

## โครงการวิศวกรรม (Engineering Project)

ชื่อหัวข้อเรื่อง(ไทย) : การประมาณค่าใช้จ่ายผลกระทบจากตลาดด้วยวิธีการ  
เรียนรู้ของเครื่อง

(อังกฤษ) : Market Impact Cost Estimation Using Machine  
Learning Algorithm

ผู้จัดทำ 1. นายโชควัฒน์ ศิลป์มานะกิจ รหัสประจำตัว 5810505937

อาจารย์ที่ปรึกษาโครงการ ผศ.ดร. อัครพงศ์ พชรรุ่งเรือง

อาจารย์ที่ปรึกษาร่วม ผศ.ดร. ยอดเยี่ยม ทิพย์สุวรรณ

ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเกษตรศาสตร์  
ปีการศึกษา 2561

ลงนามอาจารย์ที่ปรึกษาโครงการ .....

วันที่ .....

# บทที่ 1

## บทนำ

### 1.1 ความเป็นมาและความสำคัญ

ค่าใช้จ่ายผลกระทบจากตลาดหรือ Market Impact Cost นั้นเป็นค่าใช้จ่ายที่เกิดขึ้นจากการซื้อขายหลักทรัพย์ ซึ่งค่าดังกล่าวจะมีผลกระทบต่อประสิทธิภาพในการทำกำไรได้ค่อนข้างมากในผู้ลงทุนที่มีการซื้อขายหลักทรัพย์ในปริมาณมาก เช่น กองทุน ผู้บริหารพอร์ตการลงทุน เป็นต้น ค่าใช้จ่ายผลกระทบจากตลาด หรือ Market Impact Cost จึงถูกให้ความสำคัญ และติดตามอย่างใกล้ชิดจาก กองทุน และผู้บริหารพอร์ตการลงทุน โดยแต่เดิมจะมีการประมาณค่าดังกล่าว จากโมเดลสมการทางคณิตศาสตร์

สืบเนื่องจากในปัจจุบันคอมพิวเตอร์เข้ามามีบทบาทสำคัญในด้านต่างๆของชีวิตประจำวัน หนึ่งในนั้นคือการนำมาใช้วิเคราะห์ข้อมูลต่างๆ อีกทั้งยังสามารถนำมาประมาณ หรือ ทำนายเหตุการณ์หรือค่าได้จากวิธีที่ได้รับความนิยมอย่างมากในปัจจุบัน คือ การเรียนรู้ของเครื่อง หรือ Machine Learning

ผู้จัดทำจึงมีความสนใจในการใช้วิธีการเรียนรู้ของเครื่องมาสร้างโมเดลประมาณค่า Market Impact Cost แทนที่โมเดลสมการคณิตศาสตร์เดิม

### 1.2 วัตถุประสงค์และเป้าหมาย

1.2.1. ศึกษาปัจจัยที่มีผลต่อ Market Impact Cost

1.2.2. สร้างโมเดลจากวิธีการเรียนรู้ของเครื่อง เพื่อใช้ประมาณ Market Impact Cost

1.2.3. ปรับโมเดลดังกล่าวให้สามารถประมาณค่าให้แม่นยำ

### 1.3 ขอบเขตการดำเนินงาน

1.3.1. ได้โมเดลที่สามารถใช้ประมาณ Market Impact Cost ของในตลาดประเทศฮ่องกง(SEHK)

## 1.4 ระยะเวลาในการดำเนินงาน

ประกอบไปด้วย 2 ส่วนการดำเนินงาน

1.4.1 ในช่วงเดือนสิงหาคม – ธันวาคม พ.ศ. 2561 เป็นส่วนของการเรียนรู้การทำโครงการ และวางแผนการทำโครงการ โดยจะมีการศึกษาความเป็นไปได้ของโครงการที่จะทำร่วมกับอาจารย์ที่ปรึกษา ร่วมทั้งศึกษาหาความรู้ ทดลอง และหาทฤษฎีที่เกี่ยวข้องในการทำโครงการเพื่อกำหนดขอบเขต และจุดประสงค์ของการทำโครงการ

1.4.2 ในช่วงเดือนมกราคม – พฤษภาคม พ.ศ. 2562 เป็นส่วนของการดำเนินโครงการตามที่วางแผนไว้ในช่วงแรก ซึ่งจะเป็นการลงมือทดสอบ และพัฒนาโปรแกรมเพื่อให้เป็นไปตามวัตถุประสงค์ที่วางแผนไว้ รวมถึงการประเมินโครงการ

## 1.5 เครื่องมือที่ใช้ในการดำเนินงาน

### 1.5.1 ฮาร์ดแวร์(Hardware)

- คอมพิวเตอร์พกพา MacBook Pro(Retina, 13-inch, Early 2015)  
ความเร็ว 2.7 GHz intel Core i5, หน่วยความจำ 8 GB 1867 MHz  
DDR3, Graphics Intel Iris Graphics 6100 1536 MB

### 1.5.2 ซอฟต์แวร์(Software)

- Jupyter Notebook
- Python 3.6.1
- Visual Studio Code
- Microsoft Excel

### 1.5.3 ข้อมูล(Data)

- SEHK Stock Data Time Frame minutely From 2012-2018

## 1.6 ปัญหา/อุปสรรค/โอกาส

1.6.1 โมเดลและสมการคณิตศาสตร์ มีความซับซ้อนและทำความเข้าใจได้ยาก

1.6.2 เป็นโอกาสในการเรียนรู้และประยุกต์ใช้โมเดลทางคณิตศาสตร์ เพื่อสร้างโปรแกรม

### 1.6.3 เป็นโอกาสในการเรียนรู้และใช้ Machine Learning กับข้อมูลจริง

#### 1.6.4 เรียนรู้แนวทางการทำวิจัย การจัดการข้อมูล และประเมินผลโมเดล

## 1.7 ระยะเวลาการดำเนินการ

| รายละเอียดการดำเนินการ  | ปี พ.ศ. 2561 |   |   |   |         |   |   |   |        |   |   |   |           |   |   |   |         |   |   |   |
|---|--------------|---|---|---|---------|---|---|---|--------|---|---|---|-----------|---|---|---|---------|---|---|---|
|   | สิงหาคม      |   |   |   | กันยายน |   |   |   | ตุลาคม |   |   |   | พฤศจิกายน |   |   |   | ธันวาคม |   |   |   |
|   | 1            | 2 | 3 | 4 | 1       | 2 | 3 | 4 | 1      | 2 | 3 | 4 | 1         | 2 | 3 | 4 | 1       | 2 | 3 | 4 |
| ศึกษาหัวข้อที่สนใจ  | ■            | ■ | ■ | ■ |         |   |   |   |        |   |   |   |           |   |   |   |         |   |   |   |
| หาอาจารย์ที่ปรึกษา  |              | ■ | ■ | ■ |         |   |   |   |        |   |   |   |           |   |   |   |         |   |   |   |
| ศึกษางานวิจัยที่เกี่ยวข้อง  |              |   |   |   | ■       | ■ |   |   |        |   |   |   |           |   |   |   |         |   |   |   |
| ศึกษาทฤษฎีที่เกี่ยวข้อง   |              |   |   |   |         | ■ | ■ | ■ |        |   |   |   |           |   |   |   |         |   |   |   |
| ทำความเข้าใจแบบจำลองข้อมูล  |              |   |   |   |         |   |   |   | ■      | ■ | ■ | ■ | ■         | ■ |   |   |         |   |   |   |
| ทำความเข้าใจโครงสร้างข้อมูลและประมวลผลข้อมูลให้เหมาะสมก่อนนำเข้าโมเดล |              |   |   |   |         |   |   |   |        | ■ | ■ | ■ | ■         | ■ |   |   |         |   |   |   |
| ศึกษาปัจจัยที่ส่งผลต่อผลลัพธ์โมเดลและทดลองข้อมูลกับโมเดลหลายๆ โมเดล   |              |   |   |   |         |   |   |   |        |   | ■ | ■ | ■         | ■ | ■ | ■ |         |   |   |   |
| หาโมเดลที่เหมาะสมสำหรับประมาณผลลัพธ์ข้อมูลที่เหมาะสม                  |              |   |   |   |         |   |   |   |        |   |   |   |           |   | ■ | ■ | ■       | ■ | ■ | ■ |

| รายละเอียดการดำเนินการ                        | ปี พ.ศ. 2562 |   |   |   |            |   |   |   |        |   |   |   |        |   |   |   |         |   |   |   |
|---|--------------|---|---|---|------------|---|---|---|--------|---|---|---|--------|---|---|---|---------|---|---|---|
|   | มกราคม       |   |   |   | กุมภาพันธ์ |   |   |   | มีนาคม |   |   |   | เมษายน |   |   |   | พฤษภาคม |   |   |   |
|   | 1            | 2 | 3 | 4 | 1          | 2 | 3 | 4 | 1      | 2 | 3 | 4 | 1      | 2 | 3 | 4 | 1       | 2 | 3 | 4 |
| หาโมเดลที่เหมาะสมสำหรับทำนายผลลัพธ์ที่เหมาะสม |              |   |   |   |            |   |   |   |        |   |   |   |        |   |   |   |         |   |   |   |
| ปรับปรุงโมเดลพารามิเตอร์                      |              |   |   |   |            |   |   |   |        |   |   |   |        |   |   |   |         |   |   |   |
| ประเมินผล                                     |              |   |   |   |            |   |   |   |        |   |   |   |        |   |   |   |         |   |   |   |
| ปรับปรุงและพัฒนาให้ได้ผลลัพธ์ที่พึงพอใจ       |              |   |   |   |            |   |   |   |        |   |   |   |        |   |   |   |         |   |   |   |
| สรุปผลการดำเนินงาน                            |              |   |   |   |            |   |   |   |        |   |   |   |        |   |   |   |         |   |   |   |
| จัดทำเอกสาร                                   |              |   |   |   |            |   |   |   |        |   |   |   |        |   |   |   |         |   |   |   |
| จัดทำสื่อนำเสนอ                               |              |   |   |   |            |   |   |   |        |   |   |   |        |   |   |   |         |   |   |   |

## 1.8 ประโยชน์ที่คาดว่าจะได้รับ

- 1.8.1 เข้าใจในกระบวนการเพื่อให้ได้มาซึ่งโมเดลสำหรับประมาณหรือทำนายค่า
- 1.8.2 ความชำนาญในการใช้งาน library และภาษาโปรแกรมในการจัดการข้อมูล

## 1.9 แนวทางการประเมิน

- 1.9.1 วัดความแม่นยำของโมเดลโดยวัดจากความคลาดเคลื่อนของค่าที่ทำนายโดยใช้มาตรวัดต่างๆ เช่น MAE, RMS

## บทที่ 2

### ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

#### 2.1 ทฤษฎีที่เกี่ยวข้อง

##### 2.1.1 Market Impact Cost

คือ trading cost ที่เกิดขึ้นจากผลต่างระหว่างราคาที่เราคาดหวังว่าจะทำการซื้อขายกับราคาที่เกิดการซื้อขายจริง ซึ่งตัว Market Impact Cost นั้นสามารถบ่งบอกถึงสภาพคล่องของหุ้นตัวนั้นได้ เช่น หุ้นที่มีสภาพคล่องต่ำจะส่งผลให้ Market Impact Cost สูง และหุ้นที่มีสภาพคล่องสูงจะส่งผลให้ Market Impact Cost ต่ำ

##### 2.1.2 I-Star Model

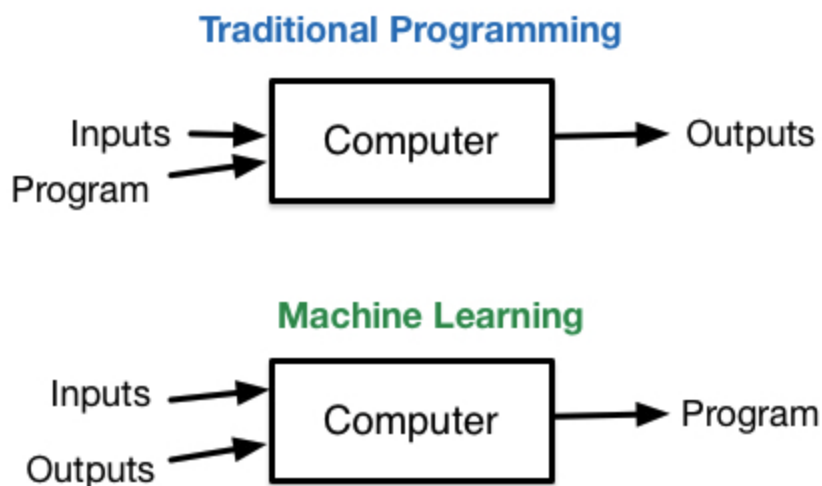
คือ model ที่ใช้ในการประมาณ Market Impact Cost ซึ่งคิดค้นโดย Kissell กับ Malamut เมื่อปี 1998 โดย model จะประมาณ Market Impact Cost จากขนาดของ order กับภาพรวมสถานะตลาด ณ ช่วงเวลานั้น ซึ่งตัว model มีสมการดังนี้

$$\begin{aligned} MI &= b_1 I^* \cdot POV^{a_1} + (1 - b_1) I^* \\ I^* &= a_1 \cdot Size^{a_2} \cdot Vol^{a_3} \end{aligned}$$

- MI คือ Market Impact Cost
- POV(Percentage Of Volume) คือ ค่าที่เกิดจากสัดส่วนขนาดของ order ที่จะมีการซื้อขายกับ volume การซื้อขายของตลาดนั้นๆ ซึ่งเป็นค่าที่สะท้อนถึงสภาพคล่องของตลาด
- Size คือ การ normalize trading order ด้วยขนาดของค่าเฉลี่ย trading order
- $b_1$ ,  $a_1$ ,  $a_2$ ,  $a_3$ ,  $a_4$  คือ parameter ของ model

### 2.1.3 Machine Learning

Machine Learning หรือ “การเรียนรู้ของเครื่องจักร” เป็นการเปลี่ยนแนวคิดของการเขียนโปรแกรมจากเดิม คือ การเขียนโปรแกรมแบบดั้งเดิม(Traditional Programming) เป็นการหา output จาก input เช่น  $y = x + 1$  ซึ่งจากโปรแกรมนี้ สิ่ง โปรแกรมเมอร์ต้องทำ คือ การเขียนโปรแกรม  $x + 1$  เพื่อให้คอมพิวเตอร์คำนวณหาค่า  $y$  มาให้ แต่ Machine Learning จะเป็นการที่โปรแกรมเมอร์ป้อนข้อมูล  $y$  กับ  $x$  จำนวนมากให้คอมพิวเตอร์ เพื่อให้คอมพิวเตอร์ คำนวณผลลัพธ์ค่า output ออกมาเป็นสมการ  $x + 1$  แทน หรือก็คือให้คอมพิวเตอร์สร้างโปรแกรมจาก input และ output แทน ดังภาพ



Machine Learning แบ่งเป็น 2 ประเภท คือ

1. Supervised Learning หรือการเรียนรู้แบบมีผู้สอน ในการเรียนรู้แบบนี้ โปรแกรมเมอร์จะสอน Machine ให้ว่า Inputs แบบนี้ แล้วจะได้ Output แบบไหน เช่น เราต้องการให้ Machine เรียนรู้เรื่อง เด็กนักเรียนว่าคนไหน สอบตก หรือ สอบผ่าน โปรแกรมเมอร์จะ ให้ Machine เรียนรู้ input เป็นคะแนน และ output เป็นสอบผ่านหรือสอบตก เพื่อให้ Machine เรียนรู้ว่าคะแนนเท่าใด เด็กจะสอบผ่าน และ คะแนนเท่าใดเด็กจะสอบตก
2. Unsupervised Learning หรือการเรียนรู้แบบไม่มีผู้สอน เป็นการป้อนข้อมูล input ให้ Machine แล้วให้ Machine ประมวลผลบางอย่างให้ ซึ่งโดยส่วนใหญ่จะเป็นการแบ่งกลุ่ม

#### 2.1.4 Pandas

เป็น library ใน Python ซึ่งช่วยอำนวยความสะดวกในการจัดการข้อมูล นิยมมากในการทำ Data Cleaning และ Data Wrangling และ Pandas นั้นสามารถนำไปใช้ประโยชน์คู่กับ Package อื่น เช่น เอาไปเตรียมข้อมูลก่อนทำ model ต่างๆใน sklearn ได้ด้วย

#### 2.1.5 SciPy

เป็น library ใน Python สำหรับการคำนวณเชิงวิทยาศาสตร์ ซึ่งมี module ที่นิยม เช่น optimization, linear algebra, integration, interpolation, FFT เป็นต้น



### 2.1.6 NumPy

เป็น library ใน Python สำหรับการจัดการข้อมูลชุด(Array) ขนาดใหญ่และเมทริกซ์ อีกทั้งยังสามารถคำนวณเชิงคณิตศาสตร์และวิทยาศาสตร์ได้อีกด้วย

### 2.1.7 Scikit-Learn

เป็น library ใน Python สำหรับที่ใช้สำหรับ Machine Learning สามารถใช้ทำ classification, regression และ clustering และมี Machine Learning Algorithm ที่นิยมให้ใช้ได้สะดวก เช่น Support Vector Machine, Random Forests, Gradient Boosting และ K-means อีกทั้งสามารถใช้งานร่วมกับ Pandas และ NumPy ได้ดี อีกด้วย

### 2.1.8 Matplotlib & Seaborn

Matplotlib เป็น library python ใช้สำหรับการ plot กราฟ 2D ได้อย่างง่ายและรวดเร็ว ตัว library มีหลาย module ให้ใช้งานเช่นการ plot histograms, power spectra, bar charts, errorcharts, scatterplots เป็นต้น

Seaborn เป็น library ที่มีพื้นฐานจาก Matplotlib แต่เพิ่มในส่วน ความสวยงาม เพื่อให้การแสดงผลข้อมูลเป็นไปได้อย่างชัดเจน

## 2.2 งานวิจัยที่เกี่ยวข้อง

### 2.2.1 Predicting Market Impact Cost Using Nonparametric Models

เป็นงานวิจัยเกี่ยวกับการสร้าง model ที่เป็น nonparametric เช่น Neural Network, Bayesian Neural Network, Gaussian Process เพื่อคาดการณ์ค่า Market Impact Cost ที่จะเกิดขึ้นกับ US stocks ซึ่งตัว model สร้างจาก data ที่เป็น transaction order ของหุ้น US

### 2.2.2 A Practical Framework for Estimating Transaction Costs and Developing Optimal Trading Strategies to Achieve Best Execution

เป็นงานวิจัยที่ศึกษาปัจจัยที่มีผลต่อ Cost ที่ต้องจ่ายเพิ่ม นอกเหนือจากราคาหุ้นและค่าทำธุรกรรมซึ่งเป็นผลจากสภาวะตลาด ซึ่งก็คือ Market Impact Cost

## Reference

- [1] Milind Paradkar Market Impact Cost  
จาก <https://www.quantinsti.com/blog/market-impact-cost/>  
(วันที่สืบค้น 28 กันยายน 2561)
- [2] Robert Kissell. หนังสือ The Science of Algorithmic Trading  
(วันที่สืบค้น 18 ตุลาคม 2561)
- [3] Kissell R, Glantz M, Malamut R. Optimal trading strategies: quantitative approaches for managing market impact and trading risk. Amacom; 2003.  
(วันที่สืบค้น 3 ตุลาคม 2561)
- [4] Saerim Park, Jaewook Lee, Youngdoo Son. Predicting Market Impact Cost Using Nonparametric Models  
จาก <https://journals.plos.org/plosone/article?id=10.1371>  
(วันที่สืบค้น 10 ตุลาคม 2561)
- [5] Dumpdatasci. Machine Learning คืออะไร? (พร้อมตัวอย่างในชีวิตจริง) จาก <https://medium.com/@dumpdatasci.th/>  
(วันที่สืบค้น 14 พฤศจิกายน 2561)
- [6] Perth@GrowthBee.com. คู่มือ Data Wrangling โดย Pandas ฉบับสมบูรณ์  
จาก <https://www.growthbee.com/cheatsheet-pandas/>  
(วันที่สืบค้น 14 พฤศจิกายน 2561)
- [7] Wanaphong. การจัดการข้อมูลด้วย Numpy เบื้องต้น  
จาก <https://python3.wannaphong.com/2015/05/numpy.html>  
(วันที่สืบค้น 14 พฤศจิกายน 2561)

[8] Wanaphong. การทำ Machine Learning ด้วย Scikit-Learn  
จาก <https://python3.wannaphong.com/2016/01/machine-learning-scikit-learn.html>

(วันที่สืบค้น 14 พฤศจิกายน 2561)

[9] <https://en.wikipedia.org/wiki/Scikit-learn>

(วันที่สืบค้น 14 พฤศจิกายน 2561)

[10] <https://matplotlib.org/>

(วันที่สืบค้น 15 พฤศจิกายน 2561)

[11] <http://seaborn.pydata.org/>

(วันที่สืบค้น 15 พฤศจิกายน 2561)