การวิเคราะห์ข้อมูลผู้บาดเจ็บและเสียชีวิตในช่วงเทศกาลปีใหม่ ประจำปี 2551 – 2558 เพื่อใช้สำหรับเผยแพร่ข้อมูลบนเว็บไซต์

นางสาวมนิศรา เทพฝั้น

นางสาวศิริขวัญ ขันนาแล

หลักสูตรระบบสารสนเทศทางธุรกิจ

สาขาบริหารธุรกิจ คณะบริหารธุรกิจและศิลปศาสตร์

มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา

ปีการศึกษา 2562

ชื่อโครงงาน	การวิเคราะห์ข้อมูลผู้บาดเจ็บและเสียชีวิตในช่วงเทศกาลปีใหม่ ประจำปี 2551 - 2558 เพื่อใช้สำหรับเผยแพร่ข้อมูลบนเว็บไซต์
โดย	นางสาวมนิศรา เทพฝั้น รหัส 61521207041-6
	นางสาวศิริขวัญ ขันนาแล รหัส 61521207045-7
หลักสูตร	ระบบสารสนเทศทางธุรกิจ
	สาขาบริหารธุรกิจ คณะบริหารธุรกิจและศิลปศาสตร์
อาจารย์ที่ปรึกษา	ผู้ช่วยศาสตราจารย์ ดร. ชัชฎาพร ปุกแก้ว

หลักสูตรระบบสารสนเทศทางธุรกิจ สาขาบริหารธุรกิจ คณะบริหารธุรกิจ และศิลปศาสตร์ มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา อนุมัติให้นับโครงงานนี้ เป็นส่วนหนึ่งของการศึกษาหลักสูตรบริหารธุรกิจบัณฑิต

(.....)

(อาจารย์สุพงศ์ แดงสุริยศรี)

หัวหน้าหลักสูตรระบบสารสนเทศทางธุรกิจ

คณะกรรมการสอบโครงงาน

.....ประธานกรรมการ

(ผู้ช่วยศาสตราจารย์ ดร. ชัชฎาพร ปุกแก้ว)

.....กรรมการ

(อาจารย์นรินทร์ จิวิตัน)

.....กรรมการ

(อาจารย์พรพิพัฒน์ ทองปรอน)

© ลิขสิทธิ์ของหลักสูตรระบบสารสนเทศทางธุรกิจ
สาขาบริหารธุรกิจ คณะบริหารธุรกิจและศิลปศาสตร์

ชื่อโครงงาน การวิเคราะห์ข้อมูลผู้บาดเจ็บและเสียชีวิตในช่วงเทศกาลปีใหม่ ประจำปี 2551 – 2558 เพื่อใช้สำหรับเผยแพร่ข้อมูลบนเว็บไซต์

โดย นางสาวมนิศรา เทพฝั้น รหัส 61521207041-6

นางสาวศิริขวัญ ขันนาแล รหัส 61521207045-7

อาจารย์ที่ปรึกษา ผู้ช่วยศาสตราจารย์ ดร. ชัชฎาพร ปุกแก้ว

หลักสูตร ระบบสารสนเทศทางธุรกิจ

สาขาบริหารธุรกิจ คณะบริหารธุรกิจและศิลปศาสตร์

ปีการศึกษา 2562

บทคัดย่อ

การจัดทำโครงงานในครั้งนี้มีวัตถุประสงค์ เพื่อวิเคราะห์ข้อมูลผู้บาดเจ็บและเสียชีวิตในเทศกาลปีใหม่ประจำปี 2551 – 2558 สำหรับเผยแพร่บนเว็บไซต์ 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 70.12% และได้กฎจำนวน 9 กฎ ซึ่งคณะผู้จัดทำได้จัดทำเว็บไซต์สำหรับเผยแพร่ข้อมูลสารสนเทศนี้บน Web browser

โดยเว็บไซต์ที่สร้างขึ้น เป็นลักษณะของการเผยแพร่ข้อมูลที่ได้จากการวิเคราะห์ข้อมูลและการแสดงผลในรูปแบบของแผนภาพโดยใช้โปรแกรม Tableau Public ซึ่งผู้ใช้งานทั่วไปสามารถกรองข้อมูลสารสนเทศผ่านระบบของ Tableau บนเว็บไซต์ที่เผยแพร่ได้ ซึ่งเว็บไซต์นี้ใช้ภาษา HTML CSS PHP และ JavaScript และมีส่วนประกอบต่าง ๆ ดังนี้ 1) ส่วนเนื้อหาเกี่ยวกับอุบัติเหตุ 2) ส่วนการวิเคราะห์ข้อมูลด้วย Decision Tree 3) ส่วนแบบฟอร์มจำลองการปฏิบัติตัวระหว่างเดินทางในช่วงเทศกาลปีใหม่ 4) ส่วนสรุปผลข้อมูลเป็นรายปี 5) ส่วนการสรุปผลข้อมูลในภาพรวม 6) ส่วนของแนวทางป้องกันการเกิดอุบัติเหตุ 7) ส่วนของแบบสอบถาม

จากการวิเคราะห์ข้อมูล และสร้างเว็บไซต์ พบว่าเว็บไซต์ที่ได้สามารถนำมาประยุกต์ใช้งานได้ตรงตามวัตถุประสงค์ที่ตั้งไว้ ทำให้เกิดประสิทธิภาพในการศึกษาค้นคว้าข้อมูลตามที่ผู้ใช้ต้องการ สรุปได้ว่าในอนาคตสามารถนำข้อมูลจากเว็บไซต์ไปปรับปรุงเพิ่มเติมต่อได้ เพื่อให้เกิดประโยชน์กับความปลอดภัยของบุคคลทั่วไปให้ได้ประสิทธิภาพมากขึ้นต่อไป

กิตติกรรมประกาศ

โครงงานครั้งนี้ สำเร็จลุล่วงไปด้วยดี เนื่องจากได้รับความกรุณาอย่างดียิ่งจาก ผู้ช่วย-ศาสตราจารย์ ดร.ชัชฎาพร ปุกแก้ว อาจารย์ที่ปรึกษาโครงงาน อาจารย์ สุพงศ์ แดงสุริยศรี หัวหน้า สาขาวิชาระบบสารสนเทศทางธุรกิจ และคณะอาจารย์ในสาขาระบบสารสนเทศทางธุรกิจ คณะบริหารธุรกิจและศิลปศาสตร์ มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนาทุกท่าน ที่ได้ให้คำปรึกษาและคำแนะนำในการจัดทำโครงงาน เพื่อบรรลุวัตถุประสงค์ของโครงงานให้เสร็จสมบูรณ์ รวมถึงช่วยแก้ไขเอกสารโครงงาน ให้มีความสมบูรณ์มากยิ่งขึ้น

ขอขอบคุณสนักงานรัฐบาลอิเล็กทรอนิกส์ (องค์การมหาชน) (สรอ.) ที่ได้พัฒนาระบบศูนย์กลางข้อมูลภาครัฐ หรือ Open Government Data of Thailand ในเว็บไซต์ data.go.th เพื่อเป็นการส่งเสริมและผลักดันให้เกิดการเปิดเผยข้อมูลภาครัฐสู่สาธารณะอย่างเป็นรูปธรรม ซึ่งเป็นเว็บไซต์ที่เก็บรวบรวมชุดข้อมูลต่าง ๆ ของประเทศไทย ทำให้คณะผู้จัดทำได้มาซึ่งข้อมูลที่ต้องการ เพื่อนำมาวิเคราะห์ตามกระบวนการและขั้นตอนต่าง ๆ

ท้ายสุดนี้ คณะผู้จัดทำโครงงานขอขอบคุณผู้มีส่วนเกี่ยวข้องทุกท่านที่กรุณาประสิทธิ์ประสาทวิชาความรู้และประสบการณ์อันมีค่ายิ่งแก่ศิษย์ และคอยชี้แนะแนวทางการ จัดทำโครงงานพร้อมทั้ง ช่วยส่งเสริมสนับสนุนให้คณะผู้จัดทำโครงงานเกิดแรงผลักดัน เป็นกำลังใจและให้ความช่วยเหลือต่าง ๆ อีกทั้งยังมอบข้อคิดเห็นที่เป็นประโยชน์ในการดำเนิน โครงงานนี้อันส่งผลให้สำเร็จลุล่วงไปด้วยดี ขอกราบขอบพระคุณเป็นอย่างสูงไว้ ณ ที่นี้

มนิศรา เทพฝั้น

ศิริขวัญ ขันนาแล

สารบัญ

	หน้า
บทคัดย่อ	ก
กิตติกรรมประกาศ	ข
สารบัญ	ค
สารบัญภาพ	จ
สารบัญตาราง	ญ
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	5
1.7 ระยะเวลาในการดำเนินการ	6
1.8 นิยามศัพท์เฉพาะ	6
1.9 บทสรุป	7
บทที่ 2 แนวคิด ทฤษฎี เครื่องมือและวรรณกรรมที่เกี่ยวข้อง	8
2.1 แนวคิด	9
2.2 ทฤษฎี	17
2.3 เครื่องมือในการออกแบบและวิเคราะห์ข้อมูล	33
2.4 วรรณกรรมที่เกี่ยวข้อง	37
2.5 บทสรุป	42
บทที่ 3 วิธีการดำเนินงานโครงงาน	43
3.1 การวิเคราะห์ข้อมูลด้วย CRISP-DM	43
3.2 การออกแบบเว็บไซต์	88
3.3 บทสรุป	92

สารบัญ (ต่อ)

	หน้า
บทที่ 4 ผลการดำเนินงาน	93
4.1 ผลการดำเนินงาน	93
4.2 การอภิปรายผล	102
4.3 บทสรุป	105
บทที่ 5 สรุปและข้อเสนอแนะ	106
5.1 บทสรุปผลโครงงาน	106
5.2 ข้อจำกัดของระบบ	107
5.3 ปัญหาและอุปสรรคของโครงงาน	108
5.4 ข้อเสนอแนะ	108
บรรณานุกรม	109
ภาคผนวก ก คู่มือการใช้งานเว็บไซต์	112
ภาคผนวก ข แบบฟอร์มและเอกสารที่ใช้ในโครงงาน	125
ประวัติผู้จัดทำ	129

สารบัญภาพ

ภาพที่		หน้า
2.1	แผนภูมิแท่ง (Bar Charts)	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)	16
2.9	เว็บที่มีโครงสร้างแบบเรียงลำดับ	23
2.10	เว็บที่มีโครงสร้างแบบลำดับขั้น	23
2.11	เว็บที่มีโครงสร้างแบบตาราง	24
2.12	เว็บที่มีโครงสร้างแบบใยแมงมุม	25
2.13	Tools Data Visualization	29
2.14	การนำเสนอแบบทิศทางหรือแนวโน้ม (Trending)	29
2.15	การนำเสนอแบบกลุ่มข้อมูล (Classification)	30
2.16	การนำเสนอเชิงเปรียบเทียบข้อมูล (Comparison)	30
2.17	การนำเสนอรูปแบบแผนที่ (Geographical)	31
2.18	กลุ่มที่ต้องพยากรณ์ล่วงหน้าและการวิเคราะห์ข้อมูลที่ซับซ้อน (Analytics)	31
2.19	Confusion Matrix)	35
2.20	การแสดงแดชบอร์ดของ Tableau Public	36
3.1	กระบวนการวิเคราะห์ข้อมูล ด้วย CRISP–DM	43
3.2	เว็บไซต์ data.go.th	44
3.3	ข้อมูลผู้บาดเจ็บและเสียชีวิตในช่วงเทศกาลปีใหม่ประจำปี 2551 – 2558	45
3.4	ข้อมูลผู้บาดเจ็บและเสียชีวิตในช่วงเทศกาลปีใหม่ประจำปี 2551 – 2558	46
3.5	การตรวจสอบข้อมูลที่ซ้ำซ้อนของข้อมูลทั้งหมด	47
3.6	ข้อมูลผิดพลาดจากการใช้รหัสโรงพยาบาลซ้ำกัน	47
3.7	ข้อมูลผิดพลาดจากการใช้รหัสโรงพยาบาลมากกว่าหนึ่งรหัส	47

สารบัญภาพ (ต่อ)

ภาพที่		หน้า
3.8	ข้อมูลผิดพลาดจากการใช้รหัสโรงพยาบาลมากกว่าหนึ่งรหัส	48
3.9	ข้อมูลผิดพลาดของชื่อข้อมูล และรหัสโรงพยาบาลที่ใช้มากกว่าหนึ่งรหัส	48
3.10	ข้อมูลผิดพลาดของชื่อข้อมูล และรหัสโรงพยาบาลที่ใช้มากกว่าหนึ่งรหัส	49
3.11	ข้อมูลผิดพลาดจากการใช้รหัสโรงพยาบาลมากกว่าหนึ่งรหัส	49
3.12	ข้อมูลผิดพลาดของชื่อข้อมูล และรหัสโรงพยาบาลที่ใช้มากกว่าหนึ่งรหัส	49
3.13	ข้อมูลผิดพลาดจากการใช้รหัสโรงพยาบาลซ้ำกัน	50
3.14	ข้อมูลผิดพลาดของช่วงข้อมูลในโรงพยาบาลห้วยทับทัน	50
3.15	ข้อมูลผิดพลาดของผู้ขับขี่	50
3.16	ข้อมูลที่ขัดแย้งกันของคนเดินเท้า	51
3.17	การลบข้อมูลที่ขัดแย้งกัน	51
3.18	ข้อมูลอายุที่เป็น Outlier	51
3.19	การหาค่าเฉลี่ยของอายุในแต่ละปี แทนข้อมูลที่เป็น 0	52
3.20	ชุดข้อมูลที่คัดเลือกมาวิเคราะห์ข้อมูล	53
3.21	แสดงการเปิดโปรแกรม Weka เวอร์ชัน 3.8.4	70
3.22	แสดงการเข้าหน้าจอโปรแกรม Weka เวอร์ชัน 3.8.4	70
3.23	แสดงการนำข้อมูลเข้าในโปรแกรม Weka เวอร์ชัน 3.8.4	71
3.24	หน้าจอแสดงข้อมูลที่นำเข้าโปรแกรม Weka เวอร์ชัน 3.8.4	71
3.25	การเลือกเทคนิคที่ใช้ในการจำแนกข้อมูลแบบ Decision Tree: J48	72
3.26	หน้าจอผลลัพธ์ของโมเดลการจำแนกข้อมูลแบบ Decision Tree: J48	72
3.27	ผลลัพธ์จากการจำแนกกลุ่มแบบ Decision Tree: J48 ในโปรแกรม Weka เวอร์ชัน 3.8.4	73
3.28	รูปแบบแผนภาพโมเดล Graph Decision Tree: J48 ในโปรแกรม Weka เวอร์ชัน 3.8.4	73
3.29	ผลลัพธ์จากการจำแนกกลุ่มแบบ Decision Tree ของโปรแกรม RapidMiner	75
3.30	รูปแบบแผนภาพโมเดล Graph Decision Tree ในโปรแกรม RapidMiner	75
3.31	ผลลัพธ์จากการจำแนกกลุ่มแบบ Decision Tree: J48 ในโปรแกรม Weka เวอร์ชัน 3.8.4	76

สารบัญภาพ (ต่อ)

ภาพที่		หน้า
3.32	กราฟแท่งแสดงข้อมูลการเกิดอุบัติเหตุแยกเป็นจังหวัด	77
3.33	กราฟแท่งแสดงช่วงเวลาในการเกิดอุบัติเหตุ	77
3.34	กราฟแท่งแสดงข้อมูลผู้ประสบอุบัติเหตุแยกเป็นช่วงอายุแบ่งเป็นเพศชายและเพศหญิง	78
3.35	แผนภาพวงกลมแสดงผลข้อมูลผู้ประสบอุบัติ	78
3.36	แผนภาพวงกลมแสดงข้อมูลถนนที่มีการเกิดอุบัติเหตุ	79
3.37	กราฟแท่งแสดงการเกิดอุบัติเหตุบนท้องถนนประเภทต่าง ๆ	79
3.38	แผนภาพแสดงสถานการณ์ประสบอุบัติเหตุของผู้ประสบอุบัติเหตุทั้งหมด	80
3.39	กราฟแท่งแสดงสถานการณ์ประสบอุบัติเหตุของผู้ประสบอุบัติเหตุของเพศชายและเพศหญิง	80
3.40	แผนภาพวงกลมแสดงผลการดื่มแอลกอฮอล์ของผู้ประสบอุบัติเหตุ	81
3.41	หน้าการแสดงผลการดื่มแอลกอฮอล์ของผู้ประสบอุบัติเหตุในแต่ละจังหวัด	81
3.42	แผนภาพแสดงผลยานพาหนะของผู้บาดเจ็บที่ประสบอุบัติเหตุ	82
3.43	แผนภาพแสดงผลยานพาหนะของคู่กรณี	82
3.44	แผนภาพวงกลมแสดงมาตรการในการป้องกันตัวเองระหว่างเดินทางของผู้ประสบอุบัติเหตุ	83
3.45	กราฟแท่งแสดงผลมาตรการในการป้องกันตัวเองระหว่างเดินทาง	83
3.46	กราฟแสดงผลผลการรักษาของผู้ประสบอุบัติเหตุ	84
3.47	แสดงผลสถานะของผู้ประสบอุบัติเหตุ	84
3.48	แดชบอร์ดอุบัติเหตุปี 2551 ในแต่ละจังหวัด	85
3.49	แดชบอร์ดแสดงอุบัติเหตุปี 2551 ยอดผู้ประสบอุบัติเหตุ	85
3.50	แดชบอร์ดแสดงถนนที่เกิดอุบัติเหตุในปี 2551	86
3.51	แดชบอร์ดแสดงยานพาหนะที่ประสบอุบัติเหตุ	86
3.52	แดชบอร์ดแสดงผลการดื่มแอลกอฮอล์	87
3.52	แดชบอร์ดแสดงข้อมูลเกี่ยวกับผู้ประสบอุบัติเหตุปี 2551	87
3.53	Wireframe หน้าของเว็บไซต์ แสดงเมนูต่าง ๆ ของหน้าเว็บ	88
3.54	Wireframe หน้าแสดงแบบฟอร์มสำหรับจำลองการปฏิบัติตัวระหว่างเดินทาง	88

สารบัญภาพ (ต่อ)

ภาพที่		หน้า
3.55	Wireframe หน้าติดต่อเรา	89
3.56	Wireframe หน้าแสดงการวิเคราะห์ข้อมูลโดยการใช้ Decision Tree	89
3.57	Wireframe หน้าแสดงผลข้อมูลแดชบอร์ดผู้ประสบอุบัติเหตุรายปี	90
3.58	Wireframe หน้าแสดงผลสรุปข้อมูลผู้ประสบอุบัติเหตุทั้งหมด	90
3.59	Wireframe หน้าแสดงผลข้อมูลแนวทางในการป้องกันอุบัติเหตุ	91
3.60	Wireframe หน้าแสดงการกรอกแบบสอบถาม	91
4.1	หน้าแรกของเว็บไซต์ เมื่อเข้าสู่เว็บไซต์การวิเคราะห์ข้อมูลอุบัติเหตุ	94
4.2	เมนูแสดงการวิเคราะห์ข้อมูลด้วย Decision Tree	94
4.3	เมนูแสดงการใช้งานโปรแกรม Weka 3.8.4	95
4.4	เมนูแสดงการใช้งานโปรแกรม Weka 3.8.4	95
4.5	เมนูแสดงผลการวิเคราะห์ข้อมูล	96
4.6	แสดงตารางสถิติข้อมูลประจำปี 2551	96
4.7	แสดงผลยอดข้อมูลแต่ละจังหวัดของผู้ประสบอุบัติเหตุปี 2551	97
4.8	แสดงผลยอดข้อมูลของผู้ประสบอุบัติเหตุปี 2551	97
4.9	แสดงผลยอดข้อมูลถนนที่เกิดอุบัติเหตุปี 2551	98
4.10	แสดงผลยอดข้อมูลยอดยานพาหนะที่ประสบอุบัติเหตุปี 2551	98
4.11	แสดงผลยอดข้อมูลการดื่มแอลกอฮอล์ของผู้ประสบอุบัติเหตุปี 2551	99
4.12	แสดงผลยอดข้อมูลผลการรักษาของผู้ประสบอุบัติเหตุปี 2551	99
4.13	แสดงตารางสถิติข้อมูลประจำปี 2551–2558	100
4.14	แสดงแดชบอร์ดของผู้ประสบอุบัติเหตุปี 2551–2558	100
4.15	แสดงแนวทางป้องกันการเกิดอุบัติเหตุ	101
4.16	แสดงส่วนของการรกรอกแบบสอบถาม	101
4.17	แสดงข้อมูลส่วนตัวของผู้จัดทำ	102
ก.1	แสดงหน้าแรกของเว็บไซต์	113
ก.2	แสดงขั้นตอนการสร้าง Decision Tree	114
ก.3	แสดงการทดสอบกับข้อมูลโดยใช้โปรแกรม Weka 3.8.4	115
ก.4	แสดงผลการวิเคราะห์ข้อมูล	115

สารบัญภาพ (ต่อ)

ภาพที่		หน้า
ก.5	แสดงแบบฟอร์มจำลองการปฏิบัติตัวระหว่างเดินทาง	116
ก.6	แสดงการแจ้งเตือนหมายเหตุจากแบบฟอร์ม	116
ก.7	แสดงผลลัพธ์ของการส่งข้อมูลจากแบบฟอร์ม	117
ก.8	แสดงตารางสถิติข้อมูลผู้บาดเจ็บและเสียชีวิต	117
ก.9	แสดงแดชบอร์ดผลสรุปของข้อมูลแต่ละด้าน	118
ก.10	แดชบอร์ดแสดงยอดผู้ประสบอุบัติเหตุ	119
ก.11	แดชบอร์ดแสดงยอดถนนที่เกิดอุบัติเหตุ	119
ก.12	แดชบอร์ดแสดงยานพาหนะที่เกิดอุบัติเหตุ	120
ก.13	แดชบอร์ดแสดงการดื่มแอลกอฮอล์	120
ก.14	แดชบอดร์ดแสดงยอดผลการรักษา	121
ก.15	แสดงผลของการสรุปข้อมูลโดยรวมจากข้อมูลจำนวน 8 ปี	122
ก.16	แสดงแนวทางการป้องกันการเกิดอุบัติเหตุ	123
ก.17	แบบสอบถามสำหรับผู้เข้าใช้งานเว็บไซต์	123
ก.18	แสดงส่วนติดต่อของคณะผู้จัดทำ	124

สารบัญตาราง

ตารางที่		หน้า
1.1	ระยะเวลาในการดำเนินการ	6
3.1	สมการการหาความสัมพันธ์ของแอตทริบิวต์	53
ข.1	แสดงข้อมูลแบบสอบถาม	126

บทที่ 1
บทนำ

โครงงานเรื่อง การวิเคราะห์ข้อมูลผู้บาดเจ็บและเสียชีวิตในช่วงเทศกาลปีใหม่ประจำปี 2551 – 2558 ในบทนี้จะเป็นการกล่าวถึงความสำคัญ และที่มาของปัญหา วัตถุประสงค์ ประโยชน์ที่จะได้รับ ขอบเขตงานของคณะผู้จัดทำและผู้ใช้งานทั่วไปบนเว็บไซต์ เครื่องมือที่ใช้ในการพัฒนาเว็บไซต์ สถานที่ใช้ในการดำเนินการศึกษาและรวบรวมข้อมูล และระยะเวลาในการดำเนินการ รวมถึงนิยามศัพท์เฉพาะของโครงงาน ดังนี้

1.1 ความสำคัญและที่มาของปัญหา

ในยุคปัจจุบันเทคโนโลยีดิจิตอลเข้ามาเป็นส่วนหนึ่งในการดำรงชีวิตประจำวันของคนเราเป็นอย่างมาก หน่วยงานต่าง ๆ ไม่ว่าจะเป็นภาคธุรกิจ องค์กรต่าง ๆ รวมถึงหน่วยงานภาครัฐเอง ต่างก็มีการนำเทคโนโลยีต่าง ๆ เข้ามาช่วยเพื่อปรับปรุงประสิทธิภาพการดำเนินงานและยกระดับการให้บริการไปสู่ประชาชนให้สามารถเข้าถึงบริการต่าง ๆ ได้สะดวก และรวดเร็วยิ่งขึ้น จึงทำให้เกิดข้อมูลสารสนเทศขึ้นอย่างมากมายมหาศาล โดยเฉพาะอย่างยิ่งข้อมูลของภาครัฐที่จัดเก็บในรูปแบบของฐานข้อมูลนับเป็นเวลาหลายสิบปี และกำลังเพิ่มขึ้นอีกในปัจจุบันและอนาคต ข้อมูลเหล่านี้สามารถนำมาใช้ประโยชน์ได้ทั้งสำหรับภาครัฐ ภาคเอกชน และประชาชนซึ่งทำให้เกิดเป็นแนวคิด วิธีการ รูปแบบทางธุรกิจ รวมถึงวิถีชีวิตใหม่ ๆ การแบ่งปันข้อมูลเพื่อนำมาใช้ประโยชน์ และสร้างนวัตกรรมนั้น ซึ่งถือเป็นสัญญาณเริ่มต้นของการขับเคลื่อนเศรษฐกิจด้วยข้อมูล อันจะช่วยให้เศรษฐกิจของประเทศเติบโตขึ้น และเพิ่มขีดความสามารถในการแข่งขันในเวทีโลกได้ ที่ผ่านมาหลาย ๆ ประเทศมีการผลักดัน และดำเนินการเกี่ยวกับการเปิดเผยข้อมูลสู่สาธารณะกันอย่างจริงจัง เช่นเดียวกันกับในประเทศไทย หลายๆ หน่วยงานก็เริ่มเห็นถึงความสำคัญ และประโยชน์ของการเปิดเผยข้อมูลกันมากขึ้น ดังนั้นสนักงานรัฐบาลอิเล็กทรอนิกส์ (องค์การมหาชน) (สรอ.) ซึ่งมีภารกิจหลักในการพัฒนารัฐบาลอิเล็กทรอนิกส์ของประเทศให้มีคุณภาพ จึงได้พัฒนาระบบศูนย์กลางข้อมูลภาครัฐ หรือ data.go.th เพื่อเป็นการส่งเสริมและผลักดันให้เกิดการเปิดเผยข้อมูลภาครัฐสู่สาธารณะอย่างเป็นรูปธรรม เป็นช่องทางให้ผู้ใช้บริการทั้งภาคประชาชน ภาคธุรกิจเอกชน รวมถึงหน่วยงานของรัฐ สามารถค้นหาและเข้าถึงข้อมูลที่มีคุณภาพของภาครัฐได้ง่าย โดยมีเป้าหมายสูงสุดเพื่อส่งเสริมให้ เกิดธรรมาภิบาล ส่งเสริมให้เกิดความร่วมมือของภาครัฐ และสร้างการมีส่วนร่วม

ของภาคประชาชน รวมถึงแก้ปัญหาการ ทุจริตคอร์รัปชั่น การเมือง และเศรษฐกิจอย่างบูรณาการร่วมกันได้การวิเคราะห์ข้อมูลผู้บาดเจ็บและเสียชีวิตในช่วงเทศกาลปีใหม่ คือการนำข้อมูลที่ได้ถูกรวบรวมไว้และถูกเผยแพร่ผ่านเว็บไซต์ Open Government Data of Thailand ซึ่งเป็นเว็บไซต์ที่เก็บรวบรวมชุดข้อมูลต่าง ๆ ของประเทศไทยซึ่งเป็นข้อมูลที่เปิดเผยได้ เพื่อให้ผู้ใช้บริการสามารถนำชุดข้อมูลไปศึกษาหรือวิเคราะห์ให้เกิดประโยชน์ต่อไปได้ การวิเคราะห์ข้อมูลปริมาณมหาศาลแสดงให้เห็นประโยชน์ต่าง ๆ ก็คือ ความรวดเร็วและประสิทธิภาพ ขณะที่หลายปีก่อนหน้านั้นหลายธุรกิจต้องรวบรวมข้อมูล ดำเนินการวิเคราะห์ และค้นหาข้อมูลที่สามารถนำมาใช้ประกอบการตัดสินใจในอนาคตได้ ส่วนในทุกวันนี้ธุรกิจสามารถระบุข้อมูลเชิงลึกสำหรับการกระทำที่ต้องการการตัดสินใจอย่างทันท่วงที ความสามารถในการทำงานได้เร็วขึ้น และยังคงไว้ซึ่งความคล่องตัว ช่วยให้หลายองค์กรสร้างความได้เปรียบด้านการแข่งขันอย่างที่ไม่เคยมีมาก่อน ซึ่งการนำข้อมูลผู้บาดเจ็บและเสียชีวิตในช่วงเทศกาลปีใหม่ประจำปี 2551 – 2558 มาทำการวิเคราะห์นั้น เนื่องจากต้องการทราบถึงข้อมูลในเชิงลึกของสาเหตุการบาดเจ็บและเสียชีวิต เช่น ปัจจัยที่นำมาซึ่งการเกิดอุบัติเหตุ อาทิเช่น การประพฤติตัวของผู้ประสบอุบัติเหตุก่อนเกิดอุบัติเหตุ และปัจจัยอื่น ๆ ที่นำไปสู่การบาดเจ็บและเสียชีวิต เพื่อหาทางป้องกันและลดอัตราการบาดเจ็บและเสียชีวิต ซึ่งการวิเคราะห์ปัจจัยต่าง ๆ จากข้อมูลที่ได้มานั้นถือเป็นเรื่องสำคัญต่อการป้องกันภัยจากอุบัติเหตุต่าง ๆ ในช่วงเทศกาล แต่ด้วยจำนวนข้อมูลที่มีปริมาณมหาศาลทำให้ยากต่อการนำเสนอให้เข้าใจได้ง่าย ซึ่งถือเป็นข้อจำกัดหากต้องการนำข้อมูลชุดนี้ไปใช้เพื่อเผยแพร่ให้แก่บุคคลภายนอกหรือกลุ่มผู้ใช้ข้อมูลได้รับรู้

จากปัญหาข้างต้นดังนั้นคณะผู้จัดทำจึงได้นำชุดข้อมูลที่ได้มานั้นเข้าสู่กระบวนการจำแนกตามรายการชุดข้อมูล โดยการใช้ขีดความสามารถในการประมวลผลของโปรแกรมคอมพิวเตอร์เพื่อเพิ่มความน่าเชื่อถือในการวิเคราะห์ และประมวลผลข้อมูลให้เหมาะสมตรงกับความต้องการ เป็นการแยกแยะสิ่งที่จะพิจารณาออกเป็นส่วนย่อยที่มีความสัมพันธ์กัน เพื่อทำความเข้าใจแต่ละส่วนให้ชัดเจน รวมทั้งการสืบค้นความสัมพันธ์ของส่วนต่าง ๆ เพื่อดูว่าส่วนประกอบปลีกย่อยนั้นสามารถเข้ากันได้หรือไม่ มีความสัมพันธ์เกี่ยวเนื่องกันอย่างไร โดยคณะผู้จัดทำได้ทำการจำแนกข้อมูลออกเป็นประเภทต่าง ๆ ด้วยเทคนิคการทำเหมืองข้อมูลแบบ Classification เป็นกระบวนการสร้างโมเดลจัดการข้อมูลให้อยู่ในกลุ่มที่กำหนดมาให้ โดยอาศัยการเรียนรู้ข้อมูลเก่าในช่วงเวลาที่ผ่านมาในรูปแบบ Descriptive Analytics คือข้อมูลเหตุการณ์ที่เกิดขึ้นแล้ว เป็นการวิเคราะห์ข้อมูลแบบพื้นฐาน โดยจะเน้นไปที่ความสามารถในการอธิบายว่าเกิดอะไรขึ้นกับเหตุการณ์ และสามารถอธิบายได้ว่าทำไมถึงเกิดขึ้น โดย Classification สร้างกฎเพื่อช่วยในการตัดสินใจจากข้อมูลที่มีอยู่ เพื่อใช้ทำนายแนวโน้มการเกิดขึ้นของปัจจัยที่ก่อให้เกิด

ผู้บาดเจ็บและเสียชีวิตจากเทคนิคการจำแนกประเภทข้อมูล ด้วยการวิเคราะห์ในรูปแบบของแผนภูมิต้นไม้ หรือที่เรียกว่า Decision tree เป็นโครงสร้างที่ใช้แสดงกฎที่ได้จากเทคนิคการจำแนกประเภทข้อมูล โดย Decision tree แต่ละโหนดแสดงคุณลักษณะ (attribute) ส่วนโหนดลูกแสดงเงื่อนไขในการทดสอบ และโหนดปลาย (leaf node) แสดงกลุ่มที่กำหนดไว้ ซึ่งการจำแนกข้อมูลออกเป็นประเภทต่าง ๆ นี้จะช่วยให้เกิดความเข้าใจต่อข้อมูลและสิ่งที่ต้องการทราบอย่างแท้จริง และทางคณะผู้จัดทำยังได้ทำการนำเสนอข้อมูลแบบ visualization ด้วยการแสดงผลข้อมูลในรูปแบบของภาพด้วยโปรแกรม Tableau Public และเผยแพร่ข้อมูลสารสนเทศนี้บน Web browser ที่เป็นที่นิยมในยุคอินเตอร์เน็ตคือการเผยแพร่ทางสื่อออนไลน์ โดยใช้ภาษา HTML ในการพัฒนาหน้าเว็บไซต์สำหรับการเปิดเผยข้อมูลภาครัฐที่ได้มาจาก data.go.th หรือใน Open Government Data of Thailand เพื่อให้ผู้ใช้งานสามารถเข้าถึงข้อมูลได้อย่างเหมาะสม ซึ่งประโยชน์ที่ผู้ใช้งานทั่วไปจะได้รับจะเป็นในเรื่องของความเข้าใจ การรับรู้ และการตระหนักถึงความปลอดภัยในเหตุการณ์ต่าง ๆ สำหรับผู้ที่ทำการค้นหาและศึกษาข้อมูลก็สามารถดำเนินการผ่าน Web browser ที่ได้เผยแพร่ข้อมูลสารสนเทศของผู้บาดเจ็บและเสียชีวิตในเทศกาลปีใหม่ประจำปี 2551 – 2558 ที่ผ่านการสรุปจำแนกข้อมูลตามประเภทต่าง ๆ ได้สะดวกรวดเร็วยิ่งขึ้น รวมถึงมีความถูกต้อง สะดวก ลดความซ้ำซ้อน เพิ่มประสิทธิภาพให้กับการศึกษาค้นคว้า และทันต่อเวลา

1.2 วัตถุประสงค์

เพื่อวิเคราะห์ข้อมูลผู้บาดเจ็บและเสียชีวิตในเทศกาลปีใหม่ประจำปี 2551 – 2558 สำหรับเผยแพร่บนเว็บไซต์

1.3 ประโยชน์ที่จะได้รับ

ได้ข้อมูลสารสนเทศของผู้บาดเจ็บและเสียชีวิตในเทศกาลปีใหม่ประจำปี 2551 – 2558 สำหรับเผยแพร่บนเว็บไซต์

1.4 ขอบเขต

1.4.1 ขอบเขตคณะผู้จัดทำ

กระบวนการวิเคราะห์ข้อมูลด้วย CRISP–DM

1.4.1.1 Business Understanding คณะผู้จัดทำทำความเข้าใจกับปัญหาให้อยู่ในรูปของการวิเคราะห์ข้อมูลทาง Data Mining โดยการวิเคราะห์ข้อมูลในประเด็นนี้ คือ ข้อมูลผู้บาดเจ็บและเสียชีวิตในช่วงเทศกาลปีใหม่ประจำปี 2551 – 2558 ซึ่งมีจำนวนข้อมูลมหาศาล

ทำให้ไม่สามารถทำความเข้าใจกับข้อมูลที่มีความซับซ้อนได้อย่างรวดเร็ว เช่น ต้องการทราบว่าช่วงอายุใดมีอัตราการเกิดอุบัติเหตุสูงสุด

1.4.1.2 Data Understanding คณะผู้จัดทำทำการรวบรวมข้อมูล เพื่อตรวจสอบรายละเอียด ปริมาณ และความน่าเชื่อถือของข้อมูลผู้บาดเจ็บและเสียชีวิตในช่วงเทศกาลปีใหม่ประจำปี 2551 – 2558 ที่ได้จากเว็บไซต์ data.go.th

1.4.1.3 Data Preparation คณะผู้จัดทำทำการคัดเลือกข้อมูล และทำการ Data Cleaning ข้อมูลเพื่อให้ได้ข้อมูลที่สมบูรณ์ ด้วยการตัดข้อมูลที่ไม่จำเป็นออกไป เช่น รหัสจังหวัด รหัสโรงพยาบาล เพื่อเหลือข้อมูลที่จำเป็นในนำไปวิเคราะห์ข้อมูล

1.4.1.4 Modeling คณะผู้จัดทำวิเคราะห์ข้อมูลด้วยการทำเหมืองข้อมูลแบบ Data Classification เพื่อใช้ทำนายแนวโน้มการเกิดขึ้นของปัจจัยที่ก่อให้เกิดผู้บาดเจ็บและเสียชีวิตจากเทคนิคการจำแนกประเภทข้อมูล ด้วยการสร้างโมเดล Decision Tree เพื่อจัดกลุ่มข้อมูลที่มีความสัมพันธ์กัน โดยใช้โปรแกรมที่ใช้ทำเหมืองข้อมูล

1.4.1.5 Evaluation คณะผู้จัดทำได้ข้อมูลที่ผ่านการวิเคราะห์ข้อมูลด้วยเทคนิค Data Mining จากการสร้างโมเดล Decision Tree และการสรุปผลข้อมูลเพื่อให้ตรงกับวัตถุประสงค์ที่ตั้งไว้

1.4.1.6 Deployment คณะผู้จัดทำนำผลข้อมูลที่ทำการวิเคราะห์แล้ว แสดงผลข้อมูลบน Web browser โดยใช้ชุดคำสั่ง HTML และ CSS3 ร่วมกับการนำเสนอข้อมูลแบบ visualization ด้วยการแสดงผลข้อมูลในรูปแบบของภาพโดยใช้โปรแกรม Tableau Public

1.4.2 ขอบเขตผู้ใช้งานทั่วไปบนเว็บบราวเซอร์

1.4.2.1 สามารถดูข้อมูลสารสนเทศของผู้บาดเจ็บและเสียชีวิตในช่วงเทศกาลปีใหม่ประจำปี 2551 – 2558 ที่เผยแพร่ผ่านทางเว็บไซต์

1.4.2.2 สามารถดูข้อมูลสารสนเทศในลักษณะตารางสรุปข้อมูลสถิติผู้บาดเจ็บและเสียชีวิต

1.4.2.3 สามารถดูข้อมูลสารสนเทศในลักษณะรูปแบบของแผนภูมิชนิดต่าง ๆ จากโปรแกรม Tableau ได้

1.4.2.4 สามารถกรองข้อมูลสารสนเทศผ่านระบบของ Tableau บน Web browser ได้

1.5 เครื่องมือที่ใช้ในการพัฒนาโปรแกรม

1.5.1 Hardware

1.5.1.1 Notebook: ASUS K556U core i7 Ram 8 GB

1.5.1.2 Notebook: LENOVO G470 core i3 Ram 12 GB

1.5.2 Software

1.5.2.1 โปรแกรม Tableau Public ใช้ในการแสดงผลข้อมูลในรูปแบบของภาพ

1.5.2.2 โปรแกรม Weka 3.8.4 ใช้ในการวิเคราะห์ข้อมูลด้วยเทคนิคเหมืองข้อมูล

1.5.2.3 โปรแกรม RapidMiner Studio 9.5.1 ใช้ในการทำเทคนิคเหมืองข้อมูล

1.5.2.4 ชุดคำสั่งภาษา HTML4, HTML5 (Hyper Text Markup Language)

1.5.2.5 ชุดคำสั่งภาษา PHP (PHP Hypertext Preprocessor)

1.5.2.6 ชุดคำสั่ง CSS และ Bootstrap ที่ใช้ในการออกแบบเว็บ

1.5.2.7 โปรแกรม Xampp 5.5.38 ใช้ในการจำลอง web server

1.5.2.8 โปรแกรม FileZilla 3.40.0 โปรแกรมที่ใช้ติดต่อกับ FTP Server เพื่ออัปโหลด (Upload) ไฟล์ (File)

1.5.2.9 โปรแกรม Atom ใช้ในการเขียนคำสั่งและเขียนโปรแกรม

1.5.2.10 โปรแกรม Adobe XD ใช้ในการออกแบบหน้าเว็บไซต์

1.5.2.11 ระบบปฏิบัติการ : Window 10 ใช้ในการเปิดซอฟต์แวร์ต่าง ๆ

1.5.2.12 Microsoft office Word 2016 ใช้ในการทำเอกสารต่าง ๆ

1.5.2.13 Microsoft office Excel 2016 ใช้ในการ Clean Data กับชุดข้อมูล

1.6 สถานที่ใช้ในการดำเนินการศึกษาและรวบรวมข้อมูล

1.6.1 สถานที่ใช้ในการดำเนินการศึกษา

มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนาภาคพายัพเชียงใหม่ 128 ถนนห้วยแก้ว ตำบลช้างเผือก อำเภอเมือง จังหวัดเชียงใหม่ 50300

1.6.2 แหล่งรวบรวมข้อมูล

เว็บไซต์ Open Government Data of Thailand หรือ data.go.th

1.7 ระยะเวลาในการดำเนินการ

ตารางที่ 1.1 ระยะเวลาในการดำเนินการ

แผนการดำเนินการ	2563				
	ม.ค.	**ก.พ.**	**มี.ค.**	**เม.ย.**	**พ.ค.**
1. ศึกษาและกำหนดความต้องการ	←	→			
2. ตรวจสอบความถูกต้องข้อมูล		←	→		
3. วิเคราะห์ข้อมูลจำแนกตามกลุ่ม		←	→		
4. ออกแบบระบบ			←	→	
5. เขียนและทดสอบระบบ			←	→	
6. ตรวจสอบระบบโดยรวม				←	→
7. ประเมินการใช้งานระบบ				←	→
8. จัดทำเอกสารประกอบโครงงาน	←				→

1.8 นิยามศัพท์เฉพาะ

1.8.1 การวิเคราะห์ข้อมูล หมายความว่า เป็นการจัดระเบียบแยกแยะส่วนต่าง ๆ ของหลักฐาน หรือข้อมูลที่ได้ออกเป็นหมวดหมู่ เพื่อหาคำตอบตามความมุ่งหมาย และตามสมมติฐานที่ได้กำหนดไว้ การวิเคราะห์ข้อมูลนี้เป็นขั้นการทำงานที่ต่อเนื่องมาจากการวัด การนับ และจัดเรียงลำดับข้อมูล ส่วนใหญ่เป็นเรื่องเกี่ยวกับการนำเอาวิธีการทางสถิติมาวิเคราะห์หาค่าตัวแปรหรือหาลักษณะของตัวแปร

1.8.2 เทศกาล หมายถึง เหตุการณ์ชนิดหนึ่ง ซึ่งตามปกติธรรมดาจัดตั้งขึ้นโดยชุมชนท้องถิ่น ที่มุ่งความสนใจและเฉลิมฉลองเอกลักษณ์บางอย่างของชุมชนนั้นและเทศกาลนั้น เทศกาลมักเกี่ยวข้องกับประเพณี ความเชื่อ หรือศาสนาของชุมชน เทศกาลส่วนใหญ่มักจัดขึ้นปีละครั้ง คือการถือเอาวันครบรอบปีของเทศกาลครั้งก่อนมาตั้งเป็นครั้งถัดไป

1.8.3 เว็บไซต์ หมายความว่า สื่อนำเสนอข้อมูลบนเครื่องคอมพิวเตอร์ หรือคือการรวบรวม หน้าเว็บเพจหลายหน้า ซึ่งเชื่อมโยงกันผ่านทางไฮเปอร์ลิงก์ ซึ่งต้องเปิดด้วยโปรแกรมเฉพาะทางที่เรียกว่า Web Browser โดยถูกจัดเก็บไว้ในเวิลด์ไวด์เว็บ และเว็บไซต์นั้นถูกสร้างขึ้นด้วยภาษาทางคอมพิวเตอร์ที่เรียกว่า HTML (Hyper Text Markup Language)

1.9 บทสรุป

จากบทนำที่ได้กล่าวมาในข้างต้นทั้งหมดนั้น คณะผู้จัดทำได้เล็งเห็นความสำคัญของการนำเทคโนโลยีมาประยุกต์ใช้ในการวิเคราะห์ข้อมูลผู้บาดเจ็บและเสียชีวิตในช่วงเทศกาลปีใหม่ประจำปี 2551 – 2558 ด้วยกระบวนการวิเคราะห์ข้อมูลด้วย CRISP-DM จากเทคนิคการทำเหมืองข้อมูลแบบ Classification ในรูปแบบของแผนภูมิต้นไม้ หรือที่เรียกว่า Decision tree และเผยแพร่ข้อมูลสารสนเทศบน web browser ที่เป็นที่นิยมในยุคอินเตอร์เน็ตคือการเผยแพร่ทางสื่อออนไลน์ โดยใช้ภาษา HTML ในการพัฒนาหน้าเว็บไซต์สำหรับการเปิดเผยข้อมูลภาครัฐที่ได้มาจาก data.go.th หรือใน Government Open Data of Thailand เนื่องจากเทคโนโลยีปัจจุบันเข้ามาช่วยในการจัดการข้อมูล หรือดำเนินงานต่าง ๆ ให้มีความสะดวกสบายเป็นอย่างมาก อีกทั้งผู้ใช้งานยังเข้าถึงข้อมูลได้ง่าย และมีประสิทธิภาพ

บทที่ 2
แนวคิด ทฤษฎี เครื่องมือและวรรณกรรมที่เกี่ยวข้อง

โครงงานเรื่อง การวิเคราะห์ข้อมูลผู้บาดเจ็บและเสียชีวิตในช่วงเทศกาลปีใหม่ประจำปี 2551 – 2558 ในบทนี้เป็นการนำเสนอเกี่ยวกับ แนวคิด ทฤษฎี เครื่องมือและวรรณกรรมที่เกี่ยวข้องของการวิเคราะห์ข้อมูลผู้บาดเจ็บและเสียชีวิตในช่วงเทศกาลปีใหม่ประจำปี 2551 – 2558 และการแสดงผลข้อมูลบนเว็บไซต์ ซึ่งได้รวบรวมการศึกษาเอกสารงานวิจัยที่เกี่ยวข้องกับการวิเคราะห์ข้อมูล เพื่อใช้เป็นแนวทางการศึกษาประกอบด้วยรายละเอียดตามลำดับ ดังนี้

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

2.3 เครื่องมือในการออกแบบและวิเคราะห์ข้อมูล

- 2.3.1 แบบจำลองต้นไม้ตัดสินใจ (Decision Tree)
- 2.3.2 การประเมินผลโมเดล (Decision Tree)
- 2.3.3 การแสดงผลแบบ visualization

2.4 วรรณกรรมที่เกี่ยวข้อง

2.5 บทสรุป

2.1 แนวคิด

2.2.1 แนวคิดเกี่ยวกับการวิเคราะห์ข้อมูล (Data analytic)

ในการดำเนินงานเรื่องการวิเคราะห์ข้อมูลผู้บาดเจ็บและเสียชีวิตในช่วงเทศกาลปีใหม่ประจำปี 2551 – 2558 และการแสดงผลข้อมูลบนเว็บไซต์ คณะผู้จัดทำได้ศึกษาหลัก การและทฤษฎีต่าง ๆ องค์ประกอบหนึ่งที่สำคัญคือการวิเคราะห์ข้อมูล ซึ่งมีรายละเอียด ดังนี้

ความหมายของการวิเคราะห์ข้อมูล

เอื้อมพร หลินเจริญ (2555) ได้ให้ความหมายการวิเคราะห์ข้อมูลเชิงคุณภาพไว้ว่า การวิเคราะห์ข้อมูลเชิงคุณภาพ นับเป็นขั้นตอนที่สำคัญในกระบวนการวิจัย วิธีการหลักที่ใช้ในการ วิเคราะห์ข้อมูลเชิงคุณภาพ เป็นวิธีการสร้างข้อสรุปจากการศึกษาจากข้อมูลจำนวนหนึ่ง ซึ่งมักไม่ใช้สถิติในการวิเคราะห์หรือถ้าใช้สถิติก็ไม่ได้ถือว่าสถิติเป็นวิธีการวิเคราะห์หลัก แต่จะถือเป็นข้อมูลเสริมในการวิเคราะห์ ข้อมูลเชิงคุณภาพนั้น คณะผู้จัดทำนับเป็นผู้ที่มีบทบาทสำคัญยิ่งและควรมีความรอบรู้ในเรื่องแนวคิดทฤษฎีอย่างกว้างขวาง มีความเป็นสหวิทยากรอยู่ในตัวเอง มีความสามารถทางภาษาสามารถเชื่อมโยงข้อความและสร้างข้อสรุปเป็นกรอบแนวคิดและตีความหมายของข้อมูลได้หลาย ๆ แบบ ทั้งนี้การวิเคราะห์ข้อมูลเชิงคุณภาพนั้นอาจไม่จำเป็นต้องใช้เฉพาะการวิจัยเชิงคุณภาพเท่านั้น แต่อาจใช้กับการวิจัยเชิงปริมาณที่มีการเก็บรวบรวมข้อมูลเชิงคุณภาพ เช่น แบบสอบถามปลายเปิด การสัมภาษณ์ การสังเกต จดบันทึกมาทำการวิเคราะห์เพื่อให้ข้อมูลที่รวบรวมมามีความหมายและตอบคำถามหรือจุดมุ่งหมายของการวิจัยสำหรับสาระในการวิเคราะห์ข้อมูลเชิงคุณภาพ โดยเทคนิคการวิเคราะห์ข้อมูลเชิงคุณภาพที่สำคัญ ๆ และนักวิจัย นิยมใช้ ได้แก่ การจำแนกหรือการจัดกลุ่มข้อมูล การเปรียบเทียบเหตุการณ์ การวิเคราะห์ส่วนประกอบ การวิเคราะห์แบบอุปนัย และการวิเคราะห์ข้อมูลที่เป็นเอกสาร

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

ของพฤติกรรมมากกว่าที่จะมุ่งหาระดับความมากน้อยหรือระดับความเข้มข้นของความสัมพันธ์ของปรากฏการณ์จึงเป็นการวิเคราะห์ข้อมูลที่อาศัยเครื่องมือทางความคิด (Conceptual tools) มากกว่าที่จะใช้เครื่องมือทางสถิติ และที่สำคัญต้องอาศัยความสามารถของนักวิจัยที่จะอธิบายให้เห็นความสัมพันธ์และความหมายของปรากฏการณ์บนพื้นฐานความเข้าใจในวัฒนธรรมและสังคมที่เป็นบริบทของปรากฏการณ์ที่ศึกษา

จุมพล หนิมพานิช และวรวลัญช์ โรจนพล (2561) ได้ให้ความหมายการวิเคราะห์ข้อมูลเชิงคุณภาพไว้ว่า การวิเคราะห์ข้อมูลเชิงคุณภาพโดยทั่วไปหมายถึง การสรุปเนื้อหาสาระของข้อมูล และหาแบบแผนของความสัมพันธ์ระหว่างข้อมูลหรือตัวแปร หรือหมายถึงการสรุปพรรณนาและการวิเคราะห์ความสัมพันธ์ระหว่างข้อมูล หรือหมายถึง การนำเอาข้อมูลที่ได้จากการค้นคว้าวิจัยมาจัดกระทำให้เป็นระบบและหาความหมาย แยกแยะ องค์ประกอบ รวมทั้งเชื่อมโยงและหาความสัมพันธ์ของข้อมูลเพื่อให้สามารถนาไปสู่ความเข้าใจต่อการดำรงอยู่และการเปลี่ยนแปลงของปรากฏการณ์ที่ศึกษา หรือหมายถึง การแยกแยะ การตีความหมาย การเปรียบเทียบ การหาความสัมพันธ์เกี่ยวกับปรากฏการณ์ การหาแบบแผนการอธิบายและการสรุปเกี่ยวกับกิจกรรมหรือพฤติกรรมในสภาพสังคมหรือปรากฏการณ์ทางสังคมที่ศึกษา

2.1.2 แนวคิดเกี่ยวกับการประยุกต์ใช้ข้อมูลทางสถิติ

ความหมายของการประยุกต์ใช้ข้อมูลทางสถิติ

จากเอกสารประกอบการสอนการวิจัยสำหรับครู ได้ให้ความหมายเกี่ยวกับสถิติไว้ว่า คำว่าสถิติ (Statistics) มาจากภาษาเยอรมันว่า Statistics มีรากศัพท์มาจาก Stat หมายถึง ข้อมูลหรือสารสนเทศ ซึ่งจะอำนวยประโยชน์ต่อการบริหารประเทศในด้านต่าง ๆ เช่น การทำสำมะโนครัวเพื่อจะทราบจำนวนพลเมืองในประเทศทั้งหมด ในสมัยต่อมา คำว่า สถิติ ได้หมายถึง ตัวเลขหรือข้อมูลที่ได้จากการเก็บรวบรวม เช่น จำนวนผู้ประสบอุบัติเหตุบนท้องถนน อัตราการเกิดของเด็กทารก ปริมาณน้ำฝนในแต่ละปี เป็นต้น สถิติในความหมายที่กล่าวมานี้เรียกอีกอย่างหนึ่งว่า 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 ที่ได้มีการรวบรวมขึ้นเพื่อบอกข้อเท็จจริงเกี่ยวกับคุณสมบัติหรือลักษณะบางสิ่งบางอย่างที่สามารถแสดงออกเป็นตัวเลขได้สถิติแบ่งเป็นประเภทใหญ่ ๆ ได้ 2 ประเภท ดังนี้ คือ

1) สถิติพรรณนา (Descriptive statistics) คือสถิติที่บรรยายถึงลักษณะของข้อมูลเฉพาะกลุ่มนั้น ๆ โดยไม่สรุปอ้างอิงไปยังประชากรกลุ่มอื่น ๆ สถิติประเภทนี้นิยมศึกษาในกลุ่มเล็กหรือกลุ่มใหญ่ก็ได้ สถิติประเภทนี้เป็นสถิติที่บรรยายลักษณะของข้อมูล เช่น ค่าเฉลี่ย ร้อยละ มัธยฐาน พิสัย ส่วนเบี่ยงเบนมาตรฐาน และการวัดความสัมพันธ์ต่าง ๆ เช่น สหสัมพันธ์

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

สุทิน ชนะบุญ (2555) ได้ให้ความหมายเกี่ยวกับสถิติไว้ว่า สถิติประกอบด้วย 2 องค์ประกอบ ดังนี้ สถิติเชิงพรรณนา (Descriptive Statistics) เป็นวิธีการทางสถิติเพื่อใช้ในการพรรณนาหรือบรรยายลักษณะของสิ่งที่ศึกษา เพื่อให้เข้าใจถึงลักษณะของข้อมูลที่เก็บรวบรวมมาได้ จะพรรณนาภายในขอบเขตของข้อมูลที่เก็บรวบรวมมาเท่านั้น ไม่สามารถจะคาดคะเนลักษณะต่าง ๆ ออกไปนอกเหนือจากข้อมูลที่มีอยู่ได้ หรือไม่มีการอ้างอิงหรืออนุมานไปถึงกลุ่มอื่น หากผู้วิจัยสามารถศึกษาทุกหน่วยของประชากรได้ ก็จะใช้สรุปหรือบรรยายลักษณะของประชากรที่ศึกษา หากผู้วิจัยไม่สามารถศึกษาสมาชิกทุกหน่วยของประชากรได้ สุ่มสมาชิกเพียงบางส่วนมาศึกษา(ศึกษาจากกลุ่มตัวอย่าง) ก็จะใช้สรุปหรือบรรยายเฉพาะกลุ่มตัวอย่างที่สุ่มมาศึกษาเท่านั้น สถิติเชิงอ้างอิงหรือสถิติอนุมาน (Inferential Statistics) เป็นวิธีการทางสถิติที่ใช้ในการสรุปลักษณะของประชากร จากผลการศึกษาข้อมูลในกลุ่ม ตัวอย่าง โดยอาศัย

ทฤษฎีความน่าจะเป็นในการอนุมานลักษณะประชากร โดยศึกษาจากกลุ่มตัวอย่าง แต่ อ้างอิงหรืออนุมานไปถึงประชากร ประกอบด้วย 2 องค์ประกอบ คือ การประมาณค่าพารามิเตอร์ (Parametric Estimation) และการทดสอบสมมติฐาน (Testing Hypothesis)

2.1.3 แนวคิดเกี่ยวกับการแสดงผลข้อมูล (Data visualization)

ความหมายการแสดงผลข้อมูล

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

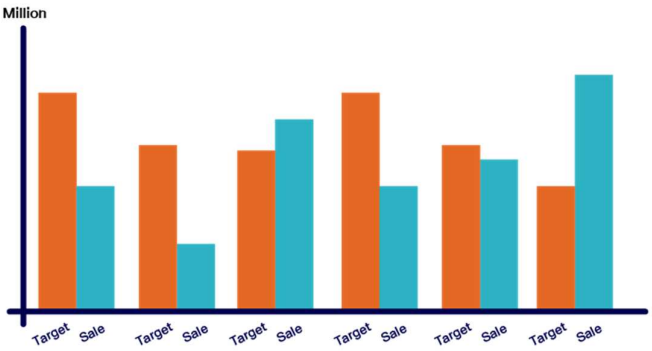
ชนาธิป ชื่นมนัส (2552) ได้ให้ความหมายของการแสดงผลข้อมูลไว้ว่า การแสดงผลข้อมูล (Data Visualization) คือวิธีการที่ใช้ในการนำข้อมูลที่เป็นนามธรรมมาแสดงให้เป็นภาพ เพื่อเพิ่มประสิทธิผลในการติดต่อสื่อสาร หรือทำให้เข้าใจข้อมูลได้ง่ายขึ้น การใช้เทคนิค Visualization มีหลายวิธี ซึ่ง แต่ละวิธีเหมาะสมที่จะใช้ในการแสดงข้อมูลที่แตกต่างกัน การเลือกใช้เทคนิคใดนั้นขึ้นกับองค์ประกอบหลายอย่าง เช่น ชนิดของข้อมูลที่จะนำมาแสดง วิธีที่ต้องการแสดงข้อมูล เป็นต้น ในปัจจุบันมีงานวิจัยที่นำเอาเทคนิค Visualization มาใช้โดยมีจุดประสงค์เพื่อเพิ่มประสิทธิผลในการแสดงข้อมูลการติดต่อสื่อสารภายในเครือข่ายคอมพิวเตอร์ และช่วยในการวิเคราะห์ความผิดปกติที่เกิดขึ้นในเครือข่าย

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

การเลือกรูปแบบ Visualization ให้เหมาะสมกับข้อมูล ในปัจจุบันเป็นยุคเทคโนโลยีเข้าถึงทุกคน ทำให้การรับรู้ข่าวสาร ข้อมูลต่าง ๆ เป็นไปได้ง่าย และรวดเร็วมากขึ้น คนที่นำเสนอข้อมูลจึงต้องนำเสนอข้อมูลให้น่าสนใจ เข้าใจง่าย และรวดเร็ว จึงเกิดการสร้าง Data Visualization ขึ้นมาData Visualization เป็นการใช้ภาพเพื่อแสดงข้อมูลในเชิงปริมาณที่วัดได้ ซึ่งอาจนำเสนอออกมาในรูปแบบ แผนภูมิ กราฟ กราฟิก และอื่น ๆ อีกมากมาย เพื่อให้เข้าใจได้

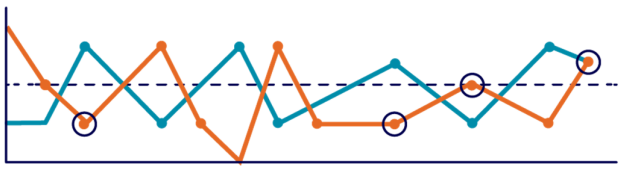
ง่าย และรวดเร็ว การเลือกรูปแบบ Visualization ให้เหมาะสมกับข้อมูลที่มีอยู่ เพื่อให้การนำเสนอข้อมูลน่าสนใจและมีประสิทธิภาพสูงสุด ซึ่งเลือกแผนภูมิยอดนิยมที่ใช้กันอย่างแพร่หลาย ดังนี้

1) แผนภูมิแท่ง (Bar Charts) เป็นแผนภูมิที่ประกอบด้วยแกนนอน แกนตั้งที่นิยมแสดงออกมาในรูปแท่งสี่เหลี่ยมที่สามารถบอกความสูงได้ เหมาะสำหรับใช้การเปรียบเทียบจำนวนของข้อมูลในแต่ละชุดเช่น รายรับในแต่ละเดือน, ยอดขายที่ขายได้จริงเปรียบเทียบกับเป้าหมายยอดขายที่ตั้งไว้ เป็นต้นซึ่งแผนภูมิแท่งยังสามารถแบ่งออกเป็นหลายประเภท • แผนภูมิแท่งแบบจัดกลุ่ม แผนภูมิแท่งแบบจัดกลุ่ม เหมาะสมกับการนำเสนอข้อมูลที่มีข้อมูลย่อย ๆ อยู่ภายใต้ข้อมูลใหญ่เป็นการเน้นให้เห็นข้อมูลย่อยนั้น ๆ แผนภูมิแท่งแบบวางซ้อนกัน แผนภูมิแท่งแบบวางซ้อนกัน เหมาะสมกับการนำเสนอข้อมูลที่ต้องการให้เห็นข้อมูลย่อยในแต่ละข้อมูลใหญ่และยังแสดงให้เห็นสัดส่วนของข้อมูลย่อยต่าง ๆ เหล่านั้นได้ด้วย ใช้แผนภูมินี้เมื่อมีชุดข้อมูล หลายชุดและต้องการเน้นผลรวมทั้งหมด



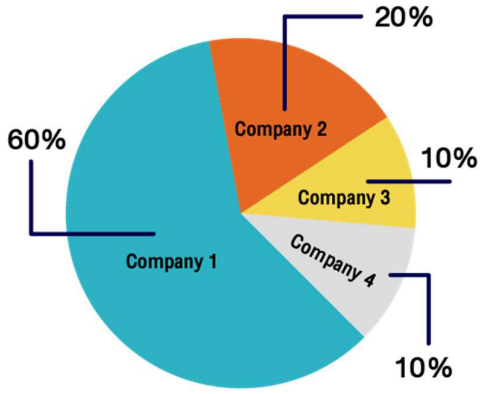
ภาพที่ 2.1 แผนภูมิแท่ง (Bar Charts)

2) แผนภูมิเส้น (Line Charts) แผนภูมิเส้น มีลักษณะคล้ายแผนภูมิแท่ง ซึ่งประกอบด้วยแกนตั้งและนอน เพียงแต่เปลี่ยนจากแท่งข้อมูลเป็นจุดบนแผนภูมินั่นเองแผนภูมิประเภทนี้เหมาะกับการนำเสนอข้อมูลตัวเลขที่มีความสัมพันธ์กับข้อมูลที่มีลักษณะเป็นช่วงใช้แสดงให้เห็นการเปลี่ยนแปลงที่เกิดขึ้นตามเวลาเพื่อดูแนวโน้ม รวมถึงสามารถใช้พยากรณ์แนวโน้มในอนาคตได้เช่น ข้อมูลของยอดขายในแต่ละปี หรือไตรมาส และนำมาวิเคราะห์เพื่อดูแนวโน้ม เป็นต้น



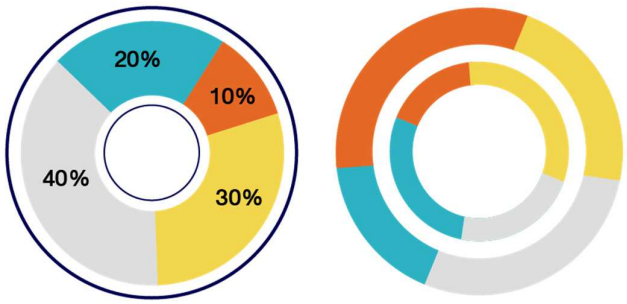
ภาพที่ 2.2 แผนภูมิเส้น (Line Charts)

3) แผนภูมิวงกลม (Pie Charts) แผนภูมิวงกลมเหมาะกับการนำเสนอข้อมูลที่มีส่วนประกอบย่อยที่รวมกันเป็นส่วนใหญ่ มีการแบ่งส่วนให้ดูง่าย และสวยงามแต่ในทางกลับกันอาจจะดูยากในเรื่องของการประมาณขนาดของแต่ละชิ้น ยิ่งถ้ามีจำนวนชิ้นมาก จะยิ่งแยกยากเพราะต้องใช้หลายสี ในการนำเสนอข้อมูล เช่น ส่วนแบ่งทางการตลาด (Market Share), ข้อมูลแสดงส่วนผสมต่าง ๆ เป็นต้น



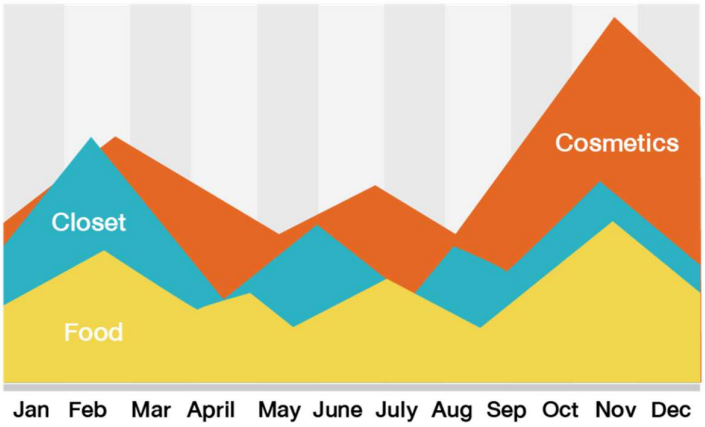
ภาพที่ 2.3 แผนภูมิวงกลม (Pie Charts)

4) แผนภูมิโดนัท (Doughnut Charts) แผนภูมิโดนัทมีหลักการออกแบบเช่นเดียวกับแผนภูมิวงกลมแต่สามารถแสดงชุดข้อมูลได้มากกว่า 1 ชุด โดยนำเสนอข้อมูลเป็นวงกลมซ้อนกันหลาย ๆ ชั้น นั่นเอง



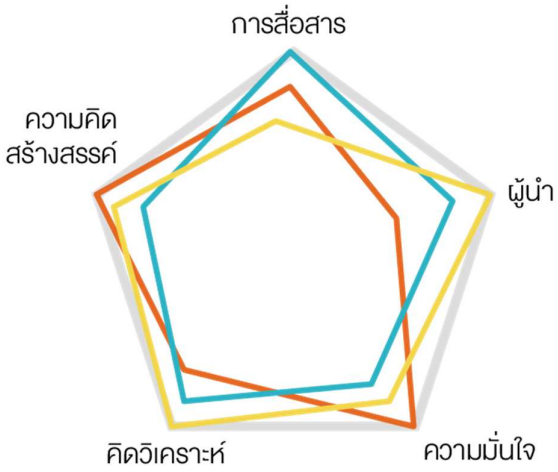
ภาพที่ 2.4 แผนภูมิโดนัท (Doughnut Charts)

5) แผนภูมิพื้นที่ (Area Charts) มีหน้าตาคล้ายแผนภูมิเส้น แต่มีการแรเงาพื้นที่ใต้เส้นข้อมูล หรือระหว่าง 2 เส้นเพื่อแสดงให้เห็นปริมาณความแตกต่างระหว่างเส้น เหมาะสำหรับเน้นความสำคัญของการเปลี่ยนแปลงในช่วงเวลาแสดงให้เห็นผลรวมของความแตกต่างระหว่างข้อมูล เช่น ข้อมูลของการซื้อสินค้าในห้างสรรพสินค้าในแต่เดือน ตามหมวดหมู่ต่าง ๆ ไล่ไป เครื่องสำอางค์ เสื้อผ้าแฟชั่น อาหาร ตามลำดับ



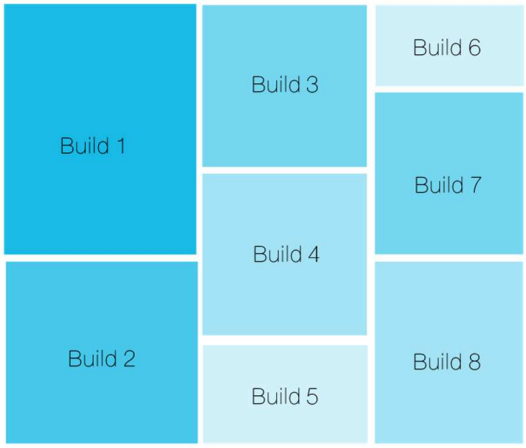
ภาพที่ 2.5 แผนภูมิพื้นที่ (Area Charts)

6) แผนภูมิเรดาร์ (Radar Charts) มีลักษณะคล้ายแผนภูมิเส้นที่มีการแสดงผลแบบวงกลมจำนวนเหลี่ยมของเรดาร์เท่ากับจำนวนหัวข้อของข้อมูล แผนภูมินี้ไม่ได้บอกถึงความต่อเนื่องของข้อมูลแต่เหมาะสำหรับการนำเสนอข้อมูลเป็นหัวข้อ แล้วนำมาวิเคราะห์หาจุดอ่อนจุดแข็งของข้อมูลเช่น นำเสนอข้อมูลเกี่ยวกับคุณสมบัติของการรับพนักงานใหม่ เพื่อดูจุดอ่อนจุดแข็งของแต่ละคน เป็นต้น



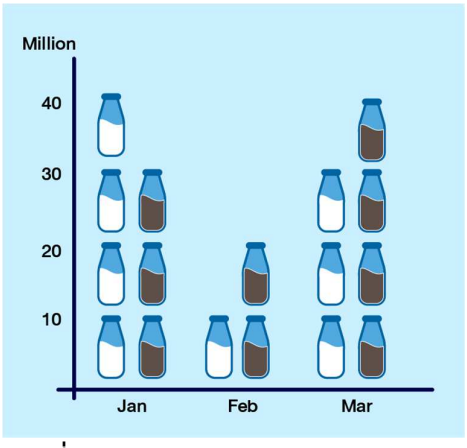
ภาพที่ 2.6 แผนภูมิเรดาร์ (Radar Charts)

7) แผนภูมิต้นไม้ (Tree Maps) คือการนำเสนอข้อมูลแบบแสดงให้เห็นพื้นที่ แสดงผลได้ในแบบลำดับชั้น เหมือนแบบโครงสร้างต้นไม้ อาจจะนำเสนอข้อมูลที่ต้องการให้เห็นถึงเขตพื้นที่ แสดงพื้นที่สีที่แตกต่างกันได้



ภาพที่ 2.7 แผนภูมิต้นไม้ (Tree Maps)

8) แผนภูมิรูปภาพ (Picture Graph) เป็นแผนภูมิที่ประกอบไปด้วยแกนนอน และแกนตั้ง แต่เลือกใช้รูปภาพ หรือไอคอนแทนจำนวนของสิ่งของนั้น ๆ เช่น การแสดงผลจำนวนของนมที่ขายได้ในแต่ละเดือน โดยนำเสนอทั้งนมรสจืด รสช็อกโกแลต เปรียบเทียบในแต่ละเดือนซึ่งมีการนำเสนอลักษณะคล้ายกับกราฟแท่ง แต่เปลี่ยนจากแท่งเป็นรูปกราฟิกของนม 2 รสชาติแทน ก็ทำให้การนำเสนอข้อมูลน่าสนใจมากยิ่งขึ้นซึ่งแนวทางการนำเสนอข้อมูลลักษณะนี้ต้องอาศัยความคุ้นชินของคนดู เพื่อแทนสัญลักษณ์ภาพลงไป เช่น เมื่อพูดถึงจำนวนคนอาจจะแทนด้วยภาพไอคอนคนหรือเมื่อพูดถึงจำนวนเงิน ควรแทนภาพเป็นเหรียญเงิน หรือแบงค์แทน ก็จะทำให้คนดูเข้าใจง่ายจากสัญลักษณ์ภาพที่คุ้นเคยอยู่แล้วและยังดึงดูดความสนใจได้มากกว่าการใช้กราฟแท่งสี่เหลี่ยมอีกด้วย



ภาพที่ 2.8 แผนภูมิรูปภาพ (Picture Graph)

2.2 ทฤษฎี

2.2.1 ทฤษฎีเกี่ยวกับการจัดการข้อมูลขนาดใหญ่

2.2.1.1 ข้อมูลขนาดใหญ่

ข้อมูลขนาดใหญ่ หมายถึงข้อมูลที่มีขนาดใหญ่ เร็ว หรือซับซ้อนจนยากหรือเป็นไปไม่ได้ที่จะประมวลผลโดยใช้วิธีการแบบเดิม การเข้าถึงและจัดเก็บข้อมูลจำนวนมากเพื่อทำการวิเคราะห์มีมานานแล้ว แต่แนวคิดเกี่ยวกับข้อมูลขนาดใหญ่เป็นที่แพร่หลายในช่วงต้นปีค.ศ. 2000 เมื่อดั๊ก ลานีย์ นักวิเคราะห์อุตสาหกรรมได้ให้คำจำกัดความที่เป็นที่เข้าใจกันในขณะนี้ว่า ข้อมูลขนาดใหญ่ประกอบด้วย3Vs ดังนี้

Volume (ปริมาณ) **:** องค์กรต่าง ๆ รวบรวมข้อมูลจากหลากหลายแหล่ง ซึ่งรวมถึงธุรกรรมของธุรกิจ อุปกรณ์อัจฉริยะ (IoT) อุปกรณ์อุตสาหกรรม วิดีโอ โซเชียลมีเดีย และอื่น ๆ ในอดีต การจัดเก็บข้อมูลถือเป็นปัญหาใหญ่ – 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 ช่วยให้องค์กรควบคุมข้อมูลของพวกเขาและใช้เพื่อระบุโอกาสใหม่ ๆ ในทางกลับกันนำไปสู่การเคลื่อนไหวทางธุรกิจที่ชาญฉลาดเพื่อการดำเนินงานที่มีประสิทธิภาพมากขึ้น ผลกำไรที่สูงขึ้นและลูกค้าที่มีความสุขมากขึ้น ในรายงาน Big Data ใน บริษัท ขนาดใหญ่โดยผู้อำนวยการฝ่ายวิจัยของ Tom Davenport ให้สัมภาษณ์ว่า มากกว่า 50 ธุรกิจใช้และทำความเข้าใจว่าพวกเขาใช้ Big Data อย่างไร และพบว่าสามารถช่วยเหลือธุรกิจได้ดังต่อไปนี้

1) ลดต้นทุน : เทคโนโลยีข้อมูลขนาดใหญ่เช่น Hadoop และการวิเคราะห์บนคลาวด์นำมาซึ่งความได้เปรียบด้านต้นทุนอย่างมีนัยสำคัญเมื่อพูดถึงการจัดเก็บข้อมูลจำนวนมากรวมทั้งสามารถระบุวิธีการทำธุรกิจที่มีประสิทธิภาพมากขึ้น

2) เร็วกว่าและตัดสินใจดีกว่า : ด้วยความเร็วของ Hadoop และการวิเคราะห์ในหน่วยความจำรวมกับความสามารถในการวิเคราะห์แหล่งข้อมูลใหม่ๆ ของธุรกิจจะสามารถสร้างข้อมูลได้ทันทีและสามารถดำเนินการต่อได้ทันทีจากการวิเคราะห์นั้น ๆ

3) ผลิตภัณฑ์และบริการใหม่ : ด้วยความสามารถในการวัดความต้องการและความพึงพอใจของลูกค้าผ่านการวิเคราะห์นำมาซึ่งสิ่งที่ลูกค้าต้องการ ซึ่งยกตัวอย่างโดยดาเวนพอร์ท สามารถชี้ให้เห็นว่าด้วยการวิเคราะห์ข้อมูลขนาดใหญ่ของบริษัทต่าง ๆ จะสามารถสร้างผลิตภัณฑ์ใหม่ ๆ 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 ให้เหมาะสำหรับการใช้ในการวิเคราะห์หรือใช้งานสำหรับธุรกิจหรือวัตถุประสงค์นั้น ๆ

การจัดการข้อมูล

ข้อมูลขนาดใหญ่ หรือ Big Data นั้นมีความต้องการสถานที่จัดเก็บขนาดใหญ่ การจัดเก็บข้อมูลมูลขนาดใหญ่จะเป็นชนิดใดก็ได้ไม่ว่าจะเป็นแบบ on premises หรือแบบ cloud ขึ้นกับความต้องการหรือความสะดวกในการใช้ ซึ่งเราสามารถใช้และประเมินผลได้เช่นเดียวกัน บางครั้งก็มีความจำเป็นที่ต้องจัดเก็บไว้ใกล้กับแหล่งข้อมูล หรือข้อมูลบางอันต้องการความยืดหยุ่นสูงและไม่ต้องการบริหารจัดการก็ใช้เป็นแบบ Cloud ซึ่งกำลังเป็นที่นิยมกันเป็นอย่างมาก

การวิเคราะห์

การลงทุนสร้างข้อมูลขนาดใหญ่ หรือ Big data จะมีประโยชน์หรือคุ้มค่าก็ต่อเมื่อคุณใช้และวิเคราะห์ข้อมูล การวิเคราะห์ข้อมูลทำให้เกิดความกระจ่างและชัดเจนในชุดข้อมูลที่คุณมีอยู่ การสำรวจข้อมูลยังทำให้เราค้นพบสิ่งใหม่ แชร์สิ่งที่ค้นพบใหม่ๆต่อคนอื่น สร้างรูปแบบจำลองข้อมูล ด้วยการเรียนรู้ของเครื่องจักรและปัญญาประดิษฐ์ AI และนำข้อมูลเหล่านั้นไปใช้งาน

2.2.1.3 เทคโนโลยีการประมวลผลข้อมูล

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

2.2.2 ทฤษฎีเกี่ยวกับการทำเหมืองข้อมูล

เหมืองข้อมูล คือกระบวนการที่กระทำกับข้อมูลจำนวนมากเพื่อค้นหารูปแบบและความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น ในปัจจุบันการทำเหมืองข้อมูลได้ถูกนำไประยุกต์ใช้ในงานหลายประเภท ทั้งในด้านธุรกิจที่ช่วยในการตัดสินใจของผู้บริหาร ในด้านวิทยาศาสตร์และการแพทย์รวมทั้งในด้านเศรษฐกิจและสังคม

2.2.2.1 การทำเหมืองข้อมูล (Data Mining) เปรียบเสมือนวิวัฒนาการหนึ่งในการจัดเก็บและตีความหมายข้อมูล จากเดิมที่มีการจัดเก็บข้อมูลอย่างง่ายๆ มาสู่การจัดเก็บในรูปแบบฐานข้อมูลที่สามารถดึงข้อมูลสารสนเทศมาใช้จนถึงการทำเหมืองข้อมูลที่สามารถค้นพบความรู้ที่ซ่อนอยู่ในข้อมูล หรือจะแยกๆ เป็นข้อๆ ได้ดังนี้

1) กระบวนการหรือการเรียงลำดับของการค้นข้อมูลจำนวนมากและเก็บข้อมูลที่เกี่ยวข้อง

2) การนำมาใช้โดยหน่วยงานทางธุรกิจและนักวิเคราะห์ทางการเงินหรือการนำมาใช้งานในด้านวิทยาศาสตร์เพื่อเอาข้อมูลขนาดใหญ่ที่สร้างโดยวิธีการทดลองและการสังเกตการณ์ที่ทันสมัย

3) การสกัดหรือแยกข้อมูลที่เป็นประโยชน์จากข้อมูลขนาดใหญ่หรือฐานข้อมูล

4) การวางแผนทรัพยากรขององค์กรโดยสามารถวิเคราะห์ทางสถิติและตรรกะของข้อมูลขนาดใหญ่เป็นการมองหารูปแบบที่สามารถช่วยการตัดสินใจได้

2.2.2.2 ขั้นตอนการทำเหมืองข้อมูล

ประกอบด้วยขั้นตอนการทำงานย่อยที่จะเปลี่ยนข้อมูลดิบให้กลายเป็นความรู้ ประกอบด้วยขั้นตอนดังนี้

1) Business Understanding เน้นไปที่การทำความเข้าใจในงาน ระบุโอกาสและหาปัญหาที่จะเกิดขึ้น กำหนดขอบเขตของข้อมูลที่จะนำวิเคราะห์ ซึ่งต้องสามารถระบุผลลัพธ์ที่มีได้

2) Data Understanding ทำความเข้าใจข้อมูลโดยการรวบรวมข้อมูลที่เกี่ยวข้อง คัดเลือกให้เหลือเพียงข้อมูลที่มีความถูกต้องและสำคัญต่องานมาทำการวิเคราะห์

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

4) Modeling การสร้างแบบจำลองเพื่อวิเคราะห์ข้อมูลที่ได้จากขั้นตอนที่ 3 พร้อมทดสอบผลลัพธ์แบบจำลองเพื่อให้ได้คำตอบที่ดีที่สุด บางครั้งอาจมีการย้อนกลับไปปรับการเตรียมข้อมูลเพื่อให้ได้แบบจำลองที่เหมาะสมที่สุด

5) Evaluation การประเมินผลลัพธ์ที่ได้ก่อนที่จะนำไปใช้จริง ว่าตรงกับวัตถุประสงค์หรือเป้าหมายที่ได้ตั้งไว้หรือมีความน่าเชื่อถือมากน้อยเพียงใด หากไม่ได้ผลลัพธ์ตามวัตถุประสงค์ต้องย้อนกลับไปปรับปรุงแก้ไขการดำเนินงานในขั้นตอนก่อนหน้า

6) Deployment การนำเอาข้อมูลที่เป็นผลลัพธ์จากทั้งหมด มาใช้ประโยชน์ตามวัตถุประสงค์ และทำการประเมินผลลัพธ์ที่ได้ว่ามีประสิทธิภาพตรงตามวัตถุประสงค์ที่ตั้งไว้หรือไม่

2.2.3 ทฤษฎีเกี่ยวกับการออกแบบเว็บไซต์

2.2.3.1 หลักในการออกแบบเว็บไซต์

หน้าเว็บเป็นสิ่งแรกที่ผู้ใช้จะได้เห็นขณะที่เปิดเข้าสู่เว็บไซต์ และยังเป็นสิ่งแรกที่แสดงถึงประสิทธิภาพในการออกแบบเว็บไซต์อีกด้วย หน้าเว็บจึงเป็นสิ่งสำคัญมาก เพราะเป็นสื่อกลางให้ผู้ชมสามารถใช้ประโยชน์จากข้อมูลของระบบงานของเว็บไซต์นั้นได้ โดยปกติหน้าเว็บจะประกอบด้วย รูปภาพ ตัวอักษร สีพื้น ระบบเนวิเกชั่น และองค์ประกอบอื่น ๆ ที่ช่วยสื่อความหมายของเนื้อหาและอำนวยความสะดวกต่อการใช้งานหลักสำคัญในการออกแบบหน้าเว็บก็คือ การใช้รูปภาพและองค์ประกอบต่าง ๆ ร่วมกันเพื่อสื่อความหมายเกี่ยวกับเนื้อหาหรือลักษณะสำคัญของเว็บไซต์ โดยมีเป้าหมายสำคัญเพื่อการสื่อความหมายที่ชัดเจนและน่าสนใจ บนพื้นฐานของความเรียบง่ายและความสะดวกของผู้ใช้ การออกแบบเว็บไซต์ ต้องคำนึงถึง

1) ความเรียบง่าย ได้แก่ มีรูปแบบที่เรียบง่าย ไม่ซับซ้อน และใช้งานได้สะดวก ไม่มีกราฟิกหรือตัวอักษรที่เคลื่อนไหวอยู่ตลอดเวลา ชนิดและสีของตัวอักษรไม่มากจนเกินไปทำให้วุ่นวาย

2) ความสม่ำเสมอ ได้แก่ ใช้รูปแบบเดียวกันตลอดทั้งเว็บไซต์ เช่น รูปแบบของหน้า สไตล์ของกราฟิก ระบบเนวิเกชันและโทนสี ควรมีความคล้ายคลึงกันตลอดทั้งเว็บไซต์

3) ความเป็นเอกลักษณ์ การออกแบบเว็บไซต์ควรคำนึงถึงลักษณะขององค์กร เพราะรูปแบบของเว็บไซต์จะสะท้อนถึงเอกลักษณ์และลักษณะขององค์กรนั้น ๆ เช่น ถ้าเป็นเว็บไซต์ของทาง ราชการ จะต้องดูน่าเชื่อถือไม่เหมือนสวนสนุก ฯลฯ

4) เนื้อหาที่มีประโยชน์ เนื้อหาเป็นสิ่งที่สำคัญที่สุดในเว็บไซต์ ดังนั้นควรจัดเตรียมเนื้อหาและข้อมูลที่ผู้ใช้ต้องการให้ถูกต้อง และสมบูรณ์ 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 สร้างความรู้สึกว่าเว็บไซต์มี....คุณภาพ ถูกต้อง และเชื่อถือได้

9) ลิงค์ต่าง ๆ จะต้องเชื่อมโยงไปหน้าที่มีอยู่จริงและถูกต้อง ระบบการทำงานต่าง ๆ ในเว็บไซต์จะต้องมีความแน่นอนและทำหน้าที่ได้อย่างถูก

2.2.3.2 องค์ประกอบของการออกแบบเว็บไซต์

1) โครงสร้างที่ชัดเจน ผู้ออกแบบเว็บไซต์ควรจัดโครงสร้างหรือจัดระเบียบของข้อมูลที่ชัดเจน แยกย่อยเนื้อหาออกเป็นส่วนต่าง ๆ ที่สัมพันธ์กันและให้อยู่ในมาตรฐานเดียวกัน จะช่วยให้น่าใช้งานและง่าย ต่อการอ่านเนื้อหาของผู้ใช้

2) การใช้งานที่ง่าย ลักษณะของเว็บที่มีการใช้งานง่ายจะช่วยให้ผู้ใช้รู้สึกสบายใจต่อการอ่านและสามารถทำความเข้าใจกับเนื้อหาได้อย่างเต็มที่ โดยไม่ต้องมาเสียเวลาอยู่กับการทำความเข้าใจ การใช้งานที่สับสนด้วยเหตุนี้ผู้ออกแบบจึงควรกำหนดปุ่มการใช้งานที่ชัดเจน เหมาะสม โดยเฉพาะปุ่มควบคุมเส้นทางการเข้าสู่เนื้อหา (Navigation) ไม่ว่าจะเป็นเดินหน้า ถอยหลัง หากเป็นเว็บไซต์ที่มีเว็บเพจจำนวนมาก ควรจะจัดทำแผนผังของเว็บไซต์ (Site Map) ที่ช่วยให้ผู้ใช้ทราบว่า ตอนนี้อยู่ ณ จุดใด หรือเครื่องมือสืบค้น (Search Engine) ที่ช่วยในการค้นหาหน้าที่ที่ต้องการ

3) การเชื่อมโยงที่ดี ลักษณะไฮเปอร์เท็กซ์ที่ใช้ในการเชื่อมโยง ควรอยู่ในรูปแบบที่เป็นมาตรฐาน ทั่วไปและต้องระวังเรื่องของตำแหน่งในการเชื่อมโยง การที่จำนวนการเชื่อมโยงมากและกระจัดกระจายอยู่ทั่วไปในหน้าอาจก่อให้เกิดความสับสน นอกจากนี้คำที่ใช้สำหรับการเชื่อมโยงจะต้องเข้าใจง่ายมีความชัดเจนและไม่สั้นจนเกินไป นอกจากนี้ในแต่ละเว็บเพจที่สร้างขึ้นมาควรมี จุดเชื่อมโยงกลับมายังหน้าแรกของเว็บไซต์ที่กำลังใช้งานอยู่ด้วย ทั้งนี้เผื่อว่าผู้ใช้เกิดหลงทาง และไม่ทราบว่าจะทำอย่างต่อไปดีจะได้มีหนทางกลับมาสู่จุดเริ่มต้นใหม่ ระวังอย่าให้มีหน้าที่ไม่มีการเชื่อมโยง (Orphan Page) เพราะจะทำให้ผู้ใช้ไม่รู้จะทำอย่างไรต่อไป

4) ความเหมาะสมในหน้าจอ เนื้อหาที่นำเสนอในแต่ละหน้าจอควรสั้นกระชับ และทันสมัย หลีกเลี่ยงการใช้หน้าจอที่มีลักษณะการเลื่อนขึ้นลง (Scrolling) แต่ถ้าจำเป็นต้องมี ควรจะให้ข้อมูลที่มี ความสำคัญอยู่บริเวณด้านบนสุดของหน้าจอ หลีกเลี่ยงการใช้กราฟิกด้านบนของหน้าจอ เพราะถึงแม้จะดูสวยงาม แต่จะทำให้ผู้ใช้เสียเวลาในการได้รับข้อมูลที่ต้องการ แต่หากต้องมีการใช้ภาพประกอบก็ควรใช้เฉพาะที่มีความสัมพันธ์กับเนื้อหาเท่านั้น นอกจากนี้การใช้รูปภาพเพื่อเป็นพื้นหลัง (Background) ไม่ควรเน้นสีสันที่ฉูดฉาดมากนัก เพราะอาจจะไปลดความเด่นชัดของเนื้อหาลง ควรใช้ภาพที่มีสีอ่อน ๆ ไม่สว่างจนเกินไป รวมไปถึงการใช้เทคนิคต่าง ๆ เช่น ภาพเคลื่อนไหว หรือตัวอักษรวิ่ง (Marquees) ซึ่งอาจจะเกิดการรบกวนการอ่านได้ ควรใช้เฉพาะที่จำเป็นจริง ๆ เท่านั้นตัวอักษรที่นำมาแสดงบนจอภาพควรเลือกขนาดที่อ่านง่าย ไม่มีสีสันและลวดลายมากเกินไป

5) ความรวดเร็ว ความรวดเร็วเป็นสิ่งสำคัญประการหนึ่งที่ส่งผลต่อการเรียนรู้ ผู้ใช้จะเกิดอาการเบื่อหน่ายและหมดความสนใจกับเว็บที่ใช้เวลาในการแสดงผลนาน สาเหตุสำคัญที่จะทำให้การแสดงผลนานคือการใช้ภาพกราฟิกหรือภาพเคลื่อนไหว ซึ่งแม้ว่าจะช่วยดึงดูดความสนใจได้ดี ฉะนั้นในการออกแบบจึงควรหลีกเลี่ยงการใช้ภาพขนาดใหญ่ หรือภาพเคลื่อนไหวที่ไม่จำเป็น และพยายามใช้กราฟิกแทนตัวอักษรธรรมดาให้น้อยที่สุด โดยไม่ควรใช้มากเกินกว่า 2 – 3 บรรทัดในแต่ละหน้าจอ

2.2.3.3 โครงสร้างเว็บไซต์

1) เว็บที่มีโครงสร้างแบบเรียงลำดับ (Sequential Structure) เป็นโครงสร้างแบบธรรมดาที่ใช้กันมากที่สุดเนื่องจากง่ายต่อการจัดระบบข้อมูล ข้อมูลที่นิยม จัดด้วยโครงสร้างแบบนี้มักเป็นข้อมูลที่มีลักษณะเป็นเรื่องราวตามลำดับของเวลา เช่น การเรียงลำดับตามตัวอักษร ดรรชนี สารานุกรม หรืออภิธานศัพท์ โครงสร้างแบบนี้ เหมาะกับเว็บไซต์ที่มีขนาดเล็ก เนื้อหาไม่ซับซ้อนใช้การลิงก์ (Link) ไปทีละหน้า ทิศทางของการเข้าสู่เนื้อหา (Navigation) ภายในเว็บจะเป็นการดำเนินเรื่องในลักษณะเส้นตรง โดยมี ปุ่มเดินหน้า–ถอยหลังเป็นเครื่องมือหลักในการกำหนดทิศทาง ข้อเสียของโครงสร้างระบบนี้คือ ผู้ใช้ไม่สามารถกำหนดทิศทางการเข้าสู่เนื้อหาของตนเองได้ ทำให้เสียเวลาเข้าสู่เนื้อ

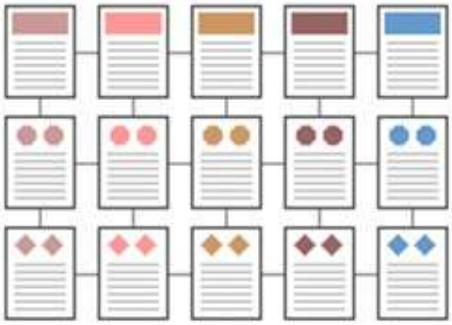
ภาพที่ 2.9 เว็บที่มีโครงสร้างแบบเรียงลำดับ

2) เว็บที่มีโครงสร้างแบบลำดับขั้น (Hierarchical Structure) เป็นวิธีที่ดีที่สุดวิธีหนึ่งในการจัดระบบโครงสร้างที่มีความซับซ้อนของข้อมูล โดยแบ่งเนื้อหา ออกเป็นส่วนต่างๆ และมีรายละเอียดย่อยๆ ในแต่ละส่วนลดหลั่นกันมาในลักษณะแนวคิดเดียวกับ แผนภูมิองค์กร จึงเป็นการง่ายต่อการทำความเข้าใจกับโครงสร้างของเนื้อหาในเว็บลักษณะนี้ ลักษณะเด่นเฉพาะของ เว็บประเภทนี้คือการมีจุดเริ่มต้นที่จุดร่วมจุดเดียว นั่นคือ โฮมเพจ (Homepage) และเชื่อมโยงไปสู่เนื้อหา ในลักษณะเป็นลำดับจากบนลงล่าง

ภาพที่ 2.10 เว็บที่มีโครงสร้างแบบลำดับขั้น

3) เว็บที่มีโครงสร้างแบบตาราง (Grid Structure) โครงสร้างรูปแบบนี้มีความซับซ้อนมากกว่ารูปแบบที่ผ่านมา การออกแบบเพิ่มความยืดหยุ่น ให้แก่การเข้าสู่เนื้อหาของผู้ใช้ โดยเพิ่มการเชื่อมโยงซึ่งกันและกันระหว่างเนื้อหาแต่ละส่วน เหมาะแก่ การแสดงให้

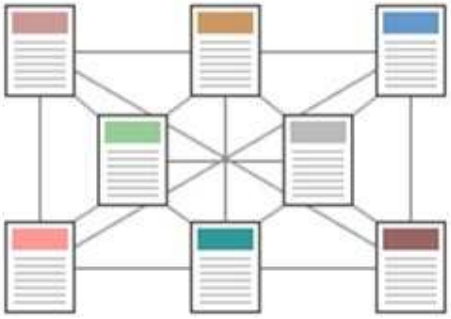
เห็นความสัมพันธ์กันของเนื้อหา การเข้าสู่เนื้อหาของผู้ใช้จะไม่เป็นลักษณะเชิงเส้นตรง เนื่องจากผู้ใช้สามารถเปลี่ยนทิศทางการเข้าสู่เนื้อหาของตนเองได้



ภาพที่ 2.11 เว็บที่มีโครงสร้างแบบตาราง

ในการจัดระบบโครงสร้างแบบนี้ เนื้อหาที่นำมาใช้แต่ละส่วนควรมีลักษณะที่เหมือนกัน และ สามารถใช้รูปแบบร่วมกัน หลักการออกแบบคือนำหัวข้อทั้งหมดมาบรรจุลงในที่เดียวกันซึ่งโดยทั่วไป จะเป็นหน้าแผนภาพ (Map Page) ที่แสดงในลักษณะเดียวกับโครงสร้างของเว็บ เมื่อผู้ใช้คลิกเลือก หัวข้อใด ก็จะเข้าไปสู่หน้าเนื้อหา (Topic Page) ที่แสดงรายละเอียดของหัวข้อนั้นๆ และภายในหน้านั้น ก็จะมีการเชื่อมโยงไปยังหน้ารายละเอียดของหัวข้ออื่นที่เป็นเรื่องเดียวกัน นอกจากนี้ยังสามารถนำ โครงสร้างแบบเรียงลำดับและแบบลำดับขั้นมาใช้ร่วมกันได้อีกด้วย ถึงแม้โครงสร้างแบบนี้ อาจจะสร้างความยุ่งยากในการเข้าใจได้ และอาจเกิดปัญหาการคงค้าง ของหัวข้อ (Cognitive Overhead) ได้ แต่จะเป็นประโยชน์ที่สุดเมื่อผู้ใช้ได้เข้าใจถึงความสัมพันธ์ ระหว่างเนื้อหา ในส่วนของการออกแบบจำเป็นจะต้องมีการวางแผนที่ดี เนื่องจากมีการเชื่อมโยงที่เกิดขึ้น ได้หลายทิศทาง นอกจากนี้การปรับปรุงแก้ไขอาจเกิดความยุ่งยากเมื่อต้องเพิ่มเนื้อหาในภายหลัง

4) เว็บที่มีโครงสร้างแบบใยแมงมุม (Web Structure) โครงสร้างประเภทนี้จะมีความยืดหยุ่นมากที่สุด ทุกหน้าในเว็บสามารถจะเชื่อมโยงไปถึงกัน ได้หมด เป็นการสร้างรูปแบบการเข้าสู่เนื้อหาที่เป็นอิสระ ผู้ใช้สามารถกำหนดวิธีการเข้าสู่เนื้อหาได้ด้วย ตนเอง การเชื่อมโยงเนื้อหาแต่ละหน้าอาศัยการโยงใยข้อความที่มีมโนทัศน์ (Concept) เหมือนกัน ของแต่ละหน้าในลักษณะของไฮเปอร์เท็กซ์หรือไฮเปอร์มีเดีย โครงสร้างลักษณะนี้จัดเป็นรูปแบบที่ ไม่มีโครงสร้างที่แน่นนอนตายตัว (Unstructured) นอกจากนี้การเชื่อมโยงไม่ได้จำกัดเฉพาะเนื้อหาภายในเว็บนั้น ๆ แต่สามารถเชื่อมโยงออกไปสู่เนื้อหาจากเว็บภายนอกได้



ภาพที่ 2.12 เว็บที่มีโครงสร้างแบบใยแมงมุม

ลักษณะการเชื่อมโยงในเว็บนั้น นอกเหนือจากการใช้ไฮเปอร์เท็กซ์หรือไฮเปอร์มีเดีย กับข้อความที่มีมโนทัศน์ (Concept) เหมือนกันของแต่ละหน้าแล้ว ยังสามารถใช้ลักษณะการเชื่อมโยง จากรายการที่รวบรวมชื่อหรือหัวข้อของเนื้อหาแต่ละหน้าไว้ ซึ่งรายการนี้จะปรากฏอยู่บริเวณใด บริเวณหนึ่งในหน้าจอ ผู้ใช้สามารถคลิกที่หัวข้อใดหัวข้อหนึ่งในรายการเพื่อเลือกที่จะเข้าไปสู่หน้าใด ๆ ก็ได้ตามความต้องการ ข้อดีของรูปแบบนี้คือง่ายต่อผู้ใช้ในการท่องเที่ยวบนเว็บ โดยผู้ใช้สามารถกำหนดทิศทาง การเข้าสู่เนื้อหาได้ด้วยตนเอง แต่ข้อเสียคือถ้ามีการเพิ่มเนื้อหาใหม่ๆ อยู่เสมอจะเป็นการยากในการ ปรับปรุง นอกจากนี้การเชื่อมโยงระหว่างข้อมูลที่มีมากมายนั้นอาจทำให้ผู้ใช้เกิดการสับสนและ เกิดปัญหาการคงค้างของหัวข้อ (Cognitive Overhead) ได้

2.2.3.3 การใช้สีในการออกแบบเว็บไซต์

การสร้างสีสันบนหน้าเว็บเป็นสิ่งที่สื่อความหมายของเว็บไซต์ได้อย่างชัดเจน การเลือกใช้สีให้เหมาะสม 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 จุดแต่ละจุดหรือพิกเซล (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 สีแต่ละสีให้ความรู้สึก อารมณ์ที่ไม่เหมือนกัน 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 สีเทา และสีน้ำตาล กลุ่มสีเหล่านี้คือ สีกลางที่สามารถนำไปผสมกับสีอื่น ๆ เพื่อให้เกิดสีกลางขึ้นมา

2.2.4 ทฤษฎีเกี่ยวกับชุดคำสั่งภาษาของการออกแบบเว็บไซต์

2.2.4.1 CSS (ย่อมาจาก Cascading Style Sheet) มักเรียกโดยย่อว่า "สไตล์ชีต" คือภาษาที่ใช้เป็นส่วนของการจัดรูปแบบการแสดงผลเอกสาร HTML โดยที่ CSS กำหนดกฎเกณฑ์ในการระบุรูปแบบ (หรือ "Style") ของเนื้อหาในเอกสาร อันได้แก่ 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

ประโยชน์ของ CSS

1) CSS มีคุณสมบัติมากกว่า tag ของ html เช่น การกำหนดกรอบให้ข้อความ รวมทั้งสี รูปแบบของข้อความที่กล่าวมาแล้ว

2) CSS นั้นกำหนดที่ต้นของไฟล์ html หรือตำแหน่งอื่น ๆ ก็ได้ และสามารถมีผล กับเอกสารทั้งหมด หมายถึงกำหนด ครั้งเดียวจุดเดียวก็มีผลกับการแสดงผลทั้งหมด ทำให้เวลาแก้ไขหรือปรับปรุงทำได้สะดวก ไม่ต้องไล่ตามแก้ tag ต่างๆ ทั่วทั้งเอกสาร

3) CSS สามารถกำหนดแยกไว้ต่างหากจาก ไฟล์เอกสาร html และสามารถนำมาใช้ร่วม กับเอกสารหลายไฟล์ได้ การแก้ไขก็แก้เพียง จุดเดียวก็มีผลกับเอกสารทั้งหมด

CSS กับ HTML / XHTML นั้นทำหน้าที่คนละอย่างกัน โดย HTML / XHTML จะทำหน้าที่ในการวางโครงร่างเอกสารอย่างเป็นรูปแบบ ถูกต้อง เข้าใจง่าย ไม่เกี่ยวข้องกับการแสดงผล ส่วน CSS จะทำหน้าที่ในการตกแต่งเอกสารให้สวยงาม เรียกได้ว่า HTML /XHTML คือส่วน coding ส่วน CSS คือส่วน design

2.2.4.2 Bootstrap คือชุดคำสั่งที่ประกอบด้วยภาษา CSS, HTML และ Javascript เป็นชุดคำสั่งที่ถูกพัฒนาขึ้นมาเพื่อกำหนดกรอบ หรือ รูปแบบการพัฒนาเว็บไซต์ในส่วนของการปฏิสัมพันธ์กับผู้ใช้งานเว็บไซต์ (User Interface) เราจึงสามารถเรียก Bootstrap ว่าเป็น Front-end framework คือใช้สำหรับ พัฒนาเว็บไซต์ส่วนการแสดงผล ซึ่งแตกต่างจากภาษาประเภท Server Side Script อย่าง PHP, Python หรือภาษาอื่น ๆ

จุดเด่นของของ Bootstrap Framework

1) มี UI เริ่มต้นแบบที่สวยงามและใช้งานง่าย

2) มีการปรับปรุงและพัฒนาอย่างต่อเนื่อง ปัจจุบันเป็นเวอร์ชั่น 3.3.0

3) เป็นที่นิยมของนักพัฒนาทั่วโลก ทำให้สามารถเรียนรู้ และแก้ปัญหาได้ง่าย

4) โค๊ดหรือชุดคำสั่งต่าง ๆ ค่อนข้างสะอาดมีโฟลเด้อต้นแบบแค่ 3 ส่วน คือ js, css, fonts

5) ประหยัดเวลาในการพัฒนาเว็บไซต์และนำไปพัฒนาต่อได้ง่าย

6) เป็น Responsive Framework พัฒนาเว็บไซต์ที่รองรับการแสดงผลได้หลากหลาย Device

2.2.5 ทฤษฎีเกี่ยวกับการ visualization

Data Visualization หรือ Information Visualization คือ การนำข้อมูลในเชิงปริมาณทั้งที่จัดเก็บไว้ในรูปแบบของข้อมูลจำนวนน้อย และข้อมูลจำนวนมหาศาล (Big Data) มาประมวลผล จากนั้นจึงนำมาแสดงผลในรูปแบบของกราฟ แผนภูมิอินฟอร์เมชันกราฟิก หรือแม้กระทั่งอินเทอร์แอคทีฟกราฟิก ที่ผู้บริโภคสามารถคลิกหรือมีปฏิสัมพันธ์กับกราฟิกนั้น ๆ ได้ ซึ่งปัจจุบันสื่อหลายสำนักในประเทศไทย เริ่มมีการนำเสนอข่าวโดยการนำข้อมูลแบบประยุกต์ด้วยภาพ (Data Visualization) การถ่ายทอดข้อมูลในเชิงปริมาณที่มีความซับซ้อนได้อย่างมีประสิทธิผล ซึ่งคำว่า “ประสิทธิผล” ในที่นี้หมายถึงมีความชัดเจน (Clarity), มีความแม่นยำ (Precision), และมีประสิทธิภาพ (Efficiency) หากไม่มีการทำ Data Visualization แล้ว อาจทำให้เราไม่สามารถค้นพบนัยยะของข้อมูลในแง่ของแนวโน้ม, รูปแบบพฤติกรรม, และความสัมพันธ์เชื่อมโยงได้

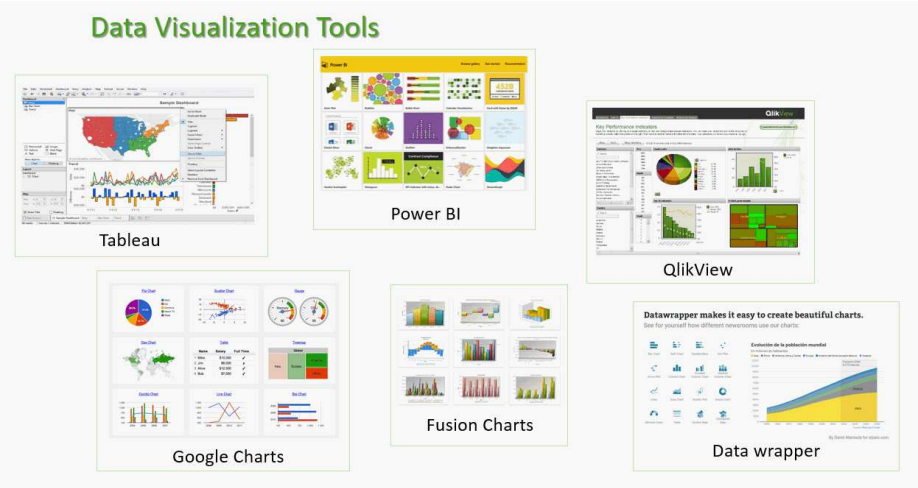
Visualization คือ การจินตนาการ หรือสร้างภาพขึ้นในความคิด ซึ่งเป็นกระบวนการทำให้เกิดความรู้ความเข้าใจสำหรับเรื่องที่จินตนาการยาก เข้าใจยาก วิธีการที่เป็นทางลัดก็คือการสร้าง ภาพ ให้เป็น บันไดความคิด ไปสู่ การใช้ความคิดอีกระดับ บันไดนี้จะช่วยตัดปริมาณข้อมูล ช่วยลดภาระการคำนวณหรือการนำไปผ่านหลากหลายกระบวนการความคิด เพื่อนำไปสู่คำตอบที่ต้องการได้โดยเร็วและถูกต้อง

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

Data Visualization เป็นการนำข้อมูลมาผสมผสานกับจินตนาการ เพื่อสร้างภาพในความคิดขึ้นมา ซึ่งมีกระบวนการนำเสนอข้อมูลที่มีความซับซ้อนหรือข้อมูลเชิงปริมาณ ให้สามารถเข้าใจได้ง่าย ในแบบของ กราฟ แผนภูมิ

2.2.5.1 โปรแกรมสำหรับการสร้างแดชบอร์ด (Data Visualization)

Tools ที่นิยมใช้ในตอนนี้ได้แก่ Tableau, Microsoft Power BI, Qlik View, Google Charts, Fusion Charts,Data wrapper และอื่น ๆ อีกมากมาย



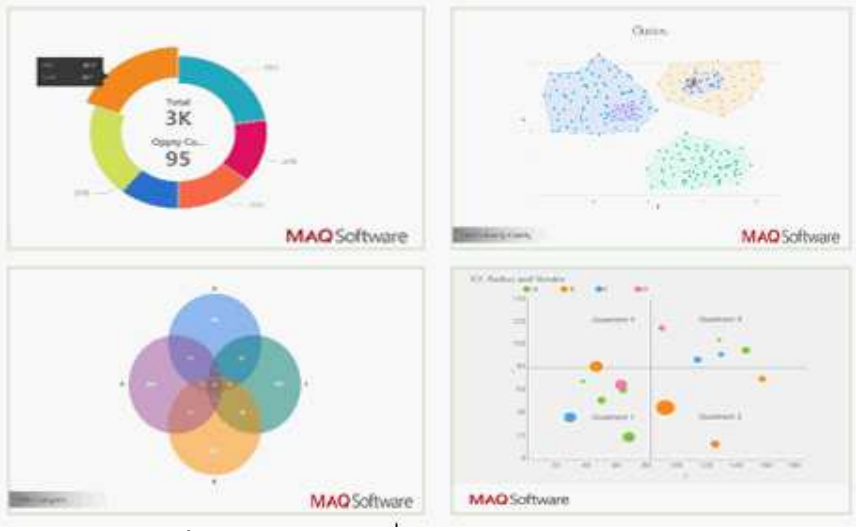
ภาพที่ 2.13 Tools Data Visualization

2.2.5.2 รูปแบบในการใช้ Data Visualization

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

ภาพที่ 2.14 การนำเสนอแบบทิศทางหรือแนวโน้ม (Trending)

การนำเสนอแบบกลุ่มข้อมูล (Classification) เป็นการนำเสนอโดยนำข้อมูลมาจัดเป็นกลุ่มๆ เช่น Donut Chart, Ring Chart, Pie Chart,

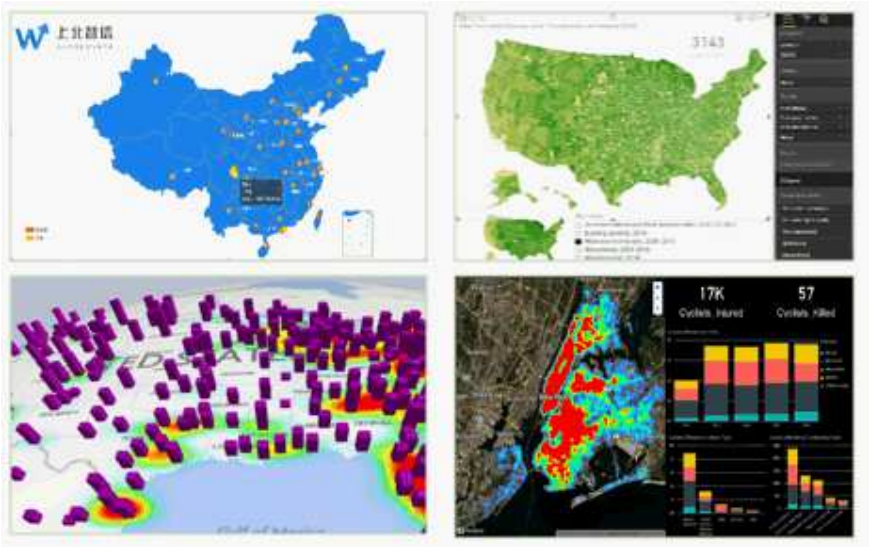


ภาพที่ 2.15 การนำเสนอแบบกลุ่มข้อมูล (Classification)

การนำเสนอเชิงเปรียบเทียบข้อมูล (Comparison) เหมาะสำหรับการนำเสนอที่ต้องการเปรียบเทียบผลการดำเนินงาน เช่นเทียบกับปีที่แล้ว(YoY) เปรียบเทียบกับเป้าที่ตั้งไว้ (Target) ซึ่งกราฟที่เหมาะสมและมักนำมาใช้ เช่น KPI Indicator, Bullet Chart, Power BI Card with state เป็นต้น

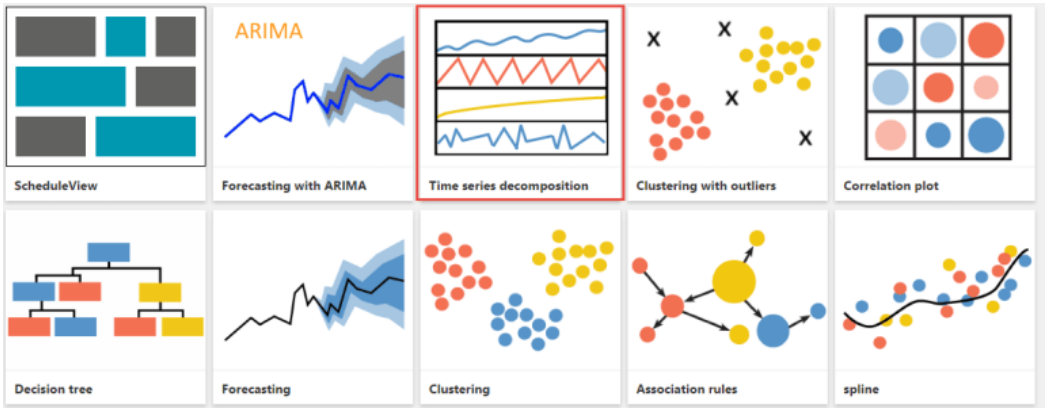
ภาพที่ 2.16 การนำเสนอเชิงเปรียบเทียบข้อมูล (Comparison)

การนำเสนอรูปแบบแผนที่ (Geographical) เหมาะสำหรับนำเสอข้อมูลบนแผนที่ โดยสามารถที่จะนำยอดขาย, รายได้, ความหนาแน่นของประชากร เพื่อFocus กลุ่มลูกค้าในแต่ละพื้นที่ที่เราสนใจ เช่น Globe Map, Google Map, Flow Map เป็นต้น



ภาพที่ 2.17 การนำเสนอรูปแบบแผนที่ (Geographical)

กลุ่มที่ต้องพยากรณ์ล่วงหน้าและการวิเคราะห์ข้อมูลที่ซับซ้อน (Analytics) เราสามารถที่จะใช้ภาษา R หรือ Python ดึงข้อมูลในอดีตมาเพื่อวิเคราะห์เพื่อพยากรณ์อนาคต และนำเสนอข้อมูลที่ได้ในรูปของกราฟ เช่น Association Rules, Clustering, Forecasting Time series, Calculation plot เป็นต้น



ภาพที่ 2.18 กลุ่มที่ต้องพยากรณ์ล่วงหน้าและการวิเคราะห์ข้อมูลที่ซับซ้อน (Analytics)

ข้อดีของ Data Visualization

- ข้อมูลที่มีปริมาณมากมายมหาศาลยากต่อการเข้าใจ เราสามารถทำให้เข้าใจง่ายได้ด้วยรูปภาพ
- ช่วยจัดระเบียบความคิดวิเคราะห์ข้อมูลให้มีความน่าเชื่อถือ
- ประหยัดเวลาในการนำเสนอ

2.2.6 ทฤษฎีเกี่ยวกับต้นไม้ตัดสินใจ

2.2.6.1 ลักษณะต้นไม้การตัดสินใจ

ต้นไม้การตัดสินใจจะทำการจัดกลุ่ม (classify) ชุดข้อมูลนำเข้าในแต่ละกรณี (Instance) แต่ละบัพ (node) ของต้นไม้การตัดสินใจคือตัวแปร (attribute) ต่าง ๆ ของชุดข้อมูล เช่นหากต้องการตัดสินใจว่าจะไปเล่นกีฬาหรือไม่ก็จะมีตัวแปรต้นที่จะต้องพิจารณาคือ ทัศนียภาพ ลม ความชื้น อุณหภูมิ เป็นต้น และมีตัวแปรตามซึ่งเป็นผลลัพธ์จากต้นไม้คือการตัดสินใจว่าจะไปเล่นกีฬารึเปล่า ซึ่งแต่ละตัวแปรนั้นก็จะมีค่าของตัวเอง (value) เกิดเป็นชุดของตัวแปร–ค่าของตัวแปร (attribute–value pair) เช่น ทัศนียภาพเป็นตัวแปร ก็อาจมีค่าได้เป็น ฝนตก แดดออก หรือการตัดสินใจว่าจะไปเล่นกีฬารึเปล่านั้นก็อาจมีค่าได้เป็นใช่ กับ ไม่ใช่ เป็นต้น การทำนายประเภทด้วยต้นไม้ตัดสินใจ จะเริ่มจากบัพราก โดยทดสอบค่าตัวแปรของบัพ แล้วจึงตามกิ่งของต้นไม้ที่กำหนดค่า เพื่อไปยังบัพลูกถัดไป การทดสอบนี้จะกระทำไปจนกระทั่งเจอบัพใบซึ่งจะแสดงผลการทำนาย

2.2.6.1 ส่วนประกอบของต้นไม้ตัดสินใจ ประกอบด้วย

1) โหนด คือ คุณสมบัติต่าง ๆ เป็นจุดที่แยกข้อมูลว่าจะให้ไปในทิศทางใด ซึ่งโหนดที่อยู่จุดสูงสุดเรียกว่า โหนดราก (Root Node)

2) กิ่ง คือ คุณสมบัติของคุณสมบัติในโหนดที่แตกออกมา โดยจานวนของกิ่งจะเท่ากับคุณสมบัติของโหนด

3) ใบ คือ กลุ่มของผลลัพธ์ในการแยกแยะข้อมูล โดยสามารถแสดงส่วนประกอบของต้นไม้ตัดสินใจ

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 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 Node

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 และถามซ้ำ ๆ ไปเรื่อย ๆ เพื่อหาต้นไม้ทั้งต้นด้วยการเขียนโปรแกรมด้วยความสัมพันธ์แบบเวียนเกิด (อังกฤษ: recursion) โดยในการเลือกว่าลักษณะใดดีที่สุดนั้นดูจากค่าของลักษณะเรียกว่าเกนความรู้ (Information gain) ก่อนที่จะรู้จักเกนความรู้จะต้องนิยามค่าหนึ่งที่ใช้บอกความไม่บริสุทธิ์ของข้อมูลก่อนเรียกว่าเอนโทรปี (Entropy) โดยนิยามเอนโทรปีของต้นไม้การตัดสินใจในตัวในเซตของตัวอย่าง S คือ E (S) ดังนี้

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

เมื่อแทนค่า S คือ ตัวอย่างที่ประกอบด้วยชุดของตัวแปรต้นและตัวแปรตามหลายๆกรณี

$p_S(j)$ 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 โดยจะมีค่าเป็นศูนย์เมื่อทุก ๆ กรณีมีผลลัพธ์เพียงแบบเดียว เช่น ใช่ทั้งหมด หรือ ไม่ใช่ทั้งหมด และจะมีค่ามากขึ้นเมื่อเริ่มมีค่าที่แตกต่างกันมากขึ้น หรือจะพูดอีกนัยหนึ่งก็คือเอนโทรปีจะมีค่ามากขึ้นหากข้อมูลไม่บริสุทธิ์ และจะตัดสินใจได้ว่าผลลัพธ์จะเป็นอะไรเมื่อเอนโทรปีเป็น 0 เท่านั้น

เกนความรู้ (Information Gain) ซึ่งจากการนิยามเอนโทรปีข้างต้น ทำให้เราสามารถนิยามลักษณะของตัวแปรต้นที่ดีได้ โดยตัวแปร A จะเป็นตัวแปรต้นที่ดีก็ต่อเมื่อหากว่าแบ่งข้อมูลตัวอย่าง (Example) ออกเป็นชุดๆมีจำนวนชุดตามจำนวนค่าของ A ที่เป็นไปได้เพื่อให้แต่ละกรณี (Instance) ในชุดนั้นมีค่า A เพียงค่าเดียวและค่าเฉลี่ยของเอนโทรปีของชุด

ข้อมูลที่ถูกแบ่งออก (partition) มานั้นต่ำที่สุด เรียกค่าคาดหวังของการลดลงของเอนโทรปีหลังจากข้อมูลถูกแบ่งด้วย A ว่าเกนความรู้ของ A นิยามโดย

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

เมื่อแทนค่า S คือตัวอย่างที่ประกอบด้วยชุดของตัวแปรต้นและตัวแปรตามหลายๆกรณี

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 2.19 Confusion Matrix)

แสดงเป็นสมการได้ ดังนั้น

Sensitivity or Recall คือค่าที่บอกว่า โปรแกรมทำนายได้ว่าจริง เป็นอัตราส่วนเท่าไหร่ของจริงทั้งหมด คำนวณได้จากสมการ

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

Precision คือค่าที่บอกว่า โปรแกรมทำนายว่าจริงถูกต้องเท่าไหร่ คำนวณได้จากสมการ

$$precision = \frac{a}{a+c}$$

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 Tableau Public

ทาโบล์ว (Tableau) Business Intelligence (BI) Software คือซอฟต์แวร์เพื่อการวิเคราะห์ข้อมูลอันหลากหลายเป็น Software ในระดับต้นๆของโลกที่ทำเรื่อง Data Visualization Tableau ทำหน้าที่แปลงข้อมูล (Data) ให้เป็นภาพ (Visualization) สามารถนำข้อมูลออกมาเป็น Chart แบบต่าง ๆ ได้อย่างง่ายดายรวมถึงการนำหลาย Chart มาสร้างเป็น Dashboard และ Story Teller ก็ได้เช่นกัน รองรับการเข้าถึงข้อมูลจากหลายฐานข้อมูล และสามารถนำข้อมูลจากฐานข้อมูลในองค์กรนั้นมาใช้งาน เช่น Excel, Access, Firebird 2.0, IBM DB2, MS SQL, Oracle, MySQL เป็นต้น

ภาพที่ 2.20 การแสดงแดชบอร์ดของ Tableau Public

(ที่มา: https://www.thestephaneandre.com/install-tableau-desktop-and-tableau-public/)

Tableau นั้นถูกพัฒนาอย่างต่อเนื่องเพื่อให้แสดงผลได้บนทุก Platform อย่างง่ายดาย ไม่ว่าจะเป็น Desktop, Mobile หรือแม้แต่ Browser โดยองค์กรสามารถเลือกติดตั้งใช้งานภายในองค์กรเองก็ได้ หรือเช่าใช้ผ่านบริการ Tableau Cloud ก็ได้เช่นกัน ทำให้การทำ Business Intelligence และ Data Analytics ไม่ใช่เรื่องยากอีกต่อไป และง่ายขึ้นจนถึงขนาดผู้บริหารก็สามารถใช้งานด้วยตัวเองได้

2.3.3.2 ลักษณะเด่นของ Tableau Public

- รูปแบบการใช้งานที่ง่ายเพียงคลิ๊กเมาส์ก็สามารถที่จะเปลี่ยนแปลงรายงานได้ตามความต้องการ
- รองรับการเข้าถึงข้อมูลจากหลายฐานข้อมูล และ สามารถนำข้อมูลจากฐานข้อมูลในองค์กรนั้นมาใช้งาน เช่น Excel, Access, Firebird 2.0, IBM DB2, MS SQL Server, Microsoft Power pivot
- รูปแบบการนำเสนอรายงานที่สวยงาม เข้าใจง่าย และง่ายในการนำเสนอต่อผู้บริหาร
- การเข้าถึงหรือการใช้งาน สามารถทำได้โดยง่าย ตั้งแต่ผู้บริหารระดับสูง ถึง พนักงานระดับปฏิบัติการ ทั้งนี้ทั้งนั้น สามารถเพิ่มทักษะและแนวคิดให้กับพนักงานระดับปฏิบัติการ ให้มีแนวคิดเชิงสถิติและการประยุกต์ใช้งานมากขึ้น
- สิ่งที่สำคัญ คือ Engine ของ Tableau Software นั้น พัฒนาจาก VizQL Technology เป็นเทคโนโลยีที่ช่วยให้เห็นภาพของข้อมูลขนาดใดก็ได้เพียงการลากและวาง (Drag and Drop) โดยใช้ข้อมูลจากฐานข้อมูล ตอบสนองในรูปแบบของ กราฟฟิก

2.4 วรรณกรรมที่เกี่ยวข้อง

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

พึงพอใจแก่ผู้บริหารให้สำมารถรับรู้ข้อมูลที่มากขึ้น และสำมารถนำข้อมูลที่ได้ไปใช้ในการสนับสนุนการตัดสินใจ

รัตนา สุวรรณวิชนีย์และปราลี มณีรัตน์ (2560) ระบบบริหารจัดการงานวิจัยแห่งชาติ (National Research Management System : NRMS) ของสำนักงานคณะกรรมการวิจัยแห่งชาติ (วช.) นำมาใช้งานในการบริหารจัดการข้อมูลงานวิจัยและงบประมาณวิจัยของหน่วยงานภาครัฐและแหล่งทุนของหน่วยงานต่าง ๆ แต่ทาง วช. ยังขาดเครื่องมือในการวิเคราะห์ข้อมูลจำนวนมากในระบบ NRMS ให้ได้ทันเวลาและนำเสนอข้อมูลในหลากหลายมิติที่ตรงตามความต้องการใช้ประโยชน์ของผู้ใช้ เพื่อนำเสนอรัฐบาลหรือหน่วยงานนโยบายสำหรับใช้ในการตัดสินใจเชิงนโยบาย ทั้งนี้จึงนำเสนอการพัฒนาระบบรายงานรูปแบบหลายมิติเพื่อสนับสนุนการตัดสินใจเชิงนโยบายของสำนักงานคณะกรรมการวิจัยแห่งชาติ โดยใช้ระบบข่าวกรองทางธุรกิจ (Business Intelligence : BI) พัฒนารายงานรูปแบบหลายมิติตามฟิลด์ เช่น จำแนกตามปีแหล่งทุน กระทรวง หน่วยงาน นโยบายและยุทธศาสตร์การวิจัย และประเด็นการวิจัย เป็นต้น อีกทั้งการนำเสนอรายงานลักษณะแดชบอร์ด (Dashboard) สนับสนุนการตัดสินใจเชิงนโยบายของเจ้าหน้าที่และผู้บริหาร วช. เพื่อให้มองเห็นภาพรวมข้อมูลวิจัยของประเทศได้ การพัฒนาระบบรายงานงานวิจัยนี้ใช้ซอฟต์แวร์ทาโบล (Tableau) ในการพัฒนาระบบรายงาน ผลการประเมินผู้ใช้งานระบบที่พัฒนาขึ้น พบว่าเจ้าหน้า วช.ที่มีความพึงพอใจต่อระบบอยู่ในระดับพอใจ ทำให้เจ้าหน้าที่ วช. สามารถนำเสนอรายงานต่อผู้บริหารได้รวดเร็วขึ้น และหลากหลายมิติมากขึ้น

วิไลลักษณ์ ตรีพืช (2562) การวิจัยนี้มีวัตถุประสงค์เพื่อการวิเคราะห์ข้อมูลผู้บาดเจ็บและเสียชีวิตจากอุบัติเหตุในช่วงเทศกาลสงกรานต์โดยใช้เทคนิคการทำเหมืองข้อมูล โดยใช้ข้อมูลผู้บาดเจ็บและเสียชีวิตในเทศกาลสงกรานต์ ประจำปี 2551 – 2557 จำแนกตามโรงพยาบาลทั่วประเทศ จากเว็บไซต์ data.go.th ซึ่งผู้วิจัยได้ทำการเตรียมข้อมูลให้เหมาะสมสำหรับการจำแนกข้อมูลโดยตัดข้อมูลที่ไม่เกี่ยวข้องออกไปและปรับให้จำนวนข้อมูลของคำตอบหรือผลลัพธ์มีสัดส่วนใกล้เคียงกัน จากนั้นทำการจำแนกข้อมูลโดยกำหนดให้แอททริบิวท์ผลการรักษาเป็น คำตอบที่ต้องการพยากรณ์สำหรับอัลกอริทึมที่ผู้วิจัยเลือกใช้ในการสร้างโมเดล ได้แก่ Naive Bayes, Generalized Linear Model (GLM), Logistic Regression, Decision Tree, Random Forest, และ Gradient Boosted Trees (XGBoost)

วิธวินท์ แสงมณ (2560) งานวิจัยนี้เป็นการประยุกต์ใช้เทคนิคการทำเหมืองข้อมูลในการ ทำนายโอกาสการกลับมารักษาตัวซ้ำของผู้ป่วยโรคเบาหวาน โดยมีวัตถุประสงค์เพื่อ 1. วิเคราะห์ปัจจัยที่ส่งผลต่อโอกาสการ กลับมารักษาซ้ำของผู้ป่วยโรคเบาหวาน 2. สร้างโมเดลสำหรับ ทำนายโอกาสการกลับเข้ามารักษาตัวซ้ำของผู้ป่วยโรคเบาหวาน โดยใช้เทคนิคการทำ

เหมืองข้อมูล 3. เปรียบเทียบประสิทธิภาพ ของโมเดลที่สร้างจากเทคนิค Decision tree ,Naïve Bays และ K–Nearest Neighbors ผลการสร้างโมเดลในการทำนายโอกาสการกลับมา รักษาตัวซ้ำของผู้ป่วยโรคเบาหวาน โดยใช้เทคนิคการทำเหมืองข้อมูล คือโมเดลที่สร้างโดยใช้เทคนิค Decision Tree มีประสิทธิภาพการทำนายสูงสุดคิดเป็นร้อยละ 85.5 ซึ่งสามารถ นำผลที่ได้ไปประยุกต์ใช้ในการสร้างระบบสนับสนุนการตัดสินใจ ในส่วนของแนวทางการรักษาของแพทย์และผู้ป่วยได้ต่อไป

อุษา บิ้กกิ้นส์ (2557) จากการวิจัยเรื่อง ปัจจัยที่มีผลต่อพฤติกรรมเมาแล้วขับของผู้ขับขี่รถจักรยานยนต์ในเขต กรุงเทพมหานคร ซึ่งมีวัตถุประสงค์เพื่อค้นหาปัจจัยที่มีผลต่อการขับขี่รถจักรยานยนต์ในขณะมึน เมา และเพื่อศึกษาพฤติกรรมการเปิดรับสื่อของผู้ขับขี่รถจักรยานยนต์ขณะมึนเมา พฤติกรรมการเปิดรับสื่อของผู้ที่เมาแล้วขับ วัยรุ่นเปิดรับโทรทัศน์และอินเทอร์เน็ตมากที่สุด พวกเขารับรู้ ข้อความจากสติ๊กเกอร์ บิลบอร์ด โปสเตอร์ และโทรทัศน์ พวกเขาสนใจการ กระตุ้นความกลัวผ่านสื่อ ดังนั้น ควรมีวิธีการแก้ปัญหา 2 วิธีคือ การขับเคลื่อนนโยบาย และการ สื่อสารสาธารณะ ได้แก่ ประชาสัมพันธ์โครงการและกิจกรรมต่าง ๆ ทางเว็บไซต์ การนำเสนอข่าว การขับรถขณะมึนเมาและการใช้การรณรงค์สื่อสารการตลาด การรณรงค์ข้อมูลข่าวสาร และการรณรงค์ด้วยการประกาศโฆษณาในบริการสาธารณะ

ภูมิพัฒน์ ดวงกลาง และรัจนา เครือแก้ว (2562) แบบจำลองการทำนายแบบอากาศยานจากข้อมูลเป้าหมายไม่ทราบฝ่ายอัตโนมัติ ผู้วิจัยได้สร้างแบบจำลองการทำนายแบบอากาศยานอัตโนมัติที่ไม่พึ่งพามนุษย์ โดยใช้ข้อมูลจากระบบวิเคราะห์ความเคลื่อนไหวกำลังทางอากาศประเทศรอบบ้านของกรมข่าวทหารอากาศเป็นข้อมูลในการวิจัยเพื่อค้นหาแบบจำลองที่เหมาะสม ด้วยเทคนิคการจำแนกประเภทข้อมูล (Classification) ได้แก่ Naïve Bayes, Deep Learning, Decision Tree และ Random Forest ซึ่งผลการวิจัยพบว่า แบบจำลองที่เหมาะสมที่จะนำไปใช้งานกับกระบวนการวิเคราะห์และระบุเป้าหมายไม่ทราบฝ่ายมากที่สุด คือ แบบจำลองที่ได้จากเทคนิค Random Forest เนื่องจากมีค่าความถูกต้องในการทำนายผลสูง ผู้วิจัยเชื่อมั่นว่าแบบจำลองดังกล่าวจะสามารถเพิ่มประสิทธิภาพให้กระบวนการตัดสินตกลงใจของผู้บังคับบัญชากระทำได้รวดเร็วและเหมาะสมยิ่งขึ้น อีกทั้งผลที่ได้จากงานวิจัยนี้จะสามารถแก้ปัญหาการขาดแคลนบุคลากรผู้เชี่ยวชาญ และลดการพึ่งพาความชำนาญของเจ้าหน้าที่เฉพาะทางลงได้ สอดคล้องกับแนวความคิดการใช้ระบบอาวุธที่ฉลาด (Smart Weapon Systems) ที่นาไปสู่การเลือกกลยุทธ์ที่ฉลาด (Smart Tactics) ของกองทัพอากาศอีกด้วย และทางวิจัยทำการสร้างแบบจำลองสำหรับวิเคราะห์และทำนายแบบอากาศยานจากข้อมูลของ

ระบบวิเคราะห์ความเคลื่อนไหวกำลังทางอากาศประเทศรอบบ้านนั้น จะอ้างอิงกระบวนการมาตรฐาน CRISP-DM

หทัยพร หวังเชย และ ณัฐพร นันทจิระพงศ์ (2561) ระบบสารสนเทศเพื่อการบริหารจัดการรายงานการดำเนินงานของโรงแรมด้วย Power BI ถูกพัฒนาขึ้นโดยมีวัตถุประสงค์เพื่อเปลี่ยนรูปแบบรายงานการวิเคราะห์และสรุปผลการดำเนินงานของโรงแรมจากเอกสารเป็นสารสนเทศประกอบการตัดสินใจของผู้บริหารในรูปแบบของหน้ากระดานที่นำเสนอผลการวิเคราะห์ข้อมูลสำหรับผู้บริหารในมุมมองต่าง ๆ ประกอบด้วยแผนภูมิหลายชนิด หลายมิติ การทำงานแบ่งเป็น 4 ส่วน ได้แก่ 1) การลงชื่อเข้าใช้งานระบบ 2) การประมวลผลข้อมูล 3) การจัดเก็บข้อมูลบน cloud และ 4) การออกรายงาน ซึ่งแบ่งตามข้อกำหนดของฝ่ายบริหารและมิติทางธุรกิจได้เป็น 7 รูปแบบ โดยหลักการและเทคโนโลยีที่นำมาประยุกต์ใช้ในการพัฒนาระบบได้แก่ การวิเคราะห์ข้อมูลขนาดใหญ่ ธุรกิจอัจฉริยะ ระบบสารสนเทศเพื่อการจัดการคลังข้อมูล และกระบวนการสกัดข้อมูล ส่วนเครื่องมือที่ใช้ในการพัฒนาระบบ คือ Power BI desktop, Microsoft SQL server และ Microsoft Visual Studio ผลการพัฒนาพบว่า ระบบสามารถทำงานได้ตามวัตถุประสงค์และขอบเขตที่กำหนดไว้ ผลการสำรวจความพึงพอใจของผู้ใช้ที่มีต่อภาพรวมของระบบพบว่า อยู่ในระดับดี (ค่าเฉลี่ย เท่ากับ 4.5 และส่วนเบี่ยงเบนมาตรฐาน เท่ากับ 0.71) จึงสรุปได้ว่า ระบบสามารถสรุปผลข้อมูลการดำเนินงานของโรงแรมและรายงานได้รวดเร็ว ทำให้ผู้บริหารสามารถวางแผน ตัดสินใจ แก้ไขปัญหา และกำหนดทิศทางการดำเนินงานของโรงแรมได้อย่างมีประสิทธิภาพ

ปัทมา เที่ยงสมบุญ และ นิเวศ จิระวิชิตชัย (2561) การพัฒนาระบบธุรกิจอัจฉริยะเพื่อสนับสนุนการพยากรณ์และการตัดสินใจของผู้บริหาร กรณีศึกษากลุ่มโรงพยาบาล งานวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาระบบธุรกิจอัจฉริยะเพื่อสนับสนุนการพยากรณ์และการตัดสินใจของผู้บริหาร กรณีศึกษากลุ่มโรงพยาบาล โดยใช้ระบบธุรกิจอัจฉริยะ (Business Intelligence) ผ่านโปรแกรมไมโครซอฟท์ เพาเวอร์ บีไอ (Microsoft Power BI) และการจัดการฐานข้อมูลด้วยออราเคิล ดาต้าเบส 11จี (Oracle Database 11g) เพื่อสนับสนุนการวิเคราะห์ข้อมูลการรักษาพยาบาลของผู้บริหาร โดยนาข้อมูลจากระบบสารสนเทศโรงพยาบาล (Hospital Information System) และข้อมูลโรคไม่ติดต่อเรื้อรัง (Non-Communicable Diseases : NCDs) มาสร้างรายงานอัจฉริยะ (Dashboard) โดยระบบสามารถพยากรณ์ข้อมูลและปรับเปลี่ยนมุมมองในการวิเคราะห์ช่วยในการตัดสินใจอย่างรวดเร็วและถูกต้อง จากการประเมินความพึงพอใจของระบบโดยผู้บริหารจานวน 30 คน ได้ค่าเฉลี่ยเท่ากับ 4.15 และค่าส่วนเบี่ยงเบนมาตรฐาน

เท่ากับ 0.67 ซึ่งสรุปได้ว่าระบบธุรกิจอัจฉริยะที่พัฒนาขึ้นมาในงานวิจัยชิ้นนี้ใช้ในการสนับสนุนการตัดสินใจของผู้บริหารได้เป็นอย่างดี

อัจฉราภรณ์ จุฑาผาด (2559) งานวิจัยเรื่องการพัฒนาระบบสารสนเทศเพื่อการพยากรณ์จำนวนนักศึกษาใหม่ โดยใช้กฎการจำแนกต้นไม้ตัดสินใจ มีวัตถุประสงค์คือ 1)เพื่อพัฒนาระบบสารสนเทศเพื่อใช้ในการพยากรณ์จานวนนักศึกษาใหม่ ที่จะเข้าศึกษาต่อในระดับปริญญาตรี คณะบริหารธุรกิจและการบัญชี มหาวิทยาลัยราชภัฏร้อยเอ็ด โดยใช้กฎการจำแนกเทคนิคต้นไม้ตัดสิน (Decision Tree)และ 2) เพื่อวัดประสิทธิภาพและความถูกต้องของตัวต้นแบบที่พัฒนาขึ้นที่นำมาใช้ในการพยากรณ์จำนวนนักศึกษาใหม่ของคณะบริหารธุรกิจและการบัญชี มหาวิทยาลัยราชภัฏร้อยเอ็ด ผลการวิจัยพบว่า 1) การพัฒนาระบบสารสนเทศเพื่อใช้ในการพยากรณ์จำนวนนักศึกษาใหม่ที่จะเข้าศึกษาต่อในระดับปริญญาตรี คณะบริหารธุรกิจและการบัญชี มหาวิทยาลัยราชภัฏร้อยเอ็ด โดยใช้กฎการจำแนกเทคนิคต้นไม้ตัดสิน (Decision Tree) ซึ่งจากการพัฒนาตัวต้นแบบในการพยากรณ์แบ่งออกเป็น 3 วิธีได้แก่ การตรวจสอบไขว้ การแบ่งข้อมูลแบบสุ่มด้วยการแบ่งร้อยละ และการแบ่งชุดข้อมูลและการทดสอบออกจากกัน จำนวนหรือปริมาณของกฎขึ้นอยู่กับความน่าจะเป็นและการกำหนดค่าความเชื่อมั่นผู้วิจัยได้กำหนดไว้ 0.25 หรือ 25 เปอร์เซ็นต์ แสดงว่าวิธีการแบ่งชุดข้อมูลและการทดสอบออกจากกันสามารถนำไปใช้ในการพัฒนาตัวต้นแบบในการพยากรณ์นักศึกษาใหม่ โดยใช้กฎการจำแนกเทคนิคต้นไม้ตัดสินใจที่มีความถูกต้องแม่นยำสูง และเหมาะสมกว่าวิธีอื่น หากนำกฎการจำแนกจำนวน 315 กฎที่ได้มาจากตัวต้นแบบการพยากรณ์มาสร้างเพื่อการทำนาย โดยมีการนำชุดข้อมูลของผู้สมัครในปีถัดไปมาทดสอบกับตัวต้นแบบหรือกฎที่ได้จะช่วยให้สามารถทราบจำนวนนักศึกษาใหม่ในปีนั้นๆได้ และยังเป็นแนวทางกำหนดรูปแบบและวิธีการ การประชาสัมพันธ์ การแนะแนวหรือรับนักศึกษาใหม่ได้อย่างเหมาะสมกับกลุ่มเป้าหมายต่อไป และ 2) การวัดค่าประสิทธิภาพต่าง ๆ ที่วัดได้จะมีค่าใกล้เคียงกันหรือมีค่าเท่ากันในบางตัวต้นแบบ โดยตัวต้นแบบการพยากรณ์ที่พัฒนาด้วยวิธีการแบ่งชุดข้อมูลและการทดสอบออกจากกัน วัดค่าความถูกต้องได้เท่ากับร้อยละ 97.34 ค่าความแม่นยาเท่ากับร้อยละ 98.56 ค่าความระลึกเท่ากับร้อยละ 97.00 และค่าความถ่วงดุลเท่ากับร้อยละ 97.13 ซึ่งมีค่าประสิทธิภาพสูงทุกค่า แสดงว่าตัวต้นแบบที่ใช้ในการพยากรณ์ที่ผู้วิจัยพัฒนาขึ้นมีความถูกต้องแม่นยำในการพยากรณ์ในการรับสมัครนักศึกษาใหม่ โดยที่ผู้วิจัยใช้กฎการจำแนกเทคนิคต้นไม้ตัดสินใจมากที่สุด

เบญจมาศ ปิยะ (2558) งานวิจัยเรื่องการศึกษาพฤติกรรมผู้ใช้งานตู้แช่เชิงพาณิชย์ด้วยเทคนิคต้นไม้ตัดสินใจ เนื่องจากปัญหาการกินพลังงานไฟฟ้าของตู้แช่ที่สูงทำให้ต้นทุนของร้านค้าสูงตามไปด้วยปัจจัยหลัก ๆ คือพฤติกรรมการใช้งานเปิดปิดตู้แช่ของลูกค้า ผู้วิจัยจึง

ทำการศึกษาพฤติกรรมการใช้งานตู้แช่ของผู้ใช้งานในร้านค้าโดยใช้อุปกรณ์ Arduino ในการบันทึกข้อมูลการใช้งานและข้อมูลการใช้พลังงานของตู้แช่ เพื่อนำข้อมูลที่ได้มาไปวิเคราะห์หาวิธีการที่จะลดการใช้พลังงานไฟฟ้าของตู้แช่ให้ได้อย่างเหมาะสมและนำเทคนิคการจำแนกข้อมูลต้นไม้การตัดสินใจมาเพื่อช่วยในการจำแนกข้อมูลแต่ละส่วน ผลที่ได้จากการดำเนินงานวิจัยนี้คือ จะได้ระบบแจ้งเตือนการเปิดประตูตู้แช่ด้วยเสียงเมื่อเปิดประตูตู้แช่เกินระยะเวลาที่กำหนดเอาไว้ จากการศึกษาพฤติกรรมการใช้งานเปิดปิดตู้แช่เพื่อนำมาปรับเพิ่มเสียงแจ้งเตือนผู้ใช้งานไม่ให้เปิดใช้งานเป็นเวลานาน และส่งผลกับการประหยัดพลังงานไฟฟ้าได้จริง โดยมีการเก็บข้อมูลการใช้งานจริงและนำเอาข้อมูลมาประมวลผลโดยโปรแกรม WEKA และได้ช่วงเวลาที่เหมาะสมกับการแจ้งเตือน เวลาที่ตั้งในการแจ้งเตือนคือไม่เกิน 2 นาทีผลที่ได้คือมีเวลาการเปิดใช้งานที่น้อยลงและทำให้ประหยัดพลังงานไฟฟ้า หลังจากได้ทำการตั้งเวลาแจ้งเตือนการเปิดปิดประตูตู้แช่ไปแล้วจากผลการเก็บข้อมูล พบว่ามีการเปิดปิดประตูตู้แช่ในระยะเวลาที่สั้นลงจากเดิมก่อนที่จะทำการติดตั้งสัญญาณเตือนการใช้งาน 18.75 % และมีผลทำให้การใช้พลังงานไฟฟ้าของตู้แช่ลดลงจากเดิม 4.4 %

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 เพื่อทำการเปรียบเทียบ และเลือกผลลัพธ์ที่ดีที่สุดมานำเสนอ จากนั้นนำข้อมูลสารสนเทศมาทำการแสดงผลแบบ visualization ในรูปแบบของภาพโดยใช้โปรแกรม Tableau Public เผยแพร่บน web browser ที่เป็นที่นิยมในยุคอินเตอร์เน็ตคือการเผยแพร่ทางสื่อออนไลน์ โดยใช้ภาษา HTML และ CSS ในการเขียนเว็บไซต์ขึ้นมา

บทที่ 3
วิธีการดำเนินงานโครงงาน

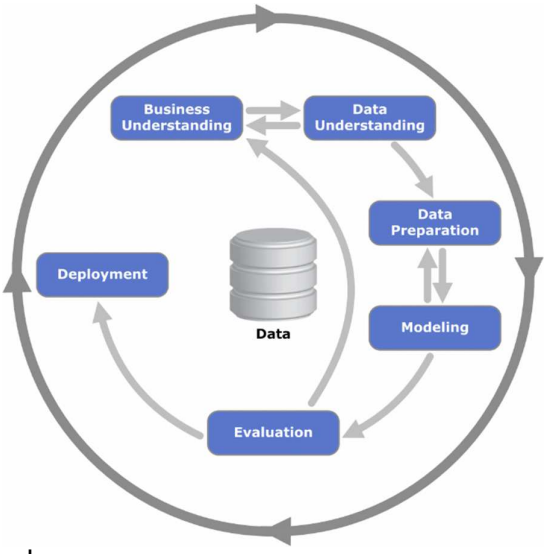
โครงงานเรื่อง การวิเคราะห์ข้อมูลผู้บาดเจ็บและเสียชีวิตในช่วงเทศกาลปีใหม่ประจำปี 2551 – 2558 เพื่อใช้สำหรับเผยแพร่ข้อมูลบนเว็บไซต์ ในบทนี้จะเป็นการวิเคราะห์ข้อมูลด้วยเทคนิคทางดาต้าไมนิ่ง ซึ่งมีกระบวนการวิเคราะห์ที่สำคัญหลายขั้นตอน เมื่อเสร็จสิ้นจากกระบวนการวิเคราะห์ข้อมูลแล้วจะเป็นการออกแบบเว็บไซต์ และออกแบบรูปแบบการแสดงผลและบทสรุปจากวิธีการดำเนินงาน

3.1 การวิเคราะห์ข้อมูลด้วย CRISP–DM

3.2 การออกแบบเว็บไซต์

3.3 บทสรุป

3.1 การวิเคราะห์ข้อมูลด้วย CRISP–DM



ภาพที่ 3.1 กระบวนการวิเคราะห์ข้อมูล ด้วย CRISP–DM

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

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

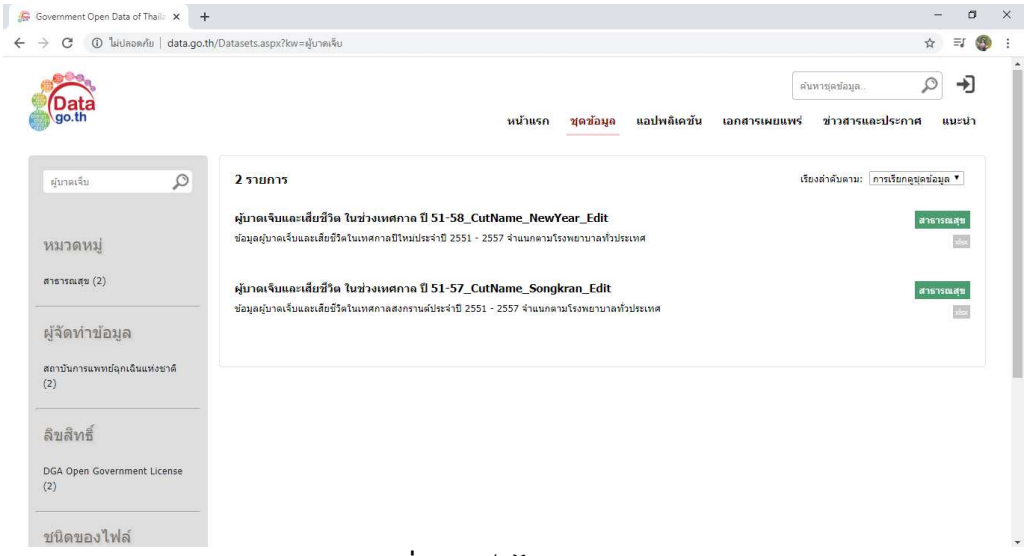
ขั้นตอนในรูปจะเป็นขั้นตอนที่ต่อเนื่องกันนั่นคือ ขั้นตอนถัดไปจะรอผลลัพธ์จากขั้นตอนก่อนหน้า ซึ่งแสดงด้วยลูกศรที่เชื่อมระหว่างกระบวนการนั้น ๆ ประกอบด้วย 6 ขั้นตอนหลัก ดังนี้

3.1.1 รู้จักและเข้าใจในธุรกิจ (Business Understanding) เป็นขั้นตอนแรกของกระบวนการที่มุ่งเน้นไปที่การทำความเข้าใจกระบวนการทางธุรกิจโดยรวม

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

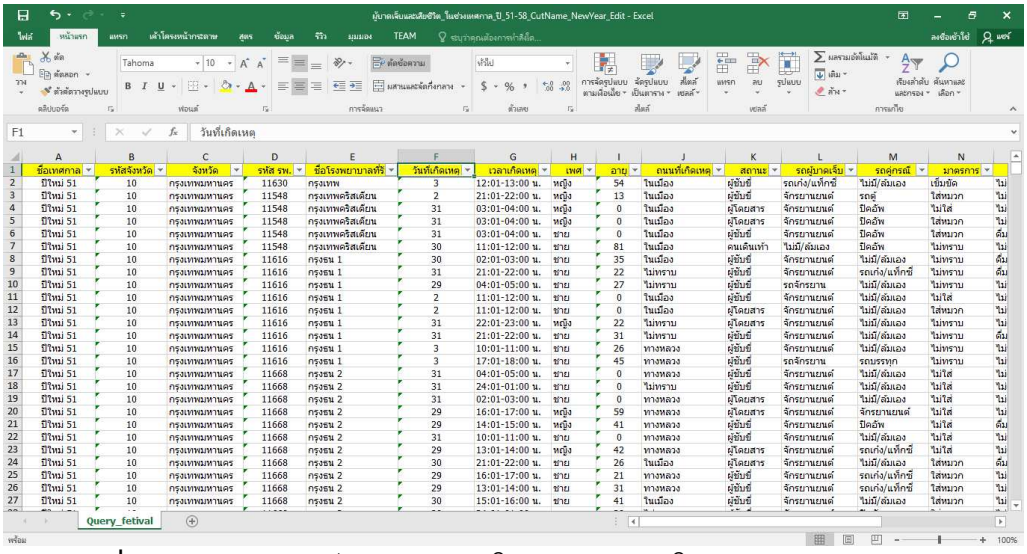
3.1.2 จัดเก็บและรวบรวมข้อมูลให้ครบ (Data Understanding) ขั้นตอนการจัดเก็บและรวบรวมข้อมูล ตลอดจนการพิจารณาตรวจสอบความถูกต้องของข้อมูลที่ได้รับ โดยเลือกว่าจะใช้ข้อมูลทั้งหมดหรือบางส่วนในการวิเคราะห์ให้สอดรับกับวัตถุประสงค์ที่กำหนดไว้

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



ภาพที่ 3.2 เว็บไซต์ data.go.th

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



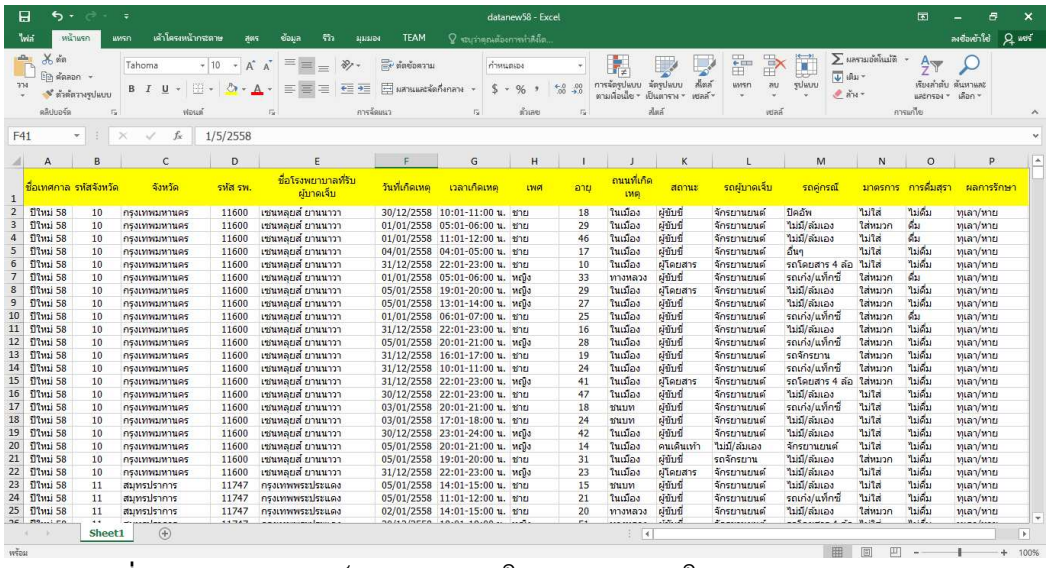
ชื่อเทศกาล	รหัสจังหวัด	จังหวัด	รหัส รพ.	ชื่อโรงพยาบาลที่รั	วันที่เกิดเหตุ	เวลาเกิดเหตุ	เพศ	อายุ	ถนนที่เกิดเหตุ	สถานะ	รถผู้บาดเจ็บ	รถคู่กรณี	มาตรการ
ปีใหม่ 51	10	กรุงเทพมหานคร	11630	กรุงเทพ	3	12:01-13:00 น.	หญิง	54	ในเมือง	ผู้ขับขี่	รถเก๋ง/แท็กซี่	ไม่มี/ดื่มเอง	เข็มขัด
ปีใหม่ 51	10	กรุงเทพมหานคร	11548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

ภาพที่ 3.3 ข้อมูลผู้บาดเจ็บและเสียชีวิตในช่วงเทศกาลปีใหม่ประจำปี 2551 – 2558

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

3.1.3.1 ทำการคัดเลือกข้อมูล (Data Selection) คือการคัดเลือกข้อมูลที่เหมาะสมเพื่อนำมาใช้ในการวิเคราะห์ข้อมูล

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



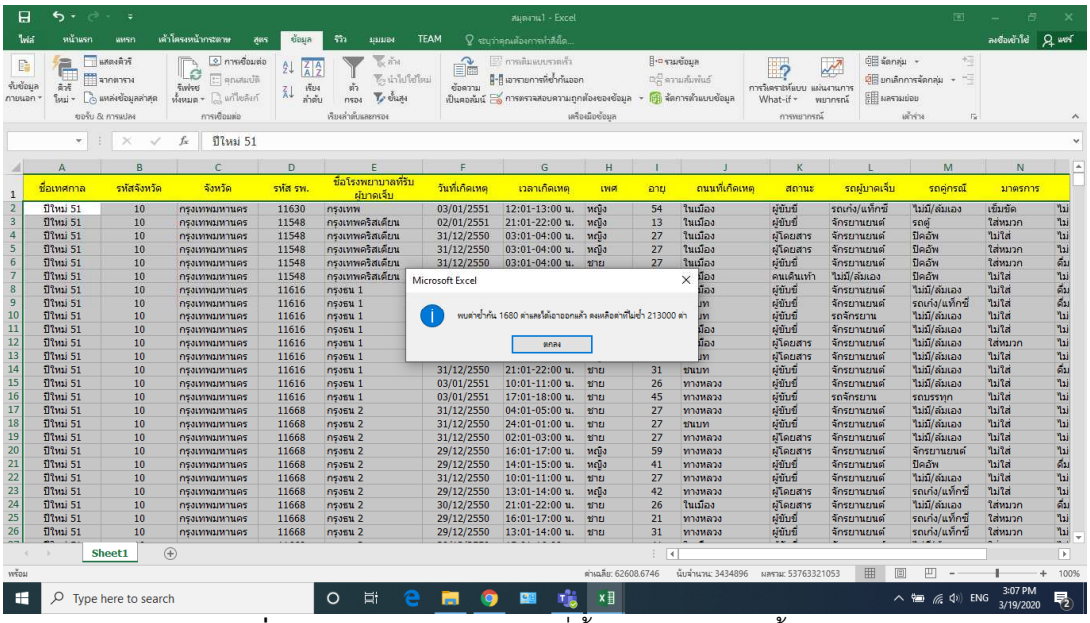
ชื่อเทศกาล	รหัสจังหวัด	จังหวัด	รหัส รพ.	ชื่อโรงพยาบาลที่รับผู้บาดเจ็บ	วันที่เกิดเหตุ	เวลาเกิดเหตุ	เพศ	อายุ	ถนนที่เกิดเหตุ	สถานะ	รถผู้บาดเจ็บ	รถคู่กรณี	มาตการ	การดื่มสุรา	ผลการรักษา
ปีใหม่ 58	10	กรุงเทพมหานคร	11600	เซนหลุยส์ ยานนาวา	30/12/2558	10:01-11:00 น.	ชาย	18	ในเมือง	ผู้ขับขี่	จักรยานยนต์	ปิคอัพ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

ภาพที่ 3.4 ข้อมูลผู้บาดเจ็บและเสียชีวิตในช่วงเทศกาลปีใหม่ประจำปี 2551 – 2558

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

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

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



ภาพที่ 3.5 การตรวจสอบข้อมูลที่ซ้ำซ้อนของข้อมูลทั้งหมด

- รหัสรพ. 10707 เป็นรหัสของโรงพยาบาลมหาสารคาม คณะผู้จัดทำพบว่ามีการใช้รหัสโรงพยาบาลซ้ำกัน จึงได้ทำการตรวจสอบข้อมูลและทำการแก้ไขให้เป็นชื่อโรงพยาบาลมหาสารคามทั้งหมด

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

ภาพที่ 3.6 ข้อมูลผิดพลาดจากการใช้รหัสโรงพยาบาลซ้ำกัน

- รหัสรพ. 13780 โดยส่วนใหญ่ใช้กับชื่อโรงพยาบาลมหาราชนครเชียงใหม่ คณะผู้จัดทำพบว่ามีการใช้รหัสโรงพยาบาลมากกว่าหนึ่งรหัส จึงทำการเปรียบเทียบรหัสที่ใช้มากที่สุด และแก้ไขข้อมูลรหัสโรงพยาบาลให้อยู่ในรหัสเดียวกัน

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

ภาพที่ 3.7 ข้อมูลผิดพลาดจากการใช้รหัสโรงพยาบาลมากกว่าหนึ่งรหัส

- รหัสรพ. 10666 เป็นรหัสของโรงพยาบาลมหาราชนครราชสีมา คณะผู้จัดทำพบว่ามีการใช้รหัสโรงพยาบาลมากกว่าหนึ่งรหัส จึงทำการเปรียบเทียบรหัสที่ใช้มากที่สุด และทำการแก้ไขเพื่อให้ได้ข้อมูลรหัสโรงพยาบาลและชื่อโรงพยาบาลที่ตรงกัน

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

ภาพที่ 3.8 ข้อมูลผิดพลาดจากการใช้รหัสโรงพยาบาลมากกว่าหนึ่งรหัส

- รหัสรพ. 11720 เป็นรหัสของโรงพยาบาลเทพธารินทร์ คณะผู้จัดทำพบว่ามีการใช้รหัสโรงพยาบาลมากกว่าหนึ่งรหัส และการใช้ชื่อโรงพยาบาลผิด จึงทำการเปรียบเทียบรหัส และชื่อที่ใช้มากที่สุด และทำการแก้ไขเพื่อให้ได้ข้อมูลรหัสโรงพยาบาลและชื่อโรงพยาบาลที่ตรงกัน

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

ภาพที่ 3.9 ข้อมูลผิดพลาดของชื่อข้อมูล และรหัสโรงพยาบาลที่ใช้มากกว่าหนึ่งรหัส

- รหัสรพ. 11730 เป็นรหัสของโรงพยาบาลไทยนครินทร์ บางนา คณะผู้จัดทำพบว่ามีการใช้รหัสโรงพยาบาลมากกว่าหนึ่งรหัส และสถานที่โรงพยาบาลเดียวกัน แต่มีการใช้ชื่อโรงพยาบาลไม่เหมือนกัน จึงทำการเปรียบเทียบรหัสที่ใช้มากที่สุด และทำการแก้ไขเพื่อให้ได้ข้อมูลรหัสโรงพยาบาลและชื่อโรงพยาบาลที่ตรงกัน

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

ภาพที่ 3.10 ข้อมูลผิดพลาดของชื่อข้อมูล และรหัสโรงพยาบาลที่ใช้มากกว่าหนึ่งรหัส

- รหัสรพ. 10669 เป็นรหัสของโรงพยาบาลสรรพสิทธิประสงค์

คณะผู้จัดทำพบว่ามีการใช้รหัสโรงพยาบาลมากกว่าหนึ่งรหัส จึงทำการเปรียบเทียบรหัสที่ใช้มากที่สุด และทำการแก้ไขเพื่อให้ได้ข้อมูลรหัสโรงพยาบาลและชื่อโรงพยาบาลที่ตรงกัน

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

ภาพที่ 3.11 ข้อมูลผิดพลาดจากการใช้รหัสโรงพยาบาลมากกว่าหนึ่งรหัส

- รหัสรพ. 11656 เป็นรหัสของโรงพยาบาลเกษมราษฎร์ บางแค

คณะผู้จัดทำพบว่ามีการใช้รหัสโรงพยาบาลมากกว่าหนึ่งรหัส และมีการใช้ชื่อโรงพยาบาลไม่เหมือนกัน จึงทำการเปรียบเทียบข้อมูลที่ใช้มากที่สุด และทำการแก้ไขเพื่อให้ได้รหัสโรงพยาบาลและชื่อโรงพยาบาลที่ตรงกัน

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

ภาพที่ 3.12 ข้อมูลผิดพลาดของชื่อข้อมูล และรหัสโรงพยาบาลที่ใช้มากกว่าหนึ่งรหัส

- รหัสรพ. 11262 เป็นรหัสของโรงพยาบาลสามง่าม คณะผู้จัดทำพบว่ามีการใช้รหัสโรงพยาบาลซ้ำกัน จึงได้ทำการตรวจสอบข้อมูลและทำการแก้ไขให้เป็นชื่อโรงพยาบาลสามง่ามทั้งหมด

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

ภาพที่ 3.13 ข้อมูลผิดพลาดจากการใช้รหัสโรงพยาบาลซ้ำกัน

- รหัสรพ. 10937 เป็นรหัสของโรงพยาบาลห้วยทับทิม คณะผู้จัดทำพบว่ามีการใช้รหัสโรงพยาบาลและชื่อโรงพยาบาลไม่ถูกต้องในกลุ่มข้อมูล จึงทำการแก้ไขเพื่อให้ได้ข้อมูลรหัสโรงพยาบาลและชื่อโรงพยาบาลที่ตรงกัน

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

ภาพที่ 3.14 ข้อมูลผิดพลาดของช่วงข้อมูลในโรงพยาบาลห้วยทับทัน

- คณะผู้จัดทำทำการแก้ไขสถานะการขับขี่ที่มีการใช้คำสะกดผิดให้ถูกต้องตามหลักของภาษาไทยจาก “ผู้ซับขี่” เป็น “ผู้ขับขี่”

สถานะ
ผู้ขับขี่
ผู้ขับขี่
ผู้ขับขี่
ผู้ขับขี่
ผู้ขับขี่
ผู้ซับขี่
ผู้ซับขี่
ผู้ซับขี่

ภาพที่ 3.15 ข้อมูลผิดพลาดของผู้ขับขี่

- คณะผู้จัดทำทำการตรวจสอบสถานะของคนเดินเท้า พบว่ามาตรการเป็นการใส่เข็มขัด, ใส่หมวก และมีรถที่ประสบอุบัติเหตุ ซึ่งเป็นข้อมูลที่ขัดแย้งกัน ทางคณะผู้จัดทำจึงได้แก้ไขมาตรการให้เป็นไม่ใส่ และรถผู้บาดเจ็บเป็นอื่น ๆ

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

ภาพที่ 3.16 ข้อมูลที่ขัดแย้งกันของคนเดินเท้า

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

- คณะผู้จัดทำทำการลบข้อมูลประเภทอายุ 1 ปีที่มีสถานะเป็นผู้ขับขี่

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

ภาพที่ 3.17 การลบข้อมูลที่ขัดแย้งกัน

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

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

ภาพที่ 3.18 ข้อมูลอายุที่เป็น Outlier

- คณะผู้จัดทำตรวจสอบข้อมูลอายุพบว่ามีข้อมูลที่เป็น 0 จึงหาค่าเฉลี่ยของอายุในแต่ละปีว่ามีการเกิดอุบัติเหตุกับบุคคลที่มีอายุในช่วงใดมากที่สุด เพื่อนำค่าเฉลี่ยที่ได้ไปแทนข้อมูลที่เป็น 0

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

ภาพที่ 3.19 การหาค่าเฉลี่ยของอายุในแต่ละปี แทนข้อมูลที่เป็น 0

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

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

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

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

คณะผู้จัดทำวิเคราะห์ข้อมูลด้วยการทำเหมืองข้อมูลแบบ Data Classification เพื่อใช้ทำนายแนวโน้มการเกิดขึ้นของปัจจัยที่ก่อให้เกิดผู้บาดเจ็บและเสียชีวิตจากเทคนิคการจำแนกประเภทข้อมูล ด้วยการสร้างโมเดล Decision Tree เพื่อจัดกลุ่มข้อมูลที่มีความสัมพันธ์กัน ด้วยชุดข้อมูลที่คัดเลือก ดังนี้

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

ภาพที่ 3.20 ชุดข้อมูลที่คัดเลือกมาวิเคราะห์ข้อมูล

จากรูปภาพที่ ประกอบด้วย 5 แอตทริบิวต์ คือ

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

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

ตารางที่ 3.1 สมการการหาความสัมพันธ์ของแอตทริบิวต์

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

การคำนวณค่าแต่ละแอตทริบิวต์เทียบกับคลาสเพื่อหาแอตทริบิวต์ที่มีค่า IG มากที่สุดมาเป็น root ของ decision tree กับจำนวนข้อมูลทั้งหมดโดยใช้ผลลัพธ์เป็นดื่ม และไม่ดื่ม ดังนี้

1) คำนวณค่า IG ของแอตทริบิวต์ เพศ จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

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

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

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

$= -[0.53 + 0.40]$

$= 0.93$

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

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

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

$= -[0.51 + 0.49]$

$= 1$

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

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

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

$= -[0.31 + 0.13]$

$= 0.44$

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

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

$= 0.93 - [0.68 + 0.14]$

$= 0.93 - 0.82$

$= 0.11$

2) คำนวณค่า IG ของแอตทริบิวต์ สถานะ จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

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

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

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

$= -[0.53 + 0.40]$

$= 0.93$

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

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

= –[0.20 × –2.32 + 0.80 × –0.32]

= –[0.46 + 0.26]

= 0.72

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

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

= –[0.41 × –1.29 + 0.59 × –0.76]

= –[0.53 + 0.45]

= 0.98

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

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

= –[0.18 × –2.47 + 0.82 × –0.29]

= –[0.44 + 0.24]

= 0.68

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

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

= 0.93 – [0.19 + 0.7 + 0.02]

= 0.93 – 0.91

= 0.02

3) คำนวณค่า IG ของแอตทริบิวต์ มาตรการ จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

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

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

= –[0.35 × –1.51 + 0.65 × –0.62]

= –[0.53 + 0.40]

= 0.93

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

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

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

$= -[0.48 + 0.28]$

$= 0.76$

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

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

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

$= -[0.48 + 0.28]$

$= 0.76$

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

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

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

$= -[0.47 + 0.3]$

$= 0.77$

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

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

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

$= -[0.53 + 0.43]$

$= 0.96$

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

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

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

$= 0.93 - 0.92$

$= 0.01$

4) คำนวณค่า IG ของแอตทริบิวต์ ผลการรักษา จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

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

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

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

= –[0.53 + 0.40]

= 0.93

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

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

= –[0.35 × –1.51 + 0.65 × –0.62]

= –[0.53 + 0.40]

= 0.93

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

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

= –[0.27 × –1.74 + 0.73 × –0.51]

= –[0.47 + 0.37]

= 0.84

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

= 0.93 – [0.98 × 0.93 + 0.02 × 0.84]

= 0.93 – [0.91 + 0.02]

= 0.93 – 0.93

= 0

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

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

1) คำนวณค่า IG ของแอตทริบิวต์เพศชายต่อสถานะ ได้ดังนี้

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

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

= –[0.48 × –1.06 + 0.52 × –0.94]

= –[0.51 + 0.49]

= 1

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

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

= –[0.35 × –1.51 + 0.65 × –0.62]

= –[0.53 + 0.40]

= 0.93

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

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

= –[0.51 × –0.97 + 0.49 × –1.03]

= –[0.49 + 0.50]

= 0.99

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

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

= –[0.25 × –2 + 0.75 × –0.42]

= –[0.5 + 0.32]

= 0.82

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

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

= 1 – [0.17 + 0.78 + 0.02]

= 1 – 0.97

= 0.03

2) คำนวณค่า IG ของแอตทริบิวต์เพศชายต่อมาตรการ ได้ดังนี้

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

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

= –[0.48 × –1.06 + 0.52 × –0.94]

= 1

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

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

= –[0.30 × –1.74 + 0.70 × –0.51]

= –[0.52 + 0.36]

= 0.88

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

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

= –[0.34 × –1.56 + 0.66 × –0.6]

= –[0.53 + 0.39]

= 0.92

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

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

= –[0.33 × –1.6 + 0.67 × –0.58]

= –[0.53 + 0.39]

= 0.92

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

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

= –[0.52 × –0.94 + 0.48 × –1.06]

= –[0.49 + 0.51]

= 1

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

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

= 1 – [0.02 + 0.06 + 0.14 + 0.76]

= 1 – 0.98

= 0.02

3) คำนวณค่า IG ของแอตทริบิวต์เพศชายต่อผลการรักษา ได้ดังนี้

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

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

= –[0.48 × –1.06 + 0.52 × –0.94]

= –[0.51 + 0.49]

= 1

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

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

= –[0.48 × –1.06 + 0.52 × –0.94]

= –[0.51 + 0.49]

= 1

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

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

= –[0.35 × –1.51 + 0.65 × –0.62]

= –[0.53 + 0.40]

= 0.93

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

= 1 – [0.98 × 1 + 0.02 × 0.93]

= 1 – [0.44 + 0.0036]

= 1 – 1

= 0

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

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

1) คำนวณค่า IG ของแอตทริบิวต์เพศชาย และสถานะ “ผู้ขับขี่” ต่อมาตรการ ได้ดังนี้

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

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

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

$= -[0.49 + 0.50]$

$= 0.99$

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

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

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

$= -[0.52 + 0.38]$

$= 0.90$

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

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

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

$= -[0.50 + 0.49]$

$= 0.99$

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

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

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

$= -[0.53 + 0.39]$

$= 0.92$

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

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

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

$= -[0.47 + 0.51]$

$= 0.98$

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

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

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

$= 0.06$

1.1) คำนวณค่า IG ของเพศชาย และสถานะ “ผู้ขับขี่” ต่อผลการรักษา ได้ดังนี้

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

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

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

$= -[0.49 + 0.50]$

$= 0.99$

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

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

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

$= -[0.48 + 0.50]$

$= 0.98$

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

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

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

$= -[0.52 + 0.40]$

$= 0.92$

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

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

$= 0.99 - [0.96 + 0.01]$

$= 0.99 - 0.97$

$= 0.02$

2) คำนวณค่า IG ของแอตทริบิวต์เพศชาย และสถานะ “ผู้โดยสาร” ต่อมาตรการ ได้ดังนี้

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

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

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

$= -[0.53 + 0.40]$

$= 0.93$

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

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

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

$= -[0.49 + 0.30]$

$= 0.79$

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

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

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

$= -[0.49 + 0.30]$

$= 0.79$

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

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

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

$= -[0.52 + 0.38]$

$= 0.90$

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

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

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

$= -[0.53 + 0.43]$

$= 0.96$

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

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

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

$= 0.93 - 0.91$

$= 0.02$

2.1) คำนวณค่า IG ของเพศชาย และสถานะ “ผู้โดยสาร” ต่อผลการรักษา ได้ดังนี้

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

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

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

$= -[0.53 + 0.40]$

$= 0.93$

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

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

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

$= -[0.53 + 0.40]$

$= 0.93$

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

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

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

$= -[0.53 + 039]$

$= 0.92$

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

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

$= 0.93 - [0.92 + 0.00]$

$= 0.93 - 0.92$

$= 0.01$

3) คำนวณค่า IG ของแอตทริบิวต์เพศชาย และสถานะ “คนเดินเท้า” ต่อมาตรการ ได้ดังนี้

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

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

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

$= -[0.50 + 0.31]$

$= 0.81$

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

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

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

$= -[0 + 0]$

$= 0$

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

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

= –[0 × 1 + 0 × 1]

= –[0 + 0]

= 0

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

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

= –[0 × 1 + 0 × 1]

= –[0 + 0]

= 0

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

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

= –[0.25 × –2 + 0.75 × –0.42]

= –[0.50 + 0.31]

= 0.81

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

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

= 0.81 – [0 + 0 + 0 + 0.81]

= 0.81 – 0.81

= 0

3.1) คำนวณค่า IG ของเพศชาย และสถานะ “คนเดินเท้า” ต่อผลการรักษา ได้ดังนี้

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

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

= –[0.25 × –2 + 0.75 × –0.42]

= –[0.50 + 0.31]

= 0.81

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

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

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

$= -[0.50 + 0.31]$

$= 0.81$

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

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

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

$= -[0.51 + 0.32]$

$= 0.83$

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

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

$= 0.81 - [0.76 + 0.04]$

$= 0.81 - 0.80$

$= 0.01$

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

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

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

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

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

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

$= 0.91$

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

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

= –[0.32 × –1.64 + 0.68 × –0.56]

= –[0.53 + 0.38]

= 0.91

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

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

= –[0.22 × –2.18 + 0.78 × –0.36]

= –[0.48 + 0.28]

= 0.76

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

= 0.91 – [0.98 × 0.91 + 0.02 × 0.76]

= 0.91 – [0.89 + 0.02]

= 0.91 – 0.91

= 0

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

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

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

= –[0.49 × –1.03 + 0.51 × –0.97]

= –[0.50 + 0.49]

= 0.99

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

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

= –[0.51 × –0.97 + 0.49 × –1.03]

= –[0.49 + 0.50]

= 0.99

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

= –[0.21 × $\log_2$(0.21) + 0.79 × $\log_2$(0.79)]

= –[0.21 × –2.25 + 0.79 × –0.34]

= 0.73

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

= 0.99 – [0.94 × 0.99 + 0.06 × 0.73]

= 0.99 – [0.93 + 0.04]

= 0.99 – 0.97

= 0.02

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

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

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

= –[0.33 × –1.6 + 0.67 × –0.58]

= –[0.53 + 0.39]

= 0.92

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

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

= –[0.33 × –1.6 + 0.67 × –0.58]

= –[0.53 + 0.39]

= 0.92

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

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

= –[0.30 × –1.75 + 0.70 × –0.51]

= –[0.53 + 0.36]

= 0.89

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

= 0.92 – [0.99 × 0.92 + 0.01 × 0.89]

= 0.92 – [0.91 + 0.01]

= 0.92 – 0.92

= 0

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

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

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

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

$= -[0.47 + 0.51]$

$= 0.98$

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

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

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

$= -[0.46 + 0.52]$

$= 0.98$

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

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

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

$= -[0.53 + 0.43]$

$= 0.96$

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

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

$= 0.98 - [0.96 + 0.01]$

$= 0.98 - 0.97$

$= 0.01$

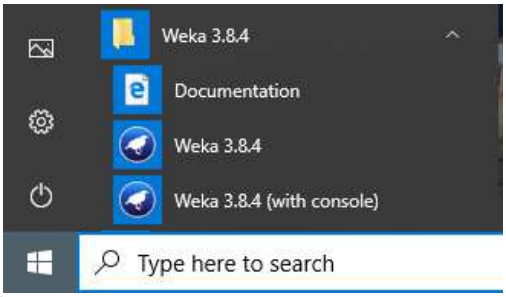
จากการคำนวณค่า IG ของแอตทริบิวต์เพศชายและสถานะ “ผู้ขับขี่” และมาตรการต่อผลการรักษาที่เป็นแอตทริบิวต์สุดท้าย คณะผู้จัดทำพบว่ามาตรการ “ไม่ใส่เข็มขัด” และ “ไม่ใส่หมวก” มีความสัมพันธ์กับผลการรักษามากที่สุดอยู่ที่ 0.02 และ 0.01 ตามลำดับ และพบว่าข้อมูลในแต่ละโหนดมีคลาสคำตอบเดียวกันแล้ว คือ ผลลัพธ์เป็นดื่ม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

3.1.5 การประเมินผล (Evaluation) เป็นขั้นตอนก่อนนำผลลัพธ์ที่ได้จากขั้นตอนที่ 4 ไปใช้งาน ด้วยการวัดประสิทธิผลของผลลัพธ์ที่ได้กับวัตถุประสงค์ที่ตั้งไว้ในขั้นตอนแรก ว่ามี

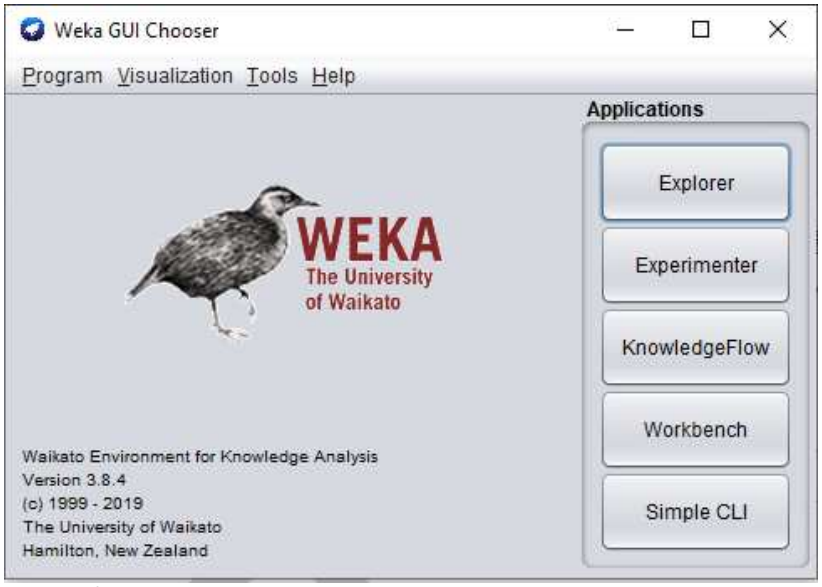
นัยสำคัญหรือความน่าเชื่อถือมากน้อยเพียงใด ด้วยการประเมินผลจากโปรแกรมว่าถูกต้องหรือไม่

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

ขั้นตอนที่ 1 คลิกเปิดโปรแกรม Weka 3.8.4 ขึ้นมา

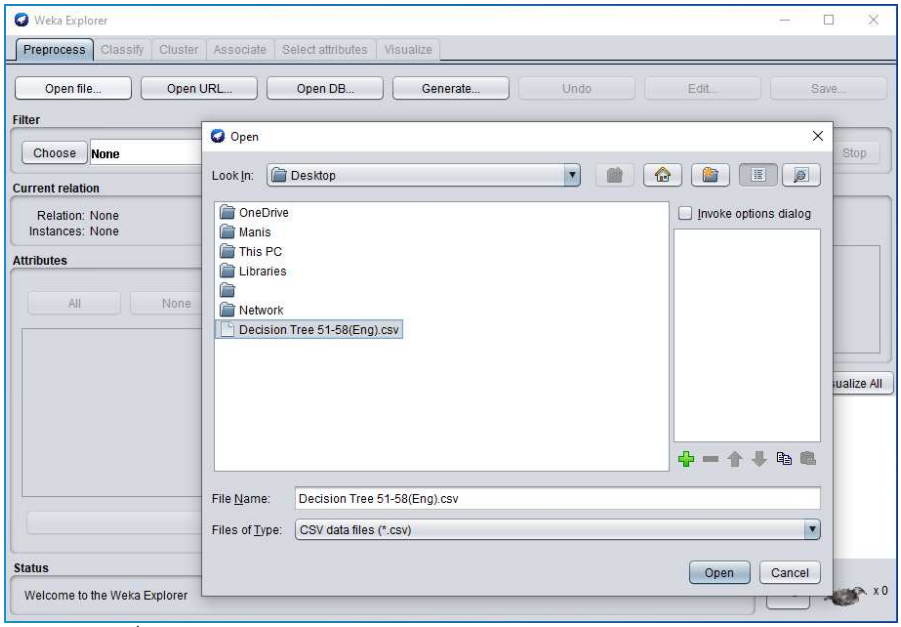


ภาพที่ **3.21** แสดงการเปิดโปรแกรม Weka เวอร์ชัน 3.8.4

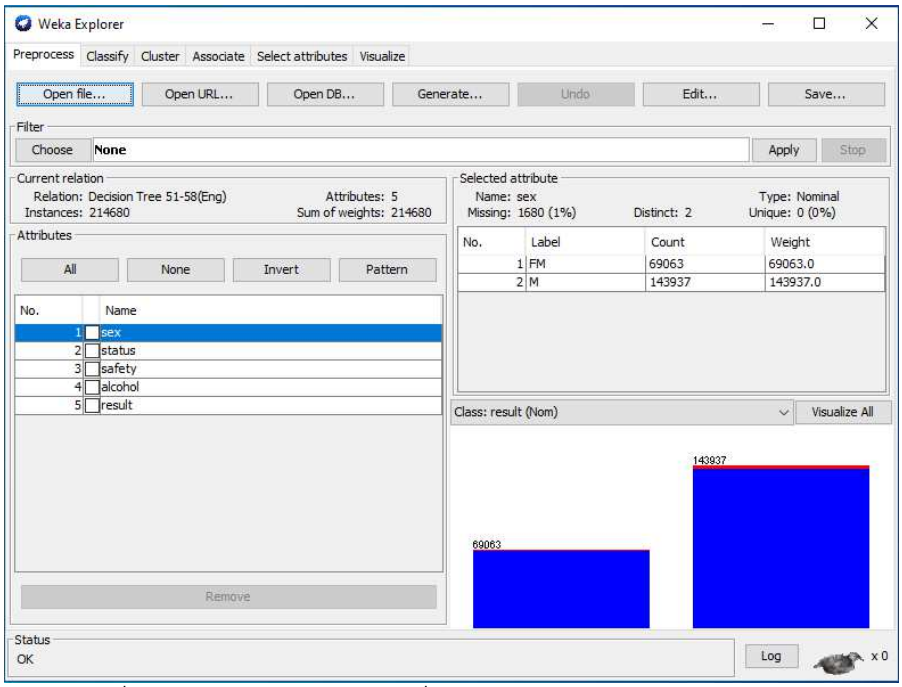


ภาพที่ **3.22** แสดงการเข้าหน้าจอโปรแกรม Weka เวอร์ชัน 3.8.4

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

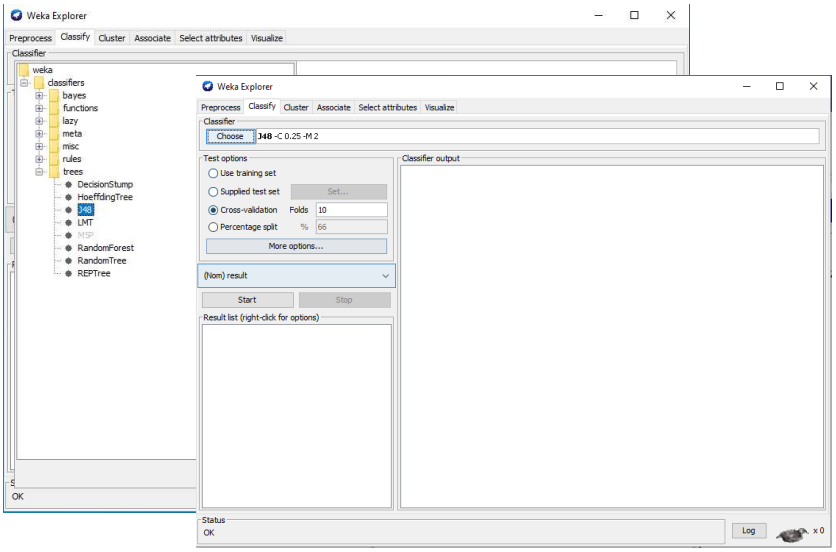


ภาพที่ **3.23** แสดงการนำข้อมูลเข้าในโปรแกรม Weka เวอร์ชัน 3.8.4

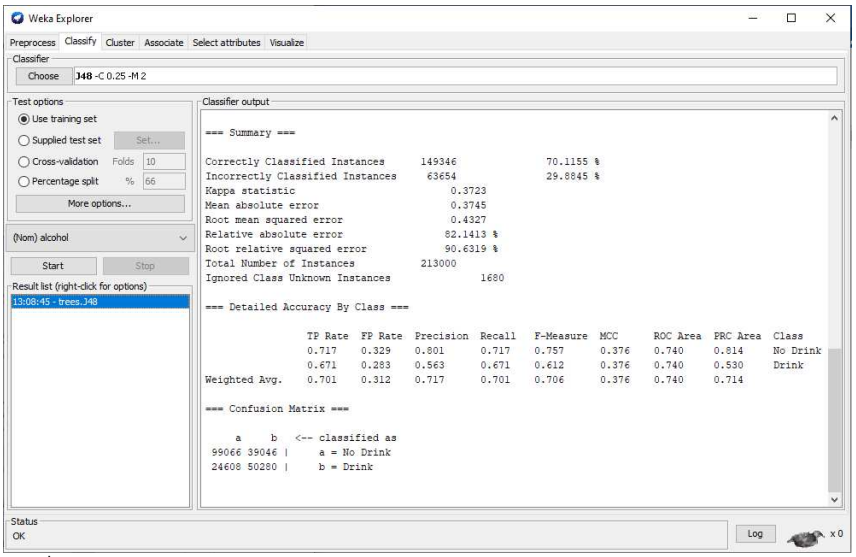


ภาพที่ **3.24** หน้าจอแสดงข้อมูลที่นำเข้าโปรแกรม Weka เวอร์ชัน 3.8.4

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



ภาพที่ **3.25** การเลือกเทคนิคที่ใช้ในการจำแนกข้อมูลแบบ Decision Tree: J48

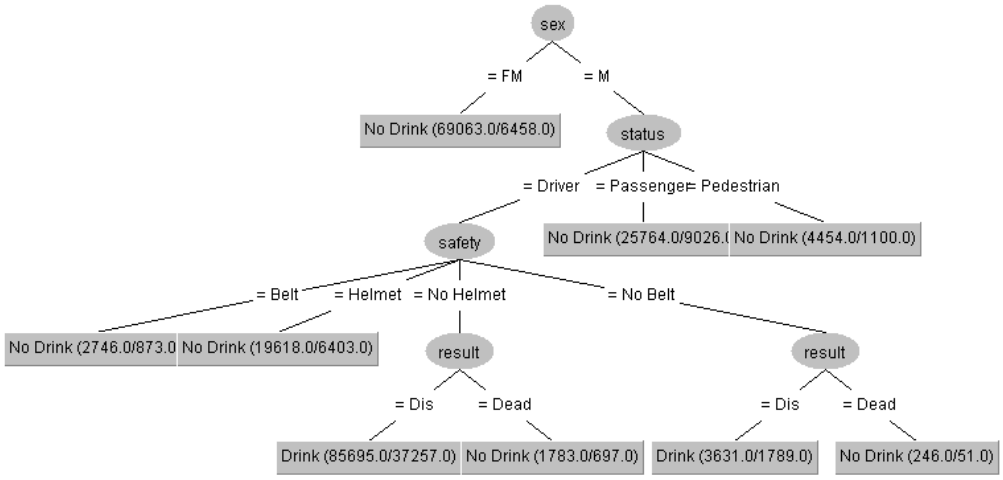


ภาพที่ **3.26** หน้าจอผลลัพธ์ของโมเดลการจำแนกข้อมูลแบบ Decision Tree: J48

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

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

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

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

Number of Leaves  :     9

Size of the tree :      14
```

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



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

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

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

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

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

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

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

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

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

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

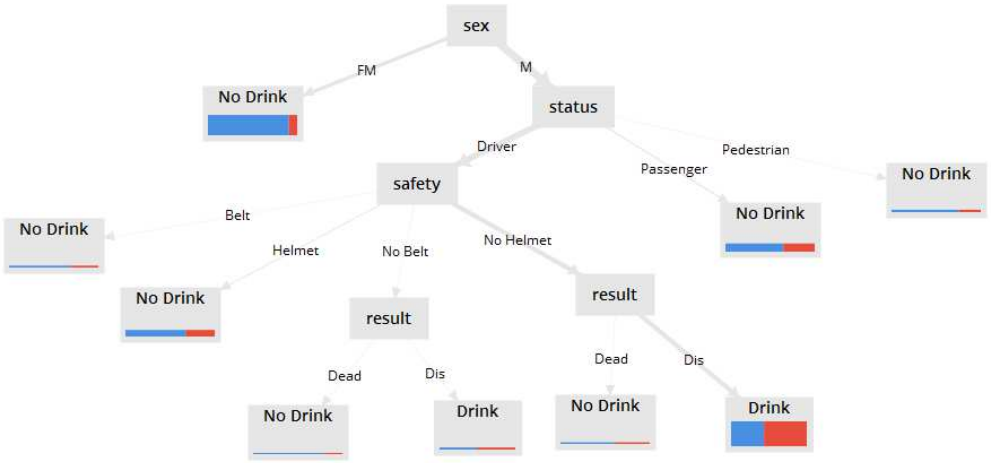
กฎข้อที่ 9 IF SEX=M And status= Driver And safety = No Helmet And result= Dead Then alcohol=No Drink หมายความว่า ถ้าเพศที่เป็นเพศชายเกี่ยวเนื่องกัน มีสถานะเป็นผู้ขับขี่ ไม่มีมาตรการป้องกันด้วยการใส่หมวกกันน็อค และมีผลการรักษาจากการเกิดอุบัติเหตุเป็นตาย ผลการพิจารณาพบว่า ประเภทของการดื่มสุรานั้นจะมีผลลัพธ์เป็น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

ทางคณะผู้จัดทำยังได้นำโมเดลของข้อมูลที่เลือกใช้ มาเปรียบเทียบกับโมเดลที่สร้างด้วยโปรแกรม RapidMiner Studio 9.5.1 ดังภาพที่ 3.29 และภาพที่ 3.30 ซึ่งคณะผู้จัดทำพบว่าได้ผลลัพธ์จากการจำแนกกลุ่ม และรูปแบบแผนภาพโมเดลที่ตรงกันสูง

Tree

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

ภาพที่ **3.29** ผลลัพธ์จากการจำแนกกลุ่มแบบ Decision Tree ของโปรแกรม RapidMiner



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

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

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

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

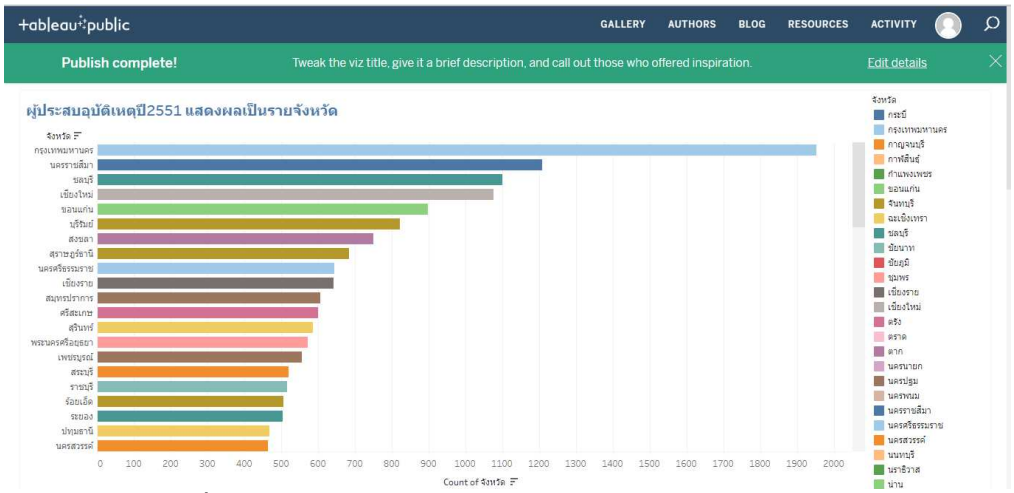
```
=== Summary ===

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

=== Detailed Accuracy By Class ===

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

=== Confusion Matrix ===

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

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

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

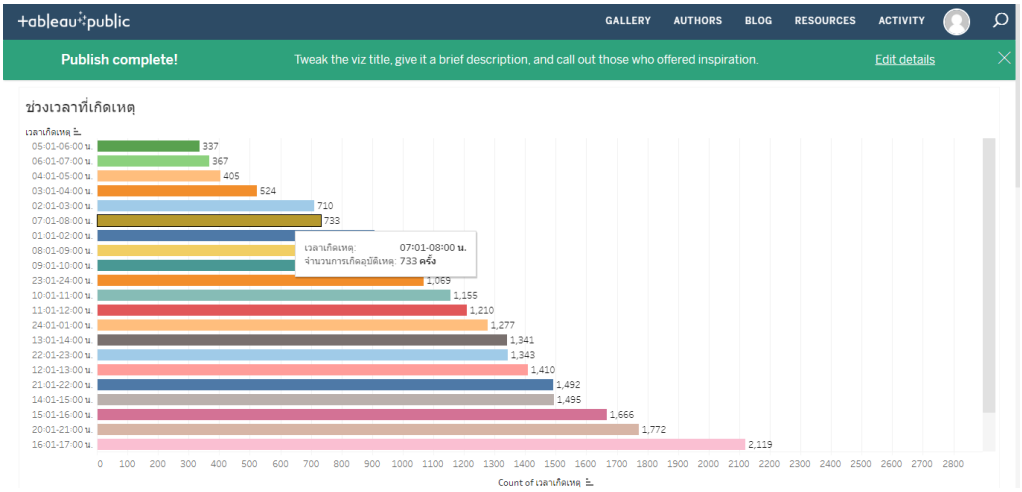
คณะผู้จัดทำนำผลข้อมูลที่ทำการวิเคราะห์แล้ว แสดงผลข้อมูลบน Web browser โดยใช้ชุดคำสั่ง HTML และ CSS3 ด้วยการนำเสนอข้อมูลแบบ visualization เป็นการแสดงผลข้อมูลในรูปแบบของภาพโดยใช้โปรแกรม Tableau Public ซึ่งทางคณะผู้จัดทำได้ยกตัวอย่างการจัดทำเป็นรูปแบบของรายงาน (Report) หรือแผนภาพ (Dashboard) กับข้อมูลผู้บาดเจ็บและเสียชีวิตในช่วงเทศกาลปีใหม่ประจำปี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

- กราฟแท่งแสดงข้อมูลการเกิดอุบัติเหตุแยกเป็นจังหวัด ผู้ใช้สามารถเลือกดูจำนวนการเกิดอุบัติเหตุได้ด้วยตัวกรองจังหวัดทางด้านขวา



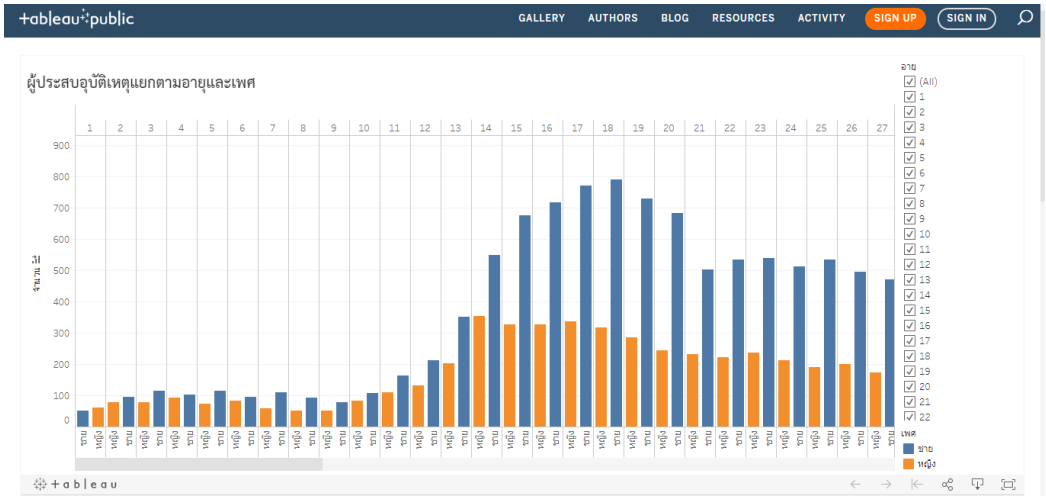
ภาพที่ 3.32 กราฟแท่งแสดงข้อมูลการเกิดอุบัติเหตุแยกเป็นจังหวัด

- กราฟแท่งแสดงช่วงเวลาในการเกิดอุบัติเหตุ ผู้ใช้สามารถเลือกดูช่วงเวลาต่าง ๆ ที่มีการเกิดอุบัติได้โดยการเลื่อนดูหรือดูช่วงเวลาการเกิดอุบัติเหตุแบบเรียงลำดับตามความต้องการได้



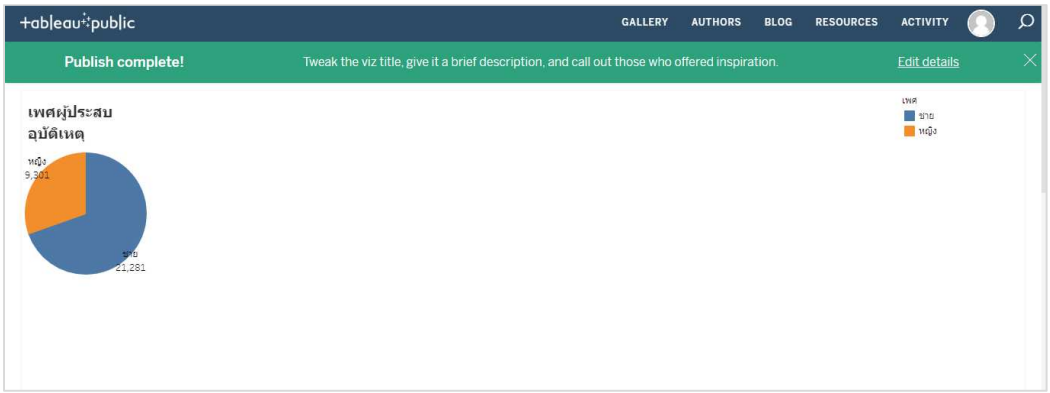
ภาพที่ 3.33 กราฟแท่งแสดงช่วงเวลาในการเกิดอุบัติเหตุ

- กราฟแท่งแสดงข้อมูลผู้ประสบอุบัติเหตุแยกเป็นช่วงอายุแบ่งเป็นเพศชายและเพศหญิง ผู้ใช้สามารถกรองดูข้อมูลได้จากตัวกรองอายุทางด้านขวา และสามารถเลือกดูความถี่การเกิดอุบัติเหตุโดยเรียงลำดับจากน้อยไปมากได้



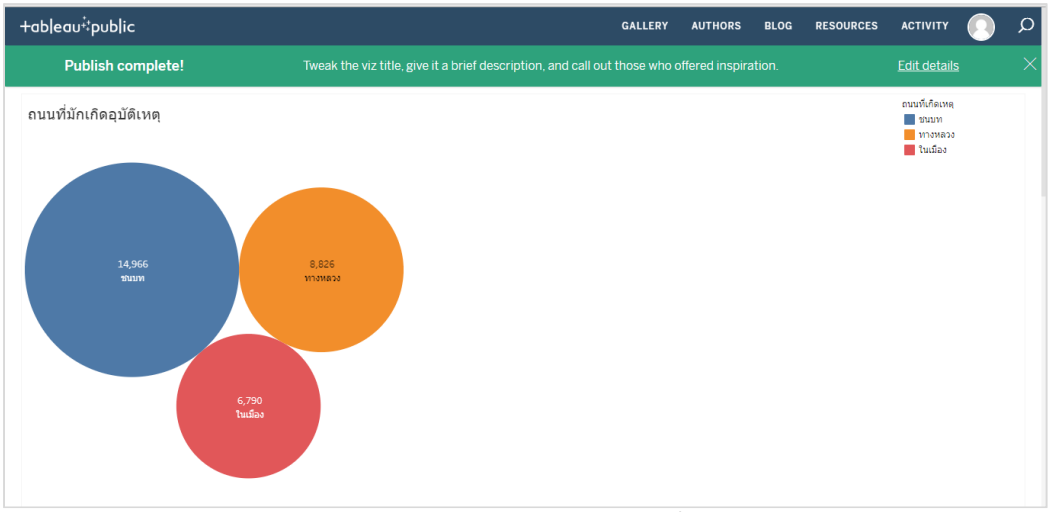
ภาพที่ 3.34 กราฟแท่งแสดงข้อมูลผู้ประสบอุบัติเหตุแยกเป็นช่วงอายุแบ่งเป็นเพศชายและเพศหญิง

- แผนภาพวงกลมแสดงผลจำนวนข้อมูลของผู้ประสบอุบัติโดยแบ่งเป็นผู้ประสบอุบัติเหตุประเภทเพศหญิงและเพศชาย



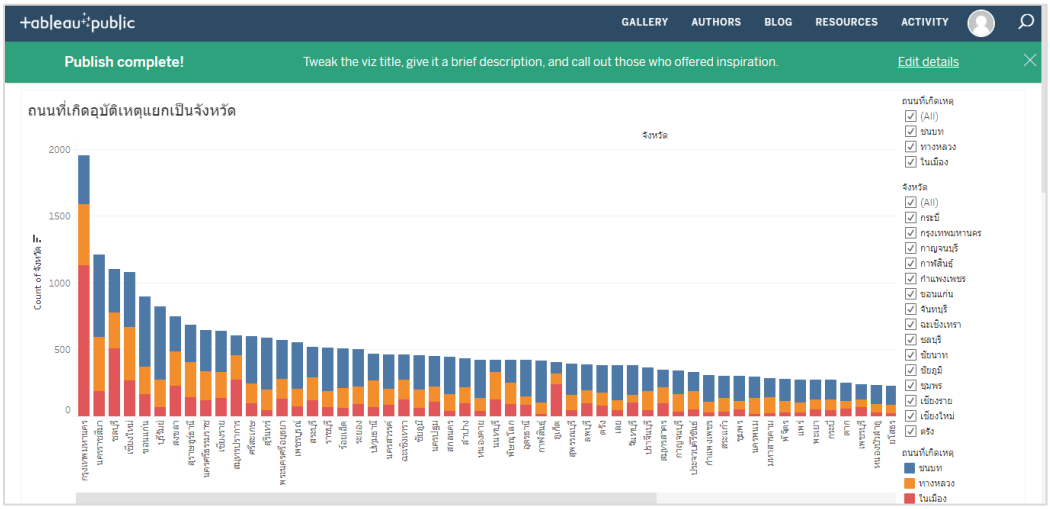
ภาพที่ 3.35 แผนภาพวงกลมแสดงผลข้อมูลผู้ประสบอุบัติ

- แผนภาพวงกลมแสดงข้อมูลถนนที่มีการเกิดอุบัติเหตุ โดยผู้ใช้สามารถกรองดูข้อมูลได้ตามความต้องการด้วยตัวกรองข้อมูลประเภทถนนทางด้านขวา



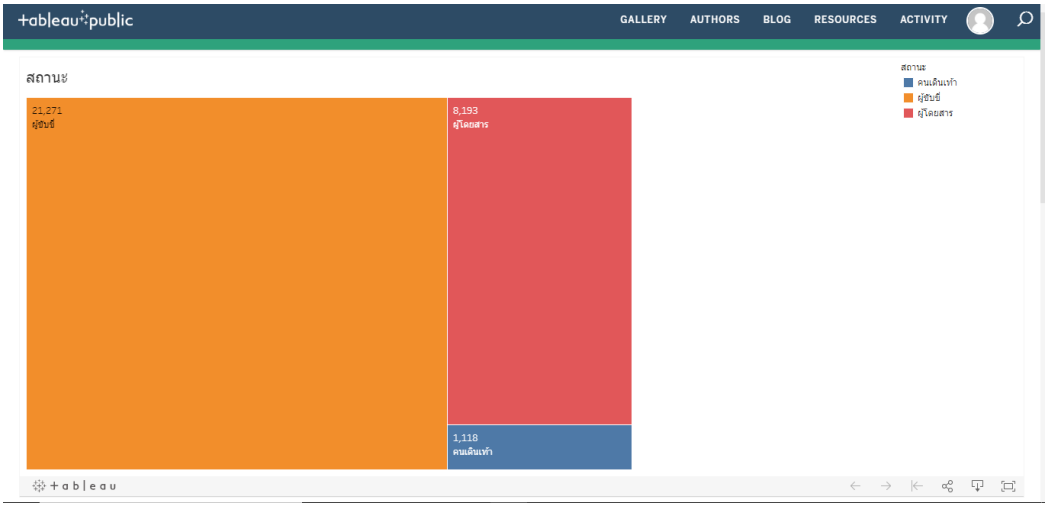
ภาพที่ 3.36 แผนภาพวงกลมแสดงข้อมูลถนนที่มีการเกิดอุบัติเหตุ

- กราฟแท่งแสดงการเกิดอุบัติเหตุบนท้องถนนประเภทต่าง ๆ โดยมีการแบ่งประเภทถนนและเรียงตามการเกิดอุบัติเหตุในแต่ละจังหวัด ผู้ใช้สามารถกรองดูข้อมูลได้จากตัวกรองด้านขวา



ภาพที่ 3.37 กราฟแท่งแสดงการเกิดอุบัติเหตุบนท้องถนนประเภทต่าง ๆ

- แผนภาพแสดงสถานะการประสบอุบัติเหตุของผู้ประสบอุบัติเหตุทั้งหมดโดยแบ่งเป็น 3 ประเภทคือ ผู้ขับขี่ ผู้โดยสารและคนเดินเท้า ผู้ใช้สามารถเลือกดูข้อมูลได้ด้วยตัวกรองสถานะทางด้านขวา

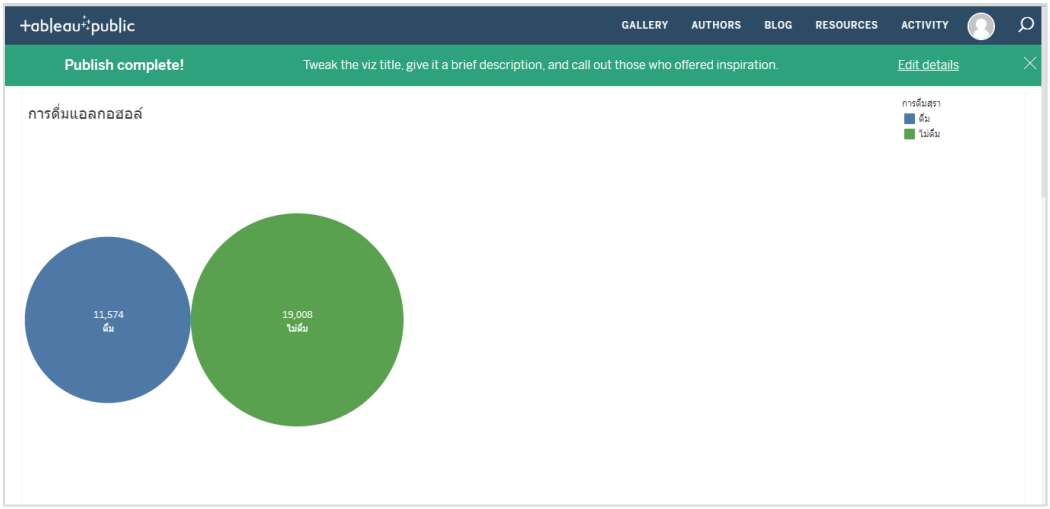


ภาพที่ 3.38 แผนภาพแสดงสถานการณ์ประสบอุบัติเหตุของผู้ประสบอุบัติเหตุทั้งหมด

- กราฟแท่งแสดงสถานะการประสบอุบัติเหตุของผู้ประสบอุบัติเหตุของเพศชายและเพศหญิงโดยแบ่งเป็น 3 ประเภทคือ ผู้ขับขี่ ผู้โดยสารและคนเดินเท้า ผู้ใช้สามารถเลือกดูข้อมูลได้ด้วยตัวกรองสถานะทางด้านขวา

ภาพที่ 3.39 กราฟแท่งแสดงสถานการณ์ประสบอุบัติเหตุของผู้ประสบอุบัติเหตุของเพศชายและเพศหญิง

- แผนภาพวงกลมแสดงผลการดื่มแอลกอฮอล์ของผู้ประสบอุบัติเหตุ โดยแบ่งข้อมูลเป็นดื่มแอลกอฮอล์และไม่ดื่มแอลกอฮอล์

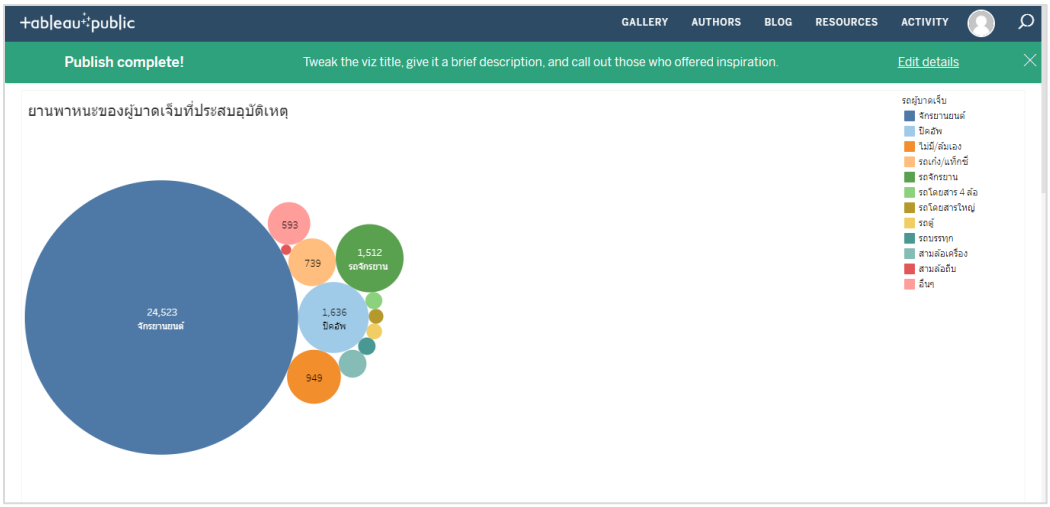


ภาพที่ 3.40 แผนภาพวงกลมแสดงผลการดื่มแอลกอฮอล์ของผู้ประสบอุบัติเหตุ

- หน้าการแสดงผลการดื่มแอลกอฮอล์ของผู้ประสบอุบัติเหตุในแต่ละจังหวัด โดยแบ่งข้อมูลเป็นดื่มแอลกอฮอล์และไม่ดื่มแอลกอฮอล์

ภาพที่ 3.41 หน้าการแสดงผลการดื่มแอลกอฮอล์ของผู้ประสบอุบัติเหตุในแต่ละจังหวัด

- แผนภาพแสดงผลยานพาหนะของผู้บาดเจ็บที่ประสบอุบัติเหตุ โดยผู้ใช้สามารถกรองดูข้อมูลประเภทยานพาหนะได้ด้วยตัวกรองประเพศยานพาหนะของผู้บาดเจ็บได้ทางด้านขวา

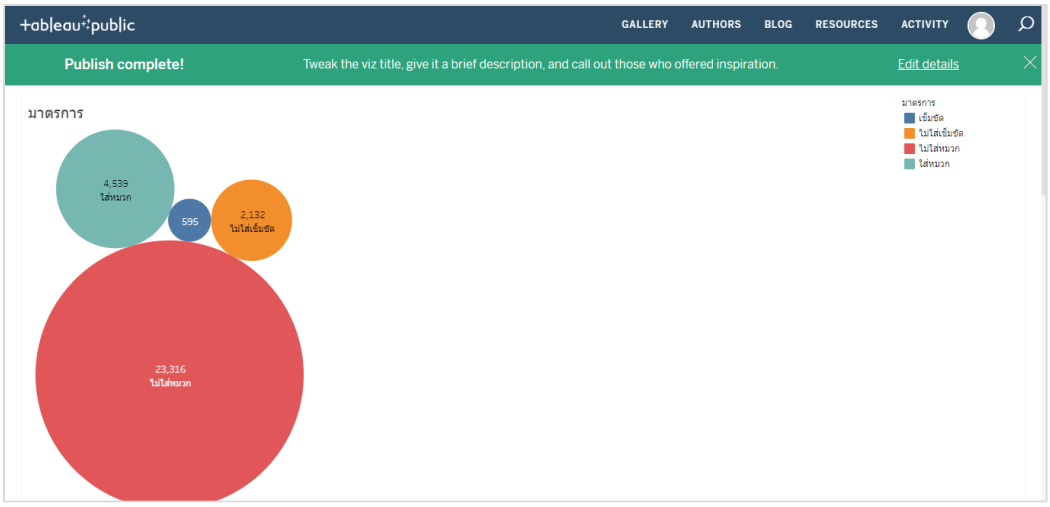


ภาพที่ 3.42 แผนภาพแสดงผลยานพาหนะของผู้บาดเจ็บที่ประสบอุบัติเหตุ

- แผนภาพแสดงผลยานพาหนะของคู่กรณี โดยผู้ใช้สามารถกรองดูข้อมูลประเภทยานพาหนะได้ด้วยตัวกรองประเพศยานพาหนะของคู่กรณีได้ทางด้านขวา

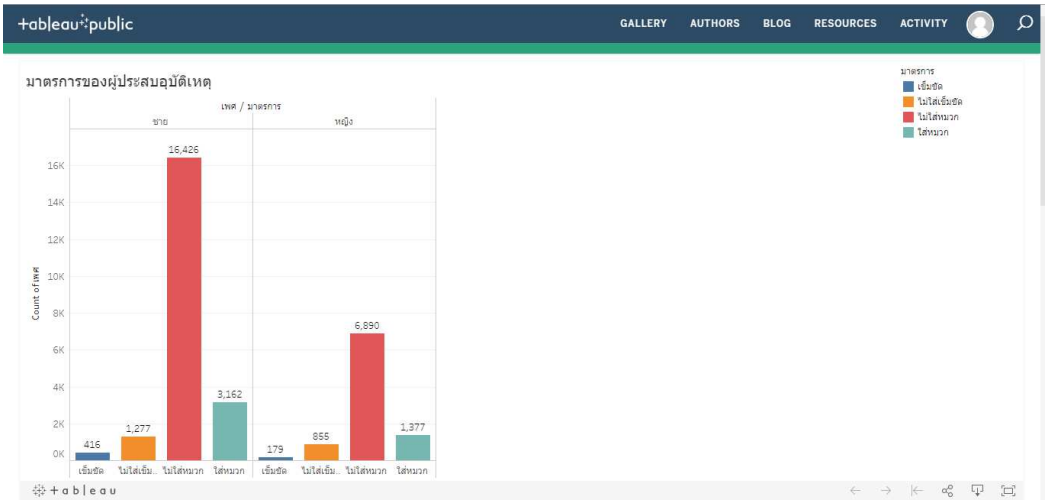
ภาพที่ 3.43 แผนภาพแสดงผลยานพาหนะของคู่กรณี

- แผนภาพวงกลมแสดงมาตรการในการป้องกันตัวเองระหว่างเดินทางของผู้ประสบอุบัติเหตุแบ่งเป็น สวมหมวกกันน็อค ไม่สวมหมวกกันน็อค คาดเข็มขัดนิรภัยและไม่คาดเข็มขัดนิรภัย



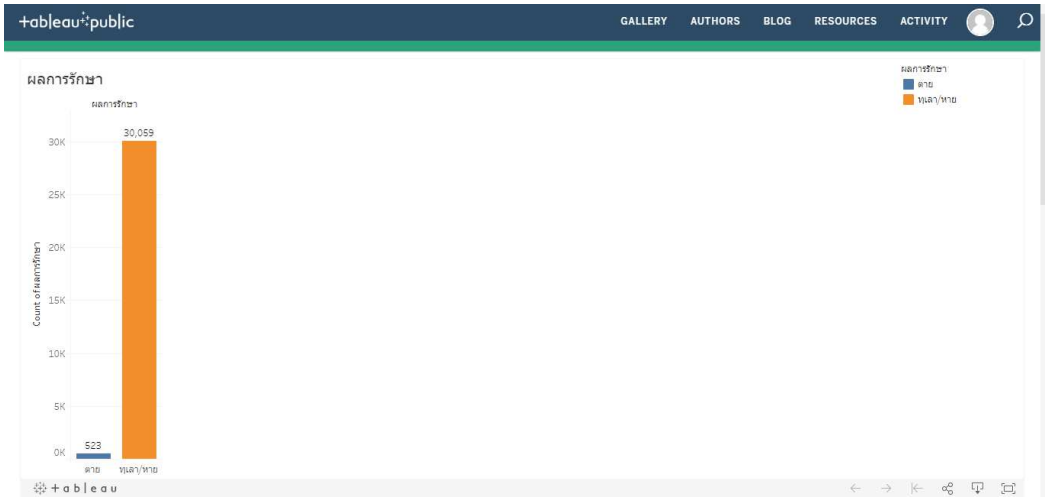
ภาพที่ 3.44 แผนภาพวงกลมแสดงมาตรการในการป้องกันตัวเองระหว่างเดินทางของผู้ประสบอุบัติเหตุ

- กราฟแท่งแสดงผลมาตรการในการป้องกันตัวเองระหว่างเดินทางของผู้ประสบอุบัติเหตุแบ่งเป็น สวมหมวกกันน็อค ไม่สวมหมวกกันน็อค คาดเข็มขัดนิรภัยและไม่คาดเข็มขัดนิรภัย โดยแยกตามเพศชายและเพศหญิง ผู้ใช้สามารถกรองดูข้อมูลมาตรการต่าง ๆ ได้



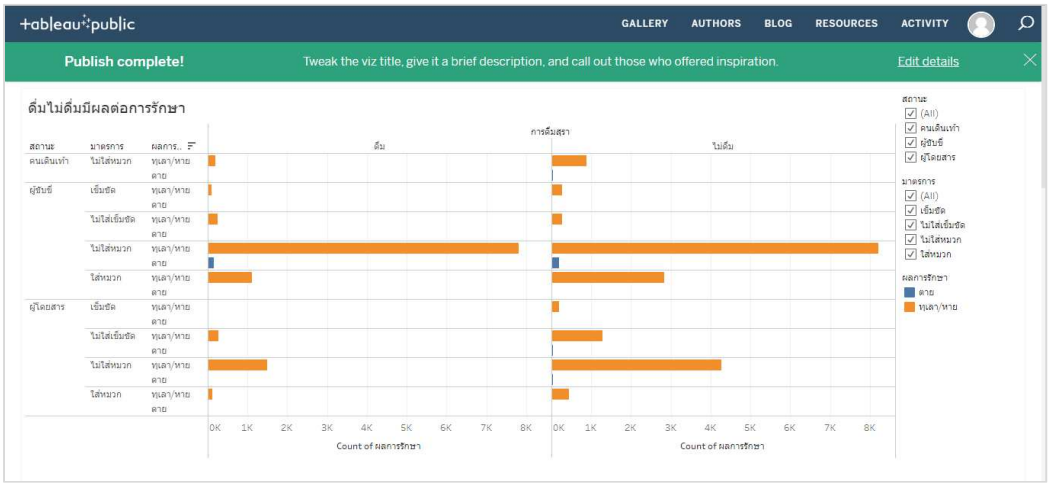
ภาพที่ 3.45 กราฟแท่งแสดงผลมาตรการในการป้องกันตัวเองระหว่างเดินทาง

- กราฟแสดงผลผลการรักษาของผู้ประสบอุบัติเหตุ แบ่งออกเป็นทุเลา/หาย และตาย โดยผู้ใช้สามารถกรองดูข้อมูลได้จากตัวกรองผลการรักษาทางด้านขวา



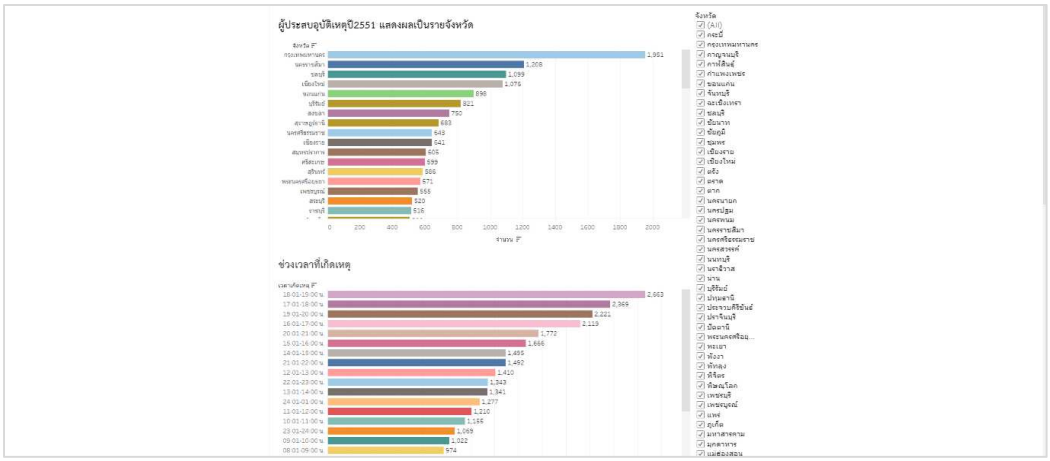
ภาพที่ 3.46 กราฟแสดงผลผลการรักษาของผู้ประสบอุบัติเหตุ

- แสดงผลสถานะของผู้ประสบอุบัติเหตุแยกเป็น ผู้ขับขี่ ผู้โดยสารและคนเดินเท้า มาตรการในการเดินทางของผู้ประสบอุบัติเหตุแยกเป็น สวมหมวกกันน็อค ไม่สวมหมวกกันน็อค คาดเข็มขัดนิรภัยและไม่คาดเข็มขัดนิรภัย การดื่มแอลกอฮอล์ และผลการรักษาของผู้ประสบอุบัติเหตุ โดยผู้ใช้สามารถกรองดูข้อมูลได้จากตัวกรองแต่ละประเภททางด้านขวา



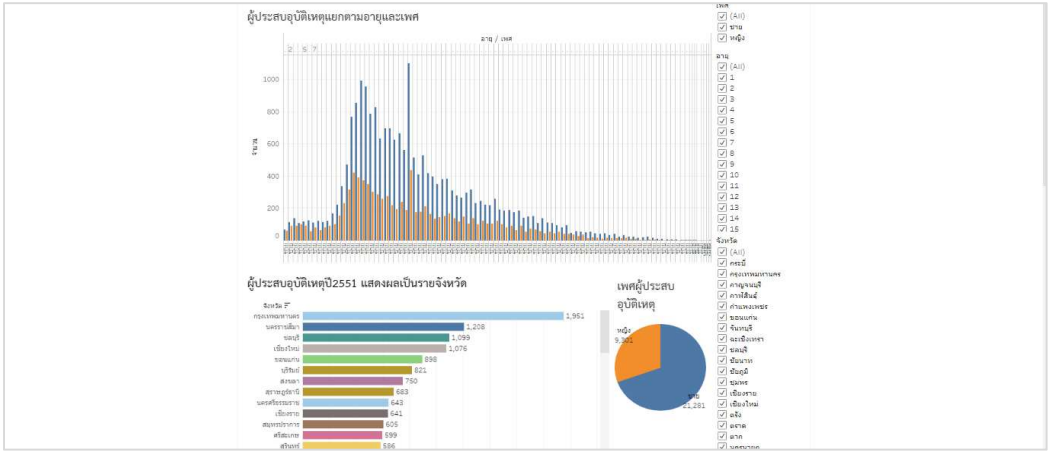
ภาพที่ 3.47 แสดงผลสถานะของผู้ประสบอุบัติเหตุ

- แดชบอร์ดอุบัติเหตุปี 2551 ในแต่ละจังหวัด แสดงผู้ประสบอุบัติเหตุทั้งหมดในปี 2551 แสดงผลโดยแยกเป็นจังหวัดและช่วงเวลาที่การเกิดอุบัติเหตุ ผู้ใช้สามารถกรองดูข้อมูลได้จากตัวกรอง



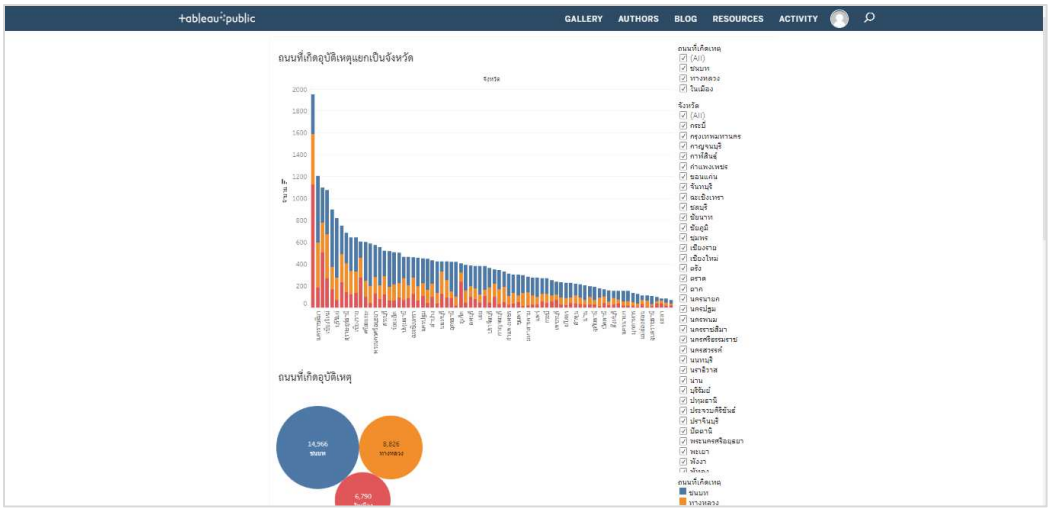
ภาพที่ 3.48 แดชบอร์ดอุบัติเหตุปี 2551 ในแต่ละจังหวัด

- แดชบอร์ดแสดงอุบัติเหตุปี 2551 ยอดผู้ประสบอุบัติเหตุ ภายในแดชบอร์ดแสดงช่วงอายุของผู้ประสบอุบัติเหตุแยกเป็นชายหญิง แสดงจำนวนผู้ประสบอุบัติเหตุทั้งหมดในปี 2551 แยกตามเพศ และแสดงการเกิดอุบัติเหตุในแต่ละจังหวัด ผู้ใช้สามารถกรองดูข้อมูลที่ต้องการดูได้จากตัวกรอง



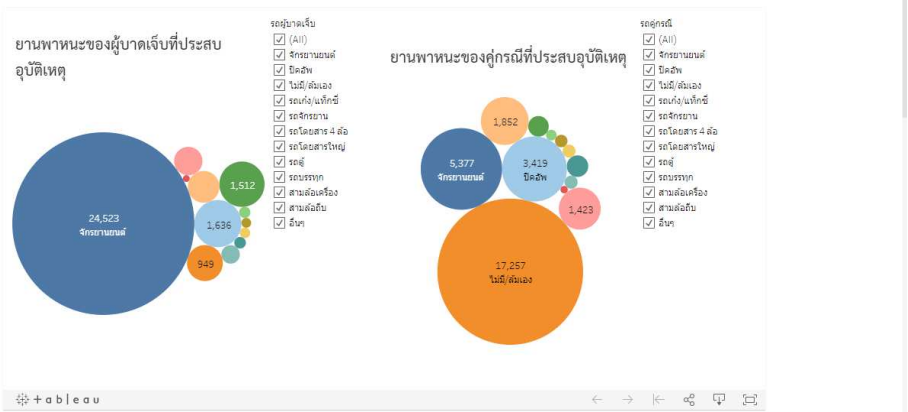
ภาพที่ 3.49 แดชบอร์ดแสดงอุบัติเหตุปี 2551 ยอดผู้ประสบอุบัติเหตุ

- แดชบอร์ดแสดงถนนที่เกิดอุบัติเหตุในปี 2551 ภายในประกอบไปด้วยกราฟแสดงถนนโดยแยกตามจังหวัด แผนภาพวงกลมแสดงจำนวนการเกิดอุบัติเหตุในถนนแต่ละเส้น ผู้ใช้สามารถกรองดูข้อมูลที่ต้องการดูได้จากตัวกรอง



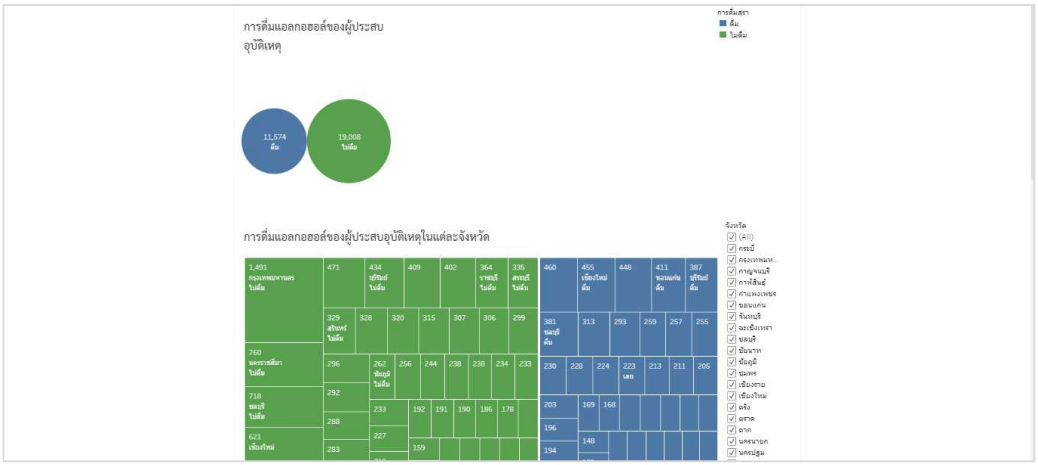
ภาพที่ 3.50 แดชบอร์ดแสดงถนนที่เกิดอุบัติเหตุในปี 2551

- แดชบอร์ดแสดงยานพาหนะที่ประสบอุบัติเหตุ ประกอบไปด้วยแผนภาพแสดงยานพาหนะของผู้บาดเจ็บที่ประสบอุบัติเหตุจะเห็นได้ว่ายานพาหนะที่ประสบอุบัติเหตุมากที่สุดคือรถจักรยานยนต์ และแผนภาพวงกลมแสดงยานพาหนะของคู่กรณีที่ประสบอุบัติเหตุ



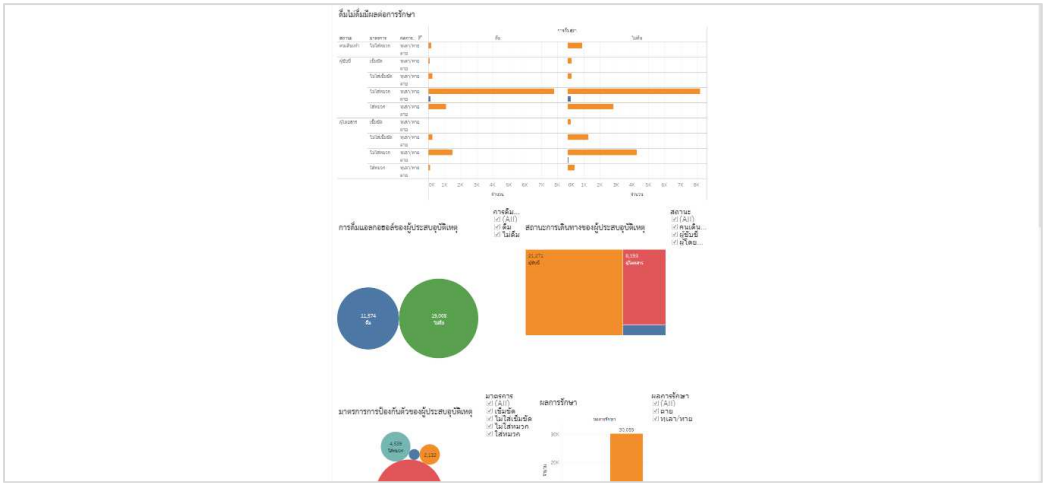
ภาพที่ 3.51 แดชบอร์ดแสดงยานพาหนะที่ประสบอุบัติเหตุ

- แดชบอร์ดแสดงผลการดื่มแอลกอฮอล์ ประกอบไปด้วย แผนภาพวงกลมแสดงจำนวนการดื่มและไม่ดื่มแอลกอฮอล์ แผนภาพแสดงจำนวนการดื่มและไม่ดื่มแอลกอฮอล์ของผู้ประสบอุบัติเหตุโดยแยกข้อมูลเป็นจังหวัด ผู้ใช้สามารถกรองดูข้อมูลได้จากตัวกรองข้อมูล



ภาพที่ 3.52 แดชบอร์ดแสดงผลการดื่มแอลกอฮอล์

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

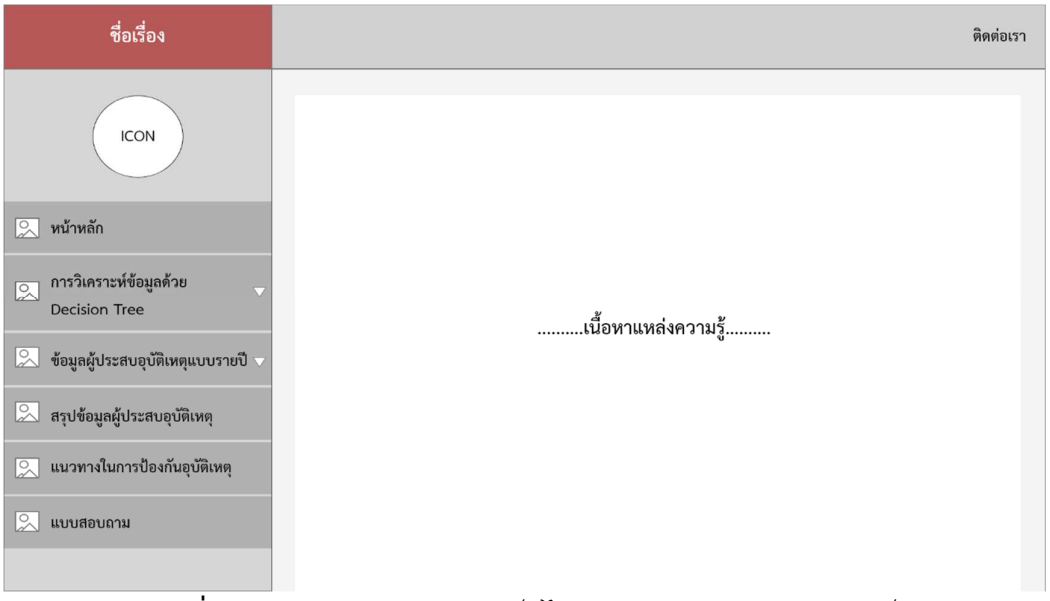


ภาพที่ 3.52 แดชบอร์ดแสดงข้อมูลเกี่ยวกับผู้ประสบอุบัติเหตุปี 2551

3.2 การออกแบบเว็บไซต์

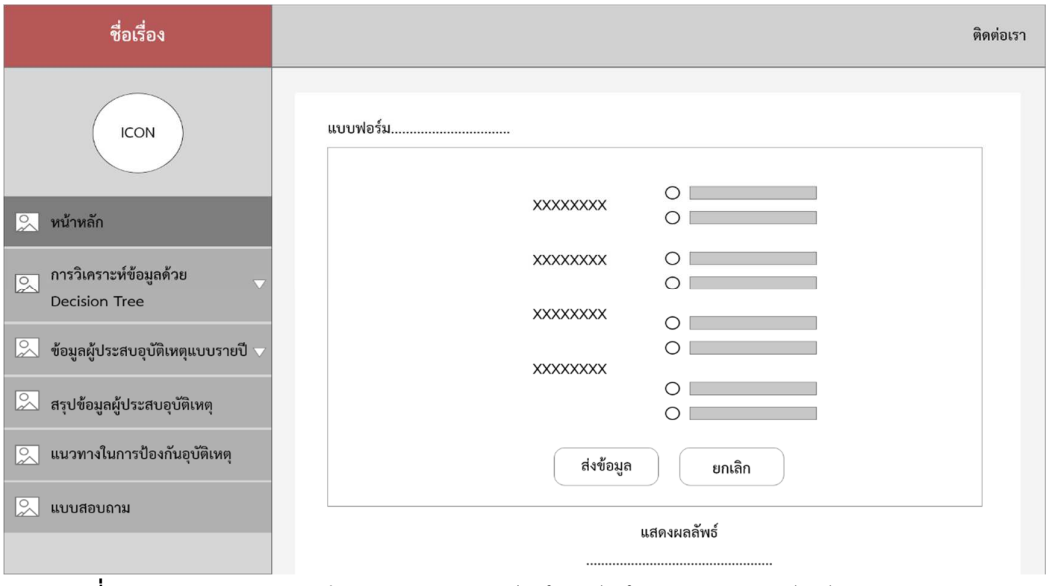
3.2.1 การออกแบบ Wireframe หน้าจอเว็บไซต์

1) หน้าของเว็บไซต์ แสดงเมนูต่าง ๆ ของหน้าเว็บ โดยจะแสดงความรู้ต่าง ๆ



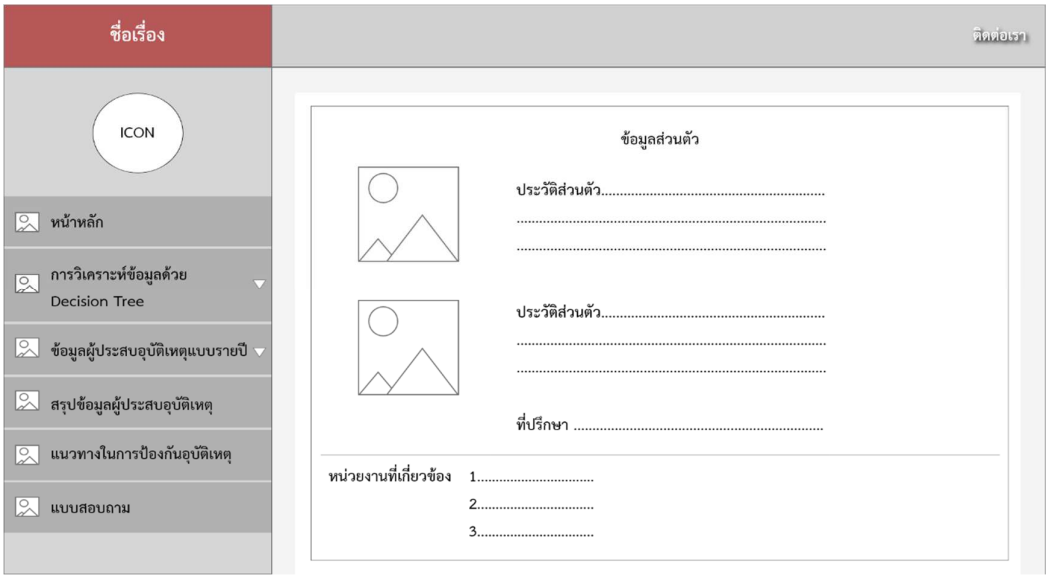
ภาพที่ 3.53 Wireframe หน้าของเว็บไซต์ แสดงเมนูต่าง ๆ ของหน้าเว็บ

2) หน้าแสดงแบบฟอร์มสำหรับจำลองการปฏิบัติตัวระหว่างเดินทางในช่วงเทศกาลปีใหม่ของผู้ใช้งานเว็บไซต์



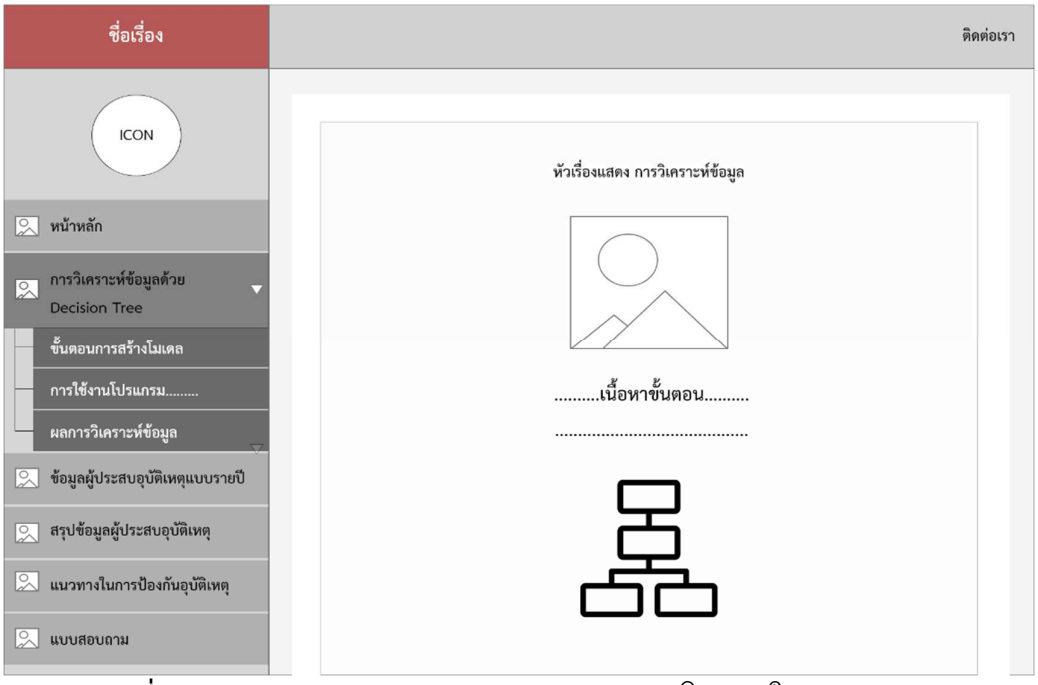
ภาพที่ 3.54 Wireframe หน้าแสดงแบบฟอร์มสำหรับจำลองการปฏิบัติตัวระหว่างเดินทาง

3) หน้า “ติดต่อเรา” จะแสดงข้อมูลประวัติผู้จัดทำ และหน่วยงานที่เกี่ยวข้องต่าง ๆ



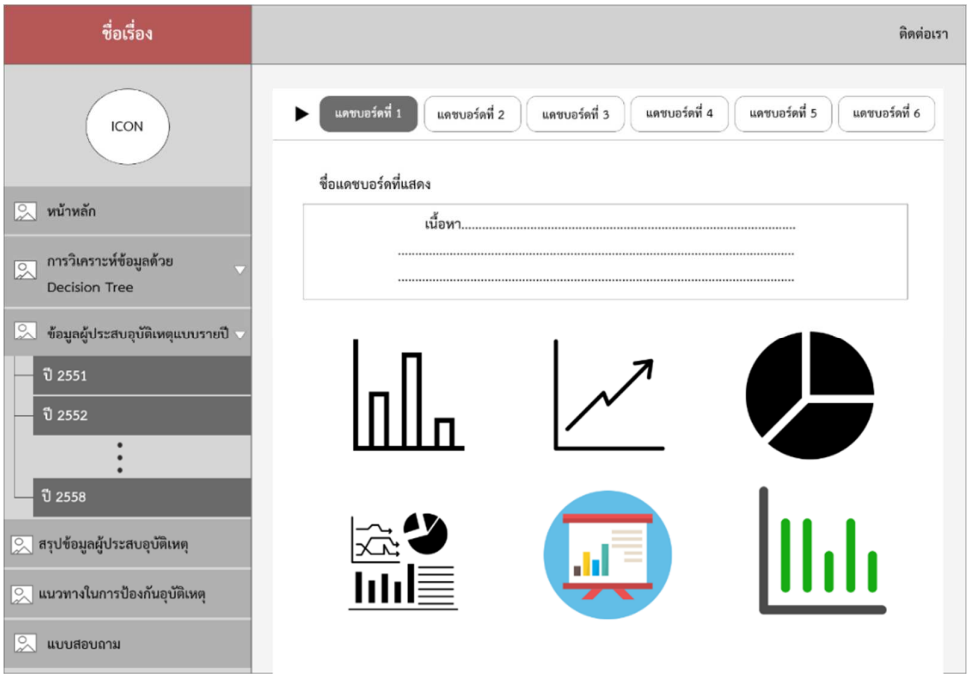
ภาพที่ 3.55 Wireframe หน้าติดต่อเรา

4) หน้าแสดงการวิเคราะห์ข้อมูลโดยการใช้ Decision Tree ที่จะมีการแสดงขั้นตอนการสร้างโมเดล การใช้งานโปรแกรม Weka 3.8.4 และผลของการวิเคราะห์ข้อมูล



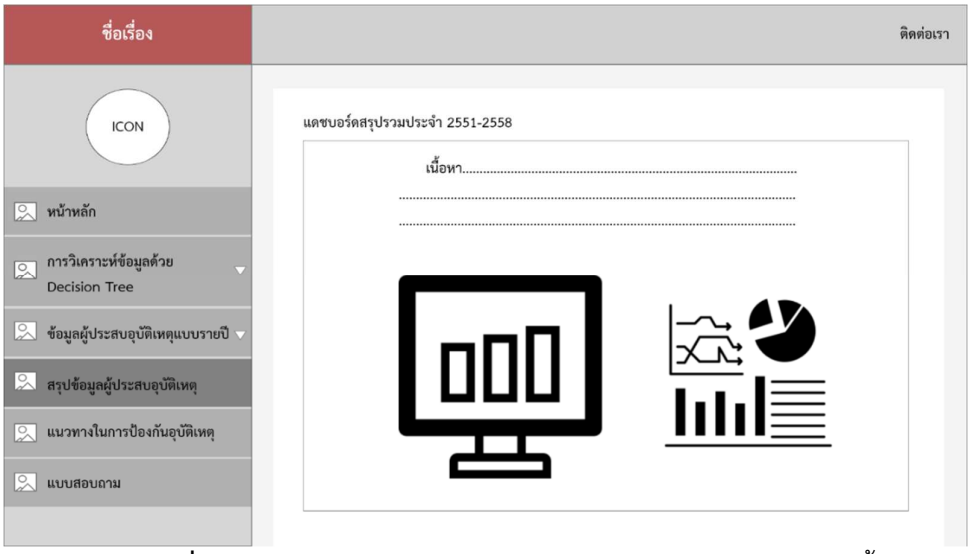
ภาพที่ 3.56 Wireframe หน้าแสดงการวิเคราะห์ข้อมูลโดยการใช้ Decision Tree

5) หน้าแสดงผลข้อมูลในรูปแบบตาราง และแดชบอร์ดผู้ประสบอุบัติเหตุรายปี ตั้งแต่ 2551–2558



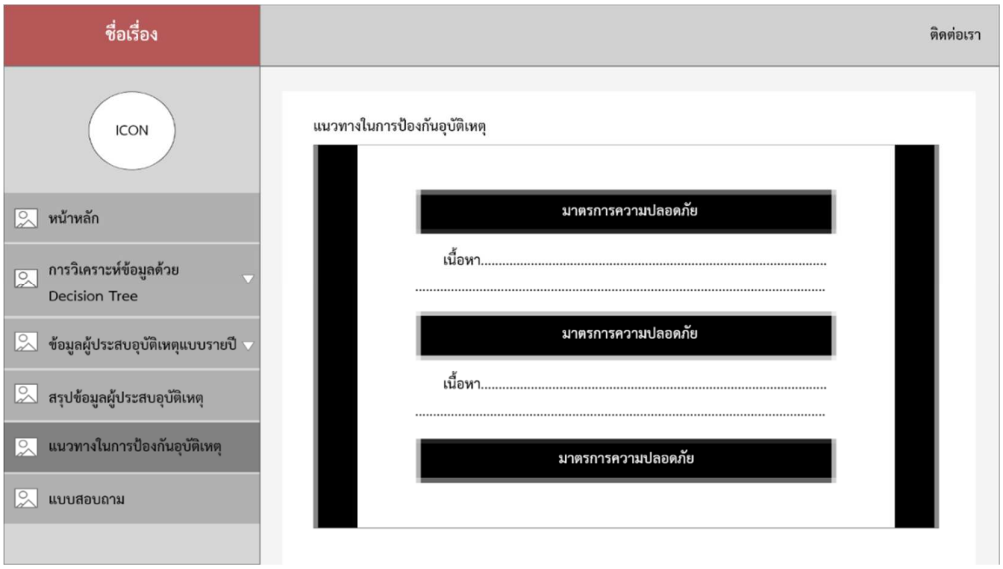
ภาพที่ 3.57 Wireframe หน้าแสดงผลข้อมูลแดชบอร์ดผู้ประสบอุบัติเหตุรายปี

6) หน้าแสดงผลสรุปข้อมูลผู้ประสบอุบัติเหตุทั้งหมด



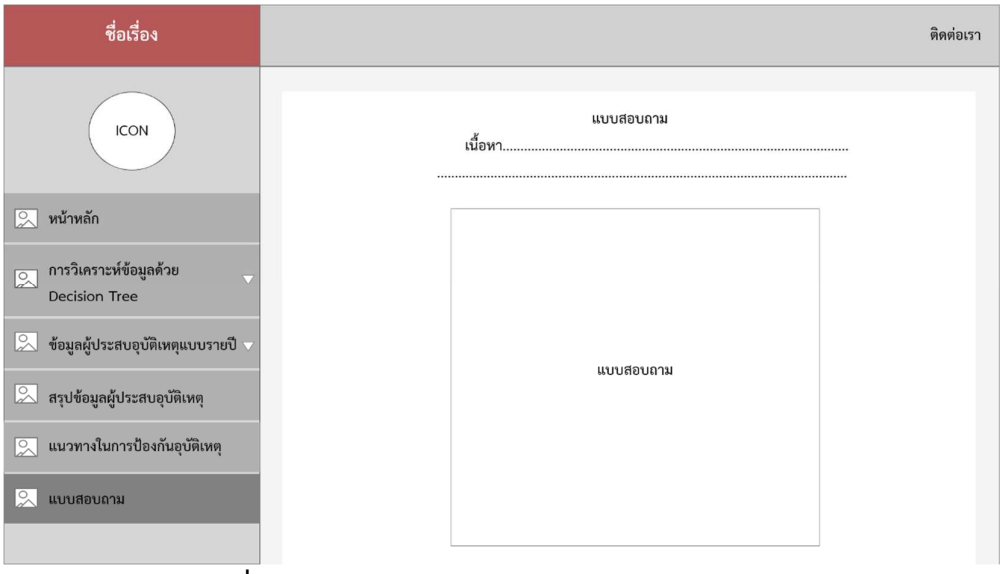
ภาพที่ 3.58 Wireframe หน้าแสดงผลสรุปข้อมูลผู้ประสบอุบัติเหตุทั้งหมด

7) หน้าแสดงผลข้อมูลแนวทางในการป้องกันอุบัติเหตุ ด้วยมาตรการต่าง ๆ



ภาพที่ 3.59 Wireframe หน้าแสดงผลข้อมูลแนวทางในการป้องกันอุบัติเหตุ

8) หน้าแสดงการกรอกแบบสอบถาม เพื่อเก็บรวบรวมข้อมูลจากผู้ใช้งานในเว็บไซต์ เพื่อการวิเคราะห์ครั้งต่อไป



ภาพที่ 3.60 Wireframe หน้าแสดงการกรอกแบบสอบถาม

3.3 บทสรุป

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

บทที่ 4
ผลการดำเนินงาน

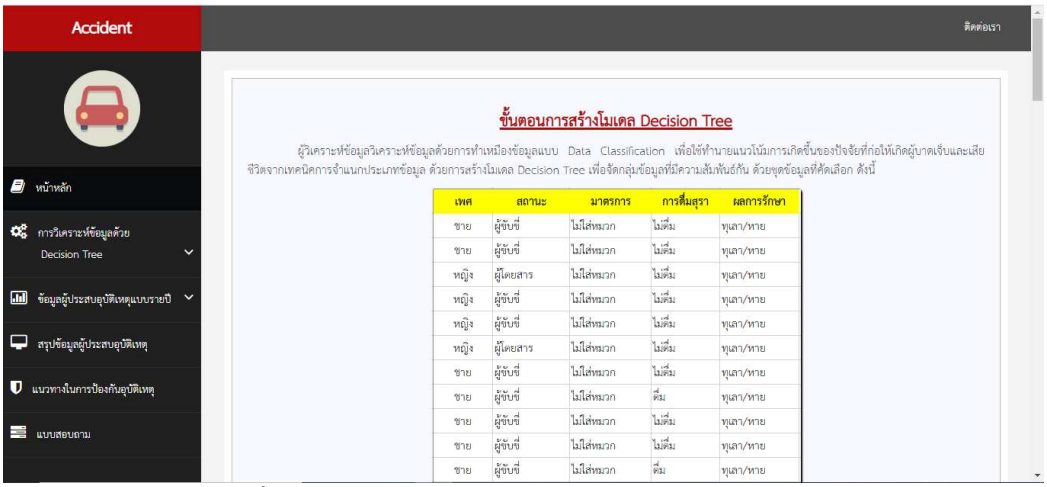
จากการวิเคราะห์ข้อมูลผู้บาดเจ็บและเสียชีวิตในช่วงเทศกาลปีใหม่ประจำปี 2551 – 2558 เพื่อใช้สำหรับเผยแพร่ข้อมูลบนเว็บไซต์ มีวัตถุประสงค์เพื่อวิเคราะห์ข้อมูลผู้บาดเจ็บและเสียชีวิตในช่วงเทศกาลปีใหม่ประจำปี 2551 – 2558 สำหรับเผยแพร่บนเว็บไซต์ เพื่อให้ได้เว็บไซต์ที่รวมแหล่งความรู้ทั่วไปเกี่ยวกับข้อมูล และแสดงกระบวนการวิเคราะห์ข้อมูล ที่สามารถนำไปใช้งานได้อย่างมีประสิทธิภาพและประสิทธิผล โดยผู้จัดทำโครงงานได้ทำการวิเคราะห์ข้อมูล และนำเทคโนโลยีสารสนเทศเข้าช่วยในการเสนอข้อมูล ทำให้มีความสะดวกเร็วและสามารถใช้งานได้จริงจนสำเร็จลุล่วงตามเป้าหมาย

4.1 ผลการดำเนินงาน

การจัดทำโครงงานเรื่อง การวิเคราะห์ข้อมูลผู้บาดเจ็บและเสียชีวิตในช่วงเทศกาลปีใหม่ประจำปี 2551 – 2558 เพื่อใช้สำหรับเผยแพร่ข้อมูลบนเว็บไซต์ ในครั้งนี้พบว่าการจำแนกข้อมูลออกเป็นประเภทต่าง ๆ ช่วยให้เกิดความเข้าใจต่อข้อมูล และสิ่งที่ต้องการทราบอย่างแท้จริง คณะผู้จัดทำได้ทำการวิเคราะห์ข้อมูลให้อยู่ในรูปแบบของ Data Mining โดยทำการจำแนกข้อมูลออกเป็นประเภทต่าง ๆ ด้วยเทคนิคการทำเหมืองข้อมูลแบบ Classification จากเทคนิคการจำแนกประเภทข้อมูลด้วยการวิเคราะห์ในรูปแบบของแผนภูมิต้นไม้ หรือที่เรียกว่า Decision tree และนำเสนอข้อมูลสารสนเทศแบบ visualization เผยแพร่ข้อมูลสารสนเทศนี้บน Web browser ให้กับผู้ใช้งาน ซึ่งทำให้เข้าใจได้อย่างสะดวกรวดเร็วยิ่งขึ้น รวมถึงมีความถูกต้อง ลดความซ้ำซ้อนจากข้อมูลที่มีจำนวนมากมหาศาล เพิ่มประสิทธิภาพให้กับการศึกษาค้นคว้าและทันต่อเวลา คณะผู้จัดทำจึงได้ทำการวิเคราะห์ข้อมูล และจัดทำเว็บไซต์ตามวัตถุประสงค์ที่ได้กำหนดไว้ และมีผลการดำเนินงาน ดังนี้

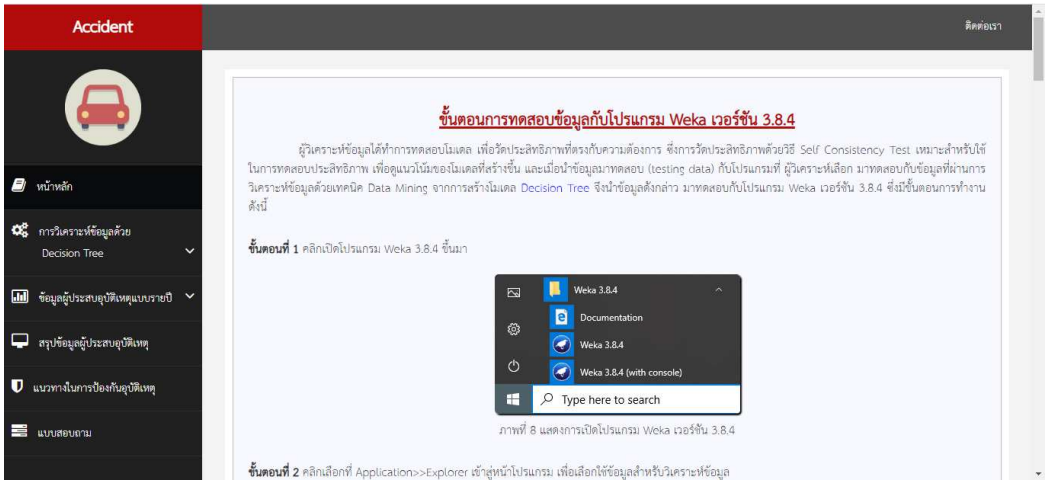
ภาพที่ 4.1 หน้าแรกของเว็บไซต์ เมื่อเข้าสู่เว็บไซต์การวิเคราะห์ข้อมูลอุบัติเหตุ

คำอธิบายการใช้งาน : แถบเมนูสามารถใช้งานได้ตามความต้องการของการใช้งาน ซึ่งเป็นหน้าสำหรับผู้ใช้งานทั่วไป สามารถเลือกดู และศึกษาค้นคว้าเข้าไปใช้งานในส่วนของเว็บไซต์ได้



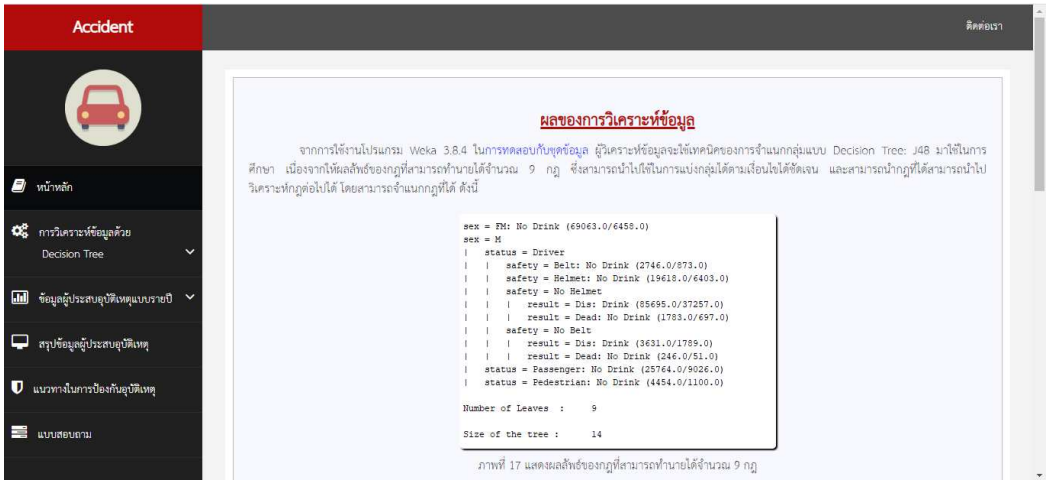
ภาพที่ 4.2 เมนูแสดงการวิเคราะห์ข้อมูลด้วย Decision Tree

คำอธิบายการใช้งาน : เป็นหน้าแสดงการอธิบายขั้นตอนในการสร้างโมเดล Decision Tree ซึ่งจะประกอบไปด้วยการคำนวณต่าง ๆ และยังแถบเมนูแสดงการใช้งานโปรแกรม Weka 3.8.4 และแถบเมนูของผลการวิเคราะห์ข้อมูล



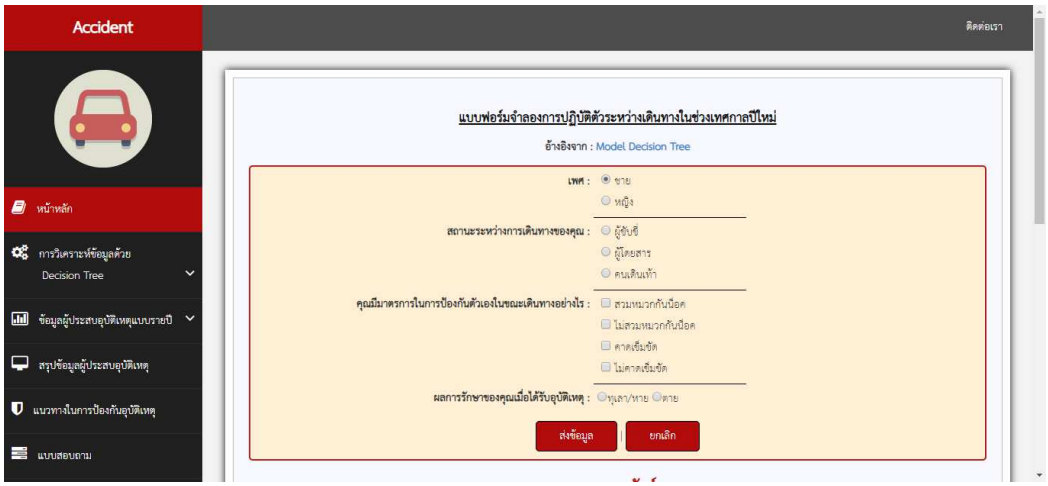
ภาพที่ 4.3 เมนูแสดงการใช้งานโปรแกรม Weka 3.8.4

คำอธิบายการใช้งาน : เป็นหน้าแสดงการอธิบายขั้นตอนการทดสอบกับข้อมูลที่คัดเลือก โดยใช้โปรแกรม Weka 3.8.4 เพื่อดูผลลัพธ์ของ Decision tree



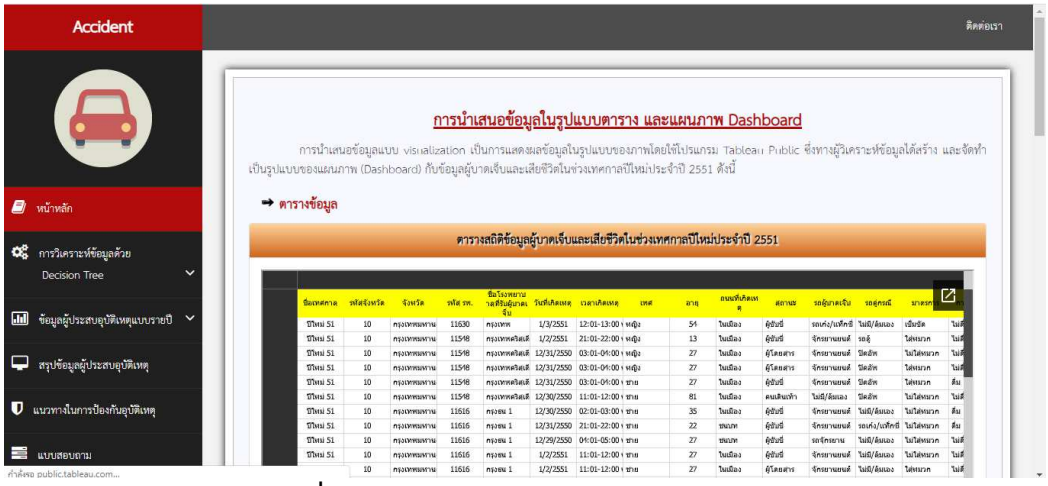
ภาพที่ 4.4 เมนูแสดงการใช้งานโปรแกรม Weka 3.8.4

คำอธิบายการใช้งาน : เป็นหน้าแสดงผลลัพธ์จากการวิเคราะห์ข้อมูล ให้อยู่ในรูปแบบของ Decision Tree ซึ่งทำนายได้จำนวณ 9 กฎ



ภาพที่ 4.5 เมนูแสดงผลการวิเคราะห์ข้อมูล

คำอธิบายการใช้งาน : เป็นหน้าจอการทำแบบฟอร์มจำลองการปฏิบัติตัวระหว่างเดินทางในช่วงเทศกาลปีใหม่ โดยจะแสดงผลลัพธ์ตามการวิเคราะห์ข้อมูล Decision Tree และให้คะแนะนำ และแนวทางการป้องกัน ซึ่งอยู่ในรูปของกฎต่าง ๆ จากข้อมูลที่เราคัดเลือก

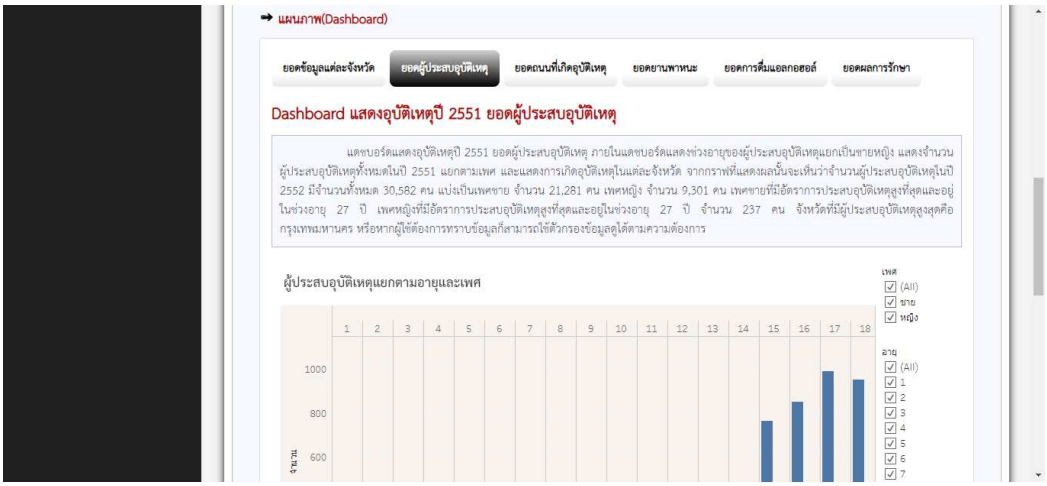


ภาพที่ 4.6 แสดงตารางสถิติข้อมูลประจำปี 2551

คำอธิบายการใช้งาน : แสดงตารางสถิติข้อมูลผู้บาดเจ็บและเสียชีวิตในช่วงเทศกาลปีใหม่ประจำปี 2551 โดยจะมีแถบเมนูการแสดงแดชบอร์ดรายปี ผู้ใช้งานทั่วไปสามารถคลิกเลือกดูข้อมูลจากตัวกรองของแดชบอร์ดได้

ภาพที่ 4.7 แสดงผลยอดข้อมูลแต่ละจังหวัดของผู้ประสบอุบัติเหตุปี 2551

คำอธิบายการใช้งาน : แสดงผลการสรุปข้อมูลประจำปี 2551 โดยหน้านี้แสดง Dashboard ยอดอุบัติเหตุในแต่ละจังหวัด ผู้ใช้งานทั่วไปสามารถคลิกเลือกดูข้อมูลจากตัวกรองของแดชบอร์ดได้



ภาพที่ 4.8 แสดงผลยอดข้อมูลของผู้ประสบอุบัติเหตุปี 2551

คำอธิบายการใช้งาน : แสดงผลการสรุปข้อมูลประจำปี 2551 โดยหน้านี้แสดง Dashboard ยอดผู้ประสบอุบัติเหตุ ผู้ใช้งานทั่วไปสามารถคลิกเลือกดูข้อมูลจากตัวกรองของแดชบอร์ดได้

ภาพที่ 4.9 แสดงผลยอดข้อมูลถนนที่เกิดอุบัติเหตุปี 2551

คำอธิบายการใช้งาน : แสดงผลการสรุปข้อมูลประจำปี 2551 โดยหน้านี้แสดง Dashboard ยอดถนนที่เกิดอุบัติเหตุสูงสุด ผู้ใช้งานทั่วไปสามารถคลิกเลือกดูข้อมูลจากตัวกรองของแดชบอร์ดได้

ภาพที่ 4.10 แสดงผลยอดข้อมูลยอดยานพาหนะที่ประสบอุบัติเหตุปี 2551

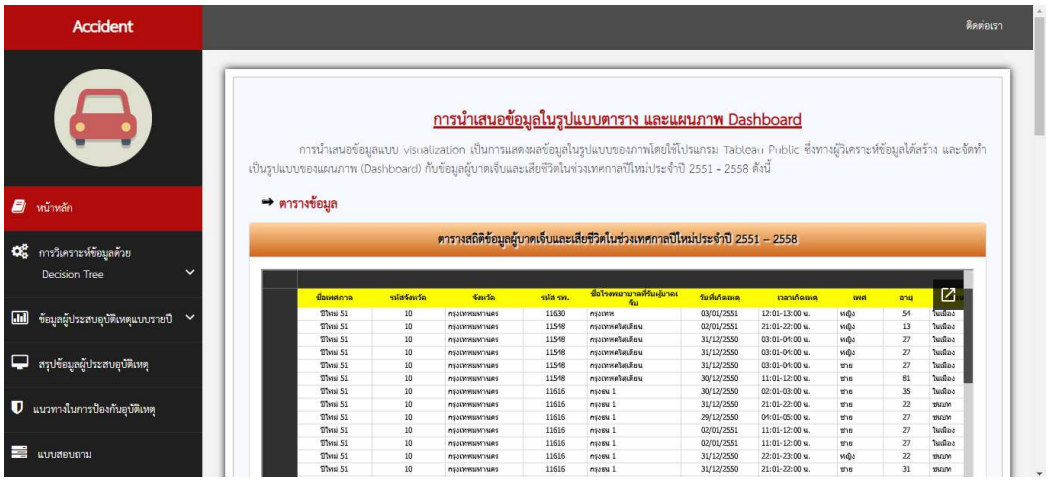
คำอธิบายการใช้งาน : แสดงผลการสรุปข้อมูลประจำปี 2551 โดยหน้านี้แสดง Dashboard ยอดยานพาหนะที่ประสบอุบัติเหตุ ผู้ใช้งานทั่วไปสามารถคลิกเลือกดูข้อมูลจากตัวกรองของแดชบอร์ดได้

ภาพที่ 4.11 แสดงผลยอดข้อมูลการดื่มแอลกอฮอล์ของผู้ประสบอุบัติเหตุปี 2551

คำอธิบายการใช้งาน : แสดงผลการสรุปข้อมูลประจำปี 2551 โดยหน้านี้แสดง Dashboard ยอดการดื่มแอลกอฮอล์ของผู้ประสบอุบัติเหตุ ผู้ใช้งานทั่วไปสามารถคลิกเลือกดูข้อมูลจากตัวกรองของแดชบอร์ดได้

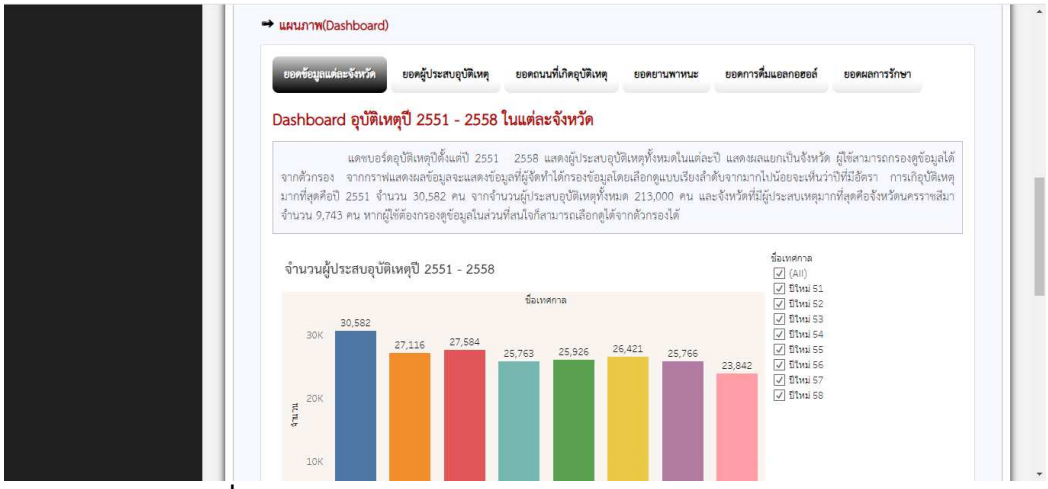
ภาพที่ 4.12 แสดงผลยอดข้อมูลผลการรักษาของผู้ประสบอุบัติเหตุปี 2551

คำอธิบายการใช้งาน : แสดงผลการสรุปข้อมูลประจำปี 2551 โดยหน้านี้แสดง Dashboard ยอดการรักษาของผู้ประสบอุบัติเหตุ ผู้ใช้งานทั่วไปสามารถคลิกเลือกดูข้อมูลจากตัวกรองของแดชบอร์ดได้



ภาพที่ 4.13 แสดงตารางสถิติข้อมูลประจำปี 2551–2558

คำอธิบายการใช้งาน : สถิติข้อมูลผู้บาดเจ็บและเสียชีวิตในช่วงเทศกาลปีใหม่ประจำปี 2551 – 2558 โดยจะมีแถบเมนูในรูปแบบเดียวกันกับการแสดงแดชบอร์ดรายปี ผู้ใช้งานทั่วไปสามารถคลิกเลือกดูข้อมูลจากตัวกรองของแดชบอร์ดได้

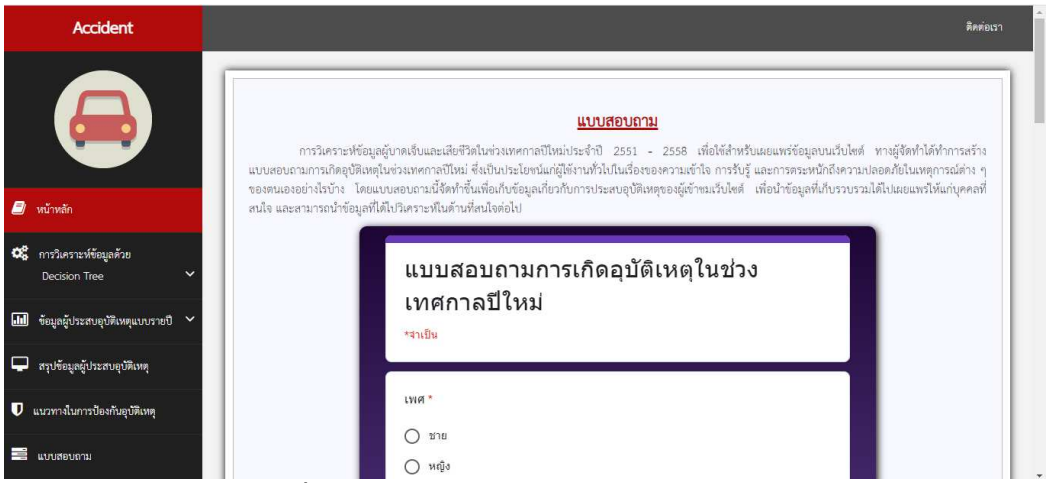


ภาพที่ 4.14 แสดงแดชบอร์ดของผู้ประสบอุบัติเหตุปี 2551–2558

คำอธิบายการใช้งาน : แสดงผลของการสรุปข้อมูลโดยรวมจากข้อมูลจำนวน 8 ปี ประจำปี 2551 – 2558 โดยจะมีแถบเมนูในรูปแบบเดียวกันกับการแสดงแดชบอร์ดรายปี ผู้ใช้งานทั่วไปสามารถคลิกเลือกดูข้อมูลจากตัวกรองของแดชบอร์ดได้

ภาพที่ 4.15 แสดงแนวทางป้องกันการเกิดอุบัติเหตุ

คำอธิบายการใช้งาน : เป็นหน้าแสดงแหล่งความรู้ของแนวทางป้องกันการเกิดอุบัติเหตุจากการใช้รถใช้ถนนในช่วงเทศกาลปีใหม่ที่เผยแพร่ให้กับผู้ใช้งานทั่วไป



ภาพที่ 4.16 แสดงส่วนของการรกรอกแบบสอบถาม

คำอธิบายการใช้งาน : เป็นหน้าจอของการกรอกแบบสอบถาม สำหรับผู้ใช้งานทั่วไปที่เข้าเว็บไซต์ เพื่อสอบถามเกี่ยวกับการปฏิบัติตัวเองในช่วงเทศกาลปีใหม่ ซึ่งคณะผู้จัดทำสามารถนำข้อมูลที่ได้มาใช้ให้เกิดประโยชน์ต่อไปได้

ภาพที่ 4.17 แสดงข้อมูลส่วนตัวของผู้จัดทำ

คำอธิบายการใช้งาน : เป็นหน้าจอแสดงประวัติผู้จัดทำเว็บไซต์ และข้อมูลที่ต้องการติดต่อ รวมถึงแสดงหน่วยงานที่เกี่ยวข้องกับงานชิ้นนี้

4.2 การอภิปรายผล

จากวัตถุประสงค์เพื่อวิเคราะห์ข้อมูลผู้บาดเจ็บและเสียชีวิตในช่วงเทศกาลปีใหม่ประจำปี 2551 – 2558 สำหรับเผยแพร่บนเว็บไซต์ คณะผู้จัดทำได้ศึกษาปัญหา และเก็บรวบรวมข้อมูล ซึ่งคณะผู้จัดทำได้ทำการวิเคราะห์ข้อมูลให้อยู่ในรูปแบบของ Data Mining โดยใช้กระบวนการวิเคราะห์ข้อมูล (Cross Industry Standard Process for Data Mining หรือ CRISP-DM) เพื่อใช้ในการวิเคราะห์ตามขอบเขตของโครงงาน ดังนี้

4.2.1 ขอบเขตคณะผู้จัดทำ

1) รู้จักและเข้าใจในธุรกิจ (Business Understanding) ผลการศึกษาพบว่าปัญหาของข้อมูล คือ ข้อมูลผู้บาดเจ็บและเสียชีวิตในช่วงเทศกาลปีใหม่ประจำปี 2551 – 2558 ซึ่งมีจำนวนข้อมูลมหาศาล จำนวนรายการทั้งหมด 214,950 รายการ ทำให้ไม่สามารถทำความเข้าใจกับข้อมูลที่มีความซับซ้อนได้อย่างรวดเร็ว และการเสนอหรือเผยแพร่ข้อมูลผู้บาดเจ็บและเสียชีวิตในช่วงเทศกาลปีใหม่ประจำปี 2551 – 2558 ให้กับบุคคลภายนอกหรือกลุ่มผู้ใช้ข้อมูลได้รับรู้ เพื่อตระหนักถึงอุบัติเหตุในช่วงเทศกาลที่ยังไม่มีประสิทธิภาพที่ดีนัก

2) จัดเก็บและรวบรวมข้อมูลให้ครบ (Data Understanding) คณะผู้จัดทำทำการรวบรวมข้อมูล เพื่อตรวจสอบรายละเอียด ปริมาณ และความน่าเชื่อถือของข้อมูลผู้บาดเจ็บและเสียชีวิตในช่วงเทศกาลปีใหม่ประจำปี 2551 – 2558 ที่ได้จากเว็บไซต์ data.go.th ซึ่งเป็นระบบศูนย์กลางข้อมูลภาครัฐ ที่เผยแพร่สู่สาธารณะอย่างเป็นรูปธรรม

3) เตรียมข้อมูลให้พร้อมใช้งาน (Data Preparation) คณะผู้จัดทำทำการคัดเลือกข้อมูล และทำการ Data Cleaning ข้อมูลเพื่อให้ได้ข้อมูลที่สมบูรณ์

4) สร้างแบบจำลอง (Modeling) คณะผู้จัดทำวิเคราะห์ข้อมูลด้วยการทำเหมืองข้อมูลแบบ Data Classification ที่จะสร้างกฎเพื่อช่วยในการตัดสินใจจากข้อมูลที่มีอยู่ เพื่อใช้ทำนายแนวโน้มการเกิดขึ้นของปัจจัยที่ก่อให้เกิดผู้บาดเจ็บและเสียชีวิตจากเทคนิคการจำแนกประเภทข้อมูล ด้วยการสร้างโมเดล Decision Tree เพื่อจัดกลุ่มข้อมูลที่มีความสัมพันธ์กัน ด้วยชุดข้อมูลที่คัดเลือก ประกอบด้วย 5 แอตทริบิวต์เพศ สถานะ มาตรการ การดื่มสุรา และผลการรักษา

5) การประเมินผล (Evaluation) คณะผู้จัดทำได้ทำการทดสอบโมเดล เพื่อวัดประสิทธิภาพที่ตรงกับความต้องการ ซึ่งการวัดประสิทธิภาพด้วยวิธี Self Consistency Test เพื่อดูแนวโน้มของโมเดลที่สร้างขึ้น ด้วยโปรแกรม Weka เวอร์ชัน 3.8.4 พบว่าเทคนิค Decision Tree: J48 ให้ผลลัพธ์การจำแนกประเภท ดื่ม(Drink) และไม่ดื่ม (No Drink) มีความถูกต้องถึง 70.12% และได้กฎจำนวน 9 กฎ ดังนี้

- กฎข้อที่ 1 IF SEX=FM Then alcohol=No Drink หมายความว่า ถ้าเพศที่เป็นเพศหญิงเกี่ยวเนื่องกัน ผลการพิจารณาพบว่า ผลการพิจารณาพบว่า ประเภทของการดื่มสุรานั้นจะมีผลลัพธ์เป็น “ไม่ดื่ม”
- กฎข้อที่ 2 IF SEX=M And status= Pedestrian Then alcohol=No Drink หมายความว่า ถ้าเพศที่เป็นเพศชายเกี่ยวเนื่องกัน และมีสถานะเป็นคนเดินเท้า ผลการพิจารณาพบว่า ประเภทของการดื่มสุรานั้นจะมีผลลัพธ์เป็น “ไม่ดื่ม”
- กฎข้อที่ 3 IF SEX=M And status= Passenger Then alcohol=No Drink หมายความว่า ถ้าเพศที่เป็นเพศชายเกี่ยวเนื่องกัน และมีสถานะเป็นผู้โดยสาร ผลการพิจารณาพบว่า ประเภทของการดื่มสุรานั้นจะมีผลลัพธ์เป็น “ไม่ดื่ม”
- กฎข้อที่ 4 IF SEX=M And status= Driver And safety = Belt Then alcohol=No Drink หมายความว่า ถ้าเพศที่เป็นเพศชายเกี่ยวเนื่องกัน มีสถานะเป็นผู้ขับขี่ และมีมาตรการป้องกันด้วยการคาดเข็มขัด ผลการพิจารณาพบว่า ประเภทของการดื่มสุรานั้นจะมีผลลัพธ์เป็น “ไม่ดื่ม”
- กฎข้อที่ 5 IF SEX = M And status = Driver And safety = Helmet Then alcohol = No Drink หมายความว่า ถ้าเพศที่เป็นเพศชายเกี่ยวเนื่องกัน มีสถานะเป็นผู้ขับขี่ และ

มีมาตรการป้องกันด้วยการใส่หมวกกันน็อค ผลการพิจารณาพบว่า ประเภทของการดื่มสุรานั้นจะมีผลลัพธ์เป็น “ไม่ดื่ม”

- กฎข้อที่ 6 IF SEX=M And status= Driver And safety = No Belt And result= Dis Then alcohol=Drink หมายความว่า ถ้าเพศที่เป็นเพศชายเกี่ยวเนื่องกัน มีสถานะเป็นผู้ขับขี่ไม่มีมาตรการป้องกันด้วยการคาดเข็มขัด และมีผลการรักษาจากการเกิดอุบัติเหตุเป็นทุเลา/หาย ผลการพิจารณาพบว่า ประเภทของการดื่มสุรานั้นจะมีผลลัพธ์เป็น “ดื่ม”
- กฎข้อที่ 7 IF SEX=M And status= Driver And safety = No Belt And result= Dead Then alcohol=No Drink หมายความว่า ถ้าเพศที่เป็นเพศชายเกี่ยวเนื่องกัน มีสถานะเป็นผู้ขับขี่ ไม่มีมาตรการป้องกันด้วยการคาดเข็มขัด และมีผลการรักษาจากการเกิดอุบัติเหตุเป็นตาย ผลการพิจารณาพบว่า ประเภทของการดื่มสุรานั้นจะมีผลลัพธ์เป็น “ไม่ดื่ม”
- กฎข้อที่ 8 IF SEX=M And status= Driver And safety = No Helmet And result= Dis Then alcohol=Drinkหมายความว่า ถ้าเพศที่เป็นเพศชายเกี่ยวเนื่องกัน มีสถานะเป็นผู้ขับขี่ ไม่มีมาตรการป้องกันด้วยการใส่หมวกกันน็อค และมีผลการรักษาจากการเกิดอุบัติเหตุเป็นทุเลา/หาย ผลการพิจารณาพบว่า ประเภทของการดื่มสุรานั้นจะมีผลลัพธ์เป็น “ดื่ม”
- กฎข้อที่ 9 IF SEX=M And status= Driver And safety = No Helmet And result= Dead Then alcohol=No Drink หมายความว่า ถ้าเพศที่เป็นเพศชายเกี่ยวเนื่องกัน มีสถานะเป็นผู้ขับขี่ ไม่มีมาตรการป้องกันด้วยการใส่หมวกกันน็อค และมีผลการรักษาจากการเกิดอุบัติเหตุเป็นตาย ผลการพิจารณาพบว่า ประเภทของการดื่มสุรานั้นจะมีผลลัพธ์เป็น “ไม่ดื่ม”

6) เผยแพร่ผลวิเคราะห์ (Deployment) คณะผู้จัดทำได้จัดทำเว็บไซต์สำหรับเผยแพร่ข้อมูลสารสนเทศนี้บน Web browser ด้วยการนำเสนอข้อมูลแบบ visualization เป็นการแสดงผลข้อมูลในรูปแบบของภาพโดยใช้โปรแกรม Tableau Public เป็นรูปแบบของรายงาน (Report) หรือแผนภาพ (Dashboard) โดยคณะผู้จัดทำได้ทำการศึกษาความรู้พื้นฐานทางด้านภาษาที่ใช้ในการพัฒนาเว็บไซต์ คือ ภาษาเอชทีเอ็มแอล (HTML) ภาษาพีเอชพี (PHP) ภาษาจาวาสคลิป (JavaScript) และ ซีเอสเอส (CSS) เมื่อศึกษาความรู้พื้นฐานในภาษาต่าง ๆ และการวิเคราะห์ข้อมูลด้วยเทคนิคดังที่กล่าวมาข้างต้นแล้วจึงทำการวิเคราะห์ข้อมูล และออกแบบเว็บไซต์ โดยเว็บไซต์มีส่วนประกอบต่าง ๆ ดังนี้ 1) ส่วนเนื้อหาเกี่ยวกับอุบัติเหตุ 2) ส่วนการวิเคราะห์ข้อมูลด้วย Decision Tree 3) ส่วนแบบฟอร์มจำลองการปฏิบัติตัวระหว่างเดินทาง

ในช่วงเทศกาลปีใหม่ 4) ส่วนสรุปผลข้อมูลเป็นรายปี 5) ส่วนการสรุปผลข้อมูลในภาพรวม 6) ส่วนของแนวทางป้องกันการเกิดอุบัติเหตุ 7) ส่วนของแบบสอบถาม

4.2.2 ขอบเขตผู้ใช้งานทั่วไปบนเว็บบราวเซอร์

1) ผู้ใช้งานทั่วไปสามารถดูข้อมูลสารสนเทศของผู้บาดเจ็บและเสียชีวิตในช่วงเทศกาลปีใหม่ประจำปี 2551 – 2558 และแหล่งความรู้ของข้อมูลทั่วไปที่เกี่ยวกับอุบัติเหตุ และการวิเคราะห์ข้อมูลที่เผยแพร่ผ่านทางเว็บไซต์

2) ผู้ใช้งานทั่วไปสามารถดูข้อมูลสารสนเทศในลักษณะตารางข้อมูลผู้บาดเจ็บและเสียชีวิตทั้งแบบรายปี และสรุปภาพรวมในทุกปี ก่อนการเลือกดูการแสดงผลข้อมูลในรูปแบบของแผนภาพ (Dashboard) จากโปรแกรม Tableau Public ในลักษณะรูปแบบของแผนภูมิชนิดต่าง ๆ อีกทั้งผู้ใช้งานทั่วไปสามารถกรองข้อมูลสารสนเทศผ่านระบบของ Tableau ผ่านทางเว็บไซต์ที่เผยแพร่ได้

คณะผู้จัดทำเพิ่มส่วนงานที่สำคัญและเป็นประโยชน์ต่อโครงงาน นอกเหนือไปจากขอบเขตของโครงงานที่ได้เสนอไว้ดังนี้

1) ส่วนของการสร้างแบบฟอร์มจำลองการปฏิบัติตัวระหว่างเดินทางในช่วงเทศกาลปีใหม่ โดยผลลัพธ์จะอ้างอิงจากข้อมูลที่ได้จากการทำ Decision tree ซึ่งจะแสดงคำแนะนำ และแนวทางการป้องกันการเกิดอุบัติเหตุ

2) ส่วนของแบบสอบถาม จัดทำขึ้นเพื่อเก็บข้อมูลเกี่ยวกับการประสบอุบัติเหตุของผู้เข้าชมเว็บไซต์ เพื่อนำข้อมูลที่เก็บรวบรวมได้ไปเผยแพร่ให้แก่บุคคลที่สนใจ และสามารถนำข้อมูลที่ได้ไปวิเคราะห์ในด้านที่สนใจต่อไป

4.3 บทสรุป

จากผลการดำเนินงานโครงงานในข้างต้นทั้งหมดนี้ คณะผู้จัดทำได้นำข้อมูลจากการวิเคราะห์ข้อมูล และการแสดงผลแบบ visualization ต่าง ๆ โดยใช้โปรแกรม Tableau Public มาเผยแพร่ให้กับบุคคลภายนอกหรือกลุ่มผู้ใช้ข้อมูลได้รับรู้ในรูปแบบของเว็บไซต์ ที่จะเป็นแหล่งความรู้ต่าง ๆ ในเรื่องของอุบัติเหตุ วิธีการและขั้นตอนในการสร้างโมเดล Decision Tree สำหรับผู้ใช้ทั่วไปที่ต้องการตัวอย่างในการศึกษาค้นคว้า และยังมีโปรแกรมประมวลผลลัพธ์ที่ตรงกับโมเดลที่สร้างขึ้นให้กับผู้ใช้ที่ต้องการคำแนะนำในการปฏิบัติตัวระหว่างเดินทางในช่วงเทศการปีใหม่ อีกทั้งยังมีแบบสอบถามสำหรับผู้ที่เข้าชมเว็บไซต์ได้ร่วมตอบคำถามในเรื่องของความปลอดภัย เพื่อสามารถนำข้อมูลไปวิเคราะห์ต่อได้

บทที่ 5
สรุปและข้อเสนอแนะ

ผลการวิเคราะห์ข้อมูลผู้ประสบอุบัติเหตุในช่วงเทศกาลตั้งแต่ปี 2551 – 2558 และจัดทำเว็บไซต์เพื่อเผยแพร่ข้อมูล ทางคณะผู้จัดทำได้ทำการประเมินผลการใช้งานเว็บไซต์จากผู้ใช้งานทั่วไป เพื่อสรุปผลการทำโครงงาน ข้อจำกัดของเว็บไซต์ ปัญหาและอุปสรรคต่าง ๆ ของการทำโครงงาน และข้อเสนอแนะเพื่อพัฒนาโครงงาน ดังนี้

5.1 บทสรุปผลโครงงาน

จากที่คณะผู้จัดทำได้วิเคราะห์ข้อมูลและจัดทำเว็บไซต์เพื่อเผยแพร่ข้อมูลผู้ประสบอุบัติเหตุในช่วงเทศกาลตั้งแต่ปี 2551 – 2558 จนแล้วเสร็จ จากนั้น จึงได้ทำการประเมินผลการใช้งานเว็บไซต์จากกลุ่มตัวอย่างจำนวน 50 คน โดยการทำแบบสอบถามความพึงพอใจในการเข้าใช้งานเว็บไซต์เผยแพร่ข้อมูลผู้ประสบอุบัติเหตุ ปี 2551 – 2558 จากกลุ่มตัวอย่างจำนวน 50 คน แบ่งเป็นเพศหญิง 34 คน คิดเป็นร้อยละ 68 เพศชายจำนวน 16 คน คิดเป็นร้อยละ 32 อายุต่ำกว่า 18 ปี 5 คน คิดเป็นร้อยละ 10 อายุตั้งแต่ 18 – 25 ปีจำนวน 42 คน คิดเป็นร้อยละ 84 อายุ 25 ปีขึ้นไป 3 คน คิดเป็นร้อยละ 6 โดยเกณฑ์การให้คะแนนของผู้ใช้งานเว็บไซต์มีดังนี้ ระดับความพึงพอใจของผู้ใช้งานเว็บไซต์ ระดับความพึงพอใจ 5 = มากที่สุด 4 = มาก 3 = ปานกลาง 2 = น้อย 1 = น้อยที่สุด

เมื่อนำคำตอบของผู้ตอบแบบสอบถามมาแจกแจงความถี่และหาค่าเฉลี่ย แล้วกำหนดระดับค่าเฉลี่ย แบ่งระดับความพึงพอใจได้ดังนี้

$$\text{ความกว้างของอันตรภาคชั้น} = \frac{(\text{คะแนนสูงสุด} - \text{คะแนนต่ำสุด})}{\text{จำนวนชั้น}} = \frac{(5-1)}{5} = 0.8$$

จากเกณฑ์ดังกล่าว สามารถแปลความหมายของความพึงพอใจได้ดังนี้

คะแนนค่าเฉลี่ย 1.00 – 1.80 หมายถึง ระดับความพึงพอใจน้อยที่สุด

คะแนนค่าเฉลี่ย 1.81 – 2.60 หมายถึง ระดับความพึงพอใจน้อย

คะแนนค่าเฉลี่ย 2.61 – 3.40 หมายถึง ระดับความพึงพอใจปานกลาง

คะแนนค่าเฉลี่ย 3.41 – 4.20 หมายถึง ระดับความพึงพอใจมาก

คะแนนค่าเฉลี่ย 4.21 – 5.00 หมายถึง ระดับความพึงพอใจมากที่สุด

(อ้างอิงจาก: สรุปแบบประเมินความพึงพอใจของผู้รับบริการศูนย์ดำรงธรรมจังหวัดราชบุรี)

เนื้อหาที่ได้ทำการประเมินและคะแนนที่ได้มีดังนี้

1) แหล่งความรู้ที่อ้างอิงมีความชัดเจน และสอดคล้องกับงาน ได้คะแนนค่าเฉลี่ย 4.41 คะแนน อยู่ในระดับพึงพอใจมากที่สุด

2) อธิบายขั้นตอนการวิเคราะห์ข้อมูล Decision Tree และอธิบายขั้นตอนการใช้งานในโปรแกรม Waka 3.8.4 ได้ดีและเข้าใจง่าย ได้คะแนนค่าเฉลี่ย 4.10 คะแนน อยู่ในระดับพึงพอใจมาก

3) ผลลัพธ์ที่ได้จากการสร้างโมเดลมีความน่าเชื่อถือ และโปรแกรมแบบจำลองของโมเดลมีประสิทธิภาพ ได้คะแนนค่าเฉลี่ย 4.22 คะแนน อยู่ในระดับพึงพอใจมากที่สุด

4) การเลือกใช้โปรแกรม Tableau Public แสดงผลแบบ Visualization ได้เหมาะสมกับข้อมูล ได้คะแนนค่าเฉลี่ย 4.31 คะแนน อยู่ในระดับพึงพอใจมากที่สุด

5) รูปแบบ visualization ที่หลากหลายของการนำเสนอข้อมูลผู้ประสบอุบัติเหตุในแต่ละปีและการนำเสนอแบบสรุปรวมทุกปี ได้คะแนนค่าเฉลี่ย 4.33 คะแนน อยู่ในระดับพึงพอใจมากที่สุด

6) แดชบอร์ดนำเสนอข้อมูลสามารถเข้าใจได้ง่าย และเลือกใช้งานแต่ละแดชบอร์ดได้สะดวกได้คะแนนค่าเฉลี่ย 4.37 คะแนน อยู่ในระดับพึงพอใจมากที่สุด

7) การตอบสนองต่อการกรองดูข้อมูลผู้ประสบอุบัติเหตุในช่วงเทศกาลปีใหม่ตามความต้องการได้อย่างรวดเร็ว ได้คะแนนค่าเฉลี่ย 4.43 คะแนน อยู่ในระดับพึงพอใจมากที่สุด

8)การนำเสนอแนวทางการป้องกันอุบัติเหตุและมาตรการในการป้องกันตัวเอง ครอบคลุมและประสิทธิภาพ ได้คะแนนค่าเฉลี่ย 4.39 คะแนน อยู่ในระดับพึงพอใจมากที่สุด

9) การออกแบบเว็บไซต์ การจัดเรียงเนื้อหา การตกแต่ง และการใช้งาน ได้คะแนนค่าเฉลี่ย 4.47 คะแนนอยู่ในระดับพึงพอใจมากที่สุด

10) ประโยชน์ที่ได้จากการใช้งานเว็บไซต์ ได้คะแนนค่าเฉลี่ย 4.41 คะแนน อยู่ในระดับพึงพอใจมากที่สุด

จากคะแนนในแต่ละด้านที่กลุ่มผู้เข้าใช้งานเว็บไซต์ได้ประเมินผ่านแบบสอบถามความพึงพอใจในการเข้าใช้งานเว็บไซต์เผยแพร่ข้อมูลผู้ประสบอุบัติเหตุ ปี 2551 – 2558 ได้คะแนนเฉลี่ยรวม 4.34 คะแนน ซึ่งอยู่ในระดับพึงพอใจมากที่สุด

5.2 ข้อจำกัดของระบบ

5.2.1 ผู้ใช้ไม่สามารถอัปโหลดชุดข้อมูลเข้าไปในเว็บไซต์ได้ เนื่องจากเป็นเว็บไซต์เผยแพร่ข้อมูลเพียงอย่างเดียว ซึ่งมีเพียงผู้จัดทำที่สามารถปรับปรุงข้อมูลต่าง ๆ ได้

5.2.2 การแสดงผลข้อมูลอาจไม่ครอบคลุมทั้งหมดทำให้เว็บไซต์เผยแพร่ข้อมูลนี้ไม่อาจตอบสนองความต้องการข้อมูลให้แก่ผู้ใช้อย่างครอบคลุม เช่น เว็บไซต์ไม่มีข้อมูลที่เป็นปัจจุบันซึ่งอาจไม่ตรงกับความต้องการทราบข้อมูลแบบทันเหตุการณ์ของผู้ใช้งาน

5.3 ปัญหาและอุปสรรคของโครงงาน

5.3.1 ข้อจำกัดของข้อมูลที่ไม่ได้ทำการจัดเก็บรวบรวมด้วยตนเองและเนื่องจากเป็นชุดข้อมูลแบบเปิดทำให้ต้องใช้ระยะเวลาในการศึกษา

5.3.2 การทำความสะอาดข้อมูลจำนวนมากทำให้ใช้ระยะเวลานานและมีความล่าช้าว่าที่คาดการณ์ไว้

5.3.3 ข้อจำกัดของโปรแกรมต่าง ๆ ที่นำมาใช้ในการวิเคราะห์ข้อมูล เนื่องจากผู้เคราะห์ข้อมูลใช้แบบเวอร์ชันทดลองทำให้โปรแกรมมีข้อจำกัดหลายอย่าง เช่น ในโปรแกรม WEKA เวอร์ชัน 3.8.4 ต้องซื้อแพ็กเก็จภาษาไทยมาติดตั้งจึงจะสามารถนำเข้าข้อมูลเป็นภาษาไทยได้

5.4 ข้อเสนอแนะ

การวิเคราะห์ข้อมูลผู้ประสบอุบัติเหตุในช่วงเทศกาลตั้งแต่ปี 2551 – 2558 เพื่อเผยแพร่บนเว็บไซต์ ที่จัดทำเสร็จสิ้นแล้วนี้แม้จะสามารถทำงานและแสดงผลข้อมูลได้ถูกต้องตรงตามวัตถุประสงค์และขอบเขตที่ตั้งไว้ แต่ก็ยังมีข้อจำกัดหลายประการ ซึ่งหากจะพัฒนาให้เว็บไซต์เผยแพร่ข้อมูลนี้แสดงผลข้อมูลและทำงานได้อย่างมีประสิทธิภาพมากขึ้น ผู้พัฒนาควรจะต้องปรับปรุงในส่วนต่าง ๆ ดังต่อไปนี้

5.4.1 หากมีชุดข้อมูลเพิ่มเติมควรมีการปรับปรุงข้อมูลใหม่ ๆ อยู่เสมอ

5.4.2 ควรจัดเก็บข้อมูลด้วยตัวเองเพื่อให้ได้ข้อมูลที่ตรงต่อความต้องการในการวิเคราะห์ข้อมูลในครั้งต่อไป

5.4.3 ควรมีการพัฒนาหน้าการแสดงผลต่าง ๆ ของเว็บไซต์ให้เหมาะสม สะดวก และเข้าใจง่ายมากยิ่งขึ้น เพื่อลดความผิดพลาดและเพิ่มประสิทธิภาพในการนำเสนอข้อมูล

บรรณานุกรม

ต้นไม้ตัดสินใจ. สืบค้นเมือง 3 มีนาคม 2563, จากวิกิพีเดีย https://th.wikipedia.org/wiki/ต้นไม้ตัดสินใจ.

ณัฏฐา กาญจนขุนดี. (2561). **Big Data ข้อมูลขนาดใหญ่คืออะไร ทำงานอย่างไร และนำไปใช้อะไรได้บ้าง**. สืบค้นเมื่อ วันที่ 8 กุมภาพันธ์ 2563, สืบค้นจาก https://www.khundee.com/ big-data/.

ธีระพงษ์ กระการดี. (มปป.). **สถิติ**. สืบค้นเมื่อวันที่ 12 กุมภาพันธ์ 2563, สืบค้นจากhttp://www.stvc.ac.th/elearning/stat/csu1.html.

เบญจมาศ ปิยะ. (2558). **การศึกษาพฤติกรรมผู้ใช้งานตู้แช่เชิงพาณิชย์ด้วยเทคนิคต้นไม้ตัดสินใจ**. กรุงเทพฯ : สาขาวิชาเทคโนโลยีสารสนเทศและการจัดการ มหาวิทยาลัยกรุงเทพ.

ปฏิพล ธีร์โรจพร. (2562). **Data Visualization ไม่ได้ทำให้สวย แต่ทำให้เข้าใจง่าย**. สืบค้นเมื่อวันที่ 5 กุมภาพันธ์ 2563, สืบค้นจาก https://blog.1moby.com/ds-ep-4-data-visualization/.

ปัทมา เที่ยงสมบุญ และ นิเวศ จิระวิชิตชัย. (2561). **การพัฒนาระบบธุรกิจอัจฉริยะเพื่อสนับสนุนการพยากรณ์และการตัดสินใจของผู้บริหาร**. วารสารวิทยาศาสตร์และเทคโนโลยี, ปีที่ 5 (ฉบับที่ 4 เดือนกรกฎาคม – สิงหาคม 2561).

ภูมิพัฒน์ ดวงกลาง และรัจนา เครือแก้ว. (2562). **แบบจำลองการทำนายแบบอากาศยานจากข้อมูลเป้าหมายไม่ทราบฝ่ายอัตโนมัติ**. สืบค้นจาก https://www.tci-thaijo.org/index.php/nkrafa-sct/article/download/164341/155011/.

รัตนา สุวรรณวิชนีย์ และปราลี มณีรัตน์. (2560). **การพัฒนาระบบรายงานรูปแบบหลายมิติเพื่อสนับสนุนการตัดสินใจ เชิงนโยบายของสำนักงานคณะกรรมการวิจัยแห่งชาติ**. ปีที่ 9, สืบค้นจาก https://www.tci-thaijo.org/index.php/spurst/article/view/122113.

รุจิรา ธรรมสมบัติ. (2554). **ระบบสนับสนุนการตัดสินใจในการเลือกใช้แพคเกจอินเทอร์เน็ตมือถือ โดยใช้ต้นไม้ตัดสินใจ**. วิทยาลัยราชพฤกษ์.

หทัยพร หวังเชย และ ณัฐพร นันทจิระพงศ์. (2561). **ระบบสารสนเทศเพื่อการบริหารจัดการรายงานการดำเนินงานของโรงแรมด้วย Power BI**. วารสารวิทย์เทคโนหัวเฉียวเฉลิมพระเกียรติ, ปีที่ 4 (ฉบับที่ 1 มกราคม – มิถุนายน 2561).

สิทธิชัย รักษาสุข. (มปป.). **CSS คืออะไร**. สืบค้นเมื่อ วันที่ 8 กุมภาพันธ์ 2563, สืบค้นจาก https://www.dwthai.com/dwarticle/.

ศจี วาณิช. (2558). **Data Mining (เหมืองข้อมูล)**. สืบค้นเมื่อ วันที่ 5 กุมภาพันธ์ 2563, สืบค้นจาก http://sajeegm301.blogspot.com/2015/11/data-mining.html.

อัจฉราภรณ์ จุฑาผาด. (2559). **งานวิจัยเรื่องการพัฒนาระบบสารสนเทศเพื่อการพยากรณ์จำนวนนักศึกษาใหม่ โดยใช้กฎการจำแนกต้นไม้ตัดสินใจ**. (ฉบับที่ 278). นเรศวรวิจัย ครั้งที่ 12: วิจัยและนวัตกรรมกับการพัฒนาประเทศ.

อภิยศ เหรียญวิพัฒน์. (ม.ป.ป.). **การวิเคราะห์ข้อมูลและการนำเสนอข้อมูลเชิงภาพ เพื่อประยุกต์ใช้กับ การแสดงข้อมูลสถิติของสำนักหอสมุดมหาวิทยาลัยเกษตรศาสตร์**. (ฉบับที่ 6), 117-126 สืบค้นจาก https://pulinet.oas.psu.ac.th/index.php/journal/article /view/381.

อาทิตย์ สิทธิบรรเจิด. (2552). **การนำ Visualization ไปประยุกต์ใช้กับการแสดงข้อมูลบนระบบ Knowledge Management**. กรุงเทพฯ: มหวิทยาลัยกรุงเทพ.

อุทัย เหตุผล. (มปป.). **หลักการออกแบบเว็บไซต์**. สืบค้นเมื่อ วันที่ 8 กุมภาพันธ์ 2563, สืบค้นจาก https://sites.google.com/a/mattayom31.go.th/npw-html/html_design.

เอกพล วิชัยสุข. **การนำเสนอแผนภาพข้อมูล (Data Visualization).** (2561, 15 พฤษภาคม). สืบค้นเมื่อ 23 กุมภาพันธ์ 2563, สืบค้นจาก 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 8 กุมภาพันธ์ 2563, สืบค้นจากhttps://uaychaiblog.com/2018/05/15/big-data-characteristics/.

codebee. **Bootstrap คืออะไร.** (2559, 3 กันยายน). สืบค้นเมื่อ 20 เมษายน 2563, สืบค้นจาก https://www.codebee.co.th/labs/bootstrap-คืออะไร/.

ภาคผนวก

ภาคผนวก ก
คู่มือการใช้งานเว็บไซต์

คู่มือการใช้งานเว็บไซต์

จากการดำเนินงานวิเคราะห์ข้อมูลผู้ประสบอุบัติเหตุในช่วงเทศกาลประจำปี 2551 – 2558 เพื่อใช้สำหรับเผยแพร่ข้อมูลบนเว็บไซต์ โดยเว็บไซต์มีส่วนใช้งาน 7 ส่วน ดังนี้ ส่วนเนื้อหาเกี่ยวกับอุบัติเหตุ ส่วนการวิเคราะห์ข้อมูลด้วย Decision Tree ส่วนแบบฟอร์มจำลองการปฏิบัติตัวระหว่างเดินทางในช่วงเทศกาลปีใหม่ ส่วนสรุปผลข้อมูลเป็นรายปี ส่วนการสรุปผลข้อมูลในภาพรวม ส่วนของแนวทางป้องกันการเกิดอุบัติเหตุ ส่วนของแบบสอบถาม ผู้จัดทำจึงได้ทำคู่มืออธิบายการใช้งานเว็บไซต์ ดังต่อไปนี้

คู่มือการใช้งานสำหรับผู้เข้าใช้งานเว็บไซต์

1) หน้าแรกของเว็บไซต์ เมื่อเข้าสู่เว็บไซต์การวิเคราะห์ข้อมูลอุบัติเหตุ

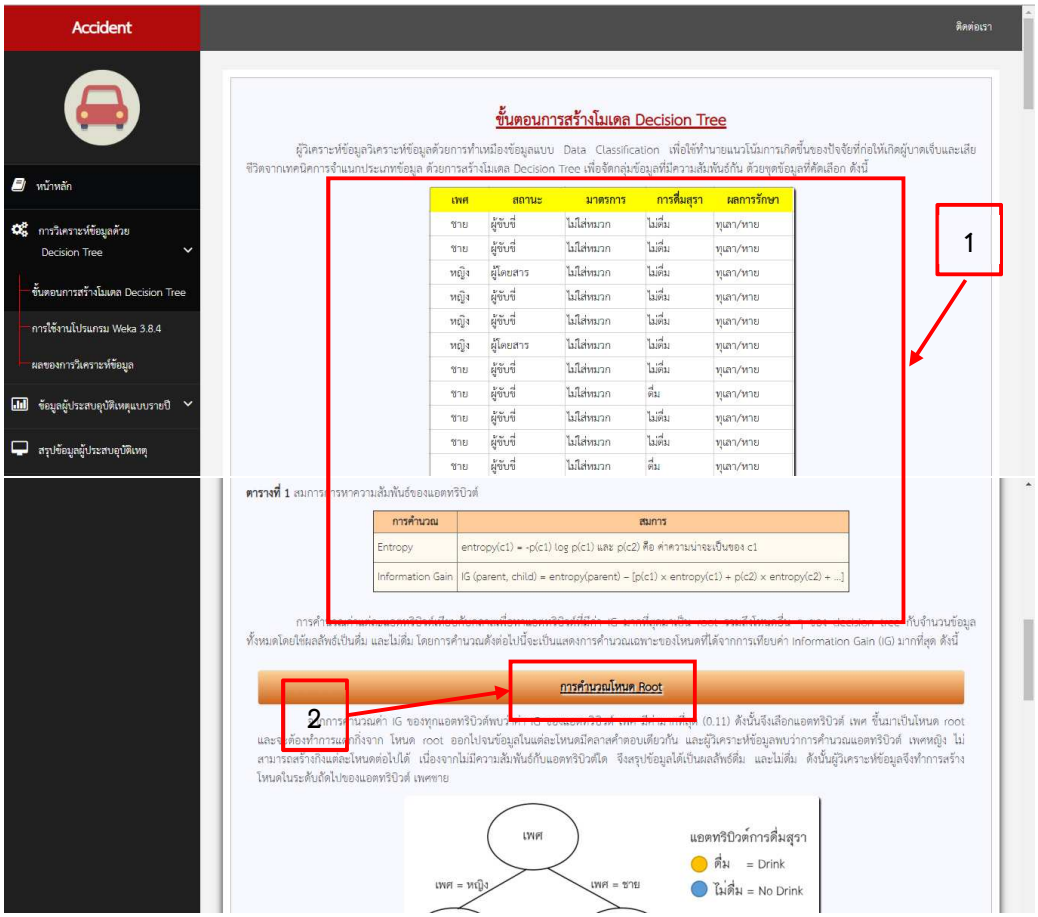
ภาพที่ ก.1 แสดงหน้าแรกของเว็บไซต์

หมายเลข 1 ส่วนของแถบเมนูด้านข้างที่ประกอบไปด้วยเมนูการใช้งานเพื่อไปยังหน้าต่าง ๆ ภายในเว็บไซต์

หมายเลข 2 ส่วนการติดต่อกับผู้จัดทำเว็บไซต์

หมายเลข 3 ส่วนแสดงเนื้อหาในหน้าแรก ที่ประกอบไปด้วยเนื้อหาเกี่ยวกับอุบัติเหตุและขั้นตอนการวิเคราะห์ข้อมูล

2) การแสดงรายละเอียด วิธีการ ขั้นตอนและผลการวิเคราะห์ข้อมูลด้วยโมเดล Decision Tree

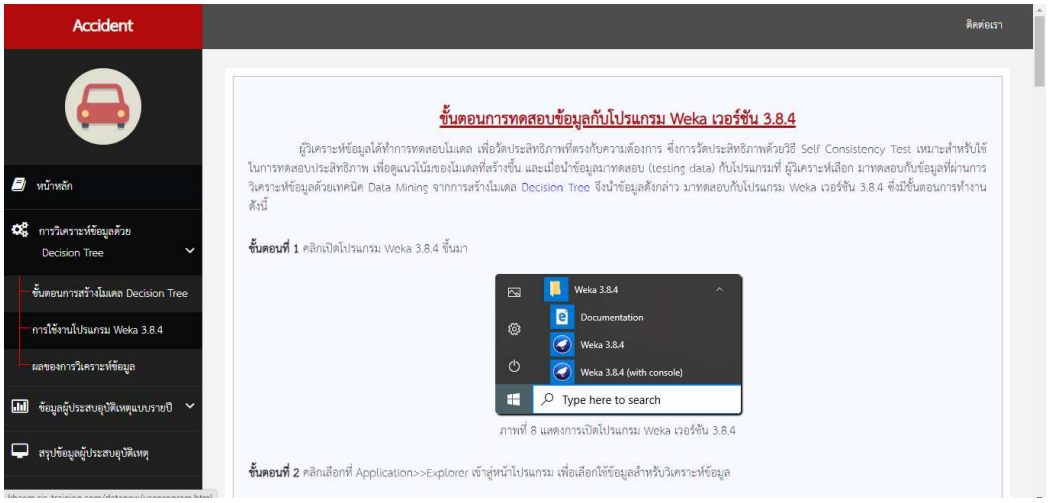


ภาพที่ ก.2 แสดงขั้นตอนการสร้าง Decision Tree

หมายเลข 1 ส่วนแสดงเนื้อหาอธิบายขั้นตอนการสร้างโมเดล Decision Tree

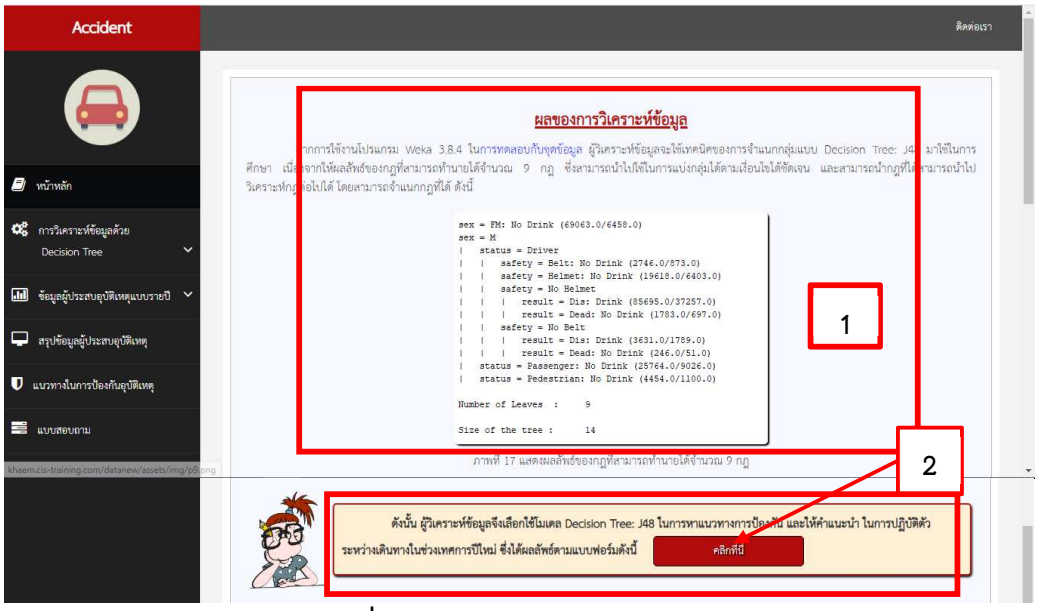
หมายเลข 2 ปุ่มอธิบายการคำนวณโหนดในระดับต่าง ๆ จากการวิเคราะห์ข้อมูล

3) การแสดงเนื้อหาการอธิบายขั้นตอนการทดสอบกับข้อมูลที่คัดเลือก โดยใช้โปรแกรม Weka 3.8.4 เพื่อดูผลลัพธ์ของ Decision tree



ภาพที่ ก.3 แสดงการทดสอบกับข้อมูลโดยใช้โปรแกรม Weka 3.8.4

4) การแสดงเนื้อหาของผลการวิเคราะห์ข้อมูล

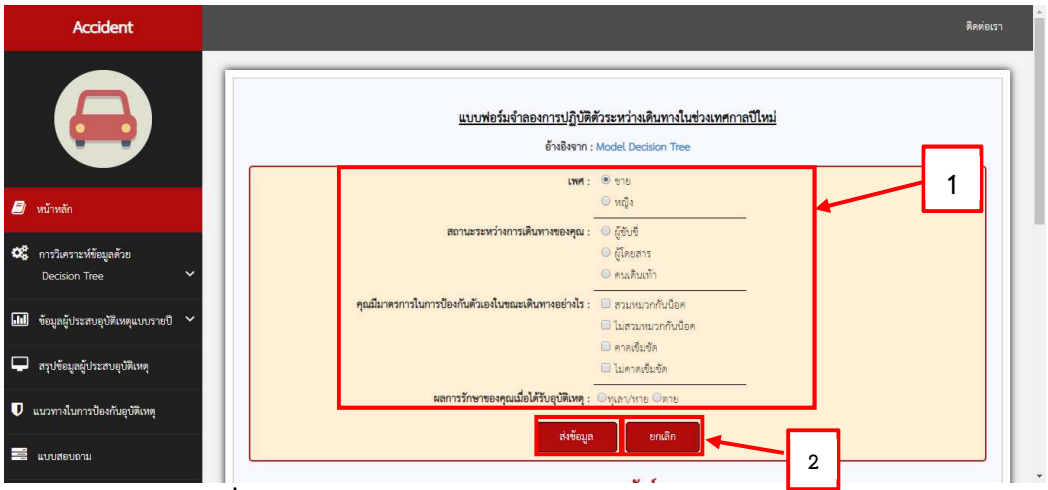


ภาพที่ ก.4 แสดงผลการวิเคราะห์ข้อมูล

หมายเลข 1 ส่วนแสดงเนื้อหาของผลการวิเคราะห์ข้อมูล Decision Tree

หมายเลข 2 ปุ่มเมนูเพื่อเข้าใช้งานแบบฟอร์มจำลองการปฏิบัติตัวระหว่างเดินทางในช่วงเทศกาลปีใหม่

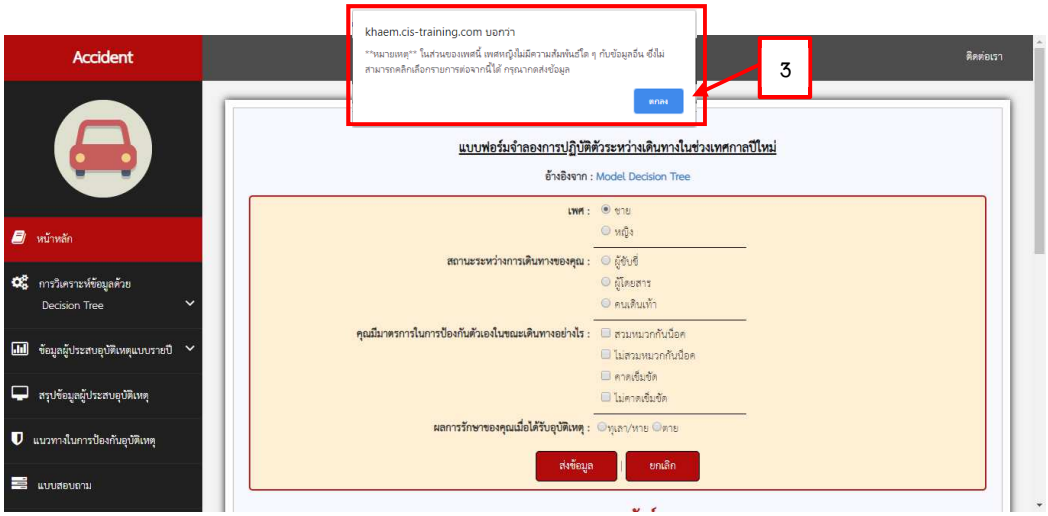
5) หน้าที่สี่แสดงแบบฟอร์มจำลองการปฏิบัติตัวระหว่างเดินทางในช่วงเทศกาลปีใหม่



ภาพที่ ก.5 แสดงแบบฟอร์มจำลองการปฏิบัติตัวระหว่างเดินทาง

หมายเลข 1 ส่วนแบบฟอร์มจำลองการปฏิบัติตัวระหว่างเดินทางในช่วงเทศกาลปีใหม่ โดยจะแสดงผลลัพธ์ตามการวิเคราะห์ข้อมูล Decision Tree

หมายเลข 2 ปุ่มส่งข้อมูลเพื่อวิเคราะห์ผลจากการเลือกตัวเลือกในฟอร์ม และปุ่มยกเลิก



ภาพที่ ก.6 แสดงการแจ้งเตือนหมายเหตุจากแบบฟอร์ม

หมายเลข 3 การแจ้งเตือนหมายเหตุเมื่อมีการเลือกรายการ เพื่อให้ผลลัพธ์เป็นไปตามเงื่อนไขของกฎที่ได้จาก Decision tree

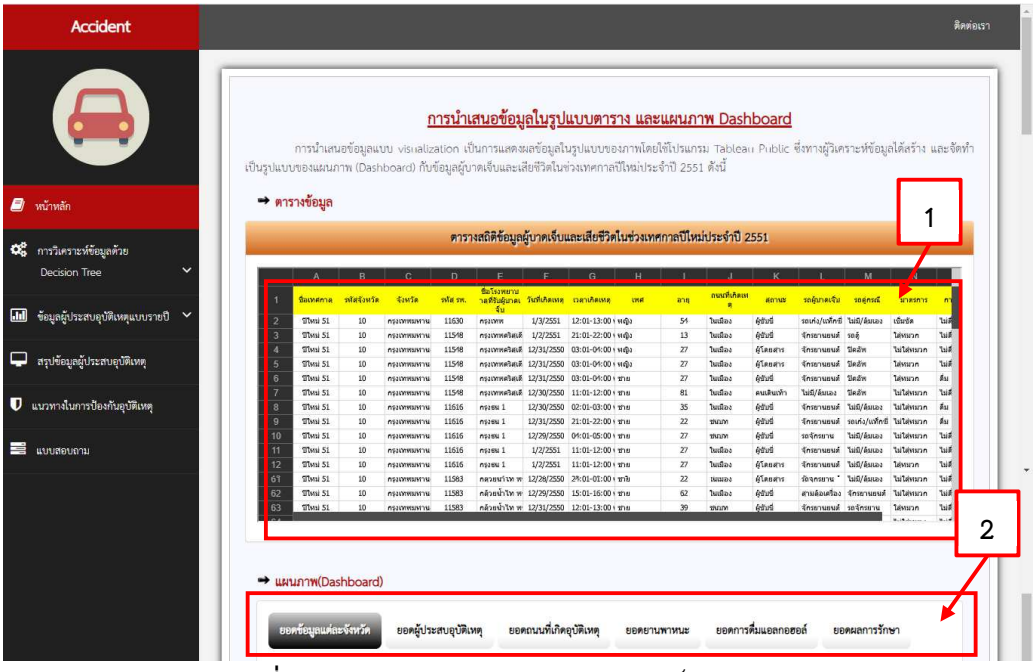
6) หน้าแสดงผลลัพธ์ต่อจากแบบฟอร์มที่ส่งข้อมูล

ภาพที่ ก.7 แสดงผลลัพธ์ของการส่งข้อมูลจากแบบฟอร์ม

หมายเลข 1 แสดงผลลัพธ์ที่ได้จากการเลือกตัวเลือกในแบบฟอร์ม ซึ่งผลลัพธ์นี้อ้างอิงมาจากการทำโมเดล Decision Tree

หมายเลข 2 แสดงคำแนะนำและแนวทางป้องกันการเกิดอุบัติเหตุตามความเหมาะสมจากผลลัพธ์ที่แสดง

7) หน้าแสดงตารางสถิติข้อมูลผู้บาดเจ็บและเสียชีวิตในช่วงเทศกาลปีใหม่



ภาพที่ ก.8 แสดงตารางสถิติข้อมูลผู้บาดเจ็บและเสียชีวิต

หมายเลข 1 แสดงตารางสถิติข้อมูลผู้บาดเจ็บและเสียชีวิตในช่วงเทศกาลปีใหม่ โดยแต่ละปีจะใช้แสดงในรูปแบบเดียวกัน

หมายเลข 2 ปุ่มเลือกดูการแสดงยอดข้อมูลผู้ประสบอุบัติเหตุในแต่ละแดชบอร์ด

8) หน้าที่ห้าหน้าแสดงผลจากแดชบอร์ดแสดงผลสรุปของข้อมูลแต่ละด้านในแต่ละปี

ภาพที่ ก.9 แสดงแดชบอร์ดผลสรุปของข้อมูลแต่ละด้าน

หมายเลข 1 เมนูแสดงข้อมูลผู้ประสบอุบัติเหตุในแต่ละปี ผู้ใช้สามารถเลือกดูข้อมูลได้ตามความต้องการ

หมายเลข 2 ปุ่มเลือกดูการแสดงยอดข้อมูลผู้ประสบอุบัติเหตุในแต่ละแดชบอร์ด

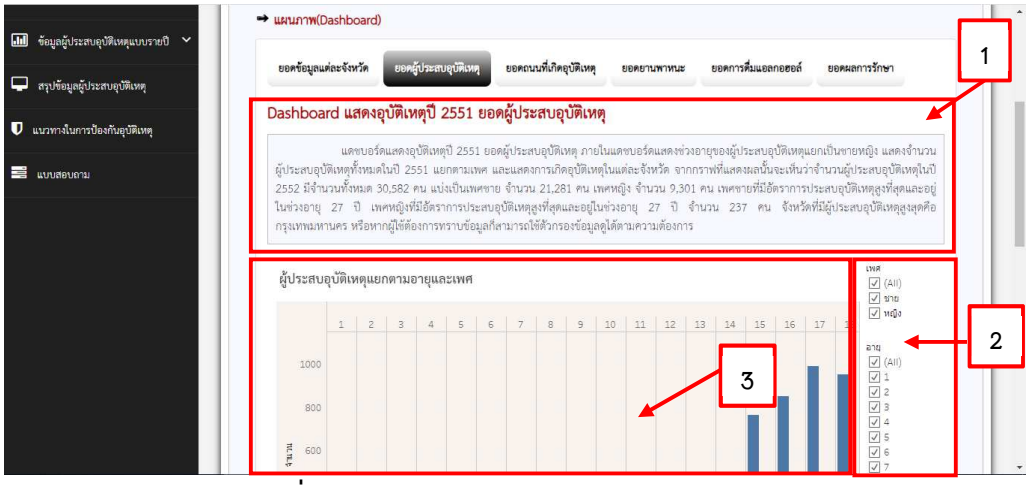
หมายเลข 3 ส่วนอธิบายข้อมูลภายในแดชบอร์ด

หมายเลข 4 ตัวกรองข้อมูลของแดชบอร์ดแสดงผล ผู้ใช้สามารถใช้ตัวกรองนี้กรองดูข้อมูลตามความต้องการได้

หมายเลข 5 หากแดชบอร์ดที่แสดงผลเป็นแบบกราฟจะมีเครื่องมือจัดเรียงข้อมูล สามารถคลิกเลือกเพื่อดูการเรียงลำดับข้อมูลได้จากตรงนี้ ทั้งเรียงจากน้อยไปมาก มากไปน้อย และเรียงข้อมูลตามตัวอักษร

หมายเลข 6 ส่วนของแดชบอร์ดแสดงผล

9) หน้าแดชบอร์ดแสดงยอดผู้ประสบอุบัติเหตุ



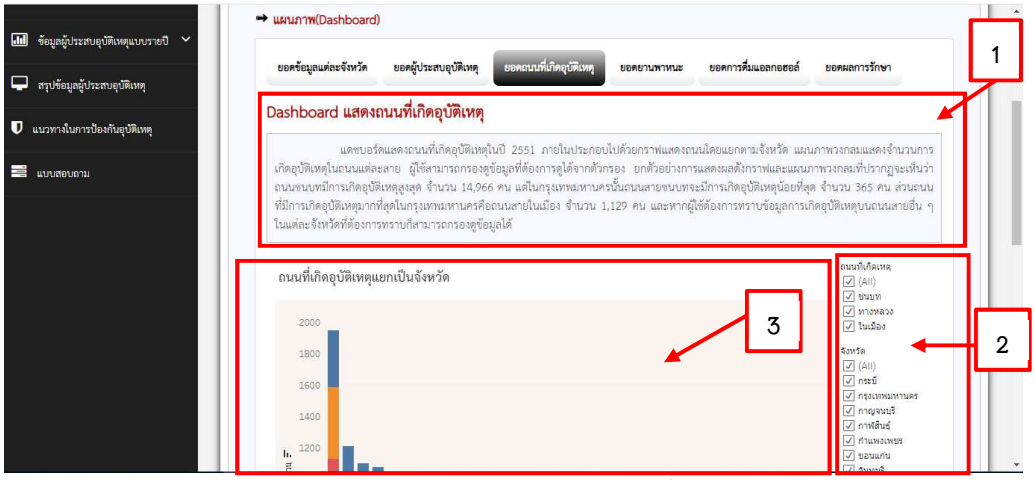
ภาพที่ ก.10 แดชบอร์ดแสดงยอดผู้ประสบอุบัติเหตุ

หมายเลข 1 ส่วนอธิบายข้อมูลภายในแดชบอร์ด

หมายเลข 2 ตัวกรองข้อมูลของแดชบอร์ดแสดงผล ผู้ใช้สามารถใช้ตัวกรองนี้กรองดูข้อมูลตามความต้องการได้

หมายเลข 3 ส่วนของแดชบอร์ดแสดงผล

10) หน้าแดชบอร์ดแสดงยอดถนนที่เกิดอุบัติเหตุ



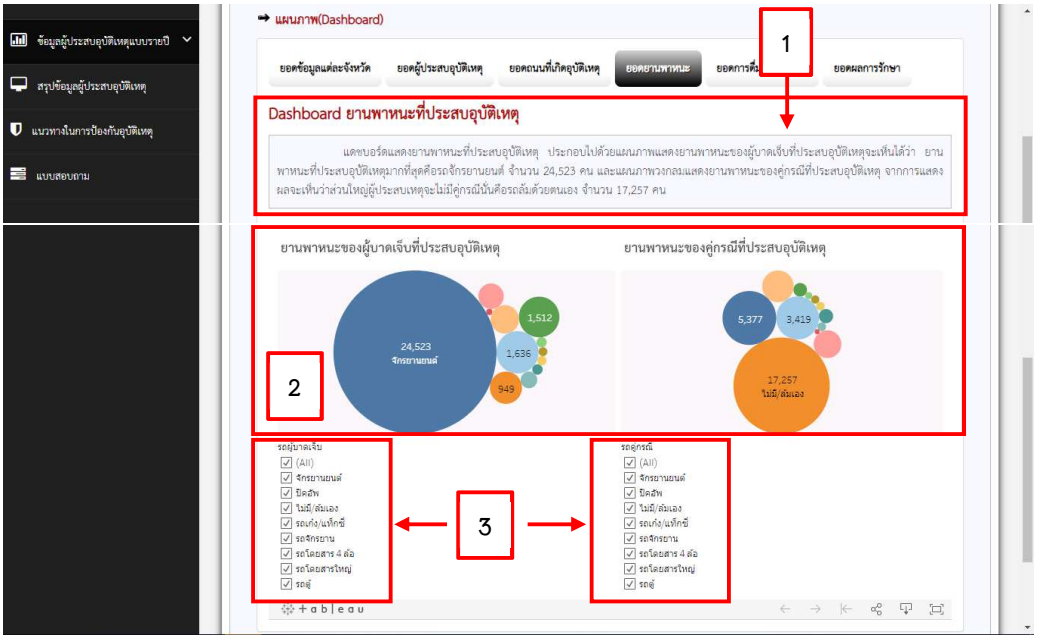
ภาพที่ ก.11 แดชบอร์ดแสดงยอดถนนที่เกิดอุบัติเหตุ

หมายเลข 1 ส่วนอธิบายข้อมูลภายในแดชบอร์ด

หมายเลข 2 ตัวกรองข้อมูลของแดชบอร์ดแสดงผล ผู้ใช้สามารถใช้ตัวกรองนี้กรองดูข้อมูลตามความต้องการได้

หมายเลข 3 ส่วนของแดชบอร์ดแสดงผล

11) หน้าแดชบอร์ดแสดงยานพาหนะที่เกิดอุบัติเหตุ



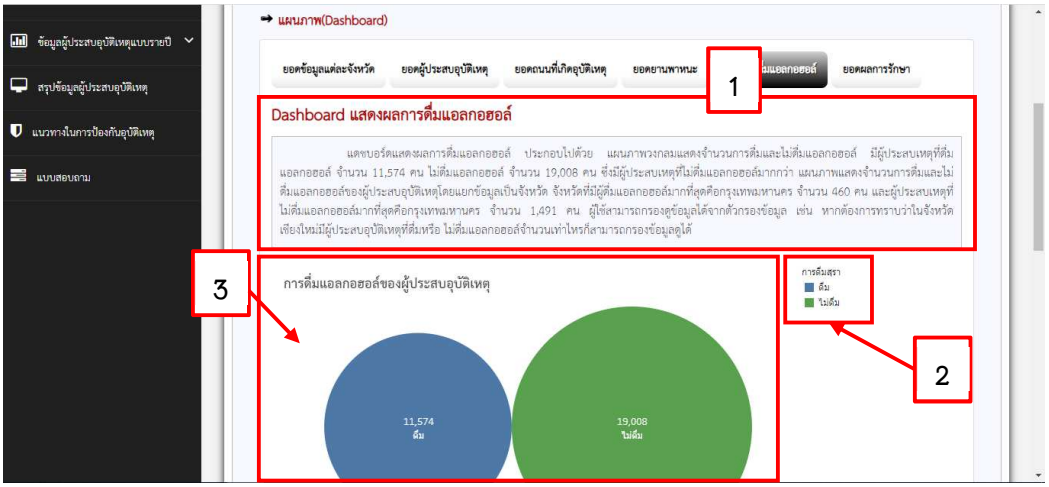
ภาพที่ ก.12 แดชบอร์ดแสดงยานพาหนะที่เกิดอุบัติเหตุ

หมายเลข 1 ส่วนอธิบายข้อมูลภายในแดชบอร์ด

หมายเลข 2 และ 3 ตัวกรองข้อมูลของแดชบอร์ดแสดงผล ผู้ใช้สามารถใช้ตัวกรองนี้กรองดูข้อมูลตามความต้องการได้

หมายเลข 4 และ 5 ส่วนของแดชบอร์ดแสดงผล

12) หน้าแดชบอร์ดแสดงการดื่มแอลกอฮอล์



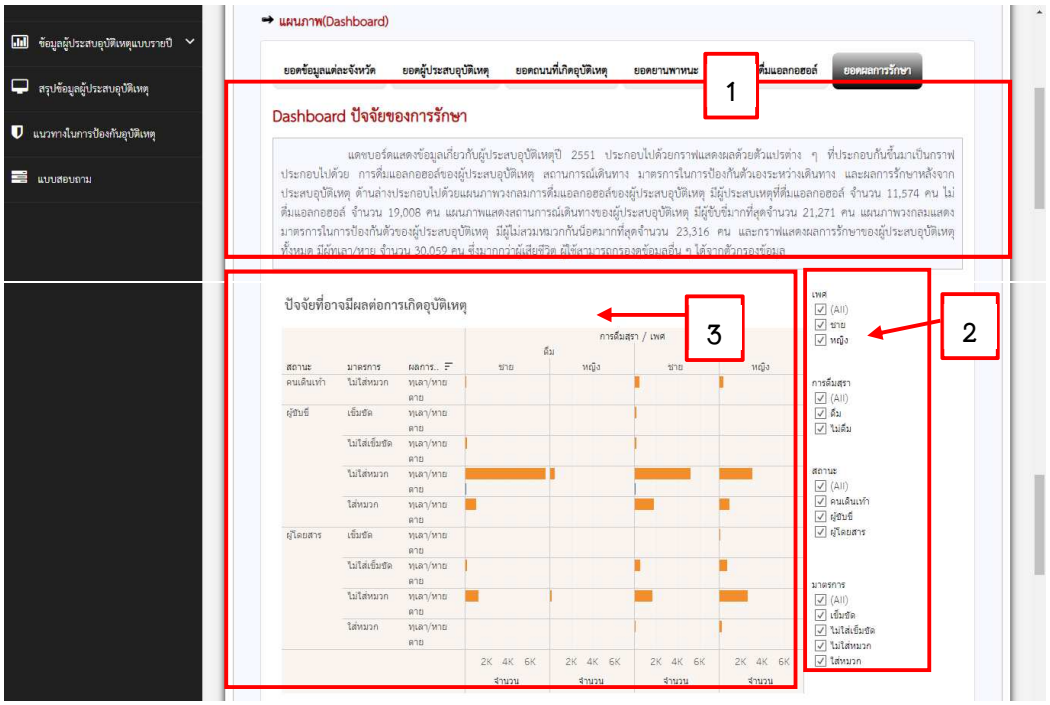
ภาพที่ ก.13 แดชบอร์ดแสดงการดื่มแอลกอฮอล์

หมายเลข 1 ส่วนอธิบายข้อมูลภายในแดชบอร์ด

หมายเลข 2 ตัวกรองข้อมูลของแดชบอร์ดแสดงผล ผู้ใช้สามารถใช้ตัวกรองนี้กรองดูข้อมูลตามความต้องการได้

หมายเลข 3 ส่วนของแดชบอร์ดแสดงผล

13) แดชบอร์ดแสดงยอดผลการรักษา



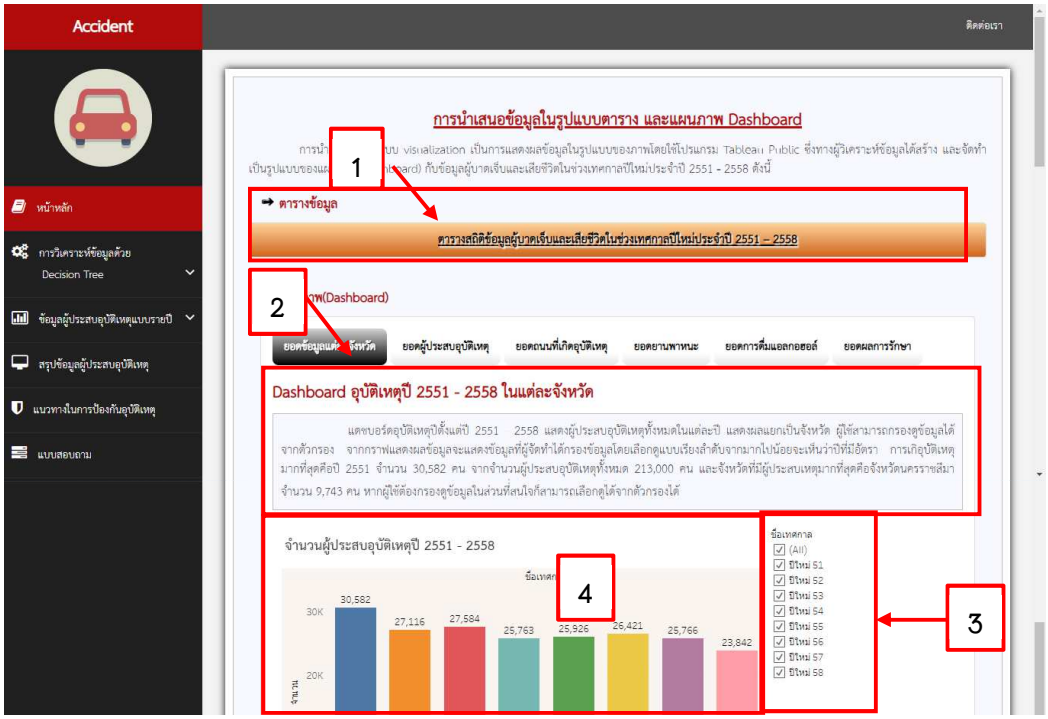
ภาพที่ ก.14 แดชบอดร์แสดงยอดผลการรักษา

หมายเลข 1 ส่วนอธิบายข้อมูลภายในแดชบอร์ด

หมายเลข 2 ตัวกรองข้อมูลของแดชบอร์ดแสดงผล ผู้ใช้สามารถใช้ตัวกรองนี้กรองดูข้อมูลตามความต้องการได้

หมายเลข 3 ส่วนของแดชบอร์ดแสดงผล

14) หน้าแสดงผลของการสรุปข้อมูลโดยรวมจากข้อมูลจำนวน 8 ปี ประจำปี 2551 – 2558 โดยจะมีแถบเมนูในรูปแบบเดียวกันกับการแสดงตาราง และแดชบอร์ดแบบรายปี ผู้ใช้งานทั่วไปสามารถคลิกเลือกดูข้อมูลจากตัวกรองของแดชบอร์ดได้



ภาพที่ ก.15 แสดงผลของการสรุปข้อมูลโดยรวมจากข้อมูลจำนวน 8 ปี

หมายเลข 1 ส่วนแสดงตารางสถิติข้อมูลผู้บาดเจ็บและเสียชีวิตในช่วงเทศกาลปีใหม่ประจำปี 2551 – 2558 สามารถเรียกดูข้อมูลได้

หมายเลข 2 ส่วนอธิบายข้อมูลภายในแดชบอร์ด

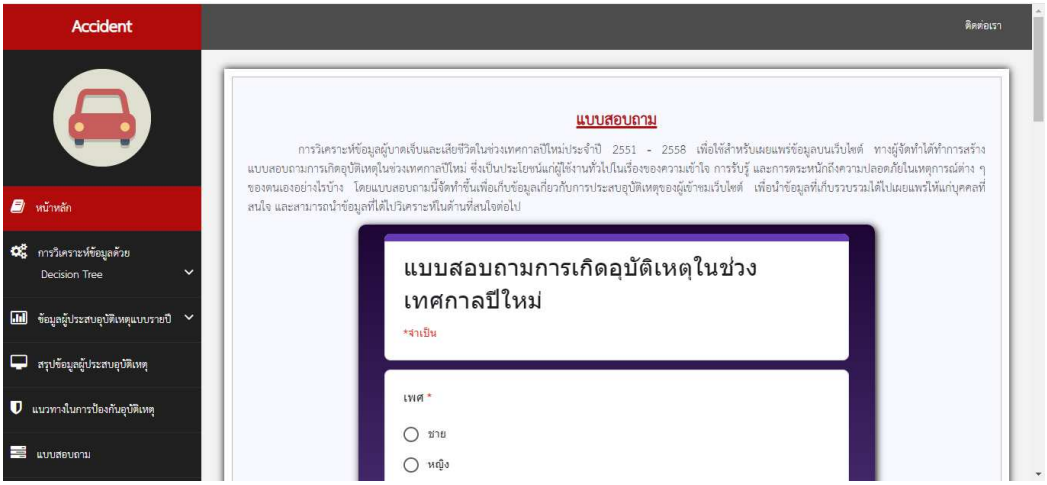
หมายเลข 3 ตัวกรองข้อมูลของแดชบอร์ดแสดงผล ผู้ใช้สามารถใช้ตัวกรองนี้กรองดูข้อมูลตามความต้องการได้

หมายเลข 4 ส่วนของแดชบอร์ดแสดงผล

15) หน้าแสดงผลแนวทางการป้องกันการเกิดอุบัติเหตุ หน้านี้จะประกอบไปด้วยเนื้อหาแนวทางการป้องกันและการปฏิบัติตนในการเดินทางในช่วงเทศกาล

ภาพที่ ก.16 แสดงแนวทางการป้องกันการเกิดอุบัติเหตุ

16) หน้าแสดงแบบสอบถามสำหรับผู้เข้าใช้งานเว็บไซต์เพื่อเก็บข้อมูลและนำข้อมูลที่ได้ไปใช้ประโยชน์ในการวิเคราะห์ข้อมูลต่อไป



ภาพที่ ก.17 แบบสอบถามสำหรับผู้เข้าใช้งานเว็บไซต์

17) เป็นหน้าจอแสดงประวัติผู้จัดทำเว็บไซต์ และข้อมูลที่ต้องการติดต่อ รวมถึงแสดงหน่วยงานที่เกี่ยวข้องกับโครงงานนี้

ภาพที่ ก.18 แสดงส่วนติดต่อของคณะผู้จัดทำ

ภาคผนวก ข

แบบฟอร์มและเอกสารที่ใช้ในโครงงาน

แบบสอบถาม

โครงงานเรื่อง : การวิเคราะห์ข้อมูลผู้บาดเจ็บและเสียชีวิตในช่วงเทศกาลปีใหม่ประจำปี 2551 – 2558 เพื่อใช้สำหรับเผยแพร่ข้อมูลบนเว็บไซต์

คำชี้แจง : แบบสอบถามประกอบด้วยชุดคำถาม 3 ส่วน คือ

ส่วนที่ 1 ข้อมูลทั่วไป

ส่วนที่ 2 การรับรู้ และความเข้าใจ

ส่วนที่ 3 ข้อเสนอแนะ/ข้อคิดเห็นอื่น ๆ

ตอนที่ 1 ข้อมูลทั่วไปของผู้ใช้งาน

ข้อชี้แจง ใส่เครื่องหมาย ✓ ลงใน ❑ หน้าคำตอบที่ตรงกับข้อมูลของผู้ใช้งานเพียงข้อเดียว

1. ระดับชั้น

❑ ชาย ❑ หญิง

2. อายุ

❑ 1.) ต่ำกว่าหรือเท่ากับ 18 ปี ❑ 2.) 18 – 25 ปี ❑ 4.) 25 ปีขึ้นไป

ตอนที่ 2 การรับรู้ และความเข้าใจ

ข้อชี้แจง ทำเครื่องหมาย ✓ ลงในช่องระดับคะแนนความคิดเห็นที่ตรงกับความคิดเห็นของผู้ใช้งานมากที่สุด โดยที่ 5 = มากที่สุด 4 = มาก 3 = ปานกลาง 2 = น้อย และ 1 = น้อยที่สุด

ตารางที่ ข.1 แสดงข้อมูลแบบสอบถาม

รายการ	ท่านเห็นด้วยในระดับใด				
	5	4	3	2	1
1. แหล่งความรู้ที่อ้างอิงมีความชัดเจน และสอดคล้องกับงาน					
2. อธิบายขั้นตอนการวิเคราะห์ข้อมูล Decision Tree และอธิบายขั้นตอนการใช้งานในโปรแกรม Waka 3.8.4 ได้ดีและเข้าใจง่าย					
3. ผลลัพธ์ที่ได้จากการสร้างโมเดลมีความน่าเชื่อถือ และโปรแกรมแบบจำลองของโมเดลมีประสิทธิภาพ					
4. การเลือกใช้โปรแกรม Tableau Public แสดงผลแบบ Visualization ได้เหมาะสมกับข้อมูล					

รายการ	ท่านเห็นด้วยในระดับใด				
	5	4	3	2	1
5.รูปแบบ visualization ที่หลากหลายของการนำเสนอข้อมูลผู้ประสบอุบัติเหตุในแต่ละปี และการนำเสนอแบบสรุปรวมทุกปี					
6. แดชบอร์ดนำเสนอข้อมูลสามารถเข้าใจได้ง่าย และเลือกใช้งานแต่ละแดชบอร์ดได้สะดวก					
7. การตอบสนองต่อการกรองดูข้อมูลผู้ประสบอุบัติเหตุในช่วงเทศกาลปีใหม่ตามความต้องการได้อย่างรวดเร็ว					
8. การนำแนวทางการป้องกันอุบัติเหตุและมาตรการในการป้องกันตัวเอง ครอบคลุมและมีประสิทธิภาพ					
9. การออกแบบเว็บไซต์ การจัดเรียงเนื้อหา การตกแต่งและการใช้งาน					
10. ประโยชน์ที่ได้จากการใช้งานเว็บไซต์					

ตอนที่ 3 ข้อเสนอแนะ

.....

.....

.....

.....

FM-CIS-10

แบบบันทึกรายละเอียดการเข้าพบอาจารย์ที่ปรึกษา

วิชา Computer Information System Project

หลักสูตรระบบสารสนเทศทางคอมพิวเตอร์ สาขาบริหารธุรกิจ

คณะบริหารธุรกิจและศิลปศาสตร์ มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา เชียงใหม่

ชื่อโครงการ การวิเคราะห์ข้อมูลผู้บาดเจ็บและเสียชีวิตในเทศกาลปีใหม่ประจำปี 2551 – 2558 เพื่อใช้สำหรับเผยแพร่ข้อมูลบนเว็บไซต์

ผู้จัดทำ 1. นางสาวมนิศรา เทพฝั้น รหัสนักศึกษา 61521207041-6

2. นางสาวศิริขวัญ ชันนาแล รหัสนักศึกษา 61521207045-7

อาจารย์ที่ปรึกษา ผู้ช่วยศาสตราจารย์ ดร. ชัชฎาพร ปุกแก้ว

ครั้งที่	วัน/เดือน/ปี	รายละเอียดการเข้าพบอาจารย์ที่ปรึกษา	ความคืบหน้าของผลงาน	ลายมืออาจารย์ที่ปรึกษา
1	20 ก.พ. 63	ส่งความคืบหน้าบทที่ 1-2	5 %	
2	4 มี.ค. 63	แก้ไขบทที่ 1-2 และส่งบทที่ 3	5 %	
3	11 มี.ค. 63	ปรึกษาวิธีการวิเคราะห์ข้อมูลในบทที่ 3	10 %	
4	18 มี.ค. 63	ส่งความคืบหน้าบทที่ 3	✓	
5	24 มี.ค. 63	ปรึกษาแนวทางการออกแบบเว็บไซต์	✓	
6	3 เม.ย. 63	ส่งความคืบหน้าการออกแบบเว็บไซต์ และแก้ไขบทที่ 3	✓	
7	10 เม.ย. 63	ส่งความคืบหน้าการนำเสนอข้อมูลด้วยโปรแกรม Tableau public	✓	
8	17 เม.ย. 63	ส่งความคืบหน้าเว็บไซต์ และแสดงผลข้อมูล นำเสนอข้อมูล ส่งบทที่ 3 ที่แก้ไข	✓	
9	24 เม.ย. 63	ส่งความคืบหน้าเว็บไซต์ และปรึกษาการอธิบายผลบทที่ 4 และแบบฟอร์ม	✓	
10	1 พ.ค. 63	ส่งความคืบหน้าบทที่ 4-5	✓	
11	7 พ.ค. 63	ตรวจเอกสารฉบับสมบูรณ์	✓	
12	8 พ.ค. 63	ตรวจเว็บไซต์ฉบับสมบูรณ์	✓	
13	15 พ.ค. 63	ส่งเอกสารที่แก้ไขเสร็จและซ้อมนำเสนอ	✓	

ลงชื่อ มนิศรา เทพฝั้น (นักศึกษา) (นางสาวมนิศรา เทพฝั้น)

ลงชื่อ นางสาวศิริขวัญ ชันนาแล (นักศึกษา) (นางสาวศิริขวัญ ชันนาแล)

ประวัติผู้จัดทำ

ชื่อ-สกุล นางสาวมนิศรา เทพฝั้น

วันเดือนปีเกิด วันอังคารที่ 14 เดือนเมษายน พ.ศ. 2541

ภูมิลำเนา 85 หมู่ที่ 8 ต.บุญนาคพัฒนา อ.เมืองลำปาง จ.ลำปาง 52000

E-Mail khae140441@gmail.com

ประวัติการศึกษา

- ระดับมัธยมศึกษาตอนต้น โรงเรียนเสด็จวนชยางค์กูลวิทยา จังหวัดลำปาง สำเร็จการศึกษาปี 2555
- ระดับประกาศนียบัตรวิชาชีพ สาขาวิชาคอมพิวเตอร์ธุรกิจ วิทยาลัยอาชีวศึกษาจังหวัดลำปาง สำเร็จการศึกษาปี 2558
- ระดับประกาศนียบัตรวิชาชีพชั้นสูง สาขาวิชาคอมพิวเตอร์ธุรกิจ วิทยาลัยอาชีวศึกษา จังหวัดลำปาง สำเร็จการศึกษาปี 2560
- ระดับปริญญาตรี สาขาวิชาระบบสารสนเทศทางธุรกิจ มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา จังหวัดเชียงใหม่ สำเร็จการศึกษาปี 2562

ชื่อ-สกุล นางสาวศิริขวัญ ขันนาแล

วันเดือนปีเกิด วันศุกร์ที่ 27 เดือนมิถุนายน พ.ศ. 2540

ภูมิลำเนา 21/1 หมู่ที่ 1 ต.กล้วยแพะ อ.เมืองลำปาง จ.ลำปาง 52000

E-Mail tk_my_love@hotmail.com

ประวัติการศึกษา

- ระดับมัธยมศึกษาตอนต้น โรงเรียนวัดหลวงวิทยา จังหวัดลำปาง สำเร็จการศึกษาปี 2555
- ระดับประกาศนียบัตรวิชาชีพ สาขาวิชาคอมพิวเตอร์ธุรกิจ วิทยาลัยอาชีวศึกษาจังหวัดลำปาง สำเร็จการศึกษาปี 2558
- ระดับประกาศนียบัตรวิชาชีพชั้นสูง สาขาวิชาคอมพิวเตอร์ธุรกิจ วิทยาลัยอาชีวศึกษา จังหวัดลำปาง สำเร็จการศึกษาปี 2560
- ระดับปริญญาตรี สาขาวิชาระบบสารสนเทศทางธุรกิจ มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา จังหวัดเชียงใหม่ สำเร็จการศึกษาปี 2563