



## รายงาน

เรื่อง การพยากรณ์ราคาดัชนีหุ้นเพื่อคาดการณ์ผลตอบแทนจากกลยุทธ์

## จัดทำโดย

นางสาวณัฐชา	สุภาพจันทร์	62090500406
นางสาวธัญนิชา	บวรวิวัฒน์ชัย	62090500411
นางสาวสุพิชชา	จำปาทอง	62090500424
นายสหัสวรรษ	ประคอง	62090500440
นางสาวเกวรินทร์	เจดีย์สถาน	62090500444
นายณพคุณ	อนันตกิจถาวร	62090500447

## เสนอ

รศ.ชูเกียรติ วรสุชีพ

ภาคเรียนที่ 1 ปีการศึกษา 2564

รายวิชา CSS 341 Introduction to Data Science

คณะวิทยาศาสตร์ ภาควิชาคณิตศาสตร์ สาขาวิทยาการคอมพิวเตอร์ประยุกต์

มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี

## บทคัดย่อ

การศึกษานี้มีวัตถุประสงค์เพื่อพยากรณ์ราคาดัชนีหุ้น ได้แก่ Dow Jones Industrial Average, Nikkei 225 Stock Average, Heng Seng Index และ SSE Composite Index และคำนวณผลตอบแทนจากการลงทุนด้วยกลยุทธ์การลงทุนถัวเฉลี่ยต้นทุนอย่างต่อเนื่องเป็นเวลา 60 เดือน ตั้งแต่เดือนมกราคม 2016 ถึง ธันวาคม 2021 โดยใช้เทคนิค Classification และ Regression ในการพยากรณ์ความเป็นไปได้ของราคาดัชนีหุ้นแล้วคำนวณเป็นอัตราผลตอบแทนของดัชนีหุ้น ผลการศึกษา การพยากรณ์ราคาดัชนีหุ้นพบว่า เทคนิค Classification

มีความสะดวกและเหมาะสมต่อการพยากรณ์ราคาดัชนีหุ้นมากที่สุดเนื่องจากมีการประเมินความแม่นยำที่ สะดวกและดีที่สุด และการคาดการณ์ผลตอบแทนของดัชนีหุ้น โดย ณ วันที่ 4 ธันวาคม 2021 ราคา Dow Jones Industrial Average, ราคา Nikkei 225 Stock Average, ราคา Heng Seng Index และ ราคา SSE Composite Index มีแนวโน้มที่จะปรับตัวลดลง ซึ่งพบว่าดัชนีหุ้นที่ควรลงทุนในวันนี้คือ Dow Jones Industrial Average, Nikkei 225 Stock Average, Heng Seng Index และ SSE Composite Index

**คำสำคัญ :** การพยากรณ์ดัชนีของหุ้น, ผลตอบแทนของการลงทุน

## บทนำ

การลงทุน หรือ การเล่นหุ้น คือการซื้อหุ้นของบริษัทที่จดทะเบียนในตลาดหลักทรัพย์ การเทรดหุ้นมีอยู่ด้วยกัน 2 ตลาด คือ ตลาดแรก การซื้อขายในตลาดแรก หรือการเทรดหุ้น IPO (Initial Public Offering) เกิดจากบริษัทต้องการระดมทุนจากนักลงทุนเพื่อขยายกิจการ โดยราคาหุ้นจะถูกกำหนดไว้ให้นักลงทุนมาจับจอง ในการซื้อหุ้น IPO นั้นจะต้องจองซื้อผ่านผู้จัดจำหน่ายเท่านั้น และตลาดรอง ซึ่งเป็นการซื้อขายหุ้นในตลาดหลักทรัพย์ ราคาของหุ้นในตลาดหลักทรัพย์จะเปลี่ยนแปลงตามผลการดำเนินการของบริษัท และสภาวะตลาดตามหลักของ demand supply

ปัจจุบันการลงทุนเป็นการเพิ่มมูลค่าของเงินในอีกช่องทางหนึ่ง ผู้คนส่วนใหญ่นิยมการลงทุนหลากหลายรูปแบบ ซึ่งการลงทุนการซื้อ-ขาย หุ้น เป็นช่องทางที่ได้รับความนิยม จากเงินทุนเป็นผลกำไรหรือผลตอบแทน โดยหลัก ๆ แล้ว มี 2 ประการ คือ Capital Gain กำไรจากส่วนต่างของราคาหุ้นที่เราซื้อเท่ากับราคาที่ขายไป และ เงินปันผล เงินส่วนแบ่งผลกำไรจากการดำเนินงานของบริษัทในระหว่างปีที่น่ามาจัดสรรให้กับผู้ถือหุ้น ในการลงทุนซื้อขายหุ้น จำเป็นที่จะต้องมีความรู้ด้านการเงินการลงทุน ความพร้อมทางด้านการเงินของตนเอง ศึกษารายละเอียดของหุ้นที่จะซื้อ ติดตามข่าวสารหุ้นอยู่เสมอ เพื่อลดความเสี่ยงที่จะเกิดขึ้นในระหว่างการลงทุน สาเหตุที่ทำให้ราคาหุ้นมีการเคลื่อนไหวขึ้นหรือลง เกิดจากความแตกต่างระหว่างความต้องการซื้อหุ้นและความต้องการขายหุ้นของคนในตลาด ถ้าช่วงใดความต้องการซื้อหุ้นของคนในตลาดมากกว่าความต้องการขายหุ้นจะส่งผลให้ราคาหุ้นปรับตัวสูงขึ้น แต่ช่วงใดราคามีความต้องการขายมากกว่าความต้องการซื้อราคาหุ้นก็จะปรับตัวลง

การศึกษาครั้งนี้จึงได้มีการเลือกนำเอาเทคนิค Classification และ Regression ในการพยากรณ์ราคาดัชนีหุ้น Dow Jones Industrial Average, Nikkei 225 Stock Average, Heng Seng Index และ SSE Composite Index ความเป็นไปได้ของราคาดัชนีหุ้นแล้วคำนวณเป็นอัตราผลตอบแทนของดัชนีหุ้น คำนวณผลตอบแทนจากการลงทุนด้วยกลยุทธ์การลงทุนถัวเฉลี่ยต้นทุนอย่างต่อเนื่องเป็นเวลา 60 เดือน ตั้งแต่เดือนมกราคม 2016 ถึง ธันวาคม 2021 เพื่อเปรียบเทียบผลตอบแทนจากราคาหุ้นที่เกิดขึ้นจริง เพื่อเป็นแนวทางในการตัดสินใจของนักลงทุนหรือผู้ที่สนใจลงทุนต่อไป

## วัตถุประสงค์

1. เพื่อพยากรณ์ราคาดัชนีหุ้น ได้แก่ Dow Jones Industrial Average, Nikkei 225 Stock Average, Heng Seng Index และ SSE Composite Index
2. คำนวณผลตอบแทนจากการลงทุนด้วยกลยุทธ์การลงทุนถัวเฉลี่ยต้นทุนอย่างต่อเนื่องเป็นเวลา 60 เดือน ตั้งแต่เดือนมกราคม 2016 ถึง ธันวาคม 2021

## วิธีดำเนินการ

### ข้อมูลและตัวแปร

การศึกษาครั้งนี้ใช้ข้อมูลตั้งแต่เดือนมกราคม 2016 ถึง ธันวาคม 2021 ของดัชนีหุ้นจำนวน 4 ตัว จากเว็บไซต์ของ Yahoo finance เพื่อสร้างตัวแบบพยากรณ์ แล้วนำไปคำนวณผลตอบแทนการลงทุนในอนาคตของดัชนีหุ้นแต่ละตัว โดยหุ้นที่ใช้ในการวิเคราะห์ ประกอบด้วย (1) Dow Jones Industrial Average หรือ DJI (2) Nikkei 225 Stock Average หรือ N225 (3) Heng Seng Index หรือ HSI และ (4) SSE Composite Index หรือ SSE

### การวิเคราะห์ข้อมูล

#### 1. Data Preparation

- 1) เริ่มด้วยการดึงข้อมูลดัชนีหุ้นมาจาก Yahoo โดยใช้ API โดยดึงข้อมูลตั้งแต่วันที่ 2016-01-01 ถึง 2021-12-31 เป็นเวลา 5 ปี
- 2) หา Technical Indicator โดยนำข้อมูลมาประมวลผลโดยใช้ Library ta-Lib

#### 2. Feature Selection

- 1) นำเอาราคาปิดของวันนี้และวันต่อมามาหาเปอร์เซ็นต์ความแตกต่างว่าหุ้นขึ้นหรือลง เพื่อใช้เป็น Target สำหรับ Classification
- 2) นำเอาราคาปิดของวันพรุ่งนี้มาใช้เป็น Target สำหรับ Regression
- 3) ใช้การ Feature Selection แบบ Recursive Feature Elimination หรือ RFE มาวิเคราะห์หา Feature สำหรับทั้ง Classification และ Regression โดยใช้ Target และ Re Target เป็นตัว Target ในการวิเคราะห์ข้อมูล โดย RFE จะเลือก Feature ที่ควรจะใช้มาเอง ไม่ได้จำกัดว่าต้องมีกี่ตัวแต่อย่างใด

#### 3. Evaluation

##### Classification

- 1) Accuracy

$$\frac{\text{correct predictions}}{\text{total predictions}} \times 100$$

- 2) Precision

$$\frac{TP}{(TP + FP)}$$

3) Recall

$$\frac{TP}{(TP + FN)}$$

4) F1-score

$$2 \times \frac{(Precision \times Recall)}{(Precision + Recall)}$$

### Regression

5) Root Mean Squared Error

$$\sqrt{((predictions - targets) \times 2).mean()}$$

6) R Squared

$$1 - \frac{(total\ sum\ of\ residuals)}{(total\ sum\ of\ squares)}$$

### 4. สร้างโมเดลในการพยากรณ์ข้อมูล

1) การพยากรณ์ด้วยวิธี Decision Tree เป็น model แบบ rule-based คือ สร้างกฎ if-else จากค่าของแต่ละ feature โดยไม่มีสมการมากำกับความสัมพันธ์ระหว่าง feature & target สิ่งที่สำคัญในการสร้าง Decision Tree คือ การเลือก split ค่า feature แต่ละครั้ง จะต้อง minimize ค่าของ cost functionให้น้อยที่สุด

วิธีการทำ Decision Tree คือ การค่อยๆ แบ่งข้อมูลออกทีละ 2 ส่วน (recursive binary split) จาก node ล่างสุดของ tree เรียกว่า root node และไล่ขึ้นมาเรื่อยๆ จนถึง leaf และทำ prediction ค่า target variable ด้วยวิธีการง่ายๆ คือ ใช้ค่า mean ของ target variable node โดยการ split ข้อมูลจาก root node จนถึง leaf node จะทำจนกว่าจะได้ condition ที่กำหนด

หลักการในการแบ่งข้อมูลในแต่ละ node สำหรับข้อมูลที่มี k feature และ n observation มีดังนี้ เลือก 1 feature จาก k feature มาทำ sorting ข้อมูล ด้วยค่าของ feature ที่เลือกมา หาจุดแบ่งข้อมูล (split point) ที่เป็นไปได้ทั้งหมด จากข้อมูล n observation สามารถหาจุดแบ่งข้อมูลที่เป็นไปได้ n-1 จุด สำหรับการแบ่งข้อมูลแต่ละแบบที่เป็นไปได้ คำนวณค่า Gini impurity ซึ่งเป็นการวัดความไม่บริสุทธิ์ หรือความไม่เพียวของ class ในแต่ละกลุ่มข้อมูลที่แบ่งตามแต่ละ split point สำหรับปัญหา classification แบบ binary ที่มี target variable เป็น 0 หรือ 1 การ split ที่ดี ควรจะได้กลุ่มข้อมูลออกมา 2 กลุ่มที่สามารถแยก class 0 กับ class 1 ออกมาได้ชัดเจนในแต่ละกลุ่ม ยังสามารถแบ่งแยก class ของ target variable ออกมาได้ดี ค่า Gini impurity ก็จะมีต่ำ เมื่อสิ้นสุดการ split แล้ว จะ predict ค่า target

$$G = \sum_{k=1}^K \hat{p}_{mk}(1 - \hat{p}_{mk}).$$

2) การพยากรณ์ด้วยวิธี Random Forest คือ แนวคิดของ Random Forest นี้คือการสร้างโมเดล ด้วยวิธีการ Decision Tree ขึ้นมาหลายๆ โมเดล โดยวิธีการสุ่มตัวแปร แล้วนำผลที่ได้แต่ละโมเดลมารวมกัน พร้อมนับจำนวนผลที่มีจำนวนซ้ำกันมากที่สุด สกัดออกมาเป็นผลลัพธ์สุดท้ายด้วยวิธีการ ของ Decision Tree คือเทคนิคที่ให้ผลลัพธ์ในลักษณะเป็นโครงสร้างของต้นไม้ภายในต้นไม้จะประกอบไปด้วยโหนด (node) ซึ่งแต่ละโหนดจะมีเงื่อนไขของคุณลักษณะเป็นตัว ทดสอบกิ่งของต้นไม้ (branch) แสดงถึงค่าที่เป็นไปได้ของ คุณลักษณะที่ถูกเลือกทดสอบ และใบ (leaf) เป็นสิ่งที่อยู่ล่างสุดของต้นไม้แสดงถึงกลุ่มของข้อมูล (class) ก็คือ ผลลัพธ์ที่ได้จากการพยากรณ์ ซึ่งข้อดีของวิธีการนี้คือให้ผลการพยากรณ์ที่ แม่นยำและเกิดปัญหา overfitting น้อย

3) การพยากรณ์ด้วยวิธี Logistic Regression คือ เป็นเทคนิคทางสถิติภายใต้การดูแลเพื่อค้นหาความ น่าจะเป็นของตัวแปรตาม (คลาสที่มีอยู่ในตัวแปร) และสร้างสมการคณิตศาสตร์เพื่อแบ่งแยก (classify) ข้อมูล ออกเป็น 2 กลุ่มคำตอบ

$$h_{\theta}(x) = \frac{1}{1 + e^{-\theta^T x}}$$

4) การพยากรณ์ด้วยวิธี XGBoost เป็น model ที่นำเอา Decision Tree มา train ต่อ ๆ กันหลาย ๆ tree โดยที่แต่ละ decision tree จะเรียนรู้จาก error ของ tree ก่อนหน้าทำให้ความแม่นยำ ในการทำ prediction จะแม่นยำมากขึ้นเรื่อยๆ เมื่อมีการเรียนรู้ของ tree ต่อเนื่องกันจนมีความลึกมากพอ และ model จะหยุดเรียนรู้เมื่อไม่เหลือ pattern ของ error จาก tree ก่อนหน้าให้เรียนรู้แล้ว ทั้ง Random Forrest และ XGBoost เป็น model แบบ ensemble คือ ใช้ model หลายๆ model มาประกอบกันเป็น model ที่ ซับซ้อน

5) การพยากรณ์ด้วยวิธี Linear Regression ก็คือ การ Fit ข้อมูลด้วย “เส้นตรง หรือ Linear” ใน การหาเส้นตรงที่จะใช้ในการสร้างโมเดลทำนายนี้จะต้องมีการคำนวณเพื่อหาฟังก์ชันเส้นตรงที่จะฟิต(พอดี)ไป กับข้อมูลได้ดีที่สุด ฟังก์ชันเส้นตรงพื้นฐาน ก็คือ

$$y = b_0 + b_1 * x_1$$

6) การพยากรณ์ด้วยวิธี Polynomial Regression คือ เป็นเทคนิคการพยากรณ์ที่พยายามอธิบายพฤติกรรมของข้อมูล โดยเรามีสมมติฐานที่ว่า ข้อมูลไม่ได้สัมพันธ์กันเป็นเส้นตรง ในการหาความสัมพันธ์เส้นตรง ของสมการ  $y = ax + b$ . สิ่งที่เราสนใจจริงๆ คือ การหาค่าสัมประสิทธิ์  $a$  ที่เหมาะสม

$$y = \alpha + \beta_1 x + \beta_1 x^2$$

### สรุปผลการวิจัย

การศึกษาในครั้งนี้ได้ใช้เทคนิค Classification และ Regression เพื่อพยากรณ์ราคาดัชนีหุ้นและคำนวณความเหมาะสมในการลงทุน โดยมีรายละเอียดดังนี้

การพยากรณ์ด้วยเทคนิค Classification 4 เทคนิค ได้แก่ Decision Tree, Random Forest, Logistic Regression และ XGBoost และเทคนิค Regression 2 เทคนิค ได้แก่ Linear Regression และ Polynomial Regression โดยใช้ข้อมูลตั้งแต่เดือนมกราคม 2016 ถึง ธันวาคม 2021 ของดัชนีหุ้นจำนวน 4 ตัว ได้แก่ (1) Dow Jones Industrial Average หรือ DJI (2) Nikkei 225 Stock Average หรือ N225 (3) Heng Seng Index หรือ HSI และ (4) SSE Composite Index หรือ SSE จากเว็บไซต์ของ Yahoo finance โดยศึกษาและคาดการณ์ราคาดัชนีหุ้น ได้ผลดังนี้

จากการศึกษาและคาดการณ์พบว่า ณ วันที่ 4 ธันวาคม 2021 ดัชนีหุ้น (1) Dow Jones Industrial Average หรือ DJI (2) Nikkei 225 Stock Average หรือ N225 (3) Heng Seng Index หรือ HSI และ (4) SSE Composite Index หรือ SSE มีแนวโน้มที่จะปรับตัวลดลงซึ่งเหมาะแก่การซื้อหรือลงทุน โดยมีเงื่อนไขว่า หากการคาดการณ์ของเทคนิคทั้งหมดเท่ากันไม่สามารถบอกได้ว่าราคาจะขึ้นหรือลง ทางระบบจะทำการประเมินผลความแม่นยำ ซึ่งเทคนิคที่มีความแม่นยำน้อยที่สุดจะถูกตัดออกและต่อมาจะทำการคาดการณ์ครั้งใหม่เพื่อให้ได้คำตอบว่า ดัชนีหุ้นทั้ง 4 ตัวนี้ควรลงทุนหรือไม่ในวันนี้

### อภิปรายผลการวิจัย

การศึกษานี้มีวัตถุประสงค์หลักเพื่อทำนายราคาดัชนีหุ้นและคำนวณผลตอบแทนจากการลงทุน ด้วยกลยุทธ์การลงทุนถัวเฉลี่ยต้นทุนอย่างต่อเนื่องเป็นเวลา 60 เดือน ตั้งแต่เดือนมกราคม 2016 ถึง ธันวาคม 2021 โดยทำการดึงข้อมูลดัชนีมาจากรายการ Yahoo finance หลังจากนั้นพยากรณ์ราคาหุ้นรายตัว ได้แก่ Dow Jones Industrial Average, Nikkei 225 Stock Average, Heng Seng Index และ SSE Composite Index ด้วยเทคนิค Classification และ Regression แล้วจึงคำนวณเป็นอัตราผลตอบแทนของดัชนีหุ้น สำหรับ Classification จะนำเอาราคาปิดของวันปัจจุบันและวันต่อไปมาหาเปอร์เซ็นต์ความแตกต่างว่าหุ้นขึ้นหรือลง เพื่อใช้เป็น Tagget ส่วน Regression จะนำเอาราคาปิดของวันถัดไป 1 วันมาใช้เป็น Target หลังจากนั้นใช้ Feature Selection แบบ Recursive Feature Elimination หรือ RFE มาวิเคราะห์หา Feature สำหรับทั้ง Classification และ Regression โดยใช้ Target (สำหรับ Classification) และ Re Target (สำหรับ Regression) เป็นตัว Target ในการวิเคราะห์ข้อมูล โดย RFE จะเลือก Feature ที่ควรจะใช้มาเอง ไม่มีการ

จำกัดว่าต้องมีกี่ตัวแต่อย่างไร เมื่อทำการเลือก Feature ได้แล้วจะทำการสร้างโมเดลในการพยากรณ์ข้อมูล ซึ่งจะมี 6 วิธี (1) การพยากรณ์ด้วยวิธี Decision Tree อยู่ในกลุ่มของ Classification (2) การพยากรณ์ด้วยวิธี Logistic Regression อยู่ในกลุ่มของ Classification (3) การพยากรณ์ด้วยวิธี Random Forest อยู่ในกลุ่มของ Classification (4) การพยากรณ์ด้วยวิธี XGBoost อยู่ในกลุ่มของ Classification (5) การพยากรณ์ด้วยวิธี Linear Regression อยู่ในกลุ่มของ Regression และ (6) การพยากรณ์ด้วยวิธี Polynomial Regression อยู่ในกลุ่มของ Regression โดยผลการศึกษาพบว่าเทคนิคที่เหมาะสมต่อการพยากรณ์มากที่สุด คือ เทคนิคของ Classification เนื่องจากการประเมินความแม่นยำที่ สะดวกและดีที่สุด โดยผลจากการศึกษาและคาดการณ์พบว่า ณ วันที่ 4 ธันวาคม 2021 ดัชนีหุ้น (1) Dow Jones Industrial Average หรือ DJI (2) Nikkei 225 Stock Average หรือ N225 (3) Heng Seng Index หรือ HSI และ (4) SSE Composite Index หรือ SSE มีแนวโน้มที่จะปรับตัวลดลงซึ่งเหมาะแก่การซื้อหรือลงทุน ณ วันนั้น ถ้าหากเกิดเหตุการณ์ที่การคาดการณ์ของเทคนิคทั้งหมดเท่ากันไม่สามารถบอกได้ว่าราคาจะขึ้นหรือลง ทางระบบจะทำการประเมินผลความแม่นยำ ซึ่งเทคนิคที่มีความแม่นยำน้อยที่สุดจะถูกตัดออกและต่อมาจะทำการคาดการณ์ครั้งใหม่เพื่อให้ได้คำตอบว่า ดัชนีหุ้นทั้ง 4 ตัวนี้ควรลงทุนหรือไม่ในวันนี้

### ข้อเสนอแนะ

ข้อเสนอแนะสำหรับการศึกษารั้งต่อไป คือ ควรมีการเปรียบเทียบเทคนิคพยากรณ์ด้วยวิธีการอื่น ๆ เช่น AdaBoost regression เป็นต้น เพื่อให้สามารถเปรียบเทียบผลตอบแทนจากการลงทุนในหุ้นอย่างเหมาะสมและมีความแม่นยำสูง



## เอกสารอ้างอิง

Bex T. **Powerful Feature Selection with Recursive Feature Elimination (RFE) of Sklearn.**

[Online]. 2018. Available from: <https://towardsdatascience.com/powerful-feature-selection-with-recursive-feature-elimination-rfe-of-sklearn-23efb2cdb54e> [27 November 2021]

Scikit learn. **Sklearn.ensemble.RandomForestClassifier.** [Online]. 2017. Available from:

<https://scikitlearn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html> [28 November 2021]

Prashanth Saravanan. **Understanding Loss Functions in Machine Learning.** [Online]. 2021.

Available from : <https://www.section.io/engineering-education/understanding-loss-functions-in-machine-learning/#loss-functions-for-regression> [28 November 2021]

Avinash Navlani. **Understanding Logistic Regression in Python.** [Online]. 2019. Available

from : <https://www.datacamp.com/community/tutorials/understanding-logistic-regression-python> [30 November 2021]

Saishruthi Swaminathan. **Logistic Regression — Detailed Overview.** [Online]. 2018. Available

from : <https://towardsdatascience.com/logistic-regression-detailed-overview-46c4da4303bc> [3 December 2021]

Witchapong Daroontham. **รู้จัก Decision Tree, Random Forest, และ XGBoost!!! — PART 1.**

[ออนไลน์]. 2018. แหล่งที่มา : <https://medium.com/@witchapongdaroontham/รู้จัก-decision-tree-random-forrest-และ-xgboost-part-1-cb49c4ac1315> [30 พฤศจิกายน 2021]

Scikit learn. **Metrics and scoring: quantifying the quality of predictions.** [Online]. 2017.

Available from : [https://scikit-learn.org/stable/modules/model\\_evaluation.html](https://scikit-learn.org/stable/modules/model_evaluation.html) [1 December 2021]

Boom626. **Confusion Matrix.** [ออนไลน์]. 2019. แหล่งที่มา :

[https://medium.com/@mirthful\\_sunset\\_cattle\\_231/confusion-matrix-48cc396b1b58](https://medium.com/@mirthful_sunset_cattle_231/confusion-matrix-48cc396b1b58) [3 ธันวาคม 2021]