

บทที่ 5

บทสรุป

สรุปผลการวิจัย

สำหรับการวิจัยเรื่อง การประยุกต์ใช้เทคนิคเหมืองข้อมูลที่เหมาะสมสำหรับจำแนกความรุนแรงของการบาดเจ็บจากอุบัติเหตุการจราจรทางบก ผู้วิจัยได้แบ่งขั้นตอนในการวิจัยออกเป็น 3 ขั้นตอน ได้แก่ ตัวพยากรณ์จำแนกระดับความรุนแรงเป็น 4 ระดับ ตัวพยากรณ์จำแนกระดับความรุนแรงเป็น 3 ระดับ และ ตัวพยากรณ์ที่ประยุกต์ใช้ตัวจำแนก 2 ตัวซ้อนกัน เพื่อให้ได้ตัวพยากรณ์ที่มีประสิทธิภาพ

ในการพัฒนาตัวพยากรณ์ปัจจัยที่มีผลต่อประสิทธิภาพของตัวพยากรณ์ คือ การเลือกลักษณะเฉพาะ การคัดเลือกข้อมูล การแบ่งชุดข้อมูล และ เทคนิคเหมืองข้อมูลที่ใช้

การพัฒนาตัวพยากรณ์แบบ 4 ระดับ โดยใช้การทดสอบไคสแควร์ (Chi – Square test) ที่ระดับนัยสำคัญเท่ากับ 0.05 ในการคัดเลือกลักษณะเฉพาะมา 9 ตัว เพื่อใช้สร้างตัวพยากรณ์ ได้ค่าความแม่นยำที่ 81% แต่การจำแนกมีความผิดพลาดค่อนข้างมากในระดับ เสียชีวิต พิกการเข้าเกณฑ์ และพิกการไม่เข้าเกณฑ์ ทั้งนี้เพราะจำนวนข้อมูลระดับปกติมีปริมาณมากกว่าเสียชีวิต พิกการเข้าเกณฑ์ และพิกการไม่เข้าเกณฑ์มาก และเมื่อเปรียบเทียบความแม่นยำในภาพรวมพบว่า อัลกอริทึม RandomTree ให้ความแม่นยำดีที่สุด

การพัฒนาตัวพยากรณ์แบบ 3 ระดับ โดยใช้การทดสอบไคสแควร์ (Chi – Square test) ที่ระดับนัยสำคัญเท่ากับ 0.05 ในการคัดเลือกลักษณะเฉพาะเช่นกัน และได้ลักษณะเฉพาะมา 15 ตัว เพื่อใช้สร้างตัวพยากรณ์ การแบ่งชุดข้อมูลแบบ Modified 10 - Folds ทำให้ตัวพยากรณ์ที่ได้มีความแม่นยำที่ดีขึ้น และเมื่อเปรียบเทียบความแม่นยำพบว่า อัลกอริทึม MLPs ให้ความแม่นยำดีที่สุด โดยให้ค่าความแม่นยำที่ 81.39%

การพัฒนาตัวจำแนกแบบ 2 ตัวซ้อนกัน เป็นการพัฒนาตัวพยากรณ์ที่สามารถจำแนกระดับความรุนแรงของการบาดเจ็บเป็น 4 ระดับ โดยการใช้อัลกอริทึม RandomTree ในการจำแนกเป็น 2 ระดับ (ปกติ และ เสียชีวิตหรือพิกการ) แล้วใช้อัลกอริทึม MLPs มาจำแนกเป็น 3 ระดับ ได้ค่าความแม่นยำ 93.67% ซึ่งดีขึ้นมากกว่าตัวพยากรณ์ 4 ระดับ

ซึ่งจากผลการทดลองสามารถสรุปได้ดังนี้

1. จากการศึกษาโดยการแบ่งระดับความรุนแรงที่แตกต่างกันมีผลต่อประสิทธิภาพของตัวพยากรณ์ที่ได้เนื่องจากข้อมูลที่ใช้ในการนำมาสร้างแบบจำลองนั้นมีจำนวนข้อมูลในแต่ละระดับความรุนแรงไม่มีความสมดุลกัน การตัดข้อมูลระดับความรุนแรงที่มีจำนวนข้อมูลมากเกินไป หรือใช้จำนวนข้อมูลในระดับความรุนแรงนั้นในปริมาณที่เหมาะสมกับชุดข้อมูลที่เหลือ จึงเป็นทางเลือกที่จะทำให้ประสิทธิภาพของตัวพยากรณ์ดีขึ้น
2. การเลือกลักษณะเฉพาะเพื่อนำมาใช้ในการสร้างตัวพยากรณ์มีความสำคัญอย่างยิ่งกับประสิทธิภาพของตัวพยากรณ์ที่ได้ ในที่นี้เลือกใช้การทดสอบไคสแควร์ (Chi – Square test) ที่ระดับนัยสำคัญเท่ากับ 0.05 ในการคัดเลือก เนื่องจากข้อมูลในแต่ละลักษณะเฉพาะมีลักษณะเป็นเชิงคุณภาพ ไม่ใช่เชิงปริมาณการทดสอบไคสแควร์จึงเหมาะสมกว่า จากการศึกษาพบว่าจำนวนลักษณะเฉพาะที่เลือกมาใช้ในการสร้างตัวพยากรณ์ถ้ามีจำนวนน้อยเกินไปจะส่งผลเสียต่อความแม่นยำในการพัฒนาตัวพยากรณ์ ดังเช่นในงานวิจัยของ J. Rungrattanaubol, A. Na-udom and A. Harfield (2012) ได้กล่าวไว้ และจากงานวิจัยนี้การใช้จำนวนลักษณะเฉพาะทั้งหมด 15 ตัว ให้ตัวพยากรณ์ที่ดีกว่าการใช้ลักษณะเฉพาะ 9 ตัว
3. การแบ่งชุดข้อมูลเพื่อนำมาสร้างตัวพยากรณ์มีผลต่อประสิทธิภาพของตัวพยากรณ์ที่ได้ ในงานวิจัยนี้ได้นำเสนอเทคนิคการแบ่งชุดข้อมูลแบบ Modified 10 – Folds ซึ่งเป็นการแบ่งชุดข้อมูลข้อมูลที่คล้ายกับการทำ 10 – Folds ปกติ แต่จะมีการแบ่งจำนวนข้อมูลของแต่ละกลุ่มคำตอบให้มีสัดส่วนใกล้เคียงกันเป็น 10 ชุด แล้วนำข้อมูลในแต่ละคำตอบมารวมกันเป็นชุดข้อมูล 1 ชุด แล้วทำชุดข้อมูลเรียนรู้ และชุดข้อมูลทดสอบเหมือนกับการทำ 10 – Folds ซึ่งในงานวิจัยนี้ได้เสนอวิธีการแบ่งชุดข้อมูลลักษณะนี้และให้ตัวพยากรณ์ที่มีความแม่นยำที่ดีขึ้นในทุกอัลกอริทึม
4. การเลือกเทคนิคเหมืองข้อมูลที่เหมาะสมกับชนิดข้อมูลส่งผลต่อประสิทธิภาพของตัวพยากรณ์ เนื่องจากข้อมูลแต่ละชุดจะมีความเหมาะสมกับเทคนิคเหมืองข้อมูลที่แตกต่างกัน สำหรับข้อมูลชุดนี้ ตัวพยากรณ์จำแนกระดับความรุนแรงเป็น 4 ระดับ อัลกอริทึม RandomTree ให้ค่าความแม่นยำดีที่สุด และตัวพยากรณ์จำแนกระดับความรุนแรงเป็น 3 ระดับ อัลกอริทึม MLPs ให้ค่าความแม่นยำดีที่สุด แต่ตัวพยากรณ์นี้ไม่สามารถบอกปัจจัยที่เป็นสาเหตุของการเกิดอุบัติเหตุจากการจราจรทางบกได้แน่ชัด แต่เทคนิคเหมืองข้อมูลแบบต้นไม้ตัดสินใจจะสามารถแปลความหมายของตัวพยากรณ์ได้ง่ายกว่าแต่ความแม่นยำที่ได้น้อยกว่า

5. จากการทำตัวพยากรณ์จำแนกระดับความรุนแรงเป็น 3 และ 4 ระดับ แล้วจึงนำตัวจำแนก 2 ตัวมาประยุกต์ใช้ร่วมกันโดยใช้อัลกอริทึม RandomTree และ MLPs ซ้อนกันดังภาพ 25 จะทำให้ได้ตัวพยากรณ์ที่มีประสิทธิภาพมากขึ้น

อภิปรายผล

ผลจากการคัดกรองข้อมูลด้วยวิธีการทางสถิติพบว่าปัจจัยที่น่าจะมีผลต่อการเกิดอุบัติเหตุการจราจรทางบกและส่งผลต่อความรุนแรงของการบาดเจ็บได้แก่ บุคคล(อายุ, เพศ, อาชีพ) ยานพาหนะ ผู้บาดเจ็บ (คนเดินเท้า, ผู้ขับขี่, ผู้โดยสาร) สาเหตุและลักษณะการบาดเจ็บ การใช้แอลกอฮอล์ การใช้ยา การใช้อุปกรณ์นิรภัย และ ลักษณะของถนนที่เกิดเหตุ ล้วนแต่เป็นปัจจัยที่ทำให้เกิดอุบัติเหตุและมีผลต่อความรุนแรงของการบาดเจ็บทั้งสิ้น ผู้วิจัยจึงได้นำปัจจัยเหล่านี้ไปใช้ในการทำการออกแบบและพัฒนาตัวพยากรณ์

การวิจัยนี้พบว่าการสร้างตัวพยากรณ์สำหรับข้อมูลผู้บาดเจ็บจากอุบัติเหตุการจราจรทางบกนั้นเทคนิคเหมืองข้อมูลแบบโครงข่ายประสาทเทียมให้ค่าความแม่นยำมากที่สุดรองลงมาคือต้นไม้ตัดสินใจ และ แบบจำลองแบบเบย์ ตามลำดับ เนื่องมาจากสาเหตุของการเกิดอุบัติเหตุและการบาดเจ็บจากการเกิดอุบัติเหตุมีปัจจัยหลายๆ ปัจจัยที่เกี่ยวข้องกัน ดังนั้น การเลือกใช้เทคนิคเหมืองข้อมูลที่เหมาะสมกับลักษณะของข้อมูลจะทำให้ตัวพยากรณ์มีประสิทธิภาพ

จากตัวพยากรณ์ที่ได้แสดงให้เห็นว่าปัจจัยหลักๆ ที่น่าจะเป็นสาเหตุที่ทำให้เกิดอุบัติเหตุและส่งผลต่อระดับความรุนแรงของการบาดเจ็บ ได้แก่ ลักษณะเส้นทาง ยานพาหนะ สาเหตุการบาดเจ็บ ลักษณะการบาดเจ็บ ผู้บาดเจ็บ และ การใช้แอลกอฮอล์

ข้อเสนอแนะ

1. การนำข้อมูลเพื่อนำมาใช้ นั้น ถ้าข้อมูลมีความสมบูรณ์ครบถ้วนจะสามารถนำข้อมูลนั้นมาใช้ได้หลากหลายมากขึ้น แต่ถ้าข้อมูลที่ได้มาไม่สมบูรณ์ การตัดลักษณะเฉพาะหรือค่าของลักษณะเฉพาะบางค่าทิ้งไปก็เป็นทางเลือกที่เหมาะสม

2. การแบ่งชุดข้อมูลเพื่อนำไปสร้างเป็นชุดข้อมูลเรียนรู้ และ ชุดข้อมูลทดสอบนั้นมีส่วนในการแบ่งอีกหลายวิธีซึ่งการแบ่งสัดส่วนนี้มีผลต่อประสิทธิภาพของตัวพยากรณ์ที่ได้ ดังนั้นถ้าเลือกแบ่งสัดส่วนของชุดข้อมูลที่ดีจะทำให้ตัวพยากรณ์ที่ได้มีประสิทธิภาพที่ดีขึ้น

3. เนื่องจากเทคนิคเหมืองข้อมูลแต่ละเทคนิคมีความเหมาะสมกับลักษณะของข้อมูลที่แตกต่างกัน ดังนั้นการเลือกใช้เทคนิคเหมืองข้อมูลที่เหมาะสมกับลักษณะของข้อมูลส่งผลต่อ

ประสิทธิภาพของแบบจำลองที่ได้ และลักษณะของข้อมูลที่ใช้ในการวิจัยนี้ยังมีเทคนิคเหมือนข้อมูลอื่นๆ ที่สามารถนำมาใช้ได้อีก

4. ตัวพยากรณ์ที่ได้มาเป็นการทำบนชุดข้อมูลที่นำมาสร้างตัวพยากรณ์ ยังไม่ได้ทำการทดสอบกับชุดข้อมูลอื่น หรือใช้กับข้อมูลจริง

5. ตัวพยากรณ์ที่ได้สามารถนำไปทำเป็นตัวพยากรณ์แบบออนไลน์เพื่อนำไปใช้พยากรณ์ต่อไปได้

