

บทที่ 3

วิธีดำเนินการวิจัย

การทำวิจัยครั้งนี้ผู้วิจัยได้ศึกษาแนวคิด ทฤษฎี หลักการ ตลอดจนเอกสาร และงานวิจัยที่เกี่ยวข้องตามที่ได้กล่าวในบทที่ 2 เพื่อเป็นพื้นฐานความรู้สำหรับการทำวิจัย และในบทนี้ได้นำเอาความรู้ที่ได้มาวิเคราะห์วิธีการทำงานของการพยากรณ์ผลสัมฤทธิ์ทางการเรียนด้วยเทคนิคเหมืองข้อมูล ผู้วิจัยจะกล่าวถึงเครื่องมือที่ใช้ในงานวิจัย ขั้นตอนการดำเนินงานวิจัย ดังหัวข้อที่จะอธิบายในลำดับต่อไป

เครื่องมือที่ใช้ในงานวิจัย

อุปกรณ์ที่ใช้ในการดำเนินงานวิจัย ประกอบด้วย เครื่องคอมพิวเตอร์โน้ตบุ๊กจำนวน 1 เครื่อง โดยมีคุณสมบัติ ดังนี้

1. ส่วนของฮาร์ดแวร์

1.1 หน่วยประมวลผลกลาง (CPU) Core(TM) i5 2.30 GHz

1.2 หน่วยความจำหลัก (RAM) 4 GB DDR3

1.3 หน่วยความจำสำรอง (HDD) 500 GB

2. ส่วนของซอฟต์แวร์

2.1 ระบบปฏิบัติการ (OS) Windows 7 Ultimate

2.2 โปรแกรมสำเร็จรูป WEKA 3.6.5 ใช้ในการหาตัวแปรที่มีอิทธิพลในการพยากรณ์ผลสัมฤทธิ์ทางการเรียน หาค่าความถูกต้องของข้อมูล และใช้สำหรับหากฎความสัมพันธ์

2.3 โปรแกรมสำเร็จรูป SPSS 11.5 for Windows ใช้สำหรับทดสอบค่าทางสถิติ

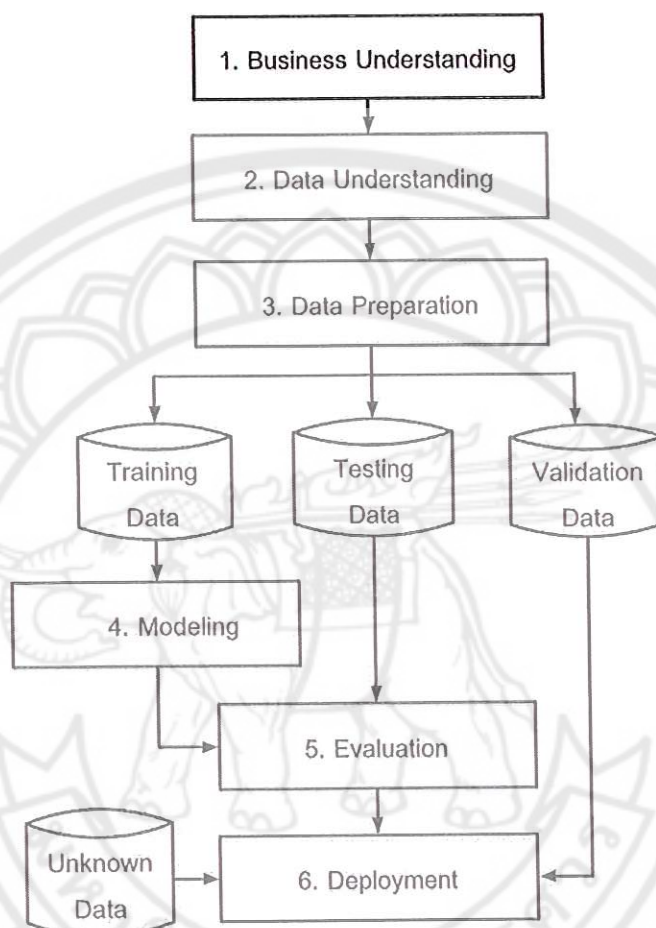
2.4 โปรแกรม Microsoft Excel 2010 ใช้เป็นเครื่องมือที่ช่วยในการรวมข้อมูล ทำความสะอาดข้อมูล แปลงข้อมูล

2.5 โปรแกรม Notepad++ ใช้เก็บข้อมูลในรูปแบบเท็กซ์ไฟล์

ขั้นตอนการดำเนินงานวิจัย

ขั้นตอนการดำเนินงานวิจัยสำหรับการพยากรณ์ผลสัมฤทธิ์ทางการเรียนด้วยเทคนิคเหมืองข้อมูลครั้งนี้ ผู้วิจัยได้ดำเนินการตามแนวทางของกระบวนการสำหรับการทำเหมืองข้อมูล

(CRISP-DM Model) (Roiger, et al., 2003; Chapman, et al., 2000) ประกอบไปด้วย 6 ขั้นตอน ดังภาพ 23 มีรายละเอียดการดำเนินการแต่ละขั้นตอน ดังนี้



ภาพ 23 แสดงขั้นตอนการดำเนินการวิจัย

ขั้นตอนที่ 1 ทำความเข้าใจโจทย์ที่ต้องการทำ

ผู้วิจัยได้ทำการศึกษาโครงสร้างหลักสูตรวิทยาศาสตรบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์ และสาขาวิชาเทคโนโลยีสารสนเทศ ของคณะวิทยาศาสตร์ มหาวิทยาลัยนเรศวร เริ่มรับนิสิตเข้าศึกษาหลักสูตรวิทยาศาสตรบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์ ตั้งแต่ พ.ศ. 2537 และเริ่มรับนิสิตสาขาวิชาเทคโนโลยีสารสนเทศ ตั้งแต่ พ.ศ. 2546 โดยมีภาควิชาวิทยาการคอมพิวเตอร์ และเทคโนโลยีสารสนเทศ เป็นหน่วยงานที่รับผิดชอบบริหารจัดการหลักสูตร และได้ปรับปรุงหลักสูตรให้ทันสมัยอยู่เสมอ โดยหลักสูตรวิทยาศาสตรบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์ และสาขาวิชาเทคโนโลยีสารสนเทศ หลักสูตรปรับปรุง พ.ศ. 2548 ดังตาราง 3 และ 4

เป็นหลักสูตรที่ใช้ผลิตบัณฑิตตั้งแต่ปี 48 เป็นต้นมา จากการสังเกตจำนวนนิสิตสาขาวิชา วิทยาการคอมพิวเตอร์ และสาขาวิชาเทคโนโลยีสารสนเทศ พบว่า มีนิสิตจำนวนหนึ่งต้องพึ่งสภาพ การเป็นนิสิต จบการศึกษาช้ากว่าเวลาที่กำหนด หรือได้ผลสัมฤทธิ์ทางการเรียนที่ต่ำกว่าที่คาดหวัง นอกจากนั้น ยังมีนิสิตจำนวนบางคนต้องการย้ายสาขา อันเนื่องมาจากการเลือกสาขาไม่ตรงกับ ความถนัดของตนเอง ดังนั้นผู้วิจัยจึงมีความต้องการที่จะทำการพยากรณ์ผลสัมฤทธิ์ทางการเรียน ด้วยเทคนิคเหมืองข้อมูล

ตาราง 3 แสดงโครงสร้างหลักสูตร ปรับปรุง พ.ศ. 2548 หลักสูตรปริญญาตรี
วิทยาศาสตร์บัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์

รายการ	เกณฑ์ สกอ.	โครงสร้างหลักสูตร
1. หมวดวิชาศึกษาทั่วไป ไม่น้อยกว่า	30	30
2. หมวดวิชาเฉพาะสาขา ไม่น้อยกว่า	84	93
2.1 กลุ่มวิชาพื้นฐานวิทยาศาสตร์ และคณิตศาสตร์		17
2.2 กลุ่มวิชาพื้นฐานเฉพาะด้าน		19
2.3 กลุ่มวิชาเอกบังคับ		39
2.4 กลุ่มวิชาเอกเลือก ไม่น้อยกว่า		18
3. หมวดวิชาเลือกเสรี ไม่น้อยกว่า	6	6
จำนวนหน่วยกิตรวมตลอดหลักสูตรไม่น้อยกว่า	120-150	129

ตาราง 4 แสดงโครงสร้างหลักสูตร ปรับปรุง พ.ศ. 2548 หลักสูตรปริญญาตรี
วิทยาศาสตร์บัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ

รายการ	เกณฑ์ สกอ.	โครงสร้างหลักสูตร
1. หมวดวิชาศึกษาทั่วไป ไม่น้อยกว่า	30	30
2. หมวดวิชาเฉพาะสาขา ไม่น้อยกว่า	84	93
2.1 กลุ่มวิชาพื้นฐานวิทยาศาสตร์ และคณิตศาสตร์		17
2.2 กลุ่มวิชาพื้นฐานเฉพาะด้าน		18
2.3 กลุ่มวิชาเอกบังคับ		43
2.4 กลุ่มวิชาเอกเลือก ไม่น้อยกว่า		15
3. หมวดวิชาเลือกเสรี ไม่น้อยกว่า	6	6
จำนวนหน่วยกิตรวมตลอดหลักสูตร	120-150	129

ขั้นตอนที่ 2 ทำความเข้าใจข้อมูลที่ต้องใช้ในการวิจัย

จากการศึกษางานวิจัยที่มีนักวิจัยเผยแพร่ผลงานมาก่อนแล้ว ผู้วิจัยที่ทำวิจัยเกี่ยวกับการศึกษาได้นำข้อมูลมาใช้อย่างหลากหลาย ซึ่งแต่ละท่านใช้ข้อมูลที่แตกต่างกันออกไป ในการทำวิจัยครั้งนี้ ผู้วิจัยจึงได้เลือกใช้ข้อมูลภูมิหลัง และข้อมูลผลการเรียนในช่วงชั้นปีที่ 1 และ 2 ของนิสิต สาขาวิชาวิทยาการคอมพิวเตอร์ และสาขาวิชาเทคโนโลยีสารสนเทศ มาใช้ในการพยากรณ์ ผลสัมฤทธิ์ทางการเรียนด้วยเทคนิคเหมืองข้อมูล

ตาราง 5 แสดงตัวแปรของนิสิตสาขาวิชาวิทยาการคอมพิวเตอร์ ช่วงชั้นปีที่ 1 และ 2

ลำดับที่	ชื่อตัวแปร	รายละเอียด	ค่าข้อมูล
1	Entrance	วิธีรับเข้า	Nominal
2	Sex	เพศ	Nominal
3	Inco_Fa	รายได้บิดา	Ordinal
4	Occu_Fa	อาชีพบิดา	Nominal
5	Inco_Mo	รายได้มารดา	Ordinal
6	Occu_Mo	อาชีพมารดา	Nominal
7	Acad_Fa	การศึกษาของบิดา	Ordinal
8	Acad_Mo	การศึกษาของมารดา	Ordinal
9	Old_GPA	เกรดเฉลี่ย ม.6	Interval
10	GENERAL	หมวดวิชาศึกษาทั่วไป	Interval
11	251100	วิชาปรัชญาวิทยาศาสตร์	Interval
12	252111	วิชาคณิตศาสตร์เบื้องต้น	Interval
13	256103	วิชาเคมีเบื้องต้น	Interval
14	258101	วิชาชีววิทยาเบื้องต้น	Interval
15	261103	วิชาฟิสิกส์เบื้องต้น	Interval
16	252112	วิชาแคลคูลัส	Interval
17	254271	วิชาการโปรแกรมเบื้องต้น	Interval
18	252351	วิชาคณิตศาสตร์เต็มหน่วย	Interval
19	254251	วิชาโครงสร้างข้อมูล	Interval
20	254261	วิชาสถาปัตยกรรมคอมพิวเตอร์	Interval
21	254275	วิชาการเขียนโปรแกรมเชิงวัตถุ	Interval
22	254351	วิชาระบบจัดการฐานข้อมูล	Interval
23	GPA	ผลสัมฤทธิ์ทางการเรียน	Interval

ตาราง 6 แสดงตัวแปรของนิสิตสาขาวิชาเทคโนโลยีสารสนเทศ ช่วงชั้นปีที่ 1 และ 2

ลำดับท	ชื่อตัวแปร	รายละเอียด	ค่าข้อมูล
1	Entrance	วิธีรับเข้า	Nominal
2	Sex	เพศ	Nominal
3	Inco_Fa	รายได้บิดา	Ordinal
4	Occu_Fa	อาชีพบิดา	Nominal
5	Inco_Mo	รายได้มารดา	Ordinal
6	Occu_Mo	อาชีพมารดา	Nominal
7	Acad_Fa	การศึกษาของบิดา	Ordinal
8	Acad_Mo	การศึกษาของมารดา	Ordinal
9	Old_GPA	เกรดเฉลี่ย ม.6	Interval
10	GENERAL	หมวดวิชาศึกษาทั่วไป	Interval
11	251100	วิชาปรัชญาวิทยาศาสตร์	Interval
12	252111	วิชาคณิตศาสตร์เบื้องต้น	Interval
13	256103	วิชาเคมีเบื้องต้น	Interval
14	258101	วิชาชีววิทยาเบื้องต้น	Interval
15	261103	วิชาฟิสิกส์เบื้องต้น	Interval
16	213100	วิชารัฐกิจเบื้องต้น	Interval
17	254271	วิชาการโปรแกรมเบื้องต้น	Interval
18	273251	วิชาการระบบสารสนเทศเบื้องต้น	Interval
19	254351	วิชาการระบบจัดการฐานข้อมูล	Interval
20	254371	วิชาการโปรแกรมบนอินเทอร์เน็ต	Interval
21	273253	วิชาการระบบสารสนเทศ และองค์กร	Interval
22	273361	วิชาการจัดการเครือข่ายสำหรับระบบสารสนเทศ	Interval
23	GPA	ผลสัมฤทธิ์ทางการเรียน	Interval

ขั้นตอนที่ 3 การเตรียมข้อมูล

ผู้วิจัยได้ทำการรวบรวมข้อมูลภูมิหลัง และข้อมูลผลการเรียนในช่วงชั้นปีที่ 1 และ 2 ของนิสิตสาขาวิชาวิทยาการคอมพิวเตอร์ และสาขาวิชาเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์ มหาวิทยาลัยนครสวรรค์ ซึ่งสำเร็จการศึกษาประจำปีการศึกษา 2551 – 2554 (รหัส 48-51) ซึ่งผู้วิจัยได้ทำการตัดข้อมูลที่ไม่มีความเกี่ยวข้องกับการพยากรณ์ออกไป เช่น สถานภาพทางการศึกษา ที่อยู่ เบอร์โทรศัพท์ เกรดรายวิชาที่อยู่ในแผนการเรียน ในชั้นปีที่ 3 และ 4 รวมทั้งข้อมูลที่มีค่าว่าง

(Missing Value) ออกไปเพราะไม่สามารถนำมาใช้ในการพยากรณ์ได้ ผลการรวบรวมข้อมูลนิสิต สาขาวิชาวิทยาการคอมพิวเตอร์ และสาขาวิชาเทคโนโลยีสารสนเทศ มีจำนวนตัวแปร 23 ตัวแปร ซึ่งจำแนกเป็นตัวแปรต้น 22 ตัวแปร และตัวแปรตาม 1 ตัวแปร ดังตาราง 5 และ 6

เมื่อรวบรวมข้อมูล ทำความสะอาดข้อมูลแล้ว ยังมีการจัดเตรียมข้อมูลเพิ่มเติมโดยการแปลงข้อมูล (Data Transformation) และการคัดเลือกตัวแปรที่สำคัญ (Attribute Selection) ดังต่อไปนี้

1. การแปลงข้อมูล

ผู้วิจัยได้ทำการแปลงข้อมูลให้มีจำนวน 9 ตัวแปร ดังนี้

วิธีรับเข้า	โครงการรับตรง = 1 ส่วนกลาง – Admission = 2
เพศ	ชาย = 1 หญิง = 2
รายได้บิดา	<150,000 บาทต่อปี = 1 150,000 – 300,000 บาทต่อปี = 2 >300,000 บาทต่อปี = 3
อาชีพบิดา	เกษตร, ประมง = 1 รับราชการ, พนักงานราชการ, รัฐวิสาหกิจ = 2 ค้าขาย, ธุรกิจส่วนตัว และอาชีพอิสระ, พนักงานหน่วยงานเอกชน = 3 อื่นๆ, ไม่ระบุ = 4
รายได้มารดา	<150,000 บาทต่อปี = 1 150,000 – 300,000 บาทต่อปี = 2 >300,000 บาทต่อปี = 3
อาชีพมารดา	เกษตร, ประมง = 1 รับราชการ, พนักงานราชการ, รัฐวิสาหกิจ = 2 ค้าขาย, ธุรกิจส่วนตัว และอาชีพอิสระ, พนักงานหน่วยงานเอกชน = 3 อื่นๆ, ไม่ระบุ = 4
การศึกษาของบิดา	ต่ำกว่าปริญญาตรี = 1 ตั้งแต่ปริญญาตรีขึ้นไป = 2
การศึกษาของมารดา	ต่ำกว่าปริญญาตรี = 1 ตั้งแต่ปริญญาตรีขึ้นไป = 2
เกรดเฉลี่ย ม.6	2.00 ถึง 2.5 = 2.5 2.51 ถึง 3.0 = 3.0 3.01 ถึง 3.5 = 3.5 3.51 ถึง 4.0 = 4.0
เพราะผู้วิจัยต้องการทำให้เป็นข้อมูลเชิงกลุ่ม	

ตัวอย่างใช้เทคนิคเคมีนในการจัดกลุ่มผลสัมฤทธิ์ทางการเรียนของนิสิต 5 คน
 $(y_1, y_2, y_3, y_4, y_5)$ ออกเป็น 3 กลุ่ม ($k = 3$)

$$y_1 = [3.33], y_2 = [2.75], y_3 = [2.10], y_4 = [2.45], y_5 = [2.99]$$

1. กำหนดค่าเริ่มต้น $k = 3$

กำหนดจุดศูนย์กลางของแต่ละกลุ่ม $M_1(1) = [3.33], M_2(1) = [2.75], M_3(1) = [2.10]$

คำนวณหาความคล้ายคลึงระหว่างข้อมูลทุกตัวกับจุดศูนย์กลางวัดระยะทางแบบ
 ยูคลิเดียน หากข้อมูลนั้นใกล้จุดศูนย์กลางกลุ่มใดมากที่สุดก็จะกำหนดให้เป็นสมาชิกของกลุ่มนั้น

รอบที่ 1

$$d(y_1, M_1) = \sqrt{(3.33 - 3.33)^2} = 0$$

$$d(y_2, M_1) = \sqrt{(2.75 - 3.33)^2} = 0.58$$

$$d(y_3, M_1) = \sqrt{(2.10 - 3.33)^2} = 1.23$$

$$d(y_4, M_1) = \sqrt{(2.45 - 3.33)^2} = 0.88$$

$$d(y_5, M_1) = \sqrt{(2.99 - 3.33)^2} = 0.34$$

$$d(y_1, M_2) = \sqrt{(3.33 - 2.75)^2} = 0.58$$

$$d(y_2, M_2) = \sqrt{(2.75 - 2.75)^2} = 0$$

$$d(y_3, M_2) = \sqrt{(2.10 - 2.75)^2} = 0.65$$

$$d(y_4, M_2) = \sqrt{(2.45 - 2.75)^2} = 0.3$$

$$d(y_5, M_2) = \sqrt{(2.99 - 2.75)^2} = 0.24$$

$$d(y_1, M_3) = \sqrt{(3.33 - 2.10)^2} = 1.23$$

$$d(y_2, M_3) = \sqrt{(2.75 - 2.10)^2} = 0.65$$

$$d(y_3, M_3) = \sqrt{(2.10 - 2.10)^2} = 0$$

$$d(y_4, M_3) = \sqrt{(2.45 - 2.10)^2} = 0.35$$

$$d(y_5, M_3) = \sqrt{(2.99 - 2.10)^2} = 0.89$$

นำค่าที่คำนวณได้ทั้งหมดมาใส่ตาราง 7 เพื่อต่อการพิจารณา

ตาราง 7 ผลการคำนวณหาความคล้ายคลึงระหว่างข้อมูลทุกตัวกับจุดศูนย์กลางวัด
ระยะทางแบบยูคลิดีเนียน

	y_1	y_2	y_3	y_4	y_5
$d(y_1, M_1(1))$	0	0.58	1.23	0.88	0.34
$d(y_1, M_2(1))$	0.58	0	0.65	0.3	0.24
$d(y_1, M_3(1))$	1.23	0.65	0	0.35	0.89
	Cl 1	Cl 2	Cl 3	Cl 2	Cl 2

ซึ่งจะได้สมาชิกของกลุ่มที่ 1 (Cl1) = $\{y_1\}$ สมาชิกของกลุ่มที่ 2 (Cl2) = $\{y_2, y_4, y_5\}$ และสมาชิกของกลุ่มที่ 3 (Cl3) คือ $\{y_3\}$

2. เมื่อกำหนดกลุ่มให้กับข้อมูลทั้งหมดแล้ว จะคำนวณค่าเฉลี่ยของสมาชิกแต่ละกลุ่ม เพื่อพิจารณาจุดศูนย์กลางชุดใหม่

$$M_1(2) = \frac{1}{1}(y_1) = \frac{1}{1}(3.33) = 3.33$$

$$M_2(2) = \frac{1}{3}(y_2, y_4, y_5) = \frac{1}{3}(2.75 + 2.45 + 2.99) = 2.73$$

$$M_3(2) = \frac{1}{1}(y_3) = \frac{1}{1}(2.10) = 2.10$$

3. ตรวจสอบระยะทางระหว่างข้อมูลกับจุดศูนย์กลางข้อมูล โดยข้อมูลที่ห่างจากจุดศูนย์กลางข้อมูลไหนสั้นที่สุดก็นำข้อมูลไปสังกัดกับกลุ่มข้อมูลดังกล่าว จากจุดนี้จะทำให้เกิดการย้ายกลุ่มข้อมูล

4. ทำการคำนวณด้วยวิธีการเดียวกันนี้ไปเรื่อย ๆ จนกระทั่งจุดศูนย์กลางในแต่ละกลุ่มไม่มีการเปลี่ยนแปลงจึงหยุดการคำนวณ และจะได้กลุ่มของข้อมูลออกเป็น 3 กลุ่มตามที่ต้องการ

เมื่อผู้วิจัยได้ศึกษาทำความเข้าใจข้อมูลผลสัมฤทธิ์ทางการเรียนของกลุ่มข้อมูลชุดฝึกสอน พบว่า มีค่าต่ำสุด และสูงสุด อยู่ระหว่าง 2.00-3.31 ผู้วิจัยได้วางแผนการศึกษาแบ่งระดับผลสัมฤทธิ์ทางการเรียนเป็น 3 ระดับ หากแบ่งเป็น 3 กลุ่ม แบบกำหนดช่วงขั้นเท่ากัน หรือแบบ

กำหนดความถี่จำนวนเท่ากัน ก็จะไม่เหมาะสม จึงได้ทำการแบ่งกลุ่มโดยใช้เทคนิคเคมีน ($k = 3$) ได้ผลการแบ่งกลุ่มผลสัมฤทธิ์ทางการเรียน คือ

กลุ่มเด็กเรียนอ่อน (L) ผลสัมฤทธิ์ตั้งแต่ 2.00-2.34

กลุ่มเด็กเรียนปานกลาง (M) ผลสัมฤทธิ์ตั้งแต่ 2.35-2.70

กลุ่มเด็กเรียนเก่ง (H) ผลสัมฤทธิ์ตั้งแต่ 2.71 ขึ้นไป

2. การคัดเลือกตัวแปรที่สำคัญ

2.1 การคัดเลือกตัวแปรด้วยเทคนิคการคัดเลือกชุดตัวแปรแบบอัตราส่วนเกินร่วมกับวิธีการค้นหาแบบจัดลำดับ

จากการศึกษางานวิจัยที่เกี่ยวข้องมีประเด็นสนใจเกี่ยวกับการสำเร็จการศึกษาของนิสิต มักจะใช้ข้อมูลหลากหลาย และมีการสรุปผลกลุ่มตัวแปรที่ไม่เหมือนกัน ผู้วิจัยได้พยายามนำข้อมูลที่เกี่ยวข้องมารวบรวมไว้สำหรับการพยากรณ์ แล้วจึงใช้เทคนิคการคัดเลือกชุดตัวแปรแบบอัตราส่วนเกินร่วมกับวิธีการค้นหาแบบจัดลำดับ ซึ่งเป็นเทคนิคที่ช่วยลดจำนวนตัวแปรที่จะใช้ในการพยากรณ์ โดยวิธีการนี้เราจะได้กลุ่มตัวแปรที่มีอำนาจการพยากรณ์ผลสัมฤทธิ์ทางการเรียนเรียงตามลำดับ ผลของการคัดเลือกตัวแปรของนิสิตสาขาวิชาวิทยาการคอมพิวเตอร์ จากจำนวนตัวแปรต้นทั้งหมด 22 ตัวแปรสามารถลดจำนวนตัวแปรที่มีอิทธิพลเหลือเพียง 11 ตัวแปร ดังภาพ 24 และในการคัดเลือกตัวแปรของนิสิตสาขาวิชาเทคโนโลยีสารสนเทศจากจำนวนตัวแปรต้นทั้งหมด 22 ตัวแปร สามารถลดจำนวนตัวแปรที่มีอิทธิพลเหลือเพียง 8 ตัวแปร ดังภาพ 25 ปรากฏว่าจำนวนตัวแปรของทั้ง 2 สาขาวิชา มีจำนวนตัวแปรที่สำคัญไม่เท่ากัน ซึ่งผู้วิจัยต้องการสร้าง 1 ตัวแบบแต่สามารถนำมาทำนายผลสัมฤทธิ์ทางการเรียนได้ทั้ง 2 สาขาวิชา ดังนั้นผู้วิจัยจึงได้ทำการรวมกลุ่มวิชา และคำนวณเกรดออกมาเป็นค่าเฉลี่ยของแต่ละกลุ่มวิชา ดังภาพ 26

ตัวอย่างการคำนวณการคัดเลือกตัวแปรด้วยเทคนิคการคัดเลือกชุดตัวแปรแบบอัตราส่วนเกณฑ์ร่วมกับวิธีการค้นหาแบบจัดลำดับ โดยใช้ข้อมูลจากตาราง 8

ตาราง 8 แสดงข้อมูลนิสิตที่เรียนในสาขาวิชาวิทยาการคอมพิวเตอร์ ช่วงชั้นปีที่ 1 และ 2

GENERAL	SCI_MATH	SPECIFIC	FORCE	Class (GPA)
4	2	2	2	Medium
4	3	4	3	High
3	4	2	3	High
4	3	3	4	High
1	1	2	2	Low
1	2	2	2	Low
2	1	2	1	Low
2	2	1	1	Low
3	2	2	3	Medium

คลาส High จำนวนสมาชิกเท่ากับ 3, คลาส Medium จำนวนสมาชิกเท่ากับ 2, คลาส Low จำนวนสมาชิกเท่ากับ 4, คลาส N จำนวนสมาชิกเท่ากับ 9

วิธีการคำนวณการคัดเลือกตัวแปรด้วยเทคนิคการคัดเลือกชุดตัวแปรแบบอัตราส่วนเกณฑ์ร่วมกับวิธีการค้นหาแบบจัดลำดับ สามารถคำนวณหาได้ดังสมการ (4) โดยเริ่มทำการคำนวณดังนี้

$$\text{คำนวณหาค่าเอนโทรปีของทั้งหมัด} = -\frac{3}{9}\log_2 \frac{3}{9} - \frac{2}{9}\log_2 \frac{2}{9} - \frac{4}{9}\log_2 \frac{4}{9} = 1.5$$

คำนวณหาค่าเอนโทรปีของแต่ละเกรดของ GENERAL

$$\text{Entropy}(\text{GENERAL}, 4) = -\frac{2}{3}\log_2 \frac{2}{3} - \frac{1}{3}\log_2 \frac{1}{3} - \frac{0}{3}\log_2 \frac{0}{3} = 0.7$$

$$\text{Entropy}(\text{GENERAL}, 3) = -\frac{1}{2}\log_2 \frac{1}{2} - \frac{1}{2}\log_2 \frac{1}{2} - \frac{0}{2}\log_2 \frac{0}{2} = 1$$

$$Entropy (GENERAL, 2) = -\frac{0}{2} \log_2 \frac{0}{2} - \frac{0}{2} \log_2 \frac{0}{2} - \frac{2}{2} \log_2 \frac{2}{2} = 0$$

$$Entropy (GENERAL, 1) = -\frac{0}{2} \log_2 \frac{0}{2} - \frac{0}{2} \log_2 \frac{0}{2} - \frac{2}{2} \log_2 \frac{2}{2} = 0$$

นำ(จำนวนสมาชิกของแต่ละคลาส)×(ค่าเกนสารสนเทศของแต่ละเกรดของ GENERAL) = $\frac{3}{9} \times 0.7 + \frac{2}{9} \times 1 + \frac{2}{9} \times 0 + \frac{2}{9} \times 0 = 0.4$

นำ(ค่าเกนสารสนเทศของทั้งหมด)-(ผลคูณของจำนวนสมาชิกของแต่ละคลาส มาคูณกับค่าเกนสารสนเทศของแต่ละเกรดของ GENERAL) = $1.5 - 0.4 = 1.1$

จากนั้นคำนวณค่าสารสนเทศของการแบ่งแยก เกรด 4 ของวิชา GENERAL มีจำนวนสมาชิกเท่ากับ 3, เกรด 3 ของวิชา GENERAL มีจำนวนสมาชิกเท่ากับ 2, เกรด 2 ของวิชา GENERAL มีจำนวนสมาชิกเท่ากับ 2, เกรด 1 ของวิชา GENERAL มีจำนวนสมาชิกเท่ากับ 2

สัดส่วนของข้อมูลของวิชา GENERAL ที่มีค่าเกรดต่อจำนวนข้อมูลทั้งหมดของ วิชา GENERAL = $-\frac{3}{9} \log_2 \frac{3}{9} - \frac{2}{9} \log_2 \frac{2}{9} - \frac{2}{9} \log_2 \frac{2}{9} - \frac{2}{9} \log_2 \frac{2}{9} = 2.0$

จากนั้นคำนวณค่ามาตรฐานอัตราส่วนเกน = $\frac{1.1}{2.0} = 0.6$

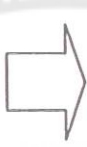
เราจะได้ค่ามาตรฐานอัตราส่วนเกน และทำการคำนวณในรายการอื่น ๆ จนครบ และนำค่ามาตรฐานอัตราส่วนเกนที่ได้มาเรียงลำดับความสำคัญ

ตัวแปรที่พยากรณ์ผลสัมฤทธิ์ทางการเรียนของนิสิตสาขาวิชาวิทยาการคอมพิวเตอร์ มี 11 ตัวแปร คือ รายวิชาในหมวดวิชาศึกษาทั่วไป คณิตศาสตร์เต็มหน่วย สถาปัตยกรรมคอมพิวเตอร์ โครงสร้างข้อมูล การโปรแกรมเบื้องต้น ระบบจัดการฐานข้อมูล แคลคูลัส คณิตศาสตร์เบื้องต้น การเขียนโปรแกรมเชิงวัตถุ ชีววิทยาเบื้องต้น และปรัชญา วิทยาศาสตร์

ตัวแปรที่พยากรณ์ผลสัมฤทธิ์ทางการเรียนของนิสิตสาขาวิชาเทคโนโลยีสารสนเทศ มี 8 ตัวแปร คือ รายวิชาในหมวดวิชาศึกษาทั่วไป ชีววิทยาเบื้องต้น การโปรแกรมเบื้องต้น ธุรกิจเบื้องต้น การโปรแกรมบนอินเทอร์เน็ต ระบบสารสนเทศ และองค์กร ระบบจัดการฐานข้อมูล และระบบสารสนเทศเบื้องต้น

ผู้วิจัยจึงได้ทำการรวมกลุ่มวิชา และคำนวณเกรดออกมาเป็นค่าเฉลี่ยของแต่ละกลุ่มวิชาออกเป็น 4 กลุ่มวิชา คือ หมวดวิชาศึกษาทั่วไป กลุ่มวิชาพื้นฐานวิทยาศาสตร์ และ คณิตศาสตร์ กลุ่มวิชาพื้นฐานเฉพาะด้าน และกลุ่มวิชาเอกบังคับ และตัวแปรตามอีก 1 ตัวแปร คือ ผลสัมฤทธิ์ทางการเรียน

Ranked	Attributes:	
0.38	10	GENERAL
0.246	20	252351
0.246	22	254261
0.232	21	254251
0.218	17	254271
0.212	19	254351
0.186	16	252112
0.171	12	252111
0.149	18	254275
0.146	14	258101
0.125	11	251100
0	4	Occu_Fa
0	5	Inco_Mo
0	1	Entrance
0	3	Inco_Fa
0	2	Sex
0	9	Old_GPA
0	13	256103
0	15	261103
0	6	Occu_Mo
0	8	Acad_Mo
0	7	Acad_Fa



Ranked	Attributes:	
0.38	10	GENERAL
0.246	20	252351
0.246	22	254261
0.232	21	254251
0.218	17	254271
0.212	19	254351
0.186	16	252112
0.171	12	252111
0.149	18	254275
0.146	14	258101
0.125	11	251100

ภาพ 24 แสดงการคัดเลือกตัวแปรของนิสิตสาขาวิชาวิทยาการคอมพิวเตอร์
ด้วยเทคนิคการคัดเลือกชุดตัวแปรแบบอัตราส่วนเกณฑ์ร่วมกับวิธีการ
ค้นหาแบบจัดลำดับ

Ranked	Attributes:	
0.451	10	GENERAL
0.43	14	258101
0.29	18	254271
0.226	17	213100
0.216	21	254371
0.159	19	273253
0.158	20	254351
0.139	16	273251
0	5	Inco_Mo
0	6	Occu_Mo
0	7	Acad_Fa
0	2	Sex
0	1	Entrance
0	4	Occu_Fa
0	3	Inco_Fa
0	13	256103
0	22	273361
0	15	261103
0	9	Old_GPA
0	8	Acad_Mo
0	12	252111
0	11	251100

➔

Ranked	Attributes:	
0.451	10	GENERAL
0.43	14	258101
0.29	18	254271
0.226	17	213100
0.216	21	254371
0.159	19	273253
0.158	20	254351
0.139	16	273251

ภาพ 25 แสดงการคัดเลือกตัวแปรของนิสิตสาขาวิชาเทคโนโลยีสารสนเทศ
ด้วยเทคนิคการคัดเลือกชุดตัวแปรแบบอัตราส่วนเกณฑ์
วิธีการค้นหาแบบจัดลำดับ

CS

No.	Attribute	No.	Attribute
1.	Entrance	13.	256103
2.	Sex	14.	258101
3.	Inco_Fa	15.	261103
4.	Occu_Fa	16.	252112
5.	Inco_Mo	17.	254271
6.	Occu_Mo	18.	252351
7.	Acad_Fa	19.	254251
8.	Acad_Mo	20.	254261
9.	Old_GPA	21.	254275
10.	GENERAL	22.	254351
11.	251100	23.	GPA
12.	252111		

การคัดเลือกตัวแปรด้วยเทคนิคการคัดเลือกชุดตัวแปรแบบอัตราส่วนเกณฑ์ร่วมกับวิธีการค้นหาแบบจัดลำดับ

No.	Attribute	No.	Attribute
1.	GENERAL	7.	252351
2.	251100	8.	254251
3.	252111	9.	254261
4.	258101	10.	254275
5.	252112	11.	254351
6.	254271	12.	GPA

No.	Attribute	No.	Attribute
1.	Entrance	13.	256103
2.	Sex	14.	258101
3.	Inco_Fa	15.	261103
4.	Occu_Fa	16.	213100
5.	Inco_Mo	17.	254271
6.	Occu_Mo	18.	273251
7.	Acad_Fa	19.	254351
8.	Acad_Mo	20.	254371
9.	Old_GPA	21.	273253
10.	GENERAL	22.	273361
11.	251100	23.	GPA
12.	252111		

IT

No.	Attribute
1.	GENERAL
2.	SCI_MATH
3.	SPECIFIC
4.	FORCE
5.	GPA

คำนวณเกรดออกมาเป็น
ค่าเฉลี่ยของแต่ละกลุ่มวิชา

No.	Attribute	No.	Attribute
1.	GENERAL	6.	254351
2.	258101	7.	254371
3.	213100	8.	273253
4.	254271	9.	GPA
5.	273251		

ภาพ 26 แสดงขั้นตอนการคัดเลือกตัวแปรด้วยเทคนิคการคัดเลือกชุดตัวแปรแบบอัตราส่วนเกณฑ์ร่วมกับวิธีการค้นหาแบบจัดลำดับ

2.2 การคัดเลือกตัวแปรด้วยสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน และสัมประสิทธิ์อันดับ

จากการศึกษางานวิจัยที่เกี่ยวข้องมีประเด็นสนใจเกี่ยวกับการสำเร็จการศึกษาของนิสิต ผู้วิจัยได้พยายามนำข้อมูลที่เกี่ยวข้องมารวบรวมไว้สำหรับการพยากรณ์ แล้วจึงใช้เทคนิคการคัดเลือกตัวแปรด้วยสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน และสัมประสิทธิ์อันดับ ซึ่งเป็นเทคนิคที่ช่วยลดจำนวนตัวแปรที่จะใช้ในการพยากรณ์ ที่มีอิทธิพลต่อการพยากรณ์ผลสัมฤทธิ์ทางการเรียน จากตาราง 12 ข้อมูลนิสิตสาขาวิชาวิทยาการคอมพิวเตอร์ ความสัมพันธ์ของตัวแปรต้นกับตัวแปรผลสัมฤทธิ์ทางการเรียน มีจำนวน 13 ตัวแปร จากตาราง 14 ข้อมูลนิสิตสาขาวิชาเทคโนโลยีสารสนเทศ ความสัมพันธ์ของตัวแปรต้นกับตัวแปรผลสัมฤทธิ์ทางการเรียน มีจำนวน 13 ตัวแปรจากตาราง 13 และ 15 พบว่า ตัวแปรทั้ง 4 ตัวแปรนั้นไม่มีความสัมพันธ์ของตัวแปรต้นกับตัวแปรผลสัมฤทธิ์ทางการเรียนเลย เนื่องจากไม่เข้าใกล้ 1 ดังนั้น สรุปได้ว่าตัวแปรที่ผ่านการคัดเลือกตัวแปรด้วยสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน และสัมประสิทธิ์อันดับของนิสิตสาขาวิชาวิทยาการคอมพิวเตอร์ และนิสิตสาขาวิชาเทคโนโลยีสารสนเทศ มีจำนวนตัวแปรทั้งหมด 13 ตัวแปรเท่ากัน และทำการคำนวณเกรดออกมาเป็นค่าเฉลี่ยของแต่ละกลุ่มวิชา ดังภาพ 27

ตัวอย่างการคำนวณหาด้วยสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน โดยใช้ข้อมูล

ตาราง 9 แสดงข้อมูลนิสิตสาขาวิชาวิทยาการคอมพิวเตอร์ ช่วงชั้นปีที่ 1 และ 2

	GENERAL	SCI_MATH	SPECIFIC	FORCE	Class (GPA)
	4	2	2	2	2.5
	4	3	4	3	3.1
	3	4	2	3	3.0
	4	3	3	4	2.8
	1	1	2	2	2.2
	1	2	2	2	2.1
	2	1	2	1	2.0
	2	2	1	1	2.2
	3	2	2	3	2.4
Σ	24	20	20	21	22.3

จากสมการ (37) สามารถคำนวณได้ดังนี้

$$\begin{aligned}\sum X_1 &= 24, \quad \sum X_2 = 20, \quad \sum X_3 = 20, \quad \sum X_4 = 21, \quad \sum Y = 22.3, \\ \sum X_1 Y &= 535.2, \quad \sum X_2 Y = 446, \quad \sum X_3 Y = 446, \\ \sum X_4 Y &= 468.3, \quad \sum X_1^2 = 286439.04, \quad \sum X_2^2 = 198916, \\ \sum X_3^2 &= 198916, \quad \sum X_4^2 = 219304.89, \quad \sum Y^2 = 247297.3441, \\ N &= 9\end{aligned}$$

$$\begin{aligned}r_{X,Y} &= \frac{9 \sum 535.2 - (24)(22.3)}{\sqrt{[9(286439.04) - (24)^2][9(247297.3441) - (22.3)^2]}} \\ &= 0.001787865\end{aligned}$$

เราจะได้อ่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน และทำการคำนวณในรายการอื่น ๆ จนครบ เพื่อนำมาทำการเปรียบเทียบความสัมพันธ์ของตัวแปรต้นกับตัวแปรผลสัมฤทธิ์ทางการเรียน

ตัวอย่างการคำนวณหาด้วยสัมประสิทธิ์อันดับ โดยใช้ข้อมูลดังตาราง 10

ตาราง 10 แสดงข้อมูลภูมิหลังของนิสิตสาขาวิชาวิทยาการคอมพิวเตอร์
ช่วงชั้นปีที่ 1 และ 2

Occu_Mo	Class (GPA)	การคำนวณเพื่อหาสัมประสิทธิ์อันดับ				
		$\bar{Y}_k$	$(Y - \bar{Y}_k)$	$(Y - \bar{Y}_k)^2$	$(Y - \bar{Y}_T)$	$(Y - \bar{Y}_T)^2$
1	2.5	3	-0.5	0.25	0.02	0.0004
1	3.1	3	0.1	0.01	0.62	0.3844
1	3.0	3	0	0	0.52	0.2704
2	2.8	2	0.8	0.64	0.32	0.1024
3	2.2	1	0.2	0.04	-0.28	0.0784
2	2.1	2	0.1	0.01	-0.38	0.1444
3	2.0	1	0	0	-0.48	0.2304
2	2.2	2	0.2	0.04	-0.28	0.0784
2	2.4	2	0.4	0.16	-0.08	0.0064
Σ	19	22.3		1.15		1.2956

$\bar{Y}_T$ (ค่าเฉลี่ยของ GPA) คือ 2.48

จากสมการ (38) คำนวณได้ดังนี้

$$\eta^2 = 1 - \frac{1.15}{1.2956}$$

$$= 0.887619636$$

$$\eta = 0.942135678$$

ตาราง 11 แสดงการคัดเลือกตัวแปรของนิสิตสาขาวิชาวิทยาการคอมพิวเตอร์
ด้วยสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน

ชื่อตัวแปร	ค่าสัมประสิทธิ์ สหสัมพันธ์แบบเพียร์สัน	ชื่อตัวแปร	ค่าสัมประสิทธิ์ สหสัมพันธ์แบบเพียร์สัน
Inco_Fa	0.126	258101	0.466**
Inco_Mo	0.019	261103	0.363**
Acad_Fa	0.173	252112	0.479**
Acad_Mo	0.061	254271	0.676**
Old_GPA	0.067	254275	0.451**
GENERAL	0.701**	254351	0.516**
251100	0.400**	252351	0.485**
252111	0.563**	254251	0.521**
256103	0.202*	254261	0.759**

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

ตาราง 12 แสดงการคัดเลือกตัวแปรของนิสิตสาขาวิชาวิทยาการคอมพิวเตอร์ด้วย
สัมประสิทธิ์อัตรา

ตัวแปร	สัมประสิทธิ์อัตรา
Entrance	.066
Sex	.156
Occu_Fa	.145
Occu_Mo	.186

ตาราง 13 แสดงการคัดเลือกตัวแปรของนิสิตสาขาวิชาเทคโนโลยีสารสนเทศด้วย
สัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน

ชื่อตัวแปร	ค่าสัมประสิทธิ์สหสัมพันธ์ แบบเพียร์สัน	ชื่อตัวแปร	ค่าสัมประสิทธิ์ สหสัมพันธ์แบบเพียร์สัน
Inco_Fa	0.124	258101	0.381**
Inco_Mo	0.051	261103	0.258**
Aca_Fa	0.074	273251	0.515**
Acad_Mo	0.075	213100	0.559**
Old_GPA	0.016	254271	0.651**
GENERAL	0.773**	273253	0.539**
251100	0.348**	254351	0.498**
252111	0.331**	254371	0.620**
256103	0.255**	273361	0.354**

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

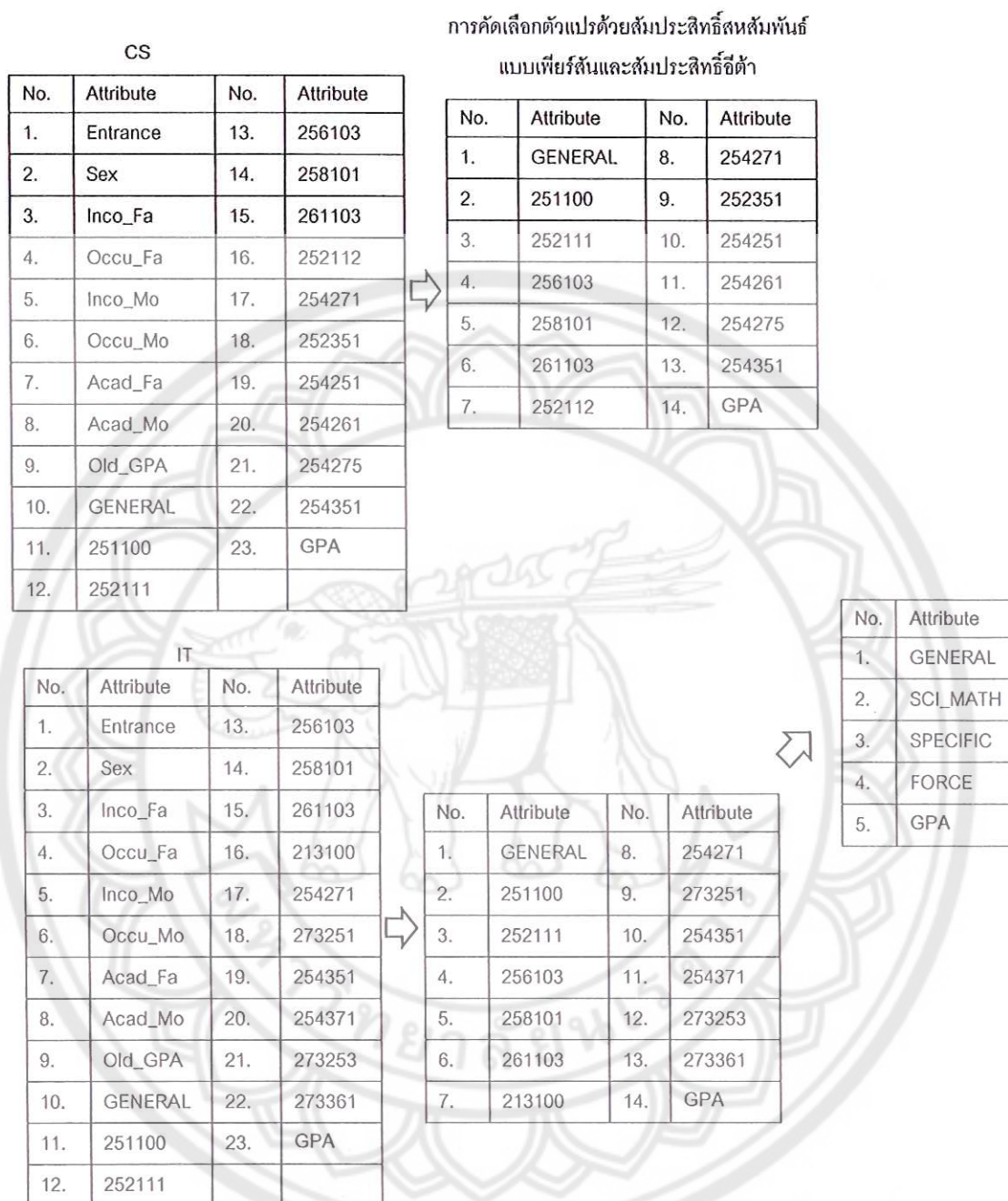
ตาราง 14 แสดงการคัดเลือกตัวแปรของนิสิตสาขาวิชาเทคโนโลยีสารสนเทศ
ด้วยสัมประสิทธิ์อัตรา

ตัวแปร	สัมประสิทธิ์อัตรา
Entrance	.045
Sex	.144
Occu_Fa	.148
Occu_Mo	.225

ตัวแปรที่พยากรณ์ผลสัมฤทธิ์ทางการเรียนของนิสิตสาขาวิชาวิทยาการคอมพิวเตอร์ มี 13 ตัวแปร คือ รายวิชาในหมวดวิชาศึกษาทั่วไป ปรัญญาวิทยาศาสตร์ คณิตศาสตร์ เบื้องต้น เคมีเบื้องต้น ชีววิทยาเบื้องต้น ฟิสิกส์เบื้องต้น แคลคูลัส การโปรแกรมเบื้องต้น การเขียนโปรแกรมเชิงวัตถุ ระบบจัดการฐานข้อมูล คณิตศาสตร์เต็มหน่วย โครงสร้างข้อมูล และสถาปัตยกรรมคอมพิวเตอร์

ตัวแปรที่พยากรณ์ผลสัมฤทธิ์ทางการเรียนของนิสิตสาขาวิชาเทคโนโลยีสารสนเทศ มี 13 ตัวแปร คือ รายวิชาในหมวดวิชาศึกษาทั่วไป ปรัญญาวิทยาศาสตร์ คณิตศาสตร์ เบื้องต้น เคมีเบื้องต้น ชีววิทยาเบื้องต้น ฟิสิกส์เบื้องต้น ระบบสารสนเทศเบื้องต้น ธุรกิจเบื้องต้น การโปรแกรมเบื้องต้น ระบบสารสนเทศ และองค์กร ระบบจัดการฐานข้อมูลการโปรแกรมบนอินเทอร์เน็ต และการจัดการเครือข่ายสำหรับระบบสารสนเทศ

ผู้วิจัยจึงได้ทำการรวมกลุ่มวิชา และคำนวณเกรดออกมาเป็นค่าเฉลี่ยของแต่ละกลุ่มวิชาออกเป็น 4 กลุ่มวิชา คือ หมวดวิชาศึกษาทั่วไป กลุ่มวิชาพื้นฐานวิทยาศาสตร์ และคณิตศาสตร์ กลุ่มวิชาพื้นฐานเฉพาะด้าน และกลุ่มวิชาเอกบังคับ และตัวแปรตามอีก 1 ตัวแปร คือ ผลสัมฤทธิ์ทางการเรียน



ภาพ 27 แสดงขั้นตอนการคัดเลือกตัวแปรด้วยสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน
และสัมประสิทธิ์อันดับ

ในการทำงานวิจัยในครั้งนี้ ผู้วิจัยได้เลือกใช้ข้อมูลนิสิตสาขาวิชาวิทยาการคอมพิวเตอร์ และสาขาวิชาเทคโนโลยีสารสนเทศ จำนวนทั้งสิ้น 489 คน ผู้วิจัยได้แบ่งข้อมูลออกเป็น 3 ชุด คือ ข้อมูลชุดฝึกสอน (Training Data) ซึ่งเป็นผู้สำเร็จการศึกษาประจำปีการศึกษา 2551 – 2552 จำนวน 236 คน ข้อมูลชุดทดสอบ (Testing Data) จำนวน 126 คน ซึ่งเป็นผู้สำเร็จการศึกษาประจำปีการศึกษา 2553 และข้อมูลชุดตรวจสอบ (Validation Data) จำนวน 127 คน ซึ่งเป็นผู้สำเร็จการศึกษาประจำปีการศึกษา 2554

ขั้นตอนที่ 4 การสร้างตัวแบบ

ข้อมูลชุดฝึกสอนที่ได้จากการคัดเลือกตัวแปร ในขั้นตอนที่ 3 มีจำนวนตัวแปรต้น 4 ตัวแปร และตัวแปรตาม 1 ตัวแปร ได้ถูกนำมาใช้วิเคราะห์ข้อมูลสำหรับการพยากรณ์โดยใช้เทคนิคเหมืองข้อมูล 4 เทคนิค ได้แก่ 1) เทคนิคเคเนียร์เรสเนเบอร์ 2) เทคนิคเนอรัลเน็ต 3) เทคนิคโครงข่ายประสาทเทียมแบบแพร่กลับ และ 4) เทคนิคต้นไม้ตัดสินใจ ผู้วิจัยได้เลือกใช้เทคนิคเคเนียร์เรสเนเบอร์ เทคนิคเนอรัลเน็ต เทคนิคต้นไม้ตัดสินใจ และเทคนิคโครงข่ายประสาทเทียมแบบแพร่กลับ มาใช้ในการพยากรณ์ผลสัมฤทธิ์ทางการเรียนของนิสิต

เทคนิคเคเนียร์เรสเนเบอร์

ตัวอย่างต้องการทราบว่านิสิตคนนี้จะได้ผลสัมฤทธิ์ทางการเรียนอยู่ในกลุ่มใด โดยพิจารณาเพื่อนบ้านที่ใกล้ที่สุด $= 3$ ($k = 3$) โดยใช้ข้อมูลดังตาราง 15

ตาราง 15 แสดงข้อมูลผลการเรียนนิสิตสาขาวิชาวิทยาการคอมพิวเตอร์

	GENERAL	SCI_MATH	SPECIFIC	FORCE	Class
q_1	3	2	2	2	Medium
q_2	4	3	3	3	High
q_3	4	4	3	3	High
q_4	4	3	3	4	High
q_5	2	2	2	2	Low
q_6	1	2	2	2	Low
q_7	2	2	1	1	Low
q_8	3	2	2	3	Medium
p	4	2	1	2	?

ทำการแปลงค่าข้อมูลดังสมการ (7)

ทำการแปลงข้อมูลจากตาราง 15 และนำข้อมูลที่แปลงได้มาใช้อย่างตาราง 16

ตาราง 16 แสดงการแปลงค่าข้อมูลผลการเรียนนิสิตสาขาวิชาวิทยาการคอมพิวเตอร์

	GENERAL	SCI_MATH	SPECIFIC	FORCE	Class
q_1	$\frac{3.0-1.0}{4.0-1.0} = 0.67$	$\frac{2.0-1.0}{4.0-1.0} = 0.33$	$\frac{2.0-1.0}{4.0-1.0} = 0.33$	$\frac{2.0-1.0}{4.0-1.0} = 0.33$	Medium
q_2	$\frac{4.0-1.0}{4.0-1.0} = 1.00$	$\frac{3.0-1.0}{4.0-1.0} = 0.67$	$\frac{3.0-1.0}{4.0-1.0} = 0.67$	$\frac{3.0-1.0}{4.0-1.0} = 0.67$	High
q_3	$\frac{4.0-1.0}{4.0-1.0} = 1.00$	$\frac{4.0-1.0}{4.0-1.0} = 1.00$	$\frac{3.0-1.0}{4.0-1.0} = 0.67$	$\frac{3.0-1.0}{4.0-1.0} = 0.67$	High
q_4	$\frac{4.0-1.0}{4.0-1.0} = 1.00$	$\frac{3.0-1.0}{4.0-1.0} = 0.67$	$\frac{3.0-1.0}{4.0-1.0} = 0.67$	$\frac{4.0-1.0}{4.0-1.0} = 1.00$	High
q_5	$\frac{2.0-1.0}{4.0-1.0} = 0.33$	$\frac{2.0-1.0}{4.0-1.0} = 0.33$	$\frac{2.0-1.0}{4.0-1.0} = 0.33$	$\frac{2.0-1.0}{4.0-1.0} = 0.33$	Low
q_6	$\frac{1.0-1.0}{4.0-1.0} = 0.00$	$\frac{2.0-1.0}{4.0-1.0} = 0.33$	$\frac{2.0-1.0}{4.0-1.0} = 0.33$	$\frac{2.0-1.0}{4.0-1.0} = 0.33$	Low
q_7	$\frac{2.0-1.0}{4.0-1.0} = 0.33$	$\frac{2.0-1.0}{4.0-1.0} = 0.33$	$\frac{1.0-1.0}{4.0-1.0} = 0.00$	$\frac{1.0-1.0}{4.0-1.0} = 0.00$	Low
q_8	$\frac{3.0-1.0}{4.0-1.0} = 0.67$	$\frac{2.0-1.0}{4.0-1.0} = 0.33$	$\frac{2.0-1.0}{4.0-1.0} = 0.33$	$\frac{3.0-1.0}{4.0-1.0} = 0.67$	Medium
p	$\frac{4.0-1.0}{4.0-1.0} = 1.00$	$\frac{2.0-1.0}{4.0-1.0} = 0.33$	$\frac{1.0-1.0}{4.0-1.0} = 0.00$	$\frac{2.0-1.0}{4.0-1.0} = 0.33$	?

เทคนิคเคเนียร์สเนเบอร์ สามารถคำนวณหาได้ดังสมการ (8)

$$d(p, q_1) = \sqrt{(1-0.67)^2 + (0.33-0.33)^2 + (0-0.33)^2 + (0.33-0.33)^2}$$

$$= 0.467$$

$$d(p, q_2) = \sqrt{(1-1)^2 + (0.33-0.67)^2 + (0-0.67)^2 + (0.33-0.67)^2}$$

$$= 0.825$$

$$d(p, q_3) = \sqrt{(1-1)^2 + (0.33-1)^2 + (0-0.67)^2 + (0.33-0.67)^2}$$

$$= 1.007$$

$$d(p, q_4) = \sqrt{(1-1)^2 + (0.33-0.67)^2 + (0-0.67)^2 + (0.33-1)^2}$$

$$= 1.007$$

$$d(p, q_5) = \sqrt{(1-0.33)^2 + (0.33-0.33)^2 + (0-0.33)^2 + (0.33-0.33)^2}$$

$$= 0.747$$

$$d(p, q_6) = \sqrt{(1-0)^2 + (0.33-0.33)^2 + (0-0.33)^2 + (0.33-0.33)^2}$$

$$= 1.053$$

$$d(p, q_7) = \sqrt{(1-0.33)^2 + (0.33-0.33)^2 + (0-0.0)^2 + (0.33-0.0)^2}$$

$$= 0.747$$

$$d(p, q_8) = \sqrt{(1-0.67)^2 + (0.33-0.33)^2 + (0-0.33)^2 + (0.33-0.67)^2}$$

$$= 0.557$$

จากค่าระยะทางที่ได้จากชุดข้อมูลของนิสิตทั้ง 8 คน จะพิจารณาเลือกจากระยะทางที่น้อยที่สุด 3 อันดับ คือ $d(p, q_1) = 0.467$, $d(p, q_5) = 0.747$ และ $d(p, q_8) = 0.557$ (เนื่องจาก $d(p, q_5)$ และ $d(p, q_7)$ มีค่าเท่ากัน ในที่นี้เลือกใช้ $d(p, q_5)$) ตามเสียงส่วนมากแล้ว ระยะทางที่คำนวณได้ จะอยู่ในกลุ่ม Medium คือ นิสิตคนนี้จะเรียนอยู่ในกลุ่มปานกลาง

ดังนั้น สรุปได้ว่า นิสิตคนนี้จะอยู่ในกลุ่ม Medium คือ นิสิตคนนี้จะเรียนอยู่กลุ่มปานกลาง

เทคนิคเนอ์ฟเบย์

ตัวอย่างต้องการทราบว่านิสิตคนนี้จะได้ผลสัมฤทธิ์ทางการเรียนอยู่ในกลุ่มใด โดยใช้ข้อมูลดังตาราง 17

ตาราง 17 แสดงข้อมูลนิสิตสาขาวิชาวิทยาการคอมพิวเตอร์

GENERAL	SCI_MATH	SPECIFIC	FORCE	Class
4	1	2	2	Medium
4	3	4	2	High
3	4	2	3	High
4	3	3	4	High
2	2	2	2	Low
1	2	2	2	Low
4	1	1	1	Low
2	2	1	1	Low
3	2	2	3	Medium
4	2	1	2	?

เทคนิคเนอ์ฟเบย์ สามารถคำนวณหาได้ดังสมการ (11)

$$P(\text{Class}=\text{"High"}) = \frac{3}{9} = 0.33$$

$$P(\text{Class}=\text{"Medium"}) = \frac{2}{9} = 0.22$$

$$P(\text{Class}=\text{"Low"}) = \frac{4}{9} = 0.44$$

$$P(\text{GENERAL}=\text{"4"}|\text{Class}=\text{"High"}) = \frac{2}{3} = 0.66$$

$$P(\text{GENERAL}=\text{"4"}|\text{Class}=\text{"Medium"}) = \frac{1}{2} = 0.5$$

$$P(\text{GENERAL}=\text{"4"}|\text{Class}=\text{"Low"}) = \frac{1}{4} = 0.25$$

$$P(\text{SCI_MATH}=\text{"2"}|\text{Class}=\text{"High"}) = \frac{0}{3} = 0$$

$$P(\text{SCI_MATH}=\text{"2"}|\text{Class}=\text{"Medium"}) = \frac{1}{2} = 0.5$$

$$P(\text{SCI_MATH}=\text{"2"}|\text{Class}=\text{"Low"}) = \frac{3}{4} = 0.75$$

$$P(\text{SPECIFIC}=\text{"1"}|\text{Class}=\text{"High"}) = \frac{0}{3} = 0$$

$$P(\text{SPECIFIC}=\text{"1"}|\text{Class}=\text{"Medium"}) = \frac{0}{2} = 0$$

$$P(\text{SPECIFIC}=\text{"1"}|\text{Class}=\text{"Low"}) = \frac{2}{4} = 0.5$$

$$P(\text{FORCE}=\text{"2"}|\text{Class}=\text{"High"}) = \frac{1}{3} = 0.33$$

$$P(\text{FORCE}=\text{"2"}|\text{Class}=\text{"Medium"}) = \frac{1}{2} = 0.5$$

$$P(\text{FORCE}=\text{"2"}|\text{Class}=\text{"Low"}) = \frac{2}{4} = 0.5$$

$$P(B|\text{Class}=\text{"High"}) = 0.66 \times 0 \times 0 \times 0.33 = 0$$

$$P(B|\text{Class}=\text{"Medium"}) = 0.5 \times 0.5 \times 0 \times 0.5 = 0$$

$$P(B|\text{Class}=\text{"Low"}) = 0.25 \times 0.75 \times 0.5 \times 0.5 = 0.046$$

$$P(B|\text{Class}=\text{"High"}) \times P(\text{Class}=\text{"High"}) = 0 \times 0.33 = 0$$

$$P(B|\text{Class}=\text{"Medium"}) \times P(\text{Class}=\text{"Medium"}) = 0 \times 0.22 = 0$$

$$P(B|\text{Class}=\text{"Low"}) \times P(\text{Class}=\text{"Low"}) = 0.046 \times 0.44 = 0.020$$

ดังนั้น สรุปได้ว่า นิสิตคนนี้จะอยู่ในกลุ่ม Low คือ กลุ่มเรียนอ่อน

เทคนิคโครงข่ายประสาทเทียมแบบแพร่กลับ

ตัวอย่าง โครงข่ายประสาทเทียมแบบแพร่กลับ โดยใช้ข้อมูลดังตาราง 18

ค่าความผิดพลาดที่เกิดขึ้นในชั้นนำออก และชั้นซ่อนก็สามารถเขียนได้ ดังสมการ (29)

ค่าน้ำหนักเส้นเชื่อมต่อในชั้นนำออกก็จะถูกปรับเปลี่ยนไป สามารถคำนวณดังสมการ (30) และ (31)

ในขณะเดียวกัน ค่าน้ำหนักเส้นเชื่อมต่อในชั้นซ่อน ก็สามารถปรับเปลี่ยนไป สามารถคำนวณดังสมการ (32) และ (33)

เทคนิคต้นไม้ตัดสินใจ

ตัวอย่าง ต้องการทราบว่านิสิตคนนี้จะได้ผลสัมฤทธิ์ทางการเรียนอยู่ในกลุ่มใด โดยใช้ข้อมูลดังตาราง 18

ตาราง 18 แสดงข้อมูลผลการเรียนนิสิตสาขาวิชาวิทยาการคอมพิวเตอร์
ช่วงชั้นปีที่ 1 และ 2

GENERAL	SCI_MATH	SPECIFIC	FORCE	Class
4	1	2	2	Medium
4	3	4	2	High
3	4	2	3	High
4	3	3	4	High
2	2	2	2	Low
1	2	2	2	Low
4	1	1	1	Low
2	2	1	1	Low
3	2	2	3	Medium
3	3	1	2	?

คลาส High จำนวนสมาชิกเท่ากับ 3, คลาส Medium จำนวนสมาชิกเท่ากับ 2, คลาส Low จำนวนสมาชิกเท่ากับ 4, คลาส N จำนวนสมาชิกเท่ากับ 9

คำนวณหาค่าเอนโทรปีของทั้งหมด

$$= -\frac{3}{9} \log_2 \frac{3}{9} - \frac{2}{9} \log_2 \frac{2}{9} - \frac{4}{9} \log_2 \frac{4}{9} = 1.53$$

คำนวณหาค่าเอนโทรปีของแต่ละเกรดของ GENERAL

$$Entropy(GENERAL, 4) = -\frac{2}{4} \log_2 \frac{2}{4} - \frac{1}{4} \log_2 \frac{1}{4} - \frac{1}{4} \log_2 \frac{1}{4} = 1.5$$

$$Entropy(GENERAL, 3) = -\frac{1}{2} \log_2 \frac{1}{2} - \frac{1}{2} \log_2 \frac{1}{2} - \frac{0}{2} \log_2 \frac{0}{2} = 1$$

$$Entropy(GENERAL, 2) = -\frac{0}{2} \log_2 \frac{0}{2} - \frac{0}{2} \log_2 \frac{0}{2} - \frac{2}{2} \log_2 \frac{2}{2} = 0$$

$$Entropy(GENERAL, 1) = -\frac{0}{1} \log_2 \frac{0}{1} - \frac{0}{1} \log_2 \frac{0}{1} - \frac{1}{1} \log_2 \frac{1}{1} = 0$$

นำ (จำนวนสมาชิกของแต่ละคลาส) × (ค่าเอนโทรปีของแต่ละเกรดของ GENERAL)

$$= \frac{4}{9} \times 1.5 + \frac{2}{9} \times 1 + \frac{2}{9} \times 0 + \frac{1}{9} \times 0 = 0.89$$

นำ (ค่าเอนโทรปีของทั้งหมด) - (ผลคูณของจำนวนสมาชิกของแต่ละคลาสคูณกับค่าเอนโทรปีของแต่ละเกรดของ GENERAL) = $1.53 - 0.89 = 0.64$

จากนั้นคำนวณค่าเอนโทรปีของการแบ่งแยก

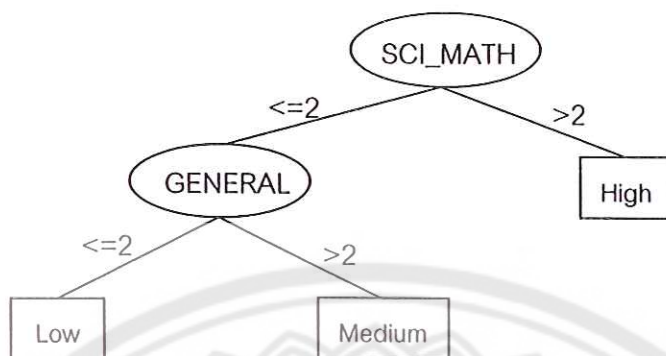
เกรด 4 ของวิชา GENERAL มีจำนวนสมาชิกเท่ากับ 4, เกรด 3 ของวิชา GENERAL มีจำนวนสมาชิกเท่ากับ 2, เกรด 2 ของวิชา GENERAL มีจำนวนสมาชิกเท่ากับ 2, เกรด 1 ของวิชา GENERAL มีจำนวนสมาชิกเท่ากับ 1

สัดส่วนของข้อมูลของวิชา GENERAL ที่มีค่าเกรดต่อจำนวนข้อมูลทั้งหมดของวิชา

$$GENERAL = -\frac{4}{9} \log_2 \frac{4}{9} - \frac{2}{9} \log_2 \frac{2}{9} - \frac{2}{9} \log_2 \frac{2}{9} - \frac{1}{9} \log_2 \frac{1}{9} = 1.84$$

$$\text{จากนั้นคำนวณค่ามาตรฐานอัตราส่วนเอน} = \frac{0.64}{1.84} = 0.35$$

เราจะได้อัตราส่วนเอน และการคำนวณในรายการอื่น ๆ จนครบ เราก็จะสามารถเลือกค่ามาตรฐานอัตราส่วนเอนสูงสุดเป็นตัวแปรเริ่มต้น และเลือกคุณสมบัติถัดไปตามค่ามาตรฐานอัตราส่วนเอนน้อยลงตามลำดับ ดังภาพ 28



ภาพ 28 แสดงใช้ค่ามาตรฐานอัตราส่วนเกินสูงสุดเป็นตัวแปรเริ่มต้น

หลังจากสร้างต้นไม้ตัดสินใจแล้ว สามารถเขียนกฎในรูปแบบของเงื่อนไข (Condition) ดังนี้

1. IF SCI_MATH = "<= 2" and GENERAL <= 2 THEN Class = "Low"
2. IF SCI_MATH = "<= 2" and GENERAL > 2 THEN Class = "Medium"
3. IF SCI_MATH = "> 2" THEN Class = "High"

ดังนั้น สรุปได้ว่า นิสิตคนนี้จะอยู่ในกลุ่ม High คือ กลุ่มเรียนเก่ง ตามกฎข้อ 3
ขั้นตอนที่ 5 การประเมินผล

ผู้วิจัยได้นำตัวแบบทุก ๆ แบบมาทดสอบด้วยค่าถูกต้อง ค่าความแม่นยำ และค่าความระลึกเพื่อให้การประเมินประสิทธิภาพของแต่ละเทคนิคมีความเชื่อมั่นสูง ในงานวิจัยนี้ได้เลือกใช้วิธี 10 - folds Cross Validation เมื่อได้ผลการประเมินของเทคนิคการพยากรณ์เป็นที่น่าพอใจแล้ว ก็จะนำข้อมูลชุดทดสอบ มาใช้กับตัวแบบการพยากรณ์ เพื่อทดสอบค่าความใช้ได้ของตัวแบบ ซึ่งจะเป็นการสร้างความมั่นใจว่าตัวแบบที่ได้สร้างขึ้นมานั้นมิได้มีความสามารถในการพยากรณ์ได้ดีเฉพาะข้อมูลชุดฝึกสอนเท่านั้น (Over-fitting)

นอกจากนั้น เนื่องจากการวัดประสิทธิภาพด้วย ค่าความถูกต้อง คำนวณจากจำนวนนับผลสัมฤทธิ์เชิงกลุ่ม ผู้วิจัยต้องการตรวจสอบความถูกต้องเชิงปริมาณ จึงได้ใช้วิธีการวัดประสิทธิภาพด้วยวิธีที่จะทำให้ทราบถึงค่าความผิดพลาดของการประมาณค่าที่ชัดเจนเนื่องจากเป็นวิธีที่ใช้ผลรวมของผลต่างระหว่างค่าของข้อมูลจริงกับค่าของข้อมูลที่คำนวณได้จากอัลกอริทึมด้วยค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง

ขั้นตอนที่ 6 นำตัวแบบมาใช้งาน

หลังจากทำการประเมินผลตัวแบบการพยากรณ์ของข้อมูลชุดฝึกสอน และข้อมูลชุดทดสอบ ได้ผลประเมินที่เหมาะสมเป็นที่น่าพอใจแล้ว ได้นำข้อมูลชุดตรวจสอบที่ได้แบ่งไว้มาตรวจสอบค่าความถูกต้องของตัวแบบอีกครั้ง ก่อนที่จะนำตัวแบบที่ได้สร้างขึ้นมาใช้ประโยชน์จริงในการพยากรณ์ผลสัมฤทธิ์ทางการเรียนของนิสิต หลักสูตรวิทยาศาสตรบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์ และสาขาวิชาเทคโนโลยีสารสนเทศ ในรุ่นต่อ ๆ ไป

