**รายงานผลการดำเนินงาน**

รายละเอียดการดำเนินงาน

1. **การเตรียมข้อมูล (Data Preprocessing)**  
   ดำเนินการแบ่งชุดข้อมูลออกเป็น 2 กลุ่ม ได้แก่ กลุ่มข้อมูล Benign จำนวน 1,082 สตริง และกลุ่มข้อมูล Malware จำนวน 10,838 สตริง โดยสุ่มคัดเลือกเพื่อแบ่งเป็นชุดข้อมูลสำหรับการฝึกสอน (Training Set) และชุดข้อมูลสำหรับการทดสอบ (Testing Set) ดังนี้:

* Benign: ข้อมูลฝึกสอนจำนวน 1,000 สตริง และข้อมูลทดสอบจำนวน 82 สตริง
* Malware: ข้อมูลฝึกสอนจำนวน 10,000 สตริง และข้อมูลทดสอบจำนวน 838 สตริง

เนื่องจากจำนวนข้อมูลฝึกสอนของกลุ่ม Malware มีมากกว่ากลุ่ม Benign อย่างมีนัยสำคัญ จึงได้ทำการสุ่มเลือกข้อมูลฝึกสอนจากกลุ่ม Malware มาใช้เพียง 1,000 สตริง เพื่อให้จำนวนข้อมูลทั้งสองกลุ่มสมดุลกันในการสร้างแบบจำลอง

1. **กระบวนการสร้างแบบจำลองต้นแบบ (Prototype Modeling)**  
   ข้อมูลฝึกสอนของแต่ละกลุ่มถูกนำมาผ่านกระบวนการ Cross Validation จำนวน 4 Fold จากนั้นดำเนินการวิเคราะห์เพื่อหาข้อมูลต้นแบบ (Prototype) จำนวน 100 สตริง โดยใช้เทคนิค *String Grammar Fuzzy C-Medians*
2. **การจำแนกข้อมูลโดยใช้ Fuzzy K-Nearest Neighbors (Fuzzy K-NN)**  
   เมื่อนำข้อมูลต้นแบบที่ได้จากแต่ละ Fold ของทั้งสองกลุ่มมารวมกัน (เช่น ข้อมูลต้นแบบของกลุ่ม Benign จาก Fold ที่ 1 รวมกับข้อมูลต้นแบบของกลุ่ม Malware จาก Fold ที่ 1) จึงทำการประเมินความแม่นยำของแบบจำลองในการจำแนกข้อมูลโดยใช้เทคนิค *Fuzzy K-Nearest Neighbors (Fuzzy K-NN)*  
   ผลการประเมินความแม่นยำของแบบจำลองในแต่ละ Fold แสดงดังตารางต่อไปนี้:

|  |  |  |  |
| --- | --- | --- | --- |
| Fold | K | M | ความแม่นยำ (%) |
| 1 | 7 | 2 | 88.20 |
| 2 | 19 | 2 | 88.80 |
| 3 | 9 | 2 | 89.20 |
| 4 | 9 | 2 | 87.80 |

1. **สรุปผลเบื้องต้น**  
   จากผลการประเมิน พบว่าแบบจำลองที่ได้จาก Fold ที่ 3 มีประสิทธิภาพสูงที่สุด โดยให้ค่าความแม่นยำในการจำแนกข้อมูลถึง 89.20%  
   ในขั้นตอนถัดไป จะนำแบบจำลองที่ได้จาก Fold ที่ 3 ไปประเมินผลกับชุดข้อมูลทดสอบที่ได้แยกไว้ในขั้นตอนที่ 1 ซึ่งในขณะนี้กำลังอยู่ระหว่างดำเนินการ