# Forecasting Gold Price Time Series

# การทำนายอนุกรมเวลาของราคาทองคำ


นางสาวกรกนก คมวิภาค

Kornkanok Komvipark

นายกิจปมาน อิงคภาวัฏร

Kijpmarn Engkapawastr


โครงงานนี้เปนสวนหนึ่งของการศึกษาหลักสูตรวิทยาศาสตรบัณฑิต

ภาควิชาวิทยาการคอมพิวเตอร์

คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ ปีการศึกษา พ.ศ. 2562

**คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ**

| | |
|---|---|
| **ชื่อหัวข้อโครงงาน** | Forecasting Gold Price Time Series |
| | การทำนายอนุกรมเวลาของราคาทองคำ |
| **นิสิต** | นางสาวกรกนก คมวิภาค 59102010896 |
| | นายกิจปมาน อิงคภาวัฏร 57107010025 |
| **ปริญญา** | วิทยาศาสตร์บัณฑิต (วท.บ.) |
| **ภาควิชา** | วิทยาการคอมพิวเตอร์ |
| **อาจารย์ที่ปรึกษาโครงงาน** | อ.ดร.ศิริสรรพ เหล่าหะเกียรติ |

ลงชื่อ.....

(อ.ดร.ศิริสรรพ เหล่าหะเกียรติ)

อาจารย์ที่ปรึกษาโครงงาน

# บทคัดย่อ

ทองคำเป็นโลหะมีค่า ที่มีบทบาทในทางเศรษฐกิจเป็นอย่างสูง ราคาทองคำเป็นตัวบ่งชี้กิจกรรมทางเศรษฐกิจที่สำคัญอย่างหนึ่ง หากเราสามารถประเมินทิศทางการเปลี่ยนแปลงของราคาทองในอนาคตได้ก็จะเป็นประโยชน์ต่อการศึกษาและทราบถึงสภาวะของเศรษฐกิจ ณ ขณะนั้นได้ บทความนี้มุ่งศึกษาเทคนิคการพยากรณ์ราคาทองคำ โดยใช้แบบจำลองสำหรับข้อมูลอนุกรมเวลา โดยจะนำปัจจัยทางเศรษฐกิจอื่นๆที่เกี่ยวข้องมาใช้เป็นอินพุตช่วยในการทำนายด้วย เช่น ราคาแพลทตินัม ราคาน้ำมันดิบ ราคาทองแดง ราคาก๊าซธรรมชาติ ก๊าซโซลีน ใช้เทคนิคในการวิเคราะห์ข้อมูลและเลือกเฟ้นปัจจัยอื่นมาช่วยเพิ่มประสิทธิภาพในการพยากรณ์ของแต่ละแบบจำลอง ทั้งแบบจำลองเชิงเส้น ได้แก่ แบบจำลอง ARIMA,HWS,SES,VAR และแบบจำลองเครือข่ายประสาทเทียมแบบ LSTM และ CNN เพื่อพยากรณ์ราคาทองคำ ซึ่งการทดสอบความน่าเชื่อถือของราคาที่พยากรณ์ออกมาสามารถทำได้ด้วยการนำราคาทองที่พยากรณ์ มาเปรียบเทียบกับราคาทองคำล่าสุด เพื่อเลือกหาแบบจำลองที่มีประสิทธิภาพสามารถนำมาใช้ในการพยากรณ์เบื้องต้นได้

# Abstract

Gold is the precious metal. It has an important role in economy. The price of gold is one of indicators of economy. If we can predict the price, it will benefit for studying and prediction economic state. This study focuses on Gold prediction by using model for Time series data by using another economic indicators to predict the price such as platinum, oil ,bronze ,natural gas ,gasoline by using data analysis in other to find another indicator for increase the performances of each model. Both linear model such as ARIMA,HWS,SES,VAR model and neural network LSTM,CNN. for predict gold price and test accuracy by compare predicted gold price with the latest gold price to find out the best preliminary forecast model.

# กิตติกรรมประกาศ

ในการศึกษาและเรียบเรียงโครงงานฉบับนี้สำเร็จลุล่วงไปด้วยดี ผู้วิจัยต้องขอขอบพระคุณ อ.ดร.ศิริสรรพ เหล่าหะเกียรติ ซึ่งเป็นอาจารย์ที่ปรึกษาของงานวิจัยนี้ที่ให้คำแนะนำปรึกษา ตรวจทาน ตลอดจนแก้ไขข้อผิดพลาดต่างๆ

ขอขอบพระคุณอาจารย์ประจำภาควิชาวิทยาการคอมพิวเตอร์ทุกท่าน ที่ได้ช่วยอบรมสั่งสอน และให้ความรู้ ให้ความสนับสนุนในการดำเนินโครงงานในครั้งนี้จนสามารถจัดทำโครงงานได้ดีและมีประสิทธิภาพมากยิ่งขึ้น

คณะผู้จัดทำหวังว่าโครงงาน การทำนายอนุกรมเวลาของราคาทองคำ (Forecasting Gold Price Time Series) จะเป็นประโยชน์ต่อการศึกษา เพื่อที่จะสามารถนำไปพัฒนาต่อยอดในอนาคตได้ดียิ่งๆ ขึ้นไป

# สารบัญ

**หัวข้อ** **หน้า**

**บทคัดย่อ** ..... **ข**

**Abstract** ..... **ค**

**กิตติกรรมประกาศ** ..... **ง**

**สารบัญ** ..... **จ**

**สารบัญรูปภาพ** ..... **ซ**

**สารบัญตาราง** ..... **ญ**

**บทที่ 1** ..... **1**

**บทนำ** ..... **1**

1.1 ที่มาและความสำคัญของโครงงาน ..... 1

1.2 วัตถุประสงค์ของโครงงาน ..... 1

1.3 ขอบเขตของโครงงาน ..... 2

1.4 ประโยชน์ที่คาดว่าจะได้รับ ..... 2

**บทที่ 2** ..... **3**

2.1 อนุกรมเวลา (Time series) ..... 3

2.2 ความนิ่งของข้อมูล (Stationary) ..... 8

2.3 การทดสอบความสัมพันธ์ (Correlation) ..... 10

2.5 วิธีการปรับให้เรียบแบบเอกซ์โพเนนเชียล Single Exponential Smoothing (SES) ..... 12

2.6 วิธีการปรับให้เรียบเอกซ์โพเนนเชียลแบบโฮลต์-วินเทอร์ (Holt-Winters exponential smoothing method) ..... 13

2.7 แบบจำลอง ARIMA (Autoregressive Intergrated Moving average) ..... 14

2.8 แบบจำลอง CNN (Convolutional Neural Network) ..... 15

2.9 แบบจำลอง LSTM (Long Short-Term Memory) ..... 19
2.10 การวัดประสิทธิภาพ (Performance Evaluation)..... 22
2.11 การแบ่ง Cross Validation แบบ Timeseriessplit..... 23
2.12 งานวิจัยที่เกี่ยวข้อง ..... 23
**บทที่ 3.....25**
3.1 ขั้นตอนการดำเนินงาน ..... 25
3.2 แผนการดำเนินงานตลอดโครงงาน ..... 26
3.3 อุปกรณ์และเครื่องมือที่ใช้..... 28
3.4 ชุดข้อมูล (Dataset)..... 28
3.5 ขั้นตอนการออกแบบระบบ..... 29
**บทที่ 4.....42**
4.1 ผลการดำเนินงานของแบบจำลอง ARIMA..... 42
4.2 ผลการดำเนินงานของแบบจำลอง SES ..... 43
4.3 ผลการดำเนินงานของแบบจำลอง HWS..... 44
4.5 ผลการดำเนินงานของแบบจำลอง CNN Allinput ..... 45
4.6 ผลการดำเนินงานของแบบจำลอง LSTM 1 input..... 46
4.7 ผลการดำเนินงานของแบบจำลอง LSTM Allinput..... 47
4.8 ผลการดำเนินงานของแบบจำลอง VAR Allinput..... 47
**บทที่ 5.....50**
5.1 สรุปผลการดำเนินการโครงการ..... 50
5.2 ปัญหาและอุปสรรค ..... 50
5.3 ข้อเสนอแนะและแนวทางในการพัฒนา..... 50

**บรรณานุกรม.....51**

**ภาคผนวก.....54**

# สารบัญรูปภาพ

รูป 1 แสดงอนุกรมเวลาของราคาทองคำตั้งแต่ปี ค.ศ.1970 - 2020....3
รูป 2 กราฟอนุกรมเวลาที่แสดงลักษณะของส่วนประกอบแนวโน้ม....5
รูป 3 รูปแสดงอนุกรมเวลาที่มีส่วนของวัฏจักร....6
รูป 4 แสดงดัชนีผลผลิตสินค้าเกษตรของประเทศหนึ่งเป็นรายเดือนตั้งแต่ปี 2548 – 2554....7
รูป 5 กราฟอนุกรมเวลาที่แสดงลักษณะของความผันผวนจากเหตุการณ์ไม่ปกติ....8
รูป 6 ตัวอย่างกราฟอนุกรมเวลาการใช้ไฟฟ้ารายเดือนของสหรัฐอเมริกา....9
รูป 7 ข้อมูลอนุกรมเวลาที่ทำการสลายตัวอนุกรมเวลา....11
รูป 8 การถ่วงน้ำหนักของวิธีการปรับให้เรียบแบบเอกซ์โพเนนเชียล....13
รูป 9 แสดงขั้นตอนการทำคอนโวลูชันระหว่างภาพนำเข้า....16
รูป 10 เปรียบเทียบการทำคอนโวลูชันแบบแคบและแบบกว้าง....16
รูป 11 การทำคอนโวลูชันโดยมีข้อมูลรับเข้าขนาด 4x4 ตัวกรองขนาด 2x2....17
รูป 12 การทำคอนโวลูชันโดยมีจำนวนตัวกรองเท่ากับ 3....17
รูป 13 แสดงการพูลลิงแบบหาค่าเฉลี่ยและหาค่ามากที่สุด....18
รูป 14 โครงสร้าง Forget หรือ Forget Gate....19
รูป 15 โครงสร้าง Write หรือ Input Gate....20
รูป 16 โครงสร้าง Update Cell State....21
รูป 17 โครงสร้าง Read หรือ Output Gate....21
รูป 18 ภาพรวมโครงสร้าง LSTM ทั้งหมด....22
รูป 19 การแบ่ง Cross Validation แบบ Timeseriessplit....23
รูป 20 ตัวอย่างข้อมูลราคาทองคำ....29
รูป 21 แสดงการทำงานของระบบ....29
รูป 22 ตรวจสอบความนิ่งของข้อมูลด้วย Dickey Fuller....31
รูป 23 ข้อมูลราคาทองหลังจากทำการ Differencing....31
รูป 24 การแบ่งย่อยข้อมูลอนุกรมเวลา....32

รูป 25 ฤดูกาลของราคาทองคำ ..... 32
รูป 26 ความสัมพันธ์ของราคาทองกับราคาอื่นๆ ..... 33
รูป 27 ใช้ auto.arima ในการหาค่าพารามิเตอร์ ..... 34
รูป 28 ตัวอย่างการปรับค่าพารามิเตอร์ของแบบจำลอง ARIMA ..... 35
รูป 29 ตัวอย่างการปรับค่าพารามิเตอร์ของแบบจำลอง SES ..... 36
รูป 30 ตัวอย่างการปรับค่าพารามิเตอร์ของแบบจำลอง SES ..... 37
รูป 31 หาค่า RMSE ของแบบจำลอง CNN 1 input ..... 45
รูป 32 หาค่า RMSE ของแบบจำลอง CNN Allinput ..... 45
รูป 33 หาค่า RMSE ของแบบจำลอง LSTM 1 input ..... 46
รูป 34 หาค่า RMSE ของแบบจำลอง LSTM Allinput ..... 47
รูป 35 ผลการทำนายราคาทองคำแสดงผล API ..... 48
รูป 36 การทำนายราคาทอง 1 สัปดาห์โดยไม่แบ่ง Cross Validation ของทุกแบบจำลอง ..... 49

# สารบัญตาราง

ตาราง 1 แผนการดำเนินโครงการวิจัย ..... 27
ตาราง 2 ตารางบันทึกผลการหาค่าพารามิเตอร์ที่ดีที่สุดของแบบจำลอง ARIMA..... 43
ตาราง 3 ตารางบันทึกผลการหาค่าพารามิเตอร์ที่ดีที่สุดของแบบจำลอง HWS..... 44
ตาราง 4 เปรียบเทียบค่าเฉลี่ย RMSE ของแต่ละแบบจำลอง..... 48

# บทที่ 1
# บทนำ

## 1.1 ที่มาและความสำคัญของโครงงาน

ปัจจุบันทองคำมีประโยชน์ในฐานะเป็นโลหะสื่อกลางแห่งการแลกเปลี่ยนเงินตรา ถูกสำรองไว้เป็นทุนสำรองเงินตราระหว่างประเทศ ถูกใช้เป็นเครื่องมือในการเก็งกำไรของตลาดการค้าเพราะทองคำมีมูลค่าในตัวเอง แต่ราคาของทองคำนั้นมีขึ้นลง มีการเปลี่ยนแปลงตลอดเวลาด้วยหลายปัจจัย หากเราสามารถประเมินทิศทางการเปลี่ยนแปลงของราคาทองในอนาคตได้ก็จะเป็นประโยชน์ต่อการศึกษาและคาดการณ์สภาวะของเศรษฐกิจ ณ ขณะนั้นได้

ข้อมูลราคาทองคำมีลักษณะเป็นข้อมูลในเชิงอนุกรมเวลา ซึ่งมีกระบวนการทาง data science สามารถเข้ามาช่วยในการพยากรณ์ราคาทองคำ และเพื่อเพิ่มความแม่นยำจากการเปลี่ยนแปลงของราคาทองคำแบบอนุกรมเวลา จะนำปัจจัยอื่นๆที่เกี่ยวข้องมาใช้เป็นอินพุตช่วยในการทำนายด้วย เช่น อัตราดอกเบี้ย ราคาน้ำมัน ค่าเงินดอลลาห์ Demand Supply เงินเฟ้อ หรือราคาของโลหะมีค่าชนิดอื่น โดยใช้เทคนิคในการวิเคราะห์ข้อมูล เพื่อเลือกเฟ้นปัจจัยอื่น มาช่วยเพิ่มประสิทธิภาพในการทำนาย และสุดท้ายคือความไม่แน่นอน เพราะทองคำนั้นถือเป็นสินทรัพย์ปลอดภัยของโลกนี้ ดังนั้นหากเกิดความไม่แน่นอนเช่น เกิดการก่อการร้าย หรือ เกิดวิกฤตทางการเงิน การเมือง ราคาทองคำก็จะสูงขึ้นนั้นเอง

ทั้งนี้การศึกษา วิเคราะห์ และทำนายข้อมูลราคาทอง จะเป็นประโยชน์ในการศึกษา และใช้เป็นเครื่องมือเบื้องต้นในการคาดการณ์สภาวะของเศรษฐกิจ โดยพิจารณาแต่เพียงการเปลี่ยนแปลงค่า ทางในเชิงการคำนวณทางสถิติเท่านั้น การทำนายข้อมูลนี้ไม่สามารถครอบคลุมถึงปัจจัยที่ซับซ้อนทางเศรษฐกิจอื่นได้ เช่น การก่อการร้ายหรือเกิดวิกฤตทางการเงิน ทำให้ค่าของการทำนายที่ได้ มิอาจมีความแม่นยำจนนำไปใช้ประโยชน์ทางเศรษฐกิจอย่างจริงจังได้

## 1.2 วัตถุประสงค์ของโครงงาน

1.2.1 ศึกษาความสัมพันธ์ระหว่างราคาทองคำกับปัจจัยทางเศรษฐกิจตัวอื่น โดยใช้เทคนิคการวิเคราะห์ข้อมูล

1.2.2 ศึกษาและเปรียบเทียบ Algorithm ของ Machine Learning เพื่อเลือกแบบจำลองที่มีประสิทธิภาพสามารถในการทำนายพยากรณ์ราคาทอง

1.2.3 เพื่อประเมินประสิทธิภาพของแบบจำลองที่ใช้ในการพยากรณ์ราคาทองคำ

## 1.3 ขอบเขตของโครงงาน

1.3.1 เลือกแบบจำลองที่มีประสิทธิภาพสามารถในการทำนายพยากรณ์ราคาทองและปรับแบบจำลอง Machine Learning เพื่อใช้กับการพยากรณ์ราคาทองคำล่วงหน้า

1.3.2 แบบจำลองที่ใช้ในการพยากรณ์ได้แก่ ARIMA,HWS,SES,VAR,CNN,LSTM

1.3.3 เลือกปัจจัยทางเศรษฐกิจที่มีผลต่อราคาทองคำ โดยใช้เทคนิคการวิเคราะห์ข้อมูล

1.3.4 ทำการตรวจสอบว่าแบบจำลองสามารถพยากรณ์ราคาทองคำได้ดีกว่าราคาทองปัจจุบันจริงหรือไม่

## 1.4 ประโยชน์ที่คาดว่าจะได้รับ

1.4.1 เป็นเครื่องมือเบื้องต้นในการพยากรณ์ทางเศรษฐกิจ

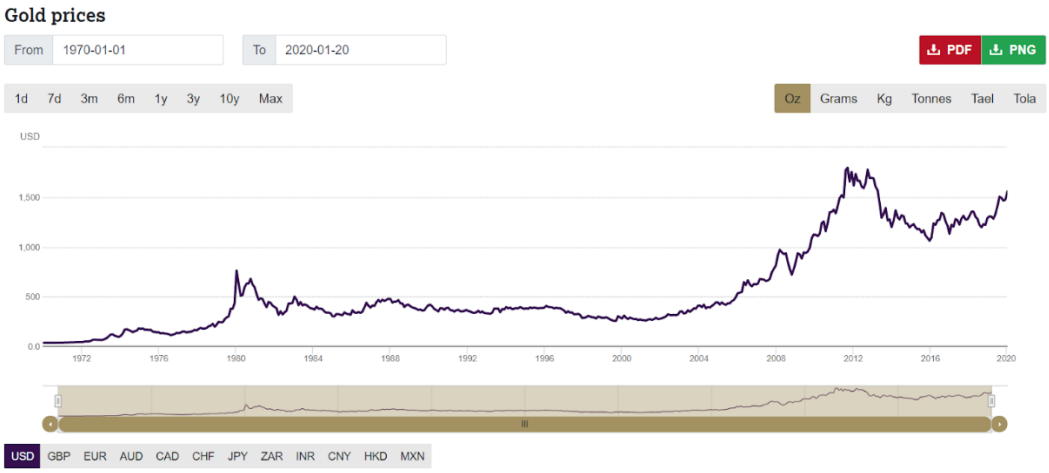
1.4.2 ผู้ที่สนใจสามารถอ่านเพิ่มเติมเพื่อประกอบการตัดสินใจในการคาดการณ์ราคาทองในอนาคตได้

1.4.3 ได้แบบจำลองที่เหมาะสมในการพยากรณ์ราคาทองคำ เพื่อเป็นแนวทางในการศึกษาแนวโน้มของราคาทองต่อไป

# บทที่ 2
# ทฤษฎีและแนวคิดที่ใช้ในการทำวิจัย

## 2.1 อนุกรมเวลา (Time series)

**2.1.1 ความหมายของอนุกรมเวลา (Time series)** หมายถึง ชุดของข้อมูลที่เก็บรวบรวมตามระยะเวลาเป็นช่วงๆอย่างต่อเนื่องกัน เช่น ข้อมูลยอดขายสินค้าที่เก็บรวบรวมต่อเนื่องกันไปเป็นระยะเวลาหลายๆเดือน ข้อมูลรายได้ ประชาชาติปีต่างๆ ที่เก็บรวบรวมต่อเนื่องกันไปเป็นระยะเวลาหลายๆปี เป็นต้น ข้อมูลอนุกรมเวลาอาจอยู่ใน ลักษณะที่เป็นข้อมูลรายปี รายไตรมาส หรือรายเดือนก็ได้ หลังจากข้อมูลของตัวแปรที่สนใจ ถูกเก็บรวบรวมมาแล้ว ไม่ว่าจะเป็นรายวัน รายสัปดาห์รายเดือน รายไตรมาส หรือรายปี ค่าทาง สถิติเบื้องต้น ได้แก่ ค่าเฉลี่ย (Mean) ความแปรปรวน (Variance) และความแปรปรวนร่วม ระหว่างช่วงเวลา (Autocovariance) จะต้องสามารถคำนวณได้เสมอ ทั้งนี้ขึ้นอยู่กับความเหมาะสมในการนำไปใช้ประโยชน์เพราะในแต่ละช่วงเวลาข้อมูลดังกล่าวสามารถเพิ่มขึ้นหรือลดลงหรือเท่าเดิมก็ได้ซึ่งไม่อาจทราบล่วงหน้าได้ เนื่องจากข้อมูลทางธุรกิจมีการเปลี่ยนแปลงตลอดเวลา ผู้นำทางธุรกิจหรือองค์กรต้องหาวิธีพัฒนาต่าง ๆ ที่สามารถนำไปใช้ประกอบการตัดสินใจวางแผนเกี่ยวกับผลที่เกิดจากความเปลี่ยนแปลงในการดำเนินการอย่างใดอย่างหนึ่ง ดังนั้นการวิเคราะห์อนุกรมเวลาจึงเข้ามามีบทบาทช่วยในการตัดสินใจ เทคนิคอย่างหนึ่งที่ใช้ช่วยในการควบคุมการดำเนินการในปัจจุบันและในการวางแผนความต้องการในอนาคต คือ การพยากรณ์(forecasting) ซึ่งการพยากรณ์นั้นทำได้หลายวิธี แต่ละวิธีต่างมีเป้าหมายเดียวกัน คือ ทำนายเหตุการณ์ในอนาคต




*รูป 1* แสดงอนุกรมเวลาของราคาทองคำตั้งแต่ปี ค.ศ.1970 - 2020

ที่มา : [ออนไลน์] https://www.gold.org/goldhub/data/gold-prices

**2.1.2 แนวคิดการการพยากรณ์อนุกรมเวลา**

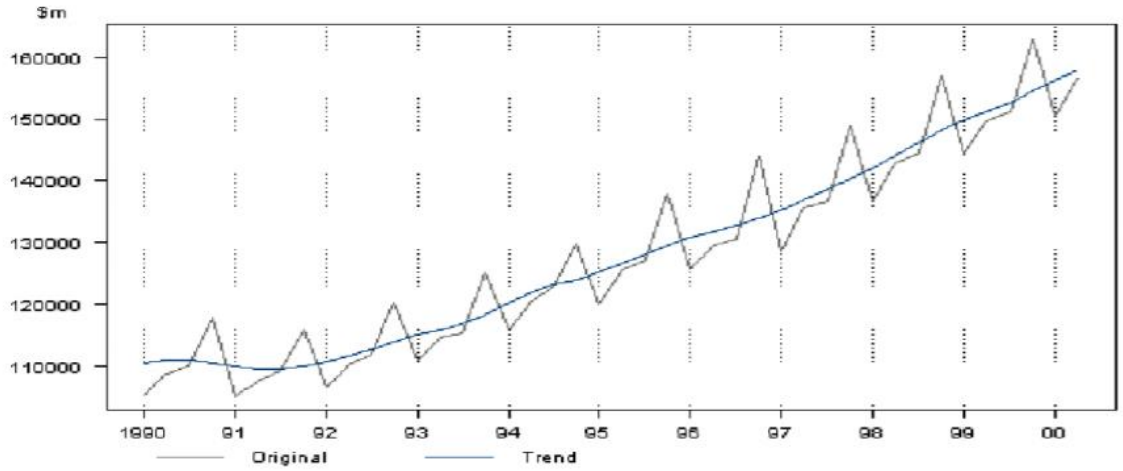
การพยากรณ์ (Forecasting) หมายถึงการคาดการณ์เหตุการณ์ในอนาคต ถ้าพิจารณาในมุมมองของนักธุรกิจมักจะมีการพยากรณ์ยอดขายสินค้าของตนเอง พยากรณ์อัตราดอกเบี้ยเงินกู้ พยากรณ์ราคาสินค้าของคู่แข่ง พยากรณ์ปริมาณการใช้วัตถุดิบ ถ้าเป็นนักการเงินจะต้องทำการพยากรณ์ราคาหุ้น พยากรณ์อัตราผลตอบแทนของหลักทรัพย์หรือถ้าเป็นนักเศรษฐศาสตร์มักต้องพยากรณ์อัตราการเจริญเติบโตทางเศรษฐกิจ พยากรณ์อัตราเงินเฟ้อ พยากรณ์อัตราการว่างงาน

ในการพยากรณ์ตัวแปรใด ๆ ก็ตามเราจะต้องใช้ข้อมูลของตัวแปรนั้นในอดีตที่ผ่านมา เช่น หากนักธุรกิจต้องการพยากรณ์ยอดขายของบริษัทตนเองในเดือนหน้า ข้อมูลที่สำคัญที่สุดที่ต้องมีก็คือยอดขายของบริษัทที่ผ่านมาในอดีต จากนั้นจะผู้บริหารจะต้องทำการวิเคราะห์ข้อมูลยอดขายในอดีตนั้นแล้วจึงนำผลการวิเคราะห์ที่ได้ไปใช้พยากรณ์ข้อมูลนั้น สาเหตุที่ต้องมีการรวบรวมข้อมูลยอดขายในอดีตเนื่องจากการวิเคราะห์ข้อมูลยอดขายในอดีตจะช่วยให้สามารถระบุถึงรูปแบบที่ค่าของตัวแปรยอดขายนั้นเป็นอยู่และการนำผลการวิเคราะห์(หรือรูปแบบที่ระบุได้) ไปใช้พยากรณ์ยอดขายของบริษัท ซึ่งจะต้องอยู่ภายใต้ข้อสมมุติ ว่า “รูปแบบที่ระบุได้จากข้อมูลยอดขายในอดีตนั้น ต้องเหมือนเดิมหรือไม่เปลี่ยนแปลงในอนาคต” จากการพยากรณ์ภายใต้ข้อสมมุติข้างต้น ทำให้เรากล่าวได้ว่าการพยากรณ์ยอดขายมี โอกาสที่เกิดความผิดพลาดได้ หากรูปแบบที่ระบุได้จากข้อมูลในอดีตไม่เหมือนเดิมหรือเปลี่ยนแปลงไปในอนาคต เช่น หลังจากเกิดแผ่นดินไหวครั้งใหญ่ หรือการจลาจลครั้งประวัติศาสตร์ เราจะใช้รูปแบบที่ระบุได้จากการวิเคราะห์ข้อมูลยอดขายบะหมี่กึ่งสำเร็จรูปในอดีต ไปใช้พยากรณ์ยอดขายบะหมี่กึงสำเร็จรูปนี้ไม่ได้ เนื่องจากพฤติกรรมผู้บริโภคเปลี่ยนไปแล้ว

**2.1.3 ส่วนประกอบของอนุกรมเวลา (Components of a Time Series)**

อนุกรมเวลาของตัวแปรหนึ่ง จะประกอบไปด้วย 4 ส่วน คือ แนวโน้ม วัฏจักร ความผัน แปรจากฤดูกาล และความผันผวนจากเหตุการณ์ไม่ปกติ ความหมายของส่วนประกอบแต่ละส่วนมี รายละเอียดดั้งนี้

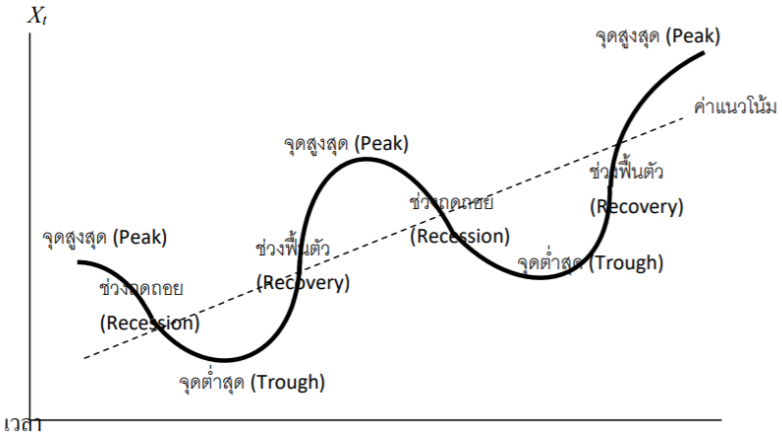
(1) แนวโน้ม (Trend) คือ ส่วนที่ทำให้อนุกรมเวลามีค่าเพิ่มขึ้นเรื่อยๆหรือลดลงเรื่อย ๆ เมื่อเวลาผ่านไป เรามักใช้แนวโน้มในการบอกว่าอนุกรมเวลาที่เก็บข้อมูลมามีอัตราการ เพิ่มขึ้นหรืออัตราการลดลงในระยะยาว




*รูป 2* กราฟอนุกรมเวลาที่แสดงลักษณะของส่วนประกอบแนวโน้ม

ที่มา : [ออนไลน์] http://www.abs.gov.au/websitedbs/OpenDocument

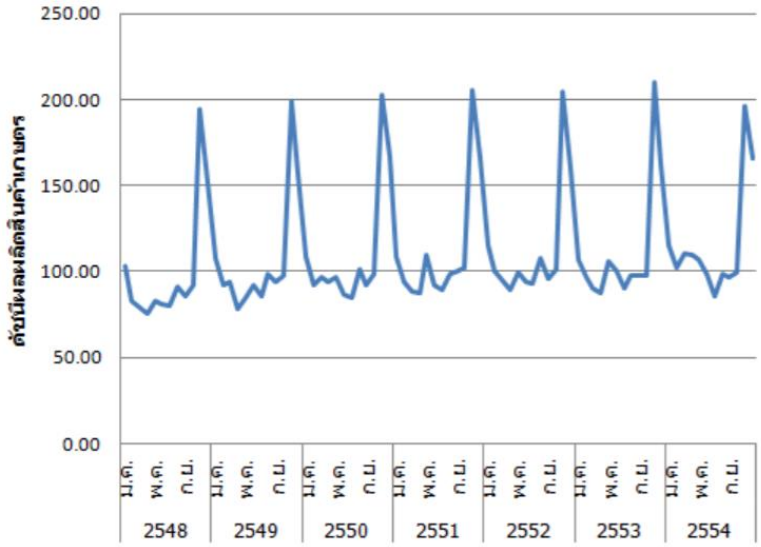
(2) วัฏจักร (Cycle) คือ ส่วนประกอบอีกส่วนประกอบหนึ่งที่มีลักษณะคล้ายกับฤดูกาลเพียงแต่ว่าส่วนประกอบฤดูกาลมีคาบหรือ ระยะเวลาสั้นกว่าส่วนประกอบวัฏจักร อย่างไรก็ตามส่วนประกอบวัฏจักรยังมีลักษณะคล้าย ๆ กับส่วนประกอบฤดูกาลคือ ส่วนประกอบวัฏจักรเป็นส่วนประกอบที่ช่วยลดความคลาดเคลื่อนอันเกิดจากการแกว่ง ตัวของข้อมูลที่มีลักษณะการเคลื่อนไหวแบบคาบเวลาตามอิทธิพลของวัฏจักรต่าง ๆ ตามธรรมชาติของข้อมูลที่ได้ทำการสำรวจส่วนมากของอนุกรมเวลาที่พบเสมอในการทำนายหรือการพยากรณ์ด้วยอนุกรมเวลา ได้แก่ วัฏจักรธุรกิจ (Business cycle) วัฏจักรเศรษฐกิจ ( Economic cycle) วัฏจักรสภาพอากาศ (Weather cycle) ซึ่งส่วนที่ทำให้อนุกรมเวลาที่เก็บข้อมูลได้มีค่าเพิ่มขึ้นและลดลงสลับกันไปรอบ ๆ ค่าแนวโน้ม (ซึ่งแสดงด้วยเส้นประดังรูปที่ 1.2) การนับระยะเวลาของส่วนวัฏจักร จะนับจุดสูงสุดหนึ่ง (peak) ไปยังอีกจุดสูงสุดหนึ่งหรือจากจุดต่ำสุดหนึ่ง (trough) ไปยังอีกจุดต่ำสุดหนึ่ง ซึ่งจะต้องกินเวลาตั้งแต่ 2 ปี ถึง 10 ปีขึ้นไป (หรือนานกว่านั้น ) ส่วนของวัฏจักรจะเริ่ม ณ เวลาใดก็ได้ ตัวอย่างของวัฏจักรแสดงได้ในรูปที่ 1.2 เมื่อส่วนของวัฏจักรอยู่ในช่วง ที่ทำให้อนุกรมเวลามีค่าลดลง เราจะเรียกว่าช่วงถดถอย (Recession) และหลังจากผ่านจุดต่ำสุดไปแล้ว ส่วนของวัฏจักรที่ทำให้อนุกรมเวลามีค่าเพิ่มขึ้น เราจะเรียกว่าช่วงฟื้นตัว (Recovery)




*รูป 3* รูปแสดงอนุกรมเวลาที่มีส่วนของวัฏจักร

ที่มา : [ออนไลน์] https://economics.utcc.ac.th/wp-content/uploads/Time-Series-for-Econ-and-Bus_Poomthan.pdf

(3) ความผันแปรจากฤดูกาล (Seasonal Variations) คือ รูปแบบในช่วงเวลาหนึ่งของอนุกรมเวลาที่จะเป็นภายใน 1 ปีและจะเป็นแบบนี้ซ้ำกันทุกปีตัวอย่าง เช่น อุณหภูมิเฉลี่ยในเดือนเมษายนจะสูงกว่าอุณหภูมิเฉลี่ยในเดือนอื่นๆ และจะเป็นเช่นนี้ซ้ำๆกันทุกปี ค่าใช้ไฟฟ้าในเดือนพฤศจิกายน–ธันวาคมจะต่ำกว่าค่าใช้ไฟฟ้าในเดือนอื่นๆและเป็นเช่นนี้ทุกปีบริษัททัวร์จะมีรายรับในช่วงปิดเทอมสูงกว่าเดือนอื่นๆและเป็นเช่นนี้ทุกปียอดขายห้างสรรพสินค้าในเดือนธันวาคม จะสูงกว่ายอดขายเดือนอื่น ๆ และเป็นเช่นนี้ทุกปี เมื่อพิจารณารูปที่ 1.3 ซึ่งแสดงข้อมูลดัชนีผลผลิตสินค้าเกษตรกรรมรายเดือนตั้งแต่ปี 2548 - 2554 ของประเทศหนึ่ง เมื่อเราสังเกต ณ ปี 2548 จะพบว่าดัชนีผลผลิตสินค้าเกษตรกรรมในช่วงเดือนพฤศจิกายนและธันวาคมสูงกว่าเดือนอื่นๆและในปีอื่นๆ ก็จะมีลักษณะเช่นนี้ ดังนั้นเรากล่าวได้ว่าดัชนีผลผลิตอุตสาหกรรมรายเดือนของประเทศนี้มีอิทธิพลของความผันแปรจากฤดูกาลเข้ามา เกี่ยวข้องในเดือนพฤศจิกายนและธันวาคมของทุกปีทั้งนี้อาจเป็นเพราะผลผลิตทางการเกษตรของ ประเทศนี้จะออกมาพร้อมๆ กันในช่วงเวลาดังกล่าว

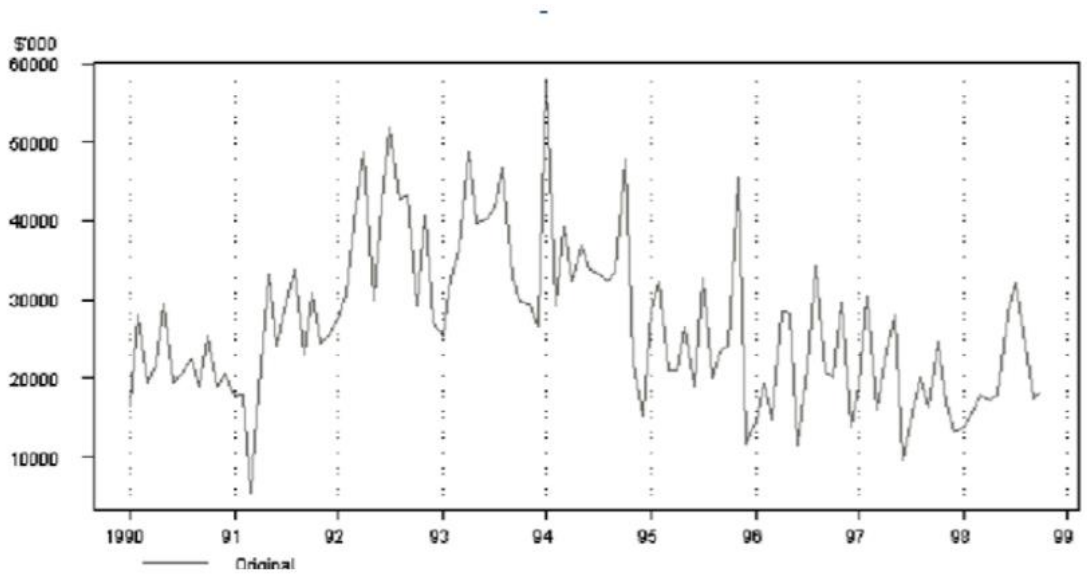



*รูป 4* แสดงดัชนีผลผลิตสินค้าเกษตรของประเทศหนึ่งเป็นรายเดือนตั้งแต่ปี 2548 – 2554

ที่มา : [ออนไลน์] https://economics.utcc.ac.th/wp-content/uploads/Time-Series-for-Econ-and-Bus_Poomthan.pdf

(4) ความผันผวนจากเหตุการณ์ไม่ปกติ (Irregular Fluctuations) คือ การเปลี่ยนแปลงที่ไม่ใช่แนวโน้ม การเปลี่ยนแปลงตามฤดูกาล และการเปลี่ยนแปลงตามวัฏจักรซึ่งมีผลต่ออนุกรม เวลาและไม่สามารถคาดการณ์ล่วงหน้าได้ ทำให้อนุกรมเวลามีค่าที่ผิดปกติไปจากรูปแบบที่เคยเป็น มักเกิดจากเหตุการณ์ไม่คาดฝัน (Shock) เช่น แผ่นดินไหว สึนามิ ระเบิด การหยุดงานประท้วง ฯลฯ ส่วนความผันผวนจากเหตุการณ์ไม่ปกติ คำนวณจากการนำค่าของส่วนแนวโน้ม ค่าของส่วนวัฏจักร และค่าของความผันแปรจากฤดูกาล ไปหักล้างค่าอนุกรมเวลานั้นนั่นเอง

ในทางปฏิบัตินั้นข้อมูลอนุกรมเวลาที่เก็บรวบรวมได้อาจประกอบด้วยส่วนใดส่วนหนึ่งหรือทั้ง 4 ส่วน โดยอาจอยู่ในรูปของผลรวมหรือผลคูณก็ได้




*รูป 5* กราฟอนุกรมเวลาที่แสดงลักษณะของความผันผวนจากเหตุการณ์ไม่ปกติ

ที่มา : [ออนไลน์] เว็บไซต์สำนักงานสถิติแห่งชาติ http://www.abs.gov.au/websitedbs/OpenDocument

## 2.2 ความนิ่งของข้อมูล (Stationary)

**2.2.1 ความหมายของความนิ่งของข้อมูล (Stationary)** หมายถึง ข้อมูลอนุกรมเวลาที่มีค่าเฉลี่ย (Mean) และ ความแปรปรวน (Variance) เท่ากันตลอดระยะเวลาที่ศึกษา

ส่วนข้อมูลที่มีลักษณะไม่นิ่ง (Non-Stationary) หมายถึงข้อมูลอนุกรมที่มีค่าเฉลี่ย (Mean) และความแปรปรวน (Variance) ไม่เท่ากันตลอดระยะเวลาที่ศึกษา หากพบว่าอนุกรมเวลาไม่มีความนิ่ง (nonstationary) แล้ว เราอาจสามารถแปลงข้อมูลนั้นให้มีความนิ่ง (stationary) ได้โดยการทำผลต่างลำดับที่ 1 (First differencing) ซึ่งสามารถคำนวณได้โดยนำข้อมูลในช่วงเวลาก่อนหน้านี้มาหักออกจากข้อมูลปัจจุบัน

**2.2.2 การปรับข้อมูลให้มีความนิ่งของข้อมูล (Stationary)**

1) ถ้าอนุกรมเวลามีแนวโน้ม ($Y_t$) ต้องปรับให้เป็นอนุกรมเวลาใหม่ที่ไม่มีแนวโน้ม ($Z_t$) โดยการใช้วิธี Regular Differencing (ใช้สัญลักษณ์ d)

- กรณีที่หาความแตกต่างแค่ลำดับเดียว (d = 1) เช่น หาผลต่างระหว่างข้อมูลเดือนใดๆกับหนึ่งเดือนก่อนหน้านั้น

$$Z_t = Y_t - Y_{t-1}$$

- ในกรณีที่หาความแตกต่างแค่ลำดับเดียวแล้วยังพบวาข้อมูลอนุกรมค่าของ $Z_t$ ที่ได้มายังคงมีแนวโน้มอยู่ให้ทำการหาความแตกต่างเพิ่มอีกลำดับหนึ่ง (d=2)

$$Z_t = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) = Y_t - 2Y_{t-1} + Y_{t-2}$$

2) ถ้าอนุกรมเวลา ($Y_t$) มีความเป็นฤดูกาล ต้องปรับให้เป็นอนุกรมเวลาใหม่ที่ไม่มีฤดูกาล ($Z_t$) โดยการใช้วิธี Seasonal Differencing(ใช้สัญลักษณ์ D)

- กรณีที่หาความแตกต่างของฤดูกาลแค่ลำดับเดียว (D = 1) เช่น ข้อมูลรายเดือนใดๆก็ให้หาความแตกต่างกับข้อมูลเมื่อ 12 เดือนก่อนหน้านั้น

$$Z_t = Y_t - Y_{t-12}$$

- ในกรณีที่หาความแตกต่างแค่ลำดับเดียวแล้วยังพบวาข้อมูลอนุกรมค่าของ $Z_t$ ที่ได้มายังคงมีความเป็นฤดูกาลอยู่ ให้ทำการหาความแตกต่างเพิ่มอีกลำดับหนึ่ง (D=2)

$$Z_t = (Y_t - Y_{t-12}) - (Y_{t-12} - Y_{t-24}) = Y_t - 2Y_{t-12} + Y_{t-24}$$

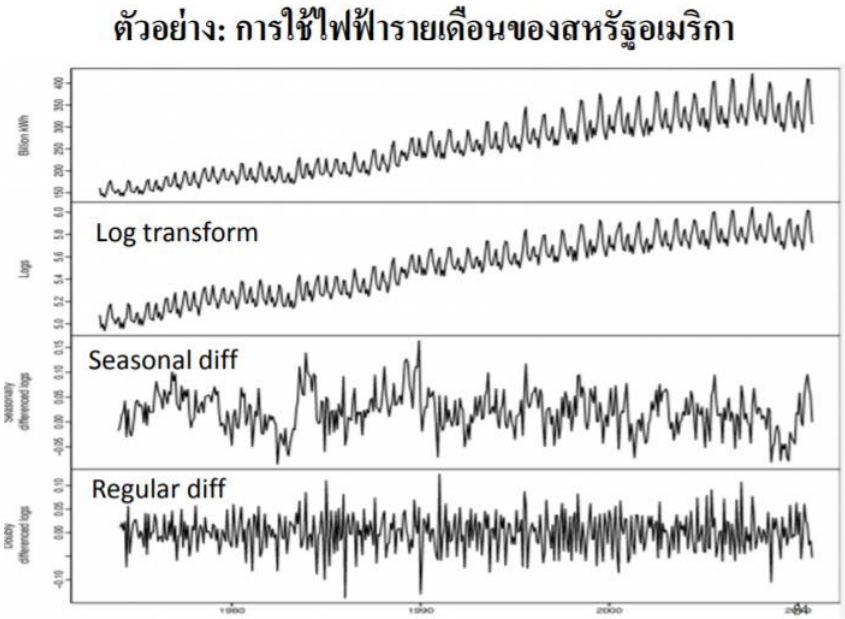
3) ถ้าอนุกรมเวลา ($Y_t$) มีทั้งแนวโน้มและมีความเป็นฤดูกาล ต้องปรับให้เป็นอนุกรมเวลาใหม่ที่ไม่มีทั้งแนวโน้มและฤดูกาล ($Z_t$) โดยการใช้วิธี Regular and Seasonal Differencing

- กรณีที่หาความแตกต่างของทั้งแนวโน้มและฤดูกาลแค่ลำดับเดียว (d= 1 และ D = 1)

$$Z_t = (Y_t - Y_{t-1}) - (Y_{t-12} - Y_{t-13}) = (Y_t - Y_{t-12}) - (Y_{t-1} - Y_{t-13}) = Y_t - Y_{t-1} - Y_{t-12} - Y_{t-13}$$

4) กรณีที่ความแปรปรวนของอนุกรมเวลาไม่คงที่( $Var(y_t)$ ไม่คงที่ในแต่ละเวลา t) จะต้องการปรับอนุกรมเวลา ($Y_t$) ให้เป็นอนุกรมเวลาใหม่($Z_t$) โดยการทำ

• Logarithmic transformation $Z_t = \log(Y_t)$ หรือ • Square root transformation $Z_t = \sqrt{Y_t}$




*รูป 6* ตัวอย่างกราฟอนุกรมเวลาการใช้ไฟฟ้ารายเดือนของสหรัฐอเมริกา

ที่มา : [ออนไลน์] https://algoaddict.com/blog/54520/arima

**รูปช่องที่ 1** : ข้อมูลจะเห็นได้ชัดเจนว่าข้อมูลดิบยังมี Trend ชัดเจนมากก็คือ เห็นได้ชัดว่า ไม่เป็น Stationary

**รูปช่องที่ 2** : ข้อมูล Log Scale ซึ่งสร้างมาจาก การนำข้อมูลดิบมา take Logarithm เข้าไป จะเห็นได้ว่า ก็ยังมี Trend ชัดเจนอยู่ ยังไม่เป็น Stationary

**รูปช่องที่ 3** : รูปนี้มีการเริ่มทำ Differencing รอบแรก จะเห็นว่าข้อมูลเริ่มเป็น Stationary มาขึ้นแล้ว (Trend เริ่มหายไป) แล้วแต่ยังมี Season หรือ Seasonality อยู่

**รูปช่องที่ 4** : นำข้อมูลจากรูปที่ 3 มา Differencing อีกรอบ และในขั้นตอนนี้เอง ข้อมูลเป็นเป็น Stationary โดยสมบูรณ์แล้ว

**2.2.3 การทดสอบความนิ่งของข้อมูล(Stationary) และการทดสอบ Unit Root**

การทดสอบอนุกรมเวลาเป็นสิ่งที่ควรกระทำก่อนที่จะนำข้อมูลอนุกรมเวลามาใช้ในการวิเคราะห์ โดยเฉพาะเงื่อนไขความนิ่งของอนุกรมเวลา (Stationary) ซึ่งเป็นเงื่อนไขที่สำคัญในการนำข้อมูลอนุกรมเวลามาใช้ เนื่องจากถ้าข้อมูลมีลักษณะไม่นิ่งจะส่งผลให้เกิดปัญหาสมการถดถอยไม่แท้จริง ซึ่งจะทำให้ค่าทางสถิติที่ประมาณการได้ไม่มีประสิทธิภาพและขาดความน่าเชื่อถือ ดังนั้นถ้าหากอนุกรมเวลาที่นำมาใช้ไม่คงที่จะต้องทำให้อนุกรมเวลาดังกล่าวคงที่ก่อน โดยการหาผลต่างของอนุกรมเวลาการคงที่ของอนุกรมเวลาหมายถึงอนุกรมเวลาที่อยู่ในสภาวะสมดุลเชิงสถิติ(Statistical equilibrium) ซึ่งก็คือการที่คุณสมบัติทางสถิติของอนุกรมเวลาไม่เปลี่ยนแปลงตามกาลเวลา การทดสอบ Unit Root เป็นการทดสอบว่าข้อมูลที่นำมาศึกษามีนิ่งหรือไม่ สามารถทำได้ โดยการทดสอบ DF (Dickey –Fuller Test) ซึ่งเสนอโดย Dickey และ Fuller ในปี 1981 และวิธีการ ทดสอบ ADF (Augmented Dickey –Fuller Test) ซึ่งเสนอโดย Said และ Dickey ในปี 1984

## 2.3 การทดสอบความสัมพันธ์ (Correlation)

Correlation หรือค่าสหสัมพันธ์เป็นการดูทิศทางความสัมพันธ์ระหว่างตัวแปร 2 ตัว โดยมี Correlation Coefficient (*r*) หรือ ค่าสัมประสิทธิ์สหสัมพันธ์ เป็นตัวบ่งชี้ถึงความสัมพันธ์นี้ ซึ่งค่าสัมประสิทธิ์สหสัมพันธ์นี้จะมีค่าอยู่ระหว่าง -1.0 ถึง +1.0 ซึ่งหากมีค่าใกล้ -1.0 นั้นหมายความว่าตัวแปรทั้งสองตัวมีความสัมพันธ์กันอย่างมากในเชิงตรงกันข้าม หากมีค่าใกล้ +1.0 นั้นหมายความว่า ตัวแปรทั้งสองมีความสัมพันธ์กันโดยตรงอย่างมาก และหากมีค่าเป็น 0 นั้นหมายความว่า ตัวแปรทั้งสองตัวไม่มีความสัมพันธ์ต่อกัน การที่ตัวแปรทั้งสองตัวมีค่าสหสัมพันธ์แสดงออกถึงความสัมพันธ์กันนั้น หมายความว่า ตัวแปรทั้งสองมีแนวโน้มจะไปในทางเดียวกัน

| ค่า | ระดับของความสัมพันธ์ |
|---|---|
| .90- 1.00 | มีความสัมพันธ์กันสูงมาก |
| .70-.90 | มีความสัมพันธ์กันในระดับสูง |
| .50-.70 | มีความสัมพันธ์กันในระดับปานกลาง |

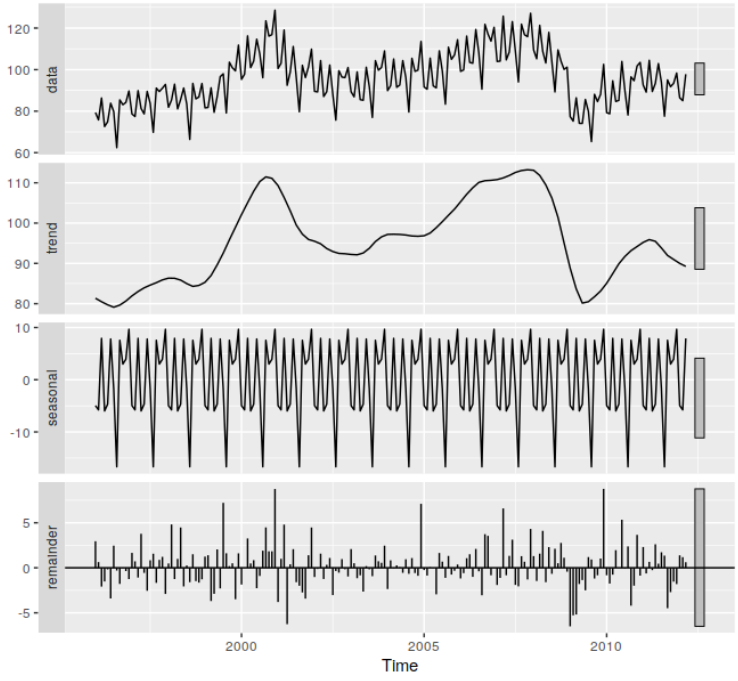
| | |
|---|---|
| .30-.50 | มีความสัมพันธ์กันในระดับต่ำ |
| .00-.30 | มีความสัมพันธ์กันในระดับต่ำมาก |

## 2.4 การสลายตัวอนุกรมเวลา (Time Series Decomposition)

Time Series Decomposition คือ การแยกอนุกรมเวลาออกเป็นส่วนต่างๆ ทำให้เราสามารถวิเคราะห์อนุกรมเวลาได้ง่ายขึ้นเพราะในหลายๆครั้ง เราต้องการพยากรณ์เฉพาะส่วนหรือต้องการวิเคราะห์เฉพาะส่วนของอนุกรมเวลา เช่น เราต้องการวิเคราะห์ข้อมูลที่ไม่มี Seasonality เราก็สามารถแยกอนุกรมเวลาออกมาเฉพาะส่วน Trend + Remainder ออกมาวิเคราะห์ ซึ่งเราเรียกข้อมูลชุดนี้ว่า Seasonally Adjusted Data ซึ่งสามารถทำใช้ในการทำนายทำนายที่เหมาะกับข้อมูล Non-seasonal เช่น Naive, Drift, ETS, ARIMA

อนุกรมเวลาซึ่งประกอบด้วยองค์ประกอบสามส่วน:

- องค์ประกอบแนวโน้ม (trend-cycle component)
- องค์ประกอบจากฤดูกาล (seasonal component)
- องค์ประกอบข้อมูลที่ไม่มีแนวโน้มและฤดูกาล (remainder component)




*รูป 7* ข้อมูลอนุกรมเวลาที่ทำการสลายตัวอนุกรมเวลา

ที่มา : [ออนไลน์] https://otexts.com/fpp2/components.html

## 2.5 วิธีการปรับให้เรียบแบบเอกซ์โพเนนเชียล Single Exponential Smoothing (SES)

วิธีการปรับให้เรียบแบบเอกซ์โพเนนเชียลเป็นวิธีที่ใช้หลักการของวิธีการหาค่าเฉลี่ยเคลื่อนที่ถ่วงน้ำหนักวิธีหนึ่ง ให้ความสำคัญลดหลั่นลงไปกับข้อมูลถัดขึ้นไปในแบบเอกซ์โพเนนเชียล คือการที่ให้ความสำคัญกับข้อมูลเก่าทุกค่า แต่ให้ความสำคัญแก่ค่าที่ใกล้ปัจจุบันมากที่สุด ค่าพยากรณ์จะตอบสนองกับข้อมูลใหม่เป็นหลัก เหมาะกับข้อมูลที่มีการเปลี่ยนแปลงและคาดเดาได้ยาก ในการนี้จะกำหนดน้ำหนักข้อมูลล่าสุดเป็น $\alpha$ ลดหลั่นลงไปตั้งแต่ค่าที่ 1 จนถึงค่าล่าสุด โดยนำเอาความคลาดเคลื่อนของค่าพยากรณ์ในอดีตมาพิจารณา ในการนี้จะกำหนดน้ำหนักข้อมูลล่าสุดเป็น $\alpha$ (Alpha) โดยให้ค่า $\alpha$ (Alpha) อยู่ระหว่าง 0-1 ถ้าค่า $\alpha$ = 1 แสดงว่าให้น้ำหนักกับข้อมูลล่าสุดมาก ค่าพยากรณ์ในช่วงเวลาถัดไปจะเท่ากับข้อมูลจริงในช่วงเวลาล่าสุด แต่ถ้า $\alpha$ มีค่าน้อยก็จะหมายความว่ายึดข้อมูลพยากรณ์ใน อดีตเป็นหลักโดยไม่คำนึงถึงข้อมูลปัจจุบัน การหาค่าพยากรณ์คำนวณได้จากสมการ การหาค่าพยากรณ์จะคำนวณโดยใช้สมการดังนี้

$$F_{t+1} = \alpha A_t + ( 1-\alpha ) F_t$$

$$= \alpha A_t + F_t - \alpha F_t$$

$$= F_t + \alpha A_t - \alpha F_t$$

$$= F_t + \alpha (A_t - F_t)$$

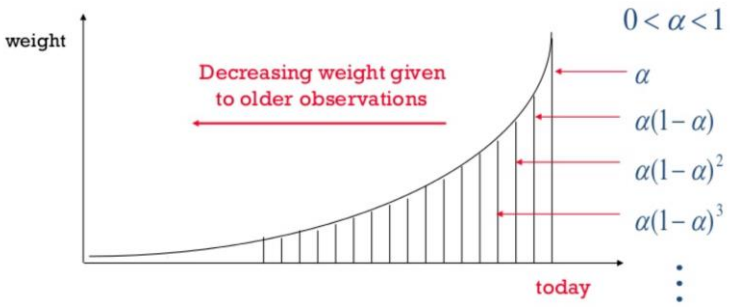
โดย $F_t$ = ค่าพยากรณ์ของช่วงเวลา t

$A_t$ = ค่าที่เกิดขึ้นจริงในช่วงเวลา t

$F_{t+1}$ = ค่าพยากรณ์ของช่วงเวลา t+1

$\alpha$ = ค่าถ่วงน้ำหนัก

สำหรับการสุ่มเลือกค่าคงที่ในการปรับเรียบควรเลือกตามการเปลี่ยนแปลงของข้อมูลหากมีการเปลี่ยนแปลงเร็วมาก ควรเลือกสุ่มค่าคงที่การปรับเรียบมากๆหรือเข้าใกล้ 1 ในทางกลับกันหากการเปลี่ยนแปลงข้อมูลน้อย ควรเลือกค่าคงที่การปรับเรียบให้มีค่าน้อยหรือเข้าใกล้ 0




*รูป 8* การถ่วงน้ำหนักของวิธีการปรับให้เรียบแบบเอกซ์โพเนนเชียล

ที่มา : [ออนไลน์] https://pt.slideshare.net/JairoMoreno15/exponential-smoothing-52500403?smtNoRedir=1

## 2.6 วิธีการปรับให้เรียบเอกซ์โพเนนเชียลแบบโฮลต์-วินเทอร์ (Holt-Winters exponential smoothing method)

วิธีการปรับให้เรียบแบบเอ็กโพเนนเชียลแบบโฮลต์-วินเทอร์ จัดเป็นเทคนิคการวิเคราะห์ข้อมูลอนุกรมเวลา (Time Series analysis) ซึ่งนำข้อมูลในอดีตมาวิเคราะห์องค์ประกอบเพื่อกำหนดตัวแบบจำลองและนำแบบจำลองนั้นมาพยากรณ์ค่าอนาคต วิธีการนี้เหมาะกับข้อมูลที่มีแนวโน้มเชิงเส้นและมีอิทธิพลของฤดูกาลซึ่งวิธีการนี้แบ่งออกเป็น 2 แบบคือวิธีการปรับให้เรียบแบบโฮลต์-วินเทอร์ที่มีอิทธิพลของฤดูกาลแบบบวก (Holt-Winters additive seasonal exponential smoothing method; Additive HWS) ซึ่งเหมาะกับการพยากรณ์ข้อมูลที่มีแนวโน้มและมีอิทธิพลของฤดูกาลที่ไม่เพิ่มขึ้นตามเวลาที่เปลี่ยนแปลงไปและวิธีการปรับให้เรียบแบบเอ็กโพเนนเชียลโฮลต์-วินเทอร์ที่มีอิทธิพลของฤดูกาลแบบคูณ (Holt-Winters multiplicative seasonal exponential smoothing method; Multiplicative HWS) ซึ่งเหมาะกับการพยากรณ์ข้อมูลที่มีแนวโน้มและมีอิทธิพลของฤดูกาลที่เพิ่มขึ้นตามลำดับเวลาที่เปลี่ยนแปลงไปงานวิจัยนี้จะกล่าวถึงวิธีการปรับเรียบเอกซ์โพเนนเชียลโฮลต์-วินเทอร์ ที่มีฤดูกาลแบบบวกเท่านั้นโดยมีสมการพยากรณ์สำหรับการพยากรณ์ข้อมูลรายเดือนคือ

$$F_{t+m} = L_t + mT_t + S_{t-12+m}$$

$$L_t = \alpha ( Y_t - S_t - 12 ) + ( 1 - \alpha) ( L_{t-1} + T_{t-1} )$$

$$T_t = \beta ( L_t - L_{t-1} ) + ( 1 - \beta ) T_{t-1}$$

$$S_t = \gamma ( Y_t - L_t ) + ( 1 - \gamma ) S_{t-12}$$

โดย $F_{t+m}$ = ค่าพยากรณ์ ณ เดือนที่ t+m ; m= 1,2...12

$Y_t$ = ค่าข้อมูลจริง ณ เดือนที่ t

$L_t$ = ค่าประมาณระดับของข้อมูล ณ เดือนที่ t

$T_t$ = ค่าประมาณแนวโน้มของข้อมูล ณ เดือนที่ t

$S_t$ = ค่าประมาณฤดูกาลของข้อมูล ณ เดือนที่ t

$\alpha$ = ค่าคงที่ปรับระดับของข้อมูลมีค่าระหว่าง 0 ถึง 1

$\beta$ = ค่าคงที่ปรับแนวโน้มของข้อมูลมีค่าระหว่าง 0 ถึง 1

$\gamma$ = ค่าคงที่ปรับฤดูกาลของข้อมูลมีค่าระหว่าง 0 ถึง 1

## 2.7 แบบจำลอง ARIMA (Autoregressive Intergrated Moving average)

ตัวแบบ ARIMA มีข้อสมมติว่าค่าปัจจุบันของค่าสังเกตเป็นฟังก์ชันเชิงเส้นของค่าสังเกตและค่าความคลาดเคลื่อนสูงในอดีตตัวแบบคือ

$$Y_t = \delta + \phi_1 Y_{t-1} + .... + \phi_P Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - ... - \theta_q \varepsilon_{t-q}$$

โดยที่ $Y_t$ = เป็นค่าสังเกตุของอนุกรมเวลาที่เวลา t

$\delta$ = เป็นค่าคงที่ในตัวแบบ

$\varepsilon_t$ = เป็นความคลาดเคลื่อนสุ่มที่เวลา t ซึ่งมีสมมติว่า $\varepsilon_t$ เป็นตัวสุ่มที่เป็นอิสระกัน ค่าเฉลี่ย 0 ความแปรปรวนคงที่

$\phi_I$ = ( i = 1,...,p) และ $\theta_j$ (j = 1,..,q) เป็นพารามิเตอร์ในตัวแบบ ส่วน p และ q เป็นจำนวนเต็มซึ่งแสดงอันดับของตัวแบบ

### 2.7.1 แบบจำลองตัวแบบ ARIMA (p, d, q) มีส่วนประกอบที่สำคัญ 3 ส่วน ได้แก่

Auto Regressive AR : (p) หมายถึง กระบวนการของค่าสังเกตค่าหนึ่ง (Y) ถูกกำหนดมาจาd ความสัมพันธ์ของตัวมันเองในอดีต ($y_{t-1}$,..., $y_{t-p}$) อย่างต่อเนื่อง (Continuous) โดยที่มี p คือลำดับของคาบเวลาในอดีตที่เหมาะสม เพื่ออธิบาย Y ณ เวลาปัจจุบัน

Integrated (I) : เป็นจำนวนครั้งที่หาผลต่าง (Difference) เหตุผลสำคัญที่ต้องหาผลต่างของอนุกรมเวลาเนื่องจากแบบจำลอง ARIMA จะต้องใช้ในการวิเคราะห์ข้อมูลอนุกรมเวลาที่มีคุณสมบัติคงที่ (Stationary) เท่านั้น

ในกรณีที่ข้อมูลอนุกรมเวลาที่ใช้ในการวิเคราะห์สมบัติไม่คงที่ (Nonstationary) จะต้องทำการแปลงข้อมูลอนุกรมเวลา ดังกล่าวให้มีคุณสมบัติคงที่ก่อน โดยการหาผลต่างของข้อมูลอนุกรมเวลา หรือการหาค่า Natural logarithm ของอนุกรมเวลาก่อนที่จะนำข้อมูลไปใช้สร้างแบบจำลอง ARIMA

Moving Average MA : (q) หมายถึง กระบวนการของค่าสังเกตค่าหนึ่ง ($\varepsilon$) ถูกกำหนดมาจากค่าความคลาดเคลื่อนในอดีต ($\varepsilon_{t-1}$ ,..., $\varepsilon_{t-}$) อย่างต่อเนื่อง (Continuous) โดยที่มี q คือลำดับของคาบเวลาในอดีตที่เหมาะสม เพื่ออธิบาย($\varepsilon$) ณ เวลาปัจจุบัน

**2.7.2 การกำหนดตัวแบบและอันดับของ p,d,q**

การกำหนดรูปแบบ p, d และ q ของ ARIMA หรือ ARMA พิจารณาออกเป็น 3 ส่วน คือ

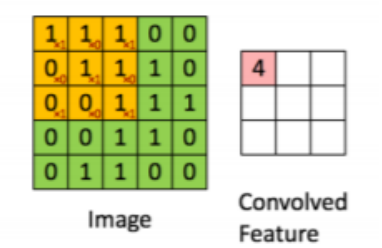
- ระบุ I(d) (Order of integration) หรืออันดับความหยุดนิ่ง d พิจารณาจากกราฟ Correlogram หรือการทดสอบ Unit root
- ระบุ AR(p) ด้วยวิธี Correlogram
  - p พิจารณาจาก Partial Autocorrelation Function (PACF), $p_k$ ($z_t$ )
  - ลำดับ p ให้พิจารณาแท่ง (Spike) ที่ยื่นออกมาประมาณ 3-5 แท่งแรก
- ระบุ MA(q) ด้วยวิธี Correlogram
  - q พิจารณาจาก Autocorrelation Function (ACF), $p_{kk}$($z_t$ )
  - ลำดับ q ให้พิจารณาแท่ง (Spike) ที่ยื่นออกมาประมาณ 3-5 แท่งแรก

## 2.8 แบบจำลอง CNN (Convolutional Neural Network)

โครงข่ายประสาทเทียมคอนโวลูชัน (Convolutional Neural Network) เป็นโครงข่ายประสาทเทียมเชิงลึกประเภทหนึ่งมีจุดเริ่มต้นมาจากงานวิจัยทางด้านการรู้จำภาพตัวอักษร. โดยมักจะใช้ข้อมูลรับเข้าเป็นเมทริกซ์จากการแปลงมาจากรูปภาพ โครงสร้างของโครงข่ายประสาทเทียมแบบคอนโวลูชันมีรายละเอียดดังนี้

1.ชั้นคอนโวลูชัน (Convolutional Layer)

ประกอบด้วยเคอร์เนลฟิลเตอร์ (Kernel Filter) ซึ่งแต่ละเคอร์เนลจะถูกกำหนดโดยการสุ่มในขั้นตอนการเรียนรู้เริ่มต้นหลังจากนั้นจะปรับค่าโดยการเรียนรู้แบบแพร่กระจายย้อนกลับจำนวนผลลัพธ์ที่ได้จากชั้นคอนโวลูชันจะเท่ากับจำนวนของเคอร์เนลฟิลเตอร์เรียกว่าฟีเจอร์แมพ (Feature map) ในชั้นของคอนโวลูชันมักจะตามด้วยฟังก์ชันกระตุ้นซึ่งเป็นฟังก์ชันแบบไม่เชิงเส้น (Non-linear Function) ขั้นตอนการทำคอนโวลูชันเป็นดังรูป



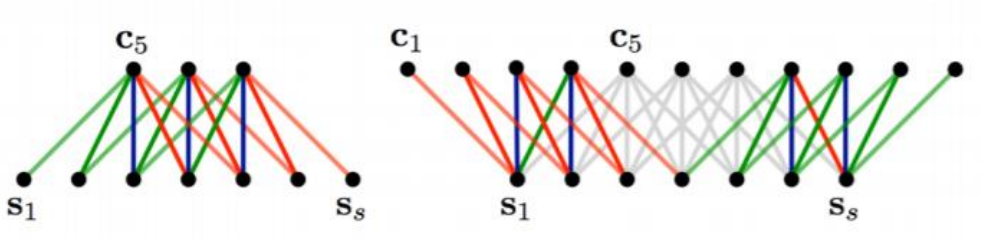

*รูป 9* แสดงขั้นตอนการทำคอนโวลูชันระหว่างภาพนำเข้า

ที่มา : [ออนไลน์] https://cuir.car.chula.ac.th/dspace/bitstream/123456789/58338/1/5870956621.pdf

- ขนาดของตัวกรอง (Filter Size) คือ ความกว้างและความสูงของตัวกรองที่จะนำมาใช้ในการทำคอนโวลูชัน

- การทำคอนโวลูชันแบบแคบ (Narrow Convolution) นิยมใช้ในการทำคอนโวลูชันโดยทั่วไปในการทำคอนโวลูชันตัวกรองที่นำไปทำการดอทเมทริกซ์นั้นจะไม่มีการกระทำเลยขอบของเมทริกซ์รับเข้ากล่าวคือการคอนโวลูชันที่มีข้อมูลรับเข้าขนาด NxN กับตัวกรองขนาด mxm จะได้เมทริกซ์ขนาด (N-m + 1) x (N-m + 1)

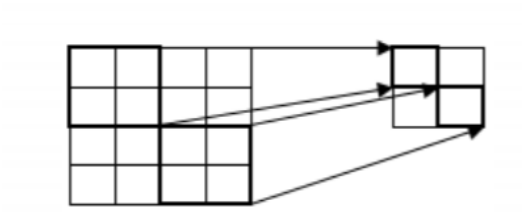
นอกจากนี้ยังมีการทำคอนโวลูกชันแบบกว้าง (Wide Convolution) จะมีการกระทำเลยขอบของเมทริกซ์รับเข้าออกไปโดยพื้นที่ที่เกินออกไปนั้นจะมีการแทนค่าของข้อมูลช่องนั้น ๆ ด้วย 0 เรียกว่าการเสริมด้วยศูนย์ (Zero Padding) กล่าวคือการทำคอนโวลูชันแบบกว้างที่มีข้อมูลรับเข้าขนาด NxN กับตัวกรองขนาด mxm จะได้เมทริกซ์ขนาด (N + m-1) x (N + m-1) การทำคอนโวลูชันแบบกว้างนี้มีจุดประสงค์เพื่อป้องกันการสูญเสียข้อมูลตรงบริเวณขอบของข้อมูลรับเข้าโดยรูปที่ 6 เป็นการเปรียบเทียบระหว่างการทำคอนโวลูชั้นแบบแคบปละแบบกว้าง




*รูป 10* เปรียบเทียบการทำคอนโวลูชันแบบแคบและแบบกว้าง

ที่มา : [ออนไลน์] https://cuir.car.chula.ac.th/dspace/bitstream/123456789/58338/1/5870956621.pdf

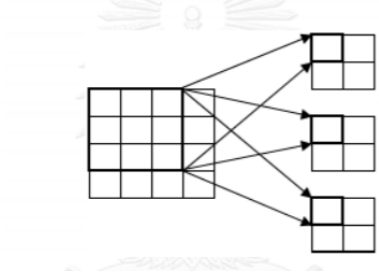
- ขนาดของการก้าวข้าม (Stride Size) ขนาดของการก้าวข้ามคือจำนวนช่องของข้อมูลรับเข้าที่จะถูกเลื่อนไปเมื่อทำการหาผลลัพธ์ของคอนโวลูชันในแต่ละช่องดังรูปที่ 7 แต่โดยทั่วไปมักใช้ขนาดของการก้าวข้ามเป็นหนึ่ง



*รูป 11* การทำคอนโวลูชันโดยมีข้อมูลรับเข้าขนาด 4x4 ตัวกรองขนาด 2x2

ที่มา : [ออนไลน์] https://cuir.car.chula.ac.th/dspace/bitstream/123456789/58338/1/5870956621.pdf

- จำนวนตัวกรอง (Number of Filters) ในการแต่ละชั้นของคอนโวลูชั้นสามารถมีตัวกรองได้มากกว่าหนึ่งดังรูปที่คโดยน้ำหนักของตัวกรองแต่ละตัวจะแยกกันจำนวนตัวกรองในชั้นคอนโวลูชันใด ๆ จะเป็นการกำหนดจำนวนช่องสัญญาณของข้อมูลรับเข้าในชั้นถัดไป



*รูป 12* การทำคอนโวลูชันโดยมีจำนวนตัวกรองเท่ากับ 3

ที่มา : [ออนไลน์] https://cuir.car.chula.ac.th/dspace/bitstream/123456789/58338/1/5870956621.pdf

-จำนวนช่องสัญญาณ (Channel) จำนวนช่องสัญญาณหรือความลึกของข้อมูลรับเข้าอาจเกิดจากจำนวนของตัวกรองในชั้นคอนโวลูชันก่อนหน้าหรืออาจมีค่ามากกว่าหนึ่งเช่นการใช้ช่องสัญญาณทั้งหมด 3 ช่องสัญญาณแทนค่าแม่สีในงานวิจัยทางด้านการใช้รูปภาพหรือสามารถเขียนสมการในการคำนวณผลลัพธ์ของชั้นคอนโวลูชันได้ดังสมการ

$$z_{ij}^{l} = \sum_{c=0}^{k-1} \sum_{a=0}^{m-1} \sum_{b=0}^{m-1} w_{a,b}^{l} a_{c,i+a,j+b}^{l-1} + b^{l}$$

โดยที่ k แทนจำนวนช่องสัญญาณ

หลังจากชั้นคอนโวลูชันจะมีการคำนวณฟังก์ชันกระตุ้นเพื่อนำผลลัพธ์ที่ได้ส่งต่อเป็นข้อมูลเข้าให้ชั้นต่อไปในเน็ตเวิร์กโดยฟังก์ชันกระตุ้นที่เหมาะกับการทำนายอายุการใช้งานคงเหลือซึ่งมีลักษณะเป็นค่าตัวเลข ได้แก่ ฟังก์ชันเรคติไฟต์เชิงเส้นและฟังก์ชันเอกซ์โพเนเชียลเชิงเส้นแต่ทั้งสองฟังก์ชันมีความแตกต่างกัน ซึ่งถ้าข้อมูลเข้ามีค่าน้อยกว่าหนึ่งฟังก์ชันเรคคิไฟต์เชิงเส้นจะมีขอบเขตของผลลัพธ์เท่ากับศูนย์ในขณะที่ฟังก์ชันเอกซ์โพเนเชียลเชิงเส้นจะให้ขอบเขตของผลลัพธ์ติดลบตามแนวโน้มแบบเอกซ์โพเนเชียลจึงส่งผลให้ค่าที่แบบจำลองทำนายได้มีความแม่นยำมากกว่า

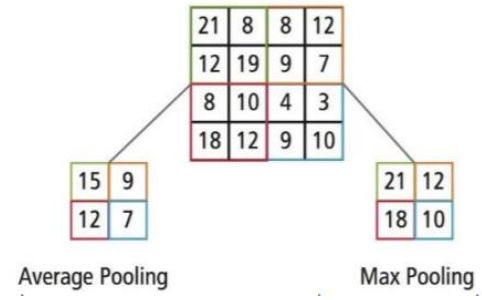
2. ชั้นพูลลิง (Pooling Layer หรือ Subsampling Layer) ในชั้นนี้มีจุดประสงค์เพื่อทำการลดขนาดของข้อมูลที่ผ่านการคอนโวลูชันมาแล้วนิยมนำมาทำต่อกับชั้นคอนโวลูชันแต่ก็อาจไม่จำเป็นต้องนำมาต่อกันเสมอไปซึ่งจะขึ้นอยู่กับการออกแบบการพูลลิงที่เป็นที่นิยมมีสองวิธีคือ

-การทำพูลลิงแบบหาค่ามากที่สุด (Max Pooling)

เป็นวิธีที่นิยมแพร่หลายมากในงานวิจัยด้านคอนโวลูชันแนลโครงข่ายประสาทเทียมในปัจจุบัน

-การทำพูลลิงแบบหาค่าเฉลี่ย (Average Pooling)

มีวิธีการพูลลิงเหมือนกับการพูลลิงแบบหาค่ามากที่สุดแต่ผลลัพธ์ที่ได้จากการพูลลิงจะเป็นค่าเฉลี่ยในรูปที่ .....ด้านล่าง แสดงให้เห็นถึงการเปรียบเทียบกันของการทำพูลลิงทั้งสองแบบ




*รูป 13* แสดงการพูลลิงแบบหาค่าเฉลี่ยและหาค่ามากที่สุด

ที่มา : [ออนไลน์] https://cuir.car.chula.ac.th/dspace/bitstream/123456789/58338/1/5870956621.pdf

3. ชั้นการเชื่อมโยงเต็มรูปแบบ (Fully Connected Layer)

เป็นชั้นสุดท้ายของนิวรอลโครงข่ายประสาทเทียมคอนโวลูชันจะเป็นการเชื่อมโยงเต็มรูปแบบหลังจากการประกอบกันของชั้นคอนโวลูชันและชั้นพูลลิงโดยในชั้นนี้ประกอบด้วยชั้นย่อย ๆ ที่มีเพอร์เซ็ปตรอนอยู่จำนวนหนึ่ง

ซึ่งเพอร์เซ็ปตรอนแต่ละตัวจะมีเส้นเชื่อมกับเพอร์เซ็ปตรอนทุกตัวในชั้นก่อนหน้าและเพอร์เซ็ปตรอนทุกตัวในชั้นถัดไปทำให้การคำนวณแบบป้อนไปข้างหน้าและการแพร่กระจายย้อนกลับสามารถทำได้ด้วยวิธีการปกติ

## 2.9 แบบจำลอง LSTM (Long Short-Term Memory)

LSTM แบบจำลองโครงข่ายประสาทเทียม เป็นแบบจำลองที่ถูกพัฒนาขึ้นจาก RNN (Recurrent Neural Network) เนื่องจาก RNN แบบจำลองเดิมจะมีปัญหา 2 อย่าง คือ ปัญหา Vanishing Gradient และปัญหา Exploding Gradient โดยที่แบบจำลอง LSTM สามารถเชื่อมต่อข้อมูลย้อนหลังทำให้แบบจำลองสามารถวิเคราะห์จุดข้อมูลเดียวได้ เช่น การวิเคราะห์ข้อมูลรูปภาพ และวิเคราะห์ข้อมูลที่เป็นลำดับได้ (Sequence)

**จุดเด่นของ LSTM**

-แบบจำลองสามารถเชื่อมต่อข้อมูลย้อนหลังได้ทำให้ เราสามารถย้อนกลับไปหาต้นทางได้ว่ามาผ่านอะไรมาบ้าง (ต่างจาก RNN ที่ไปข้างหน้าอย่างเดียว)

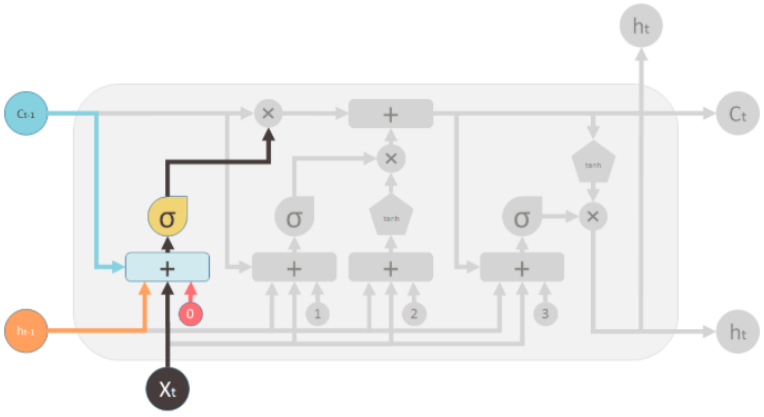
-แบบจำลองสามารถจดจำค่าในช่วงเวลาต่างๆ ได้ โดยผ่านประตู (Gate) ที่จะทำหน้าที่ควบคุมการไหลของข้อมูลเข้าและข้อมูลออกจากเซลล์

**จุดด้อยของ LSTM**

-LSTM แบบจำลองจะเป็น Blackbox ทำให้ไม่สามารถเรียนรู้วิธีการคำนวณภายในได้

**โครงสร้างของ LSTM** ประกอบไปด้วย 4 ส่วนหลัก คือ Forget,Write,Update Cell State,Read

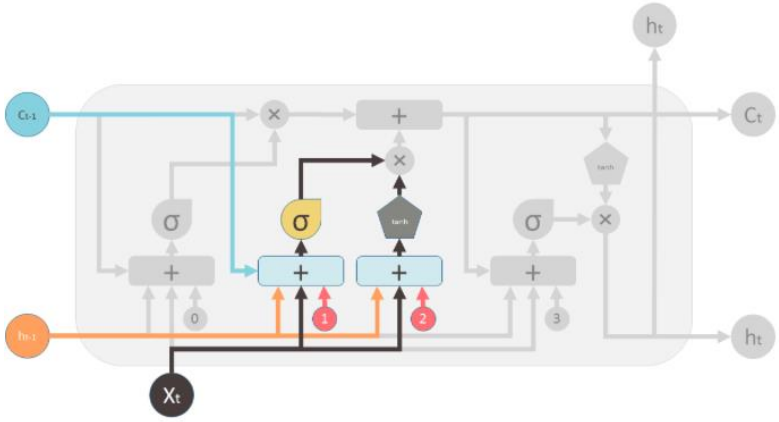
1. Forget หรือ Forget Gate เป็นการลืมหรือลบข้อมูล Cell State เพื่อเตรียมตัวในการที่จะรับข้อมูลใหม่ที่กำลังจะเข้ามา โดยที่ Forget Gate จะคอยควบคุมการไหลข้อมูล ซึ่งจะตัดสินว่าเราจะลืมข้อมูลในทิ้งไปหรือจะยังเก็บข้อมูลนั้นไว้ต่อ ฟังก์ชั่นที่ใช้ เป็น Sigmoid Function คือ ถ้า 1 ไม่ลบ แต่ถ้าเป็น 0 คือ ลบทิ้งไป




*รูป 14* โครงสร้าง Forget หรือ Forget Gate

ที่มา : [ออนไลน์] https://medium.com/@pradyasin/what-is-lstm-1e6e92ba5b4c

2. Write หรือ Input Gate การ Write มีการทำงาน 2 ขั้นตอน คือ Input Gate และ Input Modulation Gate



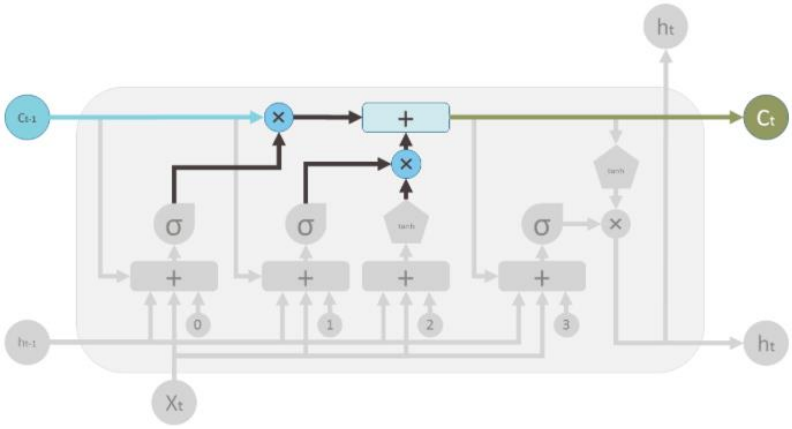

*รูป 15* โครงสร้าง Write หรือ Input Gate

ที่มา : [ออนไลน์] https://medium.com/@pradyasin/what-is-lstm-1e6e92ba5b4c

2.1 Input gate จะคำนวณว่าจะ Update State ด้วย Input data หรือไม่ โดยที่ Input Gate จะคอยควบคุมการไหลข้อมูล ซึ่งจะตัดสินว่าเราอนุญาตให้ Update State หรือไม่ ฟังก์ชั่นที่ใช้ เป็น Sigmoid Function คือ ถ้า 1 อนุญาต แต่ถ้าเป็น 0 คือ ไม่อนุญาต

2.2 Input Modulation Gate จะเป็นการปรับค่าที่ได้จาก Input Gate โดยใช้ Tanh Function ซึ่งค่าที่ได้นั้นจะถูกเรียกว่า Cell State Candidate

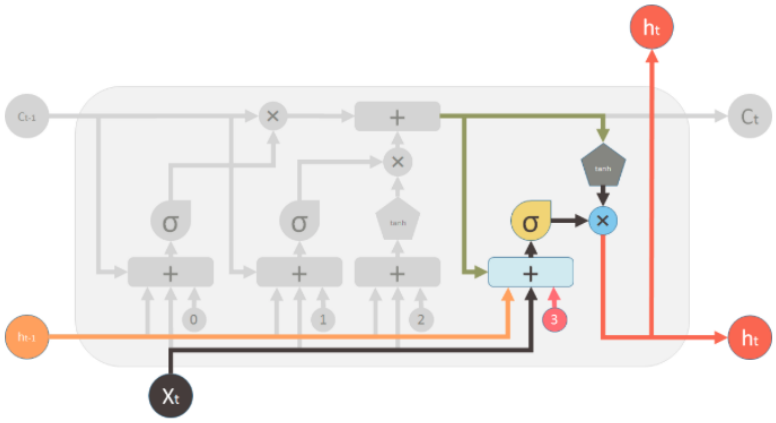
3. Update Cell State เป็นการทำงานหลังที่ได้ข้อมูลครบแล้วจาก Forget Gate, Input Gate, Input Modulation Gate



*รูป 16* โครงสร้าง Update Cell State

ที่มา : [ออนไลน์] https://medium.com/@pradyasin/what-is-lstm-1e6e92ba5b4c

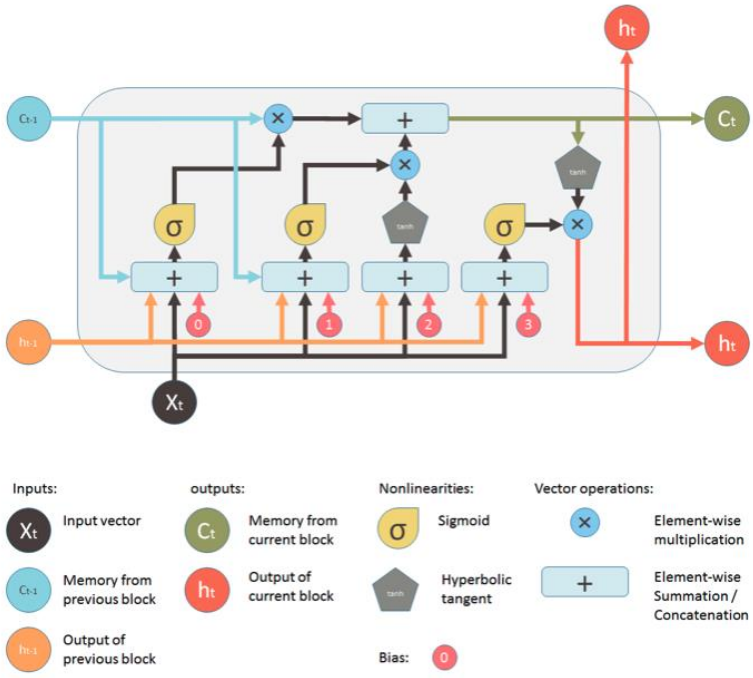
4. Read หรือ Output Gate เป็นการอนุญาตให้สามารถ Read ข้อมูล Output ได้โดยที่ Output Gate จะคอยควบคุมการไหลข้อมูล ซึ่งจะตัดสินว่าเราอนุญาตให้ Read หรือไม่ ฟังก์ชั่นที่ใช้ เป็น Sigmoid Function คือ ถ้า 1 อนุญาต แต่ถ้าเป็น 0 คือ ไม่อนุญาต



*รูป 17* โครงสร้าง Read หรือ Output Gate

ที่มา : [ออนไลน์] https://medium.com/@pradyasin/what-is-lstm-1e6e92ba5b4c

**ภาพรวมโครงสร้าง LSTM**




*รูป 18* ภาพรวมโครงสร้าง LSTM ทั้งหมด

ที่มา : [ออนไลน์] https://medium.com/@pradyasin/what-is-lstm-1e6e92ba5b4c

## 2.10 การวัดประสิทธิภาพ (Performance Evaluation)

ในการประเมินว่าแบบจำลองทำนายได้แม่นยำเพียงใด นิยมการวิเคราะห์จากความคลาดเคลื่อนระหว่างค่าทำนายได้กับค่าจริง หรือค่าผลเฉลยของชุดข้อมูล วิธีการวัดที่นิยมใช้ในการประเมินได้แก่ ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย(Root Mean Square Error; RMSE) ซึ่งเป็นค่าที่ใช้วัดความแตกต่างระหว่างค่าจริงกับค่าที่ทำนายได้ หากมีค่าน้อยแสดงว่าแบบจำลองทำนายได้ใกล้เคียงกับความจริงและเกิดความคลาดเคลื่อนน้อยดังสมการ

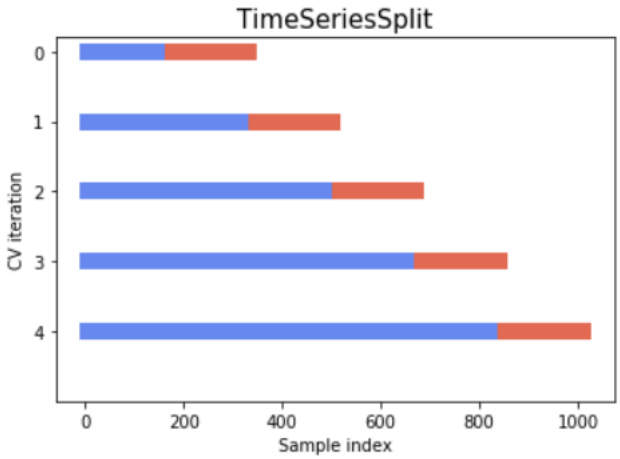
$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(Y_t^s - Y_t^a)^2}$$

โดย $Y_t^s$ แทนค่าประมาณจากแบบจำลอง

$Y_t^a$ แทนค่าจริง

$T$ แทนจำนวนคาบเวลาที่ใช้ประมาณแบบจำลอง

## 2.11 การแบ่ง Cross Validation แบบ Timeseriessplit




*รูป 19* การแบ่ง Cross Validation แบบ Timeseriessplit

ที่มา : [ออนไลน์] https://medium.com/@soumyachess1496/cross-validation-in-time-series-566ae4981ce4

- วิธีการที่สามารถใช้สำหรับตรวจสอบความถูกต้องของแบบจำลองแบบอนุกรมเวลาคือการแบ่ง Cross Validation แบบ Timeseriessplit ซึ่งเป็นการแบ่งข้อมูลออกเป็นหลายๆชุด เพื่อวัดประสิทธภาพของแบบจำลอง โดยที่การแบ่ง Train data และ Test data ซึ่งในข้อมูลชุดถัดไป Train data จะเกิดจาก Train data และ Test data ของชุดก่อนหน้า

## 2.12 งานวิจัยที่เกี่ยวข้อง

### 2.12.1 Forecasting Volatility of Gold Price with Artificial Neural Networks (2560)

ใช้ dataset จาก Gold Traders Association โดยเลือก ข้อมูลราคาทองคำสูงสุดรายวัน ตั้งแต่วันที่ 2 พฤษภาคม 2549 ถึง 30 กันยายน 2558 90% ของข้อมูลทั้งหมดนำมาใช้เป็น Train Set ส่วน 10% สุดท้ายนำมาใช้เป็น Test Set

เมื่อทำการจะแบ่งข้อมูลเรียบร้อยก็จะนำข้อมูลมาใช้งานกับ Artificial Neural Network ที่พัฒนามาจาก Biological Neuron โดยผลลัพธ์ที่ได้เป็นค่า Non-stationary = 0.05 ค่า stationary = 0.05 จากการพิจารณาความสัมพันธ์ราคาทองกับราคาน้ำมันโลกอยู่ที่ระหว่าง 0.41 ถึง 0.70 ส่วนกับเงินดอลลาร์สหรัฐที่มีค่า -0.85 และ ANN มีประสิทะธิภาพดีที่สุดเมื่อมีขนาด 20 neuron ใน hidden layer มีค่า MAE มีค่าเท่ากับ 0.0120 RMSE มีค่าเท่ากับ 0.0368 และ MAPE มีค่าเท่ากับ 1.2229

**ข้อดี** : ค่าจะใช้ในการประเมินประสิทธิภาพการพยากรณ์มีหลายค่า มีการทดลองหาจำนวน neuron ที่ดีที่สุดของ ANN จำนวนมาก(1-50 neurons)

**ข้อเสีย** : ไม่มีการทดลองเปรียบเทียบกับ Artificial Neural Network ที่มี hidden layer มากกว่า 1 layer

**2.12.2 Gold Price Forecasting Using ARIMA Model (2559)**

ใช้ dataset ราคาทองจาก Multi Commodity Exchange of India Ltd (MCX) ในช่วงเดือน พฤษจิกายน 2546 ถึง มกราคม 2557 นำมาใช้งานกับ ARIMA(p,d,q) model ขนาดต่างๆโดย p มาจากการ autocorrelation ,d เป็นจำนวนการ integrate และ q เป็นการทำ moving average ผลลัพธ์การทดลองทำให้ทราบว่า model ARIMA(1,1,1)เป็น model ที่มีความแม่นยำสูงสุด โดยมีค่า RMSE เท่ากับ 719.18 ค่า MAPE เท่ากับ 3.245 และ ค่า MAE เท่ากับ 477.330

**ข้อดี** : มีการทดสอบ ARIMA กับ parameter หลายรูปแบบ เพื่อหา model ARIMA ที่ดีที่สุด มีเกณฑ์วัดประสิทธิภาพให้การพยากรณ์หลายรูปแบบ

**ข้อเสีย** : ในการทดลองนั้น ไม่มีการเปลี่ยนแปลงค่า p ทุก model ที่ใช้งาน มีค่า p = 1

# บทที่ 3
# วิธีการดำเนินโครงงาน

## 3.1 ขั้นตอนการดำเนินงาน

3.1.1 เริ่มต้นและวางแผนโครงการ

3.1.1.1 วางแผนจัดทำโครงการ

3.1.1.2 หาข้อมูลศึกษาแนวคิดและทฤษฎี

3.1.1.3 เลือกแบบจำลอง

3.1.1.4 ศึกษาแบบจำลองและปัจจัยที่เกี่ยวข้องกับราคาทอง

3.1.2 รวบรวมข้อมูล

3.1.2.1 หา Dataset

3.1.2.2 เตรียมข้อมูลให้พร้อมกับการใช้งานในแบบจำลอง

3.1.3 การวิเคราะห์ข้อมูล

3.1.3.1 ทำความเข้าใจข้อมูล

3.1.3.2 หาความสัมพันธ์ของข้อมูล

3.1.4 ปรับใช้กับแบบจำลอง

3.1.4.1 ใช้งานแบบจำลอง

3.1.4.2 ปรับพารามิเตอร์ในแบบจำลองเพื่อปรับปรุงประสิทธิภาพ

3.1.5 ทดสอบความแม่นยำของแบบจำลอง

3.1.5.1 การเปรียบเทียบความแม่นยำของการพยากรณ์ด้วย RMSE

3.1.5.2 การแบ่งข้อมูลเพื่อนำทดสอบประสิทธิภาพของแบบจำลอง

3.1.6 นำแบบจำลองไปใช้ทำนายราคาทองคำในอนาคต

3.1.6.1 แสดงผลการทำงานของแบบจำลองผ่าน API

3.1.6 สรุปและเผยแพร่งานวิจัย

3.1.6.1 สรุปผลงานเพื่อนำเสนอ

## 3.2 แผนการดำเนินงานตลอดโครงงาน

| แผนดำเนินโครงการวิจัย | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **ขั้นตอนการดำเนินการ** | **ต.ค. 62** | **พ.ย. 62** | **ธ.ค. 62** | **ม.ค. 63** | **ก.พ. 63** | **มี.ค. 63** | **เม.ย. 63** | **พ.ค. 63** |
| **1. เริ่มต้นและวางแผนโครงการ** | | | | | | | | |
| 1.1 วางแผนจัดทำโครงการ | ← | → | | | | | | |
| 1.2 หาข้อมูลศึกษาแนวคิดและทฤษฎี | ← | → | | | | | | |
| 1.3 เลือกแบบจำลอง | ← | → | | | | | | |
| 1.4 ศึกษาแบบจำลองและปัจจัยที่เกี่ยวข้องกับราคาทอง | ← | → | | | | | | |
| **2. รวบรวมข้อมูล** | | | | | | | | |
| 2.1 ข้อมูลราคาทองคำ | | ← | → | | | | | |
| 2.2 ข้อมูลราคาทองแดง | | ← | → | | | | | |
| 2.3 ข้อมูลราคาก๊าซธรรมชาติ | | ← | → | | | | | |
| 2.4 ข้อมูลราคาน้ำมันดิบ : เบรนต์ – ยุโรป | | ← | → | | | | | |
| 2.5 ข้อมูลราคาน้ำมันดิบ : (WTI) | | ← | → | | | | | |
| 2.6 ข้อมูลราคาปิโตรเลียมและของเหลวอื่น ๆ | | ← | → | | | | | |
| 2.7 ข้อมูลราคาแพลตตินัม | | ← | → | | | | | |

| ขั้นตอนการดำเนินการ | ต.ค. 60 | พ.ย. 60 | ธ.ค. 60 | ม.ค. 61 | ก.พ. 61 | มี.ค. 61 | เม.ย. 61 | พ.ค. 61 |
|---|---|---|---|---|---|---|---|---|
| **3. การวิเคราะห์ข้อมูล** | | | | | | | | |
| 3.1 ทำความสะอาดข้อมูล | | | | | | | | |
| 3.2 ทำการแปลงข้อมูลให้อยู่ในรูปแบบที่ใช้งานง่าย | | | | | | | | |
| 3.3 พรอตกราฟดูฤดูกาลของข้อมูล | | | | | | | | |
| 3.4 พรอตกราฟดูแนวโน้มของข้อมูล | | | | | | | | |
| 3.5 หาความสัมพันธ์ของข้อมูล | | | | | | | | |
| **4. ปรับใช้กับแบบจำลอง** | | | | | | | | |
| 4.1 ปรับพารามิเตอร์ในแบบจำลอง | | | | | | | | |
| 4.2 ใช้แบบจำลอง ARIMA | | | | | | | | |
| 4.3 ใช้แบบจำลอง SARIMA | | | | | | | | |
| 4.4 ใช้แบบจำลอง VAR | | | | | | | | |
| 4.5 ใช้แบบจำลอง SES | | | | | | | | |
| 4.6 ใช้แบบจำลอง HWS | | | | | | | | |
| 4.7 ใช้แบบจำลอง CNN | | | | | | | | |
| 4.8 ใช้แบบจำลอง LSTM | | | | | | | | |
| **5. ทดสอบความแม่นยำของแบบจำลอง** | | | | | | | | |
| 5.1 การแบ่งข้อมูลเพื่อนำทดสอบประสิทธิภาพของแบบจำลองด้วยวิธี Cross-validation | | | | | | | | |
| 5.2 การเปรียบเทียบความแม่นยำของการพยากรณ์ด้วย RMSE | | | | | | | | |
| 5.3 ทดสอบสมมติฐานว่าแบบจำลอง | | | | | | | | |
| **6. สรุปและเผยแพร่งานวิจัย** | | | | | | | | |

*ตาราง 1 แผนการดำเนินโครงการวิจัย*

## 3.3 อุปกรณ์และเครื่องมือที่ใช้

3.3.1 ฮาร์ดแวร์

- ระบบปฏิบัติการ: Windows 10 (64-bit)
- หน่วยประมวลผล: Intel Core i7-1065G7
- หน่วยความจำ: RAM 16 GB

3.3.2 ซอฟต์แวร์

- Anaconda Navigator (anaconda3)
- Visual Studio Code
- Postman
- Colaboratory
- Flask

## 3.4 ชุดข้อมูล (Dataset)

3.4.1 ข้อมูลราคาทองคำตั้งแต่ปี ค.ศ.1970 – 2020

3.4.2 ข้อมูลราคาทองแดงตั้งแต่ปี ค.ศ.1959 – 2020

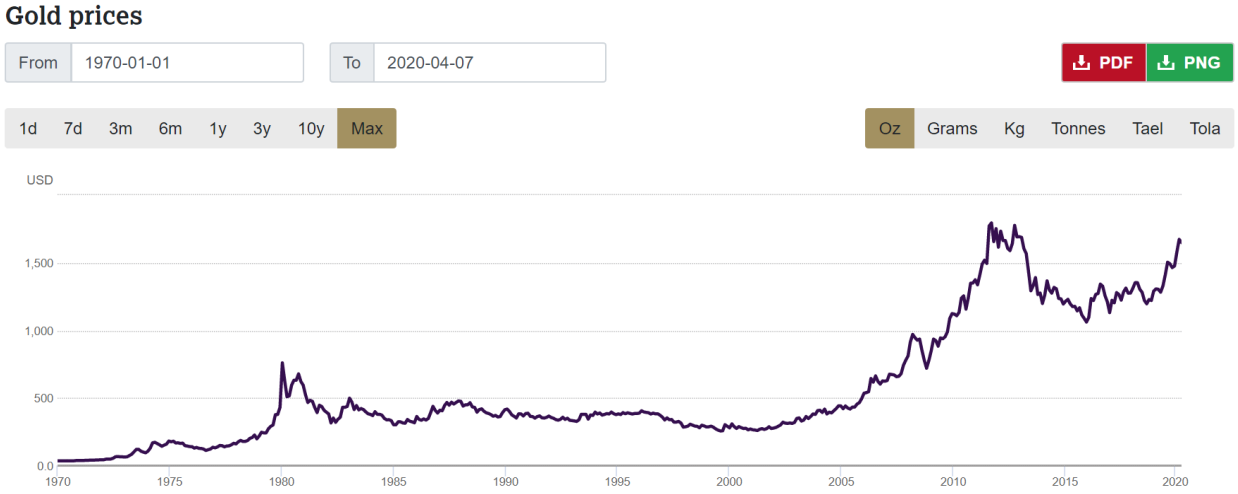
3.4.3 ข้อมูลราคาก๊าซธรรมชาติตั้งแต่ปี ค.ศ.1976- 2012

3.4.4 ข้อมูลราคาน้ำมันดิบ : เบรนต์ - ยุโรป ตั้งแต่ปี ค.ศ.1987 – 2020

3.4.5 ข้อมูลราคาน้ำมันดิบ : (WTI) ตั้งแต่ปี ค.ศ.1986 – 2020

3.4.6 ข้อมูลราคาแก๊สโซลีน ตั้งแต่ปี ค.ศ.1993 – 2020

3.4.7 ข้อมูลราคาแพลทินัม ตั้งแต่ปี ค.ศ.1969 – 2020

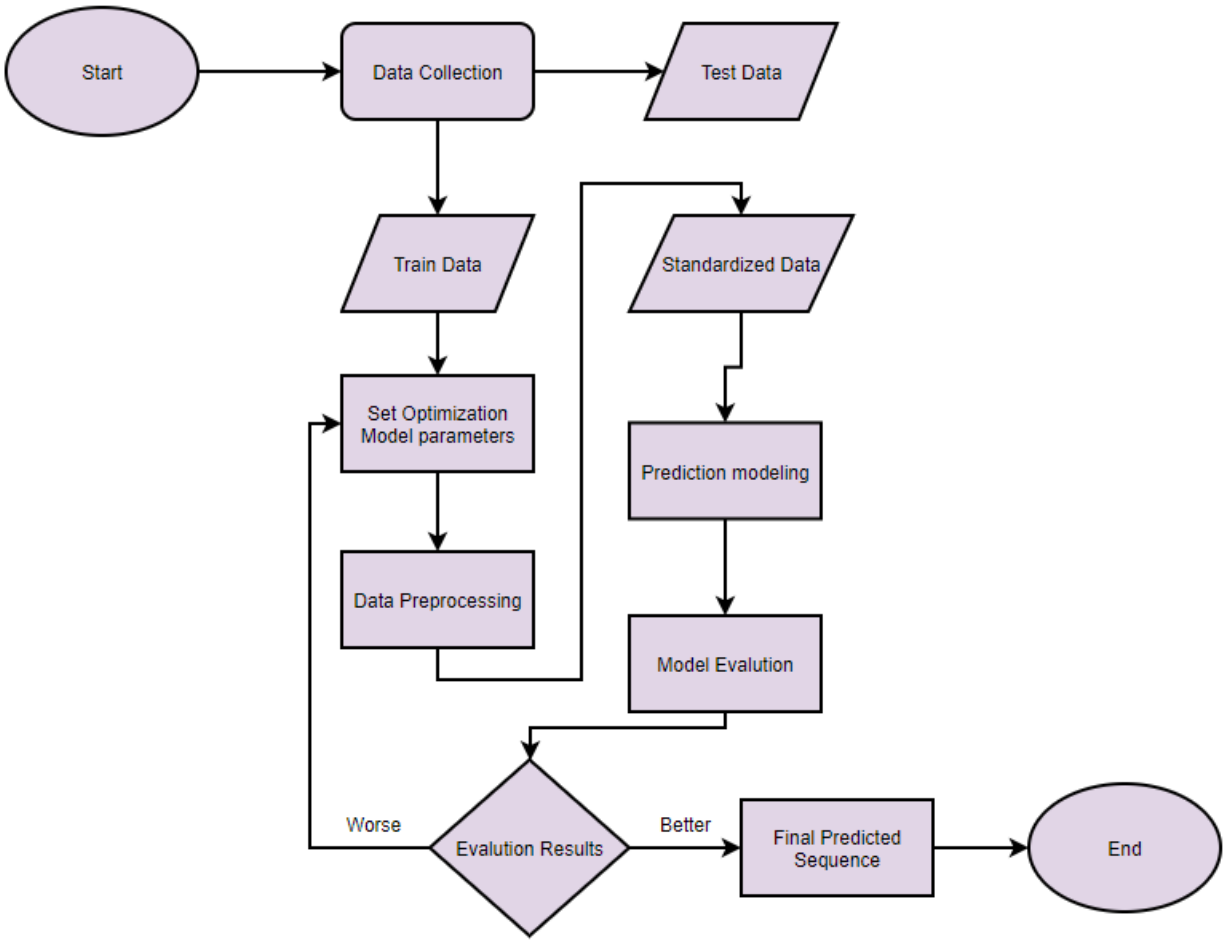


*รูป 20* ตัวอย่างข้อมูลราคาทองคำ

ที่มา : [ออนไลน์] https://www.gold.org/goldhub/data/gold-prices

## 3.5 ขั้นตอนการออกแบบระบบ

- ภาพรวมการทำงานของระบบ




*รูป 21 แสดงการทำงานของระบบ*

3.5.1 ขั้นตอนการออกแบบระบบมี 6 ขั้นตอน ดังนี้

1. การเก็บรวบรวมข้อมูล โดยใช้ราคาที่เกี่ยวข้องกับราคาทองคำ
2. ทำการแบ่งข้อมูล Cross Validation โดยใช้ข้อมูล Train ในแบบจำลองเพื่อทำนายราคาในอนาคต และข้อมูล Test ไว้ใช้สำหรับตรวจสอบประสิทธิภาพของพารามิเตอร์และแบบจำลอง
3. ทำการหาค่าพารามิเตอร์ที่ดีที่สุดของแต่ละแบบจำลองด้วยวิธีที่แตกต่างกัน
4. ลองใช้พารามิเตอร์หลากหลายแบบกับแบบจำลองเพื่อทำนายราคาในอนาคต
5. นำค่าพารามิเตอร์ที่ดีที่สุดที่หาได้มาใช้ทำนายกับแบบจำลอง
6. ประเมินแบบจำลองว่าสามารถทำนายราคาทองได้ข้อมูลที่เป็นมาตราฐานหรือไม่ โดยจะทำการตรวจสอบแบบจำลองว่าสามารถทำนายราคาทองได้ดีกว่าราคาปัจจุบันหรือไม่
7. เมื่อได้แบบจำลองที่สามารถทำนายได้แล้ว ก็จะทำการนำแบบจำลองและพารามิเตอร์ที่หาได้ไปใช้ในการทำนายราคาโดยไม่แบ่ง Cross Validation เพื่อให้ได้ผลการทำนายราคาถัดไป 1 สัปดาห์

3.5.2 ขั้นตอนการทำงานของระบบ

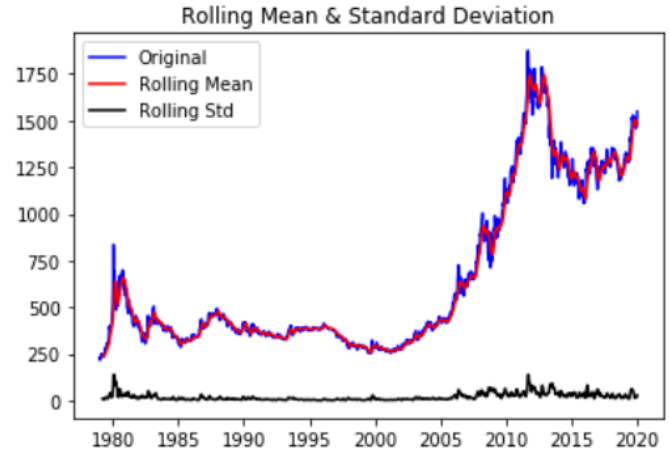
1.เตรียมข้อมูลให้พร้อมกับการใช้งานในแบบจำลอง

1.1 ทำความสะอาดข้อมูล (Clean data)

- ข้อมูลราคาต้องนำมาทำความสะอาดข้อมูลก่อน ตรวจสอบและการแก้ไขรายการข้อมูลที่ไม่ถูกต้อง ลบค่าที่ที่ซ้ำซ้อน
- drop rows ที่ไม่นำมาใช้ในแบบจำลองและทำการเลือกคอลัมน์ที่เป็นวันที่และราคา(USD) มาใช้ในแบบจำลองต่อไป
- ข้อมูลวันที่ให้เปลี่ยนเป็นชื่อ Weekly และข้อมูลราคาจะเปลี่ยนเป็นชื่อ Prices
- เปลี่ยน Weekly ให้เป็น datetime และเลือกให้เป็น index
- ในคอลัมน์ของ Prices หากช่องในมีเครื่องหมาย “,” ให้แทนที่ด้วย “”
- ตัวข้อมูลเป็นรายสัปดาห์ แต่แต่ละสัปดาห์เริ่มวันที่ไม่เหมือนกัน จึงต้องปรับให้เป็นวันศุกร์ให้เหมือนกันทั้งหมด

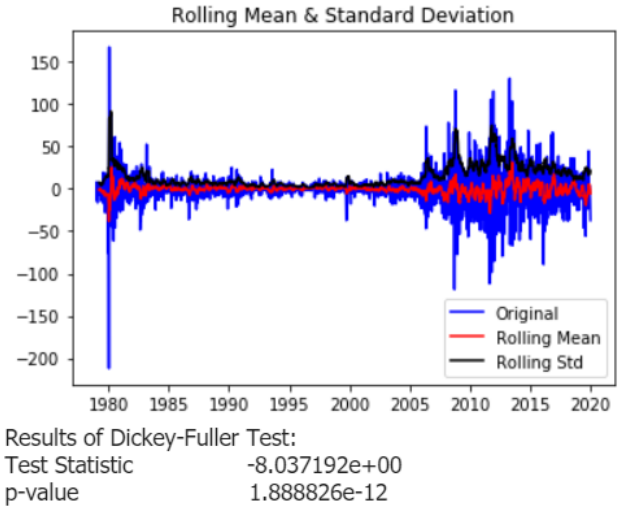
1.2 ทำให้ข้อมูลราคามีความนิ่ง (Stationary)

- ทำการตรวจสอบข้อมูลราคาว่ามีความนิ่งของข้อมูล(Stationary)หรือไม่ โดยทดลอง plot ข้อมูลเพื่อวิเคราะห์ข้อมูลใช้ Dickey Fuller แล้วดูได้จากค่า p - value ถ้าค่า p-value < 0.05 คือ ข้อมูลมีความนิ่ง(Stationary)




*รูป 22 ตรวจสอบความนิ่งของข้อมูลด้วย Dickey Fuller*

- หากข้อมูลไม่นิ่ง ต้องทำการ Differencing หรือการหาความแตกต่างระหว่างค่าอดีตกับค่าปัจจุบัน และดูการเปลี่ยนแปลงของข้อมูล
- การ Differencing จะทำการขจัดแนวโน้มและฤดูกาล เมื่อทำเสร็จจึงทำการตรวจสอบค่า ซึ่งจะทำให้เราได้ข้อมูลมีความนิ่ง(Stationary)ออกมา สังเกตได้จากค่า p-value มีค่าน้อยกว่า 0.05

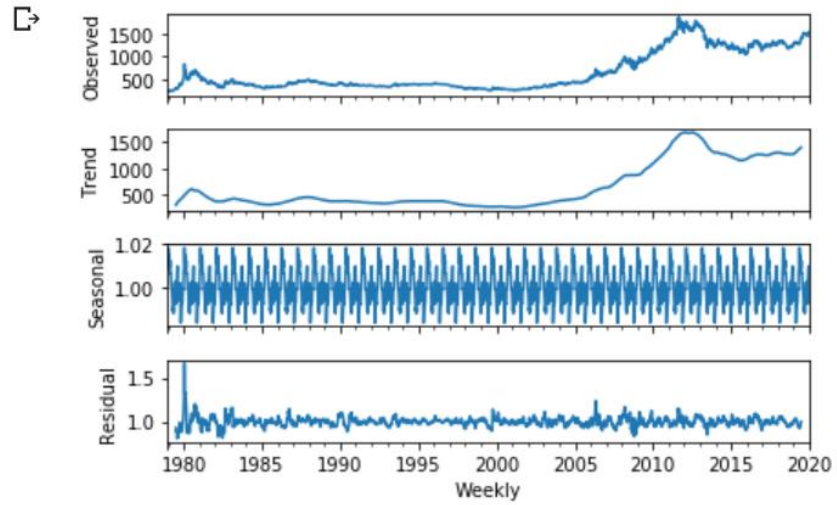



*รูป 23 ข้อมูลราคาทองหลังจากทำการ Differencing*

2.วิเคราะห์ข้อมูล

2.1 ทำความเข้าใจข้อมูล

- แบ่งย่อยข้อมูลอนุกรมเวลา(Decompose) โดยแบ่งออกมาเป็นแนวโน้ม(Trend),ฤดูกาล (Seasonal) และข้อมูลที่ไม่มีแนวโน้มและฤดูกาล(Residual) เพื่อทำการใช้วิเคราะห์ข้อมูลได้ตามส่วนย่อยๆที่ต้องการ

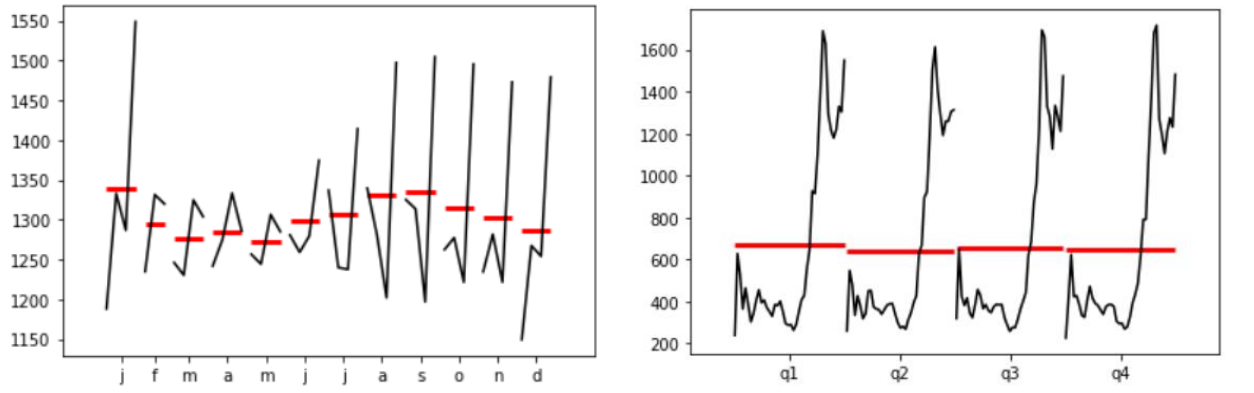



*รูป 24* การแบ่งย่อยข้อมูลอนุกรมเวลา

- ทำการพรอตฤดูกาลของราคาทองคำในแต่ละปี เพื่อดูลักษณะฤดูกาลของข้อมูลโดยแบ่งเป็นรายเดือนและรายไตรมาส

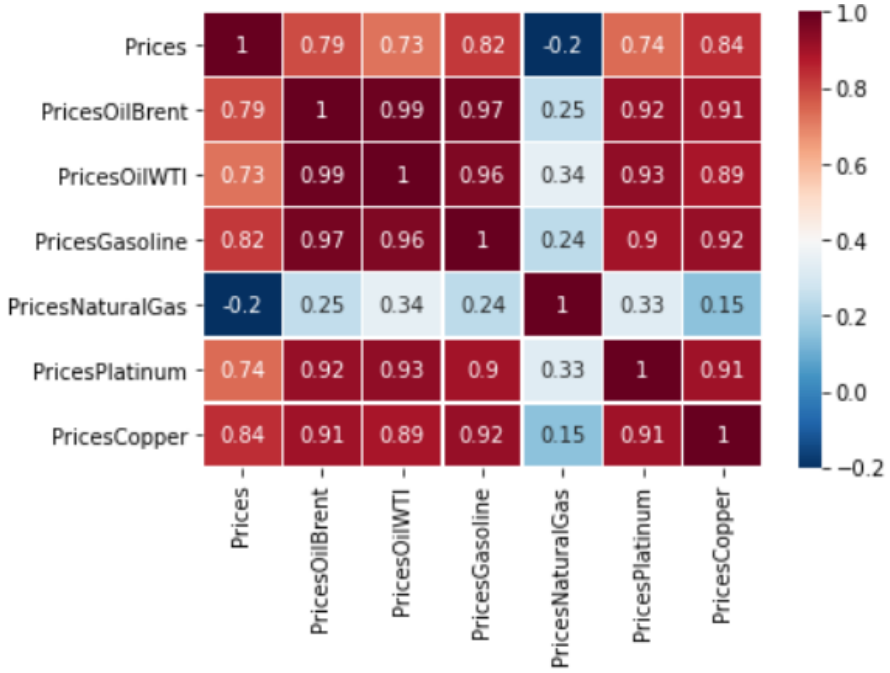
```
tsaplots.month_plot(ts.tail(200).resample(rule='M').mean())
```

```
tsaplots.quarter_plot(ts.resample(rule='Q').mean())
```




*รูป 25* ฤดูกาลของราคาทองคำ

- นำข้อมูลราคาทั้งหมดมาหาค่าความสัมพันธ์ จากรูปที่ 26 แก๊สธรรมชาติไม่มีความสัมพันธ์กับราคาทองดูจากมีค่าความสัมพันธ์น้อยกว่า 0.7 จึงจะไม่นำมาใช้ในแบบจำลองชนิดหลายตัวแปร




*รูป 26* ความสัมพันธ์ของราคาทองกับราคาอื่นๆ

3.ปรับใช้กับแบบจำลอง

3.1 ARIMA (Autoregressive Integrated Moving average)

- ข้อมูลราคาที่จะนำมาใช้ในแบบจำลอง ARIMA จะต้องมีการตรวจสอบความนิ่งของข้อมูลก่อน คือไม่มีแนวโน้มเชิงเส้นและไม่มีฤดูกาล สามารถตรวจสอบได้โดยใช้ Dickey Fuller หากแบบจำลองยังไม่มีความนิ่งของข้อมูล ก็จะต้องทำการ Differencing ก่อนเพื่อให้ได้ข้อมูลที่นำมาใช้ในแบบจำลองนี้ได้
- สำหรับวิธีการเลือก Parameter ของแบบจำลอง ARIMA สามารถใช้คำสั่ง auto.arima เลือกค่าพารามิเตอร์หรือใช้วิธีดูจาก ACF/PACF Plot ก็ได้ หลังจากลองใช้คำสั่ง auto.arima ต้องลองปรับค่า p กับ q ประมาณ +-2 ดูด้วย หลังจากนั้นนำมาเทียบค่า RMSE เพื่อหาค่าพารามิเตอร์ที่ดีที่สุด
- ทำการ Plot ค่า ACF/PACF เพื่อดูค่าพารามิเตอร์ โดยต้องลองปรับค่า p กับ q ประมาณ +-2 ดูด้วย หลังจากนั้นนำมาเทียบค่า RMSE เพื่อหาค่าพารามิเตอร์ที่ดีที่สุด

```
fit_2 = auto_arima(ts,start_p=0,start_q=0,max_p=5,max_q=5,seasonal=True,trace=True,m=25)
fit_2.summary()
```

```
Performing stepwise search to minimize aic
Fit ARIMA: (0, 1, 0)x(1, 0, 1, 25) (constant=True); AIC=18985.153, BIC=19007.846, Time=5.288 seconds
Fit ARIMA: (0, 1, 0)x(0, 0, 0, 25) (constant=True); AIC=19000.691, BIC=19012.037, Time=0.117 seconds
Fit ARIMA: (1, 1, 0)x(1, 0, 0, 25) (constant=True); AIC=18980.985, BIC=19003.677, Time=4.394 seconds
Fit ARIMA: (0, 1, 1)x(0, 0, 1, 25) (constant=True); AIC=18980.779, BIC=19003.472, Time=4.436 seconds
Fit ARIMA: (0, 1, 0)x(0, 0, 0, 25) (constant=False); AIC=19000.755, BIC=19006.428, Time=0.079 seconds
Fit ARIMA: (0, 1, 1)x(0, 0, 0, 25) (constant=True); AIC=18998.546, BIC=19015.566, Time=0.651 seconds
Fit ARIMA: (0, 1, 1)x(1, 0, 1, 25) (constant=True); AIC=18982.779, BIC=19011.145, Time=8.262 seconds
Fit ARIMA: (0, 1, 1)x(0, 0, 2, 25) (constant=True); AIC=18982.779, BIC=19011.145, Time=21.767 seconds
Fit ARIMA: (0, 1, 1)x(1, 0, 0, 25) (constant=True); AIC=18981.149, BIC=19003.842, Time=5.929 seconds
Fit ARIMA: (0, 1, 1)x(1, 0, 2, 25) (constant=True); AIC=18983.968, BIC=19018.007, Time=68.107 seconds
Fit ARIMA: (0, 1, 0)x(0, 0, 1, 25) (constant=True); AIC=18983.164, BIC=19000.183, Time=4.339 seconds
Fit ARIMA: (1, 1, 1)x(0, 0, 1, 25) (constant=True); AIC=18982.254, BIC=19010.620, Time=9.409 seconds
Fit ARIMA: (0, 1, 2)x(0, 0, 1, 25) (constant=True); AIC=18981.815, BIC=19010.182, Time=5.939 seconds
Fit ARIMA: (1, 1, 0)x(0, 0, 1, 25) (constant=True); AIC=18980.612, BIC=19003.304, Time=4.523 seconds
Fit ARIMA: (1, 1, 0)x(0, 0, 0, 25) (constant=True); AIC=18998.414, BIC=19015.433, Time=0.221 seconds
Fit ARIMA: (1, 1, 0)x(1, 0, 1, 25) (constant=True); AIC=18982.612, BIC=19010.978, Time=8.957 seconds
Fit ARIMA: (1, 1, 0)x(0, 0, 2, 25) (constant=True); AIC=18982.611, BIC=19010.978, Time=18.093 seconds
Fit ARIMA: (1, 1, 0)x(1, 0, 2, 25) (constant=True); AIC=18983.810, BIC=19017.850, Time=63.551 seconds
Fit ARIMA: (2, 1, 0)x(0, 0, 1, 25) (constant=True); AIC=18981.959, BIC=19010.325, Time=5.038 seconds
```

*รูป 27* ใช้ auto.arima ในการหาค่าพารามิเตอร์

- เริ่มต้นทำการปรับพารามิเตอร์ของแบบจำลอง ARIMA กำหนดกรอบการสุ่มเบื้องต้นของพารามิเตอร์ (p,d,q) ไว้ที่ [1,2,3] เป็นค่า range จากการทำการ plot ในขั้นตอนก่อนหน้าแล้วใช้ algorithm ของ Grid Search Optimization มาใช้เพื่อหาค่าที่เหมาะสม
- นอกจากค่า Order (p,d,q) สามารถปรับ Trend เป็นแบบ c,nc และปรับ Method เป็นแบบ css-mle,css,mle ได้ เพื่อหาค่าพารามิเตอร์ที่ดีกับแบบจำลอง
- จากการทำนายแบบจำลองนี้ จะทำการแบ่ง Cross validation เพื่อทดสอบประสิทธิภาพของแบบจำลองโดยใช้วิธีแบ่งแบบ Timeseriessplit
- โดยข้อมูล Train จะอยู่ช่วงวันที่ 1978/12/29 – 2020/01/31 และข้อมูล Test จะอยู่ช่วงวันที่ 2020/02/07-2020/04/10
- เพื่อดูผลในการ Fit ARIMA model จะนำผลลัพธ์จากการสุ่มเลขพารามิเตอร์ในแต่ละรอบมาทำการหาค่าของ RMSE ที่ค่าน้อยที่สุดของแต่ละพารามิเตอร์ที่ประมวลผลได้ โดยเทียบข้อมูลที่ทำการ Prediction กับข้อมูล Testing Data ที่เตรียมไว้ หลักสำคัญคือพยายามหา RMSE ที่น้อย เพราะสะท้อนถึงประสิทธิภาพของพารามิเตอร์ เมื่อเสร็จขั้นตอนนี้แล้วจะได้พารามิเตอร์ที่ดีที่สุด
- เมื่อได้พารามิเตอร์ที่ดีที่สุดแล้ว จะนำมาใช้ในแบบจำลองเพื่อทำนายราคาและเมื่อทำนายเสร็จแล้วต้องนำค่าที่ทำนายได้คืนค่าเดิมด้วย เนื่องจากมีการทำ Differencing ราคาในตอนแรก

- จะทำการตรวจประสิทธิภาพของแบบจำลอง ARIMA
- มาถึงขั้นตอนนี้ จะได้แบบจำลอง ARIMA ที่พร้อมทำงานกับข้อมูลแล้ว จากนั้นจะใช้ ARIMA ที่ได้พัฒนาขึ้นทำการ Fit() และ Prediction() โดยไม่แบ่ง Cross validation เพื่อหาผลการทำนายราคาทองที่จะเกิดขึ้นในอีก 1 สัปดาห์
- หากแบบจำลองมีประสิทธิภาพนำผลลัพธ์ของราคาทองที่ทำนายได้ และวันที่ทำนายมาแสดง API และ Plot กราฟเปรียบเทียบระหว่างข้อมูลในอดีตกับราคาที่ทำนายออกมา

```
tscv = TimeSeriesSplit(n_splits = 200)
rmse = []

for train_index, test_index in tscv.split(ts):
  cv_train, cv_test = ts.iloc[train_index], ts.iloc[test_index].head(1)

  model_ARIMA_tscv = ARIMA(cv_train, order=(2, 1, 1))
  results = model_ARIMA_tscv.fit(trend  = 'c' , method='css-mle')
  predictions = results.predict(cv_test.index.values[0], cv_test.index.values[-1],typ="levels")
  rmse_result = sqrt(mean_squared_error(cv_test.values, predictions))
  print("predictions: {}".format(predictions))
  print("test: {}".format(cv_test.values))
  print("RMSE: {}".format(rmse_result))
  print("---------------------------------------")

  rmse.append(rmse_result)
print("RMSE: {}".format(np.mean(rmse)))
```

*รูป 28* ตัวอย่างการปรับค่าพารามิเตอร์ของแบบจำลอง ARIMA

### 3.2 SES (Single Exponential Smoothing)

- SES จะให้ความสำคัญกับข้อมูลเก่าทุกค่า แต่ให้ความสำคัญกับค่าที่ใกล้ปัจจุบันมากที่สุด ยิ่งข้อมูลในอดีตจะยิ่งมีผลต่อการทำนายน้อยลง ค่าพยากรณ์จะตอบสนองกับข้อมูลใหม่เป็นหลัก
- การทำนายในแบบจำลองนี้จะไม่สามารถปรับพารามิเตอร์ได้ จะนำข้อมูลราคาทองมาใช้ในแบบจำลองนี้แล้วทำการหาค่า RMSE และผลการทำนายราคาทอง
- การทำนายแบบจำลองนี้ จะทำการแบ่ง Cross validation เพื่อทดสอบประสิทธิภาพของแบบจำลองโดยใช้วิธีแบ่งแบบ Timeseriessplit
- โดยข้อมูล Train จะอยู่ช่วงวันที่ 1978/12/29 – 2020/01/31 และข้อมูล Test จะอยู่ช่วงวันที่ 2020/02/07-2020/04/10

- จากการทำนายราคาทองโดยแบบจำลอง SES จะสามารถตรวจสอบประสิทธิภาพของแบบจำลองนี้ได้จากการเปรียบเทียบราคาทองในปัจจุบันและราคาทองที่ทำนายได้จากแบบจำลอง ว่าค่าไหนใกล้เคียงกับราคาทองในอนาคตมากกว่ากัน ซึ่งจะแบ่งข้อมูลโดยใช้ Cross Validation แบบ Timeseriessplit แบ่งออกเป็น 10 ชุด และในแต่ละชุดจะวัดว่าแบบจำลองสามารถทำนายราคาทองในอนาคตได้ดีหรือไม่ ถ้าหากแบบจำลองสามารถทำนายได้ดีกว่าครึ่งนึงจะถือว่าแบบจำลองนี้ มีประสิทธิภาพหรือสามารถทำนายราคาทองได้
- หากจำลองมีประสิทธิภาพนำผลลัพธ์ของราคาทองที่ทำนายได้ และวันที่ทำนายมาแสดง API และ Plot กราฟเปรียบเทียบระหว่างข้อมูลในอดีตกับราคาที่ทำนายออกมา

```
tscv = TimeSeriesSplit(n_splits = 200)
rmse = []

for train_index, test_index in tscv.split(ts):
  cv_train, cv_test = ts.iloc[train_index], ts.iloc[test_index]

  model = SimpleExpSmoothing(data)
  model_fit = model.fit()

  predictions = model_fit.predict(cv_test.index.values[0], cv_test.index.values[-1])
  rmse_result = sqrt(mean_squared_error(cv_test.values, predictions))
  rmse.append(rmse_result)

print("RMSE: {}".format(np.mean(rmse)))
```

```
RMSE: 14.191506565014084
```

*รูป 29* ตัวอย่างการปรับค่าพารามิเตอร์ของแบบจำลอง SES

3.3 HWS (Holt-Winters exponential smoothing method)

- แบบจำลอง HWS จะนำข้อมูลในอดีตมาวิเคราะห์เพื่อกำหนดตัวแบบจำลองและนำแบบจำลองนั้นมาทำนายค่าราคาในอนาคต วิธีการนี้เหมาะกับข้อมูลที่มีแนวโน้มเชิงเส้นและมีอิทธิพลของฤดูกาล
- เริ่มต้นทำการปรับพารามิเตอร์ของแบบจำลอง HWS โดยสามารถปรับ Trend และ Seasonal เพื่อใช้หาค่าที่เหมาะสมได้ โดย Trend มีแบบ mul,add และ Seasonal เป็นแบบ mul,add ซึ่งแบบจำลองสามารถใส่ Trend ค่าเดียว , Seasonal ค่าเดียว และสามารถใส่ได้พร้อมกันทั้ง Trend และ Seasonal ได้

- การทำนายแบบจำลองนี้ จะทำการแบ่ง Cross validation เพื่อทดสอบประสิทธิภาพของแบบจำลองโดยใช้วิธีแบ่งแบบ Timeseriessplit
- โดยข้อมูล Train จะอยู่ช่วงวันที่ 1978/12/29 – 2020/01/31 และข้อมูล Test จะอยู่ช่วงวันที่ 2020/02/07-2020/04/10
- เพื่อดูผลในการ Fit HWS model จะนำผลลัพธ์จากการสุ่มพารามิเตอร์ในแต่ละรอบมาทำการหาค่าของ RMSE ที่ค่าน้อยที่สุดของแต่ละพารามิเตอร์ที่ประมวลผลได้ โดยเทียบข้อมูลที่ทำการ Prediction กับข้อมูล Testing Data ที่เตรียมไว้ เมื่อเสร็จขั้นตอนนี้แล้วจะได้พารามิเตอร์ที่ดีที่สุด
- เมื่อได้พารามิเตอร์ที่ดีที่สุดแล้ว จะทำการตรวจประสิทธิภาพของแบบจำลอง HWS
- หากจำลองมีประสิทธิภาพนำผลลัพธ์ของราคาทองที่ทำนายได้ และวันที่ทำนายมาแสดง API และ Plot กราฟเปรียบเทียบระหว่างข้อมูลในอดีตกับราคาที่ทำนายออกมา

```
tscv = TimeSeriesSplit(n_splits = 200)
rmse = []

for train_index, test_index in tscv.split(ts):
  cv_train, cv_test = ts.iloc[train_index], ts.iloc[test_index]

  model_ETS = ExponentialSmoothing(ts,trend='mul')
  results = model_ETS.fit()
  predictions = results.predict(cv_test.index.values[0], cv_test.index.values[-1])

  rmse_result = sqrt(mean_squared_error(cv_test.values, predictions))
  rmse.append(rmse_result)

print("RMSE: {}".format(np.mean(rmse)))
```

```
RMSE: 14.18819507694242
```

*รูป 30* ตัวอย่างการปรับค่าพารามิเตอร์ของแบบจำลอง SES

3.1 CNN 1 input (Convolutional Neural Network)

- แบบจำลอง CNN 1 input สามารถปรับพารามิเตอร์ได้ โดยการกำหนดจำนวนวันที่ทำนาย จำนวนข้อมูลราคาและจำนวนพารามิเตอร์ LSTM
- แบบจำลองนี้ จะทำการแบ่ง Cross validation เพื่อทดสอบประสิทธิภาพของแบบจำลอง

- โดยข้อมูล Train จะอยู่ช่วงวันที่ 1978/12/29 – 2020/03/20 และข้อมูล Test จะอยู่ช่วงวันที่ 2020/03/27-2020/04/10
- ทำการ Scale ค่าข้อมูล ให้ข้อมูลมีค่าระหว่างค่าต่ำสุดและสูงสุด เพื่อรักษารูปแบบการกระจายตัวของข้อมูลให้เหมือนกัน โดยใช้ Library MinMaxScaler
- นำข้อมูลไปใส่ในแบบจำลองแล้วทำการ fit แบบจำลอง เพื่อให้แบบจำลองเรียนรู้ข้อมูล โดยใช้จำนวน epoch ที่มากพอในที่นี้คือ 2000 ให้โมเดลสามารถทำนายได้ใกล้เคียงมากที่สุด
- เมื่อแบบจำลองเรียนรู้ข้อมูลแล้ว จะนำมาใช้ในแบบจำลองเพื่อทำนายราคาและเมื่อทำนายเสร็จแล้วต้องนำค่าที่ทำนายได้คืนค่าเดิมด้วย เนื่องจากมีการทำ Scale ราคาในตอนแรก
- ตรวจสอบประสิทธิภาพของแบบจำลอง หากแบบจำลองมีประสิทธิภาพนำผลลัพธ์ของราคาทองที่ทำนายได้ และวันที่ทำนายมาแสดง API และ Plot กราฟเปรียบเทียบระหว่างข้อมูลในอดีตกับราคาที่ทำนายออกมา

3.2 CNN Allinput (Convolutional Neural Network)

- แบบจำลอง CNN allinput สามารถนำปัจจัยราคาอื่นที่มีความเกี่ยวข้องกับราคาทองคำมาใช้ในแบบจำลองด้วยได้ เพื่อเพิ่มประสิทธิภาพของการทำนาย
- แบบจำลองสามารถปรับพารามิเตอร์ได้ โดยการกำหนดจำนวนวันที่ทำนาย จำนวนข้อมูลราคาที่เกี่ยวข้องและจำนวนพารามิเตอร์ LSTM
- แบบจำลองนี้ จะทำการแบ่ง Cross validation เพื่อทดสอบประสิทธิภาพของแบบจำลอง
- โดยข้อมูล Train จะอยู่ช่วงวันที่ 1993/04/09 – 2020/04/03 และข้อมูล Test จะอยู่ช่วงวันที่ 2020/04/10
- ทำการ Scale ค่าข้อมูล ให้ข้อมูลมีค่าระหว่างค่าต่ำสุดและสูงสุด เพื่อรักษารูปแบบการกระจายตัวของข้อมูลให้เหมือนกัน โดยใช้ Library MinMaxScaler
- นำข้อมูลไปใส่ในแบบจำลองแล้วทำการ fit แบบจำลอง เพื่อให้แบบจำลองเรียนรู้ข้อมูล โดยใช้จำนวน epoch ที่มากพอในที่นี้คือ 2000 ให้โมเดลสามารถทำนายได้ใกล้เคียงมากที่สุด
- เมื่อแบบจำลองเรียนรู้ข้อมูลแล้ว จะนำมาใช้ในแบบจำลองเพื่อทำนายราคาและเมื่อทำนายเสร็จแล้วต้องนำค่าที่ทำนายได้คืนค่าเดิมด้วย เนื่องจากมีการทำ Scale ราคาในตอนแรก

- ตรวจสอบประสิทธิภาพของแบบจำลอง หากแบบจำลองมีประสิทธิภาพนำผลลัพธ์ของราคาทองที่ทำนายได้ และวันที่ทำนายมาแสดง API และ Plot กราฟเปรียบเทียบระหว่างข้อมูลในอดีตกับราคาที่ทำนายออกมา

3.3 LSTM 1 input (Long short-term memory)

- แบบจำลอง LSTM 1 input สามารถปรับพารามิเตอร์ได้ โดยการกำหนดจำนวนวันที่ทำนาย จำนวนข้อมูลราคาและจำนวนพารามิเตอร์ LSTM
- แบบจำลองนี้ จะทำการแบ่ง Cross validation เพื่อทดสอบประสิทธิภาพของแบบจำลอง
- โดยข้อมูล Train จะอยู่ช่วงวันที่ 1978/12/29 – 2020/03/20 และข้อมูล Test จะอยู่ช่วงวันที่ 2020/03/27-2020/04/10
- ทำการ Scale ค่าข้อมูล ให้ข้อมูลมีค่าระหว่างค่าต่ำสุดและสูงสุด เพื่อรักษารูปแบบการกระจายตัวของข้อมูลให้เหมือนกัน โดยใช้ Library MinMaxScaler
- นำข้อมูลไปใส่ในแบบจำลองแล้วทำการ fit แบบจำลอง เพื่อให้แบบจำลองเรียนรู้ข้อมูล โดยใช้จำนวน epoch ที่มากพอในที่นี้คือ 2000 ให้โมเดลสามารถทำนายได้ใกล้เคียงมากที่สุด
- เมื่อแบบจำลองเรียนรู้ข้อมูลแล้ว จะนำมาใช้ในแบบจำลองเพื่อทำนายราคาและเมื่อทำนายเสร็จแล้วต้องนำค่าที่ทำนายได้คืนค่าเดิมด้วย เนื่องจากมีการทำ Scale ราคาในตอนแรก
- ตรวจสอบประสิทธิภาพของแบบจำลอง หากแบบจำลองมีประสิทธิภาพนำผลลัพธ์ของราคาทองที่ทำนายได้ และวันที่ทำนายมาแสดง API และ Plot กราฟเปรียบเทียบระหว่างข้อมูลในอดีตกับราคาที่ทำนายออกมา

3.4 LSTM Allinput

- แบบจำลอง LSTM Allinput สามารถนำปัจจัยราคาอื่นที่มีความเกี่ยวข้องกับราคาทองคำมาใช้ในแบบจำลองด้วยได้ เพื่อเพิ่มประสิทธิภาพของการทำนาย
- แบบจำลองสามารถปรับพารามิเตอร์ได้ โดยการกำหนดจำนวนวันที่ทำนาย จำนวนข้อมูลราคาที่เกี่ยวข้องและจำนวนพารามิเตอร์ LSTM
- แบบจำลองนี้ จะทำการแบ่ง Cross validation เพื่อทดสอบประสิทธิภาพของแบบจำลอง
- โดยข้อมูล Train จะอยู่ช่วงวันที่ 1993/04/09 – 2020/04/03 และข้อมูล Test จะอยู่ช่วงวันที่ 2020/04/10

- ทำการ Scale ค่าข้อมูล ให้ข้อมูลมีค่าระหว่างค่าต่ำสุดและสูงสุด เพื่อรักษารูปแบบการกระจายตัวของข้อมูลให้เหมือนกัน โดยใช้ Library MinMaxScaler
- นำข้อมูลไปใส่ในแบบจำลองแล้วทำการ fit แบบจำลอง เพื่อให้แบบจำลองเรียนรู้ข้อมูล โดยใช้จำนวน epoch ที่มากพอในที่นี้คือ 2000 ให้โมเดลสามารถทำนายได้ใกล้เคียงมากที่สุด
- เมื่อแบบจำลองเรียนรู้ข้อมูลแล้ว จะนำมาใช้ในแบบจำลองเพื่อทำนายราคาและเมื่อทำนายเสร็จแล้วต้องนำค่าที่ทำนายได้คืนค่าเดิมด้วย เนื่องจากมีการทำ Scale ราคาในตอนแรก
- ตรวจสอบประสิทธิภาพของแบบจำลอง หากแบบจำลองมีประสิทธิภาพนำผลลัพธ์ของราคาทองที่ทำนายได้ และวันที่ทำนายมาแสดง API และ Plot กราฟเปรียบเทียบระหว่างข้อมูลในอดีตกับราคาที่ทำนายออกมา

3.5 VAR Allinput (Vector Autoregressive)

- แบบจำลอง VAR Allinput สามารถนำปัจจัยราคาอื่นที่มีความเกี่ยวข้องกับราคาทองคำมาใช้ในแบบจำลองด้วยได้ เพื่อเพิ่มประสิทธิภาพของการทำนาย แต่ข้อมูลที่จะใช้ต้องเป็นข้อมูลที่มีความนิ่งของข้อมูล คือไม่มีแนวโน้มและฤดูกาล หากมีจะต้องทำการ Differencing ก่อน
- แบบจำลองนี้ จะทำการแบ่ง Cross validation เพื่อทดสอบประสิทธิภาพของแบบจำลอง
- โดยข้อมูล Train จะอยู่ช่วงวันที่ 1993/04/09 – 2020/04/03 และข้อมูล Test จะอยู่ช่วงวันที่ 2020/04/10
- พารามิเตอร์ที่ดีที่สุดสามารถหาได้จากเลข Order p ที่ได้ค่า AIC ต่ำที่สุด
- เมื่อได้พารามิเตอร์ที่ดีที่สุดแล้ว นำพารามิเตอร์มาใช้ในแบบจำลองเพื่อทำนายราคาและเมื่อทำนายเสร็จแล้วต้องนำค่าที่ทำนายได้คืนค่าเดิมด้วย เนื่องจากมีการทำ Differencing
- ทำการตรวจประสิทธิภาพของแบบจำลอง VAR Allinput หากแบบจำลองมีประสิทธิภาพนำผลลัพธ์ของราคาทองที่ทำนายได้ และวันที่ทำนายมาแสดง API และ Plot กราฟเปรียบเทียบระหว่างข้อมูลในอดีตกับราคาที่ทำนายออกมา
- ทดสอบความแม่นยำของแบบจำลอง

4.ทดสอบความแม่นยำของแบบจำลอง

4.1 การเปรียบเทียบความแม่นยำของการพยากรณ์ด้วย RMSE

ในการหาค่าพารามิเตอร์ที่ดีที่สุดของแต่ละแบบจำลอง จะทำการหาค่าของ RMSE ที่ค่าน้อยที่สุดของแต่ละพารามิเตอร์ที่ประมวลผลได้ โดยเทียบข้อมูลที่ทำการ Prediction กับข้อมูล Testing Data ที่เตรียมไว้ หลักสำคัญคือพยายามหา RMSE ที่น้อยเพราะสะท้อนถึงประสิทธิภาพของพารามิเตอร์ เมื่อเสร็จขั้นตอนนี้แล้วจะได้พารามิเตอร์ที่ดีที่สุด

4.2 การแบ่งข้อมูลเพื่อนำทดสอบประสิทธิภาพของแบบจำลอง

เมื่อได้พารามิเตอร์ที่ดีที่สุดแล้ว จะทำการตรวจประสิทธิภาพของแบบจำลอง โดยจะสามารถตรวจสอบประสิทธิภาพของแบบจำลองนี้ได้จากการเปรียบเทียบราคาทองในปัจจุบันและราคาทองที่ทำนายได้จากแบบจำลอง ว่าค่าไหนใกล้เคียงกับราคาทองในอนาคตมากกว่ากัน ซึ่งจะแบ่งข้อมูลโดยใช้ Cross Validation แบบ Timeseriessplit แบ่งออกเป็น 10 ชุด และในแต่ละชุดจะวัดว่าแบบจำลองสามารถทำนายราคาทองในอนาคตได้ดีหรือไม่ ถ้าหากแบบจำลองสามารถทำนายได้ดีกว่าครึ่งนึงจะถือว่าแบบจำลองนี้ มีประสิทธิภาพหรือสามารถทำนายราคาทองได้

5.นำแบบจำลองไปใช้ทำนายราคาทองคำในอนาคต

หลังจากการทดสอบประสิทธิภาพของแบบจำลองแล้ว หากแบบจำลองมีประสิทธิภาพในการทำนายราคาทอง จะนำแบบจำลองไปใช้ในการทำนายราคาทองแบบไม่แบ่ง Cross Validation แล้วแสดงผลการทำงานของแบบจำลองผ่าน API

# บทที่ 4
# ผลการดำเนินโครงงาน

การดำเนินโครงงานในครั้งนี้ผู้จัดทำได้เสนอผลการดำเนินโครงงาน โดยแสดงผลลัพธ์ดังต่อไปนี้

## 4.1 ผลการดำเนินงานของแบบจำลอง ARIMA

| เลขพารามิเตอร์(p,d,q) | Trend , Method | ค่า RMSE |
|---|---|---|
| ARIMA (2,1,1) | c , css-mle | 10.887920304530965 |
| ARIMA (2,1,1) | c , css | 10.883551589387952 |
| ARIMA (2,1,1) | c , mle | 10.887699563880227 |
| ARIMA (2,1,1) | nc ,css -mle | 10.895499941173444 |
| ARIMA (2,1,1) | nc , css | 10.890931739619697 |
| ARIMA (2,1,1) | nc , mle | 10.895625893899972 |
| ARIMA (2,1,2) | c , css-mle | 30.835982403241722 |
| ARIMA (2,1,2) | c , css | 30.83609199550184 |
| ARIMA (2,1,2) | c , mle | 30.789128983944412 |
| ARIMA (2,1,2) | nc ,css -mle | 30.646011576767147 |
| ARIMA (2,1,2) | nc , css | 30.607558369747704 |
| ARIMA (2,1,2) | nc , mle | 30.611135899772634 |
| ARIMA (3,1,1) | c , css-mle | 31.00949925218406 |
| ARIMA (3,1,1) | c , css | 31.022330650006907 |
| ARIMA (3,1,1) | c , mle | 30.991322350799205 |
| ARIMA (3,1,1) | nc ,css -mle | 30.823031963237487 |
| ARIMA (3,1,1) | nc , css | 30.8369736009715 |
| ARIMA (3,1,1) | nc , mle | 30.80934297050093 |
| ARIMA (3,1,2) | c , css-mle | 30.86711337899664 |
| ARIMA (3,1,2) | c , css | 30.825606789972618 |
| ARIMA (3,1,2) | c , mle | 30.91172096180258 |
| ARIMA (3,1,2) | nc ,css -mle | 30.685933350352663 |
| ARIMA (3,1,2) | nc , css | 30.64516749416912 |

| | | |
|---|---|---|
| ARIMA (3,1,2) | nc , mle | 30.698456635178918 |
| ARIMA (4,1,1) | c , css-mle | 31.307666350704523 |
| ARIMA (4,1,1) | c , css | 31.26516683213244 |
| ARIMA (4,1,1) | c , mle | 31.143355991930584 |
| ARIMA (4,1,1) | nc ,css -mle | 31.073856343086064 |
| ARIMA (4,1,1) | nc , css | 31.071668716544256 |
| ARIMA (4,1,1) | nc , mle | 30.915423204532882 |

*ตาราง 2* ตารางบันทึกผลการหาค่าพารามิเตอร์ที่ดีที่สุดของแบบจำลอง ARIMA

จากการหาพารามิเตอร์ที่ดีที่สุดของแบบจำลอง ARIMA คือ ARIMA (2,1,1) c , css ค่า RMSE = 10.883551589387952

จากทำการตรวจสอบประสิทธิภาพของแบบจำลองได้ผลลัพธ์ คือ แบบจำลองสามารถทำนายราคาทองได้ดีกว่าราคาทองในปัจจุบันจำนวน 7 ชุดข้อมูล

### 4.2 ผลการดำเนินงานของแบบจำลอง SES

```
tscv = TimeSeriesSplit(n_splits = 200)
rmse = []

for train_index, test_index in tscv.split(ts):
  cv_train, cv_test = ts.iloc[train_index], ts.iloc[test_index]

  model = SimpleExpSmoothing(data)
  model_fit = model.fit()

  predictions = model_fit.predict(cv_test.index.values[0], cv_test.index.values[-1])
  rmse_result = sqrt(mean_squared_error(cv_test.values, predictions))
  rmse.append(rmse_result)

print("RMSE: {}".format(np.mean(rmse)))
```

```
RMSE: 14.191506565014084
```

*รูปที่ 30 หาค่า RMSE ของแบบจำลอง SES*

หลังจากได้ใช้แบบจำลอง SES ในการทำนายราคาทองแล้วได้ค่า RMSE = 14.191506

จากทำการตรวจสอบประสิทธิภาพของแบบจำลองได้ผลลัพธ์ คือ แบบจำลองสามารถทำนายราคาทองได้ดีกว่าราคาทองในปัจจุบันจำนวน 6 ชุดข้อมูล

**4.3 ผลการดำเนินงานของแบบจำลอง HWS**

| Trend | Seasonal | ค่า RMSE |
|---|---|---|
| Add | - | 14.10103 |
| Mul | - | 14.081040833 |
| - | Add | 17.592097 |
| - | Mul | 14.0377424 |
| Add | Add | 17.97907 |
| Add | Mul | 23.7574834 |
| Mul | Mul | - |
| Mul | Add | - |
| Mul | seasonal_periods=26 | 14.122758265 |
| Mul | seasonal_periods=52 | 24.205636972 |

*ตาราง 3* ตารางบันทึกผลการหาค่าพารามิเตอร์ที่ดีที่สุดของแบบจำลอง HWS

จากการหาพารามิเตอร์ที่ดีที่สุดของแบบจำลอง HWS คือไม่ใส่รูปแบบ Trend และ Seasonal เป็นแบบ Mul ค่า RMSE = 14.081040833

จากทำการตรวจสอบประสิทธิภาพของแบบจำลองได้ผลลัพธ์ คือ แบบจำลองสามารถทำนายราคาทองได้ดีกว่าราคาทองในปัจจุบันจำนวน 4 ชุดข้อมูล ซึ่งหมายถึงแบบจำลองไม่สามารถใช้ในการทำนายราคาทองคำได้

4.4 ผลการดำเนินงานของแบบจำลอง CNN 1 input

```
true_predictions = scaler.inverse_transform(result_predictions)
true_predictions = true_predictions.reshape(n_features,n_step)

data_test_rmse = []
data_test_rmse.append(data_test.values[0][0])
print(data_test_rmse)

true_predictions_rmse = []
true_predictions_rmse.append(true_predictions[0])
print(true_predictions_rmse)

rmse_result = sqrt(mean_squared_error(data_test_rmse, true_predictions_rmse))
print(rmse_result)
```

```
[1617.3]
[1525.744509994984]
91.55549000501605
```

*รูป 31* หาค่า RMSE ของแบบจำลอง CNN 1 input

จากการทำนายราคาทองโดยแบบจำลอง CNN 1 input ได้ค่าที่ทำนายราคาทองคือ 1525.74450 ค่า RMSE ได้ 91.55549 ส่วนราคาทองในอนาคตจะอยู่ที่ 1617.3 (ข้อมูลราคาที่ใช้เป็น Test data)

จากทำการตรวจสอบประสิทธิภาพของแบบจำลองได้ผลลัพธ์ คือ แบบจำลองสามารถทำนายราคาทองได้ดีกว่าราคาทอง

**4.5 ผลการดำเนินงานของแบบจำลอง CNN Allinput**

```
Epoch 1999/2000
1408/1408 [==============================] - 2s 1ms/step - loss: 2.9019e-04
Epoch 2000/2000
1408/1408 [==============================] - 2s 1ms/step - loss: 2.8937e-04
[1680.7]
[1536.7660479426386]
143.93395205736147
```

*รูป 32* หาค่า RMSE ของแบบจำลอง CNN Allinput

จากการทำนายราคาทองโดยแบบจำลอง CNN Allinput ได้ค่าที่ทำนายราคาทองคือ 1536.766047942 ค่า RMSE ได้ 143.933952057 ส่วนราคาทองในอนาคตจะอยู่ที่ 1680.7 (ข้อมูลราคาที่ใช้เป็น Test data)

จากทำการตรวจสอบประสิทธิภาพของแบบจำลองได้ผลลัพธ์ คือ แบบจำลองสามารถทำนายราคาทองได้แย่กว่าราคาทองในปัจจุบัน ซึ่งหมายถึงแบบจำลองไม่สามารถใช้ในการทำนายราคาทองคำได้

## 4.6 ผลการดำเนินงานของแบบจำลอง LSTM 1 input

```
true_predictions = scaler.inverse_transform(result_predictions)
true_predictions = true_predictions.reshape(n_features)
print(true_predictions)

data_test_rmse = []
data_test_rmse.append(data_test.values[0])
data_test_rmse_float = []
data_test_rmse_float.append(float(data_test_rmse[0]))
print(data_test_rmse_float)



rmse_result = sqrt(mean_squared_error(data_test_rmse_float, true_predictions))
print(rmse_result)
```

```
[1625.20058555]
[1680.7]
55.4994144499301
```

*รูป 33* หาค่า RMSE ของแบบจำลอง LSTM 1 input

จากการทำนายราคาทองโดยแบบจำลอง LSTM 1 input ได้ค่าที่ทำนายราคาทองคือ 1625.2005855 ค่า RMSE ได้ 55.4994144 ส่วนราคาทองในอนาคตจะอยู่ที่ 1680.7 (ข้อมูลราคาที่ใช้เป็น Test data)

จากทำการตรวจสอบประสิทธิภาพของแบบจำลองได้ผลลัพธ์ คือ แบบจำลองสามารถทำนายราคาทองได้ดีกว่าราคาทอง

**4.7 ผลการดำเนินงานของแบบจำลอง LSTM Allinput**

```
Epoch 1997/2000
470/470 [==============================] - 2s 4ms/step
Epoch 1998/2000
470/470 [==============================] - 2s 4ms/step
Epoch 1999/2000
470/470 [==============================] - 2s 4ms/step
Epoch 2000/2000
470/470 [==============================] - 2s 4ms/step
[1633.29880743   20.99608255   22.42808556    1.9443270
    2.21055481]
[1680.7]
[1633.2988074302675]
47.40119256973253
```

*รูป 34* หาค่า RMSE ของแบบจำลอง LSTM Allinput

จากการทำนายราคาทองโดยแบบจำลอง LSTM Allinput ได้ค่าที่ทำนายราคาทองคือ 1633.2988074 ค่า RMSE ได้ 47.40119256 ส่วนราคาทองในอนาคตจะอยู่ที่ 1680.7 (ข้อมูลราคาที่ใช้เป็น Test data)

จากทำการตรวจสอบประสิทธิภาพของแบบจำลองได้ผลลัพธ์ คือ แบบจำลองสามารถทำนายราคาทองได้ดีกว่าราคาทอง

**4.8 ผลการดำเนินงานของแบบจำลอง VAR Allinput**

จากการทำนายราคาทองโดยแบบจำลอง VAR Allinput ได้ค่าที่ทำนายราคาทองคือ 1626.356867 ค่า RMSE ได้ 54.34313 ส่วนราคาทองในอนาคตจะอยู่ที่ 1680.7 (ข้อมูลราคาที่ใช้เป็น Test data)

จากทำการตรวจสอบประสิทธิภาพของแบบจำลองได้ผลลัพธ์ คือ แบบจำลองสามารถทำนายราคาทองได้แย่กว่าราคาทองในปัจจุบัน ซึ่งหมายถึงแบบจำลองไม่สามารถใช้ในการทำนายราคาทองคำได้

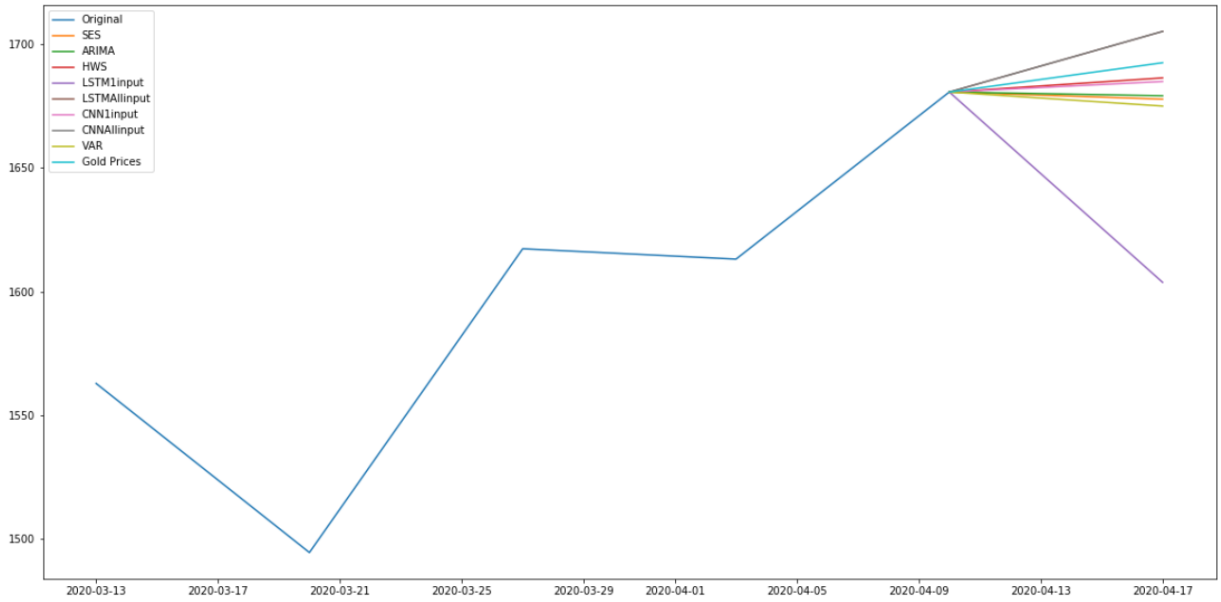
จากแบบจำลองที่ได้ทำการทำนายราคาได้ ได้แก่ SES,ARIMA,CNN 1 input,LSTM 1 input,LSTM Allinput แล้วนำแบบจำลองไปทำนายราคาทองโดยไม่แบ่ง Cross Validation จะได้ผลทำนายแสดงผล API ดังรูปที่ 35 ซึ่งใช้แบบจำลองวันที่ 17 เมษายน 2563 เพื่อทำนายราคาทองวันที่ 24 เมษายน 2563

```
"ARIMA": {
    "Date": "Fri, 01 May 2020 00:00:00 GMT",
    "Model": "ARIMA",
    "Prediction": 1612.8744520126709
},
"CNN 1 input": {
    "Date": "Fri, 01 May 2020 00:00:00 GMT",
    "Model": "CNN1input",
    "Prediction": 1716.4784371972085
},
"CNN All Input": {
    "Date": "Fri, 01 May 2020 00:00:00 GMT",
    "Model": "CnnAllinput",
    "Prediction": 1808.174997496605
},
"HWS": {
    "Date": "Fri, 01 May 2020 00:00:00 GMT",
    "Model": "HWS",
    "Prediction": 1670.8413974559542
},
"LSTM 1 input": {
    "Date": "Fri, 01 May 2020 00:00:00 GMT",
    "Model": "LSTM 1 input",
    "Prediction": 1584.9013176620008
},
"LSTMAllinput": {
    "Date": "Fri, 01 May 2020 00:00:00 GMT",
    "Model": "LSTMAllinput",
    "Prediction": 1808.174997496605
},
"SES": {
    "Date": "Fri, 01 May 2020 00:00:00 GMT",
    "Model": "SES",
    "Prediction": 1677.7990197349457
},
"VAR": {
    "Date": "Fri, 01 May 2020 00:00:00 GMT",
    "Model": "VAR",
    "Prediction": 1672.419252869331
}
```


*รูป 35* ผลการทำนายราคาทองคำแสดงผล API

| ชื่อแบบจำลอง | RMSE |
|---|---|
| ARIMA | 10.88 |
| SES | 14.19 |
| HWS | 14.08 |
| CNN 1 input | 91.55 |
| CNN Allinput | 143.93 |
| LSTM 1 input | 55.49 |
| LSTM Allinput | 47.40 |
| VAR Allinput | 54.34 |

*ตาราง 4* เปรียบเทียบค่าเฉลี่ย RMSE ของแต่ละแบบจำลอง




*รูป 36 การทำนายราคาทอง 1 สัปดาห์โดยไม่แบ่ง Cross Validation ของทุกแบบจำลอง*

# บทที่ 5
# สรุปผล อภิปรายผล และข้อเสนอแนะ

## 5.1 สรุปผลการดำเนินการโครงการ

จากการศึกษาความสัมพันธ์ระหว่างราคาทองคำกับปัจจัยทางเศรษฐกิจตัวอื่น ได้ใช้แบบจำลองทั้งหมด 7 วิธี ได้แก่ ARIMA,SES,HWS,LSTM 1 input,LSTM Allinput,CNN 1 input ,CNN Allinput ซึ่งจากการตรวจประสิทธิภาพของแบบจำลอง แบบจำลองที่สามารถใช้ทำนายราคาทองได้ดีที่สุดคือ ARIMA เนื่องจากเราได้ทำการปรับข้อมูลให้มีความนิ่งตามคำแนะนำการสร้างแบบจำลอง ในขณะที่วิธีอื่นๆไม่ได้ปรับตรงส่วนนั้น

การศึกษาวิเคราะห์และทำนายข้อมูลราคาทอง พิจารณาแต่เพียงการเปลี่ยนแปลงค่าทางในเชิงการคำนวณทางสถิติเท่านั้น การทำนายข้อมูลนี้ไม่สามารถครอบคลุมถึงปัจจัยที่ซับซ้อนทางเศรษฐกิจอื่นได้ เช่น การก่อการร้ายหรือเกิดวิกฤตทางการเงิน ทำให้ค่าของการทำนายที่ได้ มิอาจมีความแม่นยำจนนำไปใช้ประโยชน์ทางเศรษฐกิจอย่างจริงจังได้

## 5.2 ปัญหาและอุปสรรค

- หาข้อมูลของราคาที่เกี่ยวข้องยาก บางข้อมูลหากนำมาใช้ต้องเสียค่าใช้จ่าย
- ข้อมูลราคาที่จะนำมาใช้ในแบบจำลอง บางข้อมูลมีข้อมูลในอดีตน้อยเกินไป ทำให้ไม่สามารถนำมาวิเคราะห์ได้
- ตอนแรกใช้ Colaboratory ในการทำแบบจำลอง ติดปัญหาอินเตอร์เน็ตไม่เสถียร ทำให้ต้อง Train ข้อมูลใหม่บ่อยครั้ง

## 5.3 ข้อเสนอแนะและแนวทางในการพัฒนา

- ในการทำนายให้ได้ความแม่นยำต้องใช้ข้อมูลปัจจุบันที่สุด หากมีการพัฒนาในอนาคตหากพัฒนาให้ข้อมูลราคาอัพเดตอัตโนมัติได้ก็จะเป็นผลที่ดีกับการทำนาย
- เพิ่มประสิทธิภาพของแบบจำลองโดยเพิ่มจำนวนข้อมูลจำนวนที่เกี่ยวข้อง
- ทำ application ที่ใช้ในการตรวจสอบผลการทำนาย เพื่อง่ายต่อการใช้งานของผู้ใช้ในอนาคต

# บรรณานุกรม


[1] ณัฐกิต การย์เกรียงไกร. (2558). **การศึกษาความสัมพันธ์ระหว่างดัชนีหลักทรัพย์กลุ่มชิ้นส่วนอิเล็กทรอนิกส์กับปัจจัยทางเศรษฐกิจด้วย แบบจำลอสVAR.** สืบค้นเมื่อ 18 พฤศจิกายน 2563, จาก http://dspace.bu.ac.th/bitstream/123456789/1772/1/Natakit.Karn.pdf

[2] ต่อพงษ์ บวรพงษ์สกุล และ ธาตรี จันทรโคลิกา. (2554). **ผลกระทบของนโยบาบการเงินและการตลังที่มีต่อระบบเศรษฐกิจมหภาคโดยใช้แบบจำลอง VARs**. สืบค้นเมื่อ 18 พฤศจิกายน 2563, จาก http://www.tpso.moc.go.th/sites/default/files/748-img.pdf

[3] เฉลิมพล จตุพร. (2560). **คู่มือการใช้โปรแกรมสำเร็จรูปทางเศรษฐมิติ GRETL.** สืบค้นเมื่อ 25 พฤศจิกายน 2563, จาก https://cj007blog.files.wordpress.com/2017/08/06-time-series-forecasting5.pdf

[4] วราพร งามสุข. (2555). **การพยากรณ์อนุกรมเวลาโดยการเปรียบเทียบวิธีแบบฉบับและวิธีบอกซ์-เจนกินส์กรณีศึกษาจำนวนการเกิดอุบัติเหตุในประเทศไทย**. สืบค้นเมื่อ 30 พฤศจิกายน 2563, จาก http://www.lib.buu.ac.th/st/ST0002619.pdf

[5] ปณิธี ธัมมวิจยะ. (2560). **การพยากรณ์โดยการวิเคราะห์ข้อมูลอนุกรมเวลา.** สืบค้นเมื่อ 2 ธันวาคม 2563, จาก http://bps.moph.go.th/new_bps/sites/default/files/ime%20Series%20Analysis_2017.pdf

[6] Jason Brownlee. (2018). **How to Develop Convolutional Neural Network Models for Time Series Forecasting.** สืบค้นเมื่อ 15 ธันวาคม 2563, จาก https://machinelearningmastery.com/how-to-develop-convolutional-neural-network-models-for-time-series-forecasting/

[7] ปทิตญา บุญรักษา และ จารี ทองคำ. (2560). **การเปรียบเทียบประสิทธิภาพของแบบจำลองการเกิดอุบัติเหตุทางถนน โดยใช้ เทคนิคอนุกรมเวลา.** สืบค้นเมื่อ 14 มกราคม 2563, จาก http://it.rmu.ac.th/itm-journal/assets/uploads/formidable/6/39-46.pdf

[8] พงษ์ศิริ ศิริพานิช. (2550). **การพยากร์อนุกลมเวลาด้วยตัวแบบผสมระหว่าง ARIMA และเครือข่ายประสาทเทียม.** สืบค้นเมื่อ 18 มกราคม 2563, จาก http://www.sure.su.ac.th/xmlui/bitstream/handle/123456789/11717/Fulltext.pdf?sequence=1&isAllowed=y

[9] สิทธา แก้วแปงจันทร์, เสมอแข สมหอม และ ลำปาง แสนจันทร์. (2553). **แบบจำลองการพยากรณ์ การใช้ไฟฟ้าโดยใช้ วิธีผสมของวิธีการปรับให้เรียบเอกซ์โพเนนเชียลแบบโฮลต์-วินเทอร์และขั้นตอนวิธีอาณานิคมผึ้งเทียม.** สืบค้นเมื่อ 10 กุมภาพันธ์ 2563,จาก http://www2.it.kmutnb.ac.th/journal/pdf/vol11/ch03.pdf

[10]กมล ท่าเรือรักษ์ และ ตะวัน ศรีดามา. (2549). **บทบาทกรมทางหลวงกับการพัฒนาประเทศโดย Vector Autoregressive Model(VAR).** สืบค้นเมื่อ 10 กุมภาพันธ์ 2563, จาก http://winti.pte.co.th/e_document/attachment/WP.233.pdf

[11]มนัสกานต์ เสน่หา. (2559). **การทำนายอายุการใช้งานคงเหลือของเครื่องจักรด้วยเน็ตเวิร์กคอนโวลูชันเชิงลึกที่เพิ่มประสิทธิ์ภาพ**. สืบค้นเมื่อ 19 กุมภาพันธ์ 2563, จาก https://cuir.car.chula.ac.th/dspace/bitstream/123456789/58338/1/5870956621.pdf

[12]PradyaSin. (2562). **Long Short-Term Memory คืออะไร**. สืบค้นเมื่อ 22 กุมภาพันธ์ 2563, จาก https://medium.com/@pradyasin/what-is-lstm-1e6e92ba5b4c

[13]Sirinart Tangruamsub. (2560). **Long Short-Term Memory (LSTM).** สืบค้นเมื่อ 22 กุมภาพันธ์ 2563, จาก https://medium.com/@sinart.t/long-short-term-memory-lstm-e6cb23b494c6

[14]สิรินภา จิตราช. (2558). **การเปรียบเทียบความแม่นยำในการพยากรณ์ดัชนีราคาตลาดหลักทรัพย์ในกลุ่มเทคโนโลยีสารสนเทศและการสื่อสารระหว่างแบบจำลอง ARIMA และแบบจำลอง E-GARCH.** สืบค้นเมื่อ 25 กุมภาพันธ์ 2563, จาก http://mslib.kku.ac.th/elib/books/Economic2558/SIRINAPA%20JITRAT/05_ch2.pdf

[15]จินดามาส สุทธิชัยเมธี. (2554). **การประยุกต์ใช้ ARIMA Model เพื่อการวิจัย.** สืบค้นเมื่อ 25 กุมภาพันธ์ 2563, จาก https://www.dpu.ac.th/dpurc/assets/uploads/magazine/f0ati696674k8g.pdf

[16]ภูมิฐาน รังคกลูนุวัฒน์. (2562). **การวิเคราะห์อนุกรมเวลาสำหรับเศรษฐศาสตร์และธุรกิจ**. สืบค้นเมื่อ 27 กุมภาพันธ์ 2563, จาก https://economics.utcc.ac.th/wp-content/uploads/Time-Series-for-Econ-and-Bus_Poomthan.pdf

[17]อภินันท์ พัชโรภาสวัฒนกุล และคนอื่นๆ. (2560). **ผลกระทบของการเปลี่ยนแปลงสภาพภูมิอากาศต่อผลผลิตมันสำปะหลังในภาคตะวันออกเฉียงเหนือ**. สืบค้นเมื่อ 27 กุมภาพันธ์ 2563, จาก https://erp.mju.ac.th/openFile.aspx?id=Mjc1Nzk2

[18]อัครัส บรรจงศิลป์. (2550). **การพยากรณ์หารใช้พลังงานไฟฟ้าในอุตสาหกรรมขนาดใหญ่**. สืบค้นเมื่อ 5 มีนาคม 2563, จาก http://www.thapra.lib.su.ac.th/objects/thesis/fulltext/snamcn/Akkaratch_Bunchongsilp/Fulltext.pdf

[19]สุดใจ ผุดผาด. (2562). **การพยากรณ์จำนวนผู้รับบริการเครื่องแก๊สโครมาโทกราฟ-แมสสเปกโทรมิเตอร์ของศูนย์เครื่องมือวิทยาศาสตร์ คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง**. สืบค้นเมื่อ 7 มีนาคม 2563, จาก http://www.council-uast.com/journal/upload/fullpaper/19-10-2019-870234743.pdf

[20]Jairo Moreno. (2015). **Exponential smoothing**. สืบค้นเมื่อ 7 มีนาคม 2563, จาก https://pt.slideshare.net/JairoMoreno15/exponential-smoothing-52500403?smtNoRedir=1

[21]Rob J Hyndman and George Athanasopoulos. (2018). **Forecasting: Principles and Practice**. สืบค้นเมื่อ 7 มีนาคม 2563, จาก https://otexts.com/fpp2/index.html

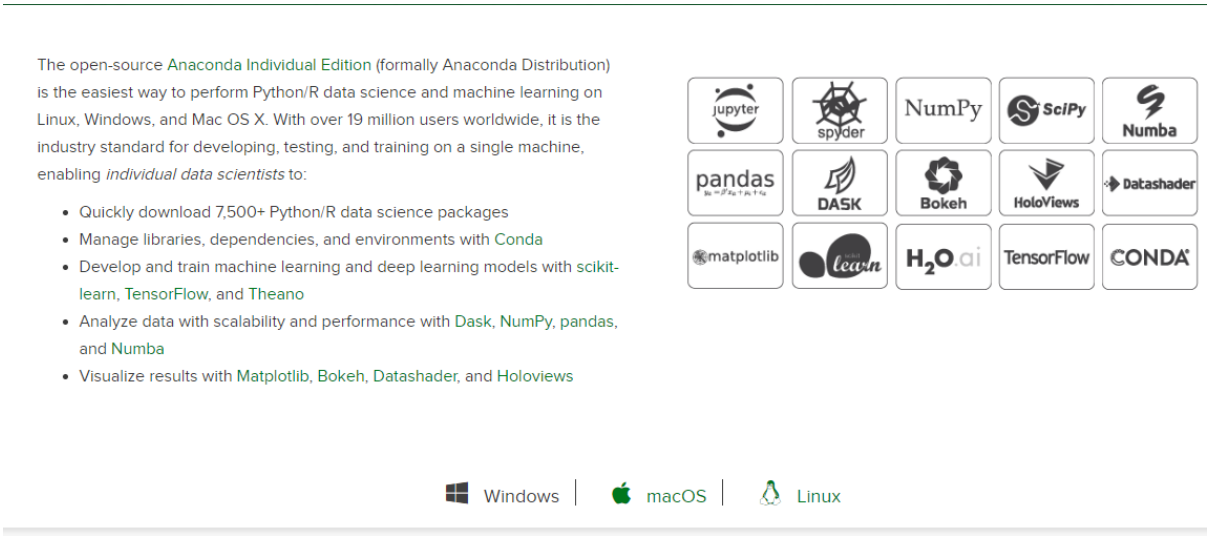
# ภาคผนวก

# ภาคผนวก ก

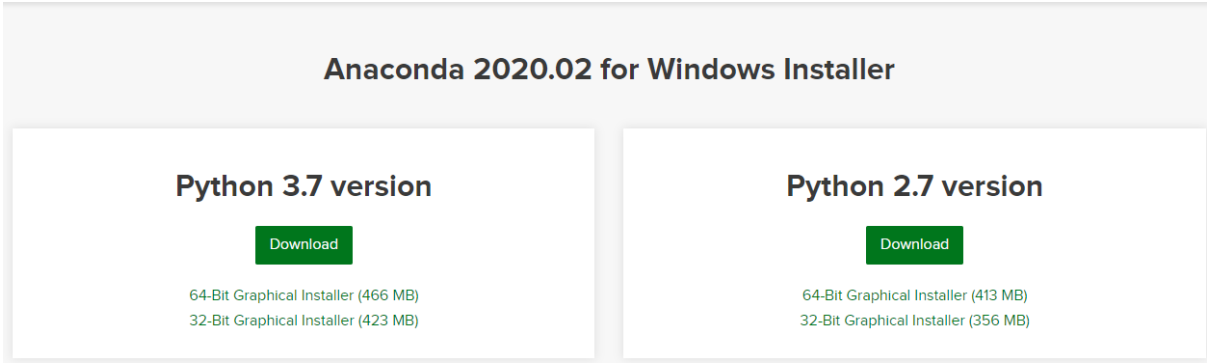
## การติดตั้งโปรแกรม

Anaconda

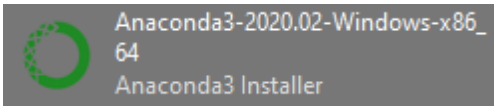
1. เข้าเว็บไซต์ https://www.anaconda.com/distribution/
2. เลือกระบบปฏิบัติการตามคอมพิวเตอร์ที่ใช้งาน (คอมพิวเตอร์ตัวอย่างใช้งานระบบปฏิบัติการ Window)



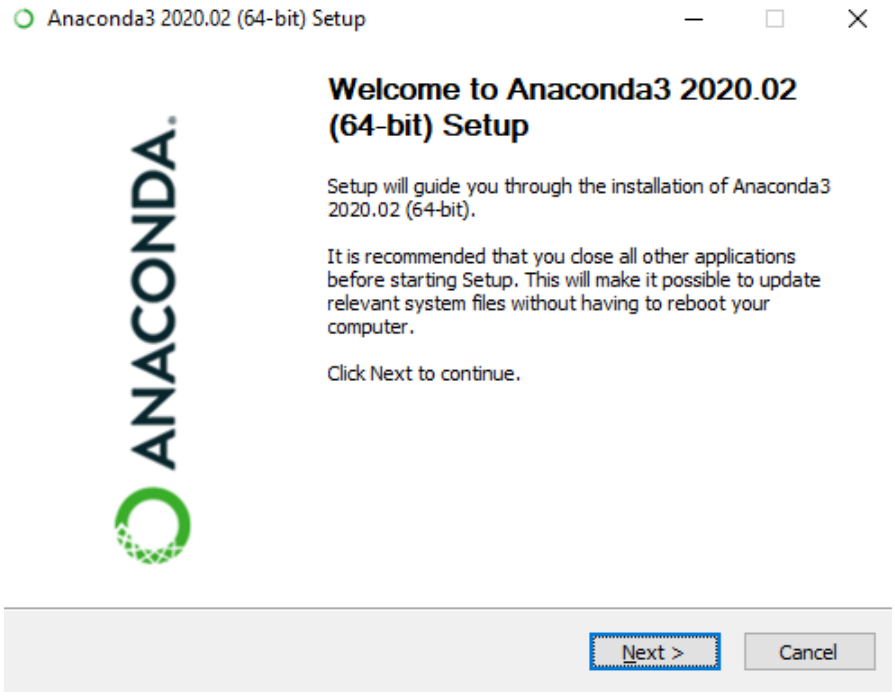

3. เลือก Download Python 3.7 Version



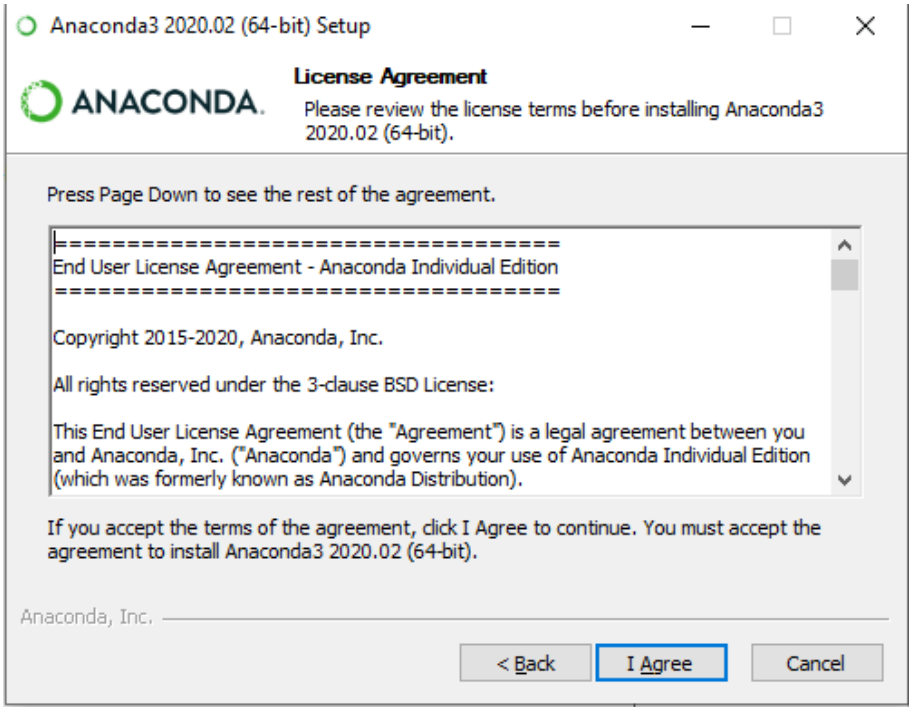

4. เลือกไฟล์ที่ Download มา เพื่อเริ่มติดตั้ง



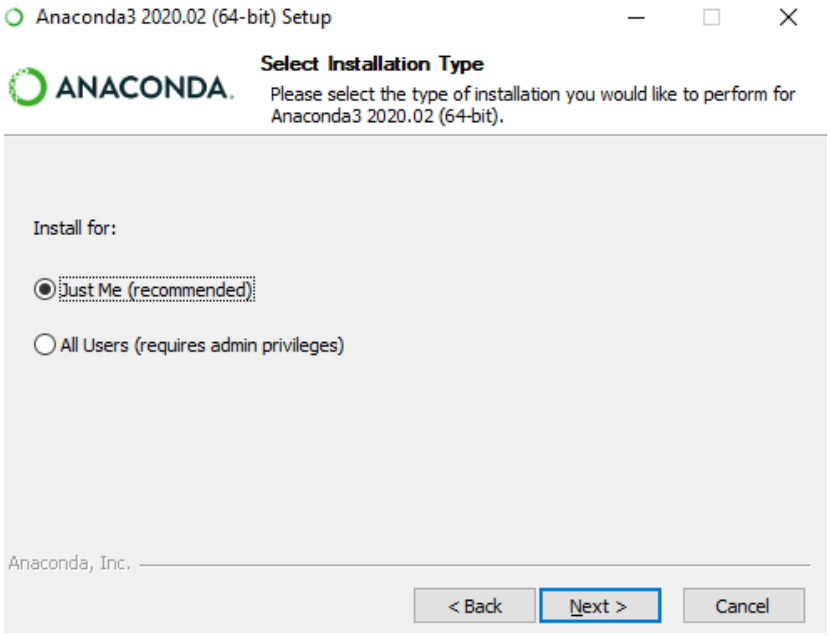
5. เลือก Next



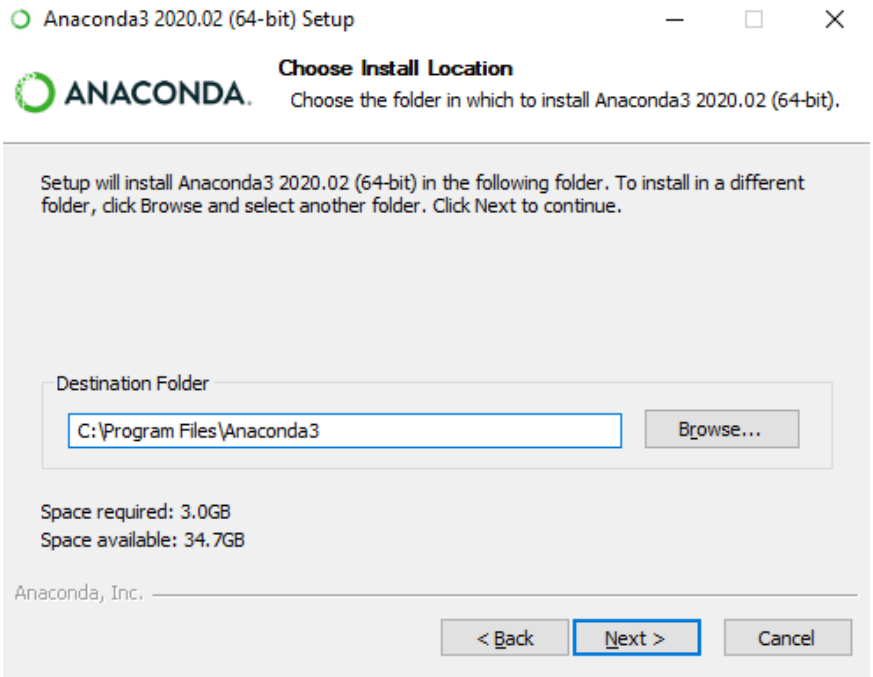

6. เลือก I Agree



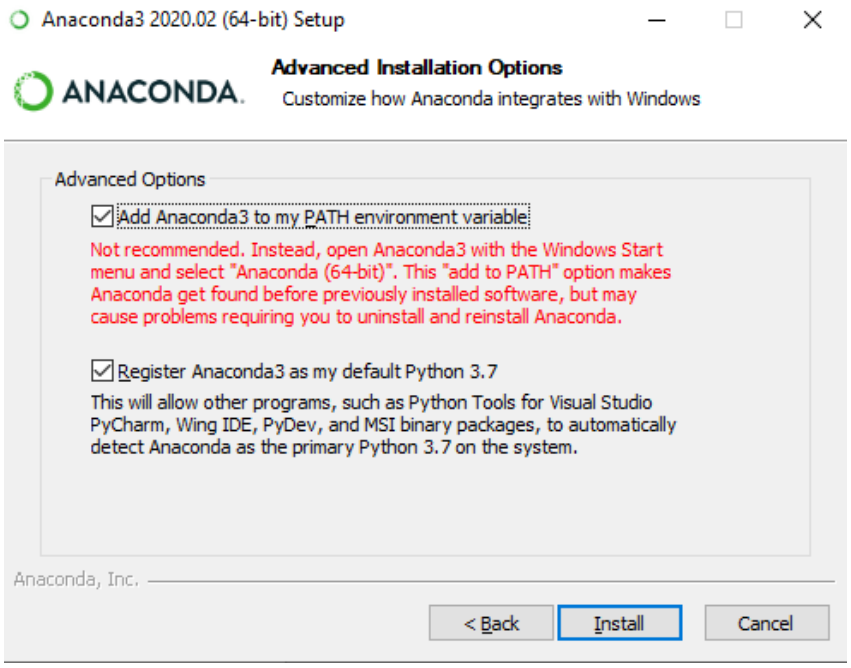
7. เลือก Just me แล้ว เลือก Next



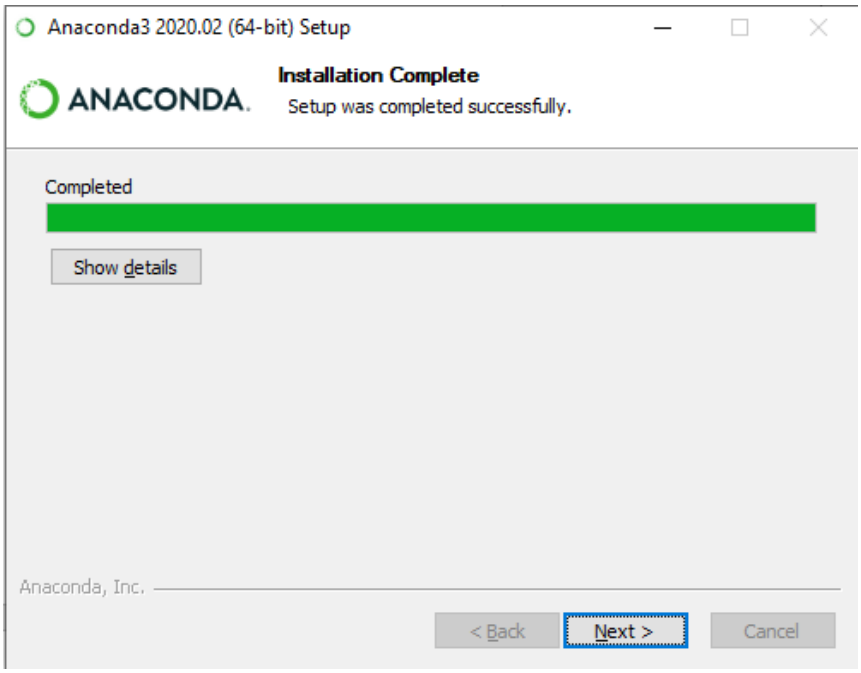

8. เลือก Browse หากต้องการเปลี่ยนจุดติดตั้งไฟล์ แล้วเลือก Next



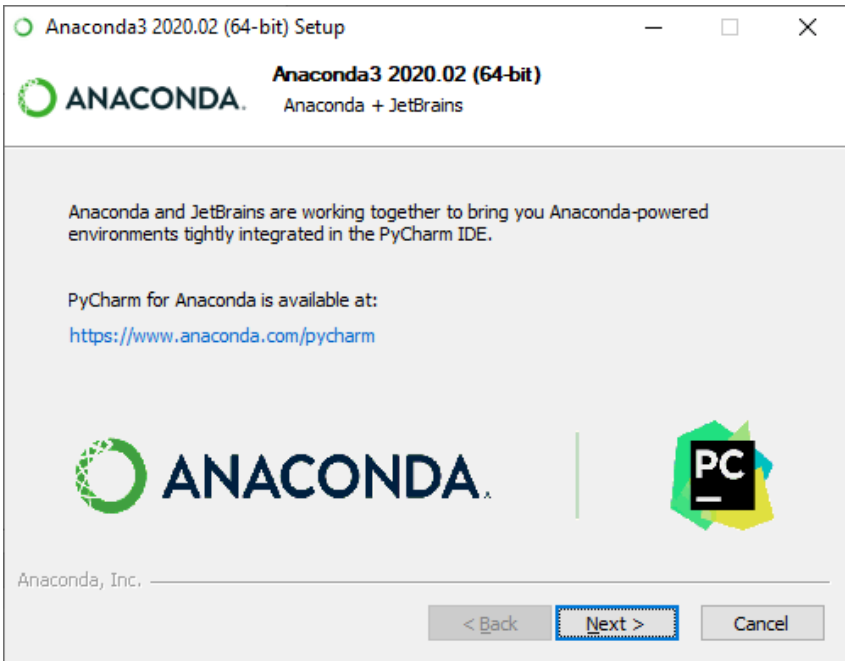
9. เลือก Add Anaconda3 to my PATH environment variable และ Register Anaconda3 as my default Python 3.7 แล้วเลือก Install



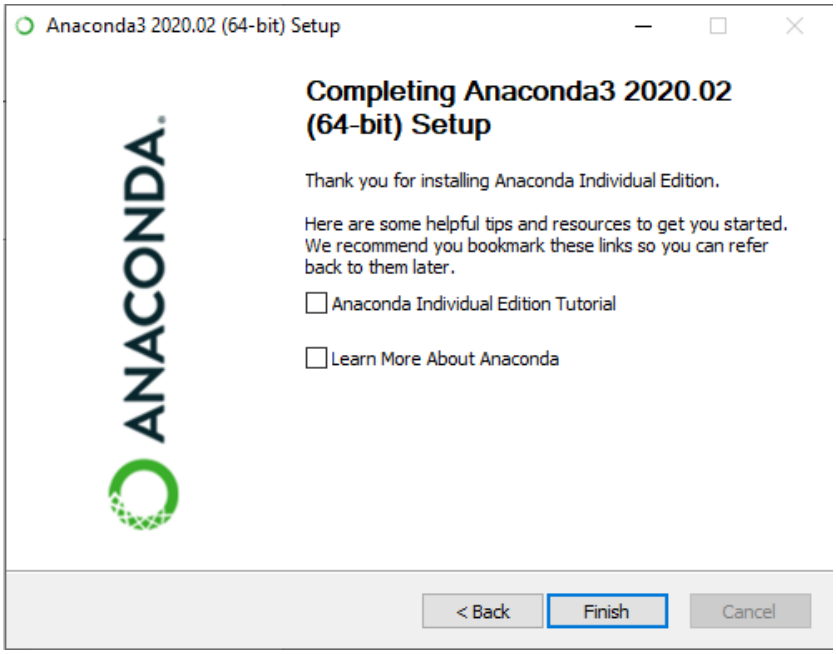

10. เมื่อโปรแกรมติดตั้งเสร็จเรียบร้อยให้เลือก Next



11. เลือก Next

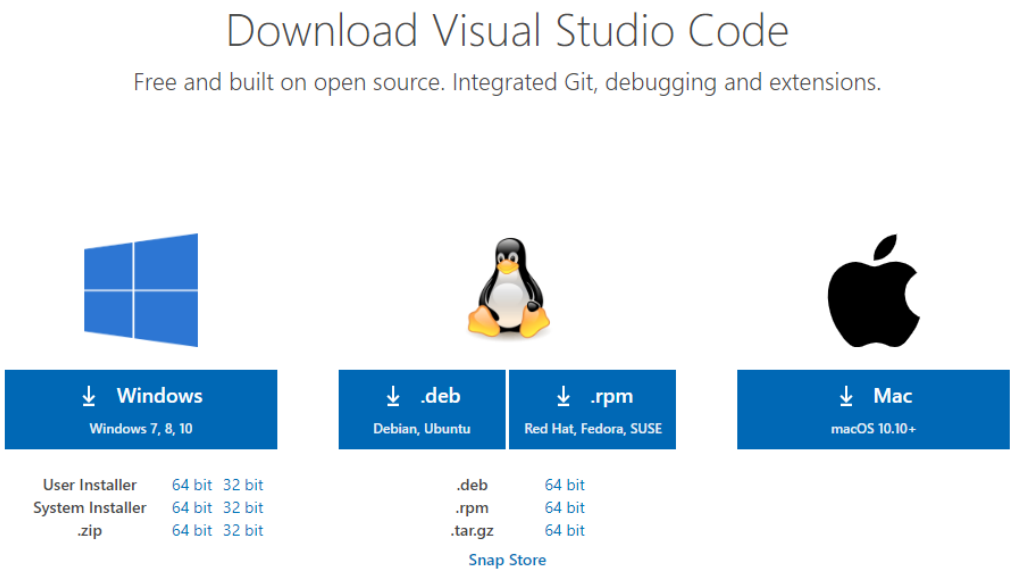



12. เลือก Finish เป็นการเสร็จสิ้นการติดตั้งโปรแกรม

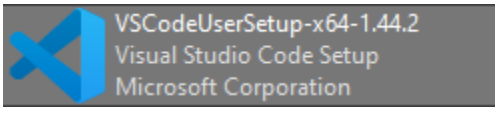


การติดตั้งโปรแกรม Visual Studio Code

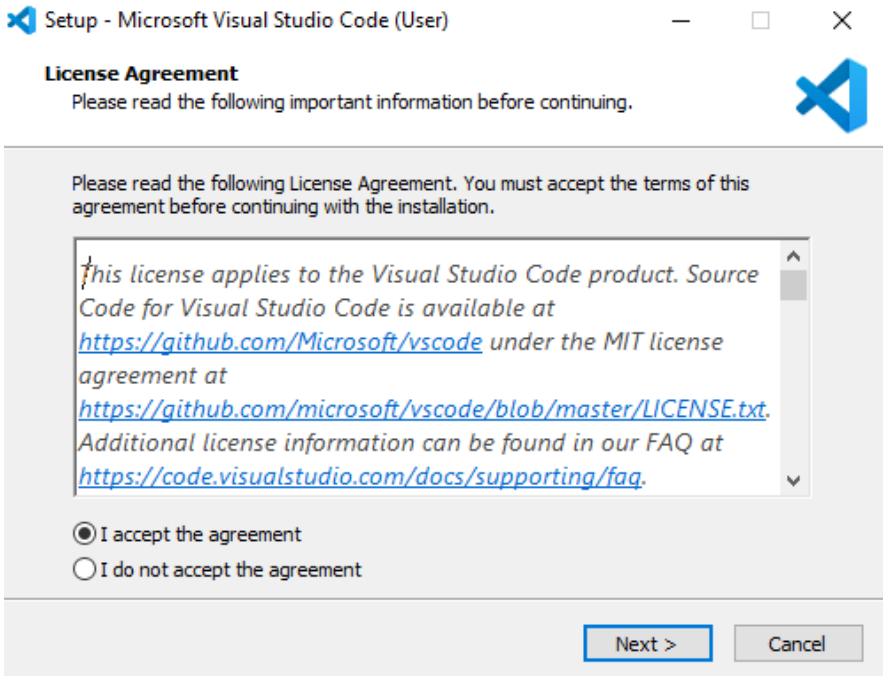
1. เข้าเว็บไซต์ https://code.visualstudio.com/Download
2. เลือก Download ไฟล์ตามระบบปฏิบัติการของคอมพิวเตอร์ที่ใช้งาน (คอมพิวเตอร์ตัวอย่างใช้งานระบบปฏิบัติการ Window)



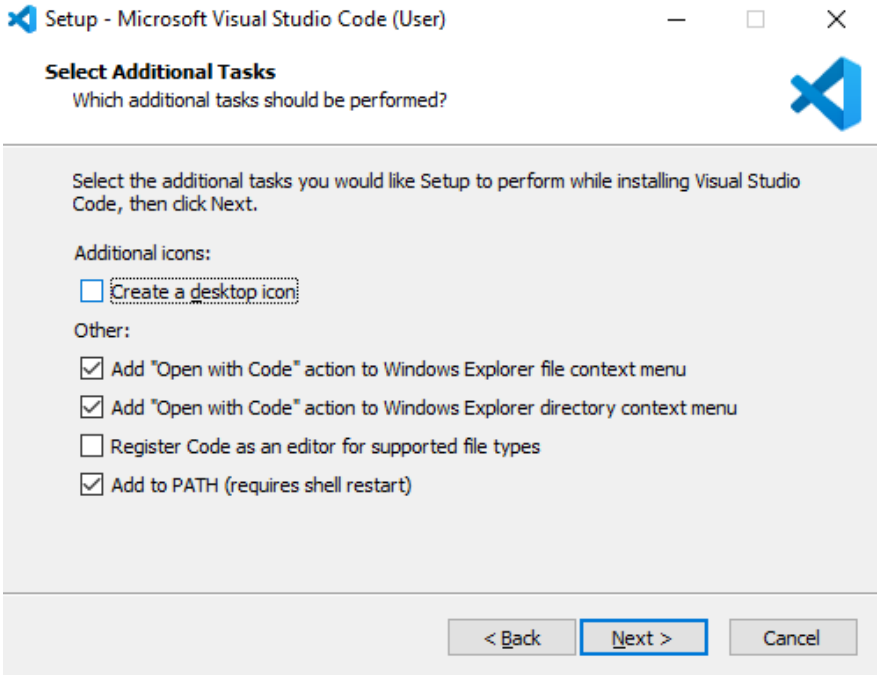

3. เลือกไฟล์ที่ Download มา เพื่อเริ่มติดตั้ง



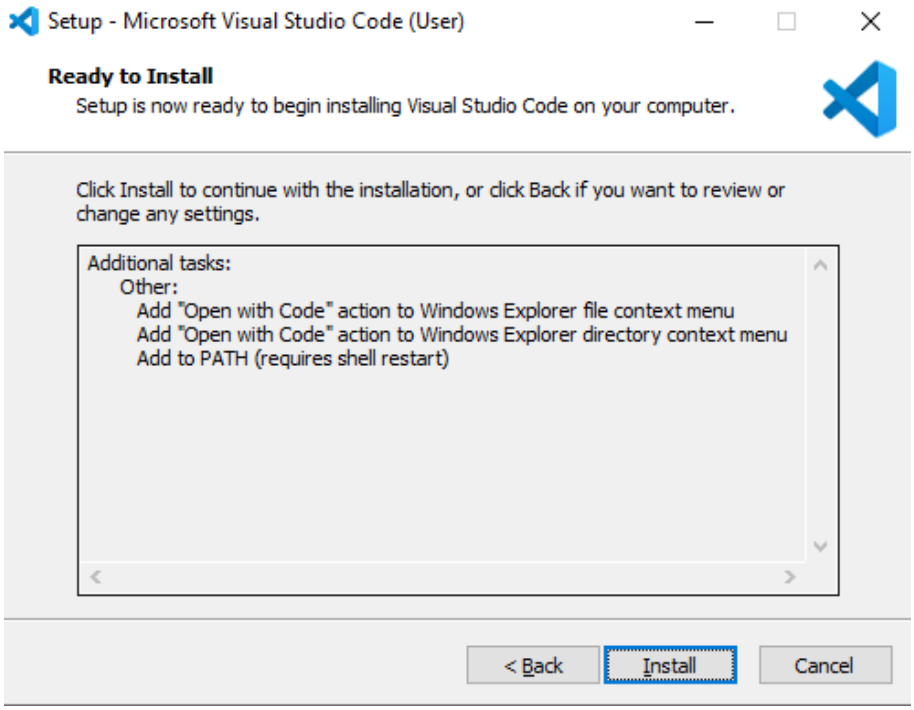
4. เลือก I accept the agreement แล้วเลือก Next



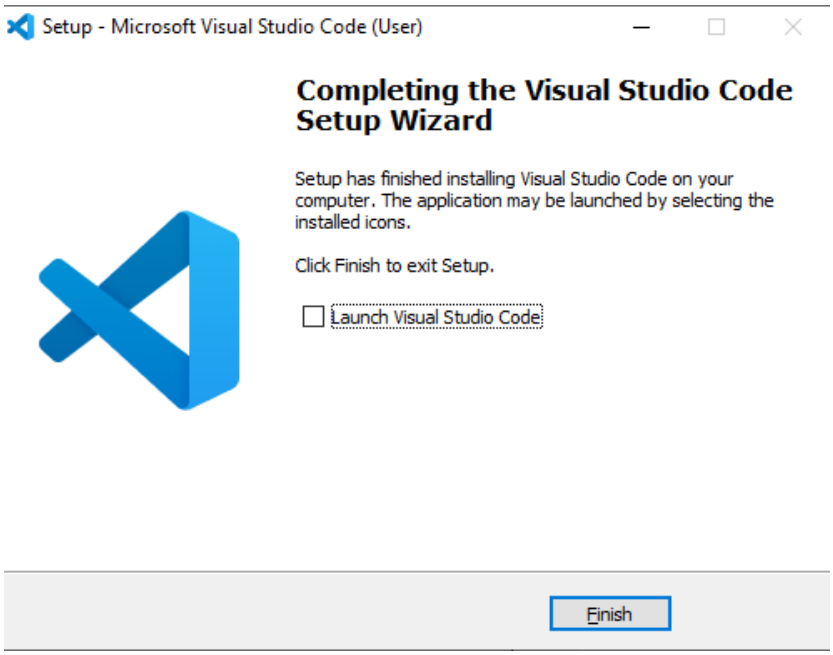

5. หากโปรแกรมไม่เลือก Add to PATH ต้องเลือกก่อน แล้วเลือก Next



6. เลือก Install

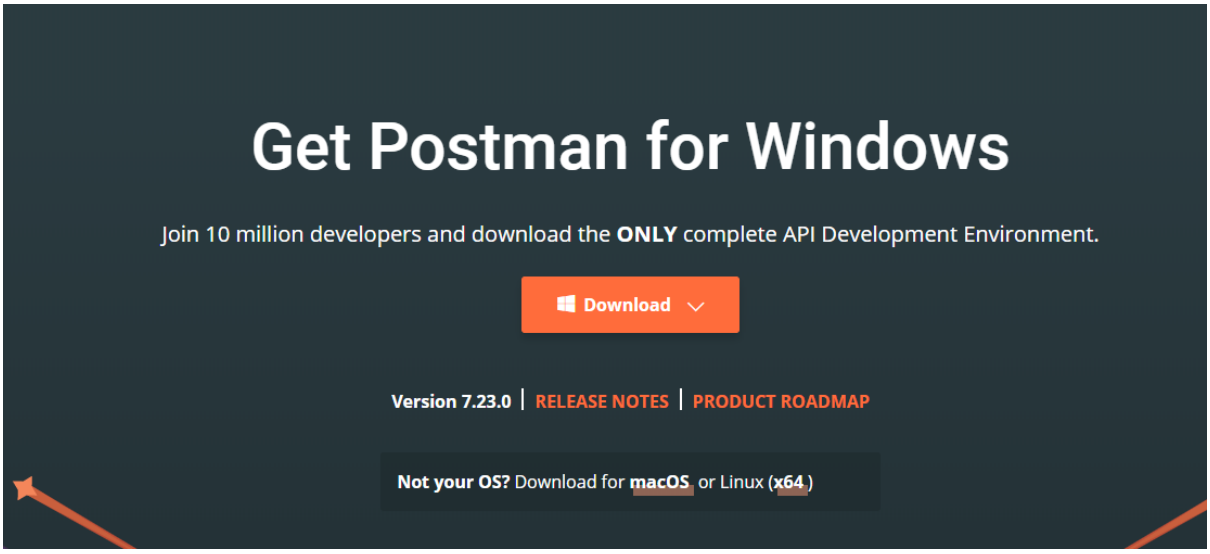



7. รอจนโปรแกรมติดตั้งเสร็จสมบูรณ์ แล้วเลือก Finish เป็นการเสร็จสิ้นการติดตั้ง

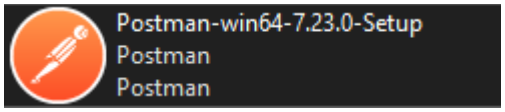


การติดตั้ง Postman

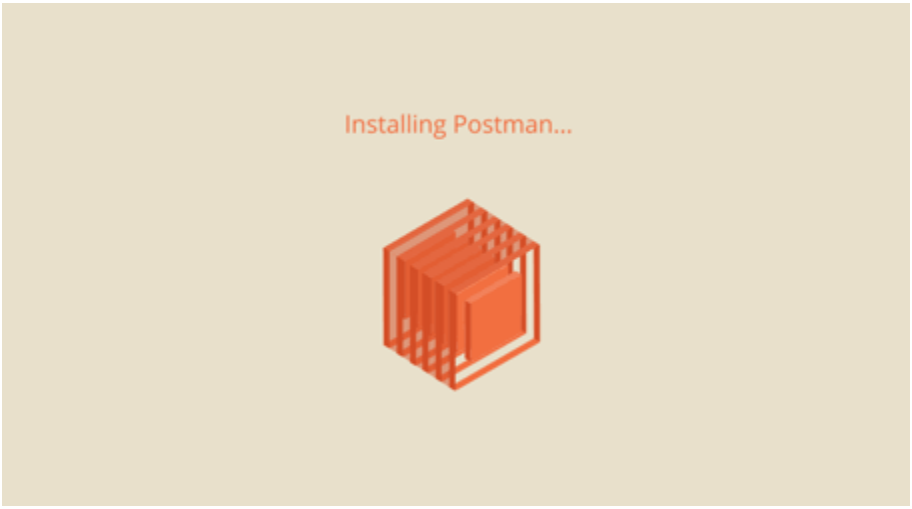
1. เข้าเว็บไซต์ https://www.postman.com/downloads/
2. เลือกระบบปฏิบัติการตามคอมพิวเตอร์ที่ใช้งาน (คอมพิวเตอร์ตัวอย่างใช้งานระบบปฏิบัติการ Window)



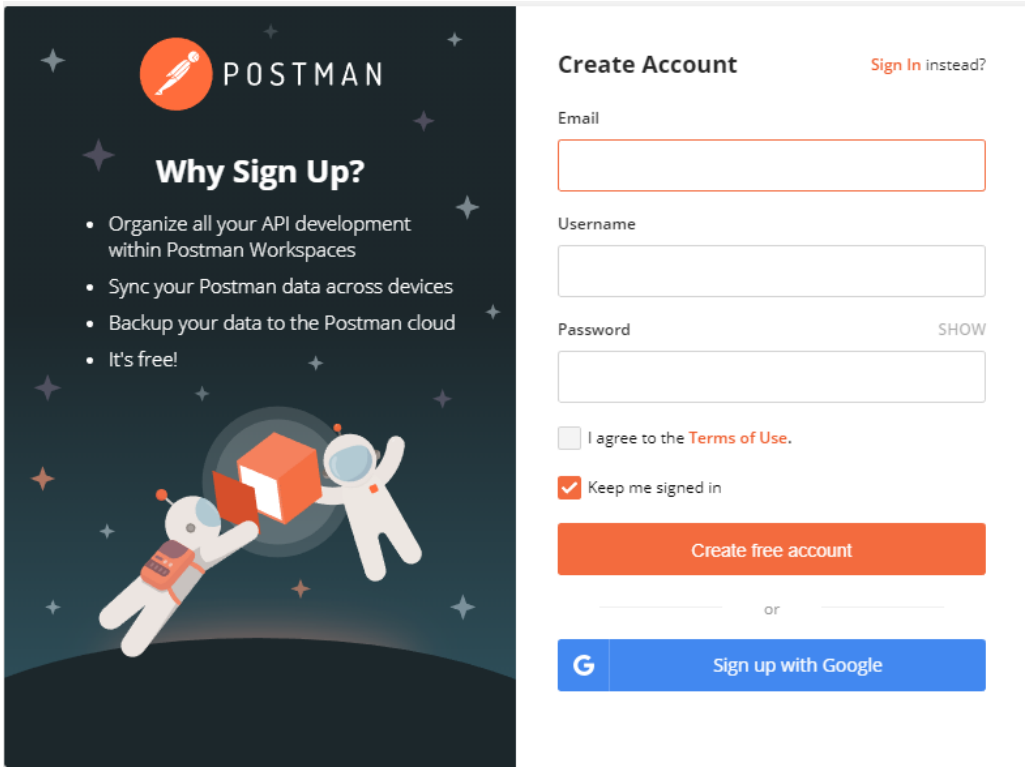

3. เลือกไฟล์ที่ Download มา เพื่อเริ่มติดตั้ง



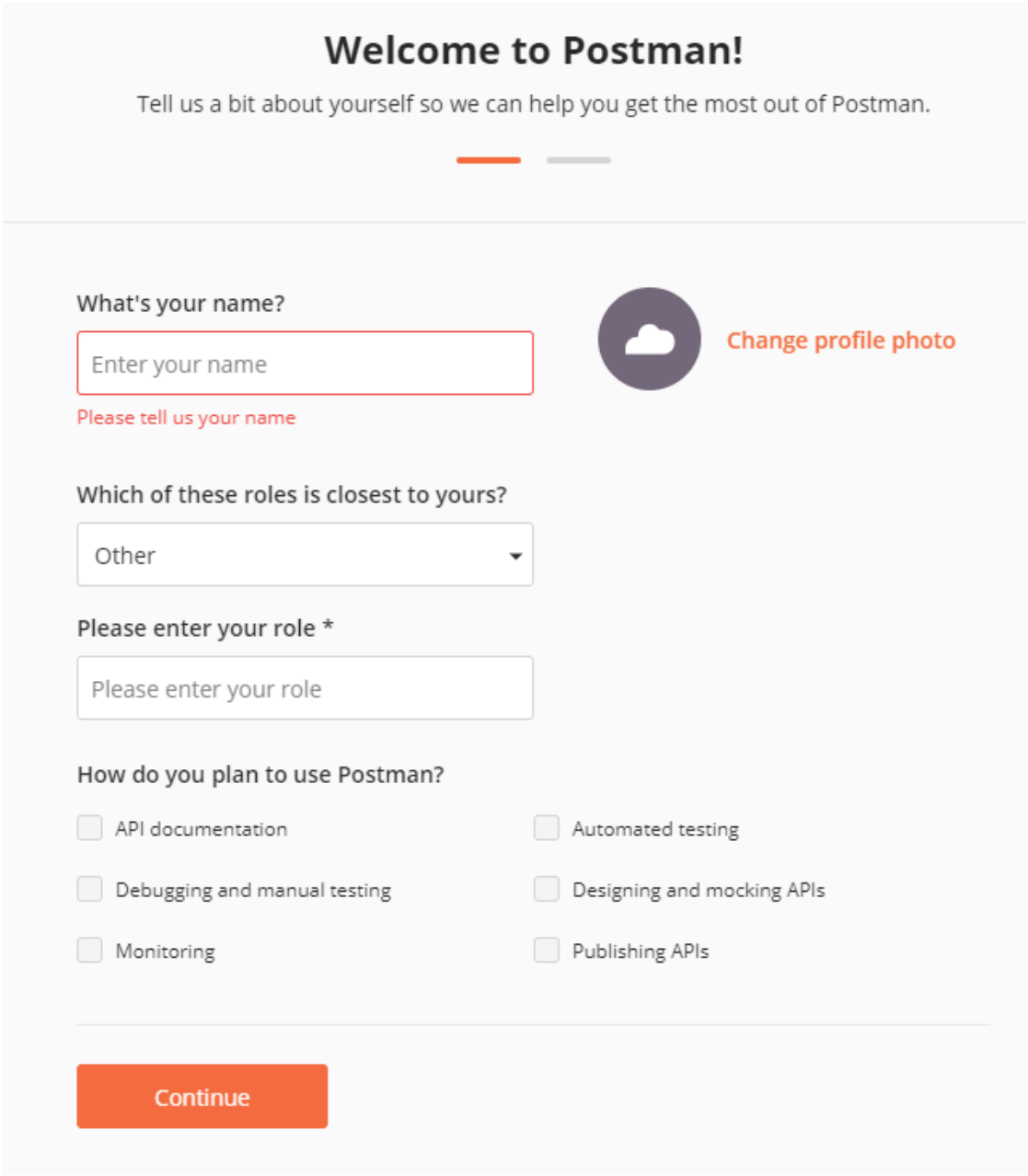
4. รอจนกว่าโปรแกรมจะติดตั้งเสร็จสมบูรณ์




5. เมื่อติดตั้งเสร็จ จะมีหน้าจอให้ Create Account กรอกข้อมูลให้ครบ เลือก I agree tp the Terms of Use แล้วเลือก Create free account(หากมี Account แล้วสามารถเลือก Sign in แทนได้เลย)

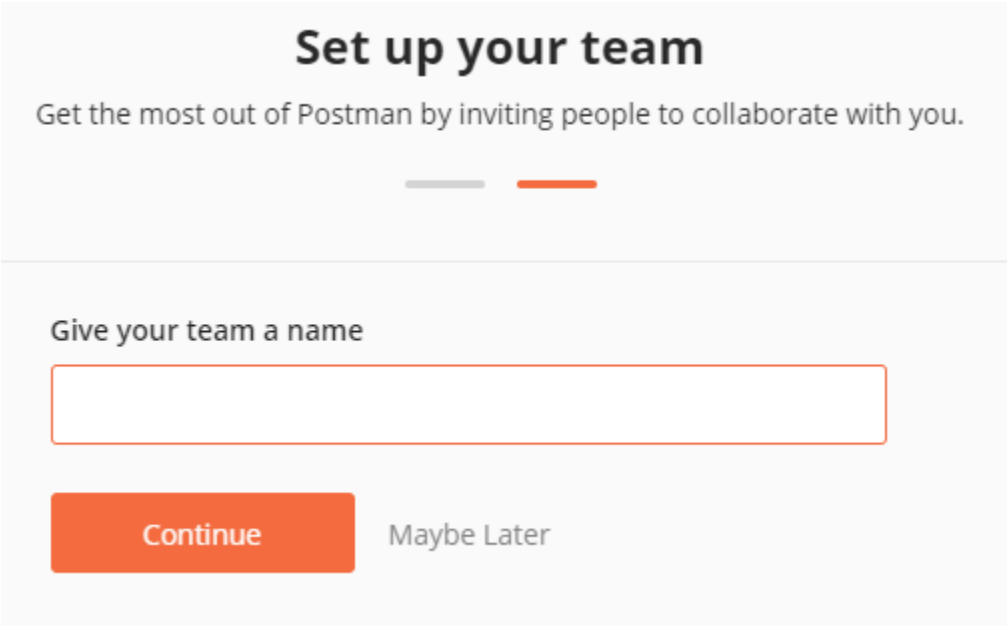


6. กรอกข้อมูลตามที่ปรากฏแล้วกด Continue

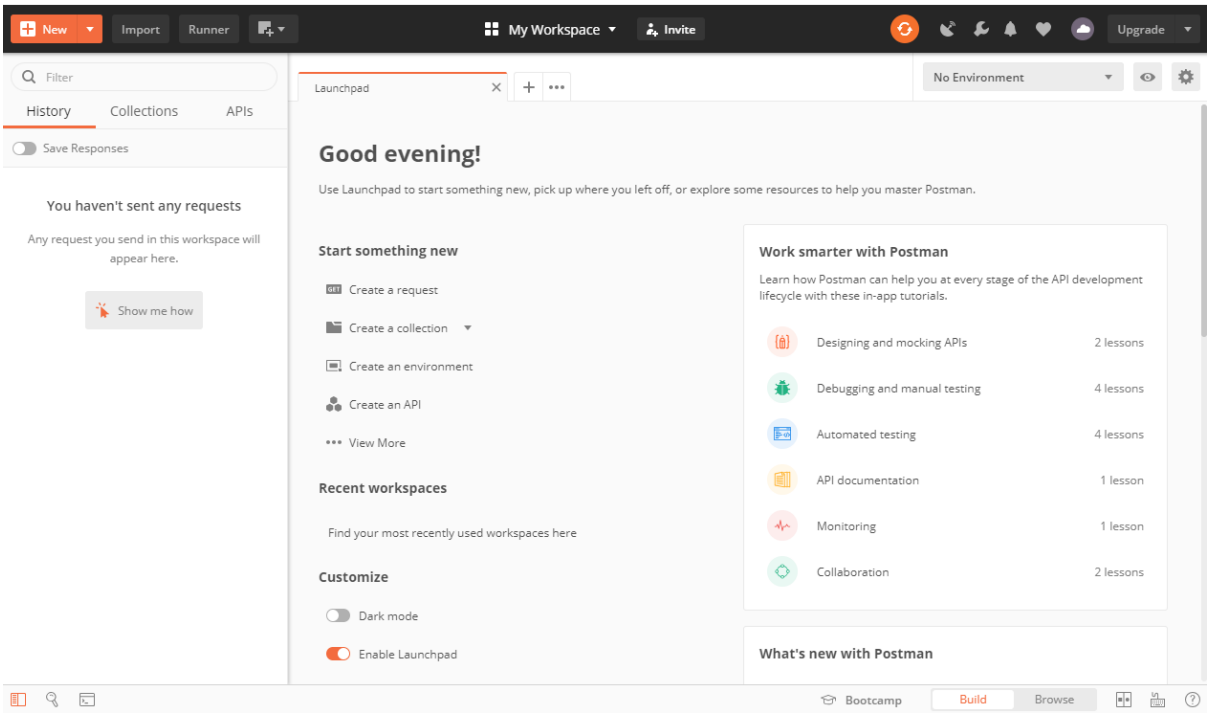


7. หากต้องการสร้างกลุ่ม ให้ตั้งชื่อกลุ่มของตัวเองแล้วกด Continue

   หากไม่ต้องการสร้างกลุ่มให้กด Maybe Later




8. เมื่อหน้าจอนี้ปรากฏเป็นการเสร็จสิ้นการติดตั้ง และ พร้อมเริ่มใช้งาน



ส่วนที่รับผิดชอบในการดำเนินงานวิจัย

นางสาวกรกนก คมวิภาค บทที่ 1,2,3,4,5 Code

นายกิจปมาน อิงคภาวัฏร บรรณานุกรม,แปลงานวิจัยที่เกี่ยวข้อง,แก้ไข Flow การทำงาน