บทที่ 3
วิธีการดำเนินงานโครงงาน

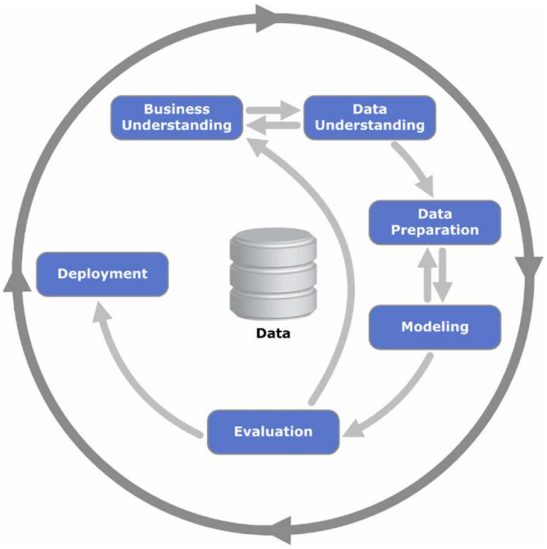
โครงงานเรื่อง การวิเคราะห์ข้อมูลผู้บาดเจ็บและเสียชีวิตในช่วงเทศกาลปีใหม่ประจำปี 2551 – 2558 เพื่อใช้สำหรับเผยแพร่ข้อมูลบนเว็บไซต์ ในบทนี้จะเป็นการวิเคราะห์ข้อมูลด้วยเทคนิคทางดาต้าไมนิ่ง ซึ่งมีกระบวนการวิเคราะห์ที่สำคัญหลายขั้นตอน เมื่อเสร็จสิ้นจากกระบวนการวิเคราะห์ข้อมูลแล้วจะเป็นการออกแบบเว็บไซต์ และออกแบบรูปแบบการแสดงผลและบทสรุปจากวิธีการดำเนินงาน

3.1 การวิเคราะห์ข้อมูลด้วย CRISP–DM

3.2 การออกแบบเว็บไซต์

3.3 บทสรุป

3.1 การวิเคราะห์ข้อมูลด้วย CRISP–DM



ภาพที่ 3.1 กระบวนการวิเคราะห์ข้อมูล ด้วย CRISP–DM

(ที่มา : http://dataminingtrend.com/2014/data–mining–techniques/crisp–dm–example/)

กระบวนการวิเคราะห์ข้อมูล (Cross Industry Standard Process for Data Mining หรือ CRISP–DM) พัฒนาขึ้นในปี ค.ศ. 1996 โดยความร่วมมือของ 3 บริษัทคือ Daimler Chrysler, SPSS และ NCR ซึ่งกระบวนการวิเคราะห์ข้อมูล CRISP–DM จะเป็นดังรูปแบบในรูปภาพที่ 3.1 แต่ละ

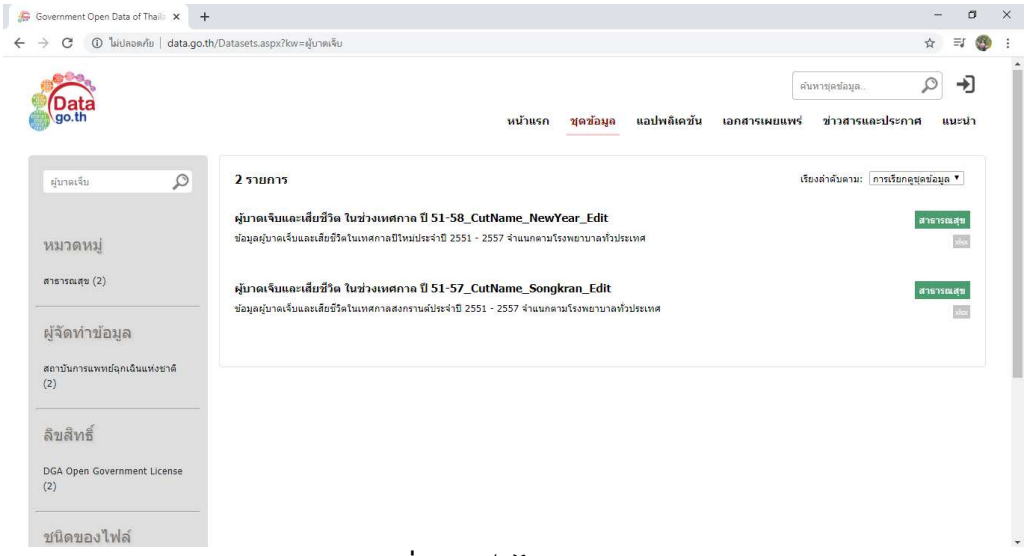
ขั้นตอนในรูปจะเป็นขั้นตอนที่ต่อเนื่องกันนั่นคือ ขั้นตอนถัดไปจะรอผลลัพธ์จากขั้นตอนก่อนหน้า ซึ่งแสดงด้วยลูกศรที่เชื่อมระหว่างกระบวนการนั้น ๆ ประกอบด้วย 6 ขั้นตอนหลัก ดังนี้

3.1.1 รู้จักและเข้าใจในธุรกิจ (Business Understanding) เป็นขั้นตอนแรกของกระบวนการที่มุ่งเน้นไปที่การทำความเข้าใจกระบวนการทางธุรกิจโดยรวม

คณะผู้จัดทำทำความเข้าใจกับปัญหาให้อยู่ในรูปของการวิเคราะห์ข้อมูลทาง Data Mining โดยการวิเคราะห์ข้อมูลในประเด็นนี้ คือ ข้อมูลผู้บาดเจ็บและเสียชีวิตในช่วงเทศกาลปีใหม่ประจำปี 2551 – 2558 ซึ่งมีจำนวนข้อมูลมหาศาล จำนวนรายการทั้งหมด 214,950 รายการ ทำให้ไม่สามารถทำความเข้าใจกับข้อมูลที่มีความซับซ้อนได้อย่างรวดเร็ว เช่น ต้องการทราบว่าช่วงอายุใดมีอัตราการเกิดอุบัติเหตุสูงสุด

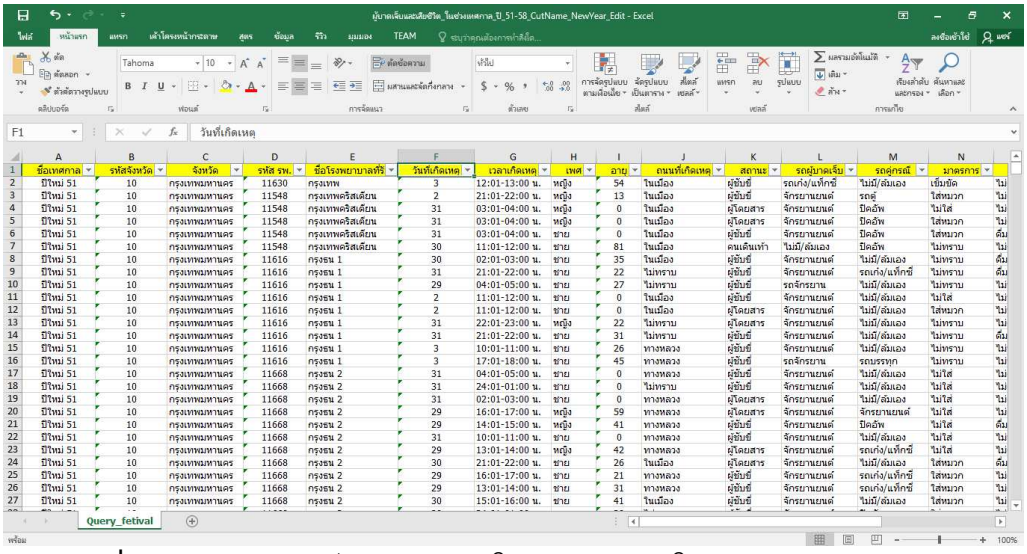
3.1.2 จัดเก็บและรวบรวมข้อมูลให้ครบ (Data Understanding) ขั้นตอนการจัดเก็บและรวบรวมข้อมูล ตลอดจนการพิจารณาตรวจสอบความถูกต้องของข้อมูลที่ได้รับ โดยเลือกว่าจะใช้ข้อมูลทั้งหมดหรือบางส่วนในการวิเคราะห์ให้สอดรับกับวัตถุประสงค์ที่กำหนดไว้

คณะผู้จัดทำทำการรวบรวมข้อมูล เพื่อตรวจสอบรายละเอียด ปริมาณ และความน่าเชื่อถือของข้อมูลผู้บาดเจ็บและเสียชีวิตในช่วงเทศกาลปีใหม่ประจำปี 2551 – 2558 ที่ได้จากเว็บไซต์ data.go.th ซึ่งเป็นระบบศูนย์กลางข้อมูลภาครัฐ ที่เผยแพร่สู่สาธารณะอย่างเป็นรูปธรรม เป็นช่องทางให้ผู้ใช้บริการทั้งภาคประชาชน ภาคธุรกิจเอกชน รวมถึงหน่วยงานของรัฐสามารถค้นหา และเข้าถึงข้อมูลที่มีคุณภาพของภาครัฐได้ง่าย



ภาพที่ 3.2 เว็บไซต์ data.go.th

ซึ่งข้อมูลผู้บาดเจ็บและเสียชีวิตในช่วงเทศกาลปีใหม่ประจำปี 2551 – 2558 มีจำนวนข้อมูล 214,951 รายการ ประกอบด้วย 19 แอตทริบิวท์ ประกอบด้วย ชื่อเทศกาล รหัสจังหวัด จังหวัด รหัสโรงพยาบาล ชื่อโรงพยาบาลที่รับผู้บาดเจ็บ วันที่เกิดเหตุ เวลาเกิดเหตุ เพศ อายุ ถนนที่เกิดเหตุ สถานะ รถผู้บาดเจ็บ รถคู่กรณี มาตรการ การดื่มสุรา การนำส่ง Refer-Admit ผลการรักษา และจำนวนวันรักษา



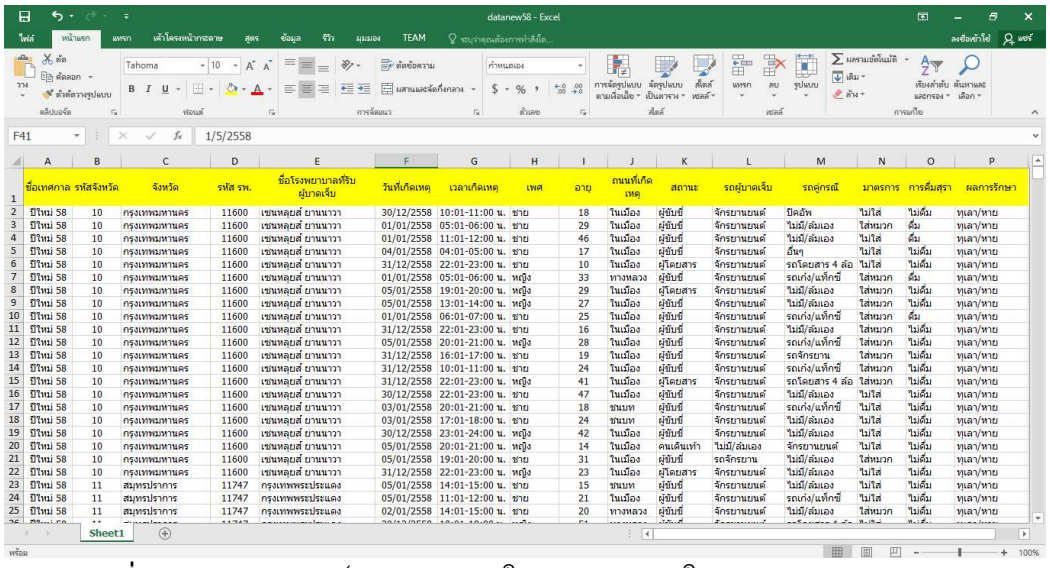
ชื่อเทศกาล	รหัสจังหวัด	จังหวัด	รหัส รพ.	ชื่อโรงพยาบาลที่รั	วันที่เกิดเหตุ	เวลาเกิดเหตุ	เพศ	อายุ	ถนนที่เกิดเหตุ	สถานะ	รถผู้บาดเจ็บ	รถคู่กรณี	มาตรการ
ปีใหม่ 51	10	กรุงเทพมหานคร	11630	กรุงเทพ	3	12:01-13:00 น.	หญิง	54	ในเมือง	ผู้บับขี่	รถเก๋ง/แท็กซี่	ไม่มี/ดื่มเอง	เข็มขัด
ปีใหม่ 51	10	กรุงเทพมหานคร	11548	กรุงเทพคริสเตียน	2	21:01-22:00 น.	หญิง	13	ในเมือง	ผู้บับขี่	จักรยานยนต์	รถตู้	ใส่หมวก
ปีใหม่ 51	10	กรุงเทพมหานคร	11548	กรุงเทพคริสเตียน	31	03:01-04:00 น.	หญิง	0	ในเมือง	ผู้โดยสาร	จักรยานยนต์	ปิคอัพ	ไม่ใส่
ปีใหม่ 51	10	กรุงเทพมหานคร	11548	กรุงเทพคริสเตียน	31	03:01-04:00 น.	หญิง	0	ในเมือง	ผู้โดยสาร	จักรยานยนต์	ปิคอัพ	ใส่หมวก
ปีใหม่ 51	10	กรุงเทพมหานคร	11548	กรุงเทพคริสเตียน	31	03:01-04:00 น.	ชาย	0	ในเมือง	ผู้บับขี่	จักรยานยนต์	ปิคอัพ	ใส่หมวก
ปีใหม่ 51	10	กรุงเทพมหานคร	11548	กรุงเทพคริสเตียน	30	11:01-12:00 น.	ชาย	81	ในเมือง	คนเดินเท้า	ไม่มี/ดื่มเอง	ปิคอัพ	ไม่ทราบ
ปีใหม่ 51	10	กรุงเทพมหานคร	11616	กรุงธน 1	30	02:01-03:00 น.	ชาย	35	ในเมือง	ผู้บับขี่	จักรยานยนต์	ไม่มี/ดื่มเอง	ไม่ทราบ
ปีใหม่ 51	10	กรุงเทพมหานคร	11616	กรุงธน 1	31	21:01-22:00 น.	ชาย	22	ไม่ทราบ	ผู้บับขี่	จักรยานยนต์	รถเก๋ง/แท็กซี่	ไม่ทราบ
ปีใหม่ 51	10	กรุงเทพมหานคร	11616	กรุงธน 1	29	04:01-05:00 น.	ชาย	27	ไม่ทราบ	ผู้บับขี่	รถจักรยาน	ไม่มี/ดื่มเอง	ไม่ทราบ
ปีใหม่ 51	10	กรุงเทพมหานคร	11616	กรุงธน 1	2	11:01-12:00 น.	ชาย	0	ในเมือง	ผู้บับขี่	จักรยานยนต์	ไม่มี/ดื่มเอง	ไม่ใส่
ปีใหม่ 51	10	กรุงเทพมหานคร	11616	กรุงธน 1	2	11:01-12:00 น.	ชาย	0	ในเมือง	ผู้โดยสาร	จักรยานยนต์	ไม่มี/ดื่มเอง	ใส่หมวก
ปีใหม่ 51	10	กรุงเทพมหานคร	11616	กรุงธน 1	31	22:01-23:00 น.	หญิง	22	ไม่ทราบ	ผู้โดยสาร	จักรยานยนต์	ไม่มี/ดื่มเอง	ไม่ทราบ
ปีใหม่ 51	10	กรุงเทพมหานคร	11616	กรุงธน 1	31	21:01-22:00 น.	ชาย	31	ไม่ทราบ	ผู้บับขี่	จักรยานยนต์	ไม่มี/ดื่มเอง	ไม่ทราบ
ปีใหม่ 51	10	กรุงเทพมหานคร	11616	กรุงธน 1	3	10:01-11:00 น.	ชาย	26	ทางหลวง	ผู้บับขี่	จักรยานยนต์	ไม่มี/ดื่มเอง	ไม่ทราบ
ปีใหม่ 51	10	กรุงเทพมหานคร	11616	กรุงธน 1	3	17:01-18:00 น.	ชาย	45	ทางหลวง	ผู้บับขี่	รถจักรยาน	รถบรรทุก	ไม่ทราบ
ปีใหม่ 51	10	กรุงเทพมหานคร	11668	กรุงธน 2	31	04:01-05:00 น.	ชาย	0	ทางหลวง	ผู้บับขี่	จักรยานยนต์	ไม่มี/ดื่มเอง	ไม่ใส่
ปีใหม่ 51	10	กรุงเทพมหานคร	11668	กรุงธน 2	31	24:01-01:00 น.	ชาย	0	ไม่ทราบ	ผู้บับขี่	จักรยานยนต์	ไม่มี/ดื่มเอง	ไม่ใส่
ปีใหม่ 51	10	กรุงเทพมหานคร	11668	กรุงธน 2	31	02:01-03:00 น.	ชาย	0	ทางหลวง	ผู้โดยสาร	จักรยานยนต์	ไม่มี/ดื่มเอง	ไม่ใส่
ปีใหม่ 51	10	กรุงเทพมหานคร	11668	กรุงธน 2	29	16:01-17:00 น.	หญิง	59	ทางหลวง	ผู้โดยสาร	จักรยานยนต์	จักรยานยนต์	ไม่ใส่
ปีใหม่ 51	10	กรุงเทพมหานคร	11668	กรุงธน 2	29	14:01-15:00 น.	หญิง	41	ทางหลวง	ผู้บับขี่	จักรยานยนต์	ปิคอัพ	ไม่ใส่
ปีใหม่ 51	10	กรุงเทพมหานคร	11668	กรุงธน 2	31	10:01-11:00 น.	ชาย	0	ทางหลวง	ผู้บับขี่	จักรยานยนต์	ไม่มี/ดื่มเอง	ไม่ใส่
ปีใหม่ 51	10	กรุงเทพมหานคร	11668	กรุงธน 2	29	13:01-14:00 น.	หญิง	42	ทางหลวง	ผู้โดยสาร	จักรยานยนต์	รถเก๋ง/แท็กซี่	ไม่ใส่
ปีใหม่ 51	10	กรุงเทพมหานคร	11668	กรุงธน 2	30	21:01-22:00 น.	ชาย	26	ในเมือง	ผู้โดยสาร	จักรยานยนต์	ไม่มี/ดื่มเอง	ใส่หมวก
ปีใหม่ 51	10	กรุงเทพมหานคร	11668	กรุงธน 2	29	16:01-17:00 น.	ชาย	21	ทางหลวง	ผู้บับขี่	จักรยานยนต์	รถเก๋ง/แท็กซี่	ใส่หมวก
ปีใหม่ 51	10	กรุงเทพมหานคร	11668	กรุงธน 2	29	13:01-14:00 น.	ชาย	31	ทางหลวง	ผู้บับขี่	จักรยานยนต์	รถเก๋ง/แท็กซี่	ใส่หมวก
ปีใหม่ 51	10	กรุงเทพมหานคร	11668	กรุงธน 2	30	15:01-16:00 น.	ชาย	41	ในเมือง	ผู้บับขี่	จักรยานยนต์	ไม่มี/ดื่มเอง	ใส่หมวก

ภาพที่ 3.3 ข้อมูลผู้บาดเจ็บและเสียชีวิตในช่วงเทศกาลปีใหม่ประจำปี 2551 – 2558

3.1.3 เตรียมข้อมูลให้พร้อมใช้งาน (Data Preparation) ขั้นตอนการแปลงข้อมูลที่ได้รวบรวมมาและเลือกไว้ ให้อยู่ในรูปแบบที่พร้อมสำหรับนำไปวิเคราะห์ในขั้นตอนต่อไปได้ โดยการทำให้เป็นข้อมูลที่ถูกต้อง (Data cleaning) มักใช้เวลาค่อนข้างมาก โดยมีขั้นตอนดังนี้

3.1.3.1 ทำการคัดเลือกข้อมูล (Data Selection) คือการคัดเลือกข้อมูลที่เหมาะสมเพื่อนำมาใช้ในการวิเคราะห์ข้อมูล

คณะผู้จัดทำทำการคัดเลือกข้อมูล และทำการ Data Cleaning ข้อมูลผู้บาดเจ็บและเสียชีวิตในช่วงเทศกาลปีใหม่ประจำปี 2551–2558 โดยตัดส่วนที่ไม่จำเป็นออกให้เหลือเฉพาะข้อมูลที่จำเป็นในการวิเคราะห์ในภาพรวม จำนวน 16 แอตทริบิวท์ ได้แก่ ชื่อเทศกาล รหัสจังหวัด จังหวัด รหัสโรงพยาบาล ชื่อโรงพยาบาลที่รับผู้บาดเจ็บ วันที่เกิดเหตุ เวลาเกิดเหตุ เพศ อายุ ถนนที่เกิดเหตุ สถานะ รถผู้บาดเจ็บ รถคู่กรณี มาตรการ การดื่มสุรา และผลการรักษา โดยใช้ข้อมูลในปี 2551–2558 จำนวน 214,951 รายการ ซึ่งเป็นข้อมูลที่จำเป็นในการนำไปวิเคราะห์ข้อมูล



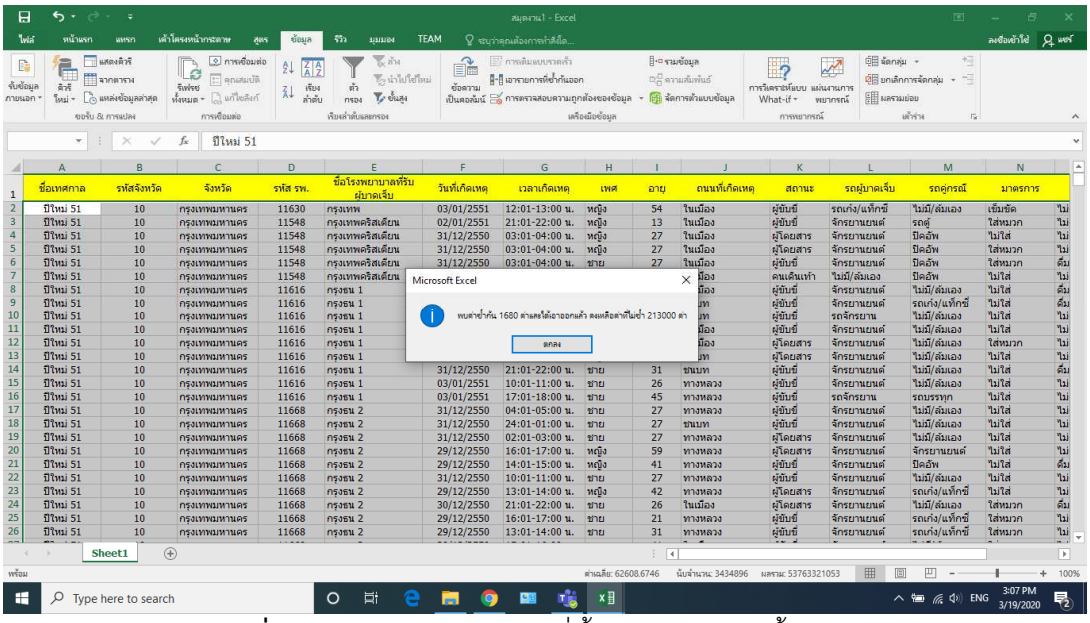
ชื่อเทศกาล	รหัสจังหวัด	จังหวัด	รหัส รพ.	ชื่อโรงพยาบาลที่รับผู้บาดเจ็บ	วันที่เกิดเหตุ	เวลาเกิดเหตุ	เพศ	อายุ	ถนนที่เกิดเหตุ	สถานะ	รถผู้บาดเจ็บ	รถคู่กรณี	มาตการ	การดื่มสุรา	ผลการรักษา
ปีใหม่ 58	10	กรุงเทพมหานคร	11600	เซนหลุยส์ ยานนาวา	30/12/2558	10:01-11:00 น.	ชาย	18	ในเมือง	ผู้ขับขี่	จักรยานยนต์	ปิคอัพ	ไม่ใส่	ไม่ดื่ม	ทุเลา/หาย
ปีใหม่ 58	10	กรุงเทพมหานคร	11600	เซนหลุยส์ ยานนาวา	01/01/2558	05:01-06:00 น.	ชาย	29	ในเมือง	ผู้ขับขี่	จักรยานยนต์	ไม่มี/ดัดแปลง	ใส่หมวก	ดื่ม	ทุเลา/หาย
ปีใหม่ 58	10	กรุงเทพมหานคร	11600	เซนหลุยส์ ยานนาวา	01/01/2558	11:01-12:00 น.	ชาย	46	ในเมือง	ผู้ขับขี่	จักรยานยนต์	ไม่มี/ดัดแปลง	ไม่ใส่	ดื่ม	ทุเลา/หาย
ปีใหม่ 58	10	กรุงเทพมหานคร	11600	เซนหลุยส์ ยานนาวา	04/01/2558	04:01-05:00 น.	ชาย	17	ในเมือง	ผู้ขับขี่	จักรยานยนต์	อื่นๆ	ไม่ใส่	ไม่ดื่ม	ทุเลา/หาย
ปีใหม่ 58	10	กรุงเทพมหานคร	11600	เซนหลุยส์ ยานนาวา	31/12/2558	22:01-23:00 น.	ชาย	10	ในเมือง	ผู้โดยสาร	จักรยานยนต์	รถโดยสาร 4 ล้อ	ไม่ใส่	ไม่ดื่ม	ทุเลา/หาย
ปีใหม่ 58	10	กรุงเทพมหานคร	11600	เซนหลุยส์ ยานนาวา	01/01/2558	05:01-06:00 น.	หญิง	33	ทางหลวง	ผู้ขับขี่	จักรยานยนต์	รถเก๋ง/แท็กซี่	ใส่หมวก	ดื่ม	ทุเลา/หาย
ปีใหม่ 58	10	กรุงเทพมหานคร	11600	เซนหลุยส์ ยานนาวา	05/01/2558	19:01-20:00 น.	หญิง	29	ในเมือง	ผู้โดยสาร	จักรยานยนต์	ไม่มี/ดัดแปลง	ใส่หมวก	ไม่ดื่ม	ทุเลา/หาย
ปีใหม่ 58	10	กรุงเทพมหานคร	11600	เซนหลุยส์ ยานนาวา	05/01/2558	13:01-14:00 น.	หญิง	27	ในเมือง	ผู้ขับขี่	จักรยานยนต์	ไม่มี/ดัดแปลง	ใส่หมวก	ไม่ดื่ม	ทุเลา/หาย
ปีใหม่ 58	10	กรุงเทพมหานคร	11600	เซนหลุยส์ ยานนาวา	01/01/2558	06:01-07:00 น.	ชาย	25	ในเมือง	ผู้ขับขี่	จักรยานยนต์	รถเก๋ง/แท็กซี่	ใส่หมวก	ดื่ม	ทุเลา/หาย
ปีใหม่ 58	10	กรุงเทพมหานคร	11600	เซนหลุยส์ ยานนาวา	31/12/2558	22:01-23:00 น.	ชาย	16	ในเมือง	ผู้ขับขี่	จักรยานยนต์	ไม่มี/ดัดแปลง	ใส่หมวก	ไม่ดื่ม	ทุเลา/หาย
ปีใหม่ 58	10	กรุงเทพมหานคร	11600	เซนหลุยส์ ยานนาวา	05/01/2558	20:01-21:00 น.	หญิง	28	ในเมือง	ผู้ขับขี่	จักรยานยนต์	รถเก๋ง/แท็กซี่	ใส่หมวก	ไม่ดื่ม	ทุเลา/หาย
ปีใหม่ 58	10	กรุงเทพมหานคร	11600	เซนหลุยส์ ยานนาวา	31/12/2558	16:01-17:00 น.	ชาย	19	ในเมือง	ผู้ขับขี่	จักรยานยนต์	รถจักรยาน	ใส่หมวก	ไม่ดื่ม	ทุเลา/หาย
ปีใหม่ 58	10	กรุงเทพมหานคร	11600	เซนหลุยส์ ยานนาวา	31/12/2558	10:01-11:00 น.	ชาย	24	ในเมือง	ผู้ขับขี่	จักรยานยนต์	รถเก๋ง/แท็กซี่	ใส่หมวก	ไม่ดื่ม	ทุเลา/หาย
ปีใหม่ 58	10	กรุงเทพมหานคร	11600	เซนหลุยส์ ยานนาวา	31/12/2558	22:01-23:00 น.	หญิง	41	ในเมือง	ผู้โดยสาร	จักรยานยนต์	รถโดยสาร 4 ล้อ	ใส่หมวก	ไม่ดื่ม	ทุเลา/หาย
ปีใหม่ 58	10	กรุงเทพมหานคร	11600	เซนหลุยส์ ยานนาวา	30/12/2558	22:01-23:00 น.	ชาย	47	ในเมือง	ผู้ขับขี่	จักรยานยนต์	ไม่มี/ดัดแปลง	ไม่ใส่	ไม่ดื่ม	ทุเลา/หาย
ปีใหม่ 58	10	กรุงเทพมหานคร	11600	เซนหลุยส์ ยานนาวา	03/01/2558	20:01-21:00 น.	ชาย	18	ชนบท	ผู้ขับขี่	จักรยานยนต์	รถเก๋ง/แท็กซี่	ไม่ใส่	ไม่ดื่ม	ทุเลา/หาย
ปีใหม่ 58	10	กรุงเทพมหานคร	11600	เซนหลุยส์ ยานนาวา	03/01/2558	17:01-18:00 น.	ชาย	24	ชนบท	ผู้ขับขี่	จักรยานยนต์	ไม่มี/ดัดแปลง	ไม่ใส่	ไม่ดื่ม	ทุเลา/หาย
ปีใหม่ 58	10	กรุงเทพมหานคร	11600	เซนหลุยส์ ยานนาวา	30/12/2558	23:01-24:00 น.	หญิง	42	ในเมือง	ผู้ขับขี่	จักรยานยนต์	ไม่มี/ดัดแปลง	ไม่ใส่	ไม่ดื่ม	ทุเลา/หาย
ปีใหม่ 58	10	กรุงเทพมหานคร	11600	เซนหลุยส์ ยานนาวา	05/01/2558	20:01-21:00 น.	หญิง	14	ในเมือง	คนเดินเท้า	ไม่มี/ดัดแปลง	จักรยานยนต์	ไม่ใส่	ไม่ดื่ม	ทุเลา/หาย
ปีใหม่ 58	10	กรุงเทพมหานคร	11600	เซนหลุยส์ ยานนาวา	05/01/2558	19:01-20:00 น.	ชาย	31	ในเมือง	ผู้ขับขี่	รถจักรยาน	ไม่มี/ดัดแปลง	ใส่หมวก	ไม่ดื่ม	ทุเลา/หาย
ปีใหม่ 58	10	กรุงเทพมหานคร	11600	เซนหลุยส์ ยานนาวา	31/12/2558	22:01-23:00 น.	หญิง	23	ในเมือง	ผู้โดยสาร	จักรยานยนต์	ไม่มี/ดัดแปลง	ไม่ใส่	ไม่ดื่ม	ทุเลา/หาย
ปีใหม่ 58	11	สมุทรปราการ	11747	กรุงเทพพระประแดง	05/01/2558	14:01-15:00 น.	ชาย	15	ชนบท	ผู้ขับขี่	จักรยานยนต์	ไม่มี/ดัดแปลง	ไม่ใส่	ไม่ดื่ม	ทุเลา/หาย
ปีใหม่ 58	11	สมุทรปราการ	11747	กรุงเทพพระประแดง	05/01/2558	11:01-12:00 น.	ชาย	21	ในเมือง	ผู้ขับขี่	จักรยานยนต์	รถเก๋ง/แท็กซี่	ไม่ใส่	ไม่ดื่ม	ทุเลา/หาย
ปีใหม่ 58	11	สมุทรปราการ	11747	กรุงเทพพระประแดง	02/01/2558	14:01-15:00 น.	ชาย	20	ทางหลวง	ผู้ขับขี่	จักรยานยนต์	ไม่มี/ดัดแปลง	ใส่หมวก	ไม่ดื่ม	ทุเลา/หาย

ภาพที่ 3.4 ข้อมูลผู้บาดเจ็บและเสียชีวิตในช่วงเทศกาลปีใหม่ประจำปี 2551 – 2558

3.1.3.2 ทำการกลั่นกรองข้อมูล (Data Cleaning) คือการทำความสะอาดข้อมูล เป็นกระบวนการตรวจสอบและการแก้ไข (หรือลบ) รายการข้อมูลที่ไม่ถูกต้องออกไปจากชุดข้อมูล ตารางหรือฐานข้อมูล ซึ่งเป็นหลักสำคัญของฐานข้อมูล ทางคณะผู้จัดทำได้ดำเนินการดังนี้

1) ข้อมูลผู้บาดเจ็บและเสียชีวิตในช่วงเทศกาลปีใหม่ประจำปี 2551–2558 คณะผู้จัดทำทำการลบข้อมูลที่ซ้ำซ้อน และแก้ไขข้อมูลที่ผิดพลาด ซึ่งคณะผู้จัดทำพบว่าบางข้อมูลรหัสโรงพยาบาลซ้ำกัน และรหัสโรงพยาบาลกับชื่อโรงพยาบาลไม่ตรงกัน จึงดูข้อมูลตามหลักความเป็นจริงโดยเปรียบเทียบข้อมูลจากส่วนใหญ่ที่ใช้ หรือข้อมูลอื่น ๆ ที่มีข้อผิดพลาด เนื่องด้วยอาจจะมาจากการคีย์ข้อมูลที่ต้นทาง ดังนั้นคณะผู้จัดทำได้ทำการเปลี่ยนข้อมูล หรือแทนที่ข้อมูลชื่อโรงพยาบาลให้ตรงกับรหัสโรงพยาบาล ดังนี้

- คณะผู้จัดทำทำการตรวจสอบข้อมูลที่ซ้ำซ้อนของข้อมูลทั้งหมด พบข้อมูลที่ซ้ำกันจำนวน 1,680 รายการ ทางคณะผู้จัดทำจึงได้ลบข้อมูลนั้นทิ้ง โดยจะมีข้อมูลในปี 2551–2558 ที่ทำการ Cleaning แล้ว จำนวน 213,000 รายการ



ภาพที่ 3.5 การตรวจสอบข้อมูลที่ซ้ำซ้อนของข้อมูลทั้งหมด

- รหัสรพ. 10707 เป็นรหัสของโรงพยาบาลมหาสารคาม คณะผู้จัดทำพบว่ามีการใช้รหัสโรงพยาบาลซ้ำกัน จึงได้ทำการตรวจสอบข้อมูลและทำการแก้ไขให้เป็นชื่อโรงพยาบาลมหาสารคามทั้งหมด

รหัส รพ.	ชื่อโรงพยาบาลที่รับผู้บาดเจ็บ
10707	มหาราชนครราชสีมา
10707	มหาสารคาม
10707	มหาสารคาม
10707	มหาสารคาม
10707	มหาสารคาม

ภาพที่ 3.6 ข้อมูลผิดพลาดจากการใช้รหัสโรงพยาบาลซ้ำกัน

- รหัสรพ. 13780 โดยส่วนใหญ่ใช้กับชื่อโรงพยาบาลมหาราชนครเชียงใหม่ คณะผู้จัดทำพบว่ามีการใช้รหัสโรงพยาบาลมากกว่าหนึ่งรหัส จึงทำการเปรียบเทียบรหัสที่ใช้มากที่สุด และแก้ไขข้อมูลรหัสโรงพยาบาลให้อยู่ในรหัสเดียวกัน

รหัส รพ.	ชื่อโรงพยาบาลที่รับผู้บาดเจ็บ
13780	มหาราชนครเชียงใหม่
13780	มหาราชนครเชียงใหม่
13780	มหาราชนครเชียงใหม่
13780	มหาราชนครเชียงใหม่
24300	มหาราชนครเชียงใหม่
24300	มหาราชนครเชียงใหม่

ภาพที่ 3.7 ข้อมูลผิดพลาดจากการใช้รหัสโรงพยาบาลมากกว่าหนึ่งรหัส

- รหัสรพ. 10666 เป็นรหัสของโรงพยาบาลมหาราชนครราชสีมา คณะผู้จัดทำพบว่ามีการใช้รหัสโรงพยาบาลมากกว่าหนึ่งรหัส จึงทำการเปรียบเทียบรหัสที่ใช้มากที่สุด และทำการแก้ไขเพื่อให้ได้ข้อมูลรหัสโรงพยาบาลและชื่อโรงพยาบาลที่ตรงกัน

รหัส รพ.	ชื่อโรงพยาบาลที่รับผู้บาดเจ็บ
10666	มหาราชนครราชสีมา
10666	มหาราชนครราชสีมา
10666	มหาราชนครราชสีมา
10666	มหาราชนครราชสีมา
10666	มหาราชนครราชสีมา
10666	มหาราชนครราชสีมา
10666	มหาราชนครราชสีมา
10666	มหาราชนครราชสีมา
10666	มหาราชนครราชสีมา
27841	มหาราชนครราชสีมา
27840	มหาราชนครราชสีมา
27840	มหาราชนครราชสีมา
27840	มหาราชนครราชสีมา
27840	มหาราชนครราชสีมา
27840	มหาราชนครราชสีมา
24060	มหาราชนครราชสีมา

ภาพที่ 3.8 ข้อมูลผิดพลาดจากการใช้รหัสโรงพยาบาลมากกว่าหนึ่งรหัส

- รหัสรพ. 11720 เป็นรหัสของโรงพยาบาลเทพธารินทร์ คณะผู้จัดทำพบว่ามีการใช้รหัสโรงพยาบาลมากกว่าหนึ่งรหัส และการใช้ชื่อโรงพยาบาลผิด จึงทำการเปรียบเทียบรหัส และชื่อที่ใช้มากที่สุด และทำการแก้ไขเพื่อให้ได้ข้อมูลรหัสโรงพยาบาลและชื่อโรงพยาบาลที่ตรงกัน

รหัส รพ.	ชื่อโรงพยาบาลที่รับผู้บาดเจ็บ
11720	เทพธารินทร์
11720	เทพธารินทร์
11720	เทพธารินทร์
11720	เทพธารินทร์
11720	เทพธารินทร์
15051	เทพธารินทร

ภาพที่ 3.9 ข้อมูลผิดพลาดของชื่อข้อมูล และรหัสโรงพยาบาลที่ใช้มากกว่าหนึ่งรหัส

- รหัสรพ. 11730 เป็นรหัสของโรงพยาบาลไทยนครินทร์ บางนา คณะผู้จัดทำพบว่ามีการใช้รหัสโรงพยาบาลมากกว่าหนึ่งรหัส และสถานที่โรงพยาบาลเดียวกัน แต่มีการใช้ชื่อโรงพยาบาลไม่เหมือนกัน จึงทำการเปรียบเทียบรหัสที่ใช้มากที่สุด และทำการแก้ไขเพื่อให้ได้ข้อมูลรหัสโรงพยาบาลและชื่อโรงพยาบาลที่ตรงกัน

จังหวัด	รหัส รพ.	ชื่อโรงพยาบาลที่รับผู้บาดเจ็บ
กรุงเทพมหานคร	11730	ไทยนครินทร์ บางนา
กรุงเทพมหานคร	11730	ไทยนครินทร์ บางนา
กรุงเทพมหานคร	11730	ไทยนครินทร์ บางนา
กรุงเทพมหานคร	11730	ไทยนครินทร์ บางนา
กรุงเทพมหานคร	14567	ไทยนรินทร์
กรุงเทพมหานคร	14567	ไทยนรินทร์
กรุงเทพมหานคร	14567	ไทยนรินทร์

ภาพที่ 3.10 ข้อมูลผิดพลาดของชื่อข้อมูล และรหัสโรงพยาบาลที่ใช้มากกว่าหนึ่งรหัส

- รหัสรพ. 10669 เป็นรหัสของโรงพยาบาลสรรพสิทธิประสงค์

คณะผู้จัดทำพบว่ามีการใช้รหัสโรงพยาบาลมากกว่าหนึ่งรหัส จึงทำการเปรียบเทียบรหัสที่ใช้มากที่สุด และทำการแก้ไขเพื่อให้ได้ข้อมูลรหัสโรงพยาบาลและชื่อโรงพยาบาลที่ตรงกัน

รหัส รพ.	ชื่อโรงพยาบาลที่รับผู้บาดเจ็บ
10669	สรรพสิทธิประสงค์
10669	สรรพสิทธิประสงค์
10669	สรรพสิทธิประสงค์
10669	สรรพสิทธิประสงค์
10669	สรรพสิทธิประสงค์
10669	สรรพสิทธิประสงค์
27967	สรรพสิทธิประสงค์
27967	สรรพสิทธิประสงค์
27967	สรรพสิทธิประสงค์

ภาพที่ 3.11 ข้อมูลผิดพลาดจากการใช้รหัสโรงพยาบาลมากกว่าหนึ่งรหัส

- รหัสรพ. 11656 เป็นรหัสของโรงพยาบาลเกษมราษฎร์ บางแค

คณะผู้จัดทำพบว่ามีการใช้รหัสโรงพยาบาลมากกว่าหนึ่งรหัส และมีการใช้ชื่อโรงพยาบาลไม่เหมือนกัน จึงทำการเปรียบเทียบข้อมูลที่ใช้มากที่สุด และทำการแก้ไขเพื่อให้ได้รหัสโรงพยาบาลและชื่อโรงพยาบาลที่ตรงกัน

รหัส รพ.	ชื่อโรงพยาบาลที่รับผู้บาดเจ็บ
14564	เกษมราษฎร์บางแค
14564	เกษมราษฎร์บางแค
14564	เกษมราษฎร์บางแค
14564	เกษมราษฎร์บางแค
11656	เกษมราษฎร์ บางแค
11656	เกษมราษฎร์ บางแค
11656	เกษมราษฎร์ บางแค
11656	เกษมราษฎร์ บางแค
11656	เกษมราษฎร์ บางแค

ภาพที่ 3.12 ข้อมูลผิดพลาดของชื่อข้อมูล และรหัสโรงพยาบาลที่ใช้มากกว่าหนึ่งรหัส

- รหัสรพ. 11262 เป็นรหัสของโรงพยาบาลสามง่าม คณะผู้จัดทำพบว่ามีการใช้รหัสโรงพยาบาลซ้ำกัน จึงได้ทำการตรวจสอบข้อมูลและทำการแก้ไขให้เป็นชื่อโรงพยาบาลสามง่ามทั้งหมด

รหัส รพ.	ชื่อโรงพยาบาลที่รับผู้บาดเจ็บ
11262	สรรพสิทธิประสงค์
11262	สามง่าม
11262	สามง่าม
11262	สามง่าม

ภาพที่ 3.13 ข้อมูลผิดพลาดจากการใช้รหัสโรงพยาบาลซ้ำกัน

- รหัสรพ. 10937 เป็นรหัสของโรงพยาบาลห้วยทับทิม คณะผู้จัดทำพบว่ามีการใช้รหัสโรงพยาบาลและชื่อโรงพยาบาลไม่ถูกต้องในกลุ่มข้อมูล จึงทำการแก้ไขเพื่อให้ได้ข้อมูลรหัสโรงพยาบาลและชื่อโรงพยาบาลที่ตรงกัน

รหัส รพ.	ชื่อโรงพยาบาลที่รับผู้บาดเจ็บ
10931	สรรพสิทธิประสงค์
10937	ห้วยทับทัน
10937	ห้วยทับทัน
10937	ห้วยทับทัน
10937	ห้วยทับทัน

ภาพที่ 3.14 ข้อมูลผิดพลาดของช่วงข้อมูลในโรงพยาบาลห้วยทับทัน

- คณะผู้จัดทำทำการแก้ไขสถานะการขับขี่ที่มีการใช้คำสะกดผิดให้ถูกต้องตามหลักของภาษาไทยจาก “ผู้ซับขี่” เป็น “ผู้ขับขี่”

สถานะ
ผู้ขับขี่
ผู้ขับขี่
ผู้ขับขี่
ผู้ขับขี่
ผู้ขับขี่
ผู้ซับขี่
ผู้ซับขี่
ผู้ซับขี่

ภาพที่ 3.15 ข้อมูลผิดพลาดของผู้ขับขี่

- คณะผู้จัดทำทำการตรวจสอบสถานะของคนเดินเท้า พบว่ามาตรการเป็นการใส่เข็มขัด, ใส่หมวก และมีรถที่ประสบอุบัติเหตุ ซึ่งเป็นข้อมูลที่ขัดแย้งกัน ทางคณะผู้จัดทำจึงได้แก้ไขมาตรการให้เป็นไม่ใส่ และรถผู้บาดเจ็บเป็นอื่น ๆ

สถานะ	รถผู้บาดเจ็บ	รถคู่กรณี	มาตรการ
คนเดินเท้า	อื่นๆ	จักรยานยนต์	เข็มขัด
คนเดินเท้า	อื่นๆ	จักรยานยนต์	เข็มขัด
คนเดินเท้า	อื่นๆ	ปิคอัพ	เข็มขัด
คนเดินเท้า	อื่นๆ	จักรยานยนต์	เข็มขัด
คนเดินเท้า	อื่นๆ	จักรยานยนต์	เข็มขัด
คนเดินเท้า	ปิคอัพ	รถบรรทุก	เข็มขัด
คนเดินเท้า	รถเก๋ง/แท็กซี่	ไม่มี/ล้มเอง	เข็มขัด

ภาพที่ 3.16 ข้อมูลที่ขัดแย้งกันของคนเดินเท้า

2) ข้อมูลผู้บาดเจ็บและเสียชีวิตในช่วงเทศกาลปีใหม่ประจำปี 2551–2558 คณะผู้จัดทำทำการหาค่าเฉลี่ยของกลุ่มข้อมูล ลบข้อมูลที่ไม่ชัดเจนทิ้ง เนื่องจากเป็นข้อมูลที่ขัดแย้ง และจัดการข้อมูลที่เป็น Outlier ที่มีค่าสูงหรือต่ำกว่าความเป็นจริง ดังนี้

- คณะผู้จัดทำทำการลบข้อมูลประเภทอายุ 1 ปีที่มีสถานะเป็นผู้ขับขี่

อายุ	ถนนที่เกิดเหตุ	สถานะ
1	ชนบท	ผู้ขับขี่
1	ชนบท	ผู้ขับขี่
1	ชนบท	ผู้ขับขี่
1	ในเมือง	ผู้ขับขี่
1	ชนบท	ผู้ขับขี่
1	ชนบท	ผู้ขับขี่
1	ชนบท	ผู้ขับขี่
1	ในเมือง	ผู้ขับขี่

ภาพที่ 3.17 การลบข้อมูลที่ขัดแย้งกัน

- คณะผู้จัดทำพบว่ามีข้อมูลที่เป็น Outlier คือ อายุ 99 ปี ซึ่งโดยเฉลี่ยจากสถิติปี 2504–2559 ในประเทศไทยนั้นมีอายุไขเฉลี่ย 75 ปี (อ้างอิงสถิติจากเว็บไซต์ bltbangkok) จึงได้ทำการลบข้อมูลประเภทอายุ 99 ปีจากชุดข้อมูล

อายุ	ถนนที่เกิดเหตุ	สถานะ
99	ชนบท	ผู้ขับขี่
99	ชนบท	ผู้ขับขี่
99	ชนบท	ผู้ขับขี่
99	ชนบท	ผู้ขับขี่
99	ในเมือง	ผู้ขับขี่

ภาพที่ 3.18 ข้อมูลอายุที่เป็น Outlier

- คณะผู้จัดทำตรวจสอบข้อมูลอายุพบว่ามีข้อมูลที่เป็น 0 จึงหาค่าเฉลี่ยของอายุในแต่ละปีว่ามีการเกิดอุบัติเหตุกับบุคคลที่มีอายุในช่วงใดมากที่สุด เพื่อนำค่าเฉลี่ยที่ได้ไปแทนข้อมูลที่เป็น 0

เพศ	อายุ
หญิง	54
หญิง	13
หญิง	0
หญิง	0
ชาย	0
ชาย	81

ภาพที่ 3.19 การหาค่าเฉลี่ยของอายุในแต่ละปี แทนข้อมูลที่เป็น 0

3.1.3.3 แปลงรูปแบบของข้อมูล (Data Transformation) เป็นขั้นตอนการแปลงข้อมูลในรูปแบบตารางฐานข้อมูลให้อยู่ในรูป item set เพื่อใช้สำหรับการวิเคราะห์ด้วยวิธีการของ data mining ทางคณะผู้จัดทำได้ดำเนินการกับข้อมูลผู้บาดเจ็บและเสียชีวิตในช่วงเทศกาลปีใหม่ประจำปี 2558 ดังนี้

1) คณะผู้จัดทำพบข้อมูล “ไม่ทราบ” ในคอลัมน์ของแต่ละประเภท คือ เวลาเกิดเหตุ, ถนนที่เกิดเหตุ, สถานะ, รถผู้บาดเจ็บ, รถคู่กรณี, มาตรการ และการดื่มสุรา จึงได้ทำการแปลงข้อความให้อยู่ในรูปแบบตัวเลขเพื่อหาฐานนิยม หรือตัวเลขที่ซ้ำกันมากที่สุดในช่วงของข้อมูล ซึ่งจะได้ตรงกับข้อความที่มีการเกิดขึ้นมากที่สุด คณะผู้จัดทำจึงนำมาแทนที่ประเภทไม่ทราบของแต่ละกลุ่มข้อมูลทั้งหมด ด้วยฟังก์ชัน MODE

2) คณะผู้จัดทำทำการแปลงรูปแบบด้วยรวมกลุ่มของข้อมูลในแอททริบิวท์ ผลการรักษา เนื่องจากปกติแล้วจะมีข้อมูลในแอททริบิวท์นี้หลากหลาย ได้แก่ ตายในตึกภายใน 24 ชม. หลังเหตุ ตายในตึกหลัง 24 ชม. – 30 วัน ตายที่เกิดเหตุ ตายที่ห้องฉุกเฉิน ตายระหว่างนำส่ง ตายระหว่างส่งต่อ ซึ่งผู้วิจัยจะขอรวมกลุ่มของการตายใน รูปแบบต่าง ๆ ให้เหลือตาย เพียงอย่างเดียว

3.1.4 สร้างแบบจำลอง (Modeling) ขั้นตอนการสร้างตัวแบบทางคณิตศาสตร์ และสถิติเพื่อการวิเคราะห์ข้อมูล โดยสามารถใช้เทคนิควิธีการต่าง ๆ อาทิ การจำแนก (Classification) การแบ่งกลุ่ม (Clustering) และการสร้างความสัมพันธ์ (Association rule)

คณะผู้จัดทำวิเคราะห์ข้อมูลด้วยการทำเหมืองข้อมูลแบบ Data Classification เพื่อใช้ทำนายแนวโน้มการเกิดขึ้นของปัจจัยที่ก่อให้เกิดผู้บาดเจ็บและเสียชีวิตจากเทคนิคการจำแนกประเภทข้อมูล ด้วยการสร้างโมเดล Decision Tree เพื่อจัดกลุ่มข้อมูลที่มีความสัมพันธ์กัน ด้วยชุดข้อมูลที่คัดเลือก ดังนี้

เพศ	สถานะ	มาตรการ	การดื่มสุรา	ผลการรักษา
ชาย	ผู้ขับขี่	ไม่ใส่หมวก	ไม่ดื่ม	ทุเลา/หาย
ชาย	ผู้ขับขี่	ไม่ใส่หมวก	ไม่ดื่ม	ทุเลา/หาย
หญิง	ผู้โดยสาร	ไม่ใส่หมวก	ไม่ดื่ม	ทุเลา/หาย
หญิง	ผู้ขับขี่	ไม่ใส่หมวก	ไม่ดื่ม	ทุเลา/หาย
หญิง	ผู้ขับขี่	ไม่ใส่หมวก	ไม่ดื่ม	ทุเลา/หาย
หญิง	ผู้โดยสาร	ไม่ใส่หมวก	ไม่ดื่ม	ทุเลา/หาย
ชาย	ผู้ขับขี่	ไม่ใส่หมวก	ไม่ดื่ม	ทุเลา/หาย
ชาย	ผู้ขับขี่	ไม่ใส่หมวก	ดื่ม	ทุเลา/หาย
ชาย	ผู้ขับขี่	ไม่ใส่หมวก	ไม่ดื่ม	ทุเลา/หาย
ชาย	ผู้ขับขี่	ไม่ใส่หมวก	ไม่ดื่ม	ทุเลา/หาย
ชาย	ผู้ขับขี่	ไม่ใส่หมวก	ดื่ม	ทุเลา/หาย
หญิง	ผู้ขับขี่	ใส่หมวก	ไม่ดื่ม	ทุเลา/หาย
ชาย	ผู้ขับขี่	ใส่หมวก	ไม่ดื่ม	ทุเลา/หาย
ชาย	ผู้ขับขี่	ใส่หมวก	ไม่ดื่ม	ทุเลา/หาย
ชาย	คนเดินเท้า	ไม่ใส่หมวก	ไม่ดื่ม	ทุเลา/หาย

ภาพที่ 3.20 ชุดข้อมูลที่คัดเลือกมาวิเคราะห์ข้อมูล

จากรูปภาพที่ ประกอบด้วย 5 แอตทริบิวต์ คือ

- เพศ แสดงเพศ ประกอบด้วย 2 ค่า คือ ชาย หญิง
- สถานะ แสดงสถานะของผู้ประสบอุบัติเหตุ ประกอบด้วย 3 ค่า คือ คนเดินเท้า ผู้โดยสาร ผู้ขับขี่
- มาตรการ แสดงมาตรการความปลอดภัยในการขับขี่ ประกอบด้วย 4 ค่า คือ เข็มขัด ไม่ใส่เข็มขัด ใส่หมวก ไม่ใส่หมวก
- การดื่มสุรา แสดงการดื่มสุรา ซึ่งเป็นคลาส ประกอบด้วย 2 ค่า คือ ดื่ม ไม่ดื่ม
- ผลการรักษา แสดงผลการรักษา ซึ่งเป็นคลาส ประกอบด้วย 2 ค่า คือ ทุเลา/หาย ตาย

การสร้างโมเดล decision tree จะทำการคัดเลือกแอตทริบิวต์ที่มีความสัมพันธ์กับคลาสมากที่สุดขึ้นมาเป็นโหนดบนสุดของ tree (root node) หลังจากนั้นก็จะหาแอตทริบิวต์ถัดไปเรื่อย ๆ ในการหาความสัมพันธ์ของแอตทริบิวต์นี้จะใช้ตัววัด ที่เรียกว่า Information Gain (IG) ค่านี้คำนวณได้จากสมการดังนี้

ตารางที่ 3.1 สมการการหาความสัมพันธ์ของแอตทริบิวต์

การคำนวณ	สมการ
Entropy	entropy(c1) = –p(c1) log p(c1) และ p(c2) คือ ค่าความน่าจะเป็นของ c1
Information Gain	IG (parent, child) = entropy(parent) – [p(c1) × entropy(c1) + p(c2) × entropy(c2) + ...]

การคำนวณค่าแต่ละแอตทริบิวต์เทียบกับคลาสเพื่อหาแอตทริบิวต์ที่มีค่า IG มากที่สุดมาเป็น root ของ decision tree กับจำนวนข้อมูลทั้งหมดโดยใช้ผลลัพธ์เป็นดื่ม และไม่ดื่ม ดังนี้

1) คำนวณค่า IG ของแอตทริบิวต์ เพศ จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) $= -p(\text{ดื่ม}) \times \log_2 p(\text{ดื่ม}) + p(\text{ไม่ดื่ม}) \times \log_2 p(\text{ไม่ดื่ม})$

$= -[0.35 \times \log_2(0.35) + 0.65 \times \log_2(0.65)]$

$= -[0.35 \times -1.51 + 0.65 \times -0.62]$

$= -[0.53 + 0.40]$

$= 0.93$

entropy(ผล = ชาย) $= - p(\text{ดื่ม}) \times \log_2 p(\text{ดื่ม}) + p(\text{ไม่ดื่ม}) \times \log_2 p(\text{ไม่ดื่ม})$

$= -[0.48 \times \log_2(0.48) + 0.52 \times \log_2(0.52)]$

$= -[0.48 \times -1.06 + 0.52 \times -0.94]$

$= -[0.51 + 0.49]$

$= 1$

entropy(ผล = หญิง) $= - p(\text{ดื่ม}) \times \log_2 p(\text{ดื่ม}) + p(\text{ไม่ดื่ม}) \times \log_2 p(\text{ไม่ดื่ม})$

$= -[0.09 \times \log_2(0.09) + 0.91 \times \log_2(0.91)]$

$= -[0.09 \times -3.47 + 0.91 \times -0.14]$

$= -[0.31 + 0.13]$

$= 0.44$

IG(parent, child) $= \text{entropy (parent)} - [p(\text{ผล} = \text{ชาย}) \times \text{entropy}(\text{ผล} = \text{ชาย}) + p(\text{ผล} = \text{หญิง}) \times \text{entropy}(\text{ผล} = \text{หญิง})]$

$= 0.93 - [0.68 \times 1 + 0.32 \times 0.44]$

$= 0.93 - [0.68 + 0.14]$

$= 0.93 - 0.82$

$= 0.11$

2) คำนวณค่า IG ของแอตทริบิวต์ สถานะ จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) $= -p(\text{ดื่ม}) \times \log_2 p(\text{ดื่ม}) + p(\text{ไม่ดื่ม}) \times \log_2 p(\text{ไม่ดื่ม})$

$= -[0.35 \times \log_2(0.35) + 0.65 \times \log_2(0.65)]$

$= -[0.35 \times -1.51 + 0.65 \times -0.62]$

$= -[0.53 + 0.40]$

$= 0.93$

entropy(ผล = ผู้โดยสาร) = – p(ดื่ม) × $\log_2$ p(ดื่ม) + p(ไม่ดื่ม) × $\log_2$ p(ไม่ดื่ม)

= –[0.20 × $\log_2$(0.20) + 0.80 × $\log_2$(0.80)]

= –[0.20 × –2.32 + 0.80 × –0.32]

= –[0.46 + 0.26]

= 0.72

entropy(ผล = ผู้ขับขี่) = – p(ดื่ม) × $\log_2$ p(ดื่ม) + p(ไม่ดื่ม) × $\log_2$ p(ไม่ดื่ม)

= –[0.41 × $\log_2$(0.41) + 0.59 × $\log_2$(0.59)]

= –[0.41 × –1.29 + 0.59 × –0.76]

= –[0.53 + 0.45]

= 0.98

entropy(ผล = คนเดินเท้า) = – p(ดื่ม) × $\log_2$ p(ดื่ม) + p(ไม่ดื่ม) × $\log_2$ p(ไม่ดื่ม)

= –[0.18 × $\log_2$(0.18) + 0.82 × $\log_2$(0.82)]

= –[0.18 × –2.47 + 0.82 × –0.29]

= –[0.44 + 0.24]

= 0.68

IG(parent, child) = entropy (parent) – [p(ผล = ผู้โดยสาร) × entropy(ผล= ผู้โดยสาร) + p(ผล = ผู้ขับขี่) × entropy(ผล = ผู้ขับขี่) + p(ผล = คนเดินเท้า) × entropy(ผล = คนเดินเท้า)]

= 0.93 – [0.26 × 0.72 + 0.71 × 0.98 + 0.03 × 0.68]

= 0.93 – [0.19 + 0.7 + 0.02]

= 0.93 – 0.91

= 0.02

3) คำนวณค่า IG ของแอตทริบิวต์ มาตรการ จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = –p(ดื่ม) × $\log_2$ p(ดื่ม) + p(ไม่ดื่ม) × $\log_2$ p(ไม่ดื่ม)

= –[0.35 × $\log_2$(0.35) + 0.65 × $\log_2$(0.65)]

= –[0.35 × –1.51 + 0.65 × –0.62]

= –[0.53 + 0.40]

= 0.93

entropy(ผล = เข็มขัด) = – p(ดื่ม) × $\log_2$ p(ดื่ม) + p(ไม่ดื่ม) × $\log_2$ p(ไม่ดื่ม)

= –[0.22 × $\log_2$(0.22) + 0.78 × $\log_2$(0.78)]

$= -[0.22 \times -2.18 + 0.78 \times -0.36]$

$= -[0.48 + 0.28]$

$= 0.76$

entropy(ผล = ไม่ใส่เข็มขัด) $= -\ p(\text{ดื่ม}) \times \log_2 p(\text{ดื่ม}) + p(\text{ไม่ดื่ม}) \times \log_2 p(\text{ไม่ดื่ม})$

$= -[0.22 \times \log_2(0.22) + 0.78 \times \log_2(0.78)]$

$= -[0.22 \times -2.18 + 0.78 \times -0.36]$

$= -[0.48 + 0.28]$

$= 0.76$

entropy(ผล = ใส่หมวก) $= -\ p(\text{ดื่ม}) \times \log_2 p(\text{ดื่ม}) + p(\text{ไม่ดื่ม}) \times \log_2 p(\text{ไม่ดื่ม})$

$= -[0.23 \times \log_2(0.23) + 0.77 \times \log_2(0.77)]$

$= -[0.22 \times -2.12 + 0.78 \times -0.38]$

$= -[0.47 + 0.3]$

$= 0.77$

entropy(ผล = ไม่ใส่หมวก) $= -\ p(\text{ดื่ม}) \times \log_2 p(\text{ดื่ม}) + p(\text{ไม่ดื่ม}) \times \log_2 p(\text{ไม่ดื่ม})$

$= -[0.39 \times \log_2(0.39) + 0.61 \times \log_2(0.61)]$

$= -[0.39 \times -1.36 + 0.61 \times -0.71]$

$= -[0.53 + 0.43]$

$= 0.96$

IG(parent, child) = entropy (parent) – [p(ผล = เข็มขัด) × entropy(ผล = เข็มขัด) + p(ผล = ไม่ใส่เข็มขัด) × entropy(ผล = ไม่ใส่เข็มขัด) + p(ผล = ใส่หมวก) × entropy(ผล = ใส่หมวก) + p(ผล = ไม่ใส่หมวก) × entropy(ผล = ไม่ใส่หมวก)]

$= 0.93 - [0.02 \times 0.76 + 0.08 \times 0.76 + 0.15 \times 0.77 + 0.75 \times 0.96]$

$= 0.93 - [0.02 + 0.06 + 0.12 + 0.72]$

$= 0.93 - 0.92$

$= 0.01$

4) คำนวณค่า IG ของแอตทริบิวต์ ผลการรักษา จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) $= -p(\text{ดื่ม}) \times \log_2 p(\text{ดื่ม}) + p(\text{ไม่ดื่ม}) \times \log_2 p(\text{ไม่ดื่ม})$

$= -[0.35 \times \log_2(0.35) + 0.65 \times \log_2(0.65)]$

$= -[0.35 \times -1.51 + 0.65 \times -0.62]$

= –[0.53 + 0.40]

= 0.93

entropy(ผล = ทุเลา/หาย) = – p(ดื่ม) × $\log_2$ p(ดื่ม) + p(ไม่ดื่ม) × $\log_2$ p(ไม่ดื่ม)

= –[0.35 × $\log_2$(0.35) + 0.65 × $\log_2$(0.65)]

= –[0.35 × –1.51 + 0.65 × –0.62]

= –[0.53 + 0.40]

= 0.93

entropy(ผล = ตาย) = – p(ดื่ม) × $\log_2$ p(ดื่ม) + p(ไม่ดื่ม) × $\log_2$ p(ไม่ดื่ม)

= –[0.30 × $\log_2$(0.30) + 0.70 × $\log_2$(0.70)]

= –[0.27 × –1.74 + 0.73 × –0.51]

= –[0.47 + 0.37]

= 0.84

IG(parent, child) = entropy (parent) – [p(ผล = ทุเลา/หาย) × entropy(ผล=ทุเลา/หาย) + p(ผล = ตาย) × entropy(ผล = ตาย)]

= 0.93 – [0.98 × 0.93 + 0.02 × 0.84]

= 0.93 – [0.91 + 0.02]

= 0.93 – 0.93

= 0

จากการคำนวณค่า IG ของทุกแอตทริบิวต์พบว่าค่า IG ของแอตทริบิวต์ เพศ มีค่ามากที่สุด (0.11) ดังนั้นจึงเลือกแอตทริบิวต์ เพศ ขึ้นมาเป็นโหนด root และจะต้องทำการแตกกิ่งจาก โหนด root ออกไปจนข้อมูลในแต่ละโหนดมีคลาสคำตอบเดียวกัน และคณะผู้จัดทำพบว่าการคำนวณแอตทริบิวต์ เพศหญิง ไม่สามารถสร้างกิ่งแต่ละโหนดต่อไปได้ เนื่องจากไม่มีความสัมพันธ์กับแอตทริบิวต์ใด จึงสรุปข้อมูลได้เป็นผลลัพธ์ดื่ม และไม่ดื่ม ดังนั้นคณะผู้จัดทำจึงทำการสร้างโหนดในระดับถัดไปของแอตทริบิวต์ เพศชาย

การคำนวณค่าแต่ละแอตทริบิวต์ในระดับที่ 2 ต่อจากโหนด root เพื่อหาค่า IG ที่มากที่สุด ของแอตทริบิวต์เพศชายต่อสถานะ มาตรการ และผลการรักษา กับจำนวนข้อมูลทั้งหมดโดยใช้ผลลัพธ์เป็นดื่ม และไม่ดื่ม ดังนี้

1) คำนวณค่า IG ของแอตทริบิวต์เพศชายต่อสถานะ ได้ดังนี้

entropy (parent) = –p(ดื่ม) × $\log_2$ p(ดื่ม) + p(ไม่ดื่ม) × $\log_2$ p(ไม่ดื่ม)

= –[0.48 × $\log_2$(0.48) + 0.52 × $\log_2$(0.52)]

= –[0.48 × –1.06 + 0.52 × –0.94]

= –[0.51 + 0.49]

= 1

entropy(ผล = ผู้โดยสาร) = – p(ดื่ม) × $\log_2$ p(ดื่ม) + p(ไม่ดื่ม) × $\log_2$ p(ไม่ดื่ม)

= –[0.35 × $\log_2$(0.35) + 0.65 × $\log_2$(0.65)]

= –[0.35 × –1.51 + 0.65 × –0.62]

= –[0.53 + 0.40]

= 0.93

entropy(ผล = ผู้ขับขี่) = – p(ดื่ม) × $\log_2$ p(ดื่ม) + p(ไม่ดื่ม) × $\log_2$ p(ไม่ดื่ม)

= –[0.51 × $\log_2$(0.51) + 0.49 × $\log_2$(0.49)]

= –[0.51 × –0.97 + 0.49 × –1.03]

= –[0.49 + 0.50]

= 0.99

entropy(ผล = คนเดินเท้า) = – p(ดื่ม) × $\log_2$ p(ดื่ม) + p(ไม่ดื่ม) × $\log_2$ p(ไม่ดื่ม)

= –[0.25 × $\log_2$(0.25) + 0.75 × $\log_2$(0.75)]

= –[0.25 × –2 + 0.75 × –0.42]

= –[0.5 + 0.32]

= 0.82

IG(parent, child) = entropy (parent) – [p(ผล = ผู้โดยสาร) × entropy(ผล=ผู้โดยสาร) + p(ผล = ผู้ขับขี่) × entropy(ผล = ผู้ขับขี่) + p(ผล = คนเดินเท้า) × entropy(ผล = คนเดินเท้า)]

= 1 – [0.18 × 0.93 + 0.79 × 0.99 + 0.03 × 0.82]

= 1 – [0.17 + 0.78 + 0.02]

= 1 – 0.97

= 0.03

2) คำนวณค่า IG ของแอตทริบิวต์เพศชายต่อมาตรการ ได้ดังนี้

entropy (parent) = –p(ดื่ม) × $\log_2$ p(ดื่ม) + p(ไม่ดื่ม) × $\log_2$ p(ไม่ดื่ม)

= –[0.48 × $\log_2$(0.48) + 0.52 × $\log_2$(0.52)]

= –[0.48 × –1.06 + 0.52 × –0.94]

= 1

entropy(ผล = เข็มขัด) = – p(ดื่ม) × $\log_2$ p(ดื่ม) + p(ไม่ดื่ม) × $\log_2$ p(ไม่ดื่ม)

= –[0.30 × $\log_2$(0.30) + 0.70 × $\log_2$(0.70)]

= –[0.30 × –1.74 + 0.70 × –0.51]

= –[0.52 + 0.36]

= 0.88

entropy(ผล = ไม่ใส่เข็มขัด) = – p(ดื่ม) × $\log_2$ p(ดื่ม) + p(ไม่ดื่ม) × $\log_2$ p(ไม่ดื่ม)

= –[0.34 × $\log_2$(0.34) + 0.66 × $\log_2$(0.66)]

= –[0.34 × –1.56 + 0.66 × –0.6]

= –[0.53 + 0.39]

= 0.92

entropy(ผล = ใส่หมวก) = – p(ดื่ม) × $\log_2$ p(ดื่ม) + p(ไม่ดื่ม) × $\log_2$ p(ไม่ดื่ม)

= –[0.33 × $\log_2$(0.33) + 0.67 × $\log_2$(0.67)]

= –[0.33 × –1.6 + 0.67 × –0.58]

= –[0.53 + 0.39]

= 0.92

entropy(ผล = ไม่ใส่หมวก) = – p(ดื่ม) × $\log_2$ p(ดื่ม) + p(ไม่ดื่ม) × $\log_2$ p(ไม่ดื่ม)

= –[0.52 × $\log_2$(0.52) + 0.48 × $\log_2$(0.48)]

= –[0.52 × –0.94 + 0.48 × –1.06]

= –[0.49 + 0.51]

= 1

IG(parent, child) = entropy (parent) – [p(ผล = เข็มขัด) × entropy(ผล = เข็มขัด) + p(ผล = ไม่ใส่เข็มขัด) × entropy(ผล = ไม่ใส่เข็มขัด) + p(ผล = ใส่หมวก) × entropy(ผล = ใส่หมวก) + p(ผล = ไม่ใส่หมวก) × entropy(ผล = ไม่ใส่หมวก)]

= 1 – [0.02 × 0.88 + 0.07 × 0.92 + 0.15 × 0.92 + 0.76 × 1]

= 1 – [0.02 + 0.06 + 0.14 + 0.76]

= 1 – 0.98

= 0.02

3) คำนวณค่า IG ของแอตทริบิวต์เพศชายต่อผลการรักษา ได้ดังนี้

entropy (parent) = –p(ดื่ม) × $\log_2$ p(ดื่ม) + p(ไม่ดื่ม) × $\log_2$ p(ไม่ดื่ม)

= –[0.48 × $\log_2$(0.48) + 0.52 × $\log_2$(0.52)]

= –[0.48 × –1.06 + 0.52 × –0.94]

= –[0.51 + 0.49]

= 1

entropy(ผล = ทุเลา/หาย) = – p(ดื่ม) × $\log_2$ p(ดื่ม) + p(ไม่ดื่ม) × $\log_2$ p(ไม่ดื่ม)

= –[0.48 × $\log_2$(0.48) + 0.52 × $\log_2$(0.52)]

= –[0.48 × –1.06 + 0.52 × –0.94]

= –[0.51 + 0.49]

= 1

entropy(ผล = ตาย) = – p(ดื่ม) × $\log_2$ p(ดื่ม) + p(ไม่ดื่ม) × $\log_2$ p(ไม่ดื่ม)

= –[0.35 × $\log_2$(0.35) + 0.65 × $\log_2$(0.65)]

= –[0.35 × –1.51 + 0.65 × –0.62]

= –[0.53 + 0.40]

= 0.93

IG(parent, child) = entropy (parent) – [p(ผล = ทุเลา/หาย) × entropy(ผล=ทุเลา/หาย) + p(ผล = ตาย) × entropy(ผล = ตาย)]

= 1 – [0.98 × 1 + 0.02 × 0.93]

= 1 – [0.44 + 0.0036]

= 1 – 1

= 0

จากการคำนวณค่า IG ของแอตทริบิวต์เพศชายต่อสถานะ มาตรการ และผลการรักษา พบว่าค่า IG ของแอตทริบิวต์ เพศชายต่อสถานะ มีค่ามากที่สุด (0.3) ดังนั้นจึงเลือกแอตทริบิวต์ สถานะ ขึ้นมาเป็นโหนดในระดับที่ 2 ต่อจากโหนด root คณะผู้จัดทำจึงทำการสร้างโหนดในระดับถัดไปของแอตทริบิวต์ เพศชายและสถานะ ออกไปจนข้อมูลในแต่ละโหนดมีคลาสคำตอบเดียวกัน

การคำนวณค่าแต่ละแอตทริบิวต์ในระดับที่ 3 ต่อจากโหนดระดับที่ 2 เพื่อหาค่า IG ที่มากที่สุด ของแอตทริบิวต์เพศชายและสถานะที่ประกอบด้วยผู้ขับขี่ ผู้โดยสาร และคนเดินเท้าต่อมาตรการ และผลการรักษา กับจำนวนข้อมูลทั้งหมดโดยใช้ผลลัพธ์เป็นดื่ม และไม่ดื่ม ดังนี้

1) คำนวณค่า IG ของแอตทริบิวต์เพศชาย และสถานะ “ผู้ขับขี่” ต่อมาตรการ ได้ดังนี้

entropy (parent) = –p(ดื่ม) × $\log_2$ p(ดื่ม) + p(ไม่ดื่ม) × $\log_2$ p(ไม่ดื่ม)

= –[0.51 × $\log_2$(0.51) + 0.49 × $\log_2$(0.49)]

$= -[0.51 \times -0.97 + 0.49 \times -1.03]$

$= -[0.49 + 0.50]$

$= 0.99$

entropy(ผล = เข็มขัด) $= -\ p(\text{ดื่ม}) \times \log_2 p(\text{ดื่ม}) + p(\text{ไม่ดื่ม}) \times \log_2 p(\text{ไม่ดื่ม})$

$= -[0.32 \times \log_2(0.32) + 0.68 \times \log_2(0.68)]$

$= -[0.32 \times -1.64 + 0.68 \times -0.56]$

$= -[0.52 + 0.38]$

$= 0.90$

entropy(ผล = ไม่ใส่เข็มขัด) $= -\ p(\text{ดื่ม}) \times \log_2 p(\text{ดื่ม}) + p(\text{ไม่ดื่ม}) \times \log_2 p(\text{ไม่ดื่ม})$

$= -[0.49 \times \log_2(0.49) + 0.51 \times \log_2(0.51)]$

$= -[0.49 \times -1.03 + 0.51 \times -0.97]$

$= -[0.50 + 0.49]$

$= 0.99$

entropy(ผล = ใส่หมวก) $= -\ p(\text{ดื่ม}) \times \log_2 p(\text{ดื่ม}) + p(\text{ไม่ดื่ม}) \times \log_2 p(\text{ไม่ดื่ม})$

$= -[0.33 \times \log_2(0.33) + 0.67 \times \log_2(0.67)]$

$= -[0.33 \times -1.6 + 0.67 \times -0.58]$

$= -[0.53 + 0.39]$

$= 0.92$

entropy(ผล = ไม่ใส่หมวก) $= -\ p(\text{ดื่ม}) \times \log_2 p(\text{ดื่ม}) + p(\text{ไม่ดื่ม}) \times \log_2 p(\text{ไม่ดื่ม})$

$= -[0.56 \times \log_2(0.56) + 0.44 \times \log_2(0.44)]$

$= -[0.56 \times -0.84 + 0.44 \times -1.18]$

$= -[0.47 + 0.51]$

$= 0.98$

IG(parent, child) = entropy (parent) – [p(ผล = เข็มขัด) × entropy(ผล = เข็มขัด) + p(ผล = ไม่ใส่เข็มขัด) × entropy(ผล = ไม่ใส่เข็มขัด) + p(ผล = ใส่หมวก) × entropy(ผล = ใส่หมวก) + p(ผล = ไม่ใส่หมวก) × entropy(ผล = ไม่ใส่หมวก)]

$= 0.99 - [0.02 \times 0.90 + 0.03 \times 0.99 + 0.17 \times 0.92 + 0.77 \times 0.98]$

$= 0.99 - [0.01 + 0.02 + 0.15 + 0.75]$

$= 0.06$

1.1) คำนวณค่า IG ของเพศชาย และสถานะ “ผู้ขับขี่” ต่อผลการรักษา ได้ดังนี้

entropy (parent) = $-p(\text{ดื่ม}) \times \log_2 p(\text{ดื่ม}) + p(\text{ไม่ดื่ม}) \times \log_2 p(\text{ไม่ดื่ม})$

$= -[0.51 \times \log_2(0.51) + 0.49 \times \log_2(0.49)]$

$= -[0.51 \times -0.97 + 0.49 \times -1.03]$

$= -[0.49 + 0.50]$

$= 0.99$

entropy(ผล = ทุเลา/หาย) = $- p(\text{ดื่ม}) \times \log_2 p(\text{ดื่ม}) + p(\text{ไม่ดื่ม}) \times \log_2 p(\text{ไม่ดื่ม})$

$= -[0.52 \times \log_2(0.52) + 0.48 \times \log_2(0.48)]$

$= -[0.52 \times -0.94 + 0.48 \times -1.06]$

$= -[0.48 + 0.50]$

$= 0.98$

entropy(ผล = ตาย) = $- p(\text{ดื่ม}) \times \log_2 p(\text{ดื่ม}) + p(\text{ไม่ดื่ม}) \times \log_2 p(\text{ไม่ดื่ม})$

$= -[0.36 \times \log_2(0.36) + 0.64 \times \log_2(0.64)]$

$= -[0.36 \times -1.47 + 0.64 \times -0.64]$

$= -[0.52 + 0.40]$

$= 0.92$

IG(parent, child) = entropy (parent) – [p(ผล = ทุเลา/หาย) × entropy(ผล = ทุเลา/หาย) + p(ผล = ตาย) × entropy(ผล = ตาย)]

$= 0.99 - [0.98 \times 0.98 + 0.02 \times 0.92]$

$= 0.99 - [0.96 + 0.01]$

$= 0.99 - 0.97$

$= 0.02$

2) คำนวณค่า IG ของแอตทริบิวต์เพศชาย และสถานะ “ผู้โดยสาร” ต่อมาตรการ ได้ดังนี้

entropy (parent) = $-p(\text{ดื่ม}) \times \log_2 p(\text{ดื่ม}) + p(\text{ไม่ดื่ม}) \times \log_2 p(\text{ไม่ดื่ม})$

$= -[0.35 \times \log_2(0.35) + 0.65 \times \log_2(0.65)]$

$= -[0.35 \times -1.51 + 0.65 \times -0.62]$

$= -[0.53 + 0.40]$

$= 0.93$

entropy(ผล = เข็มขัด) = $- p(\text{ดื่ม}) \times \log_2 p(\text{ดื่ม}) + p(\text{ไม่ดื่ม}) \times \log_2 p(\text{ไม่ดื่ม})$

$= -[0.24 \times \log_2(0.24) + 0.76 \times \log_2(0.76)]$

$= -[0.24 \times -2.06 + 0.76 \times -0.4]$

$= -[0.49 + 0.30]$

$= 0.79$

entropy(ผล = ไม่ใส่เข็มขัด) $= -\ p(\text{ดื่ม}) \times \log_2 p(\text{ดื่ม}) + p(\text{ไม่ดื่ม}) \times \log_2 p(\text{ไม่ดื่ม})$

$= -[0.24 \times \log_2(0.24) + 0.76 \times \log_2(0.76)]$

$= -[0.24 \times -2.06 + 0.76 \times -0.4]$

$= -[0.49 + 0.30]$

$= 0.79$

entropy(ผล = ใส่หมวก) $= -\ p(\text{ดื่ม}) \times \log_2 p(\text{ดื่ม}) + p(\text{ไม่ดื่ม}) \times \log_2 p(\text{ไม่ดื่ม})$

$= -[0.32 \times \log_2(0.32) + 0.68 \times \log_2(0.68)]$

$= -[0.32 \times -1.64 + 0.68 \times -0.56]$

$= -[0.52 + 0.38]$

$= 0.90$

entropy(ผล = ไม่ใส่หมวก) $= -\ p(\text{ดื่ม}) \times \log_2 p(\text{ดื่ม}) + p(\text{ไม่ดื่ม}) \times \log_2 p(\text{ไม่ดื่ม})$

$= -[0.39 \times \log_2(0.39) + 0.61 \times \log_2(0.61)]$

$= -[0.39 \times -1.36 + 0.61 \times -0.71]$

$= -[0.53 + 0.43]$

$= 0.96$

IG(parent, child) = entropy (parent) – [p(ผล = เข็มขัด) × entropy(ผล =เข็มขัด) + p(ผล = ไม่ใส่เข็มขัด) × entropy(ผล = ไม่ใส่เข็มขัด) + p(ผล = ใส่หมวก) × entropy(ผล = ใส่หมวก) + p(ผล = ไม่ใส่หมวก) × entropy(ผล = ไม่ใส่หมวก)]

$= 0.93 - [0.03 \times 0.79 + 0.22 \times 0.79 + 0.06 \times 0.90 + 0.70 \times 0.96]$

$= 0.93 - [0.02 + 0.17 + 0.05 + 0.67]$

$= 0.93 - 0.91$

$= 0.02$

2.1) คำนวณค่า IG ของเพศชาย และสถานะ “ผู้โดยสาร” ต่อผลการรักษา ได้ดังนี้

entropy (parent) $= -p(\text{ดื่ม}) \times \log_2 p(\text{ดื่ม}) + p(\text{ไม่ดื่ม}) \times \log_2 p(\text{ไม่ดื่ม})$

$= -[0.35 \times \log_2(0.35) + 0.65 \times \log_2(0.65)]$

$= -[0.35 \times -1.51 + 0.65 \times -0.62]$

$= -[0.53 + 0.40]$

$= 0.93$

entropy(ผล = ทุเลา/หาย) $= -\ p(\text{ดื่ม}) \times \log_2 p(\text{ดื่ม}) + p(\text{ไม่ดื่ม}) \times \log_2 p(\text{ไม่ดื่ม})$

$= -[0.35 \times \log_2(0.35) + 0.65 \times \log_2(0.65)]$

$= -[0.35 \times -1.51 + 0.65 \times -0.62]$

$= -[0.53 + 0.40]$

$= 0.93$

entropy(ผล = ตาย) $= -\ p(\text{ดื่ม}) \times \log_2 p(\text{ดื่ม}) + p(\text{ไม่ดื่ม}) \times \log_2 p(\text{ไม่ดื่ม})$

$= -[0.34 \times \log_2(0.34) + 0.66 \times \log_2(0.66)]$

$= -[0.34 \times -1.56 + 0.66 \times -0.6]$

$= -[0.53 + 039]$

$= 0.92$

IG(parent, child) = entropy (parent) – [p(ผล = ทุเลา/หาย) × entropy (ผล = ทุเลา/หาย) + p(ผล = ตาย) × entropy(ผล = ตาย)]

$= 0.93 - [0.99 \times 0.93 + 0.01 \times 0.92]$

$= 0.93 - [0.92 + 0.00]$

$= 0.93 - 0.92$

$= 0.01$

3) คำนวณค่า IG ของแอตทริบิวต์เพศชาย และสถานะ “คนเดินเท้า” ต่อมาตรการ ได้ดังนี้

entropy (parent) $= -p(\text{ดื่ม}) \times \log_2 p(\text{ดื่ม}) + p(\text{ไม่ดื่ม}) \times \log_2 p(\text{ไม่ดื่ม})$

$= -[0.25 \times \log_2(0.25) + 0.75 \times \log_2(0.75)]$

$= -[0.25 \times -2 + 0.75 \times -0.42]$

$= -[0.50 + 0.31]$

$= 0.81$

entropy(ผล = เข็มขัด) $= -\ p(\text{ดื่ม}) \times \log_2 p(\text{ดื่ม}) + p(\text{ไม่ดื่ม}) \times \log_2 p(\text{ไม่ดื่ม})$

$= -[0 \times \log_2(0) + 0 \times \log_2(0)]$

$= -[0 \times 1 + 0 \times 1]$

$= -[0 + 0]$

$= 0$

entropy(ผล = ไม่ใส่เข็มขัด) = – p(ดื่ม) × $\log_2$ p(ดื่ม) + p(ไม่ดื่ม) × $\log_2$ p(ไม่ดื่ม)

= –[0 × $\log_2$(0) + 0 × $\log_2$(0)]

= –[0 × 1 + 0 × 1]

= –[0 + 0]

= 0

entropy(ผล = ใส่หมวก) = – p(ดื่ม) × $\log_2$ p(ดื่ม) + p(ไม่ดื่ม) × $\log_2$ p(ไม่ดื่ม)

= –[0 × $\log_2$(0) + 0 × $\log_2$(0)]

= –[0 × 1 + 0 × 1]

= –[0 + 0]

= 0

entropy(ผล = ไม่ใส่หมวก) = – p(ดื่ม) × $\log_2$ p(ดื่ม) + p(ไม่ดื่ม) × $\log_2$ p(ไม่ดื่ม)

= –[0.25 × $\log_2$(0.25) + 0.75 × $\log_2$(0.75)]

= –[0.25 × –2 + 0.75 × –0.42]

= –[0.50 + 0.31]

= 0.81

IG(parent, child) = entropy (parent) – [p(ผล = เข็มขัด) × entropy(ผล = เข็มขัด) + p(ผล = ไม่ใส่เข็มขัด) × entropy(ผล = ไม่ใส่เข็มขัด) + p(ผล = ใส่หมวก) × entropy(ผล = ใส่หมวก) + p(ผล = ไม่ใส่หมวก) × entropy(ผล = ไม่ใส่หมวก)]

= 0.81 – [0 × 0 + 0 × 0 + 0 × 0 + 1 × 0.81]

= 0.81 – [0 + 0 + 0 + 0.81]

= 0.81 – 0.81

= 0

3.1) คำนวณค่า IG ของเพศชาย และสถานะ “คนเดินเท้า” ต่อผลการรักษา ได้ดังนี้

entropy (parent) = –p(ดื่ม) × $\log_2$ p(ดื่ม) + p(ไม่ดื่ม) × $\log_2$ p(ไม่ดื่ม)

= –[0.25 × $\log_2$(0.25) + 0.75 × $\log_2$(0.75)]

= –[0.25 × –2 + 0.75 × –0.42]

= –[0.50 + 0.31]

= 0.81

entropy(ผล = ทุเลา/หาย) = – p(ดื่ม) × $\log_2$ p(ดื่ม) + p(ไม่ดื่ม) × $\log_2$ p(ไม่ดื่ม)

= –[0.25 × $\log_2$(0.25) + 0.75 × $\log_2$(0.75)]

$= -[0.25 \times -2 + 0.75 \times -0.42]$

$= -[0.50 + 0.31]$

$= 0.81$

entropy(ผล = ตาย) $= -$ p(ดื่ม) $\times \log_2$ p(ดื่ม) + p(ไม่ดื่ม) $\times \log_2$ p(ไม่ดื่ม)

$= -[0.27 \times \log_2(0.27) + 0.73 \times \log_2(0.73)]$

$= -[0.27 \times -1.89 + 0.73 \times -0.45]$

$= -[0.51 + 0.32]$

$= 0.83$

IG(parent, child) = entropy (parent) – [p(ผล = ทุเลา/หาย) × entropy(ผล = ทุเลา/หาย) + p(ผล = ตาย) × entropy(ผล = ตาย)]

$= 0.81 - [0.95 \times 0.81 + 0.05 \times 0.83]$

$= 0.81 - [0.76 + 0.04]$

$= 0.81 - 0.80$

$= 0.01$

จากการคำนวณค่า IG ของแอตทริบิวต์เพศชาย และสถานะ ต่อมาตรการ และผลการรักษา พบว่าค่า IG ของแอตทริบิวต์เพศชาย และสถานะ “ผู้ขับขี่” ต่อมาตรการ มีค่ามากที่สุด (0.6) ดังนั้นจึงเลือกแอตทริบิวต์มาตรการของสถานะ “ผู้ขับขี่” ขึ้นมาเป็นโหนดในระดับที่ 3 และทำการแตกกิ่งจากโหนดในระดับที่ 3 ออกไปจนข้อมูลในแต่ละโหนดมีคลาสคำตอบเดียวกัน

การคำนวณค่าแต่ละแอตทริบิวต์ในระดับที่ 4 ต่อจากโหนดระดับที่ 3 เพื่อหาค่า IG ที่มากที่สุด ของแอตทริบิวต์เพศชายและสถานะ “ผู้ขับขี่” และมาตรการต่อผลการรักษา กับจำนวนข้อมูลทั้งหมดโดยใช้ผลลัพธ์เป็นดื่ม และไม่ดื่ม ดังนี้

1) คำนวณค่า IG ของแอตทริบิวต์ ชายและสถานะ “ผู้ขับขี่” และมาตรการ “เข็มขัด” ต่อผลการรักษา จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) $= -$p(ดื่ม) $\times \log_2$ p(ดื่ม) + p(ไม่ดื่ม) $\times \log_2$ p(ไม่ดื่ม)

$= -[0.32 \times \log_2(0.32) + 0.68 \times \log_2(0.68)]$

$= -[0.32 \times -1.64 + 0.68 \times -0.56]$

$= 0.91$

entropy(ผล = ทุเลา/หาย) $= -$ p(ดื่ม) $\times \log_2$ p(ดื่ม) + p(ไม่ดื่ม) $\times \log_2$ p(ไม่ดื่ม)

$= -[0.32 \times \log_2(0.32) + 0.68 \times \log_2(0.68)]$

$= -[0.32 \times -1.64 + 0.68 \times -0.56]$

$= -[0.53 + 0.38]$

$= 0.91$

entropy(ผล = ตาย) $= -\ p(\text{ดื่ม}) \times \log_2 p(\text{ดื่ม}) + p(\text{ไม่ดื่ม}) \times \log_2 p(\text{ไม่ดื่ม})$

$= -[0.22 \times \log_2(0.22) + 0.78 \times \log_2(0.78)]$

$= -[0.22 \times -2.18 + 0.78 \times -0.36]$

$= -[0.48 + 0.28]$

$= 0.76$

IG(parent, child) = entropy (parent) – [p(ผล = ทุเลา/หาย) × entropy(ผล = ทุเลา/หาย) + p(ผล = ตาย) × entropy(ผล = ตาย)]

$= 0.91 - [0.98 \times 0.91 + 0.02 \times 0.76]$

$= 0.91 - [0.89 + 0.02]$

$= 0.91 - 0.91$

$= 0$

2) คำนวณค่า IG ของแอตทริบิวต์ ชายและสถานะ “ผู้ขับขี่” และมาตรการ “ไม่ใส่เข็มขัด” ต่อผลการรักษา จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) $= -p(\text{ดื่ม}) \times \log_2 p(\text{ดื่ม}) + p(\text{ไม่ดื่ม}) \times \log_2 p(\text{ไม่ดื่ม})$

$= -[0.49 \times \log_2(0.49) + 0.51 \times \log_2(0.51)]$

$= -[0.49 \times -1.03 + 0.51 \times -0.97]$

$= -[0.50 + 0.49]$

$= 0.99$

entropy(ผล = ทุเลา/หาย) $= -\ p(\text{ดื่ม}) \times \log_2 p(\text{ดื่ม}) + p(\text{ไม่ดื่ม}) \times \log_2 p(\text{ไม่ดื่ม})$

$= -[0.51 \times \log_2(0.51) + 0.49 \times \log_2(0.49)]$

$= -[0.51 \times -0.97 + 0.49 \times -1.03]$

$= -[0.49 + 0.50]$

$= 0.99$

entropy(ผล = ตาย) $= -\ p(\text{ดื่ม}) \times \log_2 p(\text{ดื่ม}) + p(\text{ไม่ดื่ม}) \times \log_2 p(\text{ไม่ดื่ม})$

$= -[0.21 \times \log_2(0.21) + 0.79 \times \log_2(0.79)]$

$= -[0.21 \times -2.25 + 0.79 \times -0.34]$

$= 0.73$

IG(parent, child) = entropy (parent) – [p(ผล = ทุเลา/หาย) × entropy(ผล=ทุเลา/หาย) + p(ผล = ตาย) × entropy(ผล = ตาย)]

= 0.99 – [0.94 × 0.99 + 0.06 × 0.73]

= 0.99 – [0.93 + 0.04]

= 0.99 – 0.97

= 0.02

3) คำนวณค่า IG ของแอตทริบิวต์ ชายและสถานะ “ผู้ขับขี่” และมาตรการ “ใส่หมวก” ต่อผลการรักษา จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = –p(ดื่ม) × $\log_2$ p(ดื่ม) + p(ไม่ดื่ม) × $\log_2$ p(ไม่ดื่ม)

= –[0.33 × $\log_2$(0.33) + 0.67 × $\log_2$(0.67)]

= –[0.33 × –1.6 + 0.67 × –0.58]

= –[0.53 + 0.39]

= 0.92

entropy(ผล = ทุเลา/หาย) = – p(ดื่ม) × $\log_2$ p(ดื่ม) + p(ไม่ดื่ม) × $\log_2$ p(ไม่ดื่ม)

= –[0.33 × $\log_2$(0.33) + 0.67 × $\log_2$(0.67)]

= –[0.33 × –1.6 + 0.67 × –0.58]

= –[0.53 + 0.39]

= 0.92

entropy(ผล = ตาย) = – p(ดื่ม) × $\log_2$ p(ดื่ม) + p(ไม่ดื่ม) × $\log_2$ p(ไม่ดื่ม)

= –[0.30 × $\log_2$(0.30) + 0.70 × $\log_2$(0.70)]

= –[0.30 × –1.75 + 0.70 × –0.51]

= –[0.53 + 0.36]

= 0.89

IG(parent, child) = entropy (parent) – [p(ผล = ทุเลา/หาย) × entropy ผล = ทุเลา/หาย) + p(ผล = ตาย) × entropy(ผล = ตาย)]

= 0.92 – [0.99 × 0.92 + 0.01 × 0.89]

= 0.92 – [0.91 + 0.01]

= 0.92 – 0.92

= 0

4) คำนวณค่า IG ของแอตทริบิวต์ ชายและสถานะ “ผู้ขับขี่” และมาตรการ “ไม่ใส่หมวก” ต่อผลการรักษา จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = $-p(\text{ดื่ม}) \times \log_2 p(\text{ดื่ม}) + p(\text{ไม่ดื่ม}) \times \log_2 p(\text{ไม่ดื่ม})$

$= -[0.56 \times \log_2(0.56) + 0.44 \times \log_2(0.44)]$

$= -[0.56 \times -0.84 + 0.44 \times -1.18]$

$= -[0.47 + 0.51]$

$= 0.98$

entropy(ผล = ทุเลา/หาย) = $- p(\text{ดื่ม}) \times \log_2 p(\text{ดื่ม}) + p(\text{ไม่ดื่ม}) \times \log_2 p(\text{ไม่ดื่ม})$

$= -[0.57 \times \log_2(0.57) + 0.43 \times \log_2(0.43)]$

$= -[0.57 \times -0.81 + 0.43 \times -1.22]$

$= -[0.46 + 0.52]$

$= 0.98$

entropy(ผล = ตาย) = $- p(\text{ดื่ม}) \times \log_2 p(\text{ดื่ม}) + p(\text{ไม่ดื่ม}) \times \log_2 p(\text{ไม่ดื่ม})$

$= -[0.39 \times \log_2(0.39) + 0.61 \times \log_2(0.61)]$

$= -[0.39 \times -1.36 + 0.61 \times -0.71]$

$= -[0.53 + 0.43]$

$= 0.96$

IG(parent, child) = entropy (parent) – [p(ผล = ทุเลา/หาย) × entropy(ผล=ทุเลา/หาย) + p(ผล = ตาย) × entropy(ผล = ตาย)]

$= 0.98 - [0.98 \times 0.98 + 0.02 \times 0.96]$

$= 0.98 - [0.96 + 0.01]$

$= 0.98 - 0.97$

$= 0.01$

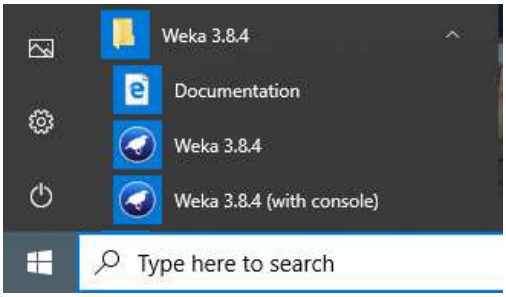
จากการคำนวณค่า IG ของแอตทริบิวต์เพศชายและสถานะ “ผู้ขับขี่” และมาตรการต่อผลการรักษาที่เป็นแอตทริบิวต์สุดท้าย คณะผู้จัดทำพบว่ามาตรการ “ไม่ใส่เข็มขัด” และ “ไม่ใส่หมวก” มีความสัมพันธ์กับผลการรักษามากที่สุดอยู่ที่ 0.02 และ 0.01 ตามลำดับ และพบว่าข้อมูลในแต่ละโหนดมีคลาสคำตอบเดียวกันแล้ว คือ ผลลัพธ์เป็นดื่ม และไม่ดื่ม

3.1.5 การประเมินผล (Evaluation) เป็นขั้นตอนก่อนนำผลลัพธ์ที่ได้จากขั้นตอนที่ 4 ไปใช้งาน ด้วยการวัดประสิทธิผลของผลลัพธ์ที่ได้กับวัตถุประสงค์ที่ตั้งไว้ในขั้นตอนแรก ว่ามี

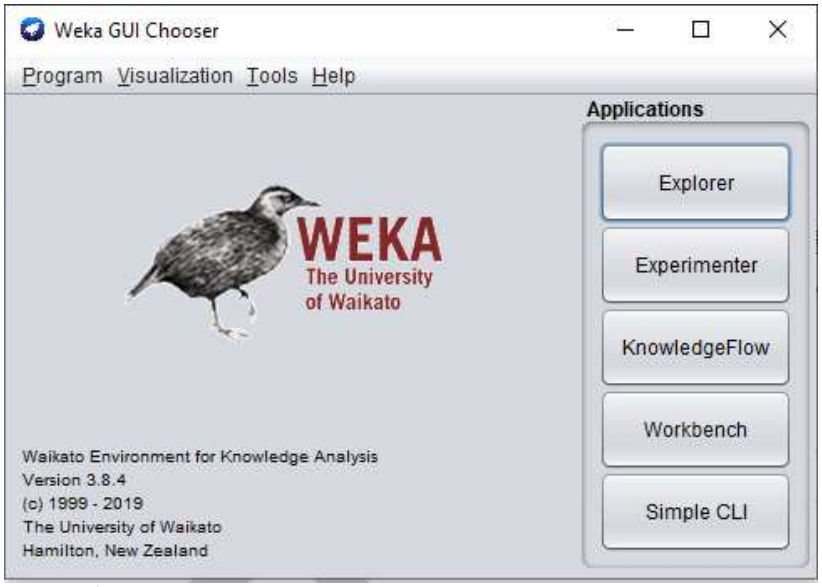
นัยสำคัญหรือความน่าเชื่อถือมากน้อยเพียงใด ด้วยการประเมินผลจากโปรแกรมว่าถูกต้องหรือไม่

คณะผู้จัดทำได้ทำการทดสอบโมเดล เพื่อวัดประสิทธิภาพที่ตรงกับความต้องการ ซึ่งการวัดประสิทธิภาพด้วยวิธี Self Consistency Test เหมาะสำหรับใช้ในการทดสอบประสิทธิภาพ เพื่อดูแนวโน้มของโมเดลที่สร้างขึ้น และเมื่อนำข้อมูลมาทดสอบ (testing data) กับโปรแกรมที่ผู้วิเคราะห์เลือก มาทดสอบกับข้อมูลที่ผ่านการวิเคราะห์ข้อมูลด้วยเทคนิค Data Mining จากการสร้างโมเดล Decision Tree จึงนำข้อมูลดังกล่าว มาทดสอบกับโปรแกรม Weka เวอร์ชัน 3.8.4 ซึ่งมีขั้นตอนการทำงาน ดังนี้

ขั้นตอนที่ 1 คลิกเปิดโปรแกรม Weka 3.8.4 ขึ้นมา

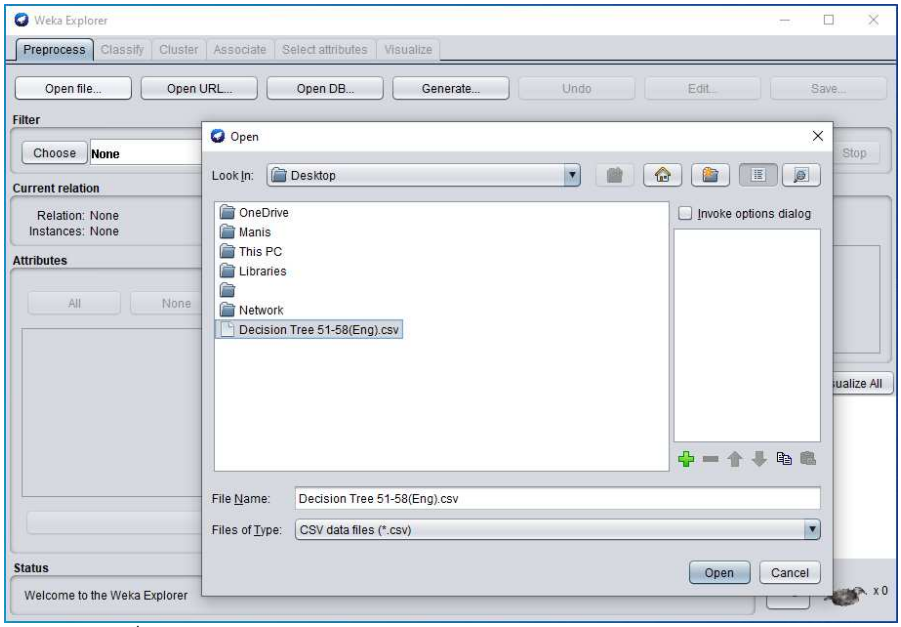


ภาพที่ **3.21** แสดงการเปิดโปรแกรม Weka เวอร์ชัน 3.8.4

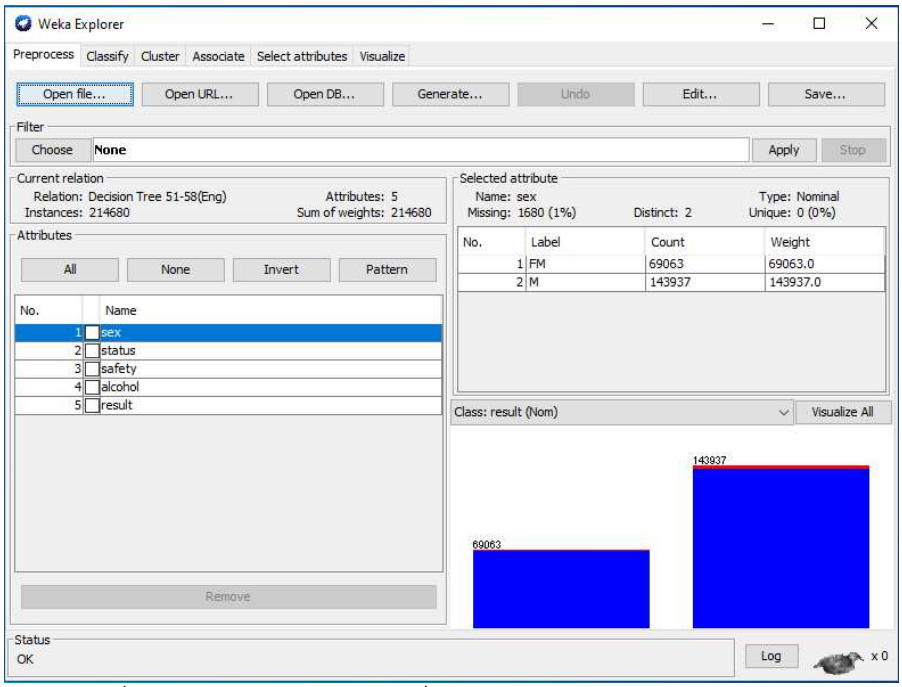


ภาพที่ **3.22** แสดงการเข้าหน้าจอโปรแกรม Weka เวอร์ชัน 3.8.4

ขั้นตอนที่ 2 นำข้อมูลที่ได้จัดเตรียมไว้ โดยเลือกที่ Application>>Explorer>>Open file เลือกไฟล์ข้อมูลที่ต้องการนำมาทดสอบตามภาพที่ 3.23 และหลังจากนั้นโปรแกรมแสดงหน้าจอข้อมูล ตามภาพที่ 3.24

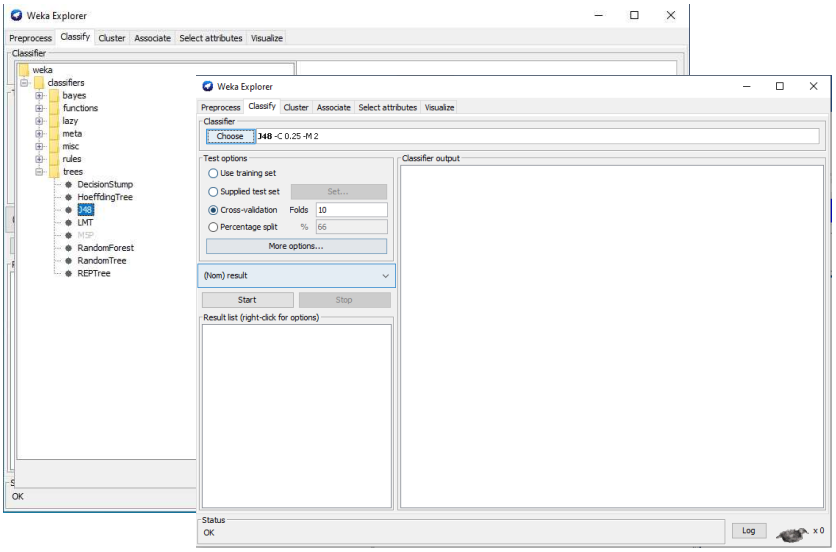


ภาพที่ **3.23** แสดงการนำข้อมูลเข้าในโปรแกรม Weka เวอร์ชัน 3.8.4

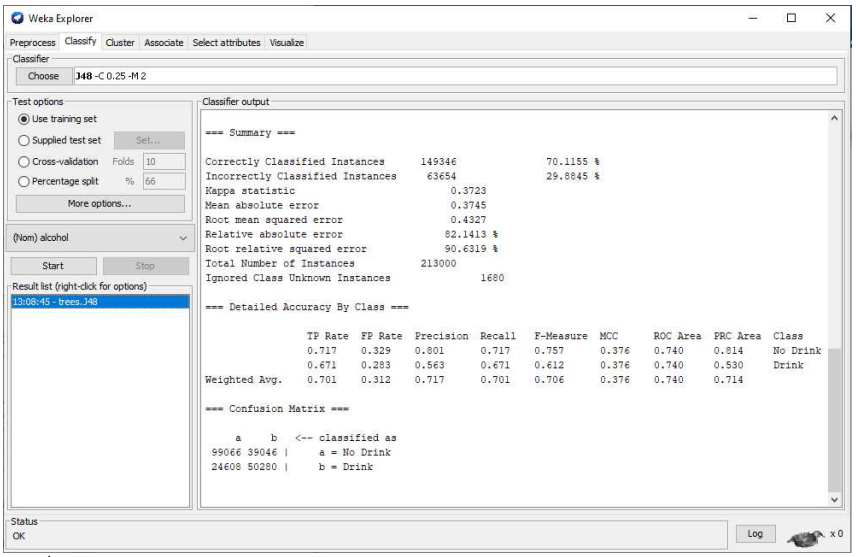


ภาพที่ **3.24** หน้าจอแสดงข้อมูลที่นำเข้าโปรแกรม Weka เวอร์ชัน 3.8.4

ขั้นตอนที่ 3 ดำเนินการเลือกเทคนิคที่ต้องการ ได้แก่ เทคนิคการจัดกลุ่มข้อมูล โดยเลือก Classification>>Choose>>tree เลือกเทคนิคที่ต้องการ ในที่นี้คณะผู้จัดทำเลือกใช้เทคนิค J48 จากนั้นให้กดปุ่ม Start ตามภาพที่ 3.25 จะแสดงผลลัพธ์ได้ตามภาพที่ 3.26



ภาพที่ **3.25** การเลือกเทคนิคที่ใช้ในการจำแนกข้อมูลแบบ Decision Tree: J48

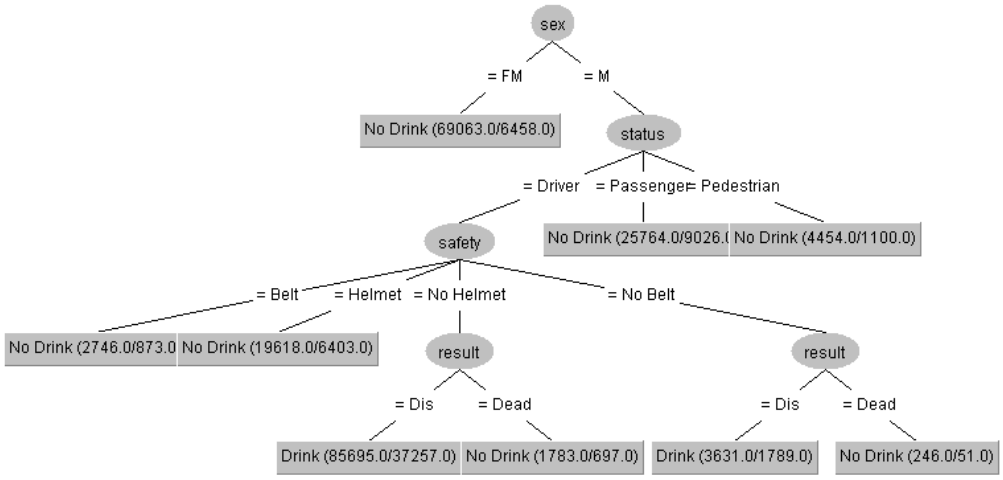


ภาพที่ **3.26** หน้าจอผลลัพธ์ของโมเดลการจำแนกข้อมูลแบบ Decision Tree: J48

จากผลการทดลองพบว่าเทคนิค Decision Tree: J48 ให้ผลลัพธ์การจำแนกประเภท ดื่ม(Drink) และไม่ดื่ม (No Drink) มีความถูกต้องถึง 70.12% และได้กฎจำนวน 9 กฎ แสดงผลลัพธ์ที่เป็นกฎในลักษณะของต้นไม้การตัดสินใจที่มีกิ่งแตกออกมา ดังภาพที่ 3.27 และภาพที่ 3.28

```
=== Classifier model (full training set) ===

J48 pruned tree
------------------

sex = FM: No Drink (69063.0/6458.0)
sex = M
|   status = Driver
|   |   safety = Belt: No Drink (2746.0/873.0)
|   |   safety = Helmet: No Drink (19618.0/6403.0)
|   |   safety = No Helmet
|   |   |   result = Dis: Drink (85695.0/37257.0)
|   |   |   result = Dead: No Drink (1783.0/697.0)
|   |   safety = No Belt
|   |   |   result = Dis: Drink (3631.0/1789.0)
|   |   |   result = Dead: No Drink (246.0/51.0)
|   status = Passenger: No Drink (25764.0/9026.0)
|   status = Pedestrian: No Drink (4454.0/1100.0)

Number of Leaves  :     9

Size of the tree :      14
```

ภาพที่ **3.27** ผลลัพธ์จากการจำแนกกลุ่มแบบ Decision Tree: J48 ในโปรแกรม Weka เวอร์ชัน 3.8.4



ภาพที่ **3.28** รูปแบบแผนภาพโมเดล Graph Decision Tree: J48 ในโปรแกรม Weka เวอร์ชัน 3.8.4

ดังนั้น คณะผู้จัดทำจะใช้เทคนิคของการจำแนกกลุ่มแบบ Decision Tree: J48 มาใช้ในการศึกษา เนื่องจากให้ผลลัพธ์ของกฎที่สามารถทำนายได้จำนวน 9 กฎ ซึ่งสามารถนำไปใช้ในการแบ่งกลุ่มได้ตามเงื่อนไขได้ชัดเจน และสามารถนำกฎที่ได้สามารถนำไปวิเคราะห์กฎต่อไปได้ โดยสามารถจำแนกกฎที่ได้ ดังนี้

กฎข้อที่ 1 IF SEX=FM Then alcohol=No Drink หมายความว่า ถ้าเพศที่เป็นเพศหญิงเกี่ยวเนื่องกัน ผลการพิจารณาพบว่า ประเภทของการดื่มสุรานั้นจะมีผลลัพธ์เป็น “ไม่ดื่ม”

กฎข้อที่ 2 IF SEX=M And status= Pedestrian Then alcohol=No Drink หมายความว่า ถ้าเพศที่เป็นเพศชายเกี่ยวเนื่องกัน และมีสถานะเป็นคนเดินเท้า ผลการพิจารณาพบว่าประเภทของการดื่มสุรานั้นจะมีผลลัพธ์เป็น “ไม่ดื่ม”

กฎข้อที่ 3 IF SEX=M And status= Passenger Then alcohol=No Drink หมายความว่า ถ้าเพศที่เป็นเพศชายเกี่ยวเนื่องกัน และมีสถานะเป็นผู้โดยสาร ผลการพิจารณาพบว่าประเภทของการดื่มสุรานั้นจะมีผลลัพธ์เป็น “ไม่ดื่ม”

กฎข้อที่ 4 IF SEX=M And status= Driver And safety = Belt Then alcohol=No Drink หมายความว่า ถ้าเพศที่เป็นเพศชายเกี่ยวเนื่องกัน มีสถานะเป็นผู้ขับขี่ และมีมาตรการป้องกันด้วยการคาดเข็มขัด ผลการพิจารณาพบว่า ประเภทของการดื่มสุรานั้นจะมีผลลัพธ์เป็น “ไม่ดื่ม”

กฎข้อที่ 5 IF SEX = M And status = Driver And safety = Helmet Then alcohol = No Drink หมายความว่า ถ้าเพศที่เป็นเพศชายเกี่ยวเนื่องกัน มีสถานะเป็นผู้ขับขี่ และมีมาตรการป้องกันด้วยการใส่หมวกกันน็อค ผลการพิจารณาพบว่า ประเภทของการดื่มสุรานั้นจะมีผลลัพธ์เป็น “ไม่ดื่ม”

กฎข้อที่ 6 IF SEX=M And status= Driver And safety = No Belt And result= Dis Then alcohol=Drink หมายความว่า ถ้าเพศที่เป็นเพศชายเกี่ยวเนื่องกัน มีสถานะเป็นผู้ขับขี่ ไม่มีมาตรการป้องกันด้วยการคาดเข็มขัด และมีผลการรักษาจากการเกิดอุบัติเหตุเป็นทุเลา/หาย ผลการพิจารณาพบว่า ประเภทของการดื่มสุรานั้นจะมีผลลัพธ์เป็น “ดื่ม”

กฎข้อที่ 7 IF SEX=M And status= Driver And safety = No Belt And result= Dead Then alcohol=No Drink หมายความว่า ถ้าเพศที่เป็นเพศชายเกี่ยวเนื่องกัน มีสถานะเป็นผู้ขับขี่ไม่มีมาตรการป้องกันด้วยการคาดเข็มขัด และมีผลการรักษาจากการเกิดอุบัติเหตุเป็นตาย ผลการพิจารณาพบว่า ประเภทของการดื่มสุรานั้นจะมีผลลัพธ์เป็น “ไม่ดื่ม”

กฎข้อที่ 8 IF SEX=M And status= Driver And safety = No Helmet And result= Dis Then alcohol=Drinkหมายความว่า ถ้าเพศที่เป็นเพศชายเกี่ยวเนื่องกัน มีสถานะเป็นผู้ขับขี่ไม่มีมาตรการป้องกันด้วยการใส่หมวกกันน็อค และมีผลการรักษาจากการเกิดอุบัติเหตุเป็นทุเลา/หาย ผลการพิจารณาพบว่า ประเภทของการดื่มสุรานั้นจะมีผลลัพธ์เป็น “ดื่ม”

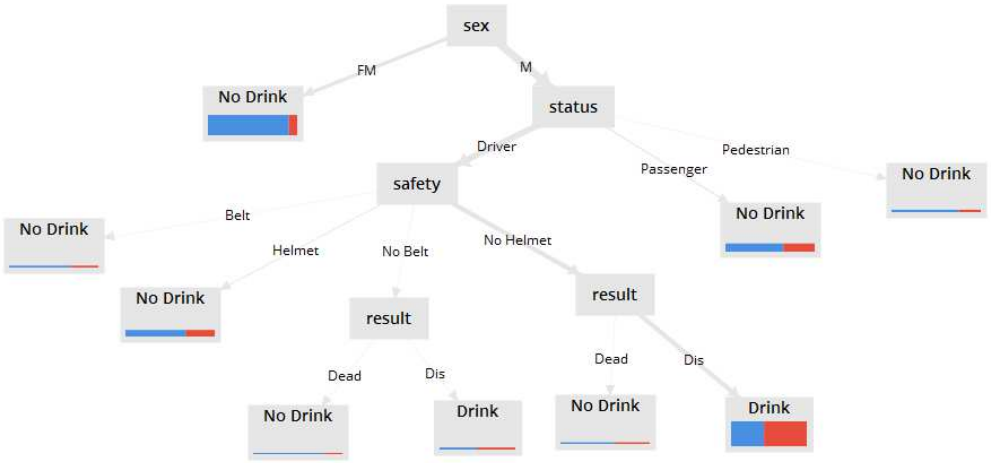
กฎข้อที่ 9 IF SEX=M And status= Driver And safety = No Helmet And result= Dead Then alcohol=No Drink หมายความว่า ถ้าเพศที่เป็นเพศชายเกี่ยวเนื่องกัน มีสถานะเป็นผู้ขับขี่ ไม่มีมาตรการป้องกันด้วยการใส่หมวกกันน็อค และมีผลการรักษาจากการเกิดอุบัติเหตุเป็นตาย ผลการพิจารณาพบว่า ประเภทของการดื่มสุรานั้นจะมีผลลัพธ์เป็น “ไม่ดื่ม”

ทางคณะผู้จัดทำยังได้นำโมเดลของข้อมูลที่เลือกใช้ มาเปรียบเทียบกับโมเดลที่สร้างด้วยโปรแกรม RapidMiner Studio 9.5.1 ดังภาพที่ 3.29 และภาพที่ 3.30 ซึ่งคณะผู้จัดทำพบว่าได้ผลลัพธ์จากการจำแนกกลุ่ม และรูปแบบแผนภาพโมเดลที่ตรงกันสูง

Tree

```
sex = FM: No Drink {No Drink=62605, Drink=6458}
sex = M
|   status = Driver
|   |   safety = Belt: No Drink {No Drink=1873, Drink=873}
|   |   safety = Helmet: No Drink {No Drink=13215, Drink=6403}
|   |   safety = No Belt
|   |   |   result = Dead: No Drink {No Drink=195, Drink=51}
|   |   |   result = Dis: Drink {No Drink=1789, Drink=1842}
|   |   safety = No Helmet
|   |   |   result = Dead: No Drink {No Drink=1086, Drink=697}
|   |   |   result = Dis: Drink {No Drink=37257, Drink=48438}
|   status = Passenger: No Drink {No Drink=16738, Drink=9026}
|   status = Pedestrian: No Drink {No Drink=3354, Drink=1100}
```

ภาพที่ **3.29** ผลลัพธ์จากการจำแนกกลุ่มแบบ Decision Tree ของโปรแกรม RapidMiner



ภาพที่ **3.30** รูปแบบแผนภาพโมเดล Graph Decision Tree ในโปรแกรม RapidMiner

หลังจากคณะผู้จัดทำเลือกการทดสอบประสิทธิภาพของ Model ด้วยวิธี Self Consistency Test หรือเรียกว่า Use Training Set เป็นวิธีการที่นำข้อมูลที่ใช้ในการสร้างโมเดล (model) และข้อมูลที่ใช้ในการทดสอบโมเดลเป็นข้อมูลชุดเดียวกัน คือข้อมูลผู้บาดเจ็บและเสียชีวิตในช่วงเทศกาลปีใหม่ประจำปี 2551–2558 ที่ได้ทำการคัดเลือกมาทั้งหมด ซึ่งคณะผู้จัดทำเลือกโปรแกรมที่ใช้นำเสนอ คือ โปรแกรม Weka 3.8.4 พบว่าการทดสอบประสิทธิภาพโมเดล Decision Tree (J48) พิจารณาได้ว่า โมเดลที่ถูกสร้างขึ้น มีค่าความถูกต้องเฉลี่ยในทุก

โมเดลเท่ากับ 70.12% มีค่าการทำนายข้อมูลไม่ถูกต้องเท่ากับ 29.88% และมีค่าความคลาดเคลื่อนเท่ากับ 0.4327 และเมื่อพิจารณาส่วนค่า Confusion Matrix ในภาพที่ 3.31 พบว่าการหาค่าของข้อมูลค่าจริง กับจำนวนข้อมูลจากการทำนาย แบ่งตามประเภทของดื่ม(Drink) และไม่ดื่ม (No Drink) และนำมาหาค่าเฉลี่ยรวมของทุก class ได้ค่าเฉลี่ยรวมเท่ากับ 69.4% มีผลลัพธ์ตรงกันอยู่ในระดับค่อนข้างดี สามารถนำโมเดลไปใช้งานได้

ดังนั้น คณะผู้จัดทำจึงเลือกใช้โมเดล Decision Tree: J48 ในการนำไปหาแนวทางการป้องกัน และให้คำแนะนำในการปฏิบัติตัวระหว่างเดินทางในช่วงเทศการปีใหม่ เพราะมีค่าความถูกต้องของโมเดล และค่าเฉลี่ย Confusion Matrix จากทุกประเภทของดื่ม(Drink) และไม่ดื่ม (No Drink) อยู่ในระดับที่ค่อนข้างดี

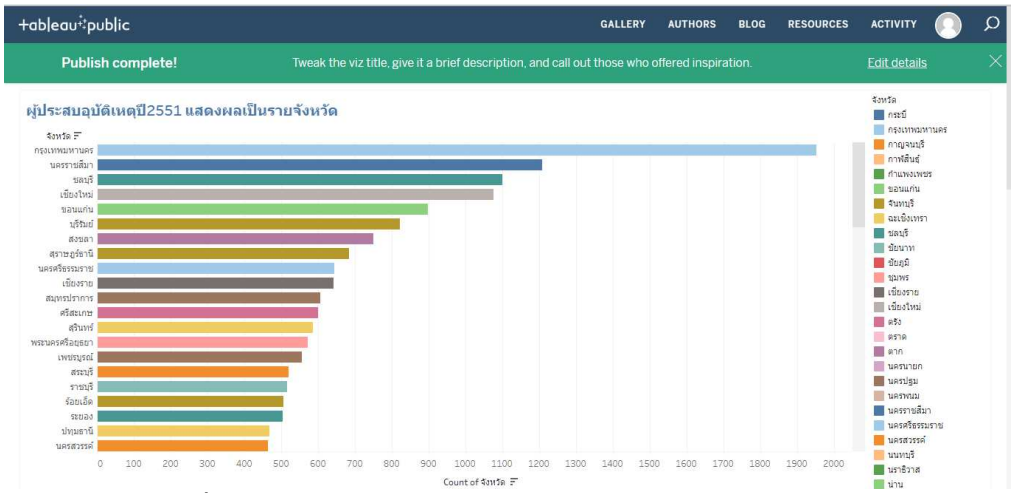
```
=== Summary ===

Correctly Classified Instances      149346               70.1155 %
Incorrectly Classified Instances     63654               29.8845 %
Kappa statistic                          0.3723
Mean absolute error                      0.3745
Root mean squared error                  0.4327
Relative absolute error                 82.1413 %
Root relative squared error             90.6319 %
Total Number of Instances           213000
Ignored Class Unknown Instances              1680

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.717    0.329    0.801      0.717    0.757      0.376    0.740     0.814     No Drink
                 0.671    0.283    0.563      0.671    0.612      0.376    0.740     0.530     Drink
Weighted Avg.    0.701    0.312    0.717      0.701    0.706      0.376    0.740     0.714

=== Confusion Matrix ===

     a     b   <-- classified as
 99066 39046 |     a = No Drink
 24608 50280 |     b = Drink
```

ภาพที่ **3.31** ผลลัพธ์จากการจำแนกกลุ่มแบบ Decision Tree: J48 ในโปรแกรม Weka เวอร์ชัน 3.8.4

3.1.6 เผยแพร่ผลวิเคราะห์ (Deployment) ขั้นตอนการนำผลลัพธ์ที่ได้ไปใช้งานเป็นการทั่วไป อาจจัดทำเป็นรูปแบบของรายงาน (Report) หรือแผนภาพ (Dashboard) ที่พร้อมให้ฝ่ายต่าง ๆ นำไปใช้ประโยชน์ในการวางแผน กำหนดกลยุทธ์ และดำเนินการต่าง ๆ ในทางธุรกิจ

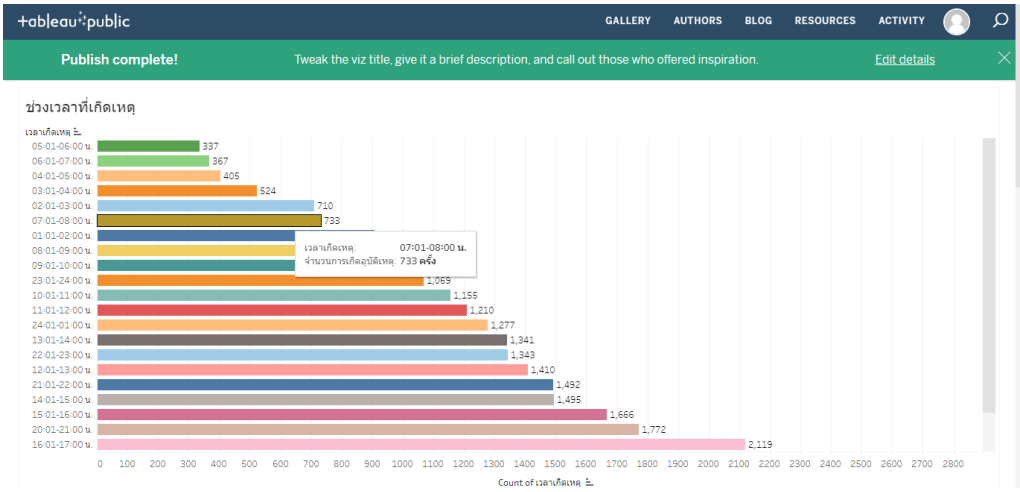
คณะผู้จัดทำนำผลข้อมูลที่ทำการวิเคราะห์แล้ว แสดงผลข้อมูลบน Web browser โดยใช้ชุดคำสั่ง HTML และ CSS3 ด้วยการนำเสนอข้อมูลแบบ visualization เป็นการแสดงผลข้อมูลในรูปแบบของภาพโดยใช้โปรแกรม Tableau Public ซึ่งทางคณะผู้จัดทำได้ยกตัวอย่างการจัดทำเป็นรูปแบบของรายงาน (Report) หรือแผนภาพ (Dashboard) กับข้อมูลผู้บาดเจ็บและเสียชีวิตในช่วงเทศกาลปีใหม่ประจำปี 2551 โดยแต่ละปีจะมีรูปแบบเดียวกัน ดังนี้

- กราฟแท่งแสดงข้อมูลการเกิดอุบัติเหตุแยกเป็นจังหวัด ผู้ใช้สามารถเลือกดูจำนวนการเกิดอุบัติเหตุได้ด้วยตัวกรองจังหวัดทางด้านขวา



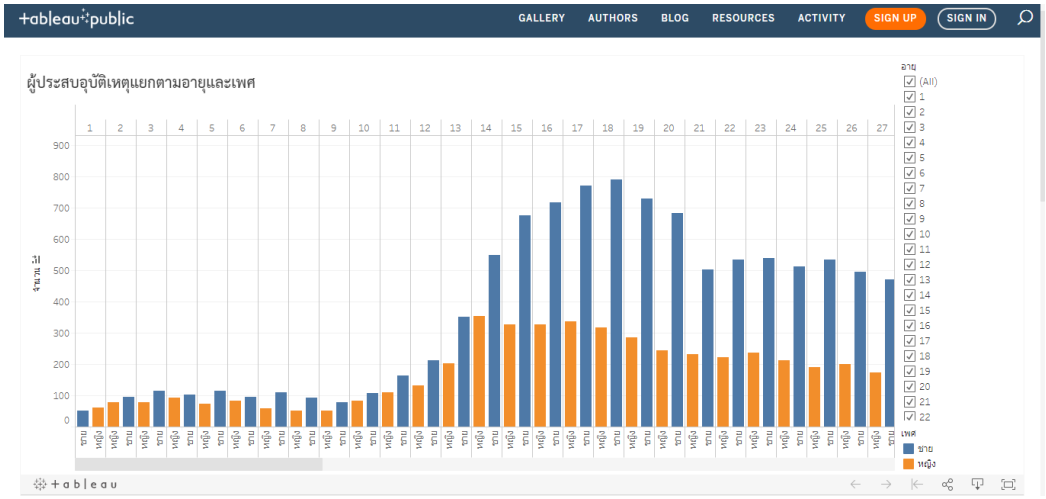
ภาพที่ 3.32 กราฟแท่งแสดงข้อมูลการเกิดอุบัติเหตุแยกเป็นจังหวัด

- กราฟแท่งแสดงช่วงเวลาในการเกิดอุบัติเหตุ ผู้ใช้สามารถเลือกดูช่วงเวลาต่าง ๆ ที่มีการเกิดอุบัติได้โดยการเลื่อนดูหรือดูช่วงเวลาการเกิดอุบัติเหตุแบบเรียงลำดับตามความต้องการได้



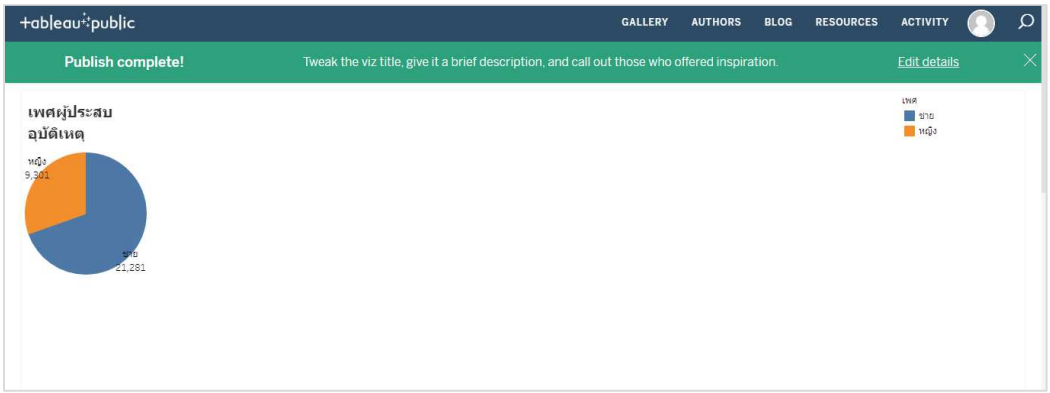
ภาพที่ 3.33 กราฟแท่งแสดงช่วงเวลาในการเกิดอุบัติเหตุ

- กราฟแท่งแสดงข้อมูลผู้ประสบอุบัติเหตุแยกเป็นช่วงอายุแบ่งเป็นเพศชายและเพศหญิง ผู้ใช้สามารถกรองดูข้อมูลได้จากตัวกรองอายุทางด้านขวา และสามารถเลือกดูความถี่การเกิดอุบัติเหตุโดยเรียงลำดับจากน้อยไปมากได้



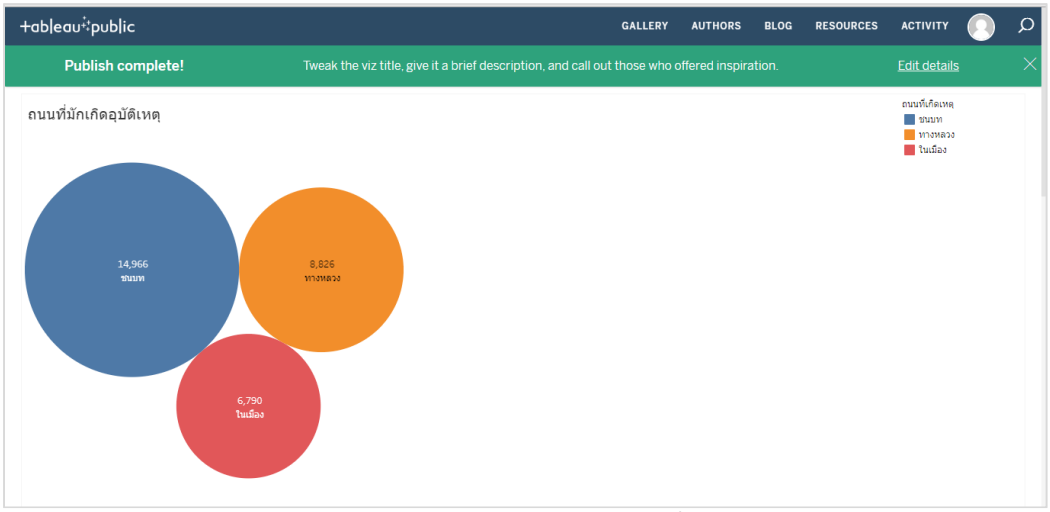
ภาพที่ 3.34 กราฟแท่งแสดงข้อมูลผู้ประสบอุบัติเหตุแยกเป็นช่วงอายุแบ่งเป็นเพศชายและเพศหญิง

- แผนภาพวงกลมแสดงผลจำนวนข้อมูลของผู้ประสบอุบัติโดยแบ่งเป็นผู้ประสบอุบัติเหตุประเภทเพศหญิงและเพศชาย



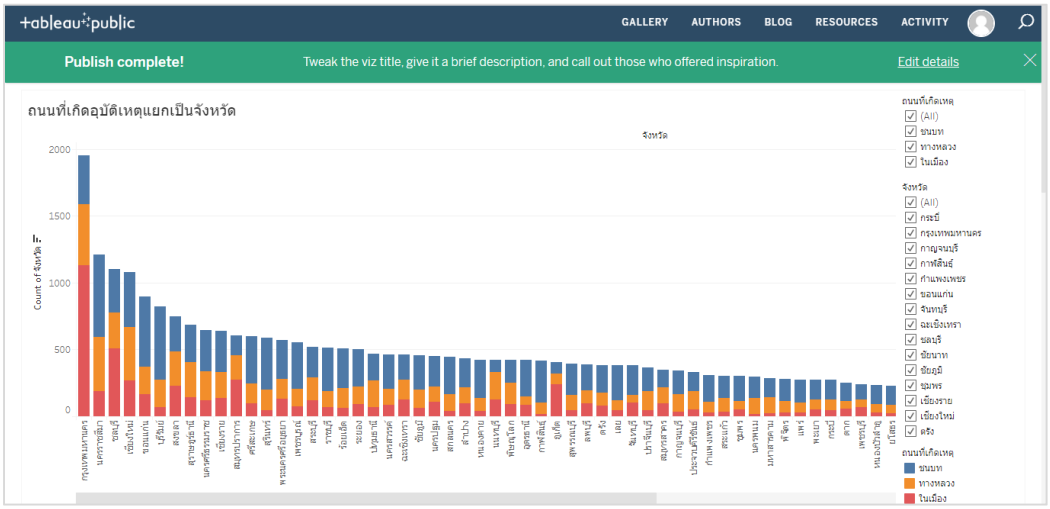
ภาพที่ 3.35 แผนภาพวงกลมแสดงผลข้อมูลผู้ประสบอุบัติ

- แผนภาพวงกลมแสดงข้อมูลถนนที่มีการเกิดอุบัติเหตุ โดยผู้ใช้สามารถกรองดูข้อมูลได้ตามความต้องการด้วยตัวกรองข้อมูลประเภทถนนทางด้านขวา



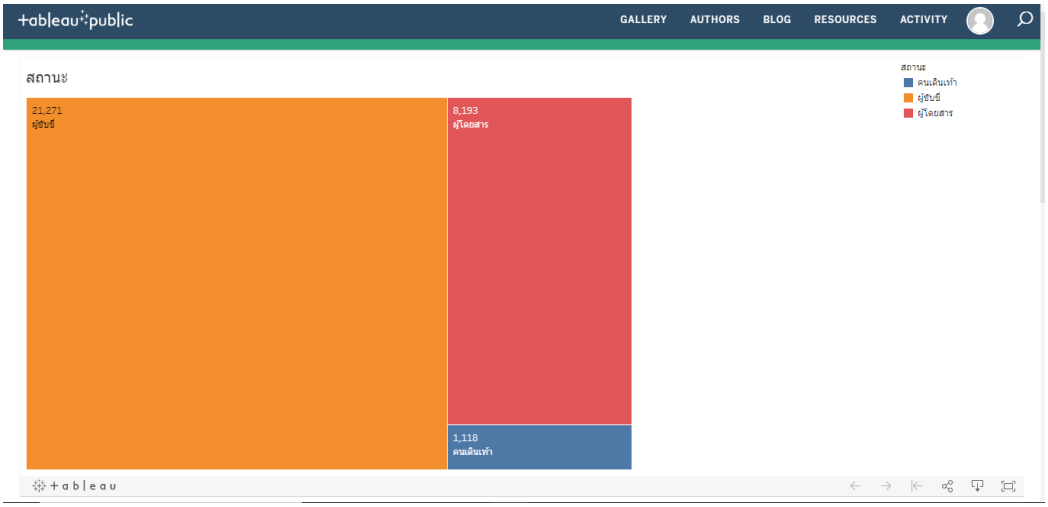
ภาพที่ 3.36 แผนภาพวงกลมแสดงข้อมูลถนนที่มีการเกิดอุบัติเหตุ

- กราฟแท่งแสดงการเกิดอุบัติเหตุบนท้องถนนประเภทต่าง ๆ โดยมีการแบ่งประเภทถนนและเรียงตามการเกิดอุบัติเหตุในแต่ละจังหวัด ผู้ใช้สามารถกรองดูข้อมูลได้จากตัวกรองด้านขวา



ภาพที่ 3.37 กราฟแท่งแสดงการเกิดอุบัติเหตุบนท้องถนนประเภทต่าง ๆ

- แผนภาพแสดงสถานะการประสบอุบัติเหตุของผู้ประสบอุบัติเหตุทั้งหมดโดยแบ่งเป็น 3 ประเภทคือ ผู้ขับขี่ ผู้โดยสารและคนเดินเท้า ผู้ใช้สามารถเลือกดูข้อมูลได้ด้วยตัวกรองสถานะทางด้านขวา

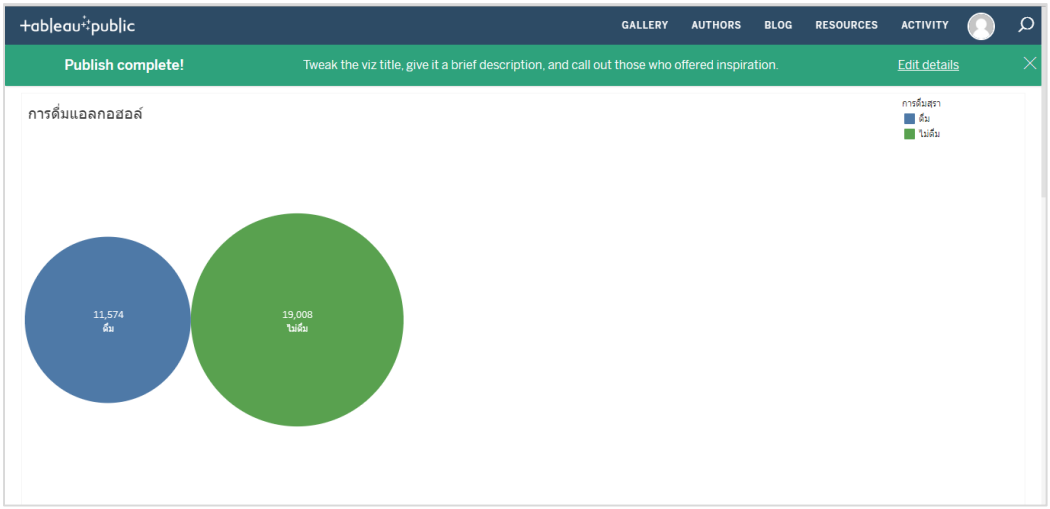


ภาพที่ 3.38 แผนภาพแสดงสถานการณ์ประสบอุบัติเหตุของผู้ประสบอุบัติเหตุทั้งหมด

- กราฟแท่งแสดงสถานะการประสบอุบัติเหตุของผู้ประสบอุบัติเหตุของเพศชายและเพศหญิงโดยแบ่งเป็น 3 ประเภทคือ ผู้ขับขี่ ผู้โดยสารและคนเดินเท้า ผู้ใช้สามารถเลือกดูข้อมูลได้ด้วยตัวกรองสถานะทางด้านขวา

ภาพที่ 3.39 กราฟแท่งแสดงสถานการณ์ประสบอุบัติเหตุของผู้ประสบอุบัติเหตุของเพศชายและเพศหญิง

- แผนภาพวงกลมแสดงผลการดื่มแอลกอฮอล์ของผู้ประสบอุบัติเหตุ โดยแบ่งข้อมูลเป็นดื่มแอลกอฮอล์และไม่ดื่มแอลกอฮอล์

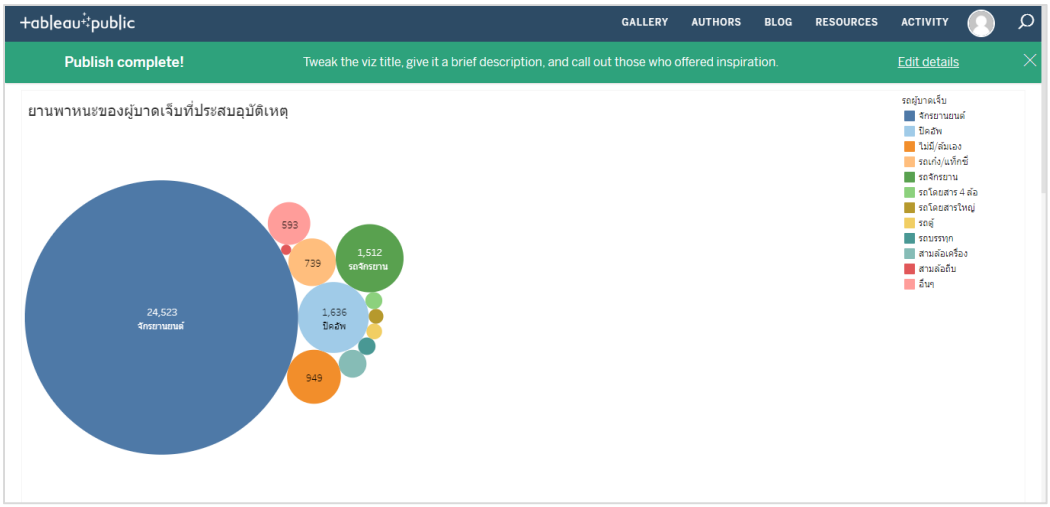


ภาพที่ 3.40 แผนภาพวงกลมแสดงผลการดื่มแอลกอฮอล์ของผู้ประสบอุบัติเหตุ

- หน้าการแสดงผลการดื่มแอลกอฮอล์ของผู้ประสบอุบัติเหตุในแต่ละจังหวัด โดยแบ่งข้อมูลเป็นดื่มแอลกอฮอล์และไม่ดื่มแอลกอฮอล์

ภาพที่ 3.41 หน้าการแสดงผลการดื่มแอลกอฮอล์ของผู้ประสบอุบัติเหตุในแต่ละจังหวัด

- แผนภาพแสดงผลยานพาหนะของผู้บาดเจ็บที่ประสบอุบัติเหตุ โดยผู้ใช้สามารถกรองดูข้อมูลประเภทยานพาหนะได้ด้วยตัวกรองประเพศยานพาหนะของผู้บาดเจ็บได้ทางด้านขวา

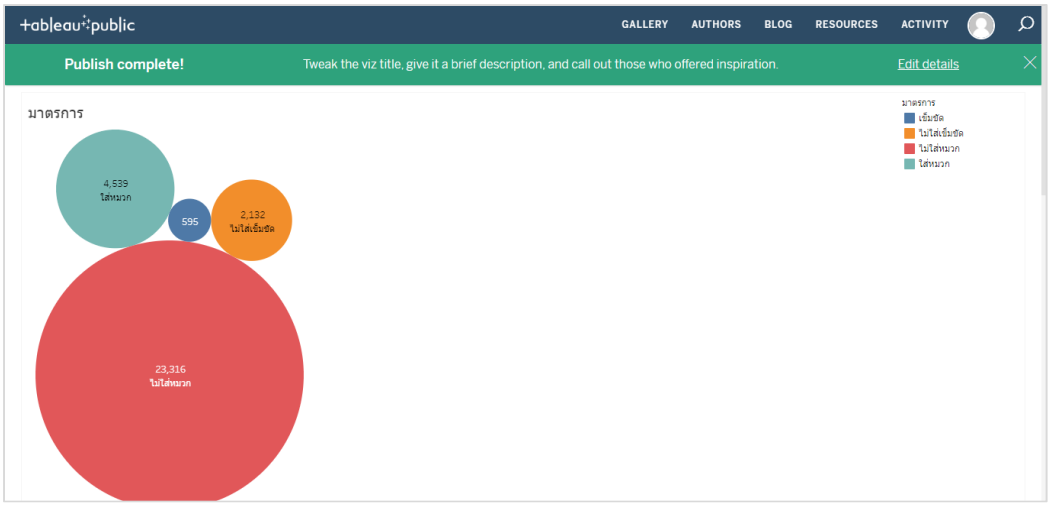


ภาพที่ 3.42 แผนภาพแสดงผลยานพาหนะของผู้บาดเจ็บที่ประสบอุบัติเหตุ

- แผนภาพแสดงผลยานพาหนะของคู่กรณี โดยผู้ใช้สามารถกรองดูข้อมูลประเภทยานพาหนะได้ด้วยตัวกรองประเพศยานพาหนะของคู่กรณีได้ทางด้านขวา

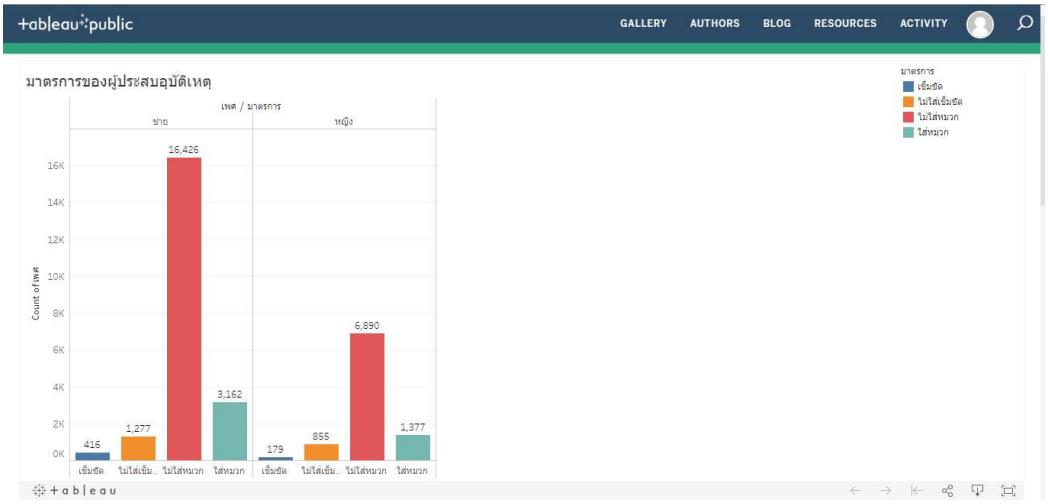
ภาพที่ 3.43 แผนภาพแสดงผลยานพาหนะของคู่กรณี

- แผนภาพวงกลมแสดงมาตรการในการป้องกันตัวเองระหว่างเดินทางของผู้ประสบอุบัติเหตุแบ่งเป็น สวมหมวกกันน็อค ไม่สวมหมวกกันน็อค คาดเข็มขัดนิรภัยและไม่คาดเข็มขัดนิรภัย



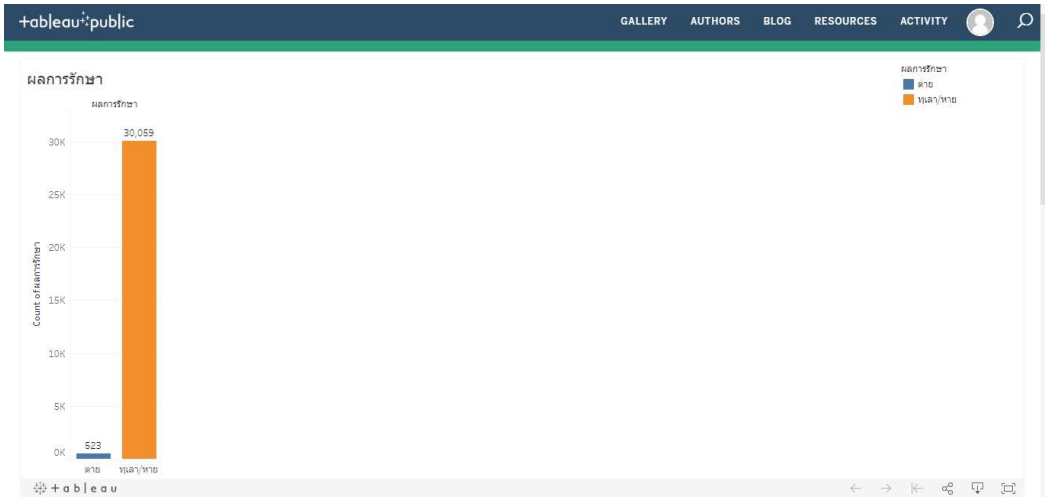
ภาพที่ 3.44 แผนภาพวงกลมแสดงมาตรการในการป้องกันตัวเองระหว่างเดินทางของผู้ประสบอุบัติเหตุ

- กราฟแท่งแสดงผลมาตรการในการป้องกันตัวเองระหว่างเดินทางของผู้ประสบอุบัติเหตุแบ่งเป็น สวมหมวกกันน็อค ไม่สวมหมวกกันน็อค คาดเข็มขัดนิรภัยและไม่คาดเข็มขัดนิรภัย โดยแยกตามเพศชายและเพศหญิง ผู้ใช้สามารถกรองดูข้อมูลมาตรการต่าง ๆ ได้



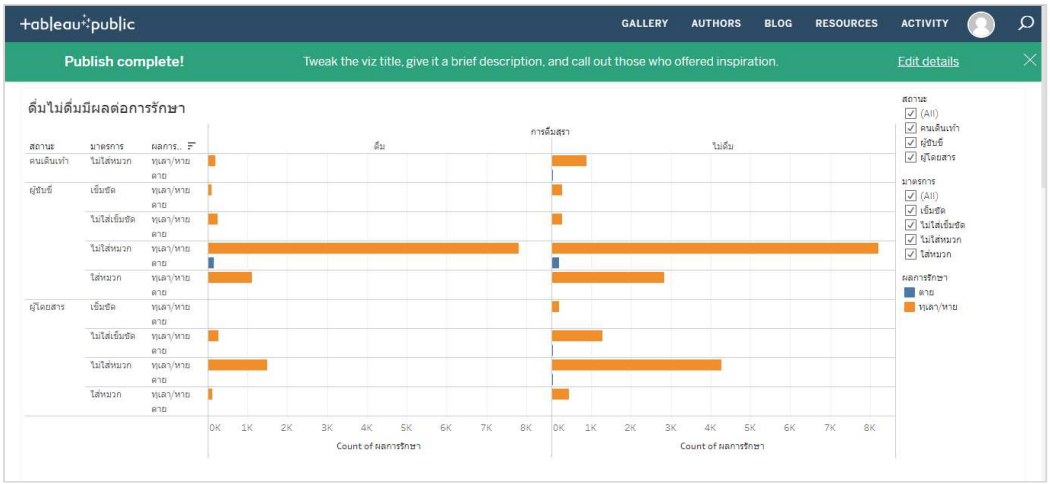
ภาพที่ 3.45 กราฟแท่งแสดงผลมาตรการในการป้องกันตัวเองระหว่างเดินทาง

- กราฟแสดงผลผลการรักษาของผู้ประสบอุบัติเหตุ แบ่งออกเป็นทุเลา/หาย และตาย โดยผู้ใช้สามารถกรองดูข้อมูลได้จากตัวกรองผลการรักษาทางด้านขวา



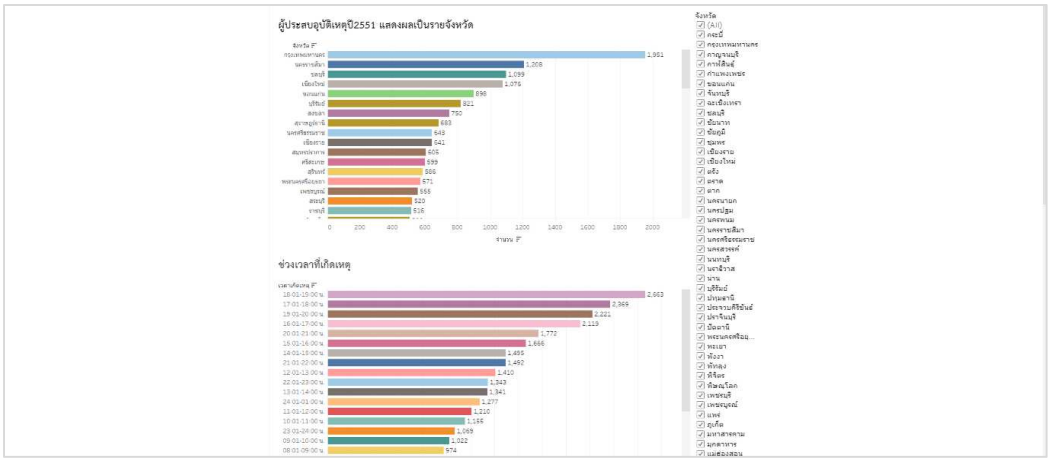
ภาพที่ 3.46 กราฟแสดงผลผลการรักษาของผู้ประสบอุบัติเหตุ

- แสดงผลสถานะของผู้ประสบอุบัติเหตุแยกเป็น ผู้ขับขี่ ผู้โดยสารและคนเดินเท้า มาตรการในการเดินทางของผู้ประสบอุบัติเหตุแยกเป็น สวมหมวกกันน็อค ไม่สวมหมวกกันน็อค คาดเข็มขัดนิรภัยและไม่คาดเข็มขัดนิรภัย การดื่มแอลกอฮอล์ และผลการรักษาของผู้ประสบอุบัติเหตุ โดยผู้ใช้สามารถกรองดูข้อมูลได้จากตัวกรองแต่ละประเภททางด้านขวา



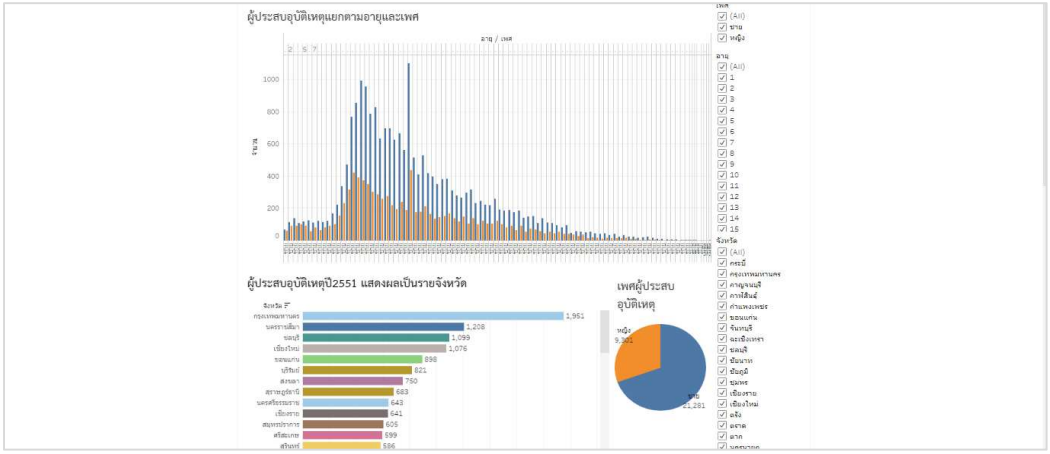
ภาพที่ 3.47 แสดงผลสถานะของผู้ประสบอุบัติเหตุ

- แดชบอร์ดอุบัติเหตุปี 2551 ในแต่ละจังหวัด แสดงผู้ประสบอุบัติเหตุทั้งหมดในปี 2551 แสดงผลโดยแยกเป็นจังหวัดและช่วงเวลาที่การเกิดอุบัติเหตุ ผู้ใช้สามารถกรองดูข้อมูลได้จากตัวกรอง



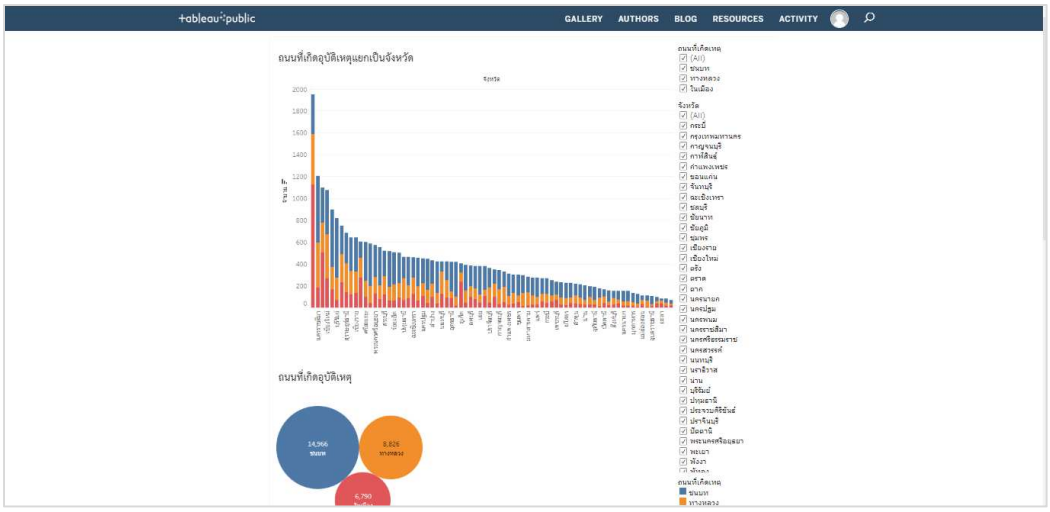
ภาพที่ 3.48 แดชบอร์ดอุบัติเหตุปี 2551 ในแต่ละจังหวัด

- แดชบอร์ดแสดงอุบัติเหตุปี 2551 ยอดผู้ประสบอุบัติเหตุ ภายในแดชบอร์ดแสดงช่วงอายุของผู้ประสบอุบัติเหตุแยกเป็นชายหญิง แสดงจำนวนผู้ประสบอุบัติเหตุทั้งหมดในปี 2551 แยกตามเพศ และแสดงการเกิดอุบัติเหตุในแต่ละจังหวัด ผู้ใช้สามารถกรองดูข้อมูลที่ต้องการดูได้จากตัวกรอง



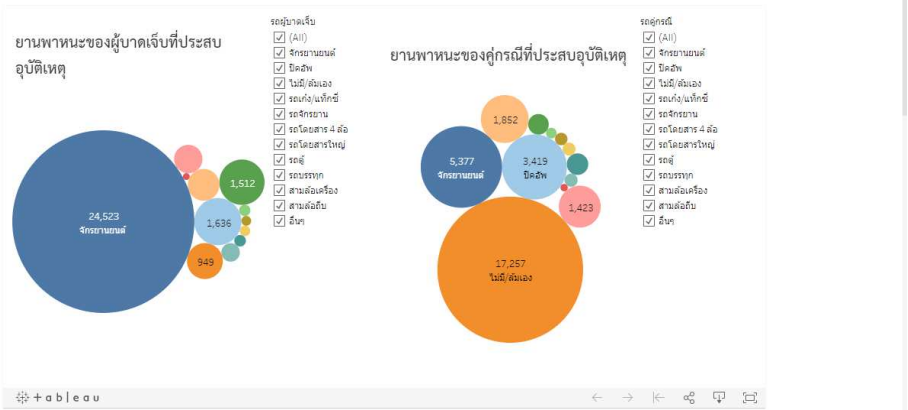
ภาพที่ 3.49 แดชบอร์ดแสดงอุบัติเหตุปี 2551 ยอดผู้ประสบอุบัติเหตุ

- แดชบอร์ดแสดงถนนที่เกิดอุบัติเหตุในปี 2551 ภายในประกอบไปด้วยกราฟแสดงถนนโดยแยกตามจังหวัด แผนภาพวงกลมแสดงจำนวนการเกิดอุบัติเหตุในถนนแต่ละเส้น ผู้ใช้สามารถกรองดูข้อมูลที่ต้องการดูได้จากตัวกรอง



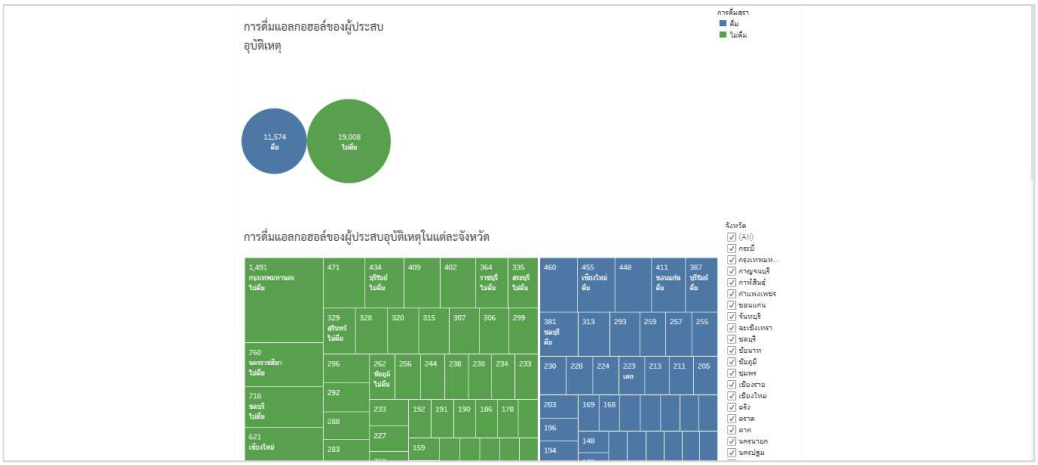
ภาพที่ 3.50 แดชบอร์ดแสดงถนนที่เกิดอุบัติเหตุในปี 2551

- แดชบอร์ดแสดงยานพาหนะที่ประสบอุบัติเหตุ ประกอบไปด้วยแผนภาพแสดงยานพาหนะของผู้บาดเจ็บที่ประสบอุบัติเหตุจะเห็นได้ว่ายานพาหนะที่ประสบอุบัติเหตุมากที่สุดคือรถจักรยานยนต์ และแผนภาพวงกลมแสดงยานพาหนะของคู่กรณีที่ประสบอุบัติเหตุ



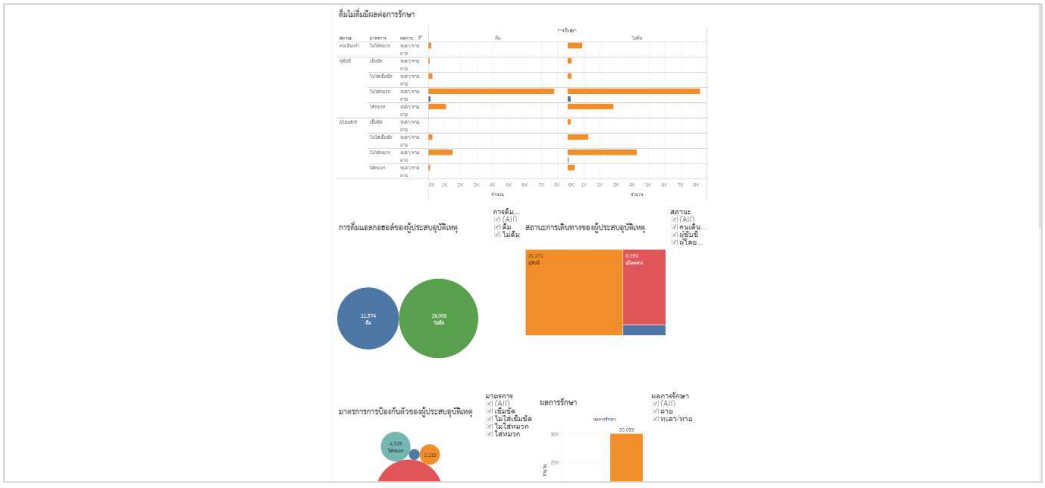
ภาพที่ 3.51 แดชบอร์ดแสดงยานพาหนะที่ประสบอุบัติเหตุ

- แดชบอร์ดแสดงผลการดื่มแอลกอฮอล์ ประกอบไปด้วย แผนภาพวงกลมแสดงจำนวนการดื่มและไม่ดื่มแอลกอฮอล์ แผนภาพแสดงจำนวนการดื่มและไม่ดื่มแอลกอฮอล์ของผู้ประสบอุบัติเหตุโดยแยกข้อมูลเป็นจังหวัด ผู้ใช้สามารถกรองดูข้อมูลได้จากตัวกรองข้อมูล



ภาพที่ 3.52 แดชบอร์ดแสดงผลการดื่มแอลกอฮอล์

- แดชบอร์ดแสดงข้อมูลเกี่ยวกับผู้ประสบอุบัติเหตุปี 2551 กราฟแสดงผลด้วยตัวแปรต่าง ๆ ที่ประกอบกันขึ้นมาจากข้อมูล การดื่มแอลกอฮอล์ของผู้ประสบอุบัติเหตุ สถานการณ์เดินทาง มาตรการในการป้องกันตัวเองระหว่างเดินทาง และผลการรักษาหลังจากประสบอุบัติเหตุ ด้านล่างประกอบไปด้วยแผนภาพวงกลมด้านบนซ้ายการดื่มแอลกอฮอล์ของผู้ประสบอุบัติเหตุ แผนภาพแสดงสถานการณ์เดินทางของผู้ประสบอุบัติเหตุ แผนภาพวงกลมแสดงมาตรการในการป้องกันตัวของผู้ประสบอุบัติเหตุ และกราฟแสดงผลการรักษาของผู้ประสบอุบัติเหตุทั้งหมด

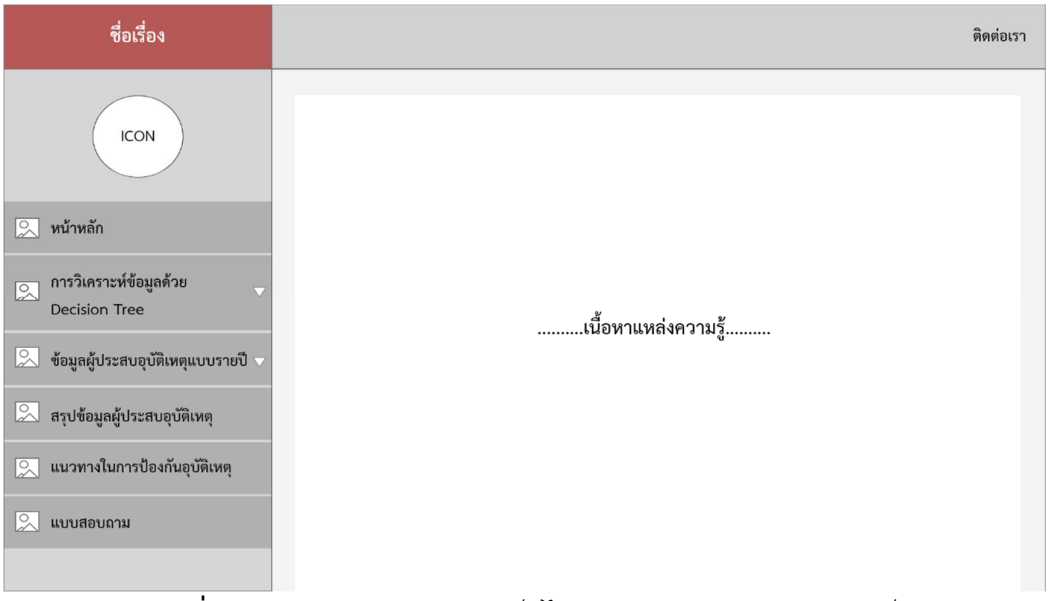


ภาพที่ 3.52 แดชบอร์ดแสดงข้อมูลเกี่ยวกับผู้ประสบอุบัติเหตุปี 2551

3.2 การออกแบบเว็บไซต์

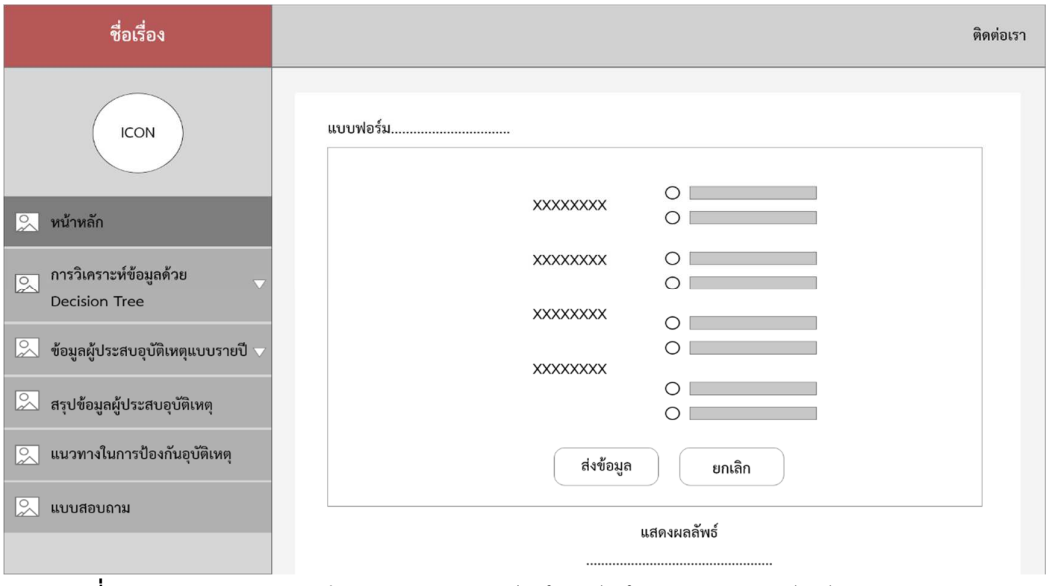
3.2.1 การออกแบบ Wireframe หน้าจอเว็บไซต์

1) หน้าของเว็บไซต์ แสดงเมนูต่าง ๆ ของหน้าเว็บ โดยจะแสดงความรู้ต่าง ๆ



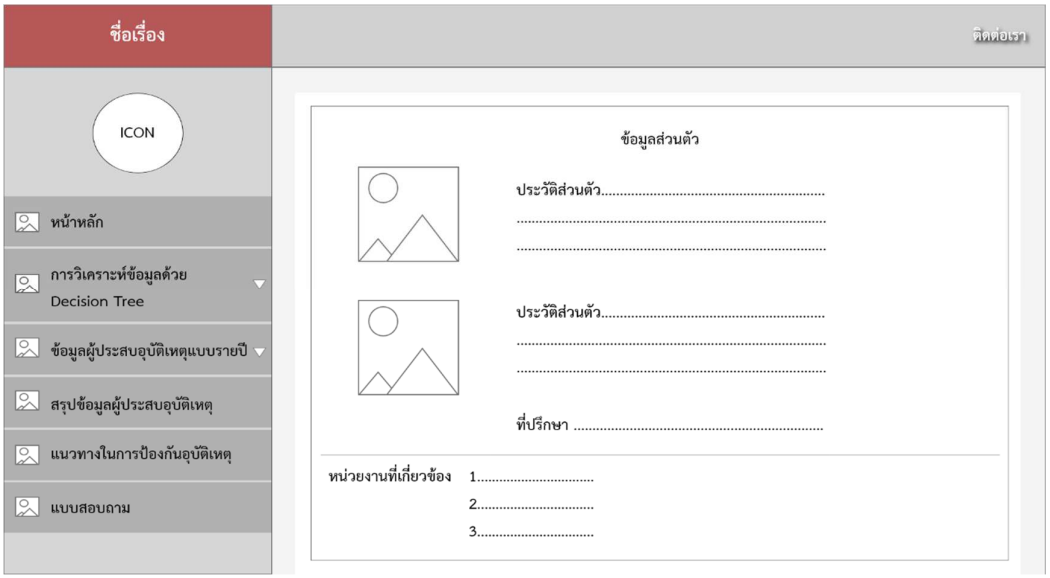
ภาพที่ 3.53 Wireframe หน้าของเว็บไซต์ แสดงเมนูต่าง ๆ ของหน้าเว็บ

2) หน้าแสดงแบบฟอร์มสำหรับจำลองการปฏิบัติตัวระหว่างเดินทางในช่วงเทศกาลปีใหม่ของผู้ใช้งานเว็บไซต์



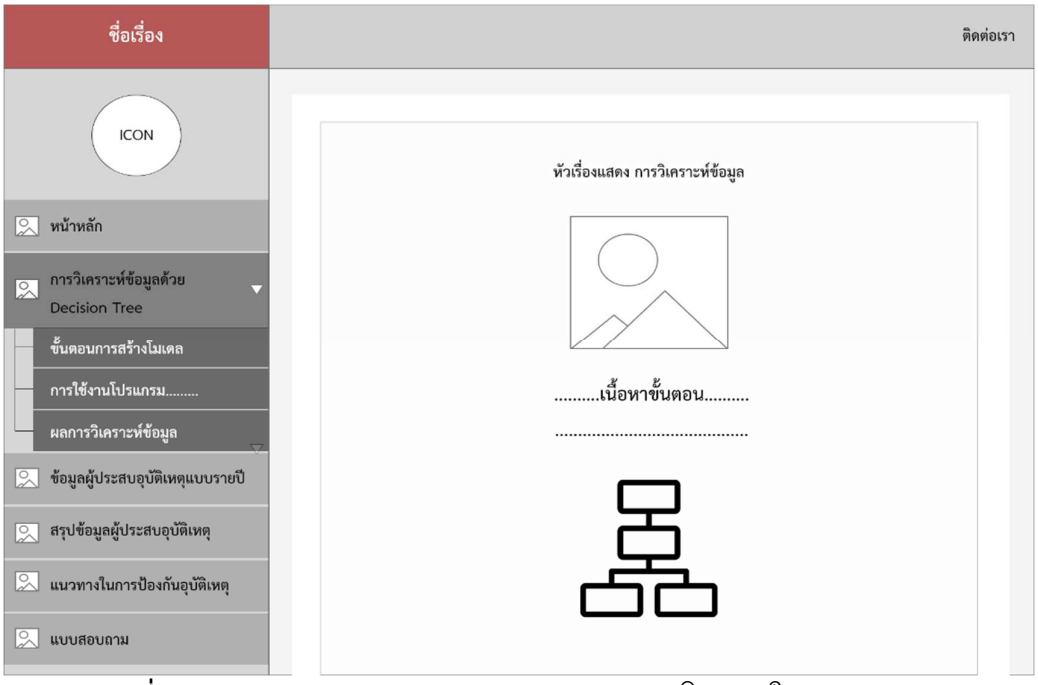
ภาพที่ 3.54 Wireframe หน้าแสดงแบบฟอร์มสำหรับจำลองการปฏิบัติตัวระหว่างเดินทาง

3) หน้า “ติดต่อเรา” จะแสดงข้อมูลประวัติผู้จัดทำ และหน่วยงานที่เกี่ยวข้องต่าง ๆ



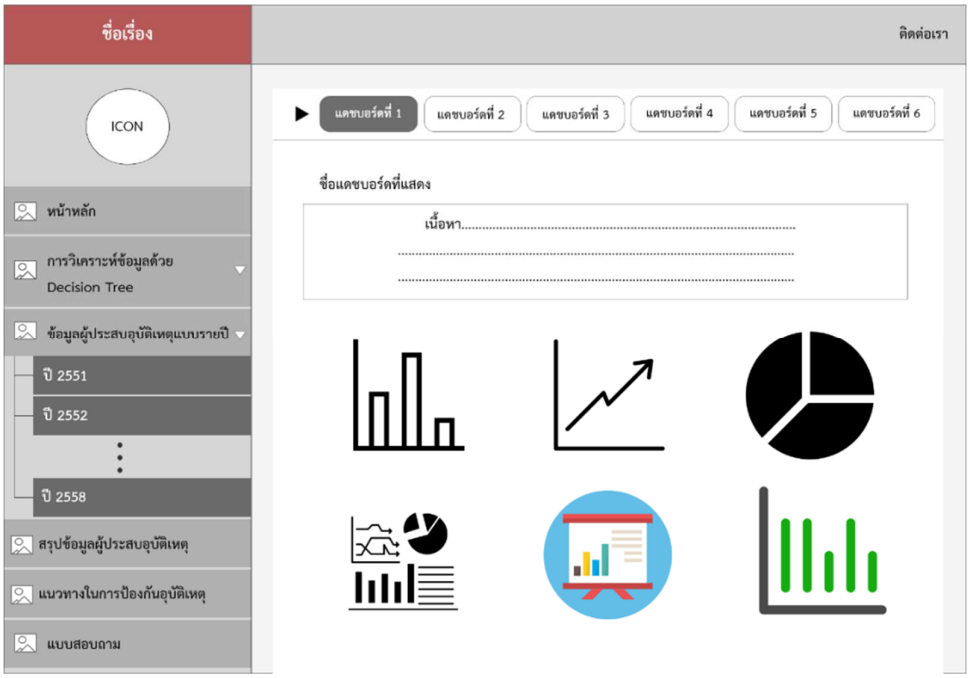
ภาพที่ 3.55 Wireframe หน้าติดต่อเรา

4) หน้าแสดงการวิเคราะห์ข้อมูลโดยการใช้ Decision Tree ที่จะมีการแสดงขั้นตอนการสร้างโมเดล การใช้งานโปรแกรม Weka 3.8.4 และผลของการวิเคราะห์ข้อมูล



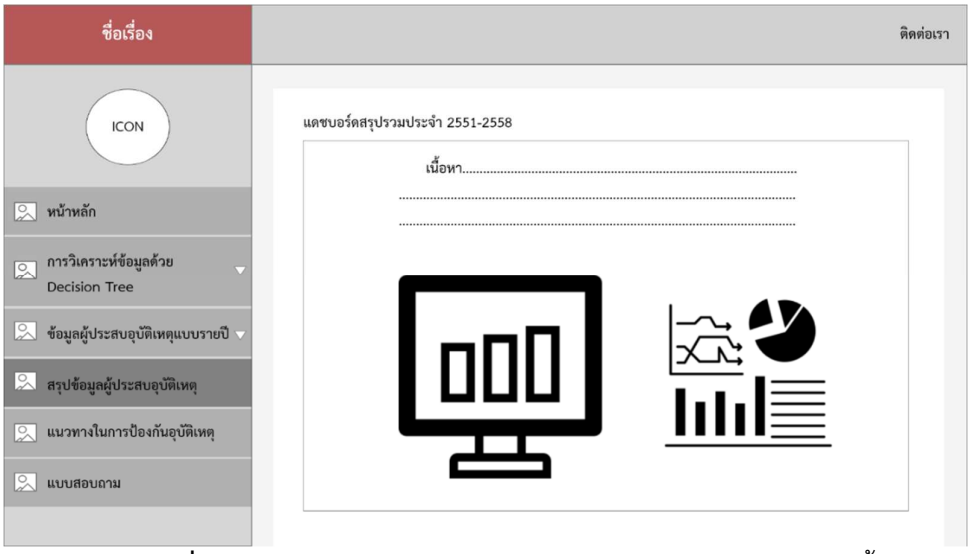
ภาพที่ 3.56 Wireframe หน้าแสดงการวิเคราะห์ข้อมูลโดยการใช้ Decision Tree

5) หน้าแสดงผลข้อมูลในรูปแบบตาราง และแดชบอร์ดผู้ประสบอุบัติเหตุรายปี ตั้งแต่ 2551–2558



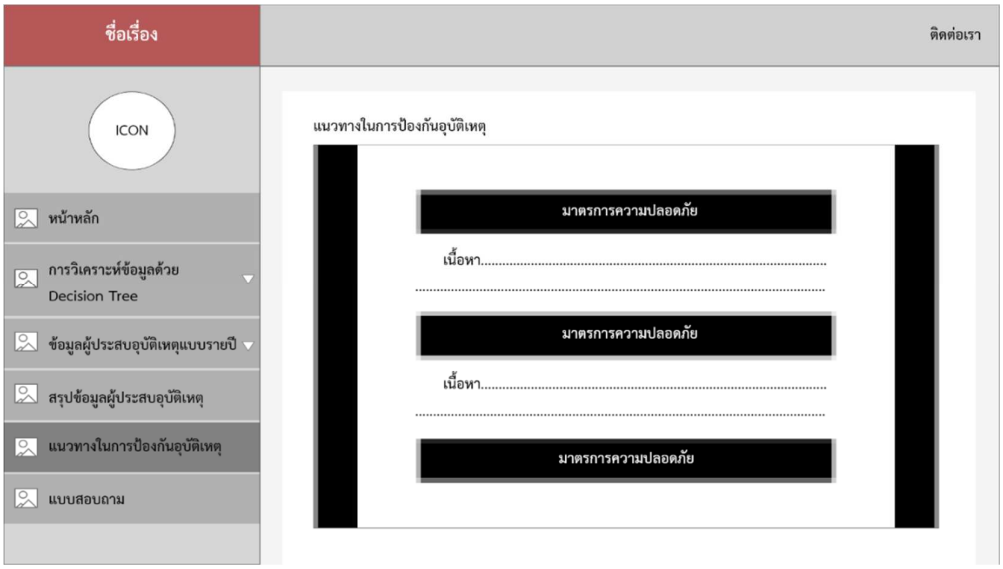
ภาพที่ 3.57 Wireframe หน้าแสดงผลข้อมูลแดชบอร์ดผู้ประสบอุบัติเหตุรายปี

6) หน้าแสดงผลสรุปข้อมูลผู้ประสบอุบัติเหตุทั้งหมด



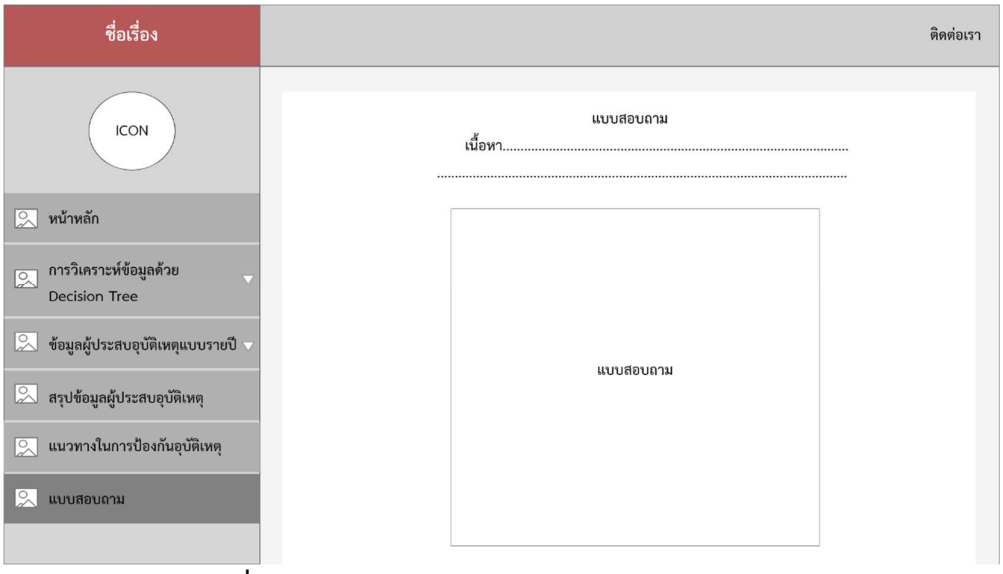
ภาพที่ 3.58 Wireframe หน้าแสดงผลสรุปข้อมูลผู้ประสบอุบัติเหตุทั้งหมด

7) หน้าแสดงผลข้อมูลแนวทางในการป้องกันอุบัติเหตุ ด้วยมาตรการต่าง ๆ



ภาพที่ 3.59 Wireframe หน้าแสดงผลข้อมูลแนวทางในการป้องกันอุบัติเหตุ

8) หน้าแสดงการกรอกแบบสอบถาม เพื่อเก็บรวบรวมข้อมูลจากผู้ใช้งานในเว็บไซต์ เพื่อการวิเคราะห์ครั้งต่อไป



ภาพที่ 3.60 Wireframe หน้าแสดงการกรอกแบบสอบถาม

3.3 บทสรุป

จากวิธีการดำเนินงานโครงงานในข้างต้นทั้งหมดนี้ คณะผู้จัดทำได้แสดงวิธีในการจัดการกับข้อมูลผู้บาดเจ็บและเสียชีวิตในช่วงเทศกาลปีใหม่ประจำปี 2551–2558 ด้วยขั้นตอนกระบวนการวิเคราะห์ข้อมูลด้วย CRISP–DM อย่างละเอียด รวมถึงการสร้างโมเดล Decision Tree จากโปรแกรมที่ใช้เลือกทำเหมืองข้อมูลเพื่อใช้นำเสนอ คือ โปรแกรม Weka 3.8.4 ในการสร้างโมเดล Decision Tree และประเมินประสิทธิภาพของโมเดล ซึ่งมีค่าความถูกต้องเฉลี่ยในทุกโมเดลเท่ากับ 70.12% เป็นการทดสอบประสิทธิภาพของ Model ด้วยวิธี Self Consistency Test พบว่ามีผลลัพธ์ตรงกันอยู่ในระดับค่อนข้างดี สามารถนำโมเดลไปใช้งานได้ และได้ผลลัพธ์ที่เหมือนกัน เมื่อเปรียบเทียบโมเดลกับโปรแกรม RapidMiner Studio 9.5.1 ทางคณะผู้จัดทำได้นำข้อมูลสารสนเทศมาทำการแสดงผลแบบ visualization โดยใช้โปรแกรม Tableau Public และออกแบบ Wireframe ของเว็บไซต์ที่จะเผยแพร่บน web browser ด้วยโปรแกรม Adobe XD