

บทที่ 3

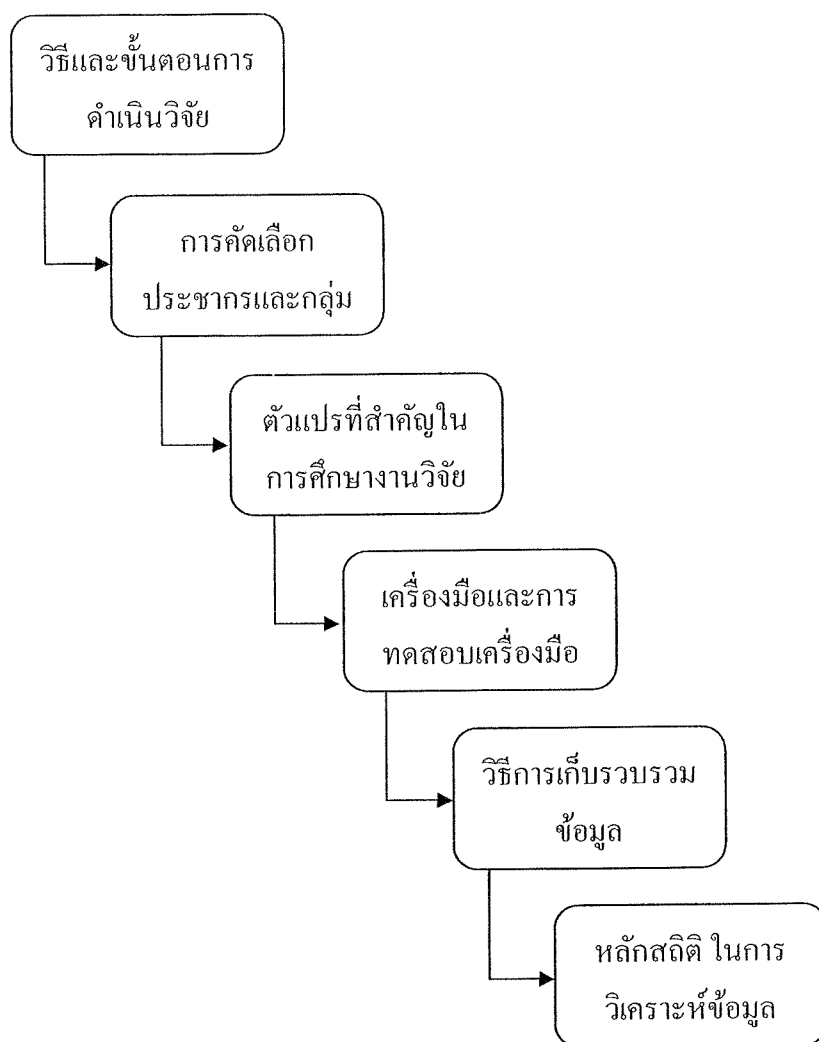
ระเบียบวิธีการวิจัย

การดำเนินการศึกษาวิจัยเรื่องการสร้างโมเดลการจัดการระบบนักศึกษาสัมพันธ์ โดยใช้เทคนิคเหมืองข้อมูล (The Development of Student Relationship Management Model using Data Mining Technique) มีวัตถุประสงค์ เพื่อศึกษาข้อมูลเกี่ยวกับหลักสูตร อาจารย์ การบริการ ต่างๆ ของมหาวิทยาลัยต่างๆ ในการศึกษาวิจัยครั้งนี้ มีระเบียบวิธีการวิจัย ดังนี้

- 3.1 วิธีการและขั้นตอนการดำเนินการวิจัย
- 3.2 การคัดเลือกประชากรและกลุ่มตัวอย่าง
- 3.3 ตัวแปรที่สำคัญในการศึกษางานวิจัย
- 3.4 เครื่องมือและการทดสอบเครื่องมือที่นำมาใช้ในการวิจัย
- 3.5 วิธีการเก็บรวบรวมข้อมูลสำหรับทำงานวิจัย
- 3.6 หลักสถิติที่ใช้ในการวิเคราะห์ข้อมูล
- 3.7 วิธีการวิจัยโดยใช้เทคนิคเหมืองข้อมูล

3.1 วิธีการและขั้นตอนการดำเนินการวิจัย

ในเนื้อหาของงานที่ทางผู้วิจัยได้ดำเนินการศึกษา เป็นรูปแบบของงานวิจัยเชิงเหตุผล (Casual Research) ซึ่งเป็นประเภทของงานวิจัยที่ระบุความสัมพันธ์ระหว่างตัวแปรเหตุและผล (Cause and Effect Relationship) โดยมีการเก็บข้อมูลจากกลุ่มตัวอย่างจากประชากร จากแบบสอบถาม ผ่านวิธีและขั้นตอน จากนั้นใช้หลักทางสถิติและพยากรณ์วิเคราะห์ข้อมูลจากงานวิจัย โดยการทำวิจัยผู้ทำการวิจัยได้วางขั้นตอนและรูปแบบการทำงาน ดังต่อไปนี้



รูปที่ 3.1 แสดงขั้นตอนดำเนินการวิจัย

3.2 การคัดเลือกประชากรและกลุ่มตัวอย่าง

ประชากร (Population) ที่ทางผู้วิจัยกำหนดเป็นกลุ่มที่ใช้ในการทำการวิจัยในครั้งนี้ เป็นนักศึกษาจากสถาบันการศึกษาที่เปิดสอนในระดับปริญญาตรี จำนวนทั้งสิ้น 2,731 คนจากสถาบันการศึกษาทั้ง 7 แห่ง การสุ่มตัวอย่าง (Sampling) สำหรับการเลือกตัวอย่างจากแต่ละกลุ่ม ผู้วิจัยใช้วิธีการตัดสินใจ ให้น่วยตัวอย่างเป็นไปตามวัตถุประสงค์ของงานวิจัย จึงเลือกวิธีเลือกสุ่มตัวอย่างแบบบังเอิญ โดยได้มีการแยกกลุ่มประชากรแยกตามสถาบันการศึกษาดังนี้

1. มหาวิทยาลัยบูรพา	จำนวน 397 คน
2. มหาวิทยาลัยกรุงเทพ	จำนวน 394 คน
3. มหาวิทยาลัยหอการค้าไทย	จำนวน 392 คน
4. มหาวิทยาลัยราชภัฏราชนครินทร์	จำนวน 374 คน
5. มหาวิทยาลัยเกษตรศาสตร์	จำนวน 397 คน
6. มหาวิทยาลัยศรีปทุม	จำนวน 393 คน
7. มหาวิทยาลัยหัวเฉียวเฉลิมพระเกียรติ	จำนวน 384 คน

3.3 ตัวแปรที่ใช้ในการวิจัย

สำหรับตัวแปรที่ทางผู้วิจัยสนใจ และสัมพันธ์กับลักษณะของประชากรมีดังต่อไปนี้

1. เจตคติต่อพฤติกรรมในความต้องการต่อระบบนักศึกษาสัมพันธ์
2. ความคล้ายคลึงต่อสิ่งอ้างอิง
3. ความเหมาะสมของหลักสูตร
4. คุณสมบัติที่จำเป็นของอาจารย์
5. สิ่งอำนวยความสะดวกในระบบการบริการ

3.4 เครื่องมือและการทดสอบเครื่องมือที่นำมาใช้ในการวิจัย

สำหรับการสร้างเครื่องมือที่ใช้ในการเก็บรวบรวมข้อมูลของการศึกษาวิจัย มีวัตถุประสงค์เพื่อทดสอบและวิเคราะห์ระหว่างตัวแปรต้นและตัวแปรตาม โดยเก็บข้อมูลของนักศึกษาจากมหาวิทยาลัยรัฐบาลและเอกชนต่างๆ ซึ่งข้อมูลระดับปฐมภูมินั้น ได้สร้างมาจากระบบสอบถาม โดยแบ่งเนื้อหาของแบบสอบถามเป็น 3 ส่วนดังนี้ คือ

ตอนที่ 1 เป็นแบบสอบถามเกี่ยวกับข้อมูลทั่วไปของประชากรที่ถูกสุ่มตัวอย่าง ในเรื่องทั่วไปเกี่ยวกับภูมิปัญญา สถานภาพ การศึกษาก่อนเข้าสู่ระบบมหาวิทยาลัย ความคาดหวังที่ต้องการได้รับการศึกษา ว่ามีรูปแบบเป็นอย่างไร เพื่อวัดตัวแปรเกี่ยวกับคุณลักษณะของนักศึกษา โดยใช้วิธี check list

ตอนที่ 2 เป็นแบบสอบถามเกี่ยวกับตัวแปรที่ใช้ในการศึกษา ทั้งหมด 3 ตัวแปร ซึ่งประกอบด้วย

2.1 ด้านเนื้อหาของหลักสูตร โดยใช้วิธีการวัดแบบ Rating Scale ตามหลักการของ Likert Scale แบบ 5 ระดับ โดยมีจำนวนคำถามทั้งหมด 12 ข้อ

2.2 ด้านคุณสมบัติของอาจารย์ โดยใช้วิธีการวัดแบบ Rating Scale ตามหลักการ

ของ Likert Scale แบบ 5 ระดับ โดยมีจำนวนคำถามทั้งหมด 11 ข้อ

2.3 ด้านสิ่งอำนวยความสะดวกต่างๆ เช่น วัสดุการศึกษา ตำรา อาคาร สถานที่และระบบการให้บริการ โดยใช้วิธีการวัดแบบ Rating Scale ตามหลักการของ Likert Scale แบบ 5 ระดับ โดยมีจำนวนคำถามทั้งหมด 12 ข้อ

ตอนที่ 3 เป็นการให้ผู้ตอบแบบสอบถามแสดงความคิดเห็นหรือข้อเสนอแนะคะแนนเฉลี่ยกับการแปลความหมายเป็นดังตารางที่ 3.1

ตารางที่ 3.1 แสดงระดับความคิดเห็น คะแนนเฉลี่ยและระดับการแปลความหมาย

ระดับความคิดเห็น	ค่าคะแนน	ค่าเฉลี่ยคะแนน	ระดับการแปลความหมาย
มากที่สุด	5	4.21-5.00	ระดับมากที่สุด
มาก	4	3.41-4.20	ระดับมาก
ปานกลาง	3	2.61-3.40	ระดับปานกลาง
น้อย	2	1.81-2.60	ระดับน้อย
น้อยที่สุด	1	1.00-1.80	ระดับน้อยที่สุด

เพื่อพิสูจน์ว่าเครื่องมือที่ใช้ในการทำวิจัย มีประสิทธิภาพในการนำไปใช้งาน จึงจำเป็นต้องมีการทดสอบเครื่องมือก่อนการนำไปใช้งาน ซึ่งทางผู้วิจัยได้กำหนดมาตรวัดด้วย 2 กระบวนการดังต่อไปนี้

1. การหาค่าความเที่ยงตรงเชิงเนื้อหา (Content Validity) ด้วยการให้ผู้เชี่ยวชาญพิจารณา ด้วยการหาค่าดัชนีความสอดคล้องของเครื่องมือระหว่างข้อคำถามกับวัตถุประสงค์ (Index of Item Objective Congruence: IOC) จากแบบสอบถามการสร้างโมเดลระบบนักศึกษาสัมพันธ์ โดยการหาค่าดัชนีความสอดคล้องของแบบจำลองดัชนีวัดผลความพอใจในระบบนักศึกษาสัมพันธ์ตามวิธีของโรวินลลิและแฮมบิลตัน โดยแบ่งคะแนนออกเป็น 3 ระดับดังนี้

แน่ใจว่ามีความสอดคล้องหรือวัดได้	มีระดับคะแนนเท่ากับ	1
ไม่แน่ใจว่ามีความสอดคล้องหรือวัดได้	มีระดับคะแนนเท่ากับ	0
แน่ใจว่าไม่มีความสอดคล้องหรือวัดได้	มีระดับคะแนนเท่ากับ	-1

หลังจากนั้นนำแบบสอบถามการสร้างโมเดลระบบนักศึกษาสัมพันธ์ ส่งให้ผู้ทรงคุณวุฒิที่มีความเชี่ยวชาญทางด้านการตรวจสอบแบบสอบถาม และนำมาหาค่าความสอดคล้องที่ได้รับมาใช้งาน โดยใช้สูตร

$$IOC = \frac{\sum R}{N}$$

เมื่อ

IOC หมายถึง คำนีความสอดคล้องมีค่าระหว่าง -1 ถึง 1

$\sum R$ หมายถึง ผลรวมของการพิจารณาของผู้เชี่ยวชาญ

N หมายถึงจำนวนผู้เชี่ยวชาญ

โดยนำผลลัพธ์ที่ได้จากสมการที่ 3.1 มาเปรียบเทียบกับเกณฑ์มาตรฐานความเที่ยงตรงของเคลิเคนดอล, กรูเบอร์ และจอร์น สัน ซึ่งได้เสนอมาตรการการประเมินผลดัชนีความสอดคล้องของแบบสอบถามการสร้างโมเดลระบบนักศึกษาสัมพันธ์ดังนี้

ค่าเฉลี่ย 0.00-0.49	ความสอดคล้องของแบบสอบถามอยู่ในเกณฑ์ต่ำ
ค่าเฉลี่ย 0.50-0.69	ความสอดคล้องของแบบสอบถามอยู่ในเกณฑ์ยอมรับ
ค่าเฉลี่ย 0.70-0.79	ความสอดคล้องของแบบสอบถามอยู่ในเกณฑ์ดี
ค่าเฉลี่ย 0.80-1.00	ความสอดคล้องของแบบสอบถามอยู่ในเกณฑ์ดีมาก

2. การหาค่าความเชื่อมั่นของเครื่องมือและข้อมูลที่ใช้ในการวิจัย ด้วยวิธีหาค่าสัมประสิทธิ์แอลฟาของครอนบาค (Cronbach alpha coefficient) ซึ่งเป็นวิธีการที่ถูกนำไปใช้กับเครื่องมือในทุกๆ ข้อคำถามที่มีคำตอบไม่ใช่แค่ 0 และ 1 เท่านั้น แต่เป็นคำตอบประเภทเรียงลำดับโดยใช้สูตร

$$\alpha = \left(\frac{k}{k-1} \right) \left(1 - \frac{\sum s_i^2}{s_t^2} \right)$$

เมื่อ

α หมายถึง ค่าความเชื่อมั่นของเครื่องมือ

k หมายถึง จำนวนคำถาม

s_i^2 หมายถึง ความแปรปรวนของคะแนนแต่ละข้อ

s_t^2 หมายถึง ความแปรปรวนของคะแนนทั้งฉบับ

การคำนวณทุกๆ ข้อคำถาม จะมีค่าความเชื่อมั่นใกล้เคียงกัน และค่าความเชื่อมั่นที่ได้จะขึ้นอยู่กับค่าความสัมพันธ์ระหว่างข้อคำถามและจำนวนคำถาม หากมีความสัมพันธ์กันสูงหรือจำนวนข้อคำถามมีมาก ค่าของความเชื่อมั่นก็จะมีค่าสูง หากค่าความเชื่อมั่นของแบบสอบถามต่ำควรที่จะเพิ่มคำถามให้มากขึ้น แต่จะต้องมีความสัมพันธ์ในทางเดียวกันกับ

ข้อคำถามเดิม แต่ข้อคำถามที่เพิ่มขึ้นมีความสัมพันธ์ในทางตรงกันข้ามกับข้อคำถามเดิม จะทำให้ค่าความเชื่อมั่นลดลงได้ ค่าเชื่อมั่นควรจะไม่น้อยกว่า 0.70

ทางผู้วิจัยได้นำแบบสอบถามที่ได้ปรับปรุงแล้วไปทำการทดสอบ (Pilot Test) เพื่อแสดงค่าความเชื่อมั่น กับกลุ่มตัวอย่างที่ใช้กลุ่มเป้าหมาย จำนวน 40 คน แล้วนำข้อมูลมาวิเคราะห์หาค่าความเชื่อมั่นด้วยวิธีครอนบาค อัลฟา ดังแสดงในตารางที่ 3.2

ตารางที่ 3.2 แสดงค่าความเชื่อมั่นด้วยวิธีของ Cronbach's Alpha

ตัวแปร	จำนวนคำถาม	Cronbach's Alpha
ด้านความเหมาะสมของหลักสูตร	12 ข้อ	0.88
ด้านคุณสมบัติของอาจารย์	11 ข้อ	0.92
ด้านสิ่งอำนวยความสะดวกของระบบการให้บริการ	12 ข้อ	0.82

ซึ่งผลลัพธ์ที่ได้มีค่า 0.88 แสดงให้เห็นว่า ข้อคำถามที่เป็นตัววัดตัวแปรแต่ละตัวมีความเหมาะสม และเชื่อถือได้ ทำให้เชื่อมั่นได้ว่าสามารถนำเครื่องมือนี้ไปใช้เพื่อทำการวิจัยกับกลุ่มตัวอย่างได้ ซึ่งจะได้นำไปเก็บข้อมูลจากกลุ่มตัวอย่างจำนวน 2,731 ชุด

3.5 วิธีการเก็บรวบรวมข้อมูลสำหรับทำงานวิจัย

จากข้อมูลข้างต้นการทำงานวิจัยครั้งนี้ ซึ่งได้กำหนดประชากร กลุ่มตัวอย่าง ซึ่งก็คือนักศึกษาระดับปริญญาตรีที่กำลังศึกษาอยู่ในสถาบันการศึกษาในมหาวิทยาลัยรัฐบาลและเอกชน จำนวน 7 แห่ง โดยการใช้เครื่องมือคือแบบสอบถาม ที่ถูกสร้างขึ้นมาตามหลักการและผ่านขั้นตอนการทดสอบจนแน่ใจว่าเป็นเครื่องมือที่มีประสิทธิภาพในการทำงานวิจัย วิธีการรวบรวมข้อมูลโดยการแจกแบบสอบถามไปยังมหาวิทยาลัยทั้งรัฐและเอกชนทั้งหมด จำนวน 5000 ชุด และได้แบบสอบถามส่งกลับมาจำนวน 2800 ชุด จากนั้นจึงนำข้อมูลที่ได้จากการเก็บรวบรวมข้อมูล นำไปใช้ในขั้นตอนการวิเคราะห์ต่อไป

3.6 หลักสถิติที่ใช้ในการวิเคราะห์ข้อมูล

หลังจากที่ได้เก็บข้อมูลจากกลุ่มตัวอย่างเรียบร้อยแล้ว ขั้นตอนต่อไปซึ่งถือเป็นขั้นตอนที่สำคัญ ก็คือการนำข้อมูลที่ได้ไปใช้วิเคราะห์ตามหลักสถิติ เพื่อให้ได้ข้อมูลสรุปสำคัญเพื่อนำไปใช้งานได้ในโอกาสต่อไป ทางผู้วิจัยได้มีการใช้ระบบคอมพิวเตอร์มาช่วยในการวิเคราะห์

โดยนำเอาโปรแกรมสำเร็จรูปที่มีความสามารถในการวิเคราะห์ข้อมูลตามหลักการทางสถิติต่างๆ โดยมีหลักการของการวิเคราะห์ดังต่อไปนี้

1. ค่าร้อยละ (Percentage) นำมาใช้ในการอธิบายข้อมูลทั่วไปสำหรับกลุ่มตัวอย่างจากประชากรผู้ตอบแบบสอบถาม
2. ค่าเฉลี่ย ($\bar{X}$) และส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation :SD) นำมาใช้ในการวิเคราะห์และอธิบายค่า และความหมายตัวแปรแต่ละตัว

3.7 วิธีการวิจัยโดยใช้เทคนิคเหมืองข้อมูล

ในงานวิจัยนี้ได้ใช้หลักการ CRIPS-DM ซึ่งเป็นกระบวนการมาตรฐานในการทำงานทางด้านเหมืองข้อมูล ประกอบด้วย 6 ขั้นตอนดังรายละเอียดต่อไปนี้

1. Business Understanding

วัตถุประสงค์ของงานวิจัยนี้ คือ เพื่อศึกษาปัจจัยที่มีผลต่อความพึงพอใจของนักศึกษาในด้านต่างๆ

2. Data Understanding

ข้อมูลที่ใช้ในงานวิจัยนี้ได้จากการจัดทำแบบสอบถาม จำนวน 2,731 โดยลักษณะแบบสอบถามเป็นดังที่ได้กล่าวมาแล้วในข้างต้น

3. Data Preparation

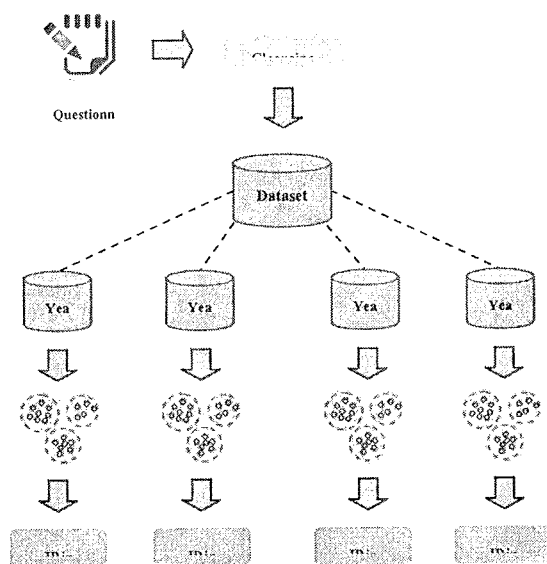
ข้อมูลที่ได้จากการตอบแบบสอบถามของนักศึกษาจำนวนทั้งหมด 2,731 คน จะถูกนำมากรองข้อมูลที่ผิดพลาดขาดหายไป (Missing Value) ก่อน โดยผู้วิจัยได้ทำการคัดกรองเฉพาะแบบสอบถามที่นักศึกษามีการทำแบบสอบถามครบทุกหัวข้อสอบถาม พบว่ามีข้อมูลจำนวน 283 แบบสอบถามที่นักศึกษาดตอบแบบสอบถามไม่ครบทุกหัวข้อ ดังนั้นจึงเหลือข้อมูลจำนวน 2,517 แบบสอบถามที่จะถูกใช้ในงานวิจัยต่อไป โดยแบ่งจำนวนข้อมูลตามชั้นปีของนักศึกษาได้ดังตารางที่ 3.3

ตารางที่ 3.3 แสดงการจำนวนนักศึกษาที่แบ่งกลุ่มในแต่ละชั้นปี

ชั้นปี	ชั้นปีที่ 1	ชั้นปีที่ 2	ชั้นปีที่ 3	ชั้นปีที่ 4
จำนวนข้อมูล	974	637	567	339

4. Modeling

ในงานวิจัยนี้มุ่งเน้นวิเคราะห์ผลสำรวจความพึงพอใจของนักศึกษาในหัวข้อต่างๆ เพื่อใช้ในการปรับปรุงคุณภาพการเรียนการสอน อุปกรณ์หรือสื่อการสอน รวมทั้งเครื่องอำนวยความสะดวกเพื่อส่งเสริมการศึกษามีคุณภาพมากยิ่งขึ้น โดยแยกพิจารณาตามชั้นปีของนักศึกษา จากนั้นจึงใช้เทคนิคการจัดกลุ่มข้อมูล (Clustering) เพื่อช่วยในการหาความสัมพันธ์ของกลุ่มนักศึกษาที่มีความพึงพอใจในแต่ละหัวข้อคล้ายๆ กัน จากนั้นผู้วิจัยจึงทำการวิเคราะห์ลงรายละเอียดว่ากลุ่มนักศึกษาลักษณะใดที่ควรให้ความสำคัญ เพื่อทำการปรับปรุงคุณภาพตามหัวข้อที่นักศึกษาในกลุ่มนั้นต้องการได้อย่างถูกต้อง นอกจากนี้ผู้วิจัยยังได้ประยุกต์ใช้เทคนิคการจำแนกประเภท JRip เพื่อช่วยหาความสัมพันธ์ของแต่ละหัวข้อแบบสอบถามในแต่ละกลุ่มนักศึกษา โดยผลลัพธ์ที่ได้จากการใช้เทคนิค JRip นั้น สามารถช่วยวิเคราะห์การปรับปรุงคุณภาพเฉพาะด้านของหัวข้อแบบสอบถามที่อาจมีความสัมพันธ์และส่งผลกระทบต่อกันได้ สำหรับกระบวนการวิจัยแสดงดังรูปที่ 3.2



รูปที่ 3.2 แสดงขั้นตอนวิจัยโดยใช้เทคนิคเหมืองข้อมูล

ในขั้นตอนการจัดกลุ่มข้อมูลนั้น ผู้วิจัยได้ทำการจัดกลุ่มข้อมูล โดยกำหนดค่า $K=3$ ถึง $K=6$ จากนั้นจึงทำการเลือกจำนวนกลุ่มข้อมูลที่เหมาะสม โดยผู้วิจัยพิจารณาจากค่าทางสถิติต่างๆ ได้แก่ ค่า Sum of Square Error (SSE) ค่า Between/Within และค่า T-Test (ดูวิธีการคำนวณจากหัวข้อที่ 5)

5. Evaluation

ในส่วนนี้จะกล่าวถึงการวัดคุณภาพของข้อมูลที่ได้ทำการจัดกลุ่ม โดยพิจารณาจากค่า Sum of Square Error (SSE) และค่า Between/Within โดยมีสูตรการคำนวณดังนี้

5.1 Sum of Squared Errors (SSE)

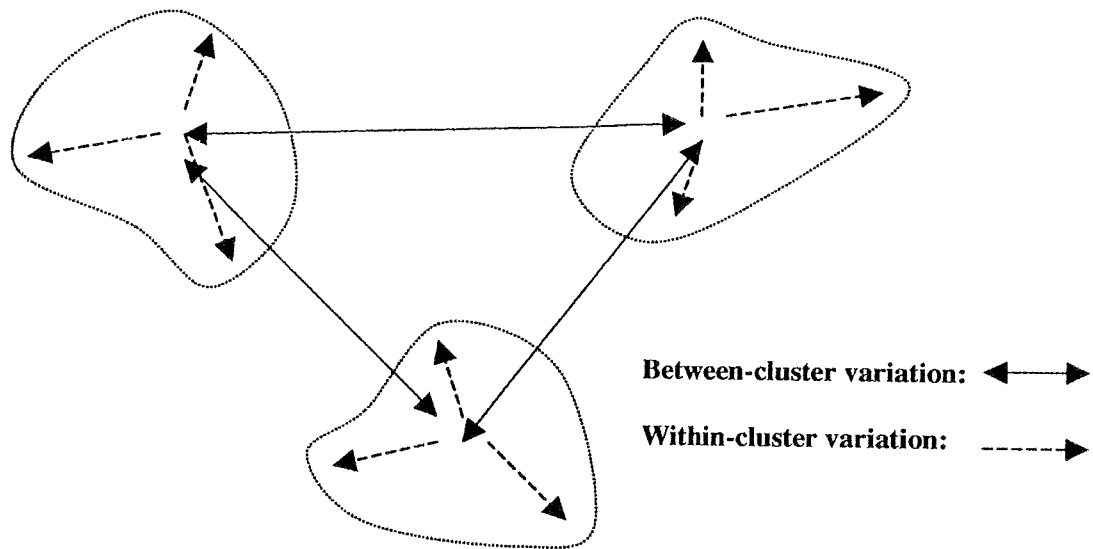
วิธีทั่วไปที่ใช้ในการวัดคุณภาพของการจัดกลุ่ม คือการหาค่าผลรวมความผิดพลาด (Error) ระหว่างจุดข้อมูลใดๆ ในกลุ่มกับจุดข้อมูลศูนย์กลางประจำกลุ่ม (Centroid of Cluster) ซึ่งค่าดังกล่าวแสดงถึงการกระจายตัวของข้อมูล โดยหากค่า Error มีค่าสูง แสดงว่าข้อมูลภายในกลุ่มมีการกระจายตัวมาก ในทางตรงกันข้ามหากพบว่า Error มีค่าต่ำ แสดงว่าข้อมูลที่ถูกจัดกลุ่มนั้น เกาะกลุ่มกันมาก หรือกล่าวอีกนัยหนึ่งคือมีความคล้ายกันหรือเหมือนกันมาก สำหรับฟังก์ชันการคำนวณค่า SSE แสดงดังสมการ

$$SSE = \sum_{i=1}^k \sum_{p \in C_i} d(p, m_i)^2$$

โดยที่ $d(p, m_i)$ คือฟังก์ชันคำนวณระยะห่างของข้อมูล
 p คือ จุดข้อมูลใดๆ ในกลุ่ม C_i
 m_i คือ จุดข้อมูลศูนย์กลางประจำกลุ่ม (Centroid of Cluster)

5.2 Between-cluster variation and Within-cluster variation

เนื่องจากการจัดกลุ่มข้อมูลนั้น ต้องการให้ข้อมูลที่ถูกจัดอยู่ในกลุ่มเดียวกันมีลักษณะคล้ายกันหรือมีการกระจายตัวน้อย และข้อมูลที่อยู่ต่างกลุ่มกันมีความต่างของข้อมูลมาก ดังนั้นวิธีหนึ่งที่นิยมใช้ในการวัดคุณภาพคลัสเตอร์คือ การพิจารณาสัดส่วนระหว่างค่าความแตกต่างระหว่างกลุ่ม (BCV: Between-cluster variation) และค่าความแตกต่างภายในกลุ่ม (WCV: Within-cluster variation) โดยการจัดกลุ่มที่ดีควรมีค่าสัดส่วนดังกล่าวมาก ดังรูปที่ 3.3



รูปที่ 3.3 ค่าความแตกต่างระหว่างกลุ่ม (BCV) ควรมีค่ามาก
และค่าความแตกต่างภายในกลุ่ม (WCV) ควรมีค่าน้อย

สำหรับการการคำนวณค่าสัดส่วนระหว่าง BCV และ WCV แสดงดังนี้

$$\frac{BCV}{WCV} = \frac{d(m_1, m_2)}{SSE}$$

โดยที่ $d(m_1, m_2)$ คือ ฟังก์ชันคำนวณระยะห่างของข้อมูล

m_1 และ m_2 คือ จุดข้อมูลศูนย์กลางประจำกลุ่ม (Centroid of Cluster) ของคลัสเตอร์ที่พิจารณา

5.3 T-Test

ในการเลือกจำนวนกลุ่มข้อมูลสำหรับเทคนิค K-Means นั้น ในงานวิจัยนี้ได้ทำการปรับเปลี่ยนค่า k เพื่อทำการพิจารณาจำนวนกลุ่มที่เหมาะสม โดยใช้วิธีการสถิติช่วยในการพิจารณาร่วมกับวิธีการในหัวข้อ 5.1-5.2 โดยใช้การพิจารณาค่า T-Test เพื่อวิเคราะห์ความแตกต่างของข้อมูลในแต่ละกลุ่มว่ามีความแตกต่างอย่างมีนัยสำคัญทางหลักสถิติหรือไม่ การคำนวณ T-Test แสดงดังสมการ

TTEST(array1,array2,tails,type)

โดยที่ Array1 คือ data set ของข้อมูลชุดที่ 1

Array2 คือ data set ของข้อมูลชุดที่ 2

Tails คือ การระบุจำนวนการกระจายของ tails

ถ้า tails = 1, TTEST uses the one-tailed distribution.

tails = 2, TTEST uses the two-tailed distribution.)

Type คือ ประเภทของ t-Test ที่จะแสดง