

การวิเคราะห์ข้อมูลผู้บาดเจ็บและเสียชีวิตจากอุบัติเหตุในช่วงเทศกาลสงกรานต์ โดยใช้เทคนิคการทำเหมืองข้อมูล

วิไลลักษณ์ ตรีพิช^{1*}

บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์เพื่อการวิเคราะห์ข้อมูลผู้บาดเจ็บและเสียชีวิตจากอุบัติเหตุในช่วงเทศกาลสงกรานต์โดยใช้เทคนิคการทำเหมืองข้อมูล โดยใช้ข้อมูลผู้บาดเจ็บและเสียชีวิตในเทศกาลสงกรานต์ ประจำปี 2551 - 2557 จำแนกตามโรงพยาบาลทั่วประเทศ จากเว็บไซต์ data.go.th ซึ่งผู้วิจัยได้ทำการเตรียมข้อมูลให้เหมาะสมสำหรับการจำแนกข้อมูลโดยตัดข้อมูลที่ไม่เกี่ยวข้องออกไปและปรับให้จำนวนข้อมูลของคำตอบหรือผลลัพธ์มีสัดส่วนใกล้เคียงกัน จากนั้นทำการจำแนกข้อมูลโดยกำหนดให้แอททริบิวต์ผลการรักษาเป็นคำตอบที่ต้องการพยากรณ์ สำหรับอัลกอริทึมที่ผู้วิจัยเลือกใช้ในการสร้างโมเดล ได้แก่ Naive Bayes, Generalized Linear Model (GLM), Logistic Regression, Decision Tree, Random Forest, และ Gradient Boosted Trees (XGBoost) ซึ่งผลการวิจัยพบว่า แอททริบิวต์ที่มีความสัมพันธ์กันมากที่สุดคือ เพศชายมีความสัมพันธ์กับการดื่มสุรา และพบว่าการไม่ใส่หมวกกันน็อกมีความสำคัญสูงที่สุดในการจำแนกข้อมูล สำหรับประสิทธิภาพของอัลกอริทึมที่ใช้ในการสร้างโมเดล พบว่า Naive Bayes มีประสิทธิภาพสูงที่สุด โดยเมื่อพิจารณาจากค่าความถูกต้อง มีค่าถึง 77.4% นอกจากนี้ยังพบว่าอุบัติเหตุส่วนใหญ่มักเกิดบนถนนชนบ ผู้ขับขี่เป็นเพศชาย ซึ่งไม่ใส่หมวกกันน็อก ไม่มีคู่มือ/ล้มเอง มักเกิดเหตุในวันที่ 13 เมษายน สำหรับมาตรการที่จะป้องกันการเสียชีวิตจากอุบัติเหตุควรมุ่งเป้าไปที่การรณรงค์ให้ใส่หมวกกันน็อก โดยเน้นบริเวณถนนทางหลวงและถนนในเมือง เนื่องจากหากไม่ใส่หมวกกันน็อกแล้วเกิดอุบัติเหตุบนถนนทางหลวงหรือถนนในเมือง ผลการวิจัยผู้ขับขี่มีโอกาสเสียชีวิตสูง

คำสำคัญ : การทำเหมืองข้อมูล การจำแนกข้อมูล เทศกาลสงกรานต์ ผู้บาดเจ็บและเสียชีวิต

¹ วิทยาลัยนวัตกรรมการจัดการและเทคโนโลยีสารสนเทศ มหาวิทยาลัยรังสิต จังหวัดปทุมธานี

* ผู้นิพนธ์หลัก e-mail : wilailak.t@rsu.ac.th

THE ANALYSIS OF INJURED AND DEAD PEOPLE DATA FROM ACCIDENT IN SONGKRAN FESTIVAL BY USING DATA MINING TECHNIQUES

Wilailak Treepuech^{1*}

Abstract

The objectives of this research is to analysis of injured and dead people data from accident in Songkran Festival by using data mining techniques. The injured and dead people data in Songkran festival from 2008 to 2014 is classified by hospital from data.go.th is used in this research. The researcher prepared the data for classification by eliminating the unrelated data and adjusting the number of answer or result to be similarly proportional. The classification of data use the treatment results attribute is set as the desired answer. The algorithms used for create the model were Naive Bayes, Generalized Linear Model (GLM), Logistic Regression, Decision Tree, Random Forest, and Gradient Boosted Trees (XGBoost). The results showed that the most relevant attributes are male have a relationship with drinking alcohol and It was found that not wearing a helmet is the highest importance in classification. For the efficiency of the algorithms used to create model, Naive Bayes has the highest efficiency when considering on the accuracy, it has 77.4% of accuracy. Moreover, it is also found that most accidents occur on rural roads, the driver is male, not wear helmet, no parties/fall down by themselves, and usually occurs on day 13 of April. Measures to prevent death from accident should be aimed at wearing a helmet, emphasize the area at highways and city streets. Due to, if not wear a helmet, then an accident occurs on highways or city streets, based on the research results, the driver has a high probability of death.

Keywords : Data Mining, Classification, Songkran Festival, Injured and Dead People

¹College of Digital Innovation and Information Technology, Rangsit University,

* Corresponding author, e-mail: wilailak.t@rsu.ac.th

บทนำ

อุบัติเหตุบนท้องถนน เป็นสาเหตุหนึ่งที่ทำให้มีผู้เสียชีวิตจำนวนมาก และเป็นปัญหาหนึ่งของประเทศไทย ซึ่งอุบัติเหตุบนท้องถนนก่อให้เกิดความสูญเสียกับประเทศไทยมากถึงกว่า 200,000 ล้านบาทต่อปี และมีผู้เสียชีวิตมากกว่า 11,000 คน นอกจากนั้นยังมียูบัติเจ็บและพิการอีกเป็นจำนวนมาก ซึ่งถือว่าเป็นความสูญเสียกับประเทศไทยเป็นอย่างมาก (กรมป้องกันและบรรเทาสาธารณภัย กระทรวงมหาดไทย, 2554) ซึ่งหน่วยงานต่างๆ ที่เกี่ยวข้องให้ความสนใจและพยายามที่จะลดอุบัติเหตุบนท้องถนนให้น้อยลง โดยเฉพาะในช่วงเทศกาลปีใหม่และเทศกาลสงกรานต์ซึ่งเป็นช่วงที่เกิดอุบัติเหตุบนท้องถนนเป็นจำนวนมากและส่งผลให้มีผู้เสียชีวิตเป็นจำนวนมากด้วยเช่นกัน เนื่องจากเป็นช่วงวันหยุดยาว ประชาชนส่วนใหญ่เดินทางไปต่างจังหวัดเป็นจำนวนมาก โดยวัตถุประสงค์ในการเดินทางหลักๆ คือ ไปเที่ยวพักผ่อนกับครอบครัว และเดินทางกลับบ้านที่ต่างจังหวัด ทำให้มีปริมาณรถยนต์บนท้องถนนเป็นจำนวนมาก ผู้ขับขี่บางส่วนพักผ่อนไม่เพียงพอ เมื่อต้องเดินทางในระยะทางที่ไกลอาจส่งผลให้เกิดอุบัติเหตุได้ง่ายขึ้น นอกจากการเดินทางดังกล่าวแล้ว ในช่วงเทศกาลจะมีการเลี้ยงสังสรรค์กัน ไม่ว่าจะเป็นภายในเครือญาติหรือระหว่างเพื่อนฝูง ซึ่งมักจะมีการดื่มเครื่องดื่มที่มีส่วนผสมของแอลกอฮอล์ด้วย ดังนั้นเมื่อมีการเดินทางเกิดขึ้นแม้ไม่ได้เป็นการเดินทางไกล ก็มีโอกาสดังกล่าวเกิดอุบัติเหตุบนท้องถนนได้ โดยเฉพาะผู้ขับขี่รถจักรยานยนต์ ทำให้หน่วยงานที่เกี่ยวข้องเรียกช่วงเทศกาลปีใหม่และเทศกาลสงกรานต์ว่า ช่วง 7 วันอันตราย โดยสรุปสถิติอุบัติเหตุตั้งแต่วันที่ 11- 17 เมษายน 2561 ซึ่งเป็นช่วงเทศกาลสงกรานต์ เกิดอุบัติเหตุรวม 3,724 ครั้ง เสียชีวิต 418 คน บาดเจ็บ 3,897 คน ซึ่งมีจำนวนผู้เสียชีวิตสูงชันมากกว่าปี 2560 โดยสาเหตุที่ทำให้เกิดอุบัติเหตุสูงสุด คือ ดื่มแล้วขับ ร้อยละ 40.28 ส่วนยานพาหนะที่เกิดอุบัติเหตุสูงสุด มาจากรถจักรยานยนต์ ร้อยละ 79.85 (ชาวไทยพีบีเอส, 2561) ส่วนเทศกาลปีใหม่ในปี 2561 นั้น มีจำนวนผู้เสียชีวิตน้อยกว่าในปี 2560 (กรมป้องกันและบรรเทาสาธารณภัย กระทรวงมหาดไทย, รายงานผู้เสียชีวิต 7 วันอันตราย เทศกาลปีใหม่, 2561)

จากรายงานผู้เสียชีวิต 7 วันอันตราย เทศกาลสงกรานต์ (กรมป้องกันและบรรเทาสาธารณภัย กระทรวงมหาดไทย, รายงานผู้เสียชีวิต 7 วันอันตราย เทศกาลสงกรานต์, 2561) มีการสรุปข้อมูลในรูปแบบของแผนภูมิวงกลมและกราฟเส้น ซึ่งแสดงข้อมูลสรุปเบื้องต้น ได้แก่ การเสียชีวิตจำแนกตามเพศ การเสียชีวิตจำแนกตามวัน การเสียชีวิตจำแนกตามเวลา การเสียชีวิตจำแนกตามอายุ จำนวนผู้เสียชีวิตเปรียบเทียบระหว่างปี 2560 และปี 2561 ซึ่งการสรุปข้อมูลในลักษณะนี้ จะสามารถวิเคราะห์ข้อมูลเพื่อนำมาสู่การกำหนดมาตรการป้องกันการเกิดอุบัติเหตุได้เพียงระดับหนึ่ง อย่างไรก็ตามผู้วิจัยพบว่า สถาบันการแพทย์ฉุกเฉินแห่งชาติได้มีการเผยแพร่ข้อมูลผู้บาดเจ็บและเสียชีวิตในเทศกาลสงกรานต์ประจำปี 2551 - 2557 จำแนกตามโรงพยาบาลทั่วประเทศผ่านเว็บไซต์ data.go.th โดยสำนักงานพัฒนารัฐบาลดิจิทัล (องค์การมหาชน) (สถาบันการแพทย์ฉุกเฉินแห่งชาติ, 2558) ซึ่งเป็นข้อมูลแบบมีโครงสร้าง (Structured Data) ในรูปแบบที่ผู้วิจัยสามารถนำมาวิเคราะห์ข้อมูลได้ง่าย โดยผู้วิจัยจะใช้เทคนิคการทำเหมืองข้อมูล (Data Mining) ซึ่งสามารถใช้ในการจำแนกประเภทข้อมูล (Classification) ได้ โดยการจำแนกประเภทข้อมูลเป็นการเรียนรู้แบบมีผู้สอน (Supervised Learning) โดยนำข้อมูลที่มีในอดีตมาวิเคราะห์โดยและสร้างโมเดล (Model) เพื่อทำนายอนาคต นอกจากนั้นเทคนิคการทำเหมืองข้อมูลยังสามารถใช้ในการวิเคราะห์เพื่อหาความสัมพันธ์ (Association) โดยหาความสัมพันธ์ของข้อมูลที่เกิดขึ้นร่วมกันบ่อยๆ นอกจากนั้นยังสามารถใช้ในการวิเคราะห์เพื่อจัดกลุ่มข้อมูล (Clustering) ซึ่งจะวิเคราะห์ข้อมูลที่มีลักษณะคล้ายๆ กันเพื่อทำการแบ่งเป็นกลุ่มได้ (เอกสิทธิ์ พัทธวงศ์ศักดิ์, 2560)

ในงานวิจัยนี้ผู้วิจัยจะนำข้อมูลผู้บาดเจ็บและเสียชีวิตในเทศกาลสงกรานต์ประจำปี 2551 - 2557 จำแนกตามโรงพยาบาลทั่วประเทศ มาทำการวิเคราะห์ข้อมูล โดยใช้เทคนิคการทำเหมืองข้อมูล เพื่อจำแนกประเภทข้อมูลและหาโมเดลที่สามารถวิเคราะห์ตัวแปรที่เกี่ยวข้องของอุบัติเหตุซึ่งนำมาสู่การเสียชีวิตได้ ซึ่งผู้วิจัยคาดว่าผลการวิจัยจะเป็นประโยชน์กับหน่วยงานที่เกี่ยวข้อง สามารถนำไปใช้ในการกำหนดนโยบายหรือมาตรการในการควบคุมอุบัติเหตุและลดจำนวนผู้เสียชีวิตในเทศกาลสงกรานต์ได้

วัตถุประสงค์ของการวิจัย

เพื่อการวิเคราะห์ข้อมูลผู้บาดเจ็บและเสียชีวิตจากอุบัติเหตุในช่วงเทศกาลสงกรานต์โดยใช้เทคนิคการทำเหมืองข้อมูล

วิธีดำเนินการวิจัย

1. ขอบเขตของการวิจัย

1.1 ข้อมูลที่นำมาวิเคราะห์เป็นข้อมูลผู้บาดเจ็บและเสียชีวิตในเทศกาลสงกรานต์ประจำปี 2551 - 2557 จำแนกตามโรงพยาบาลทั่วประเทศ

1.2 ผู้วิจัยเลือกใช้ซอฟต์แวร์ Rapid Miner มาใช้ในการวิเคราะห์ข้อมูล โดยใช้อัลกอริทึม (Algorithm) ได้แก่ Naive Bayes, Generalized Linear Model (GLM), Logistic Regression, Decision Tree, Random Forest, และ Gradient Boosted Trees (XGBoost) มาใช้ในการสร้างโมเดลเพื่อใช้ในการจำแนกประเภทข้อมูล

2. ขั้นตอนการวิจัย

2.1 การเตรียมข้อมูล (Data Preparation) ผู้วิจัยนำข้อมูลผู้บาดเจ็บและเสียชีวิตในเทศกาลสงกรานต์ประจำปี 2551 - 2557 จำแนกตามโรงพยาบาลทั่วประเทศ ที่ดาวน์โหลดมาจากเว็บไซต์ data.go.th มาปรับเปลี่ยนข้อมูลให้อยู่ในรูปแบบที่เหมาะสม โดยข้อมูลที่ได้นั้นมีข้อมูลทั้งหมด 202,299 รายการ ประกอบด้วย 19 แอททริบิวต์ (Attribute) ได้แก่ ชื่อเทศกาล รหัสจังหวัด จังหวัด รหัส รพ. ชื่อโรงพยาบาลที่รับผู้บาดเจ็บ วันที่เกิดเหตุ เวลาเกิดเหตุ เพศ อายุ ถนนที่เกิดเหตุ สถานะ รถผู้บาดเจ็บ รถคู่กรณี มาตรการ การดื่มสุรา การนำส่ง Refer-Admit ผลการรักษา และจำนวนวันรักษา ซึ่งมีแอททริบิวต์ที่ไม่มีความสำคัญในการวิเคราะห์ข้อมูลอยู่หลายแอททริบิวต์ และข้อมูลบางรายการอาจจะผิดพลาดรวมไปถึงเป็นข้อมูลที่ผู้วิจัยพิจารณาแล้วว่าไม่เกี่ยวข้องกับการวิจัย โดยดำเนินการกับข้อมูล ดังนี้

2.1.1 ผู้วิจัยลบบางแอททริบิวต์ที่ไม่มีความสำคัญในการวิเคราะห์ข้อมูล ได้แก่ ชื่อเทศกาล รหัสจังหวัด รหัส รพ. ชื่อโรงพยาบาล จำนวนวันรักษา การนำส่ง และ Refer-Admit

2.1.2 ทำการลบข้อมูลที่ไม่ชัดเจนทั้ง ได้แก่ ไม่ทราบเวลาเกิดเหตุ อายุ 0 ปี เนื่องจากมีข้อมูลที่ขัดแย้งเช่น อายุ 0 ปี แต่เป็นผู้ขับขี่ อายุ 1 ปี ข้อมูลเป็นผู้ขับขี่ ไม่ทราบถนนที่เกิดเหตุ ไม่ทราบสถานะผู้โดยสาร ไม่ทราบรถผู้บาดเจ็บ ไม่ทราบรถคู่กรณี ไม่ทราบมาตรการ และไม่ทราบว่าดื่มสุรา

2.1.3 เลือกข้อมูลเฉพาะข้อมูลที่สนใจที่จะนำไปวิเคราะห์ข้อมูล ได้แก่ รถผู้บาดเจ็บ เป็นรถจักรยานยนต์ เท่านั้น เนื่องจากข้อมูลการเกิดอุบัติเหตุส่วนใหญ่ และผู้ขับขี่เสียชีวิตประมาณ 90% เป็นรถจักรยานยนต์ เลือกสถานะเป็นผู้ขับขี่เท่านั้น เนื่องจากข้อมูลผู้เสียชีวิตส่วนใหญ่เป็นผู้ขับขี่ ซึ่งเมื่อเลือกข้อมูลของแอททริบิวต์นี้เฉพาะข้อมูลที่สนใจ จึงทำให้ข้อมูลในแอททริบิวต์ดังกล่าวมีเพียงรูปแบบเดียว ผู้วิจัยจึงทำการลบแอททริบิวต์ รถผู้บาดเจ็บและสถานะ ออกไป

2.1.4 รวมกลุ่มของข้อมูลในแอททริบิวต์ ผลการรักษา เนื่องจากปกติแล้ว จะมีข้อมูลในแอททริบิวต์นี้หลากหลาย ได้แก่ ตายในตึกภายใน 24 ชม. หลังเหตุ ตายในตึกหลัง 24 ชม. - 30 วัน ตายที่เกิดเหตุ ตายที่ห้องฉุกเฉิน ตายระหว่างนำส่ง ตายระหว่างส่งต่อ ทูเลา/หาย ซึ่งผู้วิจัยจะรวบรวมกลุ่มของการตายในรูปแบบต่างๆ ให้เหลือ ตาย เพียงอย่างเดียว

2.1.5 ทำให้ข้อมูลในแอททริบิวต์ ผลการรักษา มีความใกล้เคียงกัน เนื่องจากข้อมูลทั้งหมดมีข้อมูลผลการรักษา ส่วนใหญ่ คือ ตาย ซึ่งแอททริบิวต์ ผลการรักษา เป็นแอททริบิวต์ที่ผู้วิจัยต้องการใช้ในการทำนาย หากมีจำนวนข้อมูลที่แตกต่างกันมากจนเกินไป คือข้อมูลไม่สมดุลกันระหว่าง ตาย กับ ทูเลา/หาย จะส่งผลให้การจำแนกประเภทข้อมูลไม่มีประสิทธิภาพ จึงควรมีกระบวนการที่จะทำให้ข้อมูลเกิดความสมดุลกัน (Chawla, 2005) (จรัสศรี รุ่งรัตนอุบล, อนามัย นาอุดม, และ ดารณี สุวพันธ์, 2555) โดยผู้วิจัยได้ดำเนินการกับข้อมูลตามลำดับคือ เลือกข้อมูลเฉพาะรายการที่ ผลการรักษา คือ ตาย จากนั้นเลือกข้อมูลที่ละจังหวัด แล้วนับจำนวนรายการ จากนั้นสุ่มเลือกข้อมูลจากจังหวัดนั้น ที่ผลการรักษา คือ ทูเลา/หาย ตามจำนวนรายการ จากขั้นตอนแรก โดยพยายามสุ่มเลือกจากรายการที่มีการดื่มสุรา ครั้งหนึ่ง แต่ไม่ดื่มสุรา ครั้งหนึ่ง ซึ่งจะทำให้ข้อมูลในแอททริบิวต์ ผลการรักษา มีความสมดุลกัน

ตัวแปรที่ใช้ในการวิจัยซึ่งมีแอททริบิวต์ทั้งหมด 10 แอททริบิวต์ ซึ่งโดยส่วนใหญ่แล้วจะข้อมูลในแอททริบิวต์มีรูปแบบเป็นแบบประเภท (Category) มีเพียงแอททริบิวต์อายุเท่านั้นที่มีรูปแบบเป็นตัวเลข (Number) ส่วนแอททริบิวต์ที่มีความหลากหลายของข้อมูลซึ่งอาจจะไม่ส่งผลต่อการวิเคราะห์ข้อมูล คือ แอททริบิวต์จังหวัด แต่ผู้วิจัยยังเห็นว่าแอททริบิวต์นี้อาจจะมีความเกี่ยวข้องในการวิเคราะห์ข้อมูลจึงไม่ได้ตัดทิ้งไป โดยตัวแปรที่เกี่ยวข้องในการวิเคราะห์ข้อมูลนำเสนอตารางที่ 1 ซึ่งผู้วิจัยตั้งใจที่จะใช้แอททริบิวต์ ผลการรักษา เป็นข้อมูลหลักที่ต้องการพยากรณ์ (Predict) โดยข้อมูลทั้งหมดหลังจากขั้นตอนการเตรียมข้อมูลแล้วมีข้อมูลเหลือทั้งสิ้น 1,566 รายการ

ตารางที่ 1 ตารางแสดงตัวแปรที่ใช้ในการวิจัย

ลำดับ	แอททริบิวต์	ข้อมูลในแอททริบิวต์	จำนวนรูปแบบข้อมูลในแอททริบิวต์
1	จังหวัด	จังหวัดในประเทศไทย 77 จังหวัด	77
2	วันที่เกิดเหตุ	วันที่ 10 ถึงวันที่ 18	9
3	เวลาเกิดเหตุ	ตลอดทั้งวันเป็นช่วงละ 1 ชั่วโมง	24
4	เพศ	เพศชายและเพศหญิง	2
5	อายุ	ตั้งแต่ช่วง 10 ปี ถึง 88 ปี	ข้อมูลแบบตัวเลขในช่วงที่กำหนด
6	ถนนที่เกิดเหตุ	ในเมือง, ขนบพ, ทางหลวง	3
7	รถคู่กรณี	ไม่มี/ล้มเอง, จักรยานยนต์, ปิคอัพ, รถเก๋ง/แท็กซี่, รถโดยสาร 4 ล้อ, รถโดยสารใหญ่, รถจักรยาน, รถตู้, รถบรรทุก, สามล้อเครื่อง, สามล้อถีบ, อื่นๆ	12

ตารางที่ 1 ตารางแสดงตัวแปรที่ใช้ในการวิจัย (ต่อ)

ลำดับ	แอททริบิวต์	ข้อมูลในแอททริบิวต์	จำนวนรูปแบบข้อมูลในแอททริบิวต์
8	มาตรการ	ใส่หมวกกันน็อค, ไม่ใส่หมวกกันน็อค	2
9	การดื่มสุรา	ดื่มสุรา, ไม่ดื่มสุรา	2
10	ผลการรักษา	ทุเลา/หาย, ตาย	2

2.2 การจำแนกข้อมูล ผู้วิจัยใช้ RapidMiner Studio Educational 9 เป็นซอฟต์แวร์ในการวิเคราะห์ข้อมูล โดยเลือกข้อมูลผ่านขั้นตอนการเตรียมข้อมูลแล้วมาใช้ โดยเลือกแอททริบิวต์ ผลการรักษา เป็นแอททริบิวต์ที่ต้องการพยากรณ์ จากนั้นเลือกแอททริบิวต์ที่เหลืออีก 9 แอททริบิวต์ เป็นข้อมูลที่นำมาใช้ในการวิเคราะห์ ซึ่งสิ่งที่ผู้วิจัยต้องการให้จากการวิเคราะห์ข้อมูลเบื้องต้น คือ ความสัมพันธ์ระหว่างแอททริบิวต์และความสำคัญของแต่ละแอททริบิวต์ สำหรับอัลกอริทึมที่ผู้วิจัยใช้ในการสร้างโมเดลเพื่อพยากรณ์ ได้แก่ Naive Bayes, Generalized Linear Model (GLM), Logistic Regression, Decision Tree, Random Forest, และ Gradient Boosted Trees (XGBoost) ซึ่งผู้วิจัยจะเปรียบเทียบประสิทธิภาพของโมเดลที่สร้างขึ้นจากอัลกอริทึมต่างๆ โดยใช้ค่าความถูกต้อง (Accuracy) เป็นหลัก ส่วนค่าอื่นๆ เช่น ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าความเหวี่ยง (F-Measure) โดยแต่ละค่าสามารถอธิบายถึงประสิทธิภาพของแต่ละอัลกอริทึมในมุมมองที่ต่างกัน (ภรณ์ยา ปาลวิสุทธิ และ มงคล รอดจันทร์, 2561) ซึ่งผู้วิจัยจะนำมาใช้เพื่อพิจารณาด้วย จากนั้นจึงค่อยๆ พิจารณาว่าโมเดลที่แต่ละอัลกอริทึมสร้างมานั้นสามารถจำแนกข้อมูลหรืออธิบายอะไรเกี่ยวกับการพยากรณ์ผลการรักษาได้บ้าง

ผลการวิจัย

จากการวิเคราะห์ข้อมูลโดยใช้เทคนิคการทำเหมืองข้อมูล ผู้วิจัยขอเสนอผลการวิจัย ดังนี้

1. ความสัมพันธ์ระหว่างแอททริบิวต์ จากผลการวิจัยพบว่าแอททริบิวต์ที่มีความสัมพันธ์กันมากที่สุดคือ เพศชายมีความสัมพันธ์กับการดื่มสุรา สำหรับความสัมพันธ์ระหว่างแอททริบิวต์อื่นๆ ที่น่าสนใจแสดงดังตารางที่ 2

ตารางที่ 2 ตารางแสดงความสัมพันธ์ระหว่างแอททริบิวต์

ลำดับ	แอททริบิวต์ที่สัมพันธ์กัน	ค่าความสัมพันธ์
1	เพศชายกับการดื่มสุรา	0.406
2	การไม่ใส่หมวกกันน็อคกับการดื่มสุรา	0.116
3	วันเกิดเหตุวันที่ 13 กับการดื่มสุรา	0.108
4	ถนนที่เกิดเหตุเป็นถนนชนบทกับการดื่มสุรา	0.090
5	ถนนที่เกิดเหตุเป็นถนนชนบทกับผลการรักษาที่ทุเลา/หาย	0.083
6	การไม่ใส่หมวกกันน็อคกับเพศชาย	0.059
7	ถนนที่เกิดเหตุเป็นถนนชนบทกับการไม่ใส่หมวกกันน็อค	0.057

2. ความสำคัญของแต่ละแอททริบิวต์ที่มีผลต่อการวิเคราะห์ข้อมูล จากผลการวิจัยพบว่าการไม่ใส่หมวดกันนี้ออกมีความสำคัญสูงสุด สำหรับความสำคัญของแต่ละแอททริบิวต์ที่มีผลต่อการวิเคราะห์ข้อมูล แสดงดังตารางที่ 3

ตารางที่ 3 ตารางแสดงความสำคัญของแต่ละแอททริบิวต์ที่มีผลต่อการวิเคราะห์ข้อมูล

ลำดับ	แอททริบิวต์	ค่าน้ำหนัก
1	การไม่ใส่หมวดกันนี้ออก	1
2	ถนนที่เกิดเหตุเป็นทางหลวง	0.347
3	อายุ	0.325
4	การดื่มสุรา	0.296
5	เพศชาย	0.295

3. อัลกอริทึมที่ใช้ในการสร้างโมเดลเพื่อพยากรณ์ จากผลการวิจัยพบว่า Naive Bayes มีความถูกต้องสูงกว่าอัลกอริทึมอื่นๆ โดยมีค่าความถูกต้องเท่ากับ 77.4% ส่วนอัลกอริทึมอื่นๆ จะมีค่าความถูกต้องตั้งแต่ 70% ถึง 75% นอกจากนั้นอัลกอริทึม Naive Bayes สามารถให้ค่าความแม่นยำและค่าความเหวี่ยงได้สูงที่สุดเช่นกัน ยกเว้นค่าความระลึกที่อัลกอริทึม Decision Tree สามารถให้ค่าได้สูงที่สุด สำหรับโมเดลที่ Naive Bayes สร้างขึ้นนั้นสามารถพยากรณ์ผลการรักษาตามแอททริบิวต์ได้ดังตารางที่ 4

ตารางที่ 4 ตารางแสดงข้อมูลของโมเดลที่ได้จากอัลกอริทึม Naive Bayes

แอททริบิวต์	รายละเอียด
การดื่มสุรา	1) กรณีดื่มสุรามีโอกาสที่ผลการรักษาจะเป็นตายมากกว่าทุเลา/หาย 2) กรณีไม่ดื่มสุรามีโอกาสที่ผลการรักษาจะเป็นทุเลา/หายมากกว่าตาย
ถนนที่เกิดเหตุ	1) กรณีเป็นถนนชนบตโอกาสที่ผลการรักษาจะเป็นทุเลา/หายมากกว่าตาย 2) กรณีเป็นถนนทางหลวงโอกาสที่ผลการรักษาจะเป็นตายมากกว่าทุเลา/หาย 3) กรณีเป็นถนนในเมืองโอกาสที่ผลการรักษาจะเป็นตายมากกว่าทุเลา/หาย
มาตรการ	1) กรณีใส่หมวดกันนี้ออกโอกาสที่ผลการรักษาจะเป็นทุเลา/หายมากกว่าตายอย่างชัดเจน 2) กรณีไม่ใส่หมวดกันนี้ออกโอกาสที่ผลการรักษาจะเป็นตายมากกว่าทุเลา/หายอย่างชัดเจน
รถคู่กรณี	1) กรณีเป็นรถปิคอัพโอกาสที่ผลการรักษาจะเป็นตายมากกว่าทุเลา/หาย 2) กรณีไม่มีคู่กรณีหรือล้มเองโอกาสที่ผลการรักษาจะเป็นทุเลา/หายมากกว่าตาย 3) กรณีอื่นๆ ผลการรักษาค่อนข้างใกล้เคียงกัน
จังหวัด	1) ทุกกรณีผลการรักษาค่อนข้างใกล้เคียงกัน
วันที่เกิดเหตุ	1) ทุกกรณีผลการรักษาค่อนข้างใกล้เคียงกัน
เพศ	1) ทุกกรณีผลการรักษาค่อนข้างใกล้เคียงกัน
อายุ	1) ช่วงอายุต่ำกว่า 40 ปี โอกาสที่ผลการรักษาจะเป็นทุเลา/หายมากกว่าตาย 2) ช่วงอายุมากกว่าหรือเท่ากับ 40 ปี โอกาสที่ผลการรักษาจะเป็นตายมากกว่าทุเลา/หาย

นอกจากข้อมูลจากตารางที่ 4 ผู้วิจัยยังมีการวิจัยอื่นๆ ที่เห็นได้ชัดเจน จากโมเดลที่ Naive Bayes สร้างขึ้น เช่น อุบัติเหตุส่วนใหญ่มักเกิดบนถนนชนบท ผู้ขับขี่เป็นเพศชาย ซึ่งไม่ใช่หมวกกันน็อก ไม่มีคู่มือ/สัมภาระ มักเกิดเหตุในวันที่ 13 สำหรับโมเดลจากอัลกอริทึมอื่นๆ มีผลการวิจัยอย่างหนึ่งที่ชัดเจนและตรงกัน คือ มาตรการในการใส่หมวกกันน็อกหรือไม่ใส่หมวกกันน็อก เป็นแอททริบิวต์ที่สำคัญในการพิจารณาว่า ผลการรักษจะเป็นอย่างไร กล่าวคือหากไม่ใส่หมวกกันน็อกผลการรักษาจะเป็นตายมากกว่าทุเลา/หายอย่างชัดเจน

สรุปผลและอภิปรายผล

จากการวิเคราะห์ข้อมูลผู้บาดเจ็บและเสียชีวิตจากอุบัติเหตุในช่วงเทศกาลสงกรานต์โดยใช้เทคนิคการทำเหมืองข้อมูล พบว่าหลายแอททริบิวต์มีความสัมพันธ์กับการดื่มสุรา เช่น เพศชาย การไม่ใส่หมวกกันน็อก วันเกิดเหตุวันที่ 13 ถนนที่เกิดเหตุเป็นถนนชนบท ซึ่งจากข้อมูลดังกล่าว จึงควรมีมาตรการเกี่ยวกับการเฝ้าระวังหรือรณรงค์ไม่ให้ดื่มสุราแล้วขับขี่ยานยนต์ โดยเฉพาะในช่วงวันที่ 12 ถึงวันที่ 14 และเน้นในบริเวณถนนชนบทเป็นหลัก เพราะจากผลการวิจัยจะช่วยลดอุบัติเหตุได้ แต่อย่างไรก็ตามการเกิดเหตุในถนนชนบทส่วนใหญ่จะผู้ขับขี่จะไม่มีคู่มือ/สัมภาระ ซึ่งอาจเป็นเพราะการดื่มสุรา แต่ผู้ขับขี่ที่ดื่มสุรามักไม่ใส่หมวกกันน็อกซึ่งหากเกิดอุบัติเหตุอาจทำให้บาดเจ็บรุนแรงได้ สำหรับมาตรการที่จะป้องกันการเสียชีวิตควรมุ่งเป้าไปที่การรณรงค์ให้ใส่หมวกกันน็อก โดยเน้นบริเวณถนนทางหลวงและถนนในเมือง เนื่องจากหากไม่ใส่หมวกกันน็อกแล้วเกิดอุบัติเหตุบนถนนทางหลวงและถนนในเมือง จากผลการวิจัยผู้ขับขี่มีโอกาสเสียชีวิตสูง

สำหรับอัลกอริทึมที่ใช้ในการสร้างโมเดลเพื่อพยากรณ์ ซึ่งงานวิจัยที่เกี่ยวข้องส่วนใหญ่มักใช้อัลกอริทึม Decision Tree เป็นหลัก (ภรณ์ยา ปาลวิสุทธิ และ มงคล รอดจันทร์, 2561) (สุนิดา ศรีสุริยชัย, 2550) แต่อย่างไรก็ตามงานวิจัยที่เกี่ยวข้อง (จรัสศรี รุ่งรัตนอุบล, อนามัย นาอุดม, และ ดารณี สุวพันธ์, 2555) (สุนิดา ศรีสุริยชัย, 2550) ได้เสนอว่าควรนำอัลกอริทึม Naive Bayes มาทำการทดลองด้วย ซึ่งจากผลการวิจัยนี้พบว่า Naive Bayes มีค่าความถูกต้องสูงกว่าอัลกอริทึมอื่นๆ ซึ่งแตกต่างกับงานวิจัย (ภัทธีรา สุวรรณโค, นิศาชล จำนงศรี, และ จิตมินต์ อังสกุล, 2560) ที่ Naive Bayes มีประสิทธิภาพต่ำสุด ซึ่งนักวิจัยได้อธิบายว่า Naive Bayes เหมาะกับการวิเคราะห์ข้อมูลที่มีปริมาณไม่มาก และเหมาะกับข้อมูลที่มีตัวแปรไม่มาก แต่มีข้อดีคือสามารถใช้กับตัวแปรได้ทุกชนิด โดยข้อมูลที่ใช้ในการวิจัยดังกล่าวมีลักษณะไม่ตรงกับความต้องการของ Naive Bayes ซึ่งแตกต่างจากงานวิจัยนี้ที่ข้อมูลมีลักษณะตรงกับความต้องการของ Naive Bayes มากกว่า นอกจากนั้น Naive Bayes มีกระบวนการเรียนรู้จากข้อมูลที่มีอยู่แล้วนำมาสร้างเงื่อนไขในการจำแนกข้อมูลด้วยทฤษฎีความน่าจะเป็นแบบมีเงื่อนไข ซึ่งข้อมูลที่นำมาวิเคราะห์ในงานวิจัยนี้มีลักษณะที่มีเงื่อนไขต่อเนื่องกัน เช่น ผู้ขับขี่เพศชายมักดื่มสุรา เมื่อดื่มสุรามักไม่ใส่หมวกกันน็อก จึงทำให้ในงานวิจัยนี้ Naive Bayes สามารถสร้างโมเดลที่มีค่าความถูกต้องสูงที่สุดได้

ข้อเสนอแนะ

1. ในการสร้างโมเดลเพื่อพยากรณ์ค่าตอบหรือผลลัพธ์นั้น ปัจจัยที่มีความสำคัญซึ่งต้องคำนึงถึง คือจำนวนข้อมูลของค่าตอบหรือผลลัพธ์ต้องมีสัดส่วนใกล้เคียงกัน เพราะหากมีจำนวนข้อมูลแตกต่างกันมากจะทำให้โมเดลไม่สามารถพยากรณ์ค่าตอบหรือผลลัพธ์ได้มีประสิทธิภาพ

2. การวิจัยในครั้งนี้ทำการวิเคราะห์ข้อมูลผู้บาดเจ็บและเสียชีวิตในเทศกาลสงกรานต์เท่านั้น ดังนั้นหากนำข้อมูลผู้บาดเจ็บและเสียชีวิตในเทศกาลปีใหม่มาวิเคราะห์ด้วย น่าจะทำให้พบความเหมือนหรือความแตกต่างเกี่ยวกับการเกิดอุบัติเหตุในช่วงเทศกาลของประเทศไทย

3. เนื่องจากข้อมูลของผู้ขับขี่ในการวิจัยครั้งนี้สนใจเฉพาะผู้ขับขี่รถจักรยานยนต์เท่านั้น หากในอนาคตมีข้อมูลผู้ขับขี่พาหนะประเภทอื่นๆ จำนวนมากพอ สามารถนำข้อมูลมาวิเคราะห์เพื่อพยากรณ์ผลการรักษาได้ ซึ่งอาจพบข้อมูลที่น่าสนใจและนำไปใช้ในการกำหนดมาตรการป้องกันและลดอุบัติเหตุสำหรับพาหนะประเภทอื่นๆ ได้

เอกสารอ้างอิง

- Chawla, N. V. (2005). *Data Mining for Imbalanced Datasets: An Overview*. Data Mining and Knowledge Discovery Handbook. Springer Science+Business Media.
- เอกสิทธิ์ พัทธวงศ์ศักดิ์. (2560). สไลด์การบรรยายเรื่อง Introduction to Data Mining and Big Data Analytics. Retrieved from Data Mining Trend: <http://dataminingtrend.com/2014/>
- กรมป้องกันและบรรเทาสาธารณภัย กระทรวงมหาดไทย. (2554). *แผนที่นำทางเชิงกลยุทธ์ทศวรรษแห่งความปลอดภัยทางถนน พ.ศ.2554-2563*. จากเว็บไซต์ ศูนย์ข้อมูลอุบัติเหตุ เพื่อเสริมสร้างวัฒนธรรมการขับขี่ปลอดภัยทางถนน: http://www.thairsc.com/th/Document/strategic_map_roadsafety.pdf
- กรมป้องกันและบรรเทาสาธารณภัย กระทรวงมหาดไทย. (2561). *รายงานผู้เสียชีวิต 7 วันอันตราย เทศกาลปีใหม่*. จากเว็บไซต์ ศูนย์ข้อมูลอุบัติเหตุ เพื่อเสริมสร้างวัฒนธรรมการขับขี่ปลอดภัยทางถนน: <http://rvpreport.rvpeservice.com/viewrsc.aspx?report=0545&session=16>
- กรมป้องกันและบรรเทาสาธารณภัย กระทรวงมหาดไทย. (2561). *รายงานผู้เสียชีวิต 7 วันอันตราย เทศกาลสงกรานต์*. จากเว็บไซต์ ศูนย์ข้อมูลอุบัติเหตุ เพื่อเสริมสร้างวัฒนธรรมการขับขี่ปลอดภัยทางถนน: <http://rvpreport.rvpeservice.com/viewrsc.aspx?report=0545&session=16>
- ข่าวไทยพีบีเอส. (2561). สถิติอุบัติเหตุทางถนน “สงกรานต์ 2561” รวม 7 วัน เสียชีวิต 418 คน เพิ่มขึ้นจากปีที่แล้ว. จากเว็บไซต์ องค์การกระจายเสียงและแพร่ภาพสาธารณะแห่งประเทศไทย: <https://news.thaipbs.or.th/content/271713>
- จรัสศรี รุ่งรัตนอุบล, อนามัย นาคอุดม, & ดารณี สุวพันธ์. (2555). *การวิเคราะห์ปัจจัยที่มีผลต่อความพิการจากอุบัติเหตุจราจรทางบก ด้วยเทคนิคเหมืองข้อมูล*. มหาวิทยาลัยนครสวรรค์.
- ภรณ์ยา ปาลวิสุทธิ, & มงคล รอดจันทร์. (2561). *แบบจำลองทำนายความเสี่ยงการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่โดยเทคนิคเหมืองข้อมูล. การประชุมวิชาการระดับชาติ สังคมศาสตร์วิชาการครั้งที่ 14 มหาวิทยาลัยราชภัฏเชียงราย*, (หน้า 689-696). เชียงราย, ประเทศไทย.
- ภทธีรา สุวรรณโค, นิศาชล จานงค์, & จิตมนต์ อังสกุล. (2560). *แบบจำลองการพยากรณ์ความเสี่ยงในการเกิดอุบัติเหตุทางถนนในเทศกาลปีใหม่ด้วยการทำเหมืองข้อมูล*. วารสารวิทยาการและเทคโนโลยีสารสนเทศ, 7(2), 10-19.
- สถาบันการแพทย์ฉุกเฉินแห่งชาติ. (2558). *ผู้บาดเจ็บและเสียชีวิต ในช่วงเทศกาล ปี 51-57_CutName_Songkran_Edit*. จากเว็บไซต์ สำนักงานพัฒนารัฐบาลดิจิทัล (องค์การ

มหาชน): <https://data.go.th/DatasetDetail.aspx?id=f33db12e-bfd5-4eec-aa8a-e5b3ff6e1cb1>

สุนิดา ศรีสุริยชัย. (2550). การวิเคราะห์อุบัติเหตุการจราจรบนท้องถนนในจังหวัดนครปฐม โดยใช้เทคนิคของเหมืองข้อมูล. วิทยานิพนธ์ วิทยาศาสตรมหาบัณฑิต มหาวิทยาลัยมหิดล.