

Machine Learning Model

1.1 Model

model แนะนำยาต้านจุลชีพในระบบแบ่งออกเป็น 2 กลุ่ม คือ model แนะนำยาสำหรับเชื้อแบคทีเรียแกรมบวก (GP) และ model แนะนำยาสำหรับเชื้อแบคทีเรียแกรมลบ (GN) แต่ละกลุ่มประกอบด้วย model สำหรับยาต้านจุลชีพชนิดละ 1 model ผลลัพธ์ของ model จะบ่งชี้ว่า model แนะนำให้ใช้ยาต้านจุลชีพชนิดนั้น ๆ หรือไม่ กล่าวคือ เมื่อผลลัพธ์เป็น True แสดงว่า model แนะนำให้ใช้ยาต้านจุลชีพชนิดนี้ และเมื่อผลลัพธ์เป็น False แสดงว่า model ไม่แนะนำให้ใช้ยาต้านจุลชีพชนิดนี้

ทั้งนี้ จะทำการสร้าง model แนะนำยาต้านจุลชีพกลุ่มละ 11 model โดยเลือกจากยาต้านจุลชีพที่ถูกแนะนำมากที่สุด (พิจารณาจาก dataset ที่มี) model แนะนำยาต้านจุลชีพในระบบมีดังนี้ (เรียงตามจำนวนครั้งที่ถูกแนะนำ จากมากไปน้อย)

model แนะนำยาต้านจุลชีพ กลุ่มเชื้อแบคทีเรียแกรมบวก	
1	Amoxicillin/clavulanic acid
2	Trimethoprim/sulfamethoxazole
3	Cefovecin
4	Cefalexin
5	Clindamycin
6	Nitrofurantoin
7	Amikacin
8	Doxycycline
9	Marbofloxacin
10	Vancomycin
11	Enrofloxacin

model แนะนำยาต้านจุลชีพ กลุ่มเชื้อแบคทีเรียแกรมลบ	
1	Trimethoprim/sulfamethoxazole
2	Marbofloxacin
3	Cefovecin
4	Amikacin
5	Amoxicillin/clavulanic acid
6	Imipenem
7	Enrofloxacin
8	Cefalexin
9	Gentamicin
10	Nitrofurantoin
11	Doxycycline

1.2 Dataset

ชุดข้อมูลที่ใช้ในการสร้างและทดสอบ model แนะนำยาต้านจุลชีพได้มาจากรายงานผลการระบุชนิดจุลชีพ และการทดสอบความไวรับต่อยา (Results of Antimicrobial Susceptibility Testing : AST) ปี พ.ศ. 2559 – 2564 เป็นข้อมูลของสัตว์ป่วย 2 ประเภท คือ สุนัขและแมว รวมทั้งหมด 6,811 รายการ

ข้อมูลแบ่งออกเป็น 2 ชุด ตามชนิดของแบคทีเรีย ได้แก่ ชุดข้อมูลของแบคทีเรียแกรมบวก และ ชุดข้อมูลของแบคทีเรียแกรมลบ ชุดข้อมูลแกรมบวกมีข้อมูลจำนวน 3,385 รายการ และชุดข้อมูลแกรมลบมีข้อมูลจำนวน 3,426 รายการ ข้อมูลที่ใช้ในการสร้าง model มีดังนี้

- vitek_id คือ ชนิดของ card ที่ใช้ตรวจ
- species คือ ชนิดของสัตว์เลี้ยง
- submitted_sample คือ ตัวอย่างที่ส่งตรวจ

- bacteria_genus คือ สกุลแบคทีเรีย
- S/I/R_[ชื่อยาต้านจุลชีพ] คือ ผลตรวจความไวต่อยาต้านจุลชีพแต่ละชนิด
- ans_[ชื่อยาต้านจุลชีพ] คือ ยาต้านจุลชีพที่สัตวแพทย์แนะนำ

ชุดข้อมูลจะถูกแบ่งเป็น training dataset และ test dataset ในสัดส่วน 80:20 โดยทั้ง 2 ชุดข้อมูลมีอัตราส่วนของผลลัพธ์ (True และ False) เท่ากัน

1.3 Data Preprocessing

- สร้าง feature ใหม่ คือ bacteria_genus สร้างจาก bacteria_species โดยการตัดมาเฉพาะคำแรก
- สร้างข้อมูลใหม่ คือ other และแปลงข้อมูล submitted_sample ทุกตัวที่มีจำนวนน้อยกว่า 10 เป็น other
- เข้ารหัส feature ประเภท nominal ทั้งหมดแบบ one-hot encoding

1.4 Machine Learning Algorithm

algorithm ที่ใช้ใน model แนะนำยาต้านจุลชีพทั้ง 22 model คือ Support Vector Machine

1.5 Model Evaluation

ประเมินประสิทธิภาพของ model จาก test dataset ด้วยวิธี 10-Fold Cross-Validation โดยมีตัววัดคือ Accuracy (ความถูกต้อง), Precision (ความแม่นยำ), Recall (ความระลึก) และ F1-Score (ความถ่วงดุล)