

บทที่ 3

วิธีการดำเนินงานโครงการ

โครงการเรื่อง การพัฒนาแบบจำลองมาตรฐานการช่วยเหลือด้านการศึกษ สำหรับ นักศึกษา ด้วยเทคนิคต้นไม้ตัดสินใจ ในบทนี้จะเป็นการวิเคราะห์ข้อมูลด้วยเทคนิคทางดาต้าไมนิ่ง ซึ่งมีกระบวนการวิเคราะห์ที่สำคัญหลายขั้นตอน เมื่อเสร็จสิ้นจากกระบวนการวิเคราะห์ข้อมูลแล้ว จะเป็นการออกแบบเว็บไซต์ และออกแบบรูปแบบการแสดงผลและบทสรุปจากวิธีการดำเนินงาน

3.1 การวิเคราะห์ข้อมูลด้วย CRISP-DM

3.2 การออกแบบเว็บไซต์

3.3 บทสรุป

3.1 การวิเคราะห์ข้อมูลด้วย CRISP-DM

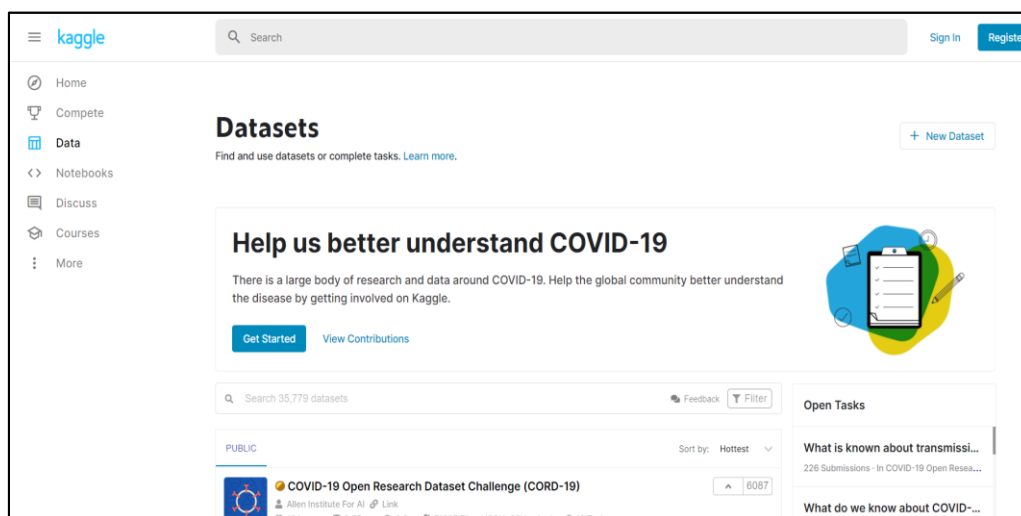
กระบวนการวิเคราะห์ข้อมูลด้วย CRISP-DM หรือ Cross Industry Standard Process for Data Mining พัฒนาขึ้นในปี ค.ศ. 1996 โดยความร่วมมือของ 3 บริษัทคือ Daimler Chrysler, SPSS และ NCR ที่มีการพัฒนาเป็น Work flow มาตรฐานสำหรับการทำเหมืองข้อมูล ประกอบด้วย 6 ขั้นตอนหลัก ดังนี้

3.1.1 รู้จักและเข้าใจในธุรกิจ (Business Understanding) เป็นขั้นตอนแรกของกระบวนการ ที่มุ่งเน้นไปที่การทำความเข้าใจกระบวนการทางธุรกิจโดยรวม

รู้จักและเข้าใจในธุรกิจ (Business Understanding) เป็นขั้นตอนแรกของกระบวนการ ที่มุ่งเน้นไปที่การทำความเข้าใจกระบวนการทางธุรกิจโดยรวม ผู้วิเคราะห์ข้อมูลทำความเข้าใจกับ ปัญหาให้อยู่ในรูปของการวิเคราะห์ข้อมูลทาง Data Mining โดยการวิเคราะห์ข้อมูลในประเด็นนี้ ซึ่งมีข้อมูลมาก ทำให้ไม่สามารถทำความเข้าใจกับข้อมูลที่มีความซับซ้อนได้อย่างรวดเร็ว เช่น ต้องการทราบว่านักศึกษาลักษณะคุณสมบัติใดที่บ่งบอกถึงความต้องการช่วยเหลือในด้านการ การศึกษา

3.1.2 จัดเก็บและรวบรวมข้อมูลให้ครบ (Data Understanding) ขั้นตอนการจัดเก็บ และรวบรวมข้อมูล ตลอดจนการพิจารณาตรวจสอบความถูกต้องของข้อมูลที่ได้รับ โดยเลือกว่า จะใช้ข้อมูลทั้งหมดหรือบางส่วนในการวิเคราะห์ให้สอดคล้องกับวัตถุประสงค์ที่กำหนดไว้

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 มากที่สุดตามลำดับและตามจำนวนที่ผู้วิเคราะห์ต้องการเลือกคอลัมน์ เพื่อนำไปใช้ในการสร้าง Model Decision tree ต่อไป

Forward Selection เป็นการสร้างโมเดลโดยการเพิ่มฟีเจอร์ทีละ 1 ฟีเจอร์ ถ้าฟีเจอร์ที่ใส่เพิ่มให้ประสิทธิภาพที่ดีก็จะเก็บไว้และเลือกฟีเจอร์อื่น ๆ มาเพิ่มต่อไปจนประสิทธิภาพของโมเดลไม่ได้ดีขึ้นก็จะหยุดทำงาน ทฤษฎีนี้ไม่สามารถวัดผลที่แน่ชัด เนื่องจากเป็นวิธีที่ใช้การสุ่มสลับคอลัมน์ไปเรื่อย ๆ จนกว่าจะได้ผลที่ดีที่สุด เป็นการวัดผลประสิทธิภาพโดยรวม ไม่ใช้การวัดผลคอลัมน์ที่จะนำมาใช้ในการสร้าง Decision tree

Backward Elimination เป็นการสร้างโมเดลที่เริ่มจากการใช้ฟีเจอร์ทั้งหมดก่อนและตัด (eliminate) ฟีเจอร์ที่ไม่สำคัญทิ้งไปทีละฟีเจอร์ถ้าประสิทธิภาพดีขึ้นก็ตัดฟีเจอร์อื่น ๆ ต่อไป ทฤษฎีนี้ไม่สามารถวัดผลที่แน่ชัด เนื่องจากเป็นวิธีที่ใช้การสุ่มสลับคอลัมน์ไปเรื่อย ๆ จนกว่าจะได้ผลที่ดีที่สุด เป็นการวัดผลประสิทธิภาพโดยรวม ไม่ใช้การวัดผลคอลัมน์ที่จะนำมาใช้ในการสร้าง Decision tree

Chi-Square เป็น สถิติที่ใช้ทดสอบความแตกต่างค่าเฉลี่ย ของกลุ่มตัวอย่างที่มีเพียงกลุ่มหรือสองกลุ่ม ข้อมูลที่นำมาทดสอบนั้นจะต้องเป็นข้อมูลที่อยู่ในระดับการวัด (Measurement Scale) ระดับอันดับภาคชั้น (Interver Scale) หรือระดับอัตราส่วน (Ratio Scale) ซึ่งข้อมูลส่วนใหญ่อยู่ในรูปแบบตัวเลข ซึ่งไม่สามารถนำมาคำนวณกับชุดข้อมูลที่มีอยู่ได้

ดังนั้นผู้วิเคราะห์ข้อมูลจึงใช้วิธี information Gain ในการคัดเลือกข้อมูล เพราะเป็นวิธีที่หาค่าน้ำหนักของคอลัมน์นั้น ๆ โดยตรงและเห็นผลที่ชัดเจนที่สุด โดยจะคำนวณในขั้นตอนต่อไป

ผู้วิเคราะห์ข้อมูลวิเคราะห์ข้อมูลด้วยการทำเหมืองข้อมูลแบบ Data Classification เพื่อใช้สำหรับเป็นมาตรการในด้านของการศึกษา จากเทคนิคการจำแนกประเภทข้อมูลด้วยการสร้างโมเดล Decision Tree เพื่อจัดกลุ่มข้อมูลที่มีความสัมพันธ์กัน โดยใช้โปรแกรมที่ใช้ทำเหมืองข้อมูลด้วยชุดข้อมูลที่คัดเลือก 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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

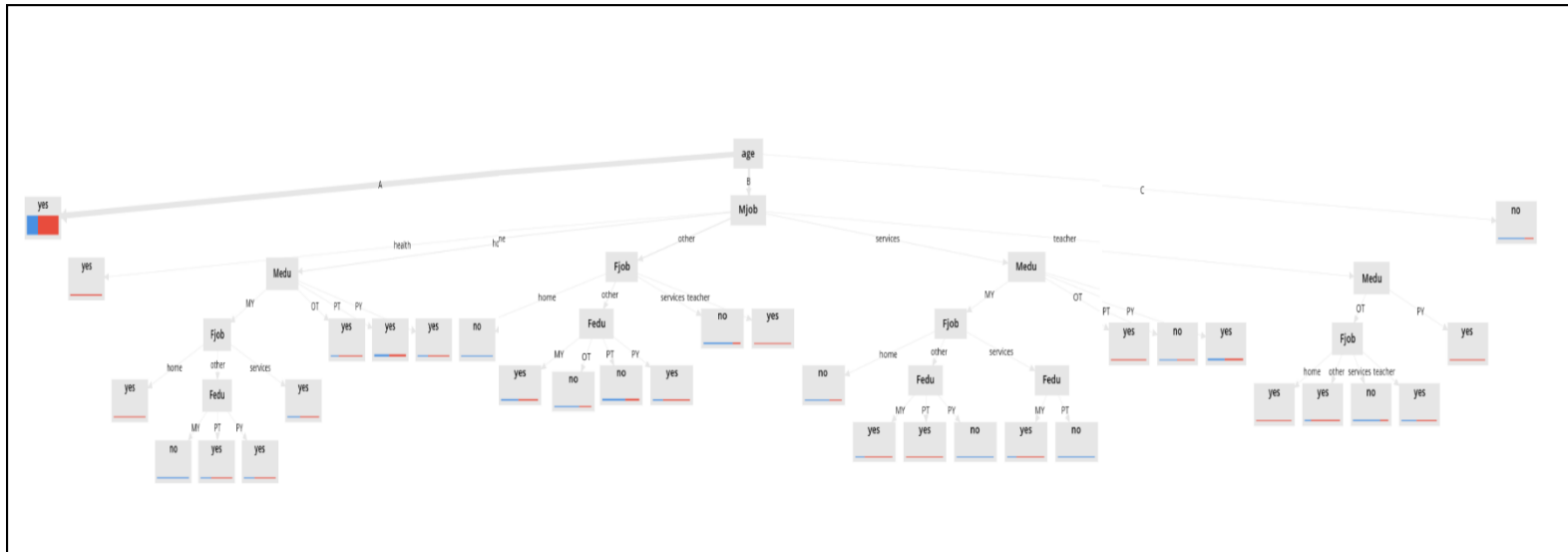
ตาราง 3.1 สรุปค่า Information Gain ของคอลัมน์ต่าง ๆ

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

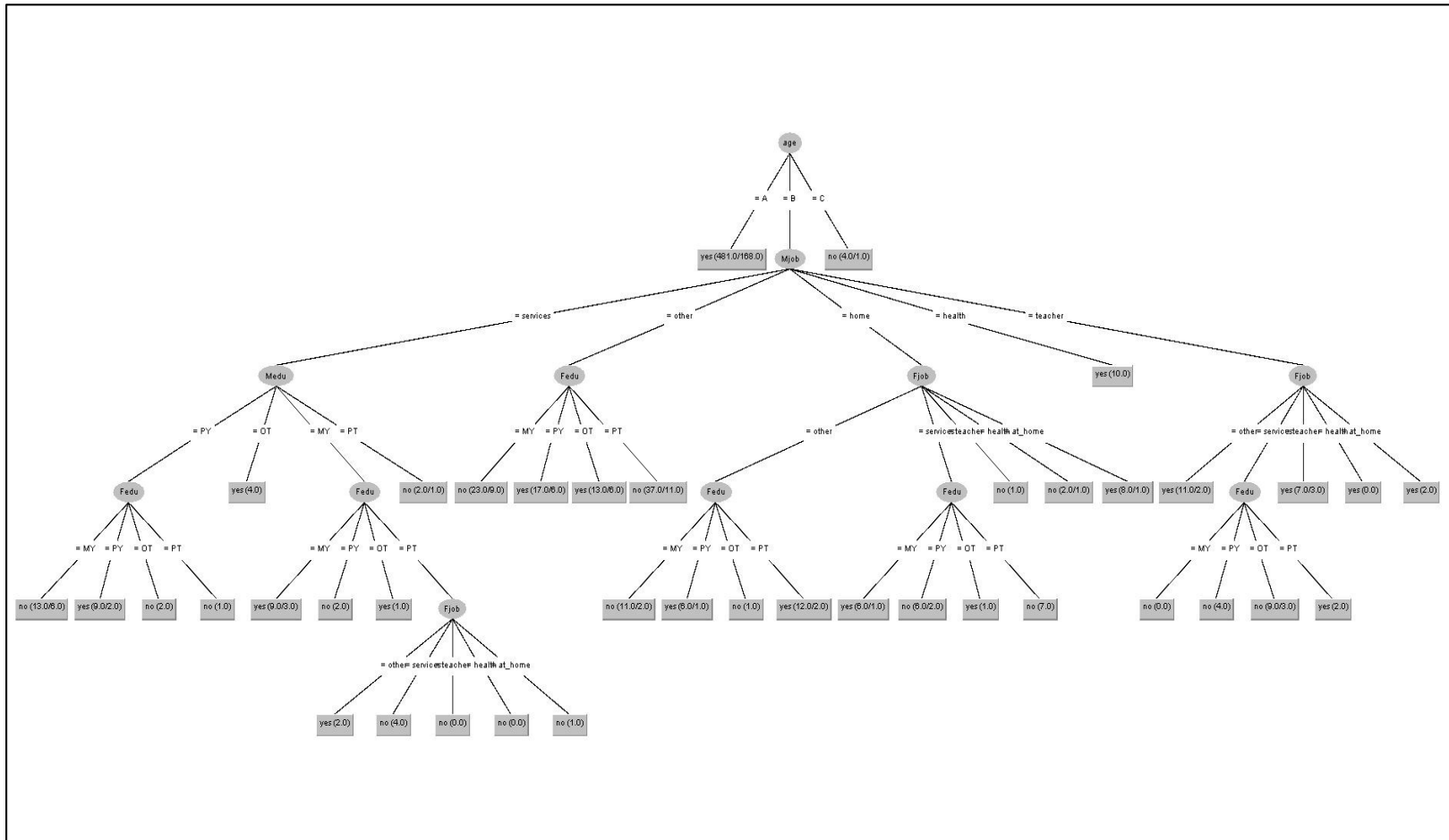

```

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”

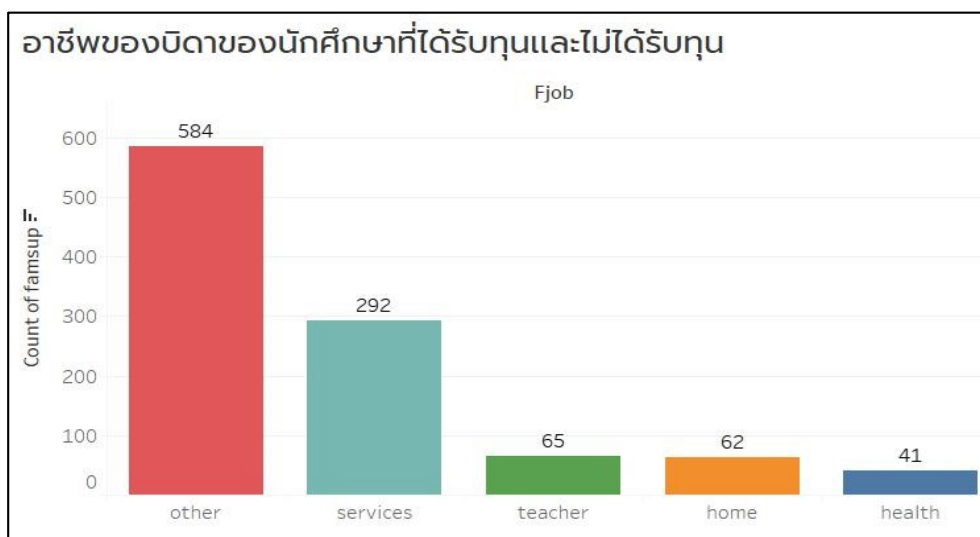
กฎข้อที่ 39 IF age=b and mjob=teacher and fjob=home then YES หมายความว่า ถ้าอายุเท่ากับ 18–20 และอาชีพมารดาเป็นครู/อาจารย์ และอาชีพบิดาเป็นอยู่บ้าน จะมีผลลัพธ์เป็น “สนับสนุน”

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

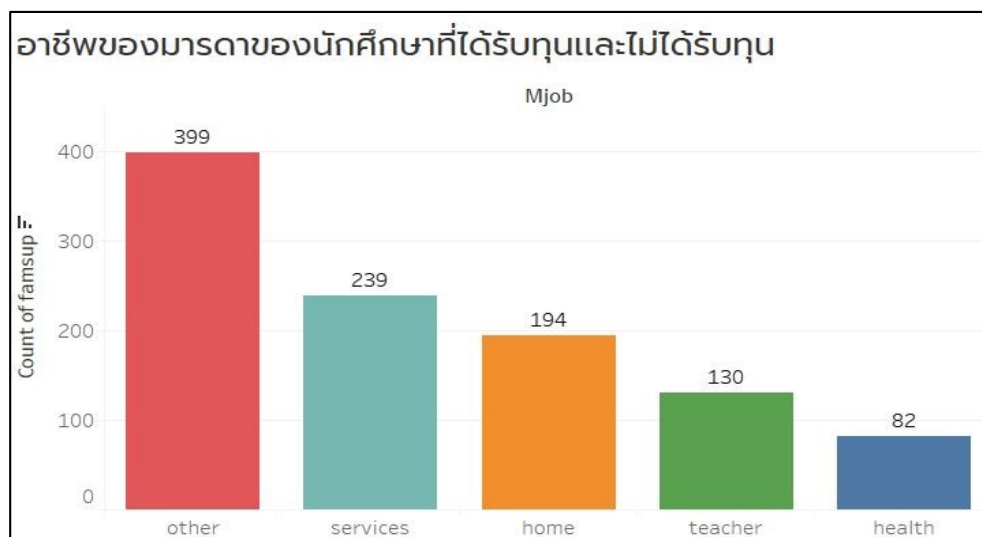
ผู้วิเคราะห์ข้อมูลนำผลข้อมูลที่ทำการวิเคราะห์แล้ว แสดงผลข้อมูลบน Web Browser โดยใช้ชุดคำสั่ง HTML CSS3 และ PHP ร่วมกับการนำเสนอข้อมูลแบบ Visualization ด้วยการแสดงผลข้อมูลในรูปแบบของภาพโดยใช้โปรแกรม Tableau Public

1) แสดงข้อมูลแสดงผลประเภทของอาชีพบิดาของนักศึกษาที่ได้รับทุนและไม่ได้รับทุนเรียงตามลำดับ ซึ่งเป็นการแสดงผลในรูปแบบแผนภูมิแท่ง ผู้ใช้สามารถเลือกดูแต่ละอาชีพของพ่อแม่ได้



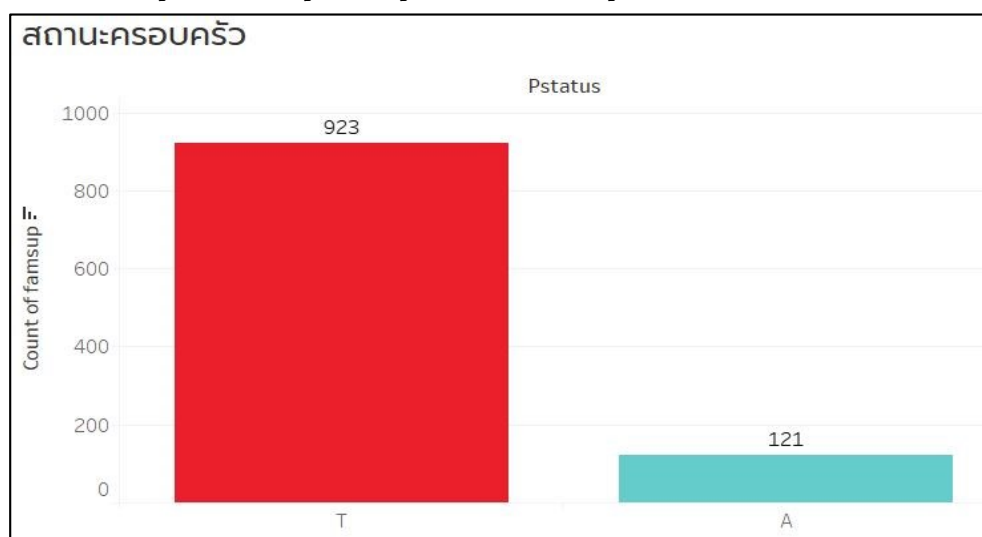
ภาพที่ 3.10 แสดงข้อมูลแสดงผลประเภทของอาชีพบิดาของนักศึกษาที่ได้รับทุนและไม่ได้รับทุน

2) แสดงข้อมูลแสดงผลประเภทของอาชีพมารดาของนักศึกษาที่ได้รับทุนและไม่ได้รับทุน เรียงตามลำดับ ซึ่งเป็นการแสดงผลในรูปแบบแผนภูมิแท่ง ผู้ใช้สามารถเลือกดูแต่ละอาชีพของมารดาได้



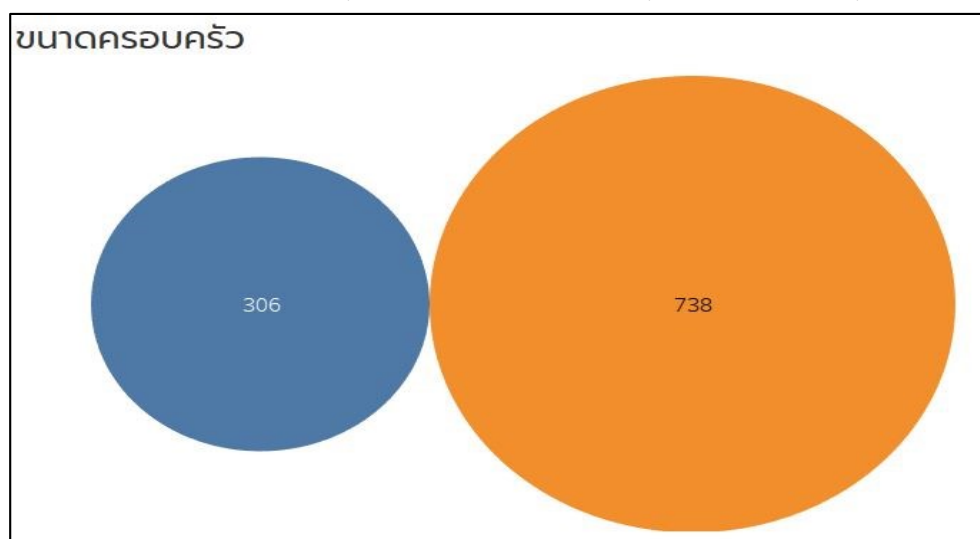
ภาพที่ 3.11 แสดงข้อมูลแสดงผลประเภทของอาชีพมารดาของนักศึกษาที่ได้รับทุนและไม่ได้รับทุน

3) แสดงข้อมูลแสดงผลประเภทของการอยู่กับครอบครัวและการอยู่ห่างจากครอบครัว ซึ่งเป็นการแสดงผลในรูปแบบแผนภูมิแท่ง ผู้ใช้สามารถเลือกดูแต่ละประเภทได้



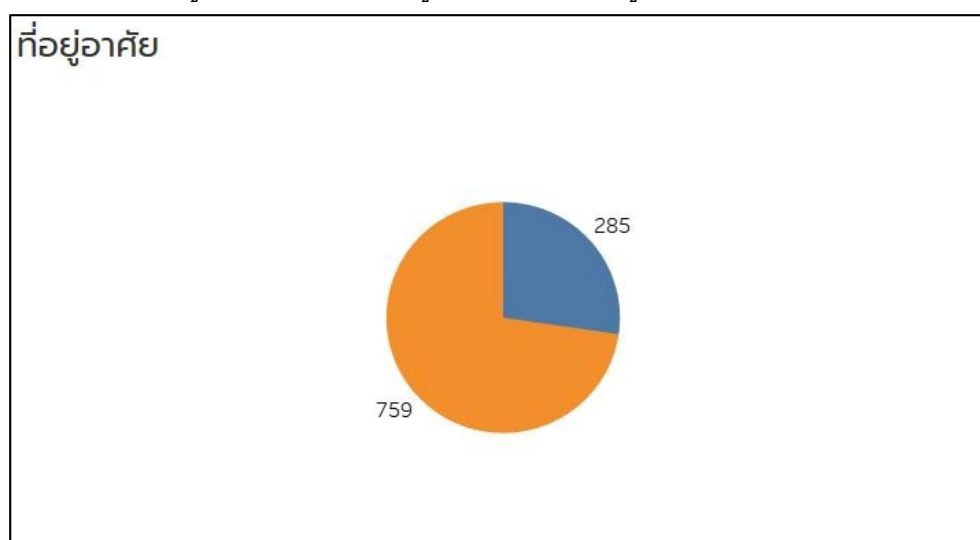
ภาพที่ 3.12 แสดงข้อมูลแสดงผลประเภทของเปอร์เซ็นต์การอยู่กับครอบครัวและการอยู่ห่าง

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



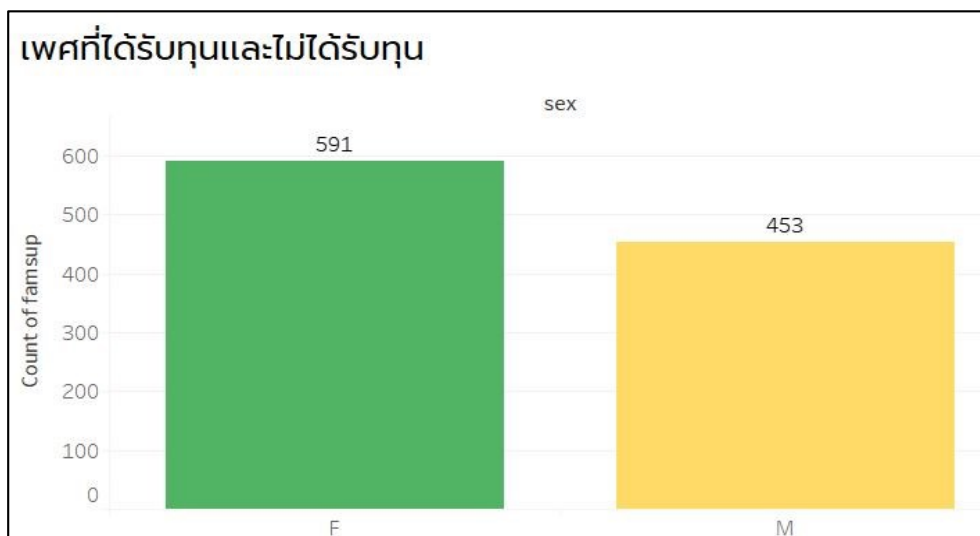
ภาพที่ 3.13 แสดงข้อมูลแสดงผลประเภทของครอบครัวตั้งแต่ 3 ขึ้นไปสนับสนุนการเรียนเท่าไร ไม่สนับสนุนการเรียนเท่าไร, ครอบครัวน้อยกว่า 3 สนับสนุนการเรียนเท่าไร ไม่สนับสนุนการเรียนเท่าไร

5) แสดงข้อมูลแสดงผลประเภทของการอยู่อาศัยในเมืองหรือนอกเมืองของนักศึกษา ซึ่งเป็นการแสดงผลในรูปแบบ Pie charts ผู้ใช้สามารถเลือกดูแต่ละประเภทได้



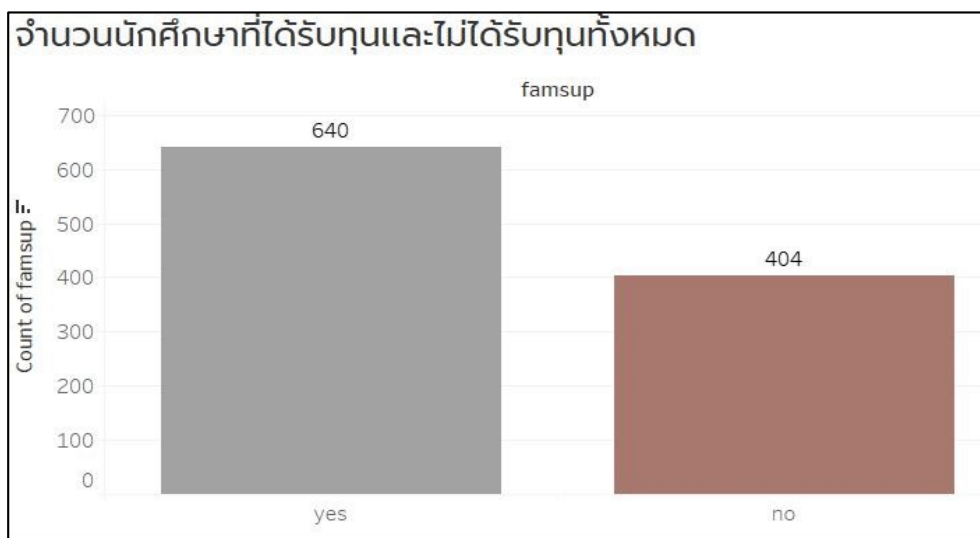
ภาพที่ 3.14 แสดงข้อมูลแสดงผลประเภทของการอยู่อาศัยในเมืองหรือนอกเมืองของนักศึกษา

6) แสดงข้อมูลแสดงผลจำนวนที่ได้รับทุนของเพศชายและเพศหญิง ซึ่งเป็นการแสดงผลในรูปแบบแผนภูมิแท่ง ผู้ใช้สามารถเลือกดูแต่ละประเภทได้



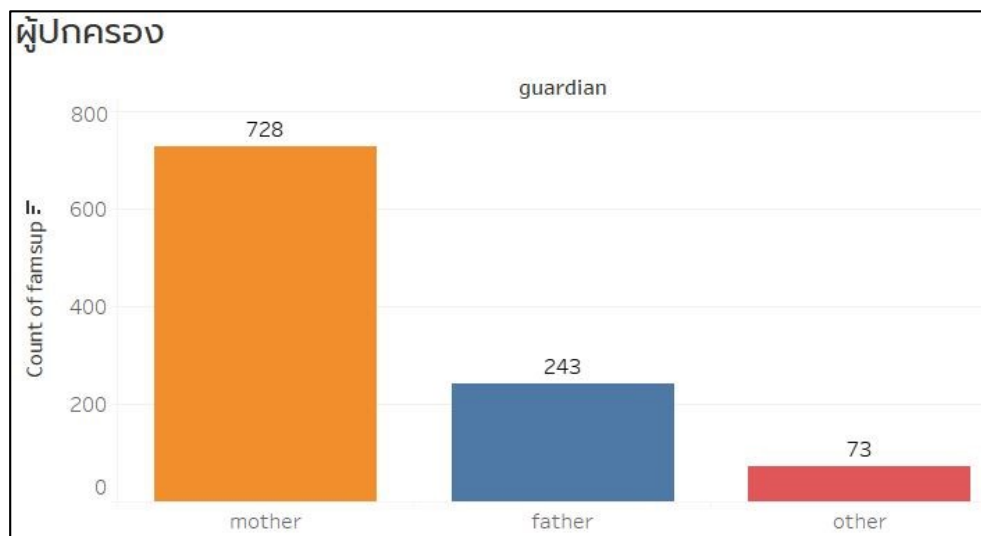
ภาพที่ 3.15 แสดงข้อมูลแสดงผลจำนวนที่ได้รับทุนของเพศชายและเพศหญิง

7) แสดงข้อมูลแสดงผลการได้รับทุนการศึกษา และไม่ได้รับทุนการศึกษาทั้งหมด ซึ่งเป็นการแสดงผลในรูปแบบแผนภูมิแท่ง



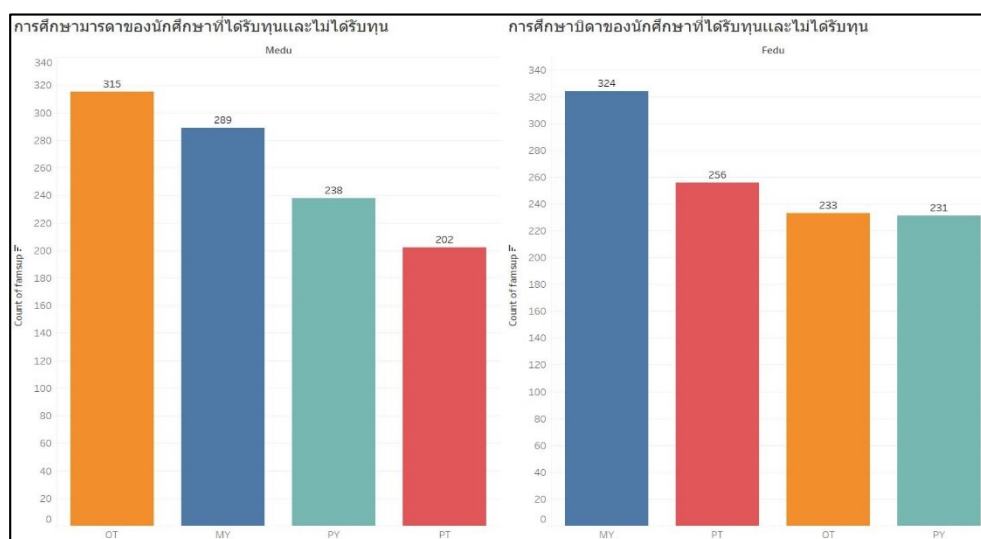
ภาพที่ 3.16 แสดงข้อมูลแสดงผลการได้รับทุนการศึกษา และไม่ได้รับทุนการศึกษาทั้งหมด

8) แสดงข้อมูลแสดงผลผู้ปกครองของนักศึกษาที่ได้รับทุนและไม่ได้รับทุน ซึ่งเป็นการแสดงผลในรูปแบบแผนภูมิแท่ง ผู้ใช้สามารถเลือกดูแต่ละประเภทได้



ภาพที่ 3.17 แสดงข้อมูลแสดงผลผู้ปกครองของนักศึกษาที่ได้รับทุนและไม่ได้รับทุน

9) แสดงข้อมูลแสดงผลประเภทของการศึกษาบิดามารดาของนักศึกษาที่ได้รับทุนและไม่ได้รับทุนเรียงตามลำดับ ซึ่งเป็นการแสดงผลในรูปแบบแผนภูมิแท่ง ผู้ใช้สามารถเลือกดูแต่ละอาชีพของพ่อแม่ได้



ภาพที่ 3.18 แสดงข้อมูลแสดงผลประเภทของการศึกษาบิดามารดาของนักศึกษาที่ได้รับทุนและไม่ได้รับทุน

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 with a dark blue sidebar on the left. The sidebar contains several navigation items, each with an 'ICON' and a Thai label: 'หน้าแรก' (Home), 'ข้อมูล' (Data), 'ข้อมูลสารสนเทศ' (Information Data), 'กระบวนการ' (Process), 'ผลจาประเมินผล' (Evaluation Results), and 'การประเมินผลสัมฤทธิ์' (Effectiveness Evaluation). The main content area is white and features a search bar at the top with the text 'Search ...'. Below the search bar, there are two buttons: 'All Information xx List' and 'Show information xx To xx'. The central part of the interface is a large table with a dark blue header and multiple empty rows. At the bottom of the main content area, there is a dark grey section labeled 'CONTACT' with three sets of horizontal lines representing contact information.

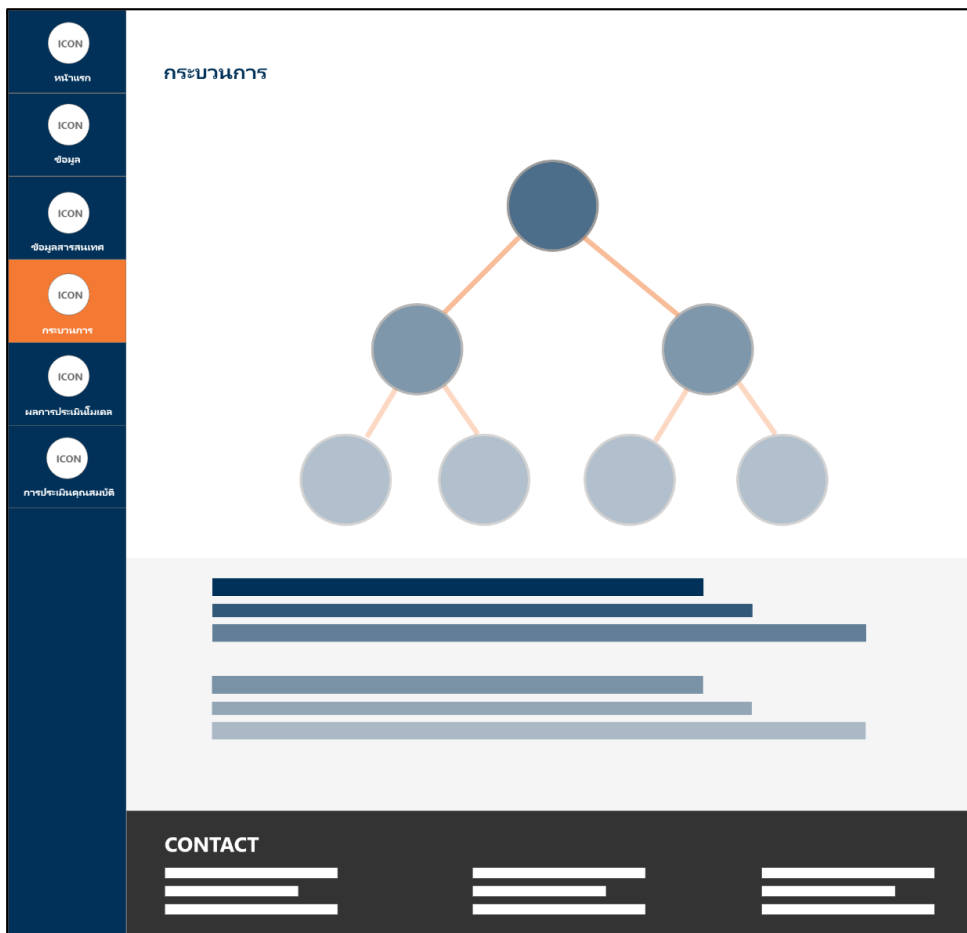
ภาพที่ 3.20 หน้าเว็บเพจที่ 2 ชุดข้อมูล

3) หน้าเว็บเพจที่ 3 ชุดข้อมูลสารสนเทศ

The screenshot shows a web application interface. On the left is a dark blue sidebar with a vertical list of navigation items, each with an 'ICON' and a label: 'หน้าแรก', 'ข้อมูล', 'ข้อมูลสารสนเทศ' (highlighted in orange), 'กระบวนการ', 'ผลการประเมินผล', and 'การประเมินคุณสมบัตินี้'. The main content area has a search bar at the top with the text 'Search ...'. Below the search bar, there are two buttons: 'All Information xx List' and 'Show information xx To xx'. The central part of the page is a table with a single orange header row and approximately 15 empty white rows. At the bottom of the page is a dark grey footer section titled 'CONTACT' with three sets of horizontal lines representing contact information.

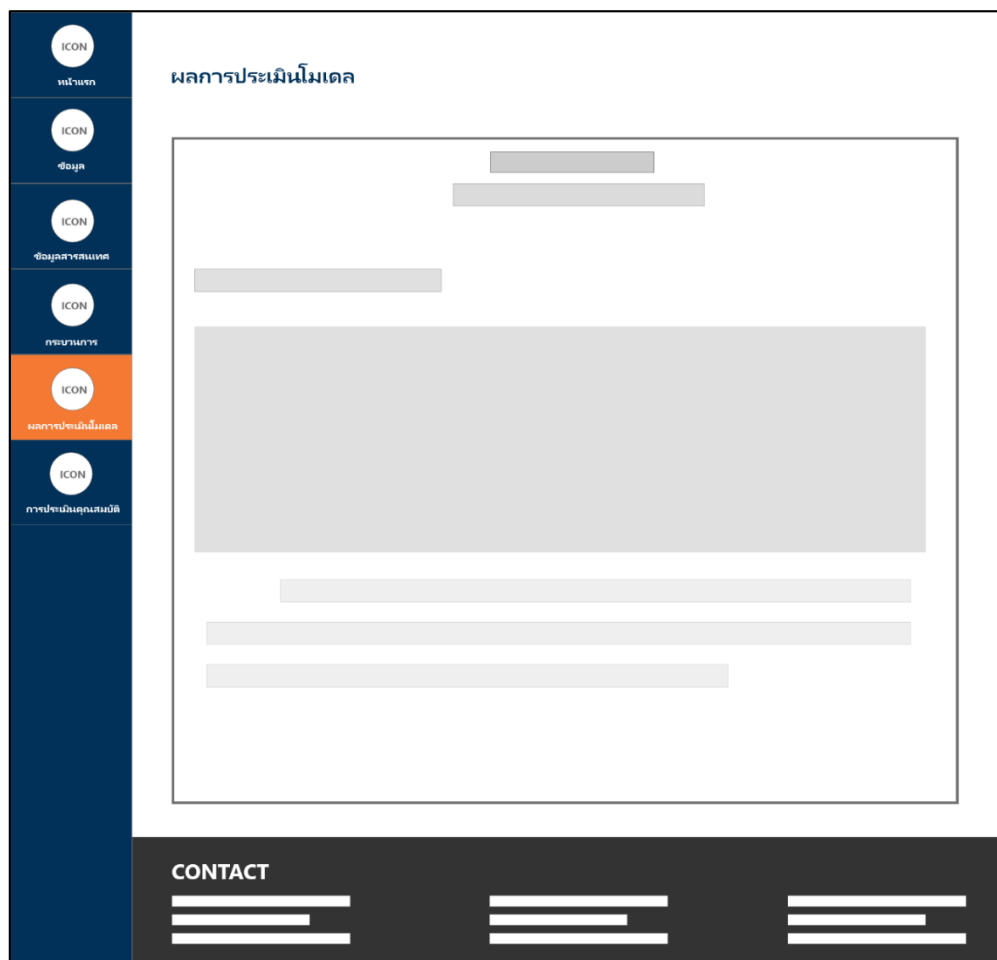
ภาพที่ 3.21 หน้าเว็บเพจที่ 3 ชุดข้อมูลสารสนเทศ

4) หน้าเว็บเพจที่ 4 Classification Model แสดงถึงขั้นตอนทำงาน



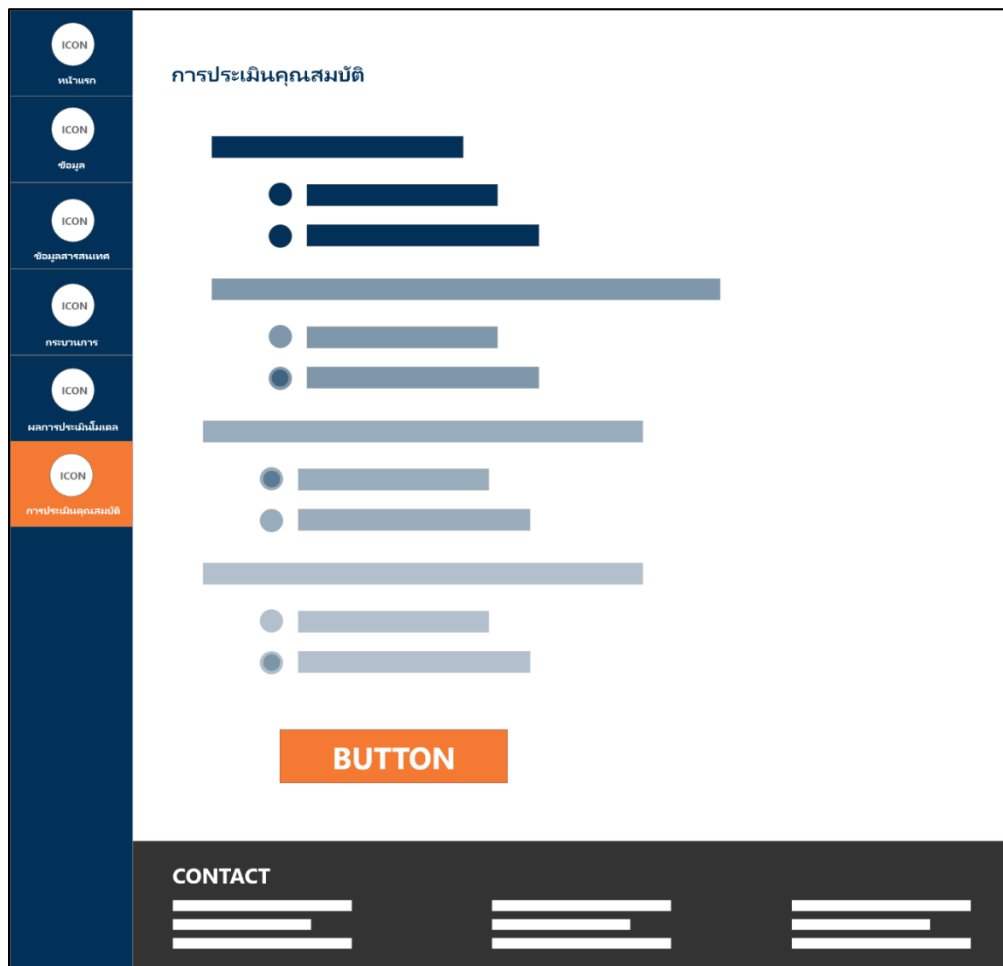
ภาพที่ 3.22 หน้าเว็บเพจที่ 4 Classification Model

6) หน้าเว็บเพจที่ 5 แสดงหน้าผลการประเมินโมเดล



ภาพที่ 3.23 หน้าเว็บเพจที่ 5 หน้าผลการประเมินโมเดล

6) หน้าเว็บเพจที่ 5 แสดงหน้าการประเมินคุณสมบัติ



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