

การพัฒนาแบบจำลองมาตรการการช่วยเหลือด้านการศึกษา
สำหรับนักศึกษา ด้วยเทคนิคต้นไม้ตัดสินใจ

A development of educational assistance measures model
for students with the decision tree technique.

นายภาณุสรณ์ ใจกลม

นายสมพงษ์ กาดาสาย

หลักสูตรระบบสารสนเทศทางธุรกิจ
สาขาบริหารธุรกิจ คณะบริหารธุรกิจและศิลปศาสตร์
มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา
ปีการศึกษา 2563

การพัฒนาแบบจำลองมาตรการการช่วยเหลือด้านการศึกษา
สำหรับนักศึกษา ด้วยเทคนิคต้นไม้ตัดสินใจ

A development of educational assistance measures model
for students with the decision tree technique.

นายภาณุสรณ์ ใจกลม

นายสมพงษ์ กาดาสาย

หลักสูตรระบบสารสนเทศทางธุรกิจ
สาขาบริหารธุรกิจ คณะบริหารธุรกิจและศิลปศาสตร์
มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา
ปีการศึกษา 2563

ชื่อโครงการ	การพัฒนาแบบจำลองมาตรการการช่วยเหลือด้านการศึกษ สำหรับนักศึกษา ด้วยเทคนิคต้นไม้ตัดสินใจ
โดย	นายภาณุสรณ์ ใจกลม รหัสนักศึกษ 61521207098-6 นายสมพงษ์ กาดาสาย รหัสนักศึกษ 61521207104-2
อาจารย์ที่ปรึกษา	อาจารย์นรินทร์ จิวิตัน
หลักสูตร	ระบบสารสนเทศทางธุรกิจ สาขาบริหารธุรกิจ คณะบริหารธุรกิจ และศิลปศาสตร์
ปีการศึกษา	2563

บทคัดย่อ

การจัดทำโครงการในครั้งนี้มีวัตถุประสงค์ เพื่อพัฒนาแบบจำลองมาตรการการช่วยเหลือด้านการศึกษ สำหรับนักศึกษา ด้วยเทคนิคต้นไม้ตัดสินใจ โดยคณะผู้จัดทำได้เลือกใช้ขั้นตอนกระบวนการวิเคราะห์ข้อมูลด้วย CRISP-DM (Cross Industry Standard Process For Data Mining) จากเทคนิคการทำเหมืองข้อมูลแบบ Classification ที่สร้างกฎด้วยเทคนิคต้นไม้ตัดสินใจ เพื่อจัดกลุ่มข้อมูลที่มีความสัมพันธ์กัน โดยใช้โปรแกรม Weka 3.8.4 โดยใช้โมเดล Decision Tree: J48 ให้ผลลัพธ์ Correctly Classified Instances

โดยเว็บไซต์ที่สร้างขึ้น เป็นลักษณะของการเผยแพร่ข้อมูลที่ได้จากการวิเคราะห์ข้อมูลและการแสดงผลในรูปแบบของแผนภาพโดยใช้โปรแกรม Tableau Public และการสร้างแบบสอบถามที่ใช้ในการตัดสินใจที่จะสนับสนุนเรื่องของการทุนการศึกษา ซึ่งเว็บไซต์นี้ใช้ภาษา HTML CSS PHP และ JavaScript และมีส่วนประกอบต่าง ๆ ดังนี้ 1) ส่วนหน้าหลัก เสนอเนื้อหาข่าวสาร และแผนภาพแดชบอร์ด 2) ส่วนของชุดข้อมูลที่นำมาใช้ในการวิเคราะห์ 3) ส่วนของชุดข้อมูลสารสนเทศที่นำมาใช้ในการวิเคราะห์ 4) ส่วนของการวิเคราะห์ข้อมูลด้วย Decision Tree 5) ส่วนของการวัดประสิทธิภาพโมเดล 6) ส่วนของแบบสอบถามที่ใช้ในการตัดสินใจในการสนับสนุนด้านการศึกษ

จากการวิเคราะห์ข้อมูล และการสร้างเว็บไซต์ พบว่าเว็บไซต์ที่ได้สามารถนำมาประยุกต์ใช้งานได้ตรงตามวัตถุประสงค์ที่ตั้งไว้ ทำให้เกิดประสิทธิภาพในการศึกษาค้นคว้าข้อมูลตามที่ต้องการ สรุปได้ว่าในอนาคตสามารถนำข้อมูลจากเว็บไซต์ไปปรับปรุงเพิ่มเติมต่อได้ เพื่อให้มีประสิทธิภาพมากขึ้นและเกิดประโยชน์สูงสุด

กิตติกรรมประกาศ

โครงการครั้งนี้ สำเร็จลุล่วงไปด้วยดี เนื่องจากได้รับความกรุณาอย่างดียิ่งจาก อาจารย์ นรินทร์ จิวิตัน อาจารย์ที่ปรึกษาโครงการ อาจารย์สุพงศ์ แดงสุริยศรี หัวหน้าหลักสูตรวิชาการระบบสารสนเทศทางธุรกิจ และคณะอาจารย์ในหลักสูตรระบบสารสนเทศทางธุรกิจ คณะบริหารธุรกิจ และศิลปศาสตร์ มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรีทุกท่าน ที่ได้ให้คำปรึกษาและคำแนะนำ ในการจัดทำโครงการ เพื่อบรรลุวัตถุประสงค์ของโครงการให้เสร็จสมบูรณ์ รวมถึงช่วยแก้ไข เอกสารโครงการ ให้มีความสมบูรณ์มากยิ่งขึ้น

ขอขอบคุณ เว็บไซต์ Kaggle.com ที่ให้ศึกษาข้อมูลและนำข้อมูลมาใช้ได้ฟรี เพื่อวิเคราะห์ ตามกระบวนการและขั้นตอนต่าง ๆ

สุดท้ายนี้ คณะผู้จัดทำโครงการขอขอบคุณผู้มีส่วนเกี่ยวข้องทุกท่านที่กรุณาประสิทธิ์ ประสาทวิชาความรู้และประสบการณ์อันมีค่ายิ่งแก่ศิษย์ และคอยชี้แนะแนวทางการจัดทำ โครงการพร้อมทั้ง ช่วยส่งเสริมสนับสนุนให้คณะผู้จัดทำโครงการเกิดแรงผลักดัน เป็นกำลังใจ และ ให้ความช่วยเหลือต่าง ๆ อีกทั้งยังมอบข้อคิดเห็นที่เป็นประโยชน์ในการดำเนิน โครงการนี้ อันส่งผล ให้สำเร็จลุล่วงไปด้วยดี ขอกราบขอบพระคุณเป็นอย่างสูงไว้ ณ ที่นี้

ภาณุสรณ์ ใจกลม

สมพงษ์ กาดาสาย

สารบัญ

	หน้า
บทคัดย่อ	ก
กิตติกรรมประกาศ	ข
สารบัญ	ค
สารบัญภาพ	จ
สารบัญตาราง	ช
บทที่ 1 บทนำ	1
1.1 ความสำคัญและที่มา	1
1.2 วัตถุประสงค์	2
1.3 ประโยชน์ที่จะได้รับ	2
1.4 ขอบเขต	2
1.5 เครื่องมือที่ใช้ในการวิเคราะห์ข้อมูลและพัฒนาโปรแกรม	4
1.6 สถานที่ใช้ในการดำเนินการศึกษาและรวบรวมข้อมูล	4
1.7 ระยะเวลาในการดำเนินการ	5
1.8 นิยามศัพท์เฉพาะ	5
1.9 บทสรุป	7
บทที่ 2 แนวคิด ทฤษฎีเอกสารที่เกี่ยวข้อง	8
2.1 แนวคิด	9
2.2 ทฤษฎี	15
2.3 เครื่องมือในการออกแบบและวิเคราะห์ข้อมูล	32
2.4 วรรณกรรมที่เกี่ยวข้อง	37
2.5 บทสรุป	41
บทที่ 3 วิธีการดำเนินงานโครงงาน	42
3.1 การวิเคราะห์ข้อมูลด้วย CRISP-DM	42
3.2 การออกแบบเว็บไซต์	80

สารบัญ(ต่อ)

	หน้าที่
3.3 บทสรุป	86
บทที่ 4 ผลการดำเนินงาน	87
4.1 ผลการดำเนินงาน	87
4.2 การอภิปรายผล	92
4.3 บทสรุป	98
บทที่ 5 สรุปและข้อเสนอแนะ	99
5.1 บทสรุปผลโครงการ	99
5.2 ข้อจำกัดของระบบ	101
5.3 ปัญหาและอุปสรรคของโครงการ	101
5.4 ข้อเสนอแนะ	101
บรรณานุกรม	102
ภาคผนวก ก คู่มือการใช้งานเว็บไซต์	105
ภาคผนวก ข แบบฟอร์มและเอกสารที่ใช้ในโครงการ	114
ประวัติผู้จัดทำ	117

สารบัญภาพ

	หน้าที่
ภาพที่ 2.1 แสดง DATA VISUALIZATION	10
ภาพที่ 2.2 แสดง NETWORK GRAPH	11
ภาพที่ 2.3 แสดง BUMP CHART	11
ภาพที่ 2.4 แสดง HEAT MAP	12
ภาพที่ 2.5 แสดง STACKED BAR CHART	12
ภาพที่ 2.6 แสดง STACKED BAR CHART	13
ภาพที่ 2.7 แสดง HISTOGRAM	13
ภาพที่ 2.8 แสดง TREE MAP	14
ภาพที่ 2.9 แสดง HEXBIN MAP	14
ภาพที่ 2.10 แสดงโครงสร้างเว็บไซต์แบบเรียงลำดับ	21
ภาพที่ 2.11 แสดงโครงสร้างเว็บไซต์แบบลำดับชั้น	22
ภาพที่ 2.12 แสดงโครงสร้างเว็บไซต์แบบตาราง	22
ภาพที่ 2.13 แสดงโครงสร้างเว็บไซต์แบบใยแมงมุม	23
ภาพที่ 2.14 ไวยากรณ์ของ HTML	25
ภาพที่ 2.15 ไวยากรณ์ของ CSS	26
ภาพที่ 2.16 แสดงซอร์สโค้ดส่วนประกอบของโครงสร้างของ CSS	27
ภาพที่ 2.17 ไวยากรณ์ของ PHP	28
ภาพที่ 2.18 แสดงตัวอย่างต้นไม้ตัดสินใจของลูกค้าที่ทำการซื้อคอมพิวเตอร์	31
ภาพที่ 2.19 แสดงตัวอย่าง RAPIDMINER PLATFORM	33
ภาพที่ 2.20 แสดงตัวอย่าง AUTO MODEL พีเจอร์สำหรับทำ MACHINE LEARNING	33
ภาพที่ 2.21 แสดงตัวอย่างหน้าโปรแกรม WEKA	34
ภาพที่ 2.22 แสดงตัวอย่างหน้าโปรแกรม ATOM	36
ภาพที่ 2.23 แสดงตัวอย่างหน้าโปรแกรม ADOBE XD	37
ภาพที่ 3.1 เว็บไซต์ Kaggle.com	43
ภาพที่ 3.2 ข้อมูลนักศึกษาเบื้องต้น	44

สารบัญภาพ(ต่อ)

	หน้าที่
ภาพที่ 3.3 ชุดข้อมูลที่คัดเลือกมาวิเคราะห์ข้อมูล	46
ภาพที่ 3.4 รูปแบบโมเดล GRAPH DECISION TREE ในโปรแกรม RAPID MINER STUDIO 9.5.1	65
ภาพที่ 3.5 คำบรรยายลักษณะงาน DECISION TREE ในโปรแกรม RAPID MINER STUDIO 9.5.1	66
ภาพที่ 3.6 รูปแบบโมเดล GRAPH DECISION TREE ในโปรแกรม WEKA 3.8.4	67
ภาพที่ 3.7 คำบรรยายลักษณะงาน DECISION TREE ในโปรแกรม WEKA 3.8.4	68
ภาพที่ 3.8 ส่วนค่า CONFUSION MATRIX ของโมเดลโดยใช้อัลกอริทึม C4.5 (J48)	69
ภาพที่ 3.9 ผลการทดสอบของโมเดลโดยใช้อัลกอริทึม C4.5 (J48)ทดสอบกับข้อมูลทดสอบ	70
ภาพที่ 3.10 แสดงข้อมูลแสดงผลประเภทของอาชีพบิดาของนักศึกษาที่ได้รับทุนและไม่ได้รับทุน	75
ภาพที่ 3.11 แสดงข้อมูลแสดงผลประเภทของอาชีพมารดาของนักศึกษาที่ได้รับ รับทุนและไม่ได้รับทุน	76
ภาพที่ 3.12 แสดงข้อมูลแสดงผลประเภทของเปอร์เซ็นต์การอยู่กับครอบครัวและการอยู่ห่าง	76
ภาพที่ 3.13 แสดงข้อมูลแสดงผลประเภทของครอบครัวตั้งแต่ 3 ขึ้นไปสนับสนุนการเรียนเท่าไร ไม่สนับสนุนการเรียนเท่าไร,ครอบครัวน้อยกว่า 3 สนับสนุนการเรียนเท่าไร ไม่สนับสนุน การเรียนเท่าไร	77
ภาพที่ 3.14 แสดงข้อมูลแสดงผลประเภทของการอยู่อาศัยในเมืองหรือนอกเมืองของนักศึกษา	77
ภาพที่ 3.15 แสดงข้อมูลแสดงผลจำนวนที่ได้รับทุนของเพศชายและเพศหญิง	78
ภาพที่ 3.16 แสดงข้อมูลแสดงผลการได้รับทุนการศึกษา และไม่ได้รับทุนการศึกษาทั้งหมด	78
ภาพที่ 3.17 แสดงข้อมูลแสดงผลผู้ปกครองของนักศึกษาที่ได้รับทุนและไม่ได้รับทุน	79
ภาพที่ 3.18 แสดงข้อมูลแสดงผลประเภทของการศึกษาบิดามารดาของนักศึกษาที่ได้รับทุนและ ไม่ได้รับทุน	79
ภาพที่ 3.19 หน้าโฮมเพจของเว็บไซต์	80
ภาพที่ 3.20 หน้าเว็บเพจที่ 2 ชุดข้อมูล	81
ภาพที่ 3.21 หน้าเว็บเพจที่ 3 ชุดข้อมูลสารสนเทศ	82
ภาพที่ 3.22 หน้าเว็บเพจที่ 4 CLASSIFICATION MODEL	83
ภาพที่ 3.23 หน้าเว็บเพจที่ 5 หน้าผลการประเมินโมเดล	84

สารบัญภาพ(ต่อ)

	หน้าที่
ภาพที่ 3.24 หน้าเว็บเพจที่ 6 แสดงหน้าการประเมินคุณสมบัติ	85
ภาพที่ 4.1 หน้าแรกของเว็บไซต์ เมื่อเข้าสู่เว็บไซต์	88
ภาพที่ 4.2 เมนูแสดงชุดข้อมูลดิบที่นำมาใช้	88
ภาพที่ 4.3 เมนูแสดงชุดข้อมูลสารสนเทศที่นำมาใช้	89
ภาพที่ 4.4 เมนูแสดงกระบวนการวิเคราะห์ ด้วย DECISION TREE	89
ภาพที่ 4.5 เมนูผลการวัดประสิทธิภาพของโมเดล	90
ภาพที่ 4.6 เมนูการประเมินคุณสมบัติ	90
ภาพที่ 4.7 หน้าเข้าสู่ระบบเฉพาะแอดมิน	91
ภาพที่ 4.8 หน้าข้อมูลส่วนตัวของผู้จัดทำ	91
ภาพที่ ก.1 แสดงหน้าแรกของเว็บไซต์	106
ภาพที่ ก.2 แสดงหน้าสรุปผลแดชบอร์ด	107
ภาพที่ ก.3 แสดงหน้าชุดข้อมูลที่นำมาใช้	107
ภาพที่ ก.4 แสดงหน้าชุดข้อมูลที่นำมาใช้	108
ภาพที่ ก.5 แสดงหน้ากระบวนการวิเคราะห์ข้อมูล	109
ภาพที่ ก.6 แสดงหน้ากระบวนการวิเคราะห์ข้อมูล	109
ภาพที่ ก.7 แสดงหน้าการประเมินคุณสมบัติ	110
ภาพที่ ก.8 แสดงหน้าผลลัพธ์การประเมินคุณสมบัติ	110
ภาพที่ ก.9 แสดงหน้าเข้าสู่ระบบ	111
ภาพที่ ก.10 แสดงแจ้งเตือนเข้าสู่ระบบสำเร็จ	111
ภาพที่ ก.11 แสดงแจ้งเตือนเข้าสู่ระบบสำเร็จ	112
ภาพที่ ก.12 แสดงการเข้าสู่ระบบของผู้ดูแลระบบ	112
ภาพที่ ก.13 แสดงส่วนติดต่อของคณะผู้จัดทำ	113

สารบัญตาราง

	หน้าที่
ตาราง 1.1 ระยะเวลาในการดำเนินงาน	5
ตาราง 3.1 สรุปค่า Information Gain ของคอลัมน์ต่าง ๆ	63

บทที่ 1

บทนำ

1.1 ความสำคัญและที่มา

ปัจจุบันอินเทอร์เน็ตและระบบสารสนเทศเข้ามามีบทบาทสำคัญอย่างมากในชีวิตประจำวันของมนุษย์มาก ความเจริญก้าวหน้าของเทคโนโลยีในด้านต่าง ๆ ได้ถูกพัฒนาไปอย่างรวดเร็วและ เทคโนโลยีสารสนเทศได้ถูกนำมาประยุกต์ใช้ภายในองค์กร อาทิเช่น ด้านการติดต่อสื่อสาร ด้านอำนวยความสะดวกด้านการศึกษา ด้านการช่วยการตัดสินใจ ด้านการพยากรณ์ ด้านนำเสนอข้อมูลหรือด้านความรู้ ประเทศไทยนั้นได้มีพัฒนาความเจริญก้าวหน้าทางด้านเทคโนโลยีและข้อมูลมาอย่างต่อเนื่องจะเห็นได้ว่าเทคโนโลยีและข้อมูลนั้นมีบทบาทสำคัญในการดำเนินงานให้ความรู้หรือช่วยเหลือประชาชนมากขึ้นซึ่งเกิดจากการพัฒนาเทคโนโลยีและข้อมูลที่รวบรวมสร้างเป็นสารสนเทศให้เกิดประโยชน์ ยิ่งในปัจจุบันจะเห็นได้ว่า เทคโนโลยีมีความสำคัญมากในหลายด้าน แต่ที่เห็นได้ชัดคือ ด้านการศึกษา เนื่องด้วยเกิดโรคระบาด COVID-19 ทำให้เทคโนโลยีมีบทบาทสำคัญมากต่อการศึกษา แต่จะมีเรื่องของค่าใช้จ่ายมากยิ่งขึ้นซึ่งนักศึกษาส่วนหนึ่ง จะต้องเจอผลกระทบหนักเรื่องของค่าใช้จ่ายต่าง ๆ และอาจส่งผลกระทบต่อการศึกษา การได้รับทุนการศึกษาจะเป็นส่วนหนึ่งที่สามารถส่งเสริมให้นักศึกษามีแรงผลักดันในการศึกษาต่อไป

ด้วยระบบการศึกษาของประเทศไทยในปัจจุบัน ถือว่ามีความเหลื่อมล้ำทางโอกาสอยู่ค่อนข้างมาก เนื่องจากไม่มีมาตรการการช่วยเหลือด้านตัดสินใจของการศึกษา ทั้งปัจจัยความเหลื่อมล้ำฐานะทางเศรษฐกิจ และความเหลื่อมล้ำฐานะทางสังคมล้วนส่งผลต่อโอกาสทางการศึกษาทั้งสิ้น ซึ่งระบบการศึกษาไทยยังคงดำเนินด้วยเงินตราเป็นหลัก และการเข้าถึงโอกาสทางการศึกษาจะเป็นตัวกำหนดฐานะทางสังคม และการประกอบอาชีพในอนาคต แนวทางการแก้ปัญหาที่ เครื่องมือสำคัญที่ช่วยให้การเข้าถึงโอกาสทางการศึกษากระจายไปสู่คนในชั้นรากฐานของสังคมมากขึ้น ทำให้ในปัจจุบันมีหลายภาคส่วนที่ส่งเสริมการกระจายทุนการศึกษา ทั้งในระดับมัธยมศึกษา อุดมศึกษา และบัณฑิตศึกษา แต่ยังไม่เห็นเทคโนโลยีในการจัดการกระบวนการในด้านของทุนการศึกษา เพื่อให้เป็นระบบ ระเบียบ การจัดสรรทุนการศึกษายังไม่เพียงพอต่อความต้องการของนักศึกษาที่มีความประสงค์จะรับทุนการศึกษา ยังมีนักศึกษาจำนวนมากที่ต้องการและมีคุณสมบัติที่ตรงกับวัตถุประสงค์การให้ทุนการศึกษา แต่ยังไม่ได้รับการดูแลช่วยเหลือ

เพราะไม่มีมาตรการช่วยเหลือนักศึกษาอย่างเป็นระบบ จึงจำเป็นต้องมีแบบจำลองมาตรการการช่วยเหลือด้านการศึกษา สำหรับนักศึกษา ด้วยเทคนิคต้นไม้ตัดสินใจ เข้ามาช่วยในการตัดสินใจ

จากปัญหาข้างต้นดังนั้นคณะผู้จัดทำจึงได้นำชุดข้อมูลตัวอย่างที่ได้มา เข้าสู่กระบวนการจำแนกตามรายการชุดข้อมูลโดยการใช้ขีดความสามารถในการประมวลผลของโปรแกรมคอมพิวเตอร์เพื่อเพิ่มความน่าเชื่อถือในการวิเคราะห์และประมวลผลข้อมูล โดยใช้เทคนิคต้นไม้ตัดสินใจในการสร้างแบบจำลองขึ้นมา เพื่อเป็นมาตรการการช่วยเหลือการตัดสินใจของด้านการศึกษาในการมอบทุนการศึกษา ให้กับนักเรียน นักศึกษาอย่างเป็นระบบ

1.2 วัตถุประสงค์

1.2.1 เพื่อพัฒนาแบบจำลองมาตรการการช่วยเหลือด้านการศึกษา สำหรับนักศึกษา ด้วยเทคนิคต้นไม้ตัดสินใจ

1.2.2 เพื่อเสนอแนวทางในการตัดสินใจที่จะมอบทุนการศึกษาให้กับนักศึกษา

1.3 ประโยชน์ที่จะได้รับ

1.3.1 ได้แบบจำลองมาตรการการช่วยเหลือด้านการศึกษา สำหรับนักศึกษา ด้วยเทคนิคต้นไม้ตัดสินใจ

1.3.2 ได้ข้อมูลแนวทางในการตัดสินใจที่จะมอบทุนการศึกษาให้กับนักศึกษา

1.4 ขอบเขต

1.4.1 ขอบเขตของข้อมูล

1.4.1.1 ข้อมูลที่นำมาวิเคราะห์ เป็นข้อมูลที่ได้จากเว็บไซต์ Kaggle.com จำนวน 1,044 ชุด

1.4.1.2 แบ่งข้อมูลที่นำมาวิเคราะห์ออกเป็น 2 กลุ่ม คือ กลุ่มข้อมูลสำหรับการเรียนรู้ (Training Data) 70% เป็นจำนวน 731 ชุด และกลุ่มข้อมูลสำหรับการทดสอบ (Testing Data) 30% จำนวน 313 ชุด

1.4.2 ขอบเขตการวิเคราะห์ข้อมูล

1.4.2.1 รวบรวมข้อมูล (Data Integration) เพื่อตรวจสอบรายละเอียดปริมาณและความน่าเชื่อถือของข้อมูลนักศึกษาเบื้องต้น เพื่อใช้สำหรับเป็นมาตรการช่วยเหลือในด้านของการศึกษา ที่ได้ชุดข้อมูลจากเว็บไซต์ Kaggle.com

1.4.2.2 คัดเลือกข้อมูล (Data Selection) โดยการเลือกคอลัมน์ที่มีความสัมพันธ์กันที่สามารถนำมาสร้างแบบจำลองต้นไม้ตัดสินใจได้ และนำมาใช้ในการนำมาวิเคราะห์

1.4.2.3 การทำความสะอาดข้อมูล (Data Cleaning) ตรวจสอบ เปลี่ยนแปลงข้อมูลให้ได้ข้อมูลที่สมบูรณ์และเหมาะสม เช่น ซึ่ขอลสถานศึกษา เป็นต้น ก่อนนำไปนำไปสร้างกฎในการตัดสินใจ

1.4.2.4 ทำเหมืองข้อมูล (Data Mining Engine) โดยใช้เทคนิคต้นไม้ตัดสินใจในการจำแนกประเภทในการ โดยการทำข้อมูลเข้าสู่โปรแกรม Weka สร้างกฎในการตัดสินใจ

1.4.2.5 ประเมินรูปแบบ (Pattern Evaluation) ที่ได้จากการทำเหมืองข้อมูลจากเทคนิคต้นไม้ตัดสินใจในขั้นตอนก่อนหน้า ซึ่งวัดค่าความยำโดยใช้ Confusion Matrix ในโปรแกรม Weka

1.4.3 ขอบเขตการปรับปรุงและการนำเสนอข้อมูล

1.4.3.1 นำเสนอชุดข้อมูลบนเว็บไซต์

1.4.3.2 นำเสนอรูปแบบโมเดล Decision Tree บนเว็บไซต์

1.4.3.3 นำเสนอข้อมูลสารสนเทศจากข้อมูลดิบ โดยใช้เครื่องมือการจัดการข้อมูล Data Visualization เพื่อนำเสนอเป็นแผนภาพแทนข้อมูล

1.4.3.4 ประยุกต์ใช้แบบจำลองมาตรการการช่วยเหลือด้านการศึกษา สำหรับนักศึกษาด้วยเทคนิคต้นไม้ตัดสินใจ บนเว็บไซต์

1.5 เครื่องมือที่ใช้ในการวิเคราะห์ข้อมูลและพัฒนาโปรแกรม

1.5.1 Hardware

1.5.1.1 Acer Nitro 5 An515-43-R1N1 AMD Ryzen 5 3550H Ram 8 GB

1.5.1.2 Acer Nitro 5 An515-43-R1N1 AMD Ryzen 5 3550H Ram 16 GB

1.5.2 Software

1.5.2.1 โปรแกรม Tableau Public หรือ ใช้ในการแสดงผลข้อมูลในรูปแบบของภาพ

1.5.2.2 ชุดคำสั่งภาษา HTML4, HTML5 (Hyper Text Markup Language)

1.5.2.3 ชุดคำสั่ง CSS3 (Cascading Style Sheets)

1.5.2.4 โปรแกรม Atom ใช้ในการเขียนคำสั่งและเขียนโปรแกรม

1.5.2.5 โปรแกรม Adobe XD ใช้ในการออกแบบหน้าเว็บไซต์

1.5.2.6 ระบบปฏิบัติการ Window 10 ใช้ในการเปิดซอฟต์แวร์ต่าง ๆ

1.5.2.7 Microsoft Office 2016 ใช้ในการทำเอกสารต่าง ๆ

1.5.2.8 โปรแกรม Weka 3.8.4 ใช้ในการวิเคราะห์ข้อมูลด้วยเทคนิคเหมืองข้อมูล

1.5.2.9 โปรแกรม Rapid Miner Studio 9.5.1 ใช้ในการทำด้วยเทคนิคเหมืองข้อมูล

1.5.2.10 โปรแกรม XAMPP 5.5.38 ใช้ในการจำลอง Web Server

1.5.2.11 โปรแกรม FileZilla 3.40.0 ใช้ในการติดต่อ Server เพื่อ Upload File

1.6 สถานที่ใช้ในการดำเนินการศึกษาและรวบรวมข้อมูล

1.6.1 สถานที่ใช้ในการดำเนินการศึกษา

มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนาภาคพายัพเชียงใหม่ 128 ถนนห้วยแก้ว ตำบลช้างเผือก อำเภอเมือง จังหวัดเชียงใหม่ 50300

1.6.2 แหล่งรวบรวมข้อมูล

เว็บไซต์ Kaggle.com

1.7 ระยะเวลาในการดำเนินการ

ตาราง 1.1 ระยะเวลาในการดำเนินงาน

แผนการดำเนินงาน	2563						
	เม.ย.	พ.ค.	มิ.ย.	ก.ค.	ส.ค.	ก.ย.	ต.ค.
1. ศึกษาและกำหนดความต้องการ	←→						
2. ตรวจสอบความถูกต้องข้อมูล		←→					
3. วิเคราะห์ข้อมูลจำแนกตามกลุ่ม		←→					
4. ออกแบบระบบ			←→				
5. เขียนและทดสอบระบบ			←→				
6. ตรวจสอบระบบโดยรวม			←→				
7. ประเมินการใช้งานระบบ			←→				
8. จัดทำเอกสารประกอบ โครงการ	←→						

1.8 นิยามศัพท์เฉพาะ

การวิเคราะห์ข้อมูล (Analysis) หมายถึงเป็นการจัดระเบียบแยกแยะส่วนต่าง ๆ ของหลักฐานหรือข้อมูลที่ได้ออกเป็นหมวดหมู่ เพื่อหาคำตอบตามความมุ่งหมาย และตามสมมติฐานที่ได้กำหนดไว้การวิเคราะห์ข้อมูลนี้เป็นขั้นการทำงานที่ต่อเนื่องมาจากการวัด การนับ และจัดเรียงลำดับข้อมูล ส่วนใหญ่เป็นเรื่องเกี่ยวกับการนำเอาวิธีการทางสถิติมาวิเคราะห์หาค่าตัวแปรหรือหาลักษณะของตัวแปร

เว็บไซต์ (Website) หมายถึง สื่อนำเสนอข้อมูลบนเครื่องคอมพิวเตอร์ หรือคือการรวบรวมหน้าเว็บเพจหลายหน้า ซึ่งเชื่อมโยงกันผ่านทางไฮเปอร์ลิงก์ ซึ่งต้องเปิดด้วยโปรแกรมเฉพาะทางที่เรียกว่า Web Browser โดยถูกจัดเก็บไว้ในเว็ลด์ไวด์เว็บ และเว็บไซต์นั้นถูกสร้างขึ้นด้วยภาษาทางคอมพิวเตอร์ที่เรียกว่า HTML (Hyper Text Markup Language)

ทุนการศึกษา (Scholarships) หมายถึง ทุนการศึกษามีมาช้านานแล้ว และมีหลายประเภท เช่น ทุนรางวัลผลการเรียนดี ให้แก่ผู้ซึ่งผ่านการสอบคัดเลือก แข่งขันความสามารถทางวิชาการ ทุนช่วยเหลือผู้ขาดแคลน ให้แก่เด็กและเยาวชนที่ยากจน หรืออยู่ในแดนทุรกันดาร ทุนส่งเสริมการศึกษาเฉพาะทาง เป็นทุนพิเศษ เพื่อส่งเสริมให้เรียนวิชาเฉพาะวิชาใดวิชาหนึ่ง ให้รู้สึกซึ่งสามารถนำวิชาไปถ่ายทอดแก่ผู้อื่น หรือปฏิบัติงานที่ต้องใช้ความเชี่ยวชาญเป็นพิเศษในเรื่องนั้น ๆ เป็นต้น

การศึกษา (Education) หมายถึง การศึกษาในความหมายทั่วไปอย่างกว้างที่สุด เป็นวิธีการส่งผ่านจุดมุ่งหมายและธรรมเนียมประเพณีให้ดำรงอยู่จากรุ่นหนึ่งสู่อีกรุ่นหนึ่ง โดยทั่วไป การศึกษาเกิดขึ้นผ่านประสบการณ์ใด ๆ ซึ่งมีผลกระทบเชิงพัฒนาต่อวิธีที่คนคนหนึ่งจะคิด รู้สึกหรือกระทำ แต่ในความหมายเทคนิคอย่างแคบ การศึกษาเป็นกระบวนการอย่างเป็นทางการซึ่งส่งผ่านความรู้ ทักษะ จารีตประเพณีและค่านิยมที่สั่งสมมาจาก รุ่นหนึ่ง ไปยังอีกรุ่นหนึ่ง นั่นคือ การสอนในสถานศึกษา สำหรับปัจจุบันนี้มีการแบ่งระดับชั้นทางการศึกษาออกเป็นชั้น ๆ เช่น การศึกษาปฐมวัย ประถมศึกษา มัธยมศึกษา ทั้งนี้รวมไปถึงระดับอาชีวศึกษา อุดมศึกษา และการฝึกงาน

แบบต้นไม้ตัดสินใจ (Decision tree) เป็นหนึ่งในวิธีการเรียนรู้ซึ่งใช้ในสถิติ, การเรียนรู้ของเครื่อง และการทำเหมืองข้อมูล โดยพิจารณาการสังเกตการแบ่งแยกข้อมูลโดยพิจารณาข้อมูลในการเรียนรู้ของเครื่อง (machine learning) ต้นไม้ตัดสินใจ เป็นโมเดลทางคณิตศาสตร์ที่ใช้ทำนายประเภทของวัตถุโดยพิจารณาจากลักษณะของวัตถุ บัพภายใน (inner node) ของต้นไม้จะแสดงตัวแปร ส่วนกิ่งจะแสดงค่าที่เป็นไปได้ของตัวแปร ส่วนบัพใบ (leaf node) จะแสดงประเภทของวัตถุ

ต้นไม้ตัดสินใจที่บัพใบแสดงถึงข้อมูลที่เป็นข้อมูลไม่ต่อเนื่อง (discrete values) จะเรียกว่า ต้นไม้ตัดสินใจแบบจำแนก (classification trees) และต้นไม้ตัดสินใจที่บัพใบเป็นข้อมูลต่อเนื่อง (continuous values) จะเรียกว่าต้นไม้ตัดสินใจแบบถดถอย (regression trees)

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

1.9 บทสรุป

จากบทนำที่ได้กล่าวมาในข้างต้นทั้งหมดนั้น ผู้วิเคราะห์ข้อมูลได้สังเกตเห็นความสำคัญของการนำเทคโนโลยีมาประยุกต์ใช้ในการวิเคราะห์ข้อมูลนักศึกษาเบื้องต้น เพื่อใช้สำหรับเป็นมาตรการช่วยเหลือในด้านของการศึกษา ด้วยกระบวนการวิเคราะห์ข้อมูลด้วย CRISP-DM จากเทคนิคการทำเหมืองข้อมูลแบบ Classification ในรูปแบบของแผนภูมิต้นไม้ หรือที่เรียกว่า Decision tree และเผยแพร่ข้อมูลสารสนเทศบน Web Browser ที่เป็นที่ยอมรับในยุคอินเทอร์เน็ตคือการเผยแพร่ทางสื่อออนไลน์ โดยใช้ภาษา HTML ในการพัฒนาเว็บไซต์สำหรับการเปิดเผยข้อมูลที่ได้มาจาก Kaggle.com เนื่องจากเทคโนโลยีปัจจุบันเข้ามาช่วยในการจัดการข้อมูลหรือดำเนินงานต่าง ๆ ให้มีความสะดวกสบายเป็นอย่างมาก อีกทั้งผู้ใช้งานยังเข้าถึงข้อมูลได้ง่ายและมีประสิทธิภาพ

บทที่ 2

แนวคิด ทฤษฎีเอกสารที่เกี่ยวข้อง

จากการศึกษาเอกสารงานการการพัฒนาแบบจำลองมาตรการการช่วยเหลือด้านการศึกษา สำหรับนักศึกษา ด้วยเทคนิคต้นไม้ตัดสินใจ ผู้ศึกษาพบว่าเอกสารแนวความคิดและโครงการต่าง ๆ ที่เกี่ยวข้องกับการศึกษาสรุปได้ ดังนี้

2.1 แนวคิด

2.1.1 แนวคิดเกี่ยวกับการวิเคราะห์ข้อมูล (Data Analytic)

2.1.2 แนวคิดเกี่ยวกับการแสดงผลข้อมูล (Data Visualization)

2.2 ทฤษฎี

2.2.1 ทฤษฎีเกี่ยวกับการจัดการข้อมูลขนาดใหญ่

2.2.2 ทฤษฎีเกี่ยวกับการทำเหมืองข้อมูล

2.2.3 ทฤษฎีเกี่ยวกับการออกแบบเว็บไซต์

2.2.4 ทฤษฎีเกี่ยวกับชุดคำสั่ง HTML

2.2.5 ทฤษฎีเกี่ยวกับชุดคำสั่ง CSS

2.2.6 ทฤษฎีเกี่ยวกับชุดคำสั่ง PHP

2.2.7 ทฤษฎีเกี่ยวกับการ Visualization

2.2.8 ทฤษฎีเกี่ยวกับการจำแนกข้อมูลด้วยต้นไม้ตัดสินใจ

2.3 เครื่องมือในการออกแบบและวิเคราะห์ข้อมูล

2.3.1 โปรแกรม Rapid Miner

2.3.2 โปรแกรม Weka

2.3.3 โปรแกรม Atom

2.3.4 โปรแกรม Adobe XD

2.4 วรรณกรรมที่เกี่ยวข้อง

2.5 บทสรุป

2.1 แนวคิด

2.1.1 แนวคิดเกี่ยวกับการวิเคราะห์ข้อมูล (Data Analytic)

Data analytics เป็นศาสตร์ของการวิเคราะห์ข้อมูลต่าง ๆ จาก Big data เพื่อช่วยในด้านธุรกิจ หรือตามวัตถุประสงค์อื่น ๆ ที่ต้องการ ซึ่งเริ่มต้นจากความสามารถของเราในการนำข้อมูลเหล่านั้น มาให้อยู่ในรูปแบบที่พร้อมจะประมวลผลได้ โดยเทคโนโลยีหรือชุดคำสั่งและแบบจำลองที่สร้างขึ้น เพื่อนำข้อมูลที่ได้ผ่านการวิเคราะห์แล้วมาใช้ หรือแปลความหมายโดยบุคคลที่ได้รับการฝึกอบรมการใช้เทคโนโลยีเหล่านั้น รูปแบบของการวิเคราะห์ข้อมูลสามารถแบ่งได้ดังนี้

1) การวิเคราะห์ข้อมูลแบบพื้นฐาน (Descriptive analytics) เป็นการวิเคราะห์เพื่อแสดงผลของรายการทางธุรกิจ เหตุการณ์ หรือกิจกรรมต่าง ๆ ที่ได้เกิดขึ้น หรืออาจกำลังเกิดขึ้นในลักษณะที่ง่ายต่อการเข้าใจ หรือต่อการตัดสินใจ ตัวอย่างเช่น รายงานการขาย รายงานผลการดำเนินงาน

2) การวิเคราะห์แบบเชิงวินิจฉัย (Diagnostic analytics) เป็นการอธิบายถึงสาเหตุของสิ่งที่เกิดขึ้นปัจจัยต่าง ๆ และความสัมพันธ์ของปัจจัยหรือตัวแปรต่าง ๆ ที่มีความสัมพันธ์ต่อกันของสิ่งที่เกิดขึ้น ตัวอย่างเช่น ความสัมพันธ์ระหว่างยอดขายต่อกิจกรรมทางการตลาดแต่ละประเภท ซึ่งเป็นก้าวใหม่ที่ช่วยเสริมให้ตัดสินใจไปในทางที่ถูกต้อง

3) การวิเคราะห์แบบพยากรณ์ (Predictive analytics) เป็นการวิเคราะห์เพื่อพยากรณ์สิ่งที่กำลังจะเกิดขึ้นหรือน่าจะเกิดขึ้น โดยใช้ข้อมูลที่ได้เกิดขึ้นแล้วกับแบบจำลองทางสถิติ หรือเทคโนโลยีปัญญาประดิษฐ์ต่าง ๆ (Artificial intelligence) ตัวอย่างเช่น การพยากรณ์ยอดขาย การพยากรณ์ผลประชามติ

4) การวิเคราะห์แบบให้คำแนะนำ (Prescriptive analytics) เป็นการวิเคราะห์ข้อมูลที่มีความซับซ้อนที่สุด เป็นทั้งการพยากรณ์สิ่งต่าง ๆ ที่จะเกิดขึ้น ข้อดี ข้อเสีย สาเหตุ และระยะเวลาของสิ่งที่เกิดขึ้น และการให้คำแนะนำ ทางเลือกต่าง ๆ ที่มีอยู่ และผลของแต่ละทางเลือก

2.1.2 แนวคิดเกี่ยวกับการแสดงผลข้อมูล (Data Visualization)

การนำข้อมูลดิบมาเปลี่ยนเป็นกราฟ แผนภูมิ หรือแม้กระทั่งวิดีโอที่ช่วยอธิบายปริมาณตัวเลข และช่วยให้คุณได้ข้อมูลเชิงลึกจากข้อมูลเหล่านั้น การใช้ Data Visualization เปลี่ยนแนวคิดของเราเกี่ยวกับการสร้างคุณค่าให้กับข้อมูล ทำให้เราค้นพบรูปแบบใหม่ ๆ และมองเห็นแนวโน้มของข้อมูลมากขึ้น

หากต้องการประสบความสำเร็จในโลกธุรกิจที่มีการแข่งขันสูง ความสามารถในการตัดสินใจที่ถูกต้องในเวลาที่เหมาะสม โดยอาศัยข้อมูลเชิงลึกที่รัดกุมเป็นสิ่งสำคัญ ปัจจุบันองค์กรธุรกิจส่วนใหญ่ในประเทศไทย เริ่มใช้ประโยชน์จากการทำ Data Visualization เพราะว่าเครื่องมือนี้เป็นกุญแจสำคัญในการแสดงผลของข้อมูลให้เข้าใจง่าย ซึ่งจะช่วยให้ทุกแผนกในฝ่ายงานได้ใช้ประโยชน์ ในการหาแนวโน้มและความสัมพันธ์เชื่อมโยง อีกทั้งทำให้ทราบข้อมูลได้เร็วขึ้น ส่งผลให้การทำงานขององค์กรมีประสิทธิภาพเพิ่มขึ้น จนดำเนินการถึงการวางกลยุทธ์ภายในองค์กร และสามารถนำไปใช้กับการทำเว็บไซต์ได้ Data Visualization ช่วยให้การสื่อสารข้อมูล มีความรวดเร็วและเข้าใจง่ายมากขึ้น

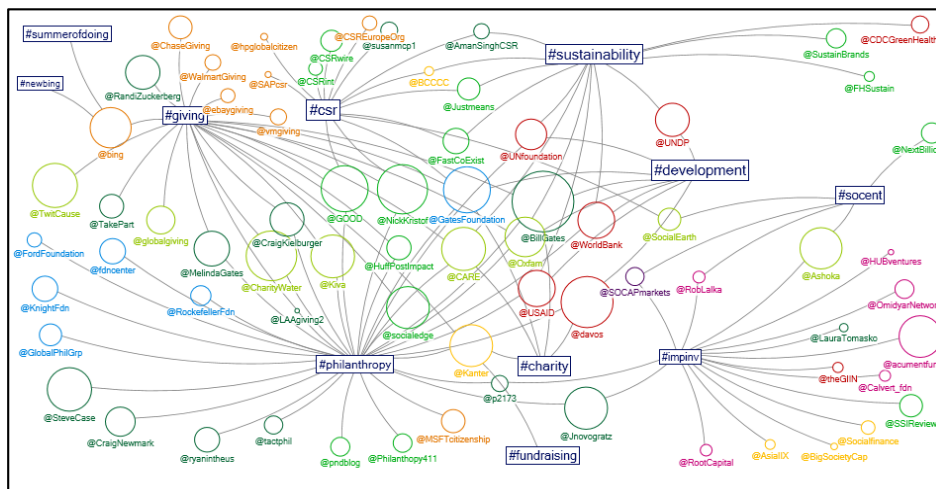


ภาพที่ 2.1 แสดง Data Visualization

พื้นฐานของการสร้าง Data Visualization คือ การ Mapping ส่วนข้อมูลกับส่วนของ Graphic เข้าด้วยกัน ซึ่งตอนนี้มีโปรแกรมสำเร็จรูปในการสร้าง Data Visualization หลากหลายโปรแกรม มีฟังก์ชันการใช้งานที่เข้าใจง่าย เช่น การสร้างฟิลเตอร์ การออกแบบเพื่อให้งานการวิเคราะห์ข้อมูล มีความยืดหยุ่นเป็นต้น ตัวอย่างรูปแบบ Data Visualization ที่แสดงผลให้อย่างสนใจ มีดังนี้

1) NETWORK/FLOW

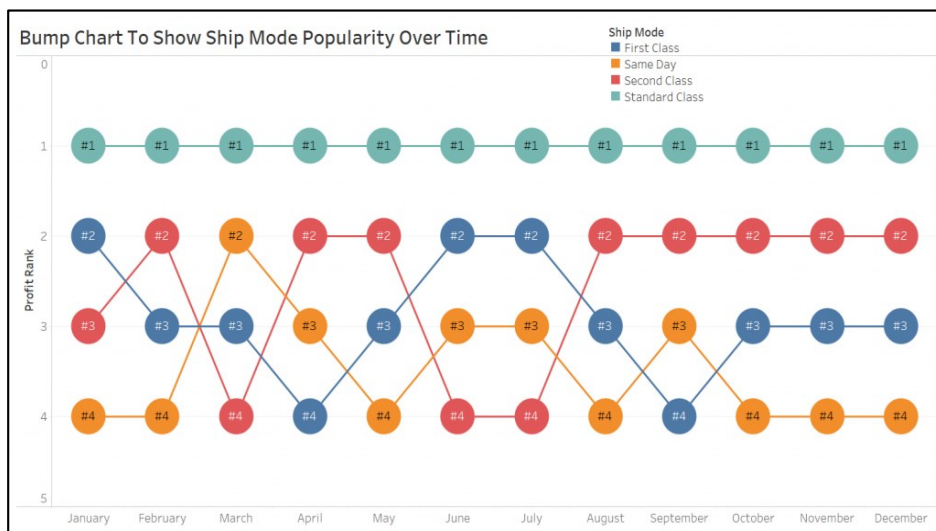
Network Graph ใช้แสดงความเชื่อมโยงของเครือข่ายหรือความสัมพันธ์ในกลุ่ม



ภาพที่ 2.2 แสดง Network Graph

2) RANKING

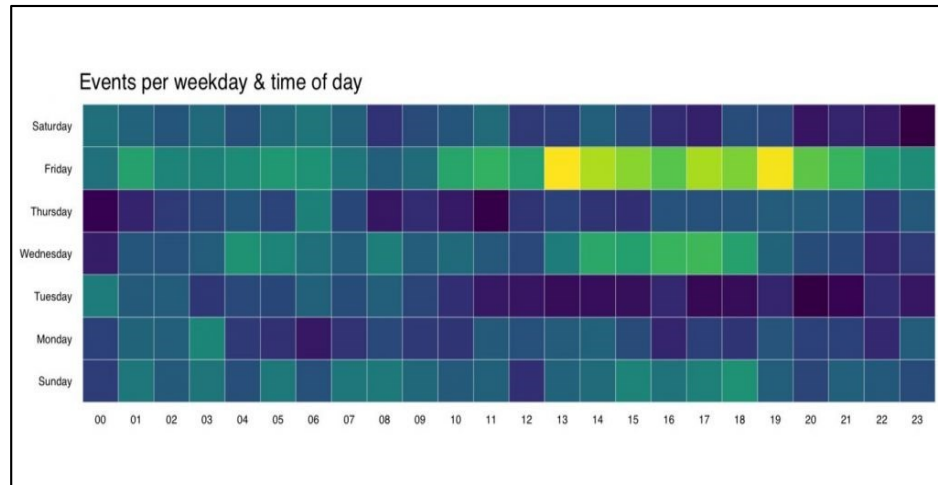
Bump Chart ใช้แสดงการเปลี่ยนแปลงลำดับของข้อมูลในหลายช่วงเวลา โดยใช้สีเพื่อช่วยให้อ่านข้อมูลได้ง่ายขึ้น



ภาพที่ 2.3 แสดง Bump Chart

3) RELATIONSHIP

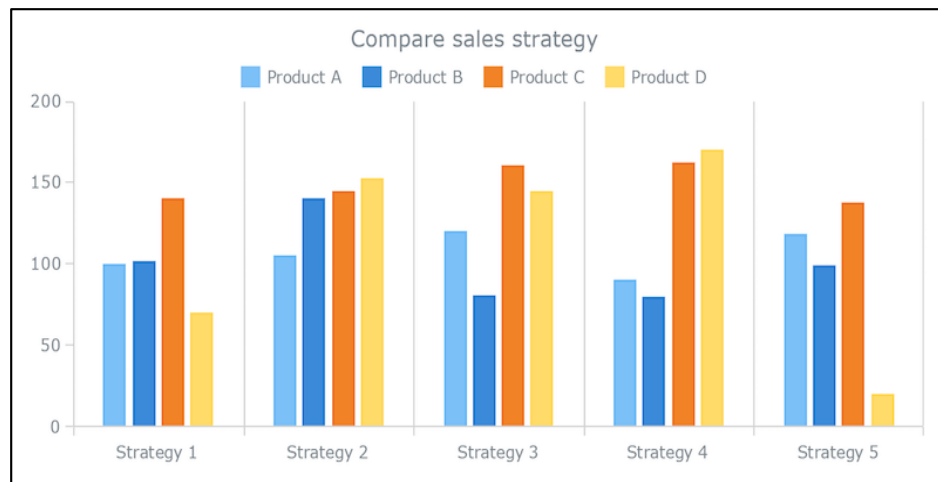
Heat map ใช้แสดงรูปแบบความสัมพันธ์ของข้อมูล โดยจะแสดงออกมาในรูปแบบของ “ สี ” ซึ่งแต่ละสีจะบ่งบอกถึงระดับความถี่ของพฤติกรรม แต่ใช้อ่านค่าความแตกต่างเล็กน้อยได้ยาก



ภาพที่ 2.4 แสดง Heat map

4) COMPARISON

Stacked Bar Chart ใช้เปรียบเทียบค่าผลรวมและสัดส่วนจากข้อมูลหลายกลุ่ม อาจอ่านสัดส่วนยากเมื่อมีข้อมูลหลายกลุ่มมากเกินไป

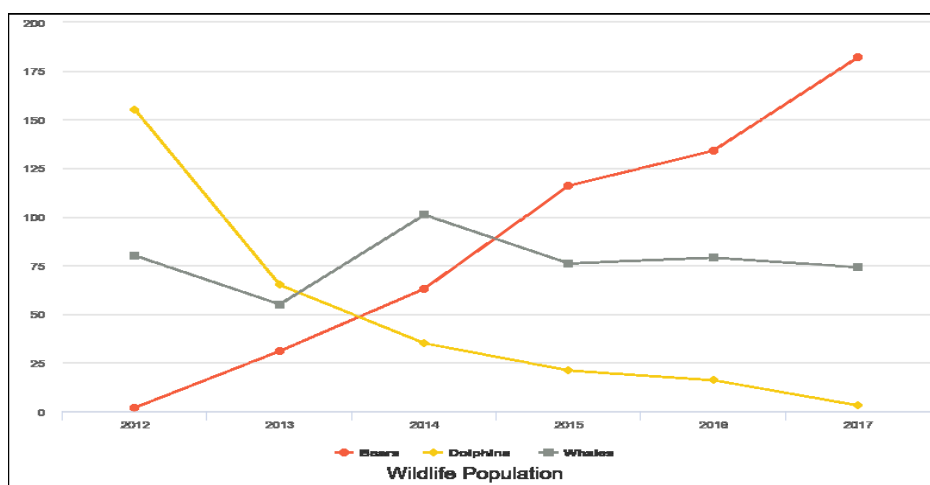


ภาพที่ 2.5 แสดง Stacked Bar Chart

5) TIME-SERIES

Line Chart ใช้เปรียบเทียบข้อมูลเพื่อดูแนวโน้ม (Trends) โดยอาจจะเทียบกับมิติของเวลา

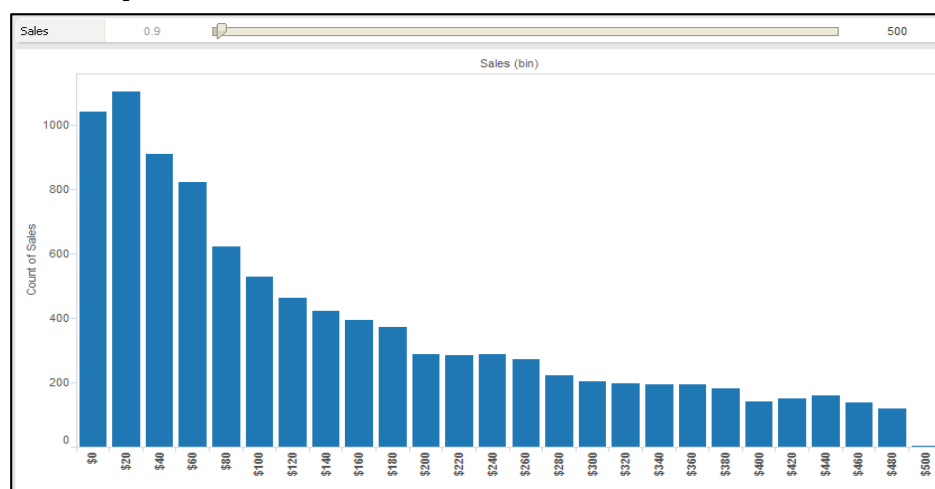
Time series



ภาพที่ 2.6 แสดง Stacked Bar Chart

6) DISTRIBUTION

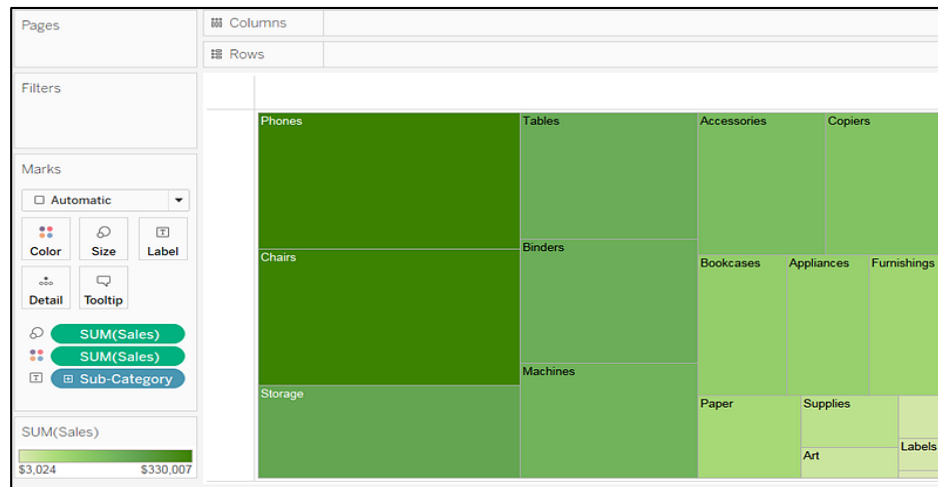
Histogram กราฟแท่งแบบเฉพาะที่แสดงความสัมพันธ์ระหว่างข้อมูลเป็นหมวดหมู่ที่เรียกว่า ชั้นข้อมูลกับความถี่ของข้อมูล เพื่อดูการกระจายของข้อมูล ลดระยะห่างระหว่างแท่งเพื่อให้เห็น “รูปทรง” ของข้อมูลชัดเจน



ภาพที่ 2.7 แสดง Histogram

7) COMPOSITION

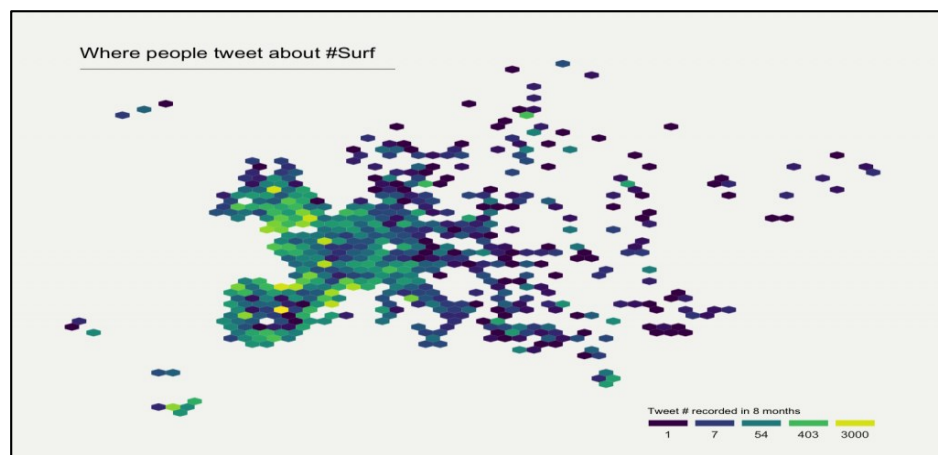
Tree map ใช้แสดงสัดส่วนของข้อมูลแบบมีลำดับชั้น (hierarchy) แสดงเป็นรูปสี่เหลี่ยม โดยจะแสดงทั้งในแง่ของสีเดียวกัน จัดเป็นกลุ่มเดียวกัน อาจอ่านสัดส่วนยากเมื่อมีกลุ่มขนาดเล็กจำนวนมาก



ภาพที่ 2.8 แสดง Tree map

8) SPATIAL

Hexbin Map เป็นการแสดงผลข้อมูลด้วยรูปทรงหกเหลี่ยมที่มีขนาดเท่ากัน โดยกำหนดสีแทนระดับความถี่ของจำนวนข้อมูล ส่วนมากนำไปใช้ลงบนตำแหน่งทางภูมิศาสตร์ เพื่อให้เห็นภาพได้ชัดเจน



ภาพที่ 2.9 แสดง Hexbin Map

จากที่ได้กล่าวไปจะเห็นได้ว่าข้อมูลแบบ Data Visualization มีพลังทางการตลาดอย่างมาก เพราะสามารถแสดงผลและเพิ่มความสนใจได้เป็นอย่างดี นอกจากนี้ยังทำให้ผู้ได้รับข้อมูลสามารถจินตนาการตามข้อมูลที่แสดงได้อีกด้วย หากองค์กรได้ลองนำการแสดงผลข้อมูลแบบ Data Visualization มาใช้ก็จะช่วยในการจัดการระดับสายงานแผนกอื่น ๆ ในบริษัท ส่งผลให้การวิเคราะห์ข้อมูลและตัดสินใจในสายงานต่าง ๆ มีประสิทธิภาพ ซึ่งดีต่อธุรกิจและทีม หรือนำมาใช้ในหน้าแรกของเว็บไซต์ เพื่อทำการโปรโมทโปรโมชันหรือข้อเสนอทางการตลาดของธุรกิจ แล้วจะพบว่า Data Visualization ช่วยให้เกิดการตอบสนองกับข้อความมากขึ้น

2.2 ทฤษฎี

2.2.1 ทฤษฎีเกี่ยวกับการจัดการข้อมูลขนาดใหญ่ (Big Data)

2.2.1.1 ข้อมูลขนาดใหญ่ (Big Data)

ข้อมูลขนาดใหญ่ (Big Data) ซึ่งสามารถนำไปประยุกต์ใช้งานให้เป็นประโยชน์ต่อองค์กรต่าง ๆ สำหรับภาคเอกชนทั่วไปก็ต้องนำมาใช้ให้เกิดประโยชน์ต่อองค์กรมากที่สุด วัตถุประสงค์การใช้งานเพื่อองค์กรของตนเป็นหลัก แต่นอกจาก Big Data จะมีความสำคัญกับภาคเอกชนแล้วสำหรับภาครัฐนั้นก็มีความสำคัญอย่างยิ่งเช่นกัน เพราะรัฐจำเป็นต้องใช้ข้อมูลของ Big Data มากก่อให้เกิดประโยชน์ต่อประชาชนทุกคน เพื่อจะกระจาย ความช่วยเหลือไปถึงทุกที่ พร้อมทั้งสามารถนำข้อมูลที่มีการวิเคราะห์ไปใช้ประโยชน์ ในการบริหาร วางแผนกำหนดทิศทางในด้านต่าง ๆ อย่างมีประสิทธิภาพที่สูงขึ้น ข้อมูลที่มีขนาดใหญ่ ซึ่งอาจจะแบ่งย่อยออกเป็นข้อมูล 3 รูปแบบคือ

1) ข้อมูลเชิงโครงสร้าง (structured data) คือตารางข้อมูลที่มีการจัดเรียงอย่างมีรูปแบบชัดเจนและเป็นระเบียบ สามารถนำมาใช้วิเคราะห์ได้เลย เช่น ข้อมูลที่เก็บไว้ในโปรแกรม spreadsheet อย่าง Microsoft Excel

2) ข้อมูลกึ่งโครงสร้าง (semi-structured data) คือข้อมูลที่ถูกจัดเก็บอย่างมีรูปแบบในระดับหนึ่งและข้อมูลที่สามารถค้นหา (search) หรือแท็ก (tag) ได้ เช่น เว็บเพจที่มีการระบุชื่อเพจ คำสำคัญในเพจ และวันที่อัปเดตข้อมูล

3) ข้อมูลที่ไม่มีโครงสร้างแน่นอน (Unstructured data) มักจะหมายถึงข้อความยาวๆ ข้อความเสียง รูปภาพ และวิดีโอ

บางองค์กรอาจจะสามารถใช้ประโยชน์จากข้อมูลที่ไม่ได้ใหญ่มาก แต่มีความสำคัญ สามารถนำมาใช้วิเคราะห์ เพื่อให้เกิดความเข้าใจเชิงลึก (insight) ทำให้ได้มาซึ่งความรู้ความเข้าใจใหม่ๆ ที่ไม่สามารถเห็นได้อย่างตรงไปตรงมา หรือเป็นข้อมูลที่รวบรวมจัดเก็บไว้ตลอดกระบวนการเชิงธุรกิจ ที่สามารถนำมาวิเคราะห์ร่วมกันได้เห็นภาพชัดเจน เช่น การจัดระบบข้อมูลเกี่ยวกับการผลิต รวบรวมข้อมูลไว้ทุกปัจจัย ตั้งแต่ข้อมูลวัตถุดิบ ถึงข้อมูลของเครื่องจักรแต่ละกระบวนการ ไปจนถึงข้อมูลผลผลิต และการขนส่ง ทำให้สามารถวิเคราะห์ข้อมูลได้ตลอดทั้งกระบวนการทฤษฎีเกี่ยวกับการทำเหมืองข้อมูล

โดยทั่วไปลักษณะพื้นฐานของ Big Data จะมีอยู่ 3 ประการ คือ ปริมาณ (Volume), ความหลากหลาย (Variety) และ ความเร็ว (Velocity) แต่วันนี้เราจะมาพูดถึงคุณลักษณะที่เพิ่มขึ้นมาอีก 3 ประการ ที่ทำให้ข้อมูลขนาดใหญ่ (Big Data) นี้มีคุณภาพสูงขึ้นและเหมาะที่จะนำมาใช้ในการวิเคราะห์เพื่อให้ได้มาซึ่งความเข้าใจเชิงลึก Big data ที่มีคุณภาพสูงควรมีลักษณะพื้นฐานอยู่ 6 ประการหลักๆ (6 Vs) ดังนี้

1) ปริมาณ (Volume) หมายถึง ปริมาณของข้อมูลควรมีจำนวนมากพอ ทำให้เมื่อนำมาวิเคราะห์แล้วจะได้ Insights ที่ตรงกับความเป็นจริง เช่น การที่เรามีข้อมูลอายุ เพศ ของลูกค้าส่วนใหญ่ ทำให้เราสามารถหา Demographic Profile ทั่วไปของลูกค้าที่ถูกต้องได้ ถ้าเรามีข้อมูลลูกค้าแค่ส่วนน้อย ค่าที่ประมาณออกมาอาจจะไม่ตรงกับความเป็นจริง

2) ความหลากหลาย (Variety) หมายถึง รูปแบบของข้อมูลควรหลากหลายแตกต่างกันออกไป ทั้งแบบโครงสร้าง, กึ่งโครงสร้าง, ไม่มีโครงสร้าง ทำให้เราสามารถนำมาวิเคราะห์ประกอบกัน จนได้ insights ครบถ้วน

3) ความเร็ว (Velocity) หมายถึง คุณลักษณะข้อมูลที่ถูกสร้างขึ้นอย่างรวดเร็ว ต่อเนื่องและทันเหตุการณ์ ทำให้เราสามารถวิเคราะห์ข้อมูลแบบ real-time นำผลลัพธ์มาทำการตัดสินใจและตอบสนองได้อย่างทันท่วงที เช่น ข้อมูล GPS ที่ใช้ติดตามตำแหน่งของรถ อาจจะสามารถวิเคราะห์โอกาสที่ทำให้เกิดอุบัติเหตุ และออกแบบระบบป้องกันอุบัติเหตุได้

4) ความถูกต้อง (Veracity) หมายถึง มีความน่าเชื่อถือของแหล่งที่มาข้อมูล และความถูกต้องของชุดข้อมูล มีกระบวนการในการตรวจสอบและยืนยันความถูกต้องของข้อมูล ซึ่งมีความเกี่ยวเนื่องโดยตรงกับผลลัพธ์การวิเคราะห์ข้อมูล

5) คุณค่า (Value) หมายถึง ข้อมูลมีประโยชน์และมีความสัมพันธ์ในเชิงธุรกิจ ซึ่งต้องเข้าใจก่อนว่าไม่ใช่ทุกข้อมูลจะมีประโยชน์ในการเก็บและวิเคราะห์ ข้อมูลที่มีประโยชน์

จะต้องเกี่ยวข้องกับวัตถุประสงค์ทางธุรกิจ เช่นถ้าต้องการเพิ่มขีดความสามารถในการแข่งขันในตลาดของผลิตภัณฑ์ที่ขาย ข้อมูลที่มีประโยชน์ที่สุดน่าจะเป็นข้อมูลผลิตภัณฑ์ของคู่แข่ง

6) ความแปรผันได้ (Variability) หมายถึง ข้อมูลสามารถในการเปลี่ยนแปลงรูปแบบไปตามการใช้งาน หรือสามารถคิดวิเคราะห์ได้จากหลายแง่มุม และรูปแบบในการจัดเก็บข้อมูลก็อาจจะต่างกันออกไปในแต่ละแหล่งของข้อมูล

คุณลักษณะเหล่านี้ทำให้การจัดการ Big Data เป็นเรื่องยาก แต่ในขณะเดียวกันก็เป็นโอกาสให้องค์กรสร้างความได้เปรียบเหนือคู่แข่งด้วยการพัฒนาขีดความสามารถในการวิเคราะห์ข้อมูล

2.2.1.2 การจัดการข้อมูลขนาดใหญ่

การรวบรวมข้อมูลมาจัดเก็บ (Storage) เป็นการรวบรวมข้อมูลของจากหลากหลายทั้งที่มา และการใช้งานที่แตกต่างกันอย่างมากมาย ซึ่งกลไกและเทคโนโลยีแบบดั้งเดิม ETL (Extract Transform and Load) ไม่สามารถทำได้ ซึ่ง Big Data หรือ ข้อมูลขนาดใหญ่ต้องการเทคนิค วิธีการ และเทคโนโลยีใหม่ในการรวบรวมข้อมูลขนาด เทราไบต์ และเพตาไบต์ ในการรวบรวมข้อมูลนั้นต้องมีการประมวลผล จัดรูปแบบ ให้เหมาะสำหรับการใช้ในการวิเคราะห์หรือใช้งานสำหรับธุรกิจหรือวัตถุประสงค์นั้น ๆ

1) การประมวลผล (Processing) ข้อมูลขนาดใหญ่ หรือ Big Data นั้นมีความต้องการสถานที่จัดเก็บขนาดใหญ่ การจัดเก็บข้อมูลขนาดใหญ่จะเป็นชนิดใดก็ได้ไม่ว่าจะเป็นแบบ On-premises หรือแบบ Cloud ขึ้นกับความต้องการ หรือความสะดวกในการใช้ ซึ่งเราสามารถใช้งานและประเมินผลได้เช่นเดียวกัน บางครั้งก็มีความจำเป็นที่ต้องจัดเก็บไว้ใกล้กับแหล่งข้อมูล หรือข้อมูลบางอันต้องการความยืดหยุ่นสูง และไม่ต้องการบริหารจัดการก็ใช้เป็นแบบ Cloud ซึ่งกำลังเป็นที่นิยมกันเป็นอย่างมาก

2) การวิเคราะห์ (Analyst) การวิเคราะห์ข้อมูลทำให้เกิดความกระจ่าง และชัดเจนในชุดข้อมูลที่มีอยู่ในการสำรวจข้อมูลยังทำให้ค้นพบสิ่งใหม่ แชร์สิ่งทีค้นพบใหม่ ๆ ต่อผู้อื่น สร้างรูปแบบจำลองข้อมูลด้วยการเรียนรู้ของเครื่องจักร และปัญญาประดิษฐ์ AI และนำข้อมูลเหล่านั้นไปใช้งาน

2.2.2 ทฤษฎีเกี่ยวกับการทำเหมืองข้อมูล

การทำเหมืองข้อมูลหรือ Data Mining เทคโนโลยีเกี่ยวกับฐานข้อมูลได้เริ่มต้นขึ้นตั้งแต่ปี 1960 โดยพัฒนามาจาก File Processing จากนั้นพัฒนามาในรูปแบบตาราง และเริ่มมีการใช้ภาษาในการเรียกข้อมูล (Query Language) ข้อมูลที่เก็บเริ่มมีจำนวนมากขึ้น ทำให้การจัดเก็บมีความซับซ้อน และมีประสิทธิภาพเพิ่มขึ้นของข้อมูลเป็นจำนวนมาก และสะสมไว้ในฐานข้อมูลนั้นมีขนาดใหญ่เกินกว่าที่กำลังคนจัดการได้ เป็นผลให้เกิดเครื่องมือที่ช่วยในการวิเคราะห์ข้อมูล และหาความเป็นไปได้ของข้อมูลที่เป็นประโยชน์

เนื่องด้วยปัจจุบันเป็นยุคที่ข้อมูลสารสนเทศมีความสำคัญ การเผยแพร่และสื่อสารข้อมูล ข่าวสาร ที่ตรงกับความต้องการของผู้ใช้จึงเป็นสิ่งจำเป็น การประยุกต์เทคโนโลยีสารสนเทศเพื่อช่วยในการสื่อสารข้อมูลจำนวนมากให้แก่ผู้ใช้ เช่น การให้บริการเว็บไซต์เพื่อเผยแพร่ข้อมูล ข่าวสาร และแลกเปลี่ยนความรู้ จึงเป็นเครื่องมือที่สำคัญในการสื่อสารข้อมูลถึงผู้ใช้งานจำนวนมาก ดังนั้น การศึกษาเกี่ยวกับพฤติกรรมของผู้ใช้บริการเว็บไซต์จะช่วยให้องค์กรสามารถนำข้อมูลมาใช้ในการวางแผนพัฒนาเว็บไซต์ ให้ตรงกับความต้องการใช้งาน หรือใช้ในการวางแผนกลยุทธ์ เพื่อสร้างความได้เปรียบทางการแข่งขัน

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

การทำเหมืองข้อมูลหากผนวกเข้ากับระบบการจัดการความรู้ที่ได้จากฐานข้อมูล จะทำให้ได้ระบบการค้นหาข้อมูลและความรู้ หรือเรียกว่า (Knowledge and Data Discovery Management System : KDDMS) เป็นการทำเหมืองข้อมูลที่ใช้เทคนิคการกลั่นกรอง วิเคราะห์ความรู้ที่ซ่อนอยู่ในฐานข้อมูลที่มีปริมาณมาก เพื่อได้ข้อมูล ที่ต้องการ และมีประโยชน์ สามารถนำข้อมูลที่ได้รับการจัดการได้อย่างมีประสิทธิภาพ โดยผลลัพธ์มีอยู่หลายแบบ เช่น รูปแบบ (Pattern) ต้นแบบ (Model) กฎ (Rule) การทำนายหรือการคาดการณ์ล่วงหน้า (Prediction) ทั้งผลลัพธ์ที่ได้มาขึ้นอยู่กับการทำเหมืองข้อมูล

2.2.3 ทฤษฎีเกี่ยวกับการออกแบบเว็บไซต์

การออกแบบเว็บไซต์นั้นมีความสำคัญอย่างมาก เนื่องจากจะทำให้ผู้ชมเกิดความประทับใจ ซึ่งชอบ จนสามารถดึงดูดความสนใจได้จนกระทั่งมีความต้องการที่จะกลับเข้าชมเว็บนั้น ๆ อีกครั้งในอนาคต สำหรับการออกแบบเว็บไซต์นั้น หมายถึง การออกแบบพัฒนาส่วนหน้าแสดงของเว็บให้เกิดความประทับใจแก่ผู้ชม และนำเสนอความเป็นเอกลักษณ์ของหน่วยงานได้ด้วยโปรแกรมฟังก์ชันการใช้งานภายในเว็บไซต์ที่ทำให้ผู้ใช้อายากเข้ามามีปฏิสัมพันธ์ต่อเว็บไซต์นั้น โดยต้องให้ความสำคัญกับปัจจัยต่าง ๆ ดังต่อไปนี้

2.2.3.1 ความสำคัญของการออกแบบเว็บไซต์

เว็บไซต์เป็นสื่อที่อยู่ในการควบคุมของผู้ใช้โดยสมบูรณ์ ดังนั้นการออกแบบเว็บจึงมีความสำคัญที่จะดึงดูดให้ผู้ใช้เข้าชม และทำกิจกรรมร่วมในเว็บไซต์ของเราให้ได้มากที่สุด ซึ่งสามารถสรุปปัญหาต่าง ๆ เกี่ยวกับผู้ใช้ซึ่งเป็นองค์ประกอบหนึ่งในการกำหนดความสำเร็จ หรือล้มเหลวของการสร้างเว็บไซต์ ดังต่อไปนี้

- 1) ผู้ใช้มักไม่อดทนกับอุปสรรค และปัญหาที่เกิดขึ้นจากการออกแบบการใช้งานผิดพลาด
- 2) ผู้ใช้มีทางเลือกมากขึ้นที่จะเข้าใช้บริการเว็บไซต์อื่น ๆ ถ้าหากการออกแบบเว็บไซต์นั้นเกิดความรู้สึกไม่ประทับใจ
- 3) ผู้ใช้หลายคนชื่นชอบเว็บไซต์ที่สวยงาม และมีความสะดวกในการใช้งานมากกว่าเว็บที่ดูสับสนวุ่นวาย มีข้อมูลจำนวนมากแต่ค้นหาข้อมูลที่ต้องการ ไม่พบ หรือใช้เวลามากในการแสดงหน้าเว็บ
- 4) หากผู้ใช้มีความประทับใจกับเว็บไซต์หนึ่งแล้ว จะทำให้อยากกลับมาเข้าใช้อีกครั้ง

2.2.3.2 องค์ประกอบของการออกแบบเว็บไซต์อย่างมีประสิทธิภาพ

การออกแบบเว็บไซต์ที่ดีนั้น จะต้องมียุทธศาสตร์ประกอบในการออกแบบให้เหมาะสมกับกลุ่มเป้าหมาย และจะต้องคำนึงถึงความสะดวกของผู้ใช้งานเป็นหลัก โดยที่ผู้ออกแบบนั้น ต้องให้ความสมดุลระหว่างความสวยงามและความสะดวก ซึ่งต้องนึกถึงองค์ประกอบดังต่อไปนี้

- 1) ความเรียบง่าย (Simplicity) คือ การสื่อเนื้อหาถึงผู้ใช้โดยจำกัดองค์ประกอบเสริมที่เกี่ยวข้องกับการนำเสนอให้เหลือแต่สิ่งที่จำเป็นเท่านั้น ทั้งการแสดงผลออกใน

ด้านเนื้อหาตัวอักษร หรือแม้แต่สีที่ใช้ ทำให้ผู้ใช้รู้สึกใช้งานง่ายและสะดวกไม่ก่อให้เกิดความรู้สึกรำคาญใด ๆ

2) ความสม่ำเสมอ (Consistency) คือ การสื่อให้ผู้ใช้รู้ว่าเป็นเนื้อหาเดียวตลอดทั้งเว็บไซต์ โดยไม่ก่อให้เกิดความสับสนแก่ผู้ใช้ ดังนั้น ผู้ออกแบบเว็บไซต์ควรออกแบบระบบ Navigation และโทนสีให้มีความสม่ำเสมอคล้ายคลึงกันตลอดทั้งเว็บไซต์

3) ความเป็นเอกลักษณ์ (Identity) คือ ต้องออกแบบเพื่อแสดงถึงความเป็นเอกลักษณ์ขององค์กร ทั้งชุดสี ตัวอักษร รูปภาพกราฟิก เป็นสื่อถึงความเป็นเอกลักษณ์แทบทั้งสิ้น

4) เนื้อหาที่ประประโยชน์ (Useful content) เนื้อที่แสดงออกของเว็บไซต์นั้น เป็นสิ่งสำคัญที่สุด ซึ่งต้องแสดงออกถึงความถูกต้องและสมบูรณ์ โดยไม่ซ้ำกับเว็บไซต์อื่น และเป็นเนื้อหา ที่ทันต่อเหตุการณ์ เนื่องจากเป็นสิ่งที่ดึงดูดผู้เข้ามาใช้งานเว็บไซต์

5) ระบบ Navigation ที่ใช้งานง่าย (User-friendly navigation) เป็นส่วนที่สำคัญของเว็บไซต์ เนื่องจากจะทำให้ผู้ใช้ไม่เกิดความสับสนเวลาใช้งานเว็บไซต์ ซึ่งผู้ออกแบบจำเป็นต้อง ออกแบบให้ผู้ใช้ใช้งานได้ง่ายและสะดวก เช่นถ้าใช้ภาพกราฟิกที่สื่อความหมายควรมีคำอธิบายที่ชัดเจนร่วมด้วย แสดงไว้ส่วนบนหรือส่วนล่างของหน้าเว็บไซต์

6) มีลักษณะที่น่าสนใจ (Visual appeal) ควรสื่อให้เว็บไซต์มีคุณภาพสัมพันธ์กับองค์ประกอบต่าง ๆ เช่น คุณภาพของกราฟิกที่แสดงบนเว็บไซต์นั้นต้องสมบูรณ์ไม่มีร่องรอยของความเสียหายเป็นจุดต่าง หรือมีขอบขั้นบันได สำหรับตัวอักษรต้องเป็นตัวอักษรที่อ่านง่าย และสบายตาพร้อมกับโทนสีที่สวยงาม

7) การใช้งานได้อย่างไม่จำกัด (Compatible) การออกแบบเว็บไซต์นั้นต้องทำให้คนส่วนใหญ่เข้าถึงให้ได้มากที่สุด โดยไม่มีการบังคับให้ติดตั้งโปรแกรมใด ๆ เพิ่มเติม สามารถแสดงผลได้ทุกระบบปฏิบัติการ และที่ความละเอียดของหน้าจอต่าง ๆ กัน

8) คุณภาพการออกแบบ (Design Stability) ต้องออกแบบให้เว็บไซต์มีเนื้อหาที่น่าเชื่อถือ เพื่อดึงดูดให้มีผู้เข้าชมเว็บไซต์เกิดความประทับใจ และมีความเชื่อถือในตัวเว็บไซต์

9) ระบบการใช้งานที่ถูกต้อง (Functional stability) ต้องออกแบบให้เว็บไซต์นั้นมีความแน่นอนทำหน้าที่ได้อย่างถูกต้อง และสามารถใช้งานได้จริง เช่น ถ้ามีแบบฟอร์มให้ผู้ใช้

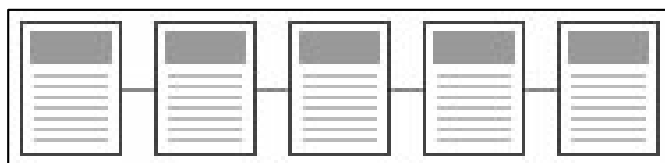
กรอกข้อมูล ก็ต้องทดสอบก่อนว่าสามารถใช้งานได้จริง หรือลิงค์ต่าง ๆ ที่แสดงอยู่บนหน้าเว็บไซต์ จะต้องเชื่อมโยงไปยังแหล่งข้อมูลนั้น ๆ ได้จริง

2.2.3.3 รูปแบบโครงสร้างของเว็บไซต์

การออกแบบโครงสร้างของเว็บไซต์ สามารถทำได้หลากหลายแบบซึ่งก็ขึ้นอยู่กับความชอบ และความถนัดของแต่ละบุคคล นอกจากนี้ยังขึ้นอยู่กับกลุ่มเป้าหมายที่ต้องการนำเสนอ เพราะจะต้องออกแบบให้เหมาะกับการใช้งานของกลุ่มเป้าหมายมากที่สุด โดยโครงสร้างของเว็บไซต์ส่วนใหญ่ก็จะประกอบไปด้วย 4 รูปแบบดังนี้

1) โครงสร้างเว็บไซต์แบบเรียงลำดับ

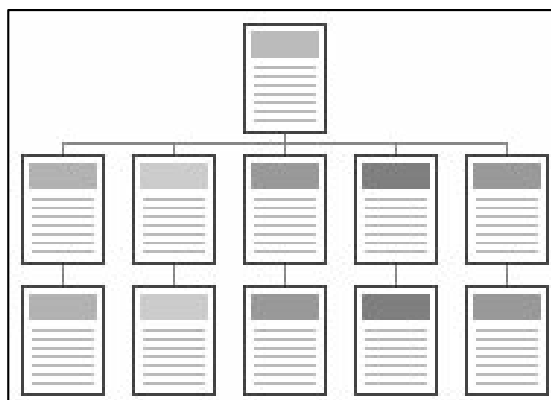
จะเป็นโครงสร้างแบบธรรมดาที่นิยมใช้งานกันมากที่สุด เนื่องจากมีความง่ายต่อการจัดระบบข้อมูล และสามารถนำเสนอเรื่องราวตามลำดับได้เป็นอย่างดี เหมาะกับเว็บไซต์ที่มีขนาดเล็ก มีเนื้อหาที่ไม่ซับซ้อน ส่วนใหญ่ก็จะเป็นพวกเว็บไซต์ที่ให้ความรู้ หรือเว็บไซต์องค์กรขนาดย่อม โดยลักษณะการลิงค์เนื้อหา ก็จะลิงค์ไปที่ละหน้า มีทิศทาง การเข้าสู่เนื้อหาต่าง ๆ ในแบบเส้นตรง ใช้ปุ่มเดินหน้า-ถอยหลัง ในการกำหนดทิศทาง จึงทำให้การใช้งานเป็นไปอย่างง่าย แต่โครงสร้างเว็บไซต์แบบเรียงลำดับก็มีข้อเสีย คือจะทำให้ผู้ใช้งานต้องเสียเวลาในการเข้าสู่เนื้อหา เพราะไม่สามารถกำหนดทิศทาง การเข้าสู่เนื้อหาด้วยตัวเองได้



ภาพที่ 2.10 แสดงโครงสร้างเว็บไซต์แบบเรียงลำดับ

2) โครงสร้างเว็บไซต์แบบลำดับชั้น

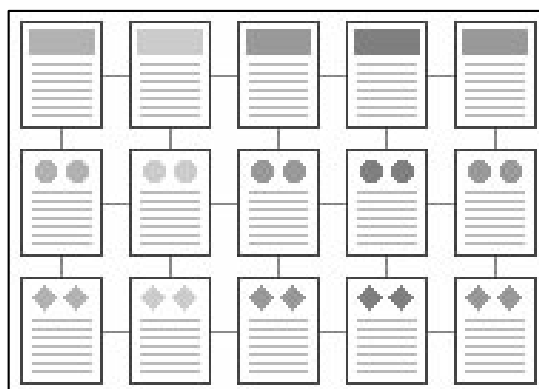
นิยมใช้กับเว็บที่มีความซับซ้อนของข้อมูล เพื่อให้สามารถเข้าถึงข้อมูลต่าง ๆ ได้ง่ายขึ้น โดยจะมีการแบ่งเนื้อหาออกเป็นส่วน ๆ และมีการนำเสนอรายละเอียดค่อย ๆ ที่ลดหลั่นกันมา ทำให้สามารถทำความเข้าใจกับโครงสร้างเนื้อหาได้ง่ายขึ้น โดยจะมีโฮมเพจเป็นจุดเริ่มต้น และจุดรวมจุดเดียวที่จะนำไปสู่การเชื่อมโยงเนื้อหาเป็นลำดับจากบนลงล่าง



ภาพที่ 2.11 แสดงโครงสร้างเว็บไซต์แบบลำดับชั้น

3) โครงสร้างเว็บไซต์แบบตาราง

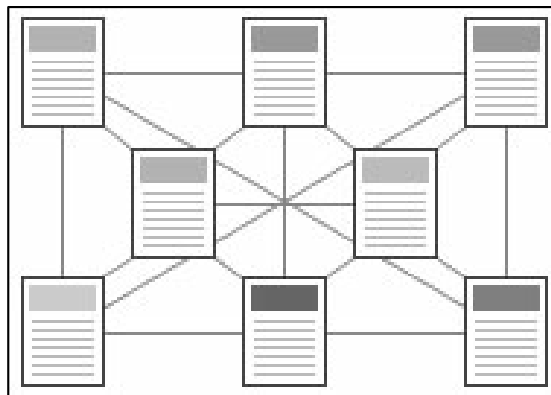
เป็นโครงสร้างการออกแบบเว็บไซต์ที่มีความซับซ้อน แต่ก็มีคามยืดหยุ่นในระดับหนึ่ง เพื่อให้ผู้ใช้งานสามารถเข้าสู่เนื้อหาต่าง ๆ ได้ง่ายขึ้น การออกแบบในลักษณะนี้จะมีการเชื่อมโยงเนื้อหาในแต่ละส่วนซึ่งกันและกัน ทำให้ผู้ใช้งานสามารถเปลี่ยนทิศทาง หรือกำหนดทิศทางในการเข้าสู่เนื้อหาด้วยตัวเองได้จึงไม่ทำให้เสียเวลา แถมยังทำให้เว็บไซต์มีความทันสมัยขึ้น



ภาพที่ 2.12 แสดงโครงสร้างเว็บไซต์แบบตาราง

4) โครงสร้างแบบใยแมงมุม

เป็นโครงสร้างที่ได้รับความนิยมเป็นอย่างมาก เพราะมีความยืดหยุ่นมากที่สุด โดยทุกหน้าเว็บจะมีการเชื่อมโยงถึงกันหมด ทำให้สามารถเข้าถึงหน้าเว็บเพจต่าง ๆ ที่ต้องการได้อย่างง่าย และมีความอิสระมากขึ้น นอกจากนี้ก็สามารถเชื่อมโยงไปสู่เว็บไซต์ภายนอกได้ดี



ภาพที่ 2.13 แสดงโครงสร้างเว็บไซต์แบบใยแมงมุม

2.2.3.4 ส่วนประกอบสำคัญของหน้าเว็บเพจบนหน้าเว็บเพจ จะมีส่วนประกอบสำคัญที่จำเป็นต้องมีอยู่ 3 ส่วน ได้แก่

1) ส่วนหัวของหน้า (Header) อยู่ตอนบนสุดของหน้าและเป็นส่วนที่สำคัญที่สุด โดยจะต้องทำให้สามารถดึงดูดผู้ชมให้รู้สึกอยากติดตามเนื้อหาในเว็บไซต์ต่อไป ซึ่งส่วนใหญ่มักจะมีการใส่ภาพกราฟิกให้ดูสวยงาม สิ่งสำคัญหลัก ๆ เลย ก็คือ โลโก้ ชื่อเว็บไซต์ และเมนูหลักที่สามารถลิงค์ไปยังเนื้อหาในหน้าเว็บเพจต่าง ๆ ได้

2) ส่วนของเนื้อหา (Body) อยู่บริเวณตอนกลางของหน้าเว็บ โดยจะแสดงข้อมูลเกี่ยวกับเนื้อหาบนเว็บแบบคร่าวๆ ซึ่งก็จะมีข้อความกราฟิก ตารางข้อมูลหรือวิดีโอประกอบอยู่ และหากมีเมนูแบบเฉพาะกลุ่มก็จะถูกจัดไว้ในหน้านี้เช่นกัน และที่สำคัญเนื้อหาในส่วนนี้ควรจะมีคามกระชับ เข้าใจง่าย มีการใช้รูปแบบตัวอักษรแบบเรียบง่ายและเป็นระเบียบ

3) ส่วนท้ายของหน้า (Footer) อยู่ล่างสุดของหน้าเว็บ ซึ่งจะมีหรือไม่มีก็ได้ ส่วนนี้จะแสดงถึงข้อมูลต่าง ๆ เพิ่มเติมเข้าไป เช่น ข้อความที่แสดงถึงการเป็นลิขสิทธิ์ ข้อมูลเจ้าของเว็บไซต์ วิธีการติดต่อและคำแนะนำต่าง ๆ เกี่ยวกับการใช้งานเว็บไซต์อย่างถูกต้อง เป็นต้น

2.2.3.5 วิธีการเลือกใช้สีสำหรับการออกแบบเว็บไซต์

การเลือกใช้สีในการออกแบบเว็บไซต์มีความสำคัญเป็นอย่างมาก เพราะสีสามารถกำหนดอารมณ์ ความรู้สึก และกระตุ้นการรับรู้ทางด้านจิตใจของมนุษย์ได้ดี ดังนั้น สีที่ใช้จึงต้องมีความสอดคล้องกับเนื้อหาและจุดประสงค์ของเว็บไซต์ว่าต้องการให้ผู้เข้าชมรู้สึกอย่างไร ต่อเนื้อหาที่ได้อ่าน โดยรูปแบบของสีที่สายตาของมนุษย์สามารถมองเห็นได้ก็แบ่งออกเป็น 3 กลุ่มดังต่อไปนี้

1) สีโทนร้อน (Warm Colors) เป็นสีแห่งความอบอุ่น ปลอดภัยและกระตุ้นความสุขได้ดี ซึ่งจะทำให้ผู้เข้าชมรู้สึกมีชีวิตชีวาและมีแรงผลักดันมากขึ้น อีกทั้งยังช่วยดึงดูดให้ผู้ชมรู้สึกอยากติดตามเนื้อหามากขึ้น

2) สีโทนเย็น (Cool Colors) เป็นสีแห่งความสงบและความอ่อนโยน ทำให้ผู้ชมรู้สึกผ่อนคลายและเพลิดเพลินมากขึ้น และยังสามารถใช้น้ำหนักจากในระยะไกลได้อีกด้วย

3.) สีโทนกลาง (Neutral Colors) สีเหล่านี้มักจะถูกนำไปผสมกับสีอื่น ๆ เพื่อให้เกิดสีที่เป็นกลางมากขึ้น และให้ความรู้สึกที่เป็นธรรมชาติ

สีสามารถสื่อถึงอารมณ์ความรู้สึกและสื่อความหมายของเว็บไซต์นั้น ๆ ได้อย่างชัดเจน ดังนั้น หากเลือกใช้สีไม่เหมาะสมกับเนื้อหาหรือจุดประสงค์ของเว็บไซต์ ก็จะทำให้เว็บดูไม่น่าสนใจ ขาดความน่าเชื่อถือและทำให้ผู้ที่เคยเข้ามาใช้บริการไม่คิดจะกลับมาใช้บริการอีก

2.2.4 ทฤษฎีเกี่ยวกับชุดคำสั่ง HTML

HTML ย่อมาจาก Hyper Text Markup Language คือ ภาษาคอมพิวเตอร์ที่ใช้ในการแสดงผลของเอกสารบน Website หรือที่เราเรียกกันว่าเว็บเพจ ถูกพัฒนาและกำหนดมาตรฐานโดยองค์กร World Wide Web Consortium (W3C) และจากการพัฒนาทางด้าน Software ของ Microsoft ทำให้ภาษา HTML เป็นอีกภาษาหนึ่งที่ใช้เขียนโปรแกรมได้หรือที่เรียกว่า HTML Application

HTML เป็นภาษาประเภท Markup สำหรับการการสร้างเว็บเพจ โดยใช้ภาษา HTML สามารถทำได้โดยใช้โปรแกรม Text Editor ต่าง ๆ เช่น Notepad, Edit plus หรือจะอาศัยโปรแกรมที่เป็นเครื่องมือช่วยสร้างเว็บเพจ เช่น Microsoft FrontPage, Dream Weaver ซึ่งอำนวยความสะดวกในการสร้างหน้า HTML ส่วนการเรียกใช้งานหรือทดสอบการทำงานของเอกสาร HTML จะใช้

โปรแกรม Web Browser เช่น IE Microsoft Internet Explorer (IE), Mozilla Firefox, Safari, Opera, และ Netscape Navigator เป็นต้น

2.2.4.1 ไวยากรณ์ของ CSS

ภาษา HTML จะแบ่งออกเป็น 2 ส่วน คือ

1) ส่วนของคำสั่ง (Tag) เป็นส่วนที่กำหนดรูปแบบของข้อความที่แสดง ซึ่งเราเรียกว่า Tag โดยจะอยู่ในเครื่องหมาย < ... >

2) ส่วนของบทความทั่ว ๆ ไป เป็นส่วนของข้อความที่เราต้องการแสดงผล ตัวอย่างการใช้งานภาษา HTML

```
<html>¶
<head>¶
    <title>·หัวข้อเรื่อง·ของหน้านี้·</title>¶
</head>¶
<body>¶
¶
    เนื้อหาที่จะแสดงใน·Web·Browser¶
¶
</body>¶
</html>¶
```

ภาพที่ 2.14 ไวยากรณ์ของ HTML

การทำงานของ Code

1) <html> </html> ในการใช้งาน HTML เราจะต้องเริ่มด้วย <html> และปิดด้วย </html> เสมอ

2) <head> </head> เป็นส่วนที่ใช้ให้รายละเอียดเกี่ยวกับเว็บเพจ หน้านี้ ซึ่งจะไม่แสดงให้เห็นในส่วนของการแสดงผลของ Web Browser แต่จะมีผลกับส่วนอื่น ๆ เช่น

การหาของ Search Engine (Google, Yahoo) การใช้งานก็จะมีคำสั่งย่อยเพื่อบรรยายรายละเอียด เช่น <title> </title>, <meta> และอื่นอีกมากมาย

3) <title> </title> ในส่วนตัวอักษรที่อยู่ในคำสั่งนี้จะอยู่ใน title bar ของ web page

4. <body> </body> ตัวอักษรที่อยู่ในคำสั่งนี้จะแสดงส่วนแสดงผล ของ Web Browser

2.2.5 ทฤษฎีเกี่ยวกับชุดคำสั่ง CSS

CSS ย่อมาจาก Cascading Style Sheet มักเรียกโดยย่อว่า "สไตลชีต" คือ ภาษาที่ใช้เป็นส่วนของการจัดรูปแบบการแสดงผลเอกสาร HTML โดยที่ CSS กำหนดกฎเกณฑ์ในการระบุรูปแบบ หรือ Style ของเนื้อหาในเอกสาร อันได้แก่ สีของข้อความ สีพื้นหลัง ประเภทตัวอักษร และการจัดวางข้อความ ซึ่งการกำหนดรูปแบบ หรือ Style นี้ใช้หลักการของการแยกเนื้อหาเอกสาร HTML ออกจากคำสั่งที่ใช้ในการจัดรูปแบบการแสดงผล กำหนดให้รูปแบบของการแสดงผลเอกสาร ไม่ขึ้นอยู่กับเนื้อหาของเอกสาร เพื่อให้ง่ายต่อการจัดรูปแบบการแสดงผลล์พ์ของเอกสาร HTML โดยเฉพาะในกรณีที่มีการเปลี่ยนแปลงเนื้อหาเอกสารบ่อยครั้ง หรือต้องการควบคุมให้รูปแบบการแสดงผลเอกสาร HTML มีลักษณะของความสม่ำเสมอทั่วกันทุกหน้าเอกสารภายในเว็บไซต์เดียวกัน โดยกฎเกณฑ์ในการกำหนดรูปแบบ (Style) เอกสาร HTML ถูกเพิ่มเข้ามาครั้งแรกใน HTML 4.0 เมื่อปีพ.ศ. 2539 ในรูปแบบของ CSS level 1 Recommendations ที่กำหนดโดย องค์การ World Wide Web Consortium หรือ W3C

2.2.5.1 ไวยากรณ์ของ CSS

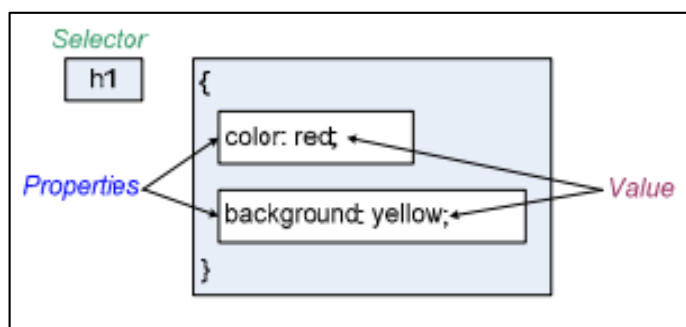
ไวยากรณ์ของ CSS นั้น มีโครงสร้างหลักของภาษาอยู่ 4 ส่วนด้วยกัน ซึ่งทั้งหมดมีความสำคัญต่อการกำหนดค่าและการใช้งาน CSS ซึ่งรูปแบบการกำหนดค่าจะประกอบไปด้วย property ตามด้วยเครื่องหมายโคลอน (:) เพื่อกำหนดค่าลงไปซึ่งอาจจะเป็นตัวเลข คีย์เวิร์ด หรือ URL และสุดท้ายของการกำหนดค่าคือ เซมิ-โคลอน (;) ที่ใช้ในการระบุว่าสิ้นสุดการกำหนดค่าให้กับproperty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

Selector [Properties: Value;]

ภาพที่ 2.15 ไวยากรณ์ของ CSS

2.2.5.2 ส่วนประกอบของโครงสร้างของ CSS มีดังนี้

- 1) Selector สามารถเป็น HTML Tag ต่างๆ ก็ได้ เช่น <body>, <p> หรือเป็น Class name 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 และ font-size:12px



ภาพที่ 2.16 แสดงซอร์ซโค้ดส่วนประกอบของโครงสร้างของ CSS

จากซอร์ซโค้ดข้างต้นนี้ สามารถพิจารณาตามรูปแบบโครงสร้างของ CSS ได้ดังนี้คือ Selector ของ CSS คือ h1, Properties ของ CSS คือ color และการใส่สีพื้นหลัง คือ Background, Value ของ CSS คือ red และ yellow

จะสังเกตเห็นได้ว่าจะมีเครื่องหมาย Colon (:) ที่เป็นตัวคั่นระหว่างชื่อ Properties กับ Value ที่เป็นการกำหนดค่าให้กับ Properties นั้น ๆ ซึ่งเมื่อสิ้นสุดแล้วแต่ละ Properties จะจบด้วย เครื่องหมาย Semi-colon (;) เสมอ ซึ่งเป็นข้อบังคับของการเขียน CSS นั่นเอง

2.2.6 ทฤษฎีเกี่ยวกับชุดคำสั่ง PHP

พีเอชพี (PHP) คือ ภาษาคอมพิวเตอร์ในลักษณะเซิร์ฟเวอร์-ไซด์ สคริปต์ โดยลิขสิทธิ์อยู่ในลักษณะโอเพนซอร์ส ภาษาพีเอชพีใช้สำหรับจัดทำเว็บไซต์ และแสดงผลออกมาในรูปแบบ HTML โดยมีรากฐานโครงสร้างคำสั่งมาจากภาษา ภาษาซี ภาษาจาวา และ ภาษาเพิร์ล ซึ่ง ภาษาพีเอชพี นั้นง่ายต่อการเรียนรู้ ซึ่งเป้าหมายหลักของภาษานี้คือให้นักพัฒนาเว็บไซต์สามารถเขียน เว็บเพจ ที่มีการตอบโต้ได้อย่างรวดเร็ว

PHP ย่อมาจากคำว่า “Personal Home Page Tool” (ปัจจุบันได้เพิ่มเติมคำย่อใหม่ โดยรวมกับตัวย่อเป็น PHP : PHP Hypertext Preprocessor) ซึ่งเป็นภาษาประเภท Script Language ที่ทำงานแบบ Server Side Script กระบวนการทำงานจะทำงานแบบโปรแกรมแปลคำสั่ง interpreter คือแปลภาษาทุกครั้งที่มีคนเรียกสคริปต์ ข้อดี คือ ไม่ต้องนำไปประมวลผลใหม่ (Compiler) เมื่อนำโปรแกรมไปใช้งาน หรือจะอัปเดตเวอร์ชันของโปรแกรม สามารถอัปเดตขึ้นไปทับไฟล์เดิมแล้วใช้งานได้ทันที ขอลือที่ต่างกันอย่างชัดเจนก็คือ กรณี Syntax ผิดจะรู้ก็ต่อเมื่อมีผู้ใช้งานเจอบั๊ก

ภาษา PHP จัดอยู่ในประเภท การเขียนโปรแกรมบนเว็บ (Web-based Programming) เพราะเราจะเก็บโค้ดคำสั่ง หรือสคริปต์ทั้งหมดที่เขียนขึ้นมาไว้บนเครื่องเซิร์ฟเวอร์ที่เดียว (Web Server) และให้ผู้ใช้งาน (Client) เรียกใช้งานโปรแกรมผ่านเว็บเบราว์เซอร์ต่าง ๆ เช่น Internet Explorer, Mozilla Firefox, Google Chrome, Opera, Safari ฯลฯ เพื่อนำข้อมูลมาแสดงผลที่หน้าจอของผู้ใช้แต่ละคนนั่นเอง

2.2.6.1 ไวยากรณ์ของ PHP

ภาษาพีเอชพี จะเป็นส่วนประกอบภายในเว็บเพจ โดยคำสั่งจะปรากฏระหว่าง `<?php ... ?>` เช่น

```
<?php
    echo "Hello, World!";
?>
```

ภาพที่ 2.17 ไวยากรณ์ของ PHP

2.2.7 ทฤษฎีเกี่ยวกับการ Visualization

Visualization เป็นส่วนประกอบสำคัญใน Cognitive System ซึ่งเป็นส่วนในการแสดงข้อมูลหรือผลลัพธ์ต่าง ๆ ในระหว่างคอมพิวเตอร์และผู้ใช้งานในรูปแบบของภาพโดยผู้ใช้งานสามารถเรียนรู้ และจดจำข้อมูลผ่านการมองเห็นได้มากกว่าการใช้ประสาทสัมผัสอื่น ๆ หรือจะกล่าวได้ว่า Visualization ก็คือ การสร้างมโนภาพของสิ่งต่าง ๆ ที่เราสนใจขึ้นมาในใจ ซึ่งต่อมาได้กลายเป็นการนำภาพมาใช้กับการนำเสนอหรือนำมาเป็นกรอบความคิดซึ่งได้นำไปใช้ในการสนับสนุนการตัดสินใจ

Visualization เป็นมากกว่า วิธีการทาง Computer Visualization เป็นการนำเสนอข้อมูลแบบหนึ่งที่ทำให้การแสดงผลให้อยู่ในรูปแบบของ Visual Form ซึ่งอาจจะเป็นการแสดงในรูปแบบของรูปภาพ, กราฟ หรือแผนภาพ ซึ่งผลลัพธ์ของการทำ Visualization คือ การนำข้อมูลต่าง ๆ ที่ซ่อนอยู่ในตัวของข้อมูลเองออกมาให้ผู้ใช้งานสามารถสัมผัสได้ซึ่งกระบวนการต่าง ๆ ในการแสดงหรือนำข้อมูลต่าง ๆ ออกมา อาจจะถูกแอบซ่อนอยู่หรือผู้ใช้ไม่ทันสังเกต แต่อย่างไรก็ตาม Visualization เป็นสิ่งที่จำเป็นในการค้นหาข้อมูล หรือใช้ในการวิเคราะห์ข้อมูลโดยจุดมุ่งหมายของ Visualization ก็คือ การถ่ายทอดข้อมูลไปสู่ระบบการรับรู้โดยภาพของผู้ใช้ระบบ เพื่อช่วยในการลดช่องว่างระหว่างผู้ใช้และตัวข้อมูล และยังช่วยผู้ใช้งานสามารถเรียนรู้ ได้อย่างมีประสิทธิภาพ

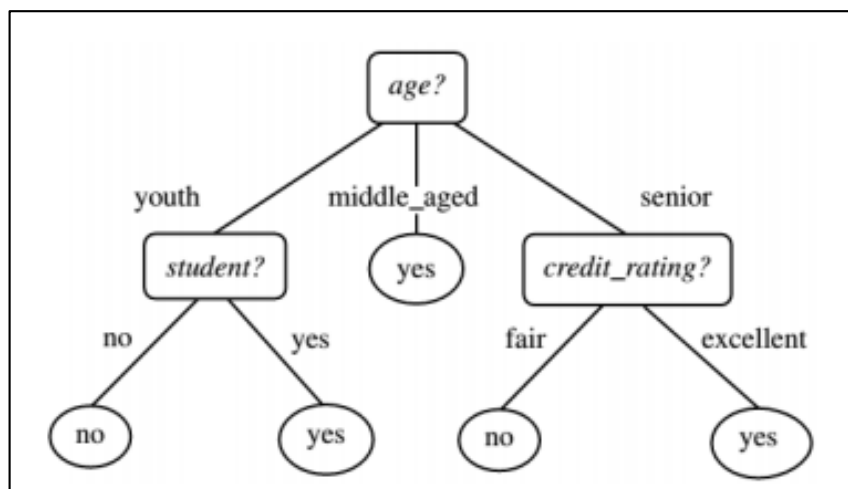
2.2.7.1 วิธีการของ Visualization

การใช้ Visualization เพื่อทำการค้นหาข้อมูลนั้นผู้ใช้งานจะทำขั้นตอนหลัก ๆ อยู่ 3 ขั้นตอน คือ Overview First, Zoom and Filter และ Detail on Demand โดยอันดับแรก ผู้ใช้ต้องการที่จะดูข้อมูลภาพรวมทั้งหมดซึ่งหลังจากดูภาพรวมทั้งหมดแล้วผู้ใช้งานก็จะทำการตัดสินใจเลือกรูปแบบหรือกลุ่มข้อมูลที่สนใจซึ่งก็จะมาถึงขั้นตอนในการวิเคราะห์ข้อมูลผู้ใช้งานก็จะทำการเจาะลึกถึงข้อมูลในรายละเอียด ซึ่ง Visualization Technology ก็จะอ้างอิงหรือพัฒนาจากขั้นตอนเหล่านี้ซึ่ง Visualization Technique มีประโยชน์มากในการแสดงภาพรวมหรือแสดงข้อมูลย่อยที่ผู้ใช้งานต้องการ โดยอาจจะใช้หลาย ๆ วิธีการรวมกัน เพื่อให้ได้ข้อมูลที่ผู้ใช้งานต้องการ ซึ่งช่วยลดช่องว่างของกิจกรรมที่ใช้ในการดึงข้อมูลต่าง ๆ ไปใช้ ซึ่งลักษณะของข้อมูลที่สามารถนำมาผ่านกระบวนการของ Visualization มีลักษณะต่าง ๆ มากมาย ดังนี้

ข้อมูล 1D ได้แก่ ข้อมูลที่เกี่ยวข้องกับช่วงเวลาใดเวลาหนึ่ง, ข้อมูล 2D ได้แก่ ข้อมูลที่เกี่ยวข้องกับแผนที่ภูมิศาสตร์, Multi-Dimensional Data ได้แก่ Relation Table, Text และ Hypertext ได้แก่ ข้อมูลหัวข้อข่าวต่าง ๆ และ Web Document, Hierarchies และ Graph ได้แก่ หมายเลขโทรศัพท์ และ Web Document, Algorithms และ Software ได้แก่ Debugging Operation ซึ่งแต่ละข้อมูลก็จะมีวิธีการที่ช่วยในการจัดการแสดงผลข้อมูลที่หลากหลาย

2.2.8 ทฤษฎีเกี่ยวกับการจำแนกข้อมูลด้วยต้นไม้ตัดสินใจ

การจำแนกข้อมูลด้วยต้นไม้ตัดสินใจจะเป็นกระบวนการสร้างต้นไม้ขึ้นเพื่อใช้ในการตัดสินใจจากข้อมูลที่มีหมวดหมู่ข้อมูลแนบอยู่ด้วย ต้นไม้ตัดสินใจจะประกอบไปด้วยโหนดต่าง ๆ ที่ไม่ใช่โหนดใบ (non-leaf node) ที่ซึ่งถูกใช้ในการแสดงถึงเงื่อนไขหรือแอทริบิวต์หนึ่ง ๆ ของข้อมูล โดยที่แต่ละกิ่งก้านของโหนดหนึ่ง ๆ จะหมายถึงค่าที่เป็นไปได้จากการทดสอบกับแอทริบิวต์นั้น ๆ และจะประกอบไปด้วยโหนดใบ (leaf node) ที่ซึ่งจะมีหมวดหมู่ข้อมูลจัดเก็บอยู่ โดยตัวอย่างต้นไม้ตัดสินใจจะถูกแสดงใน ภาพที่ 2.14 ด้านล่าง ที่จะแสดงการทำนายคุณลักษณะของลูกค้าที่จะทำการซื้อคอมพิวเตอร์จากร้านขายอุปกรณ์ไฟฟ้า โดยโหนดต่าง ๆ ที่ไม่ใช่โหนดใบจะถูกแทนด้วยสี่เหลี่ยม และโหนดใบจะถูกแทนด้วยวงรีตามลำดับ จากรูปจะเห็นว่าโหนดใบจะเป็นโหนดที่บ่งบอกถึงข้อมูลหมวดหมู่ของคำตอบที่เราต้องการ อาทิ เช่น “yes” หมายถึง ลูกค้าจะซื้อคอมพิวเตอร์ และ “no” หมายถึง ลูกค้าจะไม่ซื้อคอมพิวเตอร์โดยต้นไม้ที่ถูกสร้างขึ้น อาจเป็นต้นไม้ที่มีลักษณะเป็นไบนารีหรืออาจจะไม่เป็นไบนารีก็ได้ หลังจากทำการสร้างต้นไม้ตัดสินใจแล้ว เราจะสามารถใช้ต้นไม้ตัดสินใจในการจำแนกข้อมูลได้ โดยจะทำการจำแนกหมวดหมู่ของข้อมูลเรคคอร์ดหนึ่ง ๆ ที่ประกอบไปด้วยแอทริบิวต์ต่าง ๆ แต่เราจะไม่ทราบหมวดหมู่ข้อมูลในเรคคอร์ดนั้น ๆ ด้วยการเปรียบเทียบแอทริบิวต์ที่อยู่ในโหนดราก กับค่าของแอทริบิวต์ในเรคคอร์ดที่พิจารณา โดยจะทำการเปรียบเทียบจากโหนดรากไปจนถึงโหนดใบ เมื่อเราทราบถึงโหนดใบจะทำให้เราทราบถึงหมวดหมู่ข้อมูลของเรคคอร์ดที่ทำการพิจารณา



ภาพที่ 2.18 แสดงตัวอย่างต้นไม้ตัดสินใจของลูกค้าที่ทำการซื้อคอมพิวเตอร์

การจำแนกหมวดหมู่ข้อมูลด้วยต้นไม้ตัดสินใจได้รับความนิยมเป็นอย่างมาก และถูกประยุกต์ใช้ในหลาย ๆ งาน อาทิ เช่น การผลิตและการใช้ยา (Medicine) การผลิตสินค้าต่าง ๆ (Manufacturing and Production) การวิเคราะห์ทางการเงิน (Financial Analysis) ดาราศาสตร์ (Astronomy) อณูชีววิทยา (Molecular Biology) เป็นต้น สาเหตุที่ต้นไม้ตัดสินใจได้รับความนิยมอันเนื่องมาจากเหตุผลหลายประการด้วยกัน เช่น 1) ไม่ต้องการองค์ความรู้ใด ๆ หรือการกำหนดค่าพารามิเตอร์ใด ๆ เพื่อที่จะทำการสร้างต้นไม้ตัดสินใจ 2) สามารถจัดการกับข้อมูลที่มีหลาย ๆ มิติหรือข้อมูลที่มีหลาย ๆ แอทธิบิวได้ 3) ผลลัพธ์ที่ได้จะเป็นต้นไม้ตัดสินใจที่อยู่ในรูปแบบที่เข้าใจง่าย 4) ขั้นตอนการสร้างต้นไม้ตัดสินใจค่อนข้างง่ายและสามารถทำงานได้อย่างรวดเร็ว และ 5) มักจะให้ผลการจำแนกข้อมูลที่มีความถูกต้องค่อนข้างสูง ที่ซึ่งอาจขึ้นอยู่กับคุณลักษณะ ของข้อมูลที่ทำการพิจารณาด้วยเช่นกัน

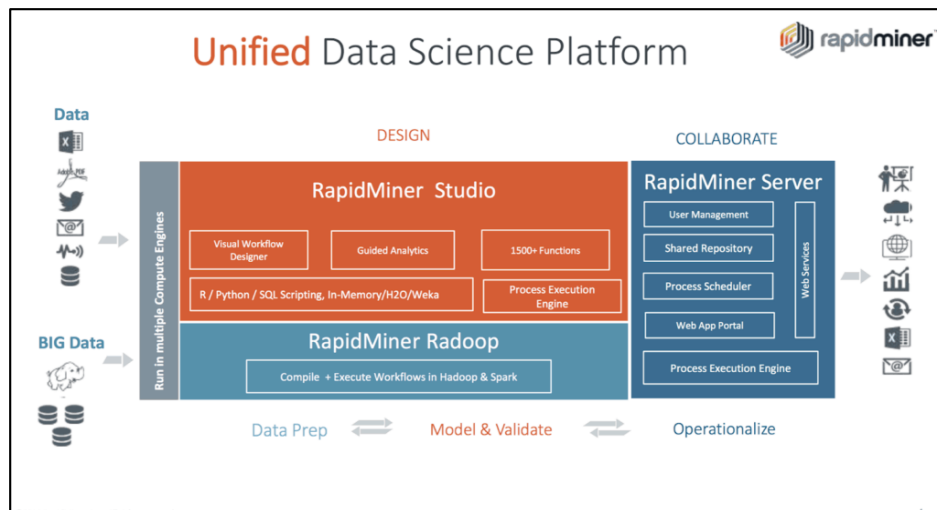
2.3 เครื่องมือในการออกแบบและวิเคราะห์ข้อมูล

2.3.1 โปรแกรม Rapid Miner

Rapid Miner เป็นวิทยาศาสตร์ข้อมูลแพลตฟอร์มซอฟต์แวร์ที่พัฒนาโดย บริษัท ที่มีชื่อเดียวกันที่ให้สภาพแวดล้อมแบบบูรณาการสำหรับการเตรียมข้อมูล, การเรียนรู้เครื่อง, การเรียนรู้ลึก, การทำเหมืองข้อความ, และการวิเคราะห์การทำนาย มันถูกใช้สำหรับธุรกิจและการใช้งานเชิงพาณิชย์ เช่นเดียวกับการวิจัยการศึกษาการฝึกอบรมการสร้างต้นแบบอย่างรวดเร็ว และการพัฒนาแอปพลิเคชัน และสนับสนุนทุกขั้นตอนของเครื่องกระบวนการการเรียนรู้รวม ถึงการเตรียมข้อมูลผลการสร้างภาพการตรวจสอบรูปแบบ และการเพิ่มประสิทธิภาพ

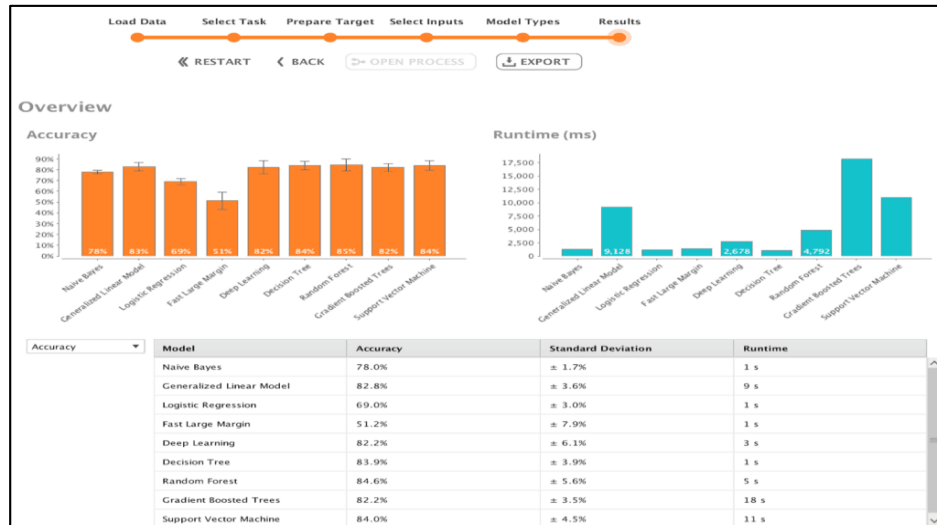
ตามการวิจัยของ Bloor Rapid Miner มอบคุณค่าการวิเคราะห์ขั้นสูง 99% ผ่านกรอบการทำงานของเทมเพลตที่ช่วยให้การจัดส่ง และลดข้อผิดพลาดโดยเกือบค่านกยุง]ไม่จำเป็นต้องเขียนโค้ด Rapid Miner มีขั้นตอนการชุดข้อมูลและการเรียนรู้ของเครื่องรวมถึง: การโหลดข้อมูลและการแปลง(ETL) การประมวลผลข้อมูลล่วงหน้าและการมองเห็นการวิเคราะห์เชิงพยากรณ์และการสร้างแบบจำลองทางสถิติการประเมินผลและการปรับใช้ Rapid Miner เขียนด้วยภาษาการเขียนโปรแกรม Java Rapid Miner มี GUI เพื่อออกแบบและดำเนินการเวิร์กโฟลว์การวิเคราะห์เวิร์กโฟลว์เหล่านั้นเรียกว่า "กระบวนการ" ใน Rapid Miner และประกอบด้วย "ตัวดำเนินการ" หลายรายการ ผู้ประกอบการแต่ละคนดำเนินงานเดี่ยวภายในกระบวนการและเอาต์พุตของผู้ประกอบการแต่ละรูปแบบการป้อนข้อมูลของหนึ่งต่อไป อีกวิธีหนึ่งเครื่องยนต์สามารถเรียกจากโปรแกรมอื่น ๆ หรือใช้เป็น API ฟังก์ชันส่วนบุคคลสามารถเรียกได้จากบรรทัดคำสั่ง Rapid Miner จัดเตรียมแผนการเรียนรู้แบบจำลองและอัลกอริทึมและสามารถขยายได้โดยใช้สคริปต์ R และ Python [6]

ฟังก์ชัน Rapid Miner สามารถขยายได้ด้วยปลั๊กอินเพิ่มเติมซึ่งมีให้ผ่านทาง Rapid Miner Marketplace Rapid Miner Marketplace เป็นแพลตฟอร์มสำหรับนักพัฒนาในการสร้างอัลกอริทึมการวิเคราะห์ข้อมูลและเผยแพร่สู่ชุมชน ซึ่ง Rapid Miner platform สามารถทำทั้งหมดนี้ได้ โดย Rapid Miner จะมี 3 โมดูลใหญ่ๆ ได้แก่ Rapid Miner Studio , Rapid Miner Server , Rapid Miner Radoop ตามรูปดังนี้



ภาพที่ 2.19 แสดงตัวอย่าง RapidMiner platform

โมดูลแรก : Rapid Miner Studio ใช้สำหรับการออกแบบการวิเคราะห์ข้อมูลผ่านทางหน้า GUI ซึ่งสามารถทำการจัดการข้อมูล และสร้างโมเดลแบบต่าง ๆ ได้โดยเวอร์ชันใหม่ ๆ ก็จะมีเรื่องของการทำงาน Auto Model ที่จะรันหลาย ๆ โมเดลในรูปแบบอัตโนมัติ ตัวอย่างรูปภาพดังนี้



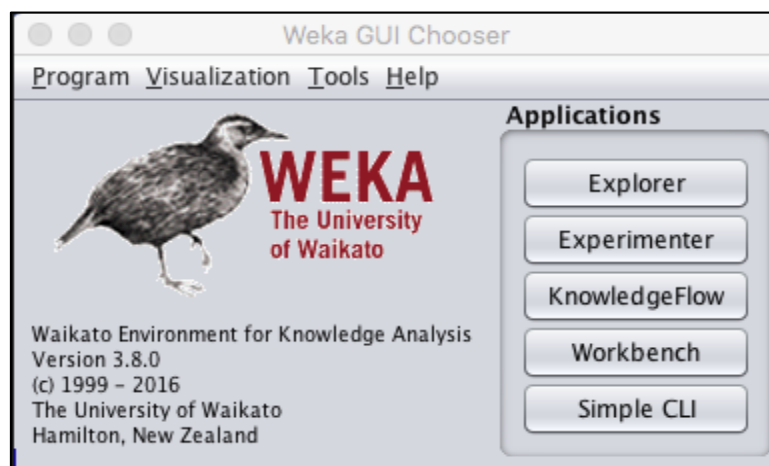
ภาพที่ 2.20 แสดงตัวอย่าง Auto Model ที่เจอรืสำหรับทำ Machine Learning

โมดูลสอง : Rapid Miner Server เป็นโมดูลที่รองรับการทำงาน of หลาย ๆ User ได้ ช่วยในเรื่องการตั้งเวลาให้ทำงาน (Scheduler) หรือสร้าง Web Service หรือ Web Application ได้ด้วย

โมดูลสาม : Rapid Miner Radoop เป็นโมดูลที่ใช้ในการจัดการข้อมูลที่มีขนาดใหญ่ ๆ แบบ Big Data โดยการทำงานจะไปประมวลผลบน Hadoop / Spark และไม่จำเป็นต้องเขียน Code

2.3.2 โปรแกรม Weka

โปรแกรม Weka (Waikato Environment for Knowledge Analysis) เริ่มพัฒนามาตั้งแต่ปี 1997 โดยมหาวิทยาลัย Waikato ประเทศนิวซีแลนด์ เป็นซอฟต์แวร์สำเร็จรูป อยู่ภายใต้การควบคุมของ GPL License โปรแกรม Weka ได้ถูกพัฒนามาจากภาษาจาวาทั้งหมด ซึ่งเขียนมาโดยเน้นกับงานทางด้านการเรียนรู้ด้วยเครื่อง (Machine Learning) และ การทำเหมืองข้อมูล (Data Mining) โปรแกรมจะประกอบไปด้วยโมดูลย่อย ๆ สำหรับใช้ในการจัดการข้อมูล และเป็นโปรแกรมที่สามารถใช้ Graphic User Interface (GUI) และใช้คำสั่งในการให้ซอฟต์แวร์ประมวลผล และสามารถรัน (Run) ได้หลายระบบปฏิบัติการ และสามารถพัฒนาต่อยอดโปรแกรมได้ เป็นเครื่องมือที่ใช้ทำงานในด้านการทำดาตาไมนนิ่งที่รวบรวมแนวคิดอัลกอริทึมมากมาย ซึ่งอัลกอริทึมสามารถเลือกใช้งานโดยตรงได้จาก 2 ทางคอกจากชุดเครื่องมือที่มีอัลกอริทึมมาให้ หรือเลือกใช้จากอัลกอริทึมที่ได้เขียนเป็นโปรแกรมลงไปเป็นชุดเครื่องมือเพิ่มเติม และชุดเครื่องมือมีฟังก์ชันสำหรับการทำงานร่วมกับข้อมูล ได้แก่ Pre-Processing, Classification, Regression, Clustering, Association rules, Selection และ Visualization ส่วนของการ Import ข้อมูล โดยข้อมูลนั้นจะต้องอยู่ในรูปแบบ ASCII อาจเป็น arff , csv , c45



ภาพที่ 2.21 แสดงตัวอย่างหน้าโปรแกรม Weka

1) กฎความสัมพันธ์ (Association Rule) แสดงความสัมพันธ์ของเหตุการณ์หรือวัตถุ ที่เกิดขึ้นพร้อมกัน ตัวอย่างของการประยุกต์ใช้กฎเชื่อมโยง เช่น การวิเคราะห์ข้อมูลการขายสินค้า โดยเก็บข้อมูลจากระบบ ณ จุดขาย (POS) หรือร้านค้าออนไลน์ แล้วพิจารณาสินค้าที่ผู้ซื้อมักจะซื้อพร้อมกัน เช่น ถ้าพบว่าคนที่ซื้อเทปวีดีโอ มักจะซื้อเทปกาวด้วย ร้านค้าก็อาจจะจัดร้านให้สินค้าสองอย่างอยู่ใกล้กัน เพื่อเพิ่มยอดขาย หรืออาจจะพบว่าหลังจากคนซื้อหนังสือ ก แล้ว มักจะซื้อหนังสือ ข ด้วย ก็สามารถนำความรู้นี้ไปแนะนำผู้ที่กำลังจะซื้อหนังสือ ก ได้

2) การจำแนกประเภทข้อมูล (Data Classification) หากกฎเพื่อระบุประเภทของวัตถุ จากคุณสมบัติของวัตถุ เช่น หาความสัมพันธ์ระหว่างผลการตรวจร่างกายต่าง ๆ กับการเกิดโรค โดยใช้ข้อมูลผู้ป่วยและการวินิจฉัยของแพทย์ที่เก็บไว้ เพื่อนำมาช่วยวินิจฉัยโรคของผู้ป่วย หรือการวิจัยทางการแพทย์ ในทางธุรกิจจะใช้เพื่อดูคุณสมบัติของผู้ที่จะก่อหนี้ดีหรือหนี้เสีย เพื่อประกอบการพิจารณาการอนุมัติเงินกู้

3) การแบ่งกลุ่มข้อมูล (Data Clustering) แบ่งข้อมูลที่มีลักษณะคล้ายกันออกเป็นกลุ่ม แบ่งกลุ่มผู้ป่วยที่เป็นโรคเดียวกันตามลักษณะอาการ เพื่อนำไปใช้ประโยชน์ในการวิเคราะห์หาสาเหตุของโรค โดยพิจารณาจากผู้ป่วยที่มีอาการคล้ายคลึงกัน

4) จินตทัศน์ (Visualization) สร้างภาพคอมพิวเตอร์กราฟิกที่สามารถนำเสนอข้อมูล มากมายอย่างครบถ้วนแทนการใช้ข้อความนำเสนอข้อมูลที่มากมาย เราอาจพบข้อมูลที่ซ้อนเร้น เมื่อดูข้อมูลชุดนั้นด้วยจินตทัศน์

การทำเหมืองข้อมูล จำเป็นต้องอาศัยบุคลากรจากหลายฝ่าย และต้องอาศัยความรู้จำนวนมาก ถึงจะได้รับการประโยชน์อย่างแท้จริง เพราะสิ่งที่ได้จากขั้นตอนวิธีเป็นเพียงตัวเลข และข้อมูล ที่อาจจะนำไปใช้ประโยชน์ได้หรือใช้ประโยชน์อะไรไม่ได้เลยก็เป็นได้ ผู้ที่ศึกษาการทำเหมืองข้อมูลจึงควรมีความรู้รอบด้าน และต้องติดต่อกับทุก ๆ ฝ่าย เพื่อให้เข้าใจถึงขอบเขตของปัญหาโดยแท้จริงก่อน เพื่อให้การทำเหมืองข้อมูลเกิดประโยชน์อย่างแท้จริง

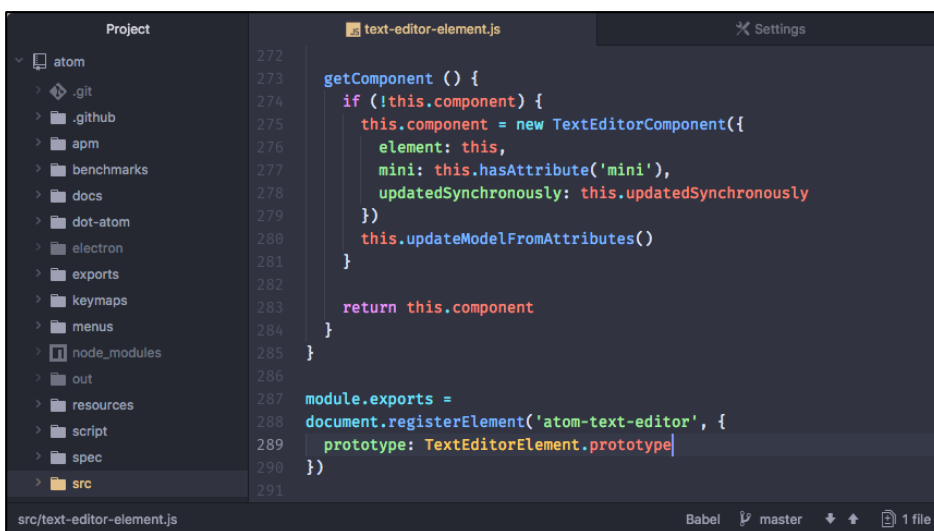
2.3.3 โปรแกรม Atom

Atom (โปรแกรม Text Editor แก้ไขข้อความ เขียนโค้ด ฟรี จาก GitHub) : สำหรับโปรแกรมนี้มีชื่อว่า โปรแกรม Atom เป็นโปรแกรมประเภทแก้ไขข้อความ (Text Editor) สามารถใช้ได้ทั้ง เขียนโค้ด รวมไปถึง แก้ไขส่วนต่าง ๆ ของโค้ดโปรแกรม หรือ แอปพลิเคชันต่าง ๆ ได้อย่างง่ายดาย ด้วยรูปแบบโปรแกรมที่เรียบง่าย พร้อมเครื่องมือ Tool สำหรับใช้งานที่หลากหลาย เหมาะสำหรับ

ผู้ที่เริ่มต้นเขียนโปรแกรมและมีอาชีพที่กำลังมองหาโปรแกรมที่ใช้งานง่าย ที่จะมาช่วยในการเขียนโค้ดง่ายและรวดเร็วมากขึ้น

โปรแกรม Atom ตัวนี้พัฒนามาจาก เว็บไซต์ชื่อดัง GitHub ที่ช่วยกันร่วมพัฒนาและเปิดให้ใช้งานได้ฟรี ๆ คุณสามารถใช้งานโปรแกรมแก้ไขเขียนโค้ด Atom เปิดใช้งาน Project ของคุณ ช่วยให้คุณดูงานเขียนโค้ดของคุณได้ง่ายมากยิ่งขึ้น มีเครื่องมือให้ใช้งาน ค้นหา คัดลอก และ แสดงตัวอย่าง หลังจากเขียนเสร็จแล้วก็สามารถ Preview ดูตัวอย่างของโค้ดที่เขียนได้เลขทันที พร้อมแสดงผลและแก้ไขส่วนข้อผิดพลาดต่าง ๆ ได้ง่ายมากยิ่งขึ้น

นอกจากนี้ โปรแกรม Atom ตัวนี้ก็ยังมีฟังก์ชัน Smart auto completion ช่วยแสดงและเติมส่วนต่าง ๆ ของโค้ดของคุณ ทำให้คุณเขียนโค้ดได้รวดเร็วมากยิ่งขึ้น พร้อมแบ่งบรรทัดการเขียน เติมแต่งสีให้เห็นชัดเจน หากผู้ใช้งานคนไหนชื่นชอบหน้าต่างโปรแกรมแบบอื่น ๆ ก็สามารถหา Theme มาติดตั้งเองจากเว็บไซต์ GitHub ได้อีกด้วย โปรแกรมนี้ รองรับการใช้งานข้ามแพลตฟอร์ม (Cross Platform) ไม่ว่าจะเป็น บนระบบปฏิบัติการ Windows Linux และ MacOS



```

Project
├── atom
├── .git
├── .github
├── apm
├── benchmarks
├── docs
├── dot-atom
├── electron
├── exports
├── keymaps
├── menus
├── node_modules
├── out
├── resources
├── script
├── spec
└── src

src/text-editor-element.js
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291

text-editor-element.js
getComponent () {
  if (!this.component) {
    this.component = new TextEditorComponent({
      element: this,
      mini: this.hasAttribute('mini'),
      updatedSynchronously: this.updatedSynchronously
    })
    this.updateModelFromAttributes()
  }
  return this.component
}

module.exports =
document.registerElement('atom-text-editor', {
  prototype: TextEditorElement.prototype
})

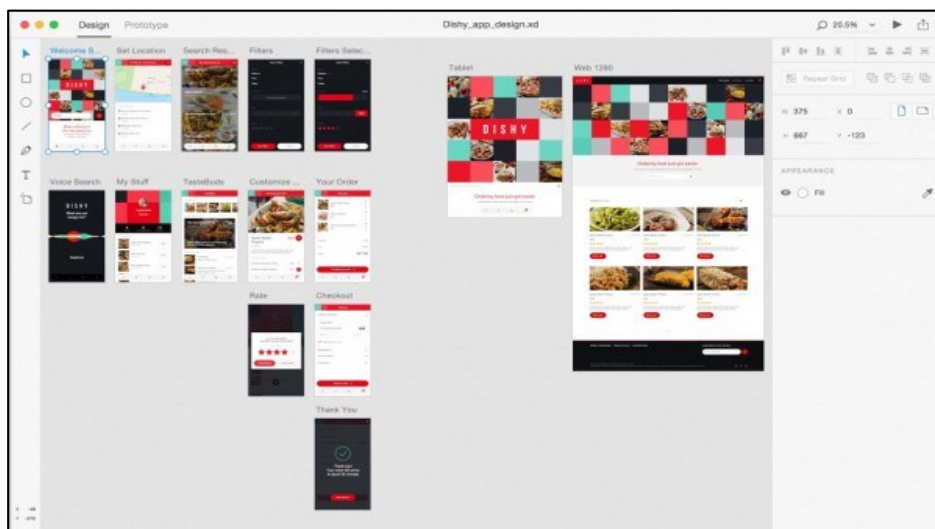
```

ภาพที่ 2.22 แสดงตัวอย่างหน้าโปรแกรม Atom

2.3.4 โปรแกรม Adobe XD

Adobe XD คือ แอปพลิเคชันบนคอมพิวเตอร์ ที่ช่วยให้นักออกแบบเริ่มทำงานตั้งแต่การสร้างไวร์เฟรม ไปจนถึงการสร้างตัวตนแบบเพื่อทดลองใช้เพื่อเก็บข้อมูลกลับมาปรับปรุงการใช้งานให้ดีขึ้น เพื่อที่ผู้จะใช้จะได้กลับมาใช้งานอีกมากขึ้น เวอร์ชันฟรีวิวแรกที่ทำให้ทดลองใช้ครั้งนี้ โหมดออกแบบ (Design) สามารถทำงานออกแบบได้อย่างง่ายดายตาย โหมดต้นแบบ (Prototype)

ก็สามารถกำหนด Hotspots และ Transitions ได้ตามต้องการ โหมดพรีวิว (Desktop Preview) สำหรับทดลองใช้งานเพื่อดูความเปลี่ยนแปลงจริงได้ นอกจากนี้ ยังสามารถแชร์พรีวิวให้ผู้ตัดสินใจและคนอื่นได้ทดลองใช้งานได้ ในบราวเซอร์บนคอมพิวเตอร์หรือบนอุปกรณ์พกพาอีกด้วย นอกจากนี้ นักออกแบบยังสามารถนำงานที่สร้างขึ้นในแอปพลิเคชันสำหรับการออกแบบอย่าง Illustrator CC และ Photoshop CC มาใช้งานร่วมกันได้ และเมื่อการออกแบบและทดลองใช้มาถึงขั้นสุดท้ายแล้ว ทุกอย่างก็สามารถเอ็กซ์พอร์ตไปให้นักพัฒนาทำงานต่อไปได้อย่างง่ายดาย



ภาพที่ 2.23 แสดงตัวอย่างหน้าโปรแกรม Adobe XD

2.4 วรรณกรรมที่เกี่ยวข้อง

อนุชิต พงศ์พรหม, พลกฤต แสงอาวูช และพูนฉัตร วิชัยดิษฐ์ (2562) การศึกษาการบริหารจัดการทุนการศึกษาของนักศึกษาระดับมหาวิทยาลัยของรัฐจังหวัดสุราษฎร์ธานีครั้งนี้มีจุดประสงค์เพื่อเปรียบเทียบ และศึกษาแนวทางในการบริหารจัดการทุนการศึกษาดังกล่าว โดยจำแนกตามปัจจัยส่วนบุคคล เนื่องจากยังมีนักศึกษาจำนวนมากที่ต้องการ และมีคุณสมบัติที่ตรงกับวัตถุประสงค์การให้ทุนการศึกษา แต่ยังไม่ได้รับการดูแลช่วยเหลือ ซึ่งกลุ่มตัวอย่างในงานวิจัยครั้งนี้เป็นกลุ่มตัวอย่างจากมหาวิทยาลัยราชภัฏสุราษฎร์ธานี และมหาวิทยาลัยสงขลานครินทร์วิทยาเขต สุราษฎร์ธานี เนื่องจากทั้ง 2 มหาวิทยาลัยเป็นองค์กรที่รับนโยบายการลดความเหลื่อมล้ำทางการศึกษา ซึ่งเป้าหมายหลักในการขับเคลื่อนมหาวิทยาลัย เพื่อผลิตบัณฑิตที่มีคุณลักษณะบัณฑิตที่พึงประสงค์สามารถเรียนรู้คิดเป็นทำเป็นและมีศักยภาพ

จากการศึกษาบทความดังกล่าว ทางผู้จัดทำเห็นว่า ทางผู้จัดทำจะนำแนวคิดและแนวทางในการบริหารจัดการทุนการศึกษา มาใช้เป็นแนวทางในการวิเคราะห์ และสร้างจำลองมาตรการช่วยเหลือด้านการศึกษา เพื่อเป็นส่วนช่วยในการตัดสินใจในการมอบทุนการศึกษาอย่างเป็นระบบ

สุมาลี ลิกเสน และมนตชัย เทียนทอง (2559)งานวิจัยนี้มีวัตถุประสงค์เพื่อวิเคราะห์ความถนัดของผู้เรียน ตามแนวทางพหุปัญญา ของ Howard Gardner ที่จำแนกความถนัดของผู้เรียนออกเป็น 3 กลุ่ม กลุ่มตัวอย่างที่ใช้ในการวิจัยครั้งนี้ คือ นักศึกษามหาวิทยาลัยราชภัฏนครปฐม ในภาคเรียนที่ 2/2556 จำนวน 431คน ทำการวิเคราะห์ข้อมูลด้วยโปรแกรม WEKA โดยใช้กฎการจำแนกเทคนิคต้นไม้ตัดสินใจ ภายใต้อัลกอริทึม j48 สร้างและทดสอบตัวแบบด้วยวิธี 10 fold Cross validation ผลการวิจัยพบว่า ตัวแปรต้นที่ส่งผลต่อความถนัดของผู้เรียน ตามแนวทางพหุปัญญา ประกอบด้วย สาขาวิชาเกรดเฉลี่ยวิชาที่ทำคะแนนได้สูงอยู่เสมอ ลักษณะการเรียนที่ชอบวิธีการเรียนที่มีความถนัดกิจกรรมที่ชอบและงานอดิเรกที่ชอบ ได้กฎการจำแนกความถนัดของผู้เรียนทั้งสิ้นจำนวน 54 กฎดังนี้ ความถนัดของผู้เรียนกลุ่มการวิเคราะห์ (Analytic) จำนวน 19 กฎ ความถนัดของผู้เรียนกลุ่มพินิจพิจารณา (Introspective) จำนวน 14 กฎ และความถนัดของผู้เรียนกลุ่มปฏิสัมพันธ์ (Interactive) จำนวน 21 กฎ ซึ่งมีค่าความถูกต้องคิดเป็นร้อยละ 85.90 ค่าความแม่นยำคิดเป็นร้อยละ 86.03 ค่าความระลึกคิดเป็นร้อยละ 85.56 และค่าถ่วงดุลคิดเป็นร้อยละ 85.89 โดยผลที่ได้จากงานวิจัยนี้ สามารถนำไปใช้ในการพัฒนาระบบ การวิเคราะห์ความถนัดของผู้เรียนตามแนวทางพหุปัญญาของ Howard Gardner โดยใช้กฎการจำแนกเทคนิคต้นไม้ตัดสินใจที่มีความถูกต้องและแม่นยำในการทำนายความถนัดของผู้เรียนได้สำหรับแบ่งกลุ่มเพื่อจัดทำระบบโครงการออนไลน์ต่อไป

จากการศึกษาบทความดังกล่าว ทางผู้จัดทำเห็นว่า ทางผู้จัดทำจะนำกฎการจำแนกด้วยเทคนิคต้นไม้ตัดสินใจมาใช้ในการสร้างแบบจำลองเพื่อเป็นมาตรการช่วยเหลือด้านการศึกษา สำหรับนักศึกษา โดยทำการวิเคราะห์ข้อมูลด้วยโปรแกรม Weka

จริยา ศรีวิจารณ์, อรัญญา ต้อยคำภีร์ และณัฐสุดา เต็มพันธ์ (2556) การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาประสบการณ์ด้านจิตใจของนักศึกษาที่ยากจนจากชนบทและได้รับทุนการศึกษา ผู้ให้ข้อมูลเป็นนักศึกษาปริญญาตรี จำนวน 10 ราย คัดเลือกแบบเจาะจงตามเกณฑ์

ที่กำหนด เก็บข้อมูลโดยการสัมภาษณ์รายบุคคล วิเคราะห์ข้อมูลโดยวิธีการวิจัยเชิงคุณภาพแบบ
 ทีมผู้วิจัยเห็นชอบร่วมกันผลการศึกษาพบ 5 ประเด็นหลัก คือ (1) การเผชิญความยากลำบากใน
 ชีวิต ได้แก่ ความเป็นอยู่ที่ขาดแคลน การไม่ยอมจำนนต่อความยากจน การมีความกดดันและรู้สึก
 ต้อยค่าจากความยากจน (2) การมีความรู้สึกสุข ระคนทุกข์เมื่อได้ทุนการศึกษา ได้แก่ ความรู้สึก
 ภาคภูมิใจร่วมกับความรู้สึกซาบซึ้งใจที่ได้รับทุนและความรู้สึกไม่มั่นใจบางขณะ (3) การปรับตัวใน
 สังคมใหม่ ได้แก่ การใช้ชีวิตในมหาวิทยาลัย การดำรงอยู่ภายใต้สิ่งแวดล้อมวัฒนธรรม และการใช้
 ชีวิตต่างถิ่นเพียงลำพัง (4) การมีทัศนคติดีและกตัญญูสำคัญในการปรับตัว ได้แก่ การมีจิตใจที่
 แข็งแกร่ง การมีสัมพันธภาพและการมีแหล่งเกื้อหนุนที่ดี และ (5) การมีจุดมุ่งหมายในชีวิต ได้แก่
 การทำสิ่งต่าง ๆ เพื่อครอบครัว เพื่อความมั่นคงของตัวเอง และเพื่อช่วยเหลือผู้อื่น จะเห็นว่าปัจจัย
 หนึ่งในที่สำคัญ ในการที่ทำให้นักศึกษาต่อสู้กับความลำบากในชีวิตจากความยากจน คือ ทัศนคติด้าน
 บวกและการมีจุดมุ่งหมาย ในชีวิตของพวกเขา

จากการศึกษาบทความดังกล่าว ทางผู้จัดทำเห็นว่า ทางผู้จัดทำจะนำแนวคิดและแนวทาง
 ในการบริหารจัดการทุนการศึกษา มาใช้เป็นแนวทางในการวิเคราะห์ และสร้างจำลองมาตรการ
 การช่วยเหลือด้านการศึกษา เพื่อเป็นส่วนช่วยในการตัดสินใจในการมอบทุนการศึกษาอย่างเป็น
 ระบบ

อภิยศ เจริญวิวัฒน์ (2562) สำนักหอสมุด มหาวิทยาลัยเกษตรศาสตร์ มีแนวคิดในการ
 ประยุกต์ใช้ข้อมูลทางสถิติเพื่อเพิ่มประสิทธิภาพในการให้บริการแก่ผู้ใช้ โดยทำการศึกษากับตัวเลข
 ของการให้บริการในด้านต่าง ๆ ที่กระจุกกระจายอยู่เป็นจำนวนมาก มีปริมาณข้อมูลสูง และยังไม่
 เคยผ่านการวิเคราะห์ทางด้านสถิติมาก่อน โดยการวิเคราะห์ข้อมูลนี้ จะวิเคราะห์ข้อมูลที่มี
 หลากหลาย ทั้งจากการเข้าใช้งานอาคารของสำนักหอสมุด และการใช้บริการพื้นฐานและบริการ
 พิเศษต่าง ๆ โดยใช้วิธีวิเคราะห์ข้อมูลแสดงความสัมพันธ์ และนำมาแสดงผลในรูปแบบของ Data
 visualization โดยใช้ภาพ ตัวเลข แผนภูมิ กราฟ และอื่น ๆ ซึ่งช่วยเพิ่มความพึงพอใจแก่ผู้บริหารให้
 สามารถรับรู้ข้อมูลที่มากขึ้น และสามารถนำข้อมูลที่ได้ไปใช้ในการสนับสนุนการตัดสินใจ

จากการศึกษาบทความดังกล่าว ทางผู้จัดทำเห็นว่า ทางผู้จัดทำจะนำวิธีวิเคราะห์ข้อมูลแสดง
 ความสัมพันธ์ โดยการทำ Data visualization เพื่อเผยแพร่ผลวิเคราะห์ ขั้นตอนการนำผลลัพธ์ที่ได้
 ไปใช้งานเป็นการทั่วไป อาจจัดทำเป็นรูปแบบของรายงาน (Report) หรือแผนภาพ (Dashboard) ที่
 พร้อมให้ฝ่ายต่าง ๆ นำไปใช้ประโยชน์ในการวางแผน กำหนดกลยุทธ์ และดำเนินการต่าง ๆ

การวิจัยครั้งนี้มีวัตถุประสงค์ 1) เพื่อเปรียบเทียบประสิทธิภาพการจำแนกข้อมูลด้วยเทคนิคต้นไม้ตัดสินใจ รุ่นที่ 3 ที่ใช้เทคนิคการแบ่งข้อมูลที่แตกต่างกัน และ 2) เพื่อทดสอบประสิทธิภาพของเทคนิคการแบ่งข้อมูลชนิดแบบไม่มีผู้สอน ชุดข้อมูลที่ใช้ในการทดลอง คือ ชุดข้อมูล KDD Cup 1999 ประสิทธิภาพ ของการจำแนกข้อมูลวัดจากอัตราการจำแนกข้อมูลถูกต้อง และอัตราการจำแนกข้อมูลผิดพลาดโดยใช้โปรแกรมเวก้า และแมทแลบในการประมวลผล จากการเฉลี่ยผลการทดลองจำนวน 10 รอบ โดยใช้เทคนิคในการแบ่งข้อมูลของชุดข้อมูลฝึกสอนให้อยู่ในรูปแบบไม้ต่อเนื่อง และจำแนกข้อมูลด้วยต้นไม้ตัดสินใจรุ่นที่ 3 พบว่าการแบ่งข้อมูลด้วยขนาดความถี่เท่ากันจำนวน 10 ชั้น มีอัตราการจำแนกข้อมูลถูกต้องสูงที่สุด เท่ากับ 99.79% รองลงมาคือ การแบ่งข้อมูลด้วยการจัดกลุ่มเคมีนส์ จำนวน 40 กลุ่ม เท่ากับ 99.75% และน้อยที่สุด คือ การแบ่งข้อมูลด้วยขนาดความถี่เท่ากัน จำนวน 20 ชั้น เท่ากับ 99.57% เมื่อเปรียบเทียบประสิทธิภาพการจำแนกข้อมูลด้วยต้นไม้ตัดสินใจ รุ่นที่ 3 โดยนำกฎที่สร้างขึ้นจากชุดข้อมูลฝึกสอนไปใช้กับชุดข้อมูลทดสอบพบว่าประสิทธิภาพของการใช้เทคนิคแบ่งข้อมูลด้วยขนาดความถี่เท่ากันและจำแนกข้อมูลด้วยต้นไม้ตัดสินใจรุ่นที่ 3 มีประสิทธิภาพสูงที่สุด โดยมีอัตราการจำแนกข้อมูลถูกต้องเท่ากับ 92.30% มีอัตราการจำแนกข้อมูลผิดพลาด เท่ากับ 4.89% และสามารถจำแนกข้อมูลที่ไม่สามารถจัดเข้ากลุ่มได้ เท่ากับ 2.81% ผลการทดลองยังแสดงให้เห็นว่าเทคนิคการแบ่งข้อมูลชนิดแบบไม่มีผู้สอน แต่ละชนิดส่งผลต่อประสิทธิภาพการจำแนกข้อมูลของต้นไม้ตัดสินใจ รุ่นที่ 3 และพบว่าการใช้เทคนิคการแบ่งข้อมูลร่วมกับเทคนิคต้นไม้ตัดสินใจ รุ่นที่ 3 สามารถช่วยกรองข้อมูลที่ไม่อยู่ในกฎที่สร้างขึ้นได้ ซึ่งเป็นสิ่งที่แตกต่างกับเทคนิคต้นไม้ตัดสินใจ รุ่นที่ C4.5 ฉะนั้นวิธีที่นำเสนอหากเลือกนำไปใช้ได้อย่างเหมาะสมจะเกิดประโยชน์อย่างมาก

จากการศึกษาบทความดังกล่าว ทางผู้จัดทำเห็นว่าจะนำวิจัยครั้งนี้มาทดสอบ เปรียบเทียบประสิทธิภาพของข้อมูลนักศึกษาด้วยเทคนิคต้นไม้ตัดสินใจ โดยการใช้เทคนิคการแบ่งข้อมูลเป็น 2 ส่วนส่วนแรก คือ การนำข้อมูลมาใช้จริง (Training Data) 70% และส่วนที่สอง คือ การนำข้อมูลมาทดสอบ (Testing Data) 30% เพื่อหาอัลกอริทึมที่เหมาะสมและให้ค่าความแม่นยำมากที่สุด ซึ่งวัดค่าความยำโดยใช้ Confusion Matrix ในโปรแกรม Weka

ดังนั้น วรรณกรรมที่เกี่ยวข้องทั้งหลายที่ได้กล่าวมาในข้างต้นทั้งหมดนั้น ผู้จัดทำโครงการได้เล็งเห็นความสำคัญของการนำเทคโนโลยีมาประยุกต์ใช้ในการวิเคราะห์ข้อมูลนักศึกษาเบื้องต้น เพื่อใช้สำหรับเป็นมาตรการช่วยเหลือในด้านของการศึกษา และเผยแพร่ข้อมูลบนเว็บไซต์บน Web

Browser เนื่องจากเทคโนโลยีปัจจุบันเข้ามาช่วยในการจัดการข้อมูล หรือดำเนินงานต่าง ๆ ให้ความสะดวกสบายเป็นอย่างมาก อีกทั้งผู้ใช้งานยังเข้าถึงข้อมูลได้ง่าย และมีประสิทธิภาพ

2.5 บทสรุป

จากแนวคิด ทฤษฎี เครื่องมือและวรรณกรรมที่เกี่ยวข้องที่ได้กล่าวมาในข้างต้นทั้งหมดนั้น ผู้วิเคราะห์ข้อมูลได้เลือกใช้ขั้นตอนกระบวนการวิเคราะห์ข้อมูลด้วย CRISP-DM จากเทคนิคการทำเหมืองข้อมูลแบบ Classification ด้วยการสร้างโมเดล Decision Tree เพื่อจัดกลุ่มข้อมูลที่มีความสัมพันธ์กัน โดยใช้โปรแกรมที่ใช้ทำเหมืองข้อมูล คือ โปรแกรม SQL Server 2014 Management Studio & SQL Server Data Tools for Visual Studio 2013, โปรแกรม Weka 3.8.4 และโปรแกรม Rapid Miner Studio 9.5.1 ในการสร้างโมเดล Decision Tree เพื่อทำการเปรียบเทียบและเลือกผลลัพธ์ที่ดีที่สุดมานำเสนอ จากนั้นนำข้อมูลสารสนเทศมาทำการแสดงผลแบบ Visualization ในรูปแบบของภาพโดยใช้โปรแกรม Tableau Public เผยแพร่บน Web Browser ที่เป็นที่ยอมรับในยุคอินเทอร์เน็ตคือการเผยแพร่ทางสื่อออนไลน์ โดยใช้ภาษา HTML และ CSS ในการเขียนเว็บไซต์ขึ้นมา

บทที่ 3

วิธีการดำเนินงานโครงการ

โครงการเรื่อง การพัฒนาแบบจำลองมาตรฐานการช่วยเหลือด้านการศึกษ สำหรับนักศึกษา ด้วยเทคนิคต้นไม้ตัดสินใจ ในบทนี้จะเป็นการวิเคราะห์ข้อมูลด้วยเทคนิคทางดาต้าไมนิ่ง ซึ่งมีกระบวนการวิเคราะห์ที่สำคัญหลายขั้นตอน เมื่อเสร็จสิ้นจากกระบวนการวิเคราะห์ข้อมูลแล้ว จะเป็นการออกแบบเว็บไซต์ และออกแบบรูปแบบการแสดงผลและบทสรุปจากวิธีการดำเนินงาน

3.1 การวิเคราะห์ข้อมูลด้วย CRISP-DM

3.2 การออกแบบเว็บไซต์

3.3 บทสรุป

3.1 การวิเคราะห์ข้อมูลด้วย CRISP-DM

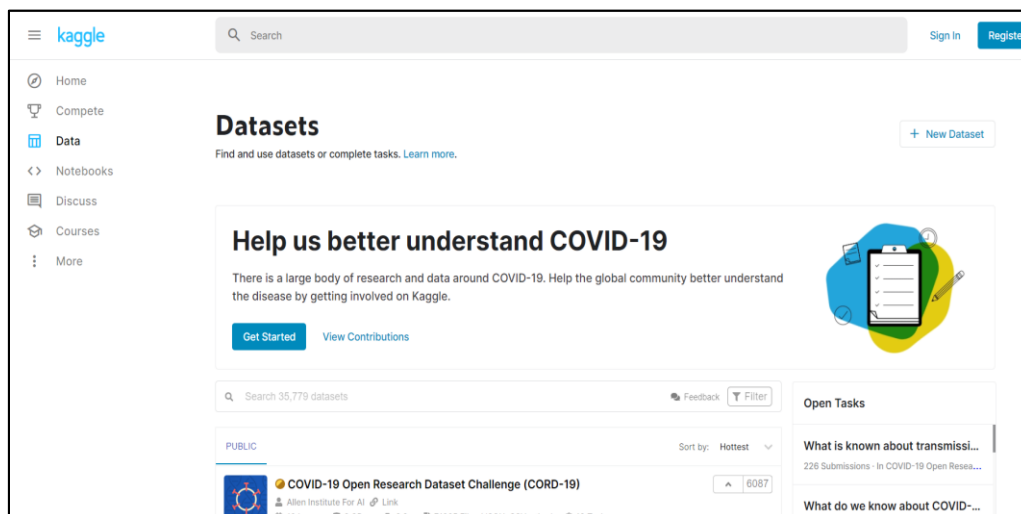
กระบวนการวิเคราะห์ข้อมูลด้วย CRISP-DM หรือ Cross Industry Standard Process for Data Mining พัฒนาขึ้นในปี ค.ศ. 1996 โดยความร่วมมือของ 3 บริษัทคือ Daimler Chrysler, SPSS และ NCR ที่มีการพัฒนาเป็น Work flow มาตรฐานสำหรับการทำเหมืองข้อมูล ประกอบด้วย 6 ขั้นตอนหลัก ดังนี้

3.1.1 รู้จักและเข้าใจในธุรกิจ (Business Understanding) เป็นขั้นตอนแรกของกระบวนการที่มุ่งเน้นไปที่การทำความเข้าใจกระบวนการทางธุรกิจโดยรวม

รู้จักและเข้าใจในธุรกิจ (Business Understanding) เป็นขั้นตอนแรกของกระบวนการที่มุ่งเน้นไปที่การทำความเข้าใจกระบวนการทางธุรกิจโดยรวม ผู้วิเคราะห์ข้อมูลทำความเข้าใจกับปัญหาให้อยู่ในรูปของการวิเคราะห์ข้อมูลทาง Data Mining โดยการวิเคราะห์ข้อมูลในประเด็นนี้ ซึ่งมีข้อมูลมาก ทำให้ไม่สามารถทำความเข้าใจกับข้อมูลที่มีความซับซ้อนได้อย่างรวดเร็ว เช่น ต้องการทราบว่านักศึกษาลักษณะคุณสมบัติใดที่บ่งบอกถึงความต้องการช่วยเหลือในด้านการศึกษา

3.1.2 จัดเก็บและรวบรวมข้อมูลให้ครบ (Data Understanding) ขั้นตอนการจัดเก็บและรวบรวมข้อมูล ตลอดจนการพิจารณาตรวจสอบความถูกต้องของข้อมูลที่ได้รับ โดยเลือกว่าจะใช้ข้อมูลทั้งหมดหรือบางส่วนในการวิเคราะห์ให้สอดคล้องกับวัตถุประสงค์ที่กำหนดไว้

ผู้วิเคราะห์ข้อมูลทำการรวบรวมข้อมูล เพื่อตรวจสอบรายละเอียด ปริมาณ และความน่าเชื่อถือของข้อมูลนักศึกษาเบื้องต้น หรือสถานะครอบครัว ที่ได้จากเว็บไซต์ Kaggle.com ซึ่งเป็นเว็บไซต์ที่เก็บรวบรวมชุดข้อมูลต่าง ๆ เป็นแหล่งรวม Datasets หรือชุดข้อมูล สำหรับฝึกสอน Machine Learning ที่ใหญ่ที่สุดในโลกแห่งหนึ่ง มีข้อมูลทุกประเภทไม่ว่าจะเป็น Datasets ในหมวดหมู่ Finance, Business, Physics, Biology, Sports, News ซึ่งเป็นข้อมูลที่เปิดเผยได้ เพื่อให้ผู้ใช้บริการสามารถนำชุดข้อมูลไปศึกษาหรือวิเคราะห์ให้เกิดประโยชน์ต่อไป



ภาพที่ 3.1 เว็บไซต์ Kaggle.com

ซึ่งข้อมูลนักศึกษาเบื้องต้น มีจำนวนข้อมูล 1,044 รายการ ประกอบด้วย 15 แอตทริบิวต์ ประกอบด้วย เพศ อายุ ที่อยู่ ขนาดครอบครัว สถานะการอยู่ร่วมกันของผู้ปกครอง การศึกษาของมารดา การศึกษาของบิดา อาชีพของมารดา อาชีพของบิดา เหตุผลในการเลือกโรงเรียนนี้ ผู้ปกครอง เวลาเดินทาง เวลาเรียน ความล้มเหลว การสนับสนุนด้านการศึกษาของครอบครัว ซึ่งทั้งหมดนี้ ผู้วิเคราะห์ข้อมูลไม่ได้นำทุก 15 แอตทริบิวต์มาวิเคราะห์ ใช้เพียงบางแอตทริบิวต์ที่จำเป็นเท่านั้น เพื่อการวิเคราะห์ข้อมูลที่มีประสิทธิภาพ

sex	age	address	famsize	Pstatus	Medu	Fedu	Mjob	Fjob	reason	guardian	traveltime	studytime	failures	famsup
F	18	U	GT3	A	4	4	at_home	teacher	course	mother	2	2	0	no
F	17	U	GT3	T	1	1	at_home	other	course	father	1	2	0	yes
F	15	U	LE3	T	1	1	at_home	other	other	mother	1	2	3	no
F	15	U	GT3	T	4	2	health	services	home	mother	1	3	0	yes
F	16	U	GT3	T	3	3	other	other	home	father	1	2	0	yes
M	16	U	LE3	T	4	3	services	other	reputation	mother	1	2	0	yes
M	16	U	LE3	T	2	2	other	other	home	mother	1	2	0	no
F	17	U	GT3	A	4	4	other	teacher	home	mother	2	2	0	yes
M	15	U	LE3	A	3	2	services	other	home	mother	1	2	0	yes
M	15	U	GT3	T	3	4	other	other	home	mother	1	2	0	yes
F	15	U	GT3	T	4	4	teacher	health	reputation	mother	1	2	0	yes
F	15	U	GT3	T	2	1	services	other	reputation	father	3	3	0	yes
M	15	U	LE3	T	4	4	health	services	course	father	1	1	0	yes
M	15	U	GT3	T	4	3	teacher	other	course	mother	2	2	0	yes
M	15	U	GT3	A	2	2	other	other	home	other	1	3	0	yes
F	16	U	GT3	T	4	4	health	other	home	mother	1	1	0	yes
F	16	U	GT3	T	4	4	services	services	reputation	mother	1	3	0	yes
F	16	U	GT3	T	3	3	other	other	reputation	mother	3	2	0	yes
M	17	U	GT3	T	3	2	services	services	course	mother	1	1	3	yes
M	16	U	LE3	T	4	3	health	other	home	father	1	1	0	no
M	15	U	GT3	T	4	3	teacher	other	reputation	mother	1	2	0	no
M	15	U	GT3	T	4	4	health	health	other	father	1	1	0	yes
M	16	U	LE3	T	4	2	teacher	other	course	mother	1	2	0	no
M	16	U	LE3	T	2	2	other	other	reputation	mother	2	2	0	yes
F	15	R	GT3	T	2	4	services	health	course	mother	1	3	0	yes
F	16	U	GT3	T	2	2	services	services	home	mother	1	1	2	yes
M	15	U	GT3	T	2	2	other	other	home	mother	1	1	0	yes
M	15	U	GT3	T	4	2	health	services	other	mother	1	1	0	no
M	16	U	LE3	A	3	4	services	other	home	mother	1	2	0	yes
M	16	U	GT3	T	4	4	teacher	teacher	home	mother	1	2	0	yes
M	15	U	GT3	T	4	4	health	services	home	mother	1	2	0	yes

ภาพที่ 3.2 ข้อมูลนักศึกษาเบื้องต้น

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

3.1.3.1 ทำการคัดเลือกข้อมูล (Data Selection) คือการคัดเลือกข้อมูลที่เหมาะสมเพื่อนำมาใช้ในการวิเคราะห์ข้อมูล

ผู้วิเคราะห์ข้อมูลทำการคัดเลือกข้อมูล และทำการ Data Cleaning ข้อมูลนักศึกษาเบื้องต้น โดยตัดส่วนที่ไม่จำเป็นออกให้เหลือเฉพาะข้อมูลที่จำเป็นในการวิเคราะห์ในภาพรวม จำนวน 15 แอตทริบิวต์ ได้แก่ เพศ อายุ ที่อยู่ ขนาดครอบครัว สถานะการอยู่ร่วมกันของผู้ปกครอง การศึกษาของมารดา การศึกษาของบิดา อาชีพของมารดา อาชีพของบิดา เหตุผลในการเลือกเรียนโรงเรียนนี้ ผู้ปกครองของนักศึกษา เวลาในการเรียน ออกไปเที่ยวข้างนอกกับเพื่อน ความล้มเหลว และการสนับสนุนด้านการศึกษาของครอบครัว โดยใช้ข้อมูล จำนวน 1,044 รายการ ซึ่งเป็นข้อมูลที่จำเป็นในการนำไปวิเคราะห์ข้อมูล

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

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

คอลัมน์ที่จะนำไปทำการวิเคราะห์ โดยการเลือกแอตทริบิวต์หรือคอลัมน์ (Future Selection) จะมีหลายวิธี เช่น Information Gain, Chi-Square, Forward Selection, Backward Elimination

Information Gain เป็นการหาค่าที่มีผลต่อ Parent node มากที่สุด โดยมีสูตรในการคำนวณ โดยหาคอลัมน์ที่มีค่าน้ำหนัก Information Gain มากที่สุดตามลำดับและตามจำนวนที่ผู้วิเคราะห์ ต้องการเลือกคอลัมน์ เพื่อนำไปใช้ในการสร้าง Model Decision tree ต่อไป

Forward Selection เป็นการสร้างโมเดลโดยการเพิ่มพีเจอร์ทีละ 1 พีเจอร์ ถ้าพีเจอร์ที่ใส่เพิ่มให้ประสิทธิภาพที่ดีก็จะเก็บไว้และเลือกพีเจอร์อื่น ๆ มาเพิ่มต่อไปจนประสิทธิภาพของโมเดล ไม่ได้ดีขึ้นก็จะหยุดทำงาน ทฤษฎีนี้ไม่สามารถวัดผลที่แน่ชัด เนื่องจากเป็นวิธีที่ใช้การสุ่มสลับ คอลัมน์ไปเรื่อย ๆ จนกว่าจะได้ผลที่ดีที่สุด เป็นการวัดผลประสิทธิภาพโดยรวม ไม่ใช้การวัดผล คอลัมน์ที่จะนำมาใช้ในการสร้าง Decision tree

Backward Elimination เป็นการสร้างโมเดลที่เริ่มจากการใช้พีเจอร์ทั้งหมดก่อนและตัด (eliminate) พีเจอร์ที่ไม่สำคัญทิ้งไปทีละพีเจอร์ถ้าประสิทธิภาพดีขึ้นก็ตัดพีเจอร์อื่น ๆ ต่อไป ทฤษฎีนี้ไม่สามารถวัดผลที่แน่ชัด เนื่องจากเป็นวิธีที่ใช้การสุ่มสลับคอลัมน์ไปเรื่อย ๆ จนกว่าจะได้ผลที่ดีที่สุด เป็นการวัดผลประสิทธิภาพโดยรวม ไม่ใช้การวัดผลคอลัมน์ที่จะนำมาใช้ในการสร้าง Decision tree

Chi-Square เป็น สถิติที่ใช้ทดสอบความแตกต่างค่าเฉลี่ย ของกลุ่มตัวอย่างที่มีเพียงกลุ่ม หรือสองกลุ่ม ข้อมูลที่นำมาทดสอบนั้นจะต้องเป็นข้อมูลที่อยู่ในระดับการวัด (Measurement Scale) ระดับอันดับภาคชั้น (Interver Scale) หรือระดับอัตราส่วน (Ratio Scale) ซึ่งข้อมูลส่วนใหญ่อยู่ใน รูปแบบตัวเลข ซึ่งไม่สามารถนำมาคำนวณกับชุดข้อมูลที่มีอยู่ได้

ดังนั้นผู้วิเคราะห์ข้อมูลจึงใช้วิธี information Gain ในการคัดเลือกข้อมูล เพราะเป็นวิธีที่หาค่าน้ำหนักของคอลัมน์นั้น ๆ โดยตรงและเห็นผลที่ชัดเจนที่สุด โดยจะคำนวณในขั้นตอนต่อไป

ผู้วิเคราะห์ข้อมูลวิเคราะห์ข้อมูลด้วยการทำเหมืองข้อมูลแบบ Data Classification เพื่อใช้สำหรับเป็นมาตรการในด้านของการศึกษา จากเทคนิคการจำแนกประเภทข้อมูลด้วยการสร้างโมเดล Decision Tree เพื่อจัดกลุ่มข้อมูลที่มีความสัมพันธ์กัน โดยใช้โปรแกรมที่ใช้ทำเหมืองข้อมูล ด้วยชุดข้อมูลที่คัดเลือก ดังนี้

sex	age	address	famsize	Pstatus	Medu	Fedu	Mjob	Fjob	reason	guardian	traveltime	studytime	failures	famsup
F	A	U	LE3	A	PY	MY	services	other	reputation	mother	Normal	Normal	Never	no
M	A	U	GT3	T	PY	PY	services	services	other	mother	Normal	Normal	Never	yes
F	A	U	GT3	T	OT	OT	services	teacher	other	father	Normal	Normal	ever	yes
F	A	R	LE3	T	MY	MY	services	services	course	mother	Normal	Much	Never	yes
F	B	R	GT3	T	OT	OT	other	teacher	other	father	Much	Normal	Never	yes
F	B	R	GT3	T	OT	OT	other	teacher	other	father	Much	Normal	Never	yes
F	A	U	GT3	T	PY	PY	other	other	home	father	Normal	Normal	Never	yes
M	B	R	LE3	T	PY	MY	services	other	course	mother	Much	Normal	Never	no
F	A	R	GT3	T	PY	PT	other	other	course	mother	Normal	Normal	ever	yes
F	B	R	LE3	A	PT	MY	home	other	course	mother	Much	Normal	Never	no
F	A	R	GT3	T	PT	MY	other	other	course	mother	Normal	Normal	Never	no
F	B	U	LE3	T	PT	PT	home	other	reputation	other	Normal	Normal	ever	yes
M	A	U	GT3	T	PY	MY	services	services	course	mother	Normal	Normal	ever	yes
M	A	R	LE3	T	PT	PY	home	other	reputation	father	Much	Normal	Never	yes
F	A	U	GT3	T	OT	MY	other	other	reputation	mother	Normal	Much	Never	yes
M	B	R	GT3	T	PY	PY	other	services	reputation	father	Normal	Normal	ever	no
F	A	R	GT3	T	PT	PT	home	other	reputation	mother	Normal	Normal	Never	yes
F	A	U	LE3	A	PY	PY	other	services	home	mother	Normal	Normal	Never	yes
F	A	U	GT3	T	MY	MY	other	other	course	mother	Normal	Normal	Never	yes
M	B	R	GT3	T	OT	MY	other	other	home	father	Normal	Normal	ever	no
M	A	U	LE3	T	OT	OT	other	teacher	home	father	Normal	Normal	Never	no
F	A	U	GT3	T	PT	PY	home	other	other	father	Normal	Normal	Never	yes
M	A	U	GT3	T	PT	MY	services	services	other	mother	Normal	Normal	Never	yes
F	A	R	GT3	T	PY	OT	services	health	course	mother	Normal	Much	Never	yes
M	A	R	LE3	T	MY	PT	home	other	course	mother	Normal	Normal	ever	no
F	A	U	GT3	T	MY	MY	services	services	home	mother	Normal	Normal	ever	yes
M	B	R	GT3	T	PT	PT	other	other	home	other	Much	Normal	ever	yes
M	B	R	GT3	T	PY	MY	other	other	course	mother	Normal	Normal	Never	yes
F	A	U	LE3	T	MY	MY	services	services	course	father	Normal	Much	Never	no
F	A	U	GT3	T	MY	PY	other	other	other	father	Normal	Normal	Never	yes
M	B	U	LE3	A	PY	OT	other	other	reputation	other	Normal	Normal	Never	no

ภาพที่ 3.3 ชุดข้อมูลที่คัดเลือกมาวิเคราะห์ข้อมูล

จากรูปภาพที่ 3.3 ประกอบด้วย 16 แอตทริบิวต์ คือ

1. เพศของนักศึกษา ประกอบด้วย ('F' คือเพศหญิง หรือ 'M' คือเพศชาย)
2. อายุของนักศึกษา ประกอบด้วย (A=อายุ 15-17 B=อายุ18-20 C=อายุ21 ขึ้นไป)
3. ที่อยู่ของนักศึกษา ประกอบด้วย ('U' คือในเมือง หรือ 'R' คือในชนบท)
4. ขนาดครอบครัว ประกอบด้วย ('LE3' คือน้อยกว่า 3 หรือ 'GT3' ตั้งแต่ 3 ขึ้นไป)
5. สถานะการอยู่ร่วมกันของผู้ปกครอง ประกอบด้วย ('T' คืออยู่ด้วยกัน หรือ 'A' คือแยกกัน)
6. การศึกษาของมารดา ประกอบด้วย ('PT'=ประถม, 'MY'=มัธยม, 'PY'=ปริญญา, 'OT'= อื่น ๆ)
7. การศึกษาของบิดา ประกอบด้วย ('PT'=ประถม, 'MY'=มัธยม, 'PY'=ปริญญา, 'OT'= อื่น ๆ)
8. อาชีพของมารดา ประกอบด้วย ('teacher'=อาจารย์, 'Health'=หมอ/พยาบาล, 'home'=อยู่บ้าน, 'services'=พนักงานทั่วไป, 'other'=อื่น ๆ)
9. อาชีพของบิดา ประกอบด้วย ('teacher'=อาจารย์, 'Health'=หมอ/พยาบาล, 'home'=อยู่บ้าน, 'services'=พนักงานทั่วไป, 'other'=อื่น ๆ)
10. เหตุผลในการเลือกโรงเรียนนี้ ('home'=ใกล้กับ, 'reputation'=ชื่อเสียง, 'course'='หลักสูตร' 'other'=อื่น ๆ)

11. ผู้ปกครองของนักศึกษา ประกอบด้วย ('mother'=มารดา, 'father'=บิดา, 'other'
=อื่น ๆ)
12. เวลาในการเดินทาง ประกอบด้วย ('Normal'=ปกติ, 'Much'=มาก)
13. ออกไปเรียน ประกอบด้วย ('Normal'=ปกติ, 'Much'=มาก)
14. ความล้มเหลว หมายถึง จำนวนความล้มเหลวของคลาสที่ผ่านมา ('ever'=เคย,
'Never'=ไม่เคย)
15. การสนับสนุนด้านการศึกษา ประกอบด้วย ('yes'=สนับสนุน, 'no'=ไม่สนับสนุน)

การสร้างโมเดล Decision tree ในการคัดเลือกแอตทริบิวต์ที่จะนำมาสร้างตัวโมเดล Decision Tree นั้น จะนำแอตทริบิวต์ที่มีค่า Information Gain มากที่สุดเรียงลำดับตามจำนวนที่ต้องการ มาใช้ โดยมีสูตรการคำนวณ ดังนี้

$$IG(\text{Parent, Child}) = \text{Entropy}(\text{Parent}) - [p(c_1) \times \text{Entropy}(c_1) + p(c_2) \times \text{Entropy}(c_2) + \dots]$$

โดยที่ $\text{Entropy}(c_1) = -p(c_1) \log_2 p(c_1)$ และ $p(c_2)$ คือ ค่าความน่าจะเป็นของ c_1

แสดงการคำนวณค่า Information Gain แต่ละแอตทริบิวต์เพื่อหาแอตทริบิวต์ที่มีค่า IG มากที่สุด ได้ดังนี้

1) คำนวณค่า IG ของแอตทริบิวต์ เพศ จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

$$p(\text{Famsup yes}) = 0.60, p(\text{No}) = 0.39$$

$$\begin{aligned} \text{entropy}(\text{parent}) &= -p(\text{Yes}) \times \log_2 p(\text{Yes}) + p(\text{No}) \times \log_2 p(\text{No}) \\ &= -[0.60 \times \log_2(0.60) + 0.39 \times \log_2(0.39)] \\ &= -[0.60 \times -0.73 + 0.39 \times -1.35] \\ &= -[(-0.43) + (-0.52)] \\ &= 0.95 \end{aligned}$$

$$\begin{aligned} \text{entropy}(M) &= -p(\text{Yes}) \times \log_2 p(\text{Yes}) + p(\text{No}) \times \log_2 p(\text{No}) \\ &= -[0.52 \times \log_2(0.52) + 0.47 \times \log_2(0.47)] \\ &= -[0.52 \times -0.94 + 0.47 \times -1.08] \\ &= -[(-0.48) + (-0.50)] \\ &= 0.98 \end{aligned}$$

$$\begin{aligned} \text{entropy}(F) &= -p(\text{Yes}) \times \log_2 p(\text{Yes}) + p(\text{No}) \times \log_2 p(\text{No}) \\ &= -[0.66 \times \log_2(0.66) + 0.33 \times \log_2(0.33)] \\ &= -[0.66 \times -0.59 + 0.33 \times -1.59] \\ &= -[(-0.38) + (-0.52)] \\ &= 0.9 \end{aligned}$$

$$IG(\text{parent, child}) = \text{entropy}(\text{parent}) - [p(M) \times \text{entropy}(M) + p(F) \times \text{entropy}(F)]$$

$$\begin{aligned}
&= 0.95 - [0.41 \times 0.99 + 0.58 \times 0.9] \\
&= 0.95 - [0.40 + 0.52] \\
&= 0.95 - 0.92 \\
&= 0.03
\end{aligned}$$

2) คำนวณค่า IG ของแอตทริบิวต์ อายุ(Age) จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

$$\begin{aligned}
\text{entropy (parent)} &= -p(\text{Yes}) \times \log_2 p(\text{Yes}) + p(\text{No}) \times \log_2 p(\text{No}) \\
&= -[0.61 \times \log_2(0.61) + 0.38 \times \log_2(0.38)] \\
&= -[0.61 \times -0.71 + 0.38 \times -1.39] \\
&= -[(-0.43) + (-0.52)] \\
&= 0.95
\end{aligned}$$

$$\begin{aligned}
\text{entropy(A)} &= -p(\text{Yes}) \times \log_2 p(\text{Yes}) + p(\text{No}) \times \log_2 p(\text{No}) \\
&= -[0.65 \times \log_2(0.65) + 0.34 \times \log_2(0.34)] \\
&= -[0.65 \times -0.62 + 0.34 \times -1.55] \\
&= -[(-0.40) + (-0.52)] \\
&= 0.92
\end{aligned}$$

$$\begin{aligned}
\text{entropy(B)} &= -p(\text{Yes}) \times \log_2 p(\text{Yes}) + p(\text{No}) \times \log_2 p(\text{No}) \\
&= -[0.52 \times \log_2(0.52) + 0.47 \times \log_2(0.47)] \\
&= -[0.52 \times -0.94 + 0.47 \times -1.08] \\
&= -[(-0.48) + (-0.50)] \\
&= 0.98
\end{aligned}$$

$$\begin{aligned}
\text{entropy(C)} &= -p(\text{Yes}) \times \log_2 p(\text{Yes}) + p(\text{No}) \times \log_2 p(\text{No}) \\
&= -[0.25 \times \log_2(0.25) + 0.75 \times \log_2(0.75)] \\
&= -[0.25 \times -2 + 0.75 \times -0.41] \\
&= -[(-0.5) + (-0.30)] \\
&= 0.80
\end{aligned}$$

$$\begin{aligned}
\text{IG(parent, child)} &= \text{entropy (parent)} - [p(A) \times \text{entropy(A)} + p(B) \times \text{entropy(B)} \\
&\quad + p(C) \times \text{entropy(C)}] \\
&= 0.95 - [0.65 \times 0.92 + 0.33 \times 0.98 + 0.005 \times 0.80]
\end{aligned}$$

$$\begin{aligned}
 &= 0.95 - [0.58 + 0.31 + 0.004] \\
 &= 0.95 - 0.89 \\
 &= 0.06
 \end{aligned}$$

- 3) คำนวณค่า IG ของแอตทริบิวต์ ที่อยู่(Address) จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

$$\begin{aligned}
 \text{entropy (parent)} &= -p(\text{Yes}) \times \log_2 p(\text{Yes}) + p(\text{No}) \times \log_2 p(\text{No}) \\
 &= -[0.61 \times \log_2(0.61) + 0.38 \times \log_2(0.38)] \\
 &= -[0.61 \times -0.71 + 0.38 \times -1.39] \\
 &= -[(-0.43) + (-0.52)] \\
 &= 0.95
 \end{aligned}$$

$$\begin{aligned}
 \text{entropy(U)} &= -p(\text{Yes}) \times \log_2 p(\text{Yes}) + p(\text{No}) \times \log_2 p(\text{No}) \\
 &= -[0.61 \times \log_2(0.61) + 0.38 \times \log_2(0.38)] \\
 &= -[0.61 \times -0.71 + 0.38 \times -1.39] \\
 &= -[(-0.43) + (-0.52)] \\
 &= 0.95
 \end{aligned}$$

$$\begin{aligned}
 \text{entropy(R)} &= -p(\text{Yes}) \times \log_2 p(\text{Yes}) + p(\text{No}) \times \log_2 p(\text{No}) \\
 &= -[0.59 \times \log_2(0.59) + 0.40 \times \log_2(0.40)] \\
 &= -[0.59 \times -0.76 + 0.40 \times -1.32] \\
 &= -[(-0.44) + (-0.52)] \\
 &= 0.96
 \end{aligned}$$

$$\begin{aligned}
 \text{IG(parent, child)} &= \text{entropy (parent)} - [p(U) \times \text{entropy(U)} + p(R) \times \text{entropy(R)}] \\
 &= 0.95 - [0.68 \times 0.95 + 0.34 \times 0.96] \\
 &= 0.95 - [0.64 + 0.32] \\
 &= 0.95 - 0.96 \\
 &= -0.01
 \end{aligned}$$

- 4) คำนวณค่า IG ของแอตทริบิวต์ จำนวนบุตร(Famsize) จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

$$\text{entropy (parent)} = -p(\text{Yes}) \times \log_2 p(\text{Yes}) + p(\text{No}) \times \log_2 p(\text{No})$$

$$\begin{aligned}
&= -[0.61 \times \log_2(0.61) + 0.38 \times \log_2(0.38)] \\
&= -[0.61 \times -0.71 + 0.38 \times -1.39] \\
&= -[(-0.43) + (-0.52)] \\
&= 0.95 \\
\text{entropy(LE3)} &= -p(\text{Yes}) \times \log_2 p(\text{Yes}) + p(\text{No}) \times \log_2 p(\text{No}) \\
&= -[0.55 \times \log_2(0.55) + 0.44 \times \log_2(0.44)] \\
&= -[0.55 \times -0.86 + 0.44 \times -1.18] \\
&= -[(-0.47) + (-0.51)] \\
&= 0.98 \\
\text{entropy(GT3)} &= -p(\text{Yes}) \times \log_2 p(\text{Yes}) + p(\text{No}) \times \log_2 p(\text{No}) \\
&= -[0.62 \times \log_2(0.62) + 0.37 \times \log_2(0.37)] \\
&= -[0.62 \times -0.68 + 0.37 \times -1.43] \\
&= -[(-0.42) + (-0.52)] \\
&= 0.94 \\
\text{IG(parent, child)} &= \text{entropy (parent)} - [p(\text{LE3}) \times \text{entropy(LE3)} + p(\text{GT3}) \times \\
&\quad \text{entropy(GT3)}] \\
&= 0.95 - [0.29 \times 0.98 + 0.70 \times 0.94] \\
&= 0.95 - [0.28 + 0.65] \\
&= 0.95 - 0.93 \\
&= 0.02
\end{aligned}$$

5) คำนวณค่า IG ของแอตทริบิวต์ สถานะครอบครัว(Pstatus) จากข้อมูลสามารถ
คำนวณค่า IG ได้ดังนี้

$$\begin{aligned}
\text{entropy (parent)} &= -p(\text{Yes}) \times \log_2 p(\text{Yes}) + p(\text{No}) \times \log_2 p(\text{No}) \\
&= -[0.61 \times \log_2(0.61) + 0.38 \times \log_2(0.38)] \\
&= -[0.61 \times -0.71 + 0.38 \times -1.39] \\
&= -[(-0.43) + (-0.52)] \\
&= 0.95 \\
\text{entropy(A)} &= -p(\text{Yes}) \times \log_2 p(\text{Yes}) + p(\text{No}) \times \log_2 p(\text{No})
\end{aligned}$$

$$\begin{aligned}
&= -[0.56 \times \log_2(0.56) + 0.43 \times \log_2(0.43)] \\
&= -[0.56 \times -0.83 + 0.43 \times -1.21] \\
&= -[(-0.46) + (-0.52)] \\
&= 0.98 \\
\text{entropy}(T) &= -p(\text{Yes}) \times \log_2 p(\text{Yes}) + p(\text{No}) \times \log_2 p(\text{No}) \\
&= -[0.61 \times \log_2(0.61) + 0.38 \times \log_2(0.38)] \\
&= -[0.61 \times -0.71 + 0.38 \times -1.39] \\
&= -[(-0.43) + (-0.52)] \\
&= 0.95 \\
\text{IG}(\text{parent, child}) &= \text{entropy}(\text{parent}) - [p(A) \times \text{entropy}(A) + p(T) \times \text{entropy}(T)] \\
&= 0.95 - [0.11 \times 0.98 + 0.88 \times 0.95] \\
&= 0.95 - [0.10 + 0.83] \\
&= 0.95 - 0.93 \\
&= 0.02
\end{aligned}$$

6) คำนวณค่า IG ของแอตทริบิวต์ การศึกษามารดา(Medu) จากข้อมูลสามารถ
คำนวณค่า IG ได้ดังนี้

$$\begin{aligned}
\text{entropy}(\text{parent}) &= -p(\text{Yes}) \times \log_2 p(\text{Yes}) + p(\text{No}) \times \log_2 p(\text{No}) \\
&= -[0.61 \times \log_2(0.61) + 0.38 \times \log_2(0.38)] \\
&= -[0.61 \times -0.71 + 0.38 \times -1.39] \\
&= -[(-0.43) + (-0.52)] \\
&= 0.95
\end{aligned}$$

$$\begin{aligned}
\text{entropy}(PT) &= -p(\text{Yes}) \times \log_2 p(\text{Yes}) + p(\text{No}) \times \log_2 p(\text{No}) \\
&= -[0.52 \times \log_2(0.52) + 0.47 \times \log_2(0.47)] \\
&= -[0.52 \times -0.94 + 0.47 \times -1.08] \\
&= -[(-0.48) + (-0.50)] \\
&= 0.98
\end{aligned}$$

$$\begin{aligned}
\text{entropy}(MY) &= -p(\text{Yes}) \times \log_2 p(\text{Yes}) + p(\text{No}) \times \log_2 p(\text{No}) \\
&= -[0.56 \times \log_2(0.56) + 0.43 \times \log_2(0.43)]
\end{aligned}$$

$$\begin{aligned}
&= -[0.56 \times -0.83 + 0.43 \times -1.21] \\
&= -[(-0.46) + (-0.52)] \\
&= 0.98 \\
\text{entropy(PY)} &= -p(\text{Yes}) \times \log_2 p(\text{Yes}) + p(\text{No}) \times \log_2 p(\text{No}) \\
&= -[0.63 \times \log_2(0.63) + 0.36 \times \log_2(0.36)] \\
&= -[0.63 \times -0.66 + 0.36 \times -1.47] \\
&= -[(-0.41) + (-0.52)] \\
&= 0.93 \\
\text{entropy(OT)} &= -p(\text{Yes}) \times \log_2 p(\text{Yes}) + p(\text{No}) \times \log_2 p(\text{No}) \\
&= -[0.68 \times \log_2(0.68) + 0.31 \times \log_2(0.31)] \\
&= -[0.68 \times -0.55 + 0.31 \times -1.68] \\
&= -[(-0.37) + (-0.52)] \\
&= 0.89 \\
\text{IG(parent, child)} &= \text{entropy (parent)} - [p(\text{PT}) \times \text{entropy}(\text{PT}) + p(\text{MY}) \\
&\quad \times \text{entropy}(\text{MY}) + p(\text{PY}) \times \text{entropy}(\text{PY})] + p(\text{OT}) \times \\
&\quad \text{entropy}(\text{OT})] \\
&= 0.95 - [0.21 \times 0.98 + 0.26 \times 0.98 + 0.22 \times 0.93 + 0.29 \times \\
&\quad 0.89] \\
&= 0.95 - [0.20 + 0.25 + 0.20 + 0.25] \\
&= 0.95 - 0.9 \\
&= 0.05
\end{aligned}$$

7) คำนวณค่า IG ของแอตทริบิวต์ การศึกษาบิดา(Fedu) จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

$$\begin{aligned}
\text{entropy (parent)} &= -p(\text{Yes}) \times \log_2 p(\text{Yes}) + p(\text{No}) \times \log_2 p(\text{No}) \\
&= -[0.61 \times \log_2(0.61) + 0.38 \times \log_2(0.38)] \\
&= -[0.61 \times -0.71 + 0.38 \times -1.39] \\
&= -[(-0.43) + (-0.52)] \\
&= 0.95
\end{aligned}$$

$$\begin{aligned}
 \text{entropy(PT)} &= - p(\text{Yes}) \times \log_2 p(\text{Yes}) + p(\text{No}) \times \log_2 p(\text{No}) \\
 &= -[0.52 \times \log_2(0.52) + 0.47 \times \log_2(0.47)] \\
 &= -[0.52 \times -0.94 + 0.47 \times -1.08] \\
 &= -[(-0.48) + (-0.50)] \\
 &= 0.98
 \end{aligned}$$

$$\begin{aligned}
 \text{entropy(MY)} &= - p(\text{Yes}) \times \log_2 p(\text{Yes}) + p(\text{No}) \times \log_2 p(\text{No}) \\
 &= -[0.58 \times \log_2(0.58) + 0.41 \times \log_2(0.41)] \\
 &= -[0.58 \times -0.78 + 0.41 \times -1.28] \\
 &= -[(-0.45) + (-0.52)] \\
 &= 0.97
 \end{aligned}$$

$$\begin{aligned}
 \text{entropy(PY)} &= - p(\text{Yes}) \times \log_2 p(\text{Yes}) + p(\text{No}) \times \log_2 p(\text{No}) \\
 &= -[0.63 \times \log_2(0.63) + 0.36 \times \log_2(0.36)] \\
 &= -[0.63 \times -0.66 + 0.36 \times -1.47] \\
 &= -[(-0.41) + (-0.52)] \\
 &= 0.93
 \end{aligned}$$

$$\begin{aligned}
 \text{entropy(OT)} &= - p(\text{Yes}) \times \log_2 p(\text{Yes}) + p(\text{No}) \times \log_2 p(\text{No}) \\
 &= -[0.69 \times \log_2(0.69) + 0.30 \times \log_2(0.30)] \\
 &= -[0.69 \times -0.53 + 0.30 \times -1.73] \\
 &= -[(-0.36) + (-0.52)] \\
 &= 0.88
 \end{aligned}$$

$$\begin{aligned}
 \text{IG(parent, child)} &= \text{entropy (parent)} - [p(\text{PT}) \times \text{entropy(PT)} + p(\text{MY}) \\
 &\times \text{entropy(MY)} + p(\text{PY}) \times \text{entropy(PY)}] + p(\text{OT}) \times \\
 &\text{entropy(OT)} \\
 &= 0.95 - [0.25 \times 0.98 + 0.31 \times 0.97 + 0.21 \times 0.93 + 0.21 \times \\
 &\quad 0.88] \\
 &= 0.95 - [0.24 + 0.30 + 0.19 + 0.18] \\
 &= 0.95 - 0.91 \\
 &= 0.04
 \end{aligned}$$

8) คำนวณค่า IG ของแอตทริบิวต์ อาชีพมารดา จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

$$\begin{aligned}
 \text{entropy (parent)} &= -p(\text{Yes}) \times \log_2 p(\text{Yes}) + p(\text{No}) \times \log_2 p(\text{No}) \\
 &= -[0.61 \times \log_2(0.61) + 0.38 \times \log_2(0.38)] \\
 &= -[0.61 \times -0.71 + 0.38 \times -1.39] \\
 &= -[(-0.43) + (-0.52)] \\
 &= 0.95
 \end{aligned}$$

$$\begin{aligned}
 \text{entropy(Services)} &= -p(\text{Yes}) \times \log_2 p(\text{Yes}) + p(\text{No}) \times \log_2 p(\text{No}) \\
 &= -[0.65 \times \log_2(0.65) + 0.34 \times \log_2(0.34)] \\
 &= -[0.65 \times -0.62 + 0.34 \times -1.55] \\
 &= -[(-0.40) + (-0.52)] \\
 &= 0.92
 \end{aligned}$$

$$\begin{aligned}
 \text{entropy(health)} &= -p(\text{Yes}) \times \log_2 p(\text{Yes}) + p(\text{No}) \times \log_2 p(\text{No}) \\
 &= -[0.75 \times \log_2(0.75) + 0.25 \times \log_2(0.25)] \\
 &= -[0.75 \times -0.41 + 0.25 \times -2] \\
 &= -[(-0.30) + (-0.5)] \\
 &= 0.8
 \end{aligned}$$

$$\begin{aligned}
 \text{entropy(Teacher)} &= -p(\text{Yes}) \times \log_2 p(\text{Yes}) + p(\text{No}) \times \log_2 p(\text{No}) \\
 &= -[0.63 \times \log_2(0.63) + 0.36 \times \log_2(0.36)] \\
 &= -[0.63 \times -0.66 + 0.36 \times -1.47] \\
 &= -[(-0.41) + (-0.52)] \\
 &= 0.93
 \end{aligned}$$

$$\begin{aligned}
 \text{entropy(Home)} &= -p(\text{Yes}) \times \log_2 p(\text{Yes}) + p(\text{No}) \times \log_2 p(\text{No}) \\
 &= -[0.61 \times \log_2(0.61) + 0.38 \times \log_2(0.38)] \\
 &= -[0.61 \times -0.71 + 0.38 \times -1.39] \\
 &= -[(-0.43) + (-0.52)] \\
 &= 0.95
 \end{aligned}$$

$$\text{entropy(other)} = -p(\text{Yes}) \times \log_2 p(\text{Yes}) + p(\text{No}) \times \log_2 p(\text{No})$$

$$\begin{aligned}
&= -[0.53 \times \log_2(0.53) + 0.46 \times \log_2(0.46)] \\
&= -[0.53 \times -0.91 + 0.46 \times -1.12] \\
&= -[(-0.48) + (-0.51)] \\
&= 0.99 \\
\text{IG(parent, child)} &= \text{entropy (parent)} - [p(\text{Services}) \times \text{entropy}(\text{Services}) + \\
&\quad p(\text{health}) \times \text{entropy}(\text{health}) + p(\text{Teacher}) \\
&\quad \times \text{entropy}(\text{Teacher}) + p(\text{home}) \times \text{entropy}(\text{home}) + \\
&\quad p(\text{other}) \times \text{entropy}(\text{other})] \\
&= 0.95 - [0.21 \times 0.92 + 0.07 \times 0.8 + 0.12 \times 0.93 + 0.20 \times \\
&\quad 0.95 + 0.37 \times 0.99] \\
&= 0.95 - [0.19 + 0.05 + 0.11 + 0.19 + 0.36] \\
&= 0.95 - 0.90 \\
&= 0.05
\end{aligned}$$

9) คำนวณค่า IG ของแอตทริบิวต์ อาชีพบิดา จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

$$\begin{aligned}
\text{entropy (parent)} &= -p(\text{Yes}) \times \log_2 p(\text{Yes}) + p(\text{No}) \times \log_2 p(\text{No}) \\
&= -[0.61 \times \log_2(0.61) + 0.38 \times \log_2(0.38)] \\
&= -[0.61 \times -0.71 + 0.38 \times -1.39] \\
&= -[(-0.43) + (-0.52)] \\
&= 0.95
\end{aligned}$$

$$\begin{aligned}
\text{entropy}(\text{Services}) &= -p(\text{Yes}) \times \log_2 p(\text{Yes}) + p(\text{No}) \times \log_2 p(\text{No}) \\
&= -[0.57 \times \log_2(0.57) + 0.42 \times \log_2(0.42)] \\
&= -[0.57 \times -0.81 + 0.42 \times -1.25] \\
&= -[(-0.46) + (-0.52)] \\
&= 0.98
\end{aligned}$$

$$\begin{aligned}
\text{entropy}(\text{health}) &= -p(\text{Yes}) \times \log_2 p(\text{Yes}) + p(\text{No}) \times \log_2 p(\text{No}) \\
&= -[0.81 \times \log_2(0.81) + 0.18 \times \log_2(0.18)] \\
&= -[0.81 \times -0.30 + 0.18 \times -2.47] \\
&= -[(-0.24) + (-0.44)]
\end{aligned}$$

$$\begin{aligned}
 &= 0.68 \\
 \text{entropy(Teacher)} &= -p(\text{Yes}) \times \log_2 p(\text{Yes}) + p(\text{No}) \times \log_2 p(\text{No}) \\
 &= -[0.66 \times \log_2(0.66) + 0.33 \times \log_2(0.33)] \\
 &= -[0.66 \times -0.59 + 0.33 \times -1.59] \\
 &= -[(-0.38) + (-0.52)] \\
 &= 0.90 \\
 \text{entropy(Home)} &= -p(\text{Yes}) \times \log_2 p(\text{Yes}) + p(\text{No}) \times \log_2 p(\text{No}) \\
 &= -[0.66 \times \log_2(0.66) + 0.34 \times \log_2(0.34)] \\
 &= -[0.66 \times -0.59 + 0.34 \times -1.55] \\
 &= -[(-0.38) + (-0.52)] \\
 &= 0.90 \\
 \text{entropy(other)} &= -p(\text{Yes}) \times \log_2 p(\text{Yes}) + p(\text{No}) \times \log_2 p(\text{No}) \\
 &= -[0.59 \times \log_2(0.59) + 0.40 \times \log_2(0.40)] \\
 &= -[0.59 \times -0.76 + 0.40 \times -1.32] \\
 &= -[(-0.44) + (-0.52)] \\
 &= 0.96 \\
 \text{IG(parent, child)} &= \text{entropy (parent)} - [p(\text{Services}) \times \text{entropy}(\text{Services}) + \\
 & \quad p(\text{health}) \times \text{entropy}(\text{health}) + p(\text{Teacher}) \\
 & \quad \times \text{entropy}(\text{Teacher}) + p(\text{home}) \times \text{entropy}(\text{home}) + \\
 & \quad p(\text{other}) \times \text{entropy}(\text{other})] \\
 &= 0.95 - [0.28 \times 0.98 + 0.03 \times 0.68 + 0.05 \times 0.90 + 0.06 \times \\
 & \quad 0.90 + 0.54 \times 0.96] \\
 &= 0.95 - [0.27 + 0.02 + 0.04 + 0.05 + 0.51] \\
 &= 0.95 - 0.89 \\
 &= 0.04
 \end{aligned}$$

10) คำนวณค่า IG ของแอตทริบิวต์ เหตุผลที่มาเรียน(reason) จากข้อมูลสามารถ
คำนวณค่า IG ได้ดังนี้

$$\text{entropy (parent)} = -p(\text{Yes}) \times \log_2 p(\text{Yes}) + p(\text{No}) \times \log_2 p(\text{No})$$

$$\begin{aligned}
&= -[0.61 \times \log_2(0.61) + 0.38 \times \log_2(0.38)] \\
&= -[0.61 \times -0.71 + 0.38 \times -1.39] \\
&= -[(-0.43) + (-0.52)] \\
&= 0.95 \\
\text{entropy(reput)} &= -p(\text{Yes}) \times \log_2 p(\text{Yes}) + p(\text{No}) \times \log_2 p(\text{No}) \\
&= -[0.66 \times \log_2(0.66) + 0.33 \times \log_2(0.33)] \\
&= -[0.66 \times -0.59 + 0.33 \times -1.59] \\
&= -[(-0.38) + (-0.52)] \\
&= 0.90 \\
\text{entropy(course)} &= -p(\text{Yes}) \times \log_2 p(\text{Yes}) + p(\text{No}) \times \log_2 p(\text{No}) \\
&= -[0.60 \times \log_2(0.60) + 0.39 \times \log_2(0.39)] \\
&= -[0.60 \times -0.73 + 0.39 \times -1.35] \\
&= -[(-0.43) + (-0.52)] \\
&= 0.95 \\
\text{entropy(home)} &= -p(\text{Yes}) \times \log_2 p(\text{Yes}) + p(\text{No}) \times \log_2 p(\text{No}) \\
&= -[0.59 \times \log_2(0.59) + 0.40 \times \log_2(0.40)] \\
&= -[0.59 \times -0.76 + 0.40 \times -1.32] \\
&= -[(-0.44) + (-0.52)] \\
&= 0.96 \\
\text{entropy(other)} &= -p(\text{Yes}) \times \log_2 p(\text{Yes}) + p(\text{No}) \times \log_2 p(\text{No}) \\
&= -[0.49 \times \log_2(0.49) + 0.50 \times \log_2(0.50)] \\
&= -[0.49 \times -1.02 + 0.50 \times -1] \\
&= -[(-0.49) + (-0.5)] \\
&= 0.99 \\
\text{IG(parent, child)} &= \text{entropy (parent)} - [p(\text{reput}) \times \text{entropy(reput)} + p(\text{course}) \\
&\quad \times \text{entropy(course)} + p(\text{home}) \times \text{entropy(home)} + \\
&\quad p(\text{other}) \times \text{entropy(other)}]
\end{aligned}$$

$$\begin{aligned}
&= 0.95 - [0.23 \times 0.90 + 0.41 \times 0.95 + 0.24 \times 0.96 + \\
&\quad 0.11 \times 0.99] \\
&= 0.95 - [0.2 + 0.38 + 0.23 + 0.11] \\
&= 0.95 - 0.92 \\
&= 0.03
\end{aligned}$$

11) คำนวณค่า IG ของแอตทริบิวต์ ผู้ปกครอง(guardian) จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

$$\begin{aligned}
\text{entropy (parent)} &= -p(\text{Yes}) \times \log_2 p(\text{Yes}) + p(\text{No}) \times \log_2 p(\text{No}) \\
&= -[0.61 \times \log_2(0.61) + 0.38 \times \log_2(0.38)] \\
&= -[0.61 \times -0.71 + 0.38 \times -1.39] \\
&= -[(-0.43) + (-0.52)] \\
&= 0.95
\end{aligned}$$

$$\begin{aligned}
\text{entropy(mother)} &= -p(\text{Yes}) \times \log_2 p(\text{Yes}) + p(\text{No}) \times \log_2 p(\text{No}) \\
&= -[0.6 \times \log_2(0.6) + 0.4 \times \log_2(0.4)] \\
&= -[0.6 \times -0.73 + 0.4 \times -1.32] \\
&= -[(-0.43) + (-0.52)] \\
&= 0.95
\end{aligned}$$

$$\begin{aligned}
\text{entropy(father)} &= -p(\text{Yes}) \times \log_2 p(\text{Yes}) + p(\text{No}) \times \log_2 p(\text{No}) \\
&= -[0.62 \times \log_2(0.62) + 0.37 \times \log_2(0.37)] \\
&= -[0.62 \times -0.68 + 0.37 \times -1.43] \\
&= -[(-0.42) + (-0.52)] \\
&= 0.94
\end{aligned}$$

$$\begin{aligned}
\text{entropy(other)} &= -p(\text{Yes}) \times \log_2 p(\text{Yes}) + p(\text{No}) \times \log_2 p(\text{No}) \\
&= -[0.61 \times \log_2(0.61) + 0.38 \times \log_2(0.38)] \\
&= -[0.61 \times -0.71 + 0.38 \times -1.39] \\
&= -[(-0.43) + (-0.52)] \\
&= 0.95
\end{aligned}$$

$$\begin{aligned}
 IG(\text{parent, child}) &= \text{entropy}(\text{parent}) - [p(\text{mother}) \times \text{entropy}(\text{mother}) + \\
 & p(\text{father}) \times \text{entropy}(\text{father}) + p(\text{other}) \times \text{entropy}(\text{other})] \\
 &= 0.95 - [0.68 \times 0.95 + 0.23 \times 0.94 + 0.08 \times 0.95] \\
 &= 0.95 - [0.64 + 0.21 + 0.07] \\
 &= 0.95 - 0.92 \\
 &= 0.03
 \end{aligned}$$

12) คำนวณค่า IG ของแอตทริบิวต์ เวลาเที่ยว(traveltime) จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

Famsup yes=0.60, No= 0.39

$$\begin{aligned}
 \text{entropy}(\text{parent}) &= -p(\text{Yes}) \times \log_2 p(\text{Yes}) + p(\text{No}) \times \log_2 p(\text{No}) \\
 &= -[0.60 \times \log_2(0.60) + 0.39 \times \log_2(0.39)] \\
 &= -[0.60 \times -0.73 + 0.39 \times -1.35] \\
 &= -[(-0.43) + (-0.52)] \\
 &= 0.95
 \end{aligned}$$

$$\begin{aligned}
 \text{entropy}(\text{Normal}) &= -p(\text{Yes}) \times \log_2 p(\text{Yes}) + p(\text{No}) \times \log_2 p(\text{No}) \\
 &= -[0.59 \times \log_2(0.59) + 0.4 \times \log_2(0.4)] \\
 &= -[0.59 \times -0.76 + 0.4 \times -1.32] \\
 &= -[(-0.44) + (-0.52)] \\
 &= 0.96
 \end{aligned}$$

$$\begin{aligned}
 \text{entropy}(\text{Much}) &= -p(\text{Yes}) \times \log_2 p(\text{Yes}) + p(\text{No}) \times \log_2 p(\text{No}) \\
 &= -[0.65 \times \log_2(0.65) + 0.34 \times \log_2(0.34)] \\
 &= -[0.65 \times -0.62 + 0.34 \times -1.55] \\
 &= -[(-0.4) + (-0.52)] \\
 &= 0.92
 \end{aligned}$$

$$\begin{aligned}
 IG(\text{parent, child}) &= \text{entropy}(\text{parent}) - [p(\text{Normal}) \times \text{entropy}(\text{Normal}) + \\
 & p(\text{Much}) \times \text{entropy}(\text{Much})] \\
 &= 0.95 - [0.89 \times 0.96 + 0.1 \times 0.92] \\
 &= 0.95 - [0.85 + 0.092]
 \end{aligned}$$

$$= 0.95 - 0.94$$

$$= 0.01$$

13) คำนวณค่า IG ของแอตทริบิวต์ เวลาเรียน(studytime) จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

Famsup yes=0.60, No= 0.39

$$\begin{aligned} \text{entropy (parent)} &= -p(\text{Yes}) \times \log_2 p(\text{Yes}) + p(\text{No}) \times \log_2 p(\text{No}) \\ &= -[0.60 \times \log_2(0.60) + 0.39 \times \log_2(0.39)] \\ &= -[0.60 \times -0.73 + 0.39 \times -1.35] \\ &= -[(-0.43) + (-0.52)] \\ &= 0.95 \end{aligned}$$

$$\begin{aligned} \text{entropy(Normal)} &= -p(\text{Yes}) \times \log_2 p(\text{Yes}) + p(\text{No}) \times \log_2 p(\text{No}) \\ &= -[0.59 \times \log_2(0.59) + 0.4 \times \log_2(0.4)] \\ &= -[0.59 \times -0.76 + 0.4 \times -1.32] \\ &= -[(-0.44) + (-0.52)] \\ &= 0.96 \end{aligned}$$

$$\begin{aligned} \text{entropy(Much)} &= -p(\text{Yes}) \times \log_2 p(\text{Yes}) + p(\text{No}) \times \log_2 p(\text{No}) \\ &= -[0.65 \times \log_2(0.65) + 0.34 \times \log_2(0.34)] \\ &= -[0.65 \times -0.62 + 0.34 \times -1.55] \\ &= -[(-0.4) + (-0.52)] \\ &= 0.92 \end{aligned}$$

$$\begin{aligned} \text{IG(parent, child)} &= \text{entropy (parent)} - [p(\text{Normal}) \times \text{entropy(Normal)} + \\ &\quad p(\text{Much}) \times \text{entropy(Much)}] \\ &= 0.95 - [0.89 \times 0.96 + 0.1 \times 0.92] \\ &= 0.95 - [0.85 + 0.092] \\ &= 0.95 - 0.94 \\ &= 0.01 \end{aligned}$$

14) คำนวณค่า IG ของแอตทริบิวต์ เวลาเรียน(studytime) จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

Famsup yes=0.60, No= 0.39

$$\begin{aligned}
 \text{entropy (parent)} &= -p(\text{Yes}) \times \log_2 p(\text{Yes}) + p(\text{No}) \times \log_2 p(\text{No}) \\
 &= -[0.60 \times \log_2(0.60) + 0.39 \times \log_2(0.39)] \\
 &= -[0.60 \times -0.73 + 0.39 \times -1.35] \\
 &= -[(-0.43) + (-0.52)] \\
 &= 0.95
 \end{aligned}$$

$$\begin{aligned}
 \text{entropy(ever)} &= -p(\text{Yes}) \times \log_2 p(\text{Yes}) + p(\text{No}) \times \log_2 p(\text{No}) \\
 &= -[0.57 \times \log_2(0.57) + 0.42 \times \log_2(0.42)] \\
 &= -[0.57 \times -0.81 + 0.42 \times -1.25] \\
 &= -[(-0.46) + (-0.52)] \\
 &= 0.98
 \end{aligned}$$

$$\begin{aligned}
 \text{entropy(Never)} &= -p(\text{Yes}) \times \log_2 p(\text{Yes}) + p(\text{No}) \times \log_2 p(\text{No}) \\
 &= -[0.61 \times \log_2(0.61) + 0.38 \times \log_2(0.38)] \\
 &= -[0.61 \times -0.71 + 0.38 \times -1.39] \\
 &= -[(-0.43) + (-0.52)] \\
 &= 0.95
 \end{aligned}$$

$$\begin{aligned}
 \text{IG(parent, child)} &= \text{entropy (parent)} - [p(\text{ever}) \times \text{entropy(ever)} + p(\text{Never}) \\
 &\quad \times \text{entropy(Never)}] \\
 &= 0.95 - [0.2 \times 0.98 + 0.79 \times 0.95] \\
 &= 0.95 - [0.19 + 0.75] \\
 &= 0.95 - 0.94 \\
 &= 0.01
 \end{aligned}$$

การคำนวณค่าแต่ละแอตทริบิวต์เทียบกับคลาสเพื่อหาแอตทริบิวต์ที่มีค่า Information Gain มากที่สุดค่า แสดงดังในตาราง ทั้งนี้ผู้วิเคราะห์ได้ทำการเลือก 5 แอตทริบิวต์ เหตุผลในที่เลือก 5 แอตทริบิวต์ เนื่องจากหากเลือกแอตทริบิวต์มากกว่านี้จะทำให้โมเดล Decision tree เยอะมาก จนเกินไปทำให้ยากต่อการนำไปใช้ในการตัดสินใจ

ดังนั้นจึงเลือกแอตทริบิวต์ ที่มีค่า Information Gain มากที่สุด 5 อันดับแรก ได้แก่ แอตทริบิวต์ Age, Medu, Fedu, Mjob, Fjob จากนั้นนำข้อมูลที่มี 5 แอตทริบิวต์ นี้ไปทดสอบประสิทธิภาพ ต่อด้วยการนำไปสร้างเป็นโมเดล Decision tree

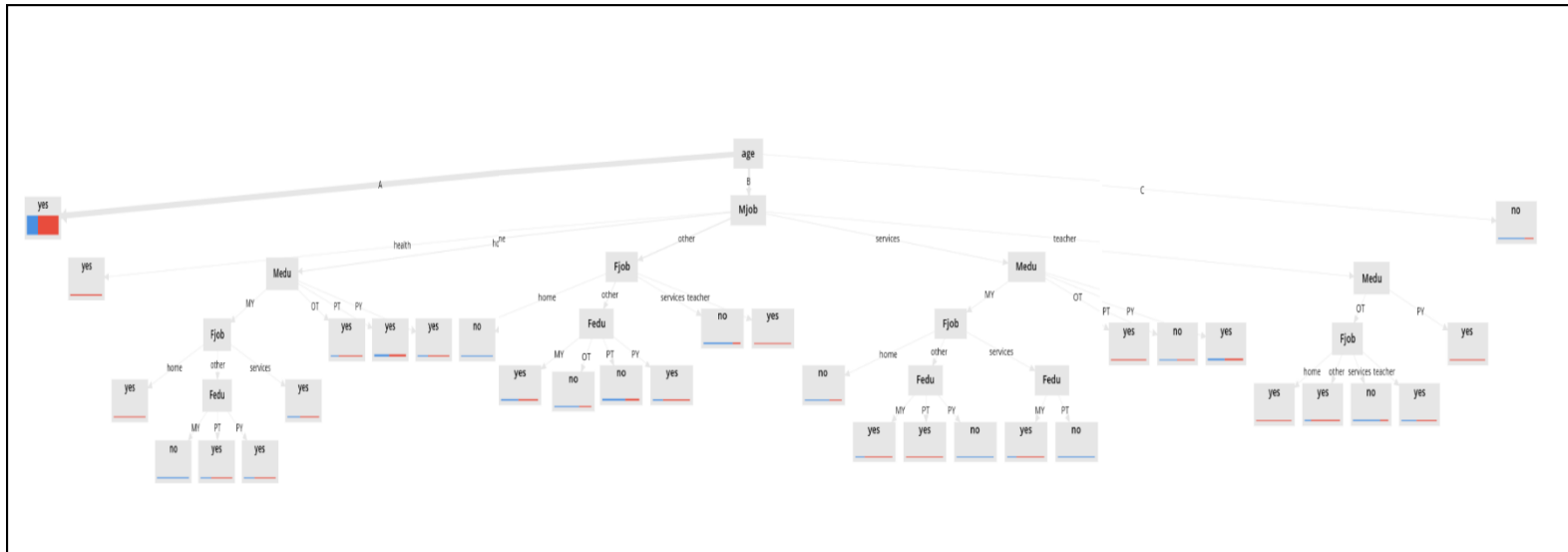
ตาราง 3.1 สรุปค่า Information Gain ของคอลัมน์ต่าง ๆ

Feature	Information Gain	Select
Sex	0.03	
Age	0.06	✓
Address	-0.01	
Famsize	0.02	
Pstatus	0.02	
Medu	0.05	✓
Fedu	0.04	✓
Mjob	0.05	✓
Fjob	0.04	✓
Reason	0.03	
guardian	0.03	
Traveltime	0.01	
Studytime	0.01	
Failures	0.01	

3.1.5 การประเมินผล (Evaluation) เป็นขั้นตอนก่อนนำผลลัพธ์ที่ได้จากขั้นตอนก่อนหน้า โดยนำไปใช้งาน ด้วยการวัดประสิทธิผลของผลลัพธ์ที่ได้กับวัตถุประสงค์ที่ตั้งไว้ในขั้นตอนแรก เพื่อแสดงให้เห็นว่ามีนัยสำคัญหรือความน่าเชื่อถือมากน้อยเพียงใด ด้วยการประเมินผลจากโปรแกรมว่าถูกต้องหรือไม่

จากกลุ่มข้อมูลสำหรับการเรียนรู้ (Training Data) ที่ได้เตรียมไว้จำนวน 731 ชุด นำมาวิเคราะห์กับแบบจำลอง โดยนำกลุ่มข้อมูลสำหรับการเรียนรู้มาทดสอบกับอัลกอริทึมต้นไม้ (Tree) คือ C4.5 (J48) ผลจากการวิเคราะห์ในโปรแกรม แสดงได้ดังนี้

1) โปรแกรม Rapid Miner Studio 9.5.1



ภาพที่ 3.4 รูปแบบโมเดล Graph Decision Tree ในโปรแกรม Rapid Miner Studio 9.5.1

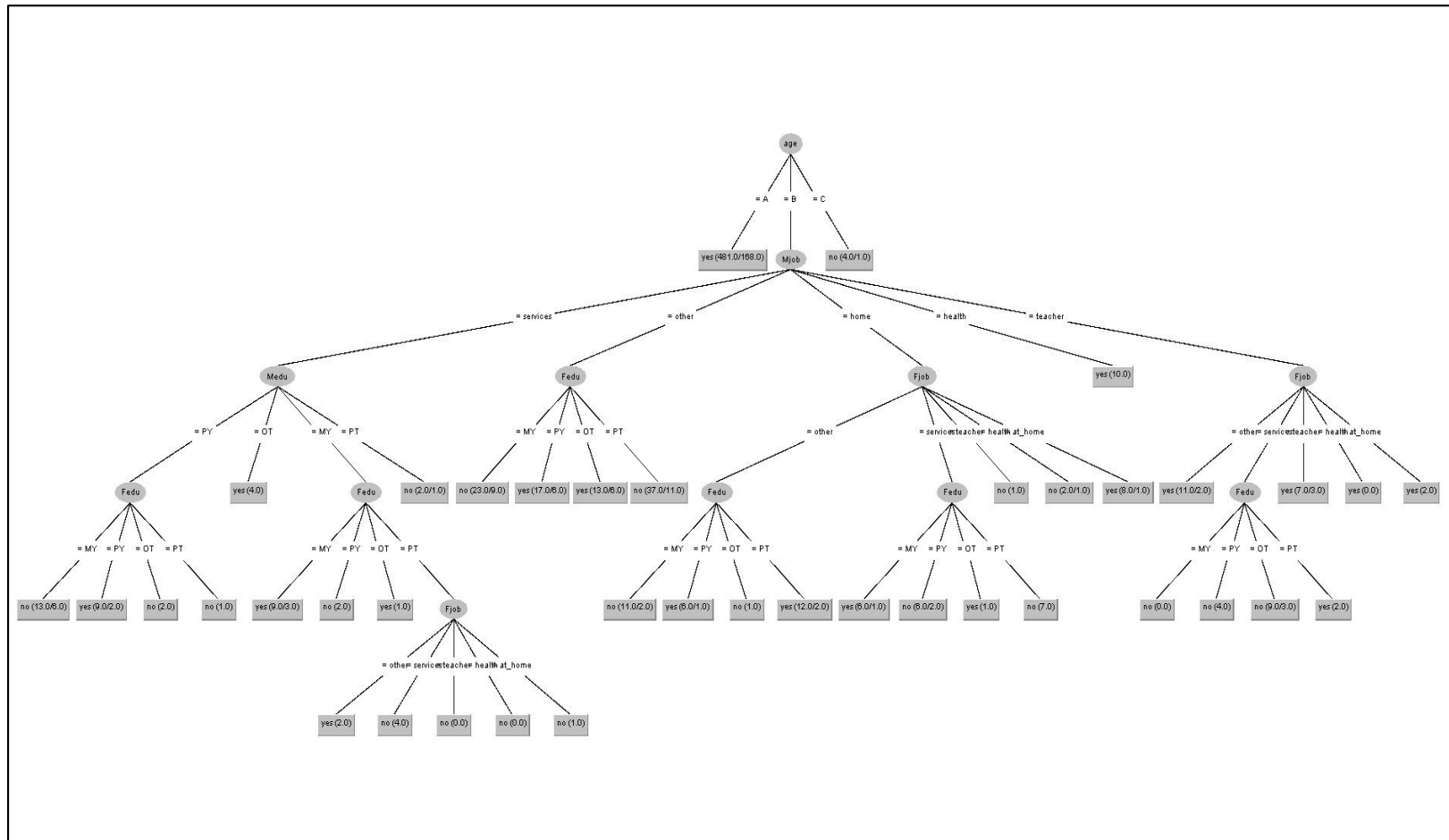
```

Tree
age = A: yes {no=168, yes=313}
age = B
| Mjob = health: yes {no=0, yes=10}
| Mjob = home
| | Medu = MY
| | | Fjob = at_home: yes {no=0, yes=2}
| | | Fjob = other
| | | | Fedu = MY: no {no=6, yes=0}
| | | | Fedu = PT: yes {no=1, yes=2}
| | | | Fedu = PY: yes {no=1, yes=2}
| | | Fjob = services: yes {no=2, yes=3}
| | Medu = OT: yes {no=1, yes=3}
| | Medu = PT: yes {no=15, yes=17}
| | Medu = PY: yes {no=2, yes=4}
| Mjob = other
| | Fjob = at_home: no {no=2, yes=0}
| | Fjob = other
| | | Fedu = MY: yes {no=8, yes=9}
| | | Fedu = OT: no {no=6, yes=3}
| | | Fedu = PT: no {no=18, yes=11}
| | | Fedu = PY: yes {no=3, yes=8}
| | Fjob = services: no {no=15, yes=4}
| | Fjob = teacher: yes {no=0, yes=3}
| Mjob = services
| | Medu = MY
| | | Fjob = at_home: no {no=2, yes=1}
| | | Fjob = other
| | | | Fedu = MY: yes {no=1, yes=3}
| | | | Fedu = PT: yes {no=0, yes=2}
| | | | Fedu = PY: no {no=2, yes=0}
| | | Fjob = services
| | | | Fedu = MY: yes {no=1, yes=3}
| | | | Fedu = PT: no {no=4, yes=0}
| | | Medu = OT: yes {no=0, yes=4}
| | | Medu = PT: no {no=1, yes=1}
| | | Medu = PY: yes {no=12, yes=13}
| Mjob = teacher
| | Medu = OT
| | | Fjob = at_home: yes {no=0, yes=2}
| | | Fjob = other: yes {no=2, yes=9}
| | | Fjob = services: no {no=10, yes=3}
| | | Fjob = teacher: yes {no=3, yes=4}
| | Medu = PY: yes {no=0, yes=2}
age = C: no {no=3, yes=1}

```

ภาพที่ 3.5 คำบรรยายลักษณะงาน Decision Tree ในโปรแกรม Rapid Miner Studio 9.5.1

2) โปรแกรม Weka 3.8.4



ภาพที่ 3.6 รูปแบบโมเดล Graph Decision Tree ในโปรแกรม Weka 3.8.4

```

age = A: yes (481.0/168.0)
age = B
| Mjob = services
| | Medu = PY
| | | Fedu = MY: no (13.0/6.0)
| | | Fedu = PY: yes (9.0/2.0)
| | | Fedu = OT: no (2.0)
| | | Fedu = PT: no (1.0)
| | Medu = OT: yes (4.0)
| | Medu = MY
| | | Fedu = MY: yes (9.0/3.0)
| | | Fedu = PY: no (2.0)
| | | Fedu = OT: yes (1.0)
| | | Fedu = PT
| | | | Fjob = other: yes (2.0)
| | | | Fjob = services: no (4.0)
| | | | Fjob = teacher: no (0.0)
| | | | Fjob = health: no (0.0)
| | | | Fjob = at_home: no (1.0)
| | Medu = PT: no (2.0/1.0)
| Mjob = other
| | Fedu = MY: no (23.0/9.0)
| | Fedu = PY: yes (17.0/6.0)
| | Fedu = OT: yes (13.0/6.0)
| | Fedu = PT: no (37.0/11.0)
| Mjob = home
| | Fjob = other
| | | Fedu = MY: no (11.0/2.0)
| | | Fedu = PY: yes (6.0/1.0)
| | | Fedu = OT: no (1.0)
| | | Fedu = PT: yes (12.0/2.0)
| | Fjob = services
| | | Fedu = MY: yes (6.0/1.0)
| | | Fedu = PY: no (6.0/2.0)
| | | Fedu = OT: yes (1.0)
| | | Fedu = PT: no (7.0)
| | Fjob = teacher: no (1.0)
| | Fjob = health: no (2.0/1.0)
| | Fjob = at_home: yes (8.0/1.0)
| Mjob = health: yes (10.0)
| Mjob = teacher
| | Fjob = other: yes (11.0/2.0)
| | Fjob = services
| | | Fedu = MY: no (0.0)
| | | Fedu = PY: no (4.0)
| | | Fedu = OT: no (9.0/3.0)
| | | Fedu = PT: yes (2.0)
| | Fjob = teacher: yes (7.0/3.0)
| | Fjob = health: yes (0.0)
| | Fjob = at_home: yes (2.0)
age = C: no (4.0/1.0)

```

=== Summary ===

Correctly Classified Instances	500	68.3995 %
Incorrectly Classified Instances	231	31.6005 %
Kappa statistic	0.2695	
Mean absolute error	0.4123	
Root mean squared error	0.4541	
Relative absolute error	86.2344 %	
Root relative squared error	92.8682 %	
Total Number of Instances	731	

ภาพที่ 3.7 คำบรรยายลักษณะงาน Decision Tree ในโปรแกรม Weka 3.8.4

ส่วนวัดประสิทธิภาพของโมเดล(Evaluation on training set) สามารถพิจารณา ผลได้ดังนี้

- 1) Correctly Classified Instances จากข้อมูล โมเดลสามารถทำนายข้อมูลถูกต้อง คิดเป็น 68.3995 % ของทั้งหมด
- 2) Incorrectly Classified Instances จากข้อมูล โมเดลทำนาย ข้อมูลไม่ถูกต้อง คิดเป็น 31.6005 % ของทั้งหมด

3) Root Mean Squared Error (RMSE) ค่าความคลาดเคลื่อนระหว่างค่าจริงและค่าที่พยากรณ์ได้ มีค่าเท่ากับ 0.4541 ดังภาพที่ 3.7

4) ส่วนค่า Confusion Matrix เป็นส่วนที่แสดงรายละเอียดผลการทำนายของข้อมูลแต่ละคลาสได้ชัดเจนขึ้น ค่าในคอลัมน์คือค่าที่ได้จากการทำนายด้วยเทคนิค C4.5 (J48) และ ค่าในแถวจะเป็นส่วนของค่าที่เป็นคำตอบของคลาสนั้นจริง ดังภาพ 3.8

```

=== Detailed Accuracy By Class ===

              TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
              0.325   0.081   0.723     0.325   0.449     0.312   0.668    0.579    no
              0.919   0.675   0.676     0.919   0.779     0.312   0.668    0.719    yes
Weighted Avg.  0.684   0.440   0.694     0.684   0.648     0.312   0.668    0.664

=== Confusion Matrix ===

  a  b  <-- classified as
94 195 |  a = no
36 406 |  b = yes

```

ภาพที่ 3.8 ส่วนค่า Confusion Matrix ของโมเดลโดยใช้อัลกอริทึม C4.5 (J48)

เมื่อได้โมเดลแล้วซึ่งก็คือโมเดลที่ถูกสร้างจากอัลกอริทึม C4.5 (J48) ขั้นตอนต่อมาคือการทดสอบ โมเดลกับชุดข้อมูลทดสอบ (Testing Data) เพื่อทดสอบความถูกต้องของโมเดล โดยนำข้อมูลที่จัดเตรียมไว้เป็นชุดข้อมูลทดสอบจำนวน 313 ชุด มาทดสอบกับโมเดลที่ถูกสร้างจากอัลกอริทึม ต้นไม้ C4.5 (J48) ผลการใช้โมเดลดังกล่าว ทดสอบกับชุดข้อมูลทดสอบจำนวน 731ชุด แสดงผลการวิเคราะห์ในโปรแกรม Weka ได้ดังภาพที่ 3.9 ผลการใช้โมเดลจากอัลกอริทึมต้นไม้ C4.5 (J48) ทดสอบกับชุดข้อมูลทดสอบจำนวน 313 ชุด ให้ผลการทดสอบโดยมีค่าความถูกต้อง (Correctly Classified Instances) เท่ากับ 64.5161 % ซึ่ง เป็นค่าความถูกต้องที่ปานกลาง ส่วนการวิเคราะห์ข้อมูลที่ทำนายไม่ถูกต้องมีค่าเท่ากับ 35.4839%


```

=== Summary ===

Correctly Classified Instances      180           64.5161 %
Incorrectly Classified Instances    99           35.4839 %
Kappa statistic                    0.0285
Mean absolute error                0.4472
Root mean squared error            0.4739
Relative absolute error            95.2123 %
Root relative squared error        98.6981 %
Total Number of Instances         279

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                0.051   0.028   0.500     0.051   0.092     0.058   0.535    0.390    no
                0.972   0.949   0.651     0.972   0.780     0.058   0.535    0.667    yes
Weighted Avg.   0.645   0.622   0.597     0.645   0.535     0.058   0.535    0.569

```

ภาพที่ 3.9 ผลการทดสอบของโมเดลโดยใช้อัลกอริทึม C4.5 (J48) ทดสอบกับข้อมูลทดสอบ

เมื่อได้โมเดล Decision tree แล้วจากนั้น จึงนำโมเดลมาใช้ในการเขียน กฎ Decision Tree เพื่อใช้ในการตัดสินใจ โดยกฎมีดังต่อไปนี้

กฎข้อที่ 1 IF age=a then YES หมายความว่า ถ้า อายุเท่ากับ 15-17 จะมีผลลัพธ์เป็น “สนับสนุน”

กฎข้อที่ 2 IF age=b and mjob=services and medu=py and fedu=my then NO หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นพนักงานทั่วไป และการศึกษามารดาเป็นปริญญา และการศึกษาบิดาเป็นมัธยม จะมีผลลัพธ์เป็น “ไม่สนับสนุน”

กฎข้อที่ 3 IF age=b and mjob=services and medu=py and fedu=py then YES หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นพนักงานทั่วไป และการศึกษามารดาเป็นปริญญา และการศึกษาบิดาเป็นปริญญา จะมีผลลัพธ์เป็น “สนับสนุน”

กฎข้อที่ 4 IF age=b and mjob=services and medu=py and fedu=ot then NO หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นพนักงานทั่วไป และการศึกษามารดาเป็นปริญญา และการศึกษาบิดาเป็นอื่น ๆ จะมีผลลัพธ์เป็น “ไม่สนับสนุน”

กฎข้อที่ 5 IF age=b and mjob=services and medu=py and fedu=pt then NO หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นพนักงานทั่วไป และการศึกษามารดาเป็นปริญญา และการศึกษาบิดาเป็นประถม จะมีผลลัพธ์เป็น “ไม่สนับสนุน”

กฎข้อที่ 6 IF age=b and mjob=services and medu=ot then YES หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นพนักงานทั่วไป และการศึกษามารดาเป็นอื่น ๆ จะมีผลลัพธ์เป็น “สนับสนุน”

กฎข้อที่ 7 IF age=b and mjob=services and medu=my and fedu=my then YES หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นพนักงานทั่วไป และการศึกษามารดาเป็นมัธยม และการศึกษาบิดาเป็นมัธยม จะมีผลลัพธ์เป็น “สนับสนุน”

กฎข้อที่ 8 IF age=b and mjob=services and medu=my and fedu=py then NO หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นพนักงานทั่วไป และการศึกษามารดาเป็นมัธยม และการศึกษาบิดาเป็นปริญญา จะมีผลลัพธ์เป็น “สนับสนุน”

กฎข้อที่ 9 IF age=b and mjob=services and medu=my and fedu=ot then YES หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นพนักงานทั่วไป และการศึกษามารดาเป็นมัธยม และการศึกษาบิดาเป็นอื่น ๆ จะมีผลลัพธ์เป็น “สนับสนุน”

กฎข้อที่ 10 IF age=b and mjob=services and medu=my and fedu=pt and fjob=other then YES หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นพนักงานทั่วไป และการศึกษามารดาเป็นมัธยม และการศึกษาบิดาเป็นประถม และอาชีพบิดาเป็นอื่น ๆ จะมีผลลัพธ์เป็น “สนับสนุน”

กฎข้อที่ 11 IF age=b and mjob=services and medu=my and fedu=pt and fjob=services then NO หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นพนักงานทั่วไป และการศึกษามารดาเป็นมัธยม และการศึกษาบิดาเป็นประถม และอาชีพบิดาเป็นพนักงานทั่วไป จะมีผลลัพธ์เป็น “ไม่สนับสนุน”

กฎข้อที่ 12 IF age=b and mjob=services and medu=my and fedu=pt and fjob=teacher then NO หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นพนักงานทั่วไป และการศึกษามารดาเป็นมัธยม และการศึกษาบิดาเป็นประถม และอาชีพบิดาเป็นครู/อาจารย์ จะมีผลลัพธ์เป็น “ไม่สนับสนุน”

กฎข้อที่ 13 IF age=b and mjob=services and medu=my and fedu=pt and fjob=health then NO หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นพนักงานทั่วไป และการศึกษามารดาเป็นมัธยม และการศึกษาบิดาเป็นประถม และอาชีพบิดาเป็นหมอ/พยาบาล จะมีผลลัพธ์เป็น “ไม่สนับสนุน”

กฎข้อที่ 14 IF age=b and mjob=services and medu=my and fedu=pt and fjob=home then NO หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นพนักงานทั่วไป และการศึกษามารดาเป็นมัธยม และการศึกษาบิดาเป็นประถม และอาชีพบิดาเป็นอยู่บ้าน จะมีผลลัพธ์เป็น “ไม่สนับสนุน”

กฎข้อที่ 15 IF age=b and mjob=services and medu=pt then NO หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นพนักงานทั่วไป และการศึกษามารดาเป็นประถม จะมีผลลัพธ์เป็น “ไม่สนับสนุน”

กฎข้อที่ 16 IF age=b and mjob=other and fedu=my then NO หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นอื่น ๆ และการศึกษาบิดาเป็นมัธยม จะมีผลลัพธ์เป็น “ไม่สนับสนุน”

กฎข้อที่ 17 IF age=b and mjob=other and fedu=py then YES หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นอื่น ๆ และการศึกษาบิดาเป็นปริญญา จะมีผลลัพธ์เป็น “ไม่สนับสนุน”

กฎข้อที่ 18 IF age=b and mjob=other and fedu=ot then YES หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นอื่น ๆ และการศึกษาบิดาเป็นอื่น ๆ จะมีผลลัพธ์เป็น “ไม่สนับสนุน”

กฎข้อที่ 19 IF age=b and mjob=other and fedu=pt then NO หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นอื่น ๆ และการศึกษาบิดาเป็นประถม จะมีผลลัพธ์เป็น “ไม่สนับสนุน”

กฎข้อที่ 20 IF age=b and mjob=home and fjob=other and fedu=my then NO หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นอยู่บ้าน และอาชีพบิดาเป็นอื่น ๆ และการศึกษาบิดาเป็นมัธยม จะมีผลลัพธ์เป็น “ไม่สนับสนุน”

กฎข้อที่ 21 IF age=b and mjob=home and fjob=other and fedu=py then YES หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นอยู่บ้าน และอาชีพบิดาเป็นอื่น ๆ และการศึกษาบิดาเป็นปริญญา จะมีผลลัพธ์เป็น “สนับสนุน”

กฎข้อที่ 22 IF age=b and mjob=home and fjob=other and fedu=ot then NO หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นอยู่บ้าน และอาชีพบิดาเป็นอื่น ๆ และการศึกษาบิดาเป็นอื่น ๆ จะมีผลลัพธ์เป็น “ไม่สนับสนุน”

กฎข้อที่ 23 IF age=b and mjob=home and fjob=other and fedu=pt then YES หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นอยู่บ้าน และอาชีพบิดาเป็นอื่น ๆ และการศึกษาบิดาเป็นประถม จะมีผลลัพธ์เป็น “สนับสนุน”

กฎข้อที่ 24 IF age=b and mjob=home and fjob=services and fedu=my then YES หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นอยู่บ้าน และอาชีพบิดาเป็นพนักงานทั่วไป และการศึกษาบิดาเป็นมัธยม จะมีผลลัพธ์เป็น “สนับสนุน”

กฎข้อที่ 25 IF age=b and mjob=home and fjob=services and fedu=py then NO หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นอยู่บ้าน และอาชีพบิดาเป็นพนักงานทั่วไป และการศึกษาบิดาเป็นปริญญา จะมีผลลัพธ์เป็น “ไม่สนับสนุน”

กฎข้อที่ 26 IF age=b and mjob=home and fjob=services and fedu=ot then NO หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นอยู่บ้าน และอาชีพบิดาเป็นพนักงานทั่วไป และการศึกษาบิดาเป็นอื่น ๆ จะมีผลลัพธ์เป็น “ไม่สนับสนุน”

กฎข้อที่ 27 IF age=b and mjob=home and fjob=services and fedu=pt then NO หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นอยู่บ้าน และอาชีพบิดาเป็นพนักงานทั่วไป และการศึกษาบิดาเป็นประถม จะมีผลลัพธ์เป็น “ไม่สนับสนุน”

กฎข้อที่ 28 IF age=b and mjob=home and fjob=teacher then NO หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นอยู่บ้าน และอาชีพบิดาเป็นครู/อาจารย์ จะมีผลลัพธ์เป็น “ไม่สนับสนุน”

กฎข้อที่ 29 IF age=b and mjob=home and fjob=health then NO หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นอยู่บ้าน และอาชีพบิดาเป็นหมอ/พยาบาล จะมีผลลัพธ์เป็น “ไม่สนับสนุน”

กฎข้อที่ 30 IF age=b and mjob=home and fjob=home then YES หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นอยู่บ้าน และอาชีพบิดาเป็นอยู่บ้านจะมีผลลัพธ์เป็น “สนับสนุน”

กฎข้อที่ 31 IF age=b and mjob=health then YES หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นหมอ/พยาบาล จะมีผลลัพธ์เป็น “สนับสนุน”

กฎข้อที่ 32 IF age=b and mjob=teacher and fjob=other then YES หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นครู/อาจารย์ และอาชีพบิดาเป็นอื่น ๆ จะมีผลลัพธ์เป็น “สนับสนุน”

กฎข้อที่ 33 IF age=b and mjob=teacher and fjob=services and fedu=my then NO หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นครู/อาจารย์ และอาชีพบิดาเป็นพนักงานทั่วไป และการศึกษาบิดาเป็นมัธยม จะมีผลลัพธ์เป็น “ไม่สนับสนุน”

กฎข้อที่ 34 IF age=b and mjob=teacher and fjob=services and fedu=py then NO หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นครู/อาจารย์ และอาชีพบิดาเป็นพนักงานทั่วไป และการศึกษาบิดาเป็นปริญญา จะมีผลลัพธ์เป็น “ไม่สนับสนุน”

กฎข้อที่ 35 IF age=b and mjob=teacher and fjob=services and fedu=ot then NO หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นครู/อาจารย์ และอาชีพบิดาเป็นพนักงานทั่วไป และการศึกษาบิดาเป็นอื่น ๆ จะมีผลลัพธ์เป็น “ไม่สนับสนุน”

กฎข้อที่ 36 IF age=b and mjob=teacher and fjob=services and fedu=pt then YES หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นครู/อาจารย์ และอาชีพบิดาเป็นพนักงานทั่วไป และการศึกษาบิดาเป็นประถม จะมีผลลัพธ์เป็น “สนับสนุน”

กฎข้อที่ 37 IF age=b and mjob=teacher and fjob=teacher then YES หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นครู/อาจารย์ และอาชีพบิดาเป็นครู/อาจารย์ จะมีผลลัพธ์เป็น “สนับสนุน”

กฎข้อที่ 38 IF age=b and mjob=teacher and fjob=health then YES หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นครู/อาจารย์ และอาชีพบิดาเป็นหมอ/พยาบาล จะมีผลลัพธ์เป็น “สนับสนุน”

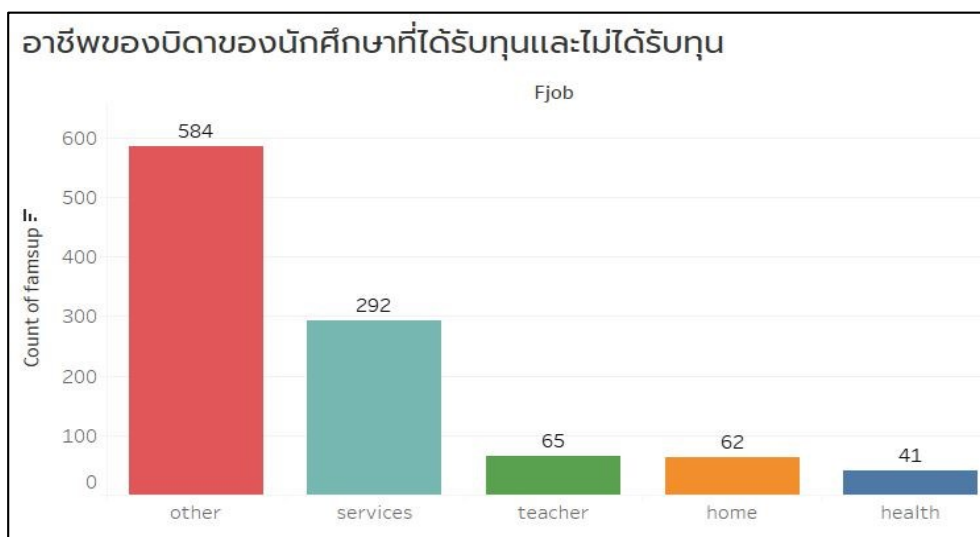
กฎข้อที่ 39 IF age=b and mjob=teacher and fjob=home then YES หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นครู/อาจารย์ และอาชีพบิดาเป็นอยู่บ้าน จะมีผลลัพธ์เป็น “สนับสนุน”

กฎข้อที่ 40 IF age=c then YES หมายความว่า ถ้าอายุเท่ากับ 21 ขึ้นไป จะมีผลลัพธ์เป็น “สนับสนุน”

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

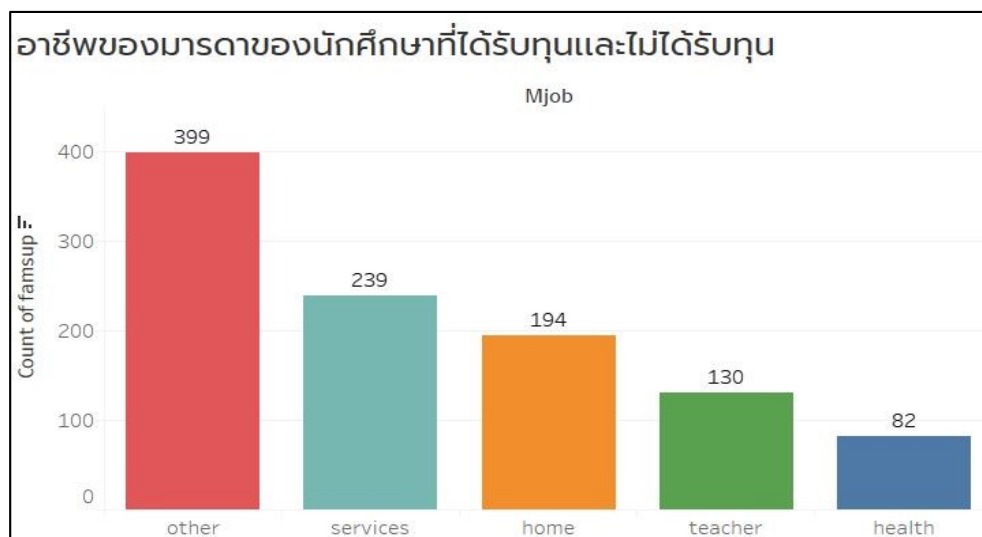
ผู้วิเคราะห์ข้อมูลนำผลข้อมูลที่ทำการวิเคราะห์แล้ว แสดงผลข้อมูลบน Web Browser โดยใช้ชุดคำสั่ง HTML CSS3 และ PHP ร่วมกับการนำเสนอข้อมูลแบบ Visualization ด้วยการแสดงผลข้อมูลในรูปแบบของภาพโดยใช้โปรแกรม Tableau Public

1) แสดงข้อมูลแสดงผลประเภทของอาชีพบิดาของนักศึกษาที่ได้รับทุนและไม่ได้รับทุนเรียงตามลำดับ ซึ่งเป็นการแสดงผลในรูปแบบแผนภูมิแท่ง ผู้ใช้สามารถเลือกดูแต่ละอาชีพของพ่อแม่ได้



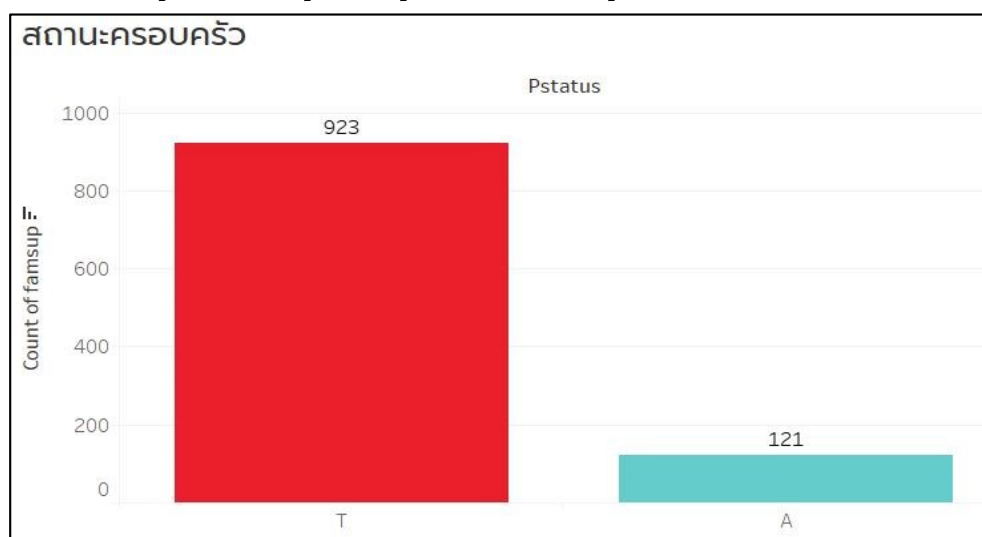
ภาพที่ 3.10 แสดงข้อมูลแสดงผลประเภทของอาชีพบิดาของนักศึกษาที่ได้รับทุนและไม่ได้รับทุน

2) แสดงข้อมูลแสดงผลประเภทของอาชีพมารดาของนักศึกษาที่ได้รับทุนและไม่ได้รับทุน เรียงตามลำดับ ซึ่งเป็นการแสดงผลในรูปแบบแผนภูมิแท่ง ผู้ใช้สามารถเลือกดูแต่ละอาชีพของมารดาได้



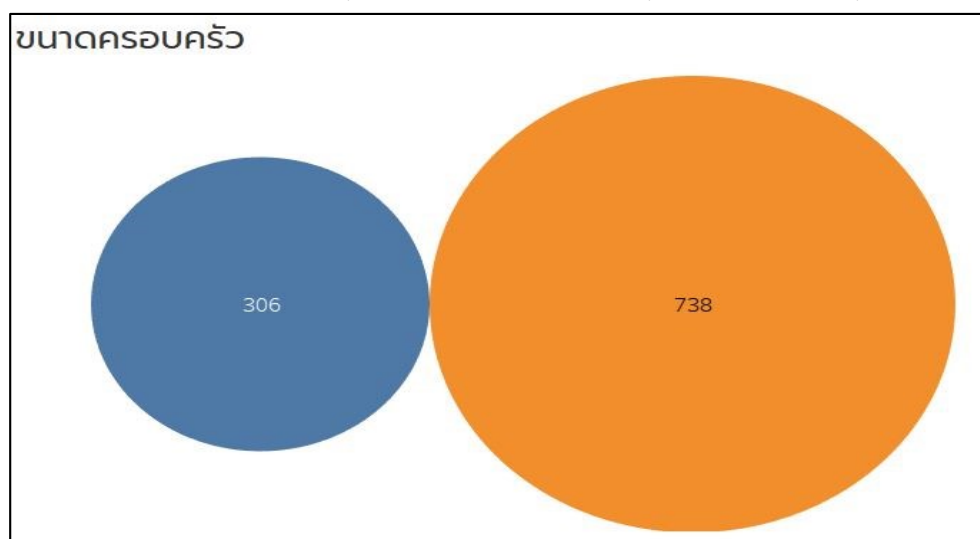
ภาพที่ 3.11 แสดงข้อมูลแสดงผลประเภทของอาชีพมารดาของนักศึกษาที่ได้รับทุนและไม่ได้รับทุน

3) แสดงข้อมูลแสดงผลประเภทของการอยู่กับครอบครัวและการอยู่ห่างจากครอบครัว ซึ่งเป็นการแสดงผลในรูปแบบแผนภูมิแท่ง ผู้ใช้สามารถเลือกดูแต่ละประเภทได้



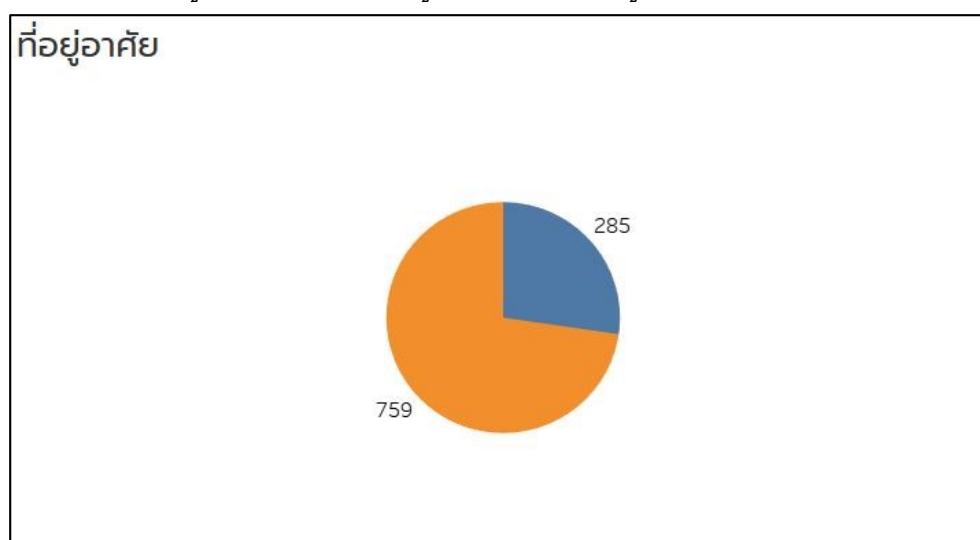
ภาพที่ 3.12 แสดงข้อมูลแสดงผลประเภทของเปอร์เซ็นต์การอยู่กับครอบครัวและการอยู่ห่าง

4) แสดงข้อมูลแสดงผลประเภทของครอบครัวมากกว่า 3 สนับสนุนการเรียนเท่าไร ไม่สนับสนุนการเรียนเท่าไร, ครอบครัวน้อยกว่า 3 สนับสนุนการเรียนเท่าไร ไม่สนับสนุนการเรียนเท่าไร ซึ่งเป็นการแสดงผลในรูปแบบ packed bubbles ผู้ใช้สามารถเลือกดูแต่ละประเภทได้



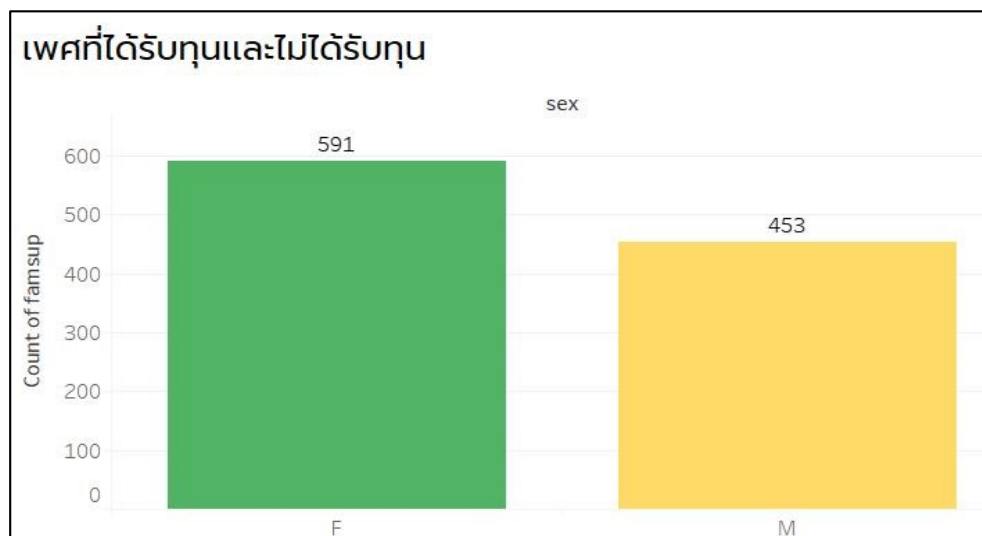
ภาพที่ 3.13 แสดงข้อมูลแสดงผลประเภทของครอบครัวตั้งแต่ 3 ขึ้นไปสนับสนุนการเรียนเท่าไร ไม่สนับสนุนการเรียนเท่าไร, ครอบครัวน้อยกว่า 3 สนับสนุนการเรียนเท่าไร ไม่สนับสนุนการเรียนเท่าไร

5) แสดงข้อมูลแสดงผลประเภทของการอยู่อาศัยในเมืองหรือนอกเมืองของนักศึกษา ซึ่งเป็นการแสดงผลในรูปแบบ Pie charts ผู้ใช้สามารถเลือกดูแต่ละประเภทได้



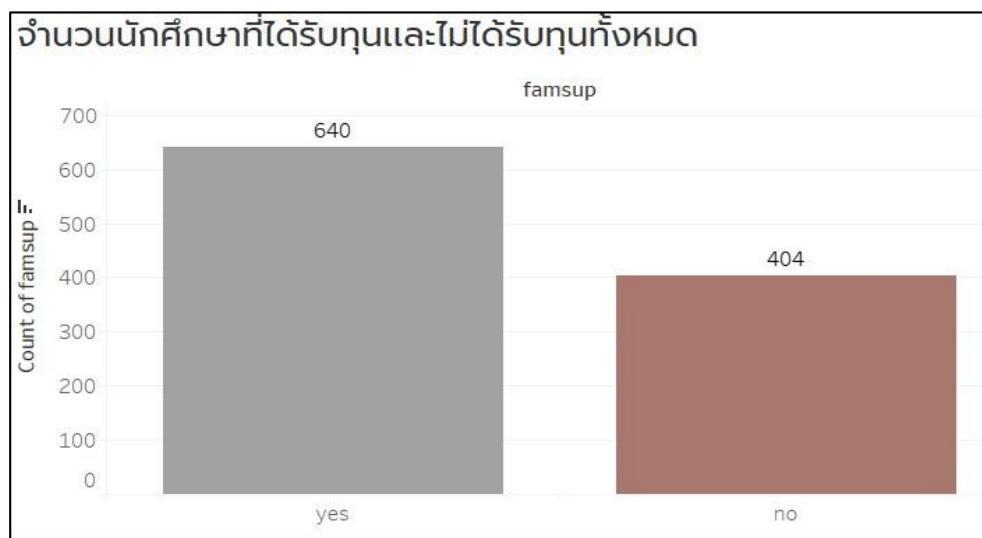
ภาพที่ 3.14 แสดงข้อมูลแสดงผลประเภทของการอยู่อาศัยในเมืองหรือนอกเมืองของนักศึกษา

6) แสดงข้อมูลแสดงผลจำนวนที่ได้รับทุนของเพศชายและเพศหญิง ซึ่งเป็นการแสดงผลในรูปแบบแผนภูมิแท่ง ผู้ใช้สามารถเลือกดูแต่ละประเภทได้



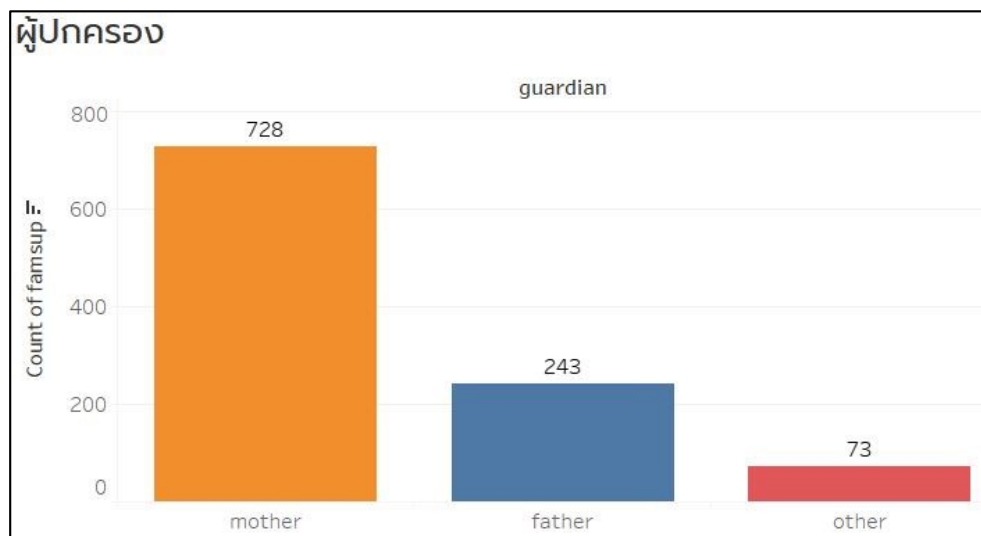
ภาพที่ 3.15 แสดงข้อมูลแสดงผลจำนวนที่ได้รับทุนของเพศชายและเพศหญิง

7) แสดงข้อมูลแสดงผลการได้รับทุนการศึกษา และไม่ได้รับทุนการศึกษาทั้งหมด ซึ่งเป็นการแสดงผลในรูปแบบแผนภูมิแท่ง



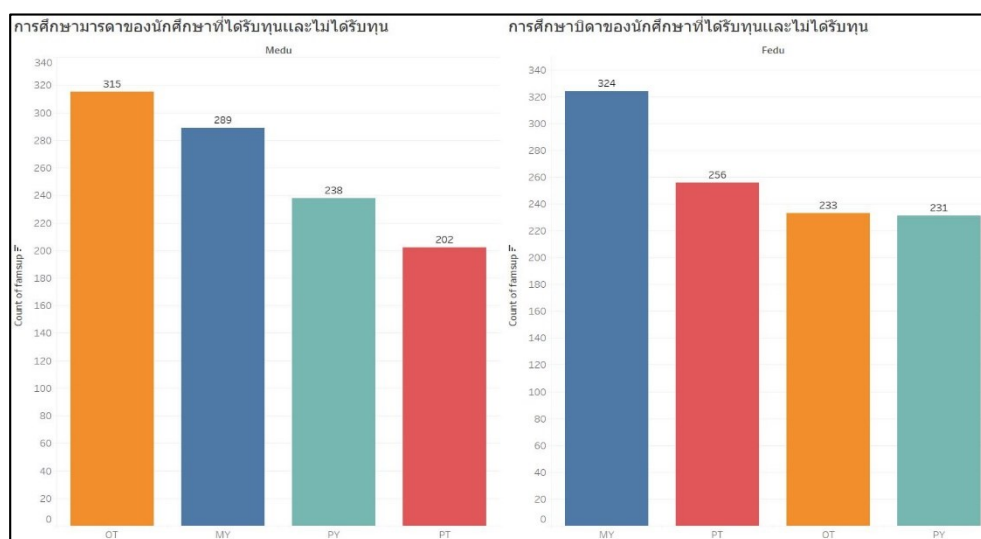
ภาพที่ 3.16 แสดงข้อมูลแสดงผลการได้รับทุนการศึกษา และไม่ได้รับทุนการศึกษาทั้งหมด

8) แสดงข้อมูลแสดงผลผู้ปกครองของนักศึกษาที่ได้รับทุนและไม่ได้รับทุน ซึ่งเป็นการแสดงผลในรูปแบบแผนภูมิแท่ง ผู้ใช้สามารถเลือกดูแต่ละประเภทได้



ภาพที่ 3.17 แสดงข้อมูลแสดงผลผู้ปกครองของนักศึกษาที่ได้รับทุนและไม่ได้รับทุน

9) แสดงข้อมูลแสดงผลประเภทของการศึกษาปีตามารดาของนักศึกษาที่ได้รับทุนและไม่ได้รับทุนเรียงตามลำดับ ซึ่งเป็นการแสดงผลในรูปแบบแผนภูมิแท่ง ผู้ใช้สามารถเลือกดูแต่ละอาชีพของพ่อแม่ได้



ภาพที่ 3.18 แสดงข้อมูลแสดงผลประเภทของการศึกษาปีตามารดาของนักศึกษาที่ได้รับทุนและไม่ได้รับทุน

3.2 การออกแบบเว็บไซต์

3.2.1 การออกแบบ Wireframe หน้าจอเว็บไซต์

1) หน้าโฮมเพจของเว็บไซต์ แสดงเมนูต่าง ๆ ของหน้าเว็บไซต์ ข่าวสาร และ Dashboard



ภาพที่ 3.19 หน้าโฮมเพจของเว็บไซต์

2) หน้าเว็บเพจที่ 2 ชุดข้อมูล

The screenshot displays a web application interface for data management. On the left, a dark blue sidebar contains navigation options: 'หน้าแรก' (Home), 'ข้อมูล' (Data), 'ข้อมูลสารสนเทศ' (Information Data), 'กระบวนการ' (Process), 'ผลการประเมินผล' (Evaluation Results), and 'การประเมินผลสัมฤทธิ์' (Performance Evaluation). The main content area features a search bar, a filter dropdown set to 'All Information', and a 'List' button. Below this is a large table with a dark blue header and multiple empty rows. At the bottom, there is a 'CONTACT' section with three placeholder icons.

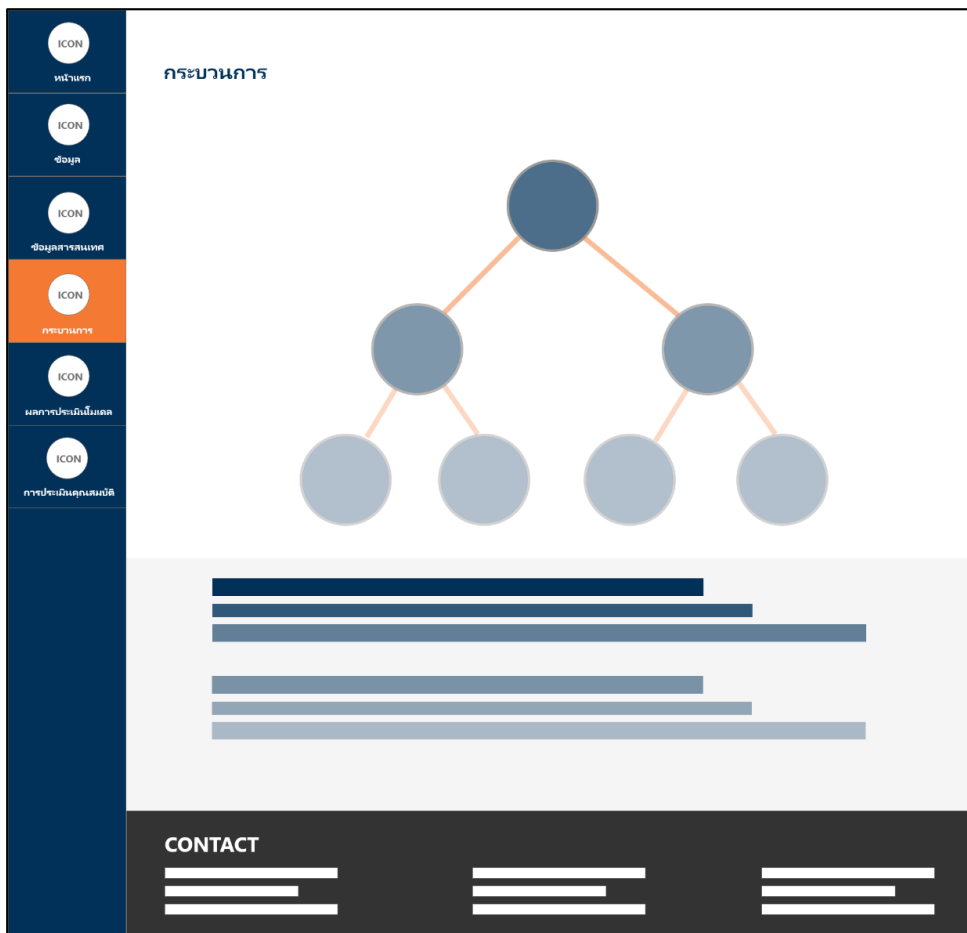
ภาพที่ 3.20 หน้าเว็บเพจที่ 2 ชุดข้อมูล

3) หน้าเว็บเพจที่ 3 ชุดข้อมูลสารสนเทศ

The screenshot shows a web application interface with a dark blue sidebar on the left containing navigation links: หน้าแรก, ข้อมูล, ข้อมูลสารสนเทศ (highlighted), กระบวนการ, ผลการประเมินผล, and การประเมินคุณสมบัตินิติ. The main content area features a search bar with the text "Search...", a list of items with "All Information xx" and "List" labels, and a "Show information" button with "xx" and "To" labels. Below this is a large table with a header row and multiple empty rows. At the bottom, there is a "CONTACT" section with three sets of horizontal lines representing contact information.

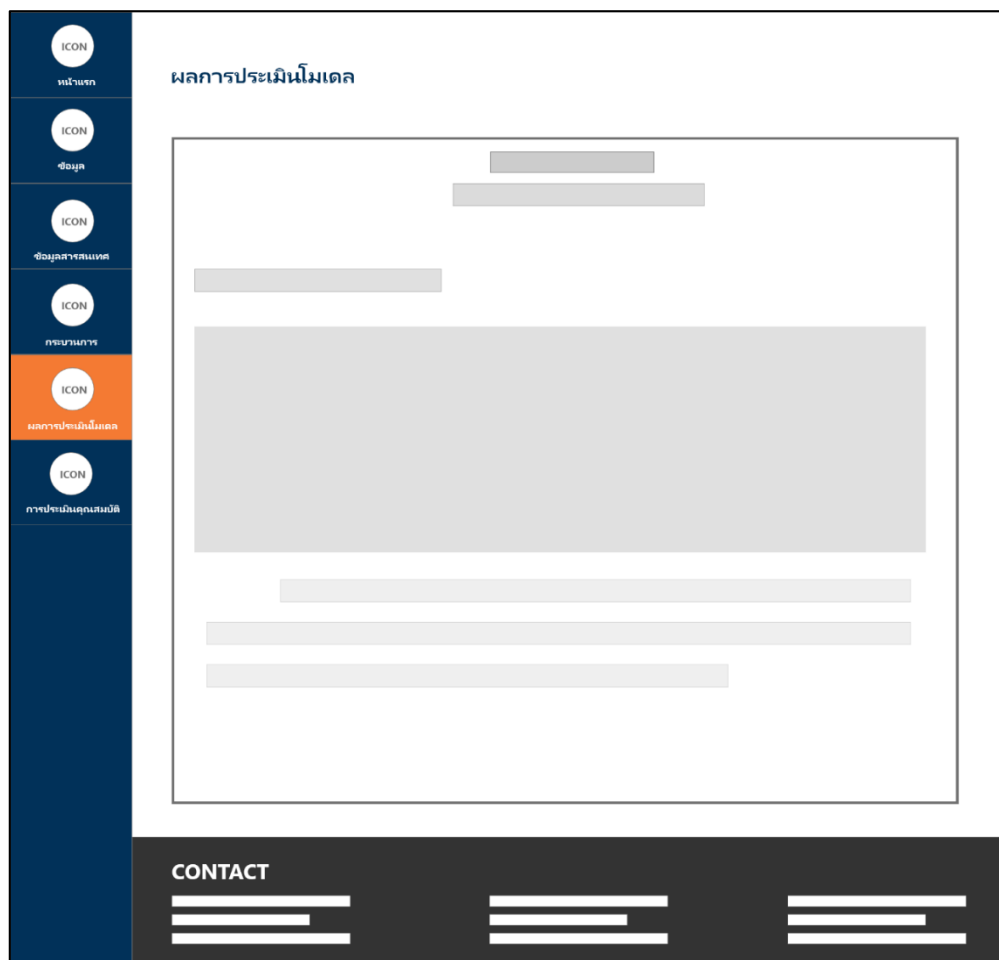
ภาพที่ 3.21 หน้าเว็บเพจที่ 3 ชุดข้อมูลสารสนเทศ

4) หน้าเว็บเพจที่ 4 Classification Model แสดงถึงขั้นตอนทำงาน



ภาพที่ 3.22 หน้าเว็บเพจที่ 4 Classification Model

6) หน้าเว็บเพจที่ 5 แสดงหน้าผลการประเมินโมเดล



ภาพที่ 3.23 หน้าเว็บเพจที่ 5 หน้าผลการประเมินโมเดล

6) หน้าเว็บเพจที่ 5 แสดงหน้าการประเมินคุณสมบัติ

ICON หน้าแรก

ICON ข้อมูล

ICON ข้อมูลสารสนเทศ

ICON กระบวนการ

ICON ผลการประเมินเบื้องต้น

ICON การประเมินคุณสมบัติ

การประเมินคุณสมบัติ

[Red Radio Button] [Grey Bar]

[Red Radio Button] [Grey Bar]

[Blue Radio Button] [Grey Bar]

[Blue Radio Button] [Grey Bar]

[Grey Bar]

[Blue Radio Button] [Grey Bar]

[Blue Radio Button] [Grey Bar]

[Grey Bar]

[Blue Radio Button] [Grey Bar]

[Blue Radio Button] [Grey Bar]

BUTTON

CONTACT

[White Lines]

[White Lines]

[White Lines]

ภาพที่ 3.24 หน้าเว็บเพจที่ 6 แสดงหน้าการประเมินคุณสมบัติ

3.3 บทสรุป

จากวิธีการดำเนินงานโครงการในข้างต้นทั้งหมดนี้ ผู้วิเคราะห์ข้อมูลการพัฒนาแบบจำลองมาตรฐานการช่วยเหลือด้านการศึกษา สำหรับนักศึกษา ด้วยเทคนิคต้นไม้ตัดสินใจ ด้วยขั้นตอนกระบวนการวิเคราะห์ข้อมูลด้วย CRISP-DM อย่างละเอียด รวมถึงการสร้างโมเดล Decision Tree จากโปรแกรมที่ใช้ทำเหมืองข้อมูล คือ โปรแกรม Rapid Miner Studio 9.5.1 และโปรแกรม Weka 3.8.4 ในการสร้างโมเดล Decision Tree ซึ่งได้ผลลัพธ์ที่เหมือนกัน และนำข้อมูลสารสนเทศมาทำการแสดงผลแบบ Visualization ในรูปแบบของภาพโดยใช้โปรแกรม Tableau Public และออกแบบ Wireframe ของเว็บไซต์ที่จะเผยแพร่บน Web Browser ด้วยโปรแกรม Adobe XD

บทที่ 4

ผลการดำเนินงาน

จากการวิเคราะห์และการพัฒนาแบบจำลองมาตรการการช่วยเหลือด้านการศึกษา สำหรับนักศึกษา ด้วยเทคนิคต้นไม้ตัดสินใจ มีวัตถุประสงค์ เพื่อพัฒนาแบบจำลองมาตรการการช่วยเหลือด้านการศึกษา สำหรับนักศึกษา ด้วยเทคนิคต้นไม้ตัดสินใจและเพื่อเสนอแนวทางในการตัดสินใจที่จะมอบทุนการศึกษาให้กับนักศึกษา โดยผู้จัดทำโครงการได้ดำเนินการตามที่วางแผนวิเคราะห์ข้อมูล และนำเทคโนโลยีสารสนเทศเข้ามาช่วยในการเสนอข้อมูล ทำให้มีความสะดวกและสามารถใช้งานได้จริงจนสำเร็จลุล่วงตามเป้าหมาย

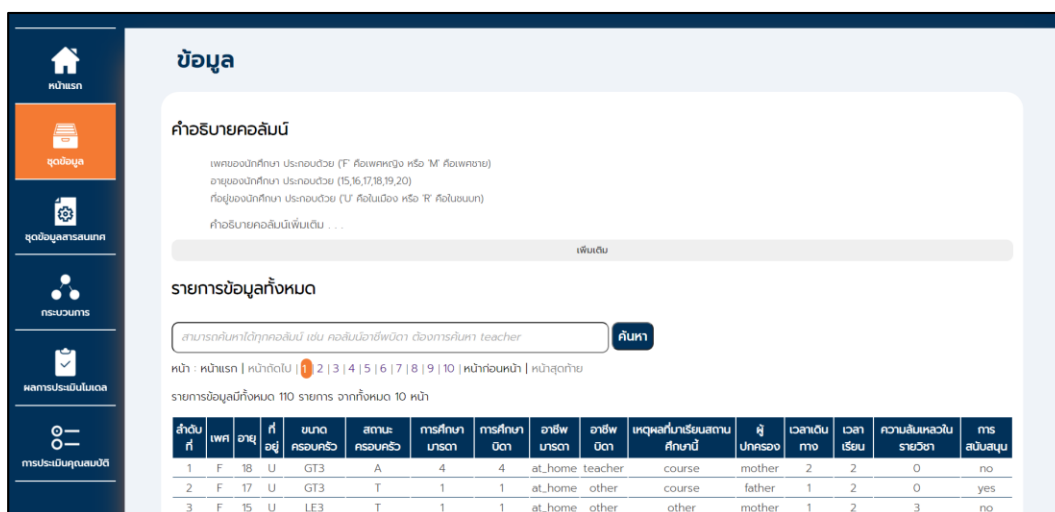
4.1 ผลการดำเนินงาน

การจัดทำโครงการเรื่อง การพัฒนาแบบจำลองมาตรการการช่วยเหลือด้านการศึกษา สำหรับนักศึกษา ด้วยเทคนิคต้นไม้ตัดสินใจ ในครั้งนี้พบว่าการจำแนกข้อมูลออกเป็นประเภทต่าง ๆ ช่วยให้เกิดความเข้าใจต่อข้อมูล และสิ่งที่ต้องการทราบอย่างแท้จริง คณะผู้จัดทำได้ทำการวิเคราะห์ข้อมูลอยู่ในรูปแบบขอบ Data Mining โดยทำการจำแนกข้อมูลออกเป็นประเภทต่าง ๆ ด้วยเทคนิคการทำเหมืองข้อมูลแบบ Classification จากเทคนิคการจำแนกประเภทข้อมูลด้วยการวิเคราะห์ข้อมูลแบบของแผนภูมิต้นไม้ หรือเรียกว่า Decision Tree และนำเสนอข้อมูลสารสนเทศแบบ Visualization เผยแพร่ข้อมูลสารสนเทศนี้บน Web Browser ให้กับผู้ใช้งาน ซึ่งทำให้เข้าใจได้อย่างสะดวกรวดเร็วยิ่งขึ้น รวมถึงมีความถูกต้องลดความซับซ้อนจากข้อมูล เพิ่มประสิทธิภาพให้กับการศึกษาค้นคว้าและทันต่อเวลา คณะผู้จัดทำจึงได้ พัฒนาแบบจำลองมาตรการการช่วยเหลือด้านการศึกษา สำหรับนักศึกษา ด้วยเทคนิคต้นไม้ตัดสินใจ ตามวัตถุประสงค์ที่ได้กำหนดไว้ และมีผลการดำเนินงาน ดังนี้



ภาพที่ 4.1 หน้าแรกของเว็บไซต์ เมื่อเข้าสู่เว็บไซต์

คำอธิบายการใช้งาน : แถบเมนูสามารถใช้งานได้ตามความต้องการของผู้ใช้งาน ซึ่งเป็นหน้าหลักของผู้ใช้งานทั่วไป สามารถเลือกดู และศึกษาค้นคว้าเข้าไปใช้งานในส่วนของเว็บไซต์ได้



ภาพที่ 4.2 เมนูแสดงชุดข้อมูลดิบที่นำมาใช้

คำอธิบายการใช้งาน : แถบเมนูชุดข้อมูล เป็นหน้าที่แสดงชุดข้อมูลที่นำมาใช้ สามารถค้นหาคำที่ต้องการให้แสดงได้

ข้อมูลสารสนเทศ

รายการข้อมูลสารสนเทศทั้งหมด

สามารถค้นหาได้ทุกคอลัมน์ เช่น คอลัมน์อายุเพศ ด้วยการพิมพ์ teacher ค้นหา

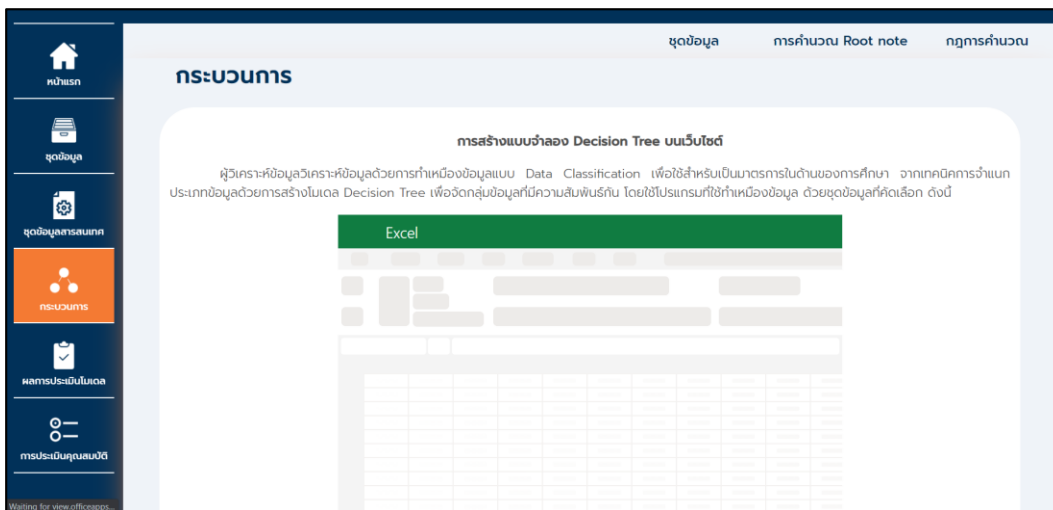
หน้า : หน้าแรก | หน้าถัดไป | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | หน้าก่อนหน้า | หน้าสุดท้าย

รายการข้อมูลสารสนเทศมีทั้งหมด 110 รายการ จากทั้งหมด 10 หน้า

ลำดับที่	เพศ	อายุ	ทศอยู่	ขนาดครอบครัว	สถานะครอบครัว	การศึกษา	การศึกษา	อาชีพ	อาชีพ	เหตุผลในการเรียน	ผู้ปกครอง	เวลา	เวลา	ความถี่	การสนับสนุน
1	หญิง	18-20	ในเมือง	3 คนขึ้นไป	แยกกัน	อื่นๆ	อื่นๆ	อยู่บ้าน	ครู/อาจารย์	หลักสูตร	มารดา	ปกติ	ปกติ	ไม่เคย	ไม่สนับสนุน
2	หญิง	15-17	ในเมือง	3 คนขึ้นไป	อยู่ด้วยกัน	ประถม	ประถม	อยู่บ้าน	อื่นๆ	หลักสูตร	บิดา	ปกติ	ปกติ	ไม่เคย	สนับสนุน
3	หญิง	15-17	ในเมือง	น้อยกว่า 3	อยู่ด้วยกัน	ประถม	ประถม	อยู่บ้าน	อื่นๆ	มารดา	ปกติ	ปกติ	ปกติ	ไม่เคย	ไม่สนับสนุน
4	หญิง	15-17	ในเมือง	3 คนขึ้นไป	อยู่ด้วยกัน	อื่นๆ	มัธยม	หอ/พยานา	พนักงานทั่วไป	ใกล้บ้าน	มารดา	ปกติ	ปกติ	ปกติ	ไม่สนับสนุน
5	หญิง	15-17	ในเมือง	3 คนขึ้นไป	อยู่ด้วยกัน	ปริญญา	ปริญญา	อื่นๆ	อื่นๆ	ใกล้บ้าน	บิดา	ปกติ	ปกติ	ไม่เคย	สนับสนุน
6	ชาย	15-17	ในเมือง	น้อยกว่า 3	อยู่ด้วยกัน	อื่นๆ	ปริญญา	พนักงานทั่วไป	อื่นๆ	ชื่อเสียง	มารดา	ปกติ	ปกติ	ไม่เคย	สนับสนุน
7	ชาย	15-17	ในเมือง	น้อยกว่า 3	อยู่ด้วยกัน	มัธยม	มัธยม	อื่นๆ	อื่นๆ	ใกล้บ้าน	มารดา	ปกติ	ปกติ	ไม่เคย	ไม่สนับสนุน
8	หญิง	15-17	ในเมือง	3 คนขึ้นไป	แยกกัน	อื่นๆ	อื่นๆ	อื่นๆ	ครู/อาจารย์	ใกล้บ้าน	มารดา	ปกติ	ปกติ	ไม่เคย	สนับสนุน
9	ชาย	15-17	ในเมือง	น้อยกว่า 3	แยกกัน	ปริญญา	มัธยม	พนักงานทั่วไป	อื่นๆ	ใกล้บ้าน	มารดา	ปกติ	ปกติ	ไม่เคย	สนับสนุน
10	ชาย	15-17	ในเมือง	3 คนขึ้นไป	อยู่ด้วยกัน	ปริญญา	อื่นๆ	อื่นๆ	อื่นๆ	ใกล้บ้าน	มารดา	ปกติ	ปกติ	ไม่เคย	สนับสนุน

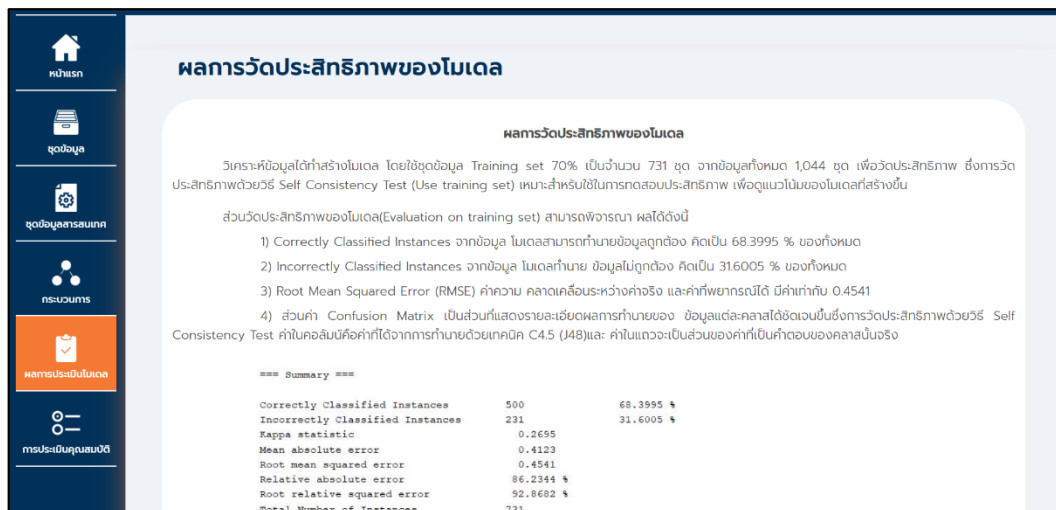
ภาพที่ 4.3 เมนูแสดงชุดข้อมูลสารสนเทศที่นำมาใช้

คำอธิบายการใช้งาน : แถบเมนูชุดข้อมูลสารสนเทศ เป็นหน้าที่แสดงชุดข้อมูลสารสนเทศที่นำมาใช้ สามารถค้นหาค่าที่ต้องการให้แสดงได้



ภาพที่ 4.4 เมนูแสดงกระบวนการวิเคราะห์ ด้วย Decision Tree

คำอธิบายการใช้งาน : เป็นหน้าที่อธิบายกระบวนการขั้นตอนในการสร้าง โมเดล Decision Tree ซึ่งจะประกอบไปด้วยการคำนวณต่าง ๆ และกฎที่ได้จากการสร้างโมเดล



ผลการวัดประสิทธิภาพของโมเดล

ผลการวัดประสิทธิภาพของโมเดล

วิเคราะห์ข้อมูลได้ทำสร้างโมเดล โดยใช้ชุดข้อมูล Training set 70% เป็นจำนวน 731 ชุด จากข้อมูลทั้งหมด 1,044 ชุด เพื่อวัดประสิทธิภาพ ซึ่งการวัดประสิทธิภาพด้วยวิธี Self Consistency Test (Use training set) เหมาะสำหรับการใช้ในการทดสอบประสิทธิภาพ เพื่อดูแนวโน้มของโมเดลที่สร้างขึ้น

ส่วนวัดประสิทธิภาพของโมเดล(Evaluation on training set) สามารถพิจารณา ผลได้ดังนี้

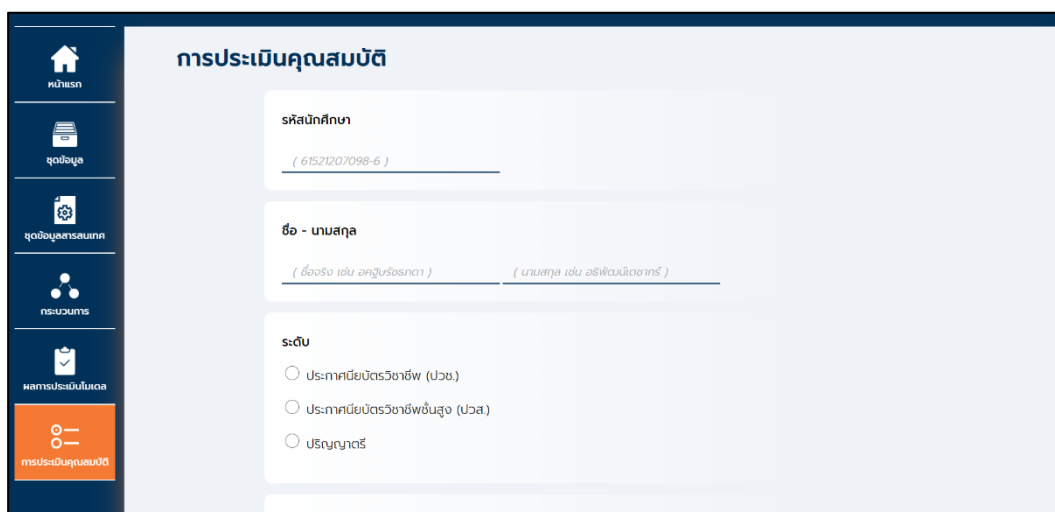
- 1) Correctly Classified Instances จากข้อมูล โมเดลสามารถทำนายข้อมูลถูกต้อง คิดเป็น 68.3995 % ของทั้งหมด
- 2) Incorrectly Classified Instances จากข้อมูล โมเดลทำนาย ข้อมูลไม่ถูกต้อง คิดเป็น 31.6005 % ของทั้งหมด
- 3) Root Mean Squared Error (RMSE) ค่าความ คลาดเคลื่อนระหว่างค่าจริง และค่าที่พยากรณ์ได้ มีค่าเท่ากับ 0.4541
- 4) ส่วนค่า Confusion Matrix เป็นส่วนที่แสดงรายละเอียดผลการทำนายของ ข้อมูลแต่ละคลาสที่ได้ชัดเจนขึ้นซึ่งการวัดประสิทธิภาพด้วยวิธี Self Consistency Test ค่าในคอลัมน์คือค่าที่ได้จากการทำนายด้วยเทคนิค C4.5 (J48)และ ค่าในแถวเป็นส่วนของค่าที่เป็นค่าของของคลาสนั้นจริง

=== Summary ===

Correctly Classified Instances	500	68.3995 %
Incorrectly Classified Instances	231	31.6005 %
Kappa statistic	0.2695	
Mean absolute error	0.4123	
Root mean squared error	0.4541	
Relative absolute error	86.2344 %	
Root relative squared error	92.8682 %	
Total Number of Instances	731	

ภาพที่ 4.5 เมนูผลการวัดประสิทธิภาพของโมเดล

คำอธิบายการใช้งาน : เป็นหน้าที่แสดงผลการวัดประสิทธิภาพต่าง ๆ ของโมเดล Decision Tree โดยแบ่งชุดข้อมูลเพื่อใช้ทดสอบสร้าง 70% และชุดข้อมูลทดสอบ 30%



การประเมินคุณสมบัติ

รศ. นิสิตศึกษา
(61521207098-6)

ชื่อ - นามสกุล
(ชื่อจริง เช่น อรุณโรสภษา) (นามสกุล เช่น อธิพัฒน์ดิษชาทร์)

ระดับ

ประกาศนียบัตรวิชาชีพ (ปวช.)

ประกาศนียบัตรวิชาชีพชั้นสูง (ปวส.)

ปริญญาตรี

ภาพที่ 4.6 เมนูการประเมินคุณสมบัติ

คำอธิบายการใช้งาน : เป็นหน้าที่ผู้เยี่ยมชมเว็บไซต์ สามารถกรอกข้อมูลต่าง ๆ ตามหัวข้อที่กำหนด และสามารถประเมินผลการมอบทุนการศึกษา รวมถึงแสดงข้อมูลของผู้ใช้งาน

ภาพที่ 4.7 หน้าเข้าสู่ระบบเฉพาะผู้ดูแลระบบ

คำอธิบายการใช้งาน : เป็นหน้าที่ผู้ดูแลระบบสามารถกรอกชื่อผู้ใช้และรหัสผ่าน เข้าสู่ระบบ เพื่อดูข้อมูลต่าง ๆ จากข้อมูลผู้ใช้งาน

ภาพที่ 4.8 หน้าข้อมูลส่วนตัวของผู้จัดทำ

คำอธิบายการใช้งาน : เป็นหน้าที่แสดงประวัติของผู้จัดทำเว็บไซต์ และข้อมูลที่ต้องการติดต่อ รวมถึงแสดงหน่วยงานที่เกี่ยวข้องกับโครงการครั้งนี้

4.2 การอภิปรายผล

จากวัตถุประสงค์ เพื่อการพัฒนาแบบจำลองมาตรฐานการการช่วยเหลือด้านการศึกษา สำหรับนักศึกษา ด้วยเทคนิคต้นไม้ตัดสินใจ คณะผู้จัดทำได้ศึกษาปัญหา และเก็บรวบรวมข้อมูล โดยใช้กระบวนการวิเคราะห์ข้อมูล CRISP-DM(Cross Industry Standard Process For Data Mining) เพื่อใช้ในการวิเคราะห์ตามขอบเขตของโครงการ ดังนี้

4.2.1 ขอบเขตของข้อมูล

1) ข้อมูลที่นำมาวิเคราะห์ เป็นข้อมูลที่ได้จากเว็บไซต์ Kaggle.com จำนวน 1,044 ชุด

2) แบ่งข้อมูลที่นำมาวิเคราะห์ออกเป็น 2 กลุ่ม คือ กลุ่มข้อมูลสำหรับการเรียนรู้ (Training Data) 70% เป็นจำนวน 731 ชุด และกลุ่มข้อมูลสำหรับการทดสอบ (Testing Data) 30% จำนวน 313 ชุด

4.2.2 ขอบเขตการวิเคราะห์ข้อมูล

1) รู้จักและเข้าใจในธุรกิจ (Business Understanding) เป็นขั้นตอนแรกของกระบวนการ ที่มุ่งเน้นไปที่การทำความเข้าใจกระบวนการทางธุรกิจโดยรวม ผู้วิเคราะห์ข้อมูลทำความเข้าใจกับปัญหาให้อยู่ในรูปของการวิเคราะห์ข้อมูลทาง Data Mining โดยการวิเคราะห์ข้อมูลในประเด็นนี้ ซึ่งมีข้อมูลมาก ทำให้ไม่สามารถทำความเข้าใจกับข้อมูลที่มีความซับซ้อนได้อย่างรวดเร็ว เช่น ต้องการทราบว่านักศึกษาลักษณะคุณสมบัติใดที่บ่งบอกถึงความต้องการช่วยเหลือในด้านการศึกษา

2) จัดเก็บและรวบรวมข้อมูลให้ครบ (Data Understanding) ขั้นตอนการจัดเก็บและรวบรวมข้อมูล ตลอดจนการพิจารณาตรวจสอบความถูกต้องของข้อมูลที่ได้รับ โดยเลือกที่จะใช้ข้อมูลทั้งหมดหรือบางส่วนในการวิเคราะห์ให้สอดคล้องกับวัตถุประสงค์ที่กำหนดไว้ ผู้วิเคราะห์ข้อมูลทำการรวบรวมข้อมูล เพื่อตรวจสอบรายละเอียด ปริมาณ และความน่าเชื่อถือของข้อมูลนักศึกษาเบื้องต้น หรือสถานะครอบครัว ที่ได้จากเว็บไซต์ Kaggle.com

3) เตรียมข้อมูลให้พร้อมใช้งาน (Data Preparation) ขั้นตอนการแปลงข้อมูลที่ได้รวบรวมมาและเลือกไว้ ให้อยู่ในรูปแบบที่พร้อมสำหรับนำไปวิเคราะห์ในขั้นตอนต่อไปได้ โดยการทำให้เป็นข้อมูลที่ถูกต้อง (Data cleaning)

4) สร้างแบบจำลอง (Modeling) ขั้นตอนการสร้างตัวแบบทางคณิตศาสตร์ และสถิติเพื่อการวิเคราะห์ข้อมูล โดยสามารถใช้เทคนิควิธีการต่าง ๆ อาทิ การจำแนก

(Classification) การแบ่งกลุ่ม (Clustering) การสร้างความสัมพันธ์ (Association rule) และการเลือกแอตทริบิวต์หรือคอลัมน์ที่จะนำไปทำการวิเคราะห์ โดยการเลือกแอตทริบิวต์หรือคอลัมน์ (Future Selection) ในวิธี Information Gain โดยได้ 5 คอลัมน์ ได้แก่ อายุ การศึกษามารดา การศึกษาบิดา อาชีพมารดา อาชีพบิดา และผลการมอบทุน

5) การประเมินผล (Evaluation) เป็นขั้นตอนก่อนนำผลลัพธ์ที่ได้จากขั้นตอนก่อนหน้าโดยนำไปใช้งาน ด้วยการวัดประสิทธิภาพของผลลัพธ์ที่ได้กับวัตถุประสงค์ที่ตั้งไว้ในขั้นตอนแรกเพื่อแสดงให้เห็นว่ามีนัยสำคัญหรือความน่าเชื่อถือมากน้อยเพียงใด โดยการประเมินผลจากโปรแกรม Weka มีความถูกต้องถึง 68.3995% และได้กฎจำนวน 40 กฎ ดังนี้

กฎข้อที่ 1 IF age=a then YES หมายความว่า ถ้า อายุเท่ากับ 15-17 จะมีผลลัพธ์เป็น “สนับสนุน”

กฎข้อที่ 2 IF age=b and mjob=services and medu=py and fedu=my then NO หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นพนักงานทั่วไป และการศึกษามารดาเป็นปริญญา และการศึกษาบิดาเป็นมัธยม จะมีผลลัพธ์เป็น “ไม่สนับสนุน”

กฎข้อที่ 3 IF age=b and mjob=services and medu=py and fedu=py then YES หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นพนักงานทั่วไป และการศึกษามารดาเป็นปริญญา และการศึกษาบิดาเป็นปริญญา จะมีผลลัพธ์เป็น “สนับสนุน”

กฎข้อที่ 4 IF age=b and mjob=services and medu=py and fedu=ot then NO หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นพนักงานทั่วไป และการศึกษามารดาเป็นปริญญา และการศึกษาบิดาเป็นอื่น ๆ จะมีผลลัพธ์เป็น “ไม่สนับสนุน”

กฎข้อที่ 5 IF age=b and mjob=services and medu=py and fedu=pt then NO หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นพนักงานทั่วไป และการศึกษามารดาเป็นปริญญา และการศึกษาบิดาเป็นประถม จะมีผลลัพธ์เป็น “ไม่สนับสนุน”

กฎข้อที่ 6 IF age=b and mjob=services and medu=ot then YES หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นพนักงานทั่วไป และการศึกษามารดาเป็นอื่น ๆ จะมีผลลัพธ์เป็น “สนับสนุน”

กฎข้อที่ 7 IF age=b and mjob=services and medu=my and fedu=my then YES หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นพนักงานทั่วไป และการศึกษามารดาเป็นมัธยม และการศึกษาบิดาเป็นมัธยม จะมีผลลัพธ์เป็น “สนับสนุน”

กฎข้อที่ 8 IF age=b and mjob=services and medu=my and fedu=py then NO
หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นพนักงานทั่วไป และการศึกษามารดาเป็น
มัธยม และการศึกษาบิดาเป็นปริญญา จะมีผลลัพธ์เป็น “สนับสนุน”

กฎข้อที่ 9 IF age=b and mjob=services and medu=my and fedu=ot then YES
หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นพนักงานทั่วไป และการศึกษามารดาเป็น
มัธยม และการศึกษาบิดาเป็นอื่น ๆ จะมีผลลัพธ์เป็น “สนับสนุน”

กฎข้อที่ 10 IF age=b and mjob=services and medu=my and fedu=pt and fjob=other
then YES หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นพนักงานทั่วไป และการศึกษา
มารดาเป็นมัธยม และการศึกษาบิดาเป็นประถม และอาชีพบิดาเป็นอื่น ๆ จะมีผลลัพธ์เป็น
“สนับสนุน”

กฎข้อที่ 11 IF age=b and mjob=services and medu=my and fedu=pt and fjob=services
then NO หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นพนักงานทั่วไป และการศึกษา
มารดาเป็นมัธยม และการศึกษาบิดาเป็นประถม และอาชีพบิดาเป็นพนักงานทั่วไป จะมีผลลัพธ์เป็น
“ไม่สนับสนุน”

กฎข้อที่ 12 IF age=b and mjob=services and medu=my and fedu=pt and fjob=teacher
then NO หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นพนักงานทั่วไป และการศึกษา
มารดาเป็นมัธยม และการศึกษาบิดาเป็นประถม และอาชีพบิดาเป็นครู/อาจารย์ จะมีผลลัพธ์เป็น
“ไม่สนับสนุน”

กฎข้อที่ 13 IF age=b and mjob=services and medu=my and fedu=pt and fjob=health
then NO หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นพนักงานทั่วไป และการศึกษา
มารดาเป็นมัธยม และการศึกษาบิดาเป็นประถม และอาชีพบิดาเป็นหมอ/พยาบาล จะมีผลลัพธ์เป็น
“ไม่สนับสนุน”

กฎข้อที่ 14 IF age=b and mjob=services and medu=my and fedu=pt and fjob=home
then NO หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นพนักงานทั่วไป และการศึกษา
มารดาเป็นมัธยม และการศึกษาบิดาเป็นประถม และอาชีพบิดาเป็นอยู่บ้าน จะมีผลลัพธ์เป็น
“ไม่สนับสนุน”

กฎข้อที่ 15 IF age=b and mjob=services and medu=pt then NO หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นพนักงานทั่วไป และการศึกษามารดาเป็นประถม จะมีผลลัพธ์เป็น “ไม่สนับสนุน”

กฎข้อที่ 16 IF age=b and mjob=other and fedu=my then NO หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นอื่น ๆ และการศึกษาบิดาเป็นมัธยม จะมีผลลัพธ์เป็น “ไม่สนับสนุน”

กฎข้อที่ 17 IF age=b and mjob=other and fedu=py then YES หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นอื่น ๆ และการศึกษาบิดาเป็นปริญญา จะมีผลลัพธ์เป็น “ไม่สนับสนุน”

กฎข้อที่ 18 IF age=b and mjob=other and fedu=ot then YES หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นอื่น ๆ และการศึกษาบิดาเป็นอื่น ๆ จะมีผลลัพธ์เป็น “ไม่สนับสนุน”

กฎข้อที่ 19 IF age=b and mjob=other and fedu=pt then NO หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นอื่น ๆ และการศึกษาบิดาเป็นประถม จะมีผลลัพธ์เป็น “ไม่สนับสนุน”

กฎข้อที่ 20 IF age=b and mjob=home and fjob=other and fedu=my then NO หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นอยู่บ้าน และอาชีพบิดาเป็นอื่น ๆ และการศึกษาบิดาเป็นมัธยม จะมีผลลัพธ์เป็น “ไม่สนับสนุน”

กฎข้อที่ 21 IF age=b and mjob=home and fjob=other and fedu=py then YES หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นอยู่บ้าน และอาชีพบิดาเป็นอื่น ๆ และการศึกษาบิดาเป็นปริญญา จะมีผลลัพธ์เป็น “สนับสนุน”

กฎข้อที่ 22 IF age=b and mjob=home and fjob=other and fedu=ot then NO หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นอยู่บ้าน และอาชีพบิดาเป็นอื่น ๆ และการศึกษาบิดาเป็นอื่น ๆ จะมีผลลัพธ์เป็น “ไม่สนับสนุน”

กฎข้อที่ 23 IF age=b and mjob=home and fjob=other and fedu=pt then YES หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นอยู่บ้าน และอาชีพบิดาเป็นอื่น ๆ และการศึกษาบิดาเป็นประถม จะมีผลลัพธ์เป็น “สนับสนุน”

กฎข้อที่ 24 IF age=b and mjob=home and fjob=services and fedu=my then YES
 หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นอยู่บ้าน และอาชีพบิดาเป็นพนักงาน
 ทั่วไป และการศึกษาบิดาเป็นมัธยม จะมีผลลัพธ์เป็น “สนับสนุน”

กฎข้อที่ 25 IF age=b and mjob=home and fjob=services and fedu=py then NO
 หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นอยู่บ้าน และอาชีพบิดาเป็นพนักงาน
 ทั่วไป และการศึกษาบิดาเป็นปริญญา จะมีผลลัพธ์เป็น “ไม่สนับสนุน”

กฎข้อที่ 26 IF age=b and mjob=home and fjob=services and fedu=ot then NO
 หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นอยู่บ้าน และอาชีพบิดาเป็นพนักงาน
 ทั่วไป และการศึกษาบิดาเป็นอื่น ๆ จะมีผลลัพธ์เป็น “ไม่สนับสนุน”

กฎข้อที่ 27 IF age=b and mjob=home and fjob=services and fedu=pt then NO
 หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นอยู่บ้าน และอาชีพบิดาเป็นพนักงาน
 ทั่วไป และการศึกษาบิดาเป็นประถม จะมีผลลัพธ์เป็น “ไม่สนับสนุน”

กฎข้อที่ 28 IF age=b and mjob=home and fjob=teacher then NO หมายความว่า ถ้าอายุ
 เท่ากับ 18-20 และอาชีพมารดาเป็นอยู่บ้าน และอาชีพบิดาเป็นครู/อาจารย์ จะมีผลลัพธ์เป็น “ไม่
 สนับสนุน”

กฎข้อที่ 29 IF age=b and mjob=home and fjob=health then NO หมายความว่า ถ้าอายุ
 เท่ากับ 18-20 และอาชีพมารดาเป็นอยู่บ้าน และอาชีพบิดาเป็นหมอ/พยาบาล จะมีผลลัพธ์เป็น
 “ไม่สนับสนุน”

กฎข้อที่ 30 IF age=b and mjob=home and fjob=home then YES หมายความว่า ถ้าอายุ
 เท่ากับ 18-20 และอาชีพมารดาเป็นอยู่บ้าน และอาชีพบิดาเป็นอยู่บ้านจะมีผลลัพธ์เป็น “สนับสนุน”

กฎข้อที่ 31 IF age=b and mjob=health then YES หมายความว่า ถ้าอายุเท่ากับ 18-20
 และอาชีพมารดาเป็นหมอ/พยาบาล จะมีผลลัพธ์เป็น “สนับสนุน”

กฎข้อที่ 32 IF age=b and mjob=teacher and fjob=other then YES หมายความว่า ถ้าอายุ
 เท่ากับ 18-20 และอาชีพมารดาเป็นครู/อาจารย์ และอาชีพบิดาเป็นอื่น ๆ จะมีผลลัพธ์เป็น
 “สนับสนุน”

กฎข้อที่ 33 IF age=b and mjob=teacher and fjob=services and fedu=my then NO
 หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นครู/อาจารย์ และอาชีพบิดาเป็นพนักงาน
 ทั่วไป และการศึกษาบิดาเป็นมัธยม จะมีผลลัพธ์เป็น “ไม่สนับสนุน”

กฎข้อที่ 34 IF age=b and mjob=teacher and fjob=services and fedu=py then NO
 หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นครู/อาจารย์ และอาชีพบิดาเป็นพนักงาน
 ทั่วไป และการศึกษาบิดาเป็นปริญญา จะมีผลลัพธ์เป็น “ไม่สนับสนุน”

กฎข้อที่ 35 IF age=b and mjob=teacher and fjob=services and fedu=ot then NO
 หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นครู/อาจารย์ และอาชีพบิดาเป็นพนักงาน
 ทั่วไป และการศึกษาบิดาเป็นอื่น ๆ จะมีผลลัพธ์เป็น “ไม่สนับสนุน”

กฎข้อที่ 36 IF age=b and mjob=teacher and fjob=services and fedu=pt then YES
 หมายความว่า ถ้าอายุเท่ากับ 18-20 และอาชีพมารดาเป็นครู/อาจารย์ และอาชีพบิดาเป็นพนักงาน
 ทั่วไป และการศึกษาบิดาเป็นประถม จะมีผลลัพธ์เป็น “สนับสนุน”

กฎข้อที่ 37 IF age=b and mjob=teacher and fjob=teacher then YES หมายความว่า ถ้า
 อายุเท่ากับ 18-20 และอาชีพมารดาเป็นครู/อาจารย์ และอาชีพบิดาเป็นครู/อาจารย์ จะมีผลลัพธ์
 เป็น “สนับสนุน”

กฎข้อที่ 38 IF age=b and mjob=teacher and fjob=health then YES หมายความว่า ถ้า
 อายุเท่ากับ 18-20 และอาชีพมารดาเป็นครู/อาจารย์ และอาชีพบิดาเป็นหมอ/พยาบาล จะมีผลลัพธ์
 เป็น “สนับสนุน”

กฎข้อที่ 39 IF age=b and mjob=teacher and fjob=home then YES หมายความว่า ถ้าอายุ
 เท่ากับ 18-20 และอาชีพมารดาเป็นครู/อาจารย์ และอาชีพบิดาเป็นอยู่บ้าน จะมีผลลัพธ์เป็น
 “สนับสนุน”

กฎข้อที่ 40 IF age=c then YES หมายความว่า ถ้าอายุเท่ากับ 21 ขึ้นไป จะมีผลลัพธ์เป็น
 “สนับสนุน”

6) เผยแพร่ผลวิเคราะห์ (Deployment) ขั้นตอนการนำผลลัพธ์ที่ได้ไปใช้งานเป็นการทั่วไป อาจจัดทำเป็นรูปแบบของรายงาน (Report) หรือแผนภาพ (Dashboard) ที่พร้อมให้ฝ่ายต่าง ๆ นำไปใช้ประโยชน์ในการวางแผน กำหนดกลยุทธ์ และดำเนินการต่าง ๆ ในทางธุรกิจ ผู้วิเคราะห์ข้อมูลนำผลข้อมูลที่ทำการวิเคราะห์แล้ว แสดงผลข้อมูลบน Web Browser โดยใช้ชุดคำสั่ง HTML CSS3 และ PHP ร่วมกับการนำเสนอข้อมูลแบบ Visualization ด้วยการแสดงผลข้อมูลในรูปแบบของภาพโดยใช้โปรแกรม Tableau Public

4.2.3 ขอบเขตการปรับปรุงและการนำเสนอข้อมูล

- 1) นำเสนอชุดข้อมูล บนเว็บไซต์
- 2) นำเสนอรูปแบบโมเดล Decision Tree บนเว็บไซต์
- 3 นำเสนอข้อมูลสารสนเทศจากข้อมูลดิบ โดยใช้เครื่องมือ Data Visualization เพื่อนำเสนอเป็นแผนภาพแทนข้อมูล
- 4) ประยุกต์ใช้แบบจำลองมาตรการการช่วยเหลือด้านการศึกษา สำหรับนักศึกษาด้วยเทคนิคต้นไม้ตัดสินใจ บนเว็บไซต์

4.3 บทสรุป

จากการดำเนินงานโครงการในข้างต้นทั้งหมดนี้ คณะผู้จัดทำได้นำข้อมูลจากการวิเคราะห์ข้อมูล และการแสดงผลแบบ Visualization ต่าง ๆ โดยใช้โปรแกรม Tableau Public มาแสดงให้กับบุคคลภายนอกหรือกลุ่มผู้ใช้ข้อมูลได้รับรู้ในรูปแบบของเว็บไซต์ ที่จะเป็นแหล่งความรู้ต่าง ๆ ในเรื่องของแบบจำลองมาตรการการช่วยเหลือด้านการศึกษา สำหรับนักศึกษา การสร้างโมเดล Decision Tree และยังมีแบบสอบถามหรือโปรแกรมประมวลผลผลลัพธ์ตรงกับโมเดลและกฎที่ได้จากโมเดล เพื่อสามารถนำไปใช้เป็นแนวทางในการตัดสินใจที่จะมอบทุนการศึกษาให้กับนักศึกษา หรือสามารถนำข้อมูลไปวิเคราะห์และไปต่อยอดได้

บทที่ 5

สรุปและข้อเสนอแนะ

ผลการพัฒนาแบบจำลองมาตรการการช่วยเหลือด้านการศึกษา สำหรับนักศึกษา ด้วยเทคนิคต้นไม้ตัดสินใจ ทางคณะผู้จัดทำ ได้ทำการประเมินผลการใช้งานเว็บไซต์ จากผู้ใช้งานทั่วไป เพื่อสรุปผลการทำโครงการ ข้อจำกัดของเว็บไซต์ ปัญหาและอุปสรรคต่าง ๆ ของการทำโครงการ และข้อเสนอแนะเพื่อพัฒนาโครงการ ดังนี้

5.1 บทสรุปผลโครงการ

จากที่คณะผู้จัดทำได้วิเคราะห์ข้อมูลและจัดทำเว็บไซต์ จึงได้ทำการประเมินผลการใช้งานเว็บไซต์จากกลุ่มตัวอย่างจำนวน 50 คน โดยการทำแบบสอบถามความพึงพอใจในการใช้งานเว็บไซต์ แบ่งเป็นเพศหญิงจำนวน 31 คน คิดเป็นร้อยละ 62% เพศชายจำนวน 19 คน คิดเป็นร้อยละ 38% อายุต่ำกว่า 18 ปี จำนวน 5 คน คิดเป็นร้อยละ 10% อายุตั้งแต่ 18-24 ปี จำนวน 37 คน คิดเป็นร้อยละ 74% อายุ 25 ปีขึ้นไป จำนวน 8 คน คิดเป็นร้อยละ 16% โดยเกณฑ์การให้คะแนนของผู้ใช้งานเว็บไซต์มีดังนี้ ระดับความพึงพอใจของผู้ใช้งานเว็บไซต์ ระดับความพึงพอใจ 5 = มากที่สุด 4 = มาก 3 = ปานกลาง 2 = น้อย 1 = น้อยที่สุด

เมื่อนำคำตอบของผู้ตอบแบบสอบถามมาแจกแจงความถี่และหาค่าเฉลี่ย แล้วกำหนดระดับค่าเฉลี่ย แบ่งระดับความพึงพอใจได้ดังนี้

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

จากเกณฑ์ดังกล่าว สามารถแปลความหมายของความพึงพอใจได้ ดังนี้

คะแนนเฉลี่ย 1.00-1.80 หมายถึง ระดับความพึงพอใจน้อยที่สุด

คะแนนเฉลี่ย 1.81-2.60 หมายถึง ระดับความพึงพอใจน้อย

คะแนนเฉลี่ย 2.61-3.40 หมายถึง ระดับความพึงพอใจปานกลาง

คะแนนเฉลี่ย 3.41-4.20 หมายถึง ระดับความพึงพอใจมาก

คะแนนเฉลี่ย 4.21-5.00 หมายถึง ระดับความพึงพอใจมากที่สุด

(อ้างอิงจาก: สรุปแบบประเมินความพึงพอใจของผู้รับบริการศูนย์ดำรง

ธรรมจังหวัดราชบุรี)

เนื้อหาที่ได้ทำการประเมินและคะแนนที่ได้มีดังนี้

- 1) แหล่งความรู้ที่อ้างอิงมีความชัดเจนและสอดคล้องกับงาน ได้คะแนนค่าเฉลี่ย 4.42 คะแนนอยู่ในระดับพึงพอใจมากที่สุด
- 2) ข้อมูลที่น่าเสนอบนเว็บไซต์มีขนาดและองค์ประกอบที่เหมาะสม ได้คะแนนค่าเฉลี่ย 4.58 คะแนนอยู่ในระดับพึงพอใจมากที่สุด
- 3) การจัดลำดับเนื้อหาเป็นขั้นตอน มีความต่อเนื่อง อ่านแล้วเข้าใจ ได้คะแนนค่าเฉลี่ย 4.48 คะแนนอยู่ในระดับพึงพอใจมากที่สุด
- 4) รูปแบบ Visualization มีความเหมาะสมกับข้อมูล ได้คะแนนค่าเฉลี่ย 4.31 คะแนนอยู่ในระดับพึงพอใจมากที่สุด
- 5) การวิเคราะห์มีความชัดเจน น่าเชื่อถือ ได้คะแนนค่าเฉลี่ย 4.68 คะแนนอยู่ในระดับพึงพอใจมากที่สุด
- 6) การจัดรูปแบบในเว็บไซต์ง่ายต่อการอ่านและสะดวกต่อการใช้งาน ได้คะแนนค่าเฉลี่ย 4.46 คะแนนอยู่ในระดับพึงพอใจมากที่สุด
- 7) เว็บไซต์มีความสวยงาม มีความทันสมัย น่าสนใจ ได้คะแนนค่าเฉลี่ย 4.32 คะแนนอยู่ในระดับพึงพอใจมากที่สุด
- 8) สีสีนในการออกแบบเว็บไซต์มีความเหมาะสม ได้คะแนนค่าเฉลี่ย 4.5 คะแนนอยู่ในระดับพึงพอใจมากที่สุด
- 9) สีพื้นหลังกับสีตัวอักษรมีความเหมาะสมต่อการอ่าน ได้คะแนนค่าเฉลี่ย 4.6 คะแนนอยู่ในระดับพึงพอใจมากที่สุด
- 10) ขนาดตัวอักษร และรูปแบบตัวอักษร มีความสวยงามและอ่านได้ง่าย ได้คะแนนค่าเฉลี่ย 4.56 คะแนน อยู่ในระดับพึง พพอใจมากที่สุด
- 11) เนื้อหามีรายละเอียดและประโยชน์เพียงพอสามารถนำไปประยุกต์ใช้ได้ ได้คะแนนค่าเฉลี่ย 4.34 คะแนน อยู่ในระดับพึง พพอใจมากที่สุด
- 12) สามารถนำไปใช้เป็นแหล่งอ้างอิงให้กับโครงการเรื่องอื่นที่เกี่ยวข้องได้ ได้คะแนนค่าเฉลี่ย 4.54 คะแนน อยู่ในระดับพึง พพอใจมากที่สุด
- 13) สามารถใช้เป็นแหล่งความรู้และแนวทางในการพัฒนาต่อไปได้ ได้คะแนนค่าเฉลี่ย 4.38 คะแนน อยู่ในระดับพึง พพอใจมากที่สุด

14) แหล่งข้อมูลที่ตรงกับความต้องการของผู้ใช้งาน ได้คะแนนค่าเฉลี่ย 4.6 คะแนน อยู่ในระดับพึงพอใจมากที่สุด

จากคะแนนในแต่ละด้านที่กลุ่มผู้ใช้งานเว็บไซต์ ได้ประเมินผ่านแบบสอบถามความพึงพอใจในการใช้งานเว็บไซต์ ได้คะแนนเฉลี่ย รวม 4.50 คะแนน ซึ่งอยู่ในระดับพึงพอใจมากที่สุด

5.2 ข้อจำกัดของระบบ

5.2.1 ผู้ใช้ไม่สามารถอัปโหลดชุดข้อมูลเข้าไปในเว็บไซต์ที่พัฒนาขึ้นได้ เนื่องจากเป็นเว็บไซต์เผยแพร่ข้อมูลและช่วยประเมินคุณสมบัติเท่านั้น

5.2.2 ผู้ดูแลระบบไม่สามารถ ลบ แก้ไขข้อมูลของผู้ที่กรอกข้อมูลได้

5.2.3 ผู้ดูแลระบบไม่สามารถอัปเดตข้อมูลแบบเรียลไทม์ได้

5.3 ปัญหาและอุปสรรคของโครงการ

5.3.1 ข้อจำกัดของข้อมูลที่ไม่ได้ทำการจัดเก็บรวบรวมด้วยตนเอง และเนื่องจากเป็นชุดข้อมูลแบบเปิดทำให้ต้องใช้ระยะเวลาในการศึกษา

5.3.2 ข้อจำกัดของโปรแกรมต่าง ๆ ที่นำมาใช้ในการพัฒนาแบบจำลอง เนื่องจากผู้วิเคราะห์ที่ใช้โปรแกรมแบบเวอร์ชันทดลองทำให้โปรแกรมมีข้อจำกัดในการทำงาน เช่น โปรแกรม Weka 3.8.4 ต้องซื้อแพ็คเกจภาษาไทยมาติดตั้งจึงจะสามารถนำเข้าชุดข้อมูลที่เป็นภาษาไทยได้

5.3.3 เนื่องด้วยในการจัดทำโครงการในครั้งนี้ อยู่ในช่วงเหตุการณ์โรค COVID19 ระบาด ทำให้ยากต่อการติดต่อสื่อสาร การประสานงานต่าง ๆ ทำให้เกิดความล่าช้าในการดำเนินงาน และการสื่อสารกันผิดพลาด

5.4 ข้อเสนอแนะ

การพัฒนาแบบจำลองมาตรการการช่วยเหลือด้านการศึกษา สำหรับนักศึกษา ด้วยเทคนิคต้นไม้ตัดสินใจ ที่จัดทำเสร็จสิ้นแล้วนี้แม้จะสามารถทำงานและแสดงผลข้อมูลได้ถูกต้องตามวัตถุประสงค์และขอบเขตที่ตั้งไว้ แต่ยังมีข้อจำกัด ซึ่งหากจะพัฒนาให้เว็บไซต์ทำงานได้อย่างมีประสิทธิภาพมากขึ้น ผู้พัฒนาควรต้องปรับปรุงงานในส่วนต่าง ๆ ดังต่อไปนี้

5.4.1 ควรพัฒนาระบบให้สามารถอัปโหลดข้อมูลเข้าไปในเว็บไซต์ได้

5.4.2 ควรพัฒนาให้ผู้ดูแลระบบสามารถ ลบ แก้ไขข้อมูลได้

5.4.3 ควรพัฒนาระบบให้ผู้ดูแลระบบสามารถอัปเดตข้อมูลแบบเรียลไทม์ได้

บรรณานุกรม

- จริยา ศรีวิจารณ์, อรัญญา ต้อยคัมภีร์และณัฐสุดา. (2556). *ประสบการณ์ด้านจิตใจของนักศึกษาที่ ยากจนจากชนบท และได้รับทุนการศึกษา*. สืบค้น 18 พฤษภาคม 2563, จาก <https://so04.tci-thaijo.org/index.php/gskkuhs/article/view/30405>
- ธนากร ปามุทา. (2557). *การวิเคราะห์พฤติกรรมการณ์มีวัสดุสารนิเทศของห้องสมุดด้วยเทคนิค การทำเหมืองข้อมูลกรณีศึกษาของสมุดมหาวิทยาลัยราชภัฏยะลา*. สืบค้น 18 พฤษภาคม 2563, จาก <https://ph01.tci-thaijo.org/index.php/rmutijo/article/view/21041/18243>
- ธนพัฒน์ ทองมา และสุชาดา บวรกิติวงศ์. (2560). *การพัฒนาโมเดลทำนายแผนการเรียนใน การศึกษาต่อระดับมัธยมศึกษาตอนปลาย*. สืบค้น 18 พฤษภาคม 2563, จาก <https://so01.tci-thaijo.org/index.php/OJED/article/view/110673/86734>
- นิรมล พันสีมา และอนันต์ เจ่าสกุล(2557) *การเปรียบเทียบการทำงานโปรแกรม R และโปรแกรม SPSS กรณีการจำแนกประเภทข้อมูลเงินยืมที่ตรงจ่ายของมหาวิทยาลัยขอนแก่น*. สืบค้น 18 พฤษภาคม 2563, จาก https://li01.tci-thaijo.org/index.php/sci_ubu/article/view/87160
- เบญจมาศ ปิยะ. (2558) *การศึกษาพฤติกรรมผู้ใช้งานตู้แช่แข็งพาณิชย์ด้วยเทคนิคต้นไม้ตัดสินใจ*. กรุงเทพฯ: สาขาวิชาเทคโนโลยีสารสนเทศและการจัดการ มหาวิทยาลัยกรุงเทพ.
- ปัทมา เทียงสมบุญ และนิเวศ จิระวิชิตชัย. (2561). *การพัฒนาระบบธุรกิจอัจฉริยะเพื่อสนับสนุน การพยากรณ์และการตัดสินใจของผู้บริหาร กรณีศึกษากลุ่มโรงพยาบาล*. สืบค้น 21 พฤษภาคม 2563, จาก <https://ph01.tci-thaijo.org/index.php/VESTSU/article/view/148906>
- พรรณธิภา เพชรบุญมี, ดวงกมล โพธิ์นาค และมนต์ชัย เทียนทอง. (2556). *การพยากรณ์ รูปแบบการเรียนรู้ตามประสบการณ์ของเดวิดโคสป์โดยใช้กฎการจำแนกเทคนิคต้นไม้ ตัดสินใจ*. สืบค้น 21 พฤษภาคม 2563, จาก <https://li01.tci-thaijo.org/index.php/tsti/article/view/14536/13308>
- รุจิรา ธรรมสมบัติ. (2554). *ระบบสนับสนุนการตัดสินใจในการเลือกใช้แพคเกจอินเทอร์เน็ตมือถือ โดยใช้ต้นไม้ตัดสินใจ*. วิทยาลัยราชพฤกษ์.

- สุมาลี ลิกเสนและมนตชัย เทียนทอง. (2559). การวิเคราะห์ความถนัดของผู้เรียนตามแนวทางพหุปัญญาโดยใช้กฎการจำแนกเทคนิคต้นไม้ตัดสินใจ. สืบค้น 21 พฤษภาคม 2563, จาก <https://so05.tci-thaijo.org/index.php/rmuj/article/view/58416>
- อนุชิต พงศ์พรหม, พลกฤต แสงอาวุธและพูนฉัตร วิชัยดิษฐ์. (2562). การศึกษาการบริหารจัดการทุนการศึกษาของนักศึกษาระดับมหาวิทยาลัยของรัฐจังหวัดสุราษฎร์ธานี. สืบค้น 21 พฤษภาคม 2563, จาก <https://so04.tci-thaijo.org/index.php/ljsj/article/view/204151/159492>
- อภิยศ เจริญวิวัฒน์. (2562). การวิเคราะห์ข้อมูลและการนำเสนอข้อมูลเชิงภาพ เพื่อประยุกต์ใช้กับการแสดงข้อมูลสถิติของสำนักหอสมุด มหาวิทยาลัยเกษตรศาสตร์. สืบค้น 21 พฤษภาคม 2563, จาก <https://pulinet.oas.psu.ac.th/index.php/journal/article/view/381>

ภาคผนวก

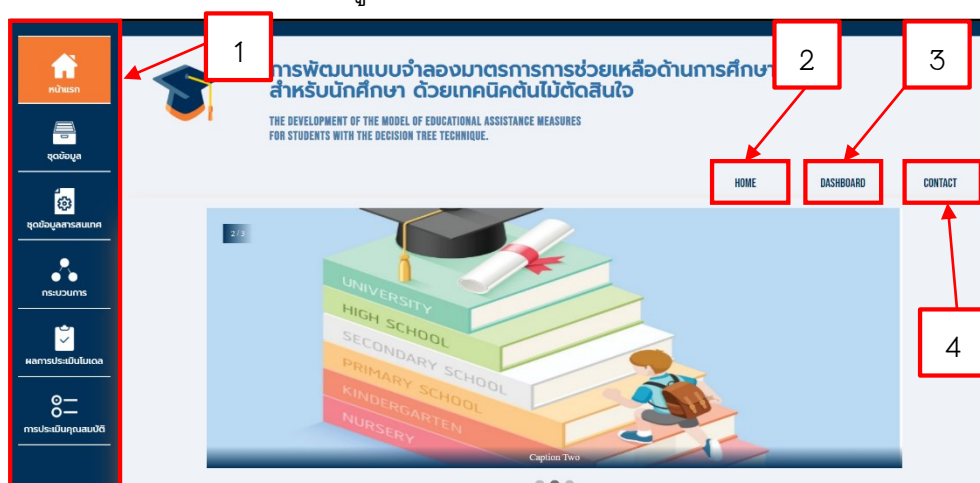
ภาคผนวก ก
คู่มือการใช้งานเว็บไซต์

คู่มือการใช้งานเว็บไซต์

จากการดำเนินงานการพัฒนาแบบจำลองมาตรการการช่วยเหลือด้านการศึกษา สำหรับนักศึกษา ด้วยเทคนิคต้นไม้ตัดสินใจ โดยเว็บไซต์ มีส่วนที่สามารถใช้งาน 6 ส่วน ดังนี้ ส่วนข้อมูลของโครงการและแดชบอร์ดสรุปข้อมูลต่าง ๆ ส่วนชุดข้อมูลที่นำมาใช้ ส่วนชุดข้อมูลสารสนเทศที่นำมาใช้ ส่วนการวิเคราะห์ข้อมูลด้วย Decision Tree ส่วนผลการประเมินโมเดล และส่วนของการประเมินคุณสมบัติ ผู้จัดทำจึงได้ทำคู่มืออธิบายการใช้งานเว็บไซต์ ดังต่อไปนี้

คู่มือการใช้งานสำหรับผู้เข้าใช้งานเว็บไซต์

- 1) หน้าแรกของเว็บไซต์เมื่อเข้าสู่เว็บไซต์



ภาพที่ ก.1 แสดงหน้าแรกของเว็บไซต์

หมายเลข 1 ส่วนของแถบเมนูด้านข้างที่ประกอบไปด้วยเมนูการใช้งานเพื่อไปยังหน้าต่าง ๆ ภายในเว็บไซต์

หมายเลข 2 ส่วนแสดงเนื้อหาในหน้าแรก ที่ประกอบไปด้วยจุดเริ่มต้นของโครงการ

หมายเลข 3 ส่วนแสดงแดชบอร์ด

หมายเลข 4 ส่วนการติดต่อกับผู้จัดทำเว็บไซต์

2) หน้าแสดงผลสรุปข้อมูลต่าง ๆ จากแดชบอร์ด



ภาพที่ ก.2 แสดงหน้าสรุปผลแดชบอร์ด

หมายเลข 1 ส่วนอธิบายข้อมูลรายละเอียดภายในแดชบอร์ด

หมายเลข 2 ส่วนของแดชบอร์ดแสดงผล

หมายเลข 3 ส่วนตัวกรองข้อมูลแดชบอร์ดแสดงผล ผู้ใช้งานสามารถใช้ตัวกรองดูข้อมูลตามความต้องการได้

3) หน้าแสดงชุดข้อมูลที่นำมาใช้

ลำดับที่	เพศ	อายุ	ที่อยู่	ขนาดครอบครัว	สถานะครอบครัว	ทรัพย์สินราคา	ทรัพย์สินราคา	อาชีพ	อาชีพ	เหตุผลในการเรียนสถานที่	ผู้ปกครอง	เวลาเดินทาง	เวลาเรียน	ความสนใจในรายวิชา	การสนับสนุน
1	F	18	U	GT3	A	4	4	at_home	teacher	course	mother	2	2	0	no
2	F	17	U	GT3	T	1	1	at_home	other	course	father	1	2	0	yes
3	F	15	U	LF3	T	1	1	at_home	other	other	mother	1	2	3	no

ภาพที่ ก.3 แสดงหน้าชุดข้อมูลที่นำมาใช้

หมายเลข 1 ส่วนของคำอธิบายคอลัมน์

หมายเลข 2 ส่วนของการค้นหา ผู้ใช้สามารถค้นหาข้อมูลที่ต้องการจากช่องค้นหาได้

หมายเลข 3 ส่วนของเลือกผลการแสดงในแต่ละหน้า ผู้ใช้สามารถเลือกใช้ตามต้องการได้

หมายเลข 4 ส่วนของการแสดงผลข้อมูลสารสนเทศที่นำมาใช้

4) หน้าแสดงชุดข้อมูลสารสนเทศที่นำมาใช้

ลำดับที่	เพศ	อายุ	ที่อยู่	ขนาดครอบครัว	สถานะครอบครัว	มรศึกษา	มรศึกษา	อาชีพประจำ	อาชีพว่าง	เหตุผลในการเรียน	ผู้ปกครอง	เวลาเข้า	เวลาเรียน	ความถี่	การสนับสนุน
1	หญิง	18-20	ในเมือง	3 คนขึ้นไป	แยกกัน	อื่นๆ	อื่นๆ	อยู่บ้าน	ครู/อาจารย์	หลักสูตร	มารดา	ปกติ	ปกติ	ไม่บ่อย	ไม่สนับสนุน
2	หญิง	15-17	ในเมือง	3 คนขึ้นไป	อยู่ด้วยกัน	ประถม	ประถม	อยู่บ้าน	อื่นๆ	หลักสูตร	บิดา	ปกติ	ปกติ	ไม่บ่อย	สนับสนุน
3	หญิง	15-17	ในเมือง	น้อยกว่า3	อยู่ด้วยกัน	ประถม	ประถม	อยู่บ้าน	อื่นๆ	อื่นๆ	มารดา	ปกติ	ปกติ	ไม่บ่อย	ไม่สนับสนุน
4	หญิง	15-17	ในเมือง	3 คนขึ้นไป	อยู่ด้วยกัน	อื่นๆ	มัธยม	หมอ/พยาบาล	พนักงานทั่วไป	หลักสูตร	มารดา	ปกติ	ปกติ	ไม่บ่อย	สนับสนุน
5	หญิง	15-17	ในเมือง	3 คนขึ้นไป	อยู่ด้วยกัน	ปริญญา	ปริญญา	อื่นๆ	อื่นๆ	หลักสูตร	บิดา	ปกติ	ปกติ	ไม่บ่อย	สนับสนุน
6	ชาย	15-17	ในเมือง	น้อยกว่า3	อยู่ด้วยกัน	อื่นๆ	ปริญญา	พนักงานทั่วไป	อื่นๆ	หลักสูตร	มารดา	ปกติ	ปกติ	ไม่บ่อย	สนับสนุน
7	ชาย	15-17	ในเมือง	น้อยกว่า3	อยู่ด้วยกัน	มัธยม	มัธยม	อื่นๆ	อื่นๆ	หลักสูตร	มารดา	ปกติ	ปกติ	ไม่บ่อย	ไม่สนับสนุน
8	หญิง	15-17	ในเมือง	3 คนขึ้นไป	แยกกัน	อื่นๆ	อื่นๆ	ครู/อาจารย์	พนักงานทั่วไป	หลักสูตร	มารดา	ปกติ	ปกติ	ไม่บ่อย	สนับสนุน
9	ชาย	15-17	ในเมือง	น้อยกว่า3	แยกกัน	ปริญญา	มัธยม	พนักงานทั่วไป	อื่นๆ	หลักสูตร	มารดา	ปกติ	ปกติ	ไม่บ่อย	สนับสนุน
10	ชาย	15-17	ในเมือง	3 คนขึ้นไป	อยู่ด้วยกัน	ปริญญา	อื่นๆ	อื่นๆ	อื่นๆ	หลักสูตร	มารดา	ปกติ	ปกติ	ไม่บ่อย	สนับสนุน

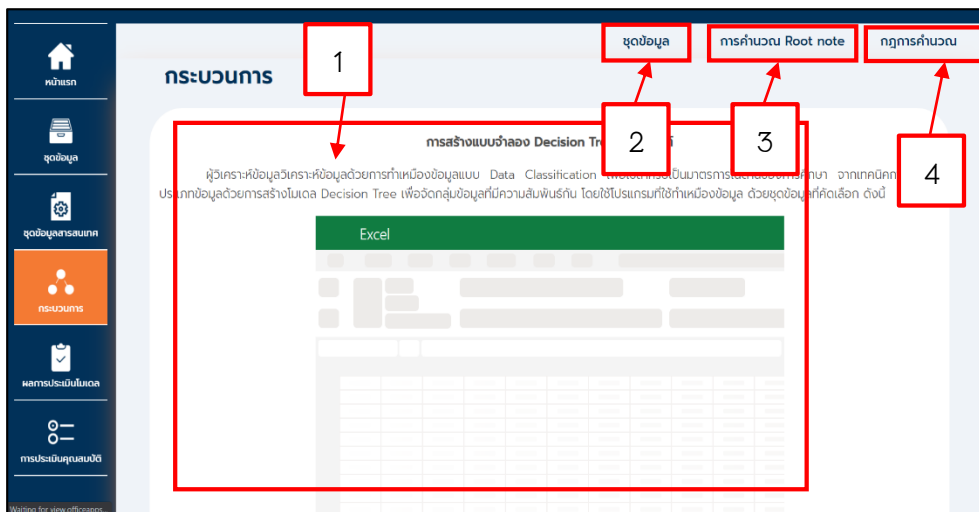
ภาพที่ ก.4 แสดงหน้าชุดข้อมูลที่นำมาใช้

หมายเลข 1 ส่วนของการค้นหา ผู้ใช้สามารถค้นหาข้อมูลที่ต้องการจากช่องค้นหาได้

หมายเลข 2 ส่วนของการเลือกผลการแสดงในแต่ละหน้า ผู้ใช้สามารถเลือกใช้ตามต้องการได้

หมายเลข 3 ส่วนของการแสดงผลข้อมูลสารสนเทศที่นำมาใช้

5) หน้าแสดงกระบวนการในการวิเคราะห์ข้อมูล



ภาพที่ ก.5 แสดงหน้ากระบวนการวิเคราะห์ข้อมูล

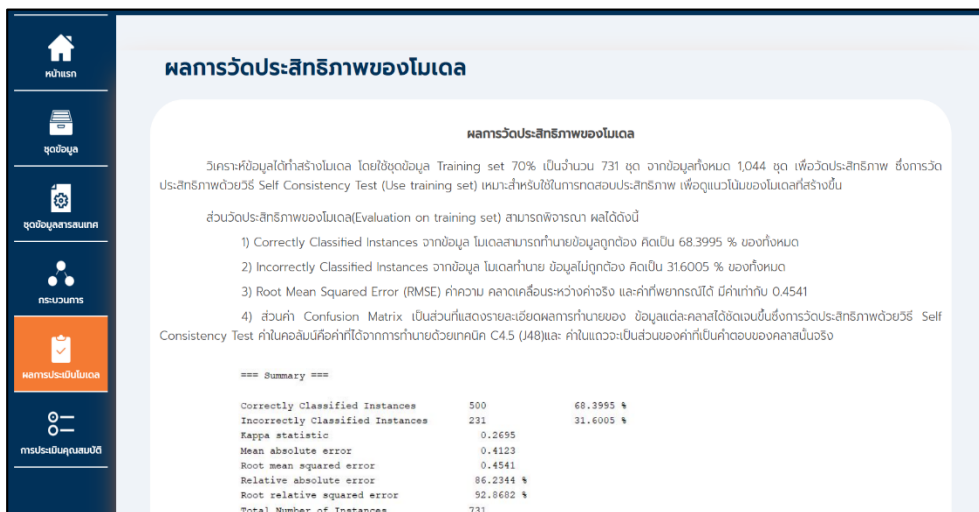
หมายเลข 1 ส่วนของการแสดงข้อมูลกระบวนการ

หมายเลข 2 ส่วนของชุดข้อมูลที่นำมาใช้วิเคราะห์

หมายเลข 3 ส่วนของการแสดงวิธีการคำนวณ Root note

หมายเลข 4 ส่วนของการกฎที่ได้จากการสร้างโมเดล Decision Tree

6) หน้าแสดงผลการวัดประสิทธิภาพของโมเดล



ภาพที่ ก.6 แสดงหน้ากระบวนการวิเคราะห์ข้อมูล

7) หน้าการประเมินคุณสมบัติ โดยผู้ใช้งานสามารถประเมินคุณสมบัติของตนเองได้

ภาพที่ ก.7 แสดงหน้าการประเมินคุณสมบัติ

8) หน้าแสดงผลลัพธ์ในการประเมินคุณสมบัติ

รหัสนักศึกษา : 6152107098-6	ชื่อ-สกุล : สุนพงษ์ ทาตาสาย
ระดับ : ประกาศนียบัตรวิชาชีพ (ปวช.)	คณะ : คณะบริหารธุรกิจและศิลปศาสตร์
ชั้นปี : 5	เพศ : ชาย
อายุ : 15-17	ที่อยู่อาศัย : ในเมือง
ที่อยู่ปัจจุบัน : ในเมือง	ขนาดครอบครัว : บิดามารดา 3
สถานะครอบครัว : หย่ากันอยู่	การศึกษาของมารดา : ประถม
การศึกษาของบิดา : ประถม	อาชีพของมารดา : อยู่บ้าน
อาชีพของบิดา : อยู่บ้าน	ผู้ปกครอง : มารดา
เหตุผลเข้าศึกษา : ฝึกงาน	ปัจจุบันได้กู้ยืมกองทุนเพื่อการศึกษาหรือไม่ : กู้ยืม
ทำงานในขณะที่เรียนหรือไม่ : ทำงาน	ยานพาหนะใช้ในการมาเรียน : รถจักรยานยนต์
เวลาในการเดินทาง : ใช้เวลาปกติ	เวลาในการเรียน : เรียนตามปกติ
ความสนใจในรายวิชา : เศรษฐ	

ภาพที่ ก.8 แสดงหน้าผลลัพธ์การประเมินคุณสมบัติ

หมายเลข 1 ส่วนการแสดงผลข้อมูลของผู้ใช้งาน

หมายเลข 2 ส่วนการแสดงผลลัพธ์ที่ได้จากการประเมินคุณสมบัติ

9) หน้าการเข้าสู่ระบบเฉพาะแอดมินเท่านั้น

The screenshot shows a web application interface for an administrator login. At the top, there is a header with a graduation cap icon and the title 'การพัฒนาแบบจำลองมาตรการการช่วยเหลือด้านการศึกษา สำหรับนักศึกษา ด้วยเทคนิคต้นไม้ตัดสินใจ' (The Development of the Model of Educational Assistance Measures for Students with the Decision Tree Technique). Below the header is a navigation bar with buttons for 'LOGIN', 'PRODUCER', 'ADVISOR', and 'DEPARTMENT'. The 'ADVISOR' button is highlighted with a red box and labeled '1'. The main content area is titled 'LOGIN' and contains a form with two input fields: 'USERNAME' (labeled '2') and 'PASSWORD' (labeled '3'). Below the input fields are two buttons: 'Login' (labeled '4') and 'Cancel' (labeled '5'). A red box also highlights the 'ADVISOR' dropdown menu in the navigation bar.

ภาพที่ ก.9 แสดงหน้าเข้าสู่ระบบ

หมายเลข 1 ส่วนของแถบเมนูเลือกแสดงผลได้ตามความต้องการของผู้ใช้งาน

หมายเลข 2 ส่วนการกรอก Username โดยจะสามารถเข้าสู่ระบบได้เฉพาะแอดมินเท่านั้น

หมายเลข 3 ส่วนการกรอก Password โดยจะสามารถเข้าสู่ระบบได้เฉพาะแอดมินเท่านั้น

หมายเลข 4 ปุ่มคลิกเพื่อเข้าสู่ระบบ

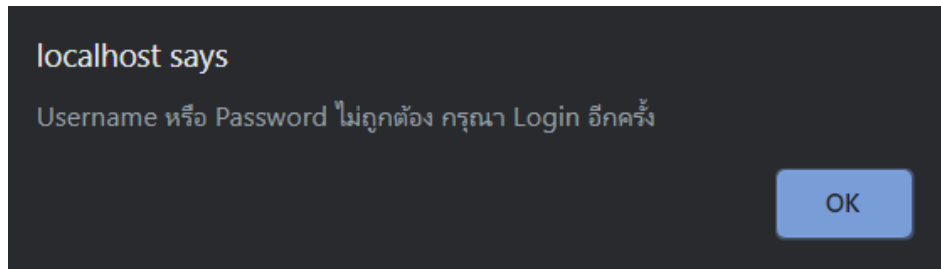
หมายเลข 5 ปุ่มคลิกเพื่อยกเลิก

10) การแจ้งเตือนเข้าสู่ระบบสำเร็จ



ภาพที่ ก.10 แสดงการแจ้งเตือนเข้าสู่ระบบสำเร็จ

11) การแจ้งเตือนเข้าสู่ระบบไม่สำเร็จ



ภาพที่ ก.11 แสดงการแจ้งเตือนเข้าสู่ระบบไม่สำเร็จ

12) แสดงผลหน้าการเข้าสู่ระบบของผู้ดูแลระบบ

ยินดีต้อนรับ User : admin เข้าสู่ระบบ

ผลลัพธ์

รายการข้อมูลทั้งหมด

สามารถค้นหาได้ทุกคอลัมน์ ค้นหา

หน้า: หน้าแรก | หน้าถัดไป | หน้าก่อนหน้า | หน้าสุดท้าย

รายการข้อมูลทั้งหมด 2 รายการ จากทั้งหมด 1 หน้า

ลำดับที่	รหัสนักศึกษา	ชื่อ	นามสกุล	ระดับ	คณะ	ชั้นปี	เพศ	อายุ	ที่อยู่อาศัย	ที่อยู่จัด
9	6152107098-6	กาญจนาณี	ใจกลอน	ประกาศนียบัตร วิชาชีพ (ปวช.)	คณะ: บริหารธุรกิจและ ศิลปศาสตร์	1	ชาย	18-20	ในเมือง	ในเมือง
10	6152107098-6	สพพณีย์	กาศาสาข	ประกาศนียบัตร วิชาชีพ (ปวช.)	คณะ: บริหารธุรกิจและ ศิลปศาสตร์	5	ชาย	15-17	ในเมือง	ในเมือง

ภาพที่ ก.12 แสดงหน้าการเข้าสู่ระบบของผู้ดูแลระบบ

หมายเลข 1 ปุ่มออกจากระบบ

หมายเลข 2 ส่วนของการค้นหา ผู้ใช้สามารถค้นหาข้อมูลที่ต้องการจากช่องค้นหาได้

หมายเลข 3 ส่วนของการเลือกผลการแสดงในแต่ละหน้า ผู้ใช้สามารถเลือกใช้ตามต้องการได้

หมายเลข 4 ส่วนของการแสดงผลข้อมูลของผู้ใช้ที่ทำการประเมินคุณสมบัติเข้ามา

13) เป็นหน้าจอแสดงประวัติผู้จัดทำเว็บไซต์ และข้อมูลการติดต่อ รวมถึงแสดงหน่วยงานที่เกี่ยวข้องกับโครงการครั้งนี้

ประวัติผู้จัดทำ

	<p>ชื่อ-สกุล : นายพานุสน์ ไวกุล</p> <p>วันเดือนปีเกิด : วันพุธ ที่ 27 เดือนพฤษภาคม พ.ศ.2541</p> <p>ภูมิลำเนา : 205 หมู่ 3 ตำบลพระบาท อำเภอเมืองลำปาง จังหวัดลำปาง 52000</p> <p>E-mail : owen.panusorn@outlook.com</p> <p>การศึกษา : เข้ารับศึกษาในระดับปริญญาตรี(เทียบโอน) พุทธศักราช 2561 ณ มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา เชียงใหม่</p> <p>สาขาวิชา : ระบบสารสนเทศทางธุรกิจ-การจัดการสารสนเทศ</p>
	<p>ชื่อ-สกุล : นายสมพงษ์ กตตาสาย</p> <p>วันเดือนปีเกิด : วันพฤหัสบดี ที่ 12 เดือนมีนาคม พ.ศ.2541</p> <p>ภูมิลำเนา : 385 หมู่ 2 ตำบลป่าเสด็จ อำเภอเมืองลำปาง จังหวัดลำปาง 52000</p> <p>E-mail : boydragorn8@gmail.com</p> <p>การศึกษา : เข้ารับศึกษาในระดับปริญญาตรี(เทียบโอน) พุทธศักราช 2561 ณ มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา เชียงใหม่</p>

ภาพที่ ก.13 แสดงส่วนติดต่อของคุณะผู้จัดทำ

ภาคผนวก ข
แบบฟอร์มและเอกสารที่ใช้ในโครงการ

แบบสอบถาม

โครงการเรื่อง : การพัฒนาแบบจำลองมาตรการการช่วยเหลือด้านการศึกษา สำหรับนักศึกษา
ด้วยเทคนิคต้นไม้ตัดสินใจ

คำชี้แจง : แบบสอบถามประกอบด้วยชุดคำถาม 3 ส่วน คือ

ส่วนที่ 1 ข้อมูลทั่วไป

ส่วนที่ 2 การรับรู้และความเข้าใจ

ส่วนที่ 3 ข้อเสนอแนะ/ข้อคิดเห็นอื่น ๆ

ตอนที่1 ข้อมูลทั่วไปของผู้ใช้งาน

ข้อชี้แจง ใส่เครื่องหมาย ลงใน หน้าคำตอบที่ตรงกับข้อมูลของผู้ใช้งานเพียงข้อเดียว

1. เพศ

ชาย

หญิง

2. อายุ

น้อยกว่า 18 ปี

18-24 ปี

25 ปีขึ้นไป

ตอนที่2 การรับรู้และความเข้าใจ

ข้อชี้แจง ใส่เครื่องหมาย ลงในช่องระดับคะแนนความคิดเห็นที่ตรงกับความคิดเห็นของผู้ใช้มากที่สุด โดยที่ 5 = มากที่สุด 4 = มาก 3 = ปานกลาง 2 = น้อย และ 1 = น้อยที่สุด

ตาราง ข.1 แสดงข้อมูลบนแบบสอบถาม

รายการ	ระดับความคิดเห็น				
	5	4	3	2	1
ด้านข้อมูล					
1. แหล่งความรู้มีความชัดเจน และสอดคล้องกับงาน					
2. ข้อมูลที่นำเสนอบนเว็บไซต์มีขนาดและองค์ประกอบที่เหมาะสม					
3. การจัดลำดับเนื้อหาเป็นขั้นตอน มีความต่อเนื่อง อ่านแล้วเข้าใจ					

รายการ	ระดับความคิดเห็น				
	5	4	3	2	1
4. รูปแบบ Visualization มีความเหมาะสมกับข้อมูล					
5. การวิเคราะห์ที่มีความชัดเจน น่าเชื่อถือ					
ด้านการออกแบบและการจัดรูปแบบเว็บไซต์					
6. การจัดรูปแบบในเว็บไซต์ต้ง่ายต่อการอ่านและสะดวกต่อการใช้งาน					
7. เว็บไซต์มีความสวยงาม มีความทันสมัย น่าสนใจ					
8. สีสีนในการออกแบบเว็บไซต์มีความเหมาะสม					
9. สีพื้นหลังกับสีตัวอักษรมีความเหมาะสมต่อการอ่าน					
10. ขนาดตัวอักษร และรูปแบบตัวอักษร มีความสวยงามและอ่านได้ง่าย					
ด้านประโยชน์และการนำไปใช้					
11. เนื้อหามีรายละเอียดและประโยชน์เพียงพอสามารถนำไปประยุกต์ใช้ได้					
12. สามารถนำไปใช้เป็นแหล่งอ้างอิงให้กับโครงการงานเรื่องอื่นที่เกี่ยวข้องได้					
13. สามารถใช้เป็นแหล่งความรู้และแนวทางในการพัฒนาต่อไปได้					
14. แหล่งข้อมูลที่ตรงกับความต้องการของผู้ใช้งาน					

ตอนที่ 3 ข้อเสนอแนะ

.....

.....

.....

.....

ประวัติผู้จัดทำ



ชื่อ-สกุล	นายภาณุสรณ์ ใจกลม
วันเดือนปีเกิด	วันพุธ ที่ 27 เดือนพฤษภาคม พ.ศ.2541
ภูมิลำเนา	205 หมู่ 3 ตำบลพระบาท อำเภอเมืองลำปาง จังหวัดลำปาง 52000
E-mail	owen.panusorn@outlook.com

ประวัติการศึกษา

- ระดับมัธยมศึกษาตอนต้น จากบุญวาทย์วิทยาลัย จังหวัดลำปาง สำเร็จการศึกษาปี 2555
- ระดับประกาศนียบัตรวิชาชีพ สาขาวิชาเทคโนโลยีสารสนเทศ จากวิทยาลัยเทคนิคลำปาง จังหวัดลำปาง สำเร็จการศึกษาปี 2558
- ระดับประกาศนียบัตรวิชาชีพชั้นสูง สาขาวิชาเทคโนโลยีสารสนเทศ จากวิทยาลัยเทคนิคลำปาง จังหวัดลำปาง สำเร็จการศึกษาปี 2560
- ระดับปริญญาตรี สาขาวิชาระบบสารสนเทศทางธุรกิจ มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา จังหวัดเชียงใหม่ สำเร็จการศึกษาปี 2563



ชื่อ-สกุล นายสมพงษ์ กาดาสาย
วันเดือนปีเกิด วันพฤหัสบดี ที่ 12 เดือนมีนาคม พ.ศ.2541
ภูมิลำเนา 385 หมู่ 2 ตำบลบ้านเสด็จ อำเภอเมืองลำปาง จังหวัดลำปาง 52000
E-mail Boydragon81@gmail.com

ประวัติการศึกษา

- ระดับมัธยมศึกษาตอนต้น จากโรงเรียนเสด็จวนชยางค์กุลวิทยา จังหวัดลำปาง สำเร็จการศึกษาปี 2555
- ระดับมัธยมศึกษาตอนปลาย จากโรงเรียนเสด็จวนชยางค์กุลวิทยา จังหวัดลำปาง สำเร็จการศึกษาปี 2558
- ระดับประกาศนียบัตรวิชาชีพชั้นสูง สาขาวิชาคอมพิวเตอร์ธุรกิจ มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา ลำปาง จังหวัดลำปาง สำเร็จการศึกษาปี 2560
- ระดับปริญญาตรี สาขาวิชาระบบสารสนเทศทางธุรกิจ มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา จังหวัดเชียงใหม่ สำเร็จการศึกษาปี 2563