



ใบรับรองวิทยานิพนธ์

บัณฑิตวิทยาลัย มหาวิทยาลัยเกษตรศาสตร์

วิทยาศาสตร์มหาบัณฑิต (สถิติ)

ปริญญา

สถิติ

สถิติ

สาขา

ภาควิชา

เรื่อง การเปรียบเทียบเทคนิคการแบ่งกลุ่มข้อมูลสำหรับการวิเคราะห์กลุ่ม

The Comparison of Clustering Techniques for Cluster Analysis

นามผู้วิจัย นางสาวปิยธิดา รุจะศิริ

ได้พิจารณาเห็นชอบโดย

อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

(ผู้ช่วยศาสตราจารย์บุญอ้อม โฉมทิ, Ph.D.)

อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม

(อาจารย์อำไพ ทองธีรภาพ, Ph.D.)

หัวหน้าภาควิชา

(อาจารย์อำไพ ทองธีรภาพ, Ph.D.)

บัณฑิตวิทยาลัย มหาวิทยาลัยเกษตรศาสตร์รับรองแล้ว

(รองศาสตราจารย์กัญจนา ชีระกุล, D.Agr.)

คณบดีบัณฑิตวิทยาลัย

วันที่ เดือน พ.ศ.

วิทยานิพนธ์

เรื่อง

การเปรียบเทียบเทคนิคการแบ่งกลุ่มข้อมูลสำหรับการวิเคราะห์กลุ่ม

The Comparison of Clustering Techniques for Cluster Analysis

โดย

นางสาวปิยธิดา รุจะศิริ

เสนอ

บัณฑิตวิทยาลัย มหาวิทยาลัยเกษตรศาสตร์
เพื่อความสมบูรณ์แห่งปริญญาวิทยาศาสตรมหาบัณฑิต (สถิติ)

พ.ศ. 2552

ปิยธิดา รุจะศิริ 2552: การเปรียบเทียบเทคนิคการแบ่งกลุ่มข้อมูลสำหรับการวิเคราะห์
กลุ่ม ปริญญาวิทยาศาสตรมหาบัณฑิต (สถิติ) สาขาสถิติ ภาควิชาสถิติ อาจารย์ที่ปรึกษา
วิทยานิพนธ์หลัก: ผู้ช่วยศาสตราจารย์บุญอ้อม โฉมทิ, Ph.D. 185 หน้า

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาและเปรียบเทียบความสามารถในการแบ่งกลุ่มของ
เทคนิคการแบ่งกลุ่ม 5 วิธี ได้แก่ วิธีการจัดกลุ่มแบบลำดับขั้น อัลกอริทึมจัดกลุ่มแบบเคมีน
วิธีจัดระบบด้วยตัวเองของโคโฮเนน อัลกอริทึมจัดกลุ่มแบบเคมคอยด์ และการใช้อัลกอริทึม
จัดกลุ่มแบบเคมคอยด์ร่วมกับวิธีวัดระยะห่างแบบไดนามิกไทม์วอร์ปปีง โดยทำการศึกษากับ
ข้อมูลจริงและข้อมูลจำลอง ซึ่งมีการแจกแจงแบบปกติหลายตัวแปร ในการจำลองข้อมูลใช้เทคนิค
มอนติคาร์โลสร้างข้อมูลจำนวน 1,000 ชุด สำหรับแต่ละชุดข้อมูลซึ่งมีขนาด 25, 100 และ 300
ตัวอย่าง และกำหนดจำนวนกลุ่มข้อมูลขนาด 2-5, 2-8 และ 2-15 กลุ่มตามลำดับ กำหนดจำนวน
ตัวแปรที่ศึกษาเป็น 3, 5 และ 7 ตัวแปร โดยพิจารณาจากเกณฑ์ค่าความแตกต่างของข้อมูลภายใน
กลุ่ม (RMSSTD) ซึ่งถ้าค่า RMSSTD ต่ำ แสดงว่าเทคนิคการแบ่งกลุ่มนั้นเป็นวิธีที่ดี และพิจารณา
จากเกณฑ์ค่าความแตกต่างของข้อมูลระหว่างกลุ่ม (RS) ซึ่งถ้าค่า RS สูง แสดงว่าเทคนิคการ
แบ่งกลุ่มนั้นเป็นวิธีที่ดี ผลการวิจัยพบว่า ทั้งการศึกษการแบ่งกลุ่มจากข้อมูลจำลองและจาก
ข้อมูลจริงให้ผลสอดคล้องกัน กล่าวคือ อัลกอริทึมจัดกลุ่มแบบเคมีนให้ค่า RMSSTD ต่ำที่สุดและ
ค่า RS สูงที่สุดในเกือบทุกกรณี ซึ่งสรุปได้ว่า อัลกอริทึมจัดกลุ่มแบบเคมีนให้ผลการวิเคราะห์
กลุ่มที่ดีที่สุด ยกเว้นกรณีที่แบ่งกลุ่มข้อมูลขนาด 25 ตัวอย่างเป็น 4 หรือ 5 กลุ่ม ซึ่งวิธีการจัดกลุ่ม
แบบลำดับขั้นด้วยวิธี Ward linkage ให้ผลการวิเคราะห์กลุ่มที่ดีที่สุด นอกจากนี้เมื่อจำนวนกลุ่ม
เพิ่มขึ้นประสิทธิภาพในการแบ่งกลุ่มของทั้ง 5 เทคนิควิธี มีแนวโน้มเพิ่มขึ้น แต่จำนวนตัวแปร
และขนาดของชุดข้อมูลที่เพิ่มขึ้นไม่มีผลต่อประสิทธิภาพในการแบ่งกลุ่มของเทคนิคการแบ่งกลุ่ม
ทั้ง 5 วิธี

Piyatida Rujasiri 2009: The Comparison of Clustering Techniques for Cluster Analysis. Master of Science (Statistics), Major Field: Statistics, Department of Statistics. Thesis Advisor: Assistant Professor Boonorm Chomtee, Ph.D. 185 pages.

The objective of this research is to compare effectiveness of the 5 clustering techniques for multivariate data; these techniques consist of Hierarchical clustering method, K-means clustering algorithm, Kohonen's Self-Organizing Maps method (SOM), K-medoids method and K-medoids method integrated with Dynamic Time Warping distance measure (DTW). To evaluate these 5 techniques, Root Mean Square Standard Deviation (RMSSTD) and R Squared (RS) are used to be the criteria. For RMSSTD, the lower value is the better technique and for RS, the higher value is the better technique. These approaches are evaluated by using both real and simulated data which are multivariate normally distributed. Each datasets were generated by Monte Carlo technique with 25, 100 and 300 sample sizes and repeated 1,000 times for 3, 5 and 7 variables. In this research, 2-15 clusters are studied. For the results, both real and simulated datasets provide the same result. That is, K-means clustering method yields the lowest RMSSTD and highest RS for almost situations. Therefore, K-means is the most suitable algorithm of clustering techniques for cluster analysis, except categorizing 25-sample dataset to 4 or 5 clusters which Ward linkage gives the best clustering result. In addition, increasing the number of clusters trends to increase the efficiency of the 5 clustering methods. However, the number of variables and the size of dataset do not affect to the efficiency of the 5 clustering methods.

Student's signature

Thesis Advisor's signature

/ /

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้จะไม่สำเร็จลุล่วงไปได้ด้วยดี หากขาดความช่วยเหลือ และคำปรึกษาตลอดระยะเวลาการวิจัยของผู้ช่วยศาสตราจารย์ ดร.บุญอ้อม โฉมทิ อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก และ ดร.อำไพ ทองธีรภาพ อาจารย์ที่ปรึกษาวิทยานิพนธ์รอง ผู้วิจัยจึงขอกราบขอบพระคุณเป็นอย่างสูงไว้ ณ โอกาสนี้ ขอบพระคุณผู้ช่วยศาสตราจารย์ ดร.กมลชนก พานิชการ ผู้ทรงคุณวุฒิ และ รองศาสตราจารย์สายสุดา สมชิต ประธานการสอบวิทยานิพนธ์ ที่กรุณาให้คำแนะนำเพิ่มเติมเพื่อความสมบูรณ์ยิ่งขึ้นของวิทยานิพนธ์เล่มนี้ และขอบพระคุณคณาจารย์ภาควิชาสถิติทุกท่าน ที่ช่วยประสิทธิ์ประสาทวิชาความรู้แก่ผู้วิจัยเป็นอย่างดี

ขอบคุณนายเดชาวุฒิ วานิชสรรพ ที่ให้คำแนะนำในการเขียนและใช้โปรแกรม Matlab รวมถึงให้คำปรึกษาเกี่ยวกับงานวิจัย ทำให้งานวิจัยชิ้นนี้สำเร็จลุล่วงไปได้ด้วยดี

ขอกราบขอบพระคุณ คุณพ่อคุณแม่ ที่ให้การสนับสนุนทุนทรัพย์ในการศึกษาเล่าเรียนดูแลเอาใจใส่ และให้กำลังใจให้ผู้วิจัยเสมอมา

ปิยธิดา รุจะศิริ

มีนาคม 2552

สารบัญ

	หน้า
สารบัญ	(1)
สารบัญตาราง	(2)
สารบัญภาพ	(13)
คำอธิบายสัญลักษณ์และคำย่อ	(16)
คำนำ	1
วัตถุประสงค์	3
การตรวจเอกสาร	9
อุปกรณ์และวิธีการ	64
อุปกรณ์	64
วิธีการ	64
ผลและวิจารณ์	70
ผล	70
วิจารณ์	148
สรุปและข้อเสนอแนะ	149
สรุป	149
ข้อเสนอแนะ	151
เอกสารและสิ่งอ้างอิง	152
ภาคผนวก	156
ภาคผนวก ก โปรแกรมที่ใช้ในการวิจัย	157
ภาคผนวก ข ข้อมูลจริงที่นำมาศึกษา	168
ประวัติการศึกษา และการทำงาน	185

สารบัญตาราง

ตารางที่		หน้า
1	ลักษณะข้อมูลจำนวน n หน่วยและ p ตัวแปร	9
2	ขั้นตอนการจัดกลุ่มแบบลำดับขั้นประเภทการเกาะกลุ่ม	19
3	ข้อมูลตัวอย่างที่ 1: A, B, C และ D และมีตัวแปรจำนวน 2 ตัวแปร	23
4	การจัดกลุ่มแบบ Single Linkage	25
5	การจัดกลุ่มแบบ Complete Linkage	27
6	การจัดกลุ่มแบบ Average Linkage Between Groups	28
7	การจัดกลุ่มแบบ Ward Linkage	30
8	ขั้นตอนการจัดกลุ่มโดยอัลกอริทึมจัดกลุ่มแบบเคมีน	32
9	ข้อมูลตัวอย่างที่ 2: A, B, C และ D และมีตัวแปรจำนวน 2 ตัวแปร	35
10	จุดศูนย์กลางเริ่มต้นของข้อมูลแต่ละกลุ่มในตัวอย่างที่ 2	35
11	ค่าเฉลี่ยของข้อมูลแต่ละกลุ่ม เมื่อข้อมูลมีการเปลี่ยนแปลงกลุ่ม	37
12	ระยะห่างระหว่างข้อมูลกับค่าเฉลี่ยของข้อมูลแต่ละกลุ่ม	37
13	ขั้นตอนการจัดกลุ่มโดยวิธีจัดระบบด้วยตัวเองของโคโฮเนน	39
14	ข้อมูลตัวอย่างที่ 3: ประกอบด้วยตัวแปร X_1 , X_2 , X_3 และ X_4	41
15	ตัวอย่างกลุ่มข้อมูลที่ได้จากการจัดกลุ่มโดยวิธีใช้ SOM อัลกอริทึม	45
16	ขั้นตอนการแบ่งกลุ่มโดยอัลกอริทึมจัดกลุ่มแบบเคมคอยด์	46
17	ข้อมูลตัวอย่างที่ 4: A, B, C, D และ E และมีตัวแปรจำนวน 2 ตัวแปร	47
18	จุดศูนย์กลางเริ่มต้นของข้อมูลแต่ละกลุ่มในตัวอย่างที่ 4	48
19	ข้อมูลความเร็วของการเดิน 5 วินาทีของนาย A และนาย B	50
20	ขั้นตอนการแบ่งกลุ่มโดยอัลกอริทึมจัดกลุ่มแบบเคมคอยด์ร่วมกับวิธีวัดระยะห่างแบบไดนามิกไทม์วอร์ปปีง	53
21	ข้อมูลตัวอย่างที่ 5: A, B, C, D, E และ G และมีตัวแปรจำนวน 5 ตัวแปร	55
22	จุดศูนย์กลางเริ่มต้นของข้อมูลแต่ละกลุ่มในตัวอย่างที่ 5	55
23	ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 1 ค่าเฉลี่ยเท่ากับ μ_{31} และค่าความแปรปรวนรวมเท่ากับ Σ_{31}	71

สารบัญตาราง (ต่อ)

ตารางที่		หน้า
24	ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 2 ค่าเฉลี่ยเท่ากับ μ_{32} และค่าความแปรปรวนรวมเท่ากับ Σ_{31}	71
25	ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 3 ค่าเฉลี่ยเท่ากับ μ_{32} และค่าความแปรปรวนรวมเท่ากับ Σ_{32}	72
26	ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 4 ค่าเฉลี่ยเท่ากับ μ_{33} และค่าความแปรปรวนรวมเท่ากับ Σ_{32}	72
27	ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 5 ค่าเฉลี่ยเท่ากับ μ_{33} และค่าความแปรปรวนรวมเท่ากับ Σ_{33}	73
28	ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 1 ค่าเฉลี่ยเท่ากับ μ_{51} และค่าความแปรปรวนรวมเท่ากับ Σ_{51}	73
29	ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 2 ค่าเฉลี่ยเท่ากับ μ_{52} และค่าความแปรปรวนรวมเท่ากับ Σ_{51}	74
30	ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 3 ค่าเฉลี่ยเท่ากับ μ_{52} และค่าความแปรปรวนรวมเท่ากับ Σ_{52}	74
31	ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 4 ค่าเฉลี่ยเท่ากับ μ_{53} และค่าความแปรปรวนรวมเท่ากับ Σ_{52}	75
32	ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 5 ค่าเฉลี่ยเท่ากับ μ_{53} และค่าความแปรปรวนรวมเท่ากับ Σ_{53}	75
33	ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 1 ค่าเฉลี่ยเท่ากับ μ_{71} และค่าความแปรปรวนรวมเท่ากับ Σ_{71}	76
34	ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 2 ค่าเฉลี่ยเท่ากับ μ_{72} และค่าความแปรปรวนรวมเท่ากับ Σ_{71}	76
35	ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 3 ค่าเฉลี่ยเท่ากับ μ_{72} และค่าความแปรปรวนรวมเท่ากับ Σ_{72}	77

สารบัญตาราง (ต่อ)

ตารางที่		หน้า
36	ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 4 ค่าเฉลี่ยเท่ากับ μ_{73} และค่าความแปรปรวนร่วมเท่ากับ Σ_{72}	77
37	ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 5 ค่าเฉลี่ยเท่ากับ μ_{73} และค่าความแปรปรวนร่วมเท่ากับ Σ_{73}	78
38	ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 1 ค่าเฉลี่ยเท่ากับ μ_{31} และค่าความแปรปรวนร่วมเท่ากับ Σ_{31}	79
39	ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 2 ค่าเฉลี่ยเท่ากับ μ_{32} และค่าความแปรปรวนร่วมเท่ากับ Σ_{31}	80
40	ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 3 ค่าเฉลี่ยเท่ากับ μ_{32} และค่าความแปรปรวนร่วมเท่ากับ Σ_{32}	80
41	ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 4 ค่าเฉลี่ยเท่ากับ μ_{33} และค่าความแปรปรวนร่วมเท่ากับ Σ_{32}	81
42	ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 5 ค่าเฉลี่ยเท่ากับ μ_{33} และค่าความแปรปรวนร่วมเท่ากับ Σ_{33}	81
43	ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 1 ค่าเฉลี่ยเท่ากับ μ_{51} และค่าความแปรปรวนร่วมเท่ากับ Σ_{51}	82
44	ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 2 ค่าเฉลี่ยเท่ากับ μ_{52} และค่าความแปรปรวนร่วมเท่ากับ Σ_{51}	82
45	ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 3 ค่าเฉลี่ยเท่ากับ μ_{52} และค่าความแปรปรวนร่วมเท่ากับ Σ_{52}	83
46	ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 4 ค่าเฉลี่ยเท่ากับ μ_{53} และค่าความแปรปรวนร่วมเท่ากับ Σ_{52}	83
47	ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 5 ค่าเฉลี่ยเท่ากับ μ_{53} และค่าความแปรปรวนร่วมเท่ากับ Σ_{53}	84

สารบัญตาราง (ต่อ)

ตารางที่		หน้า
48	ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 1 ค่าเฉลี่ยเท่ากับ μ_{71} และค่าความแปรปรวนรวมเท่ากับ Σ_{71}	84
49	ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 2 ค่าเฉลี่ยเท่ากับ μ_{72} และค่าความแปรปรวนรวมเท่ากับ Σ_{71}	85
50	ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 3 ค่าเฉลี่ยเท่ากับ μ_{72} และค่าความแปรปรวนรวมเท่ากับ Σ_{72}	85
51	ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 4 ค่าเฉลี่ยเท่ากับ μ_{73} และค่าความแปรปรวนรวมเท่ากับ Σ_{72}	86
52	ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 5 ค่าเฉลี่ยเท่ากับ μ_{73} และค่าความแปรปรวนรวมเท่ากับ Σ_{73}	86
53	ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 1 ค่าเฉลี่ยเท่ากับ μ_{31} และค่าความแปรปรวนรวมเท่ากับ Σ_{31}	88
54	ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 2 ค่าเฉลี่ยเท่ากับ μ_{32} และค่าความแปรปรวนรวมเท่ากับ Σ_{31}	89
55	ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 3 ค่าเฉลี่ยเท่ากับ μ_{32} และค่าความแปรปรวนรวมเท่ากับ Σ_{32}	90
56	ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 4 ค่าเฉลี่ยเท่ากับ μ_{33} และค่าความแปรปรวนรวมเท่ากับ Σ_{32}	91
57	ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 5 ค่าเฉลี่ยเท่ากับ μ_{33} และค่าความแปรปรวนรวมเท่ากับ Σ_{33}	92
58	ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 1 ค่าเฉลี่ยเท่ากับ μ_{51} และค่าความแปรปรวนรวมเท่ากับ Σ_{51}	93
59	ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 2 ค่าเฉลี่ยเท่ากับ μ_{52} และค่าความแปรปรวนรวมเท่ากับ Σ_{51}	94

สารบัญตาราง (ต่อ)

ตารางที่		หน้า
60	ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 3 ค่าเฉลี่ยเท่ากับ μ_{52} และค่าความแปรปรวนรวมเท่ากับ Σ_{52}	95
61	ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 4 ค่าเฉลี่ยเท่ากับ μ_{53} และค่าความแปรปรวนรวมเท่ากับ Σ_{52}	96
62	ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 5 ค่าเฉลี่ยเท่ากับ μ_{53} และค่าความแปรปรวนรวมเท่ากับ Σ_{53}	97
63	ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 1 ค่าเฉลี่ยเท่ากับ μ_{71} และค่าความแปรปรวนรวมเท่ากับ Σ_{71}	98
64	ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 2 ค่าเฉลี่ยเท่ากับ μ_{72} และค่าความแปรปรวนรวมเท่ากับ Σ_{71}	99
65	ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 3 ค่าเฉลี่ยเท่ากับ μ_{72} และค่าความแปรปรวนรวมเท่ากับ Σ_{72}	100
66	ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 4 ค่าเฉลี่ยเท่ากับ μ_{73} และค่าความแปรปรวนรวมเท่ากับ Σ_{72}	101
67	ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 5 ค่าเฉลี่ยเท่ากับ μ_{73} และค่าความแปรปรวนรวมเท่ากับ Σ_{73}	102
68	ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 1 ค่าเฉลี่ยเท่ากับ μ_{31} และค่าความแปรปรวนรวมเท่ากับ Σ_{31}	104
69	ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 2 ค่าเฉลี่ยเท่ากับ μ_{32} และค่าความแปรปรวนรวมเท่ากับ Σ_{31}	104
70	ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 3 ค่าเฉลี่ยเท่ากับ μ_{32} และค่าความแปรปรวนรวมเท่ากับ Σ_{32}	105
71	ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 4 ค่าเฉลี่ยเท่ากับ μ_{33} และค่าความแปรปรวนรวมเท่ากับ Σ_{32}	105

สารบัญตาราง (ต่อ)

ตารางที่		หน้า
72	ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 5 ค่าเฉลี่ยเท่ากับ μ_{33} และค่าความแปรปรวนร่วมเท่ากับ Σ_{33}	106
73	ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 1 ค่าเฉลี่ยเท่ากับ μ_{51} และค่าความแปรปรวนร่วมเท่ากับ Σ_{51}	106
74	ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 2 ค่าเฉลี่ยเท่ากับ μ_{52} และค่าความแปรปรวนร่วมเท่ากับ Σ_{51}	107
75	ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 3 ค่าเฉลี่ยเท่ากับ μ_{52} และค่าความแปรปรวนร่วมเท่ากับ Σ_{52}	107
76	ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 4 ค่าเฉลี่ยเท่ากับ μ_{53} และค่าความแปรปรวนร่วมเท่ากับ Σ_{52}	108
77	ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 5 ค่าเฉลี่ยเท่ากับ μ_{53} และค่าความแปรปรวนร่วมเท่ากับ Σ_{53}	108
78	ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 1 ค่าเฉลี่ยเท่ากับ μ_{71} และค่าความแปรปรวนร่วมเท่ากับ Σ_{71}	109
79	ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 2 ค่าเฉลี่ยเท่ากับ μ_{72} และค่าความแปรปรวนร่วมเท่ากับ Σ_{71}	109
80	ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 3 ค่าเฉลี่ยเท่ากับ μ_{72} และค่าความแปรปรวนร่วมเท่ากับ Σ_{72}	110
81	ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 4 ค่าเฉลี่ยเท่ากับ μ_{73} และค่าความแปรปรวนร่วมเท่ากับ Σ_{72}	110
82	ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 5 ค่าเฉลี่ยเท่ากับ μ_{73} และค่าความแปรปรวนร่วมเท่ากับ Σ_{73}	111
83	ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 1 ค่าเฉลี่ยเท่ากับ μ_{31} และค่าความแปรปรวนร่วมเท่ากับ Σ_{31}	112

สารบัญตาราง (ต่อ)

ตารางที่		หน้า
84	ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 2 ค่าเฉลี่ยเท่ากับ μ_{32} และค่าความแปรปรวนร่วมเท่ากับ Σ_{31}	113
85	ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 3 ค่าเฉลี่ยเท่ากับ μ_{32} และค่าความแปรปรวนร่วมเท่ากับ Σ_{32}	113
86	ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 4 ค่าเฉลี่ยเท่ากับ μ_{33} และค่าความแปรปรวนร่วมเท่ากับ Σ_{32}	114
87	ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 5 ค่าเฉลี่ยเท่ากับ μ_{33} และค่าความแปรปรวนร่วมเท่ากับ Σ_{33}	114
88	ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 1 ค่าเฉลี่ยเท่ากับ μ_{51} และค่าความแปรปรวนร่วมเท่ากับ Σ_{51}	115
89	ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 2 ค่าเฉลี่ยเท่ากับ μ_{52} และค่าความแปรปรวนร่วมเท่ากับ Σ_{51}	115
90	ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 3 ค่าเฉลี่ยเท่ากับ μ_{52} และค่าความแปรปรวนร่วมเท่ากับ Σ_{52}	116
91	ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 4 ค่าเฉลี่ยเท่ากับ μ_{53} และค่าความแปรปรวนร่วมเท่ากับ Σ_{52}	116
92	ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 5 ค่าเฉลี่ยเท่ากับ μ_{53} และค่าความแปรปรวนร่วมเท่ากับ Σ_{53}	117
93	ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 1 ค่าเฉลี่ยเท่ากับ μ_{71} และค่าความแปรปรวนร่วมเท่ากับ Σ_{71}	117
94	ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 2 ค่าเฉลี่ยเท่ากับ μ_{72} และค่าความแปรปรวนร่วมเท่ากับ Σ_{71}	118
95	ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 3 ค่าเฉลี่ยเท่ากับ μ_{72} และค่าความแปรปรวนร่วมเท่ากับ Σ_{72}	118

สารบัญตาราง (ต่อ)

ตารางที่		หน้า
96	ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 4 ค่าเฉลี่ยเท่ากับ μ_{73} และค่าความแปรปรวนร่วมเท่ากับ Σ_{72}	119
97	ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 5 ค่าเฉลี่ยเท่ากับ μ_{73} และค่าความแปรปรวนร่วมเท่ากับ Σ_{73}	119
98	ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 1 ค่าเฉลี่ยเท่ากับ μ_{31} และค่าความแปรปรวนร่วมเท่ากับ Σ_{31}	121
99	ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 2 ค่าเฉลี่ยเท่ากับ μ_{32} และค่าความแปรปรวนร่วมเท่ากับ Σ_{31}	122
100	ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 3 ค่าเฉลี่ยเท่ากับ μ_{32} และค่าความแปรปรวนร่วมเท่ากับ Σ_{32}	123
101	ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 4 ค่าเฉลี่ยเท่ากับ μ_{33} และค่าความแปรปรวนร่วมเท่ากับ Σ_{32}	124
102	ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 5 ค่าเฉลี่ยเท่ากับ μ_{33} และค่าความแปรปรวนร่วมเท่ากับ Σ_{33}	125
103	ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 1 ค่าเฉลี่ยเท่ากับ μ_{51} และค่าความแปรปรวนร่วมเท่ากับ Σ_{51}	126
104	ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 2 ค่าเฉลี่ยเท่ากับ μ_{52} และค่าความแปรปรวนร่วมเท่ากับ Σ_{51}	127
105	ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 3 ค่าเฉลี่ยเท่ากับ μ_{52} และค่าความแปรปรวนร่วมเท่ากับ Σ_{52}	128
106	ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 4 ค่าเฉลี่ยเท่ากับ μ_{53} และค่าความแปรปรวนร่วมเท่ากับ Σ_{52}	129
107	ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 5 ค่าเฉลี่ยเท่ากับ μ_{53} และค่าความแปรปรวนร่วมเท่ากับ Σ_{53}	130

สารบัญตาราง (ต่อ)

ตารางที่		หน้า
108	ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 1 ค่าเฉลี่ยเท่ากับ μ_{71} และค่าความแปรปรวนร่วมเท่ากับ Σ_{71}	131
109	ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 2 ค่าเฉลี่ยเท่ากับ μ_{72} และค่าความแปรปรวนร่วมเท่ากับ Σ_{71}	132
110	ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 3 ค่าเฉลี่ยเท่ากับ μ_{72} และค่าความแปรปรวนร่วมเท่ากับ Σ_{72}	133
111	ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 4 ค่าเฉลี่ยเท่ากับ μ_{73} และค่าความแปรปรวนร่วมเท่ากับ Σ_{72}	134
112	ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 5 ค่าเฉลี่ยเท่ากับ μ_{73} และค่าความแปรปรวนร่วมเท่ากับ Σ_{73}	135
113	ข้อมูลการจำแนกคนไข้ทางสายตาเป็น 3 กลุ่มคือ กลุ่มที่ 1 คนไข้เหมาะกับคอนแทคเลนส์แบบแข็ง กลุ่มที่ 2 คนไข้เหมาะกับคอนแทคเลนส์แบบอ่อน และ กลุ่มที่ 3 คนไข้ไม่เหมาะกับคอนแทคเลนส์ แบ่งตามคุณลักษณะ 4 ประเภท มีข้อมูลทั้งหมด 24 หน่วยตัวอย่าง	137
114	ค่า RMSSTD และ RS ของข้อมูลการจำแนกคนไข้ทางสายตา	138
115	ข้อมูลการจำแนกคนไข้ที่เป็นมะเร็งเต้านม 2 กลุ่มคือ กลุ่มที่ 1 คนไข้ที่มีอายุยืนมากกว่า 5 ปีขึ้นไป และกลุ่มที่ 2 คนไข้ที่ตายภายใน 5 ปี แบ่งตามคุณลักษณะ 3 ประเภท คือ อายุของผู้ป่วยในช่วงเวลาที่สำรวจ ปีที่ผู้ป่วยมาเข้ารับการรักษา (หน่วยเป็นปีค.ศ.-1900) และจำนวนก้อนเนื้อที่เกิดขึ้น มีข้อมูลทั้งหมด 33 หน่วยตัวอย่าง	139
116	ค่า RMSSTD และ RS ของข้อมูลการจำแนกคนไข้ที่เป็นมะเร็งเต้านม	141
117	ข้อมูลผลการคัดแยกสับปะรดเป็น 2 กลุ่มคือ กลุ่มที่ 1 สับปะรดอ่อน และกลุ่มที่ 2 สับปะรดสุก ทดสอบด้วยเครื่องทดสอบสมบัติทางกายภาพของผลไม้และเก็บข้อมูล 3 ลักษณะ ได้แก่ ความถี่สั่นพ้อง ค่าเฉลี่ยความหนาแน่นเนื้อ และปริมาณน้ำตาล มีข้อมูลทั้งหมด 149 หน่วยตัวอย่าง	142

สารบัญตาราง (ต่อ)

ตารางที่		หน้า
118	ค่า RMSSTD และ RS ของข้อมูลการคัดแยกสับปะรด	144
119	ข้อมูลต้นไอริส 3 กลุ่มคือ กลุ่มที่ 1 Iris Setosa กลุ่มที่ 2 Iris Versicolour และกลุ่มที่ 3 Iris Virginica แบ่งตามคุณลักษณะ 4 ประเภท คือ ความยาวของกลีบเลี้ยง ความกว้างของกลีบเลี้ยง ความยาวของกลีบดอก และความกว้างของกลีบดอก ซึ่งวัดในหน่วย ซม. มีข้อมูลทั้งหมด 150 หน่วยตัวอย่าง	145
120	ค่า RMSSTD และ RS ของข้อมูลการจำแนกต้นไอริส	147
121	สรุปผลการแบ่งกลุ่มข้อมูลจำลองด้วยค่าเฉลี่ย RMSSTD	149
122	สรุปผลการแบ่งกลุ่มข้อมูลจำลองด้วยค่าเฉลี่ย RS	150

ตารางผนวกที่

ข1	ข้อมูลการจำแนกคนไข้ทางสายตาเป็น 3 กลุ่มคือ กลุ่มที่ 1 คนไข้เหมาะกับคอนแทคเลนส์แบบแข็ง กลุ่มที่ 2 คนไข้เหมาะกับคอนแทคเลนส์แบบอ่อน และ กลุ่มที่ 3 คนไข้ไม่เหมาะกับคอนแทคเลนส์ แบ่งตามคุณลักษณะ 4 ประเภท มีข้อมูลทั้งหมด 24 หน่วยตัวอย่าง	170
ข2	ข้อมูลการจำแนกคนไข้ที่เป็นมะเร็งเต้านม 2 กลุ่มคือ กลุ่มที่ 1 คนไข้ที่มีอายุยืนมากกว่า 5 ปีขึ้นไป และกลุ่มที่ 2 คนไข้ที่ตายภายใน 5 ปี แบ่งตามคุณลักษณะ 3 ประเภท คือ อายุของผู้ป่วยในช่วงเวลาที่สำรวจ ปีที่ผู้ป่วยมาเข้ารับการรักษา (หน่วยเป็นปีค.ศ.-1900) และจำนวนก้อนเนื้อที่เกิดขึ้น มีข้อมูลทั้งหมด 33 หน่วยตัวอย่าง	171
ข3	ข้อมูลผลการคัดแยกสับปะรดเป็น 2 กลุ่มคือ กลุ่มที่ 1 สับปะรดอ่อน และกลุ่มที่ 2 สับปะรดสุก ทดสอบด้วยเครื่องทดสอบสมบัติทางกายภาพของผลไม้และเก็บข้อมูล 3 ลักษณะ ได้แก่ ความถี่สั่นพ้อง ค่าเฉลี่ยความหนาแน่นเนื้อ และปริมาณน้ำตาล มีข้อมูลทั้งหมด 149 หน่วยตัวอย่าง	173

สารบัญตาราง (ต่อ)

ตารางผนวกที่	หน้า
ข4	ข้อมูลต้นไอริส 3 กลุ่มคือ กลุ่มที่ 1 Iris Setosa กลุ่มที่ 2 Iris Versicolour และ กลุ่มที่ 3 Iris Virginica แบ่งตามคุณลักษณะ 4 ประเภท คือ ความยาวของกลีบ เลี้ยง ความกว้างของกลีบเลี้ยง ความยาวของกลีบดอก และความกว้างของ กลีบดอก ซึ่งวัดในหน่วย ซม. มีข้อมูลทั้งหมด 150 หน่วยตัวอย่าง
	179

สารบัญภาพ

ภาพที่		หน้า
1	ตัวอย่างการแบ่งกลุ่มข้อมูลเป็น 3 กลุ่มโดยการวิเคราะห์กลุ่ม	13
2	การจัดกลุ่มแบบลำดับขั้นของข้อมูลการแสดงออกของยีน	17
3	ตัวอย่างการสร้างและแบ่งกลุ่มแบบลำดับขั้นด้วยเดนโดรแกรม	18
4	ตัวอย่างการสร้างกลุ่มแบบเกาะกลุ่มและแตกกลุ่ม	19
5	การรวมกลุ่มแบบ Single Linkage	20
6	การรวมกลุ่มแบบ Complete Linkage	20
7	การรวมกลุ่มแบบ Average Linkage Between Groups	21
8	ขั้นตอนการจัดกลุ่มข้อมูลแบบลำดับขั้นประเภทการเกาะกลุ่ม	22
9	แผนภูมิต้นไม้การจัดกลุ่มแบบ Single Linkage ของข้อมูลตัวอย่างที่ 1	26
10	แผนภูมิต้นไม้การจัดกลุ่มแบบ Complete Linkage ของข้อมูลตัวอย่างที่ 1	27
11	แผนภูมิต้นไม้การจัดกลุ่มแบบ Average Linkage Between Groups ของข้อมูลตัวอย่างที่ 1	29
12	แผนภูมิต้นไม้การจัดกลุ่มแบบ Ward Linkage ของข้อมูลตัวอย่างที่ 1	31
13	การกำหนดจุดข้อมูลเพื่อใช้ในแบ่งกลุ่มข้อมูลแบบเคมีน	33
14	ตัวอย่างการเลื่อนจุดข้อมูล แสดงถึงการเปลี่ยนจุดศูนย์กลางของแต่ละกลุ่มไปยังตำแหน่งที่ครอบคลุมข้อมูลบริเวณใกล้เคียงภายในกลุ่มเดียวกันให้ได้มากที่สุด โดยใช้วิธีการหาค่าเฉลี่ยเลขคณิต (Arithmetic Means)	33
15	การแบ่งกลุ่มที่เสร็จสมบูรณ์ โดยจุดศูนย์กลางที่ได้จะอยู่ในตำแหน่งที่เป็นตัวแทนของข้อมูลแต่ละกลุ่ม	34
16	ขั้นตอนการจัดกลุ่มข้อมูล โดยใช้เคมีนอัลกอริทึม	34
17	การทำงานของระบบโครงข่ายประสาทเทียมแบบ SOM	38
18	ระบบโครงข่ายประสาทเทียมแบบมีชั้นซ่อน	38
19	ขั้นตอนการจัดกลุ่มข้อมูล โดยใช้วิธีจัดระบบด้วยตัวเองของโคโฮเนน	40
20	Input และ Output ของโครงข่ายประสาทเทียม	41
21	ขั้นตอนการจัดกลุ่มข้อมูล โดยใช้อัลกอริทึมจัดกลุ่มแบบเคเมคอยด์	47

สารบัญภาพ (ต่อ)

ภาพที่		หน้า
22	ตำแหน่งที่ใช้ในการคำนวณด้วยวิธีวัดระยะห่างแบบยุคลิด	51
23	ตำแหน่งที่ใช้ในการคำนวณด้วยวิธีวัดระยะห่างแบบไดนามิกโทมัสวอร์ปปีง	52
24	ขั้นตอนการจัดกลุ่มข้อมูลโดยใช้อัลกอริทึมจัดกลุ่มแบบเคเมคอยด์ร่วมกับวิธีวัดระยะห่างแบบไดนามิกโทมัสวอร์ปปีง	54
25	ผลรวมของค่าระยะห่างสะสมในการคำนวณค่า DTW(A,G)	57
26	ผังงานสรุปขั้นตอนการดำเนินงานวิจัย	69
27	ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 3 ค่าเฉลี่ยเท่ากับ μ_{72} และค่าความแปรปรวนรวมเท่ากับ Σ_{72} จำนวนตัวแปรอิสระ 7 ตัวแปร ขนาดตัวอย่างเท่ากับ 25	78
28	ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 5 ค่าเฉลี่ยเท่ากับ μ_{33} และค่าความแปรปรวนรวมเท่ากับ Σ_{33} จำนวนตัวแปรอิสระ 3 ตัวแปร ขนาดตัวอย่างเท่ากับ 100	87
29	ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 4 ค่าเฉลี่ยเท่ากับ μ_{53} และค่าความแปรปรวนรวมเท่ากับ Σ_{52} จำนวนตัวแปรอิสระ 5 ตัวแปร ขนาดตัวอย่างเท่ากับ 300	103
30	ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 3 ค่าเฉลี่ยเท่ากับ μ_{72} และค่าความแปรปรวนรวมเท่ากับ Σ_{72} จำนวนตัวแปรอิสระ 7 ตัวแปร ขนาดตัวอย่างเท่ากับ 25	111
31	ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 1 ค่าเฉลี่ยเท่ากับ μ_{31} และค่าความแปรปรวนรวมเท่ากับ Σ_{31} จำนวนตัวแปรอิสระ 3 ตัวแปร ขนาดตัวอย่างเท่ากับ 100	120
32	ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 4 ค่าเฉลี่ยเท่ากับ μ_{53} และค่าความแปรปรวนรวมเท่ากับ Σ_{52} จำนวนตัวแปรอิสระ 5 ตัวแปร ขนาดตัวอย่างเท่ากับ 300	136
33	แผนภูมิเปอร์เซ็นต์ความแม่นยำในการแบ่งคนไข้ทางสายตาเป็น 3 กลุ่มด้วยวิธีต่าง ๆ	138
34	แผนภูมิเปอร์เซ็นต์ความแม่นยำในการแบ่งคนไข้มะเร็งเต้านมเป็น 2 กลุ่มด้วยวิธีต่าง ๆ	140

สารบัญภาพ (ต่อ)

ภาพที่		หน้า
35	แผนภูมิเปอร์เซ็นต์ความแม่นยำในการคัดแยกสับประรดเป็น 2 กลุ่มด้วยวิธีต่าง ๆ	143
36	แผนภูมิเปอร์เซ็นต์ความแม่นยำในการแยกประเภทต้นไทรเป็น 3 กลุ่มด้วยวิธีต่าง ๆ	146

คำอธิบายสัญลักษณ์และคำย่อ

Hier. Single	=	วิธีการจัดกลุ่มแบบลำดับขั้นด้วยวิธี Single Linkage
Hier. Average	=	วิธีการจัดกลุ่มแบบลำดับขั้นด้วยวิธี Average Linkage Between Groups
Hier. Complete	=	วิธีการจัดกลุ่มแบบลำดับขั้นด้วยวิธี Complete Linkage
Hier. Ward	=	วิธีการจัดกลุ่มแบบลำดับขั้นด้วยวิธี Ward Linkage
K-mean	=	อัลกอริทึมจัดกลุ่มแบบเคมีน
SOM	=	วิธีจัดระบบด้วยตัวเองของโคโฮเนน
K-med	=	อัลกอริทึมจัดกลุ่มแบบเคเมดอยด์
K-med (DTW)	=	อัลกอริทึมจัดกลุ่มแบบเคเมดอยด์ร่วมกับวิธีวัดระยะห่างแบบไดนามิก ไทม์วอร์ปปีง

การเปรียบเทียบเทคนิคการแบ่งกลุ่มข้อมูลสำหรับการวิเคราะห์กลุ่ม

The Comparison of Clustering Techniques for Cluster Analysis

คำนำ

งานวิจัยส่วนใหญ่ไม่ว่าจะเป็นทางด้านสังคมศาสตร์ ธุรกิจ อุตสาหกรรม เกษตรกรรม หรือ วิทยาศาสตร์ จำเป็นต้องใช้ข้อมูลในการดำเนินงาน เช่น ผลการทดลอง แบบสอบถาม ข้อมูลที่อ้างอิงมาจากแหล่งอื่น เป็นต้น เมื่อพิจารณาข้อมูลเหล่านี้จะพบว่า สามารถแบ่งข้อมูลออกเป็นกลุ่มย่อย ๆ ได้หลายกลุ่มตามความคล้ายคลึงกันของสิ่งที่ศึกษา

การวิเคราะห์แบ่งกลุ่มข้อมูลหรือจำแนกกลุ่มข้อมูลนั้น ตัวแปรตามจะเป็นตัวแปรเชิงกลุ่มหรือเชิงคุณภาพ ส่วนตัวแปรอิสระสามารถเป็นได้ทั้งตัวแปรเชิงปริมาณและเชิงคุณภาพ การวิเคราะห์ข้อมูลสำหรับงานวิจัยในปัจจุบัน โดยส่วนใหญ่จะศึกษาตัวแปรหลาย ๆ ตัวไปพร้อม ๆ กัน (Multivariate Analysis) ซึ่งเทคนิคการวิเคราะห์ข้อมูลหลายตัวแปรนี้เป็นการวิเคราะห์ข้อมูลที่มีความสัมพันธ์ของตัวแปรตั้งแต่ 3 ตัวขึ้นไปพร้อม ๆ กัน ซึ่งวัดจากหน่วยตัวอย่างเดียวกัน เช่น การศึกษาพฤติกรรมจากข้อมูลส่วนบุคคล ได้แก่ เพศ รายได้ อายุ น้ำหนัก เป็นต้น ซึ่งจะศึกษาความสัมพันธ์ของตัวแปรทุกตัวที่กล่าวมาพร้อมกัน (กัลยา, 2550) ดังนั้นการคัดเลือกตัวแปรอิสระที่มีความเหมาะสมจึงเป็นสิ่งจำเป็น เพื่อให้ผลการแบ่งกลุ่มที่ได้มีความผิดพลาดน้อยที่สุด โดยความผิดพลาดนี้อาจเกิดจากการละเลยตัวแปรที่จำเป็นบางตัวไป เทคนิควิธีการแบ่งกลุ่มข้อมูลหรือจำแนกกลุ่มข้อมูลมีหลายเทคนิค เทคนิคที่นิยมนำมาใช้มี 3 เทคนิค คือ การวิเคราะห์จำแนกประเภท (Discriminant Analysis) การวิเคราะห์การถดถอยโลจิสติก (Logistic Regression Analysis) และการวิเคราะห์กลุ่ม (Cluster Analysis)

ดังนั้นงานวิจัยนี้ ผู้วิจัยจึงมีความสนใจศึกษาเทคนิคการวิเคราะห์กลุ่ม ซึ่งเป็นเทคนิคที่ใช้ในการแบ่งกลุ่มข้อมูลใด ๆ ออกเป็นกลุ่มย่อย ๆ โดยผลที่ได้จากการวิเคราะห์กลุ่มนั้น ข้อมูลที่ถูกจัดอยู่ในกลุ่มเดียวกันจะมีลักษณะที่มีความเหมือนกันหรือคล้ายกัน ส่วนข้อมูลที่ถูกจัดอยู่ต่างกลุ่มกันจะมีลักษณะที่แตกต่างกัน โดยที่การแบ่งกลุ่มข้อมูลด้วยเทคนิคการวิเคราะห์กลุ่มนั้น การเลือกใช้ตัวแปรต่าง ๆ ในแต่ละมิติของชุดข้อมูลอย่างเหมาะสม จะส่งผลต่อประสิทธิภาพที่ดีของ

การแบ่งกลุ่มที่ได้ อีกด้วย และข้อมูลทุกหน่วยจะถูกจำแนกกลุ่มอยู่ในกลุ่มใดกลุ่มหนึ่งเพียงกลุ่มเดียวเท่านั้น โดยไม่มีความซ้ำซ้อน

ในปัจจุบันเทคนิคการวิเคราะห์กลุ่มอาจแบ่งย่อยเป็น 2 ประเภทใหญ่ ๆ คือ การวิเคราะห์กลุ่มแบบลำดับขั้น (Hierarchical cluster analysis) และการวิเคราะห์กลุ่มแบบไม่เป็นลำดับขั้น (Non-hierarchical cluster analysis) ได้แก่ อัลกอริทึมจัดกลุ่มแบบเคมีน (K-Means clustering algorithm) เป็นต้น ซึ่งงานวิจัยในปัจจุบันได้พัฒนาและสร้างเทคนิควิธีใหม่ ๆ เพิ่มขึ้นมากมาย เช่น วิธีจัดระบบด้วยตัวเองของโคโฮเนน (Kohonen's Self-Organizing Maps method) อัลกอริทึมจัดกลุ่มแบบเคเมดอยด์ (K-Medoids clustering algorithm) อัลกอริทึมเอ็กซ์เปคเตชัน-แมกซิไมเซชัน (Expectation-Maximization Algorithm) อัลกอริทึมจัดกลุ่มแบบ CLARANS เป็นต้น โดยเทคนิคที่ผู้วิจัยเลือกนำมาศึกษาเปรียบเทียบกับงานวิจัยนี้จำนวน 5 วิธี เป็นเทคนิคที่ครอบคลุมทั้ง 2 ประเภทหลักดังที่กล่าวมาแล้ว ได้แก่ วิธีการจัดกลุ่มแบบลำดับขั้น (Hierarchical clustering method) อัลกอริทึมจัดกลุ่มแบบเคมีน วิธีจัดระบบด้วยตัวเองของโคโฮเนน อัลกอริทึมจัดกลุ่มแบบเคเมดอยด์ และเทคนิคที่ได้รับการพัฒนาขึ้นใหม่ คือ การใช้อัลกอริทึมจัดกลุ่มแบบเคเมดอยด์ ร่วมกับวิธีวัดระยะห่างแบบไดนามิกไทม์วอร์ปปีง (K-Medoids clustering algorithm integrated with Dynamic Time Warping distance measure) ซึ่งมีเกณฑ์พิจารณา คือ ค่าความแตกต่างของข้อมูลภายในกลุ่ม (Root Mean Square Standard Deviation) และค่าความแตกต่างของข้อมูลระหว่างกลุ่ม (R Squared) ซึ่งถ้าค่าความแตกต่างของข้อมูลภายในกลุ่มต่ำ แสดงว่ากลุ่มที่แบ่งได้มีความเหมาะสม และถ้าค่าความแตกต่างของข้อมูลระหว่างกลุ่มสูง แสดงว่ากลุ่มที่แบ่งได้มีความเหมาะสม

วัตถุประสงค์

1. เพื่อศึกษาเทคนิคการแบ่งกลุ่ม 5 วิธี คือ วิธีการจัดกลุ่มแบบลำดับชั้น อัลกอริทึมจัดกลุ่มแบบเคมีน วิธีจัดระบบด้วยตัวเองของโคโฮเนน อัลกอริทึมจัดกลุ่มแบบเคมคอยด์ และ การใช้ อัลกอริทึมจัดกลุ่มแบบเคมคอยด์ร่วมกับการใช้วิธีวัดระยะห่างแบบไดนามิกไทม์วอร์ปปีง
2. เพื่อเปรียบเทียบความสามารถในการแบ่งกลุ่มระหว่างเทคนิควิธีการแบ่งกลุ่มวิธีต่าง ๆ ทั้ง 5 วิธีดังกล่าว โดยพิจารณาจากค่า RMSSTD และ RS

ขอบเขตการวิจัย

การวิจัยครั้งนี้เป็นการศึกษาเปรียบเทียบเทคนิคการแบ่งกลุ่มข้อมูลสำหรับการวิเคราะห์กลุ่มซึ่งมีขอบเขตการศึกษาดังต่อไปนี้

1. ตัวแปรที่ศึกษา
ตัวแปรอิสระ (Independent variables; X) เป็นตัวแปรเชิงปริมาณ และตัวแปรตาม (Dependent variable; Y) เป็นตัวแปรเชิงกลุ่ม 1 ตัว และกำหนดจำนวนตัวแปรอิสระ (p) เป็น 3, 5 และ 7 ตัวแปร
2. กำหนดขนาดตัวอย่าง 25, 100 และ 300 แทนข้อมูลขนาดเล็ก, ขนาดกลาง และขนาดใหญ่ ตามลำดับ และกำหนดจำนวนกลุ่มในข้อมูลขนาดเล็กอยู่ในช่วง 2-5 กลุ่ม ขนาดกลางอยู่ในช่วง 2-8 กลุ่ม ขนาดใหญ่อยู่ในช่วง 2-15 กลุ่ม
3. กำหนดรูปแบบการแจกแจงแบบปกติหลายตัวแปร (Multivariate Normal Distribution) ที่มีค่าเฉลี่ยและค่าความแปรปรวนร่วมกรณีละ 3 ค่า และจัดเป็นรูปแบบในแต่ละจำนวนตัวแปรอิสระที่กำหนดได้ 5 รูปแบบ ดังแสดงในหัวข้อ 3.1 ถึง 3.3 รวมทั้งสิ้น 15 รูปแบบ โดยที่ μ_{ij} แทนค่าเฉลี่ยแบบที่ j ตัวแปรอิสระจำนวน i ตัวแปร และ Σ_{ij} แทนค่าความแปรปรวนร่วม (Co-variance Matrix) แบบที่ j ตัวแปรอิสระจำนวน i ตัวแปร เมื่อ $i = 3, 5, 7$ และ $j = 1, 2, 3$

3.1 จำนวนตัวแปรอิสระ 3 ตัวแปร กำหนดค่าเฉลี่ยและค่าความแปรปรวนร่วม ดังนี้

$$\mu_{31} = \begin{bmatrix} 10 \\ 20 \\ 30 \end{bmatrix}, \mu_{32} = \begin{bmatrix} 100 \\ 200 \\ 300 \end{bmatrix}, \mu_{33} = \begin{bmatrix} 1000 \\ 2000 \\ 3000 \end{bmatrix}$$

$$\Sigma_{31} = \begin{bmatrix} 2 & 3 & 4 \\ 3 & 7 & 8 \\ 4 & 8 & 15 \end{bmatrix}, \Sigma_{32} = \begin{bmatrix} 30 & 35 & 40 \\ 35 & 60 & 70 \\ 40 & 70 & 140 \end{bmatrix}, \Sigma_{33} = \begin{bmatrix} 300 & 350 & 400 \\ 350 & 600 & 700 \\ 400 & 700 & 1400 \end{bmatrix}$$

และจัดเป็นรูปแบบได้ 5 รูปแบบ ดังนี้

3.1.1 รูปแบบที่ 1 การแจกแจงแบบปกติ มีค่าเฉลี่ยเท่ากับ μ_{31} และค่าความแปรปรวนร่วมเท่ากับ Σ_{31}

3.1.2 รูปแบบที่ 2 การแจกแจงแบบปกติ มีค่าเฉลี่ยเท่ากับ μ_{32} และค่าความแปรปรวนร่วมเท่ากับ Σ_{31}

3.1.3 รูปแบบที่ 3 การแจกแจงแบบปกติ มีค่าเฉลี่ยเท่ากับ μ_{32} และค่าความแปรปรวนร่วมเท่ากับ Σ_{32}

3.1.4 รูปแบบที่ 4 การแจกแจงแบบปกติ มีค่าเฉลี่ยเท่ากับ μ_{33} และค่าความแปรปรวนร่วมเท่ากับ Σ_{32}

3.1.5 รูปแบบที่ 5 การแจกแจงแบบปกติ มีค่าเฉลี่ยเท่ากับ μ_{33} และค่าความแปรปรวนร่วมเท่ากับ Σ_{33}

3.2 จำนวนตัวแปรอิสระ 5 ตัวแปร กำหนดค่าเฉลี่ยและค่าความแปรปรวนร่วม ดังนี้

$$\mu_{s_1} = \begin{bmatrix} 10 \\ 20 \\ 30 \\ 40 \\ 50 \end{bmatrix}, \mu_{s_2} = \begin{bmatrix} 100 \\ 200 \\ 300 \\ 400 \\ 500 \end{bmatrix}, \mu_{s_3} = \begin{bmatrix} 1000 \\ 2000 \\ 3000 \\ 4000 \\ 5000 \end{bmatrix}$$

$$\Sigma_{s_1} = \begin{bmatrix} 2 & 3 & 4 & 5 & 7 \\ 3 & 8 & 10 & 12 & 13 \\ 4 & 10 & 15 & 18 & 21 \\ 5 & 12 & 18 & 25 & 30 \\ 7 & 13 & 21 & 30 & 42 \end{bmatrix}, \Sigma_{s_2} = \begin{bmatrix} 30 & 35 & 40 & 48 & 52 \\ 35 & 60 & 75 & 81 & 96 \\ 40 & 75 & 140 & 156 & 187 \\ 48 & 81 & 156 & 230 & 290 \\ 52 & 96 & 187 & 290 & 400 \end{bmatrix},$$

$$\Sigma_{s_3} = \begin{bmatrix} 300 & 350 & 400 & 460 & 520 \\ 350 & 600 & 750 & 830 & 910 \\ 400 & 750 & 1400 & 1650 & 1890 \\ 460 & 830 & 1650 & 2300 & 2900 \\ 520 & 910 & 1890 & 2900 & 4000 \end{bmatrix}$$

และจัดเป็นรูปแบบได้ 5 รูปแบบ ดังนี้

3.2.1 รูปแบบที่ 1 การแจกแจงแบบปกติ มีค่าเฉลี่ยเท่ากับ μ_{s_1} และค่าความแปรปรวนร่วมเท่ากับ Σ_{s_1}

3.2.2 รูปแบบที่ 2 การแจกแจงแบบปกติ มีค่าเฉลี่ยเท่ากับ μ_{s_2} และค่าความแปรปรวนร่วมเท่ากับ Σ_{s_1}

3.2.3 รูปแบบที่ 3 การแจกแจงแบบปกติ มีค่าเฉลี่ยเท่ากับ μ_{s_2} และค่าความแปรปรวนร่วมเท่ากับ Σ_{s_2}

3.2.4 รูปแบบที่ 4 การแจกแจงแบบปกติ มีค่าเฉลี่ยเท่ากับ μ_{s_3} และค่าความแปรปรวนร่วมเท่ากับ Σ_{s_2}

3.2.5 รูปแบบที่ 5 การแจกแจงแบบปกติ มีค่าเฉลี่ยเท่ากับ μ_{53} และค่าความแปรปรวนรวมเท่ากับ Σ_{53}

3.3 จำนวนตัวแปรอิสระ 7 ตัวแปร กำหนดค่าเฉลี่ยและค่าความแปรปรวนรวม ดังนี้

$$\mu_{71} = \begin{bmatrix} 10 \\ 20 \\ 30 \\ 40 \\ 50 \\ 60 \\ 70 \end{bmatrix}, \mu_{72} = \begin{bmatrix} 100 \\ 200 \\ 300 \\ 400 \\ 500 \\ 600 \\ 700 \end{bmatrix}, \mu_{73} = \begin{bmatrix} 1000 \\ 2000 \\ 3000 \\ 4000 \\ 5000 \\ 6000 \\ 7000 \end{bmatrix}$$

$$\Sigma_{71} = \begin{bmatrix} 2 & 3 & 4 & 5 & 7 & 9 & 10 \\ 3 & 12 & 13 & 15 & 17 & 19 & 21 \\ 4 & 13 & 25 & 27 & 29 & 32 & 34 \\ 5 & 15 & 27 & 38 & 40 & 43 & 47 \\ 7 & 17 & 29 & 40 & 51 & 53 & 58 \\ 9 & 19 & 32 & 43 & 53 & 64 & 69 \\ 10 & 21 & 34 & 47 & 58 & 69 & 75 \end{bmatrix}, \Sigma_{72} = \begin{bmatrix} 30 & 35 & 40 & 46 & 52 & 59 & 63 \\ 35 & 70 & 74 & 80 & 85 & 92 & 98 \\ 40 & 74 & 110 & 120 & 132 & 140 & 161 \\ 46 & 80 & 120 & 180 & 193 & 200 & 208 \\ 52 & 85 & 132 & 193 & 230 & 240 & 254 \\ 59 & 92 & 140 & 200 & 240 & 280 & 310 \\ 63 & 98 & 161 & 208 & 254 & 310 & 400 \end{bmatrix},$$

$$\Sigma_{73} = \begin{bmatrix} 300 & 350 & 400 & 460 & 520 & 590 & 630 \\ 350 & 700 & 740 & 800 & 850 & 920 & 980 \\ 400 & 740 & 1100 & 1200 & 1320 & 1400 & 1610 \\ 460 & 800 & 1200 & 1800 & 1930 & 2000 & 2080 \\ 520 & 850 & 1320 & 1930 & 2300 & 2400 & 2540 \\ 590 & 920 & 1400 & 2000 & 2400 & 2800 & 3100 \\ 630 & 980 & 1610 & 2080 & 2540 & 3100 & 4000 \end{bmatrix}$$

และจัดเป็นรูปแบบได้ 5 รูปแบบ ดังนี้

3.3.1 รูปแบบที่ 1 การแจกแจงแบบปกติ มีค่าเฉลี่ยเท่ากับ μ_{71} และค่าความแปรปรวนรวมเท่ากับ Σ_{71}

3.3.2 รูปแบบที่ 2 การแจกแจงแบบปกติ มีค่าเฉลี่ยเท่ากับ μ_2 และค่าความแปรปรวนรวมเท่ากับ Σ_{71}

3.3.3 รูปแบบที่ 3 การแจกแจงแบบปกติ มีค่าเฉลี่ยเท่ากับ μ_2 และค่าความแปรปรวนรวมเท่ากับ Σ_{72}

3.3.4 รูปแบบที่ 4 การแจกแจงแบบปกติ มีค่าเฉลี่ยเท่ากับ μ_3 และค่าความแปรปรวนรวมเท่ากับ Σ_{72}

3.3.5 รูปแบบที่ 5 การแจกแจงแบบปกติ มีค่าเฉลี่ยเท่ากับ μ_3 และค่าความแปรปรวนรวมเท่ากับ Σ_{73}

4. เทคนิคการแบ่งกลุ่มที่นำมาศึกษา 5 วิธี ซึ่งแบ่งเป็น 2 ประเภทใหญ่ ๆ คือ

การวิเคราะห์กลุ่มแบบลำดับขั้น คือ (1) วิธีการจัดกลุ่มแบบลำดับขั้น

และการวิเคราะห์กลุ่มแบบไม่เป็นลำดับขั้น ได้แก่

(2) อัลกอริทึมจัดกลุ่มแบบเคมีน

(3) วิธีจัดระบบด้วยตัวเองของโคโฮเนน

(4) อัลกอริทึมจัดกลุ่มแบบเคเมคอยด์

(5) การใช้อัลกอริทึมจัดกลุ่มแบบเคเมคอยด์ ร่วมกับวิธีวิเคราะห์ทางแบบไดนามิกไทม์วอร์ปปีง

5. ในการเปรียบเทียบประสิทธิภาพและวัดความสามารถของการแบ่งกลุ่ม พิจารณาจากค่าความแตกต่างของข้อมูลภายในกลุ่ม (Root Mean Square Standard Deviation; RMSSTD) และค่าความแตกต่างของข้อมูลระหว่างกลุ่ม (R Squared; RS)

6. ในการวิจัยครั้งนี้ ทำการจำลองข้อมูลด้วยเทคนิคมอนติคาร์โล (Monte Carlo Simulation Technique) และทำซ้ำ (r) 1,000 ครั้งในแต่ละสถานการณ์ของการทดลอง

ประโยชน์ที่คาดว่าจะได้รับ

เป็นแนวทางในการเลือกใช้เทคนิควิธีการแบ่งกลุ่มได้อย่างเหมาะสมกับลักษณะของข้อมูล และสามารถนำผลการวิจัยที่ได้ไปประยุกต์ใช้กับการแบ่งกลุ่มของข้อมูลจริงได้อย่างถูกต้องเหมาะสม

การตรวจเอกสาร

การตรวจเอกสารจะแบ่งออกเป็น 2 ส่วน ส่วนแรกจะกล่าวถึงวิธีการทางสถิติที่ใช้ในการศึกษาและสถิติทดสอบ ส่วนหลังจะกล่าวถึงผลงานวิจัยที่เกี่ยวข้องดังนี้

วิธีการทางสถิติ

งานวิจัยส่วนใหญ่จำเป็นต้องอาศัยสถิติเป็นเครื่องมือในการช่วยตัดสินใจสำหรับหลาย ๆ ขั้นตอนในกระบวนการวิจัย การใช้วิธีการทางสถิติในงานวิจัยจะช่วยให้ผู้วิจัยสามารถหาข้อสรุปผลการวิจัยได้อย่างเหมาะสมและมีคุณภาพน่าเชื่อถือ

1. ลักษณะข้อมูลหลายตัวแปร

ข้อมูลหลายตัวแปรนั้น รวบรวมจากหลาย ๆ ลักษณะหรือหลาย ๆ ตัวแปรจากแต่ละหน่วย โดยกำหนดให้มีตัวแปร p ตัว คือ $X_1, X_2, \dots, X_p$ เมื่อ $p > 1$ และกำหนดให้ X_{ij} แทนตัวแปรที่ j ของหน่วยที่ i เมื่อ $i = 1, 2, \dots, n$ และ $j = 1, 2, \dots, p$

ตารางที่ 1 ลักษณะข้อมูลจำนวน n หน่วยและ p ตัวแปร

หน่วยที่ i	ตัวแปรที่ j			
	1	2	...	p
1	X_{11}	X_{12}	...	X_{1p}
2	X_{21}	X_{22}	...	X_{2p}
$\vdots$	$\vdots$	$\vdots$		$\vdots$
n	X_{n1}	X_{n2}	...	X_{np}

จากตารางที่ 1 สามารถเขียนข้อมูลในรูปเมทริกซ์ได้ดังนี้

$$\mathbf{X} = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & \cdots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{np} \end{bmatrix}$$

กรณีของข้อมูลที่ได้จากการสุ่มตัวอย่าง แสดงในรูปเมทริกซ์ได้ดังนี้

$$\mathbf{X} = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & \cdots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{np} \end{bmatrix}$$

และกำหนดให้เวกเตอร์ของตัวแปรคือ

$$\mathbf{x} = [X_1, X_2, \dots, X_p]$$

เนื่องจากข้อมูลมีหลายตัวแปร ซึ่งหน่วยที่ใช้ในการวัดข้อมูลแต่ละตัวแปรอาจแตกต่างกัน ดังนั้นในการนำข้อมูลมาวิเคราะห์จึงควรปรับข้อมูลให้มีหน่วยมาตรฐานเสียก่อน ซึ่งสามารถคำนวณได้ดังนี้

$$Z - Normalization = \frac{x_i - \bar{x}}{\hat{\sigma}_x} \quad ; i = 1, 2, \dots, n$$

$$\text{เมื่อ} \quad \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{และ} \quad \hat{\sigma}_x = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$

โดยที่

x_i คือข้อมูลหน่วยที่ i

$\bar{x}$ คือค่าเฉลี่ยของข้อมูล x_i

$\hat{\sigma}_x$ คือค่าประมาณส่วนเบี่ยงเบนมาตรฐานของข้อมูล x_i

2. การทดสอบการแจกแจงแบบปกติหลายตัวแปร (Multivariate Normal Distribution Test)

วิธีการทดสอบข้อมูลหลายตัวแปรว่ามีการแจกแจงแบบปกติหรือไม่นั้นทำได้หลายวิธี ซึ่งในที่นี้นำเสนอวิธีทดสอบการแจกแจง 2 วิธี ดังนี้

2.1 Chi-square q-q plot เป็นวิธีการที่ใช้ทดสอบความปกติของข้อมูลหลายตัวแปร (Johnson and Wichern, 1992) โดยใช้ค่าระยะห่างมาฮาลาโนบิส (Mahalanobis distances) ซึ่งมีการแจกแจงแบบไคสแควร์ สามารถคำนวณดังนี้

$$d_i^2 = (x_i - \mu)' \Sigma^{-1} (x_i - \mu) ; i = 1, 2, \dots, n$$

โดยที่	x_i	คือข้อมูลหน่วยที่ i
	μ	คือค่าเฉลี่ยของข้อมูลทั้งหมด
	Σ	คือค่าความแปรปรวนร่วมของข้อมูล
	Σ^{-1}	คือเมทริกซ์ผกผัน (Inverse Matrix) ของ Σ

นำค่า d_i^2 ที่คำนวณได้ไปสร้างกราฟ q-q plot เพื่อดูแนวโน้มของข้อมูล หากข้อมูลมีลักษณะการแจกแจงแบบปกติหลายตัวแปร กราฟที่ได้จะมีลักษณะเป็นเส้นตรง และสามารถทดสอบโดยใช้ค่าสถิติไคสแควร์ด้วยค่าองศาอิสระ (degree of freedom) เท่ากับจำนวนตัวแปรอิสระ (p) โดยตั้งสมมติฐานการทดสอบหลักคือ ข้อมูลมีการแจกแจงแบบปกติหลายตัวแปร

2.2 Henze-Zirkler test เป็นวิธีที่คิดค้นขึ้นโดย Henze and Zirkler (1990) โดยใช้ค่าระยะห่างแบบนอนเนกาทิฟฟังก์ชัน (non-negative functional distance) สามารถคำนวณได้ดังนี้

$$T = \frac{1}{n} \sum_{i=1}^n \exp\left(-\frac{\beta^2}{2} d_i^2\right) - 2(1 + \beta^2)^{-\frac{\alpha}{2}} \sum_{i=1}^n \exp\left(\frac{\beta^2}{2(1 + \beta^2)} d_i^2\right) + n(1 + 2\beta^2)^{-\frac{\alpha}{2}}$$

โดยที่	d_i^2	คือค่าระยะห่างมาฮาลาโนบิส
	α	คือค่าระดับนัยสำคัญ
	β	คือค่าพารามิเตอร์ ซึ่งมีค่าอยู่ในช่วง 0 ถึง 1
	n	คือจำนวนข้อมูลทั้งหมด

ข้อมูลจะมีการแจกแจงแบบปกติหลายตัวแปร ก็ต่อเมื่อ T มีการแจกแจงใกล้เคียงการแจกแจงแบบลอการิทึม (lognormal distribution) ซึ่งคำนวณค่าคาดหวังของ T ($E[T]$) และค่าความแปรปรวนของ T ($Var[T]$) ได้ดังนี้

$$E[T] = 1 - (1 + 2\beta^2)^{-\frac{\alpha}{2}} \left[1 + \frac{\alpha\beta^2}{1 + 2\beta^2} + \frac{\alpha(\alpha + 2)\beta^4}{2(1 + 2\beta^2)^2} \right]$$

$$Var[T] = 2(1 + 4\beta^2)^{-\frac{\alpha}{2}} + 2(1 + 2\beta^2)^{-\alpha} \left[1 + \frac{2\alpha\beta^4}{(1 + 2\beta^2)^2} + \frac{3\alpha(\alpha + 2)\beta^8}{4(1 + 2\beta^2)^4} \right]$$

$$- 2 \left[(1 + \beta^2)(1 + 3\beta^2) \right]^{-\frac{\alpha}{2}} \left[1 + \frac{3\alpha\beta^4}{2(1 + \beta^2)(1 + 3\beta^2)} + \frac{\alpha(\alpha + 2)\beta^8}{2 \left[(1 + \beta^2)(1 + 3\beta^2) \right]^2} \right]$$

3. การวิเคราะห์หลายตัวแปร (Multivariate Analysis)

การวิเคราะห์ข้อมูลหลายตัวแปร สามารถนำมาใช้ในการหาความสัมพันธ์ของตัวแปรอิสระตั้งแต่ 2 ตัวขึ้นไปพร้อม ๆ กัน ซึ่งเทคนิคที่ใช้วิเคราะห์ข้อมูลหลายตัวแปร มีหลายเทคนิค เช่น การวิเคราะห์ปัจจัย การวิเคราะห์จำแนกประเภท การวิเคราะห์กลุ่ม เป็นต้น (กัลยา, 2550)

3.1 การวิเคราะห์ปัจจัย (Factor Analysis) เป็นเทคนิคที่ใช้ในการแบ่งกลุ่มตัวแปรที่มีจำนวนมาก ๆ โดยการแบ่งกลุ่มตัวแปรออกเป็นปัจจัย (Factor) หลาย ๆ ปัจจัย ซึ่งกลุ่มตัวแปรที่อยู่ในปัจจัยเดียวกันจะมีความสัมพันธ์กัน และการวัดความสัมพันธ์ของตัวแปรจะพิจารณาจากค่าสัมประสิทธิ์สหสัมพันธ์ ดังนั้นตัวแปรกลุ่มที่อยู่ในปัจจัยเดียวกันจะมีความสัมพันธ์กันมาก โดยความสัมพันธ์อาจจะอยู่ในทิศทางเดียวกัน (ค่าสัมประสิทธิ์สหสัมพันธ์เป็นบวก) หรืออยู่ในทิศทางตรงกันข้าม (ค่าสัมประสิทธิ์สหสัมพันธ์เป็นลบ) แต่กลุ่มตัวแปรที่อยู่ในต่างปัจจัยกันจะไม่มีความสัมพันธ์กันหรือมีความสัมพันธ์กันน้อยมาก

3.2 การวิเคราะห์จำแนกกลุ่ม (Discriminant Analysis) เป็นเทคนิคที่ทำการแบ่งข้อมูลหรือหน่วยตัวอย่างออกเป็นกลุ่มย่อยหลาย ๆ กลุ่ม โดยที่คนหรือหน่วยที่อยู่ในกลุ่มเดียวกันจะมีความคล้ายคลึงกันในตัวแปรที่ศึกษา ซึ่งเทคนิคนี้จะต้องทราบมาก่อนว่ามีจำนวนกี่กลุ่ม ตัวแปรตามเป็นตัวแปรเชิงกลุ่ม ส่วนตัวแปรอิสระเป็นตัวแปรเชิงปริมาณหรือตัวแปรเชิงคุณภาพ นำตัวแปรที่

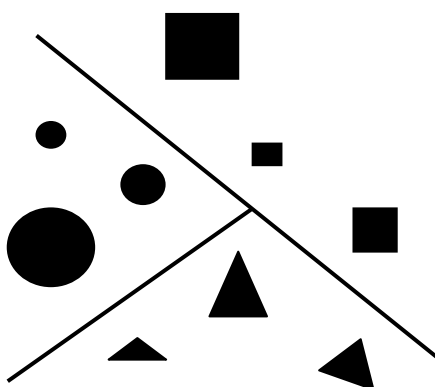
กำหนดไว้มาศึกษาหาความสัมพันธ์ที่อยู่ในรูปเชิงเส้น หลังจากนั้นจึงนำสมการเชิงเส้นที่ได้มาพยากรณ์ว่าหน่วยใหม่ควรอยู่กลุ่มใด

3.3 การวิเคราะห์กลุ่ม (Cluster Analysis) เป็นเทคนิคในการแบ่งกลุ่มข้อมูลออกเป็นกลุ่มย่อย ๆ โดยแบ่งกลุ่มข้อมูลที่คล้ายกันอยู่ในกลุ่มเดียวกัน ซึ่งไม่จำเป็นต้องทราบจำนวนกลุ่มมาก่อนหรือไม่จำเป็นต้องทราบว่าแต่ละหน่วยอยู่กลุ่มใดมาก่อน นิยมนำเทคนิคนี้มาใช้กับงานด้านการตลาด เช่น การแบ่งกลุ่มลูกค้าที่มีลักษณะหรือพฤติกรรมการบริโภคที่เหมือนกันหรือคล้ายกันให้อยู่กลุ่มเดียวกัน หรือด้านสังคมศาสตร์ เช่น ใช้แบ่งกลุ่มพฤติกรรมต่าง ๆ ของคนที่มีพฤติกรรมคล้ายกันไว้ในกลุ่มเดียวกัน

ดังนั้นในงานวิจัยนี้ ผู้วิจัยสนใจศึกษาเปรียบเทียบเทคนิคการแบ่งกลุ่มข้อมูล โดยใช้การวิเคราะห์กลุ่ม ดังรายละเอียดต่อไปนี้

4. การวิเคราะห์กลุ่ม (Cluster Analysis)

การวิเคราะห์ กลุ่มข้อมูลนั้น ประกอบด้วยส่วนที่สำคัญ 2 ส่วน คือ การวัดระยะห่าง และเทคนิควิธีการแบ่งกลุ่ม ซึ่งในปัจจุบันมีการพัฒนามีหลากหลายเทคนิควิธี



ภาพที่ 1 ตัวอย่างการแบ่งกลุ่มข้อมูลเป็น 3 กลุ่มโดยการวิเคราะห์กลุ่ม

ในภาพที่ 1 แสดงตัวอย่างการวิเคราะห์กลุ่ม โดยแบ่งข้อมูลที่มีลักษณะคล้ายคลึงกันไว้ในกลุ่มเดียวกัน และข้อมูลที่มีลักษณะแตกต่างกันจะถูกจัดอยู่ต่างกลุ่มกัน ซึ่งแบ่งได้เป็น 3 กลุ่มคือ กลุ่มข้อมูลวงกลม กลุ่มข้อมูลสามเหลี่ยม และกลุ่มข้อมูลสี่เหลี่ยม

4.1 การวัดระยะห่างหรือค่าความคล้าย

ในการจัดกลุ่มข้อมูลนั้นจะมีการวัดระยะห่างของกลุ่มข้อมูลที่ละคู่ เมื่อมีจำนวนข้อมูลเท่ากับ n จะต้องหาระยะห่างของกลุ่มข้อมูลจำนวน nC_2 คู่ ซึ่งจะแบ่งวิธีการวัดระยะห่างตามชนิดของข้อมูลได้เป็น 2 ประเภท คือ ข้อมูลเชิงปริมาณ และ ข้อมูลเชิงคุณภาพ ดังนี้

กำหนดให้	X_{ia}, X_{ja}	แทนข้อมูลตัวแปรที่ a หน่วยที่ i และ j ตามลำดับ
		เมื่อ $i, j = 1, 2, \dots, n$
	$E(X_{ia}), E(X_{ja})$	แทนค่าคาดหวังของข้อมูลตัวแปรที่ a หน่วยที่ i และ j ตามลำดับ
	Z_{ia}, Z_{ja}	แทนค่ามาตรฐานของข้อมูลตัวแปรที่ a หน่วยที่ i และ j ตามลำดับ ซึ่ง $Z_i = \frac{x_i - \bar{x}}{\hat{\sigma}_x}$
	x_i	แทนข้อมูลหน่วยที่ i เมื่อ $i = 1, 2, \dots, n$
	$\bar{x}$	แทนค่าเฉลี่ยของข้อมูล
	$\hat{\sigma}_x$	แทนค่าประมาณส่วนเบี่ยงเบนมาตรฐานของข้อมูล
	D_{ij}	แทนระยะห่างระหว่างข้อมูล
	S_{ij}	แทนค่าความคล้ายของข้อมูล
	p	คือจำนวนตัวแปรทั้งหมดของข้อมูล

4.1.1 การวัดระยะห่างสำหรับข้อมูลเชิงปริมาณ ซึ่งข้อมูลเชิงปริมาณ หมายถึงข้อมูลชนิดอันตรภาคหรือแบบช่วง (Interval) และ ข้อมูลชนิดอัตราส่วน (Ratio) ใช้ในการหาค่าระยะห่างระหว่างข้อมูล 2 หน่วย คือ หน่วยที่ i และหน่วยที่ j สามารถคำนวณหาค่าระยะห่างได้หลายวิธี ได้แก่

1) ค่าระยะห่างแบบยูคลิด (Euclidean Distance) เป็นวิธีมาตรฐานและนิยมใช้มากที่สุดในการหาค่าระยะห่าง คำนวณได้ดังนี้

$$D_{ij} = \sqrt{\sum_{a=1}^p (X_{ia} - X_{ja})^2} \quad (1)$$

2) ค่าระยะห่างแบบยุคลิดกำลังสอง (Square Euclidean Distance) :

$$D_{ij} = \sum_{a=1}^p (X_{ia} - X_{ja})^2$$

3) ค่าระยะห่างแบบซิดีบล็อก (City-block) : $D_{ij} = \sum_{a=1}^p |X_{ia} - X_{ja}|$

4) ค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สัน (Pearson Correlation) คำนวณค่าสัมประสิทธิ์สหสัมพันธ์ของ 2 เวกเตอร์ใช้ในการวัดความคล้าย ได้ดังนี้

$$S_{ij} = \frac{\sum_{a=1}^p Z_{ia} Z_{ja}}{n-1}$$

5) ค่าความคล้ายโคไซน์ (Cosine) : $S_{ij} = \frac{\sum_{a=1}^p X_{ia} X_{ja}}{\sqrt{\left(\sum_{a=1}^p X_{ia}^2\right) \left(\sum_{a=1}^p X_{ja}^2\right)}}$

4.1.2 การวัดระยะห่างสำหรับข้อมูลเชิงคุณภาพ ใช้กับข้อมูลที่อยู่ในรูปความถี่ ได้แก่

1) ค่าระยะห่างไคสแควร์ (Chi-square measure) ใช้หลักการของการทดสอบไคสแควร์ ในการทดสอบความเท่ากันของความถี่ของข้อมูล 2 หน่วย สามารถคำนวณได้ดังนี้

$$D_{ij} = \sqrt{\sum_{a=1}^p \frac{(X_{ia} - E(X_{ia}))^2}{E(X_{ia})} + \sum_{a=1}^p \frac{(X_{ja} - E(X_{ja}))^2}{E(X_{ja})}}$$

2) ค่าระยะห่างไฟสแควร์ (Phi-square measure) :

$$D_{ij} = \sqrt{\frac{\sum_{a=1}^p \frac{(X_{ia} - E(X_{ia}))^2}{E(X_{ia})} + \sum_{a=1}^p \frac{(X_{ja} - E(X_{ja}))^2}{E(X_{ja})}}{n}}$$

เนื่องจากข้อมูลที่นำมาใช้เป็นข้อมูลหลายตัวแปร ดังนั้นการคำนวณระยะห่างหรือค่าความคล้ายเพื่อนำไปใช้ในการวิเคราะห์กลุ่ม ซึ่งสร้างขึ้นในรูปเมทริกซ์ มีขั้นตอนดังนี้

ขั้นที่ 1 : สร้างเมทริกซ์ข้อมูลขนาด $n \times p$ โดยที่ O_i แทนข้อมูลหน่วยที่ i เมื่อ $i = 1, 2, \dots, n$

$$\mathbf{X} = \begin{matrix} & X_1 & X_2 & \cdots & X_p \\ \begin{matrix} O_1 \\ O_2 \\ \vdots \\ O_n \end{matrix} & \begin{bmatrix} & & & \\ & & & \\ & & & \\ & & & \\ & & & \end{bmatrix} \end{matrix}$$

ขั้นที่ 2 : แปลงเมทริกซ์ข้อมูลขนาด $n \times p$ ให้เป็นเมทริกซ์ขนาด $n \times n$ ซึ่งสร้างโดยคำนวณค่าความคล้าย (S) หรือระยะห่าง (D) ของข้อมูลแต่ละหน่วย O_i จากเมทริกซ์ $\mathbf{X}$ โดยที่ S_{ij} = ค่าความคล้ายของข้อมูลหน่วยที่ i และ j ซึ่ง $S_{ii} = 1$ และ D_{ij} = ค่าระยะห่างระหว่างข้อมูลหน่วยที่ i และ j ซึ่ง $D_{ii} = 0$ เมื่อ $i, j = 1, 2, \dots, n$

$$\mathbf{S} = \begin{matrix} & O_1 & O_2 & \cdots & O_n \\ \begin{matrix} O_1 \\ O_2 \\ \vdots \\ O_n \end{matrix} & \begin{bmatrix} S_{11} & S_{12} & \cdots & S_{1n} \\ S_{21} & S_{22} & \cdots & S_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ S_{n1} & S_{n2} & \cdots & S_{nn} \end{bmatrix} \end{matrix} \quad \mathbf{D} = \begin{matrix} & O_1 & O_2 & \cdots & O_n \\ \begin{matrix} O_1 \\ O_2 \\ \vdots \\ O_n \end{matrix} & \begin{bmatrix} D_{11} & D_{12} & \cdots & D_{1n} \\ D_{21} & D_{22} & \cdots & D_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ D_{n1} & D_{n2} & \cdots & D_{nn} \end{bmatrix} \end{matrix}$$

เมื่อได้เมทริกซ์ค่าความคล้าย (S) หรือระยะห่าง (D) ของข้อมูลแล้ว สามารถนำไปใช้กับเทคนิควิธีการแบ่งกลุ่มที่จะกล่าวต่อไป

4.2 เทคนิควิธีการแบ่งกลุ่ม

เทคนิควิธีการแบ่งกลุ่มมีหลากหลายวิธี ซึ่งวิธีที่ผู้วิจัยเลือกมาศึกษามี 5 วิธี ดังนี้

4.2.1 วิธีการจัดกลุ่มแบบลำดับขั้น (Hierarchical clustering method)

4.2.2 อัลกอริทึมจัดกลุ่มแบบเคมีน (K-Means algorithm)

4.2.3 วิธีจัดระบบด้วยตัวเองของโคโฮเนน (Kohonen's Self-Organizing Maps method)

4.2.4 อัลกอริทึมจัดกลุ่มแบบเคเมดอยด์ (K-Medoids clustering algorithm)

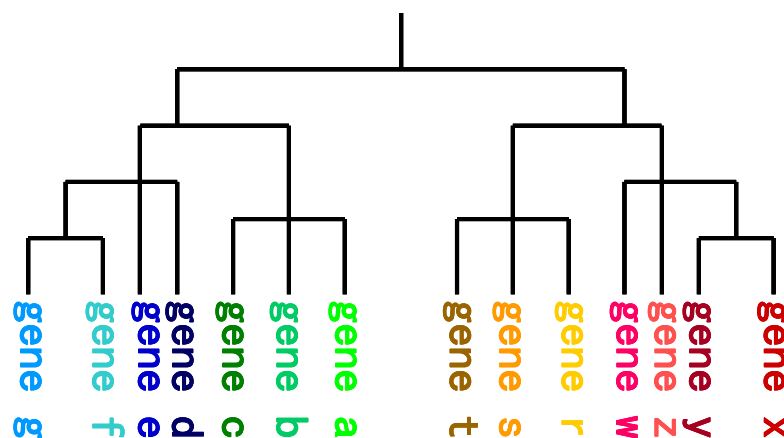
4.2.5 การใช้อัลกอริทึมจัดกลุ่มแบบเคเมดอยด์ ร่วมกับวิธีวัดระยะห่างแบบไดนามิกไทม์วอร์ปปีง (K-Medoids clustering algorithm integrated with Dynamic Time Warping distance measure)

ซึ่งจะกล่าวถึงรายละเอียดของแต่ละเทคนิควิธีในสถิติทดสอบ ดังต่อไปนี้

สถิติทดสอบ

1. วิธีการจัดกลุ่มแบบลำดับขั้น (Hierarchical clustering method)

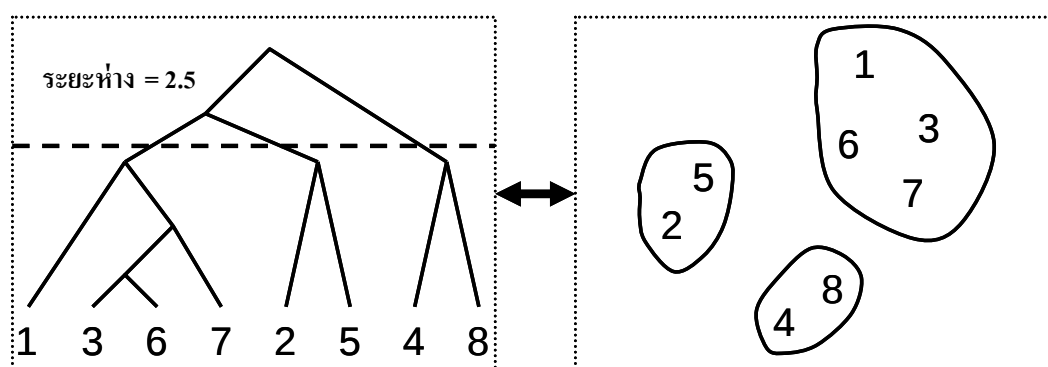
วิธีการจัดกลุ่มแบบลำดับขั้น (Johnson, 1967) เป็นเทคนิควิธีที่นิยมมากที่สุด เนื่องจากคำนวณได้ง่าย ซึ่งสามารถนำไปใช้ในการจัดกลุ่มข้อมูลที่มีคุณลักษณะซ้อน (Nested) เช่น ข้อมูลการแสดงออกของยีน (Gene Expression) (Eisen *et al.*, 1998) แสดงในภาพที่ 2



ภาพที่ 2 การจัดกลุ่มแบบลำดับขั้นของข้อมูลการแสดงออกของยีน

ที่มา: Eisen *et al.* (1998)

การจัดกลุ่มแบบลำดับขั้นสามารถแสดงในรูปแผนภูมิต้นไม้ที่เรียกว่า เดนโดแกรม (Dendrogram) โดยกำหนดจำนวนกลุ่มได้จากการลากเส้นตัดผ่านเดนโดแกรม ด้วยค่าระยะห่างหรือค่าความคล้าย



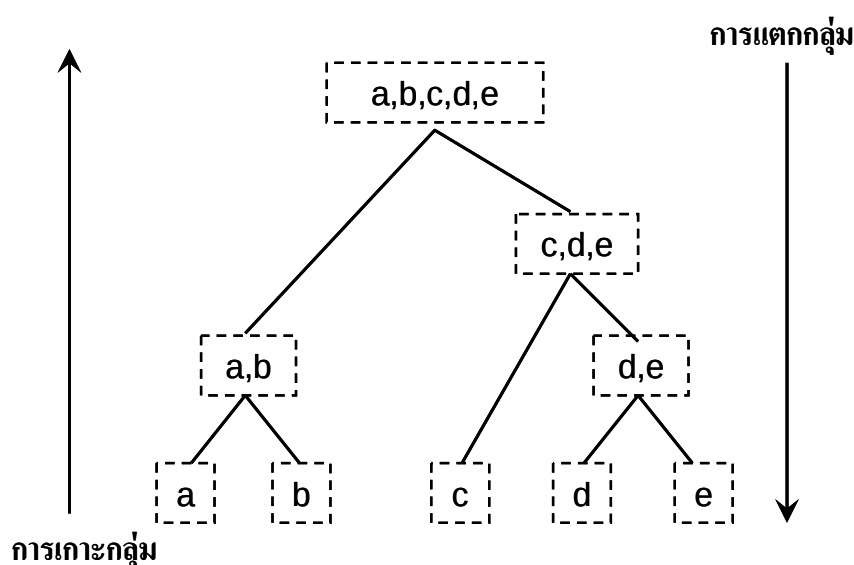
ภาพที่ 3 ตัวอย่างการสร้างและแบ่งกลุ่มแบบลำดับขั้นด้วยเดนโดแกรม

จากภาพที่ 3 ทางด้านซ้าย เมื่อกำหนดระยะห่างเท่ากับ 2.5 ลากเส้นตัดเดนโดแกรม จะสามารถแบ่งกลุ่มข้อมูลได้เป็น 3 กลุ่ม ดังแสดงในภาพที่ 3 ทางด้านขวา

การจัดกลุ่มแบบลำดับขั้นสามารถแบ่งออกเป็น 2 ประเภท คือ

1.1 การเกาะกลุ่ม (Agglomerative) มีหลักการทำงาน คือ เริ่มจากการแบ่งกลุ่มข้อมูลเป็นจำนวน n กลุ่ม จากข้อมูลทั้งหมด n ตัว ให้เป็นบัพภายนอก (External node) หลังจากนั้นทำการรวมข้อมูลที่มีค่าระยะห่างระหว่าง 2 กลุ่มน้อยที่สุด โดยรวมข้อมูลครั้งละ 2 กลุ่ม ทำซ้ำจนกระทั่งสุดท้ายรวมข้อมูลเหลือเพียง 1 กลุ่ม ซึ่งเป็นบัพราก (Root node)

1.2 การแตกกลุ่ม (Divisive) มีหลักการในการทำงานคล้ายวิธีของการเกาะกลุ่ม แต่มีความแตกต่างกันที่วิธีการแตกกลุ่มนี้ จะทำงานในทิศตรงข้ามกัน กล่าวคือการแตกกลุ่มจะเริ่มจากบัพราก ไปสู่บัพภายนอกหรือจากบนลงล่างนั่นเอง ซึ่งวิธีนี้ไม่เป็นที่นิยม เนื่องจากยุ่งยากและใช้เวลาในการคำนวณนานกว่าวิธีของการเกาะกลุ่ม (Lebart and Rajman, 2000)



ภาพที่ 4 ตัวอย่างการสร้างกลุ่มแบบเกาะกลุ่มและแตกกลุ่ม

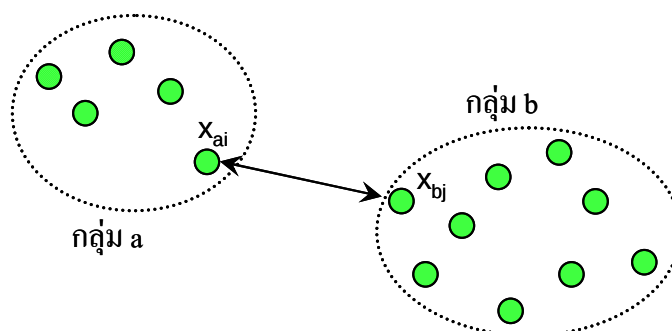
สำหรับขั้นตอนวิธีการคำนวณของการจัดกลุ่มแบบลำดับขั้น โดยใช้การเกาะกลุ่มมีวิธีการดังแสดงในตารางที่ 2

ตารางที่ 2 ขั้นตอนการจัดกลุ่มแบบลำดับขั้นประเภทการเกาะกลุ่ม

ขั้นตอนที่	การจัดกลุ่มแบบลำดับขั้นประเภทการเกาะกลุ่ม
1	แบ่งข้อมูลออกเป็น n กลุ่ม โดยในแต่ละกลุ่มมีสมาชิกเพียง 1 ตัว
2	รวมกลุ่มข้อมูล 2 กลุ่มให้อยู่ในกลุ่มเดียวกัน โดยพิจารณาจากค่าระยะห่าง และใช้หลักเกณฑ์ในการรวมกลุ่มตามที่กำหนด
3	พิจารณาว่าควรรวมข้อมูลกลุ่มที่ 3 เข้ากับข้อมูล 2 กลุ่มแรกหรือรวมเข้ากับข้อมูลกลุ่มใหม่อีกกลุ่มหนึ่ง โดยการพิจารณาจากค่าระยะห่างและใช้หลักเกณฑ์ในการรวมกลุ่มเช่นเดียวกับขั้นตอนที่ 2
4	ทำซ้ำขั้นตอนที่ 3 ไปเรื่อย ๆ จนกระทั่งข้อมูลทุกกลุ่มอยู่ในกลุ่มเดียวกัน นั่นคือสุดท้ายเหลือข้อมูลเพียง 1 กลุ่ม และกลุ่มข้อมูลใดที่ถูกจัดกลุ่มแล้วจะไม่มีการเปลี่ยนแปลงอีก
5	สร้างเดนไดรแกรม
6	กำหนดค่าระยะห่าง แล้วนำมามาลากเส้นตัดผ่านเดนไดรแกรมเพื่อแบ่งกลุ่มข้อมูล

หลักเกณฑ์ในการแบ่งกลุ่มสำหรับวิธีการเกาะกลุ่มนั้นมีหลายวิธี ที่เลือกมาใช้ได้แก่

1.1.1 Single Linkage คือ การหาระยะห่างระหว่าง 2 กลุ่มข้อมูลใด ๆ และรวม 2 กลุ่มข้อมูลที่มีระยะห่างที่ใกล้กันที่สุด โดยเลือกใช้ระยะห่างระหว่างข้อมูลหน่วยที่ใกล้กันที่สุดจากแต่ละกลุ่มมาพิจารณา

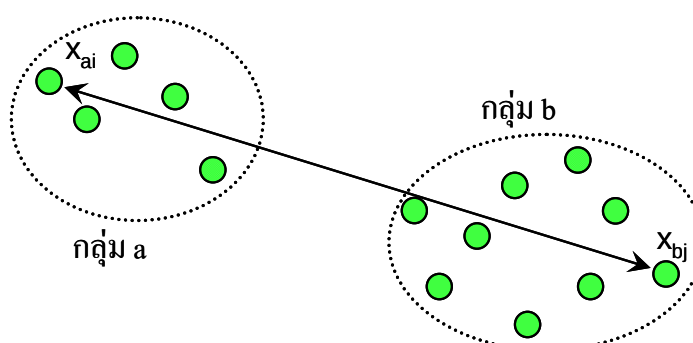


ภาพที่ 5 การรวมกลุ่มแบบ Single Linkage

$$\text{คำนวณโดยใช้สมการดังนี้} \quad d(a,b) = \min(\text{dist}(x_{ai}, x_{bj})) \quad (2)$$

โดยที่ x_{ai} แทนข้อมูลหน่วยที่ i กลุ่ม a
 x_{bj} แทนข้อมูลหน่วยที่ j กลุ่ม b

1.1.2 Complete Linkage วิธีนี้จะตรงข้ามกับวิธี Single Linkage นั่นคือ เลือกใช้ระยะห่างระหว่างข้อมูลหน่วยที่ไกลกันที่สุดจากแต่ละกลุ่มมาพิจารณา แล้วจึงรวม 2 กลุ่มข้อมูลที่มีระยะห่างใกล้กันที่สุด แต่วิธีนี้มีสิ่งรบกวน (Noise) มาก หรือ อาจมีข้อมูลที่ผิดปกติ (Outlier) ทำให้การคำนวณใช้เวลานาน



ภาพที่ 6 การรวมกลุ่มแบบ Complete Linkage

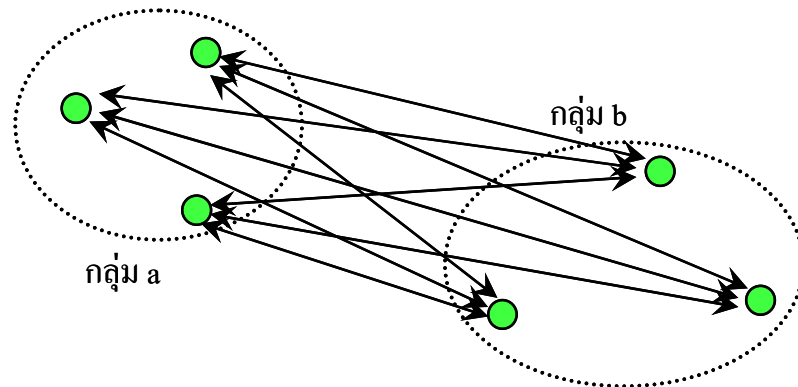
คำนวณโดยใช้สมการดังนี้ $d(a,b) = \max(\text{dist}(x_{ai}, x_{bj}))$ (3)

โดยที่ x_{ai} แทนข้อมูลหน่วยที่ i กลุ่ม a
 x_{bj} แทนข้อมูลหน่วยที่ j กลุ่ม b

1.1.3 Average Linkage Between Groups คือ การหาค่าระยะห่างเฉลี่ยของทุกคู่ข้อมูลระหว่างกลุ่มข้อมูล 2 กลุ่มใด ๆ โดยรวม 2 กลุ่มข้อมูลที่มีระยะห่างเฉลี่ยใกล้กันที่สุด

คำนวณโดยใช้สมการดังนี้ $d(a,b) = \frac{1}{n_a n_b} \sum_{i=1}^{n_a} \sum_{j=1}^{n_b} \text{dist}(x_{ai}, x_{bj})$ (4)

โดยที่ x_{ai} แทนข้อมูลหน่วยที่ i กลุ่ม a
 x_{bj} แทนข้อมูลหน่วยที่ j กลุ่ม b
 n_a แทนจำนวนข้อมูลในกลุ่ม a
 n_b แทนจำนวนข้อมูลในกลุ่ม b



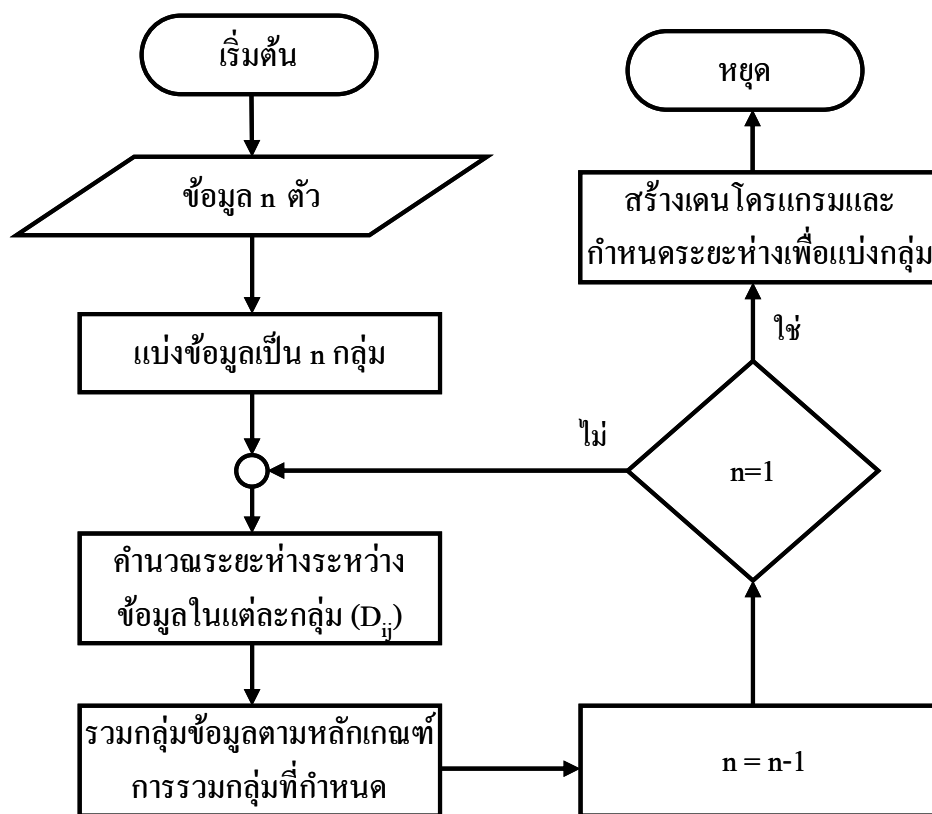
ภาพที่ 7 การรวมกลุ่มแบบ Average Linkage Between Groups

1.1.4 Ward Linkage คือ การหาค่าระยะห่างระหว่างข้อมูลที่อยู่ภายในกลุ่ม แล้วจึงรวม 2 กลุ่มข้อมูลที่ทำให้ค่าระยะห่างของผลรวมกำลังสองภายในกลุ่ม (Sum of square within-cluster distance) เพิ่มขึ้นน้อยที่สุด คำนวณดังนี้

$$d(a,b) = \left(\frac{n_a n_b}{n_a + n_b} \right)^{\frac{1}{2}} \text{dist} \left(\frac{\sum_{i=1}^{n_a} x_{ai}}{n_a}, \frac{\sum_{j=1}^{n_b} x_{bj}}{n_b} \right) \quad (5)$$

โดยที่ x_{ai} แทนข้อมูลหน่วยที่ i กลุ่ม a
 x_{bj} แทนข้อมูลหน่วยที่ j กลุ่ม b
 n_a แทนจำนวนข้อมูลในกลุ่ม a
 n_b แทนจำนวนข้อมูลในกลุ่ม b

และสามารถแสดงผังงานขั้นตอนการจัดกลุ่มข้อมูลแบบลำดับขั้นประเภทการเกาะกลุ่มได้
 ดังนี้



ภาพที่ 8 ขั้นตอนการจัดกลุ่มข้อมูลแบบลำดับขั้นประเภทการเกาะกลุ่ม

ตัวอย่างที่ 1 การแบ่งกลุ่มโดยวิธีการจัดกลุ่มแบบลำดับขั้นประเภทการเกาะกลุ่ม แสดงดังนี้

สมมติมีข้อมูลอยู่ 4 รายการคือ A, B, C และ D ในแต่ละชุดข้อมูลมีตัวแปร 2 ตัวแปร คือ X_1 และ X_2

ตารางที่ 3 ข้อมูลตัวอย่างที่ 1: A, B, C และ D และมีตัวแปรจำนวน 2 ตัวแปร

รายการข้อมูล	ตัวแปร	
	X_1	X_2
A	5	3
B	-1	1
C	1	-2
D	-3	-2

1) จำนวนค่าระยะห่างระหว่างข้อมูลแบบยุคลิด โดยใช้สมการ (1) ดังนี้

$$D_{ij} = \sqrt{\sum_{p=1}^2 (X_{pi} - X_{pj})^2}$$

นำข้อมูลตัวอย่างมาคำนวณตามสมการที่กล่าวมาข้างต้น และสามารถแสดงข้อมูลในรูปแบบเมทริกซ์ $\mathbf{X}$ ขนาด 4×2 ได้ดังนี้

$$\mathbf{X} = \begin{matrix} & \begin{matrix} X_1 & X_2 \end{matrix} \\ \begin{bmatrix} 5 & 3 \\ -1 & 1 \\ 1 & -2 \\ -3 & -2 \end{bmatrix} \end{matrix}$$

จำนวนค่าระยะห่างระหว่างข้อมูลแต่ละคู่จากเมทริกซ์ $\mathbf{X}$ ได้ดังนี้

$$D(A,B) = \sqrt{(5+1)^2 + (3-1)^2} = \sqrt{40} = 6.3246$$

$$D(A,C) = \sqrt{(5-1)^2 + (3+2)^2} = \sqrt{41} = 6.4031$$

$$D(A,D) = \sqrt{(5+3)^2 + (3+2)^2} = \sqrt{89} = 9.4340$$

$$D(B,C) = \sqrt{(-1-1)^2 + (1+2)^2} = \sqrt{13} = 3.6056$$

$$D(B,D) = \sqrt{(-1+3)^2 + (1+2)^2} = \sqrt{13} = 3.6056$$

$$D(C,D) = \sqrt{(1+3)^2 + (-2+2)^2} = \sqrt{16} = 4.0000$$

ถ้า D_{ij} ทั้งหมดที่คำนวณได้ สามารถแสดงในรูปเมทริกซ์ระยะห่าง (**D**) ดังนี้

$$\mathbf{D} = \begin{matrix} & \begin{matrix} A & B & C & D \end{matrix} \\ \begin{bmatrix} 0 & 6.3246 & 6.4031 & 9.4340 \\ 6.3246 & 0 & 3.6056 & 3.6056 \\ 6.4031 & 3.6056 & 0 & 4.0000 \\ 9.4340 & 3.6056 & 4.0000 & 0 \end{bmatrix} & \begin{matrix} A \\ B \\ C \\ D \end{matrix} \end{matrix}$$

2) ทำการแบ่งกลุ่มด้วยวิธีการแบ่งกลุ่มที่เลือกมาใช้ 4 วิธี ได้แก่

2.1) แบ่งกลุ่มแบบ Single Linkage คำนวณโดยใช้สมการ (2) ดังนี้

$$d(a,b) = \min(\text{dist}(x_{ai}, x_{bj}))$$

ครั้งที่ 1 : จากเมทริกซ์ **D** จะเห็นว่า $d(B,C) = d(B,D) = 3.6056$ เป็นสมาชิกที่มีค่าระยะห่างน้อยที่สุด ดังนั้นจึงเลือกจับกลุ่ม B กับ D ก่อน ด้วยระยะห่างเท่ากับ 3.6056

ครั้งที่ 2 : จากเมทริกซ์ **D** เปรียบเทียบค่าระยะห่างได้ดังนี้

$$d((BD),A) = \min(6.3246, 9.4340) = 6.3246$$

$$d((BD),C) = \min(3.6056, 4.0000) = 3.6056$$

$$d(A,C) = 6.4031$$

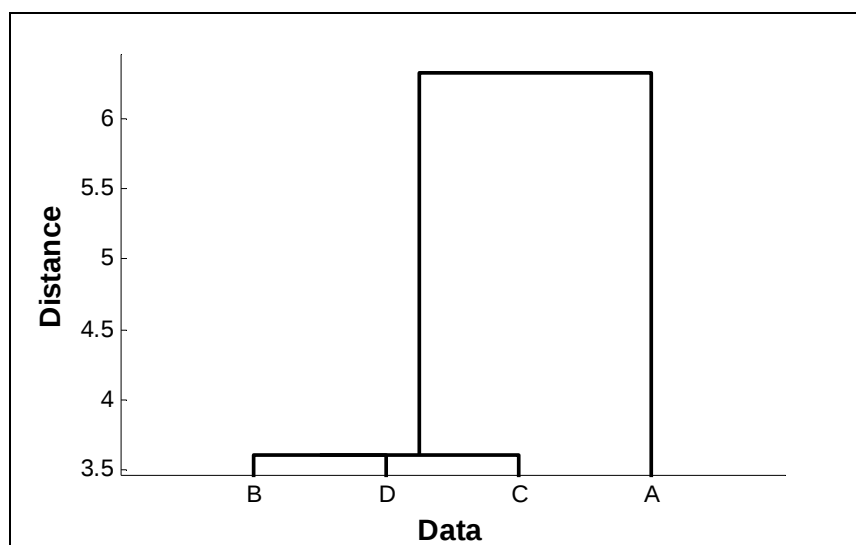
จะเห็นว่า $d((BD),C)$ มีค่าน้อยที่สุด ดังนั้นจึงเลือก C เข้าไปอยู่ในกลุ่ม (BD) ที่ระยะห่างเท่ากับ 3.6056

ครั้งที่ 3 : ทำซ้ำขั้นตอนเดียวกันกับในครั้งที่ 2 พบว่า $d((BCD),A) = \min(6.3246, 6.4031, 9.4340) = 6.3246$ ดังนั้นจึงนำ A เข้าไปอยู่ในกลุ่ม (BCD) ที่ระยะห่างเท่ากับ 6.3246

ตารางที่ 4 การจัดกลุ่มแบบ Single Linkage

ครั้งที่	กลุ่มที่ 1	กลุ่มที่ 2	ระยะห่าง
1	B	D	3.6056
2	C	(BD)	3.6056
3	A	(BCD)	6.3246

จากตารางที่ 4 สรุปได้ว่าในครั้งที่ 1 ข้อมูล B กับ D รวมกลุ่มกันที่ระยะห่างเท่ากับ 3.6056 ในครั้งที่ 2 กลุ่ม C กับ (BD) รวมกันที่ระยะห่างเท่ากับ 3.6056 และในครั้งที่ 3 กลุ่ม A กับ (BCD) รวมกันที่ระยะห่างเท่ากับ 6.3246 เมื่อข้อมูลทุกตัวรวมเป็นกลุ่มเดียวกันแล้วจึงหยุดดำเนินการ สามารถสร้างเป็นแผนภูมิต้นไม้ดังนี้



ภาพที่ 9 แผนภูมิต้นไม้การจับกลุ่มแบบ Single Linkage ของข้อมูลตัวอย่างที่ 1

2.2) แบ่งกลุ่มแบบ Complete Linkage คำนวณโดยใช้สมการ (3) ดังนี้

$$d(a,b) = \max(\text{dist}(x_{ai}, x_{bj}))$$

ครั้งที่ 1 : จากเมทริกซ์ **D** จะเห็นว่า $d(B,C) = d(B,D) = 3.6056$ เป็นสมาชิกที่มีค่าระยะห่างน้อยที่สุด ดังนั้นจะเลือกจับกลุ่ม B กับ D ก่อน ด้วยระยะห่างเท่ากับ 3.6056

ครั้งที่ 2 : จากเมทริกซ์ **D** เปรียบเทียบค่าระยะห่างได้ดังนี้

$$d((BD),A) = \max(6.3246, 9.4340) = 9.4340$$

$$d((BD),C) = \max(3.6056, 4.0000) = 4.0000$$

$$d(A,C) = 6.4031$$

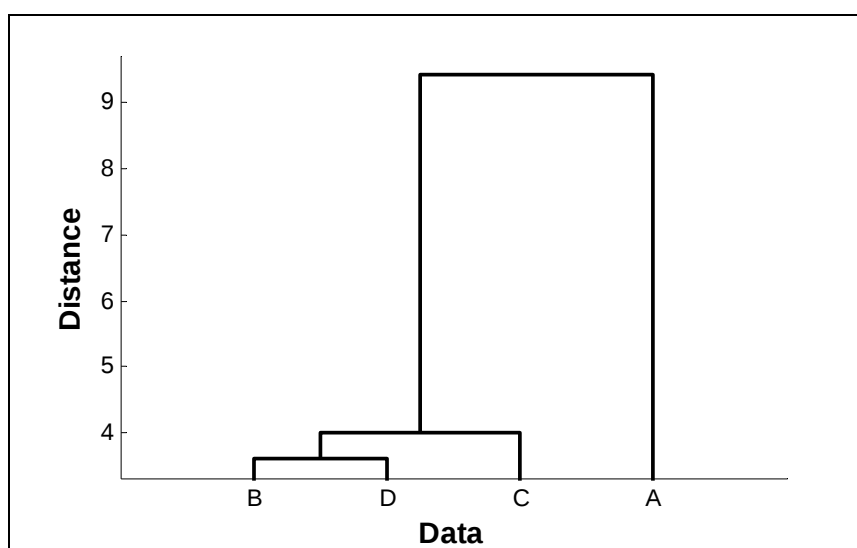
จะเห็นว่า $d((BD),C)$ มีค่าน้อยที่สุด ดังนั้นจึงเลือก C เข้าไปอยู่ในกลุ่ม (BD) ที่ระยะห่างเท่ากับ 4.0000

ครั้งที่ 3 : ทำซ้ำขั้นตอนเดียวกับในครั้งที่ 2 พบว่า $d((BCD), A) = \max(6.3246, 6.4031, 9.4340) = 9.4340$ ดังนั้นจึงนำ A เข้าไปอยู่ในกลุ่ม (BCD) ที่ระยะห่างเท่ากับ 9.4340

ตารางที่ 5 การจัดกลุ่มแบบ Complete Linkage

ครั้งที่	กลุ่มที่ 1	กลุ่มที่ 2	ระยะห่าง
1	B	D	3.6056
2	C	(BD)	4.0000
3	A	(BCD)	9.4340

จากตารางที่ 5 สรุปได้ว่าในครั้งที่ 1 ข้อมูล B กับ D รวมกลุ่มกันที่ระยะห่างเท่ากับ 3.6056 ในครั้งที่ 2 กลุ่ม C กับ (BD) รวมกันที่ระยะห่างเท่ากับ 4.0000 และในครั้งที่ 3 กลุ่ม A กับ (BCD) รวมกันที่ระยะห่างเท่ากับ 9.4340 เมื่อข้อมูลทุกตัวรวมเป็นกลุ่มเดียวกันแล้วจึงหยุดดำเนินการ สามารถสร้างเป็นแผนภูมิต้นไม้ได้ดังนี้



ภาพที่ 10 แผนภูมิต้นไม้การจัดกลุ่มแบบ Complete Linkage ของข้อมูลตัวอย่างที่ 1

2.3) แบ่งกลุ่มแบบ Average Linkage Between Groups คำนวณโดยใช้สมการ (4)

ดังนี้

$$d(a,b) = \frac{1}{n_a n_b} \sum_{i=1}^{n_a} \sum_{j=1}^{n_b} \text{dist}(x_{ai}, x_{bj})$$

ครั้งที่ 1 : จากเมทริกซ์ **D** จะเห็นว่า $d(B,C) = d(B,D) = 3.6056$ เป็นสมาชิกที่มีค่าระยะห่างน้อยที่สุด ดังนั้นจึงเลือกจับกลุ่ม B กับ D ก่อน ด้วยระยะห่างเท่ากับ 3.6056

ครั้งที่ 2 : จากเมทริกซ์ **D** นำมาคำนวณค่าระยะห่างระหว่างกลุ่มได้ดังนี้

$$d((BD),A) = \frac{1}{(2)(1)} (6.3246 + 9.4340) = 7.8793$$

$$d((BD),C) = \frac{1}{(2)(1)} (3.6056 + 4.0000) = 3.8028$$

$$d(A,C) = 6.4031$$

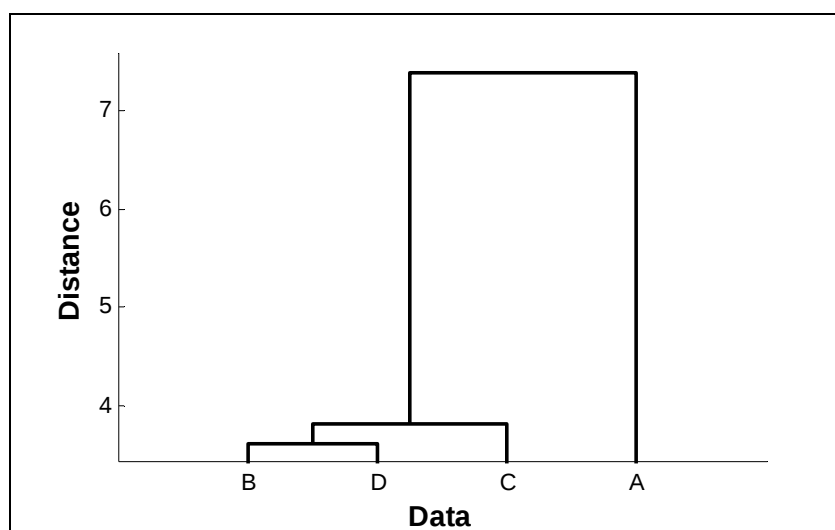
จะเห็นว่า $d((BD),C)$ มีค่าน้อยที่สุด ดังนั้นจึงเลือก C เข้าไปอยู่ในกลุ่ม (BD) ที่ระยะห่างเท่ากับ 3.8028

ครั้งที่ 3 : ทำซ้ำขั้นตอนเดียวกับในครั้งที่ 2 คำนวณค่าระยะห่างระหว่างกลุ่มได้
คือ $d((BCD),A) = \frac{1}{(3)(1)} (6.3246 + 6.4031 + 9.4340) = 7.3872$ ดังนั้นจึงนำ A เข้าไปอยู่ในกลุ่ม (BCD) ที่ระยะห่างเท่ากับ 7.3872

ตารางที่ 6 การจัดกลุ่มแบบ Average Linkage Between Groups

ครั้งที่	กลุ่มที่ 1	กลุ่มที่ 2	ระยะห่าง
1	B	D	3.6056
2	C	(BD)	3.8028
3	A	(BCD)	7.3872

จากตารางที่ 6 สรุปได้ว่า ในครั้งที่ 1 ข้อมูล B กับ D รวมกลุ่มกันที่ระยะห่างเท่ากับ 3.6056 ในครั้งที่ 2 กลุ่ม C กับ (BD) รวมกันที่ระยะห่างเท่ากับ 3.8028 และในครั้งที่ 3 กลุ่ม A กับ (BCD) รวมกันที่ระยะห่างเท่ากับ 7.3872 เมื่อข้อมูลทุกตัวรวมเป็นกลุ่มเดียวกันแล้วจึงหยุดดำเนินการ สามารถสร้างเป็นแผนภูมิต้นไม้ได้ดังนี้



ภาพที่ 11 แผนภูมิต้นไม้การจัดกลุ่มแบบ Average Linkage Between Groups ของข้อมูลตัวอย่างที่ 1

2.4) แบ่งกลุ่มแบบ Ward Linkage คำนวณโดยใช้สมการ (5) ดังนี้

$$d(a,b) = \left(\frac{n_a n_b}{n_a + n_b} \right)^{\frac{1}{2}} \text{dist} \left(\frac{\sum_{i=1}^{n_a} x_{ai}}{n_a}, \frac{\sum_{j=1}^{n_b} x_{bj}}{n_b} \right)$$

ครั้งที่ 1 : เนื่องจากแต่ละกลุ่มมีสมาชิกเริ่มต้นเพียง 1 หน่วยคือ กลุ่ม A, กลุ่ม B, กลุ่ม C และ กลุ่ม D ดังนั้นสามารถใช้ค่าในเมทริกซ์ **D** ซึ่งเป็นค่าระยะห่างระหว่างข้อมูลแต่ละหน่วย แทนระยะห่างระหว่างค่าเฉลี่ยของกลุ่มข้อมูลได้ ซึ่ง $D_{B,C} = D_{B,D} = 3.6056$ เป็นสมาชิกที่มีค่าระยะห่างน้อยที่สุด ดังนั้นจะเลือกจับกลุ่ม B กับ D ก่อน ด้วยระยะห่างเท่ากับ 2.5495 ซึ่งแสดงการคำนวณได้ดังนี้

$$d(B,D) = \sqrt{\frac{(1)(1)}{1+1}} (3.6056) = (0.7071)(3.6056) = 2.5495$$

ครั้งที่ 2 : จากเมทริกซ์ข้อมูล X สามารถคำนวณค่าเฉลี่ยข้อมูล B และ D ได้ดังนี้

$$\text{ค่าเฉลี่ยข้อมูล B และ D ในตัวแปร } X_1 : \frac{(-1-3)}{2} = -2$$

$$\text{ค่าเฉลี่ยข้อมูล B และ D ในตัวแปร } X_2 : \frac{(1-2)}{2} = -0.5$$

ดังนั้นคำนวณค่าระยะห่างระหว่างกลุ่มได้ดังนี้

$$d((BD),A) = \sqrt{\frac{(2)(1)}{2+1} [(-2-5)^2 + (-0.5-3)^2]} = 6.3901$$

$$d((BD),C) = \sqrt{\frac{(2)(1)}{2+1} [(-2-1)^2 + (-0.5+2)^2]} = 2.7386$$

$$d(A,C) = \sqrt{\frac{(1)(1)}{1+1}} (4.0000) = 2.8284$$

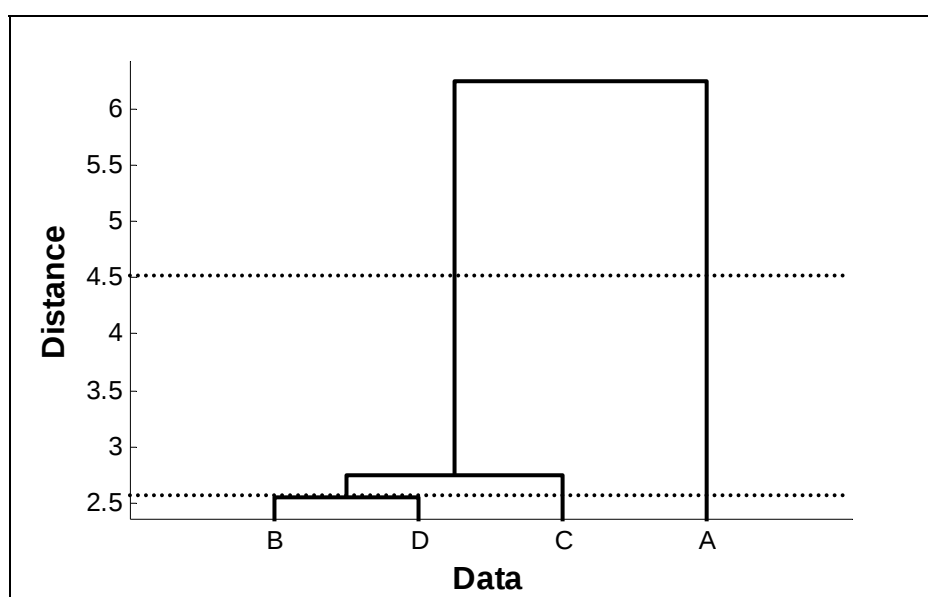
จะเห็นว่า $d((BD),C)$ มีค่าน้อยที่สุด ดังนั้นจึงเลือก C เข้าไปอยู่ในกลุ่ม (BD) ที่ค่าระยะห่างเท่ากับ 2.7386

ครั้งที่ 3 : ทำซ้ำขั้นตอนเดียวกับในครั้งที่ 2 ซึ่งจะได้ว่า A เข้าไปอยู่ในกลุ่ม (BCD) ที่ค่าระยะห่างเท่ากับ 6.2450

ตารางที่ 7 การจัดกลุ่มแบบ Ward Linkage

ครั้งที่	กลุ่มที่ 1	กลุ่มที่ 2	ระยะห่าง
1	B	D	2.5495
2	C	(BD)	2.7386
3	A	(BCD)	6.2450

จากตารางที่ 7 สรุปได้ว่าในครั้งที่ 1 ข้อมูล B กับ D รวมกลุ่มกันที่ระยะห่างเท่ากับ 2.5495 ในครั้งที่ 2 กลุ่ม C กับ (BD) รวมกันที่ระยะห่างเท่ากับ 2.7386 และในครั้งที่ 3 กลุ่ม A กับ (BCD) รวมกันที่ระยะห่างเท่ากับ 6.2450 เมื่อข้อมูลทุกตัวรวมเป็นกลุ่มเดียวกันแล้วจึงหยุดดำเนินการ สามารถสร้างเป็นแผนภูมิต้นไม้ได้ดังนี้



ภาพที่ 12 แผนภูมิต้นไม้การจับกลุ่มแบบ Ward Linkage ของข้อมูลตัวอย่างที่ 1

3) ลากเส้นตัดแกนตั้ง Distance (ค่าระยะห่าง) ของแผนภูมิเพื่อแบ่งกลุ่มข้อมูล เช่น จากภาพที่ 12 กำหนดค่าระยะห่างในการแบ่งกลุ่มเป็น 4.5000 สามารถแบ่งข้อมูลได้ 2 กลุ่มคือ กลุ่ม A และ กลุ่ม (BCD) แต่เมื่อกำหนดระยะห่างในการแบ่งกลุ่มเป็น 2.5500 สามารถแบ่งข้อมูลได้ 3 กลุ่มคือ กลุ่ม A, กลุ่ม C และกลุ่ม (BD)

2. อัลกอริทึมจัดกลุ่มแบบเคมีน (K-Means Clustering Algorithm)

เทคนิควิธีการจัดกลุ่มข้อมูลโดยอัลกอริทึมจัดกลุ่มแบบเคมีน (MacQueen, 1967) สร้างขึ้นในปี ค.ศ. 1956 ซึ่งอัลกอริทึมจัดกลุ่มแบบเคมีนนี้ สมาชิกแต่ละตัวภายในกลุ่มจะมีระยะใกล้จุดศูนย์กลางหรือตัวแทนของกลุ่ม (Centroid) นั้นมากที่สุด ขั้นตอนการจัดกลุ่มข้อมูลโดยใช้ อัลกอริทึมจัดกลุ่มแบบเคมีนประกอบด้วย การกำหนดจำนวนกลุ่มเริ่มต้น การสุ่มตัวแทนกลุ่ม การจัดข้อมูลแต่ละตัวเข้ากลุ่ม และการปรับปรุงตัวแทนกลุ่มในแต่ละกลุ่ม สรุปตามวิธีการได้ 6 ขั้นตอนดังตารางที่ 8

ตารางที่ 8 ขั้นตอนการจัดกลุ่มโดยอัลกอริทึมจัดกลุ่มแบบเคมีน

ขั้นตอนที่	อัลกอริทึมจัดกลุ่มแบบเคมีน
1	กำหนดจำนวนกลุ่ม (k) ที่ต้องการจะแบ่ง
2	สุ่มจุดศูนย์กลางของแต่ละกลุ่ม
3	คำนวณค่าระยะห่างของข้อมูลทุกตัวกับตัวแทนของกลุ่มในแต่ละกลุ่มว่าข้อมูลตัวนั้นสมควรที่จะจัดให้อยู่ในกลุ่มใด โดยพิจารณาจากค่าระยะห่างที่น้อยที่สุด
4	ปรับปรุงจุดศูนย์กลางของข้อมูลแต่ละกลุ่มใหม่ โดยคำนวณค่าเฉลี่ยของข้อมูลภายในกลุ่ม $(\bar{x}_j)$ ดังแสดงในสมการ (6) เพื่อใช้เป็นตัวแทนใหม่ของกลุ่ม
5	ทำขั้นตอนที่ 3 และ 4 ซ้ำ จนกระทั่งสมาชิกทุกตัวในกลุ่มแต่ละกลุ่มจะไม่มีการเปลี่ยนแปลงไปยังกลุ่มอื่น ๆ

ค่าเฉลี่ยของข้อมูลภายในกลุ่ม $(\bar{x}_j)$ ตามขั้นตอนที่ 4 คำนวณได้ดังนี้

$$\bar{x}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} x_{ij} \quad ; i = 1, 2, \dots, n_j ; j = 1, 2, \dots, k \quad (6)$$

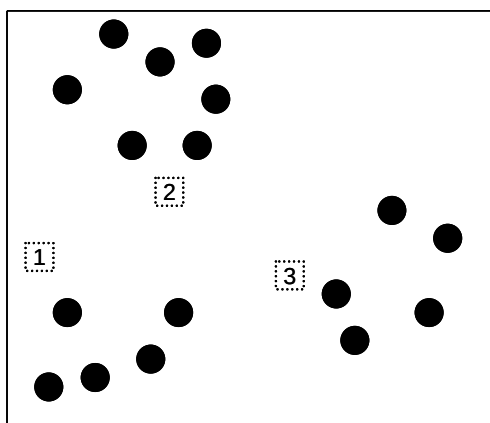
โดยที่ n_j คือ จำนวนข้อมูลในกลุ่มแต่ละครั้งของการแบ่งกลุ่ม

x_{ij} คือ ข้อมูลในหน่วยที่ i กลุ่มที่ j

$\bar{x}_j$ คือ ค่าเฉลี่ยของข้อมูลในกลุ่มที่ j

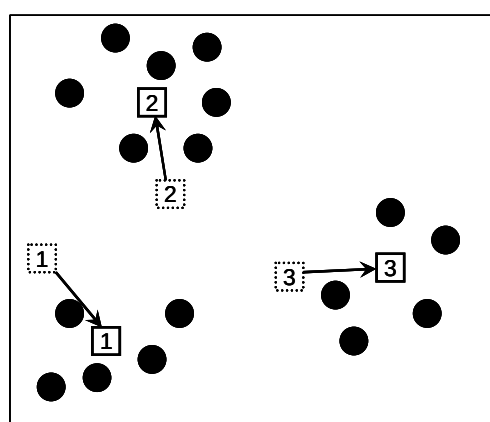
k คือ จำนวนกลุ่มทั้งหมด

ตัวอย่างการแบ่งกลุ่มโดยอัลกอริทึมจัดกลุ่มแบบเคมีน เมื่อแบ่งข้อมูลเป็น 3 กลุ่ม แสดงในภาพที่ 13-15



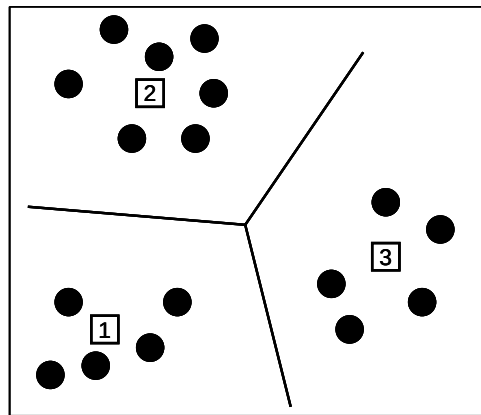
ภาพที่ 13 การกำหนดจุดข้อมูลเพื่อใช้ในแบ่งกลุ่มข้อมูลแบบเคมีน

จากภาพที่ 13 สัญลักษณ์วงกลมสีดำคือข้อมูลใด ๆ สัญลักษณ์สี่เหลี่ยมกรอบไข่ปลาที่มีหมายเลขอยู่ด้านในคือจุดข้อมูลที่สุ่มได้และใช้แทนจุดศูนย์กลางของกลุ่มข้อมูล ซึ่งในกรณีนี้กำหนดจำนวนกลุ่มเป็น 3 กลุ่ม ($k=3$)



ภาพที่ 14 ตัวอย่างการเลื่อนจุดข้อมูล แสดงถึงการเปลี่ยนจุดศูนย์กลางของแต่ละกลุ่มไปยังตำแหน่งที่ครอบคลุมข้อมูลบริเวณใกล้เคียงภายในกลุ่มเดียวกันให้ได้มากที่สุดโดยใช้วิธีการหาค่าเฉลี่ยเลขคณิต (Arithmetic Means)

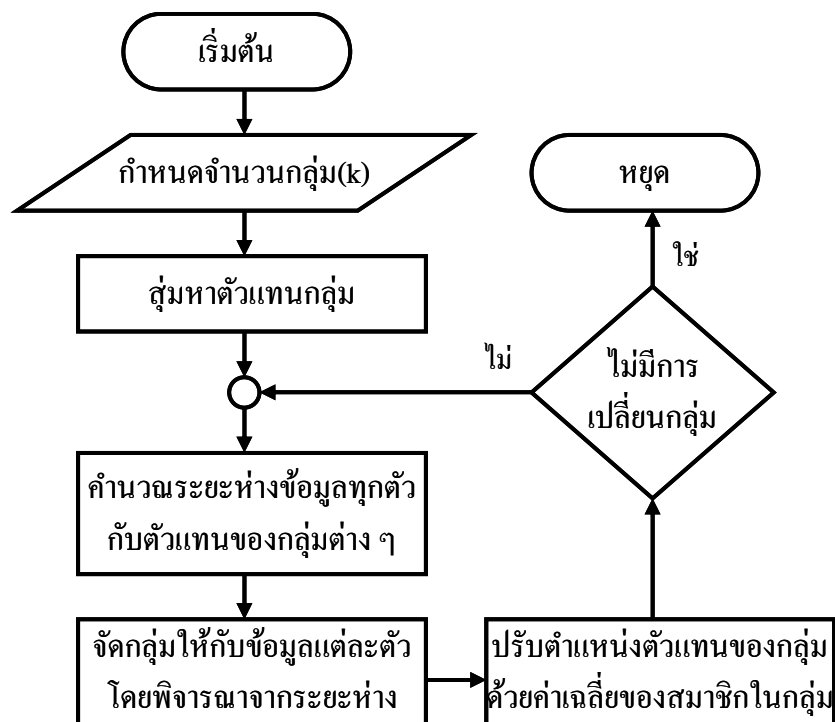
จากภาพที่ 14 สัญลักษณ์วงกลมสีดำคือข้อมูลใด ๆ สัญลักษณ์สี่เหลี่ยมกรอบไข่ปลาที่มีหมายเลขอยู่ด้านในคือจุดข้อมูลที่ทำการสุ่มเพื่อใช้แทนจุดศูนย์กลางของกลุ่มข้อมูล สัญลักษณ์สี่เหลี่ยมกรอบทึบที่มีหมายเลขอยู่ด้านในคือจุดศูนย์กลางของข้อมูลที่ได้รับการปรับให้อยู่ในตำแหน่งที่ครอบคลุมข้อมูลในกลุ่มเดียวกันมากขึ้น



ภาพที่ 15 การแบ่งกลุ่มที่เสร็จสมบูรณ์ โดยจุดศูนย์กลางที่ได้จะอยู่ในตำแหน่งที่เป็นตัวแทนของข้อมูลแต่ละกลุ่ม

จากภาพที่ 15 สัญลักษณ์วงกลมสีดำคือข้อมูลใด ๆ สัญลักษณ์สี่เหลี่ยมกรอบทึบที่มีหมายเลขอยู่ด้านในคือจุดศูนย์กลางของข้อมูลที่เป็นตัวแทนที่เหมาะสมที่สุดของข้อมูลในแต่ละกลุ่ม เส้นขีดแบ่งแสดงการแบ่งกันข้อมูลที่อยู่ต่างกลุ่มแต่ละกลุ่มออกจากกัน

ผังงานขั้นตอนการจัดกลุ่มข้อมูลโดยใช้เคมีนอัลกอริทึมแสดงดังภาพที่ 16



ภาพที่ 16 ขั้นตอนการจัดกลุ่มข้อมูลโดยใช้เคมีนอัลกอริทึม

ตัวอย่างที่ 2 การแบ่งกลุ่มโดยอัลกอริทึมจัดกลุ่มแบบเคมีน แสดงดังนี้

สมมติมีข้อมูลอยู่ 4 รายการคือ A, B, C และ D โดยแต่ละชุดข้อมูลมีตัวแปร 2 ตัวแปร คือ X_1 และ X_2 กำหนดให้มีการแบ่งข้อมูลเป็น 2 กลุ่ม ($k=2$)

ตารางที่ 9 ข้อมูลตัวอย่างที่ 2: A, B, C และ D และมีตัวแปรจำนวน 2 ตัวแปร

รายการข้อมูล	ตัวแปร	
	X_1	X_2
A	5	3
B	-1	1
C	1	-2
D	-3	-2

1) เนื่องจากต้องการแบ่งข้อมูลเป็น 2 กลุ่ม จึงสุ่มเลือกข้อมูลมา 2 จุด กำหนดให้เป็น R1 และ R2 ซึ่งมีค่าอยู่ภายในขอบเขตของข้อมูลตัวอย่าง เพื่อใช้เป็นตัวแทนของแต่ละกลุ่ม แสดงดังตารางที่ 10

ตารางที่ 10 จุดศูนย์กลางเริ่มต้นของข้อมูลแต่ละกลุ่มในตัวอย่างที่ 2

จุดข้อมูล	ตัวแทนของกลุ่ม	
	X_1	X_2
R1	2	2
R2	-1	-2

2) คำนวณค่าระยะห่างแบบยูคลิดระหว่างข้อมูลกับตัวแทนกลุ่ม (D_{ij}) เมื่อมีจำนวนตัวแปรเท่ากับ 2 ตามสมการ (1)

$$D_{ij} = \sqrt{\sum_{p=1}^2 (X_{pi} - X_{pj})^2}$$

คำนวณค่า D_{ij} ของข้อมูลตัวอย่าง ค่าที่ได้จะเป็นระยะห่างระหว่างจุดข้อมูลแต่ละตัวกับตัวแทนกลุ่ม เมื่อพิจารณาค่าระยะห่างที่จุดข้อมูล A คือ $D(A,R1)$ และ $D(A,R2)$ จะเลือกให้ข้อมูลไปอยู่ในกลุ่มเดียวกับตัวแทนกลุ่มที่มีค่าระยะห่างที่คำนวณได้น้อยที่สุด

$$D(A,R1) = \sqrt{(5-2)^2 + (3-2)^2} = \sqrt{10} = 3.16$$

$$D(A,R2) = \sqrt{(5+1)^2 + (3+2)^2} = \sqrt{61} = 7.81$$

จะเห็นว่า ค่าระยะห่างระหว่างจุดข้อมูล A กับ R1 ($D(A,R1)$) มีค่าน้อยกว่าระยะห่างระหว่างจุดข้อมูล A กับ R2 ($D(A,R2)$) ดังนั้น A จึงควรอยู่กลุ่มเดียวกับ R1

3) คำนวณค่าระยะห่างของตัวแทนกลุ่มกับข้อมูลที่เหลือจนครบทั้งหมด

$$D(B,R1) = \sqrt{(-1-2)^2 + (1-2)^2} = 3.16, D(B,R2) = \sqrt{(-1+1)^2 + (1+2)^2} = 3$$

$$D(C,R1) = \sqrt{(1-2)^2 + (-2-2)^2} = 4.12, D(C,R2) = \sqrt{(1+1)^2 + (-2+2)^2} = 2$$

$$D(D,R1) = \sqrt{(-3-2)^2 + (-2-2)^2} = 6.4, D(D,R2) = \sqrt{(-3+1)^2 + (-2+2)^2} = 2$$

จะเห็นว่า ระยะห่างระหว่างข้อมูล B กับ R2 มีค่าน้อยกว่าระยะห่างระหว่างข้อมูล B กับ R1 ดังนั้น B จึงควรอยู่ในกลุ่มเดียวกับ R2 ในทำนองเดียวกันเมื่อพิจารณาค่าระยะห่างของ C และ D กับตัวแทนกลุ่ม ซึ่งพบว่า C และ D ควรอยู่ในกลุ่มเดียวกับ R2 เช่นกัน นั่นคือ B, C และ D จะอยู่ในกลุ่มเดียวกับ R2

4) เมื่อข้อมูลทุกตัวได้รับการกำหนดกลุ่มแล้ว จะทำการปรับเปลี่ยนจุดศูนย์กลางของแต่ละกลุ่ม โดยใช้ค่าเฉลี่ยของสมาชิกภายในกลุ่ม ดังแสดงในตารางที่ 11

ตารางที่ 11 ค่าเฉลี่ยของข้อมูลแต่ละกลุ่ม เมื่อข้อมูลมีการเปลี่ยนแปลงกลุ่ม

กลุ่ม	ค่าเฉลี่ยของข้อมูลในกลุ่ม	
	$\bar{X}_1$	$\bar{X}_2$
A	5	3
(BCD)	$\frac{-1+1-3}{3} = -1$	$\frac{1-2-2}{3} = -1$

5) นำค่าเฉลี่ยของข้อมูลที่ได้มาไว้เป็นตัวแทนกลุ่ม แล้วทำซ้ำขั้นตอนที่ 2-4 จนกว่าข้อมูลจะไม่มีการเปลี่ยนแปลงกลุ่ม โดยสุดท้ายแล้วจะได้ผลดังแสดงในตารางที่ 12

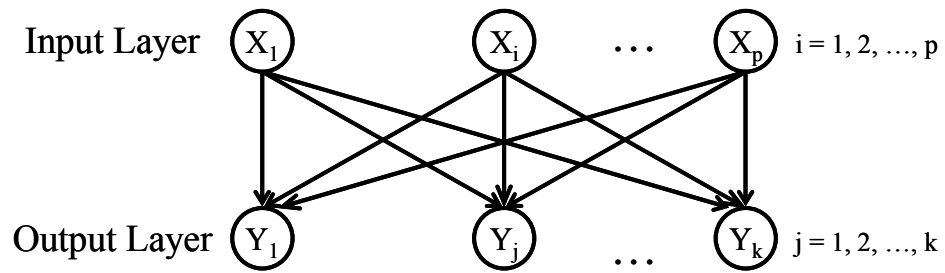
ตารางที่ 12 ระยะห่างระหว่างข้อมูลกับค่าเฉลี่ยของข้อมูลแต่ละกลุ่ม

กลุ่ม	ระยะห่างระหว่างข้อมูลกับค่าเฉลี่ยกลุ่ม			
	A	B	C	D
A	0	6.32	6.40	9.43
(BCD)	7.21	2	2.24	2.24

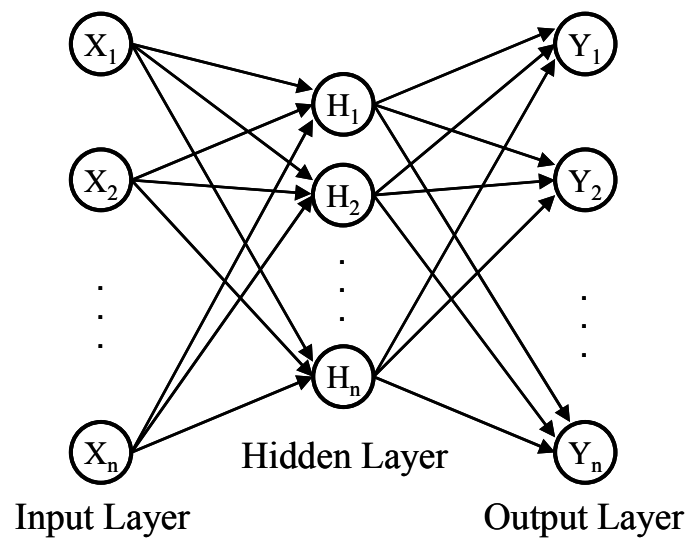
จากตารางที่ 12 พบว่าไม่มีการเปลี่ยนแปลงกลุ่มแล้วจึงหยุดดำเนินการ และสรุปได้ว่าสามารถแบ่งข้อมูลออกเป็น 2 กลุ่ม ดังนี้ กลุ่มที่ 1 คือ A และ กลุ่มที่ 2 คือ (BCD)

3. วิธีจัดระบบด้วยตัวเองของโคโฮเนน (Kohonen's Self-Organizing Maps method--SOM)

วิธีจัดระบบด้วยตัวเองของโคโฮเนน (SOM) (Kohonen, 1988) ได้รับการพัฒนาขึ้นในปี ค.ศ.1982 โดยได้แนวคิดมาจากอัลกอริทึมระบบโครงข่ายประสาทเทียม (Artificial Neural Networks Algorithm) ซึ่งเป็นเทคนิคที่ได้รับความนิยมในสาขาปัญญาประดิษฐ์ (Artificial Intelligent) และมีความสามารถในการเรียนรู้ที่คล้ายคลึงกับระบบสมองของมนุษย์ เช่น ด้านการพยากรณ์ โดยขั้นตอนของการเรียนรู้คือ อาศัยข้อมูลนำเข้าเพื่อสร้างแบบจำลองสำหรับการพยากรณ์ ข้อมูลในอนาคต วิธีจัดระบบด้วยตัวเองของโคโฮเนนเป็นระบบโครงข่ายประสาทเทียมแบบไม่มีชั้นซ่อน (Hidden layer) ดังนั้นจึงมีเพียง 2 ชั้น คือ ชั้นนำเข้าข้อมูล (Input Layer) และชั้นผลลัพธ์ที่ได้ (Output Layer) ดังภาพต่อไปนี้



ภาพที่ 17 การทำงานของระบบโครงข่ายประสาทเทียมแบบ SOM



ภาพที่ 18 ระบบโครงข่ายประสาทเทียมแบบมีชั้นซ่อน

แนวคิดของวิธีจัดระบบด้วยตัวเองของโคโฮเนน คือ การทำซ้ำของข้อมูลในแต่ละชุดข้อมูล (Jain and Dubes, 1998) เพื่อหาค่าของน้ำหนักของข้อมูลที่มีอยู่ทั้งหมดตามจำนวนกลุ่มที่ต้องการแบ่ง ซึ่งขั้นตอนแสดงในตารางที่ 13

ตารางที่ 13 ขั้นตอนการจัดกลุ่มโดยวิธีจัดระบบด้วยตัวเองของโคโฮแน

ขั้นตอนที่	วิธีจัดระบบด้วยตัวเองของโคโฮแน
1	สุ่มน้ำหนัก (w) ให้กับข้อมูล โดยกำหนดให้ค่าความผิดพลาดที่ยอมรับได้ (α) มีค่าอยู่ในช่วง 0 ถึง 1
2	เลือกข้อมูลนำเข้า เพื่อคำนวณค่าระยะห่างของข้อมูลกับน้ำหนัก (D_j) ได้ดังสมการ (7)
3	เปรียบเทียบค่าระยะห่างของข้อมูลกับน้ำหนัก (D_j) ในแต่ละกลุ่มข้อมูลว่าค่าของกลุ่มใดมีค่าน้อยที่สุด
4	นำข้อมูลของกลุ่มที่มีค่าระยะห่างน้อยที่สุดมาคำนวณหาน้ำหนักใหม่ดังสมการ (8)
5	ปรับลดค่า α
6	ทำซ้ำตั้งแต่ขั้นตอนที่ 2 ใหม่ จนกว่าจะไม่มีมีการเปลี่ยนแปลงของค่าน้ำหนัก (w)

สามารถคำนวณระยะห่างระหว่างข้อมูลกับน้ำหนัก ได้ดังนี้

$$D_j = \sqrt{\sum_{i=1}^p (w_{ij} - x_i)^2} \quad ; j = 1, 2, \dots, k \quad (7)$$

โดยที่ w_{ij} คือ น้ำหนักของข้อมูลกลุ่มที่ j ตัวแปรที่ i

x_i คือ ค่าของข้อมูลในตัวแปรที่ i

D_j คือ ระยะทางระหว่างข้อมูลกับค่าน้ำหนัก

p คือ จำนวนตัวแปรทั้งหมด

และ คำนวณหาน้ำหนักใหม่โดยใช้สูตรดังนี้

$$w_{ij}(\text{new}) = w_{ij}(\text{old}) + \alpha(x_i - w_{ij}(\text{old})) \quad (8)$$

โดยที่ x_i คือ ค่าของข้อมูลในตัวแปรที่ i

$w_{ij}(\text{new})$ คือ ค่าน้ำหนักใหม่ของข้อมูลกลุ่มที่ j ตัวแปรที่ i

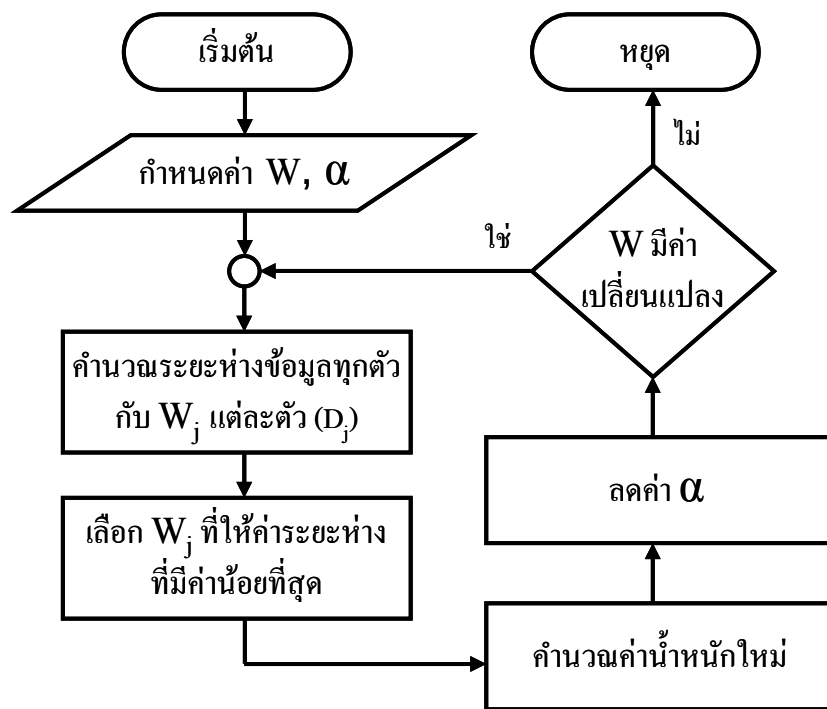
$w_{ij}(\text{old})$ คือ ค่าน้ำหนักเก่าของข้อมูลกลุ่มที่ j ตัวแปรที่ i

α คือ ค่าความผิดพลาดที่ยอมรับได้มีค่าอยู่ในช่วง 0 ถึง 1

ค่าน้ำหนักของข้อมูล (W) สามารถเขียนในรูปเมทริกซ์ได้ดังนี้

$$W = \begin{bmatrix} W_1 & W_2 & \dots & W_k \\ w_{11} & w_{12} & \dots & w_{1k} \\ w_{21} & w_{22} & \dots & w_{2k} \\ \vdots & \vdots & \dots & \vdots \\ w_{p1} & w_{p2} & \dots & w_{pk} \end{bmatrix}_{p \times k}$$

เมื่อ p คือ จำนวนตัวแปร และ k คือ จำนวนกลุ่ม ซึ่ง $W_j = W_1, W_2, \dots, W_k$



ภาพที่ 19 ขั้นตอนการจัดกลุ่มข้อมูลโดยใช้วิธีจัดระบบด้วยตัวเองของโคโฮเนน

ตัวอย่างที่ 3 การจัดกลุ่มข้อมูลโดยใช้วิธีจัดระบบด้วยตัวเองของโคโฮเนน แสดงดังนี้

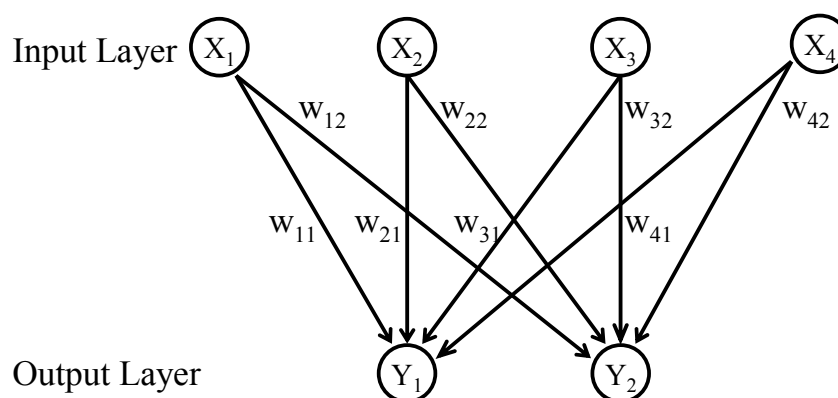
สมมติมีข้อมูลอยู่ 4 รายการคือ A, B, C และ D โดยแต่ละชุดข้อมูลมี 4 ตัวแปรคือ X_1 , X_2 , X_3 และ X_4 กำหนดให้มีการแบ่งข้อมูลเป็น 2 กลุ่ม ($k=2$)

ตารางที่ 14 ข้อมูลตัวอย่างที่ 3: ประกอบด้วยตัวแปร X_1 , X_2 , X_3 และ X_4

รายการข้อมูล	ตัวแปร			
	X_1	X_2	X_3	X_4
A	1	1	0	0
B	0	0	0	1
C	1	0	0	1
D	0	0	1	1

ทำการลดค่า α ลงในทุก ๆ รอบ รอบละ $1/2$ ของค่า α เดิม โดยกำหนดค่า α ในรอบแรกเท่ากับ 0.6 คือ $\alpha(0) = 0.6$

จากค่าที่กำหนดดังตัวอย่าง สามารถเขียนในรูปของโครงข่ายประสาทเทียมที่เป็น Input และ Output ได้ดังต่อไปนี้



ภาพที่ 20 Input และ Output ของโครงข่ายประสาทเทียม

1) สมมติสัมน้ำหนัก (W) ให้แก่ข้อมูลที่จะทำการแบ่งกลุ่ม เมื่อข้อมูลตัวอย่างมี 4 ตัวแปร และแบ่งข้อมูลเป็น 2 กลุ่ม ได้ดังเมทริกซ์ต่อไปนี้

$$\begin{array}{cc} W_1 & W_2 \\ \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \\ w_{31} & w_{32} \\ w_{41} & w_{42} \end{bmatrix} & = \begin{bmatrix} 0.2 & 0.8 \\ 0.6 & 0.4 \\ 0.5 & 0.7 \\ 0.9 & 0.3 \end{bmatrix} \end{array}$$

ค่าน้ำหนักที่สุ่มมาได้สามารถเขียนในรูปเวกเตอร์ W_1 และ W_2 ได้ดังนี้

$$W_1 = (w_{11}, w_{21}, w_{31}, w_{41}) = (0.2, 0.6, 0.5, 0.9)$$

$$W_2 = (w_{12}, w_{22}, w_{32}, w_{42}) = (0.8, 0.4, 0.7, 0.3)$$

2) ทำการคำนวณค่าระยะห่างระหว่างข้อมูลกับน้ำหนัก (D_j) ตามสมการ (7) ดังต่อไปนี้

$$D_{ij} = (w_{1j} - x_1)^2 + (w_{2j} - x_2)^2 + (w_{3j} - x_3)^2 + (w_{4j} - x_4)^2$$

นำข้อมูล A จากตารางที่ 14 คือ $(x_1, x_2, x_3, x_4) = (1, 1, 0, 0)$ มาทำการคำนวณค่าระยะห่าง (D_j) ได้ดังนี้

$$D(W_1, A) = \sqrt{(0.2 - 1)^2 + (0.6 - 1)^2 + (0.5 - 0)^2 + (0.9 - 0)^2} = \sqrt{1.86} = 1.36$$

$$D(W_2, A) = \sqrt{(0.8 - 1)^2 + (0.4 - 1)^2 + (0.7 - 0)^2 + (0.3 - 0)^2} = \sqrt{0.98} = 0.99$$

เมื่อได้ค่า $D(W_1, A)$ และ $D(W_2, A)$ แล้วเลือกระยะห่างที่มีค่าน้อยที่สุดเพื่อคำนวณหาน้ำหนักใหม่ จากผลที่ได้แสดงว่าควรเลือก W_2 มาคำนวณน้ำหนักใหม่ โดยใช้สมการ (8) คือ

$$w_{ij}(\text{new}) = w_{ij}(\text{old}) + \alpha(x_i - w_{ij}(\text{old}))$$

คำนวณค่าได้ดังต่อไปนี้

$$w_{12} = 0.8 + 0.6 (1 - 0.8) = 0.92$$

$$w_{22} = 0.4 + 0.6 (1 - 0.4) = 0.76$$

$$w_{32} = 0.7 + 0.6 (0 - 0.7) = 0.28$$

$$w_{42} = 0.3 + 0.6 (0 - 0.3) = 0.12$$

สามารถเขียนน้ำหนักใหม่ในรูปเมทริกซ์ได้ดังต่อไปนี้

$$\begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \\ w_{31} & w_{32} \\ w_{41} & w_{42} \end{bmatrix} = \begin{bmatrix} 0.2 & 0.92 \\ 0.6 & 0.76 \\ 0.5 & 0.28 \\ 0.9 & 0.12 \end{bmatrix}$$

นำข้อมูล B จากตารางที่ 14 คือ $(x_1, x_2, x_3, x_4) = (0, 0, 0, 1)$ มาคำนวณซ้ำ ด้วยวิธีเดียวกับข้อมูล A ซึ่งจะได้ดังต่อไปนี้

$$D(W_1, B) = 0.81 \text{ และ } D(W_2, B) = 1.51$$

เลือก W_1 เนื่องจาก $D(W_1, B)$ มีค่าน้อยกว่า $D(W_2, B)$ เพื่อทำการคำนวณหาน้ำหนักใหม่ ซึ่งจะได้ดังเมทริกซ์ต่อไปนี้

$$\begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \\ w_{31} & w_{32} \\ w_{41} & w_{42} \end{bmatrix} = \begin{bmatrix} 0.08 & 0.92 \\ 0.24 & 0.76 \\ 0.20 & 0.28 \\ 0.96 & 0.12 \end{bmatrix}$$

นำข้อมูล C จากตารางที่ 14 คือ $(x_1, x_2, x_3, x_4) = (1, 0, 0, 1)$ มาทำการคำนวณซ้ำ ด้วยวิธีเดียวกับข้อมูล A ซึ่งจะได้ดังต่อไปนี้

$$D(W_1, C) = 1.37 \text{ และ } D(W_2, C) = 0.82$$

เลือก W_2 เนื่องจาก $D(W_2, C)$ มีค่าน้อยกว่า $D(W_1, C)$ เพื่อทำการคำนวณหาค่าน้ำหนักใหม่ ซึ่งจะได้ดังเมทริกซ์ต่อไปนี้

$$\begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \\ w_{31} & w_{32} \\ w_{41} & w_{42} \end{bmatrix} = \begin{bmatrix} 0.08 & 0.968 \\ 0.24 & 0.304 \\ 0.20 & 0.112 \\ 0.96 & 0.048 \end{bmatrix}$$

นำค่าจากข้อมูล D จากตารางที่ 14 คือ $(x_1, x_2, x_3, x_4) = (0, 0, 1, 1)$ มาทำการคำนวณซ้ำ ด้วยวิธีเดียวกับข้อมูล A ซึ่งจะได้ดังผลต่อไปนี้

$$D(W_1, D) = 0.84 \text{ และ } D(W_2, D) = 1.65$$

เลือก W_1 เนื่องจาก $D(W_1, D)$ มีค่าน้อยกว่า $D(W_2, D)$ เพื่อทำการคำนวณหาค่าน้ำหนักใหม่ ซึ่งจะได้ดังเมทริกซ์ต่อไปนี้

$$\begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \\ w_{31} & w_{32} \\ w_{41} & w_{42} \end{bmatrix} = \begin{bmatrix} 0.032 & 0.968 \\ 0.096 & 0.304 \\ 0.680 & 0.112 \\ 0.984 & 0.048 \end{bmatrix}$$

3) ปรับลดค่า α ลงครึ่งหนึ่งดังต่อไปนี้

$$\alpha(1) = \alpha(0) / 2 = 0.6 / 2 = 0.3$$

4) หลังจากลดค่า α แล้วทำการคำนวณซ้ำขั้นตอนที่ 2 และ 3 จนกระทั่งค่าน้ำหนัก (W) ไม่มีการเปลี่ยนแปลง ซึ่งหลังจากคำนวณครบ 100 รอบ ได้ผลดังนี้

$$\begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \\ w_{31} & w_{32} \\ w_{41} & w_{42} \end{bmatrix} = \begin{bmatrix} 6.7 \times 10^{-17} & 1 \\ 2 \times 10^{-16} & 0.49 \\ 0.51 & 2.3 \times 10^{-16} \\ 1 & 1 \times 10^{-16} \end{bmatrix}$$

เมื่อปรับค่าน้ำหนัก ($\mathbf{w}$) ให้อยู่ในรูปแบบที่ง่ายขึ้น ได้ผลดังนี้

$$\begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \\ w_{31} & w_{32} \\ w_{41} & w_{42} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0.5 \\ 0.5 & 0 \\ 1 & 0 \end{bmatrix}$$

กลุ่ม1
กลุ่ม2

จากเมทริกซ์น้ำหนัก ($\mathbf{w}$) ที่ได้นี้ จะเห็นว่าสามารถแบ่งข้อมูลได้เป็น 2 กลุ่มอย่างชัดเจน โดยนำค่าน้ำหนักที่ได้มาหาค่าระยะห่างกับข้อมูลเพื่อแบ่งกลุ่ม ซึ่งแสดงตัวอย่างการคำนวณค่าระยะห่างระหว่างน้ำหนักกับข้อมูล A คือ $(x_1, x_2, x_3, x_4) = (1, 1, 0, 0)$ ได้ดังต่อไปนี้

$$D_{A1} = \sqrt{(0-1)^2 + (0-1)^2 + (0.5-0)^2 + (1-0)^2} = \sqrt{3.25} = 1.80$$

$$D_{A2} = \sqrt{(1-1)^2 + (0.5-1)^2 + (0-0)^2 + (0-0)^2} = \sqrt{0.25} = 0.5$$

เนื่องจากค่า D_{A2} มีค่าน้อยกว่า D_{A1} ดังนั้นจึงจัดข้อมูล A อยู่ในกลุ่มที่ 2 และคำนวณค่าระยะห่าง (D_j) เพื่อจัดกลุ่มให้กับข้อมูล B, C และ D ซึ่งผลที่ได้แสดงในตารางที่ 15

ตารางที่ 15 ตัวอย่างกลุ่มข้อมูลที่ได้จากการจัดกลุ่มโดยวิธีใช้ SOM อัลกอริทึม

ข้อมูล	x_1	x_2	x_3	x_4	กลุ่ม
A	1	1	0	0	2
B	0	0	0	0	1
C	1	0	0	1	2
D	0	0	1	1	1

จากข้อมูลในตัวอย่างที่ 3 สรุปได้ว่าสามารถแบ่งข้อมูลออกเป็น 2 กลุ่ม ดังนี้ กลุ่มที่ 1 คือ (AC) และ กลุ่มที่ 2 คือ (BD)

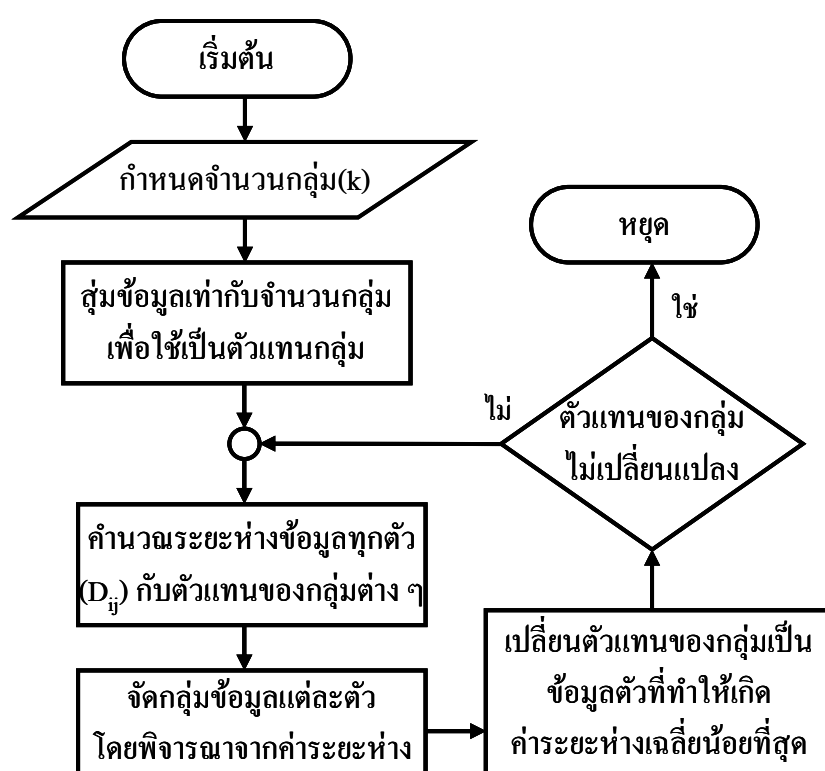
4. อัลกอริทึมจัดกลุ่มแบบเคเมดอยด์ (K-medoids clustering algorithm)

เคเมดอยด์ (Sheng and Liu, 2006) เป็นวิธีการจัดกลุ่มข้อมูลวิธีหนึ่ง ซึ่งมีความคล้ายคลึงกับอัลกอริทึมจัดกลุ่มแบบเคมิน เพราะทั้งสองอัลกอริทึมจะทำการแบ่งข้อมูลทุกหน่วยออกเป็นกลุ่มย่อย โดยมีหลักการคือผลรวมของค่าระยะห่างระหว่างข้อมูลทุกหน่วยภายในกลุ่มนั้นกับจุดศูนย์กลางของหน่วยนั้นจะต้องมีค่าน้อยที่สุด ความแตกต่างระหว่างอัลกอริทึมจัดกลุ่มแบบเคเมดอยด์กับอัลกอริทึมจัดกลุ่มแบบเคมินคือ เคเมดอยด์เลือกข้อมูลตัวใด ๆ ในชุดข้อมูล เพื่อใช้เป็นจุดศูนย์กลางของแต่ละกลุ่ม ซึ่งอัลกอริทึมจัดกลุ่มแบบเคเมดอยด์เป็นวิธีจัดกลุ่มที่สามารถทนทานต่อข้อมูลรบกวน (Noise Data) หรือข้อมูลที่ผิดปกติ (Outlier Data) ได้ดี

เมดอยด์ของแต่ละกลุ่ม คือข้อมูลหน่วยที่ได้รับการกำหนดให้เป็นตัวแทนของกลุ่มนั้น ๆ ซึ่งทำให้ค่าระยะห่างเฉลี่ยของข้อมูลหน่วยอื่น ๆ ภายในกลุ่มกับเมดอยด์มีค่าน้อยที่สุด โดยขั้นตอนการแบ่งกลุ่มโดยอัลกอริทึมจัดกลุ่มแบบเคเมดอยด์แสดงดังตารางที่ 16

ตารางที่ 16 ขั้นตอนการแบ่งกลุ่มโดยอัลกอริทึมจัดกลุ่มแบบเคเมดอยด์

ขั้นตอนที่	อัลกอริทึมจัดกลุ่มแบบเคเมดอยด์
1	กำหนดจำนวนกลุ่ม (k) ที่ต้องการจะแบ่ง
2	สุ่มข้อมูลเพื่อเป็นเมดอยด์ของกลุ่มต่าง ๆ โดยจำนวนข้อมูลที่ทำกรสุ่มเท่ากับจำนวนกลุ่ม
3	คำนวณค่าระยะห่างแบบยุคลิดระหว่างข้อมูลทุกหน่วยกับเมดอยด์แต่ละกลุ่ม (D_{ij}) ตามสมการ (1)
4	กำหนดกลุ่มให้ข้อมูลแต่ละหน่วยในขั้นตอนที่ 3 ให้อยู่ในกลุ่มเดียวกับเมดอยด์ที่ทำให้เกิดค่าระยะห่างที่น้อยที่สุด
5	ปรับเลือกเมดอยด์ของแต่ละกลุ่มใหม่ โดยเลือกข้อมูลหน่วยที่ทำให้เกิดค่าระยะห่างเฉลี่ยที่น้อยที่สุด เมื่อคำนวณค่าระยะห่างระหว่างข้อมูลหน่วยนั้นกับข้อมูลหน่วยอื่นภายในกลุ่ม
6	คำนวณขั้นตอนที่ 3-5 ซ้ำ จนกว่าเมดอยด์ของแต่ละกลุ่มจะไม่มีเปลี่ยนแปลง



ภาพที่ 21 ขั้นตอนการจัดกลุ่มข้อมูลโดยใช้อัลกอริทึมจัดกลุ่มแบบเกมคอยด์

ตัวอย่างที่ 4 การคำนวณการแบ่งกลุ่มโดยอัลกอริทึมจัดกลุ่มแบบเกมคอยด์ แสดงดังนี้

สมมติมีข้อมูลอยู่ 5 รายการคือ A, B, C, D และ E โดยแต่ละชุดข้อมูลมี 2 ตัวแปร คือ X_1 และ X_2 กำหนดให้มีการแบ่งกลุ่มข้อมูลเป็น 2 กลุ่ม ($k=2$)

ตารางที่ 17 ข้อมูลตัวอย่างที่ 4: A, B, C, D และ E และมีตัวแปรจำนวน 2 ตัวแปร

รายการข้อมูล	ตัวแปร	
	X_1	X_2
A	5	3
B	-1	1
C	1	-2
D	-3	-2
E	3	-3

1) เนื่องจากต้องการแบ่งข้อมูลเป็น 2 กลุ่ม ดังนั้นจึงสุ่มเลือกข้อมูลจากข้อมูลตัวอย่าง A, B, C, D และ E มา 2 จุดข้อมูล ซึ่งสุ่มได้จุดข้อมูล A และ D เพื่อใช้เป็นตัวแทนของแต่ละกลุ่ม ค่าที่สุ่มได้แสดงในตารางที่ 18

ตารางที่ 18 จุดศูนย์กลางเริ่มต้นของข้อมูลแต่ละกลุ่มในตัวอย่างที่ 4

จุดข้อมูล	ตัวแทนของกลุ่ม	
	X_1	X_2
A	5	3
D	-3	-2

2) คำนวณค่าระยะห่างแบบยุคลิดระหว่างข้อมูลกับตัวแทนกลุ่มตามสมการ (1) เมื่อข้อมูลมี 2 ตัวแปร ดังนี้

$$D_{ij} = \sqrt{\sum_{p=1}^2 (X_{pi} - X_{pj})^2}$$

คำนวณค่า D_{ij} ของข้อมูลตัวอย่าง ค่าที่ได้จะเป็นระยะห่างระหว่างข้อมูลแต่ละตัวกับตัวแทนกลุ่มคือ A และ D ซึ่งข้อมูลจะไปอยู่ในกลุ่มเดียวกับตัวแทนกลุ่มที่มีค่าระยะห่างน้อยที่สุด

$$D(A,B) = \sqrt{(5+1)^2 + (3-1)^2} = \sqrt{40} = 6.32$$

$$D(D,B) = \sqrt{(-3+1)^2 + (-2-1)^2} = \sqrt{13} = 3.61$$

จะเห็นว่า ระยะห่างระหว่างข้อมูล D กับ B มีค่าน้อยกว่าระยะห่างระหว่างข้อมูล A กับ B ดังนั้น B จึงควรอยู่กลุ่มเดียวกับ D

3) คำนวณค่าระยะห่างระหว่างตัวแทนกลุ่มกับข้อมูลที่เหลือจนครบทั้งหมด

$$D(C,A) = \sqrt{(1-5)^2 + (-2-3)^2} = \sqrt{41} = 6.40$$

$$D(C,D) = \sqrt{(1+3)^2 + (-2+2)^2} = \sqrt{16} = 4.00$$

$$D(E,A) = \sqrt{(3-5)^2 + (-3-3)^2} = \sqrt{40} = 6.32$$

$$D(E,D) = \sqrt{(3+3)^2 + (-3+2)^2} = \sqrt{37} = 6.08$$

จะเห็นว่า ระยะห่างระหว่างข้อมูล C กับ D มีค่าน้อยกว่าระยะห่างระหว่างข้อมูล C กับ A ดังนั้น C จึงควรอยู่ในกลุ่มเดียวกับ D และระยะห่างระหว่างข้อมูล E กับ D มีค่าน้อยกว่าระยะห่างระหว่างข้อมูล E กับ A ดังนั้น E จึงควรอยู่ในกลุ่มเดียวกับ D นั่นคือข้อมูลแบ่งได้เป็น 2 กลุ่มคือ กลุ่ม A และกลุ่ม (BCDE)

4) เมื่อข้อมูลทุกตัวได้รับการกำหนดกลุ่มข้อมูลแล้ว จะทำการปรับเปลี่ยนจุดศูนย์กลางของแต่ละกลุ่ม โดยคำนวณระยะห่างของสมาชิกแต่ละตัวภายในกลุ่ม ดังนี้

$$D(B,C) = \sqrt{(-1-1)^2 + (1+2)^2} = \sqrt{13} = 3.61$$

$$D(B,D) = \sqrt{(-1+3)^2 + (1+2)^2} = \sqrt{13} = 3.61$$

$$D(B,E) = \sqrt{(-1-3)^2 + (1+3)^2} = \sqrt{32} = 5.66$$

$$D(C,D) = \sqrt{(1+3)^2 + (-2+2)^2} = \sqrt{16} = 4.00$$

$$D(C,E) = \sqrt{(1-3)^2 + (-2+3)^2} = \sqrt{5} = 2.24$$

$$D(D,E) = \sqrt{(-3-3)^2 + (-2+3)^2} = \sqrt{37} = 6.08$$

ทำการหาค่าระยะห่างเฉลี่ยของสมาชิกแต่ละตัวภายในกลุ่มที่น้อยที่สุด ดังนี้

$$\text{ระยะห่างเฉลี่ยของ B : } [D(B,C) + D(B,D) + D(B,E)] / 3 = (3.61+3.61+5.66) / 3 = 4.29$$

$$\text{ระยะห่างเฉลี่ยของ C : } [D(B,C) + D(C,D) + D(C,E)] / 3 = (3.61+4.00+2.24) / 3 = 3.28$$

$$\text{ระยะห่างเฉลี่ยของ E : } [D(B,E) + D(C,E) + D(D,E)] / 3 = (5.66+2.24 + 6.08) / 3 = 4.66$$

จะเห็นว่า C มีระยะห่างเฉลี่ยน้อยที่สุด ซึ่งมีระยะห่างเฉลี่ยเท่ากับ 3.28 ดังนั้นทำการปรับจุดศูนย์กลางกลุ่มจาก D เป็น C

5) ให้จุดข้อมูล A และ C เป็นตัวแทนใหม่ของกลุ่ม แล้วทำซ้ำขั้นตอนที่ 2-4 ดังที่กล่าวมาข้างต้น จนกว่าข้อมูลจะไม่มีเปลี่ยนแปลงกลุ่ม โดยสุดท้ายผลที่ได้คือ C ยังคงเป็นจุดศูนย์กลางของกลุ่ม (BCDE) และไม่มีสมาชิกตัวใดย้ายกลุ่ม สามารถสรุปได้ว่าข้อมูลแบ่งออกเป็น 2 กลุ่ม ดังนี้ กลุ่มที่ 1 คือ A และ กลุ่มที่ 2 คือ (BCDE)

5. อัลกอริทึมจัดกลุ่มแบบเคเมดอยด์ร่วมกับวิธีวัดระยะห่างแบบไดนามิกไทม์วอร์ปิง

(K-medoids algorithm integrated with Dynamic Time Warping distance measure--DTW)

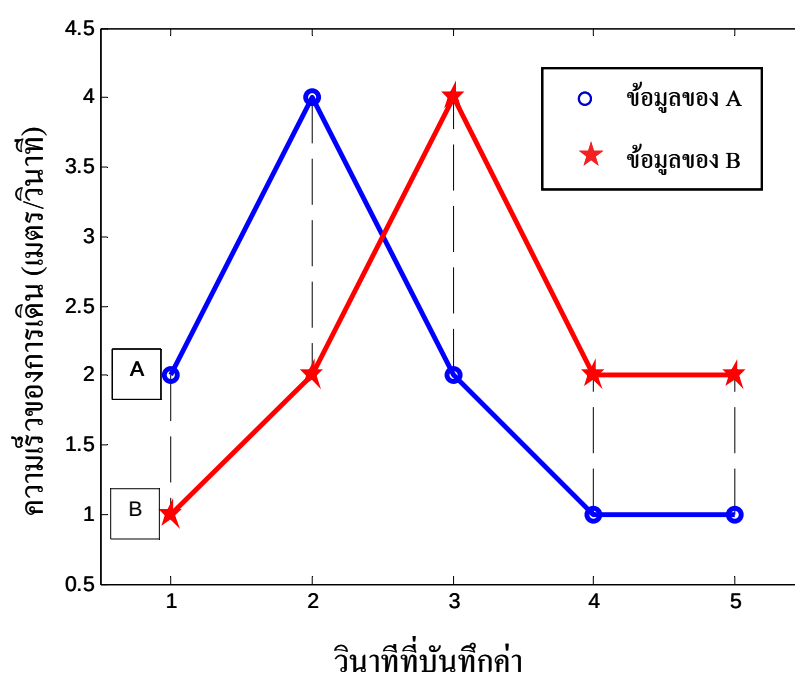
จากงานวิจัยในอดีตแสดงให้เห็นว่าอัลกอริทึมจัดกลุ่มแบบเคเมดอยด์มีประสิทธิภาพมากกว่าอัลกอริทึมจัดกลุ่มแบบเคมีน ในงานวิจัยของ Niennattrakul and Ratanamahatana (2006) ได้พัฒนาความสามารถของอัลกอริทึมจัดกลุ่มแบบเคเมดอยด์ด้วยการนำวิธีวัดระยะห่างแบบไดนามิกไทม์วอร์ปิงมาใช้แทนวิธีวัดระยะห่างแบบเดิม ซึ่งแต่เดิมคือวิธีวัดระยะห่างแบบยุคลิด

ไดนามิกไทม์วอร์ปิงมีการใช้อย่างยาวนานในวงการการรู้จำเสียงพูด (Speech Recognition Society) ซึ่งใช้วัดความคล้ายของเสียงพูดที่มีความเร็วในการพูดที่แตกต่างกัน และเป็นอัลกอริทึมที่ใช้วัดค่าระยะห่างระหว่างข้อมูลสองหน่วย ซึ่งสามารถปรับเปลี่ยนมิติของการคำนวณได้ ตัวอย่างเช่น ข้อมูลความเร็วของการเดินในแต่ละวินาทีของคนสองคนเป็นเวลา 5 วินาที ข้อมูลชุดนี้มี 5 ตัวแปร ตัวแปรแต่ละตัวคือความเร็วของการเดินในวินาทีต่าง ๆ ซึ่งข้อมูลแสดงดังนี้

ตารางที่ 19 ข้อมูลความเร็วของการเดิน 5 วินาทีของนาย A และนาย B

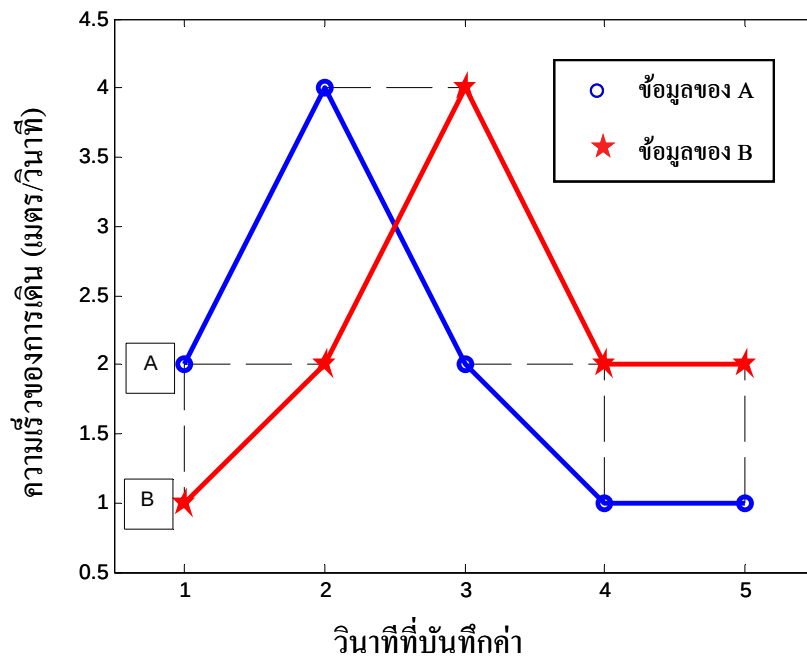
บุคคล	ความเร็วของการเดินในแต่ละวินาที (m/s)				
	วินาทีที่ 1	วินาทีที่ 2	วินาทีที่ 3	วินาทีที่ 4	วินาทีที่ 5
A	2	4	2	1	1
B	1	2	4	2	2

จากตัวอย่างนี้ หากคำนวณค่าระยะห่างระหว่างข้อมูลด้วยวิธีวัดระยะห่างแบบยุคลิด ซึ่งเป็น การคำนวณค่าระยะห่างของข้อมูลในตำแหน่งของตัวแปรที่ตรงกัน คือ (1,1), (2,2), (3,3), (4,4) และ (5,5) จากการคำนวณ ค่าระยะห่างที่ได้คือ 3.3166 การจับคู่ข้อมูลคำนวณตามตำแหน่งของตัวแปร ด้วยวิธีวัดระยะห่างแบบยุคลิดแสดงดังภาพที่ 22 ซึ่งเส้นประแสดงการจับคู่คำนวณระหว่างตัวแปร ซึ่งในที่นี้จะจับคู่คำนวณในตำแหน่งที่ตัวแปรของทั้งสองข้อมูลตรงกัน



ภาพที่ 22 ตำแหน่งที่ใช้ในการคำนวณด้วยวิธีวัดระยะห่างแบบยุคลิด

เมื่อใช้วิธีวัดระยะห่างแบบไดนามิกไทม์วอร์ปปีงด้วยข้อมูลเดียวกันในตารางที่ 19 สามารถ คำนวณรูปแบบการเดินของคนสองคนที่เหมือนกันแต่เกิดในช่วงเวลาที่แตกต่างกันได้ ซึ่งในกรณีนี้ วิธีวัดระยะห่างแบบยุคลิดไม่สามารถคำนวณค่าระยะห่างที่มีความเหมาะสมได้ การคำนวณค่า ระยะห่างของวิธีวัดระยะห่างแบบไดนามิกไทม์วอร์ปปีงนั้นใช้หลักการของกำหนดการพลวัต (Dynamic Programming) ซึ่งจะคำนวณค่าระยะห่างในทุกเส้นทางที่เป็นไปได้ เพื่อให้ได้ค่า ระยะห่างสะสมที่มีค่าน้อยที่สุดภายในเมทริกซ์ระยะห่าง (Distance Matrix) การคำนวณค่าระยะห่าง ระหว่างตัวแปรแต่ละตัวของข้อมูลทั้งสองชุดแสดงดังภาพที่ 23 ซึ่งเส้นประแสดงการจับคู่คำนวณ ระหว่างตัวแปร ซึ่งในที่นี้จะจับคู่คำนวณในตำแหน่งที่ตัวแปรของทั้งสองข้อมูลแตกต่างกัน ในกรณี นี้จะมีการจับคู่คำนวณระหว่างตัวแปรคู่ที่ (1,1), (1,2), (2,3), (3,4), (4,4) และ (5,5)



ภาพที่ 23 ตำแหน่งที่ใช้ในการคำนวณด้วยวิธีวัดระยะห่างแบบไดนามิกไทม์วอร์ปปีง

การคำนวณค่าระยะห่างของวิธีไดนามิกไทม์วอร์ปปีง (DTW(Q,C)) เป็นดังนี้ กำหนดให้ข้อมูลสองรายการคือ $Q = q_1, q_2, \dots, q_p$ และ $C = c_1, c_2, \dots, c_p$ ซึ่ง p แสดงจำนวนตัวแปรของข้อมูล Q และ C การคำนวณค่าระยะห่างแบบไดนามิกไทม์วอร์ปปีง จะใช้เมตริกซ์ระยะห่างที่มีขนาด $p \times p$ สำหรับเก็บค่าระยะห่างระหว่างจุดข้อมูลทั้งสอง ซึ่งคำนวณจากผลรวมของค่าระยะห่างจากเซลล์ (i, j) และเซลล์ข้างเคียงที่มีค่าน้อยที่สุด ดังสมการต่อไปนี้

$$\gamma(i, j) = d(q_i, c_j) + \min\{\gamma(i-1, j), \gamma(i, j-1), \gamma(i-1, j-1)\} \quad (9)$$

โดยที่ $\gamma(i, j)$ คือค่าระยะห่างสะสม ณ เซลล์ (i, j) ส่วน $d(q_i, c_j)$ คือค่าระยะห่างซึ่งคำนวณได้จาก $(q_i - c_j)^2$ และค่าระยะห่างแบบไดนามิกไทม์วอร์ปปีงสามารถคำนวณได้จากผลรวมของค่าระยะห่างบนเส้นทางที่ให้ค่าระยะห่างน้อยที่สุด ดังสมการต่อไปนี้

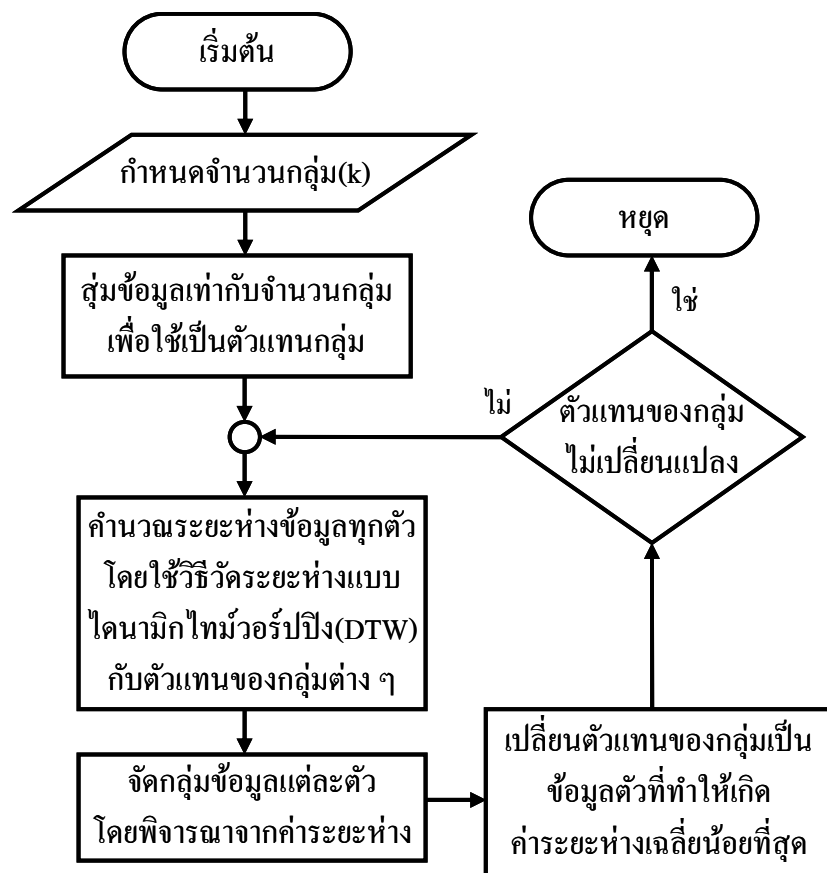
$$DTW(Q,C) = \min_{\forall w \in P} \sqrt{\sum_{l=1}^L d_{w_l}} \quad (10)$$

โดยที่ w คือ เส้นทางการวอร์ปเส้นทางหนึ่งที่มีความยาว L ส่วน P คือเซตของเส้นทางที่เป็นไปได้ทั้งหมด และ d_{w_l} คือ ค่าระยะทางของเส้นทางการวอร์ป w ในลำดับที่ l

การนำวิธีวัดระยะห่างแบบไดนามิกไทม์วอร์ปิงมาใช้กับอัลกอริทึมจัดกลุ่มแบบเคเมคอยด์สามารถทำได้โดยง่าย คือนำวิธีวัดระยะห่างแบบไดนามิกไทม์วอร์ปิงมาใช้ในการหาค่าระยะห่างแทนวิธีวัดระยะห่างแบบยูคลิดในขั้นตอนที่ 3, 4 และ 5 ดังตารางที่ 20

ตารางที่ 20 ขั้นตอนการแบ่งกลุ่มโดยอัลกอริทึมจัดกลุ่มแบบเคเมคอยด์ร่วมกับวิธีวัดระยะห่างแบบไดนามิกไทม์วอร์ปิง

ขั้นตอนที่	อัลกอริทึมจัดกลุ่มแบบเคเมคอยด์ร่วมกับวิธีวัดระยะห่างแบบไดนามิกไทม์วอร์ปิง
1	กำหนดจำนวนกลุ่ม (k) ที่ต้องการจะแบ่ง
2	สุ่มข้อมูลเพื่อเป็นเมคคอยด์ของกลุ่มต่าง ๆ โดยจำนวนข้อมูลที่ทำการสุ่มเท่ากับจำนวนกลุ่ม
3	คำนวณค่าระยะห่างระหว่างข้อมูลทุกหน่วยกับเมคคอยด์แต่ละกลุ่ม โดยนำวิธีวัดระยะห่างแบบไดนามิกไทม์วอร์ปิง (DTW) มาใช้แทนวิธีวัดระยะห่างแบบยูคลิด (D_{ij})
4	กำหนดกลุ่มให้ข้อมูลแต่ละหน่วยในขั้นตอนที่ 3 ให้อยู่ในกลุ่มเดียวกับเมคคอยด์ตัวที่ทำให้เกิดค่าระยะห่างที่น้อยที่สุด
5	ปรับเลือกเมคคอยด์ของแต่ละกลุ่มใหม่ โดยเลือกข้อมูลหน่วยที่ทำให้เกิดค่าระยะห่างเฉลี่ยที่น้อยที่สุด เมื่อคำนวณค่าระยะห่างระหว่างข้อมูลหน่วยนั้นกับข้อมูลหน่วยอื่นภายในกลุ่ม
6	คำนวณขั้นตอนที่ 3-5 ซ้ำ จนกว่าเมคคอยด์ของแต่ละกลุ่มจะไม่มีเปลี่ยนแปลงหรือจนกว่าจะครบรอบของการฝึกสอนที่กำหนดไว้



ภาพที่ 24 ขั้นตอนการจัดกลุ่มข้อมูลโดยใช้อัลกอริทึมจัดกลุ่มแบบเคเมคอยด์ร่วมกับวิธีวัดระยะห่างแบบไดนามิกไทม์วอร์ปปีง

ตัวอย่างที่ 5 การคำนวณการแบ่งกลุ่มโดยอัลกอริทึมจัดกลุ่มแบบเคเมคอยด์ร่วมกับวิธีวัดระยะห่างแบบไดนามิกไทม์วอร์ปปีง แสดงดังนี้

สมมติมีข้อมูลอยู่ 6 รายการคือ A, B, C, D, E และ G โดยแต่ละชุดข้อมูลมี 5 ตัวแปร คือ X_1 , X_2 , X_3 , X_4 และ X_5 กำหนดให้มีการแบ่งกลุ่มข้อมูลเป็น 2 กลุ่ม ($k=2$)

ตารางที่ 21 ข้อมูลตัวอย่างที่ 5: A, B, C, D, E และ G และมีตัวแปรจำนวน 5 ตัวแปร

รายการข้อมูล	ตัวแปร				
	X_1	X_2	X_3	X_4	X_5
A	1	2	1	2	1
B	1	2	1	1	1
C	1	1	1	2	1
D	1	1	2	1	1
E	1	2	3	2	1
G	2	3	4	3	2

1) เนื่องจากต้องการแบ่งข้อมูลเป็น 2 กลุ่ม ดังนั้นจึงสุ่มเลือกข้อมูลจากข้อมูลตัวอย่าง A, B, C, D, E และ G มา 2 จุดข้อมูล สุ่มได้จุดข้อมูล C และ G เพื่อใช้เป็นตัวแทนของแต่ละกลุ่ม ค่าที่สุ่มได้แสดงในตารางที่ 22

ตารางที่ 22 จุดศูนย์กลางเริ่มต้นของข้อมูลแต่ละกลุ่มในตัวอย่างที่ 5

จุดข้อมูล	ตัวแทนของกลุ่ม				
	X_1	X_2	X_3	X_4	X_5
C	1	1	1	2	1
G	2	3	4	3	2

2) คำนวณค่าระยะห่างแบบไดนามิกโทมัส (DTW) ของข้อมูลตัวอย่าง ค่าที่ได้จะเป็นระยะห่างระหว่างข้อมูลแต่ละตัวกับตัวแทนกลุ่มคือ C และ G ซึ่งข้อมูลจะไปอยู่ในกลุ่มเดียวกับตัวแทนกลุ่มที่ค่าระยะห่างที่คำนวณได้น้อยที่สุด

ในที่นี้ ยกตัวอย่างเฉพาะการหาค่า DTW(A,G) โดยคำนวณผลรวมของค่าระยะห่างสะสมตามสมการ (9) ดังนี้

$$\mathcal{V}(i,j) = d(q_i, c_j) + \min\{\mathcal{V}(i-1,j), \mathcal{V}(i,j-1), \mathcal{V}(i-1,j-1)\} \text{ โดยที่ } d(q_i, c_j) = (q_i - c_j)^2$$

คำนวณ $\gamma(i, j)$ ซึ่งเป็นค่าระยะห่างสะสม ณ เซลล์ (i, j) ในทุก ๆ เซลล์ ดังนี้

$$\gamma(1,1) = d(A_1, G_1) = (1-2)^2 = 1$$

$$\gamma(1,2) = d(A_1, G_2) + \gamma(1,1) = (1-3)^2 + 1 = 5$$

$$\gamma(2,1) = d(A_2, G_1) + \gamma(1,1) = (2-2)^2 + 1 = 1$$

$$\gamma(2,2) = d(A_2, G_2) + \min\{\gamma(1,2) + \gamma(1,1) + \gamma(2,1)\} = (2-3)^2 + \min\{5, 1, 1\} = 2$$

$$\gamma(1,3) = d(A_1, G_3) + \gamma(1,2) = (1-4)^2 + 5 = 14$$

$$\gamma(2,3) = d(A_2, G_3) + \min\{\gamma(1,3) + \gamma(1,2) + \gamma(2,2)\} = (2-4)^2 + \min\{14, 5, 2\} = 6$$

$\vdots$

$$\gamma(5,5) = d(A_5, G_5) + \min\{\gamma(4,5) + \gamma(4,4) + \gamma(5,4)\} = (1-2)^2 + \min\{8, 8, 11\} = 9$$

นำผลรวมของค่าระยะห่างสะสมทุก ๆ ค่าที่คำนวณได้ เติมลงในเซลล์ได้ผลดังภาพที่ 25 และหาค่า DTW(A,G) ตามสมการ (10) ดังนี้

$$DTW(Q, C) = \min_{\forall w \in P} \sqrt{\sum_{l=1}^L d_{w_l}}$$

A \ G	2	3	4	3	2
1	1	5	14	18	19
2	1	2	6	7	7
1	2	5	11	10	8
2	2	3	7	8	8
1	3	6	12	11	9

ภาพที่ 25 ผลรวมของค่าระยะห่างสะสมในการคำนวณค่า DTW(A,G)

จากภาพที่ 25 จะได้ว่า $DTW(A,G) = \sqrt{9} = 3$ และสรุปค่า DTW ที่คำนวณได้ทั้งหมดในรูปแบบเมทริกซ์ **DTW** ดังนี้

$$DTW = \begin{bmatrix} 0 & 1 & 1 & 1 & 1.4142 & 3 \\ 1 & 0 & 0 & 0 & 1 & 3.1623 \\ 1 & 0 & 0 & 0 & 1 & 3.1623 \\ 1 & 0 & 0 & 0 & 1 & 3.1623 \\ 1.4142 & 1 & 1 & 1 & 0 & 1.7321 \\ 3 & 3.1623 & 3.1623 & 3.1623 & 1.7321 & 0 \end{bmatrix} \begin{matrix} A \\ B \\ C \\ D \\ E \\ G \end{matrix}$$

3) เปรียบเทียบค่าระยะห่างแบบไดนามิกไทม์วอร์ปิงจากเมทริกซ์ **DTW** โดยมีข้อมูล C และ G เป็นตัวแทนของแต่ละกลุ่ม ดังนี้

$$DTW(C,A) = 1, DTW(G,A) = 3$$

$$DTW(C,B) = 0, DTW(G,B) = 3.1623$$

$$DTW(C,D) = 0, DTW(G,D) = 3.1623$$

$$DTW(C,E) = 1 \text{ , } DTW(G,E) = 1.7321$$

จากค่าระยะห่างระหว่างข้อมูล (DTW) ทุกคู่ จะเห็นว่าระยะห่างระหว่างข้อมูล C กับ A, B, D และ E มีค่าน้อยกว่าระยะห่างระหว่างข้อมูล G กับ A, B, D และ E ดังนั้นสามารถแบ่งข้อมูลได้เป็น 2 กลุ่มคือ กลุ่ม (ABCDE) และ กลุ่ม G

4) เมื่อข้อมูลทุกตัวได้รับการกำหนดกลุ่มแล้ว จะทำการปรับเปลี่ยนจุดศูนย์กลางของแต่ละกลุ่ม ด้วยระยะห่างของสมาชิกแต่ละตัวภายในกลุ่ม ดังนี้

$$DTW(A,B) = 1 \text{ , } DTW(A,C) = 1 \text{ , } DTW(A,D) = 1 \text{ , } DTW(A,E) = 1.4142$$

$$DTW(B,C) = 0 \text{ , } DTW(B,D) = 0 \text{ , } DTW(B,E) = 1 \text{ , } DTW(C,D) = 0$$

$$DTW(C,E) = 1 \text{ , } DTW(D,E) = 1$$

ทำการหาค่าระยะห่างเฉลี่ยของสมาชิกแต่ละตัวภายในกลุ่มที่น้อยที่สุด ดังนี้

$$\text{ระยะห่างเฉลี่ยของ A : } [DTW(A,B) + DTW(A,C) + DTW(A,D) + DTW(A,E)] / 4 = 1.1036$$

$$\text{ระยะห่างเฉลี่ยของ B : } [DTW(A,B) + DTW(B,C) + DTW(B,D) + DTW(B,E)] / 4 = 0.5$$

$$\text{ระยะห่างเฉลี่ยของ C : } [DTW(A,C) + DTW(B,C) + DTW(C,D) + DTW(C,E)] / 4 = 0.5$$

$$\text{ระยะห่างเฉลี่ยของ D : } [DTW(A,D) + DTW(B,D) + DTW(C,D) + DTW(D,E)] / 4 = 0.6036$$

$$\text{ระยะห่างเฉลี่ยของ E : } [DTW(A,E) + DTW(B,E) + DTW(C,E) + DTW(D,E)] / 4 = 1.1036$$

จะเห็นว่า B, C และ D มีระยะห่างเฉลี่ยน้อยที่สุด ซึ่งมีระยะห่างเฉลี่ยเท่ากับ 0.5 ดังนั้นจะเลือกใช้ C เป็นจุดศูนย์กลางของกลุ่มตามเดิม

5) ให้จุดข้อมูล C และ G เป็นตัวแทนกลุ่ม แล้วทำซ้ำขั้นตอนที่ 3 และ 4 ดังที่กล่าวมาข้างต้น จนกว่าข้อมูลจะไม่มีมีการเปลี่ยนแปลงกลุ่ม โดยสุดท้ายผลที่ได้คือ C ยังคงเป็นจุดศูนย์กลาง

ของกลุ่ม (ABCDE) และไม่มีสมาชิกตัวใดย้ายกลุ่ม สามารถสรุปได้ว่าข้อมูลแบ่งออกเป็น 2 กลุ่ม ดังนี้ กลุ่มที่ 1 คือ (ABCDE) และ กลุ่มที่ 2 คือ G

6. การวัดประสิทธิภาพการแบ่งกลุ่มข้อมูล

การวัดประสิทธิภาพของการแบ่งกลุ่มข้อมูล เพื่อเปรียบเทียบเทคนิคการแบ่งกลุ่มทั้ง 5 วิธี ว่าทำได้ดีหรือไม่นั้น สามารถทำได้หลายวิธี ซึ่งผู้วิจัยเลือกนำมาใช้ 2 วิธี ดังนี้

6.1 การวัดค่าความแตกต่างของข้อมูลภายในกลุ่ม (Root Mean Square Standard Deviation: RMSSTD)

โดยค่าความแตกต่างของข้อมูลภายในกลุ่มนั้นแสดงให้เห็นถึงประสิทธิภาพการจัดกลุ่ม ว่ามีมากน้อยเพียงใด ถ้าค่าความแตกต่างภายในกลุ่มน้อย หมายถึงการแบ่งกลุ่มที่ดี (Halkidi *et al.*, 2002)

สามารถคำนวณค่าความแตกต่างภายในกลุ่มได้ดังนี้

$$\text{RMSSTD} = \sqrt{\frac{\sum_{i=1}^k \sum_{j=1}^p (x_{ij} - \bar{x}_{ij})^2}{\sum_{i=1}^k \sum_{j=1}^p (n_{ij} - 1)}}$$

โดยที่

- k คือ จำนวนกลุ่มที่แบ่งได้ทั้งหมด
- p คือ จำนวนตัวแปรอิสระทั้งหมดภายในชุดข้อมูล
- x_{ij} คือ ข้อมูลกลุ่มที่ i ตัวแปรอิสระที่ j
- $\bar{x}_{ij}$ คือ ค่าเฉลี่ยของข้อมูลกลุ่มที่ i ตัวแปรอิสระที่ j
- n_{ij} คือ จำนวนข้อมูลในกลุ่มที่ i ตัวแปรอิสระที่ j

ในงานวิจัยนี้ทำการจำลองข้อมูลจำนวน 1,000 ซ้ำ จึงได้ค่า RMSSTD ทั้งหมด 1,000 ค่า สามารถคำนวณค่าเฉลี่ยของ RMSSTD ได้ดังนี้

$$\text{Average RMSSTD} = \frac{\sum_{i=1}^{1,000} \text{RMSSTD}_i}{1,000}$$

6.2 การวัดค่าความแตกต่างของข้อมูลระหว่างกลุ่ม (R Squared: RS)

โดยการคำนวณค่าความแตกต่างของข้อมูลระหว่างกลุ่มนั้น เป็นวิธีหนึ่งที่จะแสดงให้เห็นถึงประสิทธิภาพการจัดกลุ่มว่ามีมากน้อยเพียงใด ถ้าค่าความแตกต่างระหว่างกลุ่มมาก หมายถึงการแบ่งกลุ่มที่ดี (Halkidi *et al.*, 2002) และข้อมูลแต่ละกลุ่มมีลักษณะที่แตกต่างกัน โดยที่ค่า RS นี้จะอยู่ในช่วง [0,1]

สามารถคำนวณค่าความแตกต่างระหว่างกลุ่มได้ดังนี้

$$RS = \frac{SS_t - SS_w}{SS_t}$$

$$\text{เมื่อ } SS_t = \sum_{i=1}^k \sum_{j=1}^p (x_{ij} - \bar{x}_j)^2 \text{ และ } SS_w = \sum_{i=1}^k \sum_{j=1}^p (x_{ij} - \bar{x}_{ij})^2$$

โดยที่ SS_t คือ ผลรวมของผลต่างกำลังสองของข้อมูลทั้งหมด

SS_w คือ ผลรวมของผลต่างกำลังสองทุกข้อมูลภายในกลุ่ม

k คือ จำนวนกลุ่มที่แบ่งได้ทั้งหมด

p คือ จำนวนตัวแปรอิสระทั้งหมด

x_{ij} คือ ข้อมูลกลุ่มที่ i ตัวแปรอิสระที่ j

$\bar{x}_j$ คือ ค่าเฉลี่ยข้อมูลในตัวแปรอิสระที่ j

$\bar{x}_{ij}$ คือ ค่าเฉลี่ยของข้อมูลกลุ่มที่ i ตัวแปรอิสระที่ j

n_j คือ จำนวนข้อมูลในตัวแปรอิสระที่ j

n_{ij} คือ จำนวนข้อมูลในกลุ่มที่ i ตัวแปรอิสระที่ j

และในการทดลองนี้ทำซ้ำจำนวน 1,000 ครั้ง จึงได้ค่า RS ทั้งหมด 1,000 ค่า ซึ่งสามารถคำนวณค่าเฉลี่ยของ RS ได้ดังนี้

$$\text{Average RS} = \frac{\sum_{i=1}^{1,000} RS_i}{1,000}$$

7. การวัดความแม่นยำในการแบ่งกลุ่ม

สำหรับการศึกษากับข้อมูลจริงที่ทราบว่า ข้อมูลแต่ละค่าอยู่กลุ่มใดมาก่อนแล้ว การทดสอบประสิทธิภาพของเทคนิควิธีการแบ่งกลุ่มสามารถทำได้โดยวัดความแม่นยำในการแบ่งกลุ่มระหว่างกลุ่มที่ได้มาจากข้อมูลที่นำมาศึกษาและกลุ่มที่ได้จากการใช้เทคนิควิธีการแบ่งกลุ่มที่นำมาศึกษาทั้ง 5 วิธี เมื่อ $C_1, C_2, \dots, C_k$ คือกลุ่มข้อมูล k กลุ่ม และ m_i คือจำนวนข้อมูล que เทคนิควิธีการแบ่งกลุ่มมีการจัดกลุ่มได้ตรงกับข้อมูลที่นำมาศึกษาในกลุ่มนั้น ๆ และ n คือจำนวนข้อมูลทั้งหมดในชุดข้อมูลนั้น ๆ ซึ่งสามารถคำนวณค่าความแม่นยำในการแบ่งกลุ่มได้ถูกต้อง ดังนี้

$$\text{Accuracy}(C_1 \dots C_k) = \frac{\sum_{i=1}^k m_i}{n} \times 100$$

ผลงานวิจัยที่เกี่ยวข้อง

Kuo *et al.* (2002) นำเสนอวิธีการผสมผสาน อัลกอริทึมระหว่าง Self-Organizing Maps (SOM) และเคมีน (K-Means) เพื่อนำมาประยุกต์ใช้ในเรื่องของการแบ่งส่วนการตลาด (Market Segmentation) โดยข้อมูลที่ใช้ในงานวิจัย คือ ข้อมูลของลูกค้าที่ซื้อสินค้าในร้านขายปลีก ในประเทศจีน ได้หว่าน ซึ่งงานวิจัยนี้ใช้ปัจจัย (Factor) ในการวิเคราะห์ 2 ส่วน คือข้อมูลส่วนตัวลูกค้า และข้อมูลการซื้อสินค้า โดยทำการเปรียบเทียบผลการทดลองจากวิธีที่นำเสนอ กับ อัลกอริทึมแบบ 2 ขั้นตอน คือ การแบ่งกลุ่มโดยใช้วิธีการจัดกลุ่มแบบลำดับขั้น (Hierarchical cluster) ร่วมกับ K-Means (Conventional 2 stage method) และใช้ SOM ร่วมกับ K-Means โดยใช้ค่า Wilk's Lambda เป็นเกณฑ์วัดผลการวิจัย ผลที่ได้พบว่า วิธีการที่ผู้วิจัยนำเสนอ คือ ใช้ SOM ร่วมกับ K-Means นั้นให้ค่าการแบ่งกลุ่มได้ดีกว่าวิธี Conventional 2 stage method ถึง 2.14 เท่า

Wiwattanacharoenchai and Srivihok (2003) นำเสนองานวิจัยในเชิงการวิเคราะห์ข้อมูลเชิงธุรกิจ โดยใช้วิธีการสืบค้นข้อมูลจากฐานข้อมูล หรือที่เรียกว่า การขุดค้นข้อมูล โดยทำการแบ่งกลุ่มข้อมูลลูกค้าที่ใช้งานอินเทอร์เน็ตแบบคลั่ง ซึ่งทำการเปรียบเทียบอัลกอริทึมที่ใช้ในการแบ่งกลุ่มระหว่างอัลกอริทึมเคมีน (K-Means Algorithm) และ อัลกอริทึมระบบโครงข่ายประสาท (Neural Networks Algorithm) โดยข้อมูลที่ใช้ทำงานวิจัยมาจากคลังข้อมูลของลูกค้าที่ใช้อินเทอร์เน็ตแบบคลั่งของธนาคารพาณิชย์แห่งหนึ่งในประเทศไทย เช่น การทำรายการการชำระค่าบริการการโอนเงิน หรือรายการต่าง ๆ ผลที่ได้จากงานวิจัยทำให้เกิดประโยชน์ทางด้านการวางแผนในการรักษาลูกค้าเดิมและการเพิ่มลูกค้าใหม่ ซึ่งเป็นการนำข้อมูลดังกล่าวมาใช้ให้เกิดประโยชน์และไม่ทำให้เกิดการสูญเสียโอกาสในการทำธุรกิจ โดยใช้ค่าส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation--SD) เป็นเกณฑ์ในการวัดประสิทธิภาพวิธีการแบ่งกลุ่ม และพบว่า K-Means Algorithm แบ่งกลุ่มได้ดีกว่าเมื่อมีจำนวนข้อมูลมาก ในทางกลับกัน Neural Networks Algorithm แบ่งกลุ่มได้ดีกว่าเมื่อข้อมูลมีจำนวนน้อย และใช้กับจำนวนกลุ่มที่แบ่งอยู่ที่ช่วง 7 – 10 กลุ่ม

Niennattrakul and Ratanamahatana (2006) นำเสนองานวิจัยในการศึกษาการจัดกลุ่มข้อมูลมัลติมีเดีย โดยการแปลงให้อยู่ในรูปข้อมูลอนุกรมเวลาและใช้อัลกอริทึมการจัดกลุ่มแบบเคมดอยด์ (K-medoids method) ร่วมกับวิธีวัดระยะห่างที่แตกต่างกัน โดยวิธีวัดระยะห่างที่ใช้คือ วิธีวัดระยะห่างแบบยูคลิด (Euclidean Distance) และวิธีวัดระยะห่างแบบไดนามิกไทม์วอร์ปิง (Dynamic Time Warping distance measure--DTW) โดยเทียบเปอร์เซ็นต์ความถูกต้องกับชุดข้อมูลที่น่ามาศึกษา เพื่อใช้ค่าความแม่นยำในการแบ่งกลุ่มเป็นเกณฑ์ในการทดสอบประสิทธิภาพ และผล

การทดลองที่ได้พบว่าการใช้อัลกอริทึมจัดกลุ่มข้อมูลอนุกรมเวลาร่วมกับวิธีวัดระยะห่างแบบไดนามิกโทมวอร์ปปีงให้ผลการจัดกลุ่มที่ดีกว่า การใช้อัลกอริทึมจัดกลุ่มข้อมูลอนุกรมเวลาร่วมกับวิธีวัดระยะห่างแบบยุคลิด

Niyagas *et al.* (2006) นำเสนองานวิจัยในการศึกษาการทำธุรกรรมบนอินเทอร์เน็ตของลูกค้าสถาบันการเงินในประเทศไทย ซึ่งนำข้อมูลลูกค้าที่ใช้งานอินเทอร์เน็ตแบ่งคั้งในธนาคารพาณิชย์แห่งหนึ่ง ได้แก่ วันและเวลาที่ทำธุรกรรม สถานะและชนิดการทำธุรกรรม ค่า RFM โดยใช้เทคนิคการแบ่งกลุ่มอัลกอริทึมเคมีน (K-Means Algorithm) ร่วมกับวิธีจัดระบบด้วยตัวเองของโคโฮเนน (SOM) และเทคนิควิธีการวิเคราะห์ RFM ทางด้านการตลาด ในการแบ่งกลุ่มลูกค้า และใช้เกณฑ์ในการวัดประสิทธิภาพเทคนิควิธีแบ่งกลุ่ม คือ ค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ค่าความแตกต่างของข้อมูลภายในกลุ่ม (RMSSTD) และค่าความแตกต่างของข้อมูลระหว่างกลุ่ม (RS) และกำหนดช่วงกลุ่มที่ใช้ในการทดลองอยู่ในช่วง 2-10 กลุ่ม ผลที่ได้พบว่าจำนวนกลุ่มที่ดีที่สุดในการแบ่งกลุ่มข้อมูลลูกค้าที่ใช้งานอินเทอร์เน็ตแบ่งคั้ง คือ 8 กลุ่ม

Gelbard *et al.* (2007) ได้เปรียบเทียบวิธีการจัดกลุ่มที่ใช้กันทั่วไป 10 วิธี ได้แก่ วิธีการจัดกลุ่มแบบลำดับขั้น 7 วิธี ประกอบด้วย Average linkage within groups, Average linkage between groups, Single linkage, Complete linkage, Centroid linkage, Median linkage และ Ward linkage วิธีอื่น ๆ อีก 3 วิธีคือ การแบ่งกลุ่มแบบ 2 ขั้นตอน (Two-step), เคมีน (K-means) และ วิธีจัดระบบด้วยตัวเองของโคโฮเนน (SOM) และวิธีที่ผู้วิจัยพัฒนาขึ้นใหม่ 1 วิธีคือ Binary-Positive โดยนำไปใช้ทดสอบกับข้อมูล 4 ชุด ได้แก่ การจำแนกกลุ่มไวน์จากการวิเคราะห์ทางเคมี การแบ่งประเภทของต้นไม้ไริสตามลักษณะของดอก การวิเคราะห์ที่ตั้งของ Ecoli และการทดลองตัวแบบทางฟิสิกส์ในการทรงตัว เมื่อแบ่งกลุ่มโดยวิธีต่าง ๆ ทั้ง 11 วิธี แล้วจึงคำนวณค่าความแม่นยำในการแบ่งกลุ่ม โดยเทียบเปอร์เซ็นต์ความถูกต้องกับชุดข้อมูลที่นำมาศึกษา เพื่อใช้เป็นเกณฑ์ในการทดสอบประสิทธิภาพ ผลที่ได้แสดงว่าวิธี Binary-Positive สามารถแบ่งกลุ่มได้อยู่ในเกณฑ์ดีสำหรับข้อมูล 3 ชุด คือ การจำแนกกลุ่มไวน์จากการวิเคราะห์ทางเคมี การแบ่งประเภทของต้นไม้ไริสตามลักษณะของดอก และการวิเคราะห์ที่ตั้งของ Ecoli ซึ่งเป็นข้อมูลที่เกิดขึ้นจริง แต่ชุดข้อมูลการทดลองตัวแบบทางฟิสิกส์ในการทรงตัว ที่ได้จากการทดลองนั้น มีความคลาดเคลื่อนสูงเมื่อใช้วิธี Binary-Positive ซึ่งอาจมีสาเหตุจากความผิดพลาดทางการทดลอง

อุปกรณ์และวิธีการ

อุปกรณ์

1. เครื่องไมโครคอมพิวเตอร์ ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์
2. โปรแกรม Matlab เวอร์ชัน 7.0

วิธีการ

ในงานวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพการแบ่งกลุ่มข้อมูลโดยใช้เทคนิคการวิเคราะห์กลุ่ม ซึ่งจำลองข้อมูลขึ้นด้วยเทคนิคมอนติคาร์โล โดยโปรแกรม Matlab ซึ่งดำเนินการตามขั้นตอนดังนี้

1. จำลองข้อมูลที่มีการแจกแจงแบบปกติหลายตัวแปร (Multivariate Normal Distribution) โดยกำหนดค่าเฉลี่ยและค่าความแปรปรวนร่วมตามจำนวนตัวแปรอิสระ ดังนี้

- 1.1 จำนวนตัวแปรอิสระ 3 ตัวแปร กำหนดค่าเฉลี่ยและค่าความแปรปรวนร่วม ดังนี้

$$\mu_{31} = \begin{bmatrix} 10 \\ 20 \\ 30 \end{bmatrix}, \mu_{32} = \begin{bmatrix} 100 \\ 200 \\ 300 \end{bmatrix}, \mu_{33} = \begin{bmatrix} 1000 \\ 2000 \\ 3000 \end{bmatrix}$$

$$\Sigma_{31} = \begin{bmatrix} 2 & 3 & 4 \\ 3 & 7 & 8 \\ 4 & 8 & 15 \end{bmatrix}, \Sigma_{32} = \begin{bmatrix} 30 & 35 & 40 \\ 35 & 60 & 70 \\ 40 & 70 & 140 \end{bmatrix}, \Sigma_{33} = \begin{bmatrix} 300 & 350 & 400 \\ 350 & 600 & 700 \\ 400 & 700 & 1400 \end{bmatrix}$$

และจัดเป็นรูปแบบได้เป็น 5 กรณี ได้แก่

1.1.1 รูปแบบที่ 1 การแจกแจงแบบปกติ มีค่าเฉลี่ยเท่ากับ μ_{31} และค่าความแปรปรวนร่วมเท่ากับ Σ_{31}

1.1.2 รูปแบบที่ 2 การแจกแจงแบบปกติ มีค่าเฉลี่ยเท่ากับ μ_{32} และค่าความแปรปรวนร่วมเท่ากับ Σ_{31}

1.1.3 รูปแบบที่ 3 การแจกแจงแบบปกติ มีค่าเฉลี่ยเท่ากับ μ_{32} และค่าความแปรปรวนร่วมเท่ากับ Σ_{32}

1.1.4 รูปแบบที่ 4 การแจกแจงแบบปกติ มีค่าเฉลี่ยเท่ากับ μ_{33} และค่าความแปรปรวนร่วมเท่ากับ Σ_{32}

1.1.5 รูปแบบที่ 5 การแจกแจงแบบปกติ มีค่าเฉลี่ยเท่ากับ μ_{33} และค่าความแปรปรวนร่วมเท่ากับ Σ_{33}

1.2 จำนวนตัวแปรอิสระ 5 ตัวแปร กำหนดค่าเฉลี่ยและค่าความแปรปรวนร่วม ดังนี้

$$\mu_{51} = \begin{bmatrix} 10 \\ 20 \\ 30 \\ 40 \\ 50 \end{bmatrix}, \mu_{52} = \begin{bmatrix} 100 \\ 200 \\ 300 \\ 400 \\ 500 \end{bmatrix}, \mu_{53} = \begin{bmatrix} 1000 \\ 2000 \\ 3000 \\ 4000 \\ 5000 \end{bmatrix}$$

$$\Sigma_{51} = \begin{bmatrix} 2 & 3 & 4 & 5 & 7 \\ 3 & 8 & 10 & 12 & 13 \\ 4 & 10 & 15 & 18 & 21 \\ 5 & 12 & 18 & 25 & 30 \\ 7 & 13 & 21 & 30 & 42 \end{bmatrix}, \Sigma_{52} = \begin{bmatrix} 30 & 35 & 40 & 48 & 52 \\ 35 & 60 & 75 & 81 & 96 \\ 40 & 75 & 140 & 156 & 187 \\ 48 & 81 & 156 & 230 & 290 \\ 52 & 96 & 187 & 290 & 400 \end{bmatrix}$$

$$\Sigma_{53} = \begin{bmatrix} 300 & 350 & 400 & 460 & 520 \\ 350 & 600 & 750 & 830 & 910 \\ 400 & 750 & 1400 & 1650 & 1890 \\ 460 & 830 & 1650 & 2300 & 2900 \\ 520 & 910 & 1890 & 2900 & 4000 \end{bmatrix}$$

และจัดเป็นรูปแบบได้เป็น 5 กรณี ได้แก่

1.2.1 รูปแบบที่ 1 การแจกแจงแบบปกติ มีค่าเฉลี่ยเท่ากับ μ_{51} และค่าความแปรปรวนร่วมเท่ากับ Σ_{51}

1.2.2 รูปแบบที่ 2 การแจกแจงแบบปกติ มีค่าเฉลี่ยเท่ากับ μ_{52} และค่าความแปรปรวนร่วมเท่ากับ Σ_{51}

1.2.3 รูปแบบที่ 3 การแจกแจงแบบปกติ มีค่าเฉลี่ยเท่ากับ μ_{52} และค่าความแปรปรวนร่วมเท่ากับ Σ_{52}

1.2.4 รูปแบบที่ 4 การแจกแจงแบบปกติ มีค่าเฉลี่ยเท่ากับ μ_{53} และค่าความแปรปรวนร่วมเท่ากับ Σ_{52}

1.2.5 รูปแบบที่ 5 การแจกแจงแบบปกติ มีค่าเฉลี่ยเท่ากับ μ_{53} และค่าความแปรปรวนร่วมเท่ากับ Σ_{53}

1.3 จำนวนตัวแปรอิสระ 7 ตัวแปร กำหนดค่าเฉลี่ยและค่าความแปรปรวนร่วม ดังนี้

$$\mu_{71} = \begin{bmatrix} 10 \\ 20 \\ 30 \\ 40 \\ 50 \\ 60 \\ 70 \end{bmatrix}, \mu_{72} = \begin{bmatrix} 100 \\ 200 \\ 300 \\ 400 \\ 500 \\ 600 \\ 700 \end{bmatrix}, \mu_{73} = \begin{bmatrix} 1000 \\ 2000 \\ 3000 \\ 4000 \\ 5000 \\ 6000 \\ 7000 \end{bmatrix}$$

$$\Sigma_{71} = \begin{bmatrix} 2 & 3 & 4 & 5 & 7 & 9 & 10 \\ 3 & 12 & 13 & 15 & 17 & 19 & 21 \\ 4 & 13 & 25 & 27 & 29 & 32 & 34 \\ 5 & 15 & 27 & 38 & 40 & 43 & 47 \\ 7 & 17 & 29 & 40 & 51 & 53 & 58 \\ 9 & 19 & 32 & 43 & 53 & 64 & 69 \\ 10 & 21 & 34 & 47 & 58 & 69 & 75 \end{bmatrix}, \Sigma_{72} = \begin{bmatrix} 30 & 35 & 40 & 46 & 52 & 59 & 63 \\ 35 & 70 & 74 & 80 & 85 & 92 & 98 \\ 40 & 74 & 110 & 120 & 132 & 140 & 161 \\ 46 & 80 & 120 & 180 & 193 & 200 & 208 \\ 52 & 85 & 132 & 193 & 230 & 240 & 254 \\ 59 & 92 & 140 & 200 & 240 & 280 & 310 \\ 63 & 98 & 161 & 208 & 254 & 310 & 400 \end{bmatrix},$$

$$\Sigma_{73} = \begin{bmatrix} 300 & 350 & 400 & 460 & 520 & 590 & 630 \\ 350 & 700 & 740 & 800 & 850 & 920 & 980 \\ 400 & 740 & 1100 & 1200 & 1320 & 1400 & 1610 \\ 460 & 800 & 1200 & 1800 & 1930 & 2000 & 2080 \\ 520 & 850 & 1320 & 1930 & 2300 & 2400 & 2540 \\ 590 & 920 & 1400 & 2000 & 2400 & 2800 & 3100 \\ 630 & 980 & 1610 & 2080 & 2540 & 3100 & 4000 \end{bmatrix}$$

และจัดเป็นรูปแบบได้เป็น 5 กรณี ได้แก่

1.3.1 รูปแบบที่ 1 การแจกแจงแบบปกติ มีค่าเฉลี่ยเท่ากับ μ_{71} และค่าความแปรปรวนรวมเท่ากับ Σ_{71}

1.3.2 รูปแบบที่ 2 การแจกแจงแบบปกติ มีค่าเฉลี่ยเท่ากับ μ_{72} และค่าความแปรปรวนรวมเท่ากับ Σ_{71}

1.3.3 รูปแบบที่ 3 การแจกแจงแบบปกติ มีค่าเฉลี่ยเท่ากับ μ_{72} และค่าความแปรปรวนรวมเท่ากับ Σ_{72}

1.3.4 รูปแบบที่ 4 การแจกแจงแบบปกติ มีค่าเฉลี่ยเท่ากับ μ_{73} และค่าความแปรปรวนรวมเท่ากับ Σ_{72}

1.3.5 รูปแบบที่ 5 การแจกแจงแบบปกติ มีค่าเฉลี่ยเท่ากับ μ_{73} และค่าความแปรปรวนรวมเท่ากับ Σ_{73}

2. กำหนดขนาดตัวอย่างเท่ากับ 25, 100 และ 300

3. จำลองข้อมูลที่มีการแจกแจงแบบปกติหลายตัวแปรตามลักษณะที่กำหนดในขั้นตอนที่ 1 และ 2 จำนวน 1,000 ครั้งในแต่ละชุดข้อมูล

4. ทำการแบ่งกลุ่มข้อมูลที่จำลองขึ้น ตามขนาดตัวอย่างในขั้นตอนที่ 2 เมื่อจำนวนการแบ่งกลุ่มเท่ากับ 2-5, 2-8 และ 2-15 ตามลำดับ โดยใช้เทคนิคการแบ่งกลุ่ม 5 วิธี คือ

4.1 วิธีการจัดกลุ่มแบบลำดับขั้น (Agglomerative Hierarchical clustering) และทำการแบ่งกลุ่มข้อมูลโดยใช้เกณฑ์การแบ่งกลุ่ม 4 วิธี คือ (1) Single Linkage (2) Complete Linkage (3) Average Linkage Between Groups (4) Ward Linkage

4.2 อัลกอริทึมจัดกลุ่มแบบเคมีน (K-Means Algorithm)

4.3 วิธีจัดระบบด้วยตัวเองของโคโฮเนน (SOM)

4.4 อัลกอริทึมจัดกลุ่มแบบเคเมดอยด์ (K-Medoids clustering algorithm)

4.5 อัลกอริทึมจัดกลุ่มแบบเคเมดอยด์ ร่วมกับวิธีวัดระยะห่างแบบไดนามิกไทม์วอร์ปิง (K-Medoids clustering algorithm integrated with Dynamic Time Warping distance measure--DTW)

5. วัดประสิทธิภาพการแบ่งกลุ่มข้อมูล

5.1 คำนวณค่าความแตกต่างของข้อมูลภายในกลุ่ม (RMSSTD) หากค่าความแตกต่างภายในกลุ่มน้อย หมายถึงการแบ่งกลุ่มที่ดี

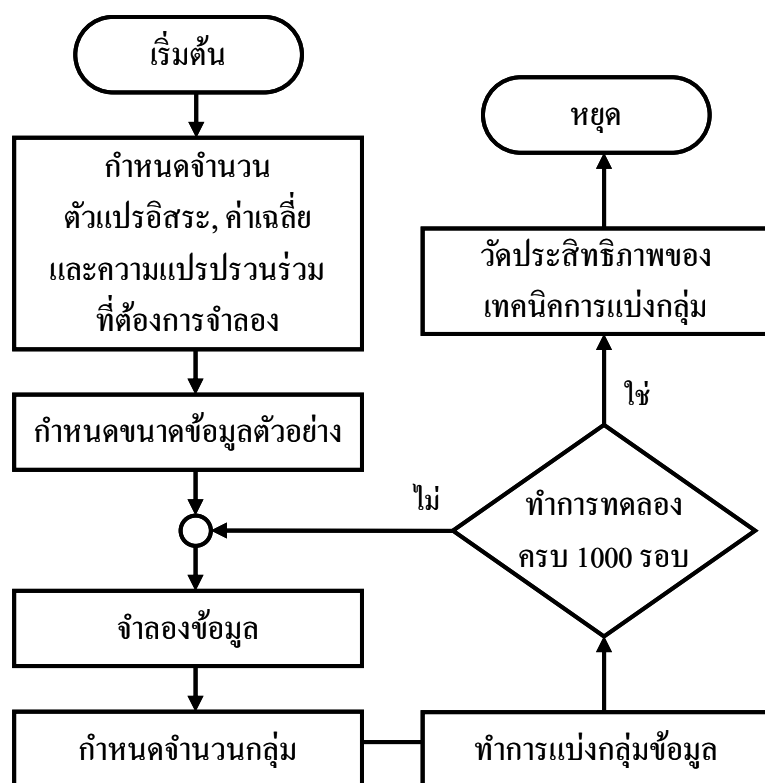
5.2 คำนวณค่าความแตกต่างของข้อมูลระหว่างกลุ่ม (RS) หากค่าความแตกต่างระหว่างกลุ่มมาก หมายถึงการแบ่งกลุ่มที่ดี

6. เปรียบเทียบประสิทธิภาพการแบ่งกลุ่มข้อมูล โดยเปรียบเทียบตามจำนวนกลุ่มที่กำหนดไว้ในช่วง 2-15 กลุ่มและเทคนิคการแบ่งกลุ่มที่ใช้ทั้ง 5 วิธี

7. สรุปผลที่ได้จากการเปรียบเทียบประสิทธิภาพการแบ่งกลุ่มข้อมูล ว่าเทคนิคการแบ่งกลุ่มแบบใดให้ผลการแบ่งกลุ่มดีที่สุด

8. นำข้อมูลจริงมาทดสอบกับเทคนิคการแบ่งกลุ่มทั้ง 5 วิธี เพื่อวัดค่าความแม่นยำในการแบ่งกลุ่ม แล้วนำมาเปรียบเทียบกับผลที่ได้จากข้อมูลจำลอง

ขั้นตอนวิธีการทดลองที่กล่าวมาข้างต้น นำมาจัดแสดงเป็นผังการดำเนินงานได้ดังภาพที่ 26



ภาพที่ 26 ผังงานสรุปขั้นตอนการดำเนินงานวิจัย

ผลและวิจารณ์

ผล

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาวิธีการทำงานของเทคนิคการแบ่งกลุ่ม 5 วิธี คือ วิธีการจัดกลุ่มแบบลำดับขั้น อัลกอริทึมจัดกลุ่มแบบเคมีน วิธีจัดระบบด้วยตัวเองของโคโฮเนน อัลกอริทึมจัดกลุ่มแบบเคมคอยด์ และการใช้อัลกอริทึมจัดกลุ่มแบบเคมคอยด์ร่วมกับการใช้วิธีวิเคราะห์ห่างแบบไดนามิกไทม์วอร์ปิง ข้อมูลที่ทำการศึกษามีจำนวนตัวแปรอิสระเท่ากับ 3, 5 และ 7 ตัวแปร และข้อมูลมีขนาดเล็ก, ขนาดกลาง และขนาดใหญ่เท่ากับ 25, 100 และ 300 ตัวอย่าง ตามลำดับ

สำหรับการแบ่งกลุ่มข้อมูลนั้น ข้อมูลขนาดต่าง ๆ จะมีการกำหนดจำนวนกลุ่มที่ต้องการแบ่งแตกต่างกัน โดยข้อมูลขนาดเล็กจะกำหนดจำนวนกลุ่มอยู่ในช่วง 2-5 กลุ่ม ข้อมูลขนาดกลางจะกำหนดจำนวนกลุ่มอยู่ในช่วง 2-8 กลุ่ม และข้อมูลขนาดใหญ่จะกำหนดจำนวนกลุ่มอยู่ในช่วง 2-15 กลุ่ม ในแต่ละชุดข้อมูลทำการทดลองจำนวน 1,000 ครั้ง เพื่อให้ผลที่ได้มีความเสถียรและเชื่อถือได้

ผลการวิจัยแบ่งเป็น 3 ส่วนคือ ส่วนที่ 1 และ 2 สำหรับข้อมูลจำลอง คือการเปรียบเทียบประสิทธิภาพการแบ่งกลุ่มด้วยค่า RMSSTD และ RS ตามลำดับ และส่วนที่ 3 สำหรับข้อมูลจริง แสดงผลการวัดค่าความแม่นยำในการแบ่งกลุ่ม

ส่วนที่ 1: ค่าความแตกต่างของข้อมูลภายในกลุ่ม (RMSSTD)

การเปรียบเทียบประสิทธิภาพการแบ่งกลุ่ม โดยพิจารณาจากค่าความแตกต่างของข้อมูลภายในกลุ่ม แสดงให้เห็นถึงประสิทธิภาพการจัดกลุ่มว่ามีมากน้อยเพียงใด ซึ่งหากค่าความแตกต่างภายในกลุ่มน้อย หมายถึงการแบ่งกลุ่มที่ดี ในการวิจัยครั้งนี้แบ่งตามขนาดตัวอย่างเป็น 3 กรณี คือ

- กำหนดให้
- * แทนเทคนิคการจัดกลุ่มที่ให้ประสิทธิภาพสูงสุด
 - ** แทนเทคนิคการจัดกลุ่มที่ให้ประสิทธิภาพสูงอันดับที่ 2

1.1 ข้อมูลขนาดเล็ก 25 หน่วยตัวอย่าง แบ่งแยกย่อยเป็น 3 กรณีคือ

1.1.1 จำนวนตัวแปรอิสระ 3 ตัวแปร

ตารางที่ 23 ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 1 ค่าเฉลี่ยเท่ากับ μ_{31} และค่าความแปรปรวนรวมเท่ากับ Σ_{31}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.88241	0.75556	0.7044	0.68632	0.67265*	0.67428**	0.72986	0.73782
3	0.79139	0.59135	0.56911	0.55626	0.55296*	0.55453**	0.61055	0.6212
4	0.71395	0.51063	0.49387	0.48385*	0.48899**	0.50077	0.54246	0.54951
5	0.64394	0.45871	0.44393**	0.4350*	0.44491	0.46091	0.4948	0.50167

ตารางที่ 24 ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 2 ค่าเฉลี่ยเท่ากับ μ_{32} และค่าความแปรปรวนรวมเท่ากับ Σ_{31}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.88447	0.75338	0.70299	0.68461	0.67032*	0.67136**	0.72625	0.73664
3	0.78955	0.58627	0.56451	0.55196	0.54739*	0.54944**	0.60766	0.61593
4	0.71268	0.50712	0.49071	0.47985*	0.48393**	0.49482	0.53657	0.54566
5	0.64357	0.4553	0.44124	0.43164*	0.4405**	0.45504	0.49032	0.49619

ตารางที่ 25 ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 3 ค่าเฉลี่ยเท่ากับ μ_{32} และค่าความแปรปรวนรวมเท่ากับ Σ_{32}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.89484	0.76409	0.71374	0.69606	0.68209*	0.68315**	0.74303	0.74565
3	0.80476	0.60753	0.5828	0.5690	0.56746*	0.56839**	0.62376	0.63274
4	0.73061	0.52564	0.50874	0.49562*	0.50175**	0.50862	0.54896	0.55999
5	0.65889	0.47022	0.45294**	0.4429*	0.45314	0.46747	0.49900	0.50691

ตารางที่ 26 ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 4 ค่าเฉลี่ยเท่ากับ μ_{33} และค่าความแปรปรวนรวมเท่ากับ Σ_{32}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.88974	0.76504	0.71625	0.69694	0.68333*	0.68451**	0.74371	0.74554
3	0.80408	0.60575	0.5829	0.56909	0.56522*	0.56775**	0.62033	0.63488
4	0.73067	0.52459	0.50696	0.4941*	0.49777**	0.50601	0.5487	0.56248
5	0.66003	0.46701	0.45261	0.44185*	0.4501**	0.46627	0.49833	0.51121

ตารางที่ 27 ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 5 ค่าเฉลี่ยเท่ากับ μ_{33} และค่าความแปรปรวนรวมเท่ากับ Σ_{33}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.88883	0.75896	0.71349	0.69463	0.68069*	0.68194**	0.74137	0.74528
3	0.80241	0.60362	0.57973	0.5661	0.56235*	0.56496**	0.6168	0.63164
4	0.72295	0.52421	0.50449	0.49288*	0.49672**	0.50709	0.54698	0.55589
5	0.65759	0.46639	0.45103	0.44119*	0.44945**	0.46517	0.49736	0.50684

1.1.2 จำนวนตัวแปรอิสระ 5 ตัวแปร

ตารางที่ 28 ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 1 ค่าเฉลี่ยเท่ากับ μ_{51} และค่าความแปรปรวนรวมเท่ากับ Σ_{51}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.87847	0.7557	0.70879	0.68786	0.67495*	0.67572**	0.72895	0.73503
3	0.78383	0.58981	0.56574	0.55239	0.54917*	0.55104**	0.60618	0.61687
4	0.71381	0.50725	0.49149	0.48207*	0.48844**	0.49818	0.53897	0.54946
5	0.64774	0.45863	0.44453**	0.4361*	0.44578	0.46534	0.49044	0.50296

ตารางที่ 29 ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 2 ค่าเฉลี่ยเท่ากับ μ_{s2} และค่าความแปรปรวนรวมเท่ากับ Σ_{s1}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.8789	0.75492	0.70509	0.68612	0.67263*	0.67368**	0.72543	0.73348
3	0.78842	0.5867	0.56357	0.55087	0.54631*	0.54802**	0.60285	0.61105
4	0.71288	0.50516	0.4903	0.48076*	0.48619**	0.49731	0.53662	0.54647
5	0.6472	0.45727	0.44309**	0.43429*	0.44444	0.46398	0.49139	0.50255

ตารางที่ 30 ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 3 ค่าเฉลี่ยเท่ากับ μ_{s2} และค่าความแปรปรวนรวมเท่ากับ Σ_{s2}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.9026	0.78048	0.73246	0.71766	0.70436*	0.70544**	0.76392	0.76499
3	0.82319	0.63874	0.61136	0.59775	0.59633*	0.5985**	0.64871	0.65874
4	0.75256	0.55898	0.53849	0.52612*	0.53013**	0.53892	0.58046	0.58855
5	0.69254	0.5024	0.48403	0.47446*	0.48394**	0.49881	0.53116	0.54032

ตารางที่ 31 ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 4 ค่าเฉลี่ยเท่ากับ μ_{53} และค่าความแปรปรวนรวมเท่ากับ Σ_{52}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.90255	0.78352	0.73461	0.71825	0.70579*	0.70674**	0.76717	0.76853
3	0.82509	0.64352	0.61178	0.5985	0.59753*	0.59781**	0.65016	0.6635
4	0.75559	0.56058	0.53912	0.5268*	0.53183**	0.53796	0.58147	0.59246
5	0.69383	0.50349	0.48685	0.47623*	0.48443**	0.50076	0.53476	0.54246

ตารางที่ 32 ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 5 ค่าเฉลี่ยเท่ากับ μ_{53} และค่าความแปรปรวนรวมเท่ากับ Σ_{53}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.90199	0.78675	0.73552	0.71692	0.70377*	0.70492**	0.76600	0.76747
3	0.82416	0.64301	0.61332	0.59903	0.59602*	0.5988**	0.65037	0.65749
4	0.75683	0.5613	0.53924	0.52618*	0.53086**	0.53826	0.58068	0.59133
5	0.69592	0.50201	0.48325	0.47274*	0.4809**	0.49802	0.52859	0.5384

1.1.3 จำนวนตัวแปรอิสระ 7 ตัวแปร

ตารางที่ 33 ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 1 ค่าเฉลี่ยเท่ากับ μ_{71} และค่าความแปรปรวนรวมเท่ากับ Σ_{71}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.89116	0.77174	0.71875	0.70317	0.69066*	0.69125**	0.74699	0.75061
3	0.81077	0.61683	0.59204	0.5794	0.5755*	0.57763**	0.62745	0.63833
4	0.74098	0.54257	0.52328	0.51365*	0.51947**	0.53056	0.56841	0.57606
5	0.68267	0.49458	0.47939**	0.47058*	0.48098	0.49836	0.5247	0.53476

ตารางที่ 34 ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 2 ค่าเฉลี่ยเท่ากับ μ_{72} และค่าความแปรปรวนรวมเท่ากับ Σ_{71}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.88828	0.76962	0.72084	0.70267	0.69067*	0.69173**	0.74321	0.75141
3	0.80594	0.61723	0.58923	0.57685	0.57401*	0.57595**	0.6311	0.64054
4	0.73453	0.53856	0.5209	0.51099*	0.51576**	0.52718	0.56618	0.57817
5	0.67459	0.49195	0.47713	0.46825*	0.47703**	0.49628	0.52475	0.53247

ตารางที่ 35 ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 3 ค่าเฉลี่ยเท่ากับ μ_{72} และค่าความแปรปรวนรวมเท่ากับ Σ_{72}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.8915	0.77661	0.73228	0.71217	0.70054*	0.70177**	0.75665	0.76252
3	0.81595	0.63047	0.60416	0.59122	0.58906*	0.58947**	0.64311	0.65654
4	0.75019	0.55515	0.53711	0.52676*	0.53105**	0.54189	0.58221	0.58955
5	0.68953	0.50782	0.49135	0.48242*	0.49074**	0.51243	0.53745	0.54622

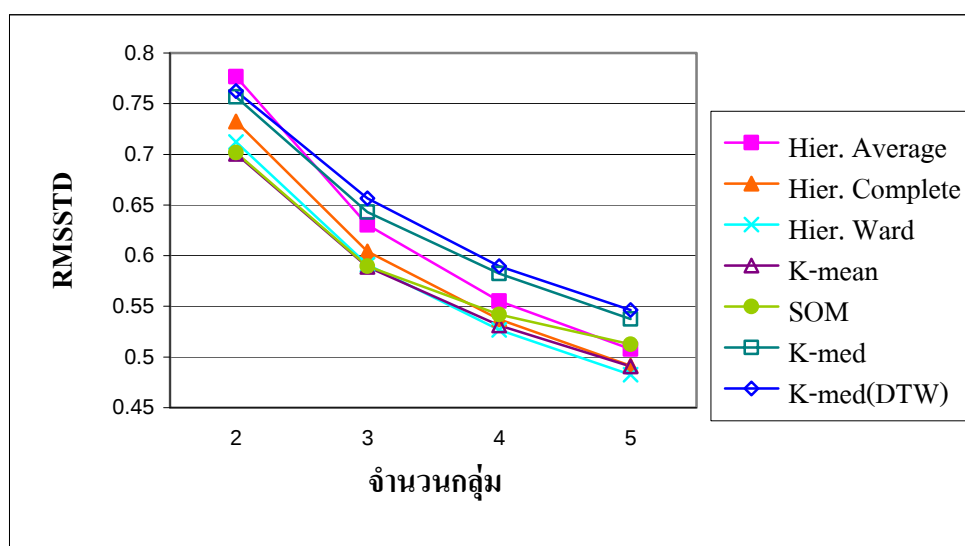
ตารางที่ 36 ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 4 ค่าเฉลี่ยเท่ากับ μ_{73} และค่าความแปรปรวนรวมเท่ากับ Σ_{72}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.8939	0.77977	0.72815	0.71105	0.69941*	0.70011**	0.75399	0.76125
3	0.81888	0.63027	0.60336	0.59138	0.58854*	0.59058**	0.6431	0.65239
4	0.75528	0.55473	0.53557	0.52564*	0.5312**	0.54328	0.58017	0.58959
5	0.69637	0.50699	0.49126**	0.48226*	0.49128	0.51066	0.53685	0.54552

ตารางที่ 37 ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 5 ค่าเฉลี่ยเท่ากับ μ_{73} และค่าความแปรปรวนรวมเท่ากับ Σ_{73}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.89884	0.77952	0.72771	0.71043	0.69883*	0.69928**	0.75223	0.76174
3	0.82342	0.62905	0.6038	0.59142	0.5895*	0.59076**	0.64389	0.64995
4	0.75663	0.55412	0.53617	0.52593*	0.53137**	0.54274	0.57818	0.59104
5	0.69776	0.50602	0.49117**	0.48236*	0.4915	0.51106	0.53655	0.54342

ตารางที่ 23-37 แสดงค่าเฉลี่ย RMSSTD ซึ่งข้อมูลมีขนาด 25 หน่วยตัวอย่าง มีจำนวนตัวแปรอิสระ 3, 5 และ 7 ตัวแปร และการแจกแจงของข้อมูลมีรูปแบบของค่าเฉลี่ยและค่าความแปรปรวนรวมที่ต่างกัน ผลการทดลองที่ได้แสดงให้เห็นว่า ในทุก ๆ ลักษณะของข้อมูล ผลที่ได้มีแนวโน้มไปในทางเดียวกัน ดังนั้นจึงเลือกใช้ข้อมูลในตารางที่ 35 เป็นตัวแทนในการสร้างกราฟเพื่อใช้อธิบายผลของข้อมูลได้ดังภาพที่ 27 ดังนี้



ภาพที่ 27 ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 3 ค่าเฉลี่ยเท่ากับ μ_{72} และค่าความแปรปรวนรวมเท่ากับ Σ_{72} จำนวนตัวแปรอิสระ 7 ตัวแปร ขนาดตัวอย่างเท่ากับ 25

จากภาพที่ 27 พบว่า เมื่อแบ่งกลุ่มข้อมูลเป็น 2 และ 3 กลุ่ม อัลกอริทึมจัดกลุ่มแบบเคมีนมีประสิทธิภาพดีที่สุด เนื่องจากให้ค่า RMSSTD ต่ำที่สุด รองลงมาคือ วิธีจัดระบบด้วยตัวเองของโคโฮเนน ซึ่งให้ค่าใกล้เคียงกัน ส่วนการแบ่งกลุ่มข้อมูลเป็น 4 และ 5 กลุ่มนั้น วิธีการจัดกลุ่มแบบลำดับขั้นด้วยวิธี Ward linkage มีประสิทธิภาพดีที่สุด รองลงมาคือ อัลกอริทึมจัดกลุ่มแบบเคมีน และวิธีการจัดกลุ่มแบบลำดับขั้นด้วยวิธี Complete linkage ให้ผลใกล้เคียงกัน และวิธีการจัดกลุ่มแบบลำดับขั้นด้วยวิธี Single linkage มีประสิทธิภาพต่ำสุดในทุกกรณีและมีค่าต่างจากวิธีอื่นมาก จึงไม่นำมารวมในการสร้างกราฟดังภาพที่ 27

1.2 ข้อมูลขนาดกลาง 100 หน่วยตัวอย่าง แบ่งแยกย่อยเป็น 3 กรณีคือ

1.2.1 จำนวนตัวแปรอิสระ 3 ตัวแปร

ตารางที่ 38 ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 1 ค่าเฉลี่ยเท่ากับ μ_{31} และค่าความแปรปรวนรวมเท่ากับ Σ_{31}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.9678	0.83916	0.72923	0.69848	0.67588*	0.68184**	0.73116	0.73999
3	0.94294	0.66448	0.60538	0.57457	0.55554*	0.55897**	0.61477	0.63026
4	0.9225	0.57023	0.53421	0.51113	0.49766*	0.49981**	0.55407	0.56471
5	0.90304	0.52179	0.4922	0.47195	0.46334*	0.46507**	0.5132	0.52387
6	0.88613	0.49174	0.46173	0.44233	0.43654*	0.43787**	0.48142	0.49215
7	0.86912	0.46799	0.43612	0.41811	0.41433*	0.41693**	0.45673	0.46618
8	0.85233	0.44584	0.41499	0.39755**	0.3953*	0.39776	0.4378	0.44618

ตารางที่ 39 ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 2 ค่าเฉลี่ยเท่ากับ μ_{32} และค่าความแปรปรวนรวมเท่ากับ Σ_{31}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.96778	0.8467	0.7294	0.69836	0.67621*	0.68153**	0.74065	0.74392
3	0.94259	0.6653	0.60258	0.57443	0.55523*	0.55904**	0.62788	0.62819
4	0.92157	0.57054	0.53461	0.5110	0.49701*	0.49963**	0.56922	0.56962
5	0.90271	0.52232	0.49213	0.4718	0.46342*	0.46437**	0.52834	0.52548
6	0.88546	0.49068	0.46095	0.44229	0.43603*	0.43726**	0.49472	0.49236
7	0.86856	0.46627	0.43621	0.41829	0.41421*	0.4164**	0.46909	0.47009
8	0.85107	0.4445	0.41506	0.39785**	0.39571*	0.39909	0.44716	0.44795

ตารางที่ 40 ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 3 ค่าเฉลี่ยเท่ากับ μ_{32} และค่าความแปรปรวนรวมเท่ากับ Σ_{32}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.96841	0.84725	0.73741	0.70962	0.68725*	0.69293**	0.75328	0.75349
3	0.94551	0.67992	0.62119	0.59163	0.57243*	0.57634**	0.64664	0.64828
4	0.9264	0.59421	0.55628	0.52934	0.51698*	0.51928**	0.58295	0.58079
5	0.9088	0.54469	0.50995	0.48651	0.47675*	0.47759**	0.53907	0.53892
6	0.89039	0.50802	0.47421	0.45318	0.44567*	0.44642**	0.5063	0.5033
7	0.87312	0.47881	0.44572	0.42638	0.42107*	0.42259**	0.47829	0.47768
8	0.85638	0.45282	0.42181	0.40382**	0.4008*	0.40396	0.45616	0.45471

ตารางที่ 41 ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 4 ค่าเฉลี่ยเท่ากับ μ_{33} และค่าความแปรปรวนรวมเท่ากับ Σ_{32}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.9676	0.85243	0.73871	0.71125	0.68874*	0.69369**	0.75346	0.75312
3	0.94341	0.68301	0.62174	0.59399	0.57434*	0.57813**	0.65028	0.6481
4	0.92253	0.59286	0.55761	0.53092	0.51792*	0.52074**	0.58294	0.58251
5	0.90322	0.54579	0.51174	0.48773	0.47782*	0.47837**	0.53847	0.53975
6	0.88435	0.51077	0.47566	0.4537	0.44674*	0.44646**	0.50541	0.50397
7	0.86819	0.48021	0.44668	0.42635	0.4215*	0.42317**	0.47702	0.477
8	0.85022	0.45271	0.42234	0.40393**	0.40078*	0.40423	0.45439	0.4553

ตารางที่ 42 ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 5 ค่าเฉลี่ยเท่ากับ μ_{33} และค่าความแปรปรวนรวมเท่ากับ Σ_{33}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.96743	0.85753	0.73992	0.71113	0.68785*	0.69337**	0.75515	0.76016
3	0.94344	0.68745	0.62333	0.59387	0.57429*	0.57772**	0.64688	0.64698
4	0.92254	0.59418	0.559	0.53166	0.51893*	0.52142**	0.58587	0.58451
5	0.90317	0.5477	0.51281	0.48851	0.4783*	0.47893**	0.54215	0.53935
6	0.8851	0.5105	0.47657	0.45454	0.44747*	0.44815**	0.50579	0.50507
7	0.86862	0.48127	0.44736	0.42715	0.42258*	0.4239**	0.47896	0.47837
8	0.85247	0.45425	0.42309	0.40459**	0.40153*	0.40552	0.45755	0.45479

1.2.2 จำนวนตัวแปรอิสระ 5 ตัวแปร

ตารางที่ 43 ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 1 ค่าเฉลี่ยเท่ากับ μ_{s1} และค่าความแปรปรวนรวมเท่ากับ Σ_{s1}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.96426	0.84341	0.72681	0.69711	0.67502*	0.6804**	0.73576	0.73482
3	0.941	0.65975	0.60385	0.57233	0.55393*	0.55727**	0.62515	0.62603
4	0.91855	0.56426	0.53186	0.50892	0.49616*	0.49812**	0.56461	0.56761
5	0.89936	0.51536	0.48852	0.47062	0.46318*	0.46516**	0.52418	0.52641
6	0.88255	0.48658	0.46039	0.4432	0.43896*	0.44022**	0.4949	0.49393
7	0.86621	0.4632	0.43804	0.4214	0.41931*	0.42128**	0.47085	0.47144
8	0.84989	0.4440	0.41874	0.40301**	0.4018*	0.40492	0.45113	0.45152

ตารางที่ 44 ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 2 ค่าเฉลี่ยเท่ากับ μ_{s2} และค่าความแปรปรวนรวมเท่ากับ Σ_{s1}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.96435	0.84222	0.72585	0.69711	0.67442*	0.68001**	0.73307	0.73327
3	0.93957	0.66094	0.6015	0.57124	0.55356*	0.55663**	0.62600	0.62316
4	0.91829	0.55836	0.53022	0.5081	0.4946*	0.49719**	0.56359	0.56302
5	0.90003	0.51444	0.48766	0.46954	0.46283*	0.46479**	0.52648	0.5237
6	0.88312	0.48445	0.459	0.44237	0.4378*	0.4402**	0.49392	0.49503
7	0.86417	0.46108	0.43645	0.42048	0.4176*	0.42023**	0.47042	0.47182
8	0.8483	0.44241	0.41728	0.40197**	0.40165*	0.40516	0.45092	0.45019

ตารางที่ 45 ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 3 ค่าเฉลี่ยเท่ากับ μ_{s2} และค่าความแปรปรวนรวมเท่ากับ Σ_{s2}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.97097	0.86562	0.75597	0.73054	0.70725*	0.71273**	0.77629	0.77685
3	0.94863	0.7136	0.64921	0.62252	0.60315*	0.60729**	0.67412	0.67506
4	0.92965	0.63407	0.58978	0.56367	0.5514*	0.55353**	0.61387	0.61439
5	0.91334	0.58331	0.54573	0.52133	0.51064*	0.51164**	0.56948	0.57029
6	0.89858	0.54688	0.51011	0.48844	0.47972*	0.48088**	0.53852	0.53605
7	0.88173	0.51639	0.48131	0.46191	0.45566*	0.45796**	0.51146	0.51086
8	0.86821	0.48973	0.45762	0.44024	0.43614*	0.44012**	0.4918	0.48893

ตารางที่ 46 ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 4 ค่าเฉลี่ยเท่ากับ μ_{s3} และค่าความแปรปรวนรวมเท่ากับ Σ_{s2}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.97191	0.85966	0.75641	0.73013	0.70714*	0.71271**	0.77502	0.77214
3	0.94899	0.71353	0.64879	0.62133	0.60187*	0.60625**	0.67287	0.6732
4	0.92947	0.63211	0.58744	0.56315	0.54983*	0.55201**	0.61245	0.61256
5	0.91187	0.58333	0.54366	0.52031	0.50957*	0.51039**	0.5689	0.57041
6	0.89465	0.5461	0.50871	0.48695	0.47869*	0.47989**	0.53739	0.5368
7	0.88045	0.51461	0.47996	0.46055	0.45463*	0.45693**	0.51186	0.51197
8	0.86544	0.4872	0.45706	0.43922**	0.43526*	0.4393	0.4902	0.48968

ตารางที่ 47 ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 5 ค่าเฉลี่ยเท่ากับ μ_{53} และค่าความแปรปรวนรวมเท่ากับ Σ_{53}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.97053	0.86926	0.75916	0.72889	0.70671*	0.71166**	0.77776	0.77571
3	0.94839	0.71643	0.64965	0.62049	0.60227*	0.60641**	0.67387	0.67313
4	0.93008	0.63121	0.58758	0.56145	0.54924*	0.55149**	0.61156	0.61284
5	0.91289	0.58304	0.54234	0.51748	0.50692*	0.50812**	0.56886	0.56838
6	0.89593	0.5421	0.50527	0.48342	0.47526*	0.47697**	0.53542	0.53379
7	0.8791	0.5113	0.47619	0.45649	0.4502*	0.45308**	0.50799	0.50822
8	0.86331	0.48377	0.45244	0.43436**	0.43057*	0.43442	0.48382	0.48598

1.2.3 จำนวนตัวแปรอิสระ 7 ตัวแปร

ตารางที่ 48 ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 1 ค่าเฉลี่ยเท่ากับ μ_{71} และค่าความแปรปรวนรวมเท่ากับ Σ_{71}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.96812	0.85416	0.74155	0.71533	0.69321*	0.69812**	0.75286	0.75029
3	0.94493	0.68273	0.62606	0.59815	0.58091*	0.58401**	0.65105	0.65103
4	0.92572	0.5902	0.5621	0.54017	0.5277*	0.52955**	0.59451	0.59218
5	0.91004	0.54959	0.52279	0.5048	0.49761*	0.49955**	0.55691	0.55543
6	0.89416	0.52149	0.49558	0.47905	0.47397*	0.47568**	0.52922	0.52518
7	0.87861	0.50083	0.47434	0.45823	0.45461*	0.45815**	0.50798	0.50449
8	0.86303	0.48212	0.45611	0.44055**	0.43881*	0.44276	0.48744	0.48702

ตารางที่ 49 ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 2 ค่าเฉลี่ยเท่ากับ μ_{72} และค่าความแปรปรวนรวมเท่ากับ Σ_{71}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.96742	0.85489	0.74039	0.71343	0.6914*	0.69635**	0.75288	0.7513
3	0.94349	0.68015	0.62669	0.59672	0.58014*	0.58297**	0.65064	0.64722
4	0.92605	0.59267	0.56129	0.53871	0.52596*	0.52873**	0.59444	0.59032
5	0.90952	0.5481	0.5211	0.50348	0.49614*	0.49807**	0.55666	0.55171
6	0.89173	0.52039	0.49458	0.47789	0.47307*	0.47517**	0.52774	0.5265
7	0.87773	0.49788	0.47298	0.45714	0.45324*	0.45634**	0.50502	0.50293
8	0.86386	0.48026	0.4545	0.43965**	0.43698*	0.44195	0.48648	0.48504

ตารางที่ 50 ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 3 ค่าเฉลี่ยเท่ากับ μ_{72} และค่าความแปรปรวนรวมเท่ากับ Σ_{72}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.97013	0.85704	0.74748	0.72109	0.70059*	0.7058**	0.76131	0.76337
3	0.94777	0.69273	0.63623	0.61007	0.59269*	0.59598**	0.66286	0.66345
4	0.93028	0.60716	0.57466	0.55435	0.54199*	0.54528**	0.60625	0.60769
5	0.91342	0.56569	0.53673	0.51877	0.51029*	0.51161**	0.56925	0.56797
6	0.89751	0.53673	0.50806	0.49189	0.48561*	0.48688**	0.54106	0.54103
7	0.8841	0.5139	0.48521	0.4702	0.46539*	0.4687**	0.51771	0.5173
8	0.86979	0.49357	0.46686	0.45218**	0.44922*	0.45355	0.49974	0.49977

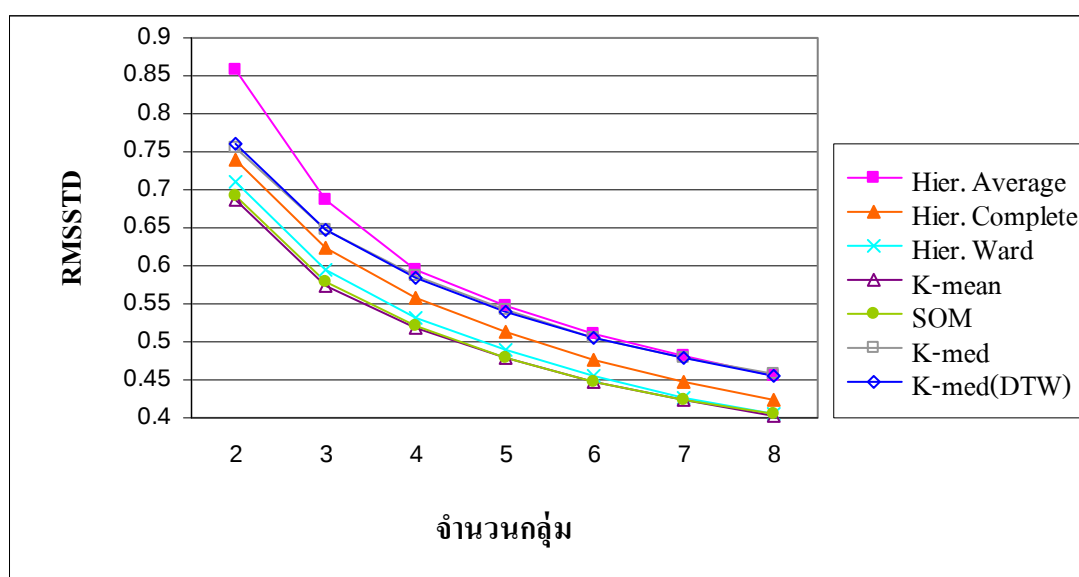
ตารางที่ 51 ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 4 ค่าเฉลี่ยเท่ากับ μ_{73} และค่าความแปรปรวนรวมเท่ากับ Σ_{72}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.96924	0.85697	0.74689	0.72117	0.69977*	0.70496**	0.76088	0.76113
3	0.94715	0.69434	0.63673	0.60898	0.59148*	0.59438**	0.66058	0.66177
4	0.92812	0.60851	0.57416	0.55289	0.54057*	0.54423**	0.60408	0.60385
5	0.91266	0.56338	0.53675	0.51699	0.50896*	0.51044**	0.56614	0.56827
6	0.8972	0.53469	0.50784	0.49004	0.48418*	0.48601**	0.53900	0.53932
7	0.88362	0.51112	0.48427	0.46851	0.46362*	0.46769**	0.51686	0.51712
8	0.87116	0.4927	0.46587	0.45072**	0.44777*	0.45246	0.49832	0.4975

ตารางที่ 52 ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 5 ค่าเฉลี่ยเท่ากับ μ_{73} และค่าความแปรปรวนรวมเท่ากับ Σ_{73}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.96858	0.85998	0.75265	0.72225	0.70127*	0.70656**	0.76398	0.7652
3	0.9465	0.69719	0.63869	0.61082	0.59301*	0.59595**	0.66498	0.66421
4	0.92854	0.61114	0.57611	0.55379	0.54132*	0.5441**	0.60574	0.60755
5	0.91261	0.56873	0.53637	0.51799	0.50995*	0.5109**	0.56876	0.56862
6	0.89746	0.53766	0.50829	0.49126	0.48438*	0.48667**	0.54009	0.5407
7	0.88304	0.51363	0.48605	0.46958	0.46528*	0.46789**	0.51820	0.51901
8	0.87026	0.49353	0.46662	0.45178**	0.4489*	0.45312	0.49921	0.49988

ตารางที่ 38-52 แสดงค่าเฉลี่ย RMSSTD ซึ่งข้อมูลมีขนาด 100 หน่วยตัวอย่าง มีจำนวนตัวแปรอิสระ 3, 5 และ 7 ตัวแปร และการแจกแจงของข้อมูลมีรูปแบบของค่าเฉลี่ยและค่าความแปรปรวนร่วมที่ต่างกัน ผลการทดลองที่ได้แสดงให้เห็นว่า ในทุก ๆ ลักษณะของข้อมูล ผลที่ได้มีแนวโน้มไปในทางเดียวกัน ดังนั้นจึงเลือกใช้ข้อมูลในตารางที่ 42 เป็นตัวแทนในการสร้างกราฟได้ดังภาพที่ 28 เพื่อใช้อธิบายผลของข้อมูล ดังนี้



ภาพที่ 28 ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 5 ค่าเฉลี่ยเท่ากับ μ_{33} และค่าความแปรปรวนร่วมเท่ากับ Σ_{33} จำนวนตัวแปรอิสระ 3 ตัวแปร ขนาดตัวอย่างเท่ากับ 100

จากภาพที่ 28 พบว่า เมื่อแบ่งกลุ่มข้อมูลตั้งแต่ 2-8 กลุ่ม อัลกอริทึมจัดกลุ่มแบบเคมีนมีประสิทธิภาพดีที่สุด เนื่องจากให้ค่า RMSSTD ต่ำที่สุด รองลงมาคือ วิธีจัดระบบด้วยตัวเองของโคโฮเนนซึ่งให้ค่าใกล้เคียงกัน ยกเว้นเมื่อแบ่งข้อมูลเป็น 8 กลุ่มนั้น วิธีการจัดกลุ่มแบบลำดับขั้นด้วยวิธี Ward linkage มีประสิทธิภาพดีกว่าวิธีจัดระบบด้วยตัวเองของโคโฮเนนเล็กน้อย และวิธีการจัดกลุ่มแบบลำดับขั้นด้วยวิธี Single linkage มีประสิทธิภาพต่ำสุดในทุกกรณีและมีค่าต่างจากวิธีอื่นมาก จึงไม่ได้นำมารวมในการสร้างกราฟดังภาพที่ 28

1.3 ข้อมูลขนาดกลาง 300 หน่วยตัวอย่าง แบ่งแยกย่อยเป็น 3 กรณีคือ

1.3.1 จำนวนตัวแปรอิสระ 3 ตัวแปร

ตารางที่ 53 ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 1 ค่าเฉลี่ยเท่ากับ μ_{31} และค่าความแปรปรวนรวมเท่ากับ Σ_{31}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.98641	0.89913	0.74443	0.70585	0.67802*	0.68561**	0.74448	0.74189
3	0.97679	0.72454	0.62965	0.5831	0.5584*	0.56491**	0.63357	0.633
4	0.96783	0.6064	0.55925	0.52224	0.49996*	0.5057**	0.57252	0.57326
5	0.96071	0.5591	0.5161	0.48556	0.46805*	0.47316**	0.53244	0.53058
6	0.95344	0.52861	0.48706	0.45946	0.44313*	0.44757**	0.50117	0.50056
7	0.94724	0.50629	0.46463	0.43812	0.42261*	0.42657**	0.47808	0.47773
8	0.94105	0.48795	0.44571	0.42001	0.40583*	0.40865**	0.45833	0.4579
9	0.93502	0.4714	0.42931	0.40429	0.39095*	0.39341**	0.44114	0.44081
10	0.92918	0.45715	0.41493	0.39042	0.37823*	0.38012**	0.42656	0.4276
11	0.92309	0.44342	0.4016	0.37799	0.36676*	0.3682**	0.41428	0.4143
12	0.91792	0.43138	0.38963	0.36677	0.35685*	0.35769**	0.40288	0.40232
13	0.91205	0.41935	0.3785	0.35654	0.34708*	0.34839**	0.39254	0.3922
14	0.9069	0.40824	0.36847	0.34711	0.33827*	0.3397**	0.38272	0.38316
15	0.90138	0.39719	0.35908	0.33844	0.33054*	0.33265**	0.37503	0.3749

ตารางที่ 54 ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 2 ค่าเฉลี่ยเท่ากับ μ_{32} และค่าความแปรปรวนรวมเท่ากับ Σ_{31}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.98674	0.89872	0.74528	0.70373	0.67798*	0.68536**	0.74247	0.73817
3	0.97719	0.71972	0.62852	0.58434	0.55852*	0.56487**	0.63611	0.63047
4	0.9687	0.60948	0.55825	0.52192	0.49977*	0.50571**	0.57351	0.57464
5	0.96109	0.55812	0.51686	0.48525	0.46772*	0.47314**	0.53108	0.53139
6	0.95383	0.52836	0.48765	0.45898	0.44297*	0.44738**	0.503	0.50282
7	0.94675	0.50699	0.46485	0.43772	0.42276*	0.4262**	0.47807	0.47943
8	0.94004	0.48768	0.44587	0.41971	0.40545*	0.40861**	0.45878	0.45783
9	0.93358	0.47195	0.42924	0.40402	0.39069*	0.39309**	0.4419	0.44187
10	0.92797	0.45805	0.41456	0.39016	0.37759*	0.37979**	0.42841	0.42786
11	0.92225	0.44419	0.40134	0.37774	0.36628*	0.36786**	0.41522	0.41455
12	0.91621	0.43142	0.38941	0.36646	0.35577*	0.3574**	0.40379	0.40339
13	0.9096	0.42013	0.37838	0.35615	0.34681*	0.34784**	0.39324	0.39296
14	0.90424	0.40873	0.36845	0.34678	0.33806*	0.33969**	0.38304	0.38334
15	0.8988	0.39767	0.35889	0.3382	0.33049*	0.33201**	0.37572	0.37435

ตารางที่ 55 ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 3 ค่าเฉลี่ยเท่ากับ μ_{32} และค่าความแปรปรวนรวมเท่ากับ Σ_{32}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.98756	0.90577	0.75015	0.71452	0.68802*	0.69542**	0.75587	0.75503
3	0.97797	0.7425	0.64132	0.5995	0.57406*	0.58105**	0.64617	0.6484
4	0.96932	0.63468	0.57854	0.54117	0.51962*	0.5267**	0.58542	0.58721
5	0.96123	0.58795	0.53603	0.5025	0.48309*	0.48815**	0.54436	0.54288
6	0.95371	0.55447	0.50345	0.47175	0.45249*	0.45737**	0.51231	0.51192
7	0.94695	0.52719	0.47667	0.44655	0.42824*	0.43213**	0.48735	0.48492
8	0.9411	0.5038	0.45396	0.42564	0.40926*	0.41229**	0.46618	0.46556
9	0.93484	0.48336	0.43478	0.40836	0.39361*	0.39616**	0.44752	0.44738
10	0.92878	0.46487	0.41867	0.39352	0.38047*	0.38242**	0.43261	0.43238
11	0.92328	0.44817	0.40453	0.38054	0.36865*	0.37023**	0.41900	0.41813
12	0.91768	0.43294	0.39194	0.36892	0.35853*	0.35958**	0.40648	0.40717
13	0.91215	0.41871	0.38079	0.35845	0.34884*	0.35033**	0.39686	0.39683
14	0.90698	0.40585	0.37026	0.34893	0.34035*	0.34184**	0.38619	0.38699
15	0.90138	0.39477	0.36076	0.34017	0.33261*	0.33416**	0.37800	0.37777

ตารางที่ 56 ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 4 ค่าเฉลี่ยเท่ากับ μ_{33} และค่าความแปรปรวนรวมเท่ากับ Σ_{32}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.98777	0.90124	0.75055	0.71508	0.68805*	0.69546**	0.75156	0.7519
3	0.9781	0.74097	0.64267	0.60111	0.57458*	0.58173**	0.64864	0.64885
4	0.9698	0.63539	0.58166	0.54232	0.52067*	0.52731**	0.58568	0.58519
5	0.96209	0.58853	0.53881	0.50329	0.48345*	0.4892**	0.54622	0.54544
6	0.95509	0.55581	0.50485	0.47257	0.45295*	0.4578**	0.5129	0.51179
7	0.94838	0.52923	0.47702	0.44745	0.42893*	0.43299**	0.4884	0.48759
8	0.942	0.50649	0.45473	0.42675	0.40983*	0.41294**	0.46539	0.46507
9	0.93564	0.48495	0.43578	0.4094	0.39432*	0.39657**	0.44879	0.44797
10	0.92956	0.46566	0.41931	0.39455	0.38076*	0.38271**	0.43344	0.43305
11	0.92416	0.44955	0.40465	0.38153	0.36894*	0.37088**	0.41993	0.41921
12	0.91823	0.43419	0.39234	0.36984	0.35887*	0.36058**	0.40735	0.40733
13	0.91204	0.41999	0.38097	0.35933	0.34921*	0.35095**	0.39829	0.39807
14	0.90648	0.40767	0.37069	0.34974	0.34101*	0.34242**	0.38738	0.38703
15	0.901	0.39629	0.36113	0.34089	0.33351*	0.33478**	0.37882	0.3784

ตารางที่ 57 ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 5 ค่าเฉลี่ยเท่ากับ μ_{33} และค่าความแปรปรวนรวมเท่ากับ Σ_{33}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.98723	0.90419	0.75403	0.71528	0.68847*	0.69637**	0.75659	0.75793
3	0.9776	0.73858	0.64423	0.60046	0.57467*	0.58188**	0.65304	0.65265
4	0.96958	0.63454	0.57868	0.5423	0.52034*	0.52739**	0.58933	0.58918
5	0.96183	0.58763	0.53781	0.50316	0.48347*	0.48895**	0.5451	0.54438
6	0.95439	0.55562	0.50428	0.47214	0.45313*	0.4576**	0.51454	0.51286
7	0.94799	0.52824	0.47694	0.44717	0.42848*	0.43271**	0.48826	0.48781
8	0.94184	0.50619	0.45444	0.42643	0.40959*	0.41281**	0.46618	0.46546
9	0.93576	0.48601	0.43612	0.40932	0.39385*	0.39673**	0.44936	0.44854
10	0.92935	0.46715	0.4197	0.39448	0.38112*	0.38303**	0.43401	0.43355
11	0.92363	0.44954	0.40569	0.38145	0.36919*	0.37115**	0.4202	0.4194
12	0.91776	0.43337	0.3932	0.36984	0.35896*	0.36049**	0.40842	0.40771
13	0.91207	0.42007	0.38208	0.35934	0.34991*	0.35107**	0.39708	0.39659
14	0.90709	0.40738	0.37135	0.34978	0.34098*	0.34267**	0.38716	0.38682
15	0.90168	0.39611	0.36179	0.34104	0.3330*	0.33509**	0.37802	0.37817

1.3.2 จำนวนตัวแปรอิสระ 5 ตัวแปร

ตารางที่ 58 ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 1 ค่าเฉลี่ยเท่ากับ μ_{s1} และค่าความแปรปรวนรวมเท่ากับ Σ_{s1}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.98627	0.89807	0.74057	0.7033	0.67652*	0.6836**	0.73948	0.7371
3	0.97581	0.7238	0.62441	0.58025	0.55661*	0.56265**	0.62918	0.62956
4	0.96721	0.6009	0.55397	0.51853	0.49766*	0.50301**	0.5697	0.57205
5	0.95916	0.5487	0.51099	0.48234	0.46518*	0.47062**	0.53125	0.53255
6	0.95231	0.51993	0.48336	0.45826	0.44434*	0.44859**	0.50438	0.50274
7	0.94577	0.49934	0.4629	0.43929	0.42638*	0.4300**	0.48192	0.48005
8	0.93941	0.48362	0.44579	0.42313	0.41106*	0.4142**	0.46069	0.4626
9	0.93379	0.46962	0.43112	0.40903	0.39751*	0.40003**	0.44642	0.44624
10	0.92837	0.45616	0.41802	0.39645	0.3858*	0.38775**	0.43267	0.43221
11	0.92287	0.44419	0.4062	0.38517	0.37501*	0.37681**	0.42097	0.42205
12	0.91726	0.43333	0.3955	0.37508	0.36584*	0.36736**	0.41081	0.41107
13	0.91211	0.42298	0.38541	0.36592	0.35728*	0.35874**	0.40075	0.40026
14	0.90756	0.41302	0.37638	0.35753	0.34933*	0.3516**	0.39251	0.3922
15	0.9025	0.40349	0.36811	0.34981	0.34244*	0.34489**	0.38506	0.38459

ตารางที่ 59 ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 2 ค่าเฉลี่ยเท่ากับ μ_{52} และค่าความแปรปรวนรวมเท่ากับ Σ_{51}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.98657	0.89432	0.74135	0.70238	0.67627*	0.68376**	0.73853	0.73752
3	0.9765	0.72103	0.62413	0.58041	0.55634*	0.56264**	0.62801	0.62777
4	0.96731	0.59666	0.55582	0.51831	0.49755*	0.5027**	0.56877	0.56947
5	0.95967	0.54847	0.51233	0.48236	0.46506*	0.47057**	0.53000	0.52976
6	0.95292	0.51866	0.48329	0.45824	0.44352*	0.44833**	0.50224	0.50191
7	0.94656	0.49909	0.46306	0.43924	0.42632*	0.42985**	0.47985	0.47905
8	0.94082	0.48255	0.44571	0.42305	0.41069*	0.41377**	0.45964	0.46004
9	0.93474	0.46806	0.43118	0.40895	0.39752*	0.39993**	0.44482	0.44489
10	0.92865	0.45555	0.41783	0.39649	0.38553*	0.38743**	0.43159	0.4318
11	0.92301	0.44413	0.40595	0.38531	0.37487*	0.37665**	0.42141	0.42125
12	0.91765	0.43322	0.39521	0.37519	0.36537*	0.36708**	0.41048	0.41015
13	0.91223	0.42264	0.38527	0.36594	0.35687*	0.35849**	0.40032	0.40053
14	0.90717	0.41262	0.37628	0.35745	0.34942*	0.35128**	0.39302	0.39253
15	0.90231	0.40324	0.3680	0.34968	0.34219*	0.34445**	0.38409	0.38434

ตารางที่ 60 ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 3 ค่าเฉลี่ยเท่ากับ μ_{s2} และค่าความแปรปรวนรวมเท่ากับ Σ_{s2}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.98863	0.91289	0.76856	0.73558	0.70893*	0.71636**	0.77408	0.77453
3	0.97951	0.7661	0.66819	0.63015	0.60505*	0.61244**	0.67654	0.67725
4	0.97151	0.67604	0.6114	0.57627	0.55544*	0.56207**	0.62128	0.6205
5	0.96504	0.62765	0.57099	0.5376	0.51611*	0.52186**	0.57853	0.57769
6	0.95836	0.59652	0.53767	0.5069	0.48628*	0.49144**	0.54573	0.54486
7	0.95253	0.56894	0.5115	0.48238	0.4632*	0.46777**	0.52216	0.5207
8	0.94739	0.54287	0.48987	0.46251	0.44509*	0.4487**	0.5016	0.50112
9	0.94172	0.52126	0.47152	0.44596	0.43028*	0.43357**	0.48512	0.48348
10	0.93685	0.50276	0.45548	0.43179	0.41753*	0.42023**	0.47052	0.46992
11	0.93149	0.48565	0.44173	0.41943	0.40665*	0.40912**	0.4577	0.4568
12	0.92631	0.46952	0.42964	0.40857	0.39662*	0.39883**	0.44645	0.4452
13	0.92082	0.45587	0.4191	0.39874	0.38792*	0.3899**	0.4364	0.43566
14	0.91582	0.44453	0.40951	0.38977	0.37992*	0.38193**	0.42726	0.426
15	0.91134	0.43367	0.40042	0.38162	0.37257*	0.37491**	0.41888	0.41796

ตารางที่ 61 ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 4 ค่าเฉลี่ยเท่ากับ μ_{53} และค่าความแปรปรวนรวมเท่ากับ Σ_{52}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.98844	0.91786	0.76528	0.73466	0.70823*	0.71549**	0.77522	0.77519
3	0.97991	0.77126	0.66708	0.62976	0.60394*	0.61123**	0.67789	0.67575
4	0.97192	0.67433	0.61212	0.57565	0.55481*	0.56145**	0.61875	0.61726
5	0.965	0.62716	0.57006	0.53657	0.51515*	0.52074**	0.57657	0.57598
6	0.9586	0.59341	0.53679	0.50598	0.48566*	0.49091**	0.54677	0.54601
7	0.95234	0.56617	0.51052	0.4817	0.46224*	0.46672**	0.52306	0.52046
8	0.94657	0.54131	0.48895	0.46192	0.44437*	0.44793**	0.50113	0.50083
9	0.94085	0.51993	0.47087	0.4455	0.42949*	0.43275**	0.48496	0.48288
10	0.93502	0.50203	0.45516	0.43128	0.41708*	0.41978**	0.46956	0.4698
11	0.92978	0.48445	0.44176	0.41885	0.40587*	0.40823**	0.45753	0.45732
12	0.92483	0.46915	0.42964	0.40782	0.39612*	0.39797**	0.44683	0.44523
13	0.91987	0.45526	0.41871	0.39795	0.38724*	0.38933**	0.43491	0.43524
14	0.91543	0.44374	0.40897	0.38902	0.37922*	0.38141**	0.42774	0.42654
15	0.91038	0.43333	0.39988	0.3809	0.37202*	0.37417**	0.41898	0.41834

ตารางที่ 62 ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 5 ค่าเฉลี่ยเท่ากับ μ_{53} และค่าความแปรปรวนรวมเท่ากับ Σ_{53}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.98852	0.91306	0.76777	0.73273	0.70639*	0.71392**	0.78016	0.77576
3	0.97929	0.77028	0.66847	0.62838	0.60245*	0.61024**	0.67558	0.67532
4	0.9711	0.67427	0.61042	0.57358	0.55238*	0.5592**	0.61515	0.61395
5	0.96436	0.62581	0.56777	0.53341	0.5115*	0.51755**	0.57397	0.57395
6	0.95807	0.59232	0.53462	0.5018	0.48102*	0.48656**	0.54238	0.5414
7	0.95177	0.56268	0.50722	0.47655	0.4576*	0.46202**	0.51764	0.51743
8	0.9459	0.53973	0.48485	0.45613	0.43886*	0.44245**	0.49568	0.49532
9	0.9403	0.51563	0.46547	0.43906	0.42351*	0.42691**	0.47939	0.47847
10	0.93502	0.49587	0.44933	0.4244	0.41039*	0.41357**	0.46487	0.46345
11	0.93033	0.47816	0.43515	0.41164	0.39924*	0.40162**	0.45188	0.45083
12	0.92524	0.46185	0.42278	0.40042	0.38905*	0.39135**	0.43936	0.43896
13	0.92081	0.44812	0.41223	0.39042	0.38005*	0.38218**	0.42913	0.42862
14	0.91575	0.43591	0.40235	0.38131	0.37196*	0.37395**	0.41992	0.41963
15	0.91074	0.42425	0.39318	0.37301	0.3646*	0.36692**	0.41061	0.4106

1.3.3 จำนวนตัวแปรอิสระ 7 ตัวแปร

ตารางที่ 63 ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 1 ค่าเฉลี่ยเท่ากับ μ_{71} และค่าความแปรปรวนรวมเท่ากับ Σ_{71}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.98762	0.90053	0.75778	0.71784	0.69312*	0.70023**	0.75476	0.75044
3	0.97796	0.73138	0.64721	0.60431	0.58126*	0.58701**	0.65036	0.65164
4	0.97008	0.62342	0.58088	0.54803	0.52743*	0.5331**	0.59848	0.59522
5	0.96359	0.57946	0.54225	0.51495	0.49823*	0.50377**	0.5614	0.55836
6	0.95719	0.5534	0.51621	0.49232	0.47741*	0.48207**	0.53485	0.53223
7	0.95123	0.53378	0.4964	0.47419	0.46006*	0.46412**	0.51191	0.5116
8	0.94611	0.51887	0.48078	0.45885	0.4453*	0.44873**	0.49775	0.49529
9	0.94084	0.50661	0.46665	0.44552	0.43275*	0.43605**	0.48265	0.48038
10	0.9359	0.49447	0.45425	0.4338	0.42186*	0.42433**	0.46957	0.46804
11	0.93091	0.48303	0.44326	0.42336	0.41196*	0.41429**	0.45917	0.45779
12	0.92631	0.47225	0.43298	0.41398	0.40315*	0.40528**	0.4489	0.44852
13	0.92165	0.46235	0.42396	0.40547	0.39507*	0.39736**	0.44163	0.43947
14	0.91692	0.4531	0.41567	0.39768	0.38828*	0.39063**	0.43399	0.43214
15	0.91271	0.44463	0.40807	0.39054	0.38162*	0.38425**	0.42669	0.42559

ตารางที่ 64 ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 2 ค่าเฉลี่ยเท่ากับ μ_{72} และค่าความแปรปรวนรวมเท่ากับ Σ_{71}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.98724	0.90775	0.75508	0.7179	0.6925*	0.69939**	0.75414	0.75196
3	0.97843	0.73705	0.64622	0.60483	0.58106*	0.58686**	0.65031	0.64865
4	0.97117	0.6248	0.58214	0.54849	0.52763*	0.53332**	0.59817	0.59676
5	0.96364	0.5795	0.54197	0.51545	0.49894*	0.5042**	0.56089	0.55906
6	0.95793	0.55402	0.5163	0.4928	0.47786*	0.48247**	0.53436	0.53224
7	0.9524	0.53517	0.4978	0.47468	0.46052*	0.46456**	0.51398	0.5121
8	0.94723	0.51877	0.48163	0.45934	0.44598*	0.44939**	0.49681	0.49539
9	0.94292	0.50529	0.46765	0.44597	0.43298*	0.43654**	0.48264	0.48106
10	0.93787	0.49341	0.45519	0.43416	0.42169*	0.42494**	0.47031	0.4694
11	0.93316	0.48249	0.44418	0.42369	0.41221*	0.41452**	0.45929	0.45823
12	0.92809	0.47228	0.43391	0.41428	0.40355*	0.40594**	0.44997	0.44833
13	0.92342	0.46213	0.42494	0.40583	0.39561*	0.39792**	0.44131	0.43996
14	0.91888	0.45298	0.41636	0.39806	0.38871*	0.39125**	0.43351	0.43251
15	0.91444	0.4445	0.40848	0.39093	0.38178*	0.38484**	0.4265	0.42531

ตารางที่ 65 ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 3 ค่าเฉลี่ยเท่ากับ μ_{72} และค่าความแปรปรวนรวมเท่ากับ Σ_{72}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.98792	0.90529	0.76043	0.72614	0.70097*	0.70781**	0.7623	0.7635
3	0.97918	0.74685	0.65718	0.61787	0.59405*	0.59982**	0.66345	0.66387
4	0.97221	0.6414	0.59628	0.56387	0.54311*	0.54889**	0.61077	0.61091
5	0.96586	0.5996	0.55841	0.53077	0.51395*	0.51914**	0.5747	0.5751
6	0.9593	0.57274	0.53181	0.50632	0.48967*	0.49452**	0.54872	0.54846
7	0.95356	0.55193	0.50992	0.48678	0.47066*	0.47527**	0.5272	0.52648
8	0.94804	0.53492	0.49284	0.47064	0.45531*	0.45936**	0.5093	0.50886
9	0.94251	0.5199	0.47793	0.45717	0.44347*	0.44649**	0.49536	0.4948
10	0.93788	0.50672	0.46538	0.44565	0.43294*	0.4357**	0.48328	0.48291
11	0.93292	0.49555	0.45458	0.4355	0.42366*	0.42635**	0.47286	0.47265
12	0.92828	0.4837	0.4451	0.42644	0.41551*	0.41811**	0.46383	0.46358
13	0.92399	0.47292	0.43622	0.4183	0.40808*	0.41051**	0.45423	0.45439
14	0.92011	0.46384	0.42831	0.41083	0.40165*	0.40378**	0.44719	0.44671
15	0.91605	0.45537	0.42102	0.40396	0.39506*	0.39795**	0.44009	0.43949

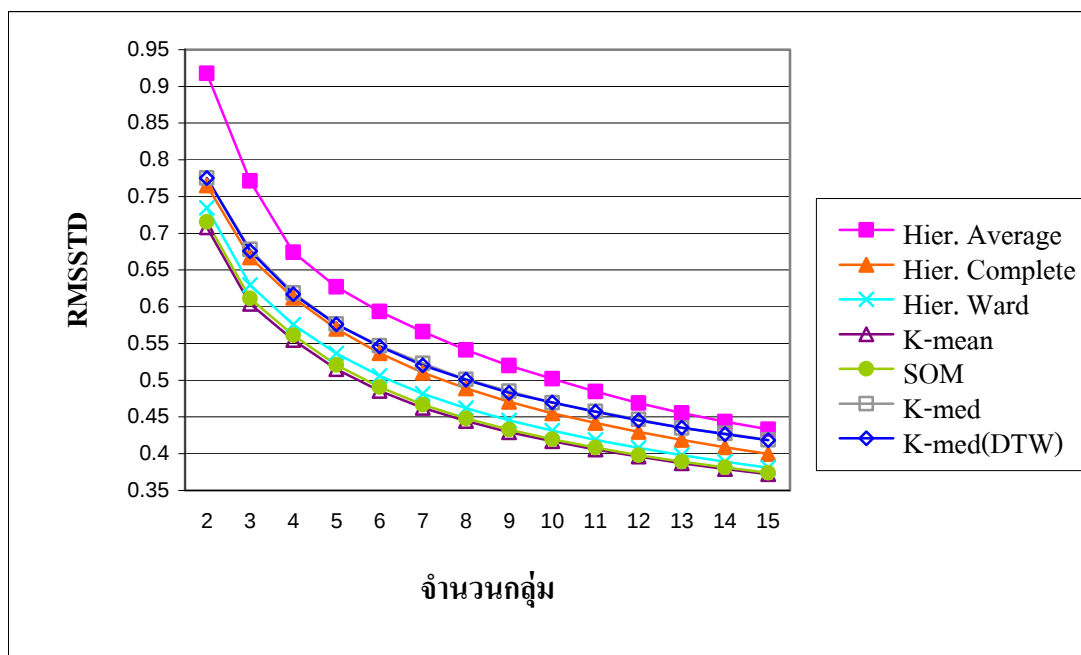
ตารางที่ 66 ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 4 ค่าเฉลี่ยเท่ากับ μ_{73} และค่าความแปรปรวนรวมเท่ากับ Σ_{72}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.98831	0.90609	0.75882	0.72497	0.70065*	0.7078**	0.76175	0.76128
3	0.97953	0.7472	0.65767	0.61657	0.59304*	0.59929**	0.66088	0.66071
4	0.97213	0.64469	0.59644	0.56266	0.54196*	0.54797**	0.60835	0.60832
5	0.96553	0.59798	0.55696	0.52994	0.5132*	0.51836**	0.57363	0.57306
6	0.95958	0.57167	0.531	0.50572	0.48886*	0.49338**	0.54571	0.54598
7	0.95439	0.5509	0.50943	0.48616	0.46987*	0.47403**	0.52614	0.52477
8	0.94925	0.53331	0.4918	0.46992	0.45461*	0.45829**	0.50842	0.50809
9	0.94422	0.51891	0.47717	0.45632	0.44245*	0.44563**	0.4951	0.49348
10	0.93953	0.50554	0.46486	0.44465	0.43204*	0.43474**	0.48208	0.48181
11	0.93462	0.49247	0.45379	0.43447	0.4231*	0.42538**	0.47169	0.47094
12	0.92983	0.48207	0.44382	0.42544	0.41442*	0.41712**	0.46184	0.46148
13	0.92548	0.47202	0.4355	0.41731	0.40734*	0.40981**	0.45261	0.45324
14	0.92126	0.46271	0.42755	0.40989	0.40039*	0.4032**	0.44549	0.44501
15	0.91733	0.45339	0.42012	0.40306	0.39448*	0.39734**	0.43798	0.43885

ตารางที่ 67 ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 5 ค่าเฉลี่ยเท่ากับ μ_{73} และค่าความแปรปรวนรวมเท่ากับ Σ_{73}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.98783	0.90583	0.76089	0.72611	0.70064*	0.70732**	0.76366	0.76439
3	0.97866	0.74792	0.65629	0.61665	0.59321*	0.59938**	0.66081	0.66191
4	0.97126	0.64069	0.59434	0.56256	0.5421*	0.5481**	0.60769	0.60767
5	0.96506	0.59945	0.5582	0.5300	0.51333*	0.51856**	0.57224	0.57155
6	0.95953	0.57088	0.53158	0.50601	0.48903*	0.49409**	0.54575	0.54564
7	0.95406	0.55145	0.51054	0.48648	0.47016*	0.47461**	0.52514	0.52539
8	0.94883	0.53387	0.49291	0.47026	0.45481*	0.45865**	0.5078	0.50808
9	0.94393	0.5189	0.47819	0.45663	0.44263*	0.44593**	0.49341	0.49292
10	0.93919	0.50527	0.46551	0.44499	0.43209*	0.43493**	0.48268	0.48204
11	0.9343	0.4937	0.45451	0.43486	0.42308*	0.42549**	0.47161	0.47129
12	0.93028	0.48298	0.44488	0.42581	0.41505*	0.41727**	0.46192	0.46141
13	0.92596	0.47278	0.43614	0.41767	0.4075*	0.41011**	0.45298	0.45297
14	0.92167	0.46331	0.42799	0.41026	0.40064*	0.40322**	0.44619	0.4457
15	0.91758	0.45466	0.42075	0.40342	0.39455*	0.39743**	0.43889	0.43871

ตารางที่ 53-67 แสดงค่าเฉลี่ย RMSSTD ซึ่งข้อมูลมีขนาด 300 หน่วยตัวอย่าง มีจำนวนตัวแปรอิสระ 3, 5 และ 7 ตัวแปร และการแจกแจงของข้อมูลมีรูปแบบของค่าเฉลี่ยและค่าความแปรปรวนรวมที่ต่างกัน ผลการทดลองที่ได้แสดงให้เห็นว่า ในทุก ๆ ลักษณะของข้อมูล ผลที่ได้มีแนวโน้มไปในทางเดียวกัน ดังนั้นจึงเลือกใช้ข้อมูลในตารางที่ 61 เป็นตัวแทนในการสร้างกราฟได้ดังภาพที่ 29 เพื่อใช้อธิบายผลของข้อมูล ดังนี้



ภาพที่ 29 ค่าเฉลี่ย RMSSTD เมื่อใช้ข้อมูลรูปแบบที่ 4 ค่าเฉลี่ยเท่ากับ μ_{53} และค่าความแปรปรวนรวมเท่ากับ Σ_{52} จำนวนตัวแปรอิสระ 5 ตัวแปร ขนาดตัวอย่างเท่ากับ 300

จากภาพที่ 29 พบว่า เมื่อแบ่งกลุ่มข้อมูลตั้งแต่ 2-15 กลุ่ม อัลกอริทึมจัดกลุ่มแบบเคมีนมีประสิทธิภาพดีที่สุด เนื่องจากให้ค่า RMSSTD ต่ำที่สุด รองลงมาคือ วิธีจัดระบบด้วยตัวเองของโคโฮแน ซึ่งให้ค่าใกล้เคียงกัน และวิธีการจัดกลุ่มแบบลำดับขั้นด้วยวิธี Single linkage มีประสิทธิภาพต่ำสุดในทุกกรณีและมีค่าต่างจากวิธีอื่นมาก จึงไม่นำมารวมในการสร้างกราฟดังภาพที่ 29

ส่วนที่ 2: ค่าความแตกต่างของข้อมูลระหว่างกลุ่ม (RS)

การเปรียบเทียบประสิทธิภาพการแบ่งกลุ่ม โดยพิจารณาจากค่าความแตกต่างของข้อมูลระหว่างกลุ่ม ซึ่งหากค่าความแตกต่างระหว่างกลุ่มมาก หมายถึงการแบ่งกลุ่มที่ดี กลุ่มแต่ละกลุ่มมีลักษณะที่แตกต่างกัน โดยที่ค่า RS นี้อยู่ในช่วง $[0,1]$ การวิจัยครั้งนี้แบ่งตามขนาดตัวอย่างเป็น 3 กรณี คือ

- กำหนดให้
- * แทนเทคนิคการจัดกลุ่มที่ให้ประสิทธิภาพสูงสุด
 - ** แทนเทคนิคการจัดกลุ่มที่ให้ประสิทธิภาพสูงอันดับที่ 2

2.1 ข้อมูลขนาดเล็ก 25 หน่วยตัวอย่าง แบ่งแยกย่อยเป็น 3 กรณีคือ

2.1.1 จำนวนตัวแปรอิสระ 3 ตัวแปร

ตารางที่ 68 ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 1 ค่าเฉลี่ยเท่ากับ μ_{31} และค่าความแปรปรวนรวมเท่ากับ Σ_{31}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.24674	0.4423	0.51988	0.54591	0.5641*	0.56197**	0.48311	0.47117
3	0.41399	0.67372	0.70016	0.71408	0.71732*	0.71552**	0.65302	0.64073
4	0.54029	0.76844	0.78418	0.7931*	0.78848**	0.7781	0.73815	0.73111
5	0.64159	0.82198	0.83377**	0.84059*	0.8332	0.82071	0.79222	0.78613

ตารางที่ 69 ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 2 ค่าเฉลี่ยเท่ากับ μ_{32} และค่าความแปรปรวนรวมเท่ากับ Σ_{31}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.24341	0.44549	0.52223	0.54823	0.56699*	0.56569**	0.48779	0.47299
3	0.41644	0.67951	0.70475	0.71835	0.72294*	0.72066**	0.65577	0.64624
4	0.54187	0.77137	0.78671	0.79641*	0.79273**	0.78322	0.74341	0.73458
5	0.64138	0.82448	0.83574	0.84298*	0.83635**	0.82519	0.79599	0.79081

ตารางที่ 70 ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 3 ค่าเฉลี่ยเท่ากับ μ_{32} และค่าความแปรปรวนร่วมเท่ากับ Σ_{32}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.22687	0.43028	0.50748	0.53319	0.55183*	0.55052**	0.46441	0.46049
3	0.39551	0.65592	0.68566	0.70092	0.70224*	0.70121**	0.63801	0.6274
4	0.51998	0.75449	0.77108	0.78314*	0.77753**	0.77137	0.73249	0.7211
5	0.62577	0.81307	0.82718**	0.83497*	0.8271	0.81575	0.78942	0.78245

ตารางที่ 71 ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 4 ค่าเฉลี่ยเท่ากับ μ_{33} และค่าความแปรปรวนร่วมเท่ากับ Σ_{32}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.23562	0.42939	0.50392	0.53199	0.5503*	0.54877**	0.46345	0.4609
3	0.3973	0.65794	0.68553	0.70095	0.70486*	0.70209**	0.64186	0.62535
4	0.52109	0.75554	0.77278	0.78453*	0.78122**	0.77377	0.73264	0.71876
5	0.62496	0.81568	0.82744	0.8358*	0.82955**	0.81674	0.78987	0.77846

ตารางที่ 72 ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 5 ค่าเฉลี่ยเท่ากับ μ_{33} และค่าความแปรปรวนร่วมเท่ากับ Σ_{33}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.23601	0.43796	0.50776	0.53511	0.55385*	0.55219**	0.4667	0.46088
3	0.39851	0.66024	0.68896	0.70401	0.70773*	0.70481**	0.6460	0.6284
4	0.52913	0.75585	0.7749	0.78552*	0.78212**	0.77265	0.73392	0.72514
5	0.62668	0.81612	0.8286	0.83623*	0.8299**	0.81764	0.79031	0.78244

2.1.2 จำนวนตัวแปรอิสระ 5 ตัวแปร

ตารางที่ 73 ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 1 ค่าเฉลี่ยเท่ากับ μ_{51} และค่าความแปรปรวนร่วมเท่ากับ Σ_{51}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.2539	0.44232	0.51412	0.54403	0.56111*	0.56024**	0.48489	0.47593
3	0.42553	0.67573	0.70383	0.71839	0.72144*	0.71937**	0.65786	0.64579
4	0.54133	0.77172	0.78655	0.79491*	0.78918**	0.78048	0.74148	0.73136
5	0.63798	0.82247	0.83357**	0.84001*	0.83264	0.81738	0.79634	0.78524

ตารางที่ 74 ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 2 ค่าเฉลี่ยเท่ากับ μ_{52} และค่าความแปรปรวนร่วมเท่ากับ Σ_{51}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.25259	0.44312	0.51903	0.54626	0.56403*	0.56271**	0.48995	0.47744
3	0.41818	0.67882	0.70574	0.71961	0.72406*	0.72211**	0.66144	0.65185
4	0.54172	0.77346	0.78725	0.79575*	0.79084**	0.78087	0.74368	0.73402
5	0.63717	0.82302	0.83435**	0.84104*	0.83335	0.81834	0.79513	0.78548

ตารางที่ 75 ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 3 ค่าเฉลี่ยเท่ากับ μ_{52} และค่าความแปรปรวนร่วมเท่ากับ Σ_{52}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.21437	0.40727	0.48223	0.50403	0.52237*	0.52095**	0.43456	0.43319
3	0.36969	0.62011	0.65454	0.67025	0.67143*	0.66901**	0.60968	0.59743
4	0.49375	0.72278	0.74395	0.75586*	0.75198**	0.74368	0.70146	0.69284
5	0.58924	0.78689	0.80291	0.81081*	0.80307**	0.79062	0.76146	0.75316

ตารางที่ 76 ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 4 ค่าเฉลี่ยเท่ากับ μ_{53} และค่าความแปรปรวนร่วมเท่ากับ Σ_{52}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.21467	0.40288	0.47903	0.50332	0.5205*	0.51923**	0.43018	0.42818
3	0.36746	0.61472	0.65393	0.66944	0.67027*	0.66982**	0.60778	0.59151
4	0.48988	0.72157	0.7435	0.7554*	0.75046**	0.74465	0.70038	0.68859
5	0.58792	0.78625	0.80074	0.80954*	0.80284**	0.78904	0.75840	0.75144

ตารางที่ 77 ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 5 ค่าเฉลี่ยเท่ากับ μ_{53} และค่าความแปรปรวนร่วมเท่ากับ Σ_{53}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.21512	0.39789	0.47753	0.50516	0.52334*	0.52179**	0.43147	0.4295
3	0.36911	0.61511	0.65237	0.66897	0.67209*	0.66887**	0.60777	0.59913
4	0.48848	0.72069	0.74347	0.75602*	0.75145**	0.74448	0.70101	0.69006
5	0.58586	0.78734	0.80374	0.81233*	0.80567**	0.79141	0.76423	0.75509

2.1.3 จำนวนตัวแปรอิสระ 7 ตัวแปร

ตารางที่ 78 ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 1 ค่าเฉลี่ยเท่ากับ μ_{71} และค่าความแปรปรวนร่วมเท่ากับ Σ_{71}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.23317	0.41959	0.50097	0.52374	0.54076*	0.53999**	0.45958	0.45392
3	0.38756	0.64607	0.6761	0.69024	0.69427*	0.69187**	0.63471	0.62132
4	0.50829	0.73893	0.75832	0.76735*	0.76187**	0.75149	0.71332	0.70544
5	0.60039	0.79356	0.80671**	0.81389*	0.80546	0.79112	0.76747	0.75815

ตารางที่ 79 ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 2 ค่าเฉลี่ยเท่ากับ μ_{72} และค่าความแปรปรวนร่วมเท่ากับ Σ_{71}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.2371	0.4231	0.49801	0.52444	0.54063*	0.53932**	0.46487	0.45268
3	0.39439	0.64581	0.67908	0.69295	0.69575*	0.69357**	0.62999	0.61815
4	0.51651	0.74299	0.76046	0.76971*	0.76518**	0.75445	0.71562	0.70299
5	0.61017	0.79588	0.80848	0.8157*	0.80855**	0.79264	0.76699	0.75992

ตารางที่ 80 ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 3 ค่าเฉลี่ยเท่ากับ μ_{72} และค่าความแปรปรวนรวมเท่ากับ Σ_{72}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.23248	0.41338	0.48209	0.51166	0.52762*	0.52601**	0.44549	0.43661
3	0.38092	0.63035	0.6625	0.67744	0.67952*	0.67902**	0.61631	0.59958
4	0.49762	0.72683	0.74529	0.75527*	0.75116**	0.74078	0.69916	0.69172
5	0.59337	0.78258	0.79692	0.8044*	0.79751**	0.77889	0.75586	0.74765

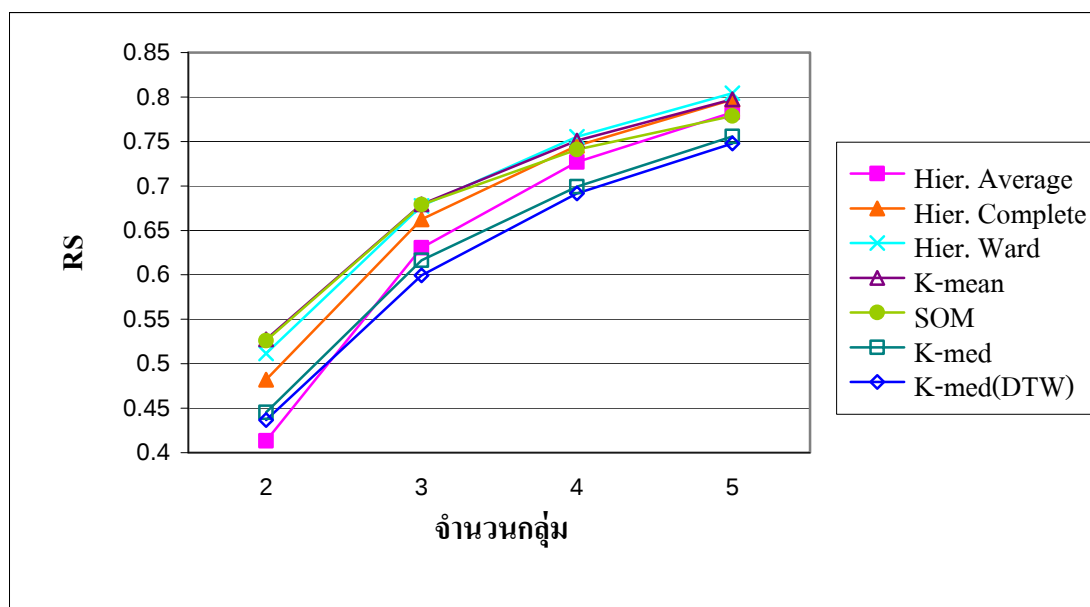
ตารางที่ 81 ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 4 ค่าเฉลี่ยเท่ากับ μ_{73} และค่าความแปรปรวนรวมเท่ากับ Σ_{72}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.22878	0.4086	0.48811	0.51336	0.52928*	0.52839**	0.44972	0.43859
3	0.37694	0.63043	0.66388	0.67755	0.68036*	0.67813**	0.61633	0.60535
4	0.49031	0.72746	0.74694	0.7565*	0.75113**	0.73953	0.70173	0.69168
5	0.58477	0.78333	0.79712	0.80463*	0.79718**	0.78046	0.75668	0.74865

ตารางที่ 82 ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 5 ค่าเฉลี่ยเท่ากับ μ_{73} และค่าความแปรปรวนร่วมเท่ากับ Σ_{73}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.22078	0.4091	0.48896	0.51429	0.53017*	0.52964**	0.45222	0.43778
3	0.37025	0.63211	0.6634	0.67749	0.67932*	0.67784**	0.61516	0.60821
4	0.48811	0.72805	0.74637	0.7562*	0.75091**	0.74011	0.70377	0.69024
5	0.58304	0.78428	0.79718**	0.80457*	0.79697	0.78024	0.75696	0.75094

ตารางที่ 68-82 แสดงค่าเฉลี่ย RS ซึ่งข้อมูลมีขนาด 25 หน่วยตัวอย่าง มีจำนวนตัวแปรอิสระ 3, 5 และ 7 ตัวแปร และการแจกแจงของข้อมูลมีรูปแบบของค่าเฉลี่ยและค่าความแปรปรวนร่วมที่ต่างกัน ผลการทดลองที่ได้แสดงให้เห็นว่า ในทุก ๆ ลักษณะของข้อมูล ผลที่ได้มีแนวโน้มไปในทางเดียวกัน ดังนั้นจึงเลือกใช้ข้อมูลในตารางที่ 80 เป็นตัวแทนในการสร้างกราฟได้ดังภาพที่ 30 เพื่อใช้อธิบายผลของข้อมูล ดังนี้



ภาพที่ 30 ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 3 ค่าเฉลี่ยเท่ากับ μ_{72} และค่าความแปรปรวนร่วมเท่ากับ Σ_{72} จำนวนตัวแปรอิสระ 7 ตัวแปร ขนาดตัวอย่างเท่ากับ 25

จากภาพที่ 30 พบว่า เมื่อแบ่งกลุ่มข้อมูลเป็น 2 และ 3 กลุ่ม อัลกอริทึมจัดกลุ่มแบบเคมีนมีประสิทธิภาพดีที่สุด เนื่องจากให้ค่า RS สูงที่สุด รองลงมาคือ วิธีจัดระบบด้วยตัวเองของโคโฮเนน ซึ่งให้ค่าใกล้เคียงกัน ส่วนการแบ่งกลุ่มข้อมูลเป็น 4 และ 5 กลุ่มนั้น วิธีการจัดกลุ่มแบบลำดับขั้นด้วยวิธี Ward linkage มีประสิทธิภาพดีที่สุด รองลงมาคือ อัลกอริทึมจัดกลุ่มแบบเคมีนและวิธีการจัดกลุ่มแบบลำดับขั้นด้วยวิธี Complete linkage ให้ผลใกล้เคียงกัน และวิธีการจัดกลุ่มแบบลำดับขั้นด้วยวิธี Single linkage มีประสิทธิภาพต่ำสุดในทุกกรณีและมีค่าต่างจากวิธีอื่นมาก จึงไม่นำมารวมในการสร้างกราฟดังภาพที่ 30

2.2 ข้อมูลขนาดกลาง 100 หน่วยตัวอย่าง แบ่งแยกย่อยเป็น 3 กรณีคือ

2.2.1 จำนวนตัวแปรอิสระ 3 ตัวแปร

ตารางที่ 83 ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 1 ค่าเฉลี่ยเท่ากับ μ_{31} และค่าความแปรปรวนรวมเท่ากับ Σ_{31}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.0723	0.29046	0.47029	0.5160	0.54724*	0.53914**	0.46674	0.45293
3	0.1276	0.55834	0.63895	0.67586	0.69705*	0.69329**	0.62689	0.60755
4	0.17304	0.68151	0.72201	0.74605	0.75918*	0.75712**	0.70047	0.68836
5	0.21535	0.73685	0.76664	0.78571	0.7934*	0.79189**	0.74596	0.73477
6	0.25197	0.76894	0.79684	0.81372	0.81855*	0.81745**	0.77896	0.76861
7	0.28737	0.79294	0.82069	0.83532	0.83827*	0.83623**	0.80318	0.7946
8	0.32131	0.81417	0.83941	0.85272**	0.85435*	0.85255	0.82112	0.81395

ตารางที่ 84 ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 2 ค่าเฉลี่ยเท่ากับ μ_{32} และค่าความแปรปรวนรวมเท่ากับ Σ_{31}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.07231	0.27849	0.47021	0.51628	0.54684*	0.53962**	0.45238	0.4469
3	0.12831	0.55713	0.64235	0.676	0.69738*	0.69319**	0.61043	0.61017
4	0.17477	0.6809	0.72162	0.74615	0.75984*	0.75729**	0.68328	0.68295
5	0.2159	0.73611	0.76667	0.78582	0.79334*	0.7925**	0.73021	0.73321
6	0.25293	0.76978	0.79752	0.81374	0.81895*	0.81794**	0.76615	0.76845
7	0.28825	0.79446	0.82061	0.83519	0.83835*	0.83664**	0.79207	0.79099
8	0.32311	0.81531	0.83937	0.85251**	0.85407*	0.85157	0.81313	0.81238

ตารางที่ 85 ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 3 ค่าเฉลี่ยเท่ากับ μ_{32} และค่าความแปรปรวนรวมเท่ากับ Σ_{32}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.071053	0.27772	0.45883	0.50048	0.5319*	0.52408**	0.43335	0.43339
3	0.12309	0.53833	0.61995	0.65626	0.67827*	0.67385**	0.58711	0.58494
4	0.16645	0.65395	0.69852	0.72761	0.74015*	0.73783**	0.66827	0.6707
5	0.20574	0.71302	0.74939	0.77229	0.78132*	0.78055**	0.71946	0.71944
6	0.24518	0.7532	0.78562	0.80448	0.81087*	0.81025**	0.75506	0.75806
7	0.28121	0.78313	0.81265	0.82875	0.83298*	0.83175**	0.78392	0.78441
8	0.3153	0.80813	0.83402	0.84804**	0.85028*	0.84791	0.80554	0.80682

ตารางที่ 86 ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 4 ค่าเฉลี่ยเท่ากับ μ_{33} และค่าความแปรปรวนรวมเท่ากับ Σ_{32}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.072616	0.2703	0.45719	0.4982	0.52987*	0.52304**	0.43355	0.43391
3	0.12686	0.53428	0.61938	0.65357	0.67617*	0.67188**	0.5824	0.58519
4	0.17318	0.65586	0.69729	0.72601	0.73925*	0.73642**	0.66828	0.66896
5	0.21508	0.71192	0.74777	0.77118	0.78038*	0.77989**	0.72008	0.71871
6	0.25494	0.75057	0.78442	0.80408	0.81002*	0.81027**	0.75604	0.75759
7	0.28904	0.78197	0.81193	0.82882	0.83266*	0.83135**	0.78518	0.78517
8	0.32503	0.80836	0.83369	0.8480**	0.85036*	0.84773	0.80708	0.80641

ตารางที่ 87 ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 5 ค่าเฉลี่ยเท่ากับ μ_{33} และค่าความแปรปรวนรวมเท่ากับ Σ_{33}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.072963	0.26164	0.45495	0.49842	0.53116*	0.52353**	0.43083	0.42284
3	0.12687	0.52797	0.61727	0.65374	0.67627*	0.67238**	0.58708	0.58679
4	0.17334	0.65413	0.69568	0.7253	0.73825*	0.73576**	0.66475	0.66667
5	0.21545	0.70999	0.7467	0.77049	0.77996*	0.7794**	0.71625	0.71925
6	0.25399	0.75087	0.78362	0.80338	0.80939*	0.80884**	0.75576	0.75644
7	0.28881	0.78104	0.81138	0.8282	0.83184*	0.83079**	0.78335	0.78379
8	0.32184	0.80709	0.83312	0.84752**	0.84977*	0.84678	0.80447	0.80673

2.2.2 จำนวนตัวแปรอิสระ 5 ตัวแปร

ตารางที่ 88 ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 1 ค่าเฉลี่ยเท่ากับ μ_{s1} และค่าความแปรปรวนรวมเท่ากับ Σ_{s1}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.078864	0.28336	0.47361	0.51791	0.54838*	0.5411**	0.4596	0.46105
3	0.13119	0.56526	0.64063	0.67834	0.69878*	0.69513**	0.61389	0.61247
4	0.17992	0.6881	0.72446	0.74824	0.76066*	0.75879**	0.68861	0.68512
5	0.22146	0.74328	0.77009	0.7869	0.79353*	0.79176**	0.73456	0.73225
6	0.25759	0.77376	0.79801	0.81298	0.8165*	0.81546**	0.76597	0.76706
7	0.29174	0.79721	0.8191	0.83272	0.83433*	0.8328**	0.79054	0.78999
8	0.32472	0.81578	0.83647	0.84864**	0.8495*	0.84717	0.80982	0.80943

ตารางที่ 89 ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 2 ค่าเฉลี่ยเท่ากับ μ_{s2} และค่าความแปรปรวนรวมเท่ากับ Σ_{s1}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.078618	0.28521	0.47497	0.51794	0.54922*	0.54166**	0.46378	0.46339
3	0.13372	0.56312	0.64348	0.67962	0.6992*	0.69586**	0.61285	0.61621
4	0.1805	0.69492	0.72619	0.74904	0.7622*	0.75968**	0.68972	0.69024
5	0.22035	0.74419	0.77095	0.78789	0.79388*	0.79211**	0.73193	0.73506
6	0.25674	0.77566	0.79923	0.8137	0.81748*	0.81547**	0.76696	0.76577
7	0.29486	0.7992	0.82042	0.83346	0.83571*	0.83362**	0.79085	0.78956
8	0.32733	0.81717	0.83763	0.84943**	0.84967*	0.84699	0.81001	0.81068

ตารางที่ 90 ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 3 ค่าเฉลี่ยเท่ากับ μ_{52} และค่าความแปรปรวนรวมเท่ากับ Σ_{52}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.066287	0.24893	0.43161	0.47073	0.50434*	0.49657**	0.39879	0.39791
3	0.11734	0.49288	0.5855	0.61959	0.64296*	0.63802**	0.55201	0.55056
4	0.16068	0.60613	0.66156	0.6913	0.7046*	0.70232**	0.6326	0.63181
5	0.198	0.67133	0.71324	0.73868	0.74929*	0.7483**	0.68726	0.68642
6	0.23145	0.71432	0.75218	0.77301	0.78103*	0.77994**	0.72333	0.72598
7	0.26735	0.74806	0.78174	0.79914	0.80452*	0.80254**	0.75317	0.75378
8	0.29699	0.77592	0.80483	0.81949	0.82283*	0.81955**	0.77428	0.77692

ตารางที่ 91 ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 4 ค่าเฉลี่ยเท่ากับ μ_{53} และค่าความแปรปรวนรวมเท่ากับ Σ_{52}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.064476	0.25868	0.43064	0.47139	0.50449*	0.49659**	0.40112	0.40536
3	0.11672	0.49273	0.58564	0.62098	0.64443*	0.63921**	0.55356	0.55343
4	0.16087	0.60832	0.66416	0.69182	0.70626*	0.70391**	0.63431	0.6343
5	0.20023	0.6709	0.7154	0.73968	0.75034*	0.74951**	0.68794	0.68618
6	0.23788	0.71498	0.75349	0.7744	0.78197*	0.78086**	0.72456	0.72511
7	0.26949	0.74978	0.78294	0.80033	0.80542*	0.8034**	0.75273	0.75268
8	0.30129	0.77822	0.8053	0.82033**	0.82353*	0.82022	0.77574	0.77619

ตารางที่ 92 ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 5 ค่าเฉลี่ยเท่ากับ μ_{53} และค่าความแปรปรวนร่วมเท่ากับ Σ_{53}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.067118	0.2429	0.42675	0.47318	0.5051*	0.4981**	0.39657	0.39979
3	0.11785	0.48865	0.58469	0.62204	0.64393*	0.63904**	0.55204	0.55317
4	0.16003	0.60992	0.66401	0.69372	0.7069*	0.70453**	0.63546	0.63378
5	0.19873	0.67137	0.71683	0.74253	0.75293*	0.75174**	0.68798	0.68848
6	0.23595	0.71911	0.75682	0.77764	0.7851*	0.7835**	0.72658	0.72816
7	0.27162	0.75293	0.78633	0.80383	0.80918*	0.8067**	0.75648	0.75624
8	0.30457	0.78126	0.80919	0.8243**	0.82733*	0.82423	0.78153	0.77953

2.2.3 จำนวนตัวแปรอิสระ 7 ตัวแปร

ตารางที่ 93 ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 1 ค่าเฉลี่ยเท่ากับ μ_{71} และค่าความแปรปรวนร่วมเท่ากับ Σ_{71}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.071685	0.26749	0.45269	0.49254	0.5238*	0.51698**	0.43484	0.43891
3	0.12404	0.53518	0.61419	0.64881	0.66885*	0.66529**	0.58154	0.58182
4	0.16745	0.6594	0.69248	0.71649	0.72944*	0.7275**	0.65521	0.65766
5	0.2034	0.7086	0.73694	0.75498	0.76188*	0.76003**	0.70073	0.70218
6	0.23855	0.74064	0.76615	0.78166	0.78624*	0.7847**	0.73273	0.73698
7	0.27219	0.76343	0.78807	0.80235	0.80543*	0.80238**	0.75655	0.75986
8	0.30476	0.78321	0.80617	0.81927**	0.82068*	0.81744	0.7782	0.7787

ตารางที่ 94 ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 2 ค่าเฉลี่ยเท่ากับ μ_{72} และค่าความแปรปรวนรวมเท่ากับ Σ_{71}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.072984	0.2658	0.45458	0.49518	0.52627*	0.51939**	0.43451	0.43713
3	0.12666	0.53918	0.61341	0.65047	0.66969*	0.66645**	0.58218	0.58676
4	0.1669	0.65633	0.69336	0.71801	0.73117*	0.72832**	0.65523	0.65983
5	0.2042	0.70984	0.73864	0.75621	0.76322*	0.7614**	0.70086	0.70636
6	0.24253	0.74151	0.76708	0.78266	0.7870*	0.7851**	0.73419	0.73543
7	0.27354	0.76605	0.78925	0.80323	0.80656*	0.80389**	0.75932	0.76136
8	0.30342	0.7847	0.8075	0.81995**	0.82213*	0.81804	0.77914	0.78037

ตารางที่ 95 ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 3 ค่าเฉลี่ยเท่ากับ μ_{72} และค่าความแปรปรวนรวมเท่ากับ Σ_{72}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.067876	0.26275	0.44416	0.48444	0.51364*	0.50633**	0.42207	0.41861
3	0.11885	0.52266	0.60164	0.63463	0.65523*	0.65138**	0.56658	0.56589
4	0.15945	0.63921	0.67861	0.70138	0.71452*	0.71101**	0.64158	0.64002
5	0.19765	0.69101	0.72269	0.74118	0.74958*	0.74827**	0.68754	0.68899
6	0.23309	0.72514	0.7542	0.76975	0.77558*	0.77438**	0.72066	0.72083
7	0.2634	0.75079	0.77822	0.79185	0.79606*	0.79313**	0.74714	0.74749
8	0.29415	0.77261	0.79689	0.80956**	0.81202*	0.80837	0.76698	0.7668

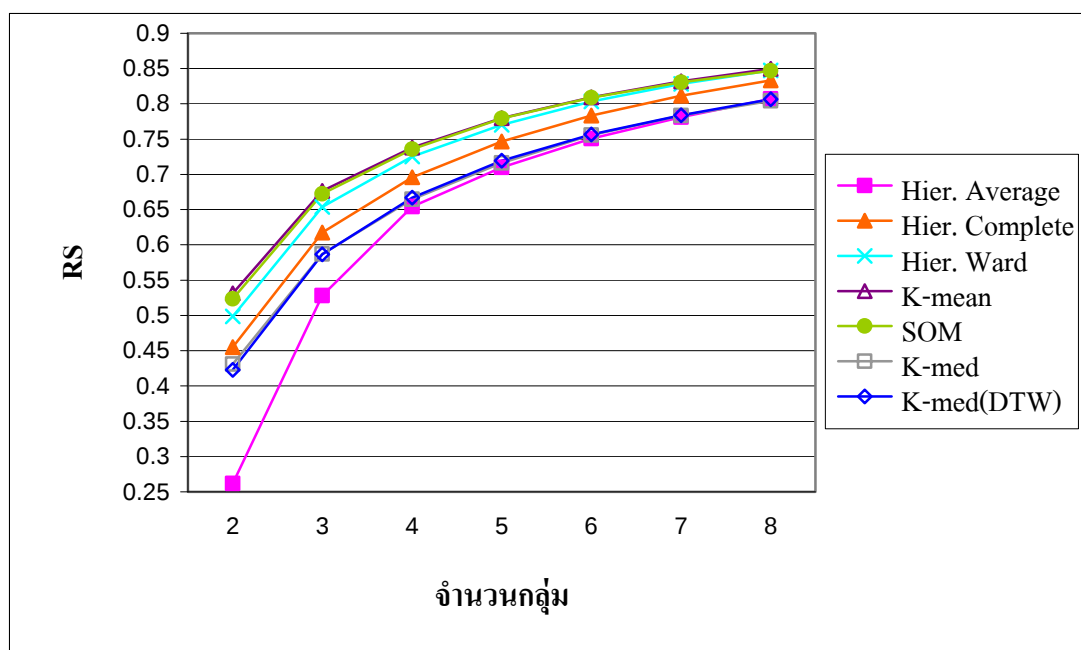
ตารางที่ 96 ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 4 ค่าเฉลี่ยเท่ากับ μ_3 และค่าความแปรปรวนรวมเท่ากับ Σ_{72}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.069556	0.26267	0.44506	0.48433	0.51479*	0.50751**	0.42295	0.42248
3	0.12008	0.5197	0.60102	0.63596	0.65663*	0.65325**	0.56955	0.56784
4	0.16326	0.63753	0.67919	0.70294	0.716*	0.71212**	0.64417	0.64448
5	0.19892	0.69348	0.72268	0.74295	0.75086*	0.74941**	0.6909	0.68852
6	0.23361	0.72706	0.75437	0.77146	0.77687*	0.77519**	0.72287	0.72252
7	0.26418	0.75345	0.77905	0.79332	0.7976*	0.79401**	0.74795	0.74776
8	0.29219	0.7734	0.79772	0.81077**	0.81322*	0.80927	0.76826	0.76893

ตารางที่ 97 ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 5 ค่าเฉลี่ยเท่ากับ μ_3 และค่าความแปรปรวนรวมเท่ากับ Σ_{73}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.070745	0.25811	0.43612	0.48273	0.51266*	0.50521**	0.41777	0.41593
3	0.12129	0.51531	0.59857	0.63372	0.65483*	0.65139**	0.56361	0.56464
4	0.16254	0.63483	0.67698	0.70197	0.71523*	0.71225**	0.64206	0.63986
5	0.19903	0.68756	0.7231	0.74197	0.74989*	0.74894**	0.68795	0.68812
6	0.23314	0.72419	0.75399	0.77035	0.77674*	0.7746**	0.72173	0.72104
7	0.26508	0.75107	0.77744	0.79241	0.79618*	0.79386**	0.74663	0.74571
8	0.2936	0.77267	0.7971	0.80991**	0.81228*	0.80873	0.76742	0.7668

ตารางที่ 83-97 แสดงค่าเฉลี่ย RS ซึ่งข้อมูลมีขนาด 100 หน่วยตัวอย่าง มีจำนวนตัวแปรอิสระ 3, 5 และ 7 ตัวแปร และการแจกแจงของข้อมูลมีรูปแบบของค่าเฉลี่ยและค่าความแปรปรวนรวมที่ต่างกัน ผลการทดลองที่ได้แสดงให้เห็นว่า ในทุก ๆ ลักษณะของข้อมูล ผลที่ได้มีแนวโน้มไปในทางเดียวกัน ดังนั้นจึงเลือกใช้ข้อมูลในตารางที่ 87 เป็นตัวแทนในการสร้างกราฟได้ดังภาพที่ 31 เพื่อใช้อธิบายผลของข้อมูล ดังนี้



ภาพที่ 31 ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 5 ค่าเฉลี่ยเท่ากับ μ_{33} และค่าความแปรปรวนรวมเท่ากับ Σ_{33} จำนวนตัวแปรอิสระ 3 ตัวแปร ขนาดตัวอย่างเท่ากับ 100

จากภาพที่ 31 พบว่า เมื่อแบ่งกลุ่มข้อมูลตั้งแต่ 2-8 กลุ่ม อัลกอริทึมจัดกลุ่มแบบเคมีนมีประสิทธิภาพดีที่สุด เนื่องจากให้ค่า RS สูงที่สุด รองลงมาคือ วิธีจัดระบบด้วยตัวเองของโคโฮเนน ซึ่งให้ค่าใกล้เคียงกัน ยกเว้นเมื่อแบ่งข้อมูลเป็น 8 กลุ่มนั้น วิธีการจัดกลุ่มแบบลำดับขั้นด้วยวิธี Ward linkage มีประสิทธิภาพดีกว่าวิธีจัดระบบด้วยตัวเองของโคโฮเนนเล็กน้อย และวิธีการจัดกลุ่มแบบลำดับขั้นด้วยวิธี Single linkage มีประสิทธิภาพต่ำสุดในทุกกรณีและมีค่าต่างจากวิธีอื่นมาก จึงไม่นำมารวมในการสร้างกราฟดังภาพที่ 31

2.3 ข้อมูลขนาดกลาง 300 หน่วยตัวอย่าง แบ่งแยกย่อยเป็น 3 กรณีคือ

2.3.1 จำนวนตัวแปรอิสระ 3 ตัวแปร

ตารางที่ 98 ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 1 ค่าเฉลี่ยเท่ากับ μ_{31} และค่าความแปรปรวนรวมเท่ากับ Σ_{31}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.030156	0.18482	0.44453	0.50278	0.54166*	0.53124**	0.44338	0.44746
3	0.052104	0.4659	0.60394	0.66194	0.69009*	0.68279**	0.59871	0.59912
4	0.072446	0.63191	0.68918	0.72971	0.75233*	0.74658**	0.67367	0.67282
5	0.089056	0.68965	0.73651	0.76712	0.78363*	0.77886**	0.71889	0.72098
6	0.10576	0.72367	0.76625	0.7922	0.80672*	0.80282**	0.75206	0.75261
7	0.12032	0.74758	0.78804	0.8117	0.8248*	0.82149**	0.77525	0.77553
8	0.13468	0.76649	0.80562	0.82754	0.83899*	0.83675**	0.79414	0.79453
9	0.1486	0.78291	0.82029	0.84075	0.85108*	0.84921**	0.81001	0.81029
10	0.16204	0.79659	0.83271	0.852	0.8611*	0.8597**	0.82303	0.82212
11	0.17575	0.80932	0.84384	0.86175	0.86985*	0.86882**	0.83369	0.83362
12	0.18771	0.82018	0.85354	0.87029	0.87721*	0.87664**	0.84321	0.84369
13	0.20076	0.83068	0.86227	0.87785	0.88424*	0.88338**	0.85172	0.852
14	0.21248	0.84008	0.86993	0.88463	0.89042*	0.88951**	0.85958	0.85924
15	0.22469	0.84917	0.87692	0.8907	0.89575*	0.89442**	0.86562	0.8657

ตารางที่ 99 ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 2 ค่าเฉลี่ยเท่ากับ μ_{32} และค่าความแปรปรวนรวมเท่ากับ Σ_{31}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.029496	0.18593	0.44315	0.50576	0.54169*	0.53155**	0.44611	0.45268
3	0.051304	0.47263	0.60539	0.66051	0.68996*	0.68283**	0.59483	0.60246
4	0.070793	0.62798	0.69033	0.73006	0.75254*	0.7466**	0.67258	0.67103
5	0.088338	0.69059	0.73577	0.76743	0.78395*	0.77891**	0.72043	0.72008
6	0.10503	0.72407	0.7657	0.79265	0.80688*	0.80302**	0.75018	0.75031
7	0.12122	0.74693	0.78785	0.81206	0.82469*	0.82183**	0.77518	0.77383
8	0.13652	0.76681	0.80552	0.8278	0.8393*	0.83679**	0.79379	0.79456
9	0.15117	0.78244	0.82039	0.84098	0.8513*	0.84947**	0.80938	0.80943
10	0.16418	0.79578	0.83305	0.85221	0.86158*	0.85997**	0.82148	0.82188
11	0.17723	0.80863	0.84408	0.86195	0.8702*	0.86908**	0.83287	0.83347
12	0.19071	0.82012	0.85372	0.87053	0.87796*	0.87685**	0.84254	0.84289
13	0.205	0.83003	0.86237	0.87813	0.88443*	0.88375**	0.85119	0.85142
14	0.21702	0.8397	0.86996	0.88487	0.89057*	0.88952**	0.85933	0.85908
15	0.22903	0.84881	0.87705	0.89087	0.89579*	0.89482**	0.8651	0.86614

ตารางที่ 100 ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 3 ค่าเฉลี่ยเท่ากับ μ_{32} และค่าความแปรปรวนรวมเท่ากับ Σ_{32}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.027926	0.17402	0.43622	0.49058	0.52804*	0.51774**	0.42588	0.42724
3	0.049823	0.43995	0.58962	0.64268	0.67245*	0.66437**	0.58267	0.57986
4	0.069615	0.59687	0.66751	0.70979	0.73246*	0.7251**	0.65900	0.65691
5	0.088083	0.65632	0.71576	0.75064	0.76957*	0.7647**	0.70641	0.70798
6	0.10526	0.69594	0.75023	0.78098	0.79852*	0.79414**	0.74104	0.74142
7	0.12086	0.72635	0.7769	0.80442	0.82014*	0.81686**	0.76647	0.7688
8	0.1346	0.7511	0.79839	0.82291	0.83627*	0.83385**	0.78709	0.78766
9	0.14897	0.77165	0.8157	0.83756	0.84908*	0.84712**	0.80461	0.80462
10	0.1628	0.78958	0.8297	0.84967	0.85946*	0.85803**	0.81800	0.81815
11	0.17547	0.80514	0.84158	0.8599	0.86852*	0.86739**	0.8299	0.8306
12	0.1882	0.8188	0.85181	0.86878	0.87606*	0.87534**	0.84048	0.83993
13	0.20071	0.83113	0.86061	0.87655	0.88308*	0.88208**	0.84846	0.84848
14	0.21245	0.84194	0.86867	0.88343	0.88909*	0.88812**	0.85701	0.85636
15	0.22477	0.85097	0.87576	0.8896	0.89444*	0.89346**	0.86348	0.86367

ตารางที่ 101 ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 4 ค่าเฉลี่ยเท่ากับ μ_{33} และค่าความแปรปรวนรวมเท่ากับ Σ_{32}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.027502	0.18142	0.43557	0.48976	0.5280*	0.51769**	0.43277	0.43223
3	0.049575	0.44217	0.58796	0.64074	0.67187*	0.66361**	0.57962	0.57936
4	0.068726	0.59606	0.66395	0.70856	0.73138*	0.72447**	0.65886	0.65941
5	0.086503	0.6557	0.71283	0.74985	0.76921*	0.76368**	0.70442	0.70529
6	0.10274	0.69438	0.74879	0.78022	0.79809*	0.79374**	0.74043	0.74158
7	0.11826	0.72412	0.77654	0.80364	0.81955*	0.81613**	0.7655	0.76623
8	0.13299	0.7483	0.79766	0.82199	0.83582*	0.83333**	0.78789	0.78814
9	0.1475	0.77008	0.81483	0.83672	0.84852*	0.8468**	0.80336	0.80408
10	0.16139	0.78884	0.82917	0.84887	0.85925*	0.85781**	0.81729	0.8176
11	0.17392	0.80391	0.84147	0.85917	0.8683*	0.86692**	0.8291	0.8297
12	0.18728	0.81773	0.85149	0.86812	0.87582*	0.87465**	0.83976	0.83975
13	0.20092	0.83008	0.86047	0.87594	0.88282*	0.88166**	0.84737	0.84751
14	0.21333	0.84044	0.86835	0.88289	0.88865*	0.88773**	0.85613	0.85639
15	0.22547	0.8498	0.8755	0.88913	0.89386*	0.89306**	0.86287	0.86317

ตารางที่ 102 ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 5 ค่าเฉลี่ยเท่ากับ μ_{33} และค่าความแปรปรวนรวมเท่ากับ Σ_{33}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.028569	0.1765	0.4301	0.48951	0.52743*	0.51641**	0.42521	0.42304
3	0.050556	0.446	0.5857	0.64153	0.67178*	0.66345**	0.57364	0.57424
4	0.069138	0.59671	0.6675	0.70859	0.73174*	0.72439**	0.65445	0.65463
5	0.087005	0.65677	0.71392	0.74999	0.7692*	0.76393**	0.70572	0.70653
6	0.10404	0.69456	0.74939	0.78061	0.79793*	0.79393**	0.73872	0.74048
7	0.11897	0.72524	0.77662	0.80388	0.81993*	0.81637**	0.76563	0.76606
8	0.13327	0.74862	0.79794	0.82225	0.83601*	0.83343**	0.78715	0.78778
9	0.14731	0.76911	0.81457	0.83679	0.84889*	0.84667**	0.80289	0.80358
10	0.16175	0.78746	0.82888	0.84893	0.85898*	0.85758**	0.8168	0.81719
11	0.17484	0.80391	0.84067	0.85923	0.86812*	0.86672**	0.82888	0.82952
12	0.18807	0.81844	0.85086	0.86813	0.87576*	0.87471**	0.8389	0.83948
13	0.20083	0.83003	0.85966	0.87594	0.88235*	0.88158**	0.84829	0.84869
14	0.21224	0.84074	0.8679	0.88287	0.88867*	0.88758**	0.85627	0.85653
15	0.22427	0.84999	0.87506	0.88904	0.89419*	0.89287**	0.8635	0.86337

2.3.2 จำนวนตัวแปรอิสระ 5 ตัวแปร

ตารางที่ 103 ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 1 ค่าเฉลี่ยเท่ากับ μ_{s1} และค่าความแปรปรวนรวมเท่ากับ Σ_{s1}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.030439	0.18611	0.45045	0.50636	0.54367*	0.53397**	0.45029	0.45417
3	0.053989	0.46649	0.61056	0.66524	0.69207*	0.68531**	0.60401	0.60356
4	0.073646	0.63854	0.69509	0.73357	0.75463*	0.7493**	0.67675	0.67408
5	0.092018	0.70079	0.74175	0.77023	0.78628*	0.78123**	0.72007	0.71869
6	0.10792	0.73289	0.76981	0.7933	0.80568*	0.80193**	0.74882	0.75045
7	0.12304	0.75464	0.78965	0.8107	0.82168*	0.81863**	0.77147	0.7734
8	0.13766	0.77072	0.80562	0.82498	0.83482*	0.83228**	0.79203	0.7903
9	0.15081	0.78463	0.81884	0.83701	0.84606*	0.8441**	0.80547	0.8056
10	0.16345	0.79754	0.83027	0.8474	0.85549*	0.85403**	0.81796	0.81832
11	0.17614	0.80874	0.84029	0.85646	0.86392*	0.86263**	0.82825	0.82736
12	0.18887	0.81864	0.84912	0.86436	0.87095*	0.86989**	0.83704	0.83684
13	0.20066	0.82781	0.85723	0.87135	0.87735*	0.87635**	0.84548	0.84588
14	0.21133	0.8364	0.86432	0.87761	0.88315*	0.88163**	0.8523	0.85256
15	0.22277	0.84441	0.87067	0.88325	0.88811*	0.8865**	0.85834	0.8587

ตารางที่ 104 ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 2 ค่าเฉลี่ยเท่ากับ μ_{52} และค่าความแปรปรวนรวมเท่ากับ Σ_{51}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.029852	0.19287	0.44904	0.50768	0.54402*	0.53375**	0.45202	0.45362
3	0.052659	0.47057	0.6111	0.66506	0.69239*	0.68534**	0.60543	0.60571
4	0.073431	0.64388	0.69303	0.73379	0.75473*	0.74962**	0.67782	0.67701
5	0.091015	0.70128	0.74037	0.77022	0.7864*	0.7813**	0.72148	0.72179
6	0.10674	0.7342	0.76988	0.79332	0.80639*	0.80216**	0.75096	0.75129
7	0.12154	0.75492	0.78951	0.81074	0.82173*	0.81875**	0.77357	0.77435
8	0.13507	0.77179	0.80568	0.82504	0.83513*	0.83264**	0.79309	0.79273
9	0.14907	0.78613	0.81877	0.83708	0.84605*	0.84418**	0.80688	0.80681
10	0.16294	0.79813	0.83041	0.84738	0.8557*	0.85428**	0.81883	0.81867
11	0.17586	0.80883	0.84048	0.85636	0.86405*	0.86275**	0.8279	0.82798
12	0.18814	0.81876	0.84935	0.86428	0.87129*	0.87008**	0.83731	0.83758
13	0.20041	0.82811	0.85732	0.87134	0.87763*	0.87652**	0.84582	0.84566
14	0.21195	0.83674	0.86439	0.87766	0.8831*	0.88185**	0.85191	0.85229
15	0.22306	0.84463	0.87074	0.88334	0.88827*	0.8868**	0.85911	0.85891

ตารางที่ 105 ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 3 ค่าเฉลี่ยเท่ากับ μ_{52} และค่าความแปรปรวนรวมเท่ากับ Σ_{52}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.025818	0.16222	0.4086	0.46015	0.49891*	0.48825**	0.39877	0.3981
3	0.046843	0.40645	0.55495	0.60518	0.63614*	0.62715**	0.54297	0.542
4	0.065451	0.54284	0.62895	0.67095	0.69435*	0.68701**	0.61622	0.61717
5	0.08094	0.60874	0.67764	0.71463	0.73702*	0.7311**	0.66865	0.66964
6	0.096616	0.64853	0.7152	0.74715	0.76733*	0.76235**	0.70632	0.70723
7	0.11058	0.68137	0.74317	0.77179	0.7896*	0.78541**	0.73213	0.73364
8	0.12311	0.71098	0.76527	0.79093	0.80638*	0.80323**	0.75365	0.7541
9	0.1365	0.73453	0.78328	0.80628	0.81967*	0.8169**	0.77045	0.77201
10	0.14831	0.75393	0.7985	0.81903	0.83077*	0.82859**	0.78483	0.78537
11	0.1609	0.77126	0.81117	0.82983	0.84003*	0.83809**	0.79708	0.7979
12	0.17302	0.78697	0.82198	0.83908	0.84835*	0.84667**	0.80766	0.80873
13	0.18557	0.79994	0.8312	0.84727	0.85544*	0.85396**	0.81684	0.81745
14	0.19717	0.81046	0.8394	0.85457	0.86182*	0.86036**	0.82506	0.82608
15	0.20775	0.82028	0.847	0.86108	0.86758*	0.86592**	0.83245	0.83319

ตารางที่ 106 ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 4 ค่าเฉลี่ยเท่ากับ μ_{53} และค่าความแปรปรวนรวมเท่ากับ Σ_{52}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.026198	0.15355	0.41393	0.46149	0.49993*	0.4895**	0.39695	0.39707
3	0.046085	0.39797	0.55654	0.60574	0.6375*	0.62865**	0.5413	0.54412
4	0.064662	0.54463	0.62817	0.6717	0.69507*	0.68769**	0.61939	0.62139
5	0.080993	0.60942	0.67871	0.71573	0.7380*	0.73227**	0.67099	0.67164
6	0.096191	0.65195	0.71618	0.74807	0.76792*	0.76286**	0.70512	0.70599
7	0.11093	0.68448	0.74418	0.77245	0.79048*	0.78639**	0.73113	0.73384
8	0.12463	0.71266	0.76615	0.79146	0.8070*	0.80391**	0.7542	0.75446
9	0.13809	0.73583	0.78387	0.80668	0.82034*	0.81759**	0.77063	0.77255
10	0.15162	0.7546	0.79878	0.81945	0.83114*	0.82895**	0.78569	0.78546
11	0.16395	0.77238	0.81111	0.8303	0.84065*	0.8388**	0.79729	0.79744
12	0.17564	0.78728	0.82197	0.83967	0.84873*	0.84732**	0.80733	0.80871
13	0.18725	0.80045	0.83151	0.84787	0.85595*	0.85438**	0.81814	0.81782
14	0.19785	0.81114	0.83982	0.85513	0.86233*	0.86074**	0.82468	0.82565
15	0.20938	0.82055	0.8474	0.8616	0.86797*	0.86645**	0.83238	0.83292

ตารางที่ 107 ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 5 ค่าเฉลี่ยเท่ากับ μ_{53} และค่าความแปรปรวนรวมเท่ากับ Σ_{53}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.026031	0.16163	0.40984	0.46436	0.50251*	0.49176**	0.38853	0.39582
3	0.047288	0.39942	0.55452	0.60746	0.6393*	0.62983**	0.54448	0.54456
4	0.066246	0.54477	0.63017	0.67407	0.69775*	0.69022**	0.62394	0.62549
5	0.082221	0.61124	0.68127	0.71908	0.7417*	0.73556**	0.67398	0.67399
6	0.097173	0.65327	0.71841	0.75224	0.77234*	0.76706**	0.70994	0.711
7	0.11199	0.68844	0.74746	0.7773	0.79466*	0.79067**	0.73678	0.73697
8	0.12584	0.71426	0.77007	0.79667	0.81177*	0.80868**	0.75953	0.75991
9	0.1391	0.74019	0.78883	0.81225	0.8253*	0.82249**	0.77585	0.77672
10	0.15163	0.76056	0.80392	0.82517	0.83653*	0.83398**	0.78998	0.79125
11	0.16295	0.77823	0.81675	0.83609	0.84581*	0.84398**	0.80221	0.80319
12	0.17489	0.79388	0.82762	0.84544	0.85409*	0.85236**	0.8137	0.81408
13	0.18559	0.80669	0.8367	0.85358	0.86125*	0.85969**	0.8229	0.82331
14	0.19727	0.81777	0.84498	0.86082	0.86755*	0.86613**	0.83103	0.83128
15	0.20874	0.82804	0.85249	0.86728	0.87319*	0.87157**	0.83904	0.83903

2.3.3 จำนวนตัวแปรอิสระ 7 ตัวแปร

ตารางที่ 108 ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 1 ค่าเฉลี่ยเท่ากับ μ_{71} และค่าความแปรปรวนรวมเท่ากับ Σ_{71}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.027789	0.18288	0.42451	0.48585	0.52102*	0.51107**	0.42826	0.43482
3	0.049828	0.45745	0.58212	0.63696	0.66422*	0.65752**	0.57727	0.57572
4	0.068171	0.61183	0.66501	0.70244	0.72444*	0.71845**	0.6436	0.64764
5	0.083614	0.66686	0.70932	0.73817	0.7549*	0.7494**	0.6878	0.69115
6	0.098759	0.69769	0.73758	0.76149	0.77572*	0.7713**	0.71786	0.7205
7	0.11291	0.71988	0.75822	0.77948	0.79244*	0.78873**	0.74254	0.74282
8	0.12539	0.7363	0.774	0.79422	0.8062*	0.80321**	0.75741	0.75984
9	0.13803	0.74954	0.78781	0.80667	0.8176*	0.8148**	0.77273	0.77495
10	0.14993	0.76227	0.79964	0.81734	0.82727*	0.82523**	0.78575	0.78709
11	0.16182	0.77398	0.80988	0.82663	0.83584*	0.83398**	0.79582	0.79706
12	0.17292	0.78472	0.81923	0.8348	0.84333*	0.84167**	0.80556	0.80587
13	0.18404	0.79437	0.8273	0.84207	0.85006*	0.84832**	0.81247	0.81431
14	0.19513	0.80321	0.83457	0.84861	0.85568*	0.85393**	0.81953	0.82107
15	0.20524	0.81119	0.84112	0.85451	0.86108*	0.85915**	0.82616	0.82706

ตารางที่ 109 ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 2 ค่าเฉลี่ยเท่ากับ μ_{72} และค่าความแปรปรวนรวมเท่ากับ Σ_{71}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.028538	0.17027	0.42887	0.48578	0.52188*	0.51222**	0.42907	0.43268
3	0.048924	0.4498	0.58336	0.63632	0.66444*	0.65767**	0.57739	0.5796
4	0.066107	0.6100	0.66351	0.70193	0.72421*	0.7182**	0.64401	0.64568
5	0.08354	0.66706	0.7097	0.73764	0.75419*	0.74896**	0.68841	0.69047
6	0.097385	0.6970	0.73752	0.76101	0.77529*	0.77092**	0.71823	0.72053
7	0.11075	0.71839	0.75684	0.77902	0.792*	0.78833**	0.7404	0.74231
8	0.12333	0.73643	0.77317	0.79377	0.80559*	0.80261**	0.75838	0.75974
9	0.13424	0.75085	0.7869	0.80627	0.81739*	0.81438**	0.77278	0.77428
10	0.14637	0.76327	0.79881	0.81703	0.82739*	0.82471**	0.78502	0.78583
11	0.15781	0.77442	0.80908	0.82634	0.83563*	0.83379**	0.79572	0.79667
12	0.16975	0.78465	0.81844	0.83455	0.84301*	0.84115**	0.80459	0.80604
13	0.18089	0.79452	0.82648	0.84178	0.84964*	0.84789**	0.81272	0.81388
14	0.19171	0.80329	0.8340	0.84831	0.85536*	0.85346**	0.81993	0.82077
15	0.20226	0.81127	0.84079	0.85421	0.86095*	0.85872**	0.82632	0.82729

ตารางที่ 110 ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 3 ค่าเฉลี่ยเท่ากับ μ_{72} และค่าความแปรปรวนรวมเท่ากับ Σ_{72}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.0272	0.1748	0.42093	0.47394	0.51011*	0.50043**	0.41698	0.41514
3	0.047487	0.43509	0.56924	0.62048	0.64928*	0.6424**	0.56038	0.55978
4	0.064092	0.58947	0.64709	0.68499	0.70778*	0.70151**	0.62894	0.62885
5	0.079369	0.64348	0.69177	0.72184	0.73919*	0.73389**	0.67299	0.67249
6	0.094848	0.67603	0.72144	0.74772	0.76406*	0.75934**	0.70293	0.7033
7	0.10864	0.70044	0.74482	0.76761	0.78276*	0.77846**	0.72697	0.72771
8	0.1219	0.71973	0.76247	0.7835	0.79738*	0.79375**	0.74613	0.74659
9	0.13502	0.7362	0.7774	0.79641	0.80843*	0.80581**	0.76073	0.76126
10	0.1464	0.75033	0.78968	0.80721	0.81805*	0.81572**	0.77306	0.77342
11	0.15826	0.76202	0.80003	0.81652	0.82636*	0.82416**	0.78348	0.78365
12	0.16947	0.77411	0.80895	0.82469	0.83356*	0.83147**	0.79242	0.79262
13	0.17995	0.78486	0.81714	0.8319	0.84002*	0.8381**	0.80162	0.8015
14	0.18962	0.79378	0.82433	0.83842	0.84555*	0.84391**	0.80842	0.80882
15	0.19952	0.80193	0.83086	0.84433	0.8511*	0.84892**	0.81509	0.8156

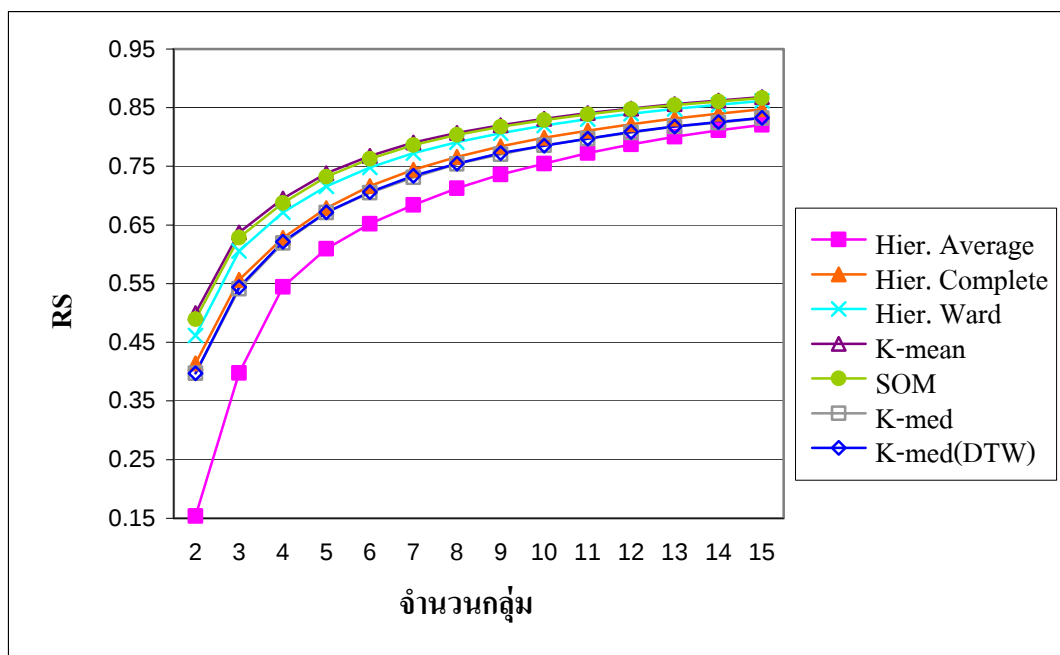
ตารางที่ 111 ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 4 ค่าเฉลี่ยเท่ากับ μ_{73} และค่าความแปรปรวนรวมเท่ากับ Σ_{72}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.02645	0.17344	0.42338	0.47563	0.51056*	0.50044**	0.41801	0.41873
3	0.046823	0.43457	0.56858	0.62209	0.65047*	0.64304**	0.56392	0.56403
4	0.064259	0.58436	0.64691	0.68635	0.70902*	0.70251**	0.63209	0.63219
5	0.07998	0.64553	0.69342	0.72271	0.73996*	0.7347**	0.67416	0.67485
6	0.094321	0.67732	0.72231	0.74833	0.76484*	0.76047**	0.70638	0.70606
7	0.10709	0.70156	0.74533	0.76821	0.78349*	0.77963**	0.72805	0.72944
8	0.11966	0.72141	0.76349	0.78417	0.79802*	0.79472**	0.74696	0.74733
9	0.13189	0.73724	0.77814	0.79718	0.80932*	0.80656**	0.76096	0.76255
10	0.14342	0.7515	0.79017	0.80809	0.81881*	0.81654**	0.77418	0.77442
11	0.1552	0.76504	0.80074	0.8174	0.82683*	0.82497**	0.78454	0.78524
12	0.16668	0.77565	0.81007	0.82552	0.83444*	0.83228**	0.7942	0.79455
13	0.1773	0.78565	0.81776	0.8327	0.8406*	0.83866**	0.80307	0.80248
14	0.18758	0.79475	0.82497	0.83917	0.84654*	0.84437**	0.80985	0.81028
15	0.19729	0.80366	0.83158	0.84502	0.85155*	0.84939**	0.81689	0.81614

ตารางที่ 112 ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 5 ค่าเฉลี่ยเท่ากับ μ_{73} และค่าความแปรปรวนร่วมเท่ากับ Σ_{73}

จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.027383	0.17416	0.42038	0.47397	0.51059*	0.50116**	0.41485	0.41373
3	0.048489	0.43344	0.57024	0.62197	0.65028*	0.64293**	0.56379	0.56229
4	0.065912	0.58993	0.64941	0.68645	0.70888*	0.70236**	0.63295	0.63301
5	0.080876	0.64348	0.692	0.72263	0.73983*	0.73448**	0.67581	0.67664
6	0.094419	0.67815	0.7217	0.74803	0.76467*	0.75977**	0.70635	0.70644
7	0.10771	0.70092	0.74421	0.76789	0.78322*	0.77909**	0.72912	0.72883
8	0.12043	0.7208	0.76239	0.78385	0.79783*	0.7944**	0.7476	0.7473
9	0.13243	0.73723	0.77716	0.79689	0.80915*	0.8063**	0.76253	0.76304
10	0.14403	0.75177	0.78956	0.80777	0.81876*	0.81637**	0.77355	0.77419
11	0.1558	0.76382	0.8001	0.81706	0.82685*	0.82487**	0.78462	0.7849
12	0.16591	0.77477	0.80913	0.8252	0.83392*	0.83215**	0.79411	0.79457
13	0.17647	0.78494	0.8172	0.8324	0.84048*	0.83843**	0.80274	0.80272
14	0.18688	0.79423	0.82458	0.83886	0.84634*	0.84435**	0.80927	0.80968
15	0.19685	0.80257	0.83106	0.84474	0.85149*	0.84931**	0.8161	0.81626

ตารางที่ 98-112 แสดงค่าเฉลี่ย RS ซึ่งข้อมูลแต่ละชุดมีขนาด 300 หน่วยตัวอย่าง มีจำนวนตัวแปรอิสระ 3, 5 และ 7 ตัวแปร และการแจกแจงของข้อมูลมีรูปแบบของค่าเฉลี่ยและค่าความแปรปรวนร่วมที่ต่างกัน ผลการทดลองที่ได้แสดงให้เห็นว่า ในทุก ๆ ลักษณะของข้อมูล ผลที่ได้มีแนวโน้มไปในทางเดียวกัน ดังนั้นจึงเลือกใช้ข้อมูลในตารางที่ 106 เป็นตัวแทนในการสร้างกราฟได้ดังภาพที่ 32 เพื่อใช้อธิบายผลของข้อมูล ดังนี้



ภาพที่ 32 ค่าเฉลี่ย RS เมื่อใช้ข้อมูลรูปแบบที่ 4 ค่าเฉลี่ยเท่ากับ μ_{53} และค่าความแปรปรวนร่วมเท่ากับ Σ_{52} จำนวนตัวแปรอิสระ 5 ตัวแปร ขนาดตัวอย่างเท่ากับ 300

จากภาพที่ 32 พบว่า เมื่อแบ่งกลุ่มข้อมูลตั้งแต่ 2-15 กลุ่ม อัลกอริทึมจัดกลุ่มแบบเคมีนมีประสิทธิภาพดีที่สุด เนื่องจากให้ค่า RS สูงที่สุด รองลงมาคือ วิธีจัดระบบด้วยตัวเองของโคโฮเนน ซึ่งให้ค่าใกล้เคียงกัน และวิธีการจัดกลุ่มแบบลำดับขั้นด้วยวิธี Single linkage มีประสิทธิภาพต่ำสุด ในทุกกรณีและมีค่าต่างจากวิธีอื่นมาก จึงไม่นำมารวมในการสร้างกราฟดังภาพที่ 32

ส่วนที่ 3: ค่าความแม่นยำสำหรับข้อมูลจริง

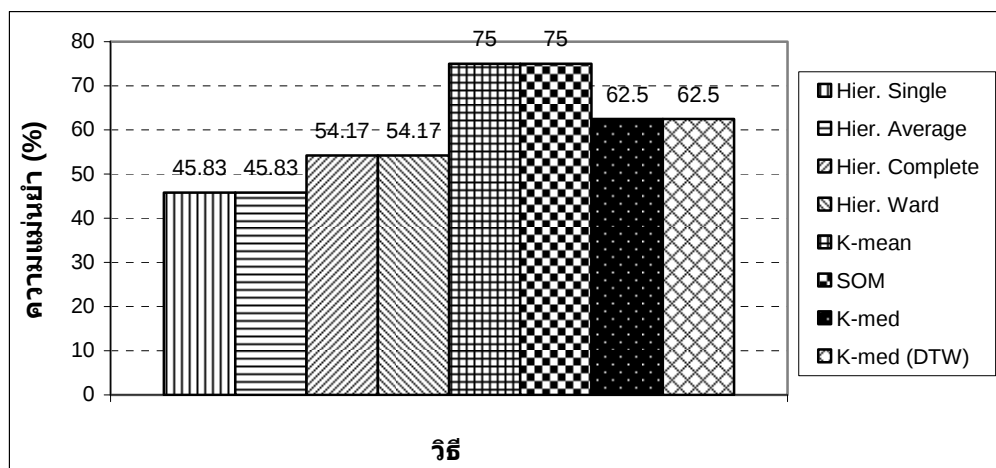
3.1 ข้อมูลคนไข้ทางด้านสายตา 24 หน่วยตัวอย่าง ซึ่งแบ่งตามความเหมาะสมในการใส่คอนแทคเลนส์เป็น 3 กลุ่ม คือ กลุ่มที่ 1 คนไข้เหมาะกับคอนแทคเลนส์แบบแข็ง กลุ่มที่ 2 คนไข้เหมาะกับคอนแทคเลนส์แบบอ่อน และกลุ่มที่ 3 คนไข้ไม่เหมาะกับคอนแทคเลนส์ จำนวน 4, 5 และ 15 คน ตามลำดับ แบ่งตามคุณลักษณะ 4 ประเภท คือ อายุ (1.วัยรุ่น 2.วัยผู้ใหญ่ 3.วัยชรา) ลักษณะของอาการผิดปกติ (1.สายตาสั้น 2.สายตายาว) สายตาเอียง (1.ไม่ 2.ใช่) อัตราน้ำหล่อเลี้ยงตา (1.ลดลง 2.ปกติ) ซึ่งตัวอย่างข้อมูลแสดงดังตารางที่ 113 (ข้อมูลเต็มแสดงไว้ในภาคผนวก)

ตารางที่ 113 ข้อมูลการจำแนกคนไข้ทางสายตาเป็น 3 กลุ่มคือ กลุ่มที่ 1 คนไข้เหมาะกับคอนแทกเลนส์แบบแข็ง กลุ่มที่ 2 คนไข้เหมาะกับคอนแทกเลนส์แบบอ่อน และ กลุ่มที่ 3 คนไข้ไม่เหมาะกับคอนแทกเลนส์ แบ่งตามคุณลักษณะ 4 ประเภท มีข้อมูลทั้งหมด 24 หน่วยตัวอย่าง

หน่วยตัวอย่าง	อายุ	ลักษณะของ อาการผิดปกติ	สายตาเอียง	อัตราน้ำหล่อ เลี้ยงตา	กลุ่ม
1	1	1	2	2	1
2	1	2	2	2	1
3	2	1	2	2	1
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
23	3	2	2	1	3
24	3	2	2	2	3

เมื่อนำข้อมูลการจำแนกคนไข้ทางสายตาทั้งหมด 24 หน่วยตัวอย่าง มาแบ่งเป็น 3 กลุ่มด้วยวิธีต่าง ๆ ทั้ง 5 วิธี แล้วนำมาเปรียบเทียบกับกลุ่มที่ได้จำแนกไว้ โดยการคำนวณค่าความแม่นยำ (%) ของการแบ่งกลุ่มในแต่ละวิธี พบว่าการแบ่งกลุ่มโดยอัลกอริทึมจัดกลุ่มแบบเคมีนและวิธีจัดระบบด้วยตัวเองของโคโฮเนนมีประสิทธิภาพสูงสุดเท่ากัน มีความแม่นยำในการแบ่งกลุ่มคิดเป็นร้อยละ 75 และวิธีการจัดกลุ่มแบบลำดับขั้นด้วยวิธี Single linkage และ Average linkage within groups มีประสิทธิภาพต่ำสุด มีความแม่นยำในการแบ่งกลุ่มร้อยละ 45.83 ดังแสดงในภาพที่ 33

ในกรณีที่ไม่ทราบจำนวนกลุ่มของข้อมูลมาก่อน จะทำการแบ่งกลุ่มข้อมูลและทดสอบประสิทธิภาพของการแบ่งกลุ่มด้วยค่า RMSSTD และ RS ได้ดังตารางที่ 114



ภาพที่ 33 แผนภูมิเปอร์เซ็นต์ความแม่นยำในการแบ่งคนไข้ทางสายตาเป็น 3 กลุ่มด้วยวิธีต่าง ๆ

ตารางที่ 114 ค่า RMSSTD และ RS ของข้อมูลการจำแนกคนไข้ทางสายตา

ค่า RMSSTD								
จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.96615	0.88549*	0.92165	0.92165	0.88549*	0.88549*	0.88549*	0.88549*
3	0.93116	0.84104	0.84104	0.84104	0.82736*	0.84104	0.84104	0.84104
4	0.86458	0.86458	0.80429	0.80429	0.77907**	0.75829*	0.84543	0.85147
5	0.80193	0.80193	0.76161**	0.76161**	0.75398*	0.7634	0.80764	0.82328
ค่า RS								
2	0.10714	0.25*	0.1875	0.1875	0.25*	0.25*	0.25*	0.25*
3	0.20833	0.35417	0.35417	0.35417	0.375*	0.35417	0.35417	0.35417
4	0.35	0.35	0.4375	0.4375	0.47222**	0.5*	0.37847	0.36957
5	0.46875	0.46875	0.52083**	0.52083**	0.53038*	0.51858	0.46115	0.44008

*เทคนิคการจัดกลุ่มที่ให้ประสิทธิภาพสูงสุด

**เทคนิคการจัดกลุ่มที่ให้ประสิทธิภาพสูงอันดับที่ 2

จากตารางที่ 114 พบว่า เมื่อแบ่งกลุ่มข้อมูลเป็น 2, 3 และ 5 กลุ่ม อัลกอริทึมจัดกลุ่มแบบเคมีนมีประสิทธิภาพดีที่สุด เนื่องจากให้ค่า RMSSTD ต่ำที่สุดและ RS สูงที่สุด ส่วนการแบ่งกลุ่มข้อมูลเป็น 4 กลุ่ม วิธีจัดระบบด้วยตัวเองของโคโฮเนนมีประสิทธิภาพดีที่สุด ซึ่งเมื่อจำนวนกลุ่มเพิ่มขึ้นประสิทธิภาพในการแบ่งกลุ่มของทั้ง 5 เทคนิควิธีมีแนวโน้มเพิ่มขึ้นด้วย

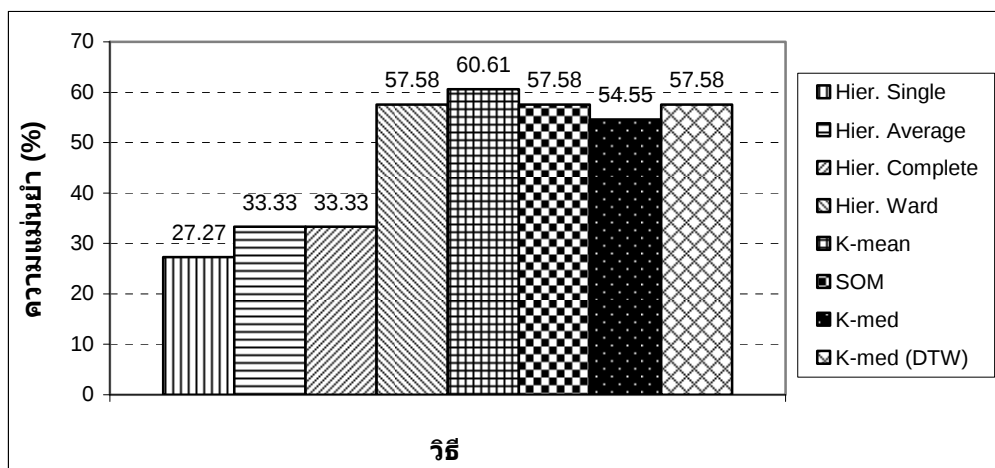
3.2 ข้อมูลคนไข้ที่เป็นมะเร็งเต้านม 33 หน่วยตัวอย่าง ซึ่งทำการศึกษาในปี ค.ศ.1958-1970 ที่โรงพยาบาลของมหาวิทยาลัยชิคาโก โดยสำรวจคนไข้ที่เข้ารับการผ่าตัดศัลยกรรมเนื่องจากเป็นมะเร็งเต้านม แบ่งคนไข้เป็น 2 กลุ่ม คือ กลุ่มที่ 1 คนไข้ที่มีอายุยืนมากกว่า 5 ปีขึ้นไปจำนวน 25 คน และกลุ่มที่ 2 คนไข้ที่ตายภายใน 5 ปีจำนวน 8 คน แบ่งตามคุณลักษณะ 3 ประเภท คือ อายุของผู้ป่วยในช่วงเวลาที่สำรวจ ปีที่ผู้ป่วยมาเข้ารับการรักษ (หน่วยเป็นปีค.ศ.-1900) และจำนวนก้อนเนื้อที่เกิดขึ้น ซึ่งตัวอย่างข้อมูลแสดงดังตารางที่ 115 (ข้อมูลเต็มแสดงไว้ในภาคผนวก)

ตารางที่ 115 ข้อมูลการจำแนกคนไข้ที่เป็นมะเร็งเต้านม 2 กลุ่มคือ กลุ่มที่ 1 คนไข้ที่มีอายุยืนมากกว่า 5 ปีขึ้นไป และกลุ่มที่ 2 คนไข้ที่ตายภายใน 5 ปี แบ่งตามคุณลักษณะ 3 ประเภท คือ อายุของผู้ป่วยในช่วงเวลาที่สำรวจ ปีที่ผู้ป่วยมาเข้ารับการรักษ (หน่วยเป็นปีค.ศ.-1900) และจำนวนก้อนเนื้อที่เกิดขึ้น มีข้อมูลทั้งหมด 33 หน่วยตัวอย่าง

หน่วยตัวอย่าง	อายุของผู้ป่วย	ปีที่ป่วย	จำนวนก้อนเนื้อ	กลุ่ม
1	34	67	7	1
2	35	64	13	1
3	38	67	5	1
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
32	70	59	8	1
33	72	67	3	1

เมื่อนำข้อมูลการจำแนกคนไข้ที่เป็นมะเร็งเต้านม ทั้งหมด 33 หน่วยตัวอย่าง มาแบ่งเป็น 2 กลุ่ม ด้วยวิธีต่าง ๆ ทั้ง 5 วิธี แล้วนำมาเปรียบเทียบกับกลุ่มที่ได้จำแนกไว้ โดยการคำนวณค่าความแม่นยำ (%) ของการแบ่งกลุ่มในแต่ละวิธี พบว่าการแบ่งกลุ่มโดยอัลกอริทึมจัดกลุ่มแบบเคมีนมีประสิทธิภาพสูงสุด มีความแม่นยำในการแบ่งกลุ่มคิดเป็นร้อยละ 60.61 และวิธีการ

จัดกลุ่มแบบลำดับขั้นด้วยวิธี Single linkage มีประสิทธิภาพต่ำสุด มีความแม่นยำในการแบ่งกลุ่มร้อยละ 27.27 ดังแสดงในภาพที่ 34



ภาพที่ 34 แผนภูมิเปอร์เซ็นต์ความแม่นยำในการแบ่งคนไข้มะเร็งเต้านมเป็น 2 กลุ่มด้วยวิธีต่าง ๆ

ในกรณีที่ ไม่ทราบจำนวนกลุ่มของข้อมูลมาก่อน จะทำการแบ่งกลุ่มข้อมูลและทดสอบประสิทธิภาพของการแบ่งกลุ่มด้วยค่า RMSSTD และ RS ได้ดังตารางที่ 116

จากตารางที่ 116 พบว่า เมื่อแบ่งกลุ่มข้อมูลเป็น 2, 3, 4 และ 5 กลุ่ม อัลกอริทึมจัดกลุ่มแบบเคมีนมีประสิทธิภาพดีที่สุด เนื่องจากให้ค่า RMSSTD ต่ำที่สุดและ RS สูงที่สุด และรองลงมาคือ วิธีการจัดกลุ่มแบบลำดับขั้นด้วยวิธี Ward linkage ซึ่งเมื่อจำนวนกลุ่มเพิ่มขึ้นประสิทธิภาพในการแบ่งกลุ่มของทั้ง 5 เทคนิควิธีมีแนวโน้มเพิ่มขึ้นด้วย

ตารางที่ 116 ค่า RMSSTD และ RS ของข้อมูลการจำแนกคนไข้ที่เป็นมะเร็งเต้านม

ค่า RMSSTD								
จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.99362	0.81608	0.81608	0.80842**	0.80509*	0.85821	0.9026	0.90953
3	0.9692	0.7424	0.75636	0.70812**	0.70529*	0.72943	0.89267	0.89267
4	0.94769	0.61501	0.67086	0.61444**	0.60508*	0.61853	0.84845	0.82346
5	0.85147	0.56498	0.56065	0.55703**	0.55081*	0.57784	0.77759	0.75801
ค่า RS								
2	0.043573	0.35483	0.35483	0.36689**	0.37209*	0.28649	0.21077	0.19861
3	0.11936	0.48329	0.46368	0.5299**	0.53365*	0.50119	0.25295	0.25295
4	0.18608	0.65722	0.59214	0.65785**	0.6682*	0.65329	0.34762	0.38548
5	0.36563	0.7207	0.72496	0.7285**	0.73453*	0.70784	0.47094	0.49725

*เทคนิคการจัดกลุ่มที่ให้ประสิทธิภาพสูงสุด

**เทคนิคการจัดกลุ่มที่ให้ประสิทธิภาพสูงอันดับที่ 2

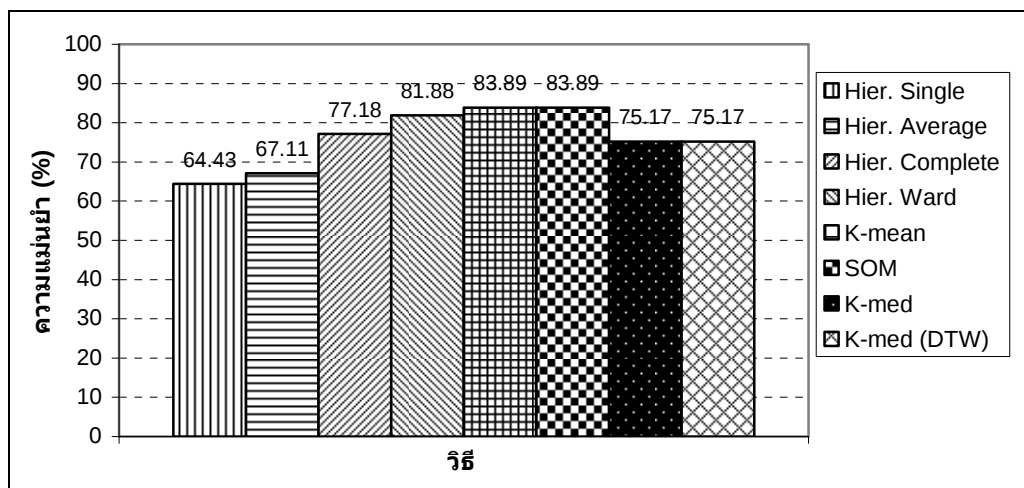
3.3 ข้อมูลการคัดแยกสับปะรดในโรงงานแปรรูป โดยใช้ระดับความสุกแก่ของสับปะรดเป็นตัวแบ่งกลุ่มคือ กลุ่มที่ 1 สับปะรดอ่อน และกลุ่มที่ 2 สับปะรดสุก ซึ่งพิจารณาจากสีภายในของสับปะรดและคุณลักษณะภายนอก จำนวน 95 และ 54 ลูก ตามลำดับ ทำการวัดด้วยเครื่องทดสอบสมบัติทางกายภาพของผลไม้และเก็บข้อมูล 3 ลักษณะคือ ความถี่สั่นพ้อง (Resonant Frequency) ซึ่งเป็นการวัดสมบัติเสียงสะท้อน ค่าเฉลี่ยความหนาแน่นเนื้อ (Average Firmness) และค่าปริมาณน้ำตาลหรือปริมาณของแข็งที่ละลายน้ำได้ (Soluble Solids Content) โดยแสดงค่าปริมาณน้ำตาลนี้ในหน่วยเปอร์เซ็นต์บริกซ์ (%Brix) เก็บข้อมูลทั้งหมด 149 หน่วยตัวอย่าง ซึ่งตัวอย่างข้อมูลแสดงดังตารางที่ 117 (ข้อมูลเต็มแสดงไว้ในภาคผนวก)

ตารางที่ 117 ข้อมูลผลการคัดแยกสับประรดเป็น 2 กลุ่มคือ กลุ่มที่ 1 สับประรดอ่อน และกลุ่มที่ 2 สับประรดสุก ทดสอบด้วยเครื่องทดสอบสมบัติทางกายภาพของผลไม้และเก็บข้อมูล 3 ลักษณะ ได้แก่ ความถี่สั่นพ้อง ค่าเฉลี่ยความหนาแน่นเนื้อ และปริมาณน้ำตาล มีข้อมูลทั้งหมด 149 หน่วยตัวอย่าง

หน่วยตัวอย่าง	ความถี่สั่นพ้อง	ค่าเฉลี่ย ความหนาแน่นเนื้อ	ปริมาณน้ำตาล (%Brix)	กลุ่ม
1	420.141	2291.83	13.48	1
2	453.130	3418.80	11.74	1
3	454.864	2048.75	12.16	1
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
147	251.738	625.23	10.74	2
148	239.583	318.37	15.1	2
149	222.223	281.93	14.22	2

เมื่อนำข้อมูลการคัดแยกสับประรดทั้งหมด 149 หน่วยตัวอย่าง มาแบ่งเป็น 2 กลุ่ม ด้วยเทคนิคการแบ่งกลุ่มที่ศึกษาทั้ง 5 วิธี แล้วนำมาเปรียบเทียบกับกลุ่มที่ได้คัดแยกไว้ โดยการคำนวณค่าความแม่นยำ (%) ของการแบ่งกลุ่มในแต่ละวิธี พบว่า การแบ่งกลุ่มโดยอัลกอริทึมจัดกลุ่มแบบเคมีนและวิธีจัดระบบด้วยตัวเองของโคโฮเนนมีประสิทธิภาพสูงสุดเท่ากัน มีความแม่นยำในการแบ่งกลุ่มคิดเป็นร้อยละ 83.89 และวิธีการจัดกลุ่มแบบลำดับขั้นด้วยวิธี Single linkage มีประสิทธิภาพต่ำสุด มีความแม่นยำในการแบ่งกลุ่มร้อยละ 64.43 ดังแสดงในภาพที่ 33

ในกรณีที่ไมทราบจำนวนกลุ่มของข้อมูลมาก่อน จะทำการแบ่งกลุ่มข้อมูลและทดสอบประสิทธิภาพของการแบ่งกลุ่มด้วยค่า RMSSTD และ RS ได้ดังตารางที่ 118



ภาพที่ 35 แผนภูมิเปอร์เซ็นต์ความแม่นยำในการคัดแยกสับปะรดเป็น 2 กลุ่มด้วยวิธีต่าง ๆ

จากตารางที่ 118 พบว่า เมื่อแบ่งกลุ่มข้อมูลเป็น 2, 3, 4 และ 5 กลุ่ม อัลกอริทึมจัดกลุ่มแบบเคมีนมีประสิทธิภาพดีที่สุด เนื่องจากให้ค่า RMSSTD ต่ำที่สุดและ RS สูงที่สุด และรองลงมาคือ วิธีจัดระบบด้วยตัวเองของโคโฮแน ซึ่งเมื่อจำนวนกลุ่มเพิ่มขึ้นประสิทธิภาพในการแบ่งกลุ่มของทั้ง 5 เทคนิควิธีมีแนวโน้มเพิ่มขึ้นด้วย

ตารางที่ 118 ค่า RMSSTD และ RS ของข้อมูลการคัดแยกสับปะรด

ค่า RMSSTD								
จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.99028	0.94678	0.78479	0.7217	0.70714*	0.70995**	0.81199	0.81199
3	0.97685	0.66607	0.63159	0.65125	0.62775*	0.62929**	0.80257	0.80257
4	0.92688	0.64264	0.57868	0.5853	0.55348*	0.56862**	0.80368	0.71116
5	0.92609	0.5624	0.54083	0.51255	0.49565*	0.50438**	0.64554	0.58259
6	0.9249	0.55765	0.48481	0.48135	0.4638*	0.47355**	0.56338	0.53575
7	0.92628	0.51895	0.46199	0.45148	0.44237*	0.44381**	0.54829	0.53324
8	0.92733	0.47134	0.43411	0.42644	0.41162*	0.42057**	0.50959	0.47032
ค่า RS								
2	0.025965	0.10967	0.38826	0.48267	0.50333*	0.49938**	0.34513	0.34513
3	0.058666	0.56234	0.60649	0.5816	0.61125*	0.60935**	0.36458	0.36458
4	0.15831	0.59538	0.67192	0.66437	0.69987*	0.68323**	0.36719	0.5045
5	0.16554	0.69226	0.71541	0.74439	0.76097*	0.75247**	0.59454	0.66976
6	0.17346	0.69953	0.7729	0.77613	0.79216*	0.78332**	0.69332	0.72267
7	0.17679	0.74161	0.79522	0.80443	0.81224*	0.81102**	0.71156	0.72719
8	0.18074	0.78835	0.82046	0.82675	0.83858*	0.83149**	0.7526	0.78926

*เทคนิคการจัดกลุ่มที่ให้ประสิทธิภาพสูงสุด

**เทคนิคการจัดกลุ่มที่ให้ประสิทธิภาพสูงอันดับที่ 2

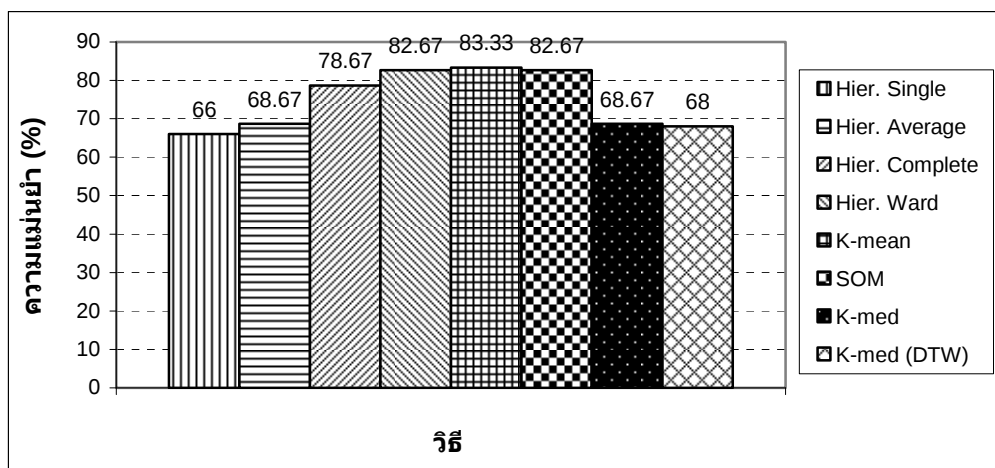
3.4 ข้อมูลต้นไทรสสายพันธุ์ต่าง ๆ 150 หน่วยตัวอย่าง ข้อมูลนี้มีการนำมาใช้ในงานวิจัยอย่างแพร่หลาย (Gelbard *et al.*, 2007) ซึ่งแบ่งตามชนิดของต้นไทรสเป็น 3 กลุ่ม คือ กลุ่มที่ 1 Iris Setosa กลุ่มที่ 2 Iris Versicolour และกลุ่มที่ 3 Iris Virginica อย่างละ 50 ตัวอย่าง แบ่งแยกตามคุณลักษณะ 4 ประเภท คือ ความยาวของกลีบเลี้ยง (Sepal) ความกว้างของกลีบเลี้ยง ความยาวของกลีบดอก (Petal) และความกว้างของกลีบดอก โดยวัดในหน่วยเซนติเมตร ซึ่งตัวอย่างข้อมูลแสดงดังตารางที่ 119 (ข้อมูลเต็มแสดงไว้ในภาคผนวก)

ตารางที่ 119 ข้อมูลต้นไอริส 3 กลุ่มคือ กลุ่มที่ 1 Iris Setosa กลุ่มที่ 2 Iris Versicolour และกลุ่มที่ 3 Iris Virginica แบ่งแยกตามคุณลักษณะ 4 ประเภท คือ ความยาวของกลีบเลี้ยง ความกว้างของกลีบเลี้ยง ความยาวของกลีบดอก และความกว้างของกลีบดอก ซึ่งวัดในหน่วย ซม. มีข้อมูลทั้งหมด 150 หน่วยตัวอย่าง

หน่วยตัวอย่าง	ความยาวของ กลีบเลี้ยง	ความกว้าง ของกลีบเลี้ยง	ความยาวของ กลีบดอก	ความกว้าง ของกลีบดอก	กลุ่ม
1	5.1	3.5	1.4	0.2	1
2	4.9	3	1.4	0.2	1
3	4.7	3.2	1.3	0.2	1
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
149	6.2	3.4	5.4	2.3	3
150	5.9	3	5.1	1.8	3

เมื่อนำข้อมูลต้นไอริสทั้งหมด 150 หน่วยตัวอย่าง มาแบ่งเป็น 3 กลุ่ม ด้วยวิธีต่าง ๆ ทั้ง 5 วิธี แล้วนำมาเปรียบเทียบกับกลุ่มที่ได้จำแนกไว้ โดยการคำนวณค่าความแม่นยำ (%) ของการแบ่งกลุ่มในแต่ละวิธี พบว่าการแบ่งกลุ่มโดยอัลกอริทึมจัดกลุ่มแบบเคมีนมีประสิทธิภาพสูงสุด มีความแม่นยำในการแบ่งกลุ่มคิดเป็นร้อยละ 83.33 และวิธีการจัดกลุ่มแบบลำดับขั้นด้วยวิธี Single linkage มีประสิทธิภาพต่ำสุด มีความแม่นยำในการแบ่งกลุ่มร้อยละ 66 ดังแสดงในภาพที่ 36

ในกรณีที่ไม่ทราบจำนวนกลุ่มของข้อมูลมาก่อน จะทำการแบ่งกลุ่มข้อมูลและทดสอบประสิทธิภาพของการแบ่งกลุ่มด้วยค่า RMSSTD และ RS ได้ดังตารางที่ 120



ภาพที่ 36 แผนภูมิเปอร์เซ็นต์ความแม่นยำในการแยกประเภทต้นไม้ไอริสเป็น 3 กลุ่มด้วยวิธีต่าง ๆ

จากตารางที่ 120 พบว่า เมื่อแบ่งกลุ่มข้อมูลเป็น 2, 3, 4, 5, 6 และ 7 กลุ่ม อัลกอริทึมจัดกลุ่มแบบเคมีนมีประสิทธิภาพดีที่สุด เนื่องจากให้ค่า RMSSTD ต่ำที่สุดและ RS สูงที่สุด ส่วนการแบ่งกลุ่มข้อมูลเป็น 8 กลุ่ม วิธีการจัดกลุ่มแบบลำดับขั้นด้วยวิธี Ward linkage มีประสิทธิภาพดีที่สุด ซึ่งเมื่อจำนวนกลุ่มเพิ่มขึ้นประสิทธิภาพในการแบ่งกลุ่มของทั้ง 5 เทคนิควิธีมีแนวโน้มเพิ่มขึ้นด้วย

ตารางที่ 120 ค่า RMSSTD และ RS ของข้อมูลการจำแนกต้นไม้

ค่า RMSSTD								
จำนวน กลุ่ม	วิธี							
	Hier.				K-mean	SOM	K-med	K-med (DTW)
	Single	Average	Complete	Ward				
2	0.61082*	0.61082*	0.70518	0.6195	0.61082*	0.61082*	0.92833	0.92863
3	0.60279	0.57861	0.50986	0.5015**	0.48601*	0.51213	0.81652	0.85204
4	0.56983	0.56983	0.46206	0.44579**	0.44114*	0.45447	0.50558	0.54047
5	0.57071	0.47697	0.42846	0.4242**	0.40338*	0.43452	0.44904	0.54044
6	0.56774	0.45284	0.41501	0.38512**	0.37482*	0.38832	0.43435	0.40174
7	0.5582	0.38954	0.39921	0.36584**	0.35828*	0.38864	0.45414	0.53212
8	0.53545	0.36901	0.35455**	0.34693*	0.37962	0.35629	0.36452	0.36965
ค่า RS								
2	0.6294*	0.6294*	0.50606	0.6188	0.6294*	0.6294*	0.14398	0.14343
3	0.64153	0.6697	0.74353	0.75187**	0.76697*	0.74125	0.34225	0.28378
4	0.68183	0.68183	0.7908	0.80527**	0.80931*	0.79761	0.74953	0.71378
5	0.68303	0.7786	0.82135	0.82489**	0.84166*	0.81626	0.80378	0.71576
6	0.68849	0.80182	0.83355	0.85666**	0.86422*	0.85427	0.81767	0.84402
7	0.70096	0.85437	0.84705	0.87155**	0.8768*	0.85504	0.80206	0.72825
8	0.72676	0.87023	0.8802**	0.8853*	0.86266	0.87902	0.87337	0.86978

*เทคนิคการจัดกลุ่มที่ให้ประสิทธิภาพสูงสุด

**เทคนิคการจัดกลุ่มที่ให้ประสิทธิภาพสูงอันดับที่ 2

วิจารณ์

จากการทดสอบประสิทธิภาพเทคนิควิธีการแบ่งกลุ่มทั้ง 5 วิธี ในแต่ละลักษณะของข้อมูล ซึ่งจำลองขึ้นตามลักษณะต่าง ๆ ที่ได้กำหนดไว้ ได้แก่ ทั้งค่าเฉลี่ย ค่าความแปรปรวนร่วม ขนาดตัวอย่าง จำนวนตัวแปรอิสระ และจำนวนกลุ่ม ผลการศึกษาพบว่า อัลกอริทึมจัดกลุ่มแบบเคมีนมีประสิทธิภาพสูงที่สุด เนื่องจากค่าความแตกต่างของข้อมูลภายในกลุ่ม (RMSSTD) ต่ำสุด และค่าความแตกต่างของข้อมูลระหว่างกลุ่ม (RS) สูงสุดเกือบทุกกรณี ยกเว้นในกรณีของข้อมูล 25 หน่วยตัวอย่าง ซึ่งแบ่งข้อมูลเป็น 4 และ 5 กลุ่ม พบว่าอัลกอริทึมจัดกลุ่มแบบเคมีนไม่ได้ให้ประสิทธิภาพสูงที่สุด สาเหตุอาจเนื่องมาจากข้อมูลมีขนาดเล็กมาก ทำให้อัลกอริทึมจัดกลุ่มแบบเคมีนไม่สามารถแบ่งกลุ่มได้อย่างมีประสิทธิภาพ และวิธีการจัดกลุ่มแบบลำดับขั้นด้วยวิธี Single linkage มีประสิทธิภาพต่ำสุดในทุกกรณี เมื่อจำนวนกลุ่มเพิ่มขึ้นประสิทธิภาพในการแบ่งกลุ่มของทั้ง 5 เทคนิควิธีมีแนวโน้มเพิ่มขึ้น แต่จำนวนตัวแปรที่เพิ่มขึ้นไม่มีผลต่อประสิทธิภาพในการแบ่งกลุ่มของทั้ง 5 เทคนิควิธี

ในงานวิจัยที่ผ่านมา ๆ มานั้น พบว่า อัลกอริทึมจัดกลุ่มแบบเคมีนมีประสิทธิภาพในการแบ่งกลุ่มดีกว่าอัลกอริทึมจัดกลุ่มแบบเคมีน ซึ่งงานวิจัยชิ้นนี้ไม่ได้ให้ผลเป็นไปตามนั้น อาจเนื่องมาจากข้อมูลที่จำลองมาศึกษานั้น มีการกำหนดให้ข้อมูลแจกแจงแบบปกติหลายตัวแปรทำให้เกิดสิ่งรบกวน (Noise) หรือมี Noise เกิดขึ้นน้อยมาก ดังนั้น ในการวิจัยครั้งนี้อัลกอริทึมจัดกลุ่มแบบเคมีนจึงให้ผลการแบ่งกลุ่มที่ดีที่สุด

และสำหรับข้อมูลจริงนั้น ผู้ศึกษาได้ทำการวัดค่าความแม่นยำ หลังจากใช้เทคนิควิธีการแบ่งกลุ่มทั้ง 5 วิธีด้วยข้อมูล 4 ชุด ผลปรากฏว่า อัลกอริทึมจัดกลุ่มแบบเคมีนมีประสิทธิภาพสูงที่สุด และวิธีการจัดกลุ่มแบบลำดับขั้นด้วยวิธี Single linkage มีประสิทธิภาพต่ำสุดในทุกชุดข้อมูล ซึ่งในข้อมูลขนาดเล็กคือข้อมูลคนไข้ทางสายตา ($n=24$, $p=4$) และข้อมูลคนไข้โรคมะเร็งเต้านม ($n=33$, $p=3$) แสดงให้เห็นว่าเมื่อจำนวนตัวแปรเปลี่ยนไป อัลกอริทึมจัดกลุ่มแบบเคมีนยังคงมีประสิทธิภาพสูงที่สุด เช่นเดียวกับในข้อมูลขนาดกลางคือข้อมูลสับปะรด ($n=149$, $p=3$) และข้อมูลต้นไทร ($n=150$, $p=4$) ซึ่งพบว่าอัลกอริทึมจัดกลุ่มแบบเคมีนยังคงให้ประสิทธิภาพสูงที่สุดในทั้งสองชุดข้อมูล และเมื่อมองในทิศทาง的增加ของจำนวนข้อมูลจากข้อมูลขนาดเล็กเป็นขนาดกลาง ผลที่ได้ยังคงสรุปได้เช่นเดิม

สรุปและข้อเสนอแนะ

สรุป

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาวิธีการทำงานของเทคนิคการแบ่งกลุ่ม 5 วิธี คือ วิธีการจัดกลุ่มแบบลำดับขั้น อัลกอริทึมจัดกลุ่มแบบเคมีน วิธีจัดระบบด้วยตัวเองของโคโฮเนน อัลกอริทึมจัดกลุ่มแบบเคเมคอยด์ และการใช้อัลกอริทึมจัดกลุ่มแบบเคเมคอยด์ร่วมกับการใช้วิธีวิเคราะห์ห่างแบบไดนามิกไทม์วอร์ปิง นำมาศึกษาร่วมกับข้อมูล 2 ชนิดคือ ข้อมูลจำลอง และข้อมูลจริง ดังนี้

1. ข้อมูลจำลอง ซึ่งมีการแจกแจงแบบปกติหลายตัวแปร ประกอบด้วยตัวอย่างขนาดเล็กเท่ากับ 25 ขนาดกลางเท่ากับ 100 ขนาดใหญ่เท่ากับ 300 กำหนดจำนวนกลุ่มในตัวอย่างขนาดเล็กตั้งแต่ 2-5 กลุ่ม ขนาดกลางตั้งแต่ 2-8 กลุ่ม ขนาดใหญ่ตั้งแต่ 2-15 กลุ่ม กำหนดจำนวนตัวแปรอิสระ 3, 5 และ 7 ตัวแปร เพื่อทำการเปรียบเทียบประสิทธิภาพการแบ่งกลุ่มทั้ง 5 เทคนิควิธี โดยพิจารณาจากค่าความแตกต่างของข้อมูลภายในกลุ่ม (RMSSTD) และค่าความแตกต่างของข้อมูลระหว่างกลุ่ม (RS) ซึ่งสรุปผลได้ดังตารางที่ 117 และ 118

ตารางที่ 121 สรุปผลการแบ่งกลุ่มข้อมูลจำลองด้วยค่าเฉลี่ย RMSSTD

ขนาด ตัวอย่าง (n)	จำนวนตัวแปร อิสระ (p)	จำนวน กลุ่ม (k)	เทคนิคการจัดกลุ่มที่ให้ประสิทธิภาพสูงที่สุด	
			อันดับที่ 1	อันดับที่ 2
25	3, 5, 7	2	KMEAN	SOM
		3	KMEAN	SOM
		4	WARD	KMEAN , COMP
		5	WARD	KMEAN , COMP
100	3, 5, 7	2-7	KMEAN	SOM
		8	KMEAN	WARD, SOM
300	3, 5, 7	2-15	KMEAN	SOM

ตารางที่ 122 สรุปผลการแบ่งกลุ่มข้อมูลจำลองด้วยค่าเฉลี่ย RS

ขนาด ตัวอย่าง	จำนวนตัวแปร อิสระ (p)	จำนวน กลุ่ม (k)	เทคนิคการจัดกลุ่มที่ให้ประสิทธิภาพสูงที่สุด	
			อันดับที่ 1	อันดับที่ 2
25	3, 5, 7	2	KMEAN	SOM
		3	KMEAN	SOM
		4	WARD	KMEAN , COMP
		5	WARD	KMEAN , COMP
100	3, 5, 7	2-7	KMEAN	SOM
		8	KMEAN	WARD, SOM
300	3, 5, 7	2-15	KMEAN	SOM

เมื่อ	COMP	แทนวิธีการจัดกลุ่มแบบลำดับขั้นแบบ Complete linkage
	WARD	แทนวิธีการจัดกลุ่มแบบลำดับขั้นแบบ Ward linkage
	SOM	แทนวิธีจัดระบบด้วยตัวเองของโคโฮเนน
	KMEAN	แทนอัลกอริทึมจัดกลุ่มแบบเคมีน

จากตารางที่ 117 และ 118 พบว่า ผลการแบ่งกลุ่มข้อมูลจำลองด้วยค่าเฉลี่ย RMSSTD และ RS ให้ผลสอดคล้องกัน คือ ในตัวอย่างขนาด 25 หน่วย เมื่อแบ่งข้อมูลเป็น 2 และ 3 กลุ่ม อัลกอริทึมจัดกลุ่มแบบเคมีนมีประสิทธิภาพในการจัดกลุ่มสูงที่สุด และเมื่อแบ่งข้อมูลเป็น 4 และ 5 กลุ่ม วิธีการจัดกลุ่มแบบลำดับขั้นด้วยวิธี Ward linkage มีประสิทธิภาพในการจัดกลุ่มสูงที่สุด ส่วนในตัวอย่างขนาด 100 และ 300 หน่วย พบว่าอัลกอริทึมจัดกลุ่มแบบเคมีนมีประสิทธิภาพในการจัดกลุ่มสูงที่สุดในทุกขนาดของจำนวนกลุ่ม

2. สำหรับข้อมูลจริง ซึ่งผ่านการทดสอบสมมติฐานแล้วว่าข้อมูลมีการแจกแจงแบบปกติหลายตัวแปรทั้ง 4 ชุด ผลการศึกษาด้วยเทคนิคการแบ่งกลุ่มปรากฏว่า ในการวัดค่าความแม่นยำของการแบ่งกลุ่มทั้ง 5 วิธีนั้น อัลกอริทึมจัดกลุ่มแบบเคมีนมีประสิทธิภาพสูงที่สุดในทุกชุดข้อมูล แม้ว่าขนาดตัวอย่างและจำนวนตัวแปรจะเปลี่ยนไปก็ตาม

ซึ่งผลที่ได้จากการศึกษาเทคนิคการแบ่งกลุ่ม 5 วิธี ในข้อมูลจำลองและข้อมูลจริงมีความสอดคล้องกัน

ข้อเสนอแนะ

ผู้วิจัยเสนอแนวทางในการศึกษาและวิจัยต่อไปสำหรับผู้สนใจ ดังนี้

1. ศึกษาและเปรียบเทียบความสามารถในการแบ่งกลุ่ม เมื่อข้อมูลมีการแจกแจงเป็นแบบอื่น ๆ
2. ทำการเปรียบเทียบความสามารถในการแบ่งกลุ่ม โดยใช้เกณฑ์การวัดประสิทธิภาพการแบ่งกลุ่มแบบอื่น ๆ นอกเหนือจากค่า RMSSTD และ RS เช่น เกณฑ์ Dunn, เกณฑ์ Davies-Bouldin (DB) เป็นต้น (Halkidi *et al.*, 2002)
3. เมื่อต้องการแบ่งกลุ่มจำนวนมากควรเพิ่มขนาดตัวอย่างให้ใหญ่ขึ้น จะช่วยให้เทคนิคการแบ่งกลุ่มมีประสิทธิภาพมากยิ่งขึ้น

เอกสารและสิ่งอ้างอิง

- กัลยา วาณิชย์บัญชา. 2550. การวิเคราะห์ข้อมูลหลายตัวแปร. พิมพ์ครั้งที่ 2. จุฬาลงกรณ์มหาวิทยาลัย, กรุงเทพฯ.
- _____. 2549. การวิเคราะห์สถิติขั้นสูงด้วย SPSS for Windows. พิมพ์ครั้งที่ 5. จุฬาลงกรณ์มหาวิทยาลัย, กรุงเทพฯ.
- มัลลิกา บุนนาค, กัลยา ครองแก้ว, วัชรารักษ์ สุริยาภิวัฒน์ และ นพรัตน์ รุ่งอุทัยศิริ. 2540. สถิติ. จุฬาลงกรณ์มหาวิทยาลัย, กรุงเทพฯ.
- วามิณี นิยาการศ. 2549. การวิเคราะห์พฤติกรรมการใช้บริการของลูกค้าอินเทอร์เน็ตแบงกิ้งด้วยเทคนิคการจัดกลุ่มแบบ 2 ขั้นตอนและ RFM Analysis. วิทยานิพนธ์ปริญญาโท, มหาวิทยาลัยเกษตรศาสตร์.
- สุภาพร บรรดาศักดิ์. 2548. การใช้เคมีนอัลกอริทึมสำหรับระบบการเรียนการสอนที่จัดเนื้อหาตามกลุ่มผู้เรียนแบบออนไลน์. วิทยานิพนธ์ปริญญาโท, มหาวิทยาลัยเกษตรศาสตร์.
- สุรัช วิวัฒน์เจริญชัย. 2546. **Data Mining of Electronic Banking in Thailand: Usage Behavior Analysis by Using K-Means Algorithm.** วิทยานิพนธ์ปริญญาโท, มหาวิทยาลัยเกษตรศาสตร์.
- อรนุช ชัยหมื่น. 2548. การศึกษาเปรียบเทียบการแบ่งกลุ่มข้อมูลลูกค้าสินค้าหัตถกรรมไทยโดยใช้วิธีการ 2 ขั้นตอนของ SOM กับ K-means Algorithm และ Hierarchical Cluster กับ K-means Algorithm. วิทยานิพนธ์ปริญญาโท, มหาวิทยาลัยเกษตรศาสตร์.
- Anderberg, M. 1973. **Cluster Analysis for Applications.** Academic Press, Washington D.C.

- Boonmung, S., B. Chomtee and K. Kanlayasiri. 2006. Evaluation of artificial neural networks for pineapple grading. **J. Text. Stud.** 37 (5): 568-579.
- Eisen, M. B., P. T. Spellman, P. O. Brown and D. Botstein. 1998. Cluster analysis and display of genome-wide expression patterns. **Proc. Natl. Acad. Sci. U.S.A.** 95 (25): 14863-14868.
- Forman, G. and B. Zhang. 2001. **Distributed Data Clustering can be Efficient and Exact.** Hewlett-Packard Labs, Data Mining Solution Dept. Boston, USA.
- Gelbard, R., O. Goldman and I. Spiegler. 2007. Investigating diversity of clustering methods: An empirical comparison. **DKE.** 63 (1): 155-166.
- Halkidi, M., Y. Batistakis and M. Vazirgiannis. 2002. Clustering validity checking methods: Part II. **SIGMOD Rec.** 31 (3): 19-27.
- Hartigan, J. A. and M. A. Wong. 1979. A K-means clustering algorithm. **Applied Stat.** 28 (1): 100-108.
- Henze, N. and B. Zirkler. 1990. A class of invariant consistent tests for multivariate normality. **Commun. Statist.-Theory Meth.** 19 (10): 3595-3618.
- Jain, A. K. and R. C. Dubes. 1988. **Algorithms for Clustering Data.** Prentice Hall Inc., New York.
- Jardine, N. and R. Sibson. 1968. The construction of hierarchic and non-hierarchic classifications. **Comp. J.** 11 (2): 177-184.
- Johnson, S. C. 1967. Hierarchical clustering schemes. **Psychometrika** 32 (3): 241-254.
- Johnson, R. A. and D. W. Wichern. 1992. **Applied Multivariate Statistical Analysis.** Prentice_Hall, Inc., New Jersey.

- Julien, B. 1990. **Lenses**. Machine Learning Repository. Available Source:
<http://archive.ics.uci.edu/ml/datasets.html>, March 3, 2009.
- Kuo, R. J., L. M. Ho. and C. M. Hu. 2002. Integration of self-organizing feature map and K-means Algorithm for market segmentation. **C&OR**. 29: 1475-1493.
- Kaufman, L. and P. J. Rousseeuw. 1990. **Finding Groups in Data: An Introduction to Cluster Analysis**. Wiley, New York.
- Kohonen, T. 1988. **Self-Organization and Associative Memory**. Springer-Verlag., New York.
- Lebart, L. and M. Rajman. 2000. Computing similarity, pp. 477-506. In R. Dale, H. Moisl and H. L. Somers, eds. **Handbook of Natural Language Processing**. Lincoln, United States.
- Lim, T. 1999. **Haberman's survival**. Machine Learning Repository. Available Source:
<http://archive.ics.uci.edu/ml/datasets.html>, March 3, 2009.
- MacQueen, J. B. 1967. Some Methods for classification and Analysis of Multivariate Observations. **Proc. Fifth Berkeley Symp. Math. Stat. Prob.** 5 (1): 281-297.
- Marshall, M. 1988. **Iris**. Machine Learning Repository. Available Source:
<http://archive.ics.uci.edu/ml/datasets.html>, March 3, 2009.
- Morrison, D. F. 2005. **Multivariate Statistical Methods**. 4th ed. McGraw-Hill Book Co., New York.
- Niennattrakul, V. and C. A. Ratanamahatana. 2006. Clustering multimedia data using time series. **ICHIT**. 1: 372-379.
- Niyagas, W., A. Srivihok and S. Kitisin. 2006. Clustering e-banking customer using data mining and marketing segmentation. **ECTI-Trans. CIT**. 2 (1): 63-69.

- Ratanamahatana, C. A. and E. Keogh. 2004. Making Time-series classification more accurate using learned constraints. **SDM**. 4: 11-22.
- Sheng, W. and X. Liu. 2006. A genetic k-medoids clustering algorithm. **J. Heuristics** 12 (6): 447-466.
- Witten, I. H. and E. Frank. 2000. **Data Mining**. Morgan Kaufmann Publishers, New York.
- Wiwattanacharoenchai, S. and A. Srivihok. 2003. Understanding online Banking in Thailand: Cluster analysis of customer usage behavior. **Asia-Australasian Regional Conference. 22-24 June**. Perth, Curtin University of Technology, Australia.

ภาคผนวก

ภาคผนวก ก
โปรแกรมที่ใช้ในการวิจัย

โปรแกรมที่ใช้ในงานวิจัยนี้ เขียนขึ้นด้วยโปรแกรม Matlab ในที่นี้จะแสดงโปรแกรมของการจำลองข้อมูล วิธีการจัดกลุ่มแบบลำดับขั้น อัลกอริทึมจัดกลุ่มแบบเคมีน วิธีจัดระบบด้วยตัวเองของโคโฮเนน อัลกอริทึมจัดกลุ่มแบบเคเมคอยด์ และวิธีวัดระยะห่างแบบไดนามิกไทม์วอร์ปปีงดังนี้

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%          การจำลองข้อมูลแจกแจงแบบปกติหลายตัวแปรขนาด 25 ตัวอย่าง          %%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

mu = [10 20 30];
sigma = [2 3 4; 3 7 8; 4 8 15;];
cases = 25;
for i=1:1000
    chkmvn = 1;
    while chkmvn == 1
        X = mvnrnd(mu,sigma,cases);
        [H,stats] = mardiatest(X,0.05);
        if H(1,1) == 0 && H(1,2) == 0 && H(1,3) == 0
            chkmvn = 0;
            dataset(:,i*3)-2 = X(:,1);
            dataset(:,i*3)-1 = X(:,2);
            dataset(:,i*3)-0 = X(:,3);
        end
        test = test +1;
    end
end
```

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%          วิธีการจัดกลุ่มแบบลำดับขั้นแบบ Single Linkage จำนวน 2-5 กลุ่ม          %%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

```

```

    for k = 2 : 5
        y = pdist(x,'euclidean');
        squareform(y);
        z = linkage(y,'single');
        T = cluster(z,k);
    end

```

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%          ฟังก์ชันอัลกอริทึมจัดกลุ่มแบบเคมีน          %%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

```

```

function [output, onc] = k_means_clustering(dataset,k);
c = size(dataset); r = c(1,1); c = c(1,2);
na(k,c) = 0; t(r,1) = 0; nc(k) = 0;
for i = 1:k
    for j = 1:c
        temp = randint(1,1,[1,r]);
        w(i,j) = dataset(temp,j);
    end
end
for i = 1:k
    temp = randint(1,1,[1,r]);
    w(i,1:c) = dataset(temp,1:c);
end
threshold = 20;
%while i > threshold
iteration = 1;
while iteration <=1
    for i = 1:r

```

```

for j = 1:k
    for a = 1:c
        w(j,c+1) = 0;
        w(j,c+1) = (w(j,c+1)+(dataset(i,c)-w(j,c))^2);
    end
    w(j,c+1) = w(j,c+1)^0.5;
end
mindist = 999999;
for j = 1:k
    if mindist > w(j,c+1)
        t(i,1) = j;
        mindist = w(j,c+1);
    end
end
end
for i = 1:r
    for j = 1:k;
        if t(i,1) == j;
            nc(j) = nc(j)+1;
            for a = 1:c
                na(j,a) = na(j,a)+dataset(i,a);
            end
        end
    end
end
end
onc = nc;
for j = 1:k
    for a = 1:c
        w(j,a) = na(j,a)/nc(j);
        na(j,a) = 0;
    end
end

```

```

        nc(j) = 0;

    end

    iteration = iteration + 1;

end

output = t;

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%                               วิธีจัดระบบด้วยตัวเองของโคโฮเนน จำนวน 2-5 กลุ่ม                               %%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

for k = 2:5

    number_of_data = size(output);
    number_of_data = number_of_data(1,1);
    alpha = 0.2;
    have_data_in_cluster = false;
    while have_data_in_cluster == false

        clear w;

        w = (randint(dimension_num,k,200)-100)*0.01;

% ----- Training Classifier -----

        data_num = 1;
        iteration_training = 1;
        while iteration_training <= 4000;
            minnest = 9999;
            for cluster_id = 1:k
                w_t = w(:,cluster_id)';
                d(cluster_id) = power(sum(power((output(data_num,:)- w_t),2)),0.5);
                if d(cluster_id) <= minnest
                    current_min = cluster_id;
                    minnest = d(cluster_id);
                end
            end
        end
    end
end

```

```

    for dimension_id = 1 : dimension_num
        w(dimension_id,current_min) = w(dimension_id,current_min) + alpha
        *(output(data_num,dimension_id)-w(dimension_id,current_min));
    end
    alpha = alpha/2;
    iteration_training = iteration_training + 1;
    data_num = data_num + 1;
    if data_num > number_of_data
        data_num = 1;
    end
end

% ----- Define Class -----

for cluster_id = 1: k
    num_of_each_cluster(cluster_id) = 0;
end

for data_num = 1 : number_of_data
    for cluster_id = 1: k
        w_t = w(:,cluster_id)';
        d(cluster_id) = power(sum(power((output(data_num,:)- w_t),2)),0.5);
    end
    minnest = 9999;
    for cluster_id = 1: k
        if d(cluster_id) <= minnest
            current_min = cluster_id;
            minnest = d(cluster_id);
        end
    end
    num_of_each_cluster(current_min) = num_of_each_cluster(current_min)+1;
    result(data_num,1) = data_num;
    result(data_num,2) = current_min;
end

```

```

        have_data_in_cluster = true;
    for cluster_id = 1: k
        if num_of_each_cluster(cluster_id) == 0;
            have_data_in_cluster = false;
        end
    end
end
end
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%                                อัลกอริทึมจัดกลุ่มแบบเคมดอยด์ จำนวน 2 กลุ่ม                                %%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

k = 2;

%----- Pick Initial Medoid -----

redundant = true;
while redundant == true
    initial_medoid = randint(1,k,number_of_data);
    initial_medoid = initial_medoid + 1;
    redundant = false;
    i = 1;
    while i <= k-1
        j = i+1;
        while j <= k
            if initial_medoid(1,i) - initial_medoid(1,j) == 0
                redundant = true;
            end
            j = j+1;
        end
        i = i+1;
    end
end
end

```

```

%----- Prepare Variable -----

iteration_training = 1;
while iteration_training <= 20
    medoid_id = 1;
    while medoid_id <= k
        number_of_medoid(medoid_id) = 0;
        medoid_id = medoid_id + 1;
    end
    clear data_in_medoid
%----- Each Element find closest Medoid -----

    data_id = 1;
    while data_id <= number_of_data
        minnest = 999999;
        medoid_id = 1;
        while medoid_id <= k
            dist = 0;
            i = 1;
            while i <= dimension_num
                dist = dist + power(output(data_id,i) - output(initial_medoid(medoid_id),i),2);
                i = i+1;
            end
            dist = power(dist,0.5);
            if dist < minnest
                T(data_id,1) = medoid_id;
                minnest = dist;
            end
            medoid_id = medoid_id + 1;
        end
        number_of_medoid(T(data_id,1)) = number_of_medoid(T(data_id,1)) + 1;
        data_in_medoid(T(data_id,1), number_of_medoid(T(data_id,1))) = data_id;
    end
end

```

```

        data_id = data_id + 1;
    end

%----- Adjust Medoid -----

    medoid_id = 1;
    while medoid_id <= k
        clear temp temp2
        if number_of_medoid(medoid_id) > 1
            i = 1;
            while i <= number_of_medoid(medoid_id)-1
                j = i+1;
                while j <= number_of_medoid(medoid_id)
                    temp(i,j) = 0;
                    dn = 1;
                    while dn <= dimension_num
                        temp(i,j) = temp(i,j) + (power(output(i,dn)-output(j,dn),2));
                        dn = dn+1;
                    end
                    temp(i,j) = power(temp(i,j),2);
                    j = j+1;
                end
                if i == 1
                    temp2(i) = sum(temp(1,:));
                temp2(number_of_medoid(medoid_id)) =
                    sum(temp(:,number_of_medoid(medoid_id)));
                else
                    temp2(i) = sum(temp(1:i-1,i));
                    temp2(i) = temp2(i) + sum(temp(i,i+1:number_of_medoid(medoid_id)));
                end
                i = i+1;
            end
        end
    end
end

```

```

minnest = 999999;
i = 1;
while i<=number_of_medoid(medoid_id)
    if temp2(i) < minnest
        initial_medoid(medoid_id) = data_in_medoid(medoid_id,i);
        minnest = temp2(i);
    end
    i = i+1;
end
end
medoid_id = medoid_id + 1;
end
iteration_training = iteration_training + 1;
end

```

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%                               ฟังก์ชันวิธีวัดระยะห่างแบบไดนามิกไทม์วอร์ปิง                               %%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

```

```

function [distDTW,matrix] = dtw(x,y);
% this is Dynamic Time Warping function
% [distDTW,matrix] = dtw(x,y);
dist = 0;
matrix(1,1)=0;
min =0;
length = size(x);
length = length(1,2);
a=2;
while a<=length+1
    matrix(a,1) = x(a-1);
    matrix(1,a) = y(a-1);

```

```

    a = a+1;
end
a = 2;
while a<=length+1
    b = 2;
    while b<=length+1
        dist = (x(a-1)-y(b-1))^2;
        if (a >= 3) && (b >= 3)
            if (matrix(a-1,b)<matrix(a,b-1))
                min=matrix(a-1,b);
            else
                min=matrix(a,b-1);
            end
            if (matrix(a-1,b-1)<min)
                min=matrix(a-1,b-1);
            end
            matrix(a,b) = dist + min;
        elseif (a==2)&&(b==2)
            matrix(a,b) = dist;
        elseif (a==2)
            matrix(a,b) = dist + matrix(a,b-1) ;
        elseif (b==2)
            matrix(a,b) = dist + matrix(a-1,b) ;
        end
        b = b+1;
    end
    a = a+1;
end
matrix = flipud(matrix);
distDTW = sqrt(matrix(1,length+1));

```

ภาคผนวก ข
ข้อมูลจริงที่นำมาศึกษา

ตารางผนวกที่ ข1 ข้อมูลการจำแนกคนไข้ทางสายตาเป็น 3 กลุ่มคือ กลุ่มที่ 1 คนไข้เหมาะกับคอนแทคเลนส์แบบแข็ง กลุ่มที่ 2 คนไข้เหมาะกับคอนแทคเลนส์แบบอ่อน และกลุ่มที่ 3 คนไข้ไม่เหมาะกับคอนแทคเลนส์ แบ่งตามคุณลักษณะ 4 ประเภท มีข้อมูลทั้งหมด 24 หน่วยตัวอย่าง

หน่วยตัวอย่าง	อายุ	ลักษณะของ อาการผิดปกติ	สายตาเอียง	อัตราน้ำหล่อ เลี้ยงตา	กลุ่ม
1	1	1	2	2	1
2	1	2	2	2	1
3	2	1	2	2	1
4	3	1	2	2	1
5	1	1	1	2	2
6	1	2	1	2	2
7	2	1	1	2	2
8	2	2	1	2	2
9	3	2	1	2	2
10	1	1	1	1	3
11	1	1	2	1	3
12	1	2	1	1	3
13	1	2	2	1	3
14	2	1	1	1	3
15	2	1	2	1	3
16	2	2	1	1	3
17	2	2	2	1	3
18	2	2	2	2	3
19	3	1	1	1	3
20	3	1	1	2	3
21	3	1	2	1	3
22	3	2	1	1	3

ตารางผนวกที่ ข1 (ต่อ)

หน่วยตัวอย่าง	อายุ	ลักษณะของ อาการผิดปกติ	สายตาเอียง	อัตราน้ำหล่อ เลี้ยงตา	กลุ่ม
23	3	2	2	1	3
24	3	2	2	2	3

ที่มา: Julien (1990)

ตารางผนวกที่ ข2 ข้อมูลการจำแนกคนไข้ที่เป็นมะเร็งเต้านม 2 กลุ่มคือ กลุ่มที่ 1 คนไข้ที่มีอายุยืนมากกว่า 5 ปีขึ้นไป และกลุ่มที่ 2 คนไข้ที่ตายภายใน 5 ปี แบ่งตามคุณลักษณะ 3 ประเภท คือ อายุของผู้ป่วยในช่วงเวลาที่สำรวจ ปีที่ผู้ป่วยมาเข้ารับการรักษา (หน่วยเป็นปีค.ศ.-1900) และจำนวนก้อนเนื้อที่เกิดขึ้น มีข้อมูลทั้งหมด 33 หน่วยตัวอย่าง

หน่วยตัวอย่าง	อายุของผู้ป่วย	ปีที่ป่วย	จำนวนก้อนเนื้อ	กลุ่ม
1	34	67	7	1
2	35	64	13	1
3	38	67	5	1
4	41	69	8	1
5	42	58	0	1
6	45	59	14	1
7	46	63	0	1
8	47	66	12	1
9	48	58	11	2
10	48	58	11	2
11	50	59	0	1
12	51	64	7	1
13	52	62	1	1

ตารางผนวกที่ ข2 (ต่อ)

หน่วยตัวอย่าง	อายุของผู้ป่วย	ปีที่ป่วย	จำนวนก้อนเนื้อ	กลุ่ม
14	53	60	9	2
15	53	65	12	2
16	53	58	1	1
17	53	61	1	1
18	54	60	3	1
19	54	63	19	1
20	56	66	3	2
21	57	64	9	1
22	58	59	0	1
23	58	58	0	1
24	59	64	4	1
25	59	64	7	1
26	59	67	3	1
27	60	65	0	2
28	62	58	0	1
29	67	63	1	2
30	67	66	0	1
31	70	58	0	2
32	70	59	8	1
33	72	67	3	1

ที่มา: Lim (1999)

ตารางผนวกที่ ข3 ข้อมูลผลการคัดแยกสับปะรดเป็น 2 กลุ่มคือ กลุ่มที่ 1 สับปะรดอ่อน และกลุ่มที่ 2 สับปะรดสุก ทดสอบด้วยเครื่องทดสอบสมบัติทางกายภาพของผลไม้และเก็บข้อมูล 3 ลักษณะ ได้แก่ ความถี่สั่นพ้อง ค่าเฉลี่ยความหนาแน่นเนื้อ และปริมาณน้ำตาล มีข้อมูลทั้งหมด 149 หน่วยตัวอย่าง

หน่วยตัวอย่าง	กลุ่ม	ความถี่สั่นพ้อง	ค่าเฉลี่ย ความหนาแน่นเนื้อ	ปริมาณน้ำตาล (%Brix)
1	1	420.141	2291.83	13.48
2	1	453.130	3418.80	11.74
3	1	454.864	2048.75	12.16
4	1	406.250	2640.48	13.44
5	1	482.643	2794.06	11.12
6	1	427.087	2254.02	14.26
7	1	371.530	2803.51	12.1
8	1	447.921	2662.34	11.4
9	1	432.292	2275.01	9.28
10	1	421.880	2056.04	11.08
11	1	453.128	2624.56	12.06
12	1	414.933	2008.71	10.28
13	1	435.768	2675.17	13.04
14	1	453.130	2596.48	11.52
15	1	439.237	2752.74	10.18
16	1	407.987	2196.05	13.16
17	1	399.306	2501.23	10.14
18	1	376.737	2182.71	10.46
19	1	432.293	1917.69	8.72
20	1	383.684	1851.20	10.12
21	1	418.406	2169.44	11.56
22	1	395.837	1696.30	11.5

ตารางผนวกที่ ข3 (ต่อ)

หน่วยตัวอย่าง	กลุ่ม	ความถี่สัมพัทธ์	ค่าเฉลี่ย ความหนาแน่นเนื้อ	ปริมาณน้ำตาล (%Brix)
23	1	442.710	2734.35	9.1
24	1	390.630	2350.13	10.62
25	1	428.822	2360.21	11.02
26	1	421.880	1828.36	9.94
27	1	307.292	1853.74	12.64
28	1	409.724	1754.02	12.34
29	1	399.307	2225.94	11.3
30	1	451.391	2309.27	11.32
31	1	357.643	2037.03	12.06
32	1	406.250	2353.69	12.58
33	1	321.183	1016.25	9.84
34	1	378.472	1782.35	13.1
35	1	342.014	2142.44	14.42
36	1	401.043	2241.78	15.16
37	1	397.572	2402.84	12.94
38	1	446.183	1936.37	13
39	1	406.250	2990.38	13.2
40	1	388.893	2397.49	10.84
41	1	416.671	2173.63	11.06
42	1	378.476	1313.69	15.48
43	1	406.253	1911.44	12.86
44	1	336.808	1699.31	11.44
45	1	355.907	2092.76	12.96
46	1	392.364	1945.69	10.92

ตารางผนวกที่ ข3 (ต่อ)

หน่วยตัวอย่าง	กลุ่ม	ความถี่สัมพัทธ์	ค่าเฉลี่ย ความหนาแน่นเนื้อ	ปริมาณน้ำตาล (%Brix)
47	1	345.489	1320.21	12.3
48	1	343.750	1737.65	12.44
49	1	338.544	1786.86	9.34
50	1	368.059	2371.94	10.58
51	1	375.001	1829.08	14.08
52	1	244.792	915.34	13.56
53	1	352.433	1499.42	13.28
54	1	362.850	1093.76	14.26
55	1	303.821	1427.86	14.28
56	1	296.877	1492.93	13.28
57	1	335.071	1528.99	12.36
58	1	338.543	1628.72	13.16
59	1	303.823	1383.60	13.3
60	1	326.391	1289.83	13.12
61	1	355.906	2054.71	10.98
62	1	274.308	1092.26	11.44
63	1	347.226	1308.18	13.18
64	1	302.084	1370.67	14.22
65	2	298.612	870.22	14.2
66	2	246.530	801.46	17.44
67	2	303.821	2437.73	13.58
68	2	276.043	1194.64	14.42
69	2	274.309	1132.83	15.28
70	2	236.114	705.86	13.64

ตารางผนวกที่ ข3 (ต่อ)

หน่วยตัวอย่าง	กลุ่ม	ความถี่สัมพัทธ์	ค่าเฉลี่ย ความหนาแน่นเนื้อ	ปริมาณน้ำตาล (%Brix)
71	2	289.933	1120.99	11.96
72	2	321.183	1334.52	13.78
73	2	321.182	1902.73	13.22
74	2	250.002	1223.62	14.6
75	2	288.197	865.32	12.9
76	2	296.877	639.35	12.62
77	2	263.890	1282.21	11.44
78	2	335.072	2133.68	12.42
79	2	309.030	1336.44	13.52
80	2	279.519	1401.39	14.62
81	2	307.293	1021.53	13.16
82	2	298.613	1400.52	14.14
83	2	246.529	1013.56	14.26
84	2	279.517	1910.94	11.5
85	2	267.362	1156.35	12.14
86	2	302.087	1159.27	12.16
87	2	331.600	1548.21	11.18
88	2	347.226	1673.42	12
89	2	237.849	656.84	10.04
90	2	262.156	646.64	14.34
91	2	239.586	373.82	14.28
92	2	265.627	471.03	8.46
93	2	215.280	436.81	9.3
94	2	216.798	959.76	13.88

ตารางผนวกที่ ข3 (ต่อ)

หน่วยตัวอย่าง	กลุ่ม	ความถี่สัมพัทธ์	ค่าเฉลี่ย ความหนาแน่นเนื้อ	ปริมาณน้ำตาล (%Brix)
95	2	246.530	578.97	11.42
96	2	229.168	732.25	13.06
97	2	256.947	262.40	9.18
98	2	217.014	558.77	7.3
99	2	234.377	446.77	14.86
100	2	203.128	342.02	14
101	1	468.751	1992.18	11.1
102	1	427.086	2050.75	12.14
103	1	435.764	1864.39	13.54
104	1	451.393	2387.87	12.5
105	1	460.070	2666.76	11.54
106	1	428.822	2429.06	14.22
107	1	432.293	2545.48	11.42
108	1	453.129	2653.89	12.08
109	1	437.500	1755.98	11.84
110	1	402.779	2105.21	10.28
111	1	423.616	1851.23	11.24
112	1	387.156	2592.79	13.04
113	1	404.514	1936.68	7.94
114	1	409.723	1818.43	11.3
115	1	381.947	2207.32	9.16
116	1	362.851	2440.23	11.96
117	1	359.379	1949.05	12.34
118	1	480.906	2136.27	12.96

ตารางผนวกที่ ข3 (ต่อ)

หน่วยตัวอย่าง	กลุ่ม	ความถี่สัมพัทธ์	ค่าเฉลี่ย ความหนาแน่นเนื้อ	ปริมาณน้ำตาล (%Brix)
119	1	387.154	1696.25	14.1
120	1	482.836	2487.08	13.2
121	1	362.851	2572.85	11.7
122	1	427.088	2161.95	12.54
123	1	461.810	3059.31	11.38
124	1	343.750	1537.87	11.26
125	1	420.143	2010.17	15.1
126	1	333.337	1575.91	14.88
127	1	328.130	2436.98	13.96
128	1	262.156	731.86	13.28
129	1	255.211	1211.21	11.52
130	1	274.308	1516.55	14.18
131	1	368.059	1646.87	13
132	2	296.879	653.41	13
133	2	321.183	1104.19	12.4
134	2	239.586	867.50	14.64
135	2	321.183	1709.57	15.26
136	2	345.491	1086.41	12.62
137	2	288.198	1328.02	14.52
138	2	230.906	749.27	14.36
139	2	321.183	1500.61	15.42
140	2	305.558	1278.13	13.6
141	2	314.238	1046.01	10.94
142	2	326.392	1568.48	12.78

ตารางผนวกที่ ข3 (ต่อ)

หน่วยตัวอย่าง	กลุ่ม	ความถี่สัมพัทธ์	ค่าเฉลี่ย ความหนาแน่นเนื้อ	ปริมาณน้ำตาล (%Brix)
143	2	307.293	941.61	13.98
144	2	250.001	992.20	14.44
145	2	236.113	525.88	14.74
146	2	232.642	892.70	14.88
147	2	251.738	625.23	10.74
148	2	239.583	318.37	15.1
149	2	222.223	281.93	14.22

ที่มา: Boonmung (2006)

ตารางผนวกที่ ข4 ข้อมูลต้นไอริส 3 กลุ่มคือ กลุ่มที่ 1 Iris Setosa กลุ่มที่ 2 Iris Versicolour และกลุ่มที่ 3 Iris Virginica แบ่งตามคุณลักษณะ 4 ประเภท คือ ความยาวของกลีบเลี้ยง ความกว้างของกลีบเลี้ยง ความยาวของกลีบดอก และความกว้างของกลีบดอก ซึ่งวัดในหน่วย ซม. มีข้อมูลทั้งหมด 150 หน่วยตัวอย่าง

หน่วยตัวอย่าง	ความยาวของ กลีบเลี้ยง	ความกว้าง ของกลีบเลี้ยง	ความยาวของ กลีบดอก	ความกว้าง ของกลีบดอก	กลุ่ม
1	5.1	3.5	1.4	0.2	1
2	4.9	3	1.4	0.2	1
3	4.7	3.2	1.3	0.2	1
4	4.6	3.1	1.5	0.2	1
5	5	3.6	1.4	0.2	1
6	5.4	3.9	1.7	0.4	1
7	4.6	3.4	1.4	0.3	1

ตารางผนวกที่ ข4 (ต่อ)

หน่วยตัวอย่าง	ความยาวของ กลีบเลี้ยง	ความกว้าง ของกลีบเลี้ยง	ความยาวของ กลีบดอก	ความกว้าง ของกลีบดอก	กลุ่ม
8	5	3.4	1.5	0.2	1
9	4.4	2.9	1.4	0.2	1
10	4.9	3.1	1.5	0.1	1
11	5.4	3.7	1.5	0.2	1
12	4.8	3.4	1.6	0.2	1
13	4.8	3	1.4	0.1	1
14	4.3	3	1.1	0.1	1
15	5.8	4	1.2	0.2	1
16	5.7	4.4	1.5	0.4	1
17	5.4	3.9	1.3	0.4	1
18	5.1	3.5	1.4	0.3	1
19	5.7	3.8	1.7	0.3	1
20	5.1	3.8	1.5	0.3	1
21	5.4	3.4	1.7	0.2	1
22	5.1	3.7	1.5	0.4	1
23	4.6	3.6	1	0.2	1
24	5.1	3.3	1.7	0.5	1
25	4.8	3.4	1.9	0.2	1
26	5	3	1.6	0.2	1
27	5	3.4	1.6	0.4	1
28	5.2	3.5	1.5	0.2	1
29	5.2	3.4	1.4	0.2	1
30	4.7	3.2	1.6	0.2	1
31	4.8	3.1	1.6	0.2	1
32	5.4	3.4	1.5	0.4	1

ตารางผนวกที่ ข4 (ต่อ)

หน่วยตัวอย่าง	ความยาวของ กลีบเลี้ยง	ความกว้าง ของกลีบเลี้ยง	ความยาวของ กลีบดอก	ความกว้าง ของกลีบดอก	กลุ่ม
33	5.2	4.1	1.5	0.1	1
34	5.5	4.2	1.4	0.2	1
35	4.9	3.1	1.5	0.2	1
36	5	3.2	1.2	0.2	1
37	5.5	3.5	1.3	0.2	1
38	4.9	3.6	1.4	0.1	1
39	4.4	3	1.3	0.2	1
40	5.1	3.4	1.5	0.2	1
41	5	3.5	1.3	0.3	1
42	4.5	2.3	1.3	0.3	1
43	4.4	3.2	1.3	0.2	1
44	5	3.5	1.6	0.6	1
45	5.1	3.8	1.9	0.4	1
46	4.8	3	1.4	0.3	1
47	5.1	3.8	1.6	0.2	1
48	4.6	3.2	1.4	0.2	1
49	5.3	3.7	1.5	0.2	1
50	5	3.3	1.4	0.2	1
51	7	3.2	4.7	1.4	2
52	6.4	3.2	4.5	1.5	2
53	6.9	3.1	4.9	1.5	2
54	5.5	2.3	4	1.3	2
55	6.5	2.8	4.6	1.5	2
56	5.7	2.8	4.5	1.3	2
57	6.3	3.3	4.7	1.6	2

ตารางผนวกที่ ข4 (ต่อ)

หน่วยตัวอย่าง	ความยาวของ กลีบเลี้ยง	ความกว้าง ของกลีบเลี้ยง	ความยาวของ กลีบดอก	ความกว้าง ของกลีบดอก	กลุ่ม
58	4.9	2.4	3.3	1	2
59	6.6	2.9	4.6	1.3	2
60	5.2	2.7	3.9	1.4	2
61	5	2	3.5	1	2
62	5.9	3	4.2	1.5	2
63	6	2.2	4	1	2
64	6.1	2.9	4.7	1.4	2
65	5.6	2.9	3.6	1.3	2
66	6.7	3.1	4.4	1.4	2
67	5.6	3	4.5	1.5	2
68	5.8	2.7	4.1	1	2
69	6.2	2.2	4.5	1.5	2
70	5.6	2.5	3.9	1.1	2
71	5.9	3.2	4.8	1.8	2
72	6.1	2.8	4	1.3	2
73	6.3	2.5	4.9	1.5	2
74	6.1	2.8	4.7	1.2	2
75	6.4	2.9	4.3	1.3	2
76	6.6	3	4.4	1.4	2
77	6.8	2.8	4.8	1.4	2
78	6.7	3	5	1.7	2
79	6	2.9	4.5	1.5	2
80	5.7	2.6	3.5	1	2
81	5.5	2.4	3.8	1.1	2
82	5.5	2.4	3.7	1	2

ตารางผนวกที่ ข4 (ต่อ)

หน่วยตัวอย่าง	ความยาวของ กลีบเลี้ยง	ความกว้าง ของกลีบเลี้ยง	ความยาวของ กลีบดอก	ความกว้าง ของกลีบดอก	กลุ่ม
83	5.8	2.7	3.9	1.2	2
84	6	2.7	5.1	1.6	2
85	5.4	3	4.5	1.5	2
86	6	3.4	4.5	1.6	2
87	6.7	3.1	4.7	1.5	2
88	6.3	2.3	4.4	1.3	2
89	5.6	3	4.1	1.3	2
90	5.5	2.5	4	1.3	2
91	5.5	2.6	4.4	1.2	2
92	6.1	3	4.6	1.4	2
93	5.8	2.6	4	1.2	2
94	5	2.3	3.3	1	2
95	5.6	2.7	4.2	1.3	2
96	5.7	3	4.2	1.2	2
97	5.7	2.9	4.2	1.3	2
98	6.2	2.9	4.3	1.3	2
99	5.1	2.5	3	1.1	2
100	5.7	2.8	4.1	1.3	2
101	6.3	3.3	6	2.5	3
102	5.8	2.7	5.1	1.9	3
103	7.1	3	5.9	2.1	3
104	6.3	2.9	5.6	1.8	3
105	6.5	3	5.8	2.2	3
106	7.6	3	6.6	2.1	3
107	4.9	2.5	4.5	1.7	3

ตารางผนวกที่ ข4 (ต่อ)

หน่วยตัวอย่าง	ความยาวของ กลีบเลี้ยง	ความกว้าง ของกลีบเลี้ยง	ความยาวของ กลีบดอก	ความกว้าง ของกลีบดอก	กลุ่ม
108	7.3	2.9	6.3	1.8	3
109	6.7	2.5	5.8	1.8	3
110	7.2	3.6	6.1	2.5	3
111	6.5	3.2	5.1	2	3
112	6.4	2.7	5.3	1.9	3
113	6.8	3	5.5	2.1	3
114	5.7	2.5	5	2	3
115	5.8	2.8	5.1	2.4	3
116	6.4	3.2	5.3	2.3	3
117	6.5	3	5.5	1.8	3
118	7.7	3.8	6.7	2.2	3
119	7.7	2.6	6.9	2.3	3
120	6	2.2	5	1.5	3
121	6.9	3.2	5.7	2.3	3
122	5.6	2.8	4.9	2	3
123	7.7	2.8	6.7	2	3
124	6.3	2.7	4.9	1.8	3
125	6.7	3.3	5.7	2.1	3
126	7.2	3.2	6	1.8	3
127	6.2	2.8	4.8	1.8	3
128	6.1	3	4.9	1.8	3
129	6.4	2.8	5.6	2.1	3
130	7.2	3	5.8	1.6	3
131	7.4	2.8	6.1	1.9	3
132	7.9	3.8	6.4	2	3

ตารางผนวกที่ ข4 (ต่อ)

หน่วยตัวอย่าง	ความยาวของ กลีบเลี้ยง	ความกว้าง ของกลีบเลี้ยง	ความยาวของ กลีบดอก	ความกว้าง ของกลีบดอก	กลุ่ม
133	6.4	2.8	5.6	2.2	3
134	6.3	2.8	5.1	1.5	3
135	6.1	2.6	5.6	1.4	3
136	7.7	3	6.1	2.3	3
137	6.3	3.4	5.6	2.4	3
138	6.4	3.1	5.5	1.8	3
139	6	3	4.8	1.8	3
140	6.9	3.1	5.4	2.1	3
141	6.7	3.1	5.6	2.4	3
142	6.9	3.1	5.1	2.3	3
143	5.8	2.7	5.1	1.9	3
144	6.8	3.2	5.9	2.3	3
145	6.7	3.3	5.7	2.5	3
146	6.7	3	5.2	2.3	3
147	6.3	2.5	5	1.9	3
148	6.5	3	5.2	2	3
149	6.2	3.4	5.4	2.3	3
150	5.9	3	5.1	1.8	3

ที่มา: Marshall (1988)

ประวัติการศึกษา และการทำงาน

ชื่อ –นามสกุล	นางสาวปิยธิดา รุจะศิริ
วัน เดือน ปี ที่เกิด	8 ธันวาคม 2526
สถานที่เกิด	อำเภอเมือง จังหวัดเพชรบุรี
ประวัติการศึกษา	ระดับปริญญาตรี คณะวิทยาศาสตร์ สาขาวิชาคณิตศาสตร์ (เกียรตินิยมอันดับ 2) มหาวิทยาลัยเกษตรศาสตร์ ระดับประกาศนียบัตรบัณฑิต สาขาวิชาการสอนวิทยาศาสตร์ มหาวิทยาลัยมหิดล
ทุนการศึกษาที่ได้รับ	โครงการส่งเสริมการผลิตครูผู้มีความสามารถพิเศษ ทางด้านวิทยาศาสตร์และคณิตศาสตร์ (สกว.) ในปี การศึกษา 2544-2548