การเพิ่มประสิทธิภาพตัวแบบการพยากรณ์ตามกฎด้วยเทคนิคการเรียนรู้แบบรวมกลุ่มสำหรับพยากรณ์ภาวะการมีงานทำของบัณฑิต

Improving Performance of Rule-Based Predictive Models Using Ensemble Learning Technique for Predicting Graduates' Employability

ภรัณยา ปาลวิสุทธิ์[1*] อภินันท์ จุ่นกรณ์[1] มงคล รอดจันทร์[2] และพุฒิยา รัตนศิริวัฒน์[3]

[1]สาขาวิชาวิทยาการข้อมูล คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏนครปฐม นครปฐม 73000
[2]สาขาวิชาเทคโนโลยีคอมพิวเตอร์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏนครปฐม นครปฐม 73000
[3]สาขาวิชาวิทยาศาสตร์และเทคโนโลยีการอาหาร คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏนครปฐม นครปฐม 73000

Paranya Palwisut[1*], Apinan Junkorn[1], Mongkol Rodjan[2] and Phutthiya Ratanasiriwat[3]

[1]Department of Data Science, Faculty of Science and Technology, Nakhon Pathom Rajabhat University, Nakhon Pathom 73000
[2]Department of Computer Technology, Faculty of Science and Technology, Nakhon Pathom Rajabhat University, Nakhon Pathom 73000
[3]Department of Food Science and Technology, Faculty of Science and Technology, Nakhon Pathom Rajabhat University, Nakhon Pathom 73000

***Corresponding author**: paranya@npru.ac.th

Received: 29 June 2023/ Revised: 20 September 2023/ Accepted: 27 September 2023

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพตัวแบบการพยากรณ์ตามกฎที่เพิ่มประสิทธิภาพด้วยเทคนิคแบบรวมกลุ่มสำหรับพยากรณ์ภาวะการมีงานทำของบัณฑิต การพยากรณ์ตามกฎที่นำมาใช้เปรียบเทียบในงานวิจัยจำนวน 4 เทคนิค ได้แก่ MODELM, PART, J48 และ Random Trees ได้เพิ่มประสิทธิภาพการพยากรณ์ตามกฎด้วยเทคนิคการเรียนรู้แบบรวมกลุ่มที่นำมาใช้ในงานวิจัยนี้ คือ Boosting Method แบบ AdaBoost ใช้ 10-fold cross validation ในการแบ่งข้อมูลออกเป็นชุดข้อมูลสอนและชุดข้อมูลทดสอบ และได้ใช้ค่าความถูกต้อง ค่าความแม่นยำ ค่าความระลึก และค่าความถ่วงดุล ในการวัดประสิทธิภาพการพยากรณ์ของตัวแบบ ผลการทดลองพบว่าการเพิ่มประสิทธิภาพตัวแบบการพยากรณ์ตามกฎด้วยเทคนิคการเรียนรู้แบบรวมกลุ่ม สามารถเพิ่มประสิทธิภาพให้ตัวแบบเพิ่มขึ้นในทุกเทคนิคที่ได้นำมาเปรียบเทียบ สรุปได้ว่าค่าความถูกต้องเพิ่มขึ้นเฉลี่ยร้อยละ 6.99 ค่าความแม่นยำเพิ่มขึ้นเฉลี่ยร้อยละ 6.90 ค่าความระลึกเพิ่มขึ้นเฉลี่ยร้อยละ 7.05 และค่าความถ่วงดุลเพิ่มขึ้นเฉลี่ยร้อยละ 7.15 โดยที่ตัวแบบพยากรณ์จากอัลกอริทึม J48 ที่เพิ่มประสิทธิภาพด้วย

วิธี AdaBoost เป็นตัวแบบที่มีประสิทธิภาพในการพยากรณ์สูงที่สุด โดยมีค่าความถูกต้องเท่ากับร้อยละ 89.47 ค่าความแม่นยำเท่ากับร้อยละ 89.50 ค่าความระลึกเท่ากับร้อยละ 89.50 และค่าความถ่วงดุลเท่ากับร้อยละ 89.50

คำสำคัญ: ตัวแบบตามกฎ การเรียนรู้แบบรวมกลุ่ม การมีงานทำของบัณฑิต

Abstract

This research proposes to compare the efficiency rule-based predictive models using ensemble learning techniques to improve the predictive performance for graduates' Employability. Four rule-based techniques were used: MODELM, PART, J48, and Random Trees, and predictive performance was improved using ensemble learning techniques based on Boosting Method (AdaBoost). Moreover, 10-fold cross-validation was utilized to split the data into the training and test set. This research has measured performance models with accuracy, precision, recall, and f-measure. The experimental result demonstrated that improving the performance of rule-based predictive models using ensemble learning techniques can increase performance for all techniques, that the accuracy increased by an average of 6.99%, the precision increased by an average of 6.90%, the recall increased by an average of 7.05%, and the f-measure increased by an average of 7.15%. The predictive model from the J48 algorithm combined with AdaBoost highest efficiency with accuracy 89.47%, precision 89.50%, recall 89.50%, and f-measure 89.50%.

Keywords: Rule-Based Models, Ensemble Learning, Graduates Employability

บทนำ

ปัจจุบันแต่ละปีจะมีบัณฑิตผู้สำเร็จการศึกษาเพื่อเข้าสู่ตลาดแรงงานในสาขาวิชาชีพตามกลุ่มอาชีพต่าง ๆ และตามประเภทงานที่บัณฑิตสนใจเป็นจำนวนมาก สถาบันอุดมศึกษาแต่ละแห่งได้มีการจัดเก็บข้อมูลภาวะการมีงานทำของบัณฑิต เช่น สถานะภาพการทำงาน ระยะเวลาในการหางานทำ ลักษณะอาชีพ เป็นต้น เพื่อนำข้อมูลที่ได้มาเป็นแนวทางในการพัฒนาหลักสูตรการเรียนการสอนเพื่อผลิตบัณฑิตที่มีคุณภาพ ตรงกับความต้องการของตลาดแรงงาน ซึ่งได้จัดเก็บเป็นประจำในแต่ละปี และข้อมูลที่ถูกจัดเก็บมีปริมาณมากขึ้นและข้อมูลเหล่านั้นมีองค์ความรู้สำคัญซ่อนอยู่ การนำข้อมูลมาใช้ให้เกิดประโยชน์สามารถช่วยในการพัฒนาหลักสูตรให้มีศักยภาพในการแข่งขันเพื่อตอบสนองความต้องการของตลาดแรงงานได้อย่างแท้จริง

การทำเหมืองข้อมูล (Data Mining) [1] คือการนำเทคนิคการเรียนรู้ของเครื่อง (Machine Learning) หรือวิธีการทางสถิติมาทำการวิเคราะห์ข้อมูลที่จัดเก็บไว้ในฐานข้อมูลหรือจัดเก็บในรูปแบบอื่น การวิเคราะห์แนวโน้มความสัมพันธ์ของข้อมูลเป็นความรู้ที่ถูกซ่อนอยู่ในฐานข้อมูล เพื่อนำความรู้สารสนเทศที่ได้มาใช้ในการวิเคราะห์ วางแผนการตัดสินใจในด้านต่างๆ การทำเหมืองข้อมูลนั้นมีเทคนิคที่สามารถประยุกต์ใช้ได้มากมาย ได้มีงานวิจัยด้านการศึกษาจากข้อมูลที่มีอยู่ในสถาบันอุดมศึกษาโดยนำเหมืองข้อมูลมาประยุกต์ใช้เป็นจำนวนมาก เช่น ตัวอย่างงานวิจัยของปริญญาและคณะ [2] ได้ทำการวิจัยเพื่อพัฒนาตัวแบบเทคโนโลยีสารสนเทศเกี่ยวกับการคาดการณ์อาชีพในอนาคตของบัณฑิตสาขาเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏนครราชสีมา สำราญและคณะ [3] ได้ทำการการศึกษาเทคนิคพยากรณ์อาชีพสำหรับนักศึกษาระดับปริญญาตรีสาขาคอมพิวเตอร์โดยใช้เทคนิคเหมืองข้อมูล

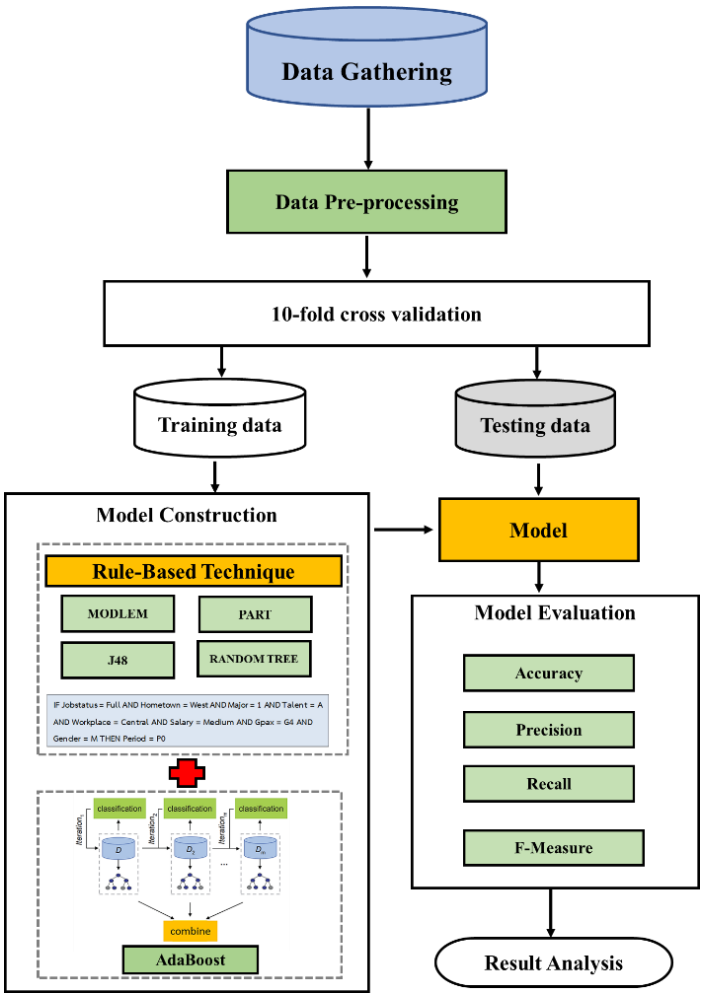
เทคนิคการพยากรณ์ตามกฎ (Rule-Based Technique) [4] ซึ่งเป็นเทคนิคหนึ่งของเหมืองข้อมูลโดยได้ใช้ชุดลำดับของกฎมาสร้างรูปแบบการแยกประเภทข้อมูล ส่วนใหญ่แล้วจะใช้กฎที่เป็น If...Then ซึ่งเป็นกฎอย่างง่าย ได้มีการนำเทคนิคการพยากรณ์ตามกฎมาใช้ในงานวิจัยด้านการศึกษา เช่น วรารินทร์ [5] ได้ประยุกต์ใช้ตัวแบบกฎการพยากรณ์ปัจจัยที่มีผลต่อการสำเร็จการศึกษาของนิสิตคณะสถาปัตยกรรมศาสตร์ ผังเมืองและนฤมิต มหาวิทยาลัยมหาสารคาม นนทวัฒน์ และคณะ [6]

ได้สร้างระบบทำนายการพ้นสภาพของนักศึกษาระดับปริญญาตรี คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏบุรีรัมย์ ด้วยเทคนิคต้นไม้ตัดสินใจ (Decision tree) มาช่วยในการสร้างกฎ เป็นต้น ในการสร้างตัวแบบกฎการพยากรณ์มีข้อดีคือง่ายต่อการแปลผลตัวแบบและมีข้อเสียคือประสิทธิภาพในการพยากรณ์ที่ไม่สูงนัก จึงได้มีงานวิจัยที่นำเสนอการรวมเทคนิคมาช่วยเพิ่มประสิทธิภาพของตัวแบบ เช่น ปพิชญาและจารี [7] ได้เพิ่มประสิทธิภาพตัวแบบตามกฎ ได้แก่ FURIA, MODLEM และ RIPPER ด้วยเทคนิคแบบรวมกลุ่ม (Ensemble Techniques) ซึ่งทำให้ค่าความถูกต้องในการพยากรณ์เพิ่มขึ้น

ดังนั้นในบทความวิจัยนี้ผู้วิจัยจึงมีแนวคิดในการใช้ประโยชน์จากข้อมูลที่ได้มีการจัดเก็บมาเข้าสู่กระบวนการทำเหมืองข้อมูล เพื่อให้ได้มาซึ่งข้อมูลที่มีประสิทธิภาพมากที่สุดด้วยการพัฒนาตัวแบบการพยากรณ์ตามกฎสำหรับพยากรณ์ภาวะการมีงานทำของบัณฑิตและเพิ่มประสิทธิภาพตัวแบบการพยากรณ์ตามกฎด้วยเทคนิคการเรียนรู้แบบรวมกลุ่มให้มีประสิทธิภาพมากยิ่งขึ้น เพื่อให้สถานศึกษาได้นำไปประยุกต์ใช้พยากรณ์ภาวการณ์มีงานทำของบัณฑิตได้อย่างแม่นยำและมีประสิทธิภาพ

วิธีดำเนินการวิจัย

การเพิ่มประสิทธิภาพตัวแบบการพยากรณ์ตามกฎด้วยเทคนิคการเรียนรู้แบบรวมกลุ่มสำหรับพยากรณ์ภาวะการมีงานทำของบัณฑิต โดยได้ดำเนินการตามกระบวนการ CRISP-DM (Cross-Industry Standard Process for Data Mining) ซึ่งเป็นกระบวนการมาตรฐานในการวิเคราะห์ข้อมูลด้านเทคนิคเหมืองข้อมูล สรุปเป็นกรอบแนวคิด ได้ดังภาพที่ 1



ภาพที่ 1 กรอบแนวคิดการวิจัย

1. การทำความเข้าใจปัญหา (Business Understanding)

คณะผู้วิจัยได้ทำการศึกษาเบื้องต้นเพื่อทำความเข้าใจปัญหาที่จะทำ และได้ทำการศึกษาเอกสารความรู้ งานวิจัยต่างๆ ที่เกี่ยวข้องกับการสร้างตัวแบบสำหรับพยากรณ์การมีงานทำของบัณฑิต ทั้งในส่วนของเทคนิคที่นำมาใช้ในการวิเคราะห์ข้อมูล โดยมีรายละเอียดดังต่อไปนี้

1.1 ศึกษาปัญหา

สถาบันอุดมศึกษาแต่ละแห่งได้มีการจัดเก็บข้อมูลภาวะการมีงานทำของบัณฑิตและข้อมูลที่ถูกจัดเก็บมีปริมาณมากขึ้นและข้อมูลเหล่านั้นมีองค์ความรู้สำคัญซ่อนอยู่ การนำข้อมูลมาใช้ให้เกิดประโยชน์สามารถช่วยในการพัฒนาหลักสูตรให้มีศักยภาพในการแข่งขันการเข้าทำงานของนักศึกษา และสามารถตอบสนองความต้องการของตลาดแรงงานได้ ซึ่งการทำเหมืองข้อมูลสามารถทำการวิเคราะห์ข้อมูลที่จัดเก็บไว้เพื่อหาองค์ความรู้ที่ถูกซ่อนอยู่ในข้อมูลมาใช้ในการวิเคราะห์ การวางแผน และการตัดสินใจในด้านต่าง ๆ ได้เป็นอย่างดี

1.2 ทฤษฎีที่เกี่ยวข้อง

1) เหมืองข้อมูล

เหมืองข้อมูล [1] เป็นหนึ่งในความรู้สมัยใหม่ที่ได้รับความนิยมอย่างมากในปัจจุบัน โดยได้ถูกประยุกต์ใช้ในแทบทุกองค์กรไม่ว่าจะเป็นภาครัฐและเอกชนเพื่อใช้ในการตัดสินใจ การวางแผนกลยุทธ์ การปรับปรุงการให้บริการและการดำเนินงานต่าง ๆ ในองค์กร เหมืองข้อมูลเป็นการทำงานที่เน้นการค้นหาสารสนเทศหรือองค์ความรู้จากข้อมูลขนาดใหญ่ เพื่อนำสิ่งที่ได้มาใช้ให้เป็นประโยชน์ โดยเหมืองข้อมูลเป็นการผสมผสานศาสตร์ทางสถิติ ปัญญาประดิษฐ์ การรู้จำ และฐานข้อมูลเข้าด้วยกัน

2) เทคนิคการพยากรณ์ตามกฎ (Rule-Based Technique)

เทคนิคการพยากรณ์ตามกฎ [7] เป็นการพยากรณ์ข้อมูลด้วยกฎ เป็นตัวแบบที่แสดงผลด้วยเซตของกฎที่มีลักษณะแบบ ‘IF-THEN’ ซึ่งกฎหนึ่งๆ จะถูกแสดงอยู่ในรูปฟอร์ม IF condition THEN conclusion จากรูปแบบฟอร์มของกฎจะประกอบไปด้วยข้อมูลสองส่วนด้วยกัน [8] คือ ส่วนแรก ส่วนเงื่อนไข ‘IF’ จะเป็นส่วนที่เรียกว่า ‘rule antecedent’ หรือ ‘precondition’ ประกอบไปด้วยเซตของแอทริบิวต่างๆประกอบกันเป็นเงื่อนไข โดยส่วนเงื่อนไขของกฎจะทำการเชื่อมโยงแอทริบิวต์ต่างๆ เข้าด้วยกันด้วยเครื่องหมาย AND เป็นการบ่งบอกว่าข้อมูลจะต้องเป็นไปตามเงื่อนไขที่ถูกกำหนดไว้ทั้งหมด ในส่วนที่สองจะเรียกว่า ‘rule consequence’ ที่จะมีหมวดหมู่ของข้อมูลบรรจุอยู่

เทคนิคการพยากรณ์ตามกฎที่นำมาใช้เปรียบเทียบในงานวิจัย จำนวน 4 เทคนิค มีดังนี้

- MODELM [9] เป็นเทคนิคที่ใช้เพื่อสร้างกฎที่กำหนดไว้สำหรับการจำแนกประเภทที่มีประสิทธิภาพและเพิ่มความถูกต้องในการจำแนกประเภทโดยการรวมตัวจำแนกประเภทต่าง ๆ อัลกอริทึมกฎการอุปนัย (Rule Induction) ที่เรียกว่าเทคนิค MODLEM ในกรอบของตัวจำแนกประเภทรวมกัน ได้แก่ bagging, n 2-classifier และ combiner aggregation

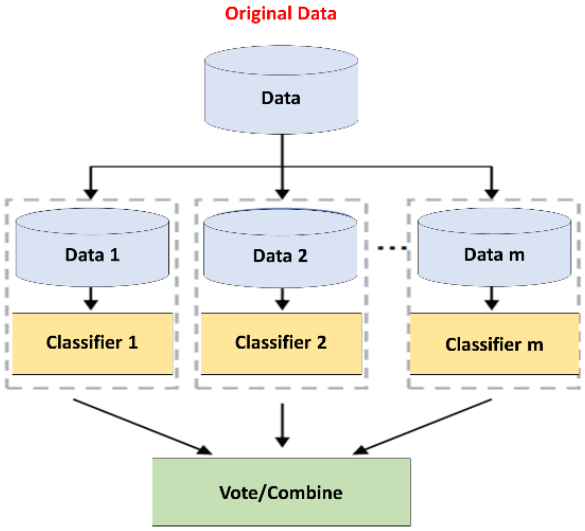
- PART [10] เป็นเทคนิคการเรียนรู้กฎที่พัฒนาจาก C4.5 และ RIPPER โดยรวมทั้ง 2 เทคนิคเข้าด้วยกัน มีจุดเด่นคือ เรียนรู้กฎได้เองเหมือนเทคนิค RIPPER และสามารถจัดการกับข้อมูลที่หายไปและคุณลักษณะทางตัวเลขที่ต่างกันได้ดี

- J48 [11] หรืออัลกอริทึมของ C4.5 เป็นอัลกอริทึมในการสร้างต้นไม้ตัดสินใจจากกลุ่มของข้อมูลฝึกสอน โดยใช้ความถูกต้องของแต่ละคุณลักษณะของข้อมูลเพื่อใช้เป็นการตัดสินใจแบ่งกลุ่มข้อมูลกลุ่มย่อย ๆ โดยพิจารณาจากค่าความแตกต่างใน Entropy ผลลัพธ์จากการเลือกคุณลักษณะสำหรับแบ่งกลุ่มข้อมูลด้วยค่า normalized information gain ที่สูงสุด

- Random Trees (RT) [12] ใช้สำหรับปัญหาการจำแนกประเภทและการถดถอย เป็นที่รวมของต้นไม้ในการพยากรณ์ (tree predictors) ที่เรียกว่าป่า เป็นหนึ่งในกลุ่มของโมเดลที่เรียกว่า Ensemble learning ที่มีหลักการคือการเทรนโมเดลที่เหมือนกันหลายๆ ครั้ง โดยหาค่าเฉลี่ยจากการพยากรณ์ของแบบจำลองที่เป็นอิสระหลายๆ แบบ

3) การเรียนรู้แบบรวมกลุ่ม (Ensemble Learning)

การเรียนรู้แบบรวมกลุ่ม [13] เป็นวิธีการรวมเอากลุ่มของตัวจำแนกประเภทข้อมูล (classifier) ที่สร้างขึ้นหลาย ๆ ตัวจำแนกและมีความเป็นอิสระต่อกันมาพิจารณาร่วมกัน เพื่อช่วยในการตัดสินใจสำหรับการหาคำตอบโดยใช้วิธีการรวม (Combine) หรือด้วยวิธีการโหวต (Voting) เพื่อให้ได้ผลลัพธ์ในการจำแนกข้อมูลที่มีประสิทธิภาพสูง ดังภาพที่ 2



ภาพที่ 2 โครงสร้างพื้นฐานของการเรียนรู้แบบรวมกลุ่ม

จากภาพที่ 2 มีขั้นตอน คือ 1) สร้างข้อมูลใหม่จากข้อมูลเดิม 2) สร้างตัวจำแนกจากข้อมูลที่สร้าง และ 3) รวมตัวจำแนกที่สร้างขึ้นเพื่อช่วยกันพิจารณาหาคำตอบ ซึ่งเทคนิคการเรียนรู้แบบรวมกลุ่มนั้น มีเทคนิคอยู่ด้วยกัน 3 แบบ คือ Bagging, Boosting และ Stacking สำหรับวิธีการที่มีประสิทธิภาพและนำมาใช้ในงานวิจัยนี้ คือ Boosting Method ซึ่งเป็นวิธีที่นิยมและใช้งานกันอย่างแพร่หลาย

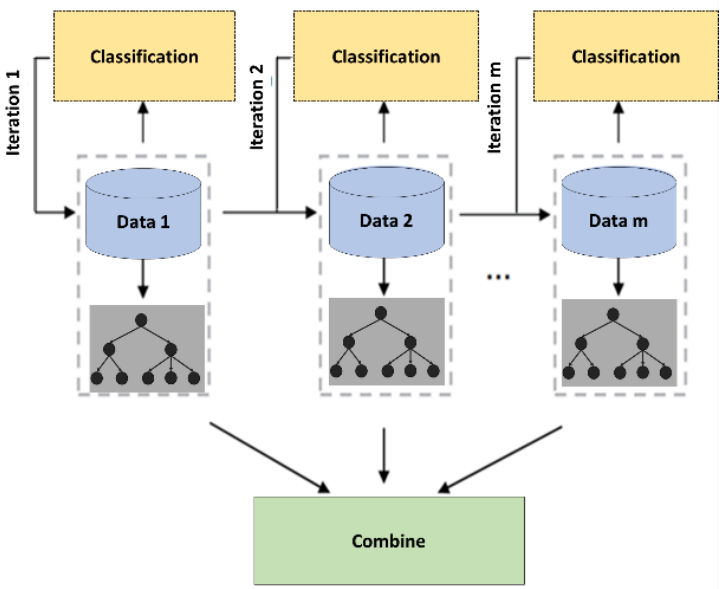
Boosting Method [14] เป็นวิธีการเรียนรู้แบบรวมกลุ่มที่นิยมนำมาสร้างตัวแบบการเรียนรู้แบบรวมกลุ่ม โดยใช้การถ่วงค่าน้ำหนักให้กับข้อมูลตัวอย่างที่ได้ทำการเรียนรู้ โดยเน้นไปที่การหาความผิดพลาดที่เกิดขึ้นจากกระบวนการเรียนรู้ข้อมูล เรียกว่า “Weak Learning” และในขั้นตอนสุดท้ายจะใช้วิธีรวมตัวจำแนกประเภทที่สร้างขึ้นหลาย ๆ ตัวจำแนกโดยพิจารณาจากค่าเฉลี่ยในการถ่วงน้ำหนัก (Mean Weight) และทำการโหวตเพื่อให้ได้ผลลัพธ์เพียงคำตอบเดียวสำหรับวิธีการของ Boosting ที่นิยมนำมาใช้ คือ Adaptive Boosting (AdaBoost) ซึ่งถูกพัฒนาขึ้นโดย Freund, Y. และ Schapire, R.E. วิธีการของ AdaBoost จะทำการถ่วงค่าน้ำหนักให้กับข้อมูลตัวอย่างที่ถูกเรียนรู้ของแต่ละรอบในการสร้างตัวแบบ โดยข้อมูลตัวอย่างที่ทำการจำแนกประเภทได้ถูกต้องจะถูกลดค่าน้ำหนักลง ส่วนข้อมูลตัวอย่างทำการจำแนกประเภทผิดพลาดจะถูกเพิ่มค่าน้ำหนักให้มีความสำคัญมากขึ้นเพื่อให้ข้อมูลนั้นมีโอกาสถูกเลือกในการเรียนรู้ของรอบต่อไป แสดงวิธีการทำงานได้ดังภาพที่ 3

2. การทำความเข้าใจข้อมูลที่จัดเก็บ (Data understanding)

เป็นขั้นตอนที่เริ่มต้นจากการเก็บรวบรวมข้อมูล หลังจากนั้นจะเป็นการตรวจสอบข้อมูลที่ได้ทำการรวบรวมมาได้เพื่อดูความถูกต้องของข้อมูล และพิจารณาว่าจะใช้ข้อมูลทั้งหมดหรือเลือกข้อมูลบางส่วนมาใช้ในการวิเคราะห์ โดยมีขั้นตอนดังนี้

2.1 การเก็บรวบรวมข้อมูล (Data gathering)

เป็นขั้นตอนที่เริ่มต้นจากการเก็บรวบรวมข้อมูลที่จะนำมาสร้างตัวแบบ ซึ่งข้อมูลที่นำมาใช้ในการสร้างตัวแบบได้ทำการรวบรวมจากข้อมูลภาวะการมีงานทำของบัณฑิตมหาวิทยาลัยราชภัฏนครปฐม ตั้งแต่ปี พ.ศ. 2557-2564 จำนวนทั้งสิ้น 13,548 ระเบียน โดยมีระเบียนที่สมบูรณ์สามารถนำมาใช้ได้จำนวน 11,971 ระเบียน และมีจำนวนคุณลักษณะ (Attribute)



ภาพที่ 3 Boosting Method แบบ AdaBoost

ทั้งหมด 20 คุณลักษณะ หลังจากนั้นจะเป็นการตรวจสอบข้อมูลที่ทำการรวบรวมมาได้ เพื่อดูความถูกต้องของข้อมูลและพิจารณาว่าจะใช้ข้อมูลทั้งหมดหรือจำเป็นต้องเลือกคุณลักษณะบางส่วนมาใช้ในการวิเคราะห์ ซึ่งคุณลักษณะที่คัดเลือกมาใช้ในการสร้างตัวแบบได้ตัดคุณลักษณะที่ไม่มีผลต่อการสร้างตัวแบบ เช่น รหัสนักศึกษา ชื่อ นามสกุล เบอร์ติดต่อ หมู่เรียน จนได้คุณลักษณะที่นำมาสร้างตัวแบบจำนวนทั้งสิ้น 13 คุณลักษณะ

2.2 การอธิบายข้อมูล (Describe data)

ทำการอธิบายข้อมูลที่นำมาสร้างตัวแบบ เช่น อธิบายถึงคุณลักษณะต่าง ๆ ในตาราง อธิบายถึงลักษณะของข้อมูลที่นำมาใช้ และทำการประเมินว่าข้อมูลที่ได้มาตรงกับความต้องการหรือไม่ แสดงรายละเอียดได้ดังตารางที่ 1

จากตารางที่ 1 คุณลักษณะที่ 13 ระยะเวลาการได้งานทำเป็นคุณลักษณะที่เป็นกลุ่มของข้อมูลที่ใช้ในการจำแนกข้อมูล โดยแบ่งออกได้ 7 กลุ่ม ได้แก่ 1) ได้งานทันที 2) ได้งานภายใน 1 – 3 เดือน 3) ได้งานภายใน 4 – 6 เดือน 4) ได้งานภายใน 7 – 9 เดือน 5) ได้งานภายใน 10 – 12 เดือน 6) ได้งานมากกว่า 1 ปี และ 7) ได้งานระหว่างศึกษา

3. การเตรียมข้อมูล (Data Preparation)

ในขั้นตอนการเตรียมข้อมูลสำหรับทำเหมืองข้อมูลที่จะนำไปใช้ในการวิเคราะห์ ผู้วิจัยได้นำข้อมูลที่ได้จากการเก็บรวบรวมข้อมูลจากขั้นตอนก่อนหน้า ดังต่อไปนี้

3.1 การทำความสะอาดข้อมูล (Cleaning data) เพื่อเพิ่มคุณภาพข้อมูลที่ถูกคัดเลือกมาใช้ในการวิเคราะห์ เช่น การแทนค่าข้อมูลที่ขาดหายไป หรือการแทนค่าเริ่มต้นที่เหมาะสมของข้อมูล

3.2 กำจัดข้อมูลที่มีความผิดปกติ (Outliers) ข้อมูลรบกวน (Noisy Data) ข้อมูลมีความไม่สอดคล้องกัน (Inconsistent Data) โดยจัดการให้มีความถูกต้องสมบูรณ์

3.3 การแปลงข้อมูล (Data Transformation) เป็นขั้นตอนการแปลงข้อมูลให้เหมาะสมสำหรับการใช้งาน ได้ทำการแปลงข้อมูลที่ได้ทำการเก็บรวบรวมมา (raw data) ให้กลายเป็นข้อมูลที่สามารถนำไปวิเคราะห์ในขั้นถัดไปได้โดยยึดการแปลงข้อมูลให้ตรงตามตารางรายละเอียดคุณลักษณะ และในงานวิจัยครั้งนี้ผู้จัดทำได้เลือกใช้โปรแกรม Weka มาใช้ในการสร้างตัวแบบซึ่งต้องจัดเตรียมแฟ้มงานให้อยู่ในรูปแบบไฟล์ arff

ตารางที่ 1 รายละเอียดคุณลักษณะ

ลำดับ	ชื่อ	คำอธิบาย
1	Gender	เพศ เช่น F = หญิง M = ชาย
2	Major	สาขาวิชา (ยึดสาขาวิชาตามระบบของ ISCED) เช่น 1 = โปรแกรมทั่วไป 2 = การศึกษา 10 = สุขภาพและสวัสดิการ
3	faculty	คณะที่ศึกษา เช่น 1 = กลุ่มสาขาวิทยาศาสตร์และเทคโนโลยี 2 = กลุ่มสาขาวิทยาศาสตร์สุขภาพ 3 = กลุ่มสาขาสังคมศาสตร์และมนุษยศาสตร์
4	Gpax	เกรดเฉลี่ยสะสม เช่น G1 = 0.00 - 2.00 G5 = 3.51 - 4.00
5	Honor	เกียรตินิยม เช่น First = เกียรตินิยมอันดับ 1 Second = เกียรตินิยมอันดับ 2 Non = สำเร็จการศึกษา
6	hometown	ภูมิภาคของจังหวัดที่เป็นภูมิลำเนา เช่น Central = ภาคกลาง West = ภาคตะวันตก
7	Jobstatus	สถานภาพการทำงาน เช่น Full = ทำงานประจำ Free = ทำงานอิสระ
8	Jobtype	ประเภทงานที่ทำ เช่น Gov = ข้าราชการ/เจ้าหน้าที่หน่วยงานรัฐ Agri = เกษตรกรรม
9	Talent	ความสามารถพิเศษ เช่น A = ภาษาต่างประเทศ B = การใช้คอมพิวเตอร์ G = การคำนวณ
10	Workplace	ภูมิภาคจังหวัดที่ได้งานทำ เช่น Central = ภาคกลาง West = ภาคตะวันตก
11	Salary	เงินเดือน/รายได้เฉลี่ยต่อเดือน เช่น Low = ต่ำกว่า 10,000 Medium = 10,001 – 30,000 High = มากกว่า 30,001 ขึ้นไป
12	Match	งานที่ทำตรงกับสาขาที่เรียน เช่น Yes = ตรง No = ไม่ตรง
13	Period	ระยะเวลาการได้งานทำ เช่น P1 = ได้งานทันที P7 = ได้งานระหว่างศึกษา

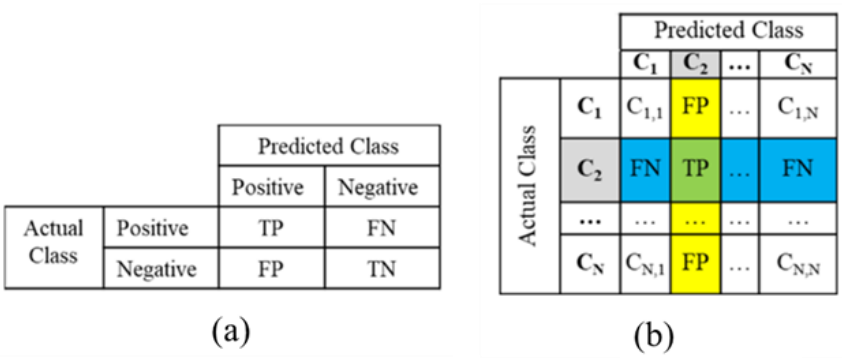
4. การพัฒนาตัวแบบพยากรณ์ (Modeling)

การพัฒนาตัวแบบพยากรณ์ ได้พัฒนาโดยทำการเปรียบเทียบเพื่อหาเทคนิคในการพยากรณ์ภาวะการมีงานทำที่เหมาะสมกับข้อมูลที่ให้ประสิทธิภาพที่ดีสุด การพยากรณ์ตามกฎที่นำมาใช้เปรียบเทียบในงานวิจัย จำนวน 4 เทคนิค ได้แก่ MODELM PART J48 และ Random Trees ได้เพิ่มประสิทธิภาพการพยากรณ์ตามกฎด้วยเทคนิคการเรียนรู้แบบรวมกลุ่มที่นำมาใช้ในงานวิจัยนี้ คือ Boosting Method แบบ AdaBoost

5. การตรวจสอบประสิทธิภาพตัวแบบ (Evaluation)

ในงานวิจัยนี้ใช้การทดสอบความแม่นตรงของตัวแบบ ด้วยเทคนิค 10-fold Cross-Validation เพื่อให้ข้อมูลทุกตัวมีโอกาสเป็นชุดทดสอบและชุดสอน โดยแบ่งข้อมูลออกเป็นชุดข้อมูลสอนและข้อมูลทดสอบ โดยแบ่งข้อมูลออกเป็น 10 ส่วนเท่าๆ กัน ใช้ 9 ส่วนเป็นชุดข้อมูลสอน และอีก 1 ส่วนเป็นชุดข้อมูลทดสอบ ซึ่งจะทำสลับกันจนครบทั้งหมด 10 รอบ แล้วทำการวัดประสิทธิภาพของตัวแบบของแต่ละเทคนิคมาทำการทดลองเปรียบเทียบ โดยวัดจากค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าความถ่วงดุล (F-measure) เพื่อหาตัวแบบที่ประสิทธิภาพดีที่สุด

การตรวจสอบประสิทธิภาพของตัวแบบในแต่ละเทคนิควิธีที่นำมาเปรียบเทียบ โดยใช้ค่าความถูกต้องของตัวแบบจะต้องทำการเลือกข้อมูลสำหรับเรียนรู้ (Training Set) และข้อมูลสำหรับทดสอบ (Testing Set) งานวิจัยนี้เลือกใช้วิธีการสุ่มเลือกข้อมูลแบบความเที่ยงตรง k กลุ่ม (k-fold Cross Validation) การวัดค่าประสิทธิภาพของตัวแบบนั้นจะอาศัย Confusion matrix [15] ในการคำนวณค่า แสดงดังภาพที่ 4



ภาพที่ 4 Confusion matrix [16]

โดยที่ True Positive (TP) เป็นจำนวนข้อมูลประเภทเป็นบวก และถูกจำแนกประเภทว่าเป็นบวก

False Positive (FP) เป็นจำนวนข้อมูลประเภทเป็นลบ แต่ถูกจำแนกประเภทว่าเป็นบวก

True Negative (TN) เป็นจำนวนข้อมูลประเภทเป็นลบ และถูกจำแนกประเภทว่าเป็นลบ

False Negative (FN) เป็นจำนวนข้อมูลประเภทเป็นบวก แต่ถูกจำแนกประเภทว่าเป็นลบ

ตัวชี้วัดที่สามารถคำนวณได้จาก Confusion Matrix

- ค่าความแม่นยำ เป็นการวัดความแม่นยำของตัวแบบ โดยพิจารณาแยกทีละคลาส ดังสมการที่ (1)

$$\text{Precision} = TP / (TP+FP) \quad (1)$$

- ค่าความระลึก เป็นการวัดความถูกต้องของตัวแบบ โดยพิจารณาแยกทีละคลาส ดังสมการที่ (2)

$$\text{Recall} = TP / (TP + FN) \quad (2)$$

- ค่าความถ่วงดุล เป็นการวัดค่าความแม่นยำและค่าความระลึกพร้อมกันของตัวแบบ โดยพิจารณาแยกทีละคลาส ดังสมการที่ (3)

F-measure = (2 x Precision x Recall) / (Precision + Recall) (3)

- ค่าความถูกต้อง เป็นการวัดความถูกต้องของตัวแบบ โดยพิจารณารวมทุกคลาส คือ จำนวน True Positive ของทุกคลาสรวมกัน ดังสมการที่ (4)

Accuracy = (TP + TN) / (TP + TN + FP + FN) (4)

6. การนำไปใช้งาน (Deployment)

นำตัวแบบที่ดีที่สุดจากการเปรียบเทียบประสิทธิภาพตัวแบบการพยากรณ์ตามกฎที่ได้เพิ่มประสิทธิภาพด้วยเทคนิคการเรียนรู้แบบรวมกลุ่มสำหรับพยากรณ์การมีงานทำของบัณฑิตไปใช้งานในสถานศึกษาเพื่อสนับสนุนการวางแผนและตัดสินใจ

ผลการวิจัย

การพัฒนาตัวแบบการพยากรณ์ตามกฎที่ได้เพิ่มประสิทธิภาพด้วยเทคนิคการเรียนรู้แบบรวมกลุ่มสำหรับพยากรณ์การมีงานทำของบัณฑิต การพยากรณ์ตามกฎที่นำมาใช้เปรียบเทียบในงานวิจัย จำนวน 4 เทคนิค ได้แก่ MODELM, PART, J48 และ Random Trees ได้เพิ่มประสิทธิภาพการพยากรณ์ตามกฎด้วยเทคนิคการเรียนรู้แบบรวมกลุ่มที่นำมาใช้ในงานวิจัยนี้ คือ Boosting Method แบบ AdaBoost ทำการตรวจสอบประสิทธิภาพของตัวแบบจากค่าความถูกต้อง ค่าความแม่นยำ ค่าความระลึก และค่าความถ่วงดุล โดยได้ทำการเปรียบเทียบการพยากรณ์ตามกฎแบบดั้งเดิมและการพยากรณ์ตามกฎที่ได้เพิ่มประสิทธิภาพการพยากรณ์ตามกฎด้วยเทคนิคการเรียนรู้แบบรวมกลุ่ม สรุปได้ดังตารางที่ 2

ตารางที่ 2 ผลการตรวจสอบประสิทธิภาพในแต่ละอัลกอริทึม (ร้อยละ)

Algorithm	Accuracy	Precision	Recall	F-measure
Rule-Based Technique				
MODELM	68.42	70.40	68.40	60.70
PART	77.19	77.00	77.20	77.10
J48	73.78	74.80	73.70	74.00
RT	78.94	78.90	78.90	78.90
Rule-Based Technique + Boosting Method				
MODELM + AdaBoost	70.16	73.00	70.20	63.70
PART + AdaBoost	80.70	80.40	80.70	80.30
J48 + AdaBoost	***89.47***	***89.50***	***89.50***	***89.50***
RT + AdaBoost	85.96	85.80	86.00	85.80

เทคนิค MODELM ได้ผลการวัดประสิทธิภาพดังนี้ ค่าความถูกต้อง เท่ากับร้อยละ 68.42 ค่าความแม่นยำเท่ากับร้อยละ 70.40 ค่าความระลึกเท่ากับร้อยละ 68.40 และค่าความถ่วงดุลเท่ากับร้อยละ 60.70 ซึ่งเทคนิค MODELM ที่เพิ่มประสิทธิภาพด้วยวิธี AdaBoost ได้ดังนี้ ค่าความถูกต้อง เท่ากับร้อยละ 70.16 ค่าความแม่นยำเท่ากับร้อยละ 73.00 ค่าความระลึกเท่ากับร้อยละ 70.20 และค่าความถ่วงดุลเท่ากับร้อยละ 63.70

เทคนิค PART ได้ผลการวัดประสิทธิภาพดังนี้ ค่าความถูกต้องเท่ากับร้อยละ 77.19 ค่าความแม่นยำเท่ากับร้อยละ 77.00 ค่าความระลึกเท่ากับร้อยละ 77.20 และค่าความถ่วงดุลเท่ากับร้อยละ 77.10 ซึ่งเทคนิค PART ที่เพิ่มประสิทธิภาพด้วยวิธี AdaBoost ได้ดังนี้ ค่าความถูกต้อง เท่ากับร้อยละ 80.70 ค่าความแม่นยำเท่ากับร้อยละ 80.40 ค่าความระลึกเท่ากับร้อยละ 80.70 และค่าความถ่วงดุลเท่ากับร้อยละ 80.30

เทคนิค J48 ได้ผลการวัดประสิทธิภาพดังนี้ ค่าความถูกต้อง เท่ากับร้อยละ 73.78 ค่าความแม่นยำเท่ากับร้อยละ 74.80 ค่าความระลึกเท่ากับร้อยละ 73.70 และค่าความถ่วงดุลเท่ากับร้อยละ 74.00 ซึ่งเทคนิค J48 ที่เพิ่มประสิทธิภาพด้วยวิธี AdaBoost ได้ดังนี้ ค่าความถูกต้อง เท่ากับร้อยละ 89.47 ค่าความแม่นยำเท่ากับร้อยละ 89.50 ค่าความระลึกเท่ากับร้อยละ 89.50 และค่าความถ่วงดุลเท่ากับร้อยละ 89.50

เทคนิค Random Tree ได้ผลการวัดประสิทธิภาพดังนี้ ค่าความถูกต้องเท่ากับร้อยละ 78.94 ค่าความแม่นยำเท่ากับร้อยละ 78.90 ค่าความระลึกเท่ากับร้อยละ 78.90 และค่าความถ่วงดุลเท่ากับร้อยละ 78.90 ซึ่งเทคนิค Random Tree ที่เพิ่มประสิทธิภาพด้วยวิธี AdaBoost ได้ดังนี้ ค่าความถูกต้องเท่ากับร้อยละ 85.96 ค่าความแม่นยำเท่ากับร้อยละ 85.80 ค่าความระลึกเท่ากับร้อยละ 85.80 และค่าความถ่วงดุลเท่ากับร้อยละ 85.80

เมื่อนำค่าความถูกต้องมาเปรียบเทียบเพื่อหาอัลกอริทึมที่มีประสิทธิภาพสูงสุดสามารถแสดงได้ดังตารางที่ 3

ตารางที่ 3 การเปรียบเทียบค่าความถูกต้องของตัวแบบ (ร้อยละ)

Algorithm	Accuracy	Algorithm	Accuracy	Improvement
MODELM	68.42	MODELM + Boost	70.16	1.74
PART	77.19	PART + Boost	80.70	3.51
J48	73.78	J48 + Boost	***89.47***	15.69
RT	78.94	RT + Boost	85.96	7.02

จากการเปรียบเทียบค่าความถูกต้องของตัวแบบ พบว่าตัวแบบที่ได้จากอัลกอริทึม J48 ที่เพิ่มประสิทธิภาพด้วยวิธี AdaBoost มีประสิทธิภาพสูงที่สุด โดยมีค่าความถูกต้องเท่ากับร้อยละ 89.47 ซึ่งเพิ่มขึ้นถึงร้อยละ 15.69 ดังนั้น คณะผู้วิจัยจึงเลือกตัวแบบจากอัลกอริทึม J48 ที่เพิ่มประสิทธิภาพด้วยวิธี AdaBoost ไปพัฒนาตัวแบบพยากรณ์ภาวะการมีงานทำของบัณฑิตต่อไป และตัวอย่างกฎที่ได้จากตัวแบบสามารถแสดงให้เห็นถึงความง่ายในการพยากรณ์ ดังภาพที่ 5

```
IF Jobstatus = Full AND Hometown = West AND Major = 1 AND Talent = A
AND Workplace = Central AND Salary = Medium AND Gpax = G4 AND
Gender = M THEN Period = P0
```

ภาพที่ 5 ตัวอย่างกฎที่ได้จากอัลกอริทึม J48 ที่เพิ่มประสิทธิภาพด้วยวิธี AdaBoost

จากภาพที่ 5 ตัวอย่างกฎที่ได้จากอัลกอริทึม J48 ที่เพิ่มประสิทธิภาพด้วยวิธี AdaBoost สามารถแปลกฎได้ดังนี้

“ถ้าสถานภาพการทำงานเป็นงานประจำและภูมิลำเนาอยู่ภาคตะวันตกและความสามารถพิเศษคือภาษาต่างประเทศ และภูมิภาคที่ได้งานทำคือภาคกลาง และเงินเดือนอยู่ระหว่าง 10,001-30,000 และเกรดเฉลี่ยสะสมอยู่ระหว่าง 3.01 - 3.50 พยากรณ์ได้ว่าระยะเวลาการได้งานทำคือ ได้งานทันทีหลังจากจบการศึกษา”

อภิปรายและสรุปผลการวิจัย

บทความวิจัยนี้ได้เสนอแนวคิดพัฒนาตัวแบบการพยากรณ์ตามกฎสำหรับพยากรณ์ภาวการณ์มีงานทำของบัณฑิตและเพิ่มประสิทธิภาพตัวแบบการพยากรณ์ตามกฎด้วยเทคนิคการเรียนรู้แบบรวมกลุ่มให้มีประสิทธิภาพมากยิ่งขึ้น เพื่อให้สถานศึกษาได้นำไปประยุกต์ในการพยากรณ์ภาวะการมีงานทำของบัณฑิตได้อย่างแม่นยำและมีประสิทธิภาพ และการศึกษาได้

ใช้ข้อมูลภาวะการมีงานทำของบัณฑิตมหาวิทยาลัยราชภัฏนครปฐม จำนวน 11,971 คน โดยมีคุณลักษณะทั้งหมด 13 คุณลักษณะ ได้แก่ เพศ สาขาวิชา คณะที่ศึกษา เกรดเฉลี่ยสะสม เกียรตินิยม ภูมิภาคของจังหวัดที่เป็นภูมิลำเนา สถานภาพการทำงาน ประเภทงานที่ทำ ความสามารถพิเศษ ภูมิภาคจังหวัดที่ได้งานทำ เงินเดือน/รายได้เฉลี่ยต่อเดือน งานที่ทำตรงกับสาขาที่เรียน และระยะเวลาการได้งานทำ โดยได้เปรียบเทียบประสิทธิภาพเพื่อหาเทคนิคในการพยากรณ์ภาวะการมีงานทำที่เหมาะสมกับข้อมูลที่ให้ประสิทธิภาพดีที่สุด โดยการพยากรณ์ตามกฎที่นำมาใช้เปรียบเทียบในงานวิจัย จำนวน 4 เทคนิค ได้แก่ MODELM, PART, J48 และ Random Trees ซึ่งได้เพิ่มประสิทธิภาพการพยากรณ์ตามกฎด้วยเทคนิคการเรียนรู้แบบรวมกลุ่มที่นำมาใช้ในงานวิจัยนี้ คือ Boosting Method แบบ AdaBoost พบว่าการเพิ่มประสิทธิภาพตัวแบบการพยากรณ์ตามกฎด้วยเทคนิคการเรียนรู้แบบรวมกลุ่ม Boosting Method แบบ AdaBoost สามารถเพิ่มประสิทธิภาพให้ตัวแบบเพิ่มขึ้นในทุกเทคนิคที่ได้นำมาเปรียบเทียบ เมื่อพิจารณาเป็นด้านสามารถสรุปได้ว่า ค่าความถูกต้องเพิ่มขึ้นเฉลี่ยร้อยละ 6.99 ค่าความแม่นยำเพิ่มขึ้นเฉลี่ยร้อยละ 6.90 ค่าความระลึกเพิ่มขึ้นเฉลี่ยร้อยละ 7.05 และค่าความถ่วงดุลเพิ่มขึ้นเฉลี่ยร้อยละ 7.15

ตัวแบบพยากรณ์จากอัลกอริทึม J48 ที่เพิ่มประสิทธิภาพด้วยวิธี AdaBoost เป็นตัวแบบที่มีประสิทธิภาพในการคาดการณ์สูงที่สุด เนื่องจากอัลกอริทึม J48 จะทำงานวนซ้ำในการเลือกคุณลักษณะที่เหมาะสมให้เป็นโหนดตัดสินใจไปเรื่อย ๆ จากหลักการของ Information Entropy จนกระทั่งโหนดสุดท้ายของแต่ละเส้นทาง ซึ่งหมายถึงคลาสที่เหมาะสมสำหรับเส้นทางการตัดสินใจแต่ละเส้นทาง และเมื่อนำอัลกอริทึม J48 มารวมกลุ่มเพื่อช่วยในการตัดสินใจสำหรับการหาคำตอบด้วยวิธีการของ AdaBoost โดยจะทำการถ่วงค่าน้ำหนักให้กับข้อมูลตัวอย่างที่ถูกเรียนรู้ของแต่ละรอบในการสร้างตัวแบบ โดยข้อมูลตัวอย่างที่ทำการจำแนกประเภทได้ถูกต้องจะถูกลดค่าน้ำหนักลง ส่วนข้อมูลตัวอย่างที่ทำการจำแนกประเภทผิดพลาดจะถูกเพิ่มค่าน้ำหนักให้มีความสำคัญมากขึ้น เพื่อให้ข้อมูลนั้นมีโอกาสถูกเลือกในการเรียนรู้ของรอบต่อไปทำให้ผลการหาคำตอบมีประสิทธิภาพมากยิ่งขึ้น ซึ่งงานวิจัยนี้สอดคล้องกับงานวิจัยอื่น เช่น วรารินทร์ [5] ที่ได้นำตัวแบบการพยากรณ์ตามกฎมาพยากรณ์การสำเร็จการศึกษาระดับปริญญาตรีโดยใช้เทคนิคในการเปรียบเทียบได้แก่ Decision Table, Decision Rule, RIPPER และ OneR ผลการทดลองพบว่าเทคนิค RIPPER มีความเหมาะสมมากที่สุด ปพิชญาและจารี [7] ได้เพิ่มประสิทธิภาพตัวแบบตามกฎ ได้แก่ FURIA, MODLEM และ RIPPER ด้วยเทคนิคแบบรวมกลุ่ม ซึ่งทำให้เพิ่มค่าความถูกต้องในการพยากรณ์เพิ่มขึ้น เป็นต้น จากผลการวิจัยนี้ตัวแบบที่ได้เหมาะสมต่อการนำไปประยุกต์ใช้สำหรับพัฒนาระบบพยากรณ์ภาวะการมีงานทำของบัณฑิตต่อไป

ข้อเสนอแนะ

ตัวแบบที่ผู้วิจัยได้นำเสนอนั้นได้มุ่งเน้นที่การสร้างตัวแบบพยากรณ์ตามกฎด้วยเทคนิคการเรียนรู้แบบรวมกลุ่ม โดยใช้ Boosting Method แบบ AdaBoost ซึ่งนอกเหนือจากเทคนิคที่นำเสนอนี้ยังมีวิธีการอื่น ๆ อีกมากสำหรับการสร้างแบบจำลองการจำแนกประเภทแบบรวมกลุ่ม ตัวอย่าง เช่น Stacking, Voting, Random Subspace และ Hybrid Experts เป็นต้น รวมถึงการนำอัลกอริทึมใหม่ ๆ มาประยุกต์ใช้สำหรับกระบวนการในการจำแนกประเภทของข้อมูล สามารถเพิ่มประสิทธิภาพความถูกต้องให้กับการจำแนกประเภทได้ เช่น อัลกอริทึมโครงข่ายประสาทเทียม (Artificial Neural Networks) แบบ Deep Learning รวมไปถึงการนำเอาอัลกอริทึมแบบผสมผสาน (Hybrid) มาประยุกต์ใช้เพื่อเพิ่มประสิทธิภาพให้กับงานด้านการจำแนกประเภทของข้อมูล

เอกสารอ้างอิง

1. จรัสศรี รุ่งรัตนาอุบล. เทคนิคเหมืองข้อมูล (DATA MINING TECHNIQUES). พิษณุโลก: สำนักพิมพ์มหาวิทยาลัยนเรศวร; 2566.

2. ปริญญา ชินจอหอ, สายสุนีย์ จับโจร, เบญจภัค จงหมื่นไวย์. การพัฒนาตัวแบบเทคโนโลยีสารสนเทศในการคาดการณ์อาชีพอนาคตของบัณฑิตสาขาเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏนครราชสีมา. วารสารวิชาการ TM 2564;8(2):20–32.
3. สำราญ วานนท์, ธรัช อารีราษฎร์, จรัญ แสนราช. การศึกษาเทคนิคพยากรณ์อาชีพสำหรับนักศึกษาระดับปริญญาตรีสาขาคอมพิวเตอร์โดยใช้เทคนิคเหมืองข้อมูล. วารสารวิชาการ TM 2561;5(1):164–71.
4. สายชล สินสมบูรณ์ทอง. การเปรียบเทียบประสิทธิภาพในการทำนายผลค่านอกเกณฑ์ด้วยการจำแนก 6 วิธี. TJST 2020;9(3): 255–68
5. วราริญทร์ ปัญญาวงษ์. การประยุกต์ใช้ตัวแบบกฎการพยากรณ์ปัจจัยที่มีผลต่อการสำเร็จการศึกษาของนิสิต คณะสถาปัตยกรรมศาสตร์ ผังเมืองและนฤมิตศิลป์ มหาวิทยาลัยมหาสารคาม โดยการทำเหมืองข้อมูล. JADC 2020;2(2):109-19.
6. นนทวัฒน์ ทวีชาติ, อรยา เพ็งประจญ, วิไลรัตน์ ยาทองไชย, ชูศักดิ์ ยาทองไทย. ระบบทำนายการพ้นสภาพของนักศึกษาระดับปริญญาตรี คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏบุรีรัมย์ ด้วยเทคนิคการทำเหมืองข้อมูล. วารสารวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏบุรีรัมย์ 2563;4(1):47–60.
7. ปพิชญา กลางนอก, จารี ทองคำ. การประยุกต์ใช้เทคนิคแบบรวมเพื่อเพิ่มประสิทธิภาพตัวแบบตามกฎในเหมืองข้อมูล. Journal of Ind Tech UBRU 2562;9(1):97–108.
8. ภรัณยา ปาลวิสุทธิ์, อภินันท์ จุ่นกรณ์, มงคล รอดจันทร์, ธานิล ม่วงพูล. การพัฒนาระบบวิเคราะห์ข้อมูลเศรษฐกิจชุมชนเพื่อสนับสนุนการตัดสินใจด้วยเทคนิคเหมืองข้อมูล. วารสารวิชาการ ITM 2563;7(2):72–83.
9. ปพิชญา กลางนอก. การประยุกต์ใช้เทคนิคแบบรวมเพื่อเพิ่มประสิทธิภาพของแบบจำลองตามกฎ. วิทยานิพนธ์ปริญญาวิทยา ศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ, มหาวิทยาลัยมหาสารคาม. มหาสารคาม; 2561.
10. ผดุง นันอำไพ, จารี ทองคำ. การตรวจจับการบุกรุกด้วยเทคนิคการจำแนกในการทำเหมืองข้อมูล. วารสารวิชาการ ITM 2562;6(2):111-8.
11. Panigrahi R, Borah S. Rank Allocation to J48 Group of Decision Tree Classifiers using Binary and Multiclass Intrusion Detection Datasets. Procedia Comput Sci 2018;132(16):323-32.
12. Addario-Berry L, Brandenberger A, Hamdan J, Kerriou C. Universal height and width bound for random trees. Electron J Probab 2022;27:1-24.
13. อัศวิน สุรวัชโยธิน, วรภัทร ไพรีเกรง. การสร้างตัวแบบการทำนายในการเลือกศึกษาต่อในระดับอุดมศึกษา โดยใช้เทคนิคแบบบูรณาการในการแก้ปัญหาการจำแนกข้อมูลไม่สมดุลของกลุ่มผู้เรียน. JIST 2564;11(1):65–74.
14. Tanha J, Abdi Y, Samadi N, Razzaghi N, Asadpour M. Boosting methods for multi-class imbalanced data classification: An experimental review. J Big Data 2020;7(1):70.
15. Markoulidakis JG, Kopsiaftis G, Rallis I, Georgoulas I. Multi-Class Confusion Matrix Reduction method and its application on Net Promoter Score classification problem. Technologies 2021;9(4):81.