

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

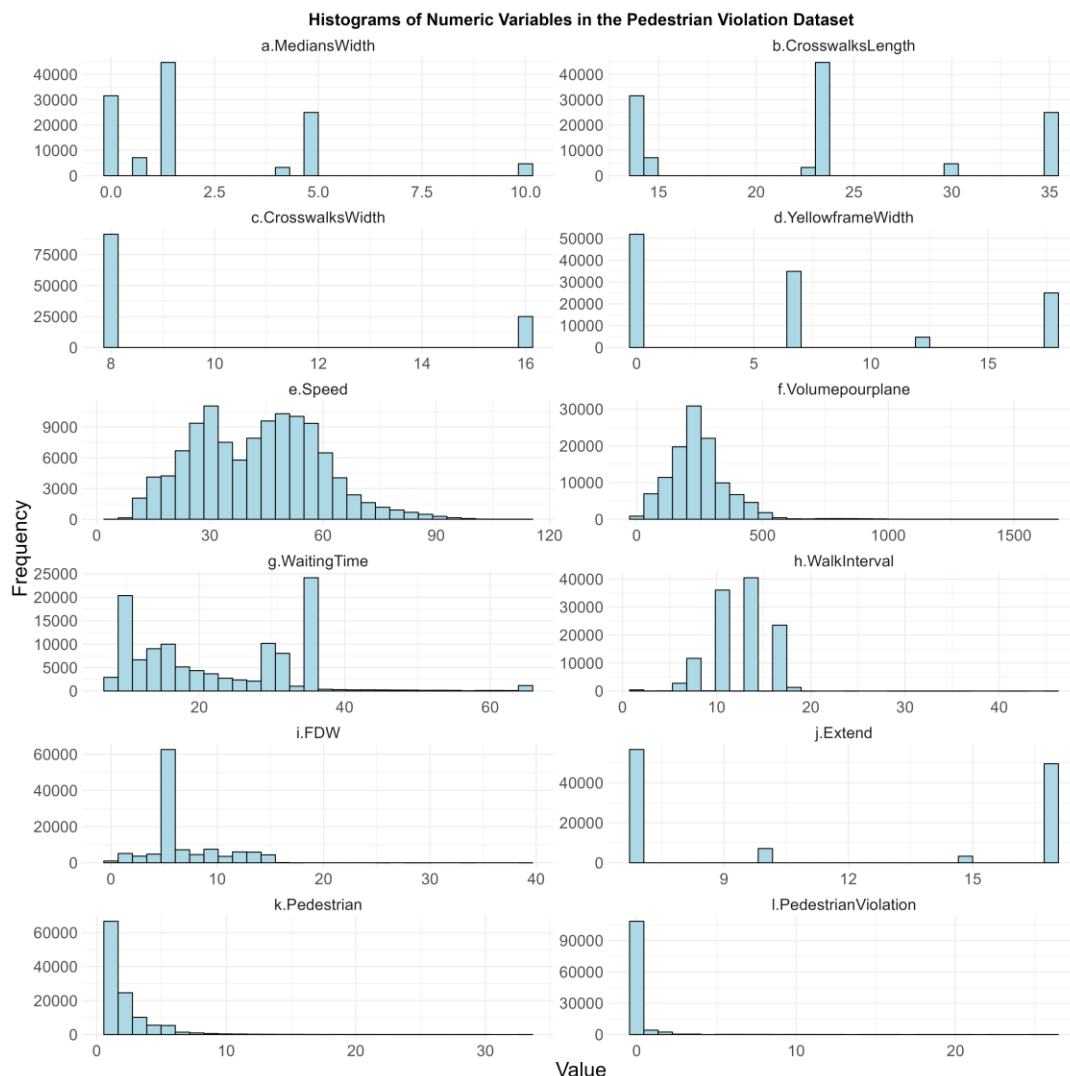
ข้อมูลลักษณะทางกายภาพของทางข้ามอัจฉริยะ								
ข้อมูลลักษณะทางกายภาพ	ชื่อตัวแปรที่ใช้ ในการวิเคราะห์	MIN	15 th Percentile	MEAN	MEDIAN	85 th Percentile	MAX	S.D.
ความกว้างของเกาะกลางถนน (เมตร)	MediansWidth	0.00	0.00	2.13	1.30	5.00	10.00	2.45
ความยาวของทางข้าม (เมตร)	CrosswalksLength	13.80	13.80	22.94	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				ชื่อตัวแปรที่ใช้ ในการวิเคราะห์	ค่าความถี่		ค่าร้อยละ	
เหตุการณ์การการฝ่าฝืนไฟสัญญาณของคนเดินข้ามที่เกิดขึ้นระหว่างรอบไฟสัญญาณ (n = 116,485 รอบไฟสัญญาณ)				VioPedestrian				
ไม่พบคนเดินข้ามฝ่าฝืนไฟสัญญาณ (0 = Not Violated)					108,490		93.14	
พบคนเดินข้ามอย่างน้อยหนึ่งคนฝ่าฝืนไฟสัญญาณ (1 = Violated)					7,995		6.86	
ช่วงเวลาของวันที่พบเหตุการณ์การการฝ่าฝืนไฟสัญญาณของคนเดินข้ามหลังจากกดปุ่ม ขอไฟสัญญาณเพื่อข้ามถนน				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 แต่การฝ่าฝืนดังกล่าวมีความเสี่ยงที่อาจจะก่อให้เกิดอุบัติเหตุที่ร้ายแรงขึ้นกับผู้ใช้รถใช้ถนน 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

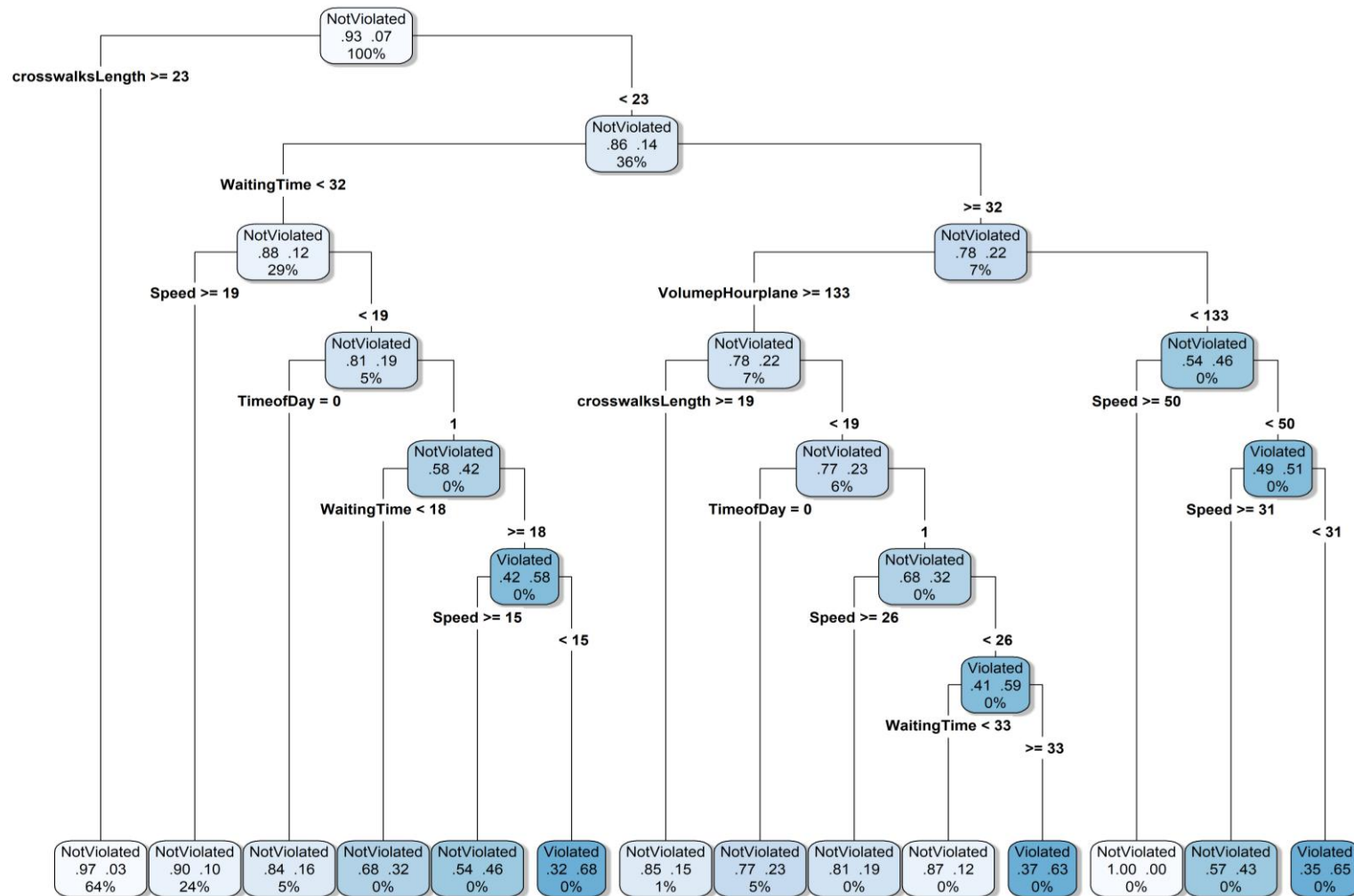
หมายเหตุ *** มีระดับนัยสำคัญทางสถิติที่ 0.001, $R^2 = 0.086$

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	Sensitivity		
		100%	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 หล่อจ๊ะระชุมห์กุล และจิราวัลย์ จิตรถเวช, 2561; อัจฉรา แผ้วบาง และสายชล สันสมบูรณ์ทอง, 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) เพื่อเป็นการตอบวัตถุประสงค์งานวิจัยและคำถามงานวิจัยในการศึกษานี้ 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 มิได้เกิดจากความผันแปรโดยบังเอิญในกระบวนการสุ่มข้อมูล 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:

ระยะเวลาการรอคอย ($B = 0.046$ และ $\text{Exp}(B) = 1.047$) พบว่า ภายใต้สภาพที่ปัจจัยอื่นมีค่าคงที่หากระยะเวลาการรอคอยเพิ่มขึ้น 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 ระยะเวลาการรอคอยที่ยาวนานขึ้นจะส่งผลให้คนเดินข้ามมีแนวโน้มที่จะฝ่าฝืนไฟสัญญาณเพิ่มมากขึ้น

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 ภายใต้อาณัติสภาพที่ปัจจัยอื่นมีค่าคงที่หากความเร็วในการขับเคลื่อนของยานพาหนะเพิ่มขึ้น 1 กม./ชม. ส่งผลให้คนเดินข้ามมีโอกาสที่จะฝ่าฝืนไฟสัญญาณลดลง 0.014 หน่วย หรือลดลงประมาณ 1.4 % ผลการวิเคราะห์นี้สอดคล้องกับงานวิจัยของ วีรพงศ์ บุญเจริญปัญญา และจำรัส พิทักษ์ศฤงคาร (2567) ซึ่งผลการศึกษาชี้ให้เห็นว่า เมื่อความเร็วของยานพาหนะเพิ่มขึ้น จะทำให้ระดับความรู้สึกปลอดภัยของคนเดินข้ามลดลงอย่างมีนัยสำคัญ ซึ่งส่งผลให้คนเดินข้ามมีโอกาสที่จะฝ่าฝืนไฟสัญญาณลดลง

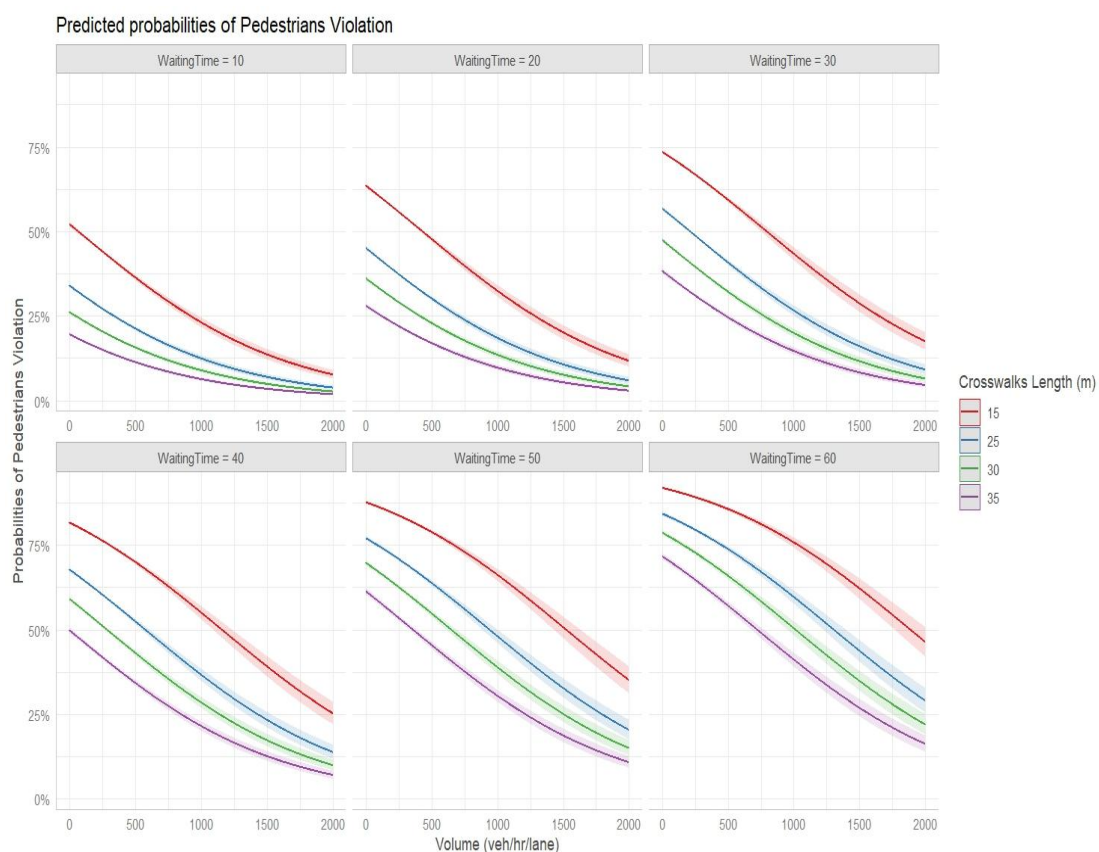
ปริมาณจราจรที่สัญจรผ่านทางข้าม ($B = -0.001$ และ $\text{Exp}(B) = 0.999$) พบว่า ภายใต้อาณัติสภาพที่ปัจจัยอื่นมีค่าคงที่หากปริมาณจราจรเพิ่มขึ้น 1 คัน/ชม/ช่องจราจร ส่งผลให้คนเดินข้ามมีโอกาสที่จะฝ่าฝืนไฟสัญญาณลดลง 0.001 หน่วย หรือลดลงประมาณ 0.1 % ผลการวิเคราะห์นี้สอดคล้องกับงานวิจัยของ Afshari and Ayati (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 ภายใต้อาณัติสภาพที่ปัจจัยอื่นมีค่าคงที่หากช่วงเวลาของวันเป็นช่วงเวลากลางคืน 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 ซึ่งมีความสำคัญสำหรับการศึกษานี้ ในครั้งนี้ 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, วิจิต หล่อจิระ 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 เพื่อพิจารณาว่าค่าเฉลี่ยของคู่ใดบ้างที่แตกต่างกัน ผลการทดสอบแสดงดังตารางที่ 4-17

ตารางที่ 4-17 การทดสอบความแตกต่างของค่าเฉลี่ย AUC จากวิธีต้นไม้ตัดสินใจรายคู่ ด้วยวิธี

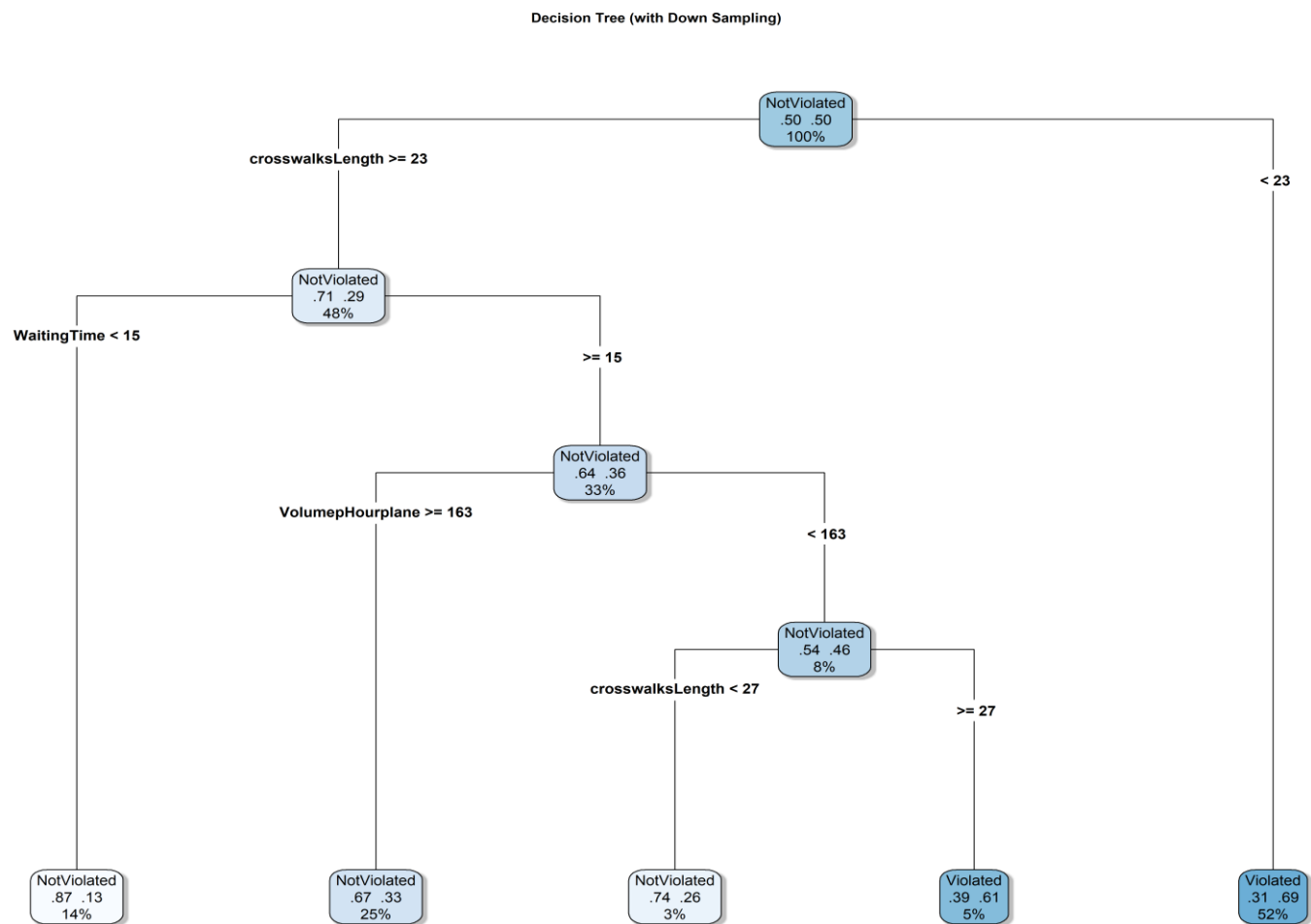
ทดสอบแบบ Dunnett T3

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) มากกว่าหรือเท่ากับ 15 วินาที และปริมาณจราจรที่สัญจรผ่านทางข้ามมากกว่าหรือเท่ากับ 163 คัน/ชั่วโมง/ช่องจราจร คนเดินข้ามจะตัดสินใจที่จะฝ่าฝืนไฟสัญญาณ 33%

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

4) ถ้าความยาวทางข้าม (CrosswalksLength) มากกว่าหรือเท่ากับ 27 เมตร เวลาที่คนเดินข้ามยืนรอไฟสัญญาณ (WaitingTime) มากกว่าหรือเท่ากับ 15 วินาที และปริมาณจราจรที่สัญจรผ่านทางข้ามน้อยกว่า 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