

การเปรียบเทียบเทคนิคการจำแนกประเภทข้อมูลสำหรับทำนายแนวโน้ม
การสำเร็จการศึกษาของนักเรียน
Classification Techniques Comparison for Predicting Graduate Trend

ศิรินตรา เสือพิทักษ์* บุญมี นิสสัยดี และ วิภาวรรณ บัวทอง

Sirintra Suepitak*, Boonmee Nissaidee, & Wipawan Buathong

สาขาวิชาเทคโนโลยีดิจิทัล คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏภูเก็ต

Digital Technology Program, Faculty of Science and Technology,

Phuket Rajabhat University

Submitted 24/8/2021 ; Revised 15/10/2021 ; Accepted 8/11/2021

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อสร้างแบบจำลองในการทำนายแนวโน้มการสำเร็จการศึกษาด้วยเทคนิคการทำเหมืองข้อมูลแบบจำแนกประเภทข้อมูล โดยใช้ 3 อัลกอริทึม ได้แก่ นาอ์ฟเบย์ ต้นไม้ตัดสินใจ และเพื่อนบ้านใกล้เคียง ชุดข้อมูลที่ใช้ในการทำวิจัยครั้งนี้คือ ผลการเรียนวิชาพื้นฐานในภาคเรียนที่ 1 ของนักเรียนระดับประกาศนียบัตรวิชาชีพแผนกวิชาคอมพิวเตอร์ธุรกิจประจำปีการศึกษา 2560 วิทยาลัยอาชีวศึกษาภูเก็ต จำนวน 175 คน ใช้วิธีการประเมินประสิทธิภาพของแบบจำลอง 3 แบบ คือ การตรวจสอบแบบข้ามจำนวน 5 กลุ่ม การตรวจสอบแบบข้ามจำนวน 10 กลุ่ม และการสุ่มด้วยการแบ่งเป็นร้อยละที่ร้อยละ 66 ผลการเปรียบเทียบประสิทธิภาพของอัลกอริทึมทั้ง 3 ชนิด แสดงให้เห็นว่าเทคนิคการทำเหมืองข้อมูลแบบจำแนกประเภทข้อมูลโดยใช้อัลกอริทึมต้นไม้ตัดสินใจให้ค่าการประเมินประสิทธิภาพของแบบจำลองที่ดีที่สุด ซึ่งผลการประเมินแบบจำลองด้วยวิธีการตรวจสอบแบบข้ามจำนวน 5 กลุ่ม และการตรวจสอบแบบข้ามจำนวน 10 กลุ่ม ให้ผลลัพธ์เท่ากัน คือ ค่าความถูกต้อง ร้อยละ 87.40 ค่าความแม่นยำ ร้อยละ 87.50 ค่าเรียกคืน ร้อยละ 87.40 และค่าประเมินผลโดยรวม ร้อยละ 86.70

คำสำคัญ: การทำเหมืองข้อมูล นาอ์ฟเบย์ ต้นไม้ตัดสินใจ เพื่อนบ้านใกล้เคียง

*ผู้ประสานงานหลัก (Corresponding Author)

E-mail: s6381423107@pkru.ac.th

Abstract

The objective of this research was to develop a model for predicting graduation trends by using classification data mining techniques. Three algorithms including Naive Bayes, Decision Tree (J48), and K-Nearest Neighbor (KNN) were used in this study. The dataset used in this research was the academic achievement of the core courses in semester 1 of the academic year 2017 of 175 students in the Department of Business Computing, Phuket Vocational College. Three methods comprising 5-folds cross-validation, 10-folds cross-validation, and the percentage split of 66, were used to evaluate the model performance. The decision tree model's assessment using 5-folds and 10-folds cross-validation revealed the model's best performance with 87.40 percent accuracy, 87.50 percent precision, 87.40 percent recall, and 86.70 percent F-measure.

Keywords: data mining, Naive Bayes, decision tree, K-nearest neighbor

บทนำ

สำนักงานคณะกรรมการการอาชีวศึกษาเป็นหน่วยงานที่จัดการศึกษาเพื่อพัฒนาคุณภาพชีวิตของคนในประเทศ โดยมุ่งเสริมสร้างความรู้ ความสามารถ ประสบการณ์ และสมรรถภาพในการประกอบอาชีพที่มีทักษะฝีมือออกสู่ตลาดแรงงาน [1] แต่อย่างไรก็ตาม พบว่ากระบวนการจัดการศึกษายังคงมีปัญหาในหลายด้าน ได้แก่ การจัดการเรียนการสอน หลักสูตร งบประมาณ ปัญหาส่วนตัวของนักเรียน รวมไปถึงโครงสร้างการบริหารงาน และระบบการจัดการภายในสถานศึกษา เป็นต้น ในปัจจุบันสถานศึกษาที่อยู่ในสังกัดกรมอาชีวศึกษามีมากกว่า 400 แห่ง จำแนกเป็นวิทยาลัยเทคนิค วิทยาลัยอาชีวศึกษา วิทยาลัยสารพัดช่าง วิทยาลัยอาชีวศึกษา วิทยาลัยเกษตรและเทคโนโลยี รวมทั้งวิทยาลัยที่เปิดการเรียนการสอนเฉพาะสาขาวิชาซึ่งเรียกชื่อตามลักษณะประเภทวิชาที่เปิดสอน [2] ที่ผ่านมามีพบว่ามึ้นักเรียนส่วนหนึ่งไม่สำเร็จการศึกษา เพราะมีปัญหาการเรียนในด้านต่าง ๆ เช่น ไม่ชอบหรือไม่ถนัดในสาขาวิชาที่เข้าเรียน หรือไม่สามารถปรับตัวให้เข้ากับสถานศึกษาแห่งใหม่ ซึ่งส่งผลกระทบต่อการผลิตกำลังคนเพื่อเข้าสู่ตลาดแรงงาน ทำให้เกิดปัญหาต่อครอบครัวของนักเรียนเอง ปัญหาการว่างงาน และปัญหาทางสังคม กล่าวคือรัฐบาลสูญเสียงบประมาณในการลงทุนเพื่อผลิตบุคลากร นอกจากนี้ยังเป็นอุปสรรคต่อการขับเคลื่อนเพื่อการบรรลุเป้าหมายยุทธศาสตร์การส่งเสริมการจัดการศึกษาขั้นพื้นฐานประเภทการอาชีวศึกษา [3] ซึ่งปัญหาการออกกลางคันของนักเรียนเป็นปัญหาเร่งด่วนที่จะต้องดำเนินการแก้ไข อันจะเป็นประโยชน์ต่อการจัดการศึกษาของสถานศึกษาในสังกัดสำนักงานคณะกรรมการการอาชีวศึกษา และหน่วยงานที่เกี่ยวข้องได้นำไปประยุกต์ใช้ และเป็นประโยชน์ต่อผู้เรียนเพื่อให้นักเรียนที่มีคุณภาพและมีคุณลักษณะที่พึงประสงค์ได้สำเร็จการศึกษาจากสถานศึกษา ซึ่งนักเรียนเหล่านี้จะเป็นบุคลากรที่มีคุณภาพ คุณธรรม และประกอบอาชีพได้ต่อไปในระยะยาว [4]

วิทยาลัยอาชีวศึกษาภูเก็ต สังกัดสถาบันการอาชีวศึกษา ในปีการศึกษา 2561 – 2563 ระหว่างการรายงานข้อมูลประจำปีงบประมาณ 2561 – 2563 มีนักเรียน ระดับประกาศนียบัตรวิชาชีพ (ปวช.) รวมทั้งสิ้น 2,662 คน และมีนักเรียนที่ออกกลางคันในระหว่างปีการศึกษา 2561 – 2563 จำนวน 107 คน จำแนกได้เป็นปีการศึกษา 2561 จำนวน 11 คน ปีการศึกษา 2562 จำนวน 45 คน และปีการศึกษา 2563 จำนวน 51 คน ซึ่งจำนวนนักเรียนที่ออกกลางคันนั้นเพิ่มมากขึ้นทุกปีและเป็นปัญหาสำคัญของสถานศึกษา

เทคโนโลยีที่ใช้วิเคราะห์ข้อมูลที่มีความสำคัญจากแหล่งเก็บข้อมูลที่มีขนาดใหญ่ เรียกว่า การทำเหมืองข้อมูล หรือการขุดค้นข้อมูล (data mining) ซึ่งหัวใจสำคัญของกระบวนการทำเหมืองข้อมูล คือ ส่วนของโปรแกรมที่ทำหน้าที่สังเคราะห์ความรู้ขึ้นมาจากข้อมูลจำนวนมากในฐานข้อมูล เรียกว่า อัลกอริทึมสำหรับการเรียนรู้ (learning algorithm) การทำเหมืองข้อมูลสามารถพัฒนาไปใช้สำหรับหลากหลายแอปพลิเคชันในภาคการศึกษาเพื่อวัตถุประสงค์ในการเพิ่มประสิทธิภาพการทำงานของนักเรียนและการรับรองของสถาบันการศึกษา [5] ซึ่งมีผู้เสนอแนวคิดและพัฒนาอัลกอริทึมที่ใช้หลักการของการสร้างต้นไม้ตัดสินใจ (decision tree) อัลกอริทึมที่ใช้หลักการทางสถิติและทฤษฎีของเบย์ ที่ชื่อว่านาอิวเบย์ (Naive Bayes) อัลกอริทึมที่ใช้หลักการของโครงข่ายประสาทเทียม (neural network) และอัลกอริทึมอื่น ๆ อีกมาก ปัจจุบันได้มีนักคอมพิวเตอร์ทดสอบเปรียบเทียบความสามารถของอัลกอริทึมแต่ละประเภท เพื่อค้นหาว่าอัลกอริทึมใดมีความสามารถสูงที่สุด ผลการทดสอบส่วนใหญ่ปรากฏว่าไม่มีอัลกอริทึมใดที่ทำงานได้ดีที่สุดในข้อมูลทุกประเภท เนื่องจากข้อมูลแต่ละประเภทมีลักษณะเฉพาะตัวที่แตกต่างกัน [6]

อัลกอริทึมนาอิวเบย์ เป็นเครื่องมือที่อาศัยหลักการของความน่าจะเป็นตามทฤษฎีของเบย์ (Bayes theorem) เป็นอัลกอริทึมที่ไม่ซับซ้อน จำแนกข้อมูลโดยการเรียนรู้ปัญหาที่เกิดขึ้นเพื่อนำมาสร้างเงื่อนไขการจำแนกข้อมูลใหม่ หลักการของนาอิวเบย์จะใช้การคำนวณหาความน่าจะเป็นซึ่งถูกใช้ในการแก้ปัญหาแบบการจำแนกที่สามารถคาดการณ์ผลลัพธ์ได้ และวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรเพื่อใช้ในการสร้างเงื่อนไขความน่าจะเป็นสำหรับแต่ละความสัมพันธ์ นาอิวเบย์จึงเป็นวิธีจำแนกกลุ่มข้อมูลที่มีประสิทธิภาพ [7] เป็นเทคนิค

การแก้ปัญหาแบบจำแนกประเภทที่สามารถคาดการณ์ผลลัพธ์ได้โดยการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปร เพื่อใช้สร้างเงื่อนไขความน่าจะเป็นสำหรับแต่ละความสัมพันธ์ เหมาะกับกรณีของเซตตัวอย่างที่มีจำนวนมาก [8] ได้มีการนำอัลกอริทึมนาอิวเพย์ไปใช้ในการวิเคราะห์หาปัจจัยที่เกี่ยวข้องในการลาออกกลางคันของนักศึกษา ระดับปริญญาตรีเพื่อเปรียบเทียบประสิทธิภาพในการจำแนกข้อมูล พบว่าแบบจำลองที่สร้างด้วยเทคนิควิธีนาอิวเพย์มีประสิทธิภาพสูงสุด มีค่าเฉลี่ยความถูกต้อง 93.58 % [9]

อัลกอริทึมต้นไม้ตัดสินใจ เป็นอัลกอริทึมที่ใช้สร้างต้นไม้ตัดสินใจ พัฒนาโดย Ross Quinlan ซึ่งปรับปรุงจากอัลกอริทึม ID3 โดย C4.5 จะสร้างผังต้นไม้ตัดสินใจมาจากกลุ่มของข้อมูลชุดฝึกที่ใช้หลักการของ ทฤษฎีสารสนเทศ (information entropy) ในการเลือกโหนดโดยใช้ค่าความถูกต้องของคุณลักษณะแต่ละตัว เพื่อชี้แยกเส้นทางตัดสินใจ โดยพิจารณาจากค่าทฤษฎีสารสนเทศ [10] ต้นไม้ตัดสินใจเป็นวิธีการแสดงข้อมูลที่รวดเร็วเมื่อเทียบกับเทคนิคอื่นและมีประสิทธิภาพในการแสดงโครงสร้างของข้อมูลผลลัพธ์ที่กระชับและง่ายขึ้น จากงานวิจัยเพื่อเปรียบเทียบประสิทธิภาพของอัลกอริทึมในการทำนายและคุณลักษณะที่มีต่อผลสัมฤทธิ์ทางการเรียนของนักศึกษาระดับอาชีวศึกษาพบว่าการใช้เทคนิคต้นไม้ตัดสินใจมีความถูกต้องเท่ากับ 83.08% [11] สามารถนำเทคนิคต้นไม้ตัดสินใจไปใช้ทำนายผลสัมฤทธิ์ทางการเรียน และเป็นแนวทางในการสอนเสริมหรือแนะแนวให้กับนักเรียนต่อไป

อัลกอริทึมเพื่อนบ้านใกล้เคียง (K – nearest neighbor : KNN) เป็นวิธีการที่ได้รับความนิยมในการใช้งานอย่างมาก เนื่องจากเป็นวิธีที่ง่ายและมีประสิทธิภาพซึ่งสามารถนำไปประยุกต์ใช้กับงานได้หลากหลาย เช่น งานทางด้านจำแนกกลุ่ม (classification) เพื่อพัฒนาแบบจำลองสำหรับการพยากรณ์ผลสัมฤทธิ์ทางการเรียนนอกจากนี้ด้วยวิธีการเรียนการสอนแบบผสมผสาน เลือกใช้เทคนิคเหมือนข้อมูลแบบเพื่อนบ้านใกล้เคียง โดยมีความถูกต้องสูงสุดเท่ากับ 86.13% [12]

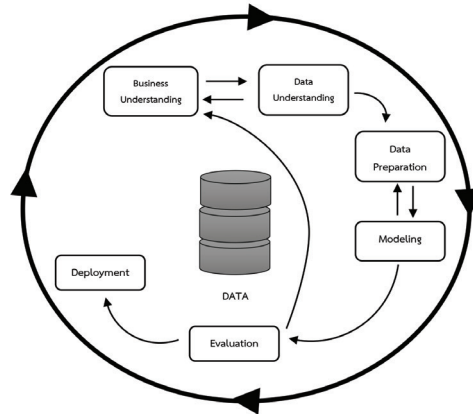
จากเหตุผลดังกล่าวผู้วิจัยจึงนำเทคนิควิธีการทำเหมืองข้อมูลด้วยเทคนิคนาอิวเพย์ ต้นไม้ตัดสินใจ และเพื่อนบ้านใกล้เคียง มาเปรียบเทียบประสิทธิภาพแบบจำลองในการทำนายผลการสำเร็จการศึกษาของนักเรียน แผนกวิชาคอมพิวเตอร์ธุรกิจ วิทยาลัยอาชีวศึกษาภูเก็ต เพื่อวิเคราะห์หาปัจจัยที่จะช่วยลดปัญหาการออกกลางคันของนักเรียน ซึ่งเป็นปัญหาที่จะต้องแก้ไขเพื่อประโยชน์ต่อการนำไปประยุกต์ใช้ในการจัดการศึกษาของสถานศึกษาในสังกัดสำนักงานคณะกรรมการการอาชีวศึกษาและหน่วยงานที่เกี่ยวข้อง รวมถึงเป็นประโยชน์ต่อผู้เรียน เพื่อให้นักเรียนเป็นผู้ที่มีคุณภาพและคุณลักษณะตามประสงค์ได้เมื่อสำเร็จการศึกษาจากสถานศึกษา

วัตถุประสงค์

เพื่อสร้างแบบจำลองในการทำนายแนวโน้มการสำเร็จการศึกษาระดับประกาศนียบัตรวิชาชีพของวิทยาลัยอาชีวศึกษาภูเก็ต ด้วยเทคนิคการทำเหมืองข้อมูลแบบจำแนกประเภทข้อมูล

วิธีดำเนินการวิจัย

เปรียบเทียบประสิทธิภาพแบบจำลองในการทำนายผลการสำเร็จการศึกษาของนักเรียนแผนกวิชาคอมพิวเตอร์ธุรกิจ วิทยาลัยอาชีวศึกษาภูเก็ต ด้วยวิธีการทำเหมืองข้อมูล ตามขั้นตอนแบบจำลองมาตรฐาน กระบวนการวิเคราะห์ข้อมูลด้วย CRISP-DM (cross industry standard process for data mining) ที่ประกอบด้วย 6 ขั้นตอน ดังภาพที่ 1



ภาพที่ 1 กระบวนการวิเคราะห์ข้อมูลด้วย CRISP-DM (ดัดแปลงจาก [13, 14])

วิธีการวิจัยในแต่ละขั้นตอนของกระบวนการวิเคราะห์ข้อมูลด้วย CRISP-DM มีรายละเอียดดังนี้

1. ความเข้าใจปัญหา (business understanding) เป็นการนำปัญหานั้นนักเรียนออกกลางคันมาแปลงให้อยู่ในรูปแบบโจทย์ของการวิเคราะห์ข้อมูล พร้อมทั้งวางแผนการดำเนินการ

2. ความเข้าใจข้อมูล (data understanding) เริ่มจากการเก็บรวบรวมข้อมูลโดยใช้ข้อมูลผลการเรียนของนักเรียนแผนกวิชาคอมพิวเตอร์ธุรกิจ ประจำปีการศึกษา 2560 – 2561 ภาคเรียนที่ 1 รายวิชาพื้นฐาน 10 รายวิชา ที่มีจำนวนข้อมูลนักเรียน 175 เรคอร์ด จำแนกเป็น 17 แอตทริบิวต์ ดังตารางที่ 1 โดยกำหนดให้แอตทริบิวต์สถานะเป็นตัวแสดงผลลัพธ์

ตารางที่ 1 รายละเอียดของข้อมูล

ลำดับ	แอตทริบิวต์	ชนิดข้อมูล	รายละเอียด
1	no	ตัวเลข	ลำดับ
2	Id_student	ตัวเลข	รหัสนักเรียน
3	Sex	ตัวอักษร	เพศ
4	Thai	ตัวเลข	เกรดวิชาภาษาไทยพื้นฐาน
5	Language	ตัวเลข	เกรดวิชาภาษาอังกฤษในชีวิตจริง
6	Science	ตัวเลข	เกรดวิชาวิทยาศาสตร์เพื่อพัฒนาทักษะชีวิต
7	Math	ตัวเลข	เกรดวิชาคณิตศาสตร์พื้นฐานอาชีพ
8	Social	ตัวเลข	เกรดวิชาประวัติศาสตร์ชาติไทย
9	Gymnastics	ตัวเลข	เกรดวิชาพลศึกษาเพื่อพัฒนาสุขภาพ
10	Computer	ตัวเลข	เกรดวิชาคอมพิวเตอร์และสารสนเทศเพื่องานอาชีพ
11	Account	ตัวเลข	เกรดวิชาการบัญชีเบื้องต้น
12	Marketing	ตัวเลข	เกรดวิชาการขายเบื้องต้น
13	Typewriter	ตัวเลข	เกรดวิชาพิมพ์ไทยเบื้องต้น
14	GPA1	ตัวเลข	คะแนนเฉลี่ยเทอม 1
15	GPA2	ตัวเลข	คะแนนเฉลี่ยเทอม 2
16	GPA_finish	ตัวเลข	คะแนนเฉลี่ย
17	Status	ตัวอักษร	สถานะ

3. การเตรียมข้อมูล (data preparation) ในขั้นตอนนี้เป็นการทำความสะอาดข้อมูล (data cleansing) โดยการแก้ไขค่าว่างให้มีตัวเลขเป็น 0 เพื่อให้รายการข้อมูลที่ไม่ถูกต้องถูกตัดออกไปจากชุดข้อมูล ตาราง หรือฐานข้อมูล แก้ไขในส่วนของหัวตารางโดยปรับให้แอตทริบิวต์เป็นภาษาอังกฤษ และลบแอตทริบิวต์ที่ไม่จำเป็นออกไปจำนวน 4 แอตทริบิวต์ คือ รหัสนักเรียน คะแนนเฉลี่ยเทอม 1 คะแนนเฉลี่ยเทอม 2 และคะแนนเฉลี่ย

4. การสร้าง (modeling) เป็นขั้นตอนการสร้างแบบจำลองเพื่อทำนายผลการสำเร็จการศึกษาของนักเรียนแผนกวิชาคอมพิวเตอร์ธุรกิจ วิทยาลัยอาชีวศึกษาภูเก็ต โดยนำข้อมูลที่ใช้ในการเรียนรู้ (training data) มาสร้างเป็นแบบจำลองด้วยเทคนิคการจำแนกประเภทข้อมูล โดยใช้อัลกอริทึมนาอิวเพย์ ต้นไม้ตัดสินใจ และเพื่อนบ้านใกล้เคียง

5. การประเมินผล (evaluation) เป็นการนำแบบจำลองที่สร้างขึ้นมาวัดประสิทธิภาพของผลลัพธ์ที่ได้ โดยใช้เกณฑ์ในการประเมินค่าประสิทธิภาพของแบบจำลอง คือ ค่าความถูกต้อง (accuracy) ค่าความแม่นยำ (precision) ค่าเรียกคืน (recall) และ ค่าประสิทธิภาพโดยรวม (F-measure) เพื่อวิเคราะห์ว่าตรงกับวัตถุประสงค์ที่ได้ตั้งไว้หรือไม่ โดยมีสมการในการวิเคราะห์ดังนี้

$$\text{โดย ค่าความถูกต้อง} = \frac{TP + TN}{(TP + TN + FP + FN)}$$

$$\text{ค่าความแม่นยำ} = \frac{TP}{(TP + FP)}$$

$$\text{ค่าเรียกคืน} = \frac{TP}{(TP + FN)}$$

$$\text{ค่าประสิทธิภาพโดยรวม} = \frac{2 \times \text{ความแม่นยำ} \times \text{ค่าเรียกคืน}}{\text{ความแม่นยำ} + \text{ค่าเรียกคืน}}$$

หมายเหตุ: True Positive (TP) คือ สิ่งที่โปรแกรมทำนายว่า “จริง” และมีค่าเป็น “จริง”
 True Negative (TN) คือ สิ่งที่โปรแกรมทำนายว่า “ไม่จริง” และมีค่า “ไม่จริง”
 False Positive (FP) คือ สิ่งที่โปรแกรมทำนายว่า “จริง” แต่มีค่าเป็น “ไม่จริง”
 False Negative (FN) คือ สิ่งที่โปรแกรมทำนายว่า “ไม่จริง” แต่มีค่าเป็น “จริง”

6. การนำแบบจำลองไปใช้งาน (deployment) นำแบบจำลองที่ได้และผลการวิเคราะห์มาเปรียบเทียบประสิทธิภาพวิธีการที่นำเชื่อถือมากที่สุดและนำไปใช้จริง

ผลการวิจัย

ผลการวิเคราะห์การทำเหมืองข้อมูลแบบการจำแนกประเภทข้อมูล สำหรับทำนายแนวโน้มการสำเร็จการศึกษาของนักเรียนแบบจำแนกประเภทข้อมูลโดยใช้อัลกอริทึม 3 ประเภท ได้แก่ นาอิวเพย์ ต้นไม้ตัดสินใจ และเพื่อนบ้านใกล้เคียง เพื่อประเมินประสิทธิภาพของแบบจำลองด้วยวิธีการตรวจสอบแบบข้ามจำนวน 5 กลุ่ม การตรวจสอบแบบข้าม จำนวน 10 กลุ่ม และการสุ่มด้วยการแบ่งเป็นร้อยละที่ร้อยละ 66 ด้วยโปรแกรมเวกา (WEKA) ผลการวิจัยมีรายละเอียดดังตารางที่ 3

ตารางที่ 3 เปรียบเทียบผลการวิเคราะห์ข้อมูลตามการจำแนกประเภทข้อมูลด้วยเทคนิคเหมืองข้อมูล

อัลกอริทึม	เกณฑ์ในการวัดค่าประสิทธิภาพ	วิธีการประเมินประสิทธิภาพของแบบจำลอง		
		การตรวจสอบแบบข้ามจำนวน 5 กลุ่ม (ร้อยละ)	การตรวจสอบแบบข้ามจำนวน 10 กลุ่ม (ร้อยละ)	การสุ่มด้วยการแบ่งเป็นร้อยละที่ร้อยละ 66 (ร้อยละ)
นาอึฟเบย์	ค่าความแม่นยำ	86.30	86.30	81.40
	ค่าความถูกต้อง	86.00	86.00	81.20
	ค่าเรียกคืน	86.30	86.30	81.40
	ค่าประสิทธิภาพโดยรวม	85.60	85.60	79.70
ต้นไม้ตัดสินใจ	ค่าความแม่นยำ	87.40	87.40	84.70
	ค่าความถูกต้อง	87.50	87.50	87.40
	ค่าเรียกคืน	87.40	87.40	84.70
	ค่าประสิทธิภาพโดยรวม	86.70	86.70	82.70
เพื่อนบ้าน	ค่าความแม่นยำ	87.40	87.40	79.70
ใกล้เคียง (IBK = 3)	ค่าความถูกต้อง	87.80	87.80	80.80
	ค่าเรียกคืน	87.40	87.40	79.70
	ค่าประสิทธิภาพโดยรวม	86.50	86.50	76.50

จากตารางที่ 3 จะเห็นได้ว่าแบบจำลองที่สร้างด้วยเทคนิคอัลกอริทึมต้นไม้ตัดสินใจมีค่าการประเมินประสิทธิภาพของแบบจำลองสูงที่สุด ด้วยวิธีการประเมินแบบจำลองแบบการตรวจสอบแบบข้ามจำนวน 5 กลุ่ม และการตรวจสอบแบบข้ามจำนวน 10 กลุ่ม ซึ่งให้ผลลัพธ์เท่ากันทั้ง 2 วิธี คือ ค่าความแม่นยำ ร้อยละ 87.40 ค่าความถูกต้อง ร้อยละ 87.50 ค่าเรียกคืน ร้อยละ 87.40 และค่าประสิทธิภาพโดยรวม ร้อยละ 86.70

อภิปรายผลการวิจัย

งานวิจัยนี้สร้างแบบจำลองในการทำนายแนวโน้มการสำเร็จการศึกษาระดับประกาศนียบัตรวิชาชีพของวิทยาลัยอาชีวศึกษาภูเก็ต ด้วยเทคนิคการทำเหมืองข้อมูลแบบจำแนกประเภทข้อมูล โดยใช้ผลการเรียนของนักเรียนแผนกวิชาคอมพิวเตอร์ธุรกิจประจำปีการศึกษา 2560 – 2561 แสดงให้เห็นว่าเทคนิคการทำเหมืองข้อมูลแบบจำแนกประเภทข้อมูลอัลกอริทึมต้นไม้ตัดสินใจ เป็นวิธีการจำแนกประเภทข้อมูลที่มีประสิทธิภาพผลการวิจัยที่ได้จากงานวิจัยนี้สอดคล้องกับงานวิจัยที่ผ่านมา ได้แก่ การเปรียบเทียบประสิทธิภาพของอัลกอริทึมในการทำนายและคุณลักษณะที่มีต่อผลสัมฤทธิ์ทางการเรียนของนักศึกษาระดับอาชีวศึกษา โดยใช้เทคนิคการจำแนกข้อมูล 3 เทคนิค ได้แก่ ต้นไม้ตัดสินใจ นาอึฟเบย์ และการสร้างกฎ ซึ่งพบว่าค่าความถูกต้องของเทคนิคต้นไม้ตัดสินใจมีค่าสูงที่สุด และสามารถนำไปใช้ในการทำนายผลสัมฤทธิ์ทางการเรียน และเป็นแนวทางในการสอนเสริมหรือแนะแนวให้กับนักศึกษาต่อไปได้ [11] การศึกษาปัจจัยที่สัมพันธ์กับการตัดสินใจลาออกและการเปรียบเทียบประสิทธิภาพตัวแบบพยากรณ์การลาออกของพนักงาน กรณีศึกษา บริษัทประกันภัย โดยใช้เทคนิคการวิเคราะห์ข้อมูล 5 เทคนิค ได้แก่ ต้นไม้ตัดสินใจ ซัพพอร์ตเวกเตอร์แมชชีน โครงข่ายประสาทเทียม เทคนิคเบย์เซียนแบบง่ายและเทคนิคเพื่อนบ้านใกล้ที่สุด แล้วเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์เหล่านั้นพบว่าผลการทดสอบตัวแบบพยากรณ์ที่เป็นต้นไม้ตัดสินใจให้ค่าความถูกต้องในการพยากรณ์สูงที่สุด ซึ่งสามารถ

นำผลที่ได้ไปประยุกต์ใช้ในการสร้างแผนกลยุทธ์ เพื่อออกแบบสวัสดิการที่เหมาะสม และรักษาพนักงานให้คงอยู่กับบริษัท อีกทั้งยังเป็นแนวทางในการสร้างระบบสารสนเทศในการสนับสนุนการตัดสินใจสำหรับงานด้านบริหารทรัพยากรบุคคล [15] และการวิเคราะห์ปัจจัยที่ส่งผลต่อการพัฒนาของนักศึกษาที่มีผลการเรียนปกติโดยใช้ต้นไม้ตัดสินใจ และใช้วิธีการประเมินประสิทธิภาพของแบบจำลองการตรวจสอบแบบข้าม จำนวน 10 กลุ่มพบว่าสามารถนำผลการจำแนกข้อมูลที่ได้มาใช้ในการพัฒนาระบบพยากรณ์การพัฒนาของนักศึกษา และสามารถใช้เป็นแนวทางให้กับผู้บริหาร อาจารย์ที่ปรึกษาในการช่วยเหลือด้านการเรียนของนักศึกษา [16]

สรุปผลการวิจัย

งานวิจัยนี้ได้ใช้เทคนิคการทำเหมืองข้อมูลแบบจำแนกประเภทข้อมูล โดยข้อมูลที่ใช้ในการเรียนรู้มีทั้งหมด 13 แอตทริบิวต์ ผลการเปรียบเทียบประสิทธิภาพทั้ง 3 อัลกอริทึม พบว่าอัลกอริทึมต้นไม้ตัดสินใจ ซึ่งสร้างแบบจำลองจากการตรวจสอบแบบข้าม จำนวน 5 กลุ่ม และการตรวจสอบแบบข้าม จำนวน 10 กลุ่ม ให้ค่าประสิทธิภาพของแบบจำลองได้ดีที่สุด คือ ค่าความแม่นยำ ร้อยละ 87.40 ค่าความถูกต้อง ร้อยละ 87.50 ค่าเรียกคืน ร้อยละ 87.40 และค่าประสิทธิภาพโดยรวม ร้อยละ 86.70 ซึ่งสามารถนำมาใช้เป็นแบบจำลองเพื่อทำนายแนวโน้มการสำเร็จการศึกษาของนักเรียนได้ อย่างไรก็ตามแบบจำลองที่สร้างขึ้นในการวิจัยครั้งนี้มีความเหมาะสมกับชุดข้อมูลผลการเรียนของนักเรียนแผนกวิชาคอมพิวเตอร์ธุรกิจประจำปีการศึกษา 2560 – 2561 ของวิทยาลัยอาชีวศึกษาภูเก็ตเท่านั้น อาจไม่เหมาะสมกับชุดข้อมูลอื่น เนื่องจากคุณลักษณะของแต่ละชุดข้อมูลมีความแตกต่างกัน จึงไม่สามารถนำผลลัพธ์ของแบบจำลองนี้ไปใช้สรุปผลกับชุดข้อมูลอื่น ๆ ได้ นอกจากนี้ผู้วิจัยเห็นว่าควรมีการพัฒนาเว็บแอปพลิเคชันที่ใช้อำนวยความสะดวกในการวางแผนการเรียนของนักเรียนเพื่อให้สำเร็จการศึกษาตามเกณฑ์ที่หลักสูตรกำหนดไว้

เอกสารอ้างอิง

- [1] สำนักงานคณะกรรมการการอาชีวศึกษา. (2556). นโยบาย เป้าหมาย ยุทธศาสตร์การผลิต และพัฒนากำลังคนอาชีวศึกษา พ.ศ. 2555-2569. [ออนไลน์] สืบค้นจาก <https://www.vec.go.th/> (6 สิงหาคม 2564).
- [2] ปริญา มีสุข. (2559). ปัจจัยที่ส่งผลต่อการตัดสินใจออกกลางคันของนักเรียนอาชีวศึกษา เขตภาคกลาง 1. *วารสารบัณฑิตศึกษา มหาวิทยาลัยราชภัฏวไลยอลงกรณ์ ในพระบรมราชูปถัมภ์*, 10(3), 72-83.
- [3] สนุ่น มีเพชร. (2563). ปัญหาการจัดการเรียนการสอนของสถาบันอาชีวศึกษาเกษตร. *วารสารสหวิทยาการมนุษยศาสตร์*, 3(3), 380-388.
- [4] สนุ่น มีเพชร. (2561). การแก้ปัญหาการออกกลางคันของผู้เรียนวิทยาลัยการอาชีพ สังกัดสถาบันการอาชีวศึกษา ภาคตะวันออกเฉียงเหนือ 3. *วารสารการบริหารและนิเทศการศึกษา มหาวิทยาลัยมหาสารคาม*, 9(1), 75-82.
- [5] Fauziastuti, V. T., & Rakhman, L. A. (2019). A review of students' graduation classification: A comparison of Naive Bayes classifier and K-nearest neighbour. *Atlantis Press Sarl*, 410, 219-221.
- [6] รุ่งโรจน์ บุญมา, และนิเวศ จิระวิชิตชัย. (2562). การจำแนกประเภทผู้ป่วยโรคเบาหวานโดยใช้เทคนิคเหมืองข้อมูลและการเลือกคุณลักษณะจากความสัมพันธ์ของข้อมูล. *วารสารวิชาการชาชนไทย มหาวิทยาลัยราชภัฏภูเก็ต*, 3(2), 11-19.
- [7] อนันต์ชัย ชูติภาสเจริญ, และจรัญ แสนราช. (2561). การเปรียบเทียบประสิทธิภาพของอัลกอริทึมและการคัดเลือกคุณลักษณะที่เหมาะสมเพื่อการพยากรณ์โอกาสความสำเร็จในการโอนเงินข้ามประเทศของบุคคลทั่วไป. *วารสารวิจัย มหาวิทยาลัยขอนแก่น (ฉบับบัณฑิตศึกษา) สาขามนุษยศาสตร์และสังคมศาสตร์*, 6(3), 105-113.

- [8] ภาพรณ์ เหล่าพิสัย, และจรัญ แสนราช. (2562). การวิเคราะห์การลาออกกลางคันของนักศึกษาระดับปริญญาตรีโดยใช้เทคนิควิธีการทำเหมืองข้อมูล. *วารสารวิทยาศาสตร์ แห่งมหาวิทยาลัยราชภัฏเพชรบุรี*, 16(2), 61-71.
- [9] ชนิดาภา บุญประสม, และจรัญ แสนราช. (2561). การวิเคราะห์การลาออกกลางคันของนักศึกษาระดับปริญญาตรีโดยใช้เทคนิควิธีการทำเหมืองข้อมูล. *วารสารวิชาการครุศาสตร์อุตสาหกรรม พระจอมเกล้าพระนครเหนือ*, 9(1), 142-151.
- [10] พุฒิพงศ์ เฟ็งศิริ, พันธนา ก้อนเชื้อรัตน์, ชัชฎา ขวรางกูร, และอัจฉราพรรณ คชเดช. (2560). การวิเคราะห์ความสัมพันธ์ของปัจจัยข้อมูลนักศึกษาที่มีผลต่อระดับผลการเรียนด้วยเทคนิคต้นไม้ตัดสินใจ กรณีศึกษานักศึกษามหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ. *การประชุมวิชาการระดับชาติมหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ*, 1, 463-471.
- [11] รัชพล กลัดชื่น, และจรัญ แสนราช. (2561). การเปรียบเทียบประสิทธิภาพอัลกอริทึมและการคัดเลือกคุณลักษณะที่เหมาะสมเพื่อการทำนายผลสัมฤทธิ์ทางการเรียนของนักศึกษาระดับอาชีวศึกษา. *วารสารวิจัยมหาวิทยาลัยเทคโนโลยีธัญบุรี*, 17(1), 1-10.
- [12] ภรียา ปาลวิสุทธิ. (2559). การเพิ่มประสิทธิภาพเทคนิคต้นไม้ตัดสินใจบนชุดข้อมูลที่ไม่สมดุล โดยวิธีการสุ่มเพิ่มตัวอย่างกลุ่มน้อยสำหรับข้อมูลการเป็นโรคติดเชื้ออินเทอร์เน็ต. *วารสารเทคโนโลยีสารสนเทศ*, 12(1), 54-63.
- [13] Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). *CRISP-DM-1.0 Step-by-step data mining guide*. USA: SPSS Inc.
- [14] Shearer, C. (2000). The CRISP-DM model: the new blueprint for data mining. *Journal of Data Warehousing*, 5(4), 13-22.
- [15] Bhatia, P. (2019). *Data mining and data warehousing*. UK: Cambridge University Press.
- [16] วทันญา นิลภาตระกูล, และชุตินา เบี้ยวไข่มุข. (2562). การศึกษาปัจจัยที่สัมพันธ์กับการตัดสินใจลาออกและการเปรียบเทียบประสิทธิภาพตัวแบบพยากรณ์การลาออกของพนักงาน กรณีศึกษา บริษัทประกันภัย. *วารสารวิชาการสมาคมสถาบันอุดมศึกษาเอกชนแห่งประเทศไทย ในพระราชูปถัมภ์ สมเด็จพระเทพรัตนราชสุดาฯ สยามบรมราชกุมารี*, 8(1), 46-63.
- [17] จิระนันต์ เจริญรัตน์. (2559). การวิเคราะห์ปัจจัยที่ส่งผลกระทบต่อสภาพของนักศึกษาที่มีผลการเรียนปกติโดยใช้ต้นไม้ตัดสินใจ. *วารสารมหาวิทยาลัยราชภัฏสกลนคร*, 8(2), 256-267.