

การทำนายจุดวิกฤติบนดัชนีตลาดหลักทรัพย์

Critical Point Prediction of Stock Market Index

นาย สารัช รุจิรานุกซ์

sarach.r@ku.th

ผศ.ดร.ยอดเยี่ยม ทิพย์สุวรรณ

yvt@ku.ac.th

ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเกษตรศาสตร์

บทคัดย่อ

ปัจจุบันการลงทุนด้วยการใช้ระบบการซื้อขายอัตโนมัติกำลังเป็นที่นิยมอย่างแพร่หลาย ด้วย การนำวิธีการวิเคราะห์ต่างๆ มาใช้ในการสร้างระบบซื้อขายอัตโนมัติ เพื่อสร้างผลกำไรให้กับนักลงทุน อย่างไรก็ตามการซื้อขายนั้นจะดีที่สุดก็ต่อเมื่อเราได้ทำการซื้อที่จุดต่ำสุดและขายที่จุดสูงสุดซึ่งในบทความนี้ได้นำระบบการเรียนรู้ของเครื่องจักรมาช่วยในการตัดสินใจซื้อหรือขายหลักทรัพย์ที่มีประสิทธิภาพและผลกำไรสูง ซึ่งสามารถทำกำไรได้สูงกว่าตลาดถึง 20 %

Abstract

Nowadays, investing by using automated trading systems is becoming widely popular. Applying various methods are used to create a lot of systems to make profit for investors. However, the best action points are to buy at the lowest prices and sell at the highest prices. In this paper, we adapted machine learning model to determine whether price is the lowest or highest point. In order to obtain a good trading system and highly profitable which outperforms the markets over 20 %.

1. บทนำ

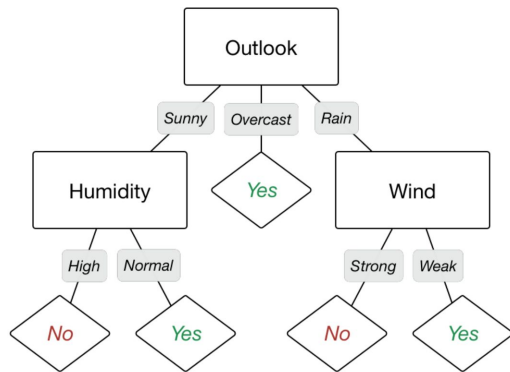
การลงทุนในหลักทรัพย์เป็นเรื่องที่น่าสนใจของนักลงทุนทั้งในและต่างประเทศ เพราะสามารถให้ผลตอบแทนที่สูงกว่าการฝากเงินไว้ในธนาคาร แต่เนื่องจากราคาหลักทรัพย์มีความผันผวนและไม่แน่นอน ทำให้นักลงทุนต้องคำนึงถึงความเสี่ยงที่อาจจะเกิดขึ้นกับการลงทุน ส่งผลให้นักลงทุนเกิดความจำเป็นที่จะต้องนำหลักการวิเคราะห์พยากรณ์ต่างๆ เช่น การวิเคราะห์ด้วย

ดัชนีชี้วัดทางเทคนิค ปัจจัยพื้นฐาน การวิเคราะห์ทางสถิติเชิงเวลา และการประเมินสถานการณ์ตามข่าวสาร เพื่อนำมาพยากรณ์ทิศทางแนวโน้มการเปลี่ยนแปลงราคาหลักทรัพย์ เพื่อลดความเสี่ยงในการลงทุนที่อาจจะเกิดขึ้น แต่ผลการพยากรณ์ดังกล่าวยังไม่สามารถที่จะใช้เป็นระบบระยะยาวได้ ซึ่งบทความนี้จะนำวิธีการวิเคราะห์ต่างๆมาต่อประยุกต์ใช้ เข้ากับระบบการเรียนรู้ของเครื่องจักรเพื่อที่จะทำการวิเคราะห์ข้อมูลได้ซับซ้อนมากขึ้น จากการเรียนรู้จากข้อมูลและสามารถทำการปรับปรุง ระบบได้ตลอดเวลาด้วยการให้ข้อมูลชุดใหม่เข้าไปกับระบบ เพื่อให้ได้ระบบที่มีความแม่นยำกับข้อมูลชุดใหม่ได้ตลอดเวลา

2. งานและทฤษฎีที่เกี่ยวข้อง

2.1 ต้นไม้การตัดสินใจ (Decision Tree)

ต้นไม้การตัดสินใจ (Decision Tree) เป็นเครื่องมือที่ใช้ในการใช้วิเคราะห์เหตุการณ์ หรือสถานการณ์เพื่อสร้างระบบการตัดสินใจได้อย่างรวดเร็วและแม่นยำ โดยจะมีลักษณะเป็นกราฟรูปต้นไม้ (tree model) ซึ่งแสดงที่ตั้งต้นที่มีรากและแตกแขนงเป็นกิ่งก้านออกมาจากรากต้นไม้ไปในทิศทางเดียว จนนำไปสู่ข้อสรุปสำหรับการตัดสินใจได้ ช่วยให้ง่ายต่อการสรุปการตัดสินใจที่มีความซับซ้อนให้ง่ายต่อความเข้าใจกับข้อมูลหลายๆ ตัวแปร โดยวิธีการสอน (training) ต้นไม้การตัดสินใจมีอยู่มากมาย โดยวิธีพื้นฐานวิธีหนึ่งซึ่งเป็นการค้นหาแบบละโมภ (greedy search) จากบนลงล่าง (top-down) ชื่อว่า ID3



2.1.1 เอนโทรปี (Entropy)

วิธีการของ ID3 จะเป็นการคำนวณหารากของต้นไม้ โดยจะทำการสร้างจากบนลงล่างด้วย ซึ่งจะเป็นการทำซ้ำๆ จนได้ค่าที่ดีที่สุด โดยจะเริ่มจากการหาค่าบอกความไม่บริสุทธิ์ของข้อมูลก่อน ซึ่งเรียกว่าเอนโทรปี (Entropy) โดยนิยามเอนโทรปีของต้นไม้การตัดสินใจในตัวในเซตของตัวอย่าง S คือ $E(S)$ ดังนี้

$$E(S) = - \sum_{j=1}^n p_S(j) \log_2 p_S(j)$$

เมื่อ

- S คือ ตัวอย่างที่ประกอบด้วยชุดของตัวแปรต้นและตัวแปรตามหลายๆกรณี
- $p_S(j)$ คือ อัตราส่วนของกรณีใน S ที่ตัวแปรตามหรือผลลัพธ์มีค่า j

ในกรณีที่ ผลลัพธ์เป็นแค่เพียงค่าตรรกะ (boolean) ใช่กับไม่ใช่เหมือนกับที่ยกมาตอนต้นของบทความนั้น จะมีเอนโทรปีคือ

$$E(S) = -p_{yes} \log_2(p_{yes}) - p_{no} \log_2(p_{no})$$

เมื่อพิจารณาเอนโทรปีแล้วจะเห็นว่าเอนโทรปีจะมีค่าอยู่ระหว่าง 0 กับ 1 โดยจะมีค่าเป็น 0 เมื่อมีผลลัพธ์เพียงแบบเดียว และจะมีค่ามากขึ้นเมื่อเริ่มมีค่าที่แตกต่างกันมากขึ้น

2.1.2 เกนความรู้ (Information Gain)

การที่จะหาตัวแปรที่เหมาะสมในการแบ่งข้อมูลนั้น สามารถระบุได้โดยการหาเกนความรู้ โดยตัวแปร A จะเป็นตัวแปรต้นที่เหมาะสมต่อการแบ่งการตัดสินใจได้ก็ต่อเมื่อสามารถแบ่งข้อมูลตัวอย่าง (Example) ออกเป็นชุดๆ มีจำนวนชุด

ตามจำนวนค่าของ A ที่เป็นไปได้เพื่อให้แต่ละกรณี (Instance) ในชุดนั้นมีค่า A เพียงค่าเดียวและค่าเฉลี่ยของเอนโทรปีของชุดข้อมูลที่ถูกแบ่งออก (partition) มานั้นต่ำที่สุด เรียกค่าคาดหวังของการลดลงของเอนโทรปีหลังจากข้อมูลถูกแบ่งด้วย A ว่าเกนความรู้ของ A นิยามโดย

$$Gain(S, A) = E(S) - \sum_{v=value(A)} \frac{|S_v|}{|S|} E(S_v)$$

เมื่อ

- S คือ ตัวอย่างที่ประกอบด้วยชุดของตัวแปรต้นและตัวแปรตาม
- E คือ เอนโทรปีของตัวอย่าง
- A คือ ตัวแปรต้นที่ต้องการพิจารณา
- $value(A)$ คือ เซตของค่าของ A ที่เป็นไปได้
- S_v คือ ตัวอย่างของ A ที่มีค่า v ทั้งหมด

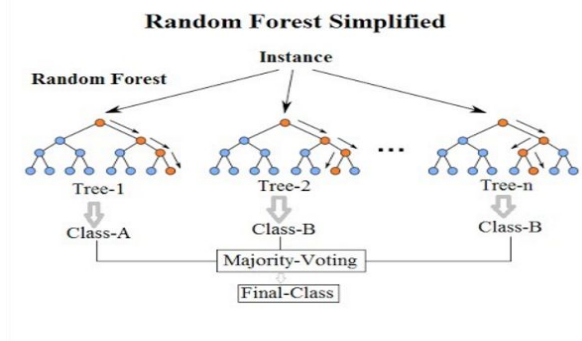
โดยเกนความรู้ที่มีค่ามากจะแสดงว่าเป็นการตัดสินใจที่ดีและมีความแม่นยำ

วิธีการสอน (training) ต้นไม้การตัดสินใจ

1. ทำการนำตัวแปรต้นไปคำนวณหาเกนความรู้
2. เลือกตัวแปรต้น ที่มีค่าเกนความรู้สูงที่สุด
3. ทำการสร้างต้นไม้ที่มีบัพรากของตัวแปรตัวนั้น

2.2 การสุ่มป่า (Random Forest)

เป็นการนำสร้างต้นไม้การตัดสินใจหลายๆต้น โดยในแต่ละต้นนั้นจะได้รับข้อมูล ตัวแปรที่ไม่เหมือนกัน ซึ่งจะทำให้การเลือกข้อมูลเพียงบางส่วนจากข้อมูลทั้งหมด ซึ่งจะทำให้ได้ ต้นไม้การตัดสินใจหลายๆแบบ และจะทำการดูผลลัพธ์ที่มีจำนวนมากที่สุดหรือดูค่าเฉลี่ยของผลลัพธ์ที่ได้จากต้นไม้การตัดสินใจทั้งหมด มาเป็นผลลัพธ์ของระบบ



3.รายละเอียดการพัฒนา

3.1 การออกแบบระบบ

การออกแบบของระบบจะเริ่มจากการทำการแปลงข้อมูลที่มีให้อยู่ในรูปแบบที่เหมาะสมกับการวิเคราะห์ โดยในบทความนี้จะเลือกใช้ข้อมูลของดัชนีตลาดหลักทรัพย์ หลังจากนั้นก็จะทำการสอนโมเดลของการสุ่มป่า (Random Forest) และทำการกำหนด วิธีการซื้อ-ขายหลักทรัพย์ จากผลลัพธ์ที่ได้

3.2 รายละเอียดของระบบที่พัฒนา

3.2.1 การกำหนดปัญหา

ทำการกำหนดเป้าหมายของโมเดล โดยจะเป็นการทำนายสถานะของราคาในวันปัจจุบันว่าเป็นราคาที่เป็นจุดสูงสุด หรือจุดต่ำสุด หรือไม่

3.2.2 การเตรียมข้อมูล

ทำการเตรียมข้อมูลราคารายวัน และทำการคำนวณหาผลต่างร้อยละของราคาดัชนีย้อนหลัง 1-60 วัน โดยใช้สมการ

$$R_t = \frac{P_t - P_{t-n}}{P_{t-n}}$$

โดย

- R_t ผลต่างร้อยละของราคาในช่วง n วัน
- P_t ราคาดัชนีที่เวลา t
- n ช่วงเวลา

จากนั้นทำการคำนวณหาส่วนเบี่ยงเบนมาตรฐาน หรือความผันผวน ย้อนหลังในช่วงตั้งแต่ 5-60 วัน

$$Vol_t = \sqrt{\frac{\sum_{i=t-n}^t (P_i - \frac{1}{n} \sum_{j=t-n}^t P_j)^2}{n}}$$

โดย

- Vol_t ความผันผวน ณ เวลา t
- P_t ราคาดัชนีที่เวลา t
- n ช่วงเวลา

จากนั้นนำข้อมูลทั้ง 2 อย่างมาต่อกันเพื่อใช้เป็นอินพุตของระบบ

3.2.3 การกำหนดเป้าหมายของโมเดล

ทำการกำหนดกลุ่มของอินพุตที่ได้รับ ซึ่งแบ่งได้เป็น 3 กลุ่มได้แก่

1. จุดสูงสุด (Local Maximum) เป็น 1
2. จุดต่ำสุด (Local Minimum) เป็น -1
3. จุดที่กำลังอยู่ช่วงแนวโน้ม (Trend) เป็น 0

ซึ่งผลลัพธ์ที่ได้จากโมเดลจะอยู่ในช่วง $[-1, 1]$

3.3 ขั้นตอนการพัฒนา

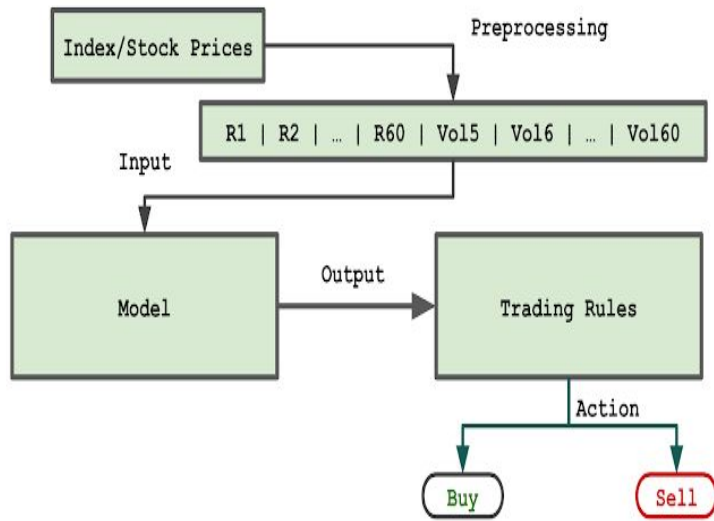
- ทำการศึกษาระบบการซื้อ-ขายหลักทรัพย์รูปแบบต่างๆ
- ทำการศึกษา Algorithm ที่ใช้
- ทำการกำหนดตัวแปร หรือ parameter ต่างๆ ที่เหมาะสมกับโมเดล และ ข้อมูล
- ทำการกำหนดและออกแบบ ระบบการตัดสินใจซื้อขาย

3.4 ขั้นตอนทำการซื้อ-ขาย

ทำการกำหนดคำสั่ง ซื้อ-ขาย เมื่อผลลัพธ์จากโมเดล มีค่าถึงเกณฑ์ที่กำหนด ดังนี้

- ทำการซื้อขายเมื่อ ผลการทำนายมีค่าน้อยกว่า -0.5
- ทำการขายเมื่อ ผลการทำนายมีค่ามากกว่า 0.5

หลังจากนั้น ทำการวัดผลด้วยการคำนวณผลกำไรสะสมที่ได้ จากการซื้อขายทั้ง โดยไม่คิดค่าธรรมเนียม



4.ผลการพัฒนาโครงการ

4.1 การเลือกข้อมูลและตัวแปรต่างๆที่ใช้สำหรับการทดลอง

ทำการเลือกข้อมูลย้อนหลังมาใช้ในการสอนและการทดสอบ โดยกำหนดให้ข้อมูลที่นำมาทดสอบต้องมีความยาวตั้งแต่ 4000-5000 วัน และต้องไม่ยาว และเก่าเกินไป และกำหนดข้อมูลสำหรับการทดสอบ 800-1000 วัน ดังนี้

4.1.1 ดัชนีหลักทรัพย์ CAC40 (FCHI)

ทำการ train ข้อมูลทั้งหมด 5000 วันย้อนหลังแล้วทำการทดสอบ 1000 วันถัดมา

4.1.2 ดัชนีหลักทรัพย์ Nikkei

ทำการ train ข้อมูลทั้งหมด 5000 วันย้อนหลังแล้วทำการทดสอบ 800 วันถัดมา

4.1.3 ดัชนีหลักทรัพย์ TWII

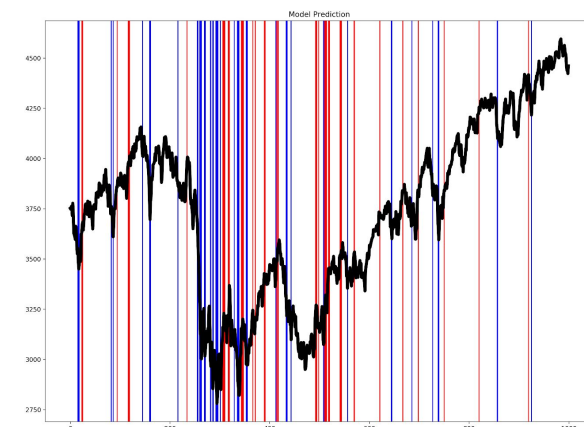
ทำการ train ข้อมูลทั้งหมด 5000 วันย้อนหลังแล้วทำการทดสอบ 1000 วันถัดมา

4.2 ผลลัพธ์จากการทำนายจุดวิกฤติ

ทำการนำโมเดลที่ได้จากการ train มาทำนายข้อมูลที่ใช้ในการทดสอบ โดยแต่ละภาพจะประกอบด้วย

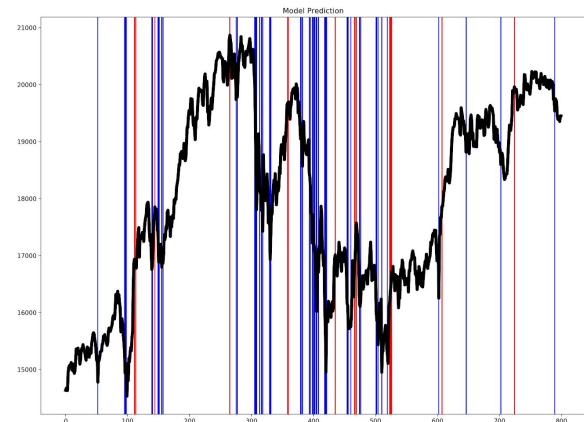
- เส้นสีแดง แสดงจุดสูงสุดของราคา จากการทำนายของโมเดล
- เส้นสีน้ำเงิน แสดงจุดต่ำสุดของราคา จากการทำนายของโมเดล
- เส้นสีดำ แสดงราคาของดัชนี

4.2.1 ดัชนีหลักทรัพย์ CAC40 (FCHI)



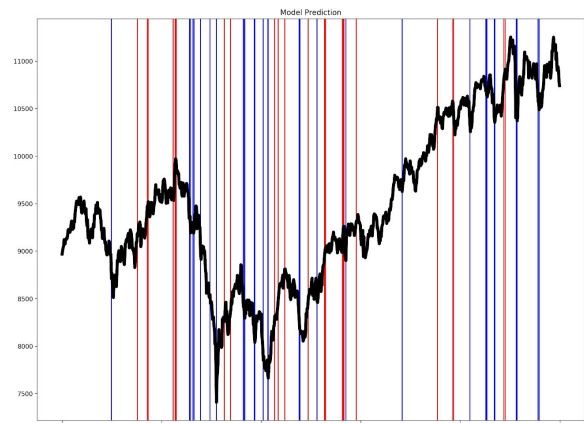
การทำนายจุดสูงสุดต่ำสุดและต่ำสุดของดัชนี CAC40 (FCHI)

4.2.2 ดัชนีหลักทรัพย์ Nikkei



การทำนายจุดสูงสุดต่ำสุดและต่ำสุดของดัชนี Nikkei

4.2.3 ดัชนีหลักทรัพย์ TWII



การทำนายจุดสูงสุดต่ำสุดและต่ำสุดของดัชนี TWII

4.3 ผลการจำลองการซื้อขายหลักทรัพย์

ทำการวัดผลด้วยการคำนวณผลกำไรสะสม บนข้อมูลที่ใช้ในการทดสอบและเปรียบเทียบผลกำไรสะสมกับตลาด โดยในแต่ละภาพจะประกอบด้วย

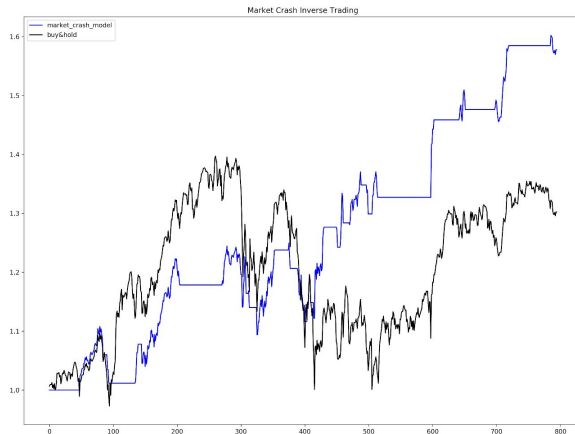
- เส้นสีน้ำเงิน - แสดงผลกำไรสะสมที่ทำได้จากการจำลองการซื้อขายของโมเดล
- เส้นสีดำ - แสดงผลกำไรสะสมที่ทำได้จากตลาด

4.3.1 ดัชนีหลักทรัพย์ CAC40 (FCHI)



ผลตอบแทนสะสมจากการจำลองการซื้อขายบนดัชนี CAC40 (FCHI)

4.3.2 ดัชนีหลักทรัพย์ Nikkei



ผลตอบแทนสะสมจากการจำลองการซื้อขายบนดัชนี Nikkei

4.3.3 ดัชนีหลักทรัพย์ TWII



ผลตอบแทนสะสมจากการจำลองการซื้อขายบนดัชนี TWII

ซึ่งเมื่อนำผลลัพธ์ที่ได้มาเปรียบเทียบกับผลลัพธ์ที่ได้จากตลาด จึงได้ข้อสรุปดังตารางดังนี้

Index	Length (Days)	Strategic Action Count	Cumulative Return	
			Model	Buy&Hold
FCHI	1000	52	1.45	1.21
Nikkei	800	56	1.58	1.31
TWII	1000	30	1.36	1.17

ตารางสรุปผลกำไรสะสม จากการจำลองการซื้อขาย

5.บทสรุป

ระบบการทำนายจุดวิกฤต สามารถทำนายจุดสูงสุดต่ำสุดของราคาดัชนี ราคาหุ้นได้ และสามารถนำประยุกต์ใช้กับการซื้อ-ขายหลักทรัพย์ได้จริง โดยจากผลการทดสอบการซื้อขายของทั้ง 3 ดัชนี สามารถกำไรได้สูงกว่าประมาณ 20 % และยังคงความเสี่ยงต่ำกว่าตลาด แต่ในข้อมูลบางจุดที่ทำให้โมเดลทำนายผิดพลาด ในข้อมูลที่เป็นแนวโน้มเป็นระยะยาว และมีความผันผวนต่ำจะทำให้โมเดลตัดสินใจผิดพลาดได้

6.เอกสารอ้างอิง

- [1] Decision Tree, Available at : <https://madlab.cpe.ku.ac.th/TR2/?itemID=238018>
- [2] Decision Tree, (2561), Available at : <https://th.wikipedia.org/wiki/ต้นไม้ตัดสินใจ>
- [3] Random Forest, (2561), Available at : <https://medium.com/@witchapongdaroontham/เจาะลึก-random-forest-part-2-of-รู้จัก-decision-tree-random-forest-และ-xgboost-79b9f41a1c1c>
- [4] Volatility Definition. (2561). Available at : <https://www.investopedia.com/terms/v/volatility.asp>
- [5] Return. (2561). Available at : <https://www.investopedia.com/terms/r/return.asp>