

ปัจจัยที่มีอิทธิพลต่อการฝ่าฝืนไฟสัญญาณของคนเดินข้ามที่ทางข้ามถนน
อัจฉริยะในประเทศไทย



นางสาวณัชชา คำมูล

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมโยธา ขนส่ง และทรัพยากรธรณี
มหาวิทยาลัยเทคโนโลยีสุรนารี
ปีการศึกษา 2567

FACTORS INFLUENCING PEDESTRIAN SIGNAL VIOLATIONS
AT MID-BLOCK SMART CROSSWALKS IN THAILAND



CHATCHA KUMMOON

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of
Master of Engineering in Civil Transportation and Geo-resources Engineering

Suranaree University of Technology

Academic Year 2024

ปัจจัยที่มีอิทธิพลต่อการฝ่าฝืนไฟสัญญาณของคนเดินข้ามที่ทางข้ามถนนอัจฉริยะ
ในประเทศไทย

มหาวิทยาลัยเทคโนโลยีสุรนารี อนุมัติให้บัณฑิตวิทยาลัยเป็นหน่วยงานหนึ่งของการศึกษา
ตามหลักสูตรปรัชญามหาบัณฑิต

คณะกรรมาการสอบวิทยานิพนธ์

(ผศ. ร.อ. ดร.สุทธิพงษ์ มียะ)

ประธานกรรมการ

(ผศ. ดร.รัฐพล ภู่บวบพาพันธ์)

กรรมการ (อาจารย์ที่ปรึกษาวิทยานิพนธ์)

(ศ. ดร.ธเนศ เสถียรนาม)

กรรมการ

(รศ. ดร.ยุพาพร รักสกุลพิวัฒน์)

รองอธิการบดีฝ่ายวิชาการและประกันคุณภาพ

(รศ. ดร.พรศิริ จงกล)

คณบดีสำนักวิชาวิศวกรรมศาสตร์

ณัชชา คำมูล : ปัจจัยที่มีอิทธิพลต่อการฝ่าฝืนไฟสัญญาณของคนเดินข้ามที่ทางข้ามถนน
อัจฉริยะในประเทศไทย (FACTORS INFLUENCING PEDESTRIAN SIGNAL VIOLATIONS
AT MID-BLOCK SMART CROSSWALKS IN THAILAND)

อาจารย์ที่ปรึกษา : ผู้ช่วยศาสตราจารย์ ดร.รัฐพล ภูบุบผาพันธ์, 89 หน้า.

คำสำคัญ : คนเดินข้าม/ความปลอดภัยของคนเดินข้าม/การฝ่าฝืนไฟสัญญาณ/ทางข้ามอัจฉริยะ/
การถดถอยโลจิสติกแบบทวิ/ต้นไม้ตัดสินใจ/ข้อมูลไม่สมดุล

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาปัจจัยที่มีอิทธิพลต่อการฝ่าฝืนไฟสัญญาณของคนเดินข้ามที่ทางข้ามถนนอัจฉริยะ (Smart Crosswalk) ในประเทศไทย ซึ่งหากทำความเข้าใจถึงปัจจัยต่าง ๆ ได้อย่างชัดเจนแล้ว จะสามารถนำไปสร้างคำแนะนำหรือมาตรการในการจัดการกับการฝ่าฝืนไฟสัญญาณของคนเดินข้าม เพื่อบรรเทาอุบัติเหตุของคนเดินข้ามที่ทางข้ามถนนอัจฉริยะทั่วประเทศได้ การศึกษานี้ดำเนินการสร้างแบบจำลองโดยใช้วิธีการถดถอยโลจิสติกแบบทวิ (Binary logistic regression) และวิธีต้นไม้ตัดสินใจ (Decision tree) เพื่อตรวจสอบปัจจัยลักษณะทางกายภาพของถนนและปัจจัยข้อมูลพฤติกรรมจากระบบทางข้ามถนนอัจฉริยะ ซึ่งตัวแปรตามที่ใช้ในการศึกษารั้งนี้แบบไบนารี ระบุเหตุการณ์ว่ามีการฝ่าฝืนไฟสัญญาณของคนเดินข้ามหรือไม่ในแต่ละรอบไฟสัญญาณจากการวิเคราะห์เบื้องต้นพบว่าจำนวนเหตุการณ์ที่พบการฝ่าฝืนไฟสัญญาณน้อยกว่าเหตุการณ์ที่ไม่พบการฝ่าฝืนไฟสัญญาณของคนเดินข้ามเป็นจำนวนมาก บ่งชี้ว่าข้อมูลในการศึกษานี้ไม่สมดุล (Imbalanced Data) ดังนั้น จึงมีการตรวจสอบเทคนิคการปรับขนาดชุดข้อมูลเพื่อจัดการกับข้อมูลที่ไม่สมดุล ผลลัพธ์การปรับขนาดชุดข้อมูลจากทั้ง 4 วิธีของทั้ง 2 แบบจำลอง พบว่าการสุ่มตัวอย่างแบบ Downsampling มีประสิทธิภาพสูงที่สุดในการพยากรณ์ตัวแปรตาม ผลลัพธ์วิธีการถดถอยโลจิสติกแบบทวิ (Binary logistic regression) เผยว่าระยะเวลาการรอคอยและช่วงเวลาของวันมีอิทธิพลในเชิงบวก และในขณะเดียวกันความยาวของทางข้าม ความเร็วของยานพาหนะ และปริมาณจราจรที่สัญจรผ่านทางข้าม มีอิทธิพลในเชิงลบต่อการฝ่าฝืนไฟสัญญาณของคนเดินข้าม สอดคล้องกับการสังเกตภาคสนามที่เผยให้เห็นว่า การฝ่าฝืนไฟสัญญาณของคนเดินข้าม มักจะเกิดขึ้นบ่อยในเวลากลางคืนเมื่อมีปริมาณจราจรที่สัญจรผ่านทางข้ามน้อย และคนเดินถนนต้องรอคอยไฟสัญญาณเป็นเวลานานหลังจากกดปุ่ม และผลลัพธ์วิธีต้นไม้ตัดสินใจ (Decision tree) เผยว่ามีตัวแปรอิสระเพียง 3 ตัวแปร ที่มีอิทธิพลต่อการฝ่าฝืนไฟสัญญาณของคนเดินข้าม ได้แก่ ระยะเวลาการรอคอย ความยาวของทางข้าม และปริมาณจราจรที่สัญจรผ่านทางข้าม ซึ่งเมื่อพิจารณาจากเกณฑ์ที่ได้จากแบบจำลอง พบว่าปัจจัยด้านระยะเวลาการรอคอยและปริมาณจราจรที่สัญจรผ่านทางข้าม มีความสมเหตุสมผลกับการสังเกตภาคสนาม อย่างไรก็ตาม ปัจจัยด้านความยาวของทางข้ามในบางกฎเกณฑ์แสดงผลที่ไม่

สมเหตุสมผลกับการสังเกตภาคสนาม ซึ่งอาจเกิดจากข้อจำกัดของวิธีต้นไม้ตัดสินใจหรืออาจเกิดจากข้อจำกัดของความหลากหลายของข้อมูลที่ยังไม่เพียงพอ จึงอาจไม่สะท้อนลักษณะที่แท้จริงของปัจจัยดังกล่าวได้อย่างครบถ้วน จากผลการวิเคราะห์ดังกล่าวผู้วิจัยจึงเสนอแนะว่า ในบริบทของข้อมูลลักษณะนี้ควรเลือกใช้วิธีถดถอยโลจิสติกแบบทวิ เนื่องจากแม้ใช้ชุดข้อมูลเดียวกัน วิธีดังกล่าวสามารถระบุปัจจัยที่มีความสัมพันธ์อย่างมีนัยสำคัญต่อพฤติกรรมการฝ่าฝืนไฟสัญญาณได้อย่างสมเหตุสมผล และสอดคล้องกับการสังเกตภาคสนามมากกว่าแบบจำลองต้นไม้ตัดสินใจ



สาขาวิชา วิศวกรรมขนส่ง
ปีการศึกษา 2567

ลายมือชื่อนักศึกษา.....*ศรัทธา คำมูล*
ลายมือชื่ออาจารย์ที่ปรึกษา.....*[Signature]*

CHATCHA KUMMOON : FACTORS INFLUENCING PEDESTRIAN SIGNAL VIOLATIONS
AT MID-BLOCK SMART CROSSWALKS IN THAILAND. THESIS ADVISOR : ASST.
PROF. RATTAPHOL PUEBOOBPAPHAN, Ph.D., 89 PP.

Keyword : PEDESTRIANS/PEDESTRIANS SAFETY/SIGNAL VIOLATION/SMART CROSSWALK/
BINARY LOGISTIC REGRESSION/DECISION TREE/IMBALANCED DATA

This study examines the factors that influence instances of pedestrians disobeying traffic signals at mid-block crossings that are equipped with an intelligent crossing device, known as the Smart Crosswalk in Thailand. Once fully comprehended, it is postulated that recommendations or interventions to address pedestrian signal violations can be formulated to mitigate pedestrian accidents at crossings nationwide. Binary logistic regression and Decision tree are utilized to create models that examine the physical road characteristics, as well as secondary data from Smart crosswalk systems. The dependent variable employed in this study is binary, indicating whether pedestrian signal is violated for each signal cycle. Our preliminary analysis of pedestrian violations at smart crosswalks indicates that the data in this study is imbalanced, i.e., the number of events that pedestrian violations are observed during any signal cycle is far less than the events of no violation. Therefore, some scaling techniques to deal with imbalanced data are investigated. According to the dataset scaling results, out of all four methods, Downsampling technique had the highest area under the curve (AUC) value for both models. The results of binary logistic regression showed a positive correlation between pedestrian violations and both waiting time and time of day, while there was a negative correlation between pedestrian violations and crossing length, traffic speed, and traffic volume. This is consistent with field observation, which reveals that violations are more likely at night when traffic volumes are low and signal waiting times are long. The results of the decision tree analysis revealed that waiting time, crosswalk length and the volume of traffic passing through the crosswalk were the independent variables influenced pedestrian traffic signal violations. Upon examining the rules derived from the model, it was found that the factors of waiting time and traffic volume passing through the crosswalk were consistent with field observations.

However, some rules related to crosswalk length showed results that were inconsistent with field observations. This inconsistency may be due to limitations of the methodology, or the insufficient variability in the dataset, which might prevent these factors from fully representative of their true effects. Based on the above analysis, the researcher recommends that, in the context of this dataset, binary logistic regression should be preferred. Although applied to the same dataset, this method can reasonably identify factors that have a statistically significant association with pedestrian traffic signal violations and demonstrates greater consistency with field observations compared to the decision tree model.



School of Transportation Engineering
Academic Year 2024

Student's Signature.....ณัฏฐา คำมูล.....
Advisor's Signature.....[Signature].....

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงด้วยดีอย่างสมบูรณ์ ผู้วิจัยขอกราบขอบพระคุณ บุคคลและกลุ่มบุคคลต่าง ๆ ที่ได้ให้ความเมตตา ช่วยเหลือดูแล ให้คำปรึกษาและคำชี้แนะต่าง ๆ ทั้งในด้านวิชาการและการดำเนินงานวิจัย

ขอกราบขอบพระคุณ ผู้ช่วยศาสตราจารย์ ดร. รัฐพล ภูบุบผาพันธ์ อาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่ให้ความรู้ทั้งในด้านวิชาการและการดำเนินงานวิจัย คอยให้คำแนะนำและช่วยเหลือในการเขียน การตรวจแก้ไขวิทยานิพนธ์หรือบทความวิจัย และสนับสนุนค่าใช้จ่ายต่าง ๆ ในการวิจัยจนเสร็จสมบูรณ์

กราบขอบพระคุณสำนักอำนวยความสะดวก กรมทางหลวง และหน่วยงานอื่น ๆ ที่เกี่ยวข้องที่อนุเคราะห์ให้ข้อมูลการทำงานของระบบทางข้ามอัจฉริยะ และข้อมูลลักษณะกายภาพถนน เพื่อนำมาใช้ในการวิเคราะห์

คุณปิยธิดา ปัญหาราช เลขานุการ สาขาวิศวกรรมขนส่ง มหาวิทยาลัยเทคโนโลยีสุรนารี ที่ให้ความช่วยเหลือในการประสานงานด้านเอกสารต่าง ๆ ระหว่างการศึกษาและการจัดทำวิทยานิพนธ์

ขอขอบคุณเพื่อนบัณฑิตทุกท่าน ที่ให้คำปรึกษา คำแนะนำต่าง ๆ ตลอดจนการแลกเปลี่ยนความรู้ การสนับสนุนและการให้กำลังใจเสมอมา

ขอขอบคุณ กองทุนสนับสนุนการวิจัยและพัฒนา มหาวิทยาลัยเทคโนโลยีสุรนารี ที่ผู้ช่วยศาสตราจารย์ ดร. รัฐพล ภูบุบผาพันธ์ ได้รับทุนวิจัยจากแหล่งทุนภายนอก (OROG) ได้ให้โอกาสมอบทุนการศึกษาสนับสนุนในการวิจัยนี้

สุดท้ายนี้ขอกราบขอบพระคุณ บิดา มารดา ที่ให้การเลี้ยงดู อบรม พร้อมทั้งส่งเสริมการศึกษาเป็นอย่างดีมาตลอด และเป็นกำลังใจให้ข้าพเจ้าจนกระทั่งสามารถดำเนินการทำวิทยานิพนธ์ฉบับนี้จนสำเร็จสมบูรณ์

ณัชชา คำมูล

สารบัญ

หน้า

บทคัดย่อ (ภาษาไทย).....	ก
บทคัดย่อ (ภาษาอังกฤษ).....	ค
กิตติกรรมประกาศ.....	จ
สารบัญ.....	ฉ
สารบัญตาราง.....	ซ
สารบัญรูป.....	ญ

บทที่

1 บทนำ.....	1
1.1 ที่มาและความสำคัญของงานวิจัย.....	1
1.2 วัตถุประสงค์ของการวิจัย.....	4
1.3 คำถามงานวิจัย.....	4
1.4 ขอบเขตของงานวิจัย.....	5
1.5 ประโยชน์ที่ได้รับจากการวิจัย.....	5
1.6 โครงสร้างของวิทยานิพนธ์.....	5
2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง.....	7
2.1 ทฤษฎีและวรรณกรรมที่เกี่ยวข้อง.....	7
2.1.1 คนเดินเท้า.....	7
2.1.2 ทางข้ามถนน.....	7
2.1.3 ทางข้ามถนนอัจฉริยะ (Smart Crosswalk).....	8
2.1.4 ข้อมูลไม่สมดุล (Imbalanced Data).....	9
2.1.5 การทดสอบความสัมพันธ์ระหว่างตัวแปร.....	13
2.1.6 การถดถอยโลจิสติก (Logistic Regression Analysis).....	14
2.1.7 ต้นไม้ตัดสินใจ (Decision Tree).....	16
2.2 งานวิจัยที่เกี่ยวข้อง.....	17
2.3 สรุปงานวิจัยที่เกี่ยวข้อง.....	21

สารบัญ (ต่อ)

หน้า

3	วิธีการดำเนินงาน	29
3.1	ขั้นตอนการดำเนินงานวิจัย	29
3.2	การกำหนดขอบเขตและพื้นที่ที่ทำการศึกษา	31
3.3	การเก็บรวบรวมข้อมูล	34
3.3.1	ข้อมูลลักษณะทางกายภาพของทางข้ามอัจฉริยะ	34
3.3.2	ข้อมูลพฤติกรรมจากระบบทางข้ามอัจฉริยะ	35
3.4	การวิเคราะห์ข้อมูล	37
3.4.1	สถิติเชิงพรรณนา	37
3.4.2	วิเคราะห์หาความสัมพันธ์ระหว่างตัวแปรอิสระ	37
3.4.3	วิเคราะห์สร้างแบบจำลองจากชุดข้อมูลดั้งเดิมโดยไม่แก้ไขคุณลักษณะที่ไม่สมดุล	37
3.4.4	ดำเนินการแก้ไขคุณลักษณะที่ไม่สมดุล (Imbalanced Data)	37
3.4.5	วิเคราะห์สร้างแบบจำลองจากชุดข้อมูลที่มีการแก้ไขคุณลักษณะที่ไม่สมดุล	38
4	ผลการวิเคราะห์ข้อมูล	39
4.1	ผลการวิเคราะห์ข้อมูลสถิติเชิงพรรณนา	39
4.2	วิเคราะห์หาความสัมพันธ์ระหว่างตัวแปรอิสระ	44
4.3	วิเคราะห์สร้างแบบจำลองจากชุดข้อมูลดั้งเดิมโดยไม่แก้ไขคุณลักษณะที่ไม่สมดุล	46
4.4	ดำเนินการแก้ไขคุณลักษณะที่ไม่สมดุล (Imbalanced Data)	51
4.5	วิเคราะห์สร้างแบบจำลองจากชุดข้อมูลที่มีการปรับแก้ไขคุณลักษณะที่ไม่สมดุล	54
4.5.1	แบบจำลองวิธีการถดถอยโลจิสติกแบบทวิ (Binary logistic regression)	54
4.5.2	แบบจำลองวิธีต้นไม้ตัดสินใจ (Decision tree)	61
5	สรุป อภิปราย และข้อเสนอแนะ	68
	รายการอ้างอิง	71
	ภาคผนวก	75
	ภาคผนวก ก งานประชุมวิชาการ	76
	ภาคผนวก ข บทความวิชาการที่ได้รับการตีพิมพ์เผยแพร่ในระหว่างศึกษา	78
	ประวัติผู้เขียน	89

สารบัญตาราง

ตารางที่	หน้า
2-1 เมตริกซ์วัดประสิทธิภาพ (Confusion Matrix)	11
2-2 ประสิทธิภาพของตัวแบบจากพื้นที่ใต้โค้ง ROC	13
2-3 สรุปงานวิจัยที่เกี่ยวข้อง	22
2-3 สรุปงานวิจัยที่เกี่ยวข้อง (ต่อ).....	23
2-3 สรุปงานวิจัยที่เกี่ยวข้อง (ต่อ).....	24
2-3 สรุปงานวิจัยที่เกี่ยวข้อง (ต่อ).....	25
2-3 สรุปงานวิจัยที่เกี่ยวข้อง (ต่อ).....	26
2-3 สรุปงานวิจัยที่เกี่ยวข้อง (ต่อ).....	27
2-3 สรุปงานวิจัยที่เกี่ยวข้อง (ต่อ).....	28
3-1 ข้อมูลลักษณะทางกายภาพของทางข้ามถนนอัจฉริยะ	35
3-2 รายละเอียดข้อมูลที่ได้จากฐานข้อมูล Traffic data.....	35
3-3 รายละเอียดข้อมูลที่ได้จากฐานข้อมูล Transaction data.....	36
3-3 รายละเอียดข้อมูลที่ได้จากฐานข้อมูล Transaction data (ต่อ)	37
4-1 ผลการวิเคราะห์ค่าสถิติพื้นฐานของข้อมูลทางข้ามถนนอัจฉริยะ	40
4-1 ผลการวิเคราะห์ค่าสถิติพื้นฐานของข้อมูลทางข้ามถนนอัจฉริยะ (ต่อ).....	41
4-1 ผลการวิเคราะห์ค่าสถิติพื้นฐานของข้อมูลทางข้ามถนนอัจฉริยะ (ต่อ).....	42
4-2 ค่าสัมประสิทธิ์สหสัมพันธ์ของตัวแปรที่ใช้ในการศึกษา	45
4-3 คำอธิบายตัวแปรตามและตัวแปรอิสระในการศึกษาครั้งนี้	46
4-3 คำอธิบายตัวแปรตามและตัวแปรอิสระในการศึกษาครั้งนี้ (ต่อ).....	47
4-4 ผลลัพธ์วิธีการถดถอยโลจิสติกแบบทวิจากชุดข้อมูลดั้งเดิมโดยไม่แก้ไขคุณลักษณะที่ไม่สมดุล	48
4-5 ผลลัพธ์การพยากรณ์ตัวแปรตามของแบบจำลองจากชุดข้อมูลดั้งเดิมโดยไม่แก้ไขคุณลักษณะที่ไม่สมดุลวิธีการถดถอยโลจิสติกแบบทวิ	48
4-6 ผลลัพธ์การพยากรณ์ตัวแปรตามของแบบจำลองจากชุดข้อมูลดั้งเดิมโดยไม่แก้ไขคุณลักษณะที่ไม่สมดุลวิธีต้นไม้ตัดสินใจ	51
4-7 ผลการวิเคราะห์ค่าสถิติพื้นฐานรายละเอียดของคลาสที่มีการแก้ไขคุณลักษณะที่ไม่สมดุลในแต่ละหลักการ	53

สารบัญตาราง (ต่อ)

ตารางที่	หน้า
4-8 ผลลัพธ์ประสิทธิภาพการพยากรณ์ตัวแปรตามของแบบจำลองด้วยวิธีการถดถอยโลจิสติกแบบทวิ จากชุดข้อมูลดั้งเดิมและชุดข้อมูลที่มีการแก้ไขคุณลักษณะที่ไม่สมดุล	55
4-9 การตรวจสอบความเท่ากันของความแปรปรวนระหว่างกลุ่มของค่า AUC จากวิธีการถดถอยโลจิสติกแบบทวิ ด้วย Levene's test (center = mean)	56
4-10 ผลการทดสอบความแตกต่างของค่าเฉลี่ย AUC ระหว่างกลุ่ม จากวิธีการถดถอยโลจิสติกแบบทวิ ด้วย Welch's ANOVA.....	57
4-11 การทดสอบความแตกต่างของค่าเฉลี่ย AUC จากวิธีการถดถอยโลจิสติกแบบทวิรายคู่ด้วยวิธีทดสอบแบบ Dunnett T3.....	57
4-12 ผลลัพธ์วิธีการถดถอยโลจิสติกแบบทวิ โดยการใช้การสุ่มตัวอย่างแบบ Downsampling.....	58
4-13 ประสิทธิภาพของแบบจำลองจากชุดข้อมูลดั้งเดิมโดยไม่แก้ไขคุณลักษณะที่ไม่สมดุล และแบบจำลองที่พัฒนาขึ้นโดยใช้วิธีการ Downsampling วิธีการถดถอยโลจิสติกแบบทวิ ด้วยชุดข้อมูลทดสอบ	60
4-14 ผลลัพธ์ประสิทธิภาพการพยากรณ์ตัวแปรตามของแบบจำลองด้วยวิธีต้นไม้ตัดสินใจ จากชุดข้อมูลดั้งเดิมและชุดข้อมูลที่มีการแก้ไขคุณลักษณะที่ไม่สมดุล	62
4-15 การตรวจสอบความเท่ากันของความแปรปรวนระหว่างกลุ่มของค่า AUC จากวิธีต้นไม้ตัดสินใจ ด้วย Levene's test (center = mean)	63
4-16 ผลการทดสอบความแตกต่างของค่าเฉลี่ย AUC ระหว่างกลุ่ม จากวิธีต้นไม้ตัดสินใจด้วย Welch's ANOVA.....	64
4-17 การทดสอบความแตกต่างของค่าเฉลี่ย AUC จากวิธีต้นไม้ตัดสินใจรายคู่ ด้วยวิธีทดสอบแบบ Dunnett T3.....	64
4-18 ประสิทธิภาพของแบบจำลองจากชุดข้อมูลดั้งเดิมโดยไม่แก้ไขคุณลักษณะที่ไม่สมดุล และแบบจำลองที่พัฒนาขึ้นโดยใช้วิธีการ Downsampling วิธีต้นไม้ตัดสินใจ ด้วยชุดข้อมูลทดสอบ	67

สารบัญรูป

รูปที่	หน้า
2-1	ตรรกะของการควบคุมสัญญาณไฟทางข้าม (Signal Control Logic)..... 9
2-2	แสดงเส้นโค้ง ROC curve..... 13
2-3	โครงสร้างต้นไม้ตัดสินใจ..... 16
3-1	ขั้นตอนการดำเนินการวิจัย..... 30
3-2	ตำแหน่งทางข้ามสัญญาณไฟอัจฉริยะบนทางหลวงหมายเลข 3242 กม. 18+110..... 31
3-3	ตำแหน่งทางข้ามสัญญาณไฟอัจฉริยะบนทางหลวงหมายเลข 306 กม. 4+970..... 32
3-4	ตำแหน่งทางข้ามสัญญาณไฟอัจฉริยะบนทางหลวงหมายเลข 3316 กม. 2+225..... 32
3-5	ตำแหน่งทางข้ามสัญญาณไฟอัจฉริยะบนทางหลวงหมายเลข 3215 กม. 11+020..... 33
3-6	ตำแหน่งทางข้ามสัญญาณไฟอัจฉริยะบนทางหลวงหมายเลข 3242 กม. 2+220..... 33
3-7	ตำแหน่งทางข้ามสัญญาณไฟอัจฉริยะบนทางหลวงหมายเลข 314 กม. 8+410..... 34
4-1	กราฟแสดงความถี่สะสมข้อมูลทางข้ามอัจฉริยะ..... 43
4-2	ผลลัพธ์ของวิธีต้นไม้ตัดสินใจจากชุดข้อมูลการฝึกเริ่มต้นโดยไม่แก้ไขคุณลักษณะที่ไม่สมดุล ..50
4-3	ความน่าจะเป็นที่คาดการณ์ว่าคนเดินข้ามจะฝ่าฝืนไฟสัญญาณ..... 60
4-4	ผลลัพธ์แบบจำลองวิธีต้นไม้ตัดสินใจ (Decision tree) โดยใช้การสุ่มตัวอย่างแบบ Downsampling 65

บทที่ 1

บทนำ

1.1 ที่มาและความสำคัญของงานวิจัย

ปัจจุบันอุบัติเหตุบนท้องถนนติดอันดับหนึ่งในสาเหตุหลักของการเสียชีวิตในทุกกลุ่มอายุทั่วโลก จากรายงานสถานการณ์ด้านความปลอดภัยทางถนนทั่วโลกประจำปี 2018 ขององค์การอนามัยโลก พบว่า จำนวนผู้เสียชีวิตจากการเกิดอุบัติเหตุบนท้องถนนทั่วโลกยังคงมีค่าสูงเกินกว่าที่ยอมรับได้ โดยมีจำนวนผู้เสียชีวิตจากการเกิดอุบัติเหตุบนท้องถนนมากถึง 1.35 ล้านรายในแต่ละปี ภูมิภาคเอเชียตะวันออกเฉียงใต้มีอัตราการเสียชีวิตจากการเกิดอุบัติเหตุบนท้องถนนเท่ากับ 20.7 รายต่อประชากร 1 แสนคน ซึ่งสูงเป็นอันดับที่สองรองจากภูมิภาคแอฟริกา และมากกว่าครึ่งหนึ่งของการเสียชีวิตและบาดเจ็บจากอุบัติเหตุบนท้องถนนทั้งหมดเกี่ยวข้องกับกลุ่มผู้ใช้ถนนที่เปราะบาง (Vulnerable Road Users) เช่น คนเดินเท้า ผู้ขับขี่จักรยาน และผู้ขับขี่จักรยานยนต์ ประเทศไทยมีอัตราการเสียชีวิตจากการเกิดอุบัติเหตุบนท้องถนนสูงเป็นอันดับที่ 9 ของโลก และสูงที่สุดในเอเชียตะวันออกเฉียงใต้ ซึ่งมีอัตราการเสียชีวิตอยู่ที่ 32.7 รายต่อประชากร 1 แสนคน ส่งผลให้เกิดความสูญเสียเป็นอย่างมากทั้งในแง่ของชีวิตและเศรษฐกิจของประเทศ และเมื่อทำการพิจารณาอัตราการเสียชีวิตตามประเภทของผู้ใช้รถใช้ถนน พบว่า ผู้ขับขี่จักรยานยนต์มีอัตราการเสียชีวิตสูงที่สุดคือร้อยละ 74 และรองลงมาคือคนเดินเท้าร้อยละ 8 (World Health Organization, 2018) จากข้อมูลการบูรณาการจำนวนผู้เสียชีวิตจากอุบัติเหตุทางถนน 3 ฐาน กรมควบคุมโรค กระทรวงสาธารณสุข พ.ศ. 2565 พบจำนวนคนเดินเท้าที่เสียชีวิตจากอุบัติเหตุทางถนนช่วงระหว่างปี พ.ศ. 2561- พ.ศ. 2565 เฉลี่ยเดือนละ 32.97 คน นั้นหมายความว่าในทุก ๆ วัน จะมีคนเดินเท้าเสียชีวิตจากอุบัติเหตุบนท้องถนนอย่างน้อยวันละ 1 คน (กรมควบคุมโรค กระทรวงสาธารณสุข, 2565)

แม้ว่าจำนวนของคนเดินเท้าที่เสียชีวิตจากอุบัติเหตุบนท้องถนนจะถือว่าไม่มากนักเมื่อเทียบกับผู้ใช้ถนนประเภทอื่น แต่สิ่งที่น่าวิตกคือแนวโน้มของความรุนแรงของอุบัติเหตุคนเดินเท้าที่มีแนวโน้มความรุนแรงเพิ่มสูงขึ้นอย่างต่อเนื่อง จากสถิติข้อมูลอุบัติเหตุบนทางหลวงระหว่างปี พ.ศ. 2555 – พ.ศ. 2564 สำนักอำนวยการความปลอดภัย กรมทางหลวง เผยว่าอุบัติเหตุที่เกิดขึ้นกับคนเดินเท้าบนทางหลวง ส่งผลทำให้คนเดินเท้าเสียชีวิตมากกว่าร้อยละ 50 และเพิ่มสูงขึ้นถึงร้อยละ 65 ในปี พ.ศ. 2564 ซึ่งสอดคล้องกับค่าดัชนีความรุนแรง (Severity Index) ในช่วง พ.ศ. 2560 - พ.ศ. 2564

(มูลนิธิไทยโรดส์, 2565) และจากการศึกษาสถานการณ์ความปลอดภัยของคนเดินเท้าทั่วโลก พบว่าสาเหตุสำคัญที่ทำให้คนเดินเท้าเสียชีวิตหรือบาดเจ็บคืออุบัติเหตุชนคนเดินเท้าขณะกำลังข้ามถนน (ศูนย์วิชาการเพื่อความปลอดภัยทางถนน, 2562)

จากที่กล่าวมาในข้างต้นจะเห็นได้ว่า ปัญหาการเกิดอุบัติเหตุยานพาหนะชนคนเดินเท้าที่ทางข้าม เมื่อเกิดขึ้นแล้วความรุนแรงของอุบัติเหตุส่งผลให้เกิดการเสียชีวิตที่สูงมาก และประเทศไทยยังคงประสบปัญหาเหล่านี้มาอย่างต่อเนื่อง กรมทางหลวง โดยสำนักอำนวยความปลอดภัยตระหนักถึงความปลอดภัยของคนเดินเท้าในการข้ามถนน จึงได้กำหนดให้มีการดำเนินงาน “โครงการศึกษาเพื่อกำหนดแนวทางการจัดทำทางข้ามถนนปลอดภัย 1 โครงการ” ขึ้นในปี พ.ศ. 2564 เพื่อยกระดับความปลอดภัยของผู้ใช้ทางทั้งประชาชนที่ต้องการข้ามถนนและผู้ขับขี่ยานพาหนะ โครงการดังกล่าวได้มีการศึกษาและเสนอแนะเทคโนโลยีต่างๆ เพื่อนำไปใช้ในการก่อสร้างหรือปรับปรุงทางข้ามถนน เพื่อพัฒนาให้เป็นทางข้ามถนนอัจฉริยะ (Smart Crosswalk) ซึ่งในปัจจุบันได้มีการนำรูปแบบของทางข้ามถนนอัจฉริยะดำเนินการติดตั้งจริงแล้วทั้งสิ้น 11 แห่ง

ทางข้ามถนนอัจฉริยะ (Smart Crosswalk) เป็นทางข้ามที่นำเอาอุปกรณ์หรือเทคโนโลยีแบบอัตโนมัติต่าง ๆ ดำเนินการติดตั้งบริเวณทางข้าม เพื่อช่วยอำนวยความสะดวก เพิ่มความปลอดภัยและประสิทธิภาพในการใช้งานทางข้ามให้มากยิ่งขึ้น หลักการทำงานของระบบทางข้ามถนนอัจฉริยะ (Smart Crosswalk) จะทำการประมวลผลความต้องการใช้ทางข้ามและวิเคราะห์ควบคุมไฟสัญญาณทางข้าม โดยอาศัยอุปกรณ์ตรวจจับคนเดินข้ามร่วมกับอุปกรณ์ตรวจจับยานพาหนะบริเวณทางข้าม หากระบบพบว่ามีคนเดินข้ามกดปุ่มขอไฟสัญญาณเพื่อข้ามถนน ระบบจะทำการคำนวณหาช่วงที่เหมาะสมในการหยุดยานพาหนะและให้ไฟสัญญาณแก่คนเดินข้าม เมื่อคนเดินข้ามได้รับไฟสัญญาณแล้ว ระบบจะทำการตรวจจับคนเดินข้ามทั้งหมดว่าได้ทำการข้ามถนนพ้นทั้งหมดแล้วหรือไม่ หากตรวจพบว่าคนเดินข้ามบางส่วนยังข้ามถนนไม่พ้น ระบบจะทำการขยายระยะเวลาไฟสัญญาณให้กับคนเดินข้าม จนกระทั่งไม่มีคนเดินข้ามอยู่ในพื้นที่ทางข้าม หรือจนกระทั่งระยะเวลาไฟสัญญาณคนข้ามเกินระยะเวลาที่ใช้ในการข้ามถนนสูงสุด จากนั้นระบบจะทำการตัดไฟสัญญาณสีเขียวให้กับยานพาหนะ เพื่อไม่ให้ส่งผลกระทบต่อการทำงานของจราจร (กรมทางหลวง, 2564) ถึงแม้จะมีการติดตั้งเทคโนโลยีระบบทางข้ามถนนอัจฉริยะ (Smart Crosswalk) แล้ว แต่ถึงอย่างนั้นก็ยังพบเห็นการฝ่าฝืนไฟสัญญาณของคนเดินข้ามและผู้ขับขี่ยานพาหนะอยู่ในทุก ๆ ทางข้าม ซึ่งการฝ่าฝืนดังกล่าวมีความเสี่ยงอย่างมากที่จะก่อเกิดอุบัติเหตุร้ายแรงทั้งต่อผู้ขับขี่ยานพาหนะและคนเดินข้าม

จากการทบทวนงานวิจัยในอดีตที่การศึกษาเกี่ยวกับปัจจัยที่ส่งผลต่อพฤติกรรมการฝ่าฝืนไฟสัญญาณจราจรของคนเดินข้าม ทั้งในประเทศและต่างประเทศ พบว่าโดยส่วนใหญ่ทำการศึกษาทางข้ามถนนที่มีสัญญาณไฟควบคุมแบบรอบเวลาคงที่ (Fixed Time Signal) และดำเนินการ

วิเคราะห์ด้วยการถดถอยโลจิสติกแบบทวิ (Binary logistic regression) เป็นหลัก (Afshari & Ayati, 2021; Raoniar et al., 2022a; Raoniar & Maurya, 2022b, 2022c; Zhang & Deng, 2019; Zhu et al., 2021; ชิตภัทร อินทรวิมลเมธา และพงษ์ศักดิ์ สุริยวานากุล, 2555; วีรพงศ์ บุญเจริญปัญญา และจรัส พิทักษ์ศฤงคาร, 2567) ซึ่งในปัจจุบันยังขาดการศึกษาเพื่อตรวจสอบและทำความเข้าใจเกี่ยวกับปัจจัยที่มีอิทธิพลต่อพฤติกรรมการฝ่าฝืนไฟสัญญาณจราจรของคนเดินข้ามที่ทางม้าลายที่มีระบบทางข้ามถนนอัจฉริยะ (Smart Crosswalk) และจากการวิเคราะห์เบื้องต้นเกี่ยวกับพฤติกรรมการฝ่าฝืนไฟสัญญาณของคนเดินข้ามที่ทางข้ามถนนอัจฉริยะ บ่งชี้ว่าข้อมูลในการศึกษานี้ไม่สมดุล (Imbalanced Data) เนื่องจากจำนวนเหตุการณ์ที่พบการฝ่าฝืนไฟสัญญาณของคนเดินข้ามในระหว่างรอบไฟสัญญาณใด ๆ น้อยกว่าเหตุการณ์ที่ไม่มีการฝ่าฝืนไฟสัญญาณของคนเดินข้ามอยู่เป็นจำนวนมาก ซึ่งหากพัฒนาแบบจำลองโดยไม่พิจารณาถึงคุณสมบัติที่ไม่สมดุลของข้อมูล ผลลัพธ์ที่คาดการณ์ไว้ของแบบจำลองอาจมีความเอนเอียงไปทางกลุ่มของคลาสส่วนมาก (Majority Class) จากการทบทวนงานวิจัยในอดีตที่เกี่ยวข้องกับการศึกษาการปรับความไม่สมดุลของชุดข้อมูลที่ดำเนินการในประเทศไทย (นพมาศ อัครจันทโชติ และดิเรก พนิตสุภาภมร, 2561; วิชญวิสิฐ เกษรสิทธิ์, วิจิต หล่อจิระขุนท์กุล และจิราวัลย์ จิตรถเวช, 2561; อัจฉรา แฝ้วบาง และสายชล สันสมบุรณ์ทอง, 2563) พบว่าโดยส่วนใหญ่แล้วจะการใช้การปรับความไม่สมดุลของชุดข้อมูล ในด้านของข้อมูลทางการแพทย์ ข้อมูลทางการเงิน ข้อมูลประชากร ข้อมูลการศึกษา เป็นต้น ซึ่งในบริบทของข้อมูลการฝ่าฝืนไฟสัญญาณของคนเดินข้ามที่มีระบบไฟสัญญาณจราจรหรือระบบทางข้ามถนนอัจฉริยะ (Smart Crosswalk) ยังไม่พบงานวิจัยที่พิจารณาถึงคุณสมบัติที่ไม่สมดุลของข้อมูล

จากช่องว่างการวิจัยที่ได้กล่าวถึงในข้างต้น การศึกษาค้นคว้าครั้งนี้จึงทำการตรวจสอบปัจจัยที่มีอิทธิพลต่อการฝ่าฝืนไฟสัญญาณของคนเดินข้ามที่ทางข้ามถนนอัจฉริยะ (Smart Crosswalk) ในประเทศไทย โดยคำนึงถึงคุณสมบัติที่ไม่สมดุลของข้อมูลร่วมด้วย โดยอาศัยข้อมูลลักษณะทางกายภาพของถนน และข้อมูลพฤติกรรมการจราจรบนทางข้ามถนนอัจฉริยะ ซึ่งประกอบด้วย 1) ข้อมูลเกี่ยวกับจราจรบริเวณทางข้าม 2) ข้อมูลเกี่ยวกับคนเดินข้าม และ 3) ข้อมูลจำเพาะเกี่ยวกับระบบไฟสัญญาณจราจรบริเวณทางข้าม อย่างไรก็ตามข้อมูลที่ได้รับจากระบบทางข้ามถนนอัจฉริยะ (Smart Crosswalk) ยังขาดข้อมูลเกี่ยวกับลักษณะเฉพาะส่วนบุคคลของคนเดินข้าม เช่น เพศ อายุ และอาชีพ ดังนั้น การศึกษานี้จึงไม่สามารถพิจารณาปัจจัยที่เกี่ยวข้องกับลักษณะเฉพาะส่วนบุคคลของคนเดินข้ามได้ และนอกจากนี้ข้อมูลการฝ่าฝืนไฟสัญญาณของคนเดินข้ามที่ถูกบันทึกในระบบมีลักษณะของข้อมูลเป็นจำนวนคนเดินข้ามทั้งหมดที่ฝ่าฝืนไฟสัญญาณในแต่ละรอบไฟสัญญาณ ดังนั้นตัวแปรตามที่ใช้ในการศึกษาค้นคว้าครั้งนี้จึงเป็นแบบไบนารี ซึ่งเป็นการระบุว่ามีการฝ่าฝืนไฟสัญญาณของคนเดินข้ามหรือไม่ในแต่ละรอบไฟสัญญาณ ผู้วิจัยจึงเลือกใช้การถดถอยโลจิสติกแบบทวิ (Binary Logistic Regression) ในการวิเคราะห์ เพื่อศึกษาว่าตัวแปรอิสระใดบ้างที่มีความสัมพันธ์กับ

ตัวแปรตาม และเพื่อสร้างสมการพยากรณ์โดยใช้ตัวแปรอิสระเป็นตัวพยากรณ์ตัวแปรตาม และนอกจากนี้ผู้วิจัยได้ทำการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ (Decision Tree) ร่วมด้วย วิธีต้นไม้ตัดสินใจเป็นตัวแทนที่เลียนแบบการตัดสินใจของมนุษย์ มีลักษณะคล้ายโครงสร้างของต้นไม้ที่แบ่งปัจจัยในการตัดสินใจออกเป็นปัจจัยย่อย ๆ แสดงให้เห็นถึงแนวโน้มพฤติกรรมการฝ่าฝืนไฟสัญญาณของคนเดินข้ามที่ทางข้ามถนนอัจฉริยะได้ชัดเจนมากยิ่งขึ้น ซึ่งเป้าหมายในการศึกษาครั้งนี้คือ เพื่อสร้างคำแนะนำหรือมาตรการเพื่อจัดการกับการฝ่าฝืนไฟสัญญาณจราจร และลดโอกาสเสี่ยงต่อการเกิดอุบัติเหตุทางถนนของคนเดินข้ามที่ทางข้ามถนนอัจฉริยะในประเทศได้อย่างมีนัยสำคัญ

1.2 วัตถุประสงค์งานวิจัย

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาปัจจัยที่มีอิทธิพลต่อการฝ่าฝืนไฟสัญญาณของคนเดินข้ามที่ทางข้ามถนนอัจฉริยะ (Smart Crosswalk) ในประเทศไทย โดยมีรายละเอียดดังนี้

1.2.1 เพื่อศึกษาปัจจัยที่มีอิทธิพลต่อการฝ่าฝืนไฟสัญญาณของคนเดินข้ามที่ทางข้ามถนนอัจฉริยะ (Smart Crosswalk) ในประเทศไทย โดยใช้วิธีการถดถอยโลจิสติกแบบทวิ (Binary Logistic Regression) และ วิธีต้นไม้ตัดสินใจ (Decision Tree)

1.2.2 เพื่อเปรียบเทียบประสิทธิภาพเทคนิคการแก้ไขคุณลักษณะที่ไม่สมดุล (Imbalanced Data) 4 วิธี ได้แก่ Upsampling, Downsampling, Random OverSampling Examples (ROSE), and Synthetic Minority Oversampling Technique (SMOTE)

1.3 คำถามงานวิจัย

งานวิจัยนี้เลือกใช้การถดถอยโลจิสติกแบบทวิ (Binary Logistic Regression) และวิธีต้นไม้ตัดสินใจ (Decision tree) ในการดำเนินการวิเคราะห์ข้อมูล โดยมีคำถามวิจัยดังนี้

1.3.1 จากชุดข้อมูลดั้งเดิมโดยไม่แก้ไขคุณลักษณะที่ไม่สมดุล ปัจจัยลักษณะทางกายภาพของถนน และปัจจัยข้อมูลทุติยภูมิจากระบบทางข้ามถนนอัจฉริยะ ส่งผลต่อการฝ่าฝืนไฟสัญญาณของคนเดินข้ามอย่างไร

1.3.2 แบบจำลองที่ได้จากชุดข้อมูลดั้งเดิมโดยไม่แก้ไขคุณลักษณะที่ไม่สมดุล มีประสิทธิภาพในการพยากรณ์การฝ่าฝืนไฟสัญญาณของคนเดินข้ามหรือไม่

1.3.3 เทคนิคการแก้ไขคุณลักษณะที่ไม่สมมูลรูปแบบใด ส่งผลทำให้แบบจำลองสามารถพยากรณ์การฝ่าฝืนไฟสัญญาณของคนเดินข้ามได้อย่างมีประสิทธิภาพมากที่สุด

1.3.4 หลังจากใช้เทคนิคการแก้ไขคุณลักษณะที่ไม่สมดุล ปัจจัยลักษณะทางกายภาพของถนน และปัจจัยข้อมูลทุติยภูมิจากระบบทางข้ามถนนอัจฉริยะ ส่งผลต่อการฝ่าฝืนไฟสัญญาณของคนเดินข้ามอย่างไร

1.3.5 จากผลลัพธ์การศึกษาระหว่างวิธีการถดถอยโลจิสติกแบบทวิ (Binary Logistic Regression) และ วิธีต้นไม้ตัดสินใจ (Decision Tree) แนวโน้มของปัจจัยที่ส่งผลต่อการฝ่าฝืนไฟสัญญาณของคนเดินข้าม ของวิธีการวิเคราะห์ทั้ง 2 วิธีมีความคล้ายคลึงหรือแตกต่างกันหรือไม่

1.4 ขอบเขตของงานวิจัย

งานวิจัยนี้ได้กำหนดพื้นที่ทางข้ามถนนอัจฉริยะ (Smart Crosswalk) ของกรมทางหลวงในการดำเนินการสำรวจและเก็บข้อมูลที่ใช้ในการวิเคราะห์ทั้งหมด 6 แห่ง จากทั้งสิ้น 11 แห่งในปัจจุบัน ได้แก่

1. ทางข้ามถนนเอกชัย 102/3 ทางหลวงหมายเลข 3242 กม. 18+110 จังหวัดกรุงเทพมหานคร
 2. ทางข้ามถนนประชาราษฎร์ ซอย 10 ทางหลวงหมายเลข 306 กม. 4+970 จังหวัดนนทบุรี
 3. ทางข้ามวัดไร่ขิง ทางหลวงหมายเลข 3316 กม. 2+225 จังหวัดนครปฐม
 4. ทางข้ามหมู่บ้าน ช.รุ่งเรือง 6 ทางหลวงหมายเลข 3215 กม. 11+020 จังหวัดนนทบุรี
 5. ทางข้ามหน้า โรงพยาบาลเอกชัย ทางหลวงหมายเลข 3242 กม. 2+220 จังหวัดสมุทรสาคร
 6. ทางข้ามเอื้ออาทรแสนฤๅช ทางหลวงหมายเลข 314 กม. 8+410 จังหวัดฉะเชิงเทรา
- ทั้งนี้สาเหตุที่ผู้วิจัยไม่นำทางข้ามอื่น ๆ อีก 5 แห่ง มาวิเคราะห์ร่วมด้วย เนื่องจากพบว่าระบบทางข้ามถนนอัจฉริยะเกิดการชำรุด และเป็นทางข้ามที่เพิ่งถูกก่อสร้างหรือปรับปรุงให้เป็นทางข้ามถนนอัจฉริยะในช่วงปลายปี 2566 ซึ่งในขณะที่ทำงานวิจัยฉบับนี้ข้อมูลที่ได้รับจากระบบทางข้ามถนนอัจฉริยะไม่เพียงพอสำหรับการนำมาใช้ในการวิเคราะห์

1.5 ประโยชน์ที่ได้รับจากงานวิจัย

ผลการศึกษาในครั้งนี้ช่วยให้สามารถระบุปัจจัยที่ส่งผลต่อพฤติกรรมการฝ่าฝืนไฟสัญญาณของคนเดินข้ามที่ทางข้ามถนนอัจฉริยะ (Smart Crosswalk) ในประเทศไทย ทั้งทางตรงและทางอ้อม องค์การภาครัฐและหน่วยงานที่เกี่ยวข้องสามารถนำข้อมูลที่ได้ไปประยุกต์ใช้ในการกำหนดนโยบายหรือมาตรการที่ต้องการจะลดโอกาสเสี่ยงในการเกิดอุบัติเหตุทางถนนสำหรับคนเดินข้าม นอกจากนี้ยังสามารถใช้ข้อมูลดังกล่าวในการปรับปรุงและพัฒนาระบบไฟสัญญาณของทางข้ามถนนอัจฉริยะให้มีประสิทธิภาพและเหมาะสมกับบริบทของแต่ละพื้นที่ ตลอดจนเพิ่มความปลอดภัยให้กับผู้ใช้รถใช้ถนนในบริเวณทางข้ามถนนอัจฉริยะมากยิ่งขึ้น

1.6 โครงสร้างของวิทยานิพนธ์

วิทยานิพนธ์ฉบับนี้ทำการศึกษาเกี่ยวกับปัจจัยที่มีอิทธิพลต่อการฝ่าฝืนไฟสัญญาณของคนเดินข้ามที่ทางข้ามถนนอัจฉริยะ (Smart Crosswalk) ในประเทศไทย เพื่อให้สามารถตอบวัตถุประสงค์และคำถามการวิจัยได้อย่างครบถ้วน วิทยานิพนธ์ฉบับนี้จึงจัดโครงสร้างเนื้อหาออกเป็น 5 บท โดยมีรายละเอียดดังนี้

บทที่ 1: บทนำ กล่าวถึงที่มาและความสำคัญของงานวิจัย, วัตถุประสงค์งานวิจัย, ขอบเขตงานวิจัย และประโยชน์ที่คาดว่าจะได้รับจากงานวิจัย

บทที่ 2: ทฤษฎีและงานวิจัยที่เกี่ยวข้อง กล่าวถึงการทบทวนความรู้ในเรื่องของคนเดินเท้า รูปแบบทางข้าม หลักการทำงานของระบบทางข้ามอัจฉริยะ (Smart Crosswalk) คำนิยามค่าพารามิเตอร์ที่ติดตั้งในปัจจุบัน หลักการทางสถิติ รวมถึงการทบทวนงานวิจัยที่เกี่ยวข้องกับการศึกษาในครั้งนี้

บทที่ 3: วิธีการดำเนินการวิจัย กล่าวถึงขั้นตอนการดำเนินการศึกษา การกำหนดขอบเขตพื้นที่ที่ทำการศึกษา การเก็บรวบรวมข้อมูล และขั้นตอนในการวิเคราะห์ข้อมูล

บทที่ 4: ผลการวิเคราะห์ข้อมูล กล่าวถึงผลการวิเคราะห์โดยใช้สถิติเชิงพรรณนาเพื่อตรวจสอบพฤติกรรมเบื้องต้นของผู้ขับขี่ยานพาหนะและคนเดินข้าม รวมถึงการทำงานของระบบทางข้ามถนนอัจฉริยะ ผลลัพธ์ของการวิเคราะห์สร้างแบบจำลองและการพยากรณ์การฝ่าฝืนไฟสัญญาณของคนเดินข้ามจากชุดข้อมูลดั้งเดิมโดยไม่แก้ไขคุณลักษณะที่ไม่สมดุล และจากชุดข้อมูลที่มีการแก้ไขคุณลักษณะที่ไม่สมดุล โดยวิธีการถดถอยโลจิสติกแบบทวิ (Binary Logistic Regression) และวิธีต้นไม้ตัดสินใจ (Decision Tree)

บทที่ 5: สรุปผล อภิปราย และข้อเสนอแนะ กล่าวถึงผลสรุปและอภิปรายการวิเคราะห์ในบทที่ 4 เกี่ยวกับปัจจัยที่มีอิทธิพลต่อการฝ่าฝืนไฟสัญญาณของคนเดินข้ามที่ทางข้ามถนนอัจฉริยะ (Smart Crosswalk) ในประเทศไทย จากวิธีการถดถอยโลจิสติกแบบทวิ (Binary Logistic Regression) และวิธีต้นไม้ตัดสินใจ (Decision Tree) และข้อเสนอแนะสำหรับการนำผลการศึกษาไปประยุกต์ใช้งาน

บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในบทนี้จะกล่าวถึงการทบทวนทฤษฎีและงานวิจัยที่เกี่ยวข้อง เพื่อเป็นแนวทางในการวิเคราะห์ข้อมูล ซึ่งประกอบด้วย ความรู้ในเรื่องของคนเดินเท้า รูปแบบทางข้าม หลักการทำงานของระบบทางข้ามอัจฉริยะ (Smart Crosswalk) คำนียามค่าพารามิเตอร์ที่ติดตั้งในปัจจุบัน หลักการทางสถิติ รวมถึงการทบทวนงานวิจัยที่เกี่ยวข้องกับการศึกษาในครั้งนี้

2.1 ทฤษฎีและวรรณกรรมที่เกี่ยวข้อง

2.1.1 คนเดินเท้า

คนเดินเท้า (Pedestrian) คือ บุคคลที่สัญจรเดิน ข้าม หรือวิ่ง บริเวณไหล่ทาง บนถนน ทางเท้าหรือขอบทาง ไปยังสถานที่ต่าง ๆ ด้วยเท้า โดยไม่มียานพาหนะเกี่ยวข้อง ซึ่งรวมถึงผู้ใช้รถเข็นสัญจรบนทางเท้าด้วย(สุเมธี สนธิกุล, 2556)

2.1.2 ทางข้ามถนน

พระราชบัญญัติจราจรทางบก พ.ศ. 2522 มาตรา 4 บัญญัติไว้ว่า “ทางข้าม หมายถึง พื้นที่ที่ทำไว้สำหรับคนเดินเท้าข้ามทางข้าม โดยทำเครื่องหมายเป็นเส้นหรือแนว หรือตอกหมุดไว้บนทาง ทั้งในระดับใต้พื้นดินหรือเหนือพื้นดิน” ได้แก่ ทางม้าลาย และสะพานลอย เป็นต้น

ทางข้ามถนน เป็นสิ่งสำคัญและมีประโยชน์ต่อคนเดินเท้าเป็นอย่างมาก ทางข้ามช่วยป้องกันและให้สิทธิในการข้ามแก่คนเดินข้าม ทำให้คนเดินเท้าสามารถข้ามถนนได้อย่างสะดวกและปลอดภัย ทางข้ามในปัจจุบันมีหลายประเภท แตกต่างกันไปตามความเหมาะสมในแต่ละพื้นที่ที่ติดตั้ง ซึ่งขึ้นอยู่กับการใช้งานและองค์ประกอบต่าง ๆ บริเวณทางข้าม เช่น ความกว้างของทางข้าม ความยาวของทางข้าม ลักษณะของทางข้าม ผิวทาง ระยะการมองเห็นทางข้ามถนน การเข้าถึงทางข้าม ลักษณะกายภาพของเกาะกลาง เป็นต้น ทางข้ามโดยส่วนใหญ่จะติดตั้งในพื้นที่เขตเมือง และสถานที่สำคัญ เช่น โรงเรียน โรงพยาบาล ย่านธุรกิจ หรือสถานที่ราชการ เป็นต้น (กรมทางหลวง , 2564)

ประเภทของทางข้าม (Pedestrian Crossing) โดยทั่วไปแบ่งออกเป็น 2 ประเภท ได้แก่

1) ทางข้ามบนผิวทาง (At-grade Pedestrian Crossing) หรือทางข้ามม้าลาย เป็นทางข้ามที่ติดตั้งอยู่ในระดับเดียวกันกับถนน ติดตั้งอยู่บริเวณช่วงถนน (Midblock) หรือบริเวณ

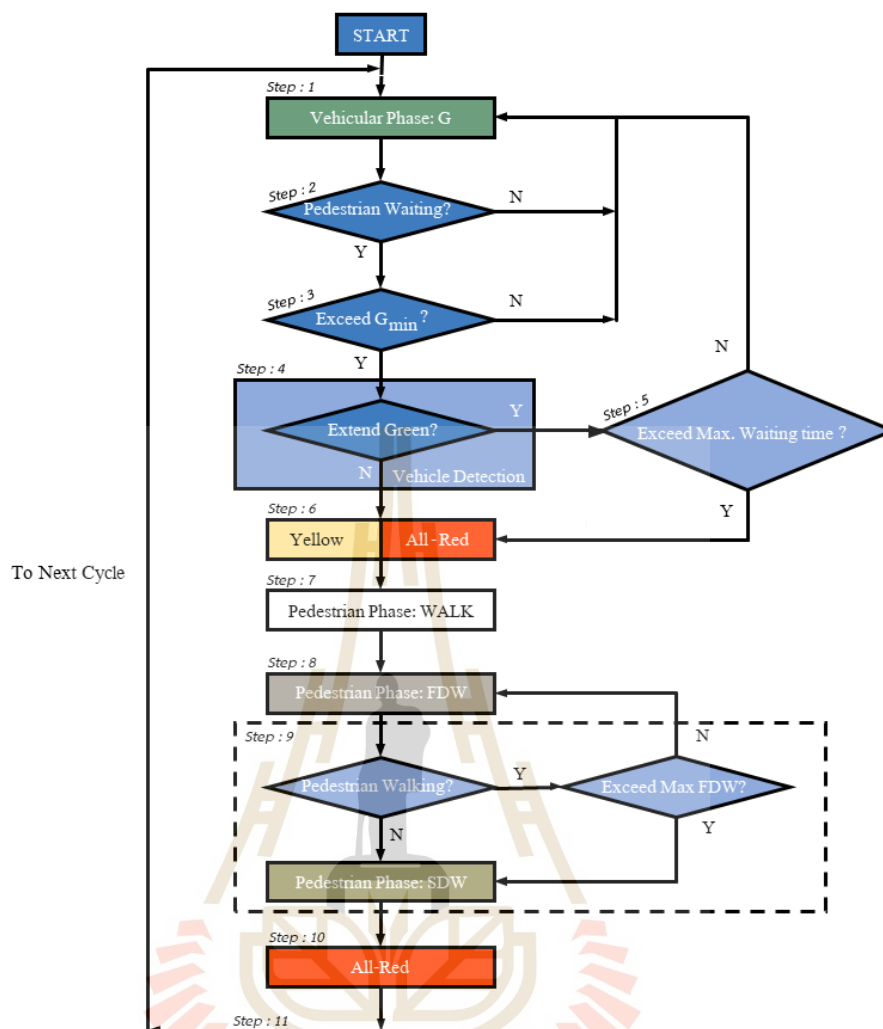
ทางแยก (Intersection) โดยส่วนใหญ่แบ่งเป็นสองรูปแบบหลัก ได้แก่ ทางข้ามระดับผิวทางแบบไม่มีไฟสัญญาณจราจร (Unsignalized Pedestrian Crossing) และทางข้ามระดับผิวทางแบบมีไฟสัญญาณจราจร (Signalized Pedestrian Crossing)

2) ทางข้ามต่างระดับ (Grade-separated Pedestrian Crossing) เป็นทางข้ามที่แยกคนเดินเท้าออกจากการสัญจรของยานพาหนะ เช่น ทางข้ามต่างระดับแบบยกข้ามหรือสะพานลอยคนข้าม (Pedestrian overpass) ทางข้ามคนลอด (Pedestrian underpass)

2.1.3 ทางข้ามถนนอัจฉริยะ (Smart Crosswalk)

จากการทบทวนงาน “โครงการศึกษาจัดทำแผนพัฒนาศูนย์บริหารจราจรกลาง (HTOC) กรมทางหลวง ปี พ.ศ. 2567 - พ.ศ. 2570 1 โครงการ” ระบุไว้ว่าทางข้ามถนนอัจฉริยะ (Smart Crosswalk) เป็นทางข้ามที่นำเอาอุปกรณ์หรือเทคโนโลยีแบบอัตโนมัติต่าง ๆ มาดำเนินการติดตั้งหรือปรับปรุงทางข้าม โดยหลักการทำงานของระบบทางข้ามถนนอัจฉริยะ (Smart Crosswalk) จะทำการประมวลผลความต้องการใช้ทางข้ามและวิเคราะห์ควบคุมไฟสัญญาณทางข้ามถนนโดยอาศัยอุปกรณ์ตรวจจับคนเดินข้ามร่วมกับอุปกรณ์ตรวจจับยานพาหนะบริเวณทางข้ามถนนเพื่อช่วยอำนวยความสะดวก เพิ่มความปลอดภัยและเพิ่มประสิทธิภาพในการใช้งานทางข้ามถนนให้มากยิ่งขึ้น

หลักการทำงานของระบบทางข้ามถนนอัจฉริยะ (Smart Crosswalk) คือ หากระบบพบว่ามีคนเดินข้ามกดปุ่มขอไฟสัญญาณเพื่อข้ามถนน ระบบจะทำการคำนวณหาช่วงที่เหมาะสม โดยพิจารณาจากระยะเวลาไฟสัญญาณสีเขียวที่ให้แก่ยานพาหนะ ความเร็วของยานพาหนะ และระยะห่างของยานพาหนะ (Gap) ที่สามารถยอมรับได้ และระยะเวลารอคอยของคนเดินข้าม ในการหยุดยานพาหนะและให้ไฟสัญญาณแก่คนเดินข้าม เมื่อตรวจสอบเงื่อนไขดังกล่าวเสร็จสิ้น ระบบจะทำการตัดไฟสัญญาณสีเขียวยานพาหนะ และเปิดไฟสัญญาณให้กับคนเดินข้าม เมื่อคนเดินข้ามได้รับไฟสัญญาณแล้ว ระบบจะทำการตรวจจับคนเดินข้ามทั้งหมดว่าได้ทำการข้ามถนนพ้นทั้งหมดแล้วหรือไม่ หากตรวจพบว่าคนเดินข้ามบางส่วนยังข้ามถนนไม่พ้น ระบบจะทำการขยายระยะเวลาไฟสัญญาณให้กับคนเดินข้าม จนกระทั่งไม่มีคนเดินข้ามอยู่ในพื้นที่ทางข้ามถนนหรือระยะเวลาไฟสัญญาณคนข้ามเกินระยะเวลาที่ใช้ในการข้ามถนนสูงสุด จากนั้นระบบจะทำการตัดไฟสัญญาณสีเขียวให้กับยานพาหนะ เพื่อไม่ให้ส่งผลกระทบต่อการทำงานของจราจร (กรมทางหลวง, 2566) แผนผังสรุปลำดับขั้นตอนพื้นฐานในการทำงานของทางข้ามถนนอัจฉริยะแสดงดังรูปที่ 2-1



รูปที่ 2-1 ตรรกะของการควบคุมไฟสัญญาณทางข้ามถนนอัจฉริยะ (Signal Control Logic)

(ที่มา: โครงการศึกษาจัดทำแผนพัฒนาศูนย์บริหารจราจรกลาง (HTOC) กรมทางหลวง ปี พ.ศ. 2567- พ.ศ. 2570 1 โครงการ, 2566)

2.1.4 ข้อมูลไม่สมดุล (Imbalanced Data)

ข้อมูลไม่สมดุล หมายถึง ข้อมูลที่มีจำนวนของข้อมูลในคลาสหนึ่งมากกว่าข้อมูลในอีกคลาสหนึ่งเป็นจำนวนมาก (Chawla, 2003) ซึ่งสาเหตุเกิดได้จากหลายปัจจัย อาทิเช่น อาจเกิดจากลักษณะทางธรรมชาติของตัวข้อมูลเอง อาจเกิดจากข้อจำกัดในการเก็บข้อมูล และอาจเกิดจากความผิดพลาดในการเก็บข้อมูล เป็นต้น จากข้อมูลทั้งหมด คลาสข้อมูลที่มีจำนวนของข้อมูลมากกว่า เรียกว่า คลาสส่วนมาก (Majority Class) และคลาสข้อมูลที่มีจำนวนของข้อมูลน้อยกว่า เรียกว่า คลาสส่วนน้อย (Minority Class) ซึ่งเมื่อชุดข้อมูลที่นำมาวิเคราะห์มีความไม่สมดุลของข้อมูลเกิดขึ้น ในกระบวนการของการสร้างแบบจำลองการเรียนรู้ (Training Model) ผลการวิเคราะห์ที่ได้ จะมีความโน้มเอียงไปในทิศทางของคลาสส่วนมาก ซึ่งจะส่งผลให้การจำแนกของคลาสส่วนน้อย มีประสิทธิภาพลดน้อยลง (กิริชาติ สุขสุทธิ์, 2559)

2.1.4.1 ระดับของความไม่สมดุล

ระดับของความไม่สมดุล คือ อัตราส่วนของจำนวนข้อมูลในคลาสส่วนมาก เทียบกับจำนวนข้อมูลในคลาสส่วนน้อย ซึ่งเป็นการจำแนกว่าชุดข้อมูลใดมีความสมดุลหรือไม่สมดุล ยิ่งชุดข้อมูลมีค่าระดับความไม่สมดุลมากกว่า 1 มากเท่าไร นั่นหมายความว่าชุดข้อมูลมีความไม่สมดุลสูงเท่านั้น หากข้อมูลมีระดับของความไม่สมดุลเท่ากับ 1 หมายความว่าในแต่ละคลาสมีจำนวนของข้อมูลที่เท่ากัน และหากข้อมูลมีระดับของความไม่สมดุลต่ำกว่า 1 หมายความว่าจำนวนข้อมูลของคลาสส่วนน้อยมากกว่าจำนวนข้อมูลในคลาสส่วนมาก (กิริชาติ สุขสุทธิ, 2559) การคำนวณระดับความไม่สมดุลสามารถคำนวณได้จากสมการที่ 2-1

$$\text{Imbalance Ratio (IR)} = \frac{n_{\text{majority}}}{n_{\text{minority}}} \quad \text{สมการที่ 2-1}$$

โดยที่ n_{majority} คือ จำนวนของข้อมูลในคลาสส่วนมาก
 n_{minority} คือ จำนวนของข้อมูลในคลาสส่วนน้อย

2.1.4.2 การจัดการกับข้อมูลไม่สมดุล

การแก้ไขปัญหาข้อมูลไม่สมดุลในระดับข้อมูล เป็นการแก้ไขและปรับปรุงกับข้อมูลโดยตรงด้วยเทคนิคการเลือกข้อมูล (Data Selection Technique) หรือเทคนิคการสุ่มเลือกข้อมูล (Data Sampling Technique) เพื่อให้ข้อมูลมีความสมดุลก่อนที่จะนำไปประมวลผล (Processing Stage) (กิริชาติ สุขสุทธิ, 2559) โดยเทคนิคการสุ่มเลือกข้อมูลโดยทั่วไปแบ่งออกเป็น 3 กลุ่ม (Modeling with R, 2019) ได้แก่

1) วิธีการสุ่มเกิน (Over Sampling) เป็นเทคนิคในการเพิ่มจำนวนข้อมูลของคลาสส่วนน้อย โดยการสุ่มเลือกจากข้อมูลเดิมหรือสร้างข้อมูลขึ้นมาใหม่จากตัวอย่างเดิมให้มีจำนวนใกล้เคียงหรือเท่ากับจำนวนข้อมูลในคลาสส่วนมาก โดยการสุ่มเลือกข้อมูลจากข้อมูลเดิม หรือสร้างข้อมูลขึ้นมาใหม่จากตัวอย่างของข้อมูลเดิม

2) วิธีการสุ่มลด (Under Sampling) เป็นเทคนิคในการลดจำนวนข้อมูลของคลาสส่วนมากลง ให้มีจำนวนใกล้เคียงหรือเท่ากับจำนวนข้อมูลในคลาสส่วนน้อย

3) วิธีผสมผสาน (Hybrid Methods) เป็นเทคนิคที่ทำงานร่วมกันระหว่างวิธีการสุ่มเกินในการเพิ่มจำนวนข้อมูลของคลาสส่วนน้อย และวิธีการสุ่มลดในการลดจำนวนข้อมูลของคลาสส่วนมาก เพื่อให้จำนวนข้อมูลของทั้งสองคลาสมีจำนวนข้อมูลที่เท่ากันหรือใกล้เคียงกัน วิธีผสมผสานที่รู้จักกันอย่างแพร่หลาย ได้แก่

Synthetic Minority Oversampling Technique (SMOTE) เป็นเทคนิคที่นำตัวอย่างจากกลุ่มข้อมูลคลาสส่วนน้อยมาพิจารณาทีละตัวจนครบทุกตัว โดยการกำหนดจำนวนเพื่อนบ้านที่ใกล้ที่สุด k ตัว แล้วจึงทำการสุ่มข้อมูลใหม่ในพื้นที่ใดๆ บนทางที่เชื่อมโยงระหว่าง

จุดข้อมูลที่กำลังพิจารณาและจุดข้อมูลเพื่อนบ้านที่ใกล้ที่สุด ซึ่งสามารถกำหนดได้ว่าจะสุ่มสร้างข้อมูลขึ้นมาใหม่มาอย่างน้อยเพียงใด

Random Over-Sampling Examples (ROSE) เป็นเทคนิคที่นำเอาตัวอย่างจากกลุ่มข้อมูลคลาสส่วนน้อยมาพิจารณาทีละตัวจนครบทุกตัว โดยการกำหนดจำนวนเพื่อนบ้านที่ใกล้ที่สุด k ตัว แล้วจึงทำการสุ่มข้อมูลใหม่บนพื้นที่รอบ ๆ ตัวอย่างกลุ่มข้อมูลคลาสส่วนน้อย ด้วยหลักการสุ่มหยิบแบบยอมรับการหยิบซ้ำ (Bootstrap) ซึ่งจะเป็นการเติมเต็มข้อมูลเข้าไปในช่วงของข้อมูลมากกว่าและน้อยกว่าเพื่อให้เกิดความราบเรียบของข้อมูล ก่อนที่จะมีการสร้างข้อมูล ซึ่งค่าที่ได้จะมีค่าอยู่ภายในช่วงเดียวกับตัวอย่างที่กำลังพิจารณา

2.1.4.3 การประเมินประสิทธิภาพสำหรับข้อมูลที่ไม่สมดุล

ในกรณีที่ข้อมูลเกิดปัญหาข้อมูลไม่สมดุล (Imbalanced Data) จำเป็นต้องมีการประเมินประสิทธิภาพในการจำแนกด้วยวิธีอื่น ๆ เพิ่มเติม นอกเหนือจากการประเมินประสิทธิภาพด้วยค่าความแม่นยำในการจำแนก (Accuracy) เนื่องจากจำนวนข้อมูลของทั้งสองคลาสมีความแตกต่างกันสูง หากผลของการพยากรณ์ข้อมูลทดสอบเป็นคลาสส่วนมากทั้งหมด ผลของค่าความแม่นยำในการจำแนกก็จะสูงด้วยเช่นกัน (กิริชาติ สุขสุทธิ, 2559)

1) เมตริกซ์วัดประสิทธิภาพ (Confusion Matrix) เป็นเมตริกซ์ที่นำมาใช้ในกรณีที่ต้องการประเมินประสิทธิภาพการพยากรณ์คลาสส่วนน้อยเป็นหลักซึ่งจะแสดงผลของการจำแนกข้อมูลที่ได้จากการทดสอบออกเป็นแต่ละคลาส โดยใช้ค่าทางสถิติดังนี้: True Positive (TP), True Negative (TN), False Positive (FP) และ False Negative (FN) ลักษณะของเมตริกซ์แสดงดังตารางที่ 2-1 สามารถนำไปคำนวณเพื่อสร้างตัวชี้วัดที่ใช้ในการตรวจสอบข้อมูลของแบบจำลอง และการประเมินประสิทธิภาพของแบบจำลอง (เบญจพร เอี่ยมประโคน 2560; กิริชาติ สุขสุทธิ, 2559)

ตารางที่ 2-1 เมตริกซ์วัดประสิทธิภาพ (Confusion Matrix)

	Predicted Positive	Predicted Negative
Actual positive	True Positive (TP)	False Negative (FN)
Actual negative	False Positive (FP)	True Negative (TN)

จากตารางที่ 2-1 แถวของเมตริกซ์จะแสดงจำนวนข้อมูลจริงในแต่ละคลาส และคอลัมน์ของเมตริกซ์จะแสดงจำนวนที่แบบจำลองพยากรณ์ได้ในแต่ละคลาส แบ่งออกเป็น 4 กรณีดังนี้

True Positive (TP) คือ เหตุการณ์ที่เกิดขึ้นจริงมีผลเป็นบวก (คลาสส่วนน้อย) และผลจากการพยากรณ์เป็นบวก (คลาสส่วนน้อย)

True Negative (TN) คือ เหตุการณ์ที่เกิดขึ้นจริงมีผลเป็นลบ (คลาสส่วนมาก) และผลจากการพยากรณ์เป็นลบ (คลาสส่วนมาก)

False Positive (FP) คือ เหตุการณ์ที่เกิดขึ้นจริงมีผลเป็นลบ (คลาสส่วนมาก) และผลจากการพยากรณ์เป็นบวก (คลาสส่วนน้อย)

False Negative (FN) คือ เหตุการณ์ที่เกิดขึ้นจริงมีผลเป็นบวก (คลาสส่วนน้อย) และผลจากการพยากรณ์เป็นลบ (คลาสส่วนมาก)

ค่าความแม่นยำในการจำแนก (Accuracy) คือ การประเมินประสิทธิภาพการจำแนกโดยรวมของทุกคลาสของแบบจำลอง สมการในการคำนวณแสดงดังสมการที่ 2-2

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \quad \text{สมการที่ 2-2}$$

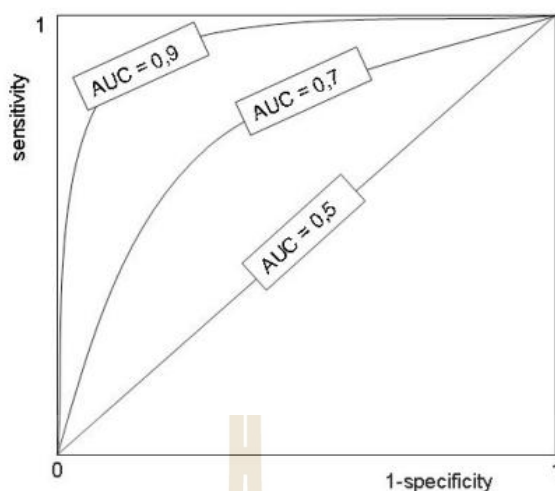
Sensitivity or True positive rate (TPR) คือ อัตราส่วนของจำนวนที่ผลจากการพยากรณ์เป็นบวก (คลาสส่วนน้อย) และเหตุการณ์ที่เกิดขึ้นจริงมีผลเป็นบวก (คลาสส่วนน้อย) ต่อจำนวนเหตุการณ์ที่เกิดขึ้นจริงมีผลเป็นบวก (คลาสส่วนน้อย) ทั้งหมด สมการในการคำนวณแสดงดังสมการที่ 2-3

$$Sensitivity = \frac{TP}{TP + FN} \quad \text{สมการที่ 2-3}$$

Specificity or True negative rate (TNR) คือ อัตราส่วนของจำนวนที่ผลจากการพยากรณ์เป็นลบ (คลาสส่วนมาก) และเหตุการณ์ที่เกิดขึ้นจริงมีผลเป็นลบ (คลาสส่วนมาก) ต่อจำนวนเหตุการณ์ที่เกิดขึ้นจริงมีผลเป็นลบ (คลาสส่วนมาก) ทั้งหมด สมการในการคำนวณแสดงดังสมการที่ 2-4

$$Specificity = \frac{TN}{FP + TN} \quad \text{สมการที่ 2-4}$$

2) Receiver Operating Characteristic curve (ROC curve) เป็นกราฟที่แสดงถึงประสิทธิภาพของ Binary classification โดยที่แนวแกน x คือค่า Sensitivity และแนวแกน y คือค่า 1-Specificity ที่จุดตัดต่าง ๆ เพื่อแบ่งผลลัพธ์การพยากรณ์ออกเป็น 2 กลุ่ม ได้แก่ กลุ่มที่เกิดเหตุการณ์ที่สนใจและกลุ่มที่ไม่เกิดเหตุการณ์ที่สนใจ ดังแสดงในรูปที่ 2-2 เพื่อบ่งชี้ถึงประสิทธิภาพหรือความน่าเชื่อถือของตัวแบบ โดยตัวแบบใดที่มีพื้นที่ใต้โค้ง ROC มากที่สุด ถือว่าเป็นตัวแบบที่ดีที่สุด (เบญจพร เอี่ยมประโคน, 2560) ยิ่งเส้นโค้งอยู่ใกล้กับมุมซ้ายบนมากเท่าไร หมายความว่าตัวแบบจะยังสามารถจำแนกกลุ่มได้ดีมากยิ่งขึ้นเท่านั้น ค่า AUC จะมีค่าอยู่ระหว่าง 0 ถึง 1 ตัวแบบที่สมบูรณ์แบบจะมีค่า AUC = 1.0 เกณฑ์ในการวัดประสิทธิภาพของตัวแบบจากพื้นที่ใต้โค้ง (Šimundić, 2009) แสดงดังตารางที่ 2-2



รูปที่ 2-2 แสดงเส้นโค้ง ROC curve
ตารางที่ 2-2 ประสิทธิภาพของตัวแบบจากพื้นที่ใต้โค้ง ROC

พื้นที่	ความแม่นยำ
0.9 – 1.0	ยอดเยี่ยม
0.8 – 0.9	ดีมาก
0.7 – 0.8	ดี
0.6 – 0.7	พอใช้
0.5 – 0.6	แย่
< 0.5	การทดสอบไม่มีประโยชน์

2.1.5 การทดสอบความสัมพันธ์ระหว่างตัวแปร

2.1.5.1 สหสัมพันธ์ (Correlation)

สหสัมพันธ์ (Correlation) เป็นสถิติที่ใช้เพื่อหาค่าความสัมพันธ์เชิงเส้นตรง (Linear Correlation) ระหว่างสองตัวแปร เพื่อป้องกันการมีความสัมพันธ์กันเองสูงระหว่างตัวแปรอิสระ (Multicollinearity) ซึ่งอาจส่งผลให้สมการตัวแบบที่ใช้ในการพยากรณ์ตัวแปรตามมีความคลาดเคลื่อน โดยกำหนดสมมติฐานทางสถิติ ได้ดังนี้

H_0 : ตัวแปรอิสระสองตัวไม่มีความสัมพันธ์กัน

H_1 : ตัวแปรอิสระสองตัวมีความสัมพันธ์กัน

ที่ระดับนัยสำคัญ 0.05

ค่าสัมประสิทธิ์สหสัมพันธ์ แทนด้วยสัญลักษณ์ r มีค่าอยู่ระหว่าง 0 ถึง 1 ยิ่งค่าสัมประสิทธิ์สหสัมพันธ์มีค่าเข้าใกล้ 1 มากเท่าไร แสดงว่าตัวแปรอิสระคู่นั้นมีความสัมพันธ์กันสูงมากเท่านั้น ค่าสหสัมพันธ์เชิงบวกและค่าสหสัมพันธ์เชิงลบในค่าเดียวกันจะมีความเข้มข้นเท่ากัน

ค่าสหสัมพันธ์เชิงบวก (positive correlation) ระหว่างตัวแปร ($r > 0$) หมายความว่า เมื่อค่าของตัวแปรหนึ่งเพิ่มขึ้น ค่าของอีกตัวแปรก็จะเพิ่มขึ้นด้วย ค่าสหสัมพันธ์เชิงลบ (negative correlation) ระหว่างตัวแปร ($r < 0$) หมายความว่า เมื่อค่าของตัวแปรหนึ่งลดลง ค่าของอีกตัวแปรก็จะลดลงด้วย (วีระศักดิ์ จินารัตน, 2557; อโนทัย ตรีวานิช, 2552) เกณฑ์ในการพิจารณาค่า r มีดังนี้

1. หาก $r < 0.20$ หมายความว่า ตัวแปรมีความสัมพันธ์กันเองน้อย
2. หาก $0.21 < r \leq 0.40$ หมายความว่า ตัวแปรมีความสัมพันธ์กันเองค่อนข้างน้อย
3. หาก $0.41 < r \leq 0.60$ หมายความว่า ตัวแปรมีความสัมพันธ์กันเองปานกลาง
4. หาก $0.61 < r \leq 0.80$ หมายความว่า ตัวแปรมีความสัมพันธ์กันเองค่อนข้างสูง
5. หาก $r > 0.80$ หมายความว่า ตัวแปรมีความสัมพันธ์กันเองสูง

2.1.6 การถดถอยโลจิสติก (Logistic Regression Analysis)

การถดถอยโลจิสติก มีวัตถุประสงค์เพื่อศึกษาว่าตัวแปรอิสระใดบ้างที่ส่งผลต่อตัวแปรตามอย่างมีนัยสำคัญทางสถิติซึ่งเป็นตัวแปรแบบทวิหรือตัวแปรแบบพหุกลุ่ม (ยุทธ ไกยวรรณ, 2555)

2.1.6.1 การถดถอยโลจิสติกแบบทวิ (Binary logistic regression)

การถดถอยโลจิสติกแบบทวิ เป็นการวิเคราะห์การถดถอยที่ตัวแปรตาม (y) มี 2 ค่า คือ เกิดเหตุการณ์ที่สนใจ ($y = 1$) หรือไม่เกิดเหตุการณ์ที่สนใจ ($y = 0$) มีความสัมพันธ์กับตัวแปรอิสระ (x) ไม่อยู่ในรูปเชิงเส้น ซึ่งในกรณีที่ตัวแปรอิสระมีมากกว่า 1 ตัว ในการวิเคราะห์จะได้ฟังก์ชันดังนี้

$$P_y = \frac{e^{b_0 + b_1x_1 + \dots + b_px_p}}{1 + e^{b_0 + b_1x_1 + \dots + b_px_p}} \quad \text{สมการที่ 2-5}$$

โดยที่ P_y คือ ความน่าจะเป็นของการเกิดเหตุการณ์ที่สนใจ

และ Q_y คือ ความน่าจะเป็นของการไม่เกิดเหตุการณ์ที่สนใจ ดังนี้

$$Q_y = 1 - P_y \quad \text{สมการที่ 2-6}$$

ผลการวิเคราะห์การถดถอยโลจิสติกแบบทวิ (Binary Logistic Regression) พบว่า ความสัมพันธ์ระหว่างตัวแปรอิสระและตัวแปรตามไม่อยู่ในรูปเชิงเส้น ดังนั้นจึงจำเป็นต้องปรับความสัมพันธ์ให้อยู่ในรูปเชิงเส้นผ่านการแปลงให้อยู่ในรูปของอัตราส่วนระหว่างโอกาส (Odds Ratio)

อัตราส่วนระหว่างโอกาส (Odds Ratio) เป็นค่าที่สะท้อนถึงการเปลี่ยนแปลงที่คาดหวังในความน่าจะเป็นของผลลัพธ์ เมื่อค่าของตัวแปรทำนายเพิ่มขึ้นหนึ่งหน่วย โดยค่า Odds Ratio นี้แสดงในรูปของค่าที่ได้จากการแปลงเอ็กซ์โปเนนเชียลของสัมประสิทธิ์ถดถอย ($\text{Exp}(B)$)

$$\text{Odds Ratio (OR)} = \frac{P_y}{Q_y} \quad \text{สมการที่ 2-7}$$

2.1.6.2 การคัดเลือกตัวแปรอิสระเข้าในสมการ

ในการเลือกตัวแปรอิสระเข้าสมการเพื่อให้ได้มาซึ่งตัวแบบที่สามารถพยากรณ์โอกาสที่จะเกิดนั้นใกล้เคียงกับความเป็นจริง (ประยูรศรี บุตรแสนคม, 2555) มีวิธีทั้งหมด 4 วิธี ดังนี้

1) การเลือกตัวแปรโดยวิธีเพิ่มตัวแปร (Forward Selection) เป็นวิธีการที่เลือกเฉพาะตัวแปรอิสระที่ดีที่สุดที่สามารถพยากรณ์ตัวแปรตามได้เท่านั้น โดยทำการคัดเลือกตัวแปรอิสระเข้าสมการทีละตัวตามลำดับของความสำคัญ และทำการทดสอบว่าตัวแปรที่เข้ามานั้นสามารถพยากรณ์ตัวแปรตามได้เพิ่มขึ้นอย่างมีนัยสำคัญทางสถิติหรือไม่ จากนั้นใช้วิธีพิจารณาเช่นเดียวกันคัดเลือกตัวแปรที่สำคัญรองลงมาต่อไปเรื่อย ๆ จนกระทั่งไม่มีตัวแปรตามเหลือ จึงจะถือเป็นการสิ้นสุดของวิธีการเพิ่มตัวแปร

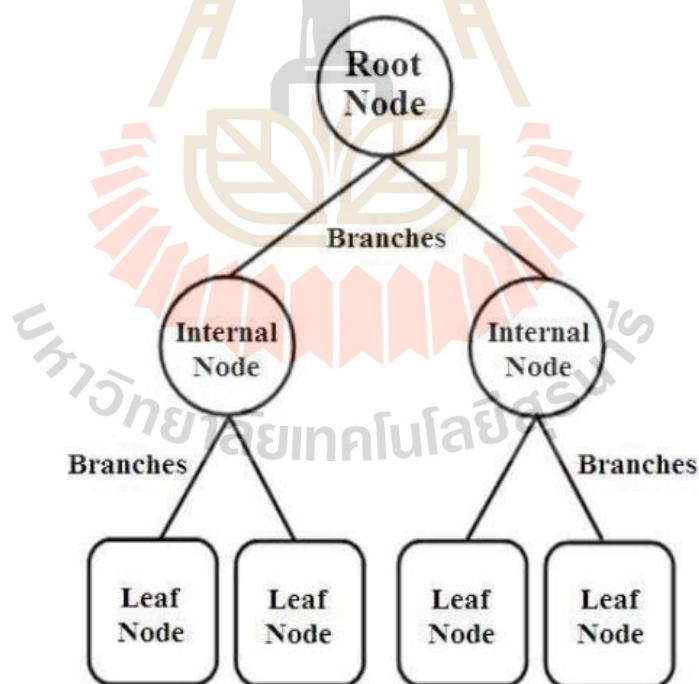
2) การเลือกตัวแปรโดยวิธีลดตัวแปร (Backward Elimination) เป็นวิธีการที่เลือกเฉพาะตัวแปรอิสระที่ดีที่สุดที่สามารถพยากรณ์ตัวแปรตามได้เท่านั้น แต่มีวิธีที่ตรงข้ามกับวิธีเพิ่มตัวแปร (Forward Selection) โดยในขั้นตอนแรกจะนำตัวแปรอิสระทุกตัวเข้ามาในสมการ และทำการพิจารณาตัวแปรอิสระที่มีค่าความสัมพันธ์กับตัวแปรตามต่ำที่สุดออกจากสมการ และทดสอบดำเนินการทดสอบ ค่า R^2 หากพบว่ามีค่าลดลงอย่างไม่มีนัยสำคัญทางสถิติ แสดงว่าตัวแปรดังกล่าวไม่ได้ทำให้การพยากรณ์ตัวแปรตามเพิ่มขึ้นสามารถขจัดออกจากสมการได้ และใช้วิธีพิจารณาเช่นเดียวกันขจัดตัวแปรอิสระที่มีความสำคัญน้อยรองลงมาออกจากสมการ การขจัดตัวแปรอิสระจะสิ้นสุดก็ต่อเมื่อพบว่า มีผลทำให้ค่า R^2 ลดลงอย่างมีนัยสำคัญทางสถิติ แสดงว่าตัวแปรดังกล่าวมีความสำคัญต่อการพยากรณ์ตัวแปรตาม ให้คงตัวแปรอิสระดังกล่าวไว้ในสมการหากขจัดตัวแปรออกจากสมการจะทำให้ประสิทธิภาพในการพยากรณ์ตัวแปรตามลดลง

3) การเลือกตัวแปรโดยวิธีเพิ่มตัวแปรอิสระแบบขั้นตอน (Stepwise Regression) เป็นวิธีที่มีขั้นตอนคล้ายกับวิธีเพิ่มตัวแปร (Forward Selection) แต่จะทำการทดสอบตัวแปรอิสระที่เข้าไปในสมการแล้วทุกครั้งที่มีการนำตัวแปรอิสระตัวใหม่เข้าไปในสมการ ดังนั้นตัวแปรอิสระที่เข้าไปในสมการแล้วก็จะสามารถถูกคัดออกจากสมการได้ หากพบว่าตัวแปรอิสระตัวนั้น ๆ ไม่ได้ส่งผลให้ค่า R^2 เพิ่มขึ้นอย่างมีนัยสำคัญทางสถิติ

4) การเลือกตัวแปรโดยวิธีนำตัวแปรเข้าทั้งหมด (Enter Regression) วิธีนี้เป็นการคัดเลือกตัวแปรอิสระเพียงขั้นตอนเดียว โดยจะนำตัวแปรอิสระทุกตัวเข้าสู่สมการ แม้ว่าตัวแปรอิสระนั้น ๆ มีนัยสำคัญทางสถิติหรือไม่ก็ตาม มักใช้ในกรณีที่ต้องการทราบถึงประสิทธิภาพการพยากรณ์ตัวแปรตามของตัวแปรอิสระแต่ละตัว ดังนั้นวิธีนี้จึงมีข้อเสียคือ เป็นวิธีการวิเคราะห์ที่ไม่ได้คัดเลือกตัวแปรอิสระเข้าสู่สมการถดถอยที่เหมาะสมให้

2.17 ต้นไม้ตัดสินใจ (Decision Tree)

ต้นไม้ตัดสินใจ (Decision Tree) เป็นการสร้างแบบจำลองในลักษณะของโครงสร้างต้นไม้ โดยทำการนำคุณลักษณะ (Attribute) ต่าง ๆ ของข้อมูลนั้นไปเทียบกับเส้นทางในต้นไม้ ชุดข้อมูลจะค่อย ๆ ถูกแบ่งย่อยออกเป็นชุดย่อยเล็ก ๆ จนกระทั่งคลาสปลายทาง ภายในต้นไม้จะประกอบไปด้วยโหนด (Node) โดยแต่ละโหนดจะมีคุณลักษณะเป็นตัวทดสอบ โหนดที่อยู่ตำแหน่งบนสุดเรียกว่า โหนดราก (root node), กิ่งของต้นไม้ (Branch) แสดงถึงค่าที่เป็นไปได้ของคุณลักษณะที่ถูกเลือกทดสอบ และใบ (Leaf) คือผลลัพธ์ของการพยากรณ์แสดงถึงกลุ่มของข้อมูล (Class) จะอยู่ตำแหน่งล่างสุดของต้นไม้ตัดสินใจ (ณัฐภัทร ฤทธิ์เดช, 2565) โครงสร้างของต้นไม้ตัดสินใจแสดงดังรูปที่ 2-3



รูปที่ 2-3 โครงสร้างต้นไม้ตัดสินใจ (Mienye & Jere, 2024)

2.2 งานวิจัยที่เกี่ยวข้อง

Raoniar and Maurya (2022b) การศึกษานี้มุ่งเน้นไปที่ปัจจัยข้อมูลทางสังคมและข้อมูลที่ไม่ใช่ข้อมูลทางสังคม ที่ส่งผลต่อการฝ่าฝืนไฟสัญญาณสีแดงของคนเดินข้ามที่ทางแยกที่มีไฟสัญญาณเมืองโกลกาตา ประเทศอินเดีย ทำการสำรวจทั้งหมด 3 แห่ง เก็บรวบรวมข้อมูลของคนเดินข้ามด้วยการตั้งกล้องถ่ายวิดีโอ และทำการวิเคราะห์ด้วยวิธี Binary logistic regression ผลการศึกษาเผยให้เห็นว่าข้อมูลที่ไม่ใช่ทางสังคม ได้แก่ ความเร็วในการข้าม เวลารอสำหรับการข้าม ปริมาณจราจร และความยาวของรอบสัญญาณไฟ มีอิทธิพลอย่างมากต่อพฤติกรรมการละเมิดไฟสัญญาณ และจากข้อมูลทางสังคมพบว่า จำนวนคนที่รอที่ทางแยก รวมถึงคนเดินข้ามที่รอข้ามถนนเมื่อเห็นคนอื่นในทิศทางเดียวกันหรือฝั่งตรงข้ามฝ่าฝืนไฟสัญญาณสีแดงสำเร็จนั้น มีอิทธิพลต่อพฤติกรรมการฝ่าฝืนไฟสัญญาณสีแดง และนอกจากนี้ผลลัพธ์ของแบบจำลองยังชี้ให้เห็นว่า หากคนเดินเท้าต้องรอนานขึ้นเพื่อข้ามถนนอย่างปลอดภัย พวกเขาจะขาดความอดทนและทำการละเมิดไฟสัญญาณในที่สุด ผลการค้นพบในปัจจุบันชี้ให้เห็นถึงข้อบกพร่องในการออกแบบไฟสัญญาณจราจรและการวางแผนโครงสร้างพื้นฐาน

Raoniar, Maqbool, Pathak, Chugh and Maurya (2022a) การศึกษานี้มุ่งเน้นไปที่การทำความเข้าใจความเสี่ยงจากการข้ามถนนของคนเดินเท้าที่ทางม้าลายที่มีสัญญาณไฟจราจร 8 แห่ง เมืองโกลกาตา ประเทศอินเดีย สังเกตคนเดินเท้าที่มาถึงทางแยกไฟแดงทั้งหมด 2,089 ราย โดยใช้ Semi-parametric Cox Proportional Hazard model เพื่อทำความเข้าใจปัจจัยต่างๆ ที่มีอิทธิพลต่อพฤติกรรมการละเมิดไฟสัญญาณจราจร และใช้แบบจำลอง Accelerated Failure Time (AFT) เพื่อทำความเข้าใจตัวแปรร่วมต่าง ๆ ที่มีผลต่อระยะเวลาการรอคอย ผลการศึกษาพบว่าคนเดินถนน 49.5% ข้ามถนนทันทีหลังจากมาถึงทางแยกในช่วงไฟสัญญาณสีแดง (ระยะเวลารอคอยน้อยกว่า 3 วินาที) บ่งชี้การตัดสินใจข้ามถนนที่กำหนดไว้ล่วงหน้าหรือความคุ้นเคยกับทางแยก ความน่าจะเป็นของคนเดินถนนที่จะข้ามถนนเมื่อไม่ปลอดภัย เช่น ขณะยานพาหนะยังติดไฟสัญญาณสีเขียวหรือสีเหลืองซึ่งจะแตกต่างกันไปตามระยะเวลารอคอย เมื่อระยะเวลารอคอยเพิ่มมากขึ้น คนเดินเท้าจะใจร้อนและตัดสินใจฝ่าฝืนไฟสัญญาณจราจร ทำให้มีความเสี่ยงที่จะทำให้เกิดอุบัติเหตุยานพาหนะชนคนเดินเท้าเพิ่มมากยิ่งขึ้น ซึ่งในช่วงเวลา 36.705 วินาที คนเดินเท้า 75% จะฝ่าฝืนไฟสัญญาณ ซึ่งแสดงให้เห็นถึงข้อบกพร่องในการออกแบบไฟสัญญาณและการวางแผนโครงสร้างพื้นฐาน การวิเคราะห์ตัวแปรร่วมโดยใช้แบบจำลอง AFT แสดงให้เห็นว่าพฤติกรรมการเลื้อมอง/การมองของคนเดินเท้า สิ่งรบกวนต่าง ๆ ความยาวรอบไฟสัญญาณ การถือสัมภาระ และปริมาณจราจร มีผลกระทบต่อนพฤติกรรมการฝ่าฝืนไฟสัญญาณของคนเดินเท้า การปรับไฟสัญญาณโดยลดความยาวของช่วงไฟสัญญาณสีแดงสำหรับคนเดินเท้า อาจเป็นวิธีที่มีประสิทธิภาพที่สุดในการลดความน่าจะเป็น

เป็นที่คนเดินเท้าจะฝ่าฝืนไฟสัญญาณ และลดความน่าจะเป็นที่จะเกิดอุบัติเหตุยานพาหนะชนคนเดินเท้า

Zhang and Deng (2019) ได้ศึกษาพฤติกรรมการฝ่าไฟแดงของคนเดินเท้าที่ทางแยกมีสัญญาณไฟในเมืองกวางโจว ประเทศจีน โดยเก็บข้อมูลจากทางแยก 3 แห่งผ่านการบันทึกวิดีโอและจำแนกพฤติกรรมการฝ่าฝืนออกเป็น 2 ประเภท ได้แก่ การฝ่าไฟแดงแบบเลียนแบบผู้อื่น (Conformity Violation) และแบบตั้งใจฝ่า (Active Violation) ดำเนินการวิเคราะห์ปัจจัยที่มีอิทธิพลจาก 3 กลุ่ม ได้แก่ ลักษณะส่วนบุคคล ลักษณะของกลุ่ม และสภาพแวดล้อมการจราจร รวม 10 ตัวแปร ผ่านการวิเคราะห์ด้วยวิธี Mixed Stepwise Selection และโมเดล Binary Logistic Regression ผลการศึกษาพบว่า ตัวแปรที่ส่งผลต่อพฤติกรรมการฝ่าไฟสัญญาณสีแดงอย่างมีนัยสำคัญ 5 ตัว ได้แก่ ระยะเวลารอคอย จำนวนคนที่รอข้าม จำนวนคนที่ฝ่าไฟแดง การแสดงเวลานับถอยหลังของไฟแดง และช่องว่างระหว่างรถ โดยที่ระยะเวลารอคอยและจำนวนคนที่รอข้ามถนนมีความสัมพันธ์ทางบวกกับโอกาสของการฝ่าไฟแดงแบบตั้งใจฝ่า จำนวนคนที่ฝ่าไฟแดงและช่องว่างระหว่างรถ มีความสัมพันธ์ทางบวกกับโอกาสของการฝ่าไฟแดงแบบเลียนแบบผู้อื่น ทั้งนี้การแสดงเวลานับถอยหลังมีอิทธิพลน้อย และมีแนวโน้มช่วยลดการฝ่าแบบเลียนแบบลงเล็กน้อย

Afshari and Ayati (2021) ศึกษาผลกระทบของปัจจัยภายนอกต่อการฝ่าฝืนไฟสัญญาณของคนเดินเท้าที่ทางแยกที่มีไฟสัญญาณ เมืองมัสฮัด ประเทศอิหร่าน ทำการสำรวจทั้งสิ้น 10 ทางข้ามจาก 6 ทางแยก ทำการวิเคราะห์ด้วยวิธี Binary logistic regression ตัวแปรอิสระที่พิจารณา ได้แก่ จำนวนคนที่ตัดสินใจฝ่าฝืนไฟสัญญาณร่วมกัน การถือสัมภาระ ปริมาณจราจรต่อช่องจราจรต่อนาที ปริมาณคนเดินข้าม สัดส่วนของคนเดินข้ามที่ฝ่าฝืนไฟสัญญาณ ความยาวของทางข้าม ระยะเวลาไฟสัญญาณสีแดง ประเภทของความขัดแย้งที่ทางข้าม ความปลอดภัยหลังการข้าม ผลการศึกษาบ่งชี้ว่า ปริมาณการจราจร จำนวนคนเดินข้ามที่ฝ่าฝืน ความยาวของทางข้าม ระยะเวลาไฟสัญญาณสีแดง และปัญหาการเคลื่อนไหวร่างกายของคนเดินข้าม ส่งผลต่อการตัดสินใจของคนเดินข้ามที่จะฝ่าฝืนไฟสัญญาณสีแดง กล่าวคือจำนวนคนเดินข้ามที่ฝ่าฝืนและระยะเวลาไฟสัญญาณสีแดง ส่งผลทางบวกต่อแนวโน้มของพฤติกรรมการฝ่าฝืนไฟสัญญาณสีแดง และในทางกลับกันการเพิ่มขึ้นของความยาวของทางม้าลายและปริมาณการจราจร ส่งผลในทางลบซึ่งจะช่วยลดความน่าจะเป็นของการฝ่าฝืนไฟสัญญาณสีแดง และนอกจากนี้ยังพบว่าคนเดินถนนที่มีความพิการทางร่างกายก็มีแนวโน้มที่จะฝ่าฝืนไฟสัญญาณสีแดงน้อยกว่าคนปกติ

Raoniar and Maurya (2022c) การศึกษานี้มุ่งเน้นไปที่ปัจจัยที่มีอิทธิพลต่อความเร็วในการข้ามถนนและพฤติกรรมของผู้ฝ่าฝืนไฟสัญญาณสีแดงที่ทางข้าม เมืองโกลกาตา ประเทศอินเดีย ทำการสำรวจทั้งสิ้น 3 แห่ง โดยได้ทำการสังเกตพฤติกรรมของคนเดินข้าม 2,360 คน ดำเนินการวิเคราะห์ข้อมูลโดยใช้วิธี OLS Regression สำหรับการวิเคราะห์ความเร็วและ safety margin และ

ใช้วิธี Binary Logistic Regression สำหรับการวิเคราะห์กรณีการฝ่าฝืนไฟสัญญาณ (violation) จากการศึกษาพบว่าปัจจัยส่วนบุคคล (วัย, การใช้อุปกรณ์, กิจกรรมขณะเดิน) มีผลต่อความเร็วและ safety margin การฝ่าฝืนไฟสัญญาณจะเพิ่มขึ้นตามระยะเวลาการรอคอยไฟสัญญาณสีแดงที่นานขึ้น กล่าวคือ ยิ่งระยะไฟแดงยาวขึ้น โอกาสที่คนเดินเท้าจะฝ่าฝืนสัญญาณก็เพิ่มขึ้นตาม

Zhu, Sze and Bai (2021) การศึกษานี้ตรวจสอบว่าปัจจัยส่วนบุคคล (ข้อมูลประชากรและพฤติกรรมของคนเดินข้าม) และสิ่งแวดล้อม (การมีอยู่และพฤติกรรมของคนเดินถนนคนอื่นๆ ระยะเวลาที่เปิดไฟสัญญาณ และสภาพการจราจร) มีอิทธิพลต่อการตัดสินใจฝ่าฝืนไฟสัญญาณสีแดงของคนเดินเท้าอย่างไร ดำเนินการสำรวจจากการสังเกตวิดีโอที่ทางข้ามที่มีไฟสัญญาณ ในเขตเมือง โดยได้รวบรวมพฤติกรรมการข้ามถนนของคนเดินถนนทั้งสิ้น 6,320 คน ดำเนินการวิเคราะห์ด้วยวิธี Binary logistic regression ผลลัพธ์แสดงให้เห็นว่าปัจจัยต่างๆ เช่น เพศ อายุ จำนวนช่องทางสำหรับคนเดินข้าม การมีเพื่อนร่วมทาง จำนวนคนเดินข้ามในบริเวณใกล้เคียง การมีผู้ฝ่าฝืนไฟสัญญาณสีแดงคนอื่นในรอบเดียวกัน ระยะเวลาไฟสัญญาณสีเขียว ระยะเวลาไฟสัญญาณสีแดง ปริมาณจราจร และเปอร์เซ็นต์ของยานพาหนะหนัก ล้วนส่งผลกระทบต่อแนวโน้มของการฝ่าฝืนไฟสัญญาณสีแดงของคนเดินเท้า

Rafe et al. (2024) การศึกษานี้ศึกษาพฤติกรรมการข้ามถนนของคนเดินเท้าที่ทางแยกที่มีไฟสัญญาณจราจร ในรัฐยูทาห์ ประเทศสหรัฐอเมริกา โดยเฉพาะพฤติกรรมการฝ่าฝืนไฟสัญญาณจราจรทั้งในมิติตำแหน่ง (spatial) และมิติเวลา (temporal) การศึกษานี้ใช้ข้อมูลการสังเกตการณ์จริงจากกล้องวิดีโอจำนวน 5,589 เหตุการณ์ ใน 47 ทางม้าลายที่สี่แยก 39 แห่ง ดำเนินการวิเคราะห์โดยใช้โมเดลถดถอยโลจิสติกแบบหลายระดับ (multilevel logistic regression) เพื่อประเมินอิทธิพลของปัจจัยรายบุคคล สิ่งแวดล้อม และปัจจัยทางสังคม ที่ส่งผลต่อการละเมิดไฟสัญญาณของคนเดินข้าม ผลการศึกษาที่สำคัญบ่งชี้ว่าปัจจัยส่วนบุคคล เช่น เพศและการใช้อุปกรณ์ช่วยเดิน ปัจจัยด้านสภาพแวดล้อม เช่น อุณหภูมิและเวลาของวัน และปัจจัยทางสังคม เช่น การมีผู้คนเดินเท้าคนอื่นๆ อยู่ด้วย มีผลกระทบต่อพฤติกรรมการละเมิดกฎจราจรอย่างมีนัยสำคัญ คนเดินข้ามที่เป็นผู้หญิงและใช้อุปกรณ์ช่วยเดินมักจะปฏิบัติตามกฎจราจรได้ดีกว่า ส่วนปัจจัยแวดล้อม เช่น อุณหภูมิสูง ช่วงเวลากลางคืน หรือช่วงเข้ามีด ส่งผลให้คนเดินข้ามมีโอกาที่จะฝ่าฝืนไฟสัญญาณเพิ่มสูงขึ้น และในส่วนของปัจจัยทางสังคม เช่น การมีคนอื่นข้ามพร้อมกัน หรือการรอลานานขึ้น มีแนวโน้มช่วยลดโอกาสที่คนเดินข้ามจะฝ่าฝืนไฟสัญญาณจราจรลงได้

วีรพงศ์ บุญเจริญปัญญา และจรัส พัทธกษัตริย์ (2567) งานวิจัยนี้มุ่งศึกษาทัศนคติของคนเดินเท้าในการข้ามถนน ตลอดจนปัจจัยที่มีอิทธิพลต่อทัศนคติเหล่านั้น กรณีสึกษาทางข้ามถนนบริเวณถนนแจ้งวัฒนะ ถนนลาดกระบัง และถนนกิ่งแก้ว โดยใช้แบบสอบถามจำนวน 2 ส่วน ได้แก่ ส่วนที่เกี่ยวกับทัศนคติในการข้ามถนน และส่วนที่เกี่ยวกับทัศนคติด้านความปลอดภัยในการข้ามถนน

รวมทั้งสิ้นจำนวน 412 ชุด ตัวแปรที่ใช้ในการศึกษาได้แก่ ตัวแปรทางสังคม ตัวแปรด้านความถี่ต่อประสบการณ์ที่ไม่ดีในการใช้งานทางข้ามถนน ตัวแปรด้านทัศนคติในการใช้งานทางข้ามถนน จากสถานที่ตัวอย่าง และตัวแปรด้านทัศนคติที่มีต่อสิ่งอำนวยความสะดวกในการข้ามถนน ในการวิเคราะห์ข้อมูล ใช้สถิติเชิงพรรณนาและการวิเคราะห์ถดถอยเชิงเส้น (Linear Regression Analysis) ผลการศึกษาพบว่า จำนวนช่องจราจร ปริมาณจราจร และความเร็วของยานพาหนะ ส่งผลเสียต่อความรู้สึกปลอดภัยในการข้ามถนนอย่างมีนัยสำคัญ ในด้านของปัจจัยทางสังคมพบว่า ปัจจัยเพศ เป็นปัจจัยหลักที่ส่งผลต่อความรู้สึกปลอดภัยอย่างมีนัยสำคัญ และผลการศึกษาในด้านของตัวแปรด้านทัศนคติที่มีต่อสิ่งอำนวยความสะดวกในการข้าม พบว่าคนเดินเท้ารู้สึกปลอดภัยเมื่อมีการเพิ่มสัญญาณไฟจราจรแบบปุ่มกดมากที่สุด รองลงมาเป็นการมีอยู่ของเกาะกลางถนน และการเพิ่มเนินชะลอความเร็ว (Rubble strip) ตามลำดับ

ชิตภัทร อินทรวินมธ และพงษ์ศักดิ์ สุริยวนากุล (2555) ได้ศึกษาความสัมพันธ์ระหว่างปัจจัยที่บ่งชี้ค่าความยากในการข้ามถนนกับระดับความยากในการข้ามถนนบริเวณทางแยก 2 ทางแยก ได้แก่ ทางแยกถนนถนนศรีนครินทร์-ถนนพัฒนาการ และทางแยกถนนพระราม 9 - ถนนวัฒนธรรม โดยทำการสุ่มตัวอย่างประชากรทั้งหมด 130 คน ซึ่งตัวแปรตามที่น่าสนใจได้แก่ ปัจจัยด้านการจราจร ปัจจัยด้านลักษณะทางกายภาพของทางแยก และปัจจัยทางเศรษฐกิจและสังคมของคนเดินเท้า โดยการสร้างแบบจำลองโพรบิตลำดับ (Ordinal/Ordered Probit Model) จากผลการวิเคราะห์พบว่า ตัวแปรเพศ ความถี่ในการข้ามถนน และความกว้างของเกาะกลางถนน ส่งผลในเชิงลบ กล่าวคือ หากตัวแปรกลุ่มนี้มีค่าเพิ่มมากขึ้นจะส่งผลทำให้ค่าความยากในการข้ามถนนบริเวณทางแยกของคนเดินเท้าต่ำ (ข้ามง่าย) ในทางกลับกันตัวแปรชนิดของเกาะกลาง จำนวนช่องจราจรรวมทั้งสองทิศทาง ปริมาณรถเลี้ยวขวาของการจราจรด้านใกล้ ส่งผลในเชิงบวก กล่าวคือ หากตัวแปรกลุ่มนี้มีค่าเพิ่มมากขึ้นจะส่งผลทำให้ค่าความยากในการข้ามถนนบริเวณทางแยกของคนเดินเท้าสูง (ข้ามยาก)

นพมาศ อัครจันทโชติ และดิเรก พินิตสุภาภมร (2561) การศึกษานี้ทำการเปรียบเทียบประสิทธิภาพการแก้ปัญหาข้อมูลไม่สมดุลด้วยวิธีการสุ่มตัวอย่างซ้ำระหว่างวิธีการสุ่มตัวอย่างเพิ่มข้อมูลเริ่มต้นอย่างสุ่ม (ROSE) และการสุ่มตัวอย่างเพิ่มกลุ่มส่วนน้อยด้วยการสังเคราะห์ (SMOTE) และทำการเปรียบเทียบประสิทธิภาพในการจำแนกข้อมูลระหว่างวิธีการถดถอยโลจิสติก และวิธีต้นไม้ตัดสินใจสำหรับการจำแนกกลุ่มรายได้ผู้ประกอบการร้านยาประเภท ข.ย. 1 จำนวน 111 ราย โดยทำการเปรียบเทียบประสิทธิภาพค่าความแม่นยำ อัตราความถูกต้องในการพยากรณ์กลุ่มส่วนน้อย อัตราความถูกต้องในการพยากรณ์กลุ่มส่วนมาก และค่าการวัดเอฟ จากผลของการเปรียบเทียบวิธีการจำแนกประเภทข้อมูล พบว่า SMOTE ให้ผลลัพธ์ที่ดีที่สุดในตัวชี้วัดเกือบทั้งหมด ยกเว้นความแม่นยำในการพยากรณ์กลุ่มที่มีจำนวนข้อมูลมาก ซึ่งยังด้อยกว่าวิธีอื่น

เมื่อเทียบกับข้อมูลเริ่มต้นและวิธี ROSE จากการพิจารณาค่าความแม่นยำ พบว่า วิธีต้นไม้ตัดสินใจ ให้ค่าความแม่นยำสูงกว่าวิธีถดถอยโลจิสติก อย่างไรก็ตาม เมื่อพิจารณาอัตราความถูกต้องในการทำนายกลุ่มส่วนน้อย ซึ่งเป็นกลุ่มเป้าหมายที่สำคัญ พบว่า หลังจากปรับขนาดข้อมูลแล้ว วิธีถดถอยโลจิสติกให้ผลลัพธ์ที่ดีกว่าหรือเทียบเท่ากับวิธีต้นไม้ตัดสินใจ

วิชญ์วิสิฐ เกษรสิทธิ์, วิชิต หล่อจิระชุมหกุล และจิราวัลย์ จิตรถเวช (2561) ได้ศึกษาเปรียบเทียบประสิทธิภาพของวิธีการแก้ปัญหาข้อมูลไม่สมดุลของข้อมูลผู้ป่วยโรคเบาหวาน โดยใช้ข้อมูลจากโรงพยาบาล 130 แห่งในประเทศสหรัฐอเมริกา โดยเปรียบเทียบวิธีการแก้ปัญหาข้อมูลไม่สมดุลที่ระดับข้อมูล 4 วิธี ได้แก่ การสุ่มเพิ่มข้อมูล, การสุ่มลดข้อมูล, การผสมผสานระหว่างการเพิ่มและลดข้อมูล และการสังเคราะห์ข้อมูลใหม่ด้วยเทคนิค (SMOTE) โดยใช้เทคนิคการจำแนกประเภทของข้อมูลสองวิธีหลัก ได้แก่ การถดถอยโลจิสติกแบบพหุนาม (Multinomial Logistic Regression) และวิธีต้นไม้ตัดสินใจ (Decision Tree) ผลการศึกษาพบว่า การแก้ปัญหาความไม่สมดุลของข้อมูลช่วยเพิ่มประสิทธิภาพในการจำแนกข้อมูลกลุ่มส่วนน้อย ซึ่งเป็นกลุ่มเป้าหมายที่สำคัญได้ถูกต้องและแม่นยำมากขึ้น

อัจฉรา แผ้วบาง และสายชล สันสมบุรณ์ทอง (2563) ได้ศึกษาเปรียบเทียบวิธีการแก้ปัญหาข้อมูลไม่สมดุล 4 วิธี คือ วิธีการสุ่มเกิน วิธีการสุ่มเกินโดยเทคนิค SMOTE วิธีการสุ่มลด และวิธีการสุ่มผสมผสาน โดยวิธีการจำแนก 5 วิธี คือ วิธีเพื่อนบ้านใกล้สุด k ตัว วิธีโครงข่ายประสาทเทียม วิธีซัพพอร์ตเวกเตอร์แมชชีน วิธีฐานกฎ และวิธีลาดลงสโตแคสติก พิจารณาจากค่าความถูกต้อง ค่าความไว ค่าความจำเพาะ ค่าคลาดเคลื่อนกำลังสองเฉลี่ย และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย จากชุดข้อมูลทั้งหมด 3 ชุด ได้แก่ ชุดข้อมูลเคมีบำบัดมะเร็งลำไส้ใหญ่ระยะ B/C ชุดข้อมูลโรคที่มีความผิดปกติของโปรตีน และชุดข้อมูลการรักษาอาการปวดศีรษะขั้นรุนแรง โดยแบ่งข้อมูลในอัตราส่วน 70, 20 และ 10 ตามลำดับ กำหนดตัวสร้างเลขสุ่มเทียม เป็น 10, 20, 30, 40 และ 50 จากผลการศึกษาของทั้ง 3 ชุดข้อมูลพบว่าวิธีที่มีประสิทธิภาพสูงสุดคือ วิธีฐานกฎโดยการปรับความไม่สมดุลด้วยวิธีการสุ่มเกินเทคนิค SMOTE

2.3 สรุปงานวิจัยที่เกี่ยวข้อง

จากการทบทวนงานวิจัยที่เกี่ยวข้อง เกี่ยวกับการศึกษาพฤติกรรมการฝ่าฝืนไฟสัญญาณของคนเดินข้ามบริเวณทางข้าม และการแก้ไขข้อมูลไม่สมดุล (Imbalanced Data) สามารถสรุปผลการทบทวนแสดงดังตารางที่ 2-3

ตารางที่ 2-3 สรุปงานวิจัยที่เกี่ยวข้อง

ผู้แต่ง	วัตถุประสงค์	วิธีการวิเคราะห์	ผลการศึกษา
Raoniar and Maurya (2022b)	- ศึกษาปัจจัยข้อมูลทางสังคมและข้อมูลที่ไม่ใช่ข้อมูลทางสังคมที่ส่งผลต่อการฝ่าฝืนไฟสัญญาณสีแดงของคนเดินข้ามที่ทางแยกที่มีไฟสัญญาณ	- วิธี Binary logistic regression	ข้อมูลที่ไม่ใช่ทางสังคม ได้แก่ ความเร็วในการข้ามเวลารอคอย ปริมาณจราจร และความยาวของรอบไฟสัญญาณ มีอิทธิพลอย่างมากต่อพฤติกรรมการฝ่าฝืนไฟสัญญาณของคนเดินข้าม ข้อมูลทางสังคม ได้แก่ จำนวนคนที่รอที่ทางแยกรวมถึงคนเดินข้ามที่รอข้ามถนนเมื่อเห็นคนอื่นในทิศทางเดียวกันหรือฝั่งตรงข้ามฝ่าฝืนไฟสัญญาณสีแดงสำเร็จ มีอิทธิพลต่อพฤติกรรมการฝ่าฝืนไฟสัญญาณสีแดง
Raoniar, Maqbool, Pathak, Chugh and Maurya (2022a)	- การศึกษานี้มุ่งเน้นการทำความเข้าใจความเสี่ยงจากการข้ามถนนของคนเดินเท้าที่ทางม้าลาย	- ใช้วิธี Semi-parametric Cox Proportional Hazard model เพื่อทำความเข้าใจปัจจัยต่างๆ ที่มีอิทธิพลต่อพฤติกรรมการละเมิดไฟสัญญาณจราจร - ใช้แบบจำลอง Accelerated Failure Time (AFT) เพื่อทำความเข้าใจตัวแปรร่วมต่างๆ ที่มีผลต่อระยะเวลาการรอคอย	ผลการศึกษาพบว่าเมื่อระยะเวลาการรอคอยเพิ่มมากขึ้น คนเดินเท้าจะใจร้อนและตัดสินใจฝ่าฝืนไฟสัญญาณจราจร ซึ่งในช่วงเวลา 36.705 วินาที คนเดินเท้า 75% จะฝ่าฝืนไฟสัญญาณ การวิเคราะห์ตัวแปรร่วมโดยใช้แบบจำลอง AFT แสดงให้เห็นว่าพฤติกรรมการเหลือบมอง/การมองของคนเดินเท้า สิ่งรบกวนต่างๆ ความยาวของรอบไฟสัญญาณ การถือสัมภาระ และปริมาณจราจร ล้วนมีผล

ตารางที่ 2-3 สรุปงานวิจัยที่เกี่ยวข้อง (ต่อ)

ผู้แต่ง	วัตถุประสงค์	วิธีการวิเคราะห์	ผลการศึกษา
			กระทบต่อพฤติกรรมฝ่าฝืนไฟสัญญาณของคนเดินเท้า
Zhang and Deng (2019)	- ศึกษาพฤติกรรมการฝ่าฝืนไฟสัญญาณสีแดงของคนเดินข้ามบริเวณทางแยกที่มีไฟสัญญาณ	- วิธี Mixed Stepwise Selection และโมเดล Binary Logistic Regression	ตัวแปรที่ส่งผลต่อพฤติกรรมการฝ่าไฟสัญญาณสีแดงอย่างมีนัยสำคัญ 5 ตัว ได้แก่ ระยะเวลารอคอย จำนวนคนที่รอข้าม จำนวนคนที่ฝ่าไฟแดง การแสดงเวลานับถอยหลังของไฟแดง และช่องว่างระหว่างรถ
Afshari and Ayati (2021)	- ศึกษาผลกระทบของปัจจัยภายนอกต่อการฝ่าฝืนไฟสัญญาณของคนเดินเท้าที่ทางแยกที่มีไฟสัญญาณ	- วิธี Binary logistic regression	ผลการศึกษาบ่งชี้ว่า จำนวนคนเดินข้ามที่ฝ่าฝืนและระยะเวลาไฟสัญญาณสีแดง ส่งผลทางบวกต่อแนวโน้มของพฤติกรรมการฝ่าฝืนไฟสัญญาณสีแดง และในทางกลับกัน การเพิ่มขึ้นของความยาวของทางม้าลายและปริมาณการจราจร ส่งผลในทางลบซึ่งจะช่วยลดความน่าจะเป็นของการฝ่าฝืนไฟสัญญาณสีแดง และนอกจากนี้ยังพบว่าคนเดินเท้าที่มีความพิการทางร่างกายก็มีแนวโน้มที่จะฝ่าฝืนไฟสัญญาณสีแดงน้อยกว่าคนปกติ
Raoniar and Maurya (2022c)	- ศึกษาปัจจัยที่มีอิทธิพลต่อความเร็วในการข้ามถนน	- ใช้วิธี OLS Regression สำหรับการวิเคราะห์ความเร็วและ safety margin - ใช้วิธี Binary Logistic Regression	จากการศึกษาพบว่าปัจจัยส่วนบุคคล (วัย, การใช้มือถือ, กิจกรรมขณะเดิน) มีผลต่อความเร็วและ safety margin การฝ่าฝืนไฟสัญญาณจะเพิ่มขึ้นตามระยะเวลาการรอคอย

ตารางที่ 2-3 สรุปงานวิจัยที่เกี่ยวข้อง (ต่อ)

ผู้แต่ง	วัตถุประสงค์	วิธีการวิเคราะห์	ผลการศึกษา
	และพฤติกรรมของผู้ฝ่าฝืนไฟสัญญาณสีแดงที่ทางข้าม	สำหรับการวิเคราะห์กรณีการฝ่าฝืนไฟสัญญาณ (violation)	ไฟสัญญาณสีแดงที่นานขึ้น กล่าวคือ ยิ่งระยะไฟแดงยาวขึ้น โอกาสที่คนเดินเท้าจะฝ่าฝืนสัญญาณก็เพิ่มขึ้นตาม
Zhu, Sze and Bai (2021)	- การศึกษานี้ตรวจสอบว่าปัจจัยส่วนบุคคล และปัจจัยสิ่งแวดล้อม มีอิทธิพลต่อการตัดสินใจฝ่าฝืนไฟสัญญาณสีแดงของคนเดินเท้าอย่างไร	- วิธี Binary logistic regression	ผลลัพธ์แสดงให้เห็นว่า เพศ อายุ จำนวนช่องทางสำหรับคนเดินข้าม การมีเพื่อนร่วมทาง จำนวนคนเดินข้ามในบริเวณใกล้เคียง การมีผู้ฝ่าฝืนไฟสัญญาณสีแดงคนอื่นในรอบเดียวกัน ระยะเวลาไฟสัญญาณสีเขียว ระยะเวลาไฟสัญญาณสีแดง ปริมาณจราจร และเปอร์เซ็นต์ของยานพาหนะหนัก ล้วนส่งผลกระทบต่อแนวโน้มของการฝ่าฝืนไฟสัญญาณสีแดงของคนเดินเท้า
Rafe et al. (2024)	- ศึกษาพฤติกรรมการข้ามถนนของคนเดินเท้าที่บริเวณทางแยกที่มีไฟสัญญาณจราจร โดยเฉพาะพฤติกรรมการฝ่าฝืนไฟสัญญาณจราจรทั้งในมิติตำแหน่ง (spatial) และมิติเวลา (temporal)	- โมเดลถดถอยโลจิสติกแบบหลายระดับ (multilevel logistic regression)	คนเดินข้ามที่เป็นผู้หญิงและใช้อุปกรณ์ช่วยเดินมักจะปฏิบัติตามกฎจราจรได้ดีกว่า ส่วนปัจจัยแวดล้อม เช่น อุณหภูมิสูง ช่วงเวลากลางคืน หรือช่วงเข้ามิด ส่งผลให้คนเดินข้ามมีโอกาสที่จะฝ่าฝืนไฟสัญญาณเพิ่มสูงขึ้น และในส่วนของปัจจัยทางสังคม เช่น การมีคนอื่นข้ามพร้อมกัน หรือการรอเวลานานขึ้น มีแนวโน้มช่วยลดโอกาสที่คนเดินข้ามจะฝ่าฝืนไฟสัญญาณจราจรลงได้

ตารางที่ 2-3 สรุปงานวิจัยที่เกี่ยวข้อง (ต่อ)

ผู้แต่ง	วัตถุประสงค์	วิธีการวิเคราะห์	ผลการศึกษา
วีรพงศ์ บุญเจริญปัญญา และ จำรัส พัทธ์ศุภการ (2567)	- ศึกษาทัศนคติในการข้ามถนน และปัจจัยที่ส่งผลกระทบทัศนคติในการข้ามถนน	- สถิติเชิงพรรณนาและการวิเคราะห์ถดถอยแบบเส้น (Linear Regression Analysis)	ผลการศึกษาพบว่า จำนวนช่องจราจร ปริมาณจราจร และความเร็วของยานพาหนะส่งผลเสียต่อความรู้สึกปลอดภัยในการข้ามถนนอย่างมีนัยสำคัญ ในด้านของปัจจัยทางสังคมพบว่า เพศ เป็นปัจจัยหลักที่ส่งผลต่อความรู้สึกปลอดภัยอย่างมีนัยสำคัญ ในด้านของตัวแปรด้านทัศนคติที่มีต่อสิ่งอำนวยความสะดวกในการข้าม พบว่าคนเดินเท้ารู้สึกปลอดภัยเมื่อมีการเพิ่มสัญญาณไฟจราจรแบบปุ่มกดมากที่สุด รองลงมาเป็นการมีอยู่ของเกาะกลางถนน และการเพิ่มเนินชะลอความเร็ว (Rubble strip) ตามลำดับ
ชิตินันท์ อินทรวิมลเมธา และ พงษ์ศักดิ์ สุริยวนากุล (2555)	- ศึกษาความสัมพันธ์ระหว่างปัจจัยที่บ่งชี้ค่าความยากในการข้ามถนนกับระดับความยากในการข้ามถนนบริเวณทางแยก	- การสร้างแบบจำลองโพรบิตลำดับ (Ordinal/Ordered Probit Model)	ตัวแปรที่มีค่าสัมประสิทธิ์เป็นลบ ได้แก่ เพศ ความถี่ในการข้ามถนน และความกว้างของเกาะกลางถนน หมายความว่า หากตัวแปรกลุ่มนี้มีค่าเพิ่มมากขึ้นจะส่งผลทำให้ค่าความยากในการข้ามถนนบริเวณทางแยกของคนเดินเท้าต่ำ (ข้ามง่าย) ตัวแปรที่มีค่าสัมประสิทธิ์เป็นบวก ได้แก่ ประเภทของ

ตารางที่ 2-3 สรุปงานวิจัยที่เกี่ยวข้อง (ต่อ)

ผู้แต่ง	วัตถุประสงค์	วิธีการวิเคราะห์	ผลการศึกษา
			เกาะกลาง จำนวนช่องจราจรรวมทั้งสองทิศทาง ปริมาณรถ เลี้ยวขวาของการจราจรด้านใกล้ หมายความว่า หากตัวแปร กลุ่มนี้มีค่าเพิ่มมากขึ้นจะส่งผลทำให้ค่าความยากในการ ข้ามถนนบริเวณทางแยกของคนเดินเท้าสูง (ข้ามยาก)
นพมาศ อัคร จันทโชติ และ ดิเรก พนิตสุภา กมล (2561)	- การศึกษานี้ดำเนินการเปรียบเทียบประสิทธิภาพการแก้ปัญหาข้อมูลไม่สมดุล - ดำเนินการเปรียบเทียบประสิทธิภาพในการจำแนกข้อมูลสำหรับการจำแนกกลุ่มรายได้ผู้ประกอบการรายประเภท ข.ย. 1	- ใช้วิธีการสุ่มตัวอย่างเพิ่มข้อมูลเริ่มต้นอย่างสุ่ม (ROSE) และการสุ่มตัวอย่างเพิ่มกลุ่มส่วนน้อยด้วยการสังเคราะห์ (SMOTE) เพื่อเปรียบเทียบประสิทธิภาพการแก้ปัญหาข้อมูลไม่สมดุล - ใช้วิธีการการถดถอยโลจิสติก และวิธี ต้นไม้ตัดสินใจ เพื่อเปรียบเทียบประสิทธิภาพในการจำแนกข้อมูล	จากผลการเปรียบเทียบวิธีการจำแนกประเภทข้อมูลพบว่า SMOTE ให้ผลลัพธ์ที่ดีที่สุดในตัวชี้วัดเกือบทั้งหมด ยกเว้นความแม่นยำในการพยากรณ์กลุ่มที่มีจำนวนข้อมูลมาก ซึ่งยังด้อยกว่าวิธีอื่นเมื่อเทียบกับข้อมูลเริ่มต้นและวิธี ROSE ผลจากการเปรียบเทียบการจำแนกประเภทข้อมูล - เมื่อพิจารณาค่าความแม่นยำ พบว่า วิธีต้นไม้ตัดสินใจให้ค่าความแม่นยำสูงกว่าวิธีถดถอยโลจิสติก - เมื่อพิจารณาอัตราความถูกต้องในการทำนายกลุ่มส่วนน้อย ซึ่งเป็นกลุ่มเป้าหมายที่สำคัญ พบว่า หลังจากปรับขนาดข้อมูลแล้ว วิธีถดถอยโลจิสติกให้ผลลัพธ์ที่ดีกว่าหรือ

ตารางที่ 2-3 สรุปงานวิจัยที่เกี่ยวข้อง (ต่อ)

ผู้แต่ง	วัตถุประสงค์	วิธีการวิเคราะห์	ผลการศึกษา
			เทียบเท่ากับวิธีต้นไม้ตัดสินใจ
วิษณุวิสิฐ เกษรสิทธิ์, วิจิต หล่อจิระขุนท์กุล และ จิราวัลย์ จิตรถเวช (2561)	- ศึกษาเปรียบเทียบประสิทธิภาพของวิธีการแก้ปัญหาข้อมูลไม่สมดุลของผู้ป่วยโรคเบาหวาน	- เปรียบเทียบวิธีการแก้ปัญหาข้อมูลไม่สมดุลที่ระดับข้อมูล 4 วิธี ได้แก่ การสุ่มเพิ่มข้อมูล, การสุ่มลดข้อมูล, การผสมผสานระหว่างการเพิ่มและลดข้อมูล และการสังเคราะห์ข้อมูลใหม่ด้วยเทคนิค (SMOTE) - ใช้เทคนิคการจำแนกประเภทของข้อมูลด้วยสองวิธีหลัก ได้แก่ การถดถอยโลจิสติกแบบพหุนาม (Multinomial Logistic Regression) และวิธีต้นไม้ตัดสินใจ (Decision Tree)	จากการศึกษาพบว่าวิธีการแก้ปัญหาข้อมูลไม่สมดุล สามารถเพิ่มประสิทธิภาพการจำแนกประเภทข้อมูลส่วนน้อยได้ถูกต้องและแม่นยำมากขึ้น
อัจฉรา แผ้วบาง และสายชล สันสมบูรณ์ทอง (2563)	- ได้ศึกษาเปรียบเทียบวิธีการแก้ปัญหาข้อมูลไม่สมดุล 4 วิธี โดยวิธีการจำแนก 5 วิธี	- ศึกษาเปรียบเทียบวิธีการปรับข้อมูลที่ไม่สมดุล 4 วิธี คือ วิธีการสุ่มเกิน วิธีการสุ่มเกินโดยเทคนิค SMOTE วิธีการสุ่มลด และวิธีการสุ่มผสมผสาน - โดยวิธีการจำแนก 5 วิธี คือ วิธีเพื่อนบ้านใกล้สุด k ตัว วิธีโครงข่ายประสาทเทียม	จากผลการศึกษาของทั้ง 3 ชุดข้อมูลพบว่าวิธีที่มีประสิทธิภาพสูงสุดคือ วิธีฐานกฎโดยการปรับความไม่สมดุลด้วยวิธีการสุ่มเกินเทคนิค SMOTE

ตารางที่ 2-3 สรุปงานวิจัยที่เกี่ยวข้อง (ต่อ)

ผู้แต่ง	วัตถุประสงค์	วิธีการวิเคราะห์	ผลการศึกษา
		วิธีซัพพอร์ตเวกเตอร์แมชชีน วิธีฐานกฎ และวิธีลาดลงสโตแคสติก	

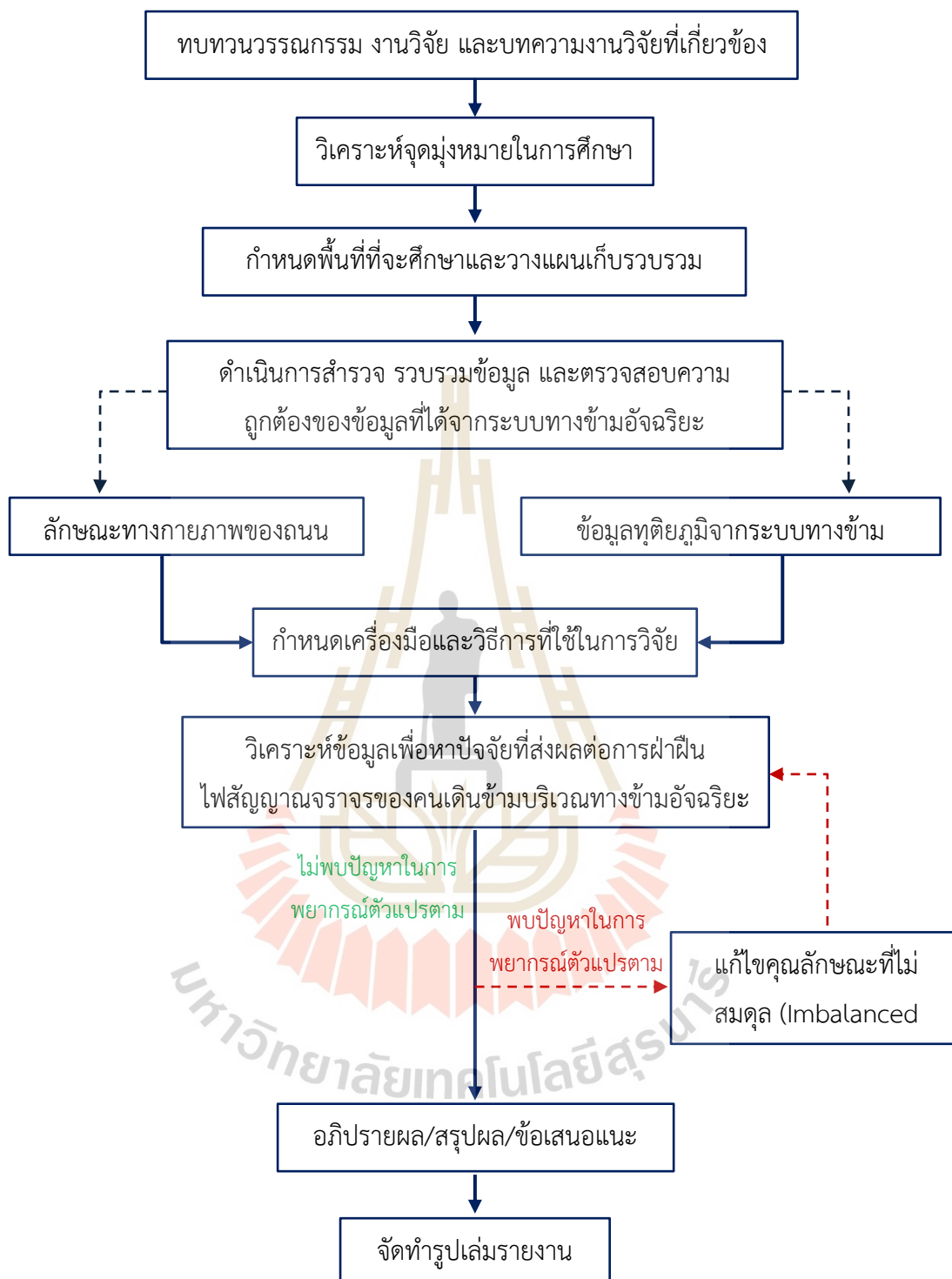


บทที่ 3

วิธีการดำเนินงาน

3.1 ขั้นตอนการดำเนินงานวิจัย

การวิจัยในครั้งนี้ทำการศึกษาปัจจัยที่มีอิทธิพลต่อการฝ่าฝืนไฟสัญญาณของคนเดินข้ามที่ทางข้ามถนนอัจฉริยะ (Smart Crosswalk) ในประเทศไทย เพื่อให้สามารถนำไปสู่จุดมุ่งหมายที่คาดหวังไว้ การศึกษานี้มีขั้นตอนการดำเนินการวิจัยแสดงดังรูปที่ 3-1 โดยในขั้นตอนแรกผู้วิจัยทำการทบทวนเอกสารงานวิจัยที่เกี่ยวข้องกับคนเดินเท้า ทางข้ามถนน ทางข้ามถนนอัจฉริยะ (Smart Crosswalk) และทฤษฎีที่เกี่ยวข้องกับการวิเคราะห์สร้างแบบจำลอง รวมถึงการทบทวนงานวิจัยที่ผ่านมาที่เกี่ยวข้องกับการศึกษาปัจจัยที่มีอิทธิพลต่อการฝ่าฝืนไฟสัญญาณของคนเดินข้าม ขั้นตอนต่อมาทำการวิเคราะห์จุดมุ่งหมายในการศึกษา และหลังจากนั้นกำหนดพื้นที่ที่จะศึกษาและวางแผนเก็บรวบรวมข้อมูล ขั้นตอนต่อมาคือดำเนินการสำรวจ รวบรวมข้อมูล และตรวจสอบความถูกต้องของข้อมูลที่ได้จากระบบทางข้ามถนนอัจฉริยะ เพื่อคัดเลือกปัจจัยที่สามารถนำไปดำเนินการวิเคราะห์ได้ หลังจากได้ทราบถึงลักษณะของข้อมูลที่จะนำมาวิเคราะห์แล้ว ขั้นตอนต่อมาคือกำหนดเครื่องมือที่ใช้ในการวิจัยเพื่อหาปัจจัยที่ส่งผลต่อการฝ่าฝืนไฟสัญญาณของคนเดินข้าม บริเวณทางข้ามถนนอัจฉริยะ แล้วจึงดำเนินการวิเคราะห์ข้อมูลที่ถูกต้องสมบูรณ์ที่ได้จากการเก็บรวบรวมข้อมูล หลังจากได้ผลจากการวิเคราะห์ข้อมูลแล้ว ทำการอภิปรายผล/สรุปผล/ข้อเสนอแนะ และจัดทำรูปแบบรายงานให้สมบูรณ์



รูปที่ 3-1 ขั้นตอนการดำเนินการวิจัย

3.2 การกำหนดขอบเขตและพื้นที่ที่ทำการศึกษา

ทางข้ามถนนอัจฉริยะ (Smart Crosswalk) ของกรมทางหลวง ที่จะดำเนินการสำรวจและเก็บข้อมูลในการศึกษาครั้งนี้ คือทางข้ามถนนอัจฉริยะทั้งหมด 6 แห่ง จากทั้งสิ้น 11 แห่งในปัจจุบัน ทั้งนี้สาเหตุที่ผู้วิจัยไม่นำทางข้ามอื่นๆอีก 5 แห่ง มาดำเนินการวิเคราะห์ร่วมด้วย เนื่องจากทางข้าม 2 แห่ง พบว่าระบบทางข้ามถนนอัจฉริยะเกิดการชำรุดเป็นระยะเวลานาน และทางข้ามอีก 3 แห่ง เป็นทางข้ามที่ได้ทำการก่อสร้างหรือปรับปรุงทางข้ามถนนให้เป็นทางข้ามถนนอัจฉริยะในช่วงปลายปี พ.ศ. 2566 ซึ่งในขณะที่ทำงานวิจัยฉบับนี้ข้อมูลที่ได้รับมานั้นไม่เพียงพอสำหรับการนำมาวิเคราะห์สำหรับทางข้ามถนนอัจฉริยะจำนวน 6 แห่งที่จะดำเนินการสำรวจและเก็บข้อมูลในการศึกษาครั้งนี้ได้แก่

1) ทางข้ามถนนเอกชัย 102/3 ทางหลวงหมายเลข 3242 กม. 18+110 จังหวัดกรุงเทพมหานคร เป็นทางข้ามบนถนน 6 ช่องจราจร แบบมีเกาะกลาง แสดงได้ดังรูปที่ 3-2



รูปที่ 3-2 ตำแหน่งทางข้ามถนนอัจฉริยะบนทางหลวงหมายเลข 3242 กม. 18+110
(ที่มา : โครงการศึกษาจัดทำแผนพัฒนาศูนย์บริหารจราจรกลาง (HTOC) กรมทางหลวง ปี พ.ศ. 2567- พ.ศ. 2570 1 โครงการ, 2566)

2) ทางข้ามถนนพระราชานุรักษ์ ซอย 10 ทางหลวงหมายเลข 306 กม. 4+970 จังหวัดนนทบุรี เป็นทางข้ามบนถนน 4 ช่องจราจร แบบไม่มีเกาะกลาง แสดงได้ดังรูปที่ 3-3



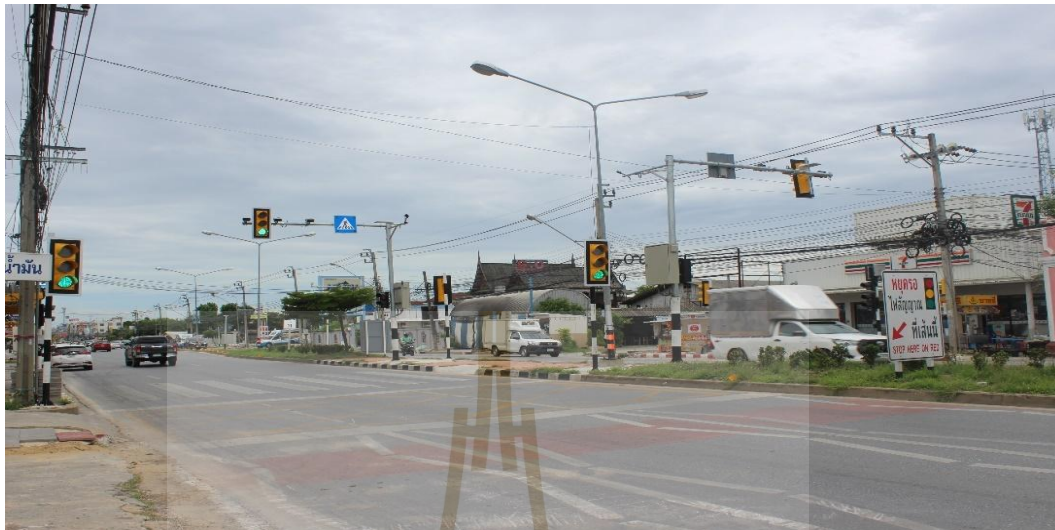
รูปที่ 3-3 ตำแหน่งทางข้ามถนนอรรถจริยะบนทางหลวงหมายเลข 306 กม. 4+970
(ที่มา : โครงการศึกษาจัดทำแผนพัฒนาศูนย์บริหารจราจรกลาง (HTOC)
กรมทางหลวง ปี พ.ศ. 2567- พ.ศ. 2570 1 โครงการ, 2566)

3) ทางข้ามวัดไร่ขิง ทางหลวงหมายเลข 3316 กม. 2+225 จังหวัดนครปฐม เป็นทางข้ามบนถนน 4 ช่องจราจร แบบไม่มีเกาะกลาง แสดงได้ดังรูปที่ 3-4



รูปที่ 3-4 ตำแหน่งทางข้ามถนนอรรถจริยะบนทางหลวงหมายเลข 3316 กม. 2+225
(ที่มา : โครงการศึกษาจัดทำแผนพัฒนาศูนย์บริหารจราจรกลาง (HTOC)
กรมทางหลวง ปี พ.ศ. 2567- พ.ศ. 2570 1 โครงการ, 2566)

4) ทางข้ามหมู่บ้าน ช.รุ่งเรือง 6 ทางหลวงหมายเลข 3215 กม. 11+020 จังหวัดนนทบุรี
เป็นถนนขนาด 6 ช่องจราจร แบบมีเกาะกลาง แสดงได้ดังรูปที่ 3-5



รูปที่ 3-5 ตำแหน่งทางข้ามถนนอัจฉริยะบนทางหลวงหมายเลข 3215 กม. 11+020
(ที่มา : โครงการศึกษาจัดทำแผนพัฒนาศูนย์บริหารจราจรกลาง (HTOC)
กรมทางหลวง ปี พ.ศ. 2567- พ.ศ. 2570 1 โครงการ, 2566)

5) ทางข้ามหน้า รพ.เอกชัย ทางหลวงหมายเลข 3242 กม. 2+220 จังหวัดสมุทรสาคร
เป็นถนนขนาด 6 ช่องจราจร แบบมีเกาะกลาง แสดงได้ดังรูปที่ 3-6



รูปที่ 3-6 ตำแหน่งทางข้ามถนนอัจฉริยะบนทางหลวงหมายเลข 3242 กม. 2+220
(ที่มา : โครงการศึกษาจัดทำแผนพัฒนาศูนย์บริหารจราจรกลาง (HTOC)
กรมทางหลวง ปี พ.ศ. 2567- พ.ศ. 2570 1 โครงการ, 2566)

6) ทางข้ามเอื้ออาทรแสนภูดาษ ทางหลวงหมายเลข 314 กม. 8+410 จังหวัดฉะเชิงเทรา เป็นถนนขนาด 6 ช่องจราจร แบบมีเกาะกลาง แสดงได้ดังรูปที่ 3-7



รูปที่ 3-7 ตำแหน่งทางข้ามถนนอัจฉริยะบนทางหลวงหมายเลข 314 กม. 8+410
(ที่มา : โครงการศึกษาจัดทำแผนพัฒนาศูนย์บริหารจราจรกลาง (HTOC)
กรมทางหลวง ปี พ.ศ. 2567- พ.ศ. 2570 1 โครงการ, 2566)

3.3 การเก็บรวบรวมข้อมูล

การศึกษานี้ใช้ข้อมูลเกี่ยวกับลักษณะทางกายภาพของถนน และข้อมูลพฤติกรรมจากระบบทางข้ามถนนอัจฉริยะ ซึ่งประกอบด้วย ข้อมูลเกี่ยวกับจราจรบริเวณทางข้าม ข้อมูลเกี่ยวกับคนเดินข้าม รวมไปถึงข้อมูลจำเพาะเกี่ยวกับระบบไฟสัญญาณบริเวณทางข้าม ซึ่งได้รวบรวมข้อมูลจากระบบทางข้ามถนนอัจฉริยะตั้งแต่เดือนมกราคม พ.ศ. 2565 ถึงเดือนสิงหาคม พ.ศ. 2565

3.3.1 ข้อมูลลักษณะทางกายภาพของทางข้ามถนนอัจฉริยะ

การสำรวจลักษณะทางกายภาพของทางข้ามถนนอัจฉริยะ ผู้วิจัยทำการรวบรวมข้อมูลจากการลงพื้นที่สำรวจที่ทางข้ามถนน และจากแบบแปลนก่อสร้างเพื่อปรับปรุงทางข้ามให้เป็นทางข้ามอัจฉริยะ โดยข้อมูลกายภาพของทางข้าม แสดงดังตารางที่ 3-1 ครอบคลุม 3 ประเด็นหลักดังนี้

- 1) ลักษณะทางกายภาพของถนน ได้แก่ การมีหรือไม่มีเกาะกลางถนน ความกว้างของเกาะกลาง และจำนวนช่องจราจร/ทิศทาง
- 2) ลักษณะทางกายภาพของทางข้าม ได้แก่ ความยาวและความกว้างของทางข้าม
- 3) อุปกรณ์ความปลอดภัยที่ทางข้าม ได้แก่ การมีหรือไม่มีและความกว้างของเส้นทแยงห้ามหยุดรถก่อนถึงทางข้าม

ตารางที่ 3-1 ข้อมูลลักษณะทางกายภาพของทางม้าลายอัจฉริยะ

ตำแหน่งทางข้าม	เกาะกลางถนน	ความกว้างเกาะกลางถนน (เมตร)	จำนวนช่องจราจร/ทิศทาง	ความยาวทางข้าม (เมตร)	ความกว้างทางข้าม (เมตร)	เส้นทแยงห้ามหยุด	ความกว้างเส้นทแยงห้ามหยุด (เมตร)
1) ทางหลวงหมายเลข 3242 กม. 18+110	มี	1.3	3	23.2	8	ไม่มี	0
2) ทางหลวงหมายเลข 306 กม. 4+970	มี	0.8	2	14.8	8	ไม่มี	0
3) ทางหลวงหมายเลข 3316 กม. 2+225	ไม่มี	0	2	13.8	8	มี	7
4) ทางหลวงหมายเลข 3215 กม. 11+020	มี	4.3	3	23	8	มี	7
5) ทางหลวงหมายเลข 3242 กม. 2+220	มี	10	3	30	8	มี	12
6) ทางหลวงหมายเลข 314 กม. 8+410	มี	5	3	35	16	มี	17.7

3.3.2 ข้อมูลทุติยภูมิจากระบบทางข้ามถนนอัจฉริยะ

ผู้วิจัยได้รับข้อมูลจากฐานข้อมูลระบบทางข้ามถนนอัจฉริยะ (Smart Crosswalk) ซึ่งเป็นระบบฐานข้อมูลส่วนกลางในการแสดงข้อมูลการทำงานของระบบทางข้ามถนนอัจฉริยะ ได้แก่ Traffic data, Speed data, Pedestrian data, and Transaction data จากข้อจำกัดบางประการของฐานข้อมูลที่ได้รับจากระบบทางข้ามถนนอัจฉริยะ สำหรับการศึกษาค้นคว้าครั้งนี้ผู้วิจัยเลือกใช้ข้อมูลจากฐานข้อมูล Traffic data และ Transaction data เป็นหลัก

1) Traffic data เป็นฐานข้อมูลที่ให้ข้อมูลเกี่ยวกับปริมาณจราจรที่สัญจรผ่านทางข้าม และความเร็วของยานพาหนะบริเวณทางข้าม รายละเอียดข้อมูลแสดงดังตารางที่ 3-2

ตารางที่ 3-2 รายละเอียดข้อมูลที่ได้จากฐานข้อมูล Traffic data

Traffic data	คำอธิบาย
ความเร็วของยานพาหนะ	คำนวณความเร็วเฉลี่ยของยานพาหนะทุก ๆ 5 นาที ที่ตรงกับรอบไฟสัญญาณ (หน่วย: กิโลเมตร/ชั่วโมง)
ปริมาณจราจรที่สัญจรผ่านทางข้าม	คำนวณปริมาณจราจรที่ขับขึ้นบริเวณทางข้ามโดยไม่แยกประเภทของยานพาหนะทุก ๆ 5 นาที ที่ตรงกับรอบไฟสัญญาณ (หน่วย: คัน/ชั่วโมง/ช่องจราจร)

2) Transaction data ให้ข้อมูลโดยละเอียดเกี่ยวกับไฟสัญญาณคนเดินข้าม ซึ่งได้แก่ ระยะเวลาที่คนเดินข้ามรอไฟสัญญาณข้ามถนน ระยะเวลาไฟสัญญาณสีเขียวรูปคนเดินข้าม (Walk Interval) ระยะเวลาการแสดงผลสัญญาณไฟรูปมือแบบกระพริบ (Flashing Don't Walk: FDW) ระยะเวลาการขยายไฟสัญญาณสีเขียวรูปคนเดินข้ามแบบกระพริบ จำนวนคนเดินข้ามทั้งหมด และจำนวนคนเดินข้ามที่ฝ่าฝืนไฟสัญญาณในแต่ละรอบไฟสัญญาณ รายละเอียดข้อมูลแสดงดังตารางที่ 3-3 อย่างไรก็ตามข้อมูลที่ได้รับการจากระบบทางข้ามถนนอัจฉริยะยังขาดข้อมูลเกี่ยวกับลักษณะเฉพาะส่วนบุคคลของคนเดินเท้า เช่น เพศ อายุ และอาชีพ ดังนั้นการศึกษานี้จึงไม่สามารถพิจารณาปัจจัยที่เกี่ยวข้องกับลักษณะเฉพาะส่วนบุคคลของคนเดินเท้าได้

และเนื่องจากลักษณะของข้อมูลการฝ่าฝืนไฟสัญญาณของคนเดินข้ามที่ถูกบันทึกในระบบ คือ จำนวนคนเดินข้ามทั้งหมดที่ฝ่าฝืนไฟสัญญาณในแต่ละรอบไฟสัญญาณ ดังนั้นตัวแปรตามที่ใช้ในการศึกษาครั้งนี้จึงเป็นแบบ Dichotomous คือ 0 และ 1 ซึ่งเป็นการระบุเหตุการณ์การฝ่าฝืนไฟสัญญาณของคนเดินข้ามที่เกิดขึ้นในแต่ละรอบไฟสัญญาณ โดยที่ 0 คือ ไม่พบคนเดินข้ามฝ่าฝืนไฟสัญญาณ (Not Violated) และ 1 คือ พบคนเดินข้ามอย่างน้อยหนึ่งคนฝ่าฝืนไฟสัญญาณ (Violated)

ตารางที่ 3-3 รายละเอียดข้อมูลที่ได้จากฐานข้อมูล Transaction data

Transaction data	คำอธิบาย
ระยะเวลารอคอย	คำนวณจากช่วงที่คนเดินข้ามกดปุ่มขอไฟสัญญาณคนเดินข้าม จนกระทั่งได้รับสัญญาณให้ข้าม (หน่วย: วินาที)
ระยะเวลาไฟสัญญาณสีเขียวรูปคนเดินข้าม (Walk Interval)	ระยะเวลาที่คำนวณจาก เวลาที่คนได้ไฟสัญญาณสีเขียวรูปคนเดินข้าม จนถึง ไฟสัญญาณรูปมือแบบกระพริบ (หน่วย: วินาที)
ระยะเวลาการแสดงผลไฟสัญญาณรูปมือแบบกระพริบ (Flashing Don't Walk: FDW)	เวลาที่คนเริ่มได้ไฟสัญญาณรูปมือแบบกระพริบ จนถึงระยะเวลาการขยายไฟสัญญาณรูปมือแบบกระพริบ (หน่วย: วินาที)
ระยะเวลาการขยายไฟสัญญาณรูปมือแบบกระพริบ	เวลาที่กล้องตรวจจับพบว่ามีคนข้ามอยู่ในพื้นที่ทางข้าม และระบบจะเพิ่มระยะเวลาไฟสัญญาณรูปมือแบบกระพริบทุก ๆ 1 วินาที จนกระทั่งไม่มีคนเดินข้ามอยู่ในพื้นที่ทางข้าม หรือระยะเวลาสัญญาณไฟคนข้ามเกินระยะเวลาที่ใช้ในการข้ามถนนสูงสุด (หน่วย: วินาที)
จำนวนคนเดินข้ามทั้งหมด	จำนวนของคนเดินข้ามที่ข้ามถนนตั้งแต่กดปุ่ม จนถึงจบรอบไฟสัญญาณ (หน่วย: คน/รอบสัญญาณไฟ)

ตารางที่ 3-3 รายละเอียดข้อมูลที่ได้จากฐานข้อมูล Transaction data (ต่อ)

Transaction data	คำอธิบาย
จำนวนคนเดินข้ามที่ฝ่าฝืนไฟสัญญาณ	จำนวนของคนเดินข้ามที่ข้ามถนนตั้งแต่กดปุ่ม จนถึงไฟสัญญาณสีแดงครั้งแรกของยานพาหนะ (หน่วย: คน/รอบสัญญาณไฟ)

3.4 การวิเคราะห์ข้อมูล

3.4.1 สถิติเชิงพรรณนา

การศึกษาครั้งนี้เริ่มต้นการวิเคราะห์โดยใช้สถิติเชิงพรรณนา เพื่อดูพฤติกรรมเบื้องต้นของผู้ขับขี่ยานพาหนะและคนเดินข้ามที่บริเวณทางข้ามถนนอัจฉริยะ รวมถึงการทำงานของระบบไฟสัญญาณทางข้ามอัจฉริยะ ซึ่งประกอบไปด้วย การหาค่าความถี่ (Frequency), ค่าร้อยละ (Percentage), ค่าเฉลี่ย (Mean), ค่าเบี่ยงเบนมาตรฐาน (S.D.), ค่าต่ำสุด (Min), ค่าสูงสุด (Max), ค่ามัธยฐาน (Median), และตำแหน่งเปอร์เซ็นต์ไทล์ (Percentile rank) เป็นต้น

3.4.2 วิเคราะห์หาความสัมพันธ์ระหว่างตัวแปรอิสระ

โดยจะทำการวิเคราะห์หาค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation) เพื่อหาความสัมพันธ์ระหว่างตัวแปรอิสระ เพื่อป้องกันการมีความสัมพันธ์กันเองสูงระหว่างตัวแปรอิสระ (Multicollinearity) ซึ่งอาจส่งผลทำให้การพยากรณ์คลาดเคลื่อน ดังนั้นจึงต้องทำการตรวจสอบเพื่อให้แน่ใจว่าตัวแปรอิสระนั้น ๆ สามารถนำมาใช้ในการวิเคราะห์สร้างแบบจำลองได้

3.4.3 วิเคราะห์สร้างแบบจำลองจากชุดข้อมูลดั้งเดิมโดยไม่แก้ไขคุณลักษณะที่ไม่สมดุล

ทำการวิเคราะห์สร้างแบบจำลองจากชุดข้อมูลดั้งเดิมโดยไม่แก้ไขคุณลักษณะที่ไม่สมดุล โดยใช้วิธีการถดถอยโลจิสติกแบบทวิ (Binary logistic regression) และวิธีต้นไม้ตัดสินใจ (Decision tree) เพื่อตรวจสอบว่าปัจจัยลักษณะทางกายภาพของถนน และปัจจัยข้อมูลทุติยภูมิจากระบบทางข้ามอัจฉริยะ ไต่บ้างที่ส่งผลต่อการฝ่าฝืนไฟสัญญาณของคนเดินข้ามและส่งผลอย่างไร รวมถึงดำเนินการประเมินประสิทธิภาพแบบจำลองในการพยากรณ์การฝ่าฝืนไฟสัญญาณของคนเดินข้าม

3.4.4 ดำเนินการแก้ไขคุณลักษณะที่ไม่สมดุล (Imbalanced Data)

หลังจากวิเคราะห์สร้างแบบจำลองจากชุดข้อมูลดั้งเดิมโดยไม่แก้ไขคุณลักษณะที่ไม่สมดุลแล้ว หากพบว่าแบบจำลองที่ได้มีประสิทธิภาพไม่เพียงพอในการพยากรณ์การฝ่าฝืนไฟสัญญาณของคนเดินข้าม ผู้วิจัยจะทำการดำเนินการแก้ไขคุณลักษณะที่ไม่สมดุล (Imbalanced Data) โดยเทคนิคการสุ่มเลือกข้อมูล (Data Sampling Technique) 4 วิธี ได้แก่ Upsampling, Downsampling, Random OverSampling Examples (ROSE), and Synthetic Minority Oversampling Technique (SMOTE) เพื่อปรับจำนวนรอบไฟสัญญาณที่มีการฝ่าฝืนไฟสัญญาณ

ของคนเดินข้ามให้ใกล้เคียงกับจำนวนรอบไฟสัญญาณที่ไม่มีการฝ่าฝืนไฟสัญญาณของคนเดินข้าม เพื่อแก้ไขและปรับปรุงความแม่นยำในการพยากรณ์การฝ่าฝืนไฟสัญญาณจราจรของคนเดินข้าม

3.4.5 วิเคราะห์สร้างแบบจำลองจากชุดข้อมูลที่มีการแก้ไขคุณลักษณะที่ไม่สมดุล

หลังจากใช้เทคนิคการแก้ไขคุณลักษณะที่ไม่สมดุลจากทั้ง 4 วิธีในข้างต้นแล้ว ดำเนินการวิเคราะห์สร้างแบบจำลองจากชุดข้อมูลที่มีการแก้ไขคุณลักษณะที่ไม่สมดุล โดยใช้วิธีการถดถอยโลจิสติกแบบทวิ (Binary logistic regression) และวิธีต้นไม้ตัดสินใจ (Decision tree) เพื่อตรวจสอบว่าปัจจัยลักษณะทางกายภาพของถนน และปัจจัยข้อมูลพฤติกรรมจากระบบทางข้ามอัจฉริยะใดบ้าง ที่ส่งผลต่อการฝ่าฝืนไฟสัญญาณของคนเดินข้ามและส่งผลอย่างไร จากนั้นคัดเลือกแบบจำลองจากชุดข้อมูลที่มีการปรับแก้ไขคุณลักษณะที่ไม่สมดุลจากทั้ง 4 วิธี ที่มีประสิทธิภาพสูงที่สุดในการพยากรณ์การฝ่าฝืนไฟสัญญาณของคนเดินข้าม เพื่อเป็นการตอบวัตถุประสงค์งานวิจัยและคำถามงานวิจัยในการศึกษานี้



บทที่ 4

ผลการวิเคราะห์ข้อมูล

4.1 ผลการวิเคราะห์ข้อมูลสถิติเชิงพรรณนา

จากการสำรวจและเก็บรวบรวมข้อมูลลักษณะทางกายภาพของถนน และข้อมูลพฤติกรรมจากระบบทางข้ามอัจฉริยะของทางข้ามอัจฉริยะทั้ง 6 แห่ง เพื่อแสดงถึงพฤติกรรมเบื้องต้นของผู้ขับขี่ยานพาหนะและคนเดินข้ามที่บริเวณทางข้ามถนนอัจฉริยะ รวมถึงการทำงานของระบบไฟสัญญาณทางข้ามถนนอัจฉริยะได้ชัดเจนมากยิ่งขึ้น ผู้วิจัยได้ทำการวิเคราะห์ข้อมูลสถิติเชิงพรรณนาของข้อมูลที่ได้รับมา ซึ่งประกอบไปด้วย การหาค่าความถี่ (Frequency), ค่าร้อยละ (Percentage), ค่าเฉลี่ย (Mean), ค่าเบี่ยงเบนมาตรฐาน (S.D.), ค่าต่ำสุด (Min), ค่าสูงสุด (Max), ค่ามัธยฐาน (Median), และตำแหน่งเปอร์เซ็นต์ไทล์ (Percentile rank) ในแต่ละปัจจัย จากข้อมูลรอบไฟสัญญาณที่นำมาวิเคราะห์ทั้งสิ้น 116,485 รอบไฟสัญญาณ ผลการวิเคราะห์ดังแสดงในตารางที่ 4-1 และสามารถสรุปผลการวิเคราะห์ความถี่สะสมของแต่ละข้อมูล ดังแสดงในรูปที่ 4-1



ตารางที่ 4-1 ผลการวิเคราะห์ค่าสถิติพื้นฐานของข้อมูลทางข้ามอัจฉริยะ

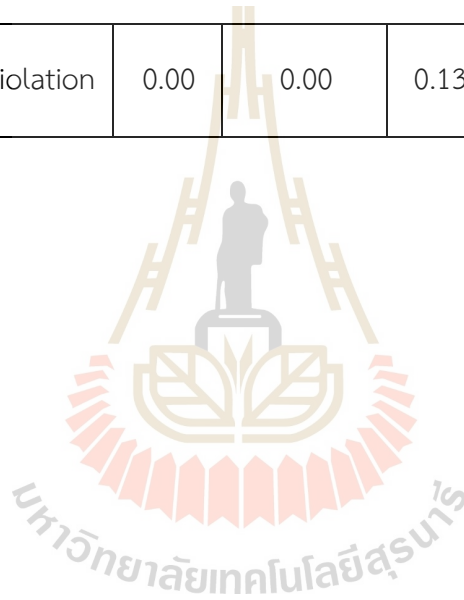
ข้อมูลลักษณะทางกายภาพของทางข้ามอัจฉริยะ								
ข้อมูลลักษณะทางกายภาพ	ชื่อตัวแปรที่ใช้ในการวิเคราะห์	MIN	15 th Percentile	MEAN	MEDIAN	85 th Percentile	MAX	S.D.
ความกว้างของเกาะกลางถนน (เมตร)	MediansWidth	0.00	0.00	2.13	1.30	5.00	10.00	2.45
ความยาวของทางข้าม (เมตร)	CrosswalksLength	13.80	13.80	22.94	23.20	35.00	35.00	7.74
ความกว้างของทางข้าม (เมตร)	CrosswalksWidth	8.00	8.00	9.72	8.00	16.00	16.00	3.28
ความกว้างเส้นทแยงห้ามหยุดรถ (เมตร)	yellowframeWidth	0.00	0.00	6.38	7.00	17.70	17.70	6.86
ข้อมูลพฤติกรรมจากระบบทางข้ามอัจฉริยะ								
ข้อมูลเกี่ยวกับคนเดินข้าม		ชื่อตัวแปรที่ใช้ในการวิเคราะห์		ค่าความถี่		ค่าร้อยละ		
เหตุการณ์การฝ่าฝืนไฟสัญญาณของคนเดินข้ามที่เกิดขึ้นระหว่างรอบไฟสัญญาณ (n = 116,485 รอบไฟสัญญาณ)		VioPedestrian						
ไม่พบคนเดินข้ามฝ่าฝืนไฟสัญญาณ (0 = Not Violated)				108,490		93.14		
พบคนเดินข้ามอย่างน้อยหนึ่งคนฝ่าฝืนไฟสัญญาณ (1 = Violated)				7,995		6.86		
ช่วงเวลาของวันที่พบเหตุการณ์การฝ่าฝืนไฟสัญญาณของคนเดินข้ามหลังจากกดปุ่มขอไฟสัญญาณเพื่อข้ามถนน		TimeofDay						
ช่วงเวลากลางคืน (1 = กลางคืน)				1,555		19.45%		
ช่วงเวลากลางวัน (0 = กลางวัน)				6,440		80.55%		

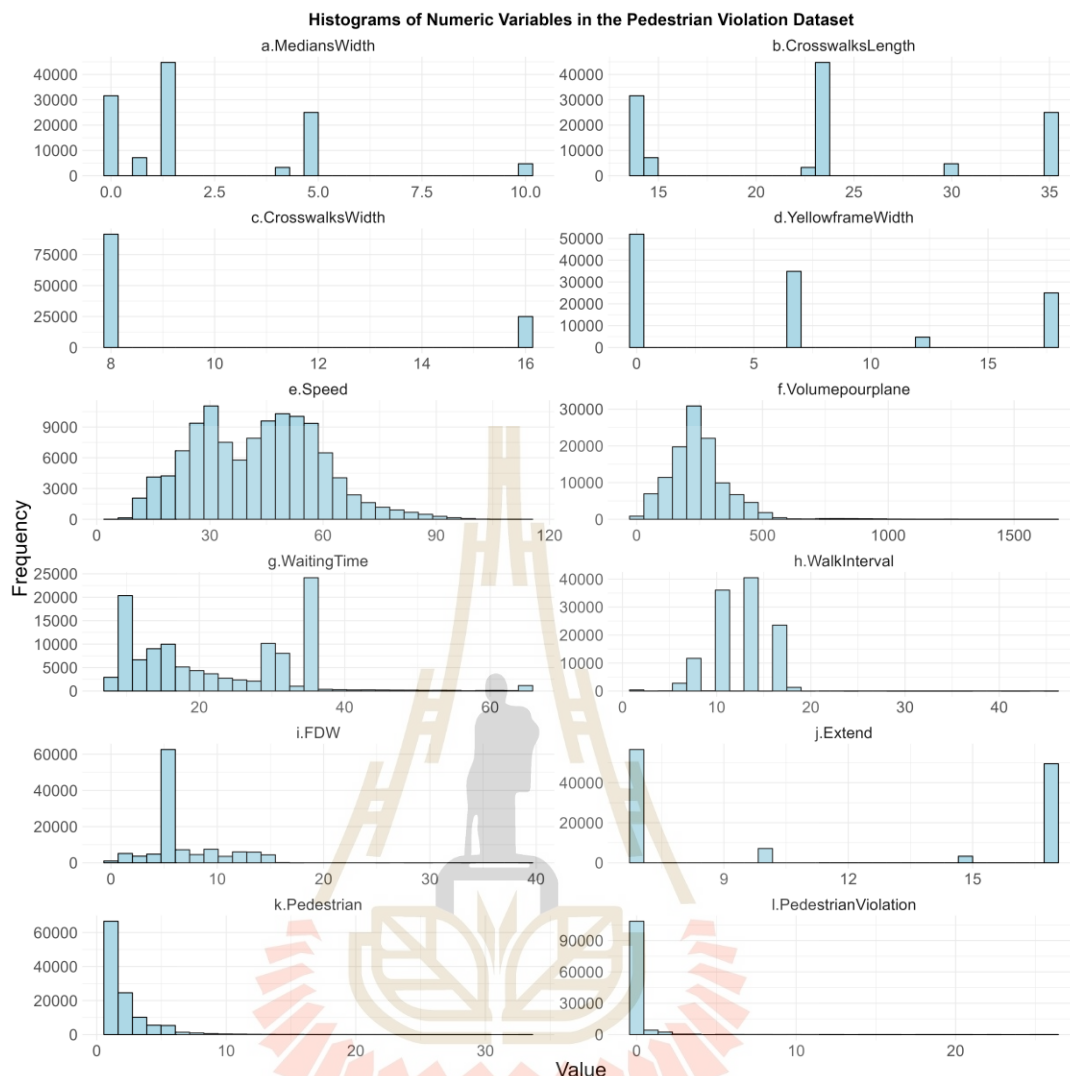
ตารางที่ 4-1 ผลการวิเคราะห์ค่าสถิติพื้นฐานของข้อมูลทางข้ามอัจฉริยะ (ต่อ)

ข้อมูลทางข้ามอัจฉริยะ	ชื่อตัวแปรที่ใช้ ในการวิเคราะห์	MIN	15 th Percentile	MEAN	MEDIAN	85 th Percentile	MAX	S.D.
ข้อมูลเกี่ยวกับจราจรบริเวณทางข้าม								
ความเร็วของยานพาหนะ (กิโลเมตร/ชั่วโมง)	Speed	5.25	24.74	42.56	43.29	59.14	115.13	16.45
ปริมาณจราจรที่สัญจรผ่านทางข้าม (คัน/ชั่วโมง/ช่องจราจร)	VolumeHourplane	2.00	136.00	246.38	234.00	350.00	1,650.00	121.67
ระยะเวลารอคอย (วินาที/รอบสัญญาณไฟ)	WaitingTime	8.00	10.00	22.75	20.00	35.00	65.00	11.54
ข้อมูลจำเพาะเกี่ยวกับระบบไฟสัญญาณบริเวณทางข้าม								
ระยะเวลาไฟสัญญาณสีเขียวรูปคนเดินข้าม (Walk Interval) (วินาที/รอบสัญญาณไฟ)	WalkInterval	1.00	10.00	12.34	13.00	17.00	45.00	3.37
ระยะเวลาการแสดงสัญญาณไฟรูปมือแบบ กระพริบ (Flashing Don't Walk: FDW) (วินาที/รอบสัญญาณไฟ)	FDW	0.00	5.00	7.07	6.00	11.00	39.00	3.21
ระยะเวลาการขยายไฟสัญญาณสีเขียวรูปมือ แบบกระพริบ (วินาที/รอบสัญญาณไฟ)	Extend	7.00	7.00	11.66	10.00	17.00	17.00	4.81
ข้อมูลเกี่ยวกับคนเดินข้าม								
จำนวนคนเดินข้ามทั้งหมด (คน/รอบสัญญาณไฟ)	Pedestrain	1.00	1.00	2.03	1.00	3.00	33.00	1.89

ตารางที่ 4-1 ผลการวิเคราะห์ค่าสถิติพื้นฐานของข้อมูลทางข้ามอัจฉริยะ (ต่อ)

ข้อมูลทางข้ามอัจฉริยะ	ชื่อตัวแปรที่ใช้ ในการวิเคราะห์	MIN	15 th Percentile	MEAN	MEDIAN	85 th Percentile	MAX	S.D.
ข้อมูลเกี่ยวกับคนเดินข้าม								
จำนวนคนเดินข้ามที่ฝ่าฝืนไฟสัญญาณ (คน/รอบสัญญาณไฟ)	PedestrianViolation	0.00	0.00	0.137	0.00	0.00	26.00	0.70





รูปที่ 4-1 กราฟแสดงความถี่สะสมข้อมูลทางข้ามอัจฉริยะ

จากตารางที่ 1 และรูปที่ 4-1 ข้อมูลเหตุการณ์การฝ่าฝืนไฟสัญญาณของคนเดินข้ามที่เกิดขึ้นระหว่างรอบไฟสัญญาณ จะเห็นได้ว่ารอบไฟสัญญาณที่พบคนเดินข้ามอย่างน้อยหนึ่งคนฝ่าฝืนไฟสัญญาณ (1 = Violated) มีเพียงร้อยละ 6.86 ซึ่งถือได้ว่ามีสัดส่วนที่ค่อนข้างน้อย เมื่อเทียบกับชุดข้อมูลทั้งหมดถือว่าเป็นเหตุการณ์ที่เกิดขึ้นไม่บ่อย (rare event) ซึ่งอาจกล่าวได้ว่าชุดข้อมูลนี้เป็นลักษณะ Imbalanced data ถึงแม้จะเป็นเหตุการณ์ที่เกิดขึ้นไม่บ่อยมากนัก แต่การฝ่าฝืนดังกล่าวมีความเสี่ยงที่อาจจะก่อให้เกิดอุบัติเหตุที่ร้ายแรงขึ้นกับผู้ขับขี่รถยนต์ ในด้านของช่วงเวลาของวันที่พบเหตุการณ์การฝ่าฝืนไฟสัญญาณของคนเดินข้ามหลังจากกดปุ่มขอไฟสัญญาณเพื่อข้ามถนน จะเห็นได้ว่าโดยส่วนใหญ่เหตุการณ์การฝ่าฝืนไฟสัญญาณของคนเดินข้ามจะเกิดขึ้นในช่วงเวลากลางวัน ร้อยละ 80.55 ด้านของความเร็วเฉลี่ยในการขับขี่ของยานพาหนะ พบว่าอยู่ที่ 42.57 กม./ชม. และค่าความเร็วที่ 85 เปอร์เซ็นไทล์อยู่ที่ 59.14 กม./ชม. ซึ่งความเร็วดังกล่าวยังมีค่าความเร็วที่สูงกว่าความเร็วที่ปลอดภัยที่แนะนำสำหรับคนเดินถนน (30– 40 กม./ชม.)

(World Health Organization, 2023) ด้านปริมาณจราจรเฉลี่ยที่สัญจรผ่านทางข้าม พบว่าอยู่ที่ 247 คัน/ชั่วโมง/ช่องจราจร ด้านระยะเวลารอคอยสัญญาณไฟเดินข้ามเฉลี่ย พบว่าอยู่ที่ 22.75 วินาที/รอบสัญญาณไฟ ด้านระยะเวลาไฟสัญญาณสีเขียวรูปคนเดินข้ามเฉลี่ย (Walk Interval) พบว่าอยู่ที่ 12.34 วินาที/รอบสัญญาณไฟ ด้านระยะเวลาการแสดงผลสัญญาณไฟรูปมือแบบกระพริบเฉลี่ย (Flashing Don't Walk: FDW) พบว่าอยู่ที่ 7.07 วินาที/รอบสัญญาณไฟ ด้านระยะเวลาการขยายไฟสัญญาณสีเขียวรูปคนเดินข้ามแบบกระพริบเฉลี่ย พบว่าอยู่ที่ 11.66 วินาที/รอบไฟสัญญาณ ด้านจำนวนคนเดินข้ามทั้งหมดในแต่ละรอบไฟสัญญาณพบคนเดินข้ามเฉลี่ย 2 คน/รอบไฟสัญญาณ สูงที่สุดอยู่ที่ 33 คน/รอบไฟสัญญาณ และในด้านของจำนวนคนเดินข้ามที่ฝ่าฝืนไฟสัญญาณพบคนเดินข้ามฝ่าฝืนไฟสัญญาณเฉลี่ย 0.137 คน/รอบไฟสัญญาณ คนเดินข้ามโดยส่วนใหญ่จะไม่ฝ่าฝืนไฟสัญญาณ ในรอบไฟสัญญาณที่พบการฝ่าฝืนของคนเดินข้าม พบคนเดินข้ามที่ฝ่าฝืนสูงสุดถึง 26 คน/รอบไฟสัญญาณ

4.2 วิเคราะห์หาความสัมพันธ์ระหว่างตัวแปรอิสระ

จากข้อมูลที่ได้รับจากระบบทางข้ามอัจฉริยะ ผู้วิจัยทำการวิเคราะห์หาค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation) ระหว่างตัวแปรที่ใช้ในการศึกษาที่ส่งผลต่อการฝ่าฝืนไฟสัญญาณของคนเดินข้ามบริเวณทางข้ามถนนอัจฉริยะ ซึ่งมีตัวแปรอิสระทั้งหมด 12 ตัวแปร เพื่อให้แน่ใจว่าไม่มีตัวแปรอิสระใดมีความสัมพันธ์กันมากเกินไป (Multicollinearity) ซึ่งอาจส่งผลทำให้การพยากรณ์ตัวแปรตามคลาดเคลื่อน ผลการวิเคราะห์หาค่าสัมประสิทธิ์สหสัมพันธ์ของตัวแปรที่ใช้ในการศึกษาแสดงดังตารางที่ 4-2

ตารางที่ 4-2 ค่าสัมประสิทธิ์สหสัมพันธ์ของตัวแปรที่ใช้ในการศึกษา

ตัวแปร	MediansWidth	Lane	CrosswalksLength	CrosswalksWidth	yellowframeWidth	Speed	VolumeHourplane	WaitingTime	WalkInterval	FDW	Extend	TimeofDay
MediansWidth	1	0.58**	0.82**	0.61**	0.65**	0.31**	0.00	0.25**	0.30**	0.35**	0.01**	0.17**
Lane		1	0.82**	0.37**	0.07**	0.63**	0.08**	0.06**	0.62**	0.18**	0.60**	0.19**
CrosswalksLength			1	0.81**	0.60**	0.52**	0.00	0.32**	0.74**	0.09**	0.06**	0.23**
CrosswalksWidth				1	0.86**	0.25**	-0.10**	0.50**	0.73**	-0.17**	-0.51**	0.19**
yellowframeWidth					1	-0.04**	-0.11**	0.54**	0.43**	-0.09**	-0.71**	0.13**
Speed						1	0.10**	0.04**	0.41**	0.10**	0.40**	0.24**
VolumeHourplane							1	0.00	-0.02**	0.01**	0.15**	-0.14**
WaitingTime								1	0.30**	-0.16**	-0.41**	0.02**
WalkInterval									1	-0.25**	-0.06**	0.17**
FDW										1	0.33**	0.06**
Extend											1	0.02**
TimeofDay												1

หมายเหตุ ** มีระดับนัยสำคัญทางสถิติที่ 0.05

จากตารางที่ 4-2 ผลการวิเคราะห์ค่าสัมประสิทธิ์สหสัมพันธ์ของตัวแปรอิสระที่ใช้ในการศึกษาทั้งหมด 13 ตัวแปร 78 คู่ เมื่อทำการพิจารณาค่า Sig.(2-tailed) ที่ระดับนัยสำคัญที่ 0.05 พบว่า ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปร 3 คู่ ไม่มีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติ ($p > 0.05$) ได้แก่ ปริมาณจราจรที่สัญจรผ่านทางข้ามกับความกว้างเกาะกลางถนน, ปริมาณจราจรที่สัญจรผ่านทางข้ามกับความยาวของทางข้าม และระยะเวลารอคอยกับปริมาณจราจรที่สัญจรผ่านทางข้าม และเมื่อทำการพิจารณาค่าสัมประสิทธิ์สหสัมพันธ์ของตัวแปรอิสระพบว่า มีคู่ของตัวแปรอิสระที่มีความสัมพันธ์อยู่ในระดับมากจำนวน 4 คู่ ได้แก่ ความยาวของทางข้ามกับความกว้างเกาะกลางถนน มีค่าความสัมพันธ์อยู่ที่ 0.82, ความยาวของทางข้ามกับจำนวนช่องจราจร มีค่าความสัมพันธ์อยู่ที่ 0.82, ความกว้างของทางข้ามกับความยาวของทางข้าม มีค่าความสัมพันธ์อยู่ที่ 0.81, และความกว้างเส้นทแยงห้ามหยุดรถกับความกว้างของทางข้าม มีค่าความสัมพันธ์อยู่ที่ 0.86

การคัดเลือกตัวแปรอิสระเพื่อใช้ในการศึกษาในครั้งนี้ ผู้วิจัยได้ทำการคัดเลือกเฉพาะข้อมูลที่คาดว่าจะเกี่ยวข้องกับคนเดินข้ามที่ฝ่าฝืนไฟสัญญาณโดยตรง โดยทำการพิจารณาจากคำนิยามของคนเดินข้ามที่ฝ่าฝืน พิจารณาเฉพาะปัจจัยที่เกิดระหว่างช่วงเวลาที่พบคนเดินข้ามกดปุ่มขอไฟสัญญาณจนถึงเวลาที่คนเดินข้ามได้ไฟสัญญาณให้ข้าม พิจารณาจากการหาสัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient) และพิจารณาจากการทบทวนวรรณกรรมที่ผ่านมา ตัวแปรอิสระที่ผ่านเกณฑ์การคัดเลือกเพื่อเข้าสู่สมการเพื่อสร้างแบบจำลอง ได้แก่ ระยะเวลารอคอย (วินาที/รอบสัญญาณไฟ), ความยาวของทางข้าม (เมตร), ความเร็วของยานพาหนะ (กิโลเมตร/ชั่วโมง), ปริมาณจราจรที่สัญจรผ่านทางข้าม (คัน/ชั่วโมง/ช่องจราจร) และช่วงเวลาของวันที่พบเหตุการณ์การฝ่าฝืนไฟสัญญาณของคนเดินข้าม (กลางวันและกลางคืน)

4.3 วิเคราะห์สร้างแบบจำลองจากชุดข้อมูลดั้งเดิมโดยไม่แก้ไขคุณลักษณะที่ไม่สมดุล

หลังจากที่ได้ทำการคัดเลือกตัวแปรอิสระที่จะนำมาใช้ในการศึกษาในครั้งนี้เรียบร้อยแล้ว รายละเอียดตัวแปรตามและตัวแปรอิสระที่ใช้ในการศึกษานี้แสดงดังตารางที่ 4-3

ตารางที่ 4-3 คำอธิบายตัวแปรตามและตัวแปรอิสระในการศึกษาครั้งนี้

ตัวแปรตาม	คำอธิบาย
เหตุการณ์การฝ่าฝืนไฟสัญญาณของคนเดินข้ามที่เกิดขึ้นระหว่างรอบไฟสัญญาณ [VioPedestrian]	0 = ไม่พบคนเดินข้ามฝ่าฝืนไฟสัญญาณ (Not Violated) 1 = พบคนเดินข้ามอย่างน้อยหนึ่งคนฝ่าฝืนไฟสัญญาณ (Violated)

ตารางที่ 4-3 คำอธิบายตัวแปรตามและตัวแปรอิสระในการศึกษาครั้งนี้ (ต่อ)

ตัวแปรตาม	คำอธิบาย
ระยะเวลารอคอย [WaitingTime]	คำนวณจากช่วงที่คนเดินข้ามกดปุ่มขอไฟสัญญาณคนเดินข้ามจนกระทั่งได้รับสัญญาณให้ข้าม (หน่วย: วินาที)
ความยาวของทางข้าม [CrosswalksLength]	ระยะทางแนวตรงจากจุดเริ่มต้นของขอบหนึ่งของทางข้ามไปยังอีกขอบหนึ่งในทิศทางที่คนเดินเท้าใช้ข้ามถนน (หน่วย: เมตร)
ความเร็วของยานพาหนะ [Speed]	คำนวณความเร็วเฉลี่ยของยานพาหนะทุก ๆ 5 นาที ที่ตรงกับรอบไฟสัญญาณ (หน่วย: กิโลเมตร/ชั่วโมง)
ปริมาณจราจรที่สัญจรผ่านทาง ข้าม [VolumepHourplane]	คำนวณปริมาณจราจรที่ขับเคลื่อนผ่านบริเวณทางข้ามโดยไม่แยกประเภทของยานพาหนะทุก ๆ 5 นาที ที่ตรงกับรอบไฟสัญญาณ (หน่วย: คัน/ชั่วโมง/ช่องจราจร)
ช่วงเวลาของวัน [TimeofDay]	ช่วงเวลาของวันที่พบเหตุการณ์การฝ่าฝืนไฟสัญญาณของคนเดินข้ามช่วงเวลากลางคืน (1 = กลางคืน) ช่วงเวลากลางวัน (0 = กลางวัน)

ทำการวิเคราะห์สร้างแบบจำลองจากชุดข้อมูลดั้งเดิมโดยไม่แก้ไขคุณลักษณะที่ไม่สมดุลโดยใช้วิธีการถดถอยโลจิสติกแบบทวิ (Binary logistic regression) และวิธีต้นไม้ตัดสินใจ (Decision tree) เพื่อตรวจสอบว่าตัวแปรอิสระใดบ้างที่ส่งผลต่อการฝ่าฝืนไฟสัญญาณของคนเดินข้ามและส่งผลอย่างไร ในการศึกษาทำการวิเคราะห์รอบไฟสัญญาณทั้งหมด 116,485 รอบไฟสัญญาณ โดยแบ่งเป็น 2 กลุ่ม ได้แก่ ชุดข้อมูลสำหรับการฝึกเพื่อสร้างแบบจำลอง 93,188 รอบไฟสัญญาณ (80%) และชุดข้อมูลสำหรับการทดสอบการพยากรณ์ของแบบจำลอง 23,297 รอบไฟสัญญาณ (20%) ผลลัพธ์การสร้างแบบจำลองและผลลัพธ์การพยากรณ์การฝ่าฝืนไฟสัญญาณของคนเดินข้ามจากชุดข้อมูลดั้งเดิมโดยไม่แก้ไขคุณลักษณะที่ไม่สมดุลวิธีการถดถอยโลจิสติกแบบทวิ (Binary logistic regression) แสดงดังตารางที่ 4-4 และตารางที่ 4-5 ตามลำดับ

ตารางที่ 4-4 ผลลัพธ์วิธีการถดถอยโลจิสติกแบบทวิจากชุดข้อมูลดั้งเดิมโดยไม่แก้ไขคุณลักษณะที่ไม่สมดุล

ตัวแปรอิสระ	Coefficient (B)	Standard Error	Exp (B)	P-value
Constant	- 1.045 ***	0.054	0.352	0.000
WaitingTime	0.041 ***	0.001	1.042	0.000
CrosswalksLength	- 0.078 ***	0.002	0.925	0.000
Speed	- 0.018 ***	0.001	0.982	0.000
VolumepHourplane	- 0.001 ***	0.000	0.999	0.000
TimeofDay	0.333 ***	0.035	1.396	0.000

หมายเหตุ *** มีระดับนัยสำคัญทางสถิติที่ 0.001, $R^2 = 0.086$

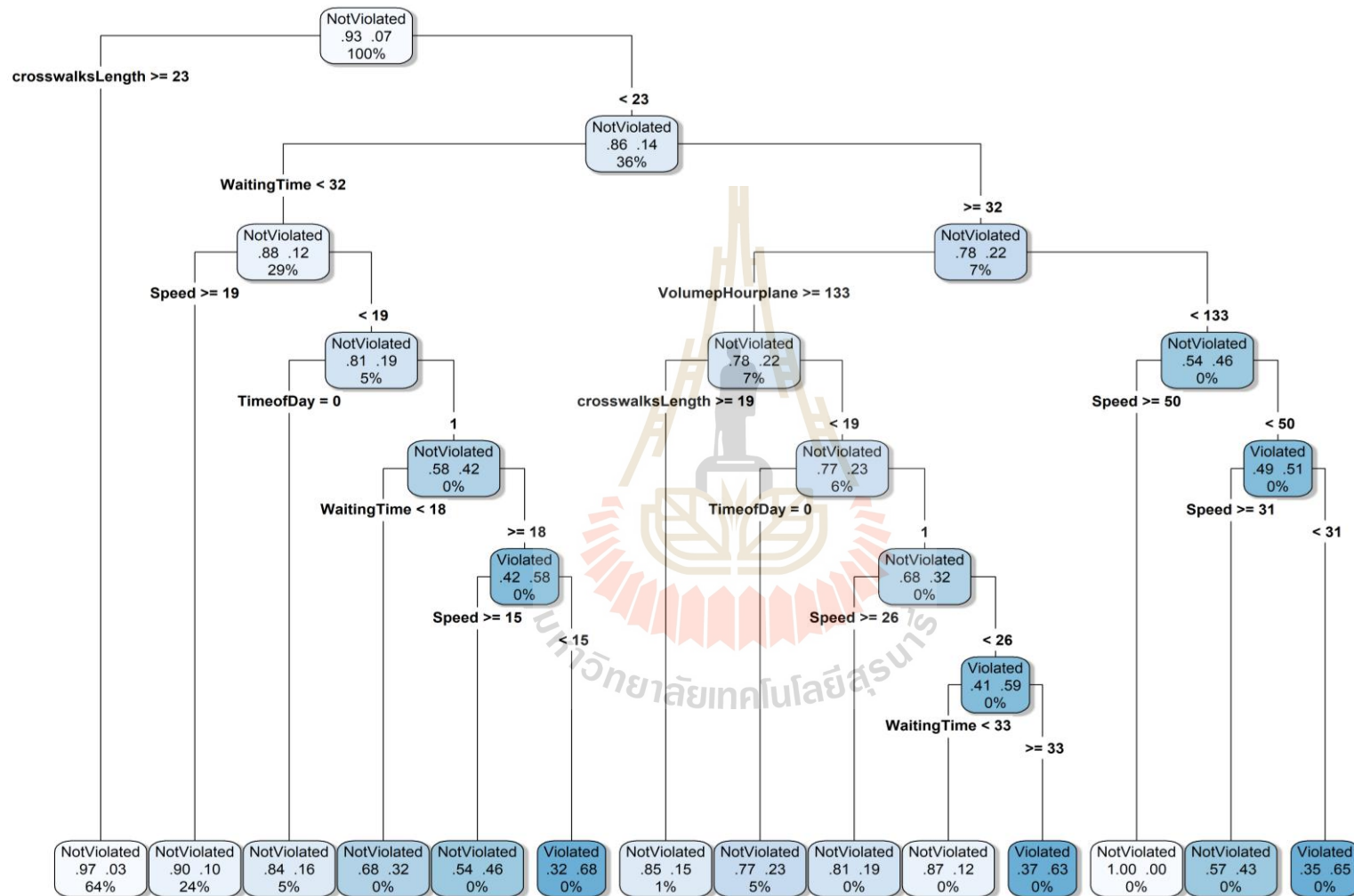
ตารางที่ 4-5 ผลลัพธ์การพยากรณ์ตัวแปรตามของแบบจำลองจากชุดข้อมูลดั้งเดิมโดยไม่แก้ไขคุณลักษณะที่ไม่สมดุลวิธีการถดถอยโลจิสติกแบบทวิ

		Predicted		ค่าความถูกต้อง ของการพยากรณ์	AUC
		Positive (Violated)	Negative (Not Violated)		
Actual	Positive (Violated)	0	1,630	93.00%	50.00%
	Negative (Not Violated)	0	21,698		
ค่าความถูกต้อง (%)		Specificity 100%	Sensitivity 0%		

จากตารางที่ 4-4 แสดงให้เห็นว่าตัวแปรอิสระทั้งหมดมีอิทธิพลทางสถิติที่สำคัญต่อการฝ่าฝืนไฟสัญญาณของคนเดินข้ามที่ระดับ 0.001 ระยะเวลารอคอย และช่วงเวลาของวันมีอิทธิพลในเชิงบวกต่อการฝ่าฝืนไฟสัญญาณของคนเดินข้าม ในขณะที่ความยาวของทางข้าม ความเร็วของยานพาหนะ และปริมาณจราจรที่สัญจรผ่านทางข้าม มีอิทธิพลในเชิงลบต่อการฝ่าฝืนไฟสัญญาณของคนเดินข้าม อย่างไรก็ตาม ตารางที่ 4-5 เผยให้เห็นว่าเมื่อใช้ชุดข้อมูลสำหรับการทดสอบเพื่อพยากรณ์ตัวแปรตาม พบว่าแบบจำลองไม่สามารถพยากรณ์เหตุการณ์การฝ่าฝืนไฟสัญญาณของคนเดินข้ามได้อย่างแม่นยำ โดยมีค่าความไว (Sensitivity) เท่ากับ 0.0% ซึ่งสะท้อนว่าแบบจำลองไม่สามารถตรวจจับเหตุการณ์เป้าหมายได้เลยในชุดข้อมูลสำหรับการทดสอบ

ผลลัพธ์การสร้างแบบจำลองและผลลัพธ์การพยากรณ์การฝ่าฝืนไฟสัญญาณของคนเดินข้าม จากชุดข้อมูลดั้งเดิมโดยไม่แก้ไขคุณลักษณะที่ไม่สมดุลวิธีต้นไม้ตัดสินใจ (Decision tree) แสดงดัง รูปที่ 4-2 และตารางที่ 4-6 ตามลำดับ





รูปที่ 4-2 ผลลัพธ์ของวิธีต้นไม้ตัดสินใจจากชุดข้อมูลดั้งเดิมโดยไม่แก้ไขคุณลักษณะที่ไม่สมดุล

ตารางที่ 4-6 ผลลัพธ์การพยากรณ์ตัวแปรตามของแบบจำลองจากชุดข้อมูลดั้งเดิมโดยไม่แก้ไข
คุณลักษณะที่ไม่สมดุลวิธีต้นไม้ตัดสินใจ

		Predicted		ค่าความถูกต้อง ของการพยากรณ์	AUC
		Positive (Violated)	Negative (Not Violated)		
Actual	Positive (Violated)	36	1,594	93.08%	51.06%
	Negative (Not Violated)	18	21,649		
ค่าความถูกต้อง (%)		Specificity	Sensitivity		
		99.92%	2.2%		

จากรูปที่ 4-2 ผลลัพธ์ของวิธีต้นไม้ตัดสินใจจากชุดข้อมูลดั้งเดิมแสดงให้เห็นว่าตัวแปรอิสระทั้งหมดมีอิทธิพลต่อการฝ่าฝืนไฟสัญญาณของคนเดินข้าม อย่างไรก็ตาม ตารางที่ 4-6 เผยให้เห็นว่าเมื่อใช้ชุดข้อมูลสำหรับการทดสอบเพื่อพยากรณ์ตัวแปรตาม พบว่าแบบจำลองไม่สามารถพยากรณ์เหตุการณ์การฝ่าฝืนไฟสัญญาณของคนเดินข้ามได้อย่างแม่นยำ โดยมีค่าความไว (Sensitivity) เท่ากับ 2.2% ซึ่งสะท้อนว่าแบบจำลองไม่สามารถตรวจจับเหตุการณ์เป้าหมายได้เลยในชุดข้อมูลสำหรับการทดสอบ

4.4 ดำเนินการแก้ไขคุณลักษณะที่ไม่สมดุล (Imbalanced Data)

จากผลลัพธ์การวิเคราะห์สร้างแบบจำลองและผลลัพธ์การพยากรณ์การฝ่าฝืนไฟสัญญาณของคนเดินข้ามจากชุดข้อมูลดั้งเดิมโดยไม่แก้ไขคุณลักษณะที่ไม่สมดุล โดยใช้วิธีการถดถอยโลจิสติกแบบทวิ (Binary logistic regression) และวิธีต้นไม้ตัดสินใจ (Decision tree) สอดคล้องกับผลการศึกษาก่อนหน้านี้ (Modeling with R, 2019; นพมาศ อัครจันทโชติ และดิเรก พินตสุภาภมร, 2561; วิษณุวิสิฐ เกษรสิทธิ์, วิชิต หล่อจ๊ะชุมห์กุล และจิราวัลย์ จิตรถเวช, 2561; อัจฉรา แผ้วบาง และสายชล สันสมบุรณ์ทอง, 2563) เกี่ยวกับชุดข้อมูลที่ไม่สมดุล เมื่อชุดข้อมูลที่นำมาวิเคราะห์มีความไม่สมดุลของข้อมูลเกิดขึ้น ในกระบวนการสร้างแบบจำลองการเรียนรู้ (Training Model) ผลการวิเคราะห์ที่ได้จะมีความโน้มเอียงไปในทิศทางของคลาสส่วนมาก เพื่อดำเนินการแก้ไขปัญหาและปรับปรุงความแม่นยำในการพยากรณ์ มีวิธีในการปรับปรุงผลการพยากรณ์ในกรณีดังกล่าวได้หลากหลายวิธี เช่น Upsampling, Downsampling, ROSE, and SMOTE ผู้วิจัยจึงทำการประยุกต์ใช้หลักการทั้ง 4 วิธี ในการปรับจำนวนของรอบไฟสัญญาณที่ไม่มีการฝ่าฝืนของคนเดินข้ามและจำนวนรอบไฟสัญญาณที่มีการฝ่าฝืนของคนเดินข้ามให้มีจำนวนรอบไฟสัญญาณที่เท่ากันหรือ

ใกล้เคียงกัน โดยทำการกำหนดสร้างเลขสุ่มเทียม ทั้งหมด 30 ค่า และทำการวิเคราะห์ข้อมูลสถิติเชิงพรรณนารายละเอียดของคลาสในแต่ละหลักการ ซึ่งประกอบไปด้วย การหาค่าความถี่ (Frequency), ค่าร้อยละ (Percentage), ค่าเฉลี่ย (Mean), ค่าเบี่ยงเบนมาตรฐาน (S.D.), ค่าต่ำสุด (Min), ค่าสูงสุด (Max), ค่ามัธยฐาน (Median), และตำแหน่งเปอร์เซ็นต์ไทล์ (Percentile rank) ชุดข้อมูลสำหรับการฝึกเพื่อสร้างแบบจำลองมีจำนวนรอบไฟสัญญาณมีทั้งหมด 93,188 รอบไฟสัญญาณ โดยพบคลาสส่วนมาก (0 = Not Violated) จำนวน 86,823 รอบไฟสัญญาณ และคลาสส่วนน้อย (1 = Violated) จำนวน 6,365 รอบไฟสัญญาณ ผลการวิเคราะห์ดังแสดงในตารางที่ 4-7



ตารางที่ 4-7 ผลการวิเคราะห์ค่าสถิติพื้นฐานรายละเอียดของคลาสที่มีการแก้ไขคุณลักษณะที่ไม่สมดุลในแต่ละหลักการ

คลาสส่วนมาก (Majority) (หน่วย: รอบไฟสัญญาณ)							
หลักการ	MIN	15 th Percentile	MEAN	MEDIAN	85 th Percentile	MAX	S.D.
Upsampling	86,823.00	86,823.00	86,823.00	86,823.00	86,823.00	86,823.00	0.00
Downsampling	6,365.00	6,365.00	6,365.00	6,365.00	6,365.00	6,365.00	0.00
ROSE	46,299.00	46,424.75	46,558.23	46,529.00	46,694.40	47,094.00	170.86
SMOTE	25,460.00	25,460.00	25,460.00	25,460.00	25,460.00	25,460.00	0.00
คลาสส่วนน้อย (Minority) (หน่วย: รอบไฟสัญญาณ)							
หลักการ	MIN	15 th Percentile	MEAN	MEDIAN	85 th Percentile	MAX	S.D.
Upsampling	86,823.00	86,823.00	86,823.00	86,823.00	86,823.00	86,823.00	0
Downsampling	6,365.00	6,365.00	6,365.00	6,365.00	6,365.00	6,365.00	0
ROSE	46,094.00	46,493.60	46,629.77	46,659.00	46,763.25	46,889.00	170.855
SMOTE	19,095.00	19,095.00	19,095.00	19,095.00	19,095.00	19,095.00	0
ผลรวม (Total) (หน่วย: รอบไฟสัญญาณ)							
หลักการ	MIN	15 th Percentile	MEAN	MEDIAN	85 th Percentile	MAX	S.D.
Upsampling	173,646.00	173,646.00	173,646.00	173,646.00	173,646.00	173,646.00	0
Downsampling	12,730.00	12,730.00	12,730.00	12,730.00	12,730.00	12,730.00	0
ROSE	93,188.00	93,188.00	93,188.00	93,188.00	93,188.00	93,188.00	0
SMOTE	44,555.00	44,555.00	44,555.00	44,555.00	44,555.00	44,555.00	0

จากตารางที่ 4-7 พบค่าเฉลี่ยของแต่ละหลักการ ดังนี้ หลักการ Upsampling พบค่าเฉลี่ยของจำนวนคลาสส่วนมากและจำนวนคลาสส่วนน้อยอยู่ที่ 86,823.00 รอบไฟสัญญาณ, หลักการ Downsampling พบค่าเฉลี่ยของจำนวนคลาสส่วนมากและจำนวนคลาสส่วนน้อยอยู่ที่ 6,365.00 รอบไฟสัญญาณ, หลักการ ROSE พบค่าเฉลี่ยของจำนวนคลาสส่วนมากอยู่ที่ 46,558.23 รอบไฟสัญญาณ และจำนวนคลาสส่วนน้อยอยู่ที่ 46,629.77 รอบไฟสัญญาณ และหลักการ SMOTE พบค่าเฉลี่ยของจำนวนคลาสส่วนมากอยู่ที่ 25,460.00 รอบไฟสัญญาณ และจำนวนคลาสส่วนน้อยอยู่ที่ 19,095.00 รอบไฟสัญญาณ

4.5 วิเคราะห์สร้างแบบจำลองจากชุดข้อมูลที่มีการปรับแก้ไขคุณลักษณะที่ไม่สมดุล

หลังจากใช้เทคนิคการแก้ไขคุณลักษณะที่ไม่สมดุลจากทั้ง 4 วิธีในข้างต้นแล้ว ดำเนินการวิเคราะห์สร้างแบบจำลองขึ้นใหม่โดยใช้วิธีการถดถอยโลจิสติกแบบทวิ (Binary logistic regression) และวิธีต้นไม้ตัดสินใจ (Decision tree) เพื่อเป็นการตอบวัตถุประสงค์งานวิจัยและคำถามงานวิจัยในการศึกษาครั้งนี้ ผู้วิจัยดำเนินการคัดเลือกแบบจำลองจากชุดข้อมูลที่มีการทดลองแก้ไขคุณลักษณะที่ไม่สมดุลทั้ง 4 วิธี ที่มีประสิทธิภาพสูงที่สุดในการพยากรณ์การฝ่าฝืนไฟสัญญาณของคนเดินข้าม

4.5.1 แบบจำลองวิธีการถดถอยโลจิสติกแบบทวิ (Binary logistic regression)

ผลลัพธ์การวิเคราะห์สถิติเชิงพรรณนาของประสิทธิภาพการพยากรณ์ตัวแปรตามของแบบจำลองจากชุดข้อมูลที่มีการแก้ไขคุณลักษณะที่ไม่สมดุล โดยใช้วิธีการถดถอยโลจิสติกแบบทวิ (Binary logistic regression) ด้วยชุดข้อมูลการทดสอบ ผ่านการกำหนดสร้างเลขสุ่มเทียมทั้งหมด 30 ค่า โดยค่าเฉลี่ย (Mean) และค่าเบี่ยงเบนมาตรฐาน (S.D.) ของแต่ละตัวชี้วัดในแต่ละวิธีการรายละเอียดแสดงดังตารางที่ 4-8

ตารางที่ 4-8 ผลลัพธ์ประสิทธิภาพการพยากรณ์ตัวแปรตามของแบบจำลองด้วยวิธีการถดถอยโลจิสติกแบบทวิ จากชุดข้อมูลดั้งเดิมและชุดข้อมูลที่มีการแก้ไข
คุณลักษณะที่ไม่สมดุล

ตัวชี้วัดประสิทธิภาพการพยากรณ์	ข้อมูล ดั้งเดิม	วิธี Upsampling		วิธี Downsampling		วิธี ROSE		วิธี SMOTE	
		MEAN	S.D.	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.
ค่าความจำเพาะ (Specificity)	100.0%	67.6%	0.05%	67.6%	0.18%	67.7%	0.20%	79.1%	0.16%
ค่าความไว (Sensitivity)	0.0%	68.1%	0.08%	68.3%	0.23%	67.9%	0.22%	52.0%	0.35%
ค่าความแม่นยำโดยรวม (Accuracy)	93.0%	67.6%	0.04%	67.7%	0.16%	67.7%	0.17%	77.2%	0.13%
พื้นที่ใต้เส้นโค้ง (AUC)	50.0%	67.9%	0.03%	68.0%	0.10%	67.8%	0.05%	65.6%	0.10%

จากตารางที่ 4-8 ผลลัพธ์ประสิทธิภาพการพยากรณ์ตัวแปรตามของแบบจำลองด้วยวิธีการถดถอยโลจิสติกแบบทวิ จากชุดข้อมูลดั้งเดิมและชุดข้อมูลที่มีการแก้ไขคุณลักษณะที่ไม่สมดุล ได้แก่ ค่าความจำเพาะ (Specificity) ค่าความไว (Sensitivity) ค่าความแม่นยำโดยรวม (Accuracy) และพื้นที่ใต้เส้นโค้ง (AUC) พบว่าการแก้ไขคุณลักษณะที่ไม่สมดุลของข้อมูลทั้ง 4 วิธี ส่งผลให้ค่าความไว (Sensitivity) ของแบบจำลองสูงขึ้น เมื่อเปรียบเทียบกับประสิทธิภาพการพยากรณ์ตัวแปรตามของแบบจำลองที่ไม่มีการแก้ไขคุณลักษณะที่ไม่สมดุล สะท้อนถึงความสามารถในการตรวจจับพฤติกรรมกรรมการฝ่าฝืนไฟสัญญาณของคนเดินข้ามได้ดีมากยิ่งขึ้น ซึ่งมีความสำคัญสำหรับการศึกษาในครั้งนี้ นอกจากนี้ เมื่อพิจารณาค่าพื้นที่ใต้เส้นโค้ง (AUC) ซึ่งผู้วิจัยให้ความสำคัญเป็นหลัก เนื่องจากค่า AUC เป็นดัชนีที่แสดงถึงความสามารถของแบบจำลองในการจำแนกประเภทได้อย่างมีประสิทธิภาพทั้งในกลุ่มของคนเดินข้ามที่ฝ่าฝืนและไม่ฝ่าฝืนไฟสัญญาณ พบว่าการแก้ไขคุณลักษณะที่ไม่สมดุลของข้อมูลทั้ง 4 วิธี ส่งผลให้ค่า AUC สูงขึ้น จึงสามารถสรุปได้ว่า การปรับสมดุลข้อมูลส่งผลให้แบบจำลองมีประสิทธิภาพในการพยากรณ์สูงขึ้น เมื่อเทียบกับประสิทธิภาพการพยากรณ์ตัวแปรตามของแบบจำลองที่ไม่มีการแก้ไขคุณลักษณะที่ไม่สมดุล ผลลัพธ์นี้สอดคล้องกับงานวิจัยที่ผ่านมา (นพมาศ อัครจันทโชติ และ ดิเรก พนิตสุภาภมร, 2561; วิชญวิสิฐ เกษรสิทธิ์, วิชิต หล่อจรัส ชุณหกุล และ จิราวัลย์ จิตตเวช, 2561) ซึ่งผลการศึกษาแสดงให้เห็นว่า การจัดการกับปัญหาข้อมูลไม่สมดุลสามารถช่วยเพิ่มความถูกต้องและความแม่นยำในการจำแนกข้อมูลกลุ่มส่วนน้อยได้ดียิ่งขึ้น

จากนั้นผู้วิจัยดำเนินการใช้สถิติ One-Way ANOVA (F-test) ในการวิเคราะห์นัยสำคัญความแตกต่างระหว่างค่าเฉลี่ย AUC ของประชากรมากกว่า 2 กลุ่ม ซึ่งช่วยยืนยันว่าความแตกต่างของผลลัพธ์ที่ได้จากแต่ละวิธีมีนัยสำคัญทางสถิติ มิได้เกิดจากความผันแปรโดยบังเอิญ ในกระบวนการสุ่มข้อมูล แสดงดังตารางที่ 4-9 ถึงตารางที่ 4-11

ตารางที่ 4-9 การตรวจสอบความเท่ากันของความแปรปรวนระหว่างกลุ่มของค่า AUC จากวิธีการถดถอยโลจิสติกแบบทวิ ด้วย Levene's test (center = mean)

	Levene Statistic (F)	df Group	df Residuals	p-value
AUC	10.45	3	116	< 0.001 ***

หมายเหตุ: * มีนัยสำคัญทางสถิติที่ระดับ 0.05, ** ที่ระดับ 0.01, *** ที่ระดับ 0.001

จากตารางที่ 4-9 ผลการทดสอบ Levene's Test พบว่า ค่า p-value < 0.001 แสดงว่าค่าความแปรปรวนของแต่ละกลุ่มแตกต่างกันอย่างมีนัยสำคัญ ดังนั้นจึงจำเป็นต้องใช้สถิติทดสอบ Welch's ANOVA ในการทดสอบความแตกต่างของค่าเฉลี่ยระหว่างกลุ่ม ผลการทดสอบแสดงดังตารางที่ 4-10

ตารางที่ 4-10 ผลการทดสอบความแตกต่างของค่าเฉลี่ย AUC ระหว่างกลุ่ม จากวิธีการถดถอย
โลจิสติกแบบทวิ ด้วย Welch's ANOVA

	F-value	df (num)	df (denom)	p-value
AUC	4,601.20	3	56.00	< 0.001 ***

หมายเหตุ: * มีนัยสำคัญทางสถิติที่ระดับ 0.05, ** ที่ระดับ 0.01, *** ที่ระดับ 0.001

จากตารางที่ 4-10 พบว่าค่าเฉลี่ย AUC ระหว่างกลุ่ม จากวิธีการถดถอยโลจิสติกแบบทวิมีความแตกต่างอย่างมีนัยสำคัญทางสถิติ [$F(3, 56.00) = 4601.20, p < 0.001$] ซึ่งบ่งชี้ว่าการสุ่มตัวอย่างของข้อมูลที่มีวิธีการที่แตกต่างกัน จะมีค่าเฉลี่ยของค่า AUC แตกต่างกัน ผู้วิจัยจึงใช้การเปรียบเทียบความแตกต่างของค่าเฉลี่ยระหว่างกลุ่ม (Multiple Comparison) โดยใช้วิธีทดสอบแบบ Dunnett's T3 test เพื่อพิจารณาว่าค่าเฉลี่ยของคู่ใดบ้างที่แตกต่างกัน ผลการทดสอบแสดงดังตารางที่ 4-11

ตารางที่ 4-11 การทดสอบความแตกต่างของค่าเฉลี่ย AUC จากวิธีการถดถอยโลจิสติกแบบทวิรายคู่ ด้วยวิธีทดสอบแบบ Dunnett T3

p-value	Upsampling	Downsampling	ROSE	SMOTE
Upsampling	-	0.114	< 0.001***	< 0.001***
Downsampling	-	-	< 0.001***	< 0.001***
ROSE	-	-	-	< 0.001***

หมายเหตุ: * มีนัยสำคัญทางสถิติที่ระดับ 0.05, ** ที่ระดับ 0.01, *** ที่ระดับ 0.001

จากตารางที่ 4-11 พบว่า ค่าเฉลี่ยของค่า AUC ของการสุ่มตัวอย่างของข้อมูลด้วยวิธี Upsampling กับ วิธี Downsampling จากวิธีการถดถอยโลจิสติกแบบทวิ ไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.001 ในขณะที่กลุ่มอื่น ๆ มีความแตกต่างอย่างมีนัยสำคัญทั้งหมด ($p < 0.001$)

จากผลลัพธ์ข้างต้น ผู้วิจัยจึงตัดสินใจใช้การสุ่มตัวอย่างแบบ Downsampling เพื่อปรับสัดส่วนของรอบไฟสัญญาณ เนื่องจากวิธีนี้ให้ค่าเฉลี่ย AUC สูงสุดที่ 68.0% และเนื่องจากข้อมูลรอบไฟสัญญาณมีจำนวนมาก การเลือกใช้วิธี Downsampling จะช่วยลดขนาดข้อมูล ทำให้การฝึกโมเดลมีประสิทธิภาพและรวดเร็วขึ้น โดยไม่สูญเสียความหลากหลายของกลุ่มคนเดินข้ามที่ฝ่าฝืนไฟสัญญาณ ผลลัพธ์วิธีการถดถอยโลจิสติกแบบทวิ โดยการใช้การสุ่มตัวอย่างแบบ Downsampling แสดงดังตารางที่ 4-12 และแสดงดังรูปที่ 4-3

ตารางที่ 4-12 ผลลัพธ์วิธีการถดถอยโลจิสติกแบบทวิ โดยใช้การสุ่มตัวอย่างแบบ Downsampling

ตัวแปรอิสระ	Coefficient (B)	Standard Error	Exp (B)	P-value
Constant	1.287 ***	0.077	3.622	0.000
WaitingTime	0.046 ***	0.002	1.047	0.000
CrosswalksLength	- 0.080 ***	0.003	0.923	0.000
Speed	- 0.014 ***	0.002	0.986	0.000
VolumepHourplane	- 0.001 ***	0.000	0.999	0.000
TimeofDay	0.339 ***	0.050	1.403	0.000

หมายเหตุ *** มีระดับนัยสำคัญทางสถิติที่ 0.001, $R^2 = 0.1149$, Specificity = 67.51%, Sensitivity = 68.83%, Accuracy = 67.60%, AUC = 68.17%

จากตารางที่ 4-12 แสดงแบบจำลองการฝ่าฝืนไฟสัญญาณของคนเดินข้ามที่พัฒนาขึ้นโดยใช้วิธีการ Downsampling แบบจำลองนี้ระบุว่าตัวแปรอิสระแต่ละตัวมีผลกระทบทางสถิติที่มีนัยสำคัญต่อการฝ่าฝืนไฟสัญญาณของคนเดินข้ามที่ระดับ 0.001 ผลการวิเคราะห์พบว่าโมเดลมีค่า R^2 เท่ากับ 0.1149 ซึ่งแสดงว่าตัวแปรอิสระที่ใช้ในโมเดลสามารถอธิบายความแปรปรวนของตัวแปรตามได้ประมาณร้อยละ 11.49 ในด้านของความสัมพันธ์ระหว่างตัวแปรอิสระและตัวแปรตามในแบบจำลองนี้ พบว่ามีแนวโน้มความสัมพันธ์คล้ายคลึงกับความสัมพันธ์ที่สังเกตได้จากแบบจำลองวิธีการถดถอยโลจิสติกแบบทวิจากชุดข้อมูลสำหรับการฝึกตั้งเดิมโดยไม่แก้ไขคุณลักษณะที่ไม่สมดุล และพบว่าแบบจำลองมีประสิทธิภาพในการทำนายคลาสส่วนน้อยได้ดีขึ้น คำอธิบายผลลัพธ์ของตัวแปรแต่ละตัวจากแบบจำลองการฝ่าฝืนไฟสัญญาณของคนเดินข้ามที่พัฒนาขึ้นโดยใช้วิธีการ Downsampling มีดังนี้:

ระยะเวลารอคอย ($B = 0.046$ และ $\text{Exp}(B) = 1.047$) พบว่า ภายใต้สภาพที่ปัจจัยอื่นมีค่าคงที่หากระยะเวลารอคอยเพิ่มขึ้น 1 วินาที ส่งผลให้คนเดินข้ามมีโอกาที่จะฝ่าฝืนไฟสัญญาณเพิ่มขึ้น 0.047 หน่วย หรือเพิ่มขึ้นประมาณ 4.7 % ผลการวิเคราะห์นี้สอดคล้องกับงานวิจัยของ Afshari and Ayati (2021), Raoniar et al. (2022a), Raoniar and Maurya (2022b, 2022c), Zhang and Deng (2019) และ Zhu et al. (2021) ซึ่งผลการศึกษาล้วนชี้ให้เห็นว่า ระยะเวลารอคอยที่ยาวนานขึ้นจะส่งผลให้คนเดินข้ามมีแนวโน้มที่จะฝ่าฝืนไฟสัญญาณเพิ่มมากขึ้น

ความยาวของทางข้าม ($B = -0.080$ และ $\text{Exp}(B) = 0.923$) พบว่า ภายใต้สภาพที่ปัจจัยอื่นมีค่าคงที่หากความยาวของทางข้ามเพิ่มขึ้น 1 เมตร ส่งผลให้คนเดินข้ามมีโอกาที่จะฝ่าฝืนไฟสัญญาณลดลง 0.077 หน่วย หรือลดลงประมาณ 7.7 % ผลการวิเคราะห์นี้สอดคล้องกับงานวิจัยของ Afshari and Ayati (2021) ซึ่งผลการศึกษาก็ชี้ให้เห็นว่า การเพิ่มขึ้นของความยาวของทางข้าม

ส่งผลในทางลบซึ่งจะช่วยลดความน่าจะเป็นของการฝ่าฝืนไฟสัญญาณสีแดงของคนเดินข้าม

ความเร็วของยานพาหนะ ($B = -0.014$ และ $\text{Exp}(B) = 0.986$) พบว่า ภายใต้อุปกรณ์ที่ปัจจัยอื่นมีค่าคงที่หากความเร็วในการขับเคลื่อนของยานพาหนะเพิ่มขึ้น 1 กม./ชม. ส่งผลให้คนเดินข้ามมีโอกาสที่จะฝ่าฝืนไฟสัญญาณลดลง 0.014 หน่วย หรือลดลงประมาณ 1.4 % ผลการวิเคราะห์นี้สอดคล้องกับงานวิจัยของ วีรพงศ์ บุญเจริญปัญญา และจำรัส พิทักษ์ศฤงคาร (2567) ซึ่งผลการศึกษาชี้ให้เห็นว่า เมื่อความเร็วของยานพาหนะเพิ่มขึ้น จะทำให้อัตราความรู้สึกลดภัยของคนเดินข้ามลดลงอย่างมีนัยสำคัญ ซึ่งส่งผลให้คนเดินข้ามมีโอกาสที่จะฝ่าฝืนไฟสัญญาณลดลง

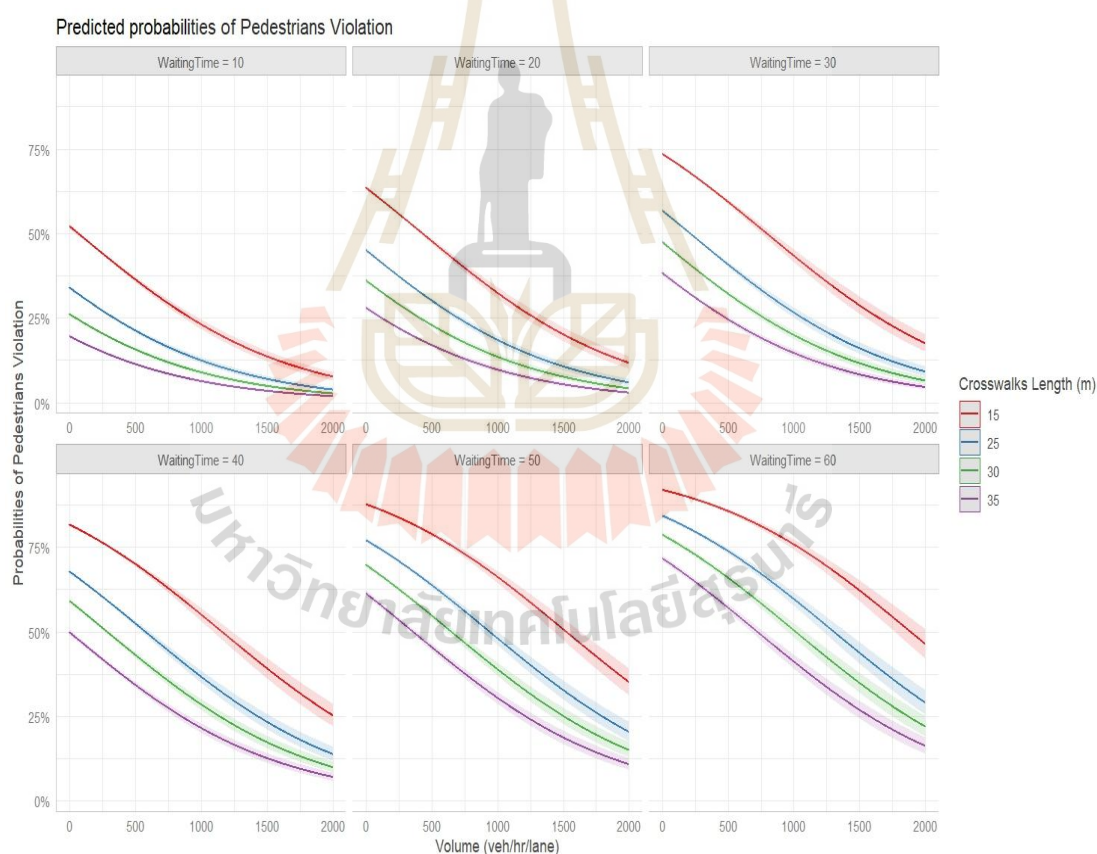
ปริมาณจราจรที่สัญจรผ่านทางข้าม ($B = -0.001$ และ $\text{Exp}(B) = 0.999$) พบว่า ภายใต้อุปกรณ์ที่ปัจจัยอื่นมีค่าคงที่หากปริมาณจราจรเพิ่มขึ้น 1 คัน/ชม/ช่องจราจร ส่งผลให้คนเดินข้ามมีโอกาสที่จะฝ่าฝืนไฟสัญญาณลดลง 0.001 หน่วย หรือลดลงประมาณ 0.1 % ผลการวิเคราะห์นี้สอดคล้องกับงานวิจัยของ Afshari and Ayati (2021), Raoniar et al. (2022a), Raoniar and Maurya (2022b), Zhu et al. (2021) รวมถึง วีรพงศ์ บุญเจริญปัญญา และจำรัส พิทักษ์ศฤงคาร (2567) ซึ่งผลการศึกษาชี้ให้เห็นว่า เมื่อปริมาณจราจรเพิ่มขึ้น คนเดินข้ามจะมีแนวโน้มที่จะฝ่าฝืนสัญญาณไฟน้อยลง เนื่องจากเมื่อมีปริมาณยานพาหนะมากขึ้น ช่องว่างปลอดภัยสำหรับการข้ามจะลดลง

และช่วงเวลาของวัน ($B = 0.339$ และ $\text{Exp}(B) = 1.403$) พบว่า ภายใต้อุปกรณ์ที่ปัจจัยอื่นมีค่าคงที่หากช่วงเวลาของวันเป็นช่วงเวลากลางคืน จะส่งผลให้คนเดินข้ามมีโอกาสที่จะฝ่าฝืนไฟสัญญาณเพิ่มขึ้น 0.403 หน่วย หรือเพิ่มขึ้นประมาณ 40.3 % ผลการวิเคราะห์นี้สอดคล้องกับงานวิจัย ของ Rafe et al. (2024) ซึ่งผลการศึกษาชี้ให้เห็นว่า ปัจจัยแวดล้อมอย่างช่วงเวลากลางคืนหรือช่วงเช้ามีด มีแนวโน้มที่จะเพิ่มโอกาสในการละเมิดไฟสัญญาณจราจรของคนเดินข้าม

ประสิทธิภาพของแบบจำลองวิธีการถดถอยโลจิสติกแบบทวิ โดยไม่แก้ไขคุณลักษณะที่ไม่สมดุลและแบบจำลองที่พัฒนาขึ้นโดยใช้วิธีการ Downsampling ด้วยชุดข้อมูลทดสอบ สามารถเปรียบเทียบได้โดยใช้ตาราง Confusion Matrix ดังที่แสดงในตารางที่ 4-13 แบบจำลองที่พัฒนาขึ้นโดยใช้วิธีการ Downsampling นั้นดีกว่าแบบจำลองจากชุดข้อมูลดั้งเดิมโดยไม่แก้ไขคุณลักษณะที่ไม่สมดุลในการพยากรณ์ผลบวกที่แท้จริง (TP) แต่ในขณะเดียวกันผลบวกปลอม (FP) ก็มีค่าเพิ่มขึ้นไปด้วย กราฟความน่าจะเป็นที่คาดการณ์ว่าคนเดินข้ามจะฝ่าฝืนไฟสัญญาณที่ได้จากแบบจำลองที่พัฒนาขึ้นโดยใช้วิธีการ Downsampling แสดงดังรูปที่ 4-3

ตารางที่ 4-13 ประสิทธิภาพของแบบจำลองจากชุดข้อมูลดั้งเดิมโดยไม่แก้ไขคุณลักษณะที่ไม่สมดุล และแบบจำลองที่พัฒนาขึ้นโดยใช้วิธีการ Downsampling วิธีการถดถอยโลจิสติกแบบทวิ ด้วยชุดข้อมูลทดสอบ

	Original model		Downsampling model	
	Predicted Positive	Predicted Negative	Predicted Positive	Predicted Negative
Actual positive	0 (TP)	1,630 (FN)	1,122 (TP)	508 (FN)
Actual negative	0 (FP)	21,698 (TN)	7,040 (FP)	14,627 (TN)



รูปที่ 4-3 ความน่าจะเป็นที่คาดการณ์ว่าคนเดินข้ามจะฝ่าฝืนไฟสัญญาณ

จากรูปที่ 4-3 แสดงกราฟความน่าจะเป็นที่คาดการณ์ว่าคนเดินข้ามจะฝ่าฝืนไฟสัญญาณ 6 กราฟย่อยโดยแต่ละกราฟแสดงระยะเวลาการรอไฟสัญญาณที่แตกต่างกัน และแสดงความสัมพันธ์ระหว่างความยาวของทางข้าม ปริมาณจราจรที่สัญจรผ่านทางข้าม และการฝ่าฝืนไฟสัญญาณของคนเดินข้าม โดยความยาวทางข้ามแบ่งออกเป็น 15, 25, 30, และ 35 เมตร จากกราฟ

แสดงให้เห็นว่าการฝ่าฝืนไฟสัญญาณของคนเดินข้ามมีแนวโน้มเพิ่มสูงขึ้นเมื่อความยาวของทางข้ามมีระยะทางสั้นลง ระยะเวลารอคอยนานขึ้น และมีปริมาณจราจรที่สัญจรผ่านทางข้ามลดลง

4.5.2 แบบจำลองวิธีต้นไม้ตัดสินใจ (Decision tree)

ผลลัพธ์การวิเคราะห์สถิติเชิงพรรณนาของประสิทธิภาพการพยากรณ์ตัวแปรตามของแบบจำลองจากชุดข้อมูลที่มีการแก้ไขคุณลักษณะที่ไม่สมดุล โดยใช้วิธีต้นไม้ตัดสินใจ (Decision tree) ด้วยชุดข้อมูลการทดสอบ ผ่านการกำหนดสร้างเลขสุ่มเทียมทั้งหมด 30 ค่า โดยค่าเฉลี่ย (Mean) และค่าเบี่ยงเบนมาตรฐาน (S.D.) ของแต่ละตัวชี้วัดในแต่ละวิธีการรายละเอียดแสดงดังตารางที่ 4-14



ตารางที่ 4-14 ผลลัพธ์ประสิทธิภาพการพยากรณ์ตัวแปรตามของแบบจำลองด้วยวิธีต้นไม้ตัดสินใจ จากชุดข้อมูลดั้งเดิมและชุดข้อมูลที่มีการแก้ไขคุณลักษณะที่ไม่สมดุล

ตัวชี้วัดประสิทธิภาพการพยากรณ์	ข้อมูลดั้งเดิม	วิธี Upsampling		วิธี Downsampling		วิธี ROSE		วิธี SMOTE	
		MEAN	S.D.	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.
ค่าความจำเพาะ (Specificity)	99.9%	62.8%	0.77%	62.4%	1.41%	62.6%	3.51%	85.8%	2.24%
ค่าความไว (Sensitivity)	2.2%	77.3%	0.81%	78.0%	2.02%	76.8%	3.99%	40.9%	4.56%
ค่าความแม่นยำโดยรวม (Accuracy)	93.1%	63.8%	0.67%	63.5%	1.17%	63.6%	2.99%	82.6%	1.77%
พื้นที่ใต้เส้นโค้ง (AUC)	51.1%	70.0%	0.14%	70.2%	0.36%	69.7%	0.30%	63.4%	1.17%

จากตารางที่ 4-14 ผลลัพธ์ประสิทธิภาพการพยากรณ์ตัวแปรตามของแบบจำลองด้วยวิธีต้นไม้ตัดสินใจ จากชุดข้อมูลดั้งเดิมและชุดข้อมูลที่มีการแก้ไขคุณลักษณะที่ไม่สมดุล ได้แก่ ค่าความจำเพาะ (Specificity) ค่าความไว (Sensitivity) ค่าความแม่นยำโดยรวม (Accuracy) และพื้นที่ใต้เส้นโค้ง (AUC) พบว่า การแก้ไขคุณลักษณะที่ไม่สมดุลของข้อมูลทั้ง 4 วิธี ส่งผลให้ค่าความไว (Sensitivity) ของแบบจำลองสูงขึ้น เมื่อเปรียบเทียบกับประสิทธิภาพการพยากรณ์ตัวแปรตามของแบบจำลองที่ไม่มีการแก้ไขคุณลักษณะที่ไม่สมดุล สะท้อนถึงความสามารถในการตรวจจับพฤติกรรมการฝ่าฝืนไฟสัญญาณของคนเดินข้ามได้ดีมากยิ่งขึ้น ซึ่งมีความสำคัญสำหรับการศึกษาในครั้งนี้ นอกจากนี้ เมื่อพิจารณาค่าพื้นที่ใต้เส้นโค้ง (AUC) ซึ่งผู้วิจัยให้ความสำคัญเป็นหลัก เนื่องจากค่า AUC เป็นดัชนีที่แสดงถึงความสามารถของแบบจำลองในการจำแนกประเภทได้อย่างมีประสิทธิภาพทั้งในกลุ่มของคนเดินข้ามที่ฝ่าฝืนและไม่ฝ่าฝืนไฟสัญญาณ พบว่าการแก้ไขคุณลักษณะที่ไม่สมดุลของข้อมูลทั้ง 4 วิธี ส่งผลให้ค่า AUC สูงขึ้น จึงสามารถสรุปได้ว่า การปรับสมดุลข้อมูลส่งผลให้แบบจำลองมีประสิทธิภาพในการพยากรณ์สูงขึ้น เมื่อเทียบกับประสิทธิภาพการพยากรณ์ตัวแปรตามของแบบจำลองที่ไม่มีการแก้ไขคุณลักษณะที่ไม่สมดุล ผลลัพธ์นี้สอดคล้องกับงานวิจัยที่ผ่านมา (นพมาศ อัครจันทโชติ และ ดิเรก พนิตสุภากมล, 2561; วิชญวิสิฐ เกษรสิทธิ์, วิชิต หล่อจรัส ชุณหกุล และ จิราวัลย์ จิตตเวช, 2561) ซึ่งผลการศึกษาแสดงให้เห็นว่า การจัดการกับปัญหาข้อมูลที่ไม่สมดุลสามารถช่วยเพิ่มความถูกต้องและความแม่นยำในการจำแนกข้อมูลกลุ่มส่วนน้อยได้ดียิ่งขึ้น

จากนั้นผู้วิจัยดำเนินการใช้สถิติ One-Way ANOVA (F-test) ในการวิเคราะห์นัยสำคัญความแตกต่างระหว่างค่าเฉลี่ย AUC ของประชากรมากกว่า 2 กลุ่ม ซึ่งช่วยยืนยันว่าความแตกต่างของผลลัพธ์ที่ได้จากแต่ละวิธีมีนัยสำคัญทางสถิติ มิได้เกิดจากความผันแปรโดยบังเอิญในกระบวนการสุ่มข้อมูล แสดงดังตารางที่ 4-15 ถึงตารางที่ 4-17

ตารางที่ 4-15 การตรวจสอบความเท่ากันของความแปรปรวนระหว่างกลุ่มของค่า AUC จากวิธีต้นไม้ตัดสินใจ ด้วย Levene's test (center = mean)

	Levene Statistic (F)	df Group	df Residuals	p-value
AUC	13.037	3	116	< 0.001 ***

หมายเหตุ: * มีนัยสำคัญทางสถิติที่ระดับ 0.05, ** ที่ระดับ 0.01, *** ที่ระดับ 0.001

จากตารางที่ 4-15 ผลการทดสอบ Levene's Test พบว่า ค่า p-value < 0.001 แสดงว่าค่าความแปรปรวนของแต่ละกลุ่มแตกต่างกันอย่างมีนัยสำคัญ ดังนั้นจึงจำเป็นต้องใช้สถิติทดสอบ Welch's ANOVA ในการทดสอบความแตกต่างของค่าเฉลี่ยระหว่างกลุ่ม ผลการทดสอบแสดงดังตารางที่ 4-16

ตารางที่ 4-16 ผลการทดสอบความแตกต่างของค่าเฉลี่ย AUC ระหว่างกลุ่ม จากวิธีต้นไม้ตัดสินใจด้วย Welch's ANOVA

	F-value	df (num)	df (denom)	p-value
AUC	323.84	3	56.86	< 0.001 ***

หมายเหตุ: * มีนัยสำคัญทางสถิติที่ระดับ 0.05, ** ที่ระดับ 0.01, *** ที่ระดับ 0.001

จากตารางที่ 4-16 พบว่าค่าเฉลี่ย AUC ระหว่างกลุ่ม จากวิธีต้นไม้ตัดสินใจมีความแตกต่างอย่างมีนัยสำคัญทางสถิติ [$F(3, 56.86) = 323.84, p < 0.001$] ซึ่งบ่งชี้ว่าการสุ่มตัวอย่างของข้อมูลที่มีวิธีการที่ต่างกัน จะมีค่าเฉลี่ยของค่า AUC แตกต่างกัน ผู้วิจัยจึงใช้การเปรียบเทียบความแตกต่างของค่าเฉลี่ยระหว่างกลุ่ม (Multiple Comparison) โดยใช้วิธีทดสอบแบบ Dunnett's T3 test เพื่อพิจารณาว่าค่าเฉลี่ยของคู่ใดบ้างที่ต่างกัน ผลการทดสอบแสดงดังตารางที่ 4-17

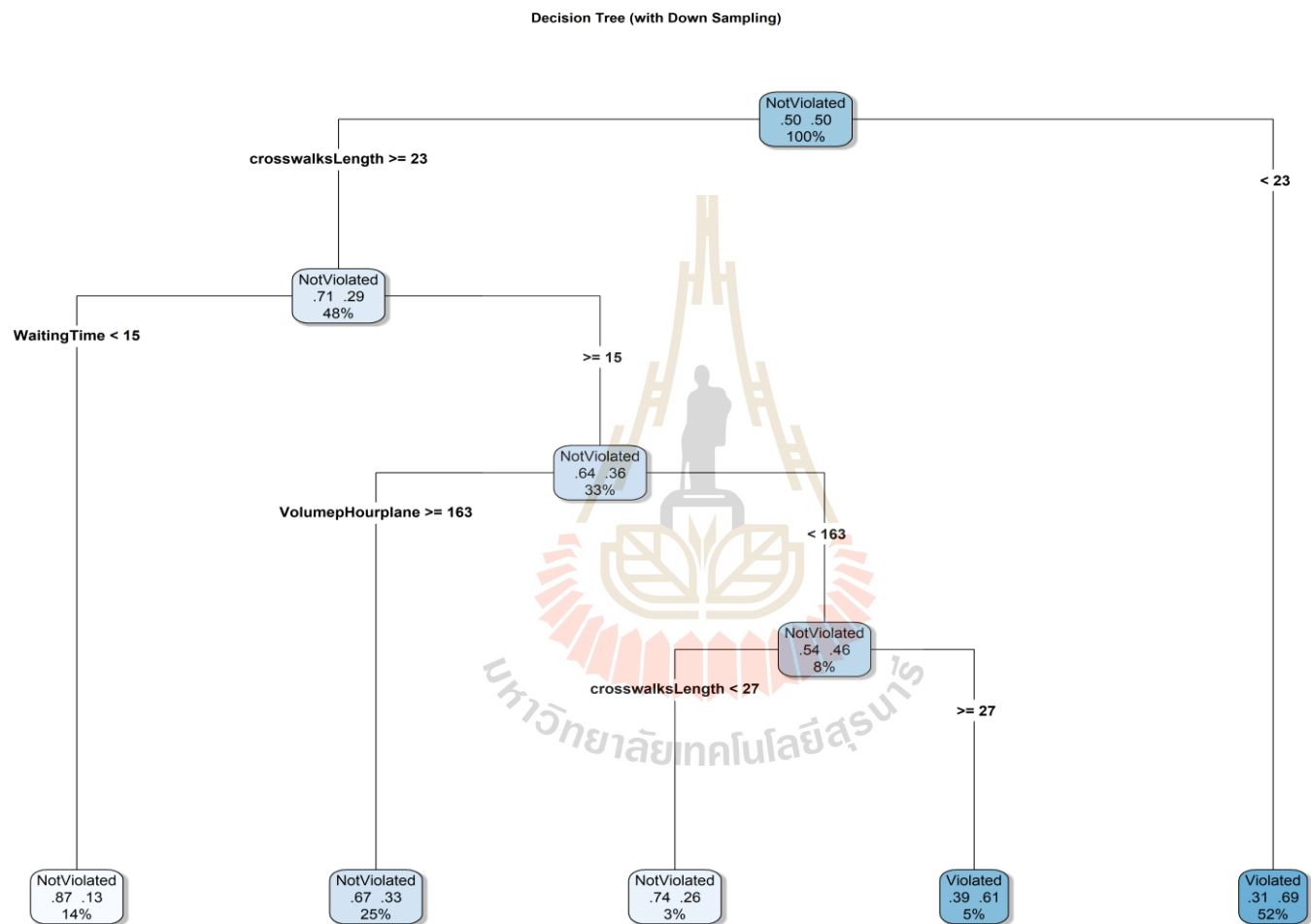
ตารางที่ 4-17 การทดสอบความแตกต่างของค่าเฉลี่ย AUC จากวิธีต้นไม้ตัดสินใจรายคู่ ด้วยวิธีทดสอบแบบ Dunnett T3

p-value	Upsampling	Downsampling	ROSE	SMOTE
Upsampling	-	0.046*	< 0.001***	< 0.001***
Downsampling	-	-	< 0.001***	< 0.001***
ROSE	-	-	-	< 0.001***

หมายเหตุ: * มีนัยสำคัญทางสถิติที่ระดับ 0.05, ** ที่ระดับ 0.01, *** ที่ระดับ 0.001

จากตารางที่ 4-17 พบว่า ค่าเฉลี่ยของค่า AUC ของการสุ่มตัวอย่างของข้อมูลทั้ง 4 วิธีแตกต่างกันอย่างมีนัยสำคัญ วิธี Upsampling กับ วิธี Downsampling แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 และค่าเฉลี่ยของค่า AUC ของการสุ่มตัวอย่างของข้อมูลคู่อื่น ๆ แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.001

จากผลลัพธ์ข้างต้น ผู้วิจัยจึงตัดสินใจใช้การสุ่มตัวอย่างแบบ Downsampling เพื่อปรับสัดส่วนของรอบไฟสัญญาณ เนื่องจากวิธีนี้ให้ค่าเฉลี่ย AUC สูงสุดที่ 70.2% และเนื่องจากข้อมูลรอบไฟสัญญาณมีจำนวนมาก การเลือกใช้วิธี Downsampling จะช่วยลดขนาดข้อมูล ทำให้การฝึกโมเดลมีประสิทธิภาพและรวดเร็วขึ้น โดยไม่สูญเสียความหลากหลายของกลุ่มคนเดินข้ามที่ฝ่าฝืนไฟสัญญาณ ผลลัพธ์จากวิธีต้นไม้ตัดสินใจ โดยใช้การสุ่มตัวอย่างแบบ Downsampling แสดงดังรูปที่ 4-4



รูปที่ 4-4 ผลลัพธ์แบบจำลองวิธีต้นไม้ตัดสินใจ (Decision tree) โดยใช้การสุ่มตัวอย่างแบบ Downsampling

จากรูปที่ 4-4 แสดงแบบจำลองต้นไม้ตัดสินใจการฝ่าฝืนไฟสัญญาณของคนเดินข้าม ที่พัฒนาขึ้นโดยใช้วิธีการ Downsampling ตัวแปรอิสระที่มีอิทธิพลต่อการฝ่าฝืนไฟสัญญาณของคนเดินข้าม ได้แก่ ระยะเวลารอคอย ความยาวของทางข้าม และปริมาณจราจรที่สัญจรผ่านทางข้าม ซึ่งแตกต่างจากผลการวิเคราะห์จากชุดข้อมูลดั้งเดิมโดยไม่แก้ไขคุณลักษณะที่ไม่สมดุล สามารถอธิบาย กฎเกณฑ์ของแบบจำลองได้ว่า

1) ถ้าความยาวทางข้าม (CrosswalksLength) มากกว่าหรือเท่ากับ 23 เมตร และเวลาที่คนเดินข้ามยืนรอไฟสัญญาณ (WaitingTime) น้อยกว่า 15 วินาที คนเดินข้ามจะตัดสินใจที่จะฝ่าฝืนไฟสัญญาณ 13%

2) ถ้าความยาวทางข้าม (CrosswalksLength) มากกว่าหรือเท่ากับ 23 เมตร เวลาที่คนเดินข้ามยืนรอไฟสัญญาณ (WaitingTime) มากกว่าหรือเท่า 15 วินาที และปริมาณจราจรที่สัญจรผ่านทางข้ามมากกว่าหรือเท่ากับ 163 คัน/ชั่วโมง/ช่องจราจร คนเดินข้ามจะตัดสินใจที่จะฝ่าฝืนไฟสัญญาณ 33%

3) ถ้าความยาวทางข้าม (CrosswalksLength) อยู่ระหว่าง 23 ถึง 27 เมตร เวลาที่คนเดินข้ามยืนรอไฟสัญญาณ (WaitingTime) มากกว่าหรือเท่า 15 วินาที และปริมาณจราจรที่สัญจรผ่านทางขำน้อยกว่า 163 คัน/ชั่วโมง/ช่องจราจร คนเดินข้ามจะตัดสินใจที่จะฝ่าฝืนไฟสัญญาณ 26%

4) ถ้าความยาวทางข้าม (CrosswalksLength) มากกว่าหรือเท่ากับ 27 เมตร เวลาที่คนเดินข้ามยืนรอไฟสัญญาณ (WaitingTime) มากกว่าหรือเท่ากับ 15 วินาที และปริมาณจราจรที่สัญจรผ่านทางขำน้อยกว่า 163 คัน/ชั่วโมง/ช่องจราจร คนเดินข้ามจะตัดสินใจที่จะฝ่าฝืนไฟสัญญาณ 61%

5) ถ้าความยาวทางข้าม (CrosswalksLength) น้อยกว่า 23 เมตร คนเดินข้ามจะตัดสินใจที่จะฝ่าฝืนไฟสัญญาณ 69%

และนอกจากนี้ยังพบว่าแบบจำลองมีประสิทธิภาพในการทำนายคลาสส่วนน้อยได้ดีขึ้น ประสิทธิภาพของแบบจำลองวิธีต้นไม้ตัดสินใจ โดยไม่แก้ไขคุณลักษณะที่ไม่สมดุลและแบบจำลองที่พัฒนาขึ้นโดยใช้วิธีการ Downsampling ด้วยชุดข้อมูลทดสอบ สามารถเปรียบเทียบได้โดยใช้ตาราง Confusion Matrix ดังที่แสดงในตารางที่ 4-18 แบบจำลองที่พัฒนาขึ้นโดยใช้วิธีการ Downsampling นั้นดีกว่าแบบจำลองจากชุดข้อมูลดั้งเดิมโดยไม่แก้ไขคุณลักษณะที่ไม่สมดุลในการพยากรณ์ผลบวกที่แท้จริง (TP) แต่ในขณะเดียวกันผลบวกปลอม (FP) ก็มีค่าเพิ่มขึ้นไปด้วย

ตารางที่ 4-18 ประสิทธิภาพของแบบจำลองจากชุดข้อมูลดั้งเดิมโดยไม่แก้ไขคุณลักษณะที่ไม่สมดุล และแบบจำลองที่พัฒนาขึ้นโดยใช้วิธีการ Downsampling วิธีต้นไม้ตัดสินใจ ด้วยชุดข้อมูลทดสอบ

	Original model		Downsampling model	
	Predicted Positive	Predicted Negative	Predicted Positive	Predicted Negative
Actual positive (1,599 cycles)	36 (TP)	1,594 (FN)	1,254 (TP)	376 (FN)
Actual negative (21,698 cycles)	18 (FP)	21,649 (TN)	7,983 (FP)	13,684 (TN)

จากผลการสร้างตัวแบบขึ้นใหม่จากชุดข้อมูลที่มีแก้ไขคุณลักษณะที่ไม่สมดุล โดยใช้การถดถอยโลจิสติกแบบทวิ (Binary logistic regression) และวิธีต้นไม้ตัดสินใจ (Decision tree) เผยให้เห็นว่าตัวแปรอิสระที่มีอิทธิพลต่อการฝ่าฝืนไฟสัญญาณของคนเดินข้ามของวิธีการทั้ง 2 วิธีแตกต่างกัน โดยผลลัพธ์วิธีการถดถอยโลจิสติกแบบทวิ (Binary logistic regression) พบว่าตัวแปรอิสระทั้งหมดมีอิทธิพลทางสถิติที่สำคัญต่อการฝ่าฝืนไฟสัญญาณของคนเดินข้ามที่ระดับ 0.001 ระยะเวลารอคอย และช่วงเวลาของวัน มีอิทธิพลในเชิงบวกต่อการฝ่าฝืนไฟสัญญาณของคนเดินข้าม ในขณะที่ความยาวของทางข้าม ความเร็วของยานพาหนะ และปริมาณจราจรที่สัญจรผ่านทางข้ามมีอิทธิพลในเชิงลบต่อการฝ่าฝืนไฟสัญญาณของคนเดินข้าม มีความสมเหตุสมผลกับการสังเกตภาคสนาม ซึ่งเผยว่าการฝ่าฝืนไฟสัญญาณของคนเดินข้ามเกิดขึ้นบ่อยในเวลากลางคืน เมื่อปริมาณจราจรที่สัญจรผ่านทางข้ามต่ำ และคนเดินข้ามรอคอยไฟสัญญาณนานหลังจากกดปุ่มขอข้าม และผลลัพธ์วิธีต้นไม้ตัดสินใจ (Decision tree) พบว่าตัวแปรอิสระที่มีอิทธิพลต่อการฝ่าฝืนไฟสัญญาณของคนเดินข้าม ได้แก่ ระยะเวลารอคอย ความยาวของทางข้าม และปริมาณจราจรที่สัญจรผ่านทางข้าม ซึ่งเมื่อพิจารณากฎเกณฑ์ที่ได้จากแบบจำลองในแต่ละข้อ ซึ่งประกอบด้วยเงื่อนไขของปัจจัยต่าง ๆ ที่มีความหลากหลาย พบว่า ปัจจัยด้านระยะเวลารอคอยและปริมาณจราจรที่สัญจรผ่านทางข้าม มีความสมเหตุสมผลกับการสังเกตภาคสนาม ซึ่งเผยว่าการฝ่าฝืนไฟสัญญาณของคนเดินข้ามเกิดขึ้นบ่อยเมื่อคนเดินข้ามรอคอยไฟสัญญาณนานหลังจากกดปุ่มขอข้าม และปริมาณจราจรที่สัญจรผ่านทางข้ามต่ำ อย่างไรก็ตาม ปัจจัยด้านความยาวของทางข้ามในบางกฎเกณฑ์แสดงผลที่ไม่สมเหตุสมผลกับการสังเกตภาคสนาม ซึ่งอาจเกิดจากข้อจำกัดของกระบวนการวิธีต้นไม้ตัดสินใจ (decision tree) ที่อาจยังไม่มีประสิทธิภาพเพียงพอในการจำแนกข้อมูลลักษณะนี้ หรืออาจเป็นผลจากข้อจำกัดของข้อมูลตัวแปรความยาวของทางข้าม ซึ่งยังมีความหลากหลายไม่เพียงพอ โดยข้อมูลที่ใช้ในการศึกษาครั้งนี้มาจากพื้นที่สำรวจของทางข้ามอัจฉริยะเพียง 6 แห่ง จึงอาจไม่สามารถสะท้อนลักษณะที่แท้จริงของปัจจัยดังกล่าวได้อย่างครอบคลุมและแม่นยำ

บทที่ 5

สรุป อภิปราย และข้อเสนอแนะ

การศึกษานี้ศึกษาปัจจัยที่มีอิทธิพลต่อการฝ่าฝืนไฟสัญญาณของคนเดินข้ามที่ทางข้ามถนนอัจฉริยะ (Smart Crosswalk) ในประเทศไทย โดยใช้ข้อมูลที่รวบรวมตั้งแต่เดือนมกราคม 2022 ถึงเดือนสิงหาคม 2022 การศึกษาในครั้งนี้ดำเนินการสร้างแบบจำลองใช้วิธีการถดถอยโลจิสติกแบบทวิ (Binary logistic regression) และวิธีต้นไม้ตัดสินใจ (Decision tree) เพื่อตรวจสอบว่าตัวแปรอิสระใดบ้างที่ส่งผลต่อการฝ่าฝืนไฟสัญญาณของคนเดินข้าม โดยมีตัวแปรอิสระทั้งหมด 5 ตัวแปร ได้แก่ ระยะเวลารอคอย (วินาที/รอบสัญญาณไฟ), ความยาวของทางข้าม (เมตร), ความเร็วของยานพาหนะ (กิโลเมตร/ชั่วโมง) ปริมาณจราจรที่สัญจรผ่านทางข้าม (คัน/ชั่วโมง/ช่องจราจร) และช่วงเวลาของวัน วิเคราะห์รอบไฟสัญญาณทั้งหมด 116,485 รอบไฟสัญญาณ โดยแบ่งเป็น 2 กลุ่ม ได้แก่ ชุดข้อมูลสำหรับการฝึก 80% และชุดข้อมูลสำหรับการทดสอบ 20% ผลลัพธ์การสร้างแบบจำลองจากชุดข้อมูลโดยไม่แก้ไขคุณลักษณะที่ไม่สมดุล โดยใช้วิธีการถดถอยโลจิสติกแบบทวิ (Binary logistic regression) และวิธีต้นไม้ตัดสินใจ (Decision tree) พบว่าตัวแปรอิสระทั้งหมดมีอิทธิพลต่อการฝ่าฝืนไฟสัญญาณของคนเดินข้าม อย่างไรก็ตาม เมื่อใช้ชุดข้อมูลสำหรับการทดสอบเพื่อพยากรณ์ตัวแปรตาม พบว่าแบบจำลองจากทั้งสองวิธีไม่สามารถพยากรณ์การฝ่าฝืนไฟสัญญาณของคนเดินข้ามได้อย่างมีประสิทธิภาพ เนื่องจากคุณลักษณะที่ไม่สมดุลของข้อมูล เพื่อดำเนินการแก้ไขปัญหานี้และปรับปรุงความแม่นยำในการพยากรณ์ ผู้วิจัยจึงทำการประยุกต์ใช้หลักการแก้ไขคุณลักษณะที่ไม่สมดุลให้ทั้ง 2 คลาส มีจำนวนรอบไฟสัญญาณที่เท่ากันหรือใกล้เคียงกัน 4 วิธี ได้แก่ Upsampling, Downsampling, ROSE, และ SMOTE ผลลัพธ์ประสิทธิภาพการของพยากรณ์ตัวแปรตามของแบบจำลองจากชุดข้อมูลที่มีการแก้ไขคุณลักษณะที่ไม่สมดุลโดยใช้วิธีการถดถอยโลจิสติกแบบทวิ (Binary logistic regression) และวิธีต้นไม้ตัดสินใจ (Decision tree) พบว่าหลักการทั้ง 4 วิธีส่งผลให้มีความไว (Sensitivity) ที่ดีขึ้น ซึ่งมีความสำคัญสำหรับการศึกษาในครั้งนี้ เมื่อเปรียบเทียบกับประสิทธิภาพการพยากรณ์ตัวแปรตามของแบบจำลองที่ไม่มีการแก้ไขคุณลักษณะที่ไม่สมดุล และการสุ่มตัวอย่างแบบ Downsampling วิธีนี้ให้ค่าพื้นที่ใต้เส้นโค้ง (AUC) สูงที่สุด ผลลัพธ์วิธีการถดถอยโลจิสติกแบบทวิ (Binary logistic regression) โดยใช้การสุ่มตัวอย่างแบบ Downsampling แบบจำลองนี้ระบุว่าตัวแปรอิสระทุกตัวมีผลกระทบทางสถิติที่มีนัยสำคัญต่อการฝ่าฝืนไฟสัญญาณของคนเดินข้ามที่ระดับ 0.001 ระยะเวลารอคอยและช่วงเวลาของวัน มีอิทธิพลในเชิงบวกต่อการฝ่าฝืนไฟสัญญาณของคนเดินข้าม ในขณะที่ความยาวของทางข้าม ความเร็ว

ของยานพาหนะ และปริมาณจราจรที่สัญจรผ่านทางข้าม มีอิทธิพลในเชิงลบต่อการฝ่าฝืนไฟสัญญาณของคนเดินข้าม หรืออาจกล่าวได้ว่าการฝ่าฝืนไฟสัญญาณของคนเดินข้ามมีแนวโน้มเพิ่มสูงขึ้นเมื่อความยาวของทางข้ามมีระยะทางสั้นลง ระยะเวลาการรอคอยนานขึ้น และมีปริมาณจราจรที่สัญจรผ่านทางข้ามลดลง ความสัมพันธ์ของตัวแปรอิสระและตัวแปรตามมีความสมเหตุสมผลและสอดคล้องกับการสังเกตภาคสนาม ซึ่งระบุว่า การฝ่าฝืนไฟสัญญาณของคนเดินข้ามมักเกิดขึ้นบ่อยในเวลากลางคืนเมื่อมีปริมาณจราจรที่สัญจรผ่านทางข้ามน้อย และคนเดินถนนต้องรอคอยไฟสัญญาณเป็นเวลานานหลังจากกดปุ่ม ผลลัพธ์วิธีต้นไม้ตัดสินใจ (Decision tree) โดยการใช้การสุ่มตัวอย่างแบบ Downsampling พบว่ามีตัวแปรอิสระเพียง 3 ตัวแปร ที่มีอิทธิพลต่อการฝ่าฝืนไฟสัญญาณของคนเดินข้าม ได้แก่ ระยะเวลาการรอคอย ความยาวของทางข้าม และปริมาณจราจรที่สัญจรผ่านทางข้าม ซึ่งเมื่อพิจารณาจากเกณฑ์ที่ได้จากแบบจำลอง พบว่า ปัจจัยด้านระยะเวลาการรอคอยและปริมาณจราจรที่สัญจรผ่านทางข้าม มีความสมเหตุสมผลกับการสังเกตภาคสนาม อย่างไรก็ตาม ปัจจัยด้านความยาวของทางข้ามในบางกฎเกณฑ์แสดงผลที่ไม่สมเหตุสมผลกับการสังเกตภาคสนาม ซึ่งอาจเกิดจากข้อจำกัดของวิธีต้นไม้ตัดสินใจ (Decision tree) หรืออาจเกิดจากข้อจำกัดของข้อมูลตัวแปรความยาวของทางข้าม ซึ่งยังมีความหลากหลายไม่เพียงพอ จึงอาจไม่สะท้อนลักษณะที่แท้จริงของปัจจัยดังกล่าวได้อย่างครบถ้วน จากผลการวิเคราะห์ดังกล่าวผู้วิจัยจึงเสนอแนะว่า ในบริบทของข้อมูลลักษณะนี้ควรเลือกใช้วิธีถดถอยโลจิสติกแบบทวี เนื่องจากแม้ใช้ชุดข้อมูลเดียวกัน วิธีดังกล่าวสามารถระบุปัจจัยที่มีความสัมพันธ์อย่างมีนัยสำคัญต่อพฤติกรรมการฝ่าฝืนไฟสัญญาณได้อย่างสมเหตุสมผล และให้ผลลัพธ์ที่สอดคล้องกับการสังเกตภาคสนามมากกว่าแบบจำลองต้นไม้ตัดสินใจ และในส่วนของข้อเสนอแนะเพื่อส่งเสริมความปลอดภัยและลดอัตราการฝ่าฝืนไฟสัญญาณของคนเดินข้าม ผู้วิจัยเสนอความคิดเห็นให้มีการปรับระยะเวลาการรอคอยของคนเดินข้ามให้สอดคล้องกับสภาพการจราจรในแต่ละช่วงเวลาอย่างเหมาะสม เพื่อเป็นการลดระยะเวลาการรอคอยและลดโอกาสในการฝ่าฝืนไฟสัญญาณของคนเดินข้าม อย่างไรก็ตามแม้ว่าแบบจำลองที่พัฒนาขึ้นโดยใช้วิธีการ Downsampling นั้นดีกว่าแบบจำลองจากชุดข้อมูลดั้งเดิมโดยไม่แก้ไขคุณลักษณะที่ไม่สมดุลในการพยากรณ์ผลบวกที่แท้จริง (TP) แต่ในขณะเดียวกันผลบวกปลอม (FP) ก็มีค่าเพิ่มขึ้นไปด้วย และถึงแม้ว่าการปรับขนาดข้อมูลจะช่วยให้ประสิทธิภาพของแบบจำลองดีขึ้นในภาพรวม แต่อัตราความไว ความจำเพาะ ความแม่นยำ และค่า AUC ยังคงอยู่ในระดับที่ค่อนข้างต่ำ (Simundić, 2009) ซึ่งอาจไม่เพียงพอสำหรับการนำไปใช้ในการพยากรณ์อย่างแม่นยำ

ข้อเสนอแนะ

- 1) ควรศึกษาทางข้ามอัจฉริยะ (Smart Crosswalk) ให้ครบทั้ง 11 จุด เพื่อให้สามารถจำลองพฤติกรรมการณ์การฝ่าฝืนไฟสัญญาณของคนเดินข้ามที่ทางข้ามอัจฉริยะ (Smart Crosswalk)

ได้ครอบคลุมมากยิ่งขึ้น

- 2) การวิจัยในอนาคตควรสำรวจเทคนิค Machine learning อื่น ๆ ร่วมด้วย เช่น Random Forests หรือ Gradient Boosting ร่วมกับวิธีการปรับขนาดข้อมูลเพื่อปรับปรุงประสิทธิภาพการพยากรณ์ และตรวจสอบผลการค้นพบเหล่านี้
- 3) การศึกษาปัจจัยที่มีอิทธิพลต่อการฝ่าฝืนไฟสัญญาณของคนเดินข้าม ควรศึกษาแยกในแต่ละพื้นที่เนื่องจากอิทธิพลของแต่ละปัจจัยในแต่ละทางข้ามนั้นมีความแตกต่างกัน ตามพฤติกรรมของคนเดินข้ามในตำแหน่งนั้นๆ และเพื่อให้สามารถนำผลลัพธ์ที่ได้ไปดำเนินการปรับปรุงและแก้ไขการฝ่าฝืนไฟสัญญาณของคนเดินข้ามในแต่ละทางข้ามได้อย่างมีประสิทธิภาพ



รายการอ้างอิง

- กรมควบคุมโรค กระทรวงสาธารณสุข. (2565). การบูรณาการจำนวนผู้เสียชีวิตจากอุบัติเหตุทางถนน 3 ฐานข้อมูล. กรุงเทพฯ: กรมควบคุมโรค. สืบค้นจาก <https://dip.ddc.moph.go.th/new>
- กรมทางหลวง. (2564). โครงการศึกษาเพื่อกำหนดแนวทางการจัดทำทางข้ามถนนปลอดภัย 1 โครงการ.
- กรมทางหลวง. (2566). โครงการศึกษาจัดทำแผนพัฒนาศูนย์บริหารจราจร (HTOC) กรมทางหลวง ปี พ.ศ. 2567 - พ.ศ. 2570 1 โครงการ.
- กิริษชาติ สุขสุทธิ. (2559). การจำแนกข้อมูลไม่สมดุลโดยใช้การปรับปรุงข้อมูลร่วมกับการหาค่าพารามิเตอร์ที่เหมาะสมด้วยขั้นตอนวิธีเชิงพันธุกรรมที่มีการเริ่มต้นใหม่. นครราชสีมา: มหาวิทยาลัยเทคโนโลยีสุรนารี. สืบค้นจาก <http://sutir.sut.ac.th:8080/jspui/handle/123456789/7213>
- ชิติภัทร อินทวิมลเมธา, และพงษ์ศักดิ์ สุริยวานากุล. (2555). แบบจำลองวัดระดับความยากในการข้ามถนนบริเวณทางแยกโดยศึกษาจากการรับรู้ของผู้เดินเท้า. *Engineering Journal of Research and Development*, 23(2), 1–8.
- ณัฐภัทร ฤทธิ์เดช, และปริพัทธ์ ศรีสมบูรณ์. (2565). การสร้างตัวแบบวิเคราะห์พฤติกรรมและสาเหตุต่อการเป็นโรคติดสุราด้วยเทคนิคต้นไม้ตัดสินใจ. *วารสารศึกษาศาสตร์ปริทัศน์*, 37(3), 202–2011.
- นพมาศ อัครจันทโชติ และดิเรก พนิตสุภาภมล. (2561). การเปรียบเทียบวิธีการแก้ปัญหาข้อมูลไม่สมดุลสำหรับการจำแนกกลุ่มรายได้ของผู้ประกอบการร้านยาประเภท ข.ย. 1. ใน *การประชุมเสนอผลงานวิจัยระดับชาติ มหาวิทยาลัยสุโขทัยธรรมาธิราช ครั้งที่ 9* (หน้า 1578–1586). นนทบุรี: มหาวิทยาลัยสุโขทัยธรรมาธิราช.
- เบญจพร เอี่ยมประโคน. (2560). วิธีการเปรียบเทียบพื้นที่ใต้โค้ง ROC สำหรับข้อมูลชุดเดียวกัน: กรณีศึกษาแบบจำลองคะแนนเครดิต. *Chulalongkorn University Theses and Dissertations (Chula ETD)*. สืบค้นจาก <https://digital.car.chula.ac.th/chulaetd/> 2020
- ประยูรศรี บุตรแสนคม. (2555). การคัดเลือกตัวแปรพยากรณ์เข้าในสมการถดถอยพหุคูณ. *วารสารการวัดผลการศึกษามหาวิทยาลัยมหาสารคาม*, 17(1), 43–60.

รายการอ้างอิง (ต่อ)

- มูลนิธิไทยโรดส์. (2565). รายงานสถานการณ์อุบัติเหตุทางถนนของประเทศไทย พ.ศ. 2561–2564. สืบค้นจาก www.thairoads.org
- ยุทธ ไกยวรรณ. (2555). หลักการและการใช้การวิเคราะห์การถดถอยโลจิสติกสำหรับการวิจัย. *วารสารวิจัยมหาวิทยาลัยเทคโนโลยีราชมงคลศรีวิชัย*, 4(1), 1–12.
- วีระศักดิ์ จินารัตน. (2557). การวิเคราะห์ข้อมูลด้วยสถิติเชิงอนุมาน. *วารสารวิชาการมหาวิทยาลัยการจัดการและเทคโนโลยีอีสเทิร์น*, หน้า 80–85.
- วีรพงศ์ บุญเจริญปัญญา, และจำรัส พิทักษ์ศฤงคาร. (2567). ทักษะคติของคนเดินเท้าต่อการใช้งานทางข้ามที่มีช่องจราจรแตกต่างกัน: กรณีศึกษาบริเวณถนนแจ้งวัฒนะ ถนนลาดกระบัง และ ถนนกิ่งแก้ว. *Engineering and Technology Horizons*, 41(1), 410111–410111.
- วิชญ์วิสิฐ เกษรสิทธิ์, วิชิต หล่อจิรคุณกุล, และ จิราวัลย์ จิตตเวช. (2561). การแก้ปัญหาข้อมูลไม่สมดุลในการจำแนกประเภทผู้ป่วยโรคเบาหวาน. *วารสารวิจัยมหาวิทยาลัยขอนแก่น*, 18(3), 11–21.
- ศูนย์วิชาการเพื่อความปลอดภัยทางถนน. (2562). เปิดมุมมองสะท้อนความจริงของผู้เปราะบางบนถนน (VRUs). สืบค้นจาก https://www.roadsafetythai.org/download_bookdetail-edoc-539.html
- สุเมธี สนธิกุล. (2556). โครงการจัดทำชุดคู่มือเสริมสร้างความปลอดภัยของคนเดินเท้า. สืบค้นจาก ศูนย์วิชาการเพื่อความปลอดภัยทางถนน: https://www.roadsafetythai.org/edoc/doc_20210204144301.pdf
- อโนทัย ตรีวานิช. (2552). สถิติธุรกิจ (พิมพ์ครั้งที่ 4). ขอนแก่น: คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น.
- อัจฉรา แผ้วบาง, และ สายชล สีนสมบูรณ์ทอง. (2563). การปรับความไม่สมดุลของข้อมูลด้วยการจำแนก 5 วิธี. *Thai Journal of Science and Technology*, 9(4), 418–435.
- Afshari, A., Ayati, E., & Barakchi, M. (2021). Evaluating the effects of external factors on pedestrian violations at signalized intersections (a case study of Mashhad, Iran). *IATSS research*, 45(2), 234-240.
- Chawla, V. (2003). C4. 5 and imbalanced data sets: investigating the effect of sampling method, probabilistic estimate, and decision tree structure. In *Proceedings of the ICML* (Vol. 3, p. 66). Toronto, ON, Canada: CIBC.

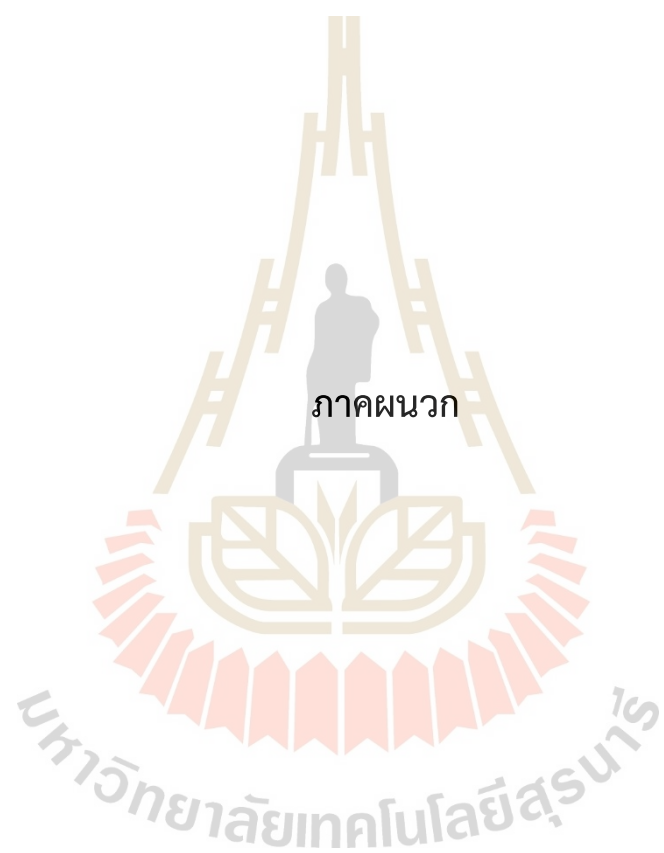
รายการอ้างอิง (ต่อ)

- Mienye, D., & Jere, N. (2024). A Survey of Decision Trees: Concepts, Algorithms, and Applications. IEEE Access.
- Modeling with R. (2019) Methods for dealing with imbalanced data., Retrieved Jun 15, 2024, from: <https://www.r-bloggers.com/2019/04/methods-for-dealing-with-imbalanced-data/>.
- Rafe, A., Singleton, P. A., Boyer, S., & Mekker, M. (2025). Pedestrian crossing behaviors at signalized intersections in Utah: Factors affecting spatial and temporal violations. *International Journal of Transportation Science and Technology*.
- Raoniar, R., Maqbool, S., Pathak, A., Chugh, M., & Maurya, A. (2022a). Hazard-based duration approach for understanding pedestrian crossing risk exposure at signalised intersection crosswalks—A case study of Kolkata, India. *Transportation research part F: traffic psychology and behaviour*, 85, 47-68.
- Raoniar, R., & Maurya, A. (2022b). Pedestrian red-light violation at signalised intersection crosswalks: Influence of social and non-social factors. *Safety science*, 147, 105583.
- Raoniar, R., & Maurya, A. (2022c). Pedestrian Crossing Behavior at Signalized Intersection Crosswalks: An Observational Study of Factors Influencing Pedestrian Walking Speed, Safety Margin, and Violation. *Journal of the Eastern Asia Society for Transportation Studies*, 14, 1456-1474.
- Šimundić, A. M. (2009). Measures of diagnostic accuracy: basic definitions. *ejifcc*, 19(4), 203.
- World Health Organization. (2018). GLOBAL STATUS REPORT ON ROAD SAFETY Retrieved from <https://www.who.int/publications/i/item/9789241565684>
- World Health Organization. (2023) Pedestrian safety: a road safety manual for decision-makers and practitioners., Retrieved Jun 30, 2024, from: <https://www.who.int/publications/i/item/9789240072497>
- Zhang, Y., & Deng, Y. (2019, July). Analysis of pedestrians' red-light violation behavior at signalized intersections. In 2019 5th International Conference on Transportation Information and Safety (ICTIS) (p.452-457). IEEE.

รายการอ้างอิง (ต่อ)

Zhu, D., Sze, N., & Bai, L. (2021). Roles of personal and environmental factors in the red-light running propensity of pedestrian: Case study at the urban crosswalks. Transportation research part F: traffic psychology and behavior, 76, 47-58.







ATRANS

Asian Transportation Research Society

CERTIFICATE OF PARTICIPATION

This is to certify

Ms. Chatcha Kummoon

Presented a paper titled

Factors affecting pedestrians' violations of traffic signals
at mid-block Smart Crosswalk in Thailand

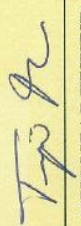
At the 17th ATRANS ANNUAL CONFERENCE; YOUNG RESEARCHER'S FORUM 2024

Given on the 30th of August 2024



DR. CHULA SUKMANOP

ATRANS - Chairperson



DR. TUENJAI FUKUDA

ATRANS Secretary-General,
Chair of ATRANS Conference
Organizing Committee

AYRF2024-027 (P)

ภาคผนวก ข
บทความวิชาการที่ได้รับการตีพิมพ์เผยแพร่ในระหว่างศึกษา

มหาวิทยาลัยเทคโนโลยีสุรนารี

รายชื่อบทความวิชาการที่ได้รับการตีพิมพ์เผยแพร่ในระหว่างศึกษา

KUMMOON, Ch., PUEBOOBPAPHAN, S., & PUEBOOBPAPHAN, R. (2024) FACTORS AFFECTING PEDESTRIANS' VIOLATIONS OF TRAFFIC SIGNALS AT MID-BLOCK SMART CROSSWALK IN THAILAND. Paper presented at 17th ATRANS Annual Conference: Young Researcher's Forum (AYRF) 2024, 30 August 2024, Bangkok, Thailand. Pp.146-154



FACTORS AFFECTING PEDESTRIANS' VIOLATIONS OF TRAFFIC SIGNALS AT MID-BLOCK SMART CROSSWALK IN THAILAND

Topic number: (3) Paper Identification number: AYRF2024-027
 Chatcha KUMMOON¹, Suthatip PUEBOOBPAPHAN², Rattaphol PUEBOOBPAPHAN^{3*}

School of Transportation Engineering, Institute of Engineering
 Suranaree University of Technology
 Telephone: 061-0281370¹, 044-224645², 044-224593^{3*}
 E-mail: M6402662@g.sut.ac.th¹, suthatip@sut.ac.th², rattaphol@sut.ac.th^{3*}

Abstract

This study examines the factors that influence instances of pedestrians disobeying traffic signals at mid-block crossings that are equipped with an intelligent crossing device, known as the Smart Crosswalk in Thailand. Once fully comprehended, it is postulated that recommendations or interventions to address pedestrian signal violations can be formulated to mitigate pedestrian accidents at crossings nationwide. The analysis examines several factors, including highway-related geometry, traffic volume, speed, and system operational parameters. The dependent variable is the binary event of whether the pedestrian signal is disobeyed or not during each cycle. Binary Logistic Regression is utilized to create models for this purpose. As the number of signal violation events is far lower than that of not being violated, the data in this study is considered imbalanced. Therefore, some scaling techniques to deal with imbalanced data are investigated and the downsampling technique is adopted for the final model development. This approach resulted in the highest Area Under the Curve (AUC) value of 71.70%. The specificity, sensitivity, and accuracy of the model were as follows: 67.56%, 68.29%, and 67.61%, respectively. The results show that while pedestrians are more likely to violate pedestrian signals as waiting time grows, the opposite is true when crossing length, traffic volume, and speed increase. Field observation reveals that violations are more likely at night when traffic volumes are low and signal waiting times are long.

Keywords: Pedestrians, Pedestrian safety, Signal violation, Smart crosswalk, Binary Logistic Regression, Imbalanced data

1. Introduction

Road accidents are one of the main causes of death in the world. According to a report by the World Health Organization on Traffic Injuries every year, road accidents result in more than 1.3 million deaths worldwide per year, and more than half of all deaths and injuries involve vulnerable road users such as pedestrians, cyclists, motorcyclists, and passengers^[1]. According to the WHO's 2018 Global Situation Report on Road Safety, Thailand has the highest death rate from road accidents in Southeast Asia and the 9th highest in the world, with 32.7 deaths per 100,000 people. Motorcycles had the highest death rate at 74%, followed by pedestrians at 8%^[2]. Out of the total number of road accidents, an average of six to eight pedestrian traffic accidents occur each day^[3]. In Thailand, road accidents kill two pedestrians on average each day. Although this is a small number compared to other types of road users, what is worrisome is the increasing trend in

the severity of pedestrian accidents. Analysis of the severity index based on highway accident statistics from the Department of Highways revealed that pedestrian accidents were approximately four times more severe than general accidents on highways. Consequently, once a pedestrian accident occurs, it leads to an exceptionally high fatality rate^[4]. The primary cause of pedestrian fatality or injury, as determined by global studies, is being hit by a vehicle while pedestrians are crossing the roadway^[5].

The Department of Highway (DOH) has implemented the Smart Crosswalk system to enhance pedestrian safety at mid-block crossings. This system incorporates pedestrian and vehicle detection devices to analyze the requirements for using the crossing and regulate the signal lights, thereby ensuring optimal safety and efficiency. Despite the installation of the intelligent crossing technology, pedestrian violations are still evident.

Previous research has primarily employed binary logistic regression analysis to investigate factors affecting pedestrian traffic signal violations. For example, research in Kolkata, India focused on various elements influencing pedestrians' red-light violations at three intersection crosswalks in different cities. The findings revealed that non-social factors such as crossing speed, waiting time for crossing signals, traffic volume, and signal cycle length significantly impact signal violation behavior. Social factors, including the number of people waiting at intersections and observing others successfully violating the red light, also positively influence this behavior^[6]. In Hong Kong, the impact of personal factors and environmental factors on the individual decision of a red-light running violation was examined at the signalized crossings in the urban area. The results indicated that pedestrian gender, age, number of lanes, presence of a companion, number of pedestrians around, presence of other violators in the same cycle, time to green, red time, traffic volume, and percentage of heavy vehicles all significantly impact the propensity of red-light running violation of pedestrians^[7]. Similarly, research in Mashhad, Iran, evaluated the impact of external factors on pedestrian violations at traffic signal intersections. The findings indicated that traffic volume, the number of violators, crosswalk length, red light duration (waiting time), and physical movement problems influence pedestrians' decisions to violate traffic signals^[8]. Prior research reviews showed that most studies, both domestically and internationally, took place at crosswalks with fixed-time signals. Currently, there is a lack of research to investigate and comprehend the factors that influence pedestrians' violation behavior at crosswalks with smart crosswalk systems.

Our preliminary analysis of pedestrian violation at smart crosswalks indicates that the data in this study is imbalanced, i.e. the number of events that pedestrian violations are observed during any signal cycle is far less than the events of no violation. If the model is developed without addressing unbalanced property of the data, the model's predicted outputs might be biased. Noppamas^[9] conducted research on comparative analysis to evaluate the effectiveness of Random OverSampling Examples (ROSE) and Synthetic Minority Oversampling Technique (SMOTE) in addressing data imbalances. The results showed that SMOTE outperformed both the initial data set and ROSE. Kesornsit^[10] compares the effectiveness of a variety

of methods for addressing the issue of imbalanced data in diabetic patients. Addressing the data imbalance can improve classification efficiency, making it more efficient than the initial data and enabling the model to classify both large and small groups of data equally or nearly equally. Previous research conducted in Thailand predominantly employed imbalance adjustments with medical data, financial data, demographic data, educational data, etc. In the context of pedestrian violation data at crosswalks with traffic signals or with smart crosswalk systems, no prior research has adjusted for imbalanced data.

Based on the aforementioned research gaps, this study investigates the factors influencing pedestrians' violations of traffic signals at mid-block smart crosswalks in Thailand by taking imbalanced property of the data into consideration. The analysis of the smart crosswalk system relies on secondary data, including traffic volume, speed, operational parameters, and road characteristics. However, the secondary data from the smart crosswalk system lacks information about the individual characteristics of pedestrians, such as gender, age, and occupation. Therefore, this study cannot consider factors related to individual characteristics. Additionally, the pedestrian violation data recorded in the system is the total number of pedestrians that violated signal for each cycle. Therefore, the dependent variable employed in this study is binary, indicating whether pedestrian signal is violated for each signal cycle. Binary Logistic Regression is therefore chosen to perform the analysis. The primary objective is to generate recommendations or measures to address signal violations and reduce pedestrian accidents.

2. Data collection

Currently, there are 11 Smart crosswalks in Thailand (Fig. 1). For this study, we selected six crossings: 1) Highway No. 3242, km 18+110, Bangkok; 2) Highway No. 306, km 4+970, Nonthaburi Province; 3) Highway No. 3316, km 2+225, Nakhon Pathom Province; 4) Highway No. 3215, km 11+020, Nonthaburi Province; 5) Highway No. 3242, km 2+220, Samut Sakhon Province; and 6) Highway No. 314, km 8+410, Chachoengsao Province. We did not consolidate the data from the other five crossings for analysis, as some of the systems were temporarily down and others had only recently been implemented with a smart crosswalk.

This study utilized data on physical road characteristics, such as the median, median width, lanes, crosswalk width, crosswalk length, yellow frame, and yellow frame width. Secondary data from smart crosswalk systems, including speed and traffic volume, as well as specifications related to traffic signal systems at crossings, were also used. The study exclusively focused on factors anticipated to have a direct correlation with pedestrians violating traffic signals, considering the system's definition of violating pedestrians, the correlation coefficient, and previous literature. Variables not listed above, such as characteristics of individual pedestrian and pedestrian behavior, were not considered as they are not available from the system database.

We collected data from the smart crosswalk system from January 2022 to August 2022.

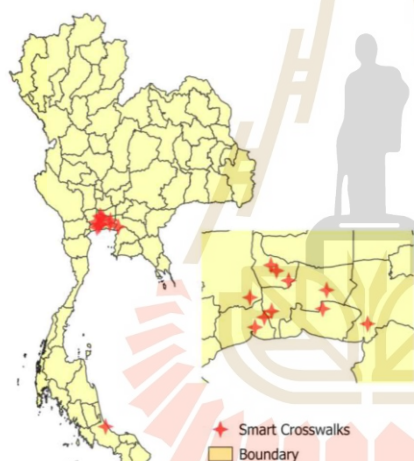


Fig. 1 Smart crosswalks in Thailand

2.1 Physical Characteristics of Smart Crosswalks

We conducted a survey of the crossing's physical characteristics, encompassing three main aspects: 1) the road's physical attributes, including the presence or absence and width of the median and the number of traffic lanes; 2) the crossing's physical features, such as its length and width; and 3) the safety equipment at the crossing, including the presence or absence and width of the yellow netting frames (Table 1).

2.2 Smart Crosswalk System database

The Smart Crosswalk System records a number of databases, including traffic, speed, pedestrian, and transaction information. For this study, we focus on the traffic and transaction databases. The traffic database provides data on traffic volume and average speed, organized into 5-minute intervals. The transaction database contains detailed information on pedestrian signals, including waiting time for crossing signals, the duration of the walk interval, the length of the flashing don't-walk signal, and any extensions of this signal. Additionally, it records the total number of pedestrians crossing and the number of pedestrians who violate the signal during each signal cycle.

The dependent variable in this study is the event of pedestrian violations occurring during signal cycles, categorized based on whether or not pedestrian violation is observed during that cycle. To mitigate issues related to multicollinearity and based on literature findings, only selected factors are considered as independent variables. Table 2 summarizes the variables included in this study.

Table 1 Physical Characteristics of Smart Crosswalks

Smart crosswalk	Lane/ direction	Crosswalk length (m)	Crosswalk width (m)	Yellow frames width (m)	Median width (m)
1) Highway No. 3242 km 18+110	3	25	8	0	1.3
2) Highway No. 306 km 4+970	2	7.8	8	0	0.8
3) Highway No. 3316 km 2+225	2	14	8	7	0
4) Highway No. 3215 km 11+020	3	23	8	7	4.3
5) Highway No. 3242 km 2+220	3	30	8	12	10
6) Highway No. 314 km 8+410	3	35	16	17.7	5

Source: Bureau of Highway Safety (2022) As Built Drawing for Implementing the Smart Crosswalk, Department of Highways, Thailand.

Table 2 Description of variables in this study

Dependent Variable	Explanation
Event of pedestrian violations that occur during signal cycles	Dichotomous Variable 0 = none of the pedestrians violated a traffic signal (Not Violated). 1 = at least one pedestrian violated a traffic signal (Violated).
Independent Variable	Explanation
Waiting time	The waiting time is calculated from the moment pedestrians press the button to request pedestrian signals until they receive the signal to cross (unit: seconds).
Crosswalk length	(unit : meters)
Vehicle speed	We calculate the vehicle average speed every 5 minutes, which corresponds to the signal cycle (unit : km/hr).
Traffic volume	We calculate the traffic volume every 5 minutes, which corresponds to the signal cycle (unit : vehicle/hr/lane).

3. Method

RStudio was employed to develop models using Binary Logistic Regression, and to process, analyze, and forecast data. We analyzed a total of 116,485 signal cycles, dividing them into two groups: 93,188 cycles (80%) for training data and 23,297 cycles (20%) for test data.

3.1 Binary Logistic Regression Analysis

The binary logistic regression is used for regression analysis where the dependent variable is dichotomous, taking on values of 0 or 1. This model investigates which independent variables lead to pedestrian violations during signal cycles. The analysis uses the following function to handle situations with multiple predictor variables:

$$P_y = \frac{e^{b_0 + b_1x_1 + \dots + b_px_p}}{1 + e^{b_0 + b_1x_1 + \dots + b_px_p}} \quad (1)$$

Where: P_y = the likelihood of the event of interest.
 and Q_y = the likelihood of the event of interest not occurring, as follows:

$$Q_y = 1 - P_y \quad (2)$$

Given the non-linear relationship between predictor and dependent variables in logistic regression, we adjust the relationship to be linear in the form of odds ratios. The odds ratio (OR), which is the expected change in probabilities for a unit increase in the predictor, is expressed as $\text{Exp}(B)^{[11]}$.

$$\text{Odds Ratio (OR)} = \frac{P_y}{Q_y} \quad (3)$$

3.2 Measures of accuracy

The performance of the model can be verified using the following statistical values: True Positive (TP), False Negative (FN), False Positive (FP), and True Negative (TN), as shown in Table 3. These values are used to create metrics to evaluate the model performance using equations (4) – (6)^[12]. Additionally, the Area Under the Curve (AUC) can be used as another model's performance indicator. A model with an AUC value closer to 1 indicates better predictive performance^[13].

Table 3 Confusion Matrix

	Predicted Positive	Predicted Negative
Actual positive	True Positive (TP)	False Negative (FN)
Actual negative	False Positive (FP)	True Negative (TN)

Source: ^[12]

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FN + FP} \quad (4)$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (5)$$

$$\text{Specificity} = \frac{TN}{FP + TN} \quad (6)$$

3.3 Imbalanced data

Imbalanced data poses a problem when the proportions of data are unequal, leading the training model to be dominated by the majority class. Consequently, the model tends to predict the majority class more accurately than the minority class, resulting in a high specificity rate but a low sensitivity rate in binary classification. Data scaling is one technique to mitigate this issue. Common scaling methods used in practice include the following^[14]:

Upsampling: This method increases the size of the minority class by sampling with replacement, making the majority and minority classes equivalent in size.

Downsampling: This method reduces the size of the majority class by sampling with replacement, making the majority and minority classes equivalent in size.

Hybrid methods: Techniques such as ROSE and SMOTE downsample the majority class while adding synthesized information to the minority class from the original dataset.

These scaling methods should be applied exclusively to the training data and not to the test data.

4. Results

This study began with an analysis using descriptive statistics to examine the behavior of drivers and pedestrians, as well as the functioning of the smart crosswalk system. The results are presented in Table 4.

Table 4 indicates that the percentage of violated signal cycle was only 6.86 percent, which is relatively small, and the dataset could be defined as imbalanced. Despite the low occurrence, such violations pose a significant risk of causing serious accidents for both drivers and pedestrians. The average speed was found to be 42.57 km/hr, with the 85th percentile speed at 59.14 km/hr, which is higher than the recommended safe speed for pedestrians (30–40 km/hr)^[15].

The results of the binary logistic regression analysis based on the original train dataset without addressing imbalanced feature are presented in Table 5.

Table 4 Descriptive statistics of the smart crosswalk data

Dependent variable	Frequency					Percentage	
1. Events of pedestrian violations that occur during signal cycles (n = 116,485 signal cycles)							
None of the pedestrians violated a signal cycle (Not Violated).	108,490					93.14 %	
At least one pedestrian violated a signal cycle (Violated).	7,995					6.86 %	
Independent variable	Min	P15	Mean	Median	P85	Max	SD
2. Waiting time (second)	8	10	22.75	20	35	65	11.54
3. Crosswalk length (meter)	13.8	13.8	22.94	23.2	35	35	7.74
4. Vehicle speed (km/hr)	5.25	24.74	42.57	43.29	59.14	115.13	16.45
5. Traffic volume (veh/hr/lane)	2	136	247	234	350	1,650	122

Table 5 Results of binary logistic regression based on original train dataset

Independent variable	Coefficient (B)	Standard Error	Exp (B)	P-value
Constant	- 1.042 ***	0.054	0.353	0.000
Waiting time (second)	0.041 ***	0.001	1.042	0.000
Crosswalk length (meter)	- 0.075 ***	0.002	0.927	0.000
Vehicle speed (km/hr)	- 0.018 ***	0.001	0.983	0.000
Traffic volume (veh/hr/lane)	- 0.001 ***	0.000	0.999	0.000

Note*** Statistical significance level at 0.001, $R^2 = 0.0859$, Specificity = 100%, Sensitivity = 0%, Accuracy = 93.14%, AUC = 71.46%

Table 5 reveals that all independent variables have a statistically significant influence on pedestrians' violations of traffic signals at the 0.001 level. However, when using the test dataset to make predictions, the model failed to accurately predict pedestrian violations of traffic signals (as sensitivity is 0.0%). This aligns with the findings of previous studies^[14] on imbalanced datasets. To address this issue and improve prediction accuracy, several methods were employed, including Upsampling, Downsampling, ROSE, and SMOTE. These methods were used to adjust the number of signal cycles with pedestrian violations to be similar to the number of signal cycles without pedestrian violations.

After applying these data scaling techniques, new models were developed. The results indicated that each independent variable continued to have a statistically significant effect on pedestrian violations of traffic signals at the 0.001 level. Furthermore, the relationship between the independent and dependent variables remained

consistent with the initial analysis. The prediction performance based on the binary logistic regression model from the new data scaling experiments are presented in Table 6.

Table 6 presents performance of the prediction results for pedestrian violations of traffic signals, including specificity, sensitivity, overall accuracy, and area under the curve (AUC). The application of data scaling principles across four methods resulted in improved sensitivity, which is crucial in this context, compared to the model's performance without data scaling. Based on these results, we decided to use the downsampling method to adjust the proportion of signal cycles. This method achieved the highest AUC value of 71.70%. The corresponding specificity, sensitivity, and accuracy were 67.56%, 68.29%, and 67.61%, respectively. Downsampling is particularly suitable for large datasets^[16]. The results of binary logistic regression with the downsampling method, are presented in Table 7 and illustrated in Fig. 2.

Table 6 Prediction performance of scaling techniques based on test dataset

Scaling techniques	Specificity	Sensitivity	Accuracy	AUC
Upsampling	67.65%	68.48%	67.71%	71.64%
Downsampling	67.56%	68.29%	67.61%	71.70%
ROSE	67.60%	68.36%	67.65%	71.65%
SMOTE	75.06%	59.29%	73.98%	71.62%

Table 7 Results of binary logistic regression based on downsampling method

Independent variable	Coefficient (B)	Standard Error	Exp (B)	P-value
Constant	1.307 ***	0.076	3.695	0.000
Waiting time (second)	0.050 ***	0.002	1.052	0.000
Crosswalk length (meter)	- 0.079 ***	0.003	0.924	0.000
Vehicle speed (km/hr)	- 0.014 ***	0.001	0.986	0.000
Traffic volume (veh/hr/lane)	- 0.001 ***	0.000	0.999	0.000

Note*** Statistical significance level at 0.001, $R^2 = 0.1232$, Specificity = 67.56%, Sensitivity = 68.29%, Accuracy = 67.61%, AUC = 71.70%

Table 7 presents a model of pedestrian traffic signal violations developed using the downsampling method. This model indicates that each independent variable has a statistically significant effect on pedestrian violations at the 0.001 level. The model accounts for 12.32% of the variance in pedestrian violations. The relationships between independent and dependent variables in this

model are similar to those observed without data scaling, but now with higher predictive capability. Performance of the original model and the downsampling model can be compared using confusion matrix as shown in Table 8. The downsampling model is superior to the original model in predicting true positive (TP) but at the cost of increasing false positive (FP). Explanation of the

results for each variable based on the downsampling method are as follows:

Waiting time ($B = 0.050$, $\text{Exp}(B) = 1.052$): For each additional second of waiting time, while holding other factors constant, the likelihood of a pedestrian signal cycle being violated increases by 5.2%.

Crosswalk length ($B = -0.079$, $\text{Exp}(B) = 0.924$): For each additional meter of crosswalk length, while holding other factors constant, the likelihood of a pedestrian signal cycle being violated decreases by 7.6%.

Average Vehicle Speed ($B = -0.014$, $\text{Exp}(B) = 0.986$): For each additional kilometer per hour increase in average vehicle speed; while holding other factors constant, the likelihood of a pedestrian signal cycle being violated decreases by 1.4%.

Traffic Volume ($B = -0.001$, $\text{Exp}(B) = 0.999$): For each additional vehicle per hour per lane, while holding other factors constant, the likelihood of a pedestrian signal cycle being violated decreases by 0.1%.

Table 8 Performance of the original model and the downsampling model with test dataset

	Original model		Downsampling model	
	Predicted Positive	Predicted Negative	Predicted Positive	Predicted Negative
Actual positive (1,599 cycles)	0 (TP)	1,599 (FN)	1,092 (TP)	507 (FN)
Actual negative (21,698 cycles)	0 (FP)	21,698 (TN)	7,040 (FP)	14,658 (TN)

Note*** Positive Class = 1 (at least one pedestrian violated a traffic signal (Violated).)

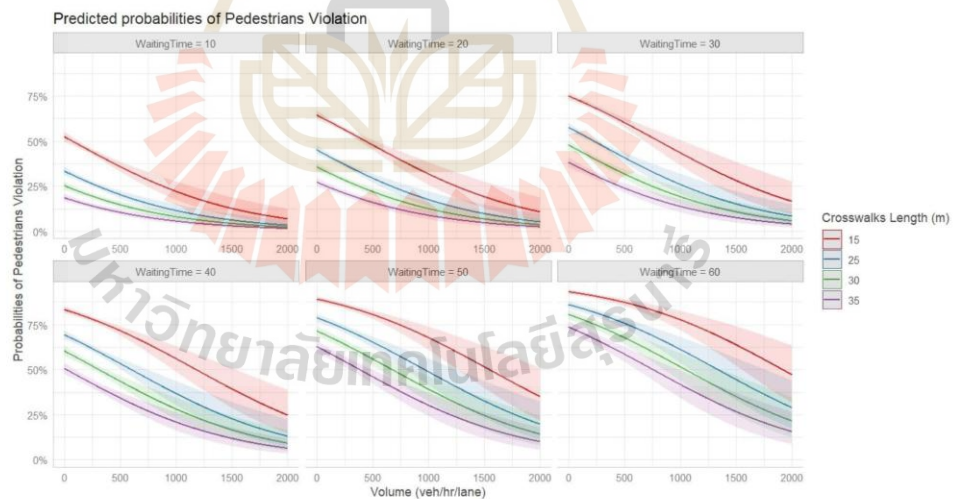


Fig. 2 Predicted probabilities of pedestrian violation

Fig. 2 presents six subgraphs, each representing different waiting times for the signal and illustrating the relationship between crosswalk

length, traffic volume, and pedestrian traffic signal violations. The crosswalk lengths are categorized into 15, 25, 30, and 35 meters. It shows that

violations are more likely when shorter crosswalk length, longer waiting time, and lower traffic flow. This finding aligns with field observation, which indicated that signal violations are more common at night when traffic flow is low, and pedestrians experience prolonged waiting times after pressing the button. To reduce violation rates, it is recommended to adjust the waiting times to better align with current traffic conditions in the smart crosswalk system.

5. Conclusion

This study investigates the factors influencing pedestrian violations of traffic signals at Smart Crosswalks in Thailand, using data collected from January 2022 to August 2022. The analysis employed binary logistic regression to assess the impact of four independent variables on pedestrian signal violations. Results revealed that each independent variable had a statistically significant effect on violations at the 0.001 level. However, when applying the model to the test set, it failed to effectively predict pedestrian violations. To address this limitation, the number of signal cycles without pedestrian violations was adjusted to match those with violations, enhancing the model's predictive capacity. The downsampling method, which achieved the highest Area Under the Curve (AUC) value of 71.70%, demonstrated improved performance compared to other techniques, with specificity, sensitivity, and accuracy values of 67.56%, 68.29%, and 67.61%, respectively. Despite these improvements, the model's overall predictive performance remained suboptimal. The analysis revealed a positive correlation between pedestrian violations and waiting time, and negative correlations with crossing length, traffic speed, and traffic volume, which aligns with field observations. To reduce violation rates, it is suggested that system waiting times be adjusted to better reflect traffic conditions. Despite the improved performance obtained from the data scaling methods, the specificity, sensitivity, accuracy, and AUC values remained relatively low^[17]. Although binary logistic regression effectively identifies relationships between variables, it may be inadequate for precise forecasting. Future research should explore alternative machine learning techniques, such as Decision Trees, Random Forests, or Gradient Boosting, together with data scaling methods to enhance predictive performance and validate these findings.

References

- [1] World Health Organization (2022) World Health Statistics 2022, Retrieved Jun 15, 2024, from: <https://www.who.int/news/item/20-05-2022-world-health-statistics-2022>.
- [2] World Health Organization (2022) GLOBAL STATUS REPORT ON ROAD SAFETY 2018., Retrieved Jun 15, 2024, from: <https://www.who.int/publications/i/item/9789241565684>.
- [3] National Statistical Office (2021) Statistics on road traffic accident cases Classified by type of road user, damage, suspects, by region 2012 - 2021., Retrieved Jun 16, 2024, from: <https://www.nso.go.th>
- [4] Thai Roads Foundation (2019) Key factor on road safety situations in thailand 2016-2017., Retrieved Jun 16, 2024, from: <http://trso.thairoads.org/statistic/national>
- [5] Thai Health Promotion Foundation (2019) Opening the perspective to reflect the reality of vulnerable people on the road (VRUs) ., Retrieved Jun 16, 2024, from: <https://data.go.th/th/dataset/rtdidi>
- [6] Raoniar, R., & Maurya, A. K. (2022) Pedestrian red-light violation at signalised intersection crosswalks: Influence of social and non-social factors., *Safety science*, 147, 105583.
- [7] Zhu, D., Sze, N. N., & Bai, L. (2021) Roles of personal and environmental factors in the red light running propensity of pedestrian: Case study at the urban crosswalks., *Transportation research part F: traffic psychology and behaviour*, 76, 47-58.
- [8] Afshari, A., Ayati, E., & Barakchi, M. (2021) Evaluating the effects of external factors on pedestrian violations at signalized intersections (a case study of Mashhad, Iran)., *IATSS research*, 45(2), 234-240.
- [9] Noppamas, D. (2018) Comparison of Imbalanced Data Problem Solving for Income Classification of Type I Pharmacies Entrepreneur., In *The 9 th STOU National Research Conference* (pp. 1578-1586).
- [10] Kesornsit, W., Lorchirachoonkul, V., & Jitthavech, J. (2018) Imbalanced data problem solving in classification of diabetes patients., *Khon Kaen University Research Journal*, 18(3), 11-21.
- [11] Kaiyawan, Y. (2012). Principle and using logistic regression analysis for research. *RMUTSV Research Journal*, 4(1), 1-12.
- [12] lamprakhon, B. (2017) A proposed method to compare areas under the ROC curves for a single

- dataset: A case of credit scoring model., Chulalongkorn University Theses and Dissertations (Chula ETD)., from: <https://digital.car.chula.ac.th/chulaetd/2020>
- [13] Šimundić, A. M. (2009). Measures of diagnostic accuracy: basic definitions. *ejifcc*, 19(4), 203.
- [14] Modeling with R. (2019) Methods for dealing with imbalanced data., Retrieved Jun 15, 2024, from: <https://www.r-bloggers.com/2019/04/methods-for-dealing-with-imbalanced-data/>.
- [15] World Health Organization. (2023) Pedestrian safety: a road safety manual for decision-makers and practitioners., Retrieved Jun 30, 2024, from: <https://www.who.int/publications/i/item/9789240072497>
- [16] Minh, H. (2018) How to Handle Imbalanced Data in Classification Problems., Retrieved Jun 15, 2024, from: <https://medium.com/@nminh.hoang1023/handling-imbalanced-data-in-classification-problems-7de598c1059f>.
- [17] Metz, C. E. (1978). Basic principles of ROC analysis., In *Seminars in nuclear medicine* (Vol. 8, No. 4, pp. 283-298). WB Saunders.



ประวัติผู้เขียน

นางสาวณัชชา คำมูล เกิดวันที่ 1 เมษายน พ.ศ. 2541 การศึกษาระดับประถมศึกษาจากโรงเรียนเหรียญทองวิทยา และมัธยมศึกษาจากโรงเรียนนครบุรี ในปีการศึกษา 2558 และสำเร็จการศึกษาระดับปริญญาตรีจากสำนักวิศวกรรมศาสตร์ สาขาวิชาวิศวกรรมขนส่ง มหาวิทยาลัยเทคโนโลยีสุรนารี ในปีการศึกษา 2562 จากนั้นเข้าทำงานในบริษัท S2R CONSULTING CO.,LTD. ในตำแหน่งวิศวกรที่ปรึกษา เป็นระยะเวลา 1 ปี 3 เดือน จากนั้นศึกษาต่อในระดับปริญญาโท สำนักวิศวกรรมศาสตร์ สาขาวิชาวิศวกรรมขนส่ง หลักสูตรวิศวกรรมโยธาขนส่ง และทรัพยากรธรณี (พ.ศ. 2563) ในปี พ.ศ. 2564 ในขณะที่ศึกษาอยู่ได้ทำหน้าที่ผู้ช่วยสอนในสาขาวิชาวิศวกรรมขนส่ง มหาวิทยาลัยเทคโนโลยีสุรนารี และเข้าร่วมงานวิจัยต่าง ๆ

