

บทที่ 3

วิธีการดำเนินงานโครงการ

โครงการเรื่อง การวิเคราะห์ข้อมูลทางโภชนาการที่มีผลต่อเกรดเฉลี่ยของนักศึกษา เพื่อใช้สำหรับเผยแพร่ข้อมูลบนเว็บไซต์ ซึ่งในบทนี้จะเป็นการวิเคราะห์ข้อมูลเทคนิค Data Mining เมื่อเสร็จสิ้นกระบวนการวิเคราะห์แล้วจะเป็นขั้นตอนการออกแบบเว็บไซต์ รูปแบบการแสดงผลและบทสรุปจากวิธีการดำเนินงาน ดังนี้

- 3.1 การวิเคราะห์ข้อมูลด้วย CRISP-DM
- 3.2 การออกแบบเว็บไซต์
- 3.3 บทสรุป

3.1 การวิเคราะห์ข้อมูลด้วย CRISP-DM

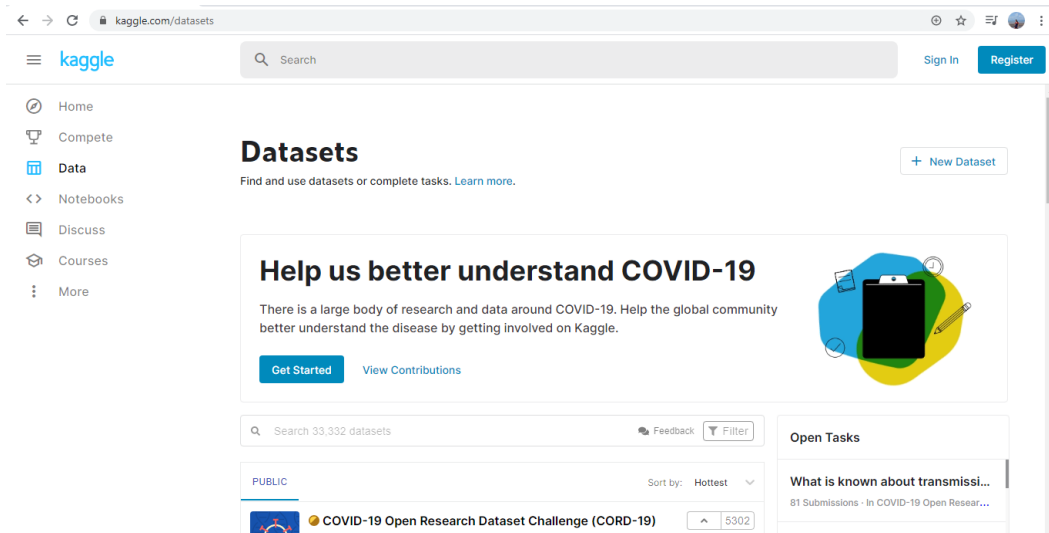
งานวิทยาการด้านข้อมูล (Data Science) ซึ่งกำลังมีบทบาทในยุคปัจจุบัน และทวีความสำคัญยิ่งขึ้นในอนาคต ได้มี CRISP-DM เป็นกระบวนการหลักในการจัดทำเหมืองข้อมูลเพื่อการวิเคราะห์และใช้ประโยชน์ในทางธุรกิจ กระบวนการวิเคราะห์ข้อมูล ด้วย CRISP-DM หรือ Cross Industry Standard Process for Data Mining พัฒนาขึ้นในปี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

3.1.1 รู้จักและเข้าใจในธุรกิจ (Business understanding) เป็นขั้นตอนแรกของกระบวนการ ที่มุ่งเน้นไปที่การทำความเข้าใจกระบวนการทางธุรกิจโดยรวม

ผู้วิเคราะห์ข้อมูลทำความเข้าใจกับปัญหาให้อยู่ในรูปของการวิเคราะห์ข้อมูลทาง Data Mining โดยการวิเคราะห์ข้อมูลทางโภชนาการที่มีผลต่อเกรดเฉลี่ยของนักศึกษา ซึ่งมีข้อมูลที่ซับซ้อนแต่สามารถทำความเข้าใจในข้อมูลได้

3.1.2 สร้างฐานข้อมูลให้ครบ (Data understanding) ขั้นตอนการจัดเก็บและรวบรวมข้อมูล ตลอดจนการพิจารณาตรวจสอบความถูกต้องของข้อมูลที่ได้รับ โดยเลือกว่าจะใช้ข้อมูลทั้งหมดหรือบางส่วนในการวิเคราะห์ให้สอดคล้องกับวัตถุประสงค์ที่กำหนดไว้

ผู้วิเคราะห์ข้อมูลได้ทำการเก็บรวบรวมข้อมูลจากเว็บไซต์ kaggle.com จากนั้นจะทำการตรวจสอบข้อมูลที่ได้ทำการรวบรวมมาได้เพื่อดูความถูกต้องของข้อมูล และพิจารณาว่าข้อมูลทางโภชนาการที่มีผลต่อเกรดเฉลี่ยของนักศึกษา จะใช้ข้อมูลทั้งหมดหรือจำเป็นต้องเลือกข้อมูลบางส่วนมาใช้ในการวิเคราะห์



ภาพที่ 3.1 เว็บไซต์ kaggle.com

ซึ่งข้อมูลการรับประทานอาหารของนักศึกษา มีจำนวน 126 รายการ รายการ ประกอบด้วย 61 แอตทริบิวต์ ข้อมูลหลัก ๆ จะประกอบด้วย เกรด เพศ แคลลอรี่ กาแฟ อาหารมื้อสะดวก อาหารในวัยเด็ก กีฬาที่เล่น การออกกำลังกาย เป็นต้น

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R					
1	GPA	Gender	breakfast	bries	chickalories	daylories	scor	coffee	omfort	foort	food	refood	reaso	cook	food_reaso	cuisine	let_current	current_cc	drink	ting	chang	changes	ch
2	2.4	2	1	430	nan	315	1	none	we dont	he	9	2	9	nan	eat good e	1	1	eat faster					
3	3.654	1	1	610	3	420	2	chocolate	Stress, bor		1	2	1	1	eat abou	2	2	eat out r					
4	3.3	1	1	720	4	420	2	frozen yog	stress, sac		1	1	1	3	toast and	3	1	sometimes					
5	3.2	1	1	430	3	420	2	Pizza, Mac	Boredom		2	2	2	2	College die	2	2	Accepting					
6	3.5	1	1	720	2	420	2	Ice cream	Stress, bor		1	1	1	2	I try to ea	2	2	I have eat					
7	2.25	1	1	610	3	980	2	Candy, br	cNone, i do		4	3	4	nan	My current	2	2	Eating rice					
8	3.8	2	1	610	3	420	2	Chocolate	, stress, bor		1	2	1	1	I eat a lot	3	1	I started e					
9	3.3	1	1	720	3	420	1	Ice cream	I eat comf		1	3	1	1	I eat a ver	1	2	Freshmen					
10	3.3	1	1	430	nan	420	1	Donuts, ki	Boredom		2	3	2	1	I eat what	1	1	snack les					
11	3.3	1	1	430	3	315	2	Mac and c	Stress, an		1	3	1	1	I eat heal	1	1	I cook a lo					
12	3.5	1	1	610	3	980	2	Pasta, gra	Boredom		2	1	2	1	I eat very l	1	2	Nun					
13	3.904	1	1	720	4	420	2	chocolate	, sadness, s		3	3	3	1	I am very	1	1	Less meat					
14	3.4	2	1	430	3	420	2	Cookies, pi	Sadness, t		3	5	3	1	I focus mc	1	2	I have bee					
15	3.6	1	1	610	3	420	2	ice cream	, stress, bo		1	2	1	1	Not as hea	2	2	not as hea					
16	3.1	2	1	610	3	420	2	Pizza, fru	Friends, er		2	3	2	1	Making sur	1	2	I knew I w					
17	nan	2	2	430	nan	980	2	cookies, di	boredom		2	4	2	1	I like a lot	2	2	none					
18	4	1	1	265	3	420	1	Salsfish, c	Stress		1	3	1	nan	I eat very	1	1	More Water					
19	3.6	2	1	430	3	980	2	chips, cook	I usually oi		2	3	2	1	My current	3	2	I would sa					
20	3.4	1	1	720	3	980	1	Chocolate	, Sadness, s		3	3	3	1	I eat lots c	1	1	I ate at th					
21	2.2	2	1	430	2	420	2	pizza, wing	boredom,		2	4	2	1	Current die	2	2	None really					
22	3.3	2	1	610	3	980	2	Fast food	, happiness,		7	5	7	1	I eat 2 me	2	2	Late night					
23	3.87	2	1	610	3	315	1	chocolate	, Mostly bor		2	3	2	3	Random. h	2	nan	Less mone					
24	3.7	2	1	610	3	420	1	burgers, cl	sadness, c		3	5	3	1	got balanced	1	1	got worse					

ภาพที่ 3.2 การวิเคราะห์ข้อมูลทางโภชนาการที่มีผลต่อเกรดเฉลี่ยของนักศึกษา

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

3.1.3.1 ทำการคัดเลือกข้อมูล (Data Selection) คือการคัดเลือกข้อมูลที่เหมาะสมเพื่อนำมาใช้ในการวิเคราะห์ข้อมูล

ผู้วิเคราะห์ข้อมูลทำการคัดเลือกข้อมูล และทำการ Data Cleaning ข้อมูลทางโภชนาการที่มีผลต่อเกรดเฉลี่ยของนักศึกษา โดยแยกข้อมูลออกและตัดส่วนที่ไม่จำเป็นออกให้เหลือเฉพาะข้อมูลที่จำเป็นในการวิเคราะห์ในภาพรวม จำนวน 8 แอดทริบิวท์ ได้แก่ เกรด เพศ อาหารเข้ากาแฟ การดื่ม อาหารมีอสังควก อาหารในวัยเด็ก การออกกำลังกาย จำนวน 391 รายการ ซึ่งเป็นข้อมูลที่จำเป็นในการนำไปวิเคราะห์ข้อมูล

	A	B	C	D	E	F	G	H
1	GPA	Gender	breakfast	coffee	drink	comfort_food	food_childhood	sports
2	C	Male	Cereal	Creamy	Orange Juice	Pizza	Rice	Yes
3	C	Male	Cereal	Creamy	Orange Juice	Chips	Chicken	Yes
4	B+	Female	Cereal	Espresso	Soda	Chocolate	Biscuits	Yes
5	B+	Female	Cereal	Espresso	Soda	Chips	Beef soup	Yes
6	B+	Female	Cereal	Espresso	Soda	Ice cream	Baked beans	Yes
7	B	Female	Cereal	Espresso	Orange Juice	Frozen yogurt	Mac and Cheese	No
8	B	Female	Cereal	Espresso	Orange Juice	Pizza	Pizza	No
9	B	Female	Cereal	Espresso	Orange Juice	Fast food	Tacos	No
10	B	Female	Cereal	Espresso	Soda	Pizza	Beef stroganoff	No
11	B	Female	Cereal	Espresso	Soda	Mac and Cheese	Tacos	No
12	B	Female	Cereal	Espresso	Soda	Ice cream	Pizza	No
13	B	Female	Cereal	Espresso	Soda	Ice cream	Pasta	Yes
14	B+	Female	Cereal	Espresso	Soda	Chocolate	Chicken fingers	Yes
15	B+	Female	Cereal	Espresso	Soda	Chips	Pizza	Yes
16	C	Female	Cereal	Espresso	Soda	Candy	French Fries	No
17	C	Female	Cereal	Espresso	Soda	Brownies	Plantain	No
18	B+	Male	Cereal	Espresso	Orange Juice	Chocolate	Grilled Cheese	Yes
19	B+	Male	Cereal	Espresso	Orange Juice	Ice cream	Burgers	Yes
20	B+	Male	Cereal	Espresso	Orange Juice	French Fries	Mac and Cheese	Yes
21	B	Female	Cereal	Creamy	Soda	Ice cream	chicken	No
22	B	Female	Cereal	Creamy	Soda	Cheeseburgers	Cheesy potatoes	No
23	B	Female	cereal	Creamy	Soda	Chips	Hot Dogs	No

ภาพที่ 3.3 ข้อมูลทางโภชนาการที่มีผลต่อเกรดเฉลี่ยของนักศึกษา

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

1) ข้อมูลทางโภชนาการที่มีผลต่อเกรดเฉลี่ยของนักศึกษา ผู้วิเคราะห์ข้อมูลได้ทำการแก้ไขข้อมูล ซึ่งผู้วิเคราะห์ข้อมูลพบว่า บางข้อมูลต้องใส่ข้อมูลแทนค่าที่มีในข้อมูล เพื่อให้ข้อมูลชัดเจนและสมบูรณ์มากยิ่งขึ้น ดังนั้นผู้วิเคราะห์ข้อมูลได้ทำการแทนที่ข้อมูล ดังนี้

- GPA ซึ่งจะเป็นข้อมูลตัวเลข ผู้วิเคราะห์พบว่าควรทำการแก้ไขเกรดให้เป็นตัวอักษรตามระดับเกรด A คือ 4.00 B+ คือ 3.99–3.50 B คือ 3.49–3.00 C+ คือ 2.99–2.50 C คือ 2.49–2.00 D+ คือ 1.99–1.50 และ D คือ 1.49–1.00 และได้ทำการแทนค่าดังนี้

GPA	GPA
2.4	C
2.4	C
3.6	B+
3.6	B+
3.6	B+
3.3	B
3.3	B
3.3	B
3.2	B
3.2	B
3.2	B

ภาพที่ 3.4 แกะไขข้อมูลแทนค่าของ GPA

- Gender ข้อมูลเพศ เลข 1 แทนด้วยค่าของ Female(เพศหญิง) เลข 2 แทนค่าของ Male(เพศชาย) ผู้วิเคราะห์จึงได้ทำการแทนค่าดังนี้

Gender	Gender
2	Male
2	Male
1	Female
1	Female
1	Female
1	Female
1	Female
1	Female
1	Female

ภาพที่ 3.5 แกะไขข้อมูลแทนค่าของ Gender

- Breakfast ข้อมูลอาหารเช้า เลข 1 แทนค่าด้วย Cereal เลข 2 แทนค่าด้วย Donut ผู้วิเคราะห์จึงได้ทำการแทนค่าดังนี้

breakfast	breakfast
1	Cereal
1	Cereal
1	Cereal
2	Donut
1	Cereal
2	Donut
2	Donut
1	Cereal
1	Cereal
2	Donut
1	Cereal

ภาพที่ 3.6 แกะไขข้อมูลแทนค่าของ Breakfast

- Coffee ข้อมูลกาแฟ เลข 1 แทนค่าด้วย Creamy เลข 2 แทนค่าด้วย Espresso ผู้วิเคราะห์จึงได้ทำการแทนค่าดังนี้

coffee	coffee
1	Creamy
1	Creamy
2	Espresso
2	Espresso
2	Espresso
2	Espresso
2	Espresso
1	Creamy
1	Creamy
2	Espresso
2	Espresso

ภาพที่ 3.7 แกไขข้อมูลแทนค่าของ Coffee

- Drink ข้อมูลการดื่ม เลข 1 แทนค่าด้วย Orange Juice เลข 2 แทนค่าด้วย Soda ผู้วิเคราะห์จึงได้ทำการแทนค่าดังนี้

drink	drink
1	Orange Juice
1	Orange Juice
2	Soda
2	Soda
2	Soda
1	Orange Juice
1	Orange Juice
1	Orange Juice
2	Soda
2	Soda

ภาพที่ 3.8 แกไขข้อมูลแทนค่าของ Drink

- Sports ข้อมูลการเล่นกีฬา เลข 1 แทนค่าด้วย Yes(ออกกำลังกาย) เลข 2 แทนค่าด้วย No(ไม่ออกกำลังกาย) ผู้วิเคราะห์จึงได้ทำการแทนค่าดังนี้

sports	sports
1	Yes
1	Yes
1	Yes
1	Yes
1	Yes
2	No
2	No
2	No
2	No

ภาพที่ 3.9 แกไขข้อมูลแทนค่าของ Sports

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

ผู้วิเคราะห์ข้อมูลวิเคราะห์ข้อมูลด้วยการทำเหมืองข้อมูลแบบ Data Classification เพื่อใช้ทำนายแนวโน้มการเกิดขึ้นของปัจจัยที่ก่อให้เกิดผู้บาดเจ็บและเสียชีวิตจากเทคนิคการจำแนกประเภทข้อมูล ด้วยการสร้างโมเดล Decision Tree เพื่อจัดกลุ่มข้อมูลที่มีความสัมพันธ์กัน โดยใช้โปรแกรมที่ใช้ทำเหมืองข้อมูล ด้วยชุดข้อมูลที่คัดเลือก ดังนี้

GPA	Gender	breakfast	coffee	drink
C	Male	Cereal	Creamy	Orange Juice
C	Male	Cereal	Creamy	Orange Juice
B+	Female	Cereal	Creamy	Soda
B+	Female	Cereal	Creamy	Soda
B+	Female	Cereal	Creamy	Soda
B	Female	Cereal	Espresso	Orange Juice
B	Female	Cereal	Espresso	Orange Juice
B	Female	Cereal	Espresso	Orange Juice
B	Female	Cereal	Espresso	Soda
B	Female	Cereal	Espresso	Soda
B	Female	Cereal	Espresso	Soda
B+	Female	Cereal	Creamy	Soda
B+	Female	Cereal	Creamy	Soda
B+	Female	Cereal	Creamy	Soda
C	Female	Cereal	Espresso	Soda
C	Female	Cereal	Espresso	Soda
B+	Male	Cereal	Creamy	Soda

ภาพที่ 3.10 ชุดข้อมูลที่คัดเลือกมาวิเคราะห์ข้อมูล

จากรูปภาพที่ ประกอบด้วย 5 แอตทริบิวต์ คือ

- GPA ประกอบด้วย A B+ B C+ C D+ และD
- Gender ประกอบด้วย Female(เพศหญิง) และMale(เพศชาย)
- Breakfast ประกอบด้วย Cereal และDonut
- Coffee ประกอบด้วย Creamy และEspresso
- Drink ประกอบด้วย Orange Juice และ Soda

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

$$IG(\text{parent}, \text{child}) = \text{entropy}(\text{parent}) - [p(c1) \times \text{entropy}(c1) + p(c2) \times \text{entropy}(c2) + \dots]$$

โดยที่ $\text{entropy}(c1) = -p(c1) \log p(c1)$ และ $p(c2)$ คือ ค่าความน่าจะเป็นของ $c1$

การคำนวณค่าแต่ละแอตทริบิวต์เทียบกับคลาสเพื่อหาแอตทริบิวต์ที่มีค่า IG มากที่สุดมาเป็น Root ของ Decision tree กับจำนวนข้อมูลทั้งหมดโดยใช้ผลลัพธ์เป็น GPA ดังนี้

1) คำนวณค่า IG ของแอตทริบิวต์ Gender จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

$$\begin{aligned} \text{Entropy (parent)} &= -p(A) \times \log_2 p(A) + p(B+) \times \log_2 p(B+) + p(B) \times \log_2 p(B) \\ &+ p(C+) \times \log_2 p(C+) + p(C) \times \log_2 p(C) \\ &= -[0.03 \times \log_2(0.03) + 0.50 \times \log_2(0.50) + 0.38 \times \\ &\log_2(0.38) + 0.07 \times \log_2(0.07) + 0.02 \times \log_2(0.02)] \\ &= -[0.03 \times -5.06 + 0.50 \times -1 + 0.38 \times -1.40 + 0.07 \times - \\ &3.84 + 0.02 \times -5.64] \end{aligned}$$

$$\begin{aligned} &= -[0.15 + 0.5 + 0.53 + 0.27 + 0.11] \\ &= 1.56 \end{aligned}$$

$$\begin{aligned} \text{Entropy (Female)} &= -p(A) \times \log_2 p(A) + p(B+) \times \log_2 p(B+) + p(B) \times \log_2 p(B) \\ &+ p(C+) \times \log_2 p(C+) + p(C) \times \log_2 p(C) \\ &= -[0.04 \times \log_2(0.04) + 0.51 \times \log_2(0.51) + 0.36 \times \\ &\log_2(0.36) + 0.08 \times \log_2(0.08) + 0.01 \times \log_2(0.01)] \\ &= -[0.04 \times -4.64 + 0.51 \times -0.97 + 0.36 \times -1.47 + 0.08 \\ &\times -3.64 + 0.01 \times -6.64] \end{aligned}$$

$$\begin{aligned} &= -[0.19 + 0.49 + 0.53 + 0.29 + 0.07] \\ &= 1.57 \end{aligned}$$

$$\begin{aligned} \text{Entropy (Male)} &= -p(A) \times \log_2 p(A) + p(B+) \times \log_2 p(B+) + p(B) \times \log_2 p(B) \\ &+ p(C+) \times \log_2 p(C+) + p(C) \times \log_2 p(C) \\ &= -[0.01 \times \log_2(0.01) + 0.48 \times \log_2(0.48) + 0.42 \times \\ &\log_2(0.42) + 0.07 \times \log_2(0.07) + 0.03 \times \log_2(0.03)] \\ &= -[0.01 \times -6.64 + 0.48 \times -1.06 + 0.42 \times -1.25 + 0.07 \times \\ &-3.84 + 0.03 \times -5.06] \end{aligned}$$

$$\begin{aligned} &= -[0.07 + 0.51 + 0.52 + 0.27 + 0.15] \\ &= 1.52 \end{aligned}$$

$$\begin{aligned} \text{IG (parent, child)} &= \text{entropy (parent)} - [p(\text{Female}) \times \text{entropy}(\text{Female}) + p(\text{Male}) \times \\ &\text{entropy}(\text{Male})] \end{aligned}$$

$$\begin{aligned} &= 1.56 - [0.62 \times 1.57 + 0.38 \times 1.52] \\ &= 1.56 - [0.97 + 0.58] \end{aligned}$$

$$= 1.56 - 1.55$$

$$= 0.01$$

2) คำนวณค่า IG ของแอตทริบิวต์ Breakfast จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

$$\begin{aligned} \text{entropy (parent)} &= -p(A) \times \log_2 p(A) + p(B+) \times \log_2 p(B+) + p(B) \times \log_2 p(B) \\ &+ p(C+) \times \log_2 p(C+) + p(C) \times \log_2 p(C) \end{aligned}$$

$$= -[0.03 \times \log_2(0.03) + 0.50 \times \log_2(0.50) + 0.38 \times \log_2(0.38) + 0.07 \times \log_2(0.07) + 0.02 \times \log_2(0.02)]$$

$$= -[0.03 \times -5.06 + 0.50 \times -1 + 0.38 \times -1.40 + 0.07 \times -3.84 + 0.02 \times -5.64]$$

$$= -[0.15 + 0.5 + 0.53 + 0.27 + 0.11]$$

$$= 1.56$$

$$\begin{aligned} \text{Entropy (Cereal)} &= -p(A) \times \log_2 p(A) + p(B+) \times \log_2 p(B+) + p(B) \times \log_2 p(B) \\ &+ p(C+) \times \log_2 p(C+) + p(C) \times \log_2 p(C) \end{aligned}$$

$$= -[0.03 \times \log_2(0.03) + 0.50 \times \log_2(0.50) + 0.38 \times \log_2(0.38) + 0.07 \times \log_2(0.07) + 0.02 \times \log_2(0.02)]$$

$$= -[0.03 \times -5.06 + 0.50 \times -1 + 0.38 \times -1.40 + 0.07 \times -3.84 + 0.02 \times -5.64]$$

$$= -[0.15 + 0.5 + 0.53 + 0.27 + 0.11]$$

$$= 1.56$$

$$\begin{aligned} \text{Entropy (Donut)} &= -p(A) \times \log_2 p(A) + p(B+) \times \log_2 p(B+) + p(B) \times \log_2 p(B) \\ &+ p(C+) \times \log_2 p(C+) + p(C) \times \log_2 p(C) \end{aligned}$$

$$= -[0 \times \log_2(0) + 0.48 \times \log_2(0.48) + 0.41 \times \log_2(0.41) + 0.11 \times \log_2(0.11) + 0 \times \log_2(0)]$$

$$= -[0 + 0.48 \times -1.06 + 0.41 \times -1.29 + 0.11 \times -3.18 + 0]$$

$$= -[0 + 0.51 + 0.53 + 0.35 + 0]$$

$$= 1.39$$

$$\begin{aligned} \text{IG (parent, child)} &= \text{entropy (parent)} - [p(\text{Cereal}) \times \text{entropy}(\text{Cereal}) + p(\text{Donut}) \\ &\times \text{entropy}(\text{Donut})] \end{aligned}$$

$$= 1.56 - [0.89 \times 1.56 + 0.11 \times 1.39]$$

$$= 1.56 - [1.39 + 0.15]$$

$$= 1.56 - 1.54$$

$$= 0.02$$

3) คำนวณค่า IG ของแอตทริบิวต์ Coffee จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

$$\begin{aligned} \text{entropy (parent)} &= -p(A) \times \log_2 p(A) + p(B+) \times \log_2 p(B+) + p(B) \times \log_2 p(B) \\ &+ p(C+) \times \log_2 p(C+) + p(C) \times \log_2 p(C)] \\ &= -[0.03 \times \log_2(0.03) + 0.50 \times \log_2(0.50) + 0.38 \times \\ &\log_2(0.38) + 0.07 \times \log_2(0.07) + 0.02 \times \log_2(0.02)] \\ &= -[0.03 \times -5.06 + 0.50 \times -1 + 0.38 \times -1.40 + 0.07 \times - \\ &3.84 + 0.02 \times -5.64] \end{aligned}$$

$$= -[0.15 + 0.5 + 0.53 + 0.27 + 0.11]$$

$$= 1.56$$

$$\begin{aligned} \text{Entropy (Creamy)} &= -p(A) \times \log_2 p(A) + p(B+) \times \log_2 p(B+) + p(B) \times \log_2 p(B) \\ &+ p(C+) \times \log_2 p(C+) + p(C) \times \log_2 p(C)] \\ &= -[0.02 \times \log_2(0.02) + 0.45 \times \log_2(0.45) + 0.43 \times \\ &\log_2(0.43) + 0.08 \times \log_2(0.08) + 0.02 \times \log_2(0.02)] \\ &= -[0.02 \times -5.64 + 0.45 \times -1.15 + 0.43 \times -1.22 + 0.08 \times \\ &-3.64 + 0.02 \times -5.64] \end{aligned}$$

$$= -[0.11 + 0.52 + 0.52 + 0.29 + 0.11]$$

$$= 1.55$$

$$\begin{aligned} \text{Entropy (Espresso)} &= -p(A) \times \log_2 p(A) + p(B+) \times \log_2 p(B+) + p(B) \times \log_2 p(B) \\ &+ p(C+) \times \log_2 p(C+) + p(C) \times \log_2 p(C)] \\ &= -[0.03 \times \log_2(0.03) + 0.52 \times \log_2(0.52) + 0.36 \times \\ &\log_2(0.36) + 0.07 \times \log_2(0.07) + 0.02 \times \log_2(0.02)] \\ &= -[0.03 \times -5.06 + 0.52 \times -0.94 + 0.36 \times -1.47 + 0.07 \\ &\times -3.84 + 0.02 \times -5.64] \end{aligned}$$

$$= -[0.15 + 0.49 + 0.53 + 0.27 + 0.11]$$

$$= 1.55$$

$$\begin{aligned} \text{IG (parent, child)} &= \text{entropy (parent)} - [p(\text{Creamy}) \times \text{entropy}(\text{Creamy}) + \\ &p(\text{Espresso}) \times \text{entropy}(\text{Espresso})] \end{aligned}$$

$$= 1.56 - [0.33 \times 1.55 + 0.67 \times 1.55]$$

$$= 1.56 - [0.51 + 1.04]$$

$$= 1.56 - 1.55$$

$$= 0.01$$

4) คำนวณค่า IG ของแอตทริบิวต์ Drink จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

$$\begin{aligned} \text{entropy (parent)} &= -p(A) \times \log_2 p(A) + p(B+) \times \log_2 p(B+) + p(B) \times \log_2 p(B) \\ &+ p(C+) \times \log_2 p(C+) + p(C) \times \log_2 p(C)] \\ &= -[0.03 \times \log_2(0.03) + 0.50 \times \log_2(0.50) + 0.38 \times \\ &\log_2(0.38) + 0.07 \times \log_2(0.07) + 0.02 \times \log_2(0.02)] \\ &= -[0.03 \times -5.06 + 0.50 \times -1 + 0.38 \times -1.40 + 0.07 \times - \\ &3.84 + 0.02 \times -5.64] \\ &= -[0.15 + 0.5 + 0.53 + 0.27 + 0.11] \\ &= 1.56 \end{aligned}$$

$$\begin{aligned} \text{Entropy (Orange Juice)} &= -p(A) \times \log_2 p(A) + p(B+) \times \log_2 p(B+) + p(B) \times \log_2 p(B) \\ &+ p(C+) \times \log_2 p(C+) + p(C) \times \log_2 p(C)] \\ &= -[0.06 \times \log_2(0.06) + 0.42 \times \log_2(0.42) + 0.38 \times \\ &\log_2(0.38) + 0.12 \times \log_2(0.12) + 0.01 \times \log_2(0.01)] \\ &= -[0.06 \times -4.06 + 0.42 \times -1.25 + 0.38 \times -1.40 + 0.12 \times \\ &-3.06 + 0.01 \times -6.64] \\ &= -[0.24 + 0.53 + 0.53 + 0.37 + 0.07] \\ &= 1.74 \end{aligned}$$

$$\begin{aligned} \text{Entropy (Soda)} &= -p(A) \times \log_2 p(A) + p(B+) \times \log_2 p(B+) + p(B) \times \log_2 p(B) \\ &+ p(C+) \times \log_2 p(C+) + p(C) \times \log_2 p(C)] \\ &= -[0.01 \times \log_2(0.01) + 0.55 \times \log_2(0.55) + 0.38 \times \\ &\log_2(0.38) + 0.04 \times \log_2(0.04) + 0.02 \times \log_2(0.02)] \\ &= -[0.01 \times -6.64 + 0.55 \times -0.86 + 0.38 \times -1.40 + 0.04 \\ &\times -4.64 + 0.02 \times -5.64] \\ &= -[0.07 + 0.47 + 0.53 + 0.19 + 0.11] \\ &= 1.37 \end{aligned}$$

$$\begin{aligned} \text{IG(parent, child)} &= \text{entropy (parent)} - [p(\text{Orange Juice}) \times \text{entropy(Orange} \\ &\text{Juice)} + p(\text{Soda}) \times \text{entropy(Soda)}] \\ &= 1.56 - [0.40 \times 1.74 + 0.60 \times 1.37] \\ &= 1.56 - [0.70 + 0.82] \\ &= 1.56 - 1.52 \end{aligned}$$

$$= 0.04$$

จากการคำนวณค่า IG ของทุกแอตทริบิวต์พบว่าค่า IG ของแอตทริบิวต์ Drink มีค่ามากที่สุด (0.04) ดังนั้นจึงเลือกแอตทริบิวต์ Drink ขึ้นมาเป็นโหนด Root และจะต้องทำการแตกกิ่งจาก โหนด root ออกไปจนข้อมูลในแต่ละโหนดมีคลาสคำตอบเดียวกัน จึงทำการสร้างโหนดในระดับถัดไปของแอตทริบิวต์ Drink

การคำนวณค่าแต่ละแอตทริบิวต์ในระดับที่ 2 กับจำนวนข้อมูลทั้งหมดโดยใช้ผลลัพธ์เป็น GPA

1) คำนวณค่า IG ของแอตทริบิวต์ Drink และ Gender ของ Orange Juice จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

$$\begin{aligned} \text{entropy (parent)} &= -p(A) \times \log_2 p(A) + p(B+) \times \log_2 p(B+) + p(B) \times \log_2 p(B) \\ &+ p(C+) \times \log_2 p(C+) + p(C) \times \log_2 p(C)] \\ &= -[0.06 \times \log_2(0.06) + 0.42 \times \log_2(0.42) + 0.38 \times \log_2(0.38) \\ &+ 0.12 \times \log_2(0.12) + 0.01 \times \log_2(0.01)] \\ &= -[0.06 \times -4.06 + 0.42 \times -1.25 + 0.38 \times -1.40 + 0.12 \times -3.06 \\ &+ 0.01 \times -6.64] \\ &= -[0.24 + 0.53 + 0.53 + 0.37 + 0.07] \\ &= 1.74 \end{aligned}$$

$$\begin{aligned} \text{Entropy (Female)} &= -p(A) \times \log_2 p(A) + p(B+) \times \log_2 p(B+) + p(B) \times \log_2 p(B) \\ &+ p(C+) \times \log_2 p(C+) + p(C) \times \log_2 p(C)] \\ &= -[0.09 \times \log_2(0.09) + 0.42 \times \log_2(0.42) + 0.35 \times \log_2(0.35) \\ &+ 0.15 \times \log_2(0.15) + 0 \times \log_2(0)] \\ &= -[0.09 \times -3.47 + 0.42 \times -1.25 + 0.35 \times -1.51 + 0.15 \times -2.74 + 0] \\ &= -[0.31 + 0.53 + 0.53 + 0.41 + 0] \\ &= 1.78 \end{aligned}$$

$$\begin{aligned} \text{Entropy (Male)} &= -p(A) \times \log_2 p(A) + p(B+) \times \log_2 p(B+) + p(B) \times \log_2 p(B) \\ &+ p(C+) \times \log_2 p(C+) + p(C) \times \log_2 p(C)] \\ &= -[0 \times \log_2(0) + 0.43 \times \log_2(0.43) + 0.47 \times \log_2(0.47) + 0.06 \times \log_2(0.06) \\ &+ 0.04 \times \log_2(0.04)] \\ &= -[0 + 0.43 \times -1.22 + 0.47 \times -1.09 + 0.06 \times -4.06 + 0.04 \times -4.64] \end{aligned}$$

$$= -[0 + 0.52 + 0.51 + 0.24 + 0.19]$$

$$= 1.46$$

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

$$= 1.74 - [0.70 \times 1.78 + 0.30 \times 1.46]$$

$$= 1.74 - [1.25 + 0.44]$$

$$= 1.74 - 1.69$$

$$= 0.05$$

1.1) คำนวณค่า IG ของแอตทริบิวต์ Drink และ Gender ของ Soda จากข้อมูลสามารถ คำนวณค่า IG ได้ดังนี้

$$\begin{aligned} \text{entropy (parent)} &= -p(A) \times \log_2 p(A) + p(B+) \times \log_2 p(B+) + p(B) \times \log_2 p(B) \\ &+ p(C+) \times \log_2 p(C+) + p(C) \times \log_2 p(C) \\ &= -[0.01 \times \log_2(0.01) + 0.55 \times \log_2(0.55) + 0.38 \times \log_2(0.38) \\ &+ 0.04 \times \log_2(0.04) + 0.02 \times \log_2(0.02)] \\ &= -[0.01 \times -6.64 + 0.55 \times -0.86 + 0.38 \times -1.40 + 0.04 \times -4.64 \\ &+ 0.02 \times -5.64] \\ &= -[0.07 + 0.47 + 0.53 + 0.19 + 0.11] \\ &= 1.37 \end{aligned}$$

$$\begin{aligned} \text{Entropy (Female)} &= -p(A) \times \log_2 p(A) + p(B+) \times \log_2 p(B+) + p(B) \times \log_2 p(B) \\ &+ p(C+) \times \log_2 p(C+) + p(C) \times \log_2 p(C) \\ &= -[0 \times \log_2(0) + 0.58 \times \log_2(0.58) + 0.38 \times \log_2(0.38) + 0.02 \times \log_2(0.02) \\ &+ 0.02 \times \log_2(0.02)] \\ &= -[0 + 0.58 \times -0.79 + 0.38 \times -1.40 + 0.02 \times -5.64 + 0.02 \times -5.64] \\ &= -[0 + 0.46 + 0.53 + 0.11 + 0.11] \\ &= 1.21 \end{aligned}$$

$$\begin{aligned} \text{Entropy (Male)} &= -p(A) \times \log_2 p(A) + p(B+) \times \log_2 p(B+) + p(B) \times \log_2 p(B) \\ &+ p(C+) \times \log_2 p(C+) + p(C) \times \log_2 p(C) \\ &= -[0.02 \times \log_2(0.02) + 0.50 \times \log_2(0.50) + 0.39 \times \log_2(0.39) \\ &+ 0.07 \times \log_2(0.07) + 0.02 \times \log_2(0.02)] \end{aligned}$$

$$= -[0.02 \times -5.64 + 0.50 \times -1 + 0.39 \times -1.36 + 0.07 \times -3.84 + 0.02 \times -5.64]$$

$$= -[0.11 + 0.5 + 0.53 + 0.27 + 0.11]$$

$$= 1.52$$

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

$$= 1.37 - [0.56 \times 1.21 + 0.44 \times 1.52]$$

$$= 1.37 - [0.68 + 0.67]$$

$$= 1.37 - 1.35$$

$$= 0.02$$

2) คำนวณค่า IG ของแอตทริบิวต์ Drink และ Breakfast ของ Orange Juice จากข้อมูล สามารถคำนวณค่า IG ได้ดังนี้

$$\text{entropy (parent)} = -p(A) \times \log_2 p(A) + p(B+) \times \log_2 p(B+) + p(B) \times \log_2 p(B) + p(C+) \times \log_2 p(C+) + p(C) \times \log_2 p(C)$$

$$= -[0.06 \times \log_2(0.06) + 0.42 \times \log_2(0.42) + 0.38 \times \log_2(0.38) + 0.12 \times \log_2(0.12) + 0.01 \times \log_2(0.01)]$$

$$= -[0.06 \times -4.06 + 0.42 \times -1.25 + 0.38 \times -1.40 + 0.12 \times -3.06 + 0.01 \times -6.64]$$

$$= -[0.24 + 0.53 + 0.53 + 0.37 + 0.07]$$

$$= 1.74$$

$$\text{Entropy (Cereal)} = -p(A) \times \log_2 p(A) + p(B+) \times \log_2 p(B+) + p(B) \times \log_2 p(B) + p(C+) \times \log_2 p(C+) + p(C) \times \log_2 p(C)$$

$$= -[0.07 \times \log_2(0.07) + 0.42 \times \log_2(0.42) + 0.39 \times \log_2(0.39) + 0.10 \times \log_2(0.10) + 0.01 \times \log_2(0.01)]$$

$$= -[0.07 \times -3.84 + 0.42 \times -1.25 + 0.39 \times -1.36 + 0.10 \times -3.32 + 0.01 \times -6.64]$$

$$= -[0.27 + 0.53 + 0.53 + 0.33 + 0.07]$$

$$= 1.73$$

$$\text{Entropy (Donut)} = -p(A) \times \log_2 p(A) + p(B+) \times \log_2 p(B+) + p(B) \times \log_2 p(B) + p(C+) \times \log_2 p(C+) + p(C) \times \log_2 p(C)$$

$$\begin{aligned}
&= -[0 \times \log_2(0) + 0.43 \times \log_2(0.43) + 0.29 \times \log_2(0.29) + \\
&0.29 \times \log_2(0.29) + 0 \times \log_2(0)] \\
&= -[0 + 0.43 \times -1.22 + 0.29 \times -1.79 + 0.29 \times -1.79 + 0] \\
&= -[0 + 0.52 + 0.52 + 0.52 + 0] \\
&= 1.56 \\
IG(\text{parent, child}) &= \text{entropy}(\text{parent}) - [p(\text{Cereal}) \times \text{entropy}(\text{Cereal}) + p(\text{Donut}) \times \\
&\text{entropy}(\text{Donut})] \\
&= 1.74 - [0.91 \times 1.73 + 0.09 \times 1.56] \\
&= 1.74 - [1.57 + 0.14] \\
&= 1.74 - 1.71 \\
&= 0.03
\end{aligned}$$

2.1) คำนวณค่า IG ของแอตทริบิวต์ Drink และ Breakfast ของ Soda จากข้อมูลสามารถ
คำนวณค่า IG ได้ดังนี้

$$\begin{aligned}
\text{entropy}(\text{parent}) &= -p(A) \times \log_2 p(A) + p(B+) \times \log_2 p(B+) + p(B) \times \log_2 p(B) \\
&+ p(C+) \times \log_2 p(C+) + p(C) \times \log_2 p(C)] \\
&= -[0.01 \times \log_2(0.01) + 0.55 \times \log_2(0.55) + 0.38 \times \\
&\log_2(0.38) + 0.04 \times \log_2(0.04) + 0.02 \times \log_2(0.02)] \\
&= -[0.01 \times -6.64 + 0.55 \times -0.86 + 0.38 \times -1.40 + 0.04 \\
&\times -4.64 + 0.02 \times -5.64] \\
&= -[0.07 + 0.47 + 0.53 + 0.19 + 0.11] \\
&= 1.37
\end{aligned}$$

$$\begin{aligned}
\text{Entropy}(\text{Cereal}) &= -p(A) \times \log_2 p(A) + p(B+) \times \log_2 p(B+) + p(B) \times \log_2 p(B) \\
&+ p(C+) \times \log_2 p(C+) + p(C) \times \log_2 p(C)] \\
&= -[0.01 \times \log_2(0.01) + 0.55 \times \log_2(0.55) + 0.37 \times \\
&\log_2(0.37) + 0.04 \times \log_2(0.04) + 0.02 \times \log_2(0.02)] \\
&= -[0.01 \times -6.64 + 0.55 \times -0.86 + 0.37 \times -1.43 + 0.04 \\
&\times -4.64 + 0.02 \times -5.64] \\
&= -[0.07 + 0.47 + 0.53 + 0.19 + 0.11] \\
&= 1.37
\end{aligned}$$

$$\begin{aligned}
\text{Entropy}(\text{Donut}) &= -p(A) \times \log_2 p(A) + p(B+) \times \log_2 p(B+) + p(B) \times \log_2 p(B) \\
&+ p(C+) \times \log_2 p(C+) + p(C) \times \log_2 p(C)]
\end{aligned}$$

$$\begin{aligned}
&= -[0 \times \log_2(0) + 0.50 \times \log_2(0.50) + 0.47 \times \log_2(0.47) + \\
&0.03 \times \log_2(0.03) + 0 \times \log_2(0)] \\
&= -[0 + 0.50 \times -1 + 0.47 \times -1.09 + 0.03 \times -5.06 + 0] \\
&= -[0 + 0.5 + 0.52 + 0.15 + 0] \\
&= 1.17
\end{aligned}$$

$$\begin{aligned}
\text{IG (parent, child)} &= \text{entropy (parent)} - [p(\text{Cereal}) \times \text{entropy}(\text{Cereal}) + p(\text{Donut}) \times \\
&\text{entropy}(\text{Donut})] \\
&= 1.37 - [0.87 \times 1.37 + 0.13 \times 1.17] \\
&= 1.37 - [1.19 + 0.15] \\
&= 1.37 - 1.34 \\
&= 0.03
\end{aligned}$$

3) คำนวณค่า IG ของแอตทริบิวต์ Drink และ Coffee ของ Orange Juice จากข้อมูล
สามารถคำนวณค่า IG ได้ดังนี้

$$\begin{aligned}
\text{entropy (parent)} &= -p(A) \times \log_2 p(A) + p(B+) \times \log_2 p(B+) + p(B) \times \log_2 p(B) \\
&+ p(C+) \times \log_2 p(C+) + p(C) \times \log_2 p(C)] \\
&= -[0.06 \times \log_2(0.06) + 0.42 \times \log_2(0.42) + 0.38 \times \\
&\log_2(0.38) + 0.12 \times \log_2(0.12) + 0.01 \times \log_2(0.01)] \\
&= -[0.06 \times -4.06 + 0.42 \times -1.25 + 0.38 \times -1.40 + 0.12 \times \\
&-3.06 + 0.01 \times -6.64] \\
&= -[0.24 + 0.53 + 0.53 + 0.37 + 0.07] \\
&= 1.74
\end{aligned}$$

$$\begin{aligned}
\text{Entropy (Creamy)} &= -p(A) \times \log_2 p(A) + p(B+) \times \log_2 p(B+) + p(B) \times \log_2 p(B) \\
&+ p(C+) \times \log_2 p(C+) + p(C) \times \log_2 p(C)] \\
&= -[0.06 \times \log_2(0.06) + 0.34 \times \log_2(0.34) + 0.42 \times \\
&\log_2(0.42) + 0.14 \times \log_2(0.14) + 0.04 \times \log_2(0.04)] \\
&= -[0.06 \times -4.06 + 0.34 \times -1.56 + 0.42 \times -1.25 + 0.14 \times \\
&-2.84 + 0.04 \times -4.64] \\
&= -[0.24 + 0.53 + 0.53 + 0.40 + 0.19] \\
&= 1.89
\end{aligned}$$

$$\begin{aligned}
\text{Entropy (Espresso)} &= -p(A) \times \log_2 p(A) + p(B+) \times \log_2 p(B+) + p(B) \times \log_2 p(B) \\
&+ p(C+) \times \log_2 p(C+) + p(C) \times \log_2 p(C)]
\end{aligned}$$

$$\begin{aligned}
&= -[0.07 \times \log_2(0.07) + 0.46 \times \log_2(0.46) + 0.36 \times \\
&\log_2(0.36) + 0.11 \times \log_2(0.11) + 0 \times \log_2(0)] \\
&= -[0.07 \times -3.84 + 0.46 \times -1.12 + 0.36 \times -1.47 + 0.11 \times \\
&-3.18 + 0] \\
&= -[0.27 + 0.52 + 0.53 + 0.35 + 0] \\
&= 1.67
\end{aligned}$$

$$\begin{aligned}
IG \text{ (parent, child)} &= \text{entropy (parent)} - [p(\text{Creamy}) \times \text{entropy}(\text{Creamy}) + \\
&p(\text{Espresso}) \times \text{entropy}(\text{Espresso})] \\
&= 1.74 - [0.32 \times 1.89 + 0.68 \times 1.67] \\
&= 1.74 - [0.60 + 1.14] \\
&= 1.74 - 1.74 \\
&= 0
\end{aligned}$$

3.1) คำนวณค่า IG ของแอตทริบิวต์ Drink และ Coffee ของ Soda จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

$$\begin{aligned}
\text{entropy (parent)} &= -p(A) \times \log_2 p(A) + p(B+) \times \log_2 p(B+) + p(B) \times \log_2 p(B) \\
&+ p(C+) \times \log_2 p(C+) + p(C) \times \log_2 p(C)] \\
&= -[0.01 \times \log_2(0.01) + 0.55 \times \log_2(0.55) + 0.38 \times \\
&\log_2(0.38) + 0.04 \times \log_2(0.04) + 0.02 \times \log_2(0.02)] \\
&= -[0.01 \times -6.64 + 0.55 \times -0.86 + 0.38 \times -1.40 + 0.04 \\
&\times -4.64 + 0.02 \times -5.64] \\
&= -[0.07 + 0.47 + 0.53 + 0.19 + 0.11] \\
&= 1.37
\end{aligned}$$

$$\begin{aligned}
\text{Entropy (Creamy)} &= -p(A) \times \log_2 p(A) + p(B+) \times \log_2 p(B+) + p(B) \times \log_2 p(B) \\
&+ p(C+) \times \log_2 p(C+) + p(C) \times \log_2 p(C)] \\
&= -[0 \times \log_2(0) + 0.53 \times \log_2(0.53) + 0.44 \times \log_2(0.44) + \\
&0.04 \times \log_2(0.04) + 0 \times \log_2(0)] \\
&= -[0 + 0.53 \times -0.92 + 0.44 \times -1.18 + 0.04 \times -4.64 + 0] \\
&= -[0 + 0.49 + 0.52 + 0.19 + 0] \\
&= 1.2
\end{aligned}$$

$$\begin{aligned}
\text{Entropy (Espresso)} &= -p(A) \times \log_2 p(A) + p(B+) \times \log_2 p(B+) + p(B) \times \log_2 p(B) \\
&+ p(C+) \times \log_2 p(C+) + p(C) \times \log_2 p(C)]
\end{aligned}$$

$$\begin{aligned}
&= -[0.01 \times \log_2(0.01) + 0.56 \times \log_2(0.56) + 0.36 \times \\
&\log_2(0.36) + 0.05 \times \log_2(0.05) + 0.03 \times \log_2(0.03)] \\
&= -[0.01 \times -6.64 + 0.56 \times -0.84 + 0.36 \times -1.47 + 0.05 \\
&\times -4.32 + 0.03 \times -5.06] \\
&= -[0.07 + 0.47 + 0.53 + 0.22 + 15] \\
&= 1.44
\end{aligned}$$

$$\begin{aligned}
IG(\text{parent, child}) &= \text{entropy}(\text{parent}) - [p(\text{Creamy}) \times \text{entropy}(\text{Creamy}) + \\
&p(\text{Espresso}) \times \text{entropy}(\text{Espresso})] \\
&= 1.37 - [0.34 \times 1.2 + 0.66 \times 1.44] \\
&= 1.37 - [0.41 + 0.95] \\
&= 1.37 - 1.36 \\
&= 0.01
\end{aligned}$$

จากการคำนวณค่า IG ของแอตทริบิวต์ Drink ต่อ Gender และ Breakfast และ Coffee พบว่า ค่า IG ของแอตทริบิวต์ Drink ต่อ Gender ของ Female มีค่ามากที่สุด (0.5) ดังนั้นจึงเลือกแอตทริบิวต์ Gender ขึ้นมาเป็นโหนดในระดับที่ 2 ต่อจากโหนด Root และทำการแตกกิ่งจากโหนดในระดับที่ 2 ออกไปจนข้อมูลในแต่ละโหนดมีคลาสคำตอบเดียวกัน

การคำนวณค่าแต่ละแอตทริบิวต์ในระดับที่ 3 กับจำนวนข้อมูลทั้งหมดโดยใช้ผลลัพธ์เป็น GPA

1) คำนวณค่า IG ของแอตทริบิวต์ Drink (Orange Juice) และ Gender (Female) และ Breakfast จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

$$\begin{aligned}
\text{entropy}(\text{parent}) &= -p(A) \times \log_2 p(A) + p(B+) \times \log_2 p(B+) + p(B) \times \log_2 p(B) \\
&+ p(C+) \times \log_2 p(C+) + p(C) \times \log_2 p(C)] \\
&= -[0.09 \times \log_2(0.09) + 0.42 \times \log_2(0.42) + 0.35 \times \\
&\log_2(0.35) + 0.15 \times \log_2(0.15) + 0 \times \log_2(0)] \\
&= -[0.09 \times -3.47 + 0.42 \times -1.25 + 0.35 \times -1.51 + 0.15 \times \\
&-2.74 + 0] \\
&= -[0.31 + 0.53 + 0.53 + 0.41 + 0] \\
&= 1.78
\end{aligned}$$

$$\begin{aligned}
\text{Entropy}(\text{Cereal}) &= -p(A) \times \log_2 p(A) + p(B+) \times \log_2 p(B+) + p(B) \times \log_2 p(B) \\
&+ p(C+) \times \log_2 p(C+) + p(C) \times \log_2 p(C)]
\end{aligned}$$

$$\begin{aligned}
&= -[0.10 \times \log_2(0.10) + 0.42 \times \log_2(0.42) + 0.37 \times \\
&\log_2(0.37) + 0.12 \times \log_2(0.12) + 0 \times \log_2(0)] \\
&= -[0.10 \times -3.32 + 0.42 \times -1.25 + 0.37 \times -1.43 + 0.12 \times \\
&-3.06 + 0] \\
&= -[0.33 + 0.53 + 0.53 + 0.37 + 0] \\
&= 1.76
\end{aligned}$$

$$\begin{aligned}
\text{Entropy (Donut)} &= -p(A) \times \log_2 p(A) + p(B+) \times \log_2 p(B+) + p(B) \times \log_2 p(B) \\
&+ p(C+) \times \log_2 p(C+) + p(C) \times \log_2 p(C)] \\
&= -[0 \times \log_2(0) + 0.43 \times \log_2(0.43) + 0 \times \log_2(0) + 0.57 \\
&\times \log_2(0.57) + 0 \times \log_2(0)] \\
&= -[0 + 0.43 \times -1.22 + 0 + 0.57 \times -0.81 + 0] \\
&= -[0 + 0.52 + 0 + 0.46 + 0] \\
&= 0.98
\end{aligned}$$

$$\begin{aligned}
\text{IG (parent, child)} &= \text{entropy (parent)} - [p(\text{Cereal}) \times \text{entropy}(\text{Cereal}) + p(\text{Donut}) \times \\
&\text{entropy}(\text{Donut})] \\
&= 1.78 - [0.66 \times 1.76 + 0.04 \times 0.98] \\
&= 1.78 - [1.16 + 0.04] \\
&= 1.78 - 1.2 \\
&= 0.58
\end{aligned}$$

2) คำนวณค่า IG ของแอตทริบิวต์ Drink (Orange Juice) และ Gender (Female) และ Coffee จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

$$\begin{aligned}
\text{entropy (parent)} &= -p(A) \times \log_2 p(A) + p(B+) \times \log_2 p(B+) + p(B) \times \log_2 p(B) \\
&+ p(C+) \times \log_2 p(C+) + p(C) \times \log_2 p(C)] \\
&= -[0.09 \times \log_2(0.09) + 0.42 \times \log_2(0.42) + 0.35 \times \\
&\log_2(0.35) + 0.15 \times \log_2(0.15) + 0 \times \log_2(0)] \\
&= -[0.09 \times -3.47 + 0.42 \times -1.25 + 0.35 \times -1.51 + 0.15 \times \\
&-2.74 + 0] \\
&= -[0.31 + 0.53 + 0.53 + 0.41 + 0] \\
&= 1.78
\end{aligned}$$

$$\begin{aligned}
\text{Entropy (Creamy)} &= -p(A) \times \log_2 p(A) + p(B+) \times \log_2 p(B+) + p(B) \times \log_2 p(B) \\
&+ p(C+) \times \log_2 p(C+) + p(C) \times \log_2 p(C)]
\end{aligned}$$

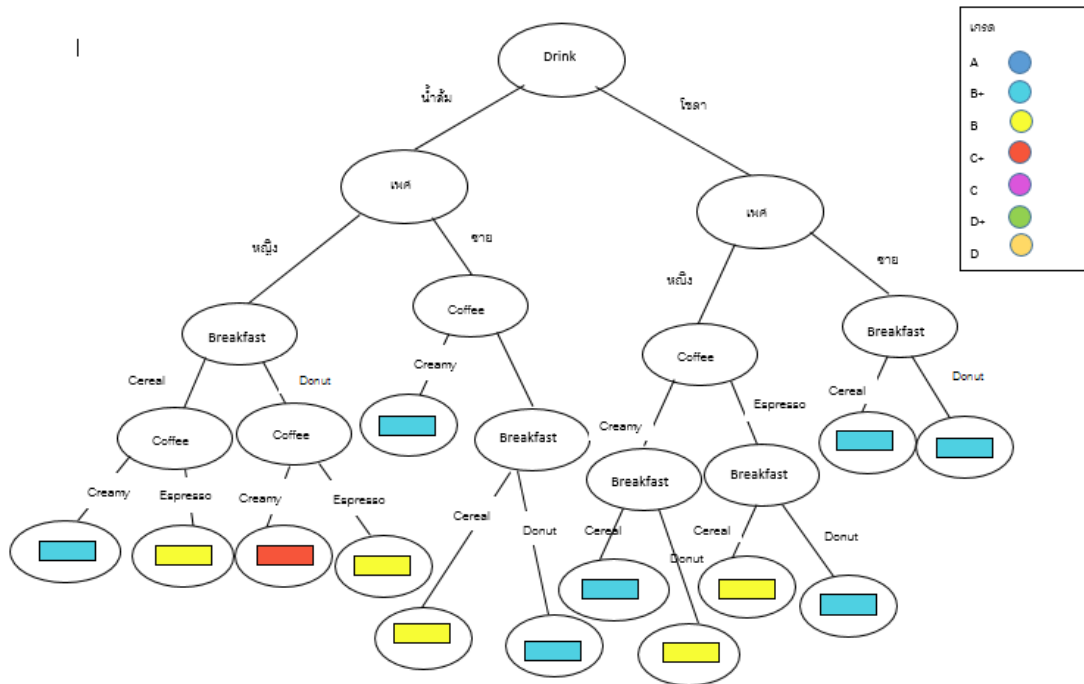
$$\begin{aligned}
&= -[0.07 \times \log_2(0.07) + 0.35 \times \log_2(0.35) + 0.42 \times \\
&\log_2(0.42) + 0.16 \times \log_2(0.16) + 0 \times \log_2(0)] \\
&= -[0.07 \times -3.84 + 0.35 \times -1.51 + 0.42 \times -1.25 + 0.16 \times \\
&-2.64 + 0] \\
&= -[0.27 + 0.53 + 0.53 + 0.42 + 0] \\
&= 1.75
\end{aligned}$$

$$\begin{aligned}
\text{Entropy (Espresso)} &= -p(A) \times \log_2 p(A) + p(B+) \times \log_2 p(B+) + p(B) \times \log_2 p(B) \\
&+ p(C+) \times \log_2 p(C+) + p(C) \times \log_2 p(C)] \\
&= -[0.10 \times \log_2(0.10) + 0.46 \times \log_2(0.46) + 0.30 \times \\
&\log_2(0.30) + 0.13 \times \log_2(0.13) + 0 \times \log_2(0)] \\
&= -[0.10 \times -3.32 + 0.46 \times -1.12 + 0.30 \times -1.74 + 0.13 \times \\
&-2.94 + 0] \\
&= -[0.33 + 0.52 + 0.52 + 0.38 + 0] \\
&= 1.75
\end{aligned}$$

$$\begin{aligned}
\text{IG (parent, child)} &= \text{entropy (parent)} - [p(\text{Creamy}) \times \text{entropy}(\text{Creamy}) + \\
&p(\text{Espresso}) \times \text{entropy}(\text{Espresso})] \\
&= 1.78 - [0.27 \times 1.75 + 0.43 \times 1.75] \\
&= 1.78 - [0.47 + 0.75] \\
&= 1.78 - 1.22 \\
&= 0.56
\end{aligned}$$

จากการคำนวณค่า IG ของแอตทริบิวต์ Drink ต่อ Gender และ Breakfast และ Coffee พบว่าค่า IG ของแอตทริบิวต์ Drink (Orange Juice) และ Gender (Female) ต่อ Breakfast (Cereal) มีค่ามากที่สุด (0.58) ดังนั้นจึงเลือกแอตทริบิวต์ Breakfast (Cereal) ขึ้นมาเป็นโหนดในระดับที่ 3 ต่อจากโหนดระดับที่ 2 และโหนดสุดท้ายคือแอตทริบิวต์ Coffee ทำการแตกกิ่งจากโหนดในระดับที่ 3 ออกไป

จากการคำนวณค่า IG ของแอตทริบิวต์ Drink และ Gender และ Breakfast และ Coffee พบว่าข้อมูลในแต่ละโหนดมีคลาสคำตอบเดียวกันแล้ว คือ ผลลัพธ์เป็น GPA

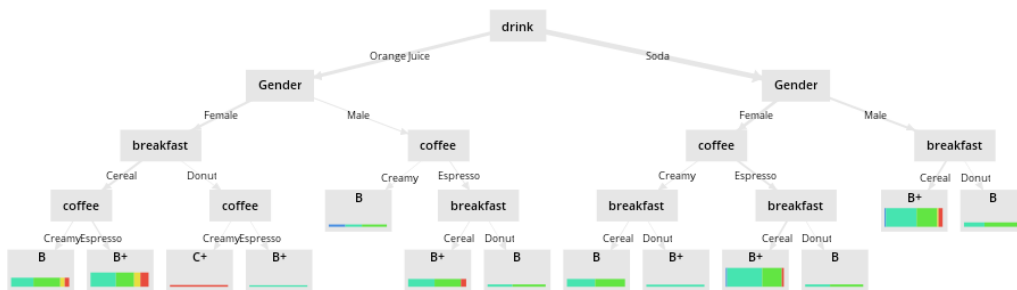


ภาพที่ 3.11 รูปแบบโมเดล Graph Decision Tree

3.1.5 การประเมินผล (Evaluation) เป็นขั้นตอนก่อนนำผลลัพธ์ที่ได้จากขั้นตอนที่ 4 ไปใช้งาน ด้วยการวัดประสิทธิภาพของผลลัพธ์ที่ได้กับวัตถุประสงค์ที่ตั้งไว้ในขั้นตอนแรก ว่ามีนัยสำคัญหรือความน่าเชื่อถือมากน้อยเพียงใด ด้วยการประเมินผลจากโปรแกรมว่าถูกต้องหรือไม่

ผู้วิเคราะห์ข้อมูลได้ข้อมูลที่ผ่านการวิเคราะห์ข้อมูลด้วยเทคนิค Data Mining จากการสร้างโมเดล Decision Tree ดังนี้

1) โปรแกรม RapidMiner Studio



ภาพที่ 3.12 รูปแบบโมเดล Graph Decision Tree ในโปรแกรม Rapid Miner

จากการจัดทำ รูปแบบโมเดล Decision Tree ทั้งในรูปแบบเขียนมือและการใช้โปรแกรม Rapid Miner เมื่อนำมาเปรียบเทียบกันแล้วพบว่าทั้ง 2 โมเดลได้ผลลัพธ์ที่มีความแม่นยำของโมเดลตรงกันสูง

ดังนั้นผู้วิเคราะห์ข้อมูลจึงเลือกใช้โมเดลจากโปรแกรม Rapid Miner ดีกว่าเพื่อจะได้มีความถูกต้องและแม่นยำกว่าการเขียนมือ และเพื่อไม่ให้เกิดข้อผิดพลาดหรืออาจซับซ้อน

จากผลลัพธ์การสร้างโมเดลด้วยเทคนิค Decision Tree ในโปรแกรม Rapid Miner สามารถทำนายได้จำนวน 13 กฎ ดังนี้

Tree

```

drink = Orange Juice
| Gender = Female
| | breakfast = Cereal
| | | coffee = Creamy: B {C=0, B+=15, B=18, A=3, C+=3}
| | | coffee = Espresso: B+ {C=0, B+=28, B=20, A=7, C+=9}
| | | breakfast = Donut
| | | coffee = Creamy: C+ {C=0, B+=0, B=0, A=0, C+=4}
| | | coffee = Espresso: B+ {C=0, B+=3, B=0, A=0, C+=0}
| Gender = Male
| | coffee = Creamy: B {C=2, B+=2, B=3, A=0, C+=0}
| | coffee = Espresso
| | | breakfast = Cereal: B+ {C=0, B+=15, B=15, A=0, C+=3}
| | | breakfast = Donut: B {C=0, B+=3, B=4, A=0, C+=0}
drink = Soda
| Gender = Female
| | coffee = Creamy
| | | breakfast = Cereal: B {C=0, B+=16, B=17, A=0, C+=0}
| | | breakfast = Donut: B+ {C=0, B+=6, B=0, A=0, C+=0}
| | | coffee = Espresso
| | | breakfast = Cereal: B+ {C=2, B+=52, B=29, A=0, C+=3}
| | | breakfast = Donut: B {C=0, B+=3, B=4, A=0, C+=0}
| Gender = Male
| | breakfast = Cereal: B+ {C=2, B+=45, B=30, A=2, C+=6}
| | breakfast = Donut: B {C=0, B+=6, B=10, A=0, C+=1}

```

ภาพที่ 3.13 คำบรรยายลักษณะงาน Decision Tree ของ Rapid Miner

ดังนั้น ผู้จัดทำได้ผลลัพธ์ของกฎที่สามารถทำนายได้จำนวน 13 กฎ และสามารถนำกฎที่ได้ไปวิเคราะห์กฎต่อไปได้ โดยสามารถจำแนกกฎได้ ดังนี้

กฎข้อที่ 1 IF drink = Orange juice = Gender = Female = breakfast = cereal = coffee = creamy = B หมายความว่า ถ้าเลือกดื่มน้ำส้ม โดยเพศหญิงแล้วเลือกทานอาหารเช้าเป็นซีเรียล และดื่มกาแฟใส่น้ำนมจะมีผลทำให้ได้เกรด “B”

กฎข้อที่ 2 IF drink = Orange juice = Gender = Female = breakfast = cereal = coffee = Espresso = B+ หมายความว่า ถ้าเลือกดื่มน้ำส้ม โดยเพศหญิงแล้วเลือกทานอาหารเช้าเป็นซีเรียล และดื่มกาแฟเอสเปรสโซจะมีผลทำให้ได้เกรด “B+”

กฎข้อที่ 3 IF drink = Orange juice = Gender = Female = breakfast = Donut = coffee = creamy = C+ หมายความว่า ถ้าเลือกดื่มน้ำส้ม โดยเพศหญิงแล้วเลือกทานอาหารเช้าเป็นโดนัท และดื่มกาแฟใส่น้ำนมจะมีผลทำให้ได้เกรด “C+”

กฎข้อที่ 4 IF drink = Orange juice = Geander = Female = breakfast = Donut = coffee = Espresso = B+ หมายความว่า ถ้าเลือกดื่มน้ำส้ม โดยเพศหญิงแล้วเลือกทานอาหารเช้าเป็นโดนัท และดื่มกาแฟดื่มกาแฟเอสเปรสโซ่อาจมีผลทำให้ได้เกรด “B+”

กฎข้อที่ 5 IF drink = Orange juice = Geander = male = coffee = creamy = B หมายความว่า ถ้าเลือกดื่มน้ำส้ม โดยเพศชาย และดื่มกาแฟใส่ครีมอาจมีผลทำให้ได้เกรด “B”

กฎข้อที่ 6 IF drink = Orange juice = Geander = male = coffee = Espresso = breakfast = Cereal = B+ หมายความว่า ถ้าเลือกดื่มน้ำส้ม โดยเพศชายแต่ดื่มกาแฟดื่มกาแฟเอสเปรสโซ่ แล้วเลือกทานอาหารเช้าเป็นซีเรียล อาจมีผลทำให้ได้เกรด “B+”

กฎข้อที่ 7 IF drink = Orange juice = Geander = male = coffee = Espresso = breakfast = Donut = B หมายความว่า ถ้าเลือกดื่มน้ำส้ม โดยเพศชายแต่ดื่มกาแฟดื่มกาแฟเอสเปรสโซ่ แล้วเลือกทานอาหารเช้าเป็นโดนัท อาจมีผลทำให้ได้เกรด “B”

กฎข้อที่ 8 IF drink = Soda = Geander = Female = coffee = Creamy = breakfast = Cereal = B หมายความว่า ถ้าเลือกดื่มโซดา โดยเพศหญิงแต่ดื่มกาแฟใส่ครีม แล้วเลือกทานอาหารเช้าเป็นซีเรียล อาจมีผลทำให้ได้เกรด “B”

กฎข้อที่ 9 IF drink = Soda = Geander = Female = coffee = Espresso = breakfast = Donut = B+ ถ้าเลือกดื่มโซดาโดยเพศหญิงแต่ดื่มกาแฟดื่มกาแฟเอสเปรสโซ่ แล้วเลือกทานอาหารเช้าเป็นโดนัท อาจมีผลทำให้ได้เกรด “B+”

กฎข้อที่ 10 IF drink = Soda = Geander = Female = coffee = Espresso = breakfast = Cereal = B+ หมายความว่า ถ้าเลือกดื่มโซดา โดยเพศหญิงแต่ดื่มกาแฟเอสเปรสโซ่ แล้วเลือกทานอาหารเช้าเป็นซีเรียล อาจมีผลทำให้ได้เกรด “B+”

กฎข้อที่ 11 IF drink = Soda = Geander = Female = coffee = Espresso = breakfast = Donut = B หมายความว่า ถ้าเลือกดื่มโซดา โดยเพศหญิงแต่ดื่มกาแฟเอสเปรสโซ่ แล้วเลือกทานอาหารเช้าเป็นโดนัท อาจมีผลทำให้ได้เกรด “B”

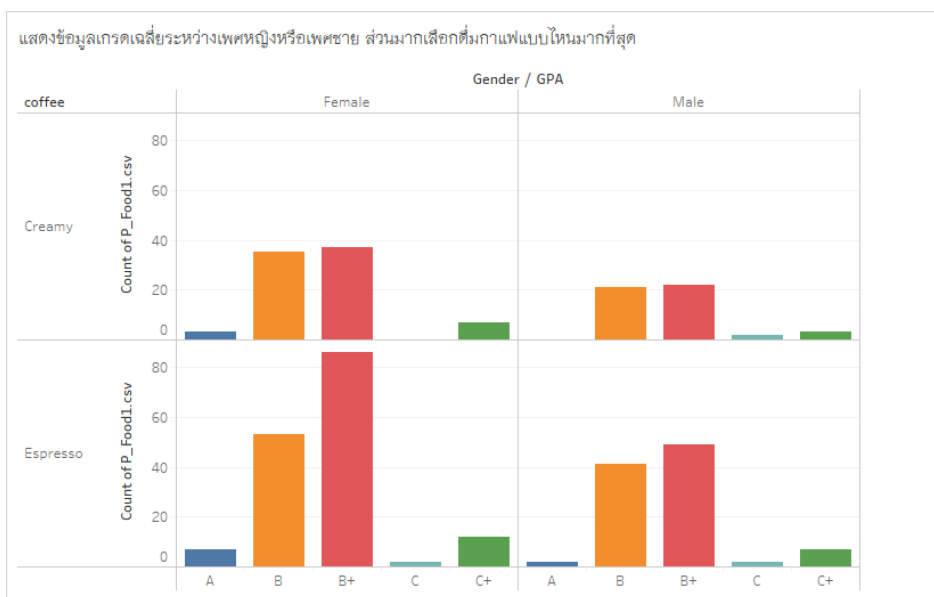
กฎข้อที่ 12 IF drink = Soda = Geander = male = breakfast = Cereal = B+ หมายความว่า ถ้าเลือกดื่มโซดา โดยเพศชาย แล้วเลือกทานอาหารเช้าเป็นซีเรียล อาจมีผลทำให้ได้เกรด “B+”

กฎข้อที่ 13 IF drink = Orange juice = Geander = Female = breakfast = Donut = B หมายความว่า ถ้าเลือกดื่มโซดา โดยเพศชาย แล้วเลือกทานอาหารเช้าเป็นโดนัท อาจมีผลทำให้ได้เกรด “B”

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

ผู้วิเคราะห์ข้อมูลนำผลข้อมูลที่ทำกรวิเคราะห์มาแสดงผลข้อมูลบนหน้า Web Browser โดยใช้ชุดคำสั่งHTML และ CSS3 ร่วมกับการนำเสนอข้อมูลแบบ visudalization ด้วยการแสดงผลข้อมูลในรูปแบบของภาพโดยใช้โปรแกรม Tableau Public ซึ่งทางผู้จัดทำได้ยกตัวอย่างการจัดทำรูปแบบของรายงาน (Report) หรือแผนภาพ (Dashboard) เกี่ยวกับข้อมูลทางโภชนาการที่มีผลต่อเกรดเฉลี่ยของนักศึกษา ดังนี้

- 1) แสดงข้อมูลเกรดเฉลี่ยระหว่างเพศหญิงหรือเพศชาย ส่วนมากเลือกดื่มกาแฟแบบไหนมากที่สุด จำแนกตามเกรดเฉลี่ยและเพศซึ่งเป็นการแสดงผลในรูปแบบแผนภูมิแท่งผู้ใช้สามารถเลือกดูการจำแนกประเภทของกาแฟได้



ภาพ3.14 กราฟแท่งแสดงผลข้อมูลเกรดเฉลี่ยระหว่างเพศหญิงหรือเพศชาย ส่วนมากเลือกดื่มกาแฟแบบไหน

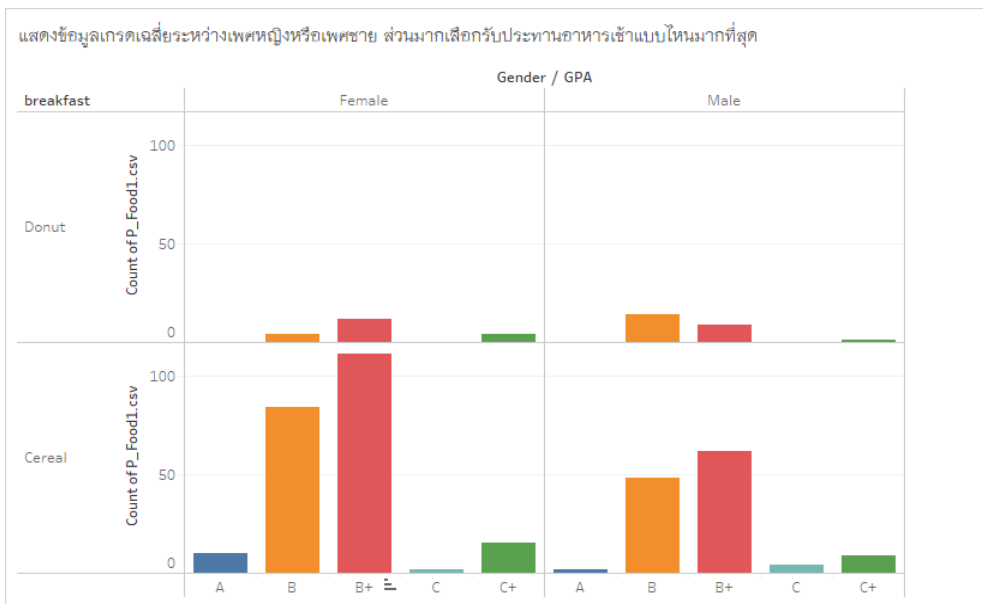
2) แสดงข้อมูลเกรดเฉลี่ยระหว่างเพศหญิงหรือเพศชาย ส่วนมากเลือกเครื่องดื่มแบบไหนมากที่สุด จำแนกตามเกรดเฉลี่ยและเพศซึ่งเป็นการแสดงผลในรูปแบบแผนภูมิแท่งผู้ใช้สามารถเลือกดูการจำแนกประเภทของเครื่องดื่มได้

แสดงข้อมูลเกรดเฉลี่ยระหว่างเพศหญิงหรือเพศชาย ส่วนมากเลือกเครื่องดื่มแบบไหนมากที่สุด



ภาพ3.15 กราฟแท่งแสดงผลข้อมูลเกรดเฉลี่ยระหว่างเพศหญิงหรือเพศชาย ส่วนมากเลือกเครื่องดื่มแบบไหน

3) แสดงข้อมูลเกรดเฉลี่ยระหว่างเพศหญิงหรือเพศชาย ส่วนมากเลือกรับประทานอาหารเช้าแบบไหนมากที่สุด จำแนกตามเกรดเฉลี่ยและเพศซึ่งเป็นการแสดงผลในรูปแบบแผนภูมิแท่งผู้ใช้สามารถเลือกดูการจำแนกประเภทของอาหารเช้าได้



ภาพ3.16 กราฟแท่งแสดงผลข้อมูลเกรดเฉลี่ยระหว่างเพศหญิงหรือเพศชาย ส่วนมากเลือกรับประทานอาหารเช้าแบบไหน

3.2 การออกแบบเว็บไซต์

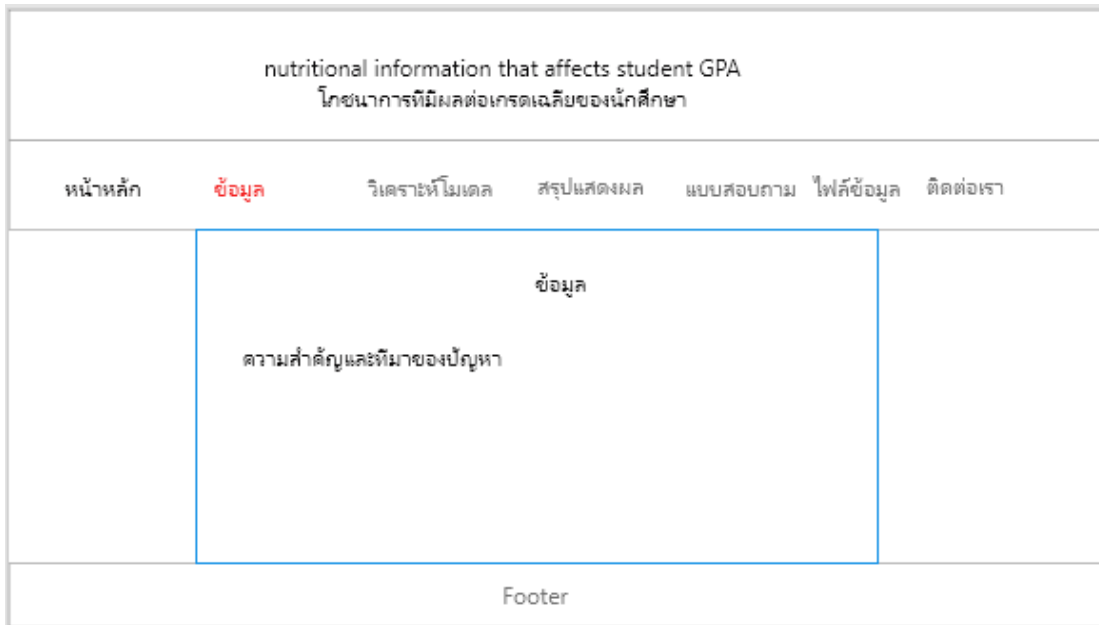
3.2.1 การออกแบบ Wireframe หน้าจอเว็บไซต์

1) หน้าแรกของเว็บไซต์

nutritional information that affects student GPA โภชนาการที่มีผลต่อเกรดเฉลี่ยของนักศึกษา							
หน้าหลัก	ข้อมูล	วิเคราะห์โมเดล	สรุปแสดงผล	แบบสอบถาม	ไฟล์ข้อมูล	ติดต่อเรา	login
	ข้อมูลโภชนาการ						
Footer							

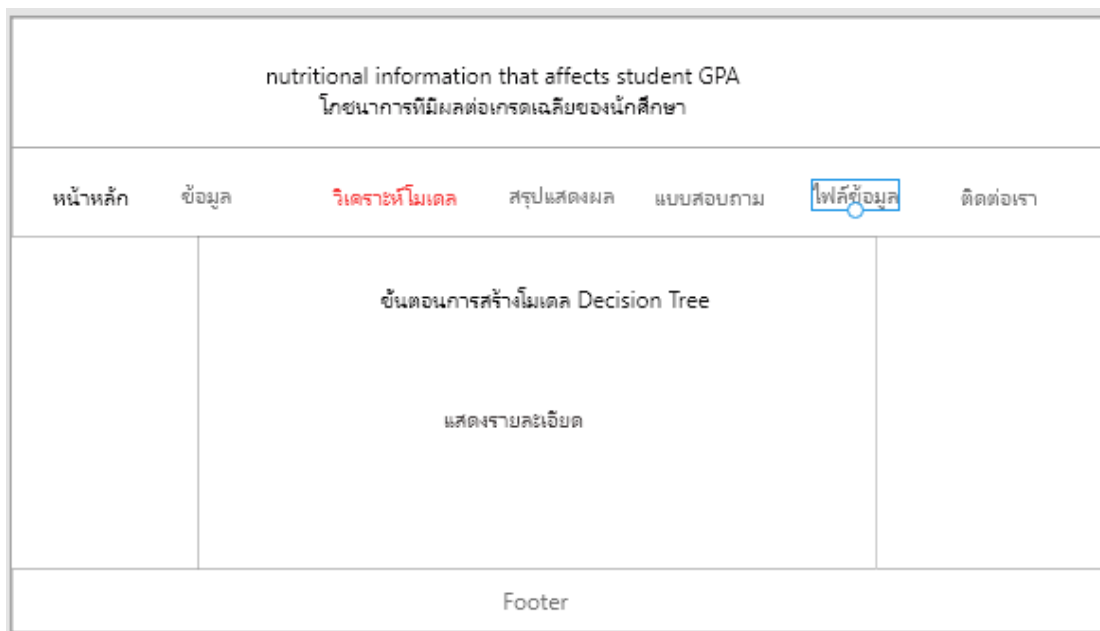
ภาพ3.17 หน้าแรก

2) หน้าแสดงข้อมูลของเว็บไซต์



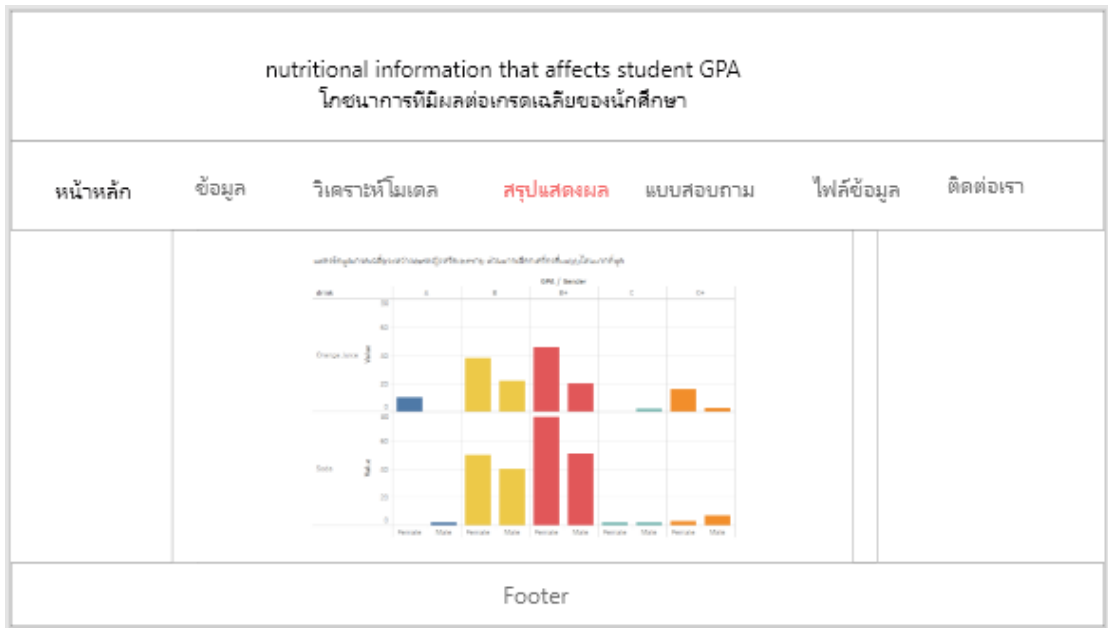
ภาพ3.18 หน้าแสดงข้อมูล

3) หน้ากฎความสัมพันธ์ของการวิเคราะห์ข้อมูลโภชนาการที่มีผลต่อเกรดเฉลี่ย



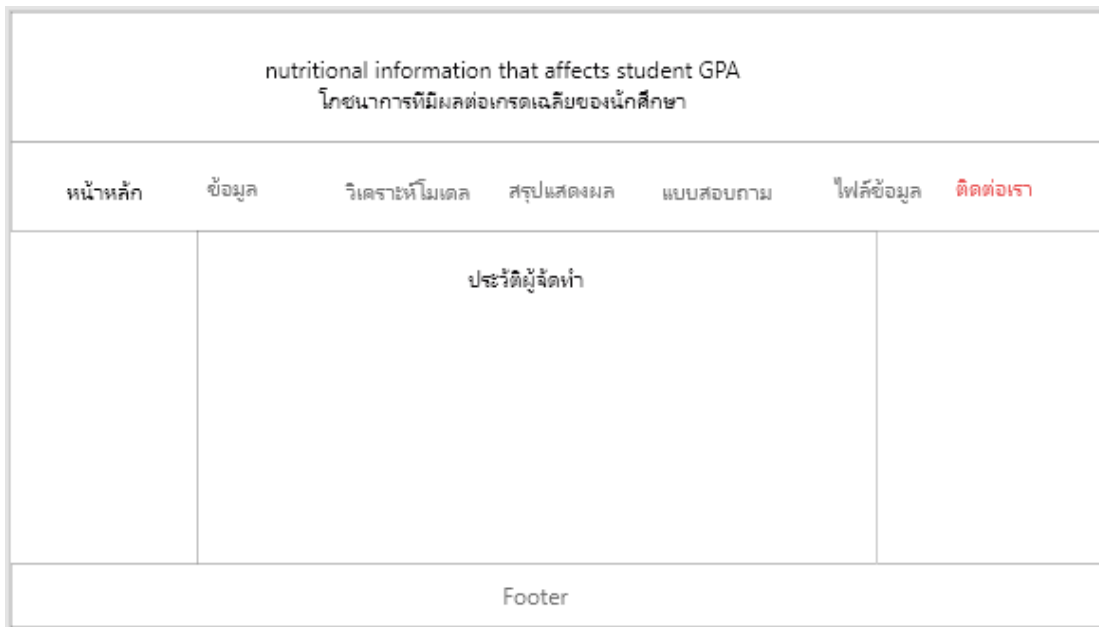
ภาพ3.19 แสดงการวิเคราะห์ Decision Tree

4) หน้าแสดงหน้า dataset ของเว็บไซต์



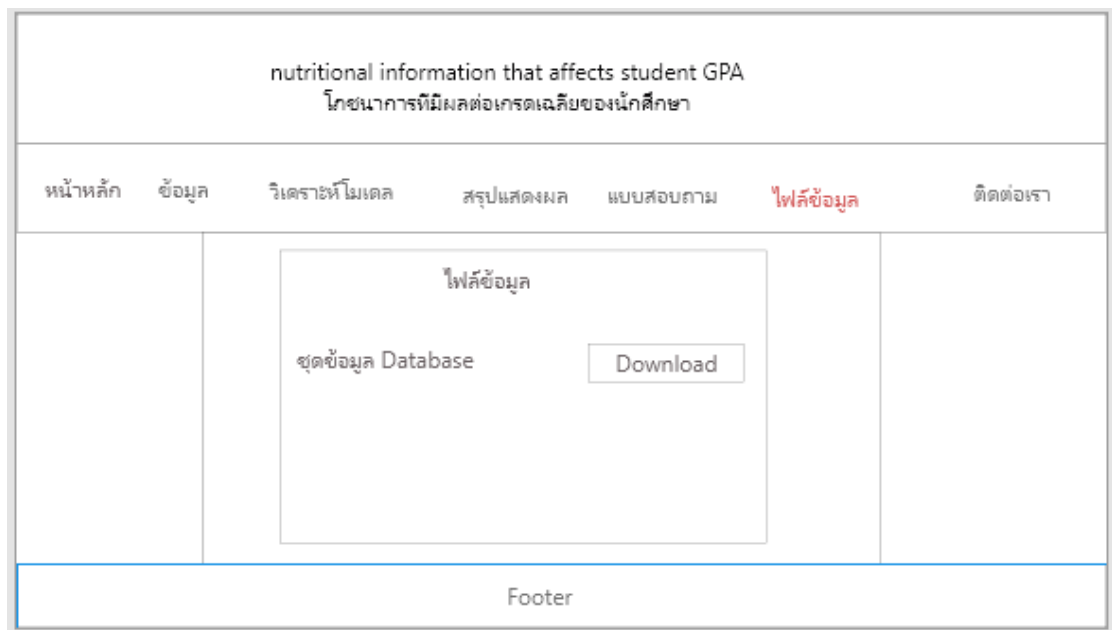
ภาพ3.20 แสดงผลลัพธ์การวิเคราะห์

5) หน้ากฎความสัมพันธ์ของการวิเคราะห์ข้อมูลโภชนาการที่มีผลต่อเกรดเฉลี่ย



ภาพ3.21 แสดงข้อมูลผู้จัดทำ

6) เป็นหน้าแสดงการรายการเอกสารที่ผู้วิเคราะห์ใช้ในการพัฒนาโครงการได้




ภาพ3.22 แสดงเอกสารสำหรับผู้ใช้งานทั่วไปดาวน์โหลดได้

7) หน้าล็อกอินสำหรับแอดมิน

nutritional information that affects student GPA โภชนาการที่มีผลต่อเกรดเฉลี่ยของนักศึกษา							
หน้าหลัก	ข้อมูล	วิเคราะห์โมเดล	สรุปแสดงผล	แบบสอบถาม	ไฟล์ข้อมูล	ติดต่อเรา	login
	<div style="border: 1px solid #ccc; padding: 10px; width: fit-content; margin: auto;"><p>sing in</p><input type="text" value="username"/> <input type="password" value="password"/> <input type="button" value="Login"/></div>						
Footer							

ภาพ3.23 หน้าล็อกอิน

8) หน้าแสดงรายการอัปโหลดไฟล์ สามารถดาวน์โหลดได้ และสามารถลบเอกสารที่ต้องการลบได้

nutritional information that affects student GPA โภชนาการที่มีผลต่อเกรดเฉลี่ยของนักศึกษา							
หน้าหลัก	ข้อมูล	วิเคราะห์โมเดล	สรุปแสดงผล	แบบสอบถาม	ไฟล์ข้อมูล	ติดต่อเรา	จกระบบ
	<div style="border: 1px solid #ccc; padding: 10px; width: fit-content; margin: auto;"><div style="text-align: center;"> ลากแล้วปล่อยไฟล์</div><div style="border: 1px solid #ccc; padding: 5px; width: fit-content; margin: 5px auto;">เลือกไฟล์ที่จะอัปโหลด</div><div style="text-align: right; margin-top: 10px;"><input type="button" value="อัปโหลด"/></div></div>						
Footer							

ภาพ3.24 หน้าแสดงรายการอัปโหลดไฟล์

3.3 บทสรุป

จากวิธีการดำเนินงานโครงการในข้างต้น ผู้วิเคราะห์ข้อมูลได้แสดงวิธีในการจัดการกับการวิเคราะห์ข้อมูลทางโภชนาการที่มีผลต่อเกรดเฉลี่ยของนักศึกษา ด้วยขั้นตอนกระบวนการวิเคราะห์ข้อมูลด้วย CRISP-DM รวมถึงการสร้างโมเดล Decision Tree จากโปรแกรมที่ใช้ทำเหมืองข้อมูลโดยโปรแกรม RapidMiner Studio ในการสร้างโมเดล Decision Tree ซึ่งได้ผลลัพธ์ที่เหมือนกัน และนำข้อมูลสารสนเทศมาทำการแสดงผลแบบ visualization ในรูปแบบของภาพโดยใช้โปรแกรม Tableau Public และออกแบบ Wireframe ของเว็บไซต์ที่จะเผยแพร่บน web browser ด้วยโปรแกรม Adobe XD ปัญหาที่พบในบทรนี้ คือการทำ Decision Tree ผ่านโปรแกรม weka ซึ่งพบปัญหาคือ ผลลัพธ์ที่ผ่านกระบวนการทำ Decision Tree นั้น รูปแบบโมเดลไม่ตรงและแสดงผลแอตทริบิวต์ออกมาไม่ครบตามที่คำนวณและ code คำบรรยายลักษณะงานก็แสดงผลออกมาไม่ครบแอตทริบิวต์ ผู้วิเคราะห์ข้อมูลจึงได้ใช้แค่โปรแกรม RapidMiner Studio ในการทำ Decision Tree