

Stock Price Direction Prediction using Machine Learning

การทำนายทิศทางของราคาหุ้นโดยใช้ Machine Learning

(Link: https://github.com/y-lims/DADS6003_Machine_Learning_Final_Project.git)

1. Introduction

1.1 Objective:

เพื่อศึกษาการทำนายทิศทางของราคาหุ้นในตลาดหลักทรัพย์แห่งประเทศไทย (SET) จำนวน 3 ตัว ได้แก่ TISCO, SPRC และ ERW โดยใช้ Machine Learning Model

1.2 Scope of Project:

- ข้อมูลราคาหุ้น 3 ตัว (TISCO, SPRC, ERW) กำหนด Window Time ทุกๆ 5 นาที เพื่อให้ข้อมูลมีการเรียนรู้ที่มากขึ้น และมีความแม่นยำมากขึ้นสำหรับการเทรน โมเดล
- กำหนดเกณฑ์ในการพิจารณาทิศทางของราคาหุ้น (Binary Classification) ดังนี้
 - Predicted Closing Price in the next 15 min > Current Closing Price : Sell (0)
 - Predicted Closing Price in the next 15 min = < Current Closing Price : Buy (1)
- ข้อมูลราคาหุ้นจาก website: <https://finance.yahoo.com>
Note: ทำการปรับเวลา Time Zone เป็นเวลาของประเทศไทย ("pendulum" library)

2. Background

2.1 ลักษณะกิจการของหุ้นที่ใช้ในการศึกษา:

- TISCO: บริษัท ทีเอสไอไฟแนนเชียลกรุ๊ป จำกัด (มหาชน)
Industry: ธุรกิจการเงิน Business group: ธุรกิจการเงิน
ลักษณะกิจการ: บริษัทประกอบธุรกิจการถือหุ้นในบริษัทอื่น (Holding Company) ของกลุ่มธุรกิจทางการเงินทีเอสไอ โดยมีธนาคารทีเอสไอเป็นหลัก ปัจจุบันให้บริการทางการเงินอื่นประกอบด้วย บริการสินเชื่อลูกค้ารายย่อยและสินเชื่อธุรกิจขนาดกลางและขนาดย่อม บริการสินเชื่อพาณิชย์ธนกิจ บริการเงินฝากรายย่อย บริการลูกค้าธนกิจ บริการตัวแทนขายประกันผ่านธนาคาร บริการจัดการการเงิน และบริการคัสโตเดียน
- SPRC: บริษัท สตาร์ ปิโตรเลียม รีไฟน์นิ่ง จำกัด (มหาชน)
Industry: ทรัพยากร Business group: พลังงานและสาธารณูปโภค
ลักษณะกิจการ: โรงกลั่นน้ำมันของบริษัทผลิตผลิตภัณฑ์ปิโตรเลียม ได้แก่ ก๊าซปิโตรเลียมเหลว น้ำมันเบนซิน ไร้สารตะกั่วเกรดพิเศษและเกรดธรรมดา น้ำมันดีเซลหมุนเร็ว น้ำมันเชื้อเพลิงอากาศยาน และน้ำมันเตา รวมทั้งผลิตภัณฑ์ปิโตรเคมีซึ่งใช้เป็นผลิตภัณฑ์ตั้งต้นในอุตสาหกรรมปิโตรเคมี
- ERW: บริษัท ดี เอราวัณ กรุ๊ป จำกัด (มหาชน)
Industry: บริการ Business group: การท่องเที่ยวและสันทนาการ
ลักษณะกิจการ: บริษัทและบริษัทย่อยประกอบธุรกิจเกี่ยวกับการลงทุนพัฒนาและดำเนินธุรกิจโรงแรมที่สอดคล้องกับทำเล สถานที่ตั้ง และกลุ่มเป้าหมาย เป็นธุรกิจหลัก โดยมีธุรกิจอื่น ได้แก่ ธุรกิจพื้นที่ให้เช่า และธุรกิจบริหารอาคาร

2.2 การวิเคราะห์ทางเทคนิค (Technical Analysis) : Technical Indicator

Indicators เป็นเครื่องมือที่ช่วยในการวิเคราะห์สภาพตลาด แนวโน้ม และทิศทางของราคาหุ้น สำหรับการศึกษา ใช้การการคำนวณและสูตรต่างๆ จาก

Library TA (TA-Lib) ร่วมกับการพิจารณาพารามิเตอร์ต่อไปนี้:

- ระดับราคา: จุดเปิด จุดปิด จุดสูงสุดและจุดต่ำสุด

- ปริมาณการซื้อขาย

Technical Indicators:

- **MOVING AVERAGE (MA) หรือเส้นค่าเฉลี่ยเคลื่อนที่:** เป็นการคำนวณค่าเฉลี่ย (Average) ของราคาหุ้น โดยใช้ข้อมูลของราคาหุ้นย้อนหลังตามระยะเวลาที่เรากำหนด (นิยมใช้ราคาปิดของหุ้นของช่วงระยะเวลาที่เราสนใจ)
 - **Exponential Moving Average (EMA) หรือ เส้นค่าเฉลี่ยราคาแบบถ่วงน้ำหนัก:**
คำนวณมาจากค่าเฉลี่ยของราคาปิดของหุ้นย้อนหลัง ตามจำนวนวันที่เราต้องการ โดยให้ราคาล่าสุด ๆ มีน้ำหนักมากกว่า
- **MOVING AVERAGE CONVERGENCE/DIVERGENCE หรือ MACD:** MOVING AVERAGE ประเภทนี้เป็น indicator ประเภท OSCILLATOR สำหรับกำหนดทิศทางที่เป็นไปได้ของ trend ราคา โดยขึ้นอยู่กับ CONVERGENCE/DIVERGENCE ของ MOVING AVERAGE แบบเอ็กซ์โปเนนเชียล
- **RELATIVE STRENGTH INDEX หรือ RSI:** แสดงการเพิ่มขึ้นโดยเฉลี่ยของอัตราส่วนอัตราส่วนลักษณะเทียบกับการลดลงเฉลี่ย นี่คือ indicator ชี้้นำ ด้วยความช่วยเหลือของมัน คุณสามารถระบุจุดเข้าสู่ตลาดได้ สัญญาณ: Overbought และ Oversold, Divergence รูปแบบกราฟ
- **ON BALANCE VOLUME หรือ OBV:** เครื่องมือที่วัดปริมาณการซื้อขายสะสม เพื่อวัดแรงซื้อแรงขาย ใช้ในการยืนยันทิศทางแนวโน้ม และหาจุดกลับตัวของราคา
- **BOLLINGER BAND INDITOR:** สามารถกำหนดตำแหน่งปัจจุบันของราคาเทียบกับช่วงการเทรดเฉลี่ย ใช้เพื่อระบุสัญญาณการกลับตัวของตลาดและความต่อเนื่องของ trend ปัจจุบัน

2.3 TA-Lib (Technical Analysis Library)

TA-Lib หรือ Technical Analysis Library เป็นเครื่องมือที่นิยมใช้ในการวิเคราะห์ทางเทคนิคในตลาดการเงิน มีฟังก์ชันหลากหลายที่มีประโยชน์สำหรับการวิเคราะห์หุ้นและเครื่องมือทางการเงิน รวมถึงตัวชี้วัดต่างๆ และมีวิธีการคำนวณที่ช่วยในการระบุแนวโน้ม, รูปแบบ, และสัญญาณซื้อ/ขายที่อาจเกิดขึ้นในข้อมูลตลาด (Financial market data) และ TA-Lib ถูกใช้โดยนักลงทุนและนักวิเคราะห์ทางการเงินเนื่องจากมีประสิทธิภาพและเชื่อถือได้ในงานวิเคราะห์ทางเทคนิค

2.4 Machine Learning Algorithm

การจำแนกประเภทด้วย Machine Learning (ML) เป็นการวิเคราะห์และทำนายข้อมูลต่างๆ แบบ Supervised Learning โดยมีเทคนิคการจำแนกประเภท (Classification Techniques) ซึ่งใช้สำหรับแบ่งข้อมูลออกเป็นหมวดหมู่ต่างๆ โดยประยุกต์ใช้ในการวินิจฉัยโรค การระบุประเภทของวัตถุ และการทำนายพฤติกรรมของผู้บริโภค ตัวอย่างของอัลกอริทึมจำแนกประเภท ได้แก่

- **Logistic Regression:** เป็นเทคนิคการถดถอยที่ใช้สำหรับการจำแนกประเภทข้อมูล เหมาะกับข้อมูลที่มีคำตอบเป็นประเภทหมวดหมู่ เช่น ใช่/ไม่ใช่ หรือ เป็น/ไม่เป็น
- **Gaussian Naive Bayes:** เป็นเทคนิคการจำแนกประเภทข้อมูลที่ใช้หลักการความน่าจะเป็น โดยสมมติข้อมูลมีการแจกแจงแบบปกติ และเลือกคลาสที่มีความน่าจะเป็นสูงสุดจากข้อมูลที่มี
- **XGBoost (Extreme Gradient Boosting):** เป็นเทคนิคการเรียนรู้ของเครื่องที่มุ่งเน้นที่ประสิทธิภาพและความเร็วในการคำนวณ มันปรับปรุงความแม่นยำโดยใช้วิธีการ 'Boosting' ซึ่งเป็นการเรียนรู้จากข้อผิดพลาดของโมเดลในขั้นตอนก่อนหน้า ทำให้โมเดลสามารถแก้ไขและปรับปรุงตัวเองได้ในแต่ละรอบของการฝึกฝน
- **SVM (Support Vector Machine):** เป็นเทคนิคที่ใช้สำหรับการจำแนกประเภทและการถดถอย ทำงานโดยการหาเส้นแบ่งที่ดีที่สุดระหว่างกลุ่มข้อมูลที่แตกต่างกัน
- **Random Forest:** เป็นเทคนิคที่รวมต้นไม้การตัดสินใจหลายต้นเข้าด้วยกันเพื่อเพิ่มความแม่นยำในการทำนาย มีประสิทธิภาพสูงในการจัดการข้อมูลที่มีความซับซ้อนและปริมาณมาก
- **Decision Tree:** เป็นแบบจำลองที่ใช้โครงสร้างเหมือนต้นไม้ในการทำนายผลลัพธ์ โดยแบ่งข้อมูลออกเป็นกลุ่มย่อยๆ ตามคุณสมบัติต่างๆ ของข้อมูล

2.5 ขั้นตอนการศึกษา: การสร้างโมเดลสำหรับทำนายทิศทางราคาหุ้น

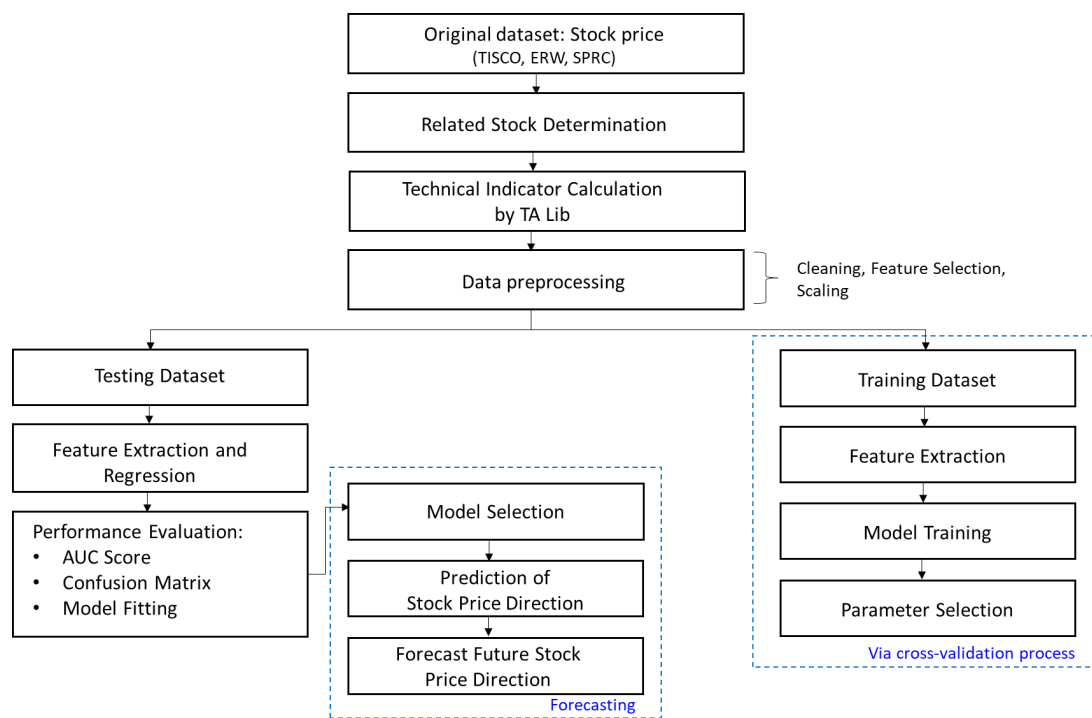


Figure 1. แผนผังแสดงขั้นตอนการสร้างโมเดลสำหรับทำนายทิศทางราคาหุ้น

3. วิธีดำเนินการศึกษา: การสร้างโมเดลสำหรับทำนายทิศทางราคาหุ้น

ขั้นตอนที่ 1: Data Understanding & Collection

1.1) ศึกษาและเก็บรวบรวมข้อมูล:

- ข้อมูลราคาหุ้นย้อนหลัง (Open, High, Low, Close, Adj Close, Volume)
- ข้อมูลราคาหุ้นและดัชนีที่เกี่ยวข้อง เช่น
 - SET Index: SET100, SET50
 - Industrial Index:
 - TISCO: Financial Index
 - ERW: Service Index
 - SPRC: Resource Index
 - Business Index:
 - TISCO: Banking Index
 - ERW: Tourism Index

■ SPRC: Energy Index

- ราคาทองคำ (Gold), ราคาเงินสกุล USD, ราคาเงินสกุล CNY, ราคาน้ำมันดิบ BRENT และ WTI
- ราคาหุ้นของบริษัทที่ดำเนินธุรกิจประเภทเดียวกัน ex. SPRC: IRPC, TOP, BCP
- Technical Indicators ต่างๆที่นิยม เช่น EMA, RSI, OBV และ Bollinger Band

1.2) ศึกษาการวิเคราะห์หุ้นด้วย Technical Indicators และเลือก Indicator Function ที่น่าสนใจ

1.3) นำเข้าโมดูลที่ใช้ในการเก็บข้อมูล คือ pandas และเก็บข้อมูลใน DataFrame

1.4) วิเคราะห์ลักษณะข้อมูลโดยการ plot กราฟ Histogram และดูค่า Outlier จากกราฟ Box plot

Note: ข้อมูลหุ้นจำนวน 2 เดือน โดยใช้ Time frame ของข้อมูลอยู่ที่ 5 นาที

ขั้นตอนที่ 2: Preprocessing

2.1) Feature Engineering

- สร้าง Feature ใหม่ โดยการสร้างคอลัมน์สำหรับ Feature ราคาหุ้นที่ 5 นาทีถัดไป เพื่อให้ง่ายต่อการนำไปวิเคราะห์ทิศทางของราคาหุ้น
- สร้าง Feature ของราคาหุ้นและดัชนีที่เกี่ยวข้องต่างๆ
- สร้าง Feature สำหรับ Technical Indicators โดยใช้ TA Lib
- สร้าง Label data (y) สำหรับทำนายทิศทางราคาหุ้น โดย
 - Predicted Closing Price in the next 5 min > Current Closing Price : Sell (0)
 - Predicted Closing Price in the next 5 min < Current Closing Price : Buy (1)
- ทำการ Clean ข้อมูลโดยการใส่ค่า Median และค่า Fill Forward (ffill) สำหรับค่า Null ในข้อมูล

2.2) Feature Selection

- วิเคราะห์หาความสัมพันธ์ระหว่างข้อมูล (Correlation) เพื่อหาว่า Feature ไດควรนำมาใช้ในการวิเคราะห์ โดยเลือกจาก Feature ที่มีค่าความสัมพันธ์ระหว่างข้อมูล มากกว่า 0.3

2.3) Normalization Data

- วิเคราะห์การกระจายตัวของข้อมูลโดยใช้ anderson test
- ทำ Feature Scaling โดยการใช้ค่า Min และ Max ค่าที่ถูก Scale จะอยู่ระหว่าง [0,1] เนื่องจากข้อมูลมีการกระจายตัวแบบไม่ปกติ

ขั้นตอนที่ 3: Model Development

3.1) Split Data

3.2) Model Training

- Grid Search และ k-fold Cross Validation ในการเทรนโมเดลจำนวน 6 โมเดล ได้แก่ Logistic Regression ,Extreme Gradient Boosting (XGBoost), Support Vector Machine (SVM), Random forest ,Decision tree ,Gaussian Naive Bayes

ขั้นตอนที่ 4: Performance Evaluation

- AUC Score และ Matrix Confusion
- Balancing Evaluation (Overfitting, Balance and Underfitting)

ขั้นตอนที่ 5: Prediction

5.1) Model Selection

5.2) ทดสอบกับค่าจริง

4. ผลการศึกษา

ผลการ Run Machine Learning ของหุ้นที่กำหนด ในแต่ละ Model พบว่าทั้ง 3 หุ้น มีประสิทธิภาพของโมเดล (AUC Score) ที่มากที่สุดคือ Logistic

Regression มีค่าเฉลี่ยอยู่ระหว่าง 60%-70%

ดังนั้นจึงนำมาเป็น Model ที่ใช้ทำนายหุ้นที่กำหนด

- TISCO:

- **Feature Selection:** tisco_close, ficial_close, bank_close, ema, obv, bb_upper, bb_lower, gold_close, rsi (เลือก Model จาก strong_corr_features และ moderate_corr_features)
- จากผลการทำนายของแต่ละโมเดล โมเดลที่มีค่า ROC สูงที่สุดคือ โมเดล **Logistic regression** เมื่อทำการประเมิน Model Fitting พบว่า โมเดล balanced
- พิจารณาเลือก โมเดล Logistic regression เนื่องจาก มีค่า ROC สูงที่สุด และ มีความ balanced ระหว่าง train set และ test set

- SPRC:

- **Feature Selection:** sprc_close, set50_close, irpc_close, top_close, ficial_close, bank_close, ema, obv, bb_upper, bb_lower, resource_close, energy_close (เลือก Model จาก strong_corr_features และ moderate_corr_features)
- จากผลการทำนายของแต่ละโมเดล โมเดลที่มีค่า ROC สูงที่สุดคือ โมเดล **Logistic regression** เมื่อทำการประเมิน Model Fitting พบว่า โมเดล balanced
- พิจารณาเลือก โมเดล Logistic regression เนื่องจาก มีค่า ROC สูงที่สุด และ มีความ balanced ระหว่าง train set และ test set

- ERW:

- **Feature Selection:** erw_close, gold_close, resource_close, ema, obv, bb_upper, bb_medium, bb_lower, cny_close, rsi (เลือก Model จาก strong_corr_features และ moderate_corr_features)
- จากผลการทำนายของแต่ละโมเดล โมเดลที่มีค่า ROC สูงที่สุดคือ โมเดล **Logistic regression** เมื่อทำการประเมิน Model Fitting พบว่า โมเดล balanced
- พิจารณาเลือก โมเดล Logistic regression เนื่องจาก มีค่า ROC สูงที่สุด และ มีความ balanced ระหว่าง train set และ test set

Conclusion and Further Study

การใช้ Machine Learning ในการวิเคราะห์ทิศทางของราคาหุ้น

ตลาดหุ้นเป็นตลาดที่มีสภาพคล่องสูงและมีการเปลี่ยนแปลงอยู่ตลอดเวลา นักลงทุนจึงจำเป็นต้องมีเครื่องมือในการช่วยวิเคราะห์ทิศทางของราคาหุ้น เพื่อประกอบการตัดสินใจลงทุน หนึ่งในเครื่องมือที่ได้รับความนิยมในปัจจุบันคือการใช้ Machine Learning

จากการศึกษาเกี่ยวกับการใช้ Machine Learning ในการวิเคราะห์ทิศทางของราคาหุ้น พบว่าโมเดลทางสถิติหรือการใช้ Machine Learning สามารถช่วยในการทำนายแนวโน้มของราคาหุ้นได้บ้าง โดยมีความแม่นยำอยู่ระดับหนึ่ง อย่างไรก็ตาม ปัจจัยที่ส่งผลกระทบต่อทิศทางราคาดังกล่าวนั้นมีหลากหลาย โมเดล Machine Learning จึงจำเป็นต้องอาศัยปัจจัยที่เหมาะสมในการช่วยทำนายให้แม่นยำยิ่งขึ้น

ปัจจัยที่ส่งผลกระทบต่อทิศทางราคาหุ้น

ปัจจัยที่ส่งผลกระทบต่อทิศทางราคาดังกล่าวนั้นมีหลายประการและมีผลต่อการขยับขึ้นหรือลงของราคาหุ้นในตลาดหุ้น ซึ่งอาจมีการเปลี่ยนแปลงตลอดเวลาขึ้นอยู่กับสภาพตลาด เช่น กำไร, รายได้, กระแสเงินสด, อัตราผลตอบแทนทางการเงิน (Return on Equity (ROE), Return on Assets (ROA), และ Debt-to-Equity

Ratio), ข้อมูลราคาย้อนหลัง, ปริมาณการซื้อขาย, ตัวชี้วัดเศรษฐกิจทั่วไปเช่น GDP, อัตราเงินเฟ้อ, ข้อมูลข่าวสารที่เกี่ยวข้องกับบริษัทหรือตลาดโลก, การใช้เทคโนโลยีทางการวิเคราะห์ข้อมูล และอื่นๆ

ความสำคัญของการตรวจสอบและปรับโมเดล

สภาพตลาดหุ้นเปลี่ยนแปลงอยู่ตลอดเวลา โมเดล Machine Learning จึงจำเป็นต้องได้รับการตรวจสอบและปรับให้เหมาะสมอยู่เสมอ เพื่อให้สามารถคาดการณ์ทิศทางราคาหุ้นได้อย่างแม่นยำยิ่งขึ้น

แนวทางการวิจัยเพิ่มเติมในการใช้ Machine Learning ในการวิเคราะห์ทิศทางของราคาหุ้น

การเพิ่มข้อมูลให้เพียงพอ เพื่อให้โมเดลสามารถเรียนรู้และคาดการณ์ได้อย่างแม่นยำยิ่งขึ้น

การเปรียบเทียบความแม่นยำของโมเดล Machine Learning กับโมเดลอื่นๆ เช่น โมเดลทางสถิติ

การพิจารณาการใช้ข้อมูลทางเทคโนโลยีสารสนเทศ (information technology) เช่น การใช้ Natural Language Processing (NLP) ในการวิเคราะห์ข้อมูลข่าวสารที่มีผลต่อตลาดหุ้น หรือการใช้ Reinforcement Learning เพื่อปรับโมเดลให้เหมาะสมกับการเปลี่ยนแปลงของตลาด

บทสรุป

การใช้ Machine Learning ในการวิเคราะห์ทิศทางของราคาหุ้นเป็นแนวทางหนึ่งที่น่าสนใจ แต่นักลงทุนควรตระหนักถึงข้อจำกัดของโมเดล Machine Learning และควรใช้อย่างระมัดระวัง

References:

- วิเคราะห์กราฟหุ้น อย่างไร ฉบับนักลงทุนสายเทคนิคเบื้องต้น:
<https://knowledge.bualuang.co.th/knowledge-base/howtouseindicators/>
- การวิเคราะห์ทางเทคนิคตลาดสำหรับผู้เริ่มต้น: พื้นฐาน วิธีการ หลักการ:
<https://www.lifefinance.org/th/blog/for-beginners/kar-wikheraah-thang-thekhnikh/>
- จับจังหวะซื้อขายหุ้นด้วย Indicators แบบไหน ยังไงดี? :
<https://www.setinvestnow.com/th/knowledge/video/159-settrade-streaming15>
- Technical Analysis Library in Python's documentation:
<https://technical-analysis-library-in-python.readthedocs.io/en/latest/>

Team Members:

- | | | |
|---------------------|--------------|-------------------------|
| 1. นางสาวณัฏฐิตา | ตั้งศรีวงศ์ | รหัสนักศึกษา 6520422014 |
| 2. นางสาวพิชญานา | ชูวัฒนวิทย์ | รหัสนักศึกษา 6520422024 |
| 3. นางสาวธัญญลักษณ์ | ลิ้มสุขวัฒน์ | รหัสนักศึกษา 6520422030 |