

บทที่ 3

วิธีการดำเนินงานโครงการ

โครงการเรื่อง ปัจจัยที่มีอิทธิพลต่อความสำเร็จของนักศึกษาโดยใช้วิธีการเหมืองข้อมูล ในบทนี้จะเป็นการวิเคราะห์ข้อมูลด้วยเทคนิคทางดาต้าไมนิ่งซึ่งมีกระบวนการวิเคราะห์ที่สำคัญหลายขั้นตอน เมื่อเสร็จสิ้นจากกระบวนการวิเคราะห์ข้อมูลแล้วจะเป็นการออกแบบเว็บไซต์และออกแบบรูปแบบการแสดงผลและบทสรุปจากวิธีการดำเนินงาน

3.1 การวิเคราะห์ข้อมูลด้วย CRISP-DM

- 1) การจัดกลุ่มรายวิชาของหลักสูตรระบบสารสนเทศทางธุรกิจ
- 2) การจำแนกข้อมูลปัจจัยที่มีผลต่อความสำเร็จของนักศึกษา
- 3) การแสดงข้อมูล

3.2 การออกแบบเว็บไซต์

3.3 บทสรุป

3.1 กระบวนการ CRISP-DM (Cross Industry Standard Process for Data Mining)

CRISP-DM เป็นกระบวนการหลักในการจัดทำเหมืองข้อมูลเพื่อการวิเคราะห์และใช้ประโยชน์ในทางธุรกิจกระบวนการวิเคราะห์ข้อมูลด้วย CRISP-DM หรือ (Cross Industry Standard Process for Data Mining) ประกอบด้วย 6 ขั้นตอนดังนี้

3.1.1 ความเข้าใจในธุรกิจ (Business Understanding)

คณะผู้จัดทำทำความเข้าใจกับปัญหาให้อยู่ในรูปแบบของการวิเคราะห์ข้อมูลทาง Data Mining โดยการวิเคราะห์ข้อมูลในประเด็นนี้ คือ ปัจจัยที่มีอิทธิพลต่อความสำเร็จของนักศึกษาโดยใช้วิธีการเหมืองข้อมูล จากข้อมูลเกรด (GPA) ของนักศึกษาหลักสูตรระบบสารสนเทศ ช่วงปีการศึกษา 2560 – 2564

3.1.2 การทำความเข้าใจข้อมูล (Data Understanding)

ขั้นตอนการจัดเก็บและรวบรวมข้อมูล ผู้จัดทำต้องพิจารณาตรวจสอบความถูกต้องของข้อมูลที่ได้รับ โดยเลือกว่าจะใช้ข้อมูลทั้งหมดหรือบางส่วนในการวิเคราะห์ให้สอดคล้องกับวัตถุประสงค์ที่กำหนดไว้วิเคราะห์ข้อมูลจึงได้ดำเนินการตามขั้นตอนดังนี้

ผู้จัดทำรวบรวมชุดข้อมูลจากฝ่ายทะเบียนกลางของมหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา เชียงใหม่ ส่งข้อมูลผ่านทาง E-mail หลังจากนั้นคลิกเลือก ดาวน์โหลด เพื่อเริ่มทำการดาวน์โหลดเก็บรวบรวมข้อมูล และตรวจสอบความถูกต้องของข้อมูล



ภาพที่ 3.1 ช่องทางการนำข้อมูล และ Download ข้อมูลจาก E-mail

ซึ่งข้อมูลได้รับมา มีจำนวนข้อมูลทั้งหมดคือ 580 รายการ 114 แอดทริบิวต์ ประกอบด้วย ประเภทการศึกษา สภาพนักศึกษา เกรดเฉลี่ย และรายวิชาที่นักศึกษาได้ลงทะเบียนเรียน

ประเภทการศึกษา	สาขาการศึกษา	เกรดเฉลี่ย	DACC2014 ค่าเฉลี่ยคะแนนเฉลี่ยชั้น 1 ถึง 2	DACC2015 ค่าปริญญ์บริหาร	BBABA2015 ค่าดัชนีผลสัมฤทธิ์	ค่ากลาง	อันดับ
ปริญญาตรี (ต่อเนื่อง/เทียบโอน)	สาขาบริหารการศึกษา	2.96		D+	TC		C
ปริญญาตรี (ต่อเนื่อง/เทียบโอน)	สาขาบริหารการศึกษา	3.37		TC	TC		A
ปริญญาตรี (ต่อเนื่อง/เทียบโอน)	สาขาบริหารการศึกษา	3.00		C+	TC		A
ปริญญาตรี (ต่อเนื่อง/เทียบโอน)	สาขาบริหารการศึกษา	2.84		D	TC		A
ปริญญาตรี (ต่อเนื่อง/เทียบโอน)	สาขาบริหารการศึกษา	2.69		D	TC		B
ปริญญาตรี (ต่อเนื่อง/เทียบโอน)	สาขาบริหารการศึกษา	2.89		D+	TC		B+
ปริญญาตรี (ต่อเนื่อง/เทียบโอน)	สาขาบริหารการศึกษา	3.14		C+	TC		B
ปริญญาตรี (ต่อเนื่อง/เทียบโอน)	สาขาบริหารการศึกษา	2.54		D	TC		B
ปริญญาตรี (ต่อเนื่อง/เทียบโอน)	สาขาบริหารการศึกษา	3.47		D	TC		C+
ปริญญาตรี (ต่อเนื่อง/เทียบโอน)	สาขาบริหารการศึกษา	2.93		D+	TC		C+
ปริญญาตรี (ต่อเนื่อง/เทียบโอน)	สาขาบริหารการศึกษา	3.30		D	C+		A
ปริญญาตรี (ต่อเนื่อง/เทียบโอน)	สาขาบริหารการศึกษา	2.71		TC	TC		C+
ปริญญาตรี (ต่อเนื่อง/เทียบโอน)	สาขาบริหารการศึกษา	2.38		D	TC		D+
ปริญญาตรี (ต่อเนื่อง/เทียบโอน)	สาขาบริหารการศึกษา	3.09		D+	TC		A
ปริญญาตรี (ต่อเนื่อง/เทียบโอน)	สาขาบริหารการศึกษา	2.72		D	C+		A
ปริญญาตรี (ต่อเนื่อง/เทียบโอน)	สาขาบริหารการศึกษา	2.26		W	TC		D
ปริญญาตรี (ต่อเนื่อง/เทียบโอน)	สาขาบริหารการศึกษา	3.20		D	TC		B+
ปริญญาตรี (ต่อเนื่อง/เทียบโอน)	สาขาบริหารการศึกษา	2.83		C	TC		A
ปริญญาตรี (ต่อเนื่อง/เทียบโอน)	สาขาบริหารการศึกษา	2.76		B	TC		A
ปริญญาตรี (ต่อเนื่อง/เทียบโอน)	สาขาบริหารการศึกษา	3.01		B	TC		C+
ปริญญาตรี (ต่อเนื่อง/เทียบโอน)	สาขาบริหารการศึกษา	2.58		TC	TC		A
ปริญญาตรี (ต่อเนื่อง/เทียบโอน)	สาขาบริหารการศึกษา	3.07		D	TC		A

ตารางที่ 3.1 ไฟล์ข้อมูลที 2 เกรด (GPA) ของนักศึกษา

3.1.3 การเตรียมข้อมูล (Data Preparation)

3.1.3.1 การเตรียมข้อมูล (Data Preparation) สำหรับการจัดกลุ่ม (Clustering)

ขั้นตอนการแปลงข้อมูลที่ได้รวบรวมมาซึ่งเป็นขั้นตอนที่ใช้เวลามาก ผู้จัดทำทำการจัดการกับข้อมูลที่ได้โดยรวบรวมข้อมูลและแปลงข้อมูลที่ได้ให้อยู่ในรูปแบบที่พร้อมสำหรับนำไปวิเคราะห์ในขั้นตอนต่อไปเพื่อลดความผิดพลาดให้น้อยที่สุด โดยใช้กระบวนการ Data cleaning เพื่อให้ได้ข้อมูลที่มีประสิทธิภาพอยู่ในรูปแบบและมาตรฐานเดียวกัน และกำจัดข้อมูลเสีย ออกดำเนินการตามขั้นตอนดังนี้

3.1.3.1.1 ทำการคัดเลือกข้อมูล (Data Selection) คือการคัดเลือกข้อมูลที่เหมาะสมเพื่อนำมาใช้ในการวิเคราะห์ข้อมูล

ข้อมูลเกรด (GPA) ของนักศึกษาหลักสูตรระบบสารสนเทศช่วงปีการศึกษา 2560 – 2564 ผู้จัดทำทำการลบข้อมูลบางส่วนที่ไม่จำเป็นต่อการนำไปใช้ในการวิเคราะห์ คือ นักศึกษาที่กำลังศึกษาอยู่ทั้งหมด เนื่องจากนักศึกษาที่กำลังศึกษาอยู่ยังมีสถานะของ ข้อมูลเกรด (GPA) ที่ยังลงเรียนรายวิชายังไม่ครบหลักสูตร จึงทำให้ข้อมูลบางส่วนของนักศึกษาที่เข้าศึกษาที่กำลังศึกษาอยู่ มีค่าว่างเป็นจำนวนมากอาจทำให้การวิเคราะห์เกิดการผิดพลาด และลบรายวิชาที่อยู่ในหมวด

รายวิชาเสรีออก หรือรายวิชาที่มีนักศึกษาลงทะเบียนเรียน 1 ถึง 10 คน รวมถึงรายวิชาที่ให้เกรด S U และ I

ผู้จัดทำทำการคัดเลือกข้อมูล และทำการ Data Cleaning ข้อมูลนักศึกษาเบื้องต้น โดยตัดส่วนที่ไม่จำเป็นออกให้เหลือเฉพาะข้อมูลที่ใช้ในการวิเคราะห์ในภาพรวม จำนวน 66 แอคทริบิวท์ จาก 114 แอคทริบิวท์ ได้แก่ สถานศึกษา รหัสนักศึกษา ปีที่เข้าศึกษา แขนงวิชา ประเภทการศึกษา เกรดเฉลี่ย และรายวิชา โดยใช้ข้อมูล จำนวน 339 จาก 579 รายการซึ่งเป็นข้อมูลที่ใช้ในการนำไปวิเคราะห์ข้อมูล

ประเภทการศึกษา	สถานศึกษา	เกรดเฉลี่ย	BAC0114 การออกแบบผลิตภัณฑ์ 2 (D)	BAC0111 ศึกษานิเทศก์ (D)	EDBA201 ศึกษานานาชาติ (D)	ภาควิชา	รายวิชา
ปริญญาตรี (ต่อเนื่อง/เทียบโอน)	สาขาการศึกษา	2.96		D+	TC		C
ปริญญาตรี (ต่อเนื่อง/เทียบโอน)	สาขาการศึกษา	3.37		TC	TC		A
ปริญญาตรี (ต่อเนื่อง/เทียบโอน)	สาขาการศึกษา	3.00		C+	TC		A
ปริญญาตรี (ต่อเนื่อง/เทียบโอน)	สาขาการศึกษา	2.64		D	TC		A
ปริญญาตรี (ต่อเนื่อง/เทียบโอน)	สาขาการศึกษา	2.69		D	TC		B
ปริญญาตรี (ต่อเนื่อง/เทียบโอน)	สาขาการศึกษา	2.89		D+	TC		B+
ปริญญาตรี (ต่อเนื่อง/เทียบโอน)	สาขาการศึกษา	3.14		C+	TC		B
ปริญญาตรี (ต่อเนื่อง/เทียบโอน)	สาขาการศึกษา	2.54		D	TC		B
ปริญญาตรี (ต่อเนื่อง/เทียบโอน)	สาขาการศึกษา	3.47		D	TC		B+
ปริญญาตรี (ต่อเนื่อง/เทียบโอน)	สาขาการศึกษา	2.93		D+	TC		C+
ปริญญาตรี (ต่อเนื่อง/เทียบโอน)	สาขาการศึกษา	3.30		D	C+		A
ปริญญาตรี (ต่อเนื่อง/เทียบโอน)	สาขาการศึกษา	2.71		TC	TC		C+
ปริญญาตรี (ต่อเนื่อง/เทียบโอน)	สาขาการศึกษา	2.28		D	TC		D+
ปริญญาตรี (ต่อเนื่อง/เทียบโอน)	สาขาการศึกษา	3.09		D+	TC		A
ปริญญาตรี (ต่อเนื่อง/เทียบโอน)	สาขาการศึกษา	2.72		D	C+		A
ปริญญาตรี (ต่อเนื่อง/เทียบโอน)	สาขาการศึกษา	2.26		D	TC		D
ปริญญาตรี (ต่อเนื่อง/เทียบโอน)	สาขาการศึกษา	3.20		D	TC		B+
ปริญญาตรี (ต่อเนื่อง/เทียบโอน)	สาขาการศึกษา	2.83		C	TC		A
ปริญญาตรี (ต่อเนื่อง/เทียบโอน)	สาขาการศึกษา	2.74		D	TC		A
ปริญญาตรี (ต่อเนื่อง/เทียบโอน)	สาขาการศึกษา	3.01		B	TC		C+
ปริญญาตรี (ต่อเนื่อง/เทียบโอน)	สาขาการศึกษา	2.54		TC	TC		A
ปริญญาตรี (ต่อเนื่อง/เทียบโอน)	สาขาการศึกษา	3.07		D	TC		A

ตารางที่ 3.2 ข้อมูลเกรด (GPA) ของนักศึกษา

3.1.3.1.2 ทำการกลั่นกรองข้อมูล (Data Cleaning) คือการทำความสะอาดข้อมูลเป็นกระบวนการตรวจสอบและแก้ไข (หรือลบ) รายการข้อมูลที่ไม่ถูกต้องออกจากชุดข้อมูลฐานข้อมูล ซึ่งเป็นหลักสำคัญของฐานข้อมูล ทางผู้วิเคราะห์ได้ดำเนินการดังนี้

1) ทำการลบวิชาที่มีนักศึกษกลงทะเบียนเรียนน้อยกว่าปกติ และเกรดที่มีจำนวนซ้ำมากกว่าปกติ ยกตัวอย่างเช่นวิชาที่มีเกรด TC และ CE เป็นจำนวนมากกว่าเกรดปกติ และลบรายวิชาที่อยู่ในหมวดรายวิชาเสรีออก รวมถึงรายวิชาที่ให้เกรด S U และ I

3.1.3.1.3 การแปลงรูปแบบข้อมูล (Data Transformation) คือ การนำไฟล์ข้อมูลที่ได้จากการรวบรวมแปลงสภาพข้อมูลให้เหมาะสมกับเครื่องมือที่ใช้ในการวิเคราะห์ในขั้นตอน Clustering

ผู้วิเคราะห์ได้นำข้อมูลเกรด (GPA) ของนักศึกษา มาจัดรูปแบบให้อยู่ในรูปแบบของการนับเกรด A, B+, B, C+, C, D+, D และ F เพื่อนำมาคำนวณค่าเฉลี่ยถ่วงน้ำหนัก (Weighted Average) เพื่อนำมาใช้ในการ Cluster เพื่อแบ่งกลุ่มความยากง่ายของวิชาเรียนเพื่อนำมาวิเคราะห์ต่อ

1) กำหนดค่าน้ำหนักของแต่ละเกรดเพื่อนำมาคำนวณค่าเฉลี่ยถ่วงน้ำหนัก
กำหนดดังนี้

เกรด	น้ำหนัก
A	4
B+	3.5
B	3
C+	2.5
C	2
D+	1.5
D	1
F	0

ตารางที่ 3.3 การกำหนดน้ำหนักให้กับเกรดของนักศึกษา

2) คำนวณค่าเฉลี่ยถ่วงน้ำหนักโดยใช้การคำนวณดังนี้

$$WA = \frac{\sum(\text{จำนวนข้อมูลแต่ละตัว} * \text{น้ำหนักของข้อมูล})}{\text{ผลรวมของจำนวนทั้งหมด}}$$

3) นับจำนวนเกรดของนักศึกษา ในแต่ละวิชาตามลำดับคอลัมน์

เกรด	BBACA111 การบัญชีการเงิน	BBABA201 องค์กรและการจัดการ
A	11	0
B+	6	6
B	9	5
C+	16	19
C	17	24
D+	18	23
D	44	19
F	22	21

ตารางที่ 3.4 ตัวอย่างการนับเกรดของแต่ละวิชาเพื่อเปลี่ยนจากข้อมูลเกรด (GPA)

Course	A	B+	B	C+	C	D+	D	F
BBACA111 การบัญชีการเงิน	11	6	9	16	17	18	44	22
BBABA201 องค์กรและการจัดการ	0	6	5	19	24	23	19	21

ตารางที่ 3.5 ตัวอย่างที่ได้จากการเปลี่ยนรูปแบบข้อมูล

4) รวมจำนวนคนของแต่ละวิชาเพื่อนำมาคำนวณค่าเฉลี่ยถ่วงน้ำหนัก

(Weighted Average) กำหนดค่าที่ได้เป็น แอตทริบิวต์ N

Course	A	B+	B	C+	C	D+	D	F	N
BBACA111 การบัญชีการเงิน	11	6	9	16	17	18	44	22	143
BBABA201 องค์กรและการจัดการ	0	6	5	19	24	23	19	21	117

ตารางที่ 3.6 ตัวอย่างการนับจำนวนนักศึกษาที่ลงทะเบียนเรียนแต่ละวิชา

5) ตัวอย่าง วิธีคำนวณค่าเฉลี่ยถ่วงน้ำหนัก

$$WA = \frac{(11 \times 4) + (6 \times 3.5) + (9 \times 3) + (16 \times 2.5) + (17 \times 2) + (18 \times 1.5) + (44 \times 1) + (143 \times 0)}{143}$$

$$WA = \frac{237}{143}$$

$$WA = 1.65$$

ภาพที่ 3.2 ตัวอย่างการคำนวณค่าเฉลี่ยถ่วงน้ำหนัก

การจำแนกข้อมูลด้วยโมเดลการตัดสินใจแบบต้นไม้ (Classification with Decision tree Model)

3.1.3.2 การเตรียมข้อมูล (Data Preparation) สำหรับกระบวนการวิเคราะห์ต้นไม้ตัดสินใจ (Decision tree)

3.1.3.2.1 การแปลงรูปแบบข้อมูล (Data Transformation)

1) ปรับข้อมูลจากรูปแบบของ การนับเกรด A, B+, B, C+, C, D+, D และ F เป็นรูปแบบของเกรดนักศึกษา

Course	A	B+	B	C+	C	D+	D	F	N	WA
BBACA111 การบัญชีการเงิน	11	6	9	16	17	18	44	22	143	1.65
BBABA201 องค์การและการจัดการ	0	6	5	19	24	23	19	21	117	1.58

ตารางที่ 3.7 ตัวอย่างข้อมูลที่ทำการปรับ

ลำดับ	BBACA111 การบัญชีการเงิน	BBABA201 องค์การและการจัดการ	เกรด
1	C	B	2.15
2	A	B+	3.00

ตารางที่ 3.8 ตัวอย่างข้อมูลที่ปรับจากรูปแบบการนับเกรด

2) ทำการเพิ่มคอลัมน์เกณฑ์เกรดเฉลี่ย ดังนี้ เกรดเฉลี่ย ต่ำกว่า 2.00 ให้เป็น Low เกรดเฉลี่ย 2.00 ถึง 2.99 ให้เป็น Moderate เกรดเฉลี่ยมากกว่า 3.00 ให้เป็น High โดยใช้คอลัมน์เกรดในการเปลี่ยน

เกณฑ์การวัด	ช่วงเกรดเฉลี่ย
High	3.00–4.00
Moderate	2.00–2.99
Low	0.00–1.99

ตารางที่ 3.9 ตัวอย่างข้อมูลที่เพิ่มในคอลัมน์เกณฑ์เกรดเฉลี่ยในฐานข้อมูล

ลำดับ	เกรด	Value	BBACA111 การบัญชีการเงิน	BBABA201 องค์การและการจัดการ
1	2.15	Moderate	C	B
2	3.00	High	A	B+

ตารางที่ 3.10 ตัวอย่างข้อมูลที่พร้อมใช้ในการวิเคราะห์

3.1.3.2.2 ทำการคัดเลือกข้อมูล (Data Selection) คือการคัดเลือกข้อมูลที่เหมาะสมเพื่อนำมาใช้สำหรับกระบวนการวิเคราะห์ต้นไม้ตัดสินใจ (Decision tree) โดยการเลือกเฉพาะกลุ่มเรียนวิชาที่ยากเท่านั้น อ้างอิงข้อมูลจากการจัดกลุ่ม (Cluster)

1) เลือกข้อมูลเกรดแต่ละวิชาของนักศึกษาเอกวิชาการจัดการ
สารสนเทศทางธุรกิจ (IM) เพื่อสร้างโมเดล Decision tree

	A	B	C	D	E	F	G
	Major	Grad	value	BACAC111 การบัญชีการเงิน	BBABA201 องค์การและการจัดการ	BBABA201 องค์การและการจัดการ	BBB3C103 การคิดและนำเสนองานในเชิงวิพากษ์
1	M	3.35	high	CE	TC	TC	N/A
2	M	2.69	moderate	CE	TC	TC	N/A
3	M	3.25	high	CE	TC	CE	N/A
4	M	2.59	moderate	CE	TC	CE	N/A
5	M	2.29	moderate	CE	TC	CE	N/A
6	M	3.17	high	CE	TC	CE	N/A
7	M	3.67	high	CE	TC	TC	N/A
8	M	3.12	high	CE	TC	TC	N/A
9	M	3.32	high	CE	TC	TC	N/A
10	M	3.05	moderate	CE	TC	TC	N/A
11	M	3.24	high	CE	TC	TC	N/A
12	M	2.07	moderate	CE	TC	CE	N/A
13	M	3.19	high	CE	TC	TC	N/A
14	M	3.45	high	CE	TC	CE	N/A
15	M	2.35	moderate	CE	TC	TC	N/A
16	M	3.08	high	CE	TC	TC	N/A
17	M	2.74	moderate	CE	CE	CE	N/A
18	M	2.75	moderate	TC	CE	CE	N/A
19	M	3.01	high	TC	TC	TC	N/A

ตารางที่ 3.11 ข้อมูลเกรดแต่ละวิชาของนักศึกษาเอกวิชาการจัดการสารสนเทศทางธุรกิจ (IM)

2) เลือกข้อมูลเกรดแต่ละวิชาของนักศึกษาเอกวิชาการพัฒนา
ซอฟต์แวร์ทางธุรกิจ (SD) เพื่อสร้างโมเดล Decision tree

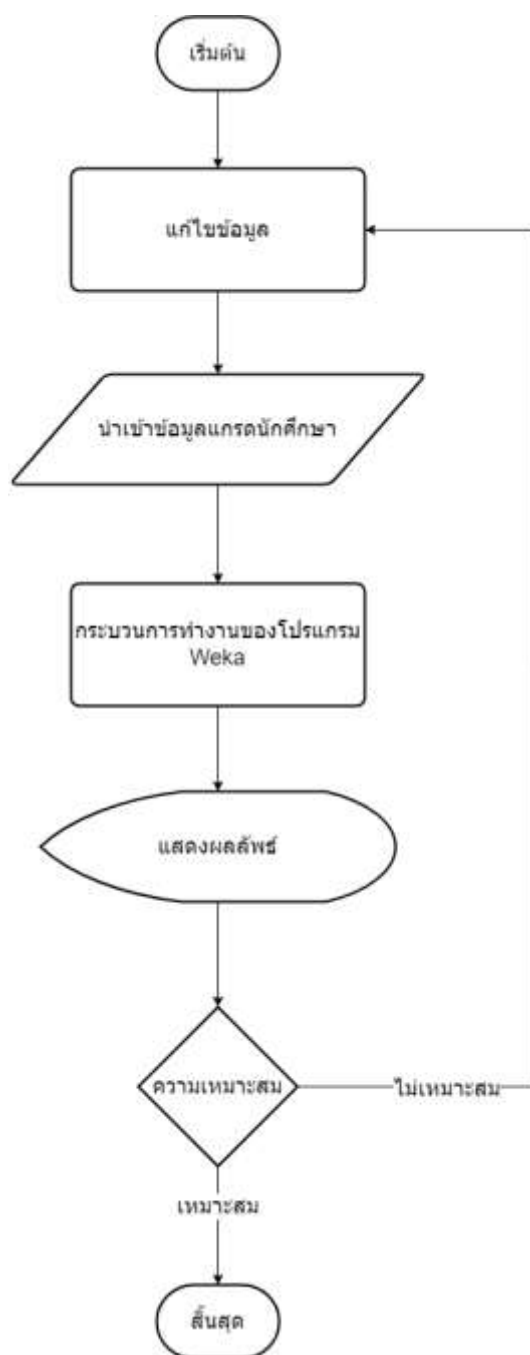
	A	B	C	D	E	F	G
	Major	Grad	value	BACAC111 การบัญชีการเงิน	BBABA201 องค์การและการจัดการ	BBACC107 การเงินธุรกิจ	BBB3C103 โครงสร้างข้อมูลและอัลกอริทึม
1	SD	2.96	moderate	D+	TC	B+	D+
2	SD	3.37	high	TC	TC	C+	C
3	SD	2.84	moderate	C+	TC	N/A	D+
4	SD	3.60	high	C+	TC	C+	B
5	SD	2.44	moderate	D	TC	D+	D+
6	SD	2.69	moderate	D	TC	D	D+
7	SD	2.89	moderate	D+	TC	C+	D+
8	SD	2.80	moderate	TC	TC	N/A	D+
9	SD	3.14	high	C+	TC	C+	B
10	SD	2.54	moderate	D	TC	D+	D+
11	SD	3.67	high	D	TC	C	B+
12	SD	2.99	moderate	D+	TC	C	C+
13	SD	3.30	high	D	C+	C+	C+
14	SD	2.59	moderate	W	C	N/A	D+
15	SD	2.71	moderate	TC	TC	D+	D+
16	SD	0.50	low	TC	TC	N/A	F
17	SD	2.79	moderate	N/A	TC	N/A	D+
18	SD	2.57	moderate	TC	TC	D	C+
19	SD	2.28	moderate	D	TC	C	C

ตารางที่ 3.12 ข้อมูลเกรดแต่ละวิชาของนักศึกษาเอกวิชาการพัฒนาซอฟต์แวร์ทางธุรกิจ (SD)

3.1.3.2.3 ทำการกลั่นกรองข้อมูล (Data Cleaning)

- 1) ลบชื่อรายวิชาภาษาไทยออกเพื่อสำหรับกระบวนการวิเคราะห์ต้นไม้ตัดสินใจ (Decision tree) โดยใช้โปรแกรม WEKA

3.1.4 สร้างแบบจำลอง (Modeling) ขั้นตอนการสร้างตัวแบบทางคณิตศาสตร์ และสถิติ เพื่อการวิเคราะห์ข้อมูล โดยสามารถใช้เทคนิควิธีการต่าง ๆ อาทิ การแบ่งกลุ่ม (Clustering) และ ต้นไม้ตัดสินใจ (Decision tree)



ภาพที่ 3.3 Flowchart การทำงานของ Cluster ของโปรแกรม WEKA

ผลลัพธ์ของการ Cluster โปรแกรม WEKA

```

Cluster output
Scheme: weka.clusterers.SimpleKMeans -init 0 -max-candidates 100 -periodic-pruning
Relation: all data cluster
Instances: 63
Attributes: 2
  Course
  AVG
Test mode: evaluate on training data

=== Clustering model (full training set) ===

KMeans
=====

Number of iterations: 5
Within cluster sum of squared errors: 60.54858432460038

Initial starting points (random):

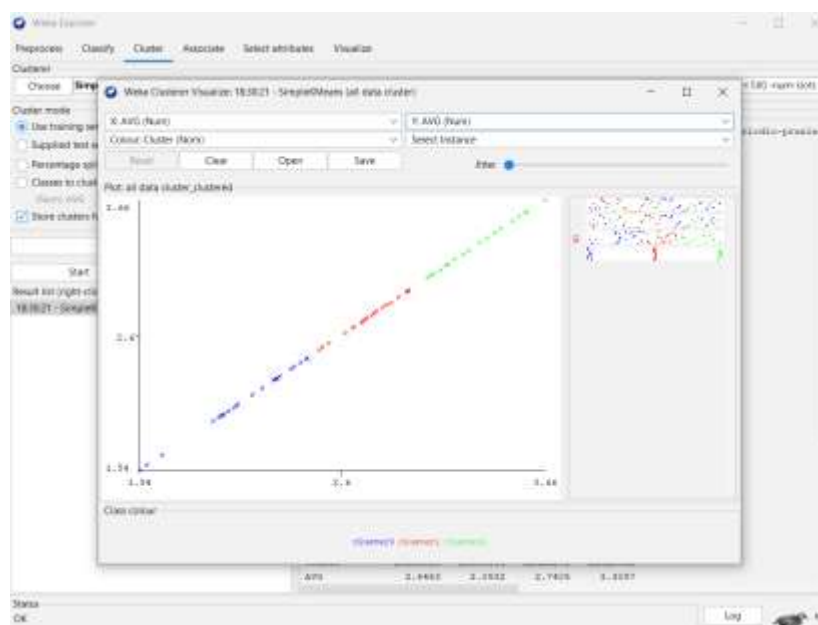
Cluster 0: BBACC110,2.782828
Cluster 1: SEBAC105,2.833333
Cluster 2: BBAB106,3.154321

Missing values globally replaced with mean/mode

Final cluster centroids:

Attribute      Full Data      Cluster#
              (63.0)        (25.0)        (20.0)        (19.0)
-----
Course         BACAC111      BACAC111      BBABA202      BBABA612
AVG            2.6463        2.0932        2.7405        3.3097
  
```

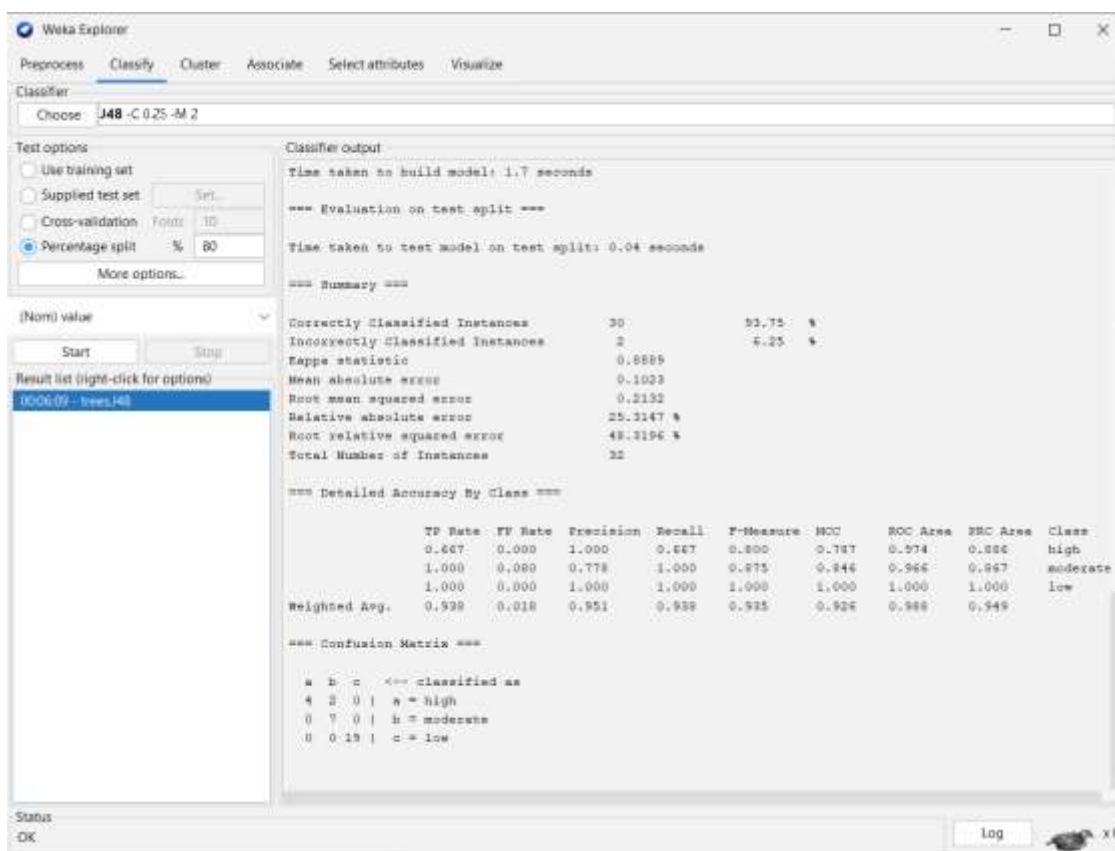
ภาพที่ 3.4 ผลการแบ่งกลุ่มแบบ SimpleKMeans ในโปรแกรม WEKA



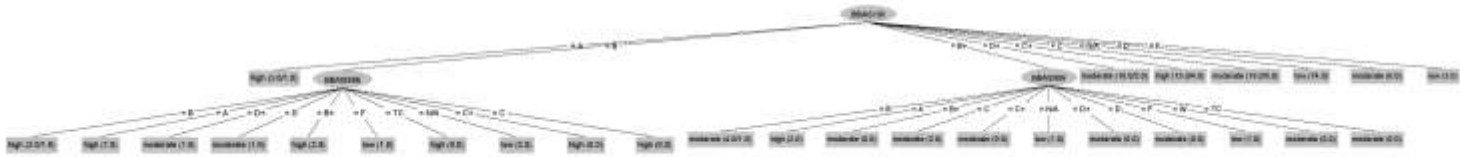
ภาพที่ 3.5 ผล Cluster Visualize จากการวิเคราะห์ข้อมูลในโปรแกรม WEKA

3.1.4.2 การสร้างโมเดล Decision tree โดยใช้ข้อมูลรายวิชาที่อยู่ในกลุ่ม ยาก ที่ได้จากการแบ่งกลุ่ม Clustering โดยสร้างโมเดล Decision tree ผ่านโปรแกรม WEKA

การสร้างโมเดล Decision tree ของเอกวิชาการจัดการสารสนเทศทางธุรกิจ แบ่งข้อมูลในการสร้างโมเดล Decision tree 80% และหาค่าพยากรณ์ (Prediction) 20% ค่าความแม่นยำของโมเดล Decision tree 93.75%



ภาพที่ 3.6 ผลสรุปของการสร้างโมเดล Decision tree เอกวิชาการจัดการสารสนเทศทางธุรกิจ



ภาพที่ 3.7 ผลลัพธ์ของโมเดล Decision tree เอกวิชาการจัดการสารสนเทศทางธุรกิจ

การสร้างโมเดล Decision tree ของเอกวิชาการพัฒนาซอฟต์แวร์ทางธุรกิจ แบ่งข้อมูลในการสร้างโมเดล Decision tree 80% และหาค่าพยากรณ์ (Prediction) 20% ค่าความแม่นยำของโมเดล Decision tree 83.33%

Classifier output

Time taken to build model: 0 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Correctly Classified Instances	15	83.3333 %
Incorrectly Classified Instances	3	16.6667 %
Kappa statistic	0.7331	
Mean absolute error	0.1423	
Root mean squared error	0.2554	
Relative absolute error	31.1517 %	
Root relative squared error	60.7411 %	
Total Number of Instances	18	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PNC Area	Class
high	0.750	0.100	0.857	0.750	0.800	0.463	0.906	0.855	high
moderate	0.667	0.133	0.500	0.667	0.571	0.478	0.811	0.600	moderate
low	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	low
Weighted Avg.	0.833	0.067	0.853	0.833	0.848	0.743	0.927	0.869	

=== Confusion Matrix ===

```

a b c  <-- classified as
6 2 0 | a = high
1 2 0 | b = moderate
0 0 7 | c = low
    
```

ภาพที่ 3.8 ผลลัพธ์ของโมเดล Decision tree เอกวิชาการพัฒนาซอฟต์แวร์ทางธุรกิจ



ภาพที่ 3.9 ผลลัพธ์ของโมเดล Decision tree เอกวิชาการพัฒนาซอฟต์แวร์ทางธุรกิจ

3.1.5 การประเมินผล (Evaluation) และกฎที่ได้

การสร้างโมเดล Decision tree ที่ได้ทำการแบ่งข้อมูลที่ใช้ในการสร้างโมเดล 80% และ 20% ใช้ในการหาค่าพยากรณ์ (Prediction)

กฎของโมเดล Decision tree เอกวิชาการจัดการสารสนเทศทางธุรกิจ

กฎที่ 1 if BBAIS105 = A than High หมายความว่า ถ้าวิชา คลังข้อมูลและเหมือนข้อมูล ได้เกรด A ผลลัพธ์ที่ได้ เกรดเฉลี่ยสูง

กฎที่ 2 if BBAIS105 = B and BBAIS806 = A,B+,B,C+,C than High หมายความว่า ถ้าวิชา คลังข้อมูลและเหมือนข้อมูล ได้เกรด B และวิชา การประยุกต์ใช้คอมพิวเตอร์ในงานสำนักงาน ได้เกรด A,B+,B,C+,C ผลลัพธ์ที่ได้ เกรดเฉลี่ยสูง

กฎที่ 3 if BBAIS105 = B and BBAIS806 = D+,D than Moderate หมายความว่า ถ้าวิชา คลังข้อมูลและเหมือนข้อมูล ได้เกรด B และวิชา การประยุกต์ใช้คอมพิวเตอร์ในงานสำนักงาน ได้เกรด D+,D ผลลัพธ์ที่ได้ เกรดเฉลี่ยปานกลาง

กฎที่ 4 if BBAIS105 = B and BBAIS806 = F,W,N/A than Low หมายความว่า ถ้าวิชา คลังข้อมูลและเหมือนข้อมูล ได้เกรด B และวิชา การประยุกต์ใช้คอมพิวเตอร์ในงานสำนักงาน ได้เกรด F,W,N/A ผลลัพธ์ที่ได้ เกรดเฉลี่ยต่ำ

กฎที่ 5 if BBAIS105 = B+ and BBAIS906 = A than High หมายความว่า ถ้าวิชา คณิตข้อมูลและเหมืองข้อมูล ได้เกรด B+ และวิชา การออกแบบและพัฒนาเว็บไซต์ ได้เกรด A ผลลัพธ์ที่ได้ เกรดเฉลี่ยสูง

กฎที่ 6 if BBAIS105 = B+ and BBAIS906 = B+,B,C+,C,D+,D,TC than Moderate หมายความว่า ถ้าวิชา คณิตข้อมูลและเหมืองข้อมูล ได้เกรด B+ และวิชา การออกแบบและพัฒนาเว็บไซต์ ได้เกรด B+,B,C+,C,D+,D,TC ผลลัพธ์ที่ได้ เกรดเฉลี่ยปานกลาง

กฎที่ 7 if BBAIS105 = B+ and BBAIS906 = F,W,N/A than Low หมายความว่า ถ้าวิชา คณิตข้อมูลและเหมืองข้อมูล ได้เกรด B+ และวิชา การออกแบบและพัฒนาเว็บไซต์ ได้เกรด F,W,N/A ผลลัพธ์ที่ได้ เกรดเฉลี่ยต่ำ

กฎที่ 8 if BBAIS105 = C+ than High หมายความว่า ถ้าวิชา คณิตข้อมูลและเหมืองข้อมูล ได้เกรด C+ ผลลัพธ์ที่ได้ เกรดเฉลี่ยสูง

กฎที่ 9 if BBAIS105 = C,D+,D than Moderate หมายความว่า ถ้าวิชา คณิตข้อมูลและเหมืองข้อมูล ได้เกรด C,D+,D ผลลัพธ์ที่ได้ เกรดเฉลี่ยปานกลาง

กฎที่ 10 if BBAIS105 = F,W,N/A than Low หมายความว่า ถ้าวิชา คณิตข้อมูลและเหมืองข้อมูล ได้เกรด F,W,N/A ผลลัพธ์ที่ได้ เกรดเฉลี่ยต่ำ

กฎของโมเดล Decision tree เอกวิชาการพัฒนาซอฟต์แวร์ทางธุรกิจ

กฎที่ 1 if BBAIS907 = C+ than High หมายความว่า ถ้าวิชา การเขียนโปรแกรมบนเว็บ ได้เกรด C+ ผลลัพธ์ที่ได้ เกรดเฉลี่ยสูง

กฎที่ 2 if BBAIS907 = A and BBAIS903 = A,B+,B than High หมายความว่า ถ้าวิชา การเขียนโปรแกรมบนเว็บ ได้เกรด A และวิชา โครงสร้างข้อมูลและอัลกอริทึม ได้เกรด A,B+,B ผลลัพธ์ที่ได้ เกรดเฉลี่ยสูง

กฎที่ 3 if BBAIS907 = A and BBAIS903 = C+,C,D+,D than Moderate หมายความว่า ถ้าวิชา การเขียนโปรแกรมบนเว็บ ได้เกรด A และวิชา โครงสร้างข้อมูลและอัลกอริทึม ได้เกรด C+,C,D+,D ผลลัพธ์ที่ได้ เกรดเฉลี่ยปานกลาง

กฎที่ 4 if BBAIS907 = A and BBAIS903 = F,W,N/A than Low หมายความว่า ถ้าวิชา การเขียนโปรแกรมบนเว็บ ได้เกรด A และวิชา โครงสร้างข้อมูลและอัลกอริทึม ได้เกรด F,W,N/A ผลลัพธ์ที่ได้ เกรดเฉลี่ยต่ำ

กฎที่ 5 if BBAIS907 = C and BBAIS904 = A,B+ than High หมายความว่า ถ้าวิชา การเขียนโปรแกรมบนเว็บ ได้เกรด C และวิชา การเขียนโปรแกรมคอมพิวเตอร์ ได้เกรด A,B+ ผลลัพธ์ที่ได้ เกรดเฉลี่ยสูง

กฎที่ 6 if BBAIS907 = C and BBAIS904 = B,C+,C,D+,D than Moderate หมายความว่า ถ้าวิชา การเขียนโปรแกรมบนเว็บ ได้เกรด C และวิชา การเขียนโปรแกรมคอมพิวเตอร์ ได้เกรด B,C+,C,D+,D ผลลัพธ์ที่ได้ เกรดเฉลี่ยปานกลาง

กฎที่ 7 if BBAIS907 = C and BBAIS904 = F,W,N/A than Low หมายความว่า ถ้าวิชา การเขียนโปรแกรมบนเว็บ ได้เกรด C และวิชา การเขียนโปรแกรมคอมพิวเตอร์ ได้เกรด F,W,N/A ผลลัพธ์ที่ได้ เกรดเฉลี่ยต่ำ

กฎที่ 8 if BBAIS907 = B+,B than High หมายความว่า ถ้าวิชา การเขียนโปรแกรมบนเว็บ ได้เกรด B+,B ผลลัพธ์ที่ได้ เกรดเฉลี่ยสูง

กฎที่ 9 if BBAIS907 = D+,D than Moderate หมายความว่า ถ้าวิชา การเขียนโปรแกรมบนเว็บ ได้เกรด D+,D ผลลัพธ์ที่ได้ เกรดเฉลี่ยปานกลาง

กฎที่ 10 if BBAIS907 = F,W,N/A than Low หมายความว่า ถ้าวิชา การเขียนโปรแกรมบนเว็บ ได้เกรด F,W,N/A ผลลัพธ์ที่ได้ เกรดเฉลี่ยต่ำ

3.1.6 เผยแพร่ผลวิเคราะห์ (Deployment) ขั้นตอนการนำผลลัพธ์ที่ได้ไปใช้งานเป็นการทั่วไป อาจจัดทำเป็นรูปแบบของรายงาน (Report) หรือแผนภาพ (Dashboard) ที่พร้อม นำไปใช้ประโยชน์ในการวางแผน กำหนดกลยุทธ์ และดำเนินการต่าง ๆ ในทางธุรกิจ คณะผู้จัดทำนำผลข้อมูลที่ทำการวิเคราะห์แล้ว แสดงผลข้อมูลบน Web browser โดยใช้ชุดคำสั่ง HTML และ CSS ด้วยการนำเสนอข้อมูลแบบ Visualization เป็นการแสดงผล ข้อมูลในรูปแบบของภาพโดยใช้โปรแกรม Power Bi ซึ่งทางคณะผู้จัดทำได้ยกตัวอย่างการ จัดทำเป็นรูปแบบของรายงาน (Report) หรือแผนภาพ (Dashboard) กับข้อมูลเกรดนักศึกษิตตามรายวิชา โดยแต่ละวิชาจะมีรูปแบบเดียวกัน ดังนี้

3.2 การออกแบบเว็บไซต์

3.2.1 หน้าแรกของเว็บไซต์เป็นรายละเอียดของเว็บไซต์

ปัจจัยที่มีอิทธิพลต่อความสำเร็จของนักศึกษาโดยใช้วิธีการเหมือนข้อมูล

เว็บไซต์นี้จัดทำขึ้นเพื่อเป็นเว็บไซต์สำหรับนักศึกษาคุณแม่วิทยาลัยและสภามหาวิทยาลัยระบบสารสนเทศเพื่อทางธุรกิจ เพื่อให้ความรู้ความเข้าใจ และกำหนดความพร้อมสำหรับการลงคะแนนเรียนรายวิชา ทั้งส่วนนักศึกษาที่ไม่ได้เข้าวิชาที่นักศึกษาต้องการจะลงคะแนนเรียนมีความยากง่ายอย่างไรบ้าง และ นักศึกษาที่คาดหวังความรู้ความเข้าใจเกี่ยวกับรายวิชาในระบบสารสนเทศ โดยเว็บไซต์นี้จะให้ความรู้และอำนวยความสะดวกในการลงคะแนนเรียนของนักศึกษาได้ 2 ทางคือ

- **ข้อเสนอแนะการเรียนของนักศึกษา**
หน้าเว็บไซต์มีข้อมูลแนะนำการเรียนของนักศึกษา 10 อันดับแรก นักศึกษาสามารถตรวจสอบผลการเรียนรายวิชา 2540 - 2541 ผลของการวิเคราะห์ผลการเรียน (Cluster) ในรูปแบบ K-Mean เพื่อจัดกลุ่มรายวิชาให้เหมาะสม ปานกลาง และง่าย และมีการนำเสนอแบบเรียลไทม์ของเกรดวิชา หรือคะแนนเรียนของรายวิชา โดยใช้โปรแกรม Power BI ในการนำเสนอ
- **เสนอข้อมูลวิเคราะห์ความสำเร็จของนักศึกษาถึงคุณครูระบบสารสนเทศ**
หน้าเว็บไซต์มีแบบฟอร์มสำหรับส่งข้อเสนอแนะถึงคุณครูระบบสารสนเทศ ซึ่งการนำข้อมูลที่ได้จากการวัดค่าโดยคอมพิวเตอร์มาวิเคราะห์ที่ระบบที่เป็น 3 มิติสามารถวิเคราะห์ได้ (Dimension free) ซึ่งค่าวิเคราะห์ที่มีประสิทธิภาพที่สุดคือค่าของการเรียนของนักศึกษาที่จุด สัมผัสทุกจุดที่นำมาเป็น เป็นแบบกราฟแสดงในการบริหารการจัดการด้านอื่นในอนาคต

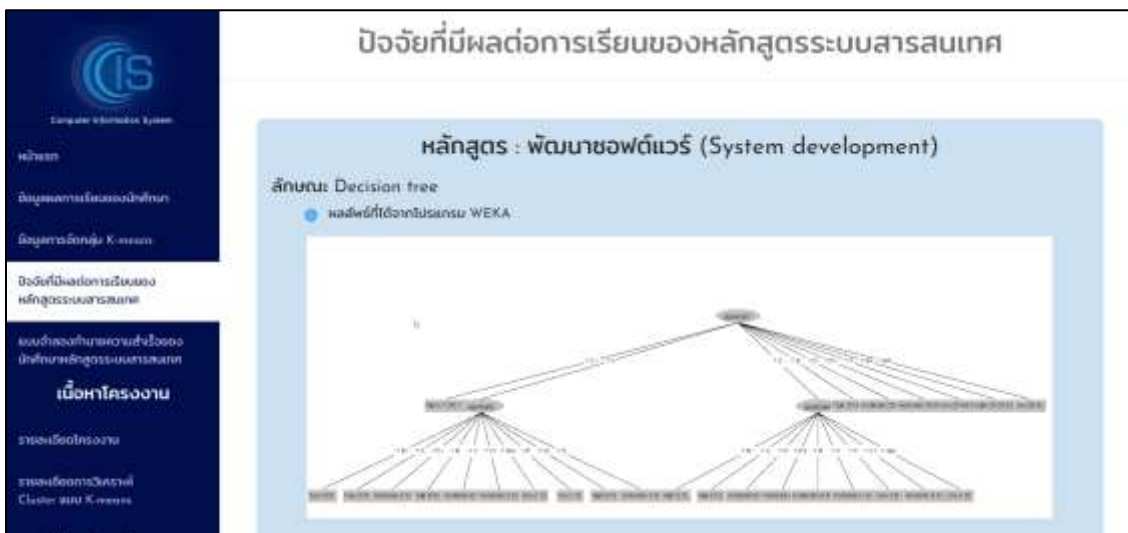
เนื้อหาโครงการงาน

รายละเอียดโครงการงาน

Cluster และ K-Mean

ภาพที่ 3.10 หน้าแรกของเว็บไซต์เป็นรายละเอียดโครงการงาน

3.2.4 หน้าเว็บไซต์ที่ 4 แสดงผลลัพธ์ผลของการวิเคราะห์ Decision tree



ภาพที่ 3.13 ผลลัพธ์ผลของการวิเคราะห์ Decision tree

3.2.5 หน้าเว็บไซต์ที่ 5 แสดงแบบจำลองทำนายความสำเร็จของนักศึกษา

The screenshot shows a web interface for a student success prediction model. The title is 'แบบจำลองทำนายความสำเร็จของนักศึกษาหลักสูตรระบบสารสนเทศ' (Student success prediction model for the Information Systems course). The main heading is 'หลักสูตร : ระบบสารสนเทศ' (Information Systems course). The interface is divided into two main sections: 'หลักสูตร : พัฒนาซอฟต์แวร์' (System development) and 'หลักสูตร : ระบบสารสนเทศ' (Information Systems). Each section contains a table with columns for 'รหัสวิชา' (Course Code), 'ชื่อวิชา' (Course Name), and 'เกรด' (Grade). Below each table is a green button labeled 'ทำนายผล' (Predict).

รหัสวิชา	ชื่อวิชา	เกรด
BBA15901	การเขียนโปรแกรมบนเว็บ	A
BBA15903	โครงสร้างข้อมูลและอัลกอริทึม	A
BBA15904	การเขียนโปรแกรมคอมพิวเตอร์	A

รหัสวิชา	ชื่อวิชา	เกรด
BBA15905	คลังข้อมูลและเหมืองข้อมูล	A
BBA15906	การออกแบบและพัฒนาเว็บไซต์	A
BBA15906	การประยุกต์ใช้คอมพิวเตอร์ในงานสำนักงาน	A

ภาพที่ 3.14 แบบจำลองทำนายความสำเร็จของนักศึกษา

3.2.8 หน้าเว็บที่ 8 แสดงขั้นตอนการวิเคราะห์ ต้นไม้การตัดสินใจ (Decision tree)

รายละเอียดการวิเคราะห์ Decision tree

ปรับข้อมูลจากรูปแบบของ การเป็นเกรด A, B+, B, C+, C, D+, D และ F เป็นรูปแบบของเกรดนักศึกษา

— ข้อมูลก่อนปรับ —

วิชา	A	B+	B	C+	C	D+	D	F	N
BBACA311 การบัญชีการเงิน	11	6	9	16	17	18	44	23	143
BBABA201 องค์การและการจัดการ	8	6	3	18	34	23	19	31	127

— ข้อมูลหลังปรับ —

ลำดับ	BBACA311 การบัญชีการเงิน	BBABA201 องค์การและการจัดการ	เกรด
1	C	B	212
2	A	B+	300

ทำการเพิ่มเงินแก่นักเรียนที่เกรดเฉลี่ย ดังนี้ เกรดเฉลี่ย ต่ำกว่า 2.00 ให้เป็น Low เกรดเฉลี่ย 2.00 ถึง 2.99 ให้เป็น Moderate เกรดเฉลี่ย มากกว่า 3.00 ให้เป็น High โดยใช้ค่าเงินเพิ่มกรณีการเปลี่ยน

— ข้อมูลที่พร้อมนำไปใช้ —

ลำดับ	เกรด	Value	BBACA311 การบัญชีการเงิน	BBABA201 องค์การและการจัดการ
1	212	Moderate	C	B
2	300	High	A	B+

ขั้นตอนการวิจัยโปรแกรม WEKA

ภาพที่ 3.17 แสดงขั้นตอนการวิเคราะห์ ต้นไม้การตัดสินใจ (Decision tree)

3.2.9 หน้าเว็บที่ 9 แสดงรายละเอียดของผู้จัดทำโครงการ

ผู้จัดทำ



ชื่อ นามสกุล อ้นนพฉัตร ชอนฉัตรคาร์น

ผู้จัดทำ



ชื่อ นามสกุล ณัฐวรงค์ ฉานเอี

ผู้จัดทำ

ภาพที่ 3.18 แสดงรายละเอียดของผู้จัดทำโครงการ

3.2.10 หน้าเว็บที่ 10 แบบสอบถามความพึงพอใจเว็บไซต์

ปัจจัยที่มีอิทธิพลต่อความสำเร็จของนักศึกษาโดยใช้วิธีเหมืองข้อมูล

แบบสอบถามความพึงพอใจ
 วัตถุประสงค์ของแบบสอบถาม: วัตถุประสงค์ของแบบสอบถาม

ชื่อ:

ชื่อจริง

เพศ *

ชาย

หญิง

อีเมล *

ภาพที่ 3.20 แบบสอบถามความพึงพอใจเว็บไซต์