

สารบัญ

บทที่

หน้า

1	บทนำ.....	1
	ความเป็นมาของปัญหา.....	1
	จุดมุ่งหมายของการวิจัย.....	5
	กรอบแนวคิดที่ใช้ในการวิจัย.....	5
	ขอบเขตของการวิจัย.....	6
	ข้อตกลงเบื้องต้น.....	8
	นิยามศัพท์เฉพาะ.....	8
	ประโยชน์ที่คาดว่าจะได้รับ.....	9
2	เอกสารและงานวิจัยที่เกี่ยวข้อง.....	10
	แนวคิดเกี่ยวกับการทำเหมืองข้อมูล.....	10
	เทคนิคเหมืองข้อมูล.....	13
	แนวคิดและทฤษฎีเกี่ยวกับตัวพยากรณ์แบบต้นไม้ตัดสินใจ.....	16
	แนวคิดและทฤษฎีเกี่ยวกับแบบจำลองแบบเบย์อย่างง่าย.....	20
	แนวคิดและทฤษฎีเกี่ยวกับตัวพยากรณ์แบบโครงข่ายประสาทเทียม.....	22
	แนวคิดและทฤษฎีเทคนิคการแบ่งชุดข้อมูลเรียนรู้และชุดข้อมูลทดสอบ.....	25
	แนวคิดและทฤษฎีตัววัดประสิทธิภาพตัวพยากรณ์.....	25
	งานวิจัยด้านอุบัติเหตุและการบาดเจ็บ.....	29
3	วิธีดำเนินการวิจัย.....	31
	ข้อมูลที่ใช้ในการวิจัย.....	31
	ขั้นตอนการดำเนินงานวิจัย.....	35
	เครื่องมือที่ใช้ในการทำวิจัย.....	36
	การพัฒนาตัวพยากรณ์.....	36
	ตัวพยากรณ์จำแนกระดับความรุนแรงเป็น 4 ระดับ.....	37

สารบัญ (ต่อ)

บทที่	หน้า
ตัวพยากรณ์จำแนกระดับความรุนแรงเป็น 3 ระดับ.....	41
ตัวพยากรณ์ที่ประยุกต์ใช้ตัวจำแนก 2 ตัวซ้อนกัน.....	45
4 ผลการวิจัย.....	47
ตัวพยากรณ์จำแนกระดับความรุนแรงเป็น 4 ระดับ.....	47
ตัวพยากรณ์จำแนกระดับความรุนแรงเป็น 3 ระดับ.....	54
ตัวพยากรณ์ประยุกต์ใช้ตัวจำแนก 2 ตัวซ้อนกัน.....	60
5 บทสรุป.....	66
สรุปผลการวิจัย.....	66
อภิปรายผล.....	68
ข้อเสนอแนะ.....	68
บรรณานุกรม.....	70
ภาคผนวก.....	76
ประวัติผู้วิจัย.....	113

สารบัญตาราง

ตาราง	หน้า
1 ตัวอย่างการแสดงความแม่นยำในการทำนายในรูปตาราง Confusion Matrix ..	26
2 แสดงลักษณะเฉพาะของข้อมูลที่คัดเลือกแล้ว 33 ตัว พร้อมรายละเอียด.....	32
3 แสดงผลของการทดสอบโคสแควร์.....	38
4 แสดงผลของการทดสอบโคสแควร์หลังจากเปลี่ยนแปลงลักษณะเฉพาะ.....	42
5 แสดงข้อมูลผู้บาดเจ็บจากอุบัติเหตุการจราจรทางถนน.....	47
6 แสดงค่าความแม่นยำของตัวพยากรณ์.....	48
7 แสดงค่าความแม่นยำของตัวพยากรณ์จากการแบ่งชุดข้อมูลแบบ 10 – Folds.....	50
8 แสดงค่าความแม่นยำของตัวพยากรณ์จากการแบ่งชุดข้อมูลแบบสัดส่วน.....	50
9 แสดงค่าความแม่นยำที่ได้จากการนำชุดข้อมูลทั้งหมดมาทดสอบตัวพยากรณ์ที่ได้จากการแบ่งชุดข้อมูลแบบสัดส่วน.....	51
10 แสดงค่าความแม่นยำของแบบจำลองบนชุดข้อมูลเรียนรู้ (958).....	55
11 แสดงค่าความแม่นยำของแบบจำลองบนชุดข้อมูลทั้งหมด (1,064).....	55
12 แสดงค่าความแม่นยำของตัวจำแนกแบบ 2 ระดับ.....	60
13 แสดงตาราง Confusion Matrix ของอัลกอริทึม RandomTree ของชุดข้อมูลทั้งหมด (9,624).....	61
14 แสดงตาราง Confusion Matrix ของอัลกอริทึม Naive Bayes ของชุดข้อมูลทั้งหมด (9,624).....	61
15 แสดงตาราง Confusion Matrix ของอัลกอริทึม C4.5 ของชุดข้อมูลทั้งหมด (9,624).....	62
16 แสดงตาราง Confusion Matrix ของอัลกอริทึม MLPs ของชุดข้อมูลทั้งหมด (9,624).....	62
17 แสดงตาราง Confusion Matrix ของการประยุกต์ใช้ตัวจำแนกของชุดข้อมูลทั้งหมด (9,624).....	64

สารบัญตาราง (ต่อ)

ตาราง	หน้า
18 แสดงค่าวัดประสิทธิภาพของการประยุกต์ใช้ตัวจำแนกของชุดข้อมูลทั้งหมด (9,624).....	64
19 แสดงตัวอย่างของลักษณะเฉพาะที่ถูกบันทึกลงคอมพิวเตอร์.....	77
20 แสดงตาราง Confusion Matrix ของอัลกอริทึม RandomTree ของการแบ่งชุด ข้อมูลแบบสุ่มข้อมูล 90%.....	92
21 แสดงค่าวัดประสิทธิภาพของอัลกอริทึม RandomTree ของการแบ่งชุดข้อมูลแบบ สุ่มข้อมูล 90%.....	92
22 แสดงตาราง Confusion Matrix ของอัลกอริทึม C4.5 ของการแบ่งชุดข้อมูลแบบ สุ่มข้อมูล 90%.....	93
23 แสดงค่าวัดประสิทธิภาพของอัลกอริทึม C4.5 ของการแบ่งชุดข้อมูลแบบสุ่มข้อมูล 90%.....	93
24 แสดงตาราง Confusion Matrix ของอัลกอริทึม Naive Baeys ของการแบ่งชุด ข้อมูลแบบสุ่มข้อมูล 90%.....	94
25 แสดงค่าวัดประสิทธิภาพของอัลกอริทึม Naive Baeys ของการแบ่งชุดข้อมูลแบบ สุ่มข้อมูล 90%.....	94
26 แสดงตาราง Confusion Matrix ของอัลกอริทึม RandomTree ของการแบ่งชุด ข้อมูลแบบ 10 – Folds.....	95
27 แสดงค่าวัดประสิทธิภาพของอัลกอริทึม RandomTree ของการแบ่งชุดข้อมูลแบบ 10 – Folds.....	95
28 แสดงตาราง Confusion Matrix ของอัลกอริทึม C4.5 ของการแบ่งชุดข้อมูลแบบ 10 – Folds.....	96
29 แสดงค่าวัดประสิทธิภาพของอัลกอริทึม C4.5 ของการแบ่งชุดข้อมูลแบบ 10 – Folds.....	96
30 แสดงตาราง Confusion Matrix ของอัลกอริทึม Naive Baeys ของการแบ่งชุด ข้อมูลแบบ 10 – Folds.....	97

สารบัญตาราง (ต่อ)

ตาราง	หน้า
31 แสดงค่าวัดประสิทธิภาพของอัลกอริทึม Naive Bayes ของการแบ่งชุดข้อมูลแบบ 10 – Folds.....	97
32 แสดงตาราง Confusion Matrix ของอัลกอริทึม RandomTree ของชุดข้อมูลเรียนรู้ (2,420).....	98
33 แสดงค่าวัดประสิทธิภาพของอัลกอริทึม RandomTree ของชุดข้อมูลเรียนรู้ (2,420).....	98
34 แสดงตาราง Confusion Matrix ของอัลกอริทึม RandomTree ของชุดข้อมูลทั้งหมด (3,064).....	99
35 แสดงค่าวัดประสิทธิภาพของอัลกอริทึม RandomTree ของชุดข้อมูลทั้งหมด (3,064).....	99
36 แสดงตาราง Confusion Matrix ของอัลกอริทึม RandomTree ของชุดข้อมูลทั้งหมด (9,624).....	100
37 แสดงตาราง Confusion Matrix ของอัลกอริทึม C4.5 ของชุดข้อมูลเรียนรู้ (2,420)	101
38 แสดงค่าวัดประสิทธิภาพของอัลกอริทึม C4.5 ของชุดข้อมูลเรียนรู้ (2,420).....	101
39 แสดงตาราง Confusion Matrix ของอัลกอริทึม C4.5 ของชุดข้อมูลทั้งหมด (3,064).....	102
40 แสดงค่าวัดประสิทธิภาพของอัลกอริทึม C4.5 ของชุดข้อมูลทั้งหมด (3,064).....	102
41 แสดงตาราง Confusion Matrix ของอัลกอริทึม C4.5 ของชุดข้อมูลทั้งหมด (9,624).....	103
42 แสดงตาราง Confusion Matrix ของอัลกอริทึม Naive Bayes ของชุดข้อมูลเรียนรู้ (2,420).....	104
43 แสดงค่าวัดประสิทธิภาพของอัลกอริทึม Naive Bayes ของชุดข้อมูลเรียนรู้ (2,420).....	104
44 แสดงตาราง Confusion Matrix ของอัลกอริทึม Naive Bayes ของชุดข้อมูลทั้งหมด (3,064).....	105

สารบัญตาราง (ต่อ)

ตาราง	หน้า
45 แสดงค่าวัดประสิทธิภาพของอัลกอริทึม Naive Bayes ของชุดข้อมูลทั้งหมด (3,064).....	105
46 แสดงตาราง Confusion Matrix ของอัลกอริทึม Naive Bayes ของชุดข้อมูลทั้งหมด (9,624).....	106
47 แสดงตาราง Confusion Matrix ของอัลกอริทึม C4.5 ของชุดข้อมูลเรียนรู้.....	107
48 แสดงค่าวัดประสิทธิภาพของอัลกอริทึม C4.5 ของชุดข้อมูลเรียนรู้.....	107
49 แสดงตาราง Confusion Matrix ของอัลกอริทึม C4.5 ของชุดข้อมูลทั้งหมด.....	108
50 แสดงค่าวัดประสิทธิภาพของอัลกอริทึม C4.5 ของชุดข้อมูลทั้งหมด.....	108
51 แสดงตาราง Confusion Matrix ของอัลกอริทึม Naive Bayes ของชุดข้อมูลเรียนรู้.....	109
52 แสดงค่าวัดประสิทธิภาพของอัลกอริทึม Naive Bayes ของชุดข้อมูลเรียนรู้.....	109
53 แสดงตาราง Confusion Matrix ของอัลกอริทึม Naive Bayes ของชุดข้อมูลทั้งหมด.....	110
54 แสดงค่าวัดประสิทธิภาพของอัลกอริทึม Naive Bayes ของชุดข้อมูลทั้งหมด.....	110
55 แสดงตาราง Confusion Matrix ของอัลกอริทึม MLPs ของชุดข้อมูลเรียนรู้.....	111
56 แสดงค่าวัดประสิทธิภาพของอัลกอริทึม MLPs ของชุดข้อมูลเรียนรู้.....	111
57 แสดงตาราง Confusion Matrix ของอัลกอริทึม MLPs ของชุดข้อมูลทั้งหมด.....	112
58 แสดงค่าวัดประสิทธิภาพของอัลกอริทึม MLPs ของชุดข้อมูลทั้งหมด.....	112

สารบัญภาพ

ภาพ	หน้า
1 ข้อมูลคดีอุบัติเหตุจากการจราจรทางบกปี พ.ศ. 2549 – 2553.....	1
2 ข้อมูลอุบัติเหตุจราจรปี พ.ศ. 2553.....	2
3 กระบวนการเหมืองข้อมูล.....	3
4 การวิเคราะห์ข้อมูล.....	6
5 โรงพยาบาลในเครือข่ายเฝ้าระวังการบาดเจ็บ 8 แห่ง.....	7
6 แสดงกระบวนการของ KDD.....	11
7 หลักการจำแนกข้อมูล.....	15
8 โครงสร้างต้นไม้ตัดสินใจ.....	17
9 โครงสร้างของชั้นเครือข่าย (Network Layer).....	23
10 จำนวนข้อมูลผู้บาดเจ็บจำแนกตามระดับความรุนแรง.....	34
11 จำนวนข้อมูลผู้บาดเจ็บจากการเลือกข้อมูลบางส่วน.....	39
12 โครงสร้างของตัวพยากรณ์โดยเลือกข้อมูลบางส่วน.....	40
13 จำนวนข้อมูลผู้บาดเจ็บที่ได้จากการเลือกข้อมูล.....	40
14 การแบ่งชุดข้อมูลวิธี 10 – Folds.....	43
15 การแบ่งชุดข้อมูลวิธี Modified 10 – Folds.....	44
16 โครงสร้างของตัวพยากรณ์จำแนกระดับความรุนแรงเป็น 3 ระดับ.....	45
17 โครงสร้างของการประยุกต์ใช้ตัวจำแนก 2 ตัวซ้อนกัน.....	46
18 ความถูกต้องในการจำแนกข้อมูลแยกตามระดับความรุนแรงสำหรับการแบ่งชุด ข้อมูลแบบสุ่มข้อมูล 90%.....	48
19 ความถูกต้องในการจำแนกข้อมูลแยกตามระดับความรุนแรงสำหรับการแบ่งชุด ข้อมูลแบบ 10 – Folds.....	49
20 ความถูกต้องในการจำแนกข้อมูลของชุดข้อมูลเรียนรู้จากการแบ่งชุดข้อมูลแบบ สัดส่วน.....	51
21 ความถูกต้องในการจำแนกข้อมูลของชุดข้อมูลทั้งหมดจากการแบ่งชุดข้อมูล แบบสัดส่วน.....	52

สารบัญภาพ (ต่อ)

ภาพ	หน้า
22 ความถูกต้องในการจำแนกข้อมูลของชุดข้อมูลทั้งหมดเมื่อนำมาทดสอบตัว พยากรณ์ที่ได้จากการแบ่งชุดข้อมูลแบบสัดส่วน.....	53
23 ความถูกต้องในการจำแนกข้อมูลของการแบ่งชุดข้อมูลแบบ Modified 10 – Folds ของชุดข้อมูลเรียนรู้.....	56
24 ความถูกต้องในการจำแนกข้อมูลของการแบ่งชุดข้อมูลแบบ Modified 10 – Folds ของชุดข้อมูลทั้งหมด.....	57
25 ตัวพยากรณ์แบบเปรียบเทียบเทคนิคเหมืองข้อมูลระหว่างต้นไม้ตัดสินใจและ แบบจำลองแบบเบย์อย่างง่าย.....	58
26 ตัวพยากรณ์จากเทคนิคโครงข่ายประสาทเทียมแบบ MLPs.....	59
27 การประยุกต์ใช้ตัวจำแนก 2 ตัวซ้อนกัน.....	63
28 ข้อมูลระดับความรุนแรงของผู้บาดเจ็บแยกตามโรงพยาบาลที่เก็บข้อมูล.....	79
29 ข้อมูลระดับความรุนแรงของผู้บาดเจ็บแยกตามช่วงเดือนที่เกิดอุบัติเหตุ.....	79
30 ข้อมูลระดับความรุนแรงของผู้บาดเจ็บแยกตามที่อยู่ปัจจุบันของผู้บาดเจ็บ.....	80
31 ข้อมูลระดับความรุนแรงของผู้บาดเจ็บแยกตามช่วงเวลาที่เกิดเหตุ.....	80
32 ข้อมูลระดับความรุนแรงของผู้บาดเจ็บแยกตามช่วงอายุของผู้บาดเจ็บ.....	81
33 ข้อมูลระดับความรุนแรงของผู้บาดเจ็บแยกตามเพศ.....	81
34 ข้อมูลระดับความรุนแรงของผู้บาดเจ็บแยกตามอาชีพของผู้บาดเจ็บ.....	82
35 ข้อมูลระดับความรุนแรงของผู้บาดเจ็บแยกตามประเภทของผู้บาดเจ็บ.....	82
36 ข้อมูลระดับความรุนแรงของผู้บาดเจ็บแยกตามยานพาหนะของผู้บาดเจ็บ.....	83
37 ข้อมูลระดับความรุนแรงของผู้บาดเจ็บแยกตามสาเหตุการบาดเจ็บ.....	83
38 ข้อมูลระดับความรุนแรงของผู้บาดเจ็บแยกตามลักษณะการบาดเจ็บ.....	84
39 ข้อมูลระดับความรุนแรงของผู้บาดเจ็บแยกตามการใช้แอลกอฮอล์.....	84
40 ข้อมูลระดับความรุนแรงของผู้บาดเจ็บแยกตามการใช้ยาเสพติด.....	85
41 ข้อมูลระดับความรุนแรงของผู้บาดเจ็บแยกตามการใช้เข็มฉีดยา.....	85
42 ข้อมูลระดับความรุนแรงของผู้บาดเจ็บแยกตามการใช้หมวกนิรภัย.....	86

สารบัญภาพ (ต่อ)

ภาพ	หน้า
43 ข้อมูลระดับความรุนแรงของผู้บาดเจ็บแยกตามผู้นำส่งโรงพยาบาล.....	86
44 ข้อมูลระดับความรุนแรงของผู้บาดเจ็บแยกตาม Severity of TBI.....	87
45 ข้อมูลระดับความรุนแรงของผู้บาดเจ็บแยกตามถนนที่เกิดเหตุ.....	87
46 ข้อมูลระดับความรุนแรงของผู้บาดเจ็บแยกตามทางลาดชัน.....	88
47 ข้อมูลระดับความรุนแรงของผู้บาดเจ็บแยกตามทางตรง.....	88
48 ข้อมูลระดับความรุนแรงของผู้บาดเจ็บแยกตามทางแยก.....	89
49 ข้อมูลระดับความรุนแรงของผู้บาดเจ็บแยกตามทางม้าลาย.....	89
50 ข้อมูลระดับความรุนแรงของผู้บาดเจ็บแยกตามทางโค้ง.....	90
51 ข้อมูลระดับความรุนแรงของผู้บาดเจ็บแยกตามทางมีสิ่งกีดขวาง.....	90
52 ข้อมูลระดับความรุนแรงของผู้บาดเจ็บแยกตามตำแหน่งการนั่ง.....	91
53 ข้อมูลระดับความรุนแรงของผู้บาดเจ็บแยกตามลักษณะการชน.....	91