

ระบบคาดการณ์ราคาหุ้นในตลาดหลักทรัพย์

Stock price prediction system

สุริยา เตชะลือ¹ กวิสรา ตักดีบุรณาเพชร²

วิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเชียงใหม่

Email: ¹suriya_techalue@cmu.ac.th, ²kavisara_sakbu@cmu.ac.th

บทคัดย่อ

ในปัจจุบันได้มีการศึกษาและพัฒนาตัวแบบพยากรณ์ราคาหุ้นโดยใช้ความรู้ด้านต่าง ๆ มาสร้างตัวแบบมากมาย แต่การที่จะพยากรณ์ได้อย่างแม่นยำนั้นตัวแบบต้องสามารถรองรับความหลากหลายของปัจจัยที่ทำให้ส่งผลต่อราคาหุ้นได้ ซึ่งในโครงการนี้ได้เสนอ การใช้ซัพพอร์ตเวกเตอร์เรกเรชัน เพื่อพยากรณ์ราคาหุ้นให้มีความแม่นยำมากขึ้น โดยได้เลือกใช้ข้อมูลประกอบด้วย ข้อมูลราคาหุ้นของแต่ละบริษัท และข้อมูลราคาน้ำมันดิบ ในการพยากรณ์ราคาหุ้นของบริษัทในกลุ่มพลังงานและสาธารณูปโภค จากนั้นทำการสร้างตัวแบบโดยใช้ซัพพอร์ตเวกเตอร์เรกเรชันโดยค้นหาค่าพารามิเตอร์ที่ให้ผลลัพธ์ค่ารากที่สองของความคลาดเคลื่อนเฉลี่ยเข้าใกล้ 0 มากที่สุด และค่าสัมประสิทธิ์แสดงการตัดสินใจเข้าใกล้ 1 มากที่สุด โดยตัวแบบที่สร้างในโครงการนี้จะทำนายราคาหุ้นของวันถัดไปที่ตลาดหลักทรัพย์เปิดทำการ ซึ่งผลการทดลองการสร้างตัวแบบที่ดีที่สุด มีค่าสัมประสิทธิ์แสดงการตัดสินใจอยู่ที่ 0.687740 – 0.971297 โดยผู้พัฒนาหวังเป็นอย่างยิ่งว่าโครงการชิ้นนี้จะสามารถนำไปใช้ประโยชน์ และให้ความรู้ความเข้าใจในการทำงานของตัวแบบที่ใช้ในการทำนายราคาหุ้นในตลาดหลักทรัพย์

คำสำคัญ: การพยากรณ์ราคาหุ้น, ปัญหาประติษฐ์กับการลงทุน, ปัญหาประติษฐ์, ทำนายราคาหุ้น

Abstract

Nowadays, Forecasting stock prices with machine learning is becoming widespread. Therefore, we want to develop a model that works precisely and prove the ability to use machine learning to forecast stock prices. However, producing an efficient algorithm requires a model with input diversity.

Hence, in this research, we develop a prediction system using the support vector regression in the prediction system as well. In this research, Stock price and Crude Oil prices are used to extract feature to predict stock price in Energy and Utility group. After that, The model was constructed using the support vector regression, searching for the parameters that yielded the lowest Root Mean Square Error (RMSE) and highest Coefficient of Determination (R-Squared). The model built in this project predicts the share price of the next day the stock market opens. The best model in this experiment yielded R-Squared around 0.687740 to 0.971297. Finally, we hope that the result of study can help people who are interested in machine learning for stock price forecasting.

Keyword: Stock price forecast, Machine learning

1. บทนำ

ปัจจุบันการพยากรณ์ราคาหุ้นได้รับความสนใจอย่างมากในหมู่นักลงทุนและนักวิเคราะห์ตลาดหุ้น มีงานศึกษาวิจัยจำนวนมากได้ทำการศึกษาค้นคว้าและพัฒนาตัวแบบพยากรณ์ให้มีประสิทธิภาพ มีความเที่ยงตรง แม่นยำ สามารถรองรับความหลากหลายของปัจจัยต่าง ๆ ได้มากยิ่งขึ้น

จากงานศึกษาวิจัยเกี่ยวกับความสัมพันธ์ของราคาหุ้นในกลุ่มพลังงานและสาธารณูปโภคพบว่า ราคาน้ำมันมีความสัมพันธ์ในทิศทางเดียวกันกับดัชนีราคาหุ้นกลุ่มพลังงานและสาธารณูปโภค[1] ดังนั้นจึงมีผลต่อราคาของหุ้นแต่ละตัวในกลุ่มนี้ด้วย และอาจมองได้ว่าราคาของหุ้นเป็นข้อมูลทางสถิติที่สามารถนำมาวิเคราะห์โดยใช้หลักการทางคณิตศาสตร์และวิศวกรรมศาสตร์เข้ามาช่วยในการหาแนวโน้มและราคาได้

เนื่องจากราคาของหุ้นเป็นข้อมูลทางสถิติซึ่งอาจจะมีรูปแบบเฉพาะซ่อนอยู่ ดังนั้นจึงเลือกใช้ตัวแบบพยากรณ์ราคาหุ้นร่วมกับการใช้ซัพพอร์ตเวกเตอร์รีเกรสชันที่เป็นเครื่องมือฝึกสอนระบบอันมีพื้นฐานมาจากทฤษฎีการฝึกสอนทางสถิติช่วยในการประมาณค่าฟังก์ชันแล้วคาดว่าจะทำให้ผลลัพธ์จากใช้ตัวแบบพยากรณ์ราคาหุ้นมีความเที่ยงตรงและแม่นยำมากยิ่งขึ้น

ทั้งหมดนี้เพื่อลดความเสี่ยงของการลงทุนในตลาดหลักทรัพย์และประหยัดเวลาวิเคราะห์ข้อมูล ทำให้สามารถวางแผนล่วงหน้าเพื่อเตรียมรับมือกับการเข้าลงทุนในหุ้น หรือหยุดการลงทุนเพื่อลดการขาดทุน ทั้งนี้เพื่อให้นักลงทุนตระหนักในความเสียง และเห็นแนวโน้มและทิศทางของการลงทุนได้ง่ายขึ้น

2. หลักการทำงานของระบบ

2.1 ความสัมพันธ์ที่เกี่ยวข้องของหุ้นกลุ่มพลังงานและสาธารณูปโภค [1]

หมวดธุรกิจพลังงานและสาธารณูปโภค [2] หมายถึง ผู้ประกอบธุรกิจผู้ผลิต สํารวจ ขุดเจาะ ก่อกั้น และ ตัวแทนจำหน่ายพลังงานธรรมชาติ ในรูปต่าง ๆ เช่น น้ำมันและก๊าซธรรมชาติ รวมถึงผู้ให้บริการสาธารณูปโภคต่าง ๆ เช่น ไฟฟ้า ประปา และแก๊ส

ซึ่งสามารถวิเคราะห์ความสัมพันธ์ทางสถิติด้วยการวิเคราะห์สมการถดถอยเชิงซ้อน (Multiple Regression) ดังสมการที่ 2.1

$$SET_{in} = b_0 + b_1(INT) + b_2(FOR) + b_3(INF) + b_4(EXC) + b_5(OIL) + b_6(GOLD) + b_7(BSI) \quad (2.1)$$

โดยที่ b_0 คือค่าคงที่ ให้ $b_1, b_2, b_3, b_4, b_5, b_6, b_7$, คือค่าสัมประสิทธิ์การถดถอยของตัวแปรอิสระ ให้ SET_{in} คือดัชนีราคาหุ้นกลุ่มอุตสาหกรรมแต่ละกลุ่ม ให้ INT คืออัตราดอกเบี้ยการกู้ยืมเงินบาท 1 ปี ให้ FOR คือปริมาณซื้อหลักทรัพย์ของนักลงทุนต่างชาติ ให้ INF คืออัตราเงินเฟ้อทั่วไป ให้ EXC คืออัตราแลกเปลี่ยนเงินบาทต่อเงินดอลลาร์สหรัฐ ให้ OIL คือราคาน้ำมันดิบ (\$/barrel) ให้ GOLD คือราคาขายทองคำแท่งเฉลี่ยรายเดือน และให้ BSI คือดัชนีความเชื่อมั่นทางธุรกิจ

จากสมการ 2.1 สามารถวิเคราะห์หุ้นกลุ่มพลังงานและสาธารณูปโภคดังสมการ 2.2 ซึ่งแสดงให้เห็นว่าราคาน้ำมันดิบมีความสัมพันธ์กับราคาหุ้นกลุ่มพลังงานและสาธารณูปโภคเป็นไปในทิศทางเดียวกัน ซึ่งเป็นไปตามสมมติฐานที่กล่าวไว้

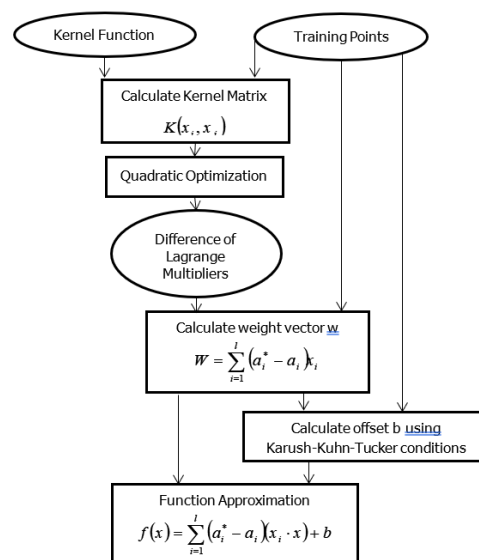
$$ENERG = 230.283 + 0.018FOR + 62.522OIL - 0.709GOLD (1.303)^{ns} (3.343)^{**} (-2.778)^{**} (-2.720)^{**} \quad (2.2)$$

สรุปได้ว่าโดยปัจจัยทางเศรษฐกิจที่มีผลต่อดัชนีราคาหุ้นกลุ่มพลังงานและสาธารณูปโภค ได้แก่ ปริมาณซื้อหลักทรัพย์ของนักลงทุนต่างชาติ ราคาน้ำมัน และราคาทองคำแท่ง โดยปริมาณซื้อหลักทรัพย์ของนักลงทุนต่างชาติ ราคาน้ำมันมีความสัมพันธ์ในทิศทางเดียวกันกับดัชนีราคาหุ้นกลุ่มพลังงานและสาธารณูปโภค ส่วนราคาทองคำแท่งมีความสัมพันธ์ในทิศทางตรงกันข้ามกับดัชนีราคาหุ้นกลุ่มพลังงานและสาธารณูปโภค

2.2 ซัพพอร์ตเวกเตอร์รีเกรสชัน (Support Vector Regression;SVR) [4]

ซัพพอร์ตเวกเตอร์รีเกรสชันมีหลักการคล้ายกับซัพพอร์ตเวกเตอร์แมชชีนแบบแบ่งกลุ่มคือใช้หาระนาบเกินที่เหมาะสมที่สุด (Optimal Hyperplane) แตกต่างกับที่ซัพพอร์ตเวกเตอร์แมชชีนแบบแบ่งกลุ่มจะสนใจเพียงค่าบวกและลบที่เกิดขึ้นจากการแบ่งกลุ่มข้อมูล แต่ซัพพอร์ตเวกเตอร์รีเกรสชันจะสนใจค่าจริงที่เกิดขึ้นจากการประมาณค่าฟังก์ชัน

ซัพพอร์ตเวกเตอร์รีเกรสชันมีอยู่ 2 ประเภท คือ แบบเชิงเส้น (Linear Regression) และแบบไม่เป็นเชิงเส้น (Nonlinear Regression) ซึ่งซัพพอร์ตเวกเตอร์แมชชีนแบบไม่เป็นเชิงเส้นจะมีขั้นตอนแตกต่างจากแบบเชิงเส้นคือจะมีการแมปข้อมูลให้อยู่ปริภูมิที่สูงกว่าเพื่อให้ได้ข้อมูลที่มีลักษณะเป็นเชิงเส้น ซึ่งขั้นตอนของซัพพอร์ตเวกเตอร์รีเกรสชัน แสดงได้ดังภาพที่ 2.1



ภาพที่ 2.1 ขั้นตอนของซัพพอร์ตเวกเตอร์รีเกรสชัน

ซึ่งการหาระนาบเงินที่เหมาะสมเป็นการหาซัพพอร์ตเวกเตอร์ที่สามารถรักษาระยะห่างมากที่สุดระหว่างข้อมูลทั้งสองกลุ่ม ซัพพอร์ตเวกเตอร์ที่ได้จะใช้เป็นฟังก์ชันประมาณค่าของกลุ่มข้อมูลทั้งหมด การหา Norm ที่น้อยสุดของ w จะทำให้ได้ค่า w ที่เหมาะสมที่สุดโดยใช้เงื่อนไขตามดังสมการที่ 2.3 และ 2.4

$$\left(\text{Minimize} \frac{1}{2} \|w\|^2 \right) \quad (2.3)$$

$$\left\{ \begin{array}{l} y_i \langle w \cdot x_i \rangle - b \leq \varepsilon \\ \langle w \cdot x_i \rangle + b - y_i \leq \varepsilon \end{array} \right\} \quad (24)$$

การสร้างระนาบเงินที่สามารถประมาณค่าได้อย่างแม่นยำนั้นสามารถกำหนดความแม่นยำได้จากการกำหนดความกว้างของระนาบที่เหมาะสมโดยพิจารณาจากค่าความคลาดเคลื่อนที่ยอมรับได้ (Error Insensitive) ในรูปฟังก์ชันการสูญเสีย (Loss Function) จากฟังก์ชันการสูญเสียแบบ $\varepsilon - \text{Insensitive}$ ดังสมการที่ 2.5

$$L(y_i, f(x)) = \begin{cases} 0 & ; |y - f(x)| \leq \varepsilon \\ |y - f(x)| - \varepsilon & ; |y - f(x)| > \varepsilon \end{cases} \quad (2.5)$$

ในฟังก์ชันการสูญเสียแบบ $\varepsilon - \text{Insensitive}$ มีการพิจารณาตัวแปรช่วย ξ (Slack) เป็นค่าความคลาดเคลื่อนของข้อมูลที่อยู่นอกกระนาบทั้งสอง ได้สมการใหม่ดังสมการที่ 2.6 และ 2.7

$$\text{Minimize} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l s_i (\xi_i + \xi_i^*) \quad (2.6)$$

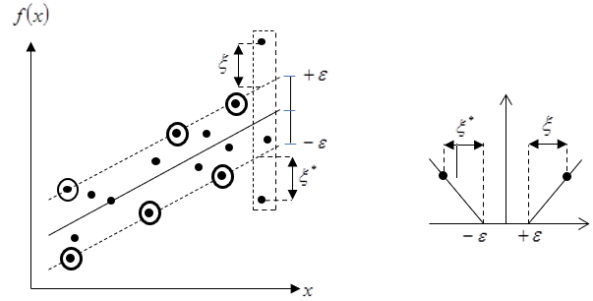
$$\left\{ \begin{array}{l} y_i - \langle w \cdot x_i \rangle - b \leq \varepsilon + \xi_i \\ \langle w \cdot x_i \rangle + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \\ 1 \leq i \leq l \end{array} \right\} \quad (2.7)$$

โดย C คือ ค่าคงที่สำหรับควบคุมค่าคลาดเคลื่อน (Regularization Parameter) ξ คือ ค่าคลาดเคลื่อนของข้อมูลจากขอบระนาบบน ξ^* คือ ค่าคลาดเคลื่อนของข้อมูลจากขอบระนาบล่าง จากสมการที่ 2.6 จะสามารถหาคำตอบได้ด้วยเงื่อนไขของสมการที่ 2.7 โดยใช้ฟังก์ชันลากรางจ์ (Lagrange Function) ได้สมการจากการเพิ่มตัวคูณลากรางจ์ (Lagrange Multipliers) ดังนี้

$$L = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l s_i (\xi_i + \xi_i^*) - \sum_{i=1}^l \alpha_i (\varepsilon + \xi_i - y_i + \langle w \cdot x_i \rangle + b)$$

$$- \sum_{i=1}^l \alpha_i^* (\varepsilon + \xi_i^* - y_i + \langle w \cdot x_i \rangle + b) - \sum_{i=1}^l (\eta_i \xi_i + \eta_i^* \xi_i^*) \quad (2.8)$$

โดย L คือ Lagrangian และ $\eta_i, \eta_i^*, \alpha_i, \alpha_i^*$ คือ ตัวคูณลากรางจ์ ซึ่ง $\eta_i, \eta_i^*, \alpha_i, \alpha_i^* \geq 0$



ภาพที่ 2.2 การหาระนาบเงินที่เหมาะสมที่สุด

จากสมการที่ 2.7 แก่สมการด้วยวิธีกำลังสอง (Quadratic Programming) โดยหาอนุพันธ์ย่อย (Partial Derivatives) เทียบกับตัวแปรที่ต้องการหาค่าโดยให้เท่ากับศูนย์ ได้คำตอบดังสมการที่ 2.9

$$\left\{ \begin{array}{l} \frac{\partial L}{\partial b} = \sum_{i=1}^l (\alpha_i^* - \alpha_i) = 0 \\ \frac{\partial L}{\partial w} = w - \sum_{i=1}^l (\alpha_i^* - \alpha_i) x_i = 0 \\ \frac{\partial L}{\partial \xi_i} = C - \alpha_i - \eta_i = 0 \\ \frac{\partial L}{\partial \xi_i^*} = C - \alpha_i^* - \eta_i^* = 0 \end{array} \right\} \quad (2.9)$$

จากสมการที่ 2.9 เมื่อนำไปแทนในฟังก์ชันลากรางจ์ได้สมการที่ 2.10

$$\text{Maximize} -\frac{1}{2} \sum_{i,j=1}^l (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) \langle x_i \cdot x_j \rangle - \varepsilon \sum_{i=1}^l (\alpha_i + \alpha_i^*) + \sum_{i=1}^l y_i (\alpha_i - \alpha_i^*) \quad (2.10)$$

ซึ่งการหาคำตอบของสมการที่ 2.10 ต้องทำภายใต้เงื่อนไข

$$\left\{ \begin{array}{l} \sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0 \\ (\alpha_i - \alpha_i^*) \in [0, s_i C] \end{array} \right.$$

$$\text{หรือ } 0 \leq \alpha_i, \alpha_i^* \leq s_i C$$

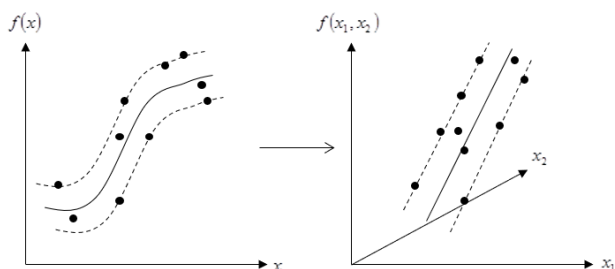
จากสมการที่ 2.9 หาก $w = \sum_{i=1}^l (\alpha_i^* - \alpha_i) x_i$ จะได้สมการระนาบเงินอันใหม่เป็น

$$f(x) = \sum_{i=1}^l (\alpha_i^* - \alpha_i) \langle x_i \cdot x \rangle + b \quad (2.11)$$

แต่หากข้อมูลที่นำมาสอนมีลักษณะไม่เป็นเชิงเส้น ต้องใช้ฟังก์ชันเคอร์เนล (Kernel Function) ส่งผ่านข้อมูลที่ไม่เป็นเชิงเส้นไปยังปริภูมิหรือมิติที่สูงขึ้นเพื่อให้ข้อมูลมีลักษณะเป็นเชิงเส้น แล้วก็จึงทำตามขั้นตอนของซัพพอร์ตเวกเตอร์รีเกรสชันแบบเชิงเส้นดังที่กล่าวมา โดยฟังก์ชันเคอร์เนลที่จะมีรูปแบบตามสมการที่ 2.12

$$K(x_i, x_j) = \langle \Phi(x_i) \cdot \Phi(x_j) \rangle \quad (2.12)$$

โดย x คือ เวกเตอร์ข้อมูลเข้า x_i คือ ซัพพอร์ตเวกเตอร์ และ Φ คือ ฟังก์ชันการส่งผ่านข้อมูล



ภาพที่ 2.3 การส่งผ่านข้อมูลจากปริภูมิข้อมูลเข้าที่ไม่เป็นเชิงเส้น

ไปยังปริภูมิลักษณะเด่นที่เป็นข้อมูลเชิงเส้น

การส่งผ่านข้อมูลด้วยฟังก์ชันเคอร์เนล จะหาค่าน้ำหนักได้สมการใหม่ดังสมการที่ 2.13

$$w = \sum_{i=1}^l (\alpha_i^* - \alpha_i) K(x, x_i) \quad (2.13)$$

หากนำ w จากสมการที่ 2.13 แทนค่าลงในสมการระนาบเงินที่เหมาะสมที่สุดจะได้สมการใหม่ดังสมการที่ 2.14

$$f(x) = \sum_{i=1}^l (\alpha_i^* - \alpha_i) K(x, x_i) + b \quad (2.14)$$

ซึ่งการหาคำตอบของสมการที่ 2.14 ต้องทำภายใต้เงื่อนไข

$$\sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0$$

$$0 \leq \alpha_i, \alpha_i^* \leq s_i C$$

ใช้หลักการของ Karush-Kuhn-Tucker (KKT) ในการปรับค่าที่อยู่ระหว่างขอบระนาบบนและขอบระนาบล่างให้เหมาะสมเพื่อหาค่าไบอัส b ที่เหมาะสม ดังสมการที่ 2.15

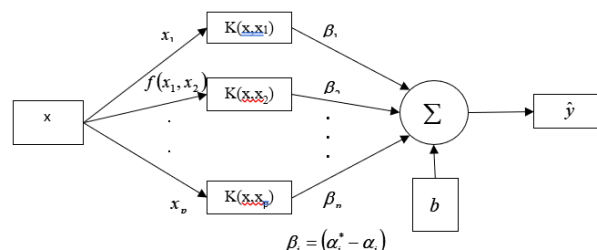
$$b = -\frac{1}{2} \sum_{i=1}^l (\alpha_i^* - \alpha_i) (K(x_i, x_r) + K(x_i, x_s)) \quad (2.15)$$

โดย x_r คือ ซัพพอร์ตเวกเตอร์ที่อยู่ระนาบบน และ x_s คือ ซัพพอร์ตเวกเตอร์ที่อยู่ระนาบล่าง

เคอร์เนลที่ใช้คือเรเดียลเบซิกฟังก์ชัน (Radial Basis Function; RBF) ดังสมการที่ 2.176

$$K(x, x_i) = \exp\left(-\|x - x_i\|^2 / 2\sigma^2\right) \quad (2.16)$$

ซึ่งภาพรวมของสถาปัตยกรรมซัพพอร์ตเวกเตอร์รีเกรสชันแสดงได้ดังรูปภาพที่ 2.4



ภาพที่ 2.4 สถาปัตยกรรมซัพพอร์ตเวกเตอร์รีเกรสชัน

2.3 การสร้างชุดฝึกสอน

โครงการนี้จัดทำขึ้นเพื่อพยากรณ์ราคาหุ้นในอนาคตโดยใช้ซัพพอร์ตเวกเตอร์รีเกรสชัน (Support Vector Regression: SVR)

ข้อมูลที่ใช้ฝึกสอนเป็นราคาหุ้นของบริษัทที่อยู่ในกลุ่มพลังงานและสาธารณูปโภค 5 บริษัท ซึ่งมีชื่อย่อบริษัทดังต่อไปนี้ BCP, IRPC, PTT, SUSCO และ TOP ตั้งแต่ วันที่ 1 มกราคม พ.ศ. 2546 ถึง 31 ธันวาคม พ.ศ. 2563 รวมทั้งสิ้น 18 ปี มีจำนวนวันทั้งสิ้น 6574 วัน และเป็นข้อมูลของราคาน้ำมันดิบในตลาดน้ำมันประเทศสหรัฐอเมริกา ตั้งแต่ วันที่ 1 มกราคม พ.ศ. 2546 ถึง 31 ธันวาคม พ.ศ. 2563 รวมทั้งสิ้น 18 ปี หรือมีจำนวนวันทั้งสิ้น 6574 วัน

2.3.1 Preprocessing

ในโครงการนี้ใช้ข้อมูลราคาหุ้นจากเว็บ Yahoo.com โดยข้อมูลราคาหุ้นเป็นแบบชนิดรายวัน ซึ่งข้อมูลแต่ละวันจะประกอบไปด้วย High คือ ราคาสูงสุดของวัน Low คือ ราคาต่ำสุดของวัน Open คือ ราคา ณ เวลาที่ตลาดหุ้นเปิดทำการของวัน Close คือ ราคา ณ เวลาที่ตลาดหุ้นปิดทำการของวัน และ Volume คือ จำนวนหุ้นที่มีการซื้อขายในวันนั้น ดังตาราง 1

และใช้ข้อมูลราคาน้ำมันดิบจากตลาดน้ำมันสหรัฐอเมริกาจากเว็บ Quandl.com ซึ่งประกอบไปด้วยราคาน้ำมันดิบของตลาดน้ำมันดิบสหรัฐอเมริกา หน่วยคือ ดอลลาร์สหรัฐต่อบาร์เรล ดังตาราง 2

ตาราง 1 ตัวอย่างข้อมูลพื้นฐานของหุ้นแต่ละบริษัท

Date	High	Low	Open	Close	Volume
2003-01-02	4.300	4.225	4.225	4.250	18449000.0
2003-01-03	4.275	4.225	4.250	4.225	15724000.0
2003-01-06	4.250	4.225	4.250	4.250	11980000.0
...					
2020-06-26	38.000	37.000	37.750	37.000	32050100.0
2020-06-29	37.500	36.500	36.500	37.250	35798700.0
2020-06-30	38.250	37.500	37.750	37.750	48960900.0

ตาราง 2 ตัวอย่างข้อมูลพื้นฐานของราคาน้ำมันดิบ

Date	Price
2003-01-02	30.05
2003-01-03	30.83
2003-01-06	30.71
...	
2020-06-26	37.99
2020-06-29	37.34
2020-06-30	38.22

เมื่อได้ข้อมูลตาราง 1 และ 2 แล้ว นำค่าที่ได้มารวมกันเป็นชุดข้อมูลสำหรับการฝึกสอนตาราง 3 โดยอาจจะมีการปรับจำนวนข้อมูลจำนวนวันของราคาหุ้นและราคาน้ำมันเพื่อความเหมาะสมและให้ได้ผลลัพธ์ที่ดีที่สุด โดยในโครงการนี้แบ่งเป็นการทดลองเป็น 3 ชุด ได้แก่ ชุดที่ 1 ใช้ข้อมูลหุ้นและราคาน้ำมันย้อนหลัง 1 วัน ชุดที่ 2 ใช้ข้อมูลหุ้นและราคาน้ำมันย้อนหลัง 3 วัน และชุดที่ 3 ใช้ข้อมูลหุ้นและราคาน้ำมันย้อนหลัง 7 วัน แล้วกำหนดให้ค่าเอาต์พุตคือ ราคาปิดของราคาหุ้นของวันถัดไปที่ตลาดหลักทรัพย์เปิดทำการ

ตาราง 3 ตัวอย่างข้อมูลชุดฝึกสอน

ลำดับข้อมูล	ข้อมูลราคาหุ้น				ข้อมูลราคาน้ำมัน				ค่าเอาต์พุต
1									
2									
3									
...									
N									

เมื่อจัดเตรียมข้อมูลสำหรับการฝึกสอนตาราง 3 แล้ว ค่าในแต่ละหลักมีขอบเขตของตัวเลขที่ต่างกันมากเกินไปจึงนำข้อมูลจากตาราง 3 ตั้งแต่หลักที่ 2 ถึง หลักสุดท้ายมาทำการ normalization โดยใช้ Standardization ดังสมการ 2.17

$$x = \frac{x - \bar{x}}{\sigma} \quad (2.17)$$

โดย x คือ ข้อมูลเดิม \bar{x} คือ ค่าเฉลี่ย และ σ คือ ค่าเบี่ยงเบนมาตรฐาน

2.3.2 การทดสอบโปรแกรมโดยใช้ชุดข้อมูลแบบบอด

เมื่อทำการสร้างตัวแบบจากชุดฝึกสอนแล้วจะนำตัวแบบที่ได้มาทดสอบกับข้อมูลแบบบอดโดยใช้ข้อมูลราคาหุ้นในกลุ่มพลังงานและสาธารณูปโภค ซึ่งในโครงการนี้ได้ทำการทดลองสร้างตัวแบบที่เหมาะสมจากการแบ่งเป็น 2 การทดลอง โดยการทดลองชุดที่ 1 ใช้ข้อมูลตั้งแต่ วันที่ 1 มกราคม พ.ศ. 2546 ถึง 31 ธันวาคม พ.ศ. 2562 เป็นข้อมูลในการฝึกสอน และใช้ข้อมูลตั้งแต่ วันที่ 1 มกราคม พ.ศ. 2563 ถึง 31 ธันวาคม พ.ศ. 2563 สำหรับการทดสอบแบบบอด และการทดลองชุดที่ 2 ใช้ข้อมูลตั้งแต่ วันที่ 1 มกราคม พ.ศ. 2546 ถึง 30 มิถุนายน พ.ศ. 2563 เป็นข้อมูลในการฝึกสอน และใช้ข้อมูลตั้งแต่ วันที่ 1 กรกฎาคม พ.ศ. 2563 ถึง 31 ธันวาคม พ.ศ. 2563 สำหรับการทดสอบแบบบอด

2.3.3 การวัดความถูกต้องจากการพยากรณ์ราคาหุ้น

การทดลองนี้วัดประสิทธิภาพการคำนวณของระบบโดยพิจารณาจาก ค่าสัมประสิทธิ์แสดงการตัดสินใจ (R^2) ดังสมการ 2.18 โดยผลลัพธ์ที่ได้จากตัวแบบจะเป็นราคาหุ้นและโปรแกรมจะทำการคาดคะเนผลลัพธ์ที่จะเกิดขึ้นในวันถัดไป

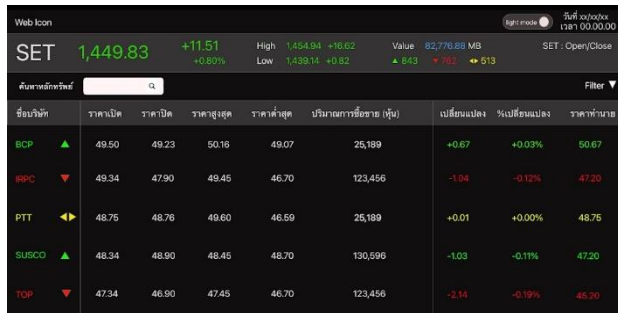
$$R^2 = 1 - \frac{\sum_{i=1}^n (Desire_i - Forecasting_i)^2}{\sum_{i=1}^n (Desire_i - \bar{Desire})^2} \quad (2.18)$$

2.4 การสร้างเว็บแอปพลิเคชัน

2.4.1 การพัฒนาส่วนการแสดงผล

ในส่วนของการพัฒนาส่วนแสดงผลเลือกใช้ React ซึ่งเป็นไลบรารี JavaScript โดยลักษณะการใช้งานเป็นรูปแบบเว็บแอปพลิเคชันที่สามารถเข้าได้ทุกอุปกรณ์ที่รองรับเว็บเบราว์เซอร์ ซึ่งมีการแสดงผลราคาของหุ้นที่ผ่านมาเป็นแบบกราฟแท่งเทียน และแสดงผลส่วนของการพยากรณ์ราคาที่จะเกิดขึ้นในอนาคตเป็นแบบกราฟเส้น ทั้งหมดแสดงผลใน

รูปแบบของวัน โดยผู้ใช้งานสามารถเปลี่ยนการแสดงผลราคาหุ้นของแต่ละบริษัทได้โดยการเลือกในส่วนของเมนู โดยมีส่วนของการแสดงผลดังรูปที่ 2.5 และ 2.6



รูปที่ 2.5 ส่วนแสดงผลหน้า Dash board



รูปที่ 2.6 ส่วนแสดงรายละเอียดราคาหุ้นย้อนหลัง

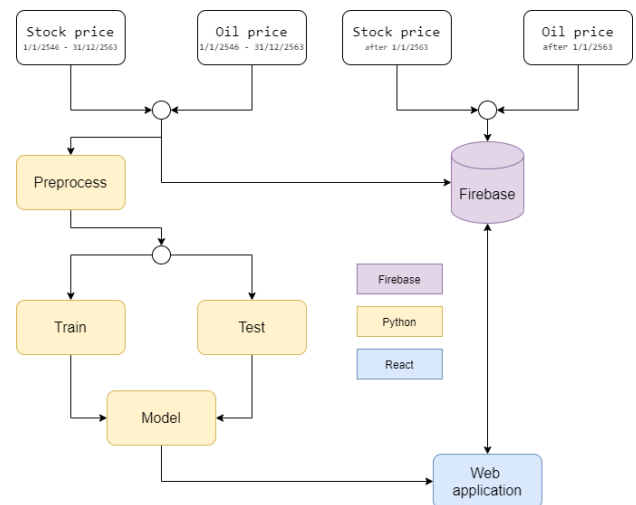
2.4.2 การพัฒนาส่วนการเก็บข้อมูล

ในส่วนของการเก็บข้อมูลเลือกใช้ Firebase ซึ่งมีการเป็นข้อมูลแบบ NoSQL ซึ่งในโครงการนี้เก็บข้อมูลราคาหุ้นประกอบด้วย ราคาเปิด-ปิด ราคาสูงสุด-ต่ำสุด ปริมาณการซื้อขาย เป็นรูปแบบรายวันตั้งแต่วันที่ 1 มกราคม พ.ศ. 2563 ถึง ปัจจุบัน และข้อมูลการพยากรณ์ราคาหุ้นซึ่งเป็นราคาปิดแบบรายวัน ตั้งแต่วันที่ 1 มกราคม พ.ศ. 2563 ถึง ปัจจุบัน

2.4.3 การ Deploy เว็บแอปพลิเคชัน

การ Deploy web hosting ใช้เครื่องมือของ Firebase ในการจัดการโดยเว็บจะทำงานบนเซิร์ฟเวอร์ของ Google firebase hosting

2.4.4 เส้นทางการไหลของข้อมูล (Data flow)



รูปที่ 2.7 เส้นทางการไหลของข้อมูล

3. ผลการทดลองและวิเคราะห์ผลการทดลอง

จากการทดลองโดยแบ่งชุดข้อมูลการทดลองเป็นสองชุดได้แก่ การทดลองชุดที่ 1 ใช้ข้อมูลตั้งแต่ วันที่ 1 มกราคม พ.ศ. 2546 ถึง 31 ธันวาคม พ.ศ. 2562 เป็นข้อมูลในการฝึกสอน และใช้ข้อมูลตั้งแต่ วันที่ 1 มกราคม พ.ศ. 2563 ถึง 31 ธันวาคม พ.ศ. 2563 สำหรับการทดสอบแบบบอด และการทดลองชุดที่ 2 ใช้ข้อมูลตั้งแต่ วันที่ 1 มกราคม พ.ศ. 2546 ถึง 30 มิถุนายน พ.ศ. 2563 เป็นข้อมูลในการฝึกสอน และใช้ข้อมูลตั้งแต่ วันที่ 1 กรกฎาคม พ.ศ. 2563 ถึง 31 ธันวาคม พ.ศ. 2563 สำหรับการทดสอบแบบบอด และในโครงการนี้แบ่งเป็นการทดลองเป็น 3 ชุด ได้แก่ ชุดที่ 1 ใช้ข้อมูลหุ้นและราคาน้ำมันย้อนหลัง 1 วัน ชุดที่ 2 ใช้ข้อมูลหุ้นและราคาน้ำมันย้อนหลัง 3 วัน และชุดที่ 3 ใช้ข้อมูลหุ้นและราคาน้ำมันย้อนหลัง 7 วัน แล้วกำหนดให้ค่าเอาต์พุตคือ ราคาปิดของราคาหุ้นของวันถัดไปที่ตลาดหลักทรัพย์เปิดทำการ ซึ่งได้ผลการทดลองตามตารางที่ 4 ดังนี้

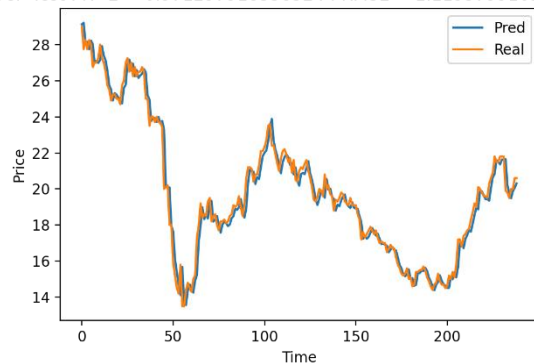
ตารางที่ 4 ผลการทดลองแสดงค่า RMSE และ ค่า R^2

หุ้น		จำนวนข้อมูลย้อนหลัง (วัน)		
		1	3	7
ชุดข้อมูล		R ²	R ²	R ²
BCP	1	0.971	0.970	0.967
	2	0.962	0.960	0.956
IRPC	1	0.688	0.050	-0.738
	2	0.641	-0.066	-0.792

PTT	1	0.932	0.928	0.889
	2	0.921	0.917	0.914
SUSCO	1	0.970	0.966	0.968
	2	0.913	0.905	0.902
TOP	1	0.955	0.943	0.904
	2	0.949	0.951	0.935

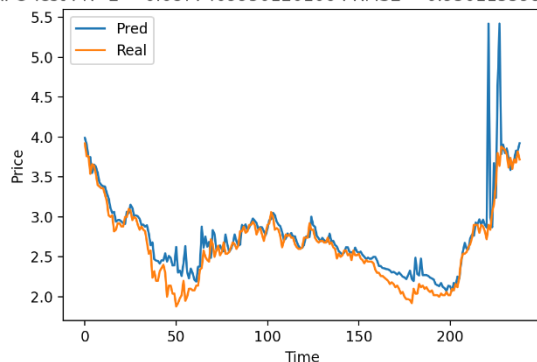
จากผลการทดลองพบว่าค่า R^2 ของตัวแบบที่ดีที่สุดเมื่อทดลองกับชุดทดสอบแบบบอดของในแต่ละบริษัทมีค่าอยู่ระหว่าง 0.688 – 0.971 ซึ่งจะเห็นได้ว่ามีความสอดคล้องและมีแนวโน้มไปในทิศทางเดียวกันเนื่องจากค่า R^2 มีค่าเข้าใกล้ 1 ดังนั้น ผู้พัฒนาจึงได้ทำการพล็อตกราฟผลการทดลองของแต่ละตัวแบบกับแต่ละชุดทดสอบแบบบอดจึงได้ผลลัพธ์ดังรูปที่ 3.1 – 3.5

BCP Test : $R^2 = 0.971297918358524$: RMSE = 1.119579916580277:



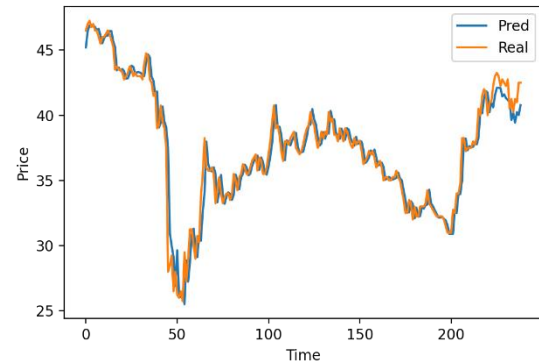
รูปที่ 3.1 ตัวแบบที่ดีที่สุดของหุ้น BCP กับ ชุดการทดสอบแบบบอด

IRPC Test : $R^2 = 0.6877409950120106$: RMSE = 0.95011559887404C



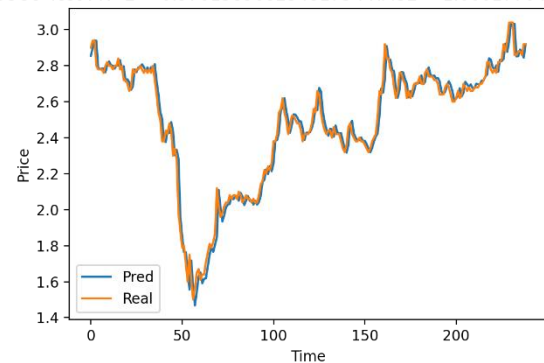
รูปที่ 3.2 ตัวแบบที่ดีที่สุดของหุ้น IRPC กับ ชุดการทดสอบแบบบอด

PTT Test : $R^2 = 0.932651667029434$: RMSE = 1.499717826429717



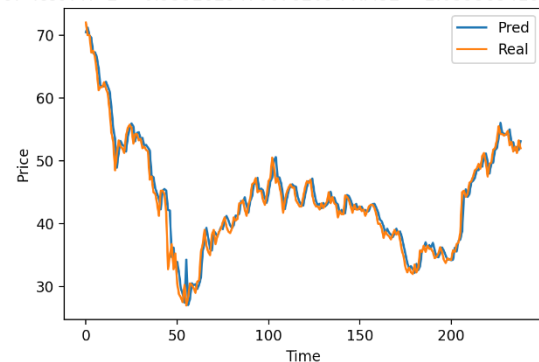
รูปที่ 3.3 ตัวแบบที่ดีที่สุดของหุ้น PTT กับ ชุดการทดสอบแบบบอด

USCO Test : $R^2 = 0.9701589682348173$: RMSE = 1.0001770466289C



รูปที่ 3.4 ตัวแบบที่ดีที่สุดของหุ้น SUSCO กับ ชุดการทดสอบแบบบอด

TOP Test : $R^2 = 0.9552823470678209$: RMSE = 2.039383418800605



รูปที่ 3.5 ตัวแบบที่ดีที่สุดของหุ้น TOP กับ ชุดการทดสอบแบบบอด

จะสังเกตได้ว่าตัวแบบที่ผู้พัฒนาได้พัฒนาขึ้นนั้นสามารถตอบรับตามแนวโน้มของราคาหุ้นได้ในทิศทางเดียวกัน แต่มีการตอบปรับที่ช้ากว่าเมื่อเทียบกับราคาจริงของตลาดหุ้น ซึ่งอาจจะเป็นผลเนื่องมาจากข้อมูลที่ต้องการทำนายเป็นรูปแบบ non stationary time series ซึ่งไม่สามารถ

ระบุค่าเฉลี่ยและค่าความแปรปรวน ได้อย่างแน่นอน อีกทั้งปัจจัยภายนอกที่มีส่วนเกี่ยวกับราคาหุ้นในระยะสั้น ส่งผลให้ตัวแบบทำนายผลได้ช้ากว่าความเป็นจริง

4. สรุป

จากการศึกษาเบื้องต้นจะพบว่า สามารถนำข้อมูลราคาหุ้นจาก Yahoo.com และราคาน้ำมันดิบในตลาดสหรัฐจาก Quandl.com ตั้งแต่วันที่ 1 มกราคม พ.ศ. 2546 ถึง 31 ธันวาคม พ.ศ. 2563 โดยทั้งหมดเป็นข้อมูลชนิดรายวัน ซึ่งข้อมูลหุ้นประกอบด้วยข้อมูล ราคาเปิด-ปิด ราคาสูงสุด-ต่ำสุด ปริมาณการซื้อขาย และข้อมูลราคาน้ำมันคือราคาขายหน่วยเป็นดอลลาร์สหรัฐต่อบาร์เรล รวมข้อมูลที่ใช้ในการทดลองทั้งสิ้น 18 ปี หรือมีจำนวนวันทั้งสิ้น 6574 วัน

จากนั้นนำเข้ามาผ่านกระบวนการ preprocessing โดยผ่านสมการ Standardization จากนั้นแบ่งการทดลอง ออกเป็น 2 ชุด โดยชุดที่ 1 ใช้ข้อมูลตั้งแต่วันที่ 1 มกราคม พ.ศ. 2546 ถึง 30 มิถุนายน พ.ศ. 2563 เป็นข้อมูลในการฝึกสอน และใช้ข้อมูลตั้งแต่วันที่ 1 กรกฎาคม พ.ศ. 2563 ถึง 31 ธันวาคม พ.ศ. 2563 สำหรับการทดสอบแบบบอด และในโครงการนี้แบ่งเป็นการทดลองเป็น 3 ชุด ได้แก่ ชุดที่ 1 ใช้ข้อมูลหุ้นและราคาน้ำมันย้อนหลัง 1 วัน ชุดที่ 2 ใช้ข้อมูลหุ้นและราคาน้ำมันย้อนหลัง 3 วัน และชุดที่ 3 ใช้ข้อมูลหุ้นและราคาน้ำมันย้อนหลัง 7 วัน แล้วกำหนดให้ค่าเอาต์พุตคือ ราคาปิดของราคาหุ้นของวันถัดไปที่ตลาดหลักทรัพย์เปิดทำการ

โดยผลการทดลองพบว่า การใช้ชุดข้อมูลที่ 1 โดยการใช้ข้อมูลหุ้นและน้ำมันดิบย้อนหลัง 1 วัน ในการฝึกสอนตัวแบบให้ผลลัพธ์ที่ดีที่สุด โดยได้ค่า R^2 ของตัวแบบที่ดีที่สุดเมื่อทดลองกับชุดทดสอบแบบบอดของในแต่ละบริษัทมีค่าอยู่ระหว่าง 0.688 – 0.971 ซึ่งจะเห็นได้ว่ามีความสอดคล้องและมีแนวโน้มเป็นไปในทิศทางเดียวกันเนื่องจาก ค่า R^2 มีค่าเข้าใกล้ 1 แต่หากดูข้อมูลจากกราฟแสดงผลในแต่ละหุ้นพบว่า ตัวแบบปรับเปลี่ยนช้ากว่าราคาในตลาดหุ้นจริง ซึ่งอาจจะเป็นผลเนื่องมาจากข้อมูลที่ต้องการทำนายเป็นรูปแบบ non stationary time series ซึ่งไม่สามารถระบุค่าเฉลี่ยและค่าความแปรปรวน ได้อย่างแน่นอน อีกทั้งปัจจัยภายนอกที่มีส่วนเกี่ยวกับราคาหุ้นในระยะสั้น ส่งผลให้ตัวแบบทำนายผลได้ช้ากว่าความเป็นจริง

ทั้งนี้ผู้พัฒนามีความคิดเห็นว่า การใช้ข้อมูลหุ้นและราคาน้ำมันดิบมาใช้ในการพยากรณ์ราคาปิดของหุ้นในวันถัดไปให้ถูกต้องเป็นไปได้ยากเนื่องจากรูปแบบของข้อมูลและปัจจัยภายนอก ในทางกลับกันหาก

ใช้ข้อมูลย้อนหลังจำนวนมากและพยากรณ์ราคาหุ้นในอนาคตที่ไกลออกไป ผลลัพธ์ที่ได้จะมีความถูกต้องที่น้อยลง

5. เอกสารอ้างอิง

- [1] สุพิชชา ตีรพัทธ์, เจียมจิตร ขวากร. (2555). “ปัจจัยทางเศรษฐกิจที่มีผลต่อการเคลื่อนไหวของดัชนีราคาหุ้นกลุ่มอุตสาหกรรมในตลาดหลักทรัพย์แห่งประเทศไทย : The Impact of Economic Factors on Industry Group Index in the Stock Exchange of Thailand”. KKU.
- [2] การจัดกลุ่มอุตสาหกรรมและหมวดหมู่ธุรกิจ. ค้นหามาเมื่อ 22 ตุลาคม 2563. Website: https://www.set.or.th/th/regulations/simplified_regulations/industry_sector_p1.html
- [3] จตุเมธ สุสุข. การวิเคราะห์ปัญหาคุณภาพไฟฟ้าโดยใช้ซัพพอร์ตเวกเตอร์แมชชีน = Power quality problem analysis using support vector machine. วิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเชียงใหม่, 2552.
- [4] นเรศ สุยะโรจน์. การทำนายค่ากำลังต้านย้อนกลับสำหรับระบบซีดีเอ็มเอ โดยใช้ซัพพอร์ตเวกเตอร์ แมชชีน = Reverse link power prediction for CDMA systems using support vector machine. วิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเชียงใหม่, 2551.