

Experimental Setting

ชุดข้อมูลทั้งหมด จะถูก normalize โดย standard scaler ซึ่งเป็น library ของ python จากนั้นการทดลองจะถูกแบ่งออกเป็นสองกลุ่มใหญ่ได้แก่

- 1) กลุ่มที่มี attribute เหมือนกับสมการของสัตว์แพทย์
['Weight', 'PCV', 'PCV\ndonor', 'Volume']
- 2) กลุ่มที่มี attribute มากกว่าสมการของสัตว์แพทย์
['Weight', 'PCV', 'PCV\ndonor', 'Volume', 'WBC', 'PLT\ng_____', 'PLATELETS', 'HGB', 'RBC', 'MCV', 'MCHC', 'MCH', 'SEGS', 'LYMPH', 'MONO', 'PROTEIN (REFRACT)', 'RDW']

จากนั้น ข้อมูลทั้งสองกลุ่ม จะถูกนำไปเทรนด้วยวิธีการดังต่อไปนี้ ซึ่งจะถูกแบ่งเป็นวิธีการที่ปรับค่าที่เหมาะสม และวิธีการที่ไม่ได้ปรับค่าของอัลกอริทึม

- 1) วิธีการที่ไม่ได้ปรับค่าอัลกอริทึม
 - 1) Linear Regression
 - 2) XG boost
- 2) วิธีการที่ปรับค่าอัลกอริทึมให้เหมาะสม
 - 1) Artificial Neural Network
 - 2) Support Vector Regression
 - 3) Ridge Regression

สำหรับอัลกอริทึมที่ต้องปรับค่าให้เหมาะสม ตามอัลกอริทึมประเภทที่ 2 ข้างต้น เราได้ใช้วิธี Double Leave One Out ในการหาค่าพารามิเตอร์ที่เหมาะสมให้กับแต่ละอัลกอริทึม โดย ANN จะได้รับการปรับค่าที่แตกต่างกันสำหรับข้อมูลที่มี attribute ไม่เท่ากัน ได้แก่ จำนวนเลเยอร์ตั้งแต่ 1-8 สำหรับข้อมูลประเภท 1 และจำนวนเลเยอร์ตั้งแต่ 1-20 สำหรับข้อมูลประเภท 2. และในกรณีของ SVR. จะได้รับการปรับค่าของ C โดยใช้ค่าตั้งแต่ 10^{-10} ถึง 10^9 โดยใช้ kernel เป็น linear และในกรณีของ Ridge Regression ใช้การปรับค่า alpha ตั้งแต่ 10^{-10} ถึง 10^9

ผลการทดลอง

กลุ่มที่ข้อมูลที่มี 1 attribute เหมือนกับสมการของสัตว์แพทย์

Formular	LR	XG Boost	ANN	SVR	Ridge
7.503460	4.776738	5.134629	21.176700	5.118797	4.735336

กลุ่มที่ข้อมูลที่มี 2 attribute มากกว่าสมการของสัตว์แพทย์

Formular	LR	XG Boost	ANN	SVR	Ridge
7.503460	7.9669	5.555596	20.345881	10.478592	7.449860

จากผลการทดลองพบว่า หากใส่จำนวนฟีเจอร์ที่มีจำนวนมากกว่าสัตว์แพทย์เข้าไป จะทำให้ Linear model ให้ผลลัพธ์ที่แย่ แต่หากใส่ฟีเจอร์เท่ากับที่สมการแพทย์ Linear model จะให้ผลลัพธ์ที่ดีกว่า แต่ในทางกลับกัน XG boost ให้ผลลัพธ์ที่ดีกว่าสัตว์แพทย์เสมอ และ ANN เป็นผลลัพธ์ที่แย่ที่สุดเสมอ