

การสร้างตัวแบบการพยากรณ์ราคากองทุนรวมอีทีเอฟทองคำด้วยเทคนิคการเรียนรู้ของเครื่อง

Forecasting Gold ETF Fund Prices Using Machine Learning Techniques

อาทิตยา ชาวดอน¹, จิรัชค ทองดี¹, ทักษพร สมพิทักษ์¹, เบญญาดา สุนทรภักดี¹, อนุปงศ์ สุขประเสริฐ*

Arthitaya Chaodon¹, Jirachoke Thongdee¹, Thaksaporn Sompithak¹, Benyada Soontornphak¹,
Anupong Sukprasert*

*Corresponding author: Anupong Sukprasert (anupong.s@acc.msu.ac.th)

บทคัดย่อ

งานวิจัยฉบับนี้มีวัตถุประสงค์เพื่อศึกษาประสิทธิภาพของเทคนิคเหมืองข้อมูล โดยการสร้างตัวแบบเพื่อพยากรณ์ราคากองทุนรวมอีทีเอฟทองคำ จำนวนข้อมูลทั้งหมด 1,718 แถว 13 ตัวแปร และนำข้อมูลมาวิเคราะห์ตามมาตรฐานในการทำเหมืองข้อมูล (CRISP-DM) การวิเคราะห์ข้อมูลครั้งนี้ผู้วิจัยทำการแบ่งชุดข้อมูลด้วยวิธีการ 10-fold Cross Validation สำหรับข้อมูลสอนและข้อมูลทดสอบ และทำการวิเคราะห์ข้อมูลด้วยเทคนิคการทำเหมืองข้อมูลทั้งหมด 4 เทคนิคประกอบด้วย เทคนิคสมการถดถอยเชิงเส้น (Linear Regression) เทคนิคเพื่อนบ้านใกล้ที่สุด (k-Nearest Neighbors) เทคนิคโครงข่ายประสาทเทียม (Neural Network) และทำการเพิ่มประสิทธิภาพของตัวแบบด้วยเทคนิควิธีการรวมกลุ่ม (Vote Ensemble) วิธีการหาค่าที่เหมาะสมที่สุด (Optimization) ด้วยวิธีด้านวิวัฒนาการ (Evolutionary Algorithms) เพื่อค้นหาค่าที่เหมาะสมที่สุดสำหรับชุดของพารามิเตอร์ในแต่ละตัวแบบและทำการเปรียบเทียบประสิทธิภาพการจำแนกประเภทข้อมูลด้วยค่ารากที่สองของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Root Means Square Error) ค่าวัดสัมประสิทธิ์การตัดสินใจ (Squared Correlation) ค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Squared Error) ค่าความคลาดเคลื่อนสัมบูรณ์ (Absolute Error) และค่าความคลาดเคลื่อนกำลังสอง (Squared Error) โดยใช้โปรแกรม RapidMiner Studio Version 10.3.1 ในการสร้างตัวแบบและการวิเคราะห์ข้อมูล ผลการวิจัยพบว่า เทคนิคการรวมกลุ่ม (Vote Ensemble) เป็นเทคนิคความเหมาะสมมากที่สุดในการสร้างตัวแบบสำหรับการพยากรณ์ราคากองทุนรวมอีทีเอฟทองคำโดยมีค่ารากที่สองของค่าคลาดเคลื่อนกำลังสองเฉลี่ย เท่ากับ 1.676 ค่าสัมประสิทธิ์การตัดสินใจ เท่ากับ 99.10% ค่าคลาดเคลื่อนกำลังสองเฉลี่ย เท่ากับ 2.81 ค่าความคลาดเคลื่อนสัมบูรณ์ เท่ากับ 1.284 ค่าความคลาดเคลื่อนกำลังสอง เท่ากับ 2.815 เทคนิคการรวมกลุ่มสามารถจัดการข้อมูลที่มีขนาดใหญ่ได้ดีและแสดงให้เห็นว่ามีความแม่นยำมากที่สุดเมื่อเปรียบเทียบกับเทคนิคอื่น ๆ ผลการศึกษานี้สามารถนำไปประยุกต์ใช้ในการตัดสินใจลงทุน รวมถึงช่วยเสริมสร้างกลยุทธ์การบริหารความเสี่ยงในตลาดการเงินที่มีความผันผวนสูงที่จะเกิดขึ้นในอนาคตได้

คำสำคัญ : การพยากรณ์ราคา, Gold ETF, เทคนิคเหมืองข้อมูล, การเรียนรู้ของเครื่อง, เทคนิคการรวมกลุ่ม, การวิเคราะห์ข้อมูล, การลงทุน, การบริหารความเสี่ยง

Abstract

This research aims to study the efficiency of data mining techniques by developing models to forecast the prices of Gold ETF funds. The dataset consists of 1,718 rows and 13 variables, analyzed according to the standard data mining process (CRISP-DM). The researcher applied a 10-fold Cross Validation method to split the dataset into training and testing sets and employed four data mining techniques: Linear Regression, k-Nearest Neighbors, Neural Networks, and an enhanced model using the Vote Ensemble technique. Additionally, model optimization was conducted using Evolutionary Algorithms to determine the optimal parameters for each model. The models' performance was compared using Root Mean Square Error (RMSE), Squared Correlation, Mean Squared Error (MSE), Absolute Error, and Squared Error. RapidMiner Studio Version 10.3.1 was used for model creation and data analysis. The results indicate that the Vote Ensemble technique is the most suitable for forecasting Gold ETF prices. It achieved an RMSE of 1.676, a Squared Correlation of 99.10%, an MSE of 2.81, an Absolute Error of 1.284, and a Squared Error of 2.815. The Vote Ensemble technique effectively handles large datasets and demonstrated the highest accuracy compared to other techniques. These findings can be applied to investment decision-making and risk management strategies in highly volatile financial markets.

Keywords: Price Forecasting, Gold ETF, Data Mining Techniques, Machine Learning, Ensemble Techniques, Data Analysis, Investment, Risk Management.

บทนำ

ทองคำเป็นหนึ่งในสินทรัพย์ที่เก่าแก่ที่สุดที่มนุษย์ใช้เพื่อรักษามูลค่า และได้รับการยอมรับว่าเป็นสินทรัพย์ปลอดภัย (safe-haven asset) ในช่วงที่เศรษฐกิจมีความผันผวน (Baur & McDermott, 2010) ในอดีต ทองคำเคยถูกใช้เป็นสกุลเงินในหลายประเทศ รวมถึงสหรัฐอเมริกา ปัจจุบัน ธนาคารกลางทั่วโลกยังคงถือครองทองคำเพื่อเป็นหลักประกันในการชำระหนี้ระหว่างประเทศและควบคุมอัตราเงินเฟ้อ ซึ่งสะท้อนถึงเสถียรภาพทางการเงินของประเทศ (World Gold Council, 2022) ในช่วงไม่กี่ปีที่ผ่านมา ประเทศเศรษฐกิจเกิดใหม่ เช่น จีน รัสเซีย และอินเดีย ได้กลายเป็นผู้ซื้อทองคำรายใหญ่ ขณะที่สหรัฐอเมริกา แอฟริกาใต้ และออสเตรเลียเป็นผู้ขายทองคำหลักในตลาด

ในช่วงสองทศวรรษที่ผ่านมา การลงทุนในทองคำได้เปลี่ยนแปลงไปอย่างมาก โดยเฉพาะหลังจากการเปิดตัว Gold Exchange-Traded Funds (Gold ETFs) ในช่วงต้นปี 2000 (Shafiee & Topal, 2010) Gold ETF เป็นเครื่องมือทางการเงินที่ช่วยให้นักลงทุนสามารถเข้าถึงตลาดทองคำโดยไม่ต้องถือครองทองคำจริง ทำให้การลงทุนสะดวกขึ้นและมีต้นทุนการทำธุรกรรมต่ำลง หนึ่งใน Gold ETF ที่ใหญ่ที่สุด คือ SPDR Gold Shares (GLD) ซึ่งเปิดตัวในปี 2004 และได้กลายเป็นเครื่องมือสำคัญสำหรับนักลงทุนในการเข้าถึงตลาดทองคำ (Kumar et al., 2018)

การพยากรณ์ราคาของ Gold ETF มีความสำคัญอย่างยิ่งสำหรับนักลงทุน เนื่องจากช่วยให้สามารถตัดสินใจได้อย่างมีประสิทธิภาพว่าจะซื้อหรือขายสินทรัพย์นี้เมื่อใด อย่างไรก็ตาม การพยากรณ์ราคาทองคำเป็นงานที่ท้าทาย เพราะราคาทองคำได้รับอิทธิพลจากหลายปัจจัย เช่น ราคาโลหะมีค่าอื่นๆ ราคาน้ำมันดิบ ดัชนีตลาดหุ้น ราคาพันธบัตร อัตราแลกเปลี่ยน และปัจจัยทางเศรษฐกิจอื่นๆ (Ghosh et al., 2004)

งานวิจัยนี้มุ่งเน้นไปที่การพัฒนาและเปรียบเทียบประสิทธิภาพของแบบจำลองการเรียนรู้ของเครื่องในการพยากรณ์ราคา Gold ETF เพื่อเพิ่มความสามารถในการวิเคราะห์การพยากรณ์ให้มีความแม่นยำมากยิ่งขึ้น ผลลัพธ์ของการวิจัยนี้คาดว่าจะประโยชน์ต่อนักลงทุน หรือนักวิเคราะห์ทางการเงิน ที่ต้องการใช้ข้อมูลการพยากรณ์ราคาทองคำในการตัดสินใจลงทุน รวมถึงช่วยเสริมสร้างกลยุทธ์การบริหารความเสี่ยงในตลาดการเงินที่มีความผันผวนสูง

วัตถุประสงค์

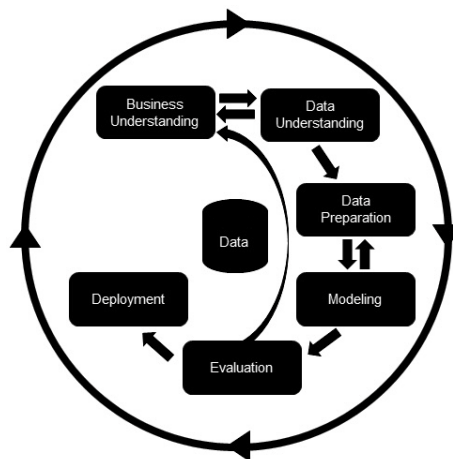
1. เพื่อสร้างตัวแบบสำหรับการพยากรณ์ราคากองทุนรวมอิตีเอฟทองคำ
2. เพื่อเปรียบเทียบตัวแบบที่ใช้สำหรับพยากรณ์ราคากองทุนรวมอิตีเอฟทองคำ
3. เพื่อเพิ่มประสิทธิภาพตัวแบบที่ใช้สำหรับพยากรณ์ราคากองทุนรวมอิตีเอฟทองคำ

วิธีการดำเนินงานวิจัย

งานวิจัยนี้ได้ทำการศึกษาและเก็บรวบรวมข้อมูลเกี่ยวกับราคาทองคำผ่านกระบวนการเตรียมข้อมูล เพื่อให้ได้ชุดข้อมูลที่เหมาะสมและมีคุณภาพพร้อมในการทำเหมืองข้อมูล จากนั้นจะดำเนินการเข้าสู่

กระบวนการสร้างตัวแบบการพยากรณ์ด้วยเทคนิคการจำแนกประเภทข้อมูล ดังนี้ เทคนิคสมการถดถอยเชิงเส้น (Linear Regression) เทคนิคโครงข่ายประสาทเทียม (Neural Network) และ เทคนิคเพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbors : K-NN) และเทคนิคการรวมกลุ่ม (Vote Ensemble) จากนั้นทำการทดสอบประสิทธิภาพตัวแบบการพยากรณ์ด้วยการแบ่งชุดข้อมูลด้วยวิธี 10-folds Cross Validation เพื่อทำการเปรียบเทียบประสิทธิภาพของตัวแบบการพยากรณ์

ข้อมูลที่ผู้วิจัยได้ดำเนินการวิจัยภายใต้กระบวนการของการทำเหมืองข้อมูล (Cross-Industry Standard Process for Data Mining : CRISP-DM) สำหรับการสร้างตัวแบบ การเพิ่มประสิทธิภาพของตัวแบบและเปรียบเทียบเทคนิคเหมืองข้อมูลสำหรับการพยากรณ์ราคาทองคำ ซึ่งประกอบไปด้วย 6 ขั้นตอนดังนี้



ภาพที่ 1 ขั้นตอนวิธีการดำเนินงานวิจัย

1. ขั้นตอนการทำความเข้าใจปัญหา(Business Understanding)

การพยากรณ์ราคากองทุนรวมอีทีเอฟทองคำมีความสำคัญต่อการตัดสินใจลงทุน เนื่องจากได้รับอิทธิพลจากราคาทองคำ ค่าเงินดอลลาร์ อัตราดอกเบี้ย และภาวะเศรษฐกิจ การพัฒนาแบบจำลองพยากรณ์ช่วยลดความเสี่ยง ปรับกลยุทธ์การลงทุน และเพิ่มประสิทธิภาพในการตัดสินใจของนักลงทุนและสถาบันการเงิน

2. ขั้นตอนการทำความเข้าใจข้อมูล (Data Understanding)

งานวิจัยนี้นำชุดข้อมูลการพยากรณ์ราคากองทุนรวมอีทีเอฟทองคำ จาก Manu Siddhartha ข้อมูลสำหรับการศึกษานี้ถูกรวบรวมตั้งแต่วันที่ 18 พฤศจิกายน 2011 ถึงวันที่ 1 มกราคม 2019 จากแหล่งข้อมูลต่างๆ ข้อมูลประกอบด้วยทั้งหมด 1,718 แถว และ 80 แอตทริบิวต์ ข้อมูลสำหรับตัวแปรต่างๆ เช่น ราคาน้ำมัน, ดัชนี S&P 500, ดัชนีดาวโจนส์, อัตราผลตอบแทนพันธบัตรสหรัฐฯ (อายุ 10 ปี), อัตราแลกเปลี่ยนเงินยูโร-ดอลลาร์สหรัฐ, ราคาของโลหะมีค่า เช่น เงินและแพลตตินัม รวมถึงโลหะอื่นๆ เช่น พาลาเดียมและโรเดียม, ราคาของดัชนีดอลลาร์สหรัฐ, Eldorado Gold Corporation และกองทุน ETF ที่เกี่ยวข้องกับเหมืองทองคำ ได้ถูกนำมารวบรวมไว้ที่อยู่ในรูปแบบไฟล์ CSV

3. ขั้นตอนการเตรียมข้อมูล (Data Preparation)

ขั้นตอนการเตรียมข้อมูลเป็นขั้นตอนที่ทำให้เกิดความเชื่อมั่นในคุณภาพข้อมูลที่นำมาใช้ แสดงถึงความเชื่อมั่นของข้อมูลก่อนจะนำไปสร้างตัวแบบการพยากรณ์ในครั้งนี้ โดยผู้วิจัยได้ทำการเตรียมข้อมูลกับชุดข้อมูลทั้งหมด 2 ขั้นตอนดังนี้

3.1. การลดขนาดข้อมูล (Data Reduction)

ผู้วิจัยได้ทำการลดปริมาณและขนาดของข้อมูลซึ่งมีจำนวนมากให้มีขนาดเล็กลงจนเหลือแค่ข้อมูลที่มีความสัมพันธ์ต่อตัวแปร 13 แอตทริบิวต์ เพื่อนำข้อมูลไปวิเคราะห์ ซึ่งมีรายละเอียดดังตารางต่อไปนี้

ตารางที่ 1: ข้อมูลที่ใช้ในงานวิจัย

No	Attribute	Type	Description
1	Date	date	วันที่
2	Adj Close	real	ราคาปิดปรับค่าแล้วของ Gold ETF
3	SP_Ajclose	real	ราคาปิดปรับค่าแล้วของ S&P 500
4	DJ_Ajclose	real	ราคาปิดปรับค่าแล้วของ Dow Jones
5	EU_Price	real	อัตราแลกเปลี่ยนยูโร-ดอลลาร์
6	OF_Price	real	ราคาน้ำมันดิบ (Oil Futures)
7	OS_Price	real	ราคาน้ำมันดิบ Brent (Oil Spot)
8	SF_Price	real	ราคา Silver Futures
9	USB_Price	real	ราคาพันธบัตรรัฐบาลสหรัฐฯ (U.S. Bonds)
10	PLT_Price	real	ราคาแพลตตินัม
11	USDI_Price	real	ดัชนีค่าเงินดอลลาร์สหรัฐฯ (U.S. Dollar Index)
12	GDX_Adj Close	real	ราคาปิดที่ปรับปรุงแล้วของกองทุน ETF ที่ติดตามหุ้นเหมืองทองคำ VanEck Gold Miners ETF (GDX)
13	USO_Adj Close	real	ราคาปิดที่ปรับปรุงแล้วของกองทุน ETF ที่ติดตามราคาน้ำมันดิบ United States Oil Fund, LP (USO)

3.2. กำหนดหน้าที่ให้กับตัวแปร

ผู้วิจัยได้กำหนดหน้าที่ให้กับตัวแปรที่ 2 ราคาปิดที่ปรับแล้วของราคากองทุนรวมอีทีเอฟทองคำ(Adj Close) กำหนดหน้าที่เป็น “Label” หรือตัวแปรตาม (Dependent Variable) เพื่อระบุผลลัพธ์ของราคากองทุนรวมอีทีเอฟทองคำในอนาคต และวันที่ (Date) กำหนดหน้าที่เป็นรหัส “ID”

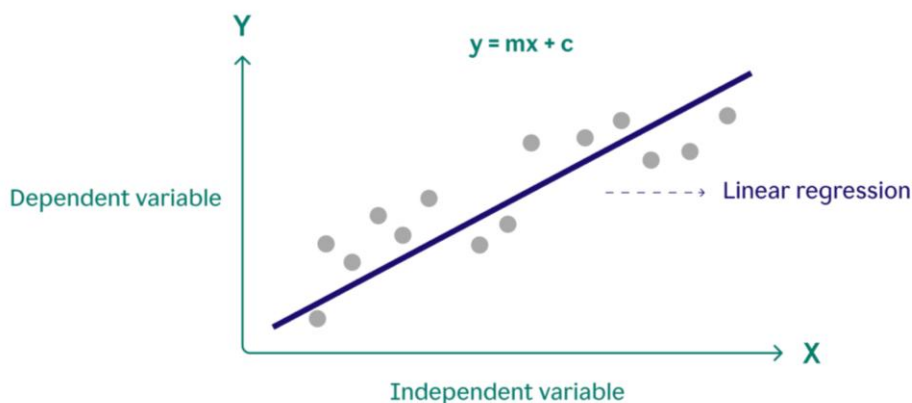
4. ขั้นตอนการสร้างแบบจำลอง (Modeling)

ขั้นตอนนี้เป็นการพัฒนาตัวแบบ ผู้วิจัยได้ใช้โปรแกรม RapidMiner Studio มาทำการสร้างการพัฒนาตัวแบบสำหรับการพยากรณ์ราคากองทุนรวมอีทีเอฟทองคำในอนาคต ด้วยเทคนิคการทำเหมืองข้อมูล 4 เทคนิค

ได้แก่ เทคนิคสมการถดถอยเชิงเส้น (Linear Regression) เทคนิคโครงข่ายประสาทเทียม (Neural Network) เทคนิคเพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbors : K-NN) และเทคนิคการรวมกลุ่ม (Vote Ensemble) วิธีการหาค่าที่เหมาะสมที่สุด (Optimization) ด้วยวิธีด้านวิวัฒนาการ (Evolutionary Algorithms) เพื่อค้นหาค่าที่เหมาะสมที่สุดสำหรับพารามิเตอร์ในแต่ละตัวแบบ โดยการกำหนดค่า K ที่เหมาะสมที่สุดสำหรับเทคนิคเพื่อนบ้านใกล้ที่สุด

4.1. เทคนิคสมการถดถอยเชิงเส้น (Linear Regression)

เทคนิคการศึกษารูปแบบความสัมพันธ์ของตัวแปร เพื่อใช้ทำนายค่าตัวแปรที่ต้องการศึกษา โดยอาศัยความรู้เกี่ยวกับค่าของตัวแปรอื่นที่เกี่ยวข้อง (ตัวแปรที่กำหนด) ซึ่งอาจจะมีหนึ่งตัวหรือมากกว่า ตัวแปรที่ต้องการศึกษา เรียกว่า ตัวแปรตาม (Dependent Variable) มักแทนด้วยตัวแปร Y ตัวแปรที่กำหนดให้ เรียกว่า ตัวแปรอิสระ (Independent variable) มักแทนด้วยตัวแปร X ดังนั้นข้อมูลที่ใช้ศึกษา ประกอบด้วยตัวแปรสองประเภท (1) ตัวแปรอิสระ (Independent variable(s) เป็นตัวแปรที่กำหนดค่าได้แน่นอน หรือสามารถวัดค่าได้โดยไม่มี ความคลาดเคลื่อน (2) ตัวแปรตาม (Dependent variable) เป็นตัวแปรที่เกิดขึ้นโดยสุ่ม รูปแบบการวิเคราะห์นี้ เป็นรูปแบบพื้นฐานที่ง่ายที่สุดของการวิเคราะห์การถดถอยโดยมีตัวแบบการถดถอย

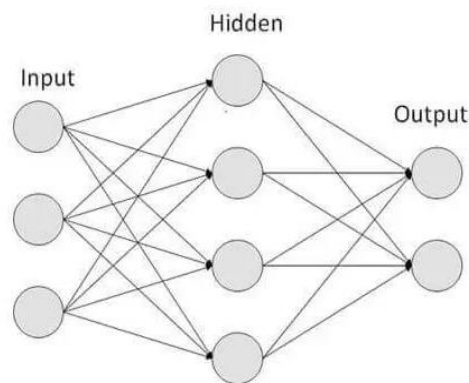


ภาพที่ 2 ตัวอย่างการทำงานสมการถดถอยเชิงเส้น(Linear Regression) (Grammarly, 2024)

4.2. เทคนิคโครงข่ายประสาทเทียม (Neural Network)

เทคนิคโครงข่ายประสาทเทียมหรือที่มักจะเรียกสั้นๆ ว่าโครงข่ายประสาท (Neural Network) คือ โมเดลคณิตศาสตร์ที่เลียนแบบการทำงานของสมองมนุษย์ โดยมีจุดประสงค์เพื่อจำลองการเรียนรู้และการทำนาย ผลงานของมันเกิดจากการเชื่อมโยงของเซลล์ประสาทหรือ “นิวรอน” (neurons) ซึ่งทำงานผ่านกระแสไฟฟ้าเคมี และสามารถเรียนรู้การจดจำรูปแบบหรือทำนายอนาคตได้ เมื่อมีการกระตุ้น เซลล์ประสาทจะส่งกระแสผ่าน “เดนไดรต์” (input) และ “แอกซอน” (output) ซึ่งเชื่อมต่อกันเป็นเครือข่าย ประสาท โดยการนำหลักการนี้มาพัฒนาเป็นโครงข่ายประสาทเทียมในลักษณะคณิตศาสตร์ที่ใช้ค่าความแข็งแรงหรือค่า

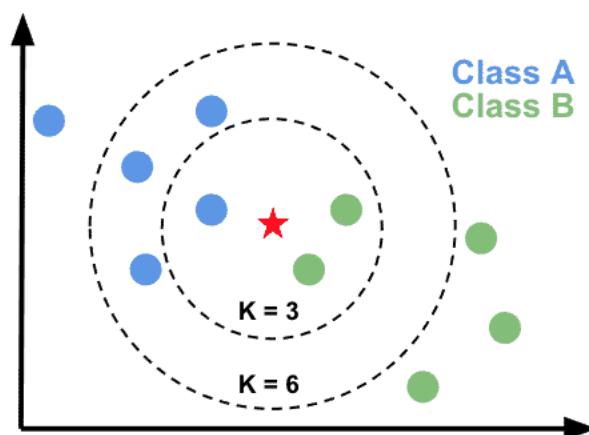
น้ำหนักในการคำนวณและส่งผลลัพธ์ผ่านฟังก์ชันถ่ายโอน (transfer function) ที่สร้างค่าเอาต์พุตจากค่าอินพุตที่ผ่านการปรับน้ำหนักแล้ว ดังแสดงดังภาพ



ภาพที่ 5 ตัวอย่างข่ายงานประสาทเทียมมีการเชื่อมต่อกันผ่านกลุ่มโหนด (Neural Network)
(clevertap,2025)

4.3. เทคนิคเพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbors : K-NN)

เทคนิคเพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbor: K-NN) เป็นวิธีการจำแนกข้อมูลที่ใช้การคำนวณระยะห่างระหว่างข้อมูลใหม่กับข้อมูลในชุดฝึกฝน โดยค้นหาข้อมูล K ตัวที่มีระยะทางใกล้เคียงที่สุดจากข้อมูลใหม่ แล้วเลือกคลาสที่มีจำนวนสมาชิกมากที่สุด ใน K ตัวอย่างนั้นเพื่อนำมาจำแนกข้อมูลใหม่ การคำนวณระยะทางมักใช้ Euclidean distance ซึ่งวัดความคล้ายคลึงกันของข้อมูล หากระยะทางน้อยแสดงว่าข้อมูลมีความคล้ายคลึงกันมาก ตัวอย่างเช่น ใน 2-NN จะค้นหา 2 ตัวอย่างที่ใกล้เคียงที่สุดเพื่อใช้ในการทำนายข้อมูลใหม่ ดังแสดงดังภาพ



ภาพที่ 6 ตัวอย่างการทำงานของเทคนิคเพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbor: K-NN)
(jcchouinard,2023)

4.4. เทคนิควิธีการรวมกลุ่ม (Vote Ensemble)

Vote Ensemble เป็นเทคนิคการเรียนรู้ของเครื่องที่รวมผลการทำนายจากโมเดลหลายตัวเข้าด้วยกัน โดยใช้หลักการโหวตเพื่อหาผลลัพธ์สุดท้าย มี 2 วิธีหลัก คือ Hard Voting ที่เลือกคลาสที่ได้คะแนนโหวตมากที่สุด และ Soft Voting ที่คำนวณค่าเฉลี่ยความน่าจะเป็นของแต่ละคลาส วิธีนี้ช่วยเพิ่มความแม่นยำในการทำนาย และลดความเสี่ยงจากข้อผิดพลาดของโมเดลเดี่ยว (Dietterich, 2000).

5. ขั้นตอนการแปรผลและประเมินผล (Evaluation)

ผู้วิจัยทำการ แบ่งข้อมูลออกเป็น 2 ส่วน เพื่อใช้สำหรับสร้างโมเดลและทำนายค่า โดยส่วนที่ 1 ใช้เพื่อสร้างโมเดล และส่วนที่ 2 ให้โมเดลทำนายค่าคลาสคำตอบออกมา ด้วยวิธีการแบ่ง 10-fold cross validation โดยแบ่งข้อมูลออกเป็น 10 กลุ่มเท่า ๆ กัน เพื่อใช้สำหรับเป็นข้อมูลในการสอนและทำการทดสอบประสิทธิภาพของตัวแบบด้วยค่า ดังสมการ 1-5 ดังนี้

- 5.1. รากที่สองของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) ถูกใช้เพื่อวัดความแม่นยำของแบบจำลองในการพยากรณ์ราคากองทุนรวมอิตาลีเอฟทองคำ โดยค่า RMSE ที่ต่ำหมายถึงแบบจำลองมีข้อผิดพลาดในการทำนายน้อย ซึ่งช่วยเพิ่มความน่าเชื่อถือของแบบจำลอง ดังสมการที่ 1

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

โดยที่ n = จำนวนตัวอย่างที่ได้ทั้งหมดที่ใช้ในการศึกษา
 y_i = ค่าจริง (Actual Data)
 \hat{y}_i = ค่าพยากรณ์ (Predictive Value)

- 5.2. ค่าสัมประสิทธิ์การตัดสินใจ (R^2) คือค่าที่ใช้เพื่อวัดความสามารถของแบบจำลองในการอธิบายความแปรปรวนของราคากองทุนรวมอิตาลีเอฟทองคำ โดยค่า R^2 ที่สูงบ่งชี้ว่าแบบจำลองสามารถอธิบายแนวโน้มของราคาทองคำได้อย่างแม่นยำดังสมการที่ 2

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2}$$

โดยที่ n = จำนวนตัวอย่างที่ได้ทั้งหมดที่ใช้ในการศึกษา
 y_i = ค่าจริง (Actual Data)
 \hat{y}_i = ค่าพยากรณ์ (Predictive Value)
 \bar{y}_i = ค่าเฉลี่ยผลลัพธ์ที่สังเกตได้

- 5.3. ค่าคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) ถูกใช้เป็นตัวชี้วัดเพื่อประเมินความแม่นยำของแบบจำลองพยากรณ์ราคากองทุนรวมอีทีเอฟทองคำ โดยค่า MSE ที่ต่ำบ่งชี้ว่าแบบจำลองมีข้อผิดพลาดในการทำนายต่ำ ซึ่งหมายถึงความสามารถของแบบจำลองในการคาดการณ์ราคาทองคำที่ใกล้เคียงกับค่าจริงดังสมการที่ 3

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

โดยที่ n = จำนวนตัวอย่างที่ได้ทั้งหมดที่ใช้ในการศึกษา

y_i = ค่าจริง (Actual Data)

\hat{y}_i = ค่าพยากรณ์ (Predictive Value)

- 5.4. ค่าความคลาดเคลื่อนสัมบูรณ์ (AE) ใช้เพื่อวัดความแตกต่างระหว่างค่าจริงของราคากองทุนรวมอีทีเอฟทองคำ ออกกับค่าที่แบบจำลองพยากรณ์ได้ โดยค่า AE ที่ต่ำหมายถึงแบบจำลองสามารถพยากรณ์ได้ใกล้เคียงกับค่าจริง ในขณะที่ค่า AE สูงบ่งชี้ว่าแบบจำลองมีข้อผิดพลาดมากขึ้นดังสมการที่ 4

$$AE = |y_i - \hat{y}_i|$$

โดยที่ y_i = ค่าจริง (Actual Data)

\hat{y}_i = ค่าพยากรณ์ (Predictive Value)

- 5.5. ค่าความคลาดเคลื่อนกำลังสอง (SE) ใช้เพื่อวัดขนาดของข้อผิดพลาดในการพยากรณ์ราคากองทุนรวมอีทีเอฟทองคำ โดยการยกกำลังสองของค่าความคลาดเคลื่อนทำให้ค่าที่เบี่ยงเบนมากมีอิทธิพลสูงขึ้น ซึ่งช่วยให้สามารถประเมินความแม่นยำของแบบจำลองได้อย่างละเอียดมากขึ้น

$$SE = (y_i - \hat{y}_i)^2$$

โดยที่ y_i = ค่าจริง (Actual Data)

\hat{y}_i = ค่าพยากรณ์ (Predictive Value)

6. การนำแบบจำลองไปใช้ (Deployment)

เมื่อทำการวิเคราะห์ตามมาตรฐานในการทำเหมืองข้อมูล (CRISP-DM) ผลการวิเคราะห์พบว่า เทคนิคที่มีความเหมาะสมมากที่สุดในการสร้างตัวแบบการพยากรณ์ราคากองทุนรวมอีทีเอฟทองคำ คือ เทคนิคการรวมกลุ่ม จากผลลัพธ์ของการสร้างแบบจำลองนี้สามารถนำไปใช้สำหรับการพยากรณ์ราคากองทุนรวมอีทีเอฟเพื่อช่วยให้นักลงทุนคาดการณ์แนวโน้มการเปลี่ยนแปลง

ผลการวิจัย (Results)

งานวิจัยนี้ได้นำข้อมูลปัจจัยที่ส่งผลทำให้เกิดการเปลี่ยนแปลงราคากองทุนรวมอียูเอฟทองคำ 13 ตัวแปร มาทำการสร้างตัวแบบสำหรับการพยากรณ์ราคากองทุนรวมอียูเอฟทองคำทำการศึกษาตามกระบวนการมาตรฐานในการทำเหมืองข้อมูล (CRISP-DM) เพื่อสร้าง ตัวแบบการพยากรณ์ราคากองทุนรวมอียูเอฟทองคำ โดยผู้วิจัยทำการแบ่งชุดข้อมูลด้วยวิธีการแบ่ง 10-fold Cross validation สำหรับการแบ่งข้อมูลออกเป็นข้อมูลสำหรับการสร้างแบบจำลองและข้อมูลสำหรับการทดสอบตัวแบบ โดยใช้เทคนิคการทำเหมืองข้อมูล 4 เทคนิค ประกอบด้วย เทคนิคสมการถดถอยเชิงเส้น เทคนิคโครงข่ายประสาทเทียม เทคนิคเพื่อนบ้านใกล้ที่สุด และเทคนิควิธีการรวมกลุ่ม จากนั้นทำการทดสอบประสิทธิภาพของการจำแนกประเภทข้อมูลด้วยค่า 5 ค่า ได้แก่ รากที่สองของค่าคลาดเคลื่อนกำลังสองเฉลี่ย ค่าวัดสัมประสิทธิ์การตัดสินใจ ค่าคลาดเคลื่อนกำลังสองเฉลี่ย ค่าความคลาดเคลื่อนสัมบูรณ์ และค่าความคลาดเคลื่อนกำลังสอง เพื่อหาตัวแบบที่มีความเหมาะสมสำหรับการพยากรณ์ราคากองทุนรวมอียูเอฟทองคำ ซึ่งในการวิเคราะห์ข้อมูลครั้งนี้ ผู้วิจัยได้ทำการเพิ่มประสิทธิภาพของตัวแบบการพยากรณ์ราคากองทุนรวมอียูเอฟทองคำด้วยวิธีการหาค่าที่เหมาะสมสำหรับพารามิเตอร์ในแต่ละเทคนิค ซึ่งผลลัพธ์ของการหาค่าที่เหมาะสมที่สุดในแต่ละเทคนิค ได้ค่าดังนี้ โดยเทคนิคโครงข่ายประสาทเทียม ค่า hidden layer ที่ทำให้ประสิทธิภาพสูงที่สุดเท่ากับ 6 เทคนิคเพื่อนบ้านใกล้ที่สุด ค่า k ที่ทำให้ประสิทธิภาพสูงที่สุดเท่ากับ 2 และเทคนิควิธีการรวมกลุ่มที่มีการรวมกลุ่มของ 3 เทคนิค ได้แก่ เทคนิคสมการถดถอยเชิงเส้น เทคนิคโครงข่ายประสาทเทียม เทคนิคเพื่อนบ้านใกล้ที่สุด โดยมีการหาค่าที่เหมาะสมที่สุดสำหรับพารามิเตอร์ของ 3 เทคนิค คือ เทคนิคเพื่อนบ้านใกล้ที่สุดค่า k ที่ทำให้ประสิทธิภาพสูงที่สุดเท่ากับ 2 ดังแสดงผลการวิเคราะห์ในตารางที่ 2

ตารางที่ 2 : การเปรียบเทียบค่าทดสอบประสิทธิภาพของตัวแบบสำหรับพยากรณ์ราคากองทุนรวมอียูเอฟทองคำ

Regression Techniques	Regression Performance				
	RMSE	R ²	MSE	AE	SE
Linear Regression	2.242	98.30%	5.03	1.775	5.04
Neural Net	2.02	99.10%	4.09	1.647	4.135
K-Nearest Neighbors	2.493	98.00%	6.22	1.6	6.259
Vote Ensemble*	1.676	99.10%	2.81	1.284	2.815

* คือ เทคนิคที่มีความเหมาะสมสำหรับนำมาสร้างตัวแบบการพยากรณ์ราคากองทุนรวมอียูเอฟทองคำ

จากตารางที่ 2 พบว่า เทคนิคการรวมกลุ่ม ให้ค่ารากที่สองของค่าคลาดเคลื่อนกำลังสองเฉลี่ย ต่ำที่สุดเท่ากับ 1.676 ค่าสัมประสิทธิ์การตัดสินใจ เท่ากับ 99.10% ค่าคลาดเคลื่อนกำลังสองเฉลี่ย เท่ากับ 2.81 ค่าความคลาดเคลื่อนสัมบูรณ์ เท่ากับ 1.284 ค่าความคลาดเคลื่อนกำลังสอง เท่ากับ 2.815 รองลงมาคือ เทคนิคโครงข่ายประสาทเทียม ให้ค่ารากที่สองของค่าคลาดเคลื่อนกำลังสองเฉลี่ย เท่ากับ 2.02 ค่าสัมประสิทธิ์การ

ตัดสินใจ เท่ากับ 99.10% ค่าคลาดเคลื่อนกำลังสองเฉลี่ย เท่ากับ 4.09 ค่าความคลาดเคลื่อนสัมบูรณ์ เท่ากับ 1.647 ค่าความคลาดเคลื่อนกำลังสอง เท่ากับ 4.135 เทคนิคสมการถดถอยเชิงเส้น ให้ค่ารากที่สองของค่าคลาดเคลื่อนกำลังสองเฉลี่ย เท่ากับ 2.242 ค่าสัมประสิทธิ์การตัดสินใจ เท่ากับ 98.30% ค่าคลาดเคลื่อนกำลังสองเฉลี่ย เท่ากับ 5.03 ค่าความคลาดเคลื่อนสัมบูรณ์ เท่ากับ 1.775 ค่าความคลาดเคลื่อนกำลังสอง เท่ากับ 5.04 และเทคนิคที่ให้ค่ารากที่สองของค่าความคลาดเคลื่อนเฉลี่ยสูงสุด คือ เทคนิคเพื่อนบ้านใกล้ที่สุด ให้ค่ารากที่สองของค่าคลาดเคลื่อนกำลังสองเฉลี่ย เท่ากับ 2.493 ค่าสัมประสิทธิ์การตัดสินใจ เท่ากับ 98.00% ค่าคลาดเคลื่อนกำลังสองเฉลี่ย เท่ากับ 6.22 ค่าความคลาดเคลื่อนสัมบูรณ์ เท่ากับ 1.6 ค่าความคลาดเคลื่อนกำลังสอง เท่ากับ 6.259 ตามลำดับ ดังนั้นเทคนิคที่มีความเหมาะสมที่สุดสำหรับการพยากรณ์ราคากองทุนรวมอีทีเอฟทองคำ คือ เทคนิคการรวมกลุ่ม

ผลสรุปและอภิปรายผลการวิจัย (Conclusion and Discussion)

งานวิจัยฉบับนี้มีวัตถุประสงค์เพื่อศึกษาประสิทธิภาพของเทคนิคเหมืองข้อมูล โดยการสร้างตัวแบบเพื่อพยากรณ์ราคากองทุนรวมอีทีเอฟทองคำ ซึ่งได้ข้อมูลจาก www.kaggle.com เก็บรวบรวมข้อมูลโดย Manu Siddhartha นำข้อมูลมาวิเคราะห์ตามมาตรฐานในการทำเหมืองข้อมูล (CRISP-DM) การวิเคราะห์ข้อมูลครั้งนี้ ผู้วิจัยทำการแบ่งชุดข้อมูลด้วยวิธีการ 10-fold Cross validation โดยแบ่งข้อมูลออกเป็น 10 กลุ่มเท่า ๆ กัน สำหรับข้อมูลสอนและข้อมูลทดสอบ และทำการวิเคราะห์ข้อมูลด้วยเทคนิคการทำเหมืองข้อมูลทั้งหมด 4 เทคนิคประกอบด้วย เทคนิคสมการถดถอยเชิงเส้น เทคนิคโครงข่ายประสาทเทียม เทคนิคเพื่อนบ้านใกล้ที่สุด และเทคนิควิธีการรวมกลุ่มมาใช้ในการสร้างตัวแบบการพยากรณ์ และทำการเพิ่มประสิทธิภาพของตัวแบบด้วยวิธีการหาค่าที่เหมาะสมที่สุด (Optimization) ด้วยวิธีด้านวิวัฒนาการ (Evolutionary Algorithms) เพื่อค้นหาค่าที่เหมาะสมที่สุดสำหรับชุดของพารามิเตอร์ในแต่ละตัวแบบและทำการเปรียบเทียบประสิทธิภาพการจำแนกประเภทข้อมูล โดยนำผลลัพธ์ของงานวิจัยนี้มาสร้างตัวแบบเพื่อพยากรณ์ราคากองทุนรวมอีทีเอฟทองคำ เพื่อเพิ่มความสามารถในการวิเคราะห์การพยากรณ์ให้มีความแม่นยำมากยิ่งขึ้นและเป็นประโยชน์ต่อนักลงทุน หรือ นักวิเคราะห์ทางการเงิน ที่ต้องการใช้ข้อมูลการพยากรณ์ราคาทองคำในการตัดสินใจลงทุน รวมถึงช่วยเสริมสร้างกลยุทธ์การบริหารความเสี่ยงในตลาดการเงินที่มีความผันผวนสูงที่จะเกิดขึ้นในอนาคตได้

จากผลการวิเคราะห์ข้อมูล พบว่า เทคนิคการรวมกลุ่ม เป็นเทคนิคที่เหมาะสมที่สุดในการนำมาสร้างตัวแบบการพยากรณ์ราคากองทุนรวมอีทีเอฟโดยมีค่ารากที่สองของค่าคลาดเคลื่อนกำลังสองเฉลี่ยเท่ากับ 1.676 ค่าสัมประสิทธิ์การตัดสินใจ เท่ากับ 99.10% ค่าคลาดเคลื่อนกำลังสองเฉลี่ย เท่ากับ 2.81 ค่าความคลาดเคลื่อนสัมบูรณ์ เท่ากับ 1.284 ค่าความคลาดเคลื่อนกำลังสอง เท่ากับ 2.815 เทคนิคการรวมกลุ่มสามารถจัดการข้อมูลที่มีขนาดใหญ่และซับซ้อนได้ดีกว่าเทคนิคอื่น ๆ โดยผลการศึกษาี้สอดคล้องกับงานวิจัยของ Alameer et al. (2019) ที่ได้นำเสนอวิธีการ ensemble ที่รวมหลายเทคนิคการเรียนรู้ของเครื่อง (Machine Learning) เข้าด้วยกันเพื่อพยากรณ์ราคาทองคำ โดยพบว่าวิธีการนี้มีประสิทธิภาพเหนือกว่าแบบจำลองทั่วไป

ข้อเสนอแนะงานวิจัย (Recommendation)

ผลการวิเคราะห์ในงานวิจัยนี้สามารถใช้ได้เฉพาะชุดข้อมูลที่ผู้วิจัยนำมาศึกษาเท่านั้น หากมีผู้ที่สนใจศึกษาการสร้างตัวแบบนี้สำหรับพยากรณ์ราคากองทุนรวมอีทีเอฟทองคำในอนาคต ควรมีการใช้เทคนิคเหมือนข้อมูลแบบอื่นที่นอกเหนือจากเทคนิคที่ผู้วิจัยได้ใช้ในงานครั้งนี้ เพื่อนำไปสร้างตัวแบบในการพยากรณ์ หรือเพิ่มตัวแปรอื่น ๆ ที่ส่งผลต่อราคากองทุนรวมอีทีเอฟทองคำที่ยังไม่ได้นำมาศึกษาในครั้งนี้ สำหรับการศึกษาต่อยอดเกี่ยวกับการพยากรณ์ราคากองทุนรวมอีทีเอฟทองคำสามารถขยายขอบเขตการวิจัยโดยการเพิ่มตัวแปรต้นที่เกี่ยวข้องอื่น ๆ เพื่อพิจารณาปัจจัยที่มีอิทธิพลต่อราคากองทุนรวมอีทีเอฟทองคำได้อย่างครอบคลุมมากขึ้น และสามารถช่วยในการพัฒนาแบบจำลองพยากรณ์ให้มีความแม่นยำและมีประสิทธิภาพสูงขึ้น

กิตติกรรมประกาศ

งานวิจัยนี้ได้รับการเอื้อเฟื้อข้อมูลที่เป็นประโยชน์จาก www.kaggle.com เก็บรวบรวมข้อมูลโดย Manu Siddhartha เพื่อนำมาใช้ในการวิเคราะห์ข้อมูล และผู้วิจัยขอขอบคุณผู้ช่วยศาสตราจารย์ ดร.อนุพงศ์ สุขประเสริฐ ที่ให้ความอนุเคราะห์ ชี้แนะแนวทางให้คำแนะนำปรึกษาตลอดจนปรับปรุงแก้ไขข้อผิดพลาดต่างๆ ด้วยความเอาใจใส่อย่างยิ่ง และคณะกรรมการบัญชีและการ จัดการมหาวิทยาลัยมหาสารคาม ที่ให้การสนับสนุนในการทำวิจัยครั้งนี้ ขอกราบขอบพระคุณเป็นอย่างสูงไว้ ณ ที่นี้

เอกสารอ้างอิง

1. Grammarly. (2024). *What is linear regression?* Grammarly.
<https://www.grammarly.com/blog/ai/what-is-linear-regression/>
2. Gunay, D. (2023, September 11). *Random forest*. Medium.
<https://medium.com/@denizgunay/random-forest-af5bde5d7e1e>
3. What is a decision tree? <https://www.ibm.com/think/topics/decision-trees>
4. Baur, D. G., & McDermott, T. K. (2010). Is gold a safe haven? International evidence. *Journal of Banking & Finance*, 34(8), 1886-1898. <https://doi.org/10.1016/j.jbankfin.2009.12.008>
5. World Gold Council. (2022). Gold demand trends.
<https://www.gold.org/goldhub/research/gold-demand-trends/gold-demand-trends-full-year-2022>
6. Shafiee, S., & Topal, E. (2010). An overview of global gold market and gold price forecasting. *Resources Policy*, 35(3), 178-189. <https://doi.org/10.1016/j.resourpol.2010.05.004>
7. Kumar, S., Pradhan, A. K., Tiwari, A. K., & Kang, S. H. (2018). Correlations and volatility spillovers between oil, natural gas, and stock prices in India. *Resources Policy*, 62, 282-291. <https://doi.org/10.1016/j.resourpol.2019.04.007>
8. Ghosh, D., Levin, E. J., Macmillan, P., & Wright, R. E. (2004). Gold as an inflation hedge? *Studies in Economics and Finance*, 22(1), 1-25. <https://doi.org/10.1108/eb043380>
9. Dietterich, T. G. (2000). Ensemble methods in machine learning. *Multiple Classifier Systems*, 1, 1-15. https://link.springer.com/chapter/10.1007/3-540-45014-9_1
10. Alameer, Z., Elaziz, M. A., Ewees, A. A., Ye, H., & Jianhua, Z. (2019). Forecasting gold price fluctuations using improved multilayer perceptron neural network and whale optimization algorithm. *Resources Policy*, 61, 250-260. <https://doi.org/10.1016/j.resourpol.2019.02.014>
11. อนุพงศ์ สุขประเสริฐ. (2564). การทำเหมืองข้อมูลด้วยโปรแกรม RapidMiner Studio (พิมพ์ครั้งที่ 5 ฉบับปรับปรุง). มหาสารคาม: สาขาคอมพิวเตอร์ธุรกิจ คณะการบัญชีและการจัดการ.
12. Siddhartha, M. (2021). Gold price prediction dataset [Dataset]. Kaggle. Retrieved March 26, 2025, from <https://www.kaggle.com/datasets/sid321axn/gold-price-prediction-dataset/data>