การพัฒนาระบบแนะนำการเลือกวิชาเอกของสาขาเทคโนโลยีสารสนเทศและการสื่อสาร โดยใช้เทคนิคต้นไม้ตัดสินใจ

Development of the recommendation system for the selection of major in Information and Communication Technology using decision tree techniques

วงกต ศรีอุไร[1*], อนุสรณ์ บรรเทิง[2] และ คมกฤษณ์ มุธาพร[3]
Wongkot Sriurai[1*], Anusorn Bunteong [2] and Komkrit Mutaporn[3]

Received: 8 September 2023 ; **Revised**: 21 December 2023 ; **Accepted**: 29 January 2024

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อ 1) พัฒนาแบบจำลองแนะนำการเลือกวิชาเอกของสาขาวิชาเทคโนโลยีสารสนเทศและการสื่อสาร 2) เปรียบเทียบประสิทธิภาพแบบจำลองด้วยวิธีต้นไม้ตัดสินใจโดยใช้ อัลกอริทึม Iterative Dichotomiser 3 อัลกอริทึม C4.5 และ Classification and Regression Trees (CART) และ 3) ศึกษาความพึงพอใจของผู้ใช้ที่มีต่อระบบที่พัฒนาขึ้น ข้อมูลที่นำมาสร้างแบบจำลองประกอบด้วย ข้อมูลผลการเรียน 10 รายวิชา ได้แก่ คอมพิวเตอร์เบื้องต้นและองค์ประกอบคอมพิวเตอร์ การเขียนโปรแกรมเบื้องต้น สถิติและการวิเคราะห์เชิงปริมาณ โครงสร้างข้อมูลและขั้นตอนวิธี การเขียนโปรแกรมเชิงอ็อบเจ็กต์ คณิตศาสตร์เต็มหน่วย ระบบฐานข้อมูล การเขียนโปรแกรมบนเว็บ การออกแบบประสบการณ์ผู้ใช้และส่วนติดต่อผู้ใช้งาน และการสื่อสารข้อมูลและเครือข่ายอินเทอร์เน็ต ข้อมูลทั้งหมดถูกนำมาปรับสมดุลด้วยวิธี SMOTE แล้วจึงนำมาพัฒนาด้วย อัลกอริทึม Iterative Dichotomiser 3, C4.5 และ Classification and Regression Trees (CART) ระบบนี้ถูกพัฒนาในรูปแบบเว็บแอปพลิเคชันด้วยภาษาพีเอชพีร่วมกับฐานข้อมูลมายเอสคิวแอล ผลการทดสอบประสิทธิภาพแบบจำลองด้วยวิธี 10-fold cross validation พบว่าแบบจำลองที่พัฒนาด้วยอัลกอริทึม C4.5 ให้ประสิทธิภาพดีที่สุด โดยมีค่าความถูกต้องเท่ากับ 93.20 % ค่าความแม่นยำเท่ากับ 93.33% และค่าระลึกเท่ากับ 93.32% ผลการประเมินความพึงพอใจจากผู้ใช้ระบบจำนวน 30 คนโดยใช้แบบสอบถาม พบว่าค่าเฉลี่ยความพึงพอใจอยู่ในระดับมาก ($\bar{X}$=3.66, SD.=0.93)

คำสำคัญ: ต้นไม้ตัดสินใจ, อัลกอริทึม Iterative Dichotomiser 3, อัลกอริทึม C4.5, Classification and Regression Trees (CART)

Abstract

The objectives of this research were: 1) to develop a recommendation model for choosing a major in Information and Communication Technology; 2) to compare the performance of decision tree methods such as Iterative Dichotomiser 3, C4.5, and Classification and Regression Trees (CART); and 3) to study the users' satisfaction with the developed system. The data used to develop this model included academic results from 10 subjects: Introduction to Computer and Computer Architecture, Introduction to Programming, Statistics and Quantitative Analysis, Data Structure and Algorithms, Object-Oriented Programming, Discrete Mathematics, Database System, Web Programming, User Experience/User Interface Design, and Data Communication and Internetworking. Data imbalance was addressed using the SMOTE method. After that, the models were constructed using Iterative Dichotomiser 3, C4.5, and CART. The system was developed as a web application using PHP with a MySQL database. The evaluation results given by 10-fold cross-validation, showed that a recommendation model for choosing a major in Information and Communication Technology developed by C4.5 provided the highest level of effectiveness with an accuracy of 93.20%, precision of 93.33%, and recall of 93.32%. The user satisfaction assessment with the proposed system was collected from 30 users through questionnaires. The results indicated that the users' satisfaction was at a high level ($\bar{X}$=3.66, SD.=0.93).

Keywords: decision tree, Iterative Dichotomiser 3, C4.5, Classification and Regression Trees (CART)

[1,2,3] ภาควิชาคณิตศาสตร์ สถิติ และคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยอุบลราชธานี อุบลราชธานี 34190
[1,2,3] Department of Mathematics Statistics and Computer, Faculty of Science, Ubon Ratchathani University, Ubon Ratchathani, 34190
[*] Corresponding author: Email wongkot.s@ubu.ac.th

บทนำ

หลักสูตรวิทยาศาตรบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศและการสื่อสาร คณะวิทยาศาสตร์ มหาวิทยาลัยอุบลราชธานี เป็นหลักสูตรที่มุ่งผลิตบัณฑิตที่มีความรู้ ความสามารถ และทักษะในงานทางด้านคอมพิวเตอร์ และการสร้างนวัตกรรมจากข้อมูล โดยหลักสูตรประกอบไปด้วย 3 มิติหลัก ได้แก่ 1) มิติด้านความรู้พื้นฐาน ซึ่งจะประกอบด้วย คณิตศาสตร์ สถิติและคอมพิวเตอร์ 2) มิติด้านความรู้ทางเทคโนโลยีสารสนเทศและการสื่อสาร ซึ่งจะเน้นความรู้หลักทั่วไปของงานด้านเทคโนโลยีสารสนเทศและการสื่อสาร 3) มิติเฉพาะตามวิชาเอกที่สนใจ ซึ่งมี 3 สาขาวิชาเอก ได้แก่ 1) วิชาเอกเทคโนโลยีเกมและการผลิตดิจิทัลคอนเทนท์ เรียนรู้เกี่ยวกับเทคโนโลยีและการพัฒนาเกมแบบ 2 มิติ 3 มิติ การออกแบบและพัฒนาสื่อและคอนเทนท์อย่างสร้างสรรค์ 2) วิชาเอกเทคโนโลยีเครือข่ายและความมั่นคงทางไซเบอร์ เรียนรู้เกี่ยวกับการจัดการการออกแบบระบบเครือข่าย การสร้างความปลอดภัยและความมั่นคงทางไซเบอร์ และ 3) วิชาเอกเทคโนโลยีดิจิทัล เรียนรู้เกี่ยวกับการวิเคราะห์ ออกแบบ และพัฒนาระบบสารสนเทศบนแพล็ตฟอร์มต่าง ๆ

ในเบื้องต้นทางหลักสูตรให้นักศึกษาสามารถเลือกสาขาวิชาเอกตามความสนใจและความถนัดของนักศึกษา จากนั้นทางหลักสูตรจะมีคณะกรรมการมาพิจารณาสาขาวิชาเอกหรือแขนงที่นักศึกษาเลือกตามความเหมาะสมอีกครั้ง เนื่องจากที่ผ่านมาทางหลักสูตรประสบปัญหามีนักศึกษาบางรายเลือกสาขาวิชาเอกตามที่ตนเองสนใจ แต่ไม่ได้พิจารณาถึงความถนัดหรือความสามารถของตนเอง ทำให้คณะกรรมการจะต้องพิจารณาเลือกวิชาเอกให้นักศึกษาใหม่ เพื่อให้นักศึกษาสามารถเรียนและสำเร็จการศึกษาตามเป้าหมายได้ หากนักศึกษาสามารถเลือกสาขาวิชาเอกได้สอดคล้องกับความถนัดและความสามารถนักศึกษา จะส่งเสริมให้นักศึกษาสามารถแสดงศักยภาพได้อย่างเต็มที่ และนำไปสู่การทำงานในอนาคต

ปัจจุบันได้มีการนำความรู้เกี่ยวกับเทคนิคการทำเหมืองข้อมูล (Data Mining) มาใช้ในการจำแนกและการพยากรณ์ในงานด้านต่าง ๆ ไม่ว่าจะเป็น ด้านธุรกิจที่ช่วยในการตัดสินใจของผู้บริหาร ด้านการแพทย์ เพื่อจำแนกผู้ป่วยโรคต่าง ๆ และด้านการศึกษา มีการนำเทคนิคเหมืองข้อมูลมาใช้ในการทำนายการพ้นสภาพของนักศึกษา (Viloria *et al*., 2018) เป็นต้น และเทคนิคต้นไม้ตัดสินใจ (Decision Tree) ก็เป็นเทคนิคหนึ่งของการทำเหมืองข้อมูลด้านการจำแนกและการพยากรณ์ที่ได้รับความนิยมจากนักวิจัยหลาย ๆ ท่าน เช่น งานวิจัยของรัตนาวดี พานทอง (2565) ได้พัฒนาแบบจำลองจำแนกประเภทข้อมูลการท่องเที่ยวของนักท่องเที่ยวในจังหวัดพะเยาโดยการ ใช้เทคนิคต้นไม้ตัดสินใจ ผลการวิจัยพบว่าแบบจำลองมีค่าความถูกต้องในการจำแนกข้อมูลเท่ากับ 81.70% ในขณะที่งานวิจัยของธนพร คล้ายทอง และชุติพนธ์ ศรีสวัสดิ์ (2566) ได้พัฒนาแบบจำลองการพยากรณ์การตกออกของนักศึกษาด้วยเทคนิคเหมืองข้อมูล โดยเปรียบเทียบ 3 เทคนิค คือ เทคนิคต้นไม้ตัดสินใจ (Decision Tree) เทคนิคนาอีฟเบย์ (Naïve Bayes) และกฎการอุปนัย (Rule Induction) ผลการวิจัยพบว่าเทคนิคต้นไม้ตัดสินใจให้ประสิทธิภาพในการพยากรณ์ดีที่สุด โดยมีค่าความถูกต้องเท่ากับ 86.29% เป็นต้น

จากปัญหาการเลือกสาขาวิชาเอกและจากการศึกษางานวิจัยที่เกี่ยวข้องกับเทคนิคเหมืองข้อมูล ทำให้คณะผู้วิจัยมีแนวคิดที่จะพัฒนาระบบแนะนำการเลือกวิชาเอกของสาขาวิชาเทคโนโลยีสารสนเทศและการสื่อสาร โดยการนำข้อมูลของนักศึกษามาวิเคราะห์ด้วยเทคนิคเหมืองข้อมูล เพื่อสร้างแบบจำลองที่ช่วยในการตัดสินใจเลือกวิชาเอกของสาขาวิชาเทคโนโลยีสารสนเทศและการสื่อสาร โดยผลของแบบจำลองนี้สามารถนำไปประยุกต์ใช้เพื่อสร้างระบบแนะนำการเลือกวิชาเอกได้ ช่วยให้นักศึกษามีข้อมูลพื้นฐานเบื้องต้นว่านักศึกษาน่าจะเหมาะกับวิชาเอกไหน และช่วยให้นักศึกษาสามารถเลือกวิชาเอกที่เหมาะสมกับตนเองมากที่สุด

วัตถุประสงค์การวิจัย

1. พัฒนาแบบจำลองเพื่อแนะนำการเลือกวิชาเอกของสาขาวิชาเทคโนโลยีสารสนเทศและการสื่อสาร

2. เปรียบเทียบประสิทธิภาพของแบบจำลองที่พัฒนาด้วยเทคนิคต้นไม้ตัดสินใจโดยใช้ อัลกอริทึม Iterative Dichotomiser 3 อัลกอริทึม C4.5 และ Classification and Regression Trees (CART)

3. ศึกษาความพึงพอใจของผู้ใช้ระบบแนะนำการเลือกสาขาวิชาเอกของสาขาวิชาเทคโนโลยีสารสนเทศและการสื่อสาร

การทำเหมืองข้อมูล

การทำเหมืองข้อมูล คือ กระบวนการค้นหาสารสนเทศหรือองค์ความรู้ที่อยู่ในฐานข้อมูลขนาดใหญ่ที่ซับซ้อน เพื่อนำความรู้ที่ได้ไปใช้ประโยชน์ในการตัดสินใจ สารสนเทศที่ได้อาจนำมาสร้างการพยากรณ์หรือสร้างตัวแบบสำหรับการจำแนกหรือแสดงความสัมพันธ์ระหว่างข้อมูลต่าง ๆ หรือให้ข้อสรุปของสาระในฐานข้อมูล การทำเหมืองข้อมูลประกอบขึ้นด้วยการนำกระบวนการทางสถิติและการเรียนรู้ผ่านระบบคอมพิวเตอร์เพื่อสร้างตัวแบบ กฎเกณฑ์ รูปแบบ การพยากรณ์ จากฐานข้อมูลขนาดใหญ่ โดยการทำเหมืองข้อมูลมีขั้นตอนการดำเนินงานหลายขั้นตอนซึ่งต้องอาศัยเทคนิคหรือวิธีการต่าง ๆ เช่น วิธีการจัดกลุ่ม การค้นหาความสัมพันธ์ การพยากรณ์ เป็นต้น (Han *et al*., 2012)

เทคนิคต้นไม้ตัดสินใจ

ต้นไม้ตัดสินใจ (Decision Tree) เป็นการนำข้อมูลมาสร้างแบบจำลองการพยากรณ์ในรูปของต้นไม้ (Tree) ซึ่งต้นไม้ตัดสินใจมีการทำงานแบบมีผู้สอน (Supervised Learning) คือ สามารถสร้างแบบจำลองจากกลุ่มตัวอย่างของข้อมูลที่ได้กำหนดไว้ก่อนล่วงหน้า หรือที่เรียกว่า ข้อมูลสำหรับใช้ในการฝึกฝน (Training dataset) และสามารถพยากรณ์กลุ่มของรายการข้อมูลที่ยังไม่เคยนำมาจัดหมวดหมู่ได้ โดยรูปแบบของต้นไม้ตัดสินใจจะประกอบโหนด (Node) แสดงคุณลักษณะ (Attribute) ที่ใช้ทดสอบข้อมูล แต่ละกิ่งแสดงผลในการทดสอบและลีฟโหนด (Leaf Node) แสดงกลุ่มหรือคลาส (Class) ที่กำหนดไว้ ซึ่งต้นไม้ตัดสินใจนี้ง่ายต่อการเข้าใจและปรับเปลี่ยนเป็นกฎการจำแนก (Classification Rules) (Han *et al*., 2012), (Sharma and Kumar, 2016) ต้นไม้ตัดสินใจ (Decision tree) มีหลากหลายวิธี ซึ่งในงานวิจัยนี้ได้ทำการเปรียบเทียบการสร้างแบบจำลองด้วยเทคนิคต้นไม้ตัดสินใจด้วย 3 วิธี ได้แก่

1. Iterative Dichotomiser 3 (ID3 Algorithm) เป็นอัลกอริทึมที่ใช้สร้างต้นไม้ตัดสินใจที่ใช้หลักการของทฤษฎีสารสนเทศ (Information Theory) โดยจะใช้ค่ามาตรฐานเกน (Gain) ในการวัดว่าจะแบ่งข้อมูลอย่างไร และค่าที่วัดได้จะนำมาใช้ตัดสินใจว่าจะใช้ตัวแปรใดในการแบ่งข้อมูล โดยวิธีการกำหนดโครงสร้างต้นไม้ตัดสินใจจะเป็นการเลือกข้อมูลตามลำดับของตัวชี้วัดหรือค่าเกน (Gain) สูงที่สุดเป็นข้อมูลเริ่มต้น และข้อมูลถัดไปที่มีค่าลดหลั่นกันตามลำดับ จากทฤษฎีสารสนเทศ ซึ่งค่าสารสนเทศของข้อมูลจะขึ้นอยู่กับความน่าจะเป็นของข้อมูล สามารถเขียนในรูปสมการที่ 1 (Ogheneovo and Nlerum, 2020) ได้ดังนี้

$$Info(D) = -\sum_{i=1}^{m} p_i \log_2(p_i) \quad (1)$$

โดยที่

P_i เป็นความน่าจะเป็นที่ข้อมูลในฐานข้อมูล D อยู่ในกลุ่ม C_i ซึ่งมีค่า $|C_{i,D}|/|D|$

m เป็นจำนวนกลุ่มทั้งหมดที่ต่างกันของข้อมูลชุดนั้น

C_i เป็นกลุ่มในลำดับที่ i โดยที่ i มีค่าระหว่าง *1* ถึง m

$|C_{i,D}|$ เป็นจำนวนข้อมูลในฐานข้อมูล D ที่อยู่ในกลุ่ม C_i

$|D|$ เป็นจำนวนข้อมูลในฐานข้อมูล D

ค่า $Info(D)$ ที่ได้นี้สามารถเรียกอีกชื่อหนึ่งว่าค่า Entropy ของฐานข้อมูล D ค่าความรู้จากทฤษฎีสารสนเทศจะช่วยในการแยกแยะข้อมูลทำให้ลดจำนวนครั้งของการทดสอบได้ อีกทั้งยังรับประกันว่าต้นไม้ตัดสินใจที่ได้จะไม่มีความซับซ้อนมากจนเกินไป เมื่อทำการพิจารณาเลือกคุณลักษณะเป็นตัวเลือกทดสอบ จะใช้ค่าความรู้จากทฤษฎีสารสนเทศของคุณลักษณะ สามารถเขียนในรูปสมการที่ 2 (Ogheneovo and Nlerum, 2020) ได้ดังนี้

$$Info_A = \sum_{j=1}^{k} \frac{|D_j|}{|D|} x Info(D_j) \quad (2)$$

โดยที่

k เป็นจำนวนค่าที่เป็นไปได้ของคุณลักษณะ

$|D|$ เป็นจำนวนข้อมูลในฐานข้อมูล D

$|D_j|$ เป็นจำนวนข้อมูลในฐานข้อมูล D ที่มีค่าที่ j ของคุณลักษณะ A

ค่ามาตรฐานเกนที่จะใช้พิจารณาเลือกคุณลักษณะ A มาเป็นโหนดของต้นไม้มีค่าเท่ากับผลต่างของความรู้จากทฤษฎีสารสนเทศ กับความรู้จากทฤษฎีสารสนเทศของคุณลักษณะ สามารถเขียนสมการที่ 3 (Ogheneovo and Nlerum, 2020) ได้ดังนี้

$$Gain(A) = Info(D) - Info_A(D) \quad (3)$$

2. อัลกอริทึม C4.5 เป็นอัลกอริทึมที่ใช้สร้างต้นไม้ตัดสินใจที่พัฒนาโดย John, R. Q. (1993) โดยมีพื้นฐานเดียวกับอัลกอริทึม ID3 โดยขั้นตอนวิธีที่ใช้สร้างต้นไม้ตัดสินใจสำหรับแบ่งกลุ่มข้อมูลจะใช้หลักการของ Information gain เช่นเดียวกับ ID3 แต่จะมีส่วนเพิ่มเติมจาก ID3 เข้ามา โดยอัลกอริทึม C4.5 ได้เพิ่มการใช้ค่ามาตรฐานอัตราส่วนเกน (Gain Ratio criterion) ในการตัดสินใจเลือกคุณลักษณะที่จะใช้เป็นโหนดรากหรือโหนดระดับต่าง ๆ บนต้นไม้ ซึ่งอัลกอริทึม C4.5 เข้ามาแก้ไขจุดด้อยของอัลกอริทึม ID3 (Han *et al*., 2012), (Sharma and Kumar, 2016) ได้ดังนี้

(1) สามารถใช้งานได้ทั้งข้อมูลแบบต่อเนื่อง (Continuous data) และแบบไม่ต่อเนื่อง (Discrete data)

(2) สามารถใช้กับชุดข้อมูลทดสอบ ที่มีค่าข้อมูลขาดหายได้ (Missing data) โดยจะแทนค่าด้วย “?” และไม่นำค่านั้นมาคำนวณในกฎของความรู้จากทฤษฎีสารสนเทศ

(3) สามารถใช้กับชุดข้อมูลทดสอบที่มีค่าผิดปกติหรือมีความเสียหายได้

(4) สามารถทำการตัดกิ่งต้นไม้ตัดสินใจในขณะสร้างได้ โดยไม่ทำให้ความถูกต้องลดลง

การเลือกคุณลักษณะที่ใช้เป็นโหนดรากหรือโหนดบนต้นไม้ตัดสินใจนั้นอัลกอริทึม ID3 จะใช้ค่าเกน (Gain) เป็นหลักในการเลือก แต่อัลกอริทึม C4.5 นั้นได้เพิ่มการใช้ค่ามาตรฐานอัตราส่วนเกน (Gain ratio criterion) ในการตัดสินใจเลือกคุณลักษณะ เนื่องจากค่าเกนจะมีการเอนเอียง (Bias) อย่างมากกับข้อมูลที่ประกอบด้วยคุณลักษณะที่มีค่าที่เป็นไปได้จำนวนมาก ๆ การแก้ไขความเอนเอียงของค่าเกนสามารถทำได้โดยการปรับค่ามาตรฐานเกนให้ถูกต้องโดยใช้ค่าสารสนเทศของการแบ่งแยก (Split information) ของคุณลักษณะแต่ละตัว สามารถเขียนสมการได้ดังนี้

$$SplitInfo(A) = -\sum_{j=1}^{k} \frac{|D_j|}{|D|} x \log_2\left(\frac{|D_j|}{|D|}\right) \quad (4)$$

ค่าสารสนเทศของการแบ่งแยก (Split information) นี้จะแสดงถึงระดับการกระจายของข้อมูล เมื่อนำค่านี้ไปหารค่าเกนจะได้ค่ามาตรฐานอัตราส่วนเกน สามารถเขียนในรูปสมการได้ดังนี้

$$GainRatio(D) = \frac{Gain(A)}{SplitInfo(A)} \quad (5)$$

ค่ามาตรฐานอัตราส่วนเกนช่วยแก้ไขความเอนเอียงของค่าเกนได้ โดยทำให้ค่ามาตรฐานอัตราส่วนเกนในแบ่งด้วยคุณลักษณะที่มีการกระจายสูงถูกปรับลดลง

3. Classification and Regression Trees (CART) ต้นไม้ตัดสินใจที่สร้างจากอัลกอริทึม CART นี้ เป็นการสร้างต้นไม้แบบ Binary ซึ่งประกอบด้วย กิ่งหรือแขนง 2 กิ่ง สำหรับแต่ละโหนด โดยอัลกอริทึม CART จะใช้ค่าดัชนีจีนี (Gini index) ในการวัดว่าจะแบ่งข้อมูลอย่างไร โดยค่า Gini index เป็นการประเมินค่าที่บ่งบอกว่าลักษณะใดเหมาะสมที่จะเป็นลักษณะสำคัญโดยวัดค่าจากแต่ละลักษณะ แล้วทำการเปรียบเทียบกับลักษณะอื่น ๆ เพื่อหาลักษณะที่มีค่า Gini index ที่น้อยที่สุดเป็นลักษณะสำคัญ ซึ่งวิธีการคำนวณ (Han *et al*., 2012), (Sharma and Kumar, 2016) ดังนี้

$$Gini(t) = 1 - \sum_j \left[p(j|t)\right]^2 \quad (6)$$

โดยที่

$p(j|t)$ คือ ค่าความถี่สัมพัทธ์ของคลาส j ในโหนด t

จากสมการที่ 6 เป็นการคำนวณค่า Gini index ของแต่ละโหนด ซึ่งในการสร้างต้นไม้ตัดสินใจโดยใช้ค่า Gini index นั้นจำเป็นจะต้องคำนวณค่า Gini index ของผลรวมของโหนดนั้นด้วย โดยเป็นการคำนวณค่าเฉลี่ยถ่วงน้ำหนัก (Weighted average) ของแต่ละโหนด ดังสมการที่ 7

$$Gini_{split} = \sum_{i=1}^{k} \frac{n_i}{n} Gini(i) \quad (7)$$

โดยที่

n_i = จำนวนของเรคคอร์ด (Record) ที่โหนด child i

n = จำนวนของเรคคอร์ด (Record) ที่โหนด P

เทคนิคการสุ่มเพิ่มตัวอย่างกลุ่มน้อย

Synthetic Minority Oversampling Technique (SMOTE) เป็นเทคนิคที่เข้ามาช่วยแก้ปัญหาการจำแนกข้อมูลที่ไม่สมดุล จากข้อมูลในแต่ละคลาสนั้นมีจำนวนที่แตกต่างกันมาก ทำให้ผลลัพธ์ในการจำแนกจะตกไปอยู่ข้อมูลที่มีกลุ่มมาก ดังนั้นการใช้เทคนิค SMOTE จึงเป็นอีกวิธีเพื่อเพิ่มจำนวนของข้อมูลคลาสน้อย ให้มีจำนวนคลาสมากขึ้น อาศัยหลักการกระจายข้อมูล ข้อมูลที่ได้เกิดความสมดุลมากขึ้น หลักการสุ่มค่าข้อมูลที่อยู่ในกลุ่มข้อมูลคลาสน้อยขึ้นมา 1 ค่า แล้วพิจารณาค่าข้อมูลที่อยู่ใกล้เคียง จำนวน K ค่า จากนั้นคำนวณค่าระยะทาง ของระหว่างค่าที่สุ่มและค่าข้อมูลที่อยู่ใกล้เคียงกันเพื่อได้ค่าระยะทางน้อยที่สุด (Chawla *et al*., 2002)

การวัดประสิทธิภาพแบบจำลอง

การวัดประสิทธิภาพของแบบจำลองจะพิจารณาจากค่าในตาราง Confusion Matrix โดยในตารางนี้จะเก็บข้อมูลจำนวนแถวที่จำแนกจากกลุ่มข้อมูลจริงและกลุ่มข้อมูลจากการทำนาย ซึ่งตารางนี้จะมีขนาด m x m โดยที่ m คือจำนวนของกลุ่ม ดัง Figure 1

		Predicted Class	
		Class=Yes	Class=No
Actual Class	Class=Yes	TP	FN
	Class=No	FP	TN

Figure 1 Confusion Matrix

อธิบายค่าต่าง ๆ ของ Confusion Matrix ได้ดังนี้

True positive (TP) หมายถึง ค่าคลาสเป้าหมาย คือ Yes และโมเดลทำนายว่า Yes

False negative (FN) หมายถึง ค่าคลาสเป้าหมาย คือ Yes และโมเดลทำนายว่า No

True negative (TN) หมายถึง ค่าคลาสเป้าหมาย คือ No และโมเดลทำนายว่า No

False positive (FP) หมายถึง ค่าคลาสเป้าหมาย คือ No และโมเดลทำนายว่า Yes

การวัดประสิทธิภาพของแบบจำลองสามารถคำนวณตามสมการ (Powers, 2011), (Han *et al*., 2012) ดังนี้

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \quad (6)$$

$$\text{Precision} = \frac{TP}{TP+FP} \times 100 \quad (7)$$

$$\text{Recall} = \frac{TP}{TP+FN} \times 100 \quad (8)$$

วิธีการดำเนินการวิจัย

งานวิจัยนี้นำเสนอการพัฒนาแบบจำลองเพื่อแนะนำการเลือกวิชาเอกของสาขาวิชาเทคโนโลยีสารสนเทศและการสื่อสาร โดยเปรียบเทียบประสิทธิภาพของแบบจำลองที่พัฒนาด้วยเทคนิคต้นไม้ตัดสินใจโดยใช้ อัลกอริทึม ID3 C4.5 และ CART โดยมีขั้นตอนดำเนินการ 4 ขั้นตอนดังนี้ 1) รวบรวมข้อมูล 2) เตรียมข้อมูล 3) พัฒนาแบบจำลองด้วยเทคนิคต้นไม้ตัดสินใจโดยใช้ ID3 C4.5 และ CART 4) วัดประสิทธิภาพแบบจำลอง และ 5) พัฒนาต้นแบบระบบแนะนำการเลือกวิชาเอกของสาขาวิชาเทคโนโลยีสารสนเทศและการสื่อสาร ซึ่งภาพการทำงานโดยรวมดัง Figure 2

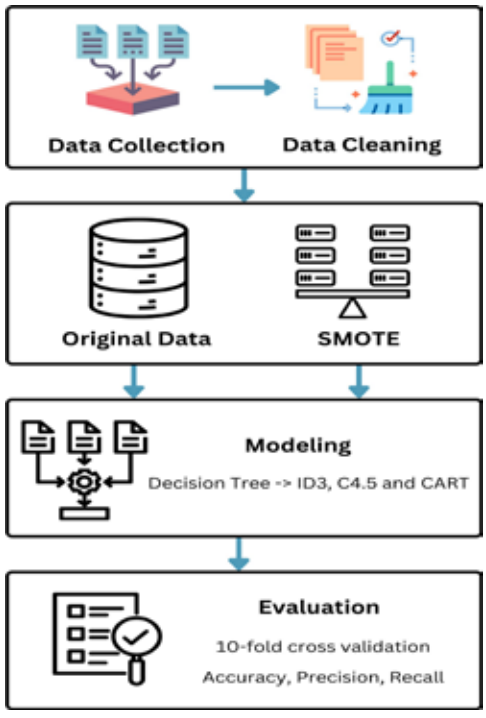


Figure 2 Research Process

1. รวบรวมข้อมูล

การวิจัยครั้งนี้ได้รวบรวมข้อมูลผลการเรียนจำนวน 10 รายวิชาของนักศึกษาในสาขาวิชาเทคโนโลยีสารสนเทศและการสื่อสาร เป็นข้อมูลระหว่างปี พ.ศ. 2563-2565 โดยชุดข้อมูลประกอบด้วยคุณลักษณะ (Attribute) ได้แก่ 1) ผลการเรียนคอมพิวเตอร์เบื้องต้นและองค์ประกอบคอมพิวเตอร์ 2) ผลการเรียนการเขียนโปรแกรมเบื้องต้น 3) ผลการเรียนสถิติและการวิเคราะห์เชิงปริมาณ 4) ผลการเรียนโครงสร้างข้อมูลและขั้นตอนวิธี 5) ผลการเรียนการเขียนโปรแกรมเชิงอ็อบเจ็กต์ 6) ผลการเรียนคณิตศาสตร์เต็มหน่วย 7) ผลการเรียนระบบฐานข้อมูล 8) ผลการเรียนการเขียนโปรแกรมบนเว็บ 9) ผลการเรียนการออกแบบประสบการณ์ผู้ใช้และส่วนติดต่อผู้ใช้งาน และ 10) ผลการเรียนการสื่อสารข้อมูลและเครือข่ายอินเทอร์เน็ต

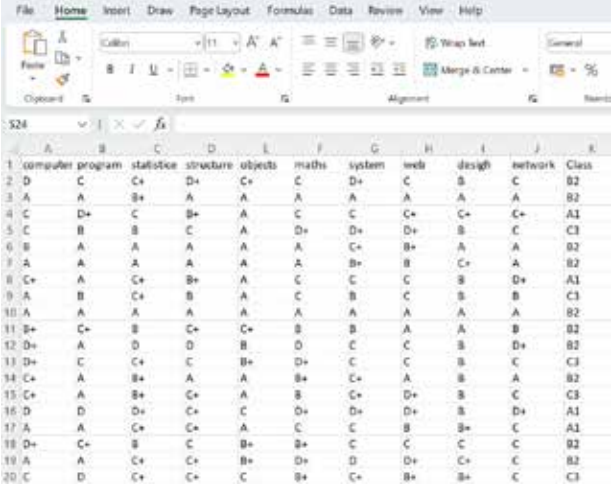
2. เตรียมข้อมูล

หลังจากที่รวบรวมข้อมูลในหัวข้อ 1. เรียบร้อยแล้ว ข้อมูลที่รวบรวมมาได้มีข้อมูลบางส่วนที่จะต้องจัดการก่อนที่จะนำไปพัฒนาแบบจำลองดังนั้นคณะผู้วิจัยจึงนำข้อมูลชุดนี้มาผ่านกระบวนการ Data Cleaning และ Data Transformation หลังจากกระบวนการดังกล่าวจะได้คุณลักษณะ (Attribute) ที่ใช้ในการสร้างแบบจำลองมีรายละเอียด ดัง Table 1

Table 1 Features used for developing model

No.	Courses	Assessment
1	Introduction to Computer and Computer Architecture	A, B+, B, C+,C, D+, D, F
2	Introduction to Programming	A, B+, B, C+,C, D+, D, F
3	Statistics and Quantitative Analysis	A, B+, B, C+,C, D+, D, F
4	Data Structure and Algorithms	A, B+, B, C+,C, D+, D, F
5	Object-Oriented Programming	A, B+, B, C+,C, D+, D, F
6	Discrete Mathematics	A, B+, B, C+,C, D+, D, F
7	Database System	A, B+, B, C+,C, D+, D, F
8	Web Programming	A, B+, B, C+,C, D+, D, F
9	User Experience/User Interface Design	A, B+, B, C+,C, D+, D, F
10	Data Communication and Inter-networking	A, B+, B, C+,C, D+, D, F
11	Class	- Game Technology and Digital Content Creation (A1) - Network Technology and Cyber Security (B2) - Digital Technology (C3)

เมื่อรวบรวมข้อมูลจาก Table 1 เรียบร้อยแล้ว คณะผู้วิจัยได้ทำการตรวจสอบความถูกต้องของข้อมูล จากนั้นแปลงข้อมูลให้อยู่ในรูปแบบไฟล์นามสกุล .CSV เพื่อที่จะนำเข้าไปสู่โปรแกรม Weka ดัง Figure 3



	computer	program	statistics	structure	objects	maths	system	web	design	network	Class
2	D	C	C+	D+	C+	C	D+	C	B	C	B2
3	A	A	B+	A	A	A	A	A	A	A	B2
4	C	D+	C	B+	A	C	C	C+	C+	C+	A1
5	C	B	B	C	A	D+	D+	D+	B	C	C3
6	B	A	A	A	A	A	C+	B+	A	A	B2
7	A	A	A	A	A	A	B+	B	C+	A	B2
8	C+	A	C+	B+	A	C	C	C	B	D+	A1
9	A	B	C+	B	A	C	B	C	B	B	C3
10	A	A	A	A	A	A	A	A	A	A	B2
11	B+	C+	B	C+	C+	B	B	A	A	B	B2
12	D+	A	D	D	B	D	C	C	B	D+	B2
13	D+	C	C+	C	B+	D+	C	C	B	C	C3
14	C+	A	B+	A	A	B+	C+	A	B	A	B2
15	C+	A	B+	C+	A	B	C+	D+	B	C	C3
16	D	D	D+	C+	C	D+	D+	D+	B	D+	A1
17	A	A	C+	C+	A	C	C	B	B+	C	A1
18	D+	C+	B	C	B+	B+	C	C	C	C	B2
19	A	A	C+	C+	B+	D+	D	D+	C+	C	B2
20	C	D	C+	C+	C	B+	C+	B+	B+	C	C3

Figure 3 The examples of data used for developing the model

3. พัฒนาแบบจำลองด้วยเทคนิคต้นไม้ตัดสินใจโดยใช้อัลกอริทึม ID3 C4.5 และ CART

หลังจากแปลงข้อมูล (Data Transformation) เรียบร้อยแล้ว คณะผู้วิจัยได้ตรวจสอบความถูกต้องและวิเคราะห์ข้อมูลแล้ว พบว่ามีข้อมูลบางกลุ่มที่มีจำนวนน้อย ทำให้ข้อมูลที่รวบรวมมาได้ขาดความสมดุล ซึ่งข้อมูลที่รวบรวมมาได้เป็นข้อมูลผลการเรียนของนักศึกษาจำนวน 196 คน แบ่งเป็นข้อมูลในกลุ่ม A1=70 B2=48 และ C3=78 เรคคอร์ด ดังนั้นคณะผู้วิจัยจึงปรับให้ข้อมูลมีความสมดุลด้วยวิธี Synthetic Minority Over-sampling TEchnique (SMOTE) ซึ่งวิธี SMOTE ใช้แก้ปัญหากรณีชุดข้อมูลที่ผลลัพธ์ของคำตอบที่จะพยากรณ์มีสัดส่วนไม่เท่ากัน (Chawla *et al*., 2002) โดยปรับค่าพารามิเตอร์ ในการเพิ่มจำนวนข้อมูลทุกกลุ่มตั้งแต่ร้อยละ 100 ถึง 300 ผลการเพิ่มจำนวนข้อมูล 100% จะได้ A1=142 B2=86 และ C3=108 เรคคอร์ด ผลการเพิ่มจำนวนข้อมูล 200% จะได้ A1=224 B2=122 และ C3=168 เรคคอร์ด ผลการเพิ่มจำนวนข้อมูล 300% จะได้ A1=324 B2=221 และ C3=274 เรคคอร์ดหลังจากปรับความไม่สมดุลของข้อมูลด้วยวิธี SMOTE แล้วนำข้อมูลชุดนี้ไปสร้างแบบจำลองด้วยเทคนิคต้นไม้ตัดสินใจ 3 วิธี ได้แก่ ID3 C4.5 และ CART

4. วัดประสิทธิภาพแบบจำลอง

ในงานวิจัยนี้ใช้วิธีการทดสอบประสิทธิภาพแบบจำลองด้วยวิธี 10-fold cross validation เพื่อให้ข้อมูลทุกตัวมีโอกาสเป็นชุดทดสอบและชุดสอน โดยแบ่งข้อมูลออกเป็น 10 ชุดเท่ากัน ใช้ 1 ชุดเป็นกลุ่มทดสอบ ข้อมูลอีก 9 ชุดเป็นกลุ่มที่ใช้ในการเรียนรู้และทำการวนซ้ำจำนวน 10 รอบ โดยเปลี่ยนกลุ่มทดสอบไปเรื่อย ๆ จนครบ ในส่วนการวัดประสิทธิภาพค่าที่ใช้วัดประสิทธิภาพแบบจำลองคือ ค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) และค่าระลึก (Recall) (Powers, 2011; Han *et al*., 2012)

5. พัฒนาต้นแบบระบบแนะนำการเลือกวิชาเอกของสาขาวิชาเทคโนโลยีสารสนเทศและการสื่อสาร

หลังจากที่พัฒนาแบบจำลองแล้ว คณะผู้วิจัยเลือกแบบจำลองที่ให้ประสิทธิภาพสูงที่สุด นำไปใช้ในการพัฒนาต้นแบบระบบแนะนำการเลือกวิชาเอกของสาขาวิชาเทคโนโลยีสารสนเทศและการสื่อสาร โดยระบบพัฒนาด้วยภาษาพีเอชพี (PHP) ร่วมกับฐานข้อมูลมายเอสคิวแอล (MySQL) โดยโครงสร้างระบบแนะนำการเลือกวิชาเอกของสาขาวิชาเทคโนโลยีสารสนเทศและการสื่อสาร ดัง Figure 4

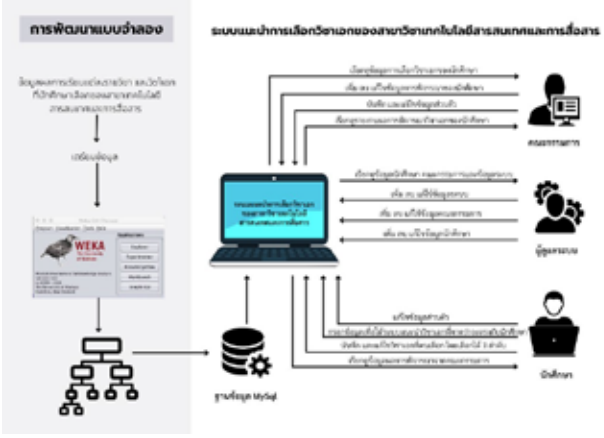


Figure 4 The structure of the recommendation system for the selection of a major in Information and Communication Technology

ผลการวิจัย

1. ผลการพัฒนาแบบจำลองเพื่อแนะนำการเลือกวิชาเอกของสาขาวิชาเทคโนโลยีสารสนเทศและการสื่อสารด้วยเทคนิคต้นไม้ตัดสินใจ

ผลการประเมินประสิทธิภาพของแบบจำลองด้วยโปรแกรม Weka การประเมินประสิทธิภาพของแบบจำลองใช้วิธีทดสอบประสิทธิภาพด้วยวิธี 10-fold cross validation ซึ่งข้อมูลที่นำมาสร้างแบบจำลองจะต้องผ่านกระบวนการปรับความไม่สมดุลของข้อมูลด้วยวิธี SMOTE ก่อน จากนั้นจึงนำข้อมูลที่ได้เข้าสู่กระบวนการจำแนกข้อมูลด้วยเทคนิคต้นไม้ตัดสินใจ 3 วิธี ได้แก่ ID3 C4.5 และ CART ซึ่งมีการทดสอบประสิทธิภาพแบบจำลองด้วยวิธี 10-fold cross validation เพื่อเปรียบเทียบค่า Precision Recall และ Accuracy สามารถสรุปผลการวัดประสิทธิภาพของแบบจำลองได้ดัง Table 2

Table 2 The results of the model's effectiveness evaluation

Method	Precision (%)	Recall (%)	Accuracy (%)
ID3(Original data)	86.21	86.23	86.20
C4.5 (Original data)	87.44	87.43	87.41
CART(Original data)	86.13	86.16	86.15
SMOTE (100%) & ID3	88.24	88.23	88.21
SMOTE (200%) & ID3	89.54	89.55	89.52
SMOTE (300%) & ID3	90.45	90.44	90.42
SMOTE (100%) & C4.5	91.42	91.41	91.40
SMOTE (200%) & C4.5	93.33	93.32	93.20
SMOTE (300%) & C4.5	92.14	92.13	92.10
SMOTE (100%) & CART	87.32	87.30	87.31
SMOTE (200%) & CART	88.63	88.64	88.61
SMOTE (300%) & CART	89.41	89.43	89.42

จาก Table 2 ผลการประเมินประสิทธิภาพของแบบจำลอง โดยข้อมูลมีการปรับความไม่สมดุลของข้อมูลด้วยวิธี SMOTE โดยเพิ่มจำนวนข้อมูลตั้งแต่ร้อยละ 100 ถึง 300 ผลการทดลองพบว่าขนาดชุดข้อมูลที่สามารถเพิ่มประสิทธิภาพของแบบจำลองได้ดีที่สุด คือร้อยละ 200 แบบจำลองที่ให้ประสิทธิภาพสูงที่สุดคือ SMOTE (200%) & C4.5 มีค่าความถูกต้อง (Accuracy) เท่ากับ 93.20% ค่าความแม่นยำ (Precision) เท่ากับ 93.33% และค่าระลึก (Recall) เท่ากับ 93.32% แบบจำลองนี้ให้ผลลัพธ์ คือกฎที่ใช้ในการตัดสินใจจำนวน 21 กฎ โดยกฎที่ได้นี้สามารถนำไปใช้ในการพัฒนาต้นแบบระบบแนะนำวิชาเอกของสาขาเทคโนโลยีสารสนเทศและการสื่อสารได้ ขอยกตัวอย่างกฎที่ได้จากแบบจำลองจำนวน 2 กฎ ดังนี้

IF network=B AND web=C THEN=C3

ถ้า ผลการเรียนการสื่อสารข้อมูลและเครือข่ายอินเทอร์เน็ตเกรด B และผลการเรียนการเขียนโปรแกรมบนเว็บเกรด C แล้วนักศึกษาจะเลือกวิชาเอกเทคโนโลยีดิจิทัล

IF network=B AND web=A THEN=A1

ถ้า ผลการเรียนการสื่อสารข้อมูลและเครือข่ายอินเทอร์เน็ตเกรด B และผลการเรียนการเขียนโปรแกรมบนเว็บ เกรด A แล้วนักศึกษาจะเลือกวิชาเอกเทคโนโลยีเกมและการผลิตดิจิทัลคอนเทนท์

2. ผลการพัฒนาต้นแบบระบบแนะนำการเลือกวิชาเอกของสาขาวิชาเทคโนโลยีสารสนเทศและการสื่อสาร

ต้นแบบระบบแนะนำการเลือกวิชาเอกของสาขาวิชาเทคโนโลยีสารสนเทศและการสื่อสาร ภาพด้านล่างเป็นภาพหน้าหลักของระบบ ดัง Figure 5

Figure 5 Screen design for main homepage

ผู้ใช้สามารถกรอกข้อมูลผลการเรียนเพื่อให้ระบบแนะนำการเลือกวิชาเอกของสาขาวิชาเทคโนโลยีสารสนเทศและการสื่อสาร ดัง Figure 6

Figure 6 Screen design for fill in the grades

หลังจากที่กรอกข้อมูลด้านบนเรียบร้อยแล้วระบบจะแนะนำวิชาเอกของสาขาวิชาเทคโนโลยีสารสนเทศและการสื่อสาร ดัง Figure 7

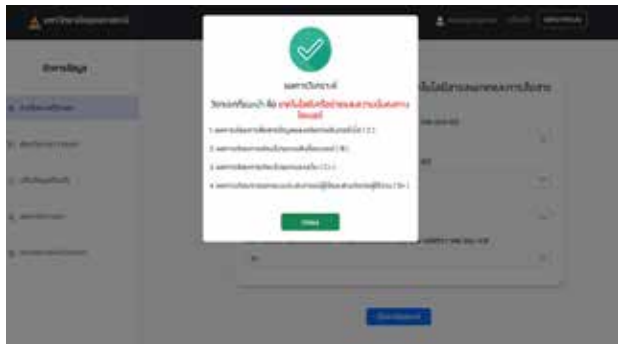


Figure 7 Screen design for recommendation of a major in Information and Communication Technology

3. ผลการประเมินความพึงพอใจของผู้ใช้ระบบ

คณะผู้วิจัยได้ศึกษาความพึงพอใจของผู้ใช้ระบบแนะนำวิชาเอกของสาขาวิชาเทคโนโลยีสารสนเทศและการสื่อสาร โดยใช้แบบสอบถามความพึงพอใจจากผู้ใช้ที่เป็นนักศึกษา จำนวน 30 คน ในแบบสอบถามใช้มาตราส่วนประมาณค่า (rating scale) 5 ระดับ คือ มากที่สุด มาก ปานกลาง น้อย และน้อยที่สุด ใช้เกณฑ์การแปล ความหมายของ (บุญชม ศรีสะอาด, 2560) ดังนี้ ผลปรากฏดัง Table 3

4.51-5.00 หมายถึง เห็นด้วยมากที่สุด

3.51-4.50 หมายถึง เห็นด้วยมาก

2.51-3.50 หมายถึง เห็นด้วยปานกลาง

1.51-2.50 หมายถึง เห็นด้วยน้อย

1.00-1.50 หมายถึง เห็นด้วยน้อยที่สุด

Table 3 The average score of user satisfaction towards the recommendation system

Assessment item	$\bar{X}$	S.D.	level
1. The ability to recommend a major in Information and Communication Technology using decision tree techniques	3.77	0.94	Good
2. The ability to display results for considering a major from the committee	3.65	0.96	Good
3. The system's ability to display the history of the recommendation for choosing a major in Information and Communication Technology	3.73	0.89	Good
4. The system's ability to display graphs comparing the number of users in each major in Information and Communication Technology	3.57	0.96	Good
5. Show the results of the recommendation for choosing a major in Information and Communication Technology correctly	3.70	0.96	Good
6. The accuracy of data recording on the web application	3.63	0.84	Good
7. The accuracy for updating data on the web application	3.53	0.94	Good
8. Ease of use of the web application	3.73	0.96	Good
9. The appropriateness of the screen design	3.57	0.89	Good
10. The appropriateness of user interaction	3.67	0.97	Good
11. The appropriateness of web application security	3.65	0.89	Good
12. Overall satisfaction with web application usage	3.72	0.96	Good
Overall assessment results	**3.66**	**0.93**	**Good**

ผลการประเมินความพึงพอใจของผู้ใช้เว็บแอปพลิเคชันแนะนำวิชาเอกของสาขาวิชาเทคโนโลยีสารสนเทศและการสื่อสาร จำนวน 30 คน ประกอบด้วยรายการประเมิน 12 ข้อ พบว่าโดยรวมอยู่ในระดับมาก ($\bar{X}$=3.66, S.D.=0.93) ข้อที่มีค่าเฉลี่ยสูงสุด 3 อันดับแรก คือ ความสามารถของการแนะนำวิชาเอกของสาขาวิชาเทคโนโลยีสารสนเทศและการสื่อสารด้วยเทคนิคต้นไม้ตัดสินใจ อยู่ในระดับความพึงพอใจมาก ($\bar{X}$=3.77, S.D.=0.94) รองลงมา คือ ความสามารถของระบบในการแสดงผลประวัติการวิเคราะห์วิชาเอก อยู่ในระดับความพึงพอใจมาก ($\bar{X}$=3.73, S.D.=0.89) และความง่ายต่อการใช้งานเว็บแอปพลิเคชัน อยู่ในระดับความพึงพอใจมาก ($\bar{X}$=3.73, S.D.=0.96) ดังแสดงใน Table 3

สรุปผลการวิจัยและข้อเสนอแนะ

บทความนี้นำเสนอแบบจำลองเพื่อแนะนำการเลือกวิชาเอกของสาขาวิชาเทคโนโลยีสารสนเทศและการสื่อสาร โดยมีการเปรียบเทียบประสิทธิภาพของแบบจำลองที่พัฒนาด้วยเทคนิคต้นไม้ตัดสินใจ 3 วิธี คือ อัลกอริทึม ID3 C4.5 และ CART และศึกษาความพึงพอใจของผู้ใช้ระบบแนะนำการเลือกสาขาวิชาเอกของสาขาวิชาเทคโนโลยีสารสนเทศและการสื่อสาร ข้อมูลที่นำมาสร้างแบบจำลองเป็นข้อมูลผลการเรียนของนักศึกษาซึ่งประกอบด้วย ผลการเรียน 10 รายวิชา เนื่องจากข้อมูลที่รวบรวมมาได้มีข้อมูลบางกลุ่มที่มีจำนวนน้อย ดังนั้นคณะผู้วิจัยจึงปรับความไม่สมดุลของข้อมูลด้วยวิธี SMOTE โดยปรับค่าพารามิเตอร์ ในการเพิ่มจำนวนข้อมูลตั้งแต่ร้อยละ 100 ถึง 300 หลังจากปรับความไม่สมดุลของข้อมูลด้วยวิธี SMOTE แล้วนำข้อมูลชุดนี้ไปสร้างแบบจำลองด้วยเทคนิคต้นไม้ตัดสินใจ 3 วิธี คือ ID3 C4.5 และ CART ซึ่งการทดสอบประสิทธิภาพแบบจำลองด้วยวิธี 10-fold cross validation ผลการทดสอบประสิทธิภาพของแบบจำลองพบว่าแบบจำลองที่พัฒนาด้วยอัลกอริทึม C4.5 ให้ประสิทธิภาพดีที่สุด โดยมีค่าความถูกต้องเท่ากับ 93.20 % ค่าความแม่นยำเท่ากับ 93.33% และค่าระลึกเท่ากับ 93.32% สำหรับผลการประเมินความพึงพอใจจากผู้ใช้งานระบบ จำนวน 30 คน พบว่าในภาพรวมผู้ใช้งานมีความพึงพอใจต่อระบบมีค่าเฉลี่ยเท่ากับ 3.66 และ ค่าส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.93 ซึ่งสรุปได้ว่าผู้ใช้งานมีความพึงพอใจอยู่ในระดับ มาก ผลการวิจัยนี้สอดคล้องกับงานวิจัยของ ธนพร คล้ายทอง และชุติพนธ์ ศรีสวัสดิ์ (2566) ที่แสดงให้เห็นว่าการประยุกต์ใช้วิธีเทคนิคต้นไม้ตัดสินใจ (Decision Tree) นำสร้างแบบจำลองในการพยากรณ์ข้อมูลให้ประสิทธิภาพในระดับดีและสามารถนำไปประยุกต์ใช้งานได้

ข้อเสนอแนะ

ข้อเสนอแนะในการทำวิจัยในอนาคต ควรนำคุณลักษณะอื่นมาใช้ในการวิเคราะห์ข้อมูล เช่น ข้อมูลความสนใจ หรือความสามารถพิเศษของนักศึกษามาวิเคราะห์เพิ่มเติม เพื่อเพิ่มประสิทธิภาพของแบบจำลอง และควรพัฒนาในรูปแบบของ Mobile Application ที่สามารถใช้งานได้ง่าย

เอกสารอ้างอิง

ธนพร คล้ายทอง และ ชุติพนธ์ ศรีสวัสดิ์. (2566). การพยากรณ์การตกออกของนักศึกษา ระดับปริญญาตรี มหาวิทยาลัยราชภัฏพิบูลสงคราม ด้วยเทคนิคเหมืองข้อมูล, *วารสารวิทยาการสารสนเทศและเทคโนโลยีประยุกต์*, *5(1)*, 1-17.

บุญชม ศรีสะอาด. (2560). *การวิจัยเบื้องต้น* (พิมพ์ครั้งที่ 10). กรุงเทพฯ: สุรีริยาสาส์น.

รัตนาวดี พานทอง. (2565). การจำแนกประเภทข้อมูลการท่องเที่ยวในจังหวัดพะเยาโดยการใช้เทคนิคต้นไม้ตัดสินใจ. *วารสารวิชาการเพื่อการพัฒนานวัตกรรมเชิงพื้นที่ (JSID)*, *3(3)*, 40-53.

Chawla, N., Bowyer, K., Hall, L., Kegelmeyer, W. (2002). SMOTE: Synthetic Minority Over-sampling TEchnique. *Journal of Artificial Intelligence Research*, *16*, 341-378.

Han, J., Kamber, M., Pei J., (2012). *Data Mining: Concepts Techniques*, 3rd ed. San Mateo, CA, USA: Morgan Kaufmann.

Ogheneovo, E.E., NIerum, P.A., (2020). Iterative Dichotomizer 3 (ID3) Decision Tree: A Machine Learning Algorithm for Data Classification and Predictive Analysis, *International Journal of Advanced Engineering Research and Science (IJAERS), 7(4)*, 514-521.

Powers, D.M.W., (2011). Evaluation : From Precision, Recall and F-measure To ROC, Informedness & Correlation. *Journal of Machine Learning Technologies*, *2(1)*, 37-63.

Sharma, H., Kumar, S., (2016). A Survey on Decision Tree Algorithms of Classification in Data Mining. *International Journal of Science and Research (IJSR)*. *5(4)*, 2094-2097.

Viloria, A. J., Padilla, C. G., Vargas-Mercado, H. Hernández-Palma, N. O. Llinas, David, M. A., (2019). Integration of Data Technology for Analyzing University Dropout. *Procedia Comput. Sci.*, 155(2018), 569-574.