

# การเปรียบเทียบประสิทธิภาพการจำแนกข้อมูลเพื่อทำนายการได้รับทุนการศึกษาของ นักศึกษาระดับปริญญาตรี โดยใช้เทคนิควิธีการทำเหมืองข้อมูล

## The Comparison of Data Classification Efficiency to Predict Scholarship Awarded for Undergraduate Students by Data Mining Techniques

พีรศุภมภ์ ทองพ่วง<sup>1\*</sup> และ จริญญา แสนราช<sup>2</sup>  
Peerasut Tongpuang<sup>1\*</sup> and Charun Sanrach<sup>2</sup>

### บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อวิเคราะห์ปัจจัยที่เกี่ยวข้องในการได้รับทุนการศึกษาของนักศึกษาระดับปริญญาตรีและเพื่อเปรียบเทียบประสิทธิภาพการจำแนกข้อมูลของตัวแบบด้วยเทคนิควิธีต้นไม้ตัดสินใจ(Decision Tree) วิธีแบบเบย์(Naïve Bayes) วิธีเคเนียร์เรสเนเบอร์(K-Nearest Neighbors) วิธีการเรียนรู้เชิงลึก(Deep Learning) และวิธีแรนดอมฟอว์เรส(Random Forest) โดยใช้ข้อมูลจากการขอทุนการศึกษาของนักศึกษาระดับปริญญาตรีคณะบริหารธุรกิจและคณะศิลปศาสตร์จากกองบริหารงานบพิตรพิมุขจักรวรรดิ มหาวิทยาลัยเทคโนโลยีราชมงคลรัตนโกสินทร์พื้นที่บพิตรพิมุขจักรวรรดิ ระหว่างปีการศึกษา 2557–2561 มีจำนวน 15 แอตทริบิวต์ และข้อมูลจำนวน 1,155 ชุด การวิเคราะห์ข้อมูลใช้โปรแกรม Rapid Miner 9.3 ในการวิเคราะห์ปัจจัยที่เกี่ยวข้องและทดสอบประสิทธิภาพด้วยวิธี 10-Fold Cross Validation กับตัวแบบ ผลการวิจัยพบว่าข้อมูลที่มีความเกี่ยวข้องกันมีจำนวน 11 แอตทริบิวต์ และการจำแนกประเภทที่มีค่าความถูกต้องสูงสุดโดยวิธีแรนดอมฟอว์เรส(Random Forest) มีค่าร้อยละ 94.28 รองลงมาคือวิธีการเรียนรู้เชิงลึก(Deep Learning) มีค่าความถูกต้องร้อยละ 93.51 วิธีต้นไม้ตัดสินใจ(Decision Tree) มีค่าความถูกต้องร้อยละ 92.64 วิธีแบบเบย์(Naïve Bayes) มีค่าความถูกต้องร้อยละ 92.47 และวิธีที่ถูกต้องน้อยที่สุดคือวิธีเคเนียร์เรสเนเบอร์(K-Nearest Neighbors) มีค่าความถูกต้องร้อยละ 89.70 ตามลำดับ งานวิจัยนี้สามารถนำไปใช้ทำนายเพื่อวิเคราะห์แนวทางในการเสนอแนะและพัฒนาระบบสารสนเทศเพื่อใช้ประกอบการพิจารณาในการให้ทุนการศึกษาแก่นักศึกษาต่อไป

คำสำคัญ: ทุนการศึกษา/ ต้นไม้ตัดสินใจ/ เบย์/ เคเนียร์เรสเนเบอร์/ เรียนรู้เชิงลึก/ แรนดอมฟอว์เรส

### Abstract

This research analyzed factors involved with the scholarship data of undergraduate students and compared the performance of data mining techniques used to analyze the data: Decision Tree, Naïve Bayes, K-Nearest Neighbors, Deep Learning and Random Forest. The data was collected from the scholarship form of undergraduate students in the Faculty of Business Administration and the Faculty of Art from Borpitpimuk Chakrawad Division, Rajamangala University of Technology Rattanakosin during 2014 to 2018 school years. The dataset had 15 attributes and 1,155 records. The RapidMiner Studio program was used to find the number

<sup>1</sup> กองบริหารงานบพิตรพิมุขจักรวรรดิ มหาวิทยาลัยเทคโนโลยีราชมงคลรัตนโกสินทร์

Borpitpimuk Chakrawad Division, Rajamangala University of Technology Rattanakosin.

<sup>2</sup> ภาควิชาคอมพิวเตอร์ศึกษา คณะครุศาสตร์อุตสาหกรรม มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

Department of Computer Education, Faculty of Technical Education, King Mongkut's University of Technology North Bangkok.

\*Corresponding author: peerasut.tho@rmutr.ac.th

Received : 8 กุมภาพันธ์ 2564 / Revised : 15 มิถุนายน 2564 / Accepted : 18 มิถุนายน 2564

of relations attributes and model, and the 10-Fold Cross Validation to evaluate the model. The results showed that 11 attributes were involved. The accuracy of Random Forest, Deep Learning, Decision Tree, Naïve Bayes and K-Nearest Neighbors were at 94.28%, 93.51%, 92.64%, 92.47%, and 89.70%, respectively. The research could be used to predict, to analyze recommendations and to develop an information system to support the decisions in awarding scholarships to students.

**Keyword:** Scholarship/ Decision Tree/ Naïve Bayes/ K-Nearest Neighbors/ Deep Learning/ Random Forest.

## 1. บทนำ

จากสภาพเศรษฐกิจในปัจจุบันประเทศไทยกำลังเผชิญหน้ากับสถานการณ์สำคัญสองประการ ได้แก่ การเปลี่ยนแปลงโครงสร้างประชากรไปสู่สังคมผู้สูงอายุ และสถานการณ์ความเหลื่อมล้ำทั้งในทางเศรษฐกิจและด้านสังคม ซึ่งจะเป็นอุปสรรคทำให้ประเทศไทยไม่สามารถก้าวไปสู่ประเทศรายได้สูงได้ การแก้ไขปัญหาและอุปสรรคคือการทำให้ประชาชนสามารถเข้าถึงการศึกษาที่มีคุณภาพอย่างทั่วถึง เพื่อให้คนไทยมีศักยภาพและผลิตภาพ การสร้างความเสมอภาคทางการศึกษาเพื่อเพิ่มผลิตภาพทางเศรษฐกิจของบุคคล ลดความเหลื่อมล้ำทางเศรษฐกิจและสังคม [1] สำหรับเด็กที่อยู่ในครอบครัวซึ่งขาดแคลนทุนทรัพย์ และด้อยโอกาสทางการศึกษา มีทั้งกลุ่มเด็กแรกเกิดจนถึงกลุ่มวัยที่อายุ 18 ปีขึ้นไป ซึ่งต้องการกลไกที่เข้ามาช่วยเหลือให้มีความพร้อมและเพิ่มโอกาสทางการศึกษาได้มากขึ้นเพื่อเป็นแรงงานที่มีคุณภาพของประเทศในอนาคต [2] จากผลกระทบด้านเศรษฐกิจมีแนวโน้มเพิ่มความเหลื่อมล้ำทางการศึกษา เด็กยากจนจะได้รับผลกระทบทางเศรษฐกิจอย่างมาก ทั้งครอบครัวที่ผู้ปกครองตกงาน ผู้ที่มีรายได้ลดน้อยลง หรือมีภาระพึ่งพิงมากขึ้น ล้วนกระทบต่อการศึกษาลูกหลาน และทำให้แนวโน้มของความเหลื่อมล้ำระหว่างคนจนกับคนรวยมีมากขึ้น เพราะอาจจะทำให้เด็กต้องขาดเรียนมากขึ้นหรือครอบครัวมีรายจ่ายเพื่ออุดหนุนการศึกษาอันน้อยลง [3] ซึ่งส่งผลกระทบต่อการศึกษาของผู้เรียน

จากปัญหาที่มีแนวโน้มที่จะเกิดขึ้นกับนักศึกษา มหาวิทยาลัยเทคโนโลยีราชมงคลรัตนโกสินทร์พื้นที่บึงพระพิบูลย์ จักรวรรดิ จึงได้ดำเนินการจัดตั้งกองบริหารงานบึงพระพิบูลย์ จักรวรรดิ โดยมีส่วนงานสวัสดิการนักศึกษา แผนกกองทุนกู้ยืมเพื่อการศึกษา และทุนการศึกษาขึ้นมาเพื่อบรรเทาปัญหาที่เกิดขึ้นกับนักศึกษาระดับปริญญาตรีที่ทำการศึกษาในอยู่ในพื้นที่ประกอบด้วยคณะบริหารธุรกิจและคณะศิลปศาสตร์ มี

หน้าที่รับผิดชอบในการจัดหาทุนการศึกษาภายนอกมหาวิทยาลัย ได้แก่ ศิษย์เก่า ห้างร้านหรือบริษัทเอกชนต่าง ๆ ในแต่ละปีการศึกษา จัดสรรและคัดเลือกทุนการศึกษาให้กับนักศึกษาที่ยื่นขอความช่วยเหลือด้านทุนการศึกษา ในขั้นการคัดเลือกนักศึกษาที่สมควรได้รับการพิจารณาได้รับทุนนั้น นักศึกษาต้องทำการส่งแบบฟอร์มการขอทุนการศึกษาประกอบด้วยข้อมูลส่วนตัวของนักศึกษา สถานะครอบครัว รายได้ของครอบครัวต่อเดือน ระดับคะแนนผลการเรียนปัจจุบัน และข้อมูลที่จำเป็นประกอบการพิจารณาความเหมาะสม จากนั้นจึงส่งข้อมูลทั้งหมดให้กับคณะกรรมการกองทุนสัมมาชีพพร้อมพิจารณาจากข้อมูลในเอกสารและข้อมูลรายละเอียดเพิ่มเติมจากคำสัมภาษณ์ของนักศึกษาหลังจากนั้นจะทำการคัดเลือกนักศึกษาที่เห็นสมควรได้รับทุนการศึกษาประจำปี ซึ่งข้อมูลที่ใช้ในการพิจารณาตัดสินใจในการให้ทุนจะนำไปสนับสนุนการวางแผนทาง การวิเคราะห์ข้อมูลที่ใช้เป็นการพิจารณาเพื่อให้สอดคล้องกับการจัดสรรทุนการศึกษาที่เหมาะสมมากขึ้น

ดังนั้นผู้วิจัยจึงได้นำข้อมูลการขอทุนการศึกษาของนักศึกษาระดับปริญญาตรี มหาวิทยาลัยเทคโนโลยีราชมงคลรัตนโกสินทร์พื้นที่บึงพระพิบูลย์ จักรวรรดิ ตั้งแต่ปีการศึกษา 2557-2561 มีจำนวนข้อมูล 1,155 ชุด มาศึกษาเทคนิควิธีการทำนายความสอดคล้องผลการได้รับทุนกับข้อมูลการขอทุน โดยใช้เทคนิคการทำเหมืองข้อมูลซึ่งเป็นกระบวนการค้นหาความรู้ ซึ่งเป็นรูปแบบและความสัมพันธ์ที่น่าสนใจในข้อมูลจากฐานข้อมูลขนาดใหญ่ที่ซับซ้อน [4] สามารถตอบปัญหาและค้นพบความรู้ใหม่จากฐานข้อมูลที่มีอยู่ [5] โดยมีเทคนิคการกำหนดกลุ่มข้อมูลได้แก่ การจัดกลุ่ม (Clustering) เป็นการรวมกลุ่มของข้อมูลที่มีลักษณะคล้ายกันหรือใกล้เคียงกันไว้ในกลุ่มเดียวกัน การหาความสัมพันธ์ (Association) เป็นการแสดงความสัมพันธ์ของข้อมูลทั้งสองชุดหรือมากกว่าไว้

ด้วยกัน การจัดหมวดหมู่ (Classification) เป็นการ จัดแบ่งประเภทของข้อมูลที่มีลักษณะเหมือนกันไว้ใน กลุ่มเดียวกัน เพื่อใช้ในการหาชุดข้อมูลต้นแบบ [6]ทำนายประเภทของข้อมูลที่ไม่มีการระบุประเภท หรือชนิดข้อมูล

งานวิจัยนี้เป็นการนำเสนอการเปรียบเทียบ ประสิทธิภาพของการจำแนกข้อมูลเพื่อการทำนาย ระหว่างผลการได้รับทุนและข้อมูลของนักศึกษา โดยใช้ เทคนิควิธีทั้งหมด 5 แบบได้แก่ วิธีต้นไม้ตัดสินใจ (Decision Tree) วิธีแบบเบย์ (Naïve Bayes) วิธีเค เนียร์เรสเนเบอร์ (K-Nearest Neighbors) วิธีการ เรียนรู้เชิงลึก (Deep Learning) และวิธีแรนดอมฟอรั เรส (Random Forest) เพื่อหาเทคนิควิธีที่มีความ ถูกต้องและมีความแม่นยำมากที่สุดและนำไปใช้ในการ พัฒนารูปแบบหรือวิธีการจัดการข้อมูลเพื่อใช้ ประกอบการพิจารณาให้มีความเหมาะสมและเป็น ประโยชน์ต่อผู้รับทุน

#### วัตถุประสงค์ของการวิจัย

1. เพื่อวิเคราะห์ปัจจัยที่เกี่ยวข้องในการได้รับ ทุนการศึกษาของนักศึกษาระดับปริญญาตรี
2. เพื่อเปรียบเทียบประสิทธิภาพการจำแนก ข้อมูลของตัวแบบด้วยเทคนิควิธีต้นไม้ตัดสินใจ วิธีแบบ เบย์ วิธีเคเนียร์เรสเนเบอร์ วิธีการเรียนรู้เชิงลึก และวิธี แรนดอมฟอรัเรส

## 2. วัสดุและวิธีการ

### ประชากรและกลุ่มตัวอย่าง

ประชากรที่ใช้ในการวิจัยคือนักศึกษาที่ยื่นคำ ร้องขอรับทุนการศึกษาของงานกองทุนเพื่อการศึกษา กองบริหารงานบพิตรพิมุขจักรวรรดิ มหาวิทยาลัย เทคโนโลยีราชมงคลรัตนโกสินทร์ โดยคัดเลือกกลุ่ม ตัวอย่างจากนักศึกษาที่ยื่นคำร้องขอรับทุนการศึกษา ตั้งแต่ปีการศึกษา 2557 ถึง 2561 จำนวนทั้งหมด 1,155 ชุด

### วิธีการดำเนินงานวิจัย

ขั้นตอนดำเนินการวิจัยได้ใช้กระบวนการทำ เหมือนข้อมูลของ CRISP-DM [7] ดังต่อไปนี้

1. การทำความเข้าใจการวิจัยและธุรกิจ โดย การศึกษาข้อมูลจากแบบคำร้องการขอรับทุนของ นักศึกษาที่งานกองทุนได้จัดทำขึ้นเพื่อใช้เป็นข้อมูล

ประกอบการพิจารณาทุนและข้อมูลผลคะแนนการ พิจารณาของคณะกรรมการกองทุน ประกอบด้วย ปี การศึกษา หลักสูตรการศึกษา รหัสนักศึกษา ชื่อ นามสกุล คณะ สาขาวิชา เกรดเฉลี่ยสะสม อายุ สัญชาติ เชื้อชาติ ศาสนา ที่อยู่ เบอร์โทรศัพท์ สถานะครอบครัว อาชีพบิดา รายได้บิดา อาชีพมารดา รายได้มารดา จำนวนบุตร เป็นบุตรลำดับที่ อาศัยอยู่กับ รายได้ผู้ อุปการะ เคยได้รับทุน เคยกู้ยืมศ. เคยทำงานพิเศษ มี สิทธิขอรับทุน การเข้าร่วมกิจกรรม การเข้ารับการ สัมภาษณ์ ครอบครัว รายได้น้อย สุขภาพของพ่อแม่ เข้าบ้าน ความเห็นอาจารย์ที่ปรึกษา สิทธิการได้รับทุน การเข้าร่วมกิจกรรม และระเบียบการแต่งกาย

2. การทำความเข้าใจข้อมูล เก็บรวบรวมข้อมูล โดยชุดข้อมูลได้จากแบบคำร้องการขอรับทุนและผล การพิจารณาทุนของงานกองทุนกู้ยืมเพื่อการศึกษา กอง บริหารงานบพิตรพิมุขจักรวรรดิ มหาวิทยาลัย เทคโนโลยีราชมงคลรัตนโกสินทร์

3. การจัดเตรียมข้อมูล เลือกข้อมูลที่ต้องการ วิเคราะห์ให้อยู่ในรูปแบบที่เหมาะสมสำหรับการ วิเคราะห์ หรือทำการแปลง โดยมีขั้นตอนดังต่อไปนี้

3.1 การคัดเลือกข้อมูล เป็นการนำข้อมูลมาผ่าน กระบวนการ [8] ตรวจสอบ เพิ่มและปรับปรุงรายการ ข้อมูลที่ไม่ถูกต้อง ข้อมูลที่ไม่มีความสัมพันธ์กันออก จากชุดข้อมูล จำนวนข้อมูลทั้งหมด 1,155 ชุด โดยการ คัดเลือกข้อมูลร่วมกับหัวหน้างานกองทุน ตัวอย่างเช่น รหัสนักศึกษา ชื่อ นามสกุล ปีการศึกษา สัญชาติ และ เชื้อชาติ เป็นต้น จนได้จำนวนข้อมูล 15 แอตทริบิวต์ ที่ ใช้ในการพิจารณา ประกอบด้วย เลขที่คำร้อง (ID) สิทธิพิจารณาได้รับทุน (stu\_List) เกรดเฉลี่ยสะสม (GPA) สถานะครอบครัว (f\_status) รายได้ (f\_income) สุขภาพพ่อแม่ (disability) จำนวนบุตร (Num\_Siblings) เช่าบ้าน (rental) สถานะการเงิน (less) ได้รับทุนการศึกษา (H\_fund) กู้ยืม กยศ. (KYS) ทำงานพิเศษ (Parttime) การร่วมกิจกรรม (activity) การสัมภาษณ์ (interview) ผลการพิจารณา (result)

3.2 การทำข้อมูลให้สมบูรณ์ จากการตรวจสอบ ข้อมูลเกรดเฉลี่ยไม่มีการบันทึกข้อมูลในคำร้องจำนวน 68 ชุด ทำให้ต้องขอข้อมูลจากงานทะเบียนนักศึกษา นำมาบันทึกให้ครบ

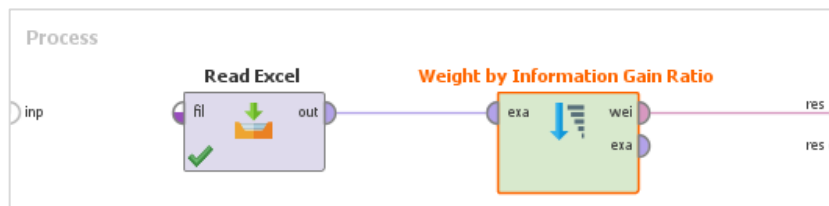
3.3 การเปลี่ยนรูปแบบข้อมูล เพื่อให้อยู่ใน รูปแบบที่เหมาะสมและพร้อมนำไปใช้งาน ตาม ตารางที่ 1

ตารางที่ 1 การแปลงข้อมูลเพื่อใช้วิเคราะห์ข้อมูล

ลำดับ	แอตทริบิวต์	รายละเอียด
1	สิทธิพิจารณาทุน (stu_List)	มีสิทธิ = yes, ไม่มีสิทธิ = no
2	การสัมภาษณ์ (interview)	เข้าสัมภาษณ์ = yes, ไม่เข้าสัมภาษณ์ = no
3	สถานะครอบครัว (f_status)	อยู่ด้วยกัน = O, หย่าร้าง = I, บิดาเสียชีวิตหรือมารดาเสียชีวิต = D1, เสียชีวิตทั้งคู่ = X, แยกกันอยู่ = T
4	สุขภาพพ่อแม่(disability)	บิดามารดาสุขภาพดี = yes, สุขภาพไม่ดี = no
5	เช่าบ้าน (rental)	เช่าบ้าน = yes, ไม่เช่า = no
6	สถานะการเงิน (less)	มั่นคง = yes, ไม่มั่นคง = no
7	ได้รับทุนการศึกษา (H_fund)	เคยได้รับทุน = yes, ไม่เคยได้รับทุน = no
8	กู้ยืมกยศ. (KYS)	กู้ยืม = yes, ไม่กู้ = no, รอผล = wait
9	ทำงานพิเศษ(Parttime)	เคยทำงานพิเศษ = yes, ไม่เคยทำงานพิเศษ = no
10	ผลพิจารณา (result)	ได้ = yes, ไม่ได้ = no

3.4 หาค่าน้ำหนักของข้อมูลที่มีความเกี่ยวข้องหรือสัมพันธ์กันของแอตทริบิวต์ โดยการวิเคราะห์ค่าน้ำหนักของแอตทริบิวต์ด้วยวิธีการ Information Gain

Ratio ได้จำนวนแอตทริบิวต์ทั้งหมด 11 แอตทริบิวต์ตามภาพที่ 1



ภาพที่ 1 วิเคราะห์ค่าน้ำหนักด้วยวิธีการ Information gain ratio

4. การสร้างตัวต้นแบบ เลือกใช้เทคนิคการสร้างต้นแบบโดยใช้แบบจำแนกข้อมูล จัดทำตัวแบบให้เป็นมาตรฐานเพื่อให้ได้ผลลัพธ์ที่ดีที่สุด เลือกใช้เทคนิคที่แตกต่างกันสำหรับปัญหาการทำเหมืองข้อมูลที่เหมือนกันได้ โดยใช้เทคนิควิธีโครงสร้างต้นไม้การตัดสินใจ(Decision Tree) วิธีแบบเบย์(Naive Bayes) วิธีเคเนียร์เรสเนเบอร์(K-Nearest Neighbors) วิธีการเรียนรู้เชิงลึก(Dep Learning) และวิธีแรน ดอมฟอรัส(Random Forest) ด้วยวิธีการ 10 - Fold Cross Validation โดยแบ่งข้อมูลออกเป็น 10 ชุดเท่ากัน ใช้ 1 ชุดเป็นกลุ่มทดสอบ ข้อมูลอีก 9 ชุดเป็นกลุ่มที่ใช้ในการเรียนรู้ และทำการวนซ้ำจำนวน 10 รอบ โดยเปลี่ยนกลุ่มทดสอบไปเรื่อย ๆ จนครบ ดังภาพที่ 2

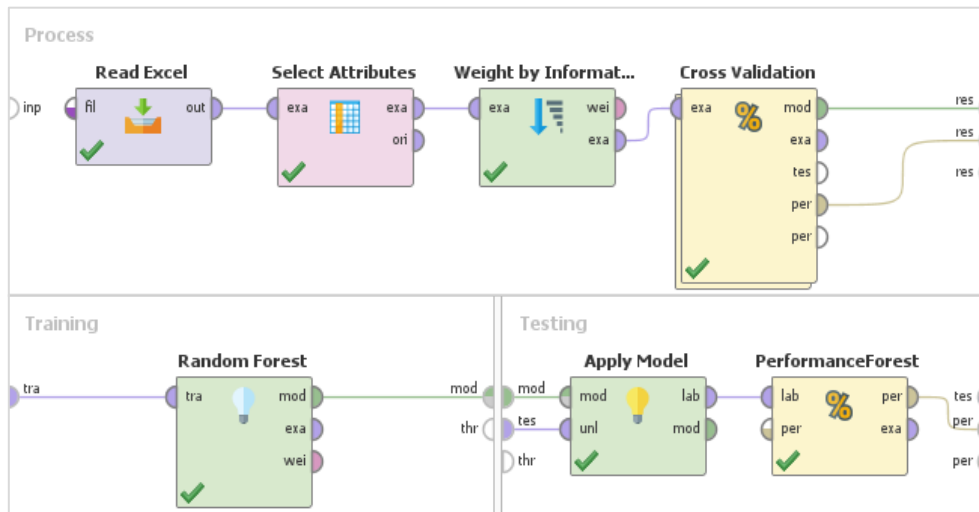
5. การประเมินตัวแบบ หาตัวแบบที่ได้บรรลุวัตถุประสงค์ สร้างกฎเกณฑ์ที่สำคัญของปัญหาการวิจัยพิจารณาเพื่อตัดสินใจจากประโยชน์ที่ได้ โดยวิเคราะห์

ด้วยวิธี 10-Fold Cross Validation เพื่อให้ได้ค่าความถูกต้อง(Accuracy) ค่าความแม่นยำ(Precision) และค่าความระลึก(Recall) ที่ได้ค่าจากประมวลผลตามตัวแบบ(Model)

6. ขั้นตอนการใช้งาน สร้างตัวแบบที่ยังไม่สามารถทำให้สมบูรณ์แบบได้ หรือการสร้างรายงานที่เหมาะสมไปใช้งานจริงเพื่อวิเคราะห์และแก้ปัญหาที่เกิดขึ้นตามความต้องการ

**เครื่องมือที่ใช้ในการเก็บข้อมูล**

1. ในการเก็บข้อมูลจากแบบฟอร์มข้อมูลการของทุนของนักศึกษาจะทำการบันทึกข้อมูลผ่านโปรแกรมเว็บที่ผู้วิจัยได้พัฒนาขึ้นมาเพื่อเก็บข้อมูลโดยบันทึกข้อมูลเก็บไว้ในฐานข้อมูลมายเอสคิวแอล (MySQL)



ภาพที่ 2 สร้างต้นแบบด้วยวิธีเรandomฟอรัส

2. ใช้โปรแกรม RapidMiner Studio 9.3 ในการวิเคราะห์ข้อมูลการทำนายการได้รับทุนการศึกษาของนักศึกษา

**การเก็บรวบรวมข้อมูล**

ในการเก็บรวบรวมข้อมูลมีขั้นตอนดังนี้

1. ผู้วิจัยได้ทำบันทึกข้อความขอความอนุเคราะห์ขอแบบฟอร์มข้อมูลการขอทุนของนักศึกษาและข้อมูลผู้ได้รับทุนการศึกษาในปีการศึกษา 2557 ถึง 2561 จำนวน 1,155 ชุด
2. เมื่อได้รับแบบฟอร์มแล้วจะทำการบันทึกข้อมูลที่อยู่ในแบบฟอร์มทั้งหมดลงในฐานข้อมูลผ่านโปรแกรมเว็บที่พัฒนาขึ้น
3. นำข้อมูลการขอทุนจากฐานข้อมูลออกมาโดยการแปลงข้อมูลให้อยู่ในรูปแบบไฟล์ .xlsx

**การวิเคราะห์ข้อมูล**

1. การวิเคราะห์ข้อมูลจะนำข้อมูลทั้งหมดทุกแอตทริบิวต์(Attribute) นำมาหาค่าน้ำหนัก(Weight) ซึ่งแสดงปัจจัยที่เกี่ยวข้องในการหาตัวแบบการได้รับทุนการศึกษา
2. นำข้อมูลมาทดสอบผลการทำนายตามตัวแบบ(Model) และเปรียบเทียบประสิทธิภาพด้วยเทคนิควิธีต้นไม้ตัดสินใจ(Decision Tree) วิธีแบบเบย์(Naïve Bayes) วิธีเคเนียร์เรสเนเบอร์(K-Nearest

Neighbors) วิธีการเรียนรู้เชิงลึก(Dep Learning) และวิธีเรandomฟอรัส(Random Forest)

**สถิติที่ใช้ในการวิจัย**

ในการพิจารณาผลการทำนายการได้รับทุนการศึกษาคาดูผลจากค่าร้อยละสูงที่สุดประกอบด้วย [9,10]

- ค่าความถูกต้อง(Accuracy) คือ จำนวนที่ทำนายถูกของทุกคลาส ใช้  $Accuracy = \frac{TP+TN}{N}$
- ค่าความแม่นยำ(Precision) คือ ผลการคำนวณที่ทำนายถูกร้อยละเท่าใด ใช้  $Precision = \frac{TP}{TP+FP}$
- ค่าความระลึก(Recall) คือจำนวนที่ทำนายถูกที่ตัว ใช้  $Recall = \frac{TP}{TP+FN}$

ค่า TP (True Positive) คือจำนวนผู้ได้รับทุนและได้รับทุนจริง ค่า FP (False Positive) คือ จำนวนผู้ได้รับทุนแต่ไม่ได้รับทุน ค่า FN (False Negative) คือ จำนวนผู้ไม่ได้รับทุนแต่ได้รับทุน ค่า TN (True Negative) คือ จำนวนผู้ไม่ได้รับทุนและไม่ได้รับทุนจริง

**3. ผลการวิจัยและอภิปรายผล**

**ผลการวิจัย**

1. การวิเคราะห์ปัจจัยที่เกี่ยวข้องในการได้รับทุนการศึกษาให้เปล่าของนักศึกษาระดับปริญญาตรี ได้ทำการคำนวณค่าน้ำหนักของชุดข้อมูลพบว่า มีจำนวน 11 แอตทริบิวต์ ที่เกี่ยวข้องกับการได้รับทุนการศึกษา

ตารางที่ 2 แสดงปัจจัยที่เกี่ยวข้องกับการได้รับทุนการศึกษาให้เปล่าของนักศึกษาระดับปริญญาตรี

ลำดับ	แอตทริบิวต์	ค่าน้ำหนัก
1	การเข้าร่วมกิจกรรม (Activity)	0.643
2	การสัมภาษณ์ (Interview)	0.616
3	สิทธิพิจารณาเงินทุน (stu_List)	0.351
4	เกรดเฉลี่ยสะสม (GPA)	0.226
5	รายได้ครอบครัว (f_Income)	0.171
6	รายได้น้อย (less)	0.090
7	สุขภาพของพ่อแม่ (disability)	0.071
8	จำนวนบุตร (Num_Sibling)	0.042
9	เช่าบ้าน (rental)	0.019
10	เคยได้รับทุน(H_fund)	0.008
11	ทำงานพิเศษ (Parttime)	0.008

จากตารางที่ 2 ปัจจัยที่เกี่ยวข้องกับการได้รับทุนการศึกษาพบว่า การเข้าร่วมกิจกรรมมีค่าน้ำหนัก 0.643 มีค่ามากที่สุด รองลงมาเป็นการเข้ารับการสัมภาษณ์มีค่าน้ำหนัก 0.616 รายชื่อผู้มีสิทธิพิจารณาเงินทุนมีค่าน้ำหนัก 0.351 เกรดเฉลี่ยสะสมมีค่าน้ำหนัก 0.226 รายได้ครอบครัวมีค่าน้ำหนัก 0.171 รายได้น้อยมีค่าน้ำหนัก 0.090 สุขภาพของพ่อแม่มีค่าน้ำหนัก 0.071 จำนวนบุตรมีค่าน้ำหนัก 0.042 เช่าบ้านมีค่าน้ำหนัก 0.019 ค่าน้ำหนักที่น้อยที่สุดคือเคยได้รับทุนมีค่าน้ำหนัก 0.008 และเคยทำงานพิเศษมีค่าน้ำหนัก 0.008 ตามลำดับ ค่าน้ำหนักมาจากการวิเคราะห์ค่าน้ำหนักของแอตทริบิวต์ด้วยวิธีการ Information Gain Ratio โดยใช้โปรแกรม Rapid Miner 9.3 ในส่วนแอตทริบิวต์สิทธิพิจารณาเงินทุน (stu\_List) หมายถึงผู้

ขอทุนมีสิทธิที่จะเข้ารับการสัมภาษณ์กับคณะกรรมการกองทุนและการสัมภาษณ์หมายถึงเข้ารับการสัมภาษณ์ตามวันที่และเวลาที่งานกองทุนกำหนดจึงทำให้ค่าน้ำหนักอยู่ในอันดับที่ 2 และ 3

2. เปรียบเทียบประสิทธิภาพการจำแนกข้อมูลของตัวแบบด้วย 5 เทคนิควิธี

เมื่อทำการเลือกปัจจัยที่เกี่ยวข้องกับชุดข้อมูลแล้ว ได้นำชุดข้อมูลทั้งหมด 1,155 ชุด มาทดสอบประสิทธิภาพของการจำแนกข้อมูลเพื่อการทำนายด้วยเทคนิควิธีต้นไม้ตัดสินใจ (Decision Tree) วิธีแบบเบย์ (Naïve Bayes) วิธีเคเนียร์เรสเนเบอร์ (K-Nearest Neighbors) วิธีการเรียนรู้เชิงลึก (Deep Learning) และวิธีแรนดอมฟอว์เรส (Random Forest) ได้ค่าประสิทธิภาพดังตารางที่ 3

ตารางที่ 3 แสดงการเปรียบเทียบประสิทธิภาพการจำแนกข้อมูลการทำนายด้วย 5 เทคนิควิธีตามลำดับ

เทคนิควิธี (Algorithm)	ค่าความถูกต้อง (Accuracy)	ค่าความแม่นยำ (Precision)	ค่าความระลึก (Recall)
Random Forest	94.28%	95.89%	83.64%
Deep Learning	93.51%	95.62%	81.21%
Decision Tree	92.64%	92.02%	81.21%
Naïve Bayes	92.47%	94.41%	78.48%
K-NN	89.70%	90.25%	71.82%

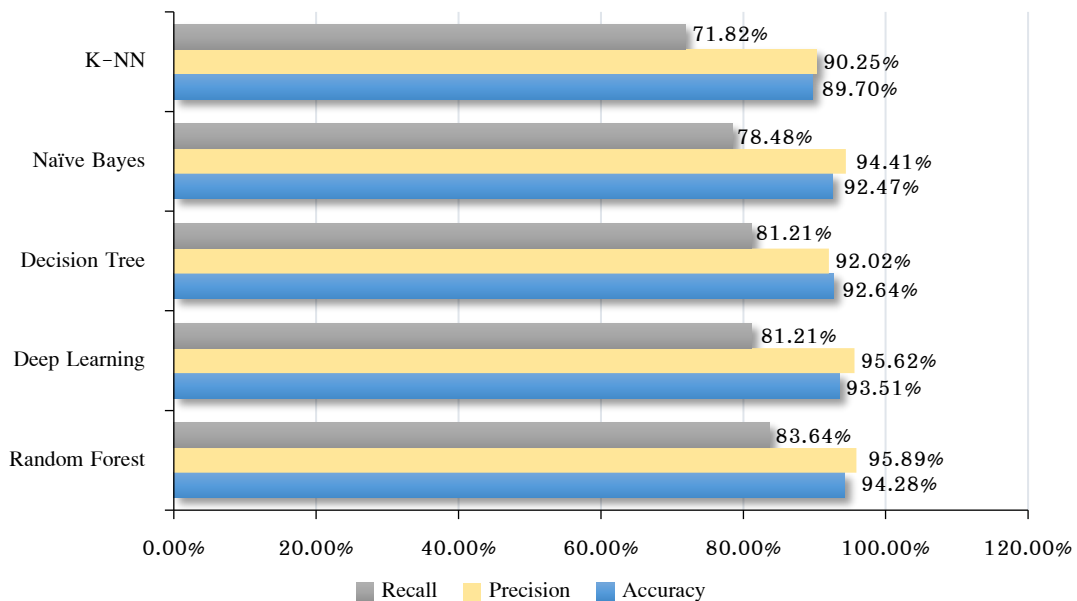
จากตารางที่ 3 พบว่าตัวแบบที่สร้างจากเทคนิควิธีแรนดอมฟอว์เรส (Random Forest) มีค่าความถูกต้องมากที่สุดคือร้อยละ 94.28 ค่าความแม่นยำร้อยละ

95.89 และค่าความระลึกร้อยละ 83.64 รองลงมาคือวิธีการเรียนรู้เชิงลึก (Deep Learning) มีค่าความถูกต้องร้อยละ 93.51 ค่าความแม่นยำร้อยละ 95.62

และค่าความระลึกร้อยละ 81.21 วิธีต้นไม้ตัดสินใจ (Decision Tree) มีค่าความถูกต้องร้อยละ 92.64 ค่าความแม่นยำร้อยละ 92.02 และค่าความระลึกร้อยละ 81.21 วิธีแบบเบย์ (Naïve Bayes) มีค่าความถูกต้องร้อยละ 92.47 ค่าความแม่นยำร้อยละ 94.41 และค่าความระลึกร้อยละ 78.48 และวิธีที่ถูกต้องน้อยที่สุดคือวิธีเคเนียร์เรสเนเบอร์ (K-Nearest Neighbors) มีค่าความถูกต้องร้อยละ 89.70 ค่าความแม่นยำร้อยละ

90.25 และค่าความระลึกร้อยละ 71.82 ตามลำดับ วิธีคำนวณหาค่าความถูกต้องยกตัวอย่างเช่น ค่าความถูกต้องของวิธีแรนดอมฟอร์เรสที่ได้ค่าร้อยละ 94.28 มาจาก ค่าจำนวนทั้งหมด (N)=1,155 ค่า TP=816 และค่า TN=273 จะคำนวณค่าความถูกต้องได้ร้อยละ 94.28 ในการจำแนกข้อมูลด้วยเทคนิควิธีทั้ง 5 แบบได้แสดงเป็นตารางเปรียบเทียบดังภาพที่ 3

เปรียบเทียบประสิทธิภาพการจำแนกข้อมูลด้วยเทคนิควิธี



ภาพที่ 3 แผนภูมิแสดงการเปรียบเทียบประสิทธิภาพการจำแนกข้อมูล 5 แบบ

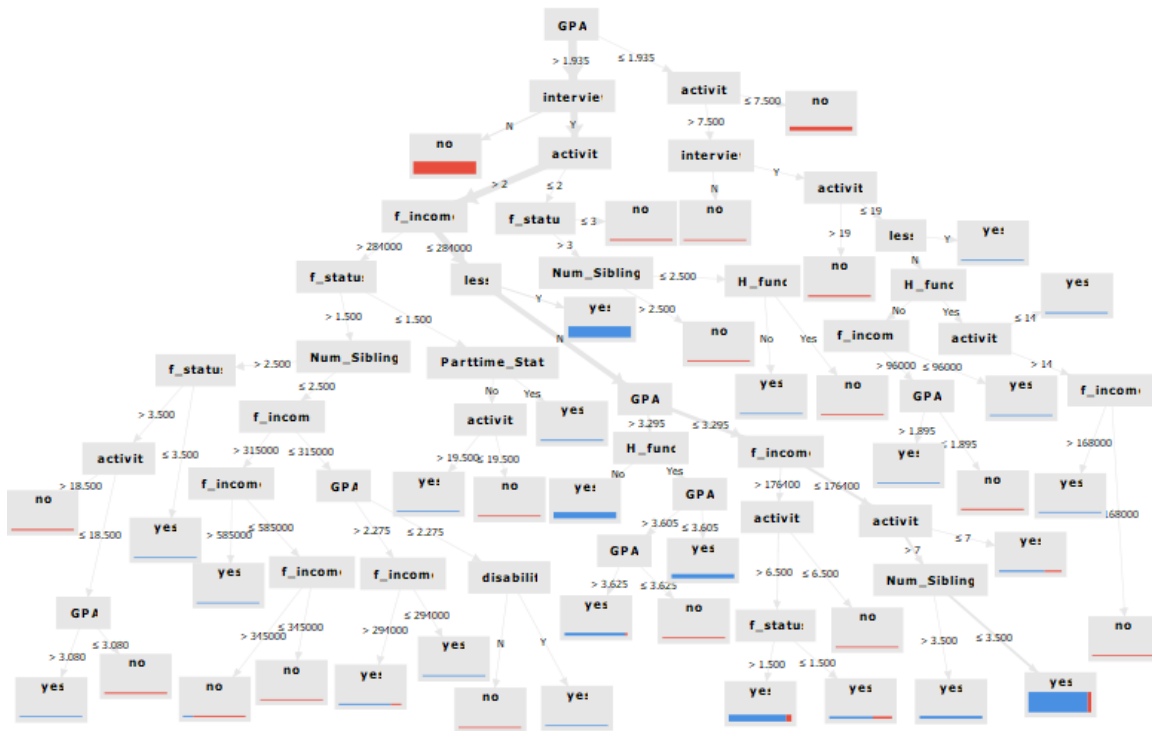
ตัวแบบที่สร้างจากเทคนิควิธีแรนดอมฟอร์เรส (Random Forest) ที่มีค่าความถูกต้อง ค่าความแม่นยำ และค่าความระลึกร้อยละมากที่สุดโดยโปรแกรม Rapid Miner 9.3 โดยใช้วิธีการแบ่งข้อมูลเพื่อทดสอบประสิทธิภาพตัวแบบด้วยวิธีการแบ่งค่าข้อมูลแบบ Cross Validation กำหนดข้อมูลค่า K=10 ตามภาพที่ 4

#### อภิปรายผล

1. ปัจจัยที่เกี่ยวข้องในการได้รับทุนการศึกษาให้เปล่าของนักศึกษาระดับปริญญาตรีที่เลือกมาทั้งหมด 11 แอตทริบิวต์ เป็นข้อมูลที่มีค่าน้ำหนักที่สามารถนำไปใช้ในการทดสอบประสิทธิภาพการจำแนกข้อมูลด้วยเทคนิควิธีต่าง ๆ เช่น เกรดเฉลี่ยสะสม รายได้ของครอบครัว สุขภาพของพ่อแม่ จำนวนบุตร ซึ่งเป็นข้อมูลที่คณะกรรมการสัมภาษณ์ทุนได้ใช้เป็นข้อมูลหลักใน

การให้คะแนนกับนักศึกษาที่ขอทุนสอดคล้องกับ [11] ได้รวบรวมเกณฑ์การพิจารณาทุนการศึกษาที่ต้องประกอบด้วยข้อมูลเกณฑ์ผลการเรียนที่อยู่ในระดับต่าง ๆ ตั้งแต่ 2.25-4.00 เกณฑ์ฐานะเศรษฐกิจในส่วนรายได้ของบิดามารดา จำนวนพี่น้องและลักษณะที่พักอาศัย

2. การเปรียบเทียบประสิทธิภาพการจำแนกข้อมูลเพื่อการทำนายการได้รับทุนการศึกษาของนักศึกษาระดับปริญญาตรีพบว่าเทคนิควิธีแรนดอมฟอร์เรส (Random Forest) มีค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) และค่าความระลึก (Recall) เป็นค่าร้อยละที่มีค่ามากที่สุด ซึ่งสอดคล้องกับ [12] พบว่าเทคนิคการจำแนกข้อมูลด้วยวิธีต้นไม้ตัดสินใจด้วยอัลกอริทึม Random Forest ให้ประสิทธิภาพในการจำแนกข้อมูลได้ดีกว่าทุกเทคนิคที่เปรียบเทียบ



ภาพที่ 4 ตัวแบบที่สร้างจากเทคนิคแรดตอมฟอ์เรสด้วยโปรแกรม Rapid Miner 9.3

**สรุปผลการวิจัย**

งานวิจัยนี้ได้ทำการวิเคราะห์ปัจจัยที่เกี่ยวข้องในการได้รับทุนการศึกษาของนักศึกษาระดับปริญญาตรี และเปรียบเทียบประสิทธิภาพการจำแนกข้อมูลการได้รับทุนการศึกษาของนักศึกษาระดับปริญญาตรี มหาวิทยาลัยเทคโนโลยีราชมงคลรัตนโกสินทร์พื้นที่บพิตรพิมุขจักรวรรดิ โดยมีข้อมูลที่ใช้ทั้งหมด 11 แอตทริบิวต์ ได้แก่ การเข้าร่วมกิจกรรม การสัมภาษณ์ สิทธิพิจารณารับทุน เกรดเฉลี่ยสะสม รายได้น้อย สุขภาพของพ่อแม่ จำนวนบุตร เข้าบ้าน เคยได้รับทุนทำงานพิเศษ และรายได้ครอบครัว โดยได้ทำการเปรียบเทียบประสิทธิภาพการจำแนกประเภทข้อมูลเพื่อการทำนายจำนวน 5 วิธี พบว่า เทคนิควิธีที่มีค่าความถูกต้องที่มากที่สุดคือวิธีแรดตอมฟอ์เรส มีค่าร้อยละ 94.28 งานวิจัยนี้สามารถนำไปใช้ทำนายเพื่อวิเคราะห์แนวทางในการเสนอแนะข้อมูลที่เป็นเพิ่มเติมของนักศึกษาเช่น วิเคราะห์ข้อมูลกลุ่มอาชีพของผู้ปกครอง ข้อมูลทางเจตคติ ทักษะคติ สภาพแวดล้อมความเป็นอยู่ที่เกี่ยวข้องกับนักศึกษา เป็นต้น และพัฒนาระบบสารสนเทศเพื่อจัดการ สนับสนุนข้อมูลให้กับนักศึกษา ผู้สนับสนุนทุนการศึกษาและมหาวิทยาลัยในการวางแผนช่วยเหลือนักศึกษาที่ประสบปัญหาการเรียนในด้านทุนทรัพย์ต่อไป

**4. เอกสารอ้างอิง**

1. กองทุนเพื่อความเสมอภาคทางการศึกษา. รายงานประจำปี 2562. กรุงเทพฯ: กสศ.; 2563.
2. พลเดช ปิ่นประทีป. กองทุนเพื่อความเสมอภาคทางการศึกษา (กสศ.) เป็นเครื่องมือใหม่ที่มีความสำคัญต่อการลดความเหลื่อมล้ำและขับเคลื่อนการปฏิรูปประเทศด้านการศึกษาในระยะยาว [อินเทอร์เน็ต]. [เข้าถึงเมื่อวันที่ 4 กุมภาพันธ์ 2564]. เข้าถึงได้จาก <https://www.eef.or.th/103/>
3. ภูมิศรันย์ ทองเหลี่ยมผด. สสำรวจผลกระทบ COVID-19 จุดเปลี่ยนครั้งสำคัญของการศึกษาโลก [อินเทอร์เน็ต]. [เข้าถึงเมื่อวันที่ 4 กุมภาพันธ์ 2564]. เข้าถึงได้จาก <https://www.eef.or.th/article1-02-01-211/>
4. วิชา เจริญภรณ์ทาร์กซ์. คลังข้อมูล เหมือนข้อมูล และธุรกิจอัจฉริยะ หน่วยที่ 8. นนทบุรี: สำนักพิมพ์มหาวิทยาลัยสุโขทัยธรรมมาธิราช; 2558.
5. วิชา ขาวฟ้า และ อรศิริ ศิลาลัย. แบบจำลองการเลือกสายการเรียนที่เหมาะสมของนักศึกษาด้วยเทคนิคการจำแนก. การประชุมวิชาการระดับชาติ



- และนานาชาติ มหาวิทยาลัยศรีปทุม ครั้งที่ 14 ประจำปี 2562; หน้า 1485–93.
6. อุดลย์ ยิ้มงาม. การทำเหมืองข้อมูล (Data Mining) [อินเทอร์เน็ต]. [เข้าถึงเมื่อวันที่ 5 กุมภาพันธ์ 2564]. เข้าถึงได้จาก <http://compcente.bu.ac.th/news-information/data-mining/>
  7. สายชล สันสมบุรณ์ทอง. การทำเหมืองข้อมูล เล่ม 1 การค้นหาความรู้จากข้อมูล. กรุงเทพฯ: จามจุรีโปรดักส์; 2560.
  8. Nutdanai Wangpratham. DATA PREPARATION การเตรียมข้อมูล [อินเทอร์เน็ต]. [เข้าถึงเมื่อวันที่ 5 กุมภาพันธ์ 2564]. เข้าถึงได้จาก <https://nutdnuy.medium.com/datapreparation-การเตรียมข้อมูล-81fba4e0b0c9>
  9. นคราญ คำวิชัย. Practical Data Mining with RapidMiner Studio 7. เชียงใหม่: ITSCI; 2559.
  10. Kasidis Satangmongkol. อธิบาย 10 Metrics พื้นฐานสำหรับวัดผลโมเดล Machine Learning [อินเทอร์เน็ต]. [เข้าถึงเมื่อวันที่ 12 เมษายน 2564]. เข้าถึงได้จาก <https://datarockie.com/2019/03/30/top-ten-machine-learning-metrics/>
  11. วิริยาภรณ์ พิชัยโชคและจันทร์จิรา พยัคฆ์เพศ. การประยุกต์ใช้กระบวนการลำดับชั้นเชิงวิเคราะห์เพื่อพิจารณาทุนการศึกษาของโรงเรียนหัวดวงราชพรหมาภรณ์ จังหวัดนครสวรรค์. NU Science Journal 2013;9(2):29–46.
  12. ปิยะนันท์ คงไฟ, จิรพันธุ์ ศรีสมพันธุ์ และจรัญแสนราช. การวิเคราะห์รูปแบบความสนใจเลือกอาชีพของผู้เรียนการแบบทดสอบ ตามทฤษฎีของจอห์น ฮอลแลนด์ ด้วยเทคนิคเหมืองข้อมูล. Science and Technology RMUTT Journal 2019;9:80–90.