บทที่ 3
วิธีการดำเนินงานโครงงาน

โครงงานเรื่อง การพัฒนาเว็บไซต์สำหรับวิเคราะห์ข้อมูลผู้เสียชีวิตจากอุบัติเหตุทางถนน ปี 2554 – 2563 ในบทนี้จะเป็นการวิเคราะห์ข้อมูลด้วยเทคนิคทางดาต้าไมนิ่ง ซึ่งมีกระบวนการวิเคราะห์ที่สำคัญหลายขั้นตอน เมื่อเสร็จสิ้นจากกระบวนการวิเคราะห์ข้อมูลแล้วจะเป็นการออกแบบเว็บไซต์และออกแบบรูปแบบการแสดงผลและบทสรุปจากวิธีการดำเนินงาน

3.1 การวิเคราะห์ข้อมูลด้วย CRISP-DM

3.2 แผนภาพกระแสข้อมูล Data Flow Diagram

3.3 การออกแบบเว็บไซต์

3.4 บทสรุป

3.1 การวิเคราะห์ข้อมูลด้วย CRISP-DM

ภาพที่ 3.1 กระบวนการวิเคราะห์ข้อมูล ด้วย CRISP-DM

ที่มา : datacubeth.ai

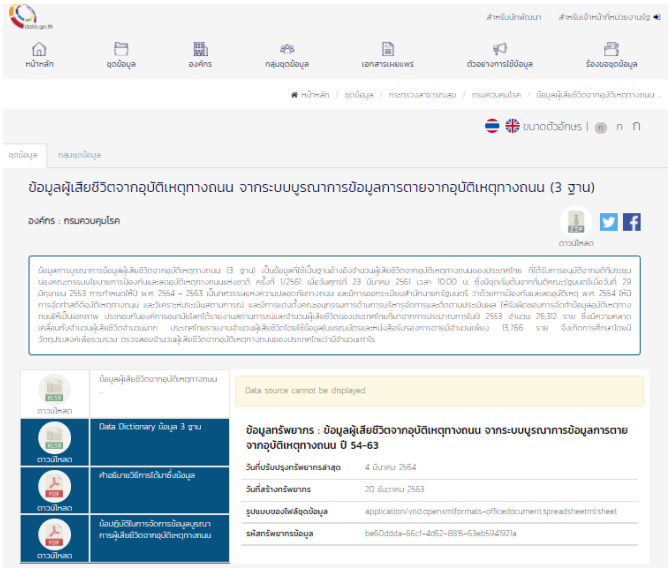
กระบวนการวิเคราะห์ข้อมูล (Cross Industry Standard Process for Data Mining หรือ CRISP. DM) พัฒนาขึ้นในปี ค.ศ. 1936 โดยความร่วมมือของ 3 บริษัทคือ Daimler Chrysler, SPSS และ NCR ซึ่งกระบวนการวิเคราะห์ข้อมูล CRISP–UM. จะเป็นดังรูปแบบในรูปภาพที่ 3.1 แต่ละขั้นตอนในรูปจะเป็นขั้นตอนที่ต่อเนื่องกันนั่นคือ ขั้นตอนถัดไปจะรอผลลัพธ์จากขั้นตอนก่อนหน้าซึ่งแสดงด้วยลูกศรที่เชื่อมระหว่างกระบวนการนั้น ๆ ประกอบด้วย 6 ขั้นตอนหลัก

3.1.1 รู้จักและเข้าใจในธุรกิจ (Business Understanding) เป็นขั้นตอนแรกขอกระบวนการที่มุ่งเน้นไปที่การทำความเข้าใจกระบวนการทางธุรกิจโดยรวม

คณะผู้จัดทำทำความเข้าใจกับปัญหาให้อยู่ในรูปของการวิเคราะห์ข้อมูลทาง Data Mining โดยการวิเคราะห์ข้อมูลในประเด็นนี้ คือ ข้อมูลผู้เสียชีวิตจากอุบัติเหตุทางถนน ปี 2554 – 2563 ซึ่งมีจำนวนข้อมูลมหาศาล จำนวนรายการทั้งหมด 206,588 รายการ ทำให้ไม่สามารถทำความเข้าใจกับข้อมูลที่มีความซับซ้อนได้อย่างรวดเร็ว เช่น ต้องการทราบว่าช่วงอายุใดมีอัตราการเกิดอุบัติเหตุหรือเสียชีวิตสูงสุด

3.1.2 จัดเก็บและรวบรวมข้อมูลให้ครบ (Data Understanding) ขั้นตอนการจัดเก็บและรวบรวมข้อมูล ตลอดจนการพิจารณาตรวจสอบความถูกต้องของข้อมูลที่ได้รับ โดยเลือกว่าจะใช้ข้อมูลทั้งหมดหรือบางส่วนในการวิเคราะห์ให้สอดรับกับวัตถุประสงค์ที่กำหนดไว

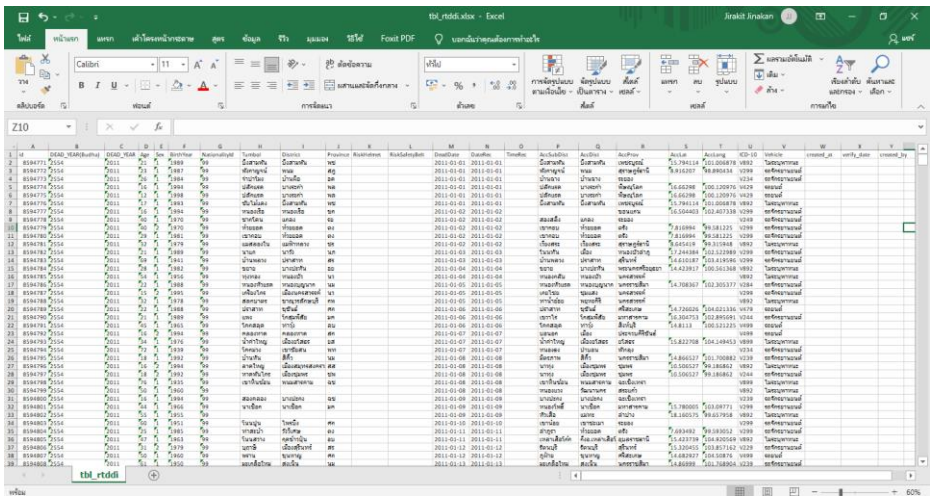
คณะผู้จัดทำทำการรวบรวมข้อมูล เพื่อตรวจสอบรายละเอียด ปริมาณ และความน่าเชื่อถือของข้อมูลผู้เสียชีวิตจากอุบัติเหตุทางถนน ปี 2554 – 2563 ที่ได้จากเว็บไซต์ data.go.th ซึ่งเป็นระบบศูนย์กลางข้อมูลภาครัฐ ที่เผยแพร่สู่สาธารณะอย่างเป็นรูปธรรม เป็นช่องทางให้ผู้ใช้บริการทั้งภาคประชาชน ภาคธุรกิจเอกชน รวมถึงหน่วยงานของรัฐสามารถค้นหาและเข้าถึงข้อมูลที่มีคุณภาพของภาครัฐได้ง่าย



ภาพที่ 3.2 เว็บไซต์ data.go.th

ที่มา : data.go.th

ซึ่งข้อมูลผู้เสียชีวิตจากอุบัติเหตุทางถนน ปี 2554 – 2563 มีจำนวนข้อมูล 206,588 รายการ ประกอบด้วย 25 แอตทริบิวท์ ประกอบด้วย รหัส ปีที่เสียชีวิต(พศ.) ปีที่เสียชีวิตของผู้ประสบเหตุ อายุผู้เสียชีวิต เพศผู้เสียชีวิต ปีเกิด สัญชาติผู้เสียชีวิต(มรณบัตร) ที่อยู่ตำบล ที่อยู่อำเภอ ที่อยู่จังหวัด หมวกนิรภัย การคาดเข็มขัดนิรภัยของผู้เสียชีวิต วันที่เสียชีวิต (ข้อมูลจากมรณบัตร) วันที่เกิดเหตุ เวลาเกิดเหตุ ตำบลเกิดเหตุ อำเภอเกิดเหตุ จังหวัดเกิดเหตุ พิกัด Latitude ของจุดที่เกิดเหตุ พิกัด Longitude ของจุดที่เกิดเหตุ รหัส ICD-10 พาหนะที่ประสบเหตุ วันที่สร้างrecord วันที่ตรวจสอบข้อมูล และผู้นำเข้าข้อมูล

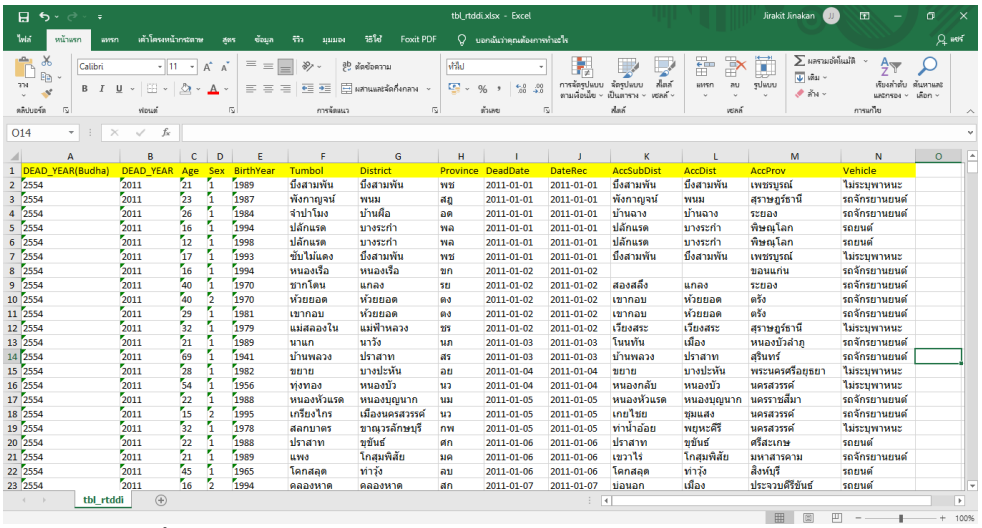


ภาพที่ 3.3 ข้อมูลผู้เสียชีวิตจากอุบัติเหตุทางถนน ปี 2554 – 2563

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 14 แอตทริบิวท์ ได้แก่ ปีที่เสียชีวิต (พศ.) ปีที่เสียชีวิตของผู้ประสบเหตุ อายุผู้เสียชีวิต เพศผู้เสียชีวิต ปีเกิด ที่อยู่ตำบล ที่อยู่อำเภอ ที่อยู่จังหวัด วันที่เสียชีวิต (ข้อมูลจากมรณบัตร) วันที่เกิดเหตุ ตำบลเกิดเหตุ อำเภอเกิดเหตุ จังหวัดเกิดเหตุ และพาหนะที่ประสบเหตุตามรหัส โดยใช้ข้อมูลในปี 2554 - 2563 จำนวน 206,588 รายการ ซึ่งเป็นข้อมูลที่จำเป็นในการนำไปวิเคราะห์ข้อมูล



DEAD_YEAR(Budha)	DEAD_YEAR	Age	Sex	BirthYear	Tumbol	District	Province	DeadDate	DateRec	AccSubDist	AccDist	AccProv	Vehicle
2554	2011	21	1	1989	บึงสามพัน	บึงสามพัน	พช	2011-01-01	2011-01-01	บึงสามพัน	บึงสามพัน	เพชรบูรณ์	ไม่ระบุพาหนะ
2554	2011	23	1	1987	พิงกาญจน์	พนม	สฎ	2011-01-01	2011-01-01	พิงกาญจน์	พนม	สุราษฎร์ธานี	รถจักรยานยนต์
2554	2011	26	1	1984	จำปาโมง	บ้านผือ	อด	2011-01-01	2011-01-01	บ้านฉาง	บ้านฉาง	ระยอง	รถจักรยานยนต์
2554	2011	16	1	1994	ปลักแรด	บางระกำ	พล	2011-01-01	2011-01-01	ปลักแรด	บางระกำ	พิษณุโลก	รถยนต์
2554	2011	12	1	1998	ปลักแรด	บางระกำ	พล	2011-01-01	2011-01-01	ปลักแรด	บางระกำ	พิษณุโลก	รถยนต์
2554	2011	17	1	1993	ชัยไม่แดง	บึงสามพัน	พช	2011-01-01	2011-01-01	บึงสามพัน	บึงสามพัน	เพชรบูรณ์	ไม่ระบุพาหนะ
2554	2011	16	1	1994	หนองเรือ	หนองเรือ	ขก	2011-01-02	2011-01-02			ขอนแก่น	รถจักรยานยนต์
2554	2011	40	1	1970	ชากโดน	แกลง	รย	2011-01-02	2011-01-02	สองสลึง	แกลง	ระยอง	รถจักรยานยนต์
2554	2011	40	2	1970	ห้วยยอด	ห้วยยอด	ตง	2011-01-02	2011-01-02	เขากอบ	ห้วยยอด	ตรัง	รถจักรยานยนต์
2554	2011	29	1	1981	เขากอบ	ห้วยยอด	ตง	2011-01-02	2011-01-02	เขากอบ	ห้วยยอด	ตรัง	รถจักรยานยนต์
2554	2011	32	1	1979	แม่สลองใน	แม่ฟ้าหลวง	ชร	2011-01-02	2011-01-02	เวียงสระ	เวียงสระ	สุราษฎร์ธานี	ไม่ระบุพาหนะ
2554	2011	21	1	1989	นานก	นาวัง	นภ	2011-01-03	2011-01-03	โนนทัน	เมือง	หนองบัวลำภู	รถจักรยานยนต์
2554	2011	69	1	1941	บ้านพลวง	ปราสาท	สร	2011-01-03	2011-01-03	บ้านพลวง	ปราสาท	สุรินทร์	รถจักรยานยนต์
2554	2011	28	1	1982	ขยาย	บางปะหัน	อย	2011-01-04	2011-01-04	ขยาย	บางปะหัน	พระนครศรีอยุธยา	ไม่ระบุพาหนะ
2554	2011	54	1	1956	ทุ่งทอง	หนองบัว	นว	2011-01-04	2011-01-04	หนองกลับ	หนองบัว	นครสวรรค์	ไม่ระบุพาหนะ
2554	2011	22	1	1988	หนองหัวแรด	หนองบุญนาก	นม	2011-01-05	2011-01-05	หนองหัวแรด	หนองบุญนาก	นครราชสีมา	รถจักรยานยนต์
2554	2011	15	2	1995	เกรียงไกร	เมืองนครสวรรค์	นว	2011-01-05	2011-01-05	เกยไชย	ชุมแสง	นครสวรรค์	รถจักรยานยนต์
2554	2011	32	1	1978	สคบาดาร	ชาณุวรลักษบุรี	กพ	2011-01-05	2011-01-05	ท่าน้ำอ้อย	พยุหะคีรี	นครสวรรค์	ไม่ระบุพาหนะ
2554	2011	22	1	1988	ปราสาท	ขุขันธ์	ศก	2011-01-06	2011-01-06	ปราสาท	ขุขันธ์	ศรีสะเกษ	รถยนต์
2554	2011	21	1	1989	แพง	โกสุมพิสัย	มค	2011-01-06	2011-01-06	เชวาไร่	โกสุมพิสัย	มหาสารคาม	รถจักรยานยนต์
2554	2011	45	1	1965	โคกสลุด	ท่าวุ้ง	ลบ	2011-01-06	2011-01-06	โคกสลุด	ท่าวุ้ง	ลิงห์บุรี	รถยนต์
2554	2011	16	2	1994	คลองหาด	คลองหาด	สก	2011-01-07	2011-01-07	บ่อนอก	เมือง	ประจวบคีรีขันธ์	รถยนต์

ภาพที่ 3.4 ข้อมูลผู้เสียชีวิตจากอุบัติเหตุทางถนน ปี 2554 – 2563

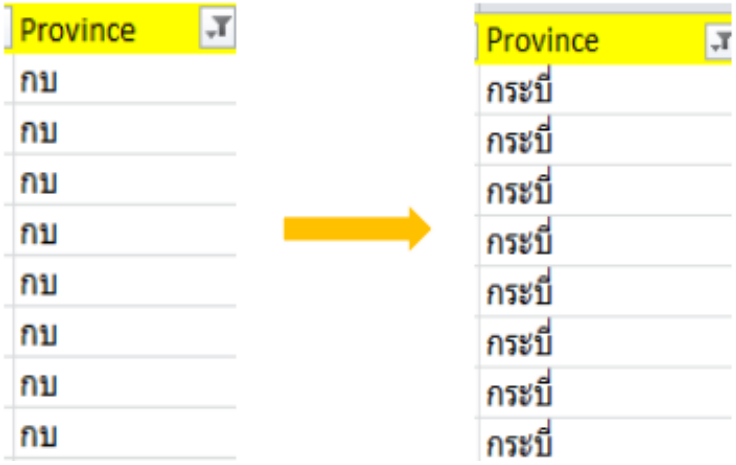
3.1.3.2 ทำการกลั่นกรองข้อมูล (Data Cleaning) คือการทำความสะอาดข้อมูลเป็นกระบวนการตรวจสอบและการแก้ไข (หรือลบ) รายการข้อมูลที่ไม่ถูกต้องออกไปจากชุดข้อมูล ตารางหรือฐานข้อมูล ซึ่งเป็นหลักสำคัญของฐานข้อมูล ทางคณะผู้จัดทำได้ดำเนินการดังนี้

1) ผู้เสียชีวิตจากอุบัติทางถนน ปี 2554 – 2563 คณะผู้จัดทำ ทำการแก้ไขคำผิดหรือชื่อย่อและแก้ไขข้อผิดพลาด ซึ่งคณะผู้จัดทำพบว่ามีข้อมูลบางส่วนมีคำผิดและ

ชื่อย่อของจังหวัด จึงดูข้อมูลตามหลักความเป็นจริงโดยเปรียบเทียบข้อมูลส่วนใหญ่ที่ใช้หรือข้อมูลที่มีคำผิด เนื่องด้วยอาจจะมาจากการคีย์ข้อมูลที่ต้นทาง ดังนั้นคณะผู้จัดทำได้ทำการแก้ไขคำผิดและชื่อย่อของจังหวัด ดังนี้

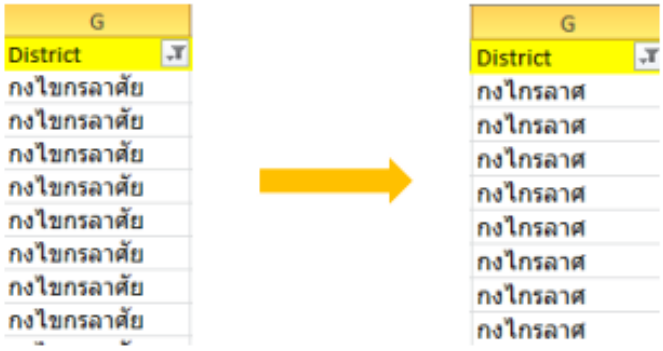
– คณะผู้จัดทำ ทำการแก้ไขข้อมูลคำผิดและชื่อย่อจังหวัดเป็นจำนวนมากไม่สามารถระบุได้ โดยใช้ข้อมูลในปี 2554 – 2563 ที่ทำการ Cleaning แล้ว จำนวน 206,583 รายการ โดยประมาณ

– แก้ไขชื่อย่อของจังหวัด โดยจะทำการเปลี่ยนชื่อย่อของแต่ละจังหวัดให้เป็นชื่อเต็มทั้งหมด ตัวอย่าง เช่น กบ จะเปลี่ยนให้เป็น กระบี่ ทั้งหมด



ภาพที่ 3.5 การแก้ไขชื่อจังหวัดให้เป็นชื่อเต็ม

– แก้ไขชื่ออำเภอที่ผิด โดยจะแก้ไขให้เป็นชื่ออำเภอที่ถูกต้องของแต่ละจังหวัด ตัวอย่าง เช่น กงไขกรลาศัย จะเปลี่ยนให้เป็น กงไกรลาศ ทั้งหมด



ภาพที่ 3.6 การแก้ไขชื่ออำเภอให้ถูกต้อง

3.1.3.3 แปลงรูปแบบของข้อมูล (Data Transformation) เป็นขั้นตอนการแปลงข้อมูลในรูปแบบตารางฐานข้อมูลให้อยู่ในรูป item set เพื่อใช้สำหรับการนำมาวิเคราะห์ด้วยวิธีการของ data mining ทางคณะผู้จัดทำได้ดำเนินการกับข้อมูลผู้เสียชีวิตจากอุบัติเหตุทางถนน ปี 2554 – 2563 ดังนี้

1) คณะผู้จัดทำพบข้อมูล “ไม่ทราบ” ในคอลัมน์ของแต่ละประเภท คือ พาหนะ, มาตรการ และการดื่มสุรา จึงได้ทำการแปลงข้อความให้อยู่ในรูปแบบตัวเลขเพื่อหาฐานนิยม หรือตัวเลขที่ซ้ำกันมากที่สุดในช่วงของข้อมูล ซึ่งจะได้ตรงกับข้อความที่มีการเกิดขึ้นมากที่สุด ทางคณะผู้จัดทำจึงนำมาแทนที่ประเภท “ไม่ทราบ” ของแต่ละกลุ่มของข้อมูลทั้งหมดด้วยฟังก์ชัน MODE

2) คณะผู้จัดทำทำการแปลงรูปแบบด้วยการรวมกลุ่มของข้อมูลในแอตทริบิวต์ ผลการรักษา (Result) เนื่องจากปกติแล้วจะมีข้อมูลในแอตทริบิวต์นี้หลากหลาย ได้แก่ ตายที่เกิดเหตุ ตายระหว่างนำส่ง ซึ่งทางคณะผู้จัดทำจะขอรวมกลุ่มของการตายในรูปแบบต่าง ๆให้เหลือเพียง “ตาย” อย่างเดียว

3) คณะผู้จัดทำทำการแปลงรูปแบบด้วยการรวมกลุ่มของข้อมูลในแอตทริบิวต์ ผลการรักษา (Result) เนื่องจากปกติแล้วจะมีข้อมูลในแอตทริบิวต์นี้หลากหลาย ได้แก่ บาดเจ็บ บาดเจ็บเล็กน้อย ICU พักฟื้น ซึ่งทางคณะผู้จัดทำจะขอรวมกลุ่มของการตายในรูปแบบต่าง ๆให้เหลือเพียง “ทุเลา/หาย” อย่างเดียว

3.1.4 สร้างแบบจำลอง (Modeling) ขั้นตอนการสร้างตัวแบบทางคณิตศาสตร์และสถิติเพื่อการวิเคราะห์ข้อมูล โดยสามารถใช้เทคนิคหรือวิธีการต่าง ๆ อาทิเช่น การแบ่งกลุ่ม (Clustering) การจำแนก (Classification) และการสร้างความสัมพันธ์ (Association rule)

คณะผู้จัดทำได้ทำการวิเคราะห์ข้อมูลด้วยการทำเหมืองข้อมูลแบบ Data Classification เพื่อใช้ทำนายแนวโน้มการเกิดขึ้นของปัจจัยที่ก่อให้เกิดผู้บาดเจ็บและเสียชีวิตจากเทคนิคการจำแนกประเภทข้อมูล ด้วยการสร้างโมเดล Decision Tree เพื่อจัดการกับกลุ่มของข้อมูลที่มีความสัมพันธ์กัน ด้วยชุดข้อมูลที่คัดเลือก ดังนี้

Sex	Vehicle	safety	alcohol	result
ชาย	รถจักรยานยนต์	ใส่หมวก	ไม่ดื่ม	ทุเลา/หาย
ชาย	รถจักรยานยนต์	ใส่หมวก	ไม่ดื่ม	ทุเลา/หาย
ชาย	รถจักรยานยนต์	ใส่หมวก	ไม่ดื่ม	ทุเลา/หาย
ชาย	รถจักรยานยนต์	ใส่หมวก	ไม่ดื่ม	ทุเลา/หาย
ชาย	รถจักรยานยนต์	ใส่หมวก	ไม่ดื่ม	ทุเลา/หาย
ชาย	รถจักรยานยนต์	ใส่หมวก	ไม่ดื่ม	ทุเลา/หาย
ชาย	รถจักรยานยนต์	ใส่หมวก	ไม่ดื่ม	ทุเลา/หาย
ชาย	รถจักรยานยนต์	ใส่หมวก	ไม่ดื่ม	ทุเลา/หาย
หญิง	รถจักรยานยนต์	ใส่หมวก	ไม่ดื่ม	ทุเลา/หาย
ชาย	รถจักรยานยนต์	ใส่หมวก	ไม่ดื่ม	ทุเลา/หาย
หญิง	รถจักรยานยนต์	ใส่หมวก	ไม่ดื่ม	ทุเลา/หาย
ชาย	รถจักรยานยนต์	ใส่หมวก	ไม่ดื่ม	ทุเลา/หาย
ชาย	รถจักรยานยนต์	ใส่หมวก	ไม่ดื่ม	ทุเลา/หาย
ชาย	รถจักรยานยนต์	ใส่หมวก	ไม่ดื่ม	ทุเลา/หาย
ชาย	รถจักรยานยนต์	ใส่หมวก	ไม่ดื่ม	ทุเลา/หาย
หญิง	รถจักรยานยนต์	ใส่หมวก	ไม่ดื่ม	ทุเลา/หาย
ชาย	รถจักรยานยนต์	ใส่หมวก	ไม่ดื่ม	ทุเลา/หาย
ชาย	รถจักรยานยนต์	ใส่หมวก	ไม่ดื่ม	ทุเลา/หาย
ชาย	รถจักรยานยนต์	ใส่หมวก	ไม่ดื่ม	ทุเลา/หาย
ชาย	รถจักรยานยนต์	ใส่หมวก	ไม่ดื่ม	ทุเลา/หาย
หญิง	รถจักรยานยนต์	ใส่หมวก	ไม่ดื่ม	ทุเลา/หาย
ชาย	รถจักรยานยนต์	ใส่หมวก	ไม่ดื่ม	ทุเลา/หาย
หญิง	รถจักรยานยนต์	ใส่หมวก	ไม่ดื่ม	ทุเลา/หาย
หญิง	รถจักรยานยนต์	ใส่หมวก	ไม่ดื่ม	ทุเลา/หาย

ภาพที่ 3.7 ชุดข้อมูลที่คัดเลือกมาทำการวิเคราะห์ข้อมูล

จากรูปภาพที่ประกอบด้วย 5 แอตทริบิวต์ คือ

- เพศ แสดงเพศ ประกอบด้วย 2 ค่า คือ ชาย หญิง
- พาหนะ แสดงสถานะของผู้ประสบอุบัติเหตุ ประกอบด้วย 3 ค่า คือ เดินเท้า รถจักรยานยนต์ รถยนต์
- มาตรการ แสดงมาตรการความปลอกภัยในการขับขี่ ประกอบด้วย 4 ค่า คือ คาดเข็มขัด ไม่คาดเข็มขัด ใส่หมวก ไม่ใส่หมวก
- การดื่มสุรา แสดงการดื่มสุรา ซึ่งเป็นคลาส ประกอบด้วย 2 ค่า คือ ดื่ม ไม่ดื่ม
- ผลการรักษา แสดงผลการรักษา ซึ่งเป็นคลาส ประกอบด้วย 2 ค่า คือ เสียชีวิต ทุเลา/หาย

การสร้างโมเดล Decision tree จะทำการคัดเลือกแอตทริบิวต์ที่มีความสัมพันธ์กับคลาสมากที่สุดขึ้นมาเป็นโหนดบนสุดของ Tree (root node) หลังจากนั้นก็จะหาแอตทริบิวต์ถัดไปเรื่อยๆ ในการหาความสัมพันต์ของแอตทริบิวต์นี้จะใช้ตัววัด 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) x entropy(c1) + p(c2) x entropy(c2) + ...]

การคำนวณค่าแต่ละแอตทริบิวต์เทียบกับคลาสเพื่อหาแอตทริบิวต์ที่มีค่า IG มากที่สุดมาเป็น root ของ Decision กับจำนวนข้อมูลทั้งหมดโดยใช้ผลลัพธ์เป็นดื่มและไม่ดื่ม ดังนี้

1. คำนวณค่า IG ของแอตทริบิวต์ เพศ จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy(parent) = −p(ดื่ม)*log2 p(ดื่ม) + p(ไม่ดื่ม)*log2 p(ไม่ดื่ม)
= −[0.583 * log2(0.583) + 0.417 * log2(0.417)]
= −[0.583 * − 0.778 + 0.417 * − 1.262]
= −[0.454 + 0.526]
= 0.98

entropy(ผล = ชาย) = −p(ดื่ม)*log2 p(ดื่ม) + p(ไม่ดื่ม)*log2 p(ไม่ดื่ม)
= −[0.582 * log2(0.582) + 0.418 * log2(0.418)]
= −[0.582 * − 0.781 + 0.418 * − 1.258]
= −[0.455 + 0.526]
= 0.981

entropy(ผล = หญิง) = −p(ดื่ม)*log2 p(ดื่ม) + p(ไม่ดื่ม)*log2 p(ไม่ดื่ม)
= −[0.584 * log2(0.584) + 0.416 * log2(0.416)]
= −[0.584 * − 0.776 + 0.416 * − 1.265]
= −[0.453 + 0.526]
= 0.979

IG(parent,child) = entropy(parent) – [p(ผล = ชาย) * entropy(ผล = ชาย) + p(ผล = หญิง) * entropy(ผล = หญิง)]

= 0.98 – [0.785 * 0.981 + 0.215 *0.979]

= 0.98 – [0.770 + 0.210]

= 0.98 – 0.98

= 0

2. คำนวณค่า IG ของแอตทริบิวต์ ยานพาหนะ จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy(parent) = –p(ดื่ม)*log2 p(ดื่ม) + p(ไม่ดื่ม)*log2 p(ไม่ดื่ม)

= –[0.583 * log2(0.583) + 0.417 * log2(0.417)]

= –[0.583 * – 0.778 + 0.417 * – 1.262]

= –[0.454 + 0.526]

= 0.98

entropy(ผล = เดินเท้า) = –p(ดื่ม)*log2 p(ดื่ม) + p(ไม่ดื่ม)*log2 p(ไม่ดื่ม)

= –[0.577 * log2(0.577) + 0..422* log2(0.422)]

= –[0.577 * – 0.793 + 0.422 * – 1.245]

= –[0.458 + 0.525]

= 0.983

entropy(ผล = รถจักรยานยนต์) = –p(ดื่ม)*log2 p(ดื่ม) + p(ไม่ดื่ม)*log2 p(ไม่ดื่ม)

= –[0.583 * log2(0.583) + 0.416 * log2(0.416)]

= –[0.583 * – 0.778 + 0.416 * – 1.265]

= –[0.454 + 0.526]

= 0.98

entropy(ผล = รถยนต์) = –p(ดื่ม)*log2 p(ดื่ม) + p(ไม่ดื่ม)*log2 p(ไม่ดื่ม)

= –[0.582 * log2(0.582) + 0.418 * log2(0.418)]

= –[0.582 * – 0.781 + 0.418 * – 1.258]

= –[0.455 + 0.526]

= 0.981

IG(parent,child) = entropy(parent) – [p(ผล = เดินเท้า) * entropy(ผล = เดินเท้า) + p(ผล = รถจักรยานยนต์) * entropy(ผล = รถจักรยานยนต์) + p(ผล = รถยนต์) * entropy(ผล = รถยนต์)]

= 0.98 – [0.035 * 0.983 + 0.411 * 0.98 + 0.554 * 0.981]

= 0.98 – [0.034 + 0.403 + 0.543]

= 0.98 – 0.98

= 0

3. คำนวณค่า IG ของแอตทริบิวต์ มาตรการ จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy(parent) = –p(ดื่ม)*log2 p(ดื่ม) + p(ไม่ดื่ม)*log2 p(ไม่ดื่ม)

= –[0.583 * log2(0.583) + 0.417 * log2(0.417)]

= –[0.583 * – 0.778 + 0.417 * – 1.262]

= –[0.454 + 0.526]

= 0.98

entropy(ผล = ใส่หมวก) = –p(ดื่ม)*log2 p(ดื่ม) + p(ไม่ดื่ม)*log2 p(ไม่ดื่ม)

= –[0.172 * log2(0.172) + 0.828 * log2(0.828)]

= –[0.172 * – 2.54 + 0.828 * – 0.272]

= –[0.437 + 0.225]

= 0.662

entropy(ผล = ไม่ใส่หมวก) = –p(ดื่ม)*log2 p(ดื่ม) + p(ไม่ดื่ม)*log2 p(ไม่ดื่ม)

= –[0.692 * log2(0.692) + 0.308 * log2(0.308)]

= –[0.692 * – 0.531 + 0.308 * – 1.70]

= –[0.367 + 0.524]

= 0.891

entropy(ผล = คาดเข็มขัด) = –p(ดื่ม)*log2 p(ดื่ม) + p(ไม่ดื่ม)*log2 p(ไม่ดื่ม)

= –[0.582 * log2(0.582) + 0.418 * log2(0.418)]

= –[0.582 * – 0.781 + 0.418 * – 1.258]

= –[0.455 + 0.526]

= 0.981

entropy(ผล = ไม่คาดเข็มขัด) = –p(ดื่ม)*log2 p(ดื่ม) + p(ไม่ดื่ม)*log2 p(ไม่ดื่ม)

= –[0.582 * log2(0.582) + 0.418 * log2(0.418)]

= –[0.582 * – 0.781 + 0.418 * – 1.258]

= –[0.455 + 0.526]

= 0.981

IG(parent,child) = entropy(parent) – [p(ผล = ใส่หมวก) * entropy(ผล = ใส่หมวก) + p(ผล = ไม่ใส่หมวก) * entropy(ผล = ไม่ใส่หมวก) + p(ผล = คาดเข็มขัด) * entropy(ผล = คาดเข็มขัด) + p(ผล = ไม่คาดเข็มขัด) * entropy(ผล = ไม่คาดเข็มขัด)]

= 0.98 – [0.093 * 0.662 + 0.353 * 0.891 + 0.381 * 0.981 + 0.173 * 0.981]

= 0.98 – [0.062 + 0.315 + 0.374 + 0.17]

= 0.98 – 0.921

= 0.059

4. คำนวณค่า IG ของแอตทริบิวต์ ผลการรักษา จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy(parent) = –p(ดื่ม)*log2 p(ดื่ม) + p(ไม่ดื่ม)*log2 p(ไม่ดื่ม)

= –[0.583 * log2(0.583) + 0.417 * log2(0.417)]

= –[0.583 * – 0.778 + 0.417 * – 1.262]

= –[0.454 + 0.526]

= 0.98

entropy(ผล = ทุเลา/หาย) = –p(ดื่ม)*log2 p(ดื่ม) + p(ไม่ดื่ม)*log2 p(ไม่ดื่ม)

= –[0.583 * log2(0.583) + 0.417 * log2(0.417)]

= –[0.583 * – 0.778 + 0.417 * – 1.262]

= –[0.454 + 0.526]

= 0.98

entropy(ผล = เสียชีวิต) = –p(ดื่ม)*log2 p(ดื่ม) + p(ไม่ดื่ม)*log2 p(ไม่ดื่ม)

= –[0.582 * log2(0.582) + 0.418 * log2(0.418)]

= –[0.582 * – 0.781 + 0.418 * – 1.258]

= –[0.455 + 0.526]

= 0.981

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

= 0.98 – [0.397 * 0.98 + 0.603 * 0.981]

= 0.98 – [0.389 + 0.592]

= 0.98 – 0.981

= 0

จากการคำนวณค่า IG ของทุกแอตทริบิวต์พบว่าค่า IG ของแอตทริบิวต์ มาตรการ มีค่ามากที่สุด (0.059) ดังนั้นจึงเลือกแอตทริบิวต์ มาตรการ ขึ้นมาเป็นโหนด root และจะต้องทำการแตกกิ่งต่าง ๆของ โหนด root ออกไปจนข้อมูลและโหนดจะมีคำตอบเดียวกัน และคณะผู้จัดทำพบว่า ใส่หมวกและไม่ใส่หมวก ไม่สามารถสร้างกิ่งแต่ละโหนดได้ เนื่องจากไม่มีความสัมพันธ์กับแอตทริบิวต์ใด จึงสรุปข้อมูลได้เป็นผลลัพธ์ดื่มและไม่ดื่ม ดังนั้นผู้จัดทำจึงทำการสร้างโหนดในระดับถัดไปของแอตทริบิวต์ มาตรการ คาดเข็มขัดและไม่คาดเข็มขัด เนื่องจากมีค่าเท่ากัน (0.981)

การคำนวณค่าแต่ละแอตทริบิวต์ในระดับที่ 2 ต่อจากโหนด root เพื่อหาค่า IG ที่มากที่สุดของแอตทริบิวต์ มาตรการต่อ ผลการรักษา เพศ และยานพาหนะ กับจำนวนทั้งหมดโดยใช้ผลลัพธ์เป็นดื่มและไม่ดื่ม ดังนี้

1. คำนวณค่า IG ของแอตทริบิวต์ของมาตรการต่อผลการรักษา ได้ดังนี้

entropy(parent) = –p(ดื่ม)*log2 p(ดื่ม) + p(ไม่ดื่ม)*log2 p(ไม่ดื่ม)

= –[0.582 * log2(0.582) + 0.418 * log2(0.418)]

= –[0.582 * – 0.781 + 0.418 * – 1.258]

= –[0.455 + 0.526]

= 0.981

entropy(ผล = ทุเลา/หาย) = –p(ดื่ม)*log2 p(ดื่ม) + p(ไม่ดื่ม)*log2 p(ไม่ดื่ม)

= –[0.564 * log2(0.564) + 0.436 * log2(0.436)]

= –[0.564 * – 0.826 + 0.436 * –1.198]

= –[0.465 + 0.522]

= 0.987

entropy(ผล = เสียชีวิต) = –p(ดื่ม)*log2 p(ดื่ม) + p(ไม่ดื่ม)*log2 p(ไม่ดื่ม)

= –[0.601 * log2(0.601) + 0.399 * log2(0.399)]

= –[0.601 * – 0.735 + 0.399 * – 1.326]

= –[0.442 + 0.529]

= 0.971

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

= 0.981 – [0.502 * 0.987 + 0.489 * 0.971]

= 0.981 – [0.495 + 0.475]

= 0.981 – 0.970

= 0.011

2. คำนวณค่า IG ของแอตทริบิวต์ของมาตรการต่อเพศ ได้ดังนี้

entropy(parent) = –p(ดื่ม)*log2 p(ดื่ม) + p(ไม่ดื่ม)*log2 p(ไม่ดื่ม)

= –[0.582 * log2(0.582) + 0.418 * log2(0.418)]

= –[0.582 * – 0.781 + 0.418 * – 1.258]

= –[0.455 + 0.526]

= 0.981

entropy(ผล = ชาย) = –p(ดื่ม)*log2 p(ดื่ม) + p(ไม่ดื่ม)*log2 p(ไม่ดื่ม)

= –[0.574 * log2(0.574) + 0.426 * log2(0.426)]

= –[0.574 * – 0.801 + 0.426 * – 1.231]

= –[0.460 + 0.524]

= 0.984

entropy(ผล = หญิง) = –p(ดื่ม)*log2 p(ดื่ม) + p(ไม่ดื่ม)*log2 p(ไม่ดื่ม)

= –[0.608 * log2(0.608) + 0.392 * log2(0.392)]

= –[0.608 * – 0.718 + 0.392 * – 1.351]

= –[0.437 + 0.530]

= 0.967

IG(parent,child) = entropy(parent) – [p(ผล = ชาย) * entropy(ผล = ชาย) + p(ผล = หญิง) * entropy(ผล = หญิง)]

= 0.981 – [0.754 * 0.984 + 0.246* 0.967]

= 0.981 – [0.733 + 0.238]

= 0.981 – 0.971

= 0.01

3. คำนวณค่า IG ของแอตทริบิวต์ของมาตรการต่อยานพาหนะ “คาดเข็มขัด” ได้ดังนี้

entropy(parent) = –p(ดื่ม)*log2 p(ดื่ม) + p(ไม่ดื่ม)*log2 p(ไม่ดื่ม)

= –[0.582 * log2(0.582) + 0.418 * log2(0.418)]

= –[0.582 * – 0.781 + 0.418 * – 1.258]

= –[0.455 + 0.526]

= 0.981

entropy(ผล = รถยนต์) = –p(ดื่ม)*log2 p(ดื่ม) + p(ไม่ดื่ม)*log2 p(ไม่ดื่ม)

= –[0.582 * log2(0.582) + 0.418 * log2(0.418)]

= –[0.582 * – 0.781 + 0.418 * – 1.258]

= –[0.455 + 0.526]

= 0.981

IG(parent,child) = entropy(parent) – [p(ผล = รถยนต์) * entropy(ผล = รถยนต์)]

= 0.981 – [1 * 0.981]

= 0.981 – 0.981

= 0

จากการคำนวณค่า IG ของทุกแอตทริบิวต์พบว่าค่า IG ของแอตทริบิวต์ มาตรการต่อการรักษา เพศ และยานพาหนะ พบว่าค่า IG ของแอตทริบิวต์ มาตรการต่อผลการรักษามีค่ามากที่สุด (0.011) ดังนั้นจึงเลือกแอตทริบิวต์ ผลการรักษา ขึ้นมาเป็นโหนดระดับที่ 2 ต่อจากโหนด root คณะผู้จัดจึงทำการสร้างโหนดในระดับถัดไปของแอตทริบิวต์ ผลการรักษาและเพศออกไปจนข้อมูลในแต่ละโหนดมีคลาสคำตอบเดียวกัน

การคำนวณค่าแต่ละแอตทริบิวต์ในระดับที่ 3 ต่อจากโหนดระดับที่ 2 เพื่อหาค่า IG ที่มากที่สุดของแอตทริบิวต์ผลการรักษาต่อเพศประกอบไปด้วย ชายและหญิง กับจำนวนทั้งหมดโดยใช้ผลลัพธ์เป็นดื่มและไมดื่ม ดังนี้

1. คำนวณค่า IG ของแอตทริบิวต์ของมาตรการและผลการรักษาต่อเพศ “ชาย” ได้ดังนี้

entropy(parent) = –p(ดื่ม)*log2 p(ดื่ม) + p(ไม่ดื่ม)*log2 p(ไม่ดื่ม)

= –[0.564 * log2(0.564) + 0.436 * log2(0.436)]

= –[0.564 * – 0.826 + 0.436 * –1.198]

= –[0.465 + 0.522]

= 0.987

entropy(ผล = ทุเลา/หาย) = –p(ดื่ม)*log2 p(ดื่ม) + p(ไม่ดื่ม)*log2 p(ไม่ดื่ม)

= –[0.564 * log2(0.564) + 0.436 * log2(0.436)]

= –[0.564 * – 0.826 + 0.436 * – 1.198]

= –[0.466 + 0.522]

= 0.988

entropy(ผล = เสียชีวิต) = –p(ดื่ม)*log2 p(ดื่ม) + p(ไม่ดื่ม)*log2 p(ไม่ดื่ม)

= –[0.585 * log2(0.585) + 0.415 * log2(0.415)]

= –[0.585 * – 0.773 + 0.415 * – 1.269]

= –[0.452 + 0.527]

= 0.979

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

= 0.987 – [0.516 * 0.988 + 0.483 *0.979]

= 0.987 – [0.510 + 0.473]

= 0.987 – 0.983

= 0.004

2. คำนวณค่า IG ของแอตทริบิวต์ของมาตรการและผลการรักษาต่อเพศ “หญิง” ได้ดังนี้

entropy(parent) = –p(ดื่ม)*log2 p(ดื่ม) + p(ไม่ดื่ม)*log2 p(ไม่ดื่ม)

= –[0.564 * log2(0.564) + 0.436 * log2(0.436)]

= –[0.564 * – 0.826 + 0.436 * –1.198]

= –[0.465 + 0.522]

= 0.987

entropy(ผล = ทุเลา/หาย) = –p(ดื่ม)*log2 p(ดื่ม) + p(ไม่ดื่ม)*log2 p(ไม่ดื่ม)

= –[0.564 * log2(0.564) + 0.436 * log2(0.436)]

= –[0.564 * – 0.826 + 0.436 * – 1.198]

= –[0.466 + 0.522]

= 0.988

entropy(ผล = เสียชีวิต) = –p(ดื่ม)*log2 p(ดื่ม) + p(ไม่ดื่ม)*log2 p(ไม่ดื่ม)

= –[0.645 * log2(0.645) + 0.355 * log2(0.355)]

= –[0.645 * – 0.633 + 0.355 * – 1.494]

= –[0.408 + 0.530]

= 0.938

IG(parent,child) = entropy(parent) – [p(ผล = ชาย) * entropy(ผล = ชาย) + p(ผล = หญิง) * entropy(ผล = หญิง)]

= 0.987 – [0.459 * 0.988 + 0.541 * 0.938]

= 0.987 – [0.453 + 0.507]

= 0.987 – 0.96

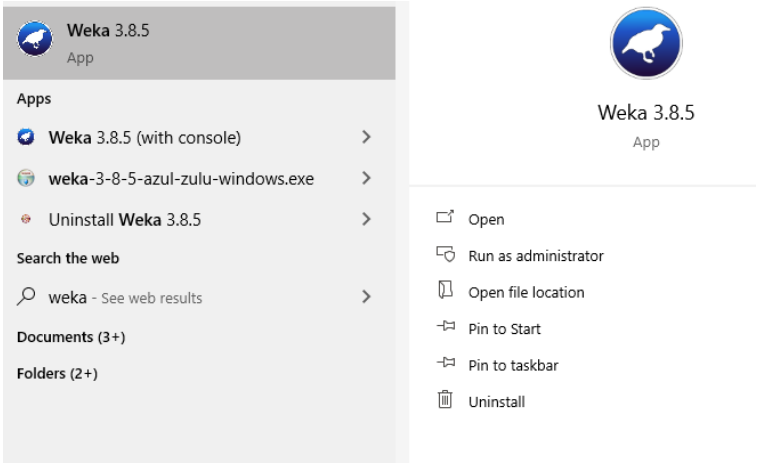
= 0.027

จากการคำนวณค่า IG ของแอตทริบิวต์มาตรการและผลการรักษาต่อเพศที่เป็นแอตทริบิวต์สุดท้าย คณะผู้จัดทำพบว่าผลการรักษา “ทุเลา/หาย” และ “เสียชีวิต” มีความสัมพันธ์กับเพศมากที่สุดที่ และพบว่าข้อมูลในแต่ละโหนดมีคำตอบเดียวกันแล้ว คือ ดื่มและไม่ดื่ม

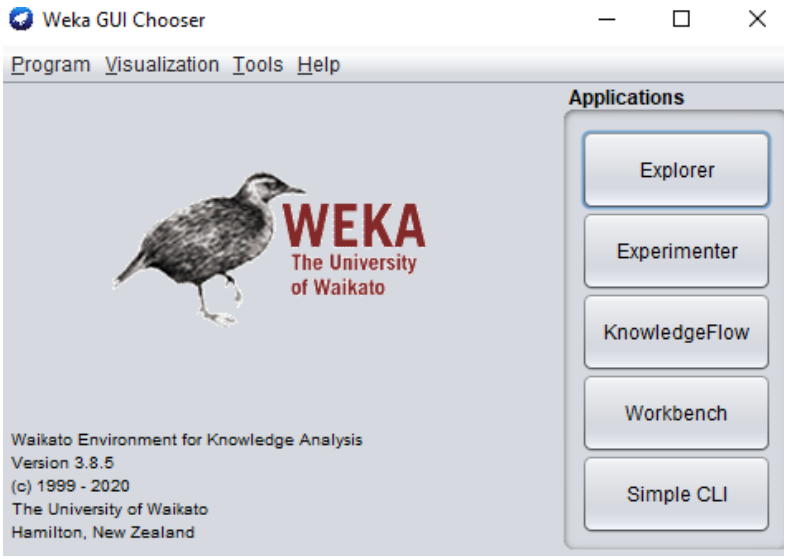
3.1.5 การประเมินผล (Evaluation) เป็นขั้นตอนก่อนนำผลลัพธ์ที่ได้จากขั้นตอนที่ 3 ไปใช้งาน ด้วยการวัดประสิทธิผลของผลลัพท์ที่ได้กับวัตถุประสงค์ที่ตั้งไว้ในขั้นตอนแรก ว่ามีนัยสำคัญหรือความน่าเชื่อถือมากน้อยเพียงใด ด้วยการประเมินผลจากโปรแกรมว่าถูกต้องหรือไม่

คณะผู้จัดทำได้ทำการทดสอบโมเดล เพื่อวัดประสิทธภาพที่ตรงกับความต้องการ ซึ่งเป็นการวัดประสิทธิภาพด้วยวิธี Self Consistency Test เหมาะสำหรับใช้ในการทอสอบประสิทธิภาพ เพื่อดูแนวโน้มของโมเดลที่สร้างขึ้นและเมื่อนำข้อมูลมาทดสอบ (Testing Data) กับโปรแกรมที่ผู้วิเคราะห์เลือกมาทอสอบกับข้อมูลที่ผ่านการวิเคราะห์ข้อมูลด้วยเทคนิค Data Mining จากการสร้างโมเดล Decision Tree จึงนำข้อมูลดังกล่าวมาทดสอบกับโปรแกรม Weka เวอร์ชั่น 3.8.5 ซึ่งมีขั้นตอน ดังนี้

ขั้นตอนที่ 1 เปิดโปรแกรม Weka ขึ้นมา

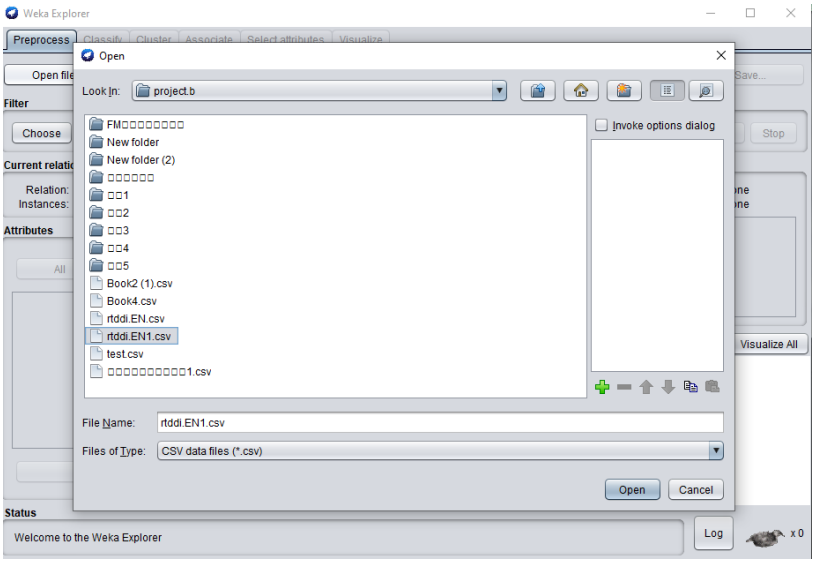


ภาพที่ 3.8 แสดงการเปิดโปรแกรม

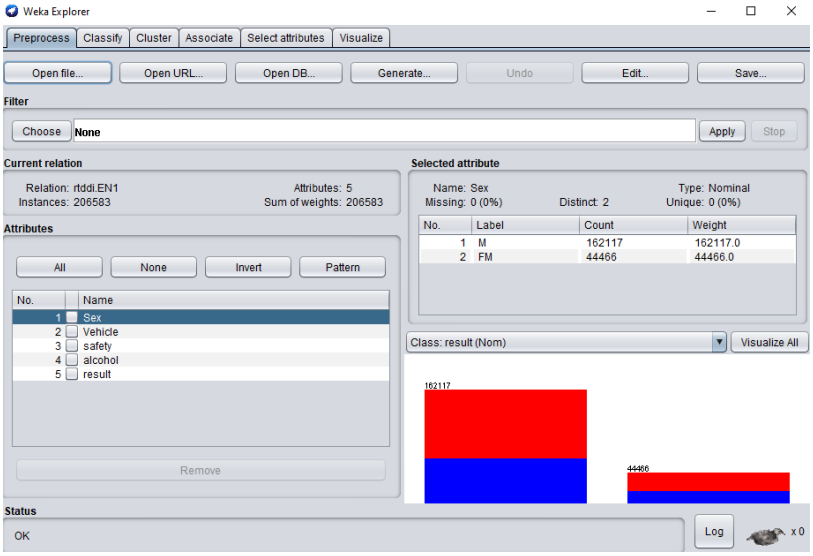


ภาพที่ 3.9 แสดงการเข้าหน้าจอโปรแกรม Weka 3.8.4

ขั้นตอนที่ 2 นำข้อมูลที่ได้รับจัดเตรียมไว้ โดยเลือกที่ Application > Explorer > Open file เลือกไฟล์ข้อมูลที่ต้องการนำมาทดสอบตามภาพที่ 3.10 และหลังจากนั้นโปรแกรมจะแสดงหน้าจอของข้อมูลตามภาพที่ 3.11

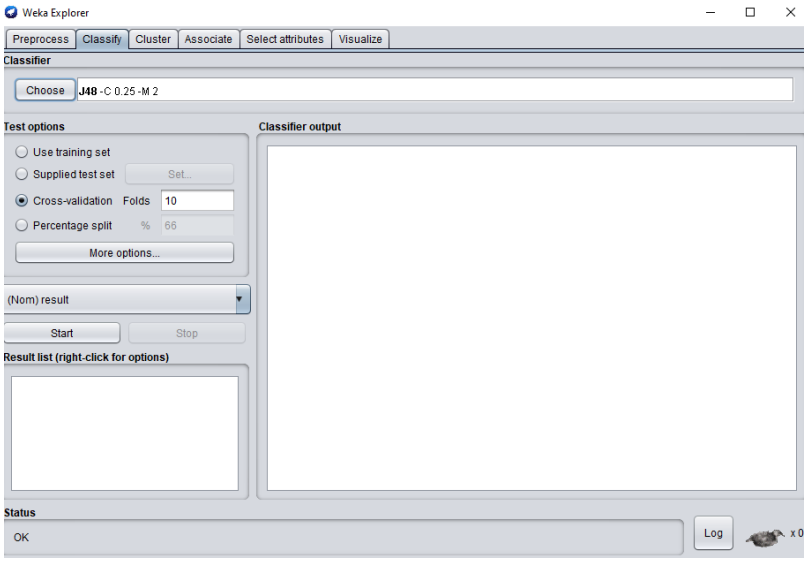


ภาพที่ 3.10 แสดงการนำข้อมูลเข้าในโปรแกรม Weka 3.8.4

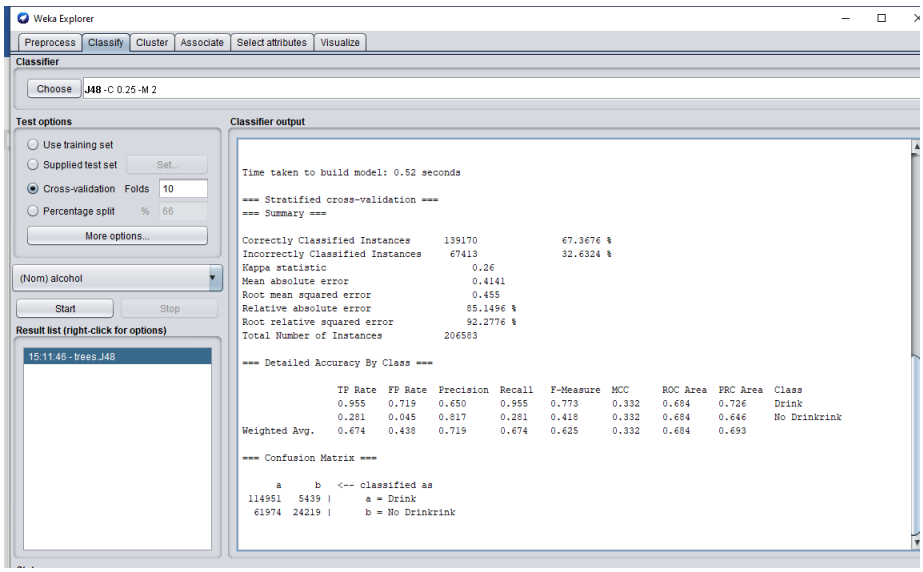


ภาพที่ 3.11 หน้าจอแสดงข้อมูลที่นำเข้าโปรแกรม Weka 3.8.4

ขั้นตอนที่ 3 การดำเนินการเลือกเทคนิคที่ต้องการ ได้แก่ เทคนิคการจัดกลุ่มของข้อมูล โดยเลือก Classification > Choose > tree เลือกเทคนิคที่ต้องการ ในที่นี้คณะผู้จัดทำเลือกใช้เทคนิค J48 จากนั้นให้กดปุ่ม Start ตามภาพที่ 3.12 จะแสดงผลลัพธ์ได้ตามภาพที่ 3.13



ภาพที่ 3.12 การเลือกเทคนิคที่ใช้ในการจำแนกข้อมูลแบบ Decision Tree: J48

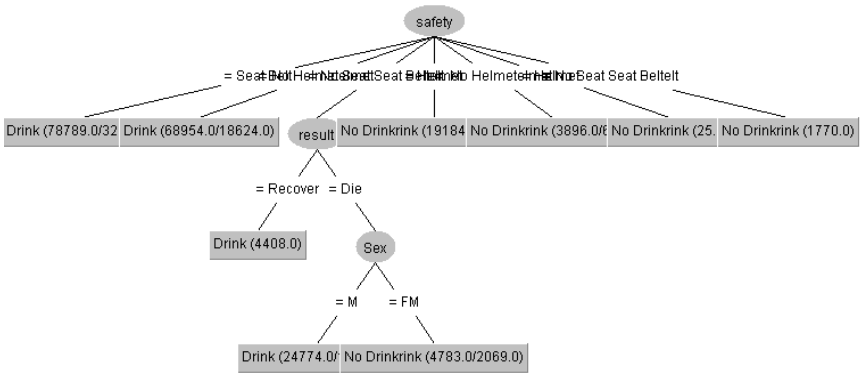


ภาพที่ 3.13 หน้าจอผลลัพธ์ของโมเดลการจำแนกข้อมูลแบบ Decision Tree: J48

จากผลการทดสอบพบว่าเทคนิค Decision Tree: J48 ให้ผลลัพธ์การจำแนกประเภท การดื่ม (Drink) และไม่ดื่ม (No Drink) มีความถูกต้องถึง 67.37 % และได้กฎจำนวน 9 กฎแสดงผลลัพธ์ที่เป็นกฎในลักษณะของต้นไม้การตัดสินใจที่มีกิ่งแตกออกมา ดังภาพที่ 3.13 และภาพที่ 3.14

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

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

safety = Seat Belt: Drink (78789.0/32905.0)
safety = No Helmetelmet: Drink (68954.0/18624.0)
safety = No Seat Seat Beltelt
|   result = Recover: Drink (4408.0)
|   result = Die
|   |   Sex = M: Drink (24774.0/10445.0)
|   |   Sex = FM: No Drinkrink (4783.0/2069.0)
safety = Helmet: No Drinkrink (19184.0/3308.0)
safety =  No Helmetelmet: No Drinkrink (3896.0/62.0)
safety = Helmet : No Drinkrink (25.0)
safety = No Seat Seat Beltelt : No Drinkrink (1770.0)

Number of Leaves  :     9

Size of the tree :      12
```

ภาพที่ 3.14 ผลลัพธ์จากการจำแนกกลุ่มแบบ Decision Tree: J48 ในโปรแกรม Weka



ภาพที่ 3.15 รูปแบบแผนภาพโมเดล Graph Decision Tree: J48 ในโปรแกรม Weka 3.8.4

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

กฎข้อที่ 1 IF safety = Seat Belt Then Alcohol = Drink หมายความว่า ถ้ามาตรการเป็นมาตรการที่ป้องกันกับการคาดเข็มขัดเกี่ยวเนื่องกัน ผลของการพิจารณาพบว่าประเภทของการดื่มสุรานั้นจะมีผลลัพธ์เป็น “ดื่ม”

กฎข้อที่ 2 IF safety = No Helmet Then Alcohol = Drink หมายความว่า ถ้ามาตรการเป็นมาตรการที่ป้องกันกับการไม่ใส่หมวกเกี่ยวเนื่องกัน ผลของการพิจารณาพบว่าประเภทของการดื่มสุรานั้นจะมีผลลัพธ์เป็น “ดื่ม”

กฎข้อที่ 3 IF safety = Helmet Then Alcohol = No Drink หมายความว่า ถ้ามาตรการเป็นมาตรการที่ป้องกันกับการใส่หมวกเกี่ยวเนื่องกัน ผลของการพิจารณาพบว่าประเภทของการดื่มสุรานั้นจะมีผลลัพธ์เป็น “ไม่ดื่ม”

กฎข้อที่ 4 IF safety = No Helmet Then Alcohol = No Drink หมายความว่า ถ้ามาตรการเป็นมาตรการที่ป้องกันกับการไม่ใส่หมวกเกี่ยวเนื่องกัน ผลของการพิจารณาพบว่าประเภทของการดื่มสุรานั้นจะมีผลลัพธ์เป็น “ไม่ดื่ม”

กฎข้อที่ 5 IF safety = Helmet Then Alcohol = No Drink หมายความว่า ถ้ามาตรการเป็นมาตรการที่ป้องกันกับการใส่หมวกเกี่ยวเนื่องกัน ผลของการพิจารณาพบว่าประเภทของการดื่มสุรานั้นจะมีผลลัพธ์เป็น “ไม่ดื่ม”

กฎข้อที่ 6 IF safety = No Seat Belt Then Alcohol = No Drink หมายความว่า ถ้ามาตรการเป็นมาตรการที่ป้องกันกับการไม่คาดเข็มขัดเกี่ยวเนื่องกัน ผลของการพิจารณาพบว่าประเภทของการดื่มสุรานั้นจะมีผลลัพธ์เป็น “ไม่ดื่ม”

กฎข้อที่ 7 IF safety = No Seat Belt And Result = Recover Then Alcohol = Drink หมายความว่า ถ้ามาตรการเป็นมาตรการที่ป้องกันกับการไม่คาดเข็มขัดและผลของการรักษากับผลเป็น ทุเลา/หาย นั้นเกี่ยวเนื่องกัน ผลของการพิจารณาพบว่า ประเภทของการดื่มสุรานั้นจะมีผลลัพธ์เป็น “ดื่ม”

กฎข้อที่ 8 IF safety = No Seat Belt And Result = Die And Sex = M Then Alcohol = Drink หมายความว่า ถ้ามาตรการเป็นมาตรการที่ป้องกันกับการไม่คาดเข็มขัดและผลของการรักษากับผลเป็น ตาย และเพศนั้นเป็นเพศชาย นั้นเกี่ยวเนื่องกัน ผลของการพิจารณาพบว่า ประเภทของการดื่มสุรานั้นจะมีผลลัพธ์เป็น “ดื่ม”

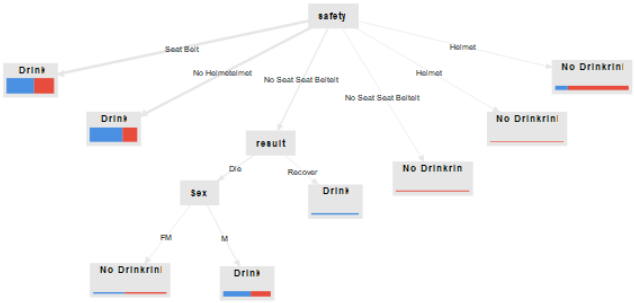
กฎข้อที่ 9 IF safety = No Seat Belt And Result = Die And Sex = FM Then Alcohol = No Drink หมายความว่า ถ้ามาตรการเป็นมาตรการที่ป้องกันกับการไม่คาดเข็มขัดและผลของการรักษากับผลเป็น ตาย และเพศนั้นเป็นเพศหญิง นั้นเกี่ยวเนื่องกัน ผลของการพิจารณาพบว่า ประเภทของการดื่มสุรานั้นจะมีผลลัพธ์เป็น “ไม่ดื่ม”

ทางคณะผู้จัดทำยังได้นำโมเดลของข้อมูลที่เลทอกใช้ มาเปรียบเทียบกับโมเดลที่สร้างด้วยโปรแกรม RapidMiner Studio ดังภาพที่ 3.15 และภาพที่ 3.16 ซึ่งทางคณะผู้จัดทำพบว่าได้ผลลัพธ์จากการจำแนกกลุ่มและรูปแบบแผนภาพโมเดลที่ตรงกันสูง

Tree

```
safety = Helmet: No Drinkrink {Drink=3308, No Drinkrink=15876}
safety = Helmet : No Drinkrink {Drink=0, No Drinkrink=25}
safety = No Helmetelmet: Drink {Drink=50392, No Drinkrink=22458}
safety = No Seat Seat Beltelt
|   result = Die
|   |   Sex = FM: No Drinkrink {Drink=2069, No Drinkrink=2714}
|   |   Sex = M: Drink {Drink=14329, No Drinkrink=10445}
|   result = Recover: Drink {Drink=4408, No Drinkrink=0}
safety = No Seat Seat Beltelt : No Drinkrink {Drink=0, No Drinkrink=1770}
safety = Seat Belt: Drink {Drink=45884, No Drinkrink=32905}
```

ภาพที่ 3.16 ผลลัพธ์จากการจำแนกกลุ่มแบบ Decision Tree ของโปรแกรม RapidMiner



ภาพที่ 3.17 รูปแบบแผนภาพโมเดล Graph Decision Tree ในโปรแกรม RapidMiner

หลักจากคณะผู้จัดทำเลือกการทดสอบประสิทธิภาพของ Model ด้วยวิธี Self Consistency Test หรือเรียกว่า Use Training Set เป็นวิธีการที่นำข้อมูลที่ใช้ในการสร้างโมเดล (Model) และข้อมูลที่ใช้ในการทดสอบโมเดลเป็นข้อมูลชุดเดียวกัน คือข้อมูลผู้เสียชีวิตจากอุบัติเหตุทางถนน ปี 2554 – 2563 ที่ได้ทำการคัดเลือกมาทั้งหมด ซึ่งคณะผู้จัดทำเลือกโปรแกรมที่ใช้นำเสนอ คือ โปรแกรม Weka พบว่าการทดสอบประสิทธิภาพโมเดล Decision Tree (J48) พิจารณาได้ว่า โมเดลที่ถูกสร้างขึ้นมีค่าความถูกต้องเฉลี่ยในทุกโมเดลเท่ากับ 67.3676% มีค่าการทำนายข้อมูลไม่ถูกต้องเท่ากับ 32.6324% และมีค่าความเคลื่อนเท่ากับ 0.4495 และเมื่อพิจารณาส่วนค่า Confusion Matrix ในภาพที่ 3.16 จะพบว่าการหาค่าของข้อมูลค่าจริงกับจำนวนข้อมูลจากการทำนายแบ่งตามประเภทของการเสียชีวิตและการหายจากการบาดเจ็บและนำมาหาค่าเฉลี่ยรวมของทุก Class ได้ค่าเฉลี่ยรวมเท่ากับ 67.4 % มีผลลัพธ์ตรงกันอยู่ในระดับค่อนข้างดี สามารถนำโมเดลไปใช้งานได้

ดังนั้น คณะผู้จัดทำจึงเลือกใช้โมเดล Decision Tree : J48 ในการนำไปหาแนวทางการป้องกันและรักษามาตรการและให้คำแนะนำในการปฏิบัติตัวระหว่างเดินทางในช่วงเทศการปีใหม่ เพราะมีค่าความถูกต้องของโมเดลและค่าเฉลี่ย Confusion Matrix จากทุกประเภทของการดื่มและไม่ได้ดื่มอยู่ในระดับที่ค่อนข้างดี

```
=== Summary ===

Correctly Classified Instances      139170               67.3676 %
Incorrectly Classified Instances     67413               32.6324 %
Kappa statistic                          0.26
Mean absolute error                      0.4141
Root mean squared error                  0.455
Relative absolute error                 85.1496 %
Root relative squared error             92.2776 %
Total Number of Instances           206583

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.955    0.719    0.650      0.955    0.773      0.332    0.684     0.726     Drink
                 0.281    0.045    0.817      0.281    0.418      0.332    0.684     0.646     No Drinkrink
Weighted Avg.    0.674    0.438    0.719      0.674    0.625      0.332    0.684     0.693

=== Confusion Matrix ===

      a      b   <-- classified as
 114951   5439 |      a = Drink
  61974  24219 |      b = No Drinkrink
```

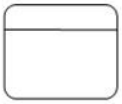
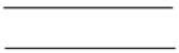
ภาพที่ 3.18 ผลลัพธ์จากการจำแนกกลุ่มแบบ Dacision Tree : J48 ในโปรแกรม Weka

3.2 แผนภาพกระแสข้อมูล Data Flow Diagram

แผนภาพกระแสข้อมูล (DFD) ย่อมาจาก Data Flow Diagramเป็นเครื่องมือเชิงโครงสร้างที่ใช้บรรยายภาพรวมของระบบโดยแสดงขั้นตอนการทำงานของระบบหรือโพรเซส(process) ระบุแหล่งกำเนิดของข้อมูล การไหลของข้อมูล ปลายทางข้อมูล การเก็บข้อมูลและการประมวลผลข้อมูล โดยจะช่วยแสดงแผนภาพ ว่าข้อมูลมาจากไหน จะไปไหน เก็บข้อมูลไว้ที่ไหน มีอะไรเกิดขึ้นกับข้อมูลระหว่างทางเรียกว่าแผนภาพกระแสข้อมูลหรือ แผนภาพแสดงความเคลื่อนไหวของข้อมูล

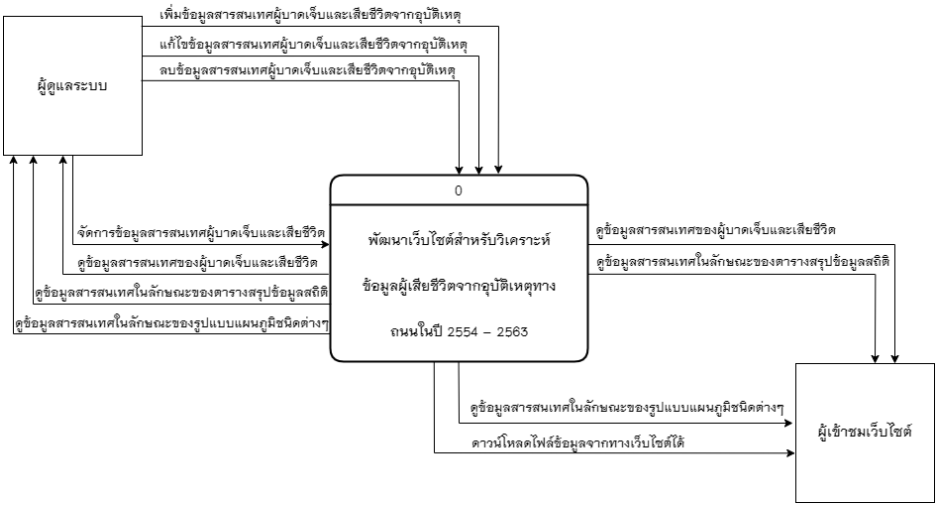
สัญลักษณ์ที่ใช้ในการเขียนแผนภาพกระแสข้อมูลนั้น ประกอบด้วย 4สัญลักษณ์ คือ การประมวลผล (Process) เส้นทางการไหลของข้อมูล (Data Flows) ตัวแทนข้อมูล (External Agents) และแหล่งจัดเก็บข้อมูล (Data Store) โดยได้มีการศึกษาคิดค้นพัฒนาวิธีการอยู่หลายแบบ

สัญลักษณ์ที่ใช้ในแผนภาพกระแสข้อมูล

DeMarco & Yourdon	Gane & Sarson	ความหมาย
		Process : ขั้นตอนการทำงานภายในระบบ
		Data Store : แหล่งข้อมูลสามารถเป็นได้ทั้งไฟล์ข้อมูลและฐานข้อมูล (File or Database)
		External Agent : ปัจจัยหรือสภาพแวดล้อมที่มีผลกระทบต่อระบบ
		Data Flow : เส้นทางการไหลของข้อมูล แสดงทิศทางของข้อมูลจากขั้นตอนการทำงานหนึ่งไปยังอีกขั้นตอนหนึ่ง

ภาพที่ 3.19 สัญลักษณ์ที่ใช้ในแผนภาพกระแสข้อมูล

แผนภาพบริบท (Context Diagram) จะแสดงภาพโดยรวมระบบของการพัฒนาเว็บไซต์สำหรับวิเคราะห์ข้อมูลผู้เสียชีวิตจากอุบัติเหตุทางถนนในปี 2554 – 2563 ดังรูปภาพ

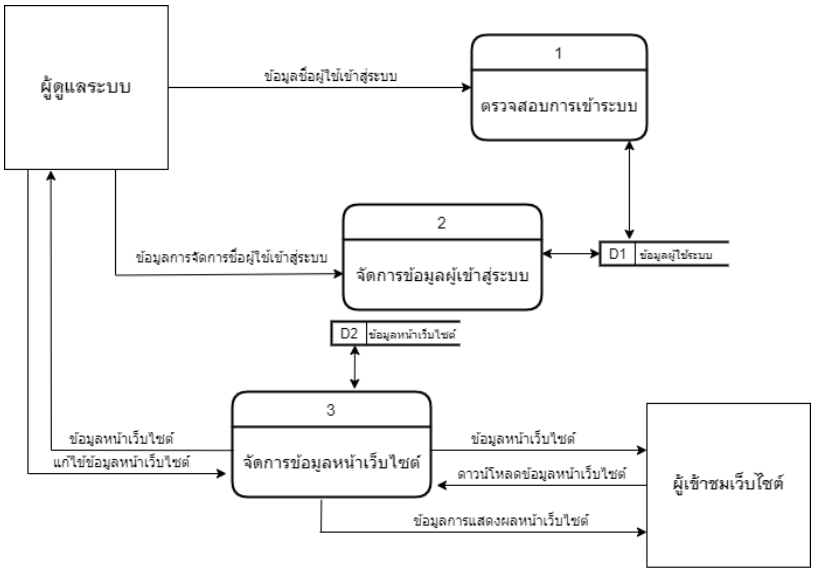


ภาพที่ 3.20 แผนภาพบริบท (Context Diagram)

จากรูปภาพที่ 3. เป็นแผนภาพบริบทระบบของการพัฒนาเว็บไซต์สำหรับวิเคราะห์ข้อมูลผู้เสียชีวิตจากอุบัติเหตุทางถนนในปี 2554 – 2563 โดยสามารถแบ่งผู้ใช้ออกเป็น 2 ประเภท ดังนี้

1) **ผู้ดูแลระบบ (System Administrator)** สามารถเพิ่ม ลบ และแก้ไขข้อมูลสารสนเทศผู้บาดเจ็บและเสียชีวิตจากอุบัติเหตุบนเว็บไซต์ได้ สามารถจัดการข้อมูลสารสนเทศผู้บาดเจ็บและเสียชีวิตจากอุบัติเหตุบนเว็บไซต์ได้ สามารถดูข้อมูลสารสนเทศของผู้บาดเจ็บและเสียชีวิตจากอุบัติเหตุทางถนน สามารถดูข้อมูลสารสนเทศในลักษณะของตารางสรุปข้อมูลสถิติผู้บาดเจ็บและผู้เสียชีวิตทางถนน และสามารถดูข้อมูลสารสนเทศในลักษณะของรูปแบบแผนภูมิชนิดต่าง ๆได้

2) **ผู้เยี่ยมชมเว็บไซต์ (Website visitor)** สามารถดูข้อมูลสารสนเทศของผู้บาดเจ็บและเสียชีวิตจากอุบัติเหตุทางถนน สามารถดูข้อมูลสารสนเทศในลักษณะของตารางสรุปข้อมูลสถิติผู้บาดเจ็บและผู้เสียชีวิตทางถนน สามารถดูข้อมูลสารสนเทศในลักษณะของรูปแบบแผนภูมิชนิดต่าง ๆได้ และสามารถดาวน์โหลดไฟล์ข้อมูลจากทางเว็บไซต์ได้



ภาพที่ 3.21 แผนภาพกระแสข้อมูลระดับที่ 0 (Data Flow Diagram Level 0)

คำอธิบายกระบวยการ

1. DFD Number 1 ตรวจสอบการเข้าระบบ

Process Description	
System	ระบบจัดการข้อมูลสารสนเทศผู้บาดเจ็บและเสียชีวิตจากอุบัติเหตุ
DFD Number	1
Process Name	ตรวจสอบการเข้าระบบ
Input Data Flow	ข้อมูลผู้ใช้และรหัสผ่าน, ข้อมูลผู้ใช้ระบบ
Output Data Flow	–
Data Store Used	ข้อมูลผู้ใช้ระบบ
Description	เป็นกระบวนการตรวจสอบ ข้อมูลและรหัสผ่านข้อผู้ใช้ สำหรับผู้ดูแลระบบ

2. DFD Number 2 จัดการข้อมูลผู้เข้าสู่ระบบ

Process Description	
System	ระบบจัดการข้อมูลสารสนเทศผู้บาดเจ็บและเสียชีวิตจากอุบัติเหตุ
DFD Number	2
Process Name	จัดการข้อมูลผู้เข้าสู่ระบบ
Input Data Flow	ข้อมูลการจัดการข้อมูลผู้ใช้ระบบ
Output Data Flow	ข้อมูลผู้ใช้ระบบ
Data Store Used	ข้อมูลผู้ใช้ระบบ
Description	เป็นกระบวนการตรวจสอบ ข้อมูลและรหัสผ่านข้อผู้ใช้ สำหรับผู้ดูแลระบบ

3. DFD Number 3 ตรวจสอบการเข้าสู่ระบบ

Process Description	
System	ระบบจัดการข้อมูลสารสนเทศผู้บาดเจ็บและเสียชีวิตจากอุบัติเหตุ
DFD Number	2
Process Name	จัดการข้อมูลหน้าเว็บไซต์
Input Data Flow	ข้อมูลหน้าเว็บไซต์,ข้อมูลภายในเว็บไซต์
Output Data Flow	ข้อมูลหน้าเว็บไซต์
Data Store Used	ข้อมูลหน้าเว็บไซต์
Description	เป็นกระบวนการจัดการหน้าเว็บไซต์ ของผู้ดูแลระบบรวมไปถึงผู้เข้าชมเว็บไซต์แต่จะมีแค่ ผู้ดูแลระบบที่สามารถแก้ไขระบบภายในเว็บไซต์ได้

3.3 ออกแบบหน้าเว็บไซต์

3.3.1 การออกแบบ Wireframe หน้าจอเว็บไซต์

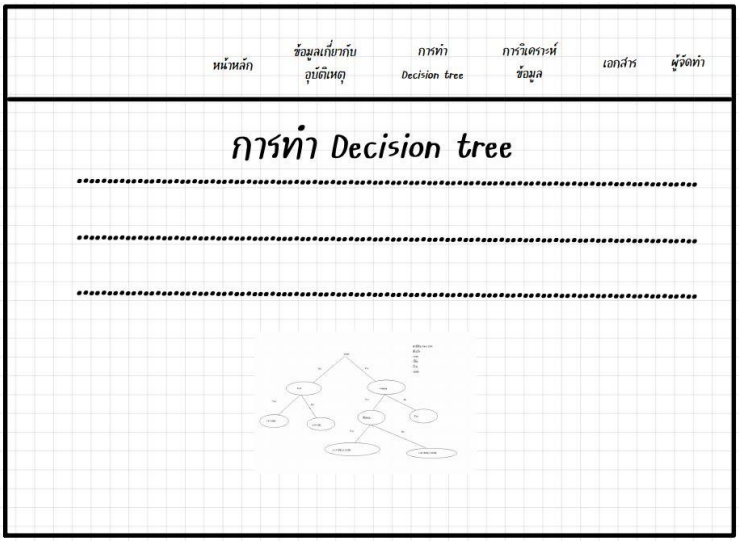
1) หน้าเว็บไซต์แสดงเมนูต่าง ๆบนหน้าเว็บไซต์

ภาพที่ 3.22 Wireframe แสดงหน้าเว็บไซต์แสดงเมนูต่าง ๆ บนหน้าเว็บไซต์

2) หน้าแสดงข้อมูลทั่วไปเกี่ยวกับการเกิดอุบัติเหตุบทท้องถนน

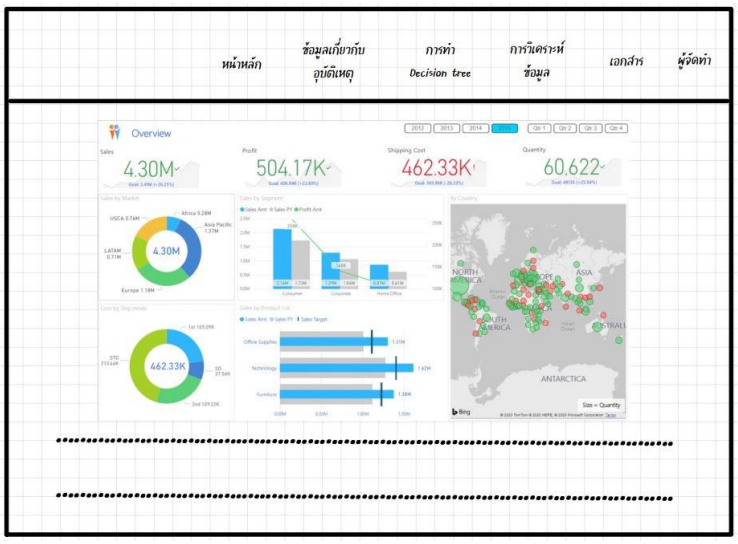
ภาพที่ 3.23 Wireframe ข้อมูลทั่วไปเกี่ยวกับการเกิดอุบัติเหตุบทท้องถนน

3) หน้าแสดงการวิเคราะห์ข้อมูลโดยใช้ Decision Tree ที่จะแสดงขั้นตอนการสร้างโมเดล โดยการใช้โปรแกรม Weka



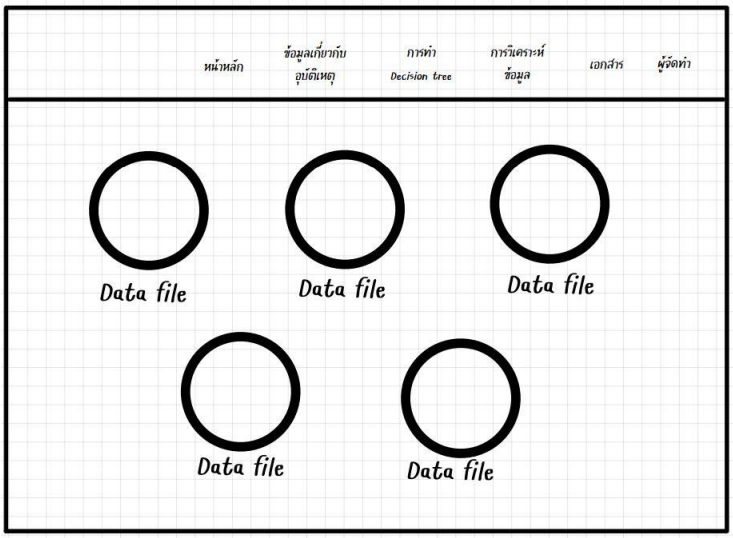
ภาพที่ 3.24 แสดงการวิเคราะห์ข้อมูลโดยใช้ Decision Tree

4) หน้าแสดงผลของข้อมูลในรูปตาราง และแดชบอร์ดผู้ประสบอุบัติเหตุบนท้องถนน



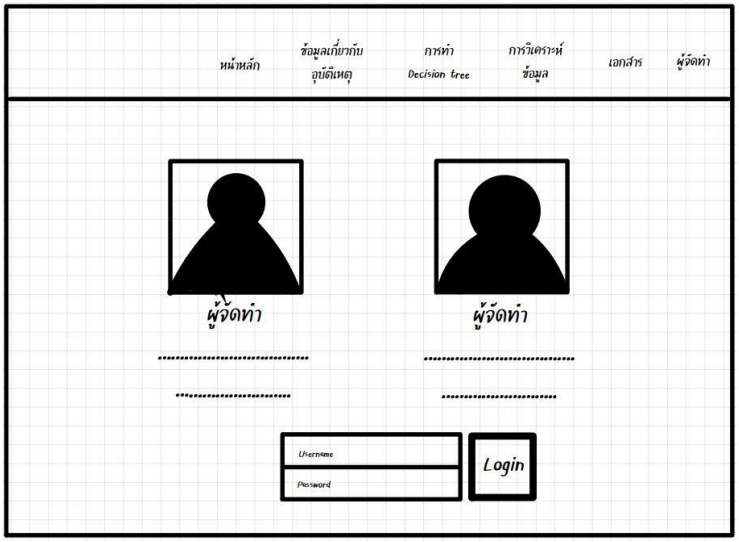
ภาพที่ 3.25 แสดงแดชบอร์ดผู้ประสบอุบัติเหตุบนท้องถนน

5) หน้าแสดง File ข้อมูลต่าง ๆ ที่สามารถดาวน์โหลดได้



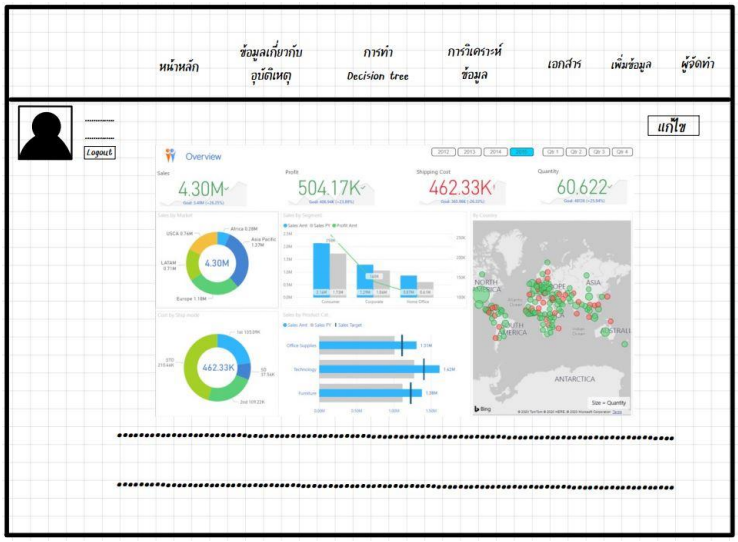
ภาพที่ 3.26 แสดง File ข้อมูลต่าง ๆ ที่สามารถดาวน์โหลดได้

6) หน้าแสดง ข้อมูลส่วนตัวผู้จัดทำ



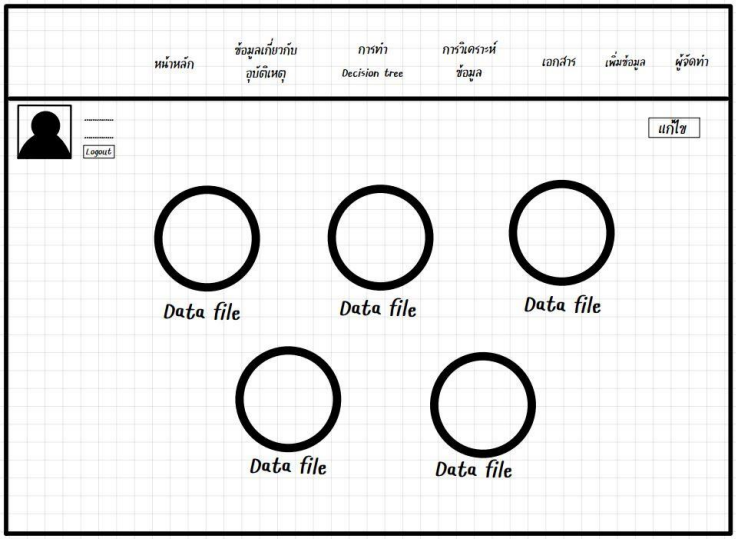
ภาพที่ 3.27 แสดงข้อมูลส่วนตัวผู้จัดทำ

7) ส่วนของ Admin หน้าแสดงหน้าแคชบอร์ดที่สามารถแก้ไขข้อมูลได้



ภาพที่ 3.28 แสดงหน้าแคชบอร์ดที่สารถมารถแก้ไขได้

8) ส่วนของ Admin หน้าแสดง File ข้อมูลต่างที่สามารถแก้ไขข้อมูลได้



ภาพที่ 3.29 แสดง File ข้อมูลต่างที่สามารถแก้ไขข้อมูลได้

9) ส่วนของ Admin หน้าแสดงการเพิ่มข้อมูล เช่น การอัพโหลดไฟล์ และการกรอกข้อมูล

ภาพที่ 3.30 แสดงการเพิ่มข้อมูล

3.4 บทสรุป

จากวิธีการดำเนินงานโครงงานในข้างต้นทั้งหมดนี้ ทางคณะผู้จัดทำได้ทำการแสดงวิธีในการจัดการกับข้อมูลของผู้เสียชีวิตจากอุบัติเหตุทางถนน ปี 2554 – 2563 ด้วยขั้นตอนการวิเคราะห์ข้อมูลด้วย CRISP-DM อย่างละเอียดรวมไปถึงการสร้างโมเดล Decision Tree จากโปรแกรมที่ใช้เลือกทำเหมืองข้อมูลเพื่อนำเสนอ คือ โปรแกรม Weka ในการสร้างโมเดล Decision Tree และตรวจสอบข้อมูลผ่านทางโปรแกรม RapidMiner Studio เพื่อความแม่นยำและประเมอนประสิทธิภาพของโมเดล ซึ่งมีค่าความถูกต้องเฉลี่ยในทุกโมเดลเท่ากับ 67.37% เป็นการทดสอบประสิทธิภาพของ Model ด้วยวิธี Self Consistency Test พบว่ามีผลลัพธ์ตรงกันอยู่ในระดับค่อนข้างดีและสามารถนำโมเดลไปใช้งานได้และได้ผลลัพธ์ที่เหมือนกัน เมื่อเปรียบเทียบโมเดลกับโปรแกรม RapidMiner Studio ทางคณะผู้จัดทำได้นำข้อมูลสารสนเทศนั้นมาทำการแสดงผลแบบ Visualization โดยใช้โปรแกรม Google data studio และออกแบบ Wireframe ของเว็บไซต์ที่จะเผยแพร่บน Web browser