

การวิเคราะห์ปัจจัยที่ส่งผลต่อผลการเรียนของบัณฑิตสาขาวิชาวิทยาการคอมพิวเตอร์ โดยใช้ชุดข้อมูลที่ไม่สมดุล

Analysis of Factors Affecting Academic Performance of Computer Science Major Graduates using Imbalanced Datasets

สุวัฒน์ บรรลือ¹, ขนิษฐา อินทแสง² และ ประยงค์ จูฑิตธนานนท์³

Suwat Banlue¹, Khanittha Inthasaeng² and PrayongThitithananon³

Received : 2 เม.ย. 2564

Revised : 14 ก.ค. 2564

Accepted : 15 ก.ค. 2564

บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์ 1) เพื่อค้นหารายวิชาที่ส่งผลต่อผลการเรียนตลอดหลักสูตรของบัณฑิตสาขาวิชาวิทยาการคอมพิวเตอร์ มหาวิทยาลัยราชภัฏอุบลราชธานี และ 2) เพื่อนำผลจากการวิจัยไปใช้เป็นแนวทางในการปรับปรุงหลักสูตรและกระบวนการจัดการเรียนการสอนในรายวิชาที่ส่งผลต่อผลการเรียน ข้อมูลที่ใช้เป็นข้อมูลทุติยภูมิจากฐานข้อมูลระบบทะเบียนมหาวิทยาลัยราชภัฏอุบลราชธานี โดยเป็นข้อมูลการลงทะเบียนเรียนของนักศึกษาสาขาวิชาวิทยาการคอมพิวเตอร์จำนวนทั้งสิ้น 399 คน ในช่วงปีการศึกษา 2558 ถึง 2562 จำนวน 6,435 รายการ ทำการคัดเลือกและปรับสมดุลข้อมูลด้วยวิธีการ ADASYN ได้ข้อมูลการลงทะเบียนทั้งสิ้น 266 คน และรายวิชาที่เกี่ยวข้อง 19 รายวิชา ทำการคัดเลือกรายวิชาด้วย Recursive Feature Elimination

ผลการวิจัยพบว่า มี 9 รายวิชาที่มีผลต่อผลการเรียนและระบุลำดับด้วย Univariate Selection และ Feature Importance พบว่ารายวิชาการออกแบบและวิเคราะห์ขั้นตอนวิธี เป็นรายวิชาที่มีความสำคัญเป็นลำดับแรกที่มีผลต่อผลการเรียน โดยนักศึกษาที่ได้รับระดับคะแนนสูงในรายวิชานี้จะมีผลการเรียนในภาพรวมที่สูงด้วย

คำสำคัญ: ชุดข้อมูลที่ไม่สมดุล, เทคนิคการสังเคราะห์ข้อมูลเพิ่ม, วิเคราะห์ข้อมูล

Abstract

The objectives of this research were to 1) search for courses that affect students' academic performance and 2) propose guidelines for curriculum revision and teaching and learning development of the courses that affect their academic performance. The data used as secondary information was retrieved from Ubon Ratchathani Rajabhat University Registration System database. The data covered 6,435 items of the registration of 399 computer science major students between the academic years 2015 to 2019. ADASYN method was used in selecting and adjusting the imbalanced dataset. Finally, 266 students and 19 related courses were obtained. The courses were selected through Recursive Feature Elimination.

¹ ผู้ช่วยศาสตราจารย์ ดร. คณะวิทยาการคอมพิวเตอร์ มหาวิทยาลัยราชภัฏอุบลราชธานี อีเมล: suwat@ubru.ac.th

² ผู้ช่วยศาสตราจารย์ คณะวิทยาการคอมพิวเตอร์ มหาวิทยาลัยราชภัฏอุบลราชธานี อีเมล: khanittha.i@ubru.ac.th

³ รองศาสตราจารย์ คณะวิทยาการคอมพิวเตอร์ มหาวิทยาลัยราชภัฏอุบลราชธานี อีเมล: prayong.t@ubru.ac.th

¹ Assistant Professor, Ph.D., Faculty of Computer Science, Ubon Ratchathani Rajabhat University, Email: suwat@ubru.ac.th

² Assistant Professor, Faculty of Computer Science, Ubon Ratchathani Rajabhat University, Email: khanittha.i@ubru.ac.th

³ Associate Professor, Faculty of Computer Science, Ubon Ratchathani Rajabhat University, Email: prayong.t@ubru.ac.th

The results of the research revealed that there were 9 courses that had great effects on academic performance. Based on Univariate Selection and Feature Importance, it was found that the Design and Analysis of Algorithms course was the first priority course that affects academic performance. Students who achieve a high grade in this course will also have high overall academic performance in other courses.

Keywords: Imbalanced datasets, Adaptive synthetic sampling, Data analysis

บทนำ

ปัญหาสำคัญของการจัดการศึกษา คือการที่ผู้สำเร็จการศึกษามีสมรรถนะความรู้และทักษะไม่ตรงตามวัตถุประสงค์ของหลักสูตร โดยสามารถแยกปัญหาดังกล่าวออกได้เป็น 2 ส่วน คือปัญหาที่เกิดจากตัวผู้เรียนเอง เช่น ความรู้พื้นฐานก่อนเข้าเรียน ระดับความสามารถ และความถนัดในสาขาที่เรียน และปัญหาที่เกิดจากตัวหลักสูตรและกระบวนการจัดการเรียนการสอนในหลักสูตร ซึ่งเป็นปัจจัยสำคัญที่ส่งผลต่อผลการเรียนของผู้เรียน ทั้งนี้เนื่องจากการจัดการเรียนการสอนนั้นจะเริ่มจากการจัดรายวิชาพื้นฐานของหลักสูตรและต่อเนื่องด้วยรายวิชาอื่น ๆ ที่เกี่ยวข้องในกลุ่มเนื้อหานั้น หากผู้เรียนไม่ประสบความสำเร็จในการเรียนในรายวิชาพื้นฐานเหล่านั้น ผลกระทบก็จะส่งไปยังการเรียนในรายวิชาอื่น ๆ และส่งผลต่อเนื่องไปยังผลการเรียนตลอดหลักสูตร และจากการที่โครงสร้างของหลักสูตรมีรายวิชาต่าง ๆ อยู่เป็นจำนวนมาก การค้นหารายวิชาที่ส่งผลต่อผลการเรียนของผู้เรียนจึงมีความสำคัญอย่างยิ่งต่อการพัฒนาปรับปรุงหลักสูตร รวมถึงการปรับปรุงกระบวนการเรียนการสอนในรายวิชาเหล่านั้น เพื่อส่งเสริมการเรียนรู้ของผู้เรียนในรายวิชาอื่นที่เกี่ยวข้องต่อไป แต่อย่างไรก็ตามการวิเคราะห์ข้อมูลการเรียนการสอนเพื่อหาปัจจัยที่ส่งผลต่อผลการเรียนรู้นักศึกษาทั้งหมดนั้นอาจจะเป็นไปไม่ได้โดยง่าย เนื่องจากมีข้อมูลจำนวนมากที่มีการส่งผลถึงผลการเรียนรู้นักศึกษาแต่ละคน (Asif, 2015: 49-61) จากการรวบรวมข้อมูลเบื้องต้นพบว่า เป็นชุดข้อมูลที่มีความไม่สมดุลกัน (Imbalanced datasets) (Maheshwari, Agrawal and Sharma, 2011: 12) โดยมีจำนวนกลุ่มนักศึกษาที่มีผลการเรียนอ่อนมาก มากกว่ากลุ่มนักศึกษาที่มีผลการเรียนดีมาก ปัญหาความไม่สมดุลของชุดข้อมูล เป็นปัญหาที่พบมากในชุดข้อมูลทั่วไปและสามารถเกิดขึ้นได้ในหลากหลายโดเมนข้อมูล เช่น ข้อมูลด้านการจัดการความเสี่ยง ข้อมูลด้านการรู้จำใบหน้า ข้อมูลด้านการศึกษา เป็นต้น ปัญหาความไม่สมดุลของชุดข้อมูลโดยทั่วไปแล้วอาจเกิดจากสาเหตุที่ข้อมูลมีการทับซ้อนกัน หรือมีความซับซ้อนของข้อมูลสูง ซึ่งทำให้ยากต่อการจำแนกคุณลักษณะ (Jaidee and Wannapee, 2020: 76) อีกปัจจัยที่มีความสำคัญในการจำแนกคุณลักษณะข้อมูล คือ กรณีที่ขนาดของชุดข้อมูลมีขนาดเล็ก (Small sample size) อาจทำให้เกิดปัญหาข้อมูลในการเรียนรู้ไม่เพียงพอ เช่น จำนวนนักศึกษาทั้งหมด 185 คน มีนักศึกษาที่มีผลการเรียนดีมากเพียง 10 คน แต่มีนักศึกษาที่ผลการเรียนอ่อน หรืออ่อนมากจำนวน 95 คน ซึ่งปัญหาความไม่สมดุลของข้อมูลในลักษณะนี้ทำให้ไม่สามารถจำแนกคุณลักษณะชุดข้อมูลได้อย่างมีประสิทธิภาพ

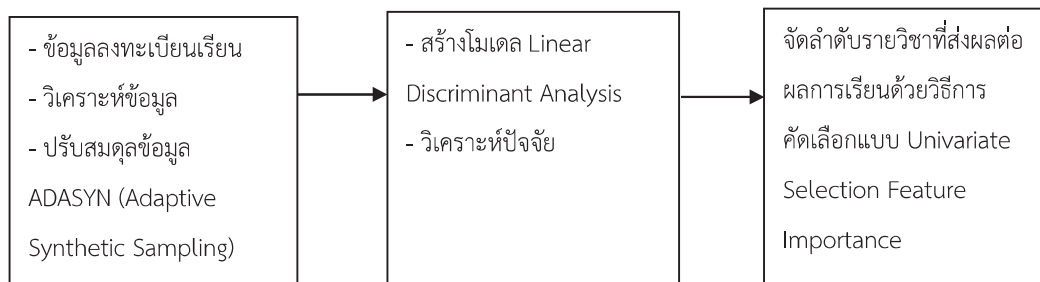
ในงานวิจัยนี้คณะผู้วิจัยเห็นว่า วิธีการสังเคราะห์ข้อมูลเพิ่มแบบ ADASYN (Adaptive Synthetic Sampling) การวิเคราะห์ปัจจัยที่ส่งผลต่อผลการเรียนด้วยวิธี Recursive Feature และการจัดลำดับรายวิชาที่ส่งผลต่อผลการเรียนด้วยการคัดเลือกแบบ Univariate Selection และ Feature Importance จะทำให้ทราบถึงปัจจัยที่ส่งผลต่อผลการเรียนของนักศึกษาได้อย่างถูกต้อง

วัตถุประสงค์

การวิจัยนี้มีวัตถุประสงค์ ดังนี้

1. เพื่อค้นหารายวิชาที่ส่งผลต่อผลการเรียนตลอดหลักสูตรของบัณฑิตสาขาวิทยาการคอมพิวเตอร์ มหาวิทยาลัยราชภัฏอุบลราชธานี
2. เพื่อนำผลจากการวิจัยไปใช้เป็นแนวทางในการปรับปรุงหลักสูตรและกระบวนการจัดการเรียนการสอนในรายวิชาที่ส่งผลต่อผลการเรียน

กรอบแนวคิด



ภาพประกอบ 1 กรอบแนวคิดและขั้นตอนการวิจัย การวิเคราะห์ปัจจัยที่ส่งผลต่อผลการเรียนของบัณฑิตสาขาวิทยาการคอมพิวเตอร์ โดยใช้ชุดข้อมูลที่ไม่สมดุล

วิธีดำเนินการวิจัย

1. การจัดเตรียมข้อมูล

ข้อมูลที่ใช้ในการดำเนินการวิจัยนี้ เป็นข้อมูลทุติยภูมิจากฐานข้อมูลระบบทะเบียนมหาวิทยาลัยราชภัฏอุบลราชธานี โดยเป็นข้อมูลการลงทะเบียนเรียน และระดับผลการเรียนรายวิชาของนักศึกษาสาขาวิทยาการคอมพิวเตอร์ จำนวน 399 คน ในช่วงปีการศึกษา 2558 ถึง 2562 จำนวน 6,435 รายการ ทำการคัดเลือกข้อมูลที่เป็นวิชาแกน และวิชาเฉพาะด้านตามโครงสร้างของหลักสูตรจำนวน 19 รายวิชา และตัดนักศึกษาบางกลุ่มที่เป็นนักศึกษาตกค้าง หรือลงทะเบียนเพื่อแก้ไขผลการเรียน ซึ่งทำให้ข้อมูลของนักศึกษาในกลุ่มดังกล่าวขาดความสมบูรณ์ของข้อมูล คงเหลือข้อมูลการลงทะเบียนของนักศึกษา 185 คน จำนวน 3,282 รายการ โดยมีรายละเอียดเบื้องต้นของข้อมูล ดังแสดงในตาราง 1

ตาราง 1 รายละเอียดเบื้องต้นของข้อมูล

รายละเอียด	จำนวน (คน)	จำนวน (รายการ)
ข้อมูลลงทะเบียนเรียน ปีการศึกษา 2558 ถึง 2562 (ทุกรายวิชา)	399	6,435
ข้อมูลลงทะเบียนเฉพาะวิชาแกน และวิชาเฉพาะด้าน	399	4,526
ข้อมูลลงทะเบียนที่คัดกรองแล้ว	185	3,282

2. การวิเคราะห์ข้อมูลเชิงสำรวจ

จากจำนวนนักศึกษา 185 คน ที่ลงทะเบียนเรียนในวิชาแกน และวิชาเฉพาะ เมื่อทำสำรวจข้อมูลเบื้องต้น และจำแนกข้อมูลโดยใช้ผลการเรียนเฉลี่ยตามเกณฑ์ที่แสดงในตาราง 2

ตาราง 2 เกณฑ์การจำแนกกลุ่มโดยใช้ระดับผลการเรียนเฉลี่ย

ผลการเรียนเฉลี่ย	ระดับ
3.50 – 4.00	ดีมาก
3.00 – 3.49	ดี
2.50 – 2.99	ปานกลาง
2.00 – 2.49	อ่อน
ต่ำกว่า 2.00	อ่อนมาก

จากเกณฑ์การจำแนกในตาราง 2 สามารถจำแนกข้อมูลออกได้เป็น 5 กลุ่ม คือ ดีมาก 10 คน ดี 39 คน ปานกลาง 40 คน อ่อน 40 คน และอ่อนมาก 55 คน ดังรายละเอียดในตาราง 3

ตาราง 3 ระดับผลการเรียน (วิชาแกนและวิชาเฉพาะ)

ระดับผลการเรียน	จำนวน (คน)
ดีมาก	10
ดี	39
ปานกลาง	40
อ่อน	40
อ่อนมาก	55
รวม	185

จากตาราง 3 พบว่านักศึกษาโดยส่วนใหญ่มีผลการเรียนเฉลี่ยในวิชาแกนและวิชาเฉพาะค่อนข้างต่ำ ซึ่งส่งผลให้ผลการเรียนเฉลี่ยตลอดหลักสูตรต่ำด้วย เนื่องจากวิชาแกนและวิชาเฉพาะมีจำนวนรายวิชารวมคิดเป็น 2 ใน 3 ของรายวิชาทั้งหมดในหลักสูตร

3. การปรับสมดุลข้อมูล

จากการที่จำนวนข้อมูลในแต่ละกลุ่มขาดความสมดุล กล่าวคือ ระดับผลการเรียนต่ำที่สุด 10 คน และสูงที่สุด 55 คน จากจำนวนข้อมูลทั้งสิ้น 185 คน จะก่อให้เกิดปัญหาการขาดความแม่นยำของโมเดลได้ ดังนั้นก่อนที่จะดำเนินการในขั้นตอนของการสร้างโมเดล จึงจำเป็นต้องมีการปรับสมดุลของข้อมูลเพื่อให้โมเดลที่สร้างขึ้นมีความเที่ยงตรงสูงขึ้น โดยใช้เทคนิคการสังเคราะห์ข้อมูลเพิ่มแบบ ADASYN (Adaptive Synthetic Sampling) (Ashfaq, Booma and Mafas, 2020: 2277-2283) โดยใช้ค่าการแจกแจงแบบถ่วงน้ำหนัก (Weight distribution) ของข้อมูลตัวอย่างในกลุ่มที่มีข้อมูลน้อย

และสร้างข้อมูลเทียมซึ่งขึ้นอยู่กับความสำคัญของข้อมูลนั้น ๆ ถ้าข้อมูลตัวใดยากต่อการแบ่งกลุ่มก็จะให้ค่าน้ำหนักข้อมูลนั้นมากและสร้างชุดข้อมูลเทียมขึ้นมาบริเวณนั้น ๆ (Alyahyan and Duştegor, 2020: 1-22) ซึ่งจะทำให้มีการปรับขอบเขตของการตัดสินใจในการแบ่งกลุ่มดีขึ้น ดังนั้นจะเห็นว่าการสร้างตัวอย่างสังเคราะห์จะขึ้นอยู่กับระดับข้อมูลที่มีจำนวนน้อย จึงทำให้ ADASYN สามารถเพิ่มประสิทธิภาพการกระจายข้อมูลโดยลดความลำเอียงในข้อมูลได้ โดยจะเปลี่ยนแปลงเฉพาะข้อมูลที่มีจำนวนน้อยเท่านั้น สำหรับข้อมูลกลุ่มอื่น ๆ จะไม่มีการเปลี่ยนแปลง (Sales, Balby and Cajueiro, 2016: 166–180) โดยได้ข้อมูลภายหลังการปรับสมดุลข้อมูลแสดงในตาราง 4

ตาราง 4 จำนวนข้อมูลภายหลังการปรับสมดุล

ระดับผลการเรียน	จำนวน (คน)
ดีมาก	56
ดี	56
ปานกลาง	47
อ่อน	52
อ่อนมาก	55
รวม	266

4. การสร้างโมเดลและการประเมินประสิทธิภาพของโมเดล

เพื่อให้ได้โมเดลมีความแม่นยำ และมีประสิทธิภาพการทำงานสูงเมื่อนำไปใช้งานกับข้อมูลจริง จะทำการคัดเลือกอัลกอริทึมด้วยการเปรียบเทียบประสิทธิภาพในการทำงานของอัลกอริทึมแบบต่าง ๆ กับข้อมูลชุดเดียวกัน โดยภายหลังจากการปรับสมดุลชุดข้อมูลแล้วจะแบ่งข้อมูลออกเป็นชุดข้อมูลเพื่อการฝึกฝนและชุดข้อมูลเพื่อการทดสอบ สำหรับสัดส่วนในการแบ่งชุดข้อมูลนั้นจะขึ้นอยู่กับจำนวนของข้อมูลที่มี ในกรณีที่ข้อมูลมีจำนวนมากจะกำหนดให้ชุดข้อมูลเพื่อการฝึกฝนมีสัดส่วนมาก และอาจมีการปรับสัดส่วนชุดข้อมูล และทำการทดสอบซ้ำเพื่อตรวจสอบประสิทธิภาพของอัลกอริทึม (Vikas, Raykar and Saha, 2015: 3-19) สำหรับงานวิจัยนี้มีข้อมูลคงเหลือหลังจากผ่านการคัดกรองมีจำนวน 3,282 รายการ ซึ่งเป็นชุดข้อมูลขนาดเล็ก จึงได้ทำการทดสอบการแบ่งชุดข้อมูลด้วยสัดส่วนต่าง ๆ และปรับสัดส่วนเพื่อตรวจสอบการทำงานและพบว่า การแบ่งชุดข้อมูลด้วยการแจกแจงร้อยละ 70/30 สามารถครอบคลุมค่าความแปรปรวนได้ จึงใช้ชุดข้อมูลที่แบ่งได้กับอัลกอริทึมรูปแบบต่าง ๆ ทั้งแบบ Linear และ Non Linear เพื่อคัดเลือกอัลกอริทึมที่มีความแม่นยำสูงที่สุด โดยทำการประเมินประสิทธิภาพของโมเดลด้วยวิธีการตรวจสอบไขว้แบบ 10 กลุ่ม เพื่อให้ข้อมูลทุกตัวมีโอกาสเป็นชุดฝึกฝน และชุดทดสอบ และจากผลการเปรียบเทียบประสิทธิภาพของโมเดลกับชุดข้อมูล พบว่าอัลกอริทึม Linear Discriminant Analysis มีความแม่นยำสูงที่สุด โดยมีผลการวัดประสิทธิภาพของโมเดล ดังแสดงในตาราง 5

ตาราง 5 ประสิทธิภาพของโมเดล

Algorithm	Accuracy	S.D.
Simple Linear		
Logistic Regression	0.607310	0.048928
Linear Discriminant Analysis	0.886550	0.078787
Non Linear		
K-Neighbors Classifier	0.816374	0.058301
Decision Tree Classifier	0.715205	0.093925
Gaussian Naïve Bayes	0.858772	0.067620
Support Vector Machine	0.870175	0.066495

นอกจากนี้แล้ว เมื่อใช้อัลกอริทึม Linear Discriminant Analysis กับข้อมูลที่ยังไม่มีการปรับสมดุลข้อมูล พบว่ามีความแม่นยำในการทำงานเป็น 0.743589 ซึ่งแสดงให้เห็นว่าการปรับสมดุลข้อมูลมีผลทำให้ประสิทธิภาพการทำงานของอัลกอริทึมสูงขึ้น

5. การวิเคราะห์ปัจจัย

การวิเคราะห์ปัจจัยที่ส่งผลต่อผลการเรียน ดำเนินการโดยระบุรายวิชาที่ส่งผลต่อผลการเรียนด้วยวิธี Recursive Feature Elimination เพื่อคัดเลือกรายวิชาที่ส่งผลต่อผลการเรียนมากที่สุด โดยไม่มีการระบุน้ำหนักหรือลำดับ และทำการจัดลำดับรายวิชาที่ส่งผลต่อผลการเรียนด้วยวิธีการคัดเลือกแบบ Univariate Selection และ Feature Importance โดย Univariate Selection เป็นการทดสอบค่าทางสถิติเพื่อประเมินคุณลักษณะที่มีความสัมพันธ์ หรือส่งผลต่อเป้าหมายมากที่สุด โดยในการวิจัยครั้งนี้ใช้ ANOVA F-value (Howard, 2018: 178-184) ดังสมการ

$$F = \left(\frac{1}{(n-1) + (m-1)} \right) (\sum_{i=1}^n (x_i - \bar{x})^2 + \sum_{j=1}^m (y_j - \bar{y})^2) \quad (1)$$

โดย x เป็นคุณลักษณะที่ศึกษา จำนวน n รายการ
 y เป็นค่าเป้าหมายที่คาดหวัง จำนวน m รายการ

สำหรับการวิเคราะห์ปัจจัยด้วยวิธีการ Feature importance โดยใช้ Extra Decision Tree ในการประเมินความสำคัญของปัจจัยที่ส่งผลต่อเป้าหมาย (Mythili & Shanavas, 2014: 63-69) โดยคำนวณจากสมการ

$$Gain(S,A) = Entropy(S) - \sum_{v \in Values(A)} \frac{S_v}{S} Entropy(S_v) \quad (2)$$

$$Entropy(S) = - \sum_{i=1}^c p_i \log_2 p_i \quad (3)$$

โดย $Entropy(S)$ คือ เอนโทรปีของชุดข้อมูล S
โดยที่ c เป็นจำนวนของค่าเป้าหมายที่เป็นไปได้
 p ข้อมูลที่มีค่าเป้าหมายเป็น c
 A เป็นปัจจัยที่ต้องการคำนวณค่าความสำคัญ

6. เครื่องมือที่ใช้ในการวิจัย

ในการวิจัยครั้งนี้ใช้ Pandas เป็นเครื่องมือในการวิเคราะห์ข้อมูลเชิงสำรวจ และจัดรูปแบบข้อมูล เพื่อการสร้างโมเดล และใช้ Scikit-Learn ในการเขียนโปรแกรมเพื่อสร้างโมเดล

7. การวิเคราะห์ข้อมูล

เมื่อนำชุดข้อมูลที่ได้รับการปรับสมดุลด้วยวิธีการ ADASYN เข้ากระบวนการจำแนกด้วย Linear Discriminant Analysis เพื่อค้นหารายวิชาที่ส่งผลให้ผลการเรียนเฉลี่ยตลอดหลักสูตรสูงขึ้น (เฉพาะวิชาแกน และวิชาเอกเฉพาะด้าน) โดยการระบุรายวิชาที่ส่งผลต่อผลการเรียนด้วยวิธี Recursive Feature Elimination พบว่ามีรายวิชาที่ส่งผลต่อผลการเรียนจำนวน 9 รายวิชา ดังแสดงในตาราง 6

ตาราง 6 แสดงรายวิชาที่ส่งผลให้ผลการเรียนเฉลี่ยตลอดหลักสูตรสูงขึ้น คัดเลือกด้วย Recursive Feature Elimination

รหัสวิชา	ชื่อวิชา
4121107	ภาษาอังกฤษเพื่องานด้านคอมพิวเตอร์และเทคโนโลยีสารสนเทศ
4121408	คณิตศาสตร์ดิสครีตและการประยุกต์
4121113	การเขียนโปรแกรมคอมพิวเตอร์เบื้องต้น
4122109	สถิติและการวิจัยเบื้องต้น
4122308	การเขียนโปรแกรมเชิงอ็อบเจกต์
4122511	โครงสร้างข้อมูล
4122512	การออกแบบและวิเคราะห์ขั้นตอนวิธี
4122704	การสื่อสารข้อมูล
4123112	การเขียนโปรแกรมบนเว็บ

สรุปผล

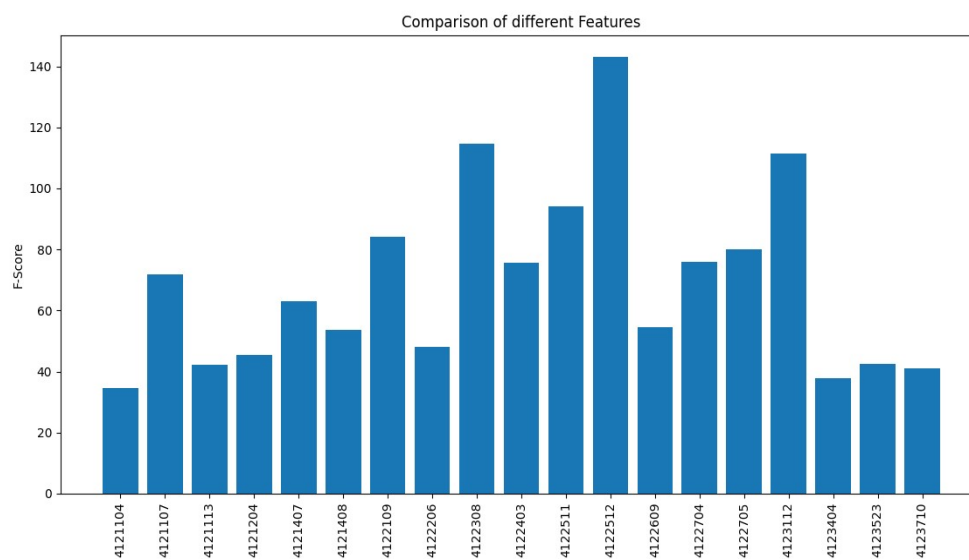
1. ผลค้นหารายวิชาที่ส่งผลต่อผลการเรียนตลอดหลักสูตรของบัณฑิตสาขาวิทยาการคอมพิวเตอร์ มหาวิทยาลัยราชภัฏอุบลราชธานี โดยใช้กระบวนการจำแนกด้วย Linear Discriminant Analysis และค้นหารายวิชาที่ส่งผลต่อผลการเรียนด้วยวิธี Recursive Feature Elimination ซึ่งพบว่ามี 9 รายวิชาที่ส่งผลต่อผลการเรียน คือ วิชาภาษาอังกฤษเพื่องานด้านคอมพิวเตอร์และเทคโนโลยีสารสนเทศ รหัสวิชา 4121107 วิชาคณิตศาสตร์ดิสครีตและการประยุกต์ รหัสวิชา 4121408 วิชาการเขียนโปรแกรมคอมพิวเตอร์เบื้องต้น รหัสวิชา 4121113 วิชาสถิติและการวิจัยเบื้องต้น รหัสวิชา 4122109 วิชาการเขียนโปรแกรมเชิงอ็อบเจกต์ รหัสวิชา 4122308 วิชาโครงสร้างข้อมูล รหัสวิชา 4122511 วิชาการออกแบบและวิเคราะห์ขั้นตอนวิธี รหัสวิชา 4122512 วิชาการสื่อสารข้อมูล รหัสวิชา 4122704 และวิชาการเขียนโปรแกรมบนเว็บ รหัสวิชา 4123112 โดยทำการระบุลำดับของรายวิชาที่มีความสำคัญและส่งผลต่อผลการเรียนด้วยวิธี Univariate Selection และ Feature Importance ซึ่งผลการวิจัยพบว่า ได้ผลการจัดลำดับที่ใกล้เคียงกันโดยรายวิชาการออกแบบและวิเคราะห์ขั้นตอนวิธี เป็นรายวิชาที่มีความสำคัญมากที่สุด

ผลการวิเคราะห์เพื่อจัดลำดับรายวิชาที่ส่งผลต่อผลการเรียนด้วยวิธีการคัดเลือกแบบ Univariate Selection และ Feature Importance แสดงในตาราง 7 และตาราง 8

ตาราง 7 แสดงรายวิชาที่ส่งผลต่อผลการเรียนที่คัดเลือกด้วย Univariate Selection

ลำดับ	รหัสวิชา	ชื่อวิชา	F Score
1	4122512	การออกแบบและวิเคราะห์ขั้นตอนวิธี	138.888
2	4122308	การเขียนโปรแกรมเชิงอ็อบเจกต์	117.899
3	4123112	การเขียนโปรแกรมบนเว็บ	115.318
4	4122511	โครงสร้างข้อมูล	96.612
5	4122109	สถิติและการวิจัยเบื้องต้น	91.667
6	4122705	โครงสร้างและสถาปัตยกรรมคอมพิวเตอร์	83.116
7	4121107	ภาษาอังกฤษเพื่องานด้านคอมพิวเตอร์และเทคโนโลยีสารสนเทศ	77.364
8	4122403	การเขียนโปรแกรมคอมพิวเตอร์กราฟิก	75.740
9	4122704	การสื่อสารข้อมูล	72.712

จากตาราง 7 พบว่ารายวิชาที่ส่งผลต่อผลการเรียนมากที่สุดคือรายวิชา การออกแบบและวิเคราะห์ขั้นตอนวิธี สำหรับรายวิชาอื่น ๆ มีลำดับความสำคัญดังแสดงในภาพประกอบ 2

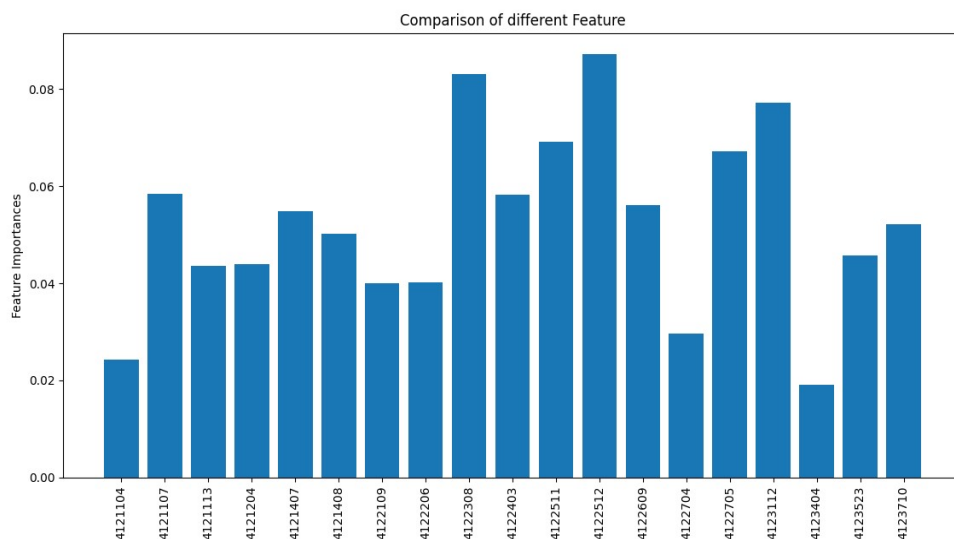


ภาพประกอบ 2 แสดงกราฟรายวิชาที่มีความสำคัญต่อผลการเรียนที่คัดเลือกด้วย Univariate Selection

ตาราง 8 แสดงรายวิชาที่ส่งผลต่อผลการเรียนที่คัดเลือกด้วย Feature Importance

ลำดับ	รหัสวิชา	ชื่อวิชา	ค่าความสำคัญ
1	4122512	การออกแบบและวิเคราะห์ขั้นตอนวิธี	0.087
2	4122308	การเขียนโปรแกรมเชิงอ็อบเจกต์	0.083
3	4123112	การเขียนโปรแกรมบนเว็บ	0.077
4	4122511	โครงสร้างข้อมูล	0.069
5	4122705	โครงสร้างและสถาปัตยกรรมคอมพิวเตอร์	0.067
6	4122403	การเขียนโปรแกรมคอมพิวเตอร์กราฟิก	0.058
6	4121107	ภาษาอังกฤษเพื่องานด้านคอมพิวเตอร์และเทคโนโลยีสารสนเทศ	0.058
8	4122609	การเขียนโปรแกรมภาษาจาวา	0.056
9	4121407	ระบบปฏิบัติการ	0.055

จากตาราง 8 พบว่ารายวิชาที่ส่งผลต่อผลการเรียนมากที่สุดคือรายวิชาการออกแบบและวิเคราะห์ขั้นตอนวิธี สำหรับรายวิชาอื่น ๆ มีลำดับความสำคัญดังแสดงในภาพประกอบ 3



ภาพประกอบ 3 แสดงกราฟรายวิชาที่มีความสำคัญต่อผลการเรียนที่คัดเลือกด้วย Feature Importance

2. จากรายวิชาในตาราง 8 พบว่ารายวิชาที่ส่งผลต่อผลการเรียนรู้ที่ดีของนักศึกษา เป็นรายวิชาที่เกี่ยวข้องกับการเขียนโปรแกรมคอมพิวเตอร์ จำนวน 8 รายวิชา และรายวิชาด้านภาษาอังกฤษ จำนวน 1 รายวิชา

อภิปรายผล

1. ผลการวิจัยการค้นหารายวิชาที่ส่งผลต่อผลการเรียนตลอดหลักสูตรของบัณฑิตสาขาวิทยาการคอมพิวเตอร์ มหาวิทยาลัยราชภัฏอุบลราชธานี โดยใช้ชุดข้อมูลที่ไม่สมดุลทำการคัดเลือกและปรับสมดุลข้อมูลด้วยวิธีการ ADASYN ทำการคัดเลือกรายวิชาด้วย Recursive Feature Elimination ทำการระบุลำดับของรายวิชาที่มีความสำคัญและส่งผลต่อผลการเรียนด้วยวิธี Univariate Selection และ Feature Importance ซึ่งผลการวิจัยพบว่า รายวิชาการออกแบบและวิเคราะห์ขั้นตอนวิธี รายวิชาการเขียนโปรแกรมเชิงอ็อบเจกต์ และรายวิชาโครงสร้างข้อมูล เป็นรายวิชาอันดับต้น ๆ ที่ส่งผลให้ผลการเรียนของนักศึกษาสาขาวิทยาการคอมพิวเตอร์ มหาวิทยาลัยราชภัฏอุบลราชธานีสูงขึ้น ซึ่งสอดคล้องกับ Jaidee and Wannapee (2020: 75-84) ศึกษาปัจจัยที่ส่งผลต่อการสำเร็จการศึกษาตามแผนการเรียนของนักศึกษาระดับปริญญาตรี โดยใช้เทคนิคการคัดเลือกคุณลักษณะบนชุดข้อมูลที่ไม่สมดุล ในการวิจัยดังกล่าวใช้เทคนิคการสุ่มเพิ่มตัวอย่างข้อมูลกลุ่มน้อย (Synthetic Minority Over-sampling Technique : SMOTE) (Zheng, Cai and Li, 2015: 1017-1037) ร่วมกับวิธีการคัดเลือกคุณลักษณะด้วยเทคนิค Chi-Square (Jaidee, 2019: 325-335), Information Gain (Rawengwan, 2017: 273-283) และเทคนิค Correlation based feature selection ทำให้ได้ปัจจัยที่ส่งผลต่อการสำเร็จการศึกษาของนักศึกษาได้อย่างถูกต้องและมีประสิทธิภาพ ผลการวิจัยดังกล่าวพบว่า มีสองรายวิชาที่ส่งผลต่อการสำเร็จการศึกษาตามแผนการเรียนของนักศึกษา คือ รายวิชาการเขียนโปรแกรมคอมพิวเตอร์ 1 และรายวิชาการเขียนโปรแกรมบนเว็บ

ดังนั้นจากผลการวิจัยเรื่องการวิเคราะห์ปัจจัยที่ส่งผลต่อผลการเรียนของบัณฑิตสาขาวิทยาการคอมพิวเตอร์ โดยใช้ชุดข้อมูลที่ไม่สมดุล และการศึกษาปัจจัยที่ส่งผลต่อการสำเร็จการศึกษาตามแผนของนักศึกษาระดับปริญญาตรี โดยใช้เทคนิคการคัดเลือกคุณลักษณะบนชุดข้อมูลที่ไม่สมดุล พบว่ารายวิชาการออกแบบและวิเคราะห์ขั้นตอนวิธี วิชาการเขียนโปรแกรมเชิง อ็อบเจกต์ รายวิชาโครงสร้างข้อมูล หรือรายวิชาการเขียนโปรแกรมคอมพิวเตอร์ 1 รายวิชาการเขียนโปรแกรมบนเว็บ ทั้งห้ารายวิชาเป็นรายวิชาที่เกี่ยวข้องกับการเขียนโปรแกรมคอมพิวเตอร์ ซึ่งเป็นรายวิชาที่พัฒนาทักษะความคิดเชิงระบบของนักศึกษา มีผลทำให้การเรียนในรายวิชาอื่น ๆ ในหลักสูตรมีผลสัมฤทธิ์การเรียนที่ดีขึ้นด้วย ดังนั้นจึงควรที่จะปรับปรุงเนื้อหาหรือเพิ่มรายวิชาที่เกี่ยวข้องกับการเขียนโปรแกรมคอมพิวเตอร์ รวมทั้งปรับกระบวนการเรียนการสอนในรายวิชาดังกล่าวให้ดียิ่งขึ้น นอกจากนี้อาจต้องปรับหรือจัดให้ผู้เรียนได้เรียนในรายวิชาเหล่านั้นในช่วงปีแรก ๆ เพื่อให้ผู้เรียนมีพื้นฐานในการเรียนรายวิชาอื่น ๆ ที่เกี่ยวข้อง

2. แนวทางการปรับปรุงหลักสูตรและกระบวนการจัดการเรียนการสอนในรายวิชาที่ส่งผลต่อผลการเรียน สำหรับหลักสูตรวิทยาการคอมพิวเตอร์ควรเน้นสมรรถนะ ด้านการเขียนโปรแกรมคอมพิวเตอร์สำหรับนักศึกษาให้มากขึ้น เนื่องจากทักษะด้านโปรแกรมคอมพิวเตอร์สามารถนำไปเป็นพื้นฐานการเรียนรู้ในรายวิชาอื่น ๆ ได้ เพราะการสร้างแนวคิดแบบขั้นตอนวิธีจะสามารถนำไปแก้ปัญหาในเรื่องอื่น ๆ ได้เป็นอย่างดี ในขณะที่เดียวกันก็ควรจัดรายวิชาภาษาอังกฤษ เพิ่มให้กับนักศึกษา เพราะหากมีทักษะด้านภาษาอังกฤษที่ดี ย่อมส่งผลให้มีทักษะการเขียนโปรแกรมคอมพิวเตอร์เร็วขึ้น สามารถเข้าใจคำสั่งเขียนคำสั่งได้ และสามารถเข้าถึงแหล่งสืบค้นที่เป็นภาษาอังกฤษได้มากขึ้น

ในประเด็นการปรับสมดุลข้อมูลโดยใช้เทคนิคการสังเคราะห์ข้อมูลเพิ่มแบบ ADASYN สามารถเพิ่มประสิทธิภาพและความแม่นยำในการทำงานของอัลกอริทึมแบบ Linear Discriminant Analysis เพื่อค้นหารายวิชาที่ส่งผลต่อผลการเรียน โดยมีระดับความแม่นยำก่อนการปรับสมดุล และหลังการปรับสมดุล เป็น 0.743589 และ 0.886550 ตามลำดับ และพบว่า รายวิชาการออกแบบและวิเคราะห์ขั้นตอนวิธี รายวิชาการเขียนโปรแกรมเชิงอ็อบเจกต์ และรายวิชาโครงสร้างข้อมูลเป็นรายวิชาอันดับต้น ๆ ที่ส่งผลให้ผลการเรียนเฉลี่ยของนักศึกษาสาขาวิทยาการคอมพิวเตอร์ มหาวิทยาลัยราชภัฏอุบลราชธานีสูงขึ้น ดังนั้น หากผู้เรียนมีความรู้ ความเข้าใจ และทักษะที่สำคัญในรายวิชาดังกล่าว จะทำให้การเรียนในรายวิชาอื่น ๆ ในหลักสูตรมีผลสัมฤทธิ์ที่ดีขึ้น

ข้อเสนอแนะ

1. ข้อเสนอแนะในการนำไปใช้

ปัจจัยที่สำคัญที่จะส่งผลถึงความแม่นยำของการวิจัยครั้งนี้ คือการปรับสมดุลชุดข้อมูล และการเลือกใช้เทคนิคการจำแนกข้อมูลที่เหมาะสม โดยในการวิจัยนี้เลือกใช้ ADASYN ในการปรับสมดุลชุดข้อมูล และใช้ Linear Discriminant Analysis ในการจำแนกข้อมูล หากมีการวิจัยที่ใช้เทคนิคการปรับสมดุลชุดข้อมูล และการจำแนกข้อมูลอื่น ๆ เพื่อเปรียบเทียบประสิทธิภาพของโมเดล จะทำให้สามารถเลือกเทคนิคที่เหมาะสมกับชุดข้อมูลได้ดียิ่งขึ้น

2. ข้อเสนอแนะในการทำวิจัยครั้งต่อไป

ประเด็นของโครงสร้างและรายวิชาในหลักสูตร การจัดการเรียนและลำดับการลงทะเบียนเรียนของรายวิชา ก็เป็นประเด็นที่น่าสนใจเป็นอย่างยิ่ง โดยในการวิจัยต่อไปจะทำการค้นหาลำดับของรายวิชาที่ส่งให้ผู้เรียนมีผลต่อผลการเรียนที่ดีขึ้น ซึ่งจะเป็นประโยชน์อย่างยิ่งต่อการปรับปรุงและพัฒนาหลักสูตรวิทยาการคอมพิวเตอร์ มหาวิทยาลัยราชภัฏอุบลราชธานี

เอกสารอ้างอิง

- Alyahyan, E. and Dustegor, D. (2020). Predicting Academic success in higher education: literature review and best practices. *International Journal of Educational Technology in Higher Education*, 17(1), 1-22.
- Ashfaq, U., Booma, P. M. and Mafas, R. (2020). Managing Student Performance: A Predictive Analytics using Imbalanced Data. *International Journal of Recent Technology and Engineering (IJRTE)*, 8(6), 2277-2283.
- Asif, R. (2015). Predicting Student Academic Performance at Degree Level: A Case Study. *International Journal of Intelligent Systems and Applications*, 01, 49-61.
- Howard, J. S. (2018). *Experimental Design and Analysis*. Department of Statistics and Data Science, Carnegie Mellon University.
- Jaidee, W. and Wannapee, N. (2020). The Study of Factors Affecting for On-time Graduation of Ungraduated Student Using Feature Selection Technique on Imbalanced Datasets. *Journal of Information Science and Technology*, 10(1), 75-84.
- Jaidee, W. (2019). A Predictive Model for On-time Graduation Using SMOTE Technique and Ensemble Model on Imbalanced Datasets for Graduation Data. *Proceeding of 12th RMUTTO Research Conference*, 325-335.
- Maheshwari, S., Agrawal, J. and Sharma, S. (2011). A New approach for Classification of Highly Imbalanced Datasets using Evolutionary Algorithms. *International Journal of Scientific & Engineering Research*, 2(7), 1-5.
- Mythili, M.S. and Shanavas, A.R.M. (2014). An Analysis of students' performance using classification algorithms. *IOSR Journal of Computer Engineering (IOSR-JCE)*, 16(1), 63-69.
- Rawengwan, P. (2017). A model for forecasting educational status of students. *Proceeding of the National and International Graduate Research Conference*, Khon Kaen University, 273-283.

- Sales, A. Balby, L. and Cajueiro, A. (2016). Exploiting Academic Records for Predicting Student Drop Out: a case study in Brazilian higher education. *Journal of Information and Data Management*, 7(2), 166–180.
- Vikas, C. Raykar, B. and Saha, A. (2015). Data Split Strategies for Evolving Predictive Models. *European Conference, ECML PKDD 2015, Porto, Portugal, September 7-11, 2015, Proceedings, Part I*, 3-19.
- Zheng, Z., Cai, Y. and Li, Y. (2015). Oversampling Method for Imbalanced Classification. *Computing and Informatics*, 34, 1017-1037.