การระบุรายวิชาศึกษาทั่วไปที่ส่งเสริมการเรียนรู้ของนักศึกษาด้วยชุดข้อมูลขนาดเล็ก
Identifying General Education Courses Enhancing Students' Learning by Using Small Datasets

สุวัฒน์ บรรลือ และประยงค์ ฐิติธนานนท์
Suwat Banlue and Prayong Thitithananon

คณะวิทยาการคอมพิวเตอร์ มหาวิทยาลัยราชภัฏอุบลราชธานี
Faculty of Computer Science, Ubon Ratchathani Rajabhat University
E-mail: suwat@ubru.ac.th, prayong.t@ubru.ac.th

บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์เพื่อ 1) ระบุรายวิชาศึกษาทั่วไปที่ส่งเสริมการเรียนรู้ของนักศึกษา และ 2) เปรียบเทียบประสิทธิภาพของอัลกอริทึมการจำแนกประเภทเพื่อการระบุรายวิชาศึกษาทั่วไปที่ส่งเสริมการเรียนรู้ของนักศึกษาด้วยชุดข้อมูลขนาดเล็ก โดยใช้ข้อมูลการลงทะเบียนเรียน ของนักศึกษาสาขาวิทยาการคอมพิวเตอร์ จำนวนทั้งสิ้น 1,302 คน ในช่วงปีการศึกษา 2550 ถึง 2562 จำนวน 26,804 รายการ ทำการคัดเลือกข้อมูล และนอมัลไลซ์ชุดข้อมูลด้วยเทคนิค MinMaxScaler ใช้ Heat map ในการแสดงผลการจัดกลุ่มตามลำดับชั้นเพื่อแสดงความสัมพันธ์ระหว่างตัวแปร และประเมินประสิทธิภาพของอัลกอริทึม โดยใช้เทคนิค LOOCV

ผลการวิจัยพบว่า

1. รายวิชา 9021103 การรู้สารสนเทศและการเรียนรู้ และรายวิชา 9022102 ภาษาอังกฤษเพื่อการเรียนรู้ข้อมูลข่าวสาร เป็นรายวิชาศึกษาทั่วไปที่ส่งเสริมการเรียนรู้ของนักศึกษาสูงที่สุด

2. พบว่า LDA เป็นอัลกอริทึมที่มีประสิทธิภาพสูงที่สุดในการจำแนกประเภทเพื่อการระบุรายวิชาศึกษาทั่วไปที่ส่งเสริมการเรียนรู้ของนักศึกษาด้วยชุดข้อมูลขนาดเล็ก

คำสำคัญ: อัลกอริทึมการจำแนกประเภท ชุดข้อมูลขนาดเล็ก

Abstract

The objectives of the research were 1) to identify general education courses enhancing students' learning and 2) to compare the efficiency of the classification algorithms used to identify general education courses enhancing students' learning by using small datasets. The research was conducted by using the data from the enrollment information of 1,302 students studying in the computer science with 26,804 enrollment information items during the academic year 2005 to 2019. The data were selected and normalized to be the datasets by using the MinMaxScaler technique that used the Heat map to show the results of the hierarchical clusters to show the relationship between variables and evaluate the efficiency of the algorithm by using the LOOCV technique.

The research findings were as follows.

1. Information and Learning Literacy (9021103) and English for Learning information (9022102) were the general education courses that enhanced the students' learning at the highest level.

2. The LDA was the algorithm that had the highest efficiency in classifying to identify the general education courses that enhanced the students' learning by using the small datasets.

Keywords: Classification Algorithm, Small Datasets

บทนำ

มาตรฐานหลักสูตรระดับปริญญาตรีกำหนดให้ผู้เรียนต้องเรียนรายวิชาศึกษาทั่วไปอย่างน้อย 30 หน่วยกิต รายวิชาเอกและรายวิชาอื่น ๆ อีกอย่างน้อย 90 หน่วยกิต ซึ่งจะเห็นว่าผู้เรียนต้องเรียนวิชาศึกษาทั่วไปประมาณ 1 ใน 4 ของรายวิชาทั้งหมด โดยมาตรฐานดังกล่าวเป็นเพียงการกำหนดกรอบกว้าง ๆ ของโครงสร้างของหลักสูตร แต่รายละเอียดของรายวิชานั้นแต่ละสถาบันจะเป็นผู้ออกแบบและกำหนด โดยส่วนของรายวิชาเอกนั้นแต่ละหลักสูตรจะเป็นผู้กำหนด สำหรับกลุ่มและรายวิชาศึกษาทั่วไปจะถูกกำหนดโดยสถาบันเพื่อให้ทุกหลักสูตรใช้ร่วมกันโดยผู้เรียนจะเป็นผู้เลือกรายวิชาศึกษาทั่วไปตามกลุ่มวิชาและจำนวนหน่วยกิตที่กำหนด ทำให้เกิดความหลากหลาย ดังนั้นหากสถาบันการศึกษาได้วิเคราะห์ถึงประสิทธิภาพของรายวิชาศึกษาทั่วไปที่ส่งเสริมการเรียนรู้ของนักศึกษาเพื่อเป็นแนวทางในการลงทะเบียนเรียนของนักศึกษาในแต่ละหลักสูตรสามารถจัดหลักสูตร หรือจัดการเรียนการสอนในรายวิชาศึกษาทั่วไปให้กับผู้เรียนได้อย่างถูกต้อง เพื่อส่งเสริมคุณภาพ และสร้างความมั่นใจให้กับผู้เรียนได้เป็นอย่างดี แต่เนื่องจากการลงทะเบียนเรียนในรายวิชาศึกษาทั่วไปนั้นมีความหลากหลาย ส่งผลให้ข้อมูลมีความกระจายมาก ทำให้ข้อมูลดังกล่าวมีขนาดเล็กและขาดความสมบูรณ์ซึ่งจะส่งผลถึงความแม่นยำในการประมวลผล เนื่องจากอัลกอริทึ่มจำแนกประเภทมีข้อมูลไม่เพียงพอที่จะทำนายผลได้อย่างแม่นยำ (Tsai and Li, 2015) โดยปัญหาดังกล่าวจะสามารถแก้ไขได้โดยการสร้างความสมบูรณ์ให้กับชุดข้อมูล และการเลือกใช้อัลกอริทึ่มแยกประเภทที่ไม่มีความซับซ้อนมาก (Chao et al, 2011)

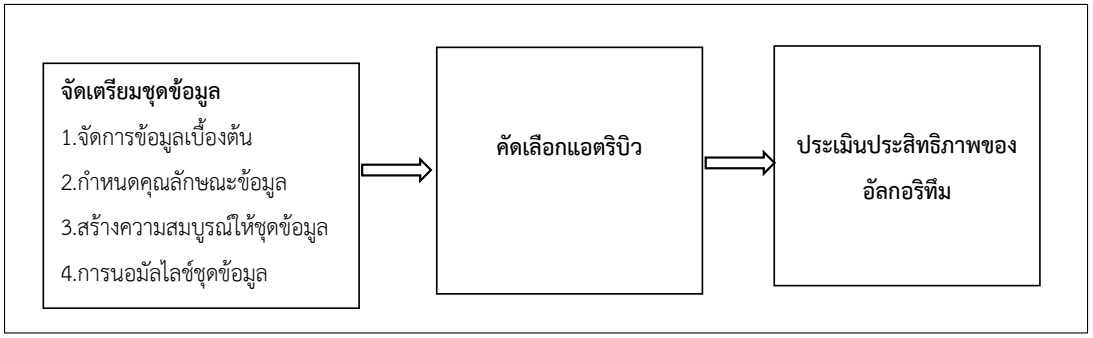
วัตถุประสงค์ของการวิจัย

1. ระบุรายวิชาศึกษาทั่วไปที่ส่งเสริมการเรียนรู้ของนักศึกษา

2. เปรียบเทียบประสิทธิภาพของอัลกอริธึมการการจำแนกประเภทเพื่อการระบุรายวิชาศึกษาทั่วไปที่ส่งเสริมการเรียนรู้ของนักศึกษาด้วยชุดข้อมูลขนาดเล็ก

วิธีดำเนินการวิจัย

การวิจัยนี้มีกรอบแนวคิดและขั้นตอนการดำเนินการวิจัยแบ่งออกเป็น 3 ขั้นตอนหลัก คือ การเตรียมชุดข้อมูล การคัดเลือกแอตริบิว และการประเมินประสิทธิภาพของอัลกอริทึม ดังแสดงในภาพที่ 1



ภาพที่ 1 กรอบแนวคิดและขั้นตอนการวิจัย

ข้อมูล

ข้อมูลที่ใช้ในการวิจัยนี้ เป็นข้อมูลการลงทะเบียนเรียนของนักศึกษาสาขาวิทยาการคอมพิวเตอร์จากฐานข้อมูลระบบทะเบียนมหาวิทยาลัยราชภัฏอุบลราชธานี จำนวน 1,302 คน ในช่วงปีการศึกษา 2550 ถึง 2562 จำนวน 26,804 รายการ โดยมีรายวิชาที่ลงทะเบียนทั้งสิ้น 450 รายวิชา

เครื่องมือในการวิจัย

ในการวิจัยครั้งนี้ใช้ Pandas เป็นเครื่องมือในการวิเคราะห์ข้อมูล และจัดรูปแบบชุดข้อมูลให้เพมาะสมกับใช้งาน ใช้ Matplotlib เพื่อสร้างกราฟ ใช้ Seaborn เพื่อสร้าง Heat map โดยใช้ ภาษา Python และ Scikit-Learn ในการเขียนโปรแกรมเพื่อสร้างโมเดล รวมถึงการวัดประสิทธิภาพของโมเดล

ขั้นตอนการวิจัย

การวิจัยนี้มีรายละเอียดการดำเนินการวิจัยในแต่ละขั้นตอน ดังนี้

การเตรียมข้อมูล

การจัดการข้อมูลเบื้องต้น

ข้อมูลการลงทะเบียนเรียนที่ได้จากระบบทะเบียนของมหาวิทยาลัย เป็นข้อมูลที่มีรายละเอียดของข้อมูลมาก และมีปัญหาเกี่ยวกับตัวข้อมูลเอง เช่น ความครบถ้วนสมบูรณ์ และ ชนิดของข้อมูลซึ่งทำให้ข้อมูลไม่พร้อมที่จะใช้ในขั้นตอนของการวิเคราะห์ข้อมูลได้ ดังนั้นจึงต้องดำเนินการจัดการข้อมูลเบื้องต้นเพื่อให้ได้ชุดข้อมูลที่เหมาะสมกับการวิเคราะห์ในขั้นตอนอื่น ๆ ต่อไป โดยใช้รายละเอียดของข้อมูลเพียงบางส่วนที่จำเป็นต่อการวิเคราะห์ข้อมูล ซึ่งประกอบด้วย รหัสประจำผู้เรียน รหัสวิชาที่ลงทะเบียนเรียนและระดับผลการเรียนในวิชานั้น ๆ ดังแสดงในตารางที่ 1

ตารางที่ 1 รายละเอียดข้อมูล

แอตริบิว	รายละเอียด
std_id	รหัสนักศึกษา
Subjectcode	รหัสวิชา
Grade	ระดับผลการเรียน

ภายหลังการกำหนดรายละเอียดข้อมูลแล้วจะยังคงเหลือข้อมูลนักศึกษา 1,302 คน จำนวน 26,804 รายการ โดยแต่ละรายการมี 3 แอตริบิวที่ได้กำหนดข้างต้น อย่างไรก็ตามจากข้อมูลการลงทะเบียนของนักศึกษาจำนวน

1,302 คน มีรายวิชาที่ลงทะเบียนทั้งหมด 450 รายวิชา เมื่อตัดรายวิชาเลือกเสรี และรายวิชาที่มีผลการเรียนเป็น I, P และ W จะคงเหลือข้อมูลนักศึกษา 1,302 คน จำนวน 200 รายวิชา นอกจากนี้แล้วยังพบว่าบางรายวิชามีการลงทะเบียนเรียนน้อย เนื่องจากนักศึกษาบางกลุ่มเป็นนักศึกษาตกค้าง หรือลงทะเบียนเพื่อแก้ไขผลการเรียน และบางรายวิชาเป็นรายวิชาที่ถูกยกเลิกเมื่อมีการปรับปรุงหลักสูตร โดยเมื่อตัดข้อมูลนักศึกษาที่มีการลงทะเบียนน้อยกว่าร้อยละ 10 ของรายวิชาทั้งหมด (น้อยกว่า 20 รายวิชา) และรายวิชาที่มีนักศึกษาลงทะเบียนเรียนน้อยกว่าร้อยละ 25 ของจำนวนนักศึกษาทั้งหมด (น้อยกว่า 325 คน) จะคงเหลือข้อมูลการลงทะเบียนของนักศึกษาที่ใช้ในการวิจัย จำนวน 814 คน จำนวน 47 รายวิชา ดังแสดงในตารางที่ 2

ตารางที่ 2 ข้อมูลการลงทะเบียนเรียน

ข้อมูล	จำนวน นักศึกษา (คน)	จำนวน รายวิชา (รายวิชา)
การลงทะเบียนเรียนจากระบบทะเบียน	1,302	450
การลงทะเบียนภายหลังการตัดรายวิชาเลือกเสรี รายวิชาที่มีผลการเรียนเป็น I , P และ W	1,302	200
นักศึกษาที่ลงทะเบียนมากกว่าร้อยละ 10 ของรายวิชาทั้งหมด (> 20 รายวิชา)	814	200
รายวิชาที่มีการลงทะเบียนมากกว่าร้อยละ 25 ของจำนวนนักศึกษาทั้งหมด (> 325 คน)	814	47

จากข้อมูลคงเหลือดังแสดงในตารางที่ 2 พบว่ามีรายวิชาการศึกษาทั่วไปจำนวน 11 รายวิชา ส่วนรายวิชาที่เหลือเป็นรายวิชาอื่น ๆ ตามโครงสร้างหลักสูตรจำนวน 36 รายวิชา ซึ่งจะใช้เป็นข้อมูลในการวิเคราะห์ความสัมพันธ์ระหว่างผลการเรียนรายวิชาศึกษาทั่วไปที่มีต่อผลการเรียนรายวิชาอื่น ๆ และค้นหารายวิชาศึกษาทั่วไปที่มีระดับการส่งผลสูงที่สุด

การกำหนดคุณลักษณะชุดข้อมูล

จากการที่อัลกอริทึมที่ใช้ในการวิเคราะห์ชุดข้อมูลขนาดเล็กจะมีประสิทธิภาพสูงเมื่อชุดข้อมูลมีชนิดข้อมูลเป็นตัวเลข เช่น LDA (Sharma and Paliwal, 2015) และ KNN (Ingrassia and Morlini, 2005; Pasini, 2015) สำหรับ SVM เป็นอัลกอริทึมที่ได้รับการออกแบบมาเพื่อให้ทำงานได้อย่างมีประสิทธิภาพกับชนิดข้อมูลที่เป็นตัวเลข ดังนั้นเพื่อให้ข้อมูลมีความเหมาะสมกับการทำงาน ในการวิจัยนี้จึงทำการกำหนดคุณลักษณะและรายละเอียดต่าง ๆ ของชุดข้อมูล โดย ชนิดข้อมูล (Data types) ของผลการเรียน (Grade) ซึ่งมีชนิดข้อมูลเป็นสตริง ดังนี้ ‘A’ ‘B+’ ‘B’ ‘C+’ ‘C’ ‘D+’ ‘D’ และ ‘E’ จะถูกกำหนดให้เป็นข้อมูลชนิดตัวเลข 4.00 3.50 3.00 2.50 2.00 2.50 1.50 1.00 และ 0.00 ตามลำดับ

การสร้างความสมบูรณ์ให้ชุดข้อมูล

แม้ว่าจะมีการดำเนินการตัดข้อมูลที่มีรายละเอียดไม่ครบถ้วนออกจากชุดข้อมูล แล้ว ชุดข้อมูลคงเหลือยังคงมีข้อมูลผลการเรียนบางรายวิชาที่ยังไม่สมบูรณ์เนื่องจากนักศึกษาแต่ละคนมีการลงทะเบียนเรียนที่แตกต่างกัน ทำให้รายวิชาที่ไม่ได้ลงทะเบียนจะไม่มีข้อมูลผลการเรียน จึงทำการสร้างความสมบูรณ์ให้ชุดข้อมูลโดยเติมเต็มข้อมูลผลการเรียนในส่วนที่ขาดหายไปโดยใช้ค่าโหมด (Mode) ของผลการเรียนในรายวิชานั้น ๆ การใช้โหมดของผลการเรียนในการแทนค่าข้อมูลที่ไม่สมบูรณ์นั้นจะทำให้ข้อมูลที่ได้อยู่ภายในกรอบของระดับผลการเรียน หากใช้ค่าเฉลี่ย (Means) หรือมัธยฐาน (Median) อาจส่งผลให้ข้อมูลไม่ตรงกับค่าผลการเรียนที่กำหนดตามตารางที่ 3 ได้ จากนั้นจะทำการคำนวนค่าผลการเรียนเฉลี่ยของนักศึกษาเฉพาะในรายวิชาอื่น ๆ ที่ไม่ใช่รายวิชาศึกษาทั่วไป เพื่อใช้ในการค้นหาความสัมพันธ์ระหว่างผลการเรียนรายวิชาศึกษาทั่วไป กับผลการเรียนเฉลี่ยในรายวิชาอื่น ๆ

ต่อไป

การนอมัลไลซ์ชุดข้อมูล

การนอมัลไลซ์ชุดข้อมูลถือเป็นขั้นตอนสำคัญก่อนส่งชุดข้อมูลเข้าทำการจำแนกประเภท ซึ่งจะทำการปรับค่าข้อมูลให้อยู่ภายในช่วง [0,1] หรือ [- 1,1] ซึ่งจะให้ผลลัพธ์ในการจำแนกประเภทที่มีความแม่นยำสูงกว่าข้อมูลที่มีค่าข้อมูลในช่วงอื่น โดยในการวิจัยนี้ใช้วิธีการนอมัลไลซ์แบบ MinMaxScaler ซึ่งจะปรับค่าของข้อมูลให้อยู่ในช่วง [0,1] โดยใช้สมการ

$$Y = \frac{X - \min(X)}{\max(X) - \min(X)}$$

การคัดเลือกแอตริบิว

การคัดเลือกแอตริบิวเป็นขั้นตอนในการการเลือกคุณสมบัติ โดยใช้ Heat map ในการแสดงผลการจัดกลุ่มตามลำดับชั้นเพื่อแสดงความสัมพันธ์ระหว่างตัวแปร และระบุตัวบ่งชี้หลักที่สามารถส่งเสริมการเรียนรู้ของนักศึกษา ซึ่งจะทำการประมวลผล และคัดเลือกแอตริบิวโดยการพิจารณาจากลักษณะของข้อมูล และจะใช้ผลการบ่งชี้ที่ได้จาก Heat map ในเปรียบเทียบกับตัวบ่งชี้ที่ระบุได้โดยอัลกอริทึมจำแนกประเภทเพื่อตรวจสอบความถูกต้องของการระบุแอตริบิว

การวัดประสิทธิภาพอัลกอริทึม

การวิจัยนี้ใช้อัลกอริทึมการจำแนกประเภทหลายแบบ ซึ่งประกอบด้วย LR, LDA, KNN, CART, NB และ SVM เพื่อวิเคราะห์ค้นหารายวิชาศึกษาทั่วไปที่ส่งเสริมการเรียนรู้ของนักศึกษา และประเมินประสิทธิภาพของอัลกอริทึมแต่ละแบบเมื่อทำงานกับชุดข้อมูลขนาดเล็ก ซึ่งมีผู้วิจัยและค้นพบว่า LDA เป็นอัลกอริทึมที่มีประสิทธิภาพสูงเมื่อทำงานกับชุดข้อมูลขนาดเล็ก โดยเฉพาะอย่างยิ่งในด้านการจดจำใบหน้า เสียงพูด และการคาดการณ์ตลาดการเงิน (Mustafa, Allen and Appiah, 2017; Pasini, 2015; Sharma and Paliwal, 2015) แต่เนื่องจากไม่มีอัลกอริทึมใดที่ถือได้ว่ามีประสิทธิภาพดีในทุกสถานการณ์ ดังนั้น งานวิจัยนี้จะตรวจสอบประสิทธิภาพอัลกอริทึมดังกล่าวข้างต้น โดยการเปรียบเทียบความแม่นยำ (Accuracy) ในการทำงานเพื่อประเมินประสิทธิภาพของอัลกอริทึม โดยใช้เทคนิค Leave-One-Out Cross (LOOCV) ซึ่งเป็นวิธีการตรวจสอบความถูกต้องที่เป็นที่นิยมใช้งานกับชุดข้อมูลขนาดเล็ก (Rao, Fung and Rosales, 2008) โดยแทนที่จะแบ่งชุดข้อมูลออกเป็นชุดฝึกฝนและชุดทดสอบ แต่จะใช้ชุดข้อมูลทั้งหมด (ยกเว้นหนึ่งรายการ) เพื่อฝึกฝนโดยจะทำการประมวลผลแบบวนซ้ำ (Cross-validation หรือ k-fold cross-validation) เพื่อทดสอบชุดข้อมูล และใช้ค่าเฉลี่ยของประสิทธิภาพ (ความแม่นยำ) ของการทดสอบทั้งหมดเป็นผลลัพธ์เพื่อระบุประสิทธิภาพของอัลกอริทึม

การวิเคราะห์ข้อมูล

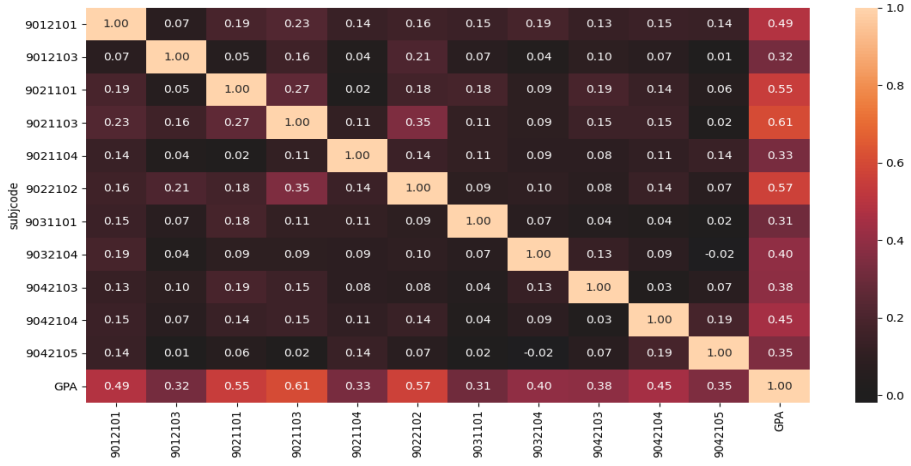
ผลจากการจัดการข้อมูลเบื้องต้นได้ข้อมูลการลงทะเบียนของนักศึกษาที่ใช้ในการวิจัยจำนวน 814 คน จำนวน 47 รายวิชา โดยเป็นรายวิชาการศึกษาทั่วไปจำนวน 11 รายวิชา และรายวิชาอื่น ๆ ตามโครงสร้างหลักสูตรจำนวน 36 รายวิชา และทำการสร้างความสมบูรณ์ให้กับชุดข้อมูลผลการเรียนในส่วนที่ขาดหายไปโดยใช้ค่าโหมดของผลการเรียนในรายวิชานั้น ๆ ทำให้ได้ชุดข้อมูลที่มีข้อมูลนักศึกษา 814 คน โดยมีผลการเรียนครบถ้วนทั้ง 47 รายวิชา และทำการคำนวนผลการเรียนเฉลี่ยเฉพาะรายวิชาที่ไม่ใช่รายวิชาศึกษาทั่วไป โดยเพิ่มข้อมูลผลการเรียนเฉลี่ยที่คำนวนได้เข้าไปในชุดข้อมูล

เมื่อทำการการนอมัลไลซ์ข้อมูลด้วยวิธีการ MinMaxScaler เพื่อปรับค่าของข้อมูลให้อยู่ในช่วง [0,1] ทำให้ได้ชุดข้อมูลใหม่ที่มีการปรับค่าข้อมูลผลการเรียนอยู่ภายในช่วง [0,1] โดยในภาพที่ 2 แสดงตัวอย่างชุดข้อมูลหลังการปรับค่าข้อมูล

subjcode	9012101	9012103	9021101	...	9042104	9042105	GPA
std_id				...			
50122425113	0.5	0.750	0.875	...	0.500	0.875	0.346286
50122425118	0.5	0.625	0.625	...	0.500	0.875	0.388343
50122425119	0.5	0.875	0.875	...	0.500	0.875	0.554286
50122425123	0.5	0.750	0.750	...	0.250	0.875	0.232000
50122425210	0.5	0.625	0.750	...	0.625	0.875	0.346743
...	...	...	...	...	...	...	...
60322420139	0.5	0.625	0.500	...	0.750	0.875	0.454629
60322420142	0.5	0.625	1.000	...	0.750	0.875	0.598629
60322420143	0.5	1.000	0.500	...	0.750	1.000	0.611886
60322420147	0.5	1.000	0.500	...	0.750	0.875	0.564343
60322420150	0.5	0.625	0.500	...	0.750	0.875	0.582171

ภาพที่ 2 แสดงชุดข้อมูลหลังการปรับค่าข้อมูล

ผลจากการใช้ heat map เพื่อแสดงผลการจัดกลุ่มตามลำดับชั้น และแสดงความสัมพันธ์ระหว่างผลการเรียนรายวิชาศึกษาทั่วไป กับผลการเรียนเฉลี่ยในรายวิชาอื่น ๆ ตามโครงสร้างหลักสูตร แสดงในภาพที่ 3 ซึ่ง พบว่ารายวิชาที่สามารถส่งเสริมการเรียนรู้ของนักศึกษามากที่สุดคือรายวิชา 9021103 การรู้สารสนเทศและการเรียนรู้ และ 9022102 ภาษาอังกฤษเพื่อการเรียนรู้ข้อมูลข่าวสาร โดยจะใช้รายวิชานี้ในเปรียบเทียบกับรายวิชาที่ระบุโดยอัลกอริทึมจำแนกประเภทแต่ละแบบเพื่อตรวจสอบความถูกต้องของการทำงานของอัลกอริทึม

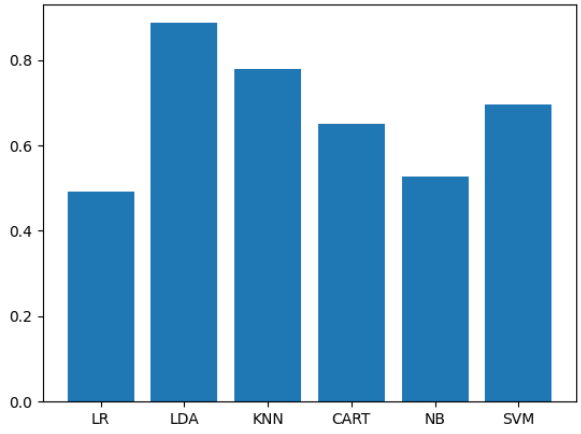


ภาพที่ 3 heat map แสดงความสัมพันธ์ของรายวิชา

เมื่อทดสอบชุดข้อมูลกับอัลกอริทึมการจำแนกประเภทแต่ละแบบและตรวจสอบความถูกต้องในการทำงานของอัลกอริทึมแต่ละแบบด้วยวิธีการ LOOCV ได้ค่าความแม่นยำ ซึ่งบ่งบอกถึงประสิทธิภาพของอัลกอริทึม ดังแสดงในตารางที่ 3 และกราฟแสดงประสิทธิภาพของอัลกอริทึม แสดงในภาพที่ 5

ตารางที่ 3 แสดงประสิทธิภาพอัลกอริทึมจำแนกประเภท

อัลกอริธึม	ประสิทธิภาพ	
	ความแม่นยำ	SD.
LR	0.471034	0.058827
LDA	0.877126	0.037894
KNN	0.774828	0.039961
CART	0.532644	0.061417
NB	0.443563	0.089254
SVM	0.641609	0.066963



ภาพที่ 5 แสดงประสิทธิภาพของอัลกอริทึม

จากผลการวิเคราะห์ข้อมูลในตารางที่ 3 พบว่า LDA มีอัตราความแม่นยำในการทำงานสูงที่สุดโดยมีความแม่นยำเท่ากับ 0.877126 เมื่อพิจารณาการระบุรายวิชาที่ส่งเสริมการเรียนรู้ของนักศึกษา โดยอัลกอริทึม LDA พบว่ารายวิชา 9021103 การรู้สารสนเทศและการเรียนรู้ มีค่าความสำคัญ 115.91710 และรายวิชา 9022102 ภาษาอังกฤษเพื่อการเรียนรู้ข้อมูลข่าวสาร มีค่าความสำคัญ 100.42111 เป็นรายวิชาศึกษาทั่วไปที่ส่งเสริมการเรียนรู้ของนักศึกษาได้ดีที่สุดและรองลงมาตามลำดับ โดยมีค่าน้ำหนักความสำคัญของรายวิชาดังแสดงในตารางที่ 4

ตารางที่ 4 แสดงค่าน้ำหนักรายวิชาที่ส่งเสริมการเรียนรู้ที่ได้จากการระบุโดย LDA

ความสำคัญของรายวิชา											
รายวิชา	**9012101**	**9012103**	**9021101**	**9021103**	**9021104**	**9022102**	**9031101**	**9032104**	**9042103**	**9042104**	**9042105**
น้ำหนัก	83.31632	63.20597	83.95628	**115.91710**	65.59424	**100.42111**	56.50617	78.18573	74.16519	79.97605	71.10661

สรุปผลการวิจัย

การวิจัยนี้มีวัตถุประสงค์เพื่อระบุรายวิชาศึกษาทั่วไปที่ส่งเสริมการเรียนรู้ของนักศึกษา และเปรียบเทียบประสิทธิภาพของอัลกอริธึมจำแนกประเภทเพื่อระบุรายวิชาศึกษาทั่วไปที่ส่งเสริมการเรียนรู้ของนักศึกษาด้วยชุดข้อมูลขนาดเล็ก ซึ่งผลจากการวิเคราะห์ข้อมูลด้วย Heat map พบว่ารายวิชา 9021103 การรู้สารสนเทศและการเรียนรู้ และ 9022102 ภาษาอังกฤษเพื่อการเรียนรู้ข้อมูลข่าวสาร เป็นรายวิชาศึกษาทั่วไปที่สามารถส่งเสริมการเรียนรู้ของนักศึกษาได้ดีที่สุด ในขณะที่ผลการระบุรายวิชาโดยอัลกอริทึมจำแนกรูปแบบพบว่ารายวิชา 9021103 การรู้สารสนเทศและการเรียนรู้ และ 9022102 ภาษาอังกฤษเพื่อการเรียนรู้ข้อมูลข่าวสาร เป็นรายวิชาศึกษาทั่วไปที่สามารถส่งเสริมการเรียนรู้ของนักศึกษาได้ดีที่สุดเช่นเดียวกัน และจากการเปรียบเทียบความแม่นยำในการทำงานพบว่า LDA เป็นอัลกอริทึมจำแนกรูปแบบที่มีประสิทธิภาพและมีความแม่นยำสูงที่สุดในการระบุรายวิชาศึกษาทั่วไปที่ส่งเสริมการเรียนรู้ของนักศึกษา

อภิปรายผลการวิจัย

รายวิชา 9021103 การรู้สารสนเทศและการเรียนรู้ และ 9022102 ภาษาอังกฤษเพื่อการเรียนรู้ข้อมูลข่าวสาร เป็นรายวิชาที่ส่งผลต่อการเรียนของนักศึกษาโดยเฉพาะอย่างยิ่งกับนักศึกษาสาขาวิทยาการคอมพิวเตอร์ ซึ่งใช้เป็นข้อมูลในการวิจัยครั้งนี้ แต่เนื่องจากรายวิชาดังกล่าวเป็นรายวิชาศึกษาทั่วไป ซึ่งนักศึกษาอาจเลือกที่จะไม่ลงทะเบียนเรียน หรือเรียนรายวิชา อื่น ๆ ทดแทนเพื่อให้จำนวนหน่วยกิตครบตามหลักสูตร ดังนั้นหากกำหนดให้เป็นวิชาบังคับให้นักศึกษาในกลุ่มเป้าหมายต้องเรียนในรายวิชานี้ ก็จะเป็นส่วนหนึ่งในการส่งเสริมการเรียนรู้ของนักศึกษาได้เป็นอย่างดี นอกจากนี้แล้วในการวิจัยนี้พบว่าอัลกอริทึมจำแนกประเภทที่มีความแม่นยำในการทำงานกับข้อมูลขนาดเล็ก คือ LDA สอดคล้องกับ (Mahmoud and Zohair, 2019) ที่พยากรณ์ประสิทธิภาพของผู้เรียนโดยใช้ชุดข้อมูลขนาดเล็ก และ (Mueen, Zafar and Manzoor, 2016) ที่สร้างโมเดลเพื่อพยากรณ์ประสิทธิภาพการเรียนซึ่งพบว่า LDA เป็นอัลกอริทึมที่มีความแม่นยำในการทำงานสูงที่สุด

ข้อเสนอแนะ

ข้อจำกัดของการวิจัยครั้งนี้คือ ชุดข้อมูลมีขนาดเล็ก และการขาดความสมบูรณ์ของข้อมูล และเมื่อดำเนินการกับข้อมูลเบื้องต้น โดยการคัดเลือกข้อมูล ก็ยิ่งทำให้ชุดข้อมูลมีขนาดเล็กลงไปอีก ดังนั้นในการวิจัยครั้งต่อไป จะใช้ข้อมูลการลงทะเบียนของนักศึกษาในทุกสาขา และเพิ่มจำนวนปีการศึกษาที่เก็บข้อมูล เพื่อให้ได้ชุดข้อมูลที่มีขนาดใหญ่ ซึ่งจะทำให้มีประสิทธิภาพและความแม่นยำสูงขึ้น

กิตติกรรมประกาศ

ขอขอบคุณคณะวิทยาการคอมพิวเตอร์ มหาวิทยาลัยราชภัฏอุบลราชธานี ที่อนุเคราะห์ข้อมูลเพื่อการวิจัย ขอขอบคุณ ผศ.ดร. เสทื้อน เทพรงทอง และ ผศ.เสรี หร่ายเจริญ ที่ได้ให้ข้อเสนอแนะ และคำแนะนำที่สำคัญ ต่อการวิจัยในครั้งนี้

เอกสารอ้าง

Chao, G. et al. “A new approach to prediction of radiotherapy of bladder cancer cells in small dataset analysis,” **Expert Syst.** 38, 7 (2011): 7963–7969.

Ingrassia, S. and I. Morlini. **Neural network modeling for small datasets.** (online) 2005 (cited 24 January 2021). Available from: https://www.researchgate.net/publication/232795375_Neural_Network_Modeling_for_Small_Datasets

Mahmoud. L and A. Zohair. “Prediction of Student’s performance by modelling small dataset size,” **International Journal of Educational Technology in Higher Education.** 16, 1 (2019): 1-18.

Mueen, A., B. Zafar and U. Manzoor. “Modeling and Predicting Students’ Academic Performance Using Data Mining Techniques,” **Modern Education and Computer Science.** 8, 11 (2016): 36-42.

Mustafa, M. K., T. Allen and K. Appiah. **A comparative review of dynamic neural networks and hidden Markov model methods for mobile on-device speech recognition.** (online) 2017 (cited 5 January 2021). Available from: https://www.researchgate.net/publication/317768552_A_comparative_review_of_dynamic_neural_networks_and_hidden_Markov_model_methods_for_mobile_on-device_speech_recognition

Pasini, A. **Artificial neural networks for small dataset analysis.** (online) 2015 (cited 12 February 2021). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4454870/

Rao, R. B., G. Fung and R. Rosales. **On the dangers of cross-validation. An experimental evaluation.** (online) 2008 (cited 12 February 2021). Available from: https://people.csail.mit.edu/romer/papers/CrossVal_SDM08.pdf

Sharma, A. and K. K. Paliwal. **Linear discriminant analysis for the small sample size problem: An** (online) 2015 (cited 25 February 2021). Available from: https://core.ac.uk/download/pdf/143905484.pdf.

Tsai, C.H and D.C. Li. **“Improving Knowledge Acquisition Capability of M5’ Model Tree on Small Datasets,”** 2015 3rd International Conference on Applied Computing and Information Technology 2nd International Conference on Computational Science and Intelligence, 2015. pp.379-386.