การวิเคราะห์ข้อมูลการจำแนกเห็ดพิษหรือเห็ดที่รับประทานได้ในอเมริกาเหนือ เพื่อเผยแพร่ข้อมูลบนเว็บไซต์

นางสาวกุลธิดา เป็นแผ่น
นายณัฐพงษ์ นันต๊ะเสน

หลักสูตรระบบสารสนเทศทางธุรกิจสาขาบริหารธุรกิจ
คณะบริหารธุรกิจและศิลปะศาสตร์มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา
ปีการศึกษา 2563

การวิเคราะห์ข้อมูลการจำแนกเห็ดพิษหรือเห็ดที่รับประทานได้ในอเมริกาเหนือ เพื่อเผยแพร่ข้อมูลบนเว็บไซต์

นางสาวกุลธิดา เป็นแผ่น
นายณัฐพงษ์ นันต๊ะเสน

หลักสูตรระบบสารสนเทศทางธุรกิจสาขาบริหารธุรกิจ
คณะบริหารธุรกิจและศิลปะศาสตร์มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา
ปีการศึกษา 2563

ชื่อโครงงาน	การวิเคราะห์ข้อมูลการจำแนกเห็ดพิษหรือเห็ดที่รับประทานได้ในอเมริกาเหนือ เพื่อเผยแพร่ข้อมูลบนเว็บไซต์
โดย	นางสาวกุลธิดา เป็นแผ่น รหัส 61521207030-9
	นายณัฐพงษ์ นันตีะเสน รหัส 61521207035-8
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 คณะบริหารธุรกิจและศิลปศาสตร์

ชื่อโครงงาน การวิเคราะห์ข้อมูลการจำแนกเห็ดพิษหรือเห็ดที่รับประทานได้ในอเมริกาเหนือ เพื่อเผยแพร่ข้อมูลบนเว็บไซต์

โดย นางสาวกุลธิดา เป็นแผ่น รหัส 61521207030-9

นายณัฐพงษ์ นันตะเสน รหัส 61521207035-8

อาจารย์ที่ปรึกษา ผู้ช่วยศาสตราจารย์ ดร.ชัชฎาพร ปุกแก้ว

หลักสูตร ระบบสารสนเทศทางธุรกิจ คณะบริหารธุรกิจและศิลปศาสตร์ ระดับปริญญาตรี

ปีการศึกษา 2563

บทคัดย่อ

การจัดทำโครงงานในครั้งนี้มีวัตถุประสงค์ เพื่อวิเคราะห์ข้อมูลการจำแนกเห็ดพิษหรือเห็ดที่รับประทานได้ในอเมริกาเหนือ เพื่อเผยแพร่ข้อมูลบนเว็บไซต์ ด้วยขั้นตอนกระบวนการการวิเคราะห์ข้อมูลด้วย CRISP-DM จากเทคนิคการทำเหมืองข้อมูลแบบ Classification ด้วยการสร้างโมเดล Decision Tree เพื่อจัดกลุ่มข้อมูลที่มีความสัมพันธ์กัน โดยใช้โปรแกรมที่ใช้ทำเหมืองข้อมูล คือ โปรแกรม Weka 3.8.4 และโปรแกรม RapidMiner Studio 9.5.1 ในการสร้างโมเดล Decision Tree เพื่อทำการเปรียบเทียบ และเลือกผลลัพธ์ที่ดีที่สุดมานำเสนอ พบว่าเทคนิค Decision Tree:J48 ให้ผลลัพธ์การจำแนก ประเภทเห็ด รับประทานได้(e) และรับประทานไม่ได้(p) มีความถูกต้องถึง 91.35% และ ให้ผลลัพธ์ของกฎที่สามารถทำนายได้จำนวน 11 กฎ จากนั้นนำข้อมูลสารสนเทศมาทำการแสดงผลแบบ Visualization ในรูปแบบของภาพ โดยใช้โปรแกรม Tableau Public เผยแพร่บน Web Browser

จากการวิเคราะห์ข้อมูล และเผยแพร่ข้อมูลบนWeb browser พบว่า สามารถนำมาประยุกต์ได้ตามจุดประสงค์ที่ตั้งไว้ สรุปได้ว่าสามารถนำข้อมูลจากเว็บไซต์ไปปรับปรุงเพิ่มเติมต่อได้ เพื่อให้เกิดประโยชน์กับความปลอดภัยในการรับประทานเห็ดของบุคคลทั่วไปให้ได้ประสิทธิภาพมากขึ้นต่อไป

กิตติกรรมประกาศ

โครงงานครั้งนี้สำเร็จลุล่วงไปได้ด้วยดี เนื่องจากได้รับความกรุณาอย่างดียิ่งจากผู้ช่วย–ศาสตราจารย์ ดร.ชัชฎาพร ปุกแก้ว อาจารย์ที่ปรึกษาโครงงาน อาจารย์ สุพงศ์ แดงสุริยศรี หัวหน้าสาขาวิชาระบบสารสนเทศทางธุรกิจ และคณะอาจารย์ในสาขาระบบสารสนเทศทางธุรกิจ คณะบริหารธุรกิจและศิลปศาสตร์ มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา ทุกท่านที่ได้ให้คำปรึกษาและคำแนะนำในการจัดทำโครงงาน เพื่อบรรลุวัตถุประสงค์ของโครงงานให้เสร็จสมบูรณ์ รวมถึงช่วยแก้ไขเอกสารโครงงานให้มีความสมบูรณ์มากขึ้น

ขอขอบพระคุณเพื่อนๆทุกคนที่ให้ส่วนช่วยแนะนำติชมการทำโครงงานในครั้งนี้ ขอบคุณกำลังใจดี ๆ จาก คุณพ่อ คุณแม่ที่คอยให้คำปรึกษา

คณะผู้จัดทำหวังว่าโครงงานฉบับนี้จะเป็นประโยชน์และเป็นตัวอย่างของโครงงานการศึกษาการวิเคราะห์ข้อมูลแก่ผู้ที่สนใจศึกษาค้นคว้าได้ไม่มากก็น้อย หากมีข้อผิดพลาดประการใด คณะผู้จัดทำได้ขออภัยมา ณ ที่นี้

กุลธิดา เป็นแผ่น

ณัฐพงษ์ นันต๊ะเสน

สารบัญ

เนื้อหา **หน้า**

บทคัดย่อ.....ก
กิตติกรรมประกาศ.....ข
สารบัญ.....ค
สารบัญภาพ.....จ
สารบัญตาราง.....ฌ
บทที่ 1.....1
1.1 ความสำคัญและที่มาของปัญหา.....1
1.2 วัตถุประสงค์.....2
1.3 ประโยชน์ที่จะได้รับ.....3
1.4 ขอบเขต.....3
1.5 เครื่องมือที่ใช้ในการพัฒนาโปรแกรม.....4
1.6 สถานที่ใช้ในการดำเนินการศึกษาและรวบรวมข้อมูล.....4
1.6 ระยะเวลาในการดำเนินการ.....5
1.7 นิยามศัพท์เฉพาะ.....5
1.8 บทสรุป.....7
บทที่ 2.....8
2.1 แนวคิด.....8
2.2 ทฤษฎี.....15
2.3 เครื่องมือในการออกแบบและวิเคราะห์ข้อมูล.....26
2.4 วรรณกรรมที่เกี่ยวข้อง.....26
2.5 บทสรุป.....30
บทที่ 3.....31
3.1 การวิเคราะห์ข้อมูลด้วย CRISP-DM.....31
3.2 การออกแบบเว็บไซต์.....72

สารบัญ (ต่อ)

เนื้อหา **หน้า**

3.3 บทสรุป....79
บทที่ 4 80
4.1 ผลการดำเนินงาน 80
4.2 การอธิปรายผล 89
4.3 บทสรุป 99
บทที่ 5 100
5.1 บทสรุปผลการดำเนินโครงงาน....100
5.2 ข้อจำกัดของโครงงาน....101
5.3 ปัญหาและอุปสรรคของโครงงาน 101
5.4 ข้อเสนอแนะ....102
บรรณานุกรม ....ญ
ภาคผนวก 105
ภาคผนวก ก 107
คู่มือการใช้งานเว็บไซต์ 107
ภาคผนวก ข 120
แบบสอบถาม 120
ภาคผนวก ค....129
เอกสารที่ใช้ในโครงงาน....129
ประวัติผู้จัดทำ....130

สารบัญภาพ

ภาพที่ **หน้า**

ภาพที่ 2. 1 แสดง Data Visualization 10
ภาพที่ 2. 2 แสดง Network Graph 10
ภาพที่ 2. 3 แสดง Bump Chart 11
ภาพที่ 2. 4 แสดง Heat map 11
ภาพที่ 2. 5 แสดง Stacked Bar Chart 12
ภาพที่ 2. 6 แสดง Line Chart 12
ภาพที่ 2. 7 แสดง Histogram 13
ภาพที่ 2. 8 แสดง Tree map 13
ภาพที่ 2. 9 แสดง Hexbin Map 14
ภาพที่ 2. 10 แสดงโครงสร้างเว็บไซต์แบบเรียงลำดับ 18
ภาพที่ 2. 11 แสดงโครงสร้างเว็บไซต์แบบลำดับขั้น 19
ภาพที่ 2. 12 แสดงโครงสร้างเว็บไซต์แบบตาราง 19
ภาพที่ 2. 13 แสดงโครงสร้างแบบใยแมงมุม 20
ภาพที่ 2. 14 แสดงซอร์จโค้ดส่วนประกอบของโครงสร้างของ CSS 23
ภาพที่ 2. 15 แสดงตัวอย่างต้นไม้ตัดสินใจสำหรับการจำแนกคุณลักษณะของลูกค้าที่ทำการซื้อคอมพิวเตอร์ 25
ภาพที่ 3. 1 เว็บไซต์ kaggle.com 32
ภาพที่ 3. 2 ข้อมูลการจำแนกเห็ดพิษหรือเห็ดที่รับประทานได้ในอเมริกาเหนือ 33
ภาพที่ 3. 3 ข้อมูลการจำแนกเห็ดพิษหรือเห็ดที่รับประทานได้ในอเมริกาเหนือ 34
ภาพที่ 3. 4 ข้อมูลผิดพลาดของชนิดของรากเห็ดที่ไม่ชัดเจนทั้งหมด 35
ภาพที่ 3. 5 ชุดข้อมูลที่คัดเลือกมาวิเคราะห์ข้อมูล 35
ภาพที่ 3. 6 แสดงรูปแบบโมเดล Decision Tree จากการคำนวณด้วยมือ 60
ภาพที่ 3. 7 แสดงการเปิดโปรแกรม Weka 3.8.4 61
ภาพที่ 3. 8 แสดงการเข้าใช้โปรแกรม Weka 3.8.4 62
ภาพที่ 3. 9 แสดงการนำไฟล์ข้อมูลเข้าโปรแกรม Weka 3.8.4 62

สารบัญภาพ(ต่อ)

ภาพที่ **หน้า**

ภาพที่ 3. 10 แสดงข้อมูลหลังจากนำเข้าโปรแกรม Weka 3.8.4 63
ภาพที่ 3. 11 แสดงการเลือกเทคนิคการจัดกลุ่มข้อมูลแบบ Decision Tree 63
ภาพที่ 3. 12 แสดงการเลือกเทคนิคการจัดกลุ่มข้อมูลแบบ Decision Tree: J48 64
ภาพที่ 3. 13 แสดงหน้าจอผลลัพธ์ของโมเดลการจัดกลุ่มข้อมูลแบบ Decision Tree: J48 64
ภาพที่ 3. 14 แสดงผลลัพธ์กฎต้นไม้การตัดสินใจที่แตกกิ่งออกมา..... 65
ภาพที่ 3. 15 แสดงรูปแบบโมเดล Graph Decision Tree ในโปรแกรม Weka 3.8.4 65
ภาพที่ 3. 16 แสดงรูปแบบโมเดล Graph Decision Tree ในโปรแกรม RapidMiner..... 66
ภาพที่ 3. 17 คำบรรยายลักษณะงาน Decision Tree ในโปรแกรม RapidMiner..... 66
ภาพที่ 3. 18 แสดงผลลัพธ์จากการจำแนกกลุ่มแบบ Decision Tree: J48 ในโปรแกรม Weka.....67
ภาพที่ 3. 19 แสดงรูปแบบโมเดล Decision Tree จากการคำนวณด้วยมือ..... 68
ภาพที่ 3. 20 แสดงรูปแบบโมเดล Graph Decision Tree ในโปรแกรม RapidMiner 68
ภาพที่ 3. 21 แสดงรูปแบบโมเดล Graph Decision Tree ในโปรแกรม Weka 3.8.4 68
ภาพที่ 3. 22 แสดงข้อมูลแสดงผลการวิเคราะห์ประเภทของเห็ดรับประทานได้และรับประทานไม่ได้จำแนกตามพื้นผิวของเห็ด.....70
ภาพที่ 3. 23 แสดงผลประเภทของเห็ดรับประทานได้และรับประทานไม่ได้จำแนกตามรอยฟกช้ำ.....71
ภาพที่ 3. 24 แสดงข้อมูลประเภทของเห็ดรับประทานได้และรับประทานไม่ได้จำแนกตามรูปทรง..... 71
ภาพที่ 3. 25 แสดงผลข้อมูลที่อยู่อาศัยของเห็ดว่าประเภทเห็ดรับประทานได้และรับประทานไม่ได้จำแนกตามที่อยู่อาศัย 72
ภาพที่ 3. 26 แสดงหน้าแรกของเว็บไซต์ แสดงเมนูต่าง ๆ ของหน้าเว็บไซต์.....72
ภาพที่ 3. 27 แสดงหน้าข้อมูลเกี่ยวกับเห็ด 73
ภาพที่ 3. 28 แสดงหน้าการทำการทำDecision tree.....73
ภาพที่ 3. 29 แสดงหน้าแสดงผลข้อมูลการวิเคราะห์จำแนกตามรูปทรงพื้นผิวและที่อยู่อาศัย.....74
ภาพที่ 3. 30 แสดงหน้าแสดงผลข้อมูลจำแนกตามรูปทรงรอยฟกช้ำและที่อยู่อาศัย 74
ภาพที่ 3. 31 แสดงหน้าแสดงผลข้อมูลการวิเคราะห์จำแนกที่อยู่อาศัยขึ้นที่ไหนมากที่สุด 75
ภาพที่ 3. 32 แสดงหน้าแสดงผลข้อมูลการวิเคราะห์จำแนกที่อยู่อาศัยขึ้นที่ไหนน้อยที่สุด 75

สารบัญภาพ(ต่อ)

ภาพที่ **หน้า**

ภาพที่ 3. 33 แสดงหน้าแบบทดสอบการให้ความรู้เกี่ยวกับเห็ด 76
ภาพที่ 3. 34 แสดงหน้าแบบสอบถามการนำ Decision Treeมาประยุกต์ 76
ภาพที่ 3. 35 แสดงหน้าแบบสอบถามความพึงพอใจการใช้เว็บไซต์ 77
ภาพที่ 3. 36 แสดงหน้าแสดงผลข้อมูลผู้จัดทำ....77
ภาพที่ 3. 37 แสดงหน้าเอกสาร....78
ภาพที่ 3. 38 แสดงหน้าล็อกอินสำหรับผู้วิเคราะห์ 78
ภาพที่ 4. 1 หน้าแรกของเว็บไซต์....81
ภาพที่ 4. 2 แสดงหน้าเมนูการวิเคราะห์ข้อมูลแบบ Visualization 81
ภาพที่ 4. 3 แสดงข้อมูลของเห็ดอเมริกาเหนือที่รับประทานได้และเห็ดที่เป็นพิษ.... 82
ภาพที่ 4. 4 แสดงรายละเอียดข้อมูลของเห็ดอเมริกาเหนือ.... 82
ภาพที่ 4. 5 แสดงข้อมูลของเห็ดไทยที่รับประทานได้และเห็ดที่เป็นพิษ 83
ภาพที่ 4. 6 แสดงรายละเอียดข้อมูลของเห็ดไทย.... 83
ภาพที่ 4. 7 แสดงความรู้และข้อมูลที่เกี่ยวกับเห็ด....84
ภาพที่ 4. 8 แสดงหน้าการทำDecision three 84
ภาพที่ 4. 9 แสดงหน้าทำแบบสอบถามการประยุกต์ใช้จาก Rules based.... 85
ภาพที่ 4. 10 แสดงหน้าแบบสอบถามความพึงพอใจการเข้าใช้งานเว็บไซต์.... 85
ภาพที่ 4. 11 แสดงหน้าแบบทดสอบเกี่ยวกับเห็ดที่รับประทานได้และเห็ดที่เป็นพิษ 86
ภาพที่ 4. 12 แสดงเอกสารที่สามารถนำไปใช้วิเคราะห์ได้ 86
ภาพที่ 4. 13 แสดงหน้าผู้จัดทำ อาจารย์ที่ปรึกษาโครงงาน และหน่วยงานที่เกี่ยวข้อง 87
ภาพที่ 4. 14 แสดงหน้าล็อกอินของผู้วิเคราะห์ 87
ภาพที่ 4. 15 แสดงหน้าอัปโหลดไฟล์.... 88
ภาพที่ 4. 16 แสดงหน้ารายการเอกสารที่อัปโหลด 88
ภาพที่ ก. 1 หน้าแรกของเว็บไซต์.... 107
ภาพที่ ก. 2 แสดงข้อมูลเนื้อหาเกี่ยวกับเห็ด 108
ภาพที่ ก. 3 แสดงข้อมูลเนื้อหาเกี่ยวกับเห็ดในอเมริกาเหนือ 108

สารบัญภาพ(ต่อ)

ภาพที่ **หน้า**

ภาพที่ ก. 4 แสดงข้อมูลเนื้อหาเกี่ยวกับเห็ดในประเทศไทย 109
ภาพที่ ก. 5 แสดงการอธิบายขั้นตอนการสร้างโมเดล Decision Tree 109
ภาพที่ ก. 6 แสดงขั้นตอนการสร้าง Decision Tree 110
ภาพที่ ก. 7 แสดงการทดสอบกับข้อมูลโดยใช้โปรแกรม Weka 3.8.4 110
ภาพที่ ก. 8 แสดงการเปรียบเทียบโมเดลของโปรแกรมและการคำนวณมือ 111
ภาพที่ ก. 9 แสดงการนำเสนอข้อมูลในรูปแบบแผนภาพDashboard 111
ภาพที่ ก. 10 แสดงการนำเสนอข้อมูลในรูปแบบแผนภาพ Dashboard 112
ภาพที่ ก. 11 แสดงผลประเภทของเห็ดรับประทานได้และรับประทานไม่ได้ 112
ภาพที่ ก. 12 แสดงแบบสอบถาม 113
ภาพที่ ก. 13 แสดงแบบสอบถาม การประยุกต์ใช้จาก Rules based 113
ภาพที่ ก. 14 แสดงแบบสอบถามความพึงพอใจในการเข้าใช้งานเว็บไซต์ 114
ภาพที่ ก. 15 แสดงแบบสอบถาม Quiz เห็ดที่รับประทานได้และเห็ดที่เป็นพิษ 114
ภาพที่ ก. 16 แสดงรายการเอกสารที่สามารถนำไปใช้วิเคราะห์ 115
ภาพที่ ก. 17 แสดงส่วนติดต่อคณะผู้จัดทำ 115
ภาพที่ ก. 18 แสดงส่วนLogin เข้าสู่ระบบสำหรับผู้ดูแลระบบ 116
ภาพที่ ก. 19 แสดงส่วนเข้าสู่ระบบของผู้ดูแลระบบ 117
ภาพที่ ก. 20 แสดงส่วนของผู้ดูแลระบบ 117
ภาพที่ ก. 21 แสดงส่วนการอัปโหลดข้อมูล 118

สารบัญตาราง

ตาราง **หน้า**

ตารางที่ 1. 1 ระยะเวลาในการดำเนินการ.....5
ตารางที่ 4. 1 แสดงข้อมูลการประมวลผลด้านการออกแบบ 95
ตารางที่ 4. 2 แสดงการประมวลผลด้านคุณภาพข้อมูล (Information Quality)..... 96
ตารางที่ 4. 3 แสดงการประเมินผลด้านประโยชน์และการนำไปใช้.....97

บทที่ 1
บทนำ

1.1 ความสำคัญและที่มาของปัญหา

เห็ดเป็นสิ่งมีชีวิตจำพวกเห็ดรา มีทั้งที่ใช้เป็นอาหารและประเภทที่มีพิษ สายพันธุ์ของเห็ดมีมากกว่า 30,000 สายพันธุ์ กระจัดกระจายอยู่ทั่วโลกในจำนวนสายพันธุ์ดังกล่าวมีสายพันธุ์ที่มนุษย์สามารถนำมาบริโภคเป็นอาหารได้และสายพันธุ์ที่เป็นเห็ดที่มีพิษหรือเห็ดเมา ซึ่งถ้าบริโภคเข้าไปอาจเป็นอันตรายถึงแก่ชีวิตได้ เห็ดที่นำมาบริโภคเป็นอาหาร ในอดีตนั้นมีเพียงไม่กี่ชนิด ซึ่งในอดีตคนส่วนใหญ่มักจะมองและคิดถึงเห็ดในลักษณะของสิ่งที่ไม่น่าปลอดภัย ลึกลับและน่ากลัว ต่อมาคนเราได้มีการศึกษาและได้รับความรู้มากขึ้นจึงสามารถที่จะเลือกชนิดของเห็ดที่ปลอดภัยเพื่อนำมารับประทาน เดิมทีเป็นการเก็บเห็ดจากธรรมชาติมาบริโภค ทั้งนี้คนที่เก็บเห็ดจะต้องมีความรู้เกี่ยวกับเห็ดด้วยว่าเป็นเห็ดที่รับประทานได้หรือไม่ เพราะบางชนิดก็มีพิษหากรับประทานเข้าไปอาจเป็นอันตรายถึงชีวิตได้ ซึ่งในอดีตจนถึงปัจจุบันมีคนจำนวนมากเสียชีวิตจากการรับประทานเห็ดที่มีพิษ เนื่องจากการขาดความรู้ ความเชี่ยวชาญในเรื่องในรูปลักษณะของเห็ดทำให้ไม่สามารถแยกแยะได้ว่าเห็ดใดมีพิษ เนื่องจากเห็ดที่สามารถนำมารับประทานได้บางชนิดมีรูปร่างลักษณะคล้ายคลึง กับเห็ดมีพิษ ดังนั้นความรู้ในเรื่องลักษณะภายนอกของเห็ดจึงมีความสำคัญอย่างมากต่อการจำแนกสายพันธุ์เห็ด

การวิเคราะห์ข้อมูลการจำแนกเห็ดพิษหรือเห็ดที่รับประทานได้ในอเมริกาเหนือ คือการนำข้อมูลที่ได้ถูกรวบรวมไว้และถูกเผยแพร่ผ่านเว็บไซต์ kaggle.com ซึ่งเป็นเว็บไซต์ที่เก็บรวบรวมชุดข้อมูลต่าง ๆ เป็นแหล่งรวม Datasets หรือ ชุดข้อมูล สำหรับฝึกสอน Machine Learning ที่ใหญ่ที่สุดในโลกแห่งหนึ่ง มีข้อมูลทุกประเภทไม่ว่าจะเป็น Datasets ในหมวดหมู่ Finance, Business, Physics, Biology, Sports, News ซึ่งเป็นข้อมูลที่เปิดเผยได้ เพื่อให้ผู้ใช้บริการสามารถนำชุดข้อมูลไปศึกษาหรือวิเคราะห์ให้เกิดประโยชน์ต่อไปได้ การวิเคราะห์ข้อมูลปริมาณ

มหาศาลแสดงให้เห็นประโยชน์ต่าง ๆ ก็คือ ความรวดเร็วและประสิทธิภาพ ดำเนินการวิเคราะห์และค้นหาข้อมูลที่สามารถนำมาใช้ประกอบการตัดสินใจในอนาคตได้ ซึ่งการจำแนกเห็ดพิษหรือเห็ดที่รับประทานได้ในอเมริกาเหนือ มาทำการวิเคราะห์นั้น เนื่องจากต้องการทราบถึงข้อมูลในเชิงลึกของคุณสมบัติใดที่บ่งบอกถึงเห็ดพิษมากที่สุด เช่น รูปทรงของหมวกเห็ด พื้นผิวของหมวกเห็ด สีของหมวกเห็ด รอยช้ำ กลิ่น เป็นต้น เห็ดถือเป็นกลุ่มสิ่งมีชีวิตที่สำคัญในระบบนิเวศ มีความสัมพันธ์ที่หลากหลาย กับสิ่งมีชีวิตชนิดอื่น ๆ ในระบบนิเวศ นอกจากนี้เห็ดยังเป็นสิ่งมีชีวิตกลุ่มหลักที่ใช้ในการชี้วัดความอุดมสมบูรณ์ของสภาพป่า และเป็นแหล่งอาหารสำหรับความหลากหลายของเห็ด ซึ่งสมควรมีการศึกษาอย่างเป็นระบบและมีการจัดทำการจำแนกประเภทของเห็ด เพื่อเก็บรวบรวมพันธุ์เห็ดต่าง ๆ ไว้ใช้ประโยชน์ในการจัดจำแนกเห็ดรับประทานได้หรือเห็ดพิษ ใช้เป็นข้อมูลที่สามารถนำไปศึกษาต่อยอดได้

ดังนั้น ทางผู้วิเคราะห์จึงมองเห็นปัญหานี้ เพื่อนำไปสู่กระบวนการวิเคราะห์ผลการรายงานผล ในการวิเคราะห์ข้อมูลการจำแนกเห็ดพิษหรือเห็ดที่รับประทานได้ในอเมริกาเหนืออย่างถูกต้อง และแม่นยำ โดยนำข้อมูลมาวิเคราะห์หาคุณลักษณะใดที่บ่งบอกถึงเห็ดมีพิษมากที่สุด โดยมีปัจจัยที่เกี่ยวข้องเช่น รูปทรงของหมวกเห็ด พื้นผิวของหมวกเห็ด สีของหมวกเห็ด รอยช้ำ กลิ่น เป็นต้น ซึ่งได้ใช้เทคนิคการทำเหมืองข้อมูลแบบ Classification เพื่อให้เข้าใจถึงความสัมพันธ์ที่เกี่ยวข้องในการจำแนกเห็ดพิษหรือเห็ดที่รับประทานได้ ให้ทราบถึงคุณลักษณะของเห็ดพิษได้อย่างชัดเจน สามารถนำข้อมูลที่วิเคราะห์ได้นั้นนำไปใช้ ในการดำเนินการแก้ไข ช่วยลดการรับประทานเห็ดพิษอย่างมีประสิทธิภาพ และทางผู้วิเคราะห์ข้อมูลยังได้ทำการนำเสนอข้อมูลแบบ visualization ด้วยการแสดงผลข้อมูลในรูปแบบของภาพด้วยโปรแกรม Tableau Public และเผยแพร่ข้อมูลสารสนเทศนี้บน Web browser ที่เป็นที่นิยมในยุคอินเตอร์เน็ตคือการเผยแพร่ทางสื่อออนไลน์ ด้วยการสรุปจำแนกข้อมูลตามประเภทต่าง ๆ ได้สะดวกรวดเร็วยิ่งขึ้น รวมถึงมีความถูกต้อง สะดวก ลดความซ้ำซ้อนเพิ่มประสิทธิภาพให้กับการศึกษาค้นคว้า และทันต่อเวลา

1.2 วัตถุประสงค์

1.2.1 เพื่อวิเคราะห์ข้อมูลการจำแนกเห็ดพิษหรือเห็ดที่รับประทานได้ในอเมริกาเหนือ

1.2.2 เพื่อเผยแพร่ข้อมูลบนเว็บไซต์

1.3 ประโยชน์ที่จะได้รับ

1.3.1 ได้ข้อมูลสารสนเทศของการจำแนกเห็ดพิษหรือเห็ดที่รับประทานได้ในอเมริกาเหนือ

1.3.2 ได้เผยแพร่ข้อมูลบนเว็บไซต์

1.4 ขอบเขต

1.4.1 ขอบเขตผู้วิเคราะห์ข้อมูล

1.4.1.1 การรวบรวมข้อมูลที่ได้จากเว็บไซต์ kaggle.com

1.4.1.2 การทำเหมืองข้อมูลแบบ Classification เพื่อใช้ทำนายแนวโน้มการเกิดขึ้นของปัจจัย ที่บ่งบอกถึงเห็ดพิษด้วยการสร้างโมเดล Decision Tree

1.4.1.3 การประมวลผลโดยใช้โปรแกรม RapidMiner Studio และ โปรแกรม Weka

1.4.1.4 การนำเสนอข้อมูลแบบ Visualization ด้วยการแสดงผลข้อมูลในรูปแบบของกราฟ ด้วยโปรแกรม Tableau Public

1.4.1.5 การนำเสนอเผยแพร่ข้อมูลสารสนเทศนี้บน Web Browser โดยการใช้ภาษา HTML

1.4.1.6 ผู้วิเคราะห์ข้อมูลสามารถเพิ่มข้อมูลสารสนเทศเกี่ยวกับเห็ดบน Web Browser ได้

1.4.2 ขอบเขตผู้ใช้งานทั่วไปบนเว็บบราวเซอร์

1.4.2.1 สามารถดูข้อมูลสารสนเทศของการจำแนกเห็ดพิษหรือเห็ดที่รับประทานได้ในอเมริกาเหนือ ที่เผยแพร่ผ่านทางเว็บไซต์

1.4.2.2 สามารถดูข้อมูลสารสนเทศในลักษณะตารางสรุปข้อมูลบ่งชี้ถึงเห็ดพิษและเห็ดที่รับประทานได้

1.4.2.3 สามารถดูข้อมูลสารสนเทศในลักษณะรูปแบบของแผนภูมิจากโปรแกรม Tableau Public ได้

1.4.2.4 สามารถกรองข้อมูลสารสนเทศผ่านระบบของ Tableau Public บนWeb Browserได้

1.5 เครื่องมือที่ใช้ในการพัฒนาโปรแกรม

1.5.1 Hardware

1.5.1.1 Notebook: LENOVO IdeapadS410 core i5 Ram 4 GB

1.5.1.2 Notebook: Acer core i5 Ram 8 GB

1.5.2 Software

1.5.2.1 โปรแกรม Tableau Public ใช้สำหรับในการแสดงผลข้อมูลรูปแบบภาพ

1.5.2.2 ชุดคำสั่งภาษา HTML, CSS, JavaScript สำหรับพัฒนาหน้าจอแสดงผล

1.5.2.3 Adobe Dreamweaver ใช้ในการเขียนคำสั่งและออกเว็บไซต์

1.5.2.4 โปรแกรม Adobe XD ใช้สำหรับการออกแบบหน้าเว็บไซต์

1.5.2.5 ระบบปฏิบัติการ : Window 10 ใช้ในการเปิดซอฟต์แวร์ต่าง ๆ

1.5.2.6 โปรแกรม Microsoft SQL Server 2014 ใช้สำหรับจัดเตรียมข้อมูล

1.5.2.7 โปรแกรม Microsoft excel ใช้สำหรับการวิเคราะห์ข้อมูล

1.5.2.8 โปรแกรม Microsoft office 2016 ใช้สำหรับจัดการข้อมูลต่าง ๆ

1.6 สถานที่ใช้ในการดำเนินการศึกษาและรวบรวมข้อมูล

- สถานที่ใช้ในการดำเนินการศึกษา

มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนาภาคพายัพเชียงใหม่ 128 ถนนห้วยแก้ว ตำบลช้างเผือก อำเภอเมือง จังหวัดเชียงใหม่ 50300

- แหล่งรวบรวมข้อมูล

เว็บไซต์ Kaggle.com

1.6 ระยะเวลาในการดำเนินการ

ตารางที่ 1. 1 ระยะเวลาในการดำเนินการ

แผนการดำเนินการ	พ.ศ.2563						
	เม.ย	พ.ค	มิ.ย	ก.ค	ส.ค	ก.ย	ต.ค
1. ศึกษาและกำหนดความต้องการ	→						
2. ศึกษาข้อมูลและค้นหาข้อมูล	→	→					
3. การเตรียมข้อมูลและตรวจสอบความถูกต้อง			→				
4. วิเคราะห์ออกแบบฐานข้อมูล			→	→			
5. ออกแบบระบบ					→		
6. เขียนและทดสอบระบบ						→	
7. ตรวจสอบระบบโดยรวม							→
8. ประเมินผลและสรุปผลการวิเคราะห์ข้อมูล							→
9. จัดทำเอกสารประกอบโครงงาน		→	→	→	→	→	→

1.7 นิยามศัพท์เฉพาะ

1.7.1 อเมริกาเหนือ หมายความว่า ภูมิภาคอเมริกาเหนือมีเพียง 2 ประเทศ คือ ประเทศแคนาดา สหรัฐอเมริกา ซึ่งส่วนใหญ่ใช้ภาษาอังกฤษเป็นภาษาราชการ ส่วนอเมริกากลางใช้ภาษาสเปนเป็นหลัก ประกอบด้วยประเทศทั้งใหญ่และเล็ก รวมถึงประเทศที่เป็นหมู่เกาะจำนวน 22 ประเทศ โดยกรีนแลนด์เป็นประเทศอยู่เหนือที่สุด และประเทศปานามาอยู่ใต้สุด มีพรมแดนติดกับทวีปอเมริกาใต้

1.7.2 Visualization หมายความว่า การทำให้เห็นได้ หรือเรียกว่า จินตทัศน์ เป็นการกล่าวถึงการสร้างภาพ แผนผัง หรือ ภาพเคลื่อนไหว ใช้ในการสื่อสารแทนข้อความ โดยวิธีการนี้สามารถใช้ได้ทั้งในทางรูปธรรม และนามธรรม โดยมีทั้งการจำลองเหตุการณ์ที่เกิดขึ้นจริงในอดีต เหตุการณ์ที่ไม่สามารถมองเห็นได้ หรือการสร้างภาพในอนาคตเพื่อใช้ในการสื่อสาร การทำให้เห็นได้เป็นหนึ่งในวิธีการทำเหมืองข้อมูล โดยการสร้างภาพคอมพิวเตอร์กราฟิก ที่สามารถนำเสนอข้อมูลมากมายอย่างครบถ้วนแทนการใช้ข้อความนำเสนอข้อมูลที่มากมาย เราอาจพบข้อมูลที่ซ้อนเร้นเมื่อดูข้อมูลชุดนั้นด้วยการทำให้เห็นได้

1.7.3 หมวกเห็ด(cap) หมายถึง เป็นส่วนประกอบปลายสุดของดอกที่เจริญเติบโตขึ้นไปในอากาศ เมื่อดอกบาน เต็มที่จะกางออก มีลักษณะรูปทรงเหมือนร่มกาง ขอบซุ้มลงหรือแบนราบ

หรือกลางหมวกเว้าลง เป็นแอ่ง มีรูปเหมือนกรวยปากกว้าง ผิวหมวก เห็ดด้านบนอาจจะเรียบ ขรุขระ มีเกล็ด (Scales) หรือมีขน แตกต่างกันไปแล้วแต่ชนิดของเห็ด เกล็ดหรือขนเป็นเนื้อเยื่อที่ยังคงเหลือติดจากกลุ่มเนื้อเยื่อห่อหุ้มดอกเห็ดอ่อน (outer veil) เนื้อของดอกเห็ดมีความหนาบางต่างกัน อาจเหนียว หรือฉีกขาดได้ง่าย สีของเนื้อเห็ดภายในและภายนอก อาจเป็นสีเดียวหรือแตกต่างกัน

1.7.4 ครีบ(gill) หมายถึง ด้านล่างของหมวกเห็ดมีครีบหรือเรียงเป็นรัศมีรอบก้านดอกห้อยแขวนลงมาจาก เนื้อของหมวกเห็ดที่อยู่ตอนบน เห็ดบางชนิดมีครีบหมวกด้านในยึดติด หรือไม่ยึดติดกับก้านดอก ด้านนอกเชื่อมติดกับขอบหมวกสองข้างของครีบหมวกเป็นที่เก็ดสปอร์ของดอกเห็ด ครีบหมวกนั้น อาจถูกย่อยให้ละลายเป็นของเหลวในเห็ดบางชนิด เช่น เห็ดหิงห้อย หรือเห็ดน้ำหมึก เห็ดแต่ละชนิดมีจำนวนครีบหมวกแตกต่างกัน และความหนาบางไม่เท่ากัน จำนวนครีบหมวกจึง ใช้เป็นลักษณะประกอบการจำแนกเห็ดด้วย สีของครับหมวกส่วนมากเป็นสีเดียวกับสปอร์ของเห็ด ซึ่งจัดเป็นลักษณะแตกต่างของเห็ดแต่ละชนิดโดยปกติมีสีขาว เหลือง ชมพู ม่วง น้ำตาลและดำ เห็ดบางสกุลไม่มีครีบแต่จะมีรู (pore หรือ tube)หรืออาจมีลักษณะคล้ายหนาม(spine) แทนครีบที่จะมีสปอร์อยู่ภายใน บางชนิดสปอร์จะถูกฝังอยู่ในเนื้อเยื่อก้อนวุ้น เช่น เห็ดหูหนู หรือมีสปอร์เกิด อยู่ในเปลือกหุ้มที่เป็นก้อนกลม เช่น เห็ดเผาะ

1.7.5 ก้านดอก(stalk หรือ stipe) หมายถึง ส่วนมากเป็นรูปทรงกระบอก บางชนิดมีโคน หรือปลายเรียวเล็ก มีขนาดและความยาวแตกต่างกัน ตอนบนยึดติดกับหมวกเห็ดหรือครับหมวกด้านใน ตอนล่าง ของเห็ดบางชนิดอาจมีเส้นใยหยาบรวมกันเป็นก้อนหรือเปลือกหุ้มโคน (volva) ซึ่งมีลักษณะ คล้ายถ้วยชาหงายรองรับอยู่ เช่น เปลือกหุ้มโคนในเห็ดฟาง ฯลฯ บนก้านดอกตอนบนของเห็ด บางชนิดมีวงแหวน (ring) หรือเยื่อบาง (ม่าน หรือ annulus) หุ้มอยู่โดยรอบ ก้านดอกเห็ดมีผิวเรียบ ขรุขระ หรือมีขน หรือมีเกล็ด เมื่อถูกสัมผัสด้วยมือหรืออากาศอาจเปลี่ยนสีได้

1.7.6 วงแหวน(ring หรือ annulus) หมายถึง เนื้อเยื่อบาง ๆ ยึดติดก้านดอกใต้หมวกเห็ดลงมาเล็กน้อย เป็นส่วนของเนื้อเยื่อห่อหุ้มครับเมื่อดอกเห็ดยังอ่อนที่เรียกว่า inner veilวงแหวนนี้อาจเลื่อนขึ้นลงได้ไม่ยึดติดกับก้านดอกในเห็ดบางชนิด

1.7.7 เปลือกหุ้ม (volva) หมายถึง เนื้อเยื่อชั้นนอกสุดที่ห่อหุ้มดอกเห็ดทั้งดอกไว้ในระยะที่เป็นดอกอ่อน หรือ outer veil นั่นเอง ซึ่งมีในเห็ดบางชนิด เช่น เห็ดฟาง และในเห็ดพิษหลายชนิดในสกุล Amanita เมื่อดอกเห็ดขยายใหญ่ขึ้นเปลือกหุ้มตอนบนจะแตกออก เพื่อให้หมวกเห็ดและ

ก้านดอก ยืดตัวชูสูงขึ้นมาในอากาศ ทิ้งให้เปลือกหุ้มอยู่ที่โคนก้าน มองดูเหมือนก้านดอกเห็ดอยู่ในถ้วย

1.8 บทสรุป

จากบทนำที่ได้กล่าวมาในข้างต้นทั้งหมดนั้น ผู้วิเคราะห์ข้อมูลได้เล็งเห็นความสำคัญของการนำเทคโนโลยีมาประยุกต์ใช้ในการวิเคราะห์ข้อมูลการจำแนกเห็ดพิษหรือเห็ดที่รับประทานได้ในอเมริกาเหนือ ด้วยกระบวนการวิเคราะห์ข้อมูลด้วย CRISP-DM จากเทคนิคการทำเหมืองข้อมูลแบบ Classification ในรูปแบบของแผนภูมิต้นไม้ หรือที่เรียกว่า Decision tree และเผยแพร่ข้อมูลสารสนเทศบน web browser ที่เป็นที่นิยมในยุคอินเตอร์เน็ตคือการเผยแพร่ทางสื่อออนไลน์ โดยใช้ภาษา HTML ในการพัฒนาหน้าเว็บไซต์สำหรับการเปิดเผยข้อมูลที่ได้มาจาก kaggle.com เนื่องจากเทคโนโลยีปัจจุบันเข้ามาช่วยในการจัดการข้อมูล หรือดำเนินงานต่าง ๆ ให้มีความสะดวกสบายเป็นอย่างมาก อีกทั้งผู้ใช้งานยังเข้าถึงข้อมูลได้ง่าย และมีประสิทธิภาพ

บทที่ 2
แนวคิด ทฤษฎี เครื่องมือและวรรณกรรมที่เกี่ยวข้อง

ในบทนี้เป็นการนำเสนอเกี่ยวกับ แนวคิด ทฤษฎี เครื่องมือและวรรณกรรมที่เกี่ยวข้องของการวิเคราะห์ข้อมูลการจำแนกเห็ดพิษหรือเห็ดที่รับประทานได้ในอเมริกาเหนือ และการแสดงผลข้อมูลบนเว็บไซต์ ซึ่งได้รวบรวมการศึกษาเอกสารงานวิจัยที่เกี่ยวข้องกับการวิเคราะห์ข้อมูล เพื่อใช้เป็นแนวทางการศึกษาประกอบด้วยรายละเอียดตามลำดับ 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)

ในการดำเนินงานเรื่องการวิเคราะห์ข้อมูลการจำแนกเห็ดพิษหรือเห็ดที่รับประทานได้ในอเมริกาเหนือ และการแสดงผลข้อมูลบนเว็บไซต์ ผู้วิเคราะห์ข้อมูลได้ศึกษาหลัก การและทฤษฎีต่าง ๆ องค์ประกอบหนึ่งที่สำคัญคือการวิเคราะห์ข้อมูล ซึ่งมีรายละเอียด ดังนี้

2.1.1.1 Data analytics เป็นศาสตร์ของการวิเคราะห์ข้อมูลต่าง ๆ จาก Big data เพื่อช่วยในด้านธุรกิจหรือตามวัตถุประสงค์อื่น ๆ ที่ต้องการ ซึ่งเริ่มต้นจากความสามารถของเรา ในการนำข้อมูลหล่านั้น มาให้อยู่ในรูปแบบที่พร้อมจะประมวลผลได้โดยเทคโนโลยีหรือชุดคำสั่งและแบบจำลองที่สร้างขึ้นเพื่อนำ ข้อมูลที่ได้ผ่านการวิเคราะห์แล้วมาใช้ หรือ แปลความหมายโดยบุคคลที่ได้รับการฝึกอบรมการใช้เทคโนโลยีเหล่านั้น รูปแบบของการวิเคราะห์ข้อมูลสามารถแบ่งได้ดังนี้

1) การวิเคราะห์ข้อมูลแบบพื้นฐาน (Descriptive analytics) เป็นการวิเคราะห์เพื่อแสดงผลของรายการทางธุรกิจ เหตุการณ์ หรือกิจกรรมต่าง ๆ ที่ได้เกิดขึ้น หรืออาจกำลังเกิดขึ้นในลักษณะที่ง่ายต่อการเข้าใจ หรือต่อการตัดสินใจตัวอย่างเช่นรายงานการขายรายงานผลการดำเนินงาน

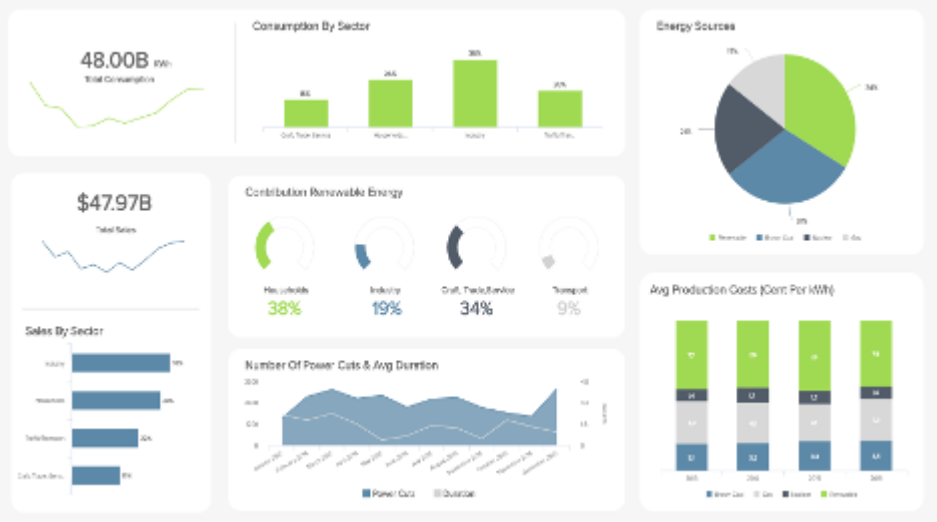
2) การวิเคราะห์แบบเชิงวินิจฉัย (Diagnostic analytics) เป็นการอธิบายถึงสาเหตุของสิ่งที่เกิดขึ้นปัจจัยต่าง ๆ และความสัมพันธ์ของปัจจัยหรือตัวแปรต่าง ๆ ที่มีความสัมพันธ์ต่อกันของสิ่งที่เกิดขึ้น ตัวอย่างเช่นความสัมพันธ์ระหว่างยอดขายต่อกิจกรรมทางการตลาดแต่ละประเภท ซึ่งเป็นก้าวใหม่ที่ช่วยเสริมให้ตัดสินใจไปในทางที่ถูกต้อง

3) การวิเคราะห์แบบพยากรณ์ (Predictive analytics) เป็นการวิเคราะห์เพื่อพยากรณ์สิ่งที่ กำลังจะเกิดขึ้นหรือน่าจะเกิดขึ้น โดยใช้ข้อมูลที่ได้เกิดขึ้นแล้วกับแบบจำลองทางสถิติ หรือ เทคโนโลยีปัญญาประดิษฐ์ต่าง ๆ (Artificial intelligence) ตัวอย่าง เช่น การพยากรณ์ยอดขาย การพยากรณ์ผลประชามติ

4) การวิเคราะห์แบบให้คำแนะนำ (Prescriptive analytics) เป็นการวิเคราะห์ข้อมูลที่มีความซับซ้อนที่สุด เป็นทั้งการพยากรณ์สิ่งต่าง ๆ ที่จะเกิดขึ้นข้อดี ข้อเสีย สาเหตุ และระยะเวลาของสิ่งที่จะเกิดขึ้นและการให้คำแนะนำทางเลือกต่าง ๆ ที่มีอยู่และผลของแต่ละทางเลือก

2.1.1.2 แนวคิดเกี่ยวกับการแสดงผลข้อมูล (Data visualization)

Data Visualization หรือ Information Visualization เป็นเครื่องมือสำคัญที่ช่วยให้เห็นภาพของข้อมูลเชิงลึกที่ผ่านการเก็บข้อมูล การจัดการข้อมูล การวิเคราะห์ข้อมูล แล้วรวบรวมได้เป็นชุดข้อมูล หากต้องการประสบความสำเร็จในโลกธุรกิจที่มีการแข่งขันสูง ความสามารถในการตัดสินใจที่ถูกต้องในเวลาที่เหมาะสม โดยอาศัยข้อมูลเชิงลึกที่รัดกุมเป็นสิ่งสำคัญ ปัจจุบันองค์กรธุรกิจส่วนใหญ่ในประเทศไทย เริ่มใช้ประโยชน์จากการทำ Data Visualization เพราะว่าเครื่องมือนี้เป็นกุญแจสำคัญในการแสดงผลของข้อมูลให้เข้าใจง่าย ซึ่งจะช่วยให้ทุกแผนกในฝ่ายงานได้ใช้ประโยชน์ ในการหาแนวโน้มและความสัมพันธ์เชื่อมโยง อีกทั้งทำให้ทราบข้อมูลได้เร็วขึ้น ส่งผลให้การทำงานขององค์กรมีประสิทธิภาพเพิ่มขึ้น จนดำเนินการถึงการวางกลยุทธ์ภายในองค์กร และสามารถนำไปใช้กับการทำเว็บไซต์ได้ Data Visualization ช่วยให้การสื่อสารข้อมูล มีความรวดเร็วและเข้าใจง่ายมากขึ้น



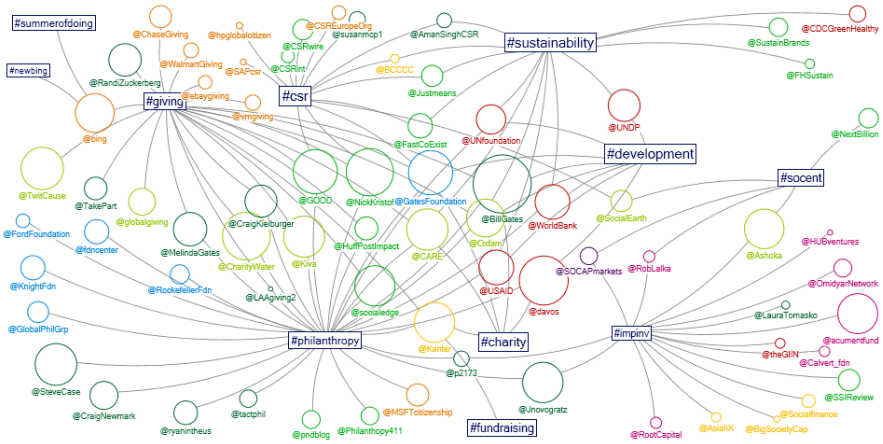
ภาพที่ 2. 1 แสดง Data Visualization

(ที่มา : https://bit.ly/31GJ4pE)

พื้นฐานของการสร้าง Data Visualization คือ การ Mapping ส่วนข้อมูลกับส่วนของ Graphic เข้าด้วยกัน ซึ่งตอนนี้มีโปรแกรมสำเร็จรูปในการสร้าง Data visualization หลากหลายโปรแกรม มีฟังก์ชันการใช้งานที่เข้าใจง่าย เช่น การสร้างฟิลเตอร์ การออกแบบเพื่อให้งานการวิเคราะห์ข้อมูลมีความยืดหยุ่น เป็นต้น ตัวอย่างรูปแบบ Data Visualization ที่แสดงผลให้อย่างสนใจ มีดังนี้

1. NETWORK/FLOW

Network Graph ใช้แสดงความเชื่อมโยงของเครือข่ายหรือความสัมพันธ์ในกลุ่ม

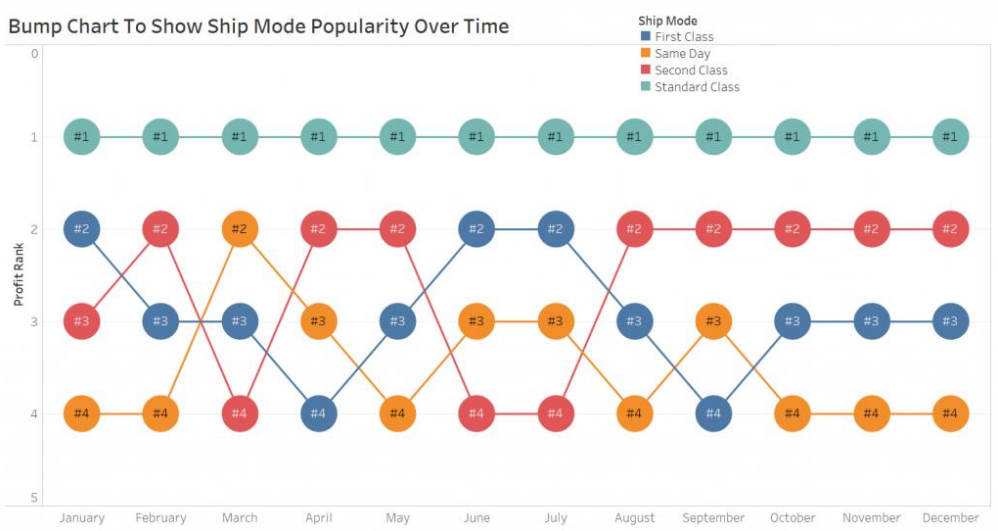


ภาพที่ 2. 2 แสดง Network Graph

(ที่มา : https://bit.ly/31GJ4pE)

2. RANKING

Bump Chart ใช้แสดงการเปลี่ยนแปลงลำดับของข้อมูลในหลายช่วงเวลา โดยใช้สีเพื่อช่วยให้อ่านข้อมูลได้ง่ายขึ้น

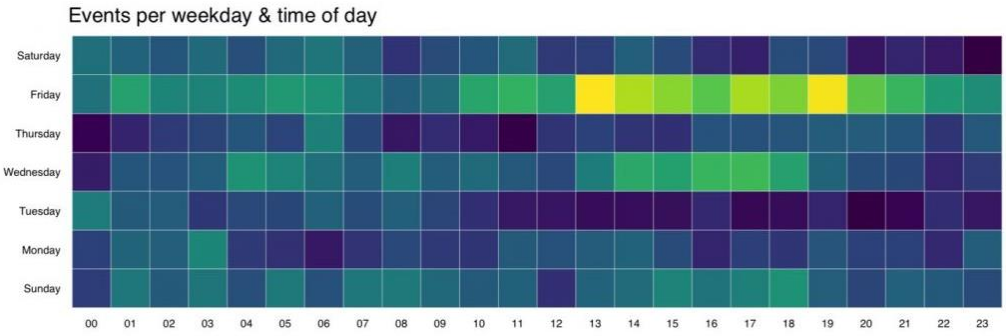


ภาพที่ 2. 3 แสดง Bump Chart

(ที่มา : https://bit.ly/31GJ4pE)

3. RELATIONSHIP

Heat map ใช้แสดงรูปแบบความสัมพันธ์ของข้อมูล โดยจะแสดงออกมาในรูปแบบของ “สี” ซึ่งแต่ละสีจะบ่งบอกถึงระดับความถี่ของพฤติกรรม แต่ใช้อ่านค่าความแตกต่างเล็กน้อยได้ยาก

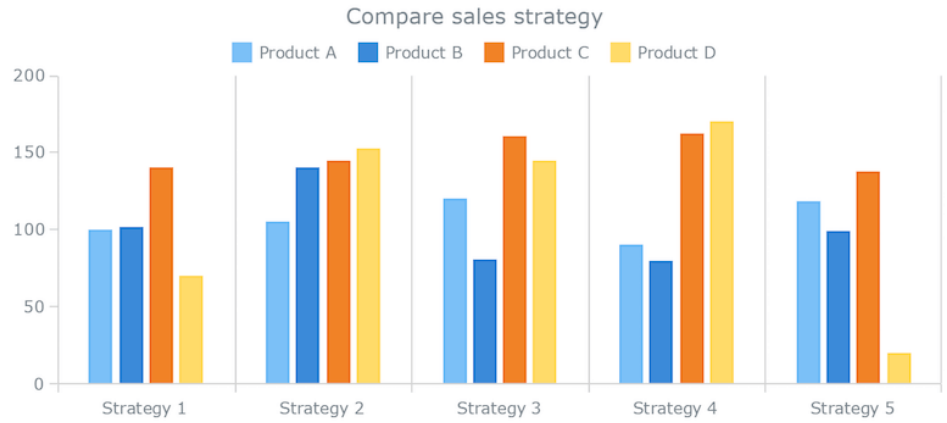


ภาพที่ 2. 4 แสดง Heat map

(ที่มา : https://bit.ly/31GJ4pE)

4. COMPARISON

Stacked Bar Chart ใช้เปรียบเทียบค่าผลรวมและสัดส่วนจากข้อมูลหลายกลุ่ม อาจอ่านสัดส่วนยากเมื่อมีข้อมูลหลายกลุ่มมากเรับประทานไป

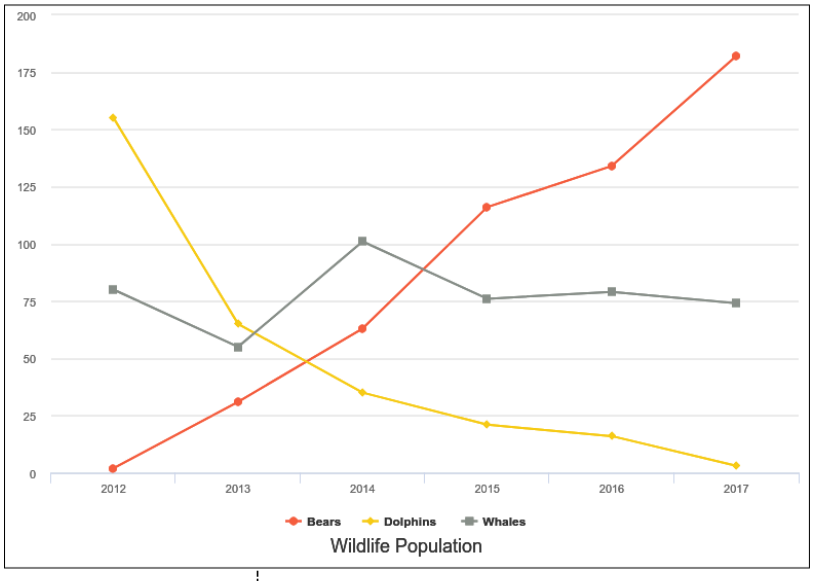


ภาพที่ 2. 5 แสดง Stacked Bar Chart

(ที่มา : https://bit.ly/31GJ4pE)

5. TIME–SERIES

Line Chart ใช้เปรียบเทียบข้อมูลเพื่อดูแนวโน้ม (Trends) โดยอาจจะเทียบกับมิติของเวลา time series

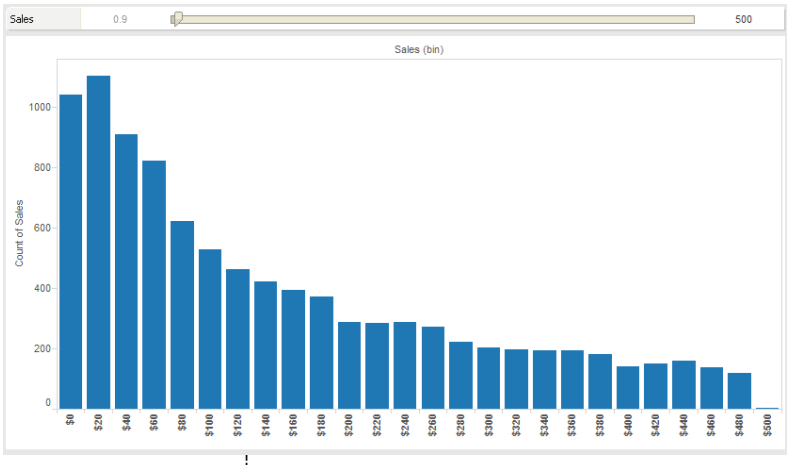


ภาพที่ 2. 6 แสดง Line Chart

(ที่มา : https://bit.ly/31GJ4pE)

6. DISTRIBUTION

Histogram กราฟแท่งแบบเฉพาะที่แสดงความสัมพันธ์ระหว่างข้อมูลเป็นหมวดหมู่ที่เรียกว่าชั้นข้อมูลกับความถี่ของข้อมูล เพื่อดูการกระจายของข้อมูล ลดระยะห่างระหว่างแท่งเพื่อให้เห็น “รูปทรง” ของข้อมูลชัดขึ้น

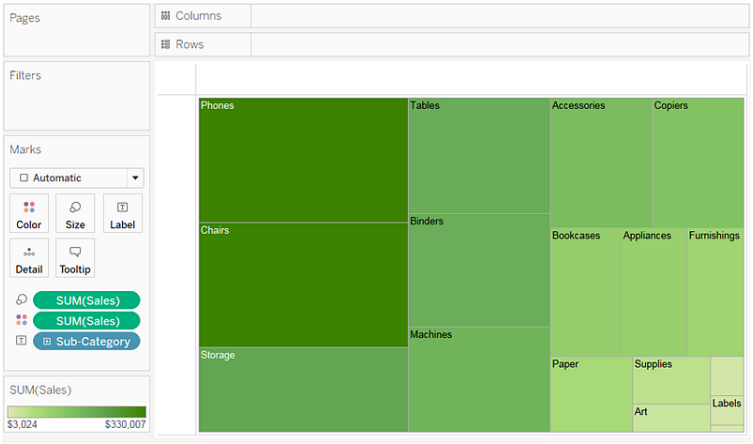


ภาพที่ 2. 7 แสดง Histogram

(ที่มา : https://bit.ly/31GJ4pE)

7. COMPOSITION

Tree map ใช้แสดงสัดส่วนของข้อมูลแบบมีลำดับชั้น (hierarchy) แสดงเป็นรูปสี่เหลี่ยม โดยจะแสดงทั้งในแง่ของสีเดียวกันจัดเป็นกลุ่มเดียวกัน อาจอ่านสัดส่วนยากเมื่อมีกลุ่มขนาดเล็กจำนวนมาก

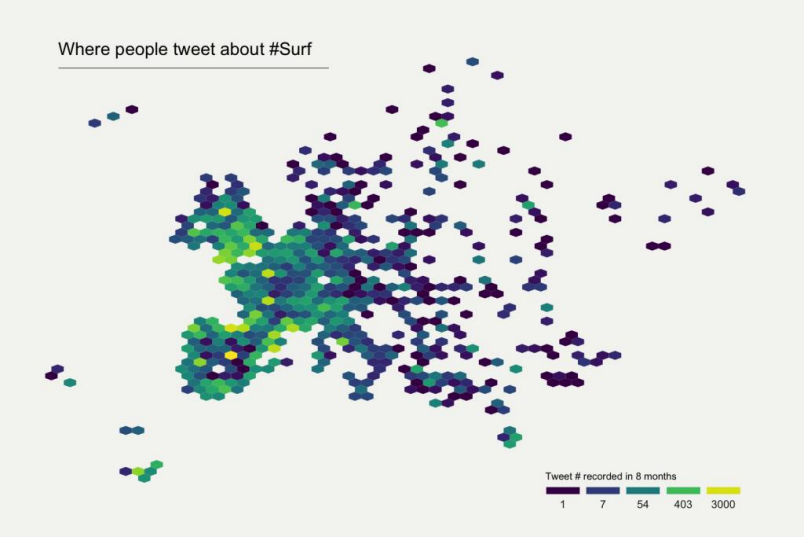


ภาพที่ 2. 8 แสดง Tree map

(ที่มา : https://bit.ly/31GJ4pE)

8. SPATIAL

Hexbin Map เป็นการแสดงผลข้อมูลด้วยรูปทรงหกเหลี่ยมที่มีขนาดเท่ากัน โดยกำหนดสีแทนระดับความถี่ของจำนวนข้อมูล ส่วนมากนำไปใช้ลงบนตำแหน่งทางภูมิศาสตร์ เพื่อให้เห็นภาพได้ชัดเจน



ภาพที่ 2. 9 แสดง Hexbin Map

(ที่มา : https://bit.ly/31GJ4pE)

จากที่ได้กล่าวไปจะเห็นได้ว่าข้อมูลแบบ Data Visualization มีพลังทางการตลาดอย่างมาก เพราะสามารถแสดงผลและเพิ่มความสนใจได้เป็นอย่างดี นอกจากนี้ยังทำให้ผู้ได้รับข้อมูลสามารถจินตนาการตามข้อมูลที่แสดงได้อีกด้วย หากองค์กรได้ลองนำการแสดงข้อมูลแบบ Data Visualization มาใช้ก็จะช่วยในการจัดการระดับสายงานแผนกอื่น ๆ ใน บริษัท ส่งผลให้การวิเคราะห์ข้อมูลและตัดสินใจในสายงานต่าง ๆ มีประสิทธิภาพ ซึ่งดีต่อธุรกิจและทีม หรือนำมาใช้ในหน้าแรกของเว็บไซต์ เพื่อทำการโปรโมทโปรโมชั่นหรือข้อเสนอทางการตลาดของธุรกิจ แล้วจะพบว่า Data Visualization ช่วยให้เกิดการตอบสนองกับข้อความมากขึ้น

2.2 ทฤษฎี

2.2.1 ทฤษฎีเกี่ยวกับการทำเหมืองข้อมูล

การทำเหมืองข้อมูลหรือ Data Mining เทคโนโลยีเกี่ยวกับฐานข้อมูลได้เริ่มต้นขึ้นตั้งแต่ปี 1960 โดยพัฒนามาจาก file processing จากนั้นพัฒนามาในรูปแบบตารางและเริ่มมีการใช้ภาษาในการเรียกข้อมูล (Query Language) ข้อมูลที่เก็บเริ่มมีจำนวนมากขึ้นทำให้การจัดเก็บมีความซับซ้อนและมีประสิทธิภาพ การเพิ่มขึ้นของข้อมูลเป็นจำนวนมากและสะสมไว้ในฐานข้อมูลนั้นมีขนาดใหญ่เรับประทานกว่าที่กำลังคนจัดการได้ เป็นผลให้เกิดเครื่องมือที่ช่วยในการวิเคราะห์ข้อมูลและหาความเป็นไปได้ของข้อมูลที่เป็นประโยชน์

วิวัฒนาการของการทำเหมืองข้อมูล

- ปี 1960 Data Collection คือ การนำข้อมูลมาจัดเก็บอย่างเหมาะสมในอุปกรณ์ที่น่าเชื่อถือและป้องกันการสูญหายได้เป็นอย่างดี
- ปี 1980 Data Access คือ การนำข้อมูลที่จัดเก็บมาสร้างความสัมพันธ์ต่อกันในข้อมูลเพื่อประโยชน์ใน การนำไปวิเคราะห์ และการตัดสินใจอย่างมีคุณภาพ
- ปี 1990 Data Warehouse & Decision Support คือ การรวบรวมข้อมูลมาจัดเก็บลงไปในฐานข้อมูลขนาดใหญ่โดยครอบคลุมทุกด้านขององค์กร เพื่อช่วยสนับสนุนการตัดสินใจ
- ปี 2000 Data Mining คือ การนำข้อมูลจากฐานข้อมูลมาวิเคราะห์และประมวลผลโดยการสร้างแบบจำลองและความสัมพันธ์ทางสถิติ

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

การทำเหมืองข้อมูล (Data Mining) คือ วิธีการค้นหาความรู้ความสัมพันธ์ และรูปแบบทั้งหมดที่ถูกซ่อนอยู่ในฐานข้อมูลที่มีปริมาณข้อมูลจำนวนมาก การทำเหมืองข้อมูลสามารถทำการสำรวจ และวิเคราะห์อย่างอัตโนมัติ จากข้อมูลที่มีอยู่ในรูปแบบความหมาย และรูปแบบของกฎ โดยที่ความสัมพันธ์แสดงให้เห็นถึงความรู้ต่าง ๆ ที่เป็นประโยชน์และน่าสนใจ ประโยชน์หลักของ

การทำเหมืองข้อมูลคือการสืบค้นความรู้ในฐานข้อมูลขนาดใหญ่เพื่อให้ได้ความรู้มาใช้ในการตัดสินใจ

การทำเหมืองข้อมูลหากผนวกเข้ากับระบบการจัดการความรู้ที่ได้จากฐานข้อมูลจะทำให้ได้ระบบการค้นหาข้อมูลและความรู้(Knowledge and Data Discovery Management System:KDDMS) เป็นการทำเหมืองข้อมูลที่ใช้เทคนิคการกลั่นกรอง วิเคราะห์ความรู้ที่ซ่อนอยู่ในฐานข้อมูลที่มีปริมาณมาก เพื่อได้ข้อมูลที่ต้องการและมีประโยชน์ สามารถนำข้อมูลที่ได้บริหารจัดการได้อย่างมีประสิทธิภาพ โดยผลลัพธ์มีอยู่หลายแบบ เช่น รูปแบบ (Pattern)ต้นแบบ (Model) กฎ (Rule) การทำนายหรือการคาดการณ์ล่วงหน้า (Prediction) ทั้งผลลัพธ์ที่ได้มาขึ้นอยู่กับการทำเหมืองข้อมูล

2.2.2 ทฤษฎีเกี่ยวกับการออกแบบเว็บไซต์

การออกแบบเว็บไซต์นั้นมีความสำคัญอย่างมากเนื่องจากจะทำให้ผู้ชมเกิดความประทับใจ ชื่นชอบ จนสามารถดึงดูดความสนใจได้จนกระทั่งมีความต้องการที่จะกลับเข้าชมเว็บนั้น ๆ อีกครั้งในอนาคต สำหรับการออกแบบเว็บไซต์นั้น หมายถึงการออกแบบพัฒนาส่วนหน้าแสดงของเว็บให้เกิดความประทับใจแก่ผู้ชม และนำเสนอความเป็นเอกลักษณ์ของหน่วยงานได้ด้วยโปรแกรมฟังก์ชันการใช้งานภายในเว็บไซต์ที่ทำให้ผู้ใช้อยากเข้ามามีปฏิสัมพันธ์ต่อเว็บไซต์นั้น โดยต้องให้ความสำคัญกับปัจจัยต่าง ๆ ดังต่อไปนี้

2.2.2.1 ความสำคัญของการออกแบบเว็บไซต์

เว็บไซต์เป็นสื่อที่อยู่ในการควบคุมของผู้ใช้โดยสมบูรณ์ ดังนั้นการออกแบบเว็บจึงมีความสำคัญที่จะดึงดูดให้ผู้ใช้เข้าชมและทำกิจกรรมร่วมในเว็บไซต์ของเราให้ได้มากที่สุด ซึ่งสามารถสรุปปัญหาต่าง ๆ เกี่ยวกับผู้ใช้ซึ่งเป็นองค์ประกอบหนึ่งในการกำหนดความสำเร็จหรือล้มเหลวของการสร้างเว็บไซต์ ดังต่อไปนี้

1. ผู้ใช้มักไม่อดทนกับอุปสรรคและปัญหาที่เกิดขึ้นจากการออกแบบการใช้งานผิดพลาด

2. ผู้ใช้มีทางเลือกมากขึ้นที่จะเข้าใช้บริการเว็บไซต์อื่น ถ้าหากการออกแบบเว็บไซต์นั้นนั้นเกิดความรู้สึกไม่ประทับใจ

3. ผู้ใช้หลายคนชื่นชอบเว็บไซต์ที่สวยงาม และมีความสะดวกในการใช้งานมากกว่าเว็บที่ดูสับสนวุ่นวาย มีข้อมูลจำนวนมากแต่ค้นหาข้อมูลที่ต้องการไม่พบ หรือใช้เวลามากในการแสดงหน้าเว็บ

4. หากผู้ใช้มีความประทับใจกับเว็บไซต์หนึ่ง ๆ แล้วจะทำให้อยากกลับมาเข้าใช้อีกครั้ง

2.2.2.2 องค์ประกอบของการออกแบบเว็บไซต์อย่างมีประสิทธิภาพ

การออกแบบเว็บไซต์ที่ดีนั้น จะต้องออกแบบให้เหมาะสมกับกลุ่มเป้าหมาย และ จะต้องคำนึงถึงความสะดวกของผู้ใช้งานเป็นหลัก โดยที่ผู้ออกแบบนั้นต้องให้ความสมดุลระหว่าง ความสวยงามและความสะดวก ซึ่งต้องนึกถึงองค์ประกอบดังต่อไปนี้

1) ความเรียบง่าย (Simplicity) คือ การสื่อเนื้อหาถึงผู้ใช้โดยจำกัดองค์ประกอบเสริมที่ เกี่ยวข้องกับการนำเสนอให้เหลือแต่สิ่งที่จำเป็นเท่านั้น ทั้งการแสดงออกในด้านเนื้อหา ตัวอักษร หรือแม้แต่สีที่ใช้ ทำให้ผู้ใช้รู้สึกใช้งานง่ายและสะดวกไม่ก่อให้เกิดความรู้สึกรำคาญใด ๆ

2) ความสม่ำเสมอ (Consistency) คือ การสื่อให้ผู้ใช้รู้ว่าเป็นเนื้อหาเดียวตลอดทั้งเว็บไซต์ โดยไม่ก็ให้เกิดความสับสนแก่ผู้ใช้ดังนั้นผู้ออกแบบเว็บไซต์ควรออกแบบระบบ Navigation และโทนสีให้มีความสม่ำเสมอคล้ายคลึงกันตลอดทั้งเว็บไซต์

3) ความเป็นเอกลักษณ์ (Identity) คือ ต้องออกแบบเพื่อแสดงถึงความเป็นเอกลักษณ์ของ องค์กร ทั้งชุดสี ตัวอักษร รูปภาพกราฟิก เป็นสื่อถึงความเป็นเอกลักษณ์แทบทั้งสิ้น

4) เนื้อหาที่ประประโยชน์ (Useful content) เนื้อที่แสดงออกของเว็บไซต์นั้นเป็นสิ่งที่ สำคัญที่สุดซึ่งต้องแสดงออกถึงความถูกต้องและสมบูรณ์ โดยไม่ฐ้ากับเว็บไซต์อื่น และเป็นเนื้อหา ที่ทันต่อเหตุการณ์ เนื่องจากเป็นสิ่งที่ดึงดูดผู้เข้ามาใช้งานเว็บไซต์

5) ระบบ Navigation ที่ใช้งานง่าย (User–friendly navigation) เป็นส่วนที่สำคัญของ เว็บไซต์เนื่องจากจะทำให้ผู้ใช้ไม่เกิดความสับสนเวลาใช้งานเว็บไซต์ ซึ่งผู้ออกแบบจำเป็นต้องออกแบบให้ผู้ใช้ใช้งานได้ง่ายและสะดวก เช่นถ้าใช้ภาพกราฟิกที่สื่อความหมายควรมีคำอธิบายที่ชัดเจนร่วมด้วย แสดงไว้ส่วนบนหรือส่วนล่างของหน้าเว็บไซต์

6) มีลักษณะที่น่าสนใจ (Visual appeal) ควรสื่อให้เว็บไซต์มีคุณภาพสัมพันธ์กับ องค์ประกอบต่าง ๆ เช่น คุณภาพของกราฟิกที่แสดงบนเว็บไซต์นั้นต้องสมบูรณ์ไม่มีร่องรอยของความเสียหายเป็นจุดด่าง หรือมีขอบขั้นบันได สำหรับตัวอักษรต้องเป็นตัวอักษรที่อ่านง่าย และ สบายตาร่วมกับโทนสีที่สวยงาม

7) การใช้งานได้อย่างไม่จำกัด (Compatible) การออกแบบเว็บไซต์นั้นต้องทำให้คนส่วน ใหญ่เข้าถึงให้ได้มากที่สุด โดยไม่มีการบังคับให้ติดตั้งโปรแกรมใด ๆ เพิ่มเติมสามารถแสดงผลได้ ทุกระบบปฏิบัติและที่ความละเอียดของหน้าจอต่าง ๆ กัน

8) คุณภาพการออกแบบ (Design Stability) ต้องออกแบบให้เว็บไซต์มีเนื้อหาที่น่าเชื่อถือ เพื่อดึงดูดให้มีผู้เข้าชมเว็บไซต์เกิดความประทับใจ และมีความเชื่อถือในตัวเว็บไซต์

9) ระบบการใช้งานที่ถูกต้อง (Functional stability) ต้องออกแบบให้เว็บไซต์นั้นมีความแน่นอน ทำหน้าที่ได้อย่างถูกต้อง และสามารถใช้งานได้จริง เช่น ถ้ามีแบบฟอร์มให้ผู้ใช้กรอก ข้อมูล ก็ต้องทดสอบก่อนว่าสามารถใช้งานได้จริง หรือลิงค์ต่าง ๆ ที่แสดงอยู่บนหน้าเว็บไซต์จะต้องเชื่อมโยงไปยังแหล่งข้อมูลนั้น ๆ ได้จริง

2.2.2.3 รูปแบบโครงสร้างของเว็บไซต์

การออกแบบโครงสร้างของเว็บไซต์ สามารถทำได้หลากหลายแบบ ซึ่งก็ขึ้นอยู่กับความชอบและความถนัดของแต่ละบุคคล นอกจากนี้ยังขึ้นอยู่กับกลุ่มเป้าหมายที่ต้องการนำเสนอ เพราะจะต้องออกแบบให้เหมาะกับการใช้งานของกลุ่มเป้าหมายมากที่สุด โดยโครงสร้างของเว็บไซต์ส่วนใหญ่ก็จะประกอบไปด้วย 4 รูปแบบดังนี้

1) โครงสร้างเว็บไซต์แบบเรียงลำดับ

จะเป็นโครงสร้างแบบธรรมดาที่นิยมใช้งานกันมากที่สุด เนื่องจากมีความง่ายต่อการจัดระบบข้อมูล และสามารถนำเสนอเรื่องราวตามลำดับได้เป็นอย่างดี เหมาะกับเว็บไซต์ที่มีขนาดเล็ก มีเนื้อหาที่ไม่ซับซ้อน ส่วนใหญ่ก็จะเป็นพวกเว็บไซต์ที่ให้ความรู้ หรือเว็บไซต์องค์กรขนาดย่อม โดยลักษณะการลิงค์เนื้อหา ก็จะลิงค์ไปทีละหน้า มีทิศทางการเข้าสู่เนื้อหาต่าง ๆ ในแบบเส้นตรง ใช้ปุ่มเดินหน้า–ถอยหลังในการกำหนดทิศทาง จึงทำให้การใช้งานเป็นไปอย่างง่าย แต่โครงสร้างเว็บไซต์แบบเรียงลำดับก็มีข้อเสีย คือจะทำให้ผู้ใช้งานต้องเสียเวลาในการเข้าสู่เนื้อหา เพราะไม่สามารถกำหนดทิศทางการเข้าสู่เนื้อหาด้วยตัวเองได้

ภาพที่ 2. 10 แสดงโครงสร้างเว็บไซต์แบบเรียงลำดับ

(ที่มา : https://bit.ly/2VI4fDW)

2) โครงสร้างเว็บไซต์แบบลำดับขั้น

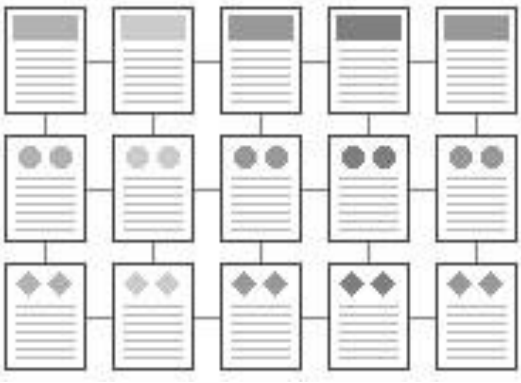
นิยมใช้กับเว็บที่มีความซับซ้อนของข้อมูล เพื่อให้สามารถเข้าถึงข้อมูลต่างๆ ได้ง่ายขึ้น โดยจะมีการแบ่งเนื้อหาออกเป็นส่วนๆ และมีการนำเสนอรายละเอียดย่อยๆ ที่ลดหลั่นกันมา ทำให้สามารถทำความข้าใจกับโครงสร้างเนื้อหาได้ง่ายขึ้น โดยจะมีโฮมเพจเป็นจุดเริ่มต้นและจุดร่วมจุดเดียวที่จะนำไปสู่การเชื่อมโยงเนื้อหาเป็นลำดับจากบนลงล่าง

ภาพที่ 2. 11 แสดงโครงสร้างเว็บไซต์แบบลำดับขั้น

(ที่มา : https://bit.ly/2VI4fDW)

3) โครงสร้างเว็บไซต์แบบตาราง

เป็นโครงสร้างการออกแบบเว็บไซต์ที่มีความซับซ้อน แต่ก็มีความยืดหยุ่นในระดับหนึ่ง เพื่อให้ผู้ใช้งานสามารถเข้าสู่เนื้อหาต่าง ๆ ได้ง่ายขึ้น การออกแบบในลักษณะนี้จะมีการเชื่อมโยงเนื้อหาในแต่ละส่วนซึ่งกันและกัน ทำให้ผู้ใช้งานสามารถเปลี่ยนทิศทาง หรือกำหนดทิศทางในการเข้าสู่เนื้อหาด้วยตัวเองได้ จึงไม่ทำให้เสียเวลา แถมยังทำให้เว็บไซต์มีความทันสมัยขึ้น

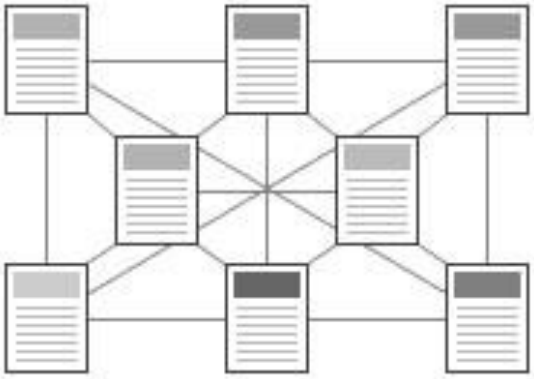


ภาพที่ 2. 12 แสดงโครงสร้างเว็บไซต์แบบตาราง

(ที่มา : https://bit.ly/2VI4fDW)

4) โครงสร้างแบบใยแมงมุม

เป็นโครงสร้างที่ได้รับความนิยมเป็นอย่างมาก เพราะมีความยืดหยุ่นมากที่สุด โดยทุกหน้าเว็บจะมีการเชื่อมโยงถึงกันหมด ทำให้สามารถเข้าถึงหน้าเว็บเพจต่าง ๆ ที่ต้องการได้อย่างง่าย และมีความอิสระมากขึ้น นอกจากนี้ก็สามารถเชื่อมโยงไปสู่เว็บไซต์ภายนอกได้ดี



ภาพที่ 2. 13 แสดงโครงสร้างแบบใยแมงมุม

(ที่มา : https://bit.ly/2VI4fDW)

ส่วนประกอบสำคัญบนหน้าเว็บเพจ

บนหน้าเว็บเพจ จะมีส่วนประกอบสำคัญที่จำเป็นต้องมีอยู่ 3 ส่วน ได้แก่

1. ส่วนหัวของหน้า (Header)

อยู่ตอนบนสุดของหน้าและเป็นส่วนที่สำคัญที่สุด โดยจะต้องทำให้สามารถดึงดูดผู้ชมให้รู้สึกอยากติดตามเนื้อหาในเว็บไซต์ต่อไป ซึ่งส่วนใหญ่ก็มักจะมีการใส่ภาพกราฟิกให้ดูสวยงาม สิ่งสำคัญหลักๆ เลย ก็คือ โลโก้ ชื่อเว็บไซต์และเมนูหลักที่สามารถลิงค์ไปยังเนื้อหาในหน้าเว็บเพจต่าง ๆ ได้

2. ส่วนของเนื้อหา (Body)

อยู่บริเวณตอนกลางของหน้าเว็บ โดยจะแสดงข้อมูลเกี่ยวกับเนื้อหาบนเว็บแบบคร่าวๆ ซึ่งก็จะมีข้อความกราฟิก ตารางข้อมูลหรือวิดีโอประกอบอยู่ และหากมีเมนูแบบเฉพาะกลุ่มก็จะถูกจัดไว้ในหน้านี้เช่นกัน และที่สำคัญเนื้อหาในส่วนนี้ควรจะมีความกระชับ เข้าใจง่าย มีการใช้รูปแบบตัวอักษรแบบเรียบง่ายและเป็นระเบียบ

3. ส่วนท้ายของหน้า (Footer)

อยู่ล่างสุดของหน้าเว็บ ซึ่งจะมีหรือไม่มีก็ได้ ส่วนนี้จะแสดงถึงข้อมูลต่าง ๆ เพิ่มเติมเข้าไป เช่น ข้อความที่แสดงถึงการเป็นลิขสิทธิ์ ข้อมูลเจ้าของเว็บไซต์ วิธีการติดต่อและคำแนะนำต่าง ๆ เกี่ยวกับการใช้งานเว็บไซต์อย่างถูกต้อง เป็นต้น

วิธีการเลือกใช้สีสำหรับการออกแบบเว็บไซต์

การเลือกใช้สีในการออกแบบเว็บไซต์มีความสำคัญเป็นอย่างมาก เพราะสีสามารถกำหนดอารมณ์ ความรู้สึกและกระตุ้นการรับรู้ทางด้านจิตใจของมนุษย์ได้ดี ดังนั้นสีที่ใช้จึงต้องมีความสอดคล้องกับเนื้อหาและจุดประสงค์ของเว็บ ว่าต้องการให้ผู้เข้าชมรู้สึกอย่างไรต่อเนื้อหาที่ได้อ่าน โดยรูปแบบของสีที่สายตาของมนุษย์สามารถมองเห็นได้ก็แบ่งออกเป็น 3 กลุ่มดังต่อไปนี้

1. สีโทนร้อน (Warm Colors) เป็นสีแห่งความอบอุ่น ปลอบโยนและกระตุ้นความสุขได้ดี ซึ่งจะทำให้ผู้เข้าชมรู้สึกมีชีวิตชีวาและมีแรงผลักดันมากขึ้น อีกทั้งยังช่วยดึงดูดให้ผู้ชมรู้สึกอยากติดตามเนื้อหามากขึ้น

2. สีโทนเย็น (Cool Colors) เป็นสีแห่งความสุภาพและความอ่อนโยน ทำให้ผู้ชมรู้สึกผ่อนคลายและเพลิดเพลินมากขึ้น และยังสามารถใช้โน้มน้าวจากในระยะไกลได้อีกด้วย

3. สีโทนกลาง (Neutral Colors) สีเหล่านี้มักจะถูกนำไปผสมกับสีอื่น ๆ เพื่อให้เกิดสีที่เป็นกลางมากขึ้น และให้ความรู้สึกที่เป็นธรรมชาติ

สีสามารถสื่อถึงอารมณ์ความรู้สึกและสื่อความหมายของเว็บไซต์นั้น ๆ ได้อย่างชัดเจน ดังนั้นหากเลือกใช้สีไม่เหมาะสมกับเนื้อหาหรือจุดประสงค์ของเว็บ ก็จะทำให้เว็บดูไม่น่าสนใจ ขาดความน่าเชื่อถือและทำให้ผู้ที่เคยเข้ามาใช้บริการไม่คิดจะกลับมาใช้บริการอีก

2.2.3 ทฤษฎีเกี่ยวกับชุดคำสั่ง CSS

CSS ย่อมาจาก Cascading Style Sheet มักเรียกโดยย่อว่า "สไตล์ชีต" คือภาษาที่ใช้เป็นส่วนของการจัดรูปแบบการแสดงผลเอกสาร HTML โดยที่ CSS กำหนดกฎเกณฑ์ในการระบุรูปแบบ (หรือ "Style") ของเนื้อหาในเอกสาร อันได้แก่ สีของข้อความ สีพื้นหลัง ประเภทตัวอักษรและการจัดวางข้อความ ซึ่งการกำหนดรูปแบบ หรือ Style นี้ใช้หลักการของการแยกเนื้อหาเอกสาร HTML ออกจากคำสั่งที่ใช้ในการจัดรูปแบบการแสดงผล กำหนดให้รูปแบบของการแสดงผลเอกสาร ไม่ขึ้นอยู่กับเนื้อหาของเอกสาร เพื่อให้ง่ายต่อการจัดรูปแบบการแสดงผลลัพธ์ของเอกสาร HTML โดยเฉพาะในกรณีที่มีการเปลี่ยนแปลงเนื้อหาเอกสารบ่อยครั้ง หรือต้องการ

ควบคุมให้รูปแบบการแสดงผลเอกสาร HTML มีลักษณะของความสม่ำเสมอทั่วกันทุกหน้าเอกสารภายในเว็บไซต์เดียวกัน โดยกฎเกณฑ์ในการกำหนดรูปแบบ (Style) เอกสาร HTML ถูกเพิ่มเข้ามาครั้งแรกใน HTML 4.0 เมื่อปีพ.ศ. 2539 ในรูปแบบของ CSS level 1 Recommendations ที่กำหนดโดย องค์กร World Wide Web Consortium หรือ W3C

ไวยากรณ์ของ CSS

ไวยากรณ์ของ CSS นั้น มีโครงสร้างหลักของภาษาอยู่ 4 ส่วนด้วยกัน ซึ่งทั้งหมดมีความสำคัญต่อการกำหนดค่าและการใช้งาน CSS ซึ่งรูปแบบการกำหนดค่าจะประกอบไปด้วย property ตามด้วยเครื่องหมายโคลอน (:) เพื่อกำหนดค่าลงไปซึ่งอาจจะเป็นตัวเลข คีย์เวิร์ด หรือ URL และสุดท้ายของการกำหนดค่าคือ เซมิ-โคลอน (;) ที่ใช้ในการระบุว่าสิ้นสุดการกำหนดค่าให้กับproperty ดังตัวอย่างที่ 1

```
Selector {
Properties: Value;
}
```

ตัวอย่างที่ 1 ไวยากรณ์ของ CSS

ส่วนประกอบของโครงสร้างของ CSS มีดังนี้

1. Selector สามารถเป็น HTML Tag ต่าง ๆ ก็ได้เช่น <body>, <p> หรือเป็น Classname หรือ ID ที่ตั้งชื่อให้ก็ได้ซึ่งจะเรียนรู้ในหัวข้อถัดไป

2. เครื่องหมายปีกกาที่ครอบ Properties กับ Value ที่ต้องการกำหนดค่า

3. ชื่อของ Properties เป็นคุณสมบัติในการจัดรูปแบบการแสดงผล เช่น color สำหรับกำหนดสี, font-size สำหรับกำหนดขนาดตัวอักษร

4. ค่า Value ที่ต้องการจะกำหนดให้กับแต่ละ Properties เช่น color: black, fontsize:12px

ภาพที่ 2. 14 แสดงซอร์จโค้ดส่วนประกอบของโครงสร้างของ CSS

(ที่มา : https://bit.ly/2Z0dt0G)

จากซอร์จโค้ดข้างต้นนี้สามารถพิจารณาตามรูปแบบโครงสร้างของ CSS ได้ดังนี้คือ Selector ของ CSS คือ h1, Properties ของ CSS คือ color และ background, Value ของ CSS คือ red และ yellow

จะสังเกตเห็นได้ว่าจะมีเครื่องหมาย Colon (:) ที่เป็นตัวคั่นระหว่างชื่อ Properties กับ Value ที่เป็นการกำหนดค่าให้กับ Properties นั้น ๆ ซึ่งเมื่อสิ้นสุดแล้ว แต่ละ Properties จะจบด้วยเครื่องหมาย semi-colon (;) เสมอ ซึ่งเป็นข้อบังคับของการเขียน CSS

2.2.4 ทฤษฎีเกี่ยวกับการ Visualization

Visualization เป็นส่วนประกอบสำคัญใน Cognitive System ซึ่งเป็นส่วนในการแสดงข้อมูลหรือผลลัพธ์ต่าง ๆ ในระหว่างคอมพิวเตอร์และผู้ใช้งานในรูปแบบของภาพ โดยผู้ใช้สามารถเรียนรู้และจดจำข้อมูลผ่านการมองได้มากกว่าการใช้ประสาทสัมผัสอื่น ๆ หรือจะกล่าวว่า Visualization ก็คือ การสร้างมโนภาพของสิ่งต่าง ๆ ที่เราสนใจขึ้นมาในใจ ซึ่งต่อมาได้กลายเป็นการนำภาพมาใช้กับการนำเสนอหรือนำมาเป็นกรอบความคิด ซึ่งได้นำไปใช้ในการสนับสนุนการตัดสินใจ

Visualization เป็นมากกว่าวิธีการทาง Computer Visualization เป็นการนำเสนอข้อมูลแบบหนึ่งที่ทำการแสดงข้อมูลให้อยู่ในรูปแบบของ Visual Form ซึ่งอาจจะเป็นการแสดงในรูปแบบของรูปภาพ ,กราฟ หรือ แผนภาพ ซึ่ง ผลลัพธ์ของการทำ Visualization คือ การนำข้อมูลต่าง ๆ ที่ซ่อนอยู่ในตัวของข้อมูลเองออกมาให้ผู้ใช้สามารถสัมผัสได้ ซึ่งกระบวนการต่าง ๆ ในการแสดงหรือนำข้อมูลต่าง ๆ ออกมา อาจจะถูกแอบซ่อนอยู่หรือผู้ใช้ไม่ทันสังเกต แต่อย่างไรก็ตาม Visualization เป็นสิ่งที่จำเป็นในการค้นหาข้อมูล หรือ ใช้ในการวิเคราะห์ข้อมูลโดยจุดมุ่งหมายของ Visualization

ก็คือ การถ่ายทอดข้อมูลไปสู่ระบบการรับรู้โดยภาพของผู้ใช้ระบบ เพื่อช่วยในการลดช่องว่างระหว่างผู้ใช้และตัวข้อมูล และยังช่วยผู้ใช้สามารถเรียนรู้ ได้อย่างมีประสิทธิภาพ

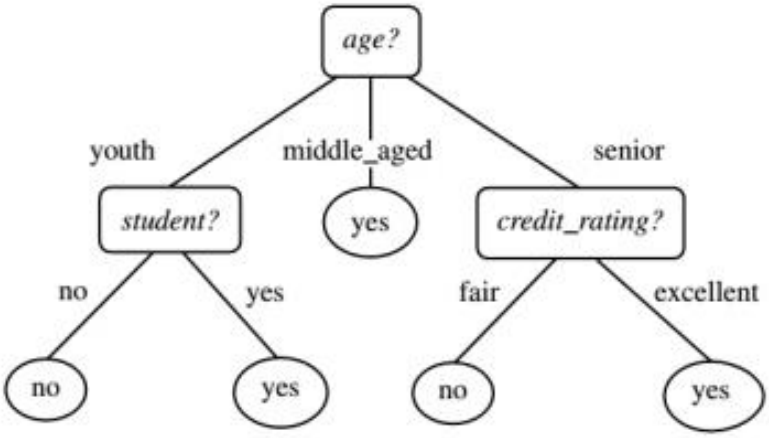
1) วิธีการของ Visualization

การใช้ Visualization เพื่อทำการค้นหาข้อมูลนั้นผู้ใช้จะทำขั้นตอนหลัก ๆ อยู่ 3 ขั้นตอน คือOverview First, Zoom and Filter และ Detail on Demand โดยอันดับแรก ผู้ใช้ต้องการที่จะดูข้อมูลภาพรวมทั้งหมดซึ่งหลังจากดูภาพรวมทั้งหมดแล้วผู้ใช้ก็จะทำการตัดสินใจเลือกรูปแบบหรือกลุ่มข้อมูลที่สนใจซึ่งก็จะมาถึงขั้นตอนในการวิเคราะห์ข้อมูลผู้ใช้ก็จะทำการเจาะลึกถึงข้อมูลในรายละเอียด ซึ่ง Visualization Technology ก็จะอ้างอิงหรือพัฒนาจากขั้นตอนเหล่านี้ซึ่งVisualization Technique มีประโยชน์มากในการแสดงภาพรวมหรือแสดงข้อมูลย่อยที่ผู้ใช้ต้องการโดยอาจจะใช้หลายๆวิธีการรวมกันเพื่อให้ได้ข้อมูลที่ผู้ใช้ต้องการ ซึ่งช่วยลดช่องว่างของกิจกรรมที่ใช้ในการดึงข้อมูลต่าง ๆ ไปใช้ ซึ่งลักษณะของข้อมูลที่สามารถนำมาผ่านกระบวนการของVisualization มีลักษณะต่าง ๆ มากมายดังนี้ ข้อมูล 1D ได้แก่ ข้อมูลที่เกี่ยวข้องกับช่วงเวลาใดเวลาหนึ่ง,ข้อมูล 2D ได้แก่ ข้อมูลที่เกี่ยวข้องกับแผนที่ภูมิศาสตร์, Multi Dimensional Data ได้แก่Relation Table, Text และ Hypertext ได้แก่ ข้อมูลหัวข้อข่าวต่าง ๆ และ Web Document,Hierarchies และ Graph ได้แก่ หมายเลขโทรศัพท์ และ Web Document ,Algorithms และ Softwareได้แก่ Debugging Operation ซึ่งแต่ละข้อมูลก็จะมีวิธีการที่ช่วยในการจัดการแสดงผลข้อมูลที่หลากหลาย

2.2.5 ทฤษฎีเกี่ยวกับการจำแนกข้อมูลด้วยต้นไม้ตัดสินใจ

การจำแนกข้อมูลด้วยต้นไม้ตัดสินใจจะเป็นกระบวนการสร้างต้นไม้ขึ้นเพื่อใช้ในการตัดสินใจจากข้อมูลที่มีหมวดหมู่ข้อมูลแนบอยู่ด้วย ต้นไม้ตัดสินใจจะประกอบไปด้วยโหนดต่าง ๆ (ที่ไม่ใช่โหนดใบ—nonleaf node) ที่ซึ่งถูกใช้ในการแสดงถึงเงื่อนไขหรือแอทริบิวหนึ่ง ๆ ของข้อมูล โดยที่แต่ละกิ่งก้านของโหนดหนึ่ง ๆ จะหมายถึงค่าที่เป็นไปได้จากการทดสอบกับแอทริบิวนั้น ๆ และจะประกอบไปด้วยโหนดใบ (leaf node) ที่ซึ่งจะมีหมวดหมู่ข้อมูลจัดเก็บอยู่ โดยตัวอย่างต้นไม้ตัดสินใจจะถูกแสดงในรูปที่ 2.14 ที่จะแสดงการทำนายคุณลักษณะของลูกค้าที่จะทำการซื้อคอมพิวเตอร์จากร้านขายอุปกรณ์ไฟฟ้า โดยโหนดต่าง ๆ ที่ไม่ใช่โหนดใบจะถูกแทนด้วยสี่เหลี่ยมและโหนดใบจะถูกแทนด้วยวงรีตามลำดับ จากรูปเราจะเห็นว่าโหนดใบจะเป็นโหนดที่ บ่งบอกถึงข้อมูลหมวดหมู่ของคำตอบที่เราต้องการ อาทิ เช่น “yes” หมายถึง ลูกค้าจะซื้อคอมพิวเตอร์ และ“no” หมายถึงลูกค้าจะไม่ซื้อคอมพิวเตอร์โดยต้นไม้ที่ถูกสร้างขึ้นอาจเป็นต้นไม้ที่

มีลักษณะเป็นไบนารี่หรืออาจจะไม่เป็นไบนารี่ก็ได้ หลังจากทำการสร้างต้นไม้ตัดสินใจแล้ว เราจะสามารถใช้ต้นไม้ตัดสินใจในการจำแนกข้อมูลได้โดยจะทำการจำแนกหมวดหมู่ของข้อมูลเรคคอร์ดหนึ่ง ๆ (ที่ประกอบไปด้วยแอทริบิวต่าง ๆ แต่เราจะไม่ทราบหมวดหมู่ข้อมูลในเรคคอร์ดนั้น ๆ) ด้วยการเปรียบเทียบแอทริบิวที่อยู่ในโหนดรากกับค่าของแอทริบิวในเรคคอร์ดที่พิจารณา โดยจะทำการเปรียบเทียบจากโหนดรากไปจนถึงโหนดใบ เมื่อเราทราบถึงโหนดใบจะทำให้เราทราบถึงหมวดหมู่ข้อมูลของเรคคอร์ดที่ทำการพิจารณา



ภาพที่ 2. 15 แสดงตัวอย่างต้นไม้ตัดสินใจสำหรับการจำแนกคุณลักษณะของลูกค้าที่ทำการซื้อคอมพิวเตอร์

(ที่มา : https://bit.ly/38r2EaK)

การจำแนกหมวดหมู่ข้อมูลด้วยต้นไม้ตัดสินใจได้รับความนิยมเป็นอย่างมาก และถูกประยุกต์ใช้ในหลาย ๆ งาน อาทิ เช่น การผลิตและการใช้ยา (medicine) การผลิตสินค้าต่าง ๆ (manufacturing and production) การวิเคราะห์ทางการเงิน (Financial analysis) ดาราศาสตร์ (astronomy) อณูชีววิทยา (molecular biology) เป็นต้น สาเหตุที่ต้นไม้ตัดสินใจได้รับความนิยมอันเนื่องมากจากเหตุผลหลายประการ ด้วยกัน เช่น 1) ไม่ต้องการองค์ความรู้ใด ๆหรือการกำหนดค่าพารามิเตอร์ใด ๆ เพื่อที่จะทำการสร้างต้นไม้ตัดสินใจ 2) สามารถจัดการกับข้อมูลที่มีหลายๆมิติหรือข้อมูลที่มีหลายๆแอทริบิวได้ 3) ผลลัพธ์ที่ได้จะเป็นต้นไม้ตัดสินใจที่อยู่ในรูปแบบที่เข้าใจง่าย 4) ขั้นตอนการสร้างต้นไม้ตัดสินใจค่อนข้างง่ายและสามารถทำงานได้อย่างรวดเร็ว และ 5) มักจะให้ผลการจำแนกข้อมูลที่มีความถูกต้องค่อนข้างสูง ที่ซึ่งอาจขึ้นอยู่กับคุณลักษณะ ของข้อมูลที่เราทำการพิจารณาด้วยเช่นกัน

2.3 เครื่องมือในการออกแบบและวิเคราะห์ข้อมูล

ในโครงงานนี้ใช้เครื่องมือสำหรับทำกระบวนการเหมืองข้อมูลที่มีชื่อว่า Weka โดย Weka เป็นโปรแกรมที่มีเทคนิคต่าง ๆ มากมายพัฒนาอยู่ภายในเช่น อัลกอริทึมการจัดหมวดหมู่:K-NN , LR , C4.5 , Support Vector Machine , การคัดเลือกคุณสมบัติ: Information Gain ,Gain Ratio เป็นต้นซึ่ง Weka เป็นที่นิยมกันอย่างแพร่หลายในกลุ่มผู้จัดทำวิจัยที่วิจัยเกี่ยวกับเหมืองข้อมูล Weka ไม่ได้มีเพียงตัวโปรแกรมที่สามารถใช้งานได้ด้วยตัวเอง (Standalone) เพียงอย่างเดียวแต่ยังมี Library สำหรับใช้คู่กับการเขียน โปรแกรมในภาษาจาวา (Java Programming) อีกด้วยซึ่ง สะดวกมากสำหรับ ผู้วิจัยที่มีความจำเป็นจะต้องเขียนโปรแกรมเกี่ยวกับการทำเหมืองข้อมูล เช่น การคัดเลือกคุณสมบัติ, การหากฎความสัมพันธ์ และผู้วิเคราะห์ยังได้ใช้การVisualization ผ่านโปรแกรมTableau Public ในการนำเสนอข้อมูลผ่านเว็บไซต์อีกด้วย ทาโบล์ว (Tableau) Business Intelligence (BI) Software คือซอฟต์แวร์เพื่อการวิเคราะห์ข้อมูลอันหลากหลายเป็น Software ในระดับต้น ๆ ของโลกที่ทำเรื่อง Data Visualization Tableau ทำหน้าที่แปลงข้อมูล (Data) ให้เป็นภาพ (Visualization) สามารถนำข้อมูลออกมาเป็น Chart แบบต่าง ๆ ได้อย่างง่ายดายรวมถึงการนำหลาย Chart มาสร้างเป็น Dashboard และ Story Teller ก็ได้เช่นกัน รองรับการเข้าถึงข้อมูลจากหลายฐานข้อมูล และ สามารถนำข้อมูลจากฐานข้อมูลในองค์กรนั้นมาใช้งาน เช่น Excel, Access, Firebird 2.0, IBM DB2, MS SQL, Oracle, MySQL เป็นต้น

2.4 วรรณกรรมที่เกี่ยวข้อง

วีณาวดี ม่วงอ้น และคณะ (2558) จากการวิจัยเรื่องระบบผู้เชี่ยวชำญสำหรับระบุพันธุ์เห็ดในประเทศไทย ผลการศึกษาพบว่า งานวิจัยนี้ได้นำเสนอระบบผู้เชี่ยวชาญ ที่มีความสามารถในการวิเคราะห์จำแนกพันธุ์เห็ด ในประเทศไทย ได้เหมาะสมตรงกับปัญหาของผู้ใช้ ระบบได้ทำการแทนค่าความรู้ในรูปแบบของกฎ ที่ถูกสร้างจากต้นไม้ตัดสินใจอัลกอริทึม ID3 และเก็บไว้ในฐานความรู้ ซึ่งเชื่อมโยงสัมพันธ์กับกฎต่าง ๆ ที่เกี่ยวข้องกับรูปร่างลักษณะพันธุ์เห็ด โดยระบบใช้วิธีการกลไกการอนุมานแบบเดินหน้าในการหาเหตุผลเพื่อค้นหาคำตอบเป้าหมายของผู้ใช้ และใช้ค่าความเชื่อมั่นในการคำนวณค่าความแตกต่างของความเชื่อมั่นในแต่ละคำตอบของผู้ใช้แล้วนำค่าความเชื่อมั่นนั้นแทนค่าออกมาเป็นร้อยละของความเป็นไปได้ที่จะเป็นสายพันธุ์เห็ดนั้นเพื่อใช้สำหรับเรียงลำดับผลลัพธ์ให้กับผู้ใช้ ระบบผู้เชี่ยวชาญระบุพันธุ์เห็ดนี้พัฒนาให้สามารถทำงานได้บนเว็บบราวเซอร์ โดยพัฒนาหน้าเว็บส่วนติดต่อผู้ใช้ด้วย ภาษา PHP HTML และ JAVASCRIPT ใช้

MySQL เป็นเครื่องมือช่วยในการจัดการฐานข้อมูลและใช้ภาษา Prolog ในการจัดเก็บฐานความรู้ของสายพันธุ์เห็ด ผลการทดสอบแสดงว่าระบบนี้สามารถให้คำตอบที่ถูกต้องอยู่ในเกณฑ์ดี

วทัญญุตา นีลาภาตระกูล และ ชุติมา เบี้ยวไข่มุข(2562) จากการวิจัยเรื่อง การศึกษาปัจจัยที่สัมพันธ์กับการตัดสินใจลาออกและการเปรียบเทียบประสิทธิภาพ ผลการศึกษาพบว่า งานวิจัยนี้การศึกษาเกี่ยวกับเทคนิคการทำ เหมืองข้อมูล เพื่อลดปัญหาอัตราการลาออกของพนักงานและลดค่าใช้จ่ายที่สูงขึ้นจากการสรรหาพนักงานใหม่ในบริษัทประกันภัยแห่งหนึ่ง ชุดข้อมูลที่ใช้วิเคราะห์เป็นข้อมูลของพนักงานลาออกและพนักงานที่ยังทำงานอยู่ ตั้งแต่ปีพ.ศ. 2556-2560 จำนวน 1,000 รายการ11 แอตทริบิวต์ เทคนิคกฎความสัมพันธ์ถูกนำ มาใช้หารูปแบบของความสัมพันธ์การลาออกของพนักงาน และสร้างตัวแบบพยากรณ์การลาออกของพนักงาน โดยใช้เทคนิคการวิเคราะห์5เทคนิคคือต้นไม้ตัดสินใจเทคนิคซัพพอร์ตเวกเตอร์แมชชีน เทคนิคโครงข่ายประสาทเทียม เทคนิคเบย์เซียนแบบง่ายและเทคนิคเพื่อนบ้านใกล้สุดและเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์เหล่านั้น การทดสอบประสิทธิภาพของตัวแบบใช้วิธีไขว้ข้อมูลโดยแบ่งข้อมูลออกเป็น 5 ชุดเท่ากัน ผลของกฎความสัมพันธ์พบปัจจัยที่มีความสัมพันธ์กับการตัดสินใจลาออก คือ เงินเดือน อายุงาน ความก้าวหน้าในอาชีพ ผลการประเมินการปฏิบัติงาน และความสัมพันธ์กับหัวหน้างาน ผลการทดสอบตัวแบบพยากรณ์ที่เป็นต้นไม้ตัดสินใจ ให้ค่าความถูกต้องในการพยากรณ์สูงที่สุดที่ร้อยละ91.03ในขณะที่ตัวแบบพยากรณ์ที่สร้างจากเทคนิคซัพพอร์ตเวกเตอร์แมชชีน ให้ค่าความถูกต้องเท่ากับร้อยละ 90.93 เทคนิคโครงข่ายประสาทเทียม ให้ค่าความถูกต้องเท่ากับ ร้อยละ 90.75 เทคนิคเบย์เซียนแบบง่ายและเทคนิคเพื่อนบ้านใกล้สุด ให้ค่าความถูกต้องเท่ากับ ร้อยละ 89.60 และ 82.10 ตามลำดับ ผลที่ได้จากการวิจัยสามารถนำผลมาประยุกต์ใช้ในการสร้างแผนกลยุทธ์สำ หรับออกแบบสวัสดิการที่เหมาะสมเพื่อรักษาพนักงานให้คงอยู่กับบริษัทและเป็นแนวทางในการสร้างระบบสารสนเทศในการสนับสนุนการตัดสินใจสำหรับงานด้านบริหารทรัพยากรบุคคลต่อไป

พีรพล คำพันธ์และคณะ(2562) จากการวิจัยเรื่องการประยุกต์ใช้เบย์เซียนและการประมวลผลภาพสำหรับการคัดแยกมะม่วง ผลการศึกษาพบว่า งานวิจัยนี้นำเสนอการประยุกต์เบย์เซียนและการประมวลผลภาพดิจิตอลเพื่อแยกประเภทมะม่วงตามลักษณะความสุก/ดิบ ด้วยวิธีการคัดแยกเชิงคุณลักษณะ (Feature-based) โดยที่คุณลักษณะที่ใช้สำหรับการแยกประเภทได้มาจากคุณลักษณะของภาพสี RGB และความเข้มข้นของค่าสีในเพลนสีแดงร่วมกับการวิเคราะห์ค่าทางสถิติ ในขั้นตอนการแยกประเภทผลมะม่วงได้ทำการทดลองเปรียบเทียบประสิทธิภาพ

ระหว่างการแยกประเภทด้วยเบย์เซียนและการแยกประเภทด้วยอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน (SVM) จากผลการทดลองพบว่าการแยกประเภทด้วยเบย์เซียนให้ผลลัพธ์ความถูกต้องแม่นยำร้อยละ 88 มีความถูกต้องแม่นยำมากกว่าการแยกประเภทด้วยอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีนซึ่งให้ผลลัพธ์ความถูกต้องแม่นยำร้อยละ 83

พิชัย ระเวงวังและพุธษดี ศิริแสงตระกูล(2560) จากการวิจัยเรื่องโมเดลเพื่อการพยากรณ์สถานภาพทางการศึกษาของนักศึกษา ผลการศึกพบว่า งานวิจัยนี้มุ่งเน้นศึกษาปัจจัยที่มีผลต่อสถานภาพการศึกษาของนักศึกษา เพื่อทำการสร้างโมเดลเพื่อพยากรณ์สถานภาพทางการศึกษา ของนักศึกษาในสถาบันอุดมศึกษาเอกชนที่ตั้งในภาคตะวันออกเฉียงเหนือ โดยการประยุกต์ใช้วิธีการเหมืองข้อมูล ในการศึกษาผู้วิจัยได้ทำการเก็บรวบรวมข้อมูลปัจจัยที่คาดวาจะมีผลต่อสถานภาพทางการศึกษาจากงานวิจัยที่เกี่ยวข้องและได้รวมรวมข้อมูลของนักศึกษาระดับปริญญาตรีที่เข้าศึกษาระหวางปี การศึกษา 2551–2553 จำนวนทั้งสิ้น 2,272 คน และวิเคราะห์ความสัมพันธ์ของปัจจัยที่มีผลต่อสถานภาพของนักศึกษาด้วยวิธีการวิเคราะห์สหสัมพันธ์ ผลจากศึกษาพบวาปัจจัยที่มีผลต่อสถานภาพของนักศึกษามี9 ปัจจัย ได้แก่การกู้ยืมในระหว่างเรียน รายได้ ผู้ปกครอง สถานภาพสมรสของผู้ปกครอง เพศ ประเภทโรงเรียนเดิมที่จบมา วุฒิการศึกษาเดิม กลุ่มโรงเรียนเดิมที่จบ อาชีพของมารดา และที่อยู่นักศึกษาระหว่างเรียน ปัจจัยดังกล่าวจะถูกนำมาใช้ในการสร้างโมเดลเพื่อทำนายสถานภาพของนักศึกษา ด้วยเทคนิคต้นไม้ตัดสินใจ เบย์เซี่ยน บีลิฟเน็ตเวิร์ค และโลจิสติกรีเกรสชัน ผลการศึกษาพบว่า ค่าความถูกต้องมีค่า 82.85%, 78.98% และ 78.50% ด้วยเทคนิคต้นไม้ตัดสินใจ จิสติกรีเกรสชัน และ การเรียนรู้แบบเบย์เซี่ยน เน็ตเวิร์ค ตามลำดับ ซึ่งพบว่าโมเดลที่สร้างด้วยวิธีต้นไม้ตัดสินใจให้ค่าความถูกต้องดีที่สุด

พัฒนพงษ์ ดลรัตน์ และ จารี ทองคำ(2560) จากการวิจัยเรื่องการเปรียบเทียบประสิทธิภาพของแบบจำลองในการพยากรณ์ความสำเร็จการศึกษาของนักเรียนระดับประกาศนียบัตรวิชาชีพ ผลการศึกษาปัจจุบันประเทศไทยนั้นอยู่สภาวะขาดแคลนแรงงานฝีมือ กระทรวงศึกษาธิการจึงได้มีนโยบายในปีพุทธศักราช 2554 การเปลี่ยนแปลงสัดส่วนของผู้เรียนอาชีวศึกษาต่อผู้เรียนสายสามัญจากเดิม 40:60 เป็น 60:40 ภายในปีพุทธศักราช 2561 ซึ่งเป็นปีสิ้นสุดของการปฏิรูปการศึกษาทศวรรษที่ 2 ซึ่งสถานศึกษาอาชีวศึกษาเอกชนก็เป็นหน่วยงานหนึ่งที่ดำเนินการสนองนโยบายของรัฐบาลเพื่อแก้ไขปัญหาสภาวะขาดแคลนแรงงานฝีมือ เพิ่มจำนวนผู้สำเร็จการศึกษา พัฒนาการศึกษาให้เกิดประสิทธิภาพสูงสุด ดังนั้นงานวิจัยนี้จึงมีจุดประสงค์เพื่อการเปรียบเทียบประสิทธิภาพของแบบจำลองในการพยากรณ์ความสำเร็จการศึกษาของนักเรียน

ระดับประกาศนียบัตรวิชาชีพ งานวิจัยนี้ได้ใช้ 6 เทคนิคที่มีประสิทธิภาพในการสร้างแบบจำลอง คือ C4.5, Random Forest, Random Tree, Reduced Error Pruning (REP Tree), k–Nearest Neighbors (k–NN) และSupport Vector Machine(SVM) และวัดประสิทธิภาพการพยากรณ์ของแบบจำลองด้วยค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) และค่าความระลึก (Recall) จากการศึกษาพบว่า แบบจำลอง C4.5 มีประสิทธิภาพในการพยากรณ์ความสำเร็จการศึกษาของนักเรียนระดับประกาศนียบัตรวิชาชีพมากที่สุดถึง 95.36%

พฤฒิพงศ์ เพ็งศิริ (2559)จากการวิจัยเรื่อง การพยากรณ์ความสัมพันธ์ของปัจจัยข้อมูลนักศึกษาที่มีผลต่อระดับผลการเรียนด้วยเทคนิคต้นไม้ตัดสินใจ กรณีศึกษานักศึกษามหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ ผลการศึกษาพบว่า การพยากรณ์โดยการประยุกต์ใช้กระบวนการตัดสินใจด้วยเทคนิคต้นไม้ตัดสินใจเป็นเทคนิคอย่างหนึ่งในการทำเหมืองข้อมูล ซึ่งอาศัยความสัมพันธ์ของปัจจัยข้อมูลนักศึกษาเป็นการบ่งชี้ถึงระดับผลการเรียนของนักศึกษามหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ ซึ่งผลลัพธ์ที่ได้นั้น จะส่งผลทำให้เอือ้ ประโยชน์กับแนวทางในการปรับปรุงบริหารจัดการและส่งเสริมการวางแผนของหลักสูตรการศึกษาที่เกี่ยวข้อง โดยนำปัจจัยข้อมูลของนักศึกษา ได้แก่ข้อมูลทั่วไป ข้อมูลส่วนตัว ข้อมูลการศึกษา ข้อมูลระดับผลการศึกษา ของนักศึกษากลุ่มที่ศึกษา มาหาความสัมพันธ์กับผลการศึกษาที่ได้ โดยวิเคราะห์จากระดับผลการเรียน (เกรดเฉลี่ย) ของการเรียนในชั้นปีที่ 1 ผลการทดลองพบว่าปัจจัยข้อมูลของนักศึกษาที่เหมาะสมในการเรียน มีทั้งหมด 7 ตัวแปร จากข้อมูลนำเข้าทั้งหมด 12 ตัวแปร ทั้งนี้ 7 ตัวแปรมาจากต้นไม้ตัดสินใจที่ได้มาสามารถสรุปเลือกเฉพาะกิ่งที่มีผลมากที่สุด โดยวัดค่าความแม่นยำ (Accuracy) ได้ค่าสูงถึง 84.78% ซึ่งถือว่าอยู่ในเกณฑ์สูง ดังนั้นตัวแปรที่น่าจะเป็นปัจจัยที่ส่งผลกับ ระดับผลการเรียนที่จบการศึกษาสูงสุดคือ ความสม่ำเสมอการเข้าเรียนในการเรียน

หนูเดือน เมืองแสนและคณะ(2556) จากการวิจัยเรื่องความหลากหลายของไลเคน เห็ดและราขนาดใหญ่ในพื้นที่ปกปักพันธุกรรมพืช อพ.สธ. เขื่อนน้ำพุง การไฟฟ้าฝ่ายผลิตแห่งประเทศไทย ผลการศึกษาพบว่า เห็ดกินได้ (edible mushroom) เป็นเห็ดที่รับประทานได้ ลักษณะ ส่วนใหญ่เจริญในทุ่งหญ้า ก้านสั้น อ้วนปูอมและไม่โป่งพองออก ผิวเรียบไม่ขรุขระ ไม่มีสะเก็ด สีผิวของหมวกส่วนใหญ่เป็นสีขาวถึงสี น้ำตาล ผิวของหมวกเห็ดเรียบจนถึงเป็นเส้นใยและ เหมือนถูกกดจนเป็นแผ่นบาง ๆ ดึงออกยาก ครีบแยกออก จากกัน ในระยะแรกเป็นสีชมพู แล้วเปลี่ยนเป็นสีน้ำตาล สปอร์สีน้ำตาลอมม่วงแก่รูปกระสวยกว้าง เช่น เห็ด ฟาง เห็ดหูหนู เห็ดหอม เห็ดเผาะ เห็ดขมิ้น ฯลฯ เห็ดพิษ (toadstool หรือ poisonous mushroom) เห็ดที่นำมาบริโภคแล้วเป็นอันตรายต่อ

ผู้บริโภคอาจร้ายแรงถึงตายได้ ลักษณะส่วนใหญ่เจริญงอกงามในป่า ก้านสูง ลำต้นโป่งพองออก โดยเฉพาะที่ ฐาน กับที่วงแหวนเห็นชัดเจน สีผิวของหมวกมีได้หลายสี เช่น สีมะนาว ถึงสีส้ม สีขาว ถึงสีเหลือง ผิวของหมวก เห็ดส่วนมากมีเยื่อหุ้มดอกเห็ดเหลืออยู่ในลักษณะที่ดึงออกได้ หรือเป็นสะเก็ดติดอยู่ ครีบแยกออกจากกัน ชัดเจน มักมีสีขาว บางชนิดสีแดงหรือสีเขียวอมเหลือง สปอร์ใหญ่มีสีขาวหรือสีอ่อน มีลักษณะใส ๆ รูปไข่กว้าง ซึ่งโดยทั่วไปมักมีลักษณะแปลก ๆ เช่น เห็ดหัวกรวด เห็ดขี้ควาย ฯลฯ

พรรณพรและคณะ (2554) จากการวิจัยเรื่อง การสำรวจความหลากหลายชนิดและคุณลักษณะพิเศษเฉพาะตัวของเห็ดป่ารับประทานได้เพื่อการจำแนกแยกเห็ดพิษ และภูมิปัญญาการใช้ประโยชน์ในจังหวัดน่าน ผลการศึกษาพบว่า เมื่อเปรียบเทียบลักษณะสัณฐานของเห็ดระโงกหิน (เห็ดพิษ) กับเห็ดไข่ขาว (เห็ดรับประทานได้) พบว่าในระยะอ่อน เห็ดระโงกหินจะมีลักษณะเนื้อแน่น เหนียวคล้ายยางพารา ก้านและโคนตัน เมื่อลอกเปลือกหุ้มออกจะเห็ดหมวกสีขาวสว่าง ขอบหมวกเรียบ ในระยะโตเต็มวัย(ดอกบาน) เห็ดระโงกหินจะมีลักษณะ เนื้อแข็ง หมวกสีขาวสว่าง ขอบหมวกมีริ้วคล้ายชายกระโปรง ก้าน และโคนตัน เปลือกหุ้มโคนแนบชิดติดก้าน

2.5 บทสรุป

จากแนวคิด ทฤษฎี เครื่องมือและวรรณกรรมที่เกี่ยวข้องที่ได้กล่าวมาในข้างต้นทั้งหมดนั้น ผู้วิเคราะห์ข้อมูลได้เลือกใช้ขั้นตอนกระบวนการวิเคราะห์ข้อมูลด้วย CRISP-DM จากเทคนิคการทำเหมืองข้อมูลแบบ Classification ด้วยการสร้างโมเดล Decision Tree เพื่อจัดกลุ่มข้อมูลที่มีความสัมพันธ์กัน โดยใช้โปรแกรมที่ใช้ทำเหมืองข้อมูล คือ, โปรแกรม Weka 3.8.4 และโปรแกรม RapidMiner Studio 9.5.1 ในการสร้างโมเดล Decision Tree เพื่อทำการเปรียบเทียบ และเลือกผลลัพธ์ที่ดีที่สุดมานำเสนอ จากนั้นนำข้อมูลสารสนเทศมาทำการแสดงผลแบบ Visualization ในรูปแบบของภาพโดยใช้โปรแกรม Tableau Public เผยแพร่บน Web Browser ที่เป็นที่นิยมในยุคอินเตอร์เน็ตคือการเผยแพร่ทางสื่อออนไลน์ โดยใช้ภาษา HTML และ CSS ในการเขียนเว็บไซต์ขึ้นมา

บทที่ 3
วิธีการดำเนินงานโครงงาน

โครงงานเรื่อง การวิเคราะห์ข้อมูลการแยกประเภทของเห็ดในอเมริกาเหนือ เพื่อเผยแพร่ข้อมูลบนเว็บไซต์ ในบทนี้จะเป็นการวิเคราะห์ข้อมูลด้วยเทคนิคทางดาต้าไมนิ่ง ซึ่งมีกระบวนการวิเคราะห์ที่สำคัญหลายขั้นตอน เมื่อเสร็จสิ้นจากกระบวนการวิเคราะห์ข้อมูลแล้วจะเป็นการออกแบบเว็บไซต์ และออกแบบรูปแบบการแสดงผลและบทสรุปจากวิธีการดำเนินงาน

3.1 การวิเคราะห์ข้อมูลด้วย CRISP-DM

3.2 การออกแบบเว็บไซต์

3.3 บทสรุป

3.1 การวิเคราะห์ข้อมูลด้วย CRISP-DM

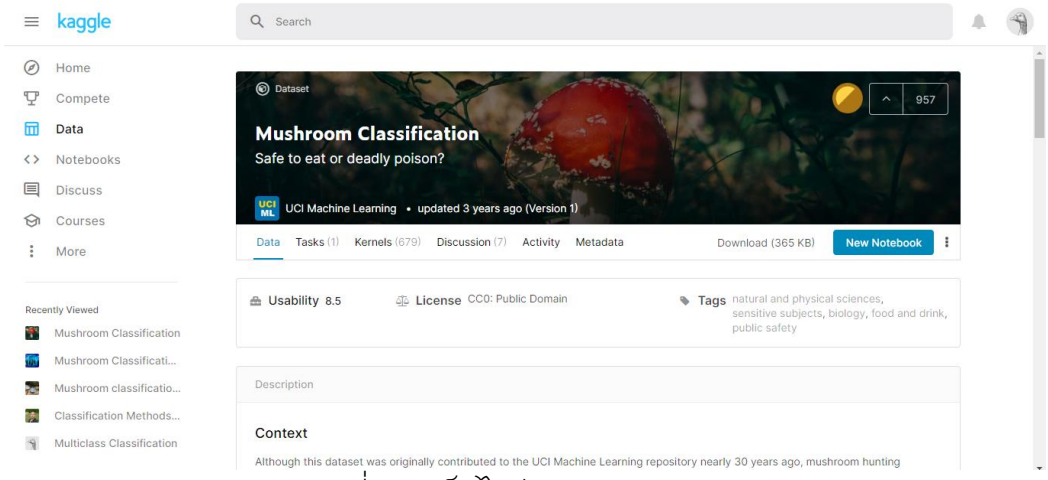
กระบวนการวิเคราะห์ข้อมูลด้วย CRISP-DM หรือ Cross Industry Standard Process for Data Mining พัฒนาขึ้นในปี ค.ศ. 1996 โดยความร่วมมือของ 3 บริษัทคือ Daimler Chrysler, SPSS และ NCR ที่มีการพัฒนาเป็น Workflow มาตรฐานสำหรับการทำเหมืองข้อมูล ประกอบด้วย 6 ขั้นตอนหลัก ดังนี้

3.1.1 รู้จักและเข้าใจในธุรกิจ (Business Understanding) เป็นขั้นตอนแรกของกระบวนการที่มุ่งเน้นไปที่การทำความเข้าใจกระบวนการทางธุรกิจโดยรวม

ผู้วิเคราะห์ข้อมูลทำความเข้าใจกับปัญหาให้อยู่ในรูปของการวิเคราะห์ข้อมูลทาง Data Mining โดยการวิเคราะห์ข้อมูลในประเด็นนี้ คือ การวิเคราะห์ข้อมูลการจำแนกเห็ดพิษหรือเห็ดที่รับประทานได้ในอเมริกาเหนือ ซึ่งมีข้อมูลมาก จำนวนรายการทั้งหมด 8,125 รายการ ทำให้ไม่สามารถทำความเข้าใจกับข้อมูลที่มีความซับซ้อนได้อย่างรวดเร็ว เช่น ต้องการทราบว่าลักษณะคุณสมบัติพื้นผิวเห็ดแบบใดที่บ่งบอกถึงเห็ดพิษมากที่สุด

3.1.2 จัดเก็บและรวบรวมข้อมูลให้ครบ (Data Understanding) ขั้นตอนการจัดเก็บและรวบรวมข้อมูล ตลอดจนการพิจารณาตรวจสอบความถูกต้องของข้อมูลที่ได้รับ โดยเลือกว่าจะใช้ข้อมูลทั้งหมดหรือบางส่วนในการวิเคราะห์ให้สอดรับกับวัตถุประสงค์ที่กำหนดไว้

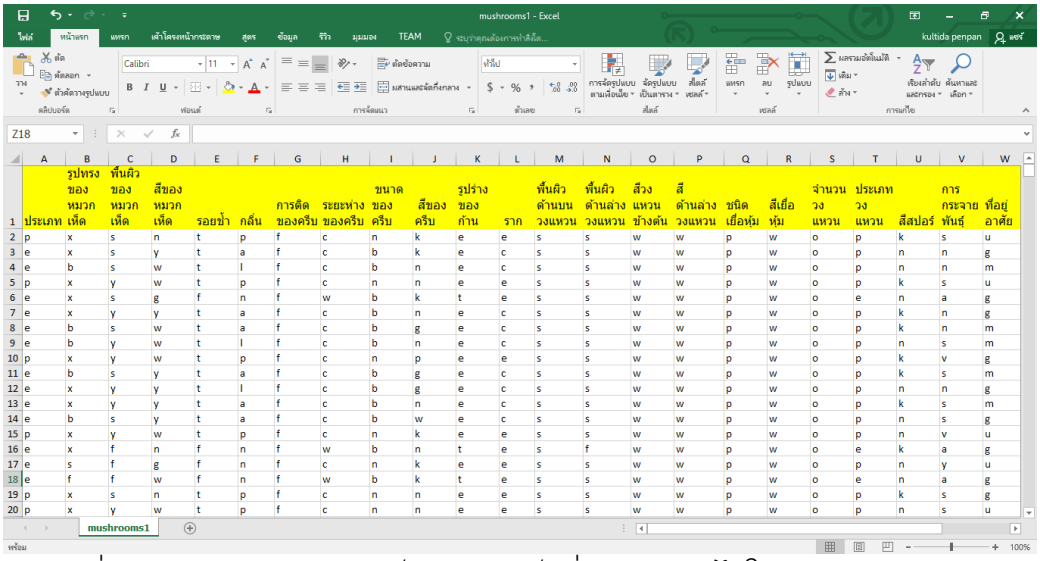
ผู้วิเคราะห์ข้อมูลทำการรวบรวมข้อมูล เพื่อตรวจสอบรายละเอียด ปริมาณ และ ความน่าเชื่อถือของข้อมูลการจำแนกเห็ดพิษหรือเห็ดที่รับประทานได้ในอเมริกาเหนือ ที่ได้จากเว็บไซต์ kaggle.com ซึ่งเป็นเว็บไซต์ที่เก็บรวบรวมชุดข้อมูลต่าง ๆ เป็นแหล่งรวม Datasets หรือ ชุดข้อมูลสำหรับฝึกสอน Machine Learning ที่ใหญ่ที่สุดในโลกแห่งหนึ่ง มีข้อมูลทุกประเภท ไม่ว่าจะเป็น Datasets ในหมวดหมู่ Finance, Business, Physics, Biology, Sports, News ซึ่งเป็นข้อมูลที่เปิดเผยได้ เพื่อให้ผู้ใช้บริการสามารถนำชุดข้อมูลไปศึกษาหรือวิเคราะห์ให้เกิดประโยชน์ต่อไป



ภาพที่ 3. 1 เว็บไซต์ kaggle.com

(ที่มา : https:www.kaggle.com)

ซึ่งข้อมูลการจำแนกเห็ดพิษหรือเห็ดที่รับประทานได้ในอเมริกาเหนือ มีจำนวนข้อมูล 8,125 รายการ ประกอบด้วย 23 แอตทริบิวท์ ประกอบด้วย ประเภท รูปทรงของหมวกเห็ด พื้นผิวของหมวกเห็ด สีของหมวกเห็ด รอยช้ำ กลิ่น การติดของครีบ ระยะห่างของครีบ ขนาดของครีบ สีของครีบ รูปร่างของก้าน ราก พื้นผิวด้านบนวงแหวน พื้นผิวด้านล่างวงแหวน สีวงแหวนข้างต้น สีด้านล่างวงแหวน ชนิดเยื่อหุ้ม สีเยื่อหุ้ม จำนวนวงแหวน ประเภทวงแหวน สีสปอร์ จำนวน การกระจายพันธุ์ ที่อยู่อาศัย



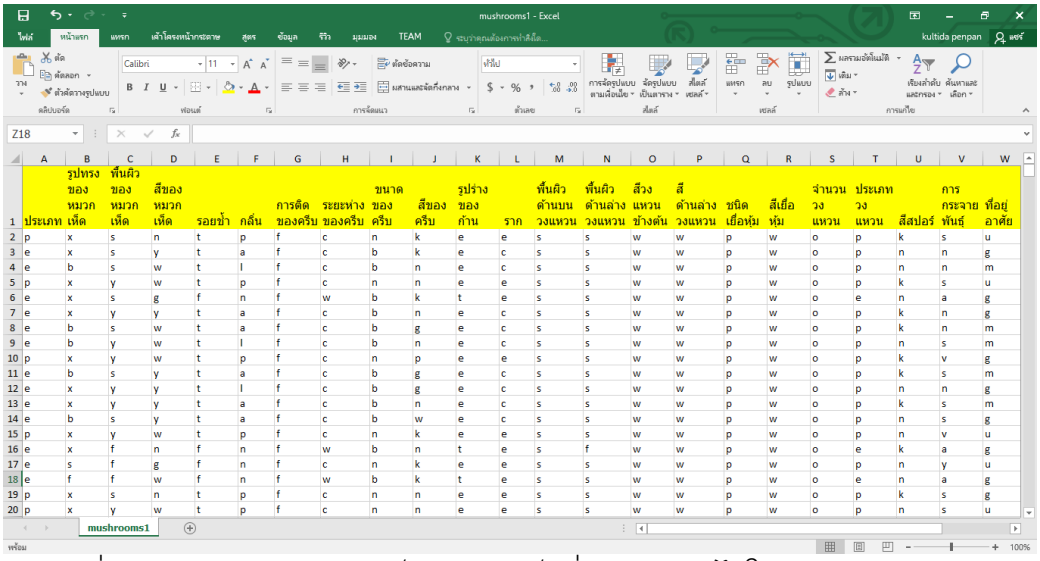
1 ประเภท	รูปทรงของหมวกเห็ด	พื้นผิวของหมวกเห็ด	สีของหมวกเห็ด	รอยช้ำ	กลิ่น	การติดของครีบ	ระยะห่างของครีบ	ขนาดของครีบ	สีของครีบ	รูปร่างของก้าน	ราก	พื้นผิวด้านบนวงแหวน	พื้นผิวด้านล่างวงแหวน	สีวงแหวนข้างต้น	สีด้านล่างวงแหวน	ชนิดเยื่อหุ้ม	สีเยื่อหุ้ม	จำนวนวงแหวน	ประเภทวงแหวน	สีสปอร์	การกระจายพันธุ์	ที่อยู่อาศัย
2 p	x	s	n	t	p	f	c	n	k	e	e	s	s	w	w	p	w	o	p	k	s	u
3 e	x	s	y	t	a	f	c	b	k	e	c	s	s	w	w	p	w	o	p	n	n	g
4 e	b	s	w	t	l	f	c	b	n	e	c	s	s	w	w	p	w	o	p	n	n	m
5 p	x	y	w	t	p	f	c	n	n	e	e	s	s	w	w	p	w	o	p	k	s	u
6 e	x	s	g	f	n	f	w	b	k	t	e	s	s	w	w	p	w	o	e	n	a	g
7 e	x	y	y	t	a	f	c	b	n	e	c	s	s	w	w	p	w	o	p	k	n	g
8 e	b	s	w	t	a	f	c	b	g	e	c	s	s	w	w	p	w	o	p	k	n	m
9 e	b	y	w	t	l	f	c	b	n	e	c	s	s	w	w	p	w	o	p	n	s	m
10 p	x	y	w	t	p	f	c	n	p	e	e	s	s	w	w	p	w	o	p	k	v	g
11 e	b	s	y	t	a	f	c	b	g	e	c	s	s	w	w	p	w	o	p	k	s	m
12 e	x	y	y	t	l	f	c	b	g	e	c	s	s	w	w	p	w	o	p	n	n	g
13 e	x	y	y	t	a	f	c	b	n	e	c	s	s	w	w	p	w	o	p	k	s	m
14 e	b	s	y	t	a	f	c	b	w	e	c	s	s	w	w	p	w	o	p	n	s	g
15 p	x	y	w	t	p	f	c	n	k	e	e	s	s	w	w	p	w	o	p	n	v	u
16 e	x	f	n	f	n	f	w	b	n	t	e	s	f	w	w	p	w	o	e	k	a	g
17 e	s	f	g	f	n	f	c	n	k	e	e	s	s	w	w	p	w	o	p	n	y	u
18 e	f	f	w	f	n	f	w	b	k	t	e	s	s	w	w	p	w	o	e	n	a	g
19 p	x	s	n	t	p	f	c	n	n	e	e	s	s	w	w	p	w	o	p	k	s	g
20 p	x	y	w	t	p	f	c	n	n	e	e	s	s	w	w	p	w	o	p	n	s	u

ภาพที่ 3. 2 ข้อมูลการจำแนกเห็ดพิษหรือเห็ดที่รับประทานได้ในอเมริกาเหนือ

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

3.1.3.1 ทำการคัดเลือกข้อมูล (Data Selection) คือการคัดเลือกข้อมูลที่เหมาะสมเพื่อนำมาใช้ในการวิเคราะห์ข้อมูล

ผู้วิเคราะห์ข้อมูลทำการคัดเลือกข้อมูล และทำการ Data Cleaning ข้อมูลการจำแนกเห็ดพิษหรือเห็ดที่รับประทานได้ในอเมริกาเหนือ โดยตัดส่วนที่ไม่จำเป็นออกให้เหลือเฉพาะข้อมูลที่จำเป็นในการวิเคราะห์ในภาพรวม จำนวน 5 แอตทริบิวท์ ได้แก่ ประเภทของเห็ด รูปทรงของหมวกเห็ด พื้นผิวของหมวกเห็ด รอยช้ำ และที่อยู่อาศัย จำนวน 3,377 รายการ เนื่องจากการศึกษาจากแหล่งอ้างอิงถึงคุณลักษณะของเห็ดพิษและเห็ดที่รับประทานได้ มักจะสังเกตได้จากรูปทรงของหมวกเห็ด พื้นผิวของหมวกเห็ด รอยช้ำ และที่อยู่อาศัย ซึ่งเป็นข้อมูลที่จำเป็นในการนำไปวิเคราะห์ข้อมูลต่อไป



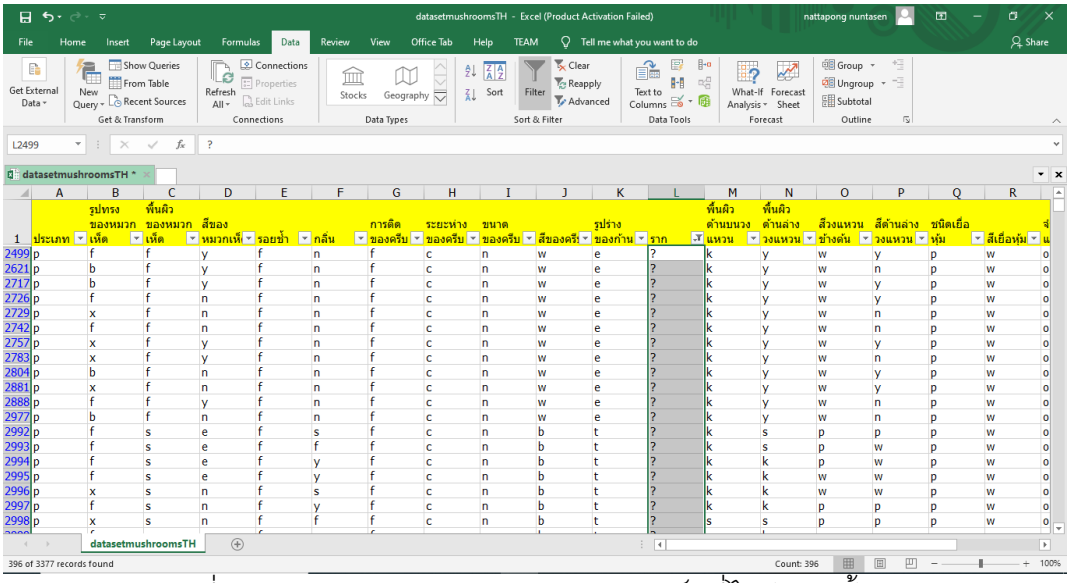
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	ประเภท	รูปทรงของหมวกเห็ด	พื้นผิวของหมวกเห็ด	สีของหมวกเห็ด	รอยช้ำ	กลิ่น	การติดของครีบ	ระยะห่างของครีบ	ขนาดของครีบ	สีของครีบ	รูปร่างของก้าน	ราก	พื้นผิวด้านบนวงแหวน	พื้นผิวด้านล่างวงแหวน	สีวงแหวนข้างต้น	สีด้านล่างวงแหวน	ชนิดเยื่อหุ้ม	สีเยื่อหุ้ม	จำนวนวงแหวน	ประเภทวงแหวน	สีสปอร์	การกระจายพันธุ์	ที่อยู่อาศัย
2	p	x	s	n	t	p	f	c	n	k	e	e	s	s	w	w	p	w	o	p	k	s	u
3	e	x	s	y	t	a	f	c	b	k	e	c	s	s	w	w	p	w	o	p	n	n	g
4	e	b	s	w	t	l	f	c	b	n	e	c	s	s	w	w	p	w	o	p	n	n	m
5	p	x	y	w	t	p	f	c	n	n	e	e	s	s	w	w	p	w	o	p	k	s	u
6	e	x	s	g	f	n	f	w	b	k	t	e	s	s	w	w	p	w	o	e	n	a	g
7	e	x	y	y	t	a	f	c	b	n	e	c	s	s	w	w	p	w	o	p	k	n	g
8	e	b	s	w	t	a	f	c	b	g	e	c	s	s	w	w	p	w	o	p	k	n	m
9	e	b	y	w	t	l	f	c	b	n	e	c	s	s	w	w	p	w	o	p	n	s	m
10	p	x	y	w	t	p	f	c	n	p	e	e	s	s	w	w	p	w	o	p	k	v	g
11	e	b	s	y	t	a	f	c	b	g	e	c	s	s	w	w	p	w	o	p	k	s	m
12	e	x	y	y	t	l	f	c	b	g	e	c	s	s	w	w	p	w	o	p	n	n	g
13	e	x	y	y	t	a	f	c	b	n	e	c	s	s	w	w	p	w	o	p	k	s	m
14	e	b	s	y	t	a	f	c	b	w	e	c	s	s	w	w	p	w	o	p	n	s	g
15	p	x	y	w	t	p	f	c	n	k	e	e	s	s	w	w	p	w	o	p	n	v	u
16	e	x	f	n	f	n	f	w	b	n	t	e	s	f	w	w	p	w	o	e	k	a	g
17	e	s	f	g	f	n	f	c	n	k	e	e	s	s	w	w	p	w	o	p	n	y	u
18	e	f	f	w	f	n	f	w	b	k	t	e	s	s	w	w	p	w	o	e	n	a	g
19	p	x	s	n	t	p	f	c	n	n	e	e	s	s	w	w	p	w	o	p	k	s	g
20	p	x	y	w	t	p	f	c	n	n	e	e	s	s	w	w	p	w	o	p	n	s	u

ภาพที่ 3. 3 ข้อมูลการจำแนกเห็ดพิษหรือเห็ดที่รับประทานได้ในอเมริกาเหนือ

3.1.3.2 ทำการกลั่นกรองข้อมูล (Data Cleaning) คือการทำความสะอาดข้อมูลเป็นกระบวนการตรวจสอบและการแก้ไข (หรือลบ) รายการข้อมูลที่ไม่ถูกต้องออกไปจากชุดข้อมูลตารางหรือฐานข้อมูล ซึ่งเป็นหลักสำคัญของฐานข้อมูล ทางผู้วิเคราะห์ข้อมูลได้ดำเนินการดังนี้

1) ข้อมูลการจำแนกเห็ดพิษหรือเห็ดที่รับประทานได้ในอเมริกาเหนือ ผู้วิเคราะห์ข้อมูลทำการแก้ไขและเปลี่ยนแปลงที่ข้อมูลที่ผิดพลาดและไม่ชัดเจนทิ้ง ซึ่งผู้วิเคราะห์ข้อมูลพบว่า บางข้อมูลราก นั้นมีค่าที่ไม่ชัดเจน จึงดูข้อมูลตามหลักความเป็นจริงโดยเปรียบเทียบข้อมูลจากส่วนใหญ่ที่ใช้ หรือข้อมูลอื่น ๆ ที่มีข้อผิดพลาด เนื่องด้วยอาจจะมาจากการคีย์ข้อมูลที่ต้นทาง ดังนั้นผู้วิเคราะห์ข้อมูลได้ทำการเปลี่ยนแปลงข้อมูลที่ไม่ชัดเจนได้ ดังนี้

– ราก. ? เป็นชนิดของรากเห็ดที่ไม่รู้จักหรืออาจจะเป็นข้อผิดพลาดของข้อมูล ผู้วิเคราะห์ข้อมูลพบว่าจะทำการเปลี่ยนแปลงชนิดของราก จึงได้ทำการตรวจสอบข้อมูลและทำการลบเปลี่ยนแปลงข้อมูลที่เป็น ? เป็นตัว n “ไม่มีราก” ทั้งหมด



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	ประเภท	รูปทรงของหมวกเห็ด	พื้นผิวของหมวกเห็ด	สีของหมวกเห็ด	รอยช้ำ	กลิ่น	การติดของครีบ	ระยะห่างของครีบ	ขนาดของครีบ	สีของครีบ	รูปร่างของก้าน	ราก	พื้นผิวด้านบนวงแหวน	พื้นผิวด้านล่างวงแหวน	สีวงแหวนข้างต้น	สีด้านล่างวงแหวน	ชนิดเยื่อหุ้ม	สีเยื่อหุ้ม
2499	p	f	f	y	f	n	f	c	n	w	e	?	k	y	w	y	p	w
2621	p	b	f	y	f	n	f	c	n	w	e	?	k	y	w	n	p	w
2717	p	b	f	y	f	n	f	c	n	w	e	?	k	y	w	y	p	w
2726	p	f	f	n	f	n	f	c	n	w	e	?	k	y	w	y	p	w
2729	p	x	f	n	f	n	f	c	n	w	e	?	k	y	w	n	p	w
2742	p	f	f	n	f	n	f	c	n	w	e	?	k	y	w	n	p	w
2757	p	x	f	y	f	n	f	c	n	w	e	?	k	y	w	y	p	w
2783	p	x	f	y	f	n	f	c	n	w	e	?	k	y	w	n	p	w
2804	p	b	f	n	f	n	f	c	n	w	e	?	k	y	w	y	p	w
2881	p	x	f	n	f	n	f	c	n	w	e	?	k	y	w	y	p	w
2888	p	f	f	y	f	n	f	c	n	w	e	?	k	y	w	n	p	w
2977	p	b	f	n	f	n	f	c	n	w	e	?	k	y	w	n	p	w
2992	p	f	s	e	f	s	f	c	n	b	t	?	k	s	p	p	p	w
2993	p	f	s	e	f	f	f	c	n	b	t	?	k	s	p	w	p	w
2994	p	f	s	e	f	y	f	c	n	b	t	?	k	k	p	w	p	w
2995	p	f	s	e	f	y	f	c	n	b	t	?	k	k	w	w	p	w
2996	p	x	s	n	f	s	f	c	n	b	t	?	k	k	w	w	p	w
2997	p	f	s	n	f	y	f	c	n	b	t	?	k	k	p	p	p	w
2998	p	x	s	n	f	f	f	c	n	b	t	?	s	s	p	p	p	w

ภาพที่ 3. 4 ข้อมูลผิดพลาดของชนิดของรากเห็ดที่ไม่ชัดเจนทั้งหมด

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

ผู้วิเคราะห์ข้อมูลวิเคราะห์ข้อมูลด้วยการทำเหมืองข้อมูลแบบ Data Classification เพื่อใช้ทำนายแนวโน้มการเกิดขึ้นของปัจจัยที่สามารถทำให้รู้ได้ว่าเห็ดชนิดไหนเป็นพิษและไม่เป็นพิษรับประทานได้หรือรับประทานไม่ได้ จากเทคนิคการจำแนกประเภทข้อมูล ด้วยการสร้างโมเดล Decision Tree เพื่อจัดกลุ่มข้อมูลที่มีความสัมพันธ์กัน โดยใช้โปรแกรมที่ใช้ทำเหมืองข้อมูลด้วยชุดข้อมูลที่คัดเลือก ดังนี้

	A	B	C	D	E
1	ประเภท	รูปทรง	พื้นผิว	รอยช้ำ	ที่อาศัย
2	p	x	s	t	u
3	e	x	s	t	g
4	e	x	s	f	g
5	e	b	s	t	g
6	e	x	f	f	g
7	e	f	f	f	g
8	p	x	s	t	g
9	p	x	s	t	u
10	p	f	s	t	g
11	e	f	f	f	u
12	e	x	s	t	d
13	e	x	f	t	d
14	e	x	f	t	d
15	e	x	f	f	u
16	e	x	s	t	g
17	p	x	s	t	u
18	e	f	f	f	g
19	e	b	s	t	g
20	e	x	s	t	g
21	e	f	s	f	g
22	e	x	s	t	g

ภาพที่ 3. 5 ชุดข้อมูลที่คัดเลือกมาวิเคราะห์ข้อมูล

จากรูปภาพที่ 3.5 ประกอบด้วย 5 แอตทริบิวต์ คือ

- ประเภท แสดงประเภทเห็ด ซึ่งเป็นค่าประกอบด้วย 2 ค่า คือ e"รับประทานได้" p" รับประทานไม่ได้"
- รูปทรง แสดงรูปทรงของเห็ด ประกอบด้วย 3 ค่า คือ b"ระฆัง" f"แบน" x"นูน"
- พื้นผิว แสดงพื้นผิวของเห็ด ประกอบด้วย 2 ค่า คือ f"เส้นใย" s"เรียบ"
- รอยฟกช้ำ แสดงรอยฟกช้ำของเห็ด ประกอบด้วย 2 ค่า คือ t"มี" f"ไม่มี"
- ที่อาศัย แสดงที่อาศัยของเห็ด ประกอบด้วย 3 ค่า คือ d"ป่า" g"ทุ่งหญ้า" u"เมือง"

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 กับจำนวนข้อมูลทั้งหมดโดยใช้ผลลัพธ์เป็นรับประทานได้ และรับประทานไม่ได้ ดังนี้

1) คำนวณค่า IG ของแอตทริบิวต์ รูปทรง จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = $-p(\text{รับประทานได้}) \times \log_2 p(\text{รับประทานได้}) + p(\text{รับประทานไม่ได้}) \times \log_2 p(\text{รับประทานไม่ได้})$

$= -[0.63 \times \log_2(0.63) + 0.37 \times \log_2(0.37)]$

$= -[0.63 \times -0.67 + 0.37 \times -1.43]$

$= -[0.42 + 0.53]$

$= 0.95$

entropy(ผล = b) = $-p(\text{รับประทานได้}) \times \log_2 p(\text{รับประทานได้}) + p(\text{รับประทานไม่ได้}) \times \log_2 p(\text{รับประทานไม่ได้})$

$= -[0.93 \times \log_2(0.93) + 0.07 \times \log_2(0.07)]$

$= -[0.93 \times -0.10 + 0.07 \times -3.84]$

$= -[0.09 + 0.27]$

= 0.36

entropy(ผล =f) = –p(รับประทานได้) × $\log_2$ p(รับประทานได้) + p(รับประทานไม่ได้) × $\log_2$ p(รับประทานไม่ได้)

= –[0.63 × $\log_2$(0.63) + 0.37 × $\log_2$(0.37)]

= –[0.63 × –0.67 + 0.37 × –1.43]

= –[0.42 + 0.53]

= 0.95

entropy(ผล =x) = –p(รับประทานได้) × $\log_2$ p(รับประทานได้) + p(รับประทานไม่ได้) × $\log_2$ p(รับประทานไม่ได้)

= –[0.60 × $\log_2$(0.60) + 0.40 × $\log_2$(0.40)]

= –[0.60 × –0.74 + 0.40 × –1.32]

= –[0.44 + 0.53]

= 0.97

IG(parent, child) = entropy (parent) – [p(ผล =b) × entropy(ผล = b) + p(ผล = f) × entropy(ผล = f)+ p(ผล = x) × entropy(ผล = x)]

= 0.95 – [0.05 × 0.36 + 0.42 × 0.95 + 0.53 × 0.97]

= 0.95 – [0.02 + 0.40 + 0.51]

= 0.95 – 0.93

= 0.02

2) คำนวณค่า IG ของแอตทริบิวต์ พื้นผิว จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = –p(รับประทานได้) × $\log_2$ p(รับประทานได้) + p(รับประทานไม่ได้) × $\log_2$ p(รับประทานไม่ได้)

= –[0.63 × $\log_2$(0.63) + 0.37 × $\log_2$(0.37)]

= –[0.63 × –0.67 + 0.37 × –1.43]

= –[0.42 + 0.53]

= 0.95

entropy(ผล = f) = −p(รับประทานได้) × $\log_2$ p(รับประทานได้) + p(รับประทานไม่ได้) × $\log_2$ p(รับประทานไม่ได้)

= −[0.73 × $\log_2$(0.73) + 0.27 × $\log_2$(0.27)]

= −[0.73 × −0.45 + 0.27 × −1.89]

= −[0.33 + 0.51]

= 0.84

entropy(ผล =s) = −p(รับประทานได้) × $\log_2$ p(รับประทานได้) + p(รับประทานไม่ได้) × $\log_2$ p(รับประทานไม่ได้)

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

= −[0.48 × −1.06 + 0.52 × −0.94]

= −[0.51 + 0.49]

= 1

IG(parent, child) = entropy (parent) – [p(ผล =f) × entropy(ผล = f) + p(ผล = s) × entropy(ผล = s)]

= 0.95 – [0.59 × 0.84 + 0.41 × 1]

= 0.95 – [0.50 + 0.41]

= 0.95 – 0.91

= 0.04

3) คำนวณค่า IG ของแอตทริบิวต์ รอยฟกช้ำ จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = −p(รับประทานได้) × $\log_2$ p(รับประทานได้) + p(รับประทานไม่ได้) × $\log_2$ p(รับประทานไม่ได้)

= −[0.63 × $\log_2$(0.63) + 0.37 × $\log_2$(0.37)]

= −[0.63 × −0.67 + 0.37 × −1.43]

= −[0.42 + 0.53]

= 0.95

entropy(ผล = f) = –p(รับประทานได้) × $\log_2$ p(รับประทานได้) + p(รับประทานไม่ได้) × $\log_2$ p(รับประทานไม่ได้)

= –[0.55 × $\log_2$(0.55) + 0.45 × $\log_2$(0.45)]

= –[0.55 × –0.86 + 0.45 × –1.15]

= –[0.47 + 0.52]

= 0.99

entropy(ผล =t) = –p(รับประทานได้) × $\log_2$ p(รับประทานได้) + p(รับประทานไม่ได้) × $\log_2$ p(รับประทานไม่ได้)

= –[0.71 × $\log_2$(0.73) + 0.29 × $\log_2$(0.27)]

= –[0.71 × –0.49 + 0.29 × –1.79]

= –[0.35 + 0.52]

= 0.87

IG(parent, child) = entropy (parent) – [p(ผล =f) × entropy(ผล = f) + p(ผล = t) × entropy(ผล = t)]

= 0.95 – [0.55 × 0.99 + 0.45 × 0.87]

= 0.95 – [0.54 + 0.39]

= 0.95 – 0.93

= 0.02

4) คำนวณค่า IG ของแอตทริบิวต์ ที่อาศัย จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = –p(รับประทานได้) × $\log_2$ p(รับประทานได้) + p(รับประทานไม่ได้) × $\log_2$ p(รับประทานไม่ได้)

= –[0.63 × $\log_2$(0.63) + 0.37 × $\log_2$(0.37)]

= –[0.63 × –0.67 + 0.37 × –1.43]

= –[0.42 + 0.53]

= 0.95

entropy(ผล = d) = –p(รับประทานได้) × $\log_2$ p(รับประทานได้) + p(รับประทานไม่ได้) × $\log_2$ p(รับประทานไม่ได้)

= –[0.61 × $\log_2$(0.61) + 0.39 × $\log_2$(0.39)]

= –[0.61 × –0.71 + 0.39 × –1.36]

= –[0.43 + 0.53]

= 0.96

entropy(ผล =g) = –p(รับประทานได้) × $\log_2$ p(รับประทานได้) + p(รับประทานไม่ได้) × $\log_2$ p(รับประทานไม่ได้)

= –[0.71 × $\log_2$(0.71) + 0.29 × $\log_2$(0.29)]

= –[0.71 × –0.49 + 0.36 × –1.79]

= –[0.35 + 0.64]

= 0.99

entropy(ผล =u) = –p(รับประทานได้) × $\log_2$ p(รับประทานได้) + p(รับประทานไม่ได้) × $\log_2$ p(รับประทานไม่ได้)

= –[0.24 × $\log_2$(0.24) + 0.76 × $\log_2$(0.76)]

= –[0.24 × –2.06 + 0.76 × –0.40]

= –[0.49 + 0.30]

= 0.79

IG(parent, child) = entropy (parent) – [p(ผล =d) × entropy(ผล = d) + p(ผล = g) × entropy(ผล = g) + p(ผล = u) × entropy(ผล = u)]

= 0.95 – [0.47 × 0.96 + 0.45 × 0.99 + 0.08 × 0.79]

= 0.95 – [0.45 + 0.45 + 0.06]

= 0.95 – 0.96

= –0.01

จากการคำนวณค่า IG ของทุกแอตทริบิวต์พบว่าค่า IG ของแอตทริบิวต์ พื้นผิว มีค่ามากที่สุด (0.04) ดังนั้นจึงเลือกแอตทริบิวต์ พื้นผิว ขึ้นมาเป็นโหนด root และจะต้องทำการแตกกิ่งจาก โหนด root ออกไปจนข้อมูลในแต่ละโหนดมีคลาสคำตอบเดียวกัน และผู้วิเคราะห์จึงทำการสร้างโหนดในระดับถัดไป

การคำนวณค่าแต่ละแอตทริบิวต์ พื้นผิว ในฝั่งของ f “เส้นใย” ในระดับที่ 2 กับจำนวนข้อมูลทั้งหมดโดยใช้ผลลัพธ์เป็นรับประทานได้ และรับประทานไม่ได้

1) คำนวณค่า IG ของแอตทริบิวต์ พื้นผิวและรูปทรง จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = $-p(\text{รับประทานได้}) \times \log_2 p(\text{รับประทานได้}) + p(\text{รับประทานไม่ได้}) \times \log_2 p(\text{รับประทานไม่ได้})$

$= -[0.73 \times \log_2(0.73) + 0.27 \times \log_2(0.37)]$

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

$= -[0.33 + 0.51]$

$= 0.84$

entropy(ผล = b) = $-p(\text{รับประทานได้}) \times \log_2 p(\text{รับประทานได้}) + p(\text{รับประทานไม่ได้}) \times \log_2 p(\text{รับประทานไม่ได้})$

$= -[0.92 \times \log_2(0.93) + 0.08 \times \log_2(0.07)]$

$= -[0.92 \times -0.12 + 0.08 \times -3.64]$

$= -[0.11 + 0.29]$

$= 0.40$

entropy(ผล =f) = $-p(\text{รับประทานได้}) \times \log_2 p(\text{รับประทานได้}) + p(\text{รับประทานไม่ได้}) \times \log_2 p(\text{รับประทานไม่ได้})$

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

$= -[0.76 \times -0.40 + 0.24 \times -2.06]$

$= -[0.30 + 0.49]$

$= 0.79$

entropy(ผล =x) = $-p(\text{รับประทานได้}) \times \log_2 p(\text{รับประทานได้}) + p(\text{รับประทานไม่ได้}) \times \log_2 p(\text{รับประทานไม่ได้})$

$= -[0.70 \times \log_2(0.70) + 0.30 \times \log_2(0.30)]$

$= -[0.70 \times -0.51 + 0.30 \times -1.74]$

$= -[0.36 + 0.52]$

$= 0.88$

IG(parent, child) = entropy (parent) – [p(ผล =b) × entropy(ผล = b) + p(ผล = f) × entropy(ผล = f)+ p(ผล = x) × entropy(ผล = x)]

= 0.84 – [0.03 × 0.40 + 0.45 × 0.79 + 0.52 × 0.88]

= 0.84 – [0.01 + 0.36 + 0.46]

= 0.84 – 0.83

= 0.01

2) คำนวณค่า IG ของแอตทริบิวต์ พื้นผิวและรอยฟกช้ำ จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = –p(รับประทานได้) × $\log_2$ p(รับประทานได้) + p(รับประทานไม่ได้) × $\log_2$ p(รับประทานไม่ได้)

= –[0.73 × $\log_2$(0.73) + 0.27 × $\log_2$(0.37)]

= –[0.73 × –0.45 + 0.27 × –1.89]

= –[0.33 + 0.51]

= 0.84

entropy(ผล = f) = –p(รับประทานได้) × $\log_2$ p(รับประทานได้) + p(รับประทานไม่ได้) × $\log_2$ p(รับประทานไม่ได้)

= –[0.50 × $\log_2$(0.50) + 0.50 × $\log_2$(0.50)]

= –[0.50 × –1 + 0.50 × –1]

= –[0.50 + 0.50]

= 1

entropy(ผล =t) = –p(รับประทานได้) × $\log_2$ p(รับประทานได้) + p(รับประทานไม่ได้) × $\log_2$ p(รับประทานไม่ได้)

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

= –[1.00 × –0.00 + 0.00 × –1]

= –[0.00 + 0.00]

= 0

IG(parent, child) = entropy (parent) – [p(ผล =f) × entropy(ผล = f) + p(ผล = t) × entropy(ผล = t)]

= 0.84 – [0.54 × 1 + 0.46 × 0]

= 0.84 – [0.54 + 0]

= 0.84 – 0.54

= 0.30

3) คำนวณค่า IG ของแอตทริบิวต์ พื้นผิวและที่อาศัย จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = –p(รับประทานได้) × $\log_2$ p(รับประทานได้) + p(รับประทานไม่ได้) × $\log_2$ p(รับประทานไม่ได้)

= –[0.73 × $\log_2$(0.73) + 0.27 × $\log_2$(0.37)]

= –[0.73 × –0.45 + 0.27 × –1.89]

= –[0.33 + 0.51]

= 0.84

entropy(ผล = d) = –p(รับประทานได้) × $\log_2$ p(รับประทานได้) + p(รับประทานไม่ได้) × $\log_2$ p(รับประทานไม่ได้)

= –[0.74 × $\log_2$(0.74) + 0.26 × $\log_2$(0.39)]

= –[0.74 × –0.43 + 0.26 × –1.94]

= –[0.32 + 0.50]

= 0.82

entropy(ผล =g) = –p(รับประทานได้) × $\log_2$ p(รับประทานได้) + p(รับประทานไม่ได้) × $\log_2$ p(รับประทานไม่ได้)

= –[0.69 × $\log_2$(0.69) + 0.31 × $\log_2$(0.31)]

= –[0.69 × –0.54 + 0.31 × –1.69]

= –[0.37 + 0.52]

= 0.89

entropy(ผล =u) = –p(รับประทานได้) × $\log_2$ p(รับประทานได้) + p(รับประทานไม่ได้) × $\log_2$ p(รับประทานไม่ได้)

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

= –[1.00 × –0.00 + 0.00 × –1.00]

= –[0.00 + 0.00]

= 0.00

IG(parent, child) = entropy (parent) – [p(ผล =d) × entropy(ผล = d) + p(ผล = g) × entropy(ผล = g) + p(ผล = u) × entropy(ผล = u)]

= 0.84 – [0.62 × 0.82 + 0.35 × 0.89 + 0.03 × 0.00]

= 0.84 – [0.51 + 0.31 + 0.00]

= 0.84 – 0.82

= 0.02

จากการคำนวณค่า IG ของแอตทริบิวต์ พื้นผิวและรูปทรง รอยฟกช้ำและที่อาศัยพบว่าค่า IG ของแอตทริบิวต์ พื้นผิวและรอยฟกช้ำ มีค่ามากที่สุด (0.30) ดังนั้นจึงเลือกแอตทริบิวต์ รอยฟกช้ำ ขึ้นมาเป็นโหนดในระดับที่ 2 ต่อจากโหนด root และทำการแตกกิ่งจากโหนดในระดับที่ 2 ออกไปจนข้อมูลในแต่ละโหนดมีคลาสคำตอบเดียวกันในฝั่งของ เส้นใย หรือ f

การคำนวณค่าแต่ละแอตทริบิวต์ในระดับที่ 3 กับจำนวนข้อมูลทั้งหมดโดยใช้ผลลัพธ์เป็นรับประทานได้ และรับประทานไม่ได้

1) คำนวณค่า IG ของแอตทริบิวต์ พื้นผิวและรอยฟกช้ำและที่อาศัย จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = –p(รับประทานได้) × $\log_2$ p(รับประทานได้) + p(รับประทานไม่ได้) × $\log_2$ p(รับประทานไม่ได้)

= –[0.50 × $\log_2$(0.50) + 0.50 × $\log_2$(0.50)]

= –[0.50 × –1 + 0.50 × –1]

= –[0.50 + 0.50]

= 1

entropy(ผล = d) = –p(รับประทานได้) × $\log_2$ p(รับประทานได้) + p(รับประทานไม่ได้) × $\log_2$ p(รับประทานไม่ได้)

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

= –[0.00 × –1.00 + 1.00 × –0.00]

= –[0.00 + 0.00]

= 0.00

entropy(ผล =g) = –p(รับประทานได้) × $\log_2$ p(รับประทานได้) + p(รับประทานไม่ได้) × $\log_2$ p(รับประทานไม่ได้)

= –[0.69 × $\log_2$(0.69) + 0.31 × $\log_2$(0.31)]

= –[0.69 × –0.54 + 0.31 × –1.69]

= –[0.37 + 0.52]

= 0.89

entropy(ผล =u) = –p(รับประทานได้) × $\log_2$ p(รับประทานได้) + p(รับประทานไม่ได้) × $\log_2$ p(รับประทานไม่ได้)

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

= –[1.00 × –0.00 + 0.00 × –1.00]

= –[0.00 + 0.00]

= 0.00

IG(parent, child) = entropy (parent) – [p(ผล =d) × entropy(ผล = d) + p(ผล = g) × entropy(ผล = g) + p(ผล = u) × entropy(ผล = u)]

= 1 – [0.30 × 0.00 + 0.64 × 0.89 + 0.06 × 0.00]

= 1 – [0.00 + 0.57 + 0.00]

= 1 – 0.57

= 0.43

2) คำนวณค่า IG ของแอตทริบิวต์ พื้นผิวและรอยฟกช้ำและรูปทรง จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = –p(รับประทานได้) × $\log_2$ p(รับประทานได้) + p(รับประทานไม่ได้) × $\log_2$ p(รับประทานไม่ได้)

= –[0.73 × $\log_2$(0.73) + 0.27 × $\log_2$(0.37)]

= –[0.73 × –0.45 + 0.27 × –1.89]

= –[0.33 + 0.51]

= 0.84

entropy(ผล = b) = −p(รับประทานได้) × $\log_2$ p(รับประทานได้) + p(รับประทานไม่ได้) × $\log_2$ p(รับประทานไม่ได้)

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

= −[1.00 × −0.00 + 0.00 × −1.00]

= −[0.00 + 0.00]

= 0.00

entropy(ผล =f) = −p(รับประทานได้) × $\log_2$ p(รับประทานได้) + p(รับประทานไม่ได้) × $\log_2$ p(รับประทานไม่ได้)

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

= −[0.00 × −1.00 + 0.00 × −1.00]

= −[0.00 + 0.00]

= 0.00

entropy(ผล =x) = −p(รับประทานได้) × $\log_2$ p(รับประทานได้) + p(รับประทานไม่ได้) × $\log_2$ p(รับประทานไม่ได้)

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

= −[0.00 × −1.00 + 0.00 × −1.00]

= −[0.00 + 0.00]

= 0.00

IG(parent, child) = entropy (parent) – [p(ผล =b) × entropy(ผล = b) + p(ผล = f) × entropy(ผล = f)+ p(ผล = x) × entropy(ผล = x)]

= 0.84 – [1.00 × 0.00 + 0.00 × 0.00 + 0.00 × 0.00]

= 0.84 – [0.00 + 0.00 + 0.00]

= 0.84 – 0.00

= 0.84

จากการคำนวณค่า IG ของแอตทริบิวต์ พื้นผิวและรอยฟกช้ำ “เส้นใย” และที่อาศัย พบว่าข้อมูลในแต่ละโหนดมีคลาสคำตอบเดียวกันแล้ว คือ ผลลัพธ์เป็นรับประทานได้ และรับประทานไม่ได้ และผู้วิเคราะห์ข้อมูลพบว่าข้อมูลในการวิเคราะห์ พื้นผิวและรอยฟกช้ำและรูปทรง นั้นผลลัพธ์ออกมาได้เป็น รับประทานได้ทั้งหมด

การคำนวณค่าแต่ละแอตทริบิวต์ พื้นผิว ในฝั่งของ s “เรียบ” ในระดับที่ 2 กับจำนวนข้อมูลทั้งหมดโดยใช้ผลลัพธ์เป็นรับประทานได้ และรับประทานไม่ได้

1) คำนวณค่า IG ของแอตทริบิวต์ พื้นผิวและรูปทรง จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = $-p(\text{รับประทานได้}) \times \log_2 p(\text{รับประทานได้}) + p(\text{รับประทานไม่ได้}) \times \log_2 p(\text{รับประทานไม่ได้})$

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

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

$= -[0.51 + 0.49]$

$= 1.00$

entropy(ผล = b) = $-p(\text{รับประทานได้}) \times \log_2 p(\text{รับประทานได้}) + p(\text{รับประทานไม่ได้}) \times \log_2 p(\text{รับประทานไม่ได้})$

$= -[0.93 \times \log_2(0.93) + 0.07 \times \log_2(0.07)]$

$= -[0.92 \times -0.10 + 0.07 \times -3.84]$

$= -[0.09 + 0.27]$

$= 0.36$

entropy(ผล = f) = $-p(\text{รับประทานได้}) \times \log_2 p(\text{รับประทานได้}) + p(\text{รับประทานไม่ได้}) \times \log_2 p(\text{รับประทานไม่ได้})$

$= -[0.41 \times \log_2(0.41) + 0.59 \times \log_2(0.59)]$

$= -[0.41 \times -1.29 + 0.59 \times -0.76]$

$= -[0.53 + 0.45]$

$= 0.98$

entropy(ผล = x) = $-p(\text{รับประทานได้}) \times \log_2 p(\text{รับประทานได้}) + p(\text{รับประทานไม่ได้}) \times \log_2 p(\text{รับประทานไม่ได้})$

$= -[0.45 \times \log_2(0.45) + 0.55 \times \log_2(0.55)]$

$= -[0.45 \times -1.15 + 0.55 \times -0.86]$

$= -[0.52 + 0.47]$

$= 0.99$

IG(parent, child) = entropy (parent) – [p(ผล =b) × entropy(ผล = b) + p(ผล = f) × entropy(ผล = f)+ p(ผล = x) × entropy(ผล = x)]

= 1.00 – [0.09 × 0.36 + 0.38 × 0.98 + 0.53 × 0.99]

= 1.00 – [0.03 + 0.37 + 0.52]

= 1.00 – 0.92

= 0.08

2) คำนวณค่า IG ของแอตทริบิวต์ พื้นผิวและรอยฟกช้ำ จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = –p(รับประทานได้) × $\log_2$ p(รับประทานได้) + p(รับประทานไม่ได้) × $\log_2$ p(รับประทานไม่ได้)

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

= –[0.48 × –1.06 + 0.52 × –0.94]

= –[0.51 + 0.49]

= 1.00

entropy(ผล = f) = –p(รับประทานได้) × $\log_2$ p(รับประทานได้) + p(รับประทานไม่ได้) × $\log_2$ p(รับประทานไม่ได้)

= –[0.63 × $\log_2$(0.63) + 0.37 × $\log_2$(0.37)]

= –[0.63 × –0.67 + 0.37 × –1.43]

= –[0.42 + 0.53]

= 0.95

entropy(ผล =t) = –p(รับประทานได้) × $\log_2$ p(รับประทานได้) + p(รับประทานไม่ได้) × $\log_2$ p(รับประทานไม่ได้)

= –[0.29 × $\log_2$(0.29) + 0.71 × $\log_2$(0.71)]

= –[0.29 × –1.79 + 0.71 × –0.49]

= –[0.52 + 0.35]

= 0.87

IG(parent, child) = entropy (parent) – [p(ผล =f) × entropy(ผล = f) + p(ผล = t) × entropy(ผล = t)]

= 1.00 – [0.56 × 0.95 + 0.44 × 0.87]

= 1.00 – [0.53 + 0.38]

= 1.00 – 0.91

= 0.09

3) คำนวณค่า IG ของแอตทริบิวต์ พื้นผิวและที่อาศัย จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = –p(รับประทานได้) × $\log_2$ p(รับประทานได้) + p(รับประทานไม่ได้) × $\log_2$ p(รับประทานไม่ได้)

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

= –[0.48 × –1.06 + 0.52 × –0.94]

= –[0.51 + 0.49]

= 1.00

entropy(ผล = d) = –p(รับประทานได้) × $\log_2$ p(รับประทานได้) + p(รับประทานไม่ได้) × $\log_2$ p(รับประทานไม่ได้)

= –[0.15 × $\log_2$(0.15) + 0.85 × $\log_2$(0.85)]

= –[0.15 × –2.74 + 0.85 × –0.23]

= –[0.41 + 0.20]

= 0.61

entropy(ผล =g) = –p(รับประทานได้) × $\log_2$ p(รับประทานได้) + p(รับประทานไม่ได้) × $\log_2$ p(รับประทานไม่ได้)

= –[0.73 × $\log_2$(0.73) + 0.27 × $\log_2$(0.27)]

= –[0.73 × –0.45 + 0.27 × –1.89]

= –[0.33 + 0.51]

= 0.84

entropy(ผล =u) = −p(รับประทานได้) × $\log_2$ p(รับประทานได้) + p(รับประทานไม่ได้) × $\log_2$ p(รับประทานไม่ได้)

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

= −[0.00 × −1.00 + 1.00 × −0.00]

= −[0.00 + 0.00]

= 0.00

IG(parent, child) = entropy (parent) – [p(ผล =d) × entropy(ผล = d) + p(ผล = g) × entropy(ผล = g) + p(ผล = u) × entropy(ผล = u)]

= 1.00 – [0.25 × 0.82 + 0.60 × 0.89 + 0.15 × 0.00]

= 1.00 – [0.21 + 0.53 + 0.00]

= 1.00 – 0.74

= 0.26

จากการคำนวณค่า IG ของแอตทริบิวต์ พื้นผิวและรูปทรง รอยฟกช้ำและที่อาศัยพบว่าค่า IG ของแอตทริบิวต์ พื้นผิวและที่อาศัย มีค่ามากที่สุด (0.26) ดังนั้นจึงเลือกแอตทริบิวต์ ที่อาศัยขึ้นมาเป็นโหนดในระดับที่ 2 ต่อจากโหนด root และทำการแตกกิ่งจากโหนดในระดับที่ 2 ออกไปจนข้อมูลในแต่ละโหนดมีคลาสคำตอบเดียวกันในฝั่งของ เรียบ หรือ s

การคำนวณค่าแต่ละแอตทริบิวต์ในระดับที่ 3 กับจำนวนข้อมูลทั้งหมดโดยใช้ผลลัพธ์เป็นรับประทานได้ และรับประทานไม่ได้ ของ กิ่ง d “ในป่า”

1) คำนวณค่า IG ของแอตทริบิวต์ พื้นผิวและที่อาศัยและรูปทรง จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = −p(รับประทานได้) × $\log_2$ p(รับประทานได้) + p(รับประทานไม่ได้) × $\log_2$ p(รับประทานไม่ได้)

= −[0.15 × $\log_2$(0.15) + 0.85 × $\log_2$(0.85)]

= −[0.15 × −2.74 + 0.85 × −0.23]

= −[0.41 + 0.20]

= 0.61

entropy(ผล = b) = −p(รับประทานได้) × $\log_2$ p(รับประทานได้) + p(รับประทานไม่ได้) × $\log_2$ p(รับประทานไม่ได้)

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

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

$= -[0.00 + 0.00]$

$= 0.00$

entropy(ผล =f) = $-p$(รับประทานได้) $\times \log_2$ p(รับประทานได้) + p(รับประทานไม่ได้) $\times \log_2$ p(รับประทานไม่ได้)

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

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

$= -[0.47 + 0.27]$

$= 0.74$

entropy(ผล =x) = $-p$(รับประทานได้) $\times \log_2$ p(รับประทานได้) + p(รับประทานไม่ได้) $\times \log_2$ p(รับประทานไม่ได้)

$= -[0.12 \times \log_2(0.12) + 0.88 \times \log_2(0.88)]$

$= -[0.12 \times -3.06 + 0.88 \times -0.18]$

$= -[0.37 + 0.16]$

$= 0.53$

IG(parent, child) = entropy (parent) – [p(ผล =b) × entropy(ผล = b) + p(ผล = f) × entropy(ผล = f) + p(ผล = x) × entropy(ผล = x)]

$= 0.61 - [0.00 \times 0.00 + 0.74 \times 0.89 + 0.64 \times 0.53]$

$= 0.61 - [0.00 + 0.66 + 0.34]$

$= 0.61 - 1.00$

$= -0.39$

2) คำนวณค่า IG ของแอตทริบิวต์ พื้นผิวและที่อาศัยและรอยฟกช้ำ จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = $-p$(รับประทานได้) $\times \log_2$ p(รับประทานได้) + p(รับประทานไม่ได้) $\times \log_2$ p(รับประทานไม่ได้)

$= -[0.15 \times \log_2(0.15) + 0.85 \times \log_2(0.85)]$

$= -[0.15 \times -2.74 + 0.85 \times -0.23]$

= –[0.41 + 0.20]

= 0.61

entropy(ผล = f) = –p(รับประทานได้) × $\log_2$ p(รับประทานได้) + p(รับประทานไม่ได้) × $\log_2$ p(รับประทานไม่ได้)

= –[0.01 × $\log_2$(0.63) + 0.99 × $\log_2$(0.99)]

= –[0.01 × –6.64 + 0.99 × –0.01]

= –[0.07 + 0.01]

= 0.08

entropy(ผล =t) = –p(รับประทานได้) × $\log_2$ p(รับประทานได้) + p(รับประทานไม่ได้) × $\log_2$ p(รับประทานไม่ได้)

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

= –[1.00 × –0.00 + 0.00 × –1.00]

= –[0.00 + 0.00]

= 0.00

IG(parent, child) = entropy (parent) – [p(ผล =f) × entropy(ผล = f) + p(ผล = t) × entropy(ผล = t)]

= 0.61 – [0.86 × 0.08 + 0.14 × 0.00]

= 0.61 – [0.07 + 0.00]

= 0.61 – 0.07

= 0.54

จากการคำนวณค่า IG ของแอตทริบิวต์ พื้นผิวและที่อาศัยและรูปทรงและรอยฟกช้ำในระดับที่2 พบว่าค่า IG ของแอตทริบิวต์ พื้นผิวและที่อาศัยและรอยฟกช้ำ มีค่ามากที่สุด (0.54) ดังนั้นจึงเลือกแอตทริบิวต์ รอยฟกช้ำ ขึ้นมาเป็นโหนดในระดับที่ 3 ต่อจากโหนด root ของ กิ่ง d และทำการแตกกิ่งจากโหนดในระดับที่ 3 ออกไปจนข้อมูลในแต่ละโหนดมีคลาสคำตอบเดียวกันในฝั่งของ เรียบ หรือ s

การคำนวณค่าแต่ละแอตทริบิวต์ในระดับที่ 3 กับจำนวนข้อมูลทั้งหมดโดยใช้ผลลัพธ์เป็นรับประทานได้ และรับประทานไม่ได้ ของ กิ่ง g “ในหญ้า”

1) คำนวณค่า IG ของแอตทริบิวต์ พื้นผิวและที่อาศัยและรูปทรง จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = $-p(\text{รับประทานได้}) \times \log_2 p(\text{รับประทานได้}) + p(\text{รับประทานไม่ได้}) \times \log_2 p(\text{รับประทานไม่ได้})$

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

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

$= -[0.33 + 0.51]$

$= 0.84$

entropy(ผล = b) = $-p(\text{รับประทานได้}) \times \log_2 p(\text{รับประทานได้}) + p(\text{รับประทานไม่ได้}) \times \log_2 p(\text{รับประทานไม่ได้})$

$= -[0.93 \times \log_2(0.93) + 0.07 \times \log_2(0.00)]$

$= -[0.93 \times -0.10 + 0.07 \times -3.84]$

$= -[0.09 + 0.27]$

$= 0.36$

entropy(ผล =f) = $-p(\text{รับประทานได้}) \times \log_2 p(\text{รับประทานได้}) + p(\text{รับประทานไม่ได้}) \times \log_2 p(\text{รับประทานไม่ได้})$

$= -[0.63 \times \log_2(0.63) + 0.37 \times \log_2(0.37)]$

$= -[0.63 \times -0.67 + 0.37 \times -1.43]$

$= -[0.42 + 0.53]$

$= 0.95$

entropy(ผล =x) = $-p(\text{รับประทานได้}) \times \log_2 p(\text{รับประทานได้}) + p(\text{รับประทานไม่ได้}) \times \log_2 p(\text{รับประทานไม่ได้})$

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

$= -[0.75 \times -0.42 + 0.25 \times -2.00]$

$= -[0.32 + 0.50]$

$= 0.82$

IG(parent, child) = entropy (parent) – [p(ผล =b) × entropy(ผล = b) + p(ผล = f) × entropy(ผล = f) + p(ผล = x) × entropy(ผล = x)]

= 0.84 – [0.15 × 0.36 + 0.37 × 0.95 + 0.49 × 0.82]

= 0.84 – [0.05 + 0.35 + 0.40]

= 0.84 – 0.80

= 0.04

2) คำนวณค่า IG ของแอตทริบิวต์ พื้นผิวและที่อาศัยและรอยฟกช้ำ จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = –p(รับประทานได้) × $\log_2$ p(รับประทานได้) + p(รับประทานไม่ได้) × $\log_2$ p(รับประทานไม่ได้)

= –[0.73 × $\log_2$(0.73) + 0.27 × $\log_2$(0.27)]

= –[0.73 × –0.45 + 0.27 × –1.89]

= –[0.33 + 0.51]

= 0.84

entropy(ผล = f) = –p(รับประทานได้) × $\log_2$ p(รับประทานได้) + p(รับประทานไม่ได้) × $\log_2$ p(รับประทานไม่ได้)

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

= –[1.00 × –0.00 + 0.00 × –1.00]

= –[0.00 + 0.00]

= 0.00

entropy(ผล =t) = –p(รับประทานได้) × $\log_2$ p(รับประทานได้) + p(รับประทานไม่ได้) × $\log_2$ p(รับประทานไม่ได้)

= –[0.36 × $\log_2$(0.36) + 0.64 × $\log_2$(0.64)]

= –[0.36 × –1.47 + 0.64 × –0.64]

= –[0.53 + 0.41]

= 0.94

IG(parent, child) = entropy (parent) – [p(ผล =f) × entropy(ผล = f) + p(ผล = t) × entropy(ผล = t)]

= 0.84 – [0.58 × 0.00 + 0.42 × 0.94]

= 0.84 – [0.00 + 0.00]

= 0.84 – 0.39

= 0.45

จากการคำนวณค่า IG ของแอตทริบิวต์ พื้นผิวและที่อาศัยและรูปทรงและรอยฟกช้ำในระดับที่2 พบว่าค่า IG ของแอตทริบิวต์ พื้นผิวและที่อาศัยและรอยฟกช้ำ มีค่ามากที่สุด (0.45) ดังนั้นจึงเลือกแอตทริบิวต์ รอยฟกช้ำ ขึ้นมาเป็นโหนดในระดับที่ 3 ต่อจากโหนด root ของ กิ่ง g และทำการแตกกิ่งจากโหนดในระดับที่ 3 ออกไปจนข้อมูลในแต่ละโหนดมีคลาสคำตอบเดียวกันในฝั่งของ เรียบ หรือ s

การคำนวณค่าแต่ละแอตทริบิวต์ในระดับที่ 4 กับจำนวนข้อมูลทั้งหมดโดยใช้ผลลัพธ์เป็นรับประทานได้ และรับประทานไม่ได้ ของ กิ่ง f “ไม่มีรอยฟกช้ำ”

1) คำนวณค่า IG ของแอตทริบิวต์ พื้นผิวและที่อาศัยและรอยฟกช้ำและรูปทรง จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = $-p$(รับประทานได้) × $\log_2$ p(รับประทานได้) + p(รับประทานไม่ได้) × $\log_2$ p(รับประทานไม่ได้)

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

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

= $-[0.00 + 0.00]$

= 0.00

entropy(ผล = b) = $-p$(รับประทานได้) × $\log_2$ p(รับประทานได้) + p(รับประทานไม่ได้) × $\log_2$ p(รับประทานไม่ได้)

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

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

= $-[0.00 + 0.00]$

= 0.00

entropy(ผล =f) = –p(รับประทานได้) × $\log_2$ p(รับประทานได้) + p(รับประทานไม่ได้) × $\log_2$ p(รับประทานไม่ได้)

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

= –[1.00 × –0.00 + 0.00 × –1.00]

= –[0.00 + 0.00]

= 0.00

entropy(ผล =x) = –p(รับประทานได้) × $\log_2$ p(รับประทานได้) + p(รับประทานไม่ได้) × $\log_2$ p(รับประทานไม่ได้)

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

= –[1.00 × –0.00 + 0.00 × –1.00]

= –[0.00 + 0.00]

= 0.00

IG(parent, child) = entropy (parent) – [p(ผล =b) × entropy(ผล = b) + p(ผล = f) × entropy(ผล = f) + p(ผล = x) × entropy(ผล = x)]

= 0.00 – [0.10 × 0.00 + 0.40 × 0.00 + 0.50 × 0.00]

= 0.00 – [0.00 + 0.00 + 0.00]

= 0.00 – 0.00

= 0.00

จากการคำนวณค่า IG ของแอตทริบิวต์ พื้นผิวและที่อาศัยและรอยฟกช้ำและรูปทรง ฝั่ง f “ไม่มีรอยฟกช้ำ” พบว่าข้อมูลในแต่ละโหนดมีคลาสคำตอบเดียวกันแล้ว คือ ผลลัพธ์เป็น รับประทานได้ และรับประทานไม่ได้ และผู้วิเคราะห์ข้อมูลพบว่าข้อมูลในการวิเคราะห์ พื้นผิวและที่อาศัยและรอยฟกช้ำและรูปทรง นั้น ผลลัพธ์ออกมาได้เป็น รับประทานได้ทั้งหมด

การคำนวณค่าแต่ละแอตทริบิวต์ในระดับที่ 4 กับจำนวนข้อมูลทั้งหมดโดยใช้ผลลัพธ์เป็นรับประทานได้ และรับประทานไม่ได้ ของ กิ่ง t “มีรอยฟกช้ำ”

1) คำนวณค่า IG ของแอตทริบิวต์ พื้นผิวและที่อาศัยและรอยฟกช้ำและรูปทรง จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = –p(รับประทานได้) × $\log_2$ p(รับประทานได้) + p(รับประทานไม่ได้) × $\log_2$ p(รับประทานไม่ได้)

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

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

$= -[0.53 + 0.41]$

$= 0.94$

entropy(ผล = b) = $-p$(รับประทานได้) $\times \log_2$ p(รับประทานได้) + p(รับประทานไม่ได้) $\times \log_2$ p(รับประทานไม่ได้)

$= -[0.88 \times \log_2(0.88) + 0.12 \times \log_2(0.12)]$

$= -[0.88 \times -0.18 + 0.12 \times -3.06]$

$= -[0.16 + 0.37]$

$= 0.53$

entropy(ผล =f) = $-p$(รับประทานได้) $\times \log_2$ p(รับประทานได้) + p(รับประทานไม่ได้) $\times \log_2$ p(รับประทานไม่ได้)

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

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

$= -[0.00 + 0.00]$

$= 0.00$

entropy(ผล =x) = $-p$(รับประทานได้) $\times \log_2$ p(รับประทานได้) + p(รับประทานไม่ได้) $\times \log_2$ p(รับประทานไม่ได้)

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

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

$= -[0.53 + 0.43]$

$= 0.96$

IG(parent, child) = entropy (parent) – [p(ผล =b) × entropy(ผล = b) + p(ผล = f) × entropy(ผล = f) + p(ผล = x) × entropy(ผล = x)]

$= 0.94 - [0.21 \times 0.53 + 0.32 \times 0.00 + 0.47 \times 0.96]$

$= 0.94 - [0.11 + 0.00 + 0.45]$

$= 0.94 - 0.56$

$= 0.38$

จากการคำนวณค่า IG ของแอตทริบิวต์ พื้นผิวและที่อาศัยและรูปทรงและรอยฟกช้ำในระดับที่3 พบว่าค่า IG ของแอตทริบิวต์ พื้นผิวและที่อาศัยและรอยฟกช้ำและรูปทรง มีค่ามากที่สุด (0.38) ดังนั้นจึงเลือกแอตทริบิวต์ รูปทรง ขึ้นมาเป็นโหนดในระดับที่ 4 ต่อจากโหนด root ของ กิ่ง t และ ผู้วิเคราะห์ข้อมูลพบว่าข้อมูลในการวิเคราะห์ พื้นผิวและที่อาศัยและรอยฟกช้ำและรูปทรง นั้นผลลัพธ์ออกมาได้เป็น รับประทานได้และรับประทานไม่ได้

การคำนวณค่าแต่ละแอตทริบิวต์ในระดับที่ 3 กับจำนวนข้อมูลทั้งหมดโดยใช้ผลลัพธ์เป็นรับประทานได้ และรับประทานไม่ได้ ของ กิ่ง u “ในเมือง”

1) คำนวณค่า IG ของแอตทริบิวต์ พื้นผิวและที่อาศัยและรูปทรง จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = $-p$(รับประทานได้) $\times \log_2$ p(รับประทานได้) + p(รับประทานไม่ได้) $\times \log_2$ p(รับประทานไม่ได้)

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

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

$= -[0.00 + 0.00]$

$= 0.00$

entropy(ผล = b) = $-p$(รับประทานได้) $\times \log_2$ p(รับประทานได้) + p(รับประทานไม่ได้) $\times \log_2$ p(รับประทานไม่ได้)

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

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

$= -[0.00 + 0.00]$

$= 0.00$

entropy(ผล =f) = $-p$(รับประทานได้) $\times \log_2$ p(รับประทานได้) + p(รับประทานไม่ได้) $\times \log_2$ p(รับประทานไม่ได้)

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

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

$= -[0.00 + 0.00]$

$= 0.00$

entropy(ผล =x) = –p(รับประทานได้) × $\log_2$ p(รับประทานได้) + p(รับประทานไม่ได้) × $\log_2$ p(รับประทานไม่ได้)

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

= –[0.00 × –1.00 + 1.00 × –0.00]

= –[0.00 + 0.00]

= 0.00

IG(parent, child) = entropy (parent) – [p(ผล =b) × entropy(ผล = b) + p(ผล = f) × entropy(ผล = f) + p(ผล = x) × entropy(ผล = x)]

= 0.00 – [0.00 × 0.00 + 0.50 × 0.00 + 0.50 × 0.00]

= 0.00 – [0.00 + 0.00 + 0.00]

= 0.00 – 0.00

= 0.00

2) คำนวณค่า IG ของแอตทริบิวต์ พื้นผิวและที่อาศัยและรอยฟกช้ำ จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = –p(รับประทานได้) × $\log_2$ p(รับประทานได้) + p(รับประทานไม่ได้) × $\log_2$ p(รับประทานไม่ได้)

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

= –[0.00 × –1.00 + 1.00 × –0.00]

= –[0.00 + 0.00]

= 0.00

entropy(ผล = f) = –p(รับประทานได้) × $\log_2$ p(รับประทานได้) + p(รับประทานไม่ได้) × $\log_2$ p(รับประทานไม่ได้)

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

= –[0.00 × –1.00 + 0.00 × –1.00]

= –[0.00 + 0.00]

= 0.00

entropy(ผล =t) = −p(รับประทานได้) × $\log_2$ p(รับประทานได้) + p(รับประทานไม่ได้) × $\log_2$ p(รับประทานไม่ได้)

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

= −[0.00 × −1.00 + 1.00 × −0.00]

= −[0.00 + 0.00]

= 0.00

IG(parent, child) = entropy (parent) – [p(ผล =f) × entropy(ผล = f) + p(ผล = t) × entropy(ผล = t)]

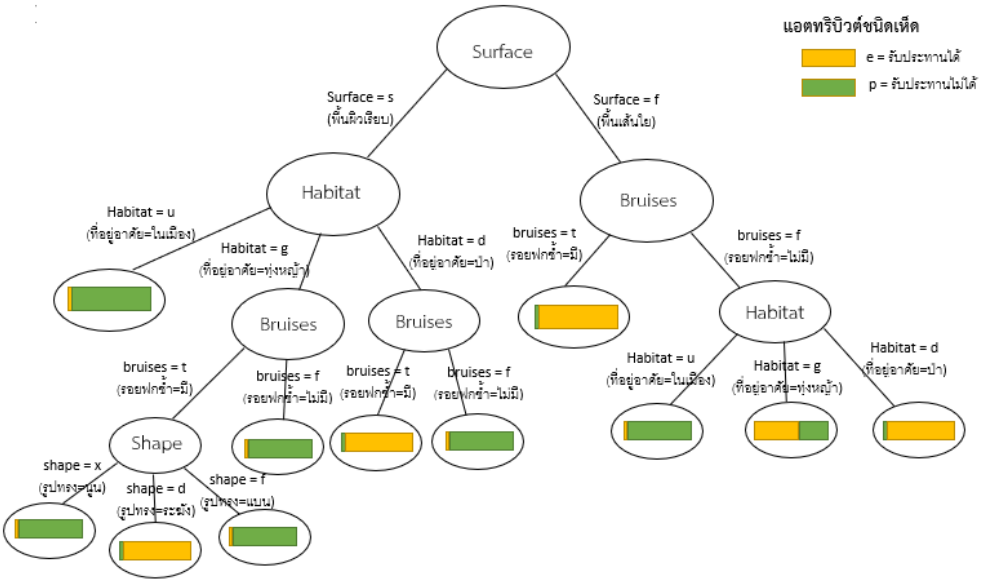
= 0.00 – [0.00 × 0.00 + 0.01 × 0.00]

= 0.00 – [0.00 + 0.00]

= 0.00 – 0.00

= 0.00

จากการคำนวณค่า IG ของแอตทริบิวต์ พื้นผิวและที่อาศัยและรูปทรงและรอยฟกช้ำในระดับที่3 พบว่าค่า IG ของแอตทริบิวต์ พื้นผิวและที่อยู่อาศัยไปรูปทรงกับรอยฟกช้ำ มีค่า IG (0.00) ซึ่งไม่สามารถแตกกิ่งได้แล้ว ดังนั้นจึงสรุปข้อมูลได้ว่า รับประทานไม่ได้ ทั้งหมด ของ กิ่ง u “ในเมือง” ตามดังรูป

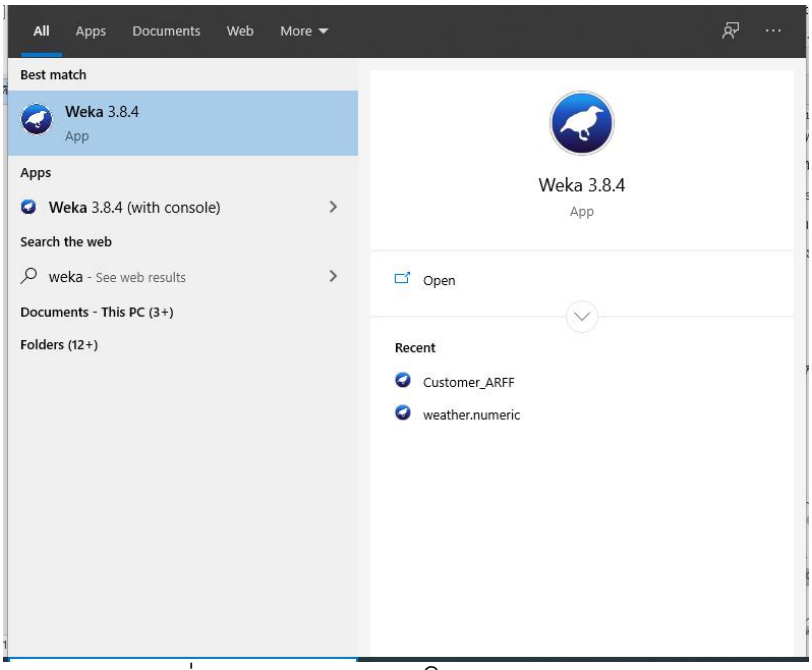


ภาพที่ 3. 6 แสดงรูปแบบโมเดล Decision Tree จากการคำนวณด้วยมือ

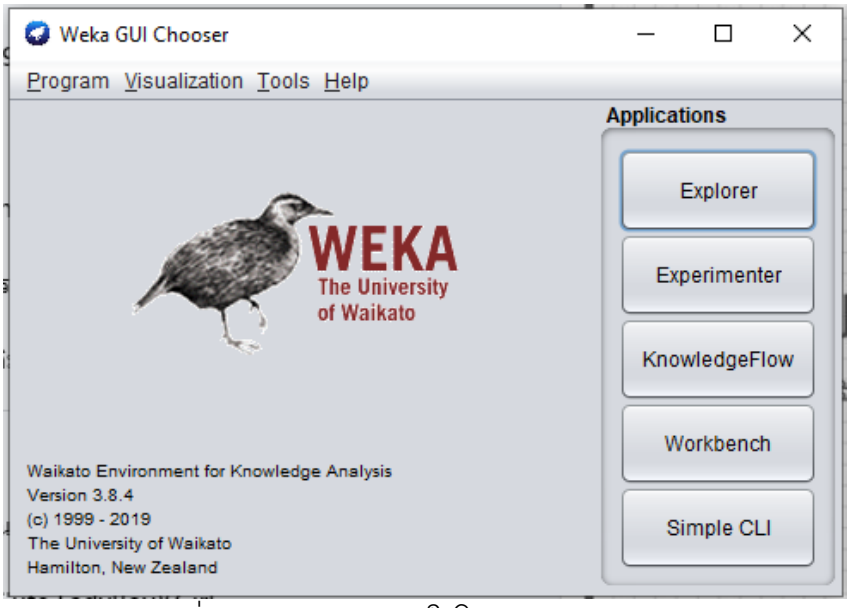
3.1.5 การประเมินผล (Evaluation) เป็นขั้นตอนก่อนนำผลลัพธ์ที่ได้จากขั้นตอนที่ 4 ไปใช้งาน ด้วยการวัดประสิทธิผลของผลลัพธ์ที่ได้กับวัตถุประสงค์ที่ตั้งไว้ในขั้นตอนแรก ว่ามีนัยสำคัญหรือความน่าเชื่อถือมากน้อยเพียงใด ด้วยการประเมินผลจากโปรแกรมว่าถูกต้องหรือไม่

ผู้วิเคราะห์ได้ทำการทดลองโมเดล เพื่อวัดประสิทธิภาพที่ตรงกับความต้องการ ซึ่งการวัดประสิทธิภาพด้วยวิธี Self-Consistency Test เหมาะสำหรับใช้ในการทดสอบประสิทธิภาพ เพื่อดูแนวโน้มของโมเดลที่สร้างขึ้น และเมื่อนำข้อมูลมาทดสอบ (Testing data) กับโปรแกรมที่ผู้วิเคราะห์เลือกมาทดสอบกับข้อมูลที่ผ่านการวิเคราะห์ข้อมูลด้วยเทคนิค Data Mining จากการสร้างโมเดล Decision Tree จึงนำข้อมูลดังกล่าว มาทดสอบกับโปรแกรม Weka เวอร์ชัน 3.8.4 ซึ่งมีขั้นตอนการทำงาน ดังนี้

ขั้นตอนที่ 1 เปิดโปรแกรม Weka 3.8.4

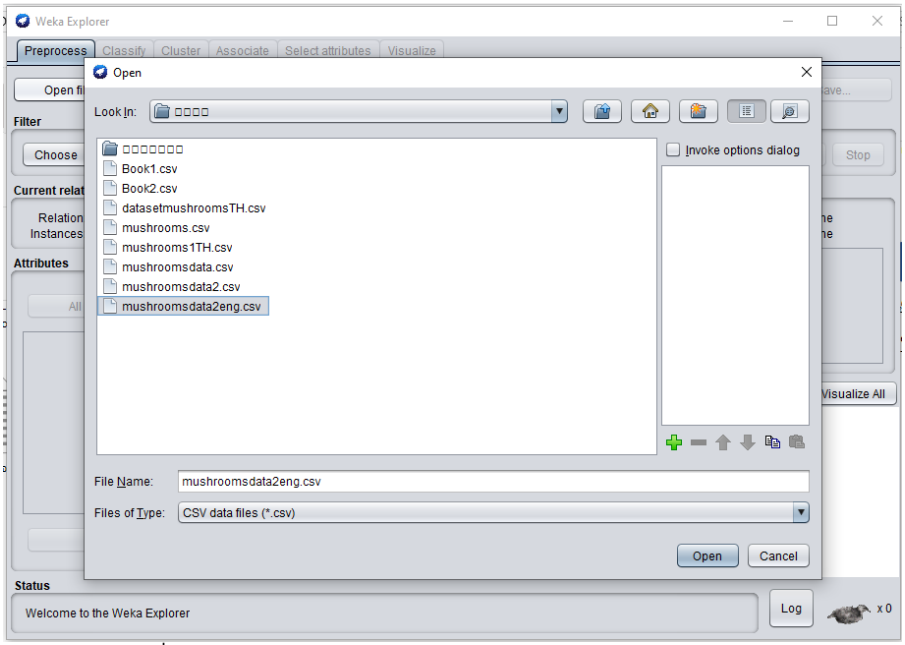


ภาพที่ 3. 7 แสดงการเปิดโปรแกรม Weka 3.8.4

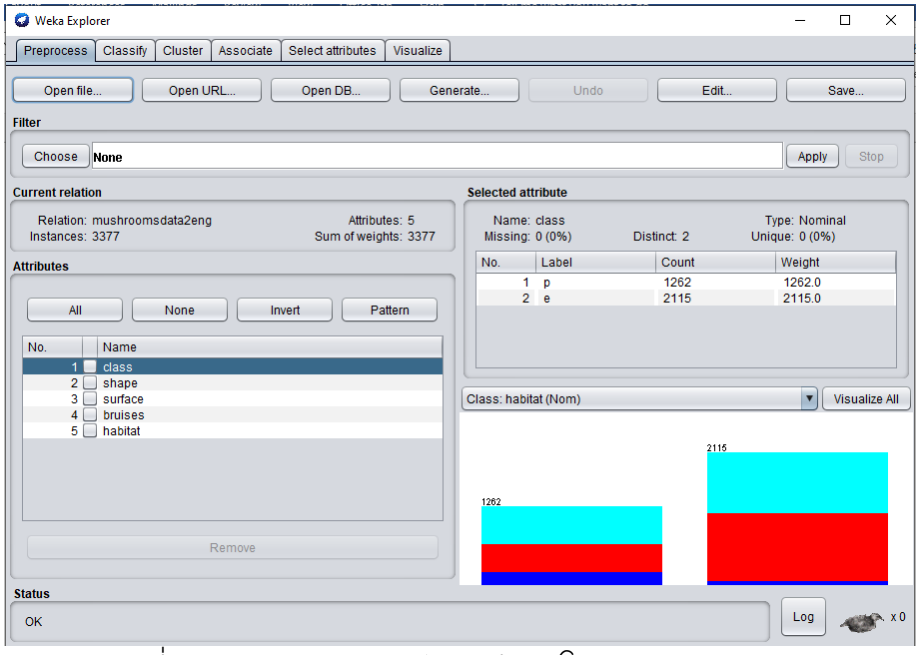


ภาพที่ 3. 8 แสดงการเข้าใช้โปรแกรม Weka 3.8.4

ขั้นตอนที่ 2 นำเข้าข้อมูลที่ได้จัดเตรียมไว้ โดยเลือกที่ Application>>Explorer>>Open file เลือกไฟล์ข้อมูลที่ต้องการนำมาทดสอบตามภาพที่ 3.7 และหลังจาก นั้นโปรแกรมแสดงหน้าจอข้อมูล ตามภาพที่ 3.8

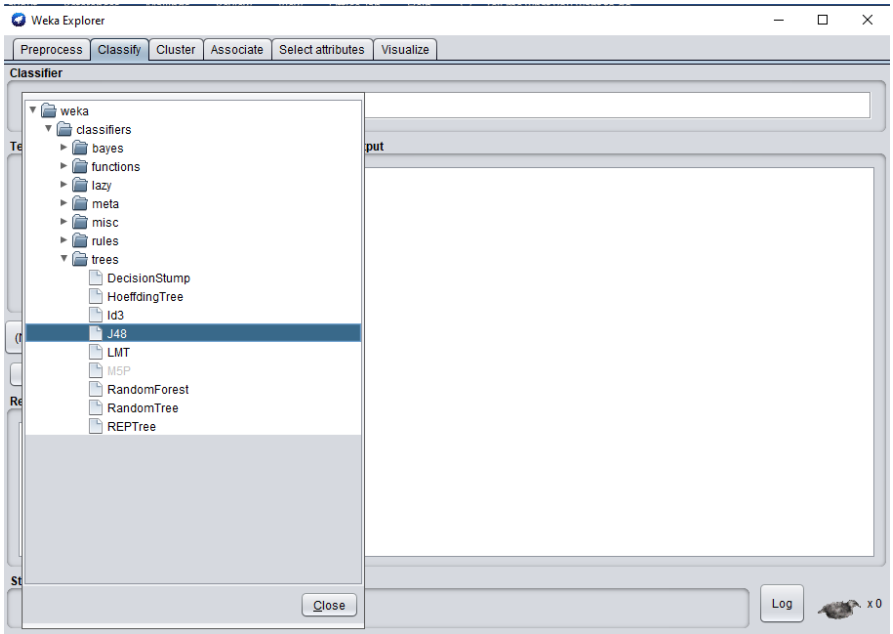


ภาพที่ 3. 9 แสดงการนำไฟล์ข้อมูลเข้าโปรแกรม Weka 3.8.4

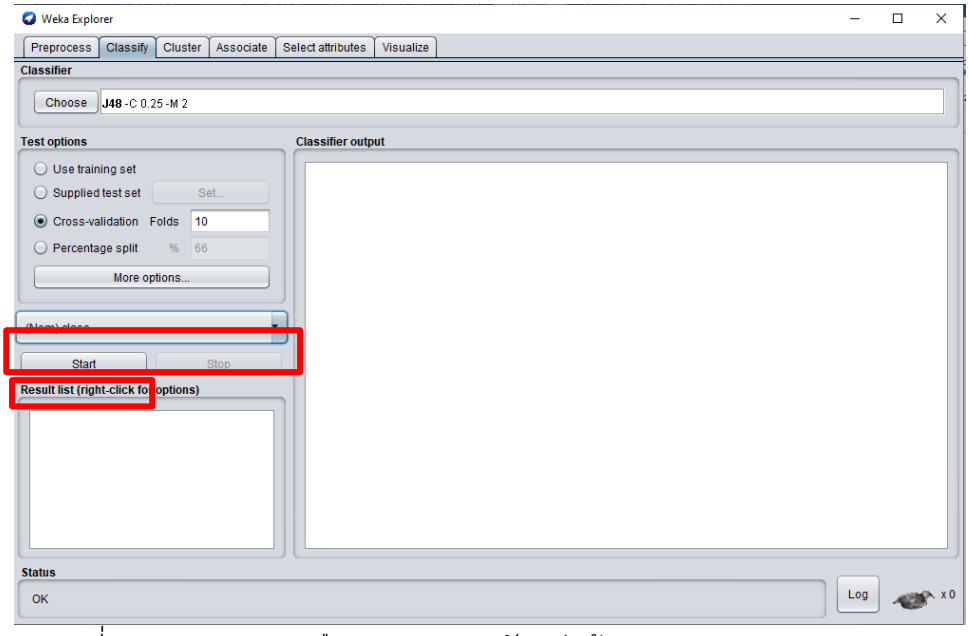


ภาพที่ 3. 10 แสดงข้อมูลหลังจากนำเข้าโปรแกรม Weka 3.8.4

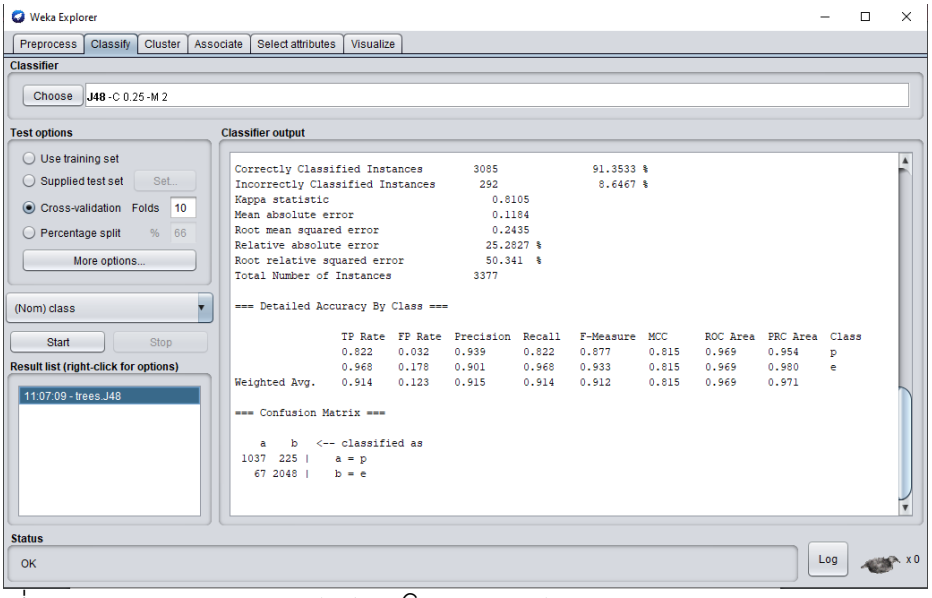
ขั้นตอนที่ 3 ดำเนินการเลือกเทคนิคการจัดกลุ่มข้อมูลแบบ Decision Tree โดย เลือกที่ Classification>>Choose>>tree และเลือกรูปแบบเป็น J48 ตามภาพที่ 3.9 จากนั้นทำการเลือก entropy (parent) แล้วกด ปุ่ม Start ตามภาพที่ 3.10 จะแสดงผลลัพธ์ที่ได้ตามภาพที่ 3.11



ภาพที่ 3. 11 แสดงการเลือกเทคนิคการจัดกลุ่มข้อมูลแบบ Decision Tree



ภาพที่ 3. 12 แสดงการเลือกเทคนิคการจัดกลุ่มข้อมูลแบบ Decision Tree: J48

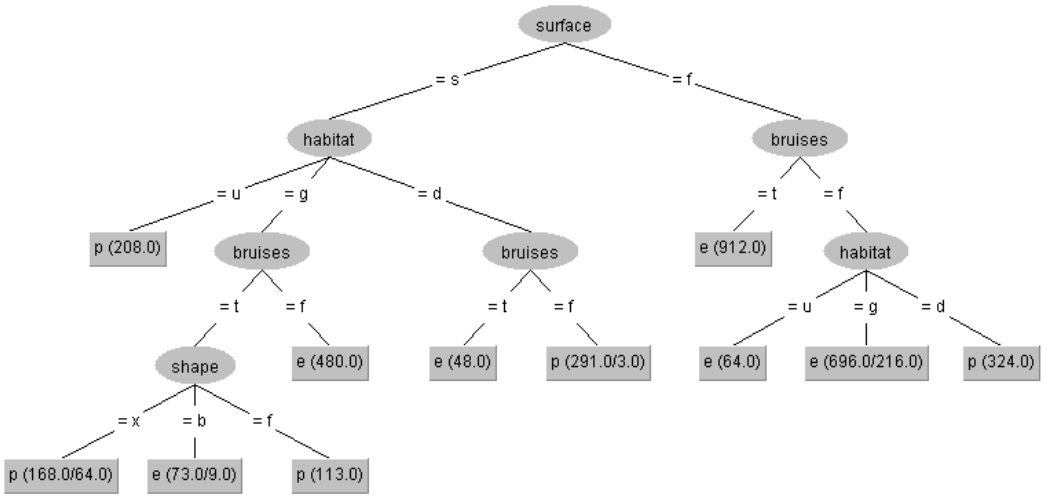


ภาพที่ 3. 13 แสดงหน้าจอผลลัพธ์ของโมเดลการจัดกลุ่มข้อมูลแบบ Decision Tree: J48

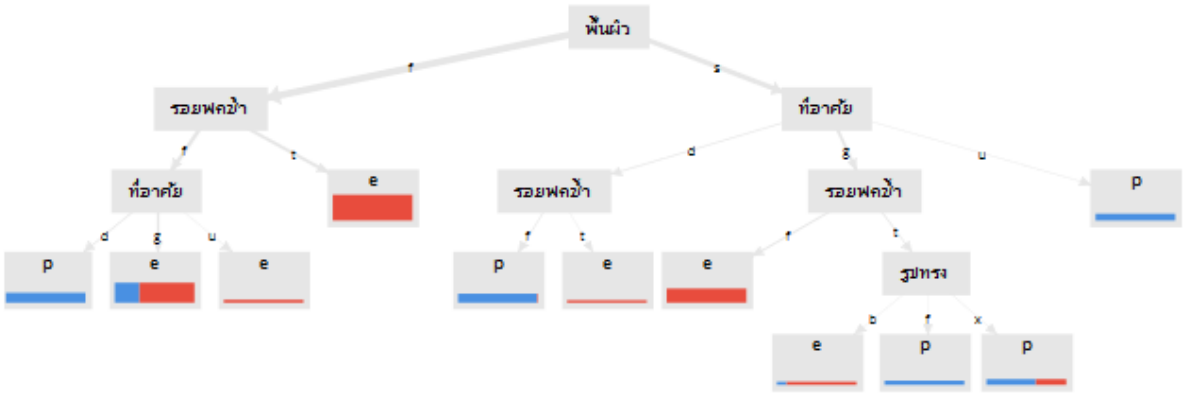
จากผลลัพธ์การทดลองพบว่าเทคนิค Decision Tree: J48 ให้ผลลัพธ์การจำแนกประเภทเห็ด รับประทานได้(e) และรับประทานไม่ได้(p) มีความถูกต้องถึง 91.35%

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

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

surface = s
|   habitat = u: p (208.0)
|   habitat = g
|   |   bruises = t
|   |   |   shape = x: p (168.0/64.0)
|   |   |   shape = b: e (73.0/9.0)
|   |   |   shape = f: p (113.0)
|   |   bruises = f: e (480.0)
|   habitat = d
|   |   bruises = t: e (48.0)
|   |   bruises = f: p (291.0/3.0)
surface = f
|   bruises = t: e (912.0)
|   bruises = f
|   |   habitat = u: e (64.0)
|   |   habitat = g: e (696.0/216.0)
|   |   habitat = d: p (324.0)

Number of Leaves  :     11

Size of the tree :      18
```

ภาพที่ 3. 14 แสดงผลลัพธ์กฎต้นไม้การตัดสินใจที่แตกกิ่งออกมา



ภาพที่ 3. 15 แสดงรูปแบบโมเดล Graph Decision Tree ในโปรแกรม Weka 3.8.4



ภาพที่ 3. 16 แสดงรูปแบบโมเดล Graph Decision Tree ในโปรแกรม RapidMiner

Tree

```
พื้นผิว = f
|   รอยฟกช้ำ = f
|   |   ที่อาศัย = d: p {p=324, e=0}
|   |   ที่อาศัย = g: e {p=216, e=480}
|   |   ที่อาศัย = u: e {p=0, e=64}
|   รอยฟกช้ำ = t: e {p=0, e=912}
พื้นผิว = s
|   ที่อาศัย = d
|   |   รอยฟกช้ำ = f: p {p=288, e=3}
|   |   รอยฟกช้ำ = t: e {p=0, e=48}
|   ที่อาศัย = g
|   |   รอยฟกช้ำ = f: e {p=0, e=480}
|   |   รอยฟกช้ำ = t
|   |   |   รูปทรง = b: e {p=9, e=64}
|   |   |   รูปทรง = f: p {p=113, e=0}
|   |   |   รูปทรง = x: p {p=104, e=64}
|   ที่อาศัย = u: p {p=208, e=0}
```

ภาพที่ 3. 17 คำบรรยายลักษณะงาน Decision Tree ในโปรแกรม RapidMiner

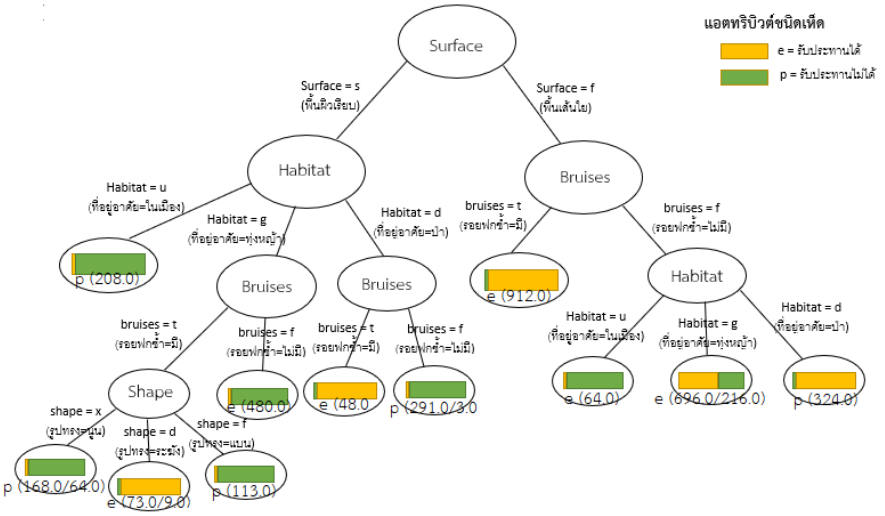
หลังจากผู้วิเคราะห์เลือกการทดสอบประสิทธิภาพของ Model ด้วยวิธี Self Consistency Test หรือเรียกว่า Use Training Set เป็นวิธีการที่นำข้อมูลที่ใช้ในการสร้างโมเดล (model) และข้อมูลที่ใช้ในการทดสอบโมเดลเป็นข้อมูลชุดเดียวกัน คือข้อมูลการจำแนกเห็ดพิษหรือเห็ดที่รับประทานได้ในอเมริกาเหนือ ที่ได้ทำการคัดเลือกมาทั้งหมด ซึ่งผู้วิเคราะห์เลือก

โปรแกรมที่ใช้นำเสนอ คือ โปรแกรม Weka 3.8.4 พบว่าการทดสอบประสิทธิภาพโมเดล Decision Tree (J48) พิจารณาได้ว่า โมเดลที่ถูกสร้างขึ้น มีค่าความถูกต้องเฉลี่ยในทุกโมเดล เท่ากับ 91.35% มีค่าการทำนายข้อมูลไม่ถูกต้องเท่ากับ 8.65% และมีค่าความคลาดเคลื่อน เท่ากับ 0.2435 และเมื่อพิจารณาส่วนค่า Confusion Matrix ในภาพที่ 3.17 พบว่าการหาค่า ของข้อมูลค่าจริง กับจำนวนข้อมูลจากการทำนาย แบ่งตามประเภทของผลการจำแนกเห็ด รับประทานได้(e) และรับประทานไม่ได้(p) และนำมาหาค่าเฉลี่ยรวมของทุก class ได้ค่าเฉลี่ยรวม เท่ากับ 0.895 มีผลลัพธ์ตรงกันอยู่ในระดับค่อนข้างดี สามารถนำโมเดลไปใช้งานได้

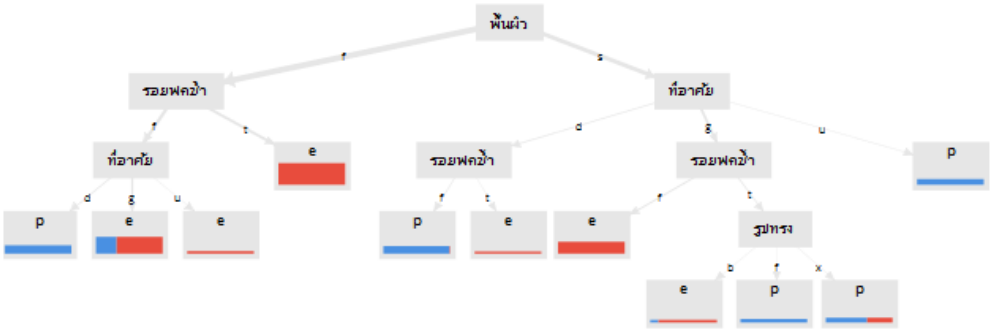
ดังนั้น ผู้วิเคราะห์จึงเลือกใช้โมเดล Decision Tree: J48 ในการนำไปหาแนวทางการวางแผนการจัดการและรับมือต่อปัจจัยเสี่ยงต่าง ๆ ของผู้ที่ประสบปัญหาด้านการจำแนกเห็ด และยังเป็นปัจจัยชี้นำประสิทธิภาพการทำงานขององค์กรการเกษตรและภาครัฐต่าง ๆ ได้อีกด้วย เพราะมีค่าความถูกต้องของโมเดล และค่าเฉลี่ย Confusion Matrix จากการจำแนกเห็ด รับประทานได้ (e) และรับประทานไม่ได้ (p) อยู่ในระดับที่ค่อนข้างดี

```
=== Summary ===

Correctly Classified Instances        3085               91.3533 %
Incorrectly Classified Instances       292                8.6467 %
Kappa statistic                          0.8105
Mean absolute error                      0.1184
Root mean squared error                  0.2435
Relative absolute error                 25.2827 %
Root relative squared error             50.341  %
Total Number of Instances             3377

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.822    0.032    0.939      0.822    0.877      0.815    0.969     0.954     p
                 0.968    0.178    0.901      0.968    0.933      0.815    0.969     0.980     e
Weighted Avg.    0.914    0.123    0.915      0.914    0.912      0.815    0.969     0.971

=== Confusion Matrix ===

    a    b   <-- classified as
 1037  225 |    a = p
   67 2048 |    b = e
```

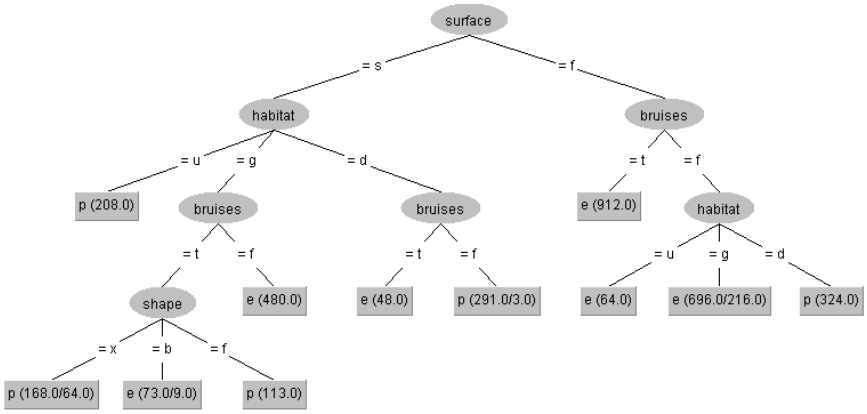
ภาพที่ 3. 18 แสดงผลลัพธ์จากการจำแนกกลุ่มแบบ Decision Tree: J48 ในโปรแกรม Weka 3.8.4



ภาพที่ 3. 19 แสดงรูปแบบโมเดล Decision Tree จากการคำนวณด้วยมือ



ภาพที่ 3. 20 แสดงรูปแบบโมเดล Graph Decision Tree ในโปรแกรม RapidMiner



ภาพที่ 3. 21 แสดงรูปแบบโมเดล Graph Decision Tree ในโปรแกรม Weka 3.8.4

ดังนั้นจะเห็นได้ว่าทางผู้วิเคราะห์ได้นำโมเดลของข้อมูลที่เลือกใช้ มาเปรียบเทียบกับโมเดลที่สร้างด้วย โปรแกรม RapidMiner Studio 9.5.1 ซึ่งผู้วิเคราะห์พบว่า ได้ผลลัพธ์จากการจำแนกกลุ่ม และรูปแบบแผนภาพโมเดลที่ตรงกันสูง ผู้วิเคราะห์จึงใช้เทคนิคของการจำแนกกลุ่มแบบ Decision Tree: J48 มาใช้ในการศึกษา เนื่องจากให้ผลลัพธ์ของกฎที่สามารถทำนายได้จำนวน 11 กฎ ซึ่งสามารถนำไปใช้ในการแบ่งกลุ่มได้ตามเงื่อนไขได้ชัดเจน และสามารถนำกฎที่ได้ไปวิเคราะห์กฎต่อไปได้ โดยสามารถจำแนกกฎได้ ดังนี้

กฎข้อที่ 1 IF surface=s Then habitat=u หมายความว่า ถ้าพื้นผิวเป็นประเภทผิวเรียบและอาศัยอยู่ในเมือง ผลการพิจารณาพบว่า เห็ดชนิดนี้รับประทานไม่ได้

กฎข้อที่ 2 IF surface=s Then habitat=g Then bruises=t Then shape=x หมายความว่า ถ้าพื้นผิวเป็นประเภทผิวเรียบและอาศัยอยู่ในทุ่งหญ้าและมีรอยฟกช้ำและมีรูปทรงนูน ผลการพิจารณาพบว่า เห็ดชนิดนี้รับประทานไม่ได้

กฎข้อที่ 3 IF surface=s Then habitat=g Then bruises=t Then shape=b หมายความว่า ถ้าพื้นผิวเป็นประเภทผิวเรียบและอาศัยอยู่ในทุ่งหญ้าและมีรอยฟกช้ำและมีรูปทรงระฆัง ผลการพิจารณาพบว่า เห็ดชนิดนี้รับประทานได้

กฎข้อที่ 4 IF surface=s Then habitat=g Then bruises=t Then shape=f หมายความว่า ถ้าพื้นผิวเป็นประเภทผิวเรียบและอาศัยอยู่ในทุ่งหญ้าและมีรอยฟกช้ำและมีรูปทรงแบน ผลการพิจารณาพบว่า เห็ดชนิดนี้รับประทานไม่ได้

กฎข้อที่ 5 IF surface=s Then habitat=g Then bruises=f หมายความว่า ถ้าพื้นผิวเป็นประเภทผิวเรียบและอาศัยอยู่ในทุ่งหญ้าและไม่มีรอยฟกช้ำ ผลการพิจารณาพบว่า เห็ดชนิดนี้รับประทานได้

กฎข้อที่ 6 IF surface=s Then habitat=d Then bruises=t หมายความว่า ถ้าพื้นผิวเป็นประเภทผิวเรียบและอาศัยอยู่ในป่าและมีรอยฟกช้ำ ผลการพิจารณาพบว่าเห็ดชนิดนี้รับประทานได้

กฎข้อที่ 7 IF surface=s Then habitat=d Then bruises=f หมายความว่า ถ้าพื้นผิวเป็นประเภทผิวเรียบและอาศัยอยู่ในป่าและไม่มีรอยฟกช้ำ ผลการพิจารณาพบว่าเห็ดชนิดนี้รับประทานไม่ได้

กฎข้อที่ 8 IF surface=f Then bruises=t หมายความว่า ถ้าพื้นผิวเป็นประเภทผิวเส้นใยและมีรอยฟกช้ำ ผลการพิจารณาพบว่าเห็ดชนิดนี้รับประทานได้

กฎข้อที่ 9 IF surface=f Then bruises=f Then habitat=u หมายความว่า ถ้าพื้นผิวเป็นประเภทผิวเส้นใยและไม่มีรอยฟกช้ำและอาศัยอยู่ในเมือง ผลการพิจารณาพบว่าเห็ดชนิดนี้รับประทานได้

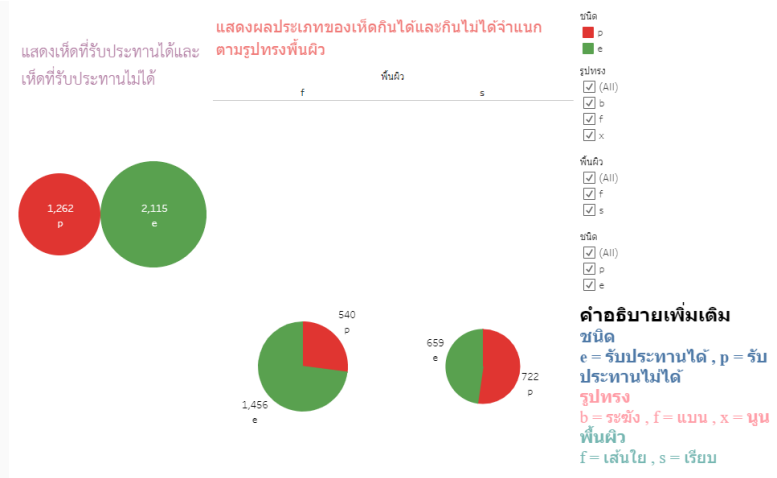
กฎข้อที่ 10 IF surface=f Then bruises=f Then habitat=g หมายความว่า ถ้าพื้นผิวเป็นประเภทผิวเส้นใยและไม่มีรอยฟกช้ำและอาศัยอยู่ในทุ่งหญ้า ผลการพิจารณาพบว่าเห็ดชนิดนี้รับประทานได้

กฎข้อที่ 11 IF surface=f Then bruises=f Then habitat=d หมายความว่า ถ้าพื้นผิวเป็นประเภทผิวเส้นใยและไม่มีรอยฟกช้ำและอาศัยอยู่ในป่า ผลการพิจารณาพบว่าเห็ดชนิดนี้รับประทานไม่ได้

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

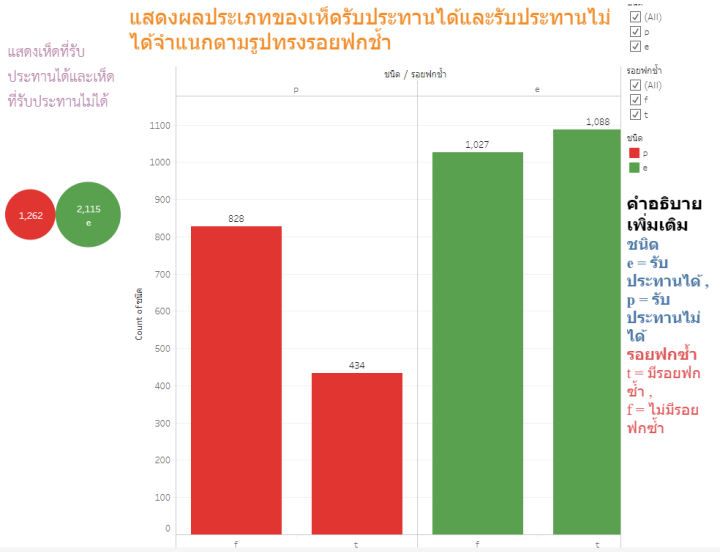
ผู้วิเคราะห์ข้อมูลนำผลข้อมูลที่ทำการวิเคราะห์แล้ว แสดงผลข้อมูลบน Web browser โดยใช้ชุดคำสั่ง HTML และ CSS3 ร่วมกับการนำเสนอข้อมูลแบบ visualization ด้วยการแสดงผลข้อมูลในรูปแบบของภาพโดยใช้โปรแกรม Tableau Public

1) แสดงข้อมูลแสดงผลการวิเคราะห์ประเภทของเห็ดรับประทานได้และรับประทานไม่ได้จำแนกตามพื้นผิว ซึ่งเป็นการแสดงผลในรูปแบบแผนภูมิวงกลม ผู้ใช้สามารถเลือกดูการจำแนกประเภทของเห็ดได้



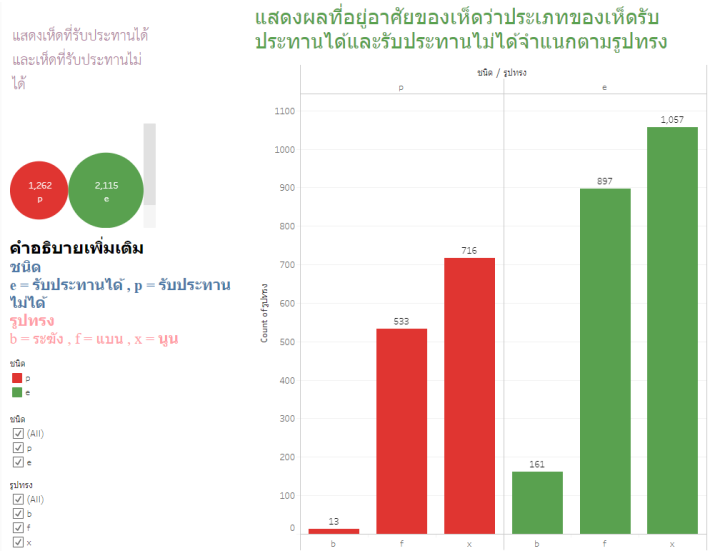
ภาพที่ 3. 22 แสดงข้อมูลแสดงผลการวิเคราะห์ประเภทของเห็ดรับประทานได้และรับประทานไม่ได้จำแนกตามพื้นผิวของเห็ด

2) แสดงข้อมูลประเภทของเห็ดรับประทานได้และรับประทานไม่ได้จำแนกตามรอยฟกช้ำ ซึ่งเป็นการแสดงผลในรูปแบบแผนภูมิแท่ง ผู้ใช้สามารถเลือกดูการจำแนกประเภทของเห็ดได้



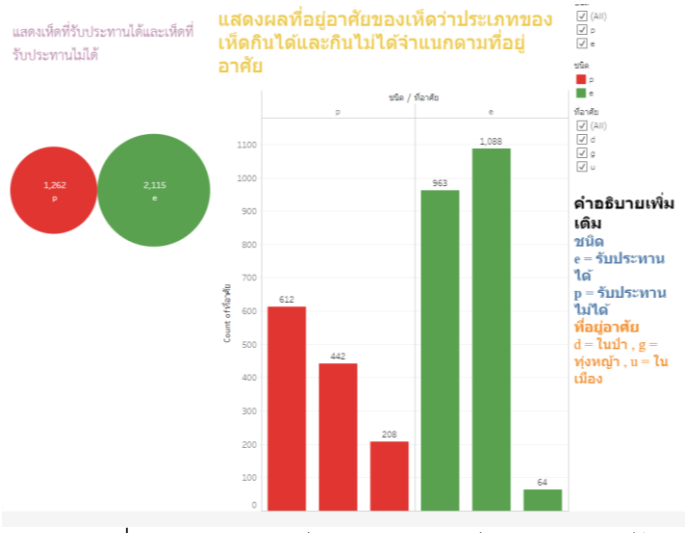
ภาพที่ 3. 23 แสดงผลประเภทของเห็ดรับประทานได้และรับประทานไม่ได้จำแนกตามรอยฟกช้ำ

3) แสดงข้อมูลประเภทของเห็ดรับประทานได้และรับประทานไม่ได้จำแนกตามรูปทรง ซึ่งเป็นการแสดงผลในรูปแบบแผนภูมิแท่ง ผู้ใช้สามารถเลือกดูการจำแนกประเภทของเห็ดได้



ภาพที่ 3. 24 แสดงข้อมูลประเภทของเห็ดรับประทานได้และรับประทานไม่ได้จำแนกตามรูปทรง

4) แสดงผลข้อมูลที่อยู่อาศัยของเห็ดว่าประเภทเห็ดรับประทานได้และรับประทานไม่ได้จำแนกตามที่อยู่อาศัย ซึ่งเป็นการแสดงผลในรูปแบบแผนภูมิแท่ง ผู้ใช้สามารถเลือกดูการจำแนกประเภทของเห็ดได้

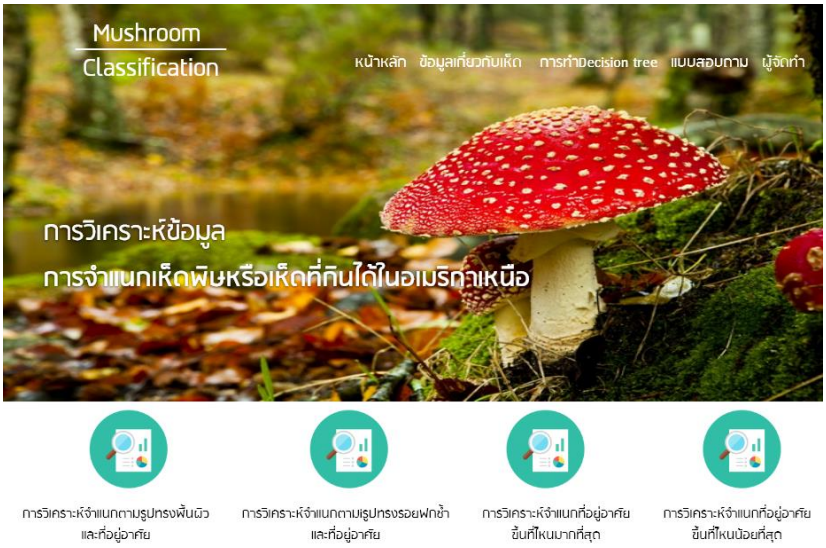


ภาพที่ 3. 25 แสดงผลข้อมูลที่อยู่อาศัยของเห็ดว่าประเภทเห็ดรับประทานได้และรับประทานไม่ได้จำแนกตามที่อยู่อาศัย

3.2 การออกแบบเว็บไซต์

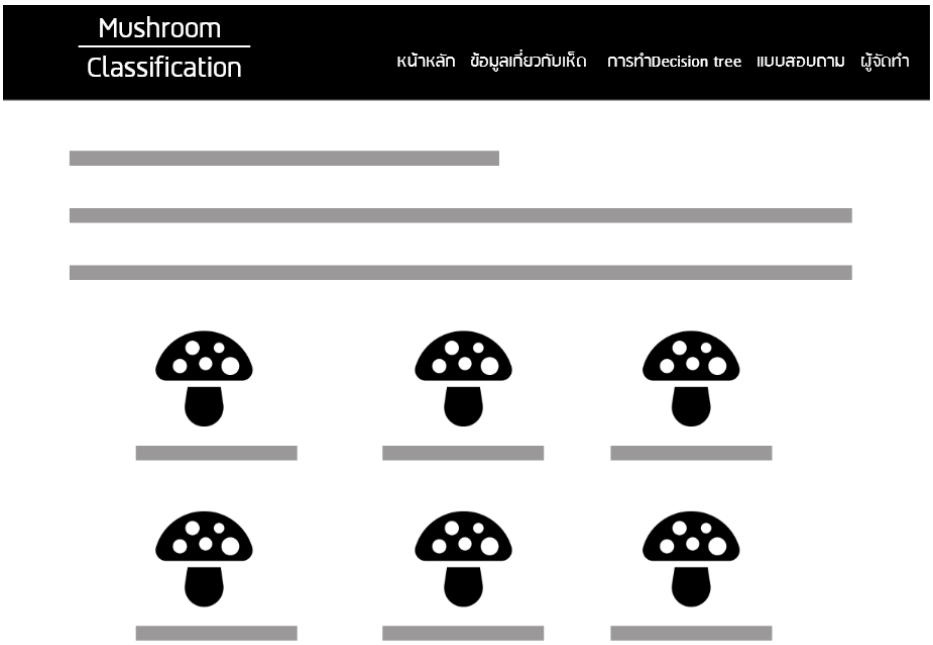
3.2.1 การออกแบบ Wireframe หน้าจอเว็บไซต์

1) หน้าแรกของเว็บไซต์ แสดงเมนูต่าง ๆ ของหน้าเว็บไซต์



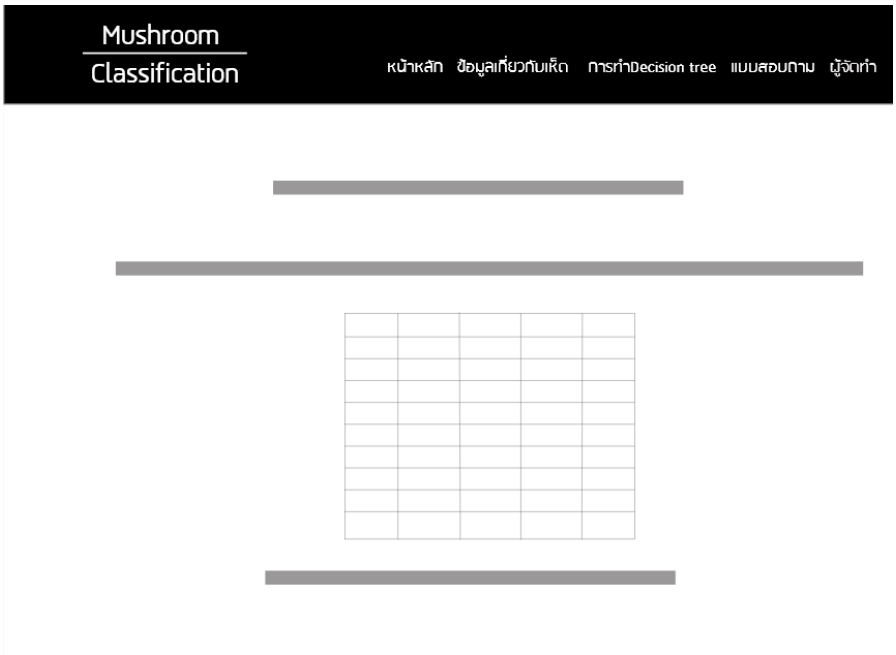
ภาพที่ 3. 26 แสดงหน้าแรกของเว็บไซต์ แสดงเมนูต่าง ๆ ของหน้าเว็บไซต์

2) หน้าข้อมูลเกี่ยวกับเห็ด



ภาพที่ 3. 27 แสดงหน้าข้อมูลเกี่ยวกับเห็ด

3) หน้าการทำการทำDecision tree

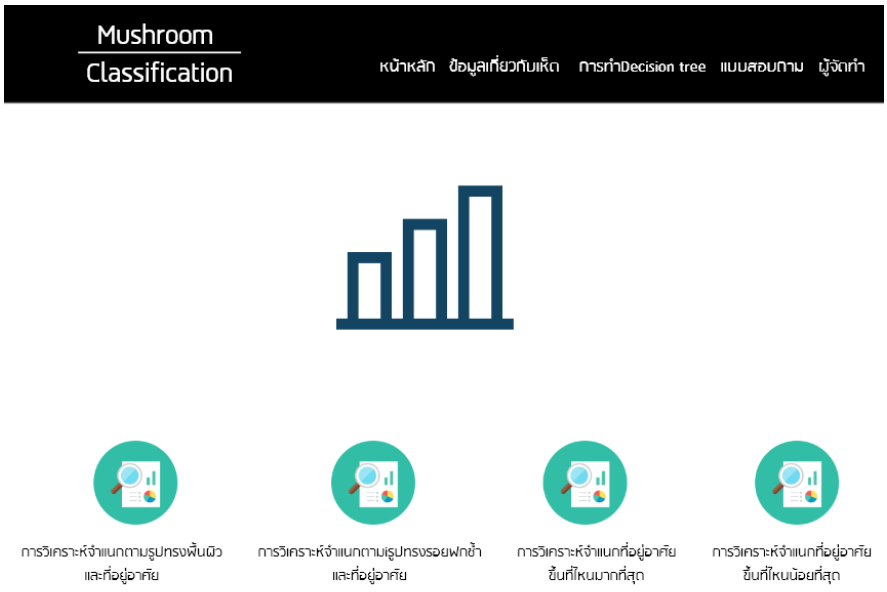


ภาพที่ 3. 28 แสดงหน้าการทำการทำDecision tree

4) หน้าแสดงผลข้อมูลการวิเคราะห์จำแนกตามรูปทรงพื้นผิวและที่อยู่อาศัย

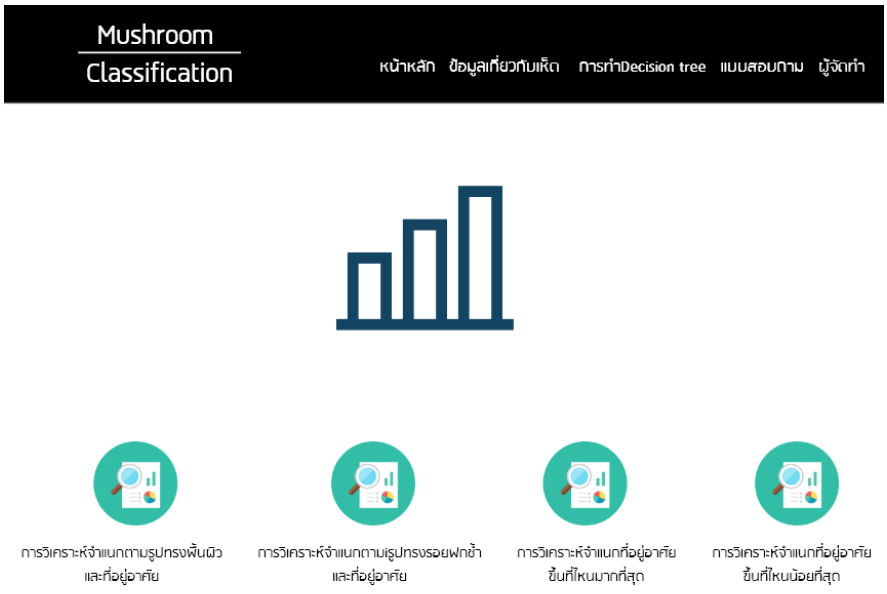
ภาพที่ 3. 29 แสดงหน้าแสดงผลข้อมูลการวิเคราะห์จำแนกตามรูปทรงพื้นผิวและที่อยู่อาศัย

5) หน้าแสดงผลข้อมูลการวิเคราะห์จำแนกตามรูปทรงรอยฟกช้ำและที่อยู่อาศัย



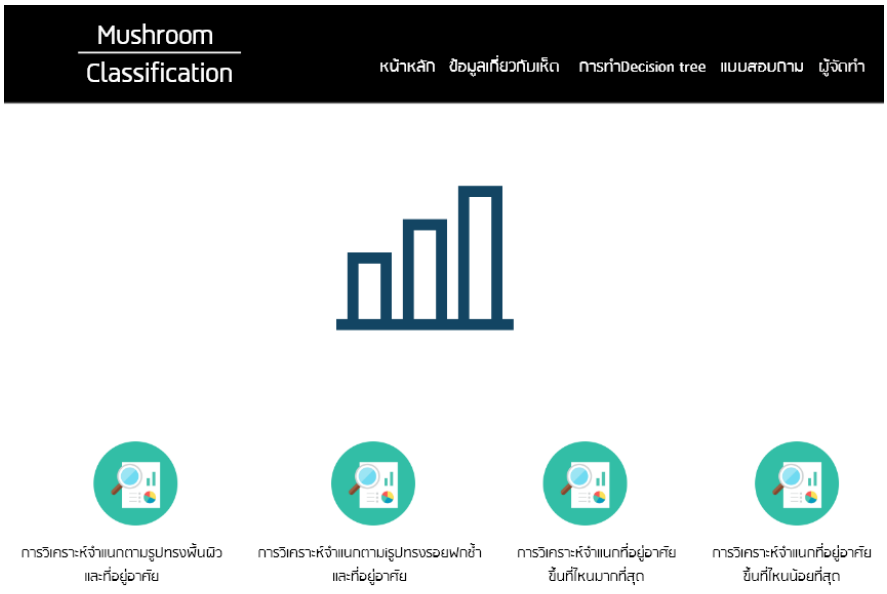
ภาพที่ 3. 30 แสดงหน้าแสดงผลข้อมูลการวิเคราะห์จำแนกตามรูปทรงรอยฟกช้ำและที่อยู่อาศัย

6) หน้าแสดงผลข้อมูลการวิเคราะห์จำแนกที่อยู่อาศัยขึ้นที่ไหนมากที่สุด



ภาพที่ 3. 31 แสดงหน้าแสดงผลข้อมูลการวิเคราะห์จำแนกที่อยู่อาศัยขึ้นที่ไหนมากที่สุด

7) หน้าแสดงผลข้อมูลการวิเคราะห์จำแนกที่อยู่อาศัยขึ้นที่ไหนน้อยที่สุด



ภาพที่ 3. 32 แสดงหน้าแสดงผลข้อมูลการวิเคราะห์จำแนกที่อยู่อาศัยขึ้นที่ไหนน้อยที่สุด

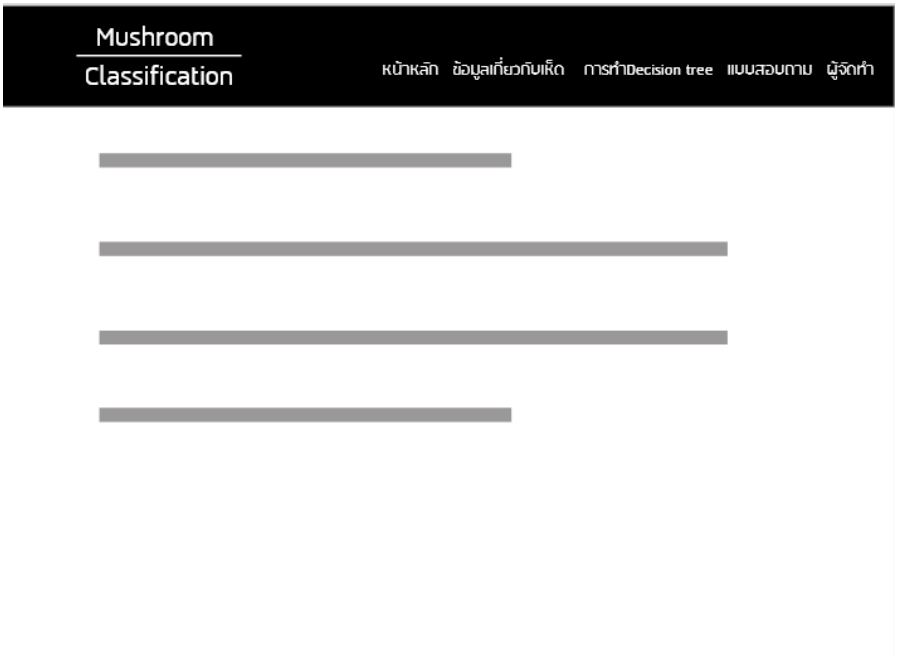
8) หน้าแบบทดสอบการให้ความรู้เกี่ยวกับเห็ด

ภาพที่ 3. 33 แสดงหน้าแบบทดสอบการให้ความรู้เกี่ยวกับเห็ด

9) หน้าแบบสอบถามการนำ Decision Treeมาประยุกต์

ภาพที่ 3. 34 แสดงหน้าแบบสอบถามการนำ Decision Treeมาประยุกต์

10) หน้าแบบสอบถามความพึ่งพอใจการใช้เว็บไซต์



ภาพที่ 3. 35 แสดงหน้าแบบสอบถามความพึ่งพอใจการใช้เว็บไซต์

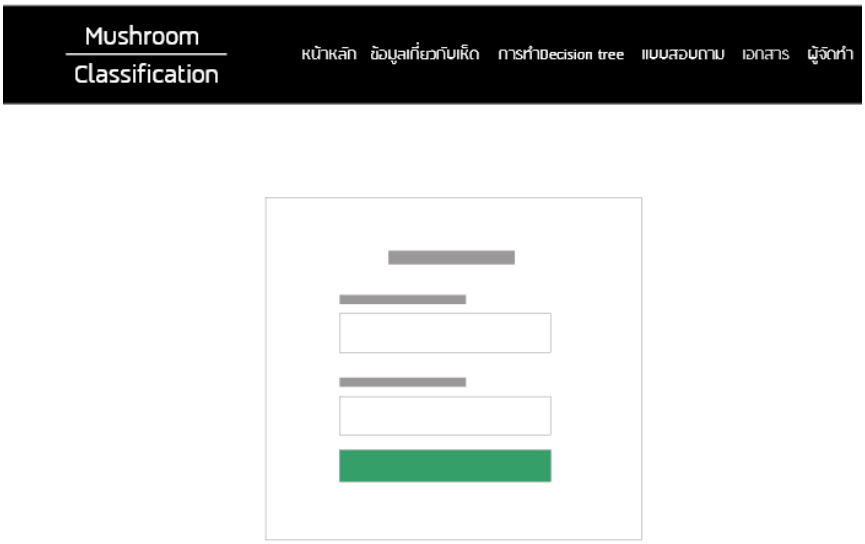
11) หน้าแสดงผลข้อมูลผู้จัดทำ

ภาพที่ 3. 36 แสดงหน้าแสดงผลข้อมูลผู้จัดทำ

12) หน้าเอกสาร

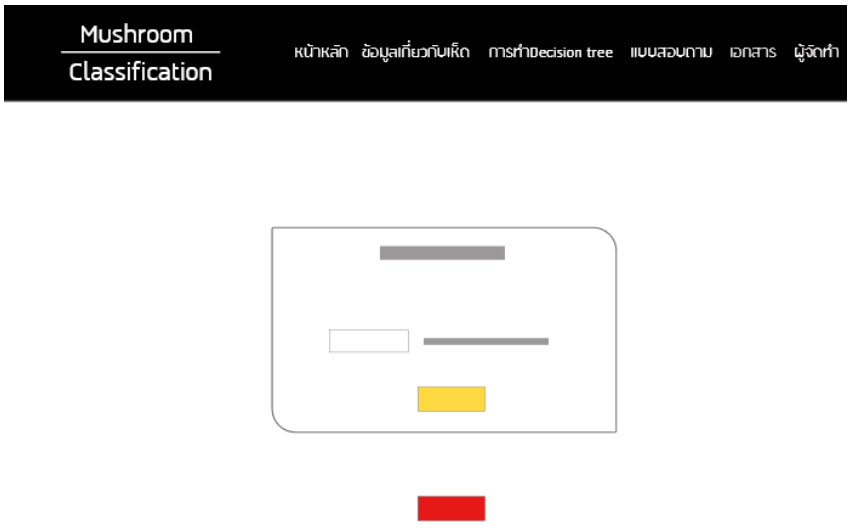
ภาพที่ 3. 37 แสดงหน้าเอกสาร

13) หน้าล็อกอินสำหรับผู้วิเคราะห์



ภาพที่ 3. 38 แสดงหน้าล็อกอินสำหรับผู้วิเคราะห์

14)หน้าอัปโหลดไฟล์



3.3 บทสรุป

จากวิธีการดำเนินงานโครงงานในข้างต้นทั้งหมดนี้ ผู้วิเคราะห์ข้อมูลการวิเคราะห์ข้อมูลการแยกประเภทของเห็ดในอเมริกาเหนือ เพื่อเผยแพร่ข้อมูลบนเว็บไซต์ ด้วยขั้นตอนกระบวนการวิเคราะห์ข้อมูลด้วย CRISP-DM อย่างละเอียด รวมถึงการสร้างโมเดล Decision Tree จากการคำนวณด้วยมือและโปรแกรมที่ใช้ทำเหมืองข้อมูล คือ, โปรแกรม Weka 3.8.4 และโปรแกรม RapidMiner Studio 9.5.1 ในการสร้างโมเดล Decision Tree ซึ่งได้ผลลัพธ์ที่เหมือนกัน และนำข้อมูลสารสนเทศมาทำการแสดงผลแบบ visualization ในรูปแบบของภาพโดยใช้โปรแกรม Tableau Public และออกแบบ Wireframe ของเว็บไซต์ที่จะเผยแพร่บน Web Browser ด้วยโปรแกรม Adobe XD

บทที่ 4
ผลการดำเนินงาน

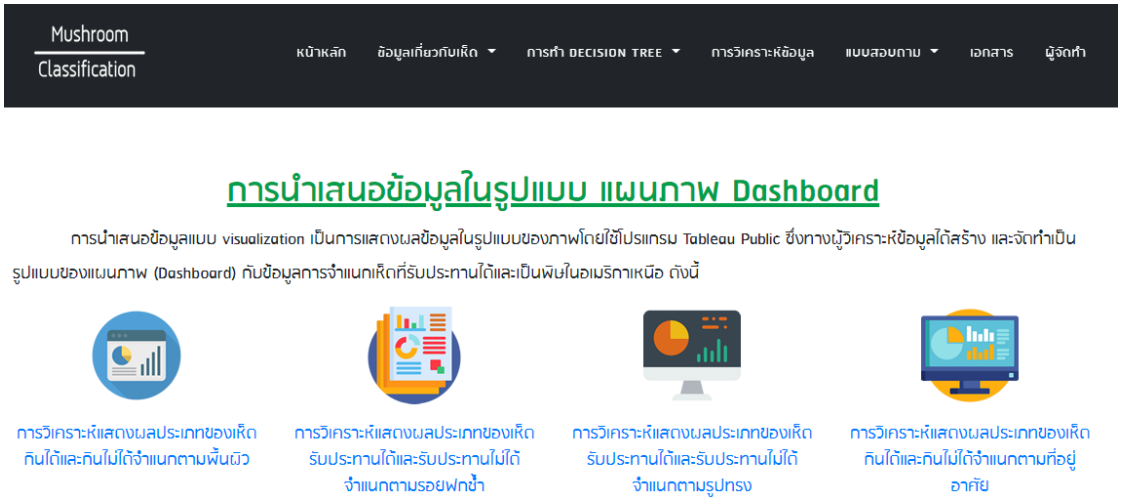
จากการวิเคราะห์ข้อมูลการวิเคราะห์ข้อมูลการจำแนกเห็ดพิษหรือเห็ดที่กินได้ในอเมริกาเหนือ เพื่อใช้สำหรับเผยแพร่ข้อมูลบนเว็บไซต์ มีวัตถุประสงค์เพื่อวิเคราะห์ข้อมูลการจำแนกเห็ดพิษหรือเห็ดที่รับประทานได้ในอเมริกาเหนือ สำหรับเผยแพร่ข้อมูลบนเว็บไซต์ เพื่อให้ได้เว็บไซต์ ที่รวมแหล่งความรู้ทั่วไปเกี่ยวกับข้อมูล และแสดงกระบวนการวิเคราะห์ข้อมูลที่สามารถนาไปใช้งานได้อย่างมีประสิทธิภาพและประสิทธิผล โดยผู้วิเคราะห์โครงงานได้ทำการวิเคราะห์ข้อมูลและนำเทคโนโลยีเข้าช่วยในการเสนอข้อมูล ทำให้มีความสะดวกรวดเร็วและสามารถใช้งานได้จริงจนสำเร็จลุล่วงตามเป้าหมาย

4.1 ผลการดำเนินงาน

การจัดทำโครงงานการจำแนกเห็ดพิษหรือเห็ดที่กินได้ในอเมริกาเหนือ เพื่อใช้สำหรับเผยแพร่ข้อมูลบนเว็บไซต์ ในครั้งนี้พบวาการจำแนกข้อมูลออกเป็นประเภทต่าง ๆ ช่วยให้เกิดความเข้าใจต่อข้อมูล และสิ่งที่ต้องการทราบอย่างแท้จริง ผู้วิเคราะห์ได้ทำการวิเคราะห์ข้อมูลให้อยู่ในรูปแบบของ Data Mining โดยทำการจำแนกข้อมูลออกเป็นประเภทต่าง ๆ ด้วยเทคนิคการทำเหมืองข้อมูลแบบ Classification จากเทคนิคการจำแนกประเภทข้อมูลด้วยการวิเคราะห์ในรูปแบบของแผนภูมิต้นไม้ หรือที่เรียกว่า Decision tree และนำเสนอข้อมูลสารสนเทศแบบ Visualization เผยแพรข้อมูลสารสนเทศนี้บน Web browser ให้กับผู้ใช้งาน ซึ่งทำให้เข้าใจได้อย่างสะดวกรวดเร็วยิ่งขึ้น รวมถึงมีความถูกต้อง ลดความซ้ำซ้อนจากข้อมูลที่มีจำนวนมากมหาศาล เพิ่มประสิทธิภาพให้กับการศึกษาค้นคว้าและทันต่อเวลา ผู้วิเคราะห์จึงได้ทำการวิเคราะห์ข้อมูล และจัดทำเว็บไซต์ตามวัตถุประสงค์ที่ได้กำหนดไว้ และมีผลการดำเนินงานดังนี้

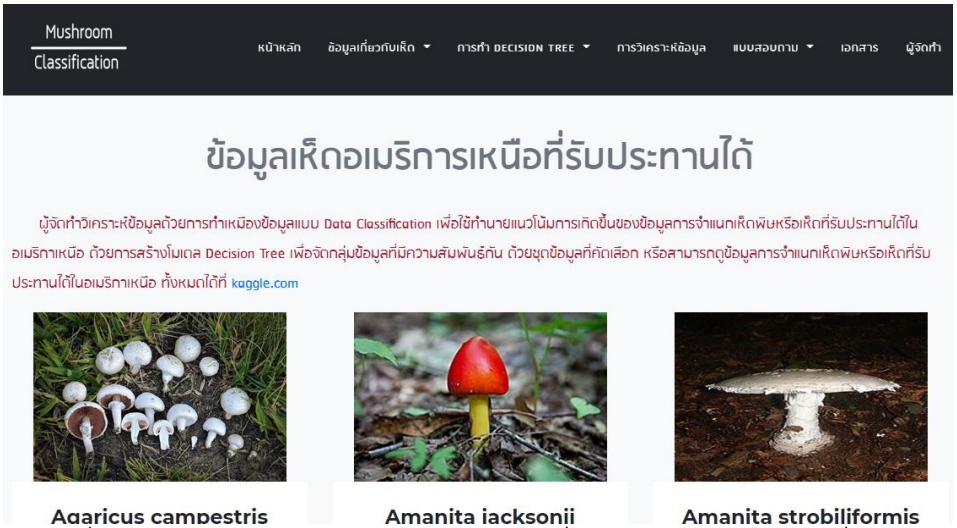
ภาพที่ 4. 1 หน้าแรกของเว็บไซต์

คำอธิบายการใช้งาน : แถบเมนูสามารถใช้งานได้ตามความต้องการของการใช้งาน ซึ่งเป็นหน้าสำหรับผู้ใช้งานทั่วไป สามารถเลือกดู และศึกษาเข้าไปใช้งานในส่วนของเว็บไซต์ได้



ภาพที่ 4. 2 แสดงหน้าเมนูการวิเคราะห์ข้อมูลแบบ Visualization

คำอธิบายการใช้งาน : เป็นหน้าที่แสดงการนำข้อมูลที่คลีนเสร็จแล้วมาวิเคราะห์

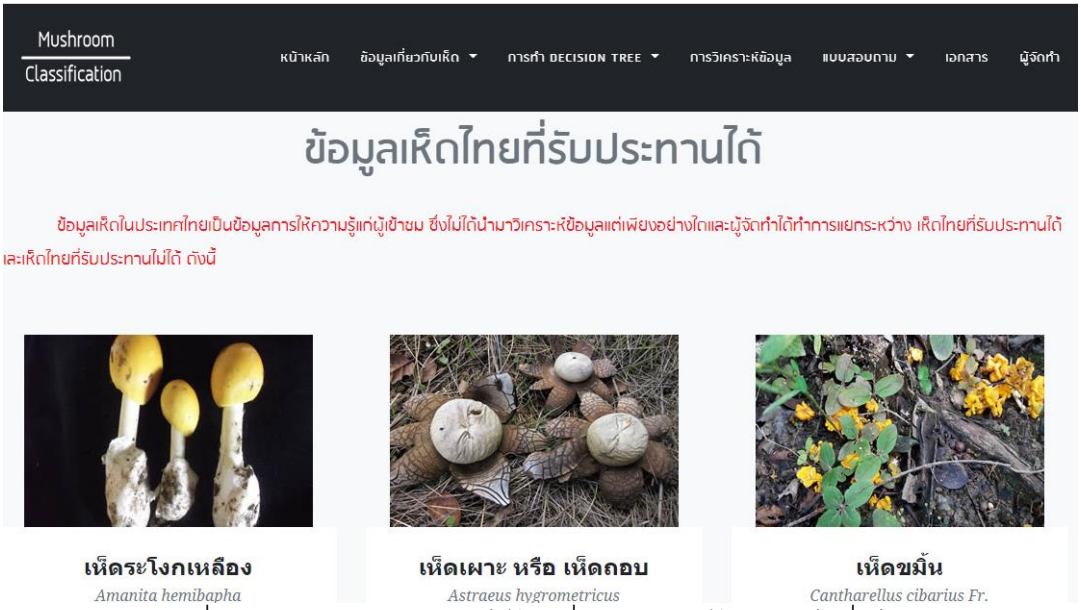


ภาพที่ 4. 3 แสดงข้อมูลของเห็ดอเมริกาเหนือที่รับประทานได้และเห็ดที่เป็นพิษ

คำอธิบายการใช้งาน : เป็นหน้าแสดงข้อมูลเกี่ยวกับเห็ดอเมริกา*เหนือ*ที่รับประทานได้และเห็ดพิษ สามารถกดเข้าดูตามรูปภาพได้

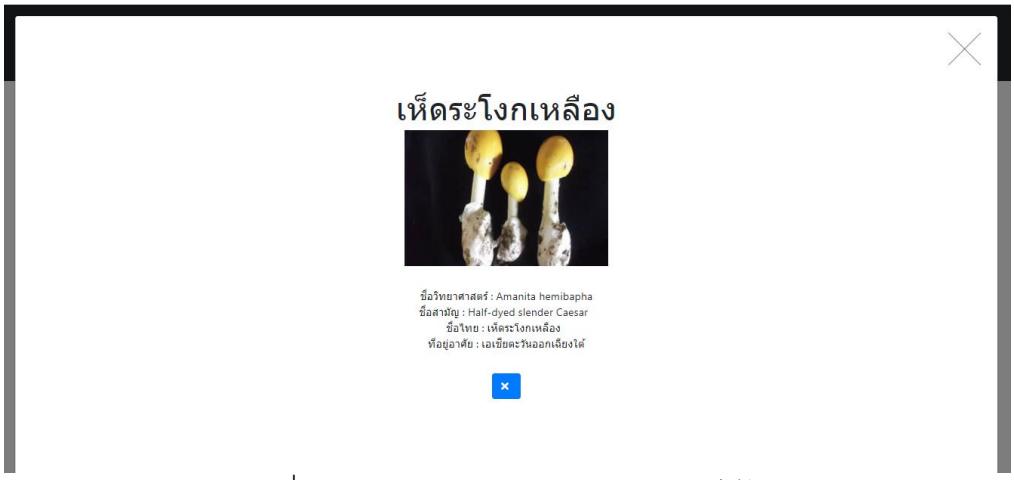
ภาพที่ 4. 4 แสดงรายละเอียดข้อมูลของเห็ดอเมริกาเหนือ

คำอธิบายการใช้งาน : เมื่อกดจากรูปข้อมูลของเห็ดก็จะปรากฏชื่อวิทยาศาสตร์ ชื่อสามัญ และการกระจายหรือที่อยู่ของเห็ดว่ามักพบเจอในทวีปใด



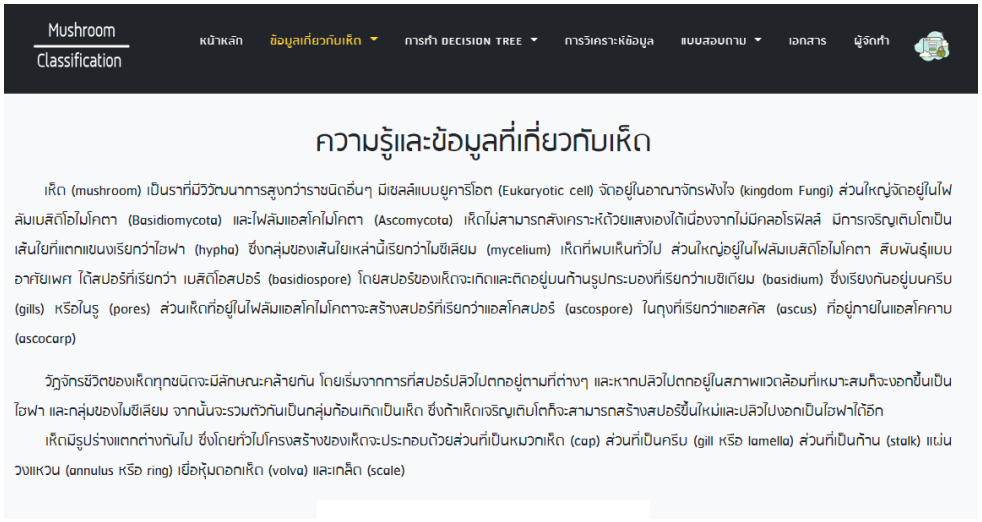
ภาพที่ 4. 5 แสดงข้อมูลของเห็ดไทยที่รับประทานได้และเห็ดที่เป็นพิษ

คำอธิบายการใช้งาน : เป็นหน้าแสดงข้อมูลเกี่ยวกับเห็ดไทยที่รับประทานได้และเห็ดที่เป็นพิษ สามารถกดเข้าดูตามรูปภาพได้



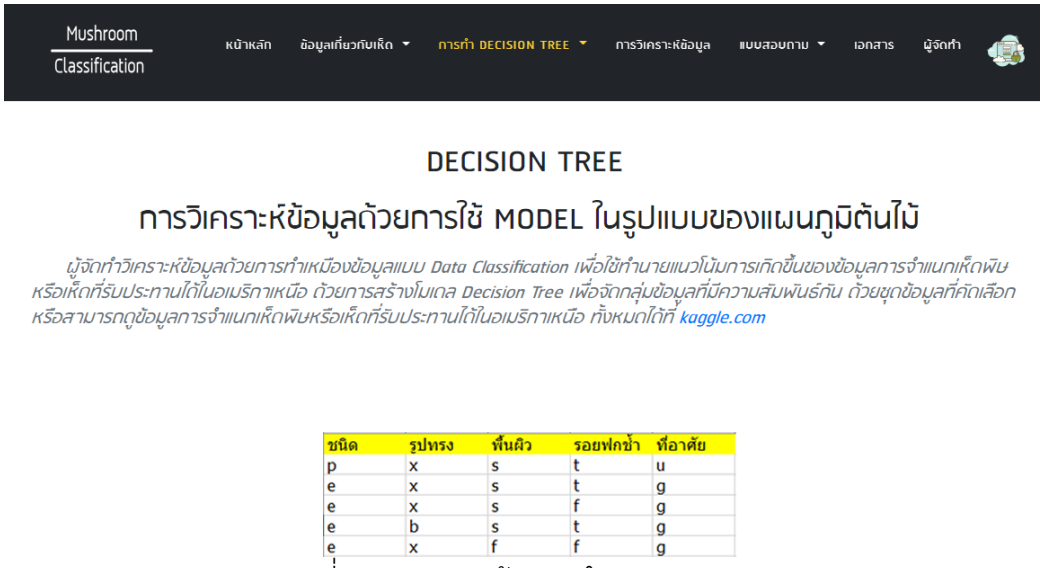
ภาพที่ 4. 6 แสดงรายละเอียดข้อมูลของเห็ดไทย

คำอธิบายการใช้งาน : เมื่อกดจากรูปข้อมูลของเห็ดก็จะปรากฏชื่อวิทยาศาสตร์ ชื่อสามัญ ชื่อไทย และการกระจายหรือที่อยู่ของเห็ดว่ามักพบเจอในทวีปใด



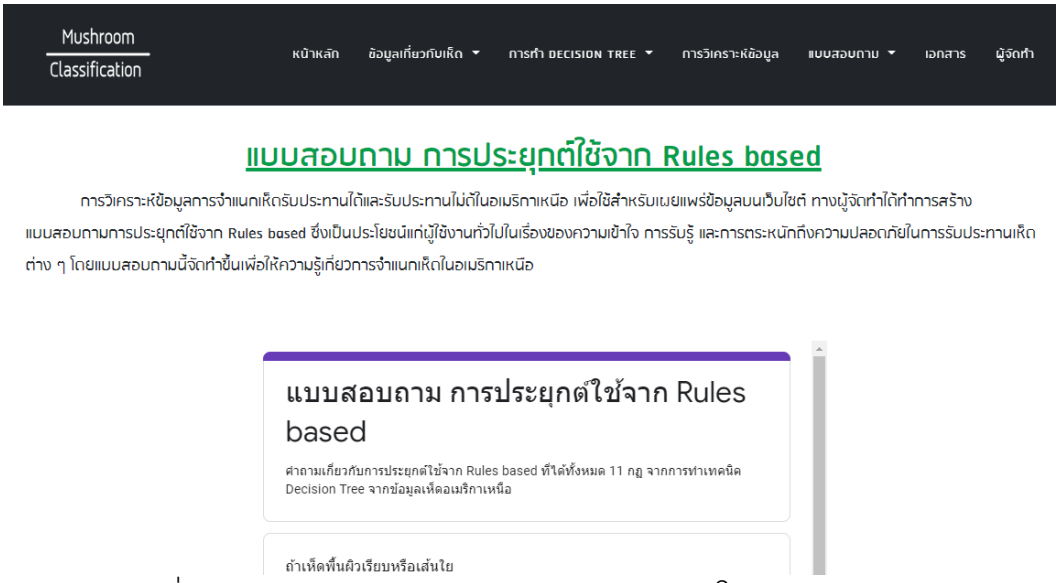
ภาพที่ 4. 7 แสดงความรู้และข้อมูลที่เกี่ยวกับเห็ด

คำอธิบายการใช้งาน : เป็นหน้าที่ให้ความรู้เกี่ยวกับเห็ดที่รับประทานได้และเห็ดพิษว่ามีลักษณะแบบใด



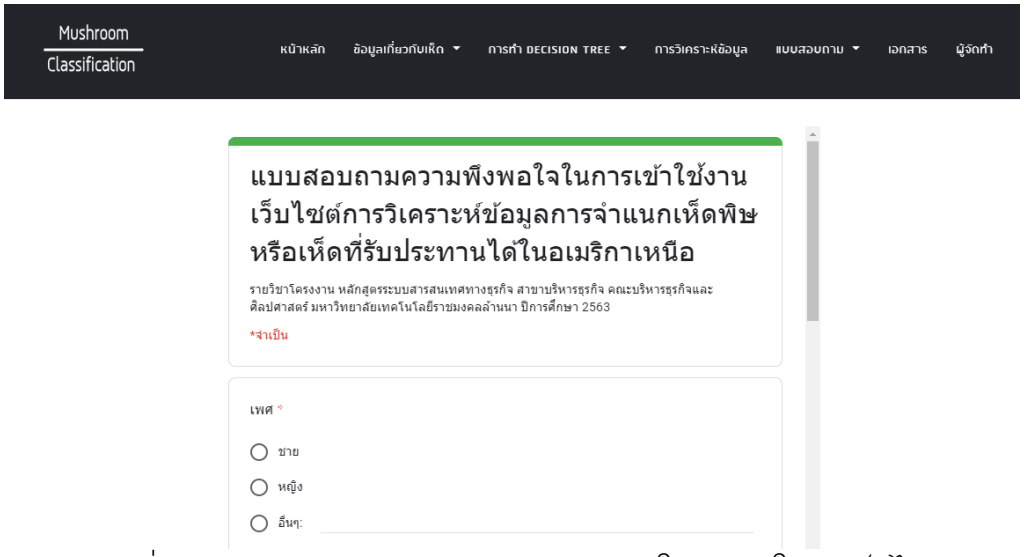
ภาพที่ 4. 8 แสดงหน้าการทำDecision three

คำอธิบายการใช้งาน : เป็นหน้าแสดงการอธิบายขั้นตอนในการสร้างโมเดล Decision Tree ซึ่งจะประกอบไปด้วยการคำนวณต่าง ๆ



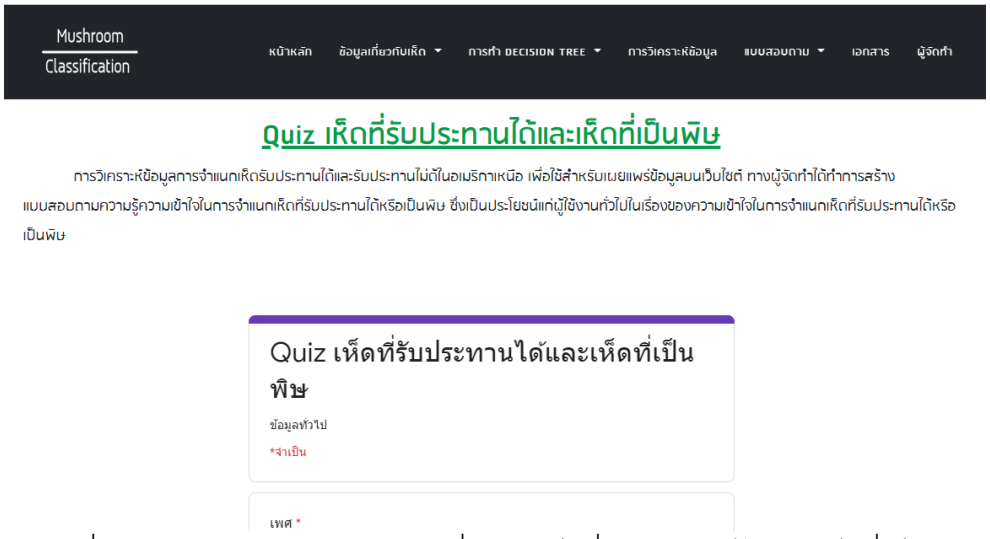
ภาพที่ 4. 9 แสดงหน้าทำแบบสอบถามการประยุกต์ใช้จาก Rules based

คำอธิบายการใช้งาน : เป็นหน้าแสดงการทำแบบสอบถามจาก Rules based ที่ได้จากกฎ 11 ข้อ ผู้ใช้งานสามารถทำแบบสอบถามตามที่เข้าใจ



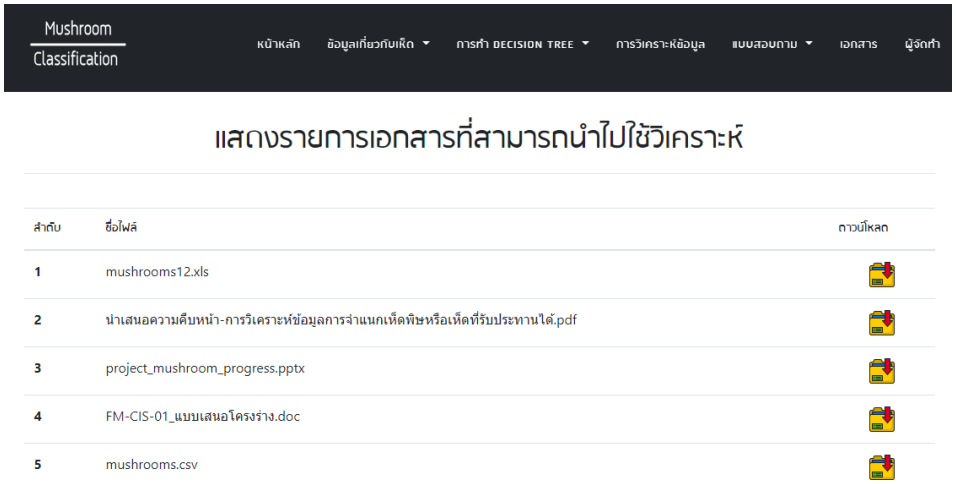
ภาพที่ 4. 10 แสดงหน้าแบบสอบถามความพึงพอใจการเข้าใช้งานเว็บไซต์

คำอธิบายการใช้งาน : เป็นหน้าแสดงการทำแบบสอบถามความพึงพอใจการเข้าใช้งานเว็บไซต์ ผู้ใช้งานสามารถทำแบบสอบถามตามที่พึงพอใจ



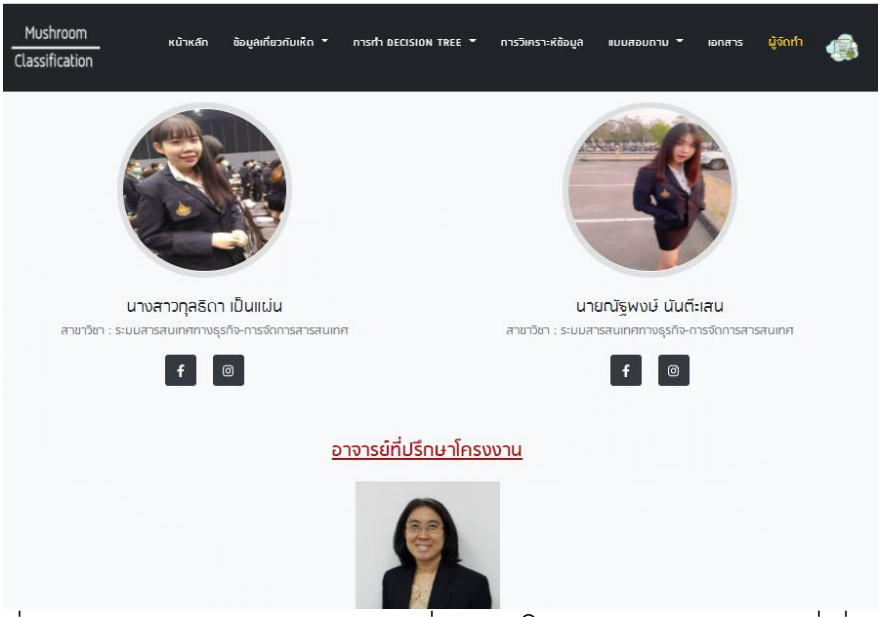
ภาพที่ 4. 11 แสดงหน้าแบบทดสอบเกี่ยวกับเห็ดที่รับประทานได้และเห็ดที่เป็นพิษ

คำอธิบายการใช้งาน : เป็นหน้าแสดงการทำแบบทดสอบเกี่ยวกับเห็ดที่รับประทานได้และเห็ดที่เป็นพิษ ผู้ใช้งานสามารถทำแบบทดสอบถามตามการรับรู้ข้อมูลภายในเว็บไซต์



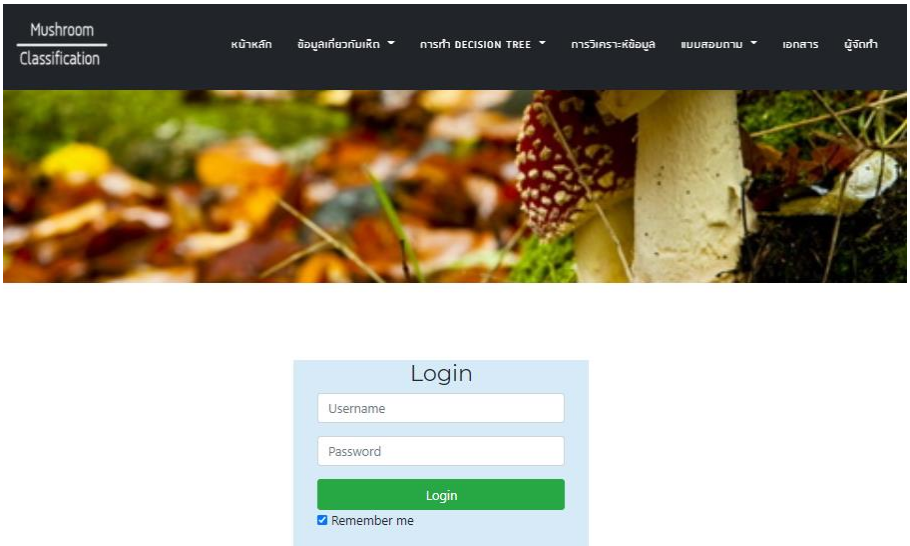
ภาพที่ 4. 12 แสดงเอกสารที่สามารถนำไปใช้วิเคราะห์ได้

คำอธิบายการใช้งาน : เป็นหน้าแสดงการรายการเอกสารที่ผู้วิเคราะห์ใช้ในการทำDecision tree ผู้ใช้งานสามารถนำไปใช้วิเคราะห์ได้



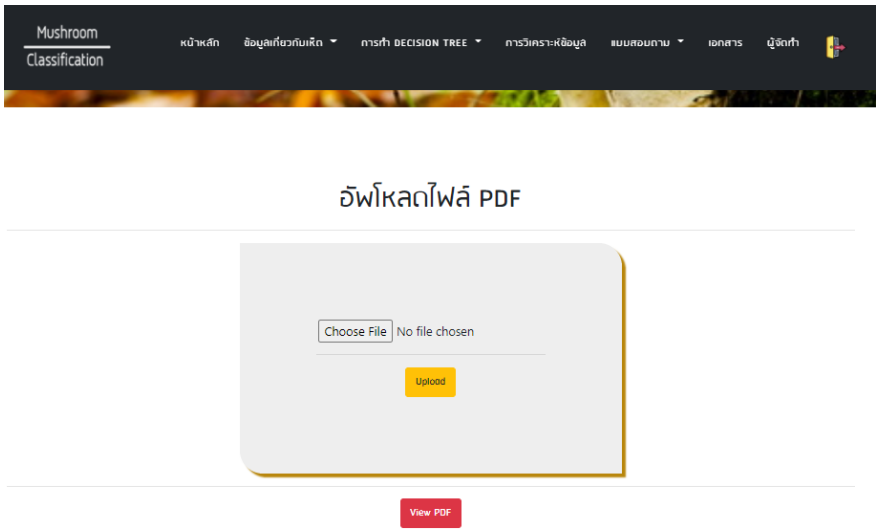
ภาพที่ 4. 13 แสดงหน้าผู้จัดทำ อาจารย์ที่ปรึกษาโครงงาน และหน่วยงานที่เกี่ยวข้อง

คำอธิบายการใช้งาน : เป็นหน้าแสดงข้อมูลติดต่อผู้จัดทำ อาจารย์ที่ปรึกษาโครงงาน และหน่วยงานที่เกี่ยวข้องกับโครงงานชิ้นนี้



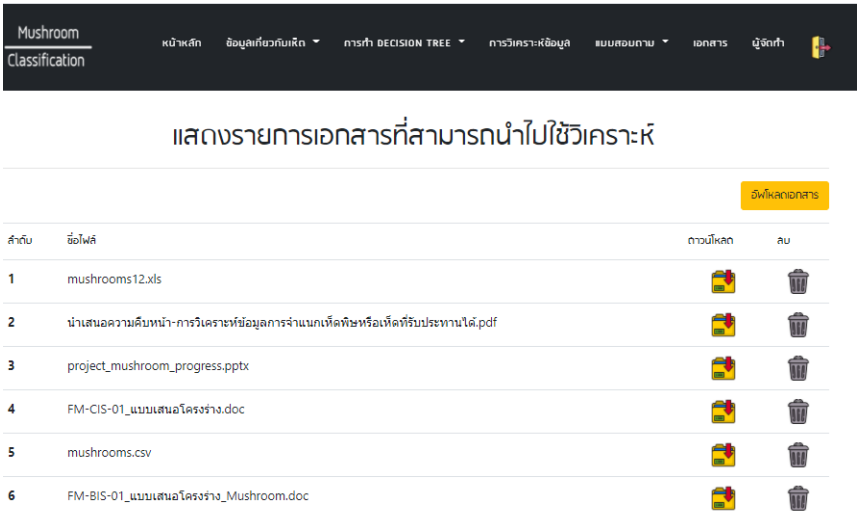
ภาพที่ 4. 14 แสดงหน้าล็อกอินของผู้วิเคราะห์

คำอธิบายการใช้งาน : เป็นหน้าล็อกอินสำหรับผู้วิเคราะห์เท่านั้น



ภาพที่ 4. 15 แสดงหน้าอัปโหลดไฟล์

คำอธิบายการใช้งาน : เป็นหน้าผู้วิเคราะห์ใช้สำหรับอัปโหลดไฟล์



ภาพที่ 4. 16 แสดงหน้ารายการเอกสารที่อัปโหลด

คำอธิบายการใช้งาน : เป็นหน้าแสดงรายการอัปโหลดไฟล์ สามารถดาวน์โหลดได้ และสามารถลบเอกสารที่ต้องการลบได้

4.2 การอธิปรายผล

จากวัตถุประสงค์เพื่อวิเคราะห์ข้อมูลการจำแนกเห็ดพิษหรือเห็ดที่รับประทานได้ในอเมริกาเหนือ เพื่อเผยแพร่ข้อมูลบนเว็บไซต์ ผู้วิเคราะห์ข้อมูลได้ทำการศึกษาถึงปัญหาเก็บรวบรวมข้อมูลและจัดการกับข้อมูลให้อยู่ในรูปแบบที่พร้อมต่อการวิเคราะห์เพื่อใช้ในวิเคราะห์ผล

4.2.1 อธิปรายผลการดำเนินโครงงาน

ผู้วิเคราะห์ข้อมูลได้ดำเนินการกับข้อมูลในการใช้การวิเคราะห์ข้อมูลตามกระบวนการของกระบวนการวิเคราะห์ข้อมูลของ CRISP-DM หรือ (Cross Industry Standard Process for Data Mining) มาใช้กับข้อมูลเห็ดพิษหรือเห็ดที่รับประทานได้ในอเมริกาเหนือซึ่งผู้วิเคราะห์ได้ดำเนินการตามกระบวน ดังนี้

4.2.1.1 ขอบเขตการวิเคราะห์ข้อมูล

1) รู้จักและเข้าใจในธุรกิจ (Business Understanding)

ผู้วิเคราะห์ข้อมูลทำความเข้าใจกับปัญหาให้อยู่ในรูปของการวิเคราะห์ข้อมูลทาง Data Mining โดยการวิเคราะห์ข้อมูลในประเด็นนี้ คือ การวิเคราะห์ข้อมูลการจำแนกเห็ดพิษหรือเห็ดที่รับประทานได้ในอเมริกาเหนือ ซึ่งมีข้อมูลมาก จำนวนรายการทั้งหมด 8,125 รายการ ทำให้ไม่สามารถทำความเข้าใจกับข้อมูลที่มีความซับซ้อนได้อย่างรวดเร็ว เช่น ต้องการทราบว่าลักษณะคุณสมบัติพื้นผิวเห็ดแบบใดที่บ่งบอกถึงเห็ดพิษมากที่สุด

2) จัดเก็บและรวบรวมข้อมูลให้ครบ (Data Understanding)

ผู้วิเคราะห์ข้อมูลทำการรวบรวมข้อมูล เพื่อตรวจสอบรายละเอียด ปริมาณ และความน่าเชื่อถือของข้อมูลการจำแนกเห็ดพิษหรือเห็ดที่รับประทานได้ในอเมริกาเหนือ ที่ได้จากเว็บไซต์ kaggle.com ซึ่งเป็นเว็บไซต์ที่เก็บรวบรวมชุดข้อมูลต่าง ๆ เป็นแหล่งรวม Datasets หรือ ชุดข้อมูล สำหรับฝึกสอน Machine Learning ที่ใหญ่ที่สุดในโลกแห่งหนึ่ง มีข้อมูลทุกประเภท ไม่ว่าจะเป็น Datasets ในหมวดหมู่ Finance, Business, Physics, Biology, Sports, News ซึ่งเป็นข้อมูลที่เปิดเผยได้ เพื่อให้ผู้ใช้บริการสามารถนำชุดข้อมูลไปศึกษาหรือวิเคราะห์ให้เกิดประโยชน์ต่อไป

ซึ่งข้อมูลการจำแนกเห็ดพิษหรือเห็ดที่รับประทานได้ในอเมริกาเหนือ มีจำนวนข้อมูล 8,125 รายการ ประกอบด้วย 23 แอตทริบิวท์ ประกอบด้วย ประเภท รูปทรงของหมวกเห็ด พื้นผิวของหมวกเห็ด สีของหมวกเห็ด รอยช้ำ กลิ่น การติดของครีบ ระยะห่างของครีบ ขนาดของครีบ สีของครีบ รูปร่างของก้าน ราก พื้นผิวด้านบนวงแหวน พื้นผิวด้านล่างวงแหวน สีวงแหวน

ข้างต้น สีด้านล่างวงแหวน ชนิดเยื่อหุ้ม สีเยื่อหุ้ม จำนวนวงแหวน ประเภทวงแหวน สีสปอร์ จำนวน การกระจายพันธุ์ ที่อยู่อาศัย

3) เตรียมข้อมูลให้พร้อมใช้งาน (Data Preparation)

ขั้นตอนการแปลงข้อมูลที่ได้รวบรวมมาและเลือกไว้ ให้อยู่ในรูปแบบที่พร้อมสำหรับนำไปวิเคราะห์ในขั้นตอนต่อไปได้ โดยการทำให้เป็นข้อมูลที่ถูกต้อง (Data cleaning)

4) สร้างแบบจำลอง (Modeling)

ผู้วิเคราะห์ข้อมูลวิเคราะห์ข้อมูลด้วยการทำเหมืองข้อมูลแบบ Data Classification เพื่อใช้ทำนายแนวโน้มการเกิดขึ้นของปัจจัยที่สามารถทำให้รู้ได้ว่าเห็ดชนิดไหนเป็นพิษและไม่เป็นพิษรับประทานได้หรือรับประทานไม่ได้ จากเทคนิคการจำแนกประเภทข้อมูล ด้วยการสร้างโมเดล Decision Tree เพื่อจัดกลุ่มข้อมูลที่มีความสัมพันธ์กัน โดยใช้โปรแกรมที่ใช้ทำเหมืองข้อมูล ด้วยชุดข้อมูลที่คัดเลือก ประกอบด้วย 5 แอตทริบิวท์ คือ ประเภท รูปทรง พื้นผิว รอยฟกช้ำ ที่อาศัย

5) การประเมินผล (Evaluation)

ผู้วิเคราะห์ได้ทำการทดลองโมเดล เพื่อวัดประสิทธิภาพที่ตรงกับความต้องการ ซึ่งการวัดประสิทธิภาพด้วยวิธี Self-Consistency Test เหมาะสำหรับใช้ในการทดสอบประสิทธิภาพเพื่อดูแนวโน้มของโมเดลที่สร้างขึ้น และเมื่อนำข้อมูลมาทดสอบ (Testing data) กับโปรแกรมที่ผู้วิเคราะห์เลือกมาทดสอบกับข้อมูลที่ผ่านการวิเคราะห์ข้อมูลด้วยเทคนิค Data Mining จากการสร้างโมเดล Decision Tree จึงนำข้อมูลดังกล่าว มาทดสอบกับโปรแกรม Weka เวอร์ชัน 3.8.4 พบว่าเทคนิค Decision Tree: J48 ให้ผลลัพธ์การจำแนก ประเภทเห็ด รับประทานได้(e) และรับประทานไม่ได้(p) มีความถูกต้องถึง 91.35% ให้ผลลัพธ์ของกฎที่สามารถทำนายได้จำนวน 11 กฎ ดังนี้

กฎข้อที่ 1 IF surface=s Then habitat=u หมายความว่า ถ้าพื้นผิวเป็นประเภทผิวเรียบและอาศัยอยู่ในเมือง ผลการพิจารณาพบว่า เห็ดชนิดนี้รับประทานไม่ได้

กฎข้อที่ 2 IF surface=s Then habitat=g Then bruises=t Then shape=xหมายความว่า ถ้าพื้นผิวเป็นประเภทผิวเรียบและอาศัยอยู่ในทุ่งหญ้าและมีรอยฟกช้ำและมีรูปทรงนูน ผลการพิจารณาพบว่า เห็ดชนิดนี้รับประทานไม่ได้

กฎข้อที่ 3 IF surface=s Then habitat=g Then bruises=t Then shape=bหมายความว่า ถ้าพื้นผิวเป็นประเภทผิวเรียบและอาศัยอยู่ในทุ่งหญ้าและมีรอยฟกช้ำและมีรูปทรงระฆัง ผลการพิจารณาพบว่า เห็ดชนิดนี้รับประทานได้

กฎข้อที่ 4 IF surface=s Then habitat=g Then bruises=t Then shape=f หมายความว่า ถ้าพื้นผิวเป็นประเภทผิวเรียบและอาศัยอยู่ในทุ่งหญ้าและมีรอยฟกช้ำและมีรูปทรงแบน ผลการพิจารณาพบว่า เห็ดชนิดนี้รับประทานไม่ได้

กฎข้อที่ 5 IF surface=s Then habitat=g Then bruises=f หมายความว่า ถ้าพื้นผิวเป็นประเภทผิวเรียบและอาศัยอยู่ในทุ่งหญ้าและไม่มีรอยฟกช้ำ ผลการพิจารณาพบว่า เห็ดชนิดนี้รับประทานได้

กฎข้อที่ 6 IF surface=s Then habitat=d Then bruises=t หมายความว่า ถ้าพื้นผิวเป็นประเภทผิวเรียบและอาศัยอยู่ในป่าและมีรอยฟกช้ำ ผลการพิจารณาพบว่าเห็ดชนิดนี้รับประทานได้

กฎข้อที่ 7 IF surface=s Then habitat=d Then bruises=f หมายความว่า ถ้าพื้นผิวเป็นประเภทผิวเรียบและอาศัยอยู่ในป่าและไม่มีรอยฟกช้ำ ผลการพิจารณาพบว่าเห็ดชนิดนี้รับประทานไม่ได้

กฎข้อที่ 8 IF surface=f Then bruises=t หมายความว่า ถ้าพื้นผิวเป็นประเภทผิวเส้นใยและมีรอยฟกช้ำ ผลการพิจารณาพบว่าเห็ดชนิดนี้รับประทานได้

กฎข้อที่ 9 IF surface=f Then bruises=f Then habitat=u หมายความว่า ถ้าพื้นผิวเป็นประเภทผิวเส้นใยและไม่มีรอยฟกช้ำและอาศัยอยู่ในเมือง ผลการพิจารณาพบว่าเห็ดชนิดนี้รับประทานได้

กฎข้อที่ 10 IF surface=f Then bruises=f Then habitat=g หมายความว่า ถ้าพื้นผิวเป็นประเภทผิวเส้นใยและไม่มีรอยฟกช้ำและอาศัยอยู่ในทุ่งหญ้า ผลการพิจารณาพบว่าเห็ดชนิดนี้รับประทานได้

กฎข้อที่ 11 IF surface=f Then bruises=f Then habitat=d หมายความว่า ถ้าพื้นผิวเป็นประเภทผิวเส้นใยและไม่มีรอยฟกช้ำและอาศัยอยู่ในป่า ผลการพิจารณาพบว่าเห็ดชนิดนี้รับประทานไม่ได้

6) เผยแพร่ผลวิเคราะห์ (Deployment)

ผู้วิเคราะห์ข้อมูลนำผลข้อมูลที่ทำการวิเคราะห์แล้ว แสดงผลข้อมูลบน Web Browser โดยใช้ชุดคำสั่ง HTML และ CSS3 ร่วมกับการนำเสนอข้อมูลแบบ visualization ด้วยการแสดงผลข้อมูลในรูปแบบของภาพโดยใช้โปรแกรม Tableau Public เป็นรูปแบบรายงานหรือ แผนภาพ (Dashboard) โดยผู้วิเคราะห์ได้ทำการศึกษาความรู้พื้นฐานทางด้านภาษาที่ใช้

ในการพัฒนาเว็บไซต์ คือ HTML PHP JavaScript และ CSS เมื่อศึกษาความรู้พื้นฐานในภาษาต่าง ๆ และการวิเคราะห์ข้อมูลด้วยเทคนิคดังที่กล่าวมาข้างต้นแล้วจึงทำการวิเคราะห์ข้อมูล และออกแบบเว็บไซต์โดยเว็บไซต์มีส่วนประกอบต่าง ๆ ดังนี้

1) ส่วนเนื้อหาเกี่ยวกับเห็ดพิษและเห็ดที่รับประทานได้ในอเมริกาเหนือ

2) ส่วนเนื้อหาเกี่ยวกับเห็ดพิษและเห็ดที่รับประทานได้ในไทย

3) ส่วนการวิเคราะห์ข้อมูล ด้วยDecision Tree

4) ส่วนความรู้และข้อมูลที่เกี่ยวกับเห็ด

5) ส่วนการวิเคราะห์ข้อมูลด้วยการใช้ MODEL

6) ส่วนขั้นตอนการทดสอบข้อมูลกับโปรแกรม WEKA

7) ส่วนผลของการวิเคราะห์ข้อมูล 8) แบบสอบถามแบบสอบถามการประยุกต์ใช้จาก Rules based

9) ส่วนแบบสอบถามความพึงพอใจในการเข้าใช้งานเว็บไซต์การวิเคราะห์ข้อมูลการจำแนกเห็ดพิษหรือเห็ดที่รับประทานได้ในอเมริกาเหนือ

10) ส่วนแบบทดสอบความรู้เกี่ยวกับเห็ด

11) ส่วนเอกสาร

4.2.1.2 ขอบเขตของผู้ใช้งานทั่วไปบนเว็บเบราเซอร์

1) ผู้ใช้งานทั่วไปสามารถดูข้อมูลสารสนเทศของการจำแนกเห็ดพิษหรือเห็ดที่รับประทานได้ในอเมริกาเหนือ ที่เผยแพร่ผ่านทางเว็บไซต์

2) ผู้ใช้งานทั่วไปสามารถดูข้อมูลสารสนเทศในลักษณะตารางสรุปข้อมูลบ่งชี้ถึงเห็ดพิษและเห็ดที่รับประทานได้

3) ผู้ใช้งานทั่วไปสามารถดูข้อมูลสารสนเทศในลักษณะรูปแบบของแผนภูมิจากโปรแกรม Tableau Public ได้

4) ผู้ใช้งานทั่วไปสามารถกรองข้อมูลสารสนเทศผ่านระบบของ Tableau Public บน Web Browserได้

คณะผู้จัดทำเพิ่มส่วนงานที่สำคัญและเป็นประโยชน์ต่อโครงงาน นอกเหนือไปจากขอบเขตโครงงานที่ได้เสนอไว้ดังนี้

1) ส่วนของแบบฟอร์มการประยุกต์ใช้จากRule Base ซึ่งจะแสดงให้เห็นถึงลักษณะของเห็ดพิษหรือเห็ดที่รับประทานได้ในอเมริกาเหนือ

2) ส่วนของแบบทดสอบความรู้เกี่ยวกับเห็ดจัดทำขึ้นเพื่อเก็บข้อมูลของผู้เข้าชมเว็บไซต์เพื่อเป็นการทบทวนความเข้าใจของผู้เข้าชมเว็บไซต์ว่ามีความรู้เกี่ยวกับเห็ดมากเพียงใด

4.2.2 การอธิปรายผลความพึงพอใจจากการตอบแบบสอบถาม

จากผลการวิเคราะห์ข้อมูลการจำแนกเห็ดพิษหรือเห็ดที่รับประทานได้ในอเมริกาเหนือ และ การจัดทำเว็บไซต์เพื่อเผยแพร่ข้อมูล แล้วจากนั้นทางผู้วิเคราะห์ข้อมูลได้ทำการประเมินผลการใช้งานเว็บไซต์จากผู้ใช้งานทั่วไป จากกลุ่มตัวอย่างจำนวน 65 คน ผู้วิเคราะห์ได้แบ่งการประเมินเป็น 3 ตอน ซึ่งมีรายละเอียดดังนี้

1) ข้อมูลทั่วไปของผู้ใช้งานเว็บไซต์

2) ข้อมูลเกี่ยวกับความพึงพอใจในการใช้บริการเว็บไซต์

3) ข้อเสนอแนะ

ในตอนที่ 1 ข้อมูลทั่วไปของผู้ใช้งานเว็บไซต์ ประกอบด้วย

ส่วนข้อมูลทั่วไปจากแบบสำรวจ ประกอบด้วยข้อมูลทั่วไป ได้แก่ เพศ อายุ ระดับการศึกษา โดยสามารถสรุปข้อมูล จากกลุ่มตัวอย่าง 65 คน แบ่งเป็น

เพศ		
ชาย	จำนวน 21 คน	คิดเป็นร้อยละ 32.3
หญิง	จำนวน 44 คน	คิดเป็นร้อยละ 67.7
อายุ		
ต่ำกว่า 20 ปี	จำนวน 24 คน	คิดเป็นร้อยละ 36.9
20–30 ปี	จำนวน 34 คน	คิดเป็นร้อยละ 52.3
31–40 ปี	จำนวน 6 คน	คิดเป็นร้อยละ 9.2
41–50 ปี	จำนวน 1 คน	คิดเป็นร้อยละ 1.5
51–60 ปี	จำนวน 0 คน	คิดเป็นร้อยละ 0
มากกว่า 60 ปี	จำนวน 0 คน	คิดเป็นร้อยละ 0
ระดับการศึกษา		
ต่ำกว่าปริญญาตรี	จำนวน 39 คน	คิดเป็นร้อยละ 60
ปริญญาตรี	จำนวน 24 คน	คิดเป็นร้อยละ 36.9
ปริญญาโท	จำนวน 2 คน	คิดเป็นร้อยละ 3.1

สูงกว่าปริญญาโท จำนวน 0 คน คิดเป็นร้อยละ 0

ในตอนที่ 2 ข้อมูลเกี่ยวกับความพึงพอใจในการใช้บริการเว็บไซต์มากน้อยเพียงใด

1) ด้านการออกแบบและการจัดรูปแบบ (Design)

2) ด้านคุณภาพข้อมูล (Information Quality)

3) ด้านประโยชน์และการนำไปใช้ (Perceived Usefulness)

การประเมินความพึงพอใจของผู้ตอบแบบสอบถามในภาพรวมด้วยค่าเฉลี่ยคะแนนความพึงพอใจ โดยมีคิดคะแนนและเกณฑ์ระดับความพึงพอใจเป็นดังนี้

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

เมื่อนำคำตอบของผู้ตอบแบบสอบถามมาแจกแจงความถี่และหาค่าเฉลี่ย แล้วกำหนดระดับค่าเฉลี่ยความพึงพอใจเป็นดังนี้

(สรุปแบบประเมินความพึงพอใจของผู้รับบริการศูนย์ดำรงธรรมจังหวัดราชบุรี หน้า 4)

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

จากเกณฑ์ดังกล่าว สามารถแปลความหมายของความพึงพอใจได้ดังนี้

คะแนนค่าเฉลี่ย 1.00 – 1.80 หมายถึง ระดับความพึงพอใจน้อยที่สุด

คะแนนค่าเฉลี่ย 1.81 – 2.60 หมายถึง ระดับความพึงพอใจน้อย

คะแนนค่าเฉลี่ย 2.61 – 3.40 หมายถึง ระดับความพึงพอใจปานกลาง

คะแนนค่าเฉลี่ย 3.41 – 4.20 หมายถึง ระดับความพึงพอใจมาก

คะแนนค่าเฉลี่ย 4.21 – 5.00 หมายถึง ระดับความพึงพอใจมากที่สุด

ดังนั้นผู้วิเคราะห์จึงนำค่าคะแนนมาคำนวณทางสถิติ คือ ค่าคะแนนเฉลี่ย ($\overline{\square}$) และค่าส่วนเบี่ยงเบนมาตรฐาน (S.D.) ผลการประเมินแสดงรายละเอียดดังนี้ การประเมินด้านด้านการออกแบบ ดังตารางที่ 4.1 การประเมินด้านการออกแบบและการจัดรูปแบบเว็บไซต์ แสดงดังตารางที่ 4.2 และการประเมินด้านประโยชน์และการนำไปใช้แสดงดังตารางที่ 4.3

ตารางที่ 4. 1 แสดงข้อมูลการประมวลผลด้านการออกแบบ

รายการประเมิน	**จำนวนผู้ตอบแบบสอบถาม**	($\bar{X}$)	**S.D.**	**ผลการประเมิน**
1.ความสวยงาม ความทันสมัย น่าสนใจของหน้าโฮมเพจ	65	4.28	0.60	ดีมาก
2.การจัดรูปแบบในเว็บไซต์ง่ายต่อการอ่านและการใช้งาน	65	4.11	0.62	ดี
3.สีสันในการออกแบบเว็บไซต์มีความเหมาะสม	65	4.23	0.75	ดีมาก
4.เมนูง่ายต่อการใช้งาน	65	4.26	0.64	ดีมาก
5.สีพื้นหลังกับสีตัวอักษรมีความเหมาะสมต่อการอ่าน	65	4.22	0.57	ดีมาก
6.ขนาดตัวอักษร และรูปแบบตัวอักษรอ่านได้ง่ายและสวยงาม	65	4.14	0.75	ดี
7.ภาพกับเนื้อหามีความสอดคล้องกันและสามารถสื่อความหมายได้	65	4.25	0.59	ดีมาก
8.โดยภาพรวมท่านมีความพึงพอใจในการออกแบบเว็บไซต์ในระดับใด	65	4.22	0.67	ดีมาก
รวม		**4.22**	**0.65**	**ดีมาก**

จากตารางที่ 4.1 พบว่าการแสดงความสวยงาม ความทันสมัย น่าสนใจของหน้าโฮมเพจ ค่าเฉลี่ยสูงสุดเท่ากับ 4.28 (S.D. = 0.60) อยู่ในเกณฑ์ระดับดีมาก เมนูง่ายต่อการใช้งาน มีค่าเฉลี่ย 4.26 (S.D. = 0.64) อยู่ในเกณฑ์ระดับดีมาก ภาพกับเนื้อหามีความสอดคล้องกันและสามารถสื่อความหมายได้ มีค่าเฉลี่ย 4.25 (S.D. = 0.59) อยู่ในเกณฑ์ระดับดีมาก สีสันในการออกแบบเว็บไซต์มีความเหมาะสม มีค่าเฉลี่ย 4.23 (S.D. = 0.75) อยู่ในเกณฑ์ระดับดีมาก สีพื้นหลังกับสีตัวอักษรมีความเหมาะสมต่อการอ่าน มีค่าเฉลี่ย 4.22 (S.D. = 0.57) อยู่ในเกณฑ์ระดับดีมาก โดยภาพรวม

ท่านมีความพึงพอใจในการออกแบบเว็บไซต์ในระดับใด มีค่าเฉลี่ย 4.22 (S.D. = 0.67) อยู่ในเกณฑ์ระดับดี ขนาดตัวอักษร และรูปแบบตัวอักษร อ่านได้ง่ายและสวยงาม มีค่าเฉลี่ย 4.14 (S.D. = 0.75) อยู่ในเกณฑ์ระดับดี จากจำนวนผู้ตอบแบบสอบถามทั้งหมด 65 คน ค่าเฉลี่ยเท่ากับ 4.22 (S.D. = 0.65) อยู่ในเกณฑ์ระดับดีมาก

ตารางที่ 4. 2 ตารางแสดงการประมวลผลด้านคุณภาพข้อมูล (Information Quality)

รายการประเมิน	**จำนวนผู้ตอบแบบสอบถาม**	**($\bar{X}$)**	**S.D.**	**ผลการประเมิน**
1. ข้อมูลในเว็บไซต์มีความเกี่ยวเนื่องกับงานของท่าน	65	4.16	0.67	ดี
2. ข้อมูลภายในเว็บไซต์เป็นข้อมูลที่ท่านสามารถทำความเข้าใจได้	65	4.20	0.62	ดี
3. ข้อมูลในแต่ละเมนูของเว็บไซต์ มีความครบถ้วนสมบูรณ์	65	4.27	0.70	ดีมาก
4. ข้อมูลภายในเว็บไซต์มีความน่าเชื่อถือ	65	4.17	0.63	ดี
5. ข้อมูลภายในเว็บไซต์ มีความทันสมัย	65	4.23	0.66	ดีมาก
6. ข้อมูลภายในเว็บไซต์มีความถูกต้อง	65	4.19	0.56	ดี
7.การนำเสนอรูปแบบ Visualization ด้วยโปรแกรม Tableau Public มีความเหมาะสมกับข้อมูล	65	4.20	0.67	ดี
รวม		**4.20**	**0.64**	**ดี**

จากตารางที่ 4.2 พบว่าข้อมูลในแต่ละเมนูของเว็บไซต์ มีความครบถ้วนสมบูรณ์ ค่าเฉลี่ยสูงสุดเท่ากับ 4.27 (S.D. = 0.70) อยู่ในเกณฑ์ระดับดีมาก ข้อมูลภายในเว็บไซต์ มีความทันสมัย มีค่าเฉลี่ย 4.23 (S.D. = 0.66) อยู่ในเกณฑ์ระดับดีมาก ข้อมูลภายในเว็บไซต์เป็นข้อมูลที่ท่านสามารถทำความเข้าใจได้ มีค่าเฉลี่ย 4.20 (S.D. = 0.62) อยู่ในเกณฑ์ระดับดี การนำเสนอรูปแบบ Visualization ด้วยโปรแกรม Tableau Public มีความเหมาะสมกับข้อมูล มีค่าเฉลี่ย 4.20(S.D.=0.67) อยู่ในเกณฑ์ระดับดี ข้อมูลภายในเว็บไซต์มีความถูกต้อง มีค่าเฉลี่ย 4.19 (S.D. = 0.56) อยู่ใน

เกณฑ์ระดับดี ข้อมูลภายในเว็บไซต์มีความน่าเชื่อถือ มีค่าเฉลี่ย 4.17 (S.D. = 0.63) อยู่ในเกณฑ์ระดับดี ข้อมูลในเว็บไซต์มีความเกี่ยวเนื่องกับงานของท่าน มีค่าเฉลี่ย 4.16 (S.D. = 0.67) อยู่ในเกณฑ์ระดับดี

ตารางที่ 4. 3 แสดงการประเมินผลด้านประโยชน์และการนำไปใช้

รายการประเมิน	**จำนวนผู้ตอบแบบสอบถาม**	($\bar{X}$)	**S.D.**	**ผลการประเมิน**
1. เนื้อหามีประโยชน์ต่อผู้ใช้งาน และสามารถนำไปประยุกต์ใช้ได้	65	4.22	0.63	ดีมาก
2. สามารถเป็นแหล่งความรู้ได้	65	4.20	0.60	ดี
3. เป็นแหล่งข้อมูลที่ตรงกับความต้องการของผู้ใช้งาน	65	4.06	0.73	ดี
4. โดยรวมท่านคิดว่าเป็นเว็บไซต์ที่มีประโยชน์	65	4.19	0.64	ดี
รวม		**4.17**	**0.65**	**ดี**

จากตารางที่ 4.3 พบว่า เนื้อหามีประโยชน์ต่อผู้ใช้งาน และสามารถนำไปประยุกต์ใช้ได้ค่าเฉลี่ยสูงสุดเท่ากับ 4.22 (S.D. = 0.63) อยู่ในเกณฑ์ระดับดีมาก สามารถเป็นแหล่งความรู้ได้ มีค่าเฉลี่ย 4.20 (S.D. = 0.60) อยู่ในเกณฑ์ระดับดี โดยรวมท่านคิดว่าเป็นเว็บไซต์ที่มีประโยชน์ มีค่าเฉลี่ย 4.19(S.D. = 0.73) อยู่ในเกณฑ์ระดับดี เป็นแหล่งข้อมูลที่ตรงกับความต้องการของผู้ใช้งาน มีค่าเฉลี่ย 4.06 (S.D. = 0.73) อยู่ในเกณฑ์ระดับดี

4.2.3 การอธิปรายผลจากการตอบแบบสอบถามการประยุกต์ใช้จาก Rules based

จากผลการวิเคราะห์ข้อมูลการจำแนกเห็ดพิษหรือเห็ดที่รับประทานได้ในอเมริกาเหนือ และ การจัดทำเว็บไซต์เพื่อเผยแพร่ข้อมูล แล้วจากนั้นทางผู้วิเคราะห์ข้อมูลได้ทำการประเมินผลการตอบแบบสอบถามการประยุกต์ใช้จาก Rules based พบว่าเป็นไปตามผลลัพธ์ของกฎที่สามารถทำนายได้ 11 กฎ เมื่อผลลัพธ์ออกมาเป็น 1 เห็ดชนิดนี้สามารถรับประทานได้ แล้วเมื่อผลลัพธ์ออกมาเป็น 0 เห็ดชนิดนี้ไม่สามารถรับประทานได้

4.2.4 การอธิปรายผลจากการทำแบบทดสอบความรู้เรื่องเห็ดที่รับประทานได้และเห็ดที่รับประทานไม่ได้

จากผลการวิเคราะห์ข้อมูลการจำแนกเห็ดพิษหรือเห็ดที่รับประทานได้ในอเมริกาเหนือ และ การจัดทำเว็บไซต์เพื่อเผยแพร่ข้อมูล แล้วจากนั้นทางผู้วิเคราะห์ข้อมูลได้ทำการทำแบบทดสอบความรู้เรื่องเห็ดที่รับประทานได้ และเห็ดที่รับประทานไม่ได้จากกลุ่มตัวอย่างจำนวน 77 คน ผู้วิเคราะห์ได้แบ่งการประเมินเป็น 3 ตอน ซึ่งมีรายละเอียดดังนี้

1) ข้อมูลทั่วไปของผู้ใช้ทำแบบทดสอบ

2) แบบสอบถามความรู้ความเข้าใจในการจำแนกเห็ดที่รับประทานได้หรือเห็ดที่รับประทานไม่ได้

ตอนที่ 1 ข้อมูลทั่วไปของผู้ใช้ทำแบบทดสอบ ประกอบด้วย

ส่วนข้อมูลทั่วไปจากแบบสอบถาม ประกอบด้วยข้อมูลทั่วไป ได้แก่ เพศ และ อาชีพ โดยสามารถสรุปข้อมูล จากกลุ่มตัวอย่าง 77 คน แบ่งเป็น

เพศ		
ชาย	จำนวน 42 คน	คิดเป็นร้อยละ 54.5
หญิง	จำนวน 34 คน	คิดเป็นร้อยละ 44.2
อาชีพ		
นักเรียน/นักศึกษา	จำนวน 53 คน	คิดเป็นร้อยละ 68.8
ครู/อาจารย์	จำนวน 9 คน	คิดเป็นร้อยละ 11.7
บุคคลทั่วไป	จำนวน 15 คน	คิดเป็นร้อยละ 19.5

ตอนที่ 2 แบบสอบถามความรู้ความเข้าใจในการจำแนกเห็ดที่รับประทานได้หรือเห็ดที่รับประทานไม่ได้

ส่วนข้อมูลทั่วไปจากแบบสอบถาม ประกอบด้วยคะแนนค่าเฉลี่ยเท่ากับ 6.82 คะแนน ค่ามัธยฐาน 7/10 คะแนน โดยสามารถสรุปข้อมูล จากกลุ่มตัวอย่าง 77 คน แบ่งเป็นช่วง1–10 คะแนน

ผู้ทำแบบทดสอบที่ได้คะแนน 1–5	จำนวน 20 คน	คิดเป็นร้อยละ 25.97
ผู้ทำแบบทดสอบที่ได้คะแนน 6–10	จำนวน 57 คน	คิดเป็นร้อยละ 74.03

4.3 บทสรุป

จากผลการดำเนินงานโครงงานในข้างต้นทั้งหมดนี้ ผู้วิเคราะห์ได้นำข้อมูลจากการวิเคราะห์ข้อมูล และการแสดงผลแบบ Visualization ต่าง ๆ โดยใช้โปรแกรม Tableau Public มาเผยแพร่ให้กับบุคคลภายนอกหรือกลุ่มผู้ใช้ข้อมูลในรูปแบบของเว็บไซต์ ที่จะเป็นแหล่งความรู้ต่าง ๆ เกี่ยวกับเห็ดพิษหรือเห็ดที่รับประทานได้ในอเมริกาเหนือ วิธีการและขั้นตอนในการสร้างโมเดล Decision Tree สำหรับผู้ใช้ทั่วไปที่ต้องการทราบว่าลักษณะแบบใดบ่งบอกถึงเห็ดที่เป็นพิษหรือ เห็ดที่รับประทานได้ อีกทั้งยังมีแบบสอบถามสำหรับผู้เข้าใช้เว็บไซต์ได้ร่วมทำแบบทดสอบเกี่ยวกับเห็ดพิษหรือเห็ดที่รับประทานได้ และส่วนของการอัปโหลดข้อมูล ผู้วิเคราะห์ข้อมูลได้ทำการวิเคราะห์ และนำผลที่ได้จากการวิเคราะห์มาแสดงผ่านหน้าเว็บไซต์ตามวัตถุประสงค์ที่ได้กำหนดไว้จนสำเร็จลุล่วงตามเป้าหมาย เพื่อที่จะได้นำผลการดำเนินงานไปสรุป และเป็นข้อเสนอแนะในบทที่ 5 ต่อไป

บทที่ 5

สรุปผลและข้อเสนอแนะ

การวิเคราะห์ข้อมูลการจำแนกเห็ดพิษหรือเห็ดที่รับประทานได้ในอเมริกาเหนือ เพื่อเผยแพร่ข้อมูลบนเว็บไซต์ ทางผู้จัดทำได้ทำการประเมินผลการใช้งานเว็บไซต์จากผู้ใช้งานทั่วไป เพื่อสรุปผลการจัดทำโครงงาน ข้อกำจัดของเว็บไซต์ ปัญหาและอุปสรรคต่าง ๆ ของการจัดทำโครงงาน และข้อเสนอแนะเพื่อพัฒนาโครงงานต่อไป ดังนี้

5.1 บทสรุปผลการดำเนินโครงงาน

จากการที่ผู้วิเคราะห์ได้ทำการศึกษาและวิเคราะห์ข้อมูลการจำแนกเห็ดพิษหรือเห็ดที่รับประทานได้ในอเมริกาเหนือ โดยใช้เทคนิคการทำเหมืองข้อมูลแบบ Data Classification เพื่อใช้ทำนายแนวโน้มการเกิดขึ้นของปัจจัยที่สามารถทำให้รู้ได้ว่าเห็ดชนิดไหนเป็นพิษและไม่เป็นพิษรับประทานได้หรือรับประทานไม่ได้ จากเทคนิคการจำแนกประเภทข้อมูล ด้วยการสร้างโมเดล Decision Tree เพื่อจัดกลุ่มข้อมูลที่มีความสัมพันธ์กัน แล้วนำมาสร้างเป็นสารสนเทศ เพื่อเผยแพร่ข้อมูลบนเว็บไซต์ ในการพัฒนาเว็บไซต์ คือ HTML PHP JavaScript และ CSS เมื่อศึกษาความรู้พื้นฐานในภาษาต่าง ๆ และการวิเคราะห์ข้อมูลด้วยเทคนิคดังที่กล่าวมาข้างต้นแล้วจึงทำการวิเคราะห์ข้อมูล จนแล้วเสร็จจากนั้นได้ทำการประเมินผลจากการใช้งานจากกลุ่มตัวอย่าง จำนวน 65 คน โดยการออกแบบสอบถามความพึงพอใจในการเข้าใช้งานเว็บไซต์การวิเคราะห์ข้อมูลการจำแนกเห็ดพิษหรือเห็ดที่รับประทานได้ในอเมริกาเหนือ จากกลุ่มตัวอย่างจำนวน 65 คน ผู้วิเคราะห์ได้แบ่งการประเมินเป็น 3 ด้านซึ่งมีรายละเอียดดังนี้

1) ด้านการออกแบบและการจัดรูปแบบ (Design)จากจำนวนผู้ตอบแบบสอบถามทั้งหมด 65 คน ค่าเฉลี่ยเท่ากับ 4.22 (S.D. = 0.65) อยู่ในเกณฑ์ระดับดีมาก

2) ด้านคุณภาพข้อมูล (Information Quality) จากจำนวนผู้ตอบแบบสอบถามทั้งหมด 65 คน ค่าเฉลี่ยเท่ากับ 4.20 (S.D. = 0.64) อยู่ในเกณฑ์ระดับดี

3) ด้านประโยชน์และการนำไปใช้ (Perceived Usefulness)จากจำนวนผู้ตอบแบบสอบถามทั้งหมด 65 คน ค่าเฉลี่ยเท่ากับ 4.17 (S.D. = 0.65) อยู่ในเกณฑ์ระดับดี

ผู้วิเคราะห์ยังได้ทำการประยุกต์ใช้จาก Rules based พบว่า เมื่อทำแบบสอบถามเสร็จแล้ว ผลลัพธ์คะแนนออกมาเป็น 1 นั้นแสดงว่าเห็ดชนิดนี้สามารถรับประทานได้ แล้วเมื่อผลลัพธ์คะแนนออกมาเป็น 0 แสดงว่าเห็ดชนิดนี้ไม่สามารถรับประทานได้ แล้วจากนั้นผู้วิเคราะห์ได้ทำแบบทดสอบความรู้ความเข้าใจในการจำแนกเห็ดที่รับประทานได้หรือเห็ดที่รับประทานไม่ได้ พบว่า ผู้ที่ทำแบบทดสอบผ่านมีจำนวน 57 คน คิดเป็นร้อยละ 74.03 ผู้ที่ทำแบบทดสอบไม่ผ่านมีจำนวน 20 คน คิดเป็นร้อยละ 25.97

5.2 ข้อจำกัดของโครงงาน

5.2.1 ผู้ใช้งานทั่วไปไม่สามารถอัปโหลดชุดข้อมูลบนเว็บไซต์ได้ เนื่องจากเว็บไซต์ใช้เผยแพร่ข้อมูลเพียงอย่างเดียว

5.2.2 ผู้ใช้งานทั่วไปไม่สามารถปรับปรุง แก้ไข เพิ่มข้อมูลบนเว็บไซต์ได้

5.2.3 ข้อมูลอาจไม่ครอบคลุม และข้อมูลอาจไม่ตอบสนองกับความต้องการของผู้เข้ามาใช้งาน เช่น ข้อมูลบนเว็บไซต์อาจมีไม่มากพอ เพราะเนื่องจากเห็ดนั้นมีมากกว่า30,000 สายพันธุ์ ทำให้ผู้วิเคราะห์เลือกเห็ดบางชนิดมาวิเคราะห์

5.2.4 ข้อมูลไม่เป็นปัจจุบัน

5.2.5 ผู้ใช้งานทั่วไปไม่สามารถปรับปรุง แก้ไข กราฟบนเว็บไซต์ได้

5.3 ปัญหาและอุปสรรคของโครงงาน

5.3.1 ชุดข้อมูลที่ไม่ได้เก็บรวบรวมเอง เนื่องจากเป็นชุดข้อมูลแบบเปิด (Open Data Set) จึงจำเป็นต้องใช้เวลาในการศึกษาและทำความเข้าใจในชุดข้อมูลที่หามาได้

5.3.2 ข้อมูลที่ได้มานั้นยังมีความผิดพลาดของข้อมูล จึงต้องทำความสะอาดชุดข้อมูลและแทนค่าชุดข้อมูลบางส่วน จึงทำให้เกิดความล่าช้ากว่าที่ผู้จัดทำกำหนดไว้

5.3.3 การจัดทำเว็บไซต์ บางหน้าหรือบางเนื้อหา เกิดข้อผิดพลาดของโค้ด และฟังก์ชันบางจุด แก้ไขแล้วเกิดการผิดพลาดเนื้อหาหารหรือไม่ตรงตามที่ผู้จัดทำต้องการ จึงเกิดการล่าช้ากว่าที่คาดการณ์ไว้

5.3.4 เนื่องจากผู้วิเคราะห์ไม่ได้ปรึกษาผู้เชี่ยวชาญหรือผู้วิจัยเรื่องเห็ดโดยตรง จึงทำให้ข้อมูลบางอย่างอาจมีความผิดพลาด

5.3.5 โปรแกรมที่ใช้ในการวิเคราะห์มีบางฟังก์ชันที่ต้องใช้ซื้อลิขสิทธิ์จึงจะใช้งานได้

5.4 ข้อเสนอแนะ

การวิเคราะห์ข้อมูลการจำแนกเห็ดพิษหรือเห็ดที่รับประทานได้ในอเมริกาเหนือ เพื่อเผยแพร่ข้อมูลผ่านเว็บไซต์ ที่จัดทำเสร็จสิ้นแล้วแม้ว่าจะสามารถจัดทำและแสดงผลได้ตรงตามวัตถุประสงค์และขอบเขตที่กำหนดไว้ แต่ยังมีอีกหลายข้อจำกัดที่ยังสามารถพัฒนาต่อได้ ซึ่งหากจะมีผู้พัฒนาให้เว็บไซต์เผยแพร่ข้อมูลแสดงผลและทำงานได้อย่างมีประสิทธิภาพมากยิ่งขึ้น ผู้พัฒนาจะต้องปรับปรุงในส่วนต่าง ๆ ดังต่อไปนี้

5.4.1 ควรมีการอัพเดตปรับปรุงข้อมูลแบบ Real time

5.4.2 ควรมีการจัดเก็บข้อมูลด้วยตนเอง จากการปรึกษาผู้เชี่ยวชาญเรื่องเห็ดหรือทางด้านการเกษตร เพื่อให้ตรงกับความต้องการในการวิเคราะห์ข้อมูลในครั้งต่อไป

5.4.3 ควรมีการศึกษาข้อมูลเพิ่มเติมเกี่ยวสารพิษในเห็ดแต่ตระกูล

5.4.4 ข้อมูลควรเป็นเห็ดในประเทศไทย เพื่อการใช้ประโยชน์สูงสุด

5.4.5 ควรจัดทำกราฟจากการวิเคราะห์ให้มีหลากหลายรูปแบบมากขึ้น

บรรณานุกรม

ชฎากัลป์ ชื่นชอบ และ ศรีนวล ตันสุวรรณ. (2560). *ความหลากหลายของเห็ดป่าและราขนาดใหญ่ บริเวณวัดป่านันทวัน บ้านมะค่า ตำบลโพนทอง อำเภอสีดา จังหวัดนครราชสีมา*. วารสารเกษตรพระจอมเกล้า. สืบค้นเมื่อ 1 เมษายน 2563 จากเว็บไซต์ https://li01.tci-thaijo.org/index.php/agritechjournal/article/view/162979

ประชาสัมพันธ์ โรงพยาบาลโพธิ์ไทร. (2560). *เตือนระวังเห็ดพิษ พร้อมวิธีสังเกต*. สืบค้นเมื่อ 30 เมษายน 2563 จากเว็บไซต์ https://bit.ly/2Flwx2l

พรรณพร กุลมา. (2558). *เห็ดพิษและจุดเสี่ยงที่เป็นปัญหาในจังหวัดน่าน* (รายงานผลการวิจัย). น่าน:มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา น่าน

พฤฒิพงศ์ เพ็งศิริ (2559). *การพยากรณ์ความสัมพันธ์ของปัจจัยข้อมูลนักศึกษาที่มีผลต่อระดับผลการเรียนด้วยเทคนิคต้นไม้ตัดสินใจ*. สืบค้นเมื่อ 4 เมษายน 2563 จากเว็บไซต์ https://bit.ly/2IGLtJI

พัฒนพงษ์ ดลรัตน์ และ จารี ทองคำ. (2560). *การเปรียบเทียบประสิทธิภาพของแบบจำลองในการพยากรณ์ความสำเร็จการศึกษาของนักเรียนระดับประกาศนียบัตรวิชาชีพ*. สืบค้นเมื่อ 1 เมษายน 2563 จากเว็บไซต์ https://bit.ly/3kb0EbM

พิชัย ระเวงวัง และ พุธษดี ศิริแสงตระกูล. (2560). *โมเดลเพื่อการพยากรณ์สถานภาพทางการศึกษาของนักศึกษา*. สืบค้นเมื่อ 30 เมษายน 2563 จากเว็บไซต์ https://gsbooks.gs.kku.ac.th/60/nigrc2017/pdf/PMP6.pdf

พีรพล คำพันธ์และคณะ. (2562). *การประยุกต์ใช้เบย์เซียนและการประมวลผลภาพสำหรับการคัดแยกมะม่วง*. สืบค้นเมื่อ 5 เมษายน 2563 จากเว็บไซต์ https://ph02.tci-thaijo.org/index.php/psru-jite/article/view/233295

วทัญญุตา นีลาภาตระกูล และ ชุติมา เบี้ยวไข่มุข. (2562). การศึกษาปัจจัยที่สัมพันธ์กับการตัดสินใจลาออกและการเปรียบเทียบประสิทธิภาพตัวแบบพยากรณ์การลาออกของพนักงาน กรณีศึกษา บริษัทประกันภัย. สืบค้นเมื่อ 1 เมษายน 2563 จากเว็บไซต์ https://bit.ly/2T4Ew73

วันบีลีฟ จำกัด. (2560).*หลักการ ออกแบบเว็บ ขั้นพื้นฐาน พร้อมองค์ประกอบและรูปแบบโครงสร้าง*. สืบค้นเมื่อ 1 เมษายน 2563 จากเว็บไซต์ https://bit.ly/2FCjJ7I

วิกิพีเดีย สารานุกรมเสรี. List of poisonous fungus species. สารานุกรมออนไลน์. สืบค้นเมื่อ 1 เมษายน 2563. จากเว็บไซต์ https://bit.ly/3lVOrbx

บรรณานุกรม(ต่อ)

วีณาวดี ม่วงอ้น และ กรัญญา สิทธิสงวน. (2558). ระบบผู้เชี่ยวชาญสำหรับระบุพันธุ์เห็ดในประเทศไทย. วารสารวิชาการ. สืบค้นเมื่อ 1 เมษายน 2563 จากเว็บไซต์ https://li01.tci-thaijo.org/index.php/ltech/article/view/49054

ศุลิเชษฐ์ ทองกล่ำ. (2560). *เห็ดกินได้และเห็ดพิษ.* สืบค้นเมื่อ 30 เมษายน 2563 จากเว็บไซต์ https://www.hrdi.or.th/Articles/Detail/20

สมาคมนักวิจัยและเพาะเห็ดแห่งประเทศไทย. 2543. *เห็ดพิษ*. บริษัท นิวธรรมดาการพิมพ์ (ประเทศไทย) จำกัด กรุงเทพฯ. 63

อนงค์ จันทร์ศรีกุล, วิรัช ชูบำรุง, กัญจนา โป๊ะเงิน, ดรุณี รัตนประภา และ ประสิทธิ์ ธนากลาง. (2529). *ชนิดของเห็ดพิษ*. รายงานประชุมวิชาการ ครั้งที่ 24 ภาคโปสเตอร์ ณ มหาวิทยาลัยเกษตรศาสตร์ วันที่ 27 – 29 มกราคม 2529. กรุงเทพฯ

9experttraining. *การเลือกรูปแบบ Visualization ให้เหมาะสมกับข้อมูล*. สืบค้นเมื่อ 1 เมษายน 2563 จากเว็บไซต์ https://bit.ly/3iQMkE7

ภาคผนวก

ภาคผนวก ก
คู่มือการใช้งานเว็บไซต์

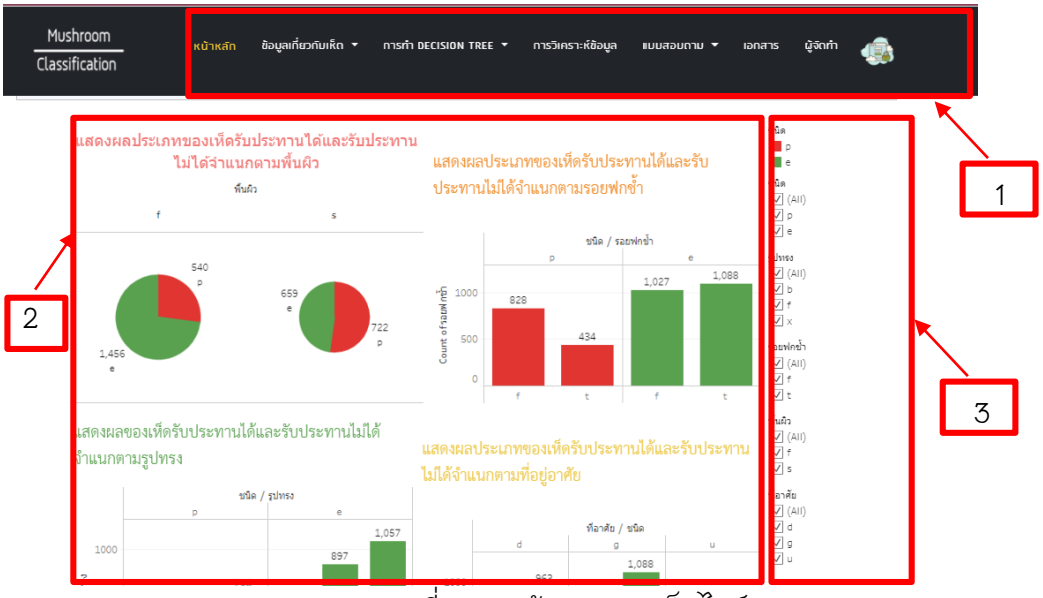
ภาคผนวก ก
คู่มือการใช้งานเว็บไซต์

คู่มือการใช้เว็บไซต์การวิเคราะห์ข้อมูลการจำแนกเห็ดพิษหรือเห็ดที่รับประทานได้ในอเมริกาเหนือ เพื่อเผยแพร่ข้อมูลบนเว็บไซต์

จากการดำเนินงานวิเคราะห์ข้อมูลการจำแนกเห็ดพิษหรือเห็ดที่รับประทานได้ในอเมริกาเหนือ เพื่อเผยแพร่ข้อมูลบนเว็บไซต์ โดยเว็บไซต์มีส่วนใช้งาน 8 ส่วน ดังนี้ ส่วนแสดงผลสรุปการวิเคราะห์การจำแนกเห็ดพิษหรือเห็ดที่รับประทานได้ ส่วนเนื้อหาข้อมูลเห็ดพิษหรือเห็ดที่รับประทานได้ ส่วนการวิเคราะห์ข้อมูลด้วย Decision Tree ส่วนการวิเคราะห์ข้อมูล ส่วนของแบบสอบถาม ส่วนของเอกสารที่เกี่ยวข้อง ส่วนของผู้จัดทำ ส่วนของการเข้าสู่ระบบ ผู้จัดทำจึงได้ทำคู่มืออธิบายการใช้งานเว็บไซต์ ดังต่อไปนี้

คู่มือการใช้สำหรับผู้เข้าใช้งานเว็บไซต์

1) หน้าแรกของเว็บไซต์ เมื่อเข้าสู่เว็บไซต์การวิเคราะห์ข้อมูลการจำแนกเห็ดพิษหรือเห็ดที่รับประทานได้ในอเมริกาเหนือ



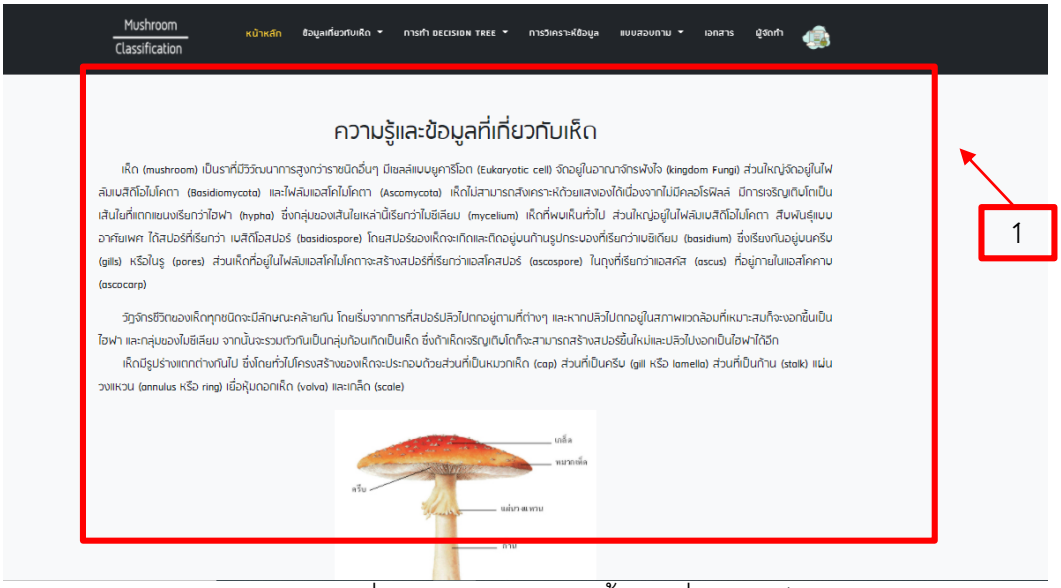
ภาพที่ ก. 1 หน้าแรกของเว็บไซต์

หมายเลข 1 ส่วนของแถบเมนูด้านบนที่ประกอบไปด้วยเมนูการใช้งานเพื่อไปยังหน้าต่าง ๆ ภายในเว็บไซต์

หมายเลข 2 ส่วนของแดชบอร์ดแสดงผล

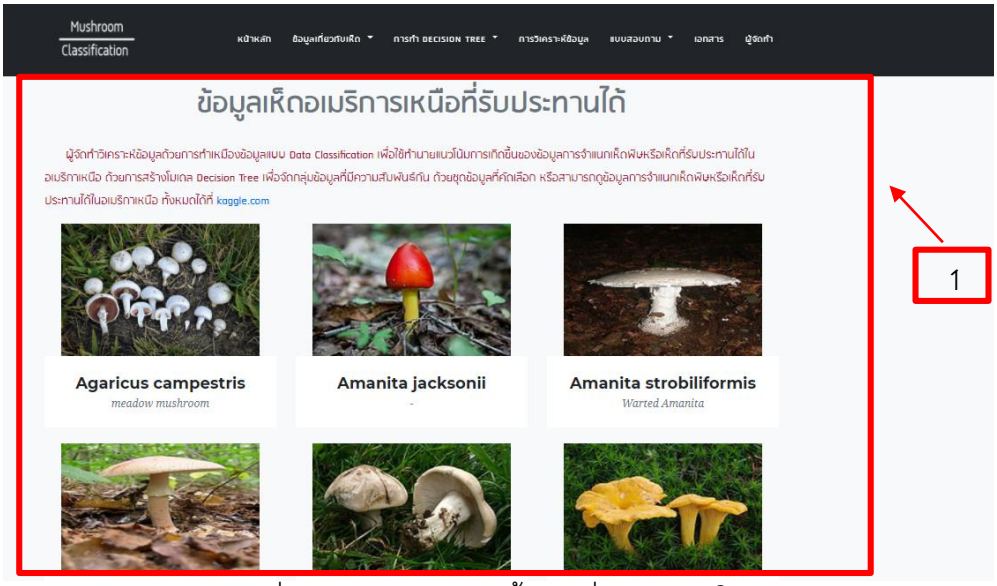
หมายเลข 3 ตัวกรองข้อมูลของแดชบอร์ดแสดงผล ผู้ใช้สามารถใช้ตัวกรองนี้กรองดู ข้อมูลตามความต้องการได้

2) ข้อมูลเนื้อหาเกี่ยวกับเห็ดและข้อมูลเห็ดพิษหรือเห็ดที่รับประทานได้ในประเทศไทยและอเมริกาเหนือ



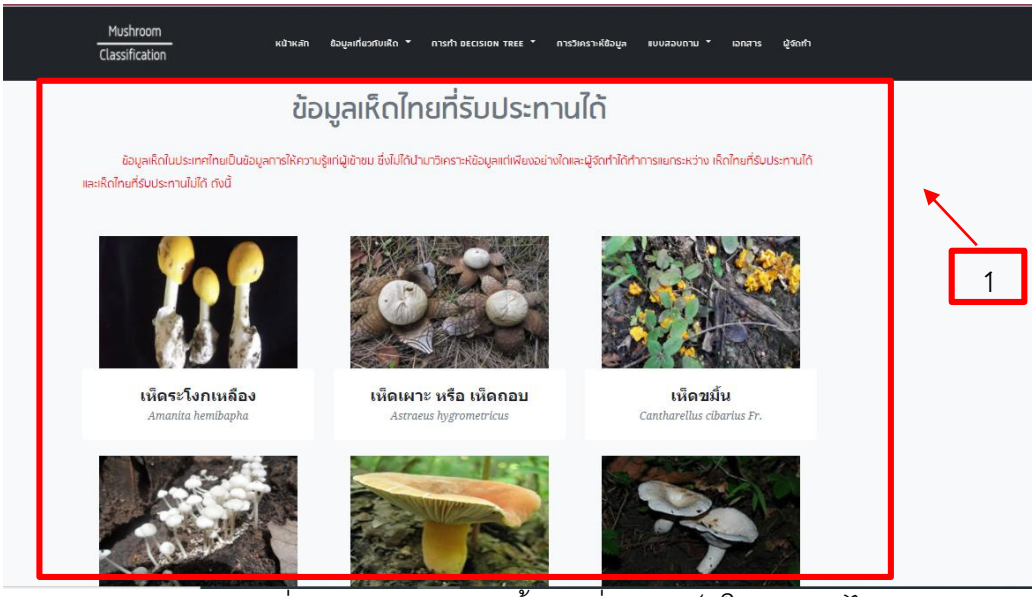
ภาพที่ ก. 2 แสดงข้อมูลเนื้อหาเกี่ยวกับเห็ด

หมายเลข 1 ส่วนของข้อมูลเนื้อหาเกี่ยวกับเห็ด



ภาพที่ ก. 3 แสดงข้อมูลเนื้อหาเกี่ยวกับเห็ดในอเมริกาเหนือ

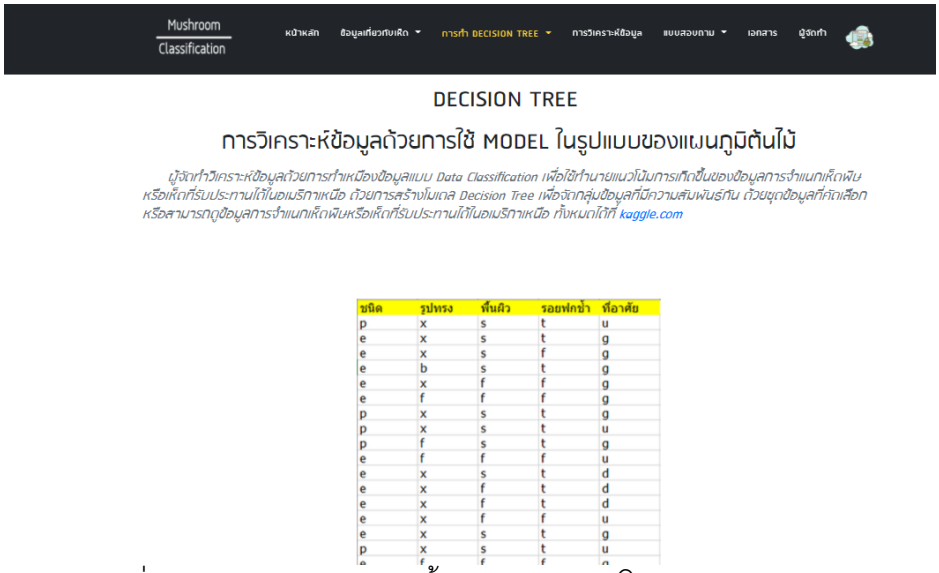
หมายเลข 1 ส่วนของข้อมูลเห็ดพิษหรือเห็ดที่รับประทานได้ในอเมริกาเหนือ



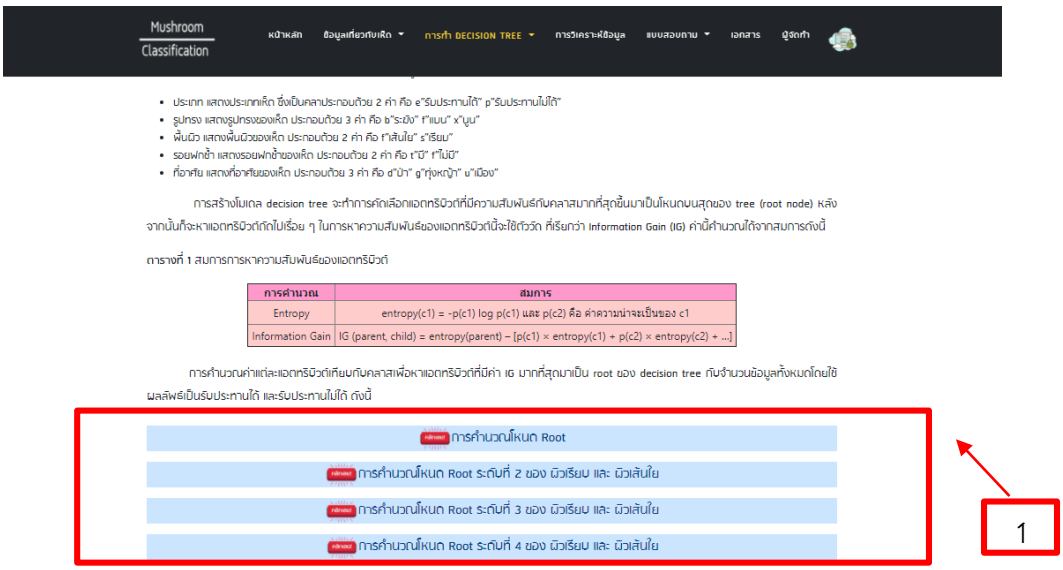
ภาพที่ ก. 4 แสดงข้อมูลเนื้อหาเกี่ยวกับเห็ดในประเทศไทย

หมายเลข 1 ส่วนของข้อมูลเห็ดพิษหรือเห็ดที่รับประทานได้ในประเทศไทย

3) การแสดงรายละเอียด วิธีการ ขั้นตอนและผลการวิเคราะห์ข้อมูลด้วยโมเดล Decision Tree

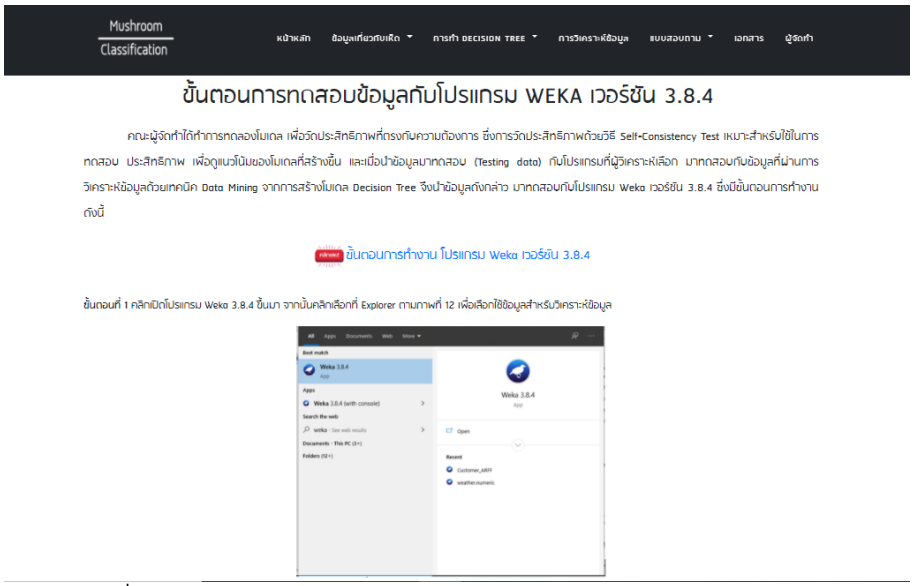


ภาพที่ ก. 5 แสดงการอธิบายขั้นตอนการสร้างโมเดล Decision Tree

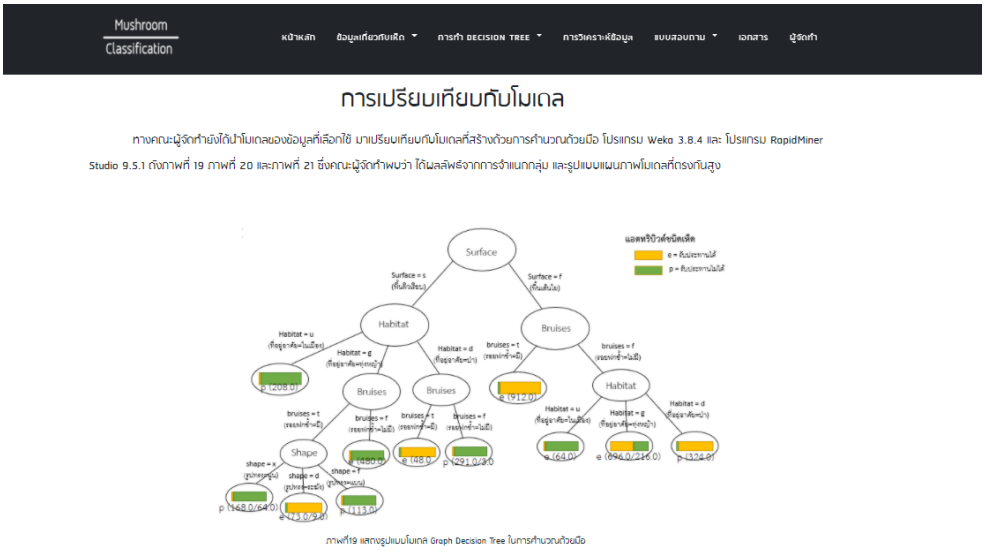


ภาพที่ ก. 6 แสดงขั้นตอนการสร้าง Decision Tree

หมายเลข 1 ปุ่มอธิบายการคำนวณค่าแต่ละแอตทริบิวต์เทียบกับคลาสเพื่อหาแอตทริบิวต์ที่มีค่า IG มากที่สุดมาเป็น root ของ decision tree

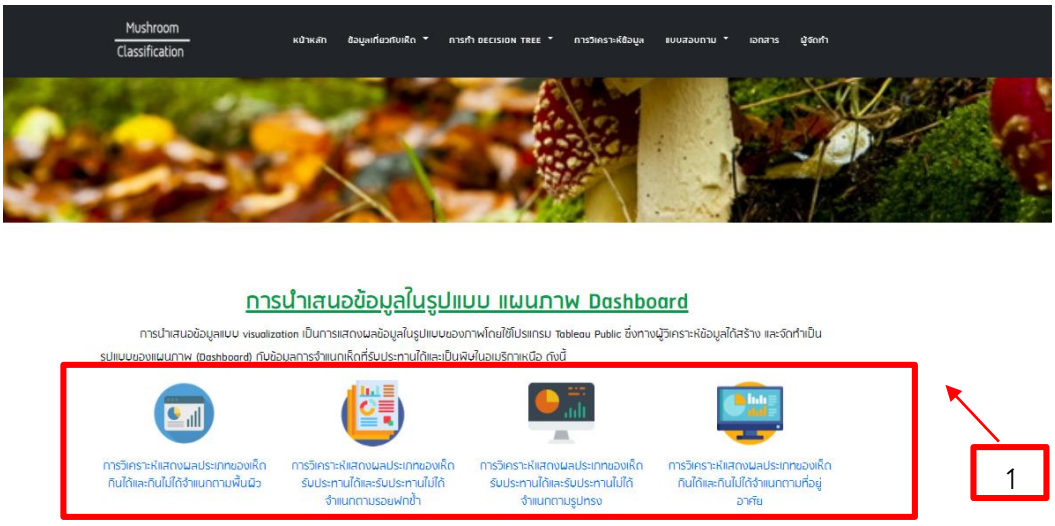


ภาพที่ ก. 7 แสดงการทดสอบกับข้อมูลโดยใช้โปรแกรม Weka 3.8.4



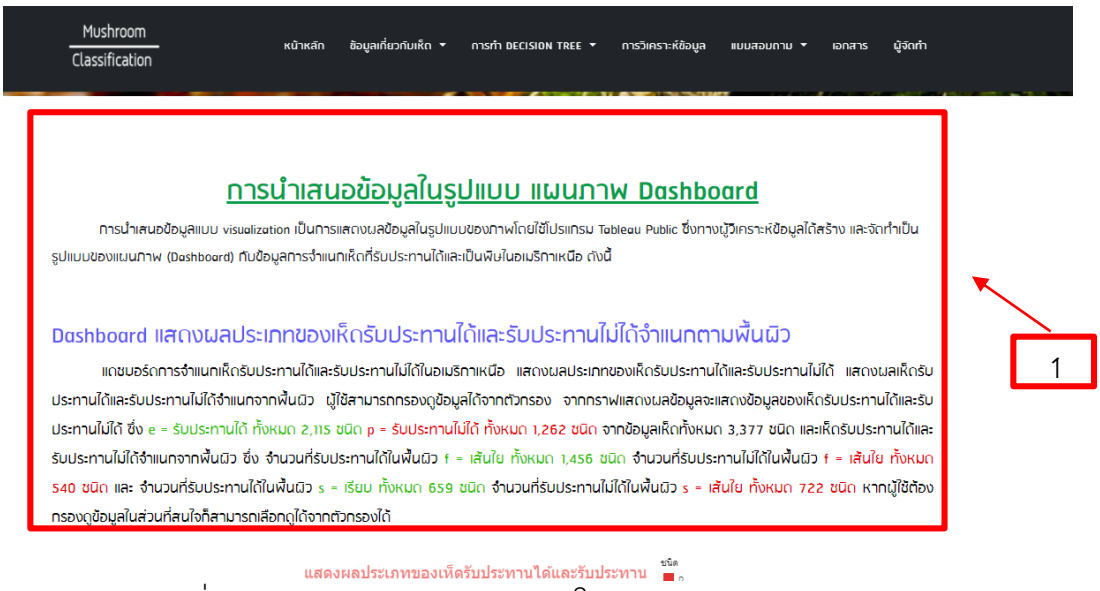
ภาพที่ ก. 8 แสดงการเปรียบเทียบโมเดลของโปรแกรมและการคำนวณมือ

4) ส่วนการวิเคราะห์ข้อมูลการจำแนกเห็ดพิษและเห็ดที่รับประทานได้ในอเมริกาเหนือ



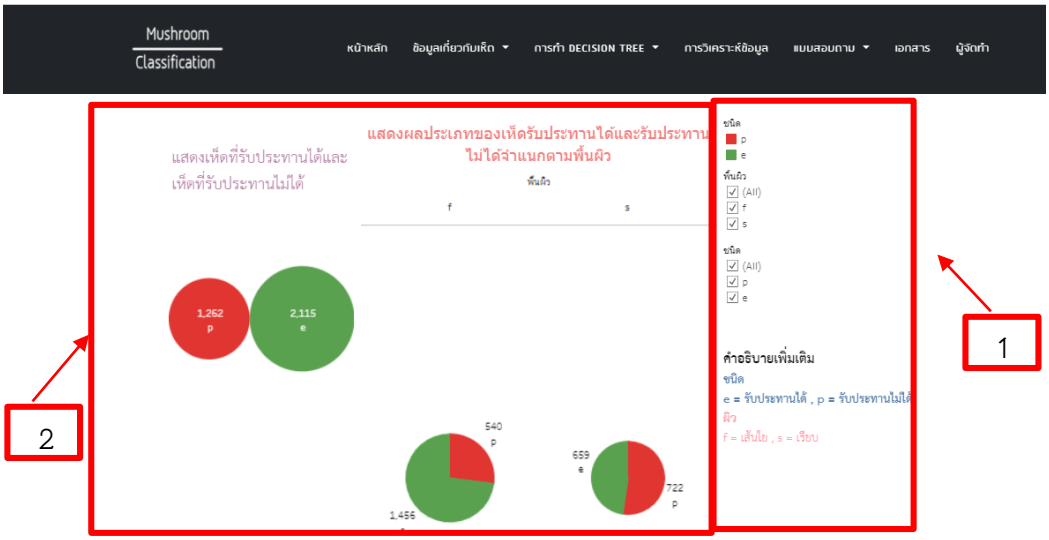
ภาพที่ ก. 9 แสดงการนำเสนอข้อมูลในรูปแบบแผนภาพDashboard

หมายเลข 1 ปุ่มแสดงการนำเสนอข้อมูลในรูปแบบ แผนภาพ Dashboard



ภาพที่ ก. 10 แสดงการนำเสนอข้อมูลในรูปแบบแผนภาพ Dashboard

หมายเลข 1 ส่วนอธิบายข้อมูลภายในแดชบอร์ด

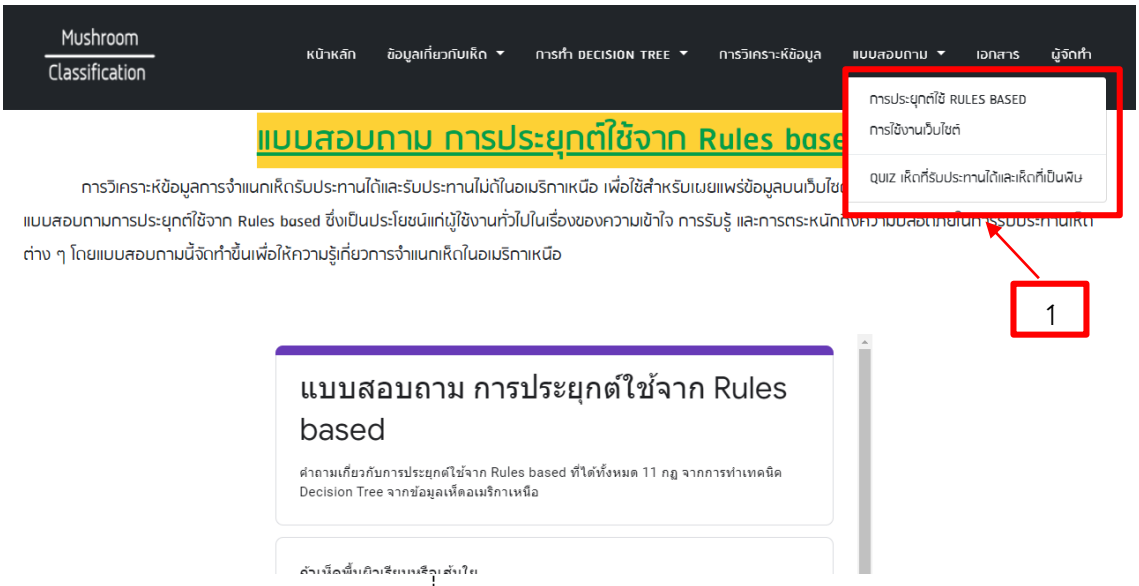


ภาพที่ ก. 11 แสดงผลประเภทของเห็ดรับประทานได้และรับประทานไม่ได้

หมายเลข 1 ตัวกรองข้อมูลของแดชบอร์ดแสดงผล ผู้ใช้สามารถใช้ตัวกรองนี้กรองดู ข้อมูลตามความต้องการได้

หมายเลข 2 ส่วนของแดชบอร์ดแสดงผล

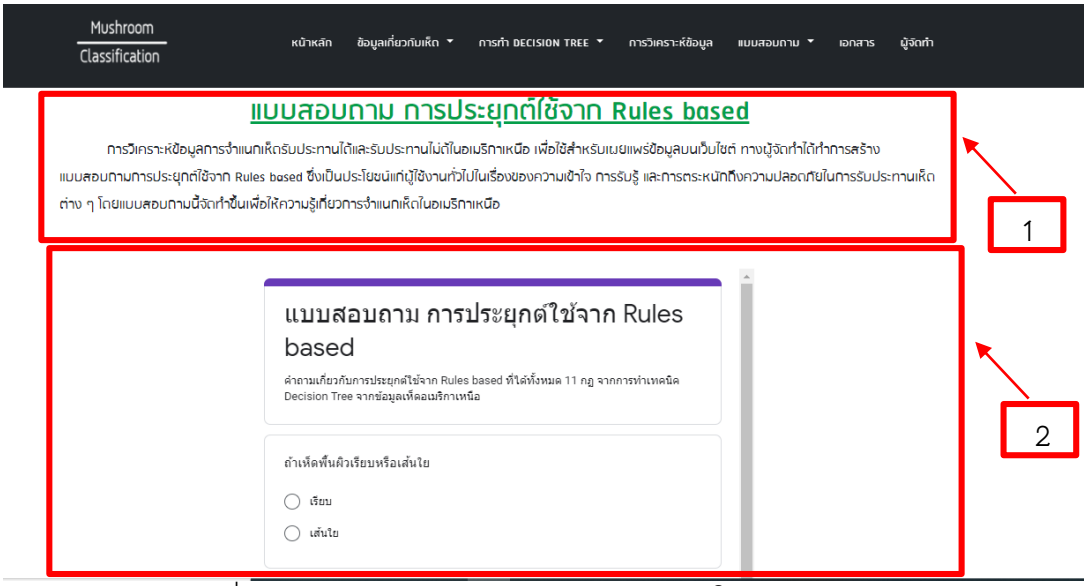
5) หน้าแสดงแบบสอบถามสำหรับผู้เข้าใช้งานเว็บไซต์เพื่อเก็บข้อมูลและนำข้อมูลที่ได้ไป ใช้ประโยชน์ในการวิเคราะห์ข้อมูลต่อไป



ภาพที่ ก. 12 แสดงแบบสอบถาม

หมายเลข 1 ปุ่มเลือกทำแบบสอบถาม ซึ่งแบบสอบถามแบ่งเป็น 3 แบบ

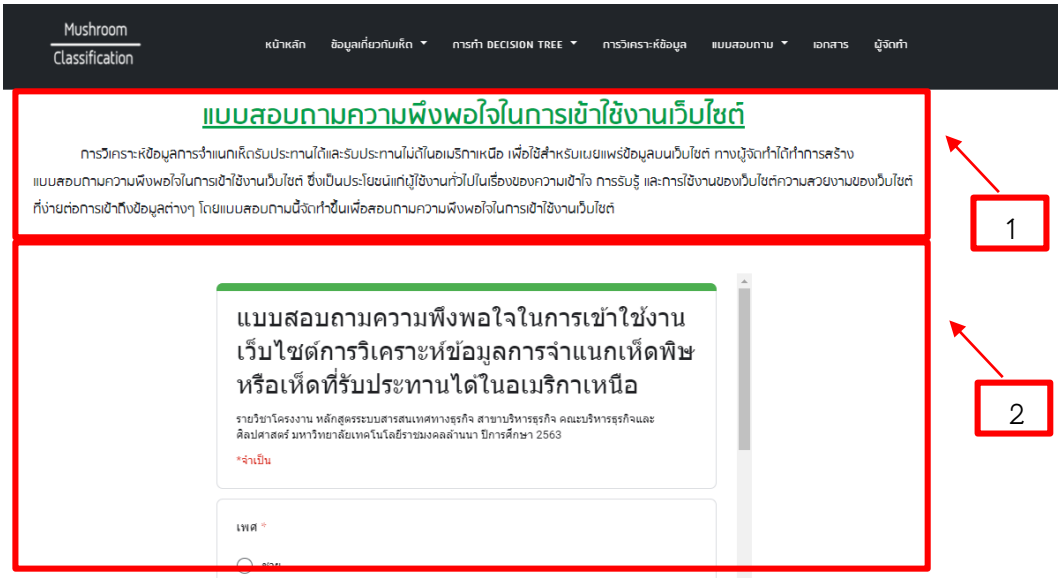
- แบบสอบถาม การประยุกต์ใช้จาก Rules based
- แบบสอบถามความพึงพอใจในการเข้าใช้งานเว็บไซต์
- แบบสอบถาม Quiz เห็ดที่รับประทานได้และเห็ดที่เป็นพิษ



ภาพที่ ก. 13 แสดงแบบสอบถาม การประยุกต์ใช้จาก Rules based

หมายเลข 1 ส่วนของการอธิบายการทำแบบสอบถาม

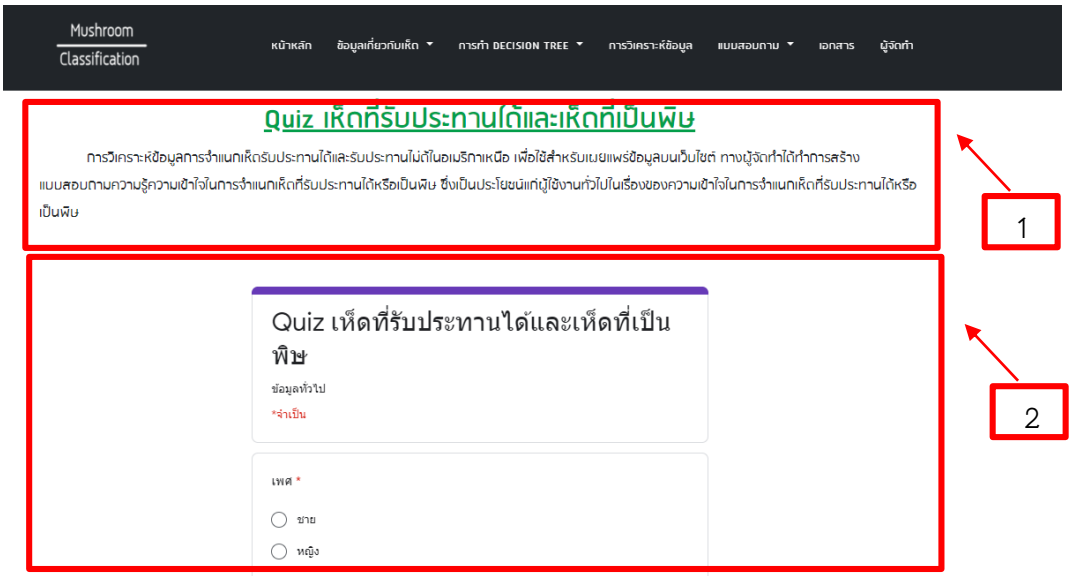
หมายเลข 2 ส่วนของแบบสอบถามการประยุกต์ใช้จาก Rules based



ภาพที่ ก. 14 แสดงแบบสอบถามความพึงพอใจในการเข้าใช้งานเว็บไซต์

หมายเลข 1 ส่วนของการอธิบายการทำแบบสอบถาม

หมายเลข 2 ส่วนของแบบสอบถามความพึงพอใจในการเข้าใช้งานเว็บไซต์

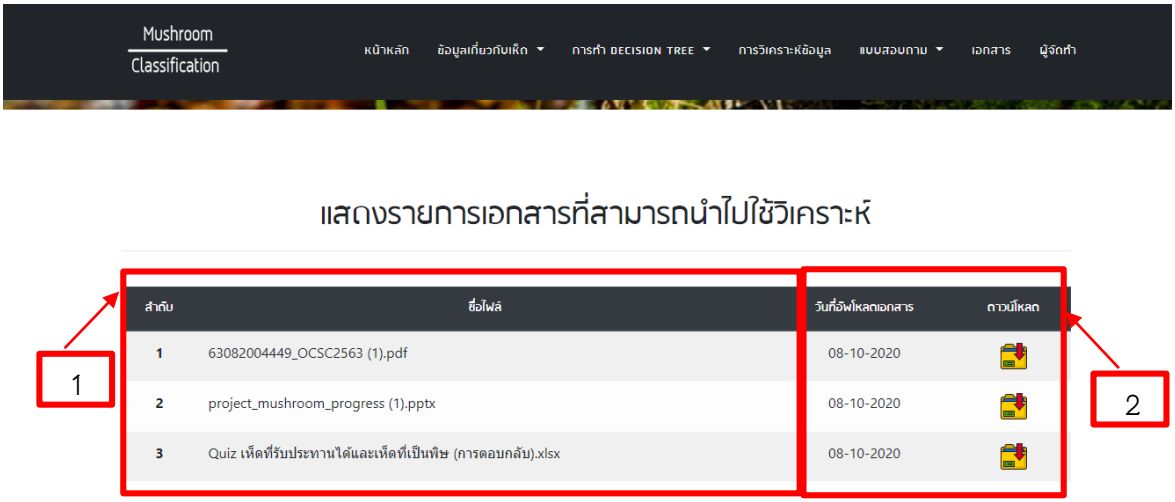


ภาพที่ ก. 15 แสดงแบบสอบถาม Quiz เห็ดที่รับประทานได้และเห็ดที่เป็นพิษ

หมายเลข 1 ส่วนของการอธิบายการทำแบบสอบถาม

หมายเลข 2 ส่วนของแบบสอบถาม Quiz เห็ดที่รับประทานได้และเห็ดที่เป็นพิษ

6) ส่วนของเอกสารที่เกี่ยวข้องและสามารถนำไปใช้วิเคราะห์

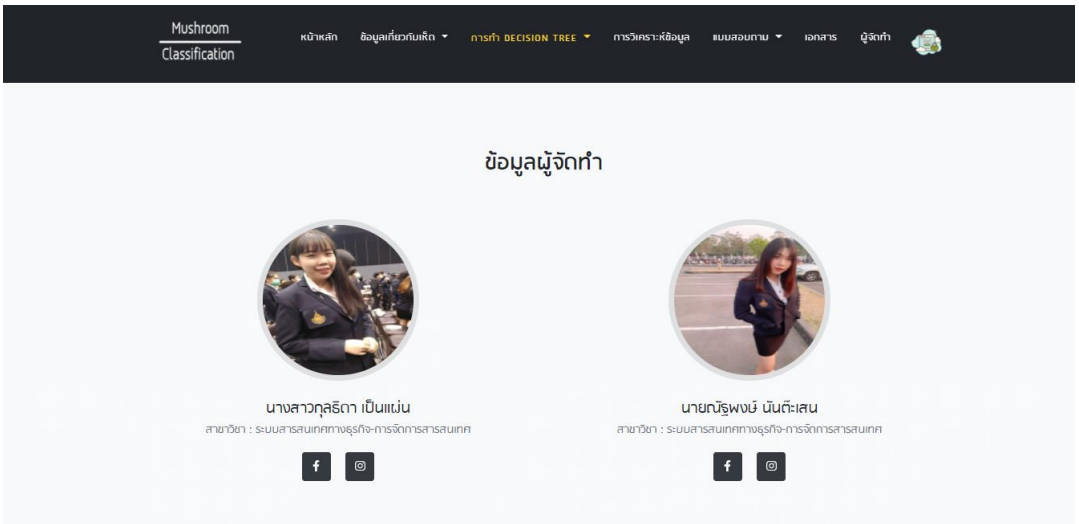


ภาพที่ ก. 16 แสดงรายการเอกสารที่สามารถนำไปใช้วิเคราะห์

หมายเลข 1 ส่วนของรายการเอกสารที่สามารถนำไปใช้วิเคราะห์

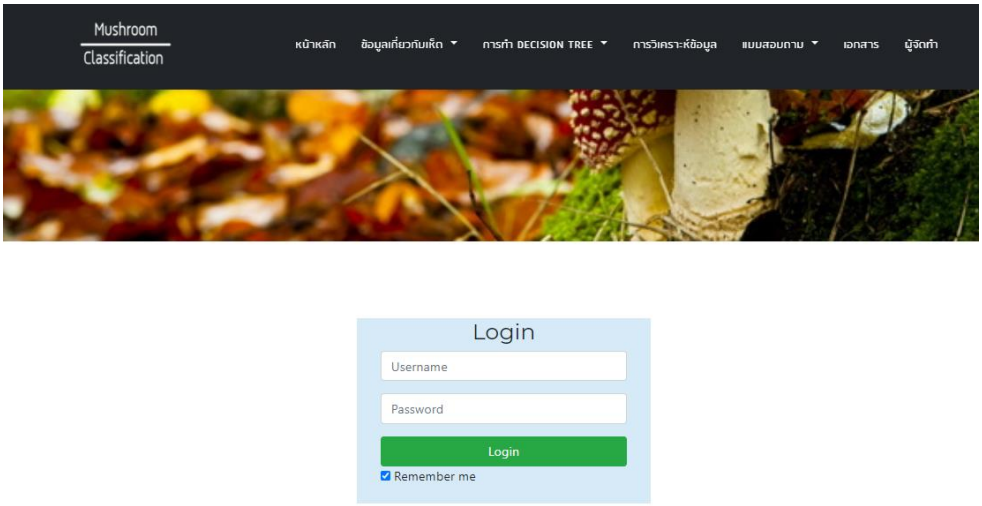
หมายเลข 2 ปุ่มสามารถให้ผู้ใช้ดาวน์โหลดเอกสารไปใช้วิเคราะห์หรือนำไปศึกษาต่อได้

7) ส่วนหน้าจอแสดงประวัติผู้จัดทำเว็บไซต์ และข้อมูลที่ต้องการติดต่อ รวมถึงแสดง หน่วยงานที่เกี่ยวข้องกับโครงงานนี้



ภาพที่ ก. 17 แสดงส่วนติดต่อคณะผู้จัดทำ

8) ส่วนของหน้า Login และ อัปโหลด ไฟล์เอกสารที่เกี่ยวข้องสำหรับแอดมิน

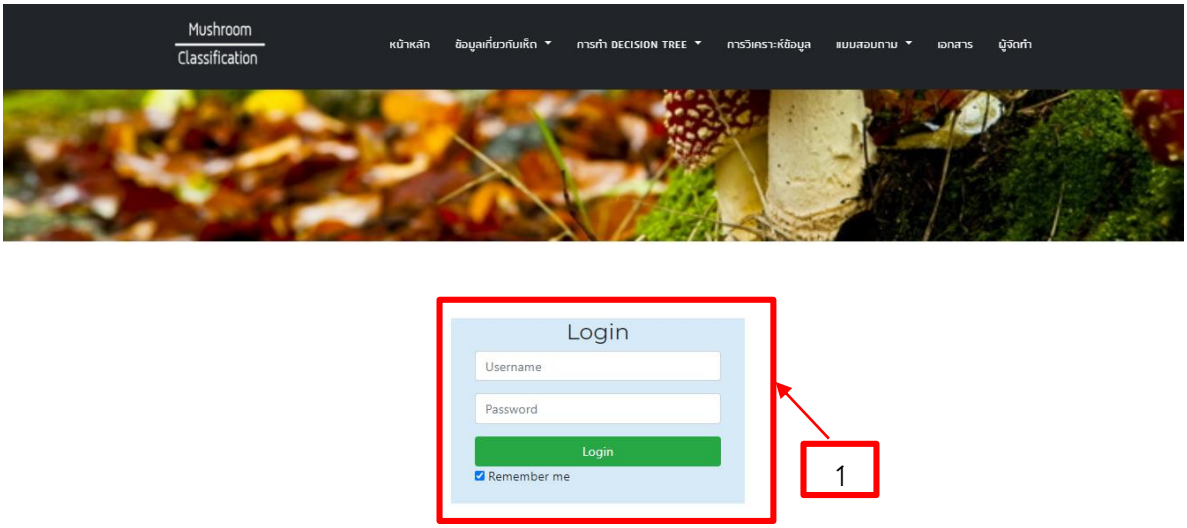


ภาพที่ ก. 18 แสดงส่วนLogin เข้าสู่ระบบสำหรับผู้ดูแลระบบ

คู่มือการใช้งานสำหรับผู้ดูแลระบบ

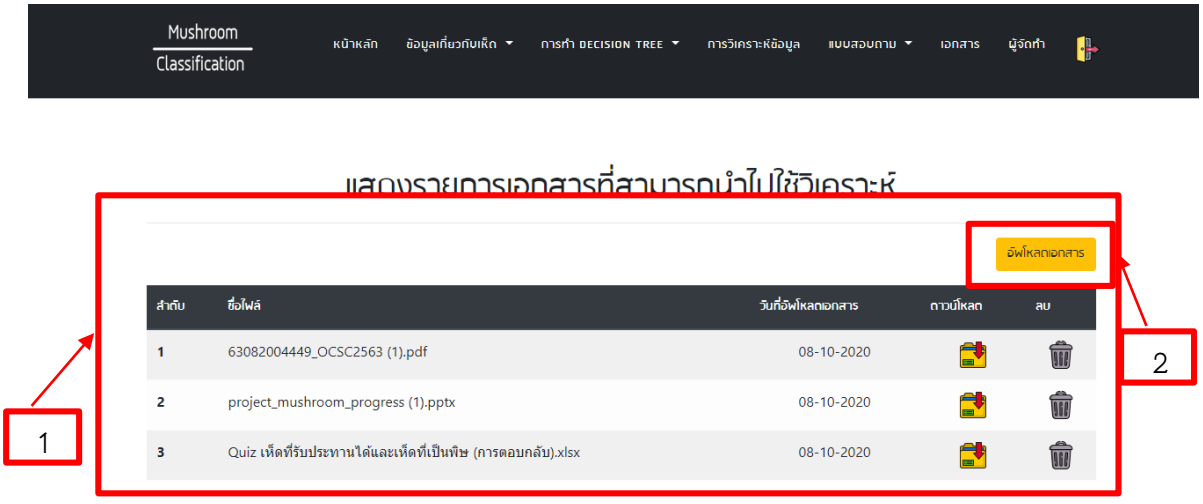
จากการดำเนินงานวิเคราะห์ข้อมูลการจำแนกเห็ดพิษหรือเห็ดที่รับประทานได้ในอเมริกาเหนือ เพื่อเผยแพร่ข้อมูลบนเว็บไซต์ โดยเว็บไซต์มีส่วนใช้งาน 2 ส่วน ดังนี้ ส่วนของการเข้าสู่ระบบ และ ส่วนของการอัปโหลดข้อมูล ผู้จัดทำจึงได้ทำคู่มืออธิบายการใช้งานเว็บไซต์สำหรับผู้ดูแลระบบ ดังต่อไปนี้

1) ส่วนของการเข้าสู่ระบบ



ภาพที่ ก. 19 แสดงส่วนเข้าสู่ระบบของผู้ดูแลระบบ

หมายเลข 1 ผู้ดูแลระบบกรอก Username และ Password เพื่อเข้าสู่ระบบ



ภาพที่ ก. 20 แสดงส่วนของผู้ดูแลระบบ

หมายเลข 1 แสดงไฟล์ที่แอดมินอัปโหลด และสามารถลบไฟล์ได้

หมายเลข 2 ปุ่มอัปโหลดไฟล์ข้อมูลและเอกสาร

2) ส่วนของการอัปโหลดข้อมูล

ภาพที่ ก. 21 แสดงส่วนการอัปโหลดข้อมูล

หมายเลข 1 แสดงหน้าอัปโหลดไฟล์

หมายเลข 2 ปุ่มเลือกไฟล์ข้อมูลและเอกสาร

หมายเลข 3 ปุ่มอัปโหลดข้อมูล

หมายเลข 4 ปุ่มย้อนกลับไปยังหน้าแสดงเอกสารของผู้ดูแลระบบ

ภาคผนวก ข

แบบสอบถาม

ภาคผนวก ข

แบบสอบถาม

ผู้วิเคราะห์ได้ทำการปรับปรุงแก้ไขคำถามในแบบสอบถามที่อ้างอิงมาจากเว็บไซต์สำนักนโยบายและแผนพลังงานกระทรวงพลังงาน และกองเทคโนโลยีสารสนเทศ สำนักงานอธิการบดี มหาวิทยาลัยมหิดล เพื่อให้สอดคล้องกับเนื้อหาและข้อมูลของเว็บไซต์การวิเคราะห์วิเคราะห์ข้อมูลการจำแนกเห็ดพิษหรือเห็ดที่รับประทานได้ในอเมริกาเหนือ เพื่อสอบถามความพึงพอใจของผู้เข้าเยี่ยมชมเว็บไซต์

แบบสอบถามความพึงพอใจการใช้เว็บไซต์

แบบสอบถามนี้มีวัตถุประสงค์เพื่อสำรวจความพึงพอใจการใช้เว็บไซต์โครงงานเรื่องการวิเคราะห์ข้อมูลการจำแนกเห็ดพิษหรือเห็ดที่รับประทานได้ในอเมริกาเหนือ เพื่อเผยแพร่ข้อมูลบนเว็บไซต์ แบ่งออกเป็น 2 ตอน คือ

ตอนที่ 1 ข้อมูลเกี่ยวกับสถานภาพและข้อมูลพื้นฐานของผู้ตอบแบบสอบถาม

คำชี้แจง โปรดให้รายละเอียดที่เกี่ยวกับตัวท่าน โดยเขียนเครื่องหมาย √ ลงใน ( ) หน้าข้อความตามความเป็นจริง

เพศ ( ) ชาย ( ) หญิง

อายุ ( ) ต่ำกว่า 20 ปี ( ) 20–30 ปี

( ) 31–40 ปี ( ) 41–50 ปี

( ) 51–60 ปี ( ) มากกว่า 60 ปี

ระดับการศึกษา ( ) ต่ำกว่าปริญญาตรี ( ) ปริญญาตรี

( ) ปริญญาโท ( ) สูงกว่าปริญญาโท

ตอนที่ 2 การประเมินความพึงพอใจ

คำชี้แจง เขียนเครื่องหมาย √ ลงในช่องระดับความพึงพอใจตามความเป็นจริง

หัวข้อประเมิน	**ระดับความพึงพอใจ**				
	มากที่สุด 5	**มาก 4**	**ปานกลาง 3**	**น้อย 2**	**น้อยที่สุด 1**
ด้านการออกแบบและการจัดรูปแบบ (Design)					
1. ความสวยงาม ความทันสมัย น่าสนใจของหน้าโฮมเพจ					
2. การจัดรูปแบบในเว็บไซต์ง่ายต่อการอ่านและการใช้งาน					
3.สีสันในการออกแบบเว็บไซต์มีความเหมาะสม					
4. เมนูง่ายต่อการใช้งาน					
5.สีพื้นหลังกับสีตัวอักษรมีความเหมาะสมต่อการอ่าน					
6. ขนาดตัวอักษร และรูปแบบตัวอักษร อ่านได้ง่ายและสวยงาม					
7. ภาพกับเนื้อหามีความสอดคล้องกันและสามารถสื่อความหมายได้					
8. โดยภาพรวมท่านมีความพึงพอใจในการออกแบบเว็บไซต์ในระดับใด					
ด้านคุณภาพข้อมูล (Information Quality)					
1. ข้อมูลในเว็บไซต์มีความเกี่ยวเนื่องกับงานของท่าน					
2. ข้อมูลภายในเว็บไซต์เป็นข้อมูลที่ท่านสามารถทำความเข้าใจได้					
3. ข้อมูลในแต่ละเมนูของเว็บไซต์ มีความครบถ้วนสมบูรณ์					

หัวข้อประเมิน	ระดับความพึงพอใจ				
	มากที่สุด 5	มาก 4	ปานกลาง 3	น้อย 2	น้อยที่สุด 1
4. ข้อมูลภายในเว็บไซต์มีความน่าเชื่อถือ					
5. ข้อมูลภายในเว็บไซต์ มีความทันสมัย					
6. ข้อมูลภายในเว็บไซต์มีความถูกต้อง					
7.การนำเสนอรูปแบบ Visualization ด้วยโปรแกรม Tableau Public มีความเหมาะสมกับข้อมูล					
ด้านประโยชน์และการนำไปใช้ (Perceived Usefulness)					
1. เนื้อหามีประโยชน์ต่อผู้ใช้งาน และสามารถนำไปประยุกต์ใช้ได้					
2. สามารถเป็นแหล่งความรู้ได้					
3. เป็นแหล่งข้อมูลที่ตรงกับความต้องการของผู้ใช้งาน					
4. โดยรวมท่านคิดว่าเป็นเว็บไซต์ที่มีประโยชน์					

ข้อเสนอแนะ

.....

.....

.....

แบบสอบถามการประยุกต์ใช้จาก Rules based

แบบสอบถามเกี่ยวกับการประยุกต์ใช้จาก Rules based ที่ได้ทั้งหมด 11 กฎ จากการทำเทคนิค Decision Tree จากข้อมูลเห็ดอเมริกาเหนือ เพื่อการจำแนกเห็ดพิษหรือเห็ดที่รับประทานได้ในอเมริกาเหนือ

ตอนที่ 1 ข้อมูลเกี่ยวกับข้อมูลลักษณะของเห็ด

คำชี้แจง โปรดให้รายละเอียดที่เกี่ยวกับเห็ดโดยเขียนเครื่องหมาย √ ลงใน ( )

1. ถ้าเห็ดมีพื้นผิวเรียบหรือเส้นใย
   ( ) เรียบ
   ( ) เส้นใย
2. ถ้าเห็ดมีที่อยู่อาศัย ในป่า ทุ่งหญ้า หรือ ในเมือง
   ( ) ในป่า
   ( ) ทุ่งหญ้า
   ( ) ในเมือง
3. ถ้าเห็ดมีรอยฟกช้ำ หรือ ไม่มีรอยฟกช้ำ
   ( ) มีรอยฟกช้ำ
   ( ) ไม่มีรอยฟกช้ำ
4. ถ้าเห็ดมีรูปทรงของหมวกเห็ด เป็นทรงระฆัง หรือ ทรงแบน หรือ ทรงนูน
   ( ) ทรงระฆัง
   ( ) ทรงแบน
   ( ) ทรงนูน

แบบทดสอบความรู้เกี่ยวกับเห็ดที่รับประทานได้และเห็ดที่รับประทานไม่ได้

แบบทดสอบความรู้เกี่ยวกับเห็ดที่รับประทานได้และเห็ดที่รับประทานไม่ได้ เพื่อเป็นการวัดประสิทธิภาพความรู้เกี่ยวกับเห็ดที่ผู้วิเคราะห์ได้นำเสนอข้อมูลบนเว็บไซต์

ตอนที่ 1 ข้อมูลเกี่ยวกับสถานภาพและข้อมูลพื้นฐานของผู้ตอบแบบสอบถาม

คำชี้แจง โปรดให้รายละเอียดที่เกี่ยวกับตัวท่าน โดยเขียนเครื่องหมาย √ ลงใน ( ) หน้าข้อความตามความเป็นจริง

เพศ ( ) ชาย ( ) หญิง

อาชีพ ( ) นักเรียน/นักศึกษา ( ) ครู/อาจารย์ ( ) บุคคลทั่วไป

ตอนที่ 2 ความรู้เกี่ยวกับเห็ด

คำชี้แจง เขียนเครื่องหมาย √ ลงใน ( ) ข้อที่ถูกต้องที่สุด

1. เห็ดชนิดนี้สามารถรับประทานได้หรือไม่

( ) รับประทานได้

( ) รับประทานไม่ได้

2. เห็ดชนิดนี้สามารถรับประทานได้หรือไม่

( ) รับประทานได้

( ) รับประทานไม่ได้

3. เห็ดชนิดนี้สามารถรับประทานได้หรือไม่

( ) รับประทานได้

( ) รับประทานไม่ได้

4. เห็ดชนิดนี้ มีชื่อเรียกไทยว่าอะไร

( ) เห็ดเผาะ

( ) เห็ดระโงกเหลือง

5. เห็ดชนิดนี้ มีชื่อเรียกไทยว่าอะไร

( ) เห็ดลม

( ) เห็ดโคน

6. เห็ดชนิดนี้ มีชื่อเรียกวิทยาศาสตร์ว่าอะไร

( ) Amanita jacksonii

( ) Amanita strobiliformis

7. เห็ดชนิดนี้ มีชื่อเรียกวิทยาศาสตร์ว่าอะไร

( ) Termitomyces robustus

( ) Russula delica Fr.

8. เห็ดชนิดนี้รับประทานได้หรือไม่

( ) รับประทานได้

( ) รับประทานไม่ได้

9. เห็ดชนิดนี้รับประทานได้หรือไม่

( ) รับประทานได้

( ) รับประทานไม่ได้

10. เห็ดชนิดนี้รับประทานได้หรือไม่

( ) รับประทานได้

( ) รับประทานไม่ได้

ภาคผนวก ค
เอกสารที่ใช้ในโครงงาน

ภาคผนวก ค
เอกสารที่ใช้ในโครงงาน

FM-BIS-10

แบบบันทึกรายละเอียดการเข้าพบอาจารย์ที่ปรึกษา

วิชา Business Information System Project

หลักสูตรระบบสารสนเทศทางคอมพิวเตอร์ สาขาบริหารธุรกิจ

คณะบริหารธุรกิจและศิลปศาสตร์ มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา เชียงใหม่

ชื่อโครงการ การวิเคราะห์ข้อมูลการจำแนกเห็ดพิษหรือเห็ดที่กินได้ในอเมริกาเหนือ เพื่อเผยแพร่ข้อมูลบนเว็บไซต์

ผู้จัดทำ 1. นางสาวกุลธิดา เป็นแผ่น รหัสนักศึกษา 61521207030-9

2. นายณัฐพงษ์ นันต๊ะเสน รหัสนักศึกษา 61521207035-8

อาจารย์ที่ปรึกษา ผู้ช่วยศาสตราจารย์ ดร. ชัชฎาพร ปุกแก้ว

ครั้งที่	วัน/เดือน/ปี	รายละเอียดการเข้าพบอาจารย์ที่ปรึกษา	ความคืบหน้าของผลงาน	ลายมืออาจารย์ที่ปรึกษา
1		ส่งความคืบหน้า บท 1-3	5%	
2		แก้ บท 1-3	5%	
3		ส่งความคืบหน้า	5%	
4		ส่งความคืบหน้าโปรแกรม Tableau + ข้อมูล	5%	
5		แก้ความคืบหน้าโปรแกรม Tableau + ข้อมูลของเว็บไซต์	5%	
6	2 ก.ย. 63	ส่งความคืบหน้าของเว็บไซต์ และ บท 1-3	15%	
7	9 ก.ย. 63	ส่งความคืบหน้าเว็บไซต์ บท 1-4 และ เพจเวอร์พอยต์ นำเสนอ	5%	
8	23 ก.ย. 63	ส่งความคืบหน้า บท 4 และความคืบหน้าเว็บไซต์	5%	
9	30 ก.ย. 63	ส่งความคืบหน้า บท 4 และ บท 5 ภาคผนวก	10%	
10	7 ต.ค. 63	แก้ไข บทที่ 4-5 หน้า Login upload	10%	

ลงชื่อ กุลธิดา เป็นแผ่น (นักศึกษา) ลงชื่อ ณัฐพงษ์ นันต๊ะเสน (นักศึกษา)

(นางสาวกุลธิดา เป็นแผ่น) (นาย ณัฐพงษ์ นันต๊ะเสน)

ประวัติผู้จัดทำ

ชื่อ-สกุล นางสาวกุลธิดา เป็นแผ่น

วันเดือนปีเกิด วันที่ 26 กุมภาพันธ์ 2541

ภูมิลำเนา 27 ม.12 ต.ชมพู อ.เมือง จ.ลำปาง 52100

E-mail kultida_pe61@live.rmutl.ac.th

ประวัติการศึกษา

- ระดับมัธยมศึกษาตอนต้น โรงเรียนลำปางกัลยาณี จังหวัดลำปาง สำเร็จการศึกษาปี 2555
- ระดับประกาศนียบัตรวิชาชีพ สาขาวิชาเทคโนโลยีสารสนเทศ วิทยาลัยเทคนิคลำปาง จังหวัดลำปาง สำเร็จการศึกษาปี 2558
- ระดับประกาศนียบัตรวิชาชีพขั้นสูง สาขาวิชาเทคโนโลยีสารสนเทศ วิทยาลัยเทคนิคลำปาง จังหวัดลำปาง สำเร็จการศึกษาปี 2560
- ระดับปริญญาตรี สาขาวิชาระบบสารสนเทศทางธุรกิจ มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา จังหวัดเชียงใหม่ สำเร็จการศึกษาปี 2563

ชื่อ-สกุล นายณัฐพงษ์ นันต๊ะเสน

วันเดือนปีเกิด วันที่ 5 กุมภาพันธ์ 2541

ภูมิลำเนา 125 หมู่.11 ต.บ้านแลง อ.เมืองลำปาง จ.ลำปาง 52000

E-mail nattapong_nu611@live.rmutl.ac.th

ประวัติการศึกษา

- ระดับมัธยมศึกษาตอนต้น โรงเรียนเสด็จวนชยางค์กูลวิทยา จังหวัดลำปาง สำเร็จการศึกษาปี 2555
- ระดับประกาศนียบัตรวิชาชีพ สาขาวิชาเทคโนโลยีสารสนเทศ วิทยาลัยเทคนิคลำปาง จังหวัดลำปาง สำเร็จการศึกษาปี 2558
- ระดับประกาศนียบัตรวิชาชีพขั้นสูง สาขาวิชาเทคโนโลยีสารสนเทศ วิทยาลัยเทคนิคลำปาง จังหวัดลำปาง สำเร็จการศึกษาปี 2560
- ระดับปริญญาตรี สาขาวิชาระบบสารสนเทศทางธุรกิจ มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา จังหวัดเชียงใหม่ สำเร็จการศึกษาปี 2563