การเพิ่มประสิทธิภาพกลยุทธ์การซื้อขายเงินตราต่างประเทศด้วยตัวชี้วัดทางเทคนิคและการเรียนรู้ของเครื่อง

ENHANCING FOREX TRADING STRATEGY WITH TECHNICAL INDICATORS AND MACHINE LEARNING

ธชณัฏฐ์ ตันพิทักษ์สิทธิ์

THACHANAT TANPITAKSIT

การค้นคว้าอิสระนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาวิทยาศาสตร์มหาบัณฑิต สาขาวิชาวิทยาการข้อมูลและการวิเคราะห์
ศูนย์วิเคราะห์ข้อมูลดิจิทัลอัจฉริยะพระจอมเกล้าลาดกระบัง
คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
พ.ศ. 2566

KMITL-2023-SC-M-017-027

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ENHANCING FOREX TRADING STRATEGY WITH TECHNICAL INDICATORS AND MACHINE LEARNING

THACHANAT TANPITAKSIT

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-017-027

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 การเพิ่มประสิทธิภาพกลยุทธ์การซื้อขายเงินตราต่างประเทศด้วยตัวชี้วัดทางเทคนิคและการเรียนรู้ของเครื่อง
ชื่อนักศึกษา นายธชณัฏฐ์ ตันพิทักษ์สิทธิ์
รหัสประจำตัว 64605047
ปริญญา วิทยาศาสตร์มหาบัณฑิต (วิทยาการข้อมูลและการวิเคราะห์)
ศูนย์วิเคราะห์ข้อมูลดิจิทัลอัจฉริยะพระจอมเกล้าลาดกระบัง
พ.ศ. 2566
อาจารย์ที่ปรึกษาการค้นคว้าอิสระ ผู้ช่วยศาสตราจารย์ ดร.ปัทมา เจริญพร

บทคัดย่อ

การซื้อขายแลกเปลี่ยนเงินตราต่างประเทศหรือฟอเร็กซ์ (Foreign Exchange, Forex) เป็นการลงทนในรูปแบบของการเก็งกำไรที่มีสภาพคล่องสูง (High-Liquidity) ซึ่งดึงดูดให้นักลงทุนจำนวนมากเข้ามาลงทุนและทำให้ตลาดแห่งนี้มีมูลค่าการลงทุนมากกว่าห้าล้านล้านดอลลาร์สหรัฐต่อวัน ส่งผลให้ฟอเร็กซ์เป็นตลาดทางการเงินที่มีขนาดใหญ่ที่สุดในโลก ในปัจจุบันทางโบรกเกอร์ (Broker) ได้นำเอาข้อมูลราคาอ้างอิงของสกุลคู่เงินต่างๆ ในตลาดฟอเร็กซ์มาใช้เป็นช่องทางในการลงทุน ในอดีตนักลงทุนนิยมใช้ตัวชี้วัดทางเทคนิค (Technical Indicators) ร่วมกับทฤษฎีการวิเคราะห์ความเคลื่อนไหวของราคา (Price Action) เป็นตัวช่วยในการตัดสินใจสำหรับการลงทุนแบบเกร็งกำไรทั้งสองรูปแบบ แต่เนื่องจากค่าของตัวชี้วัดทางเทคนิคเหล่านั้นมีการเปลี่ยนแปลงอยู่ตลอดเวลาทำให้การคาดการณ์ในด้านทิศทางการเคลื่อนไหวของราคามีความแม่นยำต่ำ ยากต่อการทำกำไร เป็นสาเหตุให้เกิดการคิดค้นและพัฒนาแบบจำลองสำหรับการทำนายทิศทางการเคลื่อนไหวของราคาโดยใช้วิธีการทางด้านปัญญาประดิษฐ์ คือ ทฤษฎีการเรียนรู้ของเครื่อง (Machine Learning) ในงานวิจัยนี้ได้ศึกษา 3 อัลกอริทึม คือ การสุ่มป่าไม้ (Random Forest), เครื่องเวกเตอร์สนับสนุน (Support Vector Machine) และการถดถอยโลจิสติก (Logistic Regression) โดยศึกษากับคู่เงิน EURUSD ที่กรอบเวลา 1 นาที อีกทั้งยังได้เพิ่มตัวชี้วัดทางเทคนิคแก่ข้อมูลที่ใช้ในการเรียนรู้ของแบบจำลอง และปรับจูนพารามิเตอร์เพื่อให้แบบจำลองมีประสิทธิภาพดีขึ้น สุดท้ายนำผลลัพธ์ที่ทำนายจากแบบจำลองไปทดสอบการเทรดแบบย้อนกลับ (Backward test) เพื่อหาแบบจำลองที่ให้ผลกำไรสูงที่สุด

คำสำคัญ : การเรียนรู้ของเครื่อง, การทดสอบแบบย้อนกลับ, ตัวชี้วัดทางเทคนิค, ฟอเร็กซ์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Independent Study Title	Enhancing Forex Trading Strategy with Technical Indicators and Machine Learning
Student Name	Thachanat Tanpitaksit
Student ID	64605047
Degree	Master of Science (Data Science and Analytics)
	KMITL-Digital Analytics and Intelligence Center
Year	2023
Independent Study Advisor	Asst. Prof. Dr. Pattama Charoenporn

Abstract

Trading in foreign exchange, also known as Forex (Foreign Exchange), is an investment strategy that capitalizes on high liquidity. This attracts a significant number of investors, resulting in a daily trading volume exceeding five trillion US dollars. As a result, Forex is the largest financial market in the world. Brokers in the present day utilize price data of various currency pairs in the Forex market as a means of investment. In the past, investors commonly used technical indicators in conjunction with price action analysis to make informed investment decisions in both profit-making strategies. However, the constant fluctuations of these technical indicators make it challenging to accurately predict price movements, resulting in difficulties in generating profits. To overcome this challenge, researchers have developed and improved models for predicting price movements using machine learning techniques. In this study, three algorithms were explored: Random Forest, Support Vector Machine, and Logistic Regression. The research focused on the EURUSD currency pair with a one-minute time frame. Additionally, technical indicators were incorporated into the learning process of the model, and parameters were adjusted to enhance its performance. Finally, the predicted results from the model were tested through a backward test to identify the model with the highest profitability.

Keywords : Backward test, Forex, Machine Learning, Technical Indicator

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

การค้นคว้าอิสระฉบับนี้สำเร็จลุล่วงได้ด้วยความกรุณาและความช่วยเหลือเป็นอย่างดียิ่งจากอาจารย์ที่ปรึกษาการค้นคว้าอิสระ อาจารย์ ผู้ช่วยศาสตราจารย์ ดร.ปัทมา เจริญพร ที่กรุณาเสียสละเวลาให้คำปรึกษา ชี้แนะวิธีการดำเนินงานที่เกี่ยวข้องกับการจัดทำการค้นคว้าอิสระนี้ ผู้วิจัยรู้สึกซาบซึ้งในความกรุณาของอาจารย์ และขอกราบขอบพระคุณเป็นอย่างสูงไว้ ณ ที่นี้

ขอขอบพระคุณท่านคณาจารย์และวิทยากรทุกท่านที่มาถ่ายทอดความรู้ตลอดระยะเวลา 2 ปี ทำให้สามารถนำความรู้ต่างๆ ที่ได้เรียนมาใช้ประโยชน์ในงานวิจัยนี้ได้อย่างเต็มที่ และความร่วมมือต่างๆ ของหลายท่าน ที่ให้การสนับสนุนผู้วิจัยตั้งแต่เริ่มต้นจนเสร็จสมบูรณ์ พร้อมทั้งขอบคุณเพื่อนๆ ทุกคนที่ให้ความช่วยเหลือทั้งในเรื่องการเรียน งานกลุ่ม การสอบ และคำแนะนำด้านเทคนิคการเขียนโปรแกรม และช่วยหาข้อมูลที่จะนำมาใช้ในการค้นคว้าอิสระให้สำเร็จลุล่วงไปด้วยดี

ขอขอบพระคุณท่านประธานและผู้บริหารบริษัท โมเดอร์น ไดสตัฟส์ แอนด์ พิคเมนท์ส จำกัด ที่ให้ความอนุเคราะห์แก่ข้าพเจ้าได้มีโอกาสศึกษาต่อในหลักสูตรนี้ ข้าพเจ้าหวังเป็นอย่างยิ่งที่จะนำความรู้ความสามารถที่มีต่อยอดและพัฒนาให้เกิดประโยชน์แก่องค์กร

สุดท้ายนี้ ผู้วิจัยขอกราบขอบพระคุณครอบครัว บิดา และมารดา ที่ให้การเลี้ยงดู สนับสนุนทุกอย่าง และอบรมส่งเสริมด้านการศึกษาเป็นอย่างดีตลอดมา

นายธชณัฏฐ์ ตันพิทักษ์สิทธิ์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

	หน้า
บทคัดย่อภาษาไทย	ก
บทคัดย่อภาษาอังกฤษ	ข
กิตติกรรมประกาศ	ค
สารบัญ	ง
สารบัญตาราง	ช
สารบัญรูป	ซ
บทที่ 1 บทนำ	1
1.1 ความเป็นมาและความสำคัญของปัญหา	1
1.2 วัตถุประสงค์ของงานวิจัย	2
1.3 ขอบเขตของงานวิจัย	3
1.4 ประโยชน์ที่คาดว่าจะได้รับ	3
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง	5
2.1 ข้อมูลอนุกรมเวลา	5
2.1.1 ลักษณะของข้อมูลแบบอนุกรมเวลา	5
2.1.2 ข้อมูลย้อนหลังของการแลกเปลี่ยนเงินตราต่างประเทศ	6
2.1.3 การลงทุนในตลาดการแลกเปลี่ยนเงินตราต่างประเทศ	7
2.2 การทำนายข้อมูลทางการเงินแบบอนุกรมเวลา	9
2.3 ตัวชี้วัดทางเทคนิค	9
2.3.1 ตัวชี้วัดทางเทคนิคโมเมนตัม	9
2.3.2 ตัวชี้วัดทางเทคนิคแบบความผันผวน	10
2.3.3 ตัวชี้วัดทางเทคนิคแบบแนวโน้มหรือทิศทาง	10
2.3.4 ตัวชี้วัดทางเทคนิคแบบปริมาณ	10
2.4 แบบจำลองการทำนายโดยใช้หลักการเรียนรู้ของเครื่อง	10
2.4.1 เครื่องเวกเตอร์สนับสนุน	11
2.4.2 การสุ่มป่าไม้	14
2.4.3 การถดถอยโลจิสติก	15
2.5 มาตรวัดประสิทธิภาพของแบบจำลอง	18
2.5.1 ความแม่นยำโดยรวม	18
2.5.2 ความแม่นยำที่สนใจเฉพาะส่วนที่ทำนาย	19
2.5.3 ค่าระลึกหรือความครอบคลุมของการทำนาย	19

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

หน้า

2.5.4 ค่าเฉลี่ยระหว่าง Precision และ Recall 19
2.6 งานวิจัยที่เกี่ยวข้อง 21
บทที่ 3 วิธีการดำเนินงานวิจัย 21
3.1 การเตรียมข้อมูล และการกำหนดคำนิยามให้กับข้อมูล 21
3.1.1 การเตรียมข้อมูล 22
3.1.2 การกำหนดคำนิยามให้กับข้อมูล 22
3.2 การพัฒนาเครื่องมือสำหรับการจำแนกประเภทแนวโน้มการเคลื่อนไหวของราคาอ้างอิง 22
3.2.1 ใช้แบบจำลองการสุ่มป่าไม้ 23
3.2.2 ใช้แบบจำลองเครื่องเวกเตอร์สนับสนุน 23
3.2.3 ใช้แบบจำลองการถดถอยโลจิสติก 23
3.3 การพัฒนาเครื่องมือสำหรับการจำแนกประเภทแนวโน้มการเคลื่อนไหวของราคาอ้างอิงโดยใช้หลักการวิเคราะห์ทางเทคนิค 23
3.4 การวัดประสิทธิภาพด้านการเงินของแบบจำลองในการซื้อขายแบบ Backward test 24
3.5 เครื่องมือที่ใช้สำหรับการวิจัย 25
บทที่ 4 ผลการวิจัยและการอภิปรายผล 26
4.1 ผลการเตรียมข้อมูล 26
4.1.1 การเตรียมข้อมูลพื้นฐาน 26
4.1.2 การแบ่งข้อมูลฝึกฝนและข้อมูลทดสอบ 27
4.1.3 กำหนดคลาสเป้าหมายให้กับข้อมูล 27
4.1.4 ปรับปรุงข้อมูลพื้นฐานด้วยรายการ Technical Indicator 28
4.2 ผลการทดสอบประสิทธิภาพแบบจำลองพื้นฐาน 31
4.2.1 แบบจำลองพื้นฐานการสุ่มป่าไม้ 32
4.2.2 แบบจำลองพื้นฐานเครื่องเวกเตอร์สนับสนุน 33
4.2.3 แบบจำลองพื้นฐานการถดถอยโลจิสติก 34
4.3 ผลการทดสอบประสิทธิภาพแบบจำลองร่วมกับหลักการวิเคราะห์ทางเทคนิค 36
4.3.1 แบบจำลองการสุ่มป่าไม้ร่วมกับหลักการวิเคราะห์ทางเทคนิค 36
4.3.2 แบบจำลองเครื่องเวกเตอร์สนับสนุนร่วมกับหลักการวิเคราะห์ทางเทคนิค 37

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

	หน้า
4.3.3 แบบจำลองการถดถอยโลจิสติกร่วมกับหลักการวิเคราะห์ทางเทคนิค	38
4.4 ผลการทดสอบประสิทธิภาพแบบจำลองโดยปรับพารามิเตอร์ที่เหมาะสม	40
4.4.1 แบบจำลองการสุ่มป่าไม้ร่วมกับหลักการวิเคราะห์ทางเทคนิค	40
4.4.2 แบบจำลองเครื่องเวกเตอร์สนับสนุนร่วมกับหลักการวิเคราะห์ทางเทคนิค	42
4.4.3 แบบจำลองการถดถอยโลจิสติกร่วมกับหลักการวิเคราะห์ทางเทคนิค	43
4.5 ผลการทดสอบแบบย้อนกลับ	44
4.6 อภิปรายผลการวิจัย	46
บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ	48
5.1 สรุปผลการวิจัย	48
5.2 ข้อเสนอแนะ	49
เอกสารอ้างอิง	50
ภาคผนวก	52
ประวัติผู้เขียน	59

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

ตารางที่	หน้า
2.1 ตัวอย่างข้อมูลย้อนหลังของการแลกเปลี่ยนเงินตราต่างประเทศของคู่เงิน EURUSD	7
2.2 ลำดับของจำนวนหลักของหน่วยต่างๆ แยกตามประเภทของบัญชีเทรด	8
2.3 ตาราง Confusion Matrix	18
3.1 แสดงรายการหลักการวิเคราะห์ทางเทคนิคและการคำนวณ	24
4.1 สรุปผลการทดสอบประสิทธิภาพแบบจำลองพื้นฐานการสุ่มป่าไม้	33
4.2 สรุปผลการทดสอบประสิทธิภาพแบบจำลองพื้นฐานเครื่องเวกเตอร์สนับสนุน	34
4.3 สรุปผลการทดสอบประสิทธิภาพแบบจำลองพื้นฐานการถดถอยโลจิสติก	35
4.4 สรุปผลการทดสอบประสิทธิภาพแบบจำลองพื้นฐานการสุ่มป่าไม้ ร่วมกับหลักการวิเคราะห์ทางเทคนิค	37
4.5 สรุปผลการทดสอบประสิทธิภาพแบบจำลองพื้นฐานเครื่องเวกเตอร์สนับสนุน ร่วมกับหลักการวิเคราะห์ทางเทคนิค	38
4.6 สรุปผลการทดสอบประสิทธิภาพแบบจำลองพื้นฐานการถดถอยโลจิสติก ร่วมกับหลักการวิเคราะห์ทางเทคนิค	40
4.7 สรุปผลการทดสอบประสิทธิภาพแบบจำลองการสุ่มป่าไม้ผ่านการปรับพารามิเตอร์	41
4.8 สรุปผลการทดสอบประสิทธิภาพแบบจำลองเครื่องเวกเตอร์สนับสนุนผ่านการปรับพารามิเตอร์	43
4.9 สรุปผลการทดสอบประสิทธิภาพแบบจำลองการถดถอยโลจิสติกผ่านการปรับพารามิเตอร์	44
4.10 สรุปผลลัพธ์การเทรดของการทดสอบแบบย้อนกลับในปี พ.ศ.2565	46

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป

รูปที่	หน้า
2.1 ตัวอย่างกราฟเส้นที่แสดงชุดข้อมูลแบบอนุกรมเวลาของการซื้อขายแลกเปลี่ยนเงินตราต่างประเทศสำหรับคู่เงิน EURUSD	6
2.2 การแบ่งแยกข้อมูลแบบเชิงเส้น (ก) และไม่เป็นเชิงเส้น (ข)	11
2.3 ตัวอย่างการแบ่งข้อมูลออกเป็น 2 กลุ่มของ SVM โดยใช้เส้น Maximum Margin Hyperplane	12
2.4 หลักการทำงานของฟังก์ชัน Kernel	13
2.5 หลักการทำงานของอัลกอริทึมแบบ Random Forest Classifier และเทคนิคการจำแนกประเภทข้อมูลแบบ Bagging (Bootstrap Aggregation)	14
2.6 ตัวอย่างการทำนายด้วยการถดถอยโลจิสติก	15
3.1 ตัวอย่างข้อมูลย้อนหลังการซื้อขายเงินตราต่างประเทศของคู่เงิน EURUSD	21
3.2 ตัวอย่างข้อมูลการกำหนดคลาสเป้าหมายของชุดข้อมูล	22
4.1 ข้อมูลที่นำเข้า Google Colab จากไฟล์ CSV	26
4.2 แผนภูมิแท่งแสดงข้อมูลที่เป็นค่าว่าง (Missing value)	27
4.3 ข้อมูลราคาอ้างอิงและคลาสเป้าหมาย	28
4.4 สัดส่วนคลาสเป้าหมาย Upward และ Downward	28
4.5 Correlation ระหว่างตัวแปรต้นกับคลาสเป้าหมาย	29
4.6 ข้อมูล Moving Average ที่ 10, 30, 200 นาที	29
4.7 ข้อมูล Exponential Moving Average ที่ 10, 30, 200 นาที	29
4.8 ข้อมูล Relative Strength Index ที่ 10, 30, 200 นาที	30
4.9 ข้อมูล Stochastic Oscillator (Fast) ที่ 10, 30, 200 นาที	30
4.10 ข้อมูล Stochastic Oscillator (Slow) ที่ 10, 30, 200 นาที	30
4.11 แผนภูมิแท่งแสดงเปอร์เซ็นค่าว่าง (Missing value) ของข้อมูลที่ใช้เรียนรู้	31
4.12 Correlation ระหว่างตัวแปรต้นที่เพิ่ม Technical Indicator กับคลาสเป้าหมาย	31
4.13 Confusion Matrix แสดงผลการทดสอบประสิทธิภาพแบบจำลองพื้นฐานการสุ่มป่าไม้	32
4.14 Confusion Matrix แสดงผลการทดสอบประสิทธิภาพแบบจำลองพื้นฐานเครื่องเวกเตอร์สนับสนุน	33
4.15 Confusion Matrix แสดงผลการทดสอบประสิทธิภาพแบบจำลองพื้นฐานการถดถอยโลจิสติก	35
4.16 Confusion Matrix แสดงผลการทดสอบประสิทธิภาพแบบจำลองการสุ่มป่าไม้ร่วมกับหลักการวิเคราะห์ทางเทคนิค	36

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป (ต่อ)

รูปที่ **หน้า**

4.17 Confusion Matrix แสดงผลการทดสอบประสิทธิภาพแบบจำลองเครื่องเวกเตอร์สนับสนุน ร่วมกับหลักการวิเคราะห์ทางเทคนิค 38

4.18 Confusion Matrix แสดงผลการทดสอบประสิทธิภาพแบบจำลองการถดถอยโลจิสติก ร่วมกับหลักการวิเคราะห์ทางเทคนิค 39

4.19 Confusion Matrix แสดงผลการทดสอบประสิทธิภาพแบบจำลองการสุ่มป่าไม้ ผ่านการปรับพารามิเตอร์ 41

4.20 Confusion Matrix แสดงผลการทดสอบประสิทธิภาพแบบจำลองเครื่องเวกเตอร์สนับสนุน ผ่านการปรับพารามิเตอร์ 42

4.21 Confusion Matrix แสดงผลการทดสอบประสิทธิภาพแบบจำลองการถดถอยโลจิสติก ผ่านการปรับพารามิเตอร์ 43

4.22 ตัวอย่างการแสดงผลการทดสอบย้อนกลับของแบบจำลองพื้นฐานการสุ่มป่าไม้ ร่วมกับตัวชี้วัดทางเทคนิค 45

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ในปัจจุบันผู้คนเริ่มมีความกังวลกับการมีรายได้ทางเดียว ทำให้คนเริ่มมีการขวนขวายหาหนทางสร้างรายได้ในรูปแบบอื่น นอกเหนือจากงานประจำ เช่น การหารายได้เสริม ขายของออนไลน์และออฟไลน์ หรือสิ่งที่คนมักจะพูดถึงคือการลงทุน การลงทุนคือการที่ใช้จ่ายเงินสดรูปแบบหนึ่งในปัจจุบัน โดยมุ่งหวังได้รับผลตอบแทนจากการใช้จ่ายนั้นในอนาคต ซึ่งผู้ลงทุนเชื่อว่าเงินสดหรือผลตอบแทนส่วนเพิ่มที่จะได้รับคืนนั้น จะสามารถชดเชยระยะเวลา อัตราเงินเฟ้อ และความเสี่ยงที่อาจเกิดขึ้นได้อย่างคุ้มค่า หรืออาจกล่าวได้ในอีกมุมหนึ่งว่าการลงทุน หมายถึงการออมเพื่อให้ได้รับผลตอบแทนที่มากขึ้น ซึ่งจะต้องยอมรับความเสี่ยงที่เพิ่มขึ้นเช่นกัน การตัดสินใจนำเงินออมมาลงทุนจึงต้องพิจารณาอย่างรอบคอบ และศึกษาหาข้อมูลที่เกี่ยวข้องเป็นอย่างดี เพื่อให้ได้รับผลตอบแทนตามที่คาดหวังและเพื่อลดความเสี่ยงที่จะเกิดขึ้นจากการลงทุน

ในตลาดการเงินปัจจุบัน มีทางเลือกสำหรับการลงทุนให้เลือกมากมาย ทั้งสินทรัพย์ทางการเงิน (Financial Assets) ประเภทพันธบัตร หุ้นกู้ หุ้นทุน และกองทุนรวมประเภทต่างๆ หรือสินทรัพย์ที่จับต้องได้ (Tangible Assets) เช่น ทองคำ ที่ดิน อาคาร เพชรนิลจินดา เครื่องประดับ เป็นต้น การมีความรู้ความเข้าใจในสินทรัพย์ที่จะลงทุนจึงมีความสำคัญมาก การลงทุนโดยขาดความรู้ หรือไม่เข้าใจในเรื่องความเสี่ยงและทางเลือกในการลงทุนที่ดีพอ ถือเป็นการลงทุนที่มีความเสี่ยงสูงที่สุด ความเสี่ยงอาจจะเกิดได้จากตัวผู้ลงทุนเอง เช่น สภาพร่างกายไม่สมบูรณ์, สติสัมปชัญญะที่ทำให้เกิด Human error ในการซื้อขาย การขาดความรู้ความเข้าใจในหลักการหรือระบบการลงทุนที่ใช้ หรือสภาพจิตใจและทัศนคติที่เป็นอุปสรรคจากหลักการลงทุนที่ดี (มนสิช, 2016) มนุษย์มีความสามารถในการตัดสินใจกรณีที่มีทางเลือกหลากหลายได้ไม่ค่อยดี เพราะไม่สามารถประเมินข้อมูลที่ใช้ประกอบการตัดสินใจเป็นปริมาณมากๆ ได้

การซื้อขายแลกเปลี่ยนเงินตราต่างประเทศหรือฟอเร็กซ์ (Foreign Exchange, Forex) เป็นตลาดที่มีการซื้อขายแลกเปลี่ยนสกุลเงินต่างๆ จากทั่วโลกและยังเป็นตลาดทางการเงินที่มีขนาดใหญ่ที่สุดในโลก (Pongsena et al., 2018) นับตั้งแต่ตลาดการลงทุนนี้ก่อตั้งขึ้นมีนักลงทุนจำนวนมากเข้ามาลงทุนในตลาดแห่งนี้ ทำให้มูลค่าการซื้อขายค่าเงินในตลาดแห่งนี้สูงมากถึงกว่าวันละ 5 ล้านล้านดอลล่าร์สหรัฐฯ (Jacque, 2013) ในตลาดฟอเร็กซ์คู่ของสกุลเงินที่ถูกนำมาซื้อขายแลกเปลี่ยนสองสกุลจะถูกกำหนดเป็นราคาอ้างอิง (Quotation) ซึ่งค่าราคาอ้างอิงนี้แสดงถึงความสัมพันธ์ระหว่างสกุลเงินหลัก (Base Currency) และสกุลเงินที่ใช้เป็นข้อมูลอ้างอิงเรียกว่าสกุลเงินที่เค้าเตอร์ (Counter Currency) คู่สกุลเงินทั้งหมดในตลาดฟอเร็กซ์ได้รับการกำหนดสัญลักษณ์อย่างเป็นระบบด้วยการเชื่อมต่อกันของรหัส ISO ของสกุลเงินหลักและสกุลเงินที่เค้าเตอร์ ตัวอย่างเช่น สัญลักษณ์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

“EURUSD” เป็นสัญลักษณ์ของคู่สกุลเงินยูโรเทียบกับสกุลเงินดอลลาร์สหรัฐ โดยที่ “EUR” แสดงถึงสกุลเงินยูโร และ “USD” แสดงถึงสกุลเงินดอลล่าร์สหรัฐ เป็นต้น

การลงทุนในตลาดฟอเร็กซ์จะเป็นการลงทุนแบบเกร็งกำไร ที่ซึ่งนักลงทุนสามารถทำกำไรจากการซื้อขายได้ไม่ว่าราคาอ้างอิงของคู่สกุลเงินที่สนใจจะมีค่าเพิ่มขึ้นหรือลดลงก็ตาม แตกต่างจากการลงทุนในตลาดหลักทรัพย์ที่นักลงทุนจะสามารถขายทำกำไรได้ก็ต่อเมื่อหลักทรัพย์ที่ลงทุนไปนั้นมีมีมูลค่าเพิ่มมากขึ้น ซึ่งลงทุนในตลาดฟอเร็กซ์นั้น นักลงทุนจะเปิดคำสั่งซื้อ (Long หรือ Buy) ราคาอ้างอิงของคู่สกุลเงินที่สนใจ หากเชื่อว่าราคาอ้างอิงของคู่สกุลเงินจะมีค่าเพิ่มมากขึ้น ในทางตรงกันข้ามนักลงทุนจะส่งคำสั่งขาย (Short หรือ Sell) คู่สกุลเงินที่สนใจ หากมั่นใจว่าราคาอ้างอิงของคู่สกุลเงินนั้นๆ จะมีค่าลดลง โดยที่การได้กำไรหรือขาดทุน ซึ่งเกิดจากคำสั่งซื้อหรือขายนั้น อธิบายได้ด้วยตัวอย่าง เช่น หากนักลงทุนกำหนดค่า Lot เท่ากับ 0.1 (ลงทุน 1 ดอลลาร์ ต่อการเคลื่อนไหวของอัตราอ้างอิง 1 จุด) ในบัญชีซื้อขายแบบ Standard ถ้านักลงทุนส่งคำสั่งซื้อราคาอ้างอิงของคู่สกุลเงิน EUR-USD ที่ 1.0500 และเมื่อราคาอ้างอิงสูงไปที่ 1.0550 หมายความว่านักลงทุนได้ทำกำไรมา 50 จุด (pips) นักลงทุนจะได้กำไร 50 ดอลลาร์ แต่ในทางกลับกันหากราคาอ้างอิงลดลงไปที่ 1.0450 หมายความว่านักลงทุนขาดทุนไปจำนวน 50 จุด ซึ่งคิดเป็นจำนวนเงิน 50 ดอลลาร์

ในปัจจุบันกล่าวได้ว่า AI และ Machine Learning เป็นเทคนิคที่ได้รับการกล่าวถึงเป็นอย่างมาก โดยที่ Machine Learning เป็นการสอนให้ระบบคอมพิวเตอร์ทำการเรียนรู้ได้ด้วยตนเอง โดยผ่านการใช้ข้อมูลที่มีอยู่ เริ่มต้นจากการเขียนโปรแกรมโดยนำเทคนิคที่ทำให้อัลกอริทึมสามารถทำงานและเรียนรู้ได้ด้วยตัวเอง เข้าใจข้อมูลและตัดสินใจวิธีการทำงาน จะไม่ได้ใช้การนำ Input ผ่าน Program แล้วหา Output เหมือนโปรแกรมทั่วๆ ไป แต่เป็นการใช้ Input และ Output เพื่อทำให้ได้ Program ที่สามารถหา Output ที่ใกล้เคียงที่สุด

ดังนั้นการลงทุนในฟอเร็กซ์ เกี่ยวข้องกับการตัดสินใจ (ซื้อ, ขาย, รอ) ที่มีความต่อเนื่องและต้องพิจารณาตัวชี้วัดประกอบการตัดสินใจในครั้งต่อๆ ไป ซึ่งเป็นลักษณะปัญหาที่เหมาะสมกับการประยุกต์เทคนิคการเรียนรู้จากเครื่อง และแนะนำการลงทุนให้กับผู้ลงทุน โดยใช้ข้อมูลอดีตในการให้คำพยากรณ์

1.2 วัตถุประสงค์ของงานวิจัย

1.2.1 เพื่อนำความรู้ด้านปัญญาประดิษฐ์มาประยุกต์ใช้ในการวิเคราะห์และทำนายข้อมูลด้านการเงิน

1.2.2 เพื่อนำเสนอแบบจำลองที่สามารถใช้เป็นตัวช่วยในการทำนายทิศทางการเคลื่อนไหวของราคาอ้างอิงของคู่สกุลเงินในตลาดการซื้อขายแลกเปลี่ยนเงินตราต่างประเทศ

1.2.3 เพื่อเปรียบเทียบการทำนายแบบจำลองระหว่างการใช้ข้อมูลดั้งเดิมจากตลาดซื้อขายแลกเปลี่ยนเงินตราต่างประเทศกับข้อมูลที่มีการเพิ่มตัวชี้วัดทางเทคนิค

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.2.4 เพื่อทดสอบประสิทธิภาพกลยุทธ์ที่ใช้ในการลงทุนด้านของการสร้างผลกำไรให้กับนักลงทุน

1.3 ขอบเขตของงานวิจัย

งานวิจัยได้นำความรู้ด้านปัญญาประดิษฐ์มาประยุกต์เพื่อพัฒนาแบบจำลองที่สามารถทำนายทิศทางของการเปลี่ยนแปลงราคาในตลาดแลกเปลี่ยนเงินตราต่างประเทศ โดยแบบจำลองที่ได้จากงานวิจัยนี้สามารถใช้เป็นเครื่องมือช่วยตัดสินใจสำหรับการซื้อขาย รวมทั้งได้นำเสนอกลยุทธ์ที่ใช้ในการลงทุน ซึ่งในส่วนของการดำเนินงานนั้น ได้มีการกำหนดขอบเขตของการวิจัยไว้ดังต่อไปนี้

1.3.1 ชุดข้อมูลที่ใช้ในการสร้างและทดสอบประสิทธิภาพของแบบจำลองในงานวิจัยนี้ คือ ข้อมูลย้อนหลังที่ประกอบด้วยราคาเปิดและราคาปิดของการซื้อขายแลกเปลี่ยนเงินตราต่างประเทศของคู่เงิน EURUSD ซึ่งเป็นคู่เงินที่นักลงทุนส่วนมากนิยมซื้อขายกันและมีมูลค่าในการซื้อขายมากกว่าร้อยละ 60 ของมูลค่าการซื้อขายทั้งหมดในตลาด

1.3.2 อัลกอริทึมหลักที่ใช้ในการพัฒนาแบบจำลองในงานวิจัยนี้ ประกอบด้วย การสุ่มป่าไม้ (Random Forest), เครื่องเวกเตอร์สนับสนุน (Support Vector Machine) และ การถดถอยโลจิสติก (Logistic Regression)

1.3.3 การทดสอบประสิทธิภาพของแบบจำลองในด้านของความถูกต้องแม่นยำในการทำนายนั้น ใช้มาตรวัดคือ Accuracy, Precision, Recall และ F-Measure

1.3.4 การทดสอบประสิทธิภาพของเครื่องมือรวมทั้งกลยุทธ์ที่ใช้ในการลงทุนในด้านการสร้างผลกำไรให้กับนักลงทุน ใช้วิธีการทดสอบแบบย้อนกลับ (Backward test)

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1.4.1. ได้นำความรู้ด้านปัญญาประดิษฐ์มาประยุกต์ใช้ในการวิเคราะห์และทำนายข้อมูลด้านการเงิน

1.4.2 ได้พัฒนาองค์ความรู้ใหม่ ทั้งในด้านการเตรียมข้อมูลที่จะถูกใช้เป็นข้อมูลนำเข้าสำหรับการสร้างแบบจำลอง และค่าพารามิเตอร์ต่างๆ ที่เหมาะสม ซึ่งเป็นสิ่งจำเป็นในกระบวนการเพิ่มประสิทธิภาพในการทำนายให้กับแบบจำลอง

1.4.3 ได้เครื่องมืออัตโนมัติที่ใช้สำหรับใช้จำแนกข้อมูลตามประเภทของแนวโน้มความเคลื่อนไหวของราคาอ้างอิงในช่วงเวลาที่กำหนด ซึ่งเครื่องมือนี้สามารถใช้จำแนกประเภทของข้อมูลเพื่อใช้ในการสร้างแบบจำลองสำหรับทำนายแนวโน้มความเคลื่อนไหวของราคาอ้างอิงในคาบเวลาถัดไปแทนการใช้ผู้เชี่ยวชาญในการจำแนก

1.4.4 ได้แบบจำลองที่สามารถใช้ทำนายทิศทางการเคลื่อนไหวของค่าเงินในตลาดการแลกเปลี่ยนเงินตราต่างประเทศได้อย่างแม่นยำ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.4.5 ได้เครื่องมือที่สามารถทำนายทิศทางของการเปลี่ยนแปลงราคาในตลาดแลกเปลี่ยนเงินตราต่างประเทศสำหรับนักลงทุนเพื่อใช้เป็นตัวช่วยการตัดสินใจในการซื้อขาย

1.4.6 ได้กลยุทธ์ที่ใช้ในการลงทุนที่เหมาะสมการลงทุนในการซื้อขาย

1.4.7 เทคนิคที่ใช้สำหรับการพัฒนาแบบจำลองในงานวิจัยนี้ สามารถนำไปประยุกต์ใช้ในการสร้างแบบจำลองสำหรับการทำนายทิศทางการเคลื่อนไหวของข้อมูลทางการเงินประเภทอื่นๆ เพื่อเพิ่มโอกาสสำหรับการลงทุนในการซื้อขายได้ ยกตัวอย่างเช่น ข้อมูลหุ้น (Stock) ราคาอ้างอิงการซื้อขายทองคำ (Gold) และข้อมูลสกุลเงินดิจิทัล (Cryptocurrency)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

เนื้อหาในบทนี้เป็นการกล่าวถึง การศึกษาในหลักการ ทฤษฎี และงานวิจัยที่เกี่ยวข้องสำหรับนำมาประยุกต์ใช้เพื่อพัฒนางานวิจัย ซึ่งประกอบด้วยเนื้อหาเกี่ยวกับข้อมูลแบบอนุกรมเวลา การทำนายข้อมูลแบบอนุกรมเวลา ตัวชี้วัดทางเทคนิค แบบจำลองสำหรับการทำนายข้อมูลอนุกรมเวลาโดยใช้หลักการเรียนรู้ของเครื่อง มาตรวัดประสิทธิภาพของแบบจำลอง และงานวิจัยที่เกี่ยวข้องตามลำดับ

2.1 ข้อมูลอนุกรมเวลา

2.1.1 ลักษณะของข้อมูลแบบอนุกรมเวลา

อนุกรมเวลา (Time Series) คือ กลุ่มของข้อมูลเชิงปริมาณที่จัดเก็บในช่วงเวลาหนึ่ง ข้อมูลอนุกรมเวลา (Time Series Data) คือ ชุดของข้อมูลที่เก็บรวบรวมตามช่วงเวลาระยะเวลาอย่างต่อเนื่องกัน ตัวอย่างเช่น ข้อมูลยอดขายสินค้าที่เก็บรวบรวมต่อเนื่องเป็นระยะเวลาหลายๆ เดือน ข้อมูลรายได้ประชาชาติปีต่างๆ ที่เก็บรวบรวมต่อเนื่องกันไปเป็นระยะเวลาหลายๆ ปี เป็นต้น ข้อมูลอนุกรมเวลาอาจอยู่ในลักษณะที่เป็นข้อมูลรายปี รายไตรมาส หรือรายเดือนก็ได้ ทั้งนี้ขึ้นอยู่กับความเหมาะสมในการนำไปใช้ประโยชน์ (Ismail et al., 2008) ตัวอย่างเช่น ดัชนีการซื้อขายในตลาดหลักทรัพย์แต่ละวัน (Stock Index) รายได้มวลรวมประชาชาติ (Gross National Income, GNI) รายรับรายจ่ายในแต่ละปีของบริษัทแห่งหนึ่ง เป็นต้น

โดยทั่วไปชุดข้อมูลแบบอนุกรมเวลาสามารถนำเสนอในรูปแบบแผนภูมิเส้นเพื่อให้ง่ายต่อการตีความหรือการวิเคราะห์ รูปที่ 2.1 แสดงตัวอย่างของข้อมูลอนุกรมเวลาของการซื้อขายเงินตราต่างประเทศสำหรับคู่เงินดอลลาร์สหรัฐอเมริกาเปรียบเทียบกับค่าเงินยูโร ชุดข้อมูลแบบอนุกรมเวลามักถูกใช้ในเชิงสถิติสำหรับการประมวลผลสัญญาณ (Signal Processing) การจดจำรูปแบบ (Pattern Recognition) การเงินทางคณิตศาสตร์ (Mathematical Finance) การพยากรณ์อากาศ การทำนายแผ่นดินไหว (Earthquake Prediction) วิศวกรรมควบคุม (Control Engineering) ดาราศาสตร์ (Astronomy) วิศวกรรมการสื่อสาร (Communications Engineering) และงานในสาขาวิทยาศาสตร์และวิศวกรรมประยุกต์ที่เกี่ยวข้องกับการวัดตามเวลา

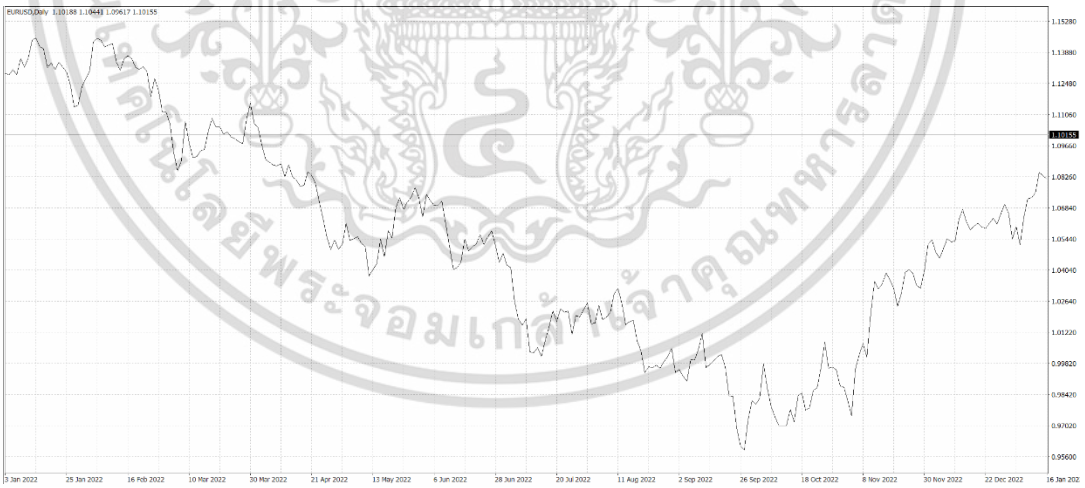
โดยธรรมชาติชุดข้อมูลแบบอนุกรมเวลาจะมีการจัดเรียงตามลำดับเวลาที่ข้อมูลนั้นๆ ถูกบันทึกไว้ สิ่งนี้ทำให้การวิเคราะห์ข้อมูลแบบอนุกรมเวลา (Time Series Data Analysis) นั้นมีความแตกต่างจากการวิเคราะห์ข้อมูลโดยทั่วไปที่ไม่ได้มีการจัดเรียงตัวตามช่วงเวลา เช่น การวิเคราะห์ข้อมูลเชิงพื้นที่ ซึ่งการสังเกตมักเกี่ยวข้องกับตำแหน่งที่ตั้งทางภูมิศาสตร์ (Geological Location) ยกตัวอย่าง เช่น การระบุราคาบ้านตามที่ตั้งและลักษณะเฉพาะของบ้าน สำหรับข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

แบบอนุกรมเวลานั้น เมื่อทำการสังเกตข้อมูลที่จัดเรียงตัวอยู่ในเวลาที่ใกล้ๆ กัน จะมีความเกี่ยวข้องกันมากกว่าข้อมูลที่จัดเรียงตัวอยู่ในเวลาที่ห่างๆ กัน นอกจากนี้การใช้ประโยชน์จากการวิเคราะห์ข้อมูลแบบอนุกรมเวลานั้นจะเป็นการวิเคราะห์แบบทางเดียว เนื่องจากบ่อยครั้งที่ข้อมูลที่เกิดขึ้นในอดีตนั้นจะสะท้อนให้เห็นแนวโน้มของข้อมูลที่จะเกิดขึ้นในอนาคต

การวิเคราะห์ข้อมูลแบบอนุกรมเวลาประกอบด้วยวิธีการสำหรับการดึงข้อมูลเชิงสถิติรวมไปถึงคุณลักษณะที่มีความหมายออกจากชุดข้อมูลนั้นๆ เพื่อตีความหรือวิเคราะห์การคาดการณ์แบบอนุกรมเวลา (Time Series Prediction) คือการใช้แบบจำลองเพื่อคาดการณ์ค่าของข้อมูลที่จะเกิดขึ้นในอนาคตโดยมีการอ้างอิงจากค่าของข้อมูลที่สังเกตก่อนหน้านี้ ในขณะที่การวิเคราะห์การถดถอย (Regression Analysis) มักใช้ในลักษณะที่จะทดสอบทฤษฎีว่าค่าปัจจุบันของชุดแบบอนุกรมเวลาอย่างน้อยหนึ่งชุด มีผลต่อค่าปัจจุบันของชุดข้อมูลแบบอนุกรมเวลาอื่นๆ สำหรับการวิเคราะห์ข้อมูลแบบอนุกรมเวลาลักษณะนี้ไม่เรียกว่า “การวิเคราะห์อนุกรมเวลา (Time Series Analysis)” ซึ่งมุ่งเน้นไปที่การเปรียบเทียบค่าของชุดข้อมูลครั้งเดียวหรือขุดเวลาหลายชุดที่ขึ้นอยู่กับเวลาที่ต่างกัน (Wei, 2006)

นอกจากนี้แบบจำลองสำหรับการวิเคราะห์ชุดข้อมูลแบบอนุกรมเวลานั้นสามารถใช้ได้กับข้อมูลที่มีค่าจริง (Real-Value) ข้อมูลที่ต่อเนื่องกัน (Continuous Data) ข้อมูลตัวเลขแบบไม่ต่อเนื่อง (Discrete Numeric Data) หรือข้อมูลสัญลักษณ์แบบไม่ต่อเนื่อง (Discrete Symbolic Data) ยกตัวอย่างเช่น ข้อมูลลำดับตัวอักษรและคำในภาษาอังกฤษ



รูปที่ 2.1 ตัวอย่างกราฟเส้นที่แสดงชุดข้อมูลแบบอนุกรมเวลาของการซื้อขายแลกเปลี่ยนเงินตราต่างประเทศสำหรับคู่เงิน EURUSD

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.1.2 ข้อมูลย้อนหลังของการแลกเปลี่ยนเงินตราต่างประเทศ

การซื้อขายเงินตราต่างประเทศ หรือสกุลเงินต่างประเทศ (Foreign Exchange, Forex) เป็นตลาดการซื้อขายล่วงหน้า (Future Markets) ที่ใหญ่ที่สุดในโลกด้วยปริมาณการซื้อขายมากกว่า 5 ล้านล้านดอลลาร์สหรัฐต่อวัน (Jacque, 2013) ตลาดการซื้อขายเงินตราต่างประเทศไม่เพียงแต่สร้างผลกำไรหรือขาดทุนสำหรับผู้ค้าเงินตราต่างประเทศ (Abbey and Doukas, 2012) แต่ยังส่งผลต่ออัตราแลกเปลี่ยนของสกุลเงินทั่วโลกอีกด้วย (Shu et al., 2015) การลงทุนตลาดการซื้อขายเงินตราต่างประเทศนั้น นักลงทุนส่วนใหญ่มักจะวิเคราะห์ข้อมูลราคาย้อนหลัง (Forex Historical Data) ของคู่เงิน (Currency Pairwise) ที่ตัวเองกำลังให้ความสนใจอยู่ เพื่อที่จะหาจุดเหมาะสมที่สุดในการส่งคำสั่งซื้อ (Buy/Long Order) หรือคำสั่งขาย (Sell/Short Order)

ข้อมูลย้อนหลังของการแลกเปลี่ยนเงินตราต่างประเทศจัดเป็นข้อมูลแบบอนุกรมเวลา เนื่องจากข้อมูลถูกจัดเก็บอย่างเป็นลำดับ (Sequence) เช่น รายนาที รายชั่วโมง รายวัน รายสัปดาห์ หรือ รายเดือน เป็นต้น ซึ่งข้อมูลที่จัดเก็บในแต่ละช่วงเวลานั้นประกอบด้วย วัน เวลา ราคาเปิด ราคาสูงที่สุดในคาบเวลานั้น ราคาต่ำที่สุดในคาบเวลานั้น และราคาปิด ดังแสดงในคารางที่ 2.1

ตารางที่ 2.1 ตัวอย่างข้อมูลย้อนหลังของการแลกเปลี่ยนเงินตราต่างประเทศของคู่เงิน EURUSD

Date	Time	Open	High	Low	Close
2022.11.10	0:00	1.00053	1.02210	0.99347	1.02077
2022.11.09	0:00	1.00707	1.00874	0.99920	1.00113
2022.11.08	0:00	1.00163	1.00956	0.99707	1.00718
2022.11.07	0:00	0.99041	1.00333	0.98995	1.00201
2022.11.04	0:00	0.97461	0.99657	0.97415	0.99570
2022.11.03	0:00	0.98187	0.98386	0.97292	0.97469
2022.11.02	0:00	0.98743	0.99752	0.98120	0.98141

2.1.3 การลงทุนในตลาดการแลกเปลี่ยนเงินตราต่างประเทศ

ในตลาดการแลกเปลี่ยนเงินตราต่างประเทศหรือฟอเร็กซ์ คู่เงินที่สามารถซื้อได้นั้นจะประกอบด้วยสกุลเงิน 2 สกุลเงิน ซึ่งจะถูกกำหนดเป็นค่าอ้างอิง (Quotation) โดยที่ค่าอ้างอิงจะเป็นการอ้างถึงความสัมพันธ์ระหว่างสกุลเงินหลัก และสกุลเงินอื่นซึ่งใช้เป็นข้อมูลอ้างอิงเรียกว่าสกุลเงินที่เคาน์เตอร์ คู่สกุลเงินทั้งหมดในตลาดฟอเร็กซ์มีการกำหนดอย่างเป็นระบบโดยการเชื่อมโยงรหัสสกุลเงิน ISO ของสกุลเงินหลักและสกุลเงินที่เคาน์เตอร์ ตัวอย่างเช่น สัญลักษณ์ “EURUSD” เป็นตัวบ่งชี้ค่าอ้างอิงของค่าเงินยูโรเปรียบเทียบกับค่าเงินดอลลาร์สหรัฐ

ยกตัวอย่างสัญลักษณ์ของคู่สกุลเงิน EURUSD ที่ถูกซื้อขายกันในตลาดฟอเร็กซ์ หมายความว่าค่าเงินยูโร (EUR) เป็นสกุลเงินหลักและมีค่าเงินดอลลาร์สหรัฐ (USD) เป็นสกุลเงิน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เคาน์เตอร์ ส่วนราคาอ้างอิงของคู่สกุลเงิน EURUSD ที่มีค่า 1.5000 หมายความว่า 1 ยูโรสามารถแลกเปลี่ยนได้ 1.5 ดอลลาร์สหรัฐ นอกจากนี้หากราคาอ้างอิงสัญลักษณ์ของคู่สกุลเงิน EURUSD เพิ่มขึ้นจาก 1.5000 เป็น 1.5100 นั้นหมายถึงมูลค่าสัมพัทธ์ของเงินยูโรเพิ่มขึ้น อาจเป็นเพราะค่าเงินยูโรแข็งค่าหรือค่าเงินดอลลาร์สหรัฐอ่อนค่าลงหรืออาจเป็นเพราะทั้งสองกรณี ในทางกลับกันหากราคา EURUSD ลดลงจาก 1.500 เป็น 1.4900 อาจเป็นเพราะค่าเงินยูโรอ่อนค่าหรือค่าเงินดอลลาร์สหรัฐแข็งค่าขึ้นหรืออาจเป็นเพราะทั้งสองกรณี เป็นต้น

การลงทุนในตลาดฟอเร็กซ์นั้น นักลงทุนจะเปิดคำสั่งซื้อ (Long หรือ Buy) ราคาอ้างอิงของคู่สกุลเงินที่พวกเขาสนใจ หากนักลงทุนเชื่อว่าราคาอ้างอิงของคู่สกุลเงินจะมีค่าเพิ่มมากขึ้น ในทางตรงกันข้ามนักลงทุนจะส่งคำสั่งขาย (Short หรือ Sell) คู่สกุลเงินที่นักลงทุนสนใจ หากนักลงทุนมั่นใจว่าราคาอ้างอิงของคู่สกุลเงินนั้นๆ จะมีค่าลดลง ซึ่งการได้กำไร หรือขาดทุน ซึ่งเกิดจากคำสั่งซื้อ หรือคำสั่งขายนั้น สามารถทำความเข้าใจได้ง่ายๆ ดังตัวอย่างต่อไปนี้

เมื่อนักลงทุนส่งคำสั่งซื้อราคาอ้างอิงของคู่สกุลเงิน EURUSD ที่ 1.0500 และราคาสูงไปที่ 1.0550 หมายความว่านักลงทุนได้ทำกำไรมา 50 จุด (pips) สำหรับค่า 1 pip ในรูปแบบเลขทศนิยมอัตราแลกเปลี่ยนของสกุลเงิน ค่าเงินอ้างอิงดอลลาร์สหรัฐ เท่ากับมูลค่าแต่ละค่า pip เท่ากับ 0.0001/1.0550 = 0.00009478 USD (ปัดเศษขึ้น) นั้นหมายความว่า นักลงทุนจะได้รับกำไรจากคำสั่งนี้ ในทางตรงกันข้ามหากราคาลดลงไปที่ 1.0450 หมายความว่านักลงทุนขาดทุนไป 50 pips ซึ่งในแต่ละ pip นักลงทุนจะขาดทุน 0.00009478 USD

สำหรับทุก pip ที่นักลงทุนทำกำไร หรือขาดทุนนั้น ในการเทรดจริงระบบซื้อขายจะถูกจัดการด้วยจำนวนหน่วยที่ใหญ่กว่า pip ที่เรียกว่าค่า lot ซึ่งหมายถึงลำดับของจำนวนหลักของหน่วยต่างๆ แยกตามประเภทของบัญชีเทรดที่นักลงทุนเปิดไว้กับตัวแทนในการเทรด (Brokers) ดังแสดงในตารางที่ 2.2

ตารางที่ 2.2 ลำดับของจำนวนหลักของหน่วยต่างๆ แยกตามประเภทของบัญชีเทรด

บัญชีเทรด	จำนวนของหน่วย (lot)
Standard	100,000
Mini	10,000
Micro	1,000

จากตารางที่ 2.2 ถ้านักลงทุนได้กำไรจากการเทรดจำนวน 50 pips หมายความว่าอัตราแลกเปลี่ยนค่าเงินอ้างอิงของสกุลเงินดอลลาร์สหรัฐเท่ากับ 1 สำหรับคู่สกุลเงิน 1 pip จะมีราคา 10 ดอลลาร์เสมอ เมื่อเราเทรด 100,000 หน่วย (1 lot ที่บัญชีเทรดแบบ Standard) ซึ่งคิดเป็นกำไรเท่ากับ 100,000*0.0001*50 = 500 USD อย่างไรก็ตามหากนักลงทุนขาดทุนจากการเทรดจำนวน 50 pips หมายความว่า นักลงทุนขาดทุนเป็นจำนวนเงิน 500 USD ด้วยเช่นกัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.2 การทำนายข้อมูลทางการเงินแบบอนุกรมเวลา

ข้อมูลราคาสินค้าโภคภัณฑ์หรือสินทรัพย์ที่มีการซื้อขายและถูกจัดเก็บในแต่ละช่วงเวลาจัดเป็นข้อมูลแบบอนุกรมเวลาซึ่งเรียกกันว่า ข้อมูลทางการเงินแบบอนุกรมเวลา (Financial Time Series Data) (Lai et al., 2009) ข้อมูลเหล่านี้ได้ถูกบันทึกและมีการศึกษาชุดข้อมูลทางการเงินที่ต่างกันเป็นเวลาหลายทศวรรษ ในปัจจุบันมีการบันทึกธุรกรรมทั้งหมดในตลาดการเงินต่างๆ เช่น ตลาดหลักทรัพย์ (Stock Market) หรือตลาดแลกเปลี่ยนเงินตราต่างประเทศ ซึ่งนำไปสู่การจัดเก็บข้อมูลจำนวนมากไม่ว่าจะเป็นข้อมูลที่เปิดเผยในอินเตอร์เน็ตหรือข้อมูลที่ใช้ในเชิงพาณิชย์

การวิเคราะห์ข้อมูลทางการเงินแบบอนุกรมเวลาเป็นเรื่องที่นักวิจัยให้ความสนใจเป็นอย่างมาก เพื่อการอนุมานและคาดการณ์ในอนาคต นอกจากนี้การจัดการความไม่แน่นอน (Uncertainty) ที่มีอยู่ในชุดข้อมูลเวลาทางการเงินและทฤษฎีที่ใช้ในการจัดการกับข้อมูลเหล่านี้เป็นเรื่องที่น่าสนใจสำหรับนักเศรษฐ์ศาสตร์ นักสถิติและนักฟิสิกส์ (Kaltwasser, 2010)

การวิเคราะห์ข้อมูลทางการเงินแบบอนุกรมเวลาเกี่ยวข้องกับทฤษฎีที่ใช้ในการประเมินมูลค่าทรัพย์สินเมื่อเวลาผ่านไปเป็นระเบียบวิธีการเชิงประจักษ์เหมือนทฤษฎีสาขาอื่นๆ ที่เป็นรากฐานในการอนุมาน (Inference) อย่างไรก็ตาม มีคุณลักษณะที่สำคัญที่ทำให้การวิเคราะห์ชุดข้อมูลทางการเงินแบบอนุกรมเวลาแตกต่างจากการวิเคราะห์ข้อมูลอนุกรมเวลาอื่นๆ ซึ่งล้วนแต่มีองค์ประกอบของความไม่แน่นอน ตัวอย่างเช่น ความผันผวนของสินทรัพย์ในตลาดหลักทรัพย์ไม่สามารถสังเกตได้โดยตรง ดังนั้นวิธีทางสถิติจึงมีบทบาทสำคัญในการวิเคราะห์ข้อมูลทางการเงินแบบอนุกรมเวลา (Tsay, 2005)

2.3 ตัวชี้วัดทางเทคนิค

ตัวชี้วัดทางเทคนิค (Technical Indicators) คือ การคำนวณทางคณิตศาสตร์ที่ขึ้นอยู่กับราคาและปริมาณการซื้อขายสินทรัพย์นั้นๆ ซึ่งนักวิเคราะห์ทางเทคนิคใช้ตัวชี้วัดเพื่อคาดการณ์การเคลื่อนไหวของราคาในอนาคตโดยการวิเคราะห์จากข้อมูลในอดีต (Lerman and Gilder, 2009) โดยส่วนใหญ่แล้วนักลงทุนในตลาดหลักทรัพย์ หรือตลาดแลกเปลี่ยนเงินตราต่างประเทศจะใช้ตัวชี้วัดทางเทคนิคในการวิเคราะห์ โมเมนตัม (Momentum) ความผันผวน (Volatility) แนวโน้ม (Trend) และปริมาณการซื้อขาย (Volume) เพื่อวิเคราะห์ความเคลื่อนไหวของราคาที่เกิดขึ้นจริงและช่วยสร้างสัญญาณซื้อขายของตัวเอง ในหัวข้อนี้อธิบายตัวชี้วัดทางเทคนิคที่ได้รับความนิยม แบ่งแยกตามหมวดหมู่ต่างๆ ดังนี้ (Trading Strategy Guide, 2019)

2.3.1 ตัวชี้วัดทางเทคนิคโมเมนตัม

โมเมนตัม คือตัวชี้วัดสำหรับการวิเคราะห์ทางเทคนิคที่ใช้งานง่ายไม่ซับซ้อน ซึ่งแสดงถึงความแตกต่างระหว่างราคาปิดของวันนี้เทียบกับราคาปิดเมื่อ N วันที่ผ่านมา โดยทั่วไปแล้วโมเมนตัมจะอ้างถึงราคาที่เคลื่อนไหวอย่างต่อเนื่องจนระบุเป็นแนวโน้มได้อย่างชัดเจน ตัวชี้วัด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โมเมนตัมแสดงถึงแนวโน้มเป็นบวกเมื่อราคาขึ้นอย่างต่อเนื่อง หรือเป็นลบในขณะที่ราคาขาลงอย่างต่อเนื่อง ในปัจจุบันมีการคิดค้นและพัฒนาตัวชี้วัดโมเมนตัมขึ้นมาอย่างมากมาย ตัวชี้วัดในหมวดหมู่นี้ที่ได้รับความนิยม ได้แก่ Relative Strength Index, Stochastic Oscillator, Williams %R, Awesome Oscillator, True strength index และ Money Flow Index เป็นต้น ซึ่งรายละเอียดของตัวชี้วัดมีดังนี้

2.3.2 ตัวชี้วัดทางเทคนิคแบบความผันผวน

ตัวบ่งชี้ทางเทคนิคประเภทความผันผวน (Volatility) มีประโยชน์อย่างมากสำหรับนักลงทุนในการตรวจสอบความผกผันของราคาสินทรัพย์ในตลาดที่อาจเกิดขึ้น ตัวบ่งชี้ประเภทนี้จะประเมินราคาสินทรัพย์จากการวัดค่าความผันผวนของราคาในตลาด แนวโน้มขาขึ้นที่แข็งแรงจะบ่งบอกถึงการลดลงของความผันผวน ในขณะที่แนวโน้มขาลงที่แข็งแรงจะบ่งบอกถึงการเพิ่มขึ้นของความผันผวน แนวโน้มของการที่จะเกิดจุดกลับตัวของราคานั้นมักจะเกิดขึ้นเมื่อความผันผวนเพิ่มขึ้น ตัวบ่งชี้ความผันผวนมีอยู่ด้วยกันหลายแบบให้นักลงทุนเลือกใช้งาน เช่น Average True Range และ Bollinger Bands

2.3.3 ตัวชี้วัดทางเทคนิคแบบแนวโน้มหรือทิศทาง

ตัวบ่งชี้ประเภทแนวโน้ม (Trend) เป็นกลุ่มของตัวชี้วัดที่ใช้วัดทิศทางหรือแนวโน้มความเคลื่อนไหวของราคาสินทรัพย์นั้นๆ ตัวชี้วัดในกลุ่มนี้มีหลากหลายแต่ตัวชี้วัดที่ได้รับความนิยมประกอบด้วย Moving Average Convergence Divergence, Average Directional Moving Index, Triple Exponential Average, Commodity Chanel Index, Detrended Price Oscillator และ Exponential Moving Average

2.3.4 ตัวชี้วัดทางเทคนิคแบบปริมาณ

ตัวชี้วัดประเภท Volume เป็นตัวชี้วัดที่ใช้สำหรับวัดว่าสินทรัพย์ทางการเงินที่ระบุมีอัตราการซื้อขายอย่างไรในช่วงเวลาที่กำหนด นักลงทุนมักใช้ข้อมูลนี้เพื่อเพิ่มผลกำไรและลดความเสี่ยงในการลงทุน ในที่งานวิจัยนี้จะยกตัวอย่างและอธิบายลักษณะของตัวชี้วัดประเภท Volume ตัวหนึ่งที่ได้รับความนิยมในหมู่นักลงทุนคือตัวชี้วัดที่มีชื่อว่า Force Index

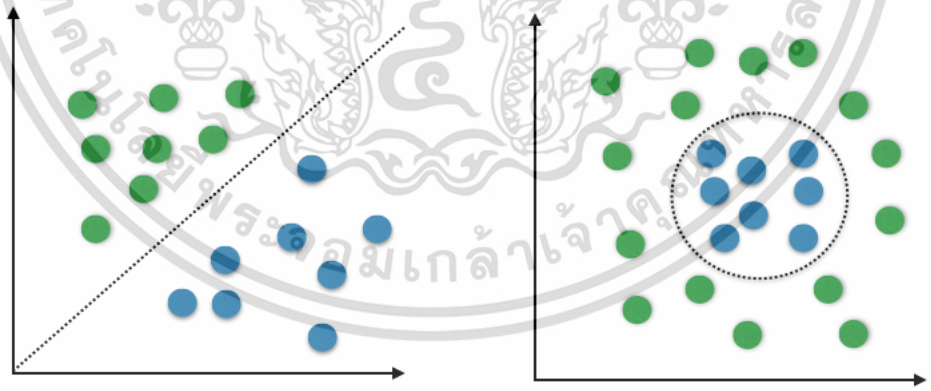
2.4 แบบจำลองการทำนายโดยใช้หลักการเรียนรู้ของเครื่อง

การพยากรณ์อนุกรมเวลาเป็นเรื่องยากเมื่อเปรียบเทียบกับการทำนายในรูปแบบอื่น เนื่องจากมีมิติเวลาเข้ามาเกี่ยวข้อง (Bontempi et al., 2012) อย่างไรก็ตาม มีแบบจำลองการเรียนรู้ของเครื่องหลายชนิดที่สามารถนำมาใช้ในการคำนวณอนุกรมเวลาได้ ยกตัวอย่าง เช่น Support Vector Machine, Random Forest และ Recurrent Neural Network ซึ่งมีรายละเอียดดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.4.1 เครื่องเวกเตอร์สนับสนุน

เครื่องเวกเตอร์สนับสนุน (Support Vector Machine, SVM) เป็นแบบจำลองการเรียนรู้แบบมีผู้สอน ใช้กับโจทย์ปัญหา 2 รูปแบบ ได้แก่ Classification และ Regression โดย SVM สามารถแก้ปัญหาได้ทั้งแบบเชิงเส้น (Linear) และไม่เป็นเชิงเส้น (Non-Linear) รวมถึงโจทย์ปัญหาชนิดอื่นๆ ได้ดี หลักการง่ายๆ ของ SVM คือ ต้องมีอัลกอลิทึมที่ใช้สร้างสมการเส้นตรงแบ่งประเภทข้อมูลที่เรียกว่าระนาบการตัดสินใจ ซึ่งใช้สำหรับโจทย์ปัญหาแบบ Classification เป้าหมายของการมีเส้นแบ่งประเภทคือ เพื่อแบ่งข้อมูลออกเป็นสองประเภท ประโยชน์ของขั้นตอนนี้คือ หลังจากมีเส้นระนาบการตัดสินใจแล้ว แบบจำลองจะสามารถเรียนรู้ได้ว่าข้อมูลชุดใหม่ที่ป้อนเข้ามาจะจัดอยู่ในประเภทใด วิธีการนี้ใช้ได้กับโจทย์ปัญหาที่ไม่ซับซ้อน เช่น ชุดข้อมูลที่มีคุณลักษณะเพียง 2 ลักษณะ แต่แท้จริงแล้ว SVM สามารถใช้ได้กับโจทย์ปัญหาที่ซับซ้อนหลายมิติได้เป็นอย่างดี (Suykens and Vandewalle, 1999) แม้แต่โจทย์ปัญหาที่มีจำนวนมิติมากกว่าจำนวนของตัวอย่าง สำหรับข้อได้เปรียบอื่นๆ ของ SVM ได้แก่ การใช้หน่วยความจำได้อย่างมีประสิทธิภาพ ทำงานได้เร็วและแม่นยำกว่าการแบ่งประเภทข้อมูลด้วยวิธีอื่น เช่น K-Nearest Neighbors หรือ Deep Neural Networks (Hsu and Lin, 2002) เครื่องเวกเตอร์สนับสนุนสามารถใช้ได้ทั้งกับข้อมูลที่เป็นข้อมูลเชิงเส้นและข้อมูลที่ไม่เป็นเชิงเส้น ความแตกต่างกันของข้อมูลทั้ง 2 คือ ข้อมูลเชิงเส้น (Linear Data) เป็นข้อมูลที่สามารถแบ่งประเภทได้ด้วยตัวจำแนกเชิงเส้น การตัดสินใจของตัวจำแนกเชิงเส้นจะขึ้นอยู่กับผลรวมเชิงเส้นของคุณลักษณะของข้อมูล โดยคุณลักษณะในที่นี้ก็คือ “Feature” ของระบบการเรียนรู้ของเครื่อง



(ก) (ข)

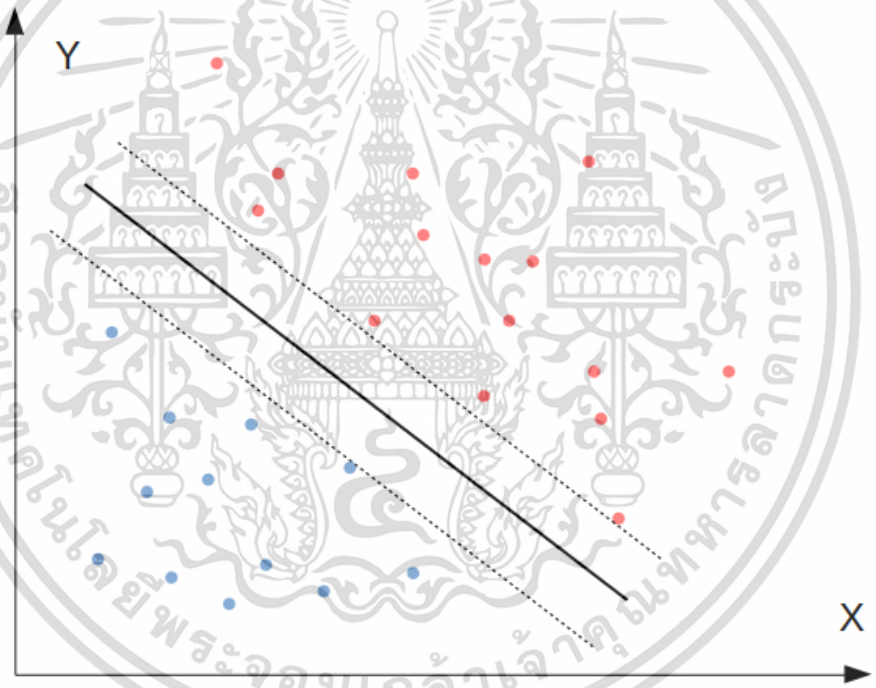
รูปที่ 2.2 การแบ่งแยกข้อมูลแบบเชิงเส้น (ก) และไม่เป็นเชิงเส้น (ข)

รูปที่ 2.2 อธิบายลักษณะของข้อมูลเชิงเส้น ที่สามารถแบ่งผลลัพธ์ของตัวอย่างที่ใช้ในการสอน (Label) ด้วยเส้นตรง (เหมือนกับที่ SVM ใช้แบ่งประเภทด้วยเส้นระนาบการตัดสินใจ) ตัวจำแนกเชิงเส้น (Linear Classifier) สามารถทำเช่นนี้ได้ด้วยการหาผลรวมเชิงเส้นของคุณลักษณะ สำหรับ SVM สามารถนำไปใช้สร้างแบบจำลองได้ แต่จะต้องใช้วิธีการจำแนกประเภทด้วยวิธีอื่น เช่น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Quadratic Classification เป็นต้น สำหรับภาพ (ข) เป็นภาพที่ใช้อธิบายข้อมูลที่ไม่เป็นเชิงเส้น ที่ไม่สามารถแบ่ง Label ได้ด้วยเส้นตรงเนื่องจากข้อมูลมีความซับซ้อนมากเกินไป ดังนั้นจึงไม่สามารถใช้วิธีเชิงเส้นในการแบ่งประเภทได้ อย่างไรก็ตาม SVM สามารถใช้ได้กับการจัดประเภทแบบเชิงเส้นและไม่เป็นเชิงเส้น ในที่นี้จะขออธิบายตัวอย่างข้อมูลเชิงเส้นอย่างง่าย จากนั้นจะอธิบายวิธีการแบ่งประเภทแบบไม่เป็นเชิงเส้นด้วย SVM และวิธีการทำงานของ SVM ด้วย Kernel

SVM ทำงานโดยการสร้างเส้นระนาบขึ้นมา 1 เส้นเพื่อแบ่งข้อมูลที่มี Label ที่แตกต่างกันออกเป็น 2 กลุ่ม เริ่มต้นจากการค้นหาระนาบที่เรียกว่า ระนาบเกินที่มีระยะห่างระหว่างข้อมูล 2 กลุ่มมากที่สุด (Maximum-Margin Hyperplane) (Noble, 2006) เนื่องจากหากเลือกเส้นตรงที่มีระยะใกล้กับข้อมูลเป้าหมายมากเกินไป หากข้อมูลชุดใหม่ที่ถูกป้อนเข้ามามีค่าคลาดเคลื่อนเพียงเล็กน้อย อาจทำให้ SVM จำแนกข้อมูลผิดกลุ่มได้



รูปที่ 2.3 ตัวอย่างการแบ่งข้อมูลออกเป็น 2 กลุ่มของ SVM โดยใช้เส้น Maximum-Margin Hyperplane

ระนาบเกิน (Hyperplane) คือ เส้นที่มีระยะห่างมากที่สุดระหว่างข้อมูล 2 กลุ่ม (รูปที่ 2.3) ซึ่งอาจเรียกว่าเป็นเส้นที่อยู่เหนือเส้นระนาบการตัดสินใจ (Decision Line) แต่ศัพท์เทคนิคทางคณิตศาสตร์เรียกว่า “ระนาบเกิน” เนื่องจากเมื่ออยู่ในมิติที่มากกว่า 2 มิติ เส้นระนาบการตัดสินใจจะไม่มีลักษณะคงรูปเส้นอีกต่อไป

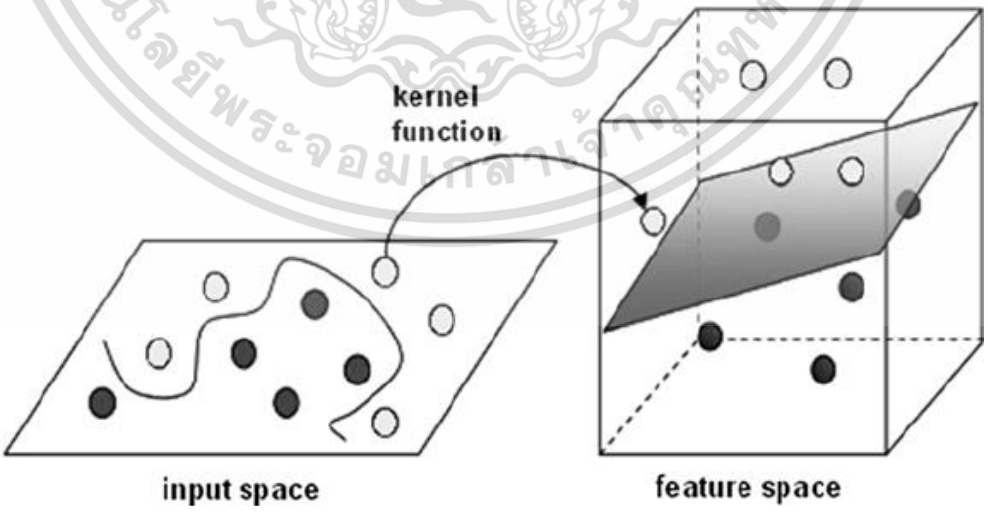
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

อัลกอริทึม SVM เรียนรู้เพื่อสร้างระนาบตัดสินใจจากชุดข้อมูลเวกเตอร์ที่เรียกว่า “ชุดข้อมูลฝึกหัด (Training Set)” โดยทุกเวกเตอร์มีค่ามิติที่ p โดยที่ p คือจำนวนคุณลักษณะ (Feature) ของชุดข้อมูลฝึกหัด ในการคำนวณหามาร์จิ้นที่โตสุดของระนาบเกินจะต้องทำให้ค่ามาร์จิ้นมีค่ามากที่สุดในทุกๆ จุดเมื่อวัดระยะจากข้อมูลเป้าหมาย ในการจำแนกข้อมูลออกเป็น 2 กลุ่ม สมมุติให้ Label ของข้อมูลทั้ง 2 กลุ่มเป็น -1 และ 1 ให้ระนาบเกิน คือกลุ่มของจุดซึ่งสามารถอธิบายได้ด้วยสมการที่ (2.3)

$$y(k) = F\left(\sum_{i=0}^{m} w_i(k) \cdot x_i(k) + b\right) \tag{2.3}$$

โดยที่ $\vec{w} * \vec{x} - b = 0$ คือ เวกเตอร์ขนาดปกติ ส่วน b คือ ค่าโน้มเอียง (Bias) และเวกเตอร์ปกติ คือ เวกเตอร์ที่ตั้งฉากกับเส้นระนาบพารามิเตอร์ในสมการที่ 2.3 ใช้คำนวณหาค่าของระนาบซึ่งเมื่อหาค่าระนาบเกินได้แล้ว จะทำให้ได้แบบจำลองสามารถจำแนกข้อมูลชุดใหม่ได้

เมื่อต้องเผชิญกับชุดข้อมูลที่ไม่เป็นเชิงเส้น การใช้ SVM ในรูปแบบปกติการคำนวณหาเส้นระนาบเกินจะไม่สามารถทำได้ ดังนั้น SVM จึงจำเป็นต้องใช้ฟังก์ชัน Kernel เข้ามาช่วย เนื่องจาก Kernel จะช่วยให้ค้นพบรูปแบบและความสัมพันธ์ที่ซ่อนอยู่ได้อย่างแม่นยำ (Scholkopf and Smola, 2001) โดยที่ฟังก์ชัน Kernel เป็นฟังก์ชันที่ใช้จับคู่ข้อมูลที่ไม่เป็นเชิงเส้นฝั่ง Input Space เข้ากับคุณลักษณะที่อยู่ในมิติที่สูงกว่า และสำหรับในกลุ่มของคุณลักษณะ (Feature Space) ซึ่งมีมิติที่สูงกว่านี้เองที่ SVM จะสามารถทำงานได้ตามปกติ ดังตัวอย่างในรูปที่ 2.4

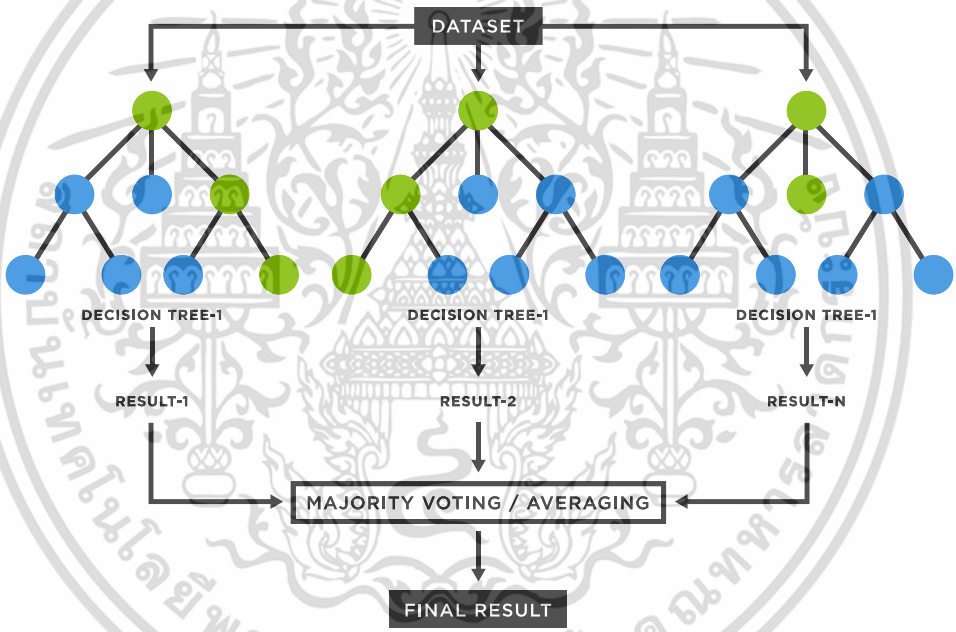


รูปที่ 2.4 หลักการทำงานของฟังก์ชัน Kernel

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.4.2 การสุ่มป่าไม้

อัลกอริทึมการสุ่มป่าไม้ (Random Forest) เป็นอัลกอริทึมที่นำแบบจำลองสำหรับการจำแนกประเภทข้อมูลหลายๆ แบบจำลองมารวมกัน (Ensemble Model) ซึ่งจะนำเอาอัลกอริทึมแบบ Decision Trees Classifier ที่ได้หลายๆ Tree มาเทรนข้อมูลร่วมกัน โดยมีหลักการนำเอาข้อมูลมาเทรนแบบจำลองหลายๆ ครั้ง บนข้อมูลชุดเดียวกัน และใช้วิธีตัดสินใจของอัลกอริทึมนั้นมาร่วมลงคะแนน หรือ โหวต (Vote) คลาสใดถูกเลือกมากที่สุดจากการสร้างอัลกอริทึมแบบ Decision Trees Classifier สำหรับกรณีการจำแนกประเภทข้อมูล หรือในกรณีของสมการเชิงเส้นกำหนดให้ทำการหาค่ากลางของข้อมูล (Mean) จากผลลัพธ์ของอัลกอริทึม Decision trees Classifier แต่ละต้น ซึ่งภายในอัลกอริทึมของต้นไม้แต่ละต้น ต้องเรียนรู้กันอย่างเป็นอิสระต่อกันมากที่สุด ดังรูปที่ 2.5



รูปที่ 2.5 หลักการทำงานของอัลกอริทึมแบบ Random Forest Classifier และเทคนิคการจำแนกประเภทข้อมูลแบบ Bagging (Bootstrap Aggregation)
(https://www.tibco.com/reference-center/what-is-a-random-forest)

โดยต้นไม้แต่ละต้นที่นำมาเทรนในแบบจำลอง จะมีตัวแปรแต่ละตัวเป็นส่วนหนึ่งของตัวแปรหรือฟีเจอร์ ซึ่งจะนำมาเทรนในรูปแบบสุ่ม (Random) และในส่วนขั้นตอนการทำนายข้อมูลนั้น จะกำหนดให้แต่ละต้นไม้ ทำนายในต้นของตัวเองและคัดเลือกผลทำนายสุดท้ายจากค่าทำนายที่ได้รับการโหวตมากที่สุด เทคนิคดังกล่าวเรียกว่า “การสุ่มตัวอย่างข้อมูล” และเทคนิคการจำแนกกระเภทข้อมูลแบบ Bagging (Bootstrap Aggregation) ดังรูปที่ 2.5 ประโยชน์ของวิธีนี้ช่วย

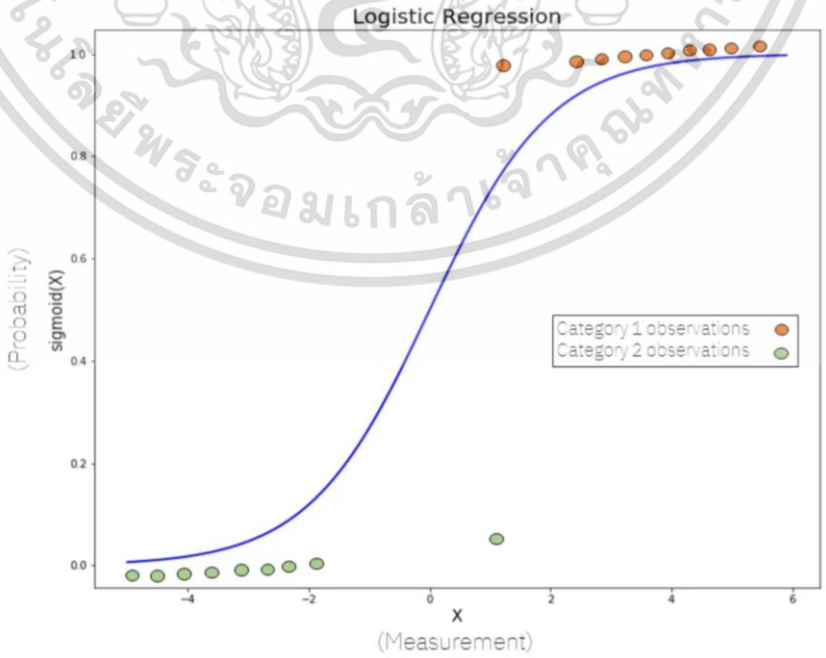
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

แก้ปัญหาแบบจำลองที่มีการตอบสนองต่อสิ่งรบกวน (Noise) มาก ทำให้เกิดตัวแปรไม่ถูกต้องเป็นผล ให้แบบจำลองที่นำมาวิเคราะห์ตัวแปรเกิดความผิดปกติของข้อมูลตัวแปร (Overfitting)

2.4.3 การถดถอยโลจีสติก

การวิเคราะห์การถดถอยโลจีสติก (Logistic Regression Analysis) เป็นเทคนิคการวิเคราะห์ตัวแปรเชิงพหุที่มีวัตถุประสงค์เพื่อประมาณค่าหรือทำนายเหตุการณ์ที่สนใจว่าจะเกิดหรือไม่เกิดเหตุการณ์นั้นภายใต้อิทธิพลของตัวปัจจัย แบบจำลองโลจีสติกประกอบด้วยตัวแปรตาม (หรือตัวแปรเกณฑ์) ที่ต้องเป็นตัวแปรแบบทวินาม (Dichotomous Variable) กล่าวคือมีได้สองค่า เช่น “เกิด” กับ “ไม่เกิด” หรือ “เสี่ยง” กับ “ไม่เสี่ยง” เป็นต้น และตัวแปรอิสระ (หรือตัวแปรทำนาย) ที่อาจมีตัวเดียวหรือหลายตัวที่เป็นได้ทั้งตัวแปรเชิงกลุ่ม (Categorical Variable) หรือตัวแปรแบบต่อเนื่อง (Continuous Variable) การวิเคราะห์การถดถอยโลจีสติกเกี่ยวข้องกับทฤษฎีความน่าจะเป็นทวินามถูกเรียกว่า Binomial Logistic Regression ถ้าตัวแปรตามเป็นพหุนามจะเรียกว่า Multinomial Logistic Regression การถดถอยโลจีสติกจัดเป็นเครื่องมือวิเคราะห์ข้อมูลในการศึกษาวิจัยที่มีวัตถุประสงค์เพื่อทำนายเหตุการณ์ หรือประเมินความเสี่ยง จึงมีการประยุกต์ใช้ในงานวิจัยหลากหลายสาขา ทั้งสาขาการแพทย์ วิศวกรรมศาสตร์ นิเวศวิทยา เศรษฐศาสตร์ และสังคมศาสตร์

การวิเคราะห์การถดถอยโลจีสติกเป็นการประมาณค่าความน่าจะเป็นของการเกิดเหตุการณ์ มีตัวแบบมาจากฟังก์ชันโลจีสติก หากมีตัวแปรอิสระเพียงตัวเดียวฟังก์ชันโลจีสติก จะแสดงความน่าจะเป็นของเหตุการณ์ที่มีดังรูปที่ 2.6



รูปที่ 2.6 ตัวอย่างการทำนายด้วยการถดถอยโลจิสติก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ความน่าจะเป็นจะมีค่าอยู่ระหว่าง 0 ถึง 1 เช่น ทิศทางการแลกเปลี่ยนเงินตราต่างประเทศจะมีแนวโน้มขาขึ้นเท่ากับ 0.25 หมายถึง ทิศทางตลาดมีโอกาสเป็นขาขึ้น 25 เปอร์เซ็นต์ แต่ในการถดถอยโลจีสติกที่ใช้ทำนายเหตุการณ์ที่น่าสนใจที่เป็นไปได้สองทาง เช่นนี้ จะใช้ค่า odds แทนความน่าจะเป็น โดย odds คืออัตราส่วนระหว่างโอกาสที่จะเกิดเหตุการณ์กับโอกาสที่จะไม่เกิดเหตุการณ์ ซึ่งจะมีค่าเป็น 0 หรือมากกว่า และให้ความหมายว่าเหตุการณ์ที่จะเกิดนั้นมีโอกาสเป็นกี่เท่าของเหตุการณ์ที่จะไม่เกิด (odds โดยทั่วไปก็คือการแสดงคู่ของตัวเลขในลักษณะอัตราต่อรอง) ดังนั้นจากตัวอย่างข้างต้นก็จะได้ว่า odds มีค่าเท่ากับ 0.25/0.75 ซึ่งเท่ากับ 0.333 เท่า (หรือ 1 ใน 3) นั่นคือโอกาสที่ทิศทางตลาดเงินตราต่างประเทศมีโอกาสที่จะเป็นขาขึ้น คือ 1/3 เท่า ในการวิเคราะห์การถดถอยโลจีสติก อัตราส่วนของโอกาสดังกล่าวแสดงสมการที่ 2.4

$$odds = \frac{P_y}{1-P_y} \tag{2.4}$$

สมการที่ 2.4 แสดงความสัมพันธ์ระหว่างตัวแปรตามและตัวแปรทำนายที่ไม่ได้อยู่ในรูปสมการเชิงเส้น จึงต้องแปลงฟังก์ชันให้อยู่ในรูปเชิงเส้น (Logistic Transformation) จะได้ Log of odds ที่เรียกว่า logit ซึ่งก็คือแบบจำลองแบบจำลองโลจิสติกที่แสดงความสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรทำนายดังสมการที่ 2.5

$$\log(odds) = ln\left(\frac{P_y}{1-P_y}\right) \tag{2.5}$$

โดยสามารถแสดงให้อยู่ในรูปสมการถดถอยเชิงเส้นได้ดังสมการที่ 2.6

$$\log(odds) = logit = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n \tag{2.6}$$

การสร้างแบบจำลองหรือสมการถดถอยโลจีสติก กระทำโดยการประมาณค่าสัมประสิทธิ์จากชุดข้อมูลสังเกตที่เก็บวัดมาได้ โดยใช้วิธีการประมาณความน่าจะเป็นสูงสุด (Maximum Likelihood Estimation; MLE) ซึ่งต่างจากการวิเคราะห์การถดถอยทั่วไปที่ใช้วิธีกำลังสองน้อยที่สุด (Least Square Method) ที่มีแนวคิดว่าต้องใช้ได้ค่าส่วนเหลือ (Residuals) มีการกระจายแบบปกติ ดังนั้นการวิเคราะห์แบบ MLE เพื่อให้ได้ตัวแบบที่เหมาะสมที่สุดจึงใช้กระบวนการวิเคราะห์แบบเวียนซ้ำ (Iterative Process) โดยเริ่มจากการประมาณค่าสัมประสิทธิ์ของตัวแปรเพื่อให้ได้สมการตั้งต้น หลังจากนั้นก็ใช้สมการเพื่อทำนายค่าแล้วนำมาคำนวณซ้ำเพื่อหาค่าสัมประสิทธิ์ใหม่ที่ทำให้ได้ความน่าจะเป็นสูงสุดเพื่อให้สามารถทำนายค่าของตัวแปรตามได้ใกล้เคียง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ค่าของข้อมูลจริงมากที่สุด จากแบบจำลอง logit แสดงให้อยู่ในรูปสมการประมาณค่าได้ดัง สมการที่ 2.7

$$\widehat{P_y} = \frac{e^{\beta_0+\beta_1+\beta_2+\ldots+\beta_n}}{1-e^{\beta_0+\beta_1+\beta_2+\ldots+\beta_n}} \quad (2.7)$$

หากทำการ antilog จะได้ odds ratio (OR) ที่มีรูปดังนี้

$$\frac{\widehat{P_y}}{1-\widehat{P_y}} = e^{\beta_0+\beta_1+\beta_2+\ldots+\beta_n} \quad (2.8)$$

ค่าสัมประสิทธิ์ (β_i) ใน linear combination ในแบบจำลองโลจีสติกนั้นมักใช้ตีความหมายที่เข้าใจได้ในเชิงทิศทางมากกว่าการบอกขนาด เพราะไม่ได้สะท้อนสิ่งที่สนใจในการวิเคราะห์นั้นๆ เท่าไรนัก โดยทั่วไปจึงใช้การตีความหมายจากค่า odds ratio เพราะให้ความหมายที่สะท้อนสิ่งที่สนใจในการวิเคราะห์ได้ดีกว่า

แบบจำลองโลจีสติกเป็นการวัดโอกาสของลอการิทึมของ odds (ล็อกของอัตราส่วนความน่าจะเป็นของเหตุการณ์) ค่าสัมประสิทธิ์ของสมการจึงมีความหมายดังนี้

ค่าสัมประสิทธิ์การถดถอย (Regression Coefficients) จะสะท้อนไปที่ค่า log odds ไม่ได้สะท้อนค่า Prob(event) โดยตรง จึงตีความหมายให้เข้าใจได้ยาก แต่สามารถบอกนัยของทิศทางอิทธิพลได้

เอ็กโปเนนเชียลของค่าสัมประสิทธิ์ (Exponential Coefficients) ในตารางผลการวิเคราะห์จะแสดงด้วยค่า Exp(B) เป็นการสะท้อนค่าของ odds ด้วย odds ratio (OR) ทำให้การตีความหมายเป็นที่เข้าใจได้ง่าย อัตราส่วน odds จะบอกให้ทราบว่า odds ของเหตุการณ์หนึ่งจะสัมพันธ์กับอีกเหตุการณ์หนึ่งอย่างไร ทำให้สามารถเปรียบเทียบชุดข้อมูลสองกลุ่มที่มีลักษณะเฉพาะบางประการหรือมีทรีตเม้นท์ต่างกันว่าจะแสดงผลต่อตัวแปรตามที่สนใจได้มากน้อยเป็นกี่เท่าของกันและกัน แสดงดังสมการที่ 2.9

$$OR = \frac{odds\ of\ a\ certain\ event}{odd\ of\ the\ another\ event} \quad (2.9)$$

ในการวิเคราะห์การถดถอยโลจีสติก จะได้สมการพยากรณ์ในรูปสมการถดถอยแบบพหุ (กรณีมีตัวแปรอิสระหลายตัว) การอธิบายตัวแปรอิสระแต่ละตัว (X_i) สามารถตีความหมายทำนองเดียวกับสมการถดถอยแบบปกติ กล่าวคือ ค่าสัมประสิทธิ์แสดงทั้งขนาดและทิศทาง (เครื่องหมาย +/-) สามารถระบุเป็นเปอร์เซ็นอิทธิพลของตัวแปรอิสระตัวนั้นๆ ที่มีต่อตัวแปรตาม เพียงแต่ต้องตระหนัก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ว่า ตัวแปรตามในที่นี้คือ logit หรือ ล็อกของอัตราส่วนระหว่างโอกาสที่จะเกิดเหตุการณ์กับโอกาสที่จะไม่เกิดเหตุการณ์ที่สนใจ ดังนั้น ถ้าต้องการทราบอิทธิพลของตัวแปรอิสระที่มีต่อตัวแปรตาม คือ โอกาสเกิดเหตุการณ์ที่สนใจ ให้พิจารณาที่ค่า e^{β_i} ซึ่งจะแสดงการเปลี่ยนแปลงของ odds โดยมีความหมายว่า เมื่อ Xi เพิ่มขึ้น 1 หน่วย จะทำให้ค่า odds เปลี่ยนแปลงไป e^{β_i} เท่าจากค่าอัตราส่วนเดิม โดยทิศทางการเปลี่ยนแปลงนั้นก็ให้พิจารณาที่เครื่องหมาย บวก-ลบ

ในการวิเคราะห์การถดถอยโลจีสติก พิจารณาที่ odds แทน Probability จึงมีความหมายในทำนองอัตราต่อรอง หรือความเป็นต่อที่จะ เกิด กับ ไม่เกิด คือต้องการคำตอบเป็น dichotomous คือมีแค่สองทางเลือก ความน่าจะเป็นมีค่าอยู่ระหว่าง 0 ถึง 1 เรามักตัดค่า (cut value) ที่ 0.5 เพื่อระบุเหตุการณ์ ดังนั้นค่า 0-0.5 ก็จะถูกตีความว่าเป็นกรณี ไม่เกิดเหตุการณ์ ถ้าเกิน 0.5 ขึ้นไปก็หมายถึงการเกิดเหตุการณ์ที่สนใจ (กาญจน์ขจร, 2018)

2.5 มาตรวัดประสิทธิภาพของแบบจำลอง

2.5.1 ความแม่นยำโดยรวม

ค่าความแม่นยำโดยรวม (Accuracy) จะเป็นตัวบ่งบอกถึงประสิทธิภาพความถูกต้องในการทำนายคลาสเป้าหมาย ซึ่งค่า Accuracy สามารถคำนวณได้จากสมการที่ 2.10

$$Accuracy = \frac{(TP+TN)}{(TP+FN+FP+TN)} \quad (2.10)$$

TP คือ จำนวนของข้อมูลที่แบบจำลองทำนายได้ตรงกับข้อมูลทดสอบซึ่งเป็นผลลัพธ์เชิงบวก

TN คือ จำนวนของข้อมูลที่แบบจำลองทำนายได้ตรงกับข้อมูลทดสอบซึ่งเป็นผลลัพธ์เชิงลบ

FP คือ จำนวนของข้อมูลที่แบบจำลองทำนายได้ผลลัพธ์เชิงลบ ซึ่งตรงข้ามกับข้อมูลทดสอบเป็นผลลัพธ์เชิงบวก

FN คือ จำนวนของข้อมูลที่แบบจำลองทำนายได้ผลลัพธ์เชิงบวก ซึ่งตรงข้ามกับข้อมูลทดสอบเป็นผลลัพธ์เชิงลบ

ค่า TP, TN, FP และ FN นั้นได้มาจากตาราง Confusion Matrix ตัวอย่างตาราง Confusion Matrix แสดงในตารางที่ 2.3

ตารางที่ 2.3 ตาราง Confusion Matrix

		Actual Data	
		Positive	Negative
Predicted Data	Positive	TP	FP
	Negative	FN	TN

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.5.2 ความแม่นยำที่สนใจเฉพาะส่วนที่ทำนาย

ความแม่นยำที่สนใจเฉพาะส่วนที่ทำนาย (Precision) คือความน่าจะเป็นที่แบบจำลองทำนายถูกต้องจากการทำนายทั้งหมด ซึ่งคำนวณได้จากสมการที่ 2.11

$$Precision = \frac{TP}{TP+FP} \tag{2.11}$$

2.5.3 ค่าระลึกหรือความครอบคลุมของการทำนาย

ค่าระลึกหรือความครอบคลุมของการทำนาย (Recall) คือความน่าจะเป็นที่แบบจำลองสามารถทำนายได้จากจำนวนข้อมูลทั้งหมดในชุดข้อมูล โดยที่ค่า Recall สามารถคำนวณได้จากสมการที่ 2.12

$$Recall = \frac{TP}{TP+FN} \tag{2.12}$$

2.5.4 ค่าเฉลี่ยระหว่าง Precision และ Recall

ค่าเฉลี่ยระหว่าง Precision และ Recall เรียกว่า F-Measure คือค่าเฉลี่ยแบบ harmonic mean ระหว่างค่า Precision และ Recall ซึ่งค่า F-Measure ถูกคิดค้นขึ้นมาเพื่อเป็น single metric ที่วัดความสามารถของแบบจำลองโดยที่ไม่ต้องเลือกระหว่างค่า Precision หรือค่า Recall เพราะเฉลี่ยให้แล้ว ค่า F-Measure สามารถคำนวณได้จากสมการที่ 2.13

$$F - Measure = 2 \times \left(\frac{Precision \times Recall}{Precision + Recall}\right) \tag{2.13}$$

2.6 งานวิจัยที่เกี่ยวข้อง

Liu และ Xiao (2009) พัฒนาระบบเทรดอัตโนมัติโดยใช้ Technical Indicator พื้นฐานร่วมกัน ได้แก่ MACD, RSI และ CCI โดยเริ่มจากแนวคิดว่าตัวชี้วัดทางเทคนิคแต่ละตัวนั้นมีจุดเด่นแตกต่างกันไป เปรียบเสมือนเซนเซอร์ที่มีความเชี่ยวชาญเฉพาะด้าน เช่น MACD จะมีความน่าเชื่อถือในการตรวจจับแนวโน้มของราคา แต่มีจุดอ่อนเรื่องสัญญาณการเทรดซื้อขาย หรือ RSI ที่ช่วยตรวจจับอัตราการซื้อขาย ณ เวลาใดเวลาหนึ่งที่มากเกินไป ซึ่งมีจุดอ่อนหากเป็นช่วงที่ราคาซื้อขายมีการเปลี่ยนแปลงอย่างรุนแรง เป็นต้น ดังนั้น ซึ่ง Liu และ Xiao ได้ทำการทดลองและวัดผลด้วยวิธี Backtesting ด้วยเงินลงทุน 10,000 ดอลลาร์สหรัฐ ณ ช่วงเวลา 2-1-2008 ถึง 31-12-2008 พบว่าหากใช้ RSI เป็นตัวชี้วัดทางเทคนิคเพียงตัวเดียว จะให้ผลลัพธ์ขาดทุนถึง -9.61% ซึ่งถ้าใช้ตัวชี้วัดร่วมกันทั้งหมด 3 ตัวแล้วจะให้ผลลัพธ์ดีที่สุด โดยได้ผลตอบแทนถึง 18.4%

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Baasher และ Fakhr (2011) วิจัยเกี่ยวกับการใช้เทคโนโลยี Machine Learning ในการทำนายแนวโน้มตลาด Forex โดยผู้วิจัยได้ทำการศึกษาและพัฒนาแบบจำลอง Machine Learning ที่สามารถทำนายแนวโน้มของตลาด Forex ได้อย่างแม่นยำ โดยผลการทดสอบบนข้อมูลจริงได้แสดงให้เห็นถึงความแม่นยำของแบบจำลองในการทำนายแบบ Binary Classification ระหว่าง Buy หรือ Sell เพื่อช่วยให้ผู้ใช้งานสามารถตัดสินใจเพื่อลงทุนให้ได้อย่างถูกต้อง เพื่อความปลอดภัยและประสิทธิภาพในการลงทุนของตัวเอง นอกจากนี้ ผู้วิจัยยังได้ใช้เทคนิคการเลือก Feature และการสกัดข้อมูลที่หลากหลาย เพื่อเพิ่มประสิทธิภาพในการทำนายและลดความผิดพลาด โดยใช้ SVM-based และ Bagging Trees เป็นเทคนิคในการเลือก Feature และใช้ MLP, RBF, และ SVM เป็น Classifier ในการสร้างแบบจำลอง Machine Learning ผลการทดลองแสดงให้เห็นว่าการใช้เทคนิค Machine Learning ในการทำนายแนวโน้มตลาด Forex มีประสิทธิภาพสูงกว่าการใช้เทคนิคอื่นๆ เช่น การวิเคราะห์เบื้องต้น (Fundamental Analysis) หรือการวิเคราะห์กราฟ (Technical Analysis) โดยมีผลลัพธ์การทำนายแบบ Binary Classification Performance ของแบบจำลอง Machine Learning ที่ได้ผลลัพธ์สูงถึง 78.6% ในขณะที่ผลลัพธ์จากการวิเคราะห์กราฟแบบเบื้องต้น (Technical Analysis) เพียง 50% เท่านั้น

Thu และ Xuan (2018) พัฒนาระบบเทรดอัตโนมัติบนพื้นฐานแบบจำลอง SVM โดยใช้ Polynomial Kernel เทียบกับการใช้ Gaussian Kernel สำหรับการเทรดคู่เงิน EUR/USD ซึ่ง Thu และ Xuan ได้ให้ข้อสรุปไว้ว่าระบบเทรดอัตโนมัติของพวกเราเป็นระบบที่มีประสิทธิภาพและมีศักยภาพที่จะสร้างผลกำไรในตลาดซื้อขายเงินตราต่างประเทศได้ ซึ่งการทดลองในงานวิจัยนี้ทำการทดสอบผ่านทางบัญชีทดลอง พบว่า Drawdown rate เท่ากับ 24.31% และอัตราผลกำไรเท่ากับ 46.3%

Chen et al., (2020) ได้ศึกษาการทำนายราคาคู่เงินเยนญี่ปุ่นกับดอลลาร์สหรัฐอเมริกาด้วยแบบจำลองการเรียนรู้ของเครื่อง พวกเขาได้เตรียมข้อมูลอัตราแลกเปลี่ยนคู่เงินดังกล่าวจาก Yahoo Finance และศึกษาทั้งหมด 3 แบบจำลองด้วยกันคือ Random Forest (RF), Support Vector Machine (SVM) และ Gradient Boosting Machine (GBM) โดยไม่ได้เพิ่ม Technical indicator ใดๆ ลงไปในข้อมูลฝึก จากผลการทดลองพบว่า SVM มีปัญหาเมื่อข้อมูลที่ Input มี Feature มากเกินไป ซึ่งได้พบลัพธ์ดีขึ้นเมื่อใช้จำนวน Feature น้อยลง และพบว่าแบบจำลอง RF กับ GBM ให้ผลลัพธ์ใกล้เคียงกันที่ Accuracy ประมาณ 75%

Parray et al., (2020) วิจัยเกี่ยวกับการใช้ Machine Learning Algorithm ในการพยากรณ์แนวโน้มตลาดหุ้น โดยผู้วิจัยได้เปรียบเทียบประสิทธิภาพของ Algorithm ต่างๆ เพื่อหาว่า Algorithm ไหนมีประสิทธิภาพสูงสุดในการพยากรณ์แนวโน้มตลาดหุ้น ดังเช่น Support Vector Machine, Perceptron และ Logistic Regression ผลการวิจัยพบว่า Support Vector Machine มีความแม่นยำในการพยากรณ์แนวโน้มตลาดหุ้นสูงที่สุด โดยมีค่าเฉลี่ยอยู่ที่ 87.35% ตามมาด้วย Logistic Regression ที่มีค่าเฉลี่ยอยู่ที่ 86.98% และ Perceptron ที่มีค่าเฉลี่ยอยู่ที่ 75.88%

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3
วิธีการดำเนินงานวิจัย

ในงานวิจัยนี้ผู้วิจัยได้แบ่งขั้นตอนในการดำเนินการวิจัยออกเป็น 4 ขั้นตอนประกอบด้วย 1) การเตรียมข้อมูล และการกำหนดคำนิยามให้กับข้อมูล 2) การพัฒนาเครื่องมือสำหรับจำแนกประเภทแนวโน้มการเคลื่อนไหวของราคาอ้างอิง 3) การพัฒนาเครื่องมือสำหรับการจำแนกประเภทแนวโน้มการเคลื่อนไหวของราคาอ้างอิงโดยใช้หลักการวิเคราะห์ทางเทคนิค 4) การวัดประสิทธิภาพด้านการเงินของแบบจำลองในการซื้อขายแบบ Backward test และ 5) เครื่องมือที่ใช้สำหรับการวิจัย

3.1 การเตรียมข้อมูล และการกำหนดคำนิยามให้กับข้อมูล

3.1.1 การเตรียมข้อมูล

ในขั้นตอนนี้จะเริ่มต้นจากการรวบรวมข้อมูลย้อนหลังของราคาอ้างอิงในการซื้อขายเงินตราต่างประเทศของคู่เงิน EURUSD ราย 1 นาที ตั้งแต่วันที่ 2 มกราคม พ.ศ.2564 ถึง 31 ธันวาคม พ.ศ.2564 ในการวิจัยนี้ ผู้วิจัยใช้วิธีการเก็บรวบรวมโดยการนำออก (Export) จากแพลตฟอร์ม Meta Trader 4 (MT4) ซึ่งเป็นแพลตฟอร์มที่ใช้ในการซื้อขายเงินตราต่างประเทศจากโบรกเกอร์ต่างๆ เช่น Exness, XM เป็นต้น โดยข้อมูลที่นำออกจากแพลตฟอร์ม MT4 นั้นจะเป็นข้อมูลที่สำคัญของกราฟแบบ OHLC ซึ่งจะอยู่ในรูปแบบไฟล์ CSV ที่ประกอบด้วย 6 คอลัมน์คือ วันที่, เวลา, ราคาเปิด, ราคาสูงสุด, ราคาต่ำสุด และราคาปิด ดังแสดงในรูปที่ 3.1

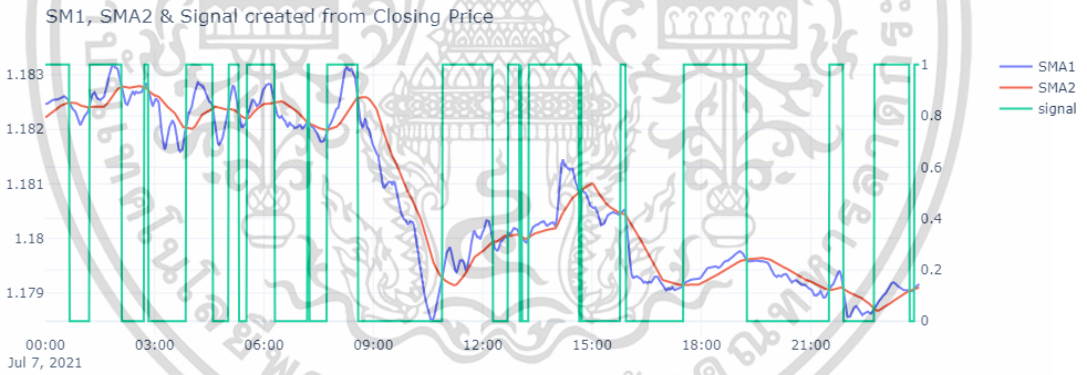
1	Date	Time	Open	High	Low	Close
2	2021.01.03	17:00	1.22396	1.22396	1.22373	1.22395
3	2021.01.03	17:01	1.22387	1.2242	1.22385	1.22395
4	2021.01.03	17:02	1.22396	1.22398	1.22382	1.22382
5	2021.01.03	17:03	1.22383	1.22396	1.22376	1.22378
6	2021.01.03	17:04	1.22378	1.22385	1.22296	1.22347
7	2021.01.03	17:05	1.22347	1.22347	1.22317	1.22337
8	2021.01.03	17:06	1.22325	1.22325	1.22312	1.22319
9	2021.01.03	17:07	1.22313	1.22313	1.22313	1.22313
10	2021.01.03	17:08	1.22298	1.2231	1.22298	1.22303
11	2021.01.03	17:09	1.22321	1.22324	1.22309	1.22322
12	2021.01.03	17:10	1.2231	1.2231	1.22309	1.2231
13	2021.01.03	17:11	1.22319	1.22319	1.22317	1.22317
14	2021.01.03	17:12	1.22315	1.22318	1.22315	1.22315
15	2021.01.03	17:15	1.22324	1.22358	1.22324	1.22358
16	2021.01.03	17:16	1.22355	1.2236	1.22352	1.22352
17	2021.01.03	17:17	1.22357	1.22361	1.22355	1.22356
18	2021.01.03	17:18	1.22359	1.22359	1.22355	1.22355
19	2021.01.03	17:19	1.22357	1.22365	1.22354	1.22356

รูปที่ 3.1 ตัวอย่างข้อมูลย้อนหลังการซื้อขายเงินตราต่างประเทศของคู่เงิน EURUSD

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.1.2 การกำหนดคำนิยามให้กับข้อมูล

การสร้างแบบจำลองสำหรับการจำแนกข้อมูลแนวโน้มของราคาในช่วงเวลาที่กำหนด โดยอาศัยการเรียนรู้ด้วยเครื่อง ข้อมูลนำเข้าจะต้องถูกกำหนดคำนิยาม (Label) ของคลาสเป้าหมาย (Target Class) เอาไว้ เพื่อให้แบบจำลองสามารถเรียนรู้ได้ตามหลักการเรียนรู้ของเครื่องแบบมีผู้แนะนำ (Supervised Learning) ซึ่งเป้าหมายของแบบจำลองสำหรับการจำแนกแนวโน้มของราคา ในช่วงเวลาที่กำหนดคือ การคัดแยกว่าข้อมูลอนุกรมชุดนี้เป็นข้อมูลที่บ่งบอกถึงเทรนหรือแนวโน้มความเคลื่อนไหวของราคาแบบใดใน 2 รูปแบบดังนี้ แนวโน้มของราคาแบบขาขึ้น (Upward Trend) และแนวโน้มของราคาแบบขาลง (Downward Trend) ดังนั้น เพื่อการกำหนดคลาสเป้าหมาย ผู้วิจัยใช้วิธีอ้างอิงจาก Daddy Trader (2019) โดยสร้างเส้น Simple Moving Average (10) และ (60) จากราคาปิด ถ้าSMA10 มากกว่า SMA60 ให้กำหนดคลาสเป้าหมายคือ Upward Trend หรือ 1 และถ้า SMA10 น้อยกว่า SMA60 ให้กำหนดคลาสเป้าหมายคือ Downward Trend หรือ 0 แสดงตัวอย่างการกำหนดคลาสเป้าหมายดังรูปที่ 3.2 หลังจากที่ระบุคลาสเป้าหมายเสร็จเรียบร้อย ข้อมูลชุดนี้จะถูกนำไปใช้เป็นข้อมูลนำเข้าสำหรับการสร้างแบบจำลองเพื่อใช้ในการเปรียบเทียบประสิทธิภาพ



รูปที่ 3.2 ตัวอย่างการกำหนดคลาสเป้าหมายของชุดข้อมูล

3.2 การพัฒนาเครื่องมือสำหรับการจำแนกประเภทแนวโน้มการเคลื่อนไหวของราคาอ้างอิง

แนวคิดของการวิจัยในขั้นตอนนี้คือ การใช้เทคโนโลยีคอมพิวเตอร์ด้านปัญญาประดิษฐ์มาสร้างเป็นเครื่องมืออัตโนมัติเพื่อใช้แทนผู้เชี่ยวชาญในกระบวนการจำแนกประเภทแนวโน้มการเคลื่อนไหวของราคาอ้างอิงในช่วงเวลาที่กำหนดและเปรียบเทียบประสิทธิภาพกับการใช้หลักการวิเคราะห์ทางเทคนิค โดยแบ่งออกเป็น 3 วิธีดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2.1 ใช้แบบจำลองการสุ่มป่าไม้

ในส่วนของการใช้แบบจำลองการสุ่มป่าไม้ คือ Random Forest Classifier โดยกำหนดค่าไฮเปอร์พารามิเตอร์ 2 วิธี คือ วิธีพื้นฐานที่แบบจำลองกำหนดค่าต่างๆ มาให้ กับวิธี GridSearchCV ที่ช่วยหาไฮเปอร์พารามิเตอร์ที่แม่นยำที่สุด โดยกำหนดช่วงไฮเปอร์พารามิเตอร์ต่างๆ ดังข้างล่าง

1) n_estimators : 50, 100, 300, 500
2) max_depth : None, 2, 4
3) min_samples_split : 2, 5
4) min_samples_leaf : 1, 2

3.2.2 ใช้แบบจำลองเครื่องเวกเตอร์สนับสนุน

ในส่วนของการใช้แบบจำลองเครื่องเวกเตอร์สนับสนุน คือ Support Vector Machine โดยกำหนดค่าไฮเปอร์พารามิเตอร์ 2 วิธี คือ วิธีพื้นฐานที่แบบจำลองกำหนดค่าต่างๆ มาให้ กับวิธี GridSearchCV ที่ช่วยหาไฮเปอร์พารามิเตอร์ที่แม่นยำที่สุด โดยกำหนดช่วงไฮเปอร์พารามิเตอร์ต่างๆ ดังข้างล่าง

1) C : 0.1, 1, 10, 100
2) Gamma : auto, scale

3.2.3 ใช้แบบจำลองการถดถอยโลจิสติก

ในส่วนของการใช้แบบจำลองการถดถอยโลจิสติก คือ Logistic Regression โดยกำหนดค่าไฮเปอร์พารามิเตอร์ 2 วิธี คือ วิธีพื้นฐานที่แบบจำลองกำหนดค่าต่างๆ มาให้ กับวิธี GridSearchCV ที่ช่วยหาไฮเปอร์พารามิเตอร์ที่แม่นยำที่สุด โดยกำหนดช่วงไฮเปอร์พารามิเตอร์ต่างๆ ดังข้างล่าง

1) Solver : newton-cg, lbfgs, liblinear
2) C : 0.01, 0.1, 1.0, 10, 100

3.3 การพัฒนาเครื่องมือสำหรับการจำแนกประเภทแนวโน้มการเคลื่อนไหวของราคาอ้างอิงโดยใช้หลักการวิเคราะห์ทางเทคนิค

หลักการวิเคราะห์ทางเทคนิคหรือ Technical Indicators เป็นการคำนวณทางคณิตศาสตร์โดยใช้ราคาในอดีตหรือปริมาณการซื้อขายหรือค่าทางสถิติ โดยอินดิเคเตอร์เป็นเครื่องมือที่นักลงทุนในปัจจุบันใช้วิเคราะห์แนวโน้มของตลาด แน่นอนว่าการใช้งานอินดิเคเตอร์จะใช้ควบคู่กับการดูกราฟของราคาเพื่อวิเคราะห์ความสัมพันธ์ระหว่างกราฟของราคากับอินดิเคเตอร์ รายการอินดิเคเตอร์ที่ใช้งาน อ้างอิงจากงานวิจัยของ (Bernardo et al., 2018) และ (Thu and Xuan 2018) ดังตารางที่ 3.1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.1 แสดงรายการหลักการวิเคราะห์ทางเทคนิคและการคำนวณ

Technical Indicators	Durations (n)	Formulas
Simple Moving Average	10,30,200	$SMA = \sum_{i=0}^{n-1} \frac{Close(i)}{n}$
Exponential Moving Average	10,30,200	$EMA = [Close(i) - SMA(i)] * \frac{2}{(n+1)} + SMA(i-1)$
Momentum	10,30,200	$MOM = Close(i)_t - Close(i)_{t-n}$
Relative Strength Index	10,30,200	$RSI = 100 - \left[\frac{100}{\left(1 + \frac{average\ positive\ price\ changes}{average\ negative\ price\ changes}\right)}\right]$
Stochastic Oscillator	10,30,200	$\%K = \frac{Close(i) - Lowest\ Low(n)}{Highest\ High(n) - Lowest\ Low(n)} * 100$ $\%D = SMA(\%K, 3)$

หลังจากเพิ่มข้อมูลอินดิเคเตอร์ต่างๆ เข้าไปใน Dataset แล้ว ให้นำไปฝึกกับแบบจำลองต่างๆ ในข้อ 3.2 อีกครั้งเพื่อเปรียบเทียบผลลัพธ์

3.4 การวัดประสิทธิภาพด้านการเงินของแบบจำลองในการซื้อขายแบบ Backward test

นอกจากแบบจำลองที่ถูกพัฒนาในงานวิจัยนี้จะถูกวัดประสิทธิภาพด้านความแม่นยำในการจำแนกและการทำนายผลแล้ว ผู้วิจัยยังได้ทดสอบในการใช้งานแบบจำลองร่วมกับการใช้กลยุทธ์การเทรดและกลยุทธ์ในการตัดขาดทุน เพื่อทดสอบประสิทธิภาพด้านการทำกำไรกับการซื้อขายอีกด้วย โดยการวัดประสิทธิภาพในด้านนี้ คือการทดสอบแบบย้อนกลับ (Backward Test) เป็นการจำลองการเทรดเพื่อวัดประสิทธิภาพในด้านการทำกำไรหรือขาดทุนของกลยุทธ์ที่นำมาใช้ในการเทรด สำหรับการทดสอบย้อนกลับที่ทำในการทดลองในงานวิจัยนี้ ได้นำข้อมูลย้อนหลังของราคาอ้างอิงอัตราแลกเปลี่ยนเงินตราต่างประเทศคู่เงิน EURUSD การจำลองการเทรดจะใช้ข้อมูลย้อนหลังเป็นระยะเวลา 1 ปี เริ่มตั้งแต่วันที่ 2 มกราคม พ.ศ. 2565 จนถึงวันที่ 30 ธันวาคม พ.ศ. 2565 มาใช้เป็นข้อมูลในการจำลองการเทรดเพื่อวัดประสิทธิภาพของแบบจำลองและกลยุทธ์ในการเทรดที่ได้นำเสนอ กำหนดเงินลงทุนเริ่มต้นคือ 1,000 ดอลลาร์

เมื่อทำการจำลองการเทรดครบ 1 ปี ผู้จัดทำทำการสรุปผลการทดลองที่ได้จากการจำลองการเทรด ซึ่งประกอบด้วย จำนวนการเทรด เปอร์เซ็นต์ผลกำไร เปอร์เซ็นต์ขาดทุนสูงสุดและจำนวนเงินคงเหลือในพอร์ต

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.5 เครื่องมือที่ใช้สำหรับการวิจัย

เครื่องมือที่ใช้ในการพัฒนางานวิจัยนี้ ประกอบด้วย

1) เครื่องคอมพิวเตอร์สำหรับพัฒนา มีรายละเอียดดังนี้

หน่วยประมวลผลกลาง : AMD Ryzen 7 5800H

หน่วยประมวลผลกราฟฟิก : NVIDIA GeForce RTX 3060 6GB

หน่วยความจำหลัก : 16 GB

หน่วยความจำรอง : 1 TB

ระบบปฏิบัติการ Windows 11 Home 64 bits

(2) ระบบโปรแกรมประยุกต์บนคลาวด์

Google Colab : ทำหน้าที่เป็นเครื่องมือเขียนโปรแกรมภาษา Python

Google Drive : ใช้ในการเก็บข้อมูลแบบชั่วคราว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4
ผลการวิจัยและการอภิปรายผล

วัตถุประสงค์หลักของงานวิจัยนี้คือการนำความรู้ทางด้านปัญญาประดิษฐ์ โดยเฉพาะเทคนิคการเรียนรู้ของเครื่อง และการเรียนรู้เชิงลึก ได้แก่ แบบจำลองการสุ่มป่าไม้, แบบจำลองเครื่องเวกเตอร์สนับสนุน และแบบจำลองการถดถอยโลจิสติก เพื่อพัฒนาเป็นแบบจำลองสำหรับทำนายทิศทางการเปลี่ยนแปลงราคาอ้างอิงในคาบเวลาถัดไปของสกุลเงินในตลาดซื้อขายแลกเปลี่ยนเงินตราต่างประเทศ ซึ่งแบบจำลองนี้จะช่วยให้นักลงทุนสามารถกำหนดทิศทางการลงทุนในแต่ละครั้งได้ นอกจากนี้ยังได้นำเสนอวิธีการเพิ่มประสิทธิภาพให้กับแบบจำลองด้วยเทคนิคต่างๆ ซึ่งจะทำให้แบบจำลองทำนายข้อมูลได้อย่างแม่นยำมากขึ้น นอกจากการวัดประสิทธิภาพด้านความแม่นยำในการทำนายแบบจำลองแล้ว งานวิจัยนี้ได้ทำการวัดประสิทธิภาพในด้านการทำกำไรของแบบจำลองต่างๆ ที่ได้นำเสนอด้วยการจำลองการเทรดโดยใช้ข้อมูลย้อนหลัง (Backward test) จากผลการทดลองทั้งหมด สามารถสรุปผลการศึกษาและวิเคราะห์ผลการศึกษาที่ได้ดังต่อไปนี้

4.1 ผลการเตรียมข้อมูล

4.1.1 การเตรียมข้อมูลพื้นฐาน

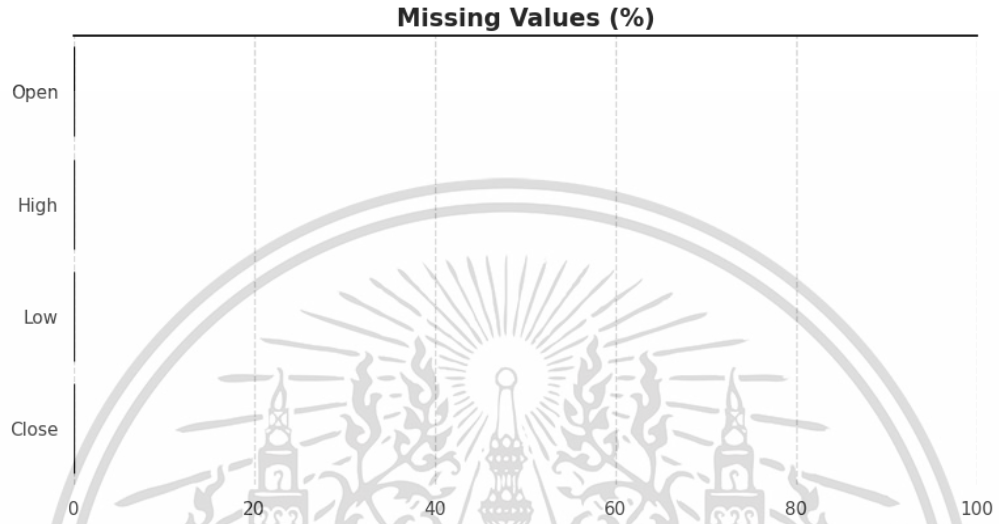
ผู้วิจัยเตรียมข้อมูลพื้นฐานโดยใช้ Library pandas ฟังก์ชัน read_csv() เพื่อทำการอ่านข้อมูลในไฟล์ CSV ที่ได้ import มาจากโปรแกรม MT4 ซึ่งได้แก่ข้อมูลย้อนหลังของราคาอ้างอิงการซื้อขายเงินตราต่างประเทศสกุลเงิน EURUSD คาบเวลา 1 นาที ตั้งแต่วันที่ 2 มกราคม พ.ศ. 2564 ถึงวันที่ 31 ธันวาคม พ.ศ. 2564 รายละเอียดดังรูปที่ 4.1

	Open	High	Low	Close	Datetime
0	1.22396	1.22396	1.22373	1.22395	2021-01-03 17:00
1	1.22387	1.22420	1.22385	1.22395	2021-01-03 17:01
2	1.22396	1.22398	1.22382	1.22382	2021-01-03 17:02
3	1.22383	1.22396	1.22376	1.22378	2021-01-03 17:03
4	1.22378	1.22385	1.22296	1.22347	2021-01-03 17:04
...	...	...	...	...	...
369497	1.13797	1.13797	1.13774	1.13774	2021-12-31 16:54
369498	1.13775	1.13781	1.13748	1.13748	2021-12-31 16:55
369499	1.13749	1.13765	1.13704	1.13707	2021-12-31 16:56
369500	1.13709	1.13740	1.13679	1.13740	2021-12-31 16:57
369501	1.13740	1.13740	1.13681	1.13686	2021-12-31 16:58

รูปที่ 4.1 ข้อมูลที่นำเข้า Google Colab จากไฟล์ CSV

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ก่อนที่จะนำเข้าข้อมูลเพื่อให้แบบจำลองเรียนรู้ต้องมีการทำความสะอาดข้อมูล (Data Cleaning) เช่นตรวจสอบประเภทของข้อมูล (Data types) และข้อมูลที่เป็นค่าว่าง (Missing value) หลังจากตรวจสอบพบว่าประเภทของข้อมูล Open High Low Close คือ Float64 และข้อมูลมีความสมบูรณ์ไม่มีข้อมูลว่างใดๆ ให้ต้องแก้ไข แสดงดังรูปที่ 4.2



รูปที่ 4.2 แผนภูมิแท่งแสดงข้อมูลที่เป็นค่าว่าง (Missing value)

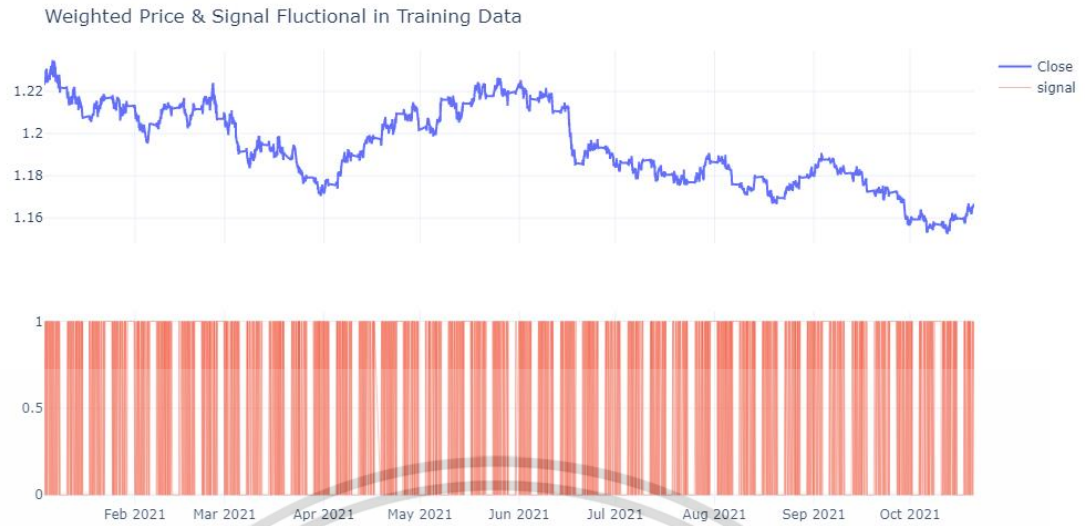
4.1.2 การแบ่งข้อมูลฝึกฝนและข้อมูลทดสอบ

ในขั้นตอนนี้ผู้วิจัยได้ทำการแบ่งข้อมูลชุดนี้ด้วยหลักการ 80/20 โดยกำหนดให้ข้อมูลที่อยู่ในช่วงเวลาตั้งแต่วันที่ 2 มกราคม พ.ศ.2564 ถึงวันที่ 21 ตุลาคม พ.ศ. 2564 เป็นข้อมูลสำหรับการเรียนรู้ของแบบจำลองทั้งหมด 295,601 แถว และข้อมูลที่อยู่ในช่วงเวลาตั้งแต่วันที่ 21 ตุลาคม พ.ศ. 2564 ถึง 31 ธันวาคม พ.ศ. 2564 เป็นข้อมูลสำหรับการทดสอบประสิทธิภาพของแบบจำลองทั้งหมด 73,901 แถว

4.1.3 กำหนดคลาสเป้าหมายให้กับข้อมูล

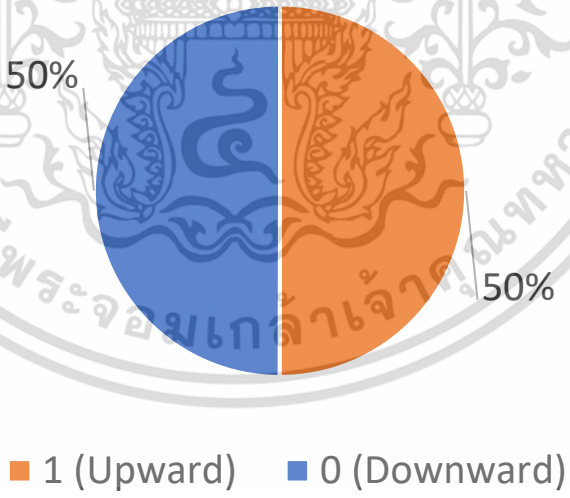
เนื่องจากงานวิจัยนี้ ผู้วิจัยศึกษาการเรียนรู้ของเครื่องแบบมีผู้แนะนำ (Supervised Learning) จึงจำเป็นต้องมีการกำหนดคลาสเป้าหมายแก่ข้อมูล (Target Class) อ้างอิงจากหัวข้อที่ 3.1.2 ให้กำหนดคลาสเป้าหมายคือ Upward Trend หรือ 1 และ Downward Trend หรือ 0 จากเส้น SMA10 และ SMA60 แสดงผลการกำหนดคลาสเป้าหมายดังรูปที่ 4.3 แสดงให้เห็นว่าข้อมูลคลาสเป้าหมายมีการกระจายตัวอยู่ทุกช่วงของข้อมูลที่ใช้เรียนรู้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.3 ข้อมูลราคาอ้างอิงและคลาสเป้าหมาย

หลังจากกำหนดคลาสเป้าหมายเรียบร้อย ต้องมีการตรวจสอบว่าข้อมูลมีปัญหา Imbalance Dataset หรือไม่ จากรูปที่ 4.4 แสดงสัดส่วนของคลาสเป้าหมายระหว่าง Upward Trend และ Downward Trend คือ 50:50 ซึ่งถือว่าข้อมูลที่ใช้ในการศึกษาครั้งนี้ไม่พบปัญหาดังกล่าว



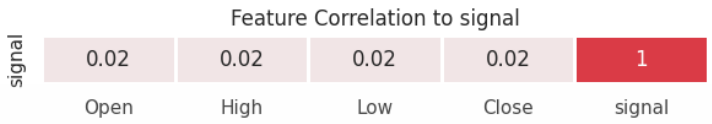
รูปที่ 4.4 สัดส่วนคลาสเป้าหมาย Upward และ Downward

4.1.4 ปรับปรุงข้อมูลพื้นฐานด้วยรายการ Technical Indicator

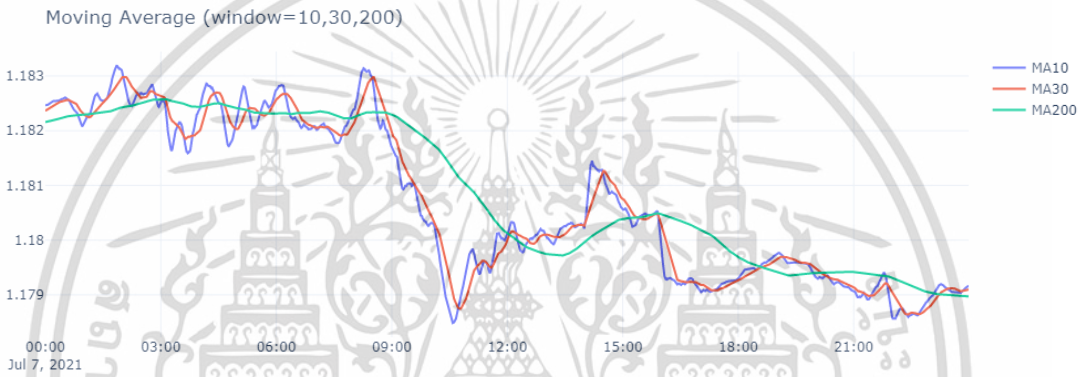
จากข้อมูลที่ใช้ในการเรียนรู้ มีข้อมูลที่เป็นตัวแปรต้น (Feature) เพียง Open, High, Low และ Close เท่านั้น ซึ่งหากดูจากค่า Correlation ดังรูปที่ 4.5 พบว่าข้อมูลดังกล่าวไม่มีความสัมพันธ์กับคลาสเป้าหมายเท่าที่ควร ดังนั้นผู้วิจัยจึงได้เพิ่ม Technical Indicator ดังข้อที่ 3.3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

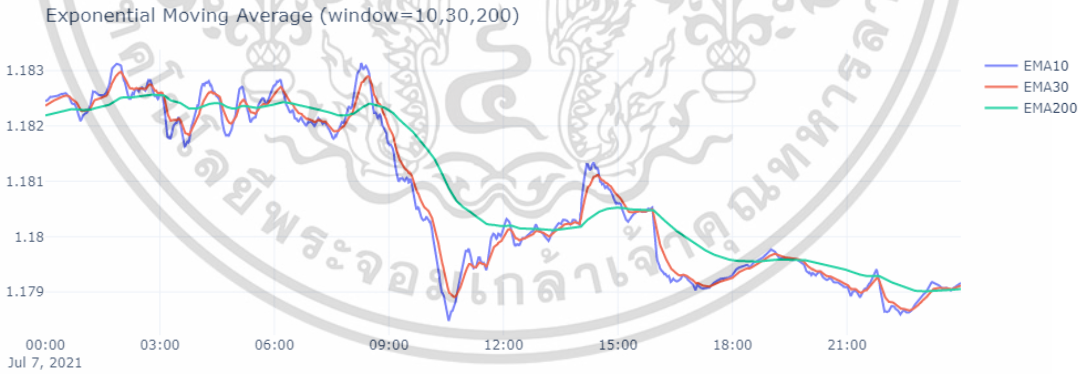
คือ Moving Average, Exponential Moving Average, Momentum, Relative Strength Index และ Stochastic Oscillator ที่ระยะสั้นเท่ากับ 10 นาที ระยะกลางเท่ากับ 30 นาที และระยะยาวเท่ากับ 200 นาที โดยรูปที่ 4.6- จะแสดงข้อมูลตัวอย่างของแต่ละ Indicator ที่ได้สร้างขึ้นเพื่อใช้กับข้อมูลเพื่อการเรียนรู้ และข้อมูลเพื่อทดสอบประสิทธิภาพ



รูปที่ 4.5 Correlation ระหว่างตัวแปรต้นกับคลาสเป้าหมาย

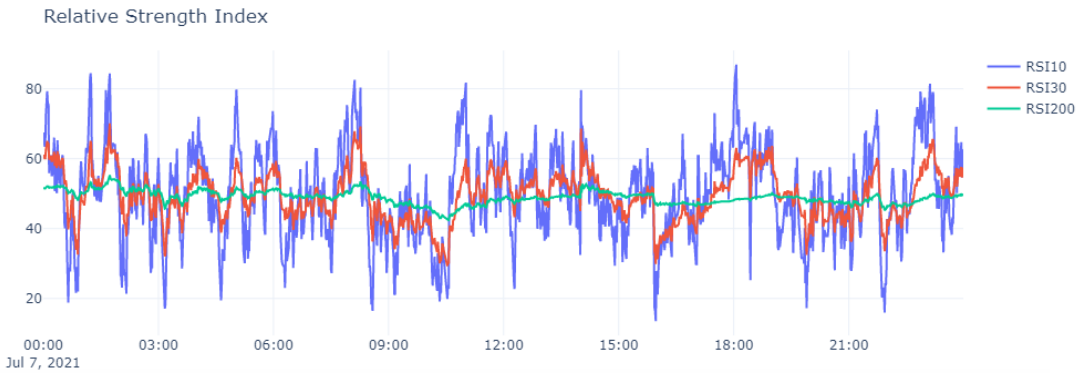


รูปที่ 4.6 ข้อมูล Moving Average ที่ 10, 30, 200 นาที

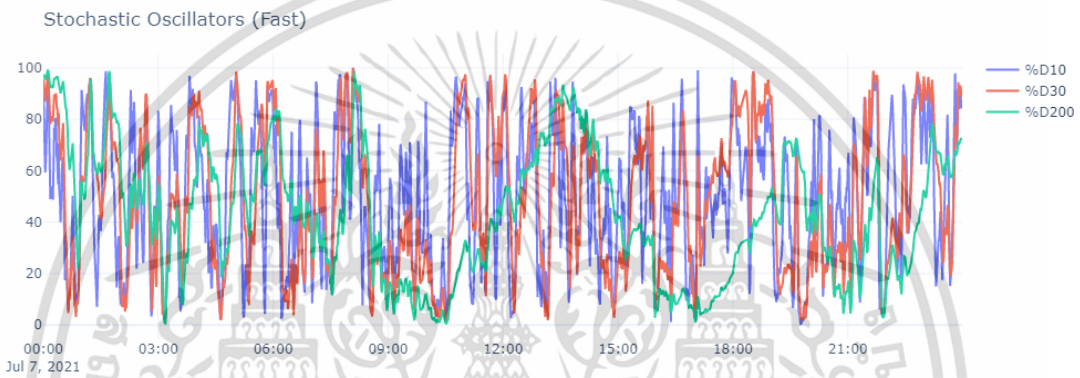


รูปที่ 4.7 ข้อมูล Exponential Moving Average ที่ 10, 30, 200 นาที

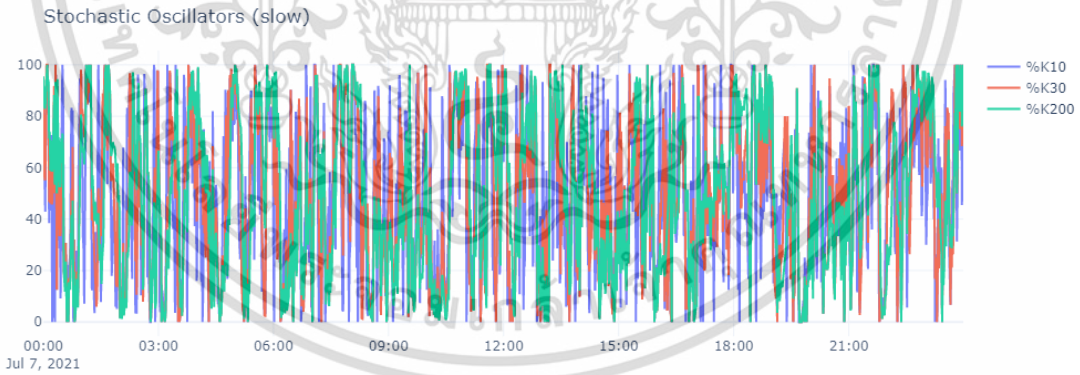
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.8 ข้อมูล Relative Strength Index ที่ 10, 30, 200 นาที



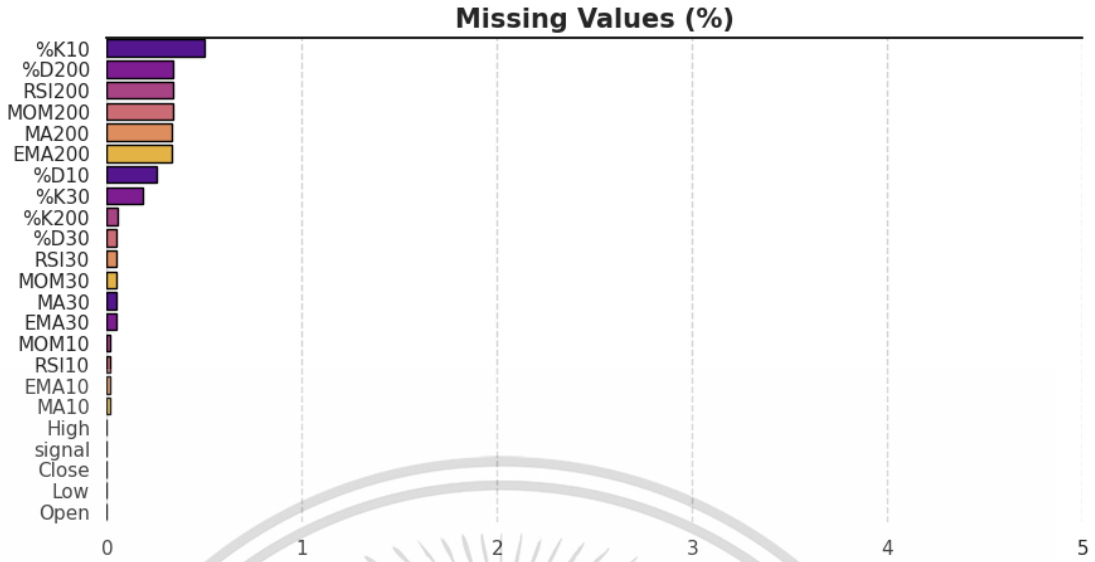
รูปที่ 4.9 ข้อมูล Stochastic Oscillator (Fast) ที่ 10, 30, 200 นาที



รูปที่ 4.10 ข้อมูล Stochastic Oscillator (Slow) ที่ 10, 30, 200 นาที

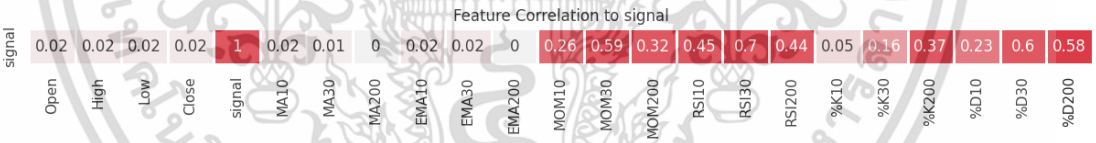
หลังจากเพิ่มตัวแปรต้นด้วย Technical Indicator เท่ากับตัวแปรต้นทั้งหมดที่ใช้ในการเรียนรู้แบบจำลอง และทดสอบประสิทธิภาพแบบจำลองคือ 22 รายการ ขั้นตอนถัดมาคือตรวจสอบข้อมูลอีกครั้งว่ามีค่าว่างหรือไม่ โดยใช้คำสั่ง DataFrame.isnull().sum() และสร้างเป็นแผนภูมิแท่งดังรูปที่ 4.11

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.11 แผนภูมิแท่งแสดงเปอร์เซ็นค่าว่าง (Missing value) ของข้อมูลที่ใช้เรียนรู้

พบว่า Technical Indicator ที่สร้างขึ้นมามีค่าว่างเกิดขึ้นไม่ถึง 1% จากข้อมูลทั้งหมด เนื่องมาจากการคำนวณด้วยคำสั่ง rolling(), diff() และ shift() เป็นต้น อย่างไรก็ตามการที่ข้อมูลมีค่าว่างนั้น อาจส่งผลต่อการเรียนรู้ของแบบจำลองได้ ดังนั้นจึงต้องกำจัดค่าว่างออกจากข้อมูลที่ใช้เรียนรู้แบบจำลอง และทดสอบประสิทธิภาพแบบจำลอง ด้วยคำสั่ง .dropna() หลังจากทำความสะอาดข้อมูลแล้ว ขั้นตอนถัดมาคือ พิจารณา Correlation ต่อคลาสเป้าหมายอีกครั้งดังรูปที่ 4.12



รูปที่ 4.12 Correlation ระหว่างตัวแปรต้นที่เพิ่ม Technical Indicator กับคลาสเป้าหมาย

พบว่าตัวแปรต้นจาก Technical Indicator นั้นมีความสัมพันธ์ต่อคลาสเป้าหมายมากขึ้น สามารถที่จะใช้เป็นข้อมูลเพื่อให้แบบจำลองเรียนรู้ได้มีประสิทธิภาพมากขึ้น ซึ่งถือเป็นหนึ่งในวิธีของนักวิทยาศาสตร์ข้อมูลคือการทำ Feature Engineering

4.2 ผลการทดสอบประสิทธิภาพแบบจำลองพื้นฐาน

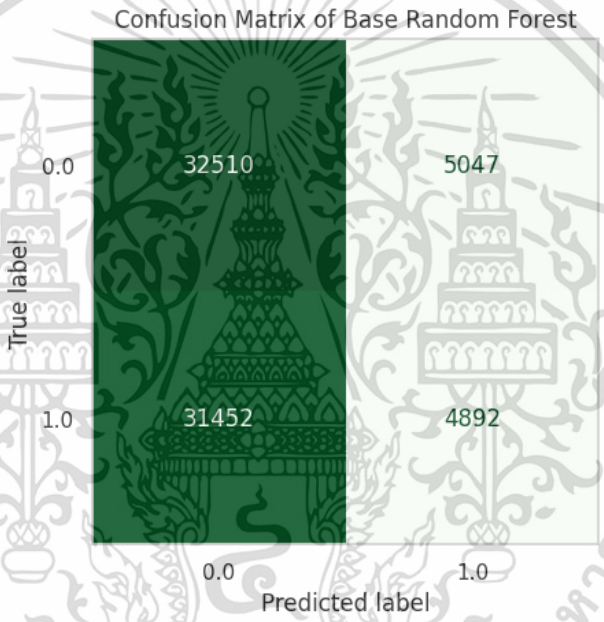
ในขั้นตอนของกระบวนการวิจัยนี้ คือการพัฒนาแบบจำลองสำหรับทำนายแนวโน้มการเปลี่ยนแปลงของราคาอ้างอิงในช่วงเวลาถัดไปโดยใช้ 3 เทคนิค ได้แก่ แบบจำลองการสุ่มป่าไม้, แบบจำลองเครื่องเวกเตอร์สนับสนุน และแบบจำลองการถดถอยโลจิสติก ซึ่งยังไม่ผ่านกระบวนการปรับปรุงประสิทธิภาพ โดยแบบจำลองนี้จะถูกใช้เป็นแบบจำลองพื้นฐานในการเปรียบเทียบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประสิทธิภาพด้านความแม่นยำในการทำนายกับแบบจำลองต่างๆ ที่ผ่านกระบวนการเพิ่มประสิทธิภาพด้วยวิธีการต่างๆ ซึ่งขั้นตอนการสร้างแบบจำลองพื้นฐานนี้ แสดงรายละเอียดดังถัดไป

4.2.1 แบบจำลองพื้นฐานการสุ่มป่าไม้

การเรียนรู้ของเครื่องด้วยแบบจำลองพื้นฐานการสุ่มป่าไม้นั้น จะใช้พารามิเตอร์พื้นฐานที่แบบจำลองกำหนดมาให้ คือ n_estimators = 100, max_depth = None, min_samples_split = 2, min_samples_leaf = 1 ใช้เวลาในการเรียนรู้ข้อมูลฝึกฝนเท่ากับ 1 นาที 50 วินาที โดยใช้มาตรวัดประสิทธิภาพของแบบจำลอง ได้แก่ Accuracy, Precision, Recall และ F-Measure ซึ่งผลการทดสอบประสิทธิภาพแสดงดังรูปที่ 4.13 และตารางที่ 4.1



รูปที่ 4.13 Confusion Matrix แสดงผลการทดสอบประสิทธิภาพแบบจำลองพื้นฐานการสุ่มป่าไม้

รูปที่ 4.13 แสดงตาราง Confusion Matrix ซึ่งบ่งบอกถึงผลลัพธ์ที่แบบจำลองการสุ่มป่าไม้พื้นฐานทำนาย จากข้อมูลที่ใช้ทดสอบทั้งสิ้นจำนวน 73,901 ข้อมูล แบบจำลองได้ทำนายได้ถูกต้องจำนวน 37,402 ข้อมูล แบ่งเป็นทำนายถูกต้องสำหรับข้อมูลแนวโน้มของราคาอ้างอิงแบบขาขึ้นจำนวน 4,892 ข้อมูล และแบบขาลงจำนวน 32,510 ข้อมูล ค่าในตาราง Confusion Matrix ถูกนำไปคิดเป็นค่าความแม่นยำในการทำนายดังแสดงในตารางที่ 4.1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

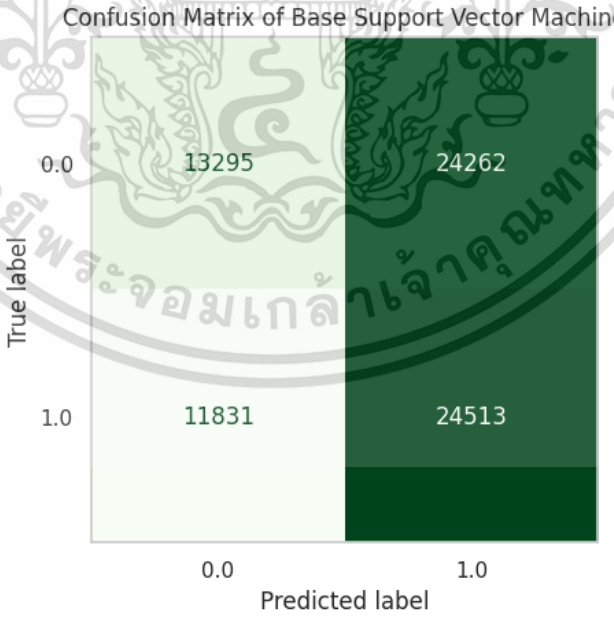
ตารางที่ 4.1 สรุปผลการทดสอบประสิทธิภาพแบบจำลองพื้นฐานการสุ่มป่าไม้

Class	Accuracy (%)	Precision (%)	Recall (%)	F-Measure
Upward (1)	51.6	49.0	13.0	21.1
Downward (0)		51.0	87.0	64.1
Average		50.0	50.0	42.6

ตารางที่ 4.1 แสดงผลลัพธ์ในด้านความแม่นยำในการทำนายของแบบจำลอง ซึ่งแบบจำลองมีความแม่นยำโดยรวม (Accuracy) อยู่ที่ 51.6% นอกจากนี้จะเห็นว่าเมื่อสนใจเฉพาะส่วนที่ทำนายแบบจำลองสามารถทำนายคลาส Down ได้ดีกว่าคลาส Up โดยพิจารณาจาก ค่า Recall และ F-Measure ของ Down สูงที่สุดเป็น 87% และ 64.1% ตามลำดับ

4.2.2 แบบจำลองพื้นฐานเครื่องเวกเตอร์สนับสนุน

การเรียนรู้ของเครื่องด้วยแบบจำลองพื้นฐานเครื่องเวกเตอร์สนับสนุนนั้น จะใช้พารามิเตอร์พื้นฐานที่แบบจำลองกำหนดมาให้ คือ C = 1.0 และ Gamma = ‘scale’ ใช้เวลาในการเรียนรู้ข้อมูลฝึกฝนเท่ากับ 1 ชั่วโมง 21 นาที 16 วินาที โดยใช้มาตรวัดประสิทธิภาพของแบบจำลอง ได้แก่ Accuracy, Precision, Recall และ F-Measure ซึ่งผลการทดสอบประสิทธิภาพแสดงดังรูปที่ 4.14 และตารางที่ 4.2



รูปที่ 4.14 Confusion Matrix แสดงผลการทดสอบประสิทธิภาพแบบจำลองพื้นฐานเครื่องเวกเตอร์สนับสนุน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปที่ 4.14 แสดงตาราง Confusion Matrix ซึ่งบ่งบอกถึงผลลัพธ์ที่แบบจำลองพื้นฐานเครื่องเวกเตอร์สนับสนุนทำนาย จากข้อมูลที่ใช้ทดสอบทั้งสิ้นจำนวน 73,901 ข้อมูล แบบจำลองได้ทำนายได้ถูกต้องจำนวน 37,808 ข้อมูล แบ่งเป็นทำนายถูกต้องสำหรับข้อมูลแนวโน้มของราคาอ้างอิงแบบขาขึ้นจำนวน 24,513 ข้อมูล และแบบขาลงจำนวน 13,295 ข้อมูล ค่าในตาราง Confusion Matrix ถูกนำไปคิดเป็นค่าความแม่นยำในการทำนายดังแสดงในตารางที่ 4.2

ตารางที่ 4.2 สรุปผลการทดสอบประสิทธิภาพแบบจำลองพื้นฐานเครื่องเวกเตอร์สนับสนุน

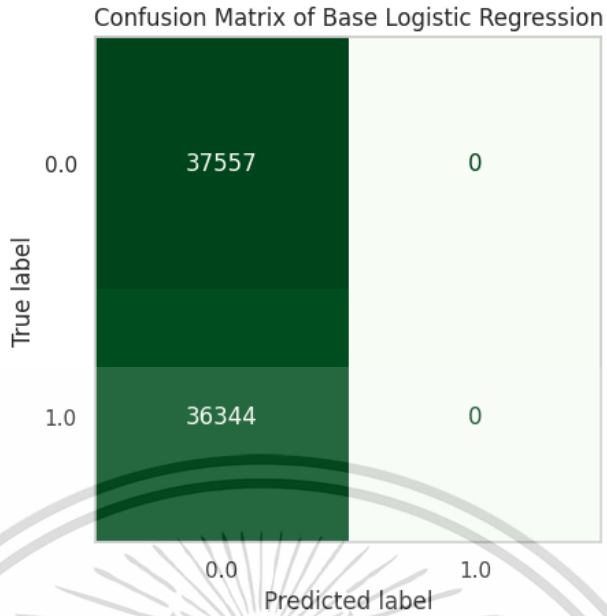
Class	Accuracy (%)	Precision (%)	Recall (%)	F-Measure
Upward (1)	51.2	50.4	67.6	58.0
Downward (0)		53.2	35.2	42.0
Average		51.8	51.4	50.0

ตารางที่ 4.2 แสดงผลลัพธ์ในด้านความแม่นยำในการทำนายของแบบจำลอง ซึ่งแบบจำลองมีความแม่นยำโดยรวม (Accuracy) อยู่ที่ 51.2% นอกจากนี้จะเห็นว่าเมื่อสนใจเฉพาะส่วนที่ทำนายแบบจำลองสามารถทำนายคลาส Upward ได้ดีกว่าคลาส Downward โดยพิจารณาจาก ค่า Recall และ F-Measure ของ Upward สูงที่สุดเป็น 67.6% และ 58.0% ตามลำดับ

4.2.3 แบบจำลองพื้นฐานการถดถอยโลจิสติก

การเรียนรู้ของเครื่องด้วยแบบจำลองพื้นฐานการถดถอยโลจิสติกนั้น จะใช้พารามิเตอร์พื้นฐานที่แบบจำลองกำหนดมาให้ คือ Solver = ‘lbfgs’ และ C = 1.0 ใช้เวลาในการเรียนรู้ข้อมูลฝึกฝนเท่ากับ 1 วินาที โดยใช้มาตรวัดประสิทธิภาพของแบบจำลอง ได้แก่ Accuracy, Precision, Recall และ F-Measure ซึ่งผลการทดสอบประสิทธิภาพแสดงดังรูปที่ 4.15 และตารางที่ 4.3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.15 Confusion Matrix แสดงผลการทดสอบประสิทธิภาพแบบจำลองพื้นฐานการถดถอยโลจิสติก

รูปที่ 4.15 แสดงตาราง Confusion Matrix ซึ่งบ่งบอกถึงผลลัพธ์ที่แบบจำลองพื้นฐานถดถอยโลจิสติกทำนาย จากข้อมูลที่ใช้ทดสอบทั้งสิ้นจำนวน 73,901 ข้อมูล แบบจำลองได้ทำนายได้ถูกต้องจำนวน 37,557 ข้อมูล แบ่งเป็นทำนายถูกต้องสำหรับข้อมูลแนวโน้มของราคาอ้างอิงแบบขาขึ้นจำนวน 0 ข้อมูล และแบบขาลงจำนวน 37,557 ข้อมูล ค่าในตาราง Confusion Matrix ถูกนำไปคิดเป็นค่าความแม่นยำในการทำนายดังแสดงในตารางที่ 4.3

ตารางที่ 4.3 สรุปผลการทดสอบประสิทธิภาพแบบจำลองพื้นฐานการถดถอยโลจิสติก

Class	Accuracy (%)	Precision (%)	Recall (%)	F-Measure
Upward (1)	50.8	0	0	0.0
Downward (0)		51.0	100	67.0
Average		24.5	50.0	33.7

ตารางที่ 4.3 แสดงผลลัพธ์ในด้านความแม่นยำในการทำนายของแบบจำลอง ซึ่งแบบจำลองมีความแม่นยำโดยรวม (Accuracy) อยู่ที่ 50.8% นอกจากนี้จะเห็นว่าเมื่อสนใจเฉพาะส่วนที่ทำนายแบบจำลองสามารถทำนายคลาส Downward ได้ดีกว่าคลาส Upward โดยพิจารณาจาก ค่า Recall และ F-Measure ของ Downward สูงที่สุดเป็น 67.6% และ 58.0% ตามลำดับ

จากผลการทดสอบประสิทธิภาพของแบบจำลองในขั้นตอนนี้ทั้งแบบจำลองการสุ่มป่าไม้, แบบจำลองเครื่องเวกเตอร์สนับสนุน และแบบจำลองการถดถอยโลจิสติก พบว่าแบบจำลองมีความ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

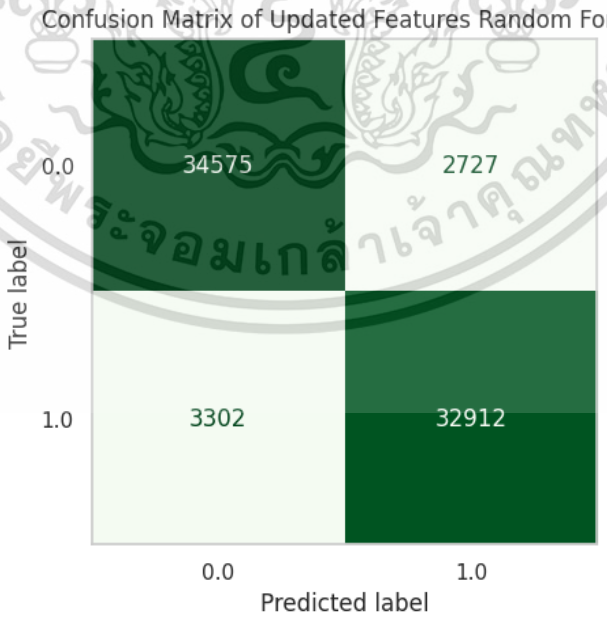
แม่นยำโดยรวม อยู่ที่ 50-51% ซึ่งความแม่นยำระดับนี้ยังไม่เพียงพอที่จะนำไปใช้เป็นเครื่องมือเพื่อใช้ทำนายทิศทางของการเปลี่ยนแปลงราคาในการเทรดได้ ดังนั้นในขั้นตอนต่อไปจะนำเสนอวิธีการต่างๆ เพื่อเพิ่มประสิทธิภาพให้กับแบบจำลอง ซึ่งผลการทดลองแสดงในหัวข้อถัดไป

4.3 ผลการทดสอบประสิทธิภาพแบบจำลองร่วมกับหลักการวิเคราะห์ทางเทคนิค

ในขั้นตอนของกระบวนการวิจัยนี้ คือการพัฒนาแบบจำลองสำหรับทำนายแนวโน้มการเปลี่ยนแปลงของราคาอ้างอิงในช่วงเวลาถัดไปโดยใช้ 3 เทคนิค ได้แก่ แบบจำลองการสุ่มป่าไม้, แบบจำลองเครื่องเวกเตอร์สนับสนุน และแบบจำลองการถดถอยโลจิสติก เช่นเดิม และได้เพิ่ม Technical Indicator ลงไปเป็นตัวแปรต้น สำหรับการเรียนรู้ และทดสอบประสิทธิภาพ ซึ่งขั้นตอนการสร้างแบบจำลองร่วมกับหลักการวิเคราะห์ทางเทคนิคนี้ แสดงรายละเอียดดังถัดไป

4.3.1 แบบจำลองการสุ่มป่าไม้ร่วมกับหลักการวิเคราะห์ทางเทคนิค

การเรียนรู้ของเครื่องด้วยแบบจำลองการสุ่มป่าไม้ร่วมกับหลักการวิเคราะห์ทางเทคนิคนั้น จะใช้พารามิเตอร์พื้นฐานที่แบบจำลองกำหนดมาให้ คือ n_estimators = 100, max_depth = None, min_samples_split = 2, min_samples_leaf = 1 ใช้เวลาในการเรียนรู้ข้อมูลฝึกฝนเท่ากับ 3 นาที 7 วินาที โดยใช้มาตรวัดประสิทธิภาพของแบบจำลอง ได้แก่ Accuracy, Precision, Recall และ F-Measure ซึ่งผลการทดสอบประสิทธิภาพแสดงดังรูปที่ 4.16 และตารางที่ 4.4



รูปที่ 4.16 Confusion Matrix แสดงผลการทดสอบประสิทธิภาพแบบจำลองการสุ่มป่าไม้ร่วมกับหลักการวิเคราะห์ทางเทคนิค

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปที่ 4.16 แสดงตาราง Confusion Matrix ซึ่งบ่งบอกถึงผลลัพธ์ที่แบบจำลองการสุ่มป่าไม้ร่วมกับหลักการวิเคราะห์ทางเทคนิคทำนาย จากข้อมูลที่ใช้ทดสอบทั้งสิ้นจำนวน 73,516 ข้อมูล แบบจำลองได้ทำนายได้ถูกต้องจำนวน 67,487 ข้อมูล แบ่งเป็นทำนายถูกต้องสำหรับข้อมูลแนวโน้มของราคาอ้างอิงแบบขาขึ้นจำนวน 32,912 ข้อมูล และแบบขาลงจำนวน 34,575 ข้อมูล ค่าในตาราง Confusion Matrix ถูกนำไปคิดเป็นค่าความแม่นยำในการทำนายดังแสดงในตารางที่ 4.4

ตารางที่ 4.4 สรุปผลการทดสอบประสิทธิภาพแบบจำลองการสุ่มป่าไม้ร่วมกับหลักการวิเคราะห์ทางเทคนิค

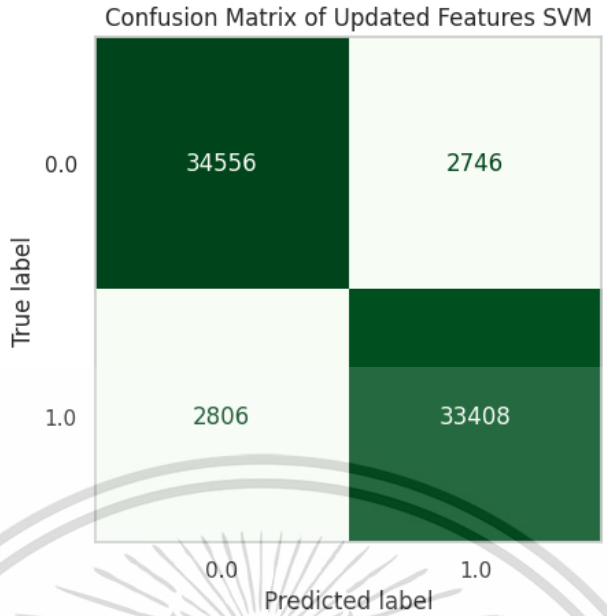
Class	Accuracy (%)	Precision (%)	Recall (%)	F-Measure
Upward (1)	91.8	92.0	91.5	91.8
Downward (0)		91.6	92.1	91.8
Average		91.8	91.8	91.8

ตารางที่ 4.4 แสดงผลลัพธ์ในด้านความแม่นยำในการทำนายของแบบจำลอง ซึ่งแบบจำลองมีความแม่นยำโดยรวม (Accuracy) อยู่ที่ 91.8% สูงขึ้นกว่าแบบจำลองพื้นฐานเดิมถึง 41.2% นอกจากนี้จะเห็นว่าเมื่อสนใจเฉพาะส่วนที่ทำนาย แบบจำลองสามารถทำนายคลาส Upward และ Downward ได้ดีขึ้นใกล้เคียงกัน โดยพิจารณาจาก ค่า F-Measure ที่ได้เท่ากันที่ 91.8%

4.3.2 แบบจำลองเครื่องเวกเตอร์สนับสนุนร่วมกับหลักการวิเคราะห์ทางเทคนิค

การเรียนรู้ของเครื่องด้วยแบบจำลองเครื่องเวกเตอร์สนับสนุนร่วมกับหลักการวิเคราะห์ทางเทคนิคนั้น จะใช้พารามิเตอร์พื้นฐานที่แบบจำลองกำหนดมาให้ คือ C = 1.0 และ Gamma = ‘scale’ ใช้เวลาในการเรียนรู้ข้อมูลฝึกฝนเท่ากับ 38 นาที 1 วินาที โดยใช้มาตรวัดประสิทธิภาพของแบบจำลอง ได้แก่ Accuracy, Precision, Recall และ F-Measure ซึ่งผลการทดสอบประสิทธิภาพแสดงดังรูปที่ 4.17 และตารางที่ 4.5

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.17 Confusion Matrix แสดงผลการทดสอบประสิทธิภาพแบบจำลองเครื่องเวกเตอร์สนับสนุนร่วมกับหลักการวิเคราะห์ทางเทคนิค

รูปที่ 4.17 แสดงตาราง Confusion Matrix ซึ่งบ่งบอกถึงผลลัพธ์ที่แบบจำลองเครื่องเวกเตอร์สนับสนุนร่วมกับหลักการวิเคราะห์ทางเทคนิคทำนาย จากข้อมูลที่ใช้ทดสอบทั้งสิ้นจำนวน 73,516 ข้อมูล แบบจำลองได้ทำนายได้ถูกต้องจำนวน 67,964 ข้อมูล แบ่งเป็นทำนายถูกต้องสำหรับข้อมูลแนวโน้มของราคาอ้างอิงแบบขาขึ้นจำนวน 33,408 ข้อมูล และแบบขาลงจำนวน 34,556 ข้อมูล ค่าในตาราง Confusion Matrix ถูกนำไปคิดเป็นค่าความแม่นยำในการทำนายดังแสดงในตารางที่ 4.5

ตารางที่ 4.5 สรุปผลการทดสอบประสิทธิภาพแบบจำลองเครื่องเวกเตอร์สนับสนุนร่วมกับหลักการวิเคราะห์ทางเทคนิค

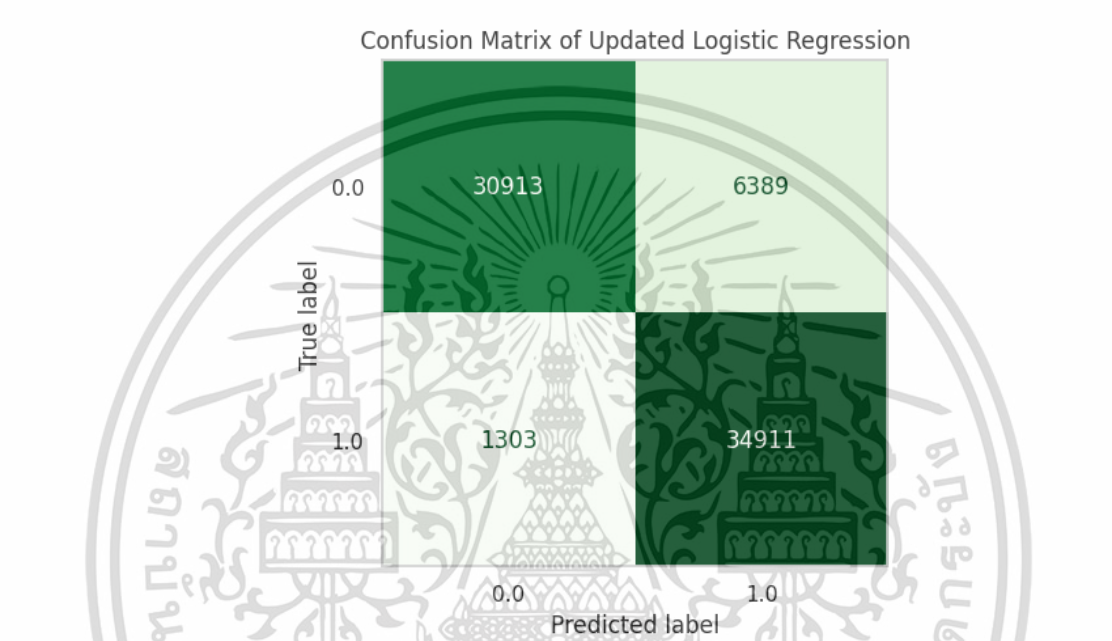
Class	Accuracy (%)	Precision (%)	Recall (%)	F-Measure
Upward (1)	92.4	92.5	92.1	92.2
Downward (0)		92.3	92.7	92.6
Average		92.4	92.4	92.4

ตารางที่ 4.5 แสดงผลลัพธ์ในด้านความแม่นยำในการทำนายของแบบจำลอง ซึ่งแบบจำลองมีความแม่นยำโดยรวม (Accuracy) อยู่ที่ 92.4% สูงขึ้นกว่าแบบจำลองพื้นฐานเดิมถึง 41.3% นอกจากนี้จะเห็นว่าเมื่อสนใจเฉพาะส่วนที่ทำนาย แบบจำลองสามารถทำนายคลาส Upward และ Downward ได้ดีขึ้นใกล้เคียงกัน โดยพิจารณาจาก ค่า F-Measure ที่ได้เท่ากันที่ 92.4%

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.3.3 แบบจำลองการถดถอยโลจิสติกร่วมกับหลักการวิเคราะห์ทางเทคนิค

การเรียนรู้ของเครื่องด้วยแบบจำลองการถดถอยโลจิสติกร่วมกับหลักการวิเคราะห์ทางเทคนิคนั้น จะใช้พารามิเตอร์พื้นฐานที่แบบจำลองกำหนดมาให้ คือ Solver = ‘lbfgs’ และ C = 1.0 ใช้เวลาในการเรียนรู้ข้อมูลฝึกฝนเท่ากับ 6 วินาที โดยใช้มาตรวัดประสิทธิภาพของแบบจำลอง ได้แก่ Accuracy, Precision, Recall และ F-Measure ซึ่งผลการทดสอบประสิทธิภาพแสดงดังรูปที่ 4.18 และตารางที่ 4.6



รูปที่ 4.18 Confusion Matrix แสดงผลการทดสอบประสิทธิภาพแบบจำลองการถดถอยโลจิสติกร่วมกับหลักการวิเคราะห์ทางเทคนิค

รูปที่ 4.18 แสดงตาราง Confusion Matrix ซึ่งบ่งบอกถึงผลลัพธ์ที่แบบจำลองเครื่องเวกเตอร์สนับสนุนร่วมกับหลักการวิเคราะห์ทางเทคนิคทำนาย จากข้อมูลที่ใช้ทดสอบทั้งสิ้นจำนวน 73,516 ข้อมูล แบบจำลองได้ทำนายได้ถูกต้องจำนวน 65,824 ข้อมูล แบ่งเป็นทำนายถูกต้องสำหรับข้อมูลแนวโน้มของราคาอ้างอิงแบบขาขึ้นจำนวน 34,911 ข้อมูล และแบบขาลงจำนวน 30,913 ข้อมูล ค่าในตาราง Confusion Matrix ถูกนำไปคิดเป็นค่าความแม่นยำในการทำนายดังแสดงในตารางที่ 4.6

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.6 สรุปผลการทดสอบประสิทธิภาพแบบจำลองการถดถอยโลจิสติกร่วมกับหลักการวิเคราะห์ทางเทคนิค

Class	Accuracy (%)	Precision (%)	Recall (%)	F-Measure
Upward (1)	89.5	85.0	96.0	90.0
Downward (0)		95.4	83.2	89.0
Average		90.2	89.6	89.5

ตารางที่ 4.6 แสดงผลลัพธ์ในด้านความแม่นยำในการทำนายของแบบจำลอง ซึ่งแบบจำลองมีความแม่นยำโดยรวม (Accuracy) อยู่ที่ 89.5% สูงขึ้นกว่าแบบจำลองพื้นฐานเดิมถึง 38.7% นอกจากนี้จะเห็นว่าเมื่อสนใจเฉพาะส่วนที่ทำนาย แบบจำลองสามารถทำนายคลาส Upward และ Downward ได้ดีขึ้นใกล้เคียงกัน โดยพิจารณาจาก ค่า F-Measure ที่ได้เท่ากับ 90.0% และ 89.0% ตามลำดับ

จากผลการทดสอบประสิทธิภาพของแบบจำลองในขั้นตอนนี้ทั้งแบบจำลองการสุ่มป่าไม้, แบบจำลองเครื่องเวกเตอร์สนับสนุน และแบบจำลองการถดถอยโลจิสติก ที่ได้ใช้หลักการวิเคราะห์ทางเทคนิคร่วมด้วย พบว่าแบบจำลองมีความแม่นยำโดยรวมอยู่ที่ 89-92% ซึ่งความแม่นยำระดับนี้ถือว่าเพียงพอที่จะนำไปใช้เป็นเครื่องมือเพื่อใช้ทำนายทิศทางของการเปลี่ยนแปลงราคาในการเทรดได้ อย่างไรก็ตามผู้ทำวิจัยคาดว่าสามารถเพิ่มประสิทธิภาพแบบจำลองให้สูงขึ้นอีกได้ เนื่องจากการเรียนรู้แบบจำลองที่กล่าวในข้างต้นนั้น ใช้พารามิเตอร์ที่ตัวระบบกำหนดมาให้ โดยไม่มีการปรับค่าใดๆ เลย ดังนั้นในขั้นตอนถัดไปนี้จะนำเสนอการปรับพารามิเตอร์ต่างๆ ให้แบบจำลองมีประสิทธิภาพสูงขึ้น ซึ่งผลการทดลองแสดงในหัวข้อถัดไป

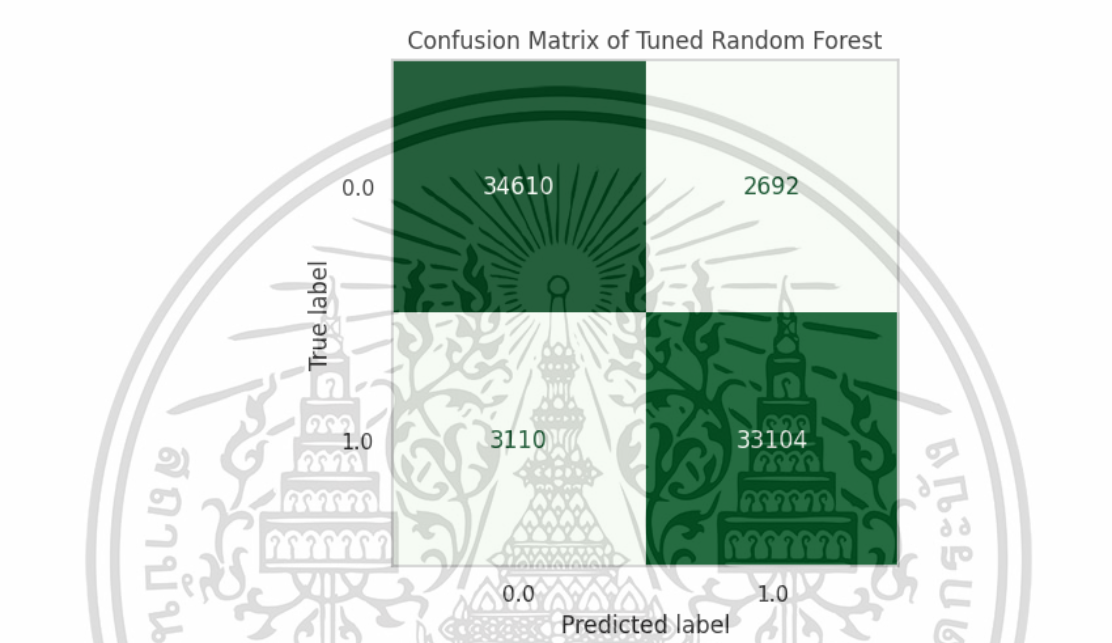
4.4 ผลการทดสอบประสิทธิภาพแบบจำลองโดยปรับพารามิเตอร์ที่เหมาะสม

ในขั้นตอนของกระบวนการวิจัยนี้ คือการพัฒนาแบบจำลองสำหรับทำนายแนวโน้มการเปลี่ยนแปลงของราคาอ้างอิงในช่วงเวลาถัดไปโดยใช้ 3 เทคนิค ได้แก่ แบบจำลองการสุ่มป่าไม้, แบบจำลองเครื่องเวกเตอร์สนับสนุน และแบบจำลองการถดถอยโลจิสติก เช่นเดิม โดยต่อยอดจากหัวข้อ 4.3 ที่ได้เพิ่มและได้เพิ่ม Technical Indicator ลงไปเป็นตัวแปรต้น ในขั้นตอนนี้จะทำการหาค่าพารามิเตอร์ที่เหมาะสมในการเรียนรู้ของแต่ละแบบจำลอง และทำให้แบบจำลองมีความแม่นยำสูงขึ้นเมื่อทดสอบประสิทธิภาพกับข้อมูลทดสอบ ซึ่งกระบวนการหาพารามิเตอร์ที่เหมาะสมนี้ จะใช้คำสั่งจาก sklearn คือ model_selection.GridSearchCV และให้แสดงพารามิเตอร์ที่ดีที่สุดด้วย best_params_ การปรับพารามิเตอร์ที่เหมาะสมที่สุด แสดงรายละเอียดดังถัดไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.4.1 แบบจำลองการสุ่มป่าไม้ร่วมกับการปรับพารามิเตอร์

การเรียนรู้ของเครื่องด้วยแบบจำลองการสุ่มป่าไม้ร่วมกับการปรับพารามิเตอร์นั้น จากหัวข้อที่ 3.2.1 ได้พารามิเตอร์ที่แบบจำลองมีประสิทธิภาพดีขึ้น คือ n_estimators = 500, max_depth = None, min_samples_split = 2, min_samples_leaf = 2 ใช้เวลาในการเรียนรู้ข้อมูลฝึกฝนเท่ากับ 14 นาที 44 วินาที โดยใช้มาตรวัดประสิทธิภาพ ได้แก่ Accuracy, Precision, Recall และ F-Measure ซึ่งผลการทดสอบประสิทธิภาพแสดงดังรูปที่ 4.19 และตารางที่ 4.7



รูปที่ 4.19 Confusion Matrix แสดงผลการทดสอบประสิทธิภาพแบบจำลองการสุ่มป่าไม้ผ่านการปรับพารามิเตอร์

รูปที่ 4.19 แสดงตาราง Confusion Matrix ซึ่งบ่งบอกถึงผลลัพธ์ที่แบบจำลองการสุ่มป่าไม้ผ่านการปรับพารามิเตอร์ จากข้อมูลที่ใช้ทดสอบทั้งสิ้นจำนวน 73,516 ข้อมูล แบบจำลองได้ทำนายได้ถูกต้องจำนวน 67,714 ข้อมูล แบ่งเป็นทำนายถูกต้องสำหรับข้อมูลแนวโน้มของราคาอ้างอิงแบบขาขึ้นจำนวน 33,104 ข้อมูล และแบบขาลงจำนวน 34,610 ข้อมูล ค่าในตาราง Confusion Matrix ถูกนำไปคิดเป็นค่าความแม่นยำในการทำนายดังแสดงในตารางที่ 4.7

ตารางที่ 4.7 สรุปผลการทดสอบประสิทธิภาพแบบจำลองการสุ่มป่าไม้ผ่านการปรับพารามิเตอร์

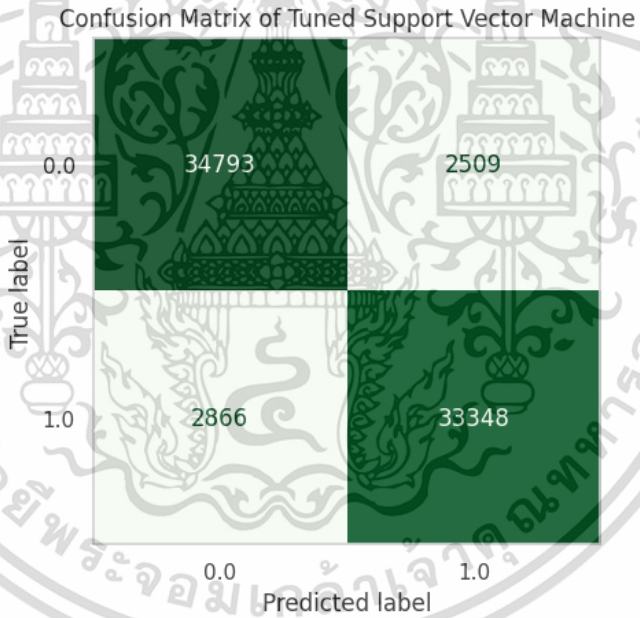
Class	Accuracy (%)	Precision (%)	Recall (%)	F-Measure
Upward (1)	92.1	92.0	91.2	92.0
Downward (0)		92.2	93.0	92.2
Average		92.1	92.1	92.1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.7 แสดงผลลัพธ์ในด้านความแม่นยำในการทำนายของแบบจำลอง ซึ่งแบบจำลองมีความแม่นยำโดยรวม (Accuracy) อยู่ที่ 92.1% สูงขึ้นกว่าแบบจำลองพื้นฐานเดิม 0.3% นอกจากนี้จะเห็นว่าเมื่อสนใจเฉพาะส่วนที่ทำนาย แบบจำลองสามารถทำนายคลาส Upward และ Downward ได้ดีขึ้นใกล้เคียงกัน โดยพิจารณาจาก ค่า F-Measure ที่ได้เท่ากับ 92.0% และ 92.2% ตามลำดับ

4.4.2 แบบจำลองเครื่องเวกเตอร์สนับสนุนร่วมกับการปรับพารามิเตอร์

การเรียนรู้ของเครื่องด้วยแบบจำลองเครื่องเวกเตอร์สนับสนุนร่วมกับการปรับพารามิเตอร์นั้น จากหัวข้อที่ 3.2.2 ได้พารามิเตอร์ที่แบบจำลองมีประสิทธิภาพดีขึ้น คือ C = 500 และ Gamma = ‘scale’ ใช้เวลาในการเรียนรู้ข้อมูลฝึกฝนเท่ากับ 1 ชั่วโมง 11 นาที 49 วินาที โดยใช้มาตรวัดประสิทธิภาพของแบบจำลอง ได้แก่ Accuracy, Precision, Recall และ F-Measure ซึ่งผลการทดสอบประสิทธิภาพแสดงดังรูปที่ 4.20 และตารางที่ 4.8



รูปที่ 4.20 Confusion Matrix แสดงผลการทดสอบประสิทธิภาพแบบจำลองเครื่องเวกเตอร์สนับสนุนผ่านการปรับพารามิเตอร์

รูปที่ 4.20 แสดงตาราง Confusion Matrix ซึ่งบ่งบอกถึงผลลัพธ์ที่แบบจำลองเครื่องเวกเตอร์สนับสนุนผ่านการปรับพารามิเตอร์ จากข้อมูลที่ใช้ทดสอบทั้งสิ้นจำนวน 73,516 ข้อมูล แบบจำลองได้ทำนายได้ถูกต้องจำนวน 68,141 ข้อมูล แบ่งเป็นทำนายถูกต้องสำหรับข้อมูลแนวโน้มของราคาอ้างอิงแบบขาขึ้นจำนวน 33,348 ข้อมูล และแบบขาลงจำนวน 34,793 ข้อมูล ค่าในตาราง Confusion Matrix ถูกนำไปคิดเป็นค่าความแม่นยำในการทำนายดังแสดงในตารางที่ 4.8

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

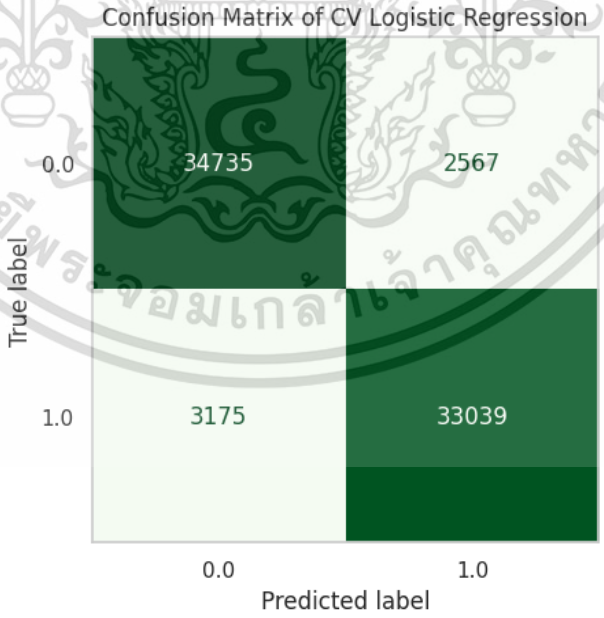
ตารางที่ 4.8 สรุปผลการทดสอบประสิทธิภาพแบบจำลองเครื่องเวกเตอร์สนับสนุนผ่านการปรับพารามิเตอร์

Class	Accuracy (%)	Precision (%)	Recall (%)	F-Measure
Upward (1)	92.7	92.8	92.6	92.7
Downward (0)		92.6	92.8	92.7
Average		92.7	92.7	92.7

ตารางที่ 4.8 แสดงผลลัพธ์ในด้านความแม่นยำในการทำนายของแบบจำลอง ซึ่งแบบจำลองมีความแม่นยำโดยรวม (Accuracy) อยู่ที่ 92.7% สูงขึ้นกว่าแบบจำลองพื้นฐานเดิม 0.3% นอกจากนี้จะเห็นว่าเมื่อสนใจเฉพาะส่วนที่ทำนาย แบบจำลองสามารถทำนายคลาส Upward และ Downward ได้ดีขึ้นเท่ากัน โดยพิจารณาจาก ค่า F-Measure ที่ได้เท่ากับ 92.7%

4.4.3 แบบจำลองการถดถอยโลจิสติกร่วมกับการปรับพารามิเตอร์

การเรียนรู้ของเครื่องด้วยแบบจำลองการถดถอยโลจิสติกร่วมกับการปรับพารามิเตอร์นั้น จากหัวข้อที่ 3.2.3 ได้พารามิเตอร์ที่แบบจำลองมีประสิทธิภาพดีขึ้น คือ Solver = ‘newton-cg’ และ C = 100 ใช้เวลาในการเรียนรู้ข้อมูลฝึกฝนเท่ากับ 31 วินาที โดยใช้มาตรวัดประสิทธิภาพของแบบจำลอง ได้แก่ Accuracy, Precision, Recall และ F-Measure ซึ่งผลการทดสอบประสิทธิภาพแสดงดังรูปที่ 4.21 และตารางที่ 4.9



รูปที่ 4.21 Confusion Matrix แสดงผลการทดสอบประสิทธิภาพแบบจำลองการถดถอยโลจิสติกผ่านการปรับพารามิเตอร์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปที่ 4.21 แสดงตาราง Confusion Matrix ซึ่งบ่งบอกถึงผลลัพธ์ที่แบบจำลองเครื่องเวกเตอร์สนับสนุนผ่านการปรับพารามิเตอร์ จากข้อมูลที่ใช้ทดสอบทั้งสิ้นจำนวน 73,516 ข้อมูล แบบจำลองได้ทำนายได้ถูกต้องจำนวน 67,774 ข้อมูล แบ่งเป็นทำนายถูกต้องสำหรับข้อมูลแนวโน้มของราคาอ้างอิงแบบขาขึ้นจำนวน 33,039 ข้อมูล และแบบขาลงจำนวน 34,735 ข้อมูล ค่าในตาราง Confusion Matrix ถูกนำไปคิดเป็นค่าความแม่นยำในการทำนายดังแสดงในตารางที่ 4.9

ตารางที่ 4.9 สรุปผลการทดสอบประสิทธิภาพแบบจำลองการถดถอยโลจิสติกผ่านการปรับพารามิเตอร์

Class	Accuracy (%)	Precision (%)	Recall (%)	F-Measure
Upward (1)	92.2	92.0	93.0	92.0
Downward (0)		93.0	91.0	92.0
Average		92.2	92.2	92.2

ตารางที่ 4.9 แสดงผลลัพธ์ในด้านความแม่นยำในการทำนายของแบบจำลอง ซึ่งแบบจำลองมีความแม่นยำโดยรวม (Accuracy) อยู่ที่ 92.2% สูงขึ้นกว่าแบบจำลองพื้นฐานเดิม 2.7% นอกจากนี้จะเห็นว่าเมื่อสนใจเฉพาะส่วนที่ทำนาย แบบจำลองสามารถทำนายคลาส Upward และ Downward ได้ดีขึ้นใกล้เคียงกัน โดยพิจารณาจาก ค่า F-Measure ที่ได้เท่ากับ 92.2% และ 92.2% ตามลำดับ

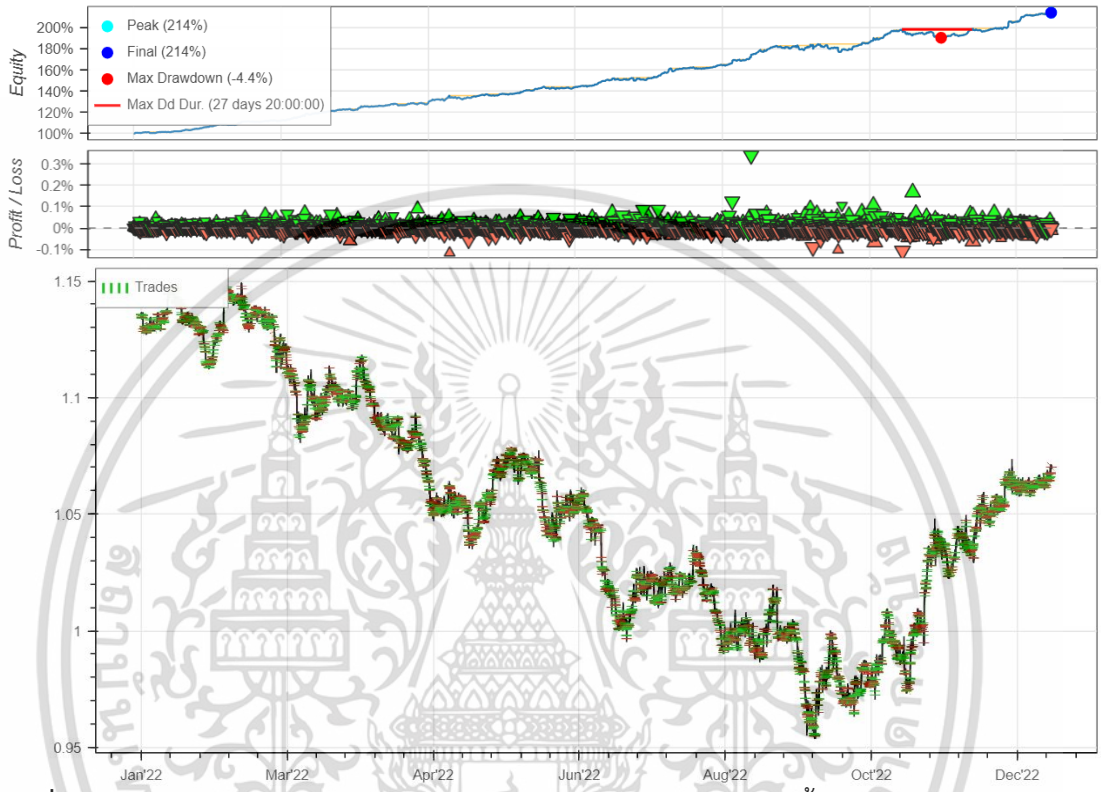
4.5 ผลการทดสอบแบบย้อนกลับ

สิ่งที่สำคัญที่สุดและเป็นเป้าหมายหลักของงานวิจัยนี้คือการที่เครื่องมือและกลยุทธ์ต่างๆ ที่ได้นำเสนอในงานวิจัยนี้สามารถใช้เพื่อสร้างผลกำไรให้กับนักลงทุนได้ ดังนั้นในขั้นตอนสุดท้ายของการวิจัยคือ การนำแบบจำลองการสุ่มป่าไม้, แบบจำลองเครื่องเวกเตอร์สนับสนุน และแบบจำลองการถดถอยโลจิสติก มาทดสอบประสิทธิภาพในการซื้อขาย โดยกระบวนการวัดประสิทธิภาพนั้นจะใช้วิธีการทดสอบแบบย้อนกลับ เป็นการวัดประสิทธิภาพในการทำกำไรของเครื่องมือต่างๆ ซึ่งเป็นกระบวนการที่นักลงทุนส่วนใหญ่นิยมทำเพื่อประเมินว่าสมควรใช้เครื่องมือและกลยุทธ์ที่คิดค้นขึ้นมานั้นไปใช้ทำกำไรในการเทรดในสภาวะแวดล้อมจริงได้หรือไม่ โดยกระบวนการทดสอบย้อนกลับนั้นจะเป็นการจำลองการเทรดจากข้อมูลย้อนหลัง ในงานวิจัยนี้ได้ใช้ข้อมูลย้อนหลังของราคาอ้างอิงของการซื้อขายและแลกเปลี่ยนเงินตราต่างประเทศคู่เงิน EURUSD คาบเวลาราย 1 นาที ตั้งแต่วันที่ 2 มกราคม พ.ศ. 2565 จนถึงวันที่ 31 ธันวาคม พ.ศ. 2565 มาใช้ทดสอบ โดยมีการกำหนดเงื่อนไขในการเทรดดังนี้

1) จำนวนเงินลงทุนเริ่มแรกเท่ากับ 1000 ดอลลาร์
2) คู่เงินที่ใช้เทรดคือ EURUSD
3) เทรดทั้งฝั่ง Buy และ Sell

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การจำลองการเทรดจะวัดผลเป็นรายปี จะวัดด้วยมาตรวัด จำนวนครั้งการเทรด (No.Trade), %Return (เปอร์เซ็นต์ผลกำไร) และ %Maximum Drawdown (เปอร์เซ็นต์ผลขาดทุนสูงสุด) ผลการทดสอบแบบย้อนกลับแต่ละแบบจำลองนำเสนอดังนี้ โดยจะแสดงตัวอย่างผลการทดสอบแบบย้อนกลับดังรูปที่ 4.22 โดยเส้นสีเขียวคือการเทรดที่ได้กำไร และเส้นสีแดงคือการเทรดที่ขาดทุน



รูปที่ 4.22 ตัวอย่างการแสดงผลการทดสอบย้อนกลับของแบบจำลองพื้นฐานการสุ่มป่าไม้ร่วมกับตัวชี้วัดทางเทคนิค

การทดสอบแบบจำลองการสุ่มป่าไม้, เครื่องเวกเตอร์สนับสนุน และการถดถอยโลจิสติก แบ่งเป็น 2 วิธีคือ 1.แบบจำลองพื้นฐานที่ร่วมกับตัวชี้วัดทางเทคนิค และ 2.แบบจำลองที่ผ่านการปรับพารามิเตอร์ โดยสรุปผลการทดสอบแบบย้อนกลับตลอดปี พ.ศ.2565 ดังตารางที่ 4.10

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.10 สรุปผลลัพธ์การเทรดของการทดสอบแบบย้อนกลับในปี พ.ศ.2565

Model	Feature	No.Trade	Return (%)	Final Equity ($)	Maximum Drawdown (%)
Random Forest	Updated Feature	19,546	114.28	2,142.8	4.36
	Parameter Turning	16,180	85.15	1,851.48	4.21
Support Vector Machine	Updated Feature	12,480	72.02	1,720.22	4.47
	Parameter Turning	13,128	63.47	1,634.73	3.63
Logistic Regression	Updated Feature	12,529	89.26	1,892.61	2.48
	Parameter Turning	12,326	58.23	1,582.29	3.88

จากผลลัพธ์ที่แสดงในตารางที่ 4.10 โดยเปรียบเทียบแบบจำลองที่มีผลกำไรสูงสุดจากมากไปน้อยดังนี้ แบบจำลองพื้นฐานการสุ่มป่าไม้ร่วมกับตัวชี้วัดทางเทคนิค มีผลกำไร 114.28% แบบจำลองพื้นฐานโลจิสติกร่วมกับตัวชี้วัดทางเทคนิค มีผลกำไร 89.26% แบบจำลองการสุ่มป่าไม้ร่วมกับการปรับพารามิเตอร์ มีผลกำไร 85.15% แบบจำลองพื้นฐานเครื่องเวกเตอร์สนับสนุนร่วมกับตัวชี้วัดทางเทคนิค มีผลกำไร 72.02% แบบจำลองเครื่องเวกเตอร์สนับสนุนร่วมกับการปรับพารามิเตอร์ มีผลกำไร 63.47% และแบบจำลองการถดถอยโลจิสติกร่วมกับการปรับพารามิเตอร์มีผลกำไร 58.23%

4.6 อภิปรายผลการวิจัย

วัตถุประสงค์หลักของงานวิจัยนี้คือการพัฒนาเครื่องมือและกลยุทธ์ที่นักลงทุนสามารถใช้เป็นตัวช่วยเพื่อทำกำไรในการซื้อขายแลกเปลี่ยนเงินตราต่างประเทศได้ ซึ่งกระบวนการวิจัยได้นำเสนอขั้นตอนในการเตรียมข้อมูลและกระบวนการที่ใช้ในการพัฒนาแบบจำลองเพื่อใช้ในการกำหนดทิศทางการลงทุนโดยเลือกใช้ทฤษฎีการเรียนรู้ของเครื่อง ได้แก่ แบบจำลองการสุ่มป่าไม้, แบบจำลองเครื่องเวกเตอร์สนับสนุน และแบบจำลองการถดถอยโลจิสติก และนำเสนอวิธีการต่างๆ ที่ใช้ในการเพิ่มประสิทธิภาพให้กับแบบจำลอง นอกจากนี้ยังได้นำเสนอการนำแบบจำลองไปทดสอบแบบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ย้อนกลับเพื่อพิจารณาผลกำไรที่ได้ จากผลลัพธ์ที่ได้จากการทดสอบทำให้สามารถสรุปเป็นประเด็นต่างๆ ได้ดังนี้

1) ผลสรุปที่ได้จากการทดสอบประสิทธิภาพแบบจำลองทั้ง 3 รูปแบบ โดยกระบวนการพื้นฐานและที่ยังไม่ได้ผ่านการเพิ่มประสิทธิภาพ พบว่าแบบจำลองมีความแม่นยำในการทำนายเฉลี่ยอยู่ในช่วง 50-51% ซึ่งยังไม่เพียงพอสำหรับการนำไปใช้งานจริง
2) ผลการทดสอบประสิทธิภาพแบบจำลองเมื่อเพิ่มตัวชี้วัดทางเทคนิคให้แก่ข้อมูลสำหรับการเรียนรู้เพื่อพัฒนาแบบจำลองพบว่า สามารถให้ความแม่นยำในการทำนายโดยรวมสูงขึ้น โดยแบบจำลองที่มีค่าความแม่นยำสูงสุด คือ เครื่องเวอเตอร์สนับสนุน เท่ากับ 92.45% ซึ่งสูงกว่างานวิจัยของ Chen et al., (2020) ที่ได้ค่าความแม่นยำอยู่ที่ 75%
3) วิธีการสุดท้ายที่ผู้วิจัยนำเสนอเพื่อเพิ่มประสิทธิภาพในการทำนายให้แก่แบบจำลอง คือ การปรับพารามิเตอร์ ด้วยเทคนิค GridSearchCV พบว่าแบบจำลองที่มีค่าความแม่นยำสูงสุดยังคงเป็นเครื่องเวอเตอร์สนับสนุน เท่ากับ 92.70% ในขณะที่แบบจำลองการถดถอยโลจิสติกมีค่าความแม่นยำเพิ่มสูงขึ้นสูงสุด จาก 89.54% เป็น 92.19% คิดเป็น 2.65%
4) จากผลการทดสอบแบบย้อนกลับสรุปได้ว่าเครื่องมือและกลยุทธ์ที่นำเสนอในงานวิจัยนี้สามารถสร้างผลกำไรได้อย่างต่อเนื่องจากการจำลองการเทรดโดยใช้ข้อมูลย้อนหลัง โดยไม่มีพอร์ตใดๆ ขาดทุนจนถึงขั้นล้างพอร์ต พบว่าแบบจำลองพื้นฐานการสุ่มป่าไม้ร่วมกับหลักการวิเคราะห์ทางเทคนิคนั้นให้ผลกำไรสูงสุดถึง 114.28% และมีผลขาดทุนสูงสุดเพียง 4.36% เท่านั้น ซึ่งสูงกว่างานวิจัยของ Thu และ Xuan (2018) ที่ได้อัตราผลกำไรเท่ากับ 46.3% และผลขาดทุนสูงสุดเท่ากับ 24.31%
5) ในขณะเดียวกันหากพิจารณาแบบจำลองอื่นๆ ที่ได้ค่าความแม่นยำสูงกว่าแบบจำลองพื้นฐานการสุ่มป่าไม้ร่วมกับหลักการวิเคราะห์ทางเทคนิค เช่นแบบจำลองเครื่องเวกเตอร์สนับสนุนร่วมกับการปรับพารามิเตอร์ แต่มีผลกำไรต่ำกว่าเนื่องจากจำนวนการเทรดที่แตกต่างกัน คือ แบบจำลองพื้นฐานการสุ่มป่าไม้ร่วมกับหลักการวิเคราะห์ทางเทคนิคมีจำนวนการเทรดสูงสุด คือ 19,546 ครั้ง ในขณะที่แบบจำลองเครื่องเวกเตอร์สนับสนุนร่วมกับการปรับพารามิเตอร์มีจำนวนการเทรดเท่ากับ 13,128 ครั้ง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5
สรุปผลการวิจัยและข้อเสนอแนะ

งานวิจัยนี้ได้นำความรู้ทางด้านปัญญาประดิษฐ์ โดยเฉพาะเทคนิคการเรียนรู้ของเครื่อง ได้แก่ แบบจำลองการสุ่มป่าไม้, เครื่องเวกเตอร์สนับสนุน และการถดถอยโลจิสติก มาประยุกต์เพื่อพัฒนาเป็นแบบจำลองสำหรับการทำนายทิศทางการเปลี่ยนแปลงของราคาอ้างอิงของสกุลเงินในตลาดการซื้อขายแลกเปลี่ยนเงินตราต่างประเทศ ซึ่งแบบจำลองนี้จะช่วยให้นักลงทุนสามารถกำหนดทิศทางการลงทุนได้ นอกจากนี้ยังได้นำเสนอวิธีการต่างๆ ในการเพิ่มประสิทธิภาพให้กับแบบจำลอง ซึ่งทำให้แบบจำลองมีประสิทธิภาพในการทำนายข้อมูลได้อย่างแม่นยำมากขึ้น ในด้านของการสร้างผลกำไรในการลงทุนซื้อขายแลกเปลี่ยนเงินตราต่างประเทศ ที่สำคัญที่สุดงานวิจัยนี้ได้วัดประสิทธิภาพด้านการจำลองการเทรดจากข้อมูลย้อนหลัง (Backward Test) จากผลการศึกษาทั้งหมด สามารถสรุปผลการวิจัยได้ดังนี้

5.1 สรุปผลการวิจัย

ในงานวิจัยนี้ผู้วิจัยได้ใช้ชุดข้อมูลราคาอ้างอิงของสกุลเงิน EURUSD ในตลาดการซื้อขายเงินตราต่างประเทศจากโปรแกรม MT4 ที่กรอบเวลา 1 นาที ช่วงเวลา 2 มกราคม พ.ศ.2564 ถึง 31 ธันวาคม พ.ศ.2564 โดยแบ่งเป็น 3 กลุ่ม ได้แก่ แบบจำลองพื้นฐาน, แบบจำลองพื้นฐานร่วมกับตัวชี้วัดทางเทคนิค และแบบจำลองที่ผ่านการปรับพารามิเตอร์ โดยมีข้อมูลเริ่มต้นเป็นราคาเปิด ราคาสูง ราคาต่ำ และราคาปิด โดยใช้อัลกอริทึมที่หลากหลาย ได้แก่ แบบจำลองการสุ่มป่าไม้, เครื่องเวกเตอร์สนับสนุน และแบบจำลองการถดถอยโลจิสติก นอกจากนี้ผู้วิจัยได้ใช้เทคนิค GridSearchCV ปรับจูนพารามิเตอร์ เพื่อให้ได้ผลลัพธ์การทำนายที่ดียิ่งขึ้น

จากการทดลองพบว่า แบบจำลองมีค่าความแม่นยำสูงขึ้น เมื่อเพิ่มตัวชี้วัดทางเทคนิคร่วมกับข้อมูลที่ใช้ในการเรียนรู้และทดสอบ และเมื่อปรับพารามิเตอร์แต่ละแบบจำลอง ส่งผลให้แบบจำลองมีค่าความแม่นยำสูงขึ้นอีกเล็กน้อย โดยแบบจำลองที่มีค่าความแม่นยำสูงสุดคือเครื่องเวกเตอร์สนับสนุน มีค่าความแม่นยำเท่ากับ 92.68% แบบจำลองที่มีค่าความแม่นยำถัดมาคือ แบบจำลองการถดถอยโลจิสติก มีค่าความแม่นยำเท่ากับ 92.19% และสุดท้ายคือแบบจำลองการสุ่มป่าไม้ มีค่าความแม่นยำเท่ากับ 92.10%

หลังจากนั้นผู้ทำวิจัยนำแบบจำลองที่ผ่านการฝึกฝนแล้วนั้น มาทำนายแนวโน้มของราคาอ้างอิงของสกุลเงิน EURUSD ช่วงเวลา 2 มกราคม พ.ศ.2565 ถึง 31 ธันวาคม พ.ศ.2565 และสร้างเป็นสัญญาณซื้อขาย เพื่อจำลองการเทรดแบบย้อนกลับที่เงินลงทุน 1,000 ดอลลาร์ ผลลัพธ์ที่ได้คือแบบจำลองการสุ่มป่าไม้ร่วมกับตัวชี้วัดทางเทคนิคที่ไม่ได้ปรับพารามิเตอร์ใดๆ ให้ผลกำไรสูงสุดที่ 114.28% ให้ผลขาดทุนสูงสุดเท่ากับ 4.36% และมีเงินทุนสุดท้ายเท่ากับ 2,142.8 ดอลลาร์ ส่วน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

แบบจำลองอื่นๆ ที่ให้ผลกำไรรองลงมาคือ แบบจำลองการถดถอยโลจิสติกร่วมกับตัวชี้วัดทางเทคนิค ให้ผลกำไรเท่ากับ 89.26%

5.2 ข้อเสนอแนะ

จากผลการศึกษาที่ได้ในงานวิจัยนี้ สามารถสรุปเป็นประเด็นต่างๆ ที่เป็นการเสนอแนะเพื่อการต่อยอดงานวิจัยได้ดังต่อไปนี้

1) ในงานวิจัยนี้ผู้วิจัยใช้ข้อมูลการซื้อขายแลกเปลี่ยนเงินตราต่างประเทศคู่สกุลเงิน EURUSD ในการทำวิจัย แต่สิ่งที่ได้จากงานวิจัยนี้สามารถนำไปประยุกต์ได้กับข้อมูลอื่นๆ เพื่อเพิ่มโอกาสในการลงทุนที่หลากหลายให้กับนักลงทุน ยกตัวอย่างเช่น ข้อมูลการซื้อขายสินทรัพย์ในตลาดหุ้นและตลาดอนุพันธ์ ข้อมูลการซื้อขายแลกเปลี่ยนโลหะมีค่า เช่น เงินและทองคำ ข้อมูลการซื้อขายแลกเปลี่ยนสกุลเงินดิจิตอล เป็นต้น
2) ในขั้นตอนการกำหนดคลาสเป้าหมายให้แก่ข้อมูล ผู้จัดทำได้ใช้เพียงเส้น MA10 และ MA60 ในการกำหนดว่าเป็นเทรนขาขึ้นหรือเทรนขาลง ซึ่งสามารถต่อยอดใช้ช่วงเวลาอื่นๆ หรือใช้ตัวชี้วัดทางเทคนิคเพิ่มเติม เพื่อให้การทำนายมีประสิทธิภาพดียิ่งขึ้น
3) ด้านการทดสอบด้วยรูปแบบการจำลองการเทรดโดยใช้ข้อมูลย้อนหลังนั้น ถึงแม้จะได้ผลลัพธ์เป็นที่น่าพอใจ แต่การเทรดในสภาวะแวดล้อมด้วยเงินลงทุนจริงนั้น จิตวิทยาในการลงทุนเป็นสิ่งที่สำคัญ นักลงทุนอาจมีความกังวลว่าจะเทรดแพ้และทำให้ขาดทุน จึงทำให้บางครั้งไม่ได้ทำตามข้อกำหนดและเงื่อนไขในการเทรดที่กลยุทธ์ได้กำหนดไว้ เป็นเหตุให้เกิดการขาดทุนได้ ดังนั้นก่อนที่นักลงทุนจะนำไปใช้งานจริง ควรทำการทดสอบแบบไปข้างหน้า (Forward test) ในพอร์ตจำลองก่อน
4) ถึงแม้ว่างานวิจัยนี้มีเป้าหมายเพื่อพัฒนาเครื่องมือและกลยุทธ์เพื่อใช้เป็นเครื่องมือในการทำกำไรให้กับนักลงทุน แต่แบบจำลองและวิธีการเพิ่มประสิทธิภาพให้แบบจำลองที่ใช้ในการทำนายทิศทางการเคลื่อนไหวของราคาที่จะเกิดขึ้นในอนาคตนั้น สามารถนำไปประยุกต์ใช้ในการลงทุนประเภทอื่นๆ ได้ แต่เนื่องจากรูปแบบการลงทุนแต่ละประเภทมีความแตกต่างกัน ดังนั้นอีกด้านหนึ่งของการต่อยอดงานวิจัยนี้ คือ การคิดค้นและพัฒนากลยุทธ์การลงทุนที่เหมาะสมกับการลงทุนประเภทอื่นๆ
5) โปรดใช้วิจารณญาณในการนำไปใช้งานเทรดลงทุนต่างๆ จากงานวิจัยนี้ จากผลการทดลองจำลองการเทรดแบบย้อนกลับที่ให้ผลตอบแทนสูงและมีความเสี่ยงต่ำนั้น ไม่ได้รับประกันว่าผู้ใช้งานจะสามารถนำไปใช้แล้วได้ผลลัพธ์ตามที่รายงาน เพราะอาจมีปัจจัยอื่นๆ ที่เกี่ยวข้อง รวมถึงผลการจำลองการเทรดในอดีต ไม่สามารถเป็นสิ่งยืนยันถึงผลลัพธ์ในอนาคตได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เอกสารอ้างอิง

กาญจน์เขจร ชูชีพ. 2018. “การถดถอยโลจิสติก.” *Remote Sensing Technical Note*, No.5

มนสิช จันทนปุ่ม. “ความเสี่ยงในการเล่นหุ้นของนักลงทุนรายย่อย.” [Online]. Avaliable : https://www.finnomena.com/mod_mangmaoclub/type-of-risk/. 2016.

Abbey, B. S., Doukas, J. A. 2012. “Is Technical Analysis Profitable for Individual Currency Traders?” *Journal of Portfolio Management*, 39(1), 142.

Baasher, A. A., Fakhr, M. W. 2012. “Forex Trend Classification using Machine Learning Techniques.” *In Recent Researches in Applied Informatics and Remote Sensing*, 41-46.

Bontempi, G., Taieb, S. B., Leborgne, Y. A. 2012. “Machine Learning Strategies for Time Series Forecasting.” *In European Business Intelligence Summer School*. 62-77.

Chen, R. C., Shiao, Y. C., Dewi, C., Liu, Q. E. 2020. “Predicting The Japanese Yen to US Dollar Exchange Rate Based on Machine Learning Models” *International Computer Symposium (ICS)*, 141-146.

Daddy Trader. “Moving Average ถ้าอยากใช้ ต้องรู้อะไรบ้าง” [Online]. Avaliable : https://www.finnomena.com/daddy-trader/moving-average/. 2018.

Hsu, C. W., Lin, C. J. 2002. “A Comparison of Methods for Multiclass Support Vector Machines.” *IEEE Transactions on Neural Networks*, 13(2), 415-425.

Ismail, Z., Jamaluddin, F., Jamaludin, F. 2008. “Time Series Regression Model for Forecasting Malaysian Electricity Load Demand.” *Asian J. Math. Stat*, 1(3), 139-149

Jacque, L. L. 2013. Management and Control of Foreign Exchange Risk. *Springer Science & Business Media.* https://doi.org/10.1007/978-94-009-1806-1

Kaltwasser, P. R. 2010. “Uncertainty About Fundamentals and Herding Behavior in The FOREX Market.” *Physica A: Statistical Mechanics and Its Applications*, 389(6), 1215-1222.

Lai, R. K., Fan, C. Y., Huang, W. H., Chang, P. C. 2009. “Evolving and Clustering Fuzzy Decision Tree for Financial Time Series Data Forecasting.” *Expert Systems with Applications*, 36(2), 3761-3773.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Lerman, K., Gilder, A. 2009. “System and Method for Forecasting Fluctuations in Future Data and Particularly for Forecasting Security Prices by News Analysis.” *Google Patents.*

Liu, Z., Xiao, D. 2009. “An Automated Trading System with Multi-Indicator Fusion Based on D-S Evidence Theory in Forex Market.” *International Conference on Fuzzy Systems and Knowledge Discovery*, 239-243.

Parray, I. R., Khurana, S. S., Kumar, M., Altalbe, A. A. 2020. “Time Series Data Analysis of Stock Price Movement Using Machine Learning Techniques.” *Soft Computing*, 24, 16509-16517.

Pongsena, P., Ditsayabut, P., Kerdprasop N., Kerdprasop. K. 2018. "An Analysis of the Co-movement of Price Change Volatility in Forex Market." *Proceedings of The World Congress on Engineering 2018*, London, U.K.

Scholkopf, B., Smola, A. J. 2001. “Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond.” *MIT Press*.

Shu, C., He, D., Cheng, X. 2015. “One Currency, Two Markets: The Renminbi’s Growing Influence in Asia-Pacific.” *China Economic Review*, 33, 163-178.

Suykens, J. A. K., Vandewalle, J. 1999. “Least Squares Support Vector Machine Classifiers.” *Neural Processing Letters*, 9(3), 293-300.

Thu, T. N. T., Xuan, V. D. 2018. “Using Support Vector Machine in Forex Predicting.” *In Proceedings of the 2018 IEEE International Conference on Innovative Research and Development,* 1-5.

Trading Strategy Guides. “Williams Percent Range Strategy – 2 Approaches.” [Online]. Available : https://tradingstrategyguides.com/williams-percent-range-strategy. 2019.

Tsay, R. S. 2005. “Analysis of Financial Time Series.” *John Wiley & Sons.* Vol. 543.

Wei, W. W. S. 2006. “Time Series Analysis.” *The Oxford Handbook of Quantitative Methods in Psychology*: Vol. 2.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

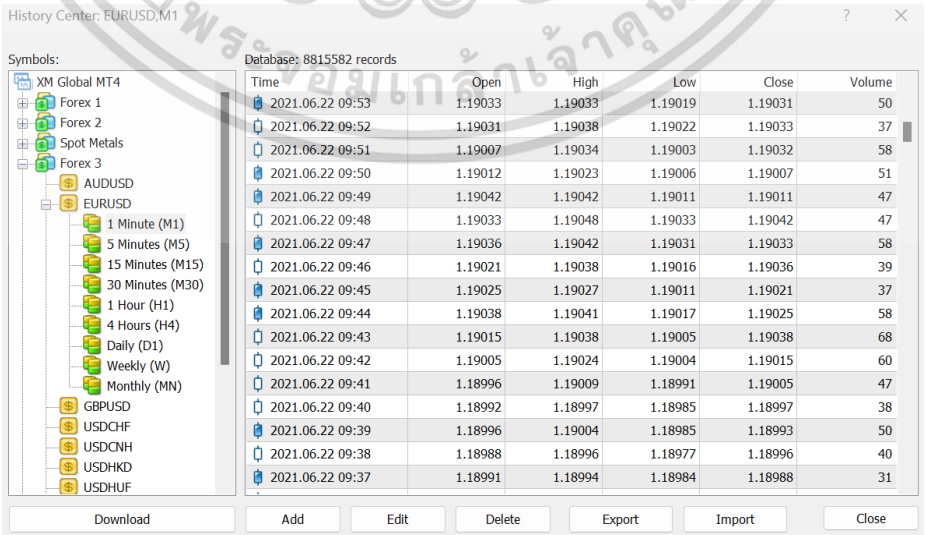
ภาคผนวก ก

ตัวอย่างการเก็บรวบรวมข้อมูล

ภาคผนวก ก.1 แสดงตัวอย่างโปรแกรมการรวบรวมข้อมูล

โปรแกรมหลักที่ใช้ในการเก็บรวบรวมข้อมูลคือ Metatrader 4 แสดงดังรูปที่ ก.1 จากนั้นไปที่คำสั่ง History Center และค้นหาคู่เงิน EURUSD จากนั้นกด Export เพื่อรวบรวมข้อมูลในรูปแบบไฟล์ CSV แสดงดังรูปที่ ก.2

รูปที่ ก.1 โปรแกรม MetaTrader 4 (MT4)



รูปที่ ก.2 Export ข้อมูลคู่เงิน EURUSD จาก History Center

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ก.2 แสดงการนำข้อมูลเข้าสู่ Google Colab

นำไฟล์ CSV ไปวางไว้ใน Google Drive ที่บัญชีเดียวกันกับ Google Colab จากนั้นใช้ Library pandas และ drive ในการเรียกข้อมูลเข้าสู่ Google Colab โดยขั้นตอนการดำเนินการแสดงดังรูปที่ ก.3

```
# import data from drive
from google.colab import drive
drive.mount('/content/drive')
```

```
Mounted at /content/drive
```

```
# Path to CSV
path = "/content/drive/MyDrive/Colab Notebooks/IS/Dataset/DAT_MT_EURUSD_M1_2021.csv"
```

```
df = pd.read_csv(path)
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 369502 entries, 0 to 369501
Data columns (total 6 columns):
 #   Column  Non-Null Count   Dtype
---  ------  --------------   -----
 0   Date    369502 non-null  object
 1   Time    369502 non-null  object
 2   Open    369502 non-null  float64
 3   High    369502 non-null  float64
 4   Low     369502 non-null  float64
 5   Close   369502 non-null  float64
dtypes: float64(4), object(2)
memory usage: 16.9+ MB
```

รูปที่ ก.3 ขั้นตอนในการนำข้อมูลเข้าสู่ Google Colab

จากรูปที่ ก.3 เมื่อนำข้อมูลดิบเข้าระบบเรียบร้อย ขั้นตอนถัดมาคือดำเนินการจัดการข้อมูลวันและเวลาให้อยู่ในคอลัมน์เดียวกัน โดยขั้นตอนการดำเนินการแสดงดังรูปที่ ก.4

```
# Set index date time
df['Datetime'] = df['Date']+ ' ' +df['Time']
df.drop(['Date', 'Time'], 1, inplace=True)

# แปลงคอลัมน์ date เป็น datetime object
df['Datetime'] = pd.to_datetime(df['Datetime'], format='%Y.%m.%d %H:%M')

# เปลี่ยนรูปแบบวันที่และเวลาเป็น YYYY-MM-DD hh:mm
df['Datetime'] = df['Datetime'].apply(lambda x: x.strftime('%Y-%m-%d %H:%M'))
```

รูปที่ ก.4 ขั้นตอนดำเนินการกับวันที่และเวลาข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ข

ตัวอย่างการทดสอบประสิทธิภาพแบบจำลอง

ภาคผนวก ข.1 แสดงตัวอย่างการกำหนดคลาสเป้าหมาย

หลังจากเตรียมข้อมูลดิบเรียบร้อย ขั้นตอนถัดมาคือการกำหนดคลาสเป้าหมายให้แก่ข้อมูล โดยใช้เส้น SMA 10 และ SMA 60 โดยขั้นตอนการดำเนินการแสดงดังรูปที่ ข.1

```
[ ]  1 def create_target(ldf,tr_id=False):
     2     ldf['SMA1'] = ldf['Close'].rolling(window=10, min_periods=1, center=False).mean()
     3     ldf['SMA2'] = ldf['Close'].rolling(window=60, min_periods=1, center=False).mean()
     4     ldf['signal'] = np.where(ldf['SMA1'] > ldf['SMA2'], 1.0, 0.0) # Create signals
     5     if(tr_id is not True):
     6         display(ldf['signal'].value_counts())
     7
     8 df_tr1 = df_tr.copy()  # Save the Baseline Model Dataframe [Training Set]
     9 df_te1 = df_te.copy() # Save the Baseline Model Dataframe [Test Set]
    10 create_target(df_tr1)  # Add target variable to Training Set
    11 create_target(df_te1,tr_id=True)  # Add target variable to Test Set
```

รูปที่ ข.1 ขั้นตอนการกำหนดคลาสเป้าหมาย

ภาคผนวก ข.2 แสดงตัวอย่างการเรียนรู้และทดสอบแบบจำลอง

หลังจากกำหนดคลาสเป้าหมายดัง ภาคผนวก ข.1 เรียบร้อย ขั้นตอนถัดมาคือนำข้อมูลให้แบบจำลองฝึกฝน และทดสอบแบบจำลอง โดยขั้นตอนการดำเนินการแสดงดังรูปที่ ข.2 และ ข.3 ตามลำดับ

```
1 %%time
2
3 # Instantiate the model
4 RF_model1 = RandomForestClassifier()
5
6 # Fit the model to the training set
7 RF_model1.fit(X_train1, y_train1)

CPU times: user 1min 45s, sys: 1.09 s, total: 1min 46s
Wall time: 1min 50s
▾ RandomForestClassifier
RandomForestClassifier()
```

รูปที่ ข.2 ตัวอย่างขั้นตอนการฝึกฝนแบบจำลองการสุ่มป่าไม้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
[ ]  1 # Predict X_test
     2
     3 y_pred1 = RF_model1.predict(X_test1)

[ ]  1 # Check Accuracy Score
     2
     3 print('Model Accuracy Score: {0:0.4f}'. format(accuracy_score(y_test1, y_pred1)))

    Model Accuracy Score: 0.5061
```

รูปที่ ข.3 ตัวอย่างขั้นตอนการทดสอบแบบจำลองการสุ่มป่าไม้

ขั้นตอนสุดท้ายคือการปรับปรุงให้แบบจำลองมีประสิทธิภาพดีขึ้นโดยปรับพารามิเตอร์ต่างๆ และใช้วิธี GridsearchCV เพื่อหาพารามิเตอร์ที่เหมาะสมที่สุด โดยขั้นตอนต่างๆ แสดงดังรูปที่ ข.4

```
     1 # Number of trees in random forest
     2 n_estimators = [50, 100, 300, 500]
     3 # Maximum number of levels in tree
     4 max_depth = [None, 2, 4]
     5 # Minimum number of samples required to split a node
     6 min_samples_split = [2, 5]
     7 # Minimum number of samples required at each leaf node
     8 min_samples_leaf = [1, 2]

[ ]  1 # Create the param grid
     2 param_grid = {'n_estimators': n_estimators,
     3                'max_depth': max_depth,
     4                'min_samples_split': min_samples_split,
     5                'min_samples_leaf': min_samples_leaf,
     6 print(param_grid)

[ ]  1 # Finding optimal hyperparameters(GridSearchCV)
     2 %%time
     3
     4 # Define model
     5 model = RandomForestClassifier()
     6
     7 # Define search
     8 search = GridSearchCV(estimator = model, param_grid = param_grid, n_jobs = -1, cv=5, verbose=2)
     9
    10 # Execute search
    11 GridSearchCV = search.fit(X_train2, y_train2)
    12
    13 # Set the clf to the best combination of parameters
    14 RF_modelcv = GridSearchCV.best_estimator_
    15
    16 # Summarize result
    17 print('Best Score: %s' % GridSearchCV.best_score_)
    18 print('Best Hyperparameters: %s' % GridSearchCV.best_params_)

    Fitting 2 folds for each of 96 candidates, totalling 192 fits
    Best Score: 0.912372472041971
    Best Hyperparameters: {'bootstrap': True, 'max_depth': None, 'min_samples_leaf': 2, 'min_samples_split': 2, 'n_estimators': 500}
    CPU times: user 16min 29s, sys: 11.1 s, total: 16min 40s
    Wall time: 5h 7min 46s
```

รูปที่ ข.4 ตัวอย่างขั้นตอนการปรับพารามิเตอร์ของแบบจำลองการสุ่มป่าไม้ด้วย GridsearchCV

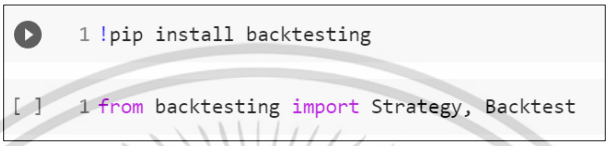
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ค

ตัวอย่างการทดสอบแบบย้อนกลับ

ภาคผนวก ค.1 แสดงตัวอย่างการนำเข้า Backtesting Library

การทดสอบแบบย้อนกลับในงานวิจัยนี้ ใช้วิธีจาก Backtesting.py ซึ่งใน Google Colab ไม่มี จึงต้องมีการ Import โดยขั้นตอนแสดงดังรูปที่ ค.1



รูปที่ ค.1 ขั้นตอนการนำ Library เข้าสู่ Google Colab

จากนั้นกำหนดคลาสการทดสอบแบบย้อนกลับ ตามขั้นตอนแสดงดังรูปที่ ค.2

```
[ ]  1 # Define class
     2 class My_backtest(Strategy):
     3     def init(self):
     4         super().init()
     5         self.signal1 = self.I(SIGNAL)
     6
     7     def next(self):
     8         super().next()
     9         if self.signal1==1 and not self.position.is_long: # เปิดไม้ Buy
    10             sl1 = self.data.Close[-1]
    11             tp1 = self.data.Close[-1]
    12             self.buy(size=0.2, sl=sl1, tp=tp1)
    13         elif self.signal1==-1 and not self.position.is_short:# เปิดไม้ Sell
    14             sl1 = self.data.Close[-1]
    15             tp1 = self.data.Close[-1]
    16             self.sell(size=0.2, sl=sl1, tp=tp1)
```

รูปที่ ค.2 ขั้นตอนการกำหนดคลาสการทดสอบแบบย้อนกลับ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หลังจากกำหนดคลาสการทดสอบแบบย้อนกลับเรียบร้อย และรันขั้นตอนเสร็จสมบูรณ์ จะได้ผลลัพธ์แสดงดังรูปที่ ค.3 โดยข้อมูลที่สำคัญต่างๆ คือ เปอร์เซ็นต์ผลกำไร (%Return), เปอร์เซ็นต์ผลขาดทุนสูงสุด (%Max Drawdown) และจำนวนการเทรด (Trade)

```
Start                     2022-01-02 20:28:00
End                       2022-12-30 16:58:00
Duration                    361 days 20:30:00
Exposure Time [%]                    9.585873
Equity Final [$]                  2142.805504
Equity Peak [$]                   2142.805504
Return [%]                          114.28055
Buy & Hold Return [%]               -5.788033
Return (Ann.) [%]                   84.973784
Volatility (Ann.) [%]               18.985409
Sharpe Ratio                         4.475742
Sortino Ratio                       16.178958
Calmar Ratio                        19.466492
Max. Drawdown [%]                   -4.365131
Avg. Drawdown [%]                   -0.114933
Max. Drawdown Duration      27 days 19:46:00
Avg. Drawdown Duration       0 days 07:20:00
# Trades                                19546
Win Rate [%]                        50.644633
Best Trade [%]                       0.377017
Worst Trade [%]                     -0.382574
Avg. Trade [%]                       0.001322
Max. Trade Duration          2 days 00:05:00
Avg. Trade Duration          0 days 00:02:00
Profit Factor                        1.24776
Expectancy [%]                       0.001324
SQN                                  6.699031
_strategy                      My_backtest
_equity_curve                          ...
_trades                      Size  Ent...
dtype: object
```

รูปที่ ค.3 ผลการทดสอบแบบจำลองพื้นฐานการสุ่มป่าไม้ร่วมกับตัวชี้วัดทางเทคนิค

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประวัติผู้เขียน

ชื่อ	นายธชณัฏฐ์ ตันพิทักษ์สิทธิ์
วัน เดือน ปี เกิด	30 กันยายน พ.ศ.2531
ที่อยู่ปัจจุบัน	บ้านเลขที่ 142/19 หมู่ 6 ตำบลหนองปรือ อำเภอบางละมุง จังหวัดชลบุรี
ประวัติการศึกษา	(2557) วิศวกรรมศาสตร์มหาบัณฑิต สาขาวิศวกรรมเคมี เกรดเฉลี่ย 3.50 (จุฬาลงกรณ์มหาวิทยาลัย)
	(2554) วิศวกรรมศาสตร์บัณฑิต สาขาวิศวกรรมเคมี เกรดเฉลี่ย 2.87 (มหาวิทยาลัยศรีนครินทรวิโรฒ)
ทุนการศึกษาที่ได้รับ	ไม่มี
ผลงานทางวิชาการ	(2558) “Effects of benzoxazine resin on property enhancement of shape memory epoxy: A dual function of benzoxazine resin as a curing agent and a stable network segment”
	(2558) “Mechanical properties and shape recovery performance of aliphatic epoxy-benzoxazine based shape memory polymers
	(2557) “Shape fixity and shape recovery characteristics of aliphatic epoxy-benzoxazine shape memory polymer”

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้