การพัฒนาแบบจำลองการแนะนำชมรมของนักศึกษา โดยประยุกต์เทคนิคเคมีนส์และต้นไม้ตัดสินใจ

THE DEVELOPMENT OF A STUDENT CLUB RECOMMENDATION MODEL USING K-MEANS CLUSTERING AND DECISION TREE ALGORITHMS

อุษณี ยี่สุ่นแก้ว[1]* และ สุดาสวรรค์ งามมงคลวงศ์[2]
Utsanee Yeesoonkaew [1] and Sudasawan Ngammongkolwong[2]

หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ
คณะเทคโนโลยีดิจิทัลและนวัตกรรม มหาวิทยาลัยเซาธ์อีสท์บางกอก[1,2]
Master of Science Program in Information Technology, Faculty of Digital Technology and Innovation,
Southeast Bangkok University[1,2]
Author email: utsanee@sbu.southeast.ac.th[1]*, sudasawan@southeast.ac.th[2]

Received: August 31, 2025
Revised: October 11, 2025
Accepted: October 21, 2025

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อ 1) ศึกษากระบวนการจัดกลุ่มนักศึกษาโดยใช้เทคนิค K-Means Clustering 2) พัฒนาโมเดล Decision Tree สำหรับการแนะนำชมรม และ 3) ประเมินประสิทธิภาพของกระบวนการจัดกลุ่มและโมเดลแนะนำชมรม กลุ่มตัวอย่าง ได้แก่ นักศึกษาระดับปริญญาตรี จำนวน 171 คน จากมหาวิทยาลัยเซาธ์อีสท์บางกอก โดยเลือกด้วยวิธีการสุ่มอย่างง่าย (Simple Random Sampling) เครื่องมือคือแบบสอบถามที่ผ่านการตรวจสอบความตรงเชิงเนื้อหา โดยผู้เชี่ยวชาญ 3 ท่าน สถิติที่ใช้ในการวิเคราะห์ประกอบด้วย การเข้ารหัสข้อมูลด้วย Label Encoding, การวิเคราะห์การจัดกลุ่มด้วย K-Means, และการประเมินจำนวนกลุ่มที่เหมาะสมด้วย Elbow Method, Davies–Bouldin Index และ Calinski–Harabasz Index ผลการวิจัยพบว่า 1) จำนวนกลุ่มที่เหมาะสมคือ 4 กลุ่ม 2) นักศึกษาแต่ละกลุ่มมีลักษณะความสนใจและทักษะที่แตกต่างกันอย่างชัดเจน และ 3) โมเดล Decision Tree มีค่า Accuracy เฉลี่ย 0.89 ซึ่งแสดงถึงความแม่นยำในระดับน่าพอใจ นักศึกษาแต่ละกลุ่มมีความสนใจและทักษะที่แตกต่างกันชัดเจน เทคนิคการเรียนรู้ของเครื่อง (ML) ช่วยเพิ่มความแม่นยำ ในการจัดกลุ่มนักศึกษาและการแนะนำชมรมให้เหมาะสมกับความสนใจ โดยสรุปการวิจัยครั้งนี้ยืนยันว่าเทคนิคการเรียนรู้ของเครื่องมีศักยภาพในการสนับสนุนการเลือกชมรมที่เหมาะสม และสามารถเพิ่มโอกาสการมีส่วนร่วมของนักศึกษาในกิจกรรมได้อย่างเป็นรูปธรรม

คำสำคัญ: การจัดกลุ่มนักศึกษา, เคมีนส์, ต้นไม้ตัดสินใจ, การแนะนำชมรม

Abstract

This research aimed to 1) study the student grouping process using the K-Means Clustering technique, 2) develop a Decision Tree model for club recommendation, and 3) evaluate the effectiveness of the clustering process and the club recommendation model. The sample consisted of 171 undergraduate students from Southeast Bangkok University, selected using a simple random sampling method. The instrument was a questionnaire that was content-validated by three experts. The statistical analysis employed Label Encoding, K-Means clustering analysis, and determination of the optimal number of clusters using the Elbow Method, Davies–Bouldin Index, and Calinski–Harabasz Index.

The results revealed that 1) the optimal number of clusters was four, 2) each group of students exhibited distinct interests and skills, and 3) the Decision Tree model achieved an average accuracy of 0.89, indicating high predictive accuracy. Machine learning (ML) techniques significantly improved the accuracy of grouping students and recommending clubs suitable for their interests. These findings suggest that machine learning techniques can effectively enhance personalized club recommendations and foster student engagement.

Keywords: Student clustering, K-Means, Decision Tree, Club recommendation

บทนำ

กิจกรรมนักศึกษาเป็นองค์ประกอบสำคัญในการพัฒนาผู้เรียนระดับอุดมศึกษา เนื่องจากช่วยเสริมสร้างทักษะชีวิต ภาวะผู้นำ ความรับผิดชอบต่อสังคม และการทำงานร่วมกับผู้อื่น ซึ่งล้วนเป็นสมรรถนะสำคัญที่จำเป็นต่อการดำรงชีวิตในศตวรรษที่ 21 [1] การเข้าร่วมชมรมถือเป็นกิจกรรมนักศึกษาที่มีบทบาทโดดเด่น เพราะเปิดโอกาสให้นักศึกษาได้ฝึกฝนความสามารถเฉพาะด้าน สร้างเครือข่ายทางสังคม และพัฒนาทักษะที่เอื้อต่อการเรียนรู้ตลอดชีวิต [2] อย่างไรก็ตาม ในทางปฏิบัติ นักศึกษาจำนวนไม่น้อยยังประสบปัญหาในการเลือกชมรมที่สอดคล้องกับความสนใจและศักยภาพของตน ส่งผลให้บางคนเลือกชมรมที่ไม่ตรงกับความต้องการ หรือไม่เข้าร่วมกิจกรรมชมรมใดเลย ซึ่งอาจทำให้พลาดโอกาสในการพัฒนาทักษะที่มีคุณค่า ดังนั้นจึงมีความจำเป็นต้องพัฒนาระบบสนับสนุนการตัดสินใจ โดยอาศัยข้อมูลจริงของนักศึกษาเพื่อช่วยแนะนำชมรมที่เหมาะสม [1], [2]

ในช่วงทศวรรษที่ผ่านมา เทคนิคการเรียนรู้ของเครื่อง (Machine Learning: ML) ได้รับความสนใจอย่างแพร่หลายในการประมวลผลและวิเคราะห์ข้อมูลเชิงซับซ้อน เนื่องจากสามารถจัดการข้อมูลจำนวนมากได้อย่างมีประสิทธิภาพ ยืดหยุ่น และให้ผลลัพธ์ที่นำไปใช้ได้จริง [3], [4] โดยเฉพาะเทคนิค K-Means Clustering ซึ่งเป็นวิธีการจัดกลุ่มข้อมูลแบบไม่กำกับ (Unsupervised Learning) ที่สามารถจำแนกข้อมูลออกเป็นกลุ่มย่อยตามความคล้ายคลึงของคุณลักษณะ และได้รับการประยุกต์ใช้อย่างกว้างขวางในหลายสาขา เช่น ธุรกิจ การแพทย์ และการศึกษา [5], [6] Zhao [7] ยังได้เสนอแนวทางปรับปรุงกระบวนการของ K-Means เพื่อเพิ่มความแม่นยำในการจัดกลุ่มข้อมูลเชิงซับซ้อน ซึ่งเป็นหลักฐานเชิงประจักษ์ที่แสดงถึงประสิทธิภาพของเทคนิคดังกล่าว

ในด้านการศึกษา มีงานวิจัยจำนวนมากที่ประยุกต์ใช้เทคนิค K-Means และ Decision Tree เพื่อสนับสนุนการวิเคราะห์และตัดสินใจทางการเรียนรู้ เช่น งานของ Jevintya et al. ที่ใช้สองเทคนิคนี้ในการวิเคราะห์ผลสัมฤทธิ์ทางการเรียนของนักศึกษาและได้ผลลัพธ์ที่ตีความและนำไปใช้ได้จริง [8] ขณะที่ Anshor และ Zy ใช้ Decision Tree

ในการพยากรณ์คุณภาพผลิตภัณฑ์และพบว่ามีความแม่นยำสูง [9] ผลลัพธ์เหล่านี้ชี้ให้เห็นว่า Decision Tree เป็นโมเดลที่เข้าใจง่าย สามารถสร้างกฎการตัดสินใจที่ตีความได้อย่างชัดเจน และมีศักยภาพในการประยุกต์ใช้ในบริบทจริงได้อย่างมีประสิทธิภาพ

อย่างไรก็ตาม แม้งานวิจัยก่อนหน้าจะใช้เทคนิค K-Means และ Decision Tree ในหลากหลายบริบท แต่ยังไม่พบงานใดที่นำสองเทคนิคนี้มาประยุกต์ใช้ร่วมกันเพื่อพัฒนา ระบบแนะนำการเลือกชมรมของนักศึกษาไทยโดยเฉพาะ ซึ่งถือเป็นช่องว่างขององค์ความรู้ (Research Gap) ที่ควรได้รับการศึกษา งานวิจัยนี้จึงมีความสำคัญในการพัฒนาระบบแนะนำชมรมที่ช่วยให้นักศึกษาเลือกชมรมได้ตรงกับความสนใจและศักยภาพของตน อีกทั้งยังสะท้อนแนวคิด การเรียนรู้เชิงข้อมูล (Data-Driven Learning Support) ที่มุ่งเน้นการใช้เทคโนโลยีเพื่อสนับสนุนการพัฒนาผู้เรียนเชิงคุณภาพ

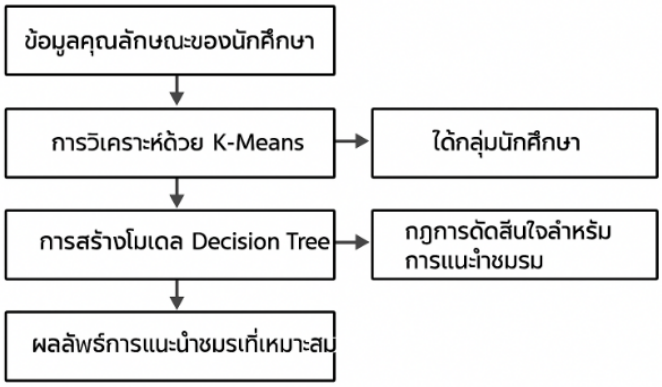
ดังนั้น งานวิจัยนี้จึงมีบทบาทสำคัญในการส่งเสริมการตัดสินใจของนักศึกษาให้มีระบบและสอดคล้องกับความสนใจ ช่วยเพิ่มโอกาสการมีส่วนร่วมในกิจกรรมนอกชั้นเรียน และเป็นแนวทางในการพัฒนาระบบแนะนำกิจกรรมนักศึกษาในสถาบันอุดมศึกษาในอนาคต การศึกษาครั้งนี้มุ่งเน้นการประยุกต์ใช้เทคนิค K-Means Clustering และ Decision Tree เพื่อจัดกลุ่มนักศึกษาและสร้างระบบแนะนำชมรมที่เหมาะสมกับแต่ละบุคคล ผลลัพธ์ที่ได้จะช่วยให้การแนะนำชมรมมีความเป็นระบบ แม่นยำ และสามารถใช้เป็นต้นแบบในการพัฒนาเครื่องมือสนับสนุนการบริหารกิจกรรมนักศึกษาในระดับมหาวิทยาลัยได้อย่างมีประสิทธิภาพ รวมทั้งเป็นกรณีศึกษาที่สะท้อนศักยภาพของเทคโนโลยีการเรียนรู้ของเครื่องในการส่งเสริมการตัดสินใจเชิงการศึกษาสำหรับผู้เรียนในอนาคต

วัตถุประสงค์การวิจัย

1. เพื่อศึกษากระบวนการจัดกลุ่มนักศึกษาและเลือกจำนวนกลุ่มที่เหมาะสมโดยใช้เทคนิค K-Means Clustering
2. เพื่อพัฒนาโมเดล Decision Tree สำหรับการแนะนำชมรมที่เหมาะสมแก่นักศึกษามหาวิทยาลัยเซาธ์อีสท์บางกอก
3. เพื่อประเมินประสิทธิภาพของกระบวนการจัดกลุ่มและโมเดลแนะนำชมรม

วิธีดำเนินการวิจัย

การวิจัยครั้งนี้ออกแบบกระบวนการโดยอิงตามแนวคิดของเทคนิคการเรียนรู้ของเครื่อง (Machine Learning) ซึ่งประกอบด้วยการจัดกลุ่มนักศึกษาด้วยเทคนิค K-Means และการสร้างโมเดล Decision Tree เพื่อแนะนำชมรมที่เหมาะสม โดยมีกรอบแนวคิดการวิจัย ดังแสดงในรูปที่ 1

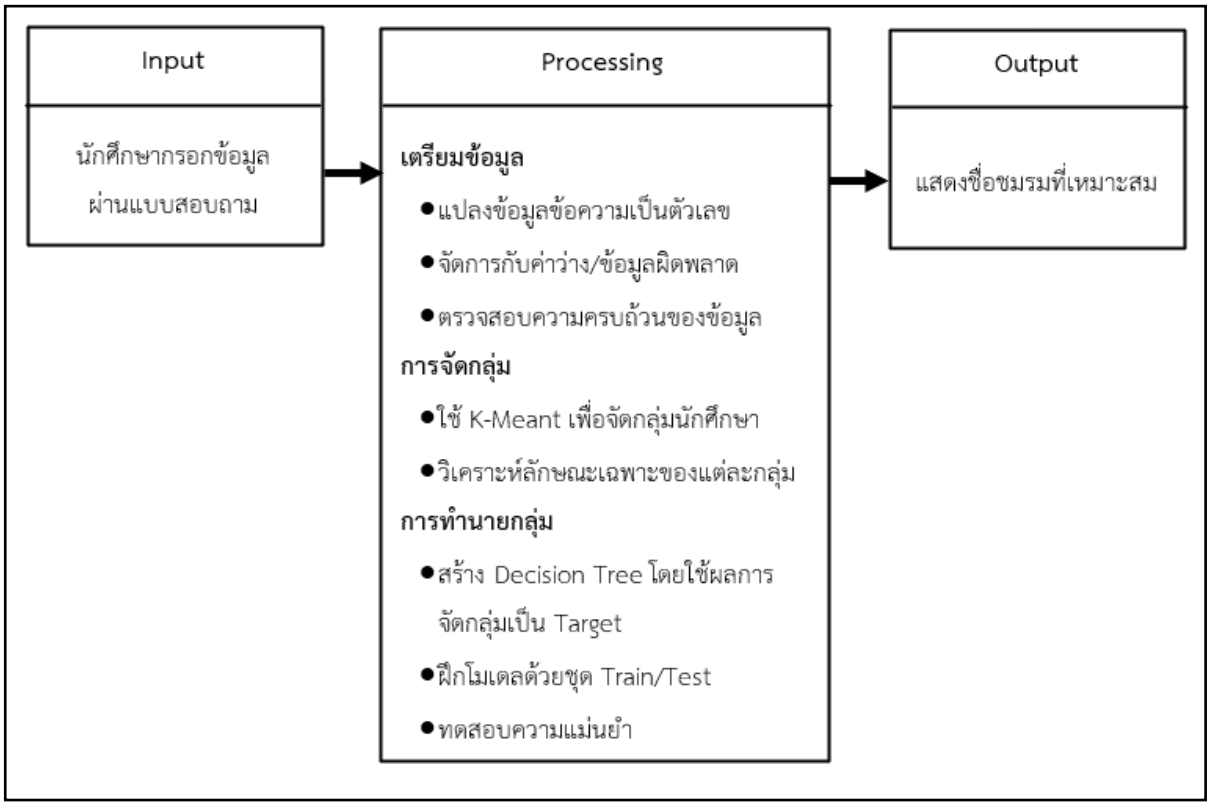


รูปที่ 1 กรอบแนวคิดในการวิจัย แสดงความสัมพันธ์ระหว่างข้อมูลนักศึกษา การวิเคราะห์ด้วย K-Means และการสร้างโมเดล Decision Tree เพื่อแนะนำชมรมที่เหมาะสม

จากรูปที่ 1 กรอบแนวคิดดังกล่าว ผู้วิจัยจึงกำหนดกระบวนการดำเนินการวิจัยออกเป็น 5 ส่วนสำคัญ ได้แก่ (1) ขั้นตอนการดำเนินการวิจัย (2) ประชากรและกลุ่มตัวอย่าง (3) เครื่องมือวิจัย (4) ขั้นตอนการเก็บข้อมูล และ (5) วิธีการวิเคราะห์ข้อมูล โดยมีรายละเอียดดังนี้

1. ขั้นตอนการดำเนินการวิจัย

เพื่อให้เห็นภาพรวมของกระบวนการดำเนินการวิจัยอย่างเป็นระบบ ผู้วิจัยได้จัดทำแผนผังแสดงลำดับขั้นตอนการดำเนินงาน ตั้งแต่การเก็บข้อมูล การเตรียมข้อมูล การจัดกลุ่มนักศึกษาด้วยเทคนิค K-Means การสร้างโมเดล Decision Tree เพื่อแนะนำชมรมที่เหมาะสม ตลอดจนการประเมินประสิทธิภาพของโมเดล ดังแสดงในรูปที่ 2



รูปที่ 2 ขั้นตอนการดำเนินการวิจัย

จากรูปที่ 2 แสดงให้เห็นกระบวนการดำเนินการวิจัยโดยเริ่มจากการเก็บรวบรวมข้อมูลจากแบบสอบถามของนักศึกษา จากนั้นดำเนินการเตรียมข้อมูลให้พร้อมสำหรับการวิเคราะห์ โดยแปลงข้อมูลเชิงข้อความให้เป็นข้อมูลเชิงตัวเลขและจัดการค่าที่ขาดหายหรือข้อมูลผิดปกติ เพื่อให้ข้อมูลมีความสมบูรณ์ก่อนนำไปใช้ในการวิเคราะห์ ขั้นต่อมาดำเนินการจัดกลุ่มนักศึกษาด้วยเทคนิค K-Means เพื่อจำแนกนักศึกษาตามลักษณะความสนใจ ทักษะ และปัจจัยที่เกี่ยวข้อง แล้วนำผลการจัดกลุ่มที่ได้มาใช้เป็นตัวแปรเป้าหมายในการสร้างโมเดล Decision Tree เพื่อทำนายและแนะนำชมรมที่เหมาะสมสำหรับนักศึกษา ก่อนทำการประเมินประสิทธิภาพของโมเดลและตรวจสอบความถูกต้องของเครื่องมือวิจัยในขั้นสุดท้าย

1.1 การวิเคราะห์ข้อมูลเบื้องต้น

ตรวจสอบความครบถ้วนของข้อมูลแบบสอบถามที่เก็บจากกลุ่มตัวอย่าง 171 คน หากพบข้อมูลที่หายไป (Missing Values) หรือข้อมูลผิดปกติ จะทำการปรับแก้ไขก่อนนำไปวิเคราะห์ จากการตรวจสอบพบว่ามีค่าข้อมูลสูญหายบางส่วน คิดเป็นไม่เกินร้อยละ 5 ของข้อมูลทั้งหมด ซึ่งอยู่ในระดับที่สามารถจัดการได้โดยไม่กระทบต่อผลการวิเคราะห์ ผู้วิจัยจึงดำเนินการแทนค่าข้อมูลที่ขาดหายด้วยวิธีแทนค่าด้วยค่าเฉลี่ย (Mean Imputation) สำหรับข้อมูลเชิงตัวเลข และแทนค่าด้วยค่าที่พบบ่อยที่สุด (Mode Imputation) สำหรับข้อมูลเชิงลักษณะ เพื่อรักษาความครบถ้วนของข้อมูลก่อนการเข้ารหัสข้อมูลเชิงข้อความให้อยู่ในรูปแบบตัวเลขด้วย Label Encoding เพื่อให้สามารถใช้ร่วมกับอัลกอริทึมการเรียนรู้ของเครื่องได้อย่างมีประสิทธิภาพ [10]

1.2 การจัดกลุ่มนักศึกษา (Clustering)

ใช้เทคนิค K-Means Clustering ในการจัดกลุ่มนักศึกษาตามความสนใจ ทักษะ และปัจจัยที่เกี่ยวข้อง การเลือกจำนวนกลุ่มที่เหมาะสมใช้ Elbow Method และตรวจสอบร่วมกับ Davies–Bouldin Index (DBI) และ Calinski–Harabasz Index (CHI) เพื่อยืนยันคุณภาพและความเหมาะสมของจำนวนกลุ่ม [5], [6] โดยใช้ค่าพารามิเตอร์ของโปรแกรมตามค่าเริ่มต้นของ scikit-learn คือ init='k-means++', n_init=10, และ max_iter=300 เพื่อให้ผลลัพธ์มีความเสถียรและสามารถทำซ้ำได้

1.3 การพัฒนาโมเดลแนะนำชมรม (Decision Tree)

นำผลการจัดกลุ่มที่ได้มาใช้เป็นตัวแปรเป้าหมาย (Target Variable) ในการสร้างโมเดล Decision Tree เพื่อกำหนดกฎการตัดสินใจ (if–then rules) สำหรับการแนะนำชมรม โดยแบ่งข้อมูลออกเป็นชุดฝึก (Training Set) และชุดทดสอบ (Testing Set) ในอัตรา 80:20 เพื่อใช้ในการประเมินสมรรถนะของโมเดล [11], [8] ทั้งนี้ในการสร้างโมเดล Decision Tree ใช้ผลการจัดกลุ่มจาก K-Means เป็นตัวแปรเป้าหมาย (Target Variable) โดยใช้เกณฑ์การแยกโหนดแบบ Gini Impurity (criterion='gini') ซึ่งเป็นวิธีการคำนวณความบริสุทธิ์ของกลุ่มข้อมูล เพื่อให้การแบ่งข้อมูลในแต่ละโหนดมีความชัดเจนและแม่นยำมากขึ้น นอกจากนี้ยังจำกัดความลึกของต้นไม้ (max_depth=5) เพื่อป้องกันการเกิด Overfitting และกำหนดให้แต่ละกิ่งมีข้อมูลอย่างน้อยสองรายการ (min_samples_leaf=2) เพื่อให้โมเดลสามารถเรียนรู้ได้อย่างมีประสิทธิภาพและคงความสอดคล้องของข้อมูลในแต่ละกลุ่ม เพื่อให้ผลลัพธ์ของการจำแนกมีความถูกต้องและสามารถนำไปใช้ในการแนะนำชมรมที่เหมาะสมได้จริง

1.4 การประเมินสมรรถนะของโมเดล

โมเดล Decision Tree ที่พัฒนาขึ้นถูกประเมินด้วยตัวชี้วัดหลัก ได้แก่ Accuracy, Precision, Recall และ F1-Score รวมถึงการใช้ Confusion Matrix เพื่อตรวจสอบข้อผิดพลาดในการจำแนกและระบุจุดแข็ง–จุดอ่อนของโมเดล [12]

1.5 การตรวจสอบคุณภาพเครื่องมือวิจัย

แบบสอบถามได้รับการตรวจสอบความตรงเชิงเนื้อหา (IOC) โดยผู้เชี่ยวชาญ 3 ท่าน พบค่าเฉลี่ย IOC = 1.00 ซึ่งเกณฑ์ยอมรับทั่วไปคือ ≥ 0.50 แสดงว่าเครื่องมือมีความน่าเชื่อถือเพียงพอสำหรับการวิจัย [13]

2. ประชากรและกลุ่มตัวอย่าง

ประชากร ได้แก่ นักศึกษาระดับปริญญาตรีของมหาวิทยาลัยเซาธ์อีสท์บางกอก จำนวน 310 คน กลุ่มตัวอย่างจำนวน 171 คน ได้มาโดยวิธีการสุ่มอย่างง่าย (Simple Random Sampling) อ้างอิงตามเกณฑ์ของ Krejcie และ Morgan เพื่อให้ได้กลุ่มตัวอย่างที่มีความเป็นตัวแทนของประชากรอย่างเหมาะสม [14]

3. เครื่องมือวิจัย

3.1 แบบสอบถาม

แบบสอบถามแบ่งออกเป็น 5 ส่วน ได้แก่ (1) ข้อมูลพื้นฐาน (2) ความสนใจ (3) ทักษะ (4) ปัจจัยที่มีผลต่อการเลือกชมรม และ (5) ความสนใจในการเข้าร่วมชมรมพร้อมข้อเสนอแนะเพิ่มเติม โดยตรวจสอบความตรงเชิงเนื้อหา (IOC) จากผู้เชี่ยวชาญ 3 ท่าน ผลการประเมินพบว่า ค่าเฉลี่ย IOC = 1.00 ซึ่งเกณฑ์ยอมรับทั่วไปคือ ≥ 0.50 แสดงว่าเครื่องมือมีความน่าเชื่อถือเพียงพอสำหรับการวิจัย [13]

3.2 เครื่องมือสำหรับการวิเคราะห์ข้อมูล

ใช้ Google Colab เป็นแพลตฟอร์มประมวลผล โดยใช้ภาษา Python และไลบรารีมาตรฐาน ได้แก่

1) pandas สำหรับการจัดการข้อมูลและ Data Cleaning [15]
2) scikit-learn สำหรับการสร้างโมเดล K-Means และ Decision Tree [16]
3) matplotlib และ seaborn สำหรับการสร้างกราฟและ Visualization [17]

การเลือกใช้ Google Colab ช่วยให้การทำงานบนระบบ Cloud เป็นไปอย่างสะดวก รองรับการทำซ้ำ (Reproducibility) และสามารถจัดการข้อมูลขนาดกลางได้อย่างมีประสิทธิภาพ [10], [18]

4. ขั้นตอนการเก็บข้อมูล

การเก็บข้อมูลเป็นกระบวนการสำคัญที่ส่งผลต่อคุณภาพและความน่าเชื่อถือของงานวิจัย ในการศึกษาครั้งนี้ได้ดำเนินการตามลำดับขั้นตอนอย่างเป็นระบบ เพื่อให้ได้ข้อมูลที่ครบถ้วน ถูกต้อง และสอดคล้องกับวัตถุประสงค์การวิจัย ดังนี้

4.1 การสร้าง ตรวจสอบ และปรับปรุงแบบสอบถาม

ผู้วิจัยออกแบบแบบสอบถามให้ครอบคลุมประเด็นที่เกี่ยวข้องกับข้อมูลพื้นฐาน ความสนใจ ทักษะ และปัจจัยที่มีผลต่อการเลือกชมรม โดยยึดตามกรอบแนวคิดของงานวิจัย จากนั้นส่งแบบสอบถามให้ผู้เชี่ยวชาญจำนวน 3 ท่าน ตรวจสอบ ความตรงเชิงเนื้อหา (Index of Item-Objective Congruence: IOC) [13] เพื่อประเมินว่ารายการคำถามมีความสอดคล้องกับวัตถุประสงค์การวิจัยหรือไม่ ค่า IOC ที่ได้อยู่ระหว่าง 0.67–1.00 ซึ่งถือว่าอยู่ในเกณฑ์ที่ยอมรับได้ แสดงว่าแบบสอบถามมีความเหมาะสมสำหรับการเก็บข้อมูลจริง หลังจากได้รับข้อเสนอแนะ ผู้วิจัยได้ปรับปรุงถ้อยคำ การจัดเรียงคำถาม และรูปแบบการนำเสนอ เพื่อให้มีความชัดเจน เข้าใจง่าย และลดความกำกวม อันจะช่วยเพิ่มความถูกต้องและความน่าเชื่อถือของคำตอบจากผู้ตอบแบบสอบถาม

4.2 การแจกแบบสอบถามแก่กลุ่มตัวอย่าง

แบบสอบถามที่ปรับปรุงแล้วถูกนำไปแจกให้แก่นักศึกษา 171 คน ซึ่งเป็นกลุ่มตัวอย่างที่สุ่มจากประชากรทั้งหมด 310 คน โดยใช้วิธีการสุ่มอย่างง่าย (Simple Random Sampling) ตามเกณฑ์ของ Krejcie และ Morgan [14] เพื่อให้กลุ่มตัวอย่างมีความเป็นตัวแทนของประชากรอย่างเหมาะสม การแจกแบบสอบถามดำเนินการภายใต้การอธิบายวัตถุประสงค์และการขอความยินยอมจากผู้เข้าร่วม (Informed Consent)

4.3 การรวบรวมและตรวจสอบข้อมูล

เมื่อสิ้นสุดการตอบแบบสอบถาม ผู้วิจัยได้รวบรวมข้อมูลทั้งหมด และตรวจสอบความสมบูรณ์ของแบบสอบถามในแต่ละชุด เพื่อตรวจหาข้อมูลที่ขาดหาย (Missing Values) หรือข้อมูลที่ไม่สมเหตุสมผล หากพบปัญหาจะมีการแก้ไขหรือกำจัดข้อมูลที่ไม่สมบูรณ์ออก โดยดำเนินการจัดการข้อมูลที่ขาดหายด้วยการแทนค่าด้วยค่าเฉลี่ย (Mean Imputation) สำหรับข้อมูลเชิงตัวเลข และการแทนค่าด้วยค่าที่พบมากที่สุด (Mode Imputation) สำหรับข้อมูลเชิงลักษณะ เพื่อให้ข้อมูลมีความครบถ้วนและสอดคล้องกับรูปแบบที่ใช้ในการวิเคราะห์ จากนั้นจึงบันทึกข้อมูลเข้าสู่ระบบประมวลผลสำหรับการวิเคราะห์ต่อไป โดยใช้โปรแกรม Google Colab และภาษา Python ร่วมกับไลบรารีมาตรฐาน [10], [15]

5. การวิเคราะห์ข้อมูล (Data Analysis)

การวิเคราะห์ข้อมูลในการวิจัยครั้งนี้ประกอบด้วย 3 ขั้นตอนสำคัญ ได้แก่ (1) การใช้สถิติเชิงพรรณนา (Descriptive Statistics) (2) การวิเคราะห์การจัดกลุ่มนักศึกษา (Clustering Analysis) และ (3) การสร้างและประเมินโมเดลแนะนำชมรม (Decision Tree Model) โดยมีรายละเอียดดังนี้

5.1 สถิติเชิงพรรณนา (Descriptive Statistics)

ใช้ในการอธิบายลักษณะทั่วไปของกลุ่มตัวอย่าง โดยคำนวณค่าทางสถิติพื้นฐาน เช่น ค่าเฉลี่ย (Mean) ค่าส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation) ค่าร้อยละ (Percentage) และค่าความถี่ (Frequency) เพื่อแสดงแนวโน้มและการกระจายของข้อมูลเบื้องต้น อันเป็นพื้นฐานสำหรับการนำไปสู่การวิเคราะห์เชิงลึกในขั้นตอนถัดไป [19]

5.2 การวิเคราะห์การจัดกลุ่มนักศึกษา (Clustering Analysis)

การจัดกลุ่มนักศึกษาใช้เทคนิค K-Means Clustering [5], [6] เพื่อแบ่งกลุ่มตามคุณลักษณะ ความสนใจ และทักษะ โดยพิจารณาจำนวนกลุ่มที่เหมาะสมจากหลายวิธีการประเมิน

ทั้งนี้ การเลือกใช้เทคนิค K-Means Clustering เนื่องจากเป็นวิธีที่สามารถจัดกลุ่มข้อมูลเชิงตัวเลขได้อย่างมีประสิทธิภาพและเข้าใจง่าย เหมาะกับข้อมูลนักศึกษาที่มีหลายมิติ ทั้งด้านความสนใจและทักษะ อีกทั้ง K-Means ยังประมวลผลได้รวดเร็วกว่าเทคนิคอื่น เช่น Hierarchical Clustering หรือ DBSCAN ซึ่งมีความซับซ้อนในการคำนวณมากกว่า

ดังนั้น K-Means จึงเป็นเทคนิคที่เหมาะสมกับลักษณะข้อมูลและวัตถุประสงค์ของงานวิจัยนี้ โดยพิจารณาจำนวนกลุ่มที่เหมาะสมจากหลายวิธีการประเมิน ได้แก่

5.2.1 Elbow Method: ใช้ค่า Sum of Squared Errors (SSE) เพื่อหาจุดที่ค่า SSE ลดลงช้าลง (Elbow Point) ซึ่งบ่งบอกจำนวนกลุ่มที่เหมาะสมที่สุด

5.2.2 Davies–Bouldin Index (DBI): ใช้ประเมินความกระชับภายในกลุ่มและความแตกต่างระหว่างกลุ่ม โดยค่าที่ต่ำกว่าหมายถึงการจัดกลุ่มที่มีคุณภาพดีกว่า

5.2.3 Calinski–Harabasz Index (CHI): ใช้เปรียบเทียบอัตราส่วนระหว่างความแปรปรวนระหว่างกลุ่มและภายในกลุ่ม โดยค่าที่สูงกว่าหมายถึงการจัดกลุ่มที่มีประสิทธิภาพมากกว่า

5.2.4 Silhouette Score (SS): ใช้ในการประเมินความเหมาะสมของจำนวนกลุ่ม (K) โดยค่าของ Silhouette Score อยู่ระหว่าง -1 ถึง 1 ยิ่งมีค่าเฉลี่ยใกล้ 1 มากเท่าใด แสดงว่าการจัดกลุ่มมีความเหมาะสมที่สุด

วิธีการทั้งสี่ถูกนำมาใช้ร่วมกัน เพื่อยืนยันความเหมาะสมและความน่าเชื่อถือของจำนวนกลุ่มที่เลือก

5.3 การสร้างและประเมินโมเดลแนะนำชมรม (Decision Tree Model)

ผลการจัดกลุ่มที่ได้จาก K-Means ถูกนำมาใช้เป็นตัวแปรเป้าหมาย (Target Variable) ในการสร้างโมเดล Decision Tree ซึ่งเป็นโมเดลที่สามารถสร้างกฎการตัดสินใจ (Decision Rules) ในรูปแบบเงื่อนไข if–then ที่เข้าใจง่ายและสามารถนำไปใช้งานจริงได้

ทั้งนี้ การเลือกใช้เทคนิค Decision Tree เนื่องจากเป็นวิธีการเรียนรู้ของเครื่อง (Machine Learning) ที่สามารถอธิบายความสัมพันธ์ระหว่างตัวแปรได้อย่างชัดเจน ให้ผลลัพธ์ในรูปแบบกฎการตัดสินใจที่เข้าใจง่ายและสามารถตีความได้โดยตรง เหมาะกับลักษณะข้อมูลเชิงพฤติกรรมของนักศึกษา ซึ่งต้องการการวิเคราะห์เชิงเหตุผลเพื่อแนะนำชมรมที่เหมาะสม นอกจากนี้ Decision Tree ยังมีความยืดหยุ่นในการประมวลผลข้อมูลหลายมิติ และไม่จำเป็นต้องอาศัยสมมติฐานทางสถิติซับซ้อนเหมือนเทคนิคอื่น เช่น Logistic Regression หรือ Neural Network ที่มีความซับซ้อนสูงและยากต่อการตีความ

การประเมินสมรรถนะของโมเดลทำโดยใช้ตัวชี้วัดหลัก ได้แก่

5.3.1 Accuracy: ร้อยละของการทำนายที่ถูกต้องทั้งหมดเมื่อเปรียบเทียบกับจำนวนตัวอย่างทั้งหมด

5.3.2 Precision: ระดับความแม่นยำในการจำแนกตัวอย่างเชิงบวกว่ามีความถูกต้องเพียงใด

5.3.3 Recall: ความสามารถของโมเดลในการตรวจจับตัวอย่างเชิงบวกทั้งหมด

5.3.4 F1-Score: ค่าเฉลี่ยเชิงฮาร์โมนิกระหว่าง Precision และ Recall เพื่อสะท้อนความสมดุลของโมเดล

5.3.5 Confusion Matrix: ตารางที่ใช้วิเคราะห์ข้อผิดพลาดในการจำแนก โดยแสดงจำนวนการทำนายที่ถูกต้องและผิดพลาดในแต่ละกลุ่มอย่างละเอียด [8], [11], [12]

5.4 การตรวจสอบความทนทานของโมเดล (Robustness & Cross-Validation)

เพื่อให้ผลประเมินมีความน่าเชื่อถือ ผู้วิจัยใช้ Stratified k-fold cross-validation (k=5) โดยคงสัดส่วนของกลุ่ม (คลัสเตอร์) ให้ใกล้เคียงกันในแต่ละพับและทำซ้ำ 3 รอบ (repeated CV) เพื่อเฉลี่ยความผันผวนจากการสุ่มแบ่งข้อมูล รายงานผลเป็นค่า เฉลี่ย ± ส่วนเบี่ยงเบนมาตรฐาน ของ Accuracy, Precision, Recall และ F1-Score นอกจากนี้เปรียบเทียบกับ baseline (เช่น ตัวจำแนกแบบทำนายกลุ่มที่พบบ่อยที่สุด) [20] เพื่อยืนยันว่าโมเดลมีประสิทธิภาพสูงกว่าค่าพื้นฐานอย่างมีนัยสำคัญ

5.5 การปรับแต่งไฮเปอร์พารามิเตอร์ (Hyperparameter Tuning)

K-Means: ใช้ค่า K=4 (จากการประเมิน Elbow, DBI, CHI และพิจารณาบริบทข้อมูล) พร้อมตั้งค่า n_init (เช่น 10–20) และ max_iter (เช่น 300) เพื่อความเสถียรของผลลัพธ์ ทั้งนี้ทำ Feature Scaling (มาตรฐานค่า Z-score) ให้กับตัวแปรเชิงปริมาณก่อนจัดกลุ่ม เพื่อลดอิทธิพลจากสเกลที่ต่างกัน [16] Decision Tree: ปรับแต่งพารามิเตอร์ด้วย Grid/Random Search เช่น criterion $\in$ {gini, entropy}, max_depth $\in$ {3–15}, min_samples_split $\in$ {2, 5, 10}, min_samples_leaf $\in$ {1, 3, 5} โดยเลือกชุดค่าที่ให้ผลเฉลี่ย F1-Score สูงสุดบนชุดตรวจสอบ (validation) ภายใต้ CV เดียวกัน [21]

5.6 การตีความและอธิบายโมเดล (Model Interpretability)

สกัด กฎการตัดสินใจ (if–then rules) จากโมเดลแนะนำตัดสินใจ พร้อมรายงาน คุณลักษณะที่สำคัญ (feature importance) เพื่ออธิบายเหตุผลของคำแนะนำชมรมในเชิงปฏิบัติ [22] เช่น

หากนักศึกษามีความสนใจด้านกีฬา ≥ เกณฑ์ A และทักษะการทำงานเป็นทีม ≥ เกณฑ์ B $\Rightarrow$ แนะนำกลุ่มชมรมกีฬา/จิตอาสา โดยเสนอตัวอย่างเส้นทางการตัดสินใจ (decision paths) 2–3 เส้นทางที่พบได้บ่อย เพื่อช่วยให้ที่ปรึกษาและผู้เรียนเข้าใจและตรวจสอบได้

5.7 ความสามารถในการทำซ้ำ (Reproducibility)

ระบุ สภาพแวดล้อมซอฟต์แวร์ ที่ใช้ (เช่น Python 3.x, scikit-learn, pandas, matplotlib, seaborn เวอร์ชัน), random seed ที่ตั้งไว้, และ สมุดงาน (notebook) / สคริปต์ ที่ใช้วิเคราะห์ โดยจัดเก็บไว้ในโครงสร้างโฟลเดอร์ที่ชัดเจน (เช่น data/, notebooks/, src/, results/) เพื่อให้การทำซ้ำผลการทดลองและการตรวจสอบภายนอกทำได้สะดวก [18]

5.8 การพิจารณาด้านจริยธรรมและการคุ้มครองข้อมูล (Ethics & Data Protection)

ดำเนินการตามหลักจริยธรรมการวิจัยในมนุษย์ [23] โดย

1) ขอความยินยอมโดยรู้ข้อมูล (Informed Consent) จากผู้เข้าร่วมทุกคน
2) เก็บรักษาข้อมูลในรูป ไม่ระบุตัวตน (de-identification) และเข้ารหัสไฟล์ข้อมูล
3) จำกัดการเข้าถึงข้อมูลเฉพาะผู้วิจัย และใช้เพื่อวัตถุประสงค์ทางวิชาการเท่านั้น
4) ปฏิบัติตามระเบียบ/กฎหมายที่เกี่ยวข้องกับการคุ้มครองข้อมูลส่วนบุคคล

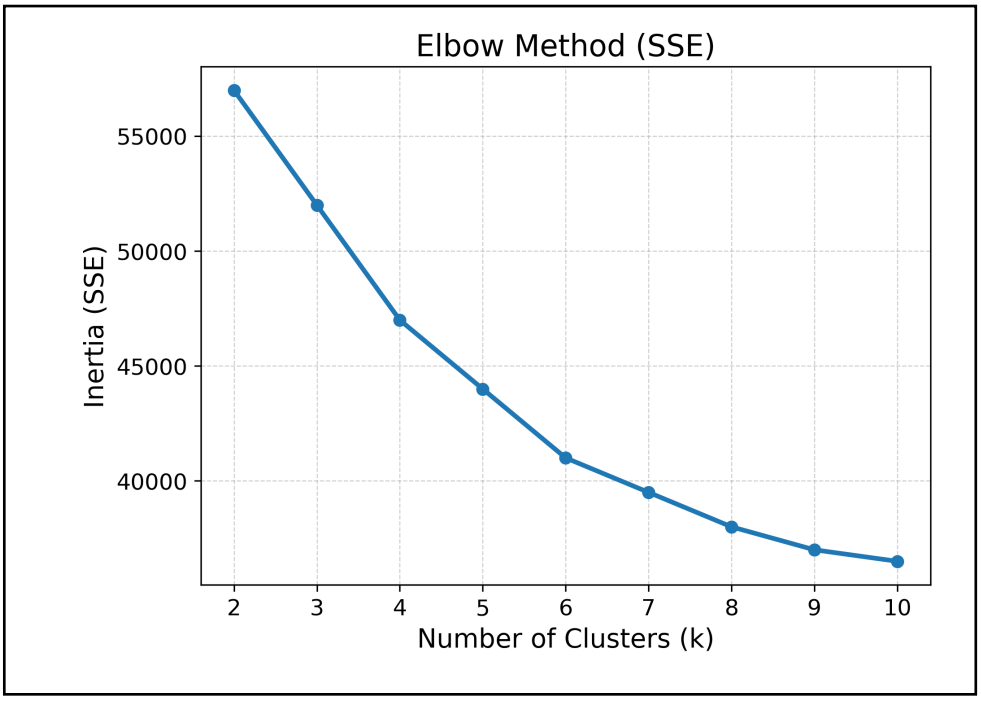
5.9 การพรรณนาลักษณะกลุ่มหลังการจัดคลัสเตอร์ (Post-Clustering Profiling)

ภายหลังการจัดกลุ่มและจำแนกด้วย Decision Tree วิเคราะห์ ลักษณะเด่นของแต่ละกลุ่ม ด้วยสถิติเชิงอธิบาย และทดสอบความแตกต่างระหว่างกลุ่มด้วย ANOVA/Chi-square (ตามชนิดตัวแปร) เพื่อยืนยันว่าแต่ละคลัสเตอร์มีเอกลักษณ์ด้านความสนใจ/ทักษะที่แตกต่างกันอย่างมีนัยสำคัญเชิงสถิติ [24]

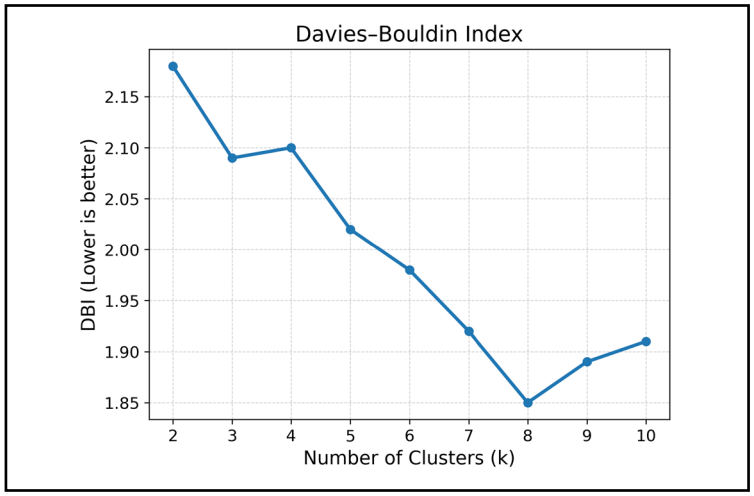
ผลการวิจัย

1. ผลการศึกษากระบวนการจัดกลุ่มนักศึกษาโดยใช้เทคนิค K-Means Clustering

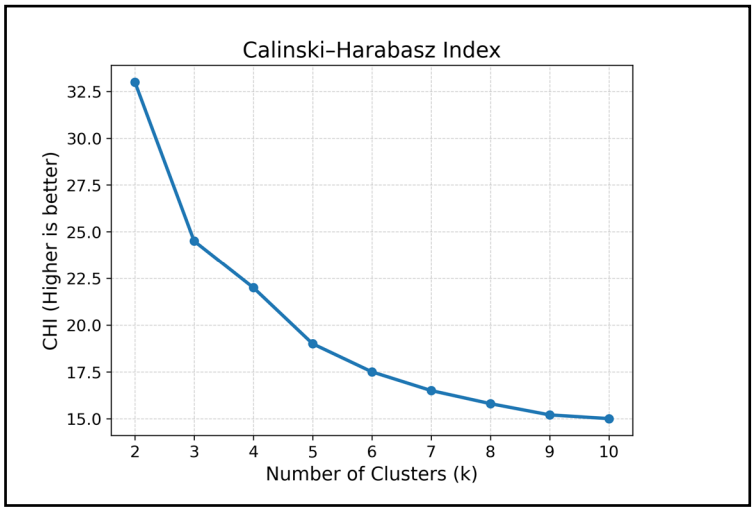
ในการวิจัยครั้งนี้ ผู้วิจัยใช้หลายวิธีในการหาค่าจำนวนกลุ่ม (K) ที่เหมาะสมสำหรับการทำ K-Means Clustering ได้แก่ Elbow Method, Davies–Bouldin Index (DBI), Calinski–Harabasz Index (CHI), และ Silhouette Score เพื่อเปรียบเทียบผลลัพธ์และพิจารณาค่าที่เหมาะสมที่สุด โดยอ้างอิงแนวทางของ Han et al. [6] แนะนำว่าการเลือกจำนวนกลุ่มที่เหมาะสมควรพิจารณาผลลัพธ์จากหลายวิธีประกอบกัน ไม่ควรอ้างอิงจากวิธีใดวิธีหนึ่งเพียงอย่างเดียว ดังแสดงในรูปที่ 3-6



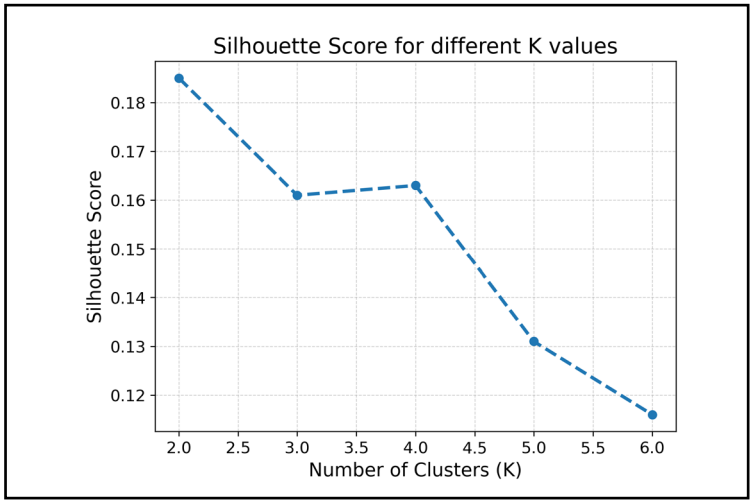
รูปที่ 3 ผลการวิเคราะห์ด้วย Elbow Method (SSE)



รูปที่ 4 ผลการวิเคราะห์ด้วย Davies–Bouldin Index (DBI)



รูปที่ 5 ผลการวิเคราะห์ด้วย Calinski-Harabasz Index (CHI)



รูปที่ 6 ผลการวิเคราะห์ด้วย Silhouette Score

ผลการวิเคราะห์พบว่า Elbow Method ชี้ไปที่ K=4, ขณะที่ Davies–Bouldin Index (DBI) ชี้ไปที่ K=8, ส่วน Calinski–Harabasz Index (CHI) และ Silhouette Score ชี้ไปที่ K=2 อย่างไรก็ตาม เมื่อพิจารณาร่วมกันทั้งเชิงสถิติและบริบทของข้อมูลนักศึกษา การเลือก K=4 ถือว่าเหมาะสมที่สุด เนื่องจากให้ค่าประสิทธิภาพที่สมดุลที่สุดระหว่างความกระชับภายในกลุ่มและความแตกต่างระหว่างกลุ่ม เมื่อเปรียบเทียบกับการจัดกลุ่มแบบ K=2 พบว่ามีการรวมกลุ่มนักศึกษาที่มีลักษณะต่างกันไว้ด้วยกันมากเกินไป ทำให้ไม่สามารถแยกคุณลักษณะเฉพาะของนักศึกษาแต่ละกลุ่มได้อย่างชัดเจน ขณะที่ K=8 มีจำนวนกลุ่มย่อยมากเกินไปจนเกิดความซ้ำซ้อนและยากต่อการตีความเชิงปฏิบัติ ดังนั้น การเลือกใช้ K=4 จึงเหมาะสมที่สุด ทั้งในเชิงสถิติและในเชิงการนำไปใช้จริง เช่น การวางแผนพัฒนากิจกรรมนักศึกษาที่ตอบโจทย์ความสนใจและศักยภาพของแต่ละกลุ่มได้อย่างชัดเจน

เมื่อกำหนดจำนวนกลุ่มที่เหมาะสมเป็น K=4 แล้ว นักศึกษาถูกแบ่งออกเป็น 4 กลุ่ม โดยแต่ละกลุ่มมีลักษณะเฉพาะดังนี้

กลุ่มที่ 1 นักศึกษาที่สนใจด้านกีฬาและจิตอาสา

กลุ่มที่ 2 นักศึกษาที่สนใจด้านวิชาการและศิลปะ

กลุ่มที่ 3 นักศึกษาที่สนใจด้านเทคโนโลยี

กลุ่มที่ 4 นักศึกษาที่สนใจด้านกีฬา ดนตรี และวิชาการ

รายละเอียดจำนวนและร้อยละของนักศึกษาแต่ละกลุ่ม ดังแสดงในตารางที่ 1

ตารางที่ 1 จำนวนและร้อยละของนักศึกษาในแต่ละกลุ่ม

กลุ่ม	จำนวนสมาชิก	ร้อยละ (%)
1	41	23.98
2	61	35.67
3	23	13.45
4	46	26.90
รวม	**171**	**100.00**

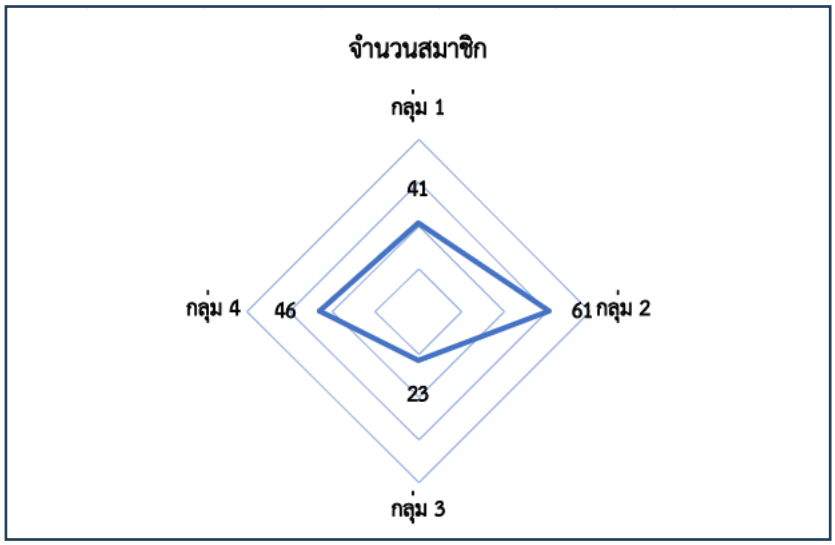
จากตารางที่ 1 สามารถวิเคราะห์คุณลักษณะของนักศึกษาแต่ละกลุ่มได้ดังนี้

1) กลุ่มที่ 1 (41 คน, 23.98%) ส่วนใหญ่เป็นเพศหญิง จากคณะเทคโนโลยีดิจิทัลและนวัตกรรม สาขาเทคโนโลยีสารสนเทศและมัลติมีเดียและอีสปอร์ต สนใจกิจกรรมด้านกีฬาและจิตอาสา ทักษะเด่นคือการสื่อสารและการทำงานเป็นทีม ซึ่งสอดคล้องกับงานของ Arslan et al. [25]

2) กลุ่มที่ 2 (61 คน, 35.67%) มีทั้งชายและหญิง โดยเพศหญิงมากกว่าเล็กน้อย ส่วนมากอยู่ในคณะบัญชีและวิทยาการจัดการ สนใจกิจกรรมด้านวิชาการและศิลปะ ทักษะเด่นคือการคิดเชิงวิเคราะห์ สอดคล้องกับ Do and Le [26]

3) กลุ่มที่ 3 (23 คน, 13.45%) ส่วนใหญ่เป็นเพศชาย จากคณะเทคโนโลยีดิจิทัลและนวัตกรรม สาขาเทคโนโลยีวิศวกรรมอุตสาหการ และเทคโนโลยีดิจิทัล สนใจกิจกรรมด้านเทคโนโลยี ทักษะเด่นคือการคิดเชิงวิเคราะห์และการใช้เทคโนโลยี ซึ่งสอดคล้องกับ Zhao [7]

4) กลุ่มที่ 4 (46 คน, 26.90%) ส่วนใหญ่เป็นเพศหญิง จากคณะโลจิสติกส์และการบิน สนใจกิจกรรมผสม เช่น กีฬา ดนตรี และวิชาการ ทักษะเด่นคือการวางแผนและการจัดการ สอดคล้องกับ Guo et al. [11]

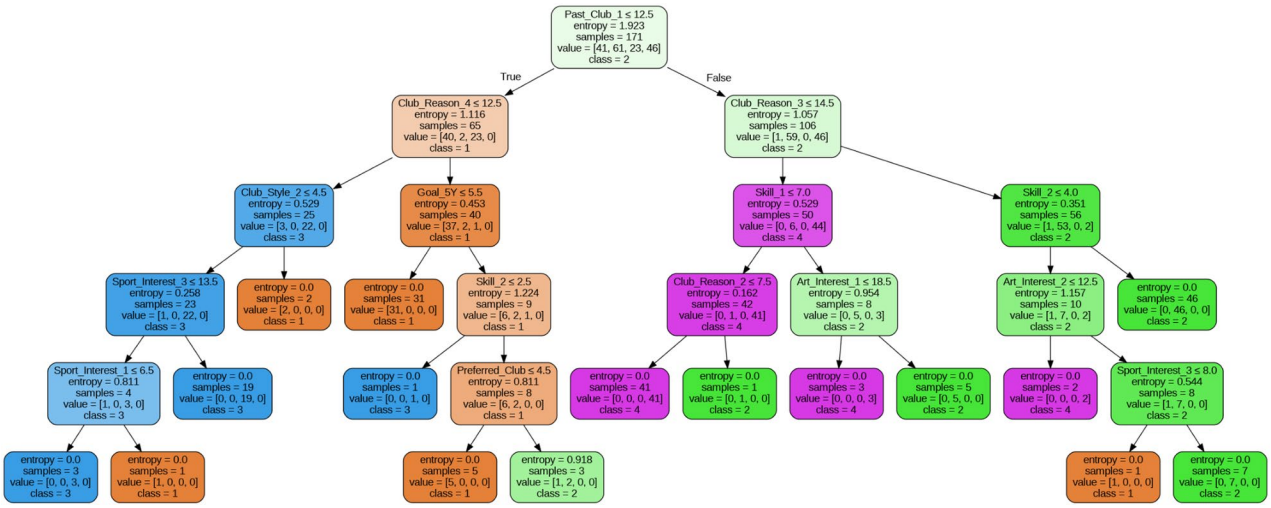


รูปที่ 7 แสดงจำนวนสมาชิกในแต่ละกลุ่ม

จากการวิเคราะห์เชิงเปรียบเทียบจากรูปที่ 7 พบว่า นักศึกษาทั้ง 4 กลุ่มมีจำนวนและลักษณะเด่นที่แตกต่างกันอย่างชัดเจน โดยกลุ่มที่ 2 มีจำนวนนักศึกษามากที่สุด (61 คน) ขณะที่กลุ่มที่ 3 มีจำนวนน้อยที่สุด (23 คน) แสดงให้เห็นถึงความหลากหลายของทักษะและความสนใจของนักศึกษาในมหาวิทยาลัย ทั้งด้านกีฬา จิตอาสา วิชาการ ศิลปะ เทคโนโลยี และการจัดการ ผลดังกล่าวสะท้อนให้เห็นว่านักศึกษาแต่ละกลุ่มมีจุดแข็งเฉพาะที่สามารถนำไปใช้เป็นแนวทางในการแนะนำชมรมและกิจกรรมที่เหมาะสมต่อไปได้

2. ผลการพัฒนาโมเดล Decision Tree สำหรับการแนะนำชมรมที่เหมาะสมแก่นักศึกษา

ผู้วิจัยได้นำผลการจัดกลุ่มจาก K-Means มาใช้เป็นตัวแปรเป้าหมายในการสร้างโมเดล Decision Tree ผลการวิเคราะห์พบว่า โมเดลสามารถจำแนกนักศึกษาได้อย่างมีประสิทธิภาพ และอธิบายปัจจัยที่มีอิทธิพลต่อการตัดสินใจได้ชัดเจน โดยรูปที่ 8 แสดงโครงสร้างของโมเดล ซึ่งปัจจัยหลักที่มีอิทธิพลมากที่สุดคือ Past_Club_1 ซึ่งในชุดข้อมูลนี้หมายถึง “ชมรมด้านความสนใจและทักษะเฉพาะทาง” ที่ได้จากการเข้ารหัสคำตอบของคำถาม “ชมรมที่เคยเข้าร่วม” ก่อนแตกแขนงไปยังปัจจัยอื่น ๆ เช่น เหตุผลในการเลือกชมรม และ เป้าหมายในอนาคตของนักศึกษา ลักษณะการแสดงผลดังกล่าวทำให้เข้าใจกระบวนการตัดสินใจได้ชัดเจน และสามารถนำไปประยุกต์ใช้จริงในการแนะนำชมรมที่เหมาะสมแก่นักศึกษาได้อย่างเหมาะสม



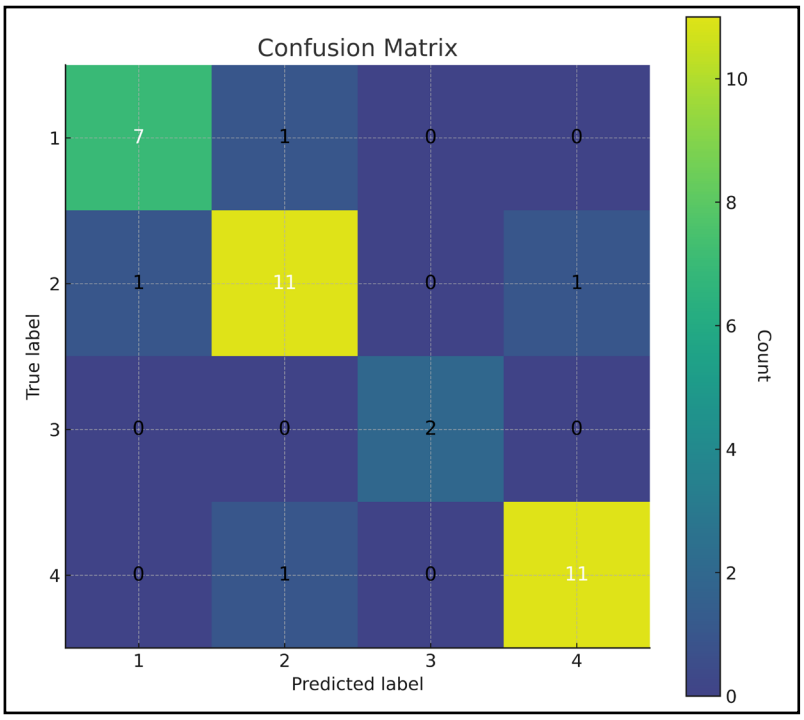
รูปที่ 8 โครงสร้างของโมเดล Decision Tree

3. ผลการประเมินประสิทธิภาพของโมเดล Decision Tree

ผลการประเมินประสิทธิภาพของโมเดล Decision Tree ดังแสดงในตารางที่ 2 และรูปที่ 9 โดยโมเดลมีค่า Accuracy เฉลี่ยเท่ากับ 0.89 ซึ่งอยู่ในระดับที่น่าพอใจ เมื่อพิจารณาเป็นรายกลุ่ม พบว่า Cluster 3 มีค่าประสิทธิภาพสูงสุดทุกตัวชี้วัด (1.00) สะท้อนถึงความชัดเจนของข้อมูล ขณะที่ Cluster 2 มีค่าต่ำที่สุด (0.85) สะท้อนถึงความซับซ้อนของข้อมูลในกลุ่มนั้น ส่วน Cluster 1 (0.88) และ Cluster 4 (0.92) อยู่ในระดับค่อนข้างสูงโดยรวม ผลการวิเคราะห์นี้สอดคล้องกับงานวิจัยของ David and Balakrishnan [27] และ Guo et al. [11] ที่ยืนยันว่า Decision Tree สามารถจำแนกข้อมูลได้อย่างแม่นยำเมื่อใช้ร่วมกับ K-Means

ตารางที่ 2 ค่าประสิทธิภาพของโมเดล Decision Tree

Cluster	Accuracy	Precision	Recall	F1-Score
Cluster 1	0.88	0.88	0.88	0.88
Cluster 2	0.85	0.85	0.85	0.85
Cluster 3	1.00	1.00	1.00	1.00
Cluster 4	0.92	0.92	0.92	0.92
Weighted Average	**0.89**	**0.89**	**0.89**	**0.89**



รูปที่ 9 Confusion Matrix ของโมเดล Decision Tree

จากรูปที่ 9 Confusion Matrix จะเห็นได้ว่า โมเดลสามารถจำแนกนักศึกษาในกลุ่ม Cluster 3 ได้ถูกต้องทั้งหมด แสดงให้เห็นว่ากลุ่มนี้มีลักษณะข้อมูลที่ชัดเจนและมีคุณลักษณะเฉพาะตัวสูง ขณะที่ Cluster 2 มีการจำแนกผิดพลาดมากที่สุด สะท้อนให้เห็นว่าข้อมูลของนักศึกษาในกลุ่มนี้มีความหลากหลายทางลักษณะและพฤติกรรม เช่น ความสนใจและทักษะที่มีการทับซ้อนกับกลุ่มอื่น ส่งผลให้ขอบเขตของกลุ่มไม่ชัดเจน ทำให้โมเดลเกิดความสับสนในการจำแนก ส่งผลให้ค่าความแม่นยำของกลุ่มนี้ต่ำกว่ากลุ่มอื่น

อย่างไรก็ตาม เมื่อพิจารณาค่าความถูกต้อง (Accuracy) และค่าเฉลี่ย F1-Score โดยรวมพบว่าโมเดลยังคงอยู่ในระดับที่น่าพอใจ แสดงให้เห็นว่าโมเดล Decision Tree สามารถจำแนกนักศึกษาได้อย่างมีประสิทธิภาพในภาพรวม

ผลการวิเคราะห์ดังกล่าวสอดคล้องกับงานวิจัยของ Arslan et al. [25] ที่พบว่าการใช้เทคนิค K-Means และต้นไม้ตัดสินใจ (Decision Tree) สามารถจำแนกกลุ่มผู้เรียนได้อย่างมีประสิทธิภาพ โดยเฉพาะในด้านพฤติกรรมและความสนใจส่วนบุคคล ซึ่งสะท้อนลักษณะการเรียนรู้ที่แตกต่างกันของผู้เรียนในแต่ละกลุ่ม

นอกจากนี้ ผลการวิจัยยังสนับสนุนแนวคิดของ Zhao [7] และ Guo et al. [11] ที่ชี้ว่าการผสมผสานระหว่างการจัดกลุ่ม (Clustering) และต้นไม้ตัดสินใจ (Decision Tree) ช่วยเพิ่มความแม่นยำของระบบแนะนำ (Recommendation System) ได้ดียิ่งขึ้น ผลลัพธ์ดังกล่าวแสดงให้เห็นว่าผลการศึกษาครั้งนี้มีแนวโน้มสอดคล้องกับงานวิจัยก่อนหน้า ทั้งในเชิงเทคนิคและในด้านพฤติกรรมของผู้เรียน

สรุปผลและอภิปรายผลการวิจัย

ผลการวิจัยนี้สะท้อนความครบถ้วนตามวัตถุประสงค์ที่กำหนด ทั้งในด้านการจัดกลุ่มนักศึกษา การพัฒนาโมเดลแนะนำ และการประเมินประสิทธิภาพของโมเดล การวิเคราะห์เพื่อหาจำนวนกลุ่มที่เหมาะสม ด้วยดัชนี Elbow Method, Davies–Bouldin Index (DBI), Calinski–Harabasz Index (CHI) และ Silhouette Score ชี้ให้เห็นค่าจำนวนกลุ่มที่แตกต่างกัน อย่างไรก็ตาม เมื่อพิจารณาผลลัพธ์เชิงสถิติร่วมกับความเหมาะสมในเชิงบริบทของข้อมูลนักศึกษา พบว่าการแบ่งเป็น 4 กลุ่ม (K = 4) ให้ความสมดุลระหว่างคุณภาพการจัดกลุ่มและความสามารถในการตีความได้ดีที่สุด สอดคล้องกับแนวทางที่แนะนำให้ใช้หลายตัวชี้วัดประกอบกัน ไม่ยึดติดวิธีใดวิธีหนึ่ง [6]

ลักษณะจำแนกกลุ่มนักศึกษาแสดงให้เห็นว่า การจัดกลุ่มแบบ K = 4 สามารถจำแนกคุณลักษณะเฉพาะของนักศึกษาอย่างชัดเจนในแต่ละกลุ่ม โดยกลุ่มที่ 1 สนใจด้านกีฬา/จิตอาสา โดดเด่นทักษะการสื่อสารและการทำงานเป็นทีม (สอดคล้องกับผลการศึกษาที่ชี้ว่ากิจกรรมภาษา/สื่อสารช่วยเสริมสมรรถนะทางสังคม) [26] ส่วนกลุ่มที่ 2 สนใจวิชาการและศิลปะ พบมากในสายบริหาร/การจัดการ มีแนวโน้มเลือกกิจกรรมเพื่อพัฒนาทักษะวิชาชีพ (สอดคล้องกับงานที่ใช้ K-Means จำแนกตามสไตล์การเรียนรู้) [25] กลุ่มที่ 3 สนใจเทคโนโลยี/การแข่งขัน เช่น หุ่นยนต์และเกม เด่นด้านการคิดวิเคราะห์และการใช้เทคโนโลยี เชื่อมโยงกับแนวคิด Experiential Learning [28] และกลุ่มที่ 4 สนใจกิจกรรมผสม (กีฬา–ดนตรี–วิชาการ) เด่นด้านการวางแผน/การจัดการ สอดคล้องกับ Humanistic Learning Theory ที่เน้นการพัฒนาศักยภาพตนเอง [29]

ผลการประเมินประสิทธิภาพของโมเดล Decision Tree ที่สร้างขึ้นจากผลการจัดกลุ่มนักศึกษา พบว่าโมเดลมีความเสถียร โดยมีค่าเฉลี่ย Accuracy/Precision/Recall/F1-Score = 0.89 (แบบถ่วงน้ำหนัก) ระดับกลุ่มพบว่า Cluster 3 ทำได้สูงสุดทุกตัวชี้วัด (=1.00) ขณะที่ Cluster 2 ต่ำสุด (≈0.85) สะท้อนความแตกต่างด้านความชัดเจนของลักษณะข้อมูลในแต่ละกลุ่ม ผลนี้สอดคล้องกับวรรณกรรมที่ระบุว่า Decision Tree ให้ความแม่นยำที่ใช้งานได้จริงและตีความได้ง่าย โดยเฉพาะเมื่อประยุกต์ใช้ร่วมกับการจัดกลุ่มล่วงหน้า [27]

แม้ดัชนีบางตัว เช่น Calinski–Harabasz Index (CHI) หรือ Silhouette Score จะชี้ไปที่จำนวนกลุ่มอื่น แต่การเลือก K = 4 มีความเหมาะสมทั้งในเชิงสถิติ (จุดศอกจาก Elbow และค่าดัชนีที่อยู่ในช่วงดี) และเชิงปฏิบัติ (ตีความง่าย วางแผนกิจกรรมรายกลุ่มได้จริง) ตามแนวคิดการใช้หลายเกณฑ์ร่วมกัน [6] นอกจากนี้การผสาน K-Means กับ Decision Tree ยังทำให้ได้ “โครงสร้างกลุ่ม” ของข้อมูล และกฎการตัดสินใจที่ชัดเจนในรูปแบบ if–then ซึ่งสามารถอธิบายได้ (explainable) และนำไปใช้ได้จริงในบริบทอุดมศึกษา (ที่ต้องสื่อสารเหตุผลการแนะนำกับนักศึกษา/อาจารย์ที่ปรึกษา) [27] อีกทั้งยังมีนัยสำคัญเชิงปฏิบัติที่มหาวิทยาลัยสามารถนำผลการจำแนกกลุ่มไปใช้ ได้แก่ ออกแบบกิจกรรมนักศึกษาแบบมุ่งเป้า (targeted) การจัดสรรทรัพยากรชมรมตามสัดส่วนความสนใจ การใช้เป็นเครื่องมือแนะแนวในช่วงปฐมนิเทศ/ที่ปรึกษา (advisory) และติดตามความสอดคล้องระหว่าง “ชมรมที่เลือก” กับ “โปรไฟล์ความสนใจ–ทักษะ” เพื่อยกระดับการมีส่วนร่วมของนักศึกษา [25], [26]

นอกจากนี้ ผลการวิจัยครั้งนี้สามารถอธิบายเชิงทฤษฎีได้ว่า การประยุกต์ใช้เทคนิค K-Means และ Decision Tree ร่วมกันในบริบทของระบบแนะนำ (Recommendation System) เป็นการผสานระหว่างการเรียนรู้แบบไม่กำกับ (Unsupervised Learning) และแบบกำกับ (Supervised Learning) โดย K-Means ทำหน้าที่จัดกลุ่มนักศึกษาตามลักษณะและความสนใจที่คล้ายกัน เพื่อลดความซับซ้อนของข้อมูล ขณะที่ Decision Tree ทำหน้าที่จำแนกและอธิบายปัจจัยสำคัญในแต่ละกลุ่ม ซึ่งส่งผลให้ระบบแนะนำสามารถให้ข้อเสนอที่เฉพาะเจาะจงกับแต่ละบุคคลได้อย่างมีประสิทธิภาพ สอดคล้องกับแนวคิดของ Zhao [7] และ Guo et al. [8] ที่ชี้ให้เห็นว่าการผสมผสานเทคนิคทั้งสองช่วยยกระดับความแม่นยำและความสามารถในการอธิบายผลลัพธ์ของระบบแนะนำได้ดียิ่งขึ้น

ข้อเสนอแนะ

1. ควรนำระบบแนะนำชมรมที่พัฒนาขึ้นไปทดลองใช้จริงกับนักศึกษาที่สนใจ เพื่อประเมินประสิทธิภาพของระบบในสภาพแวดล้อมจริง และตรวจสอบความเหมาะสมของการแนะนำชมรมให้สอดคล้องกับความสนใจและทักษะของนักศึกษา ทั้งนี้ ผลการทดลองใช้งานจริงจะช่วยให้สามารถปรับปรุงระบบให้มีความแม่นยำและใช้งานได้จริงในระดับมหาวิทยาลัย

2. ควรขยายการใช้งานระบบให้กับหน่วยงานที่เกี่ยวข้อง เช่น กองกิจการนักศึกษา หรือสโมสรนักศึกษา เพื่อใช้เป็นเครื่องมือช่วยแนะแนวการเลือกชมรมสำหรับนักศึกษาใหม่ รวมทั้งใช้ประกอบการวางแผนจัดกิจกรรมเสริมหลักสูตรให้เหมาะสมกับลักษณะของนักศึกษาแต่ละกลุ่ม

3. ควรเพิ่มจำนวนตัวแปรและเทคนิคการวิเคราะห์ในการวิจัยครั้งต่อไป เช่น ปัจจัยด้านบุคลิกภาพ แรงจูงใจ หรือผลการเรียน รวมทั้งทดลองใช้เทคนิคการเรียนรู้ของเครื่อง (Machine Learning) อื่น ๆ เช่น Random Forest, Support Vector Machine (SVM) หรือ Neural Network เพื่อเปรียบเทียบและเพิ่มความแม่นยำของโมเดล

4. ควรพัฒนาระบบให้รองรับการใช้งานจริงในรูปแบบเว็บไซต์หรือแอปพลิเคชัน เพื่อให้นักศึกษาสามารถกรอกข้อมูลและรับคำแนะนำชมรมได้ทันที รวมถึงสามารถเก็บข้อมูลเพิ่มเติมเพื่อใช้ปรับปรุงประสิทธิภาพของระบบในอนาคต

เอกสารอ้างอิง

[1] Ministry of Higher Education, Science, Research and Innovation, *Guidelines for the Development of Thai Higher Education Graduates in the 21st Century*. Bangkok, Thailand: Ministry of Higher Education, Science, Research and Innovation, 2021. (in Thai)

[2] Ministry of Education, *Bachelor's Degree Qualification Standards in Education and Teacher Education (Four-Year Program)*. Bangkok, Thailand: Ministry of Education, 2019. (in Thai)

[3] T. M. Mitchell, *Machine Learning*. New York, NY, USA: McGraw-Hill, 1997.

[4] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. New York, NY, USA: Springer, 2009.

[5] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: A review," *ACM Computing Surveys*, vol. 31, no. 3, pp. 264–323, 1999.

[6] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. San Francisco, CA, USA: Morgan Kaufmann, 2012.

[7] H. Zhao, "Design and implementation of an improved K-means clustering algorithm," *Mobile Information Systems*, Art. no. 6041484, 2022.

[8] N. R. Jevintya, U. Darusalam, and S. Abdullah, "Application of the K-means and decision tree algorithms in determining student achievement," *Jurnal Informatika dan Komputer (JIKO)*, vol. 7, no. 1, pp. 13–18, 2024.

[9] A. H. Anshor and A. T. Zy, "Implementasi metode decision tree pada sistem prediksi status kualitas produk minuman A," *Jurnal Ilmiah Informatika Global*, vol. 15, no. 1, pp. 17–22, 2024.

[10] S. Raschka and V. Mirjalili, *Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow 2*, 3rd ed. Birmingham, UK: Packt, 2019.

[11] Z. Guo, Y. Shi, F. Huang, X. Fan, and J. Huang, "Landslide susceptibility zonation method based on C5.0 decision tree and K-means cluster algorithms to improve the efficiency of risk management," *Geoscience Frontiers*, vol. 12, Art. no. 101249, 2021.

[12] D. M. W. Powers, "Evaluation: From precision, recall and F-measure to ROC, informedness, markedness, and correlation," *Journal of Machine Learning Technologies*, vol. 2, no. 1, pp. 37–63, 2011.

[13] R. A. Stemler, "A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability," *Practical Assessment, Research, and Evaluation*, vol. 9, no. 4, pp. 1–19, 2004.

[14] R. V. Krejcie and D. W. Morgan, "Determining sample size for research activities," *Educational and Psychological Measurement*, vol. 30, no. 3, pp. 607–610, 1970.

[15] W. McKinney, *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython*, 2nd ed. Sebastopol, CA, USA: O'Reilly Media, 2017.

[16] F. Pedregosa *et al*., "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[17] J. D. Hunter, "Matplotlib: A 2D graphics environment," *Computing in Science & Engineering*, vol. 9, no. 3, pp. 90–95, 2007.

[18] F. Chollet, *Deep Learning with Python*, 2nd ed. Shelter Island, NY, USA: Manning Publications, 2021.

[19] A. Agresti and B. Finlay, *Statistical Methods for the Social Sciences*, 4th ed. Upper Saddle River, NJ, USA: Pearson, 2009.

[20] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proc. Int. Joint Conf. Artificial Intelligence (IJCAI)*, pp. 1137–1145, 1995.

[21] L. Breiman, J. Friedman, R. Olshen, and C. Stone, *Classification and Regression Trees*. Belmont, CA, USA: Wadsworth, 1984.

[22] S. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems (NeurIPS)*, pp. 4765–4774, 2017.

[23] World Medical Association, "Declaration of Helsinki: Ethical principles for medical research involving human subjects," *JAMA*, vol. 310, no. 20, pp. 2191–2194, 2013.

[24] A. Field, *Discovering Statistics Using SPSS*, 5th ed. London, UK: SAGE Publications, 2018.

[25] A. Arslan, M. Kaya, and T. Demir, "Clustering students based on learning styles using K-means algorithm," *Education and Information Technologies*, vol. 29, no. 1, pp. 55–70, 2024.

[26] H. Do and P. Le, "Communication activities and student social skills development: A study in higher education," *Education and Training*, vol. 66, no. 2, pp. 123–135, 2024.

[27] J. David and R. Balakrishnan, "Predicting student performance using decision tree models in higher education," *Applied Artificial Intelligence*, vol. 38, no. 5, pp. 412–428, 2024.

[28] D. A. Kolb, *Experiential Learning: Experience as the Source of Learning and Development*, 2nd ed. Upper Saddle River, NJ, USA: Pearson, 2015.

[29] C. Rogers, *Freedom to Learn*, 3rd ed. Columbus, OH, USA: Merrill, 1994.