การวิเคราะห์ข้อมูลรายงานการฆาตกรรมปี 2011–2013 เพื่อใช้สำหรับเผยแพร่ข้อมูลบนเว็บไซต์

นางสาววรรณพร นันต๊ะเสน

หลักสูตรระบบสารสนเทศทางธุรกิจ
สาขาบริหารธุรกิจ คณะบริหารธุรกิจและศิลปศาสตร์
มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา

ปีการศึกษา 2563

การวิเคราะห์ข้อมูลรายงานการฆาตกรรมปี 2011–2013 เพื่อใช้สำหรับเผยแพร่ข้อมูลบนเว็บไซต์

นางสาววรรณพร นันต๊ะเสน

หลักสูตรระบบสารสนเทศทางธุรกิจ

สาขาบริหารธุรกิจ คณะบริหารธุรกิจและศิลปศาสตร์

มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา

ปีการศึกษา 2563

ชื่อโครงงาน	การวิเคราะห์ข้อมูลรายงานการฆาตกรรมปี 2011-2013 เพื่อใช้สำหรับเผยแพร่ข้อมูลบนเว็บไซต์
โดย	นางสาววรรณพร นันต๊ะเสน รหัสนักศึกษ 61521207042-4
หลักสูตร	ระบบสารสนเทศทางธุรกิจ สาขาบริหารธุรกิจ คณะบริหารธุรกิจและศิลปศาสตร์
อาจารย์ที่ปรึกษา	ผู้ช่วยศาสตราจารย์ ดร.ชัชฎาพร ปุกแก้ว

หลักสูตรระบบสารสนเทศทางธุรกิจ สาขาบริหารธุรกิจ คณะบริหารธุรกิจและศิลปศาสตร์ มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา อนุมัติให้นับโครงงานนี้ เป็นส่วนหนึ่งของการศึกษาหลักสูตรบริหารธุรกิจบัณฑิต

(.....)
(อาจารย์สุพงศ์ แดงสุริยศรี)
หัวหน้าหลักสูตรระบบสารสนเทศทางธุรกิจ

คณะกรรมการสอบโครงงาน

.....ประธานกรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.ชัชฎาพร ปุกแก้ว)

.....กรรมการ
(ผู้ช่วยศาสตราจารย์ สัจจธรรม สุภาจันทร์)

.....กรรมการ
(ผู้ช่วยศาสตราจารย์ หฤทัย อาษากิจ)

© ลิขสิทธิ์ของหลักสูตรระบบสารสนเทศทางธุรกิจ สาขาบริหารธุรกิจ คณะบริหารธุรกิจและศิลปะศาสตร์

ชื่อโครงงาน การวิเคราะห์ข้อมูลรายงานการฆาตกรรมปี 2011–2013 เพื่อใช้สำหรับเผยแพร่ข้อมูลบนเว็บไซต์

โดย นางสาววรรณพร นันต๊ะเสน รหัส 61521207042–4

อาจารย์ที่ปรึกษา ผู้ช่วยศาสตราจารย์ ดร.ชัชฎาพร ปุกแก้ว

หลักสูตร ระบบสารสนเทศทางธุรกิจ

สาขาบริหารธุรกิจ คณะบริหารธุรกิจและศิลปศาสตร์

ปีการศึกษา 2563

บทคัดย่อ

การจัดทำโครงงานในครั้งนี้มีวัตถุประสงค์เพื่อวิเคราะห์ข้อมูลรายงานการฆาตกรรมปี 2011–2013 เพื่อใช้สำหรับเผยแพร่ข้อมูลบนเว็บไซต์ โดยผู้จัดทำได้เลือกใช้ขั้นตอนกระบวนการวิเคราะห์ข้อมูลด้วย CRISP–DM หรือ (Cross Industry Standard Process for Data Mining) มาดำเนินการตั้งแต่ การวิเคราะห์ปัญหา จัดการรวบรวมข้อมูล และแปลงข้อมูลให้อยู่ในรูปแบบที่เหมาะสมแก่การนำไปวิเคราะห์ โดยเลือกทำเหมืองข้อมูลแบบ Classification ที่สร้างโมเดล Decision Tree โดยใช้โปรแกรม RapidMiner ในการทำเหมืองข้อมูลและข้อมูลที่ได้จากการวิเคราะห์ข้อมูลและแสดงผลในรูปแบบของแผนภาพโดยใช้โปรแกรม Tableau Public ผู้ใช้งานสามารถกรองข้อมูลผ่านเว็บไซต์โดยระบบ Tableau Public ได้ ซึ่งผู้จัดทำได้เว็บไซต์สำหรับเผยแพร่ข้อมูลสารสนเทศนี้บน Web browser

โดยเว็บไซต์ที่สร้างขึ้น ได้ใช้ภาษา HTML CSS PHP และ JavaScript ในการสร้างเว็บไซต์ขึ้นมา และมีส่วนประกอบต่าง ๆ ดังนี้ 1. ส่วนเนื้อหาปัญหาอาชญากรรม 2. ส่วนการวิเคราะห์ข้อมูลด้วย Decision Tree 3. ส่วนแสดงผลข้อมูลอาชญากรรม 4. ส่วนแบบฟอร์มแบบทดสอบ 5. ส่วนแบบฟอร์มแบบสอบถาม 6. ส่วนดาวน์โหลดข้อมูล 7.ส่วนอัปโหลดเพิ่ม ลบ แก้ไขข้อมูล

จากการศึกษาวิเคราะห์และสร้างเว็บไซต์ พบว่าเว็บไซต์ได้สามารถนำมาประยุกต์ใช้ได้ตรงตามวัตถุประสงค์ที่ตั้งไว้ สรุปได้ว่าในอนาคตสามารถนำข้อมูลจากเว็บไซต์ไปปรับปรุงเพิ่มเติมต่อได้ เพื่อให้เกิดประโยชน์มากขึ้นต่อไป

กิตติกรรมประกาศ

โครงงานครั้งนี้ สำเร็จลุล่วงไปได้ด้วยดี เนื่องจากได้รับความกรุณาอย่างดียิ่งจาก ผู้ช่วยศาสตราจารย์ ดร.ชัชฎาพร ปุกแก้ว อาจารย์ที่ปรึกษาโครงงาน อาจารย์ สุพงศ์ แดงสุริยศรี หัวหน้า สาขาวิชาระบบสารสนเทศทางธุรกิจ และคณะอาจารย์ในสาขาระบบสารสนเทศทางธุรกิจ คณะบริหารธุรกิจและศิลปะศาสตร์ มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา ทุกท่าน ที่ได้ให้คำปรึกษาและคำแนะนำในการจัดทำโครงงานเพื่อบรรลุวัตถุประสงค์ของ โครงงาน ให้เสร็จสมบูรณ์ รวมถึงช่วยแก้ไขเอกสารโครงงาน ให้มีความสมบูรณ์มากยิ่งขึ้น

ขอขอบพระคุณ หน่วยงานเจ้าของเว็บไซต์ kaggle.com ที่ให้ดาวน์โหลดชุดข้อมูล ซึ่งเป็นเว็บไซต์ที่เก็บรวบรวมชุดข้อมูลต่าง ๆ ทั้งในประเทศไทยและต่างประเทศ เพื่อที่ให้นำมาวิเคราะห์ข้อมูลตามกระบวนการและขั้นตอนต่าง ๆ

ท้ายสุดนี้ ผู้จัดทำโครงงานขอขอบพระคุณผู้ที่มีส่วนเกี่ยวข้องและทุกท่านที่ช่วยส่งเสริม สนับสนุนให้ผู้จัดทำโครงงานเกิดแรงผลักดัน และให้ความช่วยเหลือต่าง ๆ แก่ผู้จัดทำโครงงาน ในครั้งนี้ ขอกราบขอบพระคุณเป็นอย่างสูงไว้ ณ ที่นี้

วรรณพร นันต๊ะเสน

สารบัญ

	หน้า
บทคัดย่อ	ก
กิตติกรรมประกาศ	ข
สารบัญ	ค
สารบัญภาพ	จ
สารบัญตาราง	ซ
บทที่ 1 บทนำ	1
1.1 ความสำคัญและความเป็นมา	1
1.2 วัตถุประสงค์	2
1.3 ประโยชน์ที่จะได้รับ	2
1.4 ขอบเขต	3
1.5 เครื่องมือที่ใช้ในการพัฒนาโปรแกรม	4
1.6 สถานที่ใช้ในการดำเนินการศึกษาและรวบรวมข้อมูล	4
1.7 ระยะเวลาในการดำเนินการ	5
1.8 นิยามศัพท์เฉพาะ	5
1.9 บทสรุป	6
บทที่ 2 แนวคิด ทฤษฎี เครื่องมือและวรรณกรรมที่เกี่ยวข้อง	7
2.1 แนวคิด	7
2.2 ทฤษฎี	12
2.3 เครื่องมือในการออกแบบและวิเคราะห์ข้อมูล	25
2.4 วรรณกรรมที่เกี่ยวข้อง	27
2.5 บทสรุป	33
บทที่ 3 วิธีการดำเนินงานโครงงาน	34
3.1 การวิเคราะห์ข้อมูลด้วย CRISP-DM	34
3.2 การออกแบบเว็บไซต์	80
3.3 บทสรุป	85

สารบัญ(ต่อ)

	หน้า
บทที่ 4 ผลการดำเนินงาน	87
4.1 ผลการดำเนินงาน	87
4.2 การอภิปรายผลการดำเนินโครงงาน	95
4.3 บทสรุป	99
บทที่ 5 สรุปและข้อเสนอแนะ	100
5.1 บทสรุปผลโครงงาน	100
5.2 ข้อจำกัดของเว็บไซต์	101
5.3 ปัญหาและอุปสรรคของโครงงาน	101
5.4 ข้อเสนอแนะ	102
บรรณานุกรรม	103
ภาคผนวก ก คู่มือการใช้งานเว็บไซต์	107
ภาคผนวก ข แบบฟอร์มและเอกสารที่ใช้ในโครงงาน	115
ประวัติผู้จัดทำ	120

สารบัญภาพ

ภาพที่		หน้า
2.1	เครื่องมือที่ใช้ทำ Data Visualization	9
2.2	การนำเสนอแบบทิศทางหรือแนวโน้ม (Trending)	10
2.3	การนำเสนอแบบกลุ่มข้อมูล (Classification)สารบัญ	10
2.4	การนำเสนอเชิงเปรียบเทียบข้อมูล (Comparison)	11
2.5	การนำเสนอรูปแบบแผนที่ (Geographical)	11
2.6	กลุ่มที่ต้องพยากรณ์ล่วงหน้า	12
2.7	โปรแกรม Tableau Public	26
3.1	เว็บไซต์ kaggle.com	35
3.2	ข้อมมูลรายงานการฆาตกรรมทั้งหมด	35
3.3	ข้อมูลรายงานการฆาตกรรม	36
3.4	การลบรายการย่อย Agency Type(ประเภทตัวแทน)	37
3.5	การลบรายการย่อย Year(ปี)	37
3.6	การลบรายการย่อย Weapon(อาวุธ)	38
3.7	ชุดข้อมูลที่คัดเลือกมาวิเคราะห์ข้อมูล	38
3.8	รูปแบบโมเดล Graph Decision Tree ในโปรแกรม RapidMiner	74
3.9	รูปแบบโมเดล Graph Decision Tree ในการคำนวณมือ	75
3.10	คำบรรยายลักษณะงาน Decision Tree ของ RapidMiner	75
3.11	แสดงข้อมูลกราฟของประเภทหน่วยงาน	79
3.12	แสดงข้อมูลของการฆาตกรรม	79
3.13	แสดงข้อมูลกราฟของการใช้อาวุธ	80
3.14	หน้าแรก	80
3.15	หน้าข้อมูลการวิเคราะห์โมเดล Decision Tree	81
3.16	หน้าข้อมูลอาชญากรรม ของ ข้อมูลหน่วยงานตำรวจ	81
3.17	หน้าข้อมูลอาชญากรรม ของ ข้อมูลการฆาตกรรม	82
3.18	หน้าข้อมูลอาชญากรรม ของ ข้อมูลการใช้อาวุธ	82
3.19	หน้าแบบทดสอบ	83

สารบัญภาพ(ต่อ)

ภาพที่		หน้า
3.20	หน้าแบบสอบถาม การป้องกันตนเอง	83
3.21	หน้าแบบสอบถาม ความพึงพอใจ	84
3.22	หน้าดาวน์โหลดไฟล์ข้อมูล	84
3.23	หน้าเข้าสู่ระบบ	85
3.24	หน้าจัดการไฟล์ เพิ่ม ลบ แก้ไข ข้อมูล	85
4.1	หน้าแรกของเว็บไซต์	88
4.2	หน้าเมนูการวิเคราะห์โมเดล	88
4.3	หน้าข้อมูลอาชญากรรม ข้อมูลหน่วยงานตำรวจ	89
4.4	หน้าข้อมูลอาชญากรรม ข้อมูลหน่วยงานตำรวจ และแสดง Dashboard	89
4.5	หน้าข้อมูลอาชญากรรม ข้อมูลการฆาตกรรม	90
4.6	หน้าข้อมูลอาชญากรรม ข้อมูลการฆาตกรรม และแสดง Dashboard	90
4.7	หน้าข้อมูลอาชญากรรม ข้อมูลการใช้อาวุธ	91
4.8	หน้าข้อมูลอาชญากรรม ข้อมูลการใช้อาวุธ และแสดง Dashboard	91
4.9	หน้าแบบทดสอบ	92
4.10	หน้าแบบสอบถาม การป้องกันตนเอง	92
4.11	หน้าแบบสอบถาม ความพึงพอใจ	93
4.12	หน้าดาวน์โหลดข้อมูล	93
4.13	หน้า Login เข้าสู่ระบบ Admin	94
4.14	หน้าอัปโหลดข้อมูล	94
ก.1	หน้าแรกของเว็บไซต์	108
ก.2	การวิเคราะห์ข้อมูลด้วยโมเดล Decision Tree	109
ก.3	ส่วนแสดงผลข้อมูลอาชญากรรม	110
ก.4	ส่วนแบบฟอร์มแบบทดสอบ	111
ก.5	ส่วนแบบฟอร์มแบบสอบถาม	112
ก.6	ส่วนดาวน์โหลดข้อมูล	113
ก.7	เข้าสู่ระบบสำหรับแอดมิน	113

สารบัญภาพ(ต่อ)

ภาพที่ **หน้า**

ก.8 ส่วนอัปโหลดข้อมูล 114

สารบัญตาราง

ตารางที่	หน้า
1.1 ระยะเวลาในการดำเนินงาน	5
4.1 แสดงข้อมูลการประเมินผลด้านเนื้อหา	97
4.2 แสดงข้อมูลการประเมินผลด้านการออกแบบเว็บไซต์	97
4.3 แสดงข้อมูลการประเมินผลด้านประโยชน์และการนำไปใช้	98
ข.1 ข้อมูลแบบสอบถาม	117

บทที่ 1
บทนำ

1.1 ความสำคัญและที่มาของปัญหา

ปัจจุบันปัญหาอาชญากรรมในสังคม จากข้อมูลคดีอาญา ทางข่าวสารหนังสือพิมพ์ โทรทัศน์หรือ สื่อสารทางอิเล็กทรอนิกส์ ต่าง ๆ ทั่วราชอาณาจักรมีแนวโน้มที่จะรุนแรงขึ้น ทั้งในด้านสถิติตัวเลข การก่ออาชญากรรม และวิธีการประกอบอาชญากรรมของผู้กระทำความผิด ที่สามารถมองเห็นได้ เช่น คดีลักทรัพย์ ชิงทรัพย์ ปล้นทรัพย์ ข่มขืนกระทำชำเรา คดียาเสพติด เป็นต้น และที่มองไม่เห็น เช่น คดีเกี่ยวกับอาชญากรรมคอมพิวเตอร์ แก๊งคอลเซ็นเตอร์ การหลอกลวงขายสินค้าทางอินเตอร์เน็ต เป็นต้น จากการนำเสนอภาคข่าวเกี่ยวกับการกระทำผิดผ่านสื่อต่าง ๆ ทั้งจากโทรทัศน์ วิทยุ หนังสือพิมพ์ และอินเทอร์เน็ต มีจำนวนผู้กระทำความผิด และผู้ที่เป็นเหยื่อของอาชญากรรมเป็นจำนวนมาก ซึ่งข้อมูลเหล่านี้ได้ถูกจัดเก็บขึ้นมาผ่านหน่วยงานเอฟบีไอในสหรัฐอเมริกาที่มีอยู่ในปัจจุบัน ผ่านเว็บไซต์ kaggle.com ข้อมูลเหล่านี้ถูกจัดเก็บมาตั้งแต่ปี 1980–2014 เพื่อเป็นข้อมูลที่บอกสถิติในการเกิดอาชญากรรม ที่เกิดขึ้นในเหตุการณ์ต่าง ๆ ใบบางพื้นที่ เช่น แหล่งรวมศูนย์กลางของแต่ละพื้นที่เป็นพื้นฐานที่มีความหนาแน่นของประชากรมาก มีบทบาทต่อความสำคัญในการดำเนินชีวิตของประชาชน เนื่องจากเป็นศูนย์กลางของกิจกรรมต่าง ๆ ทุกด้าน และยังก่อให้เกิดการลงทุนต่าง ๆ มากมาย เป็นแหล่งที่มีผู้คนจับตามองเห็นกันและกันและอาจโดนเพ่งเล็งขึ้นได้จึงอาจทำให้เกิดอาชญากรรมขึ้นโดยอาจไม่รู้ตัว

อาชญากรรมมีแต่ความสูญเสียและก่อให้เกิดผลเสียทั้งด้านเศรษฐกิจ สังคม ความปลอดภัยในชีวิตและทรัพย์สินของประชาชนตลอดจนความมั่นคงของชาติ ความสูญเสียจากอาชญากรรม มีมูลค่าต่อทรัพย์สินและจิตใจ ในปัจจุบันอาชญากรรมมีแนวโน้มเพิ่มสูงขึ้นตามการขยายตัวของประชากรและการขยายตัวของเมือง เท่านั้นไม่พอ รูปแบบอาชญากรรมยังก้าวหน้าไปตามความเจริญของเทคโนโลยี ส่งผลให้การปราบปรามเป็นไปด้วยความยากลำบากกว่าแต่ก่อน เช่น อาชญากรรมข้ามชาติเกี่ยวกับการแอบอ้างใช้ข้อมูลส่วนตัวหรือข้อมูลทางการเงินหรือแอบอ้างใช้เอกสารโดยสำนักงานกิจการยุติธรรม กระทรวงยุติธรรม ได้กล่าวถึงอาชญากรรมพื้นฐานที่มีผลต่อการพัฒนาเศรษฐกิจและสังคมของชาติในปัจจุบันประกอบด้วย อาชญากรรมต่อทรัพย์สิน เช่น ลักทรัพย์ ปล้นทรัพย์ ทำให้เสียทรัพย์ บุกรุก

อาชญากรรมต่อชีวิตและร่างกาย เช่น ทำร้าย พยายามทำร้าย ข่มขู่ อาชญากรรมทางเพศ เช่น การข่มขืน การกระทำอนาจาร ลวนลามทางเพศ และอาชญากรรมเกี่ยวกับการแอบอ้างใช้ข้อมูลส่วนตัวหรือข้อมูลทางการเงินหรือแอบอ้าใช้เอกสาร จึงอาจกล่าวได้ว่าปัญหาอาชญากรรมเป็นปัญหาหนึ่งที่สำคัญยิ่งต่อสภาพสังคมไทยที่ส่งผลกระทบต่อชีวิตและทรัพย์สินของประชาชน เช่น คดีฆ่าผู้อื่น ทำร้ายร่างกาย ข่มขืนกระทำชำเรา ลักทรัพย์ วิ่งราวทรัพย์ ชิงทรัพย์ และปล้นทรัพย์ เป็นต้น นับได้ว่าเป็นปัญหาที่ร้ายแรงยิ่งต่อสังคมไทย เพราะส่งผลกระทบต่อชีวิตและความเป็นอยู่ที่ดีของประชาชน เป็นปัญหาที่สร้างความไม่สงบให้แก่สังคม สร้างความเดือดร้อนต่อชีวิตและทรัพย์สิน ซึ่งการวิเคราะห์ข้อมูลรายงานการฆาตกรรม เป็นการนำเอาชุดข้อมูลจากเว็บไซต์ kaggle.com มาวิเคราะห์ในเชิงลึก เพื่อให้ทราบสาเหตุที่เกิดอาชญากรรมขึ้น ว่ามีสาเหตุหรือแรงจูงใจอย่างไร หรือเกิดจากการเพ่งเล็งเหยื่อจากรูปพรรณสัณฐาน และในการก่อเหตุใช้อาวุธอะไรในการก่อเหตุของแต่ละคน

จากปัญหาดังกล่าวผู้วิเคราะห์ข้อมูลได้นำชุดข้อมูลดังกล่าวมาประมวลผลข้อมูล แยกแยะและพิจารณาข้อมูลผ่านโปรแกรมให้ตรงตามความถูกต้องการ โดยผู้วิเคราะห์ข้อมูลได้นำชุดข้อมูลมาทำกระบวนการวิเคราะข้อมูล พิจารณาการแบ่งแยกข้อมูลในการจำแนกข้อมูลด้วยการทำ Decision tree เป็นโมเดลที่ช่วยในการตัดสินใจ ซึ่งข้อมูลในแต่ละโหนดจะแบ่งแยกออกเป็นแต่ละประเภท เช่น เพสคนร้าย เพศเหยื่อ อาวุธ เป็นต้น ทำให้ชุดข้อมูลที่ได้มาในการทำ Decision tree มีความเข้าใจในข้อมูลมากขึ้น และนำชุดข้อมูลนี้ทำออกมาเป็นกราฟรูปภาพแสดงข้อมูลออกมาผ่านโปรแกรม Tableau Public และเผยแพร่ข้อมูลสารสานเทศบน Web Browser เพื่อให้ผู้ใช้งานได้ทราบถึงข้อมูลรายงานการฆาตกรรมที่เกิดขึ้น

1.2 วัตถุประสงค์

1.2.1 เพื่อวิเคราะห์ข้อมูลรายงานการฆาตกรรมปี 2011–2013

1.2.2 เพื่อเผยแพร่ข้อมูลรายงานการฆาตกรรมปี 2011–2013 บนเว็บไซต์

1.3 ประโยชน์ที่จะได้รับ

1.3.1 ได้ข้อมูลสารสนเทศของรายงานการฆาตกรรมปี 2011–2013

1.3.2 ได้เผยแพร่ข้อมูลรายงานการฆาตกรรมปี 2011–2013 บนเว็บไซต์

1.4 ขอบเขต

1.4.1 ขอบเขตผู้วิเคราะห์ข้อมูล

1.4.1.1 Business Understanding ผู้วิเคราะห์ข้อมูลทำความเข้าใจกับปัญหาให้อยู่ในรูปของการวิเคราะห์ข้อมูลทางดาต้าไมน์นิ่งโดยการวิเคราะห์ข้อมูลรายงานการฆาตกรรมปี 2011–2013 ซึ่งมีข้อมูลที่ซับซ้อนแต่สามารถทำความเข้าใจในข้อมูลได้

1.4.1.2 Data Understanding ผู้วิเคราะห์ข้อมูลทำการรวบรวมข้อมูล จากนั้นทำการตรวจสอบรายละเอียดของชุดข้อมูลที่ใช้ข้อมูลทั้งหมดหรือจำเป็นต้องเลือกข้อมูลบางส่วนมาใช้ในการวิเคราะห์รายงานการฆาตกรรมปี 2011–2013 จากเว็ปไซต์ kaggle.com

1.4.1.3 Data Preparation ผู้วิเคราะห์ข้อมูลทำการแปลงข้อมูลที่ได้ทำการเก็บรวบรวมมา ให้กลายเป็นข้อมูลที่สามารถนำไปวิเคราะห์ในขั้นถัดไปได้ โดยการทำ Data Cleaning ข้อมูล การทำกระบวนการนี้อาจจะต้องมีการทำข้อมูลให้ถูกต้อง เช่น การเติมข้อมูลที่ขาดหายไปให้ครบถ้วนให้ข้อมูลมีความสมบูรณ์ การตัดข้อมูลที่ไม่จำเป็นออกไป เป็นต้น

1.4.1.4 Modeling ผู้วิเคราะห์ข้อมูลวิเคราะห์ข้อมูลด้วยเทคนิคทางดาต้าไมน์นิงแบบการจำแนกประเภทข้อมูล (Classification) โดยการใช้โมเดลการตัดสินใจแบบต้นไม้ (Decision Trees) ซึ่งในขั้นตอนนี้จะถูกนำมาใช้เพื่อให้ได้คำตอบที่ดีที่สุด

1.4.1.5 Evaluation ผู้วิเคราะห์ข้อมูลได้ข้อมูลที่ผ่านการวิเคราะห์ข้อมูลด้วยเทคนิค ดาต้าไมน์นิง จากการสร้างโมเดล Decision Tree ซึ่งต้องมีการวัดประสิทธิภาพของผลลัพธ์ที่ได้ว่าตรงกับวัตถุประสงค์ที่ได้ตั้งไว้

1.4.1.6 Deployment ผู้วิเคราะห์ข้อมูลนำผลข้อมูลที่ทำการวิเคราะห์มาแสดงผลข้อมูลบนหน้า Web Browser โดยใช้โปรแกรม Adobe Dreamweaver ร่วมกับการนำเสนอข้อมูลแบบ visualization ด้วยการแสดงผลข้อมูลในรูปแบบของภาพโดยใช้โปรแกรม Tableau Public

1.4.2 ขอบเขตผู้ใช้งานทั่วไปบนเว็บบราวเซอร์

1.4.2.1 สามารถดูข้อมูลสารสนเทศของรายงานการฆาตกรรมปี 2011–2013 ที่เผยแพร่ผ่านทางเว็บไซต์

1.4.2.2 สามารถดูข้อมูลสารสนเทศในการสรุปข้อมูลการฆาตกรรมได้

1.4.2.3 สามารถดูข้อมูลสารสนเทศในรูปแบบของแผนภูมิ จากโปรแกรม Tableau ได้

1.4.2.4 สามารถกรองข้อมูลสารสนเทศผ่านระบบของ Tableau บน Web browser ได้

1.4.2.5 สามารถอัพโหลดข้อมูลเพิ่มเข้ามาในระบบได้

1.5 เครื่องมือที่ใช้ในการพัฒนาโปรแกรม

1.5.1 Hardware

1.5.1.1 Notebook: Acer aspire Intel Celeron

1.5.2 Software

1.5.2.1 โปรแกรม Tableau Public ใช้ในการแสดงผลข้อมูลในรูปแบบของภาพ

1.5.2.2 โปรแกรม RapidMiner 9.6 ใช้ในการสร้างโมเดล Decision Tree

1.5.2.3 โปรแกรม Adobe Dreamweaver CS6 ใช้ในการสร้างเว็บไซต์

1.5.2.4 โปรแกรม Adobe XD ใช้ในการออกแบบหน้าเว็บไซต์

1.5.2.5 Microsoft office Word 2019 ใช้ในการทำเอกสาร

1.5.2.6 Microsoft office Excel 2019 ใช้ในการจัดการชุดข้อมูล

1.5.2.7 ระบบปฏิบัติการ Window 10

1.6 สถานที่ใช้ในการดำเนินการศึกษาและรวบรวมข้อมูล

- สถานที่ใช้ในการดำเนินการศึกษา

มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนาภาคพายัพเชียงใหม่ 128 ถนนห้วยแก้ว ตำบลช้างเผือก อำเภอเมือง จังหวัดเชียงใหม่ 50300

- แหล่งรวบรวมข้อมูล

เว็บไซต์ kaggle.com

1.7 ระยะเวลาในการดำเนินการ

ตารางที่ 1.1 ระยะเวลาในการดำเนินงาน

แผนการดำเนินการ	2563						
	เม.ย.	**พ.ค.**	**มิ.ย.**	**ก.ค.**	**ส.ค.**	**ก.ย.**	**ต.ค.**
1. การวางแผนในการดำเนินการ	←	→					
2. การเก็บรวบรวมข้อมูล		←	→				
3. การแปลงข้อมูลที่ได้ทำการเก็บรวมรวมไว้แล้ว			←	→			
4. การวิเคราะห์ข้อมูล			←	→			
5. การวัดประสิทธิภาพ					←	→	
6. การนำเสนอข้อมูล						←	→
7. จัดทำเอกสารประกอบโครงงาน	←	→	←	→	←	→	→

1.8 นิยามศัพท์เฉพาะ

1.8.1 อาชญากรรม คือ การกระทำผิด โดยทำให้เกิดความเดือดร้อน เสียหาย ต่อทรัพย์สิน หรือ บุคคล เป็นการเรียกการกระทำทางคดีอาญาแบบทั่วไป ตัวอย่าง อาชญากรรม เช่น การฆ่าคนตาย ปล้นทรัพย์ ข่มขืน ซึ่งกฎหมายโดยทั่วไปของทุกสังคมมักกำหนดบทลงโทษของผู้ก่ออาชญากรรม (เรียกว่า อาชญากร)

- อาชญากรรมในความหมายอย่างกว้าง หมายถึง พฤติกรรมที่มีการกระทำผิด โดยผู้กระทำผิดมีเจตนาในการกระทำ โดยเป็นการกระทำความผิดที่มีลักษณะร้ายแรง มีความรุนแรงและเป็นอันตรายต่อสังคม ซึ่งก่อให้เกิดผลกระทบจำนวนมหาศาลต่อสังคม อันเป็นการกระทำที่มีการละเมิดต่อกฎหมายบ้านเมือง ผู้กระทำผิดจะต้องได้รับโทษทั้งที่ไม่เป็นทางการจากสมาชิกในสังคม อาทิ การตำหนิ ติเตียน การไม่คบหาสมาคมด้วย และการได้รับโทษที่เป็นทางการจากข้อกำหนดของกฎหมายบ้านเมือง โดยผู้กระทำผิดจะต้องถูกลงโทษโดยผ่านกระบวนการยุติธรรมเป็นสำคัญ

- อาชญากรรมในความหมายอย่างแคบ คือ พฤติกรรมที่เป็นการละเมิดต่อกฎหมายอาญาเท่านั้น โดยการพิจารณาพฤติกรรมการกระทำของบุคคลในสังคมตามข้อกำหนดของกฎหมายอาญาเท่านั้น ไม่ได้คำนึงถึงเจตนา หรือลักษณะของความผิดแต่อย่างใด

1.8.2 ฆาตกรรม หมายถึง เป็นการกระทำให้มนุษย์ถึงแก่ความตาย จัดเป็นอาชญากรรมประเภทหนึ่ง ทางนิติศาสตร์แบ่งเป็นสองประเภท คือ การทำให้คนตายโดยเจตนา (homicide) และการทำให้คนตายโดยไม่เจตนา (manslaughter) การฆ่าคนทั้งสองประเภท ผู้กระทำต้องระวางโทษหนักหรือเบาตามกฎหมายแล้วแต่กรณี

1.8.3 อาวุธ คือ เครื่องมือที่ใช้ระหว่างการรบเพื่อสังหารหรือทำลายทรัพย์สินและทรัพยากร จนไม่สามารถใช้ประโยชน์จากทรัพยากรได้ อาจใช้ในการโจมตี การป้องกัน เช่น การข่มขู่ ถูกแบ่งไว้ 2ประเภทคือ (1) อาวุธที่เป็นอาวุธโดยลักษณะเช่น ดาบ หอก กระบี่ สนับมือ ปืน เครื่องระเบิด เป็นต้น (2) อาวุธที่มิใช่อาวุธโดยลักษณะแต่ใช้ได้เสมือนดัง อาวุธ มีบันทึกการใช้อาวุธปรากฏในภาพเขียนบนผนังถ้ำ การศึกษากระบวนการในการต่อสู้ ทำให้เกิดทั้งศิลปะการต่อสู้และหลักยุทธศาสตร์

1.9 บทสรุป

จากบทนำที่ได้กล่าวมาในข้างต้นทั้งหมดนั้น ผู้วิเคราะห์ข้อมูลได้เล็งเห็นความสำคัญของการนำเทคโนโลยีมาประยุกต์ใช้ในการวิเคราะห์ข้อมูลรายงานการฆาตกรรมปี 2011–2013 ด้วยกระบวนการในการในการจัดทำเหมืองข้อมูลเพื่อการวิเคราะห์ชุดข้อมูลแบบ Classification ในรูปแบบ Decision Tree และเผยแพร่ข้อมูลสารสนเทศบน Web Browser จากข้อมูลที่ได้มาจาก kaggle.com เนื่องจากเทคโนโลยีในปัจจุบันมีการใช้งานที่ง่ายและมีความสะดวกสบาย อีกทั้งผู้ใช้งานยังเข้าถึงข้อมูลได้ง่าย และมีประสิทธิภาพ

บทที่ 2
แนวคิด ทฤษฎี เครื่องมือและวรรณกรรมที่เกี่ยวข้อง

ในบทนี้เป็นการนำเสนอเกี่ยวกับ แนวคิด ทฤษฎี เครื่องมือและวรรณกรรมที่เกี่ยวข้องของการวิเคราะห์ข้อมูลรายงานการฆาตกรรมปี 2011–2013 เพื่อใช้สำหรับเผยแพร่ข้อมูลบนเว็บไซต์ และการแสดงผลข้อมูลบนเว็บไซต์ ซึ่งได้รวบรวมการศึกษาเอกสารงานวิจัยที่เกี่ยวข้องกับการวิเคราะห์ข้อมูล ดังนี้

2.1 แนวคิด

2.2 ทฤษฎี

2.3 เครื่องมือในการออกแบบและวิเคราะห์ข้อมูล

2.4 วรรณกรรมที่เกี่ยวข้อง

2.5 บทสรุป

2.1 แนวคิด

2.1.1 แนวคิดเกี่ยวกับการวิเคราะห์ข้อมูล (Data analytic)

ในการดำเนินงานเรื่องการวิเคราะห์ข้อมูลรายงานการฆาตกรรมปี 2011–2013 และการแสดงผลข้อมูลบนเว็บไซต์ ผู้วิเคราะห์ข้อมูลได้ศึกษาหลักการและทฤษฎีต่าง ๆ องค์ประกอบหนึ่งที่สำคัญคือการวิเคราะห์ข้อมูล ซึ่งมีรายละเอียด ดังนี้

2.1.1.1 การวิเคราะห์ข้อมูล เป็นการจัดการข้อมูลด้วยวิธีต่าง ๆ เช่น การคำนวณ การนำเสนอข้อมูล เป็นต้น เพื่อให้ได้ผลลัพธ์ตามวัตถุประสงค์ การวิเคราะห์ เป็นการแยกแยะสิ่งที่จะพิจารณาออกเป็นส่วนย่อยที่มีความสัมพันธ์กัน เพื่อทำความเข้าใจแต่ละส่วนให้แจ่มแจ้ง รวมทั้งการสืบค้นความสัมพันธ์ของส่วนต่าง ๆ เพื่อดูว่าส่วนประกอบปลีกย่อยนั้นสามารถเข้ากันได้หรือไม่ สัมพันธ์เกี่ยวเนื่องกันอย่างไร ซึ่งจะช่วยให้เกิดความเข้าใจต่อสิ่งหนึ่งสิ่งใดอย่างแท้จริง โดยพื้นฐานแล้ว การวิเคราะห์ข้อมูล (Data Analysis) จำแนกได้ 2 แบบ ดังนี้

- การวิเคราะห์ข้อมูลเบื้องต้น ใช้สถิติเชิงพรรนา (Descriptive Statistics) คือสถิติที่ใช้เพื่ออธิบาย บรรยาย หรือสรุป ลักษณะของกลุ่มข้อมูลที่เป็นตัวเลข ที่เก็บรวบรวม

มาซึ่งไม่สามารถอ้างอิงลักษณะประชากรได้ (ยกเว้นมีการเก็บข้อมูลของประชากรทั้งหมด) ตัวอย่างสถิติเชิงพรรณนา เช่น การแจกแจงความถี่ การวัดค่ากลางของข้อมูล การวัดการกระจายของข้อมูล

- การวิเคราะห์ข้อมูลขั้นสูง ใช้สถิติเชิงอนุมาน (Inference Statistics) คือ สถิติที่ใช้ในการวิเคราะห์ข้อมูลกลุ่มตัวอย่างที่เป็นตัวแทนของประชากรซึ่งสามารถนำผลการวิเคราะห์นั้นไปสรุปอ้างถึงประชากรได้โดยใช้ทฤษฎีความน่าจะเป็น สถิติเชิงอนุมาน ประกอบด้วย การประมาณค่า และการทดสอบสมมติฐาน

2.1.2 แนวคิดเกี่ยวกับเหมืองข้อมูล (Data Mining)

กระบวนการที่กระทำกับข้อมูลจำนวนมากเพื่อค้นหารูปแบบและความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น ในปัจจุบันการทำเหมืองข้อมูลได้ถูกนำไปประยุกต์ใช้ในงานหลายประเภท ทั้งในด้านธุรกิจที่ช่วยในการตัดสินใจของผู้บริหาร ในด้านวิทยาศาสตร์และการแพทย์รวมทั้งในด้านเศรษฐกิจและสังคม

การทำเหมืองข้อมูล (Data Mining) เปรียบเสมือนวิวัฒนาการหนึ่งในการจัดเก็บและตีความหมายข้อมูล จากเดิมที่มีการจัดเก็บข้อมูลอย่างง่าย ๆ มาสู่การจัดเก็บในรูปแบบฐานข้อมูลที่สามารถดึงข้อมูลสารสนเทศมาใช้จนถึงการทำเหมืองข้อมูลที่สามารถค้นพบความรู้ที่ซ่อนอยู่ในข้อมูล หรือจะแยก ได้ดังนี้

- กระบวนการหรือการเรียงลำดับของการค้นข้อมูลจำนวนมากและเก็บข้อมูลที่เกี่ยวข้อง
- การนำมาใช้โดยหน่วยงานทางธุรกิจและนักวิเคราะห์ทางการเงินหรือการนำมาใช้งานในด้านวิทยาศาสตร์เพื่อเอาข้อมูลขนาดใหญ่ที่สร้างโดยวิธีการทดลองและการสังเกตการณ์ที่ทันสมัย
- การสกัดหรือแยกข้อมูลที่เป็นประโยชน์จากข้อมูลขนาดใหญ่หรือฐานข้อมูล
- การวางแผนทรัพยากรขององค์กรโดยสามารถวิเคราะห์ทางสถิติและตรรกะของข้อมูลขนาดใหญ่เป็นการมองหารูปแบบที่สามารถช่วยการตัดสินใจได้

ซึ่งเหตุผลที่ต้องมีการทพเหมืองข้อมูลเพราะข้อมูลที่ถูกเก็บไว้ในฐานข้อมูลหากเก็บไว้เฉย ๆ ก็จะไม่เกิดประโยชน์ดังนั้นจึงต้องมีการสกัดสารสนเทศหรือการคัดเลือกข้อมูลออกมาใช้งานส่วนที่เราต้องการในอดีตได้ใช้คนเป็นผู้สืบค้นข้อมูลต่าง ๆ ในฐานข้อมูลซึ่งผู้สืบค้นจะทำการสร้างเงื่อนไขขึ้นมาตามภูมิปัญญาของผู้สืบค้น ในปัจจุบันการวิเคราะห์ข้อมูลจากฐานข้อมูล

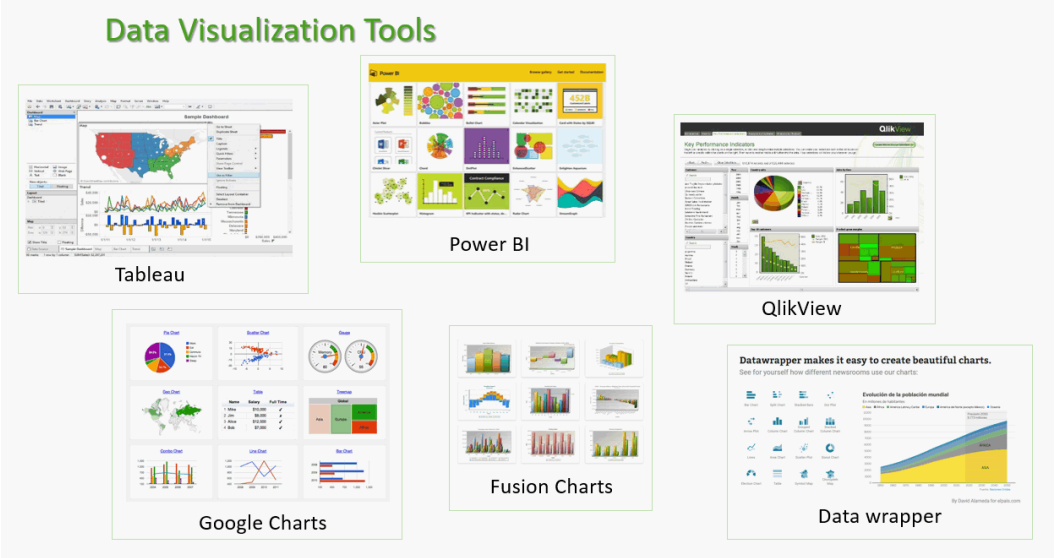
เดียวอาจไม่ให้ความรู้เพียงพอและลึกซึ้งสำหรับการดำเนินงานภายใต้ภาวะที่มีการแข่งขันสูง และมีการเปลี่ยนแปลงที่รวดเร็วจึงจำเป็นที่จะต้องรวบรวมฐานข้อมูลหลาย ๆ ฐานข้อมูลเข้าด้วยกัน เรียกว่า “ คลังข้อมูล” (Data Warehouse) ดังนั้น จึงจำเป็นต้องใช้ Data Mining ในการดึงข้อมูลจากฐานข้อมูลที่มีขนาดใหญ่ เพื่อที่จะนำข้อมูลนั่นมาใช้งานให้เกิดประโยชน์สูงที่สุด

2.1.3 แนวคิดเกี่ยวกับการแสดงผลข้อมูล (Data visualization)

Data Visualization หรือ Information Visualization คือ การถ่ายทอดข้อมูลในเชิงปริมาณที่มีความซับซ้อนได้อย่างมีประสิทธิผล ซึ่งคำว่า “ประสิทธิผล” ในที่นี้หมายถึงมีความชัดเจน (Clarity), มีความแม่นยำ (Precision), และมีประสิทธิภาพ (Efficiency) หากไม่มีการทำ Data Visualization แล้ว อาจทำให้เราไม่สามารถค้นพบนัยยะของข้อมูลในแง่ของแนวโน้ม, รูปแบบพฤติกรรม, และความสัมพันธ์เชื่อมโยงได้

Data Visualization เป็นการนำข้อมูลมาผสมผสานกับจินตนาการ เพื่อสร้างภาพในความคิดขึ้นมา ซึ่งมีกระบวนการนำเสนอข้อมูลที่มีความซับซ้อนหรือข้อมูลเชิงปริมาณ ให้สามารถเข้าใจได้ง่าย ในแบบของ กราฟ แผนภูมิ

2.1.3.1 เครื่องมือที่นิยมใช้ในตอนนี้ได้แก่ Tableau, Microsoft Power BI, Qlik View, Google Charts, Fusion Charts,Data wrapper และอื่นๆอีกมากมาย



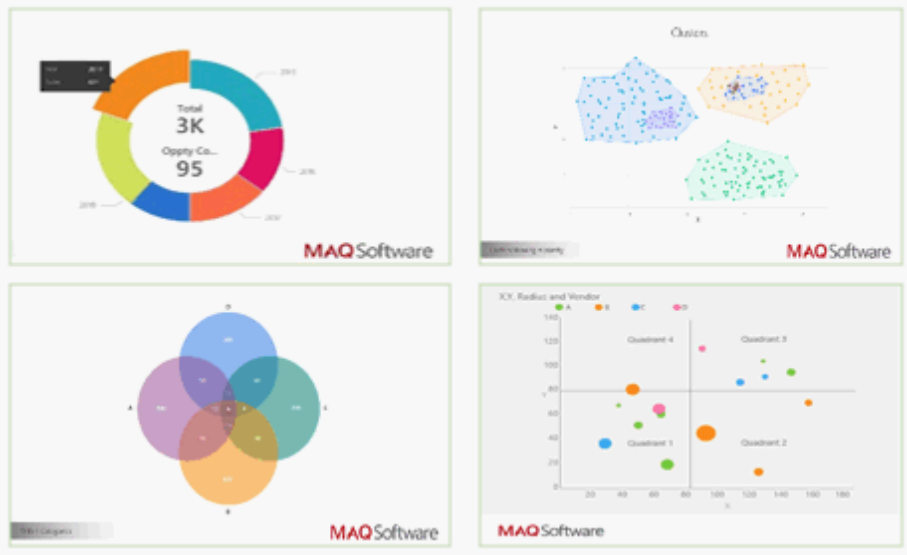
ภาพที่ **2.1** เครื่องมือที่ใช้ทำ Data Visualization

2.1.3.2 รูปแบบในการใช้ Data Visualization จำแนกได้ ดังนี้

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

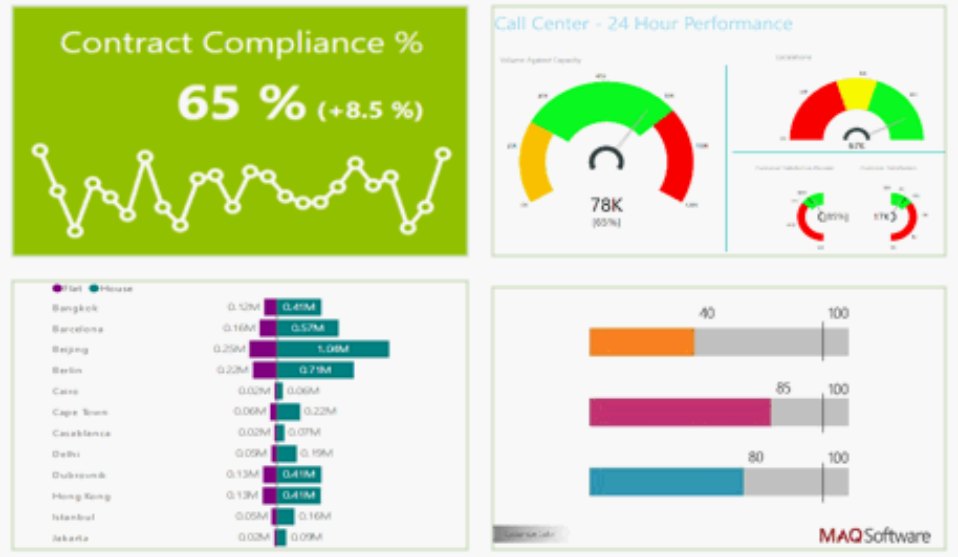
ภาพที่ **2.2** การนำเสนอแบบทิศทางหรือแนวโน้ม (Trending)

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



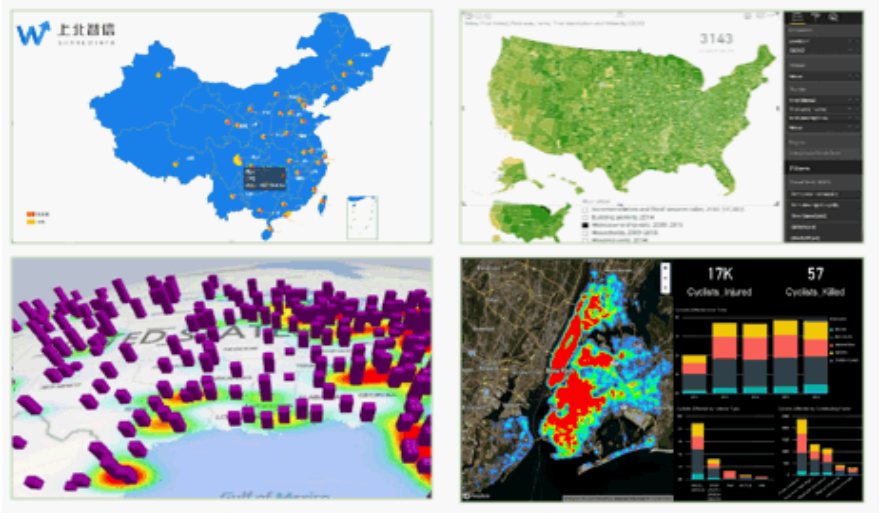
ภาพที่ **2.3** การนำเสนอแบบกลุ่มข้อมูล (Classification)

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



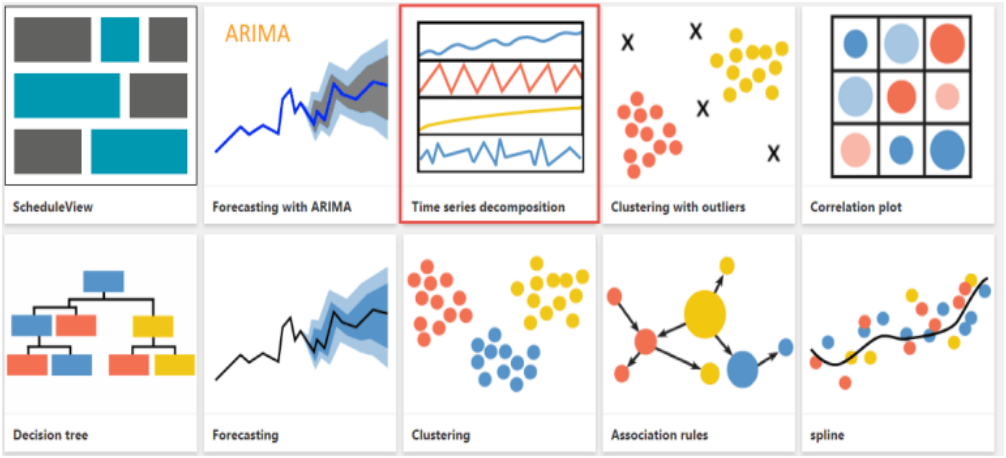
ภาพที่ **2.4** การนำเสนอเชิงเปรียบเทียบข้อมูล (Comparison)

2.1.3.2.4 การนำเสนอรูปแบบแผนที่ (Geographical) เหมาะสำหรับนำเสอข้อมูลบนแผนที่ โดยสามารถที่จะนำยอดขาย, รายได้, ความหนาแน่นของประชากร เพื่อ Focus กลุ่มผู้คนในแต่ละพื้นที่ เช่น Globe Map, Google Map, Flow Map เป็นต้น



ภาพที่ **2.5** การนำเสนอรูปแบบแผนที่ (Geographical)

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



ภาพที่ **2.6** กลุ่มที่ต้องพยากรณ์ล่วงหน้า

2.2 ทฤษฎี

2.2.1 ทฤษฎีเกี่ยวกับการจัดการข้อมูลขนาดใหญ่

2.2.1.1 ข้อมูลขนาดใหญ่

การพัฒนาอย่างรวดเร็วของเทคโนโลยีทำให้การเก็บข้อมูลที่มีปริมาณมหาศาลเป็นเรื่องที่เป็นไปได้ในปัจจุบัน ข้อมูลรูปแบบใหม่ที่จัดเก็บในปัจจุบันไม่เพียงแต่มีความใหญ่ด้านปริมาณเท่านั้น แต่มีความซับซ้อนและรูปแบบที่หลากหลายมากขึ้น เช่น เสียง ภาพนิ่ง ภาพเคลื่อนไหว เป็นต้น ซึ่งก็คือการใช้ประโยชน์จากเทคโนโลยีของอุปกรณ์ตรวจจับที่มีขนาดเล็กลง แต่มีประสิทธิภาพมากขึ้น ลักษณะข้อมูลขนาดใหญ่หรือ Big Data นี้จึงถูกให้คำนิยามว่ามีลักษณะ “The Four V’s” (IBM, Online, 2017) ซึ่งประกอบไปด้วย

1) ปริมาณ (Volume) การจัดเก็บข้อมูลในปัจจุบัน นอกจากจะมีความหลากหลายแล้วความถี่ในการจัดเก็บก็ยังเพิ่มขึ้นด้วย ข้อมูลบางประเภทที่เคยจัดเก็บรายเดือนอาจถูกเปลี่ยนเป็นจัดเก็บรายวัน หรือในบางกรณี เช่น ข้อมูลตำแหน่งของยานพาหนะในกรณีบริษัทขนส่งอาจจัดเก็บในรูปแบบ Real time เหล่านี้ล้วนทำให้ปริมาณข้อมูลเพิ่มขึ้นเป็นอย่างมาก

2) ความหลากหลาย (Variety) นอกเหนือจากข้อมูลที่ถูกจัดเก็บประจำวันระหว่างกระบวนการทำงานปกติแล้ว ปัจจุบันองค์กรต่าง ๆ ยังเริ่มจัดเก็บข้อมูลที่มีรูปแบบหลากหลายมากขึ้น เช่น ข้อมูลจาก Sensor ต่าง ๆ , ลักษณะ Posts บน Social network, รูปภาพหรือไฟล์วิดีโอ, ระยะเวลาการใช้งานบนเว็บไซต์ เป็นต้น

3) ความเร็ว (Velocity) ความเร็วในที่นี้หมายถึงความเร็วของข้อมูลที่ถูกผลิต ตัวอย่างที่เห็นได้ชัด ได้แก่ ข้อมูลราคาซื้อขายหุ้นของตลาดหลักทรัพย์ที่ราคาของหลักทรัพย์มีการแสดงผลจากธุรกรรมซื้อขายที่เกิดขึ้นในอัตราที่รวดเร็วเป็นอย่างมาก ทำให้ความต้องการใช้ข้อมูลในลักษณะ Realtime เพื่อใช้วิเคราะห์และตอบสนองต่อเหตุการณ์ที่เกิดขึ้นในตลาดหลักทรัพย์เพิ่มสูงขึ้นอย่างมาก

4) ความไม่แน่นอน (Veracity) นิยามความไม่แน่นอนถูกกำหนดให้เป็นหนึ่งในคำนิยาม Big Data เนื่องจากข้อมูลล้วนมีที่มาจากหลากหลายแหล่งไม่ว่าจะมาจากภายในหรือภายนอกองค์กร ข้อมูลภายในองค์กรหมายถึง ข้อมูลที่มาจากการประมวลผลภายในหน่วยงาน ยกตัวอย่าง เช่น สถาบันการเงินจะมีข้อมูลจำนวนบัญชีเงินฝากจากสาขา บัญชีเงินกู้จากฝ่ายสินเชื่อ ประวัติการเดินบัญชี ส่วนแหล่งข้อมูลภายนอกองค์กร เช่น สภาพการจราจร สภาพอากาศ ปริมาณนักท่องเที่ยว ด้วยสาเหตุนี้ข้อมูล Big Data จึงมีความไม่แน่นอนจากระดับความซับซ้อนและความหลากหลายของข้อมูล

2.2.1.2 เทคโนโลยีการรวบรวมและจัดเก็บข้อมูล (Data Collection and Data Storage)

เนื่องจากลักษณะข้อมูลของ Big Data มีลักษณะเฉพาะตัว การพัฒนา เทคโนโลยีจัดเก็บข้อมูลเพื่อให้เท่าทันกับลักษณะที่มีความหลากหลายและปริมาณข้อมูลที่มากขึ้น รวมทั้งการคำนึงถึงความถูกต้องและปลอดภัยของข้อมูลจึงมีความสำคัญเป็นอย่างมาก การเก็บข้อมูลในอดีต (Traditional Database) มักจะถูกเก็บอยู่ใน Rational Database ซึ่งคือการจัดเก็บข้อมูลในรูปแบบของตารางซึ่งแต่ละตารางสามารถเชื่อมโยง ความสัมพันธ์ระหว่างกันได้ อย่างไรก็ตามโครงสร้างข้อมูล Rational Database จะต้องเป็นข้อมูล ที่กำหนดรูปแบบ (Normalization) ก่อนการจัดเก็บหรือที่เรียกว่า Schema on write ในการจัดการ ระบบฐานข้อมูลรายธุรกรรม (Database Transaction) เพื่อให้ชุดกระบวนการฐานข้อมูลมี ความน่าเชื่อถือ (Data Integrity) และหลีกเลี่ยงข้อผิดพลาดที่จะเกิดขึ้นกับฐานข้อมูล จึงมี การ

กำหนดคุณลักษณะของข้อมูลที่จัดเก็บบน Rational Database 4 ประการภายใต้ชื่อ “ACID” โดยมีรายละเอียดดังนี้(Suphakit Annoppornchai, Online, 2560)

1) Atomicity ความถูกต้องของฐานข้อมูลภายใต้แนวคิด “All or Nothing” นั่นคือ แต่ละธุรกรรมจะต้องถูกดำเนินการอย่างครบถ้วนสมบูรณ์ มิฉะนั้นแล้วธุรกรรมนั้น จะไม่ถูกดำเนินการใด ๆ เลย เช่น ธุรกรรมการถอนเงินหรือโอนเงิน

2) Consistency ข้อมูลที่จะเข้าสู่ฐานข้อมูลได้จะต้องมีความสอดคล้องถูกต้อง และสมเหตุสมผลกันเท่านั้น

3) Isolation ข้อมูลเป็นอิสระต่อกัน กล่าวคือ เมื่อมีการประมวลผลข้อมูลใดข้อมูลหนึ่ง ผลจากการประมวลผลนั้นจะไม่ส่งผลกระทบกับชุดข้อมูล ถึงแม้ว่าการประมวลผลข้อมูล นั้นจะผิดพลาดหรือหยุดชะงักก็ตาม

4) Durability คือ ข้อมูลแต่ละธุรกรรมจะต้องยังคงอยู่ครบถ้วน ถึงแม้จะเป็นเหตุให้ระบบใช้การไม่ได้ก็ตาม

ต่อมาเมื่อก้าวเข้าสู่ยุคข้อมูลขนาดใหญ่ (Big Data) ที่ข้อมูลเกิดขึ้นอย่าง รวดเร็ว มีปริมาณมหาศาลและมีลักษณะเป็น Unstructured หรือ Semi-structure ซึ่งข้อมูลเหล่านี้ Relational Database ไม่สามารถรองรับได้ดีเพียงพอ (สมเกียรติ ปุ่ยสูงเนิน, ออนไลน์, 2557) เนื่องจากข้อมูลลักษณะ Unstructured หรือ Semi-structure ในทางปฏิบัติไม่สามารถกำหนดรูปแบบก่อนการจัดเก็บได้ ท าให้ Non-Relational Database เข้ามามีบทบาทมากขึ้นเนื่องจาก Non-Relational Database มีจุดเด่นด้านโครงสร้างการจัดเก็บข้อมูลที่ยืดหยุ่นเมื่อเปรียบเทียบกับ Rational Database โดยไม่จำเป็นต้องก าหนดรูปแบบข้อมูลก่อนการจัดเก็บและสามารถจัดเก็บ ข้อมูลประเภท รูปภาพ ภาพเคลื่อนไหว ไฟล์เสียงได้ กระบวนการกำหนดรูปแบบจะดำเนินการเมื่อ ผู้ใช้งานเรียกข้อมูลกลับมาใช้งานอีกครั้ง (Schema on read) และมีจุดเด่นเรื่อง Scalability และ Availability โดยฐานข้อมูลประเภทดังกล่าวมีคุณสมบัติเรียกโดยย่อว่า “BASE” (Charles Roe, 2012) มีรายละเอียดดังนี้

1) Basic Availability ระบบฐานข้อมูลพร้อมใช้งานและต้องมี การตอบสนองต่อคำร้องขอตลอดเวลาเพื่อให้เป็นไปตามทฤษฎี CAP : ซึ่งได้แก่ Consistency ความสมบูรณ์ของระบบ Availability ความพร้อมใช้งานของข้อมูล และ Partition Tolerance ความทนทานต่อการล้มเหลวของระบบ

2) Soft state สถานะของระบบสามารถเปลี่ยนแปลงได้ตลอดเวลา แม้จะไม่มีการใส่ข้อมูล (Input)

3) Eventual consistency เมื่อมีการเพิ่มข้อมูลลงในระบบ ระบบ จะทยอยปรับปรุงข้อมูล Node ในเครือข่ายทั้งหมด ดังนั้น จะมีช่วงระยะเวลาสั้น ๆ ที่สถานะ ข้อมูลไม่สอดคล้องกัน อย่างไรก็ตาม เมื่อผ่านไปสักระยะหนึ่งข้อมูลที่ถูกเก็บในระบบจะถูกปรับปรุง ให้ตรงกัน

ระบบการเก็บข้อมูลและประมวลผลแบบ Non-Relational Database ปัจจุบันเป็นที่นิยมสำหรับการเก็บข้อมูลในลักษณะ Unstructured หรือ Semi-structure ร่วมกับ การใช้บริการจากระบบประมวลผลบน Public service ที่มีความสามารถในการประมวลผลค่อนข้างสูง และมีต้นทุนที่ต่ำอย่าง Hadoop หรือ Service บน Public cloud เป็นต้น

2.2.1.3 เทคโนโลยีจัดเก็บข้อมูลแบบกระจายศูนย์ (Distributed ledger technology)

Distributed ledger technology (DLT) เป็นแนวคิดเทคโนโลยีใหม่ที่ ทำการกระจายฐานข้อมูลให้ผู้ใช้งานทุกคนในเครือข่ายสามารถถือข้อมูลที่เหมือนกันหมดทั้งระบบโดย ไม่จำเป็นต้องมีศูนย์กลางในการให้สิทธิประมวลผลหรือเก็บข้อมูล โดยบัญชีธุรกรรมอิเล็กทรอนิกส์ หรือ Ledger จะถูกเก็บไว้ในเครือข่ายเซิร์ฟเวอร์ที่เรียกว่า Node ข้อมูลบัญชีนี้จะถูกคัดลอกเก็บไว้ใน ทุก ๆ Node ของเครือข่าย โดยที่ไม่ระบุตัวตนว่าเป็นของใคร ทำให้ทุกคนในเครือข่ายสามารถเห็น 17 รายการเดินบัญชีการเงินของผู้ใช้รายอื่น ๆ ได้หากเกิดธุรกรรมใดขึ้น ข้อมูลการทำธุรกรรมนั้นจะถูก ส่งไปถึงทุกคนในเครือข่ายเพื่อช่วยยืนยันความถูกต้องก่อนเพิ่มข้อมูลเข้าไปในระบบฐานข้อมูล

Blockchain เป็นเทคโนโลยีที่มีแนวคิดต้นแบบมาจาก DLT โดย เทคโนโลยีดังกล่าวช่วยให้ข้อมูลธุรกรรมต่าง ๆ จะถูกบันทึกในฐานข้อมูลเพียงชุดเดียวและกระจาย ฐานข้อมูลหรือบัญชี(Ledger) เหล่านั้นไปตามเครื่องคอมพิวเตอร์ต่าง ๆ โดยไม่ผ่านตัวกลางใด ๆ ทั้งสิ้น ผู้ใช้งานเครือข่าย Blockchain ในแต่ละเครื่องจะเห็นข้อมูลทั้งหมดเป็นชุดเดียวกัน (Single version of truth) ดังนั้นการปลอมแปลงข้อมูลภายใต้เทคโนโลยีนี้จึงเกิดขึ้นได้ยากเนื่องจากการบันทึก ตรวจสอบ และรับรองธุรกรรมต่าง ๆ ที่เกิดขึ้นในระบบบัญชีจะต้องผ่านกระบวนการตรวจสอบและกระบวนการรับรองความถูกต้องของธุรกรรมข้อมูลจากผู้ใช้งานเครือข่ายทุก

เครื่องที่มี บัญชีบันทึกข้อมูลธุรกรรมเหล่านั้น จุดเด่นสำคัญที่ทำให้Blockchain แตกต่างจากระบบฐานข้อมูล ทั่วไปโดยสามารถสรุปได้ดังนี้(Business Insider, Online, 2016)

1) Shared Publicly : ทุก Node ในเครือข่าย Blockchain จะเห็นบัญชีธุรกรรมอิเล็กทรอนิกส์ที่เกิดขึ้นและสามารถตรวจสอบได้ว่าธุรกรรมนั้นถูกจัดเก็บใน Block ใด

2) Decentralized : ไม่มีศูนย์กลางในการจัดเก็บข้อมูล รวมถึงไม่มี หน่วยงานกลางในการกำหนดกฎเกณฑ์หรือสิทธิในการเข้าถึงข้อมูล

3) Secure : ข้อมูลที่จัดเก็บในระบบจะไม่สามารถเปลี่ยนแปลงแก้ไข หรือเอาค่าคืนกลับมาได้

4) Trusted : เมื่อมีธุรกรรมเกิดขึ้น ข้อมูลจะต้องได้รับการยอมรับจาก ผู้ใช้งานส่วนใหญ่ก่อนถูกบันทึกเข้าในระบบ ทำให้ระบบมีความน่าเชื่อถือ

5) Automated : การจัดเก็บธุรกรรมข้อมูลอยู่ในรูปแบบอัตโนมัติโดย Software ที่ถูกเขียนขึ้นทำหน้าที่ป้องกันการเขียนข้อมูลที่ซ้ำซ้อนหรือขัดแย้งกัน

เทคโนโลยีBlockchain ได้รับการจับตามองอย่างมากทั้งจากภาครัฐ และภาคธุรกิจด้วยจุดเด่นในเรื่องของความโปร่งใส ความสามารถในการตรวจสอบความถูกต้องข้อมูล และโอกาสในการปลอมแปลงข้อมูลที่น้อยมาก ทำให้องค์กรทั่วโลกต่างตื่นตัวและพยายามศึกษาเพื่อ นำเทคโนโลยีBlockchain ไปประยุกต์ใช้ให้เกิดประโยชน์เช่น รัฐบาลประเทศเอสโตเนียใช้ Blockchain เพื่อเก็บข้อมูลของประชาชนในประเทศ โดยสามารถป้องกันการปลอมแปลงตัวตน และ ทำให้ประชาชนสามารถเข้าถึงข้อมูลได้อย่างปลอดภัย (จรัล งามวิโรจน์เจริญ, ออนไลน์, 2561)

2.2.2 ทฤษฎีเกี่ยวกับการทำเหมืองข้อมูล

เนื่องด้วยปัจจุบันเป็นยุคที่ข้อมูลสารและสนเทศมีความสำคัญ การเผยแพร่และสื่อสารข้อมูล ข่าวสาร ที่ตรงกับความต้องการของผู้ใช้จึงเป็นสิ่งจำเป็น การประยุกต์เทคโนโลยีสารสนเทศเพื่อช่วยในการสื่อสารข้อมูลจำนวนมากให้แก่ผู้ ใช้ เช่น การให้บริการเว็บไซต์เพื่อเผยแพร่ข้อมูลข่าวสารและแลกเปลี่ยนความรู้ จึงเป็นเครื่องมือที่สำคัญในการสื่อสารข้อมูลถึงผู้ใช้จำนวนมาก ดังนั้นการศึกษาเกี่ยวกับพฤติกรรมของผู้ใช้บริการเว็บไซต์ จะช่วยให้องค์กรสามารถนำข้อมูลมาใช้ในการวางแผนพัฒนาเว็บไซต์ ให้ตรงกับความต้องการใช้งานหรือใช้ในการวางแผนกลยุทธ์ เพื่อสร้างความได้เปรียบทางการแข่งขัน

2.2.2.1 การทำเหมืองข้อมูล (Data Mining)

การทำเหมืองข้อมูล (Data Mining) คือกระบวนการที่กระทำกับข้อมูลจำนวนมากเพื่อค้นหารูปแบบและความสัมพันธ์ ที่ซ่อนอยู่ในชุดข้อมูลนั้น ในปัจจุบันการทำเหมืองข้อมูลได้ถูกนำไปประยุกต์ใช้ในงานหลายประเภท ทั้งในด้านธุรกิจที่ช่วยในการตัดสินใจของผู้บริหาร ในด้านวิทยาศาสตร์และการแพทย์รวมทั้งในด้านเศรษฐกิจและสังคม

การทำเหมืองข้อมูลเปรียบเสมือนวิวัฒนาการหนึ่งในการจัดเก็บและตีความหมาย ข้อมูล จากเดิมที่มีการจัดเก็บข้อมูลอย่างง่าย ๆ มาสู่การจัดเก็บในรูปฐานข้อมูลที่สามารถดึงข้อมูลสารสนเทศมาใช้จนถึงการทำเหมืองข้อมูลที่สามารถค้นพบความรู้ที่ซ่อนอยู่ในข้อมูล

2.2.2.2 ขั้นตอนการทำเหมืองข้อมูล

ประกอบด้วยขั้นตอนการทำงานย่อยที่จะเปลี่ยนข้อมูลดิบให้กลายเป็นความรู้ประกอบด้วยขั้นตอนดังนี้

- Data Cleaning เป็นขั้นตอนสำหรับการคัดข้อมูลที่ไม่เกี่ยวข้องออกไป
- Data Integration เป็นขั้นตอนการรวมข้อมูลที่มีหลายแหล่งให้เป็นข้อมูลชุดเดียวกัน
- Data Selection เป็นขั้นตอนการดึงข้อมูลสำหรับการวิเคราะห์จากแหล่งที่บันทึกไว้
- Data Transformation เป็นขั้นตอนการแปลงข้อมูลให้เหมาะสมสำหรับการใช้งาน
- Data Mining เป็นขั้นตอนการค้นหารูปแบบที่เป็นประโยชน์จากข้อมูลที่มีอยู่
- Pattern Evaluation เป็นขั้นตอนการประเมินรูปแบบที่ได้จากการทำเหมืองข้อมูล
- Knowledge Representation เป็นขั้นตอนการนำเสนอความรู้ที่ค้นพบโดยใช้เทคนิคในการนำเสนอเพื่อให้เข้าใจ

2.2.2.3 ประเภทข้อมูลที่ใช้ทำเหมืองข้อมูล

- Relational Database เป็นฐานข้อมูลที่จัดเก็บอยู่ในรูปแบบของตาราง โดยในแต่ละตารางจะประกอบไปด้วยแถวและคอลัมน์ ความสัมพันธ์ของข้อมูลทั้งหมดสามารถแสดงได้โดย Entity Relationship Model

- Data Warehouses เป็นการเก็บรวบรวมข้อมูลจากหลายแหล่งมาเก็บไว้ในรูปแบบเดียวกันและรวบรวมไว้ในที่ ๆ เดียวกัน
- Transactional Database ประกอบด้วยข้อมูลที่แต่ละทรานแซ็กชั่นแทนด้วยเหตุการณ์ในขณะใดขณะหนึ่ง เช่น ใบเสร็จรับเงิน จะเก็บข้อมูลในรูปชื่อลูกค้าและรายการสินค้าที่ลูกค้ารายซื้อ
- Advanced Database เป็นฐานข้อมูลที่จัดเก็บในรูปแบบอื่น ๆ เช่น ข้อมูลแบบ Object-Oriented ข้อมูลที่เป็น Text File ข้อมูลมัลติมีเดีย ข้อมูลในรูปของ Web

2.2.3 ทฤษฎีเกี่ยวกับการออกแบบเว็บไซต์

2.2.3.1 หลักการออกแบบเว็บไซต์

ในการออกแบบเว็บไซต์นั้นประกอบด้วยกระบวนการต่าง ๆ มากมาย เช่น การออกแบบ โครงสร้าง ลักษณะหน้าตาหรือการเขียนโปรแกรม แต่มีหลายคนที่พัฒนาเว็บไซต์โดยขาดการ วางแผนและทำงานไม่เป็นระบบ ตัวอย่าง เช่น การลงมือออกแบบโดยการใช้โปรแกรมช่วยสร้างเว็บ เนื้อหาและรูปแบบก็เป็นไปตามที่นึกขึ้นได้ขณะนั้น

การออกแบบเว็บไซต์อย่างถูกต้องจะช่วยลดความผิดพลาดเหล่านี้ และช่วยลดความเสี่ยงที่จะ ทำให้เว็บประสบความล้มเหลว การออกแบบเว็บไซต์ที่ดีต้องอาศัยการออกแบบและจัดระบบข้อมูล อย่างเหมาะสม

- องค์ประกอบของการออกแบบเว็บไซต์การออกแบบเว็บไซต์ที่มีประสิทธิภาพนั้นต้องคำนึงถึง องค์ประกอบสำคัญดังต่อไปนี้

1) ความเรียบง่าย (Simplicity) หมายถึงการจำกัดองค์ประกอบเสริมให้เหลือเฉพาะ องค์ประกอบหลัก กล่าวคือในการสื่อสารเนื้อหากบผู้ใช้นั้น เราต้องเลือกเสนอสิ่งที่เรา ต้องการนำเสนอจริง ๆ ออกมาในส่วนของกราฟิก สีสัน ตัวอักษรและภาพเคลื่อนไหว ต้องเลือกให้พอเหมาะ ถ้าหากมีมากเกินไปจะรบกวนสายตาและสร้างความรำคาญต่อ ผู้ใช้ตัวอย่างเว็บไซต์ที่ได้รับการออกแบบที่ดี ได้แก่เว็บไซต์ของบริษัทใหญ่ ๆ อยางเช่น Apple Adobe Microsoft หรือ Kokia ที่มีการออกแบบเว็บไซต์ในรูปแบบที่เรียบง่ายไม่ซับซ้อนและใช้งานอยางสะดวก

2) ความสม่ำเสมอ (Consistency) หมายถึง การสร้างความสม่ำเสมอให้เกิดขึ้นตลอดทั้งเว็บไซต์โดยอาจเลือกใช้รูปแบบเดียวกนตลอดทั้งเว็บไซต์ก็ได้ เพราะถ้าหากว่าแต่ละ หน้าในเว็บไซต์นั้นมีความแตกต่างกนมากจนเกินไป อาจทำให้ผู้ใช้เกิดความสับสนและ ไม่แน่ใจว่ากาลังอยู่ในเว็บไซต์เดิมหรือไม่ เพราะฉะนั้นการออกแบบเว็บไซต์ใน

แต่ละ หน้าควรที่จะมีรูปแบบ สไตล์ของกราฟิก ระบบเนวิเกชั่น (Navigation) และโทนสีที่มีความคล้ายคลึงกันตลอดทั้งเว็บไซต์

3) ความเป็นเอกลักษณ์ (Identity) ในการออกแบบเว็บไซต์ต้องคำนึงถึงลักษณะขององค์กร เป็นหลัก เนื่องจากเว็บไซต์จะสะท้อนถึงเอกลักษณ์และลักษณะขององค์กร การเลือกใช้ 10 ตัวอักษร ชุดสี รูปภาพหรือกราฟิก จะมีผลต่อรูปแบบของเว็บไซต์เป็นอย่างมาก ตัวอย่างเช่น ถ้าเราต้องออกแบบเว็บไซต์ของธนาคารแต่เรากลับเลือกสีสันและกราฟิก มากมาย อาจทำให้ผู้ใช้คิดว่าเป็นเว็บไซต์ของสวนสนุกซึ่งส่งผลต่อความเชื่อถือขององค์กรได้

4) เนื้อหา (Useful Content)ถือเป็นสิ่งสำคัญที่สุดในเว็บไซต์ เนื้อหาในเว็บไซต์ต้องสมบูรณ์ และได้รับการปรับปรุงพัฒนาให้ทันสมัยอยู่เสมอ ผู้พัฒนาต้องเตรียมข้อมูลและเนื้อหาที่ ผู้ใช้ต้องการให้ถูกต้องและสมบูรณ์ เนื้อหาที่สำคัญที่สุดคือเนื้อหาที่ทีมผู้พัฒนา สร้างสรรค์ขึ้นมาเอง และไม่ไปซ้ำกับเว็บอื่น เพราะจะถือเป็นสิ่งที่ดึงดูดผู้ใช้ให้เข้ามา เว็บไซต์ได้เสมอ แต่ถ้าเป็นเว็บที่ลิงค์ข้อมูลจากเว็บอื่น ๆ มาเมื่อใดก็ตามที่ผู้ใช้ทราบว่าข้อมูลนั้นมาจากเว็บใด ผู้ใช้ก็ไม่จำเป็นต้องกลับมาใช้งานลิงค์เหล่านั้นอีก

5) ระบบเนวิเกชั่น (User–Friendly Navigation) เป็นส่วนประกอบที่มีความสำคัญต่อเว็บไซต์ มาก เพราะจะช่วยไม่ให้ผู้ใช้เกิดความสับสนระหว่างดูเว็บไซต์ ระบบเนวิเกชั่นจึง เปรียบเสมือนป้ายบอกทาง ดังนั้นการออกแบบเนวิเกชั่น จึงควรให้เข้าใจง่าย ใช้งานได้ สะดวก ถ้ามีการใช้กราฟิ กก็ควรสื่อความหมาย ตำแหน่งของการวางเนวิเกชั่นก็ควรวางให้สม่ำเสมอ เช่น อยู่ตำแหน่งบนสุดของทุกหน้าเป็นต้น ซึ่งถ้าจะให้ดีเมื่อมีเนวิเกชันที่เป็นกราฟิกก็ควรเพิ่มระบบเนวิเกชั่นที่เป็นตัวอักษรไว้ส่วนล่างด้วยเพื่อช่วยอำนวยความสะดวกให้กบผู้ใช้ที่ยกเลิกการแสดงผลภาพกราฟิกบนเว็บบราวเซอร์

6) คุณภาพของสิ่งที่ปรากฏให้เห็นในเว็บไซต์ (Visual Appeal) ลักษณะที่น่าสนใจของ เว็บไซต์นั้นขึ้นอยู่กับความชอบส่วนบุคคลเป็นสำคัญ แต่โดยรวมแล้วก็สามารถสรุปได้ ว่าเว็บไซต์ที่น่าสนใจนั้นส่วนประกอบต่าง ๆ ควรมีคุณภาพ เช่น กราฟิกควรสมบูรณ์ไม่มีรอยหรือขอบขั้นบันได้ให้เห็นชนิดตัวอักษรอ่านง่ายสบายตามีการเลือกใช้โทนสีที่เข้ากันอย่างสวยงาม เป็นต้น

7) ความสะดวกของการใช้ในสภาพต่าง ๆ (Compatibility) การใช้งานของเว็บไซต์นั้นไม่ ควรมีขอบจำกัดกล่าวคือต้องสามารถใช้งานได้ดีในสภาพแวดล้อมที่หลากหลายไม่มีการบังคับให้ผู้ใช้ต้องติดตั้งโปรแกรมอื่นใดเพิ่มเติม นอกเหนือจากเว็บบราวเซอร์ ควรเป็น เว็บที่แสดงผลได้ดีในทุกระบบปฏิบัติการ สามารถแสดงผลได้ในทุกความ

ละเอียดหน้าจอ ซึ่งหากเป็นเว็บไซต์ที่มีผู้ใช้บริการมากและกลุ่มเป้าหมายหลากหลายควรให้ความสำคัญกับเรื่องนี้ให้มาก

8) ความคงที่ในการออกแบบ (Design Stability) ถ้าต้องการให้ผู้ใช้งานรู้สึกว่าเว็บไซต์มี คุณภาพถูกต้องและเชื่อถือได้ควรให้ความสำคัญกับการออกแบบเว็บไซต์เป็นอย่างมาก ต้องออกแบบวางแผนและเรียบเรียงเนื้อหาอย่างรอบคอบ ถ้าเว็บที่จัดทำขึ้นอย่างลวก ๆ ไม่มีมาตรฐานการออกแบบและระบบการจัดการข้อมูลถ้ามีปัญหามากขึ้นอาจส่งผลให้เกิดปัญหาและทำให้ผู้ใช้หมดความเชื่อถือ

9) ความคงที่ของการทำงาน (Function Stability)ระบบการทำงานต่าง ๆ ในเว็บไซต์ควรมี ความถูกต้องแน่นอน ซึ่งต้องได้รับการออกแบบสร้างสรรค์และตรวจสอบอยู่เสมอ ตัวอย่างเช่น ลิงค์ต่าง ๆ ในเว็บไซต์ ต้องตรวจสอบว่ายังสามารถลิงค์ข้อมูลได้ถูกต้อง หรือไม่ เพราะเว็บไซต์อื่นอาจมีการเปลี่ยนแปลงได้ตลอดเวลา ปัญหาที่เกิดจากลิงค์ก็คือ ลิงค์ขาด ซึ่งพบได้บ่อยเป็นปัญหาที่สร้างความรำคาญกับผู้ใช้เป็นอย่างมาก

- กำหนดเป้าหมายของเว็บไซต์ ขั้นตอนแรกของการออกแบบเว็บไซต์คือการกาหนดเป้าหมายของเว็บไซต์ให้แน่ชัดเสียก่อน เพื่อจะได้ออกแบบการใช้งานได้ตรงกับเป้าหมายที่ได้ตั้งเอาไว้ โดยทั่วไปมักจะเข้าใจว่าการทำเว็บไซต์มีจุดมุ่งหมายเพื่อบริการข้อมูลของหน่วยงานหรือองค์กรเท่านั้น แต่ในความเป็นจริงแล้ว เว็บไซต์แต่ละแห่งก็จะมีเป้าหมายของตนเองแตกต่างกันออกไป
- กาหนดกลุ่มผู้ใช้เป้าหมาย ผู้ออกแบบเว็บไซต์จำเป็นต้องทราบกลุ่มผู้ใช้เป้าหมายที่เข้ามาใช้บริการเว็บไซต์เพื่อที่จะได้ตอบสนองความต้องการของผู้ใช้ได้อย่างชัดเจน ตัวอย่างเช่นเว็บไซต์ที่มีกลุ่มผู้ใช้หลากหลาย เช่น เซิร์ชเอ็นจิน เว็บท่า และเว็บไดเรกทอรี่ แต่เว็บไซต์ส่วนใหญ่นั้นจะตอบสนองความต้องการเฉพาะ กลุ่มเท่านั้น ไม่สำหรับทุกคน เพราะคุณไม่สามารถตอบสนองความต้องการของคนที่หลากหลายได้ ในเว็บไซต์เดียว
- สิ่งที่ผู้ใช้ต้องการจากเว็บหลังจากที่ได้เป้าหมายและกลุ่มเป้าหมายของเว็บไซต์แล้ว ลำดับต่อไปคือการออกแบบเว็บไซต์เพื่อดึงดูดผู้ใช้งานให้ได้นานที่สุดด้วยการสร้างสิ่งที่น่าสนใจเพื่อดึงดูดผู้ใช้โดยทั่วไปแล้ว สิ่งที่ผู้ใช้ คาดหวังจากการเข้าชมเว็บไซต์หนึ่ง
- ข้อมูลหลักที่ควรมีอยู่ในเว็บไซต์เมื่อเราทราบถึงความต้องการที่ผู้ใช้ต้องการได้รับเมื่อเข้าชมเว็บไซต์หนึ่ง ๆ แล้วก็ออกแบบเว็บไซต์ให้มีข้อมูลที่ผู้ใช้ต้องการ

2.2.4 ทฤษฎีเกี่ยวกับการ visualization

Visualization เป็นส่วนประกอบสำคัญใน Cognitive System ซึ่งเป็นส่วนในการแสดงข้อมูลหรือผลลัพธ์ต่าง ๆ ในระหว่างคอมพิวเตอร์และผู้ใช้งานในรูปแบบของภาพ โดยผู้ใช้

สามารถ เรียนรู้และจดจำข้อมูลผ่านการมองได้มากกว่าการใช้ประสาทสัมผัสอื่น ๆ หรือจะกล่าวว่า Visualization ก็คือ การสร้างมโนภาพของสิ่งต่าง ๆ ที่เราสนใจขึ้นมาในใจ ซึ่งต่อมาได้กลายเป็นการ นำภาพมาใช้กับการนำเสนอหรือนำมาเป็นกรอบความคิด ซึ่งได้นำไปใช้ในการสนับสนุนการ ตัดสินใจ

ข้อดีของ Visualization มีดังนี้

- ช่วยในการแสดงข้อมูลที่มีปริมาณสูง
- ช่วยในการรับรู้หรือคาดคะเนสิ่งที่อาจจะเกิดขึ้นในอนาคต
- Visualization ไม่เพียงแต่แสดงรายละเอียดข้อมูลในตัวเองเท่านั้นยังสามารถแสดงปัญหา ที่เกิดขึ้นได้ด้วย
- ช่วยให้ผู้ใช้สามารถเข้าใจหรือวิเคราะห์ข้อมูลได้สะดวกขึ้นไม่ว่าขอบเขตข้อมูลนั้นจะมี ขนาดใหญ่หรือเล็ก
- ช่วยให้ผู้ใช้สามารถสันนิษฐานข้อมูลได้สะดวกขึ้น (Ware, 2004)

ซึ่ง Visualization เป็นมากกว่า วิธีการทาง Computer Visualization เป็นการนำเสนอข้อมูล แบบหนึ่งที่ทำการแสดงข้อมูลให้อยู่ในรูปแบบของ VisualForm ซึ่งอาจจะเป็นการแสดงในรูปแบบ ของรูปภาพ ,กราฟ หรือ แผนภาพ ซึ่ง ผลลัพธ์ของการทำ Visualization คือ การนำข้อมูลต่าง ๆ ที่ ซ่อนอยู่ในตัวของข้อมูลเองออกมาให้ผู้ใช้สามารถสัมผัสได้ ซึ่งกระบวนการต่าง ๆ ในการแสดงหรือ นำข้อมูลต่าง ๆ ออกมา อาจจะถูกแอบซ่อนอยู่หรือผู้ใช้ไม่ทันสังเกต แต่อย่างไรก็ตาม Visualization เป็นสิ่งที่จำเป็นในการค้นหาข้อมูล หรือ ใช้ในการวิเคราะห์ข้อมูล โดยจุดมุ่งหมายของ Visualization ก็คือ การถ่ายทอดข้อมูลไปสู่ระบบการรับรู้โดยภาพของผู้ใช้ระบบ (Diehl, 2007) เพื่อช่วยในการ ลดช่องว่างระหว่างผู้ใช้และตัวข้อมูล และยังช่วยผู้ใช้สามารถเรียนรู้ ได้อย่างมีประสิทธิภาพ (Keim, 2002)

2.2.4.1 วิธีการของ Visualization

การใช้ Visual เพื่อทำการค้นหาข้อมูลนั้นผู้ใช้จะทำขั้นตอนหลักๆอยู่ 3 ขั้นตอน คือ Overview First, Zoom and Filter และ Detail on Demand โดยอันดับแรก ผู้ใช้ต้องการที่จะดูข้อมูลภาพรวมทั้งหมดซึ่งหลังจากดูภาพรวมทั้งหมดแล้วผู้ใช้ก็จะท าการตัดสินใจเลือกรูปแบบหรือกลุ่มข้อมูลที่สนใจซึ่งก็จะมาถึงขั้นตอนในการวิเคราะห์ข้อมูลผู้ใช้ก็จะทำการเจาะลึกถึงข้อมูลในรายละเอียด ซึ่ง Visualization Technology ก็จะอ้างอิงหรือพัฒนาจากขั้นตอนเหล่านี้ซึ่งVisualization Technique มีประโยชน์มากในการแสดงภาพรวมหรือแสดงข้อมูลย่อยที่ผู้ใช้ต้องการโดยอาจจะใช้หลายๆวิธีการรวมกันเพื่อให้ได้ข้อมูลที่ผู้ใช้ต้องการ ซึ่งช่วยลดช่องว่างของกิจกรรมที่ใช้ในการดึงข้อมูลต่าง ๆ ไปใช้ ซึ่งลักษณะของข้อมูลที่สามารถ

นำมาผ่านกระบวนการของVisualization มีลักษณะต่าง ๆ มากมายดังนี้ ข้อมูล 1D ได้แก่ ข้อมูลที่เกี่ยวข้องกับช่วงเวลาใดเวลาหนึ่ง,ข้อมูล 2D ได้แก่ ข้อมูลที่เกี่ยวข้องกับแผนที่ภูมิศาสตร์, Multi Dimensional Data ได้แก่ Relation Table, Text และ Hypertext ได้แก่ ข้อมูลหัวข้อข่าวต่าง ๆ และ Web Document,Hierarchies และ Graph ได้แก่ หมายเลขโทรศัพท์ และ Web Document ,Algorithms และ Softwareได้แก่ Debugging Operation ซึ่งแต่ละข้อมูลก็จะมีวิธีการที่ช่วยในการจัดการแสดงผลข้อมูลที่หลากหลายๆ เช่น

- ประเภทที่แสดงเป็น 2D/3D เช่น แผนภูมิแท่ง และ แผนภูมิ xy
- ประเภทที่แสดงข้อมูลในลักษณะภูมิศาสตร์ เช่น Parallel Coordinates และ ภาพ Landscape
- ประเภทที่แสดงเป็น Icon-Base เช่น Needle Icon และ Star Icon
- ประเภทที่แสดงข้อมูลแบบ Dense Pixel เช่น Recursive Pattern และ Circle Segment
- ประเภทที่แสดงข้อมูลแบบ Stack เช่น Tree Stamp และ Dimension Stacking
- (Dykes, MacEachren & Kraak, 2005,Keim ,2002)

2.2.5 ทฤษฎีเกี่ยวกับต้นไม้ตัดสินใจ

เทคนิคต้นไม้ตัดสินใจเป็นเทคนิคหนึ่งของการจำแนกประเภท(Classification) ซึ่งเป็นวิธีการแบ่งประเภทหรือแยกหมวดหมู่ข้อมูลโดยการจำแนกประเภทนั้นเป็นเทคนิคหนึ่งของ

2.2.5.1 เหมืองข้อมูลคือการวิเคราะห์ข้อมูลเพื่อแยกประเภท จำแนกรูปแบบและความสัมพันธ์ของข้อมูลจากฐานข้อมูลที่มีขนาดใหญ่หรือคล้องข้อมูลโดยมีเทคนิคต่าง ๆ หลายวิธีซึ่งรูปแบบการทำเหมืองข้อมูลนั้นได้รวมความรู้จากหลายแขนงเข้าไวด้วยกันที่ประกอบด้วยระบบการเรียนรู้ของเครื่องจักรรวมกับวิทยาศาสตร์สารสนเทศ สถิติและ ระบบฐานข้อมูล โดยทั่วไปแล้วเทคนิคที่นำมาใช่ส่วนใหญ่มี5 ประเภท(ศุภชัย, 2551)

1) เทคนิคการจำแนกข้อมูล(Classification) เป็นการจำแนกกลุ่มข้อมูลด้วยคุณลักษณะต่าง ๆ ที่ได้มีการกำหนดไว้แล้วเทคนิคประเภทนี้เหมาะกบัการสร้างแบบจำลองเพื่อการพยากรณ์ค่าข้อมูล(Predictive Modeling) ในอนาคตจากการที่ได้จำแนกกลุ่มข้อมูลตัวอย่างไว้แล้วซึ่งในลักษณะดังกล่าวนี้เรียกว่าการเรียนรู้ข้อมูลแบบมีผู้สอน (Supervised Learning)เทคนิคการการจำแนกข้อมูล มี2 รูปแบบ ได้แก่แบบต้นไม้ตัดสินใจ และแบบโครงข่ายประสาทเทียม และเป็นกระบวนการสร้างแบบจำลองเพื่อจัดการข้อมูลให้อยู่ใน

กลุ่มที่กำหนดตัวอย่างเช่น การแบ่งประเภทลูกค้าว่าเชื่อถือได้หรือไม่ซึงเป็นการสร้างแบบจำลองโดยการเรียนรู้จากขอ้มูลที่ได้กำหนดกลุ่มไว้เรียบร้อยแล้ว

2) เทคนิคค้นหาความสัมพันธ์ (Association Rule Discovery) เป็นการค้นหาความสัมพันธ์ของฐานข้อมูลที่มีขนาดใหญ่ เพื่อที่จะทำการวิเคราะห์ข้อมูลและหาสิ่งที่ซ่อนอยู่ในข้อมูลนั้น เช่น การวิเคราะห์ข้อมูลการซื้อขายในซุปเปอร์มาร์เก็ต เพื่อทำการจัดทำการวางแผนเพื่อจัดการส่งเสริมการขายและเตรียมการวางแผนการเรียงชั้นวางสินค้า เช่น การวางน้ำอัดลมกับข้าวโพดคั่ว ไว้ใกล้กัน

3) เทคนิคจัดหมวดหมู่(Clustering) เป็นการลดขนาดของข้อมูลด้วยการรวมกลุ่มตัวแปรที่มีลกัษณะเดียวกันไว้ด้วยกัน ทำให้สามารถค้นหาข้อมูลที่ถูกละเลยไปได้เทคนิคนี้กับถูกใช้เป็นขั้นตอนเบื้องต้นในการทำเหมืองขอ้มูลและเหมาะกับข้อมูลที่ยังไม่มีกลุ่มอย่างชัดเจนจึงทำการจัดหมวดหมู่เพื่อหากลุ่มต่าง ๆ ของข้อมูลโดยจำนวนกลุ่มของข้อมูลมักจะใช้ตัว k แทนซึ่งผู้ที่ใช้เทคนิคนี้จะเป็นผู้กำหนดจำนวนกลุ่มวิธีนี้อาจเรียกอีกอย่างว่าเคมีน (K–mean)

4) เทคนิคหาค่าที่แตกต่างไปจากค่ามาตรฐาน(Deviation Detection) เป็นการหาค่าที่แตกต่างไปจากค่ามาตรฐาน หรือค่าที่คาดคิดไวว่าต่างไปมากน้อยเพียงใดโดยทั่วไปมักใช้วิธีการทางสถิติหรือการแสดงให้เห็นภาพสำหรับเทคนิคนี้ใช้ในการตรวจสอบลายเซ็นปลอมหรือบัตรเครดิตปลอมเป็นต้น

5) เทคนิคการวิเคราะห์ลำดับ(Sequential Analysis) เป็นการวิเคราะห์ลำดับเพื่อค้นพบรูปแบบของการปรากฏของข้อมูล ซึ่งปรากฏในรายการที่แยกออกมา เช่น ถ้าผู้ซื้อ ซื้อสินค้า A แล้วเขาจะซื้อสินค้า B ในภายหลัง เทคนิคนี้จะแตกต่างจากเทคนิคค้นหาความสัมพันธ์เพราะคำนึงถึงลำดับการซื้อด้วย

2.2.5.2 วิธีการแบ่งประเภทหรือแยกหมวดหมู่ข้อมูลการจำแนกประเภท

คือกระบวนการสร้างแบบจำลองเพื่อจัดการข้อมูลให้อยู่นกลุ่มที่กำหนดเป็นการสร้างแบบจำลองการจัดหมวดหมู่ได้จากกลุ่มตัวอย่างของข้อมูลที่ได้กำหนดไว้ล่วงหน้าและสามารถพยากรณ์กลุ่มของข้อมูลที่ยังไม่เคยนำมาจัดหมวดหมู่ได้แบบจำลองที่ได้อาจอยู่ในรูปแบบของต้นไม้ตัดสินใจหรือแบบโครงข่ายประสาทเทียมในการจัดหมวดหมู่จำเป็นต้องชุดของข้อมูลเพื่อใช้สำหรับการเรียนรู้(Training Data) และสร้างแบบจำลอง(Model Construction) ซึ่งสามารถทดสอบโดยชุดของข้อมูลสำหรับการทดสอบ(TestingData) เพื่อประเมินความ

ถูกต้องของแบบจำลอง(Model Evaluation) อีกทั้งใช้ชุดข้อมูลที่ไม่เคยเห็นมาก่อน(Unseen Data) เพื่อทำการกำหนดประเภทให้กับข้อมูลใหม่ที่ได้ทำนายค่าออกมาตามที่ต้องการ เช่น การจัดหมวดหมู่ของผู้ขอเครดิตเป็นระดับต่ำระดับกลางและระดับสูงของความเสี่ยงที่จะได้รับหรือการอนุมัติบุคคลเข้ารับทำงานในลักษณะงานต่าง ๆ

2.2.5.3 เทคนิคต้นไม้ตัดสินใจคือแบบจำลองทางคณิตศาสตร์เพื่อการหาทางเลือกที่ดีที่สุด โดยการนำข้อมูลมาสร้างแบบจำลองการพยากรณ์ในรูปแบบของโครงสร้างต้นไม้ซึ่งมีการเรียนรู้ข้อมูลแบบมีผู้สอนสามารถสร้างแบบจำลองการจัดหมวดหมู่ได้จากกลุ่มตัวอย่างของข้อมูลที่กำหนดไว้ล่วงหน้าโดยอัตโนมัติและสามารถพยากรณ์กลุ่มของรายการที่ยังไม่เคยนำมาจัดหมวดหมู่ได้อีกด้วย โดยปกติมักประกอบด้วยกฎในรูปแบบ “ถ้า เงื่อนไข แล้วผลลัพธ์” เช่น

“If Income= High and Married = No THEN Risk = Poor”

“If Income = High and Married = Yes THEN Risk = Good”

2.2.5.3.1 ส่วนประกอบของต้นไม้ตัดสินใจประกอบด้วย

- โหนด(Node) คือคุณสมบัติต่าง ๆ เป็นจุดที่แยกข้อมูลว่าจะให้ไปในทิศทาง ใดซึ่งโหนดที่อยู่สูงสุดเรียกว่าโหนดราก(Root Node)
- กิ่ง(Branch) คือ คุณสมบัติของคุณสมบัติในโหนดที่แตกออกมาโดยจำนวนของกิ่งจะเท่ากับคุณสมบัติของโหนด
- ใบ(Leaf) คือ กลุ่มของผลลัพธ์ในการแยกแยะข้อมูล

โดยสามารถแสดงส่วนประกอบของต้นไม้ตัดสินใจ แสดงตัวอย่างของต้นไม้ตัดสินใจได้

2.2.5.3.2 การสร้างต้นไม้ตัดสินใจหลักการพื้นฐานของการสร้างต้นไม้ตัดสินใจเป็นการสร้างในลักษณะจากบนลงล่าง(Top–Down) คือเริ่มจากการสร้างรากของต้นไม้ก่อนแล้วจึงแตกกิ่งไปจนถึงใบโดยแสดงขั้นตอนการสร้างต้นไม้ตัดสินใจได้ดังนี้ (ฮานและแคมเบอร์, 2001)

1) ต้นไม้เริ่มต้นโดยมีโหนดเพียงโหนดเดียวแสดงถึงชุดข้อมูลฝึก

2) ถ้าข้อมูลทั้งหมดอยู่ในกลุ่มเดียวกันแล้ว ให้โหนดนั้นเป็นใบและตั้งชื่อแยกตามกลุ่มของข้อมูลนั้น

3) ถ้าในโหนดมีข้อมูลหลายกลุ่มปะปนอยู่จะต้องวัดค่าเกน(Gain) ของแต่ละแอททริบิวต์เพื่อที่จะใช้เป็นเกณฑ์ในการคัดเลือกแอททริบิวต์ที่มีความสามารถในการแบ่งแยกข้อมูลออกเป็นกลุ่มต่าง ๆ ได้ดีที่สุด โดยแอททริบิวต์ที่มีค่าเกนมากที่สุดจะถูกเลือกให้เป็นตัวทดสอบหรือแอททริบิวต์ใช้ในการตัดสินใจ โดยแสดงในรูปของโหนดบนต้นไม้

4) กิ่งของต้นไม้ถูกสร้างขึ้นจากค่าต่าง ๆ ที่เป็นไปได้ของโหนดทดสอบ และข้อมูลจะถูกแบ่งออกตามกิ่งต่าง ๆ ที่สร้างขึ้น

5) ทำการวนซ้ำเพื่อหาแอททริบิวต์ที่มีค่าเกนมากที่สุดสำหรับข้อมูลที่ถูกแบ่งแยกออกมาในแต่ละกิ่งเพื่อนำแอททริบิวต์นี้มาสร้างเป็นโหนดตัดสินใจต่อไปโดยที่แอททริบิวต์ที่ถูกเลือกมาเป็นโหนดแล้วจะไม่ถูกเลือกมาอีก สำหรับโหนดในระดับต่อ ๆ ไป

6) ทำการวนซ้ำ เพื่อแบ่งข้อมูลและแตกกิ่งของต้นไม้ไปเรื่อย ๆ โดยการวนซ้ำจะสิ้นสุดก็ต่อเมื่อเงื่อนไขข้อใดข้อหนึ่งต่อไปนี้เป็นจริง

2.3 เครื่องมือในการออกแบบและวิเคราะห์ข้อมูล

2.3.1 โปรแกรม RapidMiner

ซอฟต์แวร์ RapidMiner สามารถจัดการข้อมูลได้ตั้งแต่การเตรียมข้อมูล (Data Preparation) สร้างโมเดล (Model & Validate) ไปจนถึงนำไปใช้งานใน production (Operationaize) ซึ่งจะแยกส่วนได้ดังนี้

• RapidMiner Radoop เป็นเวอร์ชั่นที่ทำงานบน Hadoop (ซึ่งเป็นการนคอมพิวเตอร์หลายๆ เครื่องมาช่วยประมวลผล) ทำให้สามารถรองรับการทำงานกับข้อมูลที่มีขนาดใหญ่ๆ หลาย (ร้อย) ล้านเรคอร์ดได้

• RapidMiner Studio เป็นเวอร์ชั่นที่ทำงานบนเครื่องคอมพิวเตอร์ของเราเอง (อาจจะเป็น PC หรือ Notebook ก็ได้) เป็นตัวหลักในการออกแบบโพรเซส (process) หรือ workflow เพื่อใช้ในการวิเคราะห์ข้อมูลต่าง ๆ เช่น สร้างโพรเซสในการคาดการณ์ว่าลูกค้าคนใดจะยกเลิกการใช้บริการ (churn) ด้วยโมเดล Decision Tree

• RapidMiner Server เป็นเวอร์ชันที่ทำงานบนเครื่องคอมพิวเตอร์และรองรับการทำงานที่มีผู้ใช้งานหลายๆ คนพร้อมกัน โดยเวอร์ชั่นนี้สามารถสร้างกราฟในลักษณะของ BI

(Business Intelligence) ตั้งเวลาให้ทำงาน (scheduler) และสร้าง web service เพื่อให้โปรแกรมต่าง ๆ มาติดต่อได้ด้วย

2.3.2 โปรแกรม Tableau Public

Tableau คือ โปรแกรมที่ใช้ในการแสดงผลข้อมูลในรูปแบบของภาพ(Data Visualiztion) เพื่อทำให้มีความเข้าใจในข้อมูลมากขึ้น ทำให้สามารถนำข้อมูลจำนวนมากมาวิเคราะห์และมองออกได้อย่างง่ายขึ้น ซึ่งตัวโปรแกรมTableau มีด้วยกันหลายเวอร์ชั่นด้วยกัน เช่น Tableau Desktop ,Tableau Online เป็นต้น โปรแกรม Tableau Public เป็นโปรแกรมที่ใช้งานฟรี และสามารถใช้งานได้ดี

ภาพที่ **2.7** โปรแกรม Tableau Public

คุณสมบัติและจุดเด่นของโปรแกรม Tableau

1) คุณสมบัติที่โดดเด่นอย่างเห็นได้ชัด คือ ผู้ใช้มีเครื่องมือที่ใช้วิเคราะห์ข้อมูลได้หลากหลายมิติ

2) รูปแบบการใช้งานที่ง่ายเพียงคลิกเมาส์ก็สามารถที่จะเปลี่ยนแปลงรายงานได้ตามความต้องการ

3) รองรับการเข้าถึงข้อมูลจากหลายฐานข้อมูล เช่น Microsoft Excel, Microsoft Access, Firebird 2.0, IBM DB2, MS SQL Server, Microsoft Power pivot เป็นต้น

4) รูปแบบการนำเสนอรายงานที่สวยงาม เข้าใจง่าย และง่ายในการนำเสนอต่อผู้บริหาร

5) สามารถเพิ่มทักษะและแนวคิดให้กับพนักงานระดับปฏิบัติการ ให้มีแนวคิดเชิงสถิติและการประยุกต์ใช้งานมากขึ้น

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

2.4 วรรณกรรมที่เกี่ยวข้อง

พัทธนันท์ รัตนวรเศวต (2563) วงเวลาที่มีการเกิดเหตุอาชญากรรมเกี่ยวกับบุคคลและทรัพย์สินสูงในซอยเสือใหญ่อุทิศนั้น จะมีการเกิดเหตุอาชญากรรมสูงมากในช่วงเวลา 00.01 – 04.00นาฬิกา รองลงมาคือช่วงเวลา 20.01 –00.00นาฬิกา ซึ่งในช่วงเวลาที่มีการเกิดเหตุอาชญากรรมสูงดังกล่าวนั้นเป็นช่วงเวลากลางคืน อันเป็นเวลาพักผ่อนของประชาชนโดยทั่วไป ร้านค้าและอาคารพาณิชย์ส่วนใหญ่จะปิด ทำให้มีประชาชนสัญจรไปมาในพื้นที่น้อย ส่งผลให้อาชญากรมีโอกาสก่ออาชญากรรมได้โดยสะดวก และพื้นที่เสี่ยงสูงต่อการเกิดอาชญากรรมสูงเกี่ยวกับบุคคลและทรัพย์สินในซอยเสือใหญ่อุทิศนั้น ส่วนใหญ่จะอยู่ในเขตพื้นที่ซอยรัชดา 36 แยกต่าง ๆ ซึ่งเป็นพื้นที่มีที่พักอาศัยประเภทหอพัก อาคารชุดพักอาศัย และอาคารพาณิชย์เป็นส่วนใหญ่ เนื่องจากซอยรัชดา 36หรือซอยเสือใหญ่อุทิศนั้น เป็นย่านใจกลางเมือง อยู่ใกล้สถานศึกษาที่มีชื่อเสียง ห้างสรรพสินค้าหลายแห่ง มีบริษัท ห้างฯ ร้านค้ามากมาย มีผู้คนอพยพย้ายถิ่นฐานเข้ามาหางานทำและประกอบอาชีพมากมายในพื้นที่ ทำให้มีการเกิดอาชญากรรมสูงในบริเวณพื้นที่

ชำนิ คนไว (2562) ปัญหาอาชญากรรมในสังคมไทยในปัจจุบัน จากข้อมูลคดีอาญาทางข่าวสารหนังสือพิมพ์ โทรทัศน์หรือ สื่อสารทางอิเล็กทรอนิกส์ ต่าง ๆ ทั่วราชอาณาจักรมีแนวโน้มที่จะรุนแรงขึ้น ทั้งในด้านสถิติตัวเลข การก่ออาชญากรรม และวิธีการประกอบอาชญากรรมของผู้กระทำความผิด ที่สามารถมองเห็นได้ เช่น คดีลักทรัพย์ ชิงทรัพย์ ปล้นทรัพย์ ข่มขืนกระทำชำเรา คดียาเสพติด เป็นต้น และที่มองไม่เห็น เช่น คดีเกี่ยวกับอาชญากรรมคอมพิวเตอร์ แก๊งคอลเซ็นเตอร์ การหลอกลวงขายสินค้าทางอินเตอร์เน็ต เป็นต้น จากการนำเสนอภาคข่าวเกี่ยวกับการกระทำผิดผ่านสื่อต่าง ๆ ทั้งจากโทรทัศน์ วิทยุ หนังสือพิมพ์ และอินเทอร์เน็ต มีจำนวนผู้กระทำความผิด และผู้ที่เป็นเหยื่อของอาชญากรรม

เป็นจำนวนมาก ทำให้รัฐบาล จะต้องเพิ่มงบประมาณ ด้านกำลังพล ด้านวัสดุอุปกรณ์ ด้านการสืบสวน สอบสวน การป้องกัน การปราบปรามอาชญากรรม การควบคุม การสร้างสวัสดิการ บำบัดรักษา และการแก้ไขฟื้นฟู ดังนั้น ปัญหาอาชญากรรมในสังคมไทย จะหมดสิ้นไปถ้าคนในสังคมร่วมมือสอดส่องดูแล ให้ความร่วมมือกันและกัน สนับสนุนรัฐบาลและรัฐบาลก็สนับสนุนประชาชนในการรักษาความสงบเรียบร้อย การอยู่ดีกินดี ดำเนินชีวิตเศรษฐกิจพอเพียง สังคมก็จะสงบสุข

อัชฌาพร กว้างสวาสดิ์, เพียงฤทัย หนูสวัสดิ์, วราลี คงเหมาะ, ปวีณา ทิพยากุลรักษ์ และบุษกร สังขนันท์ (2560) งานวิจัยนี้มีวัตถุประสงค์ 1) เพื่อสร้างโมเดลการทำนายระดับความเครียด ด้วยเทคนิคต้นไม้การตัดสินใจ (Decision tree) 2) เพื่อพัฒนาระบบทำนายระดับความเครียด ด้วยเทคนิคต้นไม้การตัดสินใจ (Decision tree) 3) เพื่อประเมินการยอมรับของระบบทำนายระดับความเครียด โดยมีเครื่องมือในการพัฒนาระบบ คือ ภาษา PHP ซึ่งการสร้างโมเดลการทำนายระดับความเครียดโดยใช้เทคนิคการทำเหมืองข้อมูลแบบต้นไม้การตัดสินใจ ใช้กลุ่มข้อมูลตัวอย่างจากผู้ตอบแบบสอบถาม จำนวน 2,000 ตัวอย่าง โดยมีการวัดประสิทธิภาพของกฎด้วยวิธีการแบ่งกลุ่มข้อมูลแบบ Cross-validation Test และนำกฎที่ได้มาพัฒนาเป็นระบบทำนายระดับความเครียด เมื่อได้ระบบทำนายระดับความเครียดแล้ว นำไปประเมินการยอมรับของระบบ โดยเครื่องมือที่ใช้ในงานวิจัยคือ 1) ระบบทำนายระดับความเครียด 2) แบบสอบถามวัดการยอมรับของระบบ โดยมีการวัดการยอมรับจากการใช้งานของกลุ่มตัวอย่าง จำนวน 30 ตัวอย่าง สถิติที่ใช้ได้แก่ ค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐาน ผลการวิจัยพบว่า 1) จากการสร้างโมเดลทำนายระดับความเครียดทำให้ได้กฎการทำนายจำนวน 120 กฎ ซึ่งมีค่าความถูกต้องของโมเดลอยู่ที่ร้อยละ 91.10 2) ผลการพัฒนาระบบทำนายระดับความเครียดผ่านเว็บไซต์ www.predictstress.com ซึ่งระบบสามารถวัดระดับความเครียดได้และเสนอแนะข้อปฏิบัติเป็นแนวทางแก้ไขความเครียดได้ 3) ผลการประเมินการยอมรับของระบบทำนายระดับความเครียด ด้วยเทคนิคต้นไม้การตัดสินใจ โดยมีการประเมิน 3 ด้าน คือด้านเนื้อหา ด้านการออกแบบ และด้านการใช้งาน ซึ่งภาพรวมยอมรับระดับมาก (4.21)

เด่นเดือน เลิศทยากุล, ธรา อั่งสกุล, วีรพงษ์ พลนิกรกิจและจิติมนต์ อั่งสกุล (2558) ด้วยประสิทธิภาพของสื่ออินเทอร์เน็ตที่สามารถเสนอขายสินค้า และบริการได้ตลอด 24 ชั่วโมง เข้าถึงกลุ่มผู้บริโภคได้แบบเจาะจงเป็นรายบุคคล ประกอบกับอัตราค่าใช้จ่ายในการแพร่กระจายสื่อโฆษณาทางสื่ออินเทอร์เน็ตนั้นยังต่ำกว่าสื่อโฆษณาประเภทอื่น จึงทำให้สื่อ

อินเทอร์เน็ตกลายเป็นสื่อโฆษณารูปแบบใหม่ที่ผู้ประกอบการพาณิชย์อิเล็กทรอนิกส์ และผู้ประกอบการทั่วไปให้ความสนใจมากขึ้น ซึ่งในการโฆษณาทางอินเทอร์เน็ตนั้น ผู้ประกอบการต้องคำนึงถึงเว็บไซต์ และหมวดหมู่ในเว็บไซต์ รวมไปถึงงบประมาณในการโฆษณาที่สามารถตอบสนองความต้องการของกลุ่มผู้บริโภคเป้าหมายได้ บทความนี้จึงนำเสนอระบบสนับสนุนการตัดสินใจเพื่อการเลือกโฆษณาทางเว็บไซต์ โดยคำนึงถึงปัจจัยสำคัญด้านลักษณะประชากรศาสตร์ของผู้บริโภคกลุ่มเป้าหมาย และหมวดหมู่ในเว็บไซต์ที่เหมาะสมกับกลุ่มผู้บริโภคเป้าหมาย โดยระบบได้นำเสนอการทำเหมืองข้อมูลโดยใช้แบบจำลองต้นไม้ประกอบการตัดสินใจ เพื่อใช้เลือกหมวดหมู่ในเว็บไซต์ที่เหมาะสม ซึ่งผลที่ได้จากการทดสอบพบว่า การเลือกหมวดหมู่ในการโฆษณาทางเว็บไซต์เหล่านั้นมีความถูกต้องตรงกับกลุ่มผู้บริโภคเป้าหมายได้สูงถึง 84.74% และจากการประเมินความสามารถในการใช้งานได้ของระบบจากกลุ่มตัวอย่าง 3 กลุ่ม ได้แก่ กลุ่มผู้ประกอบการพาณิชย์อิเล็กทรอนิกส์ กลุ่มผู้ประกอบการทั่วไป และกลุ่มผู้ให้บริการโฆษณาทางเว็บไซต์ จากการใช้แบบสอบถามออนไลน์โดยรวม พบว่า กลุ่มตัวอย่างมีความพึงพอใจอยู่ในระดับมาก โดยพึงพอใจมากด้านประสิทธิภาพของระบบในส่วนของระบบที่สามารถประมวลผลได้รวดเร็ว ผู้ใช้สามารถใช้งานได้ทันทีโดยไม่ต้องเรียนรู้ใหม่ เมื่อกลับมาใช้งานระบบอีกครั้ง ระบบสามารถวางแผนโฆษณาได้ตรงตามงบประมาณที่ผู้ใช้กำหนด สามารถคาดการณ์ค่าใช้จ่ายในการโฆษณาได้ ช่วยในการวางแผนการโฆษณาทางเว็บไซต์ได้

สุขุมา อรุณจิตและวุฒิพล มั่นเหมาะ (2559) อาชญากรรมทางเพศเป็นปัญหาสังคมที่ร้ายแรงการศึกษานี้มีวัตถุประสงค์เพื่อศึกษาปัจจัยการเกิดอาชญากรรมทางเพศรวมถึงแนวทางการป้องกันอาชญากรรมทางเพศตามทฤษฎีสามเหลี่ยมอาชญากรรม ด้วยวิธีการสังเคราะห์เอกสารงานวิจัยทั้งหมด 10 เรื่อง ผลการศึกษาพบว่าปัจจัยการเกิดอาชญากรรมทางเพศสามารถแบ่งได้ ดังนี้ 1) ปัจจัยด้านผู้กระทำผิด/คนร้าย ได้แก่ ฐานะทางเศรษฐกิจต่ำ เสพยาเสพติด ได้รับสิ่งยั่วยุจากสื่อ ครอบครัวมีปัญหาหรือแตกแยก บุคลิกภาพเฉพาะ เช่น การควบคุมตนเองต่ำ ขาดความมั่นใจซึมเศร้า เก็บกด มีค่านิยมชายเป็นใหญ่ เป็นผู้ใช้ชีวิตที่เสี่ยงภัย และมีกลุ่มเพื่อนที่ยั่วยุ 2) ปัจจัยด้านเหยื่อ ได้แก่ อยู่ในภาวะเสพของมึนเมารูปร่างหน้าตาดี แต่งกายล่อแหลม ครอบครัวไม่สมบูรณ์แตกแยกห่างเหิน มีช่วงวัยอยู่ในช่วงวัยรุ่นและมีฐานะทางเศรษฐกิจต่ำ 3) ปัจจัยด้านโอกาส ได้แก่ ความใกล้ชิดเนื่องจากเป็นคนรู้จักสนิทสนมกัน เหตุเกิดในเวลากลางคืน–ดึก และอยู่ในสภาพแวดล้อมเสี่ยง พื้นที่เปลี่ยว ลับตาคน

ส่วนข้อเสนอแนะในการป้องกันอาชญากรรมทางเพศแบ่งเป็น 3 ส่วนคือ ส่วนปัจจัยความเสี่ยงด้านโอกาส ได้แก่ หน่วยงานที่เกี่ยวข้องต้องสนับสนุนให้ประชาชนมีความรู้ทักษะการป้องกันตนเอง สถาบันครอบครัว สถาบันการศึกษาต้องให้ความรู้เกี่ยวกับการป้องกันอาชญากรรมทางเพศ การมีมาตรการ นโยบายตรวจตราความปลอดภัยของพื้นที่ให้มากขึ้น การอาศัยการมีส่วนร่วมของชุมชน รวมทั้งบูรณาการการปฏิบัติงานของทุกภาคส่วน ส่วนปัจจัยความเสี่ยงด้านผู้กระทำผิด ได้แก่ สถาบันทางสังคมควรอบรมสั่งสอนขัดเกลาให้บุคคลเป็นคนดี มีความประพฤติเหมาะสม สื่อมวลชนจะต้องมีความรับผิดชอบต่อสังคมในการนำเสนอสื่อต่าง ๆ การควบคุมสื่อ การมีมาตรการหรือกฎหมายควบคุมปัจจัยเสี่ยงต่อการเกิดอาชญากรรม ส่วนปัจจัยความเสี่ยงด้านเหยื่อ ได้แก่ หน่วยงานต่าง ๆ ต้องร่วมรณรงค์ ประชาสัมพันธ์ไม่ให้ประชาชนตกเป็นเหยื่อ ประชาชนเองต้องเรียนรู้วิธีการป้องกันตนเองจากอาชญากรรม การบรรจุความรู้เรื่องอาชญากรรมทางเพศในหลักสูตรการศึกษา และการก าหนดโครงการกิจกรรมควรครอบคลุมการป้องกันแก้ไขปัจจัยเสี่ยงการเกิดอาชญากรรม

สุวิมล สิทธิชาติ (2560) งานวิจัยนี้ได้ศึกษาคุณลักษณะของนักศึกษาคณะวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา เชียงใหม่ ที่มีผลต่อการเรียนแคลคูลัส 1 สำหรับวิศวกร ด้วยเทคนิคเหมืองข้อมูลเพื่อจำแนกนักศึกษาออกเป็น 2 กลุ่ม คือ กลุ่มเสี่ยงที่จะไม่ผ่านรายวิชา และกลุ่มที่ไม่มีความเสี่ยง โดยการวิเคราะห์ข้อมูลได้เก็บรวบรวมข้อมูลคุณลักษณะทั้ง 3 ด้าน รวมถึงปัจจัยส่วนบุคคลอื่น ๆ ของนักศึกษาจำนวน 453 คน แล้วดำเนินตามกระบวนการ CRISP–DM (Cross Industry Standard Process for Data Mining) โดยใช้เทคนิคโครงข่ายประสาทเทียมและต้นไม้ตัดสินใจในการจำแนกประเภทข้อมูลและคัดเลือกคุณลักษณะที่สำคัญด้วย Filter Ranker Method โดยคำนวณค่าน้ำหนักด้วย Chi–Square และ Gain Ratio จากผลการศึกษา พบว่า ความแม่นยำในการจำแนกข้อมูลจาก 50 คุณลักษณะนั้น วิธี ANN มีค่า 71.52% และ Decision Trees – J48 มีค่า 66.23% หลังจากทำการคัดเลือกคุณลักษณะแสดงให้เห็นว่าการจำแนกข้อมูลด้วยวิธี ANN จาก 5 คุณลักษณะแรกที่ได้จากการคัดเลือกโดยวิธี Filter Ranker Method ที่คำนวณค่าน้ำหนักด้วย Chi–Square ทำให้ได้ค่าความถูกต้องมีค่าสูงสุด คือ 80.13% และการจำแนกข้อมูลด้วย Decision Tree ก็ให้ผลไปในทางเดียวกัน โดยมีค่าความถูกต้องสูงสุดที่ 75.83% คุณลักษณะที่ผ่านการคัดเลือกนั้นแสดงถึงคุณลักษณะในด้านพื้นฐานความรู้ของนักศึกษาที่มีมาก่อน จากผลการวิจัยนี้แสดงให้เห็นว่า

เทคนิคเหมืองข้อมูลสามารถนำมาประยุกต์ใช้ทางด้านการศึกษาในการวิเคราะห์ข้อมูลเพื่อแก้ปัญหาเกี่ยวกับการเรียนการสอนได้

พรรณธิภา เพชรบุญมี, ดวงกมล โพธิ์นาคและมนต์ชัย เทียนทอง (2556) งานวิจัยนี้มีวัตถุประสงค์เพื่อพยากรณ์รูปแบบการเรียนรู้ตามประสบการณ์ของเดวิด โคล์ป ตัวแบบที่พัฒนาขึ้นโดยใช้กฎการจำแนกเทคนิคต้นไม้ตัดสินใจ ถูกสร้างและทดสอบตัวแบบด้วยวิธีการตรวจสอบไขว้ (K-fold cross-validation) และวิธีการแบ่งข้อมูลแบบสุ่มด้วยการแบ่งร้อยละ (percentage split) ทำการวิเคราะห์ข้อมูลด้วยโปรแกรม WEKA ผลการทดลองพบว่าได้กฎการจำแนกข้อมูลทั้งสิ้นจำนวน 8 กฎ เมื่อวัดค่าประสิทธิภาพของตัวแบบ การพัฒนาตัวแบบด้วยวิธีการแบ่งข้อมูลแบบสุ่มด้วยการแบ่งร้อยละ 66.00 จะมีค่าความถูกต้องเท่ากับร้อยละ 85 ค่าความแม่นยำเท่ากับร้อยละ 88.20 ค่าความระลึกเท่ากับร้อยละ 85.00 และค่าความถ่วงดุลเท่ากับร้อยละ 85.10 ซึ่งมีประสิทธิภาพทุกค่าสูงกว่าวิธีการอื่น โดยงานวิจัยนี้สามารถนำไปใช้ในการพัฒนาตัวแบบการพยากรณ์รูปแบบการเรียนรู้ตามประสบการณ์ของเดวิด โคล์ป โดยใช้กฎการจำแนกเทคนิคต้นไม้ตัดสินใจที่มีความถูกต้องและแม่นยำในการทำนายรูปแบบการเรียนรู้ได้เป็นอย่างดี

อนัตต์ชัย ชุติภาสเจริญและจรัญ แสนราช (2561) การวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของอัลกอริทึมในการทานายและคุณลักษณะที่มีต่อโอกาสความสำเร็จในการโอนเงินข้ามประเทศของบุคคลทั่วไปโดยทาการศึกษาข้อมูลการโอนเงินของบุคคลทั่วไปจำนวน 51,901ระเบียนทาการเก็บข้อมูลตั้งแต่ปี2559-2560 โดยใช้เทคนิคการจำแนกข้อมูลทั้งหมด3เทคนิคได้แก่เทคนิคต้นไม้ตัดสินใจ เทคนิคนาอีฟเบย์ และเทคนิคการค้นหาเพื่อนบ้านใกล้สุด ซึ่งท าการเปรียบเทียบประสิทธิภาพรูปแบบเทคนิคการทานายระหว่างการใช้คุณลักษณะทั้งหมดการทดสอบประสิทธิภาพตัวแบบทานายด้วยวิธีการCross Validation โดยใช้โปรแกรม RapidMiner Studio 8จากนั้นทาการทดลองเพื่อหาผลการทดสอบประสิทธิภาพที่มีค่าความถูกต้องที่สูงที่สุดผลการศึกษาพบว่าการใช้เทคนิคต้นไม้ตัดสินใจด้วยการเลือกคุณลักษณะทั้งหมดมีค่าความถูกต้องเท่ากับ99.90% เทคนิคการค้นหาเพื่อนบ้านใกล้สุดมีค่าความถูกต้องเท่ากับ99.55%และเทคนิคนาอีฟเบย์มีค่าความถูกต้องเท่ากับ96.71% จากผลการเปรียบเทียบประสิทธิภาพในครั้งนี้สามารถนาเทคนิคต้นไม้ตัดสินใจที่มีค่าความถูกต้องสูงสุดไปใช้ในการพยากรณ์โอกาสความสำเร็จในการโอนเงินข้ามประเทศของบุคคลทั่วไปต่อไป

อิทธิพล ดวงแก้วและสายยัญ สายยศ (2562) การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อวิเคราะห์ปัจจัยที่มีผลต่อพัฒนาการล่าช้าในแต่ละกลุ่มอายุของเด็กวัยก่อนเรียน 4 คนคือ 9, 18, 30 และ 42 เดือนโดยประเมินจากคู่มือการเฝ้าระวังและส่งเสริมการขาย การพัฒนาเด็กปฐมวัย (DSPM) โดยขอข้อมูลสนับสนุนจากสำนักงานสาธารณสุขจังหวัดขอนแก่น ข้อมูลเริ่มตั้งแต่ตุลาคม 2559 ถึงกันยายน 2560 นักวิจัยศึกษาปัจจัยที่มีผลต่อพัฒนาการในวัยเด็กจากการวิจัยทางการแพทย์และสาธารณสุขก่อนหน้านี้ รับ 24 ปัจจัย แบบจำลองนี้สร้างขึ้นโดยใช้เทคนิคการขุดข้อมูลเช่นเทคนิค Decision Tree, K–Neighbour เพื่อนบ้านและเทคนิคNaïve Bayes ทำการทดสอบโดยใช้วิธีการตรวจสอบความถูกต้องข้ามแบบ K–fold Cross validation ผลการวิจัยปรากฏว่าเทคนิคDecision Tree มีความถูกต้องแม่นยำมากที่สุดในการทานายพัฒนาการทั้ง4 ช่วง

เกรียงศักดิ์ รักภักดีและวชิระ โมราชาติ (2561) งานวิจัยครั้งนี้มีวัตถุประสงค์ 1) เพื่อศึกษาความต้องการและพฤติกรรมการตัดสินใจเลือกแหล่งท่องเที่ยวทางธรรมชาติของนักท่องเที่ยว 2) เพื่อพัฒนาฐานข้อมูลแนะนำแหล่งท่องเที่ยวทางธรรมชาติในจังหวัดอุบลราชธานี ตัวอย่างที่ใช้ในการศึกษา คือ นักท่องเที่ยวที่เดินทางมาท่องเที่ยวหรือชาวบ้านที่อาศัยอยู่บริเวณแหล่งท่องเที่ยว จำนวน 1,840 คน จากแหล่งท่องเที่ยว จำนวน 23 แห่ง โดยแบ่งแหล่งท่องเที่ยวตามกิจกรรมการท่องเที่ยวออกเป็น 5 กิจกรรม การพัฒนาฐานข้อมูลแนะนำแหล่งท่องเที่ยวดำเนินการด้วยการใช้เทคนิคการทำเหมืองข้อมูล (Data Mining) ซึ่งเทคนิคที่นำมาใช้ คือ การเรียนรู้ต้นไม้ตัดสินใจ(Decision Tree Learning) และนำเสนอระบบฐานข้อมูลของงานวิจัยในรูปแบบเว็บไซต์ ผลการศึกษาความต้องการและพฤติกรรม พบว่าการได้รับรู้ข้อมูลข่าวสารของแหล่งท่องเที่ยวจากแอปพลิเคชั่นแนะนำแหล่งท่องเที่ยว มีความสำคัญต่อการตัดสินใจที่จะเดินทางมาท่องเที่ยวเป็นอย่างมาก และผลการประเมินประสิทธิภาพของระบบฐานข้อมูลแนะนำแหล่งท่องเที่ยวในภาพรวมอยู่ในระดับดี

2.5 บทสรุป

จากแนวคิด ทฤษฎี เครื่องมือและวรรณกรรมที่เกี่ยวข้องที่ได้กล่าวมาในข้างต้นนั้น ผู้วิเคราะห์ข้อมูลได้เลือกใช้ขั้นตอนกระบวนการวิเคราะห์ทำเหมืองข้อมูลแบบ Classification ด้วยการสร้างโมเดล Decision Tree โดยใช้โปรแกรม RapidMiner Studio 9.6 ในการสร้างโมเดล Decision Tree เพื่อทำการเปรียบเทียบ จากนั้นนำข้อมูลสารสนเทศมาทำการแสดงผลแบบ visualization ในรูปแบบของภาพโดยใช้โปรแกรม Tableau Public เผยแพร่บน Web Browser

บทที่ 3
วิธีการดำเนินงานโครงงาน

โครงงานเรื่อง การวิเคราะห์ข้อมูลรายงานการฆาตกรรมปี 2011–2013 เพื่อใช้สำหรับเผยแพร่ข้อมูลบนเว็บไซต์ ซึ่งในบทนี้จะเป็นการวิเคราะห์ข้อมูลเทคนิค Data Mining เมื่อเสร็จสิ้นกระบวนการวิเคราะห์แล้วจะเป็นขั้นตอนการออกแบบเว็บไซต์ รูปแบบการแสดงผลและบทสรุปจากวิธีการดำเนินงาน ดังนี้

3.1 การวิเคราะห์ข้อมูลด้วย CRISP–DM

3.2 การออกแบบเว็บไซต์

3.3 บทสรุป

3.1 การวิเคราะห์ข้อมูลด้วย CRISP–DM

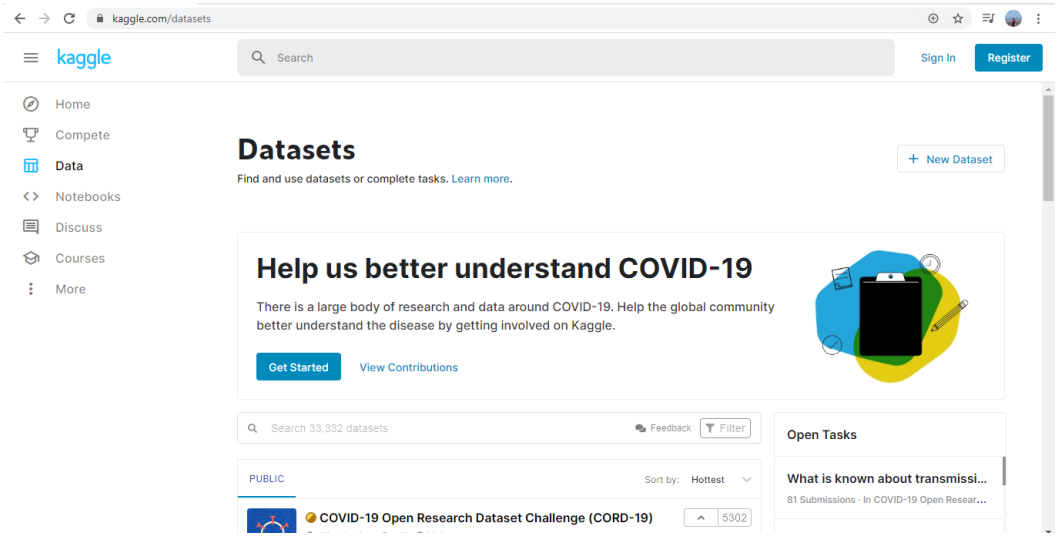
งานวิทยาการด้านข้อมูล (Data Science) ซึ่งกำลังมีบทบาทในยุคปัจจุบัน และทวีความสำคัญยิ่งขึ้นในอนาคต ได้มี CRISP–DM เป็นกระบวนการหลักในการจัดทำเหมืองข้อมูลเพื่อการวิเคราะห์และใช้ประโยชน์ในทางธุรกิจ กระบวนการวิเคราะห์ข้อมูล ด้วย CRISP–DM หรือ Cross Industry Standard Process for Data Mining พัฒนาขึ้นในปี ค.ศ. 1996 โดยความร่วมมือของ 3 บริษัทคือ Daimler Chrysler, SPSS และ NCR ประกอบด้วย 6 ขั้นตอนหลัก ได้แก่

3.1.1 รู้จักและเข้าใจในธุรกิจ (Business understanding) เป็นขั้นตอนแรกของกระบวนการที่มุ่งเน้นไปที่การทำความเข้าใจกระบวนการทางธุรกิจโดยรวม

ผู้วิเคราะห์ข้อมูลทำความเข้าใจกับปัญหาให้อยู่ในรูปของการวิเคราะห์ข้อมูลทางดาต้าไมน์นิ่งโดยการวิเคราะห์ข้อมูลรายงานการฆาตกรรมปี 2011–2013 ซึ่งมีข้อมูลทั้งหมด 638,455 รายการ ซึ่งเป็นข้อมูลที่มีจำนวนมากและทำให้ซับซ้อน

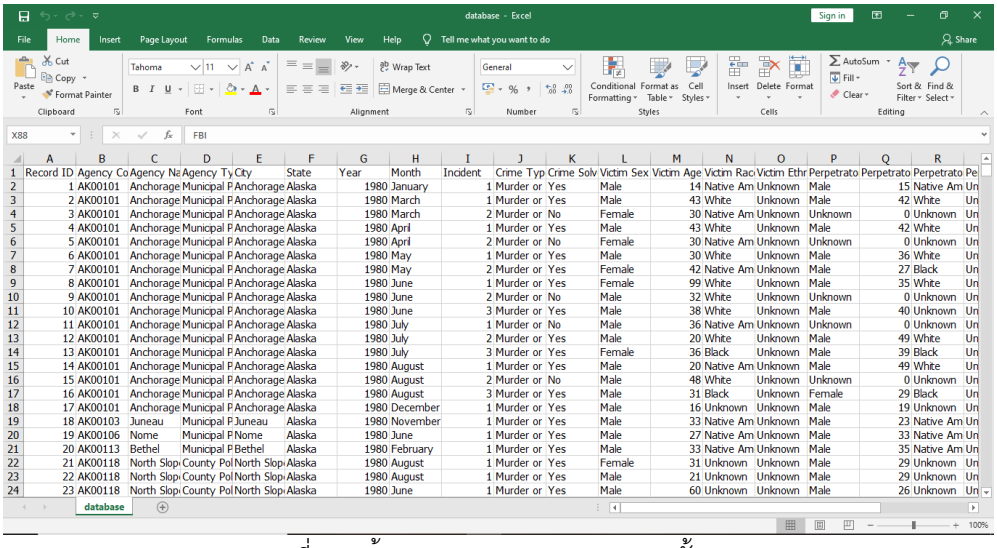
3.1.2 สร้างฐานข้อมูลให้ครบ (Data understanding) ขั้นตอนการจัดเก็บและรวบรวมข้อมูล ตลอดจนการพิจารณาตรวจสอบความถูกต้องของข้อมูลที่ได้รับ โดยเลือกว่าจะใช้ข้อมูลทั้งหมดหรือบางส่วนในการวิเคราะห์ให้สอดรับกับวัตถุประสงค์ที่กำหนดไว้

ผู้วิเคราะห์ข้อมูลได้ทำการเก็บรวบรวมข้อมูลจากเว็บไซต์ kaggle.com จากนั้นจะทำการตรวจสอบข้อมูลที่ได้ทำการรวบรวมมาได้ เพื่อดูความถูกต้องของข้อมูล และพิจารณาว่าข้อมูลอาชญากรรม จะใช้ข้อมูลทั้งหมดหรือจำเป็นต้องเลือกข้อมูลบางส่วนมาใช้ในการวิเคราะห์



ภาพที่ 3.1 เว็บไซต์ kaggle.com

ซึ่งข้อมูลอาชญากรรม มีจำนวน 638,455 รายการ รายการ ประกอบด้วย 24 แอตทริบิวท์ ข้อมูลหลัก ๆ จะประกอบด้วย ประเภทตัวแทน ปี ประเภทอาชญากรรม เพศเหยื่อ เพศผู้กระทำ อาวุธ เป็นต้น

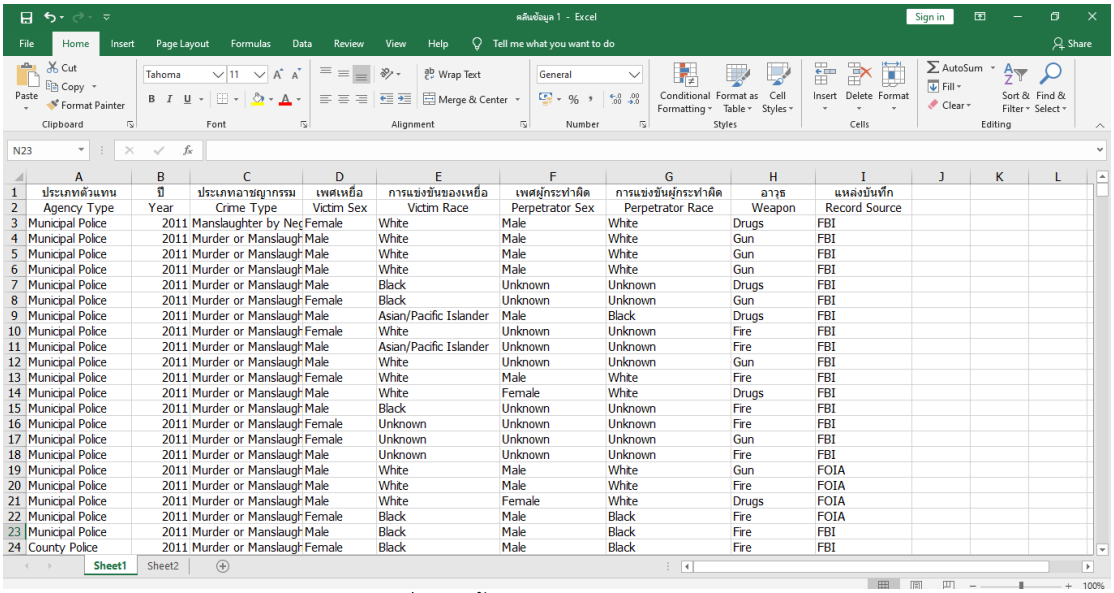


ภาพที่ 3.2 ข้อมูลรายงานการฆาตกรรมทั้งหมด

3.1.3 เตรียมข้อมูลให้พร้อมใช้ (Data preparation) ขั้นตอนการแปลงข้อมูลที่ได้รวบรวมมาและเลือกไว้ ให้อยู่ในรูปแบบที่พร้อมสำหรับนำไปวิเคราะห์ในขั้นตอนต่อไปได้ โดยการทำให้เป็นข้อมูลที่ถูกต้อง (Data cleaning) มักใช้เวลาค่อนข้างมาก

3.1.3.1 ทำการคัดเลือกข้อมูล (Data Selection) คือการคัดเลือกข้อมูลที่เหมาะสมเพื่อนำมาใช้ในการวิเคราะห์ข้อมูล

ผู้วิเคราะห์ข้อมูลทำการคัดเลือกข้อมูลโดยมี ส.ต.ต. วโรดม ไชยชนะ มาช่วยในการพิจารณา และทำการ Data Cleaning ข้อมูลรายงานการฆาตกรรมปี 2011–2013 โดยแยกข้อมูลออกและตัดส่วนที่ไม่จำเป็นออกให้เหลือเฉพาะข้อมูลที่จำเป็นในการวิเคราะห์ในภาพรวม จำนวน 9 แอตทริบิวท์ ได้แก่ ประเภทตัวแทน ปี ประเภทอาชญากรรม เพศเหยื่อ การแข่งขันของเหยื่อ เพศผู้กระทำผิด การแข่งขันผู้กระทำผิด อาวุธ แหล่งบันทึก จำนวน 640 รายการ ซึ่งเป็นข้อมูลที่จำเป็นในการนำไปวิเคราะห์ข้อมูล



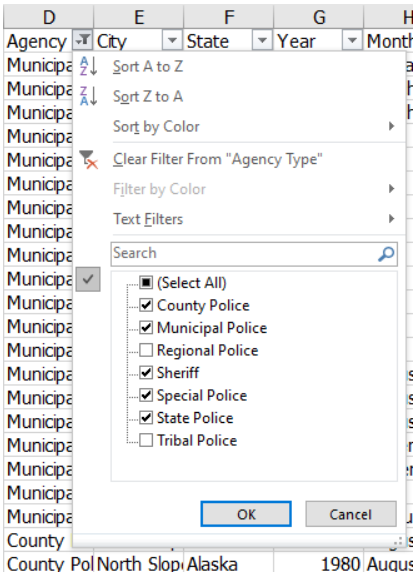
ประเภทตัวแทน	ปี	ประเภทอาชญากรรม	เพศเหยื่อ	การแข่งขันของเหยื่อ	เพศผู้กระทำผิด	การแข่งขันผู้กระทำผิด	อาวุธ	แหล่งบันทึก
Agency Type	Year	Crime Type	Victim Sex	Victim Race	Perpetrator Sex	Perpetrator Race	Weapon	Record Source
Municipal Police	2011	Manslaughter by Ne	Female	White	Male	White	Drugs	FBI
Municipal Police	2011	Murder or Manslaugh	Male	White	Male	White	Gun	FBI
Municipal Police	2011	Murder or Manslaugh	Male	White	Male	White	Gun	FBI
Municipal Police	2011	Murder or Manslaugh	Male	White	Male	White	Gun	FBI
Municipal Police	2011	Murder or Manslaugh	Male	Black	Unknown	Unknown	Drugs	FBI
Municipal Police	2011	Murder or Manslaugh	Female	Black	Unknown	Unknown	Gun	FBI
Municipal Police	2011	Murder or Manslaugh	Male	Asian/Pacific Islander	Male	Black	Drugs	FBI
Municipal Police	2011	Murder or Manslaugh	Female	White	Unknown	Unknown	Fire	FBI
Municipal Police	2011	Murder or Manslaugh	Male	Asian/Pacific Islander	Unknown	Unknown	Fire	FBI
Municipal Police	2011	Murder or Manslaugh	Male	White	Unknown	Unknown	Gun	FBI
Municipal Police	2011	Murder or Manslaugh	Female	White	Male	White	Fire	FBI
Municipal Police	2011	Murder or Manslaugh	Male	White	Female	White	Drugs	FBI
Municipal Police	2011	Murder or Manslaugh	Male	Black	Unknown	Unknown	Fire	FBI
Municipal Police	2011	Murder or Manslaugh	Female	Unknown	Unknown	Unknown	Fire	FBI
Municipal Police	2011	Murder or Manslaugh	Female	Unknown	Unknown	Unknown	Gun	FBI
Municipal Police	2011	Murder or Manslaugh	Male	Unknown	Unknown	Unknown	Fire	FBI
Municipal Police	2011	Murder or Manslaugh	Male	White	Male	White	Gun	FOIA
Municipal Police	2011	Murder or Manslaugh	Male	White	Male	White	Fire	FOIA
Municipal Police	2011	Murder or Manslaugh	Male	White	Female	White	Drugs	FOIA
Municipal Police	2011	Murder or Manslaugh	Female	Black	Male	Black	Fire	FOIA
Municipal Police	2011	Murder or Manslaugh	Male	Black	Male	Black	Fire	FBI
County Police	2011	Murder or Manslaugh	Female	Black	Male	Black	Fire	FBI

ภาพที่ 3.3 ข้อมูลรายงานการฆาตกรรม

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

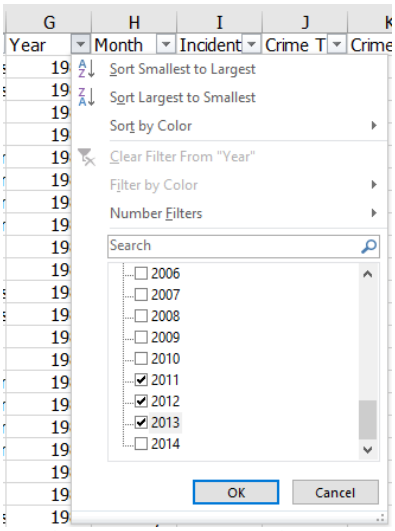
1) ข้อมูลรายงานการฆาตกรรม ผู้วิเคราะห์ข้อมูลได้ทำการแก้ไขและลบข้อมูล ซึ่งผู้วิเคราะห์ข้อมูลพบว่า ข้อมูลทั้งหมดนั้นมีจำนวนที่เยอะจึงเลือกส่วนหัวข้อย่อยในแต่ละแอตทริบิวท์มาเท่านั้น ดังนั้นผู้วิเคราะห์ข้อมูลได้ทำลบข้อมูล ดังนี้

- Agency Type(ประเภทตัวแทน) ซึ่งมี 7 หัวข้อย่อย ผู้วิเคราะห์พบว่าควรตัด Regional Police(ตำรวจท้องที่)และ Tribal Police(ตำรวจเผ่า) 2 หัวข้อย่อยเพราะ 2 หัวข้อย่อยนี้มีจำนวนที่น้อย 37 รายการ จึงทำการลบทิ้ง ดังนี้



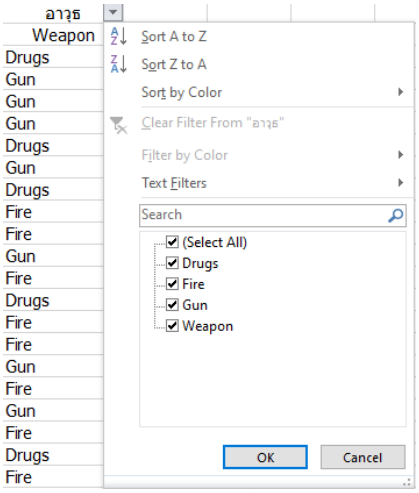
ภาพที่ 3.4 การลบรายการย่อย Agency Type(ประเภทตัวแทน)

- Year(ปี) ข้อมูลทั้งหมดมีตั้งแต่ปี 1980–2014 ผู้วิเคราะห์พบว่าคัดเลือกออกมาแสดงผล 3 ปี คือปี 2011–2013 ดังนี้



ภาพที่ 3.5 การลบรายการย่อย Year(ปี)

- Weapon(อาวุธ) มีรายการอาวุธ 14 รายการ ซึ่งมีจำนวนอาวุธ 44,199 รายการ จึงทำการเลือกอาวุธที่เป็นรายการหลัก 3 รายการ คือ Drugs(ยาเสพติด) Fire(ไฟ) Gun(ปืน) ดังนี้



ภาพที่ 3.6 การลบรายการย่อย Weapon(อาวุธ)

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

ผู้วิเคราะห์ข้อมูลวิเคราะห์ข้อมูลด้วยเทคนิคทางดาต้าไมน์นิง แบบการจำแนกประเภทข้อมูล (Classification) โดยการใช้โมเดลการตัดสินใจแบบต้นไม้ (Decision Tree) ซึ่งในขั้นตอนนี้จะถูกนำมาใช้เพื่อให้ได้คำตอบที่ดีที่สุด โดยใช้โปรแกรมที่ใช้ทำเหมืองข้อมูล ด้วยชุดข้อมูลที่คัดเลือก ดังนี้

	A	B	C	D	E	F
2	Agency Type	Year	Crime Type	Victim Sex	Perpetrator Sex	Weapon
3	Municipal Police	2011	Manslaughter by Ne	Female	Male	Drugs
4	Municipal Police	2011	Murder or Manslaugh	Male	Male	Gun
5	Municipal Police	2011	Murder or Manslaugh	Male	Male	Gun
6	Municipal Police	2011	Murder or Manslaugh	Male	Male	Gun
7	Municipal Police	2011	Murder or Manslaugh	Male	Unknown	Drugs
8	Municipal Police	2011	Murder or Manslaugh	Female	Unknown	Gun
9	Municipal Police	2011	Murder or Manslaugh	Male	Male	Drugs
10	Municipal Police	2011	Murder or Manslaugh	Female	Unknown	Fire
11	Municipal Police	2011	Murder or Manslaugh	Male	Unknown	Fire
12	Municipal Police	2011	Murder or Manslaugh	Male	Unknown	Gun
13	Municipal Police	2011	Murder or Manslaugh	Female	Male	Fire
14	Municipal Police	2011	Murder or Manslaugh	Male	Female	Drugs
15	Municipal Police	2011	Murder or Manslaugh	Male	Unknown	Fire
16	Municipal Police	2011	Murder or Manslaugh	Female	Unknown	Fire
17	Municipal Police	2011	Murder or Manslaugh	Female	Unknown	Gun
18	Municipal Police	2011	Murder or Manslaugh	Male	Unknown	Fire
19	Municipal Police	2011	Murder or Manslaugh	Male	Male	Gun
20	Municipal Police	2011	Murder or Manslaugh	Male	Male	Fire
21	Municipal Police	2011	Murder or Manslaugh	Male	Female	Drugs
22	Municipal Police	2011	Murder or Manslaugh	Female	Male	Fire
23	Municipal Police	2011	Murder or Manslaugh	Male	Male	Fire
24	County Police	2011	Murder or Manslaugh	Female	Male	Fire
25	Municipal Police	2011	Murder or Manslaugh	Male	Male	Gun

Sheet1

ภาพที่ 3.7 ชุดข้อมูลที่คัดเลือกมาวิเคราะห์ข้อมูล

จากรูปภาพที่ ประกอบด้วย 6 แอตทริบิวต์ คือ

- Agency Type(ประเภทตัวแทน) ประกอบด้วย 4 ค่า คือ County Police, Municipal Police, Special Police, Special Police
- Year(ปี) ประกอบด้วย 3 ค่า คือ ปี2011 ปี2012 ปี2013
- Crime Type(ประเภทอาชญากรรม) ประกอบด้วย 2 ค่า คือ Murder or Manslaughter, Manslaughter by Negligence
- Victim Sex(เพศเหยื่อ) ประกอบด้วย 2 ค่า คือ Male, Female
- Perpetrator Sex(เพศผู้กระทำผิด) ประกอบด้วย 3 ค่า คือ Male, Female, Unknown
- Weapon(อาวุธ) ประกอบด้วย 3 ค่า คือ Drugs, Fire, Gun

การสร้างโมเดล Decision Tree จะทำการคัดเลือกแอตทริบิวต์ที่มีความสัมพันธ์กับคลาสมากที่สุดขึ้นมาเป็นโหนดบนสุดของ Tree (root node) หลังจากนั้นก็จะหาแอตทริบิวต์ถัดไปเรื่อย ๆ ในการหาความสัมพันธ์ของแอตทริบิวต์นี้จะใช้ตัววัด ที่เรียกว่า Information Gain (IG) ค่านี้คำนวณได้จากสมการดังนี้

IG (parent, child) = entropy(parent) – [p(c1) × entropy(c1) + p(c2) × entropy(c2) + ...]

โดยที่ entropy(c1) = –p(c1) log p(c1) และ p(c2) คือ ค่าความน่าจะเป็นของ c1

การคำนวณค่าแต่ละแอตทริบิวต์เทียบกับคลาสเพื่อหาแอตทริบิวต์ที่มีค่า IG มากที่สุดมาเป็น Root ของ Decision tree กับจำนวนข้อมูลทั้งหมดโดยใช้ผลลัพธ์เป็น Perpetrator Sex(เพศผู้กระทำผิด) ชาย และหญิง ดังนี้

1) คำนวณค่า IG ของแอตทริบิวต์ Agency Type(ตัวแทนหน่วยงาน) จากข้อมูล สามารถคำนวณค่า IG ได้ดังนี้

entropy(parent) = $-p(\text{Male}) \times \log_2 p(\text{Male}) + p(\text{Female}) \times \log_2 p(\text{Female}) + p(\text{Unknown}) \times \log_2 p(\text{Unknown})$

$= -[0.53 \times \log_2(0.53) + 0.19 \times \log_2(0.19) + 0.28 \times \log_2(0.28)]$

$= -[0.53 \times -0.92 + 0.19 \times -2.40 + 0.28 \times -1.84]$

$= -[0.49 + 0.46 + 0.52]$

$= 1.47$

entropy(County Police) = $p(\text{Male}) \times \log_2 p(\text{Male}) + p(\text{Female}) \times \log_2 p(\text{Female}) + p(\text{Unknown}) \times \log_2 p(\text{Unknown})$

$= -[0.76 \times \log_2(0.76) + 0.08 \times \log_2(0.08) + 0.16 \times \log_2(0.16)]$

$= -[0.76 \times -0.40 + 0.08 \times -0.64 + 0.16 \times -2.64]$

$= -[0.30 + 0.05 + 0.42]$

$= 0.77$

entropy(Municipal Police) = $p(\text{Male}) \times \log_2 p(\text{Male}) + p(\text{Female}) \times \log_2 p(\text{Female}) + p(\text{Unknown}) \times \log_2 p(\text{Unknown})$

$= -[0.52 \times \log_2(0.52) + 0.19 \times \log_2(0.19) + 0.30 \times \log_2(0.30)]$

$= -[0.52 \times -0.94 + 0.19 \times -2.40 + 0.30 \times -1.74]$

$= -[0.49 + 0.46 + 0.52]$

$= 1.47$

entropy(Special Police) = p(Male) × $\log_2$ p(Male) + p(Female) × $\log_2$ p(Female) + p(Unknown) × $\log_2$ p(Unknown)

= −[0.25 × $\log_2$(0.25) + 0.25 × $\log_2$(0.25) + 0.50 × $\log_2$(0.50)]

= −[0.25 × −2 + 0.25 × −2 + 0.50 × −1]

= −[0.50 + 0.50 + 0.50]

= 1.50

entropy(State Police) = p(Male) × $\log_2$ p(Male) + p(Female) × $\log_2$ p(Female) + p(Unknown) × $\log_2$ p(Unknown)

= −[0.69 × $\log_2$(0.69) + 0.25 × $\log_2$(0.25) + 0.06 × $\log_2$(0.06)]

= −[0.69 × −0.53 + 0.25 × −2 + 0.06 × −4.06]

= −[0.37 + 0.50 + 0.24]

= 1.11

IG (parent, child) = entropy(parent) – [p(County Police) × entropy(County Police) + p(Municipal Police) × entropy(Municipal Police) + p(Special Police) × entropy(Special Police) + p(State Police) × entropy(State Police)]

= 1.47 – [0.04 × 0.77 + 0.90 × 1.47 + 0.01 × 1.50 + 0.06 × 1.11]

= 1.47 – [0.03 + 1.32 + 0.02 + 0.07]

= 1.47 – 1.44

= 0.03

2) คำนวณค่า IG ของแอตทริบิวต์ Year(ปี) จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy(parent) = $-p(\text{Male}) \times \log_2 p(\text{Male}) + p(\text{Female}) \times \log_2 p(\text{Female}) + p(\text{Unknown}) \times \log_2 p(\text{Unknown})$

$= -[0.53 \times \log_2(0.53) + 0.19 \times \log_2(0.19) + 0.28 \times \log_2(0.28)]$

$= -[0.53 \times -0.92 + 0.19 \times -2.40 + 0.28 \times -1.84]$

$= -[0.49 + 0.46 + 0.52]$

$= 1.47$

entropy(2011) = $-p(\text{Male}) \times \log_2 p(\text{Male}) + p(\text{Female}) \times \log_2 p(\text{Female}) + p(\text{Unknown}) \times \log_2 p(\text{Unknown})$

$= -[0.55 \times \log_2(0.55) + 0.15 \times \log_2(0.15) + 0.30 \times \log_2(0.30)]$

$= -[0.55 \times -0.86 + 0.15 \times -2.74 + 0.30 \times -1.74]$

$= -[0.47 + 0.41 + 0.52]$

$= 1.40$

entropy(2012) = $-p(\text{Male}) \times \log_2 p(\text{Male}) + p(\text{Female}) \times \log_2 p(\text{Female}) + p(\text{Unknown}) \times \log_2 p(\text{Unknown})$

$= -[0.53 \times \log_2(0.53) + 0.20 \times \log_2(0.20) + 0.28 \times \log_2(0.28)]$

$= -[0.53 \times -0.92 + 0.20 \times -2.32 + 0.28 \times -1.84]$

$= -[0.49 + 0.46 + 0.52]$

$= 1.47$

entropy(2013) = $-p(\text{Male}) \times \log_2 p(\text{Male}) + p(\text{Female}) \times \log_2 p(\text{Female}) + p(\text{Unknown}) \times \log_2 p(\text{Unknown})$

$= -[0.53 \times \log_2(0.53) + 0.20 \times \log_2(0.20) + 0.27 \times \log_2(0.27)]$

$= -[0.53 \times 0.92 + 0.20 \times -2.32 + 0.28 \times -1.89]$

$= -[0.49 + 0.46 + 0.53]$

$= 1.48$

IG (parent, child) = entropy(parent) – [p(2011) × entropy(2011) + p(2012) × entropy(2012) + p(2013) × entropy(2013)]

$= 1.47 - [0.28 \times 1.40 + 0.34 \times 1.47 + 0.38 \times 1.48]$

$= 1.47 - [0.39 + 0.50 + 0.56]$

$= 1.47 - 1.45$

$= 0.02$

3) คำนวณค่า IG ของแอตทริบิวต์ Crime Type(ประเภทอาชญากรรม) จากข้อมูล สามารถคำนวณค่า IG ได้ดังนี้

entropy(parent) = $-p(\text{Male}) \times \log_2 p(\text{Male}) + p(\text{Female}) \times \log_2 p(\text{Female}) + p(\text{Unknown}) \times \log_2 p(\text{Unknown})$

$= -[0.53 \times \log_2(0.53) + 0.19 \times \log_2(0.19) + 0.28 \times \log_2(0.28)]$

$= -[0.53 \times -0.92 + 0.19 \times -2.40 + 0.28 \times -1.84]$

$= -[0.49 + 0.46 + 0.52]$

$= 1.47$

entropy(Murder or Manslaughter) = $-p(\text{Male}) \times \log_2 p(\text{Male}) + p(\text{Female}) \times \log_2 p(\text{Female}) + p(\text{Unknown}) \times \log_2 p(\text{Unknown})$

$= -[0.54 \times \log_2(0.54) + 0.16 \times \log_2(0.16) + 0.30 \times \log_2(0.30)]$

$= -[0.54 \times -0.89 + 0.16 \times -2.64 + 0.30 \times -1.74]$

$= -[0.48 + 0.42 + 0.52]$

$= 1.42$

entropy(Manslaughter by Negligence)= $-p(\text{Male}) \times \log_2 p(\text{Male}) + p(\text{Female}) \times \log_2 p(\text{Female}) + p(\text{Unknown}) \times \log_2 p(\text{Unknown})$

$= -[0.45 \times \log_2(0.45) + 0.51 \times \log_2(0.51) + 0.04 \times \log_2(0.04)]$

$= -[0.45 \times -1.15 + 0.51 \times -0.97 + 0.04 \times -4.64]$

$= -[0.52 + 0.49 + 0.19]$

$= 1.20$

IG (parent, child) = entropy(parent) – [p(Murder or Manslaughter) × entropy(Murder or Manslaughter) + p(Manslaughter by Negligence) × entropy(Manslaughter by Negligence)]

$= 1.47 - [0.91 \times 1.42 + 0.09 \times 1.20]$

$= 1.47 - [1.29 + 0.11]$

$= 1.47 - 1.40$

$= 0.07$

4) คำนวณค่า IG ของแอตทริบิวต์ Victim Sex(เพศของเหยื่อ) จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy(parent) = $-p(\text{Male}) \times \log_2 p(\text{Male}) + p(\text{Female}) \times \log_2 p(\text{Female}) + p(\text{Unknown}) \times \log_2 p(\text{Unknown})$

$= -[0.53 \times \log_2(0.53) + 0.19 \times \log_2(0.19) + 0.28 \times \log_2(0.28)]$

$= -[0.53 \times -0.92 + 0.19 \times -2.40 + 0.28 \times -1.84]$

$= -[0.49 + 0.46 + 0.52]$

$= 1.47$

entropy(Male) = $-p(\text{Male}) \times \log_2 p(\text{Male}) + p(\text{Female}) \times \log_2 p(\text{Female}) + p(\text{Unknown}) \times \log_2 p(\text{Unknown})$

$= -[0.51 \times \log_2(0.51) + 0.17 \times \log_2(0.17) + 0.32 \times \log_2(0.32)]$

$= -[0.51 \times -0.97 + 0.17 \times -2.56 + 0.32 \times -1.64]$

$= -[0.49 + 0.44 + 0.52]$

$= 1.45$

entropy(Female) = $-p(\text{Male}) \times \log_2 p(\text{Male}) + p(\text{Female}) \times \log_2 p(\text{Female}) + p(\text{Unknown}) \times \log_2 p(\text{Unknown})$

$= -[0.59 \times \log_2(0.59) + 0.22 \times \log_2(0.22) + 0.19 \times \log_2(0.19)]$

$= -[0.59 \times -0.76 + 0.22 \times -2.18 + 0.19 \times -2.40]$

$= -[0.45 + 0.48 + 0.46]$

$= 1.39$

IG(parent, child) = entropy(parent) – [p(Male) × entropy(Male) + p(Female) × entropy(Female)]

= 1.47 – [0.71 × 1.45 + 0.29 × 1.39]

= 1.47 – [1.03 + 0.40]

= 1.47 – 1.43

= 0.04

5) คำนวณค่า IG ของแอตทริบิวต์ Weapon(อาวุธ) จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy(parent) = –p(Male) × $\log_2$ p(Male) + p(Female) × $\log_2$ p(Female) + p(Unknown) × $\log_2$ p(Unknown)

= –[0.53 × $\log_2$(0.53) + 0.19 × $\log_2$(0.19) + 0.28 × $\log_2$(0.28)]

= –[0.53 × –0.92 + 0.19 × –2.40 + 0.28 × –1.84]

= –[0.49 + 0.46 + 0.52]

= 1.47

entropy(Drugs) = –p(Male) × $\log_2$ p(Male) + p(Female) × $\log_2$ p(Female) + p(Unknown) × $\log_2$ p(Unknown)

= –[0.52 × $\log_2$(0.52) + 0.29 × $\log_2$(0.29) + 0.19 × $\log_2$(0.19)]

= –[0.52 × –0.94 + 0.29 × –1.79 + 0.19 × –2.40]

= –[0.49 + 0.52 + 0.46]

= 1.47

entropy(Fire) = $-p(\text{Male}) \times \log_2 p(\text{Male}) + p(\text{Female}) \times \log_2 p(\text{Female}) + p(\text{Unknown}) \times \log_2 p(\text{Unknown})$

$= -[0.51 \times \log_2(0.51) + 0.20 \times \log_2(0.20) + 0.29 \times \log_2(0.29)]$

$= -[0.51 \times -0.97 + 0.20 \times -2.32 + 0.29 \times -2.06]$

$= -[0.49 + 0.46 + 0.60]$

$= 1.55$

entropy(Gun) = $-p(\text{Male}) \times \log_2 p(\text{Male}) + p(\text{Female}) \times \log_2 p(\text{Female}) + p(\text{Unknown}) \times \log_2 p(\text{Unknown})$

$= -[0.56 \times \log_2(0.56) + 0.10 \times \log_2(0.10) + 0.34 \times \log_2(0.34)]$

$= -[0.56 \times -0.84 + 0.10 \times -3.32 + 0.34 \times -1.56]$

$= -[0.47 + 0.33 + 0.53]$

$= 1.33$

IG(parent, child) = entropy(parent) – [p(Drugs) × entropy(Drugs) + p(Fire) × entropy(Fire) + p(Gun) × entropy(Gun)]

$= 1.47 - [0.29 \times 1.47 + 0.33 \times 1.55 + 0.39 \times 1.33]$

$= 1.47 - [0.43 + 0.51 + 0.52]$

$= 1.47 - 1.46$

$= 0.01$

จากการคำนวณค่า IG ของทุกแอตทริบิวต์พบว่าค่า IG ของแอตทริบิวต์ Crime Type(ประเภทอาชญากรรม) มีค่ามากที่สุด (0.07) ดังนั้นจึงเลือกแอตทริบิวต์ Crime Type ขึ้นมาเป็นโหนด Root และจะต้องทำการแตกกิ่งจาก โหนด root ออกไปจนข้อมูลในแต่ละโหนดมีคลาสคำตอบเดียวกัน จึงทำการสร้างโหนดในระดับถัดไปของแอตทริบิวต์ Crime Type

การคำนวณค่าแต่ละแอตทริบิวต์ในระดับที่ 2 กับจำนวนข้อมูลทั้งหมดโดยใช้ผลลัพธ์เป็น Perpetrator Sex(เพศผู้กระทำผิด)

1) คำนวณค่า IG ของแอตทริบิวต์ Crime Type(ประเภทอาชญากรรม) และ Agency Type(ตัวแทนหน่วยงาน) ของ Murder or Manslaughter จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = $-p(\text{Male}) \times \log_2 p(\text{Male}) + p(\text{Female}) \times \log_2 p(\text{Female}) + p(\text{Unknown}) \times \log_2 p(\text{Unknown})$

$= -[0.54 \times \log_2(0.54) + 0.16 \times \log_2(0.16) + 0.30 \times \log_2(0.30)]$

$= -[0.54 \times -0.89 + 0.16 \times -2.64 + 0.30 \times -1.74]$

$= -[0.48 + 0.42 + 0.52]$

$= 1.42$

entropy(County Police) = $p(\text{Male}) \times \log_2 p(\text{Male}) + p(\text{Female}) \times \log_2 p(\text{Female}) + p(\text{Unknown}) \times \log_2 p(\text{Unknown})$

$= -[0.83 \times \log_2(0.83) + 0.00 \times \log_2(0.00) + 0.17 \times \log_2(0.17)]$

$= -[0.83 \times -0.27 + 0.00 \times -0 + 0.17 \times -2.56]$

$= -[0.22 + 0 + 0.44]$

$= 0.66$

entropy(Municipal Police) = $p(\text{Male}) \times \log_2 p(\text{Male}) + p(\text{Female}) \times \log_2 p(\text{Female}) + p(\text{Unknown}) \times \log_2 p(\text{Unknown})$

$= -[0.52 \times \log_2(0.52) + 0.16 \times \log_2(0.16) + 0.32 \times \log_2(0.32)]$

$= -[0.52 \times -0.94 + 0.16 \times -2.64 + 0.32 \times -1.64]$

$= -[0.49 + 0.42 + 0.52]$

$= 1.43$

entropy(Special Police) = p(Male) × $\log_2$ p(Male) + p(Female) × $\log_2$ p(Female) + p(Unknown) × $\log_2$ p(Unknown)

= −[0.25 × $\log_2$(0.25) + 0.25 × $\log_2$(0.25) + 0.50 × $\log_2$(0.50)]

= −[0.25 × −2 + 0.25 × −2 + 0.50 × −1]

= −[0.50 + 0.50 + 0.50]

= 1.50

entropy(State Police) = p(Male) × $\log_2$ p(Male) + p(Female) × $\log_2$ p(Female) + p(Unknown) × $\log_2$ p(Unknown)

= −[0.74 × $\log_2$(0.74) + 0.21 × $\log_2$(0.21) + 0.06 × $\log_2$(0.06)]

= −[0.74 × −0.43 + 0.21 × −2.25 + 0.06 × −4.06]

= −[0.32 + 0.47 + 0.24]

= 1.03

IG (parent, child) = entropy(parent) – [p(County Police) × entropy(County Police) + p(Municipal Police) × entropy(Municipal Police) + p(Special Police) × entropy(Special Police) + p(State Police) × entropy(State Police)]

= 1.42 – [0.03 × 0.66 + 0.90 × 1.43 + 0.01 × 1.50 + 0.06 × 1.03]

= 1.42 – [0.02 + 1.29 + 0.02 + 0.06

= 1.42 – 1.39

= 0.03

1.1) คำนวณค่า IG ของแอตทริบิวต์ Crime Type(ประเภทอาชญากรรม) และ Agency Type(ตัวแทนหน่วยงาน) ของ Manslaughter by Negligence จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = $-p(\text{Male}) \times \log_2 p(\text{Male}) + p(\text{Female}) \times \log_2 p(\text{Female}) + p(\text{Unknown}) \times \log_2 p(\text{Unknown})$

$= -[0.45 \times \log_2(0.45) + 0.51 \times \log_2(0.51) + 0.04 \times \log_2(0.04)]$

$= -[0.45 \times -1.15 + 0.51 \times -0.97 + 0.04 \times -4.64]$

$= -[0.52 + 0.49 + 0.19]$

$= 1.20$

entropy(County Police) = $p(\text{Male}) \times \log_2 p(\text{Male}) + p(\text{Female}) \times \log_2 p(\text{Female}) + p(\text{Unknown}) \times \log_2 p(\text{Unknown})$

$= -[0.57 \times \log_2(0.57) + 0.29 \times \log_2(0.29) + 0.14 \times \log_2(0.14)]$

$= -[0.57 \times -0.81 + 0.29 \times -1.79 + 0.14 \times -2.84]$

$= -[0.46 + 0.52 + 0.40]$

$= 1.38$

entropy(Municipal Police) = $p(\text{Male}) \times \log_2 p(\text{Male}) + p(\text{Female}) \times \log_2 p(\text{Female}) + p(\text{Unknown}) \times \log_2 p(\text{Unknown})$

$= -[0.46 \times \log_2(0.46) + 0.52 \times \log_2(0.52) + 0.02 \times \log_2(0.02)]$

$= -[0.46 \times -1.12 + 0.52 \times -0.94 + 0.02 \times -5.64]$

$= -[0.52 + 0.49 + 0.11]$

$= 1.12$

entropy(Special Police) = p(Male) × $\log_2$ p(Male) + p(Female) × $\log_2$ p(Female) + p(Unknown) × $\log_2$ p(Unknown)

= −[0.00 × $\log_2$(0.00) + 0.00 × $\log_2$(0.00) + 0.00 × $\log_2$(0.00)]

= −[0.00 × −0 + 0.00 × −0 + 0.00 × −0]

= −[0.0 + 0.0 + 0.0]

= 0

entropy(State Police) = p(Male) × $\log_2$ p(Male) + p(Female) × $\log_2$ p(Female) + p(Unknown) × $\log_2$ p(Unknown)

= −[0.00 × $\log_2$(0.00) + 1.00 × $\log_2$(1.00) + 0.00 × $\log_2$(0.00)]

= −[0.00 × −0.00 + 1.00 × −0 + 0.00 × −0.00]

= −[0 + 0 + 0]

= 0

IG (parent, child) = entropy(parent) – [p(County Police) × entropy(County Police) + p(Municipal Police) × entropy(Municipal Police) + p(Special Police) × entropy(Special Police) + p(State Police) × entropy(State Police)]

= 1.20 – [0.13 × 1.38 + 0.84 × 1.12 + 0 × 0 + 0 × 0]

= 1.20 – [0.18 + 0.94 + 0 + 0]

= 1.20 – 1.12

= 0.08

2) คำนวณค่า IG ของแอตทริบิวต์ Crime Type(ประเภทอาชญากรรม) และ Year(ปี) ของ Murder or Manslaughter จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = $-p(\text{Male}) \times \log_2 p(\text{Male}) + p(\text{Female}) \times \log_2 p(\text{Female}) + p(\text{Unknown}) \times \log_2 p(\text{Unknown})$

$= -[0.54 \times \log_2(0.54) + 0.16 \times \log_2(0.16) + 0.30 \times \log_2(0.30)]$

$= -[0.54 \times -0.89 + 0.16 \times -2.64 + 0.30 \times -1.74]$

$= -[0.48 + 0.42 + 0.52]$

$= 1.42$

entropy(2011) = $-p(\text{Male}) \times \log_2 p(\text{Male}) + p(\text{Female}) \times \log_2 p(\text{Female}) + p(\text{Unknown}) \times \log_2 p(\text{Unknown})$

$= -[0.54 \times \log_2(0.54) + 0.14 \times \log_2(0.14) + 0.31 \times \log_2(0.31)]$

$= -[0.54 \times -0.89 + 0.14 \times -2.84 + 0.31 \times -1.69]$

$= -[0.48 + 0.40 + 0.52]$

$= 1.40$

entropy(2012) = $-p(\text{Male}) \times \log_2 p(\text{Male}) + p(\text{Female}) \times \log_2 p(\text{Female}) + p(\text{Unknown}) \times \log_2 p(\text{Unknown})$

$= -[0.53 \times \log_2(0.53) + 0.16 \times \log_2(0.16) + 0.31 \times \log_2(0.31)]$

$= -[0.53 \times -0.92 + 0.16 \times -2.64 + 0.31 \times -1.69]$

$= -[0.49 + 0.42 + 0.52]$

$= 1.43$

entropy(2013) = $-p(\text{Male}) \times \log_2 p(\text{Male}) + p(\text{Female}) \times \log_2 p(\text{Female}) + p(\text{Unknown}) \times \log_2 p(\text{Unknown})$

$= -[0.55 \times \log_2(0.55) + 0.16 \times \log_2(0.16) + 0.29 \times \log_2(0.29)]$

$= -[0.55 \times -0.86 + 0.16 \times -2.64 + 0.29 \times -1.79]$

$= -[0.47 + 0.42 + 0.52]$

$= 1.41$

IG (parent, child) = entropy(parent) – [p(2011) × entropy(2011) + p(2012) × entropy(2012) + p(2013) × entropy(2013)]

$= 1.42 - [0.29 \times 1.40 + 0.32 \times 1.43 + 0.39 \times 1.41]$

$= 1.42 - [0.41 + 0.46 + 0.55]$

$= 1.42 - 1.42$

$= 0$

2.1) คำนวณค่า IG ของแอตทริบิวต์ Crime Type(ประเภทอาชญากรรม) และ Year(ปี) ของ Manslaughter by Negligence จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = $-p(\text{Male}) \times \log_2 p(\text{Male}) + p(\text{Female}) \times \log_2 p(\text{Female}) + p(\text{Unknown}) \times \log_2 p(\text{Unknown})$

$= -[0.45 \times \log_2(0.45) + 0.51 \times \log_2(0.51) + 0.04 \times \log_2(0.04)]$

$= -[0.45 \times -1.15 + 0.51 \times -0.97 + 0.04 \times -4.64]$

$= -[0.52 + 0.49 + 0.19]$

$= 1.20$

entropy(2011) = $-p(\text{Male}) \times \log_2 p(\text{Male}) + p(\text{Female}) \times \log_2 p(\text{Female}) + p(\text{Unknown}) \times \log_2 p(\text{Unknown})$

$= -[0.70 \times \log_2(0.70) + 0.20 \times \log_2(0.20) + 0.10 \times \log_2(0.10)]$

$= -[0.70 \times -0.51 + 0.20 \times -2.32 + 0.10 \times -3.32]$

$= -[0.36 + 0.46 + 0.33]$

$= 1.15$

entropy(2012) = $-p(\text{Male}) \times \log_2 p(\text{Male}) + p(\text{Female}) \times \log_2 p(\text{Female}) + p(\text{Unknown}) \times \log_2 p(\text{Unknown})$

$= -[0.54 \times \log_2(0.54) + 0.43 \times \log_2(0.43) + 0.04 \times \log_2(0.04)]$

$= -[0.54 \times -0.89 + 0.43 \times -1.22 + 0.04 \times -4.64]$

$= -[0.48 + 0.52 + 0.19]$

$= 1.19$

entropy(2013) = $-p(\text{Male}) \times \log_2 p(\text{Male}) + p(\text{Female}) \times \log_2 p(\text{Female}) + p(\text{Unknown}) \times \log_2 p(\text{Unknown})$

$= -[0.18 \times \log_2(0.18) + 0.82 \times \log_2(0.82) + 0.0 \times \log_2(0.0)]$

$= -[0.18 \times 2.47 + 0.82 \times -0.29 + 0.0 \times -0]$

$= -[0.44 + 0.24 + 0.0]$

$= 0.68$

IG (parent, child) = entropy(parent) – [p(2011) × entropy(2011) + p(2012) × entropy(2012) + p(2013) × entropy(2013)]

$= 1.20 - [0.18 \times 1.15 + 0.51 \times 1.19 + 0.31 \times 0.68]$

$= 1.20 - [0.21 + 0.61 + 0.21]$

$= 1.20 - 1.03$

$= 0.17$

3) คำนวณค่า IG ของแอตทริบิวต์ Crime Type(ประเภทอาชญากรรม) และ Victim Sex (เพศของเหยื่อ) ของ Murder or Manslaughter จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = $-p(\text{Male}) \times \log_2 p(\text{Male}) + p(\text{Female}) \times \log_2 p(\text{Female}) + p(\text{Unknown}) \times \log_2 p(\text{Unknown})$

$= -[0.54 \times \log_2(0.54) + 0.16 \times \log_2(0.16) + 0.30 \times \log_2(0.30)]$

$= -[0.54 \times -0.89 + 0.16 \times -2.64 + 0.30 \times -1.74]$

$= -[0.48 + 0.42 + 0.52]$

$= 1.42$

entropy(Male) = $-p(\text{Male}) \times \log_2 p(\text{Male}) + p(\text{Female}) \times \log_2 p(\text{Female}) + p(\text{Unknown}) \times \log_2 p(\text{Unknown})$

$= -[0.52 \times \log_2(0.52) + 0.14 \times \log_2(0.14) + 0.34 \times \log_2(0.34)]$

$= -[0.52 \times -0.94 + 0.14 \times -2.84 + 0.34 \times -1.56]$

$= -[0.49 + 0.40 + 0.53]$

$= 1.42$

entropy(Female) = $-p(\text{Male}) \times \log_2 p(\text{Male}) + p(\text{Female}) \times \log_2 p(\text{Female}) + p(\text{Unknown}) \times \log_2 p(\text{Unknown})$

$= -[0.60 \times \log_2(0.60) + 0.19 \times \log_2(0.19) + 0.21 \times \log_2(0.21)]$

$= -[0.60 \times -0.74 + 0.19 \times -2.40 + 0.21 \times -2.25]$

$= -[0.44 + 0.46 + 0.47]$

$= 1.37$

IG (parent, child) = entropy(parent) – [p(Male) × entropy(Male) + p(Female) × entropy(Female)]

= 1.42 – [0.72 × 1.42 + 0.28 × 1.37]

= 1.42 – [1.02 + 0.38]

= 1.42 – 1.40

= 0.02

3.1) คำนวณค่า IG ของแอตทริบิวต์ Crime Type(ประเภทอาชญากรรม) และ Victim Sex (เพศของเหยื่อ) ของ Manslaughter by Negligence จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = $-p(\text{Male}) \times \log_2 p(\text{Male}) + p(\text{Female}) \times \log_2 p(\text{Female}) + p(\text{Unknown}) \times \log_2 p(\text{Unknown})$

= $-[0.45 \times \log_2(0.45) + 0.51 \times \log_2(0.51) + 0.04 \times \log_2(0.04)]$

= $-[0.45 \times -1.15 + 0.51 \times -0.97 + 0.04 \times -4.64]$

= $-[0.52 + 0.49 + 0.19]$

= 1.20

entropy(Male) = $-p(\text{Male}) \times \log_2 p(\text{Male}) + p(\text{Female}) \times \log_2 p(\text{Female}) + p(\text{Unknown}) \times \log_2 p(\text{Unknown})$

= $-[0.37 \times \log_2(0.37) + 0.57 \times \log_2(0.57) + 0.07 \times \log_2(0.07)]$

= $-[0.37 \times -1.43 + 0.57 \times -0.81 + 0.07 \times -3.84]$

= $-[0.53 + 0.46 + 0.27]$

= 1.26

entropy(Female) = $-p(Male) \times \log_2 p(Male) + p(Female) \times \log_2 p(Female) + p(Unknown) \times \log_2 p(Unknown)$

$= -[0.58 \times \log_2(0.58) + 0.42 \times \log_2(0.42) + 0.0 \times \log_2(0.0)]$

$= -[0.58 \times 0.79 + 0.42 \times -1.25 + 0.0 \times -0]$

$= -[0.46 + 0.53 + 0.0]$

$= 0.99$

IG (parent, child) = entropy(parent) – [p(Male) × entropy(Male) + p(Female) × entropy(Female)]

$= 1.20 - [0.54 \times 1.26 + 0.46 \times 0.99]$

$= 1.20 - [0.68 + 0.46]$

$= 1.20 - 1.14$

$= 0.06$

4) คำนวณค่า IG ของแอตทริบิวต์ Crime Type(ประเภทอาชญากรรม) และ Weapon(อาวุธ) ของ Murder or Manslaughter จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = $-p(Male) \times \log_2 p(Male) + p(Female) \times \log_2 p(Female) + p(Unknown) \times \log_2 p(Unknown)$

$= -[0.54 \times \log_2(0.54) + 0.16 \times \log_2(0.16) + 0.30 \times \log_2(0.30)]$

$= -[0.54 \times -0.89 + 0.16 \times -2.64 + 0.30 \times -1.74]$

$= -[0.48 + 0.42 + 0.52]$

$= 1.42$

entropy(Drugs) = $-p(\text{Male}) \times \log_2 p(\text{Male}) + p(\text{Female}) \times \log_2 p(\text{Female}) + p(\text{Unknown}) \times \log_2 p(\text{Unknown})$

$= -[0.52 \times \log_2(0.52) + 0.27 \times \log_2(0.27) + 0.21 \times \log_2(0.21)]$

$= -[0.52 \times -0.94 + 0.27 \times -1.89 + 0.21 \times -2.25]$

$= -[0.49 + 0.51 + 0.47]$

$= 1.47$

entropy(Fire) = $-p(\text{Male}) \times \log_2 p(\text{Male}) + p(\text{Female}) \times \log_2 p(\text{Female}) + p(\text{Unknown}) \times \log_2 p(\text{Unknown})$

$= -[0.54 \times \log_2(0.54) + 0.15 \times \log_2(0.15) + 0.30 \times \log_2(0.30)]$

$= -[0.54 \times -0.89 + 0.15 \times -2.74 + 0.30 \times -1.74]$

$= -[0.49 + 0.41 + 0.52]$

$= 1.42$

entropy(Gun) = $-p(\text{Male}) \times \log_2 p(\text{Male}) + p(\text{Female}) \times \log_2 p(\text{Female}) + p(\text{Unknown}) \times \log_2 p(\text{Unknown})$

$= -[0.56 \times \log_2(0.56) + 0.08 \times \log_2(0.08) + 0.36 \times \log_2(0.36)]$

$= -[0.56 \times -0.84 + 0.08 \times -3.64 + 0.36 \times -1.47]$

$= -[0.47 + 0.29 + 0.53]$

$= 1.29$

IG (parent, child) = entropy(parent) – [p(Drugs) × entropy(Drugs) + p(Fire) × entropy(Fire) + p(Gun) × entropy(Gun)]

$= 1.42 - [0.26 \times 1.47 + 0.33 \times 1.42 + 0.40 \times 1.29]$

$= 1.42 - [0.38 + 0.47 + 0.52]$

$= 1.42 - 1.37$

$= 0.05$

4.1) คำนวณค่า IG ของแอตทริบิวต์ Crime Type(ประเภทอาชญากรรม) และ Weapon(อาวุธ) ของ Manslaughter by Negligence จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = $-p(\text{Male}) \times \log_2 p(\text{Male}) + p(\text{Female}) \times \log_2 p(\text{Female}) + p(\text{Unknown}) \times \log_2 p(\text{Unknown})$

$= -[0.45 \times \log_2(0.45) + 0.51 \times \log_2(0.51) + 0.04 \times \log_2(0.04)]$

$= -[0.45 \times -1.15 + 0.51 \times -0.97 + 0.04 \times -4.64]$

$= -[0.52 + 0.49 + 0.19]$

$= 1.20$

entropy(Drugs) = $-p(\text{Male}) \times \log_2 p(\text{Male}) + p(\text{Female}) \times \log_2 p(\text{Female}) + p(\text{Unknown}) \times \log_2 p(\text{Unknown})$

$= -[0.55 \times \log_2(0.55) + 0.41 \times \log_2(0.41) + 0.03 \times \log_2(0.03)]$

$= -[0.55 \times -0.86 + 0.41 \times -1.29 + 0.03 \times -5.06]$

$= -[0.47 + 0.53 + 0.15]$

$= 1.15$

entropy(Fire) = $-p(\text{Male}) \times \log_2 p(\text{Male}) + p(\text{Female}) \times \log_2 p(\text{Female}) + p(\text{Unknown}) \times \log_2 p(\text{Unknown})$

$= -[0.08 \times \log_2(0.08) + 0.85 \times \log_2(0.85) + 0.08 \times \log_2(0.08)]$

$= -[0.08 \times -3.64 + 0.85 \times -0.23 + 0.08 \times -3.64]$

$= -[0.29 + 0.20 + 0.29]$

$= 0.78$

entropy(Gun) = $-p(\text{Male}) \times \log_2 p(\text{Male}) + p(\text{Female}) \times \log_2 p(\text{Female}) + p(\text{Unknown}) \times \log_2 p(\text{Unknown})$

$= -[0.62 \times \log_2(0.62) + 0.38 \times \log_2(0.38) + 0.0 \times \log_2(0.0)]$

$= -[0.62 \times -0.69 + 0.38 \times -1.40 + 0.0 \times -0]$

$= -[0.43 + 0.53 + 0.0]$

$= 0.96$

IG (parent, child) = entropy(parent) – [p(Drugs) × entropy(Drugs) + p(Fire) × entropy(Fire) + p(Gun) × entropy(Gun)]

$= 1.20 - [0.53 \times 1.15 + 0.24 \times 0.78 + 0.24 \times 0.96]$

$= 1.20 - [0.61 + 0.19 + 0.23]$

$= 1.20 - 1.03$

$= 0.17$

จากการคำนวณค่า IG ของแอตทริบิวต์ Crime Type(ประเภทอาชญากรรม) ต่อ Agency Type(ตัวแทนหน่วยงาน) Year(ปี) Victim Sex (เพศของเหยื่อ) และWeapon(อาวุธ) พบว่าค่า IG ของแอตทริบิวต์ Crime Type(ประเภทอาชญากรรม) ต่อ Year(ปี)และWeapon(อาวุธ) มีค่ามากที่สุด (0.17) ดังนั้นจึงเลือกแอตทริบิวต์ Year(ปี) ขึ้นมาเป็นโหนดในระดับที่ 2 ต่อจากโหนด Root และทำการแตกกิ่งจากโหนดในระดับที่ 2 ออกไปจนข้อมูลในแต่ละโหนดมีคลาสคำตอบเดียวกัน

การคำนวณค่าแต่ละแอตทริบิวต์ในระดับที่ 3 กับจำนวนข้อมูลทั้งหมดโดยใช้ผลลัพธ์เป็น Perpetrator Sex(เพศผู้กระทำผิด)

1) คำนวณค่า IG ของแอตทริบิวต์ Crime Type(Manslaughter by Negligence)และ Year(2012) และ Agency Type จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = $-p(\text{Male}) \times \log_2 p(\text{Male}) + p(\text{Female}) \times \log_2 p(\text{Female}) + p(\text{Unknown}) \times \log_2 p(\text{Unknown})$

$= -[0.54 \times \log_2(0.54) + 0.43 \times \log_2(0.43) + 0.04 \times \log_2(0.04)]$

$= -[0.54 \times -0.89 + 0.43 \times -1.22 + 0.04 \times -4.64]$

$= -[0.48 + 0.52 + 0.19]$

$= 1.19$

entropy(County Police) = $p(\text{Male}) \times \log_2 p(\text{Male}) + p(\text{Female}) \times \log_2 p(\text{Female}) + p(\text{Unknown}) \times \log_2 p(\text{Unknown})$

$= -[1.00 \times \log_2(1.00) + 0.00 \times \log_2(0.00) + 0.00 \times \log_2(0.00)]$

$= -[1.00 \times -0.00 + 0.00 \times -0 + 0.00 \times -0]$

$= -[0 + 0 + 0]$

$= 0$

entropy(Municipal Police) = $p(\text{Male}) \times \log_2 p(\text{Male}) + p(\text{Female}) \times \log_2 p(\text{Female}) + p(\text{Unknown}) \times \log_2 p(\text{Unknown})$

$= -[0.54 \times \log_2(0.54) + 0.42 \times \log_2(0.42) + 0.04 \times \log_2(0.04)]$

$= -[0.54 \times -0.89 + 0.42 \times -1.25 + 0.04 \times -4.64]$

$= -[0.48 + 0.53 + 0.19]$

$= 1.20$

entropy(Special Police) = p(Male) × $\log_2$ p(Male) + p(Female) × $\log_2$ p(Female) + p(Unknown) × $\log_2$ p(Unknown)

= −[0.00 × $\log_2$(0.00) + 0.00 × $\log_2$(0.00) + 0.00 × $\log_2$(0.00)]

= −[0.00 × −0.00 + 0.00 × −0.00 + 0.00 × −0.00]

= −[0.00 + 0.00 + 0.00]

= 0

entropy(State Police) = p(Male) × $\log_2$ p(Male) + p(Female) × $\log_2$ p(Female) + p(Unknown) × $\log_2$ p(Unknown)

= −[0.00 × $\log_2$(0.00) + 1.00 × $\log_2$(0.00) + 0.00 × $\log_2$(0.00)]

= −[0.00 × −0.00 + 1.00 × −0.00 + 0.00 × −0.00]

= −[0.00 + 0.00 + 0.00]

= 0

IG (parent, child) = entropy(parent) – [p(County Police) × entropy(County Police) + p(Municipal Police) × entropy(Municipal Police) + p(Special Police) × entropy(Special Police) + p(State Police) × entropy(State Police)]

= 1.19 – [0.07 × 0 + 0.86 × 1.20 + 0.00 × 0.00 + 0.07 × 0]

= 1.19 – [0.00 + 1.03 + 0.00 + 0.00]

= 1.19 – 1.03

= 0.16

1.1) คำนวณค่า IG ของแอตทริบิวต์ Crime Type(Murder or Manslaughter)และ Year(2012) และ Agency Type จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = $-p(\text{Male}) \times \log_2 p(\text{Male}) + p(\text{Female}) \times \log_2 p(\text{Female}) + p(\text{Unknown}) \times \log_2 p(\text{Unknown})$

$= -[0.53 \times \log_2(0.53) + 0.16 \times \log_2(0.16) + 0.31 \times \log_2(0.31)]$

$= -[0.53 \times -0.92 + 0.16 \times -2.64 + 0.31 \times -1.69]$

$= -[0.49 + 0.42 + 0.52]$

$= 1.43$

entropy(County Police) = $p(\text{Male}) \times \log_2 p(\text{Male}) + p(\text{Female}) \times \log_2 p(\text{Female}) + p(\text{Unknown}) \times \log_2 p(\text{Unknown})$

$= -[0.67 \times \log_2(0.67) + 0.00 \times \log_2(0.00) + 0.33 \times \log_2(0.33)]$

$= -[0.67 \times -0.58 + 0.00 \times -0 + 0.33 \times -1.60]$

$= -[0.39 + 0.00 + 0.53]$

$= 0.92$

entropy(Municipal Police) = $p(\text{Male}) \times \log_2 p(\text{Male}) + p(\text{Female}) \times \log_2 p(\text{Female}) + p(\text{Unknown}) \times \log_2 p(\text{Unknown})$

$= -[0.51 \times \log_2(0.51) + 0.16 \times \log_2(0.16) + 0.33 \times \log_2(0.33)]$

$= -[0.51 \times -0.97 + 0.16 \times -2.64 + 0.33 \times -1.60]$

$= -[0.49 + 0.42 + 0.53]$

$= 1.44$

entropy(Special Police) = p(Male) × $\log_2$ p(Male) + p(Female) × $\log_2$ p(Female) + p(Unknown) × $\log_2$ p(Unknown)

= −[0.00 × $\log_2$(0.00) + 0.50 × $\log_2$(0.50) + 0.50 × $\log_2$(0.50)]

= −[0.00 × −0.00 + 0.50 × −1 + 0.50 × −1]

= −[0.00 + 0.50 + 0.50]

= 1

entropy(State Police) = p(Male) × $\log_2$ p(Male) + p(Female) × $\log_2$ p(Female) + p(Unknown) × $\log_2$ p(Unknown)

= −[0.77 × $\log_2$(0.77) + 0.15 × $\log_2$(0.15) + 0.08 × $\log_2$(0.08)]

= −[0.77 × −0.38 + 0.15 × −2.74 + 0.08 × −3.64]

= −[0.29 + 0.41 + 0.29]

= 0.99

IG (parent, child) = entropy(parent) – [p(County Police) × entropy(County Police) + p(Municipal Police) × entropy(Municipal Police) + p(Special Police) × entropy(Special Police) + p(State Police) × entropy(State Police)]

= 1.43 – [0.02 × 0.92 + 0.91 × 1.44 + 0.01 × 1 + 0.07 × 0.99]

= 1.43 – [0.02 + 1.31 + 0.01 + 0.07]

= 1.43 – 1.41

= 0.02

2) Crime Type(Manslaughter by Negligence)และ Year(2012) และ Victim Sex จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = $-p(\text{Male}) \times \log_2 p(\text{Male}) + p(\text{Female}) \times \log_2 p(\text{Female}) + p(\text{Unknown}) \times \log_2 p(\text{Unknown})$

$= -[0.54 \times \log_2(0.54) + 0.43 \times \log_2(0.43) + 0.04 \times \log_2(0.04)]$

$= -[0.54 \times -0.89 + 0.43 \times -1.22 + 0.04 \times -4.64]$

$= -[0.48 + 0.52 + 0.19]$

$= 1.19$

entropy(Male) = $-p(\text{Male}) \times \log_2 p(\text{Male}) + p(\text{Female}) \times \log_2 p(\text{Female}) + p(\text{Unknown}) \times \log_2 p(\text{Unknown})$

$= -[0.40 \times \log_2(0.40) + 0.53 \times \log_2(0.53) + 0.07 \times \log_2(0.07)]$

$= -[0.40 \times -1.32 + 0.53 \times -0.92 + 0.07 \times -3.84]$

$= -[0.53 + 0.49 + 0.27]$

$= 1.29$

entropy(Female) = $-p(\text{Male}) \times \log_2 p(\text{Male}) + p(\text{Female}) \times \log_2 p(\text{Female}) + p(\text{Unknown}) \times \log_2 p(\text{Unknown})$

$= -[0.69 \times \log_2(0.69) + 0.31 \times \log_2(0.31) + 0.00 \times \log_2(0.00)]$

$= -[0.69 \times -0.54 + 0.31 \times -1.69 + 0.00 \times -0.00]$

$= -[0.37 + 0.52 + 0.00]$

$= 0.89$

IG(parent, child) = entropy(parent) – [p(Male) × entropy(Male) + p(Female) × entropy(Female)]

= 1.19 – [0.54 × 1.29 + 0.46 × 0.89]

= 1.19 – [0.70 + 0.41]

= 1.19 – 1.11

= 0.08

2.1) Crime Type(Murder or Manslaughter)และ Year(2012) และ Victim Sex จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = –p(Male) × $\log_2$ p(Male) + p(Female) × $\log_2$ p(Female) + p(Unknown) × $\log_2$ p(Unknown)

= –[0.53 × $\log_2$(0.53) + 0.16 × $\log_2$(0.16) + 0.31 × $\log_2$(0.31)]

= –[0.53 × –0.92 + 0.16 × –2.64 + 0.31 × –1.69]

= –[0.49 + 0.42 + 0.52]

= 1.43

entropy(Male) = –p(Male) × $\log_2$ p(Male) + p(Female) × $\log_2$ p(Female) + p(Unknown) × $\log_2$ p(Unknown)

= –[0.51 × $\log_2$(0.51) + 0.15 × $\log_2$(0.15) + 0.33 × $\log_2$(0.33)]

= –[0.51 × –0.97 + 0.15 × –2.74 + 0.33 × –1.60]

= –[0.49 + 0.41 + 0.53]

= 1.43

entropy(Female) = $-p(Male) \times \log_2 p(Male) + p(Female) \times \log_2 p(Female) + p(Unknown) \times \log_2 p(Unknown)$

$= -[0.57 \times \log_2(0.57) + 0.20 \times \log_2(0.20) + 0.24 \times \log_2(0.24)]$

$= -[0.57 \times -0.81 + 0.20 \times -2.32 + 0.24 \times -2.06]$

$= -[0.46 + 0.46 + 0.49]$

$= 1.41$

IG(parent, child) = entropy(parent) – [p(Male) × entropy(Male) + p(Female) × entropy(Female)]

$= 1.43 – [0.76 \times 1.43 + 0.24 \times 1.41]$

$= 1.43 – [1.09 + 0.34]$

$= 1.43 – 1.43$

$= 0$

3) คำนวณค่า IG ของแอตทริบิวต์ Crime Type(Manslaughter by Negligence)และ Year(2012) และ Weapon จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = $-p(Male) \times \log_2 p(Male) + p(Female) \times \log_2 p(Female) + p(Unknown) \times \log_2 p(Unknown)$

$= -[0.54 \times \log_2(0.54) + 0.43 \times \log_2(0.43) + 0.04 \times \log_2(0.04)]$

$= -[0.54 \times -0.89 + 0.43 \times -1.22 + 0.04 \times -4.64]$

$= -[0.48 + 0.52 + 0.19]$

$= 1.19$

entropy(Drugs) = $-p(\text{Male}) \times \log_2 p(\text{Male}) + p(\text{Female}) \times \log_2 p(\text{Female}) + p(\text{Unknown}) \times \log_2 p(\text{Unknown})$

$= -[0.63 \times \log_2(0.63) + 0.31 \times \log_2(0.31) + 0.06 \times \log_2(0.03)]$

$= -[0.63 \times -0.67 + 0.31 \times -1.69 + 0.06 \times -4.06]$

$= -[0.42 + 0.52 + 0.24]$

$= 1.18$

entropy(Fire) = $-p(\text{Male}) \times \log_2 p(\text{Male}) + p(\text{Female}) \times \log_2 p(\text{Female}) + p(\text{Unknown}) \times \log_2 p(\text{Unknown})$

$= -[0.00 \times \log_2(0.00) + 1.00 \times \log_2(1.00) + 0.00 \times \log_2(0.00)]$

$= -[0.00 \times -0.00 + 1.00 \times -0.00 + 0.00 \times -0.00]$

$= -[0.00 + 0.00 + 0.00]$

$= 0.00$

entropy(Gun) = $-p(\text{Male}) \times \log_2 p(\text{Male}) + p(\text{Female}) \times \log_2 p(\text{Female}) + p(\text{Unknown}) \times \log_2 p(\text{Unknown})$

$= -[0.83 \times \log_2(0.83) + 0.17 \times \log_2(0.17) + 0.0 \times \log_2(0.0)]$

$= -[0.83 \times -0.27 + 0.17 \times -2.56 + 0.0 \times -0]$

$= -[0.22 + 0.44 + 0.0]$

$= 0.66$

IG (parent, child) = entropy(parent) – [p(Drugs) × entropy(Drugs) + p(Fire) × entropy(Fire) + p(Gun) × entropy(Gun)]

$= 1.19 - [0.57 \times 1.18 + 0.21 \times 0.0 + 0.21 \times 0.66]$

$= 1.19 - [0.67 + 0.0 + 0.14]$

$= 1.19 - 0.78$

$= 0.41$

3.1) คำนวณค่า IG ของแอตทริบิวต์ Crime Type(Murder or Manslaughter)และ Year(2012) และ Weapon จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = $-p(\text{Male}) \times \log_2 p(\text{Male}) + p(\text{Female}) \times \log_2 p(\text{Female}) + p(\text{Unknown}) \times \log_2 p(\text{Unknown})$

$= -[0.53 \times \log_2(0.53) + 0.16 \times \log_2(0.16) + 0.31 \times \log_2(0.31)]$

$= -[0.53 \times -0.92 + 0.16 \times -2.64 + 0.31 \times -1.69]$

$= -[0.49 + 0.42 + 0.52]$

$= 1.43$

entropy(Drugs) = $-p(\text{Male}) \times \log_2 p(\text{Male}) + p(\text{Female}) \times \log_2 p(\text{Female}) + p(\text{Unknown}) \times \log_2 p(\text{Unknown})$

$= -[0.45 \times \log_2(0.45) + 0.39 \times \log_2(0.39) + 0.16 \times \log_2(0.16)]$

$= -[0.45 \times -1.15 + 0.39 \times -1.36 + 0.16 \times -2.64]$

$= -[0.52 + 0.53 + 0.42]$

$= 1.47$

entropy(Fire) = $-p(\text{Male}) \times \log_2 p(\text{Male}) + p(\text{Female}) \times \log_2 p(\text{Female}) + p(\text{Unknown}) \times \log_2 p(\text{Unknown})$

$= -[0.55 \times \log_2(0.55) + 0.18 \times \log_2(0.18) + 0.27 \times \log_2(0.27)]$

$= -[0.55 \times -0.86 + 0.18 \times -2.47 + 0.27 \times -1.89]$

$= -[0.47 + 0.44 + 0.51]$

$= 1.42$

entropy(Gun) = $-p(\text{Male}) \times \log_2 p(\text{Male}) + p(\text{Female}) \times \log_2 p(\text{Female}) + p(\text{Unknown}) \times \log_2 p(\text{Unknown})$

$= -[0.54 \times \log_2(0.54) + 0.09 \times \log_2(0.09) + 0.37 \times \log_2(0.37)]$

$= -[0.54 \times -0.89 + 0.09 \times -3.47 + 0.37 \times -1.43]$

$= -[0.48 + 0.31 + 0.53]$

$= 1.32$

IG (parent, child) = entropy(parent) – [p(Drugs) × entropy(Drugs) + p(Fire) × entropy(Fire) + p(Gun) × entropy(Gun)]

$= 1.43 - [0.16 \times 1.47 + 0.27 \times 1.42 + 0.57 \times 1.32]$

$= 1.43 - [0.24 + 0.38 + 0.75]$

$= 1.43 - 1.37$

$= 0.06$

จากการคำนวณค่า IG ของแอตทริบิวต์ Crime Type(ประเภทอาชญากรรม) ต่อ Agency Type(ตัวแทนหน่วยงาน) Year(ปี) Victim Sex (เพศของเหยื่อ) และWeapon(อาวุธ) พบว่าค่า IG ของแอตทริบิวต์ Crime Type(ประเภทอาชญากรรม) ต่อ Year(ปี)และWeapon(อาวุธ) มีค่ามากที่สุด (0.41) ดังนั้นจึงเลือกแอตทริบิวต์ Weapon(อาวุธ) ขึ้นมาเป็นโหนดในระดับที่ 3 ต่อจากโหนด Root และทำการแตกกิ่งจากโหนดในระดับที่ 3 ออกไปจนข้อมูลในแต่ละโหนดมีคลาสคำตอบเดียวกัน

การคำนวณค่าแต่ละแอตทริบิวต์ในระดับที่ 4 กับจำนวนข้อมูลทั้งหมดโดยใช้ผลลัพธ์เป็น Perpetrator Sex(เพศผู้กระทำผิด)

1) คำนวณค่า IG ของแอตทริบิวต์ Crime Type(Manslaughter by Negligence)และ Year(2012) และ Weapon(Drugs) และAgency Type จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = $-p(\text{Male}) \times \log_2 p(\text{Male}) + p(\text{Female}) \times \log_2 p(\text{Female}) + p(\text{Unknown}) \times \log_2 p(\text{Unknown})$

$= -[0.63 \times \log_2(0.63) + 0.31 \times \log_2(0.31) + 0.06 \times \log_2(0.03)]$

$= -[0.63 \times -0.67 + 0.31 \times -1.69 + 0.06 \times -4.06]$

$= -[0.42 + 0.52 + 0.24]$

$= 1.18$

entropy(County Police) = $p(\text{Male}) \times \log_2 p(\text{Male}) + p(\text{Female}) \times \log_2 p(\text{Female}) + p(\text{Unknown}) \times \log_2 p(\text{Unknown})$

$= -[1.00 \times \log_2(1.00) + 0.00 \times \log_2(0.00) + 0.00 \times \log_2(0.00)]$

$= -[1.00 \times -0.00 + 0.00 \times -0.00 + 0.00 \times -0.00]$

$= -[0 + 0 + 0]$

$= 0$

entropy(Municipal Police) = $p(\text{Male}) \times \log_2 p(\text{Male}) + p(\text{Female}) \times \log_2 p(\text{Female}) + p(\text{Unknown}) \times \log_2 p(\text{Unknown})$

$= -[0.69 \times \log_2(0.69) + 0.23 \times \log_2(0.23) + 0.08 \times \log_2(0.08)]$

$= -[0.69 \times -0.54 + 0.23 \times -2.12 + 0.08 \times -3.64]$

$= -[0.37 + 0.49 + 0.29]$

$= 1.15$

entropy(Special Police) = p(Male) × $\log_2$ p(Male) + p(Female) × $\log_2$ p(Female) + p(Unknown) × $\log_2$ p(Unknown)

= −[0.00 × $\log_2$(0.00) + 0.00 × $\log_2$(0.00) + 0.00 × $\log_2$(0.00)]

= −[0.00 × −0.00 + 0.00 × −0.00 + 0.00 × −0.00]

= −[0 + 0 + 0]

= 0

entropy(State Police) = p(Male) × $\log_2$ p(Male) + p(Female) × $\log_2$ p(Female) + p(Unknown) × $\log_2$ p(Unknown)

= −[0.00 × $\log_2$(0.00) + 1.00 × $\log_2$(1.00) + 0.00 × $\log_2$(0.00)]

= −[0.00 × −0.00 + 1.00 × −0.00 + 0.00 × −0.00]

= −[0 + 0 + 0]

= 0

IG (parent, child) = entropy(parent) – [p(County Police) × entropy(County Police) + p(Municipal Police) × entropy(Municipal Police) + p(Special Police) × entropy(Special Police) + p(State Police) × entropy(State Police)]

= 1.18 – [0.06 × 0.00 + 0.81 × 1.15 + 0.00 × 0.00 + 0.13 × 0.00]

= 1.18 – [0 + 0.93 + 0 + 0]

= 1.18 – 0.93

= 0.25

2) คำนวณค่า IG ของแอตทริบิวต์ Crime Type(Manslaughter by Negligence)และ Year(2012) และ Weapon(Drugs) และVictim Sex จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = $-p(\text{Male}) \times \log_2 p(\text{Male}) + p(\text{Female}) \times \log_2 p(\text{Female}) + p(\text{Unknown}) \times \log_2 p(\text{Unknown})$

$= -[0.63 \times \log_2(0.63) + 0.31 \times \log_2(0.31) + 0.06 \times \log_2(0.03)]$

$= -[0.63 \times -0.67 + 0.31 \times -1.69 + 0.06 \times -4.06]$

$= -[0.42 + 0.52 + 0.24]$

$= 1.18$

entropy(Male) = $-p(\text{Male}) \times \log_2 p(\text{Male}) + p(\text{Female}) \times \log_2 p(\text{Female}) + p(\text{Unknown}) \times \log_2 p(\text{Unknown})$

$= -[0.38 \times \log_2(0.38) + 0.50 \times \log_2(0.50) + 0.13 \times \log_2(0.13)]$

$= -[0.38 \times -1.40 + 0.50 \times -1 + 0.13 \times -2.94]$

$= -[0.53 + 0.50 + 0.38]$

$= 1.41$

entropy(Female) = $-p(\text{Male}) \times \log_2 p(\text{Male}) + p(\text{Female}) \times \log_2 p(\text{Female}) + p(\text{Unknown}) \times \log_2 p(\text{Unknown})$

$= -[0.88 \times \log_2(0.88) + 0.13 \times \log_2(0.13) + 0.00 \times \log_2(0.00)]$

$= -[0.88 \times -0.18 + 0.13 \times -2.94 + 0.00 \times -0.00]$

$= -[0.16 + 0.38 + 0]$

$= 0.54$

IG(parent, child) = entropy(parent) – [p(Male) × entropy(Male) + p(Female) × entropy(Female)]

$= 1.18 - [0.50 \times 1.41 + 0.50 \times 0.54]$

$= 1.18 - [0.71 + 0.27]$

$= 1.18 - 0.98$

$= 0.2$

จากการคำนวณค่า IG ของแอตทริบิวต์ Crime Type(ประเภทอาชญากรรม) ต่อ Agency Type(ตัวแทนหน่วยงาน) Year(ปี) Victim Sex (เพศของเหยื่อ) และWeapon(อาวุธ) พบว่าค่า IG ของแอตทริบิวต์ Crime Type(ประเภทอาชญากรรม) และ Year(ปี) และWeapon(อาวุธ) ต่อ Agency Type(ตัวแทนหน่วยงาน) มีค่ามากที่สุด (0.25) ดังนั้นจึงเลือกแอตทริบิวต์ Agency Type(ตัวแทนหน่วยงาน) ขึ้นมาเป็นโหนดในระดับที่ 4 ต่อจากโหนดระดับที่ 3 และโหนดสุดท้ายคือแอตทริบิวต์ Victim Sex (เพศของเหยื่อ) ทำการแตกกิ่งจากโหนดในระดับที่ 3 ออกไป

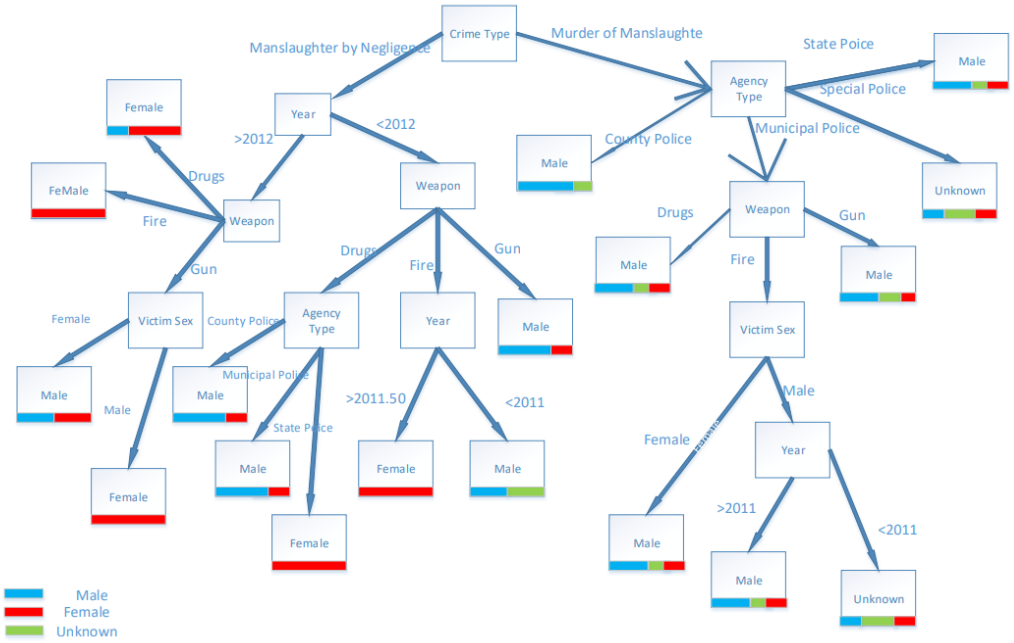
จากการคำนวณค่า IG ของแอตทริบิวต์ Crime Type และ Year และ Weapon และ Agency Type และ Victim Sex พบว่าข้อมูลในแต่ละโหนดมีคลาสคำตอบเดียวกันแล้ว คือ ผลลัพธ์เป็น Perpetrator Sex

3.1.5 การประเมินผล (Evaluation) เป็นขั้นตอนก่อนนำผลลัพธ์ที่ได้จากขั้นตอนที่ 4 ไปใช้งาน ด้วยการวัดประสิทธิผลของผลลัพธ์ที่ได้กับวัตถุประสงค์ที่ตั้งไว้ในขั้นตอนแรก ว่ามีนัยสำคัญหรือความน่าเชื่อถือมากน้อยเพียงใด ด้วยการประเมินผลจากโปรแกรมว่าถูกต้องหรือไม่

ผู้วิเคราะห์ข้อมูลได้ข้อมูลที่ผ่านการวิเคราะห์ข้อมูลด้วยเทคนิค Data Mining จากการสร้างโมเดล Decision Tree ดังนี้

1.โปรแกรม RapidMiner Studio 9.6

ภาพที่ 3.8 รูปแบบโมเดล Graph Decision Tree ในโปรแกรม RapidMiner



ภาพที่ 3.9 รูปแบบโมเดล Graph Decision Tree ในการคำนวณมือ

Tree

```
Crime Type = Manslaughter by Negligence
|   Year > 2012.500
|   |   Weapon = Drugs: Female {Male=2, Unknown=0, Female=6}
|   |   Weapon = Fire: Female {Male=0, Unknown=0, Female=5}
|   |   Weapon = Gun
|   |   |   Victim Sex = Female: Male {Male=1, Unknown=0, Female=1}
|   |   |   Victim Sex = Male: Female {Male=0, Unknown=0, Female=2}
|   Year ≤ 2012.500
|   |   Weapon = Drugs
|   |   |   Agency Type = County Police: Male {Male=2, Unknown=0, Female=1}
|   |   |   Agency Type = Municipal Police: Male {Male=12, Unknown=1, Female=3}
|   |   |   Agency Type = State Police: Female {Male=0, Unknown=0, Female=2}
|   |   Weapon = Fire
|   |   |   Year > 2011.500: Female {Male=0, Unknown=0, Female=6}
|   |   |   Year ≤ 2011.500: Male {Male=1, Unknown=1, Female=0}
|   |   Weapon = Gun: Male {Male=7, Unknown=0, Female=2}
Crime Type = Murder or Manslaughter
|   Agency Type = County Police: Male {Male=15, Unknown=3, Female=0}
|   Agency Type = Municipal Police
|   |   Weapon = Drugs: Male {Male=69, Unknown=32, Female=37}
|   |   Weapon = Fire
|   |   |   Victim Sex = Female: Male {Male=37, Unknown=13, Female=11}
|   |   |   Victim Sex = Male
|   |   |   |   Year > 2011.500: Male {Male=38, Unknown=26, Female=14}
|   |   |   |   Year ≤ 2011.500: Unknown {Male=15, Unknown=16, Female=3}
|   |   Weapon = Gun: Male {Male=117, Unknown=83, Female=18}
|   Agency Type = Special Police: Unknown {Male=1, Unknown=2, Female=1}
|   Agency Type = State Police: Male {Male=25, Unknown=2, Female=7}
```

ภาพที่ 3.10 คำบรรยายลักษณะงาน Decision Tree ของ RapidMiner

จากผลสรุปที่ได้จากการสร้างโมเดล Decision Tree จากโปรแกรม RapidMiner และจากการคำนวณมือ มีผลลัพธ์ออกมาที่สอดคล้องกัน ทำให้ข้อมูลมีความถูกต้องและน่าเชื่อถือมากยิ่งขึ้น

ดังนั้น ผู้จัดทำได้ผลลัพธ์ของกฎที่สามารถทำนายได้จำนวน 18 กฎ และสามารถนำกฎที่ได้นำไปวิเคราะห์กฎต่อไปได้ โดยสามารถจำแนกกฎได้ ดังนี้

กฎข้อที่ 1 IF Crime Type = Manslaughter by Negligence Year = >2012 Weapon = Drugs = Female หมายความว่า ถ้าประเภทอาชญากรรมเป็นการฆ่าโดยประมาท ในปีที่ต่ำกว่าปี 2012 โดย อาวุธ คือ การใช้ยา ส่วนมากผู้กระทำผิดจะเป็นเพศ “หญิง”

กฎข้อที่ 2 IF Crime Type = Manslaughter by Negligence Year = >2012 Weapon = Fire = Female หมายความว่า ถ้าประเภทอาชญากรรมเป็นการฆ่าโดยประมาท ในปีที่ต่ำกว่าปี 2012 โดย อาวุธ คือ ไฟ ส่วนมากผู้กระทำผิดจะเป็นเพศ “หญิง”

กฎข้อที่ 3 IF Crime Type = Manslaughter by Negligence Year = >2012 Weapon = Gun = Victim Sex = Female = Male หมายความว่า ถ้าประเภทอาชญากรรมเป็นการฆ่าโดยประมาท ในปีที่ต่ำกว่าปี 2012 โดย อาวุธ คือ ปืน เพศเหยื่อ ผู้หญิง ส่วนมากผู้กระทำผิดจะเป็นเพศ “ชาย”

กฎข้อที่ 4 IF Crime Type = Manslaughter by Negligence Year = >2012 Weapon = Gun = Victim Sex = Male = Female หมายความว่า ถ้าประเภทอาชญากรรมเป็นการฆ่าโดยประมาท ในปีที่ต่ำกว่าปี 2012 โดย อาวุธ คือ ปืน เพศเหยื่อ ผู้ชาย ส่วนมากผู้กระทำผิดจะเป็นเพศ “หญิง”

กฎข้อที่ 5 IF Crime Type = Manslaughter by Negligence Year = <2012 Weapon = Drugs Agency Type = County Police = Male หมายความว่า ถ้าประเภทอาชญากรรมเป็นการฆ่าโดยประมาท ในปีที่มากกว่าหรือในปี 2012 โดย อาวุธ คือ การใช้ยา ตัวแทนหน่วยงาน คือ ตำรวจนครบาล ส่วนมากผู้กระทำผิดจะเป็นเพศ “ชาย”

กฎข้อที่ 6 IF Crime Type = Manslaughter by Negligence Year = <2012 Weapon = Drugs Agency Type = Municipal Police = Male หมายความว่า ถ้าประเภทอาชญากรรมเป็นการฆ่าโดยประมาท ในปีที่มากกว่าหรือในปี 2012 โดย อาวุธ คือ การใช้ยา ตัวแทนหน่วยงาน คือ ตำรวจท้องที่ ส่วนมากผู้กระทำผิดจะเป็นเพศ “ชาย”

กฎข้อที่ 7 IF Crime Type = Manslaughter by Negligence Year = <2012 Weapon = Drugs Agency Type = State Police = Female หมายความว่า ถ้าประเภทอาชญากรรมเป็นการฆ่าโดยประมาท ในปีที่มากกว่าหรือในปี 2012 โดย อาวุธ คือ การใช้ยา ตัวแทนหน่วยงาน คือ ตำรวจรัฐ ส่วนมากผู้กระทำผิดจะเป็นเพศ “หญิง”

กฎข้อที่ 8 IF Crime Type = Manslaughter by Negligence Year = <2012 Weapon = Fire Year = >2012 = Female หมายความว่า ถ้าประเภทอาชญากรรมเป็นการฆ่าโดยประมาท ในปีที่มากกว่าหรือในปี 2012 โดย อาวุธ คือ ไฟ ในปีที่ต่ำกว่าปี 2012 ส่วนมากผู้กระทำผิดจะเป็นเพศ “หญิง”

กฎข้อที่ 9 IF Crime Type = Manslaughter by Negligence Year = <2012 Weapon = Fire Year = <2012 = Male หมายความว่า ถ้าประเภทอาชญากรรมเป็นการฆ่าโดยประมาท ในปีที่มากกว่าหรือในปี 2012 โดย อาวุธ คือ ไฟ ในปีที่มากกว่าหรือในปี 2012 ส่วนมากผู้กระทำผิดจะเป็นเพศ “ชาย”

กฎข้อที่ 10 IF Crime Type = Manslaughter by Negligence Year = <2012 Weapon = Gun = Male หมายความว่า ถ้าประเภทอาชญากรรมเป็นการการฆ่าโดยประมาท ในปีที่มากกว่าหรือในปี 2012 โดย อาวุธ คือ ปืน ส่วนมากผู้กระทำผิดจะเป็นเพศ “ชาย”

กฎข้อที่ 11 IF Crime Type = Murder or Manslaughter Agency Type = County Police = Male หมายความว่า ถ้าประเภทอาชญากรรมเป็นการฆ่าโดยเจตนา ตัวแทนหน่วยงาน คือ ตำรวจนครบาล ส่วนมากผู้กระทำผิดจะเป็นเพศ “ชาย”

กฎข้อที่ 12 IF Crime Type = Murder or Manslaughter Agency Type = State Police = Male หมายความว่า ถ้าประเภทอาชญากรรมเป็นการฆ่าโดยเจตนา ตัวแทนหน่วยงาน คือ ตำรวจรัฐ ส่วนมากผู้กระทำผิดจะเป็นเพศ “ชาย”

กฎข้อที่ 13 IF Crime Type = Murder or Manslaughter Agency Type = Special Police = Unknown หมายความว่า ถ้าประเภทอาชญากรรมเป็นการฆ่าโดยเจตนา ตัวแทนหน่วยงาน คือ ตำรวจพิเศษ ส่วนมากผู้กระทำผิดจะเป็นเพศ “ไม่ระบุเพศ”

กฎข้อที่ 14 IF Crime Type = Murder or Manslaughter Agency Type = Municipal Police Weapon = Gun = Male หมายความว่า ถ้าประเภทอาชญากรรมเป็นการฆ่าโดยเจตนา ตัวแทนหน่วยงาน คือ ตำรวจท้องที่ โดย อาวุธ คือ ปืน ส่วนมากผู้กระทำผิดจะเป็นเพศ “ชาย”

กฎข้อที่ 15 IF Crime Type = Murder or Manslaughter Agency Type = Municipal Police Weapon = Drugs = Male หมายความว่า ถ้าประเภทอาชญากรรมเป็นการฆ่าโดยเจตนา ตัวแทนหน่วยงาน คือ ตำรวจท้องที่ โดย อาวุธ คือ การใช้ยา ส่วนมากผู้กระทำผิดจะเป็นเพศ “ชาย”

กฎข้อที่ 16 IF Crime Type = Murder or Manslaughter Agency Type = Municipal Police Weapon = Fire Victim Sex = Female = Male หมายความว่า ถ้าประเภทอาชญากรรมเป็นการฆ่าโดยเจตนา ตัวแทนหน่วยงาน คือ ตำรวจท้องที่ โดย อาวุธ คือ ไฟ เพศเหยื่อ ผู้หญิง ส่วนมากผู้กระทำผิดจะเป็นเพศ “ชาย”

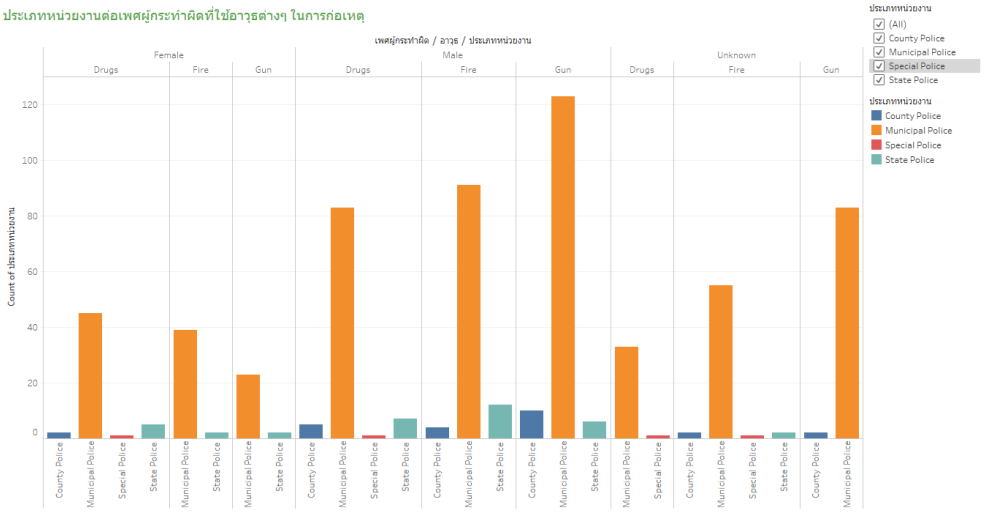
กฎข้อที่ 17 IF Crime Type = Murder or Manslaughter Agency Type = Municipal Police Weapon = Fire Victim Sex = Male Year = >2012 = Male หมายความว่า ถ้าประเภทอาชญากรรมเป็นการฆ่าโดยเจตนา ตัวแทนหน่วยงาน คือ ตำรวจท้องที่ โดย อาวุธ คือ ไฟ เพศเหยื่อ ผู้ชาย ในปีที่ต่ำกว่าปี 2012 ส่วนมากผู้กระทำผิดจะเป็นเพศ “ชาย”

กฎข้อที่ 17 IF Crime Type = Murder or Manslaughter Agency Type = Municipal Police Weapon = Fire Victim Sex = Male Year = <2012 = Unknown หมายความว่า ถ้าประเภทอาชญากรรมเป็นการฆ่าโดยเจตนา ตัวแทนหน่วยงาน คือ ตำรวจท้องที่ โดย อาวุธ คือ ไฟ เพศเหยื่อ ผู้ชาย ในปีที่มากกว่าหรือในปี 2012 ส่วนมากผู้กระทำผิดจะเป็นเพศ “ไม่ระบุเพศ”

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

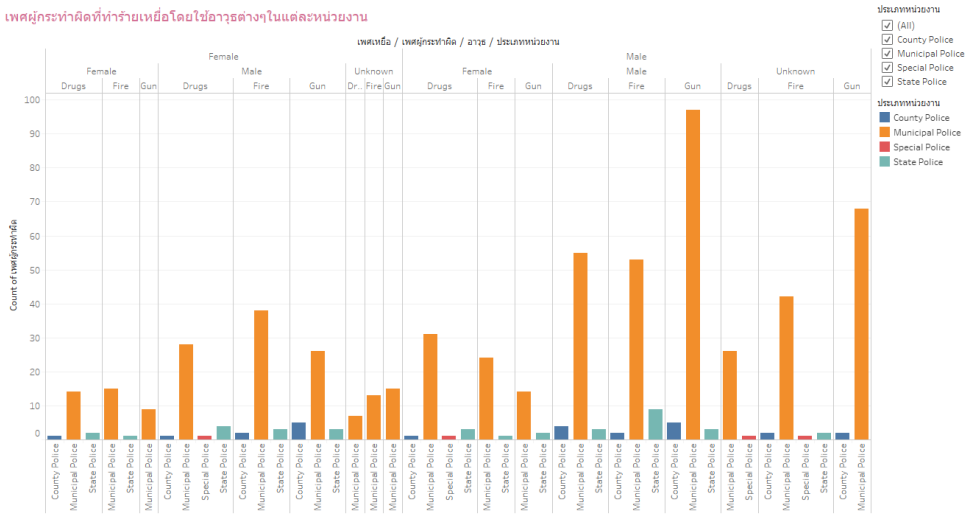
ผู้วิเคราะห์ข้อมูลนำผลข้อมูลที่ทำการวิเคราะห์มาแสดงผลข้อมูลบนหน้า Web Browser โดยใช้โปรแกรม Adobe Dreamweaver ร่วมกับการนำเสนอข้อมูลแบบ visualization ด้วยการแสดงผลข้อมูลในรูปแบบของภาพโดยใช้โปรแกรม Tableau Public

1. แสดงข้อมูลของประเภทหน่วยงาน แบบแผนภูมิแท่ง



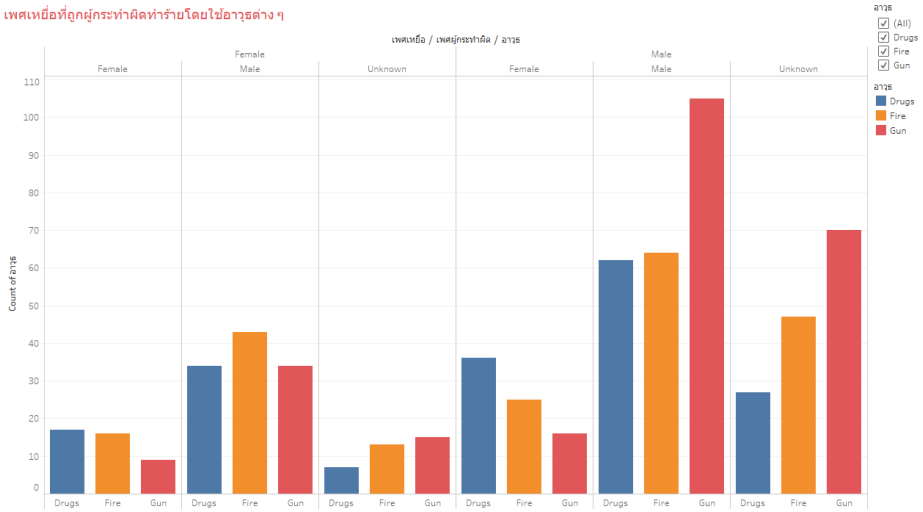
ภาพที่ 3.11 แสดงข้อมูลกราฟของประเภทหน่วยงาน

2. แสดงข้อมูลของการฆาตกรรม แบบแผนภูมิแท่ง



ภาพที่ 3.12 แสดงข้อมูลของการฆาตกรรม

3. แสดงข้อมูลของการใช้อาวุธ แบบแผนภูมิแท่ง



ภาพที่ 3.13 แสดงข้อมูลกราฟของการใช้อาวุธ

3.2 การออกแบบเว็บไซต์

3.2.1 การออกแบบ Wireframe หน้าจอเว็บไซต์

1. หน้าแรก แสดงข้อมูลเกี่ยวกับอาชญากรรม

ภาพที่ 3.14 หน้าแรก

2. หน้าข้อมูลการวิเคราะห์โมเดล Decision Tree

ภาพที่ 3.15 หน้าข้อมูลการวิเคราะห์โมเดล Decision Tree

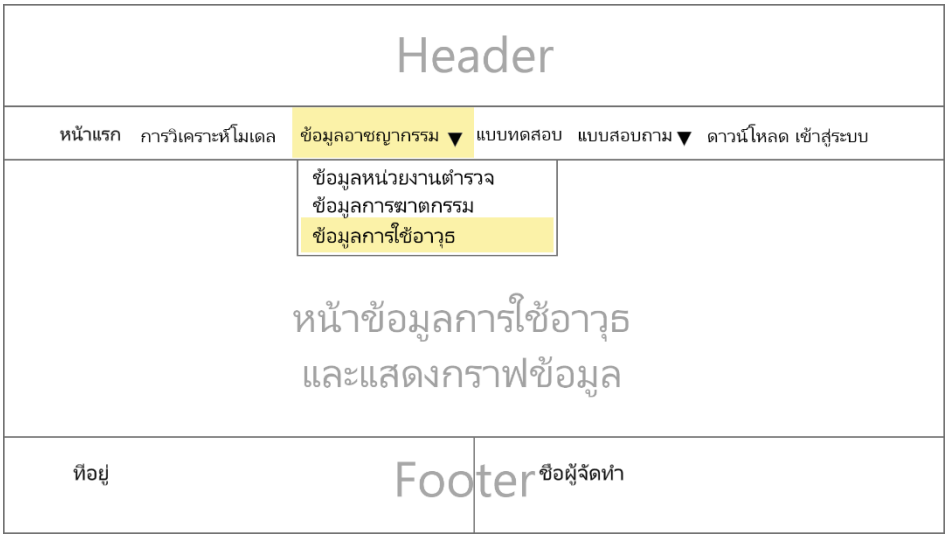
3. หน้าข้อมูลอาชญากรรม ของ ข้อมูลหน่วยงานตำรวจ

ภาพที่ 3.16 หน้าข้อมูลอาชญากรรม ของ ข้อมูลหน่วยงานตำรวจ

4. หน้าข้อมูลอาชญากรรม ของ ข้อมูลการฆาตกรรม

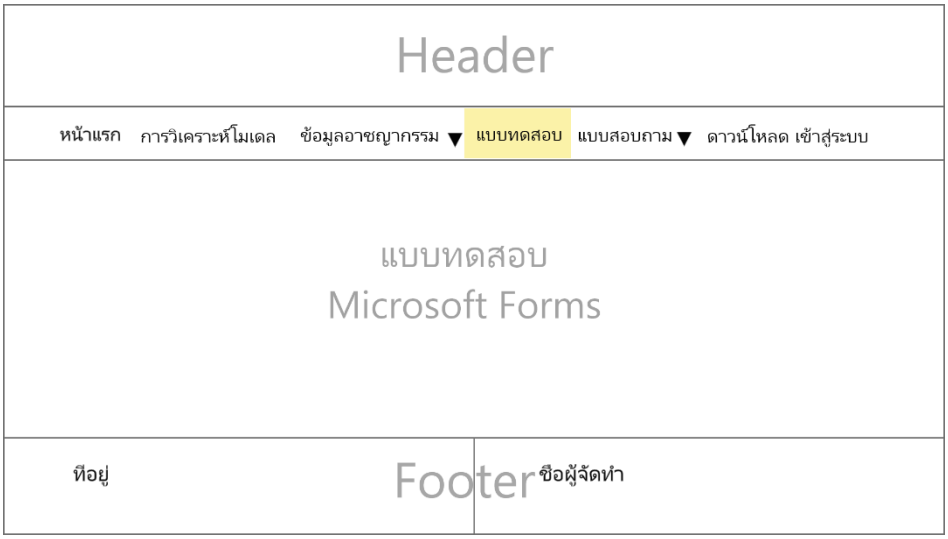
ภาพที่ 3.17 หน้าข้อมูลอาชญากรรม ของ ข้อมูลการฆาตกรรม

5. หน้าข้อมูลอาชญากรรม ของ ข้อมูลการใช้อาวุธ



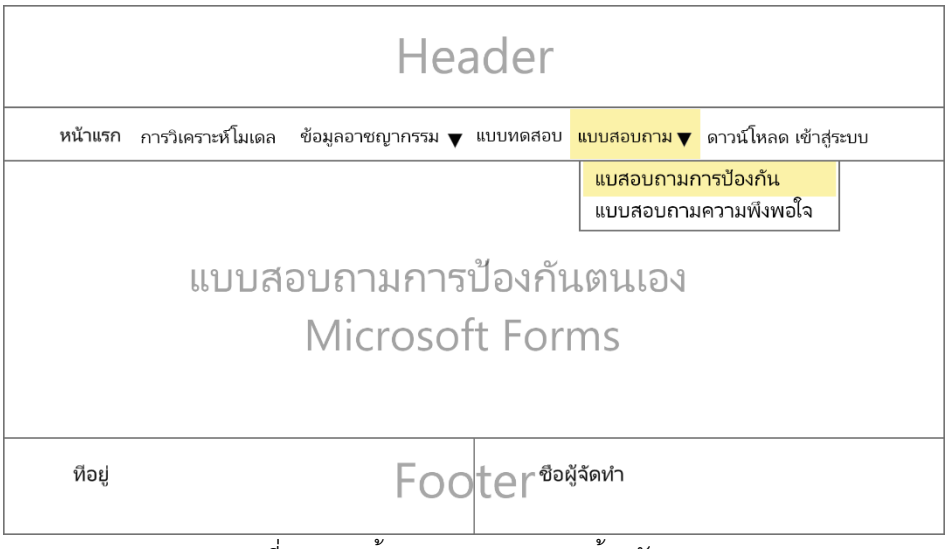
ภาพที่ 3.18 หน้าข้อมูลอาชญากรรม ของ ข้อมูลการใช้อาวุธ

6. หน้าแบบทดสอบ



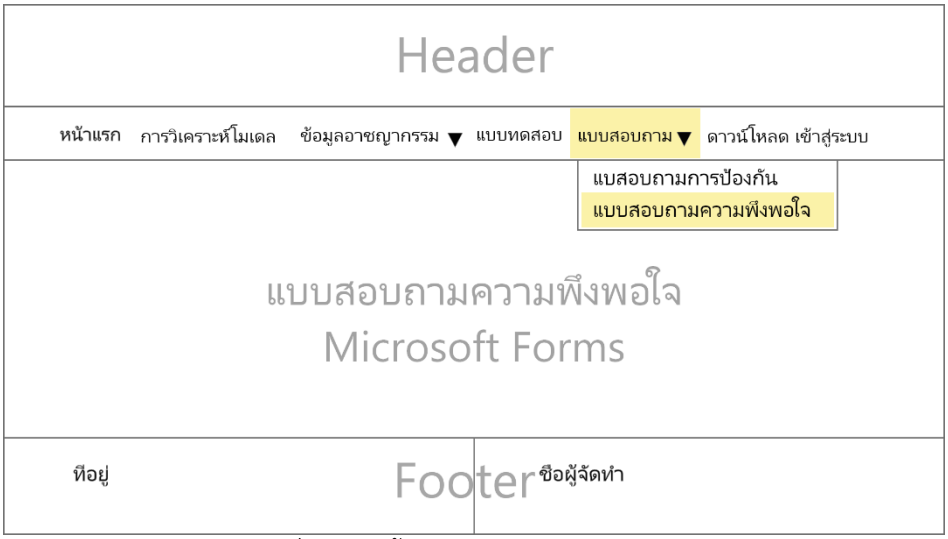
ภาพที่ 3.19 หน้าแบบทดสอบ

7. หน้าแบบสอบถาม การป้องกันตนเอง



ภาพที่ 3.20 หน้าแบบสอบถาม การป้องกันตนเอง

8. หน้าแบบสอบถาม ความพึงพอใจ

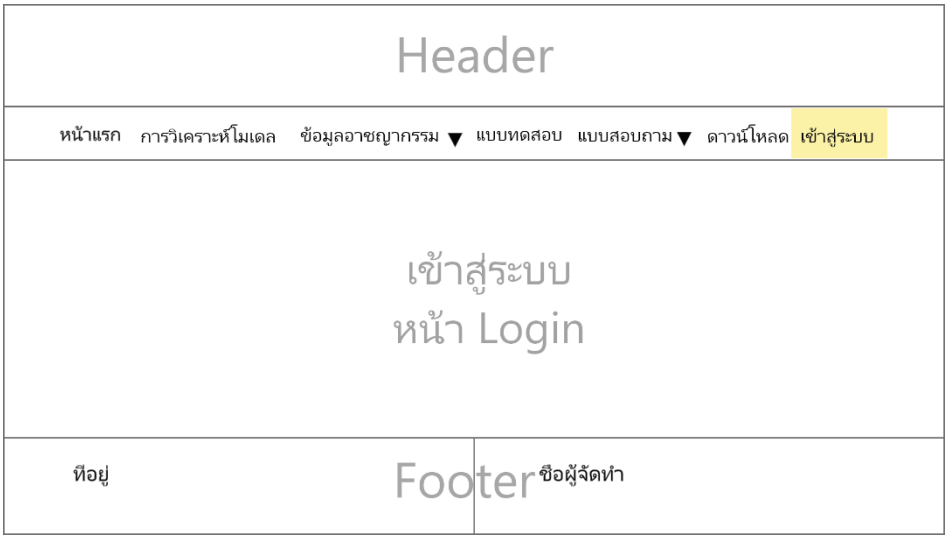


ภาพที่ 3.21 หน้าแบบสอบถาม ความพึงพอใจ

9. หน้าดาวน์โหลดไฟล์ข้อมูล

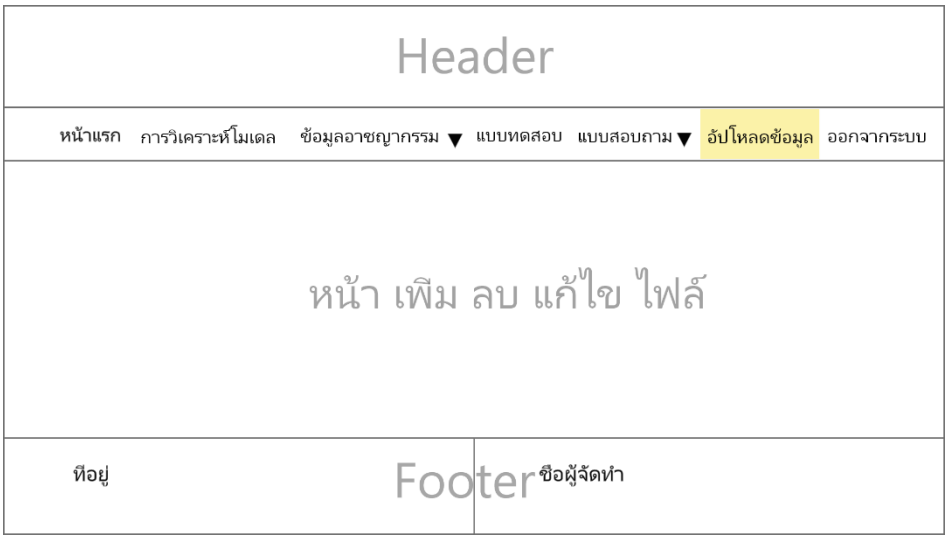
ภาพที่ 3.22 หน้าดาวน์โหลดไฟล์ข้อมูล

10. หน้าเข้าสู่ระบบ



ภาพที่ 3.23 หน้าเข้าสู่ระบบ

11. หน้าจัดการไฟล์ เพิ่ม ลบ แก้ไข ข้อมูล



ภาพที่ 3.24 หน้าจัดการไฟล์ เพิ่ม ลบ แก้ไข ข้อมูล

3.3 บทสรุป

จากวิธีการดำเนินงานโครงงานในข้างต้น ผู้วิเคราะห์ข้อมูลได้แสดงวิธีในการจัดการกับการวิเคราะห์ข้อมูลรายงานการฆาตกรรมปี 2011–2013 ด้วยขั้นตอนกระบวนการวิเคราะห์ข้อมูลด้วย CRISP–DM รวมถึงการสร้างโมเดล Decision Tree จากโปรแกรมที่ใช้ทำเหมืองข้อมูลโดยโปรแกรม RapidMiner Studio ในการสร้างโมเดล Decision Tree ซึ่งได้ผลลัพธ์ที่

เหมือนกัน และนำข้อมูลสารสนเทศมาทำการแสดงผลแบบ visualization ในรูปแบบของภาพ โดยใช้โปรแกรม Tableau Public และออกแบบ Wireframe ของเว็บไซต์ที่จะเผยแพร่บน web browser ด้วยโปรแกรม Adobe XD ปัญหาที่พบในบทนี้ คือการทำ Decision Tree ผ่านโปรแกรม weka ซึ่งพบปัญหาคือ ผลลัพธ์ที่ผ่านกระบวนการทำ Decision Tree นั้น รูปแบบโมเดลไม่ตรง และแสดงผลแอตทริบิวต์ออกมาไม่ครบตามที่คำนวณและ code คำบรรยายลักษณะงานก็แสดงผลออกมาไม่ตรบแอตทริบิวต์ ผู้วิเคราะห์ข้อมูลจึงได้ใช้แค่ โปรแกรม RapidMiner Studio ในการทำ Decision Tree

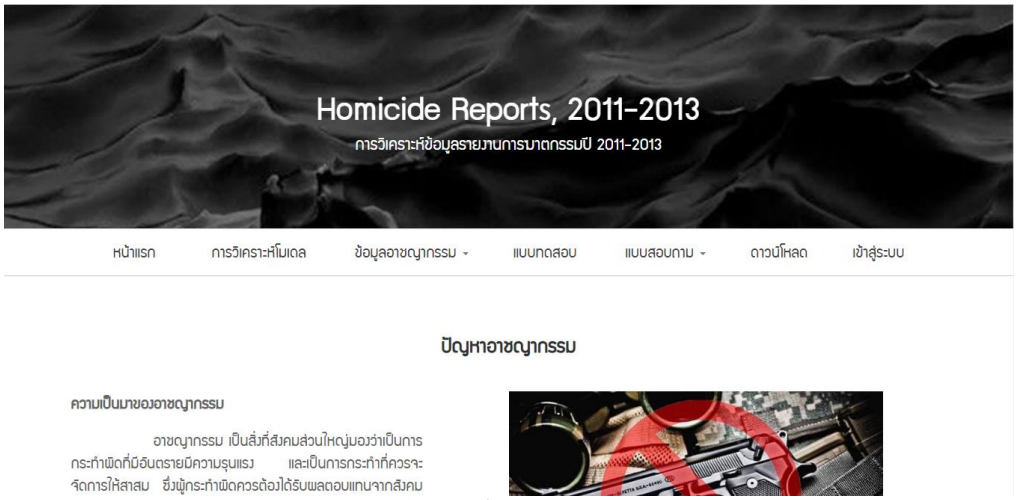
บทที่ 4

ผลการดำเนินงาน

จากการวิเคราะห์ข้อมูลรายงานการฆาตกรรมปี 2011–2013 เพื่อใช้สำหรับเผยแพร่ข้อมูลบนเว็บไซต์ มีวัตถุประสงค์เพื่อวิเคราะห์ข้อมูลรายงานการฆาตกรรมปี 2011–2013 สำหรับเผยแพร่ข้อมูลบนเว็บไซต์ เพื่อให้ได้เว็บไซต์ที่เป็นประโยชน์ในการรวบรวมข้อมูล และแสดงกระบวนการวิเคราะห์ข้อมูลที่สามารถนำไปใช้งานได้อย่างมีประสิทธิภาพและประสิทธิผล โดยผู้จัดทำโครงงานได้ทำการวิเคราะห์ข้อมูล และนำเทคโนโลยีเข้าช่วยในการเสนอข้อมูล ทำให้มีความสะดวกรวดเร็วและสามารถใช้งานได้จริงจนสำเร็จลุล่วงตามเป้าหมาย

4.1 ผลการดำเนินงาน

การจัดทำโครงงานเรื่อง การวิเคราะห์ข้อมูลรายงานการฆาตกรรมปี 2011–2013 เพื่อใช้สำหรับเผยแพร่ข้อมูลบนเว็บไซต์ ในครั้งนี้พบว่าการจำแนกข้อมูลออกเป็นประเภทต่าง ๆ ช่วยทำให้เกิดความเข้าใจต่อข้อมูลได้ง่ายมากขึ้น และสิ่งที่ต้องการทราบอย่างแท้จริง ผู้จัดทำได้ทำการวิเคราะห์ข้อมูลให้อยู่ในรูปแบบของ Data Mining โดยทำการจำแนกข้อมูลออกเป็นประเภทต่าง ๆ ด้วยเทคนิคการทำเหมืองข้อมูลแบบ Classification จากเทคนิคการวิเคราะห์ข้อมูลในรูปแบบของ Decision tree และนำเสนอข้อมูลแบบ Visualization เผยแพร่ข้อมูลสารสนเทศบน Web browser ให้กับผู้ใช้งาน ผู้จัดทำจึงได้ทำการวิเคราะห์ข้อมูล และจัดทำเว็บไซต์ตามวัตถุประสงค์ที่ได้กำหนดไว้ และมีผลต่อการดำเนินงาน ดังนี้



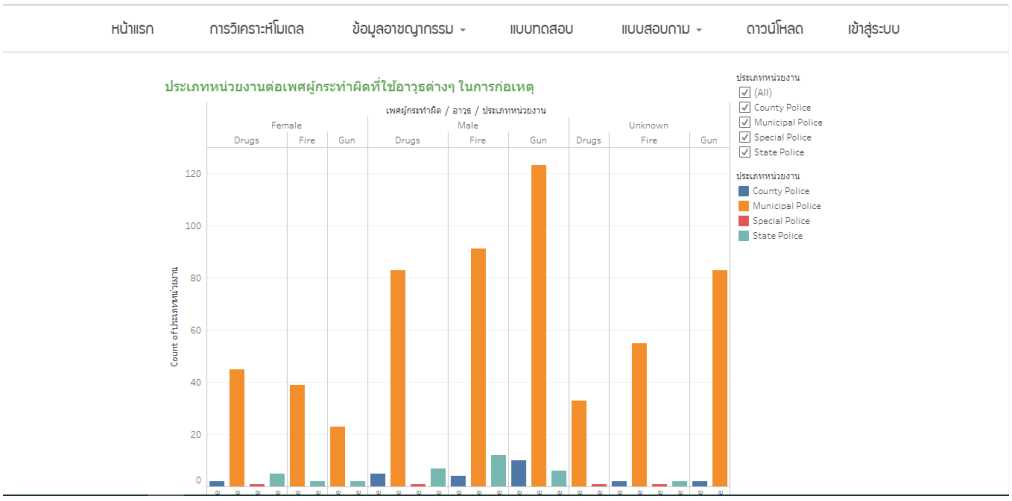
ภาพที่ 4.1 หน้าแรกของเว็บไซต์

คำอธิบายการใช้งาน : แถบเมนูผู้ใช้งานสามารถเลือกดูในแต่ละส่วนได้ตามความต้องการของผู้ใช้งานได้

ภาพที่ 4.2 หน้าเมนูการวิเคราะห์โมเดล

คำอธิบายการใช้งาน : เป็นหน้ารายละเอียดเกี่ยวกับการวิเคราะห์การทำโมเดลแบบ Decision tree มีวิธีการคิดแบบคำนวณมือและการวิเคราะห์โดยใช้โปรแกรม

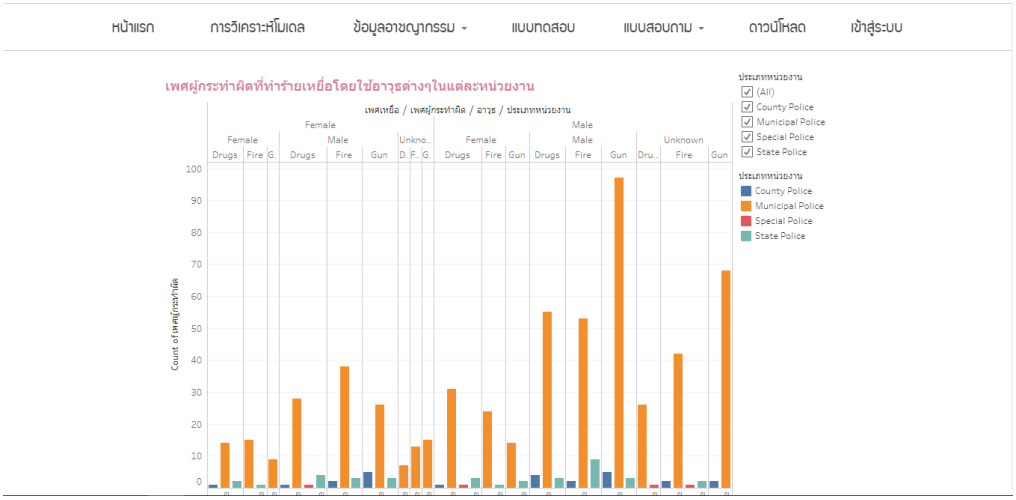
ภาพที่ 4.3 หน้าข้อมูลอาชญากรรม ข้อมูลหน่วยงานตำรวจ



ภาพที่ 4.4 หน้าข้อมูลอาชญากรรม ข้อมูลหน่วยงานตำรวจ และแสดง Dashboard

คำอธิบายการใช้งาน : เป็นหน้ารายละเอียดเกี่ยวกับหน่วยงานตำรวจและแสดง Dashboard ข้อมูลของแต่ละหน่วยงาน

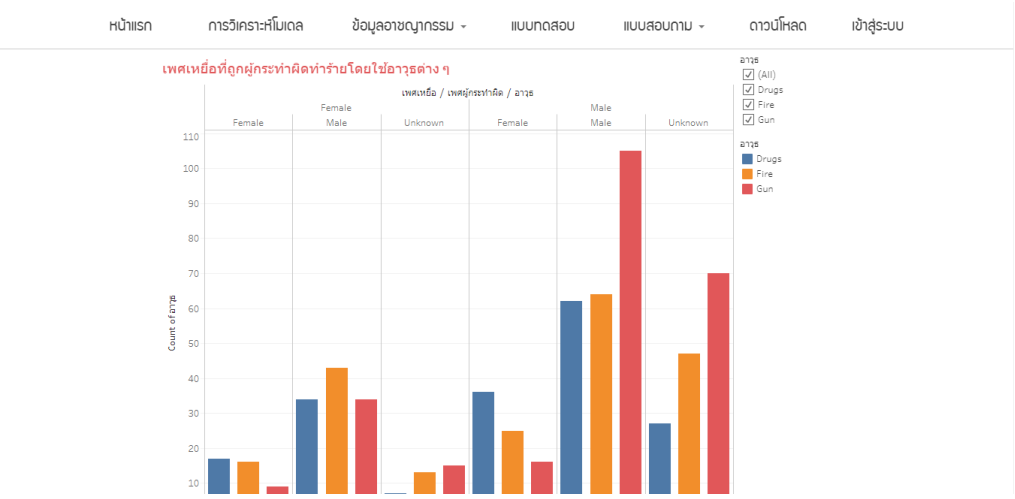
ภาพที่ 4.5 หน้าข้อมูลอาชญากรรม ข้อมูลการฆาตกรรม



ภาพที่ 4.6 หน้าข้อมูลอาชญากรรม ข้อมูลการฆาตกรรม และแสดง Dashboard

คำอธิบายการใช้งาน : เป็นหน้ารายละเอียดเกี่ยวกับการฆาตกรรมและแสดง Dashboard ข้อมูลของแต่ละหน่วยงาน ในแต่ละหน่วยงานมีผู้กระทำผิดอย่างไรบ้าง

ภาพที่ 4.7 หน้าข้อมูลอาชญากรรม ข้อมูลการใช้อาวุธ



ภาพที่ 4.8 หน้าข้อมูลอาชญากรรม ข้อมูลการใช้อาวุธ และแสดง Dashboard

คำอธิบายการใช้งาน : เป็นหน้ารายละเอียดเกี่ยวกับอาวุธที่ผู้กระทำผิดทั้งหมดก่อเหตุและแสดง Dashboard ข้อมูลอาวุธ

ภาพที่ 4.9 หน้าแบบทดสอบ

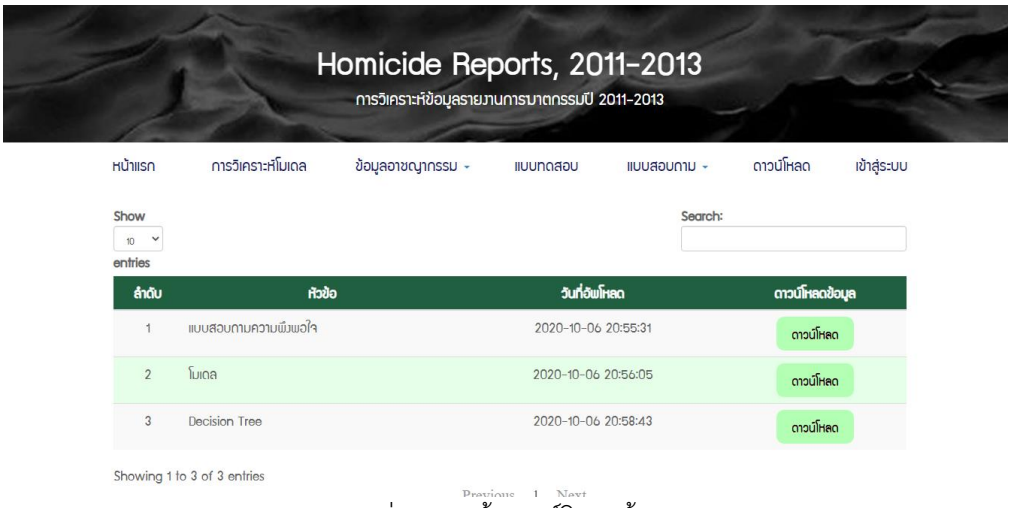
คำอธิบายการใช้งาน : หน้าแบบทดสอบความรู้เรื่องอาชญากรรม ประเมินความรู้ตนเองในเรื่องอาชญากรรม

ภาพที่ 4.10 หน้าแบบสอบถาม การป้องกันตนเอง

คำอธิบายการใช้งาน : หน้าแบบสอบถามข้อมูลการป้องกันตนเองไม่ให้ตกเป็นเหยื่อของอาชญากรรม

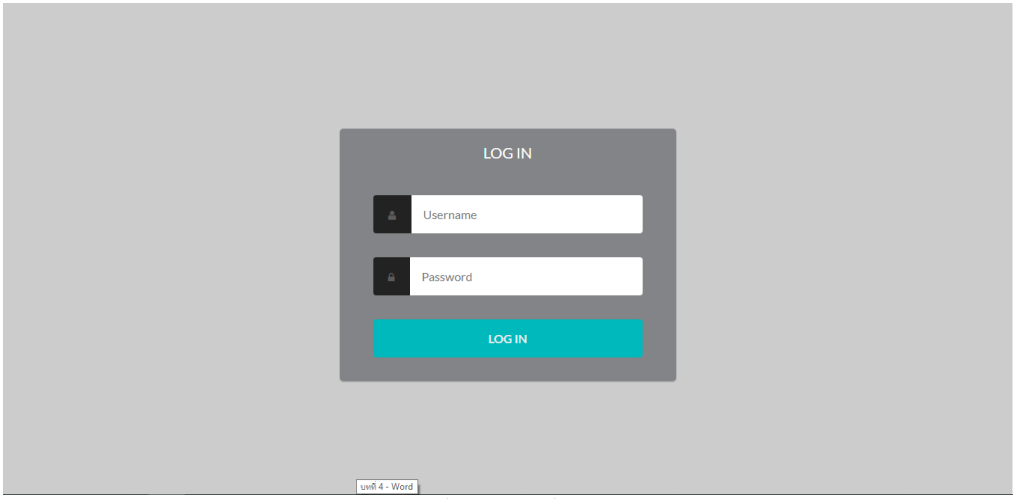
ภาพที่ 4.11 หน้าแบบสอบถาม ความพึงพอใจ

คำอธิบายการใช้งาน : หน้าแบบสอบถามความพึงพอใจในการใช้งานเว็บไซต์ การวิเคราะห์ข้อมูลรายงานการฆาตกรรมปี 2011-2013



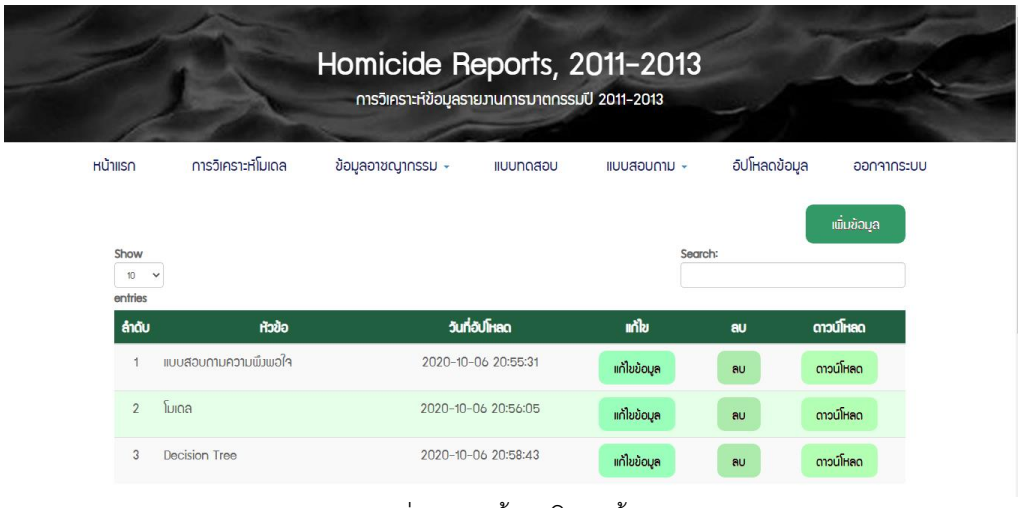
ภาพที่ 4.12 หน้าดาวน์โหลดข้อมูล

คำอธิบายการใช้งาน : หน้าดาวน์โหลดข้อมมูลที่ผู้ใช้งานสามารถดาวน์โหลดไฟล์ข้อมูลต่าง ๆ ที่ต้องการใช้งานได้



ภาพที่ 4.13 หน้า Login เข้าสู่ระบบ Admin

คำอธิบายการใช้งาน : หน้า Login เข้าสู่ระบบเฉพาะ Admin ที่สามารถเข้าใช้งานได้เฉพาะ Admin



ภาพที่ 4.14 หน้าอัปโหลดข้อมูล

คำอธิบายการใช้งาน : หน้าอัปโหลดข้อมูล ผู้เป็น Admin สามารถ เพิ่ม ลบ แก้ไขข้อมูลได้

4.2 การอภิปรายผลการดำเนินโครงงาน

จากวัตถุประสงค์เพื่อวิเคราะห์ข้อมูลรายงานการฆาตกรรมปี 2011 – 2013 เพื่อใช้สำหรับเผยแพร่ข้อมูลบนเว็บไซต์ ผู้จัดทำได้ศึกษาปัญหา และเก็บรวบรวมข้อมูล โดยใช้กระบวนการวิเคราะห์ข้อมูลด้วย CRISP–DM เพื่อใช้ในการวิเคราะห์ขอบเขตของโครงงาน ดังนี้

4.2.1 ขอบเขตผู้จัดทำ

1. รู้จักและเข้าใจในธุรกิจ (Business understanding) ผลการศึกษาพบว่า ข้อมูลรายงานการฆาตกรรมปี 2011–2013 ซึ่งมีข้อมูลทั้งหมด 638,455 รายการ ซึ่งเป็นข้อมูลที่มีจำนวนมาก ซึ่งไม่สามารถทำความเข้าใจกับข้อมูลได้อย่างรวดเร็ว และการเผยแพร่ข้อมูลรายงานการฆาตกรรมปี 2011–2013 ให้กับผู้ใช้ เพื่อตระหนักถึงการก่ออาชญากรรมที่ยังไม่ค่อยมีประสิทธิภาพที่ดีมาก

2. สร้างฐานข้อมูลให้ครบ (Data understanding) ผู้จัดทำทำการรวบรวมข้อมูลรายงานการฆาตกรรมปี 2011–2013 ที่ได้จากเว็บไซต์ kaggle.com

3. เตรียมข้อมูลให้พร้อมใช้ (Data preparation) ผู้จัดทำทำการคัดเลือกข้อมูลโดยมี ส.ต.ต. วโรดม ไชยชนะ มาช่วยในการพิจารณา และทำการ Data Cleaning เพื่อให้ได้ข้อมูลที่สมบูรณ์

4. สร้างแบบจำลอง (Modeling) ผู้จัดทำวิเคราะห์ข้อมูลด้วยเทคนิคทางดาต้าไมน์นิง แบบการจำแนกประเภทข้อมูล (Classification) โดยการใช้โมเดลการตัดสินใจแบบต้นไม้ (Decision Tree) เพื่อจัดกลุ่มข้อมูลที่มีความสัมพันธ์กัน ประกอบด้วย 6 แอตทริบิวต์ Agency Type(ประเภทตัวแทน) Year(ปี) Crime Type(ประเภทอาชญากรรม) Victim Sex(เพศเหยื่อ) Perpetrator Sex(เพศผู้กระทำผิด)และWeapon(อาวุธ)

5. การประเมินผล (Evaluation) ผู้จัดทำได้ทำการทำการทดสอบโมเดล เพื่อวัดประสิทธิภาพที่ตรงกับความต้องการ ซึ่งการวัดประสิทธิภาพด้วยโปรแกรม RapidMiner ให้ผลลัพธ์การจำแนกประเภท เพศผู้กระทำผิด และได้กฎจำนวน 18 กฎ

6. เผยแพร่ผลวิเคราะห์ (Deployment) ผู้จัดทำได้จัดทำข้อมูลที่ทำการวิเคราะห์มาแสดงผลข้อมูลบนหน้า Web Browser โดยใช้โปรแกรม Adobe Dreamweaver ร่วมกับการนำเสนอข้อมูลแบบ visualization ด้วยการแสดงผลข้อมูลในรูปแบบของภาพโดยใช้โปรแกรม Tableau Public โดยผู้จัดทำได้ใช้ภาษา HTML CSS PHP และ JavaScript ในการสร้างเว็บไซต์ขึ้นมา

4.2.1.1 ขอบเขตผู้ใช้งานทั่วไปบนเว็บบราวเซอร์

1. ผู้ใช้งานสามารถดูข้อมูลรายงานการฆาตกรรมปี 2011–2013 และปัญหาหรือการป้องกันการโดนก่อเหตุอาชญากรรมผ่านทางเว็บไซต์

2. ผู้ใช้งานสามารถดูแผนภาพและกรองข้อมูลสารสนเทศผ่านระบบ Tableau Public ผ่านทางเว็บไซต์ได้

3. ผู้ใช้งานสามารถดาวน์โหลดไฟล์ข้อมูลผ่านเว็บไซต์ได้

4.2.2 การอภิปรายผลการประเมินแบบสอบถาม

จากที่ผู้จัดทำได้ทำการวิเคราะห์ข้อมูลและจัดทำเว็บไซต์เพื่อเผยแพร่ข้อมูลรายงานการฆาตกรรมปี 2011–2013 จนแล้วเสร็จ จากนั้น จึงได้ทำการประเมินผลการใช้งานเว็บไซต์จากกลุ่มตัวอย่างจำนวน 35 คน โดยทำแบบสอบถามความพึงพอใจในการเข้าใช้งานเว็บไซต์เผยแพร่ข้อมูลรายงานการฆาตกรรมปี 2011–2013 จากกลุ่มตัวอย่างจำนวน 35 คน แบ่งเป็นเพศหญิง 22 คน คิดเป็นร้อยละ 62.9 เพศชายจำนวน 13 คน คิดเป็นร้อยละ 37.1 และ อายุตั้งแต่ 21–25 ปี 30 คน คิดเป็นร้อยละ 85.7 อายุตั้งแต่ 26–30 ปี 4 คน คิดเป็นร้อยละ 11.4 อายุตั้งแต่ 31–35 ปี 1 คน คิดเป็นร้อยละ 2.9 โดยเกณฑ์การให้คะแนนของผู้ใช้งานเว็บไซต์ มีดังนี้ ระดับความพึงพอใจ 5 = ดีมาก 4 = ดี 3 = ปานกลาง 2 = น้อย 1 = ควรปรับปรุง

เมื่อนำคำตอบของผู้ตอบแบบสอบถามมาแจกแจงความถี่และหาค่าเฉลี่ย แล้วกำหนดระดับค่าเฉลี่ย แบ่งระดับความพึงพอใจได้ดังนี้

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

จากเกณฑ์ดังกล่าว สามารถแปลความหมายของความพึงพอใจได้ ดังนี้

คะแนนค่าเฉลี่ย 1.00 – 1.80 หมายถึง ระดับความพึงพอใจควรปรับปรุง

คะแนนค่าเฉลี่ย 1.81 – 2.60 หมายถึง ระดับความพึงพอใจน้อย

คะแนนค่าเฉลี่ย 2.61 – 3.40 หมายถึง ระดับความพึงพอใจปานกลาง

คะแนนค่าเฉลี่ย 3.41 – 4.20 หมายถึง ระดับความพึงพอใจดี

คะแนนค่าเฉลี่ย 4.21 – 5.00 หมายถึง ระดับความพึงพอใจดีมาก

ดังนั้นผู้วิจัยจึงนำค่าคะแนนมาคำนวณทางสถิติ คือค่าคะแนนเฉลี่ย ($\bar{x}$) และค่าส่วนเบี่ยงเบนมาตรฐาน (S.D.) ผลการประเมินรายละเอียดดังนี้ การประเมินด้านเนื้อหา แสดงดังตารางที่ 4.1 การประเมินด้านการออกแบบเว็บไซต์ แสดงดังตารางที่ 4.2 และการประเมินด้านประโยชน์และการนำไปใช้งาน แสดงดังตารางที่ 4.3

ตารางที่ 4.1 แสดงข้อมูลการประเมินผลด้านเนื้อหา

รายการประเมิน	$\bar{x}$	S.D.	แปลผล
1.เนื้อหามีความถูกต้อง	4.43	0.50	ดีมาก
2.เนื้อหามีความกะทัดรัด เข้าใจง่าย	4.26	0.66	ดีมาก
3.ข้อมูลภายในเว็บไซต์มีความน่าเชื่อถือ	4.31	0.63	ดีมาก
4.รูปแบบการนำเสนอ visualization มีความเหมาะสมเข้าใจง่าย	4.34	0.76	ดีมาก

จากตารางที่ 4.1 พบว่าข้อมูลที่มีเนื้อหามีความถูกต้อง มีค่าเฉลี่ยสูงสุดเท่ากับ 4.43 (S.D. = 0.50) อยู่ในเกณฑ์ระดับดีมาก เนื้อหามีความกะทัดรัด เข้าใจง่าย มีค่าเฉลี่ยสูงสุดเท่ากับ 4.26 (S.D. = 0.66) อยู่ในเกณฑ์ระดับดีมาก ข้อมูลภายในเว็บไซต์มีความน่าเชื่อถือ มีค่าเฉลี่ยสูงสุดเท่ากับ 4.31 (S.D. = 0.63) อยู่ในเกณฑ์ระดับดีมาก รูปแบบการนำเสนอ visualization มีความเหมาะสมเข้าใจง่าย มีค่าเฉลี่ยสูงสุดเท่ากับ 4.34 (S.D. = 0.76) ตามลำดับ

ตารางที่ 4.2 แสดงข้อมูลการประเมินผลด้านการออกแบบเว็บไซต์

รายการประเมิน	$\bar{x}$	S.D.	แปลผล
1.เว็บไซต์ง่ายต่อการใช้งาน	4.49	0.66	ดีมาก
2.มีความเหมาะสมในการใช้ชนิด ขนาด สี ตัวอักษรบนเว็บไซต์	4.31	0.58	ดีมาก
3.มีความเหมาะสมในการจัดตำแหน่ง การออกแบบไม่ซับซ้อน	4.40	0.74	ดีมาก
4.มีความเหมาะสมในการใช้สัญลักษณ์หรือรูปถาพในการสื่อความหมาย	4.26	0.66	ดีมาก
5.มีความเหมาะสมในการจัดวางตำแหน่งของส่วนประกอบเมนูต่าง ๆ บนเว็บไซต์	4.34	0.64	ดีมาก

จากตารางที่ 4.2 พบว่าข้อมูลเว็บไซต์ง่ายต่อการใช้งาน มีค่าเฉลี่ยสูงสุดเท่ากับ 4.49 (S.D. = 0.66) อยู่ในเกณฑ์ระดับดีมาก มีความเหมาะสมในการใช้ชนิด ขนาด สีตัวอักษรบนเว็บไซต์ มีค่าเฉลี่ยสูงสุดเท่ากับ 4.31 (S.D. = 0.58) อยู่ในเกณฑ์ระดับดีมาก มีความเหมาะสมในการจัดตำแหน่ง การออกแบบไม่ซับซ้อน มีค่าเฉลี่ยสูงสุดเท่ากับ 4.40 (S.D. = 0.74) มีความเหมาะสมในการใช้สัญลักษณ์หรือรูปถาพในการสื่อความหมาย มีค่าเฉลี่ยสูงสุดเท่ากับ 4.26 (S.D. = 0.66) มีความเหมาะสมในการจัดวางตำแหน่งของส่วนประกอบเมนูต่าง ๆ บนเว็บไซต์ มีค่าเฉลี่ยสูงสุดเท่ากับ 4.34 (S.D. = 0.64) ตามลำดับ

ตารางที่ 4.3 แสดงข้อมูลการประเมินผลด้านประโยชน์และการนำไปใช้

รายการประเมิน	$\bar{x}$	S.D.	แปลผล
1.เนื้อหามีประโยชน์ต่อการนำไปประยุกต์ใช้	4.31	0.72	ดีมาก
2.สามารถนำความรู้ที่ได้รับไปประยุกต์ใช้ในชีวิตประจำวันได้	4.11	0.72	ดีมาก
3.สามารถเป็นแหล่งความรู้และแนวทางในการพัฒนาต่อไปได้	4.23	0.77	ดีมาก
4.แหล่งข้อมูลตรงกับความต้องการของผู้ใช้งาน	4.03	0.82	ดีมาก

จากตารางที่ 4.3 พบว่าข้อมูลที่มีเนื้อหามีประโยชน์ต่อการนำไปประยุกต์ใช้ มีค่าเฉลี่ยสูงสุดเท่ากับ 4.31 (S.D. = 0.72) อยู่ในเกณฑ์ระดับดีมาก สามารถนำความรู้ที่ได้รับไปประยุกต์ใช้ในชีวิตประจำวันได้ มีค่าเฉลี่ยสูงสุดเท่ากับ 4.11 (S.D. = 0.72) อยู่ในเกณฑ์ระดับดีมาก สามารถเป็นแหล่งความรู้และแนวทางในการพัฒนาต่อไปได้ มีค่าเฉลี่ยสูงสุดเท่ากับ 4.23 (S.D. = 0.77) อยู่ในเกณฑ์ระดับดีมาก แหล่งข้อมูลตรงกับความต้องการของผู้ใช้งาน มีค่าเฉลี่ยสูงสุดเท่ากับ 4.03 (S.D. = 0.82) ตามลำดับ

4.3 บทสรุป

จากผลการดำเนินงานโครงงานข้างต้น ผู้จัดทำได้นำข้อมูลมาวิเคราะห์ และแสดงผลในรูปแบบ visualization และวิธีการขั้นตอนในการสร้าง Decision Tree มาเผยแพร่ข้อมูลรายงานการฆาตกรรมปี 2011 – 2013 ในรูปแบบของเว็บไซต์ ที่เป็นแหล่งศึกษาค้นคว้าหาความรู้ และยังมีแบบทดสอบและแบบสอบถามให้ผู้เข้าชมเว็บไซต์ได้ตอบคำถามในเรื่องความรู้การป้องกันตนเองจากอาชญากรรม เพื่อสามารถนำไปวิเคราะห์ข้อมูลต่อได้

บทที่ 5
สรุปและข้อเสนอแนะ

ผลการวิเคราะห์ข้อมูลรายงานการฆาตกรรมปี 2011–2013 เพื่อใช้สำหรับเผยแพร่ข้อมูลบนเว็บไซต์ ทางผู้จัดทำได้ทำการประเมินผลการใช้งานเว็บไซต์จากผู้ใช้งานทั่วไป เพื่อสรุปผลการทำโครงงาน ข้อจำกัดของเว็บไซต์ ปัญหาและอุปสรรคต่าง ๆ ของการทำโครงงาน และข้อเสนอแนะเพื่อพัฒนาโครงงาน ดังนี้

5.1 บทสรุปผลโครงงาน

จากการที่ผู้จัดทำได้ทำการวิเคราะห์ข้อมูลรายงานการฆาตกรรมปี 2011–2013 เพื่อใช้สำหรับเผยแพร่ข้อมูลบนเว็บไซต์ จนแล้วเสร็จ จากนั้น จึงได้ทำการประเมินผลการใช้งานเว็บไซต์จากกลุ่มตัวอย่างจำนวน 35 คน โดยทำแบบสอบถามความพึงพอใจในการเข้าใช้งานเว็บไซต์เผยแพร่ข้อมูลรายงานการฆาตกรรมปี 2011–2013 จากกลุ่มตัวอย่างจำนวน 35 คน แบ่งเป็นเพศหญิง 22 คน คิดเป็นร้อยละ 62.9 เพศชายจำนวน 13 คน คิดเป็นร้อยละ 37.1 และอายุตั้งแต่ 21–25 ปี 30 คน คิดเป็นร้อยละ 85.7 อายุตั้งแต่ 26–30 ปี 4 คน คิดเป็นร้อยละ 11.4 อายุตั้งแต่ 31–35 ปี 1 คน คิดเป็นร้อยละ 2.9

ผลสรุปค่าเฉลี่ยได้ดังนี้ ข้อมูลการประเมินผลด้านเนื้อหา พิจารณาเป็นรายข้อ ข้อมูลที่มีเนื้อหามีความถูกต้อง ($\overline{X}$=4.43,S.D.=0.50) อยู่ในเกณฑ์ระดับดีมาก เนื้อหามีความกะทัดรัด เข้าใจง่าย ($\overline{X}$=4.26,S.D. = 0.66) อยู่ในเกณฑ์ระดับดีมาก ข้อมูลภายในเว็บไซต์มีความน่าเชื่อถือ ($\overline{X}$=4.31,S.D. = 0.63) อยู่ในเกณฑ์ระดับดีมาก รูปแบบการนำเสนอ visualization มีความเหมาะสมเข้าใจง่าย ($\overline{X}$=4.34,S.D. = 0.76) ตามลำดับ ข้อมูลการประเมินผลด้านการออกแบบเว็บไซต์ พิจารณาเป็นรายข้อ ข้อมูลเว็บไซต์ง่ายต่อการใช้งาน ($\overline{X}$=4.49,S.D. = 0.66) อยู่ในเกณฑ์ระดับดีมาก มีความเหมาะสมในการใช้ชนิด ขนาด สีตัวอักษรบนเว็บไซต์ ($\overline{X}$=4.31,S.D. = 0.58) อยู่ในเกณฑ์ระดับดีมาก มีความเหมาะสมในการจัดตำแหน่ง การออกแบบไม่ซับซ้อน ($\overline{X}$=4.40,S.D. = 0.74) มีความเหมาะสมในการใช้สัญลักษณ์

หรือรูปถาพในการสื่อความหมาย ($\overline{x}$=4.26,S.D. = 0.66) มีความเหมาะสมในการจัดวางตำแหน่งของส่วนประกอบเมนูต่าง ๆ บนเว็บไซต์ ($\overline{x}$=4.34,S.D. = 0.64) ตามลำดับ ข้อมูลการประเมินผลด้านประโยชน์และการนำไปใช้ พิจารณาเป็นรายข้อ ข้อมูลที่มีเนื้อหามีประโยชน์ต่อการนำไปประยุกต์ใช้ ($\overline{x}$=4.31,S.D. = 0.72) อยู่ในเกณฑ์ระดับดีมาก สามารถนำความรู้ที่ได้รับไปประยุกต์ใช้ในชีวิตประจำวันได้ ($\overline{x}$=4.11,S.D. = 0.72) อยู่ในเกณฑ์ระดับดีมาก สามารถเป็นแหล่งความรู้และแนวทางในการพัฒนาต่อไปได้ ($\overline{x}$=4.23,S.D. = 0.77) อยู่ในเกณฑ์ระดับดีมาก แหล่งข้อมูลตรงกับความต้องการของผู้ใช้งาน ($\overline{x}$=4.03,S.D. = 0.82) ตามลำดับ

5.2 ข้อจำกัดของเว็บไซต์

5.2.1 ผู้ใช้งานไม่สามารถปรับเปลี่ยนรูปแบบการแสดงกราฟข้อมูลได้ เช่น ผู้ใช้ไม่สามารถปรับเปลี่ยนกราฟจากกราฟแท่งไปเป็นกราฟเส้นหรือในรูปแบบอื่น ๆ ได้

5.2.2 การแสดงผลข้อมูลอาจไม่ครอบคลุมทั้งหมดทำให้เว็บไซต์เผยแพร่ข้อมูลไม่อาจตอบสนองความต้องการให้ผู้ใช้อย่างครอบคลุม เช่น เว็บไซต์ไม่มีข้อมูลที่เป็นปัจจุบัน

5.2.3 ผู้เข้าชมไม่สามารถอัปโหลดชุดข้อมูลเข้าไปในเว็บไซต์ได้ เนื่องจากเป็นเว็บไซต์เผยแพร่ข้อมูลได้เพียงอย่างเดียว

5.2.4 ผู้เข้าชมไม่สามารถรับการแจ้งเตือนการอัปเดทชุดข้อมูลใหม่หรือการปรับปรุงข้อมูลในเว็บไซต์ได้

5.2.5 ผู้เข้าชมไม่สามารถเพิ่มลบหรือแก้ไขชุดข้อมูลที่อัปโหลดในเว็บไซต์ได้ เนื่องจากยังขาดการเข้าถึงในส่วนของผู้เข้าชมที่จะช่วยในการอัปโหลดเพิ่ม ลบ แก้ไข ชุดข้อมูล

5.3 ปัญหาและอุปสรรคของโครงงาน

5.3.1 แหล่งข้อมูลที่ใช้ในการวิเคราะห์เป็นข้อมูลที่ไม่ได้จัดเก็บด้วยตนเองจึงต้องใช้ระยะเวลาในการศึกษาชุดข้อมูล

5.3.2 ชุดข้อมูลที่ใช้ในการวิเคราะห์เป็นข้อมูลที่ไม่เป็นปัจจุบัน และมีค่าซ้ำซ้อนจึงต้องใช้เวลาในการทำความสะอาดข้อมูล

5.3.3 การกำจัดหรือการทำความสะอาดชุดข้อมูล อาจส่งผลให้ข้อมูลบางส่วนคลาดเคลื่อนต่างไปบ้างเล็กน้อย

5.3.4 โปรแกรมบางโปรแกรมต้องจ่ายเงินเพื่อซื้อโปรแกรมมาวอเคราะห์ข้อมูลที่เป็นภาษาไทย

5.3.5 โปรแกรมต่าง ๆ ที่นำมาใช้ในการวิเคราะห์เป็นโปรแกรมที่ไม่เสียค่าใช้จ่าย จึงอาจมีข้อจำกัดในการใช้งานโปรแกรม

5.4 ข้อเสนอแนะ

การวิเคราะห์ข้อมูลรายงานการฆาตกรรมปี 2011–2013 เพื่อใช้สำหรับเผยแพร่ข้อมูลบนเว็บไซต์ ที่จัดทำเสร็จสิ้นแล้วนี้แม้จะสามารถทำงานได้ถูกต้องตรงตามวัตถุประสงค์และขอบเขตที่ตั้งไว้ แต่ก็ยังมีข้อจำกัดหลายประการ ซึ่งหากจะวิเคราะห์ข้อมูลให้เว็บไซต์ทำงานได้อย่างมีประสิทธิภาพ ผู้วิเคราะห์จะต้องปรับปรุงในประเด็นต่าง ๆ ต่อไปนี้

5.4.1 หากมีชุดข้อมูลเพิ่มเติมควรมีการอัปเดทข้อมูลใหม่ ๆ อยู่เสมอ

5.4.2 ควรมีการพัฒนาหน้าการแสดงผลต่าง ๆ ของเว็บไซต์ให้เหมาะสม เพื่อลดความผิดพลาดและเพิ่มประสิทธิภาพในการนำเสนอข้อมูล

5.4.3 ควรมีการพัฒนาให้ผู้เข้าชมสามารถอัปโหลดชุดข้อมูลเพิ่มเข้ามาในเว็บไซต์เข้ามาได้

5.4.4 ควรมีการพัฒนาการปรับเปลี่ยนรูปแบบมุมมองการแสดงกราฟข้อมูลในรูปแบบต่าง ๆ ได้

บรรณานุกรม

เกรียงศักดิ์ รักภักดีและวชิระ โมราชาติ./ (2561)./ **การพัฒนาฐานข้อมูลแหล่งท่องเที่ยวทางธรรมชาติในจังหวัดอุบลราชธานีโดยใช้เทคนิคการขุดข้อมูลจากความต้องการและพฤติกรรมของนักท่องเที่ยว** วารสารมหาวิทยาลัยราชภัฏเชียงราย,/ ปีที่ 11 (ฉบับที่ 3),/ 165–178 สืบค้นจาก https://so01.tci-thaijo.org/index.php/crrugds_ejournal/article/view/169133/121671

จำลักษณ์ ขุนพลแก้ว./ (ม.ป.ป.)./ **กระบวนการวิเคราะห์ข้อมูล**./ สืบค้นเมื่อวันที่ 17 เมษายน 2563,/ สืบค้นจาก https://www.bangkokbiznews.com/blog/detail/646353

ชำนิ คนไว./ (2562)./ **ปัญหาอาชญากรรมในสังคม** วารสารวิชาการแสงอีสาน,/ ปีที่ 14 (ฉบับที่ 2),/ 28–41 สืบค้นจาก https://so01.tci-thaijo.org/index.php/jsi/article/view/193735/134926

เด่นเดือน เลิศทยากุล, ธรา อั่งสกุล, วีรพงษ์ พลนิกรกิจและจิติมนต์ อั่งสกุล./ (2558)./ **ระบบสนับสนุนการตัดสินใจเพื่อการเลือกโฆษณาทางเว็บไซต์** วารสารวิชาการแสงอีสาน,/ ปีที่ 9 (ฉบับที่ 1),/ 28–41 สืบค้นจาก https://so05.tci-thaijo.org/index.php/sjss/article/view/34715/28863

พรรณธิภา เพชรบุญมี, ดวงกมล โพธิ์นาคและมนต์ชัย เทียนทอง./ (2556)./ **การพยากรณ์รูปแบบการเรียนรู้ตามประสบการณ์ของเดวิด โคล์ปโดยใช้กฎการจำแนกเทคนิคต้นไม้ตัดสินใจ** วารสารวิทยาศาสตร์และเทคโนโลยี,/ ปีที่ 21 (ฉบับที่ 6),/ 547–557 สืบค้นจาก https://li01.tci-thaijo.org/index.php/tstj/article/view/14536/13308

พัทธนันท์ รัตนวรเศวต./ (2563)./ **วิเคราะห์พื้นที่เสี่ยงต่ออาชญากรรมด้วยระบบสารสนเทศทางภูมิศาสตร์** วารสารบัณฑิตวิทยาลัย มหาวิทยาลัยราชภัฏจันทรเกษม,/ ปีที่ 15 (ฉบับที่ 1),/ 112–126 สืบค้นจาก https://so01.tci-thaijo.org/index.php/journalgrdcru/article/view/164818/163759

ยงยุทธ ลิขิตพัฒนะกุล./ (ม.ป.ป.)./ **Tableau คืออะไร**./ สืบค้นเมื่อวันที่ 18 เมษายน2563,/ สืบค้นจาก http://www.aiteam.co.th/what-is-tableau/

ศจี วานิช./ (2558)./ **การวิเคราะห์เหมืองข้อมูล**./ สืบค้นเมื่อวันที่ 17 เมษายน2563,/ สืบค้นจาก http://sajeegm301.blogspot.com/2015/11/data-mining.html

สุขุมา อรุณจิตและวุฒิพล มั่นเหมาะ./ (2559)./ **ปัจจัยที่ก่อให้เกิดอาชญากรรมทางเพศและทิศทางในการป้องกันอาชญากรรมทางเพศตามทฤษฎีอาชญากรรมสามเหลี่ยม** วารสารวิชาการ,/ ปีที่ 1 (ฉบับที่ 2),/ 23-34 สืบค้นจาก https://so02.tci-thaijo.org/index.php/forensic/article/view/185415/130386

สุวิมล สิทธิชาติ./ (2560)./ **การวิเคราะห์คุณลักษณะพื้นฐานทางการศึกษาด้วยเทคนิคเหมืองข้อมูล** วารสารเทคโนโลยีสารสนเทศ,/ ปีที่ 13 (ฉบับที่ 2),/ 20-28 สืบค้นจาก https://ph01.tci-thaijo.org/index.php/IT_Journal/article/view/105139/83569

อนัตต์ชัย ชุติภาสเจริญและจรัญ แสนราช./ (2561)./ **การเปรียบเทียบประสิทธิภาพของอัลกอริทึมและการคัดเลือกคุณลักษณะที่เหมาะสมเพื่อการพยากรณ์โอกาสความสำเร็จในการโอนเงินข้ามประเทศของบุคคลทั่วไป** วารสารวิจัย มข.,/ ปีที่ 6 (ฉบับที่ 3),/ 105-113 สืบค้นจาก https://so04.tci-thaijo.org/index.php/gskkuhs/article/view/156370/113480

อดุลย์ ยิ้มงาม./ (ม.ป.ป.)./ **การทำข้อมูล**./ สืบค้นเมื่อวันที่ 18 เมษายน2563,/ สืบค้นจาก http://compcenter.bu.ac.th/news-information/data-mining

อัชฌาพร กว้างสวาสดิ์, เพียงฤทัย หนูสวัสดิ์, วราสี คงเหมาะ, ปวีณา ทิพยากุลรักษ์และบุษกร สังขนันท์./ (2560)./ **ระบบทำนายระดับความเครียด ด้วยเทคนิคต้นไม้การตัดสินใจ**,/ ปีที่1 (ฉบับที่ 2),/ 13-26 สืบค้นจากhttps://ph02.tci-thaijo.org/index.php/RJST/article/view/239865/163568

อารีรัตน์ วงศ์สุวรรณ./ (2557)./ **RapidMiner Studio 6**./ สืบค้นเมื่อวันที่ 18 เมษายน2563,/ สืบค้นจาก http://www.ubu.ac.th/blog/areerat-104

อิทธิพล ดวงแก้วและสายยัญ สายยศ./ (2562)./ **วิเคราะห์ปัจจัยที่มีผลต่อการพัฒนาการล่าช้าอายุในวัยเด็กกับการทำเหมืองข้อมูลเทคนิคการ** วารสารวิทยาศาสตร์และเทคโนโลยีสารสนเทศ,/ ปีที่ 9 (ฉบับที่ 2),/ 44-55 สืบค้นจาก https://ph02.tci-thaijo.org/index.php/JIST/article/view/227026/158138

Automated./ (2561)./ **การนำเสนอข้อมูล Visualization**./ สืบค้นเมื่อวันที่ 19 เมษายน2563,/ สืบค้นจาก http://www.autosoft.in.th/data-visualization/การนำเสนอแผนภาพข้อมูล-data-visu

ภาคผนวก

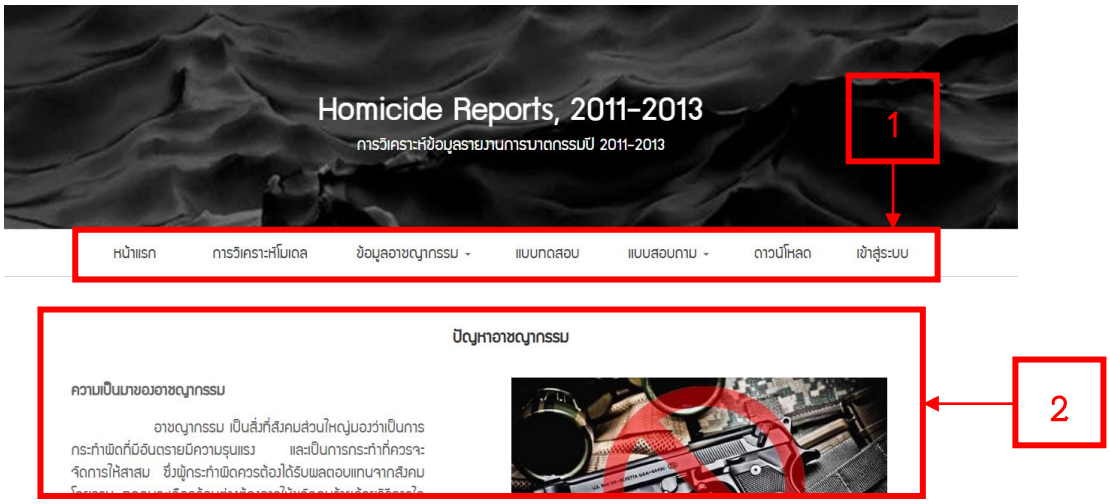
ภาคผนวก ก
คู่มือการใช้งานเว็บไซต์

ภาคผนวก ก
คู่มือการใช้งานเว็บไซต์

จากการดำเนินงานวิเคราะห์ข้อมูลรายงานการฆาตกรรมปี 2011-2013 เพื่อใช้สำหรับเผยแพร่ข้อมูลบนเว็บไซต์ มีการใช้งาน 7 ส่วน คือ ส่วนเนื้อหาปัญหาอาชญากรรม ส่วนการวิเคราะห์ข้อมูลด้วย Decision Tree ส่วนแสดงผลข้อมูลอาชญากรรม ส่วนแบบฟอร์มแบบทดสอบ ส่วนแบบฟอร์มแบบสอบถาม ส่วนดาวน์โหลดข้อมูล ส่วนอัปโหลดเพิ่ม ลบ แก้ไขข้อมูล ผู้จัดทำจึงได้ทำคู่มืออธิบายการใช้งานเว็บไซต์ ดังต่อไปนี้

คู่มือการเข้าใช้งานสำหรับผู้เข้าชมเว็บไซต์

1.หน้าแรกของเว็บไซต์ เมื่อเข้าสู่หน้าเว็บไซต์รายงานการฆาตกรรม

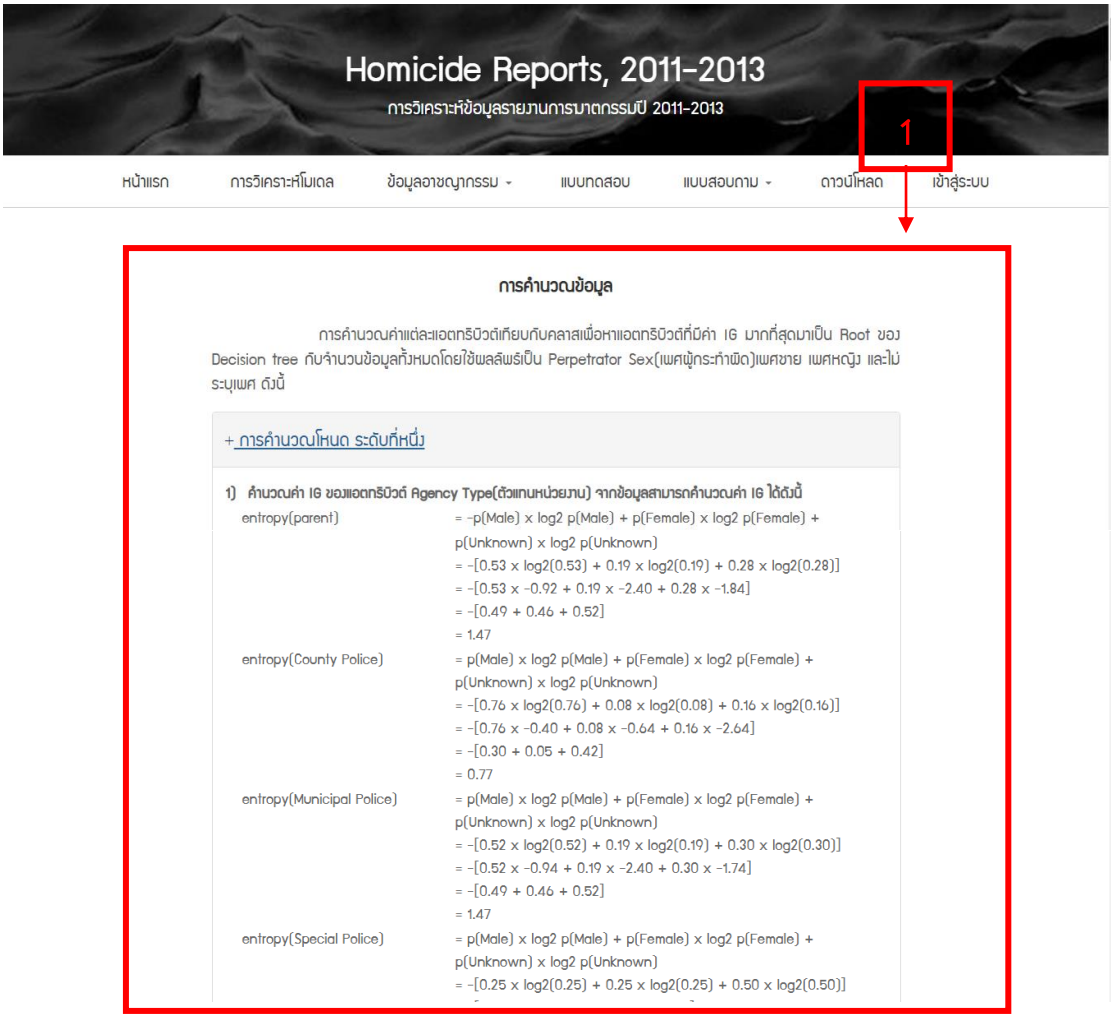


ภาพที่ ก.1 หน้าแรกของเว็บไซต์

หมายเลข 1 ส่วนของแถบเมนูที่ผู้เข้าใช้งานสามารถเลือกดูหน้าต่าง ๆ ภายในเว็บไซต์ได้

หมายเลข 2 ส่วนของเนื้อหาที่ประกอบด้วยปัญหาอาชญากรรม

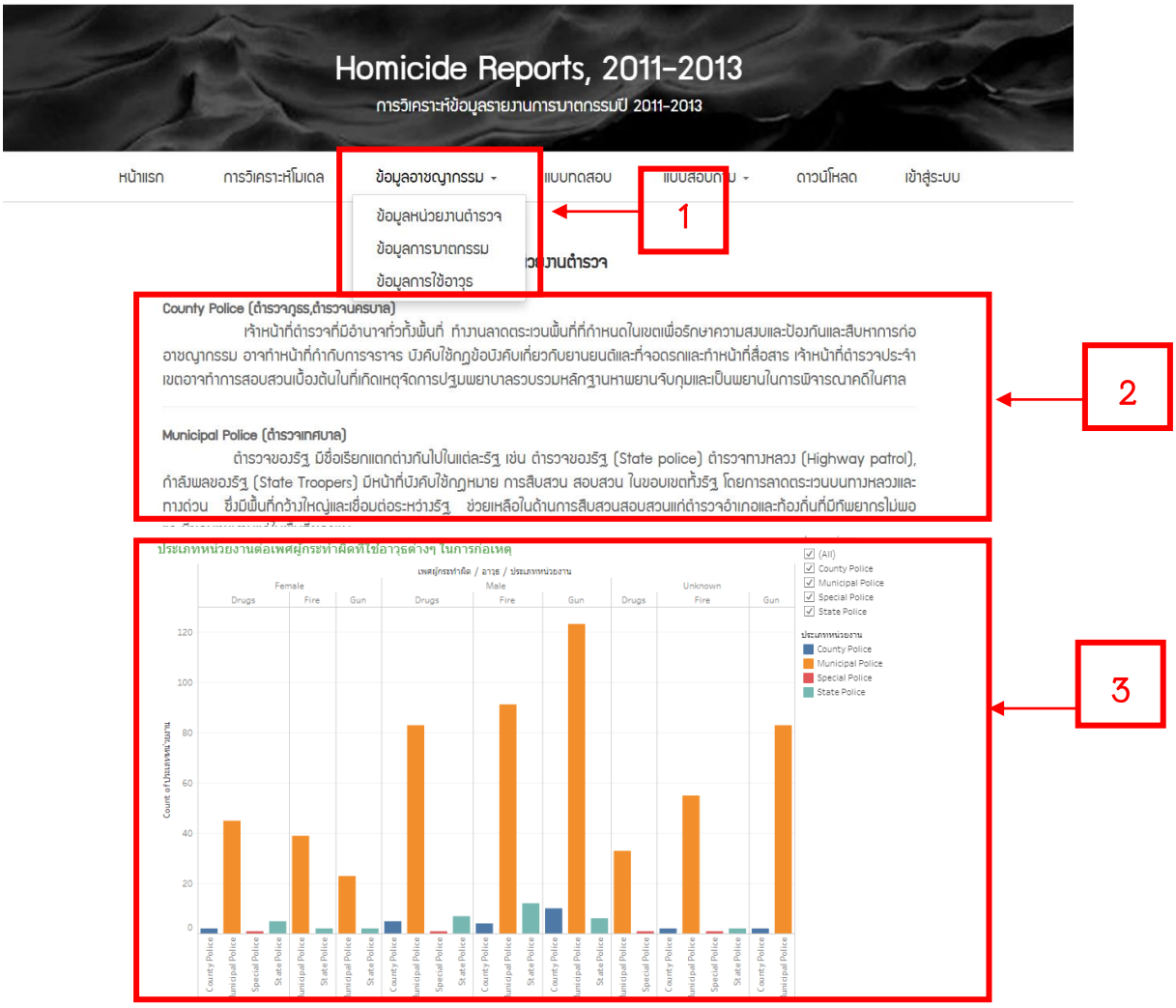
2. การแสดงขั้นตอนการวิเคราะห์ข้อมูลด้วยโมเดล Decision Tree



ภาพที่ ก.2 การวิเคราะห์ข้อมูลด้วยโมเดล Decision Tree

หมายเลข 1 ส่วนแสดงเนื้อหาขั้นตอนการคำนวณโหนดในแต่ละระดับและการสร้างโมเดล Decision Tree

3. หน้าส่วนแสดงผลข้อมูลอาชญากรรม ประกอบด้วย 3 ส่วน ข้อมูลหน่วยงานตำรวจ ข้อมูลการฆาตกรรมและข้อมูลการใช้อาวุธ แต่ละส่วนจะประกอบด้วยเนื้อหาและแดชบอร์ดเพื่อแสดงผลสรุป



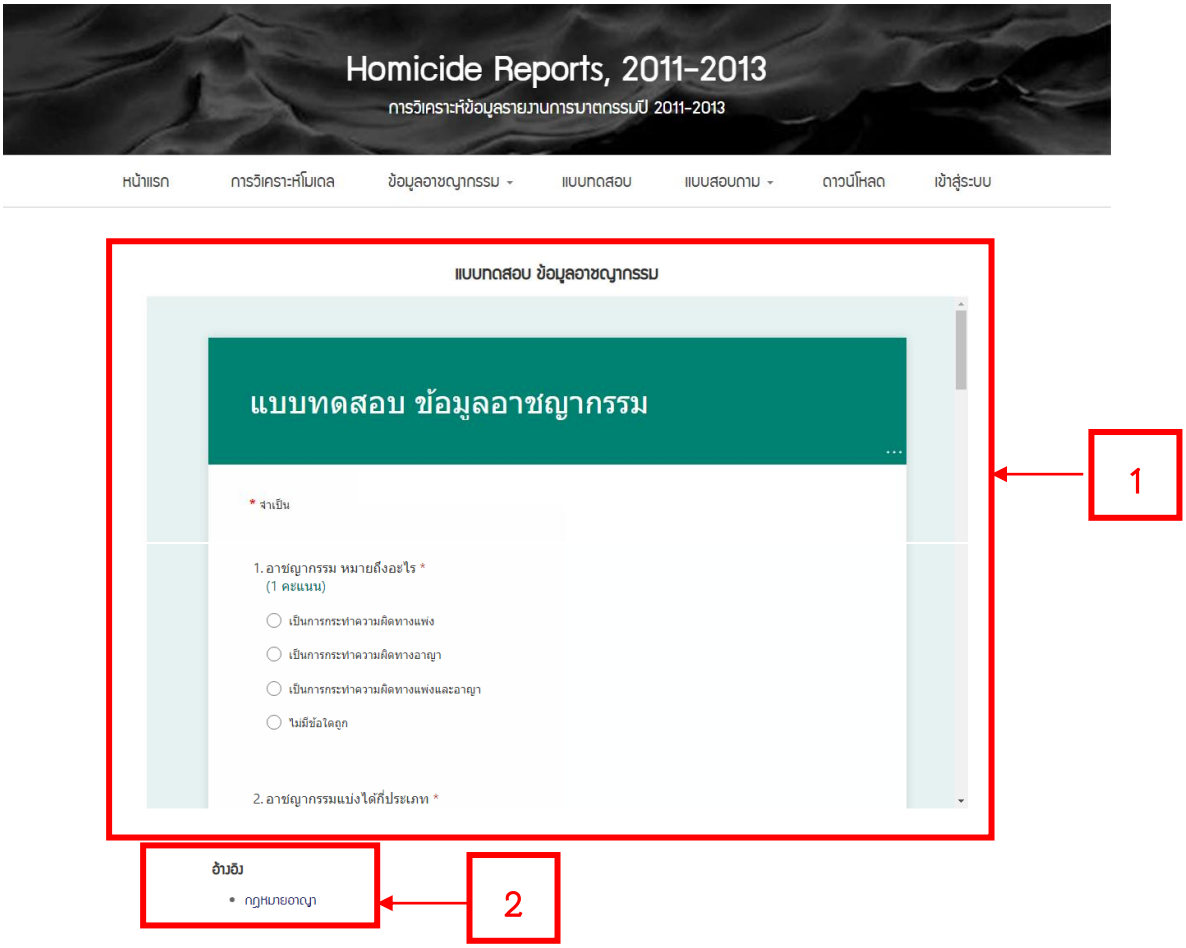
ภาพที่ ก.3 ส่วนแสดงผลข้อมูลอาชญากรรม

หมายเลข 1 ส่วนหัวข้อย่อยในแถบเมนูข้อมูลอาชญากรรม ประกอบด้วย ข้อมูลหน่วยงานตำรวจ ข้อมูลการฆาตกรรมและข้อมูลการใช้อาวุธ

หมายเลข 2 ส่วนรายละเอียดเนื้อหาในหัวข้อ ข้อมูลหน่วยงานตำรวจ

หมายเลข 3 ส่วนแดชบอร์ดแสดงข้อมูลสรุปผล ผู้เข้าใช้สามารถกรองดูข้อมูลได้ตามต้องการ

4. ส่วนแบบฟอร์มแบบทดสอบ ข้อมูลอาชญากรรม

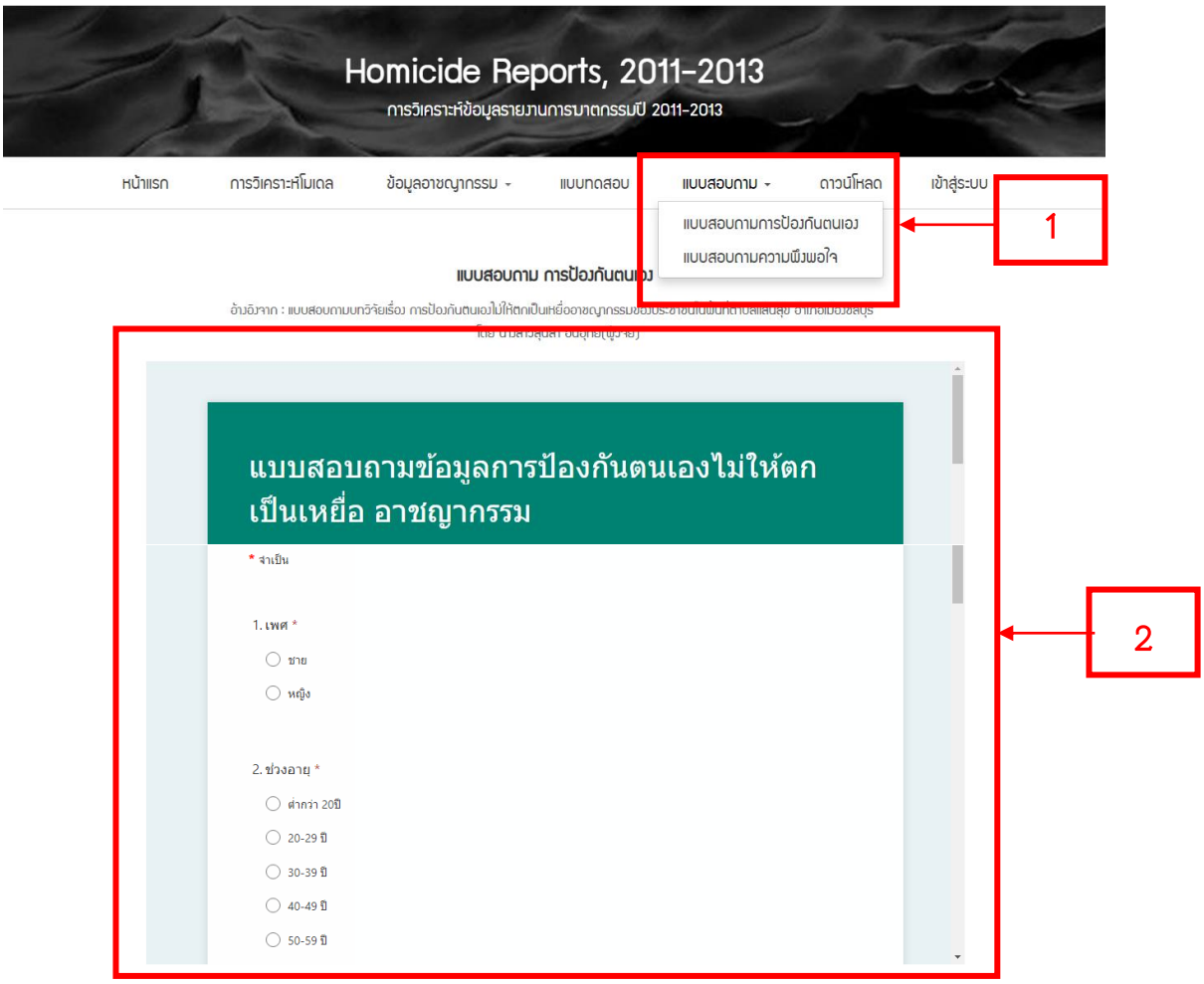


ภาพที่ ก.4 ส่วนแบบฟอร์มแบบทดสอบ

หมายเลข 1 แบบฟอร์มแบบทดสอบ ที่ผู้เข้าชมสามารถทดลองทำแบบทดสอบได้

หมายเลข 2 อ้างอิงไปยังลิงค์ข้อมูลเพื่อผู้เข้าชมได้เข้าดูแบบทดสอบเพิ่มเติมได้

5. แบบฟอร์มแบบสอบถาม ประกอบด้วย 2 ส่วน แบบสอบถาม การป้องกันตนเองและแบบสอบถามความพึงพอใจ

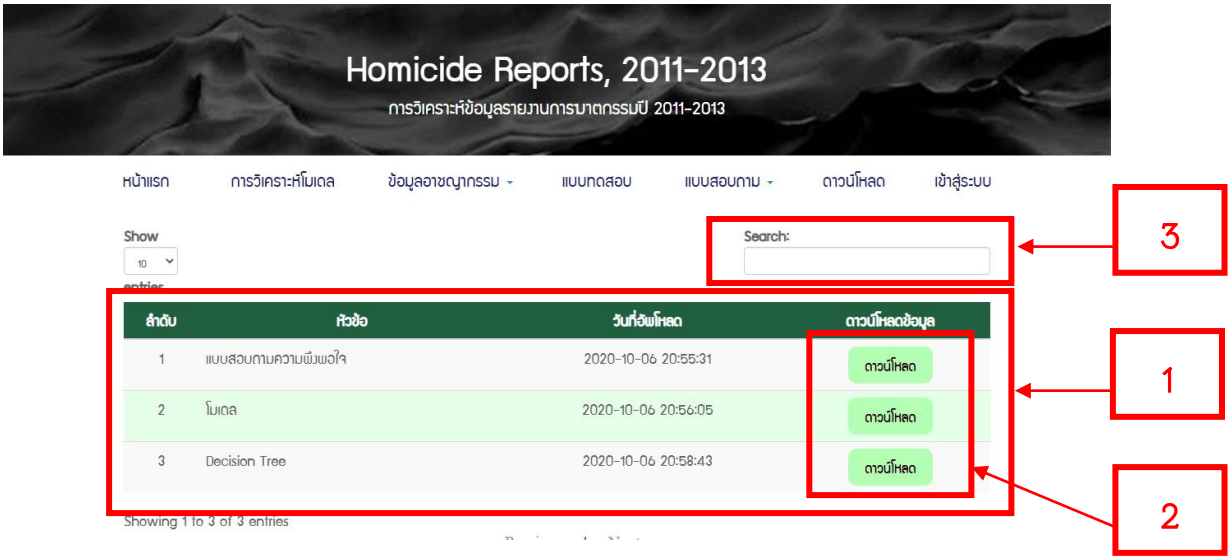


ภาพที่ ก.5 ส่วนแบบฟอร์มแบบสอบถาม

หมายเลข 1 ส่วนหัวข้อย่อยในแถบเมนูแบบฟอร์มแบบสอบถาม ประกอบด้วย แบบสอบถามข้อมูลการป้องกันตนเองไม่ให้ตกเป็นเหยื่อ อาชญากรรมและแบบสอบถามความพึงพอใจในการใช้งานเว็บไซต์

หมายเลข 2 ส่วนแบบสอบถามที่ผู้เข้าใช้สามารถกรอกแบบสอบถามเพื่อเก็บเป็นชุดข้อมูลไว้ได้

6. ส่วนดาวน์โหลดข้อมูล



ภาพที่ ก.6 ส่วนดาวน์โหลดข้อมูล

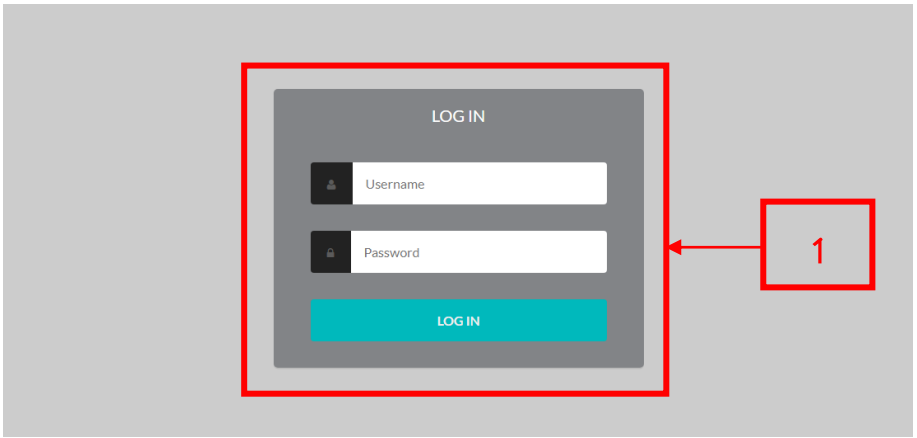
หมายเลข 1 ส่วนรายการชุดข้อมูลที่ผู้เข้าใช้สามารถดาวน์โหลดข้อมูลได้

หมายเลข 2 ปุ่มดาวน์โหลดข้อมูลในแต่ละรายการ

หมายเลข 3 ช่องค้นหาชื่อรายการที่ต้องการค้นหาในการดาวน์โหลดข้อมูล

คู่มือการเข้าใช้งานสำหรับแอดมิน

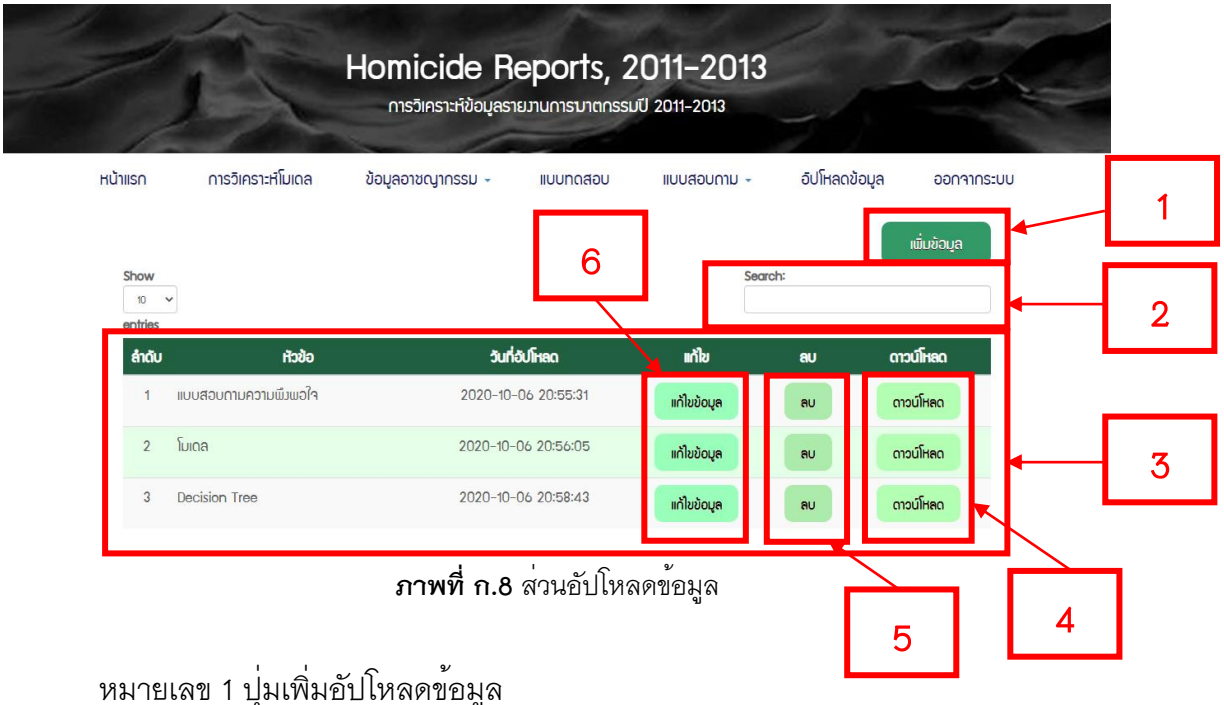
1.เข้าสู่ระบบสำหรับแอดมิน



ภาพที่ ก.7 เข้าสู่ระบบสำหรับแอดมิน

หมายเลข 1 ส่วนเข้าสู่ระบบสำหรับแอดมิน

2. ส่วนอัปโหลดข้อมูล



ภาพที่ ก.8 ส่วนอัปโหลดข้อมูล

หมายเลข 1 ปุ่มเพิ่มอัปโหลดข้อมูล

หมายเลข 2 ช่องค้นหาชื่อรายการที่ต้องการค้นหาในการดาวน์โหลดข้อมูล ลบข้อมูล หรือต้องการแก้ไขข้อมูลนั้น

หมายเลข 3 ตารางแสดงรายการข้อมูล

หมายเลข 4 ปุ่มดาวน์โหลดข้อมูล

หมายเลข 5 ปุ่มลบข้อมูล

หมายเลข 6 ปุ่มแก้ไขข้อมูล

ภาคผนวก ข
แบบฟอร์มและเอกสารที่ใช้ในโครงงาน

ภาคผนวก ข

แบบฟอร์มและเอกสารที่ใช้ในโครงงาน

โครงงานเรื่อง : การวิเคราะห์ข้อมูลรายงานการฆาตกรรมปี 2011-2013 เพื่อใช้สำหรับเผยแพร่ข้อมูลบนเว็บไซต์

คำชี้แจง : แบบสอบถามประกอบด้วยชุดคำถาม 3 ส่วน คือ

ส่วนที่ 1 ข้อมูลทั่วไป

ส่วนที่ 2 แบบสอบถามความพึงพอใจ

ส่วนที่ 3 ข้อเสนอแนะ

<u>ตอนที่ 1</u> ข้อมูลทั่วไป

<u>ข้อชี้แจง</u> ใส่เครื่องหมาย ✓ ลงใน ☐ หน้าคำตอบที่ตรงกับข้อมูลของผู้ใช้งานเพียงข้อเดียว

1. เพศ

☐ ชาย ☐ หญิง

2. อายุ

☐ ต่ำกว่า 20 ปี ☐ 21-25 ปี ☐ 26-30 ปี

☐ 31-35 ปี ☐ 36 ปีขึ้นไป

3. อาชีพ

☐ นักเรียน/นักศึกษา ☐ ครู/อาจารย์ ☐ บุคคลทั่วไป

<u>ตอนที่ 2</u> แบบสอบถามความพึงพอใจ

<u>ข้อชี้แจง</u> ใส่เครื่องหมาย ✓ ลงใน ☐ ลงในช่องระดับคะแนนความคิดเห็นที่ตรงกับความคิเห็นของผู้ใช้งานมากที่สุด 5 = ดีมาก 4 = ดี 3 = ปานกลาง 2 = น้อย 1 = ควรปรับปรุง

ตารางที่ ข.1 ข้อมูลแบบสอบถาม

รายการ	ระดับความพึงพอใจ				
	5	4	3	2	1
ด้านเนื้อหา					
1.เนื้อหามีความถูกต้อง					
2.เนื้อหามีความกะทัดรัด เข้าใจง่าย					
3.ข้อมูลภายในเว็บไซต์มีความน่าเชื่อถือ					
4.รูปแบบการนำเสนอ visualization มีความเหมาะสมเข้าใจง่าย					
ด้านการออกแบบเว็บไซต์					
1.เว็บไซต์ง่ายต่อการใช้งาน					
2.มีความเหมาะสมในการใช้ชนิด ขนาด สีตัวอักษรบนเว็บไซต์					
3.มีความเหมาะสมในการจัดตำแหน่ง การออกแบบไม่ซับซ้อน					
4.มีความเหมาะสมในการใช้สัญลักษณ์หรือรูปถาพในการสื่อความหมาย					
5.มีความเหมาะสมในการจัดวางตำแหน่งของส่วนประกอบเมนูต่าง ๆ บนเว็บไซต์					
ด้านประโยชน์และการนำไปใช้					
1.เนื้อหามีประโยชน์ต่อการนำไปประยุกต์ใช้					
2.สามารถนำความรู้ที่ได้รับไปประยุกต์ใช้ในชีวิตประจำวันได้					
3.สามารถเป็นแหล่งความรู้และแนวทางในการพัฒนาต่อไปได้					
4.แหล่งข้อมูลตรงกับความต้องการของผู้ใช้งาน					

ข้อเสนอแนะ

.....

.....

FM-BIS-10

แบบบันทึกรายละเอียดการเข้าพบอาจารย์ที่ปรึกษา

วิชา Computer Information System Project

หลักสูตรระบบสารสนเทศทางคอมพิวเตอร์ สาขาบริหารธุรกิจ

คณะบริหารธุรกิจและศิลปศาสตร์ มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา เชียงใหม่

ชื่อโครงการ การวิเคราะห์ข้อมูลรายงานการฆาตกรรมปี 2011-2013 เพื่อใช้สำหรับเผยแพร่ข้อมูลบนเว็บไซต์

ผู้จัดทำ 1. นางสาววรรณพร นันตีะเสน รหัสนักศึกษา 61521207042-4

อาจารย์ที่ปรึกษา ผู้ช่วยศาสตราจารย์ ดร. ชัชฎาพร ปุกแก้ว

ครั้งที่	วัน/เดือน/ปี	รายละเอียดการเข้าพบอาจารย์ที่ปรึกษา	ความคืบหน้าของผลงาน	ลายมืออาจารย์ที่ปรึกษา
1		ส่งความคืบหน้า บทที่ 1-3	5 %	
2		แก้บทที่ 1-3	5 %	
3		ส่งความคืบหน้าโปรแกรม Tableu + ข้อมูล	5 %	
4		ส่งความคืบหน้าโปรแกรม Tableu + ข้อมูล	5 %	
5		แก้ความคืบหน้าโปรแกรม Tableu + ข้อมูล	5 %	
		ของเว็บไซต์		
6	1 ก.ย. 63	ส่งความคืบหน้าของเว็บไซต์ และ บทที่ 1-3	5 %	
7	9 ก.ย. 63	ส่งความคืบหน้าเว็บไซต์ บทที่ 1-4	15 %	
		แสดงผลการอัพเดทข้อมูล		
8	23 ก.ย. 63	ส่งความคืบหน้าบทที่ 4 และ หน้าเว็บไซต์	5 %	
9	30 ก.ย. 63	ส่งความคืบหน้าเว็บไซต์ + บทที่ 1-5	10 %	
10	4 ต.ค. 63	แก้เอกสารโครงงาน + แก้ไขเว็บไซต์	10 %	

ลงชื่อ วรรณพร นันตีะเสน (นักศึกษา)

(นางสาววรรณพร นันตีะเสน)

ประวัติผู้จัดทำ

ชื่อ-สกุล นางสาววรรณพร นันต๊ะเสน

วันเดือนปีเกิด วัน จันทร์ ที่ 24 เดือน พฤศจิกายน พ.ศ. 2540

ภูมิลำเนา 56 หมู่ที่ 1 ตำบลท่าวังตาล อำเภอสารภี จังหวัดเชียงใหม่ 50140

E-mail veview24@gmail.com

ประวัติการศึกษา

- ระดับมัธยมศึกษาตอนต้น โรงเรียนดาราวิทยาลัย จังหวัดเชียงใหม่ สำเร็จการศึกษาปี 2555
- ระดับประกาศนียบัตรวิชาชีพ วิทยาลัยอาชีวศึกษา จังหวัดเชียงใหม่ สำเร็จการศึกษาปี 2558
- ระดับประกาศนียบัตรวิชาชีพชั้นสูง วิทยาลัยอาชีวศึกษา จังหวัดเชียงใหม่ สำเร็จการศึกษาปี 2560
- ระดับปริญญาตรี สาขาวิชาระบบสารสนเทศทางธุรกิจ มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา จังหวัดเชียงใหม่ สำเร็จการศึกษาปี 2563