

การทำนายราคาทองคำโดยแบบจำลองการถดถอย  
Gold Price Prediction using a Regression Model

สุวรรณี ทองตา

600510589

การค้นคว้าอิสระนี้เป็นส่วนหนึ่งของการศึกษาหลักสูตรวิทยาศาสตรบัณฑิต  
สาขาวิชาวิทยาการคอมพิวเตอร์  
คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่  
ปีการศึกษา 2563

การทำนายราคาทองคำโดยใช้แบบจำลองการถดถอย  
Gold Price Prediction using a Regression Model

สุวรรณี ทองตา  
600510589

การค้นคว้าอิสระนี้ได้รับการพิจารณาอนุมัติให้เป็นส่วนหนึ่งของการศึกษา  
ตามหลักสูตรวิทยาศาสตรบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์  
คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่  
ปีการศึกษา 2563

คณะกรรมการสอบการค้นคว้าอิสระ

..... ประธานกรรมการ  
(ผู้ช่วยศาสตราจารย์ ดร.ชุมพล บุญคุ้มพรภัทร)  
..... กรรมการ  
(ผู้ช่วยศาสตราจารย์ ดร.วัฒนา จินดาหลวง)  
วันที่.....เดือน.....พ.ศ.....

## กิตติกรรมประกาศ

การค้นคว้าอิสระเล่มนี้สำเร็จลุล่วงได้โดยได้รับความอนุเคราะห์จากบุคคลหลายท่าน ขอกราบขอบพระคุณ ผู้ช่วยศาสตราจารย์ ดร.ชุมพล บุญคุ้มพรภัทร ซึ่งได้กรุณาให้คำปรึกษาแนะนำแนวคิดวิธีการและเสียสละเวลาอันมีค่าแก้ไขข้อบกพร่องของเนื้อหาและสำนวนภาษาด้วยความใส่ใจยิ่ง ผู้ค้นคว้าอิสระขอกราบขอบพระคุณเป็นอย่างสูง ณ โอกาสนี้

ขอกราบขอบพระคุณ ผู้ช่วยศาสตราจารย์ ดร.วัฒนา จินดาหลวง ที่กรุณารับเป็นกรรมการสอบการค้นคว้าอิสระนี้รวมทั้งให้คำแนะนำเป็นอย่างดีมาโดยตลอด ขอขอบคุณคณาจารย์ที่ได้ให้การสนับสนุนการดำเนินการทำงานและมอบความรู้วิชาอันมีค่า เพื่อเป็นพื้นฐานในการทำการค้นคว้าอิสระในครั้งนี้ และขอบคุณทุกความช่วยเหลือในการทำการค้นคว้าอิสระนี้ให้สำเร็จลุล่วงไปด้วยดี

นางสาวสุวรรณี ทองตา

600510589

หัวข้อการค้นคว้าอิสระ	การทำนายราคาทองคำโดยใช้แบบจำลองการถดถอย	
ชื่อเจ้าของโครงการ	นางสาว สุวรรณี ทองตา	รหัสประจำตัว 600510589
วิทยาศาสตร์บัณฑิต	สาขาวิทยาการคอมพิวเตอร์	
อาจารย์ที่ปรึกษา	ผู้ช่วยศาสตราจารย์ ดร. ชุมพล บุญคุ้มพรภัทร	

## บทคัดย่อ

การทำนายราคาทองคำเป็นเทคนิคที่ใช้ประมาณราคาทองคำในอนาคต แบบจำลองการคาดการณ์ขึ้นอยู่กับวิธีการวิเคราะห์ราคาทองคำในอดีต นักวิจัยได้สร้างแบบจำลองประเภทต่าง ๆ เช่น Extra tree regressor, Catboost regressor และ K-Neighbors regressor สำหรับการทำนายการถดถอยพหุคูณ ในการทดลองโมเดลทั้งหมดจะถูกเปรียบเทียบการแสดงผลเพื่อให้ได้ข้อสรุป

ในงานวิจัยนี้ ผู้วิจัยได้ทำการสร้างและเปรียบเทียบโมเดลที่ 1, 2 และ 3 ได้แก่ โมเดลการทำนายทองคำล่วงหน้า 1 วัน, โมเดลการทำนายทองคำล่วงหน้า 14 วัน และ โมเดลการทำนายทองคำล่วงหน้า 22 วัน โดยโมเดลทั้ง 3 โมเดล มีแนวโน้มการทำนายคล้ายกัน แต่โมเดลที่ 1 ค่าความถูกต้องในการทำนายมีค่าดีกว่าค่าความถูกต้องการทำนายของโมเดลที่ 2 และ โมเดลที่ 3 ทำให้โมเดลที่ 1 ค่าที่ทำนายมีค่าใกล้เคียงกับค่าจริงมากที่สุด โดยมีค่าความถูกต้อง (Accuracy) ที่ได้จาก Stack Model ที่  $MAE = 0.0067$ ,  $MSE = 0.0001$ ,  $RMSE = 0.0097$ ,  $R2 = -0.0355$ ,  $RMSLE = 0.0088$  และ  $MAPE = -0.1903$

<b>Independent Study Title</b>	Gold Price Prediction using a Regression Model
<b>Author</b>	Ms. Suwannee Thongtha <b>Student ID</b> 600510589
<b>Bachelor of Science</b>	Computer Science
<b>Supervisor</b>	Assistant Professor Dr. Chumphol Bunkhumpornpat

## Abstract

Gold price prediction is a technique that estimates the future price of gold. A predictive model is created based on the analysis of historical gold price. The researchers use various types of models such as Extra trees regressor, Catboost regressor, and K-Neighbors regressor for the Multiple Regression Prediction. In the experiment, all models are compared their performances to get the conclusion.

This paper, researcher created and compared models 1, 2 and 3: The 1-day gold prediction model, the 14-day gold prediction model, and the 22-day gold prediction model. Model 1 prediction accuracy was better than the prediction accuracy of Model 2 and Model 3, making Model 1 the predicted value closest to the true value. The accuracy obtained from the stack model at  $MAE = 0.0067$ ,  $MSE = 0.0001$ ,  $RMSE = 0.0097$ ,  $R^2 = -0.0355$ ,  $RMSLE = 0.0088$  and  $MAPE = -0.1903$ .

## สารบัญ

	หน้า
กิตติกรรมประกาศ	ก
บทคัดย่อ	ข
Abstract	ค
สารบัญ	ง-จ
สารบัญตาราง	ฉ
สารบัญภาพ	ช
บทที่ 1 บทนำ	1
1.1 หลักการและเหตุผล	1
1.2 วัตถุประสงค์ของโครงการ	2
1.3 ประโยชน์ที่จะได้รับจากการศึกษาเชิงประยุกต์	2
1.4 ขอบเขตของโครงการ	2
1.5 แผนการดำเนินงานและระยะเวลาการดำเนินงาน	5
บทที่ 2 ทฤษฎี เอกสาร และงานวิจัยที่เกี่ยวข้อง	6
2.1 ท้องคำ	6
2.2 การถดถอยพหุคูณ	12
2.3 งานวิจัยที่เกี่ยวข้อง	15
บทที่ 3 วิธีการดำเนินงานวิจัย	19
3.1 กำหนดแนวทางที่ใช้ในการทำนายราคาทองคำ	19
3.2 การรวบรวมและเตรียมข้อมูล	19
3.3 การสำรวจลักษณะของข้อมูล	20
3.4 การทำความสะอาดข้อมูล	21
3.5 การฝึกฝนแบบจำลอง	21

3.6	การเตรียมแบบจำลอง	22
3.7	การเรียกใช้แบบจำลอง	22
3.8	การทำนายราคาทองคำล่วงหน้า	22
บทที่ 4	ผลการดำเนินงานและผลการวิเคราะห์ข้อมูล	23
4.1	การหาแบบจำลองที่เหมาะสมกับการทำนายผล	23
4.2	Compare Models	24
4.3	Create Model	24
4.4	Tune Model	27
4.5	Ensemble Model	30
4.6	Blend Model	31
4.6	Stack Model	32
บทที่ 5	สรุปผลงานวิจัย	35
5.1	ผลการทำนายราคาทองคำ	35
5.2	กราฟเปรียบเทียบผลการทำนายราคาทองคำกับราคาทองคำจริง	36
	เอกสารอ้างอิง	38

## สารบัญตาราง

ตารางที่		หน้า
1.1	ตัวอย่างข้อมูลนำเข้า	4
1.2	ขั้นตอนการดำเนินงานและระยะเวลาดำเนินงาน	5
2.1	หน่วยวัดน้ำหนัก	10
2.2	การเปรียบเทียบหน่วยวัดทองคำ	10-11
2.3	ตัวอย่างข้อมูลอนุกรมเวลาของราคาทองคำ	15
3.1	รายละเอียดข้อมูลที่ใช้อธิบายสัญลักษณ์	20



## สารบัญภาพ

ภาพที่		หน้า
2.1	ขั้นตอนการเรียนรู้เพื่อสร้างโมเดล	12
2.2	ขั้นตอนการทำนายผลจากโมเดล	13
2.3	การเปรียบเทียบระหว่างการถดถอยเชิงเส้นอย่างง่ายและการถดถอยพหุคูณ	14
4.1	การเปรียบเทียบแบบจำลอง	23
4.2	ค่าMetricsที่แสดงจากการสร้างแบบจำลอง TheilSen Regressor	24
4.3	ค่าMetricsที่แสดงจากการสร้างแบบจำลอง K Neighbors Regressor	25
4.4	ค่าMetricsที่แสดงจากการสร้างแบบจำลอง Extra Trees Regressor	25
4.5	ค่าMetricsที่แสดงจากการสร้างแบบจำลอง CatBoost Regressor	26
4.6	ค่าMetricsที่แสดงจากการสร้างแบบจำลอง Automatic Relevance Determination	26
4.7	ค่าMetricsหลังทำการ tune model ของแบบจำลอง K Neighbors Regressor โดยกำหนดจำนวนครั้งในการสุ่มค่าไฮเปอร์พารามิเตอร์เท่ากับ 150 ครั้ง	27
4.8	การแสดงผลของแบบจำลอง K Neighbors Regressor ก่อนทำการปรับค่าไฮเปอร์พารามิเตอร์	28
4.9	การแสดงผลของแบบจำลอง K Neighbors Regressor หลังทำการปรับค่าไฮเปอร์พารามิเตอร์	28
4.10	การรวมแบบจำลอง K Neighbors Regressor tuned	29
4.11	การแสดงผลการรวมแบบจำลอง K Neighbors Regressor tuned	30
4.12	การ blending แบบจำลอง knn_tuned และ แบบจำลอง et	30
4.13	การซ้อนแบบจำลองที่1	31
4.14	การซ้อนแบบจำลองที่2	31
4.15	การซ้อนแบบจำลองที่3	32
5.1	ผลการทำนายราคาทองคำในอีก 1 วัน	33
5.2	ผลการทำนายราคาทองคำในอีก 14 วัน	33
5.3	ผลการทำนายราคาทองคำในอีก 22 วัน	34
5.4	กราฟเปรียบเทียบราคาทองคำจริงกับราคาทองคำที่ทำนาย อีก1วัน	34
5.5	กราฟเปรียบเทียบราคาทองคำจริงกับราคาทองคำที่ทำนาย อีก14วัน	35
5.6	กราฟเปรียบเทียบราคาทองคำจริงกับราคาทองคำที่ทำนาย อีก22วัน	35

# บทที่ 1

## บทนำ

ในปัจจุบันนักลงทุนสามารถทำการซื้อขายทองคำได้เองผ่านทางออนไลน์ ซึ่งจะเครื่องมือ (Indicator) ที่ช่วยในการวิเคราะห์สภาพตลาด ทำให้นักลงทุนสามารถประเมินแนวโน้มตลาดสถานะต่าง ๆ ได้ง่าย และยังมีการใช้กราฟการเปลี่ยนแปลงของราคาทองคำย้อนหลังมาประกอบการตัดสินใจการลงทุนด้วย แต่ก็ยังมีความเสี่ยงในการลงทุนอยู่ ซึ่งในงานวิจัยฉบับนี้จะกล่าวถึง การทำนายราคาทองคำจากการเปลี่ยนแปลงของราคาทองคำย้อนหลังกับสินทรัพย์บางประเภทมีความสัมพันธ์กัน โดยใช้การเรียนรู้ของเครื่องจักร (Machine Learning) ในการทำนายราคาทองคำ

### 1.1 หลักการและเหตุผล

ทองคำเป็นสื่อกลางในการแลกเปลี่ยนของมนุษย์ เป็นมาตรฐานในการวัดมูลค่าของสินค้าและบริการ แม้แต่นำมาใช้ทำเป็นเครื่องประดับ สิ่งของเครื่องใช้ มานานหลายศตวรรษจนกระทั่งพัฒนามาเป็นสกุลเงินกระดาษที่ใช้กันในปัจจุบัน ทำให้ทองคำเป็นที่หมายปองของมนุษย์ โดยนำมาตีมูลค่าสำหรับการแลกเปลี่ยนระหว่างประเทศและใช้เป็นวัตถุดิบสำคัญในด้านวงการต่าง ๆ เช่น วงการอุตสาหกรรมเครื่องประดับอัญมณี ด้านความมั่นคงทางเศรษฐกิจการคลัง ด้านการคมนาคมและการสื่อสารโทรคมนาคม ด้านการแพทย์ ด้วยประโยชน์และมูลค่าทำให้ทองคำเป็นสินทรัพย์ที่มีการซื้อขายกันอย่างแพร่หลายในตลาดที่มีการแข่งขันสูง จึงมีหลายคนมองเห็นช่องทางในการลงทุนไม่ว่าจะเป็นการซื้อทองคำเพื่อเก็งกำไรและขายทองคำในเวลาที่เราค่าเพิ่มขึ้น ทำให้ในปัจจุบันการเคลื่อนไหวของราคาทองคำได้รับความสนใจมากขึ้นเพื่อที่จะเป็นเกณฑ์ในการประกอบการตัดสินใจลงทุน ดังนั้นผู้ศึกษาจึงได้ใช้เทคนิคทางสาขาวิทยาการคอมพิวเตอร์เข้ามาช่วยทำนาย โดยได้นำตัวแบบ (Model) ในการรู้จำแบบ (Pattern recognition) มาใช้เพื่อทำให้ตัวแบบทำนายได้แม่นยำที่สุด เพื่อใช้ทำนายราคาทองคำที่จะเกิดขึ้นในอนาคต เพื่อช่วยลดความเสี่ยงของการลงทุน และเป็นประโยชน์แก่นักลงทุนคนอื่น ๆ ซึ่งสุดท้ายแล้วคิดว่าจะสามารถนำตัวทำนายนี้ไปประยุกต์ใช้กับข้อมูลอื่น ๆ ได้ในอนาคต

ด้วยเหตุนี้จึงสนใจในการสร้างโมเดลการทำนายราคาทองคำ โดยทำการค้นหา ข้อมูลชุด (Dataset) รวบรวม และเรียงข้อมูลที่จะนำมาทำนาย ซึ่งข้อมูลดังกล่าวนี้ได้มาการใช้โมดูล yahoo-finance[1] ในภาษาไพทอน (Python) ดึงมาจากเว็บไซต์ Yahoo Finance เป็นบริการของ Yahoo

ซึ่งเป็นเว็บไซต์ที่ให้บริการข้อมูลตลาดหุ้นให้แก่ผู้ใช้บริการของ Yahoo สำหรับตัวแบบ (Model) เรา จะทดลองกับอัลกอริทึมที่แตกต่างกันโดยใช้ไลบรารี PyCaret ของภาษาไพทอน เพื่อให้ได้ตัวแบบ (Model) ที่มีประสิทธิภาพ และสร้าง Pipeline เพื่อนำเข้าข้อมูลใหม่อย่างต่อเนื่อง

## 1.2 วัตถุประสงค์ของโครงการ

เพื่อสร้างโมเดลสำหรับทำนายราคาทองคำล่วงหน้า

## 1.3 ประโยชน์ที่จะได้รับจากการศึกษาเชิงประยุกต์

- 1) สามารถนำโมเดลที่ได้ไปเป็นเกณฑ์การตัดสินใจในการลงทุนของนักลงทุน
- 2) เป็นการดึงข้อมูลที่มีอยู่แล้วมาประยุกต์ใช้ให้เกิดประโยชน์

## 1.4 ขอบเขตของโครงการ

ขอบเขตของโครงการประกอบด้วย ขอบเขตทางสถาปัตยกรรม ขอบเขตของระบบงาน และ ขอบเขตของข้อมูล มีรายละเอียดดังนี้

### 1) ขอบเขตทางสถาปัตยกรรม

ระบบที่ได้ทำการพัฒนาขึ้นมาเป็นแบบเดี่ยว (Standalone)

#### 1.1) ฮาร์ดแวร์ (Hardware) ที่ใช้ในการพัฒนาระบบ ประกอบด้วย

- คอมพิวเตอร์มีหน่วยประมวลผล (CPU) Intel Core i5-7200
- หน่วยความจำเข้าถึงแบบสุ่ม (Random Access Memory) 4 Gigabyte
- จานบันทึกแบบแข็ง (Hard Disk) ขนาดความจุ 200 Gigabyte

#### 1.2) ซอฟต์แวร์ (Software) ที่ใช้ในการพัฒนาแบบจำลอง ประกอบด้วย

- ระบบปฏิบัติการไมโครซอฟท์วินโดวส์ 10 โพร (Microsoft Windows 10 Pro)
- Python 3.7.0 (ภาษาสำหรับพัฒนาโปรแกรมที่ใช้ในงานวิจัย)
- Pandas 1.1.0 (ไลบรารีสำหรับการจัดการข้อมูล)
- DateTime 4.3 (ไลบรารีสำหรับการเข้าถึงวันและเวลา)
- Matplotlib 3.3.1 (ไลบรารีสำหรับการแสดงผลในรูปแบบของกราฟ)
- Numpy 1.19.1 (ไลบรารีสำหรับการคำนวณทางคณิตศาสตร์)

- Jupyterlab 6.1.3 (เป็นโอเพนซอร์สเว็บแอปพลิเคชัน สำหรับการเขียนภาษาไพทอน)
- Pycaret 2.0 (ไลบรารีสำหรับการเรียนรู้ของเครื่องโอเพนซอร์ส)
- Xlrd 1.2.0 (ไลบรารีสำหรับเปิดอ่านไฟล์เอกซ์เซล)
- Scipy 1.5.2 (ไลบรารีสำหรับการคำนวณทางคณิตศาสตร์)
- Seaborn 0.10.1 (ไลบรารีสำหรับการแสดงผลในรูปแบบของกราฟ)
- Prophet 0.1.1 (ไลบรารีสำหรับสร้างแบบจำลองการทำนาย)
- YahooFinancials 1.4.0 (เอพีไอสำหรับดึงข้อมูล) [1]

## 2) ขอบเขตของระบบงาน

ลักษณะการทำงานหลักแบ่งออกเป็นดังนี้

### 2.1) การดึงข้อมูลจากแหล่งออนไลน์

ในการทำนายราคาทองคำนั้น มีปัจจัยและความสัมพันธ์ที่เกี่ยวเนื่องอยู่มากมาย เช่น ความสัมพันธ์ของทองคำและหุ้นที่ไปในทิศทางเดียวกัน ราคาทองต่างประเทศ อัตราเงินเฟ้อ หรือแม้แต่ค่าเงินบาทต่อดอลลาร์สหรัฐ ก็เป็นปัจจัยที่เกี่ยวข้องกับราคาทองคำ จากปัจจัยและความสัมพันธ์ที่กล่าวมาข้างต้นเป็นเพียงปัจจัยบางส่วนที่ส่งผลต่อการเปลี่ยนแปลงของราคาทองคำ เนื่องจากข้อมูลดังกล่าวในปัจจุบันสามารถดึงข้อมูลได้ผ่าน API เว็บไซต์การเงินออนไลน์ โดยผู้วิจัยจะต้องทำการสืบหาข้อมูลเกี่ยวกับปัจจัยต่าง ๆ ที่เกี่ยวข้องและหาแหล่งข้อมูลเหล่านั้นใช้ในการคำนวณทำนายราคาทองคำ

### 2.2) การเตรียมข้อมูลก่อนนำข้อมูลเข้าสู่แบบจำลอง

เนื่องจากการเปลี่ยนแปลงของราคาทองคำนั้น มีหลายส่วยและสามารถสังเกตได้จากสินทรัพย์บางรายการ ทำให้ข้อมูลที่จะนำเข้านั้นมีหลายรายการ ซึ่งข้อมูลแต่ละรายการนั้นจะต้องมีการปรับแต่งให้เป็นมาตรฐานเดียวกันและแก้ไขค่าที่ไม่เหมาะสม เพื่อให้ข้อมูลที่สมบูรณ์พร้อมที่จะนำสร้างแบบจำลอง

### 2.3) การฝึกแบบจำลอง

ในการฝึกฝนแบบจำลองนั้นผู้วิจัยจะใช้ แบบจำลองการวิเคราะห์การถดถอยพหุคูณ (Multiple Regression Analysis Model) โดยจะต้องทำการแบ่งข้อมูลออกเป็นสองส่วนในการใช้สำหรับฝึกฝนและทดสอบแบบจำลอง เมื่อทำการฝึกฝนเสร็จแล้ว แบบจำลองจะให้ผลลัพธ์ที่ดีที่สุดออกมา

### 3) ขอบเขตของข้อมูล

ข้อมูลที่ได้มาจากการดึงข้อมูล API ของ yahoofinancials [1] ที่ได้จากเว็บไซต์ <https://finance.yahoo.com> ซึ่งเป็นเว็บไซต์ที่ให้บริการข้อมูลตลาดหุ้นให้แก่ผู้ใช้บริการของ Yahoo ทำให้ yahoofinancials จะใช้สัญลักษณ์ของ Yahoo ดังนั้นจึงต้องมีการนำเข้าไฟล์ที่ใช้ในการอธิบายสัญลักษณ์

#### 3.1) ข้อมูลนำเข้า

ในการทำนาย ผู้วิจัยได้เลือกนำเข้าราคาปิดของรายการสินทรัพย์บางรายการ ได้แก่ Gold, Silver, Crude Oil, S&P500, Russel 2000 Index, 10 Yr US T-Note futures, 10 Yr US T-Note futures, Platinum, Copper, Dollar Index, Volatility Index, Soybean, MSCI EM ETF, Euro USD, Euronext100, Nasdaq. มีทั้งหมด 16 แอททริบิวต์ (ไม่รวมวันที่) และแต่ละแอททริบิวต์ (Attribute) มี 2651 ทูเปิล (Tuple) โดยเริ่มเลือกช่วงเวลานำเข้าข้อมูลตั้งแต่วันที่ 1 มกราคม ปี ค.ศ.2010 เป็นต้นไป เนื่องจากสามารถเลือกช่วงเวลาของการนำเข้าข้อมูลได้ ความแม่นยำของโมเดลจะขึ้นอยู่กับจำนวนข้อมูลที่เลือกมา

ตารางที่ 1.1 ตัวอย่างข้อมูลนำเข้า

Date	Gold	Silver	Crude Oil	S&P500	...	Euro USD	Euronext100	Nasdaq
2010-01-01	1117.7	17.4	81.5	1133.0	...	1.4	697.0	2308.4
2010-01-04	1117.7	17.4	81.5	1133.0	...	1.4	697.0	2308.4
2010-01-05	1118.1	17.8	81.8	1136.5	...	1.4	697.6	2308.7

2010-01-06	1135.9	18.2	83.2	1137.1	...	1.4	698.3	2301.1
2010-01-07	1133.1	18.3	82.7	1141.7	...	1.4	697.8	2300.1

### 3.2) ข้อมูลส่งออก

ข้อมูลส่งออกจะเป็นค่าตัวเลขที่ทำนายราคาทองคำในอนาคต แสดงในรูปแบบตารางและกราฟเปรียบเทียบผล

## 1.5 แผนการดำเนินงานและระยะเวลาดำเนินงาน

การศึกษานี้เริ่มดำเนินงานตั้งแต่ เดือนกรกฎาคม พ.ศ. 2563 ถึง เดือนมีนาคม พ.ศ. 2564 แสดงรายละเอียดการดำเนินงาน ดังตารางที่ 1.2

ตารางที่ 1.2 ขั้นตอนการดำเนินงานและระยะเวลาดำเนินงาน

ขั้นตอนการดำเนินงาน	ระยะเวลาการดำเนินงาน								
	พ.ศ.2563						พ.ศ.2564		
	ก.ค.	ส.ค.	ก.ย.	ต.ค.	พ.ย.	ธ.ค.	ม.ค.	ก.พ.	มี.ค.
ศึกษาทฤษฎีหลักการที่เกี่ยวข้อง									
รวบรวมข้อมูลที่จะทำนาย									
เลือกตัวแบบที่จะใช้ในการทำนาย									
เตรียมข้อมูลที่จะทำนาย									
ออกแบบและพัฒนาโมเดล									
ทดสอบตัวแบบที่สร้างไว้และดูความแม่นยำในการทำนาย									
จัดทำเอกสารประกอบระบบ									

## บทที่ 2

### ทฤษฎี เอกสาร และงานวิจัยที่เกี่ยวข้อง

ในบทนี้จะกล่าวถึงหลักการและทฤษฎีที่เกี่ยวข้องกับการวิเคราะห์ ออกแบบ และพัฒนาแบบจำลองสำหรับการทำนายราคาทองคำ ในการทำนายด้วยการเรียนรู้ของเครื่อง ซึ่งได้ทำการศึกษาหลักการทฤษฎีที่เกี่ยวข้องและนิยามศัพท์ รวมไปถึงเอกสารและงานวิจัยครั้งนี้

#### 2.1 ทองคำ [2]

##### 2.1.1 ความรู้ทั่วไปเกี่ยวกับทองคำ

ทองคำ คือธาตุเคมีที่มีหมายเลขอะตอม 79 และสัญลักษณ์คือ Au เรียกโดยย่อว่า “ทอง” เป็นธาตุโลหะทรานซิชันสีเหลือง เกิดเป็นธาตุอิสระในธรรมชาติ มีจุดหลอมเหลวที่ 1,064 องศาเซลเซียสและจุดเดือดอยู่ที่ 2,700 องศาเซลเซียส มีความถ่วงจำเพาะ 19.33 กรัมต่อซีซี น้ำหนักอะตอม 196.67 หน่วยมวลอะตอม ลักษณะที่พบเป็นเกล็ด เม็ดกลม แบน หรือรูปร่างคล้ายกิ่งไม้ รูปผลึกแบบลูกเต๋า (Cube) หรือ ออกตะฮีดรอน (Octahedron) หรือ โดเดคะฮีดรอน (Dodecahedron)

##### 2.1.2 คุณสมบัติของทองคำ

ทองคำถูกตีมูลค่าสำหรับการแลกเปลี่ยนระหว่างประเทศ และใช้เป็นวัตถุดิบที่สำคัญสำหรับวงการเครื่องประดับ จึงทำให้ทองคำได้รับความนิยมอย่างสูงสุดในวงการเครื่องประดับ เพราะเป็นโลหะมีค่าชนิดเดียวที่มีคุณสมบัติพื้นฐาน 4 ประการซึ่งทำให้ทองคำมีลักษณะเด่นกว่าโลหะมีค่าทุกชนิดในโลก คือ

- 1) ความมันวาว (Lustre) สีที่สวามิฉันตามธรรมชาติผสมกับความมันวาวก่อให้เกิดความงามอันเป็นอมตะ ทองคำสามารถเปลี่ยนเฉดสีทองโดยการนำทองคำไปผสมกับโลหะมีค่าอื่น ๆ ช่วยเพิ่มความงดงามให้แก่ทองคำได้อีกทางหนึ่ง
- 2) ความคงทน (Durable) เนื่องจากทองคำไม่ทำปฏิกิริยากับออกซิเจน ดังนั้น เมื่อถูกสัมผัสกับอากาศ สีของทองคำจะไม่หมองและไม่ขึ้นสนิมทองคำไม่ขึ้นสนิม ไม่ผุกร่อน แม้ว่ากาลเวลาจะผ่านไปนานเท่าไรก็ตาม และเป็นโลหะที่มีความเหนียว จะยืดขยายเมื่อถูกตีหรือรีดในทุกทิศทาง
- 3) ความหายาก (Rarity) ทองคำเป็นแร่ที่หายาก กว่าจะได้ทองคำมาหนึ่งออนซ์ ต้องถลุงก้อนแร่ที่มีทองคำอยู่เป็นจำนวนหลายตัน และต้องขุดเหมืองลึก

ลงไปหลายสิบเมตรจึงทำให้มีค่าใช้จ่ายที่สูง เป็นสาเหตุให้ทองคำมีราคาแพงตามต้นทุนในการผลิต

- 4) การนำกลับไปใช้ประโยชน์ (Reuseable) เนื่องจากมีคุณสมบัติความคงทน ดังนั้นจึงทองคำเหมาะสมที่สุดต่อการนำมาทำเป็นเครื่องประดับ เพราะมีความเหนียวและอ่อนนิ่ม สามารถนำมาทำขึ้นรูปได้ง่าย อีกทั้งยังสามารถนำกลับมาใช้ใหม่ได้โดยการทำให้บริสุทธิ์ (Purified) ด้วยการหลอมได้อีกนับครั้งไม่ถ้วน

### 2.1.3 คุณสมบัติประโยชน์

นอกจากจะใช้ทองคำเป็นทุนสำรองทางการเงินของหลายประเทศแล้ว ทองคำยังมีประโยชน์ในด้านอื่น ๆ อีก เช่น

- 1) ด้านอุตสาหกรรมเครื่องประดับ ทองคำเป็นโลหะที่ได้รับความนิยมมากสุดในการนำไปใช้ทำเป็นเครื่องประดับ จากอดีตถึงปัจจุบันเครื่องประดับอัญมณีทองคำได้มีส่วนทำเป็นฐานเรือน รองรับอัญมณีมาโดยตลอด จากรูปแบบชั้นพื้นฐานของงานทองที่ง่ายที่สุด ไปสู่เทคนิคการทำทองด้วยเทคโนโลยีขั้นสูง
- 2) ด้านอวกาศ ในทางอวกาศได้มีการนำทองคำมาใช้เป็นชุดนักบินอวกาศและแคปซูลเพื่อป้องกันไม่ให้นักบินอวกาศกระทบกับรังสีในอวกาศที่มีพลังงานสูง นอกจากนี้ยังมีการใช้ทองคำบริสุทธิ์เคลือบกับเครื่องยนต์ ระบบอิเล็กทรอนิกส์ หมวกเหล็ก เกราะบังหน้า และอุปกรณ์อื่น ๆ ที่ใช้ในอวกาศ เนื่องจากทองคำที่มีความหนา 0.000006 นิ้ว จะมีคุณสมบัติช่วยสะท้อนรังสีความร้อนจากดวงอาทิตย์ไม่ให้ทำลาย หรือลดประสิทธิภาพการทำงานของอุปกรณ์เหล่านี้
- 3) ด้านทันตกรรม มีการใช้ทองคำเพื่อการครอบฟัน การอุดฟัน การเชื่อมฟัน การจัดฟัน การดัดฟัน หรือการเลี่ยมทอง และยังมีการใช้ในการผลิตฟันปลอมด้วย เนื่องจากทองคำมีความคงทนต่อการกัดกร่อน การหมองคล้ำ และยังมีความแข็งแรงอีกด้วย โดยจะใช้ทองคำผสมกับธาตุอื่น เช่น แพลทินัม (Platinum)
- 4) ด้านการแพทย์ ในสมัยโบราณได้มีความเชื่อเกี่ยวทองคำที่ว่าทองคำนั้นมีศักยภาพทำให้คนที่มีความสุขร่างกายที่แยกละกลับมาดีขึ้นได้ และด้วยความเชื่อนี้เองทางการแพทย์จึงได้นำทองคำมาทดลองโดยการนำเอาแร่ทองคำมาทดสอบ ซึ่งทองคำที่นำมาทดสอบ



นั้นจะอยู่ในรูปของเกลือ (Gold salts) แล้วก็พบว่าทองคำนั้นมีฤทธิ์ด้านอาการอักเสบและบวมซ้ำของโรคเก๊าท์ได้ นอกจากนั้นจากการทดลองนานกว่า 80 ปีของทางการแพทย์ก็พบอีกว่าแร่ทองคำนั้นสามารถต้านอนุมูลอิสระที่เกิดขึ้นจากข้อกระดูกอักเสบ และทำช่วยบรรเทาความเจ็บปวดหรืออาการบวมซ้ำได้จริง

- 5) ด้านอิเล็กทรอนิกส์ มีการนำทองคำมาใช้เป็นวัสดุที่ทำหน้าที่สัมผัสในอุปกรณ์อิเล็กทรอนิกส์ต่าง ๆ เช่น เครื่องคิดเลข โทรศัพท์ คอมพิวเตอร์ หรือโทรศัพท์มือถือ เนื่องจากทองคำมีค่าการนำไฟฟ้าสูง และมีความคงทนต่อการกัดกร่อน จึงช่วยเพิ่มประสิทธิภาพ และอายุการใช้งานของเครื่องไฟฟ้าเหล่านั้น
- 6) ด้านความมั่นคงทางเศรษฐกิจ ทองคำมีประโยชน์ในฐานะเป็นโลหะสีกลางแห่งการแลกเปลี่ยนเงินตรา ทองคำถูกสำรองไว้เป็นทุนสำรองเงินตราระหว่างประเทศ เพราะทองคำมีมูลค่าในตัวเอง ผิดกับเงินตราสกุลต่าง ๆ อาจเพิ่มหรือลดได้ ทองคำถูกใช้เป็นเครื่องมือในการเก็งกำไรของตลาดการค้า นอกจากนี้ยังได้มีการจัดทำเป็นเหรียญกษาปณ์ทองคำ หรือแสตมป์ทองคำ หรือธนบัตรทองคำ ซึ่งถูกผลิตโดยรัฐบาล หรือหน่วยงานเอกชน ในวาระโอกาสพิเศษต่าง ๆ เพื่อก่อให้เกิดกระแสค่านิยมการเก็บสะสมเป็นที่ระลึกอีกด้วย

#### 2.1.4 การเกิดของแร่ทองคำ

การเกิดของแร่ทองคำออกเป็น 2 แบบ ตามลักษณะที่พบในธรรมชาติได้ดังนี้

- 1) แบบปฐมภูมิ คือกระบวนการทางธรณีวิทยา มีการผสมทางธรณีเคมีจากน้ำแร่ร้อน ผสมผสานกับสารละลายพวกซิลิกา ทำให้เกิดการสะสมตัวของแร่ทองคำในหินต่าง ๆ เช่น หินอัคนี หินชั้น และหินแปร มีการพบการฝังตัวของแร่ทองคำในหิน หรือสายแร่ที่แทรกอยู่ในหิน ซึ่งส่วนใหญ่จะมองไม่เห็นด้วยตาเปล่า
- 2) แบบทุติยภูมิหรือลานแร่ คือการที่หินที่มีแร่ทองคำแบบปฐมภูมิได้มีการสึกกร่อน และถูกน้ำพัดพาไปสะสมตัวในที่แห่งใหม่ เช่น ตามเชิงเขา ลำห้วย หรือในตะกอนกรวดทรายในลำน้ำ

### 2.1.5 การกำหนดคุณภาพของทองคำ

เกณฑ์การกำหนดคุณภาพของทองคำยังคงใช้ความบริสุทธิ์ของทองคำในการบ่งบอกคุณภาพของทองคำ โดยการคิดเนื้อทองเป็น กะรัต หรือ ทองเค (K) เป็นหน่วยที่ใช้บอกความบริสุทธิ์ของทองคำ ที่จะบอกว่าในทองคำนั้น มีเนื้อทองคำกี่ส่วน ยิ่งตัวเลขสูงก็แสดงว่ามีทองคำอยู่มาก ซึ่งในทองคำบริสุทธิ์ 100เปอร์เซ็นต์ จะคิดเป็น 24 ส่วน หรือเรียกว่า ทองคำ 24K

- 1) ทองคำ 24K คือ มีทองคำแท้เป็นส่วนประกอบอยู่ 24 ส่วน หรือเรียกกันว่าทองคำบริสุทธิ์ 99.99 เปอร์เซ็นต์ (ในทางปฏิบัติสามารถสกัดธาตุเจือปนออกให้หมดได้จึงไม่เป็น 100 เปอร์เซ็นต์) ประเทศที่นิยมใช้ได้แก่ ไทย จีน ฮ่องกง สวิสเซอร์แลนด์ และอินโดนีเซีย
- 2) ทองคำ 22K คือ มีทองคำเป็นส่วนประกอบอยู่ 22 ส่วน หรือ 91.7 เปอร์เซ็นต์ อีก 2 ส่วน เป็นโลหะชนิดอื่นประเทศที่นิยมใช้ได้แก่ อินเดีย
- 3) ทองคำ 21K คือ มีทองคำเป็นส่วนประกอบอยู่ 21 ส่วน หรือ 84.5 เปอร์เซ็นต์ อีก 3 ส่วน เป็นโลหะชนิดอื่นประเทศที่นิยมใช้ได้แก่ กลุ่มประเทศตะวันออกกลาง
- 4) ทองคำ 18K คือ มีทองคำเป็นส่วนประกอบอยู่ 18 ส่วน หรือ 75 เปอร์เซ็นต์ อีก 6 ส่วน เป็นโลหะชนิดอื่นประเทศที่นิยมใช้ได้แก่ ประเทศในแถบยุโรป เช่น อิตาลี สวิสเซอร์แลนด์ ฝรั่งเศส ญี่ปุ่น และสหรัฐอเมริกา
- 5) ทองคำ 14K คือ มีทองคำเป็นส่วนประกอบอยู่ 14 ส่วน หรือ 58.3 เปอร์เซ็นต์ อีก 10 ส่วน เป็นโลหะชนิดอื่น ประเทศที่นิยมใช้ได้แก่ สหรัฐอเมริกา อังกฤษ อเมริกาเหนือ และเยอรมัน
- 6) ทองคำ 10K คือ มีทองคำเป็นส่วนประกอบอยู่ 10 ส่วน หรือ 41.6 เปอร์เซ็นต์ อีก 14 ส่วน เป็นโลหะชนิดอื่นประเทศที่นิยมใช้ได้แก่ สหรัฐอเมริกา และอเมริกาเหนือ
- 7) ทองคำ 9K คือ มีทองคำเป็นส่วนประกอบอยู่ 9 ส่วน หรือ 37.5 เปอร์เซ็นต์ อีก 15 ส่วน เป็นโลหะชนิดอื่นประเทศที่นิยมใช้ได้แก่ อังกฤษ
- 8) ทองคำ 8K คือ มีทองคำเป็นส่วนประกอบอยู่ 8 ส่วน หรือ 33.3 เปอร์เซ็นต์ อีก 16 ส่วน เป็นโลหะชนิดอื่นประเทศที่นิยมใช้ได้แก่ เยอรมัน

สำหรับประเทศไทยนั้นใช้มาตรฐานความบริสุทธิ์ของทองคำอยู่ที่ 96.5% หากเปรียบเทียบกะรัตแล้วจะอยู่ที่ 23.16K ซึ่งสีทองที่ได้นั้นจะเป็นสีเหลืองเข้ม และมีความแข็งของเนื้อทองพอเหมาะ เหมาะสำหรับนำมาทำเป็นเครื่องประดับมากที่สุด

## 2.1.6 หน่วยวัดน้ำหนักทอง

หน่วยการแปลงน้ำหนักของทองคำที่ใช้กันเป็นหน่วยสากล แสดงรายละเอียดดังตารางที่ 2.1

ตารางที่ 2.1 หน่วยวัดน้ำหนัก

หน่วย	ความนิยมในแต่ละประเทศ
กรัม (Grammes)	ถือว่าเป็นหน่วยวัดระดับสากล ใช้กันเป็นส่วนใหญ่
ทรอยออนซ์(Troy Ounces)	เป็นหน่วยวัดน้ำหนักที่ใช้ในการกำหนดราคาซื้อขายกันในตลาดโลก ใช้ในประเทศที่ใช้ภาษาอังกฤษ เช่น อังกฤษ สหรัฐอเมริกา ออสเตรเลีย
ตำลึง, เทล(Taels)	เป็นหน่วยวัดที่นิยมใช้กันในประเทศ จีน, ไต้หวัน และฮ่องกง
โทลา(Tolas)	เป็นหน่วยวัดที่นิยมใช้กันในประเทศ สิงคโปร์, อินเดีย, ปากีสถาน และแถบตะวันออกกลาง
ชี(Chi)	เป็นหน่วยวัดที่นิยมใช้กันในประเทศ เวียดนาม
ดอน(Don)	เป็นหน่วยวัดที่นิยมใช้กันในประเทศ เกาหลีใต้
บาท(Baht)	เป็นหน่วยวัดที่นิยมใช้กันในประเทศไทย
Mesghal	เป็นหน่วยวัดที่นิยมใช้กันในประเทศ อิหร่าน

## 2.1.7 การแปลงหน่วยวัดทองคำ

การแปลงหน่วยวัดทองคำสามารถเปรียบเทียบได้ดังตารางที่ 2.2

ตารางที่ 2.2 การเปรียบเทียบหน่วยวัดทองคำ

1 กิโลกรัม	32.1508 ทรอยออนซ์
1 ทรอยออนซ์	31.1034807 กรัม
1 ตำลึง	37.429 กรัม

1 โทลา	11.6638 กรัม
1 ชี	3.75 กรัม
1 ดอน	3.75 กรัม
1 Mesghal	4.6083 กรัม
1 บาท (ทองคำแท่ง)	15.244 กรัม
1 บาท (ทองรูปพรรณ)	15.16 กรัม
1 บาท	4 สลึง
1 สลึง	10 หุน
1 หุน	0.38 กรัม

ซึ่งการกำหนดน้ำหนักของทองคำในประเทศไทยนั้นมีหน่วยวัดเป็น บาท โดยทองคำแท่ง 1 บาทหนัก 15.244 กรัม และส่วนทองรูปพรรณ 1 บาทหนัก 15.16 กรัม

### 2.1.8 การลงทุนทองคำ [3]

การตั้งราคาทองคำในประเทศไทยอ้างอิง ปัจจัยหลัก คือ Gold spot และ USD-THB XAUUSD คือ ราคาทองคำต่างประเทศ มีการซื้อขายโดยใช้เงินสกุลดอลลาร์ หรือที่ รู้จักในชื่อ “Gold Spot”

การตั้งราคาทองในประเทศไทย มีสูตรคำนวณดังนี้

$$\text{ราคาทองคำ} = ((\text{Spot Gold} + \text{Premium}) \times 32.148 \times \text{THB} \times 0.965) / 65.6$$

$$\text{หรือ} = (\text{Spot Gold} + 2) \times \text{THB} \times 0.473 \quad \text{โดยที่}$$

THB คือ อัตราการแลกเปลี่ยนค่าเงินบาทเทียบกับเงินของประเทศที่ขาย

Premium คือ ต้นทุนในการนำเข้าทองคำจากต่างประเทศ จะมีค่าอยู่ระหว่าง 1 – 2 เหรียญ

32.148 คือ น้ำหนักของทองคำ 1 กิโลกรัม เมื่อเทียบเป็นออนซ์

ซึ่งเป็นทองคำต่างประเทศ ชนิด 99.99%

0.965 คือ ทองคำในประเทศชนิด 96.5% คิดจาก 96.5/100

65.6 คือ น้ำหนักของทองคำชนิด 96.5% 1 กิโลกรัมเมื่อเทียบกับน้ำหนัก 1 บาท

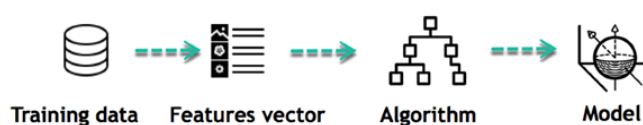
### 2.1.9 ปัจจัยที่กำหนดราคาทองคำในตลาดโลก

- 1) ค่าเงินเหรียญสหรัฐฯ : เมื่อค่าเงินเหรียญสหรัฐฯ มีสัญญาณอ่อนค่าลง ธนาคารกลางประเทศต่าง ๆ ที่ถือครองเงินเหรียญสหรัฐฯ มักจะกระจายความเสี่ยง โดยแบ่งเงินไปลงทุนในสินทรัพย์อื่น เช่น เงินสกุลอื่น ๆ รวมถึงทองคำ ส่งผลให้ราคาทองคำปรับตัวสูงขึ้น
- 2) ความกังวลเรื่องอัตราเงินเฟ้อ : ทองคำเป็นสินทรัพย์ที่จัดว่าให้ผลตอบแทนที่ชนะเงินเฟ้อ เมื่อไหร่ที่เริ่มมีความกังวลว่าเงินเฟ้อจะมากขึ้น มักส่งผลดีต่อทองคำเช่นกัน
- 3) ความเสี่ยงทางการเมืองระหว่างประเทศและระบบการเงิน : ราคาทองคำมักจะปรับตัวเพิ่มขึ้นในช่วงที่มีความตึงเครียดทางการเมืองระหว่างประเทศ และความไม่แน่นอนสูงในระบบการเงินโลก
- 4) อุปสงค์ (Demand) และอุปทาน (Supply) ในตลาด : อุปทานของทองคำหลักๆ แล้วจะมาจากผลผลิตของเหมืองแร่ธนาคารกลาง (แอฟริกาใต้เป็นผู้ผลิตรายใหญ่ 14% ของปริมาณการผลิตทองคำทั่วโลก) ตามมาด้วยเศษทองคำเก่าที่หมุนเวียนอยู่ในระบบ การขายจากหน่วยงานภาครัฐ และการขายล่วงหน้าเพื่อป้องกันความเสี่ยงของผู้ผลิต และในส่วนของอุปสงค์ มีมาจากทั้งภาคเครื่องประดับ ภาคอุตสาหกรรม และการแพทย์ และภาคการลงทุน โดยส่วนใหญ่อุปสงค์ยังคงมาจากภาคเครื่องประดับ
- 5) ค่าเงินบาทเมื่อเทียบกับค่าเงินเหรียญสหรัฐฯ

## 2.2 การถดถอยพหุคูณ (Multiple Regression) [4]

### 2.2.1 การเรียนรู้ของเครื่องจักร (Machine Learning)

การเรียนรู้ของเครื่องจักรเป็น ระบบที่สามารถเรียนรู้ได้ด้วยตนเองจากตัวอย่างข้อมูลที่ได้รับเข้ามาเพื่อวิเคราะห์หารูปแบบ (Pattern) ของข้อมูล โดยเครื่องจักร (Machine) จะเรียนรู้ผ่านการค้นพบรูปแบบซ้ำ ๆ ของข้อมูลผ่านอัลกอริทึมที่ใช้ และแปลงสิ่งที่ค้นพบให้กลายเป็นโมเดล (Model)



ภาพที่ 2.1 ขั้นตอนการเรียนรู้เพื่อสร้างโมเดล

ที่มา <https://www.thaiprogrammer.org>

เมื่อโมเดลถูกสร้างขึ้น จะถูกทดสอบกับข้อมูลที่ไม่เคยเจอมาก่อนเป็นการฝึก (train) ข้อมูลให้สามารถเข้าใจความสัมพันธ์ระหว่างข้อมูลนำเข้าและข้อมูลขาออกอย่างอัตโนมัติ เมื่อเครื่องจักรสิ้นสุดการเรียนรู้ จะสามารถทำการตัดสินใจหรือทำนายข้อมูลใหม่ได้อย่างชัดเจน



ภาพที่2.2 ขั้นตอนการทำนายผลจากโมเดล

ที่มา <https://www.thaiprogrammer.org>

การเรียนรู้ของเครื่องจักรสามารถแบ่งออกได้ 2 แบบใหญ่ๆ ได้แก่ การเรียนรู้แบบมีผู้สอน (Supervised Learning) และการเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning)

### 2.2.2 การเรียนรู้แบบมีผู้สอน (Supervised Learning)

เป็นกลุ่มของอัลกอริทึมที่เน้นสอนคอมพิวเตอร์ โดยการศึกษาจากข้อมูลตัวอย่าง เพื่อให้คอมพิวเตอร์สามารถหาคำตอบของปัญหา (การแก้ปัญหา) ได้ด้วยตัวเอง หลังจากเรียนรู้จากชุดข้อมูลตัวอย่างที่ได้ป้อนให้ไปแล้วระยะหนึ่ง ประเภทของ การเรียนรู้แบบมีผู้สอนอยู่ 2 ประเภท คือ การถดถอย (Regression) และ การแบ่งแยกประเภท (Classification) ส่วนอัลกอริทึมที่ใช้หาก็คจะมี การถดถอยเชิงเส้น (Linear Regression) , ต้นไม้ตัดสินใจ (Decision Tree) , Naïve Bayes, Gradient-boosting trees เป็นต้น

### 2.2.3 การเรียนรู้แบบไม่มีผู้สอน(Unsupervised Learning)

การเรียนรู้แบบไม่มีผู้สอน เป็นกลุ่มอัลกอริทึมที่สร้างขึ้นเพื่อพยายามหาคุณสมบัติของข้อมูล จัดแบ่งกลุ่ม และสร้างนิยามขึ้นมาเอง ประเภทของการเรียนรู้แบบไม่มีผู้สอนส่วนใหญ่จะเป็น การแบ่งกลุ่มข้อมูล (Clustering) ส่วนอัลกอริทึมที่ใช้หาก็คจะมี การแบ่งกลุ่มข้อมูลแบบเคมีน (K-means clustering) , แบบจำลองการผสมของGauss (Gaussian mixture model) , การแบ่งกลุ่มข้อมูลตามลำดับชั้น (Hierarchical clustering) เป็นต้น

## 2.2.4 การถดถอยเชิงเส้นอย่างง่าย (Simple Linear Regression) [5]

การถดถอยเชิงเส้นอย่างง่าย (Simple Linear Regression) เป็นการศึกษาความสัมพันธ์ระหว่าง 2 ตัวแปร ประกอบด้วยตัวแปรอิสระหรือตัวแปรต้น (X) ที่ทำหน้าที่พยากรณ์ค่าของตัวแปรตาม (Y) อย่างละหนึ่งตัว ว่าจะมีเท่าใดหรือมีความสัมพันธ์กันอย่างไร มีความสัมพันธ์กันในลักษณะเชิงเส้น (Linear) โดยสมการถดถอยอย่างง่ายเขียนได้ดังนี้  $Y = b_0 + b_1X$

เมื่อ  $Y$  = ตัวแปรตาม (เนื่องจากค่าของ  $Y$  ขึ้นอยู่กับค่าของ  $X$ )

$X$  = ตัวแปรอิสระหรือตัวแปรต้น

$b_0$  = ค่าคงที่ (Constant) เป็นค่าที่ติดกันกับแกน  $Y$

$b_1$  = ความชัน (Slope) ของเส้นกราฟ

## 2.2.5 การถดถอยพหุคูณ (Multiple Linear Regression)

เป็นการศึกษาความสัมพันธ์ระหว่างตัวแปรอิสระหรือตัวแปรต้นที่ทำหน้าที่พยากรณ์ตั้งแต่ 2 ตัวขึ้นไป กับตัวแปรตาม 1 ตัว

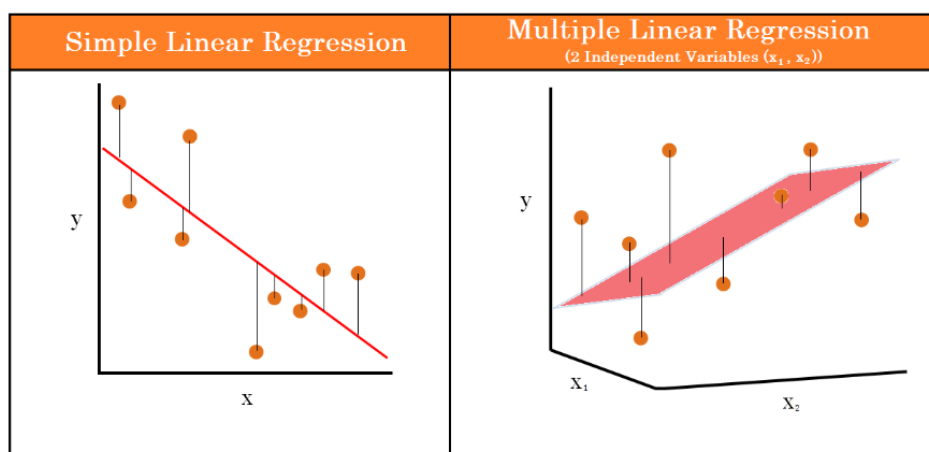
สมการถดถอยอย่างง่ายเขียนได้ดังนี้  $Y = b_0 + b_1X_1 + \dots + b_nX_n$

เมื่อ  $Y$  = ตัวแปรตาม (เนื่องจากค่าของ  $Y$  ขึ้นอยู่กับค่าของ  $X$ )

$X_i$  = ตัวแปรอิสระหรือตัวแปรต้น ตั้งแต่  $i$  ที่ 1 ถึง  $n$  ตัว

$b_0$  = ค่าคงที่ (Constant) เป็นค่าที่ติดกันกับแกน  $Y$

$b_i$  = ความชัน (Slope) ของเส้นกราฟ ตั้งแต่  $i$  ที่ 1 ถึง  $n$  ตัว



ภาพที่ 2.3 การเปรียบเทียบระหว่างการถดถอยเชิงเส้นอย่างง่ายและการถดถอยพหุคูณ

<https://www.keboola.com/blog/linear-regression-machine-learning>

## 2.2.6 อนุกรมเวลา (Time Series) [6]

อนุกรมเวลา หมายถึงข้อมูลที่ถูกจัดเรียงตามเวลาที่ข้อมูลนั้นได้ถูกบันทึก เช่น ปริมาณความชื้นในอากาศในแต่ละวัน จำนวนการเกิดอุบัติเหตุในแต่ละเดือน ราคาหุ้นในแต่ละวัน ดังนั้นในการวิเคราะห์อนุกรมเวลาจึงเป็นการนำเอาเทคนิคต่าง ๆ รวมถึงเทคนิคการทำเหมืองข้อมูล (Data Mining) และการเรียนรู้ของเครื่องจักร (Machine Learning) มาทำการศึกษาความเคลื่อนไหว เพื่อให้สามารถพยากรณ์ข้อมูลในอนาคตได้แม่นยำขึ้น ตัวอย่างข้อมูลอนุกรมเวลาที่ใช้ในงานวิจัยนี้ แสดงดังตารางที่ 2.3

ตารางที่ 2.3 ตัวอย่างข้อมูลอนุกรมเวลาของราคาทองคำ

Date	Gold
2020-09-09	1944.7
2020-09-10	1954.2
2020-09-11	1937.8
2020-09-14	1953.1
2020-09-15	1956.3

## 2.3 งานวิจัยที่เกี่ยวข้อง

### 2.3.1 การพยากรณ์ราคาทองคำด้วยการวิเคราะห์องค์ประกอบหลักและตัวแบบการ

#### ถดถอยเชิงเส้นแบบพหุคูณ [7]

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาปัจจัยที่มีความสัมพันธ์กับราคาทองคำและเพื่อหาตัวแบบที่เหมาะสมสำหรับการพยากรณ์ราคาทองคำ ได้มีการทดลองโดย ผู้ช่วยศาสตราจารย์วิวรรณ กาญจนวชิ โดยนำข้อมูลราคาปิดรายวัน ตั้งแต่วันที่ 1 เดือนมกราคม พ.ศ. 2550 จนถึงวันที่ 1 เดือนพฤศจิกายน พ.ศ. 2558 ตัวแปรที่ใช้ในการวิจัยได้แก่ ตัวแปรตามราคาทองคำ (GOLDPRICE) และตัวแปรต้นประกอบด้วย ราคาน้ำมันดิบ (NYMEX) ราคาแร่โลหะเงิน (SILVER) อัตราแลกเปลี่ยนเงินดอลลาร์สหรัฐเมื่อเทียบกับยูโร (USEURO) ดัชนีของตลาดหุ้นสหรัฐอเมริกา (DJIA) ดัชนีของตลาดหุ้น S&P (SP500) ดัชนีของตลาดหุ้นประเทศเยอรมนี (DAX) และดัชนีตลาดหุ้นประเทศอังกฤษ (FSTE)



เมื่อนำข้อมูลมาหาความสัมพันธ์พบว่า ความสัมพันธ์ของราคาทองและตัวแปรต้นที่เป็นปัจจัยนั้นมีความสัมพันธ์กัน โดยปัจจัยที่มีความสัมพันธ์กับราคาทองสูงสุดคือ ดัชนีของตลาดหุ้นสหรัฐอเมริกา เมื่อหาตัวแบบที่เหมาะสมสำหรับการพยากรณ์ราคาทองคำด้วยวิธี แวริแมกซ์ (Varimax Method) พบว่า องค์ประกอบของปัจจัยที่มีความสัมพันธ์ราคาทองคำ มีทั้งหมด 7 องค์ประกอบ แต่เมื่อพิจารณาความเหมาะสมของค่าน้ำหนักแล้ว ผู้วิจัยได้ตัดองค์ประกอบที่มีค่าน้ำหนักที่น้อยเกินไปออก ทำให้เหลือองค์ประกอบที่ใช้ได้จริง 5 องค์ประกอบ อันได้แก่ PCA1 PCA2 PCA3 PCA4 และ PCA5

เมื่อได้องค์ประกอบหลักแล้วจึงนำมาหาค่าสัมประสิทธิ์สหสัมพันธ์เพื่อพิจารณาความสัมพันธ์กับราคาทองคำพบว่า ความสัมพันธ์ของราคาทองคำและองค์ประกอบหลักที่เป็นปัจจัยนั้นมีความสัมพันธ์ที่เป็นบวก กับราคาทองคำ ได้แก่ PCA1 ( $r=0.097$ ) PCA2 ( $r=0.209$ ) PCA3 ( $r=0.864$ ) ปัจจัยมีความสัมพันธ์ที่เป็นลบกับราคาทองคำ ได้แก่ PCA4 ( $r=-0.274$ ) PCA5 ( $r=-0.146$ ) ดังนั้นองค์ประกอบหลักทั้งหมดที่มีความสัมพันธ์กับราคาทองคำอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 จึงนำมาพิจารณาสร้างสมการถดถอย เชิงเส้นพหุคูณ ซึ่งใช้วิธีการนำตัวแปรเข้าด้วยวิธี Stepwise ได้สมการที่เหมาะสมให้ค่าสัมประสิทธิ์สหสัมพันธ์พหุคูณระหว่างปัจจัยเหล่านี้กับราคาทองคำเป็น 0.947 (R) และปัจจัยเหล่านี้สามารถอธิบายความแปรปรวนของราคาทองคำได้ร้อยละ 89.6 (R<sup>2</sup>) ซึ่งเป็นสมการที่ให้ค่าเหมาะสมที่สุด ผู้วิจัยจึงเลือกเป็น สมการพยากรณ์ราคาทองคำ สามารถนำมาเขียนเป็นสมการถดถอยเชิงเส้นแบบพหุคูณ เพื่อพยากรณ์ราคา ทองคำได้ดังนี้

$$Z_{\text{GOLDPRICE}} = 0.864\text{PCA}_3 - 0.274\text{PCA}_4 + 0.209\text{PCA}_2 - 0.146\text{PCA}_5 + 0.097\text{PCA}_1$$

### 2.3.2 การพยากรณ์แนวโน้มอัตราแลกเปลี่ยนเงินตราต่างประเทศโดยใช้อัตราเวลา [6]

งานวิจัยนี้เป็นงานวิจัยที่มีจุดมุ่งหมายเพื่อพยากรณ์แนวโน้มอัตราแลกเปลี่ยนเงินตราต่างประเทศโดยใช้อัตราเวลา ได้มีการทดลองโดย สมร เหล็กหล้า และ จาริ ทองคำ การพยากรณ์อัตราแลกเปลี่ยนเงินตราจากการอ้างอิงราคาจากตลาดการแลกเปลี่ยนเงินตราต่างประเทศ โดยใช้ข้อมูล แนวโน้มขาขึ้นของอัตราแลกเปลี่ยนเงินตรา ตั้งแต่เดือนกุมภาพันธ์ ค.ศ. 2014 ถึง เดือน มกราคม ค.ศ.

2017 ในงานวิจัยนี้ 4 เทคนิคที่มีประสิทธิภาพในการสร้างแบบจำลองเพื่อพยากรณ์ คือ Linear Regression (LR) , Multi-Layer 9 Perceptron (MLP) , Support Vector Machine Regression (SVMR) และ Sequential Minimal Optimization Regression (SMOR)

ในการทดสอบประสิทธิภาพของแบบจำลองที่สร้างขึ้น Sliding Windows ได้ถูกนำมาใช้ในการแบ่งกลุ่ม ข้อมูลเป็นชุดข้อมูลการเรียนรู้ และชุดข้อมูลทดสอบ 12 รอบของ Sliding Windows ถูกนำมาใช้เพื่อลดความแปรปรวนของผลการทดลอง ค่าเฉลี่ยความคลาดเคลื่อนสัมบูรณ์ (Mean Absolute Error: MAE) และ Root Mean Square Error (RMSE) ได้ถูกนำมาใช้การวัดประสิทธิภาพของแบบจำลอง ผลการทดลองพบว่า เทคนิค SVMR ดีกว่า LR, MLP and SMOR โดยมีค่า MAE และค่า RMSE ต่ำสุดถึง  $1.11 \pm 2.10$  และ  $1.13 \pm 2.14$  ตามลำดับ

### 2.3.3 ปัจจัยที่มีผลต่อราคาทองคำในตลาดโลก [8]

การศึกษาค้นคว้านี้มีวัตถุประสงค์เพื่อศึกษาปัจจัยที่มีผลต่อราคาทองคำในตลาดโลก ได้มีการทดลองโดย สุภาวดี ศิริวัฒน์ และ นพจักร ทองเรือนดี โดยเลือกศึกษาความสัมพันธ์ระหว่างปัจจัยต่าง ๆ ได้แก่ ราคาน้ำมันดิบในตลาดโลก ราคาแร่โลหะเงินในตลาดโลก ราคาแร่โลหะแพลทินัมในตลาดโลก ราคาแร่โลหะพัลลาเดียมในตลาดโลก อัตราแลกเปลี่ยนเงินดอลลาร์ สหรัฐอัตราแลกเปลี่ยนเงินยูโร และอัตราดอกเบี้ย นโยบายประกาศโดยรัฐบาลสหรัฐอเมริกา โดยใช้ข้อมูลทุติยภูมิ (Secondary Data) แบบรายเดือน ตั้งแต่เดือนมกราคม 2550 ถึงเดือนธันวาคม 2554 เป็นจำนวน 60 เดือน มาวิเคราะห์หาความสัมพันธ์ด้วยวิธีทางเทคนิคของสมการถดถอยโดยทดสอบด้วยโปรแกรม Eviews ซึ่งเป็น โปรแกรมทางสถิติ เพื่อดำเนินการหาความสัมพันธ์ดังกล่าว

ผลการศึกษาปัจจัยที่มีผลต่อราคาทองคำในตลาดโลก พบว่า ราคาน้ำมันดิบในตลาดโลก ราคาแร่โลหะ เงินในตลาดโลก ราคาแร่โลหะแพลทินัมในตลาดโลก ราคาแร่โลหะพัลลาเดียมในตลาดโลก อัตราแลกเปลี่ยน เงินยูโร เป็นปัจจัยที่ส่งผลกระทบต่อเปลี่ยนแปลงของราคาทองคำในตลาดโลกอย่างมีนัยสำคัญที่ระดับนัยสำคัญทาง สถิติที่ 0.05 ส่วนอัตราดอกเบี้ยนโยบายประกาศโดยรัฐบาลสหรัฐอเมริกาไม่มี

ผลกระทบต่อราคาทองคำใน ตลาดโลก และยังพบว่าตัวแบบที่ได้ทำการศึกษาใน ครั้งนี้สามารถพยากรณ์ทิศทางราคาทองในตลาดโลกโดยมี ความแม่นยำ 89.75%

#### 2.3.4 การพยากรณ์ราคาทองคำโดยวิธีอาร์มา [9]

การศึกษานี้มีวัตถุประสงค์เพื่อที่จะพยากรณ์ราคาทองคำและ ทองรูปพรรณ โดยใช้ข้อมูลรายเดือน จำนวน 120 เดือน ตั้งแต่ปี 2537 ถึง 2546 วิเคราะห์ข้อมูลโดยใช้แบบจำลองอาร์มาด้วยวิธีบ็อกส์และเจนกินส์ (Box-Jenkins) จากการศึกษาพบว่าข้อมูลราคามีลักษณะไม่นิ่ง และเมื่อทดสอบความนิ่งของข้อมูล พบว่าข้อมูลนี้ระดับที่ 1 ทั้งจากการพิจารณาคอเรโลแกรมพบว่าแบบจำลองที่มีค่า Autoregressive (AR) และ Moving Average (MA) ให้ ค่าสถิติ Root Mean Squared Error (RMSE) และ Theil Inequality Coefficient (U) ที่มีค่าต่ำที่สุด เท่ากับ 0.020343 และ 0.001139 ตามลำดับ มีความเหมาะสมที่สุดที่จะเป็น ตัวแทนของราคาขายทองคำและทองรูปพรรณเพื่อการพยากรณ์ในอนาคต สามารถ สรุปได้ว่าผลจากการพยากรณ์สามารถนำไปใช้ประโยชน์ให้มีประสิทธิภาพมากยิ่งขึ้น

## บทที่ 3

### วิธีการดำเนินงานวิจัย

งานวิจัยฉบับนี้มีวัตถุประสงค์ คือ เพื่อสร้างโมเดลการทำนายราคาทองคำ ในอีก 3 อาทิตย์ข้างหน้า โดยใช้เทคนิคการเรียนรู้ของเครื่องจักรในการแก้ปัญหา ซึ่งได้มีการวางแผนดำเนินการดังนี้

- 1) กำหนดแนวทางที่ใช้ในการทำนายราคาทองคำ (Defining the Approach)
- 2) การรวบรวมและเตรียมข้อมูล (Gathering and Preparing Data)
- 3) การสำรวจลักษณะของข้อมูล (Data Visualization)
- 4) การทำความสะอาดข้อมูล (Cleaning Data)
- 5) การฝึกฝนแบบจำลอง (Training Models)
- 6) การเตรียมแบบจำลอง (Model preparation)
- 7) การเรียกใช้แบบจำลอง (Importing a model)
- 8) การทำนายราคาทองคำล่วงหน้า (Gold price prediction in advance)

#### 3.1 กำหนดแนวทางที่ใช้ในการทำนายราคาทองคำ (Defining the Approach)

เนื่องจากการทำนายราคาทองคำโดยใช้วิธีการเรียนรู้ของเครื่องจักรและเป็นการเรียนรู้แบบมีผู้สอน ดังนั้นผู้วิจัยจึงเลือกใช้ การวิเคราะห์การถดถอยพหุคูณ (Multiple Linear Regression) เป็นอัลกอริทึมที่ใช้สร้างโมเดล จากนั้นจะวิเคราะห์อนุกรมเวลาเพื่อหาความสัมพันธ์ของตัวแปรที่อยู่ห่างกันในช่วงเวลาต่าง ๆ ตามลำดับเวลาเพื่อทำนายราคาทองคำล่วงหน้า ในการทำนายจะใช้ข้อมูลในอดีตของสินทรัพย์ต่าง ๆ ที่มีผลต่อแนวโน้มของราคาทองคำมาคำนวณ ดังนั้นจึงจำเป็นต้องนำเข้าสู่ข้อมูลจำนวนมาก โดยปัจจุบันข้อมูลราคาทองคำสามารถสืบค้นได้ด้วยตนเองจากเว็บไซต์ต่าง ๆ เช่น IEX, Quandl, Yahoofinance, Google finance

#### 3.2 การรวบรวมและเตรียมข้อมูล (Gathering and Preparing Data)

เว็บไซต์ Yahoofinance มีการเก็บข้อมูลทางการเงินและการลงทุนเป็นจำนวนมาก ผู้วิจัยจึงเลือกนำเข้าราคาปิดของสินทรัพย์ประเภทต่าง ๆ หลายรายการในช่วง 10 ปีที่ผ่านมา เนื่องจากแพ็คเกจของ Yahoofinance ต้องใช้สัญลักษณ์ของ Yahoo ทำให้ยากต่อความเข้าใจ จึงต้องมีการนำเข้าไฟล์ที่ใช้อธิบายสัญลักษณ์ เมื่อได้รายการสินทรัพย์แล้ว ต้องมีการกำหนดช่วงวันที่ที่ต้องนำเข้าสู่ข้อมูล ช่วงเวลาที่เลือกคือตั้งแต่ เดือนมกราคม ปี ค.ศ.2010 จนถึง เดือนมีนาคม ปี ค.ศ.2020 สาเหตุที่ไม่ดึงข้อมูลก่อนหน้าปี ค.ศ.2010 เนื่องจากเป็นเพราะวิกฤตการเงินโลกในปี

ค.ศ.2008-ค.ศ.2009 ทำให้ภูมิทัศน์ทางเศรษฐกิจและตลาดเปลี่ยนแปลงไปอย่างมาก ความสัมพันธ์ของสินทรัพย์ในช่วงเวลานั้นอาจมีความเกี่ยวข้องกันน้อย โดยข้อมูลที่ผู้วิจัยเลือก นำเข้าคือราคาปิดของรายการสินทรัพย์ ได้แก่ Gold, Silver, Crude Oil, S&P500, Russel 2000 Index, 10 Yr US T-Note futures, 10 Yr US T-Note futures, Platinum, Copper, Dollar Index, Volatility Index, Soybean, MSCI EM ETF, Euro USD, Euronext100, Nasdaq

ใน python สามารถติดตั้ง/ลงแพ็คเกจ ได้โดยใช้คำสั่ง : Pip install yahoofinance

ตัวอย่างการนำเข้า Libraries ที่ใช้ในงานวิจัย

Import pandas as pd (ใช้สำหรับการจัดการข้อมูล)

From datetime import datetime (ใช้สำหรับการเข้าถึงวันและเวลา)

Import matplotlib.pyplot as plt (ใช้สำหรับการแสดงในรูปแบบของกราฟ)

From yahoofinancials import yahoofinancials (เป็นช่องทางที่ใช้ในการดึงข้อมูล)

ตัวอย่างรายละเอียดของชื่อสัญลักษณ์และคำอธิบายที่ใช้ในการทำนาย แสดงดังตารางที่ 3.1

ตารางที่ 3.1 รายละเอียดข้อมูลที่ใช้ทำนายสัญลักษณ์

	Ticker	Description
0	GC=F	Gold
1	SI=F	Silver
2	CL=F	Crude Oil
3	^GSPC	S&P500
4	^RUT	Russel 2000 Index
5	ZN=F	10 Yr US T-Note futures
6	ZT=F	2 Yr US T-Note Futures
7	PL=F	Platinum
8	HG=F	Copper
9	DX=F	Dollar Index
10	^VIX	Volatility Index
11	S=F	Soybean
12	EEM	MSCI EM ETF
13	EURUSD=X	Euro USD
14	^N100	Euronext100
15	^IXIC	Nasdaq

### 3.3 การสำรวจลักษณะของข้อมูล (Data Visualization)

ในขั้นตอนนี้เป็นการสำรวจลักษณะของข้อมูล โดยการนำข้อมูลดิบที่ได้มาเปลี่ยนเป็นกราฟ แผนภูมิหรือตาราง ที่ช่วยอธิบายปริมาณตัวเลขและเพื่อสร้างความเข้าใจให้แก่ผู้วิจัยว่าลักษณะข้อมูลมีรูปแบบไปในทิศทางใด เป็นการสร้างคุณค่าให้กับข้อมูล ทำให้สามารถค้นพบรูปแบบใหม่ๆ และมองเห็นแนวโน้มของสินทรัพย์มากขึ้นหรือสินทรัพย์ใดที่ส่งผลกระทบต่อ การเปลี่ยนแปลงราคาทองคำ

### 3.4 การทำความสะอาดข้อมูล (Cleaning Data)

การทำความสะอาดข้อมูลเป็นขั้นตอนหนึ่งในการเตรียมข้อมูลก่อนที่จะนำข้อมูลที่เราเลือกนั้นไปใช้งานในการฝึกฝนแบบจำลอง เนื่องจากชุดข้อมูลที่มีบางครั้งมีความไม่สมบูรณ์อยู่หรือข้อมูลที่มีความไม่สัมพันธ์กับข้อมูลอื่น ๆ จึงจำเป็นต้องมีการตรวจสอบหรือลบข้อมูลบางรายการที่ไม่ถูกต้องออกไปจากชุดข้อมูล เพื่อให้ข้อมูลมีคุณภาพ ในการทำนายราคาทองคำ ข้อมูลที่เก็บมาจะมีลักษณะเป็นราคาปิดของสินทรัพย์หลายๆรายการ และสินทรัพย์เหล่านี้อาจมีวันหยุดตามภูมิภาคและวันหยุดการซื้อขายที่แตกต่างกัน ทำให้ช่วงวันที่สำหรับการดึงข้อมูลทุกครั้งไม่เหมือนกัน ซึ่งการซื้อขายทองคำจะมีการหยุดทุก ๆ วันเสาร์อาทิตย์ ดังนั้นจึงมีการลบข้อมูลช่วงวันเสาร์อาทิตย์ออกจากสินทรัพย์ทุกรายการก่อนนำข้อมูลมาใช้

ในช่วงข้อมูลจะมีค่าว่างเกิดขึ้น ผู้วิจัยได้เลือกการแทนที่ค่าว่างเหล่านั้นด้วยค่าก่อนหน้าของค่าว่างแต่ละตัวทั้งหมด หลังจากการแทนที่ด้วยค่าว่างก่อนหน้าเสร็จแล้ว หากยังมีค่าว่างที่ยังไม่ถูกแทน (ในกรณีที่ค่าว่างอยู่บนสุดของข้อมูล) จึงมีการแทนค่าว่างด้วยค่าถัดไปของค่าว่างนั้น เพื่อให้ข้อมูลนั้นมีความสมบูรณ์. ในงานวิจัยนี้ใช้การแทนค่าว่างด้วยค่าก่อนหน้าและค่าถัดไป เพราะง่ายต่อการแทนค่า แต่ยังมีอีกหลายวิธีที่ใช้ในการแทนค่า ตัวอย่างเช่น การใส่ค่าว่างด้วยค่าสุ่ม (Random) การใส่ค่าว่างด้วยค่าเฉลี่ย (Average) เป็นต้น

### 3.5 การฝึกฝนแบบจำลอง (Training Models)

เป็นขั้นตอนการฝึกฝนแบบจำลองให้เรียนรู้ความสัมพันธ์ของชุดข้อมูล เพื่อนำไปใช้ทำนายราคาทองคำในอีก 2-3 อาทิตย์ข้างหน้า โดยแบบจำลองที่เลือกใช้ คือ Multiple Linear Regression โดยทั่วไปจะแบ่งออกเป็น 2 ส่วน มีอัตราส่วน 70:30 โดยส่วน 70% แรกใช้สำหรับเรียนรู้ และอีก 30% เป็นส่วนที่ใช้ทดสอบความแม่นยำ

จากสมการถดถอย  $Y = b_0 + b_1X_1 + \dots + b_nX_n$  เป็นการศึกษาความสัมพันธ์ระหว่างตัวแปรต้นที่ทำหน้าที่พยากรณ์ตั้งแต่ 2 ตัวขึ้นไป กับตัวแปรตาม 1 ตัว โดยในงานวิจัยนี้จะหมายถึง

$Y$  = ราคาทองคำ (GOLDPRICE) เป็นตัวแปรตาม

$X_i$  = สินทรัพย์ที่เลือกมา (Silver, Crude Oil, S&P500, Russel 2000 Index, 10 Yr US T-Note futures, 10 Yr US T-Note futures, Platinum, Copper, Dollar Index, Volatility Index, Soybean, MSCI EM ETF, Euro USD, Euronext100, Nasdaq) ทั้งหมด 16 ตัว เป็นตัวแปรต้นที่ทำหน้าที่พยากรณ์

จากสมการถดถอยเชิงเส้นแบบพหุคูณ สามารถนำมาเขียนเป็นสมการพยากรณ์ราคาทองคำ เพื่อใช้ในการพยากรณ์ ได้ดังนี้

$$\text{GOLDPRICE} = b_0 + b_1X_{\text{Silver}} + b_2X_{\text{CrudeOil}} + b_3X_{\text{S\&P500}} + b_4X_{\text{Russel}} + \dots + b_{16}X_{\text{Nasdaq}}$$

### 3.6 การเตรียมแบบจำลอง (Model preparation)

การเตรียมแบบจำลองเป็นการเตรียมไว้ใช้กับจุดประสงค์ของการทำนายในแต่ละการทำนาย เช่นหากต้องการทำนายราคาทองคำในอีก 14 วันข้างหน้า ควรเตรียมแบบจำลองการทำนายใน 14 วันข้างหน้าก่อน หรือ หากต้องการทำนายอีก 1 วันข้างหน้า ก็ควรเตรียมแบบจำลองการทำนายใน 1 วันข้างหน้าก่อน เพื่อตรงจุดประสงค์ในเวลาที่ต้องการทำนาย จากนั้นบันทึกแบบจำลองไว้ใช้ ข้อมูลที่เป็น Time Series

### 3.7 การเรียกใช้แบบจำลอง (Importing a model)

เมื่อต้องการเรียกใช้แบบจำลองกับข้อมูลใหม่ ควรเตรียมข้อมูลให้พร้อมก่อนการเรียกใช้ เพื่อให้ข้อมูลมีความถูกต้อง ครบถ้วน สมบูรณ์

### 3.8 การทำนายราคาทองคำล่วงหน้า (Gold price prediction in advance)

หลังจากได้แบบจำลองที่เหมาะสมไว้ใช้ทำนายแล้ว ในขั้นตอนนี้จะเป็นการทำนายโดยใช้แบบจำลองที่บันทึกไว้ ผลลัพธ์ที่ได้ คือ วันที่เริ่มทำนาย (Date) , ราคาทองคำในวันที่เริ่มทำนาย (Gold) , แนวโน้มของราคาทองคำ (Return) , ราคาทองคำที่ทำนายได้ (Gold-T+..) , วันที่ทำนาย(Date-T+...) และกราฟราคาทองคำที่ทำนายไว้ เทียบกับราคาทองคำจริง

## บทที่ 4

### ผลการดำเนินงานและผลการวิเคราะห์ข้อมูล

การทำนายราคาทองคำล่วงหน้าและการวิเคราะห์แบบจำลองที่เหมาะสมกับการทำนายผล ซึ่งผลวิจัยดังกล่าวจะถูกนำเสนอออกเป็นหัวข้อย่อย ๆ ดังต่อไปนี้

#### 4.1 การหาแบบจำลองที่เหมาะสมกับการทำนายผล

เมื่อทำการเตรียมข้อมูลที่จะใช้ทำนายเสร็จแล้ว จะเป็นการเลือกตัวแบบจำลองไว้ใช้ในการทำนายโดยเลือกแบบจำลอง Regressor ที่มีค่า Metrics [10] [11] [12] เหมาะสมที่สุด จากการเปรียบเทียบแบบจำลองกับข้อมูลที่ให้

	Model	MAE	MSE	RMSE	R2	RMSLE	MAPE	TT (Sec)
0	Extra Trees Regressor	0.0124	0.0003	0.0167	0.8556	0.0152	-0.0479	1.6874
1	CatBoost Regressor	0.0132	0.0003	0.0175	0.8410	0.0159	-0.0546	27.1378
2	K Neighbors Regressor	0.0132	0.0003	0.0182	0.8275	0.0155	-0.0327	0.0412
3	Light Gradient Boosting Machine	0.0138	0.0003	0.0183	0.8268	0.0164	-0.1669	0.8458
4	Random Forest	0.0154	0.0004	0.0208	0.7748	0.0189	-0.1178	4.3180
5	Extreme Gradient Boosting	0.0157	0.0004	0.0211	0.7703	0.0185	-0.1391	1.7819

ภาพที่4.1 การเปรียบเทียบแบบจำลอง

ในงานวิจัยนี้ แบบจำลองที่ผู้วิจัยได้เลือกคือ Extra Trees Regressor, CatBoost Regressor, K- Neighbors Regressor, TheilSen Regressor, และ Automatic Relevance Determination. จากนั้นทำการสร้างแบบจำลองขึ้นมาเพื่อเปรียบเทียบค่า Metrics ของแต่ละโมเดล โดยค่า Metrics คือค่าที่ได้จากการคำนวณเปรียบเทียบผลการทำนายกับค่าจริง (prediction vs. actual) ว่าโมเดลทำนายได้ถูกต้องแค่ไหน. สำหรับ Metrics ประเภท Regression [13] [14] ได้แก่ MAE, MSE, RMSE, R2, RMSLE และ MAPE

- 1) MAE (Mean Absolute Error) คือการคำนวณหาผลรวมของค่า absolute (error) แล้วคูณกับ  $1/n$  เพื่อเปลี่ยนเป็นค่าเฉลี่ย.
- 2) MSE (Mean Squared Error) คล้ายกับ MAE แต่เปลี่ยนจากการทำ absolute เป็น squared (ยกกำลังสองค่า error) ก่อนหาค่าเฉลี่ย.
- 3) RMSE (Root Mean Square Error) คือการทำ square root ค่า MSE (ถ้า RMSE เท่ากับ 2.56 แปลว่าโดยเฉลี่ยโมเดลทำนาย y ผิดไป +/- 2.56 point)



- 4) R2 คือ variance ที่โมเดลของเราอธิบายได้เป็นสัดส่วนจาก total variance ทั้งหมดของข้อมูลชุดนั้น (R2 จะมีค่าอยู่ระหว่าง 0-1 ยิ่งเข้าใกล้ 1 แปลว่าโมเดลเราทำนายผลได้ดีมาก)
- 5) RMSLE (Root Mean Squared Log Error) คือ root mean squared error (RMSE) ของ target variable ที่ take log เสร็จแล้ว.
- 6) MAPE (Mean Absolute Percentage Error) มีลักษณะคล้าย MAE แต่เปลี่ยนค่าเป็นร้อยละ (percent)

MAE, MSE, MAPE, RMSE มีค่ายิ่งต่ำยิ่งดี ถ้าเท่ากับ 0 แปลว่าโมเดลทำนายค่า  $y$  ได้ถูกต้อง 100% แต่ในทางปฏิบัติโอกาสที่จะ Train Model ได้ loss = 0 เป็นไปได้นยาก

## 4.2 Compare Models

ไวยากรณ์ คือ

```
compare_models(blacklist = None, whitelist = None, fold = 10, round = 4, sort =
'R2', n_select = 1, turbo = True, verbose = True)
```

ฟังก์ชันนี้จะเปรียบเทียบโมเดลทั้งหมดพร้อมกันและให้คะแนน(scores)โดยใช้ K-fold Cross Validation ที่ถูกตั้งค่าเริ่มต้นเท่ากับ 10 ผลลัพธ์จะพิมพ์ตารางแสดงค่า MAE, MSE, RMSE, R2, RMSLE และ MAPE ที่ดีที่สุดโดยอัตโนมัติ ตารางจะเรียงลำดับจากสูงสุดไปต่ำสุด. ตามค่าเริ่มต้นของตารางจะเรียงลำดับตามความแม่นยำโดยใช้ R2 ในการเรียง. ในการเลือกรุ่นโมเดลที่อยู่บนสุดให้ใช้พารามิเตอร์ n\_select ที่ตั้งค่าเป็น 1 ตามค่าเริ่มต้น.

## 4.3 Create Model [14] [15]

ไวยากรณ์ คือ

```
create_model(estimator = None, ensemble = False, method = None, fold = 10,
round = 4, cross_validation = True, verbose = True, system = True, **kwargs)
```

ฟังก์ชันจะสร้างโมเดลและให้คะแนน (scores) โดย K-fold Cross Validation ที่กำหนดค่าเริ่มต้นเท่ากับ 10. ผลลัพธ์จะพิมพ์ตารางแสดงค่า Metrics ที่ไว้ใช้ประเมินได้แก่ MAE, MSE, RMSE, RMSLE, R2 และ MAPE. ฟังก์ชันนี้จะส่งคืนอ็อบเจกต์โมเดลที่ได้รับการฝึก(trained model object). การสร้างโมเดลสามารถทำได้เพียงเขียน create\_model และใช้พารามิเตอร์เพียงตัวเดียวคือ Model ID. ตัวอย่างเช่น create\_model('knn') หมายถึงการสร้างโมเดล K-Neighbors Regressor.

K-Fold Cross Validation คือการที่แบ่งข้อมูลเป็นจำนวน K ส่วนเท่า ๆ กัน เพื่อสร้างและทดสอบโมเดล (train + validate) โดยการในแต่ละส่วนจะต้องมาจากสุ่ม (Random) เพื่อที่จะให้ข้อมูลกระจายเท่า ๆ กัน.

```
tr = create_model('tr')
```

	MAE	MSE	RMSE	R2	RMSLE	MAPE
0	0.0290	0.0014	0.0370	0.1839	0.0287	0.0004
1	0.0308	0.0018	0.0425	-0.2511	0.0265	-2.1669
2	0.0303	0.0018	0.0419	0.2170	0.0319	0.4041
3	0.0308	0.0017	0.0418	0.0770	0.0302	-0.5992
4	0.0281	0.0015	0.0381	0.2136	0.0282	0.1840
5	0.0314	0.0016	0.0400	0.1977	0.0324	-1.0306
6	0.0346	0.0020	0.0448	0.1455	0.0348	1.7746
7	0.0325	0.0020	0.0445	0.2371	0.0368	-0.1693
8	0.0321	0.0018	0.0424	-0.0423	0.0298	0.6993
9	0.0313	0.0016	0.0399	0.0718	0.0294	-0.3223
Mean	0.0311	0.0017	0.0413	0.1050	0.0309	-0.1226
SD	0.0017	0.0002	0.0024	0.1444	0.0030	0.9967

ภาพที่4.2 ค่า Metrics ที่แสดงจากการสร้างแบบจำลอง TheilSen Regressor

```
knn = create_model('knn')
```

	MAE	MSE	RMSE	R2	RMSLE	MAPE
0	0.0145	0.0004	0.0196	0.7704	0.0159	0.4986
1	0.0112	0.0002	0.0148	0.8479	0.0130	-0.8262
2	0.0139	0.0004	0.0205	0.8123	0.0170	-0.2190
3	0.0138	0.0004	0.0194	0.8015	0.0159	0.3063
4	0.0116	0.0002	0.0149	0.8801	0.0132	0.3527
5	0.0137	0.0003	0.0182	0.8343	0.0161	-0.8846
6	0.0137	0.0003	0.0182	0.8590	0.0159	0.1472
7	0.0155	0.0005	0.0217	0.8187	0.0184	0.1047
8	0.0124	0.0003	0.0175	0.8216	0.0147	0.4950
9	0.0121	0.0003	0.0171	0.8291	0.0144	-0.3015
Mean	0.0132	0.0003	0.0182	0.8275	0.0155	-0.0327
SD	0.0013	0.0001	0.0021	0.0292	0.0016	0.4835

ภาพที่4.3 ค่า Metrics ที่แสดงจากการสร้างแบบจำลอง K-Neighbors Regressor

```
et = create_model('et')
```

	MAE	MSE	RMSE	R2	RMSLE	MAPE
0	0.0130	0.0003	0.0176	0.8154	0.0165	0.1759
1	0.0104	0.0002	0.0139	0.8653	0.0127	-0.4533
2	0.0133	0.0003	0.0182	0.8533	0.0163	-0.1426
3	0.0130	0.0003	0.0170	0.8468	0.0157	0.3504
4	0.0107	0.0002	0.0144	0.8872	0.0135	0.0188
5	0.0121	0.0003	0.0162	0.8683	0.0148	-0.4067
6	0.0129	0.0003	0.0183	0.8575	0.0161	0.1034
7	0.0150	0.0004	0.0202	0.8424	0.0183	0.0044
8	0.0118	0.0003	0.0161	0.8493	0.0148	0.3003
9	0.0112	0.0002	0.0149	0.8701	0.0137	-0.4294
Mean	0.0124	0.0003	0.0167	0.8556	0.0152	-0.0479
SD	0.0013	0.0001	0.0019	0.0184	0.0016	0.2842

ภาพที่4.4 ค่า Metrics ที่แสดงจากการสร้างแบบจำลอง Extra Trees Regressor

```
catboost = create_model('catboost')
```

	MAE	MSE	RMSE	R2	RMSLE	MAPE
0	0.0127	0.0003	0.0166	0.8366	0.0155	0.2158
1	0.0112	0.0002	0.0153	0.8384	0.0137	-0.6891
2	0.0145	0.0004	0.0196	0.8291	0.0176	-0.4286
3	0.0138	0.0003	0.0183	0.8235	0.0169	0.4034
4	0.0119	0.0003	0.0159	0.8628	0.0146	0.0120
5	0.0125	0.0003	0.0163	0.8669	0.0148	-0.6946
6	0.0135	0.0004	0.0188	0.8487	0.0165	0.1325
7	0.0161	0.0005	0.0214	0.8235	0.0198	0.0588
8	0.0133	0.0003	0.0177	0.8172	0.0163	0.8270
9	0.0120	0.0002	0.0153	0.8635	0.0139	-0.3835
Mean	0.0132	0.0003	0.0175	0.8410	0.0159	-0.0546
SD	0.0014	0.0001	0.0019	0.0175	0.0018	0.4661

ภาพที่4.5 ค่า Metrics ที่แสดงจากการสร้างแบบจำลอง CatBoost Regressor

```
ard = create_model('ard')
```

	MAE	MSE	RMSE	R2	RMSLE	MAPE
0	0.0286	0.0013	0.0366	0.2012	0.0295	0.0947
1	0.0274	0.0013	0.0364	0.0838	0.0271	-1.8711
2	0.0304	0.0017	0.0409	0.2547	0.0356	0.1868
3	0.0302	0.0016	0.0399	0.1559	0.0302	-0.6246
4	0.0286	0.0014	0.0381	0.2141	0.0304	0.2345
5	0.0300	0.0015	0.0393	0.2234	0.0321	-0.7297
6	0.0338	0.0019	0.0440	0.1738	0.0356	1.5350
7	0.0324	0.0019	0.0439	0.2590	0.0364	-0.1962
8	0.0298	0.0015	0.0390	0.1152	0.0303	0.7929
9	0.0290	0.0014	0.0372	0.1937	0.0308	-0.4407
Mean	0.0300	0.0016	0.0395	0.1875	0.0318	-0.1018
SD	0.0018	0.0002	0.0026	0.0539	0.0029	0.8739

ภาพที่4.6 ค่า Metrics ที่แสดงจากการสร้างแบบจำลอง Automatic Relevance

Determination

#### 4.4 Tune Model [16]

ไวยากรณ์ คือ

```
tune_model(estimator = None, fold = 10, round = 4, n_iter = 10, custom_grid
= None, optimize = 'r2', choose_better = False, verbose = True)
```

ขั้นตอนต่อไปหลังจากได้แบบจำลองแล้ว ก็คือการนำโมเดลมาปรับค่าไฮเปอร์พารามิเตอร์ เพื่อให้ได้ผลลัพธ์ที่ดียิ่งขึ้น ซึ่งใน Pycaret ก็อำนวยความสะดวกโดยการใช้ฟังก์ชัน tune\_model().

ฟังก์ชันนี้จะปรับแต่งพารามิเตอร์ไฮเปอร์พารามิเตอร์ของโมเดลและให้คะแนน (scores) โดยใช้ K-fold Cross Validation ผลลัพธ์จะพิมพ์ตารางคะแนนที่แสดง MAE, MSE, RMSE, R2, RMSLE และ MAPE โดย fold ที่ค่าเริ่มต้นเท่ากับ 10 Folds(จำนวน fold ที่จะใช้ใน K-fold Cross Validation ต้องมีอย่างน้อย 2) ฟังก์ชันนี้จะส่งคืนอ็อบเจกต์โมเดลที่ได้รับการฝึกฝน. Round คือ จำนวนตำแหน่งทดสอบของเมตริก (Metrics) ในตารางคะแนนจะถูกพิเศษให้มี 4 ตำแหน่ง. n\_iter คือ จำนวนการทำซ้ำภายใน Random Grid Search. สำหรับการวนทุกครั้งแบบจำลองจะสุ่มเลือกหนึ่งค่าจาก grid of hyperparameters ที่กำหนดไว้ล่วงหน้า. Optimize

(ตัวปรับแต่ง) คือเป็นการวัดที่ใช้ในการเลือกโมเดลที่ดีที่สุดผ่านการปรับแต่งไฮเปอร์พารามิเตอร์ การปรับค่าจะปรับโดย R2 (ค่าเริ่มต้น).

ฟังก์ชัน `tune_model()` จะทำการสุ่มค่าไฮเปอร์พารามิเตอร์เพื่อหาค่าที่ดีที่สุดโดยที่ `n_iter` คือจำนวนครั้งในการสุ่มค่าไฮเปอร์พารามิเตอร์ ยิ่งมีค่าเยอะก็ยิ่งดี แต่ก็ยิ่งใช้เวลานานขึ้นเช่นกัน และผลลัพธ์ที่ได้จากขั้นตอนนี้ คือโมเดลที่มีค่าไฮเปอร์พารามิเตอร์ที่เหมาะสม ที่ให้ค่า Metrics ที่กำหนดในพารามิเตอร์ให้มีประสิทธิภาพ (optimize) ได้ดีที่สุด ซึ่งในที่นี้กำหนดให้ Metrics เป็นค่า Accuracy

ตัวอย่างการนำแบบจำลองมาปรับค่าพารามิเตอร์ แสดงให้เห็นดังภาพที่ 4.7 โดยจะเป็นการปรับค่าแบบจำลองของ K-Neighbors Regressor

```
knn_tuned = tune_model(knn,n_iter=150)
```

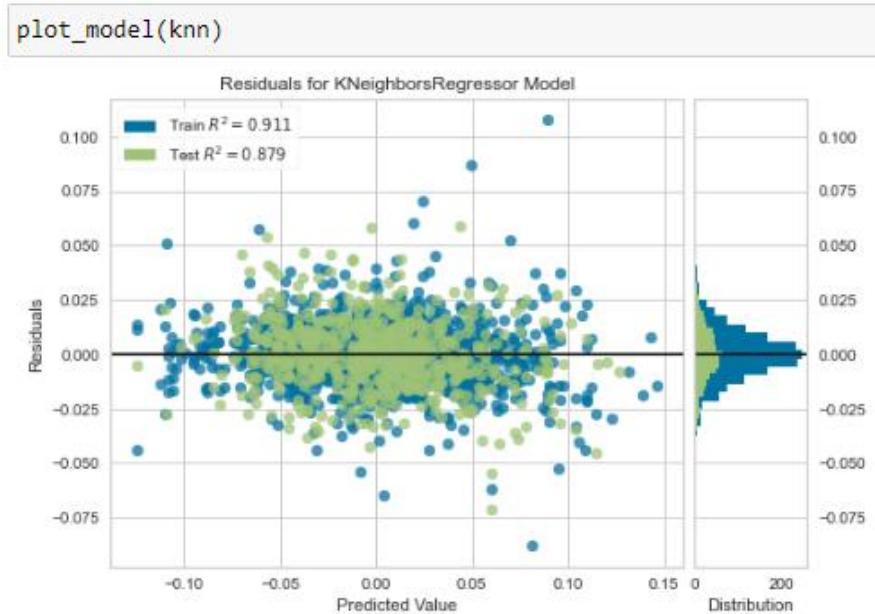
	MAE	MSE	RMSE	R2	RMSLE	MAPE
0	0.0109	0.0002	0.0151	0.8646	0.0126	0.8397
1	0.0108	0.0002	0.0146	0.8521	0.0131	-1.3066
2	0.0126	0.0004	0.0193	0.8336	0.0158	-0.1494
3	0.0109	0.0003	0.0177	0.8344	0.0132	-0.2879
4	0.0095	0.0002	0.0123	0.9180	0.0112	0.1777
5	0.0108	0.0003	0.0161	0.8700	0.0143	-0.6400
6	0.0113	0.0002	0.0155	0.8980	0.0140	0.1330
7	0.0129	0.0004	0.0200	0.8459	0.0164	0.1451
8	0.0102	0.0002	0.0143	0.8805	0.0123	1.1556
9	0.0092	0.0002	0.0130	0.9017	0.0116	-0.5973
Mean	0.0109	0.0003	0.0158	0.8699	0.0135	-0.0530
SD	0.0011	0.0001	0.0024	0.0278	0.0016	0.6827

ภาพที่4.7 ค่า Metrics หลังทำการ tune model ของแบบจำลอง K-Neighbors Regressor

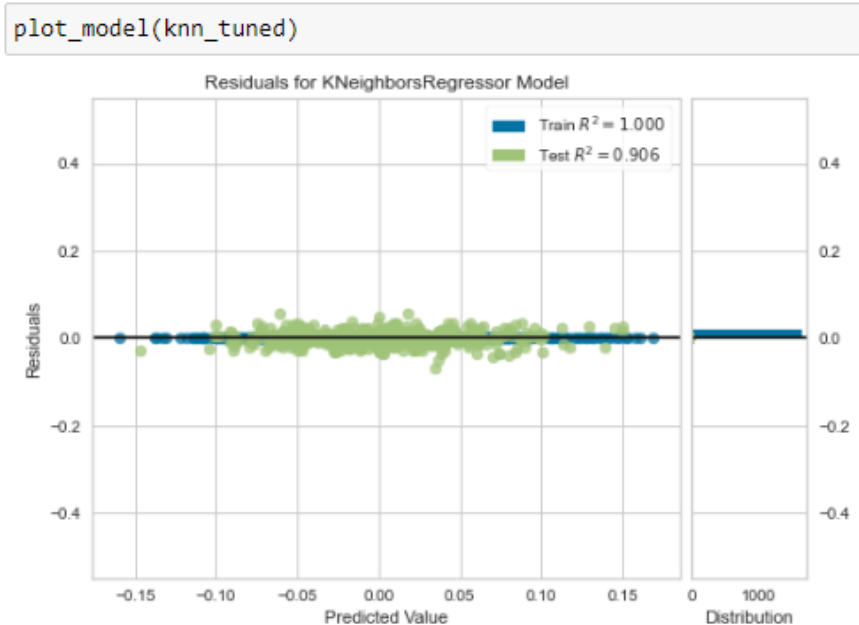
โดยกำหนดจำนวนครั้งในการสุ่มค่าไฮเปอร์พารามิเตอร์เท่ากับ 150 ครั้ง

เมื่อเปรียบเทียบการ Metrics ของแบบจำลอง [17] K-Neighbors Regressor ก่อนและหลัง การปรับค่าไฮเปอร์พารามิเตอร์ ดังภาพที่ 4.3 และ ภาพที่ 4.7 พบว่าค่า Metrics มีประสิทธิภาพ เพิ่มขึ้น ในที่นี้ค่าเริ่มต้นของจำนวนเพื่อนบ้านที่ใกล้เคียง (`n_neighbors`) ของโมเดล K-Neighbors Regressor เท่ากับ 5

ในการแสดงผลของแบบจำลองนั้น ๆ สามารถแสดงโดยใช้ฟังก์ชัน `plot_model()` [18] เพื่อให้เห็นเข้าใจภาพของแบบจำลองมากขึ้น ตัวอย่างการแสดงผลแบบจำลองดัง ภาพที่ 4.8 และภาพที่ 4.9



ภาพที่ 4.8 การแสดงผลของแบบจำลอง K-Neighbors Regressor ก่อนทำการปรับค่าไฮเปอร์พารามิเตอร์



ภาพที่ 4.9 การแสดงผลของแบบจำลอง K-Neighbors Regressor หลังทำการปรับค่าไฮเปอร์พารามิเตอร์

จากนั้นทำการสร้างและเปรียบเทียบแบบจำลองอื่น ๆ กับข้อมูลที่มีการลบค่าผิดปกติ (remove outliers) เพื่อดูว่าค่า Metrics มีการปรับเปลี่ยนค่าหรือไม่

#### 4.5 Ensemble Model [20] [21]

ไวยากรณ์ คือ

```
ensemble_model(estimator, method = 'Bagging', fold = 10, n_estimators = 10,  
round = 4, choose_better = None, optimize = 'R2', verbose = True)
```

การรวมแบบจำลองเป็นการนำหลายๆแบบจำลองมารวมกันเพื่อที่จะทำให้ผลลัพธ์ออกมาดีที่สุด โดยการทำให้ Ensembling หลักๆมี2วิธี ได้แก่

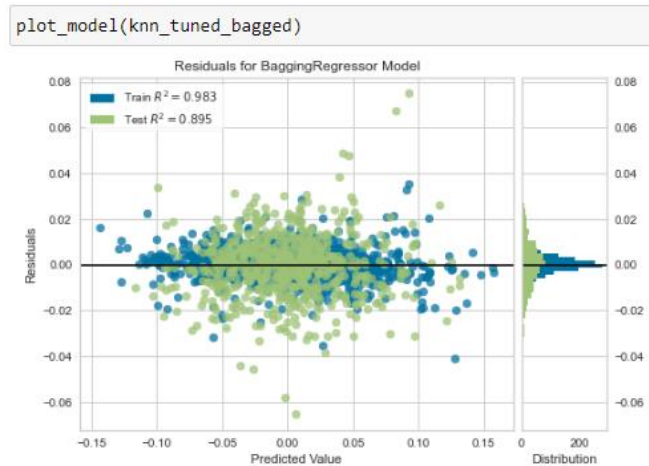
- 1) Bootstrap Aggregation (Bagging) เป็นการพยายามลดโอกาสการเกิด Overfitting ของแบบจำลองที่มีความซับซ้อนสูง
- 2) Boosting พยายามที่จะพัฒนาความสามารถในการทำนายให้มีความยืดหยุ่นมากยิ่งขึ้น

ฟังก์ชันนี้จะรวมเครื่องมือประมาณค่าพื้นฐานที่ได้รับการฝึกฝนโดยใช้วิธีการที่กำหนดไว้ในพารามิเตอร์ "method" (ค่าเริ่มต้น = "Bagging") และปรับแต่งโดย R2 ผลลัพธ์จะพิมพ์ตารางคะแนนที่แสดง MAE, MSE, RMSE, R2, RMSLE และ MAPE โดยการ fold (ค่าเริ่มต้น K-fold CV = 10 Folds). โมเดลต้องถูกสร้างโดยใช้ create\_model() หรือ tune\_model() ก่อนฟังก์ชันนี้จะส่งคืนอ็อบเจกต์โมเดลที่ได้รับการฝึกฝนแล้ว. ตัวอย่างการรวมแบบจำลองของ K-Neighbors Regressor tuned แสดงดังภาพที่4.10

```
knn_tuned_bagged = ensemble_model(knn_tuned, method='Bagging')
```

	MAE	MSE	RMSE	R2	RMSLE	MAPE
0	0.0101	0.0002	0.0127	0.9102	0.0117	0.0555
1	0.0097	0.0002	0.0127	0.9212	0.0109	-0.2502
2	0.0113	0.0003	0.0159	0.8765	0.0133	1.0827
3	0.0103	0.0002	0.0142	0.8995	0.0129	-0.4973
4	0.0107	0.0002	0.0142	0.8828	0.0127	0.0337
5	0.0120	0.0003	0.0165	0.8834	0.0143	1.3027
6	0.0109	0.0002	0.0148	0.8557	0.0131	0.4448
7	0.0104	0.0003	0.0161	0.8763	0.0138	-0.4299
8	0.0111	0.0002	0.0146	0.8949	0.0137	0.3271
9	0.0113	0.0003	0.0170	0.8370	0.0141	1.0914
Mean	0.0108	0.0002	0.0149	0.8838	0.0131	0.3161
SD	0.0006	0.0000	0.0014	0.0236	0.0010	0.6217

ภาพที่ 4.10 การรวมแบบจำลอง K-Neighbors Regressor tuned



ภาพที่ 4.11 การแสดงผลการรวมแบบจำลอง K-Neighbors Regressor tuned

#### 4.6 Blend Models [19]

ไวยากรณ์(Syntax) :

```
blend_models(estimator_list = 'All', fold = 10, round = 4, choose_better =
False, optimize = 'R2', turbo = True, verbose = True)
```

ฟังก์ชันนี้จะเป็นวิธีผสมผสานแบบจำลองที่ผ่านการฝึกอบรมซึ่งใช้ความเห็นพ้องกัน (การโหวตเสียงข้างมาก) ระหว่าง estimators เพื่อสร้างการคาดการณ์ขั้นสุดท้าย แนวคิดคือการรวมอัลกอริทึมที่แตกต่างกันและใช้การโหวตเสียงข้างมาก โดยใช้ 'R2' ในการปรับปรุงค่า ฟังก์ชันนี้จะส่งคืนตารางที่มีคะแนนเมตริก (Metrics) ที่ผ่านการตรวจสอบแล้ว k-fold พร้อมกับอ็อบเจกต์แบบจำลองที่ได้รับการฝึกฝน ใช้ K-fold Cross Validation. ผลลัพธ์จะพิมพ์ตารางคะแนนที่แสดง MAE, MSE, RMSE, R2, RMSLE และ MAPE โดย fold เท่ากับ 10 (ค่าเริ่มต้น) ฟังก์ชันนี้จะส่งคืนอ็อบเจกต์โมเดลที่ได้รับการฝึกฝน ตัวอย่างการ Blend Model โดยใช้แบบจำลอง knn\_tuned ดังภาพที่ 4.7 และ แบบจำลอง et ดังภาพที่ 4.4



```
blend_knn_et = blend_models(estimator_list=[knn_tuned,et])
```

	MAE	MSE	RMSE	R2	RMSLE	MAPE
0	0.0094	0.0001	0.0122	0.9170	0.0115	-0.0586
1	0.0098	0.0002	0.0129	0.9187	0.0116	-0.2564
2	0.0103	0.0002	0.0147	0.8943	0.0132	0.5092
3	0.0099	0.0002	0.0144	0.8966	0.0133	-0.4267
4	0.0098	0.0002	0.0130	0.9006	0.0119	-0.1062
5	0.0110	0.0002	0.0156	0.8956	0.0138	1.2262
6	0.0097	0.0002	0.0130	0.8886	0.0119	0.2616
7	0.0098	0.0002	0.0139	0.9078	0.0121	-0.4004
8	0.0102	0.0002	0.0138	0.9064	0.0128	0.2728
9	0.0097	0.0002	0.0137	0.8945	0.0120	-0.2057
Mean	0.0100	0.0002	0.0137	0.9020	0.0124	0.0816
SD	0.0004	0.0000	0.0009	0.0096	0.0007	0.4800

ภาพที่ 4.12 การ blending แบบจำลอง knn\_tuned และ แบบจำลอง et

#### 4.7 Stack Models [22]

ไวยากรณ์ คือ

```
stack_models(estimator_list, meta_model = None, fold = 10, round = 4, restack
= True, plot = False, choose_better = False, optimize = 'R2', finalize = False,
verbose = True)
```

การซ้อนโมเดลเป็นวิธีการรวมกลุ่มโมเดล เพื่อให้ผลลัพธ์ออกมาดีที่สุด แนวคิดคือการสร้างโมเดลที่สร้างการคาดการณ์ขั้นสุดท้ายโดยใช้การคาดคะเนของตัวประมาณค่าฐานหลายตัว (multiple estimators) โดยใช้ 'R2' ในการปรับปรุงค่า. การวางซ้อนโมเดลใน PyCaret ทำได้ง่ายเหมือนกับการเขียน stack\_models ฟังก์ชันนี้รับรายการโมเดลที่ได้รับการฝึกฝนโดยใช้พารามิเตอร์ Estimator\_list โมเดลทั้งหมดเหล่านี้สร้างเลเยอร์พื้นฐานของการเรียงซ้อนและการคาดคะเนถูกใช้เป็นอินพุต (Input) สำหรับโมเดลที่สามารถส่งผ่านได้โดยใช้พารามิเตอร์ meta\_model ฟังก์ชันนี้จะส่งคืนตารางที่มีคะแนนที่ผ่าน k-fold cross validation แล้ว พร้อมกับวัตถุแบบจำลองที่ผ่านการฝึก (trained model object) เพื่อใช้ในการทำนาย

```
stack1 = create_stacknet(estimator_list=[[catb,knn_tuned],[et,blend_knn_et]],restack=True)
```

	MAE	MSE	RMSE	R2	RMSLE	MAPE
0	0.0088	0.0001	0.0116	0.9260	0.0103	0.1129
1	0.0096	0.0002	0.0129	0.9191	0.0116	-0.8774
2	0.0112	0.0003	0.0158	0.8772	0.0139	0.5321
3	0.0107	0.0002	0.0151	0.8856	0.0140	-0.7142
4	0.0102	0.0002	0.0139	0.8878	0.0126	-0.2094
5	0.0106	0.0002	0.0146	0.9094	0.0133	1.1723
6	0.0100	0.0002	0.0132	0.8866	0.0118	0.4704
7	0.0101	0.0002	0.0143	0.9015	0.0126	-0.2826
8	0.0104	0.0002	0.0143	0.8993	0.0128	0.3522
9	0.0093	0.0002	0.0126	0.9096	0.0112	-0.5295
Mean	0.0101	0.0002	0.0138	0.9002	0.0124	0.0027
SD	0.0007	0.0000	0.0012	0.0151	0.0011	0.6082

ภาพที่ 4.13 การซ้อนแบบจำลองที่1

```
stack2 = create_stacknet(estimator_list=[[catb,et,knn_tuned],[blend_knn_et]], restack=True)
```

	MAE	MSE	RMSE	R2	RMSLE	MAPE
0	0.0089	0.0001	0.0118	0.9234	0.0109	-0.0306
1	0.0091	0.0001	0.0120	0.9294	0.0109	-0.6902
2	0.0100	0.0002	0.0140	0.9038	0.0128	0.4045
3	0.0102	0.0002	0.0143	0.8978	0.0133	-0.7235
4	0.0098	0.0002	0.0133	0.8967	0.0122	-0.2537
5	0.0105	0.0002	0.0147	0.9074	0.0132	1.0128
6	0.0092	0.0001	0.0121	0.9047	0.0109	0.3288
7	0.0093	0.0002	0.0135	0.9128	0.0121	-0.4048
8	0.0094	0.0002	0.0133	0.9122	0.0122	0.2921
9	0.0092	0.0001	0.0120	0.9179	0.0107	-0.6426
Mean	0.0096	0.0002	0.0131	0.9106	0.0119	-0.0707
SD	0.0005	0.0000	0.0010	0.0101	0.0010	0.5454

ภาพที่ 4.14 การซ้อนแบบจำลองที่2

```
stack3 = create_stacknet(estimator_list=[[catb,et,knn_tuned],[blend_knn_et]], restack=True.
```

	MAE	MSE	RMSE	R2	RMSLE	MAPE
0	0.0087	0.0001	0.0112	0.9309	0.0105	-0.0908
1	0.0090	0.0001	0.0121	0.9280	0.0105	-0.4437
2	0.0105	0.0002	0.0150	0.8902	0.0132	0.7666
3	0.0098	0.0002	0.0148	0.8902	0.0138	-0.4581
4	0.0098	0.0002	0.0135	0.8933	0.0122	-0.1131
5	0.0098	0.0002	0.0134	0.9231	0.0119	1.3410
6	0.0095	0.0002	0.0131	0.8872	0.0118	0.3404
7	0.0100	0.0002	0.0143	0.9015	0.0125	-0.4329
8	0.0096	0.0002	0.0127	0.9198	0.0119	0.2878
9	0.0093	0.0002	0.0132	0.9011	0.0116	-0.3318
Mean	0.0096	0.0002	0.0133	0.9065	0.0120	0.0865
SD	0.0005	0.0000	0.0011	0.0163	0.0010	0.5688

ภาพที่ 4.15 การซ้อนแบบจำลองที่3

เมื่อเปรียบเทียบ Stack model ที่ 1, 2 และ 3 จะเห็นว่า stack model ที่ 2 มีค่า Metrics ที่เหมาะสมที่สุด จากนั้นบันทึก model ไว้ใช้ในการทำนายต่อไป

## บทที่ 5

### สรุปผลงานวิจัย

หลังจากการบันทึกแบบจำลองที่ใช้ทำนาย ต่อไปเป็นการเรียกใช้แบบจำลอง เพื่อใช้ทำนายราคาทองคำล่วงหน้า ในงานวิจัยนี้ ผู้วิจัยได้ทำการทำนายล่วงหน้าแบบ 1 วัน 14วันและ22วัน แสดงดังภาพที่ 5.1, ภาพที่5.2 และ ภาพที่ 5.3

#### 5.1 ผลการทำนายราคาทองคำ

	Date	Gold	Return_1	Gold-T+1	Date-T+1
0	2019-12-17	1474.6	-0.0004	1474.0	2019-12-18
1	2019-12-18	1472.6	-0.0024	1469.1	2019-12-19
2	2019-12-19	1478.2	-0.0002	1477.9	2019-12-20
3	2019-12-20	1474.7	-0.0034	1469.7	2019-12-21
4	2019-12-23	1482.5	-0.0008	1481.3	2019-12-24
5	2019-12-24	1482.5	-0.0025	1478.8	2019-12-25
6	2019-12-25	1482.5	-0.0083	1470.2	2019-12-26

ภาพที่5.1 ผลการทำนายราคาทองคำในอีก 1 วัน

	Date	Gold	Return_14	Gold-T+14	Date-T+14
0	2019-12-17	1474.6	-0.0639	1380.4	2019-12-31
1	2019-12-18	1472.6	-0.0589	1385.9	2020-01-01
2	2019-12-19	1478.2	-0.0598	1389.8	2020-01-02
3	2019-12-20	1474.7	-0.0518	1398.3	2020-01-03
4	2019-12-23	1482.5	-0.0508	1407.2	2020-01-06
5	2019-12-24	1482.5	-0.0482	1411.0	2020-01-07
6	2019-12-25	1482.5	-0.0440	1417.3	2020-01-08

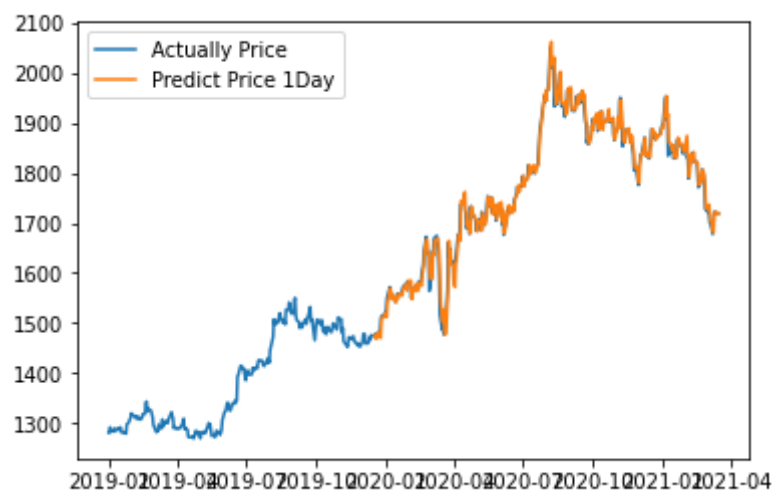
ภาพที่5.2 ผลการทำนายราคาทองคำในอีก 14 วัน

	Date	Gold	Return_22	Gold-T+22	Date-T+22
0	2019-12-17	1474.6	-0.0622	1382.9	2020-01-08
1	2019-12-18	1472.6	-0.0573	1388.2	2020-01-09
2	2019-12-19	1478.2	-0.0592	1390.7	2020-01-10
3	2019-12-20	1474.7	-0.0572	1390.3	2020-01-11
4	2019-12-23	1482.5	-0.0571	1397.8	2020-01-14
5	2019-12-24	1482.5	-0.0594	1394.4	2020-01-15
6	2019-12-25	1482.5	-0.0657	1385.1	2020-01-16

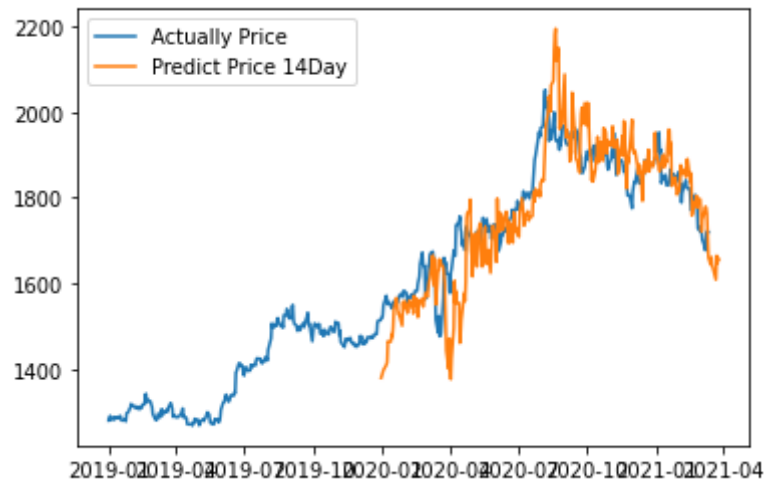
ภาพที่5.3 ผลการทำนายราคาทองคำในอีก 22 วัน

## 5.2 กราฟเปรียบเทียบผลการทำนายราคาทองคำกับราคาทองคำจริง

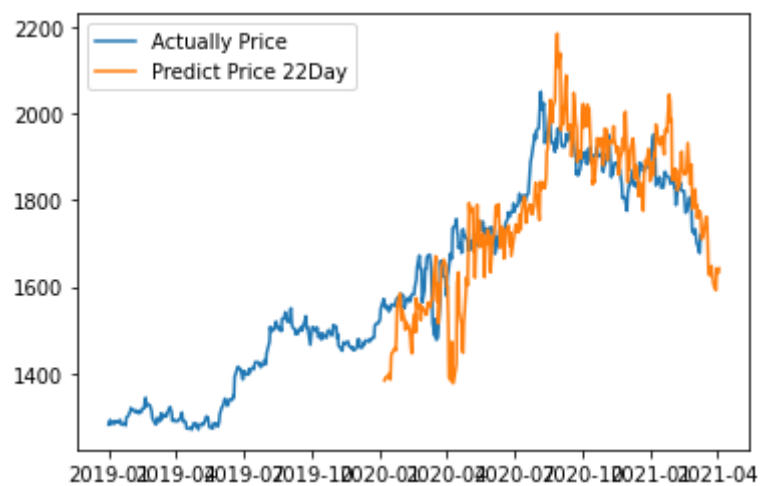
เมื่อได้ผลการทดลองแล้ว จากนั้นทำการทดสอบแบบจำลองโดยการเทียบกราฟระหว่างค่าทำนายกับค่าจริงของราคาทองคำแต่ละแบบจำลอง



ภาพที่5.4 กราฟเปรียบเทียบราคาทองคำจริงกับราคาทองคำที่ทำนาย อีก1วัน



ภาพที่5.5 กราฟเปรียบเทียบราคาทองคำจริงกับราคาทองคำที่ทำนาย อีก14วัน



ภาพที่5.6 กราฟเปรียบเทียบราคาทองคำจริงกับราคาทองคำที่ทำนาย อีก22วัน

จากการเทียบกราฟแบบจำลองพบว่ากราฟของการทำนายล่วงหน้า 1 วันทับซ้อนกับกราฟราคาทองคำจริง มากกว่ากราฟการทำนายล่วงหน้าแบบ 14 วันและแบบ 22 วัน

## เอกสารอ้างอิง

- [1] Yahoo-finance 1.4.0. (ออนไลน์) . <https://pypi.org/project/yahoo-finance/>, 28 มีนาคม 2564
- [2] สมาคมค้าทองคำ GOLD TRADERS ASSOCIATION. (ออนไลน์) . <https://www.goldtraders.or.th/>, 28 มีนาคม 2564
- [3] ราคาทองคำวันนี้. (ออนไลน์) . <https://ทองคำราคา.com/สูตรคำนวณราคาทองคำในประเทศไทย-ตามที่สมาคมค้าทองคำประกาศ/>, 28 มีนาคม 2564
- [4] มาทำความรู้จัก Machine Learning เบื้องต้น. (ออนไลน์) . <https://medium.com/@natthawatphongchit/machine-learning-basics-2b38700cb10b/>, 28 มีนาคม 2564
- [5] การวิเคราะห์การถดถอยเชิงเส้นอย่างง่าย. (pdf) . <http://www.watpon.in.th/regression/chap2.pdf/>, 28 มีนาคม 2564
- [6] สมร เหล็กกล้าและจारी ทองคำ. (2561) . *การพยากรณ์แนวโน้มอัตราแลกเปลี่ยนเงินตราต่างประเทศโดยใช้ข้อมูลเวลา* (วารสารวิชาการการจัดการเทคโนโลยีสารสนเทศและนวัตกรรม) . มหาสารคาม : คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏมหาสารคาม
- [7] วิวรรณ กาญจนวจี. (2561) . *การพยากรณ์ราคาทองคำด้วยการวิเคราะห์องค์ประกอบหลักและตัวแบบการถดถอยเชิงเส้นแบบพหุคูณ* (รายงานผลการวิจัย) . นครราชสีมา : หลักสูตรวิทยาศาสตร์บัณฑิต สาขาวิชาสถิติประยุกต์คณะวิทยาศาสตร์และเทคโนโลยีมหาวิทยาลัยราชภัฏนครราชสีมา
- [8] สุภาวดี ศิริวัฒน์และนพจักร ทองเรือนดี. *ปัจจัยที่มีผลต่อราคาทองคำในตลาดโลก* (รายงานผลการวิจัย) . กรุงเทพฯ : สาขาวิชาการบัญชี มหาวิทยาลัยนอร์ทกรุงเทพ

- [9] นริสา สมุทรสาคร. (2547) . *การพยากรณ์ราคาทองคำโดยวิธีอาร์มีนา* (รายงานผลการวิจัย) .  
เชียงใหม่ : บัณฑิตวิทยาลัย มหาวิทยาลัยเชียงใหม่
- [10] Metrics พื้นฐานสำหรับวัดประสิทธิภาพของโมเดล Machine Learning. (ออนไลน์) .  
<https://medium.com/@615162020027/metrics-พื้นฐานสำหรับวัดประสิทธิภาพของโมเดล-machine-learning-c00fcc32fa30/>, 28 มีนาคม 2564
- [11] อธิบาย 10 Metrics พื้นฐานสำหรับวัดผลโมเดล Machine Learning. (ออนไลน์) .  
<https://datarockie.com/2019/03/30/top-ten-machine-learning-metrics/>, 28 มีนาคม 2564
- [12] Pycaret — ผู้ช่วยยุคใหม่ที่จะทำให้การเลือกใช้ Machine Learning ง่ายขึ้น. (ออนไลน์) .  
<https://medium.com/c-g-datacommunity/pycaret-ผู้ช่วยยุคใหม่ที่จะทำให้การเลือกใช้-machine-learning-ง่ายขึ้น-462ddf4364b/>, 28 มีนาคม 2564
- [13] Pycaret Regression. (ออนไลน์). <https://pycaret.org/regression/>, 28 มีนาคม 2564
- [14] Pycaret ผู้ช่วยยุคใหม่ที่จะทำให้การเลือกใช้ Machine Learning ง่ายขึ้น. (ออนไลน์).  
<https://medium.com/c-g-datacommunity/pycaret-ผู้ช่วยยุคใหม่ที่จะทำให้การเลือกใช้-machine-learning-ง่ายขึ้น-462ddf4364b/>, 28 มีนาคม 2564
- [15] Pycaret, Create Model. (ออนไลน์) . <https://pycaret.org/create-model/>, 28 มีนาคม 2564
- [16] Pycaret, Tune Model. (ออนไลน์) . <https://pycaret.org/tune-model/>, 28 มีนาคม 2564
- [17] Pycaret, Compare Models. (ออนไลน์) . <https://pycaret.org/compare-models/>, 28 มีนาคม 2564
- [18] Pycaret, Plot Model. (ออนไลน์) . <https://pycaret.org/plot-model/>, 28 มีนาคม 2564
- [19] Pycaret, Blend Models. (ออนไลน์) . <https://pycaret.org/blend-model/>, 28 มีนาคม 2564



- [20] Pycaret, Ensemble Model.(ออนไลน์). <https://pycaret.org/ensemble-model/>, 28 มีนาคม 2564
- [21] Botnoi Classroom — มาทำให้ผลลัพธ์ดีขึ้นด้วย Ensemble Method กันเถอะ!. (ออนไลน์).<https://medium.com/botnoi-classroom/botnoi-classroom-มาทำให้ผลลัพธ์ดีขึ้นด้วย-ensemble-method-กันเถอะ-b5ac9acfa8d3/>, 28 มีนาคม 2564
- [22] Pycaret, Stack Models.(ออนไลน์). <https://pycaret.org/stack-models/>, 28 มีนาคม 2564