การพยากรณ์ความผันผวนของตลาดหลักทรัพย์โดยใช้ซัพพอร์ตเวกเตอร์แมชชีน

FORCASETING STOCK MARKET VOLATILITY WITH SUPPORT VECTOR MACHINE

กรอภิชิต เหง้าพันธ์[1] และ สมพร ปั่นโภชา[2]

[1]วิทยาศาสตรมหาบัณฑิต สาขาวิศวกรรมการเงิน คณะวิทยาศาสตร์ มหาวิทยาลัยหอการค้าไทย, aceskorn555@outlook.com

[2]วิทยาศาสตรมหาบัณฑิต สาขาวิศวกรรมการเงิน คณะวิทยาศาสตร์ มหาวิทยาลัยหอการค้าไทย, somporn_pun@utcc.ac.th

---

บทคัดย่อ

ซัพพอร์ตเวกเตอร์แมชชีนเป็นวิธีการเรียนรู้ของเครื่องอย่างหนึ่งที่ถูกนำไปประยุกต์ใช้อย่างแพร่หลายในหลาย ๆ ด้าน แต่การประยุกต์ใช้ทางด้านการเงินยังไม่เป็นที่แพร่หลายมากนักโดยเฉพาะกับข้อมูลทางการเงินในประเทศไทย สำหรับงานวิจัยนี้มีวัตถุประสงค์เพื่อจะศึกษาความสามารถในการพยากรณ์ความผันผวนของดัชนีตลาดหลักทรัพย์แห่งประเทศไทย (Stock Exchange of Thailand Index: SET Index) และเปรียบเทียบผลการพยากรณ์กับแบบจำลองความผันผวนอื่น ๆ โดยใช้ราคาปิดรายวันที่ปรับปรุงแล้ว (Adjusted Close Price) ของ SET Index ตั้งแต่วันที่ 1 มกราคม พ.ศ. 2553 ถึง 31 ธันวาคม พ.ศ. 2564 มาหาผลตอบแทนรายวัน จากนั้นทำการวิเคราะห์และพยากรณ์ความผันผวนด้วยแบบจำลองซัพพอร์ตเวกเตอร์แมชชีนและเปรียบเทียบผลการพยากรณ์กับแบบจำลอง GARCH ซึ่งเป็นเทคนิคที่นิยมใช้ในการวิเคราะห์และพยากรณ์ข้อมูลอนุกรมเวลา ผลการเปรียบเทียบค่า RMSE พบว่าแบบจำลองซัพพอร์ตเวกเตอร์แมชชีนมีค่ารากที่สองของค่าเฉลี่ยความผิดพลาดกำลังสอง (Root Mean Square Error: RMSE) น้อยกว่าแบบจำลองประเภท GARCH แสดงถึงความสามารถในการพยากรณ์ที่เหนือกว่า

คำสำคัญ: การเรียนรู้ของเครื่อง, ซัพพอร์ตเวกเตอร์แมชชีน, การพยากรณ์ความผันผวนของตลาดหลักทรัพย์

ABSTRACT

Support Vector Machine (SVM) is one type of machine learning that has been applied in various fields. In the financial field, although some of SVM's applications have been studied, they remain largely unexplored, especially in Thailand. There are two main objectives in this study; the first is to evaluate the accuracy of the predicted SET Index's volatility by using an SVM model; the second is to compare the accuracy of the SVM model's prediction with GARCH-type models' prediction. In this paper, the author applies the SVM method to the Stock Exchange of Thailand Index (SET index) by using adjusted close prices in the timeframe from 1st January 2010 to 31st December 2021. The comparison of RMSE between SVM models and GARCH-type models show that SVM models have more accuracy in volatility forecasting than GARCH-type models.

Keywords: Machine learning, Support vector machine, Stock market volatility forecasting

1. บทนำ

ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine: SVM) เป็นหนึ่งในวิธีการเรียนรู้ของเครื่อง (Machine learning) ประเภทแบบจำลองการเรียนรู้ภายใต้การกำกับ (Supervised learning) ด้วยขั้นตอนวิธี (Algorithm) ที่ใช้ในการวิเคราะห์ข้อมูลเพื่อจำแนกประเภท (Classification) และการวิเคราะห์การถดถอย (Regression) ซึ่งถูกพัฒนาโดย Vladimir Vapnik และเพื่อนร่วมงานที่ทำงานอยู่ใน AT&T Bell Laboratories (Corinna Cortes and Vladimir Vapnik, 1995)

SVM ได้ถูกพัฒนาในกรอบแนวคิดของทฤษฎีการเรียนรู้เชิงสถิติและถูกนำไปประยุกต์ใช้ในหลาย ๆ ด้าน ไม่ว่าจะเป็นการประยุกต์ใช้ SVM เพื่อสนับสนุนการตัดสินใจทางการแพทย์, การทำนายข้อมูลที่เป็นอนุกรมเวลา (Time series), การระบุตัวตนด้วยใบหน้า เป็นต้น (Evgeniou and Pontil, 2001) จากงานศึกษาเรื่อง การประยุกต์ใช้ SVM ในงานวิจัยทางการเงิน (Cao, Zhan and Wu, 2009) ที่ได้ศึกษางานวิจัยที่เกี่ยวข้องกับการประยุกต์ใช้ SVM ในด้านการเงินและการลงทุนพบว่ามีการนำ SVM มาประยุกต์ใช้หลัก ๆ ได้แก่ การประยุกต์ใช้ SVM เพื่อพยากรณ์การเคลื่อนไหวของตลาดหลักทรัพย์, การหาสัญญาณเตือนล่วงหน้า (Early-warning) ทางการเงินของบริษัท, การวิเคราะห์การจัดอันดับเครดิตของบริษัท

แม้ว่า SVM ถูกนำมาประยุกต์ใช้ในหลาย ๆ ด้าน รวมถึงในทางการเงินแล้วก็ตาม แต่ยังไม่ได้มีการศึกษาในวงกว้างและยังมีข้อมูลที่น่าสนใจอีกมากที่ยังไม่ได้ถูกศึกษา ดังนั้น Gavrishchaka และ Ganguli ได้มีการนำเสนอแบบจำลองที่มีพื้นฐานมาจาก SVM เพื่อที่จะใช้ในการพยากรณ์ความผันผวนจากข้อมูลตลาดแบบหลายมาตราส่วน (Multi-scale) และมีมิติสูง (High-dimensional) (Gavrishchaka and Ganguli, 2003)

ซัพพอร์ตเวกเตอร์แมชชีนกรอบแนวคิดที่มีประสิทธิภาพสำหรับการพยากรณ์ความผันผวนของตลาดหลักทรัพย์ (Gavrishchaka and Banerjee, 2006) ได้นำเสนอแบบจำลองซัพพอร์ตเวกเตอร์แมชชีนว่าเป็นแบบจำลองที่มีประสิทธิภาพในการดึงข้อมูลจากข้อมูลตลาดที่มีลักษณะหลายขนาดและมิติสูง โดยนำเสนอผลของการประยุกต์ใช้ซัพพอร์ตเวกเตอร์แมชชีนกับดัชนี SP500 พบว่าผลการพยากรณ์ด้วยแบบจำลองซัพพอร์ตเวกเตอร์แมชชีนนั้นเหนือกว่าแบบจำลองความผันผวนกระแสหลัก

งานวิจัยที่มีการศึกษาในลักษณะคล้ายกันโดย Chen (2010) ชื่อ การพยากรณ์ความผันผวนด้วยซัพพอร์ตเวกเตอร์แมชชีนที่มีพื้นฐานจากแบบจำลอง GARCH ได้ทำการศึกษาโดยใช้ข้อมูลอัตราแลกเปลี่ยน GBP/USD และข้อมูล NYSE ได้ผลลัพธ์ไปในทิศทางเดียวกันกับ Gavrishchaka คือ แบบจำลองซัพพอร์ตเวกเตอร์แมชชีนที่มีพื้นฐานมาจากแบบจำลอง GARCH มีประสิทธิภาพเหนือกว่าแบบจำลองกระแสหลักอื่น ๆ อย่างมากในการพยากรณ์ความผันผวนล่วงหน้าหนึ่งช่วง ซึ่งยืนยันข้อได้เปรียบทางทฤษฎีของแบบจำลองซัพพอร์ตเวกเตอร์แมชชีน

สำหรับงานวิจัยฉบับนี้มีลักษณะคล้ายกับงานวิจัยของ Gavrishchaka และ Banerjee (2006) โดยนำราคาปิดของดัชนีตลาดหลักทรัพย์แห่งประเทศไทยมาใช้ในการหาผลตอบแทนรายวัน และทำการวิเคราะห์และพยากรณ์ความผันผวนด้วยแบบจำลองกระแสหลัก จากนั้นทำการวิเคราะห์และพยากรณ์ความผันผวนด้วยแบบจำลองแบบจำลองซัพพอร์ตเวกเตอร์แมชชีน ในการเปรียบเทียบความสามารถในการพยากรณ์ของแต่ละแบบจำลองผู้วิจัยได้เลือกใช้ค่ารากที่สองของค่าเฉลี่ยความผิดพลาดกำลังสอง (Root Mean Square Error: RMSE) เนื่องจากเป็นค่าสถิติที่นิยมใช้เพื่อเปรียบเทียบระหว่างแบบจำลองการวิเคราะห์เชิงถดถอยหลาย ๆ แบบจำลอง

ความผันผวนในตลาดหลักทรัพย์ถือว่ามีความสำคัญอย่างมากในการวิเคราะห์และตัดสินใจทางการเงินไม่ว่าจะเป็นการใช้เพื่อหาราคาตราสารอนุพันธ์ (Derivative pricing) หรือการคำนวณหามูลค่าของความเสี่ยง (Value-at-Risk) ในการบริหารความเสี่ยงของพอร์ตฟอลิโอ และยังเป็นองค์ประกอบหนึ่งที่ใช้ในการตัดสินใจในระบบการซื้อขาย (Trading systems) โดยทางผู้จัดทำหวังว่างานวิจัยชิ้นนี้จะสามารถเป็นพื้นฐานในการประยุกต์ใช้การเรียนรู้ของเครื่องกับข้อมูลตลาดการเงินในประเทศไทยต่อไป

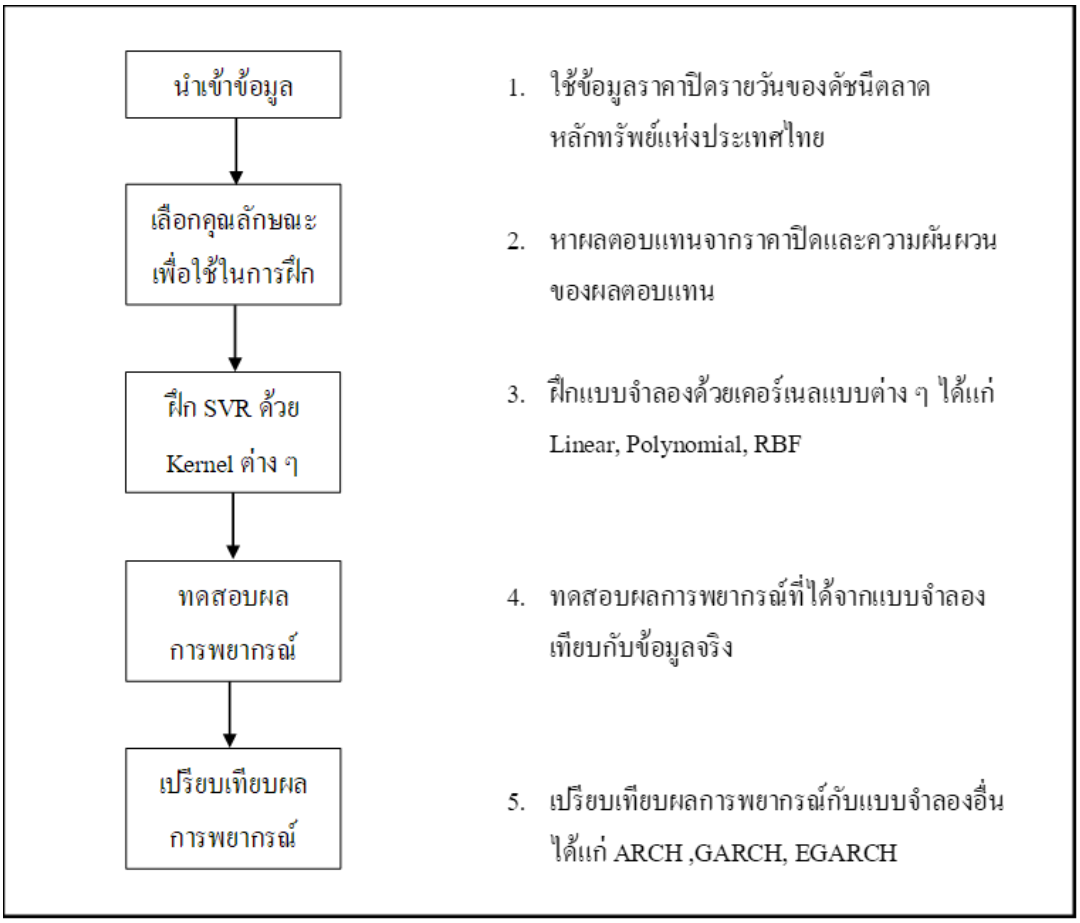
2. วัตถุประสงค์การวิจัย

1. เพื่อศึกษาการนำแบบจำลองที่มีพื้นฐานจากซัพพอร์ตเวกเตอร์แมชชีนมาประยุกต์ใช้ในการพยากรณ์ความผันผวนของดัชนีตลาดหลักทรัพย์แห่งประเทศไทย
2. เพื่อเปรียบเทียบประสิทธิภาพของแบบจำลองที่มีพื้นฐานจากซัพพอร์ตเวกเตอร์แมชชีนในการพยากรณ์ความผันผวนของดัชนีตลาดหลักทรัพย์แห่งประเทศไทย กับแบบจำลองอื่น ๆ

3. การดำเนินการวิจัย

ขั้นตอนการดำเนินงาน

1. รวบรวมข้อมูลราคาปิดรายวันของดัชนีตลาดหลักทรัพย์แห่งประเทศไทย
2. คำนวณผลตอบแทนจากราคาปิดและความผันผวนของผลตอบแทน
3. ฝึกแบบจำลองด้วยเคอร์เนลฟังก์ชันแบบต่าง ๆ ได้แก่ เชิงเส้น (Linear), แบบพหุนา (Polynomial), และแบบรัศมี (RBF)
4. ทดสอบผลการพยากรณ์ที่ได้จากแบบจำลองเปรียบเทียบกับข้อมูลจริงโดยใช้ค่า RMSE
5. เปรียบเทียบผลการพยากรณ์กับแบบจำลองอื่น ได้แก่ แบบจำลอง ARCH , แบบจำลอง GARCH และแบบจำลอง EGARCH



รูปที่ 1 ขั้นตอนวิธีการศึกษาวิจัย

โดยแต่ละแบบจำลองที่ใช้มีแนวคิดและทฤษฎีดังนี้

1. แบบจำลอง ARCH

แบบจำลอง ARCH (p) เป็นแบบจำลองความผันผวนแบบตัวแปรเดียวและอ้างอิงกับผลตอบแทนในอดีตของสินทรัพย์ตัวนั้น ๆ เมื่อ p คือลำดับของความล่าช้าของ r_t^2 สามารถเขียนได้ดังสมการที่ (1)

$$\sigma_t^2 = \omega + \sum_{k=1}^{p} \alpha_k (r_{t-k})^2 \qquad (1)$$

เมื่อ r_t คือ อัตราผลตอบแทน ณ เวลา t

σ_t คือ คือความแปรปรวน ณ เวลา t

ω คือ ค่าสัมประสิทธิ์

α_k คือ ค่าสัมประสิทธิ์ ณ เวลา k

ϵ_t คือ กระบวนการสุ่มที่มีการแจกแจงแบบปกติ ณ เวลา t

และมีสมมติฐานสำคัญคือค่าความแปรปรวนจะต้องเป็นบวกเสมอ จะได้ว่า

$$\omega > 0 \, , \alpha_k \geq 0$$

2. แบบจำลอง GARCH

เป็นแบบจำลองที่ต่อยอดมาจากแบบจำลอง ARCH โดยมีการเพิ่มความแปรปรวนแบบมีเงื่อนไขในอดีตเข้าไปด้วย แบบจำลอง GARCH (p, q)จึงเป็นแบบจำลองหลายตัวแปร สำหรับความแปรปรวนของอนุกรมเวลาเป็นแบบ ARMA (Autoregressive Moving Average) เมื่อ p เป็นลำดับของความล่าช้าของ σ_t^2 และ q เป็นลำดับความล่าช้าของ r_t^2 สามารถเขียนได้ดังสมการที่ (2)

$$\sigma_t^2 = \omega + \sum_{k=1}^{q} \alpha_k r_{t-k}^2 + \sum_{k=1}^{p} \beta_k \sigma_{t-k}^2 \qquad (2)$$

เมื่อ β_k คือ ค่าสัมประสิทธิ์ ณ เวลา k

โดยที่ $\omega, \beta,$ และ α เป็นค่าสัมประสิทธิ์ที่ประมาณขึ้นและ p และ q คือความล่าช้าสูงสุดของแบบจำลอง มีเงื่อนไขดังนี้

$$\omega > 0\ ,\ \beta \geq 0\ , \alpha \geq 0 \text{ และ } \beta + \alpha < 1$$

3. แบบจำลอง EGARCH

แบบจำลอง Exponential General AutoRegressive Conditional Heteroskedastic (EGARCH) มีสมติฐานเพิ่มเติมคือการตอบสนองของตลาดที่ไม่สมมาตรกันของข่าวสาร (Asymmetric effects of announcements) ระหว่างข่าวดีและข่าวร้ายหรือการแจกแจงของผลตอบแทนฝั่งขาดทุน (losses) มีลักษณะเป็นหางอ้วน (fat tail) มากกว่าฝั่งกำไร (gain) แบบจำลอง EGARCH (p, q) เมื่อ p เป็นลำดับของความล่าช้าของ σ_t^2 และ q เป็นลำดับความล่าช้าของ r_t^2 สามารถเขียนได้ดังสมการที่ (3)

$$\log(\sigma_t^2) = \omega + \sum_{k=1}^{p} \beta_k log\sigma_{t-k}^2 + \sum_{k=1}^{q} \alpha_i \frac{|r_{k-1}|}{\sqrt{\sigma_{t-k}^2}} + \sum_{k=1}^{q} \gamma_k \frac{r_{t-k}}{\sqrt{\sigma_{t-k}^2}} \qquad (3)$$

เมื่อ γ_k คือ ค่าสัมประสิทธิ์ ณ เวลา k

4. ซัพพอร์ตเวกเตอร์แบบถดถอย (Support Vector Regression)

ซัพพอร์ตเวกเตอร์แมชชีนเป็นแบบจำลองการเรียนรู้ของเครื่องที่สามารถวิเคราะห์ข้อมูลแบบเชิงเส้น (Linear) และไม่เชิงเส้น (Nonlinear) ทั้งแบบการจำแนกประเภท (Classification) และการวิเคราะห์การถดถอย (Regression) ซึ่งเหมาะกับข้อมูลที่มีขนาดเล็กไปจนถึงขนาดกลาง

ในการวิเคราะห์เชิงถดถอยกำหนดให้เวกเตอร์สำหรับฝึก $\boldsymbol{x}_i \in \mathbb{R}^{\mathrm{p}}, i = 1, \ldots, n$ และเวกเตอร์ $\boldsymbol{y} \in \mathbb{R}^n$ $\boldsymbol{\varepsilon}$-SVR สามารถแก้ปัญหาดั้งเดิม (Primal problem) ได้ดังสมการที่ (4)

ฟังก์ชันเป้าประสงค์: $\min \quad \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{l}(\xi_i + \xi_i^*)$ (4)

เงื่อนไขบังคับ:
$$y_i - \langle w, x_i \rangle - b \leq \varepsilon + \xi_i$$
$$\langle \omega, x_i \rangle + b - y_i \leq \varepsilon + \xi_i^*$$
$$\xi_i,\ \xi_i^* \geq 0$$

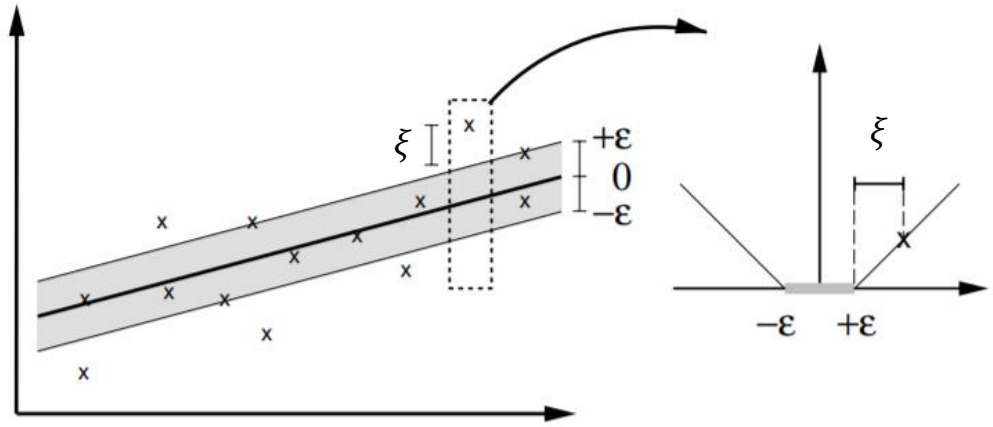
เมื่อ $\mathbf{w}$ คือ เวกเตอร์น้ำหนัก

b คือ ค่าสัมประสิทธิ์

ξ_i คือ ตัวแปรส่วนขาด (*Slack variable)*

มีฟังก์ชันสูญเสียแบบ ε- Insensitive โดยเป้าหมายในการฝึกแบบจำลองก็คือการหาฟังก์ชันการตัดสินใจที่ y_i ไม่เกินค่า ε สำหรับทุก ๆ x_i ดังแสดงในสมการที่ (5)

$$|\xi|_\varepsilon := \begin{cases} 0 & \text{if } |\xi| \le \varepsilon \\ |\xi| - \varepsilon & \text{otherwise} \end{cases} \tag{5}$$



รูปที่ 2 (ซ้าย) Support Vector Regression (ขวา) ε- Insensitive Loss Function

สามารถเขียนให้อยู่ในรูปของปัญหาควบคู่ (Dual problem) ดังนี้

ฟังก์ชันเป้าประสงค์:

$$\text{Max} \quad -\frac{1}{2}\sum_{i,j=1}^{l}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)k(x_i, x_j) - \varepsilon\sum_{i=1}^{l}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{l} y_i(\alpha_i - \alpha_i^*) \tag{6}$$

เงื่อนไขบังคับ:

$$-\varepsilon\sum_{i=1}^{l}(\alpha_i - \alpha_i^*) = 0 \qquad \alpha_i, \alpha_i^* \in [0, C]$$

สมการคาดการณ์คือ

$$f(x) = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)k(x_i, x) + b \tag{7}$$

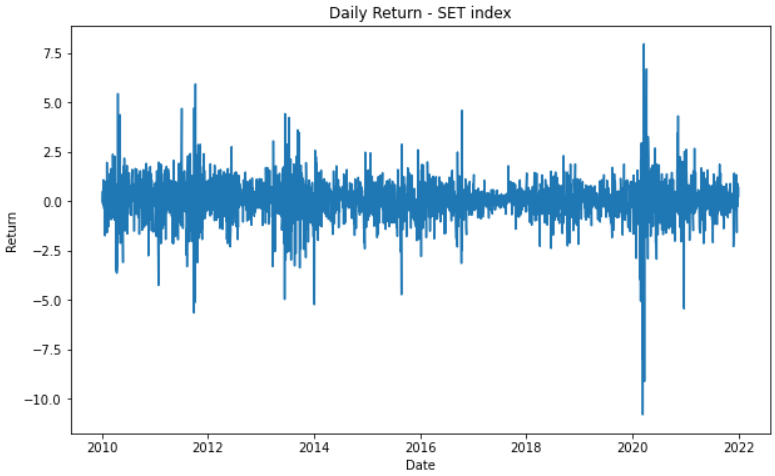
สำหรับเคอร์เนลเป็นฟังก์ชันสามารถคำนวณผลจากคูณเชิงสเกลาร์ (Dot Product) $\Phi(\mathbf{a})^T\Phi(\mathbf{b})$ ของเวกเตอร์ $\mathbf{a}$ และ $\mathbf{b}$ โดยไม่ต้องคำนวณการแปลงของ Φ ดังแสดงในสมการชุดที่ (8)

$$
\begin{aligned}
&\text{Linear:} && K(\mathbf{a},\mathbf{b}) = \mathbf{a}^T\mathbf{b} \\
&\text{Polynomial:} && K(\mathbf{a},\mathbf{b}) = (\gamma\mathbf{a}^T\mathbf{b} + r)^d \\
&\text{Gaussian RBF:} && K(\mathbf{a},\mathbf{b}) = \exp(-\gamma\|\mathbf{a}-\mathbf{b}\|^2) \\
&\text{Sigmoid:} && K(\mathbf{a},\mathbf{b}) = \tanh(\gamma\mathbf{a}^T\mathbf{b} + r)
\end{aligned}
\qquad (8)
$$

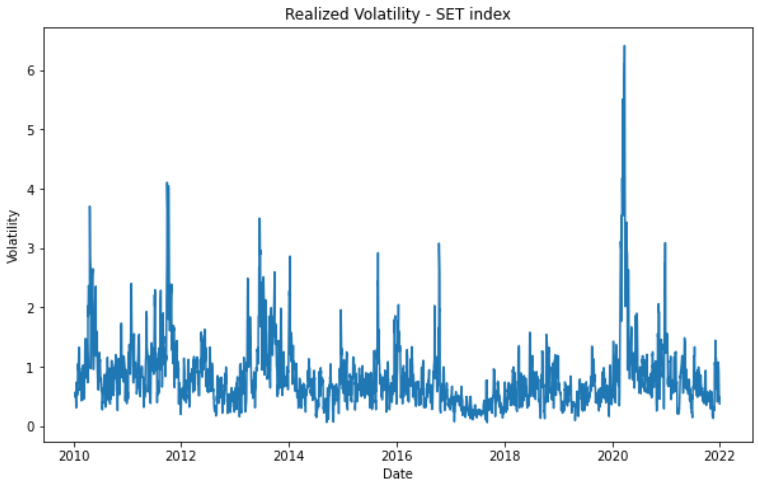
งานวิจัยนี้ได้ทำการวิเคราะห์และพยากรณ์โดยใช้ภาษาไพทอน (Python) และแบบจำลองที่ใช้ในการวิเคราะห์ทั้งหมด ผู้วิจัยได้ใช้ฟังก์ชันจากไลบรารี่ (Library) ที่ถูกพัฒนาขึ้นโดยนักพัฒนาท่านอื่น โดยมีไลบรารี่ที่สำคัญ ได้แก่ arch และ sklearn โดย code และทฤษฎีที่เกี่ยวข้องอ้างอิงมาจากหนังสือ Machine Learning for Financial Risk Management with Python (2022) โดย Abdullah Karasa

4. ผลการวิจัย

ในการจัดเตรียมข้อมูลเพื่อใช้ในการฝึกแบบจำลองการเรียนรู้ของเครื่องและการทดสอบความสามารถในการพยากรณ์ ต้องนำข้อมูลของราคาปิดที่ปรับปรุงแล้วของดัชนีตลาดหลักทรัพย์แห่งประเทศไทย (SET index) ระหว่างวันที่ 1 ม.ค. 2553 ถึง 31 ธ.ค. 2564 มาหาผลตอบแทนรายวันดังแสดงในรูปที่ 3 และหาความผันผวนของอัตราผลตอบแทนรายวันตั้งแต่วันที่ 1 ม.ค. 2553 ถึง 31 ธ.ค. 2564 ดังรูปที่ 4 จะเห็นได้ว่าผลตอบแทนรายวันของข้อมูลที่ศึกษาครอบคุมช่วงที่ตลาดมีความผันผวนหลายช่วง ไม่ว่าจะเป็นช่วงหลังวิกฤติซับไพร์มหรือช่วงที่มีการแพร่ระบาดอย่างรุนแรงของเชื้อไวรัสโควิด-19



รูปที่ 3 ผลตอบแทนรายวันของดัชนีตลาดหลักทรัพย์แห่งประเทศไทย



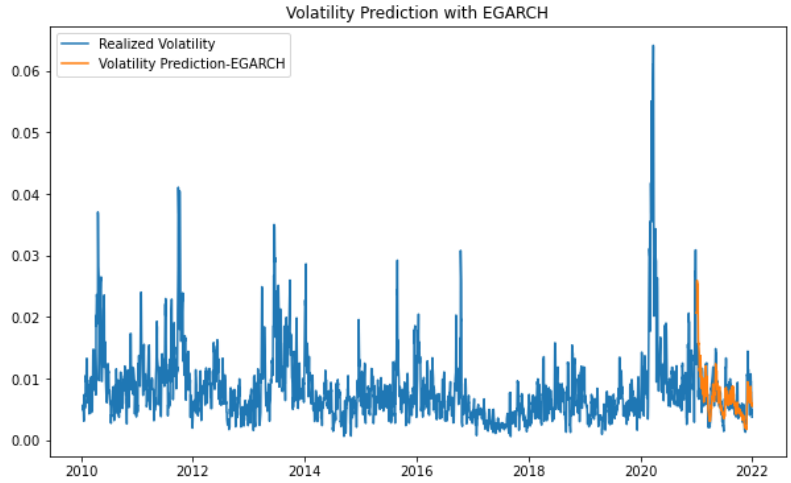
รูปที่ 4 ความผันผวนของอัตราผลตอบแทนรายวันของ SET index

จากการเปรียบเทียบผลของการพยากรณ์กับความผันผวนที่เกิดขึ้นจริงในปีพ.ศ. 2564 จากนั้นคำนวณค่า RMSE ออกมาได้ผลดังตารางที่ 1 สามารถแปลความหมายของค่า RMSE ได้ว่าหากค่า RMSE ยิ่งต่ำแสดงถึงความผิดพลาดในการพยากรณ์ที่ต่ำกว่าหรืออาจกล่าวได้ว่า แบบจำลองสามารถพยากรณ์ได้แม่นยำกว่านั่นเอง โดยสามารถเรียงลำดับแบบจำลองที่มีค่า RMSE ต่ำสุดไปหาสูงสุดได้ดังนี้ SVR RBF, SVR Linear, SVR Polynomial, EGARCH, GARCH, และ ARCH ตามลำดับ ซึ่งสรุปผลการพยากรณ์ของแต่ละแบบจำลองดังตารางที่ 1

ตารางที่ 1 ผลการพยากรณ์ของแต่ละแบบจำลอง

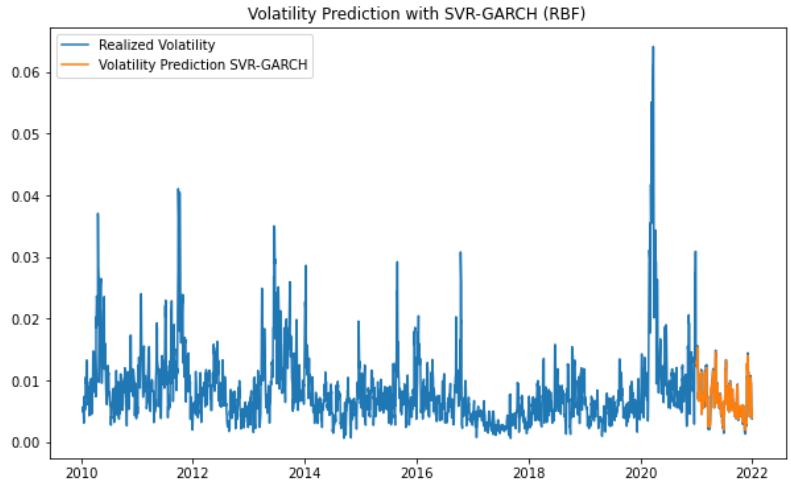
Model	RMSE score
ARCH	0.0794
GARCH	0.0783
EGARCH	0.0773
SVR Linear	0.000519
SVR Polynomial	0.001807
SVR RBF	0.000313

จากตารางที่ 1 จะเห็นได้ว่าแบบจำลองที่ดีที่สุดของแบบจำลองประเภท GARCH คือ EGARCH และแบบจำลองที่ดีที่สุดของแบบจำลองประเภท SVR คือ SVR RBF โดยพิจารณาจากค่า RMSE และเมื่อนำผลการพยากรณ์ของแบบจำลองที่ดีที่สุดทั้งสองมาเทียบกับข้อมูลจริง จะได้ผลดังรูปที่ 5 และรูปที่ 6 ตามลำดับ



รูปที่ 5 ผลการพยากรณ์ของแบบจำลอง EGARCH(1,1)

จากรูปที่ 6 แสดงการเปรียบเทียบค่าความผันผวนที่เกิดขึ้นจริง(สีฟ้า) กับความผันผวนที่ได้จากการพยากรณ์ของแบบจำลอง(สีส้ม) จะเห็นได้ว่าผลการพยากรณ์ของแบบจำลอง EGARCH(1,1) ยังมีความผันผวนน้อยกว่าข้อมูลที่เกิดขึ้นจริงในปี 2021 ถึง 2022 แต่ยังดีกว่าแบบจำลอง ARCH(4) และ GARCH(1,1) (พิจารณาจากค่า RMSE)



รูปที่ 6 ผลการพยากรณ์ด้วยแบบจำลอง SVR-GARCH (RBF)

จากรูปที่ 6 แสดงการเปรียบเทียบค่าความผันผวนที่เกิดขึ้นจริง(สีฟ้า) กับความผันผวนที่ได้จากการพยากรณ์ของแบบจำลอง(สีส้ม) จะเห็นได้ว่าผลการพยากรณ์ความผันผวนของแบบจำลอง SVR-GARCH (Linear) ใกล้เคียงกับข้อมูลที่เกิดขึ้นจริงในปี 2021 ถึง 2022 และมีค่า RMSE ต่ำกว่าแบบจำลอง ARCH, GARCH, EGARCH และยังมีค่า RMSE ต่ำที่สุดเมื่อเทียบกับแบบจำลอง SVR-GARCH ที่มีเคอร์เนลฟังก์ชันแบบอื่น ๆ

5. การอภิปรายผล

งานวิจัยนี้ใช้ซัพพอร์ตเวกเตอร์แมชชีนในการพยากรณ์ความผันผวนของผลตอบแทนของดัชนีตลาดหลักทรัพย์แห่งประเทศไทย จากนั้นทำการเปรียบเทียบความสามารถในการพยากรณ์ของแต่ละแบบจำลองด้วยค่าความผันผวนที่เกิดขึ้นจริงโดยใช้รากที่สองของค่าเฉลี่ยความผิดพลาดกำลังสองหรือ RMSE โดยแบบจำลองที่ดีที่สุดคือ SVR RBF ที่มีค่า RMSE อยู่ที่ 0.000313 ในขณะที่แบบจำลองที่แย่ที่สุดคือ ARCH ที่มีค่า RMSE อยู่ที่ 0.0794

เมื่อเทียบกับแบบจำลองประเภทเดียวกันแบบจำลองประเภท GARCH ให้ผลในการพยากรณ์หรือ RMSE ใกล้เคียงกัน ในขณะที่แบบจำลองประเภท SVR ให้ผลการพยากรณ์ที่ค่อนข้างต่างกันโดยเฉพาะ SVR Polynomial ซึ่งในที่นี้ใช้พหุนามแบบยกกำลังสองทำให้ความสามารถในการณ์พยากรณ์ต่ำเมื่อเทียบกับแบบจำลอง คือ SVR Linear และ SVR RBF

ผลการทดลองที่ได้ในงานวิจัยฉบับนี้สอดคล้องกับผลการศึกษาของ Gavrishchaka และ Banerjee (2006) ที่พบว่าแบบจำลองซัพพอร์ตเวกเตอร์แมชชีนมีความสามารถในการพยากรณ์ที่เหนือกว่าแบบจำลอง GARCH-type

6. บทสรุปและข้อเสนอแนะ

บทสรุป

จากงานวิจัยนี้พบว่าแบบจำลองที่มีพื้นฐานจากซัพพอร์ตเวกเตอร์แมชชีนที่ใช้เคอร์เนลฟังก์ชันแบบรัศมี (RBF) มีความสามารถในการพยากรณ์ความผันผวนของดัชนีตลาดหลักทรัพย์แห่งประเทศไทยได้แม่นยำที่สุดเมื่อเทียบกับเคอร์เนลฟังก์ชันแบบอื่น ๆ

และเมื่อเปรียบเทียบกับประสิทธิภาพของแบบจำลองที่มีพื้นฐานจากซัพพอร์ตเวกเตอร์แมชชีนในการพยากรณ์ความผันผวนของดัชนีตลาดหลักทรัพย์แห่งประเทศไทย กับแบบจำลอง ARCH, GARCH, และ EGARCH พบว่าแบบจำลองที่มีพื้นฐานจากซัพพอร์ตเวกเตอร์แมชชีนมีความสามารถในการพยากรณ์ที่เหนือกว่า

ข้อเสนอแนะ

1. การทดสอบความสามารถในการพยากรณ์ความผันผวนของแต่ละแบบจำลอง ใช้วิธีการเทียบค่าความผันผวนที่ได้จากการพยากรณ์กับค่าความผันผวนที่เกิดขึ้นจริงโดยใช้รากที่สองของค่าเฉลี่ยความผิดพลาดกำลังสองหรือ RMSE ในการศึกษาครั้งต่อ ๆ ไป อาจมีการวัดความสามารถในการพยากรณ์ด้วยค่าทางสถิติอื่น ๆ เช่น ค่าเปอร์เซ็นต์ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error : MAPE), ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Error : MAE) เป็นต้น

2. การเปรียบเทียบความสามารถในการพยากรณ์ความผันผวน อาจมีการเพิ่มแบบจำลองการเรียนรู้ของเครื่องอื่น ๆ เข้ามาศึกษาเปรียบเทียบด้วย เช่น โครงข่ายประสาทเทียม (Artificial Neuron Network: ANN) เป็นต้น

3. ในงานวิจัยนี้ ไม่ได้มีการทดสอบปัญหา Overfitting และ Underfitting ซึ่งเป็นปัญหาสำคัญที่มักเกิดแบบจำลองการเรียนรู้ของเครื่อง

เอกสารอ้างอิง

Abdullah Karasa. (2022). Machine Learning for Financial Risk Management with Python. 1005 Gravenstein Highway North, Sebastopol, CA 95472: Published by O'Reilly Media, Inc.,

Alex J. Smola, & Bernhard Schölkopf. (2004). A tutorial on support vector regression. Statistics and Computing, (volume 14), pp. 199–222.
Retrieved from https://link.springer.com/article/10.1023/B:STCO.0000035301.49549.88

Bingyu Cao, Deping Zhan, & Xianbin Wu. (2009). Application of SVM in Financial Research. International Joint Conference on Computational Sciences and Optimization,
Retrieved from https://ieeexplore.ieee.org/document/5194004

Corinna Cortes, and Vladimir Vapnik. (1995). Support-Vector Networks. Machine Learning, (volume 20), pp. 273–297. Retrieved from https://link.springer.com/article/10.1007/BF00994018

Theodoros Evgeniou, & Massimiliano Pontil. (2001). Support Vector Machines: Theory and Applications. Lecture Notes in Computer Science,
Retrieved from https://link.springer.com/chapter/10.1007/3-540-44673-7_12

Valeriy V. Gavrishchaka, & Supriya B. Ganguli. (2003). Volatility forecasting from multiscale and high-dimensional market data. Neurocomputing, (Volume 55), pp. 285-305.
Retrieved from https://dblp.uni-trier.de/db/journals/ijon/ijon55.html#GavrishchakaG03

Valeriy V. Gavrishchaka, & Supriya Banerjee. (2006). Support vector machine as an efficient framework for stock market volatility forecasting. Computational Management Science (CMS), pp. 147–160.
Doi: 10.1007/s10287-005-0005-5