การประยุกต์ใช้การเรียนรู้ของเครื่องในการทำนายความรุนแรงของผู้บาดเจ็บจากอุบัติเหตุทางถนนในช่วงเทศกาลปีใหม่จากข้อมูลเปิดภาครัฐของประเทศไทย

USING MACHINE LEARNING TO PREDICT INJURY SEVERITY OF ROAD TRAFFIC ACCIDENTS DURING NEW YEAR FESTIVALS FROM THAILAND'S OPEN GOVERNMENT DATA

เอกพันธ์ บุญเสริม

บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ

2563

การประยุกต์ใช้การเรียนรู้ของเครื่องในการทำนายความรุนแรงของผู้บาดเจ็บจากอุบัติเหตุทางถนนในช่วงเทศกาลปีใหม่จากข้อมูลเปิดภาครัฐของประเทศไทย

เอกพันธ์ บุญเสริม

สารนิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
วิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ
คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ
ปีการศึกษา 2563
ลิขสิทธิ์ของมหาวิทยาลัยศรีนครินทรวิโรฒ

USING MACHINE LEARNING TO PREDICT INJURY SEVERITY OF ROAD TRAFFIC ACCIDENTS DURING NEW YEAR FESTIVALS FROM THAILAND'S OPEN GOVERNMENT DATA

EAKAPAN BOONSERM

A Master's Project Submitted in Partial Fulfillment of the Requirements

for the Degree of MASTER OF SCIENCE

(Information Technology)

Faculty of Science, Srinakharinwirot University

2020

Copyright of Srinakharinwirot University

สารนิพนธ์

เรื่อง

การประยุกต์ใช้การเรียนรู้ของเครื่องในการทำนายความรุนแรงของผู้บาดเจ็บจากอุบัติเหตุทางถนน ในช่วงเทศกาลปีใหม่จากข้อมูลเปิดภาครัฐของประเทศไทย

ของ

เอกพันธ์ บุญเสริม

ได้รับอนุมัติจากบัณฑิตวิทยาลัยให้นับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ
ของมหาวิทยาลัยศรีนครินทรวิโรฒ

.....

(รองศาสตราจารย์ นายแพทย์ฉัตรชัย เอกปัญญาสกุล)
คณบดีบัณฑิตวิทยาลัย

.....

คณะกรรมการสอบปากเปล่าสารนิพนธ์

..... ที่ปรึกษาหลัก
(ผู้ช่วยศาสตราจารย์ ดร.นุวีย์ วิวัฒนวัฒนา)

..... ประธาน
(อาจารย์ ดร.รัตน์ชัยนันท์ ธรรมสุจริต)

..... กรรมการ
(อาจารย์ ดร.โสภณ มงคลลักษมี)

ชื่อเรื่อง	การประยุกต์ใช้การเรียนรู้ของเครื่องในการทำนายความรุนแรงของผู้บาดเจ็บจากอุบัติเหตุทางถนนในช่วงเทศกาลปีใหม่จากข้อมูลเปิดภาครัฐของประเทศไทย
ผู้วิจัย	เอกพันธ์ บุญเสริม
ปริญญา	วิทยาศาสตรมหาบัณฑิต
ปีการศึกษา	2563
อาจารย์ที่ปรึกษา	ผู้ช่วยศาสตราจารย์ ดร. นุวีย์ วิวัฒนวัฒนา

จำนวนยานพาหนะบนท้องถนนที่เพิ่มขึ้นอย่างรวดเร็ว ส่งผลให้จำนวนอุบัติเหตุจราจรเพิ่มมากขึ้น โดยเฉพาะในช่วงวันหยุดยาวที่ผู้คนเดินทางกลับบ้าน ความเข้าใจที่ดีขึ้นเกี่ยวกับปัจจัยเสี่ยงที่นำไปสู่อุบัติเหตุจราจร อาจช่วยลดจำนวนและป้องกันการบาดเจ็บและเสียชีวิตอย่างรุนแรง งานวิจัยเรื่องนี้กล่าวถึงปัญหาการจำแนกแบบไบนารีที่ไม่สมดุล ในการทำนายความรุนแรงของการบาดเจ็บของอุบัติเหตุจราจรบนท้องถนนในช่วงเทศกาลปีใหม่ของประเทศไทย โดยใช้ข้อมูลเปิดของรัฐบาลของประเทศไทย การสุ่มลดจำนวนตัวอย่าง การเพิ่มจำนวนตัวอย่างด้วย SMOTE และการรวมกันของทั้งการสุ่มลดจำนวนตัวอย่างและการเพิ่มจำนวนตัวอย่าง ถูกนำมาใช้เพื่อปรับความสมดุลของข้อมูลอินสแตนซ์ในข้อมูลฝึกสอน ก่อนที่จะฝึกสอนการเรียนรู้ของเครื่องด้วยอัลกอริทึม Random Forests ผลการวิจัยแสดงให้เห็นว่าการสุ่มลดจำนวนตัวอย่างให้ประสิทธิภาพที่ดีที่สุดที่ 83 เปอร์เซ็นต์ ทั้งในด้าน Recall และ Accuracy ปัจจัยสำคัญที่ทำให้การทำนายความรุนแรงของผู้บาดเจ็บจากอุบัติเหตุ ได้แก่ วิธีการนำส่งของผู้บาดเจ็บ อายุ และการดื่มแอลกอฮอล์ งานวิจัยเรื่องนี้อาจเป็นประโยชน์ต่อหน่วยงานของรัฐในการวิเคราะห์ระบุตำแหน่งความรุนของการเกิดอุบัติเหตุและปัจจัยสำคัญที่กีดกันความปลอดภัยทางถนน

คำสำคัญ : การจำแนกประเภทแบบไบนารีที่ไม่สมดุล, การเรียนรู้ของเครื่อง, อุบัติเหตุทางถนน

Title	USING MACHINE LEARNING TO PREDICT INJURY SEVERITY OF ROAD TRAFFIC ACCIDENTS DURING NEW YEAR FESTIVALS FROM THAILAND'S OPEN GOVERNMENT DATA
Author	EAKAPAN BOONSERM
Degree	MASTER OF SCIENCE
Academic Year	2020
Thesis Advisor	Assistant Professor Nuwee Wiwatwattana , Ph.D.

The fast-rising number of vehicles on the road resulted in a matching number of traffic accidents, especially during long holidays when people travel back home. A better understanding of the risk factors leading to traffic accidents may help decrease the number and to prevent severe injuries and deaths. This research addressed the imbalanced binary classification problem of predicting injury severity of road traffic accidents during New Year festivals in Thailand using open government data. Random undersampling and oversampling with SMOTE, and the combination of both undersampling and oversampling, used to rebalance training instances before fitting any Random Forests models. The results showed that random undersampling gave the best performance at 83%, with regard to both recall and accuracy. The important factors driving the prediction of severe injuries, including method, age, and alcohol consumption. This work could benefit public authorities in analyzing accident hot spots and crucial factors that discourage road safety.

Keyword : Imbalanced Classification, Machine Learning, Traffic Accident, Random Forest

กิตติกรรมประกาศ

สารนิพนธ์นี้สำเร็จลุล่วงได้ด้วยความอนุเคราะห์จาก ผศ.ดร.นุวีย์ วิวัฒนวัฒนา อาจารย์ที่ปรึกษาที่ให้คำปรึกษา คำแนะนำในการทำสารนิพนธ์ ตลอดจนสนับสนุนข้อมูลทางวิชาการ และข้อมูลสำหรับทำสารนิพนธ์นี้

ขอกราบขอบพระคุณสำนักงานพัฒนารัฐบาลดิจิทัล องค์การมหาชน ที่ให้ความอนุเคราะห์ข้อมูลผู้บาดเจ็บจากอุบัติเหตุจราจรในช่วงเทศกาลปีใหม่ 2551 ถึงปี 2558 มาใช้ในการทำสารนิพนธ์นี้

ขอกราบขอบพระคุณคณะกรรมการสอบสารนิพนธ์ที่ได้ให้คำแนะนำและข้อเสนอแนะสำหรับการปรับปรุงสารนิพนธ์

ขอกราบขอบพระคุณบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ สำหรับทุนสนับสนุนการนำเสนอผลงานวิจัยของนิสิตบัณฑิตศึกษา ทำให้ได้รับประสบการณ์นำเสนอสารนิพนธ์รวมถึงได้แลกเปลี่ยนความรู้ทางด้านเทคโนโลยีต่างๆ

เอกพันธ์ บุญเสริม

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ง

บทคัดย่อภาษาอังกฤษ ..... จ

กิตติกรรมประกาศ ..... ฉ

สารบัญ ..... ช

สารบัญตาราง ..... ฌ

สารบัญรูปภาพ ..... ญ

บทที่ 1 บทนำ 1

1.1 ความสำคัญและความเป็นมาของงานวิจัย 1

1.2 วัตถุประสงค์ 2

1.3 กรอบแนวคิดในการวิจัย 3

1.4 ประโยชน์ที่คาดว่าจะได้รับจากงานวิจัย 4

บทที่ 2 วรรณกรรมและงานวิจัยที่เกี่ยวข้อง 5

2.1 ทฤษฎีเกี่ยวกับการเตรียมข้อมูล (Data Preprocessing) 5

2.1.1 Imbalanced data 5

2.1.2 Oversampling with SMOTE 6

2.1.3 Random Undersampling 7

2.1.4 Transformation Pipelines 8

2.2 ทฤษฎีเกี่ยวกับอัลกอริทึม 8

2.2.1 Random Forest 8

2.2.2 Grid Search 10

2.2.3 การเลือกพารามิเตอร์ของแต่ละอัลกอริทึม 11

2.2.4 การวัดประสิทธิภาพ 12
2.3 งานวิจัยที่เกี่ยวข้อง 17
บทที่ 3 การดำเนินการวิจัย 28
3.1 กระบวนการทำงานของแบบจำลอง 28
3.2 การเก็บรวบรวมข้อมูล 30
3.3 การสำรวจข้อมูล Exploratory Data Analysis (EDA) 39
3.4 การเตรียมข้อมูล 73
3.5 อัลกอริทึมของแบบจำลองของการทำนาย 74
บทที่ 4 ผลการดำเนินการวิจัย 79
4.1 ผลลัพธ์ประสิทธิภาพของแบบจำลองการทำนาย 79
4.2 ผลลัพธ์การวิเคราะห์ปัจจัยในการทำนายความรุนแรงของผู้บาดเจ็บจากอุบัติเหตุทางถนน ในช่วงเทศกาลปีใหม่ 86
บทที่ 5 สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ 91
5.1 สรุปผลการวิจัย 91
5.2 อภิปรายผลการวิจัย 92
5.3 ข้อเสนอแนะ 94
บรรณานุกรม 96
ประวัติผู้เขียน 100

สารบัญตาราง

หน้า

ตาราง 1 รายละเอียดของชุดข้อมูลผู้บาดเจ็บจากอุบัติเหตุจราจรในช่วงเทศกาลปีใหม่ 2551 ถึงปี 2558 31

ตาราง 2 แสดงปัจจัยที่มีผลต่ออัตราการเสียชีวิตของผู้บาดเจ็บจากอุบัติเหตุสูงสุดในช่วงเทศกาลปีใหม่ 2551 ถึง 2558 72

ตาราง 3 แสดงข้อมูลฝึกสอนแบบจำลองการทำนาย..... 76

ตาราง 4 แสดงพารามิเตอร์ที่ดีที่สุดของแต่ละแบบจำลองการทำนายความรุนแรงของผู้บาดเจ็บจากอุบัติเหตุทางถนนในช่วงเทศกาลปีใหม่ 2551 ถึง 2558 78

สารบัญรูปภาพ

หน้า

ภาพประกอบ 1 แสดงหลักการทำงานของ SMOTE (Wei-Chiang Hong, November 11, 2013) 7

ภาพประกอบ 2 แสดงการทำงานของอัลกอริทึม Random Forest (Andreas C.Muller & Sarah Guido, October, 2016) 9

ภาพประกอบ 3 แสดงการทำงานของ Grid Search และ Cross-validation (J. Mach. Learn. Res, 2010) 11

ภาพประกอบ 4 แสดงการทำงานของ Cross Validation (J. Mach. Learn. Res, 2010) 12

ภาพประกอบ 5 แสดงตาราง Confusion Matrix (Kevin Markham, 2014) 13

ภาพประกอบ 6 แสดงแนวคิดของ Receiver Operating Characteristic Curve (Tourassi, 2018) 16

ภาพประกอบ 7 Flowchart กระบวนการสร้างแบบจำลอง 28

ภาพประกอบ 8 สัดส่วนข้อมูลจากองค์กรสถาบันการแพทย์ฉุกเฉินแห่งชาติ ผู้บาดเจ็บจากอุบัติเหตุ ในช่วงเทศกาลปีใหม่ของปี 2551 ถึงปี 2558 36

ภาพประกอบ 9 สัดส่วนข้อมูลจากองค์กรสถาบันการแพทย์ฉุกเฉินแห่งชาติ ผู้บาดเจ็บจากอุบัติเหตุ ในช่วงเทศกาลปีใหม่ของปี 2551 ถึงปี 2558 (ต่อ) 37

ภาพประกอบ 10 แสดงตัวอย่างข้อมูลจากองค์กรสถาบันการแพทย์ฉุกเฉินแห่งชาติ ผู้บาดเจ็บจากอุบัติเหตุ ในช่วงเทศกาลปีใหม่ของปี 2551 ถึงปี 2558 38

ภาพประกอบ 11 ตัวอย่างข้อมูลองค์กรสถาบันการแพทย์ฉุกเฉินแห่งชาติ ผู้บาดเจ็บจากอุบัติเหตุที่เป็นหมวดหมู่ (Categorical Data) 38

ภาพประกอบ 12 ตัวอย่างข้อมูลองค์กรสถาบันการแพทย์ฉุกเฉินแห่งชาติ ผู้บาดเจ็บจากอุบัติเหตุที่เป็นข้อมูลต่อเนื่อง (Continuous Data) 39

ภาพประกอบ 13 ข้อมูลกราฟแท่งแสดงความสัมพันธ์ระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับวันที่เกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558 40

ภาพประกอบ 14 ข้อมูลกราฟแท่งแสดงความสัมพันธ์ระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับชื่อเทศกาลปีใหม่ 2551 ถึง ปี 2558 41

ภาพประกอบ 15 ข้อมูลเปอร์เซ็นต์เปรียบเทียบระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับชื่อเทศกาลปีใหม่ 2551 ถึง ปี 2558 41

ภาพประกอบ 16 ข้อมูลเปรียบเทียบระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับชื่อเทศกาลปีใหม่และค่าอัตราการเสียชีวิตในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558..... 42

ภาพประกอบ 17 ข้อมูลกราฟแท่งแสดงความสัมพันธ์ระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับช่วงเวลารายชั่วโมงในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558 43

ภาพประกอบ 18 ข้อมูลเปอร์เซ็นต์เปรียบเทียบระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับช่วงเวลารายชั่วโมงในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558 43

ภาพประกอบ 19 ข้อมูลเปรียบเทียบระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับช่วงเวลารายชั่วโมงและค่าอัตราการเสียชีวิตในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558..... 44

ภาพประกอบ 20 ข้อมูลกราฟแท่งแสดงความสัมพันธ์ระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับอายุของผู้ประสบอุบัติเหตุในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558..... 45

ภาพประกอบ 21 ข้อมูลเปอร์เซ็นต์เปรียบเทียบระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับอายุของผู้ประสบอุบัติเหตุในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558..... 45

ภาพประกอบ 22 ข้อมูลเปรียบเทียบระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับอายุของผู้ประสบอุบัติเหตุและค่าอัตราการเสียชีวิตในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558 46

ภาพประกอบ 23 ข้อมูลกราฟแท่งแสดงความสัมพันธ์ระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับสถานะของผู้ประสบอุบัติเหตุในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558 47

ภาพประกอบ 24 ข้อมูลกราฟแท่งแสดงความสัมพันธ์ระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับยานพาหนะของผู้ขับขี่ที่ประสบอุบัติเหตุในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558 48

ภาพประกอบ 25 ข้อมูลกราฟแท่งแสดงความสัมพันธ์ระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับยานพาหนะของรถคู่กรณีของผู้ประสบอุบัติเหตุในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558..... 49

ภาพประกอบ 26 ข้อมูลกราฟแท่งแสดงความสัมพันธ์ระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับมาตรการรณรงค์ลดอุบัติเหตุในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558..... 50

ภาพประกอบ 27 ข้อมูลเปอร์เซ็นต์เปรียบเทียบระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับมาตรการรณรงค์ลดอุบัติเหตุในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558 50

ภาพประกอบ 28 ข้อมูลเปรียบเทียบระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับมาตรการรณรงค์ลดอุบัติเหตุและค่าอัตราการเสียชีวิตในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558 51

ภาพประกอบ 29 ข้อมูลกราฟแท่งแสดงความสัมพันธ์ระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับเพศของผู้ประสบอุบัติเหตุในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558 52

ภาพประกอบ 30 ข้อมูลเปอร์เซ็นต์เปรียบเทียบระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับเพศของผู้ประสบอุบัติเหตุในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558 52

ภาพประกอบ 31 ข้อมูลเปรียบเทียบระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับเพศของผู้ประสบอุบัติเหตุและค่าอัตราการเสียชีวิตในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558 52

ภาพประกอบ 32 ข้อมูลกราฟแท่งแสดงความสัมพันธ์ระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับเส้นทางถนนที่เกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558 53

ภาพประกอบ 33 ข้อมูลกราฟแท่งแสดงความสัมพันธ์ระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับโรงพยาบาลที่รักษาผู้บาดเจ็บจากอุบัติเหตุจากมากสุด 30 ลำดับแรก ในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558 54

ภาพประกอบ 34 ข้อมูลเปรียบเทียบระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับโรงพยาบาลที่รักษาผู้บาดเจ็บจากอุบัติเหตุจากมากสุด 30 ลำดับแรก และค่าอัตราการเสียชีวิตในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558 55

ภาพประกอบ 35 ข้อมูลกราฟแท่งแสดงความสัมพันธ์ระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับจำนวนวันรักษาของผู้บาดเจ็บในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558 56

ภาพประกอบ 36 ข้อมูลเปอร์เซ็นต์เปรียบเทียบระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับจำนวนวันรักษาของผู้บาดเจ็บในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558 57

ภาพประกอบ 37 ข้อมูลเปรียบเทียบระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับจำนวนวันรักษาของผู้บาดเจ็บและค่าอัตราการเสียชีวิตในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558 58

ภาพประกอบ 38 ข้อมูลกราฟแท่งแสดงความสัมพันธ์ระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับจังหวัดที่เกิดอุบัติเหตุมากสุด 10 ลำดับแรก ในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558 59

ภาพประกอบ 39 ข้อมูลเปรียบเทียบระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับจังหวัดที่เกิดอุบัติเหตุมากสุด 10 ลำดับแรกและค่าอัตราการเสียชีวิตในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558 59

ภาพประกอบ 40 ข้อมูลกราฟแท่งแสดงความสัมพันธ์ระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับการนำส่งผู้ประสบอุบัติเหตุในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558 61

ภาพประกอบ 41 ข้อมูลกราฟแท่งแสดงความสัมพันธ์ระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับการดื่มสุราของผู้ประสบอุบัติเหตุในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558 62

ภาพประกอบ 42 ข้อมูลกราฟแท่งแสดงความสัมพันธ์ระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับจังหวัดที่เกิดอุบัติเหตุในประเทศไทยมากสุด 30 ลำดับแรก ในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558 63

ภาพประกอบ 43 แสดงความสัมพันธ์ระหว่างวันที่เกิดเหตุกับเวลาเกิดเหตุของผู้ประสบอุบัติเหตุในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558 64

ภาพประกอบ 44 แสดงความสัมพันธ์ระหว่างวันที่เกิดเหตุกับเวลาเกิดเหตุของผู้ประสบอุบัติเหตุที่เสียชีวิตในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558 65

ภาพประกอบ 45 แสดงความสัมพันธ์ระหว่างวันที่เกิดเหตุกับเวลาเกิดเหตุของผู้ประสบอุบัติเหตุที่ไม่เสียชีวิตในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558 66

ภาพประกอบ 46 แสดงความสัมพันธ์ระหว่างชื่อเทศกาลปีใหม่กับวันที่เกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558 67

ภาพประกอบ 47 แสดงความสัมพันธ์ระหว่างชื่อเทศกาลปีใหม่กับวันที่เกิดอุบัติเหตุของผู้เสียชีวิตในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558 67

ภาพประกอบ 48 แสดงความสัมพันธ์ระหว่างชื่อเทศกาลปีใหม่กับวันที่เกิดอุบัติเหตุของผู้ไม่เสียชีวิตในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558 67

ภาพประกอบ 49 แสดงความรุนแรงของค่าอัตราการเสียชีวิตของการเกิดอุบัติเหตุทั้งประเทศไทยในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558 68

ภาพประกอบ 50 แสดงความรุนแรงของค่าอัตราการเสียชีวิตของการเกิดอุบัติเหตุในภาคกลางของประเทศไทยในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558 69

ภาพประกอบ 51 แสดงความรุนแรงของค่าอัตราการเสียชีวิตของการเกิดอุบัติเหตุในภาคตะวันออกของประเทศไทยในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558 69

ภาพประกอบ 52 แสดงความรุนแรงของค่าอัตราการเสียชีวิตของการเกิดอุบัติเหตุในภาคตะวันออกเฉียงเหนือของประเทศไทยในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558 70

ภาพประกอบ 53 แสดงความรุนแรงของค่าอัตราการเสียชีวิตของการเกิดอุบัติเหตุในภาคเหนือของประเทศไทยในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558 70

ภาพประกอบ 54 แสดงความรุนแรงของค่าอัตราการเสียชีวิต ของการเกิดอุบัติเหตุในภาคใต้ของประเทศไทยในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558 71

ภาพประกอบ 55 แสดงความรุนแรงของค่าอัตราการเสียชีวิตของการเกิดอุบัติเหตุในภาคตะวันตกของประเทศไทยในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558 71

ภาพประกอบ 56 แสดงจำนวนผลการรักษาผู้บาดเจ็บอุบัติเหตุ (Class Label) 74

ภาพประกอบ 57 ข้อมูล Categorical Data ที่ผ่านการแปลงด้วยเทคนิค One Hot Encoding ... 74

ภาพประกอบ 58 แสดงข้อมูลที่ใช้สำหรับการสร้างและประเมินแบบจำลองการทำนายความรุนแรงผู้บาดเจ็บจากอุบัติเหตุ 75

ภาพประกอบ 59 กราฟแท่งแสดงสัดส่วนข้อมูลฝึกสอนแบบจำลองการทำนาย 76

ภาพประกอบ 60 กราฟแท่งแสดงข้อมูลเปรียบเทียบค่าความถูกต้องของ Overall Accuracy และ Recall 79

ภาพประกอบ 61 กราฟแท่งแสดงข้อมูลเปรียบเทียบค่า Precision และ Recall 80

ภาพประกอบ 62 แสดงข้อมูลของ Confusion Matrices แกน y คือค่าข้อมูลจริง แกน x คือค่าทำนาย 80

ภาพประกอบ 63 แสดง Receiver Operating Characteristic (ROC Curve) เปรียบเทียบ 4 แบบจำลอง 83

ภาพประกอบ 64 แสดง Precision-Recall Curve เปรียบเทียบ 4 แบบจำลอง 84

ภาพประกอบ 65 แสดงข้อมูล Reported Feature importances โดยแบบจำลองที่ 1 Random Forest ที่ไม่ใช้เทคนิคจัดการ Imbalanced Data 86

ภาพประกอบ 66 แสดงข้อมูล Reported Feature importances โดยแบบจำลองที่ 2 Random Forest ที่ใช้เทคนิคจัดการ Imbalanced Data ของ Random Undersample 87

ภาพประกอบ 67 แสดงข้อมูล Reported Feature importances โดยแบบจำลองที่ 3 Random Forest ที่ใช้เทคนิคจัดการ Imbalanced Data ของ Oversample with SMOTE 88

ภาพประกอบ 68 แสดงข้อมูล Reported Feature importances โดยแบบจำลองที่ 4 Random Forest ที่ใช้เทคนิคจัดการ Imbalanced Data ของ Oversample with SMOTE และ Random Undersample 89

บทที่ 1
บทนำ

1.1 ความสำคัญและความเป็นมาของงานวิจัย

ปัจจุบันเมืองที่มีการพัฒนาเติบโตขึ้นอย่างรวดเร็วแล้วยังคงพัฒนาต่อไปในอนาคต ถือเป็นความท้าทายในการจัดการและวางแผนระบบขนส่งสาธารณะที่มีประสิทธิภาพ สะดวกและปลอดภัย จากรายงานขององค์การอนามัยโลก (World Health Organization : WHO) ผลการจัดทำรายงานสถานการณ์โลกด้านความปลอดภัยทางถนน ครั้งที่ 4 เมื่อวันที่ 20 ธันวาคม 2558 (องค์การอนามัยโลก World Health Organization, 2558) ที่ผ่านมา พบว่าจากข้อมูลปี 2558 ทั่วโลกมีจำนวนผู้เสียชีวิตจากอุบัติเหตุทางถนนประมาณ 1.35 ล้านคนต่อปีหรือเฉลี่ย 3,700 คนต่อวัน และสาเหตุอันดับ 1 ที่คร่าชีวิตของเด็กและเยาวชนทั่วโลก โดยครึ่งของผู้เสียชีวิตเป็นผู้ขับขี่รถจักรยานยนต์ จักรยาน และคนเดินถนน และสำหรับประเทศไทย องค์การอนามัยโลกได้ประมาณการอัตราการเสียชีวิตอยู่ที่ 32.8 ต่อประชากรแสนคน จัดอยู่ในอันดับที่ 9 ของประเทศสมาชิกทั้งหมด ซึ่งเป็นปัญหาที่ยากจะวางแผนป้องกันหรือคาดการณ์ล่วงหน้าได้ สำหรับในประเทศที่กำลังพัฒนา จำนวนยานพาหนะบนท้องถนนที่เพิ่มขึ้นอย่างรวดเร็วส่งผลให้จำนวนอุบัติเหตุจราจรบนท้องถนนมีจำนวนมาก โดยเฉพาะในช่วงวันหยุดยาวของเทศกาลปีใหม่ที่ผู้คนส่วนใหญ่เดินทางกลับภูมิลำเนา การทำความเข้าใจเกี่ยวกับปัจจัยเสี่ยงที่จะนำไปสู่การเกิดอุบัติเหตุจราจร อาจช่วยลดจำนวนและป้องกันการบาดเจ็บจากอุบัติเหตุที่รุนแรงถึงการเสียชีวิตได้ ในความเป็นจริงรัฐบาลของประเทศส่วนใหญ่ได้รวบรวมข้อมูลและรายงานสถิติอุบัติเหตุทางถนนเป็นประจำ อย่างไรก็ตามข้อมูลเหล่านี้อาจจะไม่ได้นำมาใช้ในการวิเคราะห์วินิจฉัยและการวิเคราะห์เชิงคาดการณ์ของอุบัติเหตุ เพื่อสร้างแผนสาธารณะในการป้องกันอุบัติเหตุ ซึ่งการพัฒนาล่าสุดในด้านวิทยาศาสตร์ข้อมูลทำให้สามารถใช้ประโยชน์จากข้อมูลอุบัติเหตุได้สูงสุด

ในงานวิจัยเรื่องนี้ทางผู้วิจัยได้วิเคราะห์ปัจจัยและสร้างแบบจำลองการทำนายความรุนแรงของผู้บาดเจ็บจากอุบัติเหตุทางถนนในช่วงเทศกาลปีใหม่ของประเทศไทย ซึ่งใช้ข้อมูลเปิดภาครัฐของประเทศไทย ในชุดข้อมูลผู้บาดเจ็บจากอุบัติเหตุจราจรในช่วงเทศกาลปีใหม่ 2551 ถึง 2558 จากองค์กรสถาบันการแพทย์ฉุกเฉินแห่งชาติ เผยแพร่โดยสำนักงานพัฒนารัฐบาลดิจิทัล องค์การมหาชน (Digital Government Development Agency (Public Organization) : DGA) (สถาบันการแพทย์ฉุกเฉินแห่งชาติ, 2558) โดยใช้เทคนิคการเรียนรู้แบบมีผู้สอน (Supervised Learning) ในแบบ Classification Algorithm เพื่อสร้างแบบจำลองการทำนายความรุนแรงของผู้บาดเจ็บจากอุบัติเหตุทางถนน อย่างไรก็ตามการฝึกฝนแบบจำลองการทำนายที่ตรงไปตรงมามี

แนวโน้มที่จะสร้างอคติกับให้กับแบบจำลองการทำนาย เนื่องจากลักษณะที่ไม่สมดุลของข้อมูลในแง่ที่ว่าจำนวนผู้บาดเจ็บจากอุบัติเหตุที่เสียชีวิตและไม่เสียชีวิตมีความแตกต่างกันมาก อุบัติเหตุทางถนนเพียงเล็กน้อยเท่านั้นที่เป็นอันตรายถึงการเสียชีวิต หากไม่มีการจัดการข้อมูลที่ไม่สมดุลอย่างเหมาะสม แบบจำลองการทำนายจะทำนายข้อมูลกลุ่มตัวอย่างส่วนมากและไม่รู้จักกลุ่มตัวอย่างของข้อมูลกลุ่มน้อย ซึ่งก็คือแบบจำลองจะมีแนวโน้มเลือกทำนายว่าอุบัติเหตุนั้นไม่เกิดคนเสียชีวิต ซึ่งแน่นอนว่านี่คือสิ่งที่ต้องได้รับการแก้ไขในงานวิจัยเรื่องนี้ เนื่องจากความสูญเสียที่ไม่สามารถตรวจจับหรือจำแนกอุบัติเหตุที่รุนแรงถึงการเสียชีวิตนั้นสูง ซึ่งก็คือการสูญเสียชีวิต

งานวิจัยเรื่องนี้มุ่งเน้นการแก้ปัญหาความไม่สมดุลของข้อมูลของการจำแนกประเภทแบบไบนารี (The Imbalanced Binary Classification Problem) ในการทำนายความรุนแรงของผู้บาดเจ็บจากอุบัติเหตุทางถนนในช่วงเทศกาลปีใหม่ 2551 ถึง 2558 ของประเทศไทย อัลกอริทึมการเรียนรู้ของเครื่อง (Machine Learning Algorithm) ที่ใช้คือ Random Forests ซึ่งแสดงให้เห็นว่ามีประสิทธิภาพในการทำนายสูง (High Predictive Power) และใช้เทคนิคการสุ่มตัวอย่างใหม่ (Resampling Techniques) 3 แบบ เพื่อปรับความสมดุลของข้อมูลอินสแตนซ์ในข้อมูลฝึกสอนของแบบจำลองการทำนาย (Rebalance Training Instances) ได้แก่ เทคนิคการลดจำนวนตัวอย่าง (Undersampling), เทคนิคการเพิ่มจำนวนตัวอย่างด้วย SMOTE (Oversampling with SMOTE) และการผสมระหว่างของทั้ง 2 เทคนิค ของเทคนิคการลดจำนวนตัวอย่าง (Undersampling) กับเทคนิคการเพิ่มจำนวนตัวอย่างด้วย SMOTE (Oversampling with SMOTE) ในส่วนการประเมินผลแบบจำลองใช้รายงาน Classification Report เพื่อวัดประสิทธิภาพของแบบจำลองในการระบุอุบัติเหตุที่รุนแรงและไม่รุนแรงถึงการเสียชีวิตจากอุบัติเหตุทางถนนในช่วงเทศกาลปีใหม่ โดยที่ไม่ได้เพียงแต่มุ่งเน้นถึงความถูกต้องของการทำนายทั้งหมด (Accuracy) เท่านั้น แต่ยังมุ่งเน้นไปที่ค่า Recall เพื่อวัดประสิทธิภาพที่แท้จริงของการระบุผู้บาดเจ็บจากอุบัติเหตุที่รุนแรงถึงการเสียชีวิตจากอุบัติเหตุทางถนนในช่วงเทศกาลปีใหม่

1.2 วัตถุประสงค์

ในการวิจัยครั้งนี้ผู้วิจัยได้ตั้งความมุ่งหมายไว้ดังนี้

1. เพื่อศึกษาปัจจัยการเกิดอุบัติเหตุที่ทำให้ผู้บาดเจ็บรุนแรงและการประยุกต์ใช้เทคนิคการเรียนรู้ของเครื่อง (Machine Learning) มาสร้างแบบจำลองการทำนายความรุนแรงของผู้บาดเจ็บจากอุบัติเหตุทางถนนในช่วงเทศกาลปีใหม่

2. เพื่อให้ทราบถึงหลักการทำงานของอัลกอริทึมที่มุ่งเน้นการแก้ปัญหาความไม่สมดุลของข้อมูลของการจำแนกประเภทแบบไบนารีและการประเมินวัดผลแบบจำลองที่มี

ประสิทธิภาพสามารถตรวจจับหรือจำแนกอุบัติเหตุที่รุนแรงถึงการเสียชีวิต ในการทำนายความรุนแรงของผู้บาดเจ็บจากอุบัติเหตุทางถนนในช่วงเทศกาลปีใหม่

1.3 กรอบแนวคิดในการวิจัย

การพัฒนาแบบจำลองการทำนายความรุนแรงของผู้บาดเจ็บจากอุบัติเหตุทางถนน แบ่งออกเป็น 3 ขั้นตอน

ขั้นตอนแรกประกอบด้วยกระบวนการรวบรวมข้อมูลและการสำรวจข้อมูล (Exploratory Data Analysis : EDA) วิเคราะห์ความสัมพันธ์ระหว่างตัวแปรต่างๆ กับผลการรักษาผู้บาดเจ็บจากอุบัติเหตุในช่วงเทศกาลปีใหม่ 2551 ถึง 2558 ที่ถูกจัดเก็บมาเรียบร้อยแล้วจากองค์กรสถาบันการแพทย์ฉุกเฉินแห่งชาติ เผยแพร่โดยสำนักงานพัฒนารัฐบาลดิจิทัล องค์การมหาชน (Digital Government Development Agency (Public Organization) : DGA) ซึ่งประกอบด้วย Feature ทั้งหมด 19 คอลัมน์ ได้แก่ ชื่อเทศกาล, รหัสจังหวัด, ชื่อจังหวัด, รหัสโรงพยาบาล, ชื่อโรงพยาบาลที่รับผู้บาดเจ็บ, วันที่เกิดเหตุ, เวลาเกิดเหตุ, เพศ, อายุ, ถนนที่เกิดเหตุ, สถานะ, รถผู้บาดเจ็บ, รถคู่กรณี, มาตรการ, การดื่มสุรา, การนำส่ง, Refer-Admit, ผลการรักษา, จำนวนวันรักษา โดยข้อมูลมีการจัดเก็บตั้งแต่วันที่ 28 ธันวาคม 2551 (เวลา 01:01 น.) ถึง วันที่ 5 มกราคม 2558 (เวลา 24:00 น.) เป็นจำนวนทั้งหมด 214,950 แถว

ขั้นตอนที่สองมุ่งเน้นไปที่กระบวนการของการเตรียมข้อมูล การแปลงข้อมูล แบ่งข้อมูล (Split Data) ออกเป็นข้อมูลฝึกสอน (Training Set) และข้อมูลทดสอบ (Test Set) นอกจากนี้ ข้อมูลฝึกสอนยังแบ่งข้อมูลออกเป็น Train Set และ Validation Set โดยใช้เทคนิค Stratify ในการ Split ข้อมูล และใช้เทคนิค Pipeline ในการจัดลำดับการทำงานของปรับความสมดุลของข้อมูลฝึกสอนและการฝึกสอนแบบจำลองการทำนาย สำหรับระดับความรุนแรงของการบาดเจ็บจากอุบัติเหตุมี 2 ระดับคือ 0 : ไม่เสียชีวิต และ 1 : เสียชีวิต

ในขั้นตอนสุดท้ายมุ่งเน้นการพัฒนาแบบจำลองการทำนายความรุนแรงของผู้บาดเจ็บจากอุบัติเหตุทางถนนในแบบการจำแนกประเภทแบบไบนารี อัลกอริทึมการเรียนรู้ของเครื่องที่ใช้คือ Random Forests และใช้เทคนิคการสุ่มตัวอย่างใหม่ (Resampling Techniques) 3 แบบ เพื่อปรับความสมดุลของข้อมูลอินสแตนซ์ในข้อมูลฝึกสอนของแบบจำลองการทำนาย (Rebalance Training Instances) ได้แก่ เทคนิคการลดจำนวนตัวอย่าง (Undersampling), เทคนิคการเพิ่มจำนวนตัวอย่างด้วย SMOTE (Oversampling with SMOTE) และการผสมระหว่างของทั้ง 2 เทคนิค ของเทคนิคการลดจำนวนตัวอย่าง (Undersampling) กับเทคนิคการเพิ่มจำนวนตัวอย่างด้วย SMOTE (Oversampling with SMOTE) แล้วทำการประเมินผลแบบจำลองใช้

รายงาน Classification Report ประกอบด้วย Precision, Recall, Accuracy, Confusion Matrix, Receiver Operating Characteristic Curve (ROC Curve) และ Precision-Recall Curve ด้วยข้อมูลทดสอบ (Test Set) และปัจจัยที่มีผลต่อการตัดสินในการทำนาย (Report Feature Importance) ของแต่ละแบบจำลอง เพื่อให้ได้ปัจจัย (Feature) ที่มีความสำคัญส่งผลต่อการทำนายความรุนแรงของผู้บาดเจ็บจากอุบัติเหตุทางถนนในช่วงเทศกาลปีใหม่ 2551 ถึง 2558

1.4 ประโยชน์ที่คาดว่าจะได้รับจากงานวิจัย

1. ทำให้สามารถนำแบบจำลองไปใช้ในการทำนายผลการรักษาผู้บาดเจ็บจากการประสบอุบัติเหตุว่าจะเสียชีวิตหรือไม่เสียชีวิต

2. ข้อมูลผลการทดลองของงานวิจัยสามารถนำไปช่วยปรับปรุงมาตราการรณรงค์ลดอุบัติเหตุจราจรในช่วงของเทศกาลปีใหม่ได้หรือใช้เป็นข้อมูลประกอบการเสนอนโยบายและแนวทางการลดอุบัติเหตุทางถนนและความรุนแรงของการเกิดอุบัติเหตุทางถนนได้

3. ทำให้ทราบว่าปัจจัยใดที่มีผลทำให้เกิดอุบัติเหตุจราจรมากที่สุด เพื่อที่จะได้เป็นประโยชน์ในการหาแนวทางป้องกันอุบัติเหตุ

4. เพื่อให้สามารถนำมาต่อยอดใช้งานและพัฒนาแบบจำลองให้แม่นยำและดียิ่งขึ้นในอนาคต

5. เพื่อใช้แบบจำลองการทำนายผลการรักษาผู้บาดเจ็บจากการประสบอุบัติเหตุในช่วงเทศกาลปีใหม่ได้อย่าง Real Time

6. เพื่อใช้เป็นข้อมูลประกอบการกำหนดมาตรการ การจัดทำระเบียบ กฎหมายและมาตรฐานด้านความปลอดภัยทางถนนต่อไป

งานวิจัยเล่มนี้มีโครงสร้างดังนี้ บทที่ 2 วรรณกรรมที่เกี่ยวข้องกับการทำนายอุบัติเหตุจราจร บทที่ 3 อธิบายการดำเนินการวิจัย บทที่ 4 กล่าวถึงผลลัพธ์และสุดท้ายบทที่ 5 ให้ข้อสรุป

บทที่ 2
วรรณกรรมและงานวิจัยที่เกี่ยวข้อง

ในการวิจัยครั้งนี้ ผู้วิจัยได้ศึกษาเอกสารทฤษฎีและงานวิจัยที่เกี่ยวข้องกับการพัฒนาแบบจำลองในการทำนายผลการรักษาผู้ประสบอุบัติเหตุ และได้นำเสนอตามหัวข้อดังต่อไปนี้

1. ทฤษฎีเกี่ยวกับการเตรียมข้อมูล (Data Preprocessing)
2. ทฤษฎีเกี่ยวกับอัลกอริทึม
3. งานวิจัยที่เกี่ยวข้อง

2.1 ทฤษฎีเกี่ยวกับการเตรียมข้อมูล (Data Preprocessing)

2.1.1 Imbalanced data

ข้อมูลที่มีปัญหา Imbalanced Data คือ ข้อมูลที่ Class หนึ่งมีจำนวนมากกว่าอีก Class หนึ่งหรือข้อมูลประเภทหนึ่งมีมากกว่าอีกประเภทหนึ่ง มีเฉพาะในโจทย์ปัญหาของการจำแนกประเภท ในส่วนของ Regression นั้นไม่มีเพราะทำนายค่าตัวเลขอย่างเดียวจึงไม่มีปัญหาที่ว่าตัวเลขไหนเยอะกว่ากัน ซึ่งปัญหาเกิดจาก Asymmetric Cost ในการได้ข้อมูลไม่เท่ากัน ตัวอย่างเช่น ถ้าต้องการได้ข้อมูลของคนป่วย ก็ต้องให้คุณหมอตรวจ พวกนี้คือค่าใช้จ่าย ดังนั้นถ้าคนไม่ได้ไปหาหมอก็จะไม่ได้ตรวจโรค ทำให้เราไม่ทราบว่าป่วยหรือไม่ เมื่อ Cost ไม่เท่ากันจำนวนคนแต่ละประเภทจะไม่เหมือนกัน และอีกปัญหาที่เกิดขึ้นคือข้อมูลในความเป็นจริงมีจำนวนไม่เท่ากันอยู่แล้ว ตัวอย่างเช่น คนที่ป่วยเป็นโรคมะเร็งมีน้อยกว่าคนที่ไม่ได้ป่วยเป็นโรคมะเร็งอยู่แล้ว ข้อมูลจึงไม่ Balanced กัน

โจทย์ปัญหาด้านการจำแนกประเภทนั้นอัลกอริทึมจะมีความ Sensitive กับข้อมูล เพราะถ้ามีข้อมูลประเภทไหนมากกว่ากว่าอีกประเภท จะมีความ Sensitive ตรงที่แบบจำลองจะตรวจจับประเภทที่มีจำนวนข้อมูลมากกว่า คือ Majority Class และจะมีความ Sensitive กับ Minority Class ซึ่งมีน้อยกว่า ถ้าไม่มุ่งเน้นถึงปัญหาของความไม่สมดุลของข้อมูล สิ่งที่เกิดขึ้นคือการจำแนกจะเป็นแบบ Majority แบบจำลองจะเลือกการทำนายประเภท Class ที่มากกว่าเข้าไว้ ทำให้มีอคติ เพราะถ้าข้อมูลประเภท Class ที่มีจำนวนน้อย อัลกอริทึมจะไม่สามารถตรวจจับได้

วิธีการทำ Imbalanced Data มี 2 แบบ คือ

แบบที่ 1 เป็นการแก้ไขข้อมูล โดยวิธีการทำ Oversampling คือเทคนิคการเพิ่มจำนวนตัวอย่าง ถ้าข้อมูลใน Class น้อย จะเพิ่มข้อมูลให้เท่ากับ Class ที่มีจำนวนมากกว่า ให้มีจำนวนเท่ากัน หรือวิธีการทำ Undersampling คือเทคนิคการลดข้อมูลจำนวนตัวอย่าง ถ้าข้อมูล

ของ Class ที่มีจำนวนมากกว่าก็ให้ลดข้อมูลให้มีจำนวนข้อมูลเท่ากับ Class ที่มีจำนวนน้อยกว่า หรือวิธีการทำ Ensemble Resampling คือ หลักการใกล้เคียงกับ Random Forest โดยใช้เทคนิค Bagging แบบสุ่มข้อมูลตัวอย่างขึ้นมาฝึกสอนแบบจำลองหลายๆ แบบจำลองพร้อมกัน

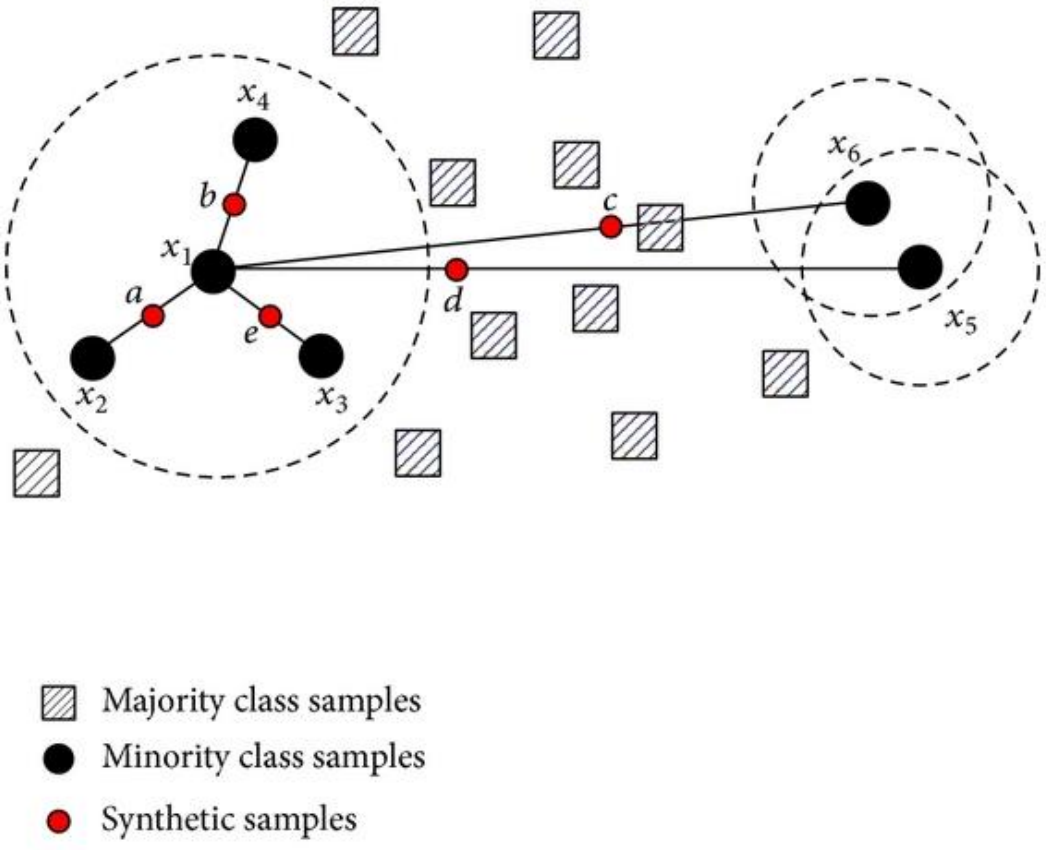
แบบที่ 2 เป็นการเปลี่ยนกระบวนการของฝึกสอนแบบจำลองที่ซับซ้อนกว่า เช่น ในการฝึกสอนแบบจำลองการเรียนรู้ของเครื่องหรือขั้นตอนของ Fit Model โดยการปรับพารามิเตอร์ของ Loss Function ของแบบจำลอง โดยไปเพิ่มค่า Weight ให้กับข้อมูลฝึกสอนแบบจำลอง โดยไป Weight ให้กับข้อมูลตัวอย่างของ Class ที่มีข้อมูลจำนวนน้อยกว่าให้มีค่า Weight เพิ่มมากขึ้น จะทำให้เหมือนกับข้อมูลมีความสมดุลกัน

2.1.2 Oversampling with SMOTE

SMOTE เป็นที่นิยมในขั้นตอนของ Oversampling Method แนวคิดหลักคือการสร้างกลุ่มข้อมูลตัวอย่างของชนกลุ่มน้อยใหม่ โดยการสร้างข้อมูลใหม่แทรกและเลือกเพื่อนบ้านที่เป็นชนกลุ่มน้อยที่อยู่ใกล้ โดยมีขั้นตอนดังนี้ ประการแรกเริ่มจากในแต่ละชนกลุ่มน้อย Minority Class : x จะได้รับ k เพื่อนบ้านที่ใกล้ที่สุดจากกลุ่มชนกลุ่มน้อยอื่นๆ ประการที่สองเราเลือกกลุ่มตัวอย่างมีในภาพประกอบ 1 ของชนกลุ่มน้อยหนึ่งกลุ่ม $\tilde{x}$ ในขอบเขตของ k เพื่อนบ้าน (K Neighbors) สุดท้ายสร้างตัวอย่างสังเคราะห์ x_{new} โดยการแทรกข้อมูลระหว่าง x และ $\tilde{x}$ โดยที่ rand (0,1) อ้างอิงถึงการสุ่มตัวเลขระหว่าง 0 และ 1 ดังในสมการที่ 1 (Jason Brownlee, 2021)

$$x_{new} = x + rand(0,1) \times (\tilde{x} - x) \quad (1)$$

SMOTE มีหลักการทำงานโดยใช้อัลกอริทึม K Nearest Neighbors เพื่อสร้างข้อมูลใหม่ โดยเริ่มต้นการทำงานด้วยการเลือกข้อมูลแบบสุ่มจากคลาสของข้อมูลชนกลุ่มน้อย จากนั้นตั้งค่า k-Nearest Neighbors เพื่อหาเพื่อนบ้านที่ใกล้ที่สุดจากข้อมูล จากนั้นข้อมูลสังเคราะห์จะสร้างขึ้นระหว่างข้อมูลสุ่มและเพื่อนบ้านที่ใกล้ที่สุดจากจำนวน k ที่เลือกแบบสุ่ม ขั้นตอนนี้ทำงานซ้ำหลายครั้งมากพอจนกระทั่งคลาสของข้อมูลชนกลุ่มน้อยมีสัดส่วนเดียวกับคลาสของข้อมูลชนกลุ่มใหญ่ ดังในรูปภาพประกอบ 1



ภาพประกอบ 1 แสดงหลักการทำงานของ SMOTE (Wei-Chiang Hong, November 11, 2013)

2.1.3 Random Undersampling

การสุ่มลบข้อมูลตัวอย่างในคลาสส่วนใหญ่ (Majority Class) เป็นการสุ่มเลือกตัวอย่างจากคลาสส่วนใหญ่ (Majority Class) และลบออกจากชุดข้อมูลการฝึกอบรม จนกว่าจะถึงการกระจายที่สมดุลแบ่งคลาสที่เท่ากัน สิ่งนี้มีผลในการลดจำนวนตัวอย่างในคลาสส่วนใหญ่ในชุดข้อมูลการฝึกอบรมของเวอร์ชันที่แปลงแล้ว กระบวนการนี้สามารถทำซ้ำได้จนกว่าจะได้การแจกแจงคลาสที่ต้องการ เช่น จำนวนตัวอย่างที่เท่ากันสำหรับแต่ละคลาส ข้อจำกัดของการสุ่มตัวอย่างแบบ Random Undersampling คือตัวอย่างจากคลาสส่วนใหญ่จะถูกลบซึ่งอาจเป็นประโยชน์มีความสำคัญหรืออาจมีความสำคัญต่อขอบเขตการตัดสินใจที่มีประสิทธิภาพ เนื่องจากตัวอย่างจะถูกลบแบบสุ่มจึงไม่มีวิธีใดที่จะตรวจจับหรือเก็บรักษา "ข้อมูลที่ดี" หรือตัวอย่างที่มีข้อมูลมากขึ้นจากคลาสส่วนใหญ่

2.1.4 Transformation Pipelines

Scikit-Learn มี Pipeline เพื่อช่วยกำหนดจัดลำดับของขั้นตอนการ Transformations โดยการทำงานของ Pipeline จะเรียกว่า ตามลำดับของ Transform ทั้งหมด ผลลัพธ์จากลำดับของการ Transform แต่ละครั้งเรียกว่าค่าพารามิเตอร์และจะถูกส่งต่อไปจนถึงการ Transform สุดท้าย การจัดลำดับจะมีประโยชน์ทำให้ไม่เกิดการรั่วไหลของข้อมูล (Data Leakage) ในระหว่างการหาพารามิเตอร์ของแบบจำลองที่ดีที่สุดและการเลือกแบบจำลอง

2.2 ทฤษฎีเกี่ยวกับอัลกอริทึม

คำศัพท์ (Terminology)

Data Instance หมายถึงเนื้อหาของข้อมูลที่เก็บอยู่ในโครงสร้างข้อมูล หรือเป็นจำนวนแถว (Row) ที่ 0, 1, 2, ไล่ไป ในแต่ละ instance ซึ่งประกอบด้วย

Attributes คือรายละเอียดของข้อมูลใน Entity หมายถึง สิ่งที่ผู้ใช้งานฐานข้อมูลต้องการจะจัดเก็บ ซึ่งอาจจะเป็นสิ่งที่เป็นรูปธรรมคือสามารถมองเห็นได้ด้วยตา หรือตัวแปรที่มีหน้าที่เก็บค่า (Value) ต่างๆ ที่กำหนด ที่ประกอบด้วย คอลัมน์ (Column) หรือฟิลด์ (Field)

Features หรือคุณลักษณะ ประกอบด้วย 2 ส่วน คือ Independent Variables เป็นข้อมูลตัวแปรอิสระหรือเรียกว่าตัวพยากรณ์ (X) และ Dependent Variables เป็นข้อมูลตัวแปรตาม (y)

Feature Vector มีที่มาจาก Features โดยเอาข้อมูล Data Instance ในแต่ละจุด Match เข้าไปใน Features Space ในการสร้างจุดและสร้าง Coordinate ขึ้นมา ซึ่งจำนวนแกนเท่ากับจำนวน Features ที่ข้อมูลมี แล้วจากนั้นเอามาสร้างจุด Match แต่ละจุด โดยใช้วิธีการ Match Data ในแต่ละข้อมูลจุดใหม่ที่เข้ามาเป็นจุด Vector Space ข้อมูลจุดที่ถูก Match เข้าไปที่จุด จุดนี้เป็นตัวแทนของข้อมูล Data Instance ในแต่ละค่า

Label เป็นคำตอบของกลุ่มข้อมูลที่เรากำหนดไว้แล้ว ว่ากลุ่มนี้มีความหมายว่าอะไร เพื่อบอกให้คอมพิวเตอร์ฝึกสอนและจดจำว่าสิ่งนี้คืออะไรและ Class เป็นกลุ่มของชุดคำตอบที่เป็นไปได้

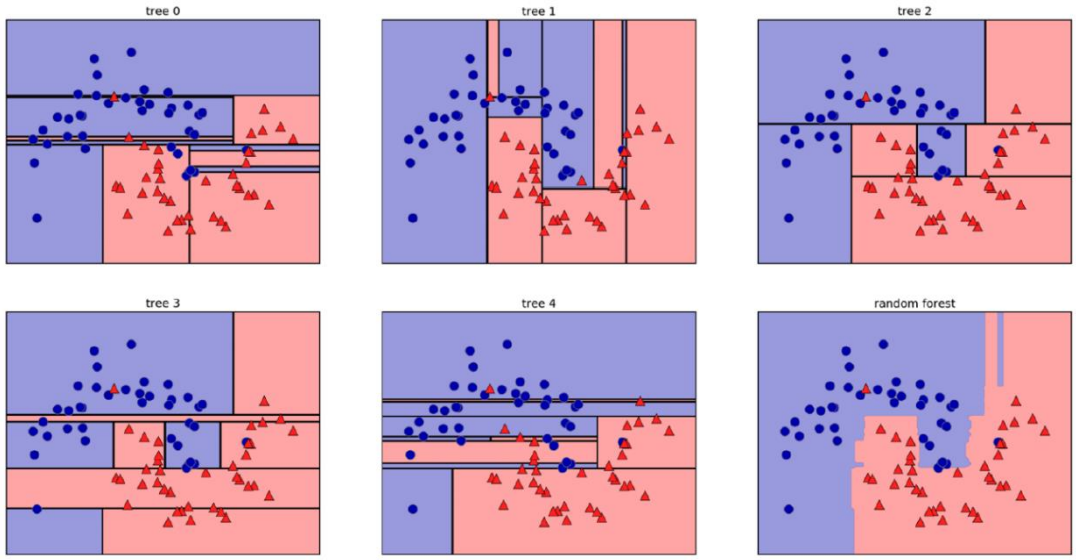
ในงานวิจัยการสร้างแบบจำลองการทำนายความรุนแรงของผู้บาดเจ็บจากอุบัติเหตุในช่วงเทศกาลปีใหม่ ซึ่งเป็นปัญหาแบบการจำแนกประเภทแบบไบนารี ได้เลือกอัลกอริทึมการเรียนรู้ของเครื่องที่ใช้คือ Random Forests ซึ่งแสดงให้เห็นว่ามีประสิทธิภาพในการทำนายสูง

2.2.1 Random Forest

อัลกอริทึม Random Forest เป็นเทคนิคที่ใช้การทำของ Bootstrap Aggregation (Bagging) กับ การเลือกคุณลักษณะแบบสุ่ม โดยที่ Pick Bootstrap เลือกมา 1 Features ที่

สามารถตัดข้อมูลได้ดีที่สุด จากการสุ่มเลือก Features ที่ได้มาจากเซตของการสุ่มของ Features ที่สุ่มขึ้นมา ในขั้นตอนก่อนเลือก Features สมมุติว่ามี Features ทั้งหมด 100 อัน วิธีการแบ่งต้นไม้ (Split Tree) แบบปกติ การเลือก Features มา 1 ครั้ง อาจเป็น Features ที่ 7 แล้วที่ได้ค่า Gini หรือการนำไปคำนวณค่า Entropy ที่ดีที่สุด สำหรับการแบ่งข้อมูล Split วิธีนี้เป็นแบบเดิมที่เลือก Features ที่ดีที่สุดมา 1 อัน จากจำนวนทั้งหมด 100 Features

แต่สำหรับ Random Forest จาก Features ทั้งหมด 100 อัน เลือกสุ่มมา 40 Features เพราะฉะนั้น Features ที่เลือกมาอาจเป็น Feature ที่ 3, 10, 11, 20, 50 และ 67 ไปจนครบทั้งหมด 40 Features แล้วเอาไปคำนวณค่า Gini ดูว่า Partition ไหนดีที่สุด โดยใช้ Feature อันที่ทำให้ดีที่สุดจากการสุ่ม 40 Features นี้ แทน 100 Features การ Split แต่ละครั้งจะเอาชุดข้อมูลของ Feature ที่แตกต่างกันมา ลักษณะนี้เป็นการใส่ Random Rate เข้าไปในอัลกอริทึม Random Forest



ภาพประกอบ 2 แสดงการทำงานของอัลกอริทึม Random Forest (Andreas C.Muller & Sarah Guido, October, 2016)

จากภาพประกอบ 2 มีต้นไม้ทั้งหมดจำนวน 5 ต้น สุดท้ายรวมกันได้ผลลัพธ์โหวดของต้นไม้การตัดสินใจทั้ง 5 ออกมา

Random Forest Error Rate คือ Correlation ระหว่างต้นไม้การตัดสินใจ ถ้ามากแสดงว่าเชิญผู้เชี่ยวชาญคล้ายๆ กัน ทำให้ไม่มีประสิทธิภาพ เพราะฉะนั้นถ้าต้องการต้นไม้การตัดสินใจ แต่ละต้นที่มีค่า Correlation น้อย และในแต่ละต้นไม้การตัดสินใจ ต้องมี Bias น้อย

เพราะจะทำให้แบบจำลองการทำนายทำงานได้อย่างมีประสิทธิภาพ ต้นไม้การตัดสินใจแต่ละต้นจะต้องมีค่า Error Rate น้อยลง

การ Tuning ของอัลกอริทึม Random Forests ประกอบด้วยค่าพารามิเตอร์ดังนี้

Parameter : Max Features คือจำนวนสูงสุดของจำนวน Feature ที่สุ่มออกมาของแต่ละ Split ถ้ามีจำนวนมาก จะไม่มีประสิทธิภาพ เพราะจำนวน Feature ที่เลือกมาในแต่ละ Split มีผล ถ้า Feature มีค่าจำนวนมาก คือถ้ามี 100 Feature ถ้าเลือกจำนวน Max Features เท่ากับ 99 Features จากทั้งหมด 100 Features ในครั้งที่ 1 กับ 2 ผลลัพธ์ที่ได้จะไม่ต่างกัน แต่มีข้อดีอยู่ที่ถ้าหากใช้ Max Feature จำนวนมาก ทำให้เชื่อมั่นได้ว่าใน Features ที่เลือกมาในแต่ละ Split จะมีให้เลือกได้จำนวนมากกว่า ทำให้โอกาสที่จะได้ต้นไม้การตัดสินใจที่มีประสิทธิภาพสูงมากกว่า เช่น 90 Features มีโอกาสที่จะหา Features ที่ดีที่สุดได้มากกว่า 30 Features

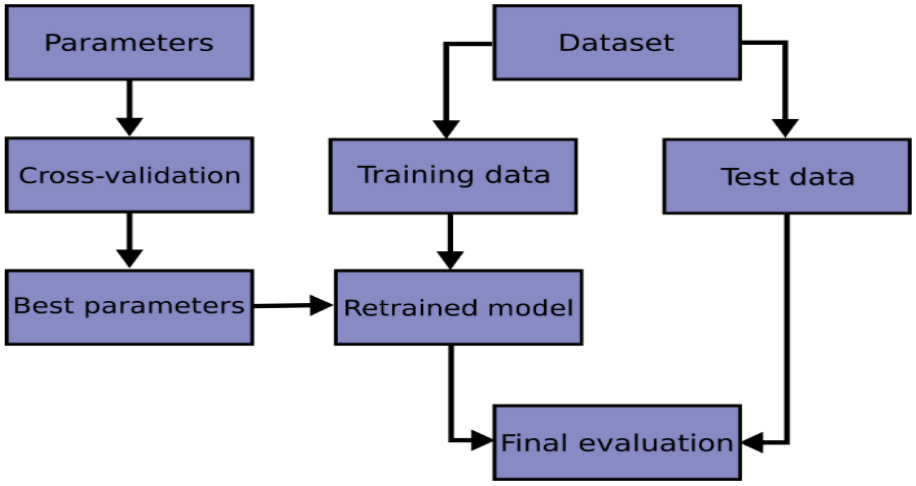
Parameter : n estimators คือ จำนวนต้นไม้ (Tree) ทั้งหมดที่ใช้ในโมเดล ถ้ามากกว่า 100 ประสิทธิภาพการทำงานมักค่อนข้างดี

Parameter : max depth คือ จำนวนลำดับชั้นของต้นไม้ (Tree) แต่ละต้น

Parameter : learning rate คือ ค่า Lambda ที่น้อยกว่าหรือเท่ากับ 1 สามารถช่วยลด Variance ลงได้

2.2.2 Grid Search

Grid Search CV ของ Scikit-Learn ช่วยในการหาค่าพารามิเตอร์ที่ดีที่สุดหรือการหาค่า Best Parameters เพื่อปรับแต่งแบบจำลองการทำนายให้ดีขึ้น สำหรับการฝึกฝนแบบจำลองเพื่อช่วยลดเวลาในการเปลี่ยนค่าพารามิเตอร์ไปทีละค่าในการปรับพารามิเตอร์ สิ่งที่แบบจำลองต้องการคือต้องการให้พารามิเตอร์ตัวใดทดลองใช้และช่วงของค่าใดที่จะลองใช้ และจะประเมินค่าของพารามิเตอร์ที่ดีที่สุดที่เป็นไปได้ทั้งหมดของค่าพารามิเตอร์ตัวนั้น



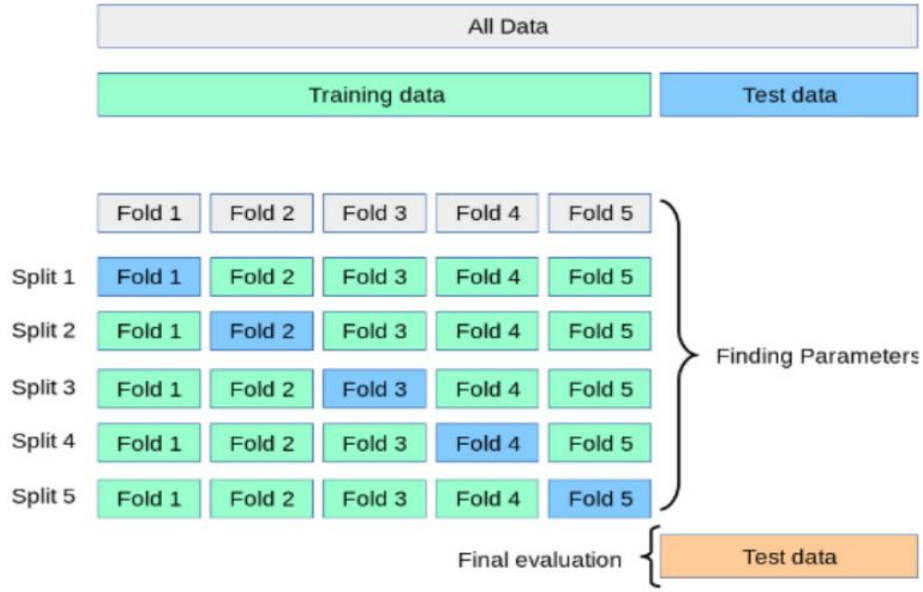
ภาพประกอบ 3 แสดงการทำงานของ Grid Search และ Cross-validation (J. Mach. Learn. Res, 2010)

2.2.3 การเลือกพารามิเตอร์ของแต่ละอัลกอริทึม

ในการวัดประสิทธิภาพของโมเดลว่าสามารถทำงานได้ดีแค่ไหน ควรจะเลือกโมเดลแบบไหนถึงจะดี บางโมเดลก็เหมาะกับข้อมูลบางชนิด ในบางสถานการณ์ เพราะฉะนั้นต้องเลือกโมเดลที่เหมาะสม ไม่ซับซ้อน และทำงานได้เร็ว

ซึ่งงานวิจัยนี้การประเมินวัดความแม่นยำและประสิทธิภาพของการทำนาย ตามปัญหาของ Classification แบบ Binary Classification สามารถอธิบายได้ดังนี้

Cross Validation เป็นขั้นตอนในการประเมินความแม่นยำของอัลกอริทึม (Model) เพื่อป้องกันปัญหา Overfit โดยการพยายามใช้ข้อมูลให้มากที่สุด โดยพยายามแบ่งเป็นข้อมูลเป็นทั้ง Training และข้อมูล Validation สลับเปลี่ยนกันไปจนครบแล้ววนซ้ำหลายครั้งตามจำนวน K-Fold ที่เราทำการแบ่งให้ทำงาน ทำให้โอกาสของ Model ได้เจอข้อมูลที่หลากหลาย เพราะข้อมูลทั้งหมดเป็นได้ทั้งข้อมูล Training และ Validation



ภาพประกอบ 4 แสดงการทำงานของ Cross Validation (J. Mach. Learn. Res, 2010)

โดยที่ขั้นตอนแรกจะทำการแบ่งข้อมูลออกเป็น Training Data และ Test Data โดยที่ Test Data เราจะเอาไว้ทดสอบขั้นตอนสุดท้าย แล้วนำข้อมูล Training Set มาแบ่งเป็น Fold แล้วในแต่ละครั้งของการ Split นำมาสร้างส่วนหนึ่งเป็นข้อมูล Validation และที่เหลือเป็นข้อมูล Training Set แล้วจะสลับเปลี่ยนกันตามจำนวนรอบของ Split ที่เราแบ่งข้อมูลตาม Fold ที่เรากำหนด ซึ่งปกตินิยมแบ่งจำนวน Fold เป็น 5 หรือ 10 ถ้ายิ่งเรากำหนดจำนวน Fold มากขึ้นเท่าไรก็ยิ่งประมวลผลได้นานขึ้น ดังภาพประกอบ 4

2.2.4 การวัดประสิทธิภาพ

ในการวัดประสิทธิภาพของแบบจำลองการทำนายว่าสามารถทำงานได้อย่างมีประสิทธิภาพ บางแบบจำลองการทำนายก็เหมาะสมกับข้อมูลบางชนิด ในบางสถานการณ์เพราะฉะนั้นต้องเลือกมาตรวัดที่เหมาะสม ซึ่งงานวิจัยนี้การประเมินวัดความแม่นยำและประสิทธิภาพของการทำนาย สามารถอธิบายได้ดังนี้

แบบจำลองการทำนายของการจำแนกประเภทในแบบไบนารี การวัดความแม่นยำและประสิทธิภาพของโมเดล ถ้าใช้ค่าความถูกต้องของ Accuracy อย่างเดียวอาจไม่มีประสิทธิภาพพอ เช่น ถ้าข้อมูลมีปัญหาของความไม่สมดุลของข้อมูล (Imbalance Data) ถ้าข้อมูล 2 กลุ่มต่างกันจำนวนมาก ยกตัวอย่างการทำนายดาวหางจะตกทุก 70 ปี เมื่อเวลาผ่านไป 20 ปี ทำนายถูกต้อง 100 เปอร์เซ็นต์ แต่อาจทำนายผิดไปแค่ 1 ครั้ง ที่ดาวหางตก ทำให้ Accuracy สูง

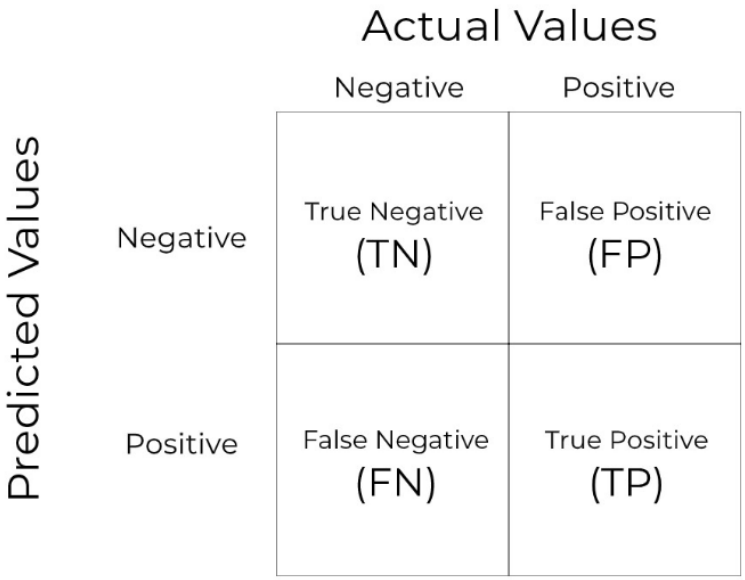
เกือบ 100 เปอร์เซ็นต์ ดังนั้นการวัดความแม่นยำและประสิทธิภาพของแบบจำลองในปัญหาแบบการจำแนกประเภทจะประกอบด้วยดังต่อไปนี้

Confusion Matrix เป็นการสร้างผลของการทำนายออกมาเป็น

Positive Class เป็นคลาสของข้อมูลที่สนใจ ต้องการทราบ ต้องการหา เช่น สนใจกลุ่มของคนที่เสียชีวิตจากอุบัติเหตุในช่วงเทศกาลปีใหม่ เป็นต้น

Negative Class เป็นคลาสของข้อมูลที่ไม่ได้สนใจ เช่น กลุ่มของคนที่ไม่เสียชีวิตจากอุบัติเหตุในช่วงเทศกาลปีใหม่ เป็นต้น

สามารถสลับเปลี่ยนกันได้ตามวัตถุประสงค์ เพราะเป็นนิยามว่าสนใจกลุ่มของข้อมูลกลุ่มไหน และตาราง Confusion Matrix เป็นเครื่องมือที่สำคัญในการประเมินวัดความแม่นยำและประสิทธิภาพของแบบจำลองที่ทำนายกับข้อมูลที่เกิดขึ้นจริง ดังภาพประกอบ 5



ภาพประกอบ 5 แสดงตาราง Confusion Matrix (Kevin Markham, 2014)

True Positive (TP) คือข้อมูลที่ทำการทำนายตรงกับสิ่งที่เกิดขึ้นจริง คือทำนายว่าจริงและสิ่งที่เกิดขึ้นก็คือจริง เช่น มีผู้บาดเจ็บจากอุบัติเหตุที่เสียชีวิตจริง และแบบจำลองทำนายว่าผู้บาดเจ็บจากอุบัติเหตุเสียชีวิต

True Negative (TN) คือข้อมูลที่ทำการทำนายตรงกับสิ่งที่เกิดขึ้นจริง คือแบบจำลองทำนายว่าไม่จริงและสิ่งที่เกิดขึ้นก็คือไม่จริง เช่น มีผู้บาดเจ็บจากอุบัติเหตุที่ไม่เสียชีวิตจริง และแบบจำลองทำนายว่าผู้บาดเจ็บจากอุบัติเหตุที่ไม่เสียชีวิต

False Positive (FP) คือข้อมูลที่แบบจำลองทำการทำนายแล้วไม่ถูกต้องกับสิ่งที่เกิดขึ้นจริง คือแบบจำลองทำนายว่าจริง แต่สิ่งที่เกิดขึ้นก็คือไม่จริง เช่น มีผู้บาดเจ็บจากอุบัติเหตุที่ไม่เสียชีวิตจริง แต่แบบจำลองทำนายว่าผู้บาดเจ็บจากอุบัติเหตุเสียชีวิต ซึ่งจริงแล้วต้องอยู่ในกลุ่มของ True Negative

False Negative (FN) คือข้อมูลที่แบบจำลองทำการทำนายแล้วไม่ถูกต้องกับสิ่งที่เกิดขึ้นจริง คือทำนายว่าไม่จริง แต่สิ่งที่เกิดขึ้นก็คือจริง เช่น ผู้บาดเจ็บจากอุบัติเหตุที่เสียชีวิตจริง แต่แบบจำลองทำนายว่าผู้บาดเจ็บจากอุบัติเหตุที่ไม่เสียชีวิต ซึ่งจริงแล้วต้องอยู่ในกลุ่มของ True Positive

เราสามารถใช้ข้อมูลของ Confusion Matrix ในตารางที่แทนด้วยค่าความถี่มาคำนวณวัดความแม่นยำและประสิทธิภาพของโมเดล ในรูปแบบต่างๆ ดังนี้

Accuracy คือ ความถูกของของกลุ่มที่ทำนายถูกต้องกับการทำนายทั้งหมด

$$Accuracy = (TP + TN)/(TP + TN + FP + FN) \quad (2)$$

Recall หรือ True Positive Rate (TPR) หรือ Sensitivity คือ การเทียบอัตราส่วนการทำนายที่ถูกต้องต่อจำนวนของที่เป็นจริงทั้งหมด หรือการที่เราเอาตัวจริงออกมาได้หรืออัตราส่วนที่ไปเลือกตัวจริง ได้มามากแค่ไหน หรือ อัตราการหา True Positive เทียบกับ True Positive ทั้งหมด ตัวอย่างเช่น สมมุติทำนายผู้บาดเจ็บจากอุบัติเหตุที่เสียชีวิต โดยเลือกผู้บาดเจ็บจากอุบัติเหตุมา 1,177 จาก 2,000 คน แล้วทำนายว่าผู้บาดเจ็บจากอุบัติเหตุจะมีโอกาสเสียชีวิต ส่วนที่เหลือ 823 คนจะไม่เสียชีวิตจากอุบัติเหตุ และเมื่อเทียบค่าจริง ผู้บาดเจ็บที่เลือกมา 1,177 คน เป็นผู้บาดเจ็บจากอุบัติเหตุที่เสียชีวิตทั้งหมดเลย ค่า Recall จะเป็น 1 แบบจำลองมีค่าระลึกที่สูงมาก เพราะสามารถเรียกสิ่งที่สนใจออกมาได้

$$Recall = TP/(TP + FN) \quad (3)$$

Precision คือ ค่าที่สนใจผลการทำนาย (Prediction) ว่าเป็นกลุ่มที่สนใจเท่าไร ก็คือกลุ่มของ Positive ผลที่ทำนายถูกเท่าไร เช่น แบบจำลองทำนายว่า 1,177 คน เป็น Positive ที่นี้จะไปดูว่า 1,177 คน ทำนายถูกต้องแค่ไหน ถ้าหากถูก 1,000 จาก 1,177 คน ถ้าหากค่า Recall เท่ากับ 1 แต่ค่า Precision จะบอกสะท้อนบอกได้ว่าการทำนายของแบบจำลองนั้นมีประสิทธิภาพ

แค่ไหน และถ้าอยากได้ค่า Precision สูงๆ คือทำให้ FP มีค่าเป็น 0 ค่า Precision จะมีค่าเป็น 1 ได้เลย แต่ในความเป็นจริงค่า FP เป็น 0 ได้ จะต้องตั้งมาตราฐานให้สูงไว้ ในการคัดเลือกกลุ่มประชากรตัวอย่างที่เป็น Positive

$$Precision = TP/(TP + FP) \quad (4)$$

False Positive Rate (FPR) คือ ค่าของอัตราการทำนายผิดเทียบกับข้อมูล Negative ทั้งหมด

$$False\ positive\ rate\ (FPR) = FP/(TN + FP) \quad (5)$$

F-Score คือเป็นค่าที่สะท้อนค่า Recall กับ Precision พร้อมกัน เป็น General Term ที่ไม่ได้ระบุค่าเบต้า (Beta) เป็นการให้ค่าน้ำหนัก (Weight) ค่า Precision กับ Recall ว่า F ตัวนี้จะสะท้อนค่าอะไรมากกว่ากัน ถ้าค่าเบต้าเท่ากับ 0.5 จะสะท้อน Precision ออกมามากกว่าและค่าเบต้าเท่ากับ 2 จะสะท้อนค่า Recall ออกมามากกว่า Precision สำหรับ F1-score คือ F-score ที่ระบุค่าเบต้าเท่ากับ 1

$$F_1 = 2\ .\ (Precision \times Recall\ )/(Precision + Recall) \quad (6)$$

Precision-Recall Curve จะแสดงให้เห็นของค่า Precision และ Recall มีการ Tradeoff กัน โดยมีสาเหตุมาจากค่า Precision สูง และค่า Recall จะน้อยลง หรือ Precision น้อย และค่า Recall จะสูง ซึ่งขึ้นอยู่กับการตั้งค่า Threshold ในการเลือกกลุ่ม Positive ที่สนใจ ถ้าตั้งค่า Threshold สูงมาก โอกาสที่จะได้กลุ่ม Positive ก็จะน้อยลง เพราะการตั้งมาตรฐานไว้สูงมาก ก็มีโอกาสทำให้พลาดในการเลือกกลุ่มที่ควรจะเป็น Positive แต่ปล่อยกลุ่มพวกนี้ไปเก็บมาไม่ครบ กรณีนี้จะทำให้ False Negative เพิ่มขึ้น และถ้าค่า Precision สูง แต่ค่า Recall จะลดลง แต่ถ้าค่า Recall สูง ค่า Precision จะลดลง

Receiver Operating Characteristic Curve (ROC Curve) เป็นวิธีที่นิยมใช้ในการวัดประสิทธิภาพความถูกต้องของการพยากรณ์ พื้นที่ใต้กราฟ ROC ที่มีค่ามากแสดงว่าแบบจำลองนั้นมีผลการทำนายที่ถูกต้องมาก ROC Curve เป็นกราฟที่มีความสัมพันธ์ระหว่างแกน

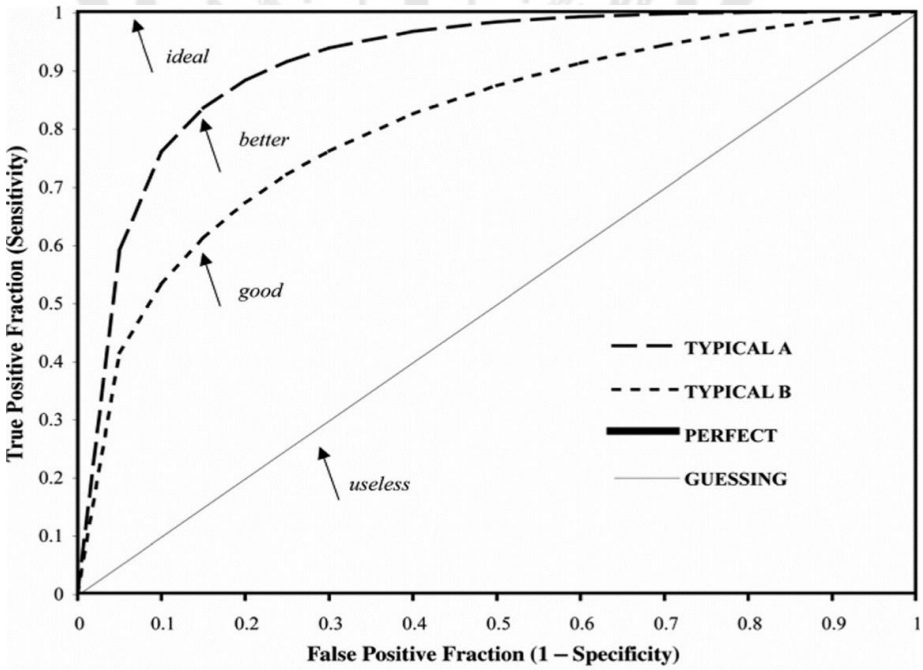
Y แทน Sensitivity (True Positive Rate) กับแกน x แทน 1-Specificity (False Positive Rate) ดังในภาพประกอบที่ 6 ถ้าค่าของ Sensitivity และ 1-Specificity มีค่าสูงกราฟ ROC Curve จะโค้งเข้าหามุมซ้ายบน ซึ่งเป็นผลการทำนายที่มีประสิทธิภาพสูง) ดังในภาพประกอบที่ 6 หรือพื้นที่ใต้กราฟโค้ง ROC Curve เป็นการบ่งชี้ความถูกต้องหรือความน่าเชื่อถือของตัวแบบจำลองการทำนาย ถ้าแบบจำลองการทำนายใดที่มีพื้นที่ใต้กราฟ ROC Curve (ROC AUC) มากสุดถือว่ามีประสิทธิภาพมากสุด

Sensitivity หรือ True Positive Rate (TPR) คือค่า Recall

$$Sensitivity = \frac{TP}{TP+FN} \tag{7}$$

Specificity หรือ True Negative Rate (TNR) คือ อัตราส่วนของผลการทำนายของกลุ่มข้อมูลที่ไม่สนใจ (Negative Class) ที่ทำนายถูกต้อง ต่อจำนวนเหตุการณ์ทั้งหมดที่เป็นกลุ่มข้อมูลที่ไม่สนใจ (Negative Class) ดังในรูปภาพประกอบ 8

$$Specificity = \frac{TN}{FP+TN} \tag{8}$$



ภาพประกอบ 6 แสดงแนวคิดของ Receiver Operating Characteristic Curve (Tourassi, 2018)

2.3 งานวิจัยที่เกี่ยวข้อง

การทบทวนวรรณกรรมของงานวิจัยนี้ได้ทำการศึกษาค้นคว้างานวิจัยที่เกี่ยวข้องกับการทำนายอุบัติเหตุการจราจรบนทางถนน โดยงานวิจัยที่เกี่ยวข้องมีรายละเอียดดังนี้

(1) บทความวิจัยเรื่อง The Traffic Accident Hotspot Prediction : Based on The Logistic Regression Method โดย Tao Lu, Yan Lixin, Zhu Dunyao, Zhang Pan China (Lu, Dunyao, Lixin, และ Pan, 2015)

งานวิจัยนี้ได้ทำการศึกษาการทำนายฮอตสปอตของอุบัติเหตุจราจรบนท้องถนน เพื่อระบุจุดเสี่ยงที่มีโอกาสเกิดอุบัติเหตุจราจรในระหว่างการขับขี่ เพื่อปรับปรุงความปลอดภัยในการขับขี่ยานพาหนะและประสิทธิภาพการจราจรผ่านการเตือนล่วงหน้า โดยใช้ข้อมูลจาก 400 ชุดข้อมูลอุบัติเหตุของถนนสายสำคัญ 10 สายในเมืองปักกิ่งจากปี 2004 ถึงปี 2007 โดยใช้ตัวแปรอิสระทั้งหมด 6 ตัว ดังนี้ 1). ข้อมูลรถในเส้นถนนตัดขวาง ประกอบด้วย มอเตอร์เลน, เลนจักรยาน, ทางเท้า, ทางม้าลาย 2). ข้อมูลทางหลวงและถนนในชนบทและในเมือง ทางหลวงประกอบด้วย ทางด่วนทางหลวงชั้นหนึ่ง, ทางหลวงชั้นสอง, ทางหลวงชั้นสาม, ทางหลวงชั้นที่สี่และถนนในชนบท ถนนในเมืองประกอบด้วย ทางด่วนในเมืองถนนสายหลักและถนนสายรอง 3). ข้อมูลสภาพพื้นผิวถนน 4). ข้อมูลสภาพการมองเห็นที่ดี คือ สำคัญอย่างยิ่งสำหรับการขับขี่อย่างปลอดภัย ในภาพที่ไม่ดีการตัดสินใจของผู้ขับขี่จะได้รับผลกระทบอย่างรุนแรง 5). ข้อมูลสภาพของยานพาหนะ 6). ข้อมูลสถานะของผู้ขับขี่ ประกอบด้วยการเมาแล้วขับ ความเมื่อยล้า การกระทำผิดกฎหมายจราจรเป็นต้น และ Label คือ จะเกิดอุบัติเหตุหรือไม่เกิดอุบัติเหตุจราจร

แล้วทำการพัฒนาแบบจำลองการทำนายฮอตสปอตของอุบัติเหตุจราจรบนท้องถนนโดยใช้ Logistic regression ผลการวิจัยพบว่าแบบจำลองนี้สามารถให้ผลที่ดีการทำนายฮอตสปอตของอุบัติเหตุจราจรบนท้องถนน ด้วยอัตราที่ถูกต้อง 86.67% สามารถใช้ผลการศึกษานี้ได้ในการช่วยเหลือความปลอดภัยของยานพาหนะในการขับขี่และแจ้งเตือนล่วงหน้าและข้อเสนอสำหรับการขับขี่อย่างปลอดภัย

(2) บทความวิจัยเรื่อง Live Prediction of Traffic Accident Risks Using Machine Learning and Google Maps โดย Meraldo Antonio (Antonio, 2019)

งานวิจัยนี้ได้ทำการศึกษาการทำนายอุบัติเหตุจราจรโดยใช้ Scikit-Learn, Google Maps API, Dark Sky API, Flask และ PythonAnywhere โดยใช้ข้อมูลจาก Kaggle ชุดข้อมูลนี้มีรายละเอียดเกี่ยวกับอุบัติเหตุทางรถยนต์ที่เกิดขึ้น 1.6 ล้านครั้ง ที่ ในสหราชอาณาจักร ระหว่างปี 2543 ถึง 2557 ประกอบด้วย 33 คอลัมน์ ที่รวบรวมรายละเอียด เช่น สถานที่เวลาความรุนแรงของ

อุบัติเหตุ รวมถึงฉากอุตุนิยมวิทยาและการจราจรต่างๆ และข้อมูลจาก Dark Sky เป็นข้อมูลจากสภาพอากาศ มีการเตรียมข้อมูลมีการใช้เทคนิคการจัดกลุ่มของข้อมูลโดยใช้อัลกอริทึม DBSCAN Clustering ในการจัดกลุ่มในพื้นที่ที่มีรัศมี 25 เมตร และมีอุบัติเหตุเกิดขึ้นในรัศมีนั้นไม่น้อยกว่า 14 ครั้ง ส่วนข้อมูลที่อยู่นอกกลุ่มจะไม่รวมอยู่ในการวิเคราะห์ครั้งนี้

การพัฒนาแบบจำลองการทำนายแบ่งชุดข้อมูลของผู้วิจัยในอัตราส่วน 70:30 เป็นชุดข้อมูลฝึกสอนและทดสอบ โดยใช้ Classification Models ประกอบด้วย Support Vector Machine (SVM), Logistic Regression และ Random Forest และวัดประสิทธิภาพและเปรียบเทียบความแม่นยำโดยใช้ Accuracy และ AUC ซึ่งแบบจำลองที่ทดสอบมีประสิทธิภาพดีที่สุดคือ Random Forest โดยใช้ 25 Features ที่เป็น Numerical Features ได้ Accuracy เท่ากับ 83.47 และ AUC เท่ากับ 87.05

การติดตั้งการใช้งานของโมเดลไปยังเว็บแอปพลิเคชันโดยใช้ Flask และ HTML 5 โดยให้ผู้ใช้งานสามารถใส่ข้อมูลต้นทางและปลายทางของผู้ใช้งานผ่านแผนที่ เมื่อใส่ข้อมูลเหล่านี้ไป แล้วจะถูกสร้างขึ้นเป็น POST request จะถูกส่งไปยังเฟรมเวิร์กเบื้องหลัง ที่นี่ข้อมูลอินพุตเหล่านี้ จะทำหน้าที่เป็นอาร์กิวเมนต์ของฟังก์ชัน Call Google ฟังก์ชั่นนี้จะเรียกว่า Google Maps API ซึ่งจะให้เส้นทางการขับขี่ที่ดีที่สุด ในการเชื่อมต่อทั้งสองเส้นทางจากต้นทางไปยังปลายทางและทำเครื่องหมายจุดตัดของถนนที่มีความเสี่ยงที่อาจเกิดอุบัติเหตุขึ้นได้

(3) บทความวิจัยเรื่อง Using Machine Learning to Predict Car Accident Risk โดย Daniel Wilson (Wilson, 2018)

งานวิจัยนี้ได้ทำการศึกษาการเรียนรู้ของเครื่อง (Machine Learning) มาช่วยชีวิตคนโดยการทำนายอุบัติเหตุบนเส้นทางถนนในรัฐยูทาห์ของสหรัฐอเมริกา โดยใช้ข้อมูลจาก NOAA ทั้งหมด 7 ปี ใน 2012 ถึงปี 2019 ประกอบด้วย Feature ของข้อมูลสภาพอากาศ เช่น อุณหภูมิ ความเร็วของลม ทัศนวิสัยการมองเห็น ปริมาณน้ำฝน หิมะตก น้ำแข็งปกคลุม ความลึกของหิมะ และข้อมูลของ ความเร็วจำกัดในท้องถนน ความโค้งของถนน ปริมาณ Traffic เฉลี่ย ทางแยกของถนน พื้นผิวถนน และปัจจัยของมนุษย์ เช่น ความหนาแน่นของประชากร ป้ายโฆษณา และข้อมูลของถนน เช่น ถนนที่เป็นศูนย์กลาง ทิศทางการไหลของจราจรบนถนน

ในการเตรียมข้อมูลมีการใช้เทคนิค Undersampling ในการแก้ไขปัญหา Imbalanced Data ในการเลือกในแต่ละ Class ของ Label ให้เท่ากัน

การพัฒนาแบบจำลองการทำนายอุบัติเหตุบนเส้นทางถนน โดยเลือกใช้โมเดลของ Decision Trees และ Gradient Boosting และวัดประสิทธิภาพและการเปรียบเทียบความแม่นยำ

โดยใช้ Accuracy, Precision, Recall และ AUC ซึ่งแบบจำลองที่ทดสอบมีประสิทธิภาพดีที่สุดคือ Gradient Boosting ได้ Accuracy เท่ากับ 68.59 Precision เท่ากับ 31.14 Recall เท่ากับ 89.07 และ AUC เท่ากับ 82.82 โดยที่โมเดลของ Decision Trees มีปัญหา Overfit ง่าย ในการฝึกสอน (Training Model) ไม่ Generalize กับ New Data ในการทำนายของแบบจำลอง แต่ Gradient Boosting ทำงานโดย รวมผลลัพธ์จาก Decision Trees ที่แตกต่างกันและทำงานประมวลผลได้เร็ว

(4) บทความวิจัยเรื่อง Analysis of Accidental Deaths During Songkran Festival Using Data Mining โดย Pornpimol Chaiwuttisak (Chaiwuttisak, 2019)

งานวิจัยนี้ได้ทำการศึกษาการวิเคราะห์การเสียชีวิตจากอุบัติเหตุในช่วงเทศกาลวันหยุดสงกรานต์และพัฒนาโมเดลสำหรับ Classification ของการเสียชีวิตในอุบัติเหตุบนทางถนน เพื่อจำแนกการบาดเจ็บหรือเสียชีวิตจากอุบัติเหตุ ข้อมูลที่ใช้ในการศึกษาวิจัยเป็นข้อมูลของผู้บาดเจ็บและเสียชีวิตจากอุบัติเหตุในระหว่างปี 2551 ถึงปี 2557 จำนวนข้อมูลที่ใช้ 2,875 คน จากข้อมูลเปิดเผยของรัฐบาล (The Digital Government Development Office) ประกอบด้วยข้อมูล 10 คอลัมน์ ได้แก่ ช่วงเวลาของการเกิดอุบัติเหตุ แบ่งเป็นช่วง ช่วงละ 6 ชั่วโมง เพศ อายุ ชนิดถนน สถานะของการขับขี่ ประเภทยานพาหนะ ประเภทยานพาหนะของคู่กรณี การสวมหมวกนิรภัย การดื่มสุรา และคนที่เสียชีวิตหรือบาดเจ็บจากการประสบอุบัติเหตุ

สถิติที่ใช้ในการทดสอบสมมติฐานทดสอบด้วยไคสแควร์ (The Chi-square) กับตัวแปรอิสระ เช่น การดื่มสุรา การไม่สวมหมวกนิรภัย สภาพแวดล้อมทางกายภาพ เช่น เวลาที่เกิดอุบัติเหตุบนท้องถนน ประเภทของถนนที่ทำให้เกิดอุบัติเหตุ และตัวแปรตามคือคนที่เสียชีวิตและบาดเจ็บจากอุบัติเหตุ จากการทดสอบสมมติฐานที่ระดับนัยสำคัญ 0.05 พบว่าตัวแปรทั้งหมดมีความสัมพันธ์กับการเสียชีวิตในช่วงวันหยุดสงกรานต์

นอกจากนี้ยังได้ใช้เทคนิคของ Data Mining มาประยุกต์ใช้ในงานวิจัยนี้ประกอบด้วย Decision Tree, Bayesian Learning, Logistic Regression และ Neural Network ในการระบุการเสียชีวิตจากอุบัติเหตุ อธิบายโดยข้อมูลตารางและเปรียบเทียบความแม่นยำของแต่ละโมเดล Classification ด้วยเทคนิคตาราง Confusion Matrix ประกอบด้วยค่า Accuracy, Precision, Recall และ F-Measure จากผลการวิจัยพบว่า Logistic Regression สามารถจำแนก Classification ได้อย่างถูกต้องสูงกว่าเทคนิคการจำแนกประเภทอื่นๆ ที่มีความแม่นยำ Precision เท่ากับ 72.20%

ในงานวิจัยที่เกี่ยวข้องของเรื่องนี้มีความแตกต่างในเรื่องของข้อมูล เพราะใช้ข้อมูลของอุบัติเหตุในช่วงเทศกาลวันหยุดสงกรานต์และเทคนิคในการจัดการความไม่สมดุลของชุดข้อมูลที่คอลัมน์ของผู้เสียชีวิตหรือบาดเจ็บจากอุบัติเหตุ (Class Label) ซึ่งแน่นอนจำนวนผู้เสียชีวิตต้องน้อยกว่าจำนวนผู้ไม่เสียชีวิตจากอุบัติเหตุหลายเท่าตัว ทำให้เกิดปัญหาความไม่สมดุลของข้อมูลขึ้น

(5) บทความวิจัยเรื่องการพยากรณ์การเสียชีวิตจากอุบัติเหตุทางถนนในช่วงเทศกาลปีใหม่ โดย นายพิเศษ คุปตินทร (คุปตินทร, 2562)

งานวิจัยนี้ได้ทำการศึกษาการศึกษาปัจจัยคุณลักษณะของข้อมูลต่างๆ และแบบจำลองการทำนายการเสียชีวิตจากอุบัติเหตุทางถนนในช่วงเทศกาลปีใหม่

ใช้ข้อมูลการเกิดอุบัติเหตุทางถนนในช่วงเทศกาลปีใหม่ จัดเก็บโดยองค์กรสถาบันการแพทย์ฉุกเฉินแห่งชาติ และเผยแพร่โดยสำนักงานพัฒนารัฐบาลดิจิทัล (EGA) ซึ่งเป็นข้อมูลแบบทุติยภูมิ โดยเป็นข้อมูลที่เกิดอุบัติเหตุทางถนนในช่วงเทศกาลปีใหม่ตั้งแต่ปี 2551 ถึง 2558 ซึ่งเป็นข้อมูลการเกิดอุบัติเหตุทั่วประเทศไทย ข้อมูลประกอบไปด้วย ข้อมูลของด้านประชากร ประกอบด้วย เพศ อายุ และสถานะของผู้ประสบอุบัติเหตุ ข้อมูลด้านสิ่งแวดล้อมประกอบด้วย ช่วงเวลา วันที่เกิดเหตุ จังหวัดที่เกิดเหตุ ประเภทของถนนที่เกิดเหตุ ประเภทยานพาหนะที่เกิดประสบเหตุ รถคู่กรณี อุณหภูมิ ความชื้น ความเร็วลม และสภาพอากาศ ข้อมูลด้านนโยบายสาธารณะประกอบด้วย ดื่มแอลกอฮอล์ มาตรการป้องกัน การนำส่ง และข้อมูลผลการรักษา

ในการสร้างแบบจำลองการพยากรณ์มีการใช้เทคนิคการเรียนรู้ของเครื่อง (Machine Learning) ทั้งแบบมีการสอนเครื่องเรียนรู้ (Supervised Learning) และไม่มีการสอนเครื่องเรียนรู้ (Unsupervised Learning) โดยใช้ Decision Tree และ Naïve Bayes ในการทำนายการเสียชีวิตจากอุบัติเหตุทางถนนในช่วงเทศกาลปีใหม่ และค่าประเมินประสิทธิภาพของแบบจำลองด้วย Accuracy และ F-measure

ในส่วนของการทำการแบ่งกลุ่มข้อมูลจะใช้วิธี K–Mean ในการจัดกลุ่มของข้อมูลตามเพศ เพื่อศึกษาว่าเพศเป็นคุณลักษณะที่มีโอกาสเสียชีวิตจากอุบัติเหตุบนทางถนนในช่วงเทศกาลปีใหม่หรือไม่

ผลการวิจัยแบบจำลองการพยากรณ์ด้วยเทคนิค Decision Tree นั้นค่าความแม่นยำของ Accuracy อยู่ที่ 98.68% และค่าประสิทธิภาพของแบบจำลอง F-measure อยู่ที่ 0.984 ส่วนเทคนิค Naïve Bayes นั้นค่าความแม่นยำของ Accuracy อยู่ที่ 98.65% และค่าประสิทธิภาพของแบบจำลอง F-measure อยู่ที่ 0.982 แสดงให้เห็นว่าแบบจำลองการทำนายการเสียชีวิตทางถนน

ในช่วงเทศกาลปีใหม่ด้วยวิธีแบบ Decision Tree ด้วยคุณสมบัติข้อมูลของ เพศ การดื่มแอลกอฮอล์ การคาดเข็มขัดนิรภัยหรือหมอกนิรภัย การนาส่ง ช่วงเวลา วันที่เกิดเหตุ ประเภทของถนนที่เกิดเหตุ ประเภทของยานพาหนะที่เกิด และรถคู่กรณี ในการสร้างแบบจำลองการพยากรณ์ให้ค่าประสิทธิภาพที่ดีกว่าแบบ Naïve Bayes

สำหรับการแบ่งลักษณะประชากรที่มีโอกาสเสียชีวิตทางถนนในช่วงเทศกาลปีใหม่ มีค่าความแม่นยำอยู่ที่ 56% โดยทำการแบ่งประชากรออกเป็น 2 กลุ่ม คือเพศชาย และเพศหญิง ซึ่งลักษณะข้อมูลที่มีความแตกต่างกันอย่างชัดเจน คือ ช่วงอายุ การดื่มแอลกอฮอล์ และรถคู่กรณี

ในงานวิจัยที่เกี่ยวข้องของเรื่องนี้ใช้ข้อมูลชุดเดียวกันกับงานของผู้วิจัย แต่มีความแตกต่างในเรื่องของเทคนิคในการจัดการความไม่สมดุลของชุดข้อมูลที่คอลัมน์ของข้อมูลผลการรักษา (Class Label) ในส่วนของการทำการแบ่งกลุ่มข้อมูลจะใช้วิธี K–Mean ในการจัดกลุ่มของข้อมูลตามเพศที่มีโอกาสเสียชีวิตทางถนนในช่วงเทศกาลปีใหม่ ผู้วิจัยไม่ได้นำเทคนิคนี้มาใช้

(6) บทความวิจัยเรื่อง Spatio-Temporal Analysis and Machine Learning for Traffic Accidents Prediction โดย Diena Al-Dogom, Nour Aburaed, Mina Al-Saad, Saeed Almansoori (Al-Dogom, Aburaed, Al-Saad, และ Almansoori, 2019)

งานวิจัยนี้ได้ทำการศึกษาการทำนายความรุนแรงของอุบัติเหตุ โดยรวมเทคนิคการเรียนรู้ของเครื่อง Machine Learning กับระบบสารสนเทศภูมิศาสตร์ (Geographic Information System : GIS) เพื่อทำการวิเคราะห์ข้อมูล Hotspot อุบัติเหตุตามระดับความรุนแรงจากการเกิดอุบัติเหตุจราจรและทำนายความรุนแรงของอุบัติเหตุทางจราจร เพื่อประโยชน์ในการระบุและค้นหาถนนที่มีความเสี่ยงได้ทันที เพื่อลดอัตราการเสียชีวิตจากอุบัติเหตุให้มากที่สุด

ในการศึกษาได้ใช้ชุดข้อมูลอุบัติเหตุทางจราจรที่ครอบคลุมสหราชอาณาจักร (The United Kingdom's : UK) ในช่วงเวลาระหว่าง 2555 - 2557 ของรัฐบาลที่เผยแพร่บน Kaggle ประกอบด้วย Features ของ ลองจิจูด, ละติจูด, ความรุนแรงของอุบัติเหตุ (แบ่งเป็น 3 ระดับ), เดือน, ชั่วโมง, ประเภทถนน, จำกัด ความเร็ว, สภาพแสง, สภาพอากาศ, สภาพผิวถนน, พื้นที่ (ในเมืองหรือในชนบท), จำนวนผู้เสียชีวิตและปี ในส่วนของขั้นตอนในการเตรียมข้อมูลมีการจัดการกับข้อมูล Missing Value โดยการลบออกจากชุดข้อมูลและมีปัญหา Imbalance Data ใน Class ของความรุนแรงของอุบัติเหตุ พบว่า 84% ของข้อมูลเป็นความรุนแรงประเภทที่ 3 และประเภทที่ 2 และ 1 คิดเป็น 14% และ 1% แล้วถูกจัดการด้วยวิธี Upsample และ Downsample แต่วิธีที่ได้ผลดีที่สุดคือ Downsample

ในการสร้างแบบจำลองการทำนายใช้ 2 เทคนิค ของการเรียนรู้ของเครื่อง (Machine Learning) เพื่อสร้างโมเดลการทำนายความรุนแรงของอุบัติเหตุ เทคนิคแรกใช้ Extreme Gradient Boosting (XGBoost) Library และเทคนิคที่สองใช้ Extra Trees Classifier (ETC) เป็นการทำงานแบบ Tree-Based Ensemble Supervised Learning ช่วยในการสุ่มลดความเสี่ยงของการเกิด Overfitting และทำให้แบบจำลองดีกว่า ถ้าเทียบกับ Random Forest และมีการทำ Feature Importance โดย Extra Trees Classifier พบว่า Feature ที่ส่งผลกระทบต่อความรุนแรงของอุบัติเหตุในปี 2555, 2556 และ 2557 คือ ชั่วโมง และ XGBoost มี Feature ที่ส่งผลกระทบต่อความรุนแรงของอุบัติเหตุในปี 2555, 2556 และ 2557 คือ ประเภทถนน และมีการวิเคราะห์ Hotspot Analysis ถือเป็นเครื่องมือที่มีประสิทธิภาพสำหรับถนนและการขนส่งเพื่อปรับปรุงการลดอุบัติเหตุและกลยุทธ์การป้องกันอุบัติเหตุ โดยใช้เทคนิคการวิเคราะห์ของ GIS เพื่อใช้ในการกำหนดฮอตสปอตของอุบัติเหตุจราจรในชุดข้อมูล มีขั้นตอนหลัก 3 ขั้นตอน สำหรับการวิเคราะห์ประเมิน Hotspots ในขั้นตอนแรก คือการเรียกใช้ฟังก์ชั่นรวบรวมเหตุการณ์ใน GIS ส่วนที่สอง คือ สถิติ Getis-Ord GI สำหรับกลุ่มการทำแผนที่ ส่วนสุดท้ายเป็นการประมาณความหนาแน่นโดยใช้ (Kernel Density Function : KDF) ในขั้นตอนเหล่านี้ดำเนินการโดยใช้เครื่องมือซอฟต์แวร์ ArcGIS

ผลการวิจัยของการวิเคราะห์ข้อมูลเชิงพื้นที่ของ Hotspot ของอุบัติเหตุตามระดับความรุนแรงและทำนายความรุนแรงของอุบัติเหตุทางจราจรในระหว่างปี 2555, 2556 และ 2557 โดยใช้ XGBoost และ Extra Trees Classifier แล้วประเมินความแม่นยำโดยใช้ค่า Accuracy ซึ่ง XGBoost แสดงความแม่นยำที่สูงขึ้นทุกปี เมื่อเทียบกับ Extra Trees Classifier ส่วนการวิเคราะห์ข้อมูลเชิงพื้นที่ของ Hotspot ของอุบัติเหตุตามระดับความรุนแรง พบว่าความรุนแรงของอุบัติเหตุจราจรมีแนวโน้มกระจายสูงในช่วงปี 2555 กับ 2556 และลดลงในปี 2557

(7) บทความวิจัยเรื่อง Machine Learning to Predict the Freeway Traffic Accidents-Based Driving Simulation โดย Rabia Emhamed Al Mamlook, Abdulla Ali, Raed Abdullah Hasan และ Haider A. Mohamed Kazim (Mamlook, Ali, Hasan, และ Kazim, 2019)

งานวิจัยนี้ได้ทำการศึกษาเพิ่มประสิทธิภาพในการสร้างแบบจำลองหลายๆ เทคนิคของการเรียนรู้ของเครื่อง ในการทำนายการเกิดอุบัติเหตุจราจรบนถนนทางด่วนในรัฐมิชิแกน สหรัฐอเมริกา จากนั้นเลือกแบบจำลองการทำนายที่แม่นยำที่สุดที่จะช่วยลดอุบัติเหตุบนถนนทาง

ด่วน ซึ่งเป็นถนนที่มีขีดจำกัดความเร็วสูงและปริมาณการใช้ถนนสูง ทำให้อุบัติเหตุจราจรที่เกิดขึ้นรุนแรงส่งผลต่อการเสียชีวิต

ข้อมูลในการศึกษานี้ถูกรวบรวมโดยใช้เครื่องมือจำลองการขับรถในทางด่วนมิชิแกนที่มหาวิทยาลัย Lawrence University เพื่อความเข้าใจที่มากขึ้นของความสัมพันธ์ระหว่างสภาพอากาศ, ยานพาหนะ, ถนน, สิ่งแวดล้อม และการเกิดอุบัติเหตุ วิธีการทั้งหมดเป็นวิธีที่สมจริง เป็นข้อมูลที่มุ่งเน้นไปที่พฤติกรรมการขับขี่ การจำลองมีผู้เข้าร่วม 100 คน (ชาย 81 คนและหญิง 19 คน) ขับรถผ่านถนนจำลองที่สถานการณ์แตกต่างกัน โดยสถานการณ์ส่วนใหญ่ ถูกกำหนดโดยสภาพอากาศ และขีดจำกัดความเร็วของถนน และหน้าจอคอมพิวเตอร์จะแสดงข้อมูลสำหรับคนขับ เช่น ความเร็ว ไมล์ต่อชั่วโมง (MPH) และรอบเครื่องต่อนาที (RPM) และแสดงสถานการณ์การขับขี่ ผ่านถนนจำลอง 10 เส้น สถานการณ์ตามสภาพอากาศ พร้อมจำกัดความเร็ว และมีการเลือกใช้ตัวแปรที่มีผลต่อการเกิดอุบัติเหตุ โดยมุ่งเน้นไปที่พฤติกรรมการขับขี่ทั้งหมด 6 Feature ที่ได้รับเลือกเป็นตัวทำนายที่เป็นไปได้ของการเกิดอุบัติเหตุ ประกอบด้วย ความเร็วจำกัดของถนนทางด่วน (Speedlimit), การเกิดอุบัติเหตุ (Crash), ค่าสัมประสิทธิ์แรงเสียดทาน (Coefficient of Fiction), ความเร็วที่คนขับใช้ (Drivers Speed), ใช้ความเร็วเกิน (Limit), เบรคหรือไม่เบรค (Break) และสภาพแวดล้อมของอากาศ ใน Dataset ทั้งหมด 1000 พบว่า ในตัวแปรที่เป็นคำตอบ 15% เป็นตัวแทนของอินสแตนซ์ที่เกิดอุบัติเหตุและส่วนที่เหลืออีก 85% เป็นตัวแทนของไม่เกิดอุบัติเหตุ แก้ปัญหาโดยใช้เทคนิค SMOTE เป็นแนวคิดของ Unbalanced Data การสุ่มตัวอย่างใหม่ โดยใช้กลยุทธ์การสุ่ม

ในการสร้างแบบจำลองการทำนาย มีการใช้ 6 อัลกอริทึมในการทำนายการเกิดอุบัติเหตุ ประกอบด้วย Bayesian Logistic Regression (BLR), Decision Tree (DT), Random Forest (RF), (KNN), Naïve Bayes Classifiers (NBC) และ Support Vector Machine (SVM) โดยแบ่งข้อมูลออกเป็น Training Dataset 70% และ Testing Dataset 30% ของชุดข้อมูลทั้งหมด และการประเมินประสิทธิภาพด้วยการทำ Cross Validation ทั้งหมด 10 Fold และในการวัดประสิทธิภาพการทำนายของแบบจำลอง ทดสอบโดยการเปรียบเทียบด้วย Confusion Matrix, Precision, Recall, F1 Score และ Area Under the ROC Curve โดยทำการทดลอง 2 แบบ คือ แบบไม่ทำเทคนิค SMOTE และไม่ทำ SMOTE

ผลการวิจัยพบว่า Random Forest ประสิทธิภาพทำงานได้ดีที่สุดใน Area Under ROC Curve เท่ากับ 0.826 และค่าความแม่นยำที่ Precision เท่ากับ 0.82 ที่ไม่มีการใช้เทคนิค SMOTE ในการจัดการข้อมู การทดลองนี้ Random Forest เป็นแบบจำลองการทำนายอุบัติเหตุ

ในทางด่วนของมิชิแกนของสหรัฐอเมริกาที่ดีที่สุด และ 2 Scenarios ที่เปรียบเทียบกันระหว่างไม่มีการใช้เทคนิค Resampling Strategy กับใช้เทคนิค Resampling Strategy ของ SMOTE ไม่มีความแตกต่างใน Area Under ROC Curve แต่มีความแตกต่างในค่า Precision, Recall, F1-score ลดลงถ้าทำเทคนิค SMOTE และเพิ่มขึ้นของค่า Recall หลังจากทำเทคนิค SMOTE ในทุกโมเดล

(8) บทความวิจัยเรื่อง Comparison of Machine Learning Algorithms for Predicting Traffic Accident Severity โดย Rabia Emhamed Al, Keneth Morgan Kwayu, Maha Reda Alkasisbeh และ Abdulbaset Ali Frefer (AlMamlook, Kwayu, Alkasisbeh, และ Frefer, 2019)

งานวิจัยนี้ได้ทำการศึกษาการสร้างแบบจำลองการทำนายความรุนแรงของการบาดเจ็บจากอุบัติเหตุบนถนนทางด่วนในรัฐมิชิแกน

ในการศึกษานี้ใช้ข้อมูลจากสำนักงานการวางแผนความปลอดภัยทางหลวง The Office of Highway Safety Planning (OHSP). ชุดข้อมูลมีข้อมูลเกี่ยวกับอุบัติเหตุบนถนนทางด่วน มิชิแกน ในช่วงเวลาตั้งแต่ปี 2010 ถึงปี 2016 ชุดข้อมูลมีจำนวน 271,563 แถว และในข้อมูลมีจำนวนของผู้เสียชีวิตในระหว่างการขับขี่บนถนนทางด่วนในปี 2016 สูงถึง 539 ราย สูงที่สุดในรอบทศวรรษของ Michigan และแนวโน้มเพิ่มขึ้นอย่างมากในช่วงหลายปีที่ผ่านมา โดยใช้ข้อมูล 10 Features ที่ได้จากการทำ Feature Selection ประกอบด้วย อายุของรถยนต์, อายุ, จำนวน ความหนาแน่นของจราจร, เพศ, การดื่มแอลกอฮอล์, สภาพแสงสว่างหรือมืด, สภาพอากาศ, ความประมาทที่เป็นการกระทำที่เป็นอันตรายต่อผู้ขับขี่, การใช้เข็มขัดนิรภัย และความรุนแรงของผู้บาดเจ็บจากอุบัติเหตุ คือ (1) แสดงถึงข้อผิดพลาดร้ายแรงเสียชีวิต และ (0) แสดงถึงข้อผิดพลาดไม่ร้ายแรงไม่เสียชีวิต ในการศึกษาครั้งนี้มีการใช้ Class-Imbalanced Data ใช้เทคนิค SMOTE Resampling Strategy เพื่อจัดการกับความไม่สมดุลของข้อมูลใน Class ของความรุนแรงของผู้บาดเจ็บจากอุบัติเหตุ เนื่องจากมีอัตราการเสียชีวิตเพียงเล็กน้อยเท่านั้น เมื่อเทียบกับการบาดเจ็บอื่นๆ SMOTE อัลกอริทึมสร้างอินสแตนซ์บวกสังเคราะห์ เพื่อเพิ่มสัดส่วนของประชากรกลุ่มน้อย

ในการสร้างแบบจำลองการทำนายความรุนแรงของการบาดเจ็บจากอุบัติเหตุเสียชีวิตหรือไม่เสียชีวิต มีการใช้ 4 อัลกอริทึมประกอบด้วย Logistic Regression (LR), Random Forest (RF), Naïve Bayesian Classifier (NBC) และAdaBoost โดยแบ่งข้อมูล 70% Training และ 30% Testing และมีการทำทดสอบประสิทธิภาพ 10 Fold Cross Validation และ

เปรียบเทียบวัดความแม่นยำของแบบจำลองการทำนายโดย Confusion Matrix, Accuracy, Precision, Recall, F-score และ The Area Under ROC Curve (AUC) โดยที่มีค่า Ratio ระหว่าง 0 และ 1 ถ้า 1 เท่ากับ Perfect Classifier

ผลการวิจัยพบว่า Random Forest ประสิทธิภาพทำงานได้ดีที่สุด สามารถทำนายความรุนแรงของผู้บาดเจ็บจากอุบัติเหตุได้อย่างแม่นยำที่แสดงค่า AUC ที่ดี Better Performance เท่ากับ 75.5%, Accuracy เท่ากับ 75.50%, Precision เท่ากับ 88.1%, Recall เท่ากับ 59.1% และ F1-Score เท่ากับ 70.7%

(9) บทความวิจัยเรื่อง Traffic Accidents Classification and Injury Severity Prediction โดย Laura García Cuenca, Nourdine Aliane, Enrique Puertas, Javier Fernandez Andres (Cuenca, Puertas, Aliane, และ Andres, 2018)

งานวิจัยนี้ได้ทำการศึกษาการสร้างแบบจำลองการทำนายความรุนแรงของอุบัติเหตุบนถนนในประเทศสเปน

ในการศึกษานี้ใช้ข้อมูลจาก The Spanish Traffic Agency (Hereafter DGT) โดยข้อมูลประกอบด้วย 3 ไฟล์ได้แก่

1. Accidents Dataset ไฟล์ที่มีข้อมูลรายปีของอุบัติเหตุจราจรที่มีโครงสร้าง 38 คุณลักษณะ เช่น ความส่องสว่าง, ประเภทถนน, สภาพอากาศ, ประเภทอุบัติเหตุ, วันในสัปดาห์, ภูมิภาค, จำนวนผู้ประสบอุบัติเหตุ, คนบาดเจ็บทั้งหมด ฯลฯ

2. Vehicles Dataset ไฟล์ที่มีข้อมูลยานพาหนะในอุบัติเหตุจราจร ประกอบด้วย ชนิดของยานพาหนะ, สถานะของยานพาหนะ, จำนวนผู้โดยสาร, ปีที่จดทะเบียน, รหัสยานพาหนะ , รหัสของอุบัติเหตุ ฯลฯ

3. People Dataset ข้อมูลของผู้ประสบอุบัติเหตุ

การเตรียมข้อมูลหลังจากการรวมข้อมูลแล้วมี 58 Features และ 1,018,204 Records และมีการเพิ่ม 2 New Aggregated Features เข้าไปใน Dataset ประกอบด้วย ข้อมูลอายุของยานพาหนะตั้งแต่วันจดทะเบียนไปจนถึงวันเกิดอุบัติเหตุ (Age of the Vehicle) และความรุนแรงของอุบัติเหตุ (Severity Accident) โดยที่ค่า 0 เท่ากับไม่รุนแรง และค่า 1 รุนแรงเสียชีวิต ที่ใช้ในการทำนายอุบัติเหตุร้ายแรงหรือไม่ และข้อมูลมีปัญหาความไม่สมดุลของข้อมูล (Unbalanced) ใน Class ของ Severity Accident สำหรับค่า 0 ของแอตทริบิวต์อุบัติเหตุไม่ร้ายแรง เมื่อเปรียบเทียบกับ 18,082 บันทึกค่า 1 ของอุบัติเหตุที่ร้ายแรงเสียชีวิต แก้ไขปัญหาข้อมูลที่ไม่สมดุลด้วยเทคนิค Subsampling เทคนิคเกี่ยวข้องกับ Majority Class และมีการเลือก

Feature Engineering โดยใช้กฎเกณฑ์ของ Correlation ต้องมากกว่า 50% เพราะมีความสัมพันธ์สูงกับ Severity Accident ใช้ในการทำนายความร้ายแรงของอุบัติเหตุ การใช้กฎเหล่านี้กับชุดข้อมูลจะได้ New Dataset ที่มีแอตทริบิวต์ 34 Feature ที่จะใช้สำหรับสร้างแบบจำลองการทำนาย ประกอบด้วย ปีที่ได้รับใบอนุญาตขับขี่ (Year of Circulation Permit), ตำแหน่ง (Position), การใช้เข็มขัดนิรภัย (Use of the Safety Belt), การใช้หมวกนิรภัย (Helmet), Shunting, การละเมิดใช้ความเร็ว(Infringement Speed), การทำผิดกฎหมาย (Infringement Opening), จำนวนการทำผิดกฎหมาย (Infringement Summary), แสงสว่าง (Infringement Lighting), อายุ, เพศ, ปี, ปีที่ยานพาหนะจดทะเบียน (Year of Vehicle Registration), เดือนที่ยานพาหนะจดทะเบียน (Month of Vehicle Registration), ชนิดของยานพาหนะ (Vehicle Type), Anomaly None, Pneumatic Anomaly, ความผิดปกติเกี่ยวกับยาง (Tire Rebound Anomaly), ทิศทางความผิดปกติ (Direction Anomaly), เบรกผิดปกติ (Brakes Anomaly), จำนวนผู้โดยสารของยานพาหนะ (Number of Occupants of the Vehicle), โซน, กลุ่มพื้นที่ (Grouped Area), ทัศนวิสัย (Restricted Visibility), เวลา, Day of the Week, ชุมชนอิสระ (Autonomous Community), เทศบาล (Municipality), ถนน, เครือข่ายถนน (Road Network), Type Via, ประเภทจุดตัดถนน (Type Intersection) และอายุของยานพาหนะ (Vehicle's age)

ในการสร้างแบบจำลองการทำนายความรุนแรงของอุบัติเหตุ เสียชีวิตหรือไม่เสียชีวิต มีการเปรียบเทียบเทคนิค Machine Learning Classification 3 เทคนิค ประกอบด้วย Gradient Boosting Trees, Deep Learning และ Naïve Bayes และประเมินประสิทธิภาพของแบบจำลองด้วย Confusion Matrix, Precision, Accuracy และ F-Measure.

ผลการวิจัยพบว่าหลังจากเปรียบเทียบผลลัพธ์ที่ต่างกัน ผลลัพธ์ที่ดีที่สุดที่ได้รับสำหรับการทำนายความรุนแรงของอุบัติเหตุ เป็นของอัลกอริทึม Deep Learning Algorithm (10 Epochs-Tanh) ที่มีค่า Accuracy เท่ากับ 87.75% Precision เท่ากับ 85.19% และ F-Measure เท่ากับ 88.18%

(10) บทความวิจัยเรื่อง Road Accident Analysis and Prediction of Accident Severity by Using Machine Learning in Bangladesh โดย Md. Farhan Labib, Ahmed Sady Rifat, Md. Mosabbir Hossain, Amit Kumar Das, Faria Nawrine (Labib, Rifat, Hossain, Das, และ Nawrine, 2019)

งานวิจัยนี้ได้ทำการศึกษาการสร้างแบบจำลองการทำนายความรุนแรงของอุบัติเหตุทางจราจรในบังคลาเทศ โดยแบ่งความรุนแรงของการเกิดอุบัติเหตุออกเป็น 4 ประเภทคือ การ

บาดเจ็บร้ายแรงเสียชีวิต, การบาดเจ็บสาหัส, การบาดเจ็บไม่รุนแรง และไม่บาดเจ็บ เป็นการชนกันของยานพาหนะ ในงานวิจัยนี้ใช้ชุดข้อมูลจาก ARI ของ BUET ซึ่งประกอบด้วยข้อมูลสถิติของอุบัติเหตุทางจราจรรวม 43,089 แถว 34 ปัจจัย ตั้งแต่ปี 2544-2558 ในประเทศบังคลาเทศ และ Missing Values 8.7% ในชุดข้อมูลทั้งหมด แทนค่าด้วย Mean Value

มีการใช้ Feature Selection เพื่อแยกความแตกต่างของคุณสมบัติ Feature ที่มีความสำคัญน้อยกว่า เพื่อให้ได้คุณสมบัติที่สำคัญที่สุด ดำเนินการทดสอบโดยใช้ 3 เทคนิค ประกอบด้วย 1) Univariate Feature Selection ทดสอบแบบ Chi-Squared Statistical 2) Recursive Feature Elimination ทำงานโดยการลบคุณสมบัติแบบวนซ้ำและใช้ความแม่นยำของแบบจำลอง เพื่อเลือกคุณสมบัติที่มีส่วนช่วยในการทำนายตัวแปรที่ต้องการมากที่สุด ในขั้นต้นจะฝึก Feature ทั้งหมด โดย Logistic Regression พยายามหาความสำคัญของแต่ละ Feature หลังจากนั้นจะตัด Feature ที่มีความสำคัญน้อยกว่าและกระบวนการนี้จะทำซ้ำจนกว่าจะถึงจำนวน Feature ที่ต้องการ 3) Feature Importance สุดท้ายจะได้ข้อมูลที่จะใช้ทั้งหมด 11 Feature ที่การทดลองทั้ง 3 นี้ เลือก Feature เหมือนกัน ประกอบด้วย Time, Traffic Control, Weather, Light, Road Geometry, Vehicle Type, Movement, Divider, Road class, Location Type และ Vehicle Defect ที่ใช้ในการสร้างแบบจำลองการทำนายความรุนแรงของอุบัติเหตุทางจราจรในบังคลาเทศ

ในการสร้างแบบจำลองการทำนายความรุนแรงของอุบัติเหตุทางจราจร มีการเปรียบเทียบเทคนิค Machine Learning Classification 4 เทคนิค ประกอบด้วย Decision Tree, AdaBoost, KNN และ Naïve Byes แล้วประเมินประสิทธิภาพของแบบจำลองด้วย Confusion Matrix, Precision, Accuracy และ F-1 Score

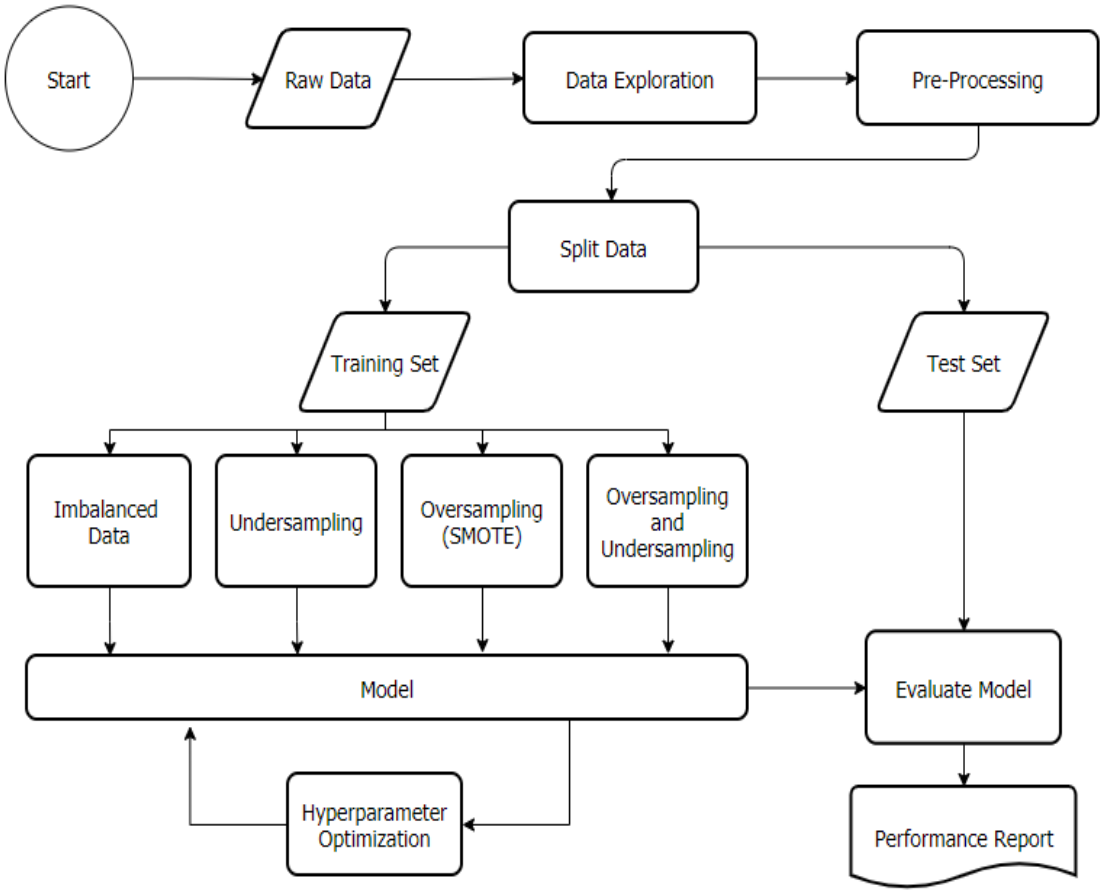
ผลการวิจัยพบว่าหลังจากเปรียบเทียบผลลัพธ์ที่ต่างกัน ผลลัพธ์ที่ดีที่สุดที่ได้รับสำหรับการทำนายความรุนแรงของอุบัติเหตุทางจราจร เป็นของอัลกอริทึม AdaBoost ให้ผลลัพธ์ประสิทธิภาพที่ดีที่สุดที่ Accuracy เท่ากับ 80% Precision เท่ากับ 68% และ F1-score เท่ากับ 73%

บทที่ 3
การดำเนินการวิจัย

ในการวิจัยครั้งนี้ ผู้วิจัยได้ดำเนินการตามขั้นตอนดังนี้

1. กระบวนการทำงานของแบบจำลอง
2. การเก็บรวบรวมข้อมูล
3. การสำรวจข้อมูล (Exploratory Data Analysis: EDA)
4. การเตรียมข้อมูล
5. อัลกอริทึมของแบบจำลองการทำนาย

3.1 กระบวนการทำงานของแบบจำลอง



ภาพประกอบ 7 Flowchart กระบวนการสร้างแบบจำลอง

จากภาพประกอบ 7 Flowchart ได้อธิบายถึงกระบวนการสร้างแบบจำลองการทำนาย โดยเริ่มจากขั้นตอนการนำเข้าข้อมูล การสำรวจข้อมูล (Exploratory Data Analysis : EDA) วิเคราะห์ความสัมพันธ์ระหว่างตัวแปรต่างๆ กับผลการรักษาผู้บาดเจ็บจากอุบัติเหตุ การเตรียมข้อมูลนำปัจจัยของชื่อเทศกาลปีใหม่ ชื่อจังหวัด และชื่อโรงพยาบาลถูกนำออกจากข้อมูลฝึกสอนแบบจำลอง ข้อมูลที่จัดเป็นกลุ่ม (Categorical Features) ถูก Encode เป็นข้อมูลตัวเลขก่อนนำเข้าฝึกสอนแบบจำลองการทำนาย

และในการสร้างแบบจำลองการทำนาย แบ่งข้อมูล (Split Data) ออกเป็นข้อมูลฝึกสอน (Training Set) และข้อมูลทดสอบ (Test Set) ในอัตราส่วน 60:40 และ Training set แบ่งข้อมูลเป็น Train : 80 และ Validation : 20 โดยใช้เทคนิค Stratify ในการ Split ข้อมูล สำหรับระดับความรุนแรงของอุบัติเหตุมี 2 ระดับคือ 0 : ไม่เสียชีวิต และ 1 : เสียชีวิต จากนั้นนำข้อมูลฝึกสอนให้แบบจำลองเรียนรู้เพื่อสร้างแบบจำลองการทำนายความรุนแรงของผู้บาดเจ็บจากอุบัติเหตุทางถนนในแบบการจำแนกประเภทแบบไบนารี อัลกอริทึมการเรียนรู้ของเครื่องที่ใช้คือ Random Forests และใช้เทคนิคการสุ่มตัวอย่างใหม่ (Resampling Techniques) 3 แบบ เพื่อปรับความสมดุลของข้อมูลอินสแตนซ์ในข้อมูลฝึกสอนของแบบจำลองการทำนาย (Rebalance Training Instances) ได้แก่ เทคนิคการลดจำนวนตัวอย่าง (Undersampling), เทคนิคการเพิ่มจำนวนตัวอย่างด้วย SMOTE (Oversampling with SMOTE) และการผสมระหว่างของทั้ง 2 เทคนิค ของเทคนิคการลดจำนวนตัวอย่าง (Undersampling) กับเทคนิคการเพิ่มจำนวนตัวอย่างด้วย SMOTE (Oversampling with SMOTE) จากนั้นทำการปรับจูนพารามิเตอร์ของแบบจำลองที่เหมาะสมด้วยวิธี Grid search แบบจำลองทั้งหมดจะในระหว่างการฝึกสอนจะถูกตรวจสอบด้วย Cross Validated ทั้งหมด 5-Fold โดยใช้การแบ่ง Fold แบบเทคนิค Stratified Shuffle Splits และ ใช้เทคนิค Pipeline ในการจัดลำดับการทำงานของปรับความสมดุลของข้อมูลฝึกสอนและการฝึกสอนแบบจำลองการทำนาย

สุดท้ายการทดสอบประสิทธิภาพของแต่ละแบบจำลองการทำนาย (Evaluate Model) โดยใช้รายงาน Classification Report ประกอบด้วย Precision, Recall, Accuracy, Confusion Matrix, Receiver Operating Characteristic Curve (ROC Curve) และ Precision-Recall Curve ถูกสร้างขึ้นเพื่อประเมินผลแบบจำลองการทำนายด้วยข้อมูลทดสอบ (Test Set) แต่ในงานวิจัยเรื่องนี้ไม่ได้เพียงแต่คำนึงถึงความถูกต้องของการทำนายทั้งหมด (Accuracy) เท่านั้น แต่ยังมุ่งเน้นไปที่ค่า Recall เพื่อวัดประสิทธิภาพที่แท้จริงของการระบุจำนวนผู้บาดเจ็บจากอุบัติเหตุที่รุนแรงถึงการเสียชีวิตจากอุบัติเหตุทางถนนในช่วงเทศกาลปีใหม่ให้ถูกต้อง และปัจจัยที่มีผลต่อ

การตัดสินในการทำนาย (Feature Importance) ของแต่ละแบบจำลอง เพื่อให้ได้ปัจจัย (Feature) ที่มีความสำคัญในการตัดสินใจส่งผลต่อการทำนายความรุนแรงของผู้บาดเจ็บจากอุบัติเหตุทางถนนในช่วงเทศกาลปีใหม่ 2551 ถึง 2558

3.2 การเก็บรวบรวมข้อมูล

ในการวิจัยครั้งนี้ใช้ข้อมูลผู้บาดเจ็บจากอุบัติเหตุจราจรในช่วงเทศกาลปีใหม่ 2551 ถึงปี 2558 ที่ถูกจัดเก็บมาเรียบร้อยแล้วจากองค์กรสถาบันการแพทย์ฉุกเฉินแห่งชาติ เผยแพร่โดยสำนักงานพัฒนารัฐบาลดิจิทัล องค์การมหาชน (Digital Government Development Agency (Public Organization) : DGA) (สถาบันการแพทย์ฉุกเฉินแห่งชาติ, 2558) ซึ่งประกอบด้วย Feature ทั้งหมด 19 คอลัมน์ ได้แก่ ชื่อเทศกาล, รหัสจังหวัด, ชื่อจังหวัด, รหัสโรงพยาบาล, ชื่อโรงพยาบาลที่รับผู้บาดเจ็บ, วันที่เกิดเหตุ, เวลาเกิดเหตุ, เพศ, อายุ, ถนนที่เกิดเหตุ, สถานะ, รถผู้บาดเจ็บ, รถคู่กรณี, มาตรการ, การดื่มสุรา, การนำส่ง, Refer-Admit, จำนวนวันรักษาและผลการรักษา โดยข้อมูลมีการจัดเก็บตั้งแต่วันที่ 27 ธันวาคม 2551 (เวลา 01:01 น.) ถึง วันที่ 5 มกราคม 2558 (เวลา 24:00 น.) มีจำนวนทั้งหมด 214,950 แถว ดังตาราง 1

ตาราง 1 รายละเอียดของชุดข้อมูลผู้บาดเจ็บจากอุบัติเหตุจราจรในช่วงเทศกาลปีใหม่ 2551 ถึงปี 2558

ลำดับ	**ข้อมูลตัวแปร** (Variable)	**คำอธิบายข้อมูล** (Description)
1	ชื่อเทศกาล	ชื่อปีของเทศการปีใหม่ของปี 2551 ถึง ปี 2558
2	รหัสจังหวัด	รหัสจังหวัดของประเทศไทย
3	ชื่อจังหวัด	ชื่อจังหวัดของประเทศไทย
4	รหัสโรงพยาบาล	รหัสโรงพยาบาลของประเทศไทย
5	ชื่อโรงพยาบาลที่รับผู้บาดเจ็บ	ชื่อของโรงพยาบาลที่รับรักษาผู้บาดเจ็บจากอุบัติเหตุ
6	วันที่เกิดเหตุ	วันที่เกิดอุบัติเหตุของช่วงเทศกาลปีใหม่
7	เวลาเกิดเหตุ	หมายถึง เวลาที่เกิดอุบัติเหตุ เช่น เวลาเกิดเหตุ 10.35 น. ให้เลือกข้อมูลช่วงเวลา 10.01-11:00 น.
8	เพศ	เพศของผู้ประสบอุบัติเหตุ
9	อายุ	อายุของผู้ประสบอุบัติเหตุเป็นจำนวนปี
10	ถนนที่เกิดเหตุ	ถนนที่เกิดอุบัติเหตุ ประกอบด้วย ถนนชนบท ถนนทางหลวง หมายถึง ถนนที่มีหมายเลขกำกับเส้นทางหลวงเป็นตัวเลข ถนนในเมือง หมายถึง ถนนภายในเขตเทศบาล (ยกเว้นทางหลวงซึ่งผ่านในเขตเทศบาลให้นับเป็นทางหลวง)
11	สถานะของผู้บาดเจ็บ	สถานะของผู้บาดเจ็บจากประสบอุบัติเหตุ ประกอบด้วย คนเดินเท้า หมายถึง ผู้บาดเจ็บหรือเสียชีวิตจากอุบัติเหตุจราจร ที่ไม่ใช่ผู้ขับขี่หรือผู้โดยสาร

ตาราง 1 (ต่อ)

ลำดับ	**ข้อมูลตัวแปร** (Variable)	**คำอธิบายข้อมูล** (Description)
		ผู้ขับขี่ หมายถึง ผู้บาดเจ็บหรือเสียชีวิตจากอุบัติเหตุจราจร ที่เป็นผู้ขับขี่ยานพาหนะ ผู้โดยสาร หมายถึง ผู้บาดเจ็บหรือเสียชีวิตจากอุบัติเหตุจราจร อยู่ในยานพาหนะที่ไม่ใช่ผู้ขับขี่ ไม่ทราบสถานะของผู้บาดเจ็บ
12	รถผู้บาดเจ็บ	ยานพาหนะขับขี่ของผู้บาดเจ็บที่เกิดอุบัติเหตุ ประกอบ ด้วย รถจักรยานยนต์ รถปิคอัพ รถจักรยาน ไม่มี / ล้มเอง รถเก๋ง / รถแท็กซี่ อื่นๆ รถสามล้อเครื่อง ไม่ทราบ รถตู้ รถโดยสารใหญ่ รถบรรทุก รถโดยสาร 4 ล้อ รถสามล้อถีบ
13	รถคู่กรณี	รถคู่กรณีของผู้ประสบอุบัติเหตุ ประกอบด้วย รถจักรยานยนต์ รถปิคอัพ รถเก๋ง / แท็กซี่ รถบรรทุก รถจักรยาน

ตาราง 1 (ต่อ)

ลำดับ	ข้อมูลตัวแปร (Variable)	คำอธิบายข้อมูล (Description)
		รถตู้ รถสามล้อเครื่อง รถโดยสารใหญ่ รถสามล้อถีบ รถโดยสาร 4 ล้อ ไม่มี / ล้มเอง ไม่ทราบ อื่นๆ
14	มาตรการ	หมายถึง ปัจจัยที่ลดความรุนแรงจากอุบัติเหตุจราจร คือการสวมหมวกนิรภัย และการคาดเข็มขัดนิรภัย โดยข้อมูลประกอบด้วย เข็มขัดนิรภัย ไม่ใส่หมวกนิรภัย ใส่หมวกนิรภัย ไม่ทราบ
15	การดื่มสุรา	หมายถึง การได้ประวัติการดื่มสุรามาก่อน หรือการตรวจพบการดื่มสุราจากการซักถาม สังเกตการได้กลิ่น การตรวจร่างกาย หรือการตรวจทางห้องปฏิบัติการ ข้อมูลประกอบด้วย ดื่มสุรา ไม่ดื่มสุรา ไม่ทราบ
16	การนำส่ง	หมายถึง วิธีการนำส่งผู้บาดเจ็บไปโรงพยาบาล โดยข้อมูลประกอบด้วย ผู้ประสบอุบัติเหตุ / ญาติ หมายถึง ผู้ประสบอุบัติเหตุมาโรงพยาบาลด้วยตนเอง พลเมืองดีนำส่ง หรือญาติของผู้บาดเจ็บ

ตาราง 1 (ต่อ)

ลำดับ	ข้อมูลตัวแปร (Variable)	คำอธิบายข้อมูล (Description)
		หรือการนำส่งจากหน่วย EMS หมายถึง การนำส่งโดยหน่วยปฏิบัติการที่ขึ้นทะเบียนในระบบการแพทย์ฉุกเฉิน มี 4 ระดับ ประกอบด้วยดังนี้ FR ย่อมาจาก First responder unit หมายถึง ชุดปฏิบัติการฉุกเฉินเบื้องต้น ซึ่งจะต้องผ่านการฝึกอบรมผู้ปฏิบัติการฉุกเฉินเบื้องต้น (First Responder curriculum) จากแหล่งฝึกอบรมที่คณะกรรมการการแพทย์ฉุกเฉินรับรอง ต้องมีขีดความสามารถในการปฏิบัติการฉุกเฉิน แจ้งเหตุ ขอความช่วยเหลือ ประเมินผู้ป่วยฉุกเฉิน จัดการทางเดินหายใจ ช่วยฟื้นคืนชีพขั้นพื้นฐาน ห้ามเลือด ช่วยคลอด เคลื่อนย้ายผู้ป่วยฉุกเฉิน คัดแยกเบื้องต้น พาหนะกู้ชีพระดับเบื้องต้น พร้อมอุปกรณ์กู้ชีพ และผู้ปฏิบัติการเป็นอาสาสมัครปฏิบัติการฉุกเฉิน BLS ชุดปฏิบัติการฉุกเฉินระดับพื้นฐาน (Basic Life Support Unit : BLS) หมายถึง ชุดปฏิบัติที่ประกอบด้วยพาหนะกู้ชีพระดับพื้นฐาน พร้อมอุปกรณ์กู้ชีพ และผู้ปฏิบัติการต่างๆ อย่างน้อยต้องเป็นเวชกรฉุกเฉินระดับต้น (EMT-B) เป็นหัวหน้าชุด ILS ชุดปฏิบัติการฉุกเฉินระดับกลาง (Intermediate Life Support Unit : ILS) หมายถึง ชุดปฏิบัติการที่ประกอบด้วย พาหนะกู้ชีพระดับกลาง พร้อมอุปกรณ์กู้ชีพ และผู้ปฏิบัติการต่างๆ อย่างน้อยต้องเป็นเวชกรฉุกเฉินระดับกลาง (EMT-I) เป็นหัวหน้าชุด

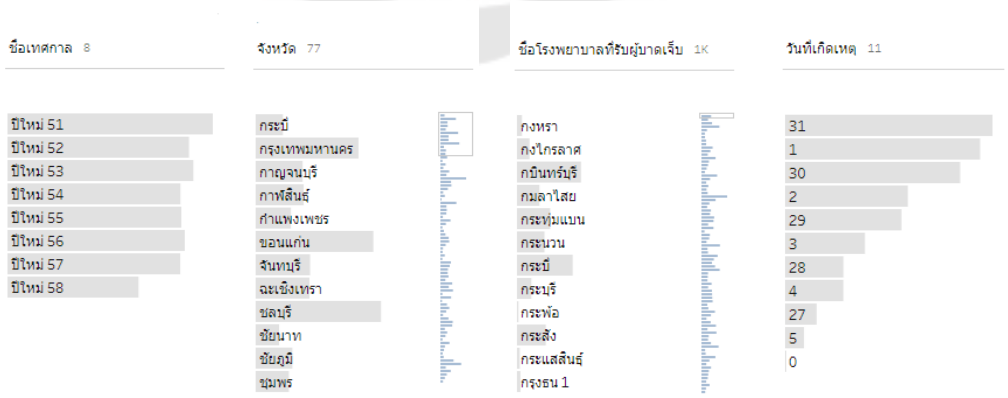
ตาราง 1 (ต่อ)

ลำดับ	ข้อมูลตัวแปร (Variable)	คำอธิบายข้อมูล (Description)
		ALS ชุดปฏิบัติการฉุกเฉินระดับสูง (Advanced Life Support Unit : ALS) หมายถึง ชุดปฏิบัติการที่ประกอบด้วยพาหนะกู้ชีพระดับสูง พร้อมอุปกรณ์กู้ชีพระดับสูงและผู้ปฏิบัติการต่างๆ อย่างน้อยต้องเป็นพยาบาลกู้ชีพ เป็นหัวหน้าชุด มูลนิธิ / อาสาสมัคร หรือเจ้าหน้าที่ตำรวจ เป็นผู้นำส่งผู้บาดเจ็บ ไม่นำส่งหรือเสียชีวิตในที่เกิดเหตุ หมายถึง เสียชีวิตในที่เกิดเหตุ
17	Refer-Admit	ผลการรักษาที่โรงพยาบาลรับผู้บาดเจ็บ โดยที่สถานภาพหรืออาการของผู้บาดเจ็บ เมื่อสิ้นสุดการรักษาเบื้องต้น จะต้องบันทึกผลการรักษาทุกรายประกอบด้วยดังนี้ ไม่ Admit/ Refer หมายถึง ผู้บาดเจ็บอาการทุเลาสามารถกลับบ้านได้ Refer หมายถึง ผู้บาดเจ็บ มีความจำเป็นต้องส่งต่อไปรับการรักษาที่โรงพยาบาลอื่น Admit หมายถึง การรับไว้รักษาต่อในโรงพยาบาล
18	จำนวนวันรักษา	จำนวนวันรักษาของผู้ประสบอุบัติเหตุ
19	ผลการรักษา	ผลการรักษาผู้ประสบอุบัติเหตุในสถานะภาพและอาการของผู้บาดเจ็บเมื่อสิ้นสุดการรักษาเบื้องต้น ประกอบด้วย ไม่เสียชีวิตหรือทุเลา/หาย หมายถึง ผู้บาดเจ็บ มีสถานะภาพและอาการเมื่อสิ้นสุดการรักษาอาการทุเลาสามารถกลับบ้านได้ เสียชีวิต ประกอบด้วย

ตาราง 1 (ต่อ)

ลำดับ	ข้อมูลตัวแปร (Variable)	คำอธิบายข้อมูล (Description)
		ตายเกิดเหตุ หมายถึง ผู้บาดเจ็บเสียชีวิตที่จุดเกิดเหตุ ตายในระหว่างนำส่ง หมายถึง ผู้บาดเจ็บเสียชีวิต ในขณะถูกนำส่งจากจุดเกิดเหตุไปโรงพยาบาล ตายที่ห้องฉุกเฉิน หมายถึง ผู้บาดเจ็บเสียชีวิตในห้องฉุกเฉินหลังถูกนำส่งไปโรงพยาบาล ตายในตึกภายใน 24 ชั่วโมง หลังเหตุ หมายถึง การเสียชีวิตในตึกผู้ป่วยในภายใน 24 ชั่วโมงแรกหลังได้รับประสบอุบัติเหตุ ตายในตึกหลัง 24 ชั่วโมง ถึง 30 วัน หมายถึง การเสียชีวิตในตึกผู้ป่วยใน หลังเวลา 24 ชั่วโมง แต่ไม่เกิน 30 วัน หลังได้รับประสบอุบัติเหตุ

สัดส่วนข้อมูลจากองค์กรสถาบันการแพทย์ฉุกเฉินแห่งชาติ ผู้บาดเจ็บจากอุบัติเหตุ ในช่วงเทศกาลปีใหม่ของปี 2551 ถึงปี 2558 ดังแสดงในภาพประกอบ 8 และ 9



ภาพประกอบ 8 สัดส่วนข้อมูลจากองค์กรสถาบันการแพทย์ฉุกเฉินแห่งชาติ ผู้บาดเจ็บจากอุบัติเหตุ ในช่วงเทศกาลปีใหม่ของปี 2551 ถึงปี 2558

ภาพประกอบ 9 สัดส่วนข้อมูลจากองค์กรสถาบันการแพทย์ฉุกเฉินแห่งชาติ ผู้บาดเจ็บจากอุบัติเหตุ ในช่วงเทศกาลปีใหม่ของปี 2551 ถึงปี 2558 (ต่อ)

สถานะ	รถผู้บาดเจ็บ	รถคู่กรณี	มาตรการ	การดื่มสุรา	การนำส่ง	Refer-Admit	ผลการรักษา	จำนวนวันรักษา
ผู้ขับขี่	รถเก๋ง/แท็กซี่	ไม่มี/ล้มเอง	เข็มขัด	ไม่ดื่ม	ผู้ประสบเหตุ/ญาติ	ไม่	ทุเลา/หาย	0
ผู้ขับขี่	จักรยานยนต์	รถตู้	ใส่หมวก	ไม่ดื่ม	มูลนิธิ/อาสาสมัคร	admit	ทุเลา/หาย	0
ผู้โดยสาร	จักรยานยนต์	ปิคอัพ	ไม่ใส่	ไม่ดื่ม	ผู้ประสบเหตุ/ญาติ	ไม่	ทุเลา/หาย	0
ผู้โดยสาร	จักรยานยนต์	ปิคอัพ	ใส่หมวก	ไม่ดื่ม	ผู้ประสบเหตุ/ญาติ	ไม่	ทุเลา/หาย	0
ผู้ขับขี่	จักรยานยนต์	ปิคอัพ	ใส่หมวก	ดื่ม	ผู้ประสบเหตุ/ญาติ	ไม่	ทุเลา/หาย	0
คนเดินเท้า	ไม่มี/ล้มเอง	ปิคอัพ	ไม่ทราบ	ไม่ดื่ม	ผู้ประสบเหตุ/ญาติ	ไม่	ทุเลา/หาย	0
ผู้ขับขี่	จักรยานยนต์	ไม่มี/ล้มเอง	ไม่ทราบ	ดื่ม	ผู้ประสบเหตุ/ญาติ	admit	ทุเลา/หาย	1
ผู้ขับขี่	จักรยานยนต์	รถเก๋ง/แท็กซี่	ไม่ทราบ	ดื่ม	FR	admit	ทุเลา/หาย	5
ผู้ขับขี่	รถจักรยาน	ไม่มี/ล้มเอง	ไม่ทราบ	ไม่ทราบ	ผู้ประสบเหตุ/ญาติ	admit	ทุเลา/หาย	2
ผู้ขับขี่	จักรยานยนต์	ไม่มี/ล้มเอง	ไม่ใส่	ไม่ทราบ	ผู้ประสบเหตุ/ญาติ	ไม่	ทุเลา/หาย	0
ผู้โดยสาร	จักรยานยนต์	ไม่มี/ล้มเอง	ใส่หมวก	ไม่ทราบ	ผู้ประสบเหตุ/ญาติ	ไม่	ทุเลา/หาย	0
ผู้โดยสาร	จักรยานยนต์	ไม่มี/ล้มเอง	ไม่ทราบ	ไม่ดื่ม	ผู้ประสบเหตุ/ญาติ	ไม่	ทุเลา/หาย	0
ผู้ขับขี่	จักรยานยนต์	ไม่มี/ล้มเอง	ไม่ทราบ	ดื่ม	ผู้ประสบเหตุ/ญาติ	ไม่	ทุเลา/หาย	0
ผู้ขับขี่	จักรยานยนต์	ไม่มี/ล้มเอง	ไม่ทราบ	ไม่ทราบ	มูลนิธิ/อาสาสมัคร	ไม่	ทุเลา/หาย	0
ผู้ขับขี่	รถจักรยาน	รถบรรทุก	ไม่ทราบ	ไม่ทราบ	มูลนิธิ/อาสาสมัคร	ไม่	ทุเลา/หาย	0

ภาพประกอบ 10 แสดงตัวอย่างข้อมูลจากองค์กรสถาบันการแพทย์ฉุกเฉินแห่งชาติ ผู้บาดเจ็บจากอุบัติเหตุ ในช่วงเทศกาลปีใหม่ของปี 2551 ถึงปี 2558

ตัวอย่างข้อมูลและสถิติเบื้องต้นของข้อมูลจากองค์กรสถาบันการแพทย์ฉุกเฉินแห่งชาติ ผู้บาดเจ็บจากอุบัติเหตุ ในช่วงเทศกาลปีใหม่ของปี 2551 ถึงปี 2558 ที่เป็นข้อมูลหมวดหมู่ (Categorical Data) ดังภาพประกอบ 11

	ชื่อเทศกาล	จังหวัด	ชื่อโรงพยาบาลที่รับผู้บาดเจ็บ	วันที่เกิดเหตุ	เวลาเกิดเหตุ	เพศ
count	214950	214950	214950	214950	214950	214950
unique	8	77	1192	11	25	2
top	ปีใหม่ 51	นครราชสีมา	บางละมุง	31	18:01-19:00 น.	ชาย
freq	30866	9835	1758	41940	16816	145146

	ถนนที่เกิดเหตุ	สถานะ	รถผู้บาดเจ็บ	รถคู่กรณี	มาตรการ	การดื่มสุรา	การนำส่ง	Refer-Admit	ผลการรักษา
count	214950	214950	214950	214950	214950	214950	214950	214950	214950
unique	4	4	13	13	4	3	9	4	7
top	ชนบท	ผู้ขับขี่	จักรยานยนต์	ไม่มี/ล้มเอง	ไม่ใส่	ไม่ดื่ม	ผู้ประสบเหตุ/ญาติ	ไม่	ทุเลา/หาย
freq	105636	149667	170992	116115	164436	126277	126951	162045	211386

ภาพประกอบ 11 ตัวอย่างข้อมูลองค์กรสถาบันการแพทย์ฉุกเฉินแห่งชาติ ผู้บาดเจ็บจากอุบัติเหตุที่เป็นหมวดหมู่ (Categorical Data)

ตัวอย่างข้อมูลและสถิติเบื้องต้นของข้อมูลจากองค์กรสถาบันการแพทย์ฉุกเฉินแห่งชาติ ผู้บาดเจ็บจากอุบัติเหตุ ในช่วงเทศกาลปีใหม่ของปี 2551 ถึงปี 2558 ที่เป็นข้อมูลต่อเนื่อง (Continuous Data) ดังภาพประกอบ 12

	อายุ	จำนวนวันรักษา
count	214950.000000	214950.000000
mean	28.861433	0.344359
std	16.402318	1.774047
min	0.000000	0.000000
25%	17.000000	0.000000
50%	25.000000	0.000000
75%	39.000000	0.000000
max	99.000000	31.000000

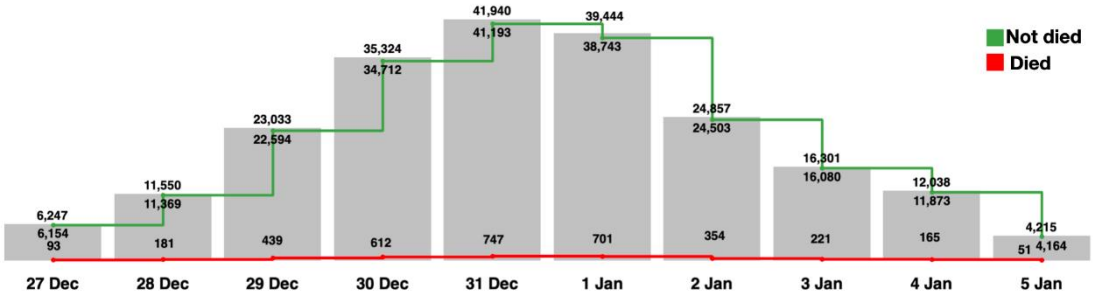
ภาพประกอบ 12 ตัวอย่างข้อมูลองค์กรสถาบันการแพทย์ฉุกเฉินแห่งชาติ ผู้บาดเจ็บจากอุบัติเหตุที่เป็นข้อมูลต่อเนื่อง (Continuous Data)

3.3 **การสำรวจข้อมูล** Exploratory Data Analysis (EDA)

การสำรวจและวิเคราะห์ความสัมพันธ์ระหว่างข้อมูลผลการรักษาของผู้ประสบอุบัติเหตุกับ Feature ตัวอื่นๆ ประกอบด้วย วันที่เกิดเหตุ, ชื่อเทศกาล, เวลาเกิดเหตุ, อายุ, สถานะ, รถผู้บาดเจ็บ, รถคู่กรณี, มาตรการ, ถนนที่เกิดเหตุ, เพศ, ชื่อโรงพยาบาลที่รับผู้บาดเจ็บ, จำนวนวันรักษา, ชื่อจังหวัด, การนำส่ง, การดื่มสุรา

จากภาพประกอบ 13 ได้ทำการวิเคราะห์ข้อมูลความสัมพันธ์ระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับวันที่เกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ของวันที่ 27 ธันวาคม ถึง วันที่ 5 มกราคม ของปี 2551 ถึง ปี 2558 จะเห็นได้ว่าแนวโน้มของการเกิดอุบัติเหตุสูงสุดเกิดในวันที่ 31 ธันวาคม และวันที่ 1 มกราคม

โดยวันที่ 31 ธันวาคม เกิดจำนวนของอุบัติเหตุทั้งหมด 41,940 ครั้ง แบ่งออกเป็นไม่เสียชีวิต 41,193 ราย และเสียชีวิต 747 ราย และในวันที่ 1 มกราคม เกิดจำนวนของอุบัติเหตุทั้งหมด 39,444 ครั้ง แบ่งออกเป็นไม่เสียชีวิต 38,743 ราย และเสียชีวิต 701 ราย เพราะในช่วงสองวันนี้เป็นวันส่งท้ายปีเก่าและเป็นวันปีใหม่ ประชากรส่วนใหญ่อยู่ในช่วงของการทำกิจกรรมฉลองเทศกาลปีใหม่ เป็นเหตุทำให้เกิดอุบัติเหตุสูงขึ้น แต่วันที่อัตราการเสียชีวิต (Mortality Rate) สูงสุดคือวันที่ 29 ธันวาคม มีค่าเท่ากับ 1.94%

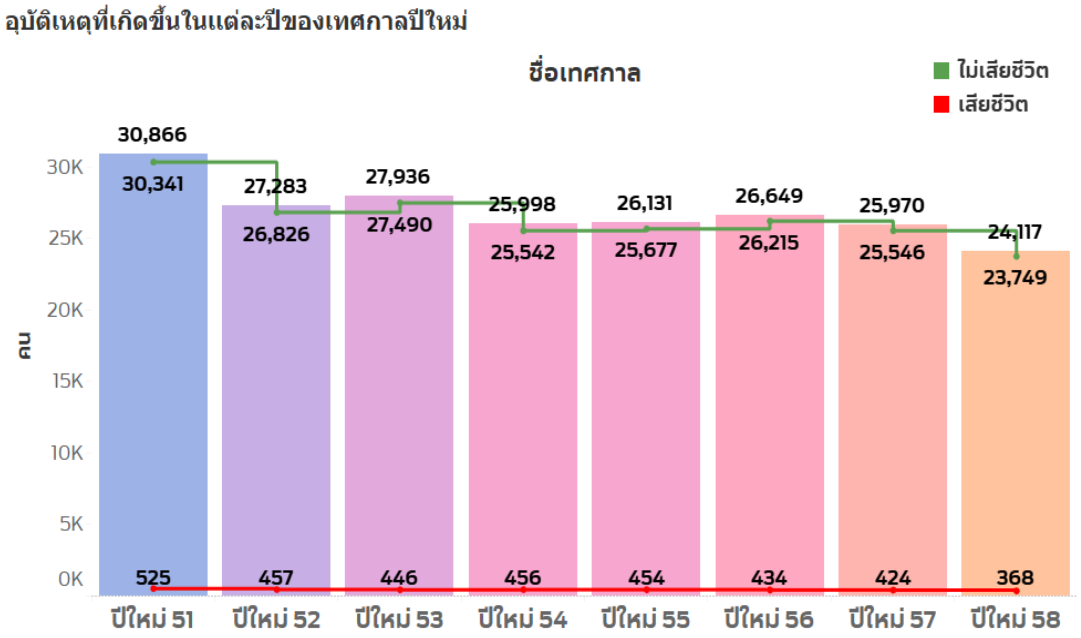


Date of the accident	Not died	Died	Total	Mortality rate (%)
27 Dec	6,154	93	6,247	1.51
28 Dec	11,369	181	11,550	1.59
29 Dec	22,594	439	23,033	1.94
30 Dec	34,712	612	35,324	1.76
31 Dec	41,193	747	41,940	1.81
1 Jan	38,743	701	39,444	1.81
2 Jan	24,503	354	24,857	1.44
3 Jan	16,080	221	16,301	1.37
4 Jan	11,873	165	12,038	1.39
5 Jan	4,164	51	4,215	1.22
Total	**211,385**	**3,564**	**214,949**	**1.69**

ภาพประกอบ 13 ข้อมูลกราฟแท่งแสดงความสัมพันธ์ระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับวันที่เกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558

จากภาพประกอบ 14,15 และ 16 ได้ทำการวิเคราะห์ข้อมูลความสัมพันธ์ระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับชื่อเทศกาลปีใหม่ของปี 2551 ถึง ปี 2558 จะเห็นได้ว่าเกิดอุบัติเหตุสูงสุดเกิดในปี 2551 และมีแนวโน้มลดลงถึงปี 2558

โดยปี 2551 เกิดจำนวนของอุบัติเหตุทั้งหมด 30,866 ครั้ง แบ่งออกเป็นไม่เสียชีวิต 30,341 ราย และเสียชีวิต 525 ราย แต่ในปี 2558 เกิดจำนวนของอุบัติเหตุทั้งหมด 24,117 ครั้ง แบ่งออกเป็นไม่เสียชีวิต 23,749 ราย และเสียชีวิต 368 ราย เพราะมีการรณรงค์บังคับใช้กฎหมายจราจรอย่างจริงจังมากขึ้น ทำให้จำนวนของอุบัติเหตุลดลง เช่นเดียวกับอัตราการเสียชีวิตที่ลดลง



ภาพประกอบ 14 ข้อมูลกราฟแท่งแสดงความสัมพันธ์ระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับชื่อเทศกาลปีใหม่ 2551 ถึง ปี 2558

ชื่อเทศกาล	ไม่เสียชีวิต	เสียชีวิต	จำนวนทั้งหมด
ปีใหม่ 51	14.12%	0.24%	14.36%
ปีใหม่ 52	12.48%	0.21%	12.69%
ปีใหม่ 54	11.88%	0.21%	12.09%
ปีใหม่ 55	11.95%	0.21%	12.16%
ปีใหม่ 53	12.79%	0.21%	13.00%
ปีใหม่ 56	12.20%	0.20%	12.40%
ปีใหม่ 57	11.88%	0.20%	12.08%
ปีใหม่ 58	11.05%	0.17%	11.22%
จำนวนทั้งหมด	98.34%	1.66%	100.00%

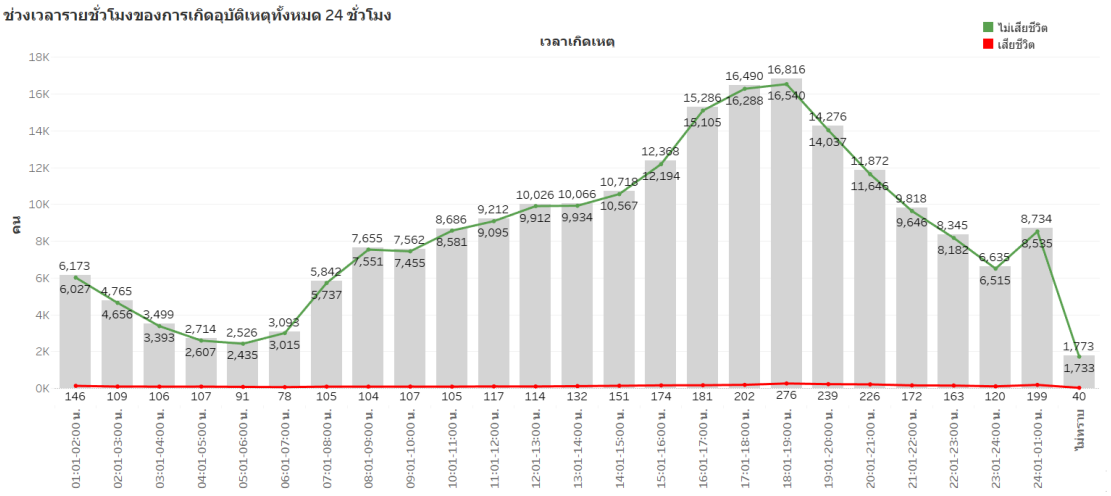
ภาพประกอบ 15 ข้อมูลเปอร์เซ็นต์เปรียบเทียบระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับชื่อเทศกาลปีใหม่ 2551 ถึง ปี 2558

ชื่อเทศกาล	ไม่เสียชีวิต	เสียชีวิต	จำนวนทั้งหมด	Mortality Rate (%)
ปีใหม่ 51	30,341	525	30,866	1.73
ปีใหม่ 52	26,826	457	27,283	1.70
ปีใหม่ 53	27,490	446	27,936	1.62
ปีใหม่ 54	25,542	456	25,998	1.79
ปีใหม่ 55	25,677	454	26,131	1.77
ปีใหม่ 56	26,215	434	26,649	1.66
ปีใหม่ 57	25,545	424	25,969	1.66
ปีใหม่ 58	23,749	368	24,117	1.55
จำนวนทั้งหมด	**211,385**	**3,564**	**214,949**	**1.69**

ภาพประกอบ 16 ข้อมูลเปรียบเทียบระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับชื่อเทศกาลปีใหม่และค่าอัตราการเสียชีวิตในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558

จากภาพประกอบ 17, 18 และ 19 ได้ทำการวิเคราะห์ข้อมูลความสัมพันธ์ระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับช่วงเวลารายชั่วโมงของการเกิดอุบัติเหตุทั้ง 24 ชั่วโมง ในช่วงเทศกาลปีใหม่ของวันที่ 27 ธันวาคม ถึง วันที่ 5 มกราคม ของปี 2551 ถึง ปี 2558 จะเห็นได้ว่าแนวโน้มของการเกิดอุบัติเหตุสูงสุดเกิดในช่วงเวลา 18:01 น. ถึง 19:00 น. แต่ช่วงการเกิดอุบัติเหตุต่ำสุดเกิดในช่วงเวลา 05:01 น. ถึง 06:00 น.

โดยในช่วงเวลา 18:01 น. ถึง 19:00 น. เกิดจำนวนของอุบัติเหตุทั้งหมด 16,816 ครั้ง แบ่งออกเป็นไม่เสียชีวิต 16,540 ราย และเสียชีวิต 276 ราย และในช่วงเวลา 05:01 น. ถึง 06:00 น. เกิดจำนวนของอุบัติเหตุทั้งหมด 2,526 ครั้ง แบ่งออกเป็นไม่เสียชีวิต 2,435 ราย และเสียชีวิต 91 ราย เพราะช่วงเวลา 18:01 น. ถึง 19:00 น. เป็นผลมาจากเส้นทางถนนมีจราจรที่หนาแน่น ประชากรส่วนใหญ่เดินทางเพื่อไปสังสรรค์ในงานเทศกาลปีใหม่ ในขณะเดียวกันอัตราการเสียชีวิตสูงที่สุดในช่วงใกล้สว่าง คาดว่าเกิดจากผู้ขับขี่ดื่มแอลกอฮอล์ รวมถึงความมืดและความเร็วในการขับรถ



ภาพประกอบ 17 ข้อมูลกราฟแท่งแสดงความสัมพันธ์ระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับช่วงเวลารายชั่วโมงในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558

เวลาเกิดเหตุ	ไม่เสียชีวิต	เสียชีวิต	จำนวนทั้งหมด
01:01-02:00 น.	2.80%	0.07%	**2.87%**
02:01-03:00 น.	2.17%	0.05%	**2.22%**
03:01-04:00 น.	1.58%	0.05%	**1.63%**
04:01-05:00 น.	1.21%	0.05%	**1.26%**
05:01-06:00 น.	1.13%	0.04%	**1.18%**
06:01-07:00 น.	1.40%	0.04%	**1.44%**
07:01-08:00 น.	2.67%	0.05%	**2.72%**
08:01-09:00 น.	3.51%	0.05%	**3.56%**
09:01-10:00 น.	3.47%	0.05%	**3.52%**
10:01-11:00 น.	3.99%	0.05%	**4.04%**
11:01-12:00 น.	4.23%	0.05%	**4.29%**
12:01-13:00 น.	4.61%	0.05%	**4.66%**
13:01-14:00 น.	4.62%	0.06%	**4.68%**
14:01-15:00 น.	4.92%	0.07%	**4.99%**
15:01-16:00 น.	5.67%	0.08%	**5.75%**
16:01-17:00 น.	7.03%	0.08%	**7.11%**
17:01-18:00 น.	7.58%	0.09%	**7.67%**
18:01-19:00 น.	7.69%	0.13%	**7.82%**
19:01-20:00 น.	6.53%	0.11%	**6.64%**
20:01-21:00 น.	5.42%	0.11%	**5.52%**
21:01-22:00 น.	4.49%	0.08%	**4.57%**
22:01-23:00 น.	3.81%	0.08%	**3.88%**
23:01-24:00 น.	3.03%	0.06%	**3.09%**
24:01-01:00 น.	3.97%	0.09%	**4.06%**
ไม่ทราบ	0.81%	0.02%	**0.82%**
จำนวนทั้งหมด	**98.34%**	**1.66%**	**100.00%**

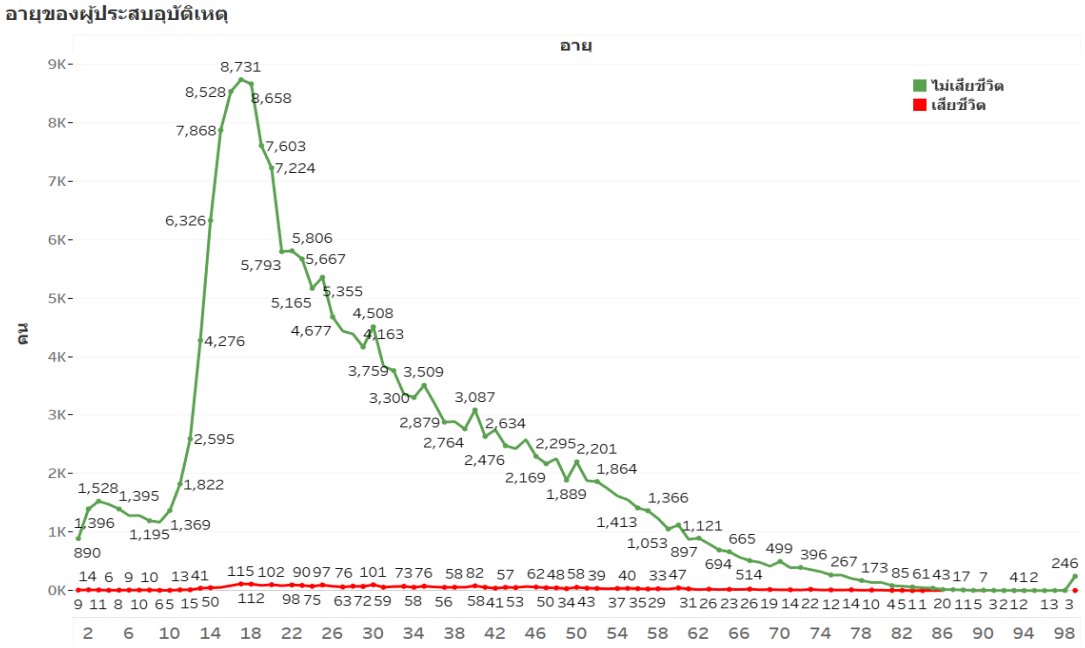
ภาพประกอบ 18 ข้อมูลเปอร์เซ็นต์เปรียบเทียบระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับช่วงเวลารายชั่วโมงในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558

เวลาเกิดเหตุ	ไม่เสียชีวิต	เสียชีวิต	จำนวนทั้งหมด	Mortality Rate (%)
01:01-02:00 น.	6,027	146	6,173	2.42
02:01-03:00 น.	4,656	109	4,765	2.34
03:01-04:00 น.	3,393	106	3,499	3.12
04:01-05:00 น.	2,607	107	2,714	4.10
05:01-06:00 น.	2,435	91	2,526	3.74
06:01-07:00 น.	3,015	78	3,093	2.59
07:01-08:00 น.	5,737	105	5,842	1.83
08:01-09:00 น.	7,551	104	7,655	1.38
09:01-10:00 น.	7,455	107	7,562	1.44
10:01-11:00 น.	8,581	105	8,686	1.22
11:01-12:00 น.	9,095	117	9,212	1.29
12:01-13:00 น.	9,912	114	10,026	1.15
13:01-14:00 น.	9,933	132	10,065	1.33
14:01-15:00 น.	10,567	151	10,718	1.43
15:01-16:00 น.	12,194	174	12,368	1.43
16:01-17:00 น.	15,105	181	15,286	1.20
17:01-18:00 น.	16,288	202	16,490	1.24
18:01-19:00 น.	16,540	276	16,816	1.67
19:01-20:00 น.	14,037	239	14,276	1.70
20:01-21:00 น.	11,646	226	11,872	1.94
21:01-22:00 น.	9,646	172	9,818	1.78
22:01-23:00 น.	8,182	163	8,345	1.99
23:01-24:00 น.	6,515	120	6,635	1.84
24:01-01:00 น.	8,535	199	8,734	2.33
ไม่ทราบ	1,733	40	1,773	2.31
จำนวนทั้งหมด	**211,385**	**3,564**	**214,949**	**1.69**

ภาพประกอบ 19 ข้อมูลเปรียบเทียบระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับช่วงเวลารายชั่วโมงและค่าอัตราการเสียชีวิตในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558

จากภาพประกอบ 20, 21 และ 22 ได้ทำการวิเคราะห์ข้อมูลความสัมพันธ์ระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับอายุของผู้ประสบอุบัติเหตุในช่วงเทศกาลปีใหม่ของวันที่ 27 ธันวาคม ถึง วันที่ 5 มกราคม ของปี 2551 ถึง ปี 2558 จะเห็นได้ว่าแนวโน้มของการเกิดอุบัติเหตุสูงสุดเกิดในช่วงอายุ 16 ถึง 20 ปี

โดยในช่วงอายุ 16 ถึง 20 ปี เกิดจำนวนของอุบัติเหตุทั้งหมด 41,249 ครั้ง แบ่งออกเป็นไม่เสียชีวิต 40,744 ราย และเสียชีวิต 505 ราย เพราะช่วงอายุของผู้ประสบอุบัติเหตุดังกล่าว เป็นช่วงของวัยรุ่นคึกคะนอง ทำให้ประมาทในการใช้เส้นทางถนนทำให้เกิดอุบัติเหตุขึ้นสูง แต่จากการวิเคราะห์อัตราการเสียชีวิต จะพบว่าผู้สูงอายุมีโอกาสเสียชีวิตสูงที่สุด เนื่องจากความเปราะบางของร่างกาย



ภาพประกอบ 20 ข้อมูลกราฟแท่งแสดงความสัมพันธ์ระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับอายุของผู้ประสบอุบัติเหตุในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558

ช่วงอายุ	ไม่เสียชีวิต	เสียชีวิต	จำนวนทั้งหมด
1-5	4.97%	0.05%	**5.02%**
6-10	2.93%	0.02%	**2.95%**
11-15	10.65%	0.08%	**10.73%**
16-20	18.96%	0.23%	**19.19%**
21-25	12.93%	0.21%	**13.13%**
26-30	10.31%	0.18%	**10.49%**
31-35	8.26%	0.16%	**8.42%**
36-40	6.89%	0.15%	**7.04%**
41-45	5.99%	0.13%	**6.11%**
46-50	5.03%	0.12%	**5.15%**
51-55	4.03%	0.09%	**4.12%**
56-60	2.88%	0.08%	**2.96%**
61-65	1.83%	0.05%	**1.88%**
66-70	1.16%	0.04%	**1.20%**
71-75	0.81%	0.03%	**0.84%**
76-80	0.43%	0.02%	**0.45%**
81-85	0.15%	0.01%	**0.15%**
86-90	0.03%	0.00%	**0.03%**
91-95	0.01%	0.00%	**0.01%**
96-100	0.12%	0.00%	**0.12%**
จำนวนทั้งหมด	**98.34%**	**1.66%**	**100.00%**

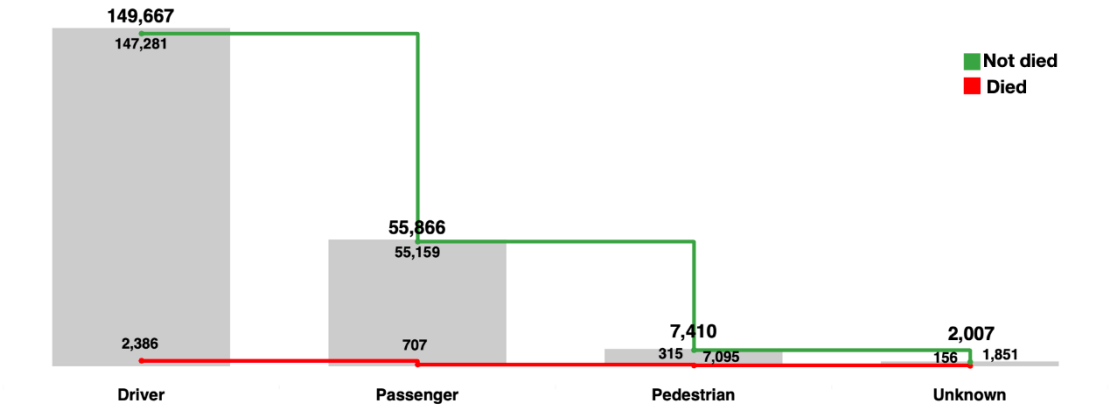
ภาพประกอบ 21 ข้อมูลเปอร์เซ็นต์เปรียบเทียบระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับอายุของผู้ประสบอุบัติเหตุในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558

ช่วงอายุ	ไม่เสียชีวิต	เสียชีวิต	จำนวนทั้งหมด	Mortality Rate (%)
1-5	10,679	115	10,794	1.08
6-10	6,299	40	6,339	0.64
11-15	22,886	175	23,061	0.76
16-20	40,744	505	41,249	1.24
21-25	27,786	445	28,231	1.60
26-30	22,167	387	22,554	1.75
31-35	17,764	336	18,100	1.89
36-40	14,820	319	15,139	2.15
41-45	12,865	279	13,144	2.17
46-50	10,810	252	11,062	2.33
51-55	8,661	192	8,853	2.22
56-60	6,183	171	6,354	2.77
61-65	3,932	115	4,047	2.92
66-70	2,484	93	2,577	3.74
71-75	1,737	70	1,807	4.03
76-80	926	48	974	5.18
81-85	316	12	328	3.80
86-90	60	5	65	8.33
91-95	13	2	15	15.38
96-100	253	3	256	1.19
จำนวนทั้งหมด	**211,385**	**3,564**	**214,949**	**1.69**

ภาพประกอบ 22 ข้อมูลเปรียบเทียบระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับอายุของผู้ประสบอุบัติเหตุและค่าอัตราการเสียชีวิตในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558

จากภาพประกอบ 23 ได้ทำการวิเคราะห์ข้อมูลความสัมพันธ์ระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับสถานะของผู้ประสบอุบัติเหตุในช่วงเทศกาลปีใหม่ของวันที่ 27 ธันวาคม ถึง วันที่ 5 มกราคม ของปี 2551 ถึง ปี 2558 จะเห็นได้ว่าแนวโน้มของการเกิดอุบัติเหตุสูงสุดเกิดในสถานะของผู้ขับขี่และรองมาเป็นผู้โดยสาร

โดยในสถานะของผู้ขับขี่เกิดจำนวนของอุบัติเหตุทั้งหมด 149,667 ครั้ง แบ่งออกเป็นไม่เสียชีวิต 147,281 ราย และเสียชีวิต 2,386 ราย และสถานะของผู้โดยสารเกิดจำนวนของอุบัติเหตุทั้งหมด 55,866 ครั้ง แบ่งออกเป็นไม่เสียชีวิต 55,159 ราย และเสียชีวิต 707 ราย ในส่วนอัตราการเสียชีวิต หากไม่นับข้อมูลที่ไม่ทราบ คนเดินถนนได้รับผลกระทบมากที่สุด

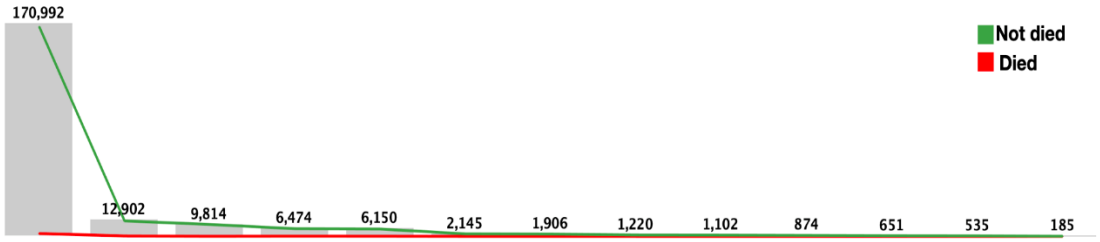


Victim status	Not died	Died	Total	Mortality rate (%)
Driver	147,280	2,386	149,666	1.62
Passenger	55,159	707	55,866	1.28
Pedestrian	7,095	315	7,410	4.44
Unknown	1,851	156	2,007	8.43
Total	**211,385**	**3,564**	**214,949**	**1.69**

ภาพประกอบ 23 ข้อมูลกราฟแท่งแสดงความสัมพันธ์ระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับสถานะของผู้ประสบอุบัติเหตุในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558

จากภาพประกอบ 24 ได้ทำการวิเคราะห์ข้อมูลความสัมพันธ์ระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับยานพาหนะของผู้ขับขี่ที่ประสบอุบัติเหตุ ในช่วงเทศกาลปีใหม่ของวันที่ 27 ธันวาคม ถึง วันที่ 5 มกราคม ของปี 2551 ถึง ปี 2558 จะเห็นได้ว่าแนวโน้มของการเกิดอุบัติเหตุสูงสุดเกิดจากรถจักรยานยนต์และรองมาเป็นรถกระบะปิคอัพ

โดยที่ยานพาหนะของผู้ขับขี่รถจักรยานยนต์เกิดจำนวนของอุบัติเหตุทั้งหมด 170,992 ครั้ง แบ่งออกเป็นไม่เสียชีวิต 168,694 ราย และเสียชีวิต 2,298 ราย และยานพาหนะของผู้ขับขี่รถกระบะปิคอัพเกิดจำนวนของอุบัติเหตุทั้งหมด 12,902 ครั้ง แบ่งออกเป็นไม่เสียชีวิต 12,476 ราย และเสียชีวิต 426 ราย

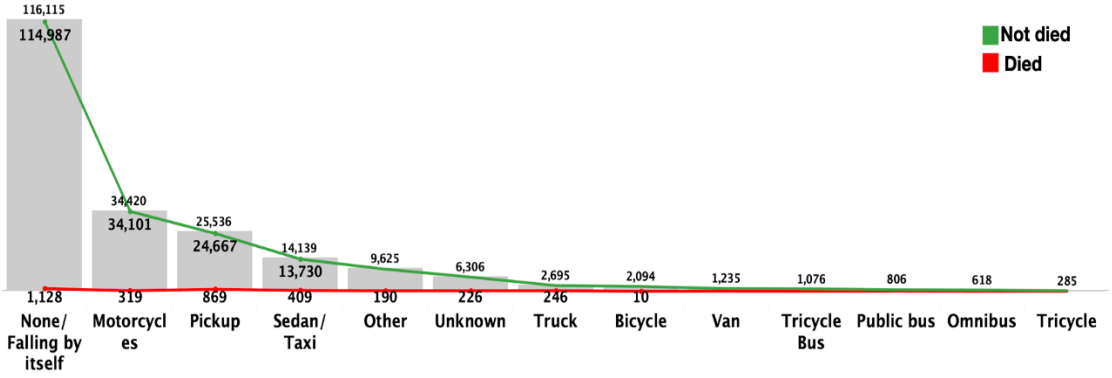


Driver's vehicle	Not died	Died	Total	Mortality rate (%)
None/Falling by itself	6,211	263	6,474	4.23
Truck	627	24	651	3.83
Other	2,071	74	2,145	3.57
Sedan/Taxi	5,944	206	6,150	3.47
Pickup	12,476	426	12,902	3.41
Unknown	1,182	38	1,220	3.21
Tricycle Bus	1,861	45	1,906	2.42
Van	1,080	22	1,102	2.04
Omnibus	526	9	535	1.71
Public bus	861	13	874	1.51
Bicycle	9,670	144	9,814	1.49
Motorcycles	168,693	2,298	170,991	1.36
Tricycle	183	2	185	1.09
Total	**211,385**	**3,564**	**214,949**	**1.69**

ภาพประกอบ 24 ข้อมูลกราฟแท่งแสดงความสัมพันธ์ระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับยานพาหนะของผู้ขับขี่ที่ประสบอุบัติเหตุในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558

จากภาพประกอบ 25 ได้ทำการวิเคราะห์ข้อมูลความสัมพันธ์ระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับยานพาหนะของรถคู่กรณีของผู้ประสบอุบัติเหตุ ในช่วงเทศกาลปีใหม่ของวันที่ 27 ธันวาคม ถึง วันที่ 5 มกราคม ของปี 2551 ถึง ปี 2558 จะเห็นได้ว่าแนวโน้มของการเกิดอุบัติเหตุสูงสุดเกิดจากยานพาหนะของรถผู้ประสบอุบัติเหตุเองไม่มีคู่กรณี และรองมาเป็นรถจักรยานยนต์

โดยที่ยานพาหนะของรถผู้ประสบอุบัติเหตุเองไม่มีคู่กรณีเกิดจำนวนของอุบัติเหตุทั้งหมด 116,115 ครั้ง แบ่งออกเป็นไม่เสียชีวิต 114,987 ราย และเสียชีวิต 1,128 ราย และยานพาหนะของรถคู่กรณีที่เป็นรถจักรยานยนต์เกิดจำนวนของอุบัติเหตุทั้งหมด 34,420 ครั้ง แบ่งออกเป็นไม่เสียชีวิต 34,101 ราย และเสียชีวิต 319 ราย หากประสบเหตุที่คู่กรณีเป็นรถใหญ่เป็นรถสาธารณะหรือรถบรรทุก จะมีอัตราการเสียชีวิตค่อนข้างสูง



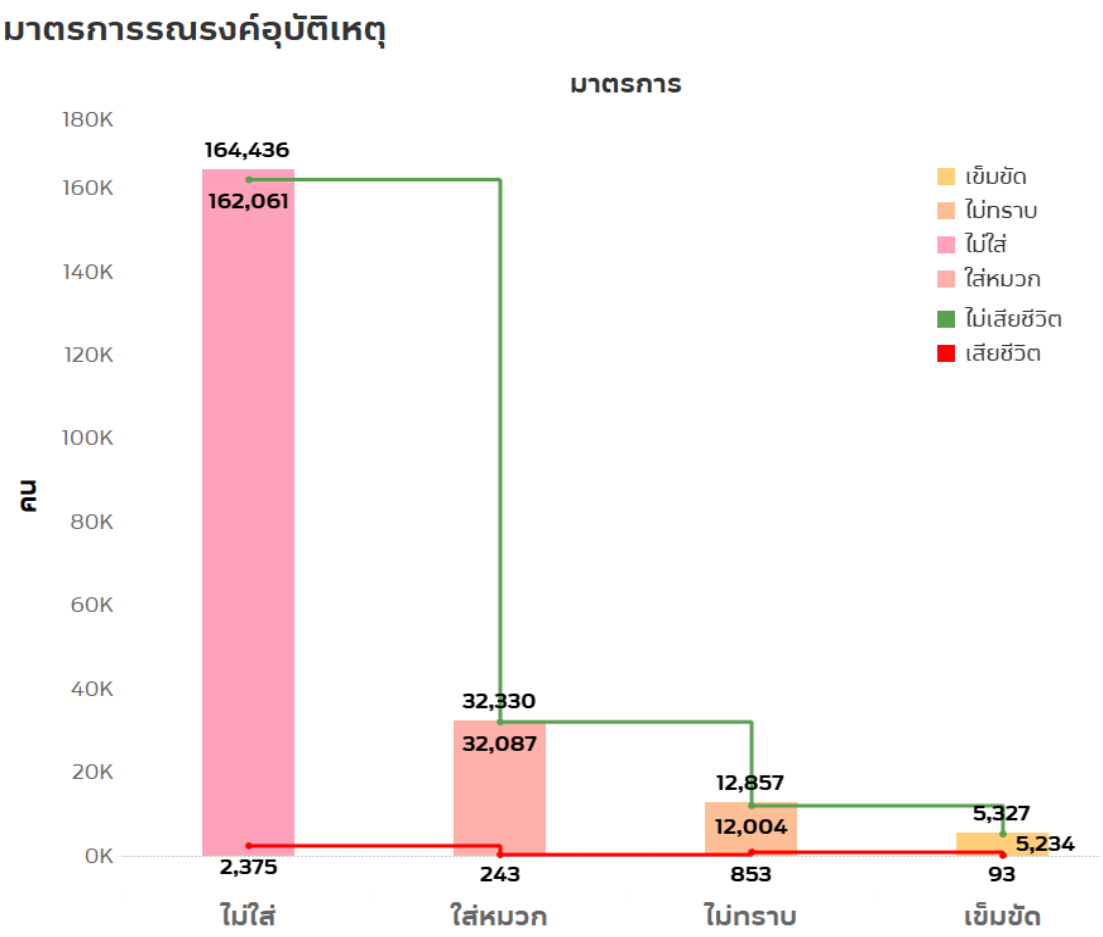
Parties' vehicle	Not died	Died	Total	Mortality rate (%)
Public bus	722	84	806	11.63
Truck	2,449	246	2,695	10.04
Van	1,175	60	1,235	5.11
Unknown	6,080	226	6,306	3.72
Pickup	24,667	869	25,536	3.52
Sedan/Taxi	13,730	409	14,139	2.98
Other	9,435	190	9,625	2.01
Omnibus	606	12	618	1.98
None/Falling by itself	114,986	1,128	116,114	0.98
Tricycle Bus	1,066	10	1,076	0.94
Motorcycles	34,101	319	34,420	0.94
Bicycle	2,084	10	2,094	0.48
Tricycle	284	1	285	0.35
Total	**211,385**	**3,564**	**214,949**	**1.69**

ภาพประกอบ 25 ข้อมูลกราฟแท่งแสดงความสัมพันธ์ระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับยานพาหนะของรถคู่กรณีของผู้ประสบอุบัติเหตุในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558

จากภาพประกอบ 26, 27 และ 28 ได้ทำการวิเคราะห์ข้อมูลความสัมพันธ์ระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับมาตรการรณรงค์ลดอุบัติเหตุในช่วงเทศกาลปีใหม่ของวันที่ 27 ธันวาคม ถึง วันที่ 5 มกราคม ของปี 2551 ถึง ปี 2558 จะเห็นได้ว่าแนวโน้มของการเกิดอุบัติเหตุสูงสุดเกิดจากไม่สวมหมวกนิรภัย และรองมาเป็นสวมหมวกนิรภัย

โดยที่ผู้ประสบอุบัติเหตุที่ไม่สวมหมวกนิรภัยเกิดจำนวนของอุบัติเหตุทั้งหมด 164,436 ครั้ง แบ่งออกเป็นไม่เสียชีวิต 162,061 ราย และเสียชีวิต 2,375 ราย และผู้ประสบอุบัติเหตุที่สวมหมวกนิรภัยเกิดจำนวนของอุบัติเหตุทั้งหมด 32,330 ครั้ง แบ่งออกเป็นไม่เสียชีวิต 32,087 ราย และเสียชีวิต 243 ราย

จะเห็นได้ว่าผู้ประสบอุบัติเหตุที่ไม่สวมหมวกนิรภัยเกิดจำนวนของอุบัติเหตุสูงกว่าผู้ประสบอุบัติเหตุที่สวมหมวกนิรภัยถึง 5 เท่า ดังนี้มาตรการรณรงค์ลดอุบัติเหตุในช่วงเทศกาลปีใหม่ได้ผลสามารถช่วยลดจำนวนอบัติเหตุลดลงและลดอัตราการเสียชีวิตได้



ภาพประกอบ 26 ข้อมูลกราฟแท่งแสดงความสัมพันธ์ระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับมาตรการรณรงค์ลดอุบัติเหตุในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558

มาตรการ	ไม่เสียชีวิต	เสียชีวิต	จำนวนทั้งหมด
เข็มขัด	2.43%	0.04%	2.48%
ไม่ทราบ	5.58%	0.40%	5.98%
ไม่ใส่	75.39%	1.10%	76.50%
ใส่หมวก	14.93%	0.11%	15.04%
จำนวนทั้งหมด	98.34%	1.66%	100.00%

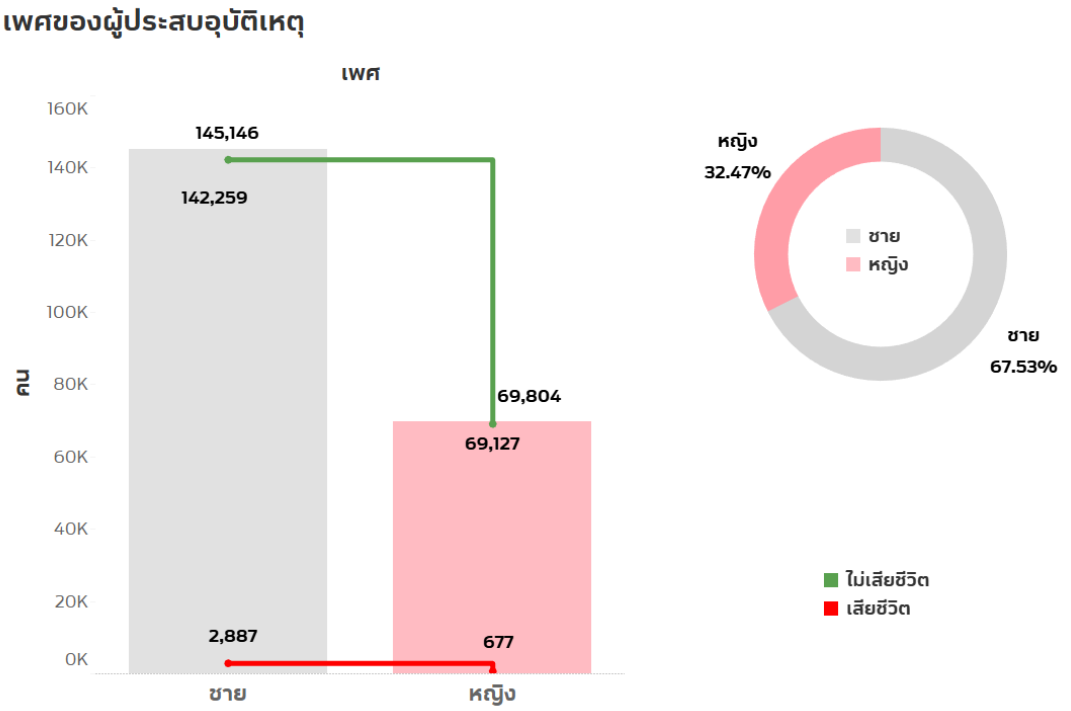
ภาพประกอบ 27 ข้อมูลเปอร์เซ็นต์เปรียบเทียบระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับมาตรการรณรงค์ลดอุบัติเหตุในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558

มาตรการ	ไม่เสียชีวิต	เสียชีวิต	จำนวนทั้งหมด	Mortality Rate (%)
ไม่ทราบ	12,004	853	12,857	7.11
เข็มขัด	5,234	93	5,327	1.78
ไม่ใส่	162,060	2,375	164,435	1.47
ใส่หมวก	32,087	243	32,330	0.76
จำนวนทั้งหมด	**211,385**	**3,564**	**214,949**	**1.69**

ภาพประกอบ 28 ข้อมูลเปรียบเทียบระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับมาตรการรณรงค์ลดอุบัติเหตุและค่าอัตราการเสียชีวิตในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558

จากภาพประกอบ 29, 30 และ 31 ได้ทำการวิเคราะห์ข้อมูลความสัมพันธ์ระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับเพศของผู้ประสบอุบัติเหตุในช่วงเทศกาลปีใหม่ของวันที่ 27 ธันวาคม ถึง วันที่ 5 มกราคม ของปี 2551 ถึง ปี 2558 จะเห็นได้ว่าแนวโน้มของการเกิดอุบัติเหตุสูงสุดเกิดจากเพศชายและรองมาเป็นเพศหญิง

โดยที่ผู้ประสบอุบัติเหตุที่เป็นเพศชายเกิดจำนวนของอุบัติเหตุทั้งหมด 145,146 ครั้ง แบ่งออกเป็นไม่เสียชีวิต 142,259 ราย และเสียชีวิต 2,887 ราย และผู้ประสบอุบัติเหตุที่เป็นเพศหญิงเกิดจำนวนของอุบัติเหตุทั้งหมด 69,804 ครั้ง แบ่งออกเป็นไม่เสียชีวิต 69,127 ราย และเสียชีวิต 677 ราย จะเห็นได้ว่าผู้ประสบอุบัติเหตุที่เป็นเพศชายเกิดจำนวนของอุบัติเหตุสูงกว่าผู้ประสบอุบัติเหตุที่เป็นเพศหญิงถึง 2 เท่า



ภาพประกอบ 29 ข้อมูลกราฟแท่งแสดงความสัมพันธ์ระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับเพศของผู้ประสบอุบัติเหตุในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558

เพศ	ไม่เสียชีวิต	เสียชีวิต	จำนวนทั้งหมด
ชาย	66.18%	1.34%	67.53%
หญิง	32.16%	0.31%	32.47%
จำนวนทั้งหมด	98.34%	1.66%	100.00%

ภาพประกอบ 30 ข้อมูลเปอร์เซ็นต์เปรียบเทียบระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับเพศของผู้ประสบอุบัติเหตุในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558

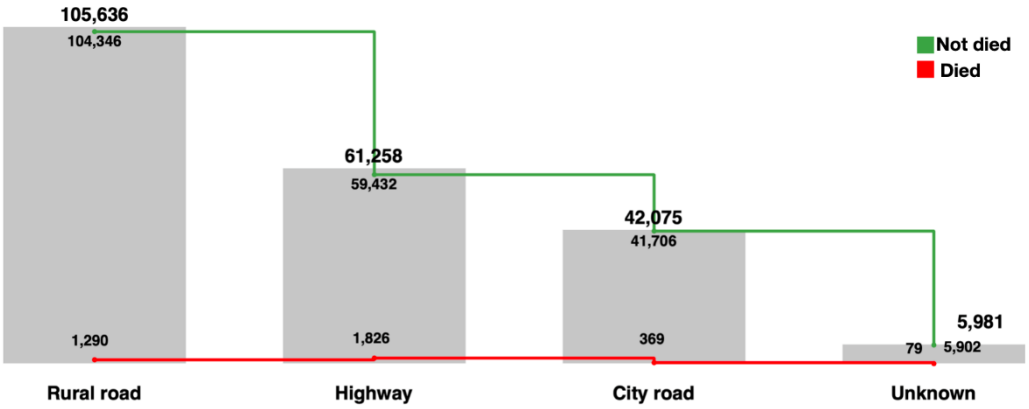
เพศ	ไม่เสียชีวิต	เสียชีวิต	จำนวนทั้งหมด	Mortality Rate (%)
ชาย	142,258	2,887	145,145	2.03
หญิง	69,127	677	69,804	0.98
จำนวนทั้งหมด	**211,385**	**3,564**	**214,949**	**1.69**

ภาพประกอบ 31 ข้อมูลเปรียบเทียบระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับเพศของผู้ประสบอุบัติเหตุและค่าอัตราการเสียชีวิตในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558

จากภาพประกอบ 32 ได้ทำการวิเคราะห์ข้อมูลความสัมพันธ์ระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับเส้นทางถนนที่เกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ของวันที่ 27 ธันวาคม ถึง วันที่ 5 มกราคม ของปี 2551 ถึง ปี 2558 จะเห็นได้ว่าแนวโน้มของการเกิดอุบัติเหตุสูงสุดเกิดจากถนนชนบทและรองมาเป็นถนนทางหลวง

โดยที่ผู้ประสบอุบัติเหตุจากถนนชนบทเกิดจำนวนของอุบัติเหตุทั้งหมด 105,636 ครั้ง แบ่งออกเป็นไม่เสียชีวิต 104,346 ราย และเสียชีวิต 1,290 ราย และผู้ประสบอุบัติเหตุจากทางหลวงเกิดจำนวนของอุบัติเหตุทั้งหมด 61,258 ครั้ง แบ่งออกเป็นไม่เสียชีวิต 59,432 ราย และเสียชีวิต 1,826 ราย

จะเห็นได้ว่าผู้ประสบอุบัติเหตุในถนนชนบทเกิดจำนวนของอุบัติเหตุสูงกว่าผู้ประสบอุบัติเหตุในถนนทางหลวงเกือบเท่าตัว แต่กลับมีผู้เสียชีวิตในถนนหลวงมากกว่าถนนชนบท ดังนั้นความเร็วในการขับขี่และความหนาแน่นของจำนวนรถบนทางถนน ส่งผลต่อการเกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ เพราะถนนหลวงสามารถใช้ความเร็วในการขับขี่ได้มากกว่าถนนชนบท และความหนาแน่นของรถในถนนหลวงมีมากกว่าถนนชนบท



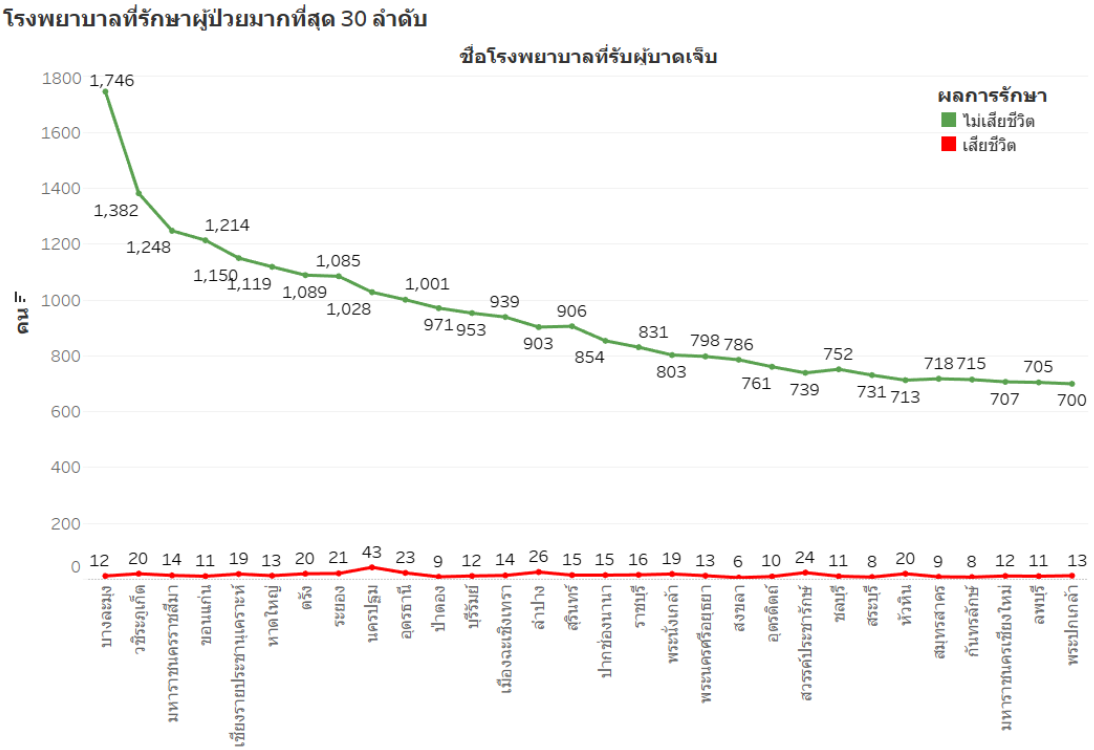
Type of road	Not died	Died	Total	Mortality rate (%)
Highway	59,432	1,826	61,258	3.07
Unknown	5,902	79	5,981	1.34
Rural road	104,345	1,290	105,635	1.24
City road	41,706	369	42,075	0.88
Total	**211,385**	**3,564**	**214,949**	**1.69**

ภาพประกอบ 32 ข้อมูลกราฟแท่งแสดงความสัมพันธ์ระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับเส้นทางถนนที่เกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558

จากภาพประกอบ 33 และ 34 ได้ทำการวิเคราะห์ข้อมูลความสัมพันธ์ระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับโรงพยาบาลที่รักษาผู้บาดเจ็บจากอุบัติเหตุ ในช่วงเทศกาลปีใหม่ของวันที่ 27 ธันวาคม ถึง วันที่ 5 มกราคม ของปี 2551 ถึง ปี 2558 จะเห็นได้ว่าแนวโน้มของการเกิดอุบัติเหตุสูงสุดเกิดที่โรงพยาบาลบางละมุง และรองลงมาเป็นโรงพยาบาลบางวชิระภูเก็ต

โดยที่ผู้บาดเจ็บจากการประสบอุบัติเหตุที่เข้ารักษาที่โรงพยาบาลบางละมุง มีจำนวนของทั้งหมด 1,758 ราย แบ่งออกเป็นไม่เสียชีวิต 1,746 ราย และเสียชีวิต 12 ราย และผู้บาดเจ็บจากการประสบอุบัติเหตุที่เข้ารักษาที่โรงพยาบาลวชิระภูเก็ตมีจำนวนของทั้งหมด 1,402 ราย แบ่งออกเป็นไม่เสียชีวิต 1,382 ราย และเสียชีวิต 20 ราย

จะเห็นได้ว่าข้อมูลสามารถวัดประสิทธิภาพของการรักษาผู้บาดเจ็บจากอุบัติเหตุในแต่ละโรงพยาบาลได้ และเป็นประโยชน์ต่อโรงพยาบาลในการเตรียมพร้อมรับมือในการรักษาผู้บาดเจ็บได้ในอนาคต



ภาพประกอบ 33 ข้อมูลกราฟแท่งแสดงความสัมพันธ์ระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับโรงพยาบาลที่รักษาผู้บาดเจ็บจากอุบัติเหตุจากมากสุด 30 ลำดับแรก ในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558

ชื่อโรงพยาบาลที่รับผู้บาดเจ็บ	ไม่เสียชีวิต	เสียชีวิต	จำนวนทั้งหมด	Mortality Rate (%)
นครปฐม	1,028	43	1,071	4.18
สวรรค์ประชารักษ์	739	24	763	3.25
ลำปาง	903	26	929	2.88
หัวหิน	713	20	733	2.81
พระนั่งเกล้า	803	19	822	2.37
อุดรธานี	1,001	23	1,024	2.30
ระยอง	1,085	21	1,106	1.94
ราชบุรี	831	16	847	1.93
พระปกเกล้า	700	13	713	1.86
ตรัง	1,089	20	1,109	1.84
ปากช่องนานา	854	15	869	1.76
มหาราชนครเชียงใหม่	707	12	719	1.70
สุรินทร์	906	15	921	1.66
เชียงรายประชานุเคราะห์	1,150	19	1,169	1.65
พระนครศรีอยุธยา	798	13	811	1.63
ลพบุรี	705	11	716	1.56
เมืองฉะเชิงเทรา	939	14	953	1.49
ชลบุรี	752	11	763	1.46
วชิระภูเก็ต	1,382	20	1,402	1.45
อุตรดิตถ์	761	10	771	1.31
บุรีรัมย์	953	12	965	1.26
สมุทรสาคร	718	9	727	1.25
หาดใหญ่	1,119	13	1,132	1.16
มหาราชนครราชสีมา	1,248	14	1,262	1.12
กันทรลักษ์	715	8	723	1.12
สระบุรี	731	8	739	1.09
ป่าตอง	971	9	980	0.93
ขอนแก่น	1,214	11	1,225	0.91
สงขลา	786	6	792	0.76
บางละมุง	1,746	12	1,758	0.69

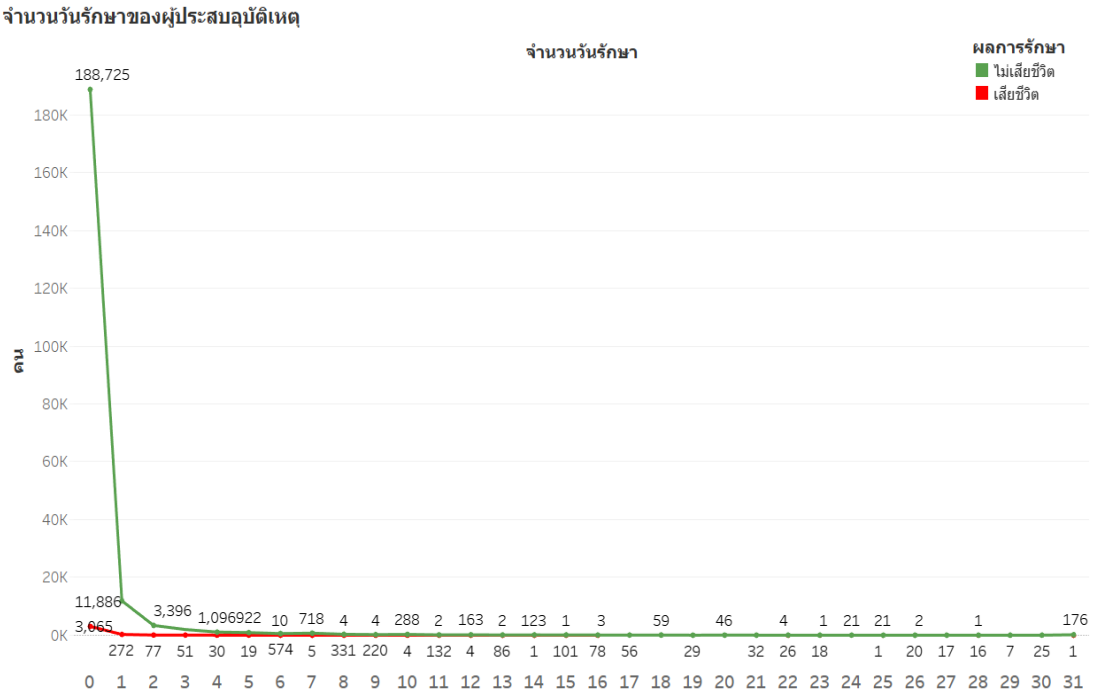
ภาพประกอบ 34 ข้อมูลเปรียบเทียบระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับโรงพยาบาลที่รักษาผู้บาดเจ็บจากอุบัติเหตุจากมากสุด 30 ลำดับแรก และค่าอัตราการเสียชีวิตในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558

จากภาพประกอบ 35, 36 และ 37 ได้ทำการวิเคราะห์ข้อมูลความสัมพันธ์ระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับจำนวนวันรักษาของผู้บาดเจ็บจากอุบัติเหตุในช่วงเทศกาลปีใหม่ของวันที่ 27 ธันวาคม ถึง วันที่ 5 มกราคม ของปี 2551 ถึง ปี 2558 จะเห็นได้ว่าแนวโน้มของจำนวนวันของการรักษาสูงสุดไม่ถึง 24 ชั่วโมง และรองมาเป็น 1 วัน

โดยที่จำนวนวันรักษาของผู้บาดเจ็บจากอุบัติเหตุไม่ถึง 24 ชั่วโมง มีจำนวนของผู้รักษาทั้งหมด 191,790 ราย แบ่งออกเป็นไม่เสียชีวิต 188,725 ราย และเสียชีวิต 3,065 ราย และจำนวนวันรักษาของผู้บาดเจ็บจากอุบัติเหตุ 1 วัน มีจำนวนของผู้รักษาทั้งหมด 12,158 ราย แบ่งออกเป็นไม่เสียชีวิต 11,886 ราย และเสียชีวิต 272 ราย

จะเห็นได้ว่าผู้ประสบอุบัติเหตุที่มีจำนวนวันของการรักษาผู้บาดเจ็บจากอุบัติเหตุไม่ถึง 24 ชั่วโมงสูงกว่าจำนวนวันของการรักษาผู้บาดเจ็บจากอุบัติเหตุ 1 วัน ถึง 15 เท่า แล้วจำนวนของการ

เสียชีวิตของผู้บาดเจ็บจากอุบัติเหตุไม่ถึง 24 ชั่วโมง สูงกว่าผู้บาดเจ็บจากอุบัติเหตุที่มีจำนวนวันของการรักษา 1 วัน ถึง 11 เท่า ทำให้เห็นว่าความรุนแรงของการเกิดอุบัติเหตุที่เกิดขึ้นทั้งหมดได้ว่าผู้เสียชีวิตส่วนใหญ่เกิดอุบัติเหตุที่รุนแรงกว่าปกติถึง 11 เท่า



ภาพประกอบ 35 ข้อมูลกราฟแท่งแสดงความสัมพันธ์ระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับจำนวนวันรักษาของผู้บาดเจ็บในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558

จำนวนวันรักษา	ไม่เสียชีวิต	เสียชีวิต	จำนวนทั้งหมด
0	87.80%	1.43%	89.23%
1	5.53%	0.13%	5.66%
2	1.58%	0.04%	1.62%
3	0.92%	0.02%	0.94%
4	0.51%	0.01%	0.52%
5	0.43%	0.01%	0.44%
6	0.27%	0.00%	0.27%
7	0.33%	0.00%	0.34%
8	0.15%	0.00%	0.16%
9	0.10%	0.00%	0.10%
10	0.13%	0.00%	0.14%
11	0.06%	0.00%	0.06%
12	0.08%	0.00%	0.08%
13	0.04%	0.00%	0.04%
14	0.06%	0.00%	0.06%
15	0.05%	0.00%	0.05%
16	0.04%	0.00%	0.04%
17	0.03%		0.03%
18	0.03%		0.03%
19	0.01%		0.01%
20	0.02%		0.02%
21	0.01%		0.01%
22	0.01%	0.00%	0.01%
23	0.01%	0.00%	0.01%
24	0.01%		0.01%
25	0.01%	0.00%	0.01%
26	0.01%	0.00%	0.01%
27	0.01%		0.01%
28	0.01%	0.00%	0.01%
29	0.00%		0.00%
30	0.01%		0.01%
31	0.08%	0.00%	0.08%
จำนวนทั้งหมด	98.34%	1.66%	100.00%

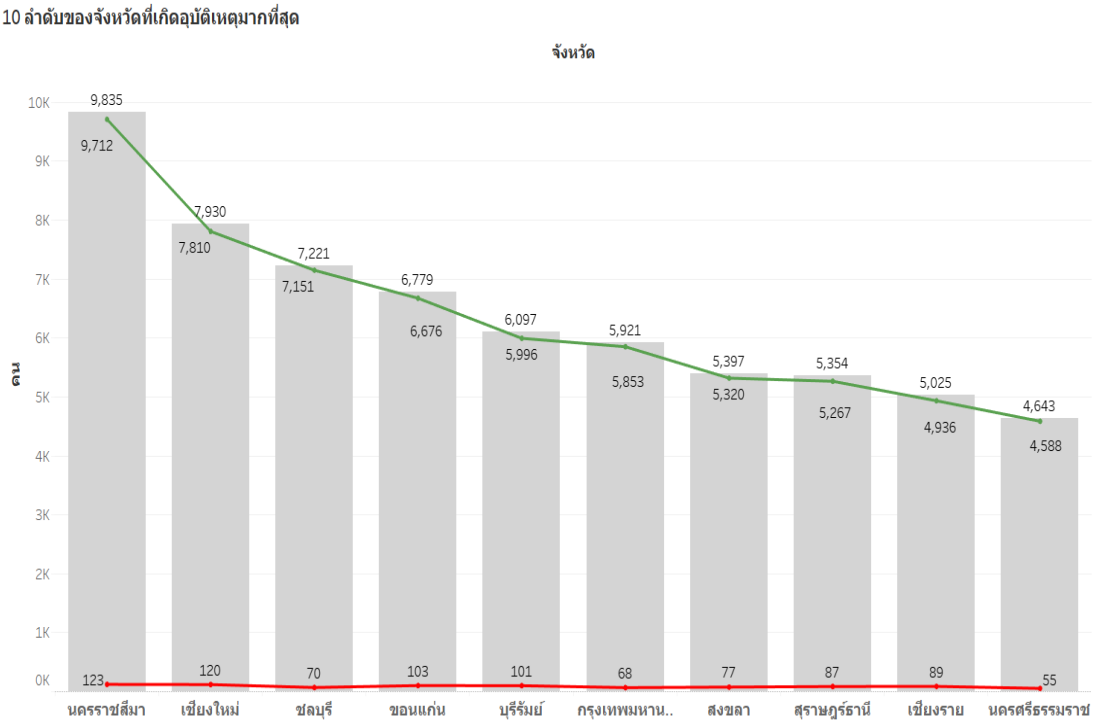
ภาพประกอบ 36 ข้อมูลเปอร์เซ็นต์เปรียบเทียบระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับจำนวนวันรักษาของผู้บาดเจ็บในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558

จำนวนวันรักษา	ไม่เสียชีวิต	เสียชีวิต	จำนวนทั้งหมด	Mortality Rate (%)
0	188,724	3,065	191,789	1.62
1	11,886	272	12,158	2.29
2	3,396	77	3,473	2.27
3	1,978	51	2,029	2.58
4	1,096	30	1,126	2.74
5	922	19	941	2.06
6	574	10	584	1.74
7	718	5	723	0.70
8	331	4	335	1.21
9	220	4	224	1.82
10	288	4	292	1.39
11	132	2	134	1.52
12	163	4	167	2.45
13	86	2	88	2.33
14	123	1	124	0.81
15	101	1	102	0.99
16	78	3	81	3.85
17	56	0	56	0.00
18	59	0	59	0.00
19	29	0	29	0.00
20	46	0	46	0.00
21	32	0	32	0.00
22	26	4	30	15.38
23	18	1	19	5.56
24	21	0	21	0.00
25	21	1	22	4.76
26	20	2	22	10.00
27	17	0	17	0.00
28	16	1	17	6.25
29	7	0	7	0.00
30	25	0	25	0.00
31	176	1	177	0.57
จำนวนทั้งหมด	**211,385**	**3,564**	**214,949**	**1.69**

ภาพประกอบ 37 ข้อมูลเปรียบเทียบระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับจำนวนวันรักษาของผู้บาดเจ็บและค่าอัตราการเสียชีวิตในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558

จากภาพประกอบ 38 และ 39 ได้ทำการวิเคราะห์ข้อมูลความสัมพันธ์ระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับจังหวัดที่เกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ของวันที่ 27 ธันวาคม ถึง วันที่ 5 มกราคม ของปี 2551 ถึง ปี 2558 จะเห็นได้ว่าแนวโน้มของจำนวนอุบัติเหตุที่เกิดขึ้นสูงสุดอยู่ที่จังหวัดนครราชสีมา และรองมาเป็นจังหวัดเชียงใหม่

โดยที่ผู้ประสบอุบัติเหตุในจังหวัดนครราชสีมา มีจำนวนของอุบัติเหตุทั้งหมด 9,835 ครั้ง แบ่งออกเป็นไม่เสียชีวิต 9,712 ราย และเสียชีวิต 123 ราย และผู้ประสบอุบัติเหตุในจังหวัดเชียงใหม่ มีจำนวนของอุบัติเหตุทั้งหมด 7,930 ครั้ง แบ่งออกเป็นไม่เสียชีวิต 7,810 ราย และเสียชีวิต 120 ราย



ภาพประกอบ 38 ข้อมูลกราฟแท่งแสดงความสัมพันธ์ระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับจังหวัดที่เกิดอุบัติเหตุมากสุด 10 ลำดับแรก ในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558

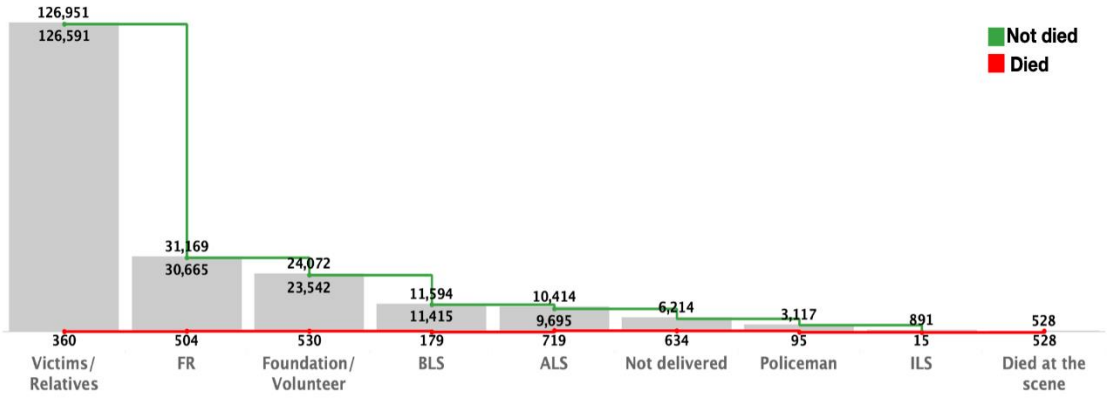
จังหวัด	ไม่เสียชีวิต	เสียชีวิต	จำนวนทั้งหมด	Mortality Rate (%)
เชียงราย	4,936	89	5,025	1.80
บุรีรัมย์	5,996	101	6,097	1.68
สุราษฎร์ธานี	5,267	87	5,354	1.65
เชียงใหม่	7,810	120	7,930	1.54
ขอนแก่น	6,676	103	6,779	1.54
สงขลา	5,320	77	5,397	1.45
นครราชสีมา	9,712	123	9,835	1.27
นครศรีธรรมราช	4,588	55	4,643	1.20
กรุงเทพมหานคร	5,853	68	5,921	1.16
ชลบุรี	7,151	70	7,221	0.98
จำนวนทั้งหมด	**63,309**	**893**	**64,202**	**1.41**

ภาพประกอบ 39 ข้อมูลเปรียบเทียบระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับจังหวัดที่เกิดอุบัติเหตุมากสุด 10 ลำดับแรกและค่าอัตราการเสียชีวิตในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558

จากภาพประกอบ 40 ได้ทำการวิเคราะห์ข้อมูลความสัมพันธ์ระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับการนำส่งผู้ประสบอุบัติเหตุในช่วงเทศกาลปีใหม่ของวันที่ 27 ธันวาคม ถึง วันที่ 5 มกราคม ของปี 2551 ถึง ปี 2558 จะเห็นได้ว่าแนวโน้มของจำนวนการนำส่งผู้ประสบอุบัติเหตุเข้าการรักษาที่โรงพยาบาล โดยผู้ประสบอุบัติเหตุเองและญาตินำส่งมีจำนวนสูงสุด และรองมาเป็นการนำส่งผู้บาดเจ็บด้วยหน่วยงานของชุดปฏิบัติการฉุกเฉินเบื้องต้น (FR) และมูลนิธิอาสาสมัคร

โดยที่การนำส่งผู้ประสบอุบัติเหตุเข้าการรักษาที่โรงพยาบาล โดยผู้ประสบอุบัติเหตุเองและญาตินำส่งมีจำนวนทั้งหมด 126,951 ราย โดยแบ่งออกเป็นไม่เสียชีวิต 126,591 ราย และเสียชีวิต 360 ราย และการนำส่งผู้บาดเจ็บของหน่วยงานของชุดปฏิบัติการฉุกเฉินเบื้องต้น (FR) มีจำนวนทั้งหมด 31,169 ราย โดยแบ่งออกเป็นไม่เสียชีวิต 30,665 ราย และเสียชีวิต 504 ราย และการนำส่งผู้บาดเจ็บของมูลนิธิอาสาสมัครมีจำนวนทั้งหมด 24,072 ราย โดยแบ่งออกเป็นไม่เสียชีวิต 23,542 ราย และเสียชีวิต 530 ราย

จะเห็นได้ว่าการนำส่งผู้ประสบอุบัติเหตุเข้าการรักษาที่โรงพยาบาล โดยผู้ประสบอุบัติเหตุเองและญาติมีจำนวนสูงกว่าการนำส่งด้วยหน่วยงานของชุดปฏิบัติการฉุกเฉินเบื้องต้น (FR) ถึง 4 เท่า และสูงกว่าการนำส่งผู้บาดเจ็บของมูลนิธิอาสาสมัครถึง 5 เท่า แต่พบว่าจำนวนของผู้เสียชีวิตสูงสุดอยู่ที่การนำส่งของหน่วยงานชุดปฏิบัติการฉุกเฉินระดับสูง (ALS) ซึ่งสูงเป็น 2 เท่า ของการนำส่งโดยผู้ประสบอุบัติเหตุเองและญาติ



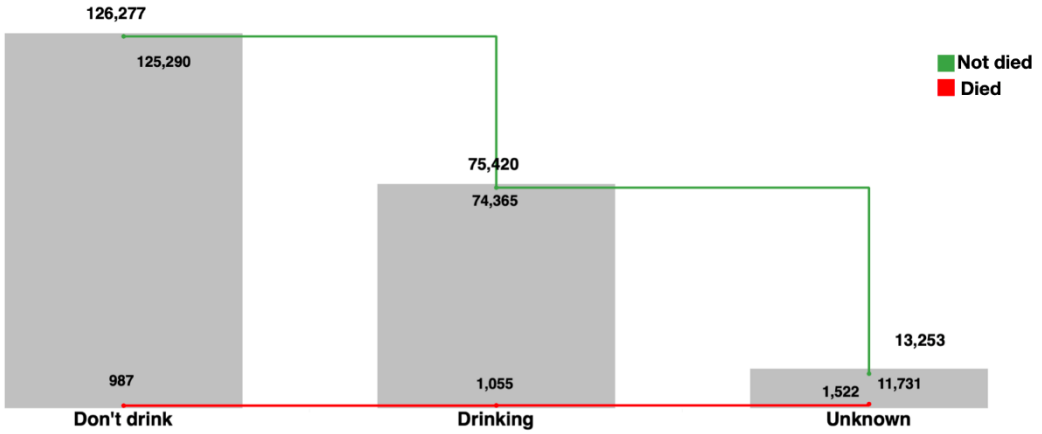
Delivering method	Not died	Died	Total	Mortality rate (%)
Not delivered	5,580	634	6,214	11.36
ALS	9,695	719	10,414	7.42
Policeman	3,022	95	3,117	3.14
Foundation/Volunteer	23,542	530	24,072	2.25
ILS	876	15	891	1.71
FR	30,665	504	31,169	1.64
BLS	11,415	179	11,594	1.57
Victims/Relatives	126,590	360	126,950	0.28
Died at the scene	0	528	528	
Total	**211,385**	**3,564**	**214,949**	**1.69**

ภาพประกอบ 40 ข้อมูลกราฟแท่งแสดงความสัมพันธ์ระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับการนำส่งผู้ประสบอุบัติเหตุในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558

จากภาพประกอบ 41 ได้ทำการวิเคราะห์ข้อมูลความสัมพันธ์ระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับการดื่มสุราของผู้ประสบอุบัติเหตุในช่วงเทศกาลปีใหม่ของวันที่ 27 ธันวาคม ถึง วันที่ 5 มกราคม ของปี 2551 ถึง ปี 2558 ซึ่งแนวโน้มของจำนวนการเกิดอุบัติเหตุสูงสุดเกิดจากไม่ดื่มสุราและรองลงมาเกิดจากการดื่มสุราของผู้ประสบอุบัติเหตุ

โดยที่การเกิดอุบัติเหตุจากการไม่ดื่มสุรามีจำนวนทั้งหมด 126,277 ราย โดยแบ่งออกเป็นไม่เสียชีวิต 125,290 ราย และเสียชีวิต 987 ราย และอุบัติเหตุจากการดื่มสุรามีจำนวนทั้งหมด 75,420 ราย โดยแบ่งออกเป็นไม่เสียชีวิต 74,365 ราย และเสียชีวิต 1,055 ราย

จะเห็นได้ว่าการอุบัติเหตุจากการไม่ดื่มสุรามีจำนวนสูงกว่าอุบัติเหตุจากการดื่มสุราถึง 1.6 เท่า ถึงแม้ว่าอุบัติเหตุของการไม่ดื่มสุราจะสูงกว่าการการดื่มสุรา แต่อัตราของการเสียชีวิตของผู้ประสบอุบัติเหตุที่ดื่มสุรากลับมีจำนวนสูงกว่าการไม่ดื่มสุรา ทำให้เห็นได้ว่าการดื่มสุรานั้นมีผลต่อการเกิดอุบัติเหตุที่ร้ายแรงกว่าอุบัติเหตุที่ผู้ประสบอุบัติเหตุไม่ดื่มสุรา



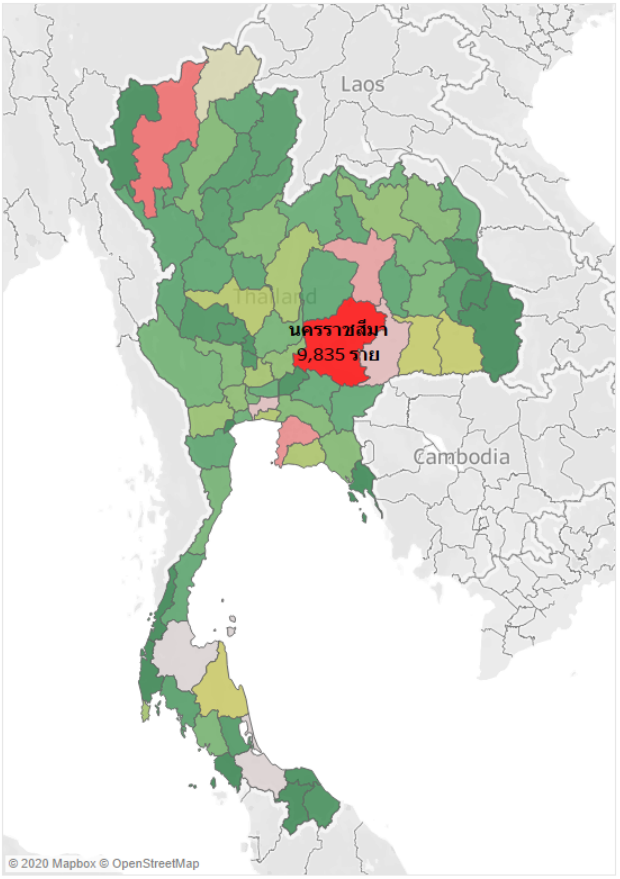
Alcohol drinking	Not died	Died	Total	Mortality rate (%)
Unknown	11,730	1,522	13,252	12.98
Drinking	74,365	1,055	75,420	1.42
Don't drink	125,290	987	126,277	0.79
Total	**211,385**	**3,564**	**214,949**	**1.69**

ภาพประกอบ 41 ข้อมูลกราฟแท่งแสดงความสัมพันธ์ระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับการดื่มสุราของผู้ประสบอุบัติเหตุในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558

จากภาพประกอบ 42 ได้ทำการวิเคราะห์ข้อมูลความสัมพันธ์ระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับจังหวัดที่เกิดอุบัติเหตุในประเทศไทยในช่วงเทศกาลปีใหม่ของวันที่ 27 ธันวาคม ถึง วันที่ 5 มกราคม ของปี 2551 ถึง ปี 2558 จะเห็นได้ว่าแนวโน้มของการเกิดอุบัติเหตุสูงสุดเกิดที่จังหวัดนครราชสีมาและรองลงมาเป็นจังหวัดเชียงใหม่

โดยที่ผู้บาดเจ็บจากการประสบอุบัติเหตุที่จังหวัดนครราชสีมา มีจำนวนของทั้งหมด 9,835 ราย แบ่งออกเป็นไม่เสียชีวิต 9,712 ราย และเสียชีวิต 123 ราย และผู้บาดเจ็บจากการประสบอุบัติเหตุที่จังหวัดเชียงใหม่ มีจำนวนของทั้งหมด 7,930 ราย แบ่งออกเป็นไม่เสียชีวิต 7,810 ราย และเสียชีวิต 120 ราย

จะเห็นได้ว่าข้อมูลสามารถบอกถึงสถิติผู้บาดเจ็บจากอุบัติเหตุในแต่ละจังหวัด และเป็นประโยชน์ต่อการวางแผนมาตรการในการเตรียมการป้องอุบัติเหตุ



535 9,835

อุบัติเหตุที่เกิดขึ้นในแต่ละจังหวัดในประเทศไทย

	ไม่เสียชีวิต	เสียชีวิต	ทั้งหมด
นครราชสีมา	9,712	123	**9,835**
เชียงใหม่	7,810	120	**7,930**
ชลบุรี	7,151	70	**7,221**
ขอนแก่น	6,676	103	**6,779**
บุรีรัมย์	5,996	101	**6,097**
กรุงเทพมหานคร	5,853	68	**5,921**
สงขลา	5,320	77	**5,397**
สุราษฎร์ธานี	5,267	87	**5,354**
เชียงราย	4,936	89	**5,025**
นครศรีธรรมราช	4,588	55	**4,643**
ศรีสะเกษ	4,459	54	**4,513**
สุรินทร์	4,432	77	**4,509**
นครสวรรค์	4,092	103	**4,195**
สมุทรปราการ	4,041	51	**4,092**
ระยอง	3,942	80	**4,022**
เพชรบูรณ์	3,860	83	**3,943**
ภูเก็ต	3,831	46	**3,877**
พระนครศรีอยุธยา	3,775	71	**3,846**
ราชบุรี	3,501	64	**3,565**
อุดรธานี	3,389	58	**3,447**
สระบุรี	3,324	51	**3,375**
นครปฐม	3,265	94	**3,359**
สกลนคร	3,143	46	**3,189**
พิษณุโลก	3,105	66	**3,171**
จันทบุรี	3,084	49	**3,133**
ฉะเชิงเทรา	3,042	55	**3,097**
ลำปาง	3,034	59	**3,093**
สุพรรณบุรี	3,022	54	**3,076**

ภาพประกอบ 42 ข้อมูลกราฟแท่งแสดงความสัมพันธ์ระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับจังหวัดที่เกิดอุบัติเหตุในประเทศไทยมากสุด 30 ลำดับแรก ในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558

จากภาพประกอบ 43, 44 และ 45 ได้ทำการวิเคราะห์ข้อมูลความสัมพันธ์ระหว่างวันที่เกิดเหตุในช่วงเทศกาลปีใหม่กับเวลาเกิดเหตุของผู้ประสบอุบัติเหตุในช่วงเทศกาลปีใหม่ของวันที่ 27 ธันวาคม ถึง วันที่ 5 มกราคม ของปี 2551 ถึง ปี 2558 ซึ่งแนวโน้มของจำนวนการเกิดอุบัติเหตุสูงสุดเกิดในช่วงเวลา 18:01 น. ถึง 19:00 น. ในช่วงของวันที่ 30 ถึง 31 ธันวาคม

โดยที่ช่วงเวลา 18:01 น. ถึง 19:00 น. ของวันที่ 30 ธันวาคม มีจำนวนอุบัติเหตุทั้งหมด 3,005 ราย และช่วงเวลา 18:01 น. ถึง 19:00 น. ของวันที่ 31 ธันวาคม มีจำนวนอุบัติเหตุทั้งหมด 3,315 ราย

	วันที่เกิดเหตุในช่วงเทศกาลปีใหม่ 2551 ถึง 2558										
เวลาเกิดเหตุ	27	28	29	30	31	1	2	3	4	5	ทั้งหมด
01:01-02:00 น.	132	256	563	783	1,042	2,128	593	352	243	81	**6,173**
02:01-03:00 น.	73	239	478	616	826	1,574	448	279	197	35	**4,765**
03:01-04:00 น.	67	149	332	458	627	1,189	305	211	127	34	**3,499**
04:01-05:00 น.	68	112	297	348	468	867	261	150	115	28	**2,714**
05:01-06:00 น.	50	145	286	336	368	668	249	182	151	91	**2,526**
06:01-07:00 น.	84	162	331	449	471	701	317	225	262	91	**3,093**
07:01-08:00 น.	261	383	717	1,015	747	819	686	506	497	211	**5,842**
08:01-09:00 น.	321	496	868	1,246	1,066	1,116	863	732	618	329	**7,655**
09:01-10:00 น.	217	377	788	1,153	1,336	1,245	1,009	752	488	197	**7,562**
10:01-11:00 น.	229	355	969	1,395	1,542	1,561	1,124	767	540	204	**8,686**
11:01-12:00 น.	235	428	948	1,483	1,664	1,623	1,232	811	599	189	**9,212**
12:01-13:00 น.	279	530	1,000	1,672	1,951	1,808	1,271	808	561	146	**10,026**
13:01-14:00 น.	239	513	1,016	1,609	2,026	1,842	1,258	841	556	166	**10,066**
14:01-15:00 น.	296	597	1,064	1,742	2,218	1,843	1,330	884	592	152	**10,718**
15:01-16:00 น.	368	658	1,257	1,981	2,355	2,181	1,523	1,061	754	230	**12,368**
16:01-17:00 น.	476	867	1,683	2,627	2,905	2,348	1,970	1,183	909	318	**15,286**
17:01-18:00 น.	568	853	1,808	2,909	3,145	2,532	1,986	1,307	1,006	376	**16,490**
18:01-19:00 น.	543	983	1,881	3,005	3,315	2,457	2,056	1,278	949	349	**16,816**
19:01-20:00 น.	450	838	1,662	2,530	2,938	2,119	1,658	1,016	761	304	**14,276**
20:01-21:00 น.	376	691	1,350	2,104	2,728	1,762	1,264	751	627	219	**11,872**
21:01-22:00 น.	315	575	1,165	1,779	2,441	1,352	993	609	427	162	**9,818**
22:01-23:00 น.	236	477	927	1,491	2,187	1,202	795	562	384	84	**8,345**
23:01-24:00 น.	169	331	634	1,125	1,761	1,273	574	423	263	82	**6,635**
24:01-01:00 น.	166	435	771	1,152	1,495	2,890	883	502	330	110	**8,734**
ไม่ทราบ	30	100	238	316	318	344	209	109	82	27	**1,773**
ทั้งหมด	**6,248**	**11,550**	**23,033**	**35,324**	**41,940**	**39,444**	**24,857**	**16,301**	**12,038**	**4,215**	**214,950**

ภาพประกอบ 43 แสดงความสัมพันธ์ระหว่างวันที่เกิดเหตุกับเวลาเกิดเหตุของผู้ประสบอุบัติเหตุในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558

ถ้าแบ่งตามผู้เสียชีวิตจากอุบัติเหตุมีจำนวนมากสุด วันที่ 31 ธันวาคม มีจำนวนผู้เสียชีวิตจำนวนทั้งหมด 747 ราย ในช่วงเวลา 18:01 ถึง 19:00 น. มีจำนวนผู้เสียชีวิตมากสุดในจำนวนทั้งหมด 52 ราย และในวันที่ 1 มกราคม มีจำนวนผู้เสียชีวิตจำนวนทั้งหมด 701 ราย ในช่วงเวลา 24:01 ถึง 01:00 น. มีจำนวนผู้เสียชีวิตมากสุดในจำนวนทั้งหมด 65 ราย และจำนวนผู้เสียชีวิตมากสุดเกิดขึ้นในช่วงเวลา 18:01 น. ถึง 19:00 น. ของวันที่ 30 มีจำนวนผู้เสียชีวิตทั้งหมด 67 ราย จะเห็นได้ว่าในช่วงเวลา 18:01 น. ถึง 19:00 น. ของวันที่ 30 ธันวาคม ถึง 1 มกราคม เป็นช่วงเวลาอันตรายที่มีจำนวนอุบัติเหตุและผู้เสียชีวิตมากที่สุด

	เสียชีวิต										
เวลาเกิดเหตุ	27	28	29	30	31	1	2	3	4	5	ทั้งหมด
01:01-02:00 น.	2	5	18	21	31	39	12	10	8		**146**
02:01-03:00 น.		8	17	13	22	27	13	7	2		**109**
03:01-04:00 น.	1	4	13	17	25	21	10	10	3	2	**106**
04:01-05:00 น.	3	5	17	14	16	35	8	5	3	1	**107**
05:01-06:00 น.	2	9	11	13	21	20	3	5	6	1	**91**
06:01-07:00 น.	1		11	8	13	16	4	6	17	2	**78**
07:01-08:00 น.	1	3	13	22	15	23	12	7	8	1	**105**
08:01-09:00 น.	4	2	10	20	28	15	12	4	7	2	**104**
09:01-10:00 น.	2	8	13	13	9	31	17	6	6	2	**107**
10:01-11:00 น.	1	1	16	15	22	19	16	8	4	3	**105**
11:01-12:00 น.	2	4	22	17	25	22	13	6	5	1	**117**
12:01-13:00 น.	3	9	6	29	20	20	18	4	3	2	**114**
13:01-14:00 น.	1	4	20	20	29	24	16	12	5	1	**132**
14:01-15:00 น.	6	6	12	29	31	28	16	12	8	3	**151**
15:01-16:00 น.	5	8	16	26	33	45	20	14	2	5	**174**
16:01-17:00 น.	6	9	22	36	42	27	16	11	10	2	**181**
17:01-18:00 น.	11	3	35	30	50	34	16	9	10	4	**202**
18:01-19:00 น.	13	14	37	67	52	41	26	16	8	2	**276**
19:01-20:00 น.	4	22	37	40	40	40	26	15	10	5	**239**
20:01-21:00 น.	6	10	30	52	51	36	18	12	7	4	**226**
21:01-22:00 น.	4	12	20	38	51	15	9	13	7	3	**172**
22:01-23:00 น.	10	15	13	25	46	22	12	9	9	2	**163**
23:01-24:00 น.	4	7	14	17	31	23	8	8	7	1	**120**
24:01-01:00 น.	1	11	11	28	36	65	27	10	8	2	**199**
ไม่ทราบ		2	5	2	8	13	6	2	2		**40**
ทั้งหมด	**93**	**181**	**439**	**612**	**747**	**701**	**354**	**221**	**165**	**51**	**3,564**

ภาพประกอบ 44 แสดงความสัมพันธ์ระหว่างวันที่เกิดเหตุกับเวลาเกิดเหตุของผู้ประสบอุบัติเหตุที่เสียชีวิตในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558

ถ้าแบ่งตามผู้ไม่เสียชีวิตจากอุบัติเหตุมีจำนวนมากสุด วันที่ 31 ธันวาคม มีจำนวนทั้งหมด 41,193 ราย ในช่วงเวลา 18:01 ถึง 19:00 น. มีจำนวนทั้งหมด 3,263 ราย และในวันที่ 1 มกราคม มีจำนวนทั้งหมด 38,743 ราย ในช่วงเวลา 24:01 ถึง 01:00 น. มีจำนวนทั้งหมด 2,825 ราย จะเห็นได้ว่าในช่วงเวลา 18:01 น. ถึง 19:00 น. ของวันที่ 30 ธันวาคม ถึง 1 มกราคม เป็นช่วงเวลาอันตรายที่มีจำนวนอุบัติเหตุมากที่สุด

	ไม่เสียชีวิต										
เวลาเกิดเหตุ	27	28	29	30	31	1	2	3	4	5	ทั้งหมด
01:01-02:00 น.	130	251	545	762	1,011	2,089	581	342	235	81	**6,027**
02:01-03:00 น.	73	231	461	603	804	1,547	435	272	195	35	**4,656**
03:01-04:00 น.	66	145	319	441	602	1,168	295	201	124	32	**3,393**
04:01-05:00 น.	65	107	280	334	452	832	253	145	112	27	**2,607**
05:01-06:00 น.	48	136	275	323	347	648	246	177	145	90	**2,435**
06:01-07:00 น.	83	162	320	441	458	685	313	219	245	89	**3,015**
07:01-08:00 น.	260	380	704	993	732	796	674	499	489	210	**5,737**
08:01-09:00 น.	317	494	858	1,226	1,038	1,101	851	728	611	327	**7,551**
09:01-10:00 น.	215	369	775	1,140	1,327	1,214	992	746	482	195	**7,455**
10:01-11:00 น.	228	354	953	1,380	1,520	1,542	1,108	759	536	201	**8,581**
11:01-12:00 น.	233	424	926	1,466	1,639	1,601	1,219	805	594	188	**9,095**
12:01-13:00 น.	276	521	994	1,643	1,931	1,788	1,253	804	558	144	**9,912**
13:01-14:00 น.	238	509	996	1,589	1,997	1,818	1,242	829	551	165	**9,934**
14:01-15:00 น.	290	591	1,052	1,713	2,187	1,815	1,314	872	584	149	**10,567**
15:01-16:00 น.	363	650	1,241	1,955	2,322	2,136	1,503	1,047	752	225	**12,194**
16:01-17:00 น.	470	858	1,661	2,591	2,863	2,321	1,954	1,172	899	316	**15,105**
17:01-18:00 น.	557	850	1,773	2,879	3,095	2,498	1,970	1,298	996	372	**16,288**
18:01-19:00 น.	530	969	1,844	2,938	3,263	2,416	2,030	1,262	941	347	**16,540**
19:01-20:00 น.	446	816	1,625	2,490	2,898	2,079	1,632	1,001	751	299	**14,037**
20:01-21:00 น.	370	681	1,320	2,052	2,677	1,726	1,246	739	620	215	**11,646**
21:01-22:00 น.	311	563	1,145	1,741	2,390	1,337	984	596	420	159	**9,646**
22:01-23:00 น.	226	462	914	1,466	2,141	1,180	783	553	375	82	**8,182**
23:01-24:00 น.	165	324	620	1,108	1,730	1,250	566	415	256	81	**6,515**
24:01-01:00 น.	165	424	760	1,124	1,459	2,825	856	492	322	108	**8,535**
ไม่ทราบ	30	98	233	314	310	331	203	107	80	27	**1,733**
ทั้งหมด	**6,155**	**11,369**	**22,594**	**34,712**	**41,193**	**38,743**	**24,503**	**16,080**	**11,873**	**4,164**	**211,386**

ภาพประกอบ 45 แสดงความสัมพันธ์ระหว่างวันที่เกิดเหตุกับเวลาเกิดเหตุของผู้ประสบอุบัติเหตุที่ไม่เสียชีวิตในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558

จากภาพประกอบ 46, 47 และ 48 ได้ทำการวิเคราะห์ข้อมูลความสัมพันธ์ระหว่างชื่อเทศกาลปีใหม่กับวันที่เกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ของวันที่ 27 ธันวาคม ถึง วันที่ 5 มกราคม ของปี 2551 ถึง ปี 2558 ซึ่งแนวโน้มของจำนวนการเกิดอุบัติเหตุสูงสุดเกิดในปี 2551 รองลงมาเป็นปี 2553

โดยที่ปี 2551 มีจำนวนอุบัติเหตุทั้งหมด 30,866 ราย ในช่วงของวันที่ 31 ธันวาคม มีจำนวนอุบัติเหตุเกิดขึ้นทั้งหมด 5,844 ราย โดยแบ่งเป็นผู้บาดเจ็บเสียชีวิตทั้งหมด 101 ราย และผู้บาดเจ็บไม่เสียชีวิตทั้งหมด 5,743 ราย และ

ในปี 2553 มีจำนวนอุบัติเหตุทั้งหมด 27,936 ราย ในช่วงของวันที่ 31 ธันวาคม มีจำนวนอุบัติเหตุเกิดขึ้นทั้งหมด 5,485 ราย โดยแบ่งเป็นผู้บาดเจ็บเสียชีวิตทั้งหมด 98 ราย และผู้บาดเจ็บไม่เสียชีวิตทั้งหมด 5,387 ราย และในวันที่ 1 มกราคม มีจำนวนอุบัติเหตุเกิดขึ้นทั้งหมด 5,332 ราย โดยแบ่งเป็นผู้บาดเจ็บเสียชีวิตทั้งหมด 96 ราย และผู้บาดเจ็บไม่เสียชีวิตทั้งหมด 5,236 ราย

	ชื่อเทศกาล								
วันที่เกิดอุบัติเหตุ	ปีใหม่ 51	ปีใหม่ 52	ปีใหม่ 53	ปีใหม่ 54	ปีใหม่ 55	ปีใหม่ 56	ปีใหม่ 57	ปีใหม่ 58	ทั้งหมด
27						2,925	3,322		**6,247**
28	4,119					3,629	3,802		**11,550**
29	4,871		3,559	3,158	3,463	4,168	3,814		**23,033**
30	5,279	4,494	4,729	4,508	4,388	4,121	3,721	4,084	**35,324**
31	5,844	5,896	5,485	5,352	5,233	4,582	4,483	5,065	**41,940**
1	5,280	5,682	5,332	4,900	4,828	4,525	4,179	4,719	**39,445**
2	3,113	3,467	3,548	3,221	3,191	2,699	2,649	2,969	**24,857**
3	2,360	2,806	2,972	2,708	2,795			2,660	**16,301**
4		2,787	2,311	2,151	2,233			2,556	**12,038**
5		2,151						2,064	**4,215**
ทั้งหมด	**30,866**	**27,283**	**27,936**	**25,998**	**26,131**	**26,649**	**25,970**	**24,117**	**214,950**

ภาพประกอบ 46 แสดงความสัมพันธ์ระหว่างชื่อเทศกาลปีใหม่กับวันที่เกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558

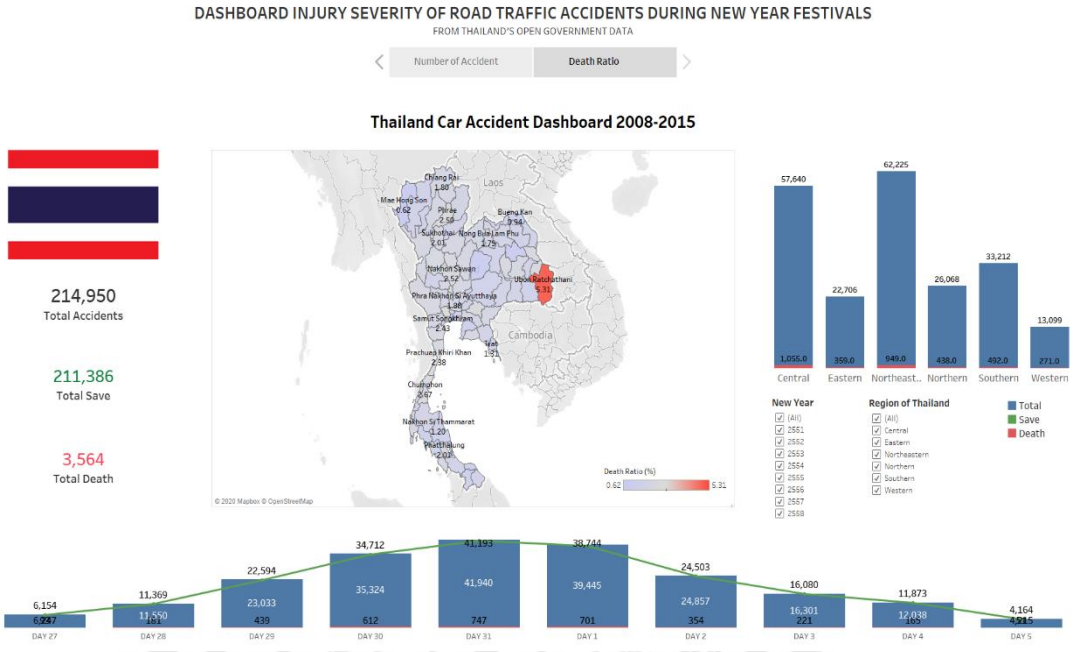
	เสียชีวิต								ทั้งหมด
วันที่เกิดอุบัติเหตุ	ปีใหม่ 51	ปีใหม่ 52	ปีใหม่ 53	ปีใหม่ 54	ปีใหม่ 55	ปีใหม่ 56	ปีใหม่ 57	ปีใหม่ 58	
27						42	51		**93**
28	63					52	66		**181**
29	109		58	43	65	91	73		**439**
30	102	77	80	80	80	65	58	70	**612**
31	101	128	98	98	95	63	71	93	**747**
1	91	91	96	100	97	83	73	70	**701**
2	28	68	50	55	46	38	32	37	**354**
3	31	36	32	49	38			35	**221**
4		36	32	31	33			33	**165**
5		21						30	**51**
ทั้งหมด	**525**	**457**	**446**	**456**	**454**	**434**	**424**	**368**	**3,564**

ภาพประกอบ 47 แสดงความสัมพันธ์ระหว่างชื่อเทศกาลปีใหม่กับวันที่เกิดอุบัติเหตุของผู้เสียชีวิตในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558

	ไม่เสียชีวิต								ทั้งหมด
วันที่เกิดอุบัติเหตุ	ปีใหม่ 51	ปีใหม่ 52	ปีใหม่ 53	ปีใหม่ 54	ปีใหม่ 55	ปีใหม่ 56	ปีใหม่ 57	ปีใหม่ 58	
27						2,883	3,271		**6,154**
28	4,056					3,577	3,736		**11,369**
29	4,762		3,501	3,115	3,398	4,077	3,741		**22,594**
30	5,177	4,417	4,649	4,428	4,308	4,056	3,663	4,014	**34,712**
31	5,743	5,768	5,387	5,254	5,138	4,519	4,412	4,972	**41,193**
1	5,189	5,591	5,236	4,800	4,731	4,442	4,106	4,649	**38,744**
2	3,085	3,399	3,498	3,166	3,145	2,661	2,617	2,932	**24,503**
3	2,329	2,770	2,940	2,659	2,757			2,625	**16,080**
4		2,751	2,279	2,120	2,200			2,523	**11,873**
5		2,130						2,034	**4,164**
ทั้งหมด	**30,341**	**26,826**	**27,490**	**25,542**	**25,677**	**26,215**	**25,546**	**23,749**	**211,386**

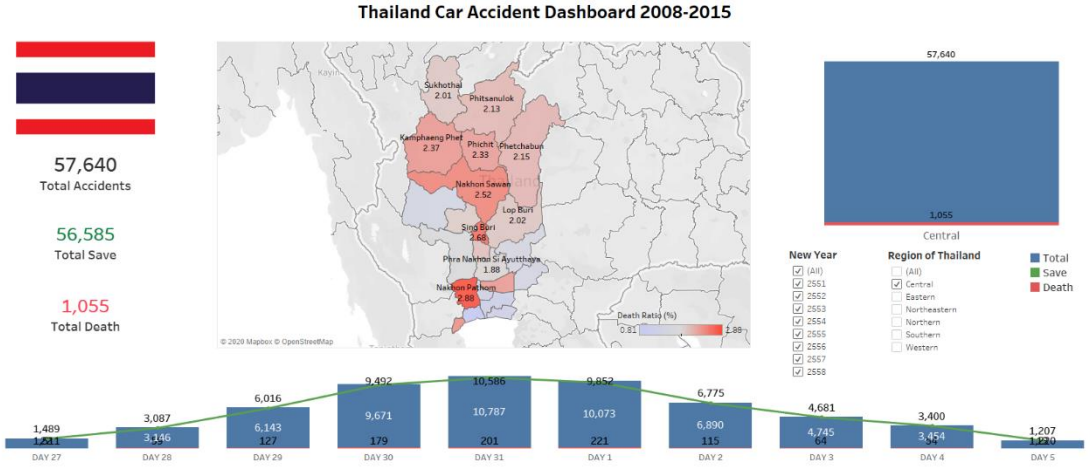
ภาพประกอบ 48 แสดงความสัมพันธ์ระหว่างชื่อเทศกาลปีใหม่กับวันที่เกิดอุบัติเหตุของผู้ไม่เสียชีวิตในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558

จากภาพประกอบ 49, 50, 51, 52, 53, 54 และ 55 ได้ทำการวิเคราะห์ข้อมูลความสัมพันธ์ระหว่างค่าอัตราการเสียชีวิตกับจังหวัดในประเทศไทย ที่เกิดอุบัติเหตุในช่วงเทศกาลปีใหม่ของวันที่ 27 ธันวาคม ถึง วันที่ 5 มกราคม ของปี 2551 ถึง ปี 2558 ซึ่งการเกิดอุบัติเหตุที่มีความรุนแรงสูงสุด เกิดขึ้นในจังหวัดอุบลราชธานีที่มีค่าอัตราการเสียชีวิต เท่ากับ 5.31% ซึ่งเป็นผู้เสียชีวิตจากอุบัติเหตุจำนวน 44 ราย จากอุบัติเหตุทั้งหมด 872 ครั้ง



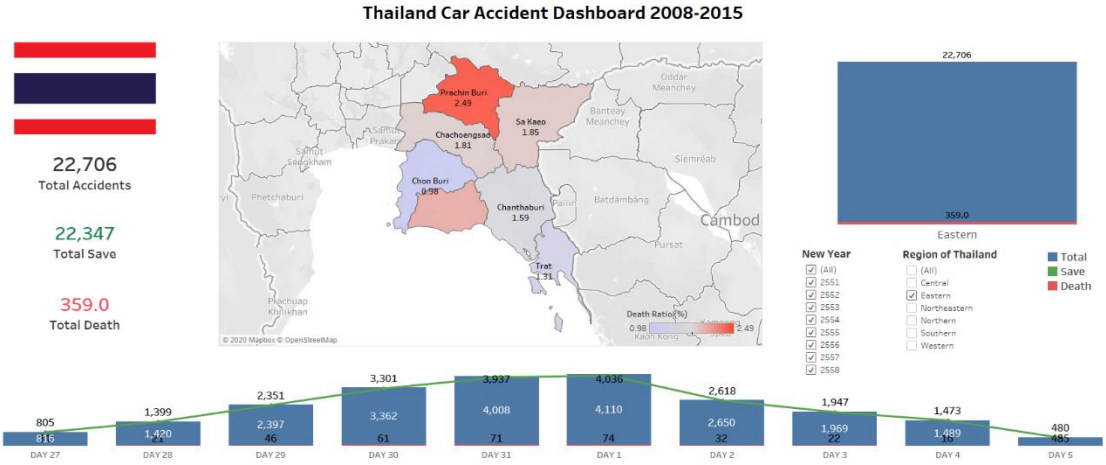
ภาพประกอบ 49 แสดงความรุนแรงของค่าอัตราการเสียชีวิตของการเกิดอุบัติเหตุทั้งประเทศไทยในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558

ถ้าแบ่งตามในภาคของประเทศไทย โดยพบว่าในภาคกลางเกิดอุบัติเหตุที่มีความรุนแรงสูงสุดเกิดขึ้นในจังหวัดนครปฐม ที่มีค่าอัตราการเสียชีวิต เท่ากับ 2.88% ซึ่งเป็นผู้เสียชีวิตจากอุบัติเหตุจำนวน 94 ราย จากอุบัติเหตุทั้งหมด 3,359 ครั้ง



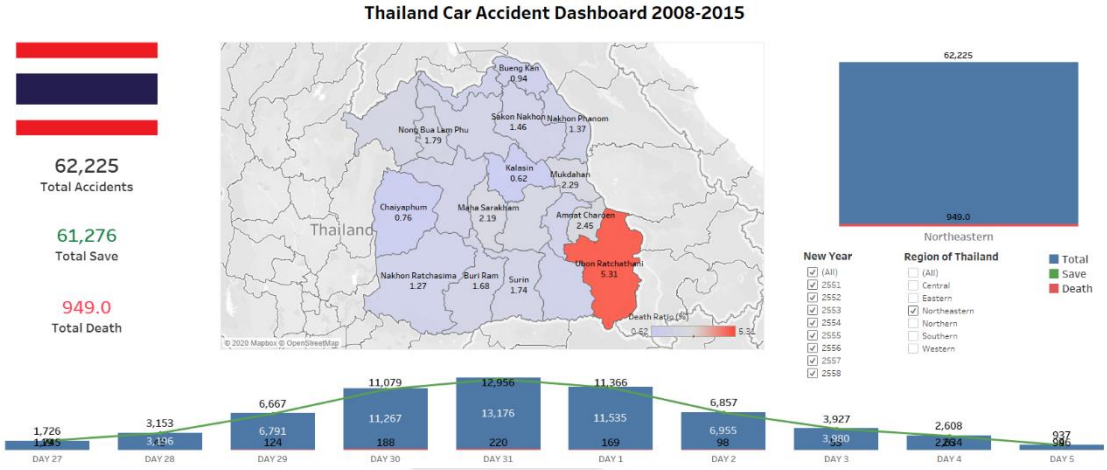
ภาพประกอบ 50 แสดงความรุนแรงของค่าอัตราการเสียชีวิตของการเกิดอุบัติเหตุในภาคกลางของประเทศไทยในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558

ในภาคตะวันออกเกิดอุบัติเหตุที่มีความรุนแรงสูงสุดเกิดขึ้นในจังหวัดปราจีนบุรี ที่มีค่าอัตราการเสียเท่ากับ 2.49% ซึ่งเป็นผู้เสียชีวิตจากอุบัติเหตุจำนวน 55 ราย จากอุบัติเหตุทั้งหมด 2,261 ครั้ง



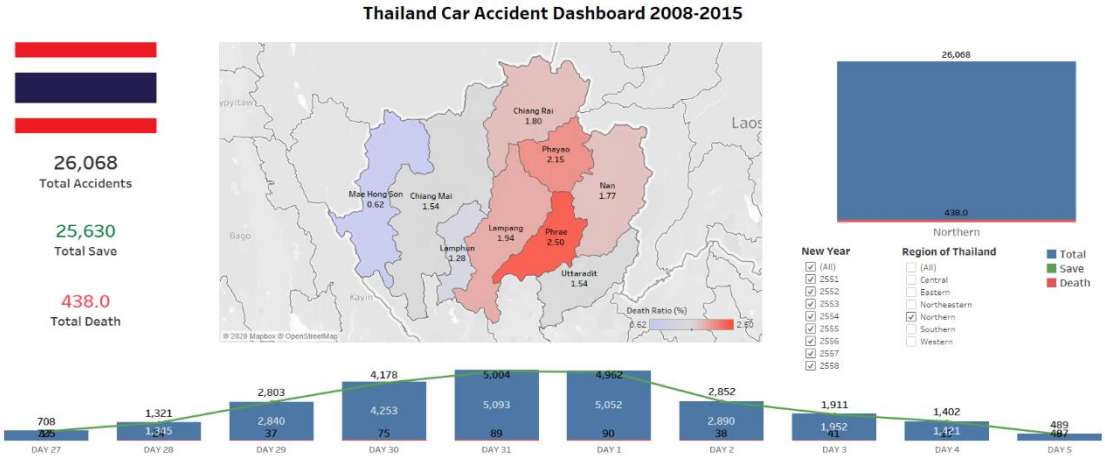
ภาพประกอบ 51 แสดงความรุนแรงของค่าอัตราการเสียชีวิตของการเกิดอุบัติเหตุในภาคตะวันออกของประเทศไทยในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558

ในภาคตะวันออกเฉียงเหนือเกิดอุบัติเหตุที่มีความรุนแรงสูงสุดเกิดขึ้นในจังหวัดอุบลราชธานี ที่มีค่าอัตราการเสียชีวิตเท่ากับ 5.31% ซึ่งเป็นผู้เสียชีวิตจากอุบัติเหตุจำนวน 44 ราย จากอุบัติเหตุทั้งหมด 872 ครั้ง



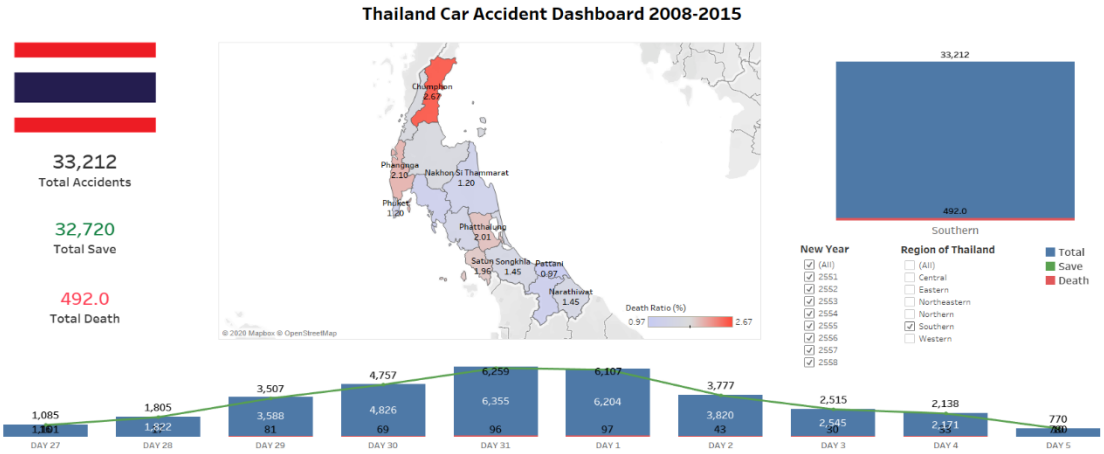
ภาพประกอบ 52 แสดงความรุนแรงของค่าอัตราการเสียชีวิตของการเกิดอุบัติเหตุในภาคตะวันออกเฉียงเหนือของประเทศไทยในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558

ในภาคเหนือเกิดอุบัติเหตุที่มีความรุนแรงสูงสุดเกิดขึ้นในจังหวัดปาย ที่มีค่าอัตราการเสียชีวิตเท่ากับ 2.50% ซึ่งเป็นผู้เสียชีวิตจากอุบัติเหตุจำนวน 41 ราย จากอุบัติเหตุทั้งหมด 1,680 ครั้ง



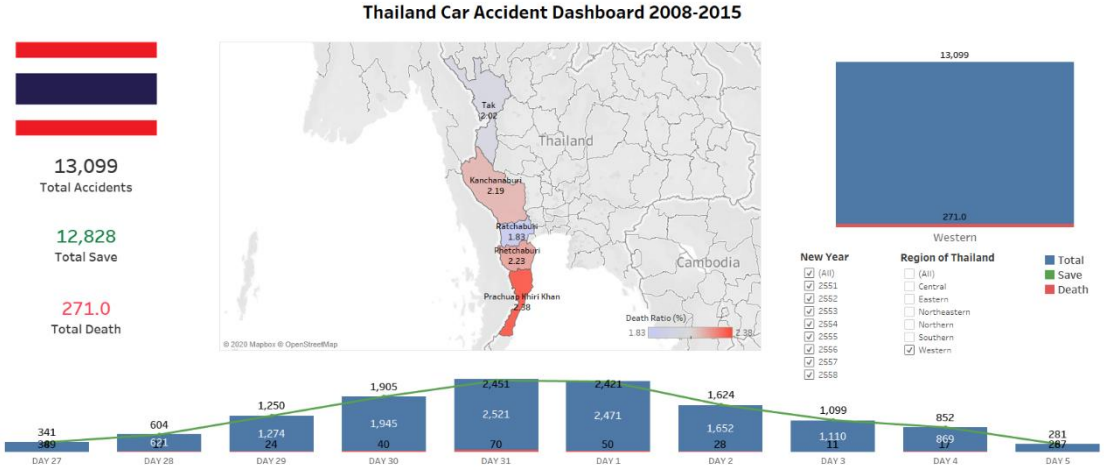
ภาพประกอบ 53 แสดงความรุนแรงของค่าอัตราการเสียชีวิตของการเกิดอุบัติเหตุในภาคเหนือของประเทศไทยในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558

ในภาคใต้เกิดอุบัติเหตุที่มีความรุนแรงสูงสุดเกิดขึ้นในจังหวัดชุมพร ที่มีค่าอัตราการเสียชีวิตเท่ากับ 2.67% ซึ่งเป็นผู้เสียชีวิตจากอุบัติเหตุจำนวน 55 ราย จากอุบัติเหตุทั้งหมด 2,118 ครั้ง



ภาพประกอบ 54 แสดงความรุนแรงของค่าอัตราการเสียชีวิต ของการเกิดอุบัติเหตุในภาคใต้ของประเทศไทยในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558

ในภาคตะวันตกเกิดอุบัติเหตุที่มีความรุนแรงสูงสุดเกิดขึ้นในจังหวัดประจวบคีรีขันธ์ ที่มีค่าอัตราการเสียชีวิตเท่ากับ 2.38% ซึ่งเป็นผู้เสียชีวิตจากอุบัติเหตุจำนวน 65 ราย จากอุบัติเหตุทั้งหมด 2,794 ครั้ง



ภาพประกอบ 55 แสดงความรุนแรงของค่าอัตราการเสียชีวิตของการเกิดอุบัติเหตุในภาคตะวันตกของประเทศไทยในช่วงเทศกาลปีใหม่ 2551 ถึง ปี 2558

จากการวิเคราะห์ข้อมูลความสัมพันธ์ระหว่างผลการรักษาของผู้ประสบอุบัติเหตุกับปัจจัยอื่นๆ ทั้งหมดในชุดข้อมูลผู้บาดเจ็บจากอุบัติเหตุในช่วงเทศกาลปีใหม่ 2551 ถึง 2558 ปัจจัยที่มีผลต่ออัตราการเสียชีวิตของผู้บาดเจ็บจากอุบัติเหตุสูงสุดจากค่าอัตราการเสียชีวิต สรุปได้ดังนี้

ตาราง 2 แสดงปัจจัยที่มีผลต่ออัตราการเสียชีวิตของผู้บาดเจ็บจากอุบัติเหตุสูงสุดในช่วงเทศกาลปีใหม่ 2551 ถึง 2558

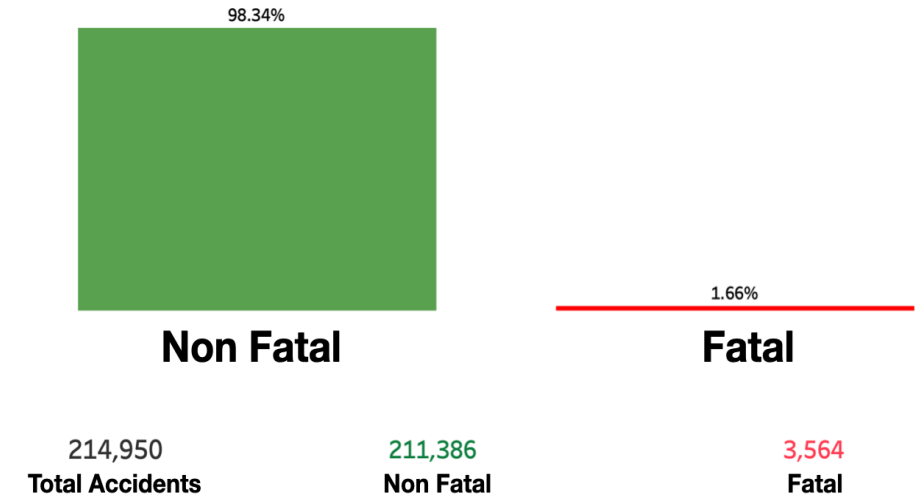
ลำดับ	ข้อมูลตัวแปร (Variable)	สิ่งที่อันตรายต่อการเสียชีวิตจากอุบัติเหตุ	ไม่เสียชีวิต	เสียชีวิต	รวมทั้งหมด	อัตราการเสียชีวิต (%)
1	ช่วงอายุ	อายุ 91-95	13	2	15	15.38
2	รถคู่กรณี	รถโดยสารใหญ่	722	84	806	11.63
3	การนำส่ง	ไม่นำส่ง	5,580	634	6,214	11.36
4	สถานะ	คนเดินเท้า	7095	315	7,410	4.44
5	รถผู้บาดเจ็บ	ไม่มี/ล้มเอง	6,211	263	6,474	4.23
6	เวลาเกิดเหตุ	04:01-05:00 น.	2,607	107	2,714	4.10
7	ถนนที่เกิดเหตุ	ทางหลวง	59,432	1,826	61,258	3.07
8	เวลาเกิดเหตุกับวันที่เกิดเหตุ	เวลา 18:01-19:00 น. ในวันที่ 30	2,938	67	3,005	2.28
9	เพศ	ชาย	142,258	2,887	145,145	2.03
10	วันที่เกิดเหตุ	วันที่ 29	22,594	439	23,033	1.94
11	จังหวัด	เชียงราย	4,936	89	5,025	1.80
12	ชื่อเทศกาล	ปีใหม่ 2554	25,542	456	25,998	1.79
13	มาตราการ	เข็มขัด	5,234	93	5,327	1.78
14	จำนวนวันรักษา	ไม่ถึง 24 ชั่วโมง	188,724	3,065	191,789	1.62
15	การดื่มสุรา	ดื่ม	74,365	1,055	75,420	1.42

3.4 การเตรียมข้อมูล

การเตรียมข้อมูลสำหรับการพัฒนาแบบจำลองการทำนายความรุนแรงของผู้บาดเจ็บจากอุบัติเหตุทางถนนในช่วงเทศกาลปีใหม่ 2551 ถึงปี 2558 ข้อมูลที่ใช้ในการสร้างแบบจำลองการทำนายมีทั้งหมด 214,950 แถว ประกอบด้วย 14 Column ได้แก่ วันที่เกิดเหตุ, อายุ, จำนวนวันรักษา, เวลาเกิดเหตุ, เพศ, ถนนที่เกิดเหตุ, สถานะ, รถผู้บาดเจ็บ, รถคู่กรณี, มาตรการ,การดื่มสุรา, การนำส่ง, Refer-Admit และผลการรักษา โดยที่ชื่อปีเทศกาลปีใหม่ ชื่อจังหวัด ชื่อโรงพยาบาล จะไม่ถูกนำมาใช้เป็นข้อมูลสำหรับฝึกสอนแบบจำลองการทำนาย

เนื่องจากในการวิจัยครั้งนี้เป็นการพัฒนาแบบจำลองการทำนายความรุนแรงของผู้บาดเจ็บจากอุบัติเหตุทางถนนในช่วงเทศกาลปีใหม่เป็นปัญหาแบบ Binary Classification มี 2 ประเภทความรุนแรงของผู้บาดเจ็บ คือ 0 และ 1 จึงได้จัดการทำการรวมกลุ่มข้อมูลใน Class ของผลการรักษาผู้บาดเจ็บ โดยที่ 0 เป็นอาการบาดเจ็บที่ไม่รุนแรงถึงชีวิต คือเป็นผลการรักษาผู้ประสบอุบัติเหตุในสถานะภาพและอาการของผู้บาดเจ็บเมื่อสิ้นสุดการรักษาเบื้องต้น ประกอบด้วย ไม่เสียชีวิตหรือทุเลา/หาย หมายถึง ผู้บาดเจ็บ มีสถานะภาพและอาการเมื่อสิ้นสุดการรักษาอาการทุเลาสามารถกลับบ้านได้ และ 1 เป็นอาการบาดเจ็บรุนแรงถึงการเสียชีวิต ประกอบด้วย ตาย ณ จุดเกิดเหตุ, ตายในระหว่างนำส่ง, ตายที่ห้องฉุกเฉิน, ตายในตึกภายใน 24 ชั่วโมง หลังเหตุ และตายในตึกหลัง 24 ชั่วโมง ถึง 30 วัน

หลังจากที่ทำการจัดการรวมกลุ่มของผลการรักษาผู้บาดเจ็บ (Class Label) อีกปัญหาของข้อมูลในชุดนี้คือเรื่อง Imbalanced Data คือข้อมูลของ Class คำตอบของผลการรักษาผู้บาดเจ็บ ซึ่งแบ่งออกเป็น 2 Class ประกอบด้วย อาการบาดเจ็บที่ไม่รุนแรงถึงชีวิตกับอาการบาดเจ็บรุนแรงถึงการเสียชีวิต มีข้อมูลทั้งหมด 214,950 แถว แบ่งเป็นอาการบาดเจ็บที่ไม่รุนแรงถึงชีวิตมีข้อมูลทั้งหมด 211,386 แถว และอาการบาดเจ็บรุนแรงถึงการเสียชีวิตมีข้อมูลทั้งหมด 3,564 แถว จะเห็นได้ว่าข้อมูล Class ผู้บาดเจ็บที่ไม่เสียชีวิต (Non Fatal) มีจำนวนมากกว่าอีก Class ที่เป็นผู้บาดเจ็บที่เสียชีวิต (Fatal) ดังภาพประกอบ 56



ภาพประกอบ 56 แสดงจำนวนผลการรักษาผู้บาดเจ็บอุบัติเหตุ (Class Label)

โดยข้อมูลที่เป็น Categorical Data ของข้อมูลชุดนี้ประกอบด้วย เวลาเกิดเหตุ, เพศ, ถนนที่เกิดเหตุ, สถานะ, รถผู้บาดเจ็บ, รถคู่กรณี, มาตรการ,การดื่มสุรา, การนำส่ง, Refer-Admit และผลการรักษา ถูกแปลงด้วยเทคนิคของ One Hot Encoding ดังภาพประกอบ 57

รหัสจังหวัด	รหัสรพ.	วันที่เกิดเหตุ	อายุ	ผลการรักษา	จำนวนวันรักษา	เวลาเกิดเหตุ_cat	เพศ_cat	ถนนที่เกิดเหตุ_cat	สถานะ_cat	รถผู้บาดเจ็บ_cat	รถคู่กรณี_cat	มาตรการ_cat	การดื่มสุรา_cat	การนำส่ง_cat	Refer-Admit_cat
10	11630	3	54	0	0	11	1	2	1	5	12	0	1	4	3
10	11548	2	13	0	0	20	1	2	1	0	3	1	1	5	0
10	11548	31	0	0	0	2	1	2	2	0	1	3	1	4	3
10	11548	31	0	0	0	2	1	2	2	0	1	1	1	4	3
10	11548	31	0	0	0	2	0	2	1	0	1	1	0	4	3

ภาพประกอบ 57 ข้อมูล Categorical Data ที่ผ่านการแปลงด้วยเทคนิค One Hot Encoding

3.5 อัลกอริทึมของแบบจำลองของการทำนาย

ในการสร้างแบบจำลองการทำนายความรุนแรงผู้บาดเจ็บจากอุบัติเหตุ แบ่งข้อมูล (Split Data) ออกเป็นข้อมูลฝึกสอน (Training Set) และข้อมูลทดสอบ (Test Set) ในอัตราส่วน 60:40 และ Training set แบ่งข้อมูลเป็น Train : 80 และ Validation : 20 โดยใช้เทคนิค Stratify ในการ Split ข้อมูล สำหรับระดับความรุนแรงของอุบัติเหตุ 2 ระดับคือ 0 : ไม่เสียชีวิต และ 1 : เสียชีวิต ดังรูปภาพประกอบที่ 58

<table>
<tr><th colspan="2">Training Set (60%)</th><th>Test Set (40%)</th></tr>
<tr><td colspan="2">128,970
(0:126,832, 1:2,138)</td><td rowspan="3">85,980
(0:84,554, 1:1,426)</td></tr>
<tr><th>Train (80%)</th><th>Validation (20%)</th></tr>
<tr><td>103,176
(0:101,466, 1:1,710)</td><td>25,794
(0:25,366, 1:428)</td></tr>
</table>

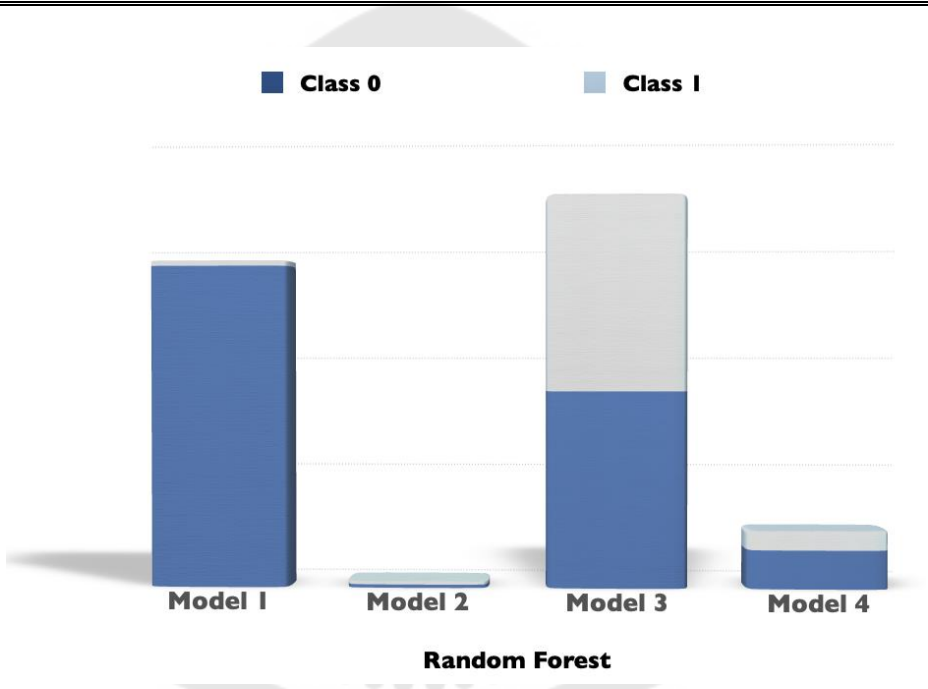
ภาพประกอบ 58 แสดงข้อมูลที่ใช้สำหรับการสร้างและประเมินแบบจำลองการทำนายความรุนแรงผู้บาดเจ็บจากอุบัติเหตุ

ในงานวิจัยนี้การพัฒนาแบบจำลองการทำนายความรุนแรงของผู้บาดเจ็บจากอุบัติเหตุทางถนนในแบบการจำแนกประเภทแบบไบนารี อัลกอริทึมการเรียนรู้ของเครื่องที่ใช้คือ Random Forests และใช้เทคนิคการสุ่มตัวอย่างใหม่ (Resampling Techniques) 3 แบบ เพื่อปรับความสมดุลของข้อมูลอินสแตนซ์ในข้อมูลฝึกสอนของแบบจำลองการทำนาย (Rebalance Training Instances) ได้แก่ เทคนิคการลดจำนวนตัวอย่าง (Undersampling), เทคนิคการเพิ่มจำนวนตัวอย่างด้วย SMOTE (Oversampling with SMOTE) และการผสมระหว่างของทั้ง 2 เทคนิค ของเทคนิคการลดจำนวนตัวอย่าง (Undersampling) กับเทคนิคการเพิ่มจำนวนตัวอย่างด้วย SMOTE (Oversampling with SMOTE)

แบบจำลองการทำนายทั้ง 4 ได้รับการฝึกฝนและปรับพารามิเตอร์ของอัลกอริทึม Random Forests ผ่านการจัดลำดับด้วย Pipeline แบบ Imbalanced โดยข้อมูลที่ผ่านการใช้เทคนิค Resampling จะมีข้อมูลดังตารางที่ 3 และภาพประกอบ 59

ตาราง 3 แสดงข้อมูลฝึกสอนแบบจำลองการทำนาย

แบบจำลองการทำนาย	**เทคนิค** Resampling
แบบจำลองที่ 1 (Model 1)	No Resampling (ไม่ใช้) (Class 0: 126,832, Class 1: 2,138)
แบบจำลองที่ 2 (Model 2)	Random Undersampling (Class 0: 2,123, Class 1: 2,138)
แบบจำลองที่ 3 (Model 3)	Oversampling with SMOTE (Class 0: 126,832, Class 1: 126,832)
แบบจำลองที่ 4 (Model 4)	Oversampling with SMOTE และ Random Undersampling (Class 0: 25,366, Class 1: 12,683)



ภาพประกอบ 59 กราฟแท่งแสดงสัดส่วนข้อมูลฝึกสอนแบบจำลองการทำนาย

แบบจำลองที่ 1 (Model 1) การสร้างแบบจำลอง Random Forest ที่ไม่ใช้เทคนิค Resampling ข้อมูลที่ใช้ในการฝึกสอนแบบจำลองการทำนายทั้งหมด 128,970 แถว มีความไม่สมดุล แบ่งเป็น Class Target 0 : ผู้บาดเจ็บจากอุบัติเหตุไม่เสียชีวิต ทั้งหมด 126,832 แถว และ Class Target 1 : ผู้บาดเจ็บจากอุบัติเหตุเสียชีวิต ทั้งหมด 2,138 แถว การปรับจูนพารามิเตอร์ของแบบจำลองที่เหมาะสมด้วยวิธี Grid Search เพื่อหาค่าพารามิเตอร์ที่ดีที่สุด ซึ่งได้ค่า max_depth = 70, ค่า min_samples_split = 2, ค่า max_features = ‘auto’ และค่า min_samples_leaf = 1

แบบจำลองที่ 2 (Model 2) การสร้างแบบจำลอง Random Forest ที่ใช้เทคนิค Resampling แบบ Random Undersampling ข้อมูลที่ใช้ในการฝึกสอนแบบจำลองการทำนาย

ทั้งหมด 4,276 แถว แบ่งเป็น Class Target 0 : ผู้บาดเจ็บจากอุบัติเหตุไม่เสียชีวิต ทั้งหมด 2,138 แถว และ Class Target 1 : ผู้บาดเจ็บจากอุบัติเหตุเสียชีวิต ทั้งหมด 2,138 แถว การปรับจูนพารามิเตอร์ของแบบจำลองที่เหมาะสมด้วยวิธี Grid Search เพื่อหาค่าพารามิเตอร์ที่ดีที่สุด ซึ่งได้ค่า max_depth = 63 , ค่า min_samples_split = 2, ค่า max_features = 28, และค่า min_samples_leaf = 1

แบบจำลองที่ 3 (Model 3) การสร้างแบบจำลอง Random Forest ที่ใช้เทคนิค Resampling แบบ Oversampling with SMOTE ข้อมูลที่ใช้ในการฝึกสอนแบบจำลองการทำนายทั้งหมด 253,664 แถว แบ่งเป็น Class Target 0 : ผู้บาดเจ็บจากอุบัติเหตุไม่เสียชีวิต ทั้งหมด 126,832 แถว และ Class Target 1 : ผู้บาดเจ็บจากอุบัติเหตุเสียชีวิต ทั้งหมด 126,832 แถว การปรับจูนพารามิเตอร์ของแบบจำลองที่เหมาะสมด้วยวิธี Grid Search เพื่อหาค่าพารามิเตอร์ที่ดีที่สุด ซึ่งได้ค่า max_depth = ‘None’, ค่า min_samples_split = 0.78, ค่า max_features = ‘auto’ และค่า min_samples_leaf = 0.32

แบบจำลองที่ 4 (Model 4) การสร้างแบบจำลอง Random Forest ที่ใช้เทคนิค Resampling แบบผสมระหว่าง Random Undersampling และ Oversampling with SMOTE จะเหลือข้อมูลที่ใช้ในการฝึกสอนแบบจำลองการทำนายทั้งหมด 38,049 แถว แบ่งเป็น Class Target 0 : ผู้บาดเจ็บจากอุบัติเหตุไม่เสียชีวิต ทั้งหมด 25,366 แถว และ Class Target 1 : ผู้บาดเจ็บจากอุบัติเหตุเสียชีวิต ทั้งหมด 12,683 แถว การปรับจูนพารามิเตอร์ของแบบจำลองที่เหมาะสมด้วยวิธี Grid Search เพื่อหาค่าพารามิเตอร์ที่ดีที่สุด ซึ่งได้ค่า max_depth = 5, ค่า min_samples_split = 0.12, ค่า max_features = 101, และค่า min_samples_leaf = 0.01

โดยที่ทั้งสี่แบบจำลองใช้จำนวนต้นไม้เท่ากับ 100 ต้น และแบบจำลองการทำนายทั้งหมดได้รับการทำ Cross Validated เท่ากับ 5 Fold และใช้ Stratified Shuffle Splits ในการสุ่มแบ่งข้อมูล ดังในตาราง 4 แสดงการสรุปพารามิเตอร์ทั้งหมด

ตาราง 4 แสดงพารามิเตอร์ที่ดีที่สุดของแต่ละแบบจำลองการทำนายความรุนแรงของผู้บาดเจ็บจากอุบัติเหตุทางถนนในช่วงเทศกาลปีใหม่ 2551 ถึง 2558

Random Forest	Max Depth	Min Samples Split	Max Features	Min Samples Leaf	N Estimators
แบบจำลองที่ 1 (No Resampling ไม่ใช้)	70	2	Auto	1	100
แบบจำลองที่ 2 (Random Undersampling)	63	2	28	1	100
แบบจำลองที่ 3 (Oversampling with SMOTE)	None	0.78	Auto	0.32	100
แบบจำลองที่ 4 (Oversampling with SMOTE และ Random Undersampling)	5	0.12	101	0.01	100

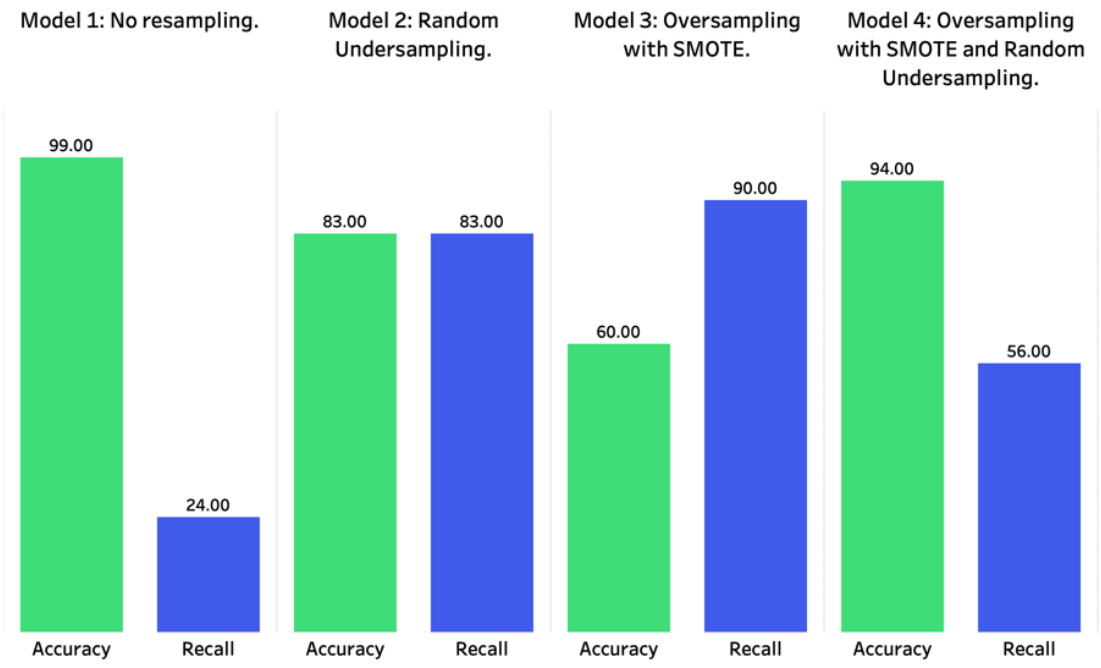
แบบจำลองที่ผ่านการปรับค่าพารามิเตอร์ที่ดีที่สุดมาแล้วใช้เทคนิค Pipeline ในการจัดลำดับการทำงานของปรับความสมดุลของข้อมูลฝึกสอนและการฝึกสอนแบบจำลองการทำนายเพื่อป้องกันและหลีกเลี่ยงการถูกสอดแนม รั่วไหลของข้อมูลทดสอบในระหว่างการฝึกสอนแบบจำลอง (Data Leakage) ในขั้นตอนสุดท้ายการทดสอบประสิทธิภาพของแต่ละแบบจำลองการทำนาย โดยใช้ Classification Report ประกอบด้วย Precision, Recall, Accuracy, F1 และ Confusion Matrix ถูกสร้างขึ้นเพื่อประเมินผลแบบจำลองการทำนายด้วยข้อมูลทดสอบ (Test Set) แต่ในงานวิจัยเรื่องนี้ไม่ได้เพียงแต่คำนึงถึงความถูกต้องของการทำนายทั้งหมด (Accuracy) เท่านั้น แต่ยังมุ่งเน้นไปที่ค่า Recall เพื่อวัดประสิทธิภาพที่แท้จริงของการระบุจำนวนผู้บาดเจ็บจากอุบัติเหตุที่รุนแรงถึงเสียชีวิตจากอุบัติเหตุทางถนนในช่วงเทศกาลปีใหม่ 2551 ถึง 2558 ให้ถูกต้อง

บทที่ 4
ผลการดำเนินการวิจัย

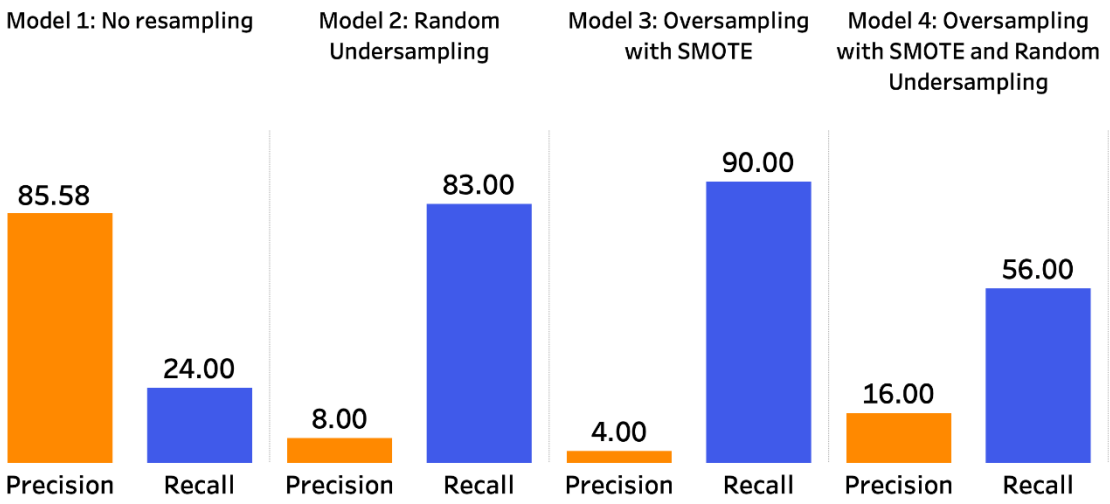
ในการวิจัยทำนายความรุนแรงของผู้บาดเจ็บจากอุบัติเหตุทางถนนในช่วงเทศกาลปีใหม่ ซึ่งใช้ข้อมูลผู้บาดเจ็บจากอุบัติเหตุในช่วงเทศกาลปีใหม่ 2551 ถึง 2558 โดยใช้เทคนิคการเรียนรู้ของเครื่อง ผู้วิจัยได้ดำเนินการวิจัยโดยการศึกษาตามขบวนการและขั้นตอนต่างๆ ตลอดจนการประเมินผลของแบบจำลองการทำนาย เพื่อให้บรรลุวัตถุประสงค์ของการวิจัยที่ได้กำหนดไว้ดังนี้

1. ผลลัพธ์ประสิทธิภาพของแบบจำลองการทำนาย
2. ผลลัพธ์การวิเคราะห์ปัจจัยแบบจำลองการทำนายความรุนแรงของผู้บาดเจ็บจากอุบัติเหตุทางถนนในช่วงเทศกาลปีใหม่

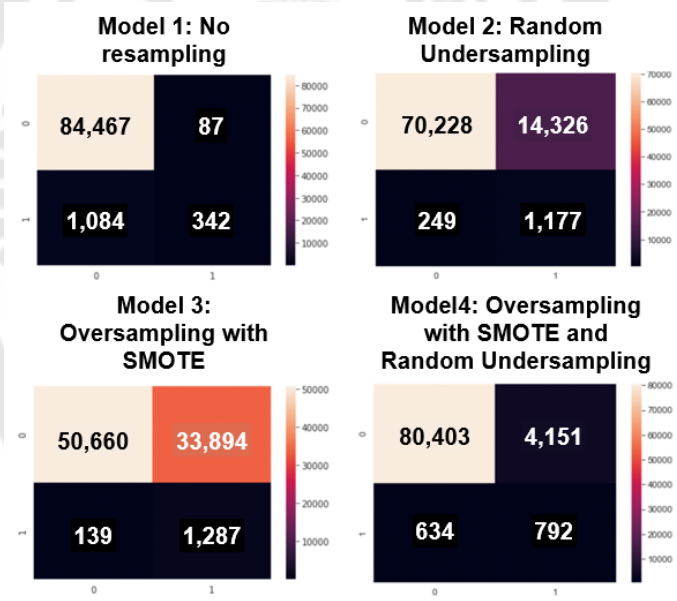
4.1 ผลลัพธ์ประสิทธิภาพของแบบจำลองการทำนาย



ภาพประกอบ 60 กราฟแท่งแสดงข้อมูลเปรียบเทียบค่าความถูกต้องของ Overall Accuracy และ Recall



ภาพประกอบ 61 กราฟแท่งแสดงข้อมูลเปรียบเทียบค่า Precision และ Recall



ภาพประกอบ 62 แสดงข้อมูลของ Confusion Matrices แกน y คือค่าข้อมูลจริง แกน x คือค่าทำนาย

จากผลการทดลองดังกล่างของงานวิจัยเรื่องนี้ ที่ได้ทำการที่ใช้อัลกอริทึมของ Random Forest เพราะเนื่องจากได้ทำการทดลองทั้งหมด 6 อัลกอริทึม ประกอบด้วย Random Forest, AdaBoost, Gaussian Naive Bayes, LightGBM, XGBoost และ Gradient Boosting ที่ใช้เทคนิค Resampling แบบ Random Undersampling ในการลดข้อมูลตัวอย่างก่อนฝึกฝนแบบจำลอง และผลลัพธ์ที่ออกมาดีที่สุดเป็นของ Random Forest จึงได้เลือกใช้อัลกอริทึม

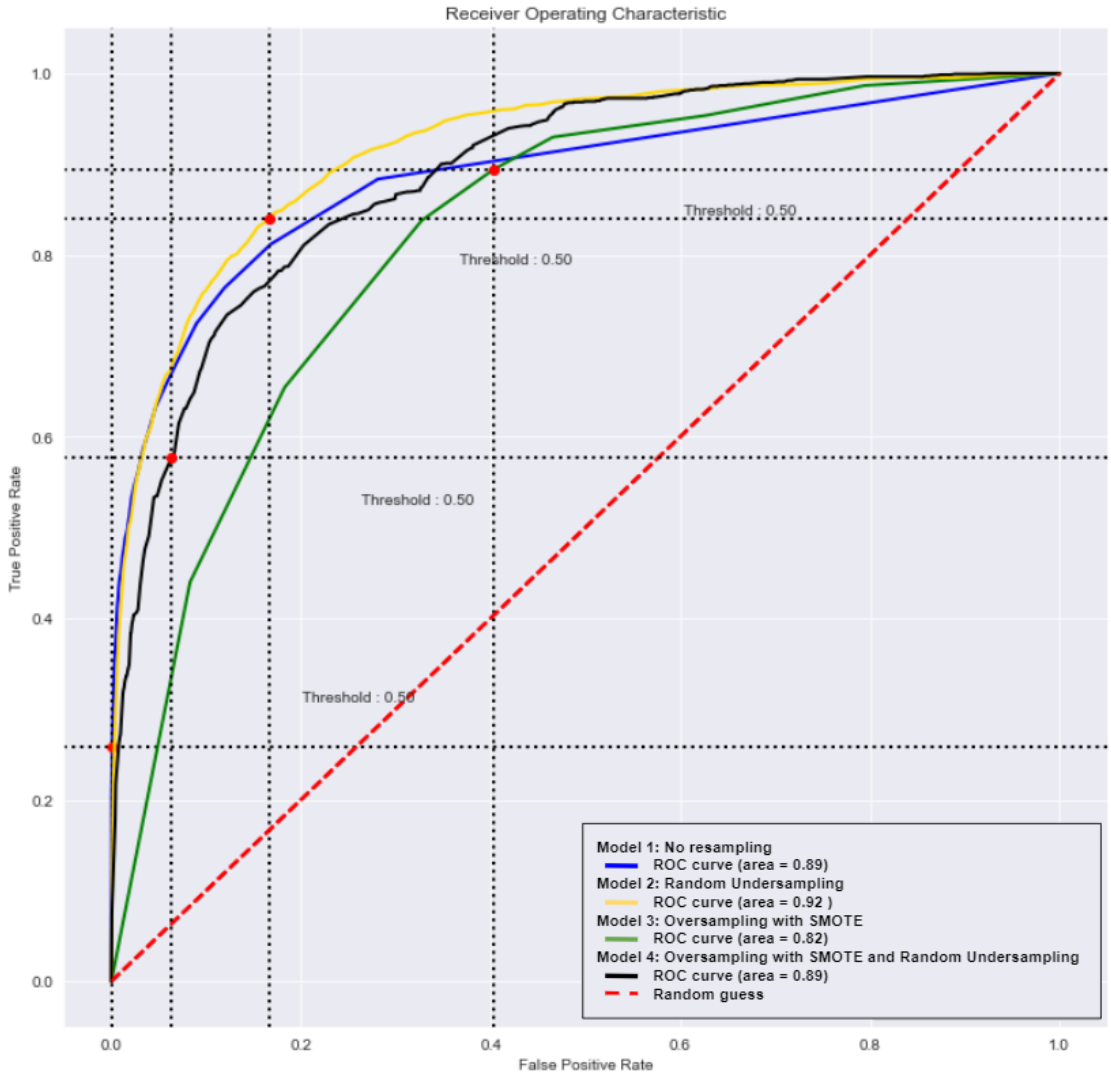
Random Forest มาใช้ในการทดลองสร้างแบบจำลอง ซึ่งจากภาพประกอบ 60 ถึง 62 แบบจำลองที่ 1 Random Forest ที่ไม่ใช้เทคนิค Resampling ในจัดการปัญหา Imbalanced Data สามารถทำนายความรุนแรงของผู้บาดเจ็บจากอุบัติเหตุทางถนนในช่วงเทศกาลปีใหม่ได้ที่ความถูกต้องทั้งหมดของการทำนายที่ค่า Accuracy เท่ากับ 99 เปอร์เซ็นต์ ค่า Recall เท่ากับ 24 เปอร์เซ็นต์ และมีค่า Precision เท่ากับ 85.58 เปอร์เซ็นต์ จากข้อมูลของ Confusion Matrix แสดงให้เห็นว่าการทำนายคนที่เสียชีวิตถูกต้อง 342 คน เทียบกับอัตราส่วนของจำนวนคนเสียชีวิตจริงที่เป็น True Positive ทั้งหมด 1,426 คน เป็นสิ่งที่ยังไม่สามารถทำนายตรวจจับคนที่เสียชีวิตจากอุบัติเหตุได้ ผลของการทำนายว่าอุบัติเหตุส่วนใหญ่ผู้บาดเจ็บจากอุบัติเหตุไม่เสียชีวิต ถึงแม้ว่าความถูกต้องของการทำนาย (Accuracy) และค่า Precision นั้นสูง ดังนั้นค่า Accuracy จึงไม่สามารถใช้ในการวัดประสิทธิภาพของแบบจำลองนี้ได้ เนื่องจากข้อมูลฝึกสอนแบบจำลองยังมีปัญหาความไม่สมดุลของข้อมูลเพราะจำนวนข้อมูลของผู้ไม่เสียชีวิตจากอุบัติเหตุมีจำนวนมากกว่าผู้เสียชีวิตหลายเท่า ส่วนค่า Recall สามารถบอกผลการทำนายของจำนวนผู้เสียชีวิตที่ถูกต้องต่อการทำนายผู้เสียชีวิตทั้งหมด ทำให้แบบจำลองนี้ยังไม่สามารถตรวจจับจำนวนผู้เสียชีวิตได้อย่างมีประสิทธิภาพ

แบบจำลองที่ 2 Random Forest ที่ใช้เทคนิค Resampling แบบ Random Undersampling สามารถทำนายความรุนแรงของผู้บาดเจ็บจากอุบัติเหตุทางถนนในช่วงเทศกาลปีใหม่ได้ที่ความถูกต้องทั้งหมดของการทำนายที่ค่า Accuracy เท่ากับ 83 เปอร์เซ็นต์ ค่า Recall เท่ากับ 83 เปอร์เซ็นต์ และมีค่า Precision เท่ากับ 8 เปอร์เซ็นต์ จากข้อมูลของ Confusion Matrix แสดงให้เห็นว่าการทำนายคนที่เสียชีวิตถูกต้อง 1,177 คน เทียบกับอัตราส่วนของจำนวนคนเสียชีวิตจริงที่เป็น True Positive ทั้งหมด 1,426 คน จะเห็นได้ว่าที่ค่า Precision นั้นต่ำมาก เพราะจำนวนข้อมูลของผู้ไม่เสียชีวิตจากอุบัติเหตุแต่ถูกทำนายว่าเสียชีวิตเป็นจำนวนมาก ส่วนค่า Recall สามารถบอกผลการทำนายของจำนวนผู้เสียชีวิตที่ถูกต้องต่อการทำนายผู้เสียชีวิตทั้งหมด ทำให้ผลของแบบจำลองการทำนายเป็นสิ่งที่สามารถระบุจำนวนผู้บาดเจ็บจากอุบัติเหตุที่รุนแรงถึงการเสียชีวิตในช่วงเทศกาลปีใหม่ได้

แบบจำลองที่ 3 Random Forest ที่ใช้เทคนิค Resampling แบบ Oversample with SMOTE สามารถทำนายความรุนแรงของผู้บาดเจ็บจากอุบัติเหตุทางถนนในช่วงเทศกาลปีใหม่ได้ที่ความถูกต้องทั้งหมดของการทำนายที่ค่า Accuracy เท่ากับ 60 เปอร์เซ็นต์ ค่า Recall เท่ากับ 90 เปอร์เซ็นต์ และมีค่า Precision เท่ากับ 4 เปอร์เซ็นต์ จากข้อมูลของ Confusion Matrices แสดงให้เห็นว่าการทำนายคนที่เสียชีวิตถูกต้อง 1,287 คน เทียบกับอัตราส่วนของจำนวนคนเสียชีวิตจริง

ที่เป็น True Positive ทั้งหมด 1,426 คน ผลของการทำนายเป็นสิ่งที่มีข้อได้เปรียบเล็กน้อยในค่า Recall เมื่อเทียบกับแบบจำลอง Random Forest ที่ใช้เทคนิคของ Random Undersampling ที่ค่า Recall เท่ากับ 90 เปอร์เซ็นต์ และ 83 เปอร์เซ็นต์ ตามลำดับ อย่างไรก็ตามแบบจำลอง Random Forest ที่ใช้เทคนิคของ Oversample with SMOTE สามารถเห็นได้จาก Confusion Matrix ข้อมูลเกือบครึ่งหนึ่งของอุบัติเหตุที่มีผู้บาดเจ็บไม่เสียชีวิตทั้งหมด 33,894 คน แต่ได้รับการทำนายว่าเป็นอุบัติเหตุที่มีความรุนแรงถึงการเสียชีวิต จะเห็นได้ว่าที่ค่า Precision นั้นต่ำมาก เพราะจำนวนข้อมูลของผู้ไม่เสียชีวิตจากอุบัติเหตุแต่ถูกทำนายว่าเสียชีวิตเป็นจำนวนมากกว่าครึ่ง ส่วนค่า Recall สามารถบอกผลการทำนายของจำนวนผู้เสียชีวิตที่ถูกต้องต่อการทำนายผู้เสียชีวิตทั้งหมด ทำให้แบบจำลองนี้สามารถตรวจจับจำนวนผู้เสียชีวิตได้มากที่สุด แต่ก็ทำนายผู้ไม่เสียชีวิตจากอุบัติเหตุแต่ถูกทำนายว่าเสียชีวิตมากด้วยเช่นกัน จะทำให้ความสามารถของแบบจำลองโดยรวมลดลง เพราะเป็นที่ยอมรับไม่ได้ที่ค่าความถูกต้องเท่ากับ 60 เปอร์เซ็นต์

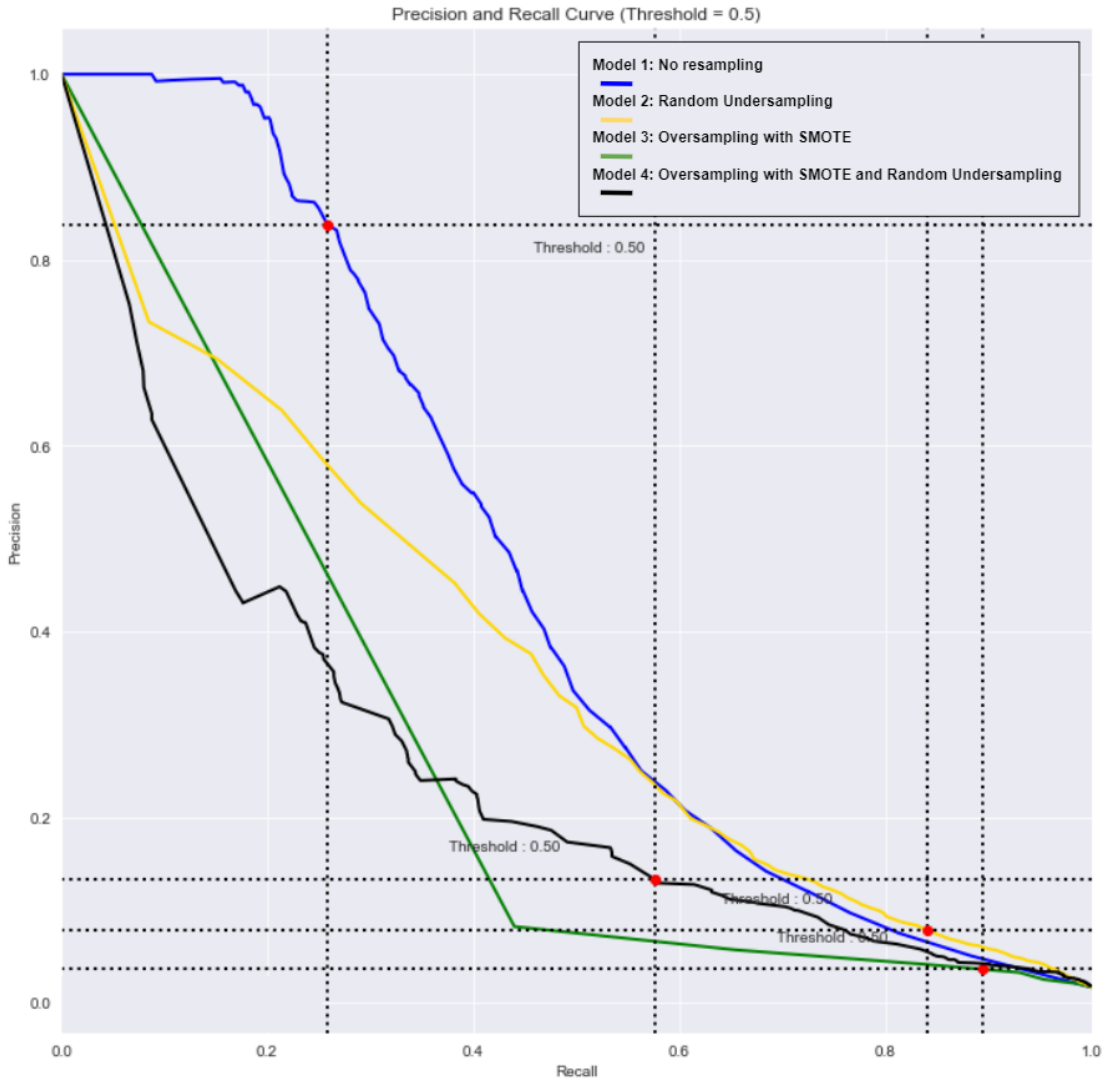
แบบจำลองที่ 4 Random Forest ที่ใช้เทคนิค Resampling แบบ Oversampling with SMOTE และ Random Undersampling สามารถทำนายความรุนแรงของผู้บาดเจ็บจากอุบัติเหตุทางถนนในช่วงเทศกาลปีใหม่ได้ที่ความถูกต้องทั้งหมดของการทำนายที่ค่า Accuracy เท่ากับ 94 เปอร์เซ็นต์ ค่า Recall เท่ากับ 56 เปอร์เซ็นต์ และมีค่า Precision เท่ากับ 16 เปอร์เซ็นต์ จากข้อมูลของ Confusion Matrix แสดงให้เห็นว่าการทำนายคนที่เสียชีวิตถูกต้อง 792 คน เทียบกับอัตราส่วนของจำนวนคนเสียชีวิตจริงที่เป็น True Positive ทั้งหมด 1,426 คน ผลของการทำนายซึ่งแนวทางไฮบริด(Hybrid) ของเทคนิค Oversampling และ Undersampling ยังคงมีอคติในข้อมูลฝึกสอนแบบจำลองการทำนาย ข้อมูลฝึกสอนแบบจำลองการทำนายยังมีปัญหาของความไม่สมดุล ทำให้ผลลัพธ์ที่ได้ เห็นถึงการปรับปรุงที่เพิ่มขึ้นเกือบสองเท่าในค่า Recall เมื่อเปรียบเทียบกับแบบจำลองการทำนายที่ไม่ใช้เทคนิค Resampling แต่ก็ยังสามารถทำนายจำนวนผู้เสียชีวิตและไม่เสียชีวิตได้ดี แต่ความถูกต้องของการทำนายที่สามารถตรวจจับจำนวนผู้เสียชีวิตได้ยังไม่ดีพอ เพราะสามารถทำนายจำนวนผู้เสียชีวิตที่ถูกต้องได้ครึ่งหนึ่งของจำนวนผู้เสียชีวิตจริงทั้งหมด



ภาพประกอบ 63 แสดง Receiver Operating Characteristic (ROC Curve) เปรียบเทียบ 4 แบบจำลอง

จากภาพประกอบ 63 แสดง Receiver Operating Characteristic Curve (ROC Curve) เป็นวิธีที่นิยมใช้ในการวัดประสิทธิภาพความถูกต้องของการพยากรณ์ เป็นกราฟที่มีความสัมพันธ์ระหว่างแกน Y แทน Sensitivity (True Positive Rate) กับแกน X แทน 1-Specificity (False Positive Rate) แบบจำลองที่ 2 Random Forest ที่ใช้เทคนิค Resampling แบบ Random Undersampling สามารถทำนายความรุนแรงของผู้บาดเจ็บจากอุบัติเหตุทางถนนในช่วงเทศกาลปีใหม่ ที่ได้ผลลัพธ์ของค่าพื้นที่ใต้กราฟ ROC Curve เท่ากับ 0.92 สูงที่สุดจากแบบจำลองการทำนายทั้งหมด ซึ่งการมีค่าพื้นที่ใต้กราฟโค้ง ROC Curve สูง เป็นการบ่งชี้ถึงความถูกต้องหรือความน่าเชื่อถือของตัวแบบจำลองการทำนายที่มีประสิทธิภาพมากสุด อีกทั้งที่ Threshold 50 %

ยังได้ False Positive Rate ที่ต่ำกว่า 20 เปอร์เซ็นต์ และ True Positive Rate ที่สูงกว่า 80 เปอร์เซ็นต์ นอกจากนี้ยังใช้จำนวนข้อมูลฝึกสอนแบบจำลองน้อยลง ทำให้เวลาของการฝึกสอนแบบจำลองการทำนายความรุนแรงของผู้บาดเจ็บจากอุบัติเหตุทางถนนในช่วงเทศกาลปีใหม่ปี 2551 ถึง 2558 ใช้เวลาน้อยลง



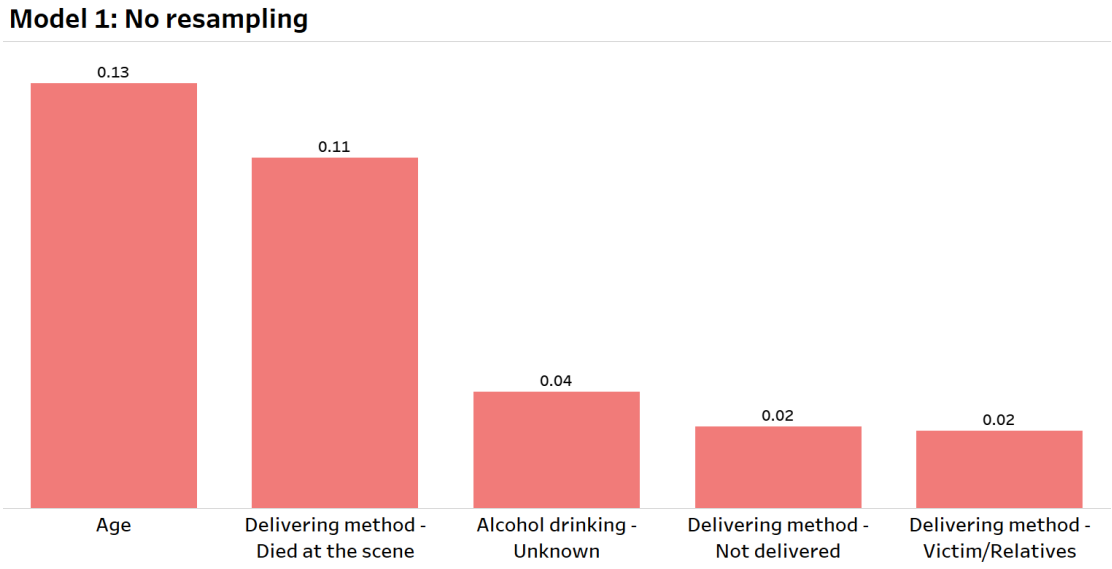
ภาพประกอบ 64 แสดง Precision-Recall Curve เปรียบเทียบ 4 แบบจำลอง

จากภาพประกอบ 64 แสดง Precision-Recall Curve จะแสดงให้เห็นของค่า Precision ซึ่งเป็นการแสดงให้เห็นถึงการทำนายของแบบจำลองในการทำนายจำนวนผู้เสียชีวิตจากอุบัติเหตุ (Positive Class) ว่าถูกต้องแค่ไหน โดยมาจากจำนวนของผลการทำนายจำนวนผู้เสียชีวิตที่ถูกต้อง (True Positives) เทียบกับผลรวมของการทำนายจำนวนผู้เสียชีวิตที่ถูกต้อง (True

Positives) กับผลรวมของการทำนายที่ไม่ถูกต้องคือผู้บาดเจ็บไม่เสียชีวิตแต่ทำนายว่าเสียชีวิตจากอุบัติเหตุ (False Positives) และ Recall ซึ่งเป็นการแสดงให้เห็นถึง Sensitivity ของแบบจำลองการทำนาย ที่มาจากจำนวนของผลรวมของการทำนายจำนวนผู้เสียชีวิตที่ถูกต้อง (True Positives) เทียบกับผลรวมของการทำนายจำนวนผู้เสียชีวิตที่ถูกต้อง (True Positives) กับการทำนายที่ผิดผลาดคือผู้บาดเจ็บที่เสียชีวิตแต่ทำนายว่าไม่เสียชีวิตจากอุบัติเหตุ (False negative) ซึ่งทั้ง 2 ค่าจะมีการ Tradeoff กันที่ Threshold 50 เปอร์เซ็นต์ แบบจำลองที่ 2 และแบบจำลองที่ 3 ที่ได้ค่า Recall ดีสุด แม้ค่า Precision ค่อนข้างต่ำ

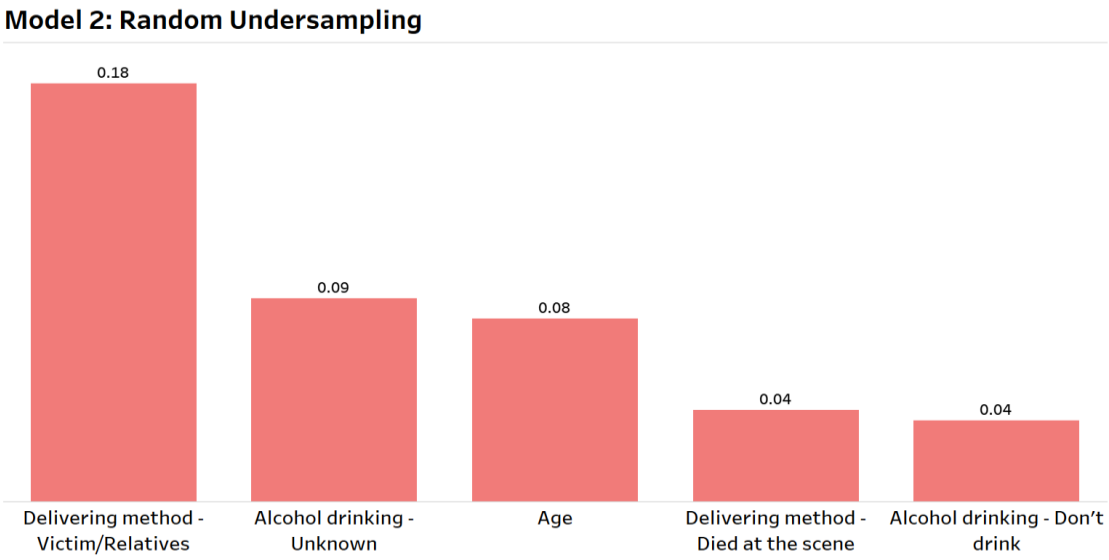
4.2 ผลลัพธ์การวิเคราะห์ปัจจัยในการทำนายความรุนแรงของผู้บาดเจ็บจากอุบัติเหตุทางถนนในช่วงเทศกาลปีใหม่

การวิเคราะห์ปัจจัยที่มีผลต่อการตัดสินในการทำนายความรุนแรงของผู้บาดเจ็บจากอุบัติเหตุทางถนนในช่วงเทศกาลปีใหม่ใช้ Feature Importance ของแต่ละแบบจำลอง Random Forest เพื่อให้ได้ปัจจัย (Feature) ที่มีความสำคัญส่งผลต่อการทำนายความรุนแรงของผู้บาดเจ็บจากอุบัติเหตุทางถนนในช่วงเทศกาลปีใหม่ 2551 ถึง 2558 ดังภาพประกอบ 65 ถึง 68



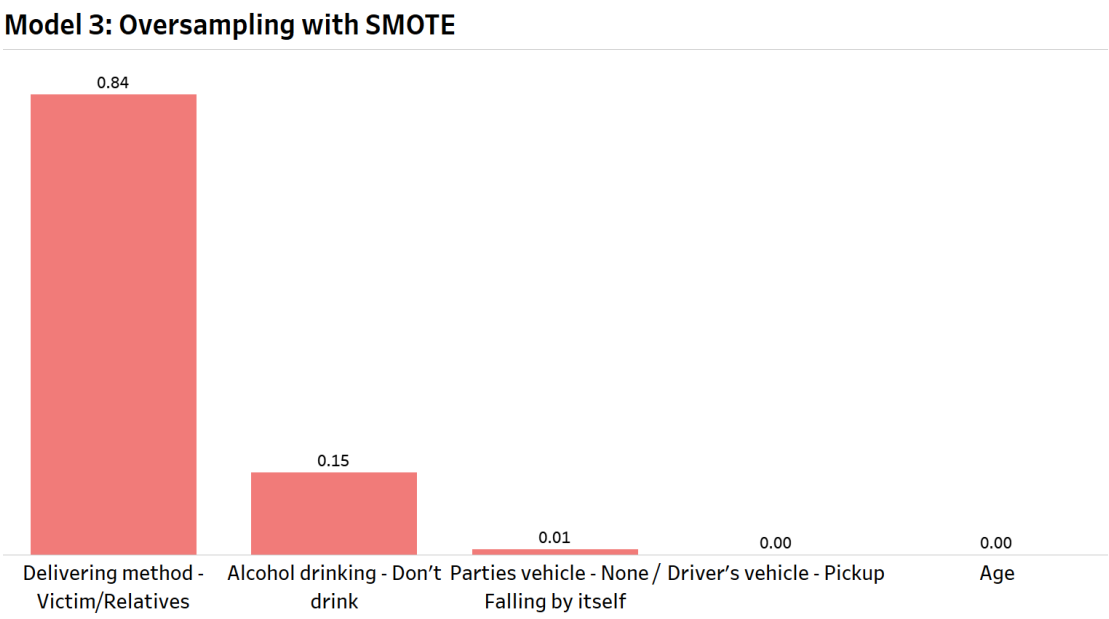
ภาพประกอบ 65 แสดงข้อมูล Reported Feature importances โดยแบบจำลองที่ 1 Random Forest ที่ไม่ใช้เทคนิคจัดการ Imbalanced Data

แบบจำลองที่ 1 Random Forest ที่ไม่ใช้เทคนิค Resampling ปัจจัยที่มีผลต่อการทำนายมากที่สุดคืออายุ (Age) ที่มีค่าสัดส่วน 0.13 การนำส่งในลักษณะที่เสียชีวิตในที่เกิดเหตุ (Delivering Method – Died at The Scene) ให้ค่าสัดส่วน 0.11 การไม่ทราบดื่มแอลกอฮอล์ (Alcohol Drinking-Unknown) ให้ค่าสัดส่วน 0.04 การนำส่งที่ไม่มีการนำส่งผู้บาดเจ็บจากอุบัติเหตุ (Delivering Method – Not Delivered)ให้ค่าสัดส่วน 0.02 และการนำส่งในลักษณะของผู้ประสบอุบัติเหตุเองหรือญาติ (Delivering Method – Victim/Relatives) ให้ค่าสัดส่วน 0.02



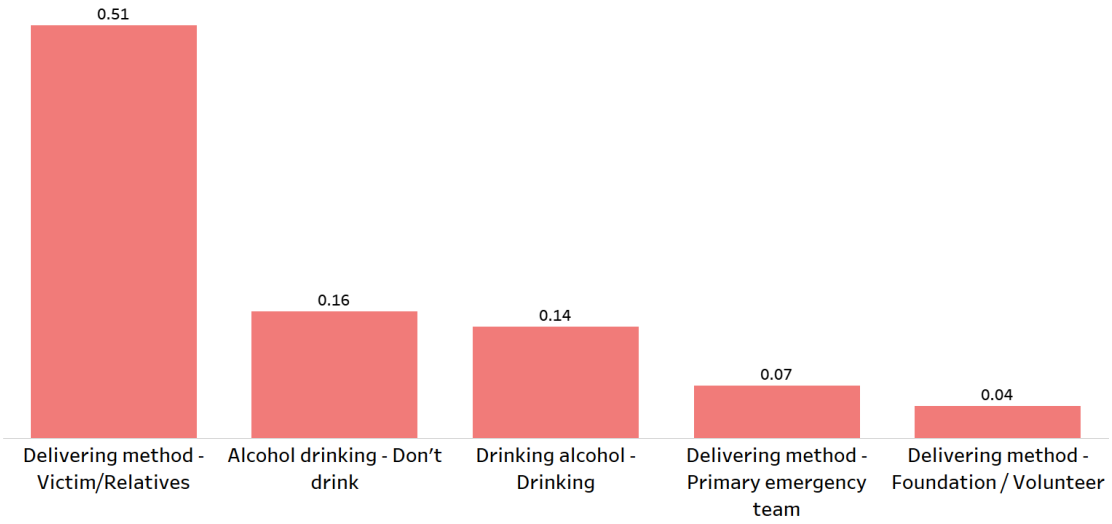
ภาพประกอบ 66 แสดงข้อมูล Reported Feature importances โดยแบบจำลองที่ 2 Random Forest ที่ใช้เทคนิคจัดการ Imbalanced Data ของ Random Undersample

แบบจำลองที่ 2 Random Forest ที่ใช้เทคนิค Resampling แบบ Random Undersampling ปัจจัยที่มีผลต่อการทำนายแนวทิศทางเดียวกันมากที่สุดคือ การนำส่งในลักษณะของผู้ประสบอุบัติเหตุเองหรือญาติ (Delivering Method – Victim/Relatives) ให้ค่าสัดส่วน 0.18 การไม่ทราบดื่มแอลกอฮอล์ (Alcohol Drinking-Unknown) ให้ค่าสัดส่วน 0.09 อายุ (Age) ที่มีค่าสัดส่วน 0.08 การนำส่งในลักษณะที่เสียชีวิตในที่เกิดเหตุ (Delivering Method – Died at The Scene) ให้ค่าสัดส่วน 0.04 และการไม่ดื่มแอลกอฮอล์ (Alcohol Drinking – Don't drink) ให้ค่าสัดส่วน 0.04



ภาพประกอบ 67 แสดงข้อมูล Reported Feature importances โดยแบบจำลองที่ 3 Random Forest ที่ใช้เทคนิคจัดการ Imbalanced Data ของ Oversample with SMOTE

แบบจำลองที่ 3 Random Forest ที่ใช้เทคนิค Resampling แบบ Oversampling with SMOTE ปัจจัยที่มีผลต่อการทำนายมากที่สุดคือการนำส่งในลักษณะของผู้ประสบอุบัติเหตุเองหรือญาติ (Delivering Method – Victim/Relatives) ให้ค่าสัดส่วน 0.84 การไม่ดื่มแอลกอฮอล์ (Alcohol Drinking – Don't drink) ให้ค่าสัดส่วน 0.15 รถคู่กรณีในแบบที่เกิดอุบัติเหตุโดยไม่มีรถคู่กรณี (Parties Vehicle – None/Falling by Itself) ให้ค่าสัดส่วน 0.01 ยานพาหนะของผู้ขับขี่ของรถปิคอัพ (Driver's Vehicle-Pickup) และอายุ (Age) ให้ค่าน้อยลดตามลำดับ



ภาพประกอบ 68 แสดงข้อมูล Reported Feature importances โดยแบบจำลองที่ 4 Random Forest ที่ใช้เทคนิคจัดการ Imbalanced Data ของ Oversample with SMOTE และ Random Undersample

แบบจำลองที่ 4 Random Forest ที่ใช้เทคนิค Resampling แบบ Oversampling with SMOTE และ Random Undersampling ปัจจัยที่มีผลต่อการทำนายมากที่สุดคือการนำส่งโดยผู้ประสบอุบัติเหตุเองหรือญาติ (Delivering Method – Victim/Relatives) ให้ค่าสัดส่วน 0.51 การไม่ดื่มแอลกอฮอล์ (Alcohol Drinking – Don't drink) ให้ค่าสัดส่วน 0.16 การดื่มแอลกอฮอล์ (Alcohol Drinking – Drinking) ให้ค่าสัดส่วน 0.14 การนำส่งโดยชุดปฏิบัติการฉุกเฉินเบื้องต้น (Delivering Method – Primary Emergency Team) ให้ค่าสัดส่วน 0.07 และการนำส่งโดยมูลนิธิหรืออาสาสมัคร (Delivering Method – Foundation/Volunteer) ให้ค่าสัดส่วน 0.04

ซึ่งเห็นได้ชัดเจนว่าแบบจำลองที่ 1 อายุ (Age) และการนำส่งในลักษณะที่เสียชีวิตในที่เกิดเหตุ (Delivering Method – Died at The Scene) เป็นปัจจัยที่มีผลมากที่สุด แน่นอนว่าอุบัติเหตุที่เกิดขึ้นกับคนวัยชราแล้วเสียชีวิตในที่เกิดเหตุ เป็นปัจจัยที่อันตรายถึงการเสียชีวิต ตามความเป็นจริงแบบจำลองที่ 1 ยังไม่สามารถระบุจำนวนผู้บาดเจ็บจากอุบัติเหตุที่มีความรุนแรงถึงการเสียชีวิตได้ ในแบบจำลองที่ 2 ถึง 4 ที่ใช้เทคนิค Resampling ไม่เห็นด้วย การนำส่งโดยผู้ประสบอุบัติเหตุหรือญาติ (Delivering Method – Victim/Relatives) มีส่วนช่วยในการตัดสินใจมากที่สุด สัดส่วนของการมีส่วนร่วมของปัจจัยนี้ มีมากในแบบจำลองที่ 3 Random Forest ที่ใช้เทคนิค Oversampling with SMOTE สิ่งนี้อาจทำให้เกิดการทำนายอุบัติเหตุที่ไม่รุนแรงถึงการ

เสียชีวิตเป็นการเสียชีวิต ดังที่แสดงในรูปของ Confusion Matrices ปัจจัยที่สำคัญของแบบจำลองการทำนายที่ 2 Random Forest ที่ใช้เทคนิคของ Random Undersampling เป็นการกระจายระหว่างปัจจัยของการนำส่ง (Delivering Method) อายุ (Age) และการดื่มแอลกอฮอล์ (Alcohol Drinking) ในขณะที่แบบจำลองการทำนายที่ 4 Random Forest ที่ใช้เทคนิคของ Oversampling with SMOTE และ Random Undersampling (Hybrid Model) ปัจจัยที่มีผลมากที่สุดคือการดื่มแอลกอฮอล์ (Alcohol Drinking - Drinking) และการนำส่งโดยผู้บาดเจ็บจากอุบัติเหตุหรือญาติ (Delivering Method – Victim/Relatives)

บทที่ 5
สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

ในการวิจัยการทำนายความรุนแรงของผู้บาดเจ็บจากอุบัติเหตุทางถนนในช่วงเทศกาลปีใหม่ของประเทศไทย ซึ่งใช้ข้อมูลเปิดภาครัฐของประเทศไทยในชุดข้อมูลผู้บาดเจ็บจากอุบัติเหตุจราจรในช่วงเทศกาลปีใหม่ 2551 ถึง 2558 จากองค์กรสถาบันการแพทย์ฉุกเฉินแห่งชาติ โดยใช้เทคนิคการเรียนรู้แบบมีผู้สอน (Supervised Learning) ในแบบ Classification Algorithm เพื่อสร้างแบบจำลองที่มุ่งเน้นการแก้ปัญหาความไม่สมดุลของข้อมูลของการจำแนกประเภทแบบไบนารี (The Imbalanced Binary Classification Problem) ในการทำนายความรุนแรงของผู้บาดเจ็บจากอุบัติเหตุทางถนน ในส่วนการประเมินผลแบบจำลองใช้รายงาน Classification Report เพื่อวัดประสิทธิภาพของแบบจำลองในการระบุอุบัติเหตุที่รุนแรงและไม่รุนแรงถึงการเสียชีวิตจากอุบัติเหตุทางถนนในช่วงเทศกาลปีใหม่ โดยสามารถแบ่งหัวข้อในการสรุปผลได้ดังต่อไปนี้

1. สรุปผลการวิจัย
2. อภิปรายผลการวิจัย
3. ข้อเสนอแนะ

5.1 สรุปผลการวิจัย

การทำนายความรุนแรงของผู้บาดเจ็บจากอุบัติเหตุทางถนนในช่วงเทศกาลปีใหม่ของประเทศไทยโดยใช้การเรียนรู้ของเครื่องในการสร้างแบบจำลองการทำนาย อย่างไรก็ตามการฝึกฝนแบบจำลองการทำนายที่ตรงไปตรงมามีแนวโน้มที่จะเกิดอคติกับข้อมูลฝึกฝนแบบจำลองการทำนาย เนื่องจากลักษณะที่ไม่สมดุลของข้อมูลอินสแตนซ์ในข้อมูลฝึกสอน (Class Target) ผู้บาดเจ็บจากอุบัติเหตุที่เสียชีวิตและไม่เสียชีวิต ซึ่งอุบัติเหตุทางถนนเพียงเล็กน้อยเท่านั้นที่เป็นอันตรายถึงการเสียชีวิต หากไม่มีการจัดการข้อมูลที่ไม่สมดุลอย่างเหมาะสม แบบจำลองการทำนายจะทำนายข้อมูลกลุ่มตัวอย่างส่วนมากและไม่รู้จักกลุ่มตัวอย่างของข้อมูลกลุ่มน้อย ซึ่งแน่นอนว่านี่คือสิ่งที่ต้องได้รับการแก้ไขปัญหาในงานวิจัยของเรื่องนี้ เนื่องจากค่าใช้จ่ายที่ไม่สามารถตรวจพบอุบัติเหตุที่รุนแรงถึงการเสียชีวิตนั้นสูง ซึ่งก็คือการสูญเสียชีวิต

อัลกอริทึมการเรียนรู้ของเครื่อง (Machine Learning Algorithm) ที่ใช้คือ Random Forests ซึ่งแสดงให้เห็นว่ามีประสิทธิภาพในการทำนายสูง (High Predictive Power) และใช้เทคนิคการสุ่มตัวอย่างใหม่ (Resampling Techniques) 3 แบบ เพื่อปรับความสมดุลของข้อมูล

อินสแตนซ์ในข้อมูลฝึกสอนของแบบจำลองการทำนาย (Rebalance Training Instances) ได้แก่ เทคนิคการลดจำนวนตัวอย่าง (Undersampling), เทคนิคการเพิ่มจำนวนตัวอย่างด้วย SMOTE (Oversampling with SMOTE) และการผสมระหว่างของทั้ง 2 เทคนิค ของเทคนิคการลดจำนวนตัวอย่าง (Undersampling) กับเทคนิคการเพิ่มจำนวนตัวอย่างด้วย SMOTE (Oversampling with SMOTE)

จากผลการทดลองสรุปได้ว่าแบบจำลองการทำนายของอัลกอริทึม Random Forest ที่ใช้เทคนิคของ Random Undersampling ในการทดลองครั้งนี้ให้ค่าความถูกต้องโดยรวมของการทำนายค่า Accuracy และค่า Recall เท่ากับ 83 เปอร์เซ็นต์ และผลลัพธ์ของค่าพื้นที่ใต้กราฟ ROC Curve เท่ากับ 0.92 สูงที่สุดจากแบบจำลองการทำนายทั้งหมด ซึ่งการมีค่าพื้นที่ใต้กราฟโค้ง ROC Curve สูง เป็นการบ่งชี้ถึงความถูกต้องหรือความน่าเชื่อถือของตัวแบบจำลองการทำนายที่มีประสิทธิภาพมากสุด และปัจจัยที่มีผลต่อการตัดสินใจในการทำนายมากที่สุดคือการนำส่ง (Delivering Method) การดื่มแอลกอฮอล์ (Alcohol Drinking) และอายุ (Age)

5.2 อภิปรายผลการวิจัย

งานวิจัยเรื่องนี้ได้วิเคราะห์ปัจจัยและสร้างแบบจำลองการทำนายความรุนแรงของผู้บาดเจ็บจากอุบัติเหตุทางถนนในช่วงเทศกาลปีใหม่ของประเทศไทย ซึ่งใช้ข้อมูลเปิดภาครัฐของประเทศไทย ในชุดข้อมูลผู้บาดเจ็บจากอุบัติเหตุจราจรในช่วงเทศกาลปีใหม่ 2551 ถึง 2558 โดยใช้เทคนิคการเรียนรู้แบบมีผู้สอน (Supervised Learning) ในแบบ Classification Algorithm เพื่อสร้างแบบจำลองการทำนายความรุนแรงของผู้บาดเจ็บจากอุบัติเหตุทางถนน อย่างไรก็ตามการฝึกฝนแบบจำลองการทำนายที่ตรงไปตรงมามีแนวโน้มที่จะสร้างอคติกับให้กับแบบจำลองการทำนาย เนื่องจากลักษณะที่ไม่สมดุลของข้อมูลในแง่ที่ว่าจำนวนผู้บาดเจ็บจากอุบัติเหตุที่เสียชีวิตและไม่เสียชีวิตมีความแตกต่างกันมาก อุบัติเหตุทางถนนเพียงเล็กน้อยเท่านั้นที่เป็นอันตรายถึงการเสียชีวิต หากไม่มีการจัดการข้อมูลที่ไม่สมดุลอย่างเหมาะสม แบบจำลองการทำนายจะทำนายข้อมูลกลุ่มตัวอย่างส่วนมากและไม่รู้จักกลุ่มตัวอย่างของข้อมูลกลุ่มน้อย ซึ่งก็คือแบบจำลองจะมีแนวโน้มเลือกทำนายว่าอุบัติเหตุนั้นไม่เกิดคนเสียชีวิต ในการแก้ไขปัญหาความไม่สมดุลของข้อมูลของการจำแนกประเภทแบบไบนารี ในการทำนายความรุนแรงของผู้บาดเจ็บจากอุบัติเหตุทางถนนในช่วงเทศกาลปีใหม่ 2551 ถึง 2558 ของประเทศไทย อัลกอริทึมการเรียนรู้ของเครื่อง ที่ใช้คือ Random Forests และใช้เทคนิคการสุ่มตัวอย่างใหม่ 3 แบบ เพื่อปรับความสมดุลของข้อมูลอินสแตนซ์ในข้อมูลฝึกสอนของแบบจำลองการทำนาย ได้แก่ เทคนิคการลดจำนวนตัวอย่าง (Undersampling), เทคนิคการเพิ่มจำนวนตัวอย่างด้วย SMOTE (Oversampling

with SMOTE) และการผสมระหว่างของทั้ง 2 เทคนิค ของเทคนิคการลดจำนวนตัวอย่าง (Undersampling) กับเทคนิคการเพิ่มจำนวนตัวอย่างด้วย SMOTE (Oversampling with SMOTE) ในส่วนการประเมินผลแบบจำลองใช้รายงาน Classification Report โดยที่ไม่ได้เพียงแต่มุ่งเน้นถึงความถูกต้องของการทำนายทั้งหมด (Accuracy) เท่านั้น แต่ต้องมุ่งเน้นไปที่ค่า Recall, Receiver Operating Characteristic Curve (ROC Curve) และ Precision-Recall Curve เพื่อวัดประสิทธิภาพของแบบจำลองในการระบุอุบัติเหตุที่รุนแรงถึงการเสียชีวิตจากอุบัติเหตุทางถนนในช่วงเทศกาลปีใหม่

จากผลการทดลองแบบจำลองการทำนายที่ถูกสร้างด้วยอัลกอริทึม Random Forest ที่ใช้เทคนิคของ Random Undersampling นั้นมีประสิทธิภาพในการทำนายค่าออกมาดี เมื่อเปรียบเทียบความสูญเสียของการทำนายผู้บาดเจ็บจากอุบัติเหตุ ความแตกต่างของ False Nagative ที่ทำนายผู้บาดเจ็บไม่เสียชีวิต แต่ข้อมูลจริงเสียชีวิต กับ False Positive ที่ทำนายผู้บาดเจ็บเสียชีวิต แต่ข้อมูลจริงไม่เสียชีวิต ซึ่ง Fales Negative มีความความสูญเสียมากกว่ามาก เพราะคือการเสียชีวิต ซึ่งแบบจำลองการทำนายของอัลกอริทึม Random Forest ที่ใช้เทคนิคของ Random Undersampling ในการปรับความสมดุลของข้อมูลที่ใช้ในการฝึกสอนแบบจำลองการเรียนรู้ของเครื่อง ซึ่งมีลักษณะการทำงานโดยการสุ่มข้อมูลใหม่ของข้อมูลฝึกสอนแบบจำลองการเรียนรู้ในการลดจำนวนข้อมูลของกลุ่มตัวอย่าง Majority Class ให้เท่ากับข้อมูลของกลุ่มตัวอย่าง Minority Class ให้ค่า False Nagative กับ False Positive น้อยที่สุด สามารถระบุจำนวนผู้เสียชีวิตจากอุบัติเหตุในช่วงเทศกาลปีใหม่ให้ถูกต้อง นอกจากนี้ยังใช้จำนวนข้อมูลฝึกสอนแบบจำลองน้อยลง ทำให้เวลาของการฝึกสอนแบบจำลองการทำนายใช้เวลาน้อยลง ดังนั้นจึงเหมาะกับการนำมาใช้การทำนายความรุนแรงของผู้บาดเจ็บจากอุบัติเหตุทางถนนในช่วงเทศกาลปีใหม่ 2551 ถึง 2558 ของประเทศไทย แต่ประสิทธิภาพที่ได้จากการทำนายของอัลกอริทึม Random Forest ด้วยเทคนิค Random Undersampling นั้นยังให้ผลลัพธ์ออกมาไม่ถึงกับดีที่สุด จึงควรมีการพัฒนาแบบจำลองให้มีประสิทธิภาพให้สูงขึ้น เพื่อช่วยในการทำนายความรุนแรงของผู้บาดเจ็บจากอุบัติเหตุทางถนนในช่วงเทศกาลปีใหม่นั้นมีความแม่นยำและมีประสิทธิภาพเพิ่มมากขึ้น อีกหนึ่งวิธีที่สร้างแบบจำลองการทำนายด้วยอัลกอริทึม Random Forest โดยวิธีปรับค่าพารามิเตอร์ของแบบจำลองแทนการข้อมูลสุ่มตัวอย่างใหม่ก่อนฝึกฝนแบบจำลองการทำนายด้วยค่าพารามิเตอร์ของ Class Weight เท่ากับ Balance โดยปกติอัลกอริทึม Random Forest ทำงานเหมือนอัลกอรึทึมของ Decision Tree โดยการสร้าง Bootstrapping ซึ่งการใส่ค่าพารามิเตอร์ของ Class Weight จะทำงานโดยการ Weight ตาม Class เพราะฉะนั้นจะช่วย

แก้ไขปัญหาของความไม่สมดุลของข้อมูลตัวอย่าง โดยที่ไม่ต้องใช้วิธีของ Oversampling และ Undersampling จากการทดลองฝึกสอนแบบจำลอง Random Forest โดยการใส่ Class Weight เท่ากับ Balanced ผลลัพธ์ที่ได้ใกล้เคียงกับ Random Forest โดยใช้การสุ่มตัวอย่างใหม่ของข้อมูลฝึกสอนแบบจำลองแบบ Oversampling และข้อดีกว่า Oversampling คือใช้ข้อมูลตัวอย่างเท่าเดิมและเวลาในการประมวลผลสร้างแบบจำลองน้อยกว่า Oversampling

5.3 ข้อเสนอแนะ

1. เนื่องในการวิจัยนี้ได้ใช้อัลกอริทึมเพียง 1 แบบ คือ Random Forest แล้วใช้เทคนิคเทคนิคการสุ่มตัวอย่างใหม่ (Resampling Techniques) ทั้ง 3 แบบ คือ Random Undersampling, Oversampling with SMOTE, และการผสมระหว่างของทั้งสองเทคนิคของ Undersampling กับ Oversampling with SMOTE เพื่อปรับความสมดุลของข้อมูลฝึกสอนของแบบจำลองการเรียนรู้ในจัดการปัญหาความไม่สมดุลของข้อมูล ดังนั้นอาจจะมีอัลกอริทึมและเทคนิคการปรับความสมดุลของข้อมูลฝึกสอนแบบจำลองการเรียนรู้อื่นที่สามารถเรียนรู้ข้อมูลแล้วทำนายค่าออกมาได้มีความแม่นยำและประสิทธิภาพที่ดีกว่าการวิจัยครั้งนี้

2. เนื่องจากในการวิจัยนี้ในขั้นตอนการใช้ข้อมูลนั้น เนื่องด้วยใช้ข้อมูลจากภาครัฐที่จัดเก็บข้อมูลไว้อยู่แล้ว ถ้าหากมีการเก็บข้อมูลเพิ่มเติมใส่ส่วนของ Feature ที่มีการเน้นไปที่พฤติกรรมของการขับขี่มากขึ้น ก็จะช่วยเพิ่มให้แบบจำลองสามารถระบุความรุนแรงของการเกิดอุบัติเหตุได้แม่นยำและมีประสิทธิภาพในการทำนายมากขึ้น

3. ผู้วิจัยอาจเปลี่ยนปัญหาจากไบนารี (Binary Classification) เป็น Multiple Class ได้เพื่อให้ผลการทำนายที่ละเอียดขึ้น เปลี่ยนจากข้อมูลมี “ผลการรักษา” ที่มีหลายรูปแบบ

4. ผู้วิจัยอาจเปลี่ยนจากอัลกอริทึมจาก Random Forest ไปใช้อัลกอริทึมที่อาจประมวลผลได้เร็วขึ้นและแม่นยำยิ่งกว่า เช่น อัลกอริทึม XGBoost หรือ CatBoost Classifier

5. ในอนาคตสามารถนำไปต่อยอดเกี่ยวกับการทำนายอุบัติเหตุ ซึ่งในงานวิจัยนี้สามารถการระบุพื้นที่และวัน ช่วงเวลาที่มีอัตราการเสียชีวิตจากอุบัติเหตุได้ สามารถลงไปสำรวจปัจจัยเพิ่มเติม เช่น พื้นที่นี้มีอัตราการเสียชีวิตสูงสุด อาจใช้ข้อมูลของสภาพถนนเพิ่มเติมว่าเป็นไปทิศทางเดียวกันหรือไม่ เช่น ความโค้งของถนน สภาพถนน ความขรุขระของผิวถนน ซึ่งข้อมูลเหล่านี้อาจสามารถนำไปเป็นปัจจัยช่วยในการสร้างแบบจำลองการทำนายความรุนแรงของอุบัติเหตุได้แม่นยำยิ่งขึ้น

6. ผู้วิจัยอาจทำให้ผลลัพธ์ของค่า Precision ดีขึ้นของอัลกอริทึม Random Forest ด้วยการปรับค่า Calibration ของค่า Threshold อาจช่วยให้ผลลัพธ์ของแบบจำลองมีประสิทธิภาพ

เพราะปกติค่ามาตรฐานของ Threshold เท่ากับ 0.5 เป็นการเลือกค่าสูงสุดของค่า AUC ถ้ามีการปรับเพิ่มขึ้นหรือลดลง อาจช่วยให้ค่า Precision และค่า Recall ดีขึ้นในกราฟของ ROC Curve

บรรณานุกรม

Al-Dogom, D., Aburaed, N., Al-Saad, M., และ Almansoori, S. (2019). Spatio-temporal Analysis and Machine Learning for Traffic Accidents Prediction. *2019 2nd International Conference on Signal Processing and Information Security (ICSPIS)*, 1-4.

AlMamlook, R. E., Kwayu, K. M., Alkasisbeh, M. R., และ Frefer, A. A. (2019). Comparison of Machine Learning Algorithms for Predicting Traffic Accident Severity. *2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT)*, 272-276.

Andreas C.Muller & Sarah Guido. (October, 2016). Introduction to Machine Learning with Python. Retrieved from https://www.academia.edu/34432405/Introduction_to

Antonio, M. (2019). Live Prediction of Traffic Accident Risks Using Machine Learning and Google Maps. Retrieved from http://towardsdatascience.com/live-prediction-of-traffic-accident-risks-using-machine-learning-and-google-maps-d2eeffb9389e

Chaiwuttisak, P. (2019). *Analysis of Accidental Deaths During Songkran Festival Using Data Mining.* (Doctoral dissertation, King Mongkut's Institute of Technology Ladkrabang Bangkok, Thailand). Retrieved from http://ieomsociety.org/pilsen2019/papers/329.pdf

Cuenca, L. G., Puertas, E., Aliane, N., และ Andres, J. F. (2018). Traffic Accidents Classification and Injury Severity Prediction. *2018 3rd IEEE International Conference on Intelligent Transportation Engineering (ICITE)*, 52-57.

J. Mach. Learn. Res. (2010). Scikit-learn Machine Learning in Python. Retrieved from https://scikit-learn.org/stable/modules/cross_validation.html

Jason Brownlee. (2021). Imbalanced Classification with Python. Retrieved from https://machinelearningmastery.com/smote-oversampling-for-imbalanced-classification/

Kevin Markham. (2014). Simple guide to confusion matrix terminology. Retrieved from https://www.dataschool.io/simple-guide-to-confusion-matrix-terminology/

Labib, M. F., Rifat, A. S., Hossain, M. M., Das, A. K., และ Nawrine, F. (2019). Road Accident Analysis and Prediction of Accident Severity by Using Machine Learning in Bangladesh. *2019 7th International Conference on Smart Computing & Communications (ICSCC)*, 1-5.

Lu, T., Dunyao, Z., Lixin, Y., และ Pan, Z. (2015). The traffic accident hotspot prediction: Based on the logistic regression method. *2015 International Conference on Transportation Information and Safety (ICTIS)*, 107-110.

Mamlook, R. E. A., Ali, A., Hasan, R. A., และ Kazim, H. A. M. (2019). Machine Learning to Predict the Freeway Traffic Accidents-Based Driving Simulation. *2019 IEEE National Aerospace and Electronics Conference (NAECON)*, 630-634.

Tourassi, G. (2018). Receiver Operating Characteristic Analysis: Basic Concepts and Practical Applications. *The Handbook of Medical Image Perception and Techniques*, 227-244. Retrieved from https://www.cambridge.org/core/books/handbook-of-medical-image-perception-and-techniques/receiver-operating-characteristic-analysis-basic-concepts-and-practical-applications/FF2BC4BEC0B24A799218564AF2C09D2B

Wei-Chiang Hong. (November 11, 2013). A Novel Boundary Oversampling Algorithm Based on Neighborhood Rough Set Model: NRSBoundary-SMOTE. Retrieved from https://www.hindawi.com/journals/mpe/2013/694809/#abstract

Wilson, D. (2018). Using Machine Learning to Predict Car Accident Risk. Retrieved from http://medium.com/geoai/using-machine-learning-to-predict-car-accident-risk-4d92c91a7d57

คุปตินทร, พ. (2562). การพยากรณ์การเสียชีวิตจากอุบัติเหตุทางถนนในช่วงเทศกาลปีใหม่. (ปริญญานิพนธ์ปริญญามหาบัณฑิต (Master's thesis)). สืบค้นจาก http://ethesisarchive.library.tu.ac.th/thesis/2018/TU_2018_6023036087_9543_9818.pdf

สถาบันการแพทย์ฉุกเฉินแห่งชาติ. (2558). ข้อมูลผู้บาดเจ็บและเสียชีวิตในเทศกาลปีใหม่ประจำปี 2551 - 2558. สืบค้นจาก http://data.go.th/dataset/item_7d61f508-d2e1-4f0c-8408-dfde29f111f5

องค์การอนามัยโลก World Health Organization. (2558). รายงานสถานการณ์โลกด้านความปลอดภัยทางถนน. สืบค้นจาก https://www.who.int/docs/default-source/thailand/roadsafety/2015-report-th-summary-(1).pdf?sfvrsn=ed2bec82_2

ประวัติผู้เขียน

ชื่อ-สกุล	นายเอกพันธ์ บุญเสริม
วัน เดือน ปี เกิด	10 กรกฎาคม 2536
สถานที่เกิด	ปทุมธานี
วุฒิการศึกษา	พ.ศ. 2555 วิทยาศาสตรบัณฑิต สาขาภูมิสารสนเทศศาสตร์ มหาวิทยาลัยบูรพา พ.ศ. 2561 วิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ มหาวิทยาลัยศรีนครินทรวิโรฒ
ที่อยู่ปัจจุบัน	7/3 หมู่ 2 ตำบลโผงเผง อำเภอป่าโมก จังหวัดอ่งทอง 14130