บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

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

2.1 ทฤษฎีและวรรณกรรมที่เกี่ยวข้อง

2.1.1 คนเดินเท้า

คนเดินเท้า (Pedestrian) คือ บุคคลที่สัญจรเดิน ข้าม หรือวิ่ง บริเวณไหล่ทางบนถนน 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 เป็นสิ่งที่สำคัญและมีประโยชน์ต่อคนเดินเท้าเป็นอย่างมาก ทางข้ามช่วยป้องกันและให้สิทธิในการข้ามแก่คนเดินข้าม ทำให้คนเดินเท้าสามารถข้ามถนนได้อย่างสะดวกและปลอดภัย ทางข้ามในปัจจุบันมีหลายประเภท แตกต่างกันไปตามความเหมาะสมในแต่ละพื้นที่ที่ติดตั้ง ซึ่งขึ้นอยู่กับการใช้งานและองค์ประกอบต่าง ๆ บริเวณทางข้าม เช่น ความกว้างของทางข้าม ความยาวของทางข้าม ลักษณะของทางข้าม ผิวทาง ระยะการมองเห็นทางข้ามถนน การเข้าถึงทางข้าม ลักษณะกายภาพของเกาะกลาง เป็นต้น ทางข้ามโดยส่วนใหญ่จะติดตั้งในพื้นที่เขตเมืองและสถานที่สำคัญ 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

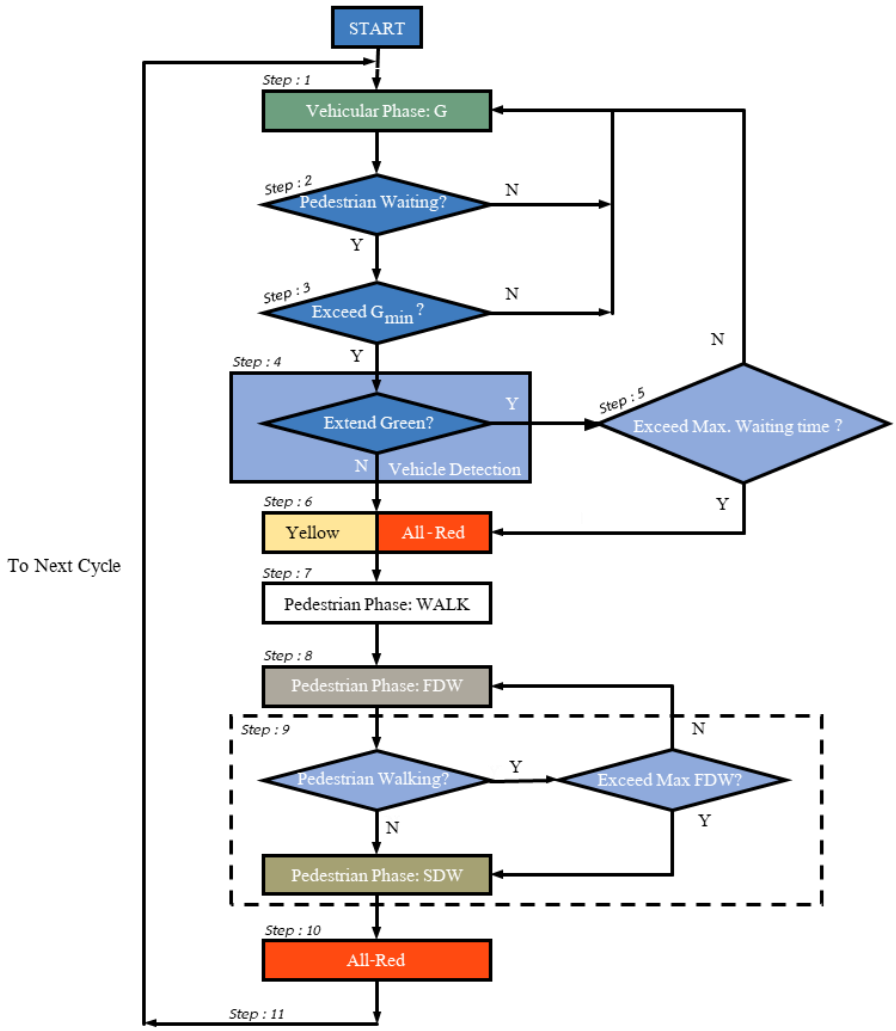
ทางแยก (Intersection) โดยส่วนใหญ่แบ่งเป็นสองรูปแบบหลัก ได้แก่ ทางข้ามระดับผิวทางแบบไม่มีไฟสัญญาณจราจร (Unsignalized Pedestrian Crossing) และทางข้ามระดับผิวทางแบบมีไฟสัญญาณจราจร (Signalized Pedestrian Crossing)

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 เพื่อไม่ให้ส่งผลกระทบต่อการเคลื่อนตัวของการจราจร (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 คลาสข้อมูลที่มีจำนวนของข้อมูลมากกว่าเรียกว่า คลาสส่วนมาก (Majority Class) และคลาสข้อมูลที่มีจำนวนของข้อมูลน้อยกว่า เรียกว่าคลาสส่วนน้อย (Minority Class) ซึ่งเมื่อชุดข้อมูลที่นำมาวิเคราะห์มีความไม่สมดุลของข้อมูลเกิดขึ้นในกระบวนของการสร้างแบบจำลองการเรียนรู้ (Training Model) ผลการวิเคราะห์ที่ได้จะมีความโน้มเอียงไปในทิศทางของคลาสส่วนมาก ซึ่งจะส่งผลให้การจำแนกของคลาสส่วนน้อยมีประสิทธิภาพลดน้อยลง (กีระชาติ สุขสุทธิ์, 2559)

2.1.4.1 ระดับของความไม่สมดุล

ระดับของความไม่สมดุล คือ อัตราส่วนของจำนวนข้อมูลในคลาสส่วนมากเทียบกับจำนวนข้อมูลในคลาสส่วนน้อย ซึ่งเป็นการจำแนกว่าชุดข้อมูลใดมีความสมดุลหรือไม่สมดุล ยิ่งชุดข้อมูลมีค่าระดับความไม่สมดุลมากกว่า 1 มากเท่าไร นั่นหมายความว่าชุดข้อมูลมีความไม่สมดุลสูงเท่านั้น หากข้อมูลมีระดับของความไม่สมดุลเท่ากับ 1 หมายความว่าในแต่ละคลาสมีจำนวนของข้อมูลที่เท่ากัน และหากข้อมูลมีระดับของความไม่สมดุลตต่ำกว่า 1 หมายความว่าจำนวนข้อมูลของคลาสส่วนน้อยมากกว่าจำนวนข้อมูลในคลาสส่วนมาก (กีระชาติ สุขสุทธิ์, 2559) การคำนวณระดับความไม่สมดุลสามารถคำนวณได้จากสมการที่ 2-1

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

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) (กีระชาติ สุขสุทธิ์, 2559) โดยเทคนิคการสุ่มเลือกข้อมูลโดยทั่วไปแบ่งออกเป็น 3 กลุ่ม (Modeling with R, 2019) ได้แก่

1) วิธีการสุ่มเกิน (Over Sampling) เป็นเทคนิคในการเพิ่มจำนวนข้อมูลของคลาสส่วนน้อย โดยการสุ่มเลือกจากข้อมูลเดิมหรือสร้างข้อมูลขึ้นมาใหม่จากตัวอย่างเดิมให้มีจำนวนใกล้เคียงหรือเท่ากันกับจำนวนข้อมูลในคลาสส่วนมาก โดยการสุ่มเลือกข้อมูลจากข้อมูลเดิม หรือสร้างข้อมูลขึ้นมาใหม่จากตัวอย่างของข้อมูลเดิม

2) วิธีการสุ่มลด (Under Sampling) เป็นเทคนิคในการลดจำนวนข้อมูลของคลาสส่วนมากลง ให้มีจำนวนใกล้เคียงหรือเท่ากันกับจำนวนข้อมูลในคลาสส่วนน้อย

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

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

จุดข้อมูลที่กำลังพิจารณาและจุดข้อมูลเพื่อนบ้านที่ใกล้ที่สุด ซึ่งสามารถกำหนดได้ว่าจะสุ่มสร้างข้อมูลขึ้นมาใหม่มากน้อยเพียงใด

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

2.1.4.3 การประเมินประสิทธิภาพสำหรับข้อมูลที่ไม่สมดุล

ในกรณีที่ข้อมูลเกิดปัญหาข้อมูลไม่สมดุล (Imbalanced Data) จำเป็นต้องมีการประเมินประสิทธิภาพในการจำแนกด้วยวิธีอื่น ๆ เพิ่มเติม นอกเหนือจากการประเมินประสิทธิภาพด้วยค่าความแม่นยำในการจำแนก (Accuracy) เนื่องจากจำนวนข้อมูลของทั้งสองคลาสมีความแตกต่างกันสูง หากผลของการพยากรณ์ข้อมูลทดสอบเป็นคลาสส่วนมากทั้งหมด ผลของค่าความแม่นยำในการจำแนกก็จะสูงด้วยเช่นกัน (กีระชาติ สุขสุทธิ์, 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; กีระชาติ สุขสุทธิ์, 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

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

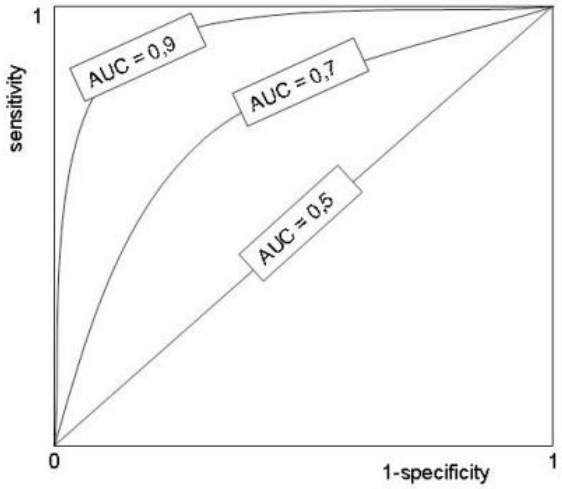
Sensitivity or True positive rate (TPR) คือ อัตราส่วนของจำนวนที่ผลจากการพยากรณ์เป็นบวก (คลาสส่วนน้อย) และเหตุการณ์ที่เกิดขึ้นจริงมีผลเป็นบวก (คลาสส่วนน้อย) ต่อจำนวนเหตุการณ์ที่เกิดขึ้นจริงมีผลเป็นบวก (คลาสส่วนน้อย) ทั้งหมด สมการในการคำนวณแสดงดังสมการที่ 2-3

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

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

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

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 มากที่สุดถือว่าเป็นตัวแบบที่ดีที่สุด (เบญจพร เอี่ยมประโคน, 2560) ยิ่งเส้นโค้งอยู่ใกล้กับมุมซ้ายบนมากเท่าไร หมายความว่าตัวแบบจะยิ่งสามารถจำแนกกลุ่มได้ดีมากยิ่งขึ้นเท่านั้น 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

H0 : ตัวแปรอิสระสองตัวไม่มีความสัมพันธ์กัน

H1 : 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 ≤ 0.40 หมายความว่า ตัวแปรมีความสัมพันธ์กันเองค่อนข้างน้อย
3. หาก 0.41 < r ≤ 0.60 หมายความว่า ตัวแปรมีความสัมพันธ์กันเองปานกลาง
4. หาก 0.61 < r ≤ 0.80 หมายความว่า ตัวแปรมีความสัมพันธ์กันเองค่อนข้างสูง
5. หาก r > 0.80 หมายความว่า ตัวแปรมีความสัมพันธ์กันเองสูง

2.1.6 การถดถอยโลจิสติค (Logistic Regression Analysis)

การถดถอยโลจิสติค มีวัตถุประสงค์เพื่อศึกษาว่าตัวแปรอิสระใดบ้างที่ส่งผลต่อตัวแปรตามอย่างมีนัยสำคัญทางสถิติซึ่งเป็นตัวแปรแบบทวิหรือตัวแปรแบบพหุกลุ่ม (ยุทธ ไกยวรรณ, 2555)

2.1.6.1 การถดถอยโลจิสติคแบบทวิ (Binary logistic regression)

การถดถอยโลจิสติคแบบทวิ เป็นการวิเคราะห์การถดถอยที่ตัวแปรตาม (y) มี 2 ค่า คือ เกิดเหตุการณ์ที่สนใจ (y = 1) หรือไม่เกิดเหตุการณ์ที่สนใจ (y = 0) มีความสัมพันธ์กับตัวแปรอิสระ (x) ไม่อยู่ในรูปเชิงเส้น ซึ่งในกรณีที่ตัวแปรอิสระมีมากกว่า 1 ตัว ในการวิเคราะห์จะได้ฟังก์ชั่นดังนี้

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

โดยที่ P_y คือ ความน่าจะเป็นของการเกิดเหตุการณ์ที่สนใจ

และ Q_y คือ ความน่าจะเป็นของการไม่เกิดเหตุการณ์ที่สนใจ ดังนี้

$$\mathrm{Q}_y = 1 - \mathrm{P}_y \qquad \text{สมการที่ 2-6}$$

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

อัตราส่วนระหว่างโอกาส (Odds Ratio) เป็นค่าที่สะท้อนถึงการเปลี่ยนแปลงที่คาดหวังในความน่าจะเป็นของผลลัพธ์ เมื่อค่าของตัวแปรทำนายเพิ่มขึ้นหนึ่งหน่วย โดยค่า Odds Ratio นี้แสดงในรูปของค่าที่ได้จากการแปลงเอ็กซ์โปเนนเชียลของสัมประสิทธิ์ถดถอย (Exp(B))

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

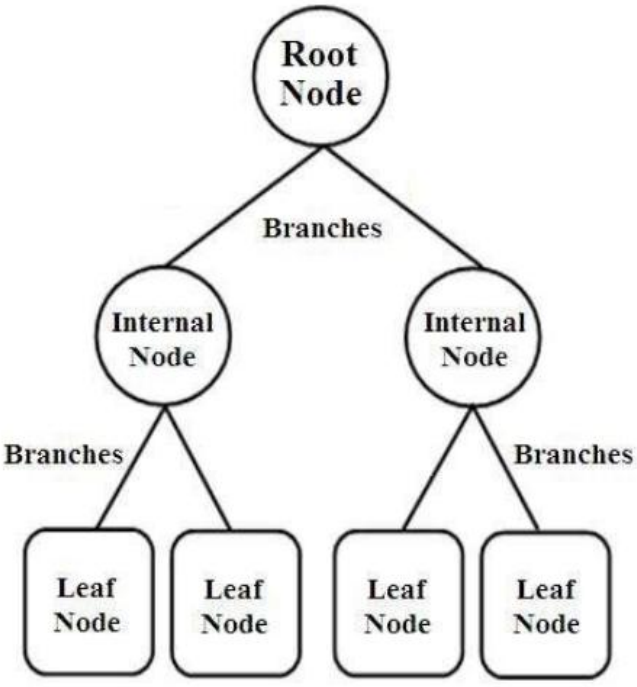
2) การเลือกตัวแปรโดยวิธีลดตัวแปร (Backward Elimination) เป็นวิธีการที่เลือกเฉพาะตัวแปรอิสระที่ดีที่สุดที่สามารถพยากรณ์ตัวแปรตามได้เท่านั้น แต่มีวิธีที่ตรงข้ามกับวิธีเพิ่มตัวแปร (Forward Selection) โดยในขั้นตอนแรกจะนำตัวแปรอิสระทุกตัวเข้ามาในสมการ และทำการพิจารณาตัวแปรอิสระที่มีค่าความสัมพันธ์กับตัวแปรตามที่ต่ำที่สุดออกจากสมการ และทดสอบดำเนินการทดสอบ ค่า R2 หากพบว่ามีค่าลดลงอย่างไม่มีนัยสำคัญทางสถิติ แสดงว่าตัวแปรดังกล่าวไม่ได้ทำให้การพยากรณ์ตัวแปรตามเพิ่มขึ้นสามารถขจัดออกจากสมการได้ และใช้วิธีพิจารณาเช่นเดียวกันขจัดตัวแปรอิสระที่มีความสำคัญน้อยรองลงมาออกจากสมการ การขจัดตัวแปรอิสระจะสิ้นสุดก็ต่อเมื่อพบว่า มีผลทำให้ค่า R2 ลดลงอย่างมีนัยสำคัญทางสถิติ แสดงว่าตัวแปรดังกล่าวมีความสำคัญต่อการพยากรณ์ตัวแปรตาม ให้คงตัวแปรอิสระดังกล่าวไว้ในสมการหากขจัดตัวแปรออกจากสมการจะทำให้ประสิทธิภาพในการพยากรณ์ตัวแปรตามลดลง

3) การเลือกตัวแปรโดยวิธีเพิ่มตัวแปรอิสระแบบขั้นตอน (Stepwise Regression) เป็นวิธีที่มีขั้นตอนคล้ายกับวิธีเพิ่มตัวแปร (Forward Selection) แต่จะทำการทดสอบตัวแปรอิสระที่เข้าไปในสมการแล้วทุกครั้งที่มีการนำตัวแปรอิสระตัวใหม่เข้าไปในสมการ ดังนั้นตัวแปรอิสระที่เข้าไปในสมการแล้วก็จะสามารถถูกคัดออกจากสมการได้ หากพบว่าตัวแปรอิสระตัวนั้น ๆ ไม่ได้ส่งผลให้ค่า R2 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 เก็บรวบรวมข้อมูลของคนเดินข้ามด้วยการตั้งกล้องถ่ายวีดิโอ 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 แสดงให้เห็นว่าพฤติกรรมการเหลือบมอง/การมองของคนเดินเท้า สิ่งรบกวนต่าง ๆ ความยาวรอบไฟสัญญาณ การถือสัมภาระ และปริมาณจราจรมีผลกระทบต่อพฤติกรรมการฝ่าฝืนไฟสัญญาณของคนเดินเท้า การปรับไฟสัญญาณโดยลดความยาวของช่วงไฟสัญญาณสีแดงสำหรับคนเดินเท้า อาจเป็นวิธีที่มีประสิทธิภาพที่สุดในการลดความน่าจะ

เป็นที่คนเดินเท้าจะฝ่าฝืนไฟสัญญาณ และลดความน่าจะเป็นที่จะเกิดอุบัติเหตุยานพาหนะชนคนเดินเท้า

Zhang and Deng (2019) ได้ศึกษาพฤติกรรมการฝ่าไฟแดงของคนเดินเท้าที่ทางแยกมีสัญญาณไฟในเมืองกวางโจว ประเทศจีน โดยเก็บข้อมูลจากทางแยก 3 แห่งผ่านการบันทึกวิดีโอ และจำแนกพฤติกรรมการฝ่าฝืนออกเป็น 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 เมืองมัชฮัด 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

ใช้วิธีBinary Logistic Regression สำหรับการวิเคราะห์กรณีการฝ่าฝืนไฟสัญญาณ (violation) จากการศึกษาพบว่าปัจจัยส่วนบุคคล (วัย, การใช้มือถือ, 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 และปัจจัยทางด้านสังคม เช่น การมีผู้คนเดินเท้าคนอื่นๆ อยู่ด้วย มีผลกระทบต่อพฤติกรรมการละเมิดกฎจราจรอย่างมีนัยสำคัญ คนเดินข้ามที่เป็นผู้หญิงและใช้อุปกรณ์ช่วยเดินมักจะปฏิบัติตามกฎจราจรได้ดีกว่า ส่วนปัจจัยแวดล้อม เช่น อุณหภูมิสูง ช่วงเวลากลางคืน หรือช่วงเช้ามืด ส่งผลให้คนเดินข้ามมีโอกาสที่จะฝ่าฝืนไฟสัญญาณเพิ่มสูงขึ้น และในส่วนของปัจจัยทางสังคม เช่น การมีคนอื่นข้ามพร้อมกัน หรือการรอเวลานานขึ้น มีแนวโน้มช่วยลดโอกาสที่คนเดินข้ามจะฝ่าฝืนไฟสัญญาณจราจรลงได้

วีรพงศ์ บุญเจริญปัญญา และจำรัส พิทักษ์ศฤงคาร (2567) งานวิจัยนี้มุ่งศึกษาทัศนคติของคนเดินเท้าในการข้ามถนน ตลอดจนปัจจัยที่มีอิทธิพลต่อทัศนคติเหล่านั้น กรณีศึกษาทางข้ามถนนบริเวณถนนแจ้งวัฒนะ ถนนลาดกระบัง และถนนกิ่งแก้ว โดยใช้แบบสอบถามจำนวน 2 ส่วน ได้แก่ ส่วนที่เกี่ยวกับทัศนคติในการข้ามถนน และส่วนที่เกี่ยวกับทัศนคติด้านความปลอดภัยในการข้ามถนน

รวมทั้งสิ้นจำนวน 412 ชุด ตัวแปรที่ใช้ในการศึกษาได้แก่ ตัวแปรทางสังคม ตัวแปรด้านความถี่ต่อประสบการณ์ที่ไม่ดีในการใช้งานทางข้ามถนน ตัวแปรด้านทัศนคติในการใช้งานทางข้ามถนนจากสถานที่ตัวอย่าง และตัวแปรด้านทัศนคติที่มีต่อสิ่งอำนวยความสะดวกในการข้ามถนน ในการวิเคราะห์ข้อมูล ใช้สถิติเชิงพรรณนาและการวิเคราะห์ถดถอยเชิงเส้น (Linear Regression Analysis) ผลการศึกษาพบว่า จำนวนช่องจราจร ปริมาณจราจร และความเร็วของยานพาหนะส่งผลเสียต่อความรู้สึกปลอดภัยในการข้ามถนนอย่างมีนัยสำคัญ ในด้านของปัจจัยทางสังคมพบว่าปัจจัยเพศ เป็นปัจจัยหลักที่ส่งผลต่อความรู้สึกปลอดภัยอย่างมีนัยสำคัญ และผลการศึกษาในด้านของตัวแปรด้านทัศนคติที่มีต่อสิ่งอำนวยความสะดวกในการข้าม พบว่าคนเดินเท้ารู้สึกปลอดภัยเมื่อมีการเพิ่มสัญญาณไฟจราจรแบบปุ่มกดมากที่สุด รองลงมาเป็นการมีอยู่ของเกาะกลางถนน และการเพิ่มเนินชะลอความเร็ว (Rubble strip) ตามลำดับ

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

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

เมื่อเทียบกับข้อมูลเริ่มต้นและวิธี ROSE จากการพิจารณาค่าความแม่นยำ พบว่า วิธีต้นไม้ตัดสินใจให้ค่าความแม่นยำสูงกว่าวิธีถดถอยโลจิสติก อย่างไรก็ตาม เมื่อพิจารณาอัตราความถูกต้องในการทำนายกลุ่มส่วนน้อย ซึ่งเป็นกลุ่มเป้าหมายที่สำคัญ พบว่า หลังจากปรับขนาดข้อมูลแล้ว วิธีถดถอยโลจิสติกให้ผลลัพธ์ที่ดีกว่าหรือเทียบเท่ากับวิธีต้นไม้ตัดสินใจ

วิชญ์วิสิฐ เกษรสิทธิ์, วิชิต หล่อจีระชุณห์กุล 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 สินสมบูรณ์ทอง (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 ความเร็วในการข้าม เวลารอคอย 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	ผลการศึกษาพบว่าเมื่อระยะเวลารอคอยเพิ่มมากขึ้น 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 และในทางกลับกันการเพิ่มขึ้นของความยาวของทางม้าลายและปริมาณการจราจร 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 หรือช่วงเช้ามืด ส่งผลให้คนเดินข้ามมีโอกาสที่จะฝ่าฝืนไฟสัญญาณเพิ่มสูงขึ้น และในส่วนของปัจจัยทางสังคม เช่น การมีคนอื่นข้ามพร้อมกัน หรือการรอเวลานานขึ้น มีแนวโน้มช่วยลดโอกาสที่คนเดินข้ามจะฝ่าฝืนไฟสัญญาณจราจรลงได้

ตารางที่ 2-3 สรุปงานวิจัยที่เกี่ยวข้อง (ต่อ)

ผู้แต่ง	วัตถุประสงค์	วิธีการวิเคราะห์	ผลการศึกษา
วีรพงศ์ บุญเจริญปัญญา และ จำรัส พิทักษ์ศฤงคาร (2567)	- ศึกษาทัศนคติในการข้ามถนน และปัจจัยที่ส่งผลกระทบทัศนคติในการข้ามถนน	- สถิติเชิงพรรณนาและการวิเคราะห์ถดถอยแบบเส้น (Linear Regression Analysis	ผลการศึกษาพบว่า จำนวนช่องจราจร ปริมาณจราจร และความเร็วของยานพาหนะส่งผลเสียต่อความรู้สึกปลอดภัยในการข้ามถนนอย่างมีนัยสำคัญ ในด้านของปัจจัยทางสังคมพบว่า เพศ เป็นปัจจัยหลักที่ส่งผลต่อความรู้สึกปลอดภัยอย่างมีนัยสำคัญ ในด้านของตัวแปรด้านทัศนคติที่มีต่อสิ่งอำนวยความสะดวกในการข้าม พบว่าคนเดินเท้ารู้สึกปลอดภัยเมื่อมีการเพิ่มสัญญาณไฟจราจรแบบปุ่มกดมากที่สุด รองลงมาเป็นการมีอยู่ของเกาะกลางถนน และการเพิ่มเนินชะลอความเร็ว (Rubble strip) ตามลำดับ
ชิติภัทธ์ 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 ปริมาณรถเลี้ยวขวาของการจราจรด้านใกล้ หมายความว่า หากตัวแปรกลุ่มนี้มีค่าเพิ่มมากขึ้นจะส่งผลทำให้ค่าความยากในการข้ามถนนบริเวณทางแยกของคนเดินเท้าสูง (ข้ามยาก)
นพมาศ อัครจันทโชติ และดิเรก พนิตสุภากมล (2561)	- การศึกษานี้ดำเนินการเปรียบเทียบประสิทธิภาพการแก้ปัญหาข้อมูลไม่สมดุล - ดำเนินการเปรียบเทียบประสิทธิภาพในการจำแนกข้อมูลสำหรับการจำแนกกลุ่มรายได้ผู้ประกอบการร้านยาประเภท ข.ย. 1	- ใช้วิธีการสุ่มตัวอย่างเพิ่มข้อมูลเริ่มต้นอย่างสุ่ม (ROSE) และการสุ่มตัวอย่างเพิ่มกลุ่มส่วนน้อยด้วยการสังเคราะห์ (SMOTE) เพื่อเปรียบเทียบประสิทธิภาพการแก้ปัญหาข้อมูลไม่สมดุล - ใช้วิธีการการถดถอยโลจิสติค และวิธีต้นไม้ตัดสินใจ เพื่อเปรียบเทียบประสิทธิภาพในการจำแนกข้อมูล	จากผลการเปรียบเทียบวิธีการจำแนกประเภทข้อมูล พบว่า SMOTE ให้ผลลัพธ์ที่ดีที่สุดในตัวชี้วัดเกือบทั้งหมด ยกเว้นความแม่นยำในการพยากรณ์กลุ่มที่มีจำนวนข้อมูลมาก ซึ่งยังด้อยกว่าวิธีอื่นเมื่อเทียบกับข้อมูลเริ่มต้นและวิธี ROSE ผลจากการเปรียบเทียบการจำแนกประเภทข้อมูล - เมื่อพิจารณาค่าความแม่นยำ พบว่า วิธีต้นไม้ตัดสินใจให้ค่าความแม่นยำสูงกว่าวิธีถดถอยโลจิสติก - เมื่อพิจารณาอัตราความถูกต้องในการทำนายกลุ่มส่วนน้อย ซึ่งเป็นกลุ่มเป้าหมายที่สำคัญ พบว่า หลังจากปรับขนาดข้อมูลแล้ว วิธีถดถอยโลจิสติกให้ผลลัพธ์ที่ดีกว่าหรือ

ตารางที่ 2-3 สรุปงานวิจัยที่เกี่ยวข้อง (ต่อ)

ผู้แต่ง	วัตถุประสงค์	วิธีการวิเคราะห์	ผลการศึกษา
			เทียบเท่ากับวิธีต้นไม้ตัดสินใจ
วิชญ์วิสิฐ เกษรสิทธิ์, วิชิต หล่อจีระชุณห์กุล และ จิราวัลย์ จิตรถเวช (2561)	- ศึกษาเปรียบเทียบประสิทธิภาพของวิธีการแก้ปัญหาข้อมูลไม่สมดุลของข้อมูลผู้ป่วยโรคเบาหวาน	- เปรียบเทียบวิธีการแก้ปัญหาข้อมูลไม่สมดุลที่ระดับข้อมูล 4 วิธี ได้แก่ การสุ่มเพิ่มข้อมูล, การสุ่มลดข้อมูล, การผสมผสานระหว่างการเพิ่มและลดข้อมูล และการสังเคราะห์ข้อมูลใหม่ด้วยเทคนิค (SMOTE) - ใช้เทคนิคการจำแนกประเภทของข้อมูลด้วยสองวิธีหลัก ได้แก่ การถดถอยโลจิสติกแบบพหุนาม (Multinomial Logistic Regression) และวิธีต้นไม้ตัดสินใจ (Decision Tree)	จากการศึกษาพบว่าการแก้ปัญหาข้อมูลไม่สมดุล สามารถเพิ่มประสิทธิภาพการจำแนกประเภทข้อมูลส่วนน้อยได้ถูกต้องและแม่นยำมากขึ้น
อัจฉรา แผ้วบาง และสายชล สินสมบูรณ์ทอง (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	