การพัฒนาเว็บไซต์สำหรับวิเคราะห์ข้อมูลผู้เสียชีวิต
จากอุบัติเหตุทางถนน ปี 2554 – 2563

นายจิรกิตติ์ จินะกัณฑ์ รหัสนักศึกษา 60541207073-8

นางสาวสุรางคณา หมื่นคำแปง รหัสนักศึกษา 60541207095-1

หลักสูตรระบบสารสนเทศทางธุรกิจ

สาขาบริหารธุรกิจ คณะบริหารธุรกิจและศิลปศาสตร์

มหาวิทยาลัยเทคโนโลยีราชมลคลล้านนา

ปีการศึกษา 2564

การพัฒนาเว็บไซต์สำหรับวิเคราะห์ข้อมูลผู้เสียชีวิต
จากอุบัติเหตุทางถนน ปี 2554 – 2563

นายจิรกิตติ์ จินะกัณฑ์ รหัสนักศึกษา 60541207073-8

นางสาวสุรางคณา หมื่นคำแปง รหัสนักศึกษา 60541207095-1

หลักสูตรระบบสารสนเทศทางธุรกิจ

สาขาบริหารธุรกิจ คณะบริหารธุรกิจและศิลปศาสตร์

มหาวิทยาลัยเทคโนโลยีราชมลคลล้านนา

ปีการศึกษา 2564

ชื่อโครงการ การพัฒนาเว็บไซต์สำหรับวิเคราะห์ข้อมูลผู้เสียชีวิตจากอุบัติเหตุทางถนน ปี 2554 – 2563

โดย นายจิรกิตติ์ จินะกัณฑ์ รหัสนักศึกษา 60541207073-8

นางสาวสุรางคณา หมื่นคำแปง รหัสนักศึกษา 60541207095-1

หลักสูตร ระบบสารสนเทศทางธุรกิจ สาขาบริหารธุรกิจ

คณะบริหารธุรกิจและศิลปศาสตร์

อาจารย์ที่ปรึกษา ผู้ช่วยศาสตราจารย์ ดร.ชัชฎาพร ปุกแก้ว

หลักสูตรระบบสารสนเทศทางธุรกิจ สาขาบริหารธุรกิจ คณะบริหารธุรกิจและศิลปศาสตร์มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา อนุมัติให้โครงกรนี้ เป็นส่วนหนึ่งของการศึกษาหลักสูตร บริหารธุรกิจบัณฑิต

(.....)

(อาจารย์นรินทร์ จิวิตัน)

หัวหน้าหลักสูตรระบบสารสนเทศทางธุรกิจ

คณะกรรมการสอบโครงงาน

.....ประธานกรรมการ

(ผู้ช่วยศาสตราจารย์ ดร.ชัชฎาพร ปุกแก้ว)

.....กรรมการ

ลงนามด้วยลายมือชื่ออิเล็กทรอนิกส์ 19 ตุลาคม 2564 เวลา 16.20 น.

(อาจารย์นพณัฐ วรรณภีร์)

.....กรรมการ

(อาจารย์พรพิพัฒน์ ทองปรอน)

© ลิขสิทธิ์ของหลักสูตรระบบสารสนเทศทางธุรกิจ
สาขาบริหารธุรกิจ คณะบริหารธุรกิจและศิลปศาสตร์

ชื่อโครงการ การพัฒนาเว็บไซต์สำหรับวิเคราะห์ข้อมูลผู้เสียชีวิตจากอุบัติเหตุทางถนน ปี 2554 – 2563

โดย นายจิรกิตติ์ จินะกัณฑ์ รหัสนักศึกษา 60541207073-8
นางสาวสุรางคณา หมื่นคำแปง รหัสนักศึกษา 60541207095-1

อาจารย์ที่ปรึกษา ผู้ช่วยศาสตราจารย์ ดร.ชัชฎาพร ปุกแก้ว

หลักสูตร ระบบสารสนเทศทางธุรกิจ สาขาบริหารธุรกิจ
คณะบริหารธุรกิจและศิลปศาสตร์

ปีการศึกษา 2564

บทคัดย่อ

การจัดทำโครงงานในครั้งนี้มีวัตถุประสงค์ เพื่อวิเคราะห์ข้อมูลผู้เสียชีวิตจากอุบัติเหตุทางถนน ปี 2554 – 2563 โดยคณะผู้จัดทำได้เลือกใช้ขั้นตอนกระบวนการวิเคราะห์ข้อมูลด้วย CRISP-DM (Cross Industry Standard Process for Data Mining) จากเทคนิคการทำเหมืองข้อมูลแบบ Classification ที่จะสร้างกฎด้วยการสร้างโมเดล Decision Tree เพื่อจัดกลุ่มข้อมูลที่มีความสัมพันธ์กัน โดยใช้โปรแกรมที่ใช้ทำเหมืองข้อมูล คือ โปรแกรม Weka 3.8.4 พบว่าเทคนิค Decision Tree: J48 ให้ผลลัพธ์การจำแนกประเภท ดื่ม (Drink) และไม่ดื่ม (No Drink) มีความถูกต้องถึง 67.37 % และได้กฎจำนวน 9 กฎซึ่งคณะผู้จัดทำได้จัดทำเว็บไซต์สำหรับเผยแพรข้อมูลสารสนเทศนี้บน Web browser

โดยเว็บไซต์ที่สร้างขึ้น เป็นลักษณะของการเผยแพร่ข้อมูลที่ได้จากการวิเคราะห์ข้อมูลและการแสดงผลในรูปแบบของแผนภาพโดยใช้ Google Data Studio ซึ่งเว็บไซต์นี้ใช้ภาษา HTML CSS PHP และ JavaScript และมีส่วนประกอบต่าง ๆ

จากการวิเคราะห์ข้อมูล และสร้างเว็บไซต์ พบว่าเว็บไซต์ที่ได้สามารถนำมาประยุกต์ใช้งานได้ตรงตามวัตถุประสงค์ที่ตั้งไว้ ทำให้เกิดประสิทธิภาพในการศึกษาค้นคว้าข้อมูลตามที่ผู้ใช้ต้องการ สรุปได้ว่าในอนาคตสามารถนำข้อมูลจากเว็บไซต์ไปปรับปรุงเพิ่มเติมต่อได้ เพื่อให้เกิดประโยชน์กับความปลอดภัยของบุคคลทั่วไปให้ได้ประสิทธิภาพมากขึ้นต่อไป

กิตติกรรมประกาศ

การพัฒนาเว็บไซต์สำหรับวิเคราะห์ข้อมูลผู้เสียชีวิตจากอุบัติเหตุทางถนน ปี 2554 – 2563 ฉบับนี้สำเร็จลุล่วงไปได้ด้วยดี ด้วยความช่วยเหลือจากบุคคลหลาย ๆ ฝ่ายที่ให้คำแนะนำ คำปรึกษา และข้อเสนอแนะต่าง ๆ อันเป็นประโยชน์อย่างยิ่งต่อการดำเนินการในครั้งนี้จึงขอกล่าวขอบคุณ ดังนี้

ครั้งนี้ สำเร็จลุล่วงด้วยความอนุเคราะห์จากคณาจารย์ผู้มีพระคุณหลายท่าน โดยเฉพาะอย่างยิ่ง ผู้ช่วยศาสตร์จารย์ ดร.ชัชฎาพร ปุกแก้ว ที่ปรึกษาโครงงานที่ได้สละเวลาอันมีค่าได้ให้ ความรู้ คำแนะนำ ข้อเสนอแนะที่มีคุณค่าต่อการศึกษา ตลอดจนช่วยแก้ไขข้อบกพร่องต่าง ๆ ในการ ทำโครงงานฉบับนี้ ให้มีความถูกต้อง สมบูรณ์แบบตลอดระยะเวลาที่ทำโครงงาน ผู้จัดทำจึงขอกราบ ขอบพระคุณเป็นอย่างสูงไว้ ณ โอกาสนี้

ขอกราบขอบพระคุณ คณาจารย์ประจำหลักสูตรระบบสารสนเทศทางธุรกิจ สาขาบริหารธุรกิจ คณะบริหารธุรกิจและศิลปศาสตร์ มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนาเป็นอย่างยิ่ง ที่ได้สละเวลาอันมีค่าในการให้ความรู้ คำแนะนำต่าง ๆ อันเป็นแนวทางให้กับผู้จัดทำโครงงาน ตลอดจนโครงงานสำเร็จลุล่วง ถูกต้อง สมบูรณ์

สุดท้ายนี้ ผู้จัดทำโครงการขอกล่าวขอบพระคุณบิดา มารดา เป็นอย่างสูง ที่เป็นผู้อุปการะ เลี้ยงดู และสนับสนุนให้โอกาสในทุก ๆ ด้าน ขอขอบพระคุณอาจารย์ทุกท่านที่ให้วิชาความรู้มาจนกระทั่งมีวันนี้ และขอบคุณเพื่อน ๆ ทุกคน ที่ให้ความช่วยเหลือ ให้คำแนะนำ และให้กำลังใจที่ดีเสมอมาและทุก ๆ ท่านที่ไม่ได้เอ่ยนามมาที่ช่วยเหลือจนกระทั่งโครงการฉบับนี้สำเร็จลุล่วงไปได้

คณะผู้จัดทำ

สารบัญ

	หน้า
บทคัดย่อ	ก
กิตติกรรมประกาศ	ข
สารบัญ	ค
สารบัญตาราง	จ
สารบัญรูปภาพ	ฉ
บทที่ 1 บทนำ	1
1.1 ความสำคัญและที่มาของปัญหา	1
1.2 วัตถุประสงค์	3
1.3 ประโยชน์ที่จะได้รับ	3
1.4 ขอบเขต	3
1.5 เครื่องมือที่ใช้ในการพัฒนาโปรแกรม	5
1.6 สถานที่ใช้ในการดำเนินการศึกษาและรวบรวมข้อมูล	6
1.7 ระยะเวลาในการดำเนินการ	7
บทที่ 2 แนวคิด ทฤษฎี เครื่องมือและวรรณกรรมที่เกี่ยวข้อง	8
2.1 แนวคิด	9
2.2 ทฤษฎี	17
2.3 เครื่องมือในการออกแบบและวิเคราะห์ข้อมูล	28
2.4 วรรณกรรมที่เกี่ยวข้อง	35
2.5 บทสรุป	37
บทที่ 3 วิธีการดำเนินงานโครงงาน	38
3.1 การวิเคราะห์ข้อมูลด้วย CRISP–DM	38
3.2 แผนภาพกระแสข้อมูล Data Flow Diagram	63
3.3 ออกแบบหน้าเว็บไซต์	67
3.4 บทสรุป	71

สารบัญ (ต่อ)

	หน้า
บทที่ 4 ผลการดำเนินงาน	72
4.1 ผลการดำเนินงาน	72
4.2 การอภิปรายผล	79
4.3 บทสรุป	87
บทที่ 5 สรุปและข้อเสนอแนะ	88
5.1 บทสรุปโครงงาน	88
5.2 ข้อจำกัดของเว็บไซต์	89
5.3 ปัญหาและอุปสรรคของโครงงาน	89
บรรณานุกรม	91
ภาคผนวก ก	94
ภาคผนวก ข	103
ภาคผนวก ค	107
ประวัติผู้จัดทำ	109

สารบัญตาราง

ตารางที่		หน้า
1.1	ระยะเวลาในการดำเนินการ	7
3.1	สมการการหาความสัมพันธ์ของแอตทริบิวต์	45
4.1	แสดงข้อมูลการประมวลผลด้านการออกแบบ (Design)	84
4.2	ตารางแสดงการประมวลผลด้านคุณภาพข้อมูล (Information Quality)	85
4.3	แสดงการประเมินผลด้านประโยชน์และการนำไปใช้ (Perceived Usefulness)	86

สารบัญภาพ

ภาพที่		หน้า
2.1	แผนภูมิแท่ง (Bar Chart)	13
2.2	แผนภูมิเส้น (Line Charts)	13
2.3	แผนภูมิวงกลม (Pie Charts)	14
2.4	แผนภูมิโดนัท (Doughnut Charts)	14
2.5	แผนภูมิพื้นที่ (Area Charts)	15
2.6	แผนภูมิเรดาร์ (Radar Charts)	15
2.7	แผนภูมิต้นไม้ (Tree Maps)	16
2.8	แผนภูมิรูปภาพ (Picture Graph)	17
2.9	เว็บที่มีโครงสร้างแบบเรียงลำดับ	22
2.10	เว็บที่มีโครงสร้างแบบลำดับขั้น	23
2.11	เว็บที่มีโครงสร้างแบบตาราง	24
2.12	เว็บที่มีโครงสร้างแบบใยแมงมุม	24
2.13	การนำเสนอแบบทิศทางหรือแนวโน้ม (Trending)	28
2.14	Confusion Matrix	31
2.15	การแสดงแดชบอร์ดของ Google Chart	33
2.16	สัญลักษณ์ที่ใช้ในแผนภาพกระแสข้อมูล	34
3.1	กระบวนการวิเคราะห์ข้อมูล ด้วย CRISP-DM	38
3.2	เว็บไซต์ data.go.th	40
3.3	ข้อมูลผู้เสียชีวิตจากอุบัติเหตุทางถนน ปี 2554 – 2563	40
3.4	ข้อมูลผู้เสียชีวิตจากอุบัติเหตุทางถนน ปี 2554 – 2563	41
3.5	การแก้ไขชื่อจังหวัดให้เป็นชื่อเต็ม	42
3.6	การแก้ไขชื่ออำเภอให้ถูกต้อง	42
3.7	ชุดข้อมูลที่คัดเลือกมาทำการวิเคราะห์ข้อมูล	44
3.8	แสดงการเปิดโปรแกรม	55
3.9	แสดงการเข้าหน้าจอโปรแกรม Weka 3.8.4	56
3.10	แสดงการนำข้อมูลเข้าในโปรแกรม Weka 3.8.4	56

สารบัญภาพ (ต่อ)

ภาพที่		หน้า
3.11	หน้าจอแสดงข้อมูลที่นำเข้าโปรแกรม Weka 3.8.4	57
3.12	การเลือกเทคนิคที่ใช้ในการจำแนกข้อมูลแบบ Decision Tree: J48	57
3.13	หน้าจอผลลัพธ์ของโมเดลการจำแนกข้อมูลแบบ Decision Tree: J48	58
3.14	ผลลัพธ์จากการจำแนกกลุ่มแบบ Decision Tree: J48 ในโปรแกรม Weka	58
3.15	รูปแบบแผนภาพโมเดล Graph Decision Tree: J48 ในโปรแกรม Weka 3.8.4	59
3.16	ผลลัพธ์จากการจำแนกกลุ่มแบบ Decision Tree ของโปรแกรม RapidMiner	61
3.17	รูปแบบแผนภาพโมเดล Graph Decision Tree ในโปรแกรม RapidMiner	61
3.18	ผลลัพธ์จากการจำแนกกลุ่มแบบ Dacision Tree : J48 ในโปรแกรม Weka	62
3.19	สัญลักษณ์ที่ใช้ในแผนภาพกระแสข้อมูล	63
3.20	แผนภาพบริบท (Context Diagram)	64
3.21	แผนภาพกระแสข้อมูลระดับที่ 0 (Data Flow Diagram Level 0)	65
3.22	Wireframe แสดงหน้าเว็บไซต์แสดงเมนูต่าง ๆ บนหน้าเว็บไซต์	67
3.23	Wireframe ข้อมูลทั่วไปเกี่ยวกับการเกิดอุบัติเหตุบทท้องถนน	67
3.24	แสดงการวิเคราะห์ข้อมูลโดยใช้ Decision Tree	68
3.25	แสดงแดชบอร์ดผู้ประสบอุบัติเหตุบนท้องถนน	68
3.26	แสดง File ข้อมูลต่าง ๆ ที่สามารถดาวน์โหลดได้	69
3.27	แสดงข้อมูลส่วนตัวผู้จัดทำ	69
3.28	แสดงหน้าแดชบอร์ดที่สารถมารถแก้ไขได้	70
3.29	แสดง File ข้อมูลต่างที่สามารถแก้ไขข้อมูลได้	70
3.30	แสดงการเพิ่มข้อมูล	71
4.1	หน้าแรกของเว็บไซต์ เมื่อเข้าสู่เว็บไซต์	73
4.2	หน้าแสดงแบบสอบถามความพึงพอใจ	73
4.3	แสดงหน้าข้อมูลการเกิดอุบัติเหตุและการป้องกันการเกิดอุบัติเหตุ	74
4.4	แสดงข้อมูลสารสนเทศในรูปแบบตาราง	74

สารบัญภาพ (ต่อ)

ภาพที่		หน้า
4.5	แสดงหน้าการทำ Decision Tree	75
4.6	แสดงขั้นตอนการทดสอบข้อมูลกับโปรแกรม WEKA	75
4.7	แสดงข้อมูลในรูปแบบ แผนภาพ Dashboard	76
4.8	แสดงเอกสารที่นำมาใช้วิเคราะห์	76
4.9	แสดงข้อมูลผู้จัดทำ	77
4.10	แสดงหน้าเข้าสู่ระบบสำหรับแอดมิน	77
4.11	แสดงหน้าแบบฟอร์มเพิ่มข้อมูล	78
4.12	แสดงหน้าการเลือกแก้ไขข้อมูล	78
4.13	แสดงหน้าออกจากระบบสำหรับแอดมิน	79
ก.1	แสดงหน้าแรกของเว็บไซต์	95
ก.2	แสดงหน้าแบบสอบถามความพึงพอใจ	96
ก.3	แสดงเข้าสู่ระบบสำหรับผู้ดูแลระบบ	96
ก.4	แสดงข้อมูลเกี่ยวกับการเกิดอุบัติบนท้องถนนและการป้องกัน	97
ก.5	แสดงข้อมูลสารสนเทศในรูปแบบตาราง	97
ก.6	แสดงการวิเคราะห์ข้อมูลด้วยการใช้ MODEL โดย Decision Tree	98
ก.7	แสดงขั้นตอนการทดสอบข้อมูลกับโปรแกรม WEKA เวอร์ชั่น 3.8.4	99
ก.8	แสดงการนำเสนอข้อมูลในรูปแบบ แผนภาพ Dashboard	99
ก.9	แสดงหน้าเอกสารที่นำมาวิเคราะห์	100
ก.10	แสดงผู้จัดทำเว็บไซต์	100
ก.11	แสดงหน้าที่สามารถแก้ไขข้อมูลได้	101
ก.12	แสดงหน้าที่สามารถแก้ไขข้อมูลได้	102
ก.13	แสดงออกจากระบบสำหรับผู้ดูแลระบบ	102

บทที่ 1
บทนำ

1.1 ความสำคัญและที่มาของปัญหา

ในยุคปัจจุบันเทคโนโลยีดิจิตอลเข้ามาเป็นส่วนหนึ่งในการดำรงชีวิตประจำวันของคนเราเป็นอย่างมากทั้งหน่วยงานต่าง ๆไม่ว่าจะเป็นภาคธุรกิจ องค์กรต่าง ๆ รวมถึงหน่วยงานของภาครัฐเองต่างก็มีการนำเทคโนโลยีต่าง ๆเข้ามาช่วยเพื่อปรับปรุงประสิทธิภาพของการดำเนินงานและยกระดับการให้บริการไปสู่ประชาชนให้สามารถเข้าถึงบริการต่าง ๆได้สะดวกและรวดเร็วเป็นอย่างยิ่ง จึงทำให้เกิดข้อมูลสารสนเทศขึ้นมาอย่างมากมาย โดยเฉพาะอย่างยิ่งข้อมูลของภาครัฐที่จัดเก็บในรูปแบบของฐานข้อมูลนับเป็นเวลาหลายปี และกำลังเพิ่มขึ้นอีกในปัจจุบันและในอนาคต ข้อมูลเหล่านี้สามารถนำมาใช้ประโยชน์ได้ทั้งสำหรับภาครัฐ ภาคเอกชน และในส่วนของประชาชนซึ่งทำให้เกิดเป็นแนวคิด วิธี รูปแบบทางธุรกิจ รวมไปจนถึงวิถีชีวิตใหม่ ๆ การแบ่งปันข้อมูลเพื่อนำมาใช้ประโยชน์ และสร้างนวัตกรรมนั้น ซึ่งถือว่าเป็นสัญญาณเริ่มต้นของการขับเคลื่อนเศรษฐกิจด้วยข้อมูล อันจะช่วยให้เศรษฐกิจของประเทศเติบโตขึ้นและเพิ่มขีดความสามารถในการแข่งขันกันในเวทีระดับโลกได้ ที่ผ่านมาหลายๆ ประเทศมีการผลัดดัน และดำเนินการเกี่ยวกับการเปิดเผยข้อมูลสู่สาธารณะกันอย่างจริงจัง เช่นเดียวกันกับในประเทศไทยหลายๆหน่วยงานก็เริ่มเห็นถึงความสำคัญ และประโยชน์ของการเปิดเผยข้อมูลกันมากขึ้น ดังนั้นสำนักงานรัฐบาลอิเล็กทรอนิกส์ (องค์การมหาชน) (สรอ.) ซึ่งมีภารกิจหลักในการพัฒนารัฐบาอิเล็กทรอนิกส์ของประเทศให้มีคุณภาพ จึงได้พัฒนาระบบศูนย์กลางข้อมูลภาครัฐหรือ Data.go.th เพื่อเป็นการส่งเสริมและผลักดันให้เกิดการเปิดเผยข้อมูลของภาครัฐสู่สาธารณะอย่างเป็นรูปธรรม เป็นช่องทางให้ผู้ใช้บริการทั้งภาครัฐ ภาคเอกชน และในส่วนของประชาชน สามารถค้นหาและเข้าถึงข้อมูลที่มีคุณภาพของภาครัฐได้ง่าย โดยมีเป้าหมายสูงสุดเพื่อที่จะส่งเสริมให้เกิดความร่วมมือของภาครัฐภาคเอกชนและประชาชน รวมถึงการแก้ไขปัญหาในส่วนต่าง ๆ การวิเคราะห์ข้อมูลของผู้เสียชีวิต

จากอุบัติเหตุทางถนนในปี 2554 – 2563 คือการนำเอาข้อมูลที่ได้ถูกรวบรวมไว้และถูกเผยแพร่ผ่านทางเว็บไซต์ data.go.th ซึ่งเป็นเว็บไซต์ที่เก็บรวบรวมชุดข้อมูลต่าง ๆของประเทศไทย ซึ่งเป็นข้อมูลที่เปิดเผยได้

เพื่อให้ผู้ใช้บริการสามารถนำชุดข้อมูลไปศึกษาหรือวิเคราะห์ให้เกิดประโยชน์ต่อไปได้ การวิเคราะห์ข้อมูลที่มีปริมาณมหาศาลแสดงให้เห็นประโยชน์ต่าง ๆก็คือ ความรวดเร็วและมีประสิทธิภาพ ขณะที่หลายปีก่อนหน้านั้น ธุรกิจต้องรวบรวมข้อมูล การดำเนินการวิเคราะห์ และค้นหาข้อมูลที่สามารถนำมาใช้ประกอบการตัดสินใจในอนาคตได้ ส่วนในทุกวันนี้ธุรกิจสามารถระบุข้อมูลเชิงลึกสำหรับการกระทำที่ต้องการการการตัดสินใจอย่างทันท่วงที ความสามารถในการทำงานได้เร็วยิ่งขึ้นและยังคงไว้ซึ่งความคล่องตัว ช่วยให้หลายองค์กรสามารถสร้างความได้เปรียบด้านการแข่งขันอย่างที่ไม่เคยมีมาก่อน ซึ่งการนำข้อมูลผู้เสียชีวิตจากอุบัติเหตุทางถนนในปี 2554 – 2563 มาทำการวิเคราะห์นั้น เนื่องจากต้องการทราบถึงข้อมูลในเชิงลึกของสาเหตุการเสียชีวิต เช่น ปัจจัยที่นำมาซึ่งเกิดอุบัติเหตุ อาทิเช่น การประพฤติตัวของผู้ที่ประสบอุบัติเหตุหรือปัจจัยอื่นๆ ที่นำไปสู่การเสียชีวิตและเกิดการบาดเจ็บ เพื่อที่จะหาทางป้องกันและลดอัตราการบาดเจ็บหรือเสียชีวิต ซึ่งการวิเคราะห์ปัจจัยต่าง ๆ จากข้อมูลที่ได้รับมานั้นถือว่าเป็นข้อมูลที่มีความสำคัญต่อการป้องกันจากอุบัติเหตุต่าง ๆทางท้องถนน แต่ด้วยจำนวนของข้อมูลที่มีปริมาณที่มากจึงทำให้ยากต่อการนำเสนอให้เข้าใจได้ง่าย ซึ่งถือเป็นข้อมูลที่มีความจำกัดหากต้องนำเสนอข้อมูลชุดนี้ไปใช้เพื่อเผยแพร่ให้แก่กลุ่มผู้ใช้ข้อมูลที่ต้องการศึกษาหรือบุคคลภายนอก

จากปัญหาข้างต้นคณะผู้จัดทำจึงได้นำชุดข้อมูลที่ได้รับมานั้นเข้าสู่กระบวนการจำแนกตามรายการชุดของข้อมูล โดยการใช้ขีดความสามารถในการประมวลผลของโปรแกรมคอมพิวเตอร์เพื่อเพิ่มความน่าเชื่อถือในการวิเคราะห์ข้อมูลและประมวลผลข้อมูลให้เหมาะสมตรงกับความต้องการ เป็นการแยกแยะสิ่งที่จะพิจารณาออกเป็นส่วนย่อยที่มีความสัมพันธ์กัน เพื่อทำความเข้าใจกับข้อมูลที่ได้รับและข้อมูลในแต่ละส่วนให้ชัดเจน รวมทั้งการสืบค้นความสัมพันธ์ของส่วนต่าง ๆ เพื่อดูว่าส่วนประกอบปลีกย่อยนั้นสามารถเข้ากันได้หรือไม่ มีความสัมพันธ์เกี่ยวเนื่องกันอย่างไร โดยคณะผู้จัดทำได้ทำการจำแนกข้อมูลออกเป็นประเภทต่าง ๆ ด้วยเทคนิคการทำเหมืองข้อมูลแบบ Classification เป็นกระบวนการสร้างโมเดลจัดการข้อมูลให้อยู่ในกลุ่มที่กำหนดให้โดยอาศัยการเรียนรู้ข้อมูลเก่าในช่วงเวลาที่ผ่านมาในรูปแบบ Descriptive Analytics คือข้อมูล

เหตุการณ์ที่เกิดขึ้นแล้วมาทำการวิเคราะห์ข้อมูลแบบพื้นฐาน โดยจะเน้นไปที่ความสามารถในการอธิบายว่าเกิดอะไรขึ้นกับเหตุการณ์และสามารถอธิบายได้ว่าทำไมถึงเกิดขึ้นโดยใช้ Classification สร้างกฎเพื่อช่วยในการตัดสินใจจากข้อมูลที่มีอยู่ เพื่อใช้ในการทำนายแนวโน้มการเกิดขึ้นของปัจจัยที่ก่อให้เกิดผู้บาดเจ็บและผู้เสียชีวิตจากเทคนิคการจำแนกประเภทของข้อมูล ด้วยการวิเคราะห์ในรูปแบบของแผนภูมิต้นไม้หรือเรียกว่า Decision tree เป็นโครงสร้างที่ใช้แสดงกฎที่ได้จากเทคนิคการจำแนกประเภทของข้อมูลโดย Decision tree แต่ละโหนดแสดงถึงคุณลักษณะ (attribute) ส่วนโหนดลูกจะแสดงเงื่อนไขในการทดสอบและโหนดปลาย (leaf node) แสดงกลุ่มที่กำหนดไว้ ซึ่งการจำแนกข้อมูลออกเป็นประเภทต่าง ๆ นี้จะช่วยให้เกิดความเข้าใจต่อข้อมูลและสิ่งที่ต้องการทราบอย่างแท้จริงและทางคณะผู้จัดทำยังได้ทำการนำเสนอข้อมูลแบบ visualization ด้วยการแสดงผลข้อมูลในรูปแบบของภาพด้วย Google Data Studio และเผยแพร่ข้อมูลสารสนเทศนี้บน Web browser ที่นิยมในการใช้เพื่อเผยแพร่ทางสื่อออนไลน์ โดยใช้ภาษา HTML ในการพัฒนาหน้าเว็บไซต์สำหรับการเปิดเผยข้อมูลภาคในภาครัฐที่ได้รับข้อมูลมาจาก data.go.th เพื่อให้ผู้ใช้งานสามารถเข้าถึงข้อมูลได้อย่างเหมาะสม ซึ่งประโยชน์ที่ผู้ใช้งานทั่วไปจะได้รับเกี่ยวกับเรื่องของความเข้าใจ การรับรู้และการะตระหนักถึงความปลอดภัยในเหตุการณ์ต่าง ๆ สำหรับผู้ที่ทำการค้นหาและศึกษาข้อมูลก็สามารถดำเนินการผ่าน Web browser ที่ได้มีการเผยแพร่ข้อมูลสารสนเทศของผู้บาดเจ็บและเสียชีวิตบนท้องถนนในปี 2554 – 2563 ที่ผ่านการสรุปจำแนกข้อมูลตามประเภทต่าง ๆ ได้สะดวกรวดเร็วยิ่งขึ้นรวมไปจนถึงมีความถูกต้อง

1.2 วัตถุประสงค์

เพื่อพัฒนาเว็บไซต์สำหรับวิเคราะห์ข้อมูลผู้เสียชีวิตจากอุบัติเหตุทางถนนในปี 2554 – 2563

1.3 ประโยชน์ที่จะได้รับ

ได้เว็บไซต์สำหรับวิเคราะห์ข้อมูลผู้เสียชีวิตจากอุบัติเหตุทางถนนในปี 2554 – 2563

1.4 ขอบเขต

การพัฒนาเว็บไซต์สำหรับวิเคราะห์ข้อมูลผู้เสียชีวิตจากอุบัติเหตุทางถนน โดยมีขอบเขตในการทำงานอยู่ 3 ส่วน ซึ่งจะมีส่วนแรกเป็น ขอบเขตผู้วิเคราะห์ข้อมูล ส่วนต่อมาเป็นส่วนของขอบเขตของผู้ใช้ ส่วนสุดท้ายเป็นส่วนสำหรับผู้เข้าชมเว็บไซต์

1. ขอบเขตผู้วิเคราะห์ข้อมูล

กระบวนการวิเคราะห์ข้อมูลด้วย CRISP–DM

1.1 เข้าใจ (Business Understanding) ปัญหาในรูปแบบของการวิเคราะห์ข้อมูลทาง Data Mining โดยการวิเคราะห์ข้อมูลในประเด็น ข้อมูลผู้เสียชีวิตจากอุบัติเหตุทางถนน

1.2 เข้าใจข้อมูล (Data Understanding) และทำการรวบรวมข้อมูลเพื่อตรวจสอบรายละเอียดปริมาณและความน่าเชื่อถือของข้อมูลผู้ที่ได้รับบาดเจ็บและผู้เสียชีวิตจากอุบัติเหตุทางถนน ปี 2554 – 2563 จากเว็บไซต์ data.go.th

1.3 เตรียมข้อมูล (Data Preparation) และทำการคัดเลือกข้อมูลและทำการ Data Cleaning ข้อมูลเพื่อให้ได้ข้อมูลที่สมบูรณ์ ด้วยการตัดข้อมูลในส่วนที่ไม่จำเป็นออกไป เพื่อเหลือข้อมูลที่จำเป็นในการนำไปวิเคราะห์ข้อมูล

1.4 สร้างแบบจำลอง (Modeling) ผู้จัดทำการวิเคราะห์ข้อมูลด้วยการทำเหมืองข้อมูลแบบ Data Classification เพื่อใช้ในการทำนายแนวโน้มการเกิดขึ้นของปัจจัยที่ก่อให้เกิดผู้เสียชีวิตและผู้ที่ได้รับบาดเจ็บจากเทคนิคการจำแนกประเภทข้อมูล ด้วยการสร้างโมเดล Decision Tree เพื่อจัดกลุ่มข้อมูลที่มีความสัมพันธ์กัน โดยใช้โปรแกรมที่ใช้ทำเหมืองข้อมูล

1.5 การประเมินผล (Evaluation) ผู้จัดทำได้ข้อมูลที่ผ่านการวิเคราะห์ข้อมูลด้วยเทคนิค Data Mining จากการสร้างโมเดล Decision Tree และการสรุปผลข้อมูลเพื่อให้ตรงกับวัตถุประสงค์ที่ตั้งไว้

1.6 การนำไปใช้ (Deployment) ผู้จัดทำนำผลของข้อมูลที่ได้รับจากการวิเคราะห์แล้วมาแสดงผลข้อมูลผ่านทาง Web Browser โดยใช้ชุดคำสั่ง HTML CSS3 และ JavaScript ร่วมกับการนำเสนอข้อมูลแบบ Visualization

2. ขอบเขตของผู้ใช้

2.1 สำหรับผู้ดูแลระบบ

2.1.1 สามารถเพิ่ม ลบ และแก้ไขข้อมูลสารสนเทศผู้บาดเจ็บและเสียชีวิตจากอุบัติเหตุบนเว็บไซต์ได้

2.1.2 สามารถจัดการข้อมูลสารสนเทศผู้บาดเจ็บและเสียชีวิตจากอุบัติเหตุบนเว็บไซต์ได้

2.1.3 สามารถดูข้อมูลสารสนเทศของผู้บาดเจ็บและเสียชีวิตจากอุบัติเหตุทางถนน

2.1.4 สามารถดูข้อมูลสารสนเทศในลักษณะของตารางสรุปข้อมูลสถิติผู้บาดเจ็บและผู้เสียชีวิตทางถนน

2.1.5 สามารถดูข้อมูลสารสนเทศในลักษณะของรูปแบบแผนภูมิชนิดต่าง ๆ ได้

2.2 สำหรับผู้เข้าชมเว็บไซต์

2.2.1 สามารถดูข้อมูลสารสนเทศของผู้บาดเจ็บและเสียชีวิตจากอุบัติเหตุทางถนน ปี 2554 – 2563

2.2.2 สามารถดูข้อมูลสารสนเทศในลักษณะของตารางสรุปข้อมูลสถิติผู้บาดเจ็บและผู้เสียชีวิตทางถนน

2.2.3 สามารถดูข้อมูลสารสนเทศในลักษณะของรูปแบบแผนภูมิชนิดต่าง ๆได้

2.2.4 สามารถดาวน์โหลดไฟล์ข้อมูลจากทางเว็บไซต์ได้

1.5 เครื่องมือที่ใช้ในการพัฒนาโปรแกรม

1. Hardware

1.1 เครื่องคอมพิวเตอร์โน้ตบุ๊ค ASUS K46C Series

- CPU : Intel Core i5–3337U (1.80 GHz, 3 MB L3 Cache, up to 2.70 GHz)
- GPU : NVIDIA GEFORCE GT 740M (2 GB GDDR3)
- RAM : 4 GB DDR3
- Storage : HDD 750 GB 5400 RPM
  SSD 240 GB
- Display : 14 inch WXGA (1366x768) LED
- OS : Windows 10 Pro (64 Bit)

1.2 เครื่องคอมพิวเตอร์โน้ตบุ๊ค ACER Aspire F15

- CPU : Intel Core i5–7200U (2.50 GHz, 2.70 GHz)
- GPU : NVIDIA GEFORCE GTX 950M (4 GB GDDR4)
- RAM : 4 GB DDR4 Memory
- Storage : HDD 1000 GB

– Display : 14 inch WXGA (1366x768) LED

– OS : Windows 10 Education

2. Software

2.1 โปรแกรม RapidMiner Studio ใช้ในการทำเทคนิคเหมืองข้อมูล

2.2 โปรแกรม Weka ใช้ในการวิเคราะห์ข้อมูลด้วยเทคนิคเหมืองข้อมูล

2.3 โปรแกรม Microsoft Excel ใช้จัดรูปแบบของข้อมูล

2.4 โปรแกรม Microsoft Word ใช้ในการทำเอกสารต่าง ๆ

2.5 โปรแกรม Xampp เพื่อใช้เป็นเซิร์ฟเวอร์ในการจำลองสำหรับทดสอบและทดลองเว็บไซต์

2.6 โปรแกรม FileZila โปรแกรมที่ใช้ติดต่อกับ FTP Server เพื่ออับโหลดไฟล์

2.7 โปรแกรม Adobe XD ใช้ในการออกแบบเพื่อจำลองรูปแบบของหน้าเว็บไซต์

2.8 โปรแกรม Dreamweaver CS6 ใช้สำหรับในการทำเว็บไซต์

2.9 เว็บไซต์ Google Data Studio ใช้ในการเขียนโปรแกรมเพื่อทำการเชื่อมต่อกับการแสดง Dashboard

2.10 ภาษา HTML ใช้ในการแสดงผลรูปแบบภาพนิทัศน์ของข้อมูล (Data Visualization)

2.11 ภาษา CSS ใช้ในการตกแต่งเว็บไซต์

2.12 ชุดคำสั่ง JavaScript ใช้ในการแสดงผลข้อมูลในรูปแบบของภาพ

1.6 สถานที่ใช้ในการดำเนินการศึกษาและรวบรวมข้อมูล

มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา ตั้งที่อยู่ที่ 128 ถนนห้วยแก้ว ตำบลช้างเผือก อำเภอเมืองเชียงใหม่ จังหวัดเชียงใหม่ 50300

1.7 ระยะเวลาในการดำเนินการ

ตารางที่ 1.1 ระยะเวลาในการดำเนินการ

แผนการดำเนินการ	2564				
	มิ.ย.	ก.ค.	ส.ค.	ก.ย.	ต.ค.
1. ศึกษาและกำหนดความต้องการ	←	—			
2. วิเคราะห์ออกแบบระบบและสร้างฐานข้อมูล		←	—		
3. เขียนและทดสอบโปรแกรม			←	—	
4. ติดตั้ง ทดสอบ และปรับปรุงระบบ			←	—	
5. ตรวจสอบระบบโดยรวม			←	—	
6. ประเมินการใช้งานระบบ				←	—
7. จัดทำคู่มือการใช้งาน				←	—
8. จัดทำเอกสารประกอบโครงงาน	←	—	—	—	—

บทที่ 2
แนวคิด ทฤษฎี เครื่องมือและวรรณกรรมที่เกี่ยวข้อง

โครงงานเรื่อง การพัฒนาเว็บไซต์สำหรับวิเคราะห์ข้อมูลผู้เสียชีวิตจากอุบัติเหตุทางถนน ปี 2554 – 2563 ในบทนี้จะเป็นการนำเสนอเกี่ยวกับ แนวคิด ทฤษฎี เครื่องมือและวรรณกรรมที่เกี่ยวข้องของการวิเคราะห์ข้อมูลของผู้เสียชีวิตจากอุบัติเหตุทางถนน ปี 2554 – 2563 และการแสดงผลข้อมูลบนเว็บไซต์ ซึ่งได้รวบรวมการศึกษาเอกสารงานวิจัยที่เกี่ยวข้องกับการวิเคราะห์ข้อมูล เพื่อใช้เป็นแนวทางการศึกษาประกอบด้วยรายละเอียดตามลำดับ

2.1 แนวคิด

- 2.1.1 แนวคิดเกี่ยวกับการวิเคราะห์ข้อมูล (Data Analytic)
- 2.1.2 แนวคิดเกี่ยวกับการประยุกต์ใช้ข้อมูลทางสถิติ
- 2.1.3 แนวคิดเกี่ยวกับการแสดงผลข้อมูล (Data Visualization)

2.2 ทฤษฎี

- 2.2.1 ทฤษฎีเกี่ยวกับการจัดการข้อมูลขนาดใหญ่
- 2.2.2 ทฤษฎีเกี่ยวกับการทำเหมืองข้อมูล
- 2.2.3 ทฤษฎีเกี่ยวกับการออกแบบเว็บไซต์
- 2.2.4 ทฤษฎีเกี่ยวกับชุดคำสั่ง CSS และ Bootstrap
- 2.2.5 ทฤษฎีเกี่ยวกับการ Visualization
- 2.2.6 ทฤษฎีเกี่ยวกับต้นไม้ตัดสินใจ (Decision Tree)

2.3 เครื่องมือในการออกแบบและวิเคราะห์ข้อมูล

- 2.3.1 แบบจำลองต้นไม้ตัดสินใจ (Decision Tree)
- 2.3.2 การประเมินผลของโมเดล (Decision Tree)
- 2.3.3 การแสดงผลแบบ Visualization
- 2.3.4 แผนภาพกระแสข้อมูล (Data Flow Diagram : DFD)

2.4 วรรณกรรมที่เกี่ยวข้อง

2.5 บทสรุป

2.1 แนวคิด

2.2.1 แนวคิดเกี่ยวกับการวิเคราะห์ข้อมูล (Data Analytic)

ในการดำเนินงานเรื่องการวิเคราะห์ข้อมูลผู้เสียชีวิตจากอุบัติเหตุทางถนน ปี 2554 – 2563 และการแสดงผลข้อมูลเว็บไซต์ คณะผู้จัดทำได้ศึกษาหลักการและทฤษฎีต่าง ๆ องค์ประกอบหนึ่งที่สำคัญคือการวิเคราะห์ข้อมูล ซึ่งมีรายละเอียด ดังนี้

ความหมายของการวิเคราะห์ข้อมูล

เอื้อมพร หลินเจริญ (2555) การวิเคราะห์ข้อมูลเชิงคุณภาพ เป็นวิธีการสร้างข้อสรุปจากข้อมูลจำนวนหนึ่งซึ่งมักไม่ใช้สถิติในการวิเคราะห์ ทั้งนี้การวิเคราะห์ข้อมูลเชิงคุณภาพ อาจใช้กับการวิจัยเชิงปริมาณที่ผู้วิจัยมีการเก็บรวบรวมข้อมูลเชิงคุณภาพ เช่น แบบสอบถามปลายเปิด การสัมภาษณ์ การสังเกต จดบันทึก สำหรับสาระในบทความนี้ ผู้เขียนมุ่งนำ เสนอสาระ เกี่ยวกับเทคนิควิเคราะห์ข้อมูลเชิงคุณภาพที่นักวิจัยนิยมใช้ ได้แก่ การจำแนกหรือ การจัดกลุ่มข้อมูล การเปรียบเทียบเหตุการณ์ การวิเคราะห์ส่วนประกอบ การวิเคราะห์แบบอุปนัย และการวิเคราะห์ข้อมูลที่เป็นเอกสาร

จุมพล หนิมพานิช และวรวลัญช์ โรจนพล (2561) การวิเคราะห์ข้อมูลการวิจัยเชิงคุณภาพในทางรัฐศาสตร์ในภาพรวมมีลักษณะเหมือนกับการวิเคราะห์ข้อมูลเชิงคุณภาพโดยทั่วไป คือ เป็นเรื่องของการนำข้อมูลที่ได้จากการค้นคว้าวิจัยมาจัดกระทำให้เป็นระบบและหาความหมาย แยกแยะองค์ประกอบ รวมทั้งเชื่อมโยงและหาความสัมพันธ์ของข้อมูล

หัสพร ทองแดง, พิศุทธิภา เมธีกุล, สมคิด ทุมวงศ์ และแพรว สมบัติใหม่(2559) ได้ให้ความหมายการวิเคราะห์ข้อมูลเชิงคุณภาพไว้ว่า การวิเคราะห์ข้อมูลเป็นการจำแนกข้อมูลให้เป็นระบบ เพื่อทำให้เกิดความเข้าใจในความหลากหลาย ความหมายและความสัมพันธ์ของข้อมูลในบริบทของสังคมและวัฒนธรรมหนึ่ง ๆ ในการวิเคราะห์ข้อมูลเชิงคุณภาพนี้หลังจากที่มีการแยกข้อมูลออกได้เป็นหมวดหมู่แล้ว นักวิจัยต้องใช้วิธีการนำเสนอโดยการพรรณนาให้เห็นสภาพเงื่อนไข กระบวนการ ขั้นตอนการสัมพันธ์ต่าง ๆ รวมทั้งพยายามหาความหมายทางวัฒนธรรมของปรากฏการณ์ในทัศนะของบุคคลหรือกลุ่มตัวอย่างซึ่งเป็นผู้ให้ข้อมูล และอาจมีความเข้าใจและตีความแตกต่างกันออกไปหรือแตกต่างไปจากของผู้วิจัยเอง การวิเคราะห์ข้อมูลสำหรับการวิจัยเชิงคุณภาพ จึงเป็นเสมือนการทำความเข้าใจในแบบแผนและความหมายของพฤติกรรมมากกว่าที่จะมุ่งหาระดับความมากน้อยหรือระดับความเข้มข้นของความสัมพันธ์

ของปรากฏการณ์จึงเป็นการวิเคราะห์ข้อมูลที่อาศัยเครื่องมือทางความคิด (Conceptual tools) มากกว่าที่จะใช้เครื่องมือทางสถิติ และที่สำคัญต้องอาศัยความสามารถของนักวิจัยที่จะอธิบายให้เห็นความสัมพันธ์และความหมายของปรากฏการณ์บนพื้นฐานความเข้าใจในวัฒนธรรมและสังคมที่เป็นบริบทของปรากฏการณ์ที่ศึกษา

2.1.2 แนวคิดเกี่ยวกับการประยุกต์ใช้ข้อมูลทางสถิติ

ความหมายของการประยุกต์ใช้ข้อมูลทางสถิติ จากเอกสารประกอบการสอนการวิจัยสำหรับครู ได้ให้ความหมายเกี่ยวกับสถิติไว้ว่า คำว่าสถิติ (Statistics) มาจากภาษาเยอรมันว่า Statistics มีรากศัพท์มาจาก Stat หมายถึง ข้อมูลหรือสารสนเทศ ซึ่งจะอำนวยประโยชน์ต่อการบริหารประเทศในด้านต่าง ๆ เช่น การทำ สำมะโนครัวเพื่อจะทราบจำนวนพลเมืองในประเทศทั้งหมดในสมัยต่อมา คำว่า สถิติ ได้หมายถึงตัวเลขหรือข้อมูลที่ได้จากการเก็บรวบรวม เช่น จำนวนผู้ประสบอุบัติเหตุบนท้องถนน อัตราการเกิดของเด็กทารก ปริมาณน้ำฝนในแต่ละปี เป็นต้น สถิติในความหมายที่กล่าวมานี้ เรียกอีกอย่างหนึ่งว่า ข้อมูลทางสถิติ (Statistical data)

1) ค่ากึ่งกลาง (Median) เป็นการวัดแนวโน้มเข้าสู่ส่วนกลางวิธีหนึ่งที่ใช้การเรียงค่าการสะท้อนของจุดภาพจากค่าน้อยที่สุดไปหาค่ามากที่สุด โดยค่ากึ่งกลางเป็นค่าที่อยู่ในตำแหน่งกึ่งกลางของข้อมูลทั้งชุด ค่ากึ่งกลางจึงเป็นตัวแทนค่าการสะท้อนของจำนวนจุดภาพทั้งหมดในช่วงคลื่นหนึ่งๆ ที่แสดงให้ทราบว่ามีจำนวนจุดภาพที่มีค่าการสะท้อนมากกว่าและน้อยกว่าค่ากึ่งกลางอยู่ประมาณร้อยละ 50

2) ค่าฐานนิยม (Mode) เป็นการวัดแนวโน้มเข้าสู่ส่วนกลางอีกวิธีหนึ่ง โดยดูจากจำนวนความถี่ของค่าการสะท้อนซึ่งมีความถี่สูงที่สุด นิยมนำมาใช้กับข้อมูลที่เป็นนามบัญญัติ เช่น ค่าของประเภทข้อมูลหลังจากการจำแนกประเภทแล้ว ถือเป็นค่าการสะท้อนที่แสดงการใช้ที่ดินประเภทต่าง ๆ ไม่ใช่ค่าการสะท้อนของวัตถุอีกต่อไป

3) ความเบี่ยงเบนมาตรฐาน (Standard Deviation : S.D.) เป็นการวัดการกระจายที่นิยมใช้มากที่สุดการคำนวณใช้วิธียกกำลังสองของผลต่างระหว่างค่าการสะท้อนของทุกจุดภาพในแต่ละช่วงคลื่นกับค่าเฉลี่ยเลขคณิตของช่วงคลื่นนั้น

ภัทรธิดา ผลงาม (2558) ได้ให้ความหมายเกี่ยวกับสถิติไว้ว่า สถิติ คือ ตัวเลขต่าง ๆ ที่ได้มีการรวบรวมขึ้นเพื่อบอกข้อเท็จจริงเกี่ยวกับคุณสมบัติหรือลักษณะบางสิ่งบางอย่างที่สามารถแสดงออกเป็นตัวเลขได้สถิติแบ่งเป็นประเภทใหญ่ๆ ได้ 2 ประเภท ดังนี้ คือ

1) สถิติพรรณนา (Descriptive statistics) คือสถิติที่บรรยายถึงลักษณะของข้อมูลเฉพาะกลุ่มนั้นๆ โดยไม่สรุปอ้างอิงไปยังประชากรกลุ่มอื่นๆ สถิติประเภทนี้นิยมศึกษาในกลุ่มเล็กหรือกลุ่มใหญ่ก็ได้สถิติประเภทนี้เป็นสถิติที่บรรยายลักษณะของข้อมูล เช่น ค่าเฉลี่ย ร้อยละ มัธยฐาน พิสัย ส่วนเบี่ยงเบนมาตรฐาน และการวัดความสัมพันธ์ต่าง ๆ เช่น สหสัมพันธ์

2) สถิติอ้างอิงหรือสถิติอนุมาน (Inferential or inductive statistics) คือ สถิติที่นำค่าสถิติพรรณนามาสรุปอ้างอิงไปยังประชากรหรือเป็นสถิติที่ศึกษาจากกลุ่มตัวอย่างแล้วนำผลอ้างอิงไปยังกลุ่มประชากร เช่น การศึกษาความคิดเห็นต่อการเลือกตั้งของประชาชนไม่จำเป็นต้องศึกษาจากประชาชนทุกคน แต่สามารถเลือกศึกษาจากประชาชนบางกลุ่มซึ่งจะเป็นตัวแทนของประชาชนทั้งหมด แล้วจึงสรุปว่าประชาชนมีความคิดเห็นอย่างไรต่อการเลือกตั้งได้ ดังนั้นในการใช้สถิติอ้างอิงนี้จึงจำเป็นต้องเลือกลุ่มตัวอย่างที่เหมาะสม สถิติอ้างอิงนี้ก่อนนำไปอ้างอิงกลุ่มประชากรต้องมีการทดสอบทางสถิติก่อนทุกครั้งจึงสามารถอ้างอิงประชากรได้อย่างมีประสิทธิภาพ

2.1.3 แนวคิดเกี่ยวกับการแสดงผลข้อมูล (Data visualization)

ความหมายการแสดงผลข้อมูล

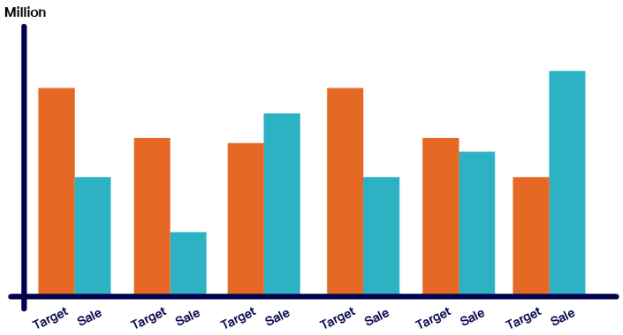
กานต์ ยงศิริวิทย์ และภาคภูมิ ชัยศิริประเสริฐ (2560) ได้ให้ความหมายของการแสดงผลข้อมูลไว้ว่า การแสดงผลข้อมูล (Data Visualization) ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลควรนำเสนอในรูปแบบที่น่าสนใจ และเข้าใจง่าย ดังนั้นเครื่องมือควรที่จะสามารถนำเสนอข้อมูลในรูปแบบของกราฟและแสดงผลลัพธ์บนแผนที่ในขั้นพื้นฐานได้ โดยไม่ควรที่จะต้องมีการติดตั้งซอฟต์แวร์หรือตัวช่วยเพิ่มเติม

ชนาธิป ชื่นมนัส (2552) ได้ให้ความหมายของการแสดงผลข้อมูลไว้ว่า การแสดงผลข้อมูล (Data Visualization) คือวิธีการที่ใช้ในการนำข้อมูลที่เป็นนามธรรมมาแสดงให้เป็นภาพเพื่อเพิ่มประสิทธิผลในการติดต่อสื่อสารหรือทำให้เข้าใจข้อมูลได้ง่ายขึ้น การใช้เทคนิค Visualization มีหลายวิธีซึ่งแต่ละวิธีเหมาะสมที่จะใช้ในการแสดงข้อมูลที่แตกต่างกัน การเลือกใช้เทคนิคใดนั้นขึ้นกับองค์ประกอบหลายอย่าง เช่น ชนิดของข้อมูลที่จะนำมาแสดง วิธี ที่ต้องการแสดงข้อมูล เป็นต้น ในปัจจุบันมีงานวิจัยที่นำเอาเทคนิค Visualization มาใช้โดยมีจุดประสงค์เพื่อเพิ่มประสิทธิผลในการแสดงข้อมูลการติดต่อสื่อสารภายในเครือข่ายคอมพิวเตอร์ และช่วยในการวิเคราะห์ความผิดปกติที่เกิดขึ้นในเครือข่าย

อาทิตย์ สิทธิบรรเจิด (2552) ได้ให้ความหมายของการแสดงผลข้อมูลไว้ว่า การแสดงผลข้อมูล (Data Visualization) เป็นส่วนประกอบสำคัญใน Cognitive System ซึ่งเป็นส่วนในการแสดงข้อมูลหรือผลลัพธ์ต่าง ๆ ในระหว่างคอมพิวเตอร์และผู้ใช้งานในรูปแบบของภาพ โดยผู้ใช้สามารถเรียนรู้และจดจำข้อมูลผ่านการมองได้มากกว่าการใช้ประสาทสัมผัสอื่นๆ หรือจะกล่าวว่า Visualization ก็คือ การสร้างมโนภาพของสิ่งต่าง ๆ ที่เราสนใจขึ้นมาในใจ ซึ่งต่อมาได้กลายเป็นการนำภาพมาใช้กับการนำเสนอหรือนามาเป็นกรอบความคิด ซึ่งได้นำไปใช้ในการสนับสนุนการตัดสินใจ

การเลือกรูปแบบ Visualization ให้เหมาะสมกับข้อมูลในปัจจุบันเป็นยุคเทคโนโลยีเข้าถึงทุกคนทำให้การรับรู้ข่าวสาร ข้อมูลต่าง ๆ เป็นไปได้ง่ายและรวดเร็วมากขึ้น คนที่นำเสนอข้อมูลจึงต้องนำเสนอข้อมูลให้น่าสนใจ เข้าใจง่าย และรวดเร็ว จึงเกิดการสร้าง Data Visualization ขึ้นมา Data Visualization เป็นการใช้ภาพเพื่อแสดงข้อมูลในเชิงปริมาณที่วัดได้ ซึ่งอาจนำเสนอออกมาในรูปแบบ แผนภูมิ กราฟ กราฟิก และอื่นๆ อีกมากมาย เพื่อให้เข้าใจได้

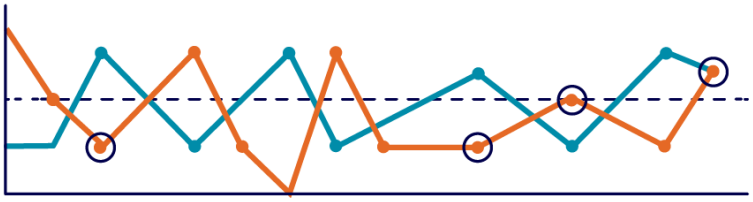
1) แผนภูมิแท่ง (Bar chart) คือกราฟที่ประกอบด้วยรูปสี่เหลี่ยมผืนผ้า (เรียกว่าแท่ง) จำนวนหนึ่ง โดยมีลักษณะเป็นแท่งสูงหรือยาวที่เปลี่ยนแปลงตามขนาด แต่มีความกว้างเท่ากันหมด เราอาจเรียงแท่งเหล่านี้ในทางตั้งหรือทางนอนก็ได้ โดยเว้นระยะช่องว่างตามสมควรและจะต้องเขียนโครงเรื่องจำแนกแต่ละแท่งให้ชัดเจนด้วย แผนภูมิแท่งอาจจะมีการระบายสีหรือแรเงาเพื่อให้ดูเด่น และในกรณีที่มีการเปรียบเทียบกันหลายแท่ง เช่น แผนภูมิแท่งซับซ้อนหรือเชิงประกอบ จำเป็นจะต้องระบายสีหรือแรเงา เพื่อจำแนกความแตกต่างของแผนภูมิแต่ละชุดที่นำมาเปรียบเทียบกันนั้น วัตถุประสงค์ของการนำเสนอข้อมูลในรูปแผนภูมิแท่งเพื่อแสดงความเปลี่ยนแปลงและการเปรียบเทียบข้อมูล



ภาพที่ 2.1 แผนภูมิแท่ง (Bar Chart)

ที่มา : 9experttraining.com

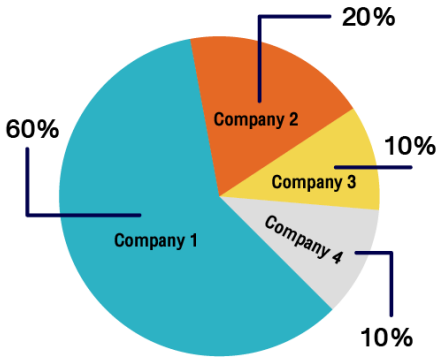
2) แผนภูมิเส้น (Line Charts) แผนภูมิเส้น มีลักษณะคล้ายแผนภูมิแท่ง ซึ่งประกอบด้วยแกนตั้งและนอน เพียงแต่เปลี่ยนจากแท่งข้อมูลเป็นจุดบนแผนภูมินั่นเองแผนภูมิประเภทนี้เหมาะกับการนำเสนอข้อมูลตัวเลขที่มีความสัมพันธ์กับข้อมูลที่มีลักษณะเป็นช่วงใช้แสดงให้เห็นการเปลี่ยนแปลงที่เกิดขึ้นตามเวลาเพื่อดูแนวโน้ม รวมถึงสามารถใช้พยากรณ์แนวโน้มในอนาคตได้ เช่น ข้อมูลของยอดขายในแต่ละปี หรือไตรมาส และนำมาวิเคราะห์เพื่อดูแนวโน้มเป็นต้น



ภาพที่ 2.2 แผนภูมิเส้น (Line Charts)

ที่มา : 9experttraining.com

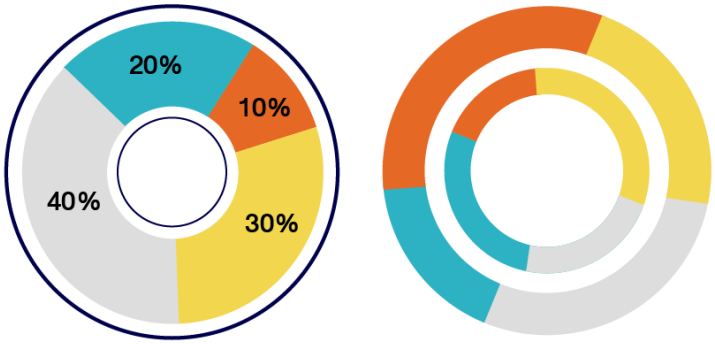
3) แผนภูมิวงกลม (Pie Charts) แผนภูมิวงกลมเหมาะกับการนำเสนอข้อมูลที่มีส่วนประกอบย่อยที่รวมกันเป็นส่วนใหญ่ มีการแบ่งส่วนให้ดูง่ายและสวยงามแต่ในทางกลับกันอาจจะดูยากในเรื่องของการประมาณขนาดของแต่ละชิ้นยิ่งถ้ามีจำนวนชิ้นมากจะยิ่งแยกยากเพราะต้องใช้หลายสีในการนำเสนอข้อมูล เช่น ส่วนแบ่งทางการตลาด (Market Share) ข้อมูลแสดงส่วนผสมต่าง ๆ เป็นต้น



ภาพที่ 2.3 แผนภูมิวงกลม (Pie Charts)

ที่มา : 9experttraining.com

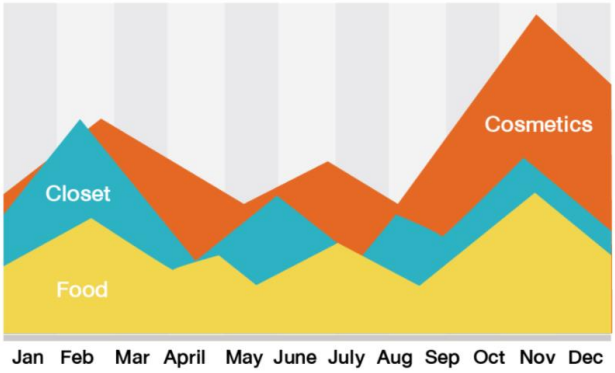
4) แผนภูมิโดนัท (Doughnut Charts) แผนภูมิโดนัทมีหลักการออกแบบเช่นเดียวกับแผนภูมิวงกลมแต่สามารถแสดงชุดข้อมูลได้มากกว่า 1 ชุด โดยนำเสนอข้อมูลเป็นวงกลมซ้อนกันหลายๆ ชั้น



ภาพที่ 2.4 แผนภูมิโดนัท (Doughnut Charts)

ที่มา : 9experttraining.com

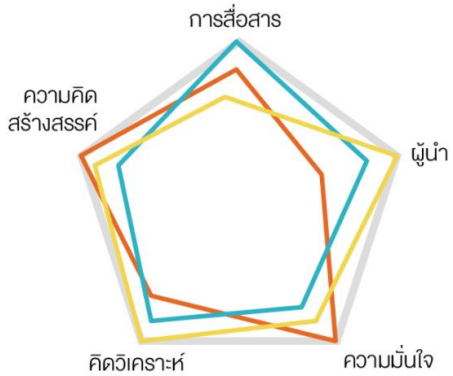
5) แผนภูมิพื้นที่ (Area Charts) มีหน้าตาคล้ายแผนภูมิเส้น แต่มีการแรเงาพื้นที่ใต้เส้นข้อมูล หรือระหว่าง 2 เส้นเพื่อแสดงให้เห็นปริมาณความแตกต่างระหว่างเส้นเหมาะสำหรับเน้นความสำคัญของการเปลี่ยนแปลงในช่วงเวลาแสดงให้เห็นผลรวมของความแตกต่างระหว่างข้อมูล เช่น ข้อมูลของการซื้อสินค้าในห้างสรรพสินค้าในแต่เดือนตามหมวดหมู่ต่าง ๆ ไล่ไป เสื้อผ้าแฟชั่น อาหาร ตามลำดับ



ภาพที่ 2.5 แผนภูมิพื้นที่ (Area Charts)

ที่มา : 9experttraining.com

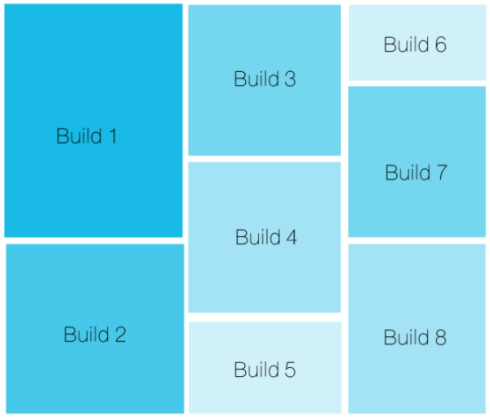
6) แผนภูมิเรดาร์ (Radar Charts) มีลักษณะคล้ายแผนภูมิเส้นที่มีการแสดงผลแบบวงกลมจำนวนเหลี่ยมของเรดาร์เท่ากับจำนวนหัวข้อของข้อมูล แผนภูมินี้ไม่ได้บอกถึงความต่อเนื่องของข้อมูลแต่เหมาะสำหรับการนำเสนอข้อมูลเป็นหัวข้อ แล้วนำมาวิเคราะห์หาจุดอ่อนจุดแข็งของข้อมูล เช่น นำเสนอข้อมูลเกี่ยวกับคุณสมบัติของการรับพนักงานใหม่ เพื่อดูจุดอ่อนจุดแข็งของแต่ละคน เป็นต้น



ภาพที่ 2.6 แผนภูมิเรดาร์ (Radar Charts)

ที่มา : 9experttraining.com

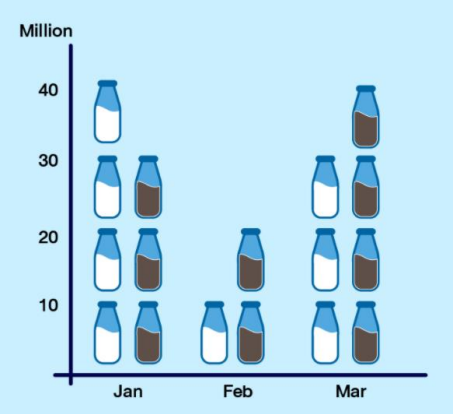
7) แผนภูมิต้นไม้ (Tree Maps) คือการนำเสนอข้อมูลแบบแสดงให้เห็นพื้นที่แสดงผลได้ในแบบลำดับชั้นเหมือนแบบโครงสร้างต้นไม้ อาจจะนำเสนอข้อมูลที่ต้องการให้เห็นถึงเขตพื้นที่ แสดงพื้นที่สีที่แตกต่างกันได้



ภาพที่ 2.7 แผนภูมิต้นไม้ (Tree Maps)

ที่มา : 9experttraining.com

8) แผนภูมิรูปภาพ (Picture Graph) เป็นแผนภูมิที่ประกอบไปด้วยแกนนอน และแกนตั้ง แต่เลือกใช้รูปภาพหรือไอคอนแทนจำนวนของสิ่งของนั้นๆ เช่น การแสดงผลจำนวนของนมที่ขายได้ในแต่ละเดือน โดยนำเสนอทั้งนมรสจืด รสช็อกโกแลต เปรียบเทียบในแต่ละเดือนซึ่งมีการนำเสนอลักษณะคล้ายกับกราฟแท่ง แต่เปลี่ยนจากแท่งเป็นรูปกราฟิกของนม 2 รสชาติแทน ก็ทำให้การนำเสนอข้อมูลน่าสนใจมากยิ่งขึ้นซึ่งแนวทางการนำเสนอข้อมูลลักษณะนี้ต้องอาศัยความคุ้นชินของคนดู เพื่อแทนสัญลักษณ์ภาพลงไป เช่น เมื่อพูดถึงจำนวนคนอาจจะแทนด้วยภาพไอคอนคนหรือเมื่อพูดถึงจำนวนเงิน ควรแทนภาพเป็นเหรียญเงิน หรือแบงค์ ก็จะทำให้คนดูเข้าใจง่ายจากสัญลักษณ์ภาพที่คุ้นเคย



ภาพที่ 2.8 แผนภูมิรูปภาพ (Picture Graph)

ที่มา : 9experttraining.com

2.2 ทฤษฎี

2.2.1 ทฤษฎีเกี่ยวกับการจัดการข้อมูลขนาดใหญ่

2.2.1.1 ข้อมูลขนาดใหญ่

ข้อมูลขนาดใหญ่ หมายถึงข้อมูลที่มีขนาดใหญ่ เร็ว หรือซับซ้อนจนยาก หรือเป็นไปไม่ได้ที่จะประมวลผลโดยใช้วิธีการแบบเดิม การเข้าถึงและจัดเก็บข้อมูลจำนวนมาก เพื่อทำการวิเคราะห์มีมานานแล้ว แต่แนวคิดเกี่ยวกับข้อมูลขนาดใหญ่เป็นที่แพร่หลายในช่วงต้นปีค.ศ. 2000 เมื่อดัก ลานีย์ นักวิเคราะห์อุตสาหกรรมได้ให้คำจำกัดความที่เป็นที่เข้าใจกันใน ขณะนี้ว่า ข้อมูลขนาดใหญ่ประกอบด้วย SVs ดังนั้น

Volume (ปริมาณ) : องค์กรต่าง ๆ รวบรวมข้อมูลจากหลากหลายแหล่ง ซึ่งรวมถึงธุรกรรมของธุรกิจ อุปกรณ์อัจฉริยะ (IoT) อุปกรณ์อุตสาหกรรม วิดีโอ โซเชียลมีเดีย และอื่นๆในอดีต การจัดเก็บข้อมูลมือเป็นปัญหาใหญ่ –แต่เมื่อค่าใช้จ่ายในการจัดเก็บบนแพลตฟอร์มต่าง ๆ เช่น พื้นที่จัดเก็บข้อมูลส่วนกลาง (Data Lake) และ Hadoop ลดลง ภาระนี้จึงบรรเทาลง

Velocity (ความเร็ว) : ด้วยการเติบโตของ Internet of Things ข้อมูลจะถูกส่งไปยังธุรกิจต่าง ๆ ด้วยความเร็วที่ไม่เคยมีมาก่อนและต้องได้รับการจัดการในเวลาที่เหมาะสม แท็ก RFID, เซ็นเซอร์ และสมาร์ทมิเตอร์ช่วยผลักดันความต้องการในการจัดการกับกระแสข้อมูลเหล่านี้ในแบบเรียลไทม์

Variety (ความหลากหลาย) : ข้อมูลมีในทุกรูปแบบ นับตั้งแต่ข้อมูลที่มีโครงสร้าง ตัวเลขในฐานข้อมูลแบบดั้งเดิม ไปจนถึงเอกสารข้อความ อีเมล วิดีโอ เสียง ข้อมูลหุ้น และธุรกรรมทางการเงิน

การวิเคราะห์ Big Data ช่วยให้องค์กรควบคุมข้อมูลของพวกเขาและใช้เพื่อระบุโอกาสใหม่ๆ ในทางกลับกันนำไปสู่การเคลื่อนไหวทางธุรกิจที่ชาญฉลาดเพื่อการดำเนินงานที่มีประสิทธิภาพมากขึ้น ผลกำไรที่สูงขึ้นและลูกค้าที่มีความสุขมากขึ้น ในรายงาน Big Data ในบริษัทขนาดใหญ่โดยผู้อำนวยการฝ่ายวิจัยของ Tom Davenport ให้ สัมภาษณ์ว่ามากกว่า 50 ธุรกิจใช้และทำความเข้าใจว่าพวกเขาใช้ Big Data อย่างไร และ พบว่าสามารถช่วยเหลือธุรกิจได้ดังต่อไปนี้

ลดต้นทุน : เทคโนโลยีข้อมูลขนาดใหญ่เช่น Hadoop และการวิเคราะห์บนคลาวด์นำมาซึ่งความได้เปรียบด้านต้นทุนอย่างมีนัยสำคัญเมื่อพูดถึงการจัดเก็บข้อมูลจำนวนมากรวมทั้งสามารถระบุวิธีการทำธุรกิจที่มีประสิทธิภาพมากขึ้น

เร็วกว่าและตัดสินใจดีกว่า : ด้วยความเร็วของ Hadoop และการวิเคราะห์ในหน่วยความจำรวมกับความสามารถในการวิเคราะห์แหล่งข้อมูลใหม่ๆ ของธุรกิจจะสามารถสร้างข้อมูลได้ทันทีและสามารถดำเนินการต่อได้ทันทีจากการวิเคราะห์นั้น

ผลิตภัณฑ์และบริการใหม่ : ด้วยความสามารถในการวัดความต้องการและความพึงพอใจของลูกค้าผ่านการวิเคราะห์นำมาซึ่งสิ่งที่ลูกค้าต้องการ ซึ่งยกตัวอย่างโดยดาเวนพอร์ท สามารถชี้ให้เห็นว่าด้วยการวิเคราะห์ข้อมูลขนาดใหญ่ของบริษัทต่าง ๆ จะสามารถสร้างผลิตภัณฑ์ใหม่ๆ เพื่อตอบสนองความต้องการของลูกค้าได้เป็นอย่างดี

2.2.1.2 การจัดการข้อมูลขนาดใหญ่

การรวบรวมข้อมูล

การรวบรวมข้อมูลของ Big Data เป็นการรวบรวมข้อมูลของจากหลากหลายทั้งที่มาและการใช้งานที่แตกต่างกันอย่างมากมาย ซึ่งกลไกและเทคโนโลยีแบบดั้งเดิม ETL (extract, transform, and load) ไม่สามารถทำได้ ซึ่ง Big Data หรือ ข้อมูลขนาดใหญ่ต้องการเทคนิค วิธีการ และเทคโนโลยีใหม่ในการรวบรวมข้อมูลขนาด เทราไบต์ และอาจจะเป็นระดับเพธาไบต์เลยก็มี ในการรวบรวมข้อมูลนั้นต้องมีการประมวลผล จัดรูปแบบ ให้เหมาะสำหรับการใช้ในการวิเคราะห์หรือใช้งานสำหรับธุรกิจหรือวัตถุประสงค์นั้นๆ

การจัดการข้อมูลข้อมูลขนาดใหญ่ หรือ Big Data นั้นมีความต้องการสถานที่จัดเก็บขนาดใหญ่ การจัดเก็บข้อมูลมูลขนาดใหญ่จะเป็นชนิดใดก็ได้ไม่ว่าจะเป็นแบบon premises หรือ แบบ cloud ขึ้นกับความต้องการหรือความสะดวกในการใช้ ซึ่งเราสามารถใช้และประเมินผลได้ เช่นเดียวกัน บางครั้งก็มีความจำเป็นที่ต้องจัดเก็บไว้ใกล้กับแหล่งข้อมูลหรือข้อมูลบางอัน ต้องการความยืดหยุ่นสูงและไม่ต้องการบริหารจัดการก็ใช้เป็นแบบ Cloudซึ่งกำลังเป็นที่นิยม กันเป็นอย่างมาก

การวิเคราะห์การลงทุนสร้างข้อมูลขนาดใหญ่ หรือ Big data จะมีประโยชน์หรือคุ้มค่า ก็ต่อเมื่อคุณใช้และวิเคราะห์ข้อมูล การวิเคราะห์ข้อมูลทำให้เกิดความกระจ่างและชัดเจนในชุด ข้อมูลที่คุณมีอยู่ การสำรวจข้อมูลยังทำให้เราค้นพบสิ่งใหม่ แชร์สิ่งที่ค้นพบใหม่ๆต่อคนอื่น สร้างรูปแบบจำลองข้อมูล ด้วยการเรียนรู้ของเครื่องจักรและปัญญาประดิษฐ์ AI และนำข้อมูล เหล่านั้นไปใช้งาน

2.2.1.3 เทคโนโลยีการประมวลผลข้อมูล

การประมวลผลข้อมูลที่เป็น Big Data จะมีทั้งการวิเคราะห์ข้อมูลที่เป็น business intelligence (BI) เพื่อที่จะดึงข้อมูลมานำเสนอ หรือการทำ Predictive Analyticsโดยใช้หลักการของ Data Science ความยากของการประมวลผลคือต้องการความเร็วในการประมวลผลข้อมูลที่นอกจากมีขนาดใหญ่แล้วบางครั้งยังเป็นข้อมูลที่ไม่มีโครงสร้าง ดังนั้นต้องจึงมีการนำเทคโนโลยีหรือภาษาต่าง ๆ มาเพื่อให้สามารถประมวลผลข้อมูลได้ ซึ่งในบางครั้งหน่วยงานอาจต้องพิจารณาต้องเลือกใช้

2.2.2 ทฤษฎีเกี่ยวกับการทำเหมืองข้อมูล

เหมืองข้อมูล คือกระบวนการที่กระทำกับข้อมูลจำนวนมากเพื่อค้นหารูปแบบและความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น ในปัจจุบันการทำเหมืองข้อมูลได้ถูกนำไประยุกต์ใช้ในงานหลายประเภท ทั้งในด้านธุรกิจที่ช่วยในการตัดสินใจของผู้บริหาร ในด้านวิทยาศาสตร์และการแพทย์รวมทั้งในด้านเศรษฐกิจและสังคม

2.2.2.1 การทำเหมืองข้อมูล (Data Mining) เปรียบเสมือนวิวัฒนาการหนึ่งในการจัดเก็บและตีความหมายข้อมูลจากเดิมที่มีการจัดเก็บข้อมูลอย่างง่ายๆ มาสู่การจัดเก็บในรูปแบบฐานข้อมูลที่สามารถดึงข้อมูลสารสนเทศมาใช้จนถึงการทำเหมืองข้อมูลที่สามารถค้นพบความรู้ที่ซ่อนอยู่ในข้อมูลหรือจะแยกๆ เป็นข้อๆ ได้ดังนี้

1) กระบวนการหรือการเรียงลำดับของการค้นข้อมูลจำนวนมากและเก็บข้อมูลที่เกี่ยวข้อง

2) การนำมาใช้โดยหน่วยงานทางธุรกิจและนักวิเคราะห์ทางการเงินหรือการนำมาใช้งานในด้านวิทยาศาสตร์เพื่อเอาข้อมูลขนาดใหญ่ที่สร้างโดยวิธีการทดลองและการสังเกตการณ์ที่ทันสมัย

3) การสกัดหรือแยกข้อมูลที่เป็นประโยชน์จากข้อมูลขนาดใหญ่หรือฐานข้อมูล

4) การวางแผนทรัพยากรขององค์กรโดยสามารถวิเคราะห์ทางสถิติและตรรกะของข้อมูลขนาดใหญ่เป็นการมองหารูปแบบที่สามารถช่วยการตัดสินใจได้

2.2.2.2 ขั้นตอนการทำเหมืองข้อมูล

ประกอบด้วยขั้นตอนการทำงานย่อยที่จะเปลี่ยนข้อมูลดิบให้กลายเป็นความรู้ประกอบด้วยขั้นตอนดังนี้

1) Business Understanding เน้นไปที่การทำความเข้าใจในงาน ระบุโอกาสและหาปัญหาที่จะเกิดขึ้น กำหนดขอบเขตของข้อมูลที่จะนำวิเคราะห์ ซึ่งต้องสามารถระบุ ผลลัพธ์ที่มีได้

2) Data Understanding ทำความเข้าใจข้อมูลโดยการรวบรวมข้อมูลที่เกี่ยวข้อง คัดเลือกให้เหลือเพียงข้อมูลที่มีความถูกต้องและสำคัญต่องานมาทำการวิเคราะห์

3) Data Preparation ทำการแปลงข้อมูล (Raw Data) ให้กลายเป็นข้อมูลที่สามารถนำมาช่วยในการวิเคราะห์ต่อไปได้ ขั้นตอนนี้จะใช้เวลามากที่สุดในทุกขั้นตอน เพราะคุณภาพของงานที่ได้จะดีเพียงใดขึ้นอยู่กับคุณภาพข้อมูลที่จัดเตรียมในขั้นนี้ การเตรียมข้อมูลประกอบด้วยการคัดเลือกข้อมูล การกลั่นกรองข้อมูล และแปลงรูปแบบของข้อมูล

4) Modeling การสร้างแบบจำลองเพื่อวิเคราะห์ข้อมูลที่ได้จากขั้นตอนที่ 3 พร้อมทดสอบผลลัพธ์แบบจำลองเพื่อให้ได้คำตอบที่ดีที่สุด บางครั้งอาจมีการย้อนกลับไปปรับ การเตรียมข้อมูลเพื่อให้ได้แบบจำลองที่เหมาะสมที่สุด

5) Evaluation การประเมินผลลัพธ์ที่ได้ก่อนที่จะนำไปใช้จริง ว่าตรงกับวัตถุประสงค์หรือเป้าหมายที่ได้ตั้งไว้หรือมีความน่าเชื่อถือมากน้อยเพียงใด หากไม่ได้ผลลัพธ์ตามวัตถุประสงค์ต้องย้อนกลับไปปรับปรุงแก้ไขการดำเนินงานในขั้นตอนก่อนหน้า

6) Deployment การนำเอาข้อมูลที่เป็นผลลัพธ์จากทั้งหมดมาใช้ประโยชน์ตามวัตถุประสงค์และทำการประเมินผลลัพธ์ที่ได้ว่ามีประสิทธิภาพตรงตามวัตถุประสงค์ที่ตั้งไว้หรือไม่

2.2.3 ทฤษฎีเกี่ยวกับการออกแบบเว็บไซต์

2.2.3.1 หลักในการออกแบบเว็บไซต์

หน้าเว็บเป็นสิ่งแรกที่ผู้ใช้จะได้เห็นขณะที่เปิดเข้าสู่เว็บไซต์และยังเป็นสิ่งแรกที่แสดงถึงประสิทธิภาพในการออกแบบเว็บไซต์อีกด้วย หน้าเว็บจึงเป็นสิ่งสำคัญมากเพราะเป็นสื่อกลางให้ผู้ชมสามารถใช้ประโยชน์จากข้อมูลของระบบงานของเว็บไซต์นั้นได้ โดยปกติหน้าเว็บจะประกอบด้วย รูปภาพ ตัวอักษร สีพื้น ระบบเนวิเกชั่น และองค์ประกอบอื่นๆ ที่ช่วยสื่อความหมายของเนื้อหาและอำนวยความสะดวกต่อการใช้งานหลักสำคัญในการออกแบบหน้าเว็บก็คือ การใช้รูปภาพและองค์ประกอบต่าง ๆ ร่วมกันเพื่อสื่อความหมายเกี่ยวกับเนื้อหาหรือลักษณะสำคัญของเว็บไซต์ โดยมีเป้าหมายสำคัญเพื่อการสื่อความหมายที่ชัดเจนและน่าสนใจ บนพื้นฐานของความเรียบง่ายและความสะดวกของผู้ใช้ การออกแบบเว็บไซต์ต้องคำนึงถึง

1) ความเรียบง่าย ได้แก่ มีรูปแบบที่เรียบง่าย ไม่ซับซ้อน และใช้งานได้สะดวก ไม่มีกราฟิกหรือตัวอักษรที่เคลื่อนไหวอยู่ตลอดเวลา ชนิดและสีของตัวอักษรไม่มากจนเกินไปทำให้วุ่นวาย

2) ความสม่ำเสมอ ได้แก่ ใช้รูปแบบเดียวกันตลอดทั้งเว็บไซต์ เช่น รูปแบบของหน้า สไตล์ของกราฟิก ระบบเนวิเกชั่นและโทนสี ควรมีความคล้ายคลึงกันตลอดทั้งเว็บไซต์

3) ความเป็นเอกลักษณ์ การออกแบบเว็บไซต์ควรคำนึงถึงลักษณะขององค์กรเพราะรูปแบบของเว็บไซต์จะสะท้อนถึงเอกลักษณ์และลักษณะขององค์กรนั้น ๆ เช่น ถ้าเป็นเว็บไซต์ของทาง ราชการ จะต้องดูน่าเชื่อถือไม่เหมือนสวนสนุก ฯลฯ

4) เนื้อหาที่มีประโยชน์ เนื้อหาเป็นสิ่งที่สำคัญที่สุดในเว็บไซต์ ดังนั้นควรจัดเตรียมเนื้อหาและข้อมูลที่ผู้ใช้ต้องการให้ถูกต้องและสมบูรณ์ มีการปรับปรุงและเพิ่มเติมให้ทันเหตุการณ์อยู่เสมอ เนื้อหาไม่ควรซ้ำกับเว็บไซต์อื่น จึงจะดึงดูดความสนใจ

5) ระบบเนวิเกชันที่ใช้งานง่าย ต้องออกแบบให้ผู้ใช้เข้าใจง่ายและใช้งานสะดวก ใช้กราฟิกที่สื่อความหมายร่วมกับคำอธิบายที่ชัดเจน มีรูปแบบและลำดับของรายการที่สม่ำเสมอ เช่น วางไว้ ตำแหน่งเดียวกันของทุกหน้า

6) ลักษณะที่น่าสนใจ หน้าตาของเว็บไซต์จะต้องมีความสัมพันธ์กับคุณภาพขององค์ประกอบต่าง ๆ เช่น คุณภาพของกราฟิกที่จะต้องสมบูรณ์ การใช้สี การใช้ตัวอักษรที่อ่านง่าย สบายตา การใช้โทนสีที่เข้ากันลักษณะหน้าตาที่น่าสนใจนั้นขึ้นอยู่กับความชอบของแต่ละบุคคล

7) การใช้งานอย่างไม่จำกัด ผู้ใช้ส่วนใหญ่สามารถเข้าถึงได้มากที่สุด เลือกใช้บราวเซอร์ชนิดใดก็ได้ในการเข้าถึงเนื้อหาสามารถแสดงผลได้ทุกระบบปฏิบัติการและความละเอียดหน้าจอต่าง ๆ กันอย่างไม่มีปัญหาเป็นลักษณะสำคัญสำหรับผู้ใช้ที่มีจำนวนมาก

8) คุณภาพในการออกแบบ การออกแบบและเรียบเรียงเนื้อหาอย่างรอบคอบ สร้างความรู้สึกว่าเว็บไซต์มีคุณภาพ ถูกต้อง และเชื่อถือได้

9) ลิงค์ต่าง ๆ จะต้องเชื่อมโยงไปหน้าที่มีอยู่จริงและถูกต้อง ระบบการทำงานต่าง ๆ ในเว็บไซต์จะต้องมีความแน่นอนและทำหน้าที่ได้อย่างถูก

2.2.3.2 โครงสร้างเว็บไซต์

1) เว็บที่มีโครงสร้างแบบเรียงลำดับ (Sequential Structure) เป็นโครงสร้างแบบธรรมดาที่ใช้กันมากที่สุดเนื่องจากง่ายต่อการจัดระบบข้อมูล ข้อมูลที่นิยม จัดด้วยโครงสร้างแบบนี้มักเป็นข้อมูลที่มีลักษณะเป็นเรื่องราวตามลำดับของเวลา เช่น การเรียงลำดับตามตัวอักษร ดรรชนี สารานุกรม หรืออภิธานศัพท์ โครงสร้างแบบนี้เหมาะกับเว็บไซต์ที่มีขนาดเล็กเนื้อหาไม่ซับซ้อนใช้การลิงก์ (Link) ไปทีละหน้า ทิศทางของการเข้าสู่เนื้อหา (Navigation) ภายในเว็บจะเป็นการดำเนินเรื่องในลักษณะเส้นตรงโดยมีปุ่มเดินหน้าถอยหลังเป็นเครื่องมือหลักในการกำหนดทิศทาง ข้อเสียของโครงสร้างระบบนี้คือ ผู้ใช้ไม่สามารถกำหนดทิศทางการเข้าสู่เนื้อหาของตนเองได้ทำให้เสียเวลาเข้าสู่เนื้อ

ภาพที่ 2.9 เว็บที่มีโครงสร้างแบบเรียงลำดับ

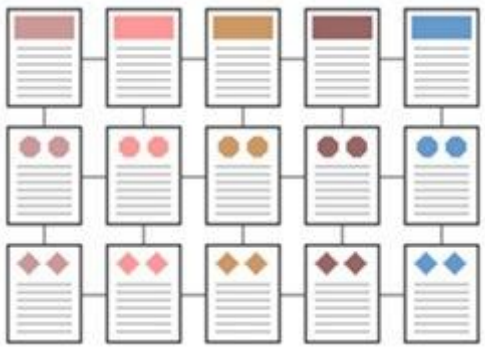
ที่มา : sites.google.com

2) เว็บที่มีโครงสร้างแบบลำดับขั้น (Hierarchical Structure) เป็นวิธีที่ดีที่สุดวิธีหนึ่งในการจัดระบบโครงสร้างที่มีความซับซ้อนของข้อมูล โดยแบ่งเนื้อหา ออกเป็นส่วนต่าง ๆ และมีรายละเอียดย่อย ในแต่ละส่วนลดหลั่นกันมาในลักษณะแนวคิดเดียวกับ แผนภูมิองค์กร จึงเป็นการง่ายต่อการทำความเข้าใจกับโครงสร้างของเนื้อหาในเว็บลักษณะนี้ ลักษณะเด่นเฉพาะของเว็บประเภทนี้คือการมีจุดเริ่มต้นที่จุดร่วมจุดเดียวนั่นคือ โฮมเพจ (Homepage) และเชื่อมโยงไปสู่เนื้อหา ในลักษณะเป็นลำดับจากบนลงล่าง

ภาพที่ 2.10 เว็บที่มีโครงสร้างแบบลำดับขั้น

ที่มา : sites.google.com

3) เว็บที่มีโครงสร้างแบบตาราง (Grid Structure) โครงสร้างรูปแบบนี้มีความซับซ้อนมากกว่ารูปแบบที่ผ่านมา การออกแบบเพิ่มความยืดหยุ่น ให้แก่การเข้าสู่เนื้อหาของผู้ใช้ โดยเพิ่มการเชื่อมโยงซึ่งกันและกันระหว่างเนื้อหาแต่ละส่วน เหมาะแก่การแสดงให้เห็นความสัมพันธ์กันของเนื้อหา การเข้าสู่เนื้อหาของผู้ใช้จะไม่เป็นลักษณะเชิงเส้นตรง เนื่องจากผู้ใช้สามารถเปลี่ยนทิศทางการเข้าสู่เนื้อหาของตนเองได้

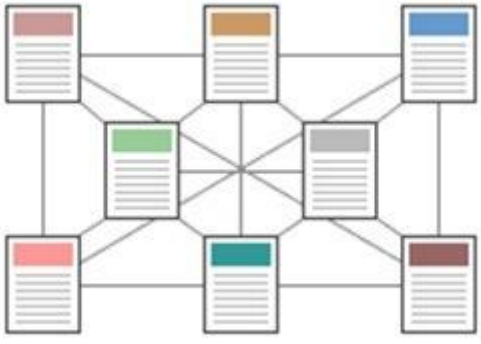


ภาพที่ 2.11 เว็บที่มีโครงสร้างแบบตาราง

ที่มา : sites.google.com

ในการจัดระบบโครงสร้างแบบนี้ เนื้อหาที่นำมาใช้แต่ละส่วนควรมีลักษณะที่เหมือนกันและสามารถใช้รูปแบบร่วมกัน หลักการออกแบบคือนำหัวข้อทั้งหมดมาบรรจุลงในที่เดียวกันซึ่งโดยทั่วไป จะเป็นหน้าแผนภาพ (Map Page) ที่แสดงในลักษณะเดียวกับโครงสร้างของเว็บ เมื่อผู้ใช้คลิกเลือก หัวข้อใด ก็จะเข้าไปสู่หน้าเนื้อหา (Topic Page) ที่แสดงรายละเอียดของหัวข้อนั้น ๆ และภายในหน้านั้น ก็จะมีการเชื่อมโยงไปยังหน้ารายละเอียดของหัวข้ออื่นที่เป็นเรื่องเดียวกัน นอกจากนี้ยังสามารถนำโครงสร้างแบบเรียงลำดับและแบบลำดับขั้นมาใช้ร่วมกันได้อีกด้วย ถึงแม้โครงสร้างแบบนี้ อาจจะสร้างความยุ่งยากในการเข้าใจได้และอาจเกิดปัญหาการคงค้าง ของหัวข้อ (Cognitive Overhead) ได้ แต่จะเป็นประโยชน์ที่สุดเมื่อผู้ใช้ได้เข้าใจถึงความสัมพันธ์ระหว่างเนื้อหา ในส่วนของการออกแบบจำเป็นจะต้องมีการวางแผนที่ดี เนื่องจากมีการเชื่อมโยงที่เกิดขึ้นได้หลายทิศทาง นอกจากนี้การปรับปรุงแก้ไขอาจเกิดความยุ่งยากเมื่อต้องเพิ่มเนื้อหาในภายหลัง

4) เว็บที่มีโครงสร้างแบบใยแมงมุม (Web Structure) โครงสร้างประเภทนี้จะมีความยืดหยุ่นมากที่สุด ทุกหน้าในเว็บสามารถจะเชื่อมโยงไปถึงกัน ได้หมด เป็นการสร้างรูปแบบการเข้าสู่เนื้อหาที่เป็นอิสระ ผู้ใช้สามารถกำหนดวิธีการเข้าสู่เนื้อหาได้ด้วยตนเอง การเชื่อมโยงเนื้อหาแต่ละหน้าอาศัยการโยงใยข้อความที่มีมโนทัศน์ (Concept) เหมือนกันของแต่ละหน้าในลักษณะของไฮเปอร์เท็กซ์หรือไฮเปอร์มีเดีย โครงสร้างลักษณะนี้จัดเป็นรูปแบบที่ ไม่มีโครงสร้างที่แน่นนอนตายตัว (Unstructured) นอกจากนี้การเชื่อมโยงไม่ได้จำกัดเฉพาะเนื้อหาภายในเว็บนั้น ๆ แต่สามารถเชื่อมโยงออกไปสู่เนื้อหาจากเว็บภายนอกได้



ภาพที่ 2.12 เว็บที่มีโครงสร้างแบบใยแมงมุม

ที่มา : sites.google.com

ลักษณะการเชื่อมโยงในเว็บนั้น นอกเหนือจากการใช้ไฮเปอร์เท็กซ์หรือไฮเปอร์มีเดียกับข้อความที่มีมโนทัศน์ (Concept) เหมือนกันของแต่ละหน้าแล้ว ยังสามารถใช้ลักษณะการเชื่อมโยง จากรายการที่รวบรวมชื่อหรือหัวข้อของเนื้อหาแต่ละหน้าไว้ ซึ่งรายการนี้จะปรากฏอยู่บริเวณใดบริเวณหนึ่งในหน้าจอ ผู้ใช้สามารถคลิกที่หัวข้อใดหัวข้อหนึ่งในรายการเพื่อเลือกที่จะเข้าไปสู่หน้าใดๆ ก็ได้ตามความต้องการ ข้อดีของรูปแบบนี้คือง่ายต่อผู้ใช้ในการท่องเที่ยวบนเว็บ โดยผู้ใช้สามารถกำหนดทิศทางการเข้าสู่เนื้อหาได้ด้วยตนเอง แต่ข้อเสียคือถ้ามีการเพิ่มเนื้อหาใหม่ๆ อยู่เสมอจะเป็นการยากในการปรับปรุง นอกจากนี้การเชื่อมโยงระหว่างข้อมูลที่มีมากมายนั้นอาจทำให้ผู้ใช้เกิดการสับสนและเกิดปัญหาการคงค้างของหัวข้อ (Cognitive Overhead) ได้

2.2.3.3 การใช้สีในการออกแบบเว็บไซต์

การสร้างสีสันบนหน้าเว็บเป็นสิ่งที่สื่อความหมายของเว็บไซต์ได้อย่างชัดเจนการเลือกใช้สีให้เหมาะสม กลมกลืน ไม่เพียงแต่จะสร้างความพึงพอใจให้กับผู้ใช้ แต่ยังสามารถทำให้เห็นถึงความแตกต่างระหว่างเว็บไซต์ได้ สีเป็นองค์ประกอบหลักสำหรับการตกแต่งเว็บ จึงจำเป็นอย่างยิ่งที่จะต้องทำความเข้าใจเกี่ยวกับการใช้สีระบบสีที่แสดงบนจอคอมพิวเตอร์ มีระบบการแสดงผลผ่านหลอดลำแสงที่เรียกว่า CRT (Cathode ray tube) โดยมีลักษณะระบบสีแบบบวก อาศัยการผสมของของแสงสีแดง สีเขียว และสีน้ำเงิน หรือระบบสี RGB สามารถกำหนดค่าสีจาก 0 ถึง 255 ได้ จากการรวมสีของแม่สีหลักจะทำให้เกิดแสงสีขาว มีลักษณะเป็นจุดเล็ก ๆ บนหน้าจอไม่สามารถมองเห็นด้วยตาเปล่าได้ จะมองเห็นเป็นสีที่ถูกผสมเป็นเนื้อสีเดียวกันแล้ว จุดแต่ละจุด หรือ พิกเซล (Pixel) เป็นส่วนประกอบของภาพบนหน้าจอคอมพิวเตอร์ โดยจำนวนบิตที่ใช้ในการกำหนดความสามารถของการแสดงสีต่าง ๆ เพื่อสร้างภาพบนจอนั้นเรียกว่า บิตเด็ป (Bit depth) ในภาษา HTML มีการกำหนดสีด้วยระบบเลขฐานสิบหก ซึ่งมีเครื่องหมาย (#) อยู่ด้านหน้าและตามด้วยเลขฐานสิบหกจำนวนอักษรอีก 6 หลัก โดยแต่ละไบต์ (byte) จะมีตัวอักษรสองตัว แบ่งออกเป็น 3 กลุ่ม เช่น #FF12AC การใช้ตัวอักษรแต่ละไบต์นี้เพื่อกำหนดระดับความเข้มของแม่สีแต่ละสีของชุดสี RGB โดย 2 หลักแรกแสดงถึงความเข้มของสีแดง 2 หลักต่อมา แสดงถึงความเข้มของสีเขียว 2 หลักสุดท้ายแสดงถึงความเข้มของสีน้ำเงิน

สีมีอิทธิพลในเรื่องของอารมณ์การสื่อความหมายที่เด่นชัด กระตุ้นการรับรู้ทางด้านจิตใจมนุษย์ สีแต่ละสีให้ความรู้สึกถึงอารมณ์ที่ไม่เหมือนกัน สีบางสีให้ความรู้สึกสงบ บางสีให้ความรู้สึกตื่นเต้นรุนแรง สีจึงเป็นปัจจัยสำคัญอย่างยิ่งต่อการออกแบบเว็บไซต์ ดังนั้นการเลือกใช้โทนสีภายในเว็บไซต์เป็นการแสดงถึงความแตกต่างของสีที่แสดงออกทางอารมณ์ มีชีวิตชีวาหรือเศร้าโศก รูปแบบของสีที่สายตาของมนุษย์มองเห็น สามารถแบ่งออกเป็น 3 กลุ่ม คือ

1) สีโทนร้อน (Warm Colors) เป็นกลุ่มสีที่แสดงถึงความสุข ความปลอบโยน ความอบอุ่น และดึงดูดใจ สีกลุ่มนี้เป็นกลุ่มสีที่ช่วยให้หายจากความเฉื่อยชา มีชีวิตชีวามากยิ่งขึ้น

2) สีโทนเย็น (Cool Colors) แสดงถึงความที่ดูสุภาพ อ่อนโยน เรียบร้อย เป็นกลุ่มสีที่มีคนชอบมากที่สุด สามารถโน้มนาวในระยะไกลได้

3) สีโทนกลาง (Neutral Colors) สีที่เป็นกลาง ประกอบด้วย สีดำ สีขาว สี เทา และสีน้ำตาล กลุ่มสีเหล่านี้คือ สีกลางที่สามารถนำไปผสมกับสีอื่นๆ เพื่อให้เกิดสีกลางขึ้นมา

2.2.4 ทฤษฎีเกี่ยวกับชุดคำสั่งภาษาของการออกแบบเว็บไซต์

2.2.4.1 CSS (ย่อมาจาก Cascading Style Sheet) มักเรียกโดยย่อว่า "สไตล์ชีต" คือภาษาที่ใช้เป็นส่วนของการจัดรูปแบบการแสดงผลเอกสาร HTML โดยที่ CSS กำหนดกฎเกณฑ์ในการระบุรูปแบบ (หรือ "Style") ของเนื้อหาในเอกสารอันได้แก่ สีของข้อความ สีพื้นหลัง ประเภทตัวอักษร และการจัดวางข้อความ ซึ่งการกำหนดรูปแบบ หรือ Style นี้ใช้หลักการของการแยกเนื้อหาเอกสาร HTML ออกจากคำสั่งที่ใช้ในการจัดรูปแบบการแสดงผล กำหนดให้รูปแบบของการแสดงผลเอกสาร ไม่ขึ้นอยู่กับเนื้อหาของเอกสาร เพื่อให้ง่ายต่อการจัดรูปแบบการแสดงผลลัพธ์ของเอกสาร HTML โดยเฉพาะในกรณีที่มีการเปลี่ยนแปลงเนื้อหาเอกสารบ่อยครั้ง หรือต้องการควบคุมให้รูปแบบการแสดงผลเอกสาร HTML มีลักษณะของความสม่ำเสมอทั่วกันทุกหน้าเอกสารภายในเว็บไซต์เดียวกัน โดยกฎเกณฑ์ในการกำหนดรูปแบบ (Style) เอกสาร HTML ถูกเพิ่มเข้ามาครั้งแรกใน HTML 4.0 เมื่อปีพ.ศ. 2539 ในรูปแบบของ CSS level 1 Recommendations ที่กำหนดโดย องค์กร World Wide Web Consortium หรือ W3C

2.2.4.2 Bootstrap คือชุดคำสั่งที่ประกอบด้วยภาษา CSS, HTML และ Java script เป็นชุดคำสั่งที่ถูกพัฒนาขึ้นมาเพื่อกำหนดกรอบหรือรูปแบบการพัฒนาเว็บไซต์ในส่วนของการปฏิสัมพันธ์กับผู้ใช้งานเว็บไซต์ (User Interface) เราจึงสามารถเรียก Bootstrap ว่าเป็น Front-end framework คือใช้สำหรับ พัฒนาเว็บไซต์ส่วนการแสดงผล ซึ่งแตกต่างจากภาษาประเภท Server Side Script อย่าง PHP, Python หรือภาษาอื่นๆ

2.2.5 ทฤษฎีเกี่ยวกับการ visualization

Data Visualization หรือ Information Visualization คือ การนำข้อมูลในเชิงปริมาณ ทั้งที่จัดเก็บไว้ในรูปแบบของข้อมูลจำนวนน้อย และข้อมูลจำนวนมหาศาล (Big Data) มา ประมวลผล จากนั้นจึงนำมาแสดงผลในรูปแบบของกราฟ แผนภูมิอินฟอร์เมชันกราฟิกหรือแม้กระทั่งอินเทอร์แอคทีฟกราฟิก ที่ผู้บริโภคสามารถคลิกหรือมีปฏิสัมพันธ์กับกราฟิกนั้นๆ ได้ ซึ่งปัจจุบันสื่อหลายสำนักในประเทศไทย เริ่มมีการนำเสนอข่าวโดยการนำข้อมูลแบบประยุกต์ด้วยภาพ (Data Visualization) การถ่ายทอดข้อมูลในเชิงปริมาณที่มีความซับซ้อนได้อย่างมีประสิทธิผล ซึ่งคำว่า “ประสิทธิผล” ในที่นี้หมายถึงมีความชัดเจน (Clarity), มีความแม่นยำ (Precision), และมีประสิทธิภาพ (Efficiency) หากไม่มีการทำ Data Visualization แล้วอาจทำาให้ เราไม่สามารถค้นพบนัยยะของข้อมูลในแง่ของแนวโน้ม, รูปแบบพฤติกรรม, และความสัมพันธ์ เชื่อมโยงได้

Visualization คือ การจินตนาการ หรือสร้างภาพขึ้นในความคิด ซึ่งเป็นกระบวนการ ทำให้เกิดความรู้ความเข้าใจสำหรับเรื่องที่จินตนาการยาก เข้าใจยาก วิธีการที่เป็นทางลัดก็คือ การสร้าง ภาพ ให้เป็น บันไดความคิด ไปสู่ การใช้ความคิดอีกระดับ บันไดนี้จะช่วยตัดปริมาณ ข้อมูล ช่วยลดภาระการคำนวณหรือการนำไปผ่านหลากหลายกระบวนการความคิด เพื่อ นำไปสู่คำตอบที่ต้องการได้โดยเร็วและถูกต้อง

Visualization System คือระบบ ซึ่งประกอบด้วยกระบวนการต่าง ๆ ที่ออกแบบมา เพื่อสร้าง รักษา นำไปใช้ และปรับปรุงทัศนสนเทศ เพื่อทำให้เกิดความรู้ความเข้าใจ อย่างถูกต้อง รวดเร็ว และได้ผลเป็นอย่างสูง

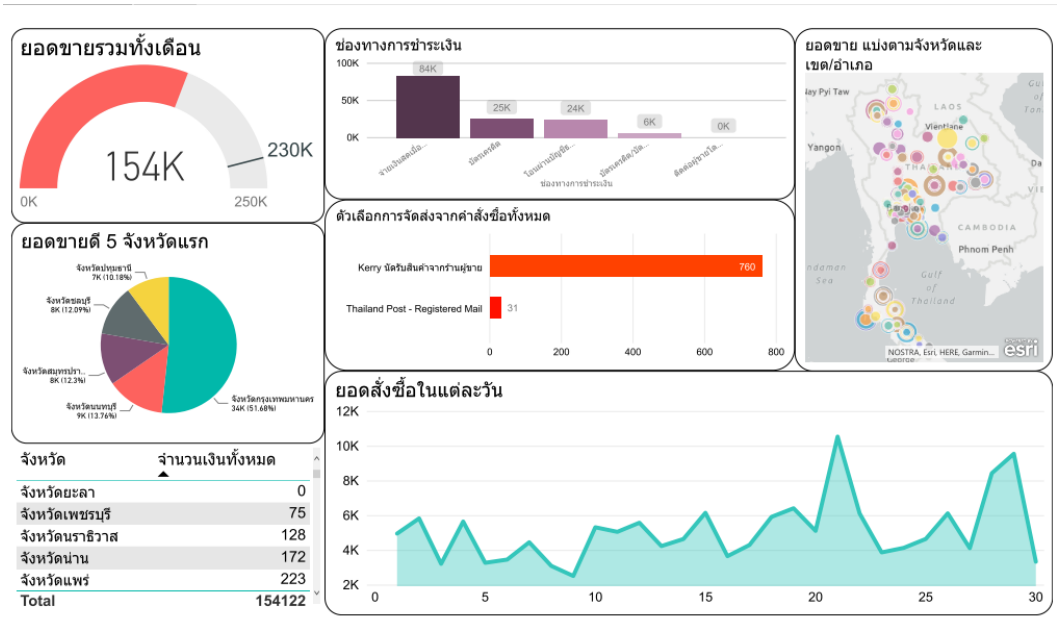
Data Visualization เป็นการนำข้อมูลมาผสมผสานกับจินตนาการ เพื่อสร้างภาพใน ความคิดขึ้นมา ซึ่งมีกระบวนการนำเสนอข้อมูลที่มีความซับซ้อนหรือข้อมูลเชิงปริมาณให้สามารถเข้าใจได้ง่าย ในแบบของ กราฟ แผนภูมิ

2.2.5.1 โปรแกรมสำหรับการสร้างแดชบอร์ด (Data Visualization)

Tools ที่นิยมใช้ในตอนนี้ได้แก่ RapidMiner Studio , Weka , Power BI , Microsoft Excel , Google Chart , Google Data Studio และอื่นๆ อีกมากมาย

2.2.5.2 รูปแบบในการใช้ Data Visualization

การนำเสนอแบบทิศทางหรือแนวโน้ม (Trending) เราใช้กราฟที่แสดงผลแบบทิศทางหรือแนวโน้ม เพื่อนำเสนอข้อมูลให้เห็นจำนวนข้อมูลที่เกิดขึ้นในแต่ละช่วงเวลา (period) รวมถึงเน้นข้อมูลที่ต้องการนำเสนอ เช่น Line Chart, Bar Chart, Radar Chart, Area Chart เป็นต้น



ภาพที่ 2.13 การนำเสนอแบบทิศทางหรือแนวโน้ม (Trending)

ที่มา : medium.com

2.3 เครื่องมือในการออกแบบและวิเคราะห์ข้อมูล

2.3.1 แบบจำลองต้นไม้ตัดสินใจ (Decision Tree)

2.3.1.1 เทคนิคต้นไม้ตัดสินใจ (Decision Tree)

เป็นเทคนิคหนึ่งของ Classification ซึ่งเป็นวิธีการแบ่งประเภทหรือแยกหมวดหมู่ข้อมูล โดย Classification นั้นเป็นเทคนิคหนึ่งของเหมืองข้อมูล (Data Mining)

การเรียนรู้แบบต้นไม้ตัดสินใจ (อังกฤษ: decision tree learning) เป็นหนึ่งในวิธีการเรียนรู้ซึ่งใช้ในสถิติ, การเรียนรู้ของเครื่อง และการทำเหมืองข้อมูล โดยพิจารณาการสังเกตการแบ่งแยกข้อมูลโดยพิจารณาข้อมูล

ในการเรียนรู้ของเครื่อง (machine learning) ต้นไม้ตัดสินใจ เป็นโมเดลทางคณิตศาสตร์ที่ใช้ทำนายประเภทของวัตถุโดยพิจารณาจากลักษณะของวัตถุ บัพภายใน (inner node) ของต้นไม้จะแสดงตัวแปร ส่วนกิ่งจะแสดงค่าที่เป็นไปได้ของตัวแปร ส่วนบัพใบ (leaf node) จะแสดงประเภทของวัตถุ

ต้นไม้ตัดสินใจที่บัพใบแสดงถึงข้อมูลที่เป็นข้อมูลไม่ต่อเนื่อง (discrete values) จะเรียกว่าต้นไม้ตัดสินใจแบบจำแนก (classification trees) และต้นไม้ตัดสินใจที่บัพใบเป็นข้อมูลต่อเนื่อง (continuous values) จะเรียกว่าต้นไม้ตัดสินใจแบบถดถอย (regression trees)

ต้นไม้การตัดสินใจในการบริหารธุรกิจ เป็นแผนผังต้นไม้ช่วยในการตัดสินใจ โดยแสดงถึงมูลค่าของทรัพยากรที่จะใช้ ความเสี่ยงในการลงทุนและผลลัพธ์ที่มีโอกาสเกิดขึ้น ต้นไม้ตัดสินใจสร้างขึ้นเพื่อช่วยการตัดสินใจเพื่อใช้ในการสร้างแผนงาน นิยมใช้มากในการบริหารความเสี่ยง (risk management) ต้นไม้ตัดสินใจเป็นส่วนหนึ่งของทฤษฎีการตัดสินใจ (decision theory) และ ทฤษฎีกราฟ ต้นไม้ตัดสินใจเป็นวิธีการพื้นฐานอย่างหนึ่งสำหรับการทำเหมืองข้อมูล

2.3.1.2 ขั้นตอนในการสร้าง Decision Tree เพื่อใช้จำแนกข้อมูล มีดังนี้

1) เลือก Attribute ที่ทำหน้าที่เป็น Root Node34

2) จาก Root Node สร้างเส้นเชื่อมโยงไปยังโหนดลูก จำนวนเส้นเชื่อมโยงจะเท่ากับจำนวนค่าที่เป็นไปได้ทั้งหมดของ Attribute ที่เป็น root node

3) ถ้าโหนดลูกเป็นกลุ่มของข้อมูลที่อยู่ในคลาสเดียวกันทั้งหมด ให้หยุดสร้างต้นไม้ แต่ถ้าโหนดลูกมีข้อมูลของหลายคลาสปะปนกันอยู่ ต้อง สร้าง subtree เพื่อจำแนกข้อมูลต่อไป โดยเลือก subtree มาทำหน้าที่ เป็น root node ของ subtree มาทำซ้ำในขั้นตอนที่ 2.) , 3.) ซึ่งการคำนวณมีดังนี้

เอนโทรปี (Entropy) การสร้างต้นไม้การตัดสินใจจากบนลงล่างด้วยการถามว่าลักษณะใด ควรจะเป็นรากของต้นไม้การตัดสินใจต้นนี้ และถามซ้ำๆ ไปเรื่อยๆ เพื่อหาต้นไม้ทั้งต้นด้วยการเขียนโปรแกรมด้วยความสัมพันธ์แบบเวียนเกิด (อังกฤษ: recursion) โดยในการเลือกว่าลักษณะใดดีที่สุดนั้นดูจากค่าของลักษณะเรียกว่าเกนความรู้ (Information gain) ก่อนที่จะรู้จักเกนความรู้จะต้องนิยามค่าหนึ่งที่ใช้บอกความไม่บริสุทธิ์ของข้อมูลก่อน เรียกว่าเอนโทรปี (Entropy) โดยนิยามเอนโทรปีของต้นไม้การตัดสินใจในตัวในเซตของตัวอย่าง S คือ E (S) ดังนี้

$$E(S) = -\sum_{j=1}^{n} p_S(j) \log_2 p_S(j)$$

เมื่อแทนค่า S คือ ตัวอย่างที่ประกอบด้วยชุดของตัวแปรต้นและตัวแปรตามหลายๆ กรณี$Ps_{(j)}$ คือ อัตราส่วนของกรณีใน S ที่ตัวแปรตามหรือผลลัพธ์มีค่า j

โดยสำหรับต้นไม้การตัดสินใจที่มีผลลัพธ์เป็นแค่เพียงค่าตรรกะ (Boolean) ใช่กับไม่ใช่เหมือนกับที่ยกมาตอนต้นของบทความนั้น จะมีเอนโทรปี คือ

$$E(S) = -p_{yes} log_2(p_{yes}) - p_{no} log_2(p_{no})$$

เมื่อพิจารณาเอนโทรปีแล้วจะเห็นว่าเอนโทรปีจะมีค่าอยู่ระหว่าง 0 กับ 1 โดยจะมีค่าเป็นศูนย์เมื่อทุกๆ กรณีมีผลลัพธ์เพียงแบบเดียว เช่น ใช่ทั้งหมด หรือ ไม่ใช่ทั้งหมด และจะมีค่ามากขึ้นเมื่อเริ่มมีค่าที่แตกต่างกันมากขึ้น หรือจะพูดอีกนัยหนึ่งก็คือเอนโทรปีจะมีค่ามากขึ้นหากข้อมูลไม่บริสุทธิ์ และจะตัดสินใจได้ว่าผลลัพธ์จะเป็นอะไรเมื่อเอนโทรปี เป็น 0 เท่านั้น

เกนความรู้ (Information Gain) ซึ่งจากการนิยามเอนโทรปีข้างต้น ทำให้เราสามารถนิยามลักษณะของตัวแปรต้นที่ดีได้ โดยตัวแปร A จะเป็นตัวแปรต้นที่ดีก็ต่อเมื่อหากว่าแบ่งข้อมูลตัวอย่าง (Example) ออกเป็นชุด มีจำนวนชุดตามจำนวนค่าของ A ที่เป็นไปได้ เพื่อให้แต่ละกรณี (Instance) ในชุดนั้นมีค่า A เพียงค่าเดียวและค่าเฉลี่ยของเอนโทรปีของชุดข้อมูลที่ถูกแบ่งออก (partition) มานั้นต่ำที่สุด เรียกค่าคาดหวังของการลดลงของเอนโทรปีหลังจากข้อมูลถูกแบ่งด้วย A ว่าเกนความรู้ของ A นิยามโดย

$$Gain(S, A) = E(S) - \sum_{v=value(A)} \frac{|S_v|}{|S|} E(S_v)$$

เมื่อแทนค่า □ คือตัวอย่างที่ประกอบด้วยชุดของตัวแปรต้นและตัวแปรตามหลายๆ กรณี

E คือเอนโทรปีของตัวอย่าง

A คือตัวแปรต้นที่พิจารณา

Value(A) คือเซตของค่าของ A ที่เป็นไปได้

S_v คือตัวอย่างที่ A มีค่า v ทั้งหมด

จะเห็นว่าหากเกนความรู้ของ A ยิ่งมากแสดงว่าหลังจากแบ่งตัวอย่าง S ด้วย A แล้วในแต่ละชุดที่แบ่งได้จะมี Entropy เข้าใกล้ศูนย์มากยิ่งขึ้น ทำให้ใกล้ที่จะตัดสินใจได้มากขึ้น เกนความรู้จึงเป็นค่าที่ดีที่จะบอกความดีของตัวแปรต้นที่นำมาพิจารณา

2.3.2 การประเมินผลโมเดล (Decision Tree)

สถิติในการทดสอบหาค่าความแม่นยำของอัลกอริทึม และทดสอบความถูกต้องและความน่าเชื่อถือของโมเดลโดยใช้การประเมินค่าความแม่นยา Confusion Matrix

Confusion Matrix คือ การประเมินผลลัพธ์การทำนาย (หรือผลลัพธ์จากโปรแกรม)

เปรียบเทียบกับผลลัพธ์จริงโดยที่

a, d คือ จำนวนข้อมูลที่มีในการทำนายถูก

b, c คือ จำนวนข้อมูลที่มีในการทำนายผิด

		PREDICTED CLASS	
		Class=Yes	***Class=No***
ACTUAL CLASS	***Class=Yes***	***a***	***b***
	Class=No	***c***	***d***

ภาพที่ 2.14 Confusion Matrix

แสดงเป็นสมการได้ ดังนั้น

Sensitivity or Recall คือค่าที่บอกว่า โปรแกรมทำนายได้ว่าจริง เป็นอัตราส่วนเท่าไหร่ของจริงทั้งหมด คำนวณได้จากสมการ

$$Recall = \frac{a}{a+b}$$

Precision คือค่าที่บอกว่า โปรแกรมทำนายว่าจริงถูกต้องเท่าไหร่ คำนวณได้จากสมการ

$$\text{Precision} = \frac{a}{a+c}$$

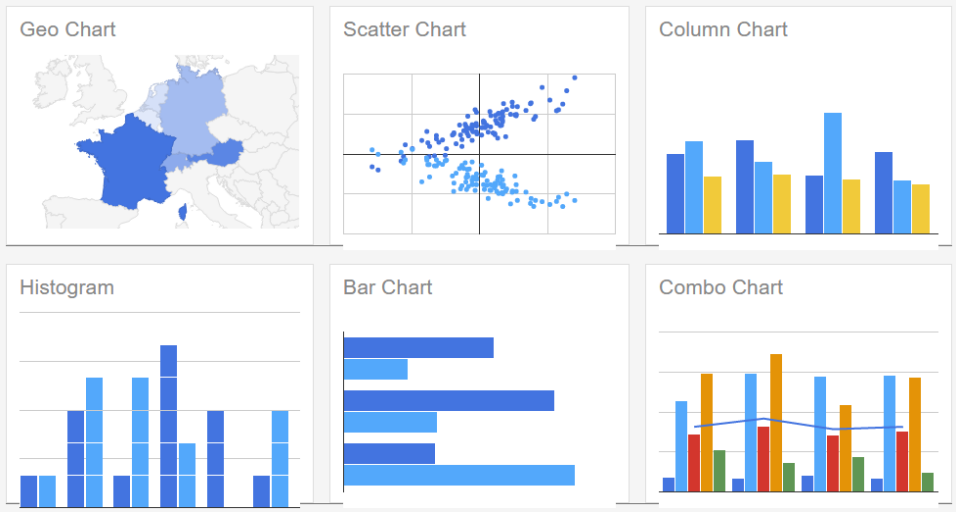
Correctly Classified Instances คือค่าที่บอกว่ามีการทำนายข้อมูลถูกต้อง และมีค่าความแม่นยำเท่าไหร่ในการทำนาย คำนวณได้จากสมการ

$$\text{Correctly Classified Instances } = \frac{(a+d)}{(a+b+c+d)}$$

2.3.3 การแสดงผลแบบ visualization

2.3.3.1 การแสดงผลแบบ visualization ด้วย Google Charts , Google Data Studio

google ได้เตรียมหรือทำขึ้นมาให้เราสามารถนำ script ที่เป็นภาษา java script มาประยุกต์ใช้กับระบบงานของเราได้ตามต้องการ เช่น line chart, bar chart, scatter chart, maps, gauge chart เป็นต้น โดยเราสามารถดูรายละเอียดเพิ่มเติมได้จาก Chart Gallery นอกจากนี้เรายังสามารถทดสอบ run script ของเราก่อน โดยคลิกไปที่ปุ่ม ซึ่งอยู่ใต้ chart และเราสามารถทดลองปรับแต่ง script ซึ่งมี 4 ส่วนคือ html, java script, css และ result เพื่อลองรันก่อนนำ script ไปใช้งาน เป็นต้น



ภาพที่ 2.15 การแสดงแดชบอร์ดของ Google Chart

ที่มา : https://bit.ly/3v4snQk

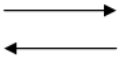
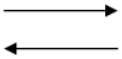
2.3.4 แผนภาพกระแสข้อมูล (Data Flow Diagram : DFD)

แผนภาพกระแสข้อมูล (DFD) เป็นเครื่องมือที่ใช้กันอย่างแพร่หลายในการเขียนแบบระบบใหม่ในการเขียนแผนภาพจำลองการทำงานของกระบวนการ (Process) ต่าง ๆ ในระบบ โดยเฉพาะกับระบบที่ "หน้าที่" ของระบบมีความสำคัญและมีความสลับซับซ้อนมากกว่าข้อมูลที่ไหลเข้า

วัตถุประสงค์ของการสร้างแผนภาพกระแสข้อมูล

1. เป็นแผนภาพที่สรุปรวมข้อมูลทั้งหมดที่ได้จากการวิเคราะห์ในลักษณะของรูปแบบที่เป็นโครงสร้าง
2. เป็นข้อตกลงร่วมกันระหว่างนักวิเคราะห์ระบบและผู้ใช้งาน
3. เป็นแผนภาพที่ใช้ในการพัฒนาต่อในขั้นตอนของการออกแบบระบบ
4. เป็นแผนภาพที่ใช้ในการอ้างอิง หรือเพื่อใช้ในการพัฒนาต่อในอนาคต
5. ทราบที่มาที่ไปของข้อมูลที่ไหลไปในกระบวนการต่าง ๆ (Data and Process)

สัญลักษณ์ที่ใช้ในแผนภาพกระแสข้อมูล

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

ภาพที่ 2.16 สัญลักษณ์ที่ใช้ในแผนภาพกระแสข้อมูล

ที่มา : http://analysisanddesignn.blogspot.com/p/42-4.html

Process หรือ ขั้นตอนการดำเนินงาน คือ งานที่ดำเนินการ/ตอบสนองข้อมูลที่รับเข้า หรือดำเนินการ/ตอบสนองต่อเงื่อนไข/ สภาวะใด ๆ ที่เกิดขึ้น ไม่ว่าขั้นตอนการดำเนินงานนั้นจะกระทำโดยบุคคล หน่วยงาน หุ่นยนต์ เครื่องจักร หรือ เครื่องคอมพิวเตอร์ก็ตาม โดยจะเป็นกริยา (Verb)

เส้นทางการไหลของข้อมูล (Data Flows) เป็นการสื่อสารระหว่างขั้นตอนการทำงาน (Process) ต่าง ๆ และสภาพแวดล้อมภายนอกหรือภายในระบบ โดยแสดงถึงข้อมูลที่นำเข้าไปในแต่ละ Process และข้อมูลที่ส่งออกจาก Process ใช้ในการแสดงถึงการบันทึกข้อมูล การลบข้อมูล การแก้ไขข้อมูลต่าง ๆ สัญลักษณ์ที่ใช้อธิบายเส้นทางการไหลของข้อมูลคือเส้นตรงที่ประกอบด้วยหัวลูกศรตรงปลายเพื่อบอกทิศทางการเดินทางหรือการไหลของข้อมูล

ตัวแทนข้อมูล (External Agents) หมายถึง บุคคล หน่วยงานในองค์กร องค์กรอื่น ๆ หรือระบบงานอื่น ๆ ที่อยู่ภายนอกขอบเขตของระบบ แต่มีความสัมพันธ์กับระบบ โดยมีการส่งข้อมูลเข้าสู่ระบบเพื่อดำเนินงาน และรับข้อมูลที่ผ่านการดำเนินงานเรียบร้อยแล้วจาก

ระบบ สัญลักษณ์ที่ใช้อธิบาย คือ สี่เหลี่ยมจัตุรัส หรือสี่เหลี่ยมผืนผ้า ภายในจะต้องแสดงชื่อของ External Agent โดยสามารถทำการซ้ำ (Duplicate) ได้ด้วยการใช้เครื่องหมาย \ (back slash) ตรงมุมล่างซ้าย

แหล่งจัดเก็บข้อมูล (Data Store) เป็นแหล่งเก็บ/บันทึกข้อมูล เปรียบเสมือนคลังข้อมูล (เทียบเท่ากับไฟล์ข้อมูล และฐานข้อมูล) โดยอธิบายรายละเอียดและคุณสมบัติเฉพาะตัวของสิ่งที่ต้องการเก็บ/บันทึก สัญลักษณ์ที่ใช้อธิบายคือสี่เหลี่ยมเปิดหนึ่งข้าง แบ่งออกเป็นสองส่วน ได้แก่ ส่วนที่ 1 ทางด้านซ้ายใช้แสดงรหัสของ Data Store อาจจะเป็นหมายเลขลำดับหรือตัวอักษรได้เช่น D1, D2 เป็นต้น สำหรับส่วนที่ 2 ทางด้านขวา ใช้แสดงชื่อ Data Store หรือชื่อไฟล์

2.4 วรรณกรรมที่เกี่ยวข้อง

อภิยศ เหรียญวิพัฒน์ (2560) สำนักหอสมุด มหาวิทยาลัยเกษตรศาสตร์ มีแนวคิดในการประยุกต์ใช้ข้อมูลทางสถิติเพื่อเพิ่มประสิทธิภาพในการให้บริการแก่ผู้ใช้ โดยทำการศึกษากับตัวเลขของการให้บริการในด้านต่าง ๆ ที่กระจัดกระจายอยู่เป็นจำนวนมาก มีปริมาณข้อมูลสูง และยังไม่เคยผ่านการวิเคราะห์ทางด้านสถิติมาก่อน โดยการวิเคราะห์ข้อมูลนี้ จะวิเคราะห์ข้อมูลที่มีหลากหลาย ทั้งจากการเข้าใช้งานอาคาร ของสำนักหอสมุด และการใช้บริการพื้นฐานและบริการพิเศษต่าง ๆ โดยใช้วิธีวิเคราะห์ข้อมูล แสดงความสัมพันธ์ และนำมาแสดงผลใน รูปแบบของ Data visualization โดยใช้ภาพ ตัวเลข แผนภูมิ กราฟ และอื่น ๆ ซึ่งช่วยเพิ่มความพึงพอใจแก่ผู้บริหารให้สำมารถรับรู้ข้อมูลที่มากขึ้น และสำมารถนำข้อมูลที่ได้ไปใช้ในการสนับสนุนการตัดสินใจ

วิไลลักษณ์ ตรีพืช (2562) การวิจัยนี้มีวัตถุประสงค์เพื่อการวิเคราะห์ข้อมูลผู้บาดเจ็บและเสียชีวิตจากอุบัติเหตุในช่วงเทศกาลสงกรานต์โดยใช้เทคนิคการทำเหมืองข้อมูล โดยใช้ข้อมูลผู้บาดเจ็บและเสียชีวิตในเทศกาลสงกรานต์ ประจำปี 2551 – 2557 จำแนกตามโรงพยาบาลทั่วประเทศ จากเว็บไซต์ data.go.th ซึ่งผู้วิจัยได้ทำการเตรียมข้อมูลให้เหมาะสมสำหรับการจำแนกข้อมูลโดยตัดข้อมูลที่ไม่เกี่ยวข้องออกไปและปรับให้จำนวนข้อมูลของคำตอบหรือผลลัพธ์มีสัดส่วนใกล้เคียงกัน จากนั้นทำการจำแนกข้อมูลโดยกำหนดให้แอตทริบิวต์ผลการรักษาเป็น คำตอบที่ต้องการพยากรณ์สำหรับอัลกอริทึมที่ผู้วิจัยเลือกใช้ในการสร้าง

โมเดล ได้แก่ Naive Bayes, Generalized Linear Model (GLM), Logistic Regression, Decision Tree, Random Forest, และ Gradient Boosted Trees (XGBoost)

Taamneh (2560) ได้ศึกษาเรื่อง เทคนิคการทำเหมืองข้อมูลสำหรับการสร้างแบบจำลองการเกิดอุบัติเหตุการจราจรและการคาดการณ์ในสหรัฐอาหรับเอมิเรตส์พบวา มีหลายปัจจัยที่นำไปสู่การบาดเจ็บที่รุนแรงรวมถึงปัจจัยที่เกี่ยวข้องกบการขับรถ ปัจจัยทางถนนที่เกี่ยวข้องกับการเกิดอุบัติเหตุและปัจจัยที่เกี่ยวข้องอื่น ๆ โดยใช้ต้นไม้ตัดสินใจ (DT) (J48) ซึ่งผลการศึกษาพบว่าความถูกต้องโดยรวมของตัวจำแนกประเภท DT J48 ตัวจำแนกประเภท PART class และตัวจำแนกประเภท MLP ในการทำนายความรุนแรงของการบาดเจ็บอันเป็นผลมาจากอุบัติเหตุทางรถโดยใช้การมากกวาผู้โดยสารและคนเดินเท้า ผู้ขับขี่ชายมีส่วนร่วมในอุบัติเหตุจราจรมากกวาผู้ขับขี่หญิงมากขึ้น

ศุภามณ จันทร์สกุล (2562) เทคนิคเหมืองข้อมูลในการวิเคราะห์ข้อมูลทางการพยาบาล ปัจจุบันมี ข้อมูลต่าง ๆ และฐานข้อมูลที่เกี่ยวข้องในในวิชาชีพการพยาบาลเป็นจำนวนมากแต่การนำข้อมูลต่าง ๆ และฐานข้อมูลเหล่านั้นมาใช้ประโยชน์โดยค้นหารูปแบบความสัมพันธ์ของข้อมูลซึ่งเรียกว่า “การวิเคราะห์เหมืองข้อมูลยังมีอยู่น้อย วัตถุประสงค์ของบทความวิชาการเพื่อนำเสนอความรู้เบื้องต้น ในการทำเหมืองข้อมูลสำหรับ การวิเคราะห์เหมืองข้อมูลแบ่งออกเป็นสองกลุ่มใหญ่คือ (1) การ วิเคราะห์เหมืองข้อมูลเพื่อการทำนายมีการเรียนรู้แบบมีการสอน และ (2) การวิเคราะห์เหมืองข้อมูล เพื่อการอธิบายมีการเรียนรู้แบบไม่มีการสอน เทคนิคเหมืองข้อมูล 3 เทคนิค ที่ได้รับความนิยมได้แก่การจำแนกประเภท (ต้นไม้ตัดสินใจ) การจัดกลุ่มข้อมูลและการค้นหา กฎ ความสัมพันธ์ของข้อมูล ผลการวิเคราะห์ที่ได้จากเหมืองข้อมูลสร้างองค์ความรู้และเป็นข้อค้นพบที่ เป็นประโยชน์ ในการนำไปใช้เพื่อการตัดสินใจ

2.5 บทสรุป

จากแนวคิด ทฤษฎี เครื่องมือและวรรณกรรมที่เกี่ยวข้องที่ได้กล่าวมาในข้างต้นทั้งหมดนั้น คณะผู้จัดทำได้เลือกใช้ขั้นตอนกระบวนการวิเคราะห์ข้อมูลด้วย CRISP-DM จากเทคนิคการทำเหมืองข้อมูลแบบ Classification ด้วยการสร้างโมเดล Decision Tree เพื่อจัดกลุ่มข้อมูลที่มีความสัมพันธ์กัน โดยใช้โปรแกรมที่ใช้ทำเหมืองข้อมูล คือ โปรแกรม Weka 3.8.4 และโปรแกรม RapidMiner Studio 9.5.1 ในการสร้างโมเดล Decision Tree เพื่อทำการเปรียบเทียบและเลือกผลลัพธ์ที่ดีที่สุดมานำเสนอ จากนั้นนำข้อมูลสารสนเทศมาทำการแสดงผลแบบ visualization ในรูปแบบของภาพโดยใช้โปรแกรม Google Chart เผยแพร่บน web browser ที่เป็นที่นิยมในยุคอินเตอร์เน็ตคือการเผยแพร่ทางสื่อออนไลน์ โดยใช้ภาษา HTML และ CSS ในการเขียนเว็บไซต์ขึ้นมา

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

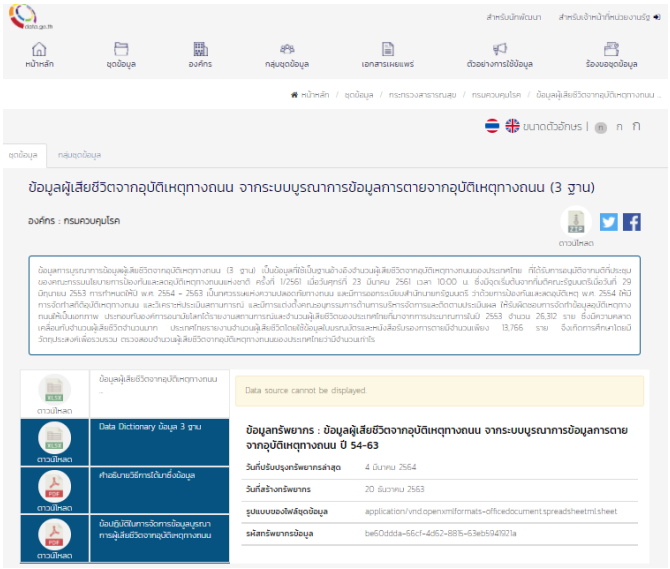
กระบวนการวิเคราะห์ข้อมูล (Cross Industry Standard Process for Data Mining หรือ CRISP. DM) พัฒนาขึ้นในปี ค.ศ. 1936 โดยความร่วมมือของ 3 บริษัทคือ Daimler Chrysler, SPSS และ NCR ซึ่งกระบวนการวิเคราะห์ข้อมูล CRISP–UM. จะเป็นดังรูปแบบในรูปภาพที่ 3.1 แต่ละขั้นตอนในรูปจะเป็นขั้นตอนที่ต่อเนื่องกันนั่นคือ ขั้นตอนถัดไปจะรอผลลัพธ์จากขั้นตอนก่อนหน้าซึ่งแสดงด้วยลูกศรที่เชื่อ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

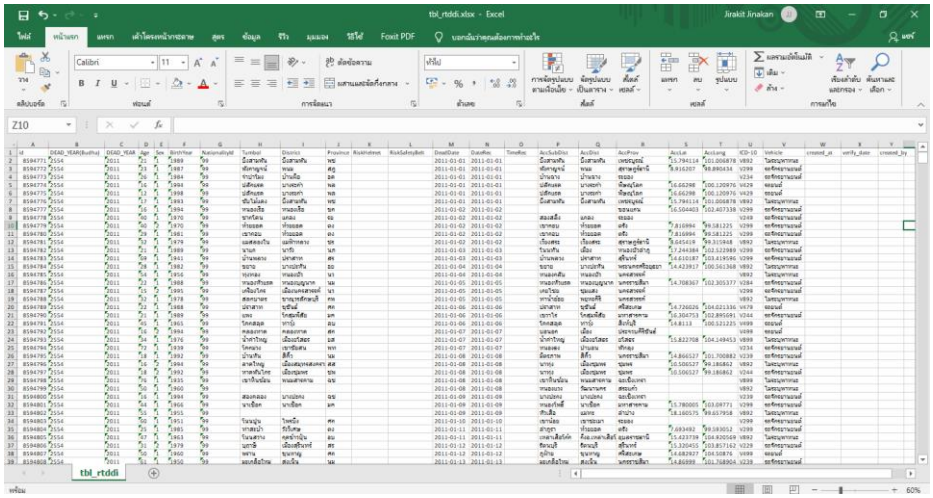
คณะผู้จัดทำทำการรวบรวมข้อมูล เพื่อตรวจสอบรายละเอียด ปริมาณ และความน่าเชื่อถือของข้อมูลผู้เสียชีวิตจากอุบัติเหตุทางถนน ปี 2554 – 2563 ที่ได้จากเว็บไซต์ data.go.th ซึ่งเป็นระบบศูนย์กลางข้อมูลภาครัฐ ที่เผยแพร่สู่สาธารณะอย่างเป็นรูปธรรม เป็นช่องทางให้ผู้ใช้บริการทั้งภาคประชาชน ภาคธุรกิจเอกชน รวมถึงหน่วยงานของรัฐสามารถค้นหาและเข้าถึงข้อมูลที่มีคุณภาพของภาครัฐได้ง่าย



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

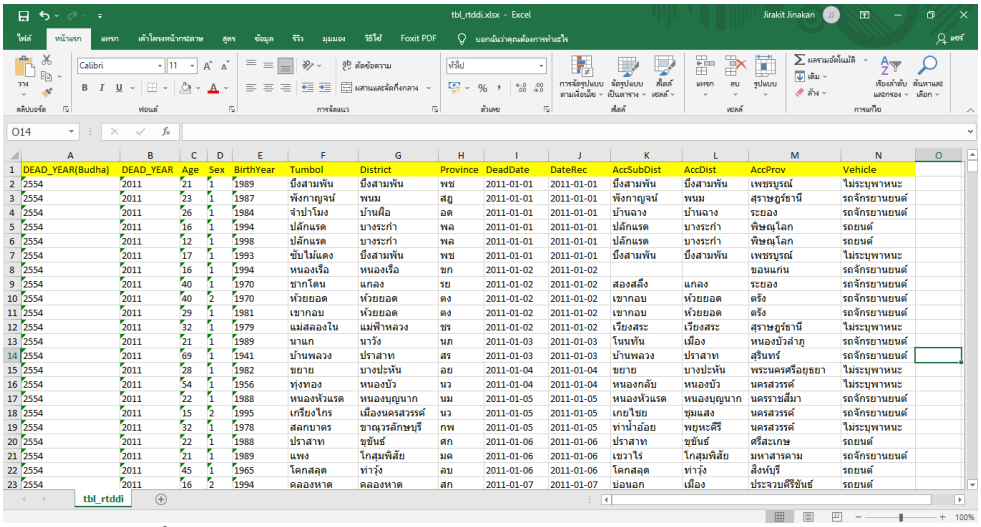


ภาพที่ 3.3 ข้อมูลผู้เสียชีวิตจากอุบัติเหตุทางถนน ปี 2554 – 2563

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	ขาณุวรลักษบุรี	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	เชวไร่	โกสุมพิสัย	มหาสารคาม	รถจักรยานยนต์
2554	2011	45	1	1965	โคกสลุด	ท่าวุ้ง	ลบ	2011-01-06	2011-01-06	โคกสลุด	ท่าวุ้ง	ลพบุรี	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

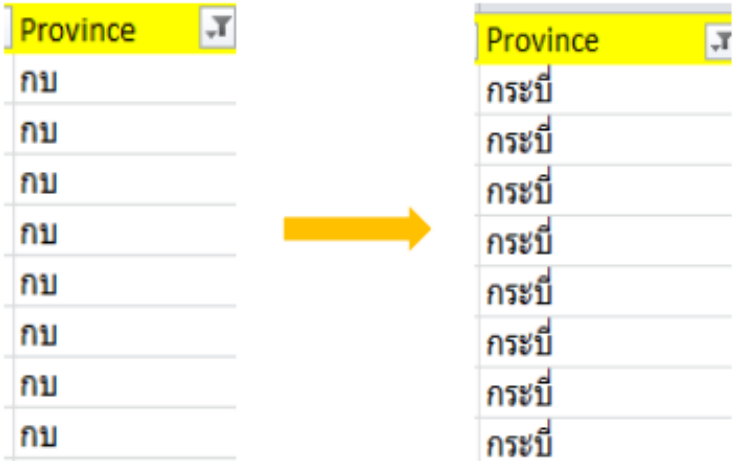
3.1.3.2 ทำการกลั่นกรองข้อมูล (Data Cleaning) คือการทำความสะอาดข้อมูลเป็นกระบวนการตรวจสอบและการแก้ไข (หรือลบ) รายการข้อมูลที่ไม่ถูกต้องออกไปจากชุดข้อมูล ตารางหรือฐานข้อมูล ซึ่งเป็นหลักสำคัญของฐานข้อมูล ทางคณะผู้จัดทำได้ดำเนินการดังนี้

1) ผู้เสียชีวิตจากอุบัติทางถนน ปี 2554 – 2563 คณะผู้จัดทำ ทำการแก้ไขคำผิดหรือชื่อย่อและแก้ไขข้อผิดพลาด ซึ่งคณะผู้จัดทำพบว่ามีข้อมูลบางส่วนมีคำผิดและ

ชื่อย่อของจังหวัด จึงดูข้อมูลตามหลักความเป็นจริงโดยเปรียบเทียบข้อมูลส่วนใหญ่ที่ใช้หรือข้อมูลที่มีคำผิด เนื่องด้วยอาจจะมาจากการคีย์ข้อมูลที่ต้นทาง ดังนั้นคณะผู้จัดทำได้ทำการแก้ไขคำผิดและชื่อย่อของจังหวัด ดังนี้

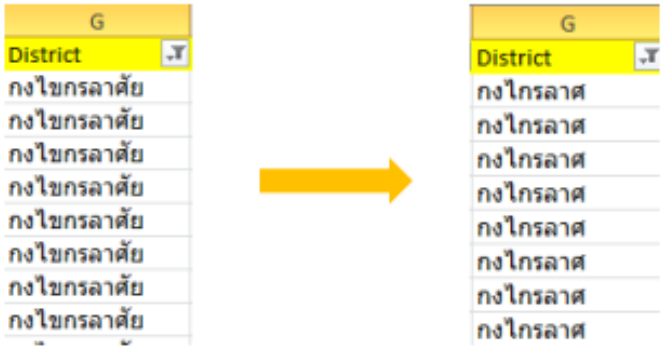
– คณะผู้จัดทำ ทำการแก้ไขข้อมูลคำผิดและชื่อย่อจังหวัดเป็นจำนวนมากไม่สามารถระบุได้ โดยใช้ข้อมูลในปี 2554 – 2563 ที่ทำการ Cleaning แล้ว จำนวน 206,583 รายการ โดยประมาณ

– 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) + ...]

การคำนวณค่าแต่ละแอตทริบิวต์เทียบกับคลาสเพื่อหารแอตทริบิวต์ที่มีค่า 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 กับจำนวนทั้งหมด โดยใช้ผลลัพธ์เป็นดื่มและไมดื่ม 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

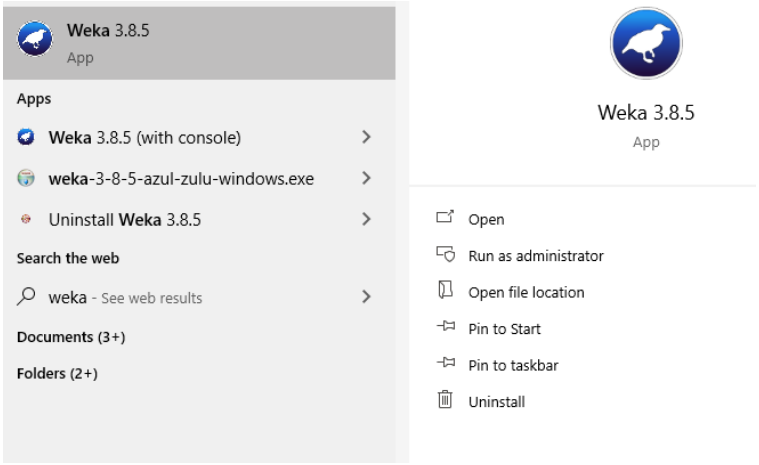
= 0.027

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

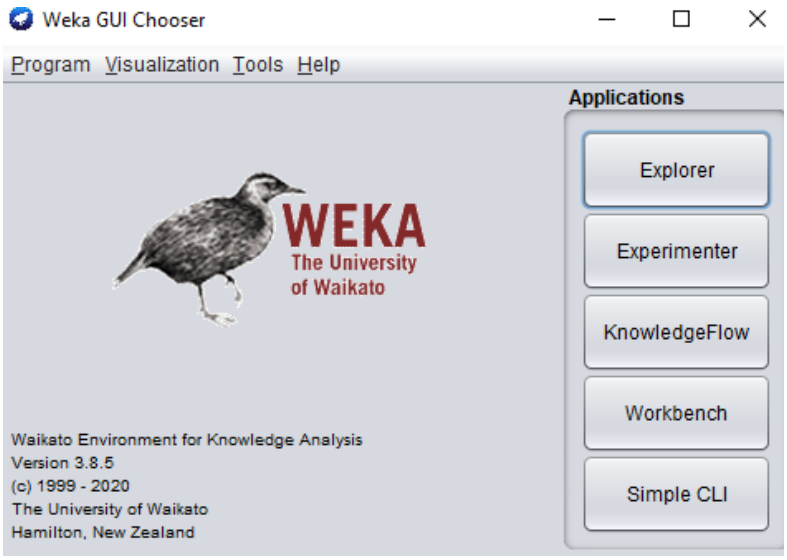
3.1.5 การประเมินผล (Evaluation) เป็นขั้นตอนก่อนนำผลลัพธ์ที่ได้จากขั้นตอนที่ 3 ไปใช้งาน ด้วยการวัดประสิทธิผลของผลลัพท์ที่ได้กับวัตถุประสงค์ที่ตั้งไว้ในขั้นตอนแรก ว่ามีนัยสำคัญหรือความน่าเชื่อถือมากน้อยเพียงใด ด้วยการประเมินผลจากโปรแกรมว่าถูกต้องหรือไม่

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

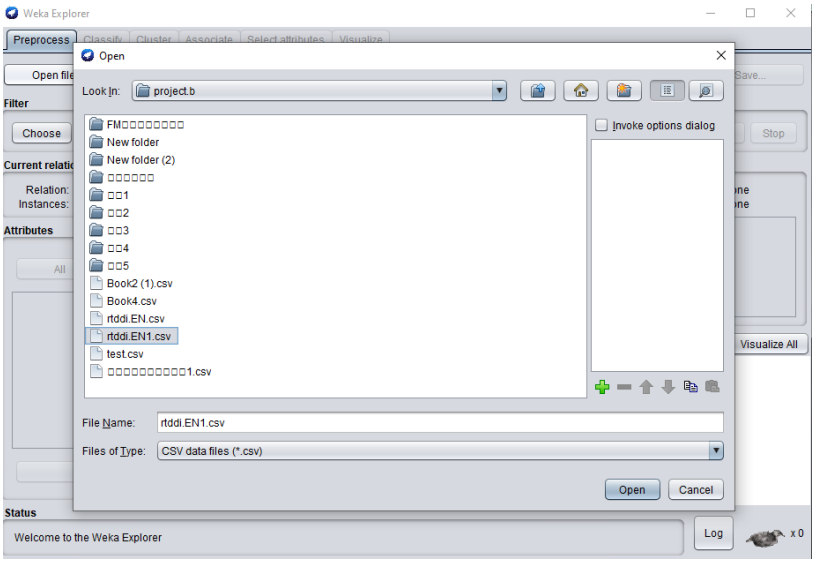


ภาพที่ 3.8 แสดงการเปิดโปรแกรม

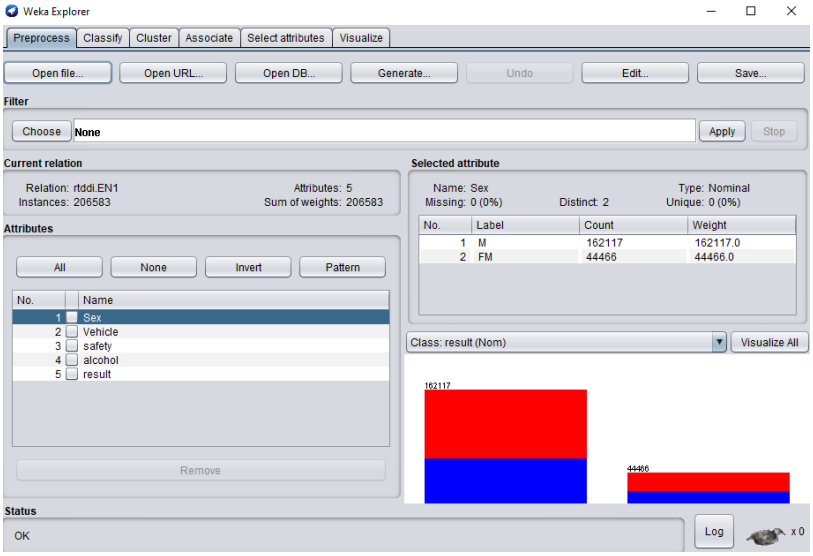


ภาพที่ 3.9 แสดงการเข้าหน้าจอโปรแกรม Weka 3.8.4

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

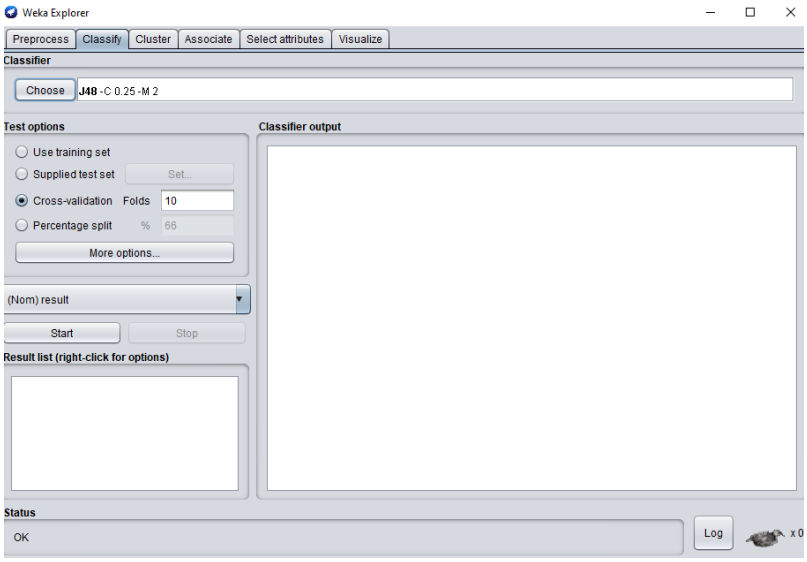


ภาพที่ 3.10 แสดงการนำข้อมูลเข้าในโปรแกรม Weka 3.8.4

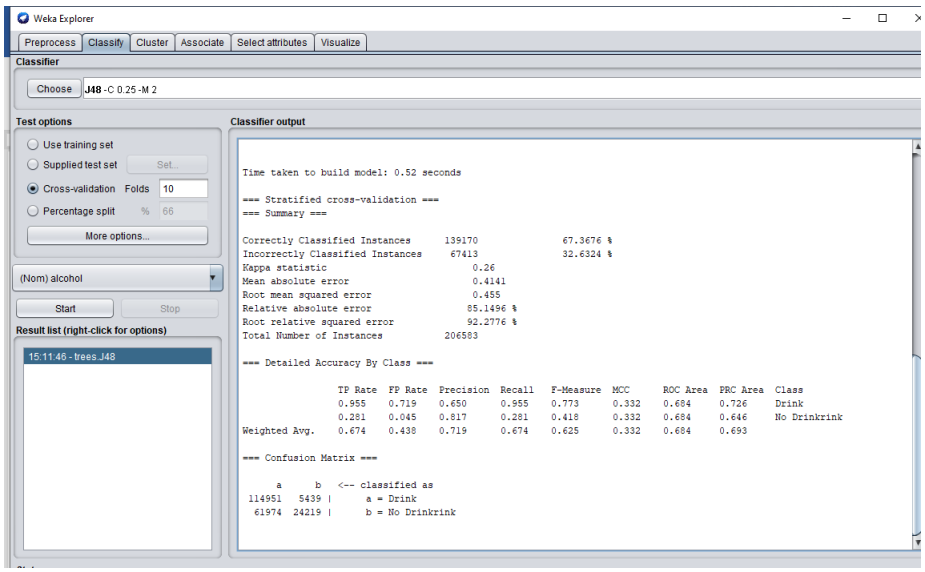


ภาพที่ 3.11 หน้าจอแสดงข้อมูลที่นำเข้าโปรแกรม Weka 3.8.4

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

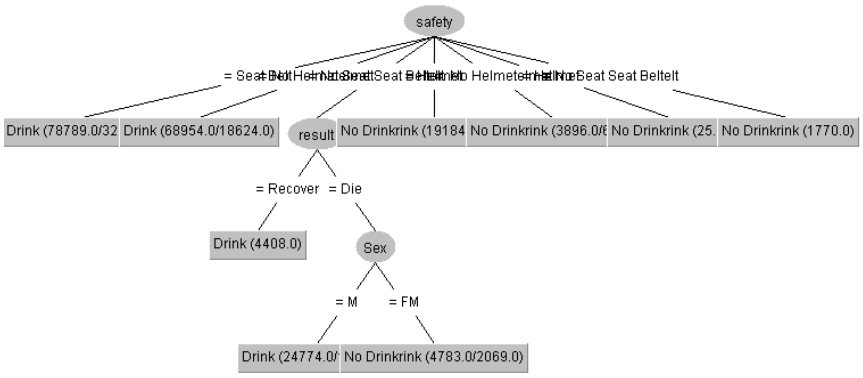


ภาพที่ 3.13 หน้าจอผลลัพธ์ของโมเดลการจำแนกข้อมูลแบบ Decision Tree: J48

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”

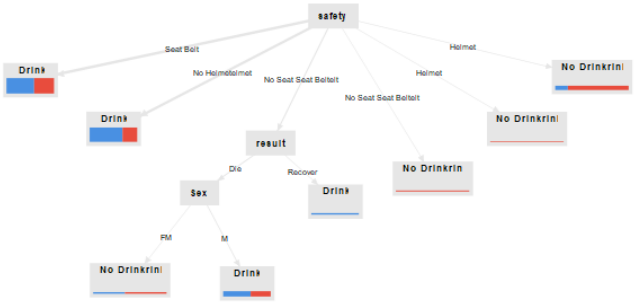
กฎข้อที่ 9 IF safety = No Seat Belt And Result = Die And Sex = FM Then Alcohol = No Drink หมายความว่า ถ้ามาตรการเป็นมาตรการที่ป้องกันกับการไม่คาดเข็มขัดและผลของการรักษากับผลเป็น ตาย และเพศนั้นเป็นเพศหญิง นั้นเกี่ยวเนื่องกัน ผลของการพิจารณาพบว่า ประเภทของการดื่มสุรานั้นจะมีผลลัพธ์เป็น “ไม่ดื่ม”

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
```

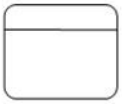
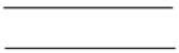
ภาพที่ 3.18 ผลลัพธ์จากการจำแนกกลุ่มแบบ Dacision Tree : J48 ในโปรแกรม Weka

3.2 แผนภาพกระแสข้อมูล Data Flow Diagram

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

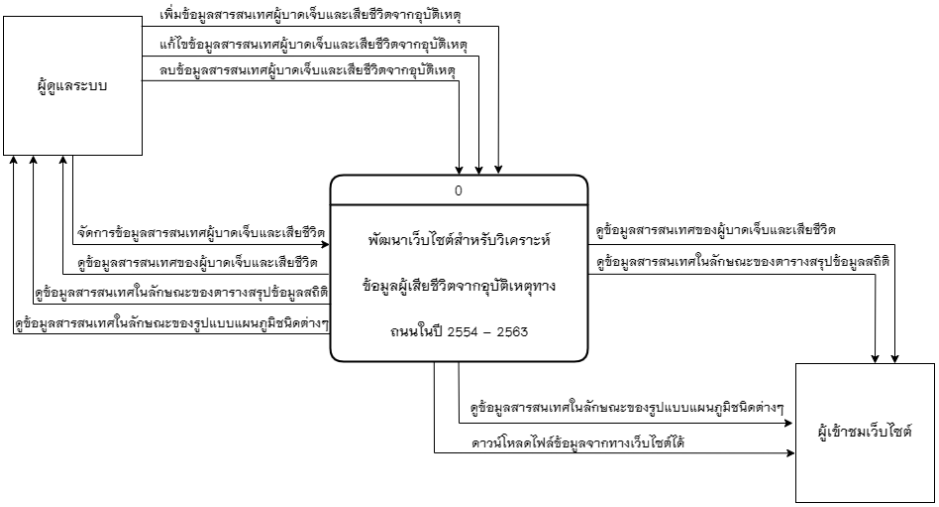
สัญลักษณ์ที่ใช้ในการเขียนแผนภาพกระแสข้อมูลนั้น ประกอบด้วย 4สัญลักษณ์ คือ การประมวลผล (Process) เส้นทางการไหลของข้อมูล (Data Flows) ตัวแทนข้อมูล (External Agents) และแหล่งจัดเก็บข้อมูล (Data Store) โดยได้มีการศึกษาคิดค้นพัฒนาวิธีการอยู่หลายแบบ

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

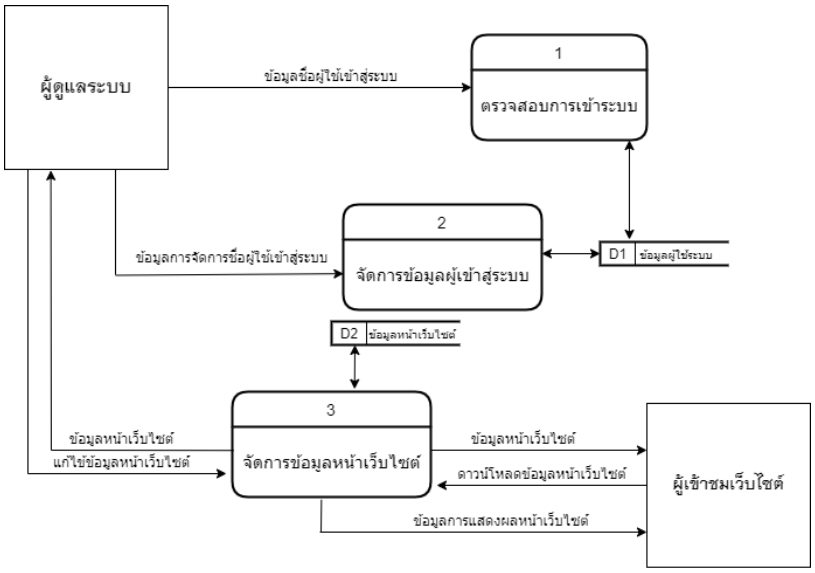


ภาพที่ 3.20 แผนภาพบริบท (Context Diagram)

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

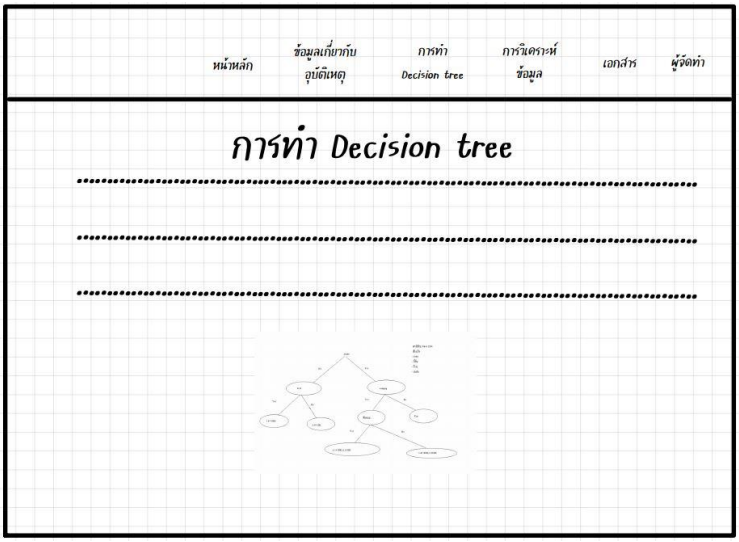
1) หน้าเว็บไซต์แสดงเมนูต่าง ๆบนหน้าเว็บไซต์

ภาพที่ 3.22 Wireframe แสดงหน้าเว็บไซต์แสดงเมนูต่าง ๆ บนหน้าเว็บไซต์

2) หน้าแสดงข้อมูลทั่วไปเกี่ยวกับการเกิดอุบัติเหตุบทท้องถนน

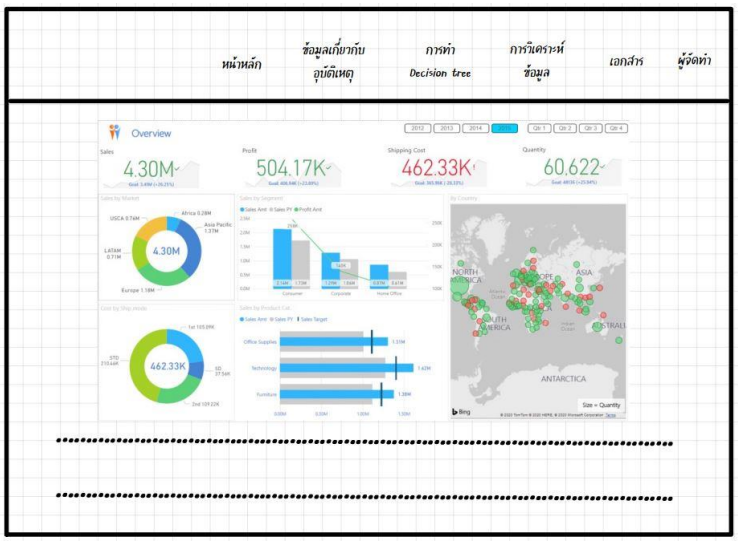
ภาพที่ 3.23 Wireframe ข้อมูลทั่วไปเกี่ยวกับการเกิดอุบัติเหตุบทท้องถนน

3) หน้าแสดงการวิเคราะห์ข้อมูลโดยใช้ Decision Tree ที่จะแสดงขั้นตอนการสร้างโมเดล โดยการใช้โปรแกรม Weka



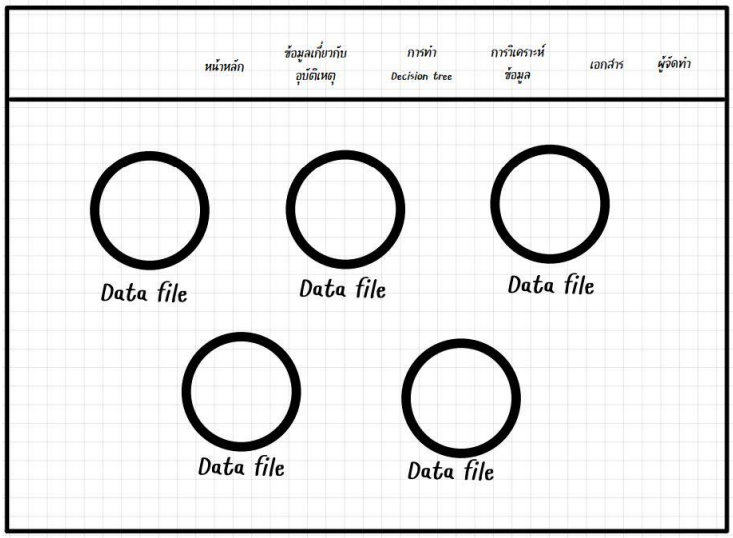
ภาพที่ 3.24 แสดงการวิเคราะห์ข้อมูลโดยใช้ Decision Tree

4) หน้าแสดงผลของข้อมูลในรูปตาราง และแดชบอร์ดผู้ประสบอุบัติเหตุบนท้องถนน



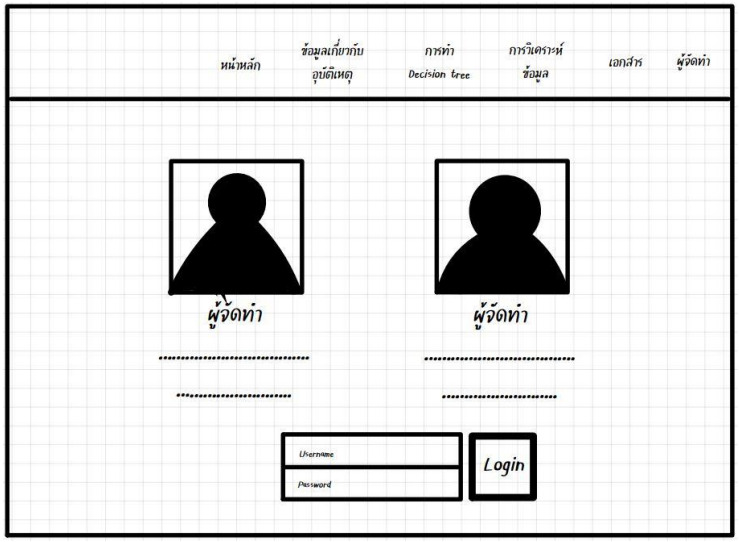
ภาพที่ 3.25 แสดงแดชบอร์ดผู้ประสบอุบัติเหตุบนท้องถนน

5) หน้าแสดง File ข้อมูลต่าง ๆ ที่สามารถดาวน์โหลดได้



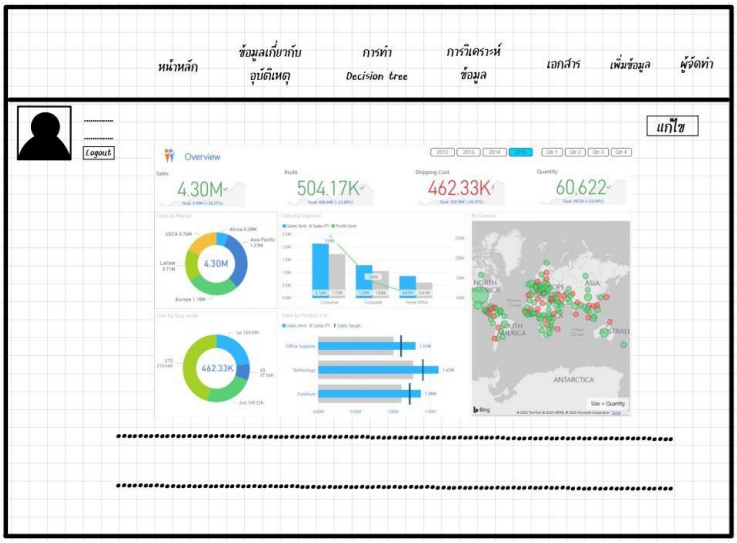
ภาพที่ 3.26 แสดง File ข้อมูลต่าง ๆ ที่สามารถดาวน์โหลดได้

6) หน้าแสดง ข้อมูลส่วนตัวผู้จัดทำ



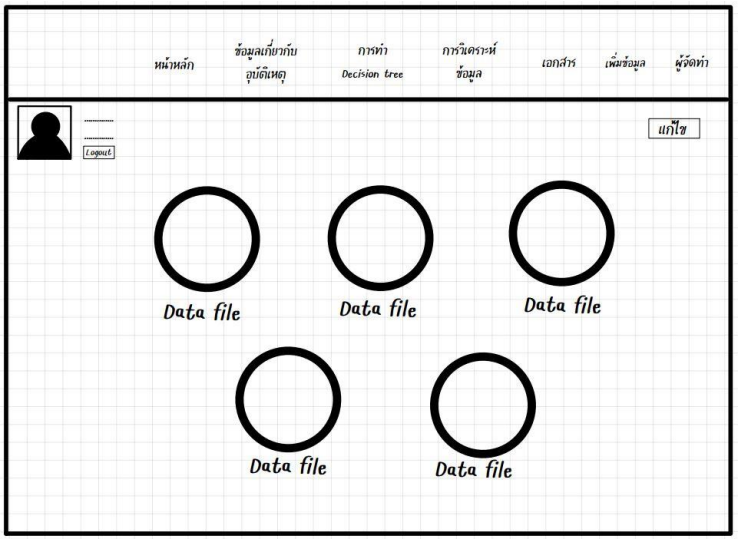
ภาพที่ 3.27 แสดงข้อมูลส่วนตัวผู้จัดทำ

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

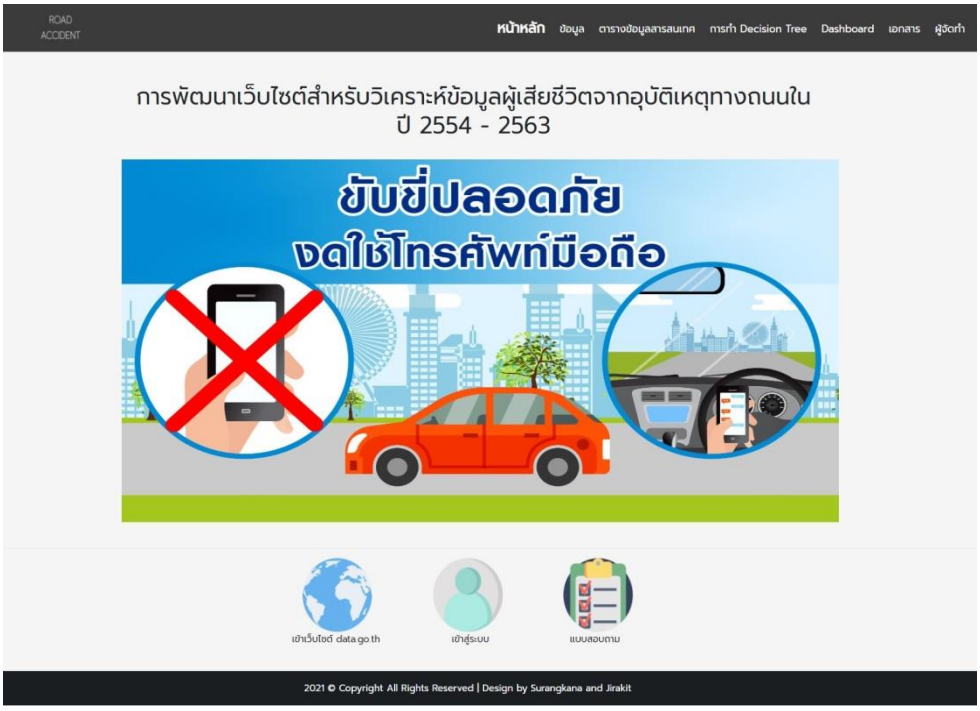
จากวิธีการดำเนินงานโครงงานในข้างต้นทั้งหมดนี้ ทางคณะผู้จัดทำได้ทำการแสดงวิธีในการจัดการกับข้อมูลของผู้เสียชีวิตจากอุบัติเหตุทางถนน ปี 2554 – 2563 ด้วยขั้นตอนการวิเคราะห์ข้อมูลด้วย CRISP-DM อย่างละเอียดรวมไปถึงการสร้างโมเดล Decision Tree จากโปรแกรมที่ใช้เลือกทำเหมืองข้อมูลเพื่อนำเสนอ คือ โปรแกรม Weka 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

บทที่ 4
ผลการดำเนินงาน

จากการวิเคราะห์วิเคราะห์ข้อมูลผู้เสียชีวิตจากอุบัติเหตุทางถนนในปี 2554 – 2563 เพื่อใช้สำหรับเผยแพร่ข้อมูลบนเว็บไซต์ มีวัตถุประสงค์เพื่อพัฒนาเว็บไซต์สำหรับวิเคราะห์ข้อมูลผู้เสียชีวิตจากอุบัติเหตุทางถนนในปี 2554 – 2563 เพื่อให้ได้เว็บไซต์ที่รวมแหล่งข้อมูลความรู้ทั่วไปเกี่ยวกับข้อมูล และแสดงกระบวนการวิเคราะห์ข้อมูล ที่สามารถนำไปใช้งานได้อย่างประสิทธิภาพ โดยผู้จัดทำโครงการได้ทำการวิเคราะห์ข้อมูล และนำเทคโนโลยีสารสนเทศเข้าช่วยในการนำเสนอข้อมูล ทำให้มีความสะดวกรวดเร็วและสามารถใช้งานได้จริงจนสำเร็จลุล่วงตามเป้าหมาย

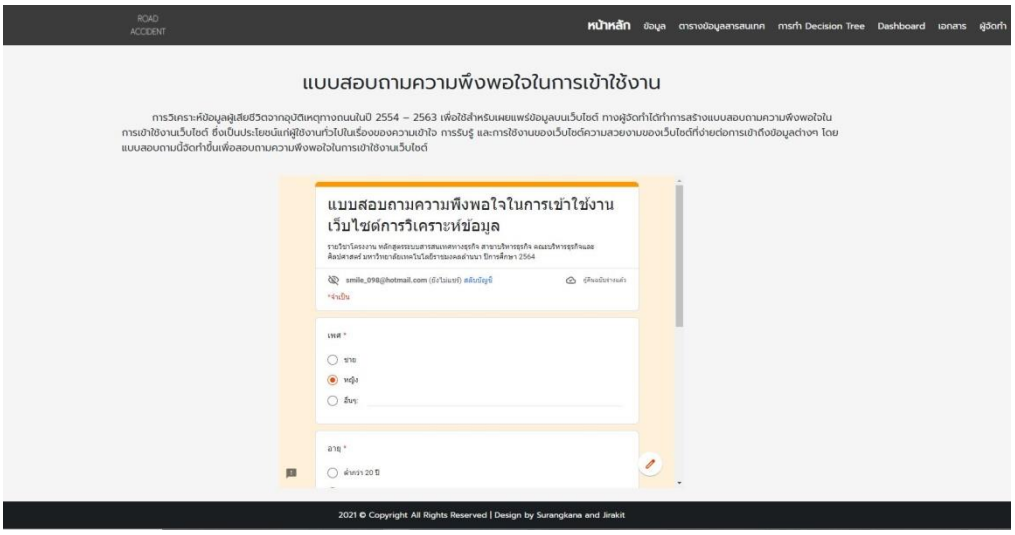
4.1 ผลการดำเนินงาน

การจัดทำโครงงานเรื่อง การพัฒนาเว็บไซต์สำหรับวิเคราะห์ข้อมูลผู้เสียชีวิตจากอุบัติเหตุทางถนนในปี 2554 – 2563 เพื่อพัฒนาเว็บไซต์สำหรับวิเคราะห์ข้อมูลผู้เสียชีวิตจากอุบัติเหตุทางถนนในปี 2554 – 2563 **ใน**ครั้งนี้พบว่าการจำแนกข้อมูลออกเป็นประเภทต่าง ๆ ช่วยให้เกิดความเข้าใจต่อข้อมูลและสิ่งที่ต้องการทราบอย่างแท้จริง คณะผู้จัดทำได้ทำการวิเคราะห์ข้อมูลในรูปแบบของ Data mining โดยทำการจำแนกข้อมูลเป็นประเภทต่าง ๆ ด้วยเทคนิคการทำเหมืองข้อมูลแบบ Classification จากเทคนิคการจำแนกข้อมูลด้วยการวิเคราะห์ในรูปแบบของแผนภูมิต้นไม้ หรือเรียกว่า Decision Tree และการนำเสนอสารสนเทศแบบ Visualization เผยแพร่ข้อมูลสารสนเทศขึ้นบน Web Browser ใช้กับผู้ใช้งาน ซึ่งจะทำให้เข้าใจได้อย่างรวดเร็ว เพื่อเพิ่มประสิทธิภาพให้กับการศึกษาค้นคว้าและทันต่อเวลา คณะผู้จัดทำจึงได้ทำการวิเคราะห์ข้อมูล และจัดทำเว็บไซต์ตามวัตถุประสงค์ที่กำหนดไว้และมีผลการดำเนินงาน ดังนี้



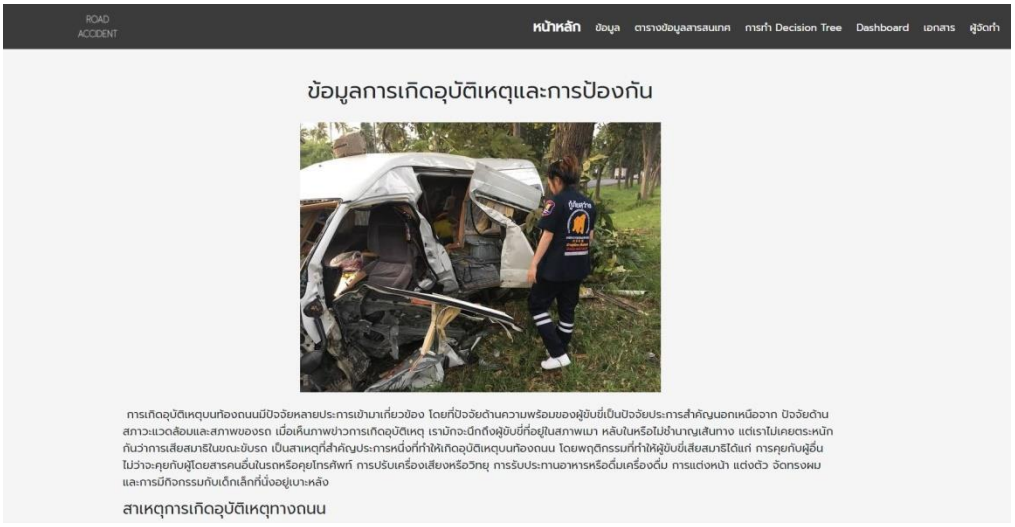
ภาพที่ 4.1 หน้าแรกของเว็บไซต์ เมื่อเข้าสู่เว็บไซต์

คำอธิบายใช้งาน : แถบเมนูสามารถใช้งานได้ตามความต้องการของการใช้งาน ซึ่งเป็นหน้าสำหรับผู้เข้าชมทั่วไป สามารถเลือกดู และศึกษาเข้าไปใช้งานในส่วนของเว็บไซต์ได้



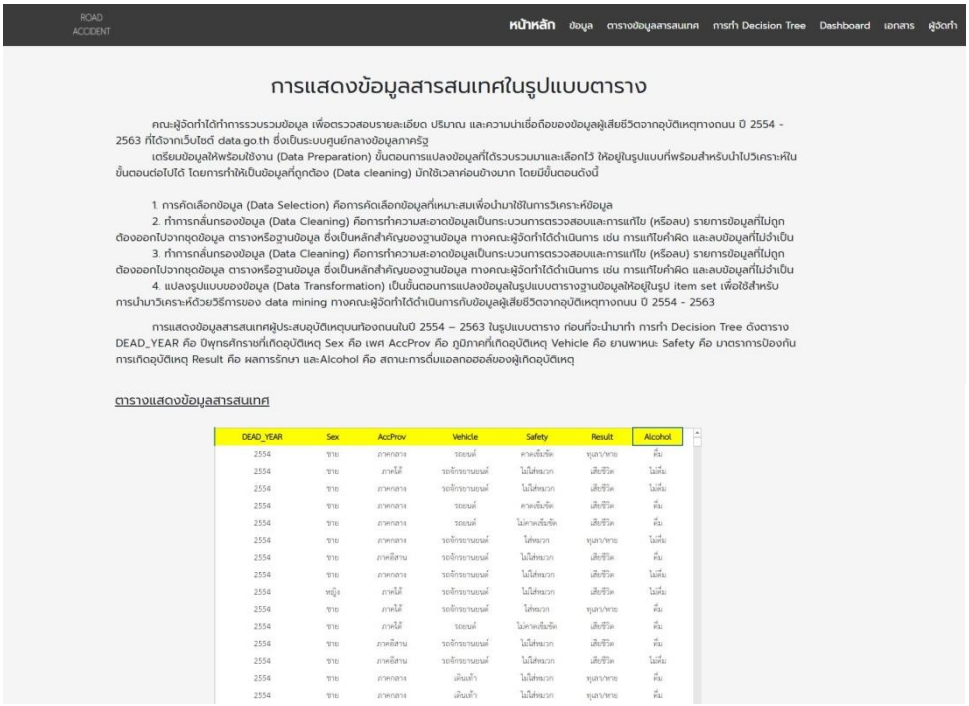
ภาพที่ 4.2 หน้าแสดงแบบสอบถามความพึงพอใจ

คำอธิบายใช้งาน : เป็นหน้าแสดงการทำแบบสอบถามความพึงพอใจของผู้เข้าชมเว็บไซต์ของผู้ใช้งานสามารถทำแบบสอบถามตามที่พึงพอใจได้



ภาพที่ 4.3 แสดงหน้าข้อมูลการเกิดอุบัติเหตุและการป้องกันการเกิดอุบัติเหตุ

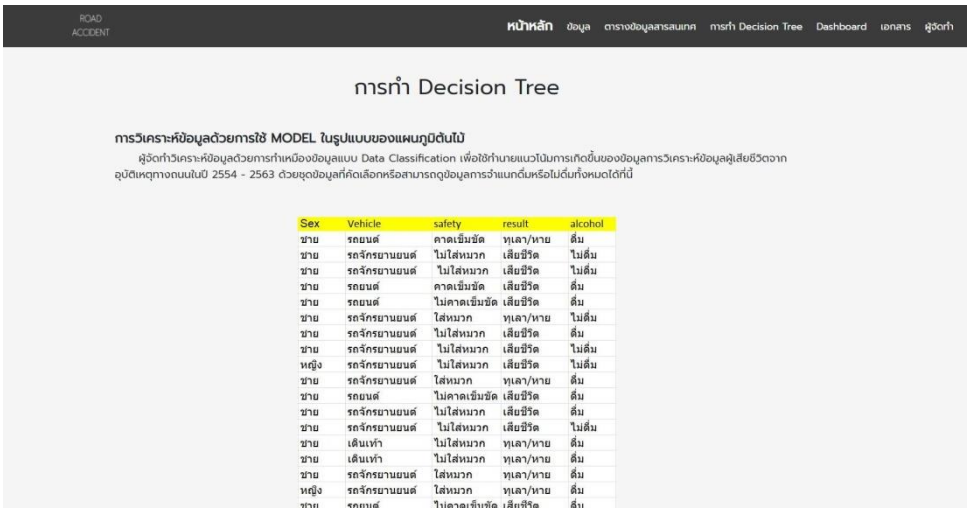
คำอธิบายใช้งาน : เป็นหน้าแสดงข้อมูลเกี่ยวกับการเกิดอุบัติเหตุและการป้องกันการเกิดอุบัติเหตุบนท้องถนน



ภาพที่ 4.4 แสดงข้อมูลสารสนเทศในรูปแบบตาราง

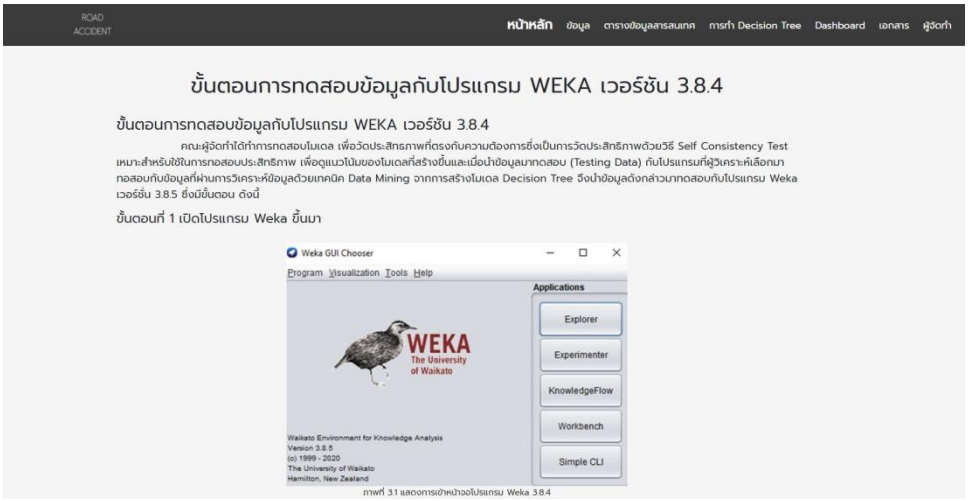
คำอธิบายใช้งาน : การแสดงข้อมูลสารสนเทศผู้ประสบอุบัติเหตุบนท้องถนนในปี 2554 – 2563 ในรูปแบบตาราง ก่อนที่จะนำมาทำ การทำ Decision Tree ดังตาราง DEAD_YEAR คือ ปีพุทธศักราชที่เกิดอุบัติเหตุ Sex คือ เพศ AccProv คือ ภูมิภาคที่เกิดอุบัติเหตุ

Vehicle คือ ยานพาหนะ Safety คือ มาตรการป้องกันการเกิดอุบัติเหตุ Result คือ ผลการรักษา และAlcohol คือ สถานการณ์ดื่มแอลกอฮอล์ของผู้เกิดอุบัติเหตุ



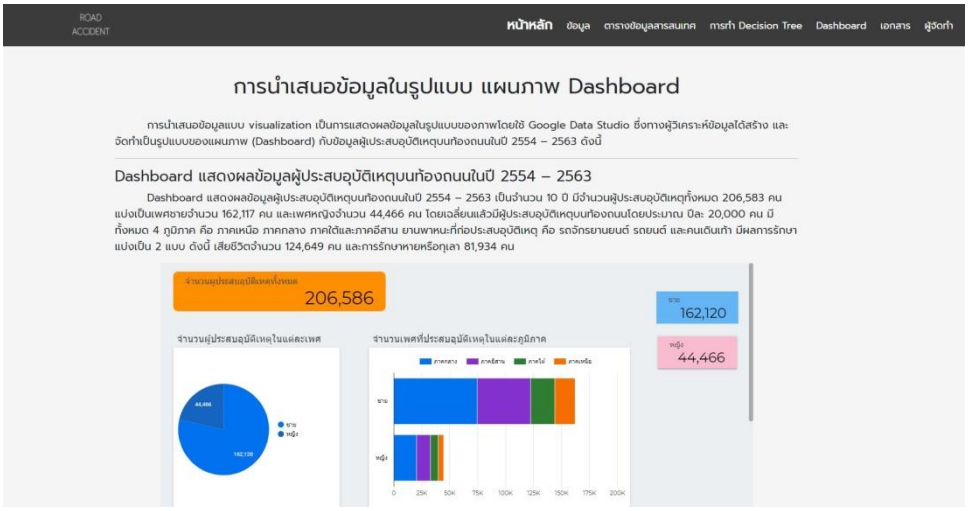
ภาพที่ 4.5 แสดงหน้าการทำ Decision Tree

คำอธิบายใช้งาน : เป็นหน้าแสดงการทำ Decision Tree และการเปรียบเทียบโมเดลที่ทำการทดลองด้วยโปรแกรม Weka 3.8.4 และโปรแกรม RapidMiner Studio 9.5.1 ซึ่งมีค่าที่ตรงกัน



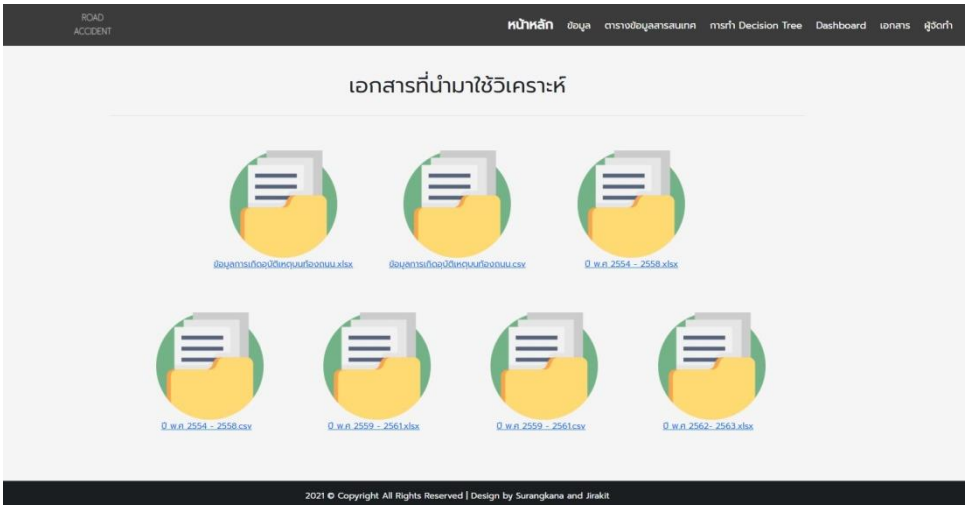
ภาพที่ 4.6 แสดงขั้นตอนการทดสอบข้อมูลกับโปรแกรม WEKA

คำอธิบายใช้งาน : เป็นหน้าแสดงขั้นตอนการทดสอบข้อมูลโดยโปรแกรม WEKA เวอร์ชั่น 3.8.4 เพื่อวัดประสิทธิภาพที่ตรงกับความต้องการซึ่งเป็นการวัดประสิทธิภาพด้วยวิธี Self Consistency Test เหมาะสำหรับใช้ในการทอสอบประสิทธิภาพ



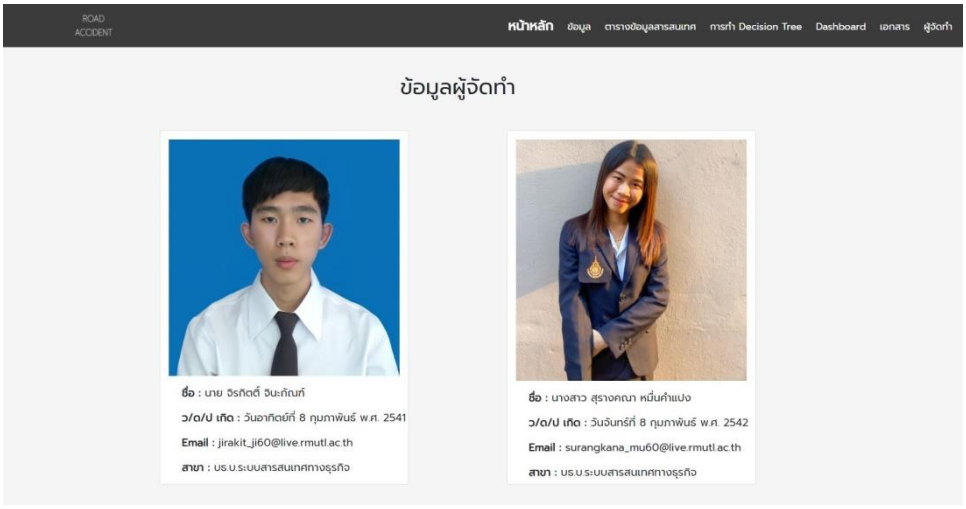
ภาพที่ 4.7 แสดงข้อมูลในรูปแบบ แผนภาพ Dashboard

คำอธิบายใช้งาน : เป็นหน้าแสดงที่การนำเสนอข้อมูลแบบ visualization เป็นการแสดงผลข้อมูลในรูปแบบของภาพโดยใช้ Google Data Studio ซึ่งทางผู้วิเคราะห์ข้อมูลได้สร้างและจัดทำเป็นรูปแบบของแผนภาพ (Dashboard)



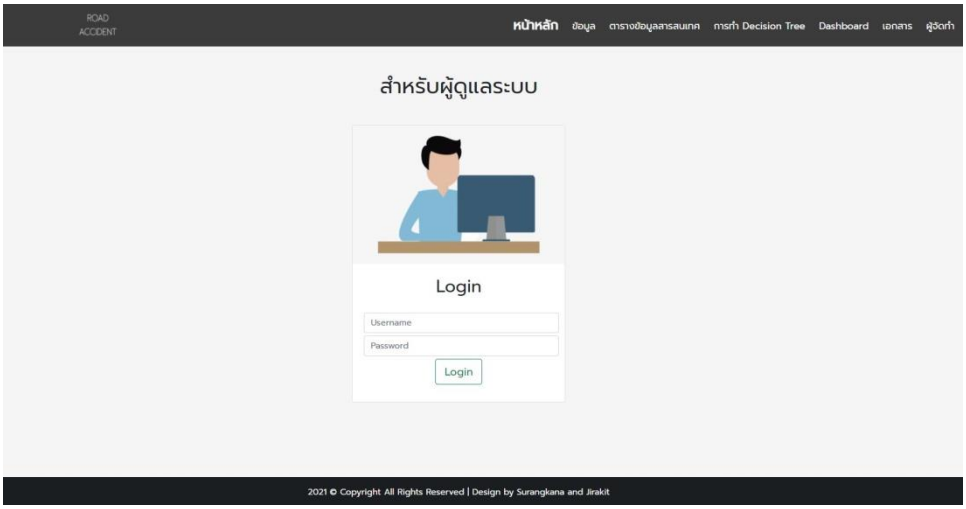
ภาพที่ 4.8 แสดงเอกสารที่นำมาใช้วิเคราะห์

คำอธิบายใช้งาน : เป็นหน้าแสดงเอกสารที่นำมาใช้ในการวิเคราะห์และผู้เข้าชมทั่วไปสามารถดาวน์โหลดไปใช้ได้



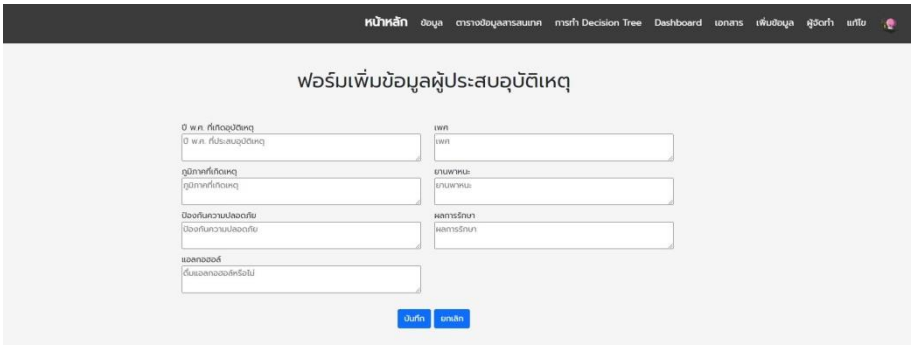
ภาพที่ 4.9 แสดงข้อมูลผู้จัดทำ

คำอธิบายใช้งาน : เป็นหน้าแสดงข้อมูลผู้จัดทำและอาจารย์ที่ปรึกษาโครงงาน



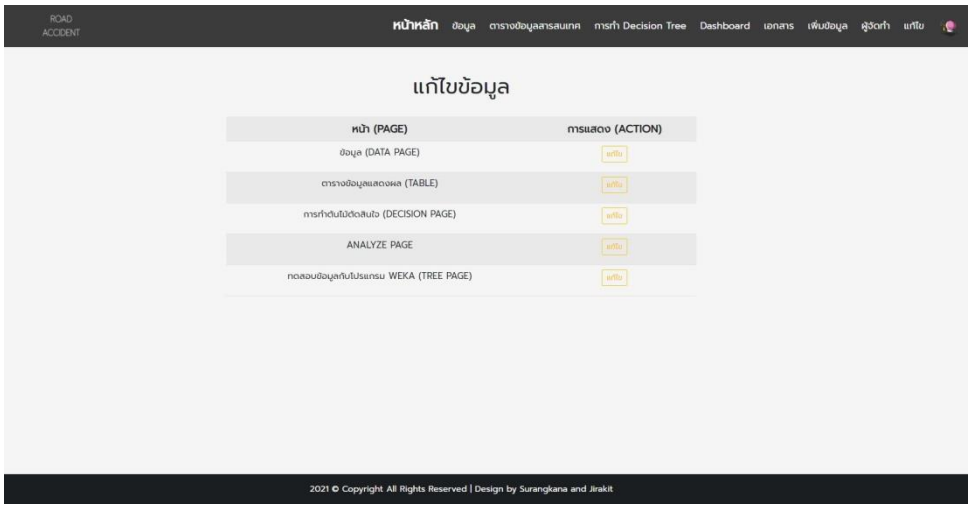
ภาพที่ 4.10 แสดงหน้าเข้าสู่ระบบสำหรับแอดมิน

คำอธิบายใช้งาน : เป็นหน้าแสดงการเข้าสู่ระบบเพื่อทำการแก้ไข ลบ ข้อมูลภายในเว็บสำหรับแอดมินเท่านั้น



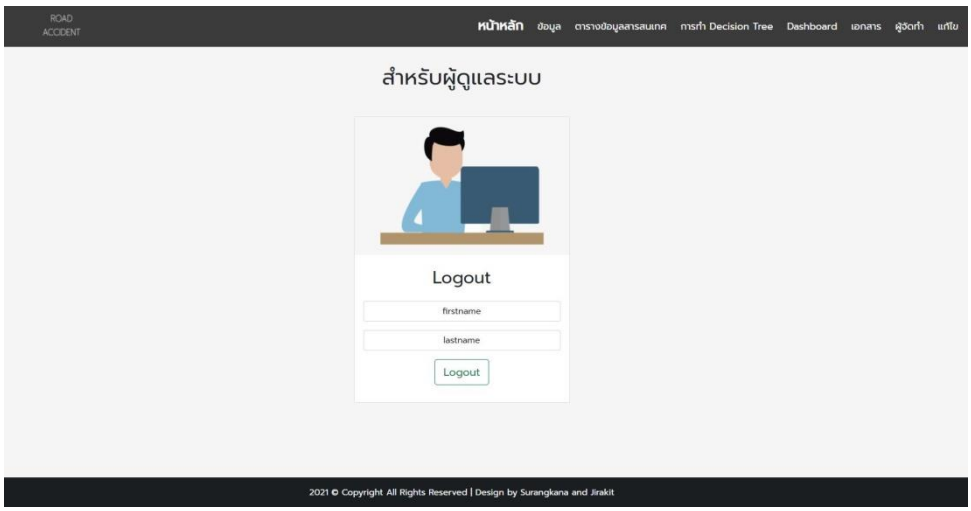
ภาพที่ 4.11 แสดงหน้าแบบฟอร์มเพิ่มข้อมูล

คำอธิบายใช้งาน : เป็นหน้าแสดงแบบฟอร์มเพิ่มข้อมูลไปยัง Google Sheet สำหรับหรับแอดมินหรือผู้ดูแลเว็บไซต์เท่านั้น



ภาพที่ 4.12 แสดงหน้าการเลือกแก้ไขข้อมูล

คำอธิบายใช้งาน : เป็นหน้าแสดงการเลือกแก้ไขข้อมูลในแต่ละหน้าของเว็บไซต์ผู้ที่จะสามารถแก้ไขข้อมูลได้จะต้องเป็นแอดมินเท่านั้น



ภาพที่ 4.13 แสดงหน้าออกจากระบบสำหรับแอดมิน

คำอธิบายใช้งาน : เป็นหน้าแสดงการออกจากระบบหลังจากที่ทำการแก้ไข ลบข้อมูลภายในเว็บเสร็จแล้วสำหรับแอดมินเท่านั้น

4.2 การอภิปรายผล

จากวัตถุประสงค์เพื่อพัฒนาเว็บไซต์สำหรับวิเคราะห์ข้อมูลผู้เสียชีวิตจากอุบัติเหตุทางถนนในปี 2554 – 2563 คณะผู้จัดทำได้ศึกษาปัญหา และเก็บรวบรวมข้อมูล ซึ่งคณะผู้จัดทำทำการวิเคราะห์ข้อมูลอยู่ในรูปแบบ Data Mining โดยใช้กระบวนการวิเคราะห์ข้อมูล (Cross-industry standard process for data mining หรือ CRISP-DM) เพื่อใช้ในการวิเคราะห์ข้อมูลตามขอบเขตของโครงงาน ดังนี้

ขอบเขตผู้วิเคราะห์ข้อมูล

4.2.1 กระบวนการวิเคราะห์ข้อมูลด้วย CRISP-DM

1. เข้าใจ (Business Understanding) ปัญหาในรูปแบบของการวิเคราะห์ข้อมูลทาง Data Mining โดยการวิเคราะห์ข้อมูลในประเด็น ข้อมูลผู้เสียชีวิตจากอุบัติเหตุทางถนน

2. เข้าใจข้อมูล (Data Understanding) และทำการรวบรวมข้อมูลเพื่อตรวจสอบรายละเอียดปริมาณและความน่าเชื่อถือของข้อมูลผู้ที่ได้รับบาดเจ็บและผู้เสียชีวิตจากอุบัติเหตุทางถนน ปี 2554 – 2563 จากเว็บไซต์ data.go.th

3. เตรียมข้อมูล (Data Preparation) และทำการคัดเลือกข้อมูลและทำการ Data Cleaning ข้อมูลเพื่อให้ได้ข้อมูลที่สมบูรณ์ ด้วยการตัดข้อมูลในส่วนที่ไม่จำเป็นออกไป เพื่อเหลือข้อมูลที่จำเป็นในการนำไปวิเคราะห์ข้อมูล

4. สร้างแบบจำลอง (Modeling) ผู้จัดทำการวิเคราะห์ข้อมูลด้วยการทำเหมืองข้อมูลแบบ Data Classification เพื่อใช้ในการทำนายแนวโน้มการเกิดขึ้นของปัจจัยที่ก่อให้เกิดผู้เสียชีวิตและผู้ที่ได้รับบาดเจ็บจากเทคนิคการจำแนกประเภทข้อมูล ด้วยการสร้างโมเดล Decision Tree เพื่อจัดกลุ่มข้อมูลที่มีความสัมพันธ์กัน โดยใช้โปรแกรมที่ใช้ทำเหมืองข้อมูล

5. การประเมินผล (Evaluation) ผู้จัดทำจะใช้เทคนิคของการจำแนกกลุ่มแบบ 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 หมายความว่าถ้ามาตรการเป็นมาตรการที่ป้องกันกับการไม่ใส่หมวกเกี่ยวเนื่องกัน 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 หมายความว่าถ้ามาตรการเป็นมาตรการที่ป้องกันกับการไม่คาดเข็มขัดเกี่ยวเนื่องกัน ผลของการพิจารณาพบว่า ประเภทของการดื่มสุรานั้นจะมีผลลัพธ์เป็น “ไม่ดื่ม”
- กฎข้อที่ 7 IF safety = No Seat Belt And Result = Recover Then Alcohol = Drinkหมายความว่า ถ้ามาตรการเป็นมาตรการที่ป้องกันกับการไม่คาดเข็มขัดและผลของการ

รักษากับผลเป็น 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”

6. การนำไปใช้ (Deployment) ผู้จัดทำนำผลของข้อมูลที่ได้รับจากการวิเคราะห์แล้วมาแสดงผลข้อมูลผ่านทาง Web Browser โดยนำเสนอข้อมูลแบบ Visualization เป็นการแสดงผลข้อมูลในรูปแบบของภาพโดยใช้ Google Data Studio เป็นรูปแบบของรายงาน (Report) หรือแผนภาพ (Dashboard) โดยคณะผู้จัดทำได้ทำการศึกษาความรู้พื้นฐานทางด้านภาษาที่ใช้ในการพัฒนาเว็บไซต์ คือ ใช้ชุดคำสั่ง HTML CSS3 และ JavaScript เมื่อศึกษาข้อมูลความรู้พื้นฐานในภาษาต่าง ๆ และการวิเคราะห์ข้อมูลด้วยเทคนิคดังที่กล่าวมาข้างต้นแล้วจึงทำการวิเคราะห์ข้อมูล และออกแบบเว็บไซต์ โดยเว็บไซต์มีส่วนประกอบต่าง ๆ ดังนี้

1. ส่วนเนื้อหาเกี่ยวกับการเกิดอุบัติเหตุและการป้องกันการเกิดอุบัติเหตุ
2. ส่วนของการทำ Decision Tree การเปรียบเทียบโมเดล
3. ส่วนของขั้นตอนการทดสอบข้อมูลกับโปรแกรม WEKA เวอร์ชั่น 3.8.4
4. ส่วนการนำเสนอข้อมูลในรูปแบบ แผนภาพ Dashboard
5. ส่วนเอกสาร
6. ส่วนของแบบสอบถาม

4.2.2 ขอบเขตผู้ใช้งานทั่วไปบนเว็บไซต์

1) ผู้ใช้งานทั่วไปสามารถดูข้อมูลสารสนเทศของผู้บาดเจ็บและเสียชีวิตบนท้องถนนในปี 2554 – 2563 และแหล่งความรู้ของข้อมูลทั่วไปที่เกี่ยวกับอุบัติเหตุและการวิเคราะห์ข้อมูลเผยแพร่บนเว็บไซต์

2) ผู้ใช้งานทั่วไปสามารถดูข้อมูลสารสนเทศในลักษณะของตารางของผู้บาดเจ็บและเสียชีวิตบนท้องถนนในปี 2554 – 2563 ในรูปแบบแผนภาพ (Dashboard) จาก Google Data Studio ในลักษณะรูปแบบของแผนภูมิชนิดต่าง ๆ

4.2.3 การอธิปรายผลความพึงพอใจจากการตอบแบบสอบถาม

จากผลการพัฒนาเว็บไซต์สำหรับวิเคราะห์ข้อมูลผู้เสียชีวิตจากอุบัติเหตุทางถนนปี 2554 – 2563 แล้วจากนั้นทางผู้วิเคราะห์ข้อมูลได้ทำการประเมินผลการใช้งานเว็บไซต์จากผู้ใช้งานทั่วไป จากกลุ่มตัวอย่างจำนวน 61 คน ผู้วิเคราะห์ได้แบ่งการประเมินเป็น 3 ตอน ซึ่งมีรายละเอียดดังนี้

1) ข้อมูลทั่วไปของผู้ใช้งานเว็บไซต์

2) ข้อมูลเกี่ยวกับความพึงพอใจในการใช้บริการเว็บไซต์

3) ข้อเสนอแนะ

ในตอนที่ 1 ข้อมูลทั่วไปของผู้ใช้งานเว็บไซต์ ประกอบด้วย

ส่วนข้อมูลทั่วไปจากแบบสำรวจ ประกอบด้วยข้อมูลทั่วไป ได้แก่ เพศ อายุ ระดับการศึกษาโดยสามารถสรุปข้อมูล จากกลุ่มตัวอย่าง 61 คน แบ่งเป็น

เพศ

ชาย	จำนวน 24 คน	คิดเป็นร้อยละ 39.3
หญิง	จำนวน 37 คน	คิดเป็นร้อยละ 60.7

อายุ

ต่ำกว่า 20 ปี	จำนวน 12 คน	คิดเป็นร้อยละ 19.7
20–30 ปี	จำนวน 43 คน	คิดเป็นร้อยละ 70.5
31–40 ปี	จำนวน 6 คน	คิดเป็นร้อยละ 9.8
41–50 ปี	จำนวน 0 คน	คิดเป็นร้อยละ 0
51–60 ปี	จำนวน 0 คน	คิดเป็นร้อยละ 0
มากกว่า 60 ปี	จำนวน 0 คน	คิดเป็นร้อยละ 0

ระดับการศึกษา

ต่ำกว่าปริญญาตรี	จำนวน 39 คน	คิดเป็นร้อยละ 27.9
ปริญญาตรี	จำนวน 24 คน	คิดเป็นร้อยละ 68.9

ปริญญาโท	จำนวน 2 คน	คิดเป็นร้อยละ 3.3
สูงกว่าปริญญาโท	จำนวน 0 คน	คิดเป็นร้อยละ 0

ในตอนที่ 2 ข้อมูลเกี่ยวกับความพึงพอใจในการใช้บริการเว็บไซต์มากน้อยเพียงใด

1) ด้านการออกแบบและการจัดรูปแบบ (Design)

2) ด้านคุณภาพข้อมูล (Information Quality)

3) ด้านประโยชน์และการนำไปใช้ (Perceived Usefulness)

การประเมินความพึงพอใจของผู้ตอบแบบสอบถามในภาพรวมด้วยค่าเฉลี่ยคะแนนความพึงพอใจ โดยมีคิดคะแนนและเกณฑ์ระดับความพึงพอใจเป็นดังนี้

ระดับความพึงพอใจ	น้อยที่สุด	มีค่าคะแนน 1
ระดับความพึงพอใจ	น้อย	มีค่าคะแนน 2
ระดับความพึงพอใจ	ปานกลาง	มีค่าคะแนน 3
ระดับความพึงพอใจ	มาก	มีค่าคะแนน 4
ระดับความพึงพอใจ	มากที่สุด	มีค่าคะแนน 5

เมื่อนำคำตอบของผู้ตอบแบบสอบถามมาแจกแจงความถี่และหาค่าเฉลี่ย แล้วกำหนดระดับค่าเฉลี่ยความพึงพอใจเป็นดังนี้

(สรุปแบบประเมินความพึงพอใจของผู้รับบริการศูนย์ดำรงธรรมจังหวัดราชบุรี หน้า 4)

จากเกณฑ์ดังกล่าว สามารถแปลความหมายของความพึงพอใจได้ดังนี้

คะแนนค่าเฉลี่ย 1.00 – 1.49 หมายถึง ระดับความพึงพอใจน้อยที่สุด

คะแนนค่าเฉลี่ย 1.50 – 2.49 หมายถึง ระดับความพึงพอใจน้อย

คะแนนค่าเฉลี่ย 2.50 – 3.49 หมายถึง ระดับความพึงพอใจปานกลาง

คะแนนค่าเฉลี่ย 3.50 – 4.49 หมายถึง ระดับความพึงพอใจมาก

คะแนนค่าเฉลี่ย 4.50 – 5.00 หมายถึง ระดับความพึงพอใจมากที่สุด

ดังนั้นผู้วิเคราะห์จึงนำค่าคะแนนมาคำนวณทางสถิติ คือ ค่าคะแนนเฉลี่ย $(\bar{x})$ และค่าส่วนเบี่ยงเบนมาตรฐาน (S.D.) ผลการประเมินแสดงรายละเอียดดังนี้ การประเมินด้านด้านการออกแบบ ดังตารางที่ 4.1 การประเมินด้านการออกแบบและการจัดรูปแบบเว็บไซต์แสดงดังตารางที่ 4.2 และการประเมินด้านประโยชน์และการนำไปใช้แสดงดังตารางที่ 4.3

ตารางที่ 4. 1 แสดงข้อมูลการประมวลผลด้านการออกแบบ (Design)

รายการประเมิน	**จำนวนผู้ตอบแบบสอบถาม**	**($\bar{x}$)**	**S.D.**	**ผลการประเมิน**
1. ความสวยงาม ความทันสมัย น่าสนใจของหน้าโฮมเพจ	61	3.79	0.86	ดี
2. การจัดรูปแบบในเว็บไซต์ง่ายต่อการอ่านและการใช้งาน	61	4.00	0.75	ดี
3. สีสันในการออกแบบเว็บไซต์มีความเหมาะสม	61	4.03	0.71	ดี
4. เมนูง่ายต่อการใช้งาน	61	3.87	0.85	ดี
5. สีพื้นหลังกับสีตัวอักษรมีความเหมาะสมต่อการอ่าน	61	3.69	0.83	ดี
6. ขนาดตัวอักษร และรูปแบบตัวอักษรอ่านได้ง่ายและสวยงาม	61	4.00	0.77	ดี
7. ภาพกับเนื้อหามีความสอดคล้องกันและสามารถสื่อความหมายได้	61	4.02	0.74	ดี
8. โดยภาพรวมท่านมีความพึงพอใจในการออกแบบเว็บไซต์ในระดับใด	61	3.98	0.87	ดี
รวม		**3.92**	**0.80**	**ดี**

จากตารางที่ 4.1 พบว่าการแสดงสีสันในการออกแบบเว็บไซต์มีความเหมาะสม ค่าเฉลี่ยสูงสุดเท่ากับ 4.03 (S.D. = 0.71) อยู่ในเกณฑ์ระดับดี ภาพกับเนื้อหามีความสอดคล้องกันและสามารถสื่อความหมายได้ มีค่าเฉลี่ย 4.02 (S.D. = 0.74) อยู่ในเกณฑ์ระดับดี การจัดรูปแบบในเว็บไซต์ง่ายต่อการอ่านและการใช้งาน มีค่าเฉลี่ย 4.00 (S.D. = 0.75) อยู่ในเกณฑ์ระดับดี ขนาดตัวอักษร และรูปแบบตัวอักษร อ่านได้ง่ายและสวยงาม มีค่าเฉลี่ย 4.00 (S.D. = 0.77) โดยภาพรวมท่านมีความพึงพอใจในการออกแบบเว็บไซต์ในระดับใด มีค่าเฉลี่ย 3.98 (S.D. = 0.87) อยู่ในเกณฑ์ระดับดี เมนูง่ายต่อการใช้งาน มีค่าเฉลี่ย 3.87 (S.D. = 0.

.85) อยู่ในเกณฑ์ระดับดี ความสวยงาม ความทันสมัย น่าสนใจของหน้าโฮมเพจ มีค่าเฉลี่ย 3.79 (S.D. = 0.86) อยู่ในเกณฑ์ระดับดี และสีพื้นหลังกับสีตัวอักษรมีความเหมาะสมต่อการอ่าน มีค่าเฉลี่ย 3.69 (S.D. = 0.83) อยู่ในเกณฑ์ระดับดี จากจำนวนผู้ตอบแบบสอบถามทั้งหมด 61 คน ค่าเฉลี่ยเท่ากับ 3.92 (S.D. = 0.80) อยู่ในเกณฑ์ระดับดี

ตารางที่ 4. 2 ตารางแสดงการประมวลผลด้านคุณภาพข้อมูล (Information Quality)

รายการประเมิน	**จำนวนผู้ตอบแบบสอบถาม**	$(\bar{x})$	**S.D.**	**ผลการประเมิน**
1. ข้อมูลในเว็บไซต์มีความเกี่ยวเนื่องกับงานของท่าน	61	4.00	0.77	ดี
2. ข้อมูลภายในเว็บไซต์เป็นข้อมูลที่ท่านสามารถทำความเข้าใจได้	61	4.15	0.68	ดี
3. ข้อมูลในแต่ละเมนูของเว็บไซต์ มีความครบถ้วนสมบูรณ์	61	4.44	0.70	ดี
4. ข้อมูลภายในเว็บไซต์มีความน่าเชื่อถือ	61	3.56	1.07	ดี
5. ข้อมูลภายในเว็บไซต์ มีความทันสมัย	61	3.95	0.67	ดี
6. ข้อมูลภายในเว็บไซต์มีความถูกต้อง	61	3.98	0.76	ดี
7.การนำเสนอรูปแบบ Visualization ด้วยโปรแกรม Google Data Studio มีความเหมาะสมกับข้อมูล	61	4.03	0.77	ดี
รวม		**4.02**	**0.78**	**ดี**

จากตารางที่ 4.2 พบว่าข้อมูลในแต่ละเมนูของเว็บไซต์ มีความครบถ้วนสมบูรณ์ ค่าเฉลี่ยสูงสุดเท่ากับ 4.44 (S.D. = 0.70) อยู่ในเกณฑ์ระดับดี ข้อมูลภายในเว็บไซต์เป็นข้อมูลที่ท่านสามารถทำความเข้าใจได้ มีค่าเฉลี่ย 4.15 (S.D. = 0.68) อยู่ในเกณฑ์ระดับดี การนำเสนอรูปแบบ Visualization ด้วยโปรแกรม Google Data Studio มีความเหมาะสมกับข้อมูล มีค่าเฉลี่ย 4.03 (S.D. = 0.77) อยู่ในเกณฑ์ระดับดี ข้อมูลในเว็บไซต์มีความเกี่ยวเนื่องกับงานของท่าน มีค่าเฉลี่ย 4.00 (S.D. = 0.77) อยู่ในเกณฑ์ระดับดี ข้อมูลภายในเว็บไซต์มีความถูกต้อง มี

ค่าเฉลี่ย 3.98 (S.D.=0. 0.76) ข้อมูลภายในเว็บไซต์ มีความทันสมัย มีค่าเฉลี่ย 3.95 (S.D.=0. 0.67) อยู่ในเกณฑ์ระดับดี และข้อมูลภายในเว็บไซต์มีความน่าเชื่อถือ มีค่าเฉลี่ย 3.56 (S.D. = 1.07) อยู่ในเกณฑ์ระดับดี จากจำนวนผู้ตอบแบบสอบถามทั้งหมด 61 คน ค่าเฉลี่ยเท่ากับ 4.02 (S.D. = 0.78) อยู่ในเกณฑ์ระดับดี

ตารางที่ 4. 3 แสดงการประเมินผลด้านประโยชน์และการนำไปใช้ (Perceived Usefulness)

รายการประเมิน	**จำนวนผู้ตอบแบบสอบถาม**	$(\bar{x})$	**S.D.**	**ผลการประเมิน**
1. เนื้อหามีประโยชน์ต่อผู้ใช้งาน และสามารถนำไปประยุกต์ใช้ได้	61	4.25	0.62	ดี
2. สามารถเป็นแหล่งความรู้ได้	61	4.21	0.69	ดี
3. เป็นแหล่งข้อมูลที่ตรงกับความต้องการของผู้ใช้งาน	61	3.89	0.95	ดี
4. โดยรวมท่านคิดว่าเป็นเว็บไซต์ที่มีประโยชน์	61	4.18	0.72	ดี
รวม		**4.13**	**0.74**	**ดี**

จากตารางที่ 4.3 พบว่า เนื้อหามีประโยชน์ต่อผู้ใช้งาน และสามารถนำไปประยุกต์ใช้ได้ ค่าเฉลี่ยสูงสุดเท่ากับ 4.25 (S.D. = 0.62) อยู่ในเกณฑ์ระดับดี สามารถเป็นแหล่งความรู้ได้ มีค่าเฉลี่ย 4.21 (S.D. = 0.69) อยู่ในเกณฑ์ระดับดี เป็นแหล่งข้อมูลที่ตรงกับความต้องการของผู้ใช้งาน มีค่าเฉลี่ย 3.89 (S.D. = 0.95) อยู่ในเกณฑ์ระดับดี โดยรวมท่านคิดว่าเป็นเว็บไซต์ที่มีประโยชน์ มีค่าเฉลี่ย 4.18 (S.D. = 0.72) อยู่ในเกณฑ์ระดับดี จากจำนวนผู้ตอบแบบสอบถามทั้งหมด 61 คน ค่าเฉลี่ยเท่ากับ 4.13 (S.D. = 0.74) อยู่ในเกณฑ์ระดับดี

4.3 บทสรุป

จากผลการดำเนินงานโครงการในข้างต้นทั้งหมดนี้ คณะผู้จัดทำได้นำข้อมูลการวิเคราะห์และแสดงผลแบบ Visualization ต่าง ๆ โดยใช้ Google Data Studio มาเผยแพร่ให้กับบุคคลภายนอกหรือกลุ่มผู้ใช้ข้อมูลได้รับในรูปแบบของเว็บไซต์ ที่จะเป็นแหล่งความรู้ต่าง ๆ ในเรื่องของอุบัติเหตุ วิธีการและขั้นตอนในการสร้างโมเดล Decision Tree สำหรับผู้ใช้ทั่วไปที่ต้องการตัวอย่างในการศึกษาค้นคว้า

บทที่ 5
สรุปและข้อเสนอแนะ

วิเคราะห์ข้อมูลผู้เสียชีวิตจากอุบัติเหตุทางถนน ปี 2554 – 2563 ทางคณะผู้จัดทำได้ทำการประเมินการใช้งานเว็บไซต์จากผู้ใช้งานทั่วไปเพื่อสรุปผลการจัดทำโครงงาน ข้อจำกัดของเว็บไซต์ ปัญหาและอุปสรรคต่าง ๆ ของการจัดทำโครงงาน และข้อเสนอแนะเพื่อพัฒนาโครงงานต่อไป ดังนี้

5.1 บทสรุปโครงงาน

จากการที่คณะผู้วิเคราะห์ได้ทำการศึกษาและวิเคราะห์ข้อมูลผู้เสียชีวิตจากอุบัติเหตุทางถนน ปี 2554 – 2563 โดยใช้เทคนิคการทำเหมืองข้อมูลแบบ Data Classification เพื่อใช้ทำแนวโน้มการเกิดขึ้นของปัจจัยที่สามารถทำให้เกิดอุบัติเหตุ จากเทคนิคการจำแนกประเภทข้อมูล ด้วยการสร้างโมเดล Decision Tree เพื่อจัดกลุ่มข้อมูลที่มีความสัมพันธ์กัน แล้วนำมาสร้างเป็นสารสนเทศ เพื่อเผยแพร่ข้อมูลบนเว็บไซต์ ในการพัฒนาเว็บไซต์ คือ HTML PHP JavaScript และ CSS เมื่อศึกษาความรู้พื้นฐานในภาษาต่าง ๆ และการวิเคราะห์ข้อมูลด้วยเทคนิคดังที่กล่าวมาข้างต้นแล้วจึงทำการวิเคราะห์ข้อมูลจนแล้วเสร็จจากนั้นได้ทำการประเมินผลจากการใช้งานจากกลุ่มตัวอย่าง จำนวน 61 คน ผู้วิเคราะห์ได้แบ่งการประเมินเป็น 3 ด้าน ซึ่งมีรายละเอียดดังนี้

1) ด้านการออกแบบและการจัดรูปแบบ (Design) จากจำนวนผู้ตอบแบบสอบถามทั้งหมด 61 คน ค่าเฉลี่ยเท่ากับ 3.92 (S.D. = 0.80) อยู่ในเกณฑ์ระดับดี

2) ด้านคุณภาพข้อมูล (Information Quality) จากจำนวนผู้ตอบแบบสอบถามทั้งหมด 61 คน ค่าเฉลี่ยเท่ากับ 4.02 (S.D. = 0.78) อยู่ในเกณฑ์ระดับดี

3) ด้านประโยชน์และการนำไปใช้ (Perceived Usefulness) จากจำนวนผู้ตอบแบบสอบถามทั้งหมด 61 คน ค่าเฉลี่ยเท่ากับ 4.13 (S.D. = 0.74) อยู่ในเกณฑ์ระดับดี

5.2 ข้อจำกัดของเว็บไซต์

5.2.1 ผู้ใช้งานทั่วไปไม่สามารถปรับเปลี่ยน แก้ไข เพิ่มข้อมูลบนเว็บไซต์ได้

5.2.2 ผู้ใช้งานทั่วไปไม่สามารถปรับเปลี่ยน แก้ไข กราฟบนเว็บไซต์ได้

5.2.3 การดาวน์โหลดชุดข้อมูลที่ผู้ใช้สามารถดาวน์โหลดได้ จะใช้เวลาค่อนข้างเยอะ เนื่องจากข้อมูลมีจำนวนมาก

5.2.4 ผู้ใช้งานไม่สามารถแยกดูข้อมูลในแต่ละจังหวัดได้

5.3 ปัญหาและอุปสรรคของโครงงาน

5.3.1 ข้อมูลที่จำนวนมากและมีความซับซ้อน คณะผู้จัดทำจึงต้องใช้เวลานานในการทำความเข้าใจและจัดการข้อมูลที่มีอยู่ให้อยู่ในรูปแบบที่เหมาะสม

5.3.2 ข้อมูลที่ได้มานั้นยังมีความผิดพลาดของข้อมูล จึงต้องทำความสะอาดชุดข้อมูล และ แทนค่าชุดข้อมูลบางส่วน

5.3.3 ข้อจำกัดของโปรแกรมต่าง ๆ ที่นำมาใช้ในการวิเคราะห์ข้อมูล เนื่องจากผู้เคราะห์ข้อมูลใช้แบบเวอร์ชั่นทดลองทำให้โปรแกรมมีข้อจำกัดหลายอย่าง เช่น ในโปรแกรม WEKA เวอร์ชั่น 3.8.4 ต้องซื้อแพ็คเกจภาษาไทยมาติดตั้งจึงจะสามารถนำเข้าข้อมูลเป็นภาษาไทยได้

5.3.4 การจัดทำเว็บไซต์ บางหน้าหรือบางเนื้อหา เกิดข้อผิดพลาดของโค้ด และฟังก์ชันบางจุด แก้ไขแล้วเกิดการผิดพลาดเนื้อหาหรือไม่ตรงตามที่ผู้จัดทำต้องการ

5.3.6 เนื่องด้วยสถานการณ์โควิด-19 ทำให้การจัดทำโครงงานของคณะผู้จัดทำล่าช้า เพราะเนื่องจากการติดต่อสื่อสารกับสมาชิกในการทำโครงงานและอาจารย์ที่ปรึกษาโครงงานต้องติดต่อสื่อสารกันผ่านทางออนไลน์ บางกรณีอินเตอร์เน็ตไม่เสถียรจึงเป็นปัจจัยที่สำคัญของปัญหาและอุปสรรคของการจัดทำโครงงานในครั้งนี้

5.4 ข้อเสนอแนะ

การวิเคราะห์ข้อมูลผู้เสียชีวิตจากอุบัติเหตุทางถนน ปี 2554 – 2563 ที่จัดทำเสร็จสิ้นแล้ว แม้ว่าจะสามารถจัดทำและแสดงผลได้ตรงตามวัตถุประสงค์และขอบเขตที่กำหนดไว้ แต่ยังมีอีกหลายข้อจำกัดที่ยังสามารถพัฒนาต่อได้ ซึ่งหากจะมีผู้พัฒนาให้เว็บไซต์เผยแพร่ข้อมูลแสดงผลและทำงานได้อย่างมีประสิทธิภาพมากยิ่งขึ้นผู้พัฒนาจะต้องปรับปรุงในส่วนต่าง ๆ

5.4.1 ควรมีการพัฒนาหน้าการแสดงผลต่าง ๆ ของเว็บไซต์ให้เหมาะสม สะดวก และเข้าใจง่ายมากยิ่งขึ้น เพื่อลดความผิดพลาดและเพิ่มประสิทธิภาพในการนำเสนอข้อมูล

5.4.2 ควรมีการพัฒนาในส่วนของเว็บไซต์ให้มีการรองรับบนโทรศัพท์ได้ในอนาคต

5.4.3 ควรมีการพัฒนาการปรับเปลี่ยนรูปแบบมุมมองการแสดงกราฟข้อมูลในรูปแบบต่าง ๆ ได้

5.4.4 ควรทำให้ระบบสามารถแยกดูรายงานยอดสรุปของแต่ละจังหวัดได้

บรรณานุกรม

ณัฏฐา กาญจนขุนดี. (2561). Big Data ข้อมูลขนาดใหญ่คืออะไร ทำงานอย่างไร และนำไปใช้อะไรได้บ้าง. สืบค้นเมื่อ วันที่ 5 มิถุนายน 2564, จาก https://www.khundee.com/big-data/.

เบญจมาศ ปิยะ. (2559). การศึกษาพฤติกรรมผู้ใช้งานตู้แช่เชิงพาณิชย์ด้วยเทคนิคต้นไม้ตัดสินใจ. กรุงเทพฯ : สาขาวิชาเทคโนโลยีสารสนเทศและการจัดการ มหาวิทยาลัยกรุงเทพ.สืบค้นเมื่อ วันที่ 5 มิถุนายน 2564, จาก http://dspace.bu.ac.th/handle/123456789/2313

ปฏิพล ธีร์โรจพร. (2562). Data Visualization ไม่ได้ทำให้สวย แต่ทำให้เข้าใจง่าย. สืบค้นเมื่อ วันที่ 12 มิถุนายน 2564, จาก https://blog.1moby.com/ds-ep-4-data-visualization/.

ภูมิพัฒน์ ดวงกลาง และรัจนา เครือแก้ว. (2562). แบบจำลองการทำนายแบบอากาศยานจากข้อมูลเป้าหมายไม่ราบฝ่ายอัตโนมัติ. สืบค้นเมื่อ วันที่ 12 มิถุนายน 2564, จาก https://www.tci-thaijo.org/index.php/nkrafa-sct/article/download/164341/155011/.

รุจิรา ธรรมสมบัติ. (2554). ระบบสนับสนุนการตัดสินใจในการเลือกใช้แพ็คเกจอินเทอร์เน็ตมือถือ โดยใช้ต้นไม้ตัดสินใจ. วิทยาลัยราชพฤกษ์. สืบค้นเมื่อ วันที่ 14 มิถุนายน 2564,

ศจี วาณิช. (2558). Data Mining (เหมืองข้อมูล). สืบค้นเมื่อ วันที่ 15 มิถุนายน 2564, จาก http://sajeegm301.blogspot.com/2015/11/data-mining.html.

สิทธิชัย รักษาสุข. (มปป.). CSS คืออะไร. สืบค้นเมื่อ วันที่ 14 มิถุนายน 2564, จาก https://www.dwthai.com/dwarticle/.

อภิยศ เหรียญวิพัฒน์. (ม.ป.ป.). การวิเคราะห์ข้อมูลและการนำเสนอข้อมูลเชิงภาพ เพื่อประยุกต์ใช้กับ การแสดงข้อมูลสถิติของสำนักหอสมุดมหาวิทยาลัยเกษตรศาสตร์. (ฉบับที่ 6), จาก https://pulinet.oas.psu.ac.th/index.php/journal/article/view/381

อาทิตย์ สิทธิบรรเจิด. (2552). การนำ Visualization ไปประยุกต์ใช้กับการแสดงข้อมูลบนระบบ Knowledge Management. กรุงเทพฯ: มหาวิทยาลัยกรุงเทพ. สืบค้นเมื่อ วันที่ 18 มิถุนายน 2564, จาก http://dspace.bu.ac.th/handle/123456789/667

บรรณานุกรม

อัจฉราภรณ์ จุฑาผาด. (2559). งานวิจัยเรื่องการพัฒนาระบบสารสนเทศเพื่อการ พ ย า ก ร ณ์จำนวนนักศึกษาใหม่ โดยใช้กฎการจำแนกต้นไม้ตัดสินใจ. (ฉบับที่278). นเรศวรวิจัยครั้งที่ 12: วิจัยและนวัตกรรมกับการพัฒนาประเทศ

อุทัย เหตุผล. (2558). หลักการออกแบบเว็บไซต์. สืบค้นเมื่อ วันที่ 18 มิถุนายน 2564, จาก https://sites.google.com/a/mattayom31.go.th/npw-html/html_design

เอกพล วิชัยสุข. (2561). การนำเสนอแผนภาพข้อมูล (Data Visualization). สืบค้นเมื่อวันที่ 24 มิถุนายน 2564, จาก http://www.autosoft.in.th/data-visualization/การนำเสนอแผนภาพข้อมูล-data-visu/.

เอื้อมพร หลินเจริญ. (2555). เทคนิคการวิเคราะห์ข้อมูลเชิงคุณภาพ วารสารการวัดผลการศึกษา, ปีที่ 17 (ฉบับที่ 1 กรกฎาคม 2555), 17-29.

อวยชัย โชติจรัสวาณิชย์. (2561). Big Data Characteristics ลักษณะของ Big Data.สืบค้นเมื่อวันที่ 27 มิถุนายน 2563, จาก https://uaychaiblog.com/2018/05/15/big-data-characteristics/.

ภาคผนวก

ภาคผนวก ก

คู่มือการใช้งานเว็บ

คู่มือการใช้งานเว็บไซต์

จากการดำเนินงานการพัฒนาเว็บไซต์สำหรับวิเคราะห์ข้อมูลผู้เสียชีวิตจากอุบัติเหตุทางถนน ปี 2554 – 2563 โดยเว็บไซต์มีส่วนใช้งาน 6 ส่วน ดังนี้ ส่วนเนื้อหาเกี่ยวกับการเกิดอุบัติเหตุและการป้องกันการเกิดอุบัติเหตุ ส่วนของการทำ Decision Tree การเปรียบเทียบโมเดล ส่วนของขั้นตอนการทดสอบข้อมูลกับโปรแกรม WEKA เวอร์ชัน 3.8.4 ส่วนการนำเสนอข้อมูลในรูปแบบ แผนภาพ Dashboard ส่วนเอกสาร และส่วนของแบบสอบถาม ผู้จัดทำจึงได้ทำคู่มืออธิบายการใช้งานเว็บไซต์ ดังต่อไปนี้

คู่มือการใช้งานสำหรับผู้เข้าใช้งานเว็บไซต์

1) หน้าแรกของเว็บไซต์ เมื่อเข้าสู่เว็บไซต์การพัฒนาเว็บไซต์สำหรับวิเคราะห์ข้อมูลผู้เสียชีวิตจากอุบัติเหตุทางถนน

ภาพที่ ก.1 แสดงหน้าแรกของเว็บไซต์

หมายเลข 1 ส่วนของแถบเมนูด้านข้างที่ประกอบไปด้วยเมนูการใช้งานเพื่อไปยังหน้าต่าง ๆ ภายในเว็บไซต์

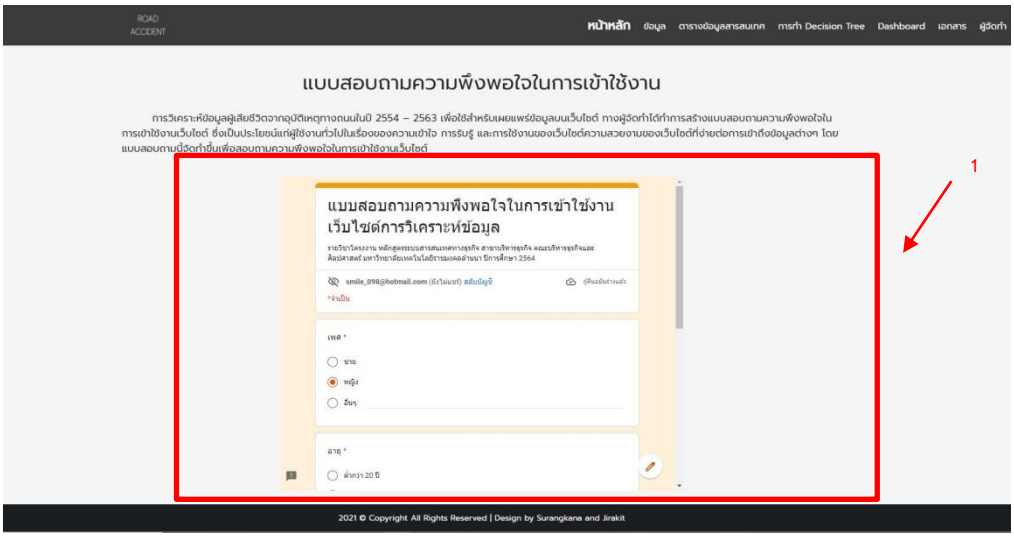
หมายเลข 2 ส่วนแสดงเนื้อหาในหน้าแรก แสดงรูปภาพการเกิดอุบัติเหตุ

หมายเลข 3 ส่วนของเว็บไซต์ ที่นำข้อมูลมาวิเคราะห์

หมายเลข 4 ส่วนเข้าสู่ระบบ เข้าได้เฉพาะผู้ดูแลระบบเท่านั้น

หมายเลข 5 ส่วนของแบบสอบถามความพึงพอใจเกี่ยวกับเว็บไซต์

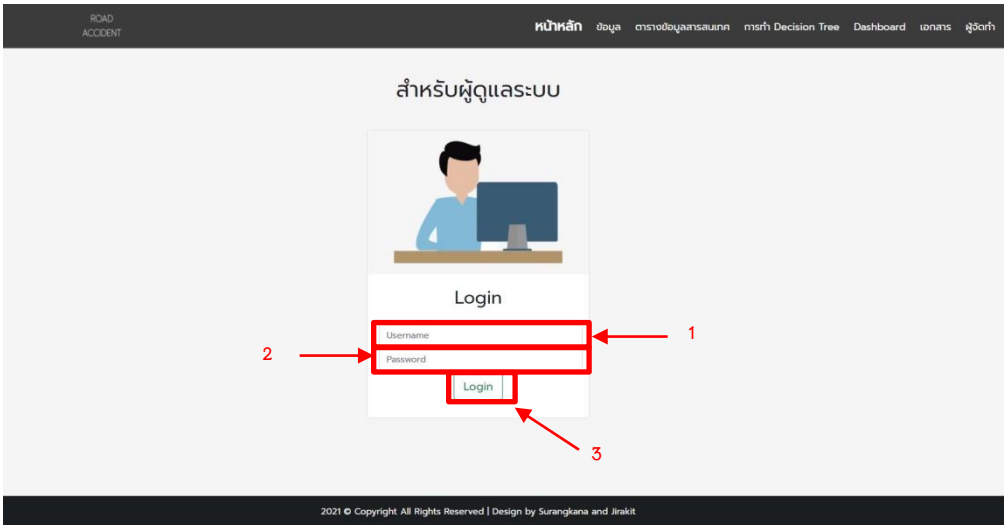
2) แสดงหน้าแบบสอบถามความพึงพอใจในการเข้าใช้งาน



ภาพที่ ก.2 แสดงหน้าแบบสอบถามความพึงพอใจ

หมายเลข 1 ส่วนแบบฟอร์มแบบสอบถามความพึงพอใจในการเข้าใช้เว็บไซต์

3) แสดงหน้าเข้าสู่ระบบสำหรับผู้ดูแลระบบเท่านั้น



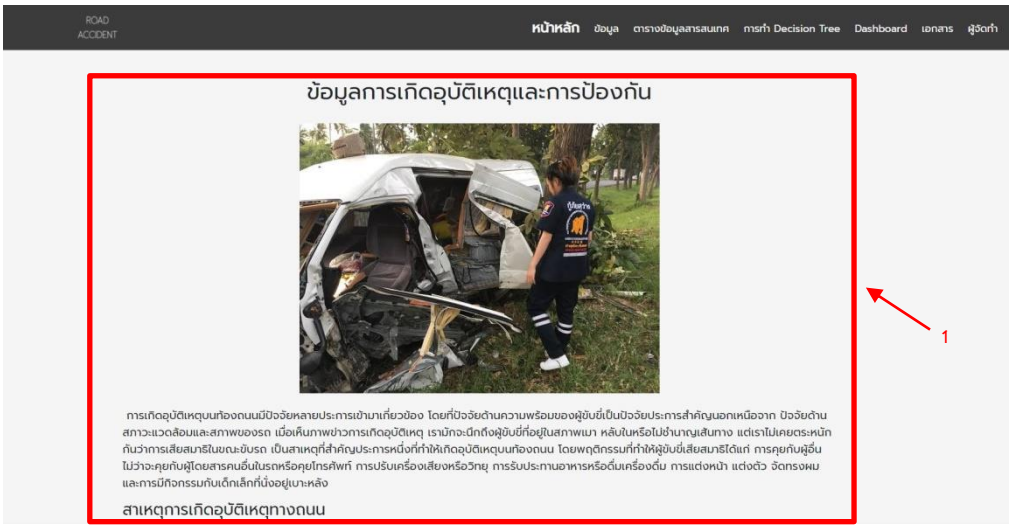
ภาพที่ ก.3 แสดงเข้าสู่ระบบสำหรับผู้ดูแลระบบ

หมายเลข 1 ใส่ชื่อผู้ใช้ (Username) ของผู้ดูแลระบบ

หมายเลข 2 ใส่รหัส (Password) ของผู้ดูแลระบบ

หมายเลข 3 ปุ่มเข้าสู่ระบบ (Login)

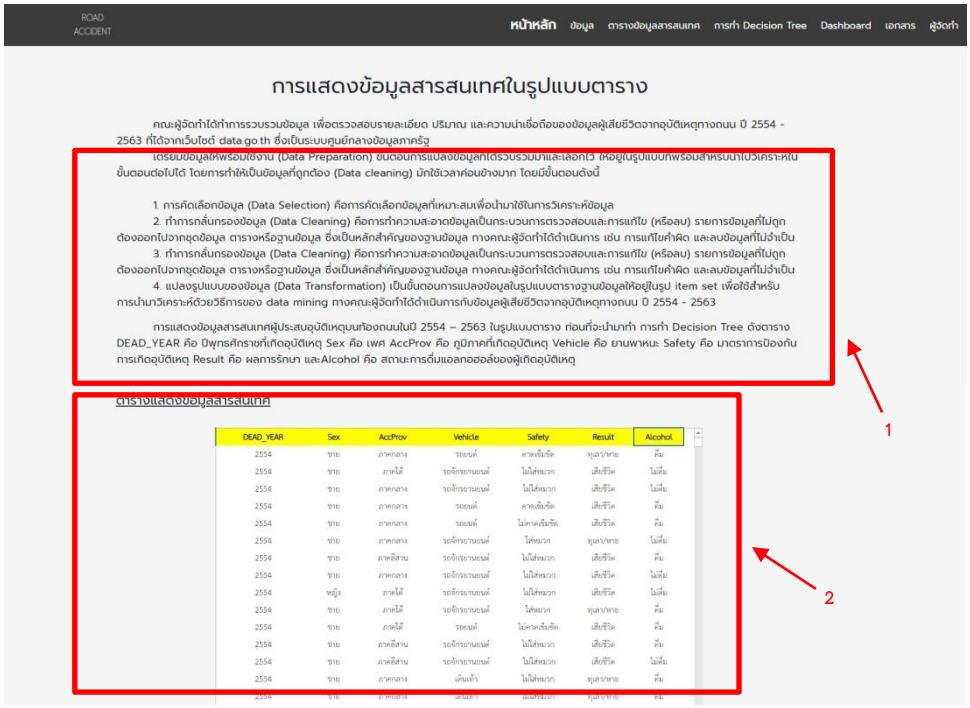
4) การแสดงข้อมูลเกี่ยวกับการเกิดอุบัติบนท้องถนนและการป้องกัน



ภาพที่ ก.4 แสดงข้อมูลเกี่ยวกับการเกิดอุบัติบนท้องถนนและการป้องกัน

หมายเลข 1 แสดงข้อมูลข้อมูลเกี่ยวกับการเกิดอุบัติบนท้องถนนและการป้องกัน

5) แสดงหารเตรียมข้อมูลให้พร้อมใช้งานและแสดงข้อมูลสารสนเทศในรูปแบบตาราง

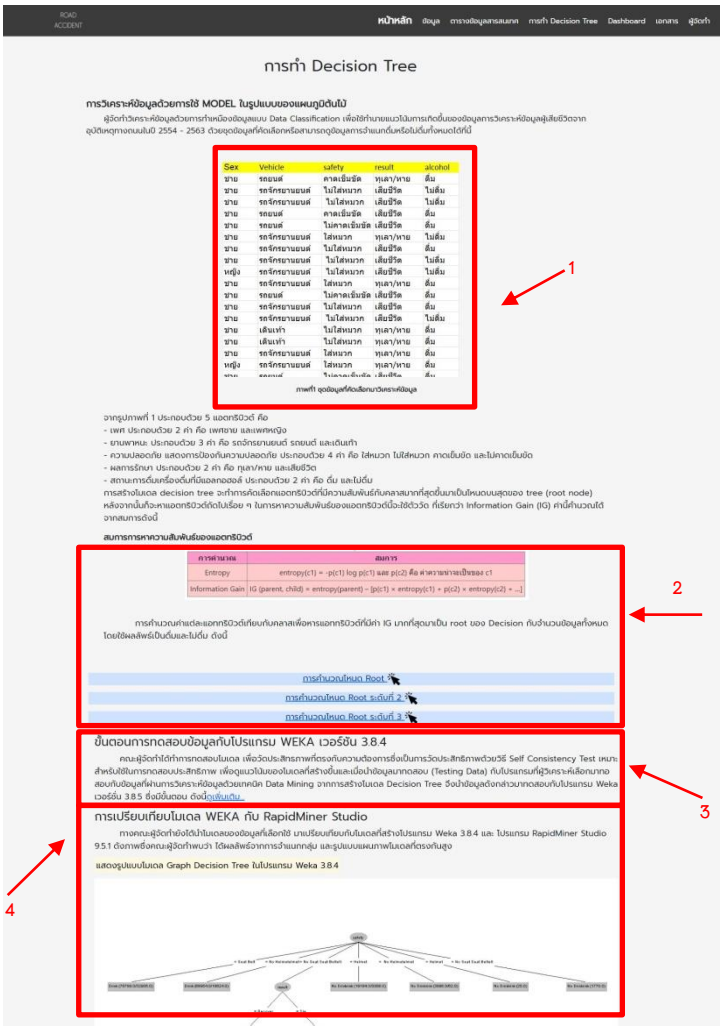


ภาพที่ ก.5 แสดงข้อมูลสารสนเทศในรูปแบบตาราง

หมายเลข 1 แสดงขั้นตอนการเตรียมข้อมูลเพื่อให้พร้อมใช้งานในการวิเคราะห์

หมายเลข 2 แสดงข้อมูลสารสนเทศในรูปแบบตาราง

6) การวิเคราะห์ข้อมูลด้วยการใช้ MODEL ในรูปแบบของแผนภูมิต้นไม้



ภาพที่ ก.6 แสดงการวิเคราะห์ข้อมูลด้วยการใช้ MODEL โดย Decision Tree

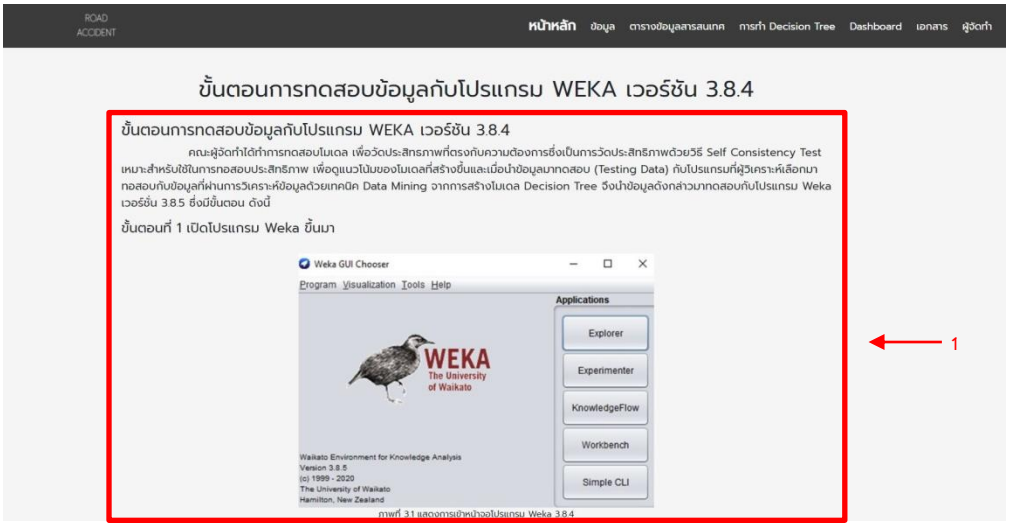
หมายเลข 1 แสดงข้อมูลในรูปแบบตารางที่นำมาวิเคราะห์

หมายเลข 2 สมการการหาความสัมพันธ์ของแอตทริบิวต์ และการคำนวณหาแอตทริบิวต์ ค่าที่มีค่ามากที่สุดมาเป็น Root และการหา Root ระดับที่ 2 และ ระดับที่ 3 ของ Decision Tree

หมายเลข 3 ขั้นตอนการทดสอบข้อมูลกับโปรแกรม WEKA เวอร์ชัน 3.8.4

หมายเลข 4 การเปรียบเทียบโมเดล WEKA กับ RapidMiner Studio

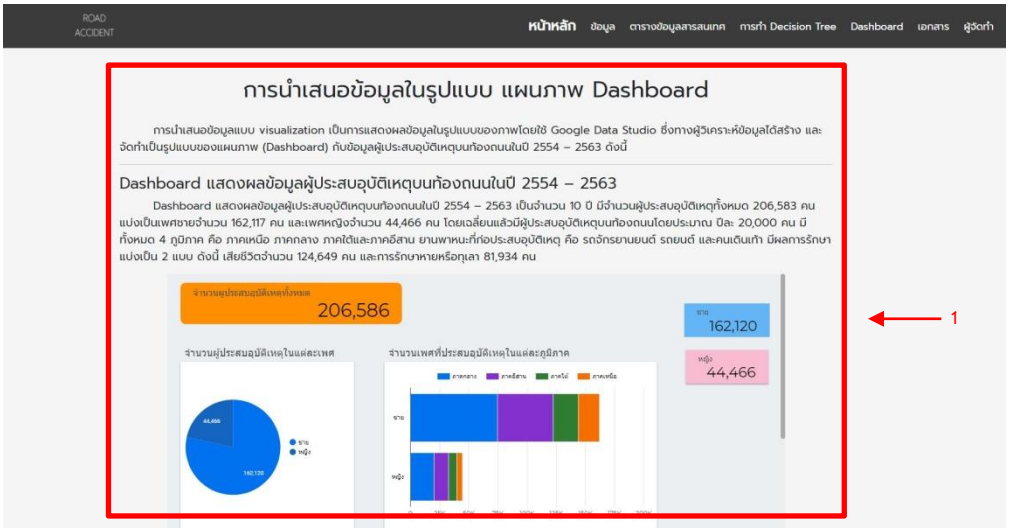
7) แสดงขั้นตอนการทดสอบข้อมูลกับโปรแกรม WEKA เวอร์ชัน 3.8.4



ภาพที่ ก.7 แสดงขั้นตอนการทดสอบข้อมูลกับโปรแกรม WEKA เวอร์ชัน 3.8.4

หมายเลข 1 แสดงขั้นตอนการนำข้อมูลที่วิเคราะห์เรียบร้อยแล้ว มาทำการทดสอบกับโปรแกรม WEKA เวอร์ชัน 3.8.4

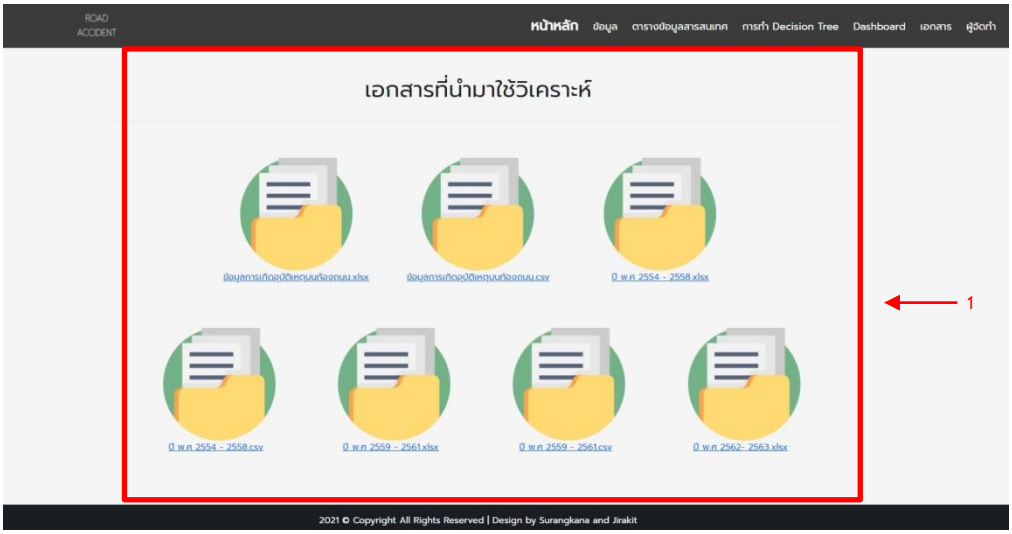
8) แสดงการนำเสนอข้อมูลในรูปแบบ แผนภาพ Dashboard



ภาพที่ ก.8 แสดงการนำเสนอข้อมูลในรูปแบบ แผนภาพ Dashboard

หมายเลข 1 แสดงการนำเสนอข้อมูลในรูปแบบ แผนภาพ Dashboard โดย Google Data Studio สามารถกรองดูข้อมูลได้

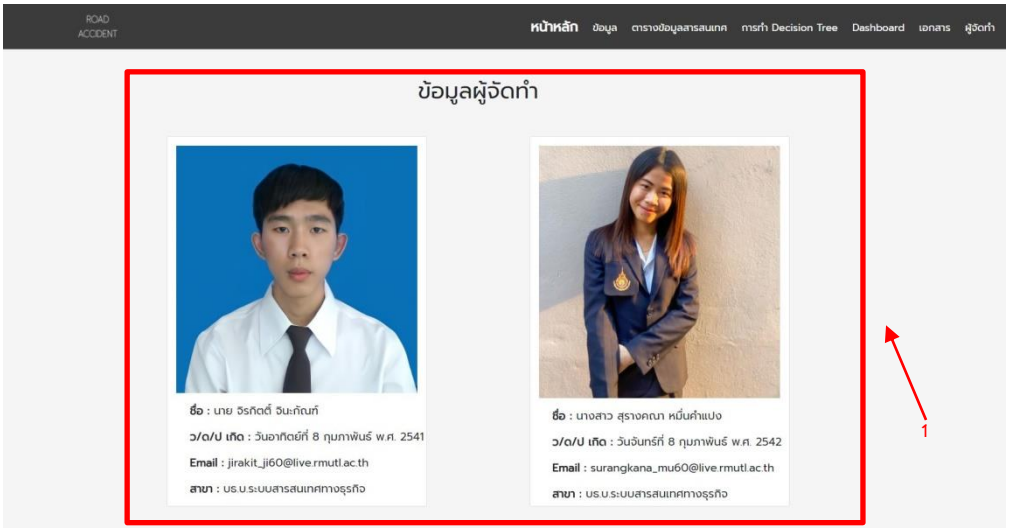
9) หน้าเอกสารที่นำมาวิเคราะห์



ภาพที่ ก.9 แสดงหน้าเอกสารที่นำมาวิเคราะห์

หมายเลข 1 แสดงหน้าเอกสารที่นำมาวิเคราะห์ และสามารถดาวน์โหลดได้

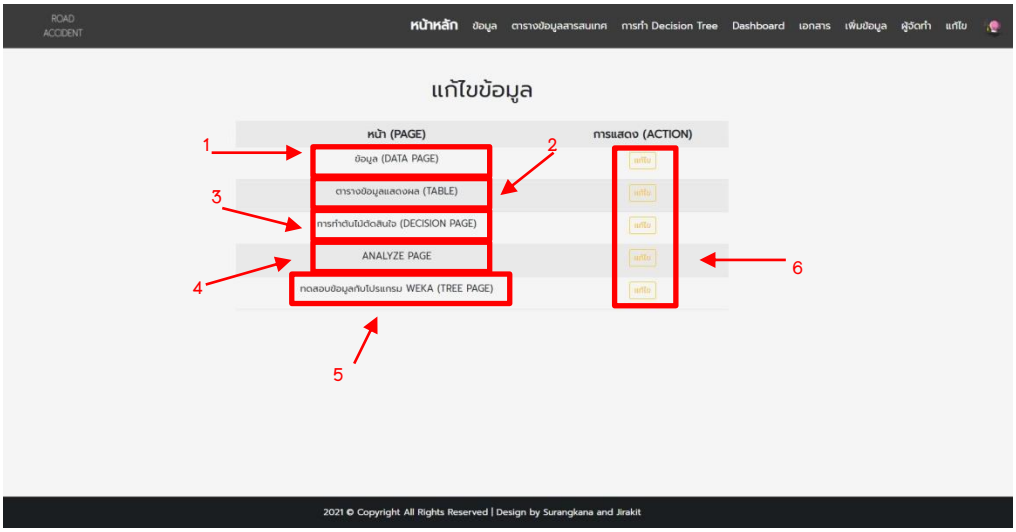
10) หน้าแสดงผู้จัดทำเว็บไซต์



ภาพที่ ก.10 แสดงผู้จัดทำเว็บไซต์

หมายเลข 1 แสดงรูปภาพและข้อมูลผู้จัดทำเว็บไซต์

11) แสดงหน้าที่สามารถแก้ไขข้อมูลได้ สำหรับผู้ดูแลระบบเท่านั้น



ภาพที่ ก.11 แสดงหน้าที่สามารถแก้ไขข้อมูลได้

หมายเลข 1 ข้อมูลการเกิดอุบัติเหตุและการป้องกัน (DATA PAGE)

หมายเลข 2 ตารางข้อมูลแสดงผล (TABLE)

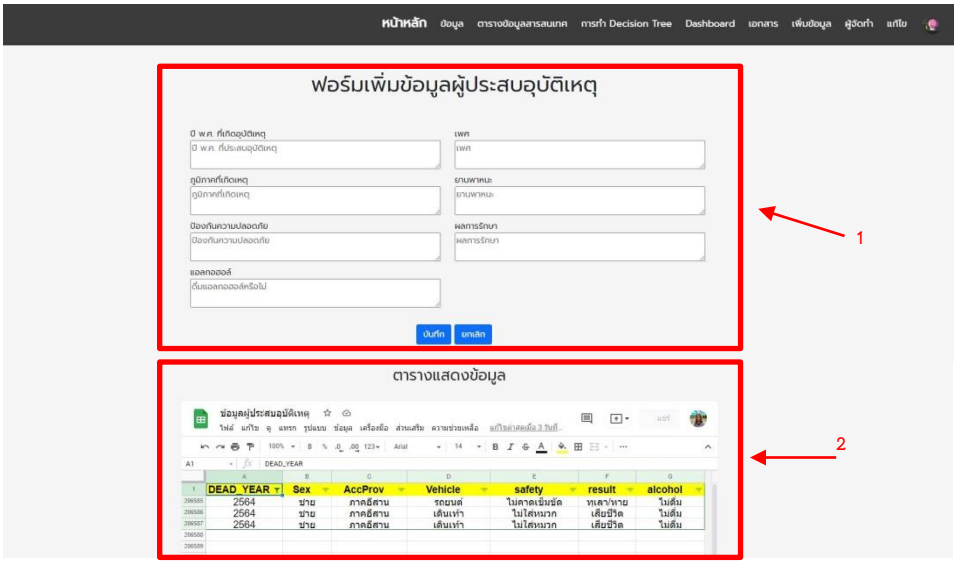
หมายเลข 3 การทำต้นไม้ตัดสินใจ (DECISION PAGE)

หมายเลข 4 ANALYZE PAGE

หมายเลข 5 ทดสอบข้อมูลกับโปรแกรม WEKA (TREE PAGE)

หมายเลข 6 ปุ่มสำหรับกดแก้ไขข้อมูลในแต่ละหน้า

12) หน้าฟอร์มเพิ่มข้อมูลผู้ประสบอุบัติเหตุและแสดงข้อมูล สำหรับผู้ดูแลระบบเท่านั้น

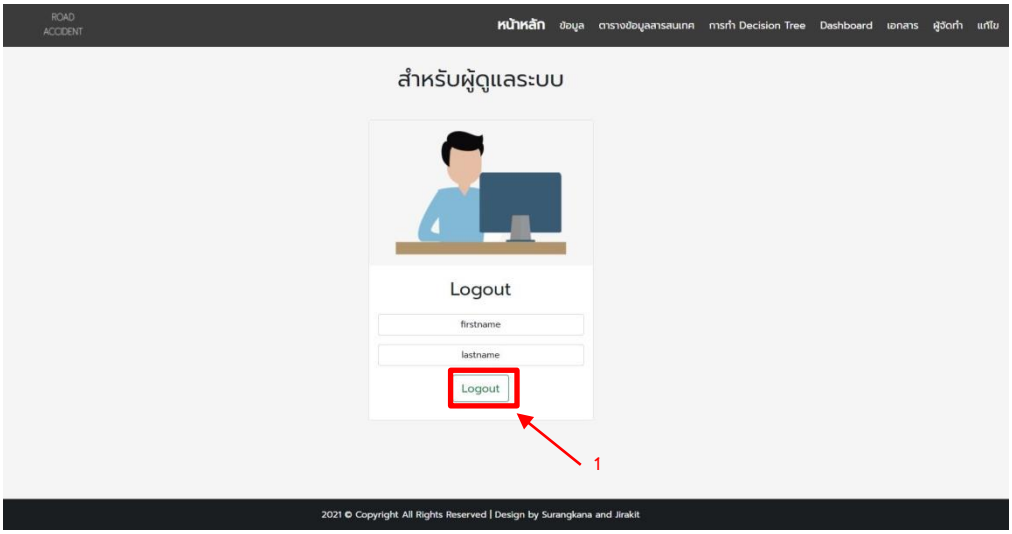


ภาพที่ ก.12 แสดงหน้าที่สามารถแก้ไขข้อมูลได้

หมายเลข 1 ฟอร์มเพิ่มข้อมูลผู้ประสบอุบัติเหตุ

หมายเลข 2 แสดงข้อมูลที่เพิ่มเข้าไป และสามารถ เพิ่ม ลบ แก้ไข ข้อมูลภายในตารางที่แสดงได้เลย **ต้องเข้าสู่ระบบของ Google ก่อน**

13) แสดงหน้าออกจากระบบสำหรับผู้ดูแลระบบเท่านั้น



ภาพที่ ก.13 แสดงออกจากระบบสำหรับผู้ดูแลระบบ

หมายเลข 3 ปุ่มออกจากระบบ (Logout)

ภาคผนวก ข

แบบสอบถาม

แบบสอบถามความพึงพอใจการใช้เว็บไซต์

แบบสอบถามนี้มีวัตถุประสงค์เพื่อสำรวจความพึงพอใจการใช้เว็บไซต์โครงงาน เรื่อง การพัฒนาเว็บไซต์สำหรับวิเคราะห์ข้อมูลผู้เสียชีวิตจากอุบัติเหตุทางถนน ปี 2554 – 2563 แบ่งออกเป็น 2 ตอน คือ

ตอนที่ 1 ข้อมูลเกี่ยวกับสถานภาพและข้อมูลพื้นฐานของผู้ตอบแบบสอบถาม

คำชี้แจง โปรดให้รายละเอียดที่เกี่ยวกับตัวท่าน โดยเขียนเครื่องหมาย √ ลงใน ( ) หน้าข้อความตามความเป็นจริง

เพศ ( ) ชาย ( ) หญิง

อายุ ( ) ต่ำกว่า 20 ปี ( ) 20–30 ปี
( ) 31–40 ปี ( ) 41–50 ปี
( ) 51–60 ปี ( ) มากกว่า 60 ปี

ระดับการศึกษา ( ) ต่ำกว่าปริญญาตรี ( ) ปริญญาตรี
( ) ปริญญาโท ( ) สูงกว่าปริญญาโท

ตอนที่ 2 การประเมินความพึงพอใจ

คำชี้แจง เขียนเครื่องหมาย √ ลงในช่องระดับความพึงพอใจตามความเป็นจริง

หัวข้อประเมิน	**ระดับความพึงพอใจ**				
	มากที่สุด (5)	**มาก (4)**	**ปานกลาง (3)**	**น้อย (2)**	**น้อยที่สุด (1)**
ด้านการออกแบบและการจัดรูปแบบ (Design)					
1. ความสวยงาม ความทันสมัย น่าสนใจของหน้าโฮมเพจ					
2. เมนูง่ายต่อการใช้งาน					
3. ภาพกับเนื้อหามีความสอดคล้องกันและสามารถสื่อความหมายได้					

หัวข้อประเมิน	ระดับความพึงพอใจ				
	มากที่สุด (5)	มาก (4)	ปานกลาง (3)	น้อย (2)	น้อยที่สุด (1)
4. โดยภาพรวมท่านมีความพึงพอใจในการออกแบบเว็บไซต์ในระดับใด					
5. สีสันในการออกแบบเว็บไซต์มีความเหมาะสม					
6. ขนาดตัวอักษร และรูปแบบตัวอักษร อ่านได้ง่ายและสวยงาม					
7. การจัดรูปแบบในเว็บไซต์ง่ายต่อการอ่านและการใช้งาน					
8. สีพื้นหลังกับสีตัวอักษรมีความเหมาะสมต่อการอ่าน					
ด้านคุณภาพข้อมูล (Information Quality)					
1. ข้อมูลภายในเว็บไซต์ มีความทันสมัย					
2. ข้อมูลภายในเว็บไซต์มีความถูกต้อง					
3. การนำเสนอรูปแบบ Visualization ด้วย Google Data Studio มีความเหมาะสมกับข้อมูล					

หัวข้อประเมิน	ระดับความพึงพอใจ				
	มากที่สุด (5)	มาก (4)	ปานกลาง (3)	น้อย (2)	น้อยที่สุด (1)
4. ข้อมูลในเว็บไซต์มีความเกี่ยวเนื่องกับงานของท่าน					
5. ข้อมูลภายในเว็บไซต์เป็นข้อมูลที่ท่านสามารถทำความเข้าใจได้					
6. ข้อมูลในแต่ละเมนูของเว็บไซต์ มีความครบถ้วนสมบูรณ์					
7. ข้อมูลภายในเว็บไซต์มีความน่าเชื่อถือ					
ด้านประโยชน์และการนำไปใช้ (Perceived Usefulness)					
1. เนื้อหามีประโยชน์ต่อผู้ใช้งานและสามารถนำไปประยุกต์ใช้ได้					
2. สามารถเป็นแหล่งความรู้ได้					
3. เป็นแหล่งข้อมูลที่ตรงกับความต้องการของผู้ใช้งาน					
4. โดยรวมท่านคิดว่าเป็นเว็บไซต์ที่มีประโยชน์					

ข้อเสนอแนะต่อการพัฒนา/ปรับปรุงเว็บไซต์

.....

.....

.....

.....

.....

ภาคผนวก ค

แบบบันทึกรายละเอียดการเข้าพบอาจารย์ที่ปรึกษา

ภาคผนวก ค

แบบบันทึกรายละเอียดการเข้าพบอาจารย์ที่ปรึกษา

FM-BIS-10

แบบบันทึกรายละเอียดการเข้าพบอาจารย์ที่ปรึกษา
วิชา Business Information System Project
หลักสูตรระบบสารสนเทศทางธุรกิจ สาขาบริหารธุรกิจ
คณะบริหารธุรกิจและศิลปศาสตร์ มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา เชียงใหม่

ชื่อโครงการ การพัฒนาเว็บไซต์สำหรับวิเคราะห์ข้อมูลผู้เสียชีวิตจากอุบัติเหตุทางถนน ปี 2554 – 2563

ผู้จัดทำ 1. นายจิรกิตติ์ จินะกัณฑ์ รหัสนักศึกษา 60541207073-8
2. นางสาวสุรางคณา หมื่นคำแปง รหัสนักศึกษา 60541207095-1

อาจารย์ที่ปรึกษา ผู้ช่วยศาสตร์จารย์ ดร.ชัชฎาพร ปุกแก้ว

ครั้งที่	วัน/เดือน/ปี	รายละเอียดการเข้าพบอาจารย์ที่ปรึกษา	ความคืบหน้าของผลงาน	ลายมืออาจารย์ที่ปรึกษา
1	18/06/64	ตรวจรูปเล่มรายงานบทที่ 1-2	30%	ชัชฎาพร ปุกแก้ว
2	23/06/64	ตรวจความคืบหน้ารายงานและส่งบทที่ 3 และหน้าเว็บไซต์	40%	ชัชฎาพร ปุกแก้ว
3	5/08/64	ตรวจความคืบหน้ารายงานฉบับแก้ไข บทที่ 1-3 และหน้าเว็บไซต์	50%	ชัชฎาพร ปุกแก้ว
4	27/08/64	ตรวจความคืบหน้าของงานก่อนเข้าสอบความคืบหน้า	80%	ชัชฎาพร ปุกแก้ว
5	16/09/64	รายงานความคืบหน้าหลังสอบความคืบหน้า และตรวจรายงานบทที่ 4-5	85%	ชัชฎาพร ปุกแก้ว
6	24/09/64	รายงานความคืบหน้ารายงานบทที่ 4-5 ฉบับแก้ไขและหน้าเว็บไซต์	90%	ชัชฎาพร ปุกแก้ว
7	1/10/64	ตรวจรูปเล่มบทที่ 1 – 5 และเว็บไซต์ก่อนเข้าสอบ	100%	ชัชฎาพร ปุกแก้ว

ลงชื่อ.....จิรกิตติ์ จินะกัณฑ์..... (นักศึกษา) ลงชื่อ .....สุรางคณา หมื่นคำแปง.....(นักศึกษา)
(นายจิรกิตติ์ จินะกัณฑ์) (นางสาวสุรางคณา หมื่นคำแปง)

ประวัติผู้จัดทำ

ผู้จัดทำ : นายจิรกิตติ์ จินะกัณฑ์

ที่อยู่อาศัย : 68/1 หมู่ 1 ถนน เวียงใหม่ ตำบลแม่สะเรียง
อำเภอแม่สะเรียง จังหวัดแม่ฮ่องสอน 58110

การศึกษา : มหาวิทยาลัยเทคโนโลยีราชมงคล เชียงใหม่
คณะบริหารธุรกิจและศิลปศาสตร์ สาขาบริหารธุรกิจ
สาขาวิชา บธ.บ.ระบบสารสนเทศทางธุรกิจ

E-Mail : Jirakit_ji60@live.rmutl.ac.th

เบอร์โทรศัพท์ : 093-1849779

ประวัติผู้จัดทำ

ผู้จัดทำ : นางสาวสุรางคณา หมื่นคำแปง

ที่อยู่อาศัย : 314/2 หมู่ 7 ตำบลห้างฉัตร อำเภอห้างฉัตร
จังหวัดลำปาง 52190

การศึกษา : มหาวิทยาลัยเทคโนโลยีราชมงคล เชียงใหม่
คณะบริหารธุรกิจและศิลปศาสตร์ สาขาบริหารธุรกิจ
สาขาวิชา บธ.บ.ระบบสารสนเทศทางธุรกิจ

E-Mail : surangkana_mu60@live.rmutl.ac.th

เบอร์โทรศัพท์ : 061-3806918