รายงานความก้าวหน้า

การประมวลผลการจำแนกข้อมูลมัลแวร์จากข้อมูลทดสอบ โดยใช้อัลกอริทึม Fuzzy K-Nearest Neighbors (Fuzzy K-NN)

1. เครื่องมือและข้อมูลในการทดสอบ
   1. แบบจำลอง

แบบจำลองถูกพัฒนาโดยการฝึกด้วยชุดข้อมูลมัลแวร์ (Malware) จำนวน 1,000 ตัวอย่าง และข้อมูลเบนไญน์ (Benign) จำนวน 1,000 ตัวอย่าง เพื่อสร้างชุดข้อมูลต้นแบบ (Prototypes) ด้วยอัลกอริทึม **String Grammar Fuzzy C-Medians** ซึ่งส่งผลให้ได้ข้อมูลต้นแบบของข้อมูลมัลแวร์ (Malware) จำนวน 100 ตัวอย่าง และข้อมูลต้นแบบของข้อมูลเบนไญน์ (Benign) อีก 100 ตัวอย่าง

* 1. ภายหลังจากการฝึกแบบจำลอง ได้ดำเนินการทดสอบการตรวจสอบความถูกต้อง (Validation Test) ด้วยอัลกอริทึม **Fuzzy K-Nearest Neighbors (K = 9, m = 2)** ซึ่งได้ค่าความแม่นยำของแบบจำลองอยู่ที่ **89.20%**
  2. ข้อมูลทดสอบ เป็นข้อมูลที่ประกอบไปด้วยข้อมูล Malware จำนวน 838 ตัวอย่าง และข้อมูล Benign จำนวน 82 ตัวอย่าง รวมทั้งสิ้น 920 ตัวอย่าง

1. ผลการจำแนกข้อมูลมัลแวร์จากชุดข้อมูลทดสอบ ด้วยอัลกอริทึม Fuzzy K-NN

Confusion Matrix

รูปภาพประกอบด้วย ข้อความ, ภาพหน้าจอ, แผนภาพ, จำนวน

เนื้อหาที่สร้างโดย AI อาจไม่ถูกต้อง

ตารางประเมินผลการจำแนกข้อมูล (Evaluation Metrics)

(*กำหนดให้ข้อมูลมัลแวร์เป็นคลาสบวก — True Positive = TP และข้อมูลเบนไญน์เป็นคลาสลบ — True Negative = TN*)

|  |  |  |  |
| --- | --- | --- | --- |
| Metric | Malware (TP = 1) | Benign (TN = 0) | ความหมาย |
| Precision | 0.9471 | 0.1190 | ความแม่นยำในการทำนายว่าเป็น class นั้น |
| |  | | --- | |  |   Recall (Sensitivity) | 0.4702 | 0.7317 | ความสามารถในการจับ class นั้นให้ครบ (เช่น ตรวจ malware ได้ครบแค่ไหน) |
| F1-Score | 0.6284 | 0.2048 | สมดุลระหว่าง Precision และ Recall |
| Specificity | - | 0.7317 | ความสามารถในการระบุ benign ได้ถูกต้อง |
| False Positive Rate | - | 0.2683 | Benign ที่พลาด ทำนายผิดว่าเป็น malware |
| False Negative Rate | 0.5298 | - | Malware ที่พลาด ทำนายผิดว่าเป็น benign |
| Support | 838 | 82 | จำนวนตัวอย่างในแต่ละ class |
| Accuracy รวม (%) | 49.35% | | ความถูกต้องรวมทั้งหมด |