

การพยากรณ์ลักษณะความเหมาะสมของนักศึกษาใหม่ โดยใช้กฎการจำแนกประเภทเชิงความสัมพันธ์แบบเต็มรูป A Forecasting Model to Evaluate a Freshman's Ability to Succeed by Using Particular Full-Scaled Class Association Rules (PFSCARs)

จามรกุล เหล่าเกียรติกุล
ภาควิชาเทคโนโลยีสารสนเทศ
มหาวิทยาลัยเทคโนโลยีพระจอมเกล้า
พระนครเหนือ
jamornkul@chandra.ac.th

ณัฐวี อดตฤกษ์
ภาควิชาการจัดการเทคโนโลยีสารสนเทศ
มหาวิทยาลัยเทคโนโลยีพระจอมเกล้า
พระนครเหนือ
nattaveeu@kmutnb.ac.th

พยุง มีสังข์
ภาควิชาครุศาสตร์ไฟฟ้า
มหาวิทยาลัยเทคโนโลยีพระจอมเกล้า
พระนครเหนือ
pym@kmutnb.ac.th

บทคัดย่อ

การวิจัยนี้เป็นการสร้างตัวแบบสำหรับการพยากรณ์ลักษณะความเหมาะสมต่อการเข้าศึกษาของนักศึกษาใหม่ ด้วยเทคนิคกฎการจำแนกเชิงความสัมพันธ์แบบเต็มรูป โดยมีวัตถุประสงค์เพื่อพยากรณ์คุณลักษณะความเหมาะสมของนักศึกษา โดยใช้ข้อมูลนักศึกษาสาขาเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏจันทรเกษมเป็นกรณีศึกษา และใช้ข้อมูลระดับผลการเรียนปัจจุบัน เป็นปัจจัยสำหรับกำหนดกลุ่มระดับความเหมาะสมของนักศึกษาใหม่ ซึ่งจำแนกออกเป็น 3 กลุ่มได้แก่ กลุ่มนักศึกษาระดับดี ระดับปานกลาง และระดับต่ำ ผลการทดลองพบว่าการจำแนกประเภทเชิงความสัมพันธ์แบบเต็มรูป สามารถจำแนกระดับความเหมาะสมของนักศึกษาได้ในระดับดี (ร้อยละ 78) ซึ่งสามารถนำไปประยุกต์ใช้ในการพยากรณ์ลักษณะความเหมาะสม ของนักศึกษาใหม่ในสาขาวิชาต่างๆ ได้ ทั้งนี้เพื่อใช้ประโยชน์ในการจัดการเรียนการสอนให้เหมาะสมกับผู้เรียน รวมถึงการพัฒนากระบวนการจัดการเรียนการสอนอื่นๆ ต่อไป

คำสำคัญ

ความเหมาะสมของนักศึกษา, กฎการจำแนกเชิงความสัมพันธ์, การประเมิน

Abstract

This study was to create a forecasting model for evaluate freshmen's ability to succeed with using the longest rules from CARs technique as called a particular full-scaled class association rules (PFSCARs). The purposed of this study was to create a classifier tool to evaluate freshmen's ability. This study used demographic data of students in Information Technology program Chandrakasem Rajabhat University and current grade levels of students as equalized levels. The equalized levels consist of 3

classes: good, fair and poor. The result of this study that proposed that the forecasting model to evaluate freshmen quality with PFSCARs performed at a good level of performance with an accuracy rate of 79% of the students equalized model. Finally, the research discovered that the forecasting model to evaluate freshmen quality can be a guideline for academic advisors or other relevant persons to help new students, to manage an appropriated study plan and could help them to improve course or curriculum in the future as well.

Keyword

Freshmen's Quality, Class Association Rules, Evaluation

1. บทนำ

ในขั้นตอนการคัดเลือกนักศึกษาใหม่ ของสถานศึกษาระดับ อุดมศึกษา เพื่อรับเข้าศึกษาในสาขาวิชาต่างๆ อาจใช้วิธีคัดเลือกกับผู้สมัครเพื่อเข้าศึกษาด้วยการทดสอบความสามารถทางวิชาการ หรือการสอบสัมภาษณ์ ซึ่งเมื่อมีผู้ผ่านการคัดเลือกเข้ามาเป็นนักศึกษาแล้ว พบว่ามีส่วนหนึ่งพบสภาพการเป็นนักศึกษาในเวลาต่อมา ดังนั้นหากมีการนำเอาข้อมูลส่วนบุคคลที่นักศึกษาให้ไว้ก่อนเข้าเป็นนักศึกษาของนักศึกษาปัจจุบันและอดีต นำมาวิเคราะห์เพื่อค้นหาปัจจัยที่ส่งผลกระทบต่อคุณลักษณะที่เหมาะสมต่อสาขาวิชา จะทำให้สามารถใช้ในคาดการณ์เพื่อจัดจำแนกกลุ่มนักศึกษาใหม่เพื่อปรับ หรือจัดการเรียนการสอนให้ เหมาะสมกับนักศึกษาต่อไป

งานวิจัยนี้มีวัตถุประสงค์เพื่อ สร้างตัวแบบในการประเมินและจำแนก ลักษณะความเหมาะสมต่อการเข้าศึกษา โดยประยุกต์ใช้เทคนิคการสร้างกฎการจำแนกเชิงความสัมพันธ์ เพื่อการวิเคราะห์รูปแบบความสัมพันธ์ของข้อมูลส่วนบุคคลของนักศึกษาที่ส่งผลกระทบต่อระดับผลการเรียน นำมาสร้างเป็นกฎเพื่อเชื่อมโยงความสัมพันธ์เฉพาะกฎที่ยาวที่สุด ซึ่งงานวิจัยนี้

เรียกว่า “กฎแบบเต็มรูป” (Particular Full-scaled CARs) โดยเลือกศึกษาข้อมูลคุณลักษณะส่วนบุคคลของนักศึกษา สาขาวิชาเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏจันทรเกษม เป็นกรณีศึกษา และใช้กฎแบบเต็มรูปดังกล่าว เพื่อการคาดการณ์ลักษณะความเหมาะสม โดยกำหนดเป็น 3 กลุ่มได้แก่ กลุ่มนักศึกษาระดับดี กลุ่มนักศึกษาระดับปานกลาง และกลุ่มนักศึกษาระดับต่ำ ซึ่งรูปแบบที่ได้สามารถนำไปใช้ประโยชน์ในการปรับแผนการเรียน หรือจัดกลุ่มเรียน รวมถึงการเรียนการสอน ให้เหมาะสมกับนักศึกษาในลักษณะต่างๆ ต่อไป รวมไปถึงพัฒนาระบบสนับสนุนงานด้านการเรียนการสอน ถือเป็นการสร้างโอกาสให้แก่ผู้สนใจเข้าศึกษาในสาขาวิชาเทคโนโลยีสารสนเทศ ได้มีโอกาสในการศึกษาได้อย่างเหมาะสม และเพิ่มโอกาสสำเร็จการศึกษายิ่งขึ้น

2. งานวิจัยที่เกี่ยวข้อง

จากการค้นคว้างานวิจัย ในลักษณะของการค้นหาปัจจัยจากข้อมูลคุณลักษณะของบุคคล ที่ส่งผลกระทบต่อผลสัมฤทธิ์ทางการศึกษา พบงานวิจัยที่ศึกษาในระดับการศึกษาลากหลายระดับด้วยกัน เช่น งานวิจัยในการวิเคราะห์ปัจจัยที่ส่งผลกระทบต่อผลการเลือกเข้าศึกษาในระดับมหาวิทยาลัย ที่ค้นพบว่า ระดับผลการเรียนเดิม อาชีพของผู้ปกครอง ระดับการศึกษาของผู้ปกครอง และรายได้ของครอบครัว ส่งผลกระทบต่อผลสัมฤทธิ์ทางการศึกษา [1-5] งานวิจัยเกี่ยวกับปัจจัย ที่ส่งผลกระทบต่อผลสัมฤทธิ์ทางการเรียนในรายวิชา หรือสาขาวิชาต่างๆ ที่ค้นพบปัจจัยเกี่ยวกับความสนใจ การเข้าห้องเรียน และคะแนนเก็บ [2] งานวิจัยที่ศึกษาปัจจัยที่ส่งผลกระทบต่อผลการเรียน [6] ซึ่งใช้การวิเคราะห์ข้อมูลทางสถิติ มาใช้ในการทำวิจัย และงานวิจัยที่ใช้เทคนิคต่างๆ ของกรรมวิธีเหมืองข้อมูลในการวิเคราะห์ข้อมูลนักศึกษา [7-9] รวมไปถึงการศึกษาปัจจัยที่ส่งผลกระทบต่อคัดเลือกบุคคลเข้ารับตำแหน่งในสาขาวิชาชีพ[2]

ผลจากงานวิจัยเกี่ยวกับคุณลักษณะของบุคคลข้างต้นแสดงให้เห็นว่าคุณลักษณะส่วนบุคคลส่งผลกระทบต่อความสามารถในการศึกษาในระดับต่างๆ แตกต่างกันไป

ในขณะที่มีงานวิจัยที่ประยุกต์ใช้ เทคนิคการสร้างกฎการเชื่อมโยงความสัมพันธ์ รวมถึงเทคนิคกฎการจำแนกประเภทเชิงความสัมพันธ์ มากมายหลายด้านด้วยกัน เช่น งานวิจัยที่ประยุกต์ใช้กฎการเชื่อมโยงความสัมพันธ์กับการวิเคราะห์ข้อมูลอาการหัวใจล้มเหลว [10] การประยุกต์ใช้กับการจัดการข้อมูลภายในห้องสมุด [11] การประยุกต์ใช้เทคนิคจำแนกเชิงความสัมพันธ์ในการวิเคราะห์การติดเชื้อ HIV-1 [12] เป็นต้น

จากงานวิจัยข้างต้น จะเห็นได้ว่า การประยุกต์ใช้เทคนิคกฎการจำแนกประเภทเชิงความสัมพันธ์ สามารถนำมาใช้ในงานด้านต่างๆ มากมาย เนื่องจากใช้หน่วยความจำ รวมถึงประสิทธิภาพของหน่วยประมวลผลไม่สูง อีกทั้งยังง่ายต่อการนำมาประยุกต์ใช้งาน ผู้วิจัยจึงมีแนวคิดในการนำเทคนิคดังกล่าว มาใช้วิเคราะห์ข้อมูลของนักศึกษา เพื่อค้นหาความสัมพันธ์ของข้อมูลคุณลักษณะส่วนบุคคล และนำเอาปัจจัย

ดังกล่าว มาประยุกต์ใช้ในการจำแนกระดับความเหมาะสมของผู้เรียน โดยเชื่อมั่นว่าเทคนิคดังกล่าวจะสามารถวิเคราะห์ปัจจัยที่ส่งผลกระทบต่อความเหมาะสมของนักศึกษา และสามารถนำมาใช้คาดการณ์อย่างมีประสิทธิภาพ

3. กฎการจำแนกประเภทเชิงความสัมพันธ์ (CARs)

กฎการจำแนกประเภทเชิงความสัมพันธ์ (CARs) เป็นเทคนิคของกรรมวิธีเหมืองข้อมูล ที่ผสมผสานระหว่างเทคนิคการวิเคราะห์เพื่อสร้างกฎเชื่อมโยงความสัมพันธ์ (Association -Rules Discovery) เข้ากับวิธีการจำแนกกลุ่มข้อมูล (Data Classification) โดยมีวัตถุประสงค์เพื่อใช้ประโยชน์ กฎการเชื่อมโยงความสัมพันธ์ที่ถูกสร้างขึ้นในการจำแนกข้อมูล โดยกฎที่ถูกคัดเลือกเพื่อนำมาใช้ในจำแนกข้อมูลนี้จะถูกเรียกว่า Class Association Rules (CARs) [8]

3.1 เทคนิควิเคราะห์เพื่อสร้างกฎเชื่อมโยงความสัมพันธ์ (Association Rules Discovery)

เทคนิคการวิเคราะห์เพื่อสร้างกฎเชื่อมโยงความสัมพันธ์นี้ เป็นเทคนิคหนึ่งในกรรมวิธีเหมืองข้อมูล ที่ได้รับความนิยมอย่างกว้างขวางในการวิจัยเชิงประยุกต์ใช้เหมืองข้อมูล กับข้อมูลด้านต่างๆ กฎของความสัมพันธ์ระหว่างข้อมูลนั้นจะถูกสร้างขึ้นโดยการค้นหาความถี่ที่เกิดขึ้น (Frequency items set) ที่อยู่ภายใต้ค่าสนับสนุนหรือค่าที่ยอมรับได้ (Support) และค่าความเป็นไปได้ของแต่ละปัจจัยที่นำมาพิจารณา (Confidence) [13]

การประยุกต์ กฎเชื่อมโยงความสัมพันธ์ เพื่อใช้เป็นกฎการจำแนกประเภทเชิงความสัมพันธ์ จะเลือกใช้เฉพาะกฎที่ส่งผลกระทบต่อประเภทของข้อมูลที่ต้องการ

3.2 การหาและสร้างกฎจำแนกประเภทเชิงความสัมพันธ์

กฎจำแนกประเภทเชิงความสัมพันธ์ เป็นกฎที่แตกต่างจากการสร้างกฎความสัมพันธ์ของข้อมูล (Association rules) ที่บอกความสัมพันธ์ของข้อมูลทั้งหมด โดยไม่เฉพาะเจาะจงกับข้อมูลที่กำหนด ในขณะที่กฎจำแนกประเภทเชิงความสัมพันธ์ นั้นจะบอกความสัมพันธ์ของข้อมูลต่อประเภทของข้อมูลนั้นๆ (Class) ซึ่งเป็นการค้นหาเฉพาะความสัมพันธ์ที่เฉพาะเจาะจงกับประเภทของข้อมูล โดยไม่คำนึงถึงความสัมพันธ์กับข้อมูลอื่นๆ [14] โดยมีลักษณะ ดังแสดงในสมการที่ 1

$$\{item_1, \dots, item_j\} \rightarrow Class : (Supp.), (Conf.) \dots \dots \dots (1)$$

เมื่อ item i-j คือ ข้อมูลที่สัมพันธ์กัน และ Class คือประเภทของข้อมูลนั้นๆ โดยที่ต้องคำนวณหาค่าสนับสนุน (Supp.) จากจำนวนสมาชิกที่สอดคล้องตามกฎต่อจำนวนข้อมูลทั้งหมด และค่าความเชื่อมั่น (Conf.) จากจำนวนข้อมูลที่สอดคล้องตามกฎต่อจำนวนสมาชิกในกฎนั้น ดังแสดงในสมการที่ 2

$$Conf. = \frac{Supp.(X \cap Y)}{Supp.(X)} \dots\dots\dots (2)$$

กฎที่ได้จากกระบวนการดังกล่าว จะมีเป็นจำนวนขึ้นอยู่กับปัจจัย และข้อมูลที่นำมาผ่านกระบวนการ ซึ่งจะได้นำมาผ่านกระบวนการในการคัดสรร กฎที่ยอมรับได้ภายใต้เงื่อนไขที่ผู้ใช้เทคนิคเป็นผู้กำหนดได้แก่ ค่าสนับสนุนขั้นต่ำ และค่าความเชื่อมั่นขั้นต่ำจัดลำดับค่าความสำคัญ (Precedence) ของกฎเพื่อให้ง่ายต่อการนำไปใช้จำแนกข้อมูล โดยพิจารณาให้ค่าความสำคัญจากกฎที่มีค่าความเชื่อมั่นที่มากกว่า และ/หรือ กฎที่มีค่าสนับสนุนที่มากกว่า ให้มีค่าความสำคัญมากกว่า

4. ตัวแบบพยากรณ์ลักษณะความเหมาะสมของนักศึกษาใหม่

4.1 จัดเตรียมข้อมูล

งานวิจัยนี้ จัดเตรียมข้อมูลโดยใช้ข้อมูลนักศึกษาสาขาวิชาเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏจันทรเกษม เป็นกรณีศึกษา โดยเตรียมข้อมูลลักษณะส่วนบุคคลจากฐานข้อมูล ที่ถูกจัดเก็บไว้ตามข้อกำหนดของสำนักคณะกรรมการการอุดมศึกษา (สกอ.) ตั้งแต่ปีการศึกษา 2547 - 2550 โดยตัดปัจจัยเช่น ชื่อ - สกุล เลขประจำตัวประชาชน และใช้ในการวิจัยทั้งสิ้น 18 ปัจจัย ดังแสดงในตารางที่ 1 ซึ่งพบ 1,003 ระเบียบที่มีข้อมูลสมบูรณ์เพียงพอที่สามารถนำมาใช้ในการสร้างตัวแบบฯ

นำข้อมูลดังกล่าวมาจัดเตรียมเป็นข้อมูลสำหรับการเรียนรู้ (Training data set) จำนวน 699 ระเบียบ และข้อมูลสำหรับการทดสอบตัวแบบ (Testing data set) 335 ระเบียบ โดยการปรับข้อมูลบางประเภทให้เป็นช่วงข้อมูล เพื่อให้ง่ายต่อการวิเคราะห์ เช่น ข้อมูลเกี่ยวกับอายุ ข้อมูลจังหวัดภูมิลำเนา และข้อมูลระดับผลการเรียน ดังแสดงในตารางที่ 1-2

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	พื้นที่จังหวัดภูมิลำเนา
5	1 st _GPA*	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
11	FAT_REVENUE_ID	NUMBER	รหัสรายได้บิดา
12	FAT_OCCUP_ID	NUMBER	รหัสอาชีพบิดา
13	MOT_STATUS_ID	NUMBER	รหัสสถานภาพมารดา
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

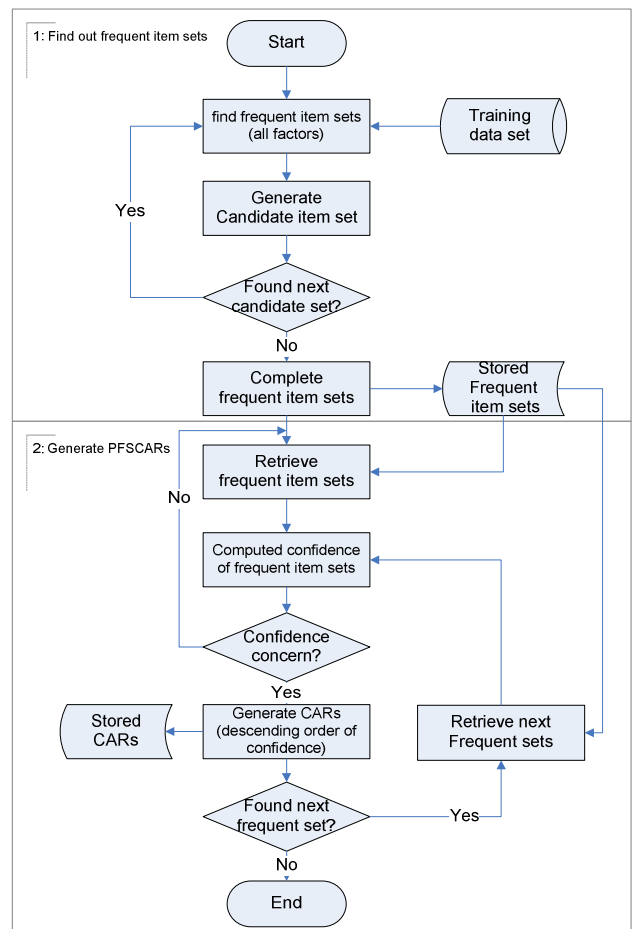
*1st_GPA ใช้เป็น CLASS ในการจัดระดับความเหมาะสมของผู้เรียน

ตารางที่ 2. ชุดข้อมูลที่เตรียมเพื่อการเรียนรู้ของระบบ

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

4.2 การสร้าง PFSCARs

ในงานวิจัยนี้ ได้มุ่งให้ความสำคัญและสร้างเฉพาะกฎเต็มรูปแบบ ที่มีคุณลักษณะเต็มรูปแบบ หรือกฎที่ยาวที่สุดที่พบ ซึ่งจะทำให้ไม่ต้องมีการวนซ้ำไปที่ละปัจจัยจนครบทั้งครั้งโดยเริ่มจาก นำข้อมูลที่จัดเตรียมสำหรับกระบวนการเรียนรู้ นำเข้าสู่กระบวนการวิเคราะห์และเชื่อมโยงความสัมพันธ์ เพื่อค้นหาความเชื่อมโยงของข้อมูลในทุกปัจจัย กับประเภทของข้อมูลที่ต้องการ ซึ่งในกระบวนการจะกำหนดให้เป็นค่าคงที่ตามจำนวนประเภทที่ต้องการคือระดับผลการเรียน 3 ระดับ ดังนั้นจึงพิจารณาและสร้างเฉพาะกฎที่มีปัจจัยต่างๆ ส่งผลต่อระดับผลการเรียน ดังแสดงในรูปที่ 1



รูปที่ 1. กระบวนการหาและกฎจำแนกประเภทเชิงความสัมพันธ์แบบเต็มรูปแบบ (PFSCARs)

ขั้นตอนในการสร้างกฎจำแนกประเภทเชิงสัมพันธ์ สามารถแบ่งออกเป็น 2 ขั้นตอนหลัก ซึ่งมีลักษณะการทำงานดังนี้

1) นำเข้าข้อมูลสำหรับการเรียนรู้ และค้นหากลุ่มข้อมูลในทุกปัจจัยที่มีความสัมพันธ์กับข้อมูล สำหรับจำแนกประเภท ทำการนับค่าความถี่ของชุดข้อมูลที่พบ (Frequent Itemsets) และนำเข้าข้อมูลรอบถัดไปสำหรับการจำแนกประเภท แล้วนับความถี่ของชุดข้อมูลจนกระทั่งครบ

2) สร้างกฎแบบเต็มรูปจาก Frequent Itemsets ที่ได้จากขั้นตอนข้างต้น โดยคำนวณค่าสนับสนุน (Supp) และค่าความเชื่อมั่น (Conf) ที่ได้จากการนับ (Frequent Itemsets) และคัดเลือกกฎโดยพิจารณาค่าสนับสนุนขั้นต่ำ (MinSupp) และค่าความเชื่อมั่นขั้นต่ำ (MinConf) ที่ยอมรับได้ ซึ่งในงานวิจัยนี้กำหนดไว้ที่ ร้อยละ 0.1 และร้อยละ 45 ตามลำดับ

จากเทคนิคกฎการจำแนกประเภทเชิงความสัมพันธ์ ทำให้ได้กฎสำหรับใช้จำแนก ทำการบันทึกกฎทั้งหมดไว้ในรูปแบบตารางในฐานข้อมูล ดังแสดงตัวอย่างในตารางที่ 3

ตารางที่ 3. ตัวอย่าง PFSCARs

PFSCARs	Supp	Conf
m_id='43', GENDER_ID = '2', CURR_AGE = '20 - 24', ADMIT_AGE = '< 20', PROVINCE_ID = '1', FUND_STATUS_ID = '0', TALENT = 'No', SON_NUM = 'None', SON_STDNUM = 'None', FAT_STATUS_ID = '1', FAT_REVENUE_ID = '2', FAT_OCCUP_ID = '2', MOT_STATUS_ID = '1', MOT_REVENUE_ID = '2', MOT_OCCUP_ID = '2', PAR_STATUS_ID = '1', PAR_REVENUE_ID = '2', PAR_OCCUP_ID = '2' ➔ 1st_GPA = 'Fair'	65.5%	8.7%
m_id='43', GENDER_ID = '1', CURR_AGE = '20 - 24', ADMIT_AGE = '20 - 24', PROVINCE_ID = '1', FUND_STATUS_ID = '0', TALENT = 'No', SON_NUM = 'None', SON_STDNUM = 'None', FAT_STATUS_ID = '1', FAT_REVENUE_ID = '2', FAT_OCCUP_ID = '2', MOT_STATUS_ID = '1', MOT_REVENUE_ID = '2', MOT_OCCUP_ID = '2', PAR_STATUS_ID = '1', PAR_REVENUE_ID = '2', PAR_OCCUP_ID = '2' ➔ 1st_GPA = 'Fair'	78.4%	11%
m_id='43', GENDER_ID = '1', CURR_AGE = '25 - 30', ADMIT_AGE = '20 - 24', PROVINCE_ID = '8', FUND_STATUS_ID = '0', TALENT = 'No', SON_NUM = 'None', SON_STDNUM = 'None', FAT_STATUS_ID = '1', FAT_REVENUE_ID = '2', FAT_OCCUP_ID = '2', MOT_STATUS_ID = '1', MOT_REVENUE_ID = '2', MOT_OCCUP_ID = '2', PAR_STATUS_ID = '1', PAR_REVENUE_ID = '2', PAR_OCCUP_ID = '2', 1st_GPA = 'Poor', Conf='33.3', Supp='0.4', frequency='1', D_status = 'Y'	33.3%	0.4%
...

4.3 การใช้ PFSCARs

การนำ PFSCARs ที่ได้ มาประยุกต์ใช้ เพื่อพยากรณ์ลักษณะความเหมาะสมของนักศึกษาใหม่ โดยพิจารณาจากระดับผลการเรียน ทำได้โดยการนำ PFSCARs ที่ได้ มาคัดเลือกตามระดับค่าความสำคัญ (Precedence) จากมากไปน้อย โดยจะพิจารณาค่าความเชื่อมั่นที่ได้ และค่าสนับสนุนที่ทำการคำนวณไว้ของกฎที่พบก่อนหน้าเปรียบเทียบกับกฎถัดไป ดังแสดงกระบวนการพิจารณาในรูปแบบที่ 2

```

For (i=0; i< TotalRules; i++)
{
    Prec[i] = 0;
    If (Conf[i-1] < Conf[i]){ Prec[i]++; }
    Else{
        If Conf[i-1] = Conf[i]
        {
            If(Supp[i-1] < Supp[i]){ Prec[i]++; }
            Else{ Prec[i-1]++; }
        }
        Else{
            Prec[i-1]++;
        }
    }
}
Return Max (Prec);
    
```

รูปที่ 2. กระบวนการพิจารณาและกำหนดค่าความสำคัญของกฎจำแนกประเภทเชิงสัมพันธ์แบบเต็มรูป

จากภาพ ถ้าค่าความเชื่อมั่นที่พบในกฎก่อนหน้า มากกว่าค่าความเชื่อมั่นที่พบในกฎถัดไปให้กฎที่พบก่อนหน้า มีความสำคัญมากกว่ากฎถัดไป ถ้าค่าความเชื่อมั่นที่พบในกฎก่อนหน้า เท่ากับค่าความเชื่อมั่นที่พบในกฎลำดับถัดไป ให้พิจารณาค่าสนับสนุนของกฎทั้งสอง ถ้าค่าสนับสนุนของกฎที่พบก่อนหน้า มากกว่าค่าสนับสนุนของกฎถัดไป ให้กฎที่พบก่อนหน้า มีความสำคัญมากกว่ากฎถัดไป และถ้าค่าความเชื่อมั่นและค่าสนับสนุนของกฎที่พบก่อนหน้า และกฎถัดไปมีค่าเท่ากันให้กฎที่มาก่อนมีค่าความสำคัญมากกว่า

5. ผลการดำเนินงาน

จากการสร้างตัวแบบพยากรณ์คุณลักษณะความเหมาะสมของนักศึกษา สาขาเทคโนโลยีสารสนเทศ ด้วยเทคนิคกฎการจำแนกเชิงความสัมพันธ์แบบเต็มรูป ผลการดำเนินการพบว่ามีจำนวนกฎที่ถูกสร้างขึ้นทั้งสิ้น 266 กฎ ซึ่งจำแนกเป็นกฎสำหรับจำแนกระดับความเหมาะสมระดับดี 52 กฎ ระดับปานกลาง 110 กฎ และระดับอ่อน 104 กฎ โดยมีกฎที่เป็นไปตามค่าความเชื่อมั่น ขั้นต่ำที่กำหนด (45%) 157 กฎ ซึ่งจำแนกเป็นกฎสำหรับจำแนกระดับความเหมาะสมจากระดับผลการเรียน ในระดับดี ระดับปานกลาง และระดับอ่อน จำนวน 43, 61 และ 40 กฎ ตามลำดับ ดังแสดงในตารางที่ 4

ตารางที่ 4. กฎแบบเต็มรูปที่ได้ (จำแนกตามผลการเรียน)

Equalized level (1 st _GPA)	Training data set	PFSCARs	PFSCARs within MinConf (45%)
Poor (< 2.20)	146	52	43
Fair (2.20 -2.75)	416	110	61
Good (> 2.75)	106	104	40
Overalls	698	266	157

5.1 การทดสอบกฎจำแนกประเภทเชิงความสัมพันธ์

ในส่วนของการทดสอบข้อมูล ใช้ชุดข้อมูลซึ่งเตรียมไว้สำหรับการทดสอบ (Testing data set) จำนวน 335 ระเบียบนำมาผ่านกระบวนการของตัวแบบที่ถูกสร้างขึ้น โดยที่สมมติระเบียบข้อมูลทดสอบเป็นข้อมูลนักศึกษาเข้าใหม่ ที่ต้องการคาดการณ์ว่าจะมีระดับความเหมาะสมอยู่ในระดับใด ทำการทดสอบตามกระบวนการของตัวแบบ ที่แบ่งเป็น 2 ขั้นตอน ได้แก่

- นำเข้าข้อมูลนักศึกษาเข้าใหม่ โดยจัดรูปแบบข้อมูลให้พร้อมสำหรับการเปรียบเทียบกฎ
- คัดเลือกกฎที่สร้างขึ้น โดยเรียงลำดับตามค่าความสำคัญ และเนื่องจากได้มีการจัดเก็บกฎไว้ในฐานข้อมูล จึงทำให้การเรียงลำดับค่าความสำคัญของกฎสามารถทำได้ด้วยภาษา SQL

กฎที่ถูกคัดเลือกด้วยการเรียงลำดับค่าความสำคัญแล้ว คือผลการคาดการณ์ระดับความเหมาะสมของข้อมูลที่ได้ ทำการทดสอบจนครบทุกระเบียบข้อมูล เปรียบเทียบผลการคาดการณ์กับ ระดับผลการเรียนจริง จากนั้นจึงคำนวณหาความถูกต้องของการคาดการณ์ (ร้อยละ) ของระดับผลการเรียนที่แท้จริง ดังในตารางที่ 5

ตารางที่ 5. ผลการทดสอบกฎจำแนกตามประเภท

Equalized level (1 st _GPA)	Training data set	Testing data set	Accuracy	
			Record	%
Good (> 2.75)	146	54	34	63
Fair (2.20 -2.75)	416	207	184	89
Poor (< 2.20)	106	74	46	62
Overalls	698	335	264	79

เมื่อทำการทดสอบกฎจำแนกประเภทเชิงสัมพันธ์ที่สร้างขึ้น ด้วยชุดข้อมูลสำหรับทดสอบพบว่า กฎสำหรับจำแนกระดับความเหมาะสมของนักศึกษาในระดับปานกลาง เป็นกฎที่มีความสามารถในการจำแนกข้อมูลได้ดีที่สุด (ร้อยละ 89) ในขณะที่กฎจำแนกระดับความเหมาะสมระดับดี และระดับต่ำ มีความสามารถในการจำแนกได้ไม่ดีเท่าที่ควร ทั้งนี้เนื่องจากข้อมูลสำหรับเรียนรู้ของระดับดังกล่าวมีน้อย ทำให้การจำแนกกลุ่มไม่ดีเท่าที่ควร โดยเมื่อคำนวณประสิทธิภาพในการพยากรณ์โดยรวมแล้ว กฎที่ได้มีความสามารถจำแนกระดับความเหมาะสมของนักศึกษาอยู่ในระดับดี (ร้อยละ 79)

6. บทสรุป ข้อเสนอแนะ และแนวทางพัฒนาต่อ

งานวิจัยนี้เป็นการพัฒนาตัวแบบ เพื่อพยากรณ์ลักษณะความเหมาะสมต่อการเข้าศึกษา ในสาขาวิชาเทคโนโลยีสารสนเทศ ด้วยเทคนิคกฎการจำแนกประเภทเชิงความสัมพันธ์ ใช้ข้อมูลนักศึกษาสาขาวิชาเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏจันทรเกษม เป็นกรณีศึกษา โดยกำหนดกลุ่มจำแนกระดับความเหมาะสมของนักศึกษาใหม่ออกเป็น 3 กลุ่มตามระดับผลการเรียนอันได้แก่ ระดับดี ระดับปานกลาง และระดับต่ำ ผลการทดลองพบว่า สามารถจำแนกระดับความเหมาะสมของนักศึกษาได้ในระดับดี (ร้อยละ 79) แม้ว่าพบข้อจำกัดในเรื่องของการนำเข้าข้อมูลสำหรับเรียนรู้ ที่ระดับดีและระดับต่ำ น้อยเกินไป ทำให้ประสิทธิภาพการพยากรณ์ความเหมาะสม ของนักศึกษาในประเภทดังกล่าวไม่ดีเท่าที่ควร ซึ่งหากมีการนำเข้าข้อมูลสำหรับการเรียนรู้ ได้มากใกล้เคียงกับข้อมูลจริง ตัวแบบเพื่อพยากรณ์ลักษณะความเหมาะสมของนักศึกษาใหม่ จะสามารถใช้พยากรณ์คุณลักษณะความเหมาะสมของนักศึกษาได้ดียิ่งขึ้น

ทั้งนี้กฎการจำแนกประเภทเชิงความสัมพันธ์ ระหว่างปัจจัยที่ค้นพบ รวมถึงลักษณะความเหมาะสมที่ได้ สามารถนำไปประยุกต์ใช้กับข้อมูลนักศึกษาในสาขาอื่นๆ ได้ และเป็นแนวทางสำหรับการจัดการเรียนกลุ่มเรียน รวมถึงรายวิชาในแผนการเรียนในหลักสูตรที่เหมาะสมกับผู้เรียนต่อไป

7. เอกสารอ้างอิง

- [1] E. Tovar, J. Carrillo, and R. Colomo, "Proposal of an educational model for technical courses in the context of the European Convergence in Higher Education," presented at Frontiers in education conference - global engineering: knowledge without borders, opportunities without passports, 2007. FIE '07. 37th annual, 2007.
- [2] W. Green, "Demographic and psycholocial factors influencing academic success in a college level human anatomy course," in *The school of Human Resource Education and Workforce Development*: Louisiana State University., 2007.
- [3] J. C. S. Wise, D.; Colledge, T., "Student assessment of faculty performance: an alternate approach," *Frontiers in Education*, 2002. FIE 2002. 32nd Annual, vol. 1, pp. T3B-1 - T3B-4, 2002.
- [4] J. M. Brown, "Key factors that influence students' success in postsecondary vocational education programs " *Journal of Career Development*, vol. 13, 2005.

- [5] P. Jongmesuk, "Factors that influenced graduates of srinakharinwirot university, Bangsaen campus to choose self-employment."
- [6] R. Anthony, "The demographics of candidates for faculty positions in computer science," *Commun. ACM*, vol. 39, pp. 78-84, 1996.
- [7] จิราพร ยิ่งกว่าชาติ, รศ.ดร.ประสงค์ ปราณีตพลกรัง, "การทำนายผลสำเร็จการศึกษาของนักศึกษาในระดับอุดมศึกษาด้วยการเรียนรู้แบบเบย์และเหมืองข้อมูล," 2549.
- [8] W. H. Bing Liu., Yiming Ma., "Integrating Classification and Association Rule Mining," presented at Fourth International Conference on Knowledge Discovery and Data Mining(Kdd-98). , 1998.
- [9] ชิดชนก สงศิริ, กฤษณะ ไวยมัย และ ธนาวิทย์ รักธรรมานนท์, "การใช้เทคนิคดาต้าไมนิ่งเพื่อพัฒนาคุณภาพการศึกษาระดับอุดมศึกษา," *NECTEC Technical Journal*, vol. 3, pp. 134-142, 2001.
- [10] C. Ordonez and C. Ordonez, "Association rule discovery with the train and test approach for heart disease prediction" *Information Technology in Biomedicine, IEEE Transactions on*, vol. 10, pp. 334-343, 2006.
- [11] A. Michail and A. Michail, "Data mining library reuse patterns using generalized association rules" presented at Software Engineering, 2000. Proceedings of the 2000 International Conference on, 2000.
- [12] A. Srisawat and B. Kijirikul, "Using associative classification for predicting HIV-1 drug resistance," presented at Hybrid Intelligent Systems, 2004. HIS '04. Fourth International Conference on, 2004.
- [13] M. J. A. B. a. G. Linoff, *Data Mining Techniques*: John Wiley & Sons., 1997.
- [14] K. H. Shimada, K. Jinglu Hu, "Class Association Rule Mining with Chi-Squared Test Using Genetic Network Programming," presented at International Conference on Systems, Man and Cybernetics, 2006. ICSMC '06., 2006.