

# สารบัญ

|                                                                    | หน้า |
|--------------------------------------------------------------------|------|
| บทคัดย่อภาษาไทย .....                                              | I    |
| บทคัดย่อภาษาอังกฤษ .....                                           | II   |
| กิตติกรรมประกาศ .....                                              | III  |
| สารบัญ .....                                                       | IV   |
| สารบัญตาราง .....                                                  | VIII |
| สารบัญรูป .....                                                    | IX   |
| บทที่ 1 บทนำ .....                                                 | 1    |
| 1.1 ความเป็นมาและความสำคัญของปัญหา .....                           | 1    |
| 1.2 ความมุ่งหมายและวัตถุประสงค์ของการศึกษา .....                   | 2    |
| 1.3 สมมติฐานของการศึกษา .....                                      | 2    |
| 1.4 ทฤษฎีหรือแนวความคิดที่ใช้ในการวิจัย .....                      | 2    |
| 1.5 ขอบเขตการวิจัย .....                                           | 2    |
| 1.6 ขั้นตอนของการศึกษา .....                                       | 3    |
| 1.7 ข้อจำกัดของการศึกษา .....                                      | 3    |
| 1.8 คำจำกัดความที่ใช้ในการศึกษา .....                              | 3    |
| 1.9 เครื่องมือและอุปกรณ์ที่ใช้ในงานวิจัย .....                     | 4    |
| 1.10 โครงสร้างของวิทยานิพนธ์ .....                                 | 4    |
| บทที่ 2 งานวิจัยที่เกี่ยวข้อง .....                                | 5    |
| 2.1 เทคนิคต่างๆในการกรองอีเมลขยะ .....                             | 5    |
| 2.1.1 เทคนิคการกรองอีเมลขยะโดยใช้การกระตุ้นและตอบสนอง .....        | 5    |
| 2.1.2 เทคนิคการกรองอีเมลขยะโดยใช้การวิเคราะห์ส่วนหัวของอีเมล ..... | 5    |
| 2.1.3 เทคนิคการกรองอีเมลขยะโดยใช้ลายเซ็นดิจิทัล .....              | 6    |
| 2.1.4 เทคนิคการกรองอีเมลขยะโดยใช้บัญชีที่อยู่ .....                | 6    |
| 2.1.5 เทคนิคการกรองอีเมลขยะโดยใช้รายการคำสำคัญ ตอบสนอง.....        | 6    |
| 2.1.6 เทคนิคการกรองอีเมลขยะโดยการพิจารณาโอเพนรีเลย์.....           | 7    |
| 2.1.7 เทคนิคการกรองอีเมลขยะ โดยการพิจารณาโอเพนพร็อกซี่ .....       | 7    |

## สารบัญ (ต่อ)

หน้า

|                                                                         |    |
|-------------------------------------------------------------------------|----|
| 2.2 การกรอกรออีเมลล์ขยะโดยใช้เครื่องมือเรียนรู้และวิธีการทางสถิติ ..... | 7  |
| 2.2.1 การกรอกรออีเมลล์ขยะโดยใช้ตัวกรองเบย์เซียน .....                   | 7  |
| 2.2.2 การกรอกรออีเมลล์ขยะโดยใช้การแบ่งแยกของมาร์คอฟเวียน .....          | 8  |
| บทที่ 3 ความรู้พื้นฐานที่เกี่ยวข้องกับงานวิจัย .....                    | 9  |
| 3.1 ทฤษฎีที่เกี่ยวข้องกับอีเมลล์ .....                                  | 9  |
| 3.1.1 ความเป็นมาของอีเมลล์ .....                                        | 9  |
| 3.1.2 ลักษณะการทำงานของระบบรับ-ส่งอีเมลล์ .....                         | 9  |
| 3.1.3 สถาปัตยกรรมของระบบเมลล์ .....                                     | 10 |
| 3.1.4 โพรโทคอล .....                                                    | 12 |
| 3.1.4.1 SMTP .....                                                      | 12 |
| 3.1.4.2 POP .....                                                       | 13 |
| 3.1.4.3 IMAP .....                                                      | 14 |
| 3.1.5 คำนิยามของอีเมลล์ขยะ .....                                        | 14 |
| 3.2 องค์ประกอบของอีเมลล์ .....                                          | 15 |
| 3.2.1 ส่วนหัวเรื่อง .....                                               | 15 |
| 3.2.2 ส่วนเนื้อเรื่อง .....                                             | 16 |
| 3.3 ผลกระทบของอีเมลล์ขยะ .....                                          | 16 |
| 3.4 ความรู้พื้นฐานเกี่ยวกับทฤษฎีเบย์เซียน .....                         | 17 |
| 3.4.1 ทฤษฎีเบย์เซียน .....                                              | 18 |
| 3.4.2 การนำทฤษฎีเบย์เซียนมาใช้ในการกรอกรออีเมลล์ขยะ .....               | 19 |
| 3.5 ความรู้พื้นฐานเกี่ยวกับแขนงการตัดสินใจ.....                         | 20 |
| 3.6 ความรู้พื้นฐานเกี่ยวกับเจเนติกอัลกอริทึม .....                      | 21 |
| 3.6.1 พันธุศาสตร์ทางชีววิทยากับเจเนติกอัลกอริทึม .....                  | 22 |
| 3.6.1.1 พันธุศาสตร์ทางชีววิทยา .....                                    | 23 |
| 3.6.1.2 พันธุศาสตร์ทางเจเนติกอัลกอริทึม .....                           | 23 |
| 3.6.2 ขั้นตอนการทำงานของเจเนติกอัลกอริทึม .....                         | 24 |
| 3.6.2.1 การกำหนดรูปแบบโครโมโซม .....                                    | 24 |
| 3.6.2.2 ประชากร .....                                                   | 25 |

## สารบัญ (ต่อ)

|                                                                                                                     | หน้า |
|---------------------------------------------------------------------------------------------------------------------|------|
| 3.6.2.3 กำหนดฟังก์ชันความเหมาะสม .....                                                                              | 25   |
| 3.6.2.4 การวิเคราะห์ค่าความเหมาะสม .....                                                                            | 25   |
| 3.6.2.5 การคัดเลือก .....                                                                                           | 25   |
| 3.6.2.6 การครอสโอเวอร์.....                                                                                         | 28   |
| 3.6.2.7 การมิวเทชัน .....                                                                                           | 29   |
| 3.6.2.8 การสร้างประชากรรุ่นใหม่.....                                                                                | 30   |
| 3.6.2.9 การกำหนดค่าตัวแปรต่างๆ.....                                                                                 | 31   |
| <br>บทที่ 4 ตัวกรองอีเมลล์ขยะด้วยทฤษฎีเจเนติกอัลกอริทึม .....                                                       | 33   |
| 4.1 การเตรียมข้อมูล.....                                                                                            | 34   |
| 4.2 การสร้างฐานข้อมูลของค่า .....                                                                                   | 34   |
| 4.3 การประยุกต์ใช้เจเนติกอัลกอริทึม.....                                                                            | 35   |
| 4.3.1 การกำหนดรูปแบบของโครโมโซม.....                                                                                | 35   |
| 4.3.2 การประเมินค่าความเหมาะสม.....                                                                                 | 38   |
| 4.3.3 การคัดเลือกประชากร.....                                                                                       | 39   |
| 4.3.4 การครอสโอเวอร์และการมิวเทชัน.....                                                                             | 40   |
| 4.3.5 การทดสอบประสิทธิภาพของระบบ.....                                                                               | 41   |
| <br>บทที่ 5 ผลการทดลองและการวิเคราะห์ .....                                                                         | 44   |
| 5.1 ขั้นตอนการทดลองตัวกรองอีเมลล์ขยะโดยใช้เจเนติกอัลกอริทึม .....                                                   | 44   |
| 5.2 การปรับพารามิเตอร์ต่างๆ ในงานวิจัย .....                                                                        | 44   |
| 5.2.1 การเปรียบเทียบวิธีการคัดเลือกโครโมโซมระหว่างการคัดเลือกแบบ<br>วงล้อถ่วงน้ำหนักและการคัดเลือกแบบจัดลำดับ ..... | 45   |
| 5.2.2 การปรับเปอร์เซ็นต์การเลือกคู่โครโมโซมพ่อแม่ .....                                                             | 48   |
| 5.2.3 การปรับค่ามิวเทชันพารามิเตอร์.....                                                                            | 50   |
| 5.2.4 การปรับครอสโอเวอร์พารามิเตอร์.....                                                                            | 53   |
| 5.2.5 การปรับค่าอินเทอร์โรว์ .....                                                                                  | 54   |
| 5.3 ผลการทดลองการกรองอีเมลล์ขยะ โดยใช้เจเนติกอัลกอริทึม .....                                                       | 59   |
| 5.4 เปรียบเทียบตัวกรองอีเมลล์ขยะที่นำเสนอกับตัวกรองอีเมลล์ขยะซึ่งประยุกต์                                           |      |

## สารบัญ (ต่อ)

|                                               | หน้า |
|-----------------------------------------------|------|
| ใช้ทฤษฎีเบย์เซียน .....                       | 61   |
| บทที่ 6 สรุปผลการทดลองและข้อเสนอแนะ .....     | 64   |
| 6.1 สรุปและวิเคราะห์ผลการดำเนินงานวิจัย ..... | 64   |
| 6.2 ปัญหาที่พบในวิทยานิพนธ์.....              | 65   |
| 6.3 ข้อเสนอแนะและแนวทางในการพัฒนาต่อ .....    | 66   |
| เอกสารอ้างอิง .....                           | 67   |
| ภาคผนวก ก ผลงานวิจัยที่ได้รับการตีพิมพ์.....  | 69   |
| ประวัติผู้เขียน .....                         | 75   |

## สารบัญตาราง

| ตารางที่                                                                                                                                                     | หน้า |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------|------|
| 3.1 คำสำคัญที่ใช้ในกระบวนการเจเนติกอัลกอริทึม .....                                                                                                          | 24   |
| 3.2 ตัวอย่างการใช้แบบจำลองวงล้อถ่วงน้ำหนัก .....                                                                                                             | 27   |
| 4.1 แสดงตัวอย่างของค่าเฉพาะในแต่ละกลุ่ม.....                                                                                                                 | 35   |
| 4.2 แสดงการคิดค่าเฉลี่ยความน่าจะเป็นของค่าเฉพาะในอีเมล์ตัวอย่าง.....                                                                                         | 36   |
| 4.3 แสดงการแปลงค่าความน่าจะเป็นก่อนปรับและหลังปรับและค่าความน่าจะเป็นหลังปรับ<br>เมื่อแปลงเป็นเลขฐานสอง.....                                                 | 37   |
| 4.4 แสดงตัวอย่างโครโมโซมแม่แบบ และค่าความเหมาะสมของโครโมโซม.....                                                                                             | 39   |
| 5.1 แสดงการเปรียบเทียบค่า Accuracy, Recall และ Precision ระหว่างการคัดเลือกแบบวง<br>ล้อถ่วงน้ำหนัก (Roulette Wheel) และการคัดเลือกแบบจัดลำดับ (Ranking)..... | 46   |
| 5.2 แสดงค่า Accuracy, Recall และ Precision ที่การคัดเลือกคู่โครโมโซมพ่อแม่ที่<br>เปอร์เซ็นต์ต่างๆ.....                                                       | 49   |
| 5.3 แสดงค่า Accuracy, Recall และ Precision ของการมิวเตชันเมื่อผ่านไป 20, 40, 60,<br>80 Generation และเมื่อไม่ทำการมิวเตชันเลย (No).....                      | 50   |
| 5.4 แสดงค่า Accuracy, Recall และ Precision ของการมิวเตชันเมื่อปรับจำนวน<br>โครโมโซมที่เปอร์เซ็นต์ต่างๆ.....                                                  | 51   |
| 5.5 แสดงค่า Accuracy, Recall และ Precision ของการครอสโอเวอร์ทั้งสองแบบ.....                                                                                  | 53   |
| 5.6 แสดงค่า Accuracy, Recall และ Precision ของยีนเรโซว์ที่แตกต่างกัน.....                                                                                    | 55   |
| 5.7 แสดงค่า Accuracy, Recall และ Precision ของการปรับจำนวนคู่โครโมโซมพ่อแม่<br>และยีนเรโซว์ที่ค่าต่างๆ.....                                                  | 57   |
| 5.8 แสดงค่าเฉลี่ย Accuracy, Recall และ Precision ของการทดลองในแต่ละเซตและค่าเฉลี่ย<br>Accuracy, Recall และ Precision ของทุกๆ เซต.....                        | 59   |
| 5.9 แสดงค่า Accuracy, Recall และ Precision ของเซตข้อมูลทั้ง 5 เซต.....                                                                                       | 61   |

# สารบัญรูป

| รูปที่                                                                                                                                                       | หน้า |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------|------|
| 2.1 แสดงลำดับชั้นในกระบวนการกระตุ้น-ตอบสนองและชนิดของการตอบสนองที่<br>ต้องการ .....                                                                          | 5    |
| 3.1 แสดงลักษณะการทำงานของระบบอีเมล .....                                                                                                                     | 10   |
| 3.2 แสดงสถาปัตยกรรมในทีซีพี/ไอพี .....                                                                                                                       | 11   |
| 3.3 แสดงโปรโตคอลที่ใช้ในการรับส่งเมล .....                                                                                                                   | 12   |
| 3.4 แสดงตัวอย่างของอีเมลขยะ .....                                                                                                                            | 15   |
| 3.5 แสดงตัวอย่างของแขนงการตัดสินใจ.....                                                                                                                      | 21   |
| 3.6 แสดงการเข้ารหัสแบบไบนารี .....                                                                                                                           | 24   |
| 3.7 แสดงการเข้ารหัสแบบสลับลำดับ .....                                                                                                                        | 24   |
| 3.8 แสดงการเข้ารหัสแบบค่า .....                                                                                                                              | 24   |
| 3.9 แสดงการเข้ารหัสแบบทรี .....                                                                                                                              | 25   |
| 3.10 แสดงการเลือกโครโมโซมตามค่าความเหมาะสม.....                                                                                                              | 26   |
| 3.11 แสดงกระบวนการครอสโอเวอร์แบบไบนารี .....                                                                                                                 | 29   |
| 3.12 แสดงกระบวนการครอสโอเวอร์แบบตัวอักษร .....                                                                                                               | 29   |
| 3.13 การครอสโอเวอร์ของโครโมโซมแบบทรี .....                                                                                                                   | 29   |
| 3.14 แสดงกระบวนการไบนารีมิวเตชัน .....                                                                                                                       | 30   |
| 3.15 แสดงกระบวนการสร้างประชากรในรุ่นถัดไป .....                                                                                                              | 31   |
| 4.1 แสดงบล็อกไดอะแกรมของระบบ.....                                                                                                                            | 33   |
| 4.2 แสดงรูปแบบของโครโมโซม.....                                                                                                                               | 36   |
| 4.3 แสดงการเก็บผลการจำแนกของแต่ละแม่แบบ (Template) สำหรับใช้คำนวณหาค่า<br>ความเหมาะสม.....                                                                   | 38   |
| 4.4 แสดงแม่แบบของอีเมลทั้งหมดที่ได้จากกระบวนการเจเนติกอัลกอริทึมในรุ่นนี้.....                                                                               | 41   |
| 4.5 แสดงการจำแนกอีเมลโดยพิจารณาเปรียบเทียบจากค่าความน่าจะเป็นเฉลี่ยในชั้น.....                                                                               | 42   |
| 5.1 แสดงการเปรียบเทียบค่า Accuracy, Recall และ Precision ระหว่างการคัดเลือกแบบวง<br>ล้อถ่วงน้ำหนัก (Roulette Wheel) และการคัดเลือกแบบจัดลำดับ (Ranking)..... | 46   |
| 5.2 แสดงค่า Accuracy ของการคัดเลือกแบบวงล้อถ่วงน้ำหนัก และการคัดเลือก<br>แบบจัดลำดับ .....                                                                   | 47   |
| 5.3 แสดงค่า Recall ของการคัดเลือกแบบวงล้อถ่วงน้ำหนัก และการคัดเลือก<br>แบบจัดลำดับ .....                                                                     | 47   |

## สารบัญรูป(ต่อ)

| รูปที่                                                                                                                               | หน้า |
|--------------------------------------------------------------------------------------------------------------------------------------|------|
| 5.4 แสดงค่า Precision ของการการคัดเลือกแบบวงล้อถ่วงน้ำหนัก และการคัดเลือกแบบจัดลำดับ .....                                           | 48   |
| 5.5 แสดงค่า Accuracy, Recall และ Precision ที่การคัดเลือกคู่โครโมโซมพ่อแม่ที่เปอร์เซ็นต์ต่างๆ.....                                   | 48   |
| 5.6 แสดงค่า Accuracy, Recall และ Precision ของการมิวเตชันเมื่อผ่านไป 20, 40, 60, 80 Generation และเมื่อไม่ทำการมิวเตชันเลย (No)..... | 52   |
| 5.7 แสดงค่า Accuracy, Recall และ Precision ของการมิวเตชันเมื่อปรับจำนวนโครโมโซมที่เปอร์เซ็นต์ต่างๆ.....                              | 52   |
| 5.8 แสดงค่า Accuracy, Recall และ Precision ของการครอสโอเวอร์ทั้งสองแบบ.....                                                          | 54   |
| 5.9 แสดงค่า Accuracy, Recall และ Precision ของยีนเรโซว์ที่แตกต่างกัน.....                                                            | 56   |
| 5.10 แสดงค่า Accuracy ของเซตข้อมูลทั้ง 5 เซตเมื่อใช้พารามิเตอร์ที่ได้จากการทดลอง .....                                               | 60   |
| 5.11 แสดงค่า Precision ของเซตข้อมูลทั้ง 5 เซตเมื่อใช้พารามิเตอร์ที่ได้จากการทดลอง .....                                              | 60   |
| 5.12 แสดงค่า Precision ของเซตข้อมูลทั้ง 5 เซตเมื่อใช้พารามิเตอร์ที่ได้จากการทดลอง.....                                               | 61   |
| 5.13 แสดงการเปรียบเทียบค่า Accuracy ระหว่างเจเนติกอัลกอริทึมและเบย์เซียน.....                                                        | 62   |
| 5.14 แสดงการเปรียบเทียบค่า Recall ระหว่างเจเนติกอัลกอริทึมและเบย์เซียน.....                                                          | 62   |
| 5.15 แสดงการเปรียบเทียบค่า Precision ระหว่างเจเนติกอัลกอริทึมและเบย์เซียน.....                                                       | 62   |