

บทที่ 4

ผลการวิจัย

ในการวิจัยเรื่อง การประยุกต์ใช้เทคนิคเหมืองข้อมูลที่เหมาะสมสำหรับจำแนกความรุนแรงของการบาดเจ็บจากอุบัติเหตุการจราจรทางบก ผู้วิจัยได้แบ่งขั้นตอนในการวิจัยดังนี้

1. ตัวพยากรณ์จำแนกระดับความรุนแรงเป็น 4 ระดับ
2. ตัวพยากรณ์จำแนกระดับความรุนแรงเป็น 3 ระดับ
3. ตัวพยากรณ์ที่ประยุกต์ใช้ตัวจำแนก 2 ตัวซ้อนกัน

ข้อมูลผู้บาดเจ็บจากอุบัติเหตุการจราจรทางถนนที่กรองข้อมูลเรียบร้อยแล้ว มีทั้งหมด 9,624 ระเบียบ โดยมีจำนวนข้อมูลแยกตามระดับความรุนแรงของการบาดเจ็บดังตาราง 5 และใช้ลักษณะเฉพาะ 19 ตัว จากนั้นนำข้อมูลมาใช้ในการพัฒนาตัวพยากรณ์ที่เหมาะสม

ตาราง 5 แสดงข้อมูลผู้บาดเจ็บจากอุบัติเหตุการจราจรทางถนน

ระดับความรุนแรง	ปกติ	เสียชีวิต	พิการเข้าเกณฑ์	พิการไม่เข้าเกณฑ์	รวม
4 ระดับ	8,560	641	306	117	9,624
3 ระดับ	-	641	306	117	1,064

ตัวพยากรณ์จำแนกระดับความรุนแรงเป็น 4 ระดับ

ชุดข้อมูลที่นำมาใช้สร้างตัวพยากรณ์โดยแบ่งระดับความรุนแรงของการบาดเจ็บออกเป็น 4 ระดับ คือ ปกติ เสียชีวิต พิการเข้าเกณฑ์ และ พิการไม่เข้าเกณฑ์

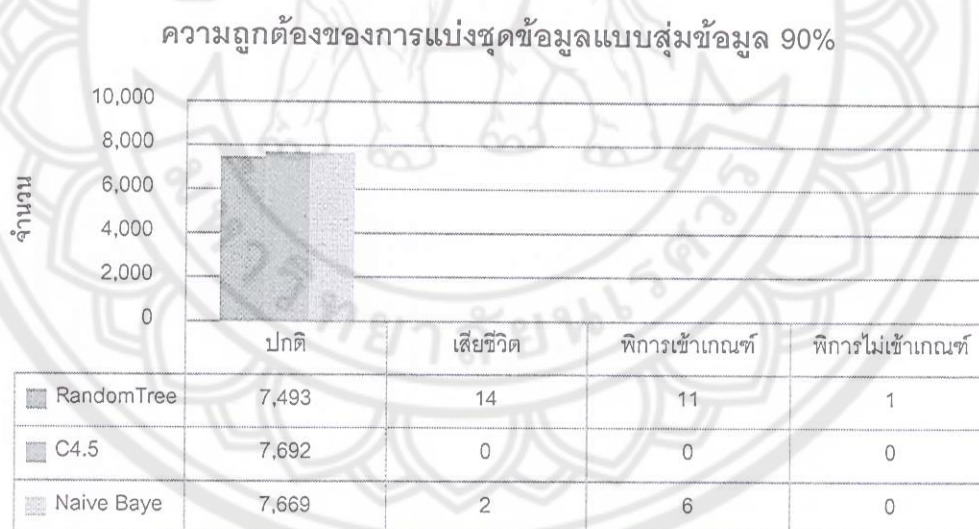
การสร้างตัวพยากรณ์เบื้องต้นด้วย Weka

จากการทดลองพัฒนาตัวพยากรณ์เบื้องต้นด้วย Weka (Waikato University, n.d.) ซึ่งใช้ข้อมูล 9,624 ระเบียบ ลักษณะเฉพาะ 9 ตัว และเลือกใช้การแบ่งชุดข้อมูล 2 แบบ คือ แบบสุ่มเลือกข้อมูล 90 % และ แบบ 10 – Folds โดยเลือกใช้เทคนิคการทำเหมืองข้อมูลแบบการจำแนกเพื่อจำแนกความรุนแรงของการบาดเจ็บ 3 เทคนิค คือ เทคนิคเหมืองข้อมูลแบบต้นไม้ตัดสินใจ (RandomTree และ C4.5) และ แบบจำลองแบบเบย์อย่างง่าย (Naive Bayes) ได้ค่าความแม่นยำของตัวพยากรณ์ดังแสดงในตาราง 6

ตาราง 6 แสดงค่าความแม่นยำของตัวพยากรณ์

การแบ่งชุดข้อมูล	เทคนิคเหมือนข้อมูล		
	RandomTree	C4.5	Naïve Bayes
สุ่มข้อมูล 90%	86.80%	88.89%	88.63%
10 – Folds	88.28%	88.93%	88.90%

จากค่าความแม่นยำของตัวพยากรณ์ในตาราง 6 จะแสดงให้เห็นว่าการแบ่งชุดข้อมูลแบบสุ่มเลือกข้อมูล 90 % และ แบบ 10 – Folds และเทคนิคการทำเหมือนข้อมูลทั้ง 3 แบบที่เลือกมาให้ค่าความแม่นยำใกล้เคียงกัน โดยแสดงความถูกต้องในการจำแนกข้อมูลโดยแยกตามระดับความรุนแรงสำหรับการแบ่งชุดข้อมูลแบบสุ่มข้อมูล 90% ดังภาพ 18 และสำหรับการแบ่งชุดข้อมูลแบบ 10 – Folds ดังภาพ 19 หรือตาราง Confusion Matrix และค่าวัดประสิทธิภาพของตัวพยากรณ์โดยรายละเอียดแสดงไว้ในตาราง 20 – 31 ในภาคผนวก ข

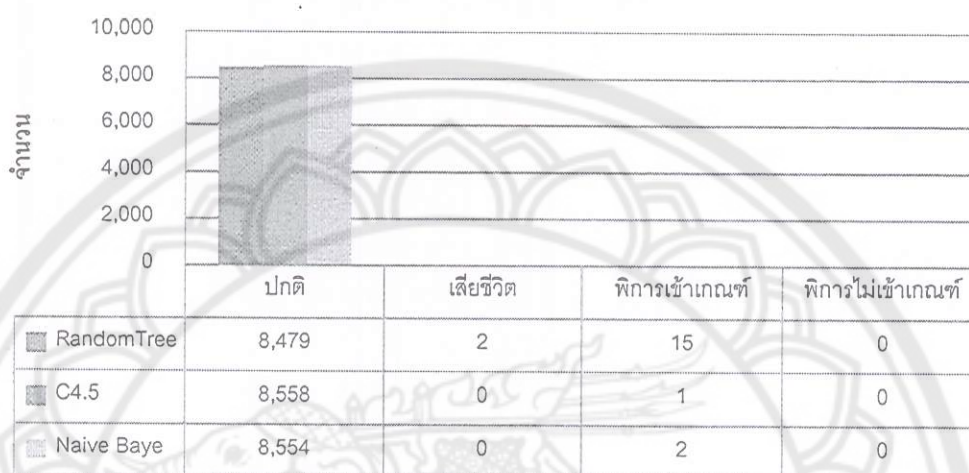


ภาพ 18 ความถูกต้องในการจำแนกข้อมูลแยกตามระดับความรุนแรงสำหรับการแบ่งชุดข้อมูลแบบสุ่มข้อมูล 90%

ภาพ 18 แสดงให้เห็นว่าตัวพยากรณ์ที่ได้จะจำแนกระดับความรุนแรงปกติได้ถูกต้องมากที่สุด ในขณะที่ระดับความรุนแรงเสียชีวิต พิกการเข้าเกณฑ์ และพิการไม่เข้าเกณฑ์ นั้นจำแนกได้

ถูกต้องน้อยมาก หรือ บางเทคนิคไม่สามารถจำแนกระดับความรุนแรง 3 ระดับนี้ได้เลย โดยในที่นี้ Random Tree ให้ตัวพยากรณ์ที่มีความแม่นยำดีที่สุด

ค่าความถูกต้องของการแบ่งชุดข้อมูลแบบ 10 - Folds



ภาพ 19 ความถูกต้องในการจำแนกข้อมูลแยกตามระดับความรุนแรงสำหรับการแบ่งชุดข้อมูลแบบ 10 - Folds

ภาพ 19 แสดงให้เห็นว่า ตัวพยากรณ์ที่ได้สามารถจำแนกระดับความรุนแรงปกติได้ถูกต้องมากที่สุด แต่ในระดับเสียชีวิต พิกัดเข้าเกณฑ์ และพิกัดไม่เข้าเกณฑ์ จำแนกได้ถูกต้องน้อยมาก หรือไม่ได้เลย แต่

เมื่อพิจารณาในภาพรวม RandomTree ให้ลักษณะการพยากรณ์ที่แม่นยำกว่า C4.5 และ Naive Bayes โดยการแบ่งชุดข้อมูลแบบ 10 - Folds ให้ค่าความแม่นยำในการพยากรณ์ที่ดีกว่าเล็กน้อย

การสร้างตัวพยากรณ์โดยเลือกข้อมูลบางส่วน

จากการทดลองการสร้างตัวพยากรณ์เบื้องต้นด้วย Weka จะเห็นว่าตัวพยากรณ์ที่ได้ไม่ค่อยมีประสิทธิภาพในการจำแนกข้อมูลระดับเสียชีวิต พิกัดเข้าเกณฑ์ และพิกัดไม่เข้าเกณฑ์ ดังนั้นผู้วิจัยจึงประยุกต์ใช้เทคนิคการเลือกข้อมูลบางส่วนเพื่อนำมาพัฒนาตัวพยากรณ์ โดยในที่นี้จะใช้ข้อมูล 3,064 ระเบียน จากการสุ่มเลือกข้อมูลของผู้บาดเจ็บที่ระดับความรุนแรงเป็นปกตินี้มา 2,000 ระเบียนจากทั้งหมด 8,560 ระเบียน รวมกับข้อมูลผู้บาดเจ็บที่มีระดับความรุนแรงเป็นเสียชีวิต พิกัดเข้าเกณฑ์ และ พิกัดไม่เข้าเกณฑ์ จำนวน 1,064 ระเบียน และเลือกใช้การแบ่งชุดข้อมูล 2 แบบ คือ การแบ่งชุดข้อมูลแบบ 10 - Folds และ การแบ่งชุดข้อมูลแบบสัดส่วน ซึ่งมีการ

แบ่งสัดส่วนของชุดข้อมูลดังแสดงในภาพ 13 แล้วใช้เทคนิคการทำเหมืองข้อมูลเพื่อจำแนกความรุนแรงของการบาดเจ็บ 3 เทคนิค คือ เทคนิคเหมืองข้อมูลแบบต้นไม้ตัดสินใจ (RandomTree และ C4.5) และ แบบจำลองแบบเบย์อย่างง่าย (Naive Bayes) ได้ค่าความแม่นยำของตัวพยากรณ์ดังแสดงในตาราง 7 สำหรับการแบ่งชุดข้อมูลแบบ 10 – Folds และตาราง 8 สำหรับการแบ่งชุดข้อมูลแบบสัดส่วน

ตาราง 7 แสดงค่าความแม่นยำของตัวพยากรณ์จากการแบ่งชุดข้อมูลแบบ 10 – Folds

ข้อมูล	เทคนิคเหมืองข้อมูล		
	RandomTree	C 4.5	Naive Bayes
10-fold	64.52%	67.66%	67.46%

จากตาราง 7 ค่าความแม่นยำของตัวพยากรณ์ที่ได้ค่อนข้างต่ำ โดยที่ในการแบ่งชุดข้อมูลแบบ 10 – Folds นี้ อัลกอริทึม C4.5 ให้ค่าความแม่นยำดีที่สุดจากอัลกอริทึมที่เลือกมาใช้สร้างตัวพยากรณ์

ตาราง 8 แสดงค่าความแม่นยำของตัวพยากรณ์จากการแบ่งชุดข้อมูลแบบสัดส่วน

ข้อมูล	เทคนิคเหมืองข้อมูล		
	RandomTree	C 4.5	Naive Bayes
ชุดข้อมูลเรียนรู้	71.16%	65.62%	65.00%
ชุดข้อมูลทดสอบ	70.19%	74.38%	77.48%
ชุดข้อมูลทั้งหมด	70.95%	67.46%	67.62%

จากตาราง 8 ค่าความแม่นยำของตัวพยากรณ์ที่ได้ดีขึ้นกว่าการแบ่งชุดข้อมูลแบบ 10 – Folds โดยในส่วนของชุดข้อมูลเรียนรู้ RandomTree ให้ค่าความแม่นยำที่ดีที่สุดคือ 71.16% รองลงมาคือ C4.5 และ Naive Bayes

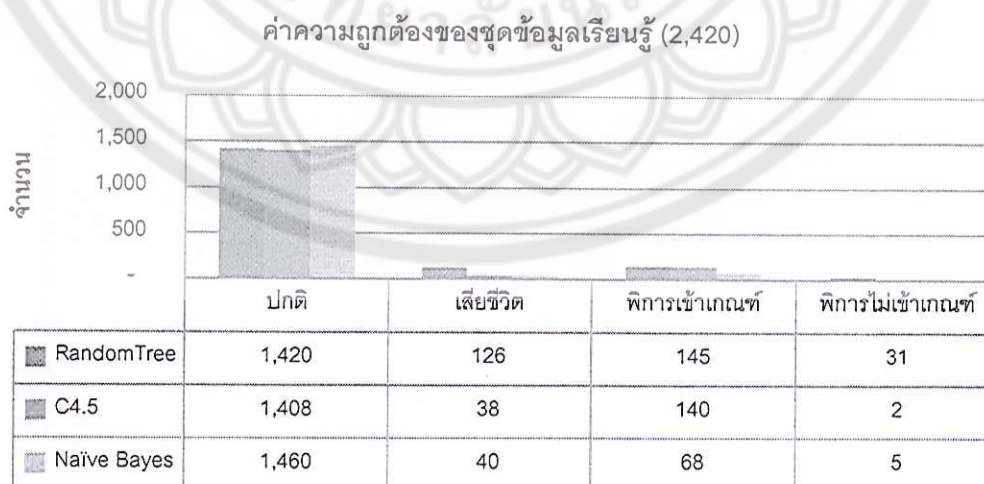
และเมื่อนำชุดข้อมูลทั้งหมด 9,624 ระเบียบมาทดสอบตัวพยากรณ์ที่ได้จากการแบ่งชุดข้อมูลแบบสัดส่วนแล้วได้ค่าความแม่นยำดังแสดงในตาราง 9

ตาราง 9 แสดงค่าความแม่นยำที่ได้จากการนำชุดข้อมูลทั้งหมดมาทดสอบตัวพยากรณ์ที่ได้จากการแบ่งชุดข้อมูลแบบสัดส่วน

ข้อมูล	เทคนิคเหมืองข้อมูล		
	RandomTree	C 4.5	Naive Bayes
ข้อมูลทั้งหมด (9,624)	79.40%	81.73%	83.13%

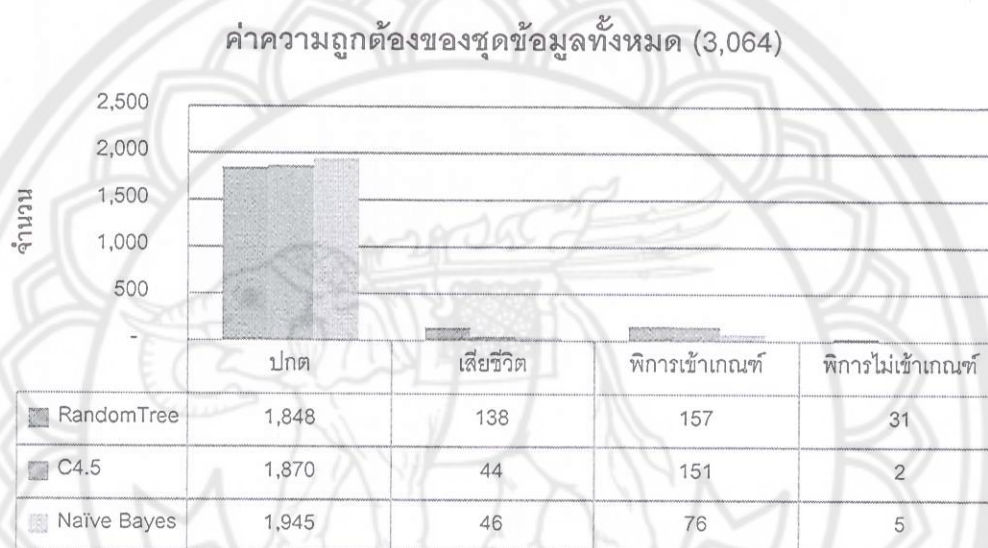
จากตาราง 9 แสดงให้เห็นว่าเมื่อนำข้อมูลทั้งหมด 9,624 ระเบียบ มาทดสอบตัวพยากรณ์ที่ได้จากการแบ่งชุดข้อมูลแบบสัดส่วนแล้วพบว่า Naive Bayes จะให้ค่าความแม่นยำดีที่สุดที่ 83.13% รองลงมาคือ C4.5 และ RandomTree ซึ่งต่างจากผลที่ได้จากตาราง 8

จากค่าความแม่นยำของตัวพยากรณ์ในตาราง 8 และ ตาราง 9 แสดงค่าความถูกต้องในการจำแนกข้อมูลโดยแยกตามระดับความรุนแรงของชุดข้อมูลเรียนรู้จากการแบ่งชุดข้อมูลแบบสัดส่วนได้ดังภาพ 20 ของชุดข้อมูลทั้งหมดจากการแบ่งชุดข้อมูลแบบสัดส่วนได้ดังภาพ 21 และเมื่อนำชุดข้อมูลทั้งหมดมาทดสอบตัวพยากรณ์ที่ได้จากการแบ่งชุดข้อมูลแบบสัดส่วนดังภาพ 22 โดยรายละเอียดของตาราง Confusion Matrix และค่าวัดประสิทธิภาพของตัวพยากรณ์แสดงไว้ในตาราง 32 – 46 ในภาคผนวก ข ซึ่งจะเห็นว่าเมื่อนำมาใช้กับข้อมูลทั้งหมด 9,624 ระเบียบ ความแม่นยำในการพยากรณ์ในระดับเสียชีวิต พิกการเข้าเกณฑ์ และพิกการไม่เข้าเกณฑ์ มีความแม่นยำที่ดีขึ้นกว่าแบบการสร้างตัวพยากรณ์เบื้องต้นด้วย Weka



ภาพ 20 ความถูกต้องในการจำแนกข้อมูลของชุดข้อมูลเรียนรู้จากการแบ่งชุดข้อมูลแบบสัดส่วน

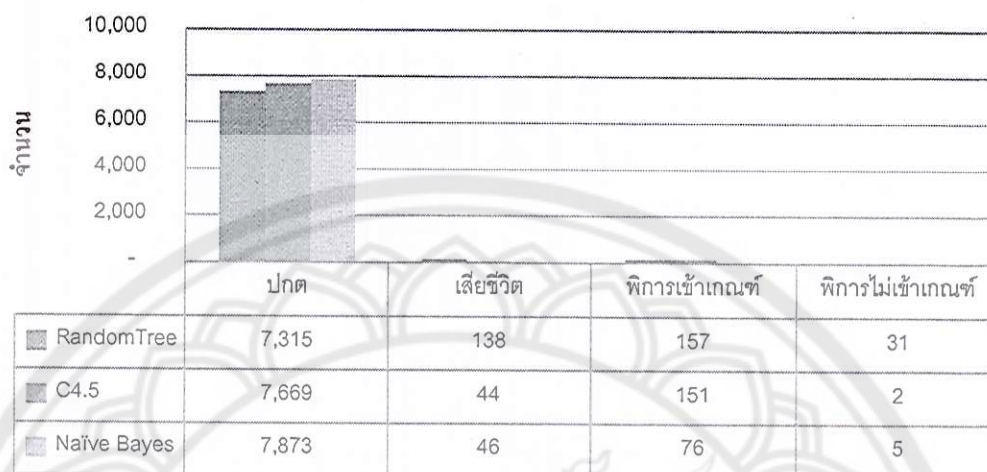
ภาพ 20 แสดงให้เห็นว่าตัวพยากรณ์ที่ได้จำแนกระดับความรุนแรงปกติได้ดีที่สุด และจำแนกระดับความรุนแรงเสียชีวิต พิกการเข้าเกณฑ์ และพิกการไม่เข้าเกณฑ์ ได้ถูกต้องมากขึ้นกว่าการแบ่งชุดข้อมูลแบบ 10 – Folds และ RandomTree จำแนกระดับความรุนแรงเสียชีวิต พิกการเข้าเกณฑ์ และพิกการไม่เข้าเกณฑ์ ได้ดีที่สุดจากอัลกอริทึมที่เลือกมาทั้งหมด รองลงมาคือ C4.5 และ Naive Bayes ตามลำดับ



ภาพ 21 ความถูกต้องในการจำแนกข้อมูลของชุดข้อมูลทั้งหมดจากการแบ่งชุดข้อมูลแบบสัดส่วน

ภาพ 21 แสดงให้เห็นว่าเมื่อนำข้อมูลทั้งหมดจากการเลือกข้อมูลบางส่วนมาทดสอบพบว่า RandomTree จำแนกข้อมูลในแต่ละระดับได้ถูกต้องมากที่สุด

ค่าความถูกต้องของชุดข้อมูลทั้งหมด (9,624)



ภาพ 22 ความถูกต้องในการจำแนกข้อมูลของชุดข้อมูลทั้งหมดเมื่อนำมาทดสอบตัวพยากรณ์ที่ได้จากการแบ่งชุดข้อมูลแบบสัดส่วน

ภาพ 22 แสดงให้เห็นเมื่อนำข้อมูลทั้งหมดมาทดสอบตัวพยากรณ์ที่ได้แล้ว พบว่าเมื่อนำไปเปรียบเทียบกับตัวพยากรณ์เบื้องต้นแล้ว การจำแนกข้อมูลในระดับความรุนแรงปกติมีความถูกต้องน้อยลง

เมื่อนำตาราง 6 และ ตาราง 9 มาเปรียบเทียบกันแล้วจะเห็นได้ว่าค่าความแม่นยำที่ได้จากการทดสอบตัวพยากรณ์ที่ได้จากการแบ่งชุดข้อมูลแบบสัดส่วนน้อยกว่าตัวพยากรณ์เบื้องต้น แต่เมื่อเปรียบเทียบจากความถูกต้องในการจำแนกในแต่ละระดับความรุนแรงด้วยแล้ว จะเห็นว่าตัวพยากรณ์ที่ได้จากการแบ่งชุดข้อมูลแบบสัดส่วนสามารถจำแนกระดับความรุนแรงเสียชีวิต พิกการเข้าเกณฑ์ และพิกการไม่เข้าเกณฑ์ได้ดีกว่าตัวพยากรณ์เบื้องต้น

จากค่าความถูกต้องของตัวพยากรณ์ที่ได้แสดงให้เห็นว่า อัลกอริทึม RandomTree สามารถจำแนกข้อมูลในทุกระดับได้ดี ทั้งนี้เพราะธรรมชาติของ RandomTree จะพยายามแตกกิ่งกระจายออกไปให้มากที่สุดเพื่อได้คำตอบในทุกๆ จุดของการเรียนรู้ถูกต้องมากที่สุด ดังนั้นเมื่อนำข้อมูลทั้งหมดมาทดสอบตัวพยากรณ์ทำให้ได้ค่าความแม่นยำน้อยที่สุดเมื่อเทียบกับอัลกอริทึมทั้งหมดที่ใช้ เพราะตัวพยากรณ์ที่ได้ค่อนข้างจะมีความเฉพาะเจาะจงเกินไป (Overfitting) ขณะที่อัลกอริทึม Naive Bayes จำแนกข้อมูลในระดับความรุนแรงปกติได้ถูกต้องมากที่สุดเนื่องจาก Naive Bayes ใช้หลักความน่าจะเป็นในการจำแนกซึ่งทำให้ตัวพยากรณ์ที่ได้สามารถพยากรณ์คำตอบในกลุ่มของข้อมูลที่เป็นกลุ่มใหญ่ได้ดี และอัลกอริทึม C4.5 จำแนกข้อมูลในระดับความ

รุนแรงปกติได้ถูกต้องมากเช่นกัน เพราะธรรมชาติของ C4.5 จะพยายามให้คำตอบในกลุ่มของข้อมูลที่เป็นกลุ่มใหญ่ให้ได้มากที่สุด แต่จะจำแนกระดับความรุนแรงเสียชีวิต พิกการเข้าเกณฑ์ และ พิกการไม่เข้าเกณฑ์ ได้ไม่ค่อยดี เนื่องจากตัวพยากรณ์ไม่มีความเฉพาะเจาะจง

จากผลการทดลองการสร้างตัวพยากรณ์ เมื่อเปรียบเทียบความแม่นยำในภาพรวมจะพบว่า อัลกอริทึมแบบ RandomTree ให้ค่าความแม่นยำดีที่สุด รองลงมาคือ อัลกอริทึมแบบ C4.5 และ Naive Bayes ตามลำดับ และจากการทดลองจะเห็นว่าค่าความแม่นยำค่อนข้างจะขึ้นอยู่กับความถูกต้องในการจำแนกข้อมูลระดับปกติซึ่งมีสัดส่วนของข้อมูลเป็นจำนวนมากเกินกว่า 85% ของข้อมูลทั้งหมด ส่งผลให้ตัวพยากรณ์ที่ได้ไม่ค่อยมีประสิทธิภาพเท่าที่ควรในการจำแนกข้อมูลในระดับ เสียชีวิต พิกการเข้าเกณฑ์ และ พิกการไม่เข้าเกณฑ์ ดังนั้นผู้วิจัยจึงทำการปรับรูปแบบการสร้างตัวพยากรณ์ใหม่โดยจะพิจารณาการจำแนกเป็น 3 ระดับแทน (ตัดระดับความรุนแรงแบบปกติออก)

ตัวพยากรณ์จำแนกระดับความรุนแรงเป็น 3 ระดับ

ชุดข้อมูลที่นำมาใช้สร้างตัวพยากรณ์นี้จะแบ่งระดับความรุนแรงของการบาดเจ็บเป็น 3 ระดับ คือ เสียชีวิต พิกการเข้าเกณฑ์ และ พิกการไม่เข้าเกณฑ์ โดยตัดข้อมูลที่มีระดับความรุนแรงของการบาดเจ็บเป็นปกติออก เพราะตัวพยากรณ์จะเอนเอียงไปทางระดับปกติมากเกินไป ทำให้มีข้อมูลทั้งหมด 1,064 ระเบียน ลักษณะเฉพาะ 15 ตัว และใช้การแบ่งชุดข้อมูล 3 แบบ คือ แบบสุ่มเลือกข้อมูล 90 % แบบ 10 – Folds และแบบ Modified 10 – Folds โดยเลือกใช้เทคนิคการทำเหมืองข้อมูลแบบแบบต้นไม้ตัดสินใจ (C4.5) แบบจำลองแบบเบย์ (Naive Bayes) และ โครงข่ายประสาทเทียม (MLPs)

เนื่องจากตัวพยากรณ์ที่จำแนกระดับความรุนแรง 4 ระดับ มีการใช้เทคนิคเหมืองข้อมูลแบบต้นไม้ตัดสินใจ 2 แบบ คือ อัลกอริทึมแบบ RandomTree และ C4.5 แต่ในการทำตัวพยากรณ์นี้เลือกอัลกอริทึม C4.5 มาใช้แบบเดียว เนื่องจากอัลกอริทึม RadomTree มีการแตกกิ่งที่มากเกินไปทำให้การจำแนกใกล้เคียงกับชุดข้อมูลเรียนรู้มากเกินไป ทำให้ต้นไม้ที่ได้นั้นไม่เหมาะสมกับข้อมูลทดสอบ หรือชุดข้อมูลอื่นๆ ที่ไม่ได้อยู่ในชุดเรียนรู้ จึงตัดอัลกอริทึมนี้ออก แล้วเพิ่มเทคนิคเหมืองข้อมูลแบบโครงข่ายประสาทเทียมเข้ามา

ค่าความแม่นยำเฉลี่ยของตัวพยากรณ์บนชุดข้อมูลเรียนรู้แสดงในตาราง 10 และค่าความแม่นยำเฉลี่ยของตัวพยากรณ์บนชุดข้อมูลทั้งหมดแสดงในตาราง 11

ตาราง 10 แสดงค่าความแม่นยำของแบบจำลองบนชุดข้อมูลเรียนรู้ (958)

ข้อมูล	เทคนิคเหมืองข้อมูล		
	C4.5	Naive Bayes	MLPs
Sampling 90%	74.08%	73.63%	81.07%
10-Folds	74.08%	73.74%	81.26%
Modified 10-Folds	74.21%	74.43%	81.39%

จากตาราง 10 แสดงให้เห็นว่าการแบ่งชุดข้อมูลแบบ Modified 10 – Folds ช่วยทำให้ค่าความแม่นยำของตัวพยากรณ์ที่ดีขึ้นในทุกๆ อัลกอริทึมที่เลือกมาใช้เมื่อเปรียบเทียบกับ การแบ่งชุดข้อมูลแบบอื่น และเทคนิคเหมืองข้อมูล MLPs ให้ค่าความแม่นยำที่ดีที่สุด

ตาราง 11 แสดงค่าความแม่นยำของแบบจำลองบนชุดข้อมูลทั้งหมด (1,064)

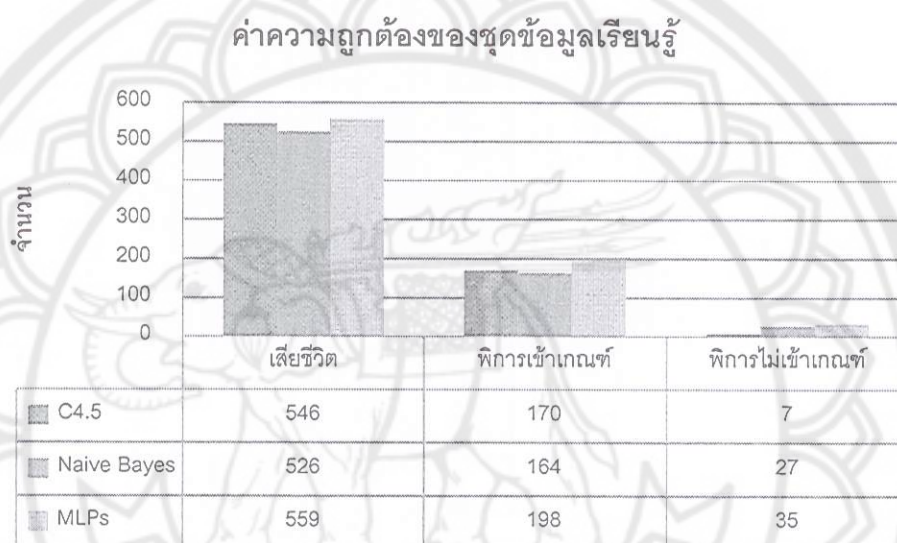
ข้อมูล	เทคนิคเหมืองข้อมูล		
	C4.5	Naive Bayes	MLPs
Sampling 90%	74.10%	73.59%	80.09%
10-Folds	74.02%	73.59%	80.08%
Modified 10-Folds	74.13%	73.67%	80.05%

จากตาราง 11 เมื่อนำชุดข้อมูลทั้งหมดมาทดสอบตัวพยากรณ์ที่ได้พบว่าการแบ่งชุดข้อมูลแบบ Modified 10 – Folds ให้ค่าความแม่นยำของตัวพยากรณ์ที่ดีขึ้น ยกเว้นตัวพยากรณ์จากเทคนิคเหมืองข้อมูล MLPs ที่ให้ค่าความแม่นยำลดลงเล็กน้อย และเมื่อเปรียบเทียบในแต่ละเทคนิคเหมืองข้อมูล MLPs ให้ตัวพยากรณ์ที่มีค่าความแม่นยำดีที่สุด

ในส่วนของชุดข้อมูลเรียนรู้ MLPs ให้ค่าความแม่นยำที่ดีที่สุดคือ 81.39% รองลงมาคือ Naive Bayes และ C4.5 แต่เมื่อนำไปใช้กับข้อมูลทั้งหมดคือ 1,064 ระเบียบแล้ว MLPs จะให้ค่าความแม่นยำที่ดีที่สุดที่ 80.05% รองลงมาคือ C4.5 และ Naive Bayes

จากค่าความแม่นยำของตัวพยากรณ์ในตาราง 10 และตาราง 11 แสดงให้เห็นว่าความถูกต้องในการจำแนกข้อมูลของการแบ่งชุดข้อมูลแบบ Modified 10 – Folds ช่วยทำให้ค่าความ

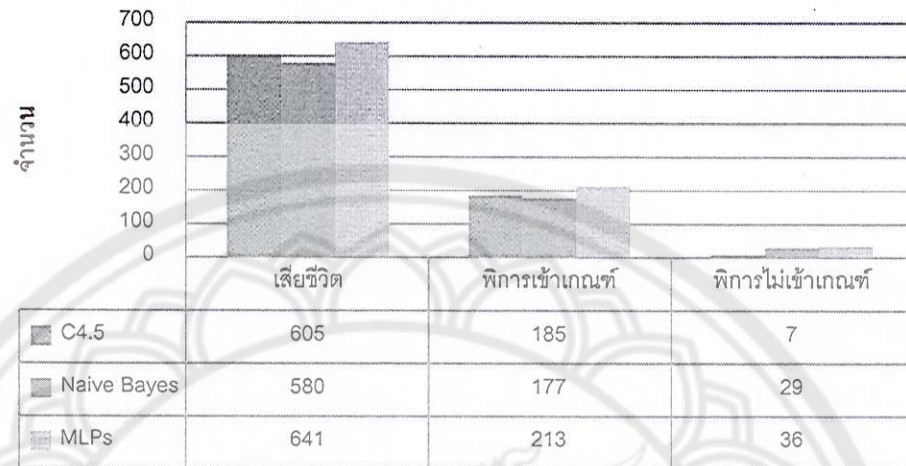
แม่นยำของตัวพยากรณ์ที่ได้ดีขึ้นในทุกๆ อัลกอริทึมที่เลือกใช้ การแสดงผลของการจำแนกโดยใช้การแบ่งชุดข้อมูลแบบ Modified 10 - Folds ของชุดข้อมูลเรียนรู้โดยแยกตามระดับความรุนแรงของชุดข้อมูลเรียนรู้ได้ดังภาพ 23 และของชุดข้อมูลทั้งหมดดังภาพ 24 โดยรายละเอียดของตาราง Confusion Matrix และค่าวัดประสิทธิภาพของตัวพยากรณ์ ของชุดข้อมูลเรียนรู้และชุดข้อมูลทดสอบแสดงดังตาราง 47 - 58 ในภาคผนวก ข



ภาพ 23 ความถูกต้องในการจำแนกข้อมูลของการแบ่งชุดข้อมูลแบบ Modified 10 - Folds ของชุดข้อมูลเรียนรู้

ภาพ 23 แสดงให้เห็นว่าตัวพยากรณ์ที่ได้จำแนกระดับความรุนแรงเสียชีวิต พิกการเข้าเกณฑ์ และพิกการไม่เข้าเกณฑ์ ได้ในระดับดี และ MLPs จำแนกได้ดีที่สุดรองลงมา Naive Bayes และ C4.5 ตามลำดับ

ค่าความถูกต้องของชุดข้อมูลทั้งหมด



ภาพ 24 ความถูกต้องในการจำแนกข้อมูลของการแบ่งชุดข้อมูลแบบ Modified 10 - Folds ของชุดข้อมูลทั้งหมด

ภาพ 24 แสดงให้เห็นว่าเมื่อนำข้อมูลทั้งหมดมาทดสอบตัวพยากรณ์ที่ได้แล้ว ค่าความถูกต้องที่ได้สอดคล้องกับค่าความถูกต้องที่ได้จากชุดข้อมูลเรียนรู้ และ MLPs จำแนกได้ดีที่สุด

ในการสร้างตัวพยากรณ์ของโครงข่ายประสาทเทียมอัลกอริทึม MLPs จะมีการตั้งค่าพารามิเตอร์ต่างๆ ก่อนที่จะสร้างแบบจำลอง โดยค่าของพารามิเตอร์ที่ได้ตั้งค่าก่อนสร้างแบบจำลองมีดังนี้

1. ค่า Learning rate ในการทดลองนี้ตั้งค่าไว้ที่ 0.2
2. ค่า Momentum ในการทดลองนี้ตั้งค่าไว้ที่ 0.1
3. จำนวนรอบในการ Train ในการทดลองนี้ตั้งค่าไว้ที่ 500 รอบ
4. จำนวนชั้นข้อมูลซ่อน (Hidden Node) ในการทดลองนี้ตั้งค่าไว้ที่ 9 Node หรือ ตาม

ค่าพื้นฐานของโปรแกรม Weka

และจากผลการทดลองการสร้างตัวพยากรณ์ เมื่อเปรียบเทียบความแม่นยำในภาพรวมจะพบว่า อัลกอริทึมแบบ MLPs ให้ค่าความแม่นยำดีที่สุด รองลงมาคือ อัลกอริทึมแบบ Naive Bayes และ C4.5 ตามลำดับ

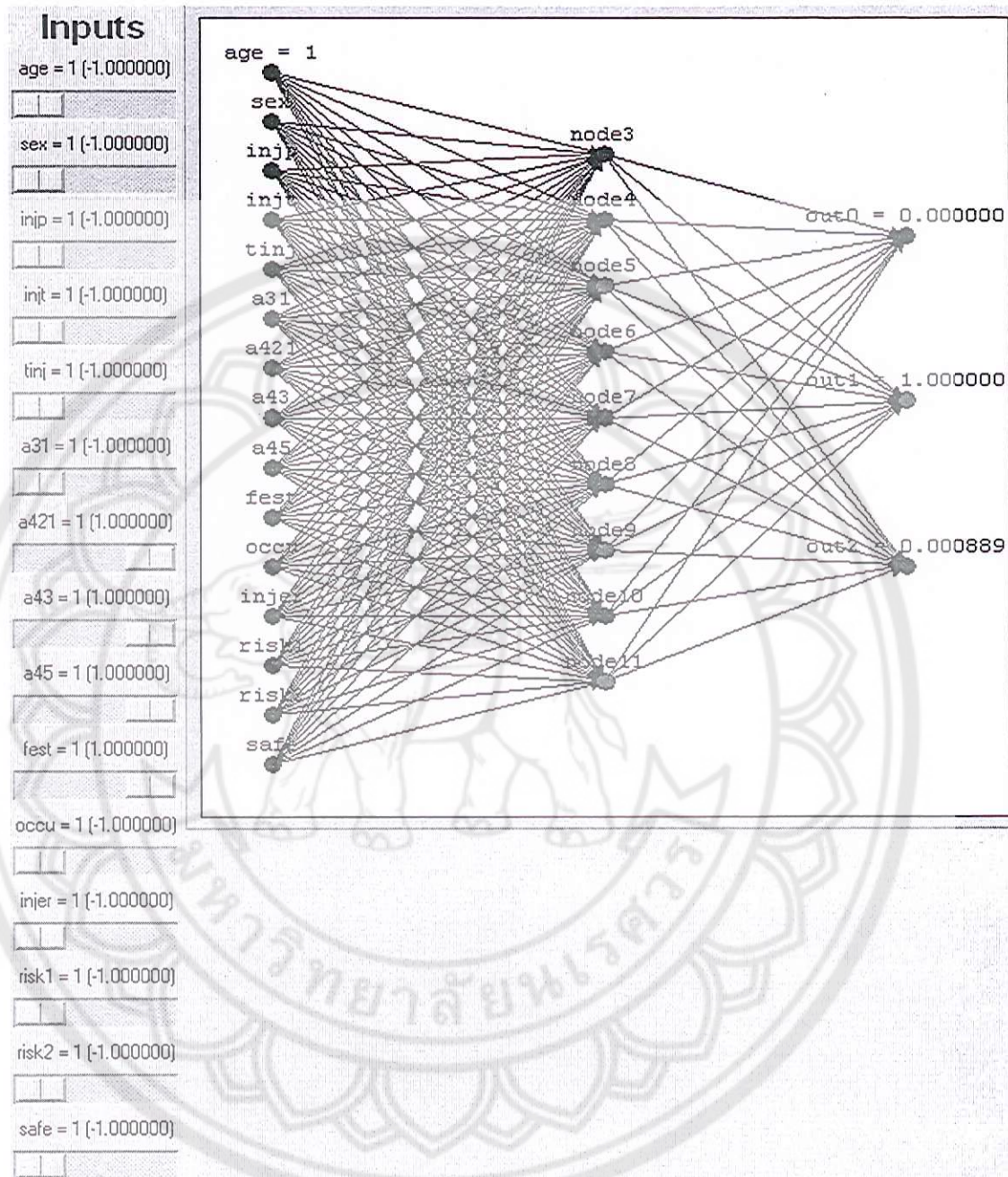
หลังจากการทดสอบแบบจำลองด้วยวิธีต่างๆ แล้วจึงได้ชุดข้อมูลและเทคนิคเหมืองข้อมูลที่มีประสิทธิภาพมากที่สุดในแต่ละอัลกอริทึมและแต่ละชุดข้อมูล แล้วนำชุดข้อมูลและเทคนิค

เหมือนข้อมูลที่ได้นั้น มาสร้างเป็นตัวพยากรณ์ ได้ตัวพยากรณ์แบบเปรียบเทียบเทคนิคเหมือน ข้อมูลโดยเปรียบเทียบจากเทคนิคต้นไม้ตัดสินใจและแบบจำลองแบบเบย์อย่างง่ายดังภาพ 25 ซึ่ง เป็นการพยากรณ์แบบ 4 ระดับ ตามการสร้างตัวพยากรณ์ และ ตัวพยากรณ์จากเทคนิคโครงข่าย ประสาทเทียมแบบ MLPs เป็นการพยากรณ์แบบ 3 ระดับ ดังภาพ 26

ข้อมูลการบาดเจ็บจากอุบัติเหตุ	ผลการทำนาย
อายุ : กรุณาเลือก เพศ : ☐ ชาย ☐ หญิง ผู้บาดเจ็บเป็น : กรุณาเลือก พาหนะของผู้บาดเจ็บ : กรุณาเลือก ลักษณะการบาดเจ็บ : กรุณาเลือก ข้อมูลที่เกิดเหตุ (ลักษณะเส้นทาง) ถนนที่เกิดเหตุ : กรุณาเลือก ทางตรง : ☐ ใช่ ☐ ไม่ใช่ ทางแยก : ☐ ใช่ ☐ ไม่ใช่ ทางโค้ง : ☐ ใช่ ☐ ไม่ใช่ ทำนายระดับความรุนแรง	Decision Tree J48 : RandomTree : Bayes Classifier Naive Bayes : Prediction Prediction : Delete Value

ภาพ 25 ตัวพยากรณ์แบบเปรียบเทียบเทคนิคเหมือนข้อมูลระหว่างต้นไม้ตัดสินใจและแบบจำลองแบบเบย์อย่างง่าย

ภาพ 25 เป็นการแสดงโปรแกรมพยากรณ์ 4 ระดับ ซึ่งพัฒนามาบนเว็บซึ่งพัฒนาโดยใช้ ภาษา PHP และใช้ AppServ เป็นเว็บเซิร์ฟเวอร์ (web server)



ภาพ 26 ตัวพยากรณ์จากเทคนิคโครงข่ายประสาทเทียมแบบ MLPs

ภาพ 26 เป็นโปรแกรมพยากรณ์ 3 ระดับ ซึ่งพัฒนามาจาก tkEden (EDEN - the Engine for DEfinitive Notations) (Empirical Modelling, 2013) ที่ให้ความยืดหยุ่น และโปรแกรมที่ได้มีการเรียนรู้แบบโต้ตอบ

ตัวพยากรณ์ที่ประยุกต์ใช้ตัวจำแนก 2 ตัวซ้อนกัน

ชุดข้อมูลก็นำมาใช้สร้างตัวพยากรณ์โดยแบ่งระดับความรุนแรงของการบาดเจ็บออกเป็น 4 ระดับ คือ ปกติ เสียชีวิต พิกัดเข้าเกณฑ์ และ พิกัดไม่เข้าเกณฑ์ ใช้ข้อมูล 9,624 ระเบียบ และ ลักษณะเฉพาะ 15 ตัว เนื่องจากข้อมูลของผู้บาดเจ็บที่นำมาใช้ในการสร้างตัวพยากรณ์นี้มีระดับความรุนแรงของการบาดเจ็บที่เป็นปกติจำนวนมาก ทำให้ตัวพยากรณ์จะเอนเอียงไปทางระดับปกติมากเกินไป ส่งผลให้ตัวพยากรณ์ที่ได้จำแนกระดับความรุนแรงเสียชีวิต พิกัดเข้าเกณฑ์ และ พิกัดไม่เข้าเกณฑ์ ได้น้อยหรืออาจจะไม่ได้เลย ทำให้ตัวพยากรณ์ที่ได้ไม่มีประสิทธิภาพ ดังนั้น ผู้วิจัยจึงได้ประยุกต์ใช้ตัวจำแนก 2 ตัว คือ ตัวจำแนกที่จำแนกระดับความรุนแรงเป็น 2 ระดับ และ ตัวจำแนกที่จำแนกระดับความรุนแรงเป็น 3 ระดับมาใช้ในการสร้างตัวพยากรณ์ โดยมีโครงสร้างของตัวพยากรณ์ดังแสดงในภาพ 17

การพัฒนาตัวจำแนกนี้จะจำแนกระดับความรุนแรงเป็น 2 ระดับ คือ ระดับปกติ และ ระดับพิกัดหรือเสียชีวิต ใช้ข้อมูลทั้งหมด 9,624 ระเบียบ และลักษณะเฉพาะ 15 ตัวมาสร้างตัวพยากรณ์ โดยทดลองใช้เทคนิคเหมืองข้อมูลแบบต้นไม้ตัดสินใจ อัลกอริทึม RandomTree และ C4.5 เทคนิคเหมืองข้อมูลแบบจำลองแบบเบย์ อัลกอริทึม Naive Bayes และโครงข่ายประสาทเทียม อัลกอริทึม MLPs ได้ค่าความแม่นยำดังตาราง 12

ตาราง 12 แสดงค่าความแม่นยำของตัวจำแนกแบบ 2 ระดับ

ข้อมูล	เทคนิคเหมืองข้อมูล			
	RandomTree	C 4.5	Naive Bayes	MLPs
Data Set (9,624)	94.86%	88.94%	88.07%	89.92%

ตาราง 12 แสดงให้เห็นว่าเมื่อใช้ข้อมูลทั้งหมด 9,624 ระเบียบทำเป็นชุดข้อมูลเรียนรู้เพื่อนำไปสร้างตัวพยากรณ์ อัลกอริทึม RandomTree ให้ค่าความแม่นยำที่ดีที่สุดคือ 94.86%

จากค่าความแม่นยำของตัวพยากรณ์ที่ได้ในตาราง 12 เมื่อแสดงในลักษณะตาราง Confusion Matrix จะได้ผลดังแสดงในตาราง 13 - 16

ตาราง 13 แสดงตาราง Confusion Matrix ของอัลกอริทึม RandomTree ของชุดข้อมูลทั้งหมด (9,624)

	ระดับ	ค่าพยากรณ์		
		ปกติ	เสียชีวิตหรือพิการ	รวม
ค่าจริง	ปกติ	8,531	29	8,560
	เสียชีวิตหรือพิการ	466	598	1,064
	รวม	8,997	627	9,624

ตาราง 13 แสดงให้เห็นว่าตัวจำแนกที่ได้สามารถจำแนกระดับความรุนแรงทั้งปกติ และเสียชีวิตหรือพิการได้ดี

ตาราง 14 แสดงตาราง Confusion Matrix ของอัลกอริทึม Naive Bayes ของชุดข้อมูลทั้งหมด (9,624)

	ระดับ	ค่าพยากรณ์		
		ปกติ	เสียชีวิตหรือพิการ	รวม
ค่าจริง	ปกติ	8,255	305	8,560
	เสียชีวิตหรือพิการ	843	221	1,064
	รวม	9,098	526	9,624

ตาราง 14 แสดงให้เห็นว่าตัวจำแนกที่ได้สามารถจำแนกระดับความรุนแรงทั้งปกติได้ดี แต่ระดับเสียชีวิตหรือพิการจำแนกได้ถูกต้องเพียงเล็กน้อย

ตาราง 15 แสดงตาราง Confusion Matrix ของอัลกอริทึม C4.5 ของชุดข้อมูลทั้งหมด (9,624)

	ระดับ	ค่าพยากรณ์		
		ปกติ	เสียชีวิตหรือพิการ	รวม
ค่าจริง	ปกติ	8,560	0	8,560
	เสียชีวิตหรือพิการ	1,064	0	1,064
	รวม	9,624	0	9,624

ตาราง 15 แสดงให้เห็นว่าตัวจำแนกที่ได้สามารถจำแนกระดับความรุนแรงทั้งปกติได้ดี แต่ระดับเสียชีวิตหรือพิการจำแนกไม่สามารถพยากรณ์ได้เลย

ตาราง 16 แสดงตาราง Confusion Matrix ของอัลกอริทึม MLPs ของชุดข้อมูลทั้งหมด (9,624)

	ระดับ	ค่าพยากรณ์		
		ปกติ	เสียชีวิตหรือพิการ	รวม
ค่าจริง	ปกติ	8,458	102	8,560
	เสียชีวิตหรือพิการ	868	196	1,064
	รวม	9,326	298	9,624

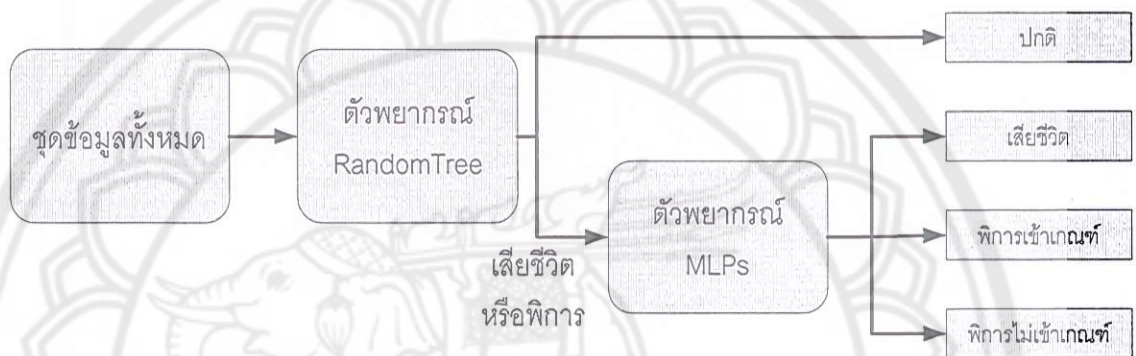
ตาราง 16 แสดงให้เห็นว่าตัวจำแนกที่ได้สามารถจำแนกระดับความรุนแรงทั้งปกติได้ดี แต่ระดับเสียชีวิตหรือพิการจำแนกได้ถูกต้องเพียงเล็กน้อย

จากค่าความแม่นยำ และ ตาราง Confusion Matrix ที่ได้จากการสร้างตัวจำแนก ผู้วิจัยเลือกใช้เทคนิคเหมืองข้อมูลแบบต้นไม้ตัดสินใจ อัลกอริทึม RandomTree ในการนำมาสร้างเป็นตัวจำแนกเนื่องจากให้ค่าความแม่นยำ และ Confusion Matrix ที่มีประสิทธิภาพมากที่สุดจากอัลกอริทึมทั้งหมดที่เลือกมาใช้

ตัวจำแนกที่จำแนกระดับความรุนแรงเป็น 3 ระดับ

การพัฒนาตัวจำแนกระดับความรุนแรงเป็น 3 ระดับ ในที่นี้ได้นำผลลัพธ์ที่ได้จากการพัฒนาตัวพยากรณ์จำแนกระดับความรุนแรง 3 ระดับ มาใช้ ซึ่งตัวพยากรณ์ที่ใช้คือตัวพยากรณ์ที่ได้จากใช้ข้อมูลจากการ Modified 10 – Flods และ เทคนิคเหมืองข้อมูลแบบโครงข่ายประสาทเทียม อัลกอริทึม MLPs จากการศึกษาที่ผ่านมา

จากนั้นนำตัวจำแนกที่ได้ทั้ง 2 ตัว มาประยุกต์ใช้ร่วมกันดังภาพ 27



ภาพ 27 การประยุกต์ใช้ตัวจำแนก 2 ตัวซ้อนกัน

จากภาพ 27 เมื่อนำข้อมูลทั้งหมด 9,624 ระเบียบมาทดสอบตัวพยากรณ์ได้ค่าความแม่นยำ 93.67% และให้ตาราง Confusion Matrix ดังตาราง 17 และ ค่าวัดประสิทธิภาพดังตาราง

ตาราง 17 แสดงตาราง Confusion Matrix ของการประยุกต์ใช้ตัวจำแนกของชุดข้อมูลทั้งหมด (9,624)

	ระดับ	ค่าพยากรณ์				รวม
		ปกติ	เสียชีวิต	พิการ เข้าเกณฑ์	พิการไม่ เข้าเกณฑ์	
ค่าจริง	ปกติ	8,531	18	8	3	8,560
	เสียชีวิต	299	308	27	7	641
	พิการเข้าเกณฑ์	107	43	155	1	306
	พิการไม่เข้าเกณฑ์	60	21	16	20	117
	รวม	8,997	390	206	31	9,624

ตาราง 18 แสดงค่าวัดประสิทธิภาพของการประยุกต์ใช้ตัวจำแนกของชุดข้อมูลทั้งหมด (9,624)

เกณฑ์วัดประสิทธิภาพ	ค่าระดับ				ค่าเฉลี่ย
	ปกติ	เสียชีวิต	พิการ เข้าเกณฑ์	พิการไม่ เข้าเกณฑ์	
ค่าความระลึก (Recall)	0.997	0.480	0.507	0.171	0.937
ค่าความแม่นยำ (Precision)	0.948	0.790	0.752	0.645	0.928
ค่าความเที่ยง (F-Measure)	0.972	0.597	0.606	0.270	0.927
ค่าผลบวกจริง (TP Rate)	0.997	0.480	0.507	0.171	0.937
ค่าผลบวกเท็จ (FP Rate)	0.438	0.009	0.005	0.001	0.390

จากตาราง 17 และ 18 แสดงให้เห็นว่าตัวพยากรณ์ที่ประยุกต์ใช้ตัวจำแนก 2 ตัวซ้อนกัน นั้น สามารถจำแนกระดับความรุนแรงในระดับปกติ เสียชีวิต พิการเข้าเกณฑ์ และพิการไม่เข้าเกณฑ์ได้ดีขึ้นเมื่อนำไปเปรียบเทียบกับผลที่ได้จากการสร้างตัวพยากรณ์เบื้องต้น และการสร้างตัวพยากรณ์โดยเลือกข้อมูลบางส่วน

โดยที่การพัฒนาตัวพยากรณ์นี้ใช้อัลกอริทึม RandomTree ในการจำแนกเป็น 2 ระดับ ก่อน แล้วใช้อัลกอริทึม MLPs ที่ได้จากการพัฒนาตัวพยากรณ์ที่จำแนกเป็น 3 ระดับ มาจำแนกต่อ เป็นตัวพยากรณ์แบบประยุต์ที่ให้ค่าความแม่นยำที่ 93.67%

