การวิเคราะห์ความเสี่ยงที่จะเสียชีวิตในเส้นทางการเดินทางจากอุบัติเหตุจราจรทางบก โดยใช้เทคนิคเหมืองข้อมูล

Analysis of the risk of death in road traffic accidents using data mining techniques

ภัณฑิรา มาติยา[1*] ภุมรัตน์ อยู่สุข[2] และสุพาพร บรรดาศักดิ์[3*]

Puntira Matiya[1] Poommarat Yoosuk[2] and Supaporn Bundasak[3*]

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์ในการวิเคราะห์ความเสี่ยงที่จะเสียชีวิต เพื่อทำการเปรียบเทียบความเหมาะสมของตัวพยากรณ์กับข้อมูลที่นำมาวิจัย และเพื่อเป็นแนวทางในการป้องกันการเกิดอุบัติเหตุ ส่งเสริมให้อัตราการเสียชีวิตลดลง โดยผู้วิจัยได้รวบรวมข้อมูลที่เกี่ยวข้องจากระบบรายงานข้อมูลอุบัติเหตุของกระทรวงคมนาคม ใช้ข้อมูลในการวิเคราะห์ตั้งแต่วันที่ 13 เมษายน 2564 - 31 มีนาคม 2565 จำนวน 20 แอตทริบิวต์ จำนวนข้อมูล 12,500 ข้อมูล โดยใช้เทคนิคเหมืองข้อมูลทั้งหมด 4 โมเดล ได้แก่ การเรียนรู้แบบต้นไม้ตัดสินใจ วิธีการเรียนรู้แบบเบย์อย่างง่าย วิธีการเพื่อนบ้านใกล้ที่สุด และซัพพอร์ตเวกเตอร์แมชชีน ผลการวิจัยครั้งนี้พบว่าเทคนิคที่มีความเหมาะสมกับข้อมูลมากที่สุดคือวิธีการเรียนรู้แบบเบย์อย่างง่ายมีค่าประสิทธิภาพเท่ากับ 75%

คำสำคัญ : อุบัติเหตุ การจราจรทางบก เหมืองข้อมูล พยากรณ์

ABSTRACT

The purpose of this research was to analyze the risk of death to compare the appropriateness of the predictor with the research data and to be a guideline for accident prevention to encourage the death rate to decrease. By collecting data from the accident reporting system of the Ministry of Transport, data for analysis from

[1,2,3] สาขาวิชาเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตศรีราชา

[*] ผู้ประสานงานหลัก

Received: Feb 22, 2023 **Revised:** Apr 22, 2023 **Accepted:** Jun 12, 2023

13 April 2021 - 31 March 2022 20 features 12,500 data were used. By using four models of data mining techniques, which were decision tree, Naïve Bayes, K-nearest neighbors, and support vector machine. The results showed that the most suitable technique for information was the Naïve Bayes with 75%

Keywords : accidents, road traffic, data mining, forecasting

บทนำ

ข้อมูลจากกลุ่มสถิติการขนส่งกองแผนงานกรมการขนส่งทางบก ด้วยความอนุเคราะห์ข้อมูลจากสำนักงานตำรวจแห่งชาติ จากสถิติการเกิดอุบัติเหตุจราจรทางบกทั่วประเทศในปี พ.ศ. 2564 มีผู้ที่ได้รับบาดเจ็บและเสียชีวิตจำนวน 59,909 ราย เฉลี่ยเดือนละ 4,992 ราย เพิ่มขึ้นเป็นร้อยละ 2.55 เมื่อเปรียบเทียบกับปี พ.ศ. 2563 หากจำแนกตามสถิติผู้ที่เสียชีวิตจากการเกิดอุบัติเหตุ มีจำนวน 6,024 ราย คิดเป็นร้อยละ 10.05 มีผู้ที่ได้รับความเสียหายจากการเกิดอุบัติเหตุมากที่สุดในช่วงเดือนมีนาคม พ.ศ. 2564 โดยผู้ที่ได้รับความเสียหายมีจำนวน 7,360 ราย เพิ่มขึ้นถึงร้อยละ 36.22 เมื่อทำการจำแนกตามสถิติผู้ที่ได้รับความเสียหายจากการเกิดอุบัติเหตุเป็นรายภาคพบว่า ส่วนกลาง (กรุงเทพมหานคร) มีจำนวน 15,615 ราย คิดเป็นร้อยละ 26.06 ในส่วนภูมิภาคมีจำนวน 44,294 ราย (ร้อยละ 73.94) ภาคที่มีผู้ได้รับความเสียหายมากที่สุดคือ ภาคตะวันออกเฉียงเหนือ มีจำนวน 19,469 ราย คิดเป็นร้อยละ 43.95 รองลงมาคือภาคใต้ มีจำนวน 7,813 ราย คิดเป็นร้อยละ 17.64 และภาคเหนือ มีจำนวน 7,145 ราย คิดเป็นร้อยละ 16.13 ผู้จัดทำนำข้อมูลการเกิดอุบัติเหตุจากระบบรายงานข้อมูลอุบัติเหตุของกระทรวงคมนาคม (transport accident management systems: TRAMS) (ศูนย์เทคโนโลยีสารสนเทศและการสื่อสาร สำนักงานปลัดกระทรวงคมนาคม, 2565) ตั้งแต่วันที่ 13 เมษายน พ.ศ. 2564 - 31 มีนาคม พ.ศ. 2565 จำนวน 20 แอตทริบิวต์ จำนวนข้อมูล 12,500 ข้อมูล โดยจำแนกตามเวลาที่เกิดเหตุ จังหวัด อำเภอ ตำบล ประเภทรถ สภาพอากาศ ผู้เสียชีวิต โดยใช้เทคนิคเหมืองข้อมูล ทั้งหมด 4 โมเดล ได้แก่ การเรียนรู้แบบต้นไม้ตัดสินใจ วิธีการเรียนรู้แบบเบย์อย่างง่าย วิธีการเพื่อนบ้านใกล้ที่สุด และซัพพอร์ตเวกเตอร์แมชชีน จากนั้นนำข้อมูลมาวิเคราะห์เพื่อหาความเสี่ยงที่จะเสียชีวิตจากการเกิดอุบัติเหตุจราจรทางบก นำผลที่ได้จากการวิเคราะห์มาคาดการณ์ความรุนแรงที่เกิดขึ้นจากการเกิดอุบัติเหตุ และสำหรับผู้ที่มีความสนใจหรือทำการวิเคราะห์ข้อมูลเกี่ยวกับอุบัติเหตุจราจรสามารถนำไปประกอบการตัดสินใจในการวิเคราะห์ข้อมูลได้

โดยผู้วิจัยได้มีการศึกษาทฤษฎีเกี่ยวกับเหมืองข้อมูลและกระบวนการขุดเหมืองเพื่อช่วยให้ผู้วิเคราะห์คาดการณ์แนวโน้มในอนาคต นำหลักการของกระบวนการวิเคราะห์ข้อมูลด้วย CRISP-DM มาประยุกต์ใช้กับการทำเหมืองข้อมูล ศึกษาทฤษฎีของเทคนิคเหมืองข้อมูลที่นำมาใช้ทั้ง 4 โมเดล อีกทั้งยังได้มีการศึกษางานวิจัยที่มีความเกี่ยวข้องกับการวิเคราะห์ความเสี่ยงในการเกิดอุบัติเหตุ จากปัญหาการเกิดอุบัติเหตุจราจรทางบกที่ส่งผลกระทบ และความเสียหายต่อประชากรรวมไปถึงความเสียหายต่อเส้นทางที่ใช้ในการสัญจร การวิจัยครั้งนี้จึงมีวัตถุประสงค์ในการวิเคราะห์ความเสี่ยงที่จะเสียชีวิต เพื่อทำการเปรียบเทียบความเหมาะสมของตัวพยากรณ์กับข้อมูลที่นำมาวิจัย และเพื่อเป็นแนวทางในการป้องกันการเกิดอุบัติเหตุ ส่งเสริมให้อัตราการเสียชีวิตลดลง

ศราวุธ แดงมาก (2560) ได้กล่าวไว้ว่า การทำเหมืองข้อมูลเป็นการเรียบเรียงชุดข้อมูลขนาดใหญ่เพื่อระบุรูปแบบความสัมพันธ์ มีเทคนิคและเครื่องมือต่างๆในการทำเหมืองข้อมูล เพื่อช่วยให้ผู้วิเคราะห์คาดการณ์แนวโน้มในอนาคต เป็นสิ่งที่ช่วยในการตัดสินใจอย่างมีข้อมูลมากขึ้น กระบวนการขุดเหมืองขอมูลมี 6 ขั้นตอน ดังนี้

1. เลือกข้อมูล (data) ที่จะนำมาทำเป็นข้อมูลเป้าหมาย เพื่อใช้ในการทำเหมืองข้อมูลจากฐานข้อมูลหรือแหล่งข้อมูลต่างๆ

2. ข้อมูลเป้าหมาย (target data) ส่วนใหญ่จะเป็นข้อมูลที่มีรูปแบบชัดเจน เช่น อายุ เพศ วันเดือนปีเกิด เป็นต้น

3. การเตรียมข้อมูล (preprocessed data) ทำการสำรวจข้อมูล ปรับปรุงข้อมูล ล้างข้อมูลที่มีความผิดพลาด

4. การแปลงข้อมูล (transformed data) เพื่อลดการกระจายของข้อมูลให้เหมาะกับการวิเคราะห์

5. รูปแบบของข้อมูล (patterns) อาศัยการทำเหมืองข้อมูลในลักษณะต่างๆ โดยใช้โปรแกรม Weka ในการวิเคราะห์ข้อมูล

6. การนำเสนอผลลัพธ์หรือความรู้ใหม่ (knowledge) ที่ค้นพบได้จากการทำเหมืองข้อมูลหรือนำข้อมูลไปพัฒนาต่อได้

นคร ละลอกน้ำ (2562) ได้กล่าวถึงกระบวนการวิเคราะห์ข้อมูลด้วย CRISP-DM ไว้ดังต่อไปนี้ CRISP-DM เป็นคำที่มาจากคำว่า cross-industry standard process for data mining เป็นแนวทางในการทำเหมืองข้อมูล เป็นขั้นตอนการทำงานที่ต่อเนื่องกัน ขั้นตอนถัดไปจะรอผลลัพธ์จากขั้นตอนก่อนหน้า ประกอบด้วย 6 ขั้นตอน ได้แก่

1. การทำความเข้าใจปัญหา (business understanding) กำหนดวัตถุประสงค์ในมุมมองทางธุรกิจ

2. การรวบรวมข้อมูล (data understanding) การทำความเข้าใจข้อมูล ตรวจสอบคุณภาพข้อมูลเพื่อหาข้อมูลที่เหมาะกับโจทย์ที่จะนำไปวิเคราะห์

3. การเตรียมข้อมูล (data preparation) แปลงข้อมูลให้มีความถูกต้อง (cleaning data) โดยการลบข้อมูลที่ไม่สมบูรณ์ มีความผิดพลาด ให้สามารถนำไปวิเคราะห์ได้

4. การเลือกเทคนิคต่าง ๆ (modeling) และทดลองสร้างแบบจำลองหลากหลายรูปแบบ เพื่อให้ได้มาซึ่งแบบจำลองที่เหมาะสมที่สุดในการแก้ปัญหา

5. การวัดประสิทธิภาพ (evaluation) ของผลลัพธ์จากการทดลองสร้างแบบจำลองหลากหลายรูปแบบ

6. การนำแบบจำลองที่เหมาะสมที่สุดไปใช้งาน (deployment) เพื่อวิเคราะห์ คาดการณ์ หรือแก้ปัญหาตามที่ผู้วิเคราะห์ต้องการ

ภัทธิรา สุวรรณโค, นิศาชล จำนงศรี และจิติมนต์ อั่งสกุล (2560) ได้ทำการรวบรวมข้อมูลเพื่อสร้างแบบจำลองพยากรณ์ความเสี่ยงในการเกิดอุบัติเหตุทางถนนเฉพาะในช่วงเทศกาลปีใหม่ โดยใช้ข้อมูลระหว่างปี พ.ศ. 2551 - พ.ศ. 2558 จำนวน 214,952 ข้อมูล ประกอบด้วยแอตทริบิวท์ 10 แอตทริบิวท์ จากข้อมูลของแอตทริบิวท์ จุดเกิดเหตุ สามารถจำแนกได้เป็น 4 คลาส คือ ไม่ทราบจุดเกิดเหตุมีอัตราส่วนเท่ากับ 2.43 เกิดเหตุในเมือง มีอัตราส่วนเท่ากับ 16.96 เกิดเหตุบนทางหลวง มีอุตราส่วนเท่ากับ 24.62 เกิดเหตุในชนบทมีอัตราส่วนเท่ากับ 41.46 ผู้วิจัยใช้หลักการแยกด้วยวิธีการ 10 - fold cross validation เลือกเทคนิค 3 เทคนิคมาประกอบการวิจัย ดังนี้ 1) Naïve Bayes 2) multilayer perceptron 3) meta bagging เมื่อสร้างแบบจำลองพยากรณ์ความเสี่ยงแล้ว ผู้วิจัยทำการวัดประสิทธิภาพโดยวัดค่าความถ่วงดุลและพื้นที่ใต้เส้นกราฟ ROC(AUC) ของแบบจำลองพยากรณ์ ผลสรุปว่าเทคนิค meta bagging มีความเหมาะสมที่จะนำไปใช้ในการกยากรณ์ต่อไป เนื่องจากมีค่าประสิทธิภาพมากที่สุดคิดเป็นร้อยละ 77.3

ฮิชาม อาแว และคณะ (2565) ได้ทำการรวบรวมข้อมูลจากตัวแทนสมาชิกครอบครัวจากชุมชนในพื้นที่จุดเสี่ยงและผู้ใช้รถจักรยานยนต์จำนวน 30 คน โดยใช้ข้อมูลระหว่างปี พ.ศ. 2563 - 2564 เก็บข้อมูลโดยการสนทนาและการสังเกต เครื่องมือในการวิจัยคือ แบบสัมภาษณ์ แบบประเมินพฤติกรรม จากนั้นนำข้อมูลมาวิเคราะห์โดยใช้สถิติเชิงพรรณนา ทำการเปรียบเทียบคะแนนเฉลี่ยก่อนและหลังทำแบบสอบ ผลการวิจัยพบว่าปี พ.ศ. 2564 อัตราการบาดเจ็บของชุมชนคือ 102 คนหรือลดลงร้อยละ 7.53 จากปี พ.ศ. 2563 ซึ่งมีอัตราการบาดเจ็บ 197 คน และคะแนนเฉลี่ยพฤติกรรมการรับรู้และป้องกันอุบัติเหตุทางสถิติ หลังทำแบบสอบถามคือ ($t = -10.04$, $p < .001$) ซึ่งสูงกว่าก่อนพัฒนาอย่างมีนัยสำคัญ

Gupta, Solanki, Singh, and García-Díaz (2018) ได้ทำการรวบรวมงานวิจัยการวิเคราะห์ปัญหาความรุนแรงของอุบัติเหตุทางจราจรด้วยเทคนิคเหมืองข้อมูล พบว่ามีการนำเทคนิคเหมือง

ข้อมูลมาใช้ในงานวิจัยตั้งแต่ ปี 1971 เป็นต้นมา เทคนิควิธีที่ใช้ได้แก่ K-mean algorithm K-modes clustering Naïve Bayes Bayesian classifier decision tree ANN ARM DT C-means C4.5 C-RT CS-MS4 Decision List และID3

Husin, Afdzal, Azmi, and Hamadi (2021) ได้ทำการรวบรวมชุดข้อมูลจำนวนอุบัติเหตุทางถนนของประเทศมาเลเซีย นำข้อมูลมาจาก Polis Diraja Malaysia (PDRM) มาใช้ในการวิเคราะห์เพื่อสร้างแบบจำลองพยากรณ์และคาดกาณ์แนวโน้มของกรณีอุบัติเหตุทางถนน โดยใช้ข้อมูลตั้งแต่เดือนมกราคม พ.ศ. 2544 - ธันวาคม พ.ศ. 2562 ผู้วิจัยนำแบบจำลอง Box-Jenkins และ State space มาประกอบการวิเคราะห์จากผลการศึกษาพบว่า การคาดการณ์ล่วงหน้าในอีก 10 ปีข้างหน้าตั้งแต่เดือนมกราคม พ.ศ. 2563 - ธันวาคม พ.ศ. 2573 มีแนวโน้มการเกิดอุบัติเหตุคงที่

จากการศึกษาทฤษฎีและงานวิจัยที่เกี่ยวข้องที่กล่าวมาข้างต้นนี้ ผู้วิจัยจึงได้เริ่มดำเนินการวิจัยโดยนำทฤษฎีและงานวิจัยที่ได้ศึกษามาปรับใช้กับงานวิจัยในครั้งนี้

วิธีดำเนินการวิจัย

วิธีการดำเนินงานในการสร้างโมเดลการวิเคราะห์ความเสี่ยงที่จะเสียชีวิตในเส้นทางการเดินทางจากอุบัติเหตุจราจรทางบก โดยนำกระบวนการวิเคราะห์ข้อมูลด้วย CRISP-DM มาประยุกต์ใช้ ดังนี้

1. ปัญหาที่พบและการทำความเข้าใจปัญหา

ทำความเข้าใจปัญหาด้านความเสี่ยงที่จะเสียชีวิตในเส้นทางการเดินทางจากอุบัติเหตุจราจรทางบก เช่น ปัญหาในเรื่องลักษณะหรือเส้นทางการจราจร สาเหตุ มูลเหตุสันนิษฐานจากการเกิดอุบัติเหตุจราจร สภาพอากาศขณะเกิดอุบัติเหตุ ช่วงเวลาที่เกิดอุบัติเหตุ สถานที่ที่เกิดอุบัติเหตุ รถที่เกิดอุบัติเหตุ

2. ทำความเข้าใจข้อมูล

มีการรวบรวมข้อมูลจากระบบรายงานข้อมูลอุบัติเหตุของกระทรวงคมนาคม (ศูนย์เทคโนโลยีสารสนเทศและการสื่อสาร สำนักงานปลัดกระทรวงคมนาคม, 2565) ใช้ข้อมูลในการวิเคราะห์ตั้งแต่วันที่ 13 เมษายน พ.ศ. 2564 - 31 มีนาคม พ.ศ. 2565 จำนวน 20 แอตทริบิวต์ จำนวนข้อมูล 12,500 ข้อมูล นำมาใช้จริงจำนวน 7 แอตทริบิวต์ จำนวนข้อมูล 9,350 ข้อมูล ได้แก่ เวลาที่เกิดเหตุ จังหวัด อำเภอ ตำบล รถที่เกิดอุบัติเหตุ สภาพอากาศ และมีผู้ที่เสียชีวิตใช่หรือไม่ ดังตารางที่ 1

ตารางที่ 1 ข้อมูลที่ใช้สำหรับงานวิจัย

ชื่อคอลัมน์	รายละเอียด
Time	เวลาที่เกิดเหตุ
Province	จังหวัด
District	อำเภอ
Subdistrict	ตำบล
Cars	รถที่เกิดอุบัติเหตุ
Climate	สภาพอากาศ
Die	เสียชีวิตหรือไม่เสียชีวิต

3. การเตรียมข้อมูล

จัดเตรียมข้อมูลอุบัติเหตุที่รวบรวมมาจากระบบรายงานข้อมูลอุบัติเหตุของกระทรวงคมนาคม (ศูนย์เทคโนโลยีสารสนเทศและการสื่อสาร สำนักงานปลัดกระทรวงคมนาคม, 2565) ก่อนที่จะนำไปวิเคราะห์ในโปรแกรม Weka 3.9.6 ดังรูปที่ 1 และ 2

- การแปลงข้อมูล จากข้อมูลที่ทำการรวบรวมมา ซึ่งข้อมูลที่ได้มาเป็นภาษาไทย จึงต้องทำการแปลงข้อมูลให้เป็นภาษาอังกฤษให้ข้อมูลอยู่ในรูปแบบที่โปรแกรม Weka สามารถวิเคราะห์ได้

	A	B	C	D	E	F	G	H	I
1	ลำดับ	วันที่เกิดเห	เวลา	จังหวัด	อำเภอ	ตำบล	รถคันที่ 1	บริเวณที่เกิดเหตุ/ลักษณะทาง	จำนวนทาง
2	3	1/1/2022	0:03	ชุมพร	ThaSae	NaKratam	รถยนต์นั่งส่วนบุคคล/รถยนต์นั่งสาธารณะ	ทางตรง+ไม่มีความลาดชัน	7681
3	4	1/1/2022	0:05	ภูเก็ต	Kathu	Kamala	รถจักรยานยนต์	ทางตรง+ไม่มีความลาดชัน	7681
4	5	1/1/2022	0:05	อุดรธานี	MueangUd	NongKhonKwang	รถปิคอัพบรรทุก 4 ล้อ	ทางตรง+ไม่มีความลาดชัน	7681
5	7	1/1/2022	0:08	อยุธยา	BangPa-in	BanPho	รถยนต์นั่งส่วนบุคคล/รถยนต์นั่งสาธารณะ	ทางตรง+ไม่มีความลาดชัน	7681
6	8	1/1/2022	0:10	อุทัยธานี	LanSak	Rabam	รถจักรยานยนต์	ทางตรง+ไม่มีความลาดชัน	7681
7	11	1/1/2022	0:20	นครราชสีม	KhamThal	PongDaeng	รถจักรยานยนต์	ทางตรง+ไม่มีความลาดชัน	7681
8	13	1/1/2022	0:25	เชียงใหม่	WiangHae	MueangHaeng	รถจักรยานยนต์	ทางตรง+ไม่มีความลาดชัน	7681
9	14	1/1/2022	0:25	ชัยนาท	Sankhabur	DongKhon	รถจักรยานยนต์	ทางตรง+ไม่มีความลาดชัน	7681
10	15	1/1/2022	0:30	ลำปาง	HangChat	WiangTan	รถบรรทุกมากกว่า 6 ล้อ ไม่เกิน 10 ล้อ	ทางตรง+ไม่มีความลาดชัน	7681
11	16	1/1/2022	0:30	พิษณุโลก	MueangPh	BanKrang	รถยนต์นั่งส่วนบุคคล/รถยนต์นั่งสาธารณะ	ทางตรง+ไม่มีความลาดชัน	7681
12	17	1/1/2022	0:30	นครราชสีม	MueangNa	KhokKruat	รถจักรยานยนต์	ทางตรง+ไม่มีความลาดชัน	7681
13	18	1/1/2022	0:30	บึงกาฬ	Seka	Seka	รถบรรทุกมากกว่า 10 ล้อ (รถพ่วง)	ทางตรง+ไม่มีความลาดชัน	7681
14	19	1/1/2022	0:30	สมุทรสงคร	MueangSa	BangKaeo	รถบรรทุกมากกว่า 6 ล้อ ไม่เกิน 10 ล้อ	ทางตรง+ไม่มีความลาดชัน	7681
15	21	1/1/2022	0:30	อุตรดิตถ์	Phichai	BanKhon	รถจักรยานยนต์	ทางตรง+ไม่มีความลาดชัน	7681
16	22	1/1/2022	0:40	สุพรรณบุรี	NongYaSai	NongKham	รถจักรยานยนต์	ทางตรง+ไม่มีความลาดชัน	7681
17	23	1/1/2022	1:00	หนองบัวลำ	MueangNc	NaMafueang	รถยนต์นั่งส่วนบุคคล/รถยนต์นั่งสาธารณะ	ทางตรง+ไม่มีความลาดชัน	7681
18	24	1/1/2022	1:00	นนทบุรี	PakKret	BanMai	รถยนต์นั่งส่วนบุคคล/รถยนต์นั่งสาธารณะ	ทางตรง+ไม่มีความลาดชัน	7681
19	25	1/1/2022	1:00	เพชรบุรี	Cha-am	Cha-am	รถยนต์นั่งส่วนบุคคล/รถยนต์นั่งสาธารณะ	ทางตรง+ไม่มีความลาดชัน	7681
20	29	1/1/2022	1:20	อ่างทอง	PaMok	Ekkarat	รถยนต์นั่งส่วนบุคคล/รถยนต์นั่งสาธารณะ	ทางตรง+ไม่มีความลาดชัน	7681
21	30	1/1/2022	1:20	นครศรีธรรม	ThungSon	NongHong	รถจักรยานยนต์	ทางตรง+ไม่มีความลาดชัน	7681
22	31	1/1/2022	1:22	สุราษฎร์ธา	Phunphin	Maluan	รถยนต์นั่งส่วนบุคคล/รถยนต์นั่งสาธารณะ	ทางตรง+ไม่มีความลาดชัน	7681
23	32	1/1/2022	1:30	สระบุรี	PhraPhuttl	PhuKhang	รถยนต์นั่งส่วนบุคคล/รถยนต์นั่งสาธารณะ	ทางตรง+ไม่มีความลาดชัน	7681
24	34	1/1/2022	1:30	เชียงใหม่	MueangCh	ThaSala	รถปิคอัพบรรทุก 4 ล้อ	ทางตรง+ไม่มีความลาดชัน	7681
25	35	1/1/2022	1:30	สุพรรณบุรี	SongPhiNc	SongPhiNong	รถยนต์นั่งส่วนบุคคล/รถยนต์นั่งสาธารณะ	ทางตรง+ไม่มีความลาดชัน	7681
26	36	1/1/2022	1:35	ชัยนาท	MueangCh	BanKluai	รถจักรยานยนต์	ทางตรง+ไม่มีความลาดชัน	7681
27	37	1/1/2022	1:38	สุราษฎร์ธา	Phunphin	Phunphin	รถยนต์นั่งส่วนบุคคล/รถยนต์นั่งสาธารณะ	ทางตรง+ไม่มีความลาดชัน	7681
28	38	1/1/2022	1:40	อุดรธานี	BanPhue	KhueaNam	รถจักรยานยนต์	ทางตรง+ไม่มีความลาดชัน	7681
29	39	1/1/2022	1:40	สุพรรณบุรี	BangPlaMa	TaKha	รถจักรยานยนต์	ทางตรง+ไม่มีความลาดชัน	7681
30	42	1/1/2022	2:00	ขอนแก่น	MueangKh	KhonKaen	รถปิคอัพบรรทุก 4 ล้อ	ทางตรง+ไม่มีความลาดชัน	7681
31	43	1/1/2022	2:00	นครราชสีม	NonSung	DonChomphu	รถโดยสารขนาดใหญ่	ทางตรง+ไม่มีความลาดชัน	7681

รูปที่ 1 ตัวอย่างข้อมูลก่อนการทำความสะอาดข้อมูล

- คัดข้อมูลที่ไม่สมบูรณ์ออก ได้แก่ ข้อมูลที่ไม่ทราบค่า ข้อมูลที่กรอกว่าอื่น ๆ ข้อมูลที่บันทึกไม่ครบทั้งแถวข้อมูล และทำการทำความสะอาดข้อมูล โดยกำจัดค่าที่เป็น outlier หรือข้อมูลที่มีค่าว่าง

- การแปลงนามสกุลของไฟล์ หลังจากทำการทำความสะอาดข้อมูลแล้ว ทำการแปลงนามสกุลของไฟล์ให้เป็น .csv เพื่อให้อยู่ในรูปแบบที่โปรแกรม Weka สามารถวิเคราะห์ได้

	A	B	C	D	E	F	G	H	I
1	time	province	district	subdistrict	cars	climate	die		
2	t2	p12	d1	s1	car1	climate1	yes		
3	t2	p42	d2	s2	car2	climate1	no		
4	t2	p74	d3	s3	car3	climate1	no		
5	t2	p33	d4	s4	car1	climate1	no		
6	t2	p76	d5	s5	car2	climate1	no		
7	t2	p21	d6	s6	car2	climate1	no		
8	t2	p14	d7	s7	car2	climate1	no		
9	t2	p10	d8	s8	car2	climate1	no		
10	t2	p53	d9	s9	car5	climate1	no		
11	t2	p38	d10	s10	car1	climate1	no		
12	t2	p21	d11	s11	car2	climate1	yes		
13	t2	p27	d12	s12	car6	climate1	no		
14	t2	p61	d13	s13	car5	climate1	no		
15	t2	p75	d14	s14	car2	climate1	no		
16	t2	p67	d15	s15	car2	climate1	no		
17	t2	p71	d16	s16	car1	climate1	no		
18	t2	p24	d17	s17	car1	climate1	no		
19	t2	p39	d18	s18	car1	climate1	no		
20	t2	p72	d19	s19	car1	climate1	no		

รูปที่ 2 ตัวอย่างข้อมูลหลังการทำความสะอาดข้อมูล

4. สร้างแบบจำลองวิเคราะห์ข้อมูล

กระบวนการในการสร้างตัวแบบพยากรณ์ โดยใช้โปรแกรม Weka จากข้อมูลที่ผ่านการทำความสะอาดแล้วได้ 9,350 ข้อมูล จะถูกแบ่งออกเป็น 2 ส่วน ได้แก่ ชุดข้อมูลฝึกฝน (training Set) จำนวน 9,250 ข้อมูล และชุดข้อมูลทดสอบ (test set) จำนวน 100 ข้อมูลเพื่อเอาไว้วัดประสิทธิภาพความแม่นยำของตัวแบบพยากรณ์ที่นำมาใช้ โดยใช้ตัวแบบพยากรณ์ทั้งหมด 4 โมเดล ได้แก่

- การเรียนรู้แบบต้นไม้ตัดสินใจ (decision tree)
- วิธีการเรียนรู้แบบเบย์อย่างง่าย (Naïve Bayes)
- วิธีการเพื่อนบ้านใกล้ที่สุด (K-nearest neighbors)
- ซัพพอร์ตเวกเตอร์แมชชีน (support vector machine)

5. การประเมินผลลัพธ์

การวัดประสิทธิภาพของตัวแบบพยากรณ์ในรายงานฉบับนี้ประกอบด้วยค่าความถูกต้อง และค่าประสิทธิภาพ โดยทำการวัดประสิทธิภาพหลายครั้งเพื่อหาค่าที่ดีและเหมาะสมที่สุดก่อนที่จะนำผลลัพธ์ไปใช้

6. นำผลลัพธ์ไปใช้

นำผลการิเคราะห์จากตัวแบบพยากรณ์ทั้ง 4 โมเดลมาจัดทำเป็นตาราง และกราฟแท่ง นำข้อมูลจำนวนผู้เสียชีวิตมาสร้างแผนที่ (map chart) โดยการทำตาราง การทำกราฟแท่งและการสร้างแผนที่ทำให้ง่ายต่อการดูภาพรวมของผลลัพธ์และผลสรุปของการวิเคราะห์ข้อมูล

- การทำตาราง สร้างตารางนำเสนอข้อมูลด้วยโปรแกรม Microsoft Power Point นำผลลัพธ์จากการวิเคราะห์ข้อมูลด้วยตัวแบบพยากณ์ทั้ง 4 โมเดล นำมาทำเป็นตารางเพื่อเปรียบเทียบค่าความถูกต้องและค่าประสิทธิภาพของตัวแบบพยากรณ์

- การทำกราฟแท่ง สร้างกราฟแท่งนำเสนอข้อมูลด้วยโปรแกรม Microsoft office Excel นำผลลัพธ์จากการวิเคราะห์ข้อมูลด้วยตัวแบบพยากณ์ทั้ง 4 โมเดล นำมาทำเป็นกราฟแท่งเพื่อเปรียบเทียบค่าความถูกต้องและค่าประสิทธิภาพของตัวแบบพยากรณ์

- การสร้างแผนที่ สร้างแผนที่ด้วยโปรแกรม Microsoft Power BI โดยนำข้อมูลจำนวนผู้ที่เสียชีวิตมาเปรียบเทียบกับสถานที่ที่เกิดอุบัติเหตุในแต่ละจุดทั่วประเทศไทย จากระบบรายงานข้อมูลอุบัติเหตุของกระทรวงคมนาคม โดยเป็นข้อมูลที่ผ่านการทำความสะอาดแล้วได้ 9,350 ข้อมูล นำมาสร้างแผนที่ เพื่อแสดงจุดที่เกิดอุบัติเหตุรวมถึงจำนวนผู้เสียชีวิตในแต่ละจุด ช่วงเวลาที่เกิดอุบัติเหตุ สภาพอากาศขณะเกิดอุบัติเหตุ ประเภทรถที่เกิดอุบัติเหตุ มูลเหตุสันนิษฐานจากการเกิดอุบัติเหตุ ลักษณะทางที่เกิดอุบัติเหตุและลักษณะการเกิดอุบัติเหตุที่ส่งผลให้มีผู้เสียชีวิต

ผลการวิจัยและอภิปรายผล

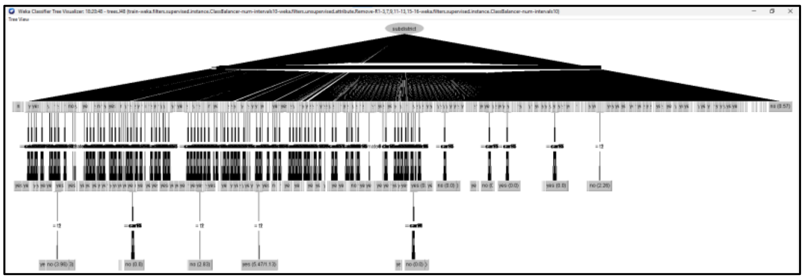
สร้างแบบจำลองด้วยโปรแกรม Weka 3.9.6 โดยใช้เทคนิคการเรียนรู้แบบต้นไม้ตัดสินใจ วิธีการเรียนรู้แบบเบย์อย่างง่าย วิธีการเพื่อนบ้านใกล้ที่สุดและซัพพอร์ตเวกเตอร์แมชชีน ทำ class balancer ก่อนที่จะทำการวิเคราะห์ข้อมูล จากนั้นใช้เทคนิค 10 - fold cross validation เพื่อหาคำตอบที่ดีที่สุด

1. การเรียนรู้แบบต้นไม้ตัดสินใจ

นำชุดข้อมูลฝึกฝนเข้าสู่โปรแกรม Weka จากนั้นใช้ model tree.J 48 ในการวิเคราะห์ มีค่าความถูกต้องเท่ากับ 54.2575% ดังรูปที่ 3 และ 4

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances       5018.8189         54.2575 %
Incorrectly Classified Instances     4231.1811         45.7425 %
Kappa statistic                         0.0852
Mean absolute error                     0.4586
Root mean squared error                 0.6149
Relative absolute error                91.7239 %
Root relative squared error           122.9796 %
Total Number of Instances            9250

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.364    0.279    0.566      0.364    0.443      0.091    0.558     0.550     yes
                 0.721    0.636    0.531      0.721    0.612      0.091    0.558     0.536     no
Weighted Avg.    0.543    0.457    0.549      0.543    0.528      0.091    0.558     0.543

=== Confusion Matrix ===

    a       b      <-- classified as
 1684.98 2940.02 |       a = yes
 1291.16 3333.84 |       b = no
```

รูปที่ 3 ผลการวิเคราะห์ด้วยเทคนิคการเรียนรู้แบบต้นไม้ตัดสินใจ



รูปที่ 4 แผนภาพการเรียนรู้แบบต้นไม้ตัดสินใจ

ยกตัวอย่างการอธิบายกฎการจำแนก If-Then ได้ 5 ข้อ ดังนี้

- If subdistrict = s565 And cars = car2 Then Yes

อธิบายผล : ถ้าเป็นตำบล/แขวงข่วงเปา และรถที่เกิดอุบัติเหตุเป็นรถจักรยานยนต์ คำตอบคือเสียชีวิต

- If subdistrict = s565 And cars = car3 And time = t2 Then Yes

อธิบายผล : ถ้าเป็นตำบล/แขวงข่วงเปา และรถที่เกิดอุบัติเหตุเป็นรถบรรทุก 4 ล้อ และเวลาที่เกิดอุบัติเหตุเป็นช่วงเวลากลางคืน คำตอบคือเสียชีวิต

- If subdistrict = s90 And cars = car1 And time = t2 Then Yes

อธิบายผล : ถ้าเป็นตำบล/แขวงศาลาธรรมสพน์ และรถที่เกิดอุบัติเหตุเป็นรถรถยนต์นั่งส่วนบุคคล/รถยนต์นั่งสาธารณะและเวลาที่เกิดอุบัติเหตุเป็นช่วงเวลากลางคืน คำตอบคือเสียชีวิต

- If subdistrict = s308 And climate = climate1 Then Yes

อธิบายผล : ถ้าเป็นตำบล/แขวงลาดสวาย และสภาพอากาศขณะที่เกิดอุบัติเหตุเป็นอากาศมืดครึ้ม คำตอบคือเสียชีวิต

- If subdistrict = s1344 And climate = climate6 Then Yes

อธิบายผล : ถ้าเป็นตำบล/แขวงหนองไผ่แก้ว และสภาพอากาศขณะที่เกิดอุบัติเหตุเป็นภัยธรรมชาติ เช่น พายุ น้ำท่วม คำตอบคือเสียชีวิต

2. วิธีการเรียนรู้แบบเบย์อย่างง่าย

นำข้อมูล Training Set เข้าสู่โปรแกรม Weka จากนั้นใช้ Naïve Bayes ในการวิเคราะห์ มีค่าความถูกต้องเท่ากับ 63.777% ดังรูปที่ 5

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        5899.3756          63.777  %
Incorrectly Classified Instances      3350.6244          36.223  %
Kappa statistic                          0.2755
Mean absolute error                      0.3891
Root mean squared error                  0.5096
Relative absolute error                 77.8284 %
Root relative squared error            101.9131 %
Total Number of Instances             9250

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.548    0.273    0.668      0.548    0.602      0.280    0.693     0.684     yes
                 0.727    0.452    0.617      0.727    0.668      0.280    0.693     0.688     no
Weighted Avg.    0.638    0.362    0.642      0.638    0.635      0.280    0.693     0.686

=== Confusion Matrix ===

    a       b      <-- classified as
 2536.15 2088.85 |       a = yes
 1261.77 3363.23 |       b = no
```

รูปที่ 5 ผลการวิเคราะห์ด้วยเทคนิควิธีการเรียนรู้แบบเบย์อย่างง่าย

3. วิธีการเพื่อนบ้านใกล้ที่สุด

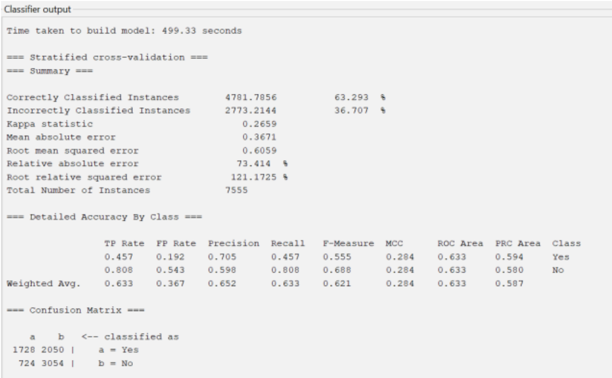
นำชุดข้อมูลฝึกฝนเข้าสู่โปรแกรม Weka จากนั้นใช้ Lazy เลือก IBk เลือก K ที่ดีที่สุด ในที่นี้ K ที่ดีที่สุดคือ K-19 ในการวิเคราะห์ มีค่าความถูกต้องเท่ากับ 65.3525% ดังรูปที่ 6

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        6045.1102          65.3525 %
Incorrectly Classified Instances      3204.8898          34.6475 %
Kappa statistic                          0.3071
Mean absolute error                      0.4139
Root mean squared error                  0.466
Relative absolute error                 82.783  %
Root relative squared error             93.2098 %
Total Number of Instances             9250

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.561    0.254    0.689      0.561    0.618      0.312    0.718     0.722     yes
                 0.746    0.439    0.629      0.746    0.683      0.312    0.718     0.702     no
Weighted Avg.    0.654    0.346    0.659      0.654    0.651      0.312    0.718     0.712

=== Confusion Matrix ===

    a       b      <-- classified as
 2592.61 2032.39 |       a = yes
 1172.5  3452.5  |       b = no
```

รูปที่ 6 ผลการวิเคราะห์ด้วยเทคนิควิธีการเพื่อนบ้านใกล้ที่สุด

4. ซัพพอร์ตเวกเตอร์แมชชีน

นำชุดข้อมูลฝึกฝนเข้าสู่โปรแกรม Weka จากนั้นใช้ function เลือก SMO เลือก calibrator เป็น logistic แล้วเลือก kernel เป็น PolyKernel ในการวิเคราะห์มีค่าความถูกต้องเท่ากับ 63.2930% ดังรูปที่ 7



```
Classifier output
Time taken to build model: 499.33 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        4781.7856          63.293  %
Incorrectly Classified Instances      2773.2144          36.707  %
Kappa statistic                          0.2659
Mean absolute error                      0.3671
Root mean squared error                  0.6059
Relative absolute error                 73.414  %
Root relative squared error            121.1725 %
Total Number of Instances             7555

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.457    0.192    0.705      0.457    0.555      0.284    0.633     0.594     Yes
                 0.808    0.543    0.598      0.808    0.688      0.284    0.633     0.580     No
Weighted Avg.    0.633    0.367    0.652      0.633    0.621      0.284    0.633     0.587

=== Confusion Matrix ===

    a    b   <-- classified as
 1728 2050 |    a = Yes
  724 3054 |    b = No
```

รูปที่ 7 ผลการวิเคราะห์ด้วยเทคนิคซัพพอร์ตเวกเตอร์แมชชีน

เมื่อได้ผลการวิเคราะห์แล้ว จึงนำตัวแบบพยากรณ์มาเปรียบเทียบค่าประสิทธิภาพและค่าความถูกต้องทั้ง 4 ตัวแบบพยากรณ์ เปรียบเทียบความเหมาะสมของตัวพยากรณ์กับข้อมูลที่นำมาวิจัย ดังตารางที่ 2

ตารางที่ 2 เปรียบเทียบค่าความถูกต้องและค่าประสิทธิภาพของตัวพยากรณ์

ตัวแบบพยากรณ์	ค่าความถูกต้อง	ค่าประสิทธิภาพ
การเรียนรู้แบบต้นไม้ตัดสินใจ	54.26%	66%
วิธีการเรียนรู้แบบเบย์อย่างง่าย	63.78%	75%
วิธีการเพื่อนบ้านใกล้ที่สุด	65.35%	69%
ซัพพอร์ตเวกเตอร์แมชชีน	63.29%	71%

สรุปผล

งานวิจัยนี้ได้พัฒนาตัวแบบพยากรณ์การวิเคราะห์ความเสี่ยงที่จะเสียชีวิตจากอุบัติเหตุจราจรทางบก เพื่อเป็นแนวทางในการป้องกันการเกิดอุบัติเหตุ และส่งเสริมให้อัตราการเสียชีวิตลดลง โดยคาดการณ์ความรุนแรงจากการวิเคราะห์ข้อมูลจากอุบัติเหตุที่เคยเกิดขึ้นแล้ว เพื่อทำการ

เปรียบเทียบความเหมาะสมของตัวพยากรณ์กับข้อมูลที่นำมาวิจัย หากมีผู้วิจัยท่านอื่นนำงานวิจัยนี้ไปต่อยอดสามารถพัฒนาเป็นระบบที่ช่วยในเรื่องการเตือนภัยเกี่ยวกับอุบัติเหตุ หรือระบบป้องกันช่วยเหลือการเกิดอุบัติเหตุจราจรทางบก

การประเมินการวิจัยครั้งนี้ได้นำข้อมูลจำนวนทั้งหมด 12,500 ข้อมูล มาทำความสะอาด คัดข้อมูล และเลือกข้อมูลที่เป็นปัจจัยสำคัญที่ส่งผลให้เกิดความเสี่ยงต่อการเสียชีวิตไว้ทำให้ได้ข้อมูลที่ใช้ในการวิเคราะห์ทั้งสิ้น 9,350 ข้อมูล เพื่อวิเคราะห์หาปัจจัยที่เสี่ยงต่อการเสียชีวิตจากอุบัติเหตุจราจรทางบก ผลการวิเคราะห์ได้ตัวแบบพยากรณ์ที่มีความเหมาะสมในการนำมาทำการวิเคราะห์ข้อมูล คือ เทคนิควิธีการเรียนรู้แบบเบย์อย่างง่าย เนื่องจากผลประเมินประสิทธิภาพของข้อมูลที่เป็น unknow data จำนวน 100 ข้อมูลมีค่าประสิทธิภาพเท่ากับ 75% ปัจจัยสำคัญที่มีผลต่อการเสียชีวิตคือพื้นที่ที่ขับขี่ ช่วงเวลาที่ผู้คนขับขี่ สภาพอากาศที่แปรปรวน ประเภทของยานพาหนะ ลักษณะทาง ลักษณะการเกิดอุบัติเหตุ สิ่งเหล่านี้ล้วนเป็นปัจจัยหลักที่ส่งผลให้เกิดความเสี่ยงต่อการเสียชีวิตได้

จากการวิเคราะห์ข้อมูลจำนวน 9,350 ข้อมูล มีการเกิดอุบัติเหตุที่ส่งผลให้เสียชีวิตทั้งหมด 973 ครั้ง เมื่อนำผลมาจัดอันดับ 3 อันดับแรกแล้วพบว่า ตำบลที่เกิดอุบัติเหตุและเสียชีวิตมากที่สุดอันดับหนึ่งคือตำบล/แขวงบางพระ รองลงมาคือตำบล/แขวงสะพานสูงและบ้านกล้วย ตามลำดับ ช่วงเวลาที่ส่งผลให้เกิดอุบัติเหตุและเสียชีวิตมากที่สุดอันดับหนึ่งคือช่วงเวลากลางวัน รองลงมาเป็นช่วงเวลากลางคืนตามลำดับ สภาพอากาศที่ส่งผลให้เกิดอุบัติเหตุและเสียชีวิตมากที่สุดอันดับหนึ่งคือ อากาศแจ่มใส รองลงมาคือฝนตกและมีหมอก/ควัน/ฝุ่นตามลำดับ ประเภทรถที่เกิดอุบัติเหตุและเสียชีวิตมากที่สุดอันดับหนึ่งคือรถบรรทุก 4 ล้อ รองลงมาคือรถยนต์นั่งส่วนบุคคล/รถยนต์นั่งสาธารณะ และรถจักรยานยนต์ตามลำดับ บริเวณหรือลักษณะทางที่ส่งผลให้เกิดอุบัติเหตุและเสียชีวิตอันดับหนึ่งคือทางตรงที่ไม่มีความลาดชัน รองลงมาคือทางโค้งกว้างที่ไม่มีความลาดชันและทางโค้งกว้างที่ลาดชันตามลำดับ ลักษณะการเกิดอุบัติเหตุที่ส่งผลให้เสียชีวิตอันดับหนึ่งคือพลิกคว่ำ/ตกถนนในทางตรง รองลงมาคือชนท้ายและพลิกคว่ำ/ตกถนนในทางโค้งตามลำดับ ซึ่งสาเหตุส่วนใหญ่มาจากการขับรถเร็วเกินอัตรากำหนด รองลงมาคือคน/รถ/สัตว์ตัดหน้ากระชั้นชิดและหลับในตามลำดับ

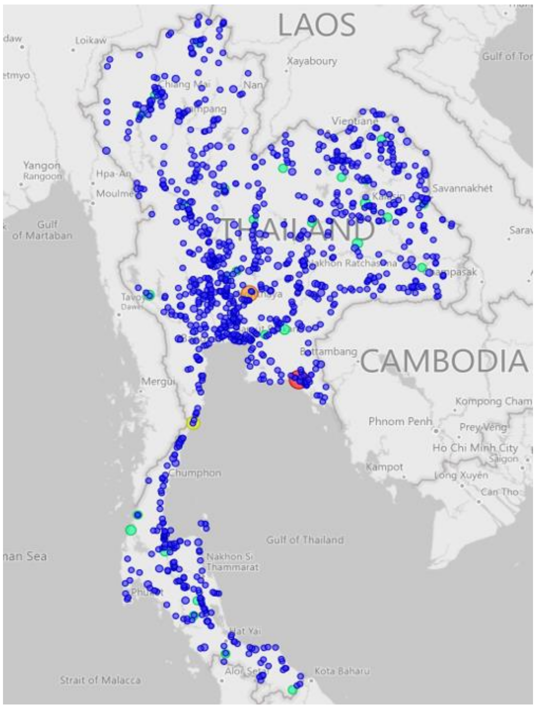
นำข้อมูลมาแสดงผลเป็นแผนที่ด้วยโปรแกรม Microsoft Power BI Desktop แสดงจำนวนผู้เสียชีวิตจากอุบัติเหตุการจราจรทางบกในพื้นที่ทั่วประเทศไทย เป็นการดูพิกัดละติจูด ลองจิจูดในพื้นที่ที่เกิดอุบัติเหตุเพื่อให้มองเห็นภาพรวมของผู้ที่เสียชีวิตในพื้นที่นั้น ๆ โดยแทนจำนวนผู้ที่เสียชีวิตเป็นสีต่างๆ ดังนี้

สีน้ำเงิน มีผู้เสียชีวิต 1 - 2 คน
สีเขียว มีผู้เสียชีวิต 3 - 4 คน
สีเหลือง มีผู้เสียชีวิต 5 - 6 คน

สีส้ม มีผู้เสียชีวิต 7 - 8 คน

สีแดง มีผู้เสียชีวิตมากที่สุด

แสดงให้เห็นว่าแต่ละจุดจะพบผู้เสียชีวิตส่วนใหญ่เป็นจำนวน 1 - 2 ราย อ้างอิงตามจุดสีน้ำเงินที่แทนจำนวนผู้เสียชีวิตในแต่ละครั้ง และมีจุดที่มีผู้เสียชีวิตครั้งละมากๆ ดังนี้ จุดสีเหลืองในครั้งนั้นเกิดอุบัติเหตุจากการขับรถบรรทุก 4 ล้อเร็วเกินอัตรากำหนด ทำให้เสียหลักพุ่งชนสิ่งกีดขวางบนผิวจราจร ส่งผลให้มีผู้เสียชีวิต 6 ราย จุดสีส้มในครั้งนั้นเกิดอุบัติเหตุจากการขับรถโดยสารขนาดใหญ่เร็วเกินอัตรากำหนด ทำให้รถโดยสารพลิกคว่ำตกถนนในทางตรง ส่งผลให้มีผู้เสียชีวิต 8 ราย และจุดสีแดง ในครั้งนั้นเกิดอุบัติเหตุจากการขับรถจักรยานยนต์ชนกับรถบรรทุก 4 ล้อ เนื่องจากคน/รถ/สัตว์ตัดหน้ากระชั้นชิด ส่งผลให้มีผู้เสียชีวิตมากถึง 10 รายในการเกิดอุบัติเหตุเพียงครั้งเดียว ดังรูปที่ 8



รูปที่ 8 แผนที่แสดงจำนวนผู้เสียชีวิตทั่วประเทศไทย

ข้อเสนอแนะ

เพื่อให้ผลการพยากรณ์มีความถูกต้องแม่นยำ มากขึ้น ผู้วิจัยควรพิจารณาปัจจัยแวดล้อมอื่นๆ ที่มีผลต่อความเสี่ยงของการเกิด อุบัติเหตุเพิ่มขึ้น เช่น ปัจจัยจากตัวผู้ขับขี่ การดื่มแอลกอฮอล์/พักผ่อนไม่เพียงพอ/ช่วงอายุของผู้ขับขี่/การประมาท เป็นต้น และข้อมูลที่นำมาใช้ในการพยากรณ์

ควรเป็นช่วงข้อมูลจากการจราจรปกติไม่ควรนำช่วงเวลาเทศกาลที่สำคัญ เช่น ปีใหม่ สงกรานต์มาร่วมพยากรณ์ด้วย เนื่องจากช่วงเวลาดังกล่าว มีสภาพการจราจรที่ไม่เป็นปกติ มีการเกิดอุบัติเหตุจราจรมากเป็นพิเศษ จึงอาจทำให้ผลการพยากรณ์คลาดเคลื่อนได้

เอกสารอ้างอิง

นคร ละลอกน้ำ. (2562). การใช้เทคนิคดาต้าไมน์นิงเพื่อการศึกษา. **วารสารการศึกษาและการพัฒนาสังคม**, 14(2), หน้า 8 - 9.

ภัทธิรา สุวรรณโค, นิศาชล จำนงศรี และจิติมนต์ อั่งสกุล. (2560). แบบจำลองการพยากรณ์ความเสี่ยงในการเกิดอุบัติเหตุทางถนนในเทศกาลปีใหม่ด้วยการทำเหมืองข้อมูล. **วารสารวิทยาการและเทคโนโลยีสารสนเทศ,** 7(2), หน้า 10 - 19.

ศูนย์เทคโนโลยีสารสนเทศและการสื่อสาร สำนักงานปลัดกระทรวงคมนาคม. (2565). **อุบัติเหตุบนโครงข่ายถนนของกระทรวงคมนาคม**. ค้นเมื่อ 23 ตุลาคม 2565, จากhttps://datagov.mot.go.th/dataset/roadaccident

ศราวุธ แดงมาก. (2560). โลกแห่งดิจิทัล ยุคแห่งข้อมูลข่าวสารความมหาศาลของข้อมูลทางการแพทย์สู่การค้นพบความรู้ที่มีค่าด้วยการทำเหมืองข้อมูล. **วารสารเวชบันทึกศิริราช**, 10(1), หน้า 30 - 32.

ฮิชาม อาแว, นิรุสณีย์ อากาจิ, มูฮัมหมัดไซด์ ซา, มามะเพาซี สือแม, ไฮย์ทรง นาวา และวนิดา อาแว. (2565). การพัฒนารูปแบบการ ป้องกันอุบัติเหตุโดยมีชุมชนเป็นฐานแบบบูรณาการเชิงพื้นที่: บริบทพื้นที่จังหวัดนราธิวาส. **วารสารมหาวิทยาลัยนราธิวาสราชนครินทร์**. 14(3). 282 - 299

Gupta, M., Solanki, V. K., Singh, V. K., & García-Díaz, V. (2018). 4Data mining approach of accident occurrences identification with effective methodology and implementation. **International Journal of Electrical and Computer Engineering**, 8(5), pp. 4033 - 4040.

Husin, W. Z. W., Afdzal, A. S., Azmi, N. L. H., & Hamadi, S. A. T. S. (2021) Box-Jenkins and State Space Model in Forecasting Malaysia Road Accident Cases. **Journal of Physics: Conference Series**, 2084, pp. 1 - 14.