



# ตัวแบบพยากรณ์ลักษณะความเหมาะสมของนักศึกษาใหม่ สาขาวิชาเทคโนโลยีสารสนเทศ ด้วยกฎการจำแนกประเภทเชิงความสัมพันธ์

จามรกุล เหล่าเกียรติกุล\* และ จิรรัตน์ สิทธิวรชาติ\*\*

## บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์เพื่อสร้างตัวแบบในการพยากรณ์ลักษณะความเหมาะสมต่อการเข้าศึกษาของนักศึกษาใหม่ในสาขาวิชาเทคโนโลยีสารสนเทศด้วยเทคนิคกฎการจำแนกเชิงความสัมพันธ์ (CARs: Class Association Rules) โดยใช้ข้อมูลนักศึกษาสาขาเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏจันทรเกษม เป็นกรณีศึกษา เพื่อพยากรณ์คุณลักษณะความเหมาะสมของนักศึกษา โดยใช้ระดับผลการเรียนปัจจุบันกำหนดกลุ่มระดับความเหมาะสมของนักศึกษาใหม่ ออกเป็น 4 กลุ่มได้แก่ กลุ่มนักศึกษาระดับดีเยี่ยม ระดับดี ระดับปานกลาง และระดับต่ำ ผลการทดลองพบว่า กฎการจำแนกประเภทเชิงความสัมพันธ์ที่สร้างขึ้น สามารถจำแนกระดับความเหมาะสมของนักศึกษาได้ในระดับดี (ร้อยละ 83.81) ซึ่งสามารถนำไปใช้ในการพยากรณ์ลักษณะความเหมาะสมของนักศึกษาใหม่ในสาขาวิชาเทคโนโลยีสารสนเทศ ทั้งนี้เพื่อใช้ประโยชน์ในการจัดการเรียนการสอนให้เหมาะสมกับผู้เรียนต่อไป

**คำสำคัญ:** ความเหมาะสมของนักศึกษา กฎการจำแนกเชิงความสัมพันธ์ การพยากรณ์

## 1. บทนำ

การคัดเลือกนักศึกษาใหม่เพื่อรับเข้าศึกษาในสาขาวิชาต่างๆ ในระดับอุดมศึกษาด้วยวิธีการรับโดยตรง อาจใช้วิธีคัดเลือกกับผู้สมัครเพื่อเข้าศึกษาด้วยการทดสอบความสามารถทางวิชาการ หรือการสอบสัมภาษณ์ ซึ่งเมื่อมีผู้ผ่านการคัดเลือกเข้ามาเป็นนักศึกษาแล้ว จะพบว่ามีส่วนหนึ่งพบสภาพการเป็นนักศึกษาในเวลาต่อมา ดังนั้นหากมีการนำเอา

ข้อมูลส่วนบุคคลที่นักศึกษาให้ไว้ก่อนเข้าเป็นนักศึกษาของนักศึกษาปัจจุบันและอดีต นำมาวิเคราะห์เพื่อค้นหาปัจจัยที่ส่งผลต่อคุณลักษณะที่เหมาะสมต่อสาขาวิชา จะทำให้สามารถใช้ในคาดการณ์เพื่อจัดจำแนกกลุ่มนักศึกษาใหม่เพื่อปรับ หรือจัดการเรียนการสอนให้เหมาะสมกับนักศึกษาต่อไป งานวิจัยนี้มีวัตถุประสงค์เพื่อ สร้างตัวแบบในการวิเคราะห์และจำแนกคุณลักษณะความเหมาะสมต่อการเข้าศึกษาในสาขาวิชาเทคโนโลยีสารสนเทศ และตัวแบบต้องสามารถจำแนกกลุ่มนักศึกษาได้ตามคุณลักษณะที่เหมาะสม โดยเลือกศึกษาข้อมูลคุณลักษณะส่วนบุคคลของนักศึกษาในสาขาวิชาเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏจันทรเกษม เป็นกรณีศึกษา โดยใช้เทคนิคกฎการจำแนกเชิงความสัมพันธ์ (CARs) ในการค้นหาความสัมพันธ์ของข้อมูลนักศึกษาในสาขาวิชาเทคโนโลยีสารสนเทศ ที่ส่งผลต่อระดับผลการเรียน สร้างเป็นกฎเพื่อเชื่อมโยงความสัมพันธ์และใช้กฎดังกล่าว เพื่อการคาดการณ์ลักษณะความเหมาะสม โดยกำหนดเป็น 4 กลุ่มได้แก่ กลุ่มนักศึกษาระดับดีเยี่ยม กลุ่มนักศึกษาระดับดี กลุ่มนักศึกษาระดับปานกลาง และกลุ่มนักศึกษาระดับต่ำ ซึ่งสามารถนำไปใช้ประโยชน์ในการปรับแผนการเรียน หรือจัดกลุ่มเรียน รวมถึงการเรียนการสอนให้เหมาะสมกับนักศึกษาในลักษณะต่างๆ ต่อไป ถือเป็นการสร้างโอกาสให้แก่ผู้สนใจเข้าศึกษาในสาขา วิชาเทคโนโลยีสารสนเทศ ได้มีโอกาสในการศึกษาได้อย่างเหมาะสม และเพิ่มโอกาสสำเร็จการศึกษายิ่งขึ้น

## 2. งานวิจัยที่เกี่ยวข้อง

จากการค้นคว้างานวิจัย ในลักษณะของการค้นหาปัจจัยจากข้อมูลคุณลักษณะของบุคคล (Demographic Data)

\* ภาควิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

\*\* ภาควิชาการจัดการเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ



ที่ส่งผลต่อผลสัมฤทธิ์ทางการศึกษา พบงานวิจัยที่ศึกษาในระดับการศึกษาหลากหลายระดับด้วยกัน เช่น งานวิจัยในการวิเคราะห์ปัจจัยที่ส่งผลต่อการเลือกเข้าศึกษาในระดับมหาวิทยาลัยที่ค้นพบว่า ระดับผลการเรียนเดิม อาชีพของผู้ปกครอง ระดับการศึกษาของผู้ปกครอง และรายได้ของครอบครัว ส่งผลต่อผลสัมฤทธิ์ทางการศึกษา [1-4] งานวิจัยเกี่ยวกับปัจจัย ที่ส่งผลต่อผลสัมฤทธิ์ทางการเรียนในรายวิชาหรือสาขาวิชาต่างๆ ที่ค้นพบปัจจัยเกี่ยวกับความสนใจ การเข้าห้องเรียน และคะแนนเก็บ [5-7] งานวิจัยที่ศึกษาปัจจัยที่ส่งผลกระทบต่อการเรียน [5, 6] ซึ่งใช้การวิเคราะห์ข้อมูลทางสถิติมาใช้ในการทำวิจัย และงานวิจัยที่ใช้เทคนิคต่างๆ ของกรรมวิธีเหมืองข้อมูลในการวิเคราะห์ข้อมูลนักศึกษา [1, 8, 9] รวมไปถึงการศึกษาปัจจัยที่ส่งผลต่อการคัดเลือกบุคคลเข้ารับตำแหน่งในสาขาวิชาชีพ [10]

ผลจากงานวิจัยเกี่ยวกับคุณลักษณะของบุคคลข้างต้น แสดงให้เห็นว่าคุณลักษณะส่วนบุคคลส่งผลต่อความสามารถในการศึกษาในระดับต่างๆ แตกต่างกันไป ในขณะที่มีงานวิจัยที่ประยุกต์ใช้ เทคนิคการสร้างกฎการเชื่อมโยงความสัมพันธ์ รวมถึงเทคนิคกฎการจำแนกประเภทเชิงความสัมพันธ์มากมายหลายด้านด้วยกัน เช่น งานวิจัยที่ประยุกต์ใช้กฎการเชื่อมโยงความสัมพันธ์กับการวิเคราะห์ข้อมูลการหัวใจล้มเหลว [11] ประยุกต์ใช้กับการจัดการข้อมูลภายในห้องสมุด [12] เป็นต้น

จะเห็นได้ว่า มีงานวิจัยที่นำเอาเทคนิควิธีต่างๆ ของกรรมวิธีเหมืองข้อมูลมาวิเคราะห์ข้อมูลเกี่ยวกับนักศึกษาแตกต่างกันไป ในขณะที่มีการประยุกต์ใช้เทคนิคกฎการจำแนกประเภทเชิงความสัมพันธ์ในด้านต่างๆ มากมาย เนื่องจากใช้หน่วยความจำ รวมถึงประสิทธิภาพของหน่วยประมวลผลไม่สูง อีกทั้งยังง่ายต่อการนำมาประยุกต์ใช้งาน ผู้วิจัยจึงมีแนวคิดในการนำเทคนิคดังกล่าว มาใช้วิเคราะห์ข้อมูลของนักศึกษา เพื่อค้นหาความสัมพันธ์ของข้อมูลคุณลักษณะส่วนบุคคล และนำเอาปัจจัยดังกล่าวมาประยุกต์ใช้ในการจำแนกระดับความเหมาะสมของผู้เรียน โดยเชื่อมั่นว่าเทคนิคดังกล่าวจะสามารถวิเคราะห์ปัจจัยที่ส่งผลต่อระดับความเหมาะสมของนักศึกษา และสามารถนำมาใช้คาดการณ์อย่างมีประสิทธิภาพ

### 3. กฎการจำแนกประเภทเชิงความสัมพันธ์ (CARs)

กฎการจำแนกประเภทเชิงความสัมพันธ์ (CARs) เป็นเทคนิคของกรรมวิธีเหมืองข้อมูล ที่ผสมผสานระหว่าง

เทคนิคการวิเคราะห์เพื่อสร้างกฎเชื่อมโยงความสัมพันธ์ (Association -Rules Discovery) เข้ากับวิธีการจำแนกกลุ่มข้อมูล (Data -Classification) โดยมีวัตถุประสงค์เพื่อใช้ประโยชน์กฎการเชื่อมโยงความสัมพันธ์ที่ถูกสร้างขึ้นในการจำแนกข้อมูล โดยกฎที่ถูกคัดเลือกเพื่อนำมาใช้ในจำแนกข้อมูลนี้จะถูกเรียกว่า Class Association Rules (CARs) [13]

#### 3.1 เทคนิควิเคราะห์เพื่อสร้างกฎเชื่อมโยงความสัมพันธ์ (Association Rules Discovery)

เทคนิคการวิเคราะห์เพื่อสร้างกฎเชื่อมโยงความสัมพันธ์นี้เป็นเทคนิคหนึ่งในกรรมวิธีเหมืองข้อมูล ที่ได้รับความนิยมอย่างกว้างขวางในการวิจัยเชิงประยุกต์ใช้เหมืองข้อมูลกับข้อมูลด้านต่างๆ กฎของความสัมพันธ์ระหว่างข้อมูลนั้นจะถูกสร้างขึ้นโดยการค้นหาความถี่ที่เกิดขึ้น (Frequency items set) ที่อยู่ภายใต้ค่าสนับสนุนหรือค่าที่ยอมรับได้ (Support) และค่าความเป็นไปได้ของแต่ละปัจจัยที่นำมาพิจารณา (Confidence) [14]

การประยุกต์ กฎเชื่อมโยงความสัมพันธ์ เพื่อใช้เป็นกฎการจำแนกประเภทเชิงความสัมพันธ์ จะเลือกใช้เฉพาะกฎที่ส่งผลต่อประเภทของข้อมูลที่ต้องการ

#### 3.2 การหาและสร้างกฎจำแนกประเภทเชิงความสัมพันธ์

กฎจำแนกประเภทเชิงความสัมพันธ์ เป็นกฎที่แตกต่างจากการสร้างกฎความสัมพันธ์ของข้อมูล (Association rules) ที่บอกความสัมพันธ์ของข้อมูลทั้งหมด โดยไม่เฉพาะเจาะจงกับข้อมูลที่กำหนด ในขณะที่กฎจำแนกประเภทเชิงความสัมพันธ์นั้นจะบอกความสัมพันธ์ของข้อมูลต่อประเภทของข้อมูลนั้นๆ (Class) ซึ่งเป็นการค้นหาเฉพาะความสัมพันธ์ที่เฉพาะเจาะจงกับประเภทของข้อมูลโดยไม่คำนึงถึงความสัมพันธ์กับข้อมูลอื่นๆ [15] โดยมีลักษณะคือ

$$\{item\ i, \dots item\ j\} \Rightarrow Class : Supt.\ Conf.$$

เมื่อ item i-j คือ ข้อมูลที่สัมพันธ์กัน และ Class คือประเภทของข้อมูลนั้นๆ โดยที่ต้องคำนวณหาค่าสนับสนุน (Supt.) จากจำนวนสมาชิกที่สอดคล้องตามกฎต่อจำนวนข้อมูลทั้งหมด และค่าความเชื่อมั่น (Conf.) จากจำนวนข้อมูลที่สอดคล้องตามกฎต่อจำนวนสมาชิกในกฎนั้น โดยที่ต้องนำ



มาผ่านกระบวนการในการจัดลำดับค่าความสำคัญ (Precedence) ของกฎเพื่อให้ง่ายต่อการนำไปใช้จำแนกข้อมูล โดยพิจารณาให้ค่าความสำคัญจากกฎที่มีค่าความเชื่อมั่นที่มากกว่า และ/หรือกฎที่มีค่าสนับสนุนที่มากกว่า ให้มีค่าความสำคัญมากกว่า

#### 4. ตัวแบบพยากรณ์ลักษณะความเหมาะสมของนักศึกษาสาขาวิชาเทคโนโลยีสารสนเทศ

##### 4.1 การจัดเตรียมข้อมูล

งานวิจัยนี้ จัดเตรียมข้อมูลโดยใช้ข้อมูลนักศึกษาสาขาวิชาเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏจันทรเกษม จากฐานข้อมูล ตามมาตรฐานของสำนักงานคณะกรรมการการอุดมศึกษา (สกอ.) เท่าที่มีการบันทึกข้อมูลส่วนบุคคลไว้ในฐานข้อมูลดังกล่าว จำนวน 2,620 คน โดยกำหนดตัวแปรที่จะใช้ในการวิจัย ซึ่งตัวแปรในงานวิจัยนี้มีจำนวนทั้งสิ้น 18 ตัวแปรดังแสดงในตารางที่ 1 ซึ่งหลังจากรวบรวมข้อมูลแล้วพบ 1,250 ระเบียบที่มีข้อมูลสมบูรณ์เพียงพอที่สามารถนำมาใช้ในการสร้างตัวแบบฯ โดยนำข้อมูลดังกล่าวมาจัดเตรียมข้อมูลสำหรับการเรียนรู้ (Training data set) จำนวน 935 ระเบียบ และข้อมูลสำหรับการทดสอบตัวแบบ (Testing data set) 315 ระเบียบ ให้เหมาะสมกับการวิเคราะห์ด้วยเทคนิคการเชื่อมโยงความสัมพันธ์ โดยการปรับข้อมูลบางประเภทให้เป็นช่วงข้อมูล เพื่อให้ง่ายต่อการวิเคราะห์ เช่น ข้อมูลเกี่ยวกับอายุ จัดเป็นช่วงอายุ ข้อมูลจังหวัดภูมิสำเนา จัดเป็นเขตจังหวัด ภูมิสำเนา และข้อมูลระดับผลการเรียน จัดเป็นกลุ่มระดับผลการเรียน ดังแสดงในตารางที่ 2

ตารางที่ 1 แสดงรายละเอียดของแอททริบิวต์ที่เกี่ยวข้อง

ลำดับ	ชื่อแอททริบิวต์	ชนิดข้อมูล	รายละเอียด
1	GENDER_ID	NUMBER	รหัสเพศ
2	CURR_AGE	TEXT	ช่วงอายุปัจจุบัน คำนวณจากปีปัจจุบัน - ปีเกิด
3	ADMIT_AGE	TEXT	ช่วงอายุเมื่อเข้าศึกษา (ปีที่เข้าศึกษา - ปีเกิด)
4	PROVINCE_AREA	NUMBER	พื้นที่จังหวัดภูมิสำเนา
5	GPA_STATUS*	TEXT	ระดับผลการเรียน (ช่วงเกรดเฉลี่ยสะสม)
6	FUND_STATUS_ID	NUMBER	รหัสสถานภาพการรับทุน (ทุนทุกประเภท)
7	TALENT	TEXT	ความถนัด / ความสนใจพิเศษ
8	SON_NUM	NUMBER	จำนวนพี่น้องทั้งหมด
9	SON_STDNUM	NUMBER	จำนวนพี่น้องที่กำลังศึกษาอยู่
10	FAT_STATUS_ID	NUMBER	รหัสสถานภาพของบิดา

ลำดับ	ชื่อแอททริบิวต์	ชนิดข้อมูล	รายละเอียด
11	FAT_REVENUE_ID	NUMBER	รหัสรายได้บิดา
12	FAT_OCCUP_ID	NUMBER	รหัสอาชีพบิดา
13	MOT_STATUS_ID	NUMBER	รหัสสถานภาพของมารดา
14	MOT_REVENUE_ID	NUMBER	รหัสรายได้มารดา
15	MOT_OCCUP_ID	NUMBER	รหัสอาชีพมารดา
16	PAR_STATUS_ID	NUMBER	รหัสสถานภาพบิดา-มารดา
17	PAR_REVENUE_ID	NUMBER	รหัสรายได้ผู้ปกครอง
18	PAR_OCCUP_ID	NUMBER	รหัสอาชีพผู้ปกครอง

\*GPA\_STATUS ใช้เป็น CLASS ในการจัดระดับความเหมาะสมของผู้เรียน

ตารางที่ 2 แสดงชุดข้อมูลที่ได้เตรียมการเพื่อการเรียนรู้ของระบบ

GENDER_ID	CURR_AGE	ADMIT_AGE	PROVINCE_AREA	PAR_OCCUP_ID	...	GPA_STATUS
2	25-30	20-25	3	07	...	Good
1	20-25	<20	1	04	...	Poor
1	>35	>35	1	07	...	Excellent
1	20-25	<20	2	01	...	Good
1	20-25	<20	1	03	...	Good
...	...	...	...	...	...	...

##### 4.2 การสร้างกฎจำแนกประเภทเชิงความสัมพันธ์

นำข้อมูลที่ได้จัดเตรียมสำหรับกระบวนการเรียนรู้ นำเข้าสู่เทคนิคการวิเคราะห์และเชื่อมโยงความสัมพันธ์ เพื่อค้นหาความเชื่อมโยงของปัจจัยต่างๆ กับประเภทของข้อมูลที่ต้องการซึ่งในกระบวนการจะกำหนดให้เป็นค่าคงที่ตามจำนวนประเภทที่ต้องการคือระดับผลการเรียน 4 ระดับ ดังนั้นจึงพิจารณาและสร้างเฉพาะกฎที่มีปัจจัยต่างๆ ส่งผลกระทบต่อระดับผลการเรียน โดยนำเข้าสู่ข้อมูลสู่กระบวนการวิเคราะห์ความสัมพันธ์เพื่อสร้างกฎเชื่อมโยงความสัมพันธ์ ดังแสดงในภาพที่ 1

ขั้นตอนในการสร้างกฎจำแนกประเภทเชิงสัมพันธ์สามารถแบ่งออกเป็น 2 ขั้นตอนหลัก ซึ่งมีลักษณะการทำงานดังนี้

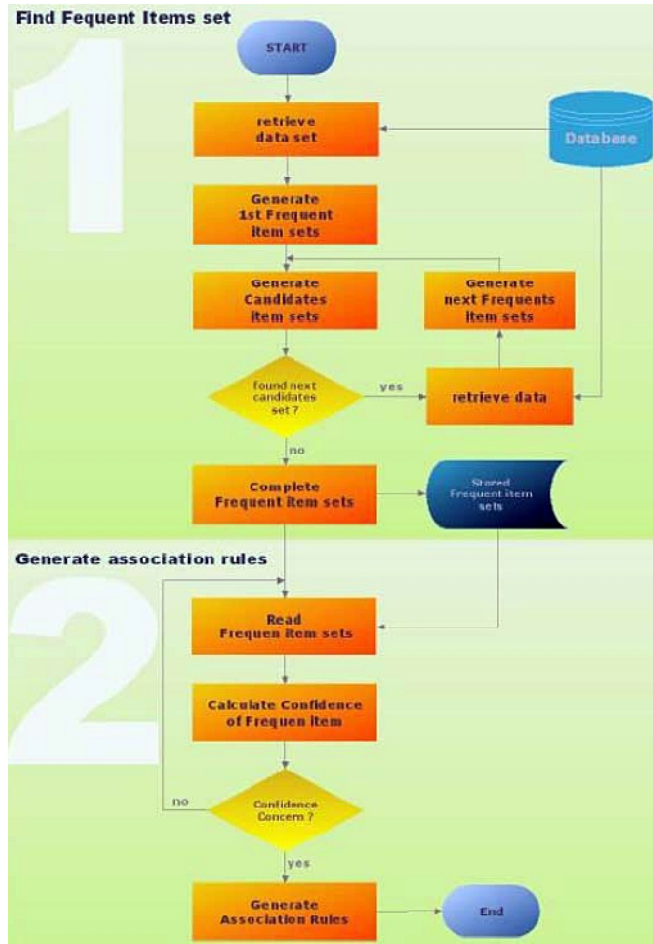
1) นำเข้าสู่ข้อมูลสำหรับการเรียนรู้ และค้นหากลุ่มข้อมูลที่มีความสัมพันธ์กับข้อมูล สำหรับจำแนกประเภท ทำการนับค่าความถี่ของชุดข้อมูลที่พบ (Frequent Itemsets) และนำเข้าสู่ข้อมูลรอบถัดไปสำหรับการจำแนกประเภท แล้วนับความถี่ของชุดข้อมูลจนกระทั่งครบ

2) สร้างกฎจาก Frequent Itemsets ที่ได้จากการดำเนินการในขั้นตอนข้างต้น โดยคำนวณหาค่าน้ำหนัก คือ ค่า



สนับสนุน (Supt.) และค่าความเชื่อมั่น (Conf.) ที่ได้จากการนับ (Frequent Itemsets)

จากเทคนิคกฎการจำแนกประเภทเชิงความสัมพันธ์ ทำให้ได้กฎสำหรับใช้จำแนก ทำการบันทึกกฎทั้งหมดไว้ในรูปแบบตารางในฐานข้อมูล ดังแสดงตัวอย่างในตารางที่ 3



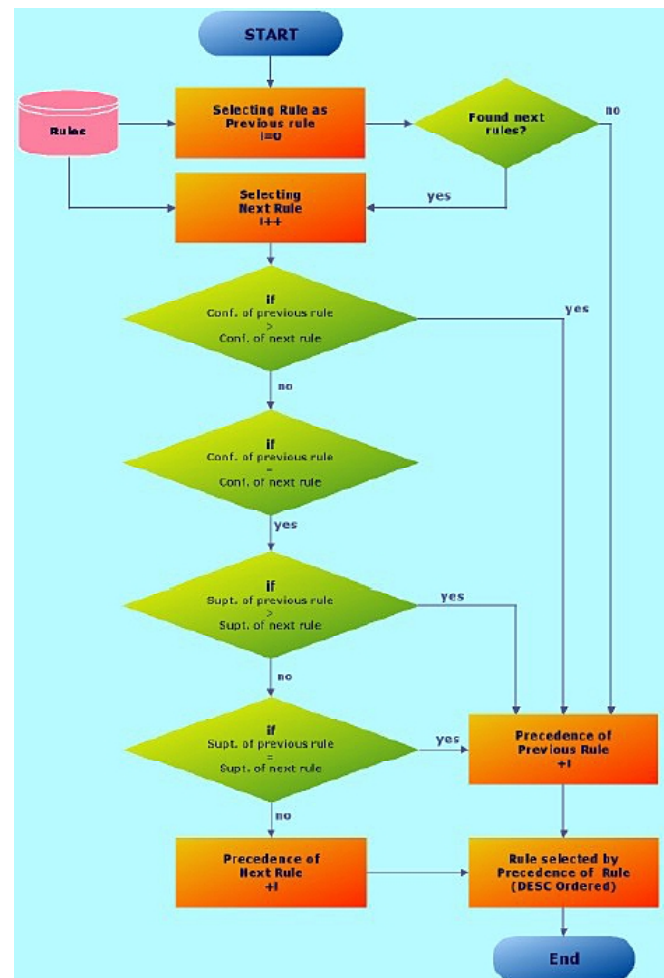
ภาพที่ 1 กระบวนการหาและกฎจำแนกประเภทเชิงความสัมพันธ์

ตารางที่ 3 แสดงตัวอย่างกฎที่บันทึกอยู่ในรูปแบบตารางในฐานข้อมูล

Rules	Supt	Conf
[GENDER_ID=2]+[CURR_AGE=20-25]+[ADMIT_AGE=ต่ำกว่า 20]+[PROVINCE_AREA=2]+[FUND_STATUS_ID=0] --> [GPA_STATUS=Good]	0.01	1
[GENDER_ID=2]+[CURR_AGE=20-25]+[ADMIT_AGE=ต่ำกว่า 20]+[PROVINCE_AREA=2]+[FUND_STATUS_ID=0]+[FAT_STATUS_ID=1] --> [GPA_STATUS=Good]	0.01	1
[GENDER_ID=2]+[CURR_AGE=20-25]+[ADMIT_AGE=ต่ำกว่า 20]+[PROVINCE_AREA=2] --> [GPA_STATUS=Good]	0.01	0.88
...	...	...

### 4.3 การใช้กฎจำแนกประเภทเชิงความสัมพันธ์

การนำกฎจำแนกประเภทเชิงความสัมพันธ์ที่ถูกสร้างขึ้นมาประยุกต์ใช้เพื่อพยากรณ์ลักษณะความเหมาะสมของนักศึกษาใหม่ โดยพิจารณาจากระดับผลการเรียนทำได้โดยการนำกฎการจำแนกประเภทเชิงความสัมพันธ์ที่ได้ นำมาคัดเลือกตามระดับค่าความสำคัญ (Precedence) จากมากไปน้อย โดยจะพิจารณาค่าความเชื่อมั่นที่ได้ และค่าสนับสนุนที่ทำการคำนวณไว้ของกฎที่พบก่อนหน้าเปรียบเทียบกับกฎถัดไป ดังแสดงกระบวนการพิจารณาในภาพที่ 2



ภาพที่ 2 กระบวนการพิจารณาและกำหนดค่าความสำคัญของกฎจำแนกประเภทเชิงสัมพันธ์

ถ้าค่าความเชื่อมั่นที่พบในกฎก่อนหน้า มากกว่าค่าความเชื่อมั่นที่พบในกฎถัดไป ให้กฎที่พบก่อนหน้ามีค่าความสำคัญมากกว่ากฎถัดไป ถ้าค่าความเชื่อมั่นที่พบในกฎก่อนหน้า เท่ากับค่าความเชื่อมั่นที่พบในกฎถัดไป ให้พิจารณา ค่าสนับสนุนของกฎทั้งสอง ถ้าค่าสนับสนุนของกฎที่พบก่อนหน้า มากกว่าค่าสนับสนุนของกฎถัดไป ให้กฎที่พบ



ก่อนหน้า มีค่าความสำคัญมากกว่ากฎถัดไป และถ้าค่าความเชื่อมั่นและค่าสนับสนุนของกฎที่พบก่อนหน้า และกฎถัดไป มีค่าเท่ากันให้กฎที่มีก่อนมีค่าความสำคัญมากกว่า

เนื่องจากเครื่องมือที่ได้พัฒนาขึ้นนั้น ได้ทำการจัดเก็บกฎการจำแนกประเภทเชิงความสัมพันธ์อยู่ในรูปแบบของตารางไว้ในฐานข้อมูล ดังนั้นการเรียงลำดับค่าความสำคัญของกฎสามารถทำได้ด้วยคำสั่ง SQL ซึ่งกฎที่ค้นได้จากคำสั่ง SQL ดังกล่าวจะผ่านการจัดเรียงค่าความสำคัญแล้ว จึงสามารถนำมาใช้เป็นกฎสำหรับจำแนกข้อมูลตามประเภทที่ผ่านการจัดระดับความสำคัญที่พร้อมสำหรับนำไปใช้ต่อไป

## 5. ผลการดำเนินงาน

จากการสร้างตัวแบบพยากรณ์คุณลักษณะความเหมาะสมของนักศึกษาสาขาเทคโนโลยีสารสนเทศ ด้วยเทคนิคกฎการจำแนกเชิงความสัมพันธ์ ผลการดำเนินการพบว่ามีจำนวนกฎที่ถูกสร้างขึ้นทั้งสิ้น 8,308 กฎ ซึ่งจำแนกเป็นกฎสำหรับจำแนกระดับความเหมาะสม จากระดับผลการเรียนในระดับดีเยี่ยม 2,239 กฎ ระดับดี 2,586 กฎ ระดับปานกลาง 2,272 กฎ และระดับอ่อน 1,211 กฎ โดยมีกฎที่เป็นไปตามค่าความเชื่อมั่น ขั้นต่ำที่กำหนด (0.9) 157 กฎ ซึ่งจำแนกเป็นกฎสำหรับจำแนกระดับความเหมาะสมจากระดับผลการเรียน ในระดับดีเยี่ยมจำนวน 13 กฎ ระดับดี 40 กฎ ระดับปานกลาง 61 กฎ และระดับอ่อน 43 กฎ ดังแสดงในตารางที่ 4

ตารางที่ 4 แสดงกฎจำแนกตามประเภท (ระดับผลการเรียน)

ระดับความเหมาะสม (ระดับผลการเรียน)	จำนวนชุดข้อมูลเรียนรู้ (Training data set)	จำนวนกฎที่ได้	จำนวนกฎที่ได้เมื่อกำหนดค่าความเชื่อมั่นขั้นต่ำ (Min. Conf = 0.9)
Poor (ต่ำกว่า 2.00)	92	1,211	43
Fair (2.00-2.49)	346	2,272	61
Good (2.50-2.99)	307	2,586	40
Excellent (3.00 ขึ้นไป)	190	2,239	13
รวม	935	8,308	157

### การทดสอบกฎจำแนกประเภทเชิงความสัมพันธ์

ในส่วนของการทดสอบข้อมูล ใช้ชุดข้อมูลซึ่งเตรียมไว้สำหรับการทดสอบ (Testing dataset) จำนวน 315 ระเบียบนำมาผ่านกระบวนการของตัวแบบฯ ที่ถูกสร้างขึ้น โดยที่สมมติระเบียบข้อมูลทดสอบเป็นข้อมูลนักศึกษาเข้าใหม่ที่ต้องการคาดการณ์ว่าจะมีระดับความเหมาะสมอยู่ในระดับใด ทำการทดสอบเปรียบเทียบกับกฎที่ใกล้เคียง ตามกระบวนการ

การของตัวแบบฯ ที่แบ่งเป็น 2 ขั้นตอน ได้แก่

1) นำเข้าข้อมูลนักศึกษาเข้าใหม่ โดยจัดรูปแบบข้อมูลให้พร้อมสำหรับการเปรียบเทียบกับกฎ

2) คัดเลือกกฎที่สร้างขึ้น โดยเรียงลำดับตามค่าความสำคัญ และเนื่องจากได้มีการจัดเก็บกฎไว้ในฐานข้อมูล จึงทำให้การเรียงลำดับค่าความสำคัญของกฎสามารถทำได้ด้วยภาษา SQL

กฎที่ถูกคัดเลือกด้วยการเรียงลำดับค่าความสำคัญแล้ว คือผลการคาดการณ์ระดับความเหมาะสมของข้อมูลที่ได้ทำการทดสอบจนครบทุกระเบียนข้อมูล และเปรียบเทียบผลการคาดการณ์ของตัวแบบกับ ระดับผลการเรียนจริง จากนั้นจึงคำนวณหาความถูกต้องของการคาดการณ์ (ร้อยละ) ของระดับผลการเรียนที่แท้จริง ดังในตารางที่ 5

ตารางที่ 5 แสดงผลการทดสอบกฎจำแนกตามประเภท

ระดับความเหมาะสม (ระดับผลการเรียน)	จำนวนชุดข้อมูลเรียนรู้ (Training data set)	จำนวนชุดข้อมูลทดสอบ (Testing data set)	ความถูกต้องของผลที่ได้	
			ระเบียบ	ร้อยละ
Poor (ต่ำกว่า 2.00)	92	20	5	25.00
Fair (2.00-2.49)	346	84	64	76.19
Good (2.50-2.99)	307	195	193	98.97
Excellent (3.00 ขึ้นไป)	190	16	2	12.50
รวม	935	315	264	83.81

เมื่อทำการทดสอบกฎจำแนกประเภทเชิงสัมพันธ์ที่สร้างขึ้น ด้วยชุดข้อมูลทดสอบพบว่า กฎสำหรับจำแนกระดับความเหมาะสมของนักศึกษาในระดับเป็นกฎที่มีความสามารถในการจำแนกข้อมูลได้ดีที่สุด (ร้อยละ 98.97) ในขณะที่กฎจำแนกระดับความเหมาะสมระดับดีเยี่ยม และระดับต่ำ มีความสามารถในการจำแนกได้ไม่ดีเท่าที่ควร ทั้งนี้เนื่องจากข้อมูลสำหรับเรียนรู้ของระดับความเหมาะสมประเภทดังกล่าวมีน้อย ทำให้ไม่สามารถจำแนกไม่ดีเท่าที่ควร โดยเมื่อคำนวณประสิทธิภาพในการพยากรณ์โดยรวมแล้ว กฎที่ได้มีความสามารถจำแนกระดับความเหมาะสมของนักศึกษาอยู่ในระดับดี (ร้อยละ 83.81)

## 6. บทสรุปและข้อเสนอแนะ และแนวทางพัฒนาต่อ

งานวิจัยนี้เป็นการพัฒนาตัวแบบ เพื่อพยากรณ์ลักษณะความเหมาะสมต่อการเข้าศึกษาในสาขาวิชาเทคโนโลยีสารสนเทศ ด้วยเทคนิคกฎการจำแนกประเภทเชิงความสัมพันธ์ ใช้ข้อมูลนักศึกษาสาขาวิชาเทคโนโลยีสารสนเทศ



มหาวิทยาลัยราชภัฏจันทรเกษม เป็นกรณีศึกษา โดยกำหนดกลุ่มจำแนกระดับความเหมาะสมของนักศึกษาใหม่ออกเป็น 4 กลุ่มตามระดับผลการเรียนอันได้แก่ ระดับดีเยี่ยม ระดับดี ระดับปานกลาง และระดับต่ำ ผลการทดลองพบว่า สามารถจำแนกระดับความเหมาะสมของนักศึกษาได้ในระดับดี (ร้อยละ 83.81) แม้ว่าจะพบข้อจำกัดในเรื่องของการนำเข้าข้อมูลสำหรับเรียนรู้ที่ระดับความเหมาะสมดีเยี่ยมและระดับต่ำน้อยเกินไป ทำให้ประสิทธิภาพการพยากรณ์ความเหมาะสมของนักศึกษาในประเภทดังกล่าวไม่ดีเท่าที่ควร ซึ่งหากมีการนำเข้าข้อมูลสำหรับการเรียนรู้ ได้มากใกล้เคียงกับข้อมูลจริงตัวแบบเพื่อพยากรณ์ลักษณะความเหมาะสมของนักศึกษาใหม่ จะสามารถใช้พยากรณ์คุณลักษณะความเหมาะสมของนักศึกษาได้ดียิ่งขึ้น

ทั้งนี้กฎการจำแนกประเภทเชิงความสัมพันธ์ ระหว่างปัจจัยที่ค้นพบ รวมถึงลักษณะความเหมาะสมที่ได้ สามารถนำไปประยุกต์ใช้กับข้อมูลนักศึกษาในสาขาอื่นๆ ได้ และเป็นแนวทางสำหรับการจัดการเรียน กลุ่มเรียน รวมถึงรายวิชาในแผนการเรียนในหลักสูตรให้เหมาะสมกับผู้เรียน ซึ่งเป็นกระบวนการที่จะเกิดขึ้นในงานวิจัยฉบับต่อไป

## 7. เอกสารอ้างอิง

- [1] จิราพร ยิ่งกว่าชาติ, ประสงค์ ปราณิตพลกรัง. การทำนายผลสำเร็จการศึกษาของนักศึกษาระดับอุดมศึกษาด้วยการเรียนรู้แบบเบย์และเหมืองข้อมูล, มหาวิทยาลัยศรีปทุม, 2549.
- [2] ประชุม ชนะ. ปัจจัยที่มีผลต่อการศึกษาต่อในระดับปริญญาตรีของนักเรียนระดับมัธยมศึกษาตอนปลายปีที่ 6 และนักศึกษาระดับประกาศนียบัตรวิชาชีพชั้นสูงปีที่ 2 ในสถานศึกษาจังหวัดชลบุรี, มหาวิทยาลัยศรีปทุม วิทยาเขตชลบุรี, 2545.
- [3] ปิณฑิต สุภากุล. "การศึกษาปัจจัยที่มีผลต่อการผันสภาพของนักศึกษามหาวิทยาลัยพายัพ"วารสารมหาวิทยาลัยพายัพ, vol. 17, pp. 95-109, 2547.
- [4] W. Green. "Demographic and psychological factors influencing academic success in a college level human anatomy course," **The school of Human Resource Education and Workforce Development:** Louisiana State University., 2007.
- [5] เพียงพบ มนต์นวลปรารงค์ การศึกษาปัจจัยที่ส่งผล
- [6] พร พรมมหาราช. การศึกษาผลสัมฤทธิ์และปัจจัยที่มีผลต่อผลสัมฤทธิ์ทางการเรียนของนักศึกษาในระดับปริญญาตรีคณะวิทยาศาสตร์. มหาวิทยาลัยเชียงใหม่, 2541.
- [7] B. W. Amy, M. H. Hisham, and F. Z. Anita. "Towards a model of student success in programming courses." **Proceedings of the 43rd annual Southeast regional conference - Volume 1.** Kennesaw, Georgia: ACM,2005.
- [8] M. Yiming, L. Bing, W. Ching Kian, S. Y. Philip, and L. Shuik Ming. "Targeting the right students using data mining." in **Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining.** Boston, Massachusetts, United States: ACM, 2000.
- [9] กฤษณะ ไวยมัย, ธนาวิพันธ์ รักธรรมานนท์. "การใช้เทคนิคดาต้าไมนิ่งเพื่อพัฒนาคุณภาพการศึกษาคณะวิศวกรรมศาสตร์," **NECTEC Technical Journal**, vol. 3, pp. 134-142, 2001.
- [10] R. Anthony. "The demographics of candidates for faculty positions in computer science," **Commun. ACM**, vol. 39, pp. 78-84, 1996.
- [11] C. Ordonez and C. Ordonez. "Association rule discovery with the train and test approach for heart disease prediction Association rule discovery with the train and test approach for heart disease prediction," **IEEE Transactions on Information Technology in Biomedicine**, vol. 10, pp. 334-343, 2006.
- [12] A. Michail and A. Michail. "Data mining library reuse patterns using generalized association rules Data mining library reuse patterns using generalized association rules," **Proceedings of the 2000 International Conference on Software Engineering**, 2000.
- [13] W. H. Bing Liu, Yiming Ma. "Integrating Classifica-



tion and Association Rule Mining.” **Knowledge Discovery and Data Mining**, 1998.

[14] M. J. A. B. a. G. Linoff. **Data Mining Techniques**, John Wiley & Sons., 1997.

[15] K. H. Shimada, K. Jinglu Hu. “Class Association Rule Mining with Chi-Squared Test Using Genetic Network Programming,” **Proceeding of International Conference on Systems, Man and Cybernetics, (ICSMC’06)**, 2006.

