การจัดกลุ่มนักศึกษาสำหรับวางแผนอบรมเสริมความรู้ก่อนจบการศึกษา สาขาคอมพิวเตอร์ศึกษา มหาวิทยาลัยราชภัฏบ้านสมเด็จเจ้าพระยา ด้วยเทคนิคการเรียนรู้ของเครื่อง

Student Clustering for Planning Training to Enhance Knowledge before Graduation Computer Education Bansomdetchaopraya Rajabhat University with Machine Learning Techniques

อังคาร ปริญญาชัยศักดิ์ [1]
aungkarn.pr@bsru.ac.th
นิธิวดี พะเทพ [1]
nitivadee.pa@bsru.ac.th
เกียรติขร โสภณาภรณ์ [1]
Kiatikhorn.so@bsru.ac.th
[1] สังกัด สาขาคอมพิวเตอร์ศึกษา มหาวิทยาลัยราชภัฏบ้านสมเด็จเจ้าพระยา

บทคัดย่อ

การวิจัยในครั้งนี้มีวัตถุประสงค์เพื่อหาจำนวนกลุ่มที่เหมาะสม และวิเคราะห์คุณลักษณะของกลุ่มที่ทำการจัดไว้เพื่อใช้วางแผนการอบรมเสริมความรู้ก่อนจบการศึกษา มีกลุ่มตัวอย่างเป็นนักศึกษาชั้นปีที่ 4 สาขาคอมพิวเตอร์ศึกษา มหาวิทยาลัยราชภัฏบ้านสมเด็จเจ้าพระยาที่เข้ารับการสอบวัดความรู้ทางคอมพิวเตอร์ จำนวน 54 คน นำข้อมูลผลการสอบมาทำการหาจำนวนกลุ่มที่เหมาะสมด้วยวิธี Eblow ได้จำนวนกลุ่มที่เหมาะสมเป็น 4 กลุ่ม (แบ่งนักศึกษาเป็นกลุ่ม 0 ,1 ,2 และกลุ่ม 3) และวิธี Silhouette Coefficient ได้จำนวนกลุ่มที่เหมาะสมเป็น 4 กลุ่ม (Silhouette Score = 0.80) ทำการแบ่งนักศึกษาเป็นกลุ่มโดยใช้เทคนิคการเรียนรู้ของเครื่อง ด้วยอัลกอริทึม K-Means ได้ผลดังนี้ นักศึกษากลุ่ม 0 จำนวน 20 คน นักศึกษากลุ่มที่ 1 จำนวน 8 คน นักศึกษากลุ่มที่ 2 จำนวน 13 คน และกลุ่มที่ 3 จำนวน 13 คน เมื่อทำการวิเคราะห์คุณลักษณะของกลุ่มพบว่า นักศึกษาที่ถูกจัดอยู่ในกลุ่ม 1-2 ควรได้รับการอบรมเพิ่มเติมความรู้ในด้านองค์การและระบบสารสนเทศ ในด้านเทคโนโลยีเพื่องานประยุกต์พบว่านักศึกษาทุกกลุ่มมีผลคะแนนที่ดีอยู่จึงไม่ต้องทำการจัดอบรม ในด้านเทคโนโลยีและวิธีการทางซอฟต์แวร์พบว่านักศึกษาที่ถูกจัดอยู่ในกลุ่ม 2 ควรได้รับการจัดอบรม ในด้านโครงสร้างพื้นฐานระบบพบว่านักศึกษาที่ถูกจัดอยู่ในกลุ่ม 1 และกลุ่ม 3 ควรได้รับการอบรม ในด้านฮาร์ดแวร์และสถาปัตยกรรมคอมพิวเตอร์พบว่านักศึกษาที่ถูกจัดอยู่ในกลุ่ม 0 กลุ่ม 2 และกลุ่ม 3 ควรได้รับการจัดอบรม

คำสำคัญ: การจัดกลุ่ม, การเรียนรู้ของเครื่อง

Abstract

This research focuses on finding the optimal number of groups and analyzing the characteristics of the group for planning the training course to enhance college student knowledge before they graduate. The sample group was 4th year students in the field of Computer Education Bansomdetchaopraya Rajabhat University who took the computer literacy test 54 students. The results of the test were used to find the optimal number of groups by Eblow and Silhouette Coefficient method. The optimal number of groups was divided into 4 groups Silhouette Score = 0.80 (dividing students into groups 0, 1, 2 and group 3). Divide students into groups by use machine learning techniques with the K-Means algorithm. The results were as follows : 20 students in Group 0, 8 students in Group 1, 13 students in Group 2 and 13 students in Group 3. When analyzing the characteristics of the group, it was found that students placed in groups 1-2 should receive additional training in organization and information systems. In the field of applied technology, it was found that all groups of students had good grades, so they did not need to organize training. In the field of technology and software methods, it was found that students in Group 2 should receive training. In the field of system infrastructure, it was found that students in Group 1, Group 3 should be trained. In the field of hardware and computer architecture, it was found that students in Group 0, Group 2 and Group 3 should receive training.

Keyword: clustering, machine Learning

บทนำ

การผลิตบัณฑิตเป็นพันธกิจหลักของหลักสูตรคอมพิวเตอร์ศึกษา มหาวิทยาลัยราชภัฏบ้านสมเด็จเจ้าพระยา ซึ่งในการจะผลิตบัณฑิตให้มีความพร้อมในการทำงานและเป็นที่ต้องการขององค์กรต่างๆ ภายใต้งบประมาณที่มีอย่างจำกัดนั้นจำเป็นจะต้องมีข้อมูลที่ดีและมีประสิทธิภาพเพื่อใช้สำหรับการวางแผนในการพัฒนานักศึกษาโดยเฉพาะนักศึกษาครูที่จะต้องออกไปทำหน้าที่ในการสอนเด็กนักเรียนเพื่อให้เติบโตขึ้นมาเป็นกำลังสำคัญในการพัฒนาประเทศชาติในอนาคต

การนำเทคนิคการเรียนรู้ของเครื่องซึ่งเป็นส่วนหนึ่งของปัญญาประดิษฐ์มาใช้ในการวางแผนในการอบรมเสริมความรู้ให้นักศึกษาก่อนจบการศึกษา จะทำให้หลักสูตรสามารถประเมินสถานการณ์และวางแผนได้อย่างแม่นยำ และมีประสิทธิภาพ โดยในการวิจัยนี้ผู้วิจัยได้ใช้เทคนิคการเรียนรู้ของเครื่องแบบไม่ใช้ผู้สอน เพื่อทำการจัดกลุ่มผู้เรียนสำหรับนำข้อมูลไปวางแผนในการจัดอบรมเสริมความรู้ให้นักศึกษาก่อนจบการศึกษา ในหลักสูตรคอมพิวเตอร์ศึกษา มหาวิทยาลัยราชภัฏบ้านสมเด็จเจ้าพระยา

วัตถประสงค์การวิจัย

1. เพื่อหาจำนวนกลุ่มที่เหมาะสมในการจัดอบรมเสริมความรู้ก่อนจบการศึกษา
2. เพื่อวิเคราะห์คุณลักษณะของกลุ่มเพื่อใช้วางแผนการอบรมเสริมความรู้ก่อนจบการศึกษา
3.

ขอบเขตการวิจัย

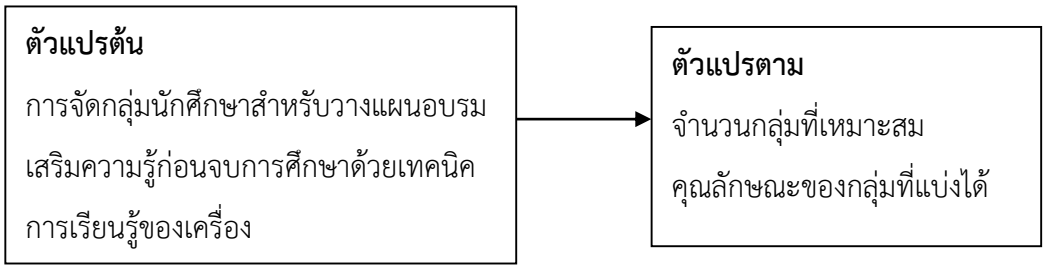
งานวิจัยนี้ มีกลุ่มประชากร ได้แก่นักศึกษาชั้นปีที่ 4 สาขาคอมพิวเตอร์ศึกษา มหาวิทยาลัยราชภัฏบ้านสมเด็จเจ้าที่เข้ารับการสอบวัดความรู้ทางคอมพิวเตอร์ จำนวน 54 คน โดยจะใช้กลุ่มประชากรทั้งหมดเป็นกลุ่มตัวอย่าง

การวิจัยครั้งนี้ใช้ข้อมูลการสอบวัดความรู้ทางคอมพิวเตอร์ศึกษาของนักศึกษาชั้นปีที่ 4 สาขาคอมพิวเตอร์ศึกษา และทำการจัดกลุ่มด้วยการใช้เทคนิคการเรียนรู้ของเครื่อง (Machine Learning) แบบไม่มีผู้สอน (Unsupervised Learning) โดยใช้ภาษา Python และใช้เครื่องมือ Jupyter Lab ในการประมวลผลและนำเสนอข้อมูล อัลกอริทึมที่ใช้ในการแบ่งกลุ่มในการวิจัยนี้จะใช้อัลกอริทึม K-means

ข้อมูลผลคะแนนการสอบวัดความรู้ทางคอมพิวเตอร์ที่นำมาใช้ในการประมวลผลจะแบ่งออกเป็น 5 ด้านโดยแบ่งตาม มาตรฐานคุณวุฒิระดับปริญญาตรี สาขาคอมพิวเตอร์ พ.ศ. 2552 และใช้เทคนิคการปรับลดมิติของข้อมูลด้วยเทคนิคการวิเคราะห์องค์ประกอบหลัก (PCA : Principle Component Analysis) เพื่อให้ง่ายต่อการวิเคราะห์ผล และการนำข้อมูลไปใช้งาน

วิธีดำเนินการวิจัย

1. กรอบแนวคิดการวิจัย



2. ขั้นตอนการวิจัย

2.1 เตรียมข้อมูล Data Preprocessing

เตรียมข้อมูลสำหรับการทำ Machine Learning โดยขั้นตอนนี้จะทำการเตรียมข้อมูลให้เหมาะสมต่อการนำไปจัดกลุ่ม เช่น ตัดเอาคอลัมน์ที่ไม่ได้ใช้งานและนำชื่อข้อมูลส่วนตัวของผู้เข้าสอบออก ให้คงเหลือไว้แต่ส่วนของคะแนนต่างๆ จากขั้นทำการปรับชื่อคอลัมน์ให้เหมาะแก่การนำข้องมูลไปประมวลผล

2.2 ทำการปรับลดมิติของข้อมูล (Dimensional Reduction)

ทำการปรับลดมิติของข้อมูลจาก 5 มิติให้เหลือ 2 มิติเพื่อให้ง่ายต่อการประมวลผลและการนำเสนอข้อมูลด้วยเทคนิคการวิเคราะห์องค์ประกอบหลัก (PCA : Principle Component Analysis) ผลคะแนนทั้ง 5 ด้านจะถูกแปลงเป็นข้อมูลใหม่ 2 คอลัมน์ที่ยังคงลักษณะความสัมพันธ์ของข้อมูลเดิมไว้

2.3 หาจำนวนกลุ่มที่เหมาะสม (Optimal Cluster Number)

ทำการคำนวณหาจำนวนกลุ่มที่เหมาะสมโดยการทดลองแบ่งกลุ่มนักศึกษาออกเป็น 2-7 กลุ่มและหาจำนวนกลุ่มที่เหมาะสมด้วยวิธี Eblow และ Silhouette Coefficient เพื่อนำข้อมูลไปใช้ในการจัดกลุ่ม

2.4 จัดกลุ่ม (Clustering)

ทำการแบ่งกลุ่มนักศึกษาตามจำนวนกลุ่มที่เหมาะสมด้วยอัลกอริทึม K-Means

2.5 วิเคราะห์หาคุณลักษณะของกลุ่ม (Cluster Analysis)

ทำการวิเคราะห์หาคุณลักษณะของกลุ่มที่แบ่งได้เพื่อนำข้อมูลที่ได้ไปใช้วางแผนในการจัดอบรมเสริมความรู้ก่อนจบการศึกษา

3. การวิเคราะห์ข้อมูล

การหาจำนวนกลุ่มที่เหมาะสม

3.1 วิธี Eblow

$$SSE = \sum_{K=1}^{K} \sum_{x_i \in S_K} \|X_i - C_k\|_2^2$$

เป็นวิธีหาค่ากลุ่มที่เหมาะสมในการกำหนดจำนวนกลุ่ม โดยใช้การวัดข้อผิดพลาดของผลรวมระยะห่างระหว่าง Object กับค่ากลางของแต่ละกลุ่ม (Centroid) เมื่อมีการจัดกลุ่มเป็นจำนวนมากขึ้นจะเกิดตัวแทนกลุ่ม (Centroid) ทำให้เกิดระยะห่างระหว่างข้อมูลและตัวแทนกลุ่มน้อยลงจนความชันของกราฟเข้าสู่ความราบเรียบจนเกิดจุดที่เป็นมุมเหมือนข้อศอก โดยจุดนี้จะเป็นจุดที่ให้ค่าจำนวนกลุ่ม (Cluster) ที่ดีที่สุด

3.2 วิธี Silhouette Coefficient

$$sil(i) = \frac{b(i) - a(i)}{max\{a(i), b(i)\}}$$

คะแนน Silhouette ที่ได้มีค่าตั้งแต่ 1 ถึง -1 สามารถแปลความดังนี้

ค่าเข้าใกล้ 1 หมายถึง ข้อมูลถูกจัดอยู่ในกลุ่มที่เหมาะสมแล้ว

ค่าเข้าใกล้ 0 หมายถึง ข้อมูลอาจเหมาะกับการจัดอยู่กลุ่มอื่น

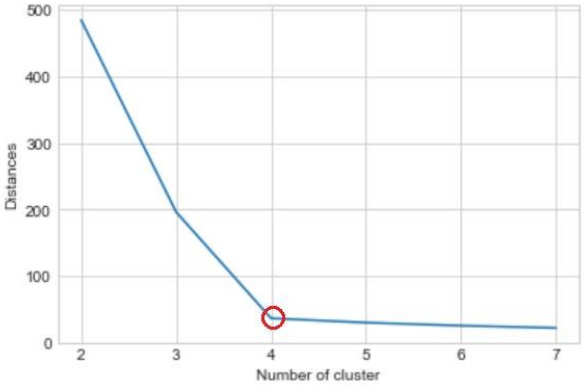
ค่าเข้าใกล้ -1 หมายถึง ข้อมูลอาจถูกจัดอยู่ผิดกลุ่ม

ผลการวิจัย

จำนวนกลุ่มที่เหมาะสมในการจัดอบรมเสริมความรู้ก่อนจบการศึกษา ผู้วิจัยได้ทำการทดลองแบ่งกลุ่มตั้งแต่ 2 กลุ่มไปจนถึง 7 กลุ่มและทำการคำนวนหาจำนวนกลุ่มที่เหมาะสมที่สุดด้วยวิธี Eblow และ Silhouette Coefficient ได้ผลดังนี้

ผลการหาจำนวนกลุ่มที่เหมาะสมด้วยวิธี Eblow

ผลการหาจำนวนกลุ่มที่เหมาะสมด้วยวิธี Eblow โดยใช้การวัดข้อผิดพลาดของผลรวมระยะห่างระหว่างข้อมูลแต่ละรายการ กับค่ากลางของแต่ละกลุ่ม (Centroid) โดยค่าที่เหมาะสมก่อนกราฟจะเข้าสู่ความราบเรียบจากการแบ่งกลุ่มมากเกินไป ได้จำนวนกลุ่มที่เหมาะสมคือการแบ่งนักศึกษาออกเป็น 4 กลุ่ม



ภาพที่ 1 กราฟแสดงจำนวนกลุ่มที่เหมาะสมด้วยวิธี Eblow

ผลการหาจำนวนกลุ่มที่เหมาะสมด้วยวิธี Silhouette Coefficient

ทำการคำนวนค่า Silhouette score ซึ่งเป็นคะแนนที่ใช้วัดว่าข้อมูลที่เราทำการจัดกลุ่มไว้ว่ามีความคล้ายกับกลุ่ม (Cluster) ที่มันอยู่มากน้อยเพียงใดเมื่อเทียบกับกลุ่ม (Cluster) อื่น

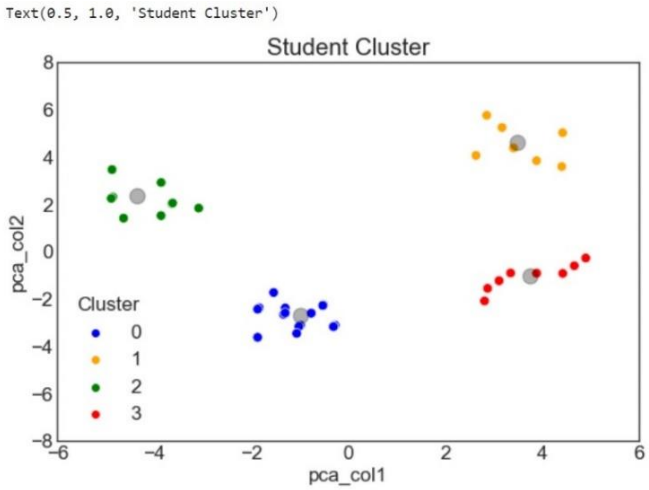
ตารางที่ 1 ผลคะแนน Silhouette score จากการทดลองแบ่งกลุ่ม 2 – 7 กลุ่ม

จำนวนกลุ่ม	Silhouette score
2	0.53
3	0.68
4	0.80
5	0.71
6	0.55
7	0.51

จากผลการคำนวณด้วยวิธี Silhouette Coefficient พบว่าการแบ่งกลุ่มนักศึกษาออกเป็น 4 กลุ่มมีความเหมาะสมที่สุดเนื่องจากมีค่า Silhouette score สูงที่สุดที่ 0.80 โดยผลที่ได้จากทั้ง 2 วิธีเป็นไปในแนวทางเดียวกันคือ การแบ่งกลุ่มนักศึกษาในการจัดอบรมเสริมทักษะทางคอมพิวเตอร์ก่อนบการศึกษาเป็น 4 กลุ่มมีความเหมาะสมมากที่สุด

ผลการวิเคราะห์คุณลักษณะของกลุ่ม (Cluster Analysis)

ในการวิเคราะห์คุณลักษณะของกลุ่มผู้วิจัยจะใช้เทคนิคการเรียนรู้ของเครื่อง (Machine Learning) แบบไม่ใช้ผู้สอน (Unsupervised Learning) และอัลกอริทึมที่ใช้ในการแบ่งกลุ่มในการวิจัยนี้จะใช้อัลกอริทึม K-meansในการแบ่งกลุ่ม ได้ผลการจัดกลุ่มดังนี้



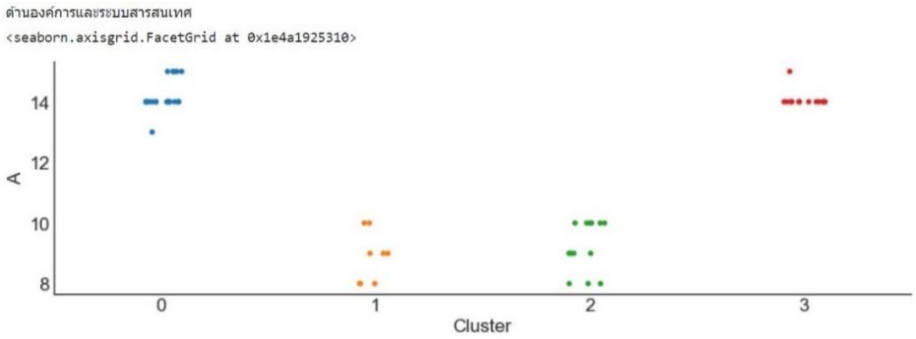
ภาพที่ 2 ผลการแบ่งกลุ่มนักศึกษาด้วยอัลกอริทึม K-Means

จากรูปกราฟแสดงผลการจัดกลุ่มของคะแนนสอบที่ผ่านการลดมิติข้อมูลเหลือ 2 มิติ โดยแบ่งเป็น 4 กลุ่มตามจำนวนกลุ่มที่คำนวณได้ โดยแต่ละจุดสีจะแสดงถึงข้อมูลผลการสอบนักศึกษาแต่ละคน และจุดสีเทาแสดงถึงจุดกึ่งกลางของกลุ่ม

ตารางที่ 2 จำนวนนักศึกษาที่ถูกจัดไว้ในแต่ละกลุ่ม

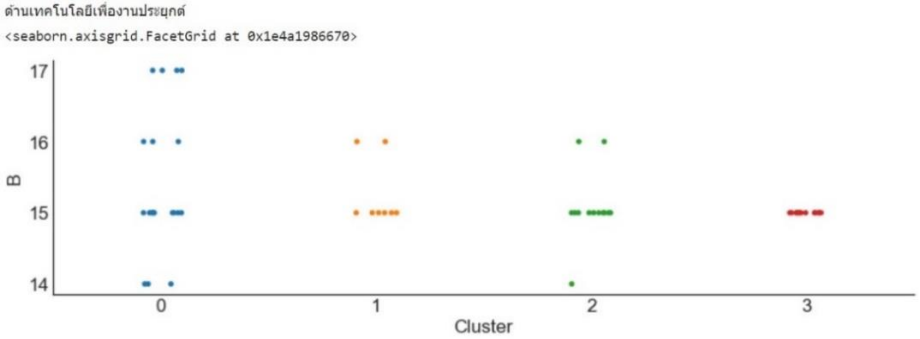
กลุ่ม	จำนวนคน
0	20
1	8
2	13
3	13
รวม	54

ข้อมูลในตารางแสดงจำนวนนักศึกษาในแต่ละกลุ่มที่ถูกจัดกลุ่มด้วยอัลกอริทึม K-Means เมื่อทำการวิเคราะห์คุณลักษณะของกลุ่มแต่ละกลุ่มตามทักษะของการสอบทั้ง 5 ด้านได้ผลดังนี้



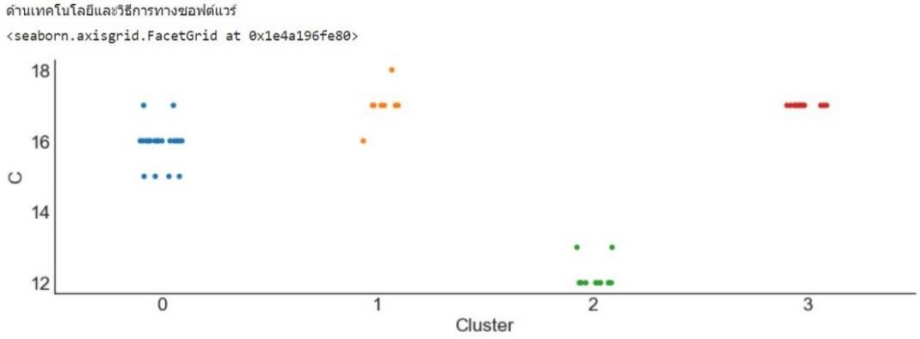
ภาพที่ 3 ผลการวิเคราะห์คุณลักษณะของกลุ่มแบ่งตามทักษะด้านองค์การและระบบสารสนเทศ

ในด้านองค์การและระบบสารสนเทศพบว่านักศึกษาที่ถูกจัดอยู่ในกลุ่ม 1-2 ควรได้รับการอบรมเพิ่มเติมความรู้ในด้านองค์การและระบบสารสนเทศ



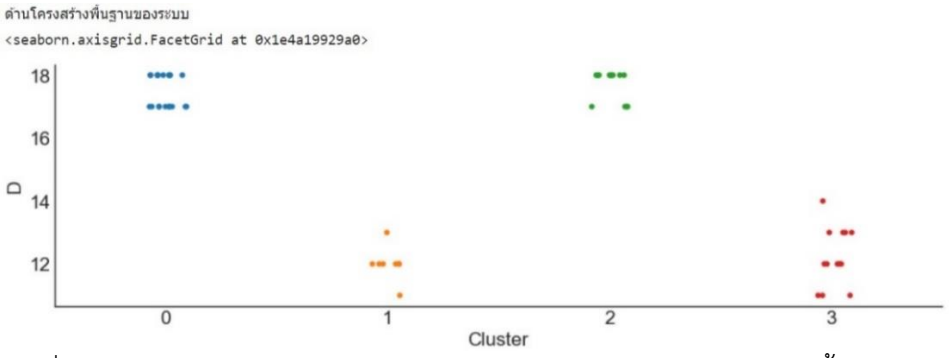
ภาพที่ 4 ผลการวิเคราะห์คุณลักษณะของกลุ่มแบ่งตามทักษะด้านเทคโนโลยีเพื่องานประยุกต์

ในด้านเทคโนโลยีเพื่องานประยุกต์พบว่านักศึกษาทุกกลุ่มมีผลคะแนนที่ดีอยู่แล้วจึงไม่ต้องวางแผนเพื่ออบรมเสริมความรู้อีก



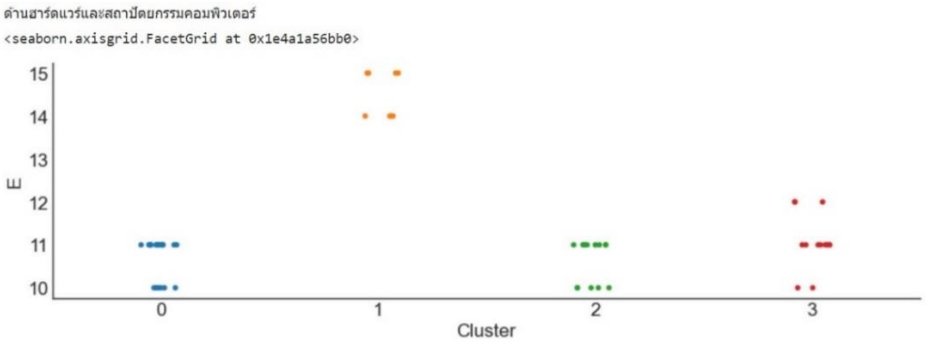
ภาพที่ 5 ผลการวิเคราะห์คุณลักษณะของกลุ่มแบ่งตามทักษะด้านเทคโนโลยีและวิธีการทางซอฟต์แวร์

ในด้านเทคโนโลยีและวิธีการทางซอฟต์แวร์พบว่านักศึกษาที่ถูกจัดอยู่ในกลุ่ม 2 ควรได้รับการจัดอบรมเสริมความรู้ในด้านนี้เพิ่มเติมก่อนจบการศึกษา



ภาพที่ 6 ผลการวิเคราะห์คุณลักษณะของกลุ่มแบ่งตามทักษะด้านโครงสร้างพื้นฐานระบบ

ในด้านโครงสร้างพื้นฐานระบบพบว่านักศึกษาที่ถูกจัดอยู่ในกลุ่ม 1 และกลุ่ม 3 ควรได้รับการอบรมเสริมความรู้ในด้านนี้เพิ่มเติมก่อนจบการศึกษา



ภาพที่ 7 ผลการวิเคราะห์คุณลักษณะของกลุ่มแบ่งตามทักษะด้านฮาร์ดแวร์และสถาปัตยกรรมคอมพิวเตอร์

ในด้านฮาร์ดแวร์และสถาปัตยกรรมคอมพิวเตอร์พบว่านักศึกษาที่ถูกจัดอยู่ในกลุ่ม 0 กลุ่ม 2 และกลุ่ม 3 ควรได้รับการจัดอบรมเสริมความรู้ในด้านนี้เพิ่มเติมก่อนจบการศึกษา

อภิปรายผล

จากการจัดกลุ่มนักศึกษาสำหรับวางแผนอบรมเสริมความรู้ก่อนจบการศึกษา สาขาคอมพิวเตอร์ศึกษา มหาวิทยาลัยราชภัฏบ้านสมเด็จเจ้าพระยา ด้วยเทคนิคการเรียนรู้ของเครื่อง พบว่าจำนวนกลุ่มที่เหมาะสมที่สุดในการจัดกลุ่มคือ 4 กลุ่ม เมื่อนำกลุ่มทั้ง 4 กลุ่มมาวิเคราะห์คุณลักษณะของกลุ่มเพื่อนำข้อมูลไปทำการวางแผนในการจัดฝึกอบรมเสริมความรู้ให้กับนักศึกษาสรุปได้ดังนี้

ตารางที่ 3 ข้อมูลจำนวนนักศึกษาที่จะต้องเข้ารับการอบรมในแต่ละทักษะ

องค์ความรู้	กลุ่ม	จำนวนคน
ด้านองค์การและระบบสารสนเทศ	0,1	28 (20+8)
ด้านเทคโนโลยีเพื่องานประยุกต์	-	-
ด้านเทคโนโลยีและวิธีการทางซอฟต์แวร์	2	13
ด้านโครงสร้างพื้นฐานระบบ	1,3	21 (13+8)
ด้านฮาร์ดแวร์และสถาปัตยกรรมคอมพิวเตอร์	0,2,3	46 (20+13+13)

กลุ่มที่จะต้องเข้ารับการอบรมเสริมความรู้ทางด้านองค์การและระบบสารสนเทศ ได้แก่ นักศึกษากลุ่มที่ 0,1 จำนวนทั้งหมด 28 คน ด้านเทคโนโลยีเพื่องานประยุกต์ไม่มีนักศึกษาที่จะต้องเข้ารับการอบรมเสริมความรู้ นักศึกษาที่ต้องเข้ารับการอบรมเสริมความรู้ด้านเทคโนโลยีและวิธีการทางซอฟต์แวร์ได้แก่กลุ่มที่ 2 จำนวน 13 คน นักศึกษาที่ต้องเข้ารับการอบรมเสริมความรู้ด้านโครงสร้างพื้นฐานระบบ ได้แก่กลุ่มที่ 1 และกลุ่มที่ 3 จำนวน 21 คน และนักศึกษาที่ต้องเข้ารับการอบรมเสริมความรู้ด้านฮาร์ดแวร์และสถาปัตยกรรมคอมพิวเตอร์ ได้แก่ นักศึกษากลุ่มที่ 0 กลุ่มที่ 2 และกลุ่มที่ 3 จำนวน 46 คน

ข้อเสนอแนะ

งานวิจัยในครั้งนี้ใช้ผลคะแนนการสอบวัดความรู้ทางคอมพิวเตอร์นักศึกษาชั้นปีที่ 4 ในการนำข้อมูลมาใช้เพื่อวางแผนในการจัดการอบรม อาจจะทำให้ไม่สามารถพัฒนาทักษะได้อย่างมีประสิทธิภาพเนื่องจากเป็นนักศึกษาชั้นปีสุดท้าย หากสามารถนำไปประยุกต์ใช้กับนักศึกษาชั้นปีที่ 1 เพื่อทำการวางแผนจัดการเรียนการสอนระยะยาวจะทำให้ได้ประโยชน์สูงสุด

เอกสารอ้างอิง

Fei Wang & Hector-Hugo & Franco-Penya & John D. Kelleher & John Pugh and Robert Ross (2017). An Analysis of the Application of Simplified Silhouette to the Evaluation of k-means Clustering Validity. 13th International Conference on Machine Learning and Data Mining MLDM Retrieved December, 2021, from https://www.researchgate.net/publication/318109824

Humaira, H. & Rasyidah R. (2020). Determining The Appropiate Cluster Number Using Elbow Method for K-Means Algorithm. 2nd Workshop on Multidisciplinary and Applications (WMA). Retrieved December, 2021, form https://www.researchgate.net/publication/339670247

Ian, T. J. & Jorge, C. (2016). Principal component analysis: a review and recent developments. The royal Society Publishing. Retrieved January, 2022, from https://royalsociety publishing.org/

Ketan, R. S. & Charles, N.(2020). Cluster Quality Analysis Using Silhouette Score. 2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA). Retrieved December, 2021, from https://mdsoar.org/handle/11603/20251

Rena, N. & Resianta, P. & Emma, S. & Feriani, A. T. (2018). Improved the Performance of the K-Means Cluster Using the Sum of Squared Error (SSE) optimized by using the Elbow Method. Journal of Physics : Conference Series. Retrieved January, 2022, from https://iopscience.iop.org/article/10.1088/1742-6596/1361/1/012015

Sidharth, P. M. & Uttam, S. & Subhash, T. & Sanjay, D. & Devi, P. S., Reshma, S. & Sasmita P. and Menalsh L. (2017). Multivariate Statistical Data Analysis- Principal Component Analysis (PCA). International Journal of Livestock Research. Retrieved January, 2022, from https://www.researchgate.net/publication/316652806