

แบบจำลองการพยากรณ์ความเสี่ยงในการเกิดอุบัติเหตุทางถนนในเทศกาลปีใหม่ ด้วยการทำเหมืองข้อมูล

A Prediction Model for Road Accident Risk in the New Year with Data Mining

ภัทธีรา สุวรรณโค¹, ดร.นิสาชล จำนงศรี² และ ผศ. ดร.จิตติมนต์ อังสกุล³

^{1,2} นักศึกษามหาบัณฑิตศึกษา สาขาวิชาเทคโนโลยีสารสนเทศ สำนักวิชาเทคโนโลยีสังคม มหาวิทยาลัยเทคโนโลยีสุรนารี

³ อาจารย์ประจำ สาขาวิชาเทคโนโลยีสารสนเทศ สำนักวิชาเทคโนโลยีสังคม มหาวิทยาลัยเทคโนโลยีสุรนารี

ABSTRACT-- Thailand's road death rate is the world's highest in 2015, most of which comes from pedestrians. This research introduces a model to predict the risk of road accidents in the New Year day. Data mining technique is used with data based on accident information, casualties and deaths collected during the 2008-2015 by the Government Information Center. By comparing the performance of Naïve Bayes Multilayer Perceptron and Meta bagging, a popular technique is used to forecast. The experiments show that Meta bagging technique is more effective than Multilayer Perceptron and Naïve Bayes.

KEY WORDS –Prediction, Data Mining, Meta Bagging, Multilayer Perceptron, Naïve Bayes.

บทคัดย่อ-- เนื่องด้วยอัตราการเสียชีวิตด้วยอุบัติเหตุทางถนนของไทยนั้นอยู่ในอันดับหนึ่งของโลกในปี 2558 ซึ่งสาเหตุส่วนใหญ่มาจากผู้ใช้นถนนโดยเฉพาะบนทางหลวงชนบท งานวิจัยนี้จึงได้สร้างแบบจำลองเพื่อพยากรณ์ความเสี่ยงในการเกิดอุบัติเหตุทางถนนในเทศกาลปีใหม่ด้วยเหมืองข้อมูล จากข้อมูลการเกิดอุบัติเหตุผู้บาดเจ็บและเสียชีวิต ในช่วงเทศกาลระหว่างปี 2551- 2558 ของศูนย์กลางข้อมูลภาครัฐ โดยการเปรียบเทียบประสิทธิภาพของการเรียนรู้ด้วยนาอ็ฟเบย์ส เพอร์เซ็ปตรอนหลายชั้น และ กูจจำแนก ซึ่งเป็นเทคนิคที่ได้รับความนิยมนำมาใช้เพื่อพยากรณ์ จากการทดลองพบว่า เทคนิคกูจจำแนก สามารถสร้างแบบจำลองที่มีประสิทธิภาพมากกว่าเพอร์เซ็ปตรอนหลายชั้น และ การเรียนรู้ด้วยนาอ็ฟเบย์ส

คำสำคัญ –การพยากรณ์, เหมืองข้อมูล, กูจจำแนก, เพอร์เซ็ปตรอนหลายชั้น, การเรียนรู้ด้วยนาอ็ฟเบย์ส.

บทนำ

ข้อมูลจากสำนักนโยบายและยุทธศาสตร์ กระทรวงสาธารณสุข รายงานอัตราการเสียชีวิตของประชากรไทยในปี 2558 (สำนักนโยบาย กระทรวงสาธารณสุข, 2558) [1] จำแนกตามสาเหตุที่สำคัญต่อประชากร 100,000 คน พบว่า คนไทยเสียชีวิตด้วยอุบัติเหตุจากคมนาคมขนส่งทางบก 22.3 คนต่อประชากรหนึ่งแสนคน รองจากการเสียชีวิตด้วยมะเร็งทุกชนิด โรคหลอดเลือดในสมอง ปอดอักเสบ และโรคหัวใจขาดเลือด ซึ่งเป็นโรคที่รักษาได้ยาก และมีความรุนแรง จึงถือว่าได้ว่าอัตราการเสียชีวิตของคนไทยยังอยู่ในระดับที่สูง ในขณะที่ค่าเฉลี่ยของทั่วโลกอยู่ที่ 18 คนต่อประชากรหนึ่งแสนคน ซึ่งเมื่อเทียบกับทั่วโลกแล้ว ในปี 2557 คนไทยเสียชีวิตด้วยอุบัติเหตุบนท้องถนนมากที่สุดเป็นอันดับสองของโลก รองจากประเทศลิเบีย (Libya) ที่มีอัตราการเสียชีวิต 36.2 คนต่อประชากรหนึ่งแสนคน (wongkarnpat, 2559) [2] ต่อมาในปี 2558 ประเทศไทยขยับขึ้นมาเป็นอันดับหนึ่ง (องค์การอนามัยโลก, 2558) [3] โดยมีอัตราคนเสียชีวิตจากอุบัติเหตุทางถนนมากที่สุดในโลก หรือเสียชีวิต 36.2 คนต่อประชากรหนึ่งแสนคน โดยในปี พ.ศ. 2558 จังหวัดที่มีการเสียชีวิตด้วยอุบัติเหตุทางถนนมากที่สุดคือ จังหวัดพระนครศรีอยุธยา มีอัตราการเสียชีวิต 21.7 คนต่อประชากรหนึ่งแสนคน รองลงมาคือ จังหวัดเพชรบุรี มีอัตราการเสียชีวิต 20.3 คนต่อประชากรหนึ่งแสนคน (หน่วยเฝ้าระวังและสะท้อนสถานการณ์ความปลอดภัยทางถนน, 2558) [4] แต่ในปี พ.ศ. 2559 จำนวนอัตราการเสียชีวิตเพิ่มขึ้น โดยจังหวัดที่มีอัตราการเสียชีวิตด้วยอุบัติเหตุทางถนนมากที่สุดคือ จังหวัดสิงห์บุรี อัตราการเสียชีวิต 30.4 คนต่อประชากรหนึ่งแสนคน รองลงมาคือ จังหวัดสระบุรี อัตราการเสียชีวิต 25.5 คนต่อประชากรหนึ่งแสนคน จังหวัดสระแก้ว อัตราการเสียชีวิต 25.8 คนต่อประชากรแสนคน จังหวัดเพชรบุรี อัตราการเสียชีวิต 23.7 คนต่อประชากรหนึ่งแสนคน และจังหวัดตราดอัตราการเสียชีวิต 23.1 คนต่อประชากรหนึ่งแสนคน (หน่วยเฝ้า

ระวังและสะท้อนสถานการณ์ความปลอดภัยทางถนน, 2559) [5] ทั้งนี้อัตราการเสียชีวิตจากอุบัติเหตุบนท้องถนนมีจำนวนสูงในช่วงเทศกาลสำคัญที่มีการเดินทางใช้ถนนเป็นจำนวนมาก โดยเฉพาะเทศกาลปีใหม่ ซึ่งข้อมูลจากศูนย์อำนวยความสะดวก (2560) [6] ได้ชี้ให้เห็นถึงปัญหาที่น่าวิตกคือความรุนแรงของอุบัติเหตุที่เกิดขึ้น จากข้อมูลที่บ่งชี้โดยตัวเลขดัชนีความรุนแรง (Severity Index) และดัชนีการเสียชีวิต (Fatality Index) ที่มีแนวโน้มเพิ่มสูงอย่างต่อเนื่อง

ปัจจุบันได้มีการทำเหมืองข้อมูลเพื่อการวิเคราะห์ข้อมูลการเกิดอุบัติเหตุบนถนนโดยการรวบรวมข้อมูลเกี่ยวกับอุบัติเหตุบนถนนจากหน่วยงานต่าง ๆ และนำข้อมูลมาวิเคราะห์เพื่อให้ได้แบบจำลองสำหรับพยากรณ์ความเสี่ยงการเกิดอุบัติเหตุบนถนน ซึ่งจากการศึกษาวรรณกรรมที่เกี่ยวข้องพบว่า มีการนำเทคนิคเหมืองข้อมูลมาใช้งานวิจัยตั้งแต่ปี 1971 เป็นต้นมา มีการศึกษาที่ทำการวิเคราะห์การเกิดอุบัติเหตุทางจราจรด้วยเทคนิควิธีที่ใช้ได้แก่ AdaBoostM1, ANN, ARM, Bayesian classifier, C4.5, C-RT, C-means, CS-MS4, Decision List, Decision Tree, ID3, K-mean algorithm, Meta classifier, Multilayer Perceptron (MLP), Naïve Bayes, Meta bagging, PART Rule classifier, Random Forest Tree classifier ซึ่งจากการศึกษาที่ผ่านมาพบว่า เทคนิคที่ได้รับความนิยมในการนำมาใช้เพื่อสร้างโมเดลการพยากรณ์ ได้แก่ Naïve Bayes Multilayer Perceptron และ Meta bagging ดังนั้น งานวิจัยนี้จึงสร้างแบบจำลองเพื่อทำนายความเสี่ยงการเกิดอุบัติเหตุบนถนนให้มีประสิทธิภาพของแบบจำลองมากยิ่งขึ้น โดยการเปรียบเทียบประสิทธิภาพของเทคนิคข้างต้น เพื่อหาเทคนิคที่ดีที่สุดในการสร้างโมเดลเพื่อพยากรณ์การเกิดอุบัติเหตุทางถนนอันจะส่งผลต่อความปลอดภัยในชีวิตและทรัพย์สินของผู้ที่สัญจรทางบกได้

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

1. เหมืองข้อมูล

เหมืองข้อมูล (Data Mining) เป็นการวิเคราะห์ข้อมูลเพื่อหารูปแบบ (patterns) หรือความสัมพันธ์

(relation) ระหว่างข้อมูลในฐานะข้อมูลขนาดใหญ่ (Linoff, 2011) [11] สารสนเทศที่ได้เอามาสร้างการพยากรณ์ หรือสร้างตัวแบบสำหรับการจำแนกหน่วยหรือกลุ่ม หรือแสดงความสัมพันธ์ระหว่างหน่วยต่าง ๆ หรือให้ข้อสรุปของสาระในการสร้างฐานข้อมูล การดำเนินงานมักอยู่ในลักษณะของการสร้างตัวแบบ (Modeling) ที่อธิบายความเป็นไปหรือสภาพการณ์หนึ่งที่เกิดขึ้น แล้วนำตัวแบบนี้มาใช้อธิบายสถานการณ์ที่ยังไม่เกิดหรือไม่ทราบคำตอบ (สุชาติ ธีระนันท์, 2554) [12]

2. เทคนิคเหมืองข้อมูล

ในปัจจุบันเทคนิคเหมืองข้อมูลมีหลายรูปแบบ ซึ่งผู้วิจัยได้ศึกษาเทคนิคที่ใช้ทำเหมืองข้อมูลเพื่อการทำนาย ได้แก่ จาก Naïve Bayes Multilayer Perceptron และ Meta Bagging ดังนี้

1. Naïve Bayes เป็นเทคนิคการสร้างโมเดลโดยอาศัยการคำนวณความน่าจะเป็นของข้อมูลต่าง ๆ (wipawan, 2017) [13] แล้วทำการวิเคราะห์ข้อมูลที่ใช้ในการสร้างแบบจำลอง ทำให้แบบจำลองสามารถเปลี่ยนแปลงได้ โดยข้อมูลในการสร้างแบบจำลองเริ่มจากศูนย์ จากหลักการนี้ทำให้แบบจำลองมีประสิทธิภาพมากขึ้น (Zhang and alt., 2013) [14] เทคนิคนี้ไม่ไวต่อจำนวนของตัวแปร และใช้ได้กับตัวแปรหลากหลาย (Kim, Hand, and Rim, 2006) [15] โดยใช้หลักการคำนวณดังสมการที่ 1 ในการเลือกตัวแปรที่ต้องการทำนาย

$$\begin{aligned} P(X|C_i) &= P(x_1|C_i) \times P(x_2|C_i) \times \dots \times P(x_n|C_i) \\ P(X|C_i) &= \prod_{k=1}^n P(x_k|C_i) \end{aligned} \quad (1)$$

เมื่อ P คือ ค่าความน่าจะเป็นของแต่ละตัวแปรตาม

X คือ แถวของข้อมูล ตัวอย่าง (X_1, X_2, \dots, X_n)

n คือ จำนวนของข้อมูล

C_i คือ ตัวแปรตามที่มีค่าเป็น ในข้อมูล

2. Multilayer Perceptron เป็นการสร้างโมเดลโดยอาศัยการคำนวณทางคณิตศาสตร์ โมเดลเป็นรูปแบบ

ของสมการ (wipawan, 2017) [13] เป็นโครงสร้างของข่ายงานระบบประสาทอย่างง่ายจึงเป็นโครงสร้าง ที่ไม่ยุ่งยากและซับซ้อนต่อผู้ใช้จึงมีผู้วิจัยให้ความสนใจศึกษากันมากทั้งแผนภูมิควบคุมสำหรับหนึ่งตัวแปรและหลายตัวแปร สำหรับโครงสร้างของข่ายงานระบบประสาทแบบ MLP สำหรับแผนภูมิควบคุมหนึ่งตัวแปร (จตุภัทร เมฆพาฬ และ กิดาพร สายชน, 2554) [16]

3. Meta Bagging เป็นการสร้างโมเดลโดยอาศัยเทคนิคของ Classifications หลาย ๆ เทคนิค เพื่อเพิ่มความถูกต้องในการจำแนกประเภทข้อมูล (wipawan, 2017) (classification) หรือคาดเดา (prediction) โดยที่เทคนิคนี้จะไปทำการลดความผันผวนของผลการทำนายของแบบจำลอง จะเห็นว่า Bagging จำเป็นต้องมีการทดลองหลาย ๆ ครั้ง เพื่อค้นหาแบบจำลองที่ดีที่สุด (Breiman, 1996) [17] ดังภาพประกอบ 1 (อมรภัทร์ หาญโคกกรวด, 2556) [18]

```

Input: training set S, number of bagging T

Bagging (T,S)
for i = 1 to T {
    S'_1 = sample from class 1 in S (with replacement)
    S'_2 = sample from class 2 in S (with replacement)
    S' = S'_1 + S'_2
    train a decision tree C_i from S'
}
Output: T classifiers
    
```

ภาพประกอบ 1 Bagging Algorithm

3. การวัดประสิทธิภาพ

การวัดประสิทธิภาพแบบจำลองพยากรณ์ความเสี่ยงในการเกิดอุบัติเหตุทางถนนในเทศกาลปีใหม่นับด้วยเหมืองข้อมูล ในด้านเหมืองข้อมูลได้ออกแบบการทดลอง โดยใช้วิธี 10-Fold cross validation วิธีนี้เป็นวิธีที่นิยมในการทำงานวิจัย เพื่อใช้ในการทดสอบประสิทธิภาพของโมเดล เนื่องจากผลที่ได้มีความน่าเชื่อถือ การวัด ประสิทธิภาพด้วยวิธี Cross-validation นี้จะทำการแบ่งข้อมูลออกเป็นหลายส่วน (มักจะแสดงด้วยค่า k) เช่น 5-fold cross-validation คือ ทำการแบ่งข้อมูลออกเป็น 5 ส่วน โดยที่แต่ละส่วนมีจำนวนข้อมูลเท่ากัน หรือ 10-fold cross-validation คือ การแบ่งข้อมูลออกเป็น 10 ส่วน โดยที่แต่ละ

ส่วนมีจำนวนข้อมูลเท่ากัน หลังจากนั้นข้อมูลหนึ่งส่วนจะ
ใช้เป็นตัวทดสอบประสิทธิภาพของโมเดล ทำวนไปเช่นนี้
จนครบจำนวนที่แบ่งไว้ (เอกสิทธิ์ พชรวงศ์ศักดิ์, 2557ข)
[19] ในแต่ละขั้นตอนหาค่าความถ่วงดุล และค่า ROC
(AUC) แล้วจึงหาค่าเฉลี่ยของค่าความถ่วงดุล และ ROC
(AUC) ซึ่งข้อมูลจะถูกทดสอบ

ค่าความถ่วงดุล (F-measure) คือ เป็นการเฉลี่ยค่า
ความแม่นยำในการตรวจพบและค่าความครบถ้วนในการ
ตรวจพบเข้าด้วยกัน จึงเปรียบเสมือนค่าวัดความแม่นยำ
โดยรวม มีสูตรในการคำนวณ ดังนี้ (เอกสิทธิ์ พชรวงศ์
ศักดิ์, 2557ก) [20]

$$F - Measure = \frac{2 \times (recall \times Precision)}{Recall + Precision}$$

ค่า ROC คือ สร้างกราฟความสัมพันธ์ระหว่าง true
positive rate (Sensitivity) กับ false positive rate (1-
Specificity) โดยการแปรค่าจุดตัด (cut-off point) ที่ใช้
ต่าง ๆ กัน (อดิพร องค์สาริต, 2560) [21] AUC เป็นการหา
พื้นที่ใต้กราฟของ ROC เพื่อใช้ในการคำนวณความถูกต้อง
ยิ่งสูงยิ่งดี (จิราวรรณ รอนราญ, 2559) [22]

4. งานวิจัยที่เกี่ยวข้อง

งานวิจัยของ Krishnaveni, and Hemalatha
(2011) [7] ได้ทำการวิเคราะห์การเกิดอุบัติเหตุทางจราจร
ด้วยเทคนิคเหมืองข้อมูล พบว่า รูปแบบการจัดหมวดหมู่
เพื่อทำนายความรุนแรงของการบาดเจ็บที่เกิดขึ้นระหว่าง
การเกิดอุบัติเหตุจราจรด้วย Naive Bayesian AdaBoostM1
Meta classifier, PART Rule classifier, J48 Decision Tree
classifier และ Random Forest Tree classifier เพื่อจำแนก
ประเภท ประเภทของความรุนแรงของการบาดเจ็บของ
อุบัติเหตุจราจรต่างๆ ผลสุดท้ายแสดงให้เห็นว่า Random
Forest ทำได้ดีกว่ามากกว่าอัลกอริทึมอื่น ๆ Gupta (2017)
[8] ได้ทำการรวบรวมงานวิจัยการวิเคราะห์ปัญหาความ
รุนแรงของอุบัติเหตุทางจราจรด้วยเทคนิคเหมืองข้อมูล
พบว่า มีการนำเทคนิคเหมืองข้อมูลมาใช้ในการวิจัยตั้งแต่

ปี 1971 เป็นต้นมา เทคนิควิธีที่ใช้ได้แก่ K-mean
algorithm, K-modes clustering, Naïve Bayes Bayesian
classifier, Decision Tree, ANN, ARM, DT, C-means, C
4.5, C-RT, CS-MS4, Decision List, ID3 และ RndTree
Liling (2017) [9] ได้ศึกษาการวิเคราะห์สถิติและ
อัลกอริทึมการทำเหมืองข้อมูลในชุดข้อมูลอุบัติเหตุ
ร้ายแรงของ FARS เพื่อแก้ปัญหา ความสัมพันธ์ระหว่าง
อัตราการเสียชีวิตกับคุณลักษณะอื่น ๆ ได้แก่ ลักษณะการ
ชน สภาพอากาศ สภาพผิว สภาพแสงและคนขับมา โดย
ใช้อัลกอริทึมอะพริโอรีในการแบ่งกลุ่ม และใช้ Naive
Bayes classified จัดแยก , และจัดกลุ่มในรูปแบบ simple
K-means clustering algorithm ในการวิเคราะห์ข้อมูล
Taamneh (2017) [10] ได้ศึกษาเรื่อง เทคนิคการทำเหมือง
ข้อมูลสำหรับการสร้างแบบจำลองการเกิดอุบัติเหตุ
การจราจรและการคาดการณ์ในสหรัฐอเมริกาอเมริกา
พบว่า มีหลายปัจจัยที่นำไปสู่การบาดเจ็บที่รุนแรงรวมถึง
ปัจจัยที่เกี่ยวข้องกับการขับรถ ปัจจัยทางถนนที่เกี่ยวข้อง
กับการเกิดอุบัติเหตุและปัจจัยที่เกี่ยวข้องอื่น ๆ โดยใช้
ต้นไม้ตัดสินใจ (DT) (J48) กฎการเหินยวน่า (Part), Naïve
Bayes (NB) และMultilayer Perceptron (MLP) ซึ่งผล
การศึกษาพบว่าความถูกต้องโดยรวมของตัวจำแนก
ประเภท DT J48 ตัวจำแนกประเภท PART class และตัว
จำแนกประเภท MLP ในการทำนายความรุนแรงของการ
บาดเจ็บอันเป็นผลมาจากอุบัติเหตุทางรถโดยใช้การ
ตรวจสอบแบบ cross-valid 10 เท่ามีความคล้ายคลึงกัน
ผลการวิจัยชี้ให้เห็นว่าปัจจัยที่สำคัญที่สุดที่เกี่ยวข้องกับ
ความรุนแรงของชีวิต ได้แก่ อายุ เพศ สัญชาติ ปีที่เกิด
อุบัติเหตุ สถานะความเสียหาย และประเภทการปะทะกัน
กลุ่มอายุ 18 ถึง 30 ปีเป็นกลุ่มอายุที่เสี่ยงที่สุดที่จะเกิด
อุบัติเหตุทางจราจร มีแนวโน้มที่ชัดเจนในการลดอุบัติเหตุ
ในช่วงเวลาเรียน ผู้ขับขี่มีส่วนเกี่ยวข้องกับอุบัติเหตุจราจร

มากกว่าผู้โดยสารและคนเดินเท้า ผู้ขับขี่ชายมีส่วนร่วมในอุบัติเหตุจราจรมากกว่าผู้ขับขี่หญิงมากขึ้น

วิธีการดำเนินการวิจัย

แบบจำลองพยากรณ์ความเสี่ยงในการเกิดอุบัติเหตุทางถนนในเทศกาลปีใหม่ด้วยเหมืองข้อมูล มีวิธีการดำเนินการวิจัย ดังนี้

1. การเตรียมข้อมูล

ข้อมูลในการวิจัยครั้งนี้ได้มาจาก ศูนย์กลางข้อมูลภาครัฐ มีลักษณะเป็นข้อมูลเปิด ซึ่งมีขั้นตอนการเตรียมข้อมูลดังนี้

1. ทำการรวบรวมข้อมูลการเกิดอุบัติเหตุ ผู้บาดเจ็บและเสียชีวิต ในช่วงเทศกาลระหว่างปี 2551-2558 จากเว็บไซต์ Data.go.th หรือศูนย์กลางข้อมูลภาครัฐ ซึ่งอยู่ในรูปแบบของไฟล์คอมพิวเตอร์ (.xls) จำนวน 214,951 รายการ ประกอบด้วย 19 แอดทริบิวต์ ประกอบด้วย ชื่อเทศกาล รหัสจังหวัด จังหวัด รหัสโรงพยาบาล ชื่อโรงพยาบาลที่รับผู้บาดเจ็บ วันที่เกิดเหตุ เวลาเกิดเหตุ เพศ อายุ ถนนที่เกิดเหตุ สถานะ รถผู้บาดเจ็บ รถคู่กรณี มาตรการ การดื่มสุรา การนำส่ง Refer-Admit ผลการรักษา จำนวนวันรักษา

2. ทำการเตรียมชุดข้อมูลเพื่อนำเข้าสู่โปรแกรม Weka จากข้อมูล จำนวน 214,951 รายการ ประกอบด้วย 19 แอดทริบิวต์ ผู้วิจัยได้ทำการตัดส่วนที่ไม่จำเป็นออก เหลือเฉพาะข้อมูลที่ใช้ในการพยากรณ์ในภาพรวม จำนวน 10 แอดทริบิวต์ ได้แก่ เพศ อายุ เวลาที่เกิดเหตุ ถนนที่เกิดเหตุ สถานะ รถผู้บาดเจ็บ รถคู่กรณี มาตรการ การดื่มสุรา และ ผลการรักษา โดยใช้ข้อมูลในปี 2551 – 2557 จำนวน 190,834 รายการ เป็น Training Set ใช้ข้อมูลในปี 2558 จำนวน 24,118 รายการ เป็น Test Set

ข้อมูลในการวิจัยฉบับนี้สามารถแบ่งออกเป็น 4 classes จากแอทริบิวต์ จุดเกิดเหตุ คือ ไม่ทราบ อัตราส่วนเท่ากับ 2.43 ในเมือง อัตราส่วนเท่ากับ 16.96 ทางหลวง อัตราส่วนเท่ากับ 24.62 ชนบท อัตราส่วนเท่ากับ 41.46 สรุปได้ดังตาราง 1

ตาราง 1 คลาสในการวิจัย

จุดเกิดเหตุ	จำนวน	ร้อยละ
ไม่ทราบ	5418	2.43
ในเมือง	37878	16.96
ทางหลวง	54970	24.62
ชนบท	92567	41.46
รวม	223293	100.00

ประกอบด้วยตัวแปรทั้งหมด 10 ตัวแปร ดังตาราง 2

ตาราง 2 ตัวแปรในการวิจัย

ลำดับ	Attributes	Description	Values
1	Gender	เพศ	1=หญิง 2=ชาย
2	Age	อายุ	Actual age

ลำดับ	Attributes	Description	Values
3	Time	เวลาที่เกิดเหตุ	1=01.01-02.00น. 2=02.01-03.00น. 3=03.01-04.00น. 4=04.01-05.00น. 5=05.01-06.00น. 6=06.01-07.00น. 7=07.01-08.00น. 8=08.01-09.00น. 9=09.01-10.00น. 10=10.01-11.00น. 11=11.01-12.00น. 12=12.01-13.00น. 13=13.01-14.00น. 14=14.01-15.00น. 15=15.01-16.00น. 16=16.01-17.00น. 17=17.01-18.00น. 18=18.01-19.00น. 19=19.01-20.00น. 20=20.01-21.00น. 21=21.01-22.00น. 22=22.01-23.00น. 23=23.01-24.00น. 24=24.01-00.00น.
4	Scene	จุดเกิดเหตุ	0=ไม่ทราบ 1=ในเมือง 2=ทางหลวง 3=ชนบท
5	Status	สถานะ	1=คนเดินเท้า 2=ผู้โดยสาร 3=ผู้ขับขี่
6	Vehicle Type	ยานพาหนะ	0=ไม่มี/ตัวเอง 1=จักรยานยนต์ 2=ปิกอัพ 3=รถเก๋ง/แท็กซี่

			4=รถโดยสาร 4 ล้อ 5= รถโดยสารขนาดใหญ่ 6=รถจักรยาน 7=รถตู้ 8=รถบรรทุก 9 =สามล้อเครื่อง 10= สามล้อถีบ 11=อื่น ๆ
7	Parties Vehicle	คู่กรณี	1=จักรยานยนต์ 2=ปิกอัพ 3=รถเก๋ง/แท็กซี่ 4=รถโดยสาร 4 ล้อ 5= รถโดยสารขนาดใหญ่ 6=รถจักรยาน 7=รถตู้ 8=รถบรรทุก 9 =สามล้อเครื่อง 10= สามล้อถีบ 11=อื่น ๆ
8	Measure	มาตรการ	0= ไม่ใส่หมวก 1=ใส่หมวก 3=คาดเข็มขัด
9	Drinking	การดื่มสุรา	0=ไม่ดื่ม 1=ดื่ม
10	Severity	ผลการรักษา	24 Hrs After 24 Hrs Immediately ER-Room Transferring Get Well

2. การสร้างแบบจำลอง

ในการสร้างแบบจำลอง งานวิจัยนี้ใช้โปรแกรม Weka 3.8.1 ซึ่งเป็นเครื่องมือที่มีเทคนิคในการทำเหมืองข้อมูลหลายเทคนิค และมีการวัดผลที่มีประสิทธิภาพของแบบจำลอง งานวิจัยนี้ได้เลือกเทคนิควิธีดังต่อไปนี้

1) Naïve Bayes 2) Multilayer Perceptron 3) Meta Bagging ซึ่งกำหนดให้ทำ 10 รอบ

3. การวิเคราะห์ประสิทธิภาพ

การวิเคราะห์ประสิทธิภาพแบบจำลองพยากรณ์ความเสี่ยงในการเกิดอุบัติเหตุทางถนนในเทศบาลปีใหม่ด้วยเหมืองข้อมูล วัดประสิทธิภาพของแบบจำลองโดยวัดค่าความถ่วงดุล และพื้นที่ใต้เส้นกราฟ ROC(AUC) ทำให้ทราบค่าความถูกต้องของวิธีการที่ทำการวิจัย และพัฒนาตรงตามวัตถุประสงค์ที่ตั้งไว้

ผลการดำเนินงานวิจัย

ผลการดำเนินงานวิจัยฉบับนี้ได้วิเคราะห์ถึงค่าแม่นยำและค่า AUC ของแบบจำลองพยากรณ์ความเสี่ยงในการเกิดอุบัติเหตุทางถนนในเทศบาลปีใหม่ด้วยเหมืองข้อมูล ในการวิเคราะห์ได้ใช้ขั้นตอนการทดลองจากการทำเหมืองข้อมูล โดยใช้ข้อมูลจากศูนย์กลางข้อมูลภาครัฐระหว่างปี 2550 – 2558 ใช้หลักการแยกด้วยวิธีการ 10-fold cross validation เทคนิคที่นำมาใช้ในการสร้างแบบจำลองคือ

1) Naïve Bayes 2) Multilayer Perceptron 3) Meta Bagging

1. การวัดประสิทธิภาพโดยวิธีการหาค่าความถ่วงดุล ในการพยากรณ์ความเสี่ยงในการเกิดอุบัติเหตุทางถนนในเทศบาลปีใหม่ด้วยเหมืองข้อมูลเป็นพื้นฐานการวัด

ประสิทธิภาพ และประสิทธิผลของแบบจำลอง ซึ่งสามารถแสดงค่าความถ่วงดุลของการพยากรณ์ได้ดัง ตาราง 3

ตาราง 3 ค่าความถ่วงดุลของการพยากรณ์

แบบจำลอง	ค่าความถ่วงดุล(ร้อยละ)
Naïve Bayes	33.9
Multilayer Perceptron	97.5
Meta Bagging	97.5

จากตาราง 3 แสดงถึงค่าความถ่วงดุล(F-Measure) ในการพยากรณ์ความเสี่ยงในการเกิดอุบัติเหตุทางถนนในเทศบาลปีใหม่ด้วยเหมืองข้อมูล จากผลการทดลองปรากฏว่า ค่าประสิทธิภาพจาก Multilayer Perceptron และ Meta Bagging มีค่าความถ่วงดุลเท่ากันโดยคิดเป็นร้อยละ 97.5 ส่วน Naïve Bayes มีค่าความถ่วงดุลน้อยกว่า คิดเป็นร้อยละ 33.9

2. การวัดประสิทธิภาพโดยวิธีการหาพื้นที่ใต้เส้นกราฟ ROC(AUC) ในการพยากรณ์ความเสี่ยงในการเกิดอุบัติเหตุทางถนนในเทศบาลปีใหม่ด้วยเหมืองข้อมูล เป็นพื้นฐานการวัดประสิทธิภาพ และประสิทธิผลของแบบจำลอง สามารถแสดงผลค่า AUC ได้ ดังตาราง 4

ตาราง 4 ค่าพื้นที่ใต้เส้นกราฟของการพยากรณ์

แบบจำลอง	ค่าพื้นที่ใต้เส้นกราฟ(ร้อยละ)
Naïve Bayes	50.5
Multilayer Perceptron	72.4
Meta Bagging	77.3

จากตาราง 4 แสดงค่าพื้นที่ใต้เส้นกราฟของการพยากรณ์ความเสี่ยงในการเกิดอุบัติเหตุทางถนนในเทศบาลปีใหม่ด้วยเหมืองข้อมูลเป็นพื้นฐานการวัดประสิทธิภาพ จากผลการทดลอง ปรากฏว่า Meta Bagging ให้ค่าประสิทธิภาพมากที่สุด โดยคิดเป็นร้อยละ 77.3 รองลงมาคือ

Multilayer Perceptron คิดเป็นร้อยละ 72.4 และ Naïve Bayes ให้ค่าประสิทธิภาพน้อยที่สุด โดยคิดเป็นร้อยละ 50.5

สรุปผลการวิจัย

จากจุดประสงค์ในการวิจัยครั้งนี้ ผู้วิจัยได้ทำการทดลองแบบจำลองให้เหมาะสมกับชุดข้อมูลการเกิดอุบัติเหตุทางถนนในช่วงเทศกาลปีใหม่ โดยใช้เทคนิค Naïve Bayes Multilayer Perceptron และ Meta Bagging ซึ่งสามารถนำเสนอผลได้ดังนี้ การสร้างแบบจำลองเหมืองข้อมูล โดยใช้ Multilayer Perceptron และ Meta Bagging มีค่าความถ่วงดุลเท่ากัน โดยคิดเป็นร้อยละ 97.5 ซึ่งถือว่ามากที่สุด แต่เมื่อพิจารณาค่า AUC ซึ่งใช้ประกอบการประเมินประสิทธิภาพ พบว่า Meta Bagging ให้ค่าประสิทธิภาพมากที่สุด โดยคิดเป็นร้อยละ 77.3 ซึ่งสามารถสรุปได้ว่าเทคนิค Meta Bagging นี้มีประสิทธิภาพในการพยากรณ์มากที่สุด ทั้งนี้สามารถอภิปรายผลได้ว่า เนื่องจากชุดข้อมูลมีปริมาณที่ค่อนข้างมาก และเทคนิควิธีที่เลือกมานั้น Naïve Bayes ไม่เหมาะกับจำนวนของตัวแปรมาก ๆ แม้จะใช้ได้กับตัวแปรทุกชนิด (Kim, Han, and Rim, 2006)[23] ในขณะที่ Multilayer Perceptron และ Meta Bagging สามารถทำงานกับชุดตัวแปรปริมาณมาก ๆ ได้ ถึงแม้ว่าสองเทคนิคนี้จะมีค่าประสิทธิภาพใกล้เคียงกัน แต่ Meta Bagging มีค่าประสิทธิภาพมากกว่า เนื่องด้วย Meta Bagging สร้างโมเดลโดยอาศัยเทคนิคการจำแนกข้อมูลหลาย ๆ เทคนิค เพื่อเพิ่มความถูกต้อง (wipawan, 2017) [13] และมีการทดลองหลายๆ ครั้ง เพื่อค้นหาแบบจำลองที่ดีที่สุด (Breiman, 1996) [17] ดังนั้น แบบจำลองพยากรณ์ความเสี่ยงในการเกิดอุบัติเหตุทางถนนในเทศกาลปีใหม่ ด้วยเหมืองข้อมูลโดยใช้เทคนิค Meta Bagging จึงมีความเหมาะสมในการนำไปใช้ในการพยากรณ์ต่อไป

เอกสารอ้างอิง

- [1] สำนักนโยบาย กระทรวงสาธารณสุข (2558). รายงานสถิติสาธารณสุข สำนักนโยบายและยุทธศาสตร์ กระทรวงสาธารณสุข. จำนวนและอัตราตายต่อประชากร 100,000 คน จำแนกตามสาเหตุที่สำคัญ ปี พ.ศ. 2537 - 2558 รายปี. Retrieved from http://social.nesdb.go.th/SocialStat/StatReport_Fin al.aspx?reportid=367&template=1R2C& yeartype=M&subcatid=15
- [2] wongkarnpat. (2559). อัตราการเสียชีวิตของชาวไทย บนท้องถนน. นิตยสารวงการแพทย์. Retrieved from <http://www.wongkarnpat.com/viewpat.php?id=21 69#.WezUR2i0M2y>
- [3] องค์การอนามัยโลก. (2558). รายงานสถานการณ์โลก ด้านความปลอดภัยทางถนน พ.ศ. 2558. Retrieved from http://www.who.int/violence_injury_prevention/ro ad_safety_status/2015/GSRRS2015_Summary_Th ai.pdf?ua=1
- [4] หน่วยเฝ้าระวังและสะท้อนสถานการณ์ความปลอดภัยทางถนน. (2558). อัตราการเสียชีวิตจากอุบัติเหตุทางถนน ต่อ ประชากร 100,000 คน แยกราย จังหวัด. Retrieved from <http://trso.thairoads.org/statistic/watch/detail/137>
- [5] หน่วยเฝ้าระวังและสะท้อนสถานการณ์ความปลอดภัยทางถนน. (2559). อัตราการเสียชีวิตจากอุบัติเหตุทางถนน ต่อ ประชากร 100,000 คน แยกราย จังหวัด. Retrieved from <http://trso.thairoads.org/statistic/watch/detail/137>
- [6] ศูนย์อำนวยการความปลอดภัยทางถนน. (2560). สรุป รายงานผลการปฏิบัติงานป้องกันและลดอุบัติเหตุทางถนนในช่วงเทศกาลปีใหม่ พ.ศ. 2560. Retrieved from

- http://122.155.1.141/upload/minisite/file_attach/196/594b37bb52a68.zip
- [7] S.Krishnaveni, and M.Hemalatha. (2011). A Perspective Analysis of Traffic Accident using Data Mining Techniques. *International Journal of Computer Applications, Volume 23*(7, June 2011), 40-48.
- [8] Meenu Gupta, Vijender Kumar Solanki, and Vijay Kumar Singh. (2017). Analysis of Datamining Technique for Traffic Accident Severity Problem: A Review. *Proceedings of the Second International Conference on Research in Intelligent and Computing in Engineering, Vol. 10*, 197–199. doi:10.15439/2017R121
- [9] Liling Li, Sharad Shrestha, and Gongzhu Hu. (2017). Analysis of road traffic fatal accidents using data mining techniques T2. *IEEE 15th International Conference on Software Engineering Research, Management and Applications (SERA)*. doi:10.1109/SERA.2017.7965753
- [10] Madhar Taamneh, Sharaf Alkheder and Salah Taameh. (2017). Data-mining techniques for traffic accident modeling and prediction in the United Arab Emirates. *Journal of Transportation Safety & Security*, 9, 2017 (2). doi:19439962.2016.1152338
- [11] Gordon S. Linoff, Michael J.A.Berry. (2011). *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management* (4 ed.). IN: Wiley Publishing.
- [12] สุชาดา กิระนันท์. (2554). คำถามคำตอบเกี่ยวกับการทำเหมืองข้อมูลเบื้องต้น. Retrieved from https://home.kku.ac.th/wichuda/Knowledge/6DataMining/Datamining_Suchada.pdf
- [13] wipawan. (2017). data mining. *CLASSIFICATION STEPS* (1). Retrieved from https://wipawanblog.files.wordpress.com/2013/08/lab_datamining.pdf
- [14] Jun Zhang and alt. (2013). Internet Traffic Classification by Aggregating Correlated Naive Bayes Predictions. *IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY*, 8(1), 5-15.
- [15] Sang Bum Kim, Kyoung Soo Han, and Hae Chang Rim. (2006). *Some effective techniques for naive bayes text classification*. *IEEE Transactions on Knowledge and Data Engineering*, 18(11), 1457-1466.
- [16] จตุภัทร เมฆพชัย และ กิดาการ สายชน. (2554). สมรรถนะของข่ายงานระบบประสาทแบบ Multi-Layer Perceptron และ Radial Basis Function สำหรับแผนภูมิควบคุมคุณภาพหลายตัวแปร. *วารสารวิทยาศาสตร์บูรพา*, 16(2554)(2), 97-106.
- [17] Breiman, Leo. (1996). Bagging Predictors. *Machine Learning*, 24(2), 123-140.
- [18] อมรภัทร์ หาญโลกกรวด และคนอื่น ๆ. (2556). แบบจำลองทำนายความเสี่ยงการเกิดอุบัติเหตุบนท้องถนนโดยเทคนิคถ่วงจำแนก. *วารสารวิทยาศาสตร์และเทคโนโลยีมหาวิทยาลัยมหาสารคาม*, 9, 253-262.
- [19] เอกสิทธิ์ พัทธวงศ์ศักดิ์. (2557ข). การวิเคราะห์ข้อมูลด้วยเทคนิคค้ำไม่มันนิ่งเบื้องต้น. Retrieved from http://dataminingtrend.com/2014/wp-content/uploads/2014/08/intro_data_mining_previ ew.pdf
- [20] เอกสิทธิ์ พัทธวงศ์ศักดิ์. (2557ก). การแบ่งข้อมูลเพื่อนำทดสอบประสิทธิภาพของโมเดล. *Data Mining Trend*. Retrieved from

- <http://dataminingtrend.com/2014/data-mining-techniques/cross-validation/>
- [21] อติพร องค์สำริด. (2560). เอกสารประกอบการสอน
หลักการพิจารณางานวิจัยเกี่ยวกับการตรวจ
วินิจฉัยมาประยุกต์ในเวชปฏิบัติ. Retrieved from
http://med.mahidol.ac.th/fammed/sites/default/files/public/pdf/EBM_Diagnostic_study.pdf
- [22] จีราวรรณ รอนราญ. (2559). การทำงานในรูปแบบ
Data Mining. Retrieved from
<http://www.erp.mju.ac.th/articleDetail.aspx?qid=502>
- [23] Sang Bum Kim, Kyoung Soo Han, and Hae Chang
Rim. (2006). Some effective techniques for naive
bayes text classification. *IEEE Transactions on
Knowledge and Data Engineering*, 18(11), 1457-
1466.