

บทที่ 3

วิธีดำเนินงานวิจัย

ในการวิจัยเรื่อง การประยุกต์ใช้เทคนิคเหมืองข้อมูลที่เหมาะสมสำหรับจำแนกความรุนแรงของการบาดเจ็บจากอุบัติเหตุการจราจรทางบก ข้อมูลที่ใช้ในงานวิจัยคือข้อมูลการเกิดอุบัติเหตุการจราจรทางถนน ที่เก็บรวบรวมไว้โดยการเข้ารับบริการในโรงพยาบาลในเครือข่ายการเฝ้าระวังการบาดเจ็บ จำนวน 8 แห่ง ทั่วประเทศไทย ซึ่งผู้วิจัยได้กำหนดการดำเนินงานวิจัยดังนี้

1. ข้อมูลที่ใช้ในการวิจัย
2. ขั้นตอนการดำเนินงานวิจัย
3. เครื่องมือที่ใช้ในการทำวิจัย
4. การพัฒนาตัวพยากรณ์

ข้อมูลที่ใช้ในการวิจัย

ข้อมูลที่นำมาใช้ในการศึกษาวิจัย คือข้อมูลผู้บาดเจ็บจากอุบัติเหตุการจราจรทางถนนที่รวบรวมไว้โดยการเข้ารับบริการในโรงพยาบาลในเครือข่ายการเฝ้าระวังการบาดเจ็บ จำนวน 8 แห่ง ทั่วประเทศไทย ได้แก่ โรงพยาบาลเลิดสิน กรุงเทพมหานคร โรงพยาบาลพระนั่งเกล้า จังหวัดนนทบุรี โรงพยาบาลศูนย์ราชบุรี จังหวัดราชบุรี โรงพยาบาลศูนย์ชลบุรี จังหวัดชลบุรี โรงพยาบาลพุทธชินราช จังหวัดพิษณุโลก โรงพยาบาลมหาราชนครราชสีมา จังหวัดนครราชสีมา โรงพยาบาลขอนแก่น จังหวัดขอนแก่น และ โรงพยาบาลมหาราชนครศรีธรรมราช จังหวัดนครศรีธรรมราช ทำการสำรวจไว้ตั้งแต่วันที่ 26 ธันวาคม พ.ศ. 2548 – 25 มิถุนายน พ.ศ. 2549 โดยใช้แบบสัมภาษณ์ 2 ชุด สำหรับกรณีบาดเจ็บไม่รุนแรง (ไม่ต้องนอนพักในโรงพยาบาล : non admit) และ กรณีบาดเจ็บรุนแรง (ต้องนอนพักในโรงพยาบาล : admit) และมีการติดตามผลของการบาดเจ็บนั้นต่อเพื่อสังเกตการเปลี่ยนแปลงที่เกิดขึ้นหลังจากออกจากโรงพยาบาล ในแบบสัมภาษณ์ที่มีการเก็บบันทึกข้อมูลที่สามารถนำมาใช้ศึกษาถึงสาเหตุที่ทำให้เกิดอุบัติเหตุ ได้แก่ ข้อมูลทั่วไปของผู้บาดเจ็บ (อายุ เพศ อาชีพ) วันและช่วงเวลาที่เกิดอุบัติเหตุ สาเหตุของการบาดเจ็บ ยานพาหนะ ลักษณะการบาดเจ็บ พฤติกรรมเสี่ยงของผู้บาดเจ็บ (การใส่ยา แอลกอฮอล์ การใช้เข็มขัดนิรภัย (สำหรับรถยนต์) และ การใช้หมวกนิรภัย (สำหรับรถจักรยานยนต์) ลักษณะ

ของถนนที่เกิดเหตุ สถานที่เกิดเหตุ ตำแหน่งการนั่งของผู้บาดเจ็บ ลักษณะการชน เป็นต้น ข้อมูลที่เก็บรวบรวมมามีจำนวนทั้งสิ้น 24,435 รายการ

จากแบบสัมภาษณ์ข้างต้นมีการแปลงข้อมูลเพื่อนำไปบันทึกเป็นข้อมูลที่สามารถนำไปวิเคราะห์โดยคอมพิวเตอร์ได้ โดยจัดเก็บข้อมูลทั้งหมด 287 ลักษณะเฉพาะ ดังตัวอย่าง ลักษณะเฉพาะแสดงในตาราง 19 ในภาคผนวก ก

ในงานวิจัยนี้ ผู้วิจัยจะใช้ข้อมูลเฉพาะกรณีบาดเจ็บรุนแรง (ต้องนอนพักในโรงพยาบาล : admit) ซึ่งข้อมูลมีจำนวนทั้งสิ้น 9,738 ระเบียน แล้วตัดข้อมูลที่เป็นค่าว่างและจัดเก็บไม่สมบูรณ์ เหลือข้อมูลทั้งหมด 9,624 ระเบียน และได้คัดเลือกลักษณะเฉพาะให้เหลือเพียง 33 ตัว โดยคัดเลือกตามปัจจัยที่เกี่ยวข้องกับการเกิดอุบัติเหตุ 4 ปัจจัย คือ คน ยานพาหนะ ถนน และสิ่งแวดล้อม ได้ลักษณะเฉพาะทั้ง 33 ตัว ดังแสดงในตาราง 2

ตาราง 2 แสดงลักษณะเฉพาะของข้อมูลที่คัดเลือกแล้ว 33 ตัว พร้อมรายละเอียด

ลักษณะเฉพาะ	ลักษณะของข้อมูล
เลขแบบสอบถาม	เลขจำนวนเต็ม เป็นลำดับที่
โรงพยาบาล	{ชลบุรี, มหาราชนครราชสีมา, ขอนแก่น, มหาราชนครศรีธรรมราช, พุทธชินราช, พระนั่งเกล้า, ราชบุรี, เลิดลิน}
สิทธิการรักษา	{ชำระเงินเอง, เบิกจากหน่วยงานต้นสังกัด, โครงการประกันสุขภาพ, บัตรสุขภาพนักเรียน, ผู้ประกันภัยตามพรบ. ประกันสังคม, กองทุนทดแทน, ประกันสุขภาพเอกชน, ผู้มีสิทธิลดหย่อนตามระเบียบสาธารณสุข, ทหารผ่านศึก, ภิกษุ/สามเณร, ผสส./อสส., กำนัน/ผู้ใหญ่บ้าน, ผู้มีรายได้น้อย, ผู้สูงอายุ, ผู้พิการ, ผู้ป่วยโรคเอดส์, ผู้มีสิทธิพรบ.คุ้มครองผู้ประสบภัยจากรถ, อื่นๆ}
ช่วงเดือนที่เกิดเหตุ	{มกราคม, กุมภาพันธ์, มีนาคม, เมษายน, พฤษภาคม, 1-25 มิถุนายน, 26-31 ธันวาคม}
เทศกาลปีใหม่	{29 ธ.ค. 48, 30 ธ.ค. 48, 31 ธ.ค. 48, 1 ม.ค. 49, 2 ม.ค. 49, 3 ม.ค. 49, 4 ม.ค. 49}
เทศกาลสงกรานต์	{7 เม.ย. 49, 8 เม.ย. 49, 9 เม.ย. 49, 10 เม.ย. 49, 11 เม.ย. 49, 12 เม.ย. 49, 13 เม.ย. 49, 14 เม.ย. 49, 15 เม.ย. 49, 16 เม.ย. 49}
ที่อยู่ปัจจุบัน	{ในจังหวัด, นอกจังหวัด, ไม่ทราบ}

ตาราง 2 (ต่อ)

ลักษณะเฉพาะ	ลักษณะของข้อมูล
ช่วงเวลาที่เกิดเหตุ	{6.01 – 9.00, 9.01 – 12.00, 12.01 – 15.00, 15.01 – 18.00, 18.01 – 21.00, 21.01 – 0.00, 0.01 – 3.00, 3.01 – 6.00}
ช่วงอายุ	{0-4, 5-14, 15-29, 30-44, 45-59, 60-69, 70-79, มากกว่า 80}
เพศ	{ชาย, หญิง}
อาชีพ	{ข้าราชการ, ตำรวจ/ทหาร, พนักงานรัฐวิสาหกิจ, พนักงานบริษัท, ผู้ใช้แรงงาน, ค้าขาย, เกษตรกรรม, นักเรียน/นักศึกษา, อื่นๆ}
ผู้บาดเจ็บ	{คนเดินเท้า, ผู้ขับขี่, ผู้โดยสาร, ไม่ทราบ}
ยานพาหนะ	{ไม่ใช้, รถจักรยานยนต์ ¹ , รถยนต์ ² , รถบรรทุก ³ , อื่นๆ}
สาเหตุการบาดเจ็บ	{พาหนะล้ม/คว่ำ/ตก/จม, ถูกชนหรือชนกับ, อื่นๆ}
ลักษณะการบาดเจ็บ	{blunt, penetrating, blunt and penetrating, อื่นๆ}
แอลกอฮอล์	{ไม่ใช้, ใช้, ไม่ทราบ}
ยา	{ไม่ใช้, ใช้, ไม่ทราบ}
เข็มขัดนิรภัย	{ไม่ใช้, ใช้, ไม่ทราบ}
หมวกนิรภัย	{ไม่ใช้, ใช้, ไม่ทราบ}
ผู้นำส่งคือ	{หน่วยบริการฉุกเฉินทางการแพทย์, เจ้าหน้าที่มูลนิธิ, ตำรวจ, อื่นๆ}
bp1	ความดันโลหิต แสดงเป็นเลขจำนวนเต็ม
bp2	ความดันโลหิต แสดงเป็นเลขจำนวนเต็ม
Severity of TBI	{Severe TBI, Moderate TBI, Mild TBI}
ถนนที่เกิดเหตุ	{ถนน อบต./หมู่บ้าน, ถนนในเมือง/เทศบาล, ถนนกรมทางหลวง, ถนนกรมทางหลวงชนบท, ถนนอื่นๆ, ไม่ทราบ}
ทางลาดชัน	{ไม่ใช้, ใช่}
ทางตรง	{ไม่ใช้, ใช่}
ทางแยก	{ไม่ใช้, ใช่}
ทางม้าลาย	{ไม่ใช้, ใช่}
ทางโค้ง	{ไม่ใช้, ใช่}
ทางมีสิ่งกีดขวาง	{ไม่ใช้, ใช่}

ตาราง 2 (ต่อ)

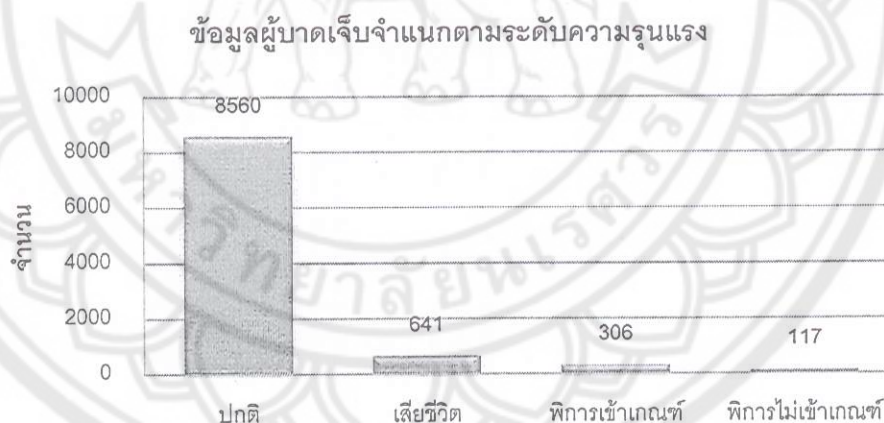
ลักษณะเฉพาะ	ลักษณะของข้อมูล
ตำแหน่งการนั่งของ ผู้บาดเจ็บ	{ผู้ขับขี่, ผู้โดยสารตอนหน้า(ในรถ), ผู้โดยสารตอนหลัง(ในรถ), ผู้โดยสาร นั่งภายนอกตัวรถ}
ลักษณะการชน	{ชนด้านหน้า, ชนด้านข้าง, ชนด้านหลัง, อื่นๆ}
ระดับความรุนแรง	{ปกติ, เสียชีวิต, พิกัดเข้าเกณฑ์, พิกัดไม่เข้าเกณฑ์}

หมายเหตุ: ¹ ใช้แทนรถจักรยาน รถสามล้อ รถจักรยานยนต์ และ รถสามล้อเครื่อง

² ใช้แทนรถเก๋ง รถปิกอัพ และ รถตู้

³ ใช้แทนรถบรรทุก รถพ่วง รถโดยสารสองแถว และ รถโดยสารบัส

โดยเมื่อจำแนกตามระดับความรุนแรง 4 ระดับ คือ ปกติ เสียชีวิต พิกัดเข้าเกณฑ์ และ พิกัดไม่เข้าเกณฑ์ ดังแสดงในภาพ 10



ภาพ 10 จำนวนข้อมูลผู้บาดเจ็บจำแนกตามระดับความรุนแรง

จากภาพ 10 จะเห็นได้ว่า เมื่อจำแนกข้อมูลผู้บาดเจ็บตามระดับความรุนแรงแล้ว ส่วนมากหลังจากทำการรักษาการบาดเจ็บที่เกิดจากอุบัติเหตุการจราจรทางบกแล้ว ข้อมูลระดับความรุนแรงของการบาดเจ็บภายหลังระดับปกติจะมีจำนวนข้อมูลมากที่สุด คือ 8,560 ระเบียบรองลงมาคือระดับเสียชีวิต 641 ระเบียบ ระดับพิกัดเข้าเกณฑ์ 306 ระเบียบ และ พิกัดไม่เข้าเกณฑ์ 117 ระเบียบ

และจากภาพ 28 - 53 ในภาคผนวก ก แสดงให้เห็นว่า โรงพยาบาลมหาวิทยาลัยนครราชสีมา จังหวัดนครราชสีมา โรงพยาบาลขอนแก่น จังหวัดขอนแก่น และ โรงพยาบาลมหาวิทยาลัยนครราชสีมา จังหวัดนครราชสีมา มีจำนวนผู้บาดเจ็บมากที่สุดทั้งนี้อาจเป็นเพราะ 3 จังหวัดนี้เป็นจังหวัดขนาดใหญ่และมีประชากรหนาแน่น ช่วงเดือนที่เกิดอุบัติเหตุมากที่สุดคือ เดือนมีนาคม เมษายน และ มกราคม ตามลำดับ ผู้บาดเจ็บส่วนมากแล้วจะเป็นคนที่อยู่ในจังหวัดที่เกิดอุบัติเหตุ และมักเกิดอุบัติเหตุในช่วงเวลา 21.01 น. ถึง 0.00 น. ผู้บาดเจ็บส่วนใหญ่จะอยู่ในช่วงอายุ 15 - 29 ปี และ เพศชายเกิดการบาดเจ็บมากกว่าเพศหญิง ส่วนใหญ่ผู้บาดเจ็บเป็นผู้ใช้แรงงาน รองลงมาคือนักเรียนหรือนักศึกษา และส่วนมากเป็นผู้ขับขี่ ยานพาหนะที่เกิดอุบัติเหตุมากที่สุดอยู่ในกลุ่มของรถจักรยานยนต์ สาเหตุของการบาดเจ็บเกิดจากการชน ลักษณะการบาดเจ็บส่วนใหญ่เป็น Blunt ผู้บาดเจ็บส่วนมากไม่ได้ใช้แอลกอฮอล์ หรือ ยาเสพติด และผู้บาดเจ็บส่วนมากไม่ได้ใช้เข็มขัดนิรภัย หรือ หมวกนิรภัย ผู้นำส่งผู้บาดเจ็บส่วนใหญ่เป็นหน่วยบริการฉุกเฉินทางการแพทย์ Severity of TBI ส่วนมากจะเป็นแบบ Mild TBI และ ถนนที่เกิดอุบัติเหตุบ่อยที่สุดคือ ถนนกรมทางหลวง เมื่อดูจากแต่ละลักษณะเส้นทาง (ทางลาดชัน ทางตรง ทางแยก ทางม้าลาย ทางโค้ง และทางมีสิ่งกีดขวาง) แล้วพบว่าทางตรงมีมีบาดเจ็บมากที่สุด รองลงมาคือทางโค้ง และ ทางแยก จากตำแหน่งการนั่งของผู้บาดเจ็บ และ ลักษณะการชน พบว่าทั้ง 2 ลักษณะเฉพาะ มีข้อมูลสูญหายเป็นจำนวนมาก

ขั้นตอนการดำเนินงานวิจัย

1. ทำข้อมูลให้สมบูรณ์ (Data Cleansing) เนื่องจากข้อมูลที่ได้มานั้นยังไม่สามารถนำมาใช้ในกระบวนการเหมืองข้อมูลได้ ดังนั้นจึงต้องมีการจัดการข้อมูลโดยทำการปรับชื่อของลักษณะเฉพาะและทำการจัดข้อมูลให้เป็นแบบเดียวกัน จากนั้นทำการจัดการข้อมูลที่ไม่เหมาะสมหรือ ที่เป็นค่าว่าง ให้เหมาะสมกับลักษณะเฉพาะของข้อมูลนั้นๆ

2. คัดเลือกข้อมูล (Data Selection) หรือการคัดเลือกลักษณะเฉพาะ คือ การคัดเลือกข้อมูลเพื่อนำมาสร้างตัวพยากรณ์ในงานวิจัยนี้ได้คัดเลือกลักษณะเฉพาะที่เป็นปัจจัยที่ทำให้เกิดอุบัติเหตุและส่งผลกระทบต่อระดับความรุนแรงของการบาดเจ็บ

3. การแบ่งชุดข้อมูล การนำข้อมูลเพื่อนำมาสร้างตัวพยากรณ์นั้น จะแบ่งข้อมูลเป็น 2 ส่วน คือ ชุดข้อมูลเรียนรู้ (Training Data Set) และ ชุดข้อมูลทดสอบ (Testing Data Set) ซึ่งมีวิธีการแบ่งที่นิยมใช้กัน 2 แบบ คือ แบบสุ่มเลือกข้อมูล (Sampling Data) และ แบบ k - folds Cross-Validation

4. เทคนิคเหมืองข้อมูล (Data Mining Technique) เป็นกระบวนการที่นำอัลกอริทึมแบบต่างๆ มาใช้สร้างตัวพยากรณ์ หรือหาความสัมพันธ์ข้อมูล หรือจัดกลุ่มข้อมูล เพื่อนำมาซึ่งตัวพยากรณ์ หรือทำนายแนวโน้มการเกิดขึ้น หรือผลลัพธ์ของข้อมูลที่ยังไม่เกิดขึ้น

5. การวัดประสิทธิภาพ เป็นการประเมินค่าความถูกต้อง หรือ ค่าความแม่นยำของตัวพยากรณ์ที่ได้มา วิธีการประเมินความแม่นยำของตัวจำแนกในกลไกทางเหมืองข้อมูล ประกอบด้วย ค่าความถูกต้องในการจำแนกกลุ่ม ค่าความระลึก หรือ ค่าผลบวกจริง ค่าความแม่นยำ ค่าความเหวี่ยง และ ค่าผลบวกเท็จ

เครื่องมือที่ใช้ในการทำวิจัย

งานวิจัยนี้ได้ใช้เครื่องมือในการดำเนินงานวิจัย ประกอบด้วย

1. ด้านโปรแกรมคอมพิวเตอร์ หรือ ซอฟต์แวร์ (Software)
 - 1.1. โปรแกรม SPSS for windows version 17 เพื่อทำการกรองข้อมูล และ การสุ่มเลือกข้อมูล
 - 1.2. โปรแกรม Microsoft Office Excel 2007 เพื่อทำการกรองข้อมูล
 - 1.3. โปรแกรม Weka 3.6.0 เพื่อสร้างและทดสอบตัวพยากรณ์
 - 1.4. โปรแกรม Eden 1.73 เพื่อสร้างตัวพยากรณ์
2. ด้านอุปกรณ์คอมพิวเตอร์ หรือ ฮาร์ดแวร์ (Hardware)
 - 2.1. หน่วยประมวลผล (CPU) Inter Core 2 Duo P7450 2.13 GHz
 - 2.2. หน่วยความจำหลัก (Ram) DDR 3 4 GB
 - 2.3. หน่วยความจำสำรอง (Hddisk) 320 GB
 - 2.4. ระบบปฏิบัติการ (OS) Microsoft Windows XP Service Pack 3

การพัฒนาตัวพยากรณ์

ในงานวิจัยได้ทำการพัฒนาตัวแบบพยากรณ์ โดยแบ่งเป็น 3 ประเด็นหลัก คือ

1. การพัฒนาตัวพยากรณ์จำแนกระดับความรุนแรงเป็น 4 ระดับ คือ ปกติ เสียชีวิต พิการเข้าเกณฑ์ และ พิการไม่เข้าเกณฑ์
2. การพัฒนาตัวพยากรณ์จำแนกระดับความรุนแรงเป็น 3 ระดับ คือ เสียชีวิต พิการเข้าเกณฑ์ และ พิการไม่เข้าเกณฑ์
3. ตัวพยากรณ์ที่ประยุกต์ใช้ตัวจำแนก 2 ตัวซ้อนกัน จะมีการจำแนกระดับความรุนแรงเป็น 4 ระดับ คือ ปกติ เสียชีวิต พิการเข้าเกณฑ์ และ พิการไม่เข้าเกณฑ์

ตัวพยากรณ์จำแนกระดับความรุนแรงเป็น 4 ระดับ

ตัวพยากรณ์จำแนกระดับความรุนแรงเป็น 4 ระดับ คือ ปกติ เสียชีวิต พิกการเข้าเกณฑ์ และ พิกการไม่เข้าเกณฑ์ โดยใช้ข้อมูลผู้บาดเจ็บจากอุบัติเหตุการจราจรทางถนน ที่ได้ทำการคัดเลือกลักษณะเฉพาะเหลือเพียง 33 ตัว และมีข้อมูลทั้งหมด 9,624 ระเบียบ แล้วนำข้อมูลที่ได้มาไปสร้างตัวพยากรณ์

การคัดเลือกข้อมูล

การคัดเลือกลักษณะเฉพาะของข้อมูลเพื่อนำมาใช้สร้างโมเดลนั้น หลังจากได้ทำการคัดเลือกลักษณะเฉพาะ 33 ตัวแล้ว เมื่อสังเกตลักษณะเฉพาะบางตัวจะไม่มีมีความเกี่ยวข้องกับ ความบาดเจ็บ เช่น หมายเลขแบบสอบถาม โรงพยาบาล สิทธการรักษา ช่วงเดือนที่เกิดเหตุ ที่อยู่ ปัจจุบัน ผู้นำส่ง ความดันโลหิต Severity of TBI เป็นต้น และลักษณะเฉพาะที่ข้อมูลมีค่าว่างเป็นจำนวนมากอย่างเช่น ตำแหน่งการนั่งของผู้บาดเจ็บ ลักษณะการชน เป็นต้น ดังนั้นจึงลดลักษณะเฉพาะลงเหลือเพียง 19 ตัว คือ ช่วงอายุ เพศ อาชีพ ช่วงเทศกาล (เทศกาลปีใหม่ และ เทศกาลสงกรานต์) ผู้บาดเจ็บ ยานพาหนะ ลักษณะการบาดเจ็บ สาเหตุการบาดเจ็บ ช่วงเวลาที่ เกิดเหตุ การใส่ยา การใช้แอลกอฮอล์ การใช้เข็มขัดนิรภัย การใช้หมวกนิรภัย ถนนที่เกิดเหตุ ลักษณะเส้นทาง (ทางตรง ทางแยก ทางโค้ง) และ ระดับความรุนแรงของการบาดเจ็บ จากนั้นทำการคัดเลือกลักษณะเฉพาะโดยใช้การทดสอบไคสแควร์ (Chi – Square test) ที่ระดับนัยสำคัญ เท่ากับ 0.05 เพื่อพิจารณาระดับความสัมพันธ์ของลักษณะเฉพาะที่มีต่อระดับความรุนแรงของการบาดเจ็บโดยใช้โปรแกรม SPSS for Windows Version 17.0 ในการวิเคราะห์ผล ดังแสดงในตาราง 3 ผลที่ได้จากการวิเคราะห์พบว่า มีลักษณะเฉพาะจำนวน 9 ตัวที่มีความสัมพันธ์กับระดับความรุนแรงของการบาดเจ็บอย่างมีนัยสำคัญทางสถิติ ซึ่งลักษณะเฉพาะดังกล่าวจะถูกนำไปสร้างตัวพยากรณ์ต่อไป

ตาราง 3 แสดงผลของการทดสอบไคสแควร์

ลักษณะเฉพาะ		ลักษณะของข้อมูล	Chi-Square	p-value
บุคคล	อายุ	{0-4, 5-14, 15-29, 30-44, 45-59, 60-69, 70-79, มากกว่า 80}	183.738	<.001*
	เพศ	{ชาย, หญิง}	21.478	<.001*
ยานพาหนะ	ประเภท	{จักรยานยนต์, รถยนต์, รถบรรทุก, อื่นๆ}	16.945	<.152
	ผู้บาดเจ็บ	{คนเดินเท้า, คนขับ, ผู้โดยสาร, ไม่ทราบ}	102.445	<.001*
การบาดเจ็บ	ลักษณะ	{ blunt, penetrating, blunt and penetrating, other}	39.573	<.001*
ถนน	ถนนที่เกิดเหตุ	{หมู่บ้าน, ในเมือง, กรมทางหลวง, กรมทางหลวงชนบท, อื่นๆ, ไม่ทราบ}	43.653	<.001*
	ทางตรง	{ไม่ใช่, ใช่}	658.718	<.001*
	ทางแยก	{ไม่ใช่, ใช่}	158.827	<.001*
	ทางโค้ง	{ไม่ใช่, ใช่}	103.008	<.001*
ระดับความรุนแรง		{ปกติ, เสียชีวิต, พิกัดเข้าเกณฑ์, พิกัดไม่เข้าเกณฑ์}		

การสร้างตัวพยากรณ์เบื้องต้นด้วย Weka

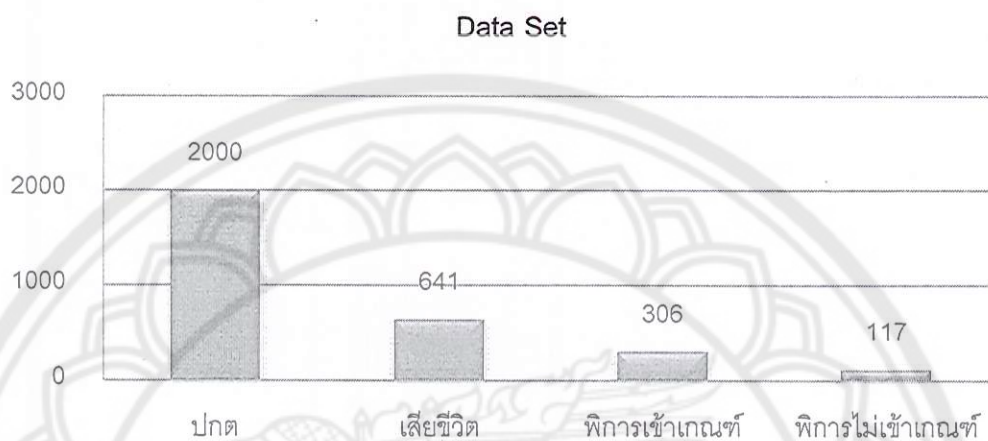
ในเบื้องต้นจากข้อมูล 9,624 ระเบียน ได้ทำการทดลองพัฒนาตัวพยากรณ์โดยใช้โปรแกรม Weka (Waikato University, n.d.) และเลือกใช้การแบ่งชุดข้อมูล 2 แบบ คือ แบบสุ่มเลือกข้อมูล 90 % และ แบบ 10 – Folds ภายในโปรแกรม Weka

โดยเลือกใช้เทคนิคการทำเหมืองข้อมูลแบบการจำแนกเพื่อจำแนกความรุนแรงของการบาดเจ็บ 3 เทคนิค คือ เทคนิคเหมืองข้อมูลแบบต้นไม้ตัดสินใจ (RandomTree และ C4.5) และแบบจำลองแบบเบย์อย่างง่าย (Naive Bayes) แล้วทำการวัดประสิทธิภาพของตัวพยากรณ์โดยใช้ค่าความแม่นยำในการจำแนก แล้วนำค่าความแม่นยำที่ได้จากตัวพยากรณ์ทั้ง 3 แบบมาเปรียบเทียบกัน

การสร้างตัวพยากรณ์โดยเลือกข้อมูลบางส่วน

เนื่องจากข้อมูลของผู้บาดเจ็บที่มีระดับความรุนแรงของการบาดเจ็บเป็นปกติมีเป็นจำนวนมากทำให้ข้อมูลไม่มีความสมดุลกัน ผู้วิจัยจึงทำการสุ่มเลือกข้อมูลของผู้บาดเจ็บที่ระดับความรุนแรงเป็นปกตินี้มา 2,000 ระเบียนจากทั้งหมด 8,560 ระเบียน เพื่อนำมารวมกับข้อมูล

ผู้บาดเจ็บที่มีระดับความรุนแรงเป็น เสียชีวิต พิการเข้าเกณฑ์ และ พิการไม่เข้าเกณฑ์ ซึ่งมีจำนวน 1,064 ระเบียบ ได้ชุดข้อมูลใหม่ที่มีจำนวนชุดข้อมูลทั้งหมด 3,064 ระเบียบ ข้อมูลที่ได้จากการเลือกข้อมูลบางส่วนนั้นมีสัดส่วนของข้อมูลดังภาพ 11

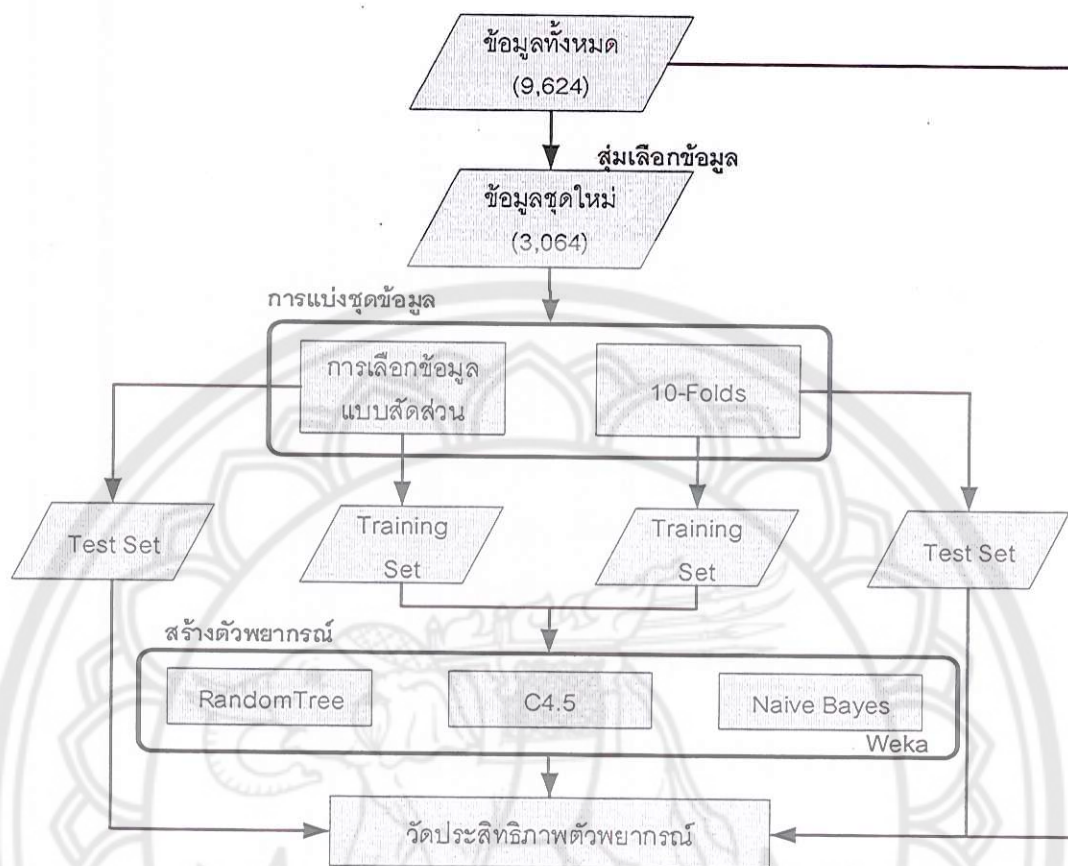


ภาพ 11 จำนวนข้อมูลผู้บาดเจ็บจากการเลือกข้อมูลบางส่วน

จากข้อมูล 3,064 ระเบียบที่สุ่มมาทำการทดลองพัฒนาตัวพยากรณ์โดยเลือกใช้การแบ่งชุดข้อมูลแบบ 10 – Folds และ การแบ่งชุดข้อมูลแบบสัดส่วนโดยการสุ่มเลือกข้อมูลระดับปกติ 1,500 ระเบียบ ระดับเสียชีวิต 530 ระเบียบ ระดับพิการเข้าเกณฑ์ 280 ระเบียบ และ ระดับพิการไม่เข้าเกณฑ์ 110 ระเบียบ เพื่อทำเป็นชุดข้อมูลเรียนรู้ และใช้ข้อมูลที่เหลือจากการสุ่มเลือกทำเป็นชุดข้อมูลทดสอบ การพัฒนาตัวพยากรณ์ในขั้นนี้จะมีการจัดสรรจำนวนข้อมูลระดับความรุนแรงปกติ เสียชีวิต พิการเข้าเกณฑ์ และ พิการไม่เข้าเกณฑ์ ในชุดข้อมูลเรียนรู้เป็นสัดส่วนที่ใกล้เคียงกันมากขึ้น

จากนั้นเลือกใช้เทคนิคการทำเหมืองข้อมูลแบบการจำแนกเพื่อจำแนกความรุนแรงของการบาดเจ็บ 3 เทคนิค คือ เทคนิคเหมืองข้อมูลแบบต้นไม้ตัดสินใจ (RandomTree และ C4.5) และแบบจำลองแบบเบย์อย่างง่าย (Naive Bayes) แล้วทำการวัดประสิทธิภาพของตัวพยากรณ์โดยใช้ค่าความแม่นยำในการจำแนก แล้วนำค่าความถูกต้องที่ได้จากตัวพยากรณ์ทั้ง 3 แบบมาเปรียบเทียบกัน

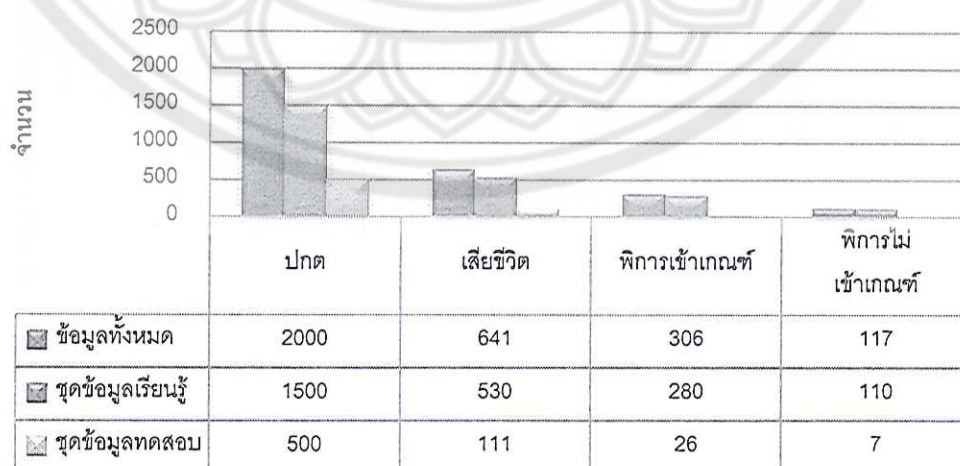
ตัวพยากรณ์โดยเลือกข้อมูลบางส่วนมีโครงสร้างการทำงานดังแสดงในภาพ 12



ภาพ 12 โครงสร้างของตัวพยากรณ์โดยเลือกข้อมูลบางส่วน

โดยข้อมูลชุดใหม่จะมีจำนวนข้อของแต่ละความรุนแรงดังภาพ 13

ชุดข้อมูลผู้บาดเจ็บ



ภาพ 13 จำนวนข้อมูลผู้บาดเจ็บที่ได้จากการเลือกข้อมูล

ตัวพยากรณ์จำแนกระดับความรุนแรงเป็น 3 ระดับ

ตัวพยากรณ์จำแนกระดับความรุนแรงเป็น 3 ระดับ คือ เสียชีวิต พิการเข้าเกณฑ์ และ พิการไม่เข้าเกณฑ์ เพื่อนำมาไปสร้างตัวพยากรณ์ โดยตัดข้อมูลที่มีระดับความรุนแรงของการบาดเจ็บเป็นปกติออกเพราะข้อมูลมีจำนวนมากและส่งผลต่อความแม่นยำในการจำแนกข้อมูล โดยมีขั้นตอนในการสร้างตัวพยากรณ์ดังนี้

1. การทำข้อมูลให้สมบูรณ์

ข้อมูลที่ถูกคัดกรองเบื้องต้นแล้วมีทั้งหมด 9,624 ระเบียน ตัดเป็นข้อมูลที่มีระดับความรุนแรงเป็นปกติ 8,560 ระเบียนออก เหลือข้อมูลที่ใช้สร้างตัวพยากรณ์ทั้งหมด 1,064 ระเบียน

2. การคัดเลือกข้อมูล

จากการคัดเลือกลักษณะเฉพาะ 19 ตัว แล้วรวมข้อมูลเทศบาลปีใหม่ และ เทศกาลสงกรานต์ ไว้ด้วยกันเพราะลักษณะเฉพาะทั้ง 2 ตัวนี้เป็นช่วงเทศกาลเหมือนกันแล้วตั้งชื่อลักษณะเฉพาะเป็นช่วงเทศกาล และรวมข้อมูลการใช้เข็มขัดนิรภัยและการใช้หมวกนิรภัยไว้ด้วยกันเนื่องจากเป็นอุปกรณ์นิรภัยเหมือนกันเพียงแต่ใช้แตกต่างกันตามยานพาหนะ แล้วตั้งชื่อลักษณะเฉพาะเป็นการใช้อุปกรณ์นิรภัย เหลือลักษณะเฉพาะ 17 ตัว

จากนั้นทำการคัดเลือกลักษณะเฉพาะโดยใช้การทดสอบไคสแควร์ (Chi – Square test) ที่ระดับนัยสำคัญเท่ากับ 0.05 เพื่อพิจารณาระดับความสัมพันธ์ของลักษณะเฉพาะที่มีต่อระดับความรุนแรงของการบาดเจ็บโดยใช้โปรแกรม SPSS for Windows Version 17.0 ในการวิเคราะห์ผล ดังแสดงในตาราง 4 ผลที่ได้จากการวิเคราะห์พบว่ามีลักษณะเฉพาะจำนวน 15 ตัวที่มีความสัมพันธ์กับระดับความรุนแรงของการบาดเจ็บอย่างมีนัยสำคัญทางสถิติ ซึ่งลักษณะเฉพาะดังกล่าวจะถูกนำไปสร้างตัวพยากรณ์ต่อไป

ตาราง 4 แสดงผลของการทดสอบไคสแควร์หลังจากเปลี่ยนแปลงลักษณะเฉพาะ

			Pearson	
ลักษณะเฉพาะ			Chi-Square	p-value
บุคคล	อายุ	{0-4, 5-14, 15-29, 30-44, 45-59, 60-69, 70-79, มากกว่า 80}	188.79	<.001*
	เพศ	{ชาย, หญิง}	21.25	<.001*
	อาชีพ	{ข้าราชการ, ตำรวจ, รัฐวิสาหกิจ, พนักงานบริษัท, แรงงาน, ค้าขาย, เกษตรกรรม, นักเรียน, อื่นๆ}	83.16	<.001*
	แอลกอฮอล์	{ไม่ใช้, ใช้, ไม่ทราบ}	1899.36	<.001*
	ยาเสพติด	{ไม่ใช้, ใช้, ไม่ทราบ}	1213.59	<.001*
	เครื่องนิรภัย	{ไม่ใช้, ใช้, ไม่ทราบ}	549.23	<.001*
สิ่งแวดล้อม	ช่วงเวลา	{6.01 – 9.00, 9.01 – 12.00, 12.01 – 15.00, 15.01 – 18.00, 18.01 – 21.00, 21.01 – 0.00, 0.01 – 3.00, 3.01 – 6.00}	22.66	.363
	เทศกาล	{ไม่ใช้, ใช้}	7.94	.047*
	ยานพาหนะ	{จักรยานยนต์, รถยนต์, รถบรรทุก, อื่นๆ}	145.37	<.001*
การบาดเจ็บ	ผู้บาดเจ็บ	{คนเดินเท้า, คนขับ, ผู้โดยสาร, ไม่ทราบ}	285.36	<.001*
	สาเหตุ	{ตก, ชน, อื่นๆ}	66.64	<.001*
	ลักษณะ	{blunt, penetrating, blunt and penetrating, other}	76.10	<.001*
ถนน	ถนนที่เกิดเหตุ	{หมู่บ้าน, ในเมือง, กรมทางหลวง, กรมทางหลวงชนบท, อื่นๆ, ไม่ทราบ}	1302.43	<.001*
	ทางตรง	{no, yes}	496.62	<.001*
	ทางแยก	{no, yes}	113.36	<.001*
	ทางโค้ง	{no, yes}	80.01	<.001*
ระดับความรุนแรง		{เสียชีวิต, พิกัดเข้าเกณฑ์, พิกัดไม่เข้าเกณฑ์}		

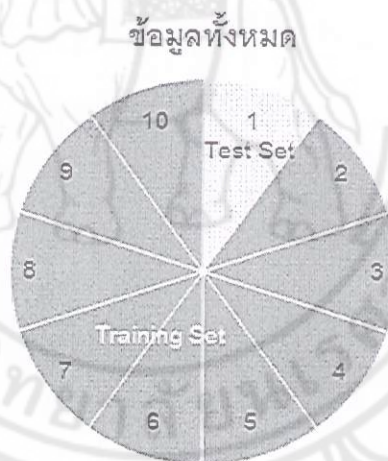
จากตาราง 4 แสดงให้เห็นว่าลักษณะเฉพาะช่วงเวลา จะเป็นลักษณะเฉพาะที่ไม่ถูกเลือกมาใช้ ดังนั้นเหลือลักษณะเฉพาะที่นำไปใช้สร้างตัวพยากรณ์ 15 ตัว

3. การแบ่งชุดข้อมูล

ตัวพยากรณ์จำแนกระดับความรุนแรง 3 ระดับนี้ใช้การแบ่งจำนวนชุดข้อมูลเพื่อนำมาสร้างตัวพยากรณ์ทั้งหมด 3 วิธี คือ

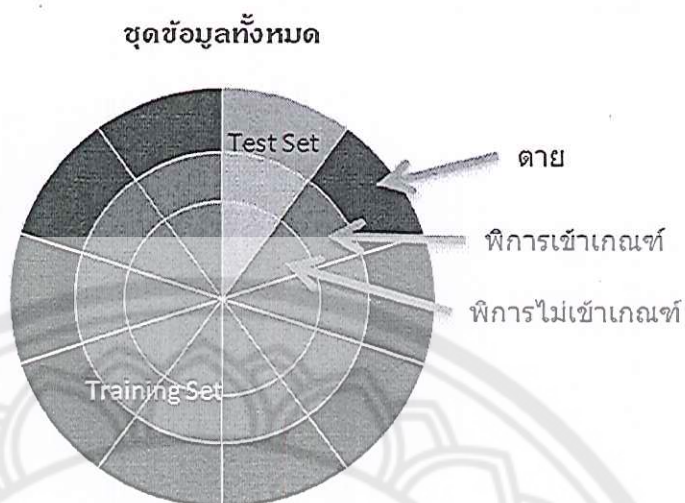
3.1. วิธีสุ่มเลือกข้อมูล 90% คือ การสุ่มเลือกข้อมูลมา 90 % จากข้อมูลทั้งหมด มาทำเป็นชุดข้อมูลเรียนรู้เพื่อนำไปสร้างแบบจำลอง และใช้ข้อมูลที่เหลือทำเป็นชุดข้อมูลทดสอบ โดยทำการสุ่มซ้ำ 10 ครั้ง แล้วนำค่าความแม่นยำที่ได้จากการทดสอบแบบจำลองมาหาค่าเฉลี่ย

3.2. วิธี 10 – Folds คือ การแบ่งชุดข้อมูลออกเป็น 10 ชุดเท่าๆ กันโดยการสุ่มเลือกข้อมูล แล้วนำข้อมูล 9 ชุดมารวมกันใช้เป็นชุดข้อมูลเรียนรู้ และข้อมูลชุดที่เหลือใช้เป็นชุดข้อมูลทดสอบ ดังภาพ 14 โดยจะทำทั้งหมด 10 รอบ โดยจะสลับเปลี่ยนชุดข้อมูลกันจนครบ แล้วนำค่าความแม่นยำที่ได้จากการทดสอบแบบจำลองมาหาค่าเฉลี่ย



ภาพ 14 การแบ่งชุดข้อมูลวิธี 10 – Folds

3.3. วิธี Modified 10 – Folds คือ การแบ่งข้อมูลเป็น 10 ชุด คล้ายกับการทำ 10 – Folds แต่จะมีการแบ่งจำนวนข้อมูลของแต่ละคำตอบให้มีจำนวนสัดส่วนที่ใกล้เคียงกันเป็น 10 ชุดเท่าๆ กัน ดังภาพ 15 แล้วนำข้อมูลในแต่ละคำตอบมารวมกันได้เป็นชุดข้อมูล 1 ชุด แล้วทำชุดข้อมูลเรียนรู้ และ ชุดข้อมูลทดสอบ เหมือนกับการทำ 10 – Folds



ภาพ 15 การแบ่งชุดข้อมูลวิธี Modified 10 – Folds

4. เทคนิคเหมืองข้อมูล

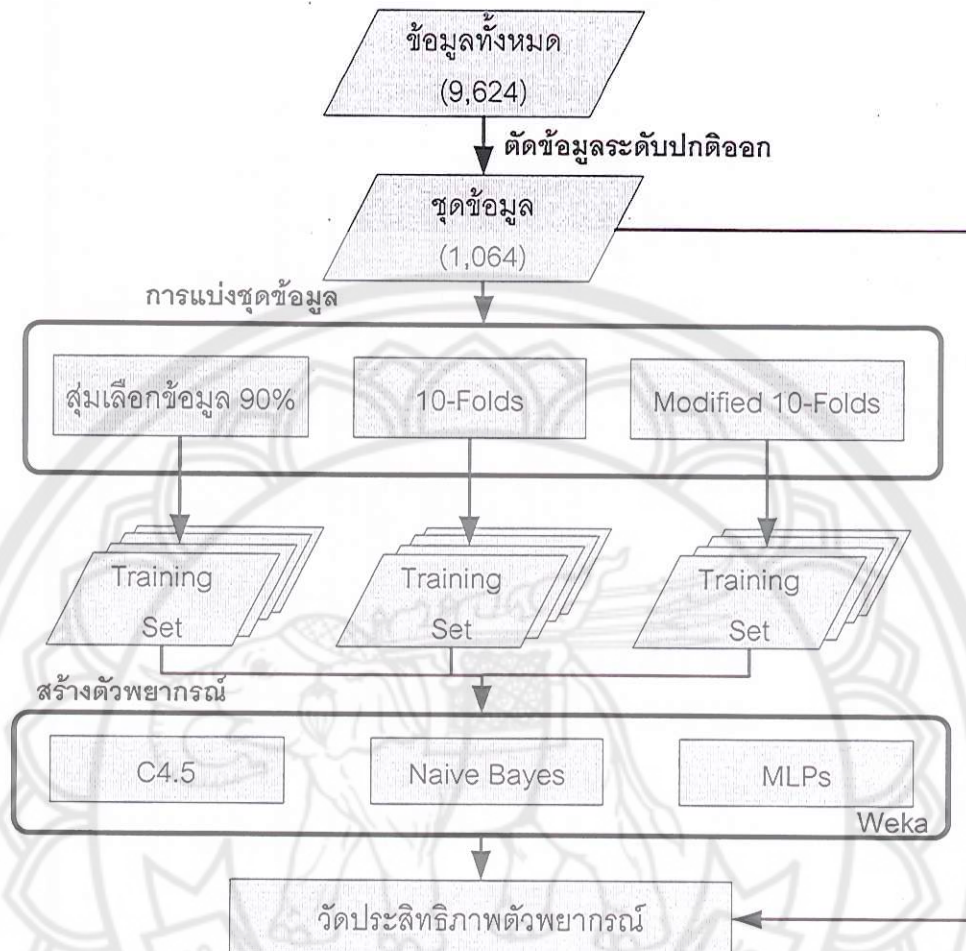
เทคนิคเหมืองข้อมูลที่ผู้วิจัยเลือกใช้เพื่อจำแนกความรุนแรงสำหรับข้อมูลชุดนี้นั้น ผู้วิจัยเลือกใช้และเลือกใช้เทคนิคเหมืองข้อมูลแบบต้นไม้ตัดสินใจ (C4.5) แบบจำลองแบบเบย์อย่างง่าย (Naive Bayes) และ โครงข่ายประสาทเทียม (MLPs)

5. วัดประสิทธิภาพ

โดยการคำนวณถูกต้องที่ได้จากตัวพยากรณ์ทั้ง 3 แบบมาเปรียบเทียบกัน เพื่อหาตัวพยากรณ์ที่ดีที่สุด

ตัวพยากรณ์จำแนกระดับความรุนแรงเป็น 3 ระดับมีโครงสร้างการทำงานดังแสดงใน

ภาพ 16



ภาพ 16 โครงสร้างของตัวพยากรณ์จำแนกระดับความรุนแรงเป็น 3 ระดับ

ตัวพยากรณ์ที่ประยุกต์ใช้ตัวจำแนก 2 ตัวซ้อนกัน

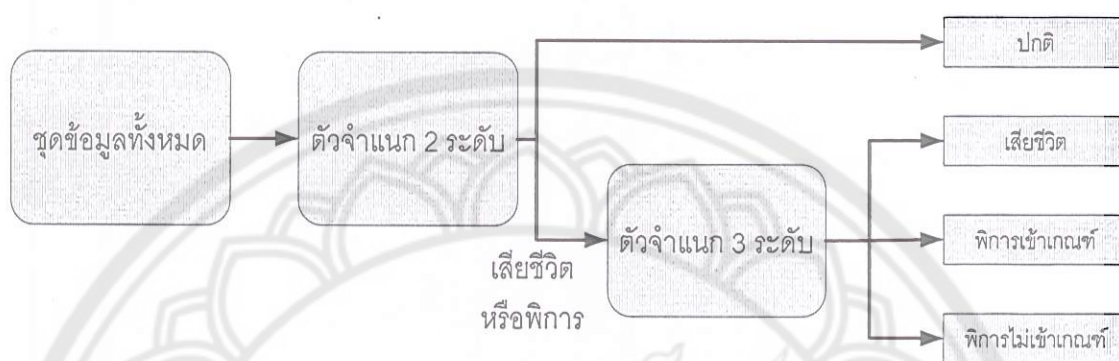
ตัวพยากรณ์ที่ประยุกต์ใช้ตัวจำแนก 2 ตัวซ้อนกัน

จำแนกระดับความรุนแรงเป็น 4 ระดับ คือ ปกติ เสียชีวิต พิกัดเข้าเกณฑ์ และ พิกัดไม่เข้าเกณฑ์ โดยใช้ข้อมูลผู้บาดเจ็บจากอุบัติเหตุการจราจรทางถนน ที่มีข้อมูลทั้งหมด 9,624 ระเบียน และใช้ลักษณะเฉพาะ 15 ตัว จากการคัดเลือกที่กล่าวไว้ก่อนหน้านี้ที่ใช้การทดสอบไคสแควร์ (Chi – Square test) ที่ระดับนัยสำคัญเท่ากับ 0.05 เนื่องจากข้อมูลของผู้บาดเจ็บที่นำมาใช้ในการสร้างตัวพยากรณ์นี้มีระดับความรุนแรงของการบาดเจ็บที่เป็นปกติจำนวนมาก ผู้วิจัยจึงได้ประยุกต์ใช้ตัวจำแนก 2 ตัว มาใช้ในการสร้างตัวพยากรณ์ ได้แก่

1. ตัวจำแนกที่จำแนกระดับความรุนแรงเป็น 2 ระดับ คือ ปกติ และ เสียชีวิตหรือพิกัด

2. ตัวจำแนกที่จำแนกระดับความรุนแรงเป็น 3 ระดับ คือ เสียชีวิต พิกการเข้าเกณฑ์ และ พิกการไม่เข้าเกณฑ์

โดยมีโครงสร้างของตัวพยากรณ์ดังภาพ 17



ภาพ 17 โครงสร้างของการประยุกต์ใช้ตัวจำแนก 2 ตัวซ้อนกัน

ตัวพยากรณ์ที่ประยุกต์ใช้ตัวจำแนก 2 ตัวซ้อนกัน มีการทำงานโดยนำข้อมูลไปจำแนกด้วยตัวจำแนก 2 ระดับก่อน ถ้าจำแนกได้ระดับปกติ จะเป็นคำตอบของตัวพยากรณ์ แต่ถ้าจำแนกได้ระดับเสียชีวิตหรือพิกการ จะใช้ตัวจำแนก 3 ระดับมาจำแนกต่อ คำตอบที่ได้จากตัวจำแนกนี้จะเป็น เสียชีวิต พิกการเข้าเกณฑ์ หรือ พิกการไม่เข้าเกณฑ์

โดยเลือกใช้เทคนิคการทำเหมืองข้อมูลแบบการจำแนกเพื่อจำแนกความรุนแรงของการบาดเจ็บ 4 เทคนิค คือ เทคนิคเหมืองข้อมูลแบบต้นไม้ตัดสินใจ (RandomTree และ C4.5) แบบจำลองแบบเบย์อย่างง่าย (Naive Bayes) และ โครงข่ายประสาทเทียม (MLPs) ในการสร้างตัวพยากรณ์ แล้ววัดประสิทธิภาพของตัวพยากรณ์โดยใช้ค่าความแม่นยำในการจำแนก แล้วนำค่าความแม่นยำที่ได้จากตัวพยากรณ์ทั้ง 4 แบบมาเปรียบเทียบกัน