**กิตติกรรมประกาศ**

โครงงานการจัดกลุ่มข้อมูล K-means แบบกึ่งมีผู้สอนประสบความสำเร็จไปได้ด้วยดี เนื่องจากได้รับการสนับสนุนเป็นอย่างดีตลอดระยะเวลาของการทำโครงงาน ผู้จัดทำโครงงานขอขอบพระคุณในความอนุเคราะห์จากอาจารย์และบุคคลต่าง ๆ ด้วยกัน

ผู้จัดทำขอขอบคุณ ผศ.ดร ศิริเพ็ญ วิกัยสุขสกุล และ อาจารย์สุจรรยา บุญประดิษฐ์ ซึ่งเป็นอาจารย์ที่ปรึกษาโครงงานที่คอยให้คำปรึกษา คำแนะนำ ข้อคิดเห็น แนวทางแก้ไขปัญหา คำชี้แนะที่เป็นประโยชน์ และยังติดตามความคืบหน้าของโครงงานอยู่เสมอ เพื่อสามารถแก้ไขปัญหาต่าง ๆ ที่เกิดขึ้นระหว่างการทำโครงงาน ทำให้โครงงานสำเร็จไปได้ด้วยดี

ขอขอบคุณ คณาอาจารย์ทุกท่านที่คอยอบรมสั่งสอน ทำให้มีความรู้และความสามารถในการทำโครงงาน

ขอขอบคุณ คณะกรรมการสอบโครงงานทุกท่านที่เปิดโอกาสในการนำเสอนโครงงาน ที่ช่วยตรวจแก้ไข ปรับปรุงและให้คำแนะนำที่ดีในการจัดทำโครงงาน ทำให้โครงงานนี้สมบูรณ์ยิ่งขึ้น

ขอขอบคุณ ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยสงขลานครินทร์ วิทยาเขตปัตตานี ที่ได้ให้ความอนุเคราะห์ห้องปฏิบัติการคอมพิวเตอร์และอุปกรณ์ต่าง ๆ เพื่อใช้ในการทำโครงงานรายงานบุคคลฉบับนี้

ขอขอบพระคุณ บิดา มารดา ผู้ที่ให้การสนับสนุน คอยดูแลและเป็นกำลังใจให้กับผู้จัดทำโครงงานอยู่เสมอ ทำให้สามารถดำเนินโครงงานนี้สำเร็จไปได้ด้วยดี

ขอขอบคุณ เพื่อนๆ นักศึกษาสาขาคณิตศาสตร์ประยุกต์ทุกคนที่ให้คำปรึกษาและข้อเสนอแนะ เป็นกำลังใจให้ในการทำโครงงานรายบุคคลนี้สำเร็จไปได้ด้วยดี

Mr. Sovannarith Phan

***ชื่อเรื่อง* การจัดกลุ่มข้อมูล** K-means **แบบกึ่งมีผู้สอน**

**ผู้เขียน**  Mr. Sovannarith Phan

**ชื่อปริญญา วิทยาศาสตรบัณฑิต**

**สาขาวิชา คณิตศาสตร์ประยุกต์**

**ปีการศึกษา** 2559

**อาจารย์ที่ปรึกษา ผศ.ดร.ศิริเพ็ญ วิกัยสุขสกุล และ อาจารย์สุจรรยา บุญประดิษฐ์**

**บทคัดย่อ**

ขั้นตอนวิธีการจัดกลุ่มข้อมูล K-means แบบกึ่งมีผู้สอน (semi-supervised K-means clustering) เป็นการจัดกลุ่มข้อมูลโดยใช้ชุดข้อมูลที่กำกับกลุ่มจำนวนหนึ่งเพื่อนำไปจัดกลุ่มข้อมูล โดยคำนวณหาระยะห่างระหว่างข้อมูลด้วยเกณฑ์วัดระยะห่างแบบยูคลิเดียน (Euclidean distance) การวัดระยะห่างแบบ Euclidean เป็นการคำนวณหาค่าระยะห่างระหว่างหน่วยตัวอย่างกับจุดศูนย์กลางซึ่งคำนวณมาจากค่าเฉลี่ยของสมาชิกในกลุ่ม หากจุดศูนย์กลางของกลุ่มมีระยะห่างที่ใกล้เคียงกัน แสดงว่ากลุ่มของข้อมูลอาจมีการซ้อนทับกัน ทำให้จัดกลุ่มหน่วยตัวอย่างได้ไม่ถูกต้อง โครงงานนี้ได้นำเกณฑ์วัดระยะห่างแบบ Mahalanobis มาใช้ในการจัดกลุ่มข้อมูลด้วยวิธี K-means แบบกึ่งมีผู้สอน โดยการวัดระยะห่างแบบ Mahalanobis เป็นการคำนวณหาระยะห่างระหว่างหน่วยตัวอย่างกับจุดศูนย์กลางที่ใช้ค่าเมทริกซ์ความแปรปรวนร่วมด้วย เกณฑ์วัดระยะห่างทั้งสองถูกนำมาใช้ในการจัดกลุ่มข้อมูลแบบกึ่งมีผู้สอนซึ่งมีสองวิธีคือ seeded K-means และ constrained K-means โครงงานนี้ได้พัฒนาโปรแกรมจากขั้นตอนวิธีและพัฒนาวิธีการจัดกลุ่มด้วยโปรแกรม R โดยใช้เกณฑ์วัดระยะห่างทั้งสองในวิธีการจัดกลุ่มดังกล่าว ได้โปรแกรมการจัดกลุ่ม 4 วิธี และนำไปทดสอบประสิทธิภาพการทำงานกับชุดข้อมูล 5 ชุดจาก UCI dataset ผลการวัดประสิทธิภาพแสดงให้เห็นว่าการใช้เกณฑ์วัดระยะห่างMahalanobis มีประสิทธิภาพการจัดกลุ่มข้อมูลดีกว่าการใช้เกณฑ์วัดระยะห่างEuclidean ในกรณีที่กลุ่มของข้อมูลมีการซ้อนทับกันมาก และมีประสิทธิภาพใกล้เคียงกันหากกลุ่มข้อมูลค่อนข้างแยกจากกัน และการเพิ่มจำนวนข้อมูลที่กำกับกลุ่มส่วนใหญ่ก็มีผลต่อการจัดกลุ่มข้อมูลทั้งสี่วิธี ทั้งนี้จำนวนข้อมูลที่กำกับกลุ่ม 30% ก็เพียงพอต่อการจัดกลุ่มข้อมูล

Title Semi-supervised K-means clustering

Author Mr. Sovannarith Phan

Program Bachelor of Science

Major Program Applied Mathematics

Academic 2016

Advisor Assist.Prof.Dr. Siripen Wikaisuksakul

Sujunya Boonpradit

Abstract

Semi-supervised K-means clustering uses some labeled data to aid the clustering of unlabeled data with Euclidean distance. The measures of Euclidean calculates the distance between a sample and a cluster center which is computed as the mean of each variable of samples within a cluster. If centers between clusters are close, it shows that the clusters of data maybe overlapped, and therefore, samples maybe in the wrong cluster. This work presents semi-supervised K-means clustering using Mahalanobis distance which calculates the distance between samples and cluster centers using covariance matrix. Two methods of semi-supervised K-means clustering are considered. They are seeded K-means and constrained K-means applying with the above distance measures, were developed coding with R program and experimented five dataset from UCI dataset. The clustering algorithms are coded with R programming. Four clustering programs are obtained and evaluated using five dataset from UCI repository. The results show that the performance of clustering data using Mahalanobis distance is better than the performance from Euclidean distance in case of overlapped data clusters and provides comparable performance in case of separated data. Increasing the amount of labeled data affects the performance of the four clustering methods, however, amount of labeled data 30% of the dataset is sufficient for clustering.

**สารบัญ**

**เรื่อง หน้า**

กิตติกรรมประกาศ………………………………………………………………………………………………………………….i

บทคัดย่อ (ภาษาไทย)……………………………………………………………………………………………………………..ii

บทคัดย่อ (ภาษาอังกฤษ)………………………………………………………………………………………………………..iii

สารบัญ………………………………………………………………………………………………………………………………..iv

สารบัญตาราง………………………………………………………………………………………………………………………viii

สารบัญภาพประกอบ …………………………………………………………………………………………………………..xiii

**บทที่ 1 บทนำ**

1.1 ความสำคัญและที่มาของโครงงาน…………………………………………………………………………1

1.2 วัตถุประสงค์ของการศึกษา…………………………………………………………………………………..1

1.3 ขอบเขตของการศึกษา ………………………………………………………………………………………..2

1.4 ประโยชน์ที่คาดว่าจะได้รับ …………………………………………………………………………………..2

1.5 ระยะเวลาในการดำเนินงาน …………………………………………………………………………………3

**บทที่ 2 ความรู้พื้นฐานและงานวิจัยที่เกี่ยวข้อง**

2.1 ความรู้พื้นฐานของการจัดกลุ่มข้อมูล …………………………………………………………………….4

2.1.1 หลักทั่วไปของการจัดกลุ่มข้อมูล ………………………………………………………………....4

2.1.2 การจัดกลุ่มข้อมูลแบบ K-means ………………………………………………………………..5

2.1.3 เกณฑ์การวัดระยะห่าง ……………………………………………………………………………….6

2.1.3.1 Euclidean distance ………………………………………………………………………6

2.1.3.2 Mahalanobis distance ………………………………………………………………….6

2.1.4 โครงสร้างข้อมูลนำเข้า ……………………………………………………………………………….6

2.1.5 การจัดกลุ่มข้อมูลแบบ semi-supervised K-means clustering ………………….7

2.1.5.1 ขั้นตอนวิธี seeded K-means …………………………………………………………7

**สารบัญ(ต่อ)**

**เรื่อง หน้า**

2.1.5.1 ขั้นตอนวิธี constrained K-means ……………………………………………..8

2.1.6 การวัดประสิทธิภาพของวิธีการจัดกลุ่มข้อมูลโดยการใช้ confusion matrix..8

2.1.6.1 Overall accuracy………………………………………………………………………9

2.1.6.2 Class accuracy …………………………………………………………………………9

2.2 งานวิจัยที่เกี่ยวข้อง ………………………………………………………………………………………..9

**บทที่ 3 วิธีการดำเนินงาน**

3.1 ขั้นตอนการดำเนินงาน …………………………………………………………………..……………..11

3.2 ชุดข้อมูลที่นำมาศึกษา ………………………………………………………………….……………….12

3.2.1 ชุดข้อมูล iris ……………………………………………………………………………………….12

3.2.2 ชุดข้อมูล seeds ………………………………………………………………………………….13

3.2.3 ชุดข้อมูล wine ……………………………………………………………………………………13

3.2.4 ชุดข้อมูล banknote authentication ………………………………………………….13

3.2.5 ชุดข้อมูล user knowledge modeling ………………………………………………..13

3.3 วิธีการเตรียมชุดข้อมูล …………………………………………………………………………………..14

3.4 การออกแบบการทดลอง ………………………………………………………………………………..15

3.4.1 การทดลองที่ 1 ……………………………………………………………………………………15

3.4.1 การทดลองที่ 2 ……………………………………………………………………………………16

3.5 การวัดประสิทธิภาพของวิธีการจัดกลุ่มข้อมูล ……………………………………………………16

**บทที่ 4 ผลและวิจารณ์ผลการทดลอง**

4.1 การพัฒนาขั้นตอนวิธีการจัดกลุ่มข้อมูล ……………………………………………………………17

4.1.1 ขั้นตอนวิธี seeded K-means with Mahalanobis (SKM) …………………….17

4.1.2 ขั้นตอนวิธี constrained K-means with Mahalanobis (CKM) ……………..19

4.2 ลักษณะของชุดข้อมูล ……………………………………………………………………………………….21

4.2.1 ชุดข้อมูล iris …………………………………………………………………………………………..21

4.2.2 ชุดข้อมูล seeds ……………………………………………………………………………………..22

4.2.3 ชุดข้อมูล wine ……………………………………………………………………………………….23

4.2.4 ชุดข้อมูล banknote authentication ………………………………………………………24

4.2.5 ชุดข้อมูล user knowledge modeling ……………………………………………………25

4.3 ผลการทดลองที่ 1 การใช้เกณฑ์วัดระยะห่างแบบ Euclidean ……………………………….26

4.3.1 ชุดข้อมูล iris ……………………………………………………………………………………........26

4.3.2 ชุดข้อมูล seeds ……………………………………………………………………………………..28

4.3.3 ชุดข้อมูล wine ………………………………………………………………………………….......30

4.3.4 ชุดข้อมูล banknote authentication ………………………………………………........32

4.3.5 ชุดข้อมูล user knowledge modeling ……………………………………………………33

4.4 ผลการทดลองที่ 2 การใช้เกณฑ์วัดระยะห่างแบบ Mahalanobis………………………….35

4.4.1 ชุดข้อมูล iris……………………………………………………………………………………………35

4.4.2 ชุดข้อมูล seeds ……………………………………………………………………………………..37

4.4.3 ชุดข้อมูล wine ……………………………………………………………………………………….39

4.4.4 ชุดข้อมูล banknote authentication ………………………………………………….....41

4.4.5 ชุดข้อมูล user knowledge modeling ……………………………………………………42

4.5 สรุปผลการทดลอง ……………………………………………………………………………………………44

4.5.1 สรุปผลการทดลองที่ 1 …………………………………………………………………………….44

4.5.2 สรุปผลการทดลองที่ 2……………………………………………………………………………..44

4.6 อภิปรายผล …………………………………………………………………………………………………….44

**บทที่ 5 สรุปผลการดำเนินงานและข้อแนะนำ**

5.1 สรุปการดำเนินงาน…………………………………………………………………………………………….46

**สารบัญ(ต่อ)**

**เรื่อง หน้า**

5.2 ข้อเสนอแนะ……………………………………………………………………………………………………..46

**บรรณานุกรม** ……………………………………………………………………………………………………………………48

**ภาคผนวก ก** …………………………………………………………………………………………………………………….49

ก.1 เมทริกซ์ค่าเฉลี่ยและเมทริกซ์ความแปรปรวนของชุดข้อมูล ………………………………….49

**ภาคผนวก ข** …………………………………………………………………………………………………………………….52

ข.1 ผลการทดลองที่ 1 การใช้เกณฑ์วัดระยะห่างแบบ Euclidean……………………………….52

ข.1.1 ผลการทดลองกับชุดข้อมูล iris...........................................................................52

ข.1.2 ผลการทดลองกับชุดข้อมูล seeds……………………………………………………………55

ข.1.3 ผลการทดลองกับชุดข้อมูล wine……………………………………………………………..58

ข.1.4 ผลการทดลองกับชุดข้อมูล banknote authentication……………………………62

ข.1.5 ผลการทดลองกับชุดข้อมูล user knowledge modeling………………………….65

ข.2 ผลการทดลองที่ 2 การใช้เกณฑ์วัดระยะห่างแบบ Mahalanobis………………………….69

ข.2.1 ผลการทดลองกับชุดข้อมูล iris………………………………………………………………….69

ข.2.2 ผลการทดลองกับชุดข้อมูล seeds……………………………………………………………72

ข.2.3 ผลการทดลองกับชุดข้อมูล wine……………………………………………………………..75

ข.2.4 ผลการทดลองกับชุดข้อมูล banknote authentication…………………………….79

ข.2.5 ผลการทดลองกับชุดข้อมูล user knowledge modeling…………………………..82

**สารบัญตาราง**

**ตารางที่ หน้า**

1.1 แผนการดำเนินงาน………………………………………………………………....……….........………………………3

3.1 สรุปลักษณะของข้อมูลทั้ง 5 ชุดได้แก่ จำนวนหน่วยตัวอย่างทั้งหมด ()

จำนวนกลุ่มข้อมูล () จำนวนหน่วยตัวอย่างแต่ละกลุ่ม () และจำนวนตัวแปร ()…………….14

4.1ค่าระยะห่างระหว่างจุดศูนย์กลางของแต่ละกลุ่มหน่วยตัวอย่างของตัวแปรทั้งหมดโดยใช้เกณฑ์วัดระยะห่าง Euclidean และ Mahalanobis กับชุดข้อมูล iris …………………………………………………...22

4.2ค่าระยะห่างระหว่างจุดศูนย์กลางของแต่ละกลุ่มหน่วยตัวอย่างของตัวแปรทั้งหมดโดยใช้เกณฑ์วัดระยะห่าง Euclidean และ Mahalanobis กับชุดข้อมูล seeds ………………………….…..............…..23

4.3ค่าระยะห่างระหว่างจุดศูนย์กลางของแต่ละกลุ่มหน่วยตัวอย่างของตัวแปรทั้งหมดโดยใช้เกณฑ์วัดระยะห่าง Euclidean และ Mahalanobis กับชุดข้อมูล wine …………………………………................23

4.4ค่าระยะห่างระหว่างจุดศูนย์กลางของแต่ละกลุ่มหน่วยตัวอย่างของตัวแปรทั้งหมดโดยใช้เกณฑ์วัดระยะห่าง Euclidean และ Mahalanobis กับชุดข้อมูล banknote authentication ……………...25

4.5ค่าระยะห่างระหว่างจุดศูนย์กลางของแต่ละกลุ่มหน่วยตัวอย่างของตัวแปรทั้งหมดโดยใช้เกณฑ์วัดระยะห่าง Euclidean และ Mahalanobis กับชุดข้อมูล user knowledge modeling …………....25

ก.1 เมทริกซ์ค่าเฉลี่ยและเมทริกซ์ความแปรปรวนของข้อมูล 3 กลุ่มในชุดข้อมูล iris …………………...49

ก.2 เมทริกซ์ค่าเฉลี่ยและเมทริกซ์ความแปรปรวนของข้อมูล 3 กลุ่มในชุดข้อมูล seeds ………………49

ก.3 เมทริกซ์ค่าเฉลี่ยและเมทริกซ์ความแปรปรวนของข้อมูล 3 กลุ่มในชุดข้อมูล wine…………………50

ก.4 เมทริกซ์ค่าเฉลี่ยและเมทริกซ์ความแปรปรวนของข้อมูล 3 กลุ่มในชุดข้อมูล banknote authentication …………………………………………………………………………………………………………………51

ก.5 เมทริกซ์ค่าเฉลี่ยและเมทริกซ์ความแปรปรวนของข้อมูล 3 กลุ่มในชุดข้อมูล user knowledge modeling …………………………………………………………………………………………………………………………51

ข.1 ค่า overall accuracy กับ class accuracy ของวิธี SKE กับวิธี CKE ของชุดข้อมูล iris ด้วยจำนวนข้อมูลที่กำกับกลุ่ม () = 5% 10% 20% 30% 40% และ 50% ……………………………..52

ข.2 จำนวนรอบของวิธี SKE กับวิธี CKE ของชุดข้อมูล iris ด้วยจำนวนข้อมูลที่กำกับกลุ่ม () = 5% 10% 20% 30% 40% และ 50% ……………………………………………………………………………………..53

**สารบัญตาราง(ต่อ)**

**ตารางที่ หน้า**

ข.3 ค่า WSCD ของวิธี SKE กับวิธี CKE ของชุดข้อมูล iris ด้วยจำนวนข้อมูลที่กำกับกลุ่ม () = 5% 10% 20% 30% 40% และ 50% ……………………………………………………………………………………..53

ข.4 ค่า  ของวิธี SKE กับวิธี CKE ของชุดข้อมูล iris ด้วยจำนวนข้อมูลที่กำกับกลุ่ม () = 5% 10% 20% 30% 40% และ 50% ……………………………………......54

ข.5 ค่า overall accuracy กับ class accuracy ของวิธี SKE กับวิธี CKE ของชุดข้อมูล seeds ด้วยจำนวนข้อมูลที่กำกับกลุ่ม () = 5% 10% 20% 30% 40% และ 50% ……………………………....55

ข.6 จำนวนรอบของวิธี SKE กับวิธี CKE ของชุดข้อมูล seeds ด้วยจำนวนข้อมูลที่กำกับกลุ่ม () = 5% 10% 20% 30% 40% และ 50% ……………………………………………………………………………….56

ข.7 ค่า WSCD ของวิธี SKE กับวิธี CKE ของชุดข้อมูล seeds ด้วยจำนวนข้อมูลที่กำกับกลุ่ม () = 5% 10% 20% 30% 40% และ 50% ……………………………………………………………………………….56

ข.8 ค่า  ของวิธี SKE กับวิธี CKE ของชุดข้อมูล seeds ด้วยจำนวนข้อมูลที่กำกับกลุ่ม () = 5% 10% 20% 30% 40% และ 50% ……………………………57

ข.9 ค่า overall accuracy กับ class accuracy ของวิธี SKE กับวิธี CKE ของชุดข้อมูล wine ด้วยจำนวนข้อมูลที่กำกับกลุ่ม () = 5% 10% 20% 30% 40% และ 50% ………………………………58

ข.10 จำนวนรอบของวิธี SKE กับวิธี CKE ของชุดข้อมูล wine ด้วยจำนวนข้อมูลที่กำกับกลุ่ม () = 5% 10% 20% 30% 40% และ 50% ……………………………………………………………………………….59

ข.11 ค่า WSCD ของวิธี SKE กับวิธี CKE ของชุดข้อมูล wine ด้วยจำนวนข้อมูลที่กำกับกลุ่ม () = 5% 10% 20% 30% 40% และ 50% ………………………………………………………………………………59

ข.12 ค่า  ของวิธี SKE กับวิธี CKE ของชุดข้อมูล wine ด้วยจำนวนข้อมูลที่กำกับกลุ่ม () = 5% 10% 20% 30% 40% และ 50% ……………………………..60

ข.13 ค่า overall accuracy กับ class accuracy ของวิธี SKE กับวิธี CKE ของชุดข้อมูล banknote authentication ด้วยจำนวนข้อมูลที่กำกับกลุ่ม () = 5% 10% 20% 30% 40% และ 50% ………………………………………………………………………………………………………………………..62

**สารบัญตาราง(ต่อ)**

**ตารางที่ หน้า**

ข.14 จำนวนรอบของวิธี SKE กับวิธี CKE ของชุดข้อมูล banknote authentication ด้วยจำนวนข้อมูลที่กำกับกลุ่ม () = 5% 10% 20% 30% 40% และ 50% ………………………………………....63

ข.15 ค่า WSCD ของวิธี SKE กับวิธี CKE ของชุดข้อมูล banknote authentication s ด้วยจำนวนข้อมูลที่กำกับกลุ่ม () = 5% 10% 20% 30% 40% และ 50% ................................................63

ข.16 ค่า  ของวิธี SKE กับวิธี CKE ของชุดข้อมูล banknote authentication ด้วยจำนวนข้อมูลที่กำกับกลุ่ม () = 5% 10% 20% 30% 40% และ 50%...64

ข.17ค่า overall accuracy กับ class accuracy ของวิธี SKE กับวิธี CKE ของชุดข้อมูล user knowledge modeling ด้วยจำนวนข้อมูลที่กำกับกลุ่ม () = 5% 10% 20% 30% 40% และ 50% ………………………………………………………………………………………………………………………………….65

ข.18 จำนวนรอบของวิธี SKE กับวิธี CKE ของชุดข้อมูล user knowledge modeling ด้วยจำนวนข้อมูลที่กำกับกลุ่ม () = 5% 10% 20% 30% 40% และ 50% ……………………………….66

ข.19 ค่า WSCD ของวิธี SKE กับวิธี CKE ของชุดข้อมูล user knowledge modeling ด้วยจำนวนข้อมูลที่กำกับกลุ่ม () = 5% 10% 20% 30% 40% และ 50% ………………………………………….66

ข.20 ค่า  ของวิธี SKE กับวิธี CKE ของชุดข้อมูล user knowledge modeling ด้วยจำนวนข้อมูลที่กำกับกลุ่ม () = 5% 10% 20% 30% 40% และ 50% ………………………………………………………………………………………………………………………………....67

ข.21 ค่า overall accuracy กับ class accuracy ของวิธี SKM กับวิธี CKM ของชุดข้อมูล iris ด้วยจำนวนข้อมูลที่กำกับกลุ่ม () = 5% 10% 20% 30% 40% และ 50% ……………………………….69

ข.22 จำนวนรอบของวิธี SKM กับวิธี CKM ของชุดข้อมูล iris ด้วยจำนวนข้อมูลที่กำกับกลุ่ม () = 5% 10% 20% 30% 40% และ 50% ………………………………………………………………………………70

ข.23 ค่า WSCD ของวิธี SKM กับวิธี CKM ของชุดข้อมูล iris ด้วยจำนวนข้อมูลที่กำกับกลุ่ม () = 5% 10% 20% 30% 40% และ 50% ………………………………………………………………………………70

ข.24 ค่า  ของวิธี SKM กับวิธี CKM ของชุดข้อมูล iris ด้วยจำนวนข้อมูลที่กำกับกลุ่ม () = 5% 10% 20% 30% 40% และ 50% ………………………………71

**สารบัญตาราง(ต่อ)**

**ตารางที่ หน้า**

ข.25 ค่า overall accuracy กับ class accuracy ของวิธี SKM กับวิธี CKM ของชุดข้อมูล seeds ด้วยจำนวนข้อมูลที่กำกับกลุ่ม () = 5% 10% 20% 30% 40% และ 50% …………………………72

ข.26 จำนวนรอบของวิธี SKM กับวิธี CKM ของชุดข้อมูล seeds ด้วยจำนวนข้อมูลที่กำกับกลุ่ม () = 5% 10% 20% 30% 40% และ 50% ……………………………………………………………………………73

ข.27 ค่า WSCD ของวิธี SKM กับวิธี CKM ของชุดข้อมูล seeds ด้วยจำนวนข้อมูลที่กำกับกลุ่ม () = 5% 10% 20% 30% 40% และ 50% …………………………………………………………………………….73

ข.28 ค่า  ของวิธี SKM กับวิธี CKM ของชุดข้อมูล seeds ด้วยจำนวนข้อมูลที่กำกับกลุ่ม () = 5% 10% 20% 30% 40% และ 50% …………………………….…74

ข.29 ค่า overall accuracy กับ class’accuracy ของวิธี SKM กับวิธี CKM ของชุดข้อมูล wine ด้วยจำนวนข้อมูลที่กำกับกลุ่ม () = 5% 10% 20% 30% 40% และ 50% …………………………75

ข.30 จำนวนรอบของวิธี SKM กับวิธี CKM ของชุดข้อมูล wine ด้วยจำนวนข้อมูลที่กำกับกลุ่ม () = 5% 10% 20% 30% 40% และ 50% ……………………………………………………………………………76

ข.31 ค่า WSCD ของวิธี SKM กับวิธี CKM ของชุดข้อมูล wine ด้วยจำนวนข้อมูลที่กำกับกลุ่ม () = 5% 10% 20% 30% 40% และ 50% ……………………………………………………………………………76

ข.32 ค่า  ของวิธี SKM กับวิธี CKM ของชุดข้อมูล wine ด้วยจำนวนข้อมูลที่กำกับกลุ่ม () = 5% 10% 20% 30% 40% และ 50% ………………………………77

ข.33 ค่า overall accuracy กับ class accuracy ของวิธี SKM กับวิธี CKM ของชุดข้อมูล banknote authentication ด้วยจำนวนข้อมูลที่กำกับกลุ่ม () = 5% 10% 20% 30% 40% และ 50% …………………………………………………………………………………………………………………………79

ข.34 จำนวนรอบของวิธี SKM กับวิธี CKM ของชุดข้อมูล banknote authentication ด้วยจำนวนข้อมูลที่กำกับกลุ่ม () = 5% 10% 20% 30% 40% และ 50%................................................80

ข.35 ค่า WSCD ของวิธี SKM กับวิธี CKM ของชุดข้อมูล banknote authentication ด้วยจำนวนข้อมูลที่กำกับกลุ่ม () = 5% 10% 20% 30% 40% และ 50%.................................................80

**สารบัญตาราง(ต่อ)**

**ตารางที่ หน้า**

ข.36 ค่า  ของวิธี SKM กับวิธี CKM ของชุดข้อมูล banknote authentication ด้วยจำนวนข้อมูลที่กำกับกลุ่ม () = 5% 10% 20% 30% 40% และ 50% ..81

ข.37 ค่า overall accuracy กับ class accuracy ของวิธี SKM กับวิธี CKM ของชุดข้อมูล user knowledge modeling ด้วยจำนวนข้อมูลที่กำกับกลุ่ม () = 5% 10% 20% 30% 40% และ 50% ………………………………………………………………………………………………………………………………..82

ข.38 จำนวนรอบของวิธี SKM กับวิธี CKM ของชุดข้อมูล user knowledge modeling ด้วยจำนวนข้อมูลที่กำกับกลุ่ม () = 5% 10% 20% 30% 40% และ 50% ………………………………………...83

ข.39 ค่า WSCD ของวิธี SKM กับวิธี CKM ของชุดข้อมูล user knowledge modeling ด้วยจำนวนข้อมูลที่กำกับกลุ่ม () = 5% 10% 20% 30% 40% และ 50% …………………………………………83

ข.40 ค่า  ของวิธี SKM กับวิธี CKM ของชุดข้อมูล user knowledge modeling ด้วยจำนวนข้อมูลที่กำกับกลุ่ม () = 5% 10% 20% 30% 40% และ 50% …………………………………………………………………………………………………………………………………84

**สารบัญภาพประกอบ**

**รูปที่ หน้า**

2.1 ตาราง confusion matrix ที่มีขนาด …………………………………………………………............8

3.1 ภาพรวมสำหรับขั้นตอนดำเนินงาน………………………………………………………………………………….11

3.2 การแบ่งชุดข้อมูลเพื่อทำการทดลอง………………………………………………………………………………...15

4.1 แผนภาพการทำงานของขั้นตอนวิธี SKM……………………………………………………………………….…18

4.2 แผนภาพการทำงานของขั้นตอนวิธี CKM………………………………………………………………………….20

4.3แผนภาพการกระจายของชุดข้อมูล iris………………………………………………………………………….…21

4.4แผนภาพการกระจายของชุดข้อมูล seeds……………………………………………………………………….22

4.5แผนภาพการกระจายของชุดข้อมูล wine…………………………………………………………………………24

4.6แผนภาพการกระจายของชุดข้อมูล banknote authentication………………………………………..24

4.7แผนภาพการกระจายของชุดข้อมูล user knowledge modeling………………………………………26

4.8 ค่า overall accuracy ของวิธี SKE กับวิธี CKEของชุดข้อมูล iris……………………………………….26

4.9 (a)-(c) class accuracy ของวิธี SKE กับวิธี CKE ของชุดข้อมูล iris……………………………………..27

4.10 ค่า overall accuracy ของวิธี SKE กับวิธี CKEของชุดข้อมูล seeds………………………………..28

4.11 (a)-(c) class accuracy ของวิธี SKE กับวิธี CKE ของชุดข้อมูล seeds………………………………29

4.12 ค่า overall accuracy ของวิธี SKE กับวิธี CKEของชุดข้อมูล wine.......................................30

4.13 (a)-(c) class accuracy ของวิธี SKE กับวิธี CKE ของชุดข้อมูล wine……………………………....31

4.14 ค่า overall accuracy ของวิธี SKE กับวิธี CKEของชุดข้อมูล banknote authentication................................................................................................................................32

4.15 (a)-(c) class accuracy ของวิธี SKE กับวิธี CKE ของชุดข้อมูล banknote authentication………………………………………………………………………………………………………………….32

4.16 ค่า overall accuracy ของวิธี SKE กับวิธี CKEของชุดข้อมูล user knowledge modeling………………………………………………………………………………………………………………………….33

**สารบัญภาพประกอบ(ต่อ)**

**รูปที่ หน้า**

4.17 (a)-(c) class accuracy ของวิธี SKE กับวิธี CKE ของชุดข้อมูล user knowledge modeling………………………………………………………………………………………………………………………….34

4.18 ค่า overall accuracy ของวิธี SKM กับวิธี CKMของชุดข้อมูล iris………………………………....35

4.19 (a)-(c) class accuracy ของวิธี SKM กับวิธี CKM ของชุดข้อมูล iris………………………………..36

4.20 ค่า overall accuracy ของวิธี SKM กับวิธี CKMของชุดข้อมูล seeds………………………….…37

4.21 (a)-(c) class accuracy ของวิธี SKM กับวิธี CKM ของชุดข้อมูล seeds…………………………..38

4.22 ค่า overall accuracy ของวิธี SKM กับวิธี CKMของชุดข้อมูล wine……………………………...39

4.23 (a)-(c) class accuracy ของวิธี SKM กับวิธี CKM ของชุดข้อมูล wine………………………….…40

4.24 ค่า overall accuracy ของวิธี SKM กับวิธี CKMของชุดข้อมูล banknote authentication………………………………………………………………………………………………………………...41

4.25 (a)-(c) class accuracy ของวิธี SKM กับวิธี CKM ของชุดข้อมูล banknote authentication……....................................................................................................................…..41

4.26 ค่า overall accuracy ของวิธี SKM กับวิธี CKMของชุดข้อมูล user knowledge modeling………...............................................................................................................................42

4.27 (a)-(c) class accuracy ของวิธี SKE กับวิธี CKE ของชุดข้อมูล user knowledge modeling………...............................................................................................................................43