

สรุปงานวิจัยที่เกี่ยวข้อง

นำเสนอ อ. วันที่ 6 มิ.ย. 2568 >>> บ่าย 2 โมง

- List รายชื่อที่อ่านไปแล้ว
- 1. ชื่องานวิจัย: การทำเหมืองข้อมูลจากงบการเงินเพื่อทำนายราคาตลาดสำหรับหุ้นในกลุ่ม SET50
กรณีศึกษาตลาดหลักทรัพย์แห่งประเทศไทย
- 2. ผู้เขียน: ฐะปะนีย์ ตรีรัตนภรณ์ และ วชรวรรษ พรหมมา
- 3. ปัญหางานวิจัย: ขณะนี้ประเทศไทยกำลังเข้าสู่ยุคสังคมผู้สูงอายุ ดังนั้นจึงมีการวางแผนการเงินเพื่อเอาไว้ใช้จ่ายยามเกษียณ การออมเงินหรือการนำเงินมาลงทุนสามารถใช้เป็นเครื่องมือวางแผนในยามเกษียณได้ แต่ยังมีข้อสงสัยว่าถ้าลงทุนในตลาดหลักทรัพย์จะตัดสินใจได้จากอะไรแล้วจะรู้ได้อย่างไรว่าหุ้นไหนเหมาะสมกับการลงทุน จึงเป็นปัจจัยที่ส่งผลให้นำมาสู่งานวิจัย
- 4. บทคัดย่อ: ทำนายราคาหุ้นในกลุ่ม SET50 ใช้ข้อมูลงบการเงิน, 56-1 ปี 2559-2563 มาทำการวิเคราะห์
- 5. ตัวแปรต้น:ทุนจดทะเบียน จำนวนหุ้นสามัญ ส่วนของผู้ถือหุ้น รายได้ กำไรสุทธิ กำไรต่อส่วนของผู้ถือหุ้น และการจ่ายเงินปันผล → ตัวแปรทั้ง 7 มีผลต่อราคาในหุ้นสามัญกลุ่ม SET50
- 6. เทคนิค: วิเคราะห์ด้วยเทคนิคการทำเหมืองข้อมูล 3 รูปแบบ
 - a. Generalized Regression Model
 - b. Decision Tree Model
 - c. Support Vector Machine
- 7. เครื่องมือ: RapidMiner Go
- 8. พิจารณาผลการทำนายจาก: RSME (Root Mean Square Error) ถ้ามีค่าน้อยแสดงว่ามีความผิดพลาดน้อย
- 9. ผลลัพธ์ที่ได้:
 - a. เทคนิค Decision Tree มีค่า 28.19 แสดงผลเป็นแผนภูมิต้นไม้
 - b. เทคนิค Support Vector Machine มีค่า 35.05 แสดงผลเป็นค่าน้ำหนัก Weight
 - c. เทคนิค Generalized Regression Model มีค่า 40.78 แสดงผลเป็นค่าสัมประสิทธิ์
- 10. ทฤษฎีที่เกี่ยวข้อง

หลักการของ การทำเหมืองข้อมูลคือค้นหาแบบ Pattern หรือ ความสัมพันธ์ (Relation) ที่ซ่อนอยู่

เทคนิคแบ่งออกเป็น 3 ประเภท

 1. หาความสัมพันธ์ข้อมูล จากการวัดค่าความสัมพันธ์ 2 ชุดขึ้นไป และผลที่ได้เป็น

- a. ค่าสนับสนุน ผลลัพธ์เป็น % ที่ข้อมูลมีความสัมพันธ์กันถูกต้องตามกฎ
- b. ค่าความมั่นใจ จำนวนที่ถูกต้อง

ใช้วิธี Apriori Algorithm และ FP-Growth Algorithm

สรุปทฤษฎีแต่ละโมเดล

1. **SVM** เป็นอัลกอริทึมที่ใช้จำแนกข้อมูล โดยอาศัยชุดข้อมูลฝึก (Train) ที่มีคุณลักษณะ (Features) และป้ายกำกับ (Labels) แล้วหาค่าที่เหมาะสมที่สุดของไฮเปอร์เพลน เพื่อแยกข้อมูลแต่ละคลาสออกจากกัน

Support Vector Machine (SVM)

SVM เป็นอัลกอริทึมการเรียนรู้แบบมีผู้สอน (Supervised Learning) ที่ใช้สำหรับงานจำแนกประเภท (Classification)

- **ข้อมูลที่ใช้:**
อัลกอริทึมจะได้รับ **ชุดข้อมูลฝึก (Training Data)** ซึ่งประกอบด้วย
 - **Features (X):** คุณลักษณะของข้อมูล
 - **Labels (y):** ป้ายกำกับว่าข้อมูลแต่ละตัวอยู่ในคลาสใด
- **หลักการทำงาน:**
SVM จะหาเส้นหรือระนาบที่เรียกว่า **ไฮเปอร์เพลน (Hyperplane)** ซึ่งสามารถแยกข้อมูลแต่ละคลาสออกจากกันได้ดีที่สุด
โดยไฮเปอร์เพลนนี้จะพยายามอยู่ในตำแหน่งที่ ห่างจากข้อมูลของแต่ละคลาสมากที่สุด เพื่อให้สามารถจำแนกข้อมูลใหม่ได้แม่นยำ
- **ในกรณีสองมิติ:**
ไฮเปอร์เพลนจะเป็นเส้นที่แบ่งระนาบออกเป็นสองฝั่ง ข้อมูลแต่ละคลาสจะอยู่คนละด้านของเส้นนี้

**** กำหนด อัลกอริทึม (**Training Data**), **Features (X)**, **Labels (y)**

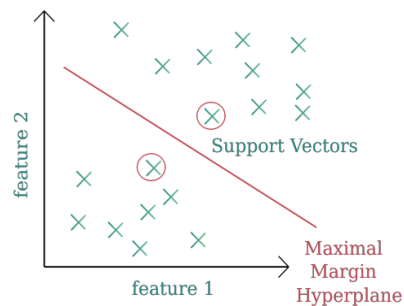
Parameter: กำหนดเงื่อนไขปรับมาตรฐาน พารามิเตอร์ → การปรับเงื่อนไขจะทำให้การจำแนกประเภทแบบไม่เชิงเส้น **Non-Linear** มีความแม่นยำมากขึ้น

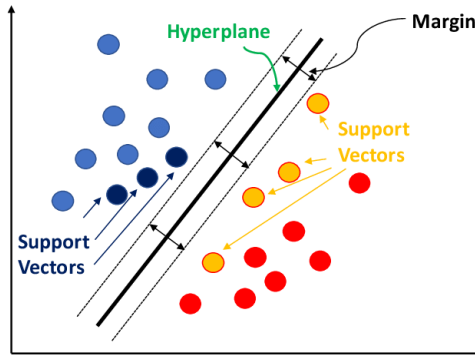
พารามิเตอร์

การตั้งค่าพารามิเตอร์ใน SVM ช่วยเพิ่มความแม่นยำในการจำแนกข้อมูลแบบไม่เชิงเส้น (Non-Linear) โดยพิจารณาจาก:

- **C (Regularization):**
 - C ต่ำ → มารจิ้นกว้าง ยอมให้ผิดบางจุด → ลด Overfitting
 - C สูง → เน้นความแม่นยำ → เสี่ยง Overfitting
- **Gamma (ใน RBF Kernel):**
 - Gamma สูง → เส้นแบ่งซับซ้อน → Overfitting
 - Gamma ต่ำ → เส้นแบ่งเรียบ → Underfitting
- **Kernel:** ใช้แปลงข้อมูลไปมิติสูงโดยไม่ต้องแปลงจริง ช่วยจำแนกข้อมูลซับซ้อน
- **Margin:** ระยะห่างระหว่างเส้นแบ่งกับข้อมูลที่ใกล้ที่สุด → มารจิ้นกว้างช่วยให้จำแนกแม่นยำและทั่วไปได้ดี

Support Vector Machines





🖼️ ภาพ: การจำแนกข้อมูลด้วย SVM

คำอธิบาย:

ภาพนี้แสดงหลักการทำงานของ Support Vector Machine (SVM) ในการจำแนกข้อมูลออกเป็น 2 กลุ่มโดยใช้ **เส้นแบ่ง (Hyperplane)**

- 📍 **จุดข้อมูล (Data Points):** จุดกลุ่มสีต่าง ๆ แสดงข้อมูลจากคนละคลาส
- 📏 **มาร์จิน (Margin):** ระยะห่างจากเส้นแบ่งกลางไปยัง **Support Vectors** ซึ่ง SVM พยายามให้กว้างที่สุด เพื่อให้จำแนกได้แม่นยำและทั่วไปได้ดี
- ⬤ **Support Vectors:** จุดข้อมูลที่อยู่ใกล้เส้นแบ่งที่สุด มีผลต่อการกำหนดตำแหน่งของเส้น
- 📊 **Kernel Trick (กรณีข้อมูลไม่เป็นเส้นตรง):** หากข้อมูลไม่สามารถแบ่งด้วยเส้นตรงได้ SVM จะใช้ Kernel แปลงข้อมูลไปยังมิติที่สูงขึ้น เพื่อให้สามารถแบ่งได้

🔗 คำอธิบายเพิ่มเติมเกี่ยวกับพารามิเตอร์

- **C (Regularization):** ปรับความเข้มงวดในการจำแนก
→ C ต่ำ = มาร์จินกว้าง / C สูง = จำแนกเป๊ะ แต่เสี่ยง Overfitting
- **Gamma (ใน RBF Kernel):**
→ Gamma สูง = เส้นแบ่งซับซ้อน / Gamma ต่ำ = เส้นแบ่งเรียบ