The Sci J of Phetchaburi Rajabhat University.
2019. 16(2): 1-10

การศึกษาเทคนิคพยากรณ์การได้รับปัจจัยพื้นฐานนักเรียนยากจนของนักเรียนโรงเรียนวัดพระขาว (ประชานุเคราะห์) ด้วยเทคนิคเหมืองข้อมูล

A Study of Techniques in Predicting of a Fund Receiving for Poor Students of Phrakhao School Students by Using Data Mining Technique

อัครพล พูลสวัสดิ์[1*] และ จรัญ แสนราช[2]

Akarapon Poonsawad[1*] and Charun Sanrach[2]

[1]*โรงเรียนวัดพระขาว (ประชานุเคราะห์) อำเภอบางบาล พระนครศรีอยุธยา*

[2]*คณะครุศาสตร์อุตสาหกรรม มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ กรุงเทพ*

[1] *Watphrakhao Prachanukroh School, Bang Ban, Phra Nakhon Si Ayutthaya*

[2] *Faculty of Technical Education King Mongkut's University of Technology North Bangkok, Bangkok*

***Corresponding author; E-mail:** akara382@gmail.com

Received: 21 July 2019/ Revised: 17 October 2019 / Accepted: 21 October 2019

บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์เพื่อ 1) เปรียบเทียบประสิทธิภาพของเทคนิคพยากรณ์การได้รับปัจจัยพื้นฐานนักเรียนยากจน ด้วยเทคนิคเหมืองข้อมูล และ 2) เพื่อค้นหาตัวแบบของการได้รับปัจจัยพื้นฐานนักเรียนยากจนด้วยเทคนิคเหมืองข้อมูลที่เหมาะสม โดยมีชุดข้อมูลที่ใช้ในการวิจัยเป็นข้อมูลการการคัดกรองนักเรียนยากจนของนักเรียนระดับชั้นประถมศึกษาปีที่ 1 ถึงมัธยมศึกษาปีที่ 3 โรงเรียนวัดพระขาว (ประชานุเคราะห์) สังกัดสำนักงานเขตพื้นที่การศึกษาประถมศึกษาพระนครศรีอยุธยา เขต 2 ในปีการศึกษา 2560 และ 2561 จำนวน 470 ชุดข้อมูล มีขั้นตอนการวิจัย 5 ขั้นตอน ตามแนวคิด CRIPS-DM โดยใช้เทคนิคการจำแนกข้อมูล 3 เทคนิค ได้แก่ ต้นไม้ตัดสินใจ (Decision tree) แรนดอมฟอเรสต์ (Random forest) และเทคนิคแบ็กกิ้ง (Bagging) ผลการวิจัยพบว่า เทคนิคการจำแนกข้อมูลด้วยเทคนิค Random forest ให้ค่าประสิทธิภาพมากที่สุด (ความถูกต้องเท่ากับ 94.12% ความแม่นยำเท่ากับ 94.37% ความระลึกเท่ากับ 94.12% และค่า F-measure = 94.24%) และได้ตัวแบบของการทำนายการได้รับปัจจัยพื้นฐานนักเรียนยากจน ซึ่งได้จากเทคนิคแรนดอมฟอเรสต์ที่มีค่าประสิทธิภาพมากที่สุดเพื่อเป็นแนวทางในพยากรณ์การได้รับปัจจัยพื้นฐานนักเรียนยากจนต่อไป

คำสำคัญ: เทคนิคพยากรณ์ ปัจจัยพื้นฐานนักเรียนยากจน เทคนิคเหมืองข้อมูล

Abstract

The purposes of this research study were 1) to compare the efficiency of data analysis for a fund receiving of poor students by using the data mining technique, and 2) to find a model of a fund receiving of poor students by using the appropriate data mining technique. All of 470 data sets of students in grade 1 to grade 9 in the academic year 2017 and 2018 at Phrakhao School were analyzed. The research procedures were divided into five phrases according to CRIPS-DM (Cross-Industry Standard Process for Data Mining) and using data mining three techniques including Decision tree, Random forest and Bagging. The results of this research were as follows: 1) the data classification techniques using the Random forest technique provided the most effective (Accuracy = 94.12%, Precision = 94.37%, Recall = 94.12% and F-measure value = 94.24%) and 2) the model of prediction was obtained from the Random forest technique which showed the most effective. The model can be used as a guide to predict receiving further fund for poor students.

Keywords: Predicting technique, Receive fund for poor students, Data mining technique

บทนำ

การพัฒนาประเทศให้มีความเจริญต่อเนื่องและยั่งยืน รากฐานสำคัญหนึ่งคือการศึกษา การศึกษาเป็นกลไกในการพัฒนาคนให้เป็นกำลังสำคัญในการพัฒนาประเทศชาติ ซึ่งพระราชบัญญัติการศึกษาแห่งชาติ พ.ศ. 2542 และที่แก้ไขเพิ่มเติม (ฉบับที่ 3) พุทธศักราช 2553 ได้กล่าวถึงความหมายของการศึกษาว่าเป็นกระบวนการเรียนรู้เพื่อความเจริญงอกงามของบุคคลและสังคม โดยการถ่ายทอดความรู้ การฝึกอบรม การสืบสานทางวัฒนธรรม และการสร้างสรรค์จรรโลงความก้าวหน้าทางวิชาการ การสร้างองค์ความรู้อันเกิดจากการจัดสภาพแวดล้อม สังคม การเรียนรู้ และปัจจัยเกื้อหนุนให้บุคคลเรียนรู้อย่างต่อเนื่องตลอดชีวิต [1] ซึ่งกระบวนการเรียนรู้เป็นกระบวนการหลักของการศึกษาที่จะสร้างองค์ความรู้ให้ผู้เรียน แต่ที่ขาดไม่ได้อีกส่วนหนึ่งก็คือการจัดสภาพแวดล้อม สังคม การเรียนรู้และปัจจัยเกื้อหนุน ซึ่งเป็นสิ่งที่ส่งเสริมและสนับสนุนการเรียนรู้ของผู้เรียนอีกทางหนึ่ง โดยเฉพาะอย่างยิ่งเงินปัจจัยสนับสนุนการศึกษา ซึ่งเป็นเหตุหนึ่งของความเหลื่อมล้ำทางการศึกษา ดังนั้นกองทุนเพื่อความเสมอภาคทางการศึกษาจึงได้ก่อนตั้งขึ้นและทำงานร่วมกับสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานเพื่อให้ความช่วยเหลือ ส่งเสริม พัฒนา และสนับสนุนเงินและค่าใช้จ่ายให้แก่เด็กและเยาวชนซึ่งขาดแคลนทุนทรัพย์และด้อยโอกาส และผู้ด้อยโอกาส จนสำเร็จ

การศึกษาขั้นพื้นฐาน โดยคัดกรองผ่านระบบการคัดกรองปัจจัยพื้นฐานนักเรียนยากจน [2]

โรงเรียนวัดพระขาว (ประชานุเคราะห์) เป็นโรงเรียนสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานที่ได้รับการจัดสรรเงินปัจจัยพื้นฐานนักเรียนยากจนโดยได้รับพิจารณาจากการคัดกรองข้อมูลนักเรียนในด้านต่าง ๆ เช่น รายได้ครอบครัว ข้อมูลสถานะของครัวเรือน ยานพาหนะ ที่ดิน เป็นต้น ทั้งนี้ข้อมูลดังกล่าวเป็นที่น่าสนใจในการวิเคราะห์เพื่อพยากรณ์ แนวโน้มการได้รับปัจจัยพื้นฐานนักเรียนยากจนซึ่งวิธีการวิเคราะห์พยากรณ์ที่เป็นที่นิยมคือ การทำเหมืองข้อมูล (Data mining) เป็นการค้นหาหรือสกัดความรู้จากฐานข้อมูลขนาดใหญ่ เพื่อค้นหารูปแบบ แนวทาง และความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้นโดยอาศัยหลักการทางคณิตศาสตร์ สถิติ การรู้จำ การเรียนรู้ทางเครื่องจักร เพื่อนำความรู้ที่ได้นั้นมาใช้ในการแก้ปัญหา วางแผน หรือดำเนินกลยุทธ์ขององค์กรให้ประสบความสำเร็จสูงสุด และสามารถดึงข้อมูลมาใช้ โดยการคัดเลือกข้อมูลออกมาใช้งานในส่วนที่ต้องการ [3, 4]

ผู้วิจัยจึงมีความสนใจที่จะศึกษาและนำเสนอเทคนิคเหมืองข้อมูลมาเรียนรู้เพื่อเลือกอัลกอริทึมในการจำแนกข้อมูลและหาปัจจัยที่มีผลต่อการได้รับปัจจัยพื้นฐานนักเรียนยากจนโรงเรียนวัดพระขาว (ประชานุเคราะห์) ด้วยวิธีการจำแนกข้อมูล 3 เทคนิค ได้แก่ ต้นไม้ตัดสินใจ (Decision tree) เทคนิคแรนดอมฟอเรสต์ (Random forest) และเทคนิคแบ็กกิ้ง (Bagging) เพื่อเป็นข้อมูลพื้นฐานสำหรับส่งเสริมและสนับสนุนทางด้านการศึกษาของผู้เรียนต่อไป

วัตถุประสงค์การวิจัย

1. เพื่อเปรียบเทียบประสิทธิภาพของเทคนิคพยากรณ์การได้รับปัจจัยพื้นฐานนักเรียนยากจนด้วยเทคนิคเหมืองข้อมูล

2. เพื่อค้นหาตัวแบบของการได้รับปัจจัยพื้นฐานนักเรียนยากจน ด้วยเทคนิคเหมืองข้อมูลที่เหมาะสม

วิธีดำเนินการวิจัย

1. เครื่องมือที่ใช้ในการวิจัย ผู้วิจัยได้ใช้เครื่องมือในการบันทึกและเก็บข้อมูล คือ โปรแกรม Microsoft Excel และใช้เครื่องมือเพื่อวิเคราะห์ข้อมูล คือ Rapid Miner Education

2. ชุดข้อมูลที่ใช้ในการวิจัย ผู้วิจัยได้ใช้ชุดข้อมูลจากการเก็บข้อมูลตามแบบฟอร์มนักเรียนยากจน (นร.01)[5] ของนักเรียนระดับชั้นประถมศึกษาปีที่ 1 ถึงมัธยมศึกษาปีที่ 3 โรงเรียนวัดพระขาว (ประชานุเคราะห์) สังกัดสำนักงานเขตพื้นที่การศึกษาประถมศึกษาพระนครศรีอยุธยา เขต 2 ในปีการศึกษา 2560 และ 2561 จำนวน 470 ชุดข้อมูล

3. ขั้นตอนการวิจัย

ผู้วิจัยได้ดำเนินการตามขั้นตอนการทำเหมืองข้อมูลตามแนวคิดคริสป์-ดีเอ็ม (CRIPS-DM : Cross-Industry Standard Process for Data Mining) [6] ดังนี้

3.1 การทำความเข้าใจสภาพปัญหา เป็นการวิเคราะห์และศึกษาเกี่ยวกับข้อมูลการคัดกรองนักเรียนเพื่อรับปัจจัยพื้นฐานนักเรียนยากจน ของโรงเรียนวัดพระขาว (ประชานุเคราะห์) จากการศึกษาได้แอททริบิวต์ที่สำคัญจากแบบฟอร์มขอทุนนักเรียนยากจน (นร.01) ที่ทางกองทุนเพื่อความ

เสมอภาคทางการศึกษา ร่วมกับสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน จัดทำขึ้นเป็นมาตรฐานเพื่อจัดเก็บเป็นข้อมูลพื้นฐานของนักเรียนทั่วประเทศในการพิจารณาให้ทุนนักเรียนยากจน ผ่านระบบปัจจัยพื้นฐานนักเรียนยากจน จำนวน 17 แอททริ-บิวต์ โดยที่แอททริบิวต์ที่ใช้เป็นผลในการพยากรณ์ คือ ความยากจน (Poverty) ซึ่งมีคลาส 3 คลาส ได้แก่ Normal (ปกติ) Poor (ยากจน) และ Very poor (ยากจนพิเศษ) โดยรายละเอียดของแอททริบิวต์ที่ใช้วิเคราะห์ข้อมูลการได้รับปัจจัยพื้นฐานนักเรียนยากจน ดังแสดงใน Table 1

Table 1. Seventeen Attributes used for analysis

ที่	แอททริบิวต์	คำอธิบาย
1	Nationality	สัญชาติไทยหรือไม่ (Yes, No)
2	Welfare registration programe	ลงทะเบียนคนจนหรือไม่ (Yes, No)
3	Family member	จำนวนสมาชิกในครัวเรือน
4	Income	รายได้ครัวเรือน
5	Disable person	มีคนพิการหรือไม่ (Yes, No)
6	Elderly people	มีคนอายุเกิน 60 ปีหรือไม่ (Yes, No)
7	Single parent	เป็นพ่อ/แม่เลี้ยงเดี่ยว (Yes, No)
8	Unemployed	มีคนอายุ 15-60 ปีที่ว่างงาน และไม่ใช่นักเรียนนักศึกษา (Yes, No)
9	Residence	ประเภทที่อยู่อาศัย (Owner, Rent, Other)
10	House condition	สภาพบ้านทรุดโทรมหรือไม่ (Yes, No)
11	Toilet	มีห้องน้ำหรือไม่ (Yes, No)
12	Private car	มีรถยนต์ส่วนบุคคลหรือไม่ (Yes, No)
13	Other private vehicles	มีรถปิกอัพ/บรรทุก/ตู้หรือไม่ (Yes, No)
14	Agricultural vehicles	มีรถไถ/เกี่ยวข้าวอื่น ๆ หรือไม่ (Yes, No)
15	Land	มีที่ดินทำกิน (None, Less than 1 rai, Others)
16	Living	ลักษณะบ้านที่ได้รับการเยี่ยม (with parent, with employer, with relative)
17	Poverty	ความยากจน (Normal, Poor, Very poor)

3.2 การทำความเข้าใจข้อมูลและแหล่งที่มา

แหล่งที่มาของข้อมูลที่ใช้งานงานวิจัยได้มาจาก โรงเรียนวัดพระขาว (ประชานุเคราะห์) สังกัดสำนักงานเขตพื้นที่การศึกษาประถมศึกษาพระนครศรีอยุธยา เขต 2 โดยใช้ข้อมูลในช่วงปีการศึกษา 2560-2561 นักเรียนระดับประถมศึกษาปีที่ 1 ถึงมัธยมศึกษาปีที่ 3 จำนวน 470 ชุดข้อมูล เพื่อวิเคราะห์หาปัจจัยที่ส่งผลให้นักเรียนได้รับ

ปัจจัยพื้นฐานนักเรียนยากจน โดยเป็นลักษณะของเอกสาร ผู้วิจัยจึงนำมากรอกข้อมูลในโปรแกรม Microsoft Excel ตามแอททริบิวต์ที่กำหนด โดยแอททริบิวต์ที่ต้องการทำนาย คือ แอททริบิวต์ Poverty ซึ่งคลาส Normal มีจำนวน 397 ชุดข้อมูล คลาส Poor มีจำนวน 42 ชุดข้อมูล และคลาส Very poor มีจำนวน 31 ชุดข้อมูล

3.3 การจัดเตรียมข้อมูล

การเตรียมข้อมูลเป็นขั้นตอนสำคัญที่จะต้องดำเนินการก่อนที่จะนำข้อมูลไปวิเคราะห์ โดยมีขั้นตอนดังนี้

1. การทำความสะอาดข้อมูล (Data cleansing) โดยได้ดำเนินตรวจสอบข้อมูลหาค่าสูญหาย (Missing value) หรือมีค่าข้อมูลที่กรอกผิดพลาด โดยหากข้อมูลมีปัญหาจะดำเนินการตรวจสอบจากเอกสารเพื่อให้ข้อมูลมีความถูกต้อง

2. ปรับเปลี่ยนรูปแบบข้อมูล (Data transformation) ได้ดำเนินการแปลงไฟล์เป็นรูปแบบของ CSV เพื่อให้สามารถนำไปวิเคราะห์กับโปรแกรม Rapid Miner ได้

3.4 การสร้างโมเดล (Modeling) แล้วเลือกเทคนิคที่เหมาะสม

เนื่องจากชุดข้อมูลมีลักษณะไม่สมดุล กล่าวคือ คลาสที่ต้องการที่ทำนาย มีจำนวนชุดข้อมูลแต่ละคลาสไม่เท่ากัน และต่างกันมาก จึงได้ปรับสมดุลของชุดข้อมูลด้วยวิธีการสุ่มเพิ่มตัวอย่างกลุ่มน้อย (Synthetic Minority Over-Sampling Technique: SMOTE) ซึ่งเป็นเทคนิคที่ใช้ในการแก้ปัญหาที่ต้องการจำแนกข้อมูลที่ไม่สมดุล ซึ่งข้อมูลมีจำนวนตัวอย่างแตกต่างกันมากในแต่ละคลาส เมื่อทำการจำแนกประเภท จะทำให้มีการเรียนรู้แต่ข้อมูลกลุ่มที่มาก ผลที่ได้ก็จะจำแนกไปในข้อมูลกลุ่มมาก วิธี SMOTE เป็นวิธีการเพิ่มจำนวนข้อมูลประเภทที่มีข้อมูลน้อย ให้เพิ่มปริมาณข้อมูลใกล้เคียงกับประเภทที่มีมากที่สุด โดยสุ่มค่าขึ้นมาหนึ่งค่า และหาค่าระยะห่างระหว่างค่าที่เลือกกับทุก ๆ ค่า แล้วเลือกค่าที่ใกล้เคียงที่สุด เช่น กำหนดไว้ 5 ค่า สุ่มค่าจากที่เลือก 1 ใน 5 หาค่าอยู่ระหว่างค่าที่เลือกตอนแรกและค่าที่สุ่มมาตอนหลัง เพื่อนำค่าที่ได้มาเพิ่มจำนวนข้อมูล [7]

ผู้วิจัยจึงใช้เทคนิควิธี SMOTE จนได้จำนวนข้อมูลแต่ละคลาสเท่ากัน คือคลาสละ 397 ชุดข้อมูล รวมทั้งสิ้น 1,191 ชุดข้อมูล

จากนั้นได้แบ่งข้อมูลเป็น Training data จำนวนร้อยละ 70 เพื่อใช้สร้างโมเดล และ Testing data จำนวนร้อยละ 30 เพื่อใช้ทดสอบโมเดล เนื่องจากเป็นการกำหนดอัตราส่วนพื้นฐานของการทดสอบข้อมูล [8] โดยสร้างโมเดลด้วยเทคนิคเหมืองข้อมูล แบบ Classification ด้วย 3 เทคนิค คือ Decision tree, Random forest และ Bagging เพราะเป็นเทคนิคเหมืองข้อมูลในการจำแนกข้อมูลที่มีความเหมาะสม ได้รับความนิยมและมีค่าความแม่นยำในการพยากรณ์ [9, 10] เพื่อหาเทคนิคที่มีประสิทธิภาพมากที่สุด โดยใช้โปรแกรม Rapid Miner ซึ่งแต่ละเทคนิค มีการดำเนินการ ดังนี้

เทคนิค Decision tree เป็นเทคนิควิธีที่คัดเลือกคุณลักษณะที่มีความสัมพันธ์กับคลาสมากที่สุดขึ้นมาเป็นโหนดบนสุดของ Tree (Root node) หลังจากนั้นก็จะหาคุณลักษณะถัดไปเรื่อย ๆ ในการหาความสัมพันธ์ของคุณลักษณะนี้จะใช้ตัววัดที่เรียกว่า Information Gain (IG) การคำนวณค่า Information Gain จะใช้ค่า Entropy ซึ่งเป็นการวัด

ความแตกต่างหรือการกระจัดกระจายของข้อมูล ถ้าข้อมูลมีความแตกต่างกันมากค่า Entropy ก็จะมีค่าสูง ในทางตรงข้ามถ้าข้อมูลมีความคล้ายกันมากค่า Entropy ก็จะมีค่าต่ำ [11] ซึ่งผู้วิจัยได้กำหนดค่าพารามิเตอร์ ดังนี้ Criterion มีค่าเป็น information_gain, Maximal depth มีค่าเท่ากับ 8, Apply pruning มีค่าเป็น True, ค่า Confidence มีค่าเท่ากับ 0.25

เทคนิค Random forest เป็นชุดของการจำแนกประเภทแบบไม่ตัดแต่งกิ่ง หรือต้นไม้ถดถอย ซึ่งถูกสร้างจากการนำข้อมูลฝึกสอนไปสุ่มตัวอย่างข้อมูลและคุณลักษณะข้อมูลแล้วนำมาสร้างเป็นต้นไม้ตัดสินใจ ซึ่งมีตัวอย่างส่วนหนึ่งที่ไม่ถูกเลือกเรียกข้อมูลส่วนนี้ว่า Out-of-Bag จะถูกนำมาในการทดสอบต้นไม้ตัดสินใจ [12] ซึ่งผู้วิจัยได้กำหนดค่าพารามิเตอร์ ดังนี้ Number of tree มีค่าเท่ากับ 50, Criterion มีค่าเป็น Information_gain, Maximal depth มีค่าเท่ากับ 9, Apply pruning มีค่าเป็น True และค่า Confidence มีค่าเท่ากับ 0.25

เทคนิค Bagging เป็นเทคนิคย่อยหนึ่งในเทคนิคการรวมกลุ่มตัวจำแนกประเภท (Ensemble Classify) โดยมีวิธีการสุ่มข้อมูลย่อยจากชุดข้อมูลฝึกสอนแบบสุ่มซ้ำได้ (Bootstraps replacement) จำนวนหลายเซตย่อย นำมาสร้างตัวจำแนกที่แตกต่างกันแล้วจึงนำผลการทำนายของแต่ละตัวจำแนกย่อยมาพิจารณาร่วมกัน [13] ซึ่งผู้วิจัยได้กำหนดค่าพารามิเตอร์ ดังนี้ Sample ratio มีค่าเท่ากับ 0.9, Iteration มีค่าเท่ากับ 7 และ Average Confidences มีค่าเป็น True

3.5 การวัดประสิทธิภาพและความแม่นยำของโมเดล (Evaluation)

ผู้วิจัยได้ดำเนินการวัดประสิทธิภาพของการจำแนกข้อมูลแต่ละเทคนิคโดยหาค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) ค่า F-measure และค่า AUC (Area Under Curve) [14] ในการเปรียบเทียบประสิทธิภาพ โดยมีวิธีการคำนวณ ดังนี้

ความถูกต้อง (Accuracy) มีสูตรคำนวณ คือ

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

ความแม่นยำ (Precision) มีสูตรคำนวณ คือ

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

ความระลึก (Recall) มีสูตรคำนวณ คือ

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

F-measure มีสูตรคำนวณ คือ

$$\text{F-measure} = \frac{2 \times (\text{Precision} \times \text{Recall})}{\text{Precision} \times \text{Recall}} \quad (4)$$

โดยที่ True Positive (TP) คือ จำนวนข้อมูลของสิ่งที่ทำนายว่าจริง และในความเป็นจริงก็เป็นเรื่องจริง

True Negative (TN) คือ จำนวนข้อมูลของสิ่งที่ทำนายว่าไม่จริง และในความเป็นจริงก็เป็นเรื่องไม่จริง

False Positive (FP) คือ จำนวนข้อมูลของสิ่งที่ทำนายว่าจริง แต่ในความเป็นจริงเป็นเรื่องไม่จริง

False Negative (FN) คือ จำนวนข้อมูลของสิ่งที่ทำนายว่าไม่จริง แต่ในความเป็นจริงเป็นเรื่องจริง

เมื่อคำนวณหาค่าประสิทธิภาพและความแม่นยำของโมเดลครบทุกเทคนิคแล้ว จึงนำค่าประสิทธิภาพทั้ง 3 เทคนิค มาเปรียบเทียบหาโมเดลที่มีประสิทธิภาพสูงสุด

ผลการวิจัย

ในการศึกษาเทคนิคพยากรณ์การได้รับปัจจัยพื้นฐานนักเรียนยากจนโรงเรียนวัดพระขาว (ประชานุเคราะห์) ด้วยเทคนิคเหมืองข้อมูล โดยมีข้อมูลในการวิจัยจำนวน 470 ข้อมูล ซึ่งหลังจากการทำให้ชุดข้อมูลมีลักษณะสมดุลด้วยวิธี SMOTE แล้วได้จำนวนข้อมูลเป็น 1,191 ชุดข้อมูล โดยศึกษาและเปรียบเทียบประสิทธิภาพของเทคนิควิธี และศึกษาแบบจำลองที่เหมาะสม ได้ผลดังนี้

1. การเปรียบเทียบประสิทธิภาพของการวิเคราะห์ข้อมูลการได้รับปัจจัยพื้นฐานนักเรียนยากจน ด้วยเทคนิคเหมืองข้อมูล

จากการทดสอบประสิทธิภาพของการจำแนกข้อมูลการได้รับปัจจัยพื้นฐานนักเรียนยากจนโรงเรียนวัดพระขาว (ประชานุเคราะห์) ด้วยเทคนิคเหมืองข้อมูลด้วยวิธี Decision tree, Random forest และ Bagging ได้ผลดังแสดงในตารางที่ 2

จาก Table 2 การทดสอบประสิทธิภาพของการจำแนกข้อมูลการได้รับปัจจัยพื้นฐานนักเรียนยากจนโรงเรียนวัดพระขาว(ประชานุเคราะห์) ด้วยเทคนิคเหมืองข้อมูล พบว่า เทคนิควิธี Random forest มีค่าความถูกต้อง (Accuracy) มีค่ามากที่สุด ร้อยละ 94.12 รองลงมาได้แก่ เทคนิควิธี Bagging และ เทคนิควิธี Decision tree โดยมีค่าความถูกต้องร้อยละ 91.21 และ 90.76 ตามลำดับ

Table 2. Identification efficiency by using 3 algorithms

Performance	Algorithm		
	Decision tree	Random forest	Bagging
Accuracy	90.76%	94.12%	91.60%
Precision	90.86%	94.37%	91.81%
Recall	90.76%	94.12%	91.60%
F-measure	90.81%	94.24%	91.70%

ส่วนการหาค่าความแม่นยำ (Precision) พบว่าเทคนิควิธี Random forest มีค่ามากที่สุดเช่นกัน ร้อยละ 94.37 รองลงมาได้แก่ เทคนิควิธี Bagging และ เทคนิควิธี Decision tree โดยมีค่าความถูกต้องร้อยละ 91.81 และ 90.86 ตามลำดับ

2. การค้นหาตัวแบบของการได้รับปัจจัยพื้นฐานนักเรียนยากจน ด้วยเทคนิคเหมืองข้อมูลที่เหมาะสม

ผู้วิจัยได้วิเคราะห์หาตัวแบบของการได้รับปัจจัยพื้นฐานนักเรียนยากจน ดังแสดงใน Figure 1 ซึ่งเป็นตัวแบบของการได้รับปัจจัยพื้นฐานนักเรียนยากจน ด้วยเทคนิควิธี Random forest ซึ่งมีประสิทธิภาพในการหาค่าความถูกต้อง และความแม่นยำมากกว่า เทคนิควิธี Decision tree และ Bagging จึงนำมาเป็นตัวแบบในการวิจัยในครั้งนี้

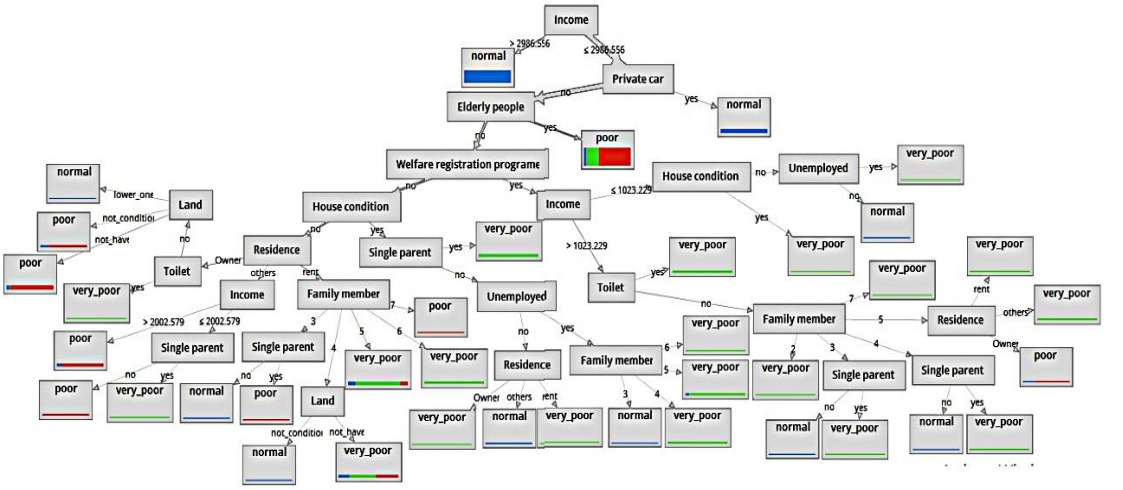

Figure 1. A model of a fund receiving of poor students by using Random forest technique

อภิปรายผล

จากผลการวิจัยสามารถอภิปรายผลได้ ดังนี้

1. ผู้วิจัยได้จัดเตรียมข้อมูลโดยการปรับสมดุลให้กับข้อมูลด้วยวิธี SMOTE ซึ่งทำให้ข้อมูลแต่ละคลาสเท่ากัน [7] เพื่อทำการจำแนกข้อมูลการได้รับปัจจัยพื้นฐานนักเรียนยากจนโรงเรียนวัดพระขาว(ประชานุเคราะห์) ด้วยเทคนิคเหมืองข้อมูล ได้แก่ Decision tree, Random forest และ Bagging พบว่าเทคนิควิธี Random forest มีค่าความถูกต้อง และความแม่นยำมากที่สุด สอดคล้องกับภรัณยา [15] ที่ได้ศึกษาการเพิ่มประสิทธิภาพเทคนิคต้นไม้ตัดสินใจบนชุดข้อมูลที่ไม่สมดุลโดยวิธีการสุ่มเพิ่มตัวอย่างกลุ่มน้อยสำหรับข้อมูลการเป็นโรคติดอินเทอร์เน็ต ที่ประสิทธิภาพของเทคนิค Random forest มีค่าสูงกว่าเทคนิควิธี Decision trees (J48), Iterative Dichotomiser 3 (ID3), Logistic Model Trees (LMT) และ Classification and Regression Trees (CART) แต่อย่างไรก็ตามค่าความถูกต้องและความแม่นยำของเทคนิควิธี Decision tree และ Bagging ก็ยังมีประสิทธิภาพในเกณฑ์สูง สามารถนำไปใช้ในการวิจัยได้เช่นกัน

2. ตัวแบบของการได้รับปัจจัยพื้นฐานนักเรียนยากจน เป็นตัวแบบที่ได้ด้วยเทคนิควิธี Random forest ซึ่งได้เลือกตัวแบบนี้เพราะ มีค่าประสิทธิภาพสูงสุด ในข้อมูลการวิจัยชุดนี้ซึ่งเป็นชุดข้อมูลของปีการศึกษา 2560 และ 2561 ดังนั้นสามารถนำตัวแบบที่ได้นี้ไปใช้ในการพยากรณ์การได้รับปัจจัยพื้นฐานนักเรียนยากจนในปีการศึกษาต่อไปได้

สรุปผล

ประสิทธิภาพของเทคนิคการจำแนกข้อมูลการได้รับปัจจัยพื้นฐานนักเรียนยากจน โรงเรียนวัดพระขาว (ประชานุเคราะห์) ทั้ง 3 วิธี ให้ค่าประสิทธิภาพระดับสูง เป็นที่ยอมรับได้ แต่เทคนิค Random forest เป็นเทคนิคที่มีประสิทธิภาพมากที่สุด จึงนำมาหาตัวแบบเพื่อทำนายการได้รับปัจจัยพื้นฐานนักเรียนยากจน แต่หากให้มีความคงตัวของตัวแบบมากยิ่งขึ้น ควรเก็บข้อมูลเพื่อใช้

ในการทดสอบให้เพิ่มขึ้น เช่น เก็บข้อมูลอีกครั้งในปีการศึกษาต่อ ๆ ไป แล้วทดสอบหาค่าประสิทธิภาพและค้นหาตัวแบบอีกครั้งจนกว่าตัวแบบจะมีความคงตัวเพื่อที่จะได้นำไปใช้ประโยชน์ต่อไป

เอกสารอ้างอิง

1. สำนักงานคณะกรรมการการศึกษาแห่งชาติ. 2553. พระราชบัญญัติการศึกษาแห่งชาติ พุทธศักราช 2542 ฉบับแก้ไขเพิ่มเติม 2553. กรุงเทพมหานคร : พริกหวานกราฟฟิค.
2. ไทย. กฎหมาย, พระราชบัญญัติ ฯลฯ. 2561. พระราชบัญญัติกองทุนเพื่อความเสมอภาคทางการศึกษา พ.ศ. 2561. กองทุนเพื่อความเสมอภาคทางการศึกษา.
3. อรนุช พันโท. 2559. การสังเคราะห์โมเดลการเรียนรู้ด้วยอีเลิร์นนิ่งแบบปรับเหมาะที่มีระบบพี่เลี้ยงอิเล็กทรอนิกส์โดยใช้รูปแบบการเรียนรู้ VARK ที่วิเคราะห์ด้วยวิธีเหมืองข้อมูล. วิทยานิพนธ์ปรัชญาดุษฎีบัณฑิต สาขาเทคโนโลยีสารสนเทศ ภาควิชาเทคโนโลยีบัณฑิตวิทยาลัย มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ.
4. รัตนา ลีรุ่งนาวารัตน์ ดวงกมล โพธิ์นาค และนัยนพัศ อินจวงจิรกิตต์. 2562. ระบบแนะนำสถานที่ฝึกประสบการณ์วิชาชีพเชิงพื้นที่ตามถนัดด้านเทคโนโลยีสารสนเทศของนักศึกษาฝึกประสบการณ์วิชาชีพด้วยเทคนิคทางด้านเหมืองข้อมูล. *วารสารวิทยาศาสตร์แห่งมหาวิทยาลัยราชภัฏเพชรบุรี*. 16(1): 42-54.
5. สำนักงานกองทุนเพื่อความเสมอภาคทางการศึกษา. คู่มือการดำเนินงานระบบคัดกรองปัจจัยพื้นฐานนักเรียนยากจน และคู่มือการรับเงินอุดหนุนสำหรับนักเรียนยากจนพิเศษแบบมีเงื่อนไข. [ออนไลน์] เข้าถึงได้จาก: https://www.eef.or.th/wp-content/uploads/2018/11/aw14.pdf. 2561.
6. Chapman, P., Clinton, J., Kerber, R. Khabaza, T., Reinartz, T., Shearer, C. and Wirth R. 2000. *CRISP-DM 1.0 Step-by-step data mining guide, Technical report*, SPSS inc.
7. Chawla, N. V., Bowyer, K. W., Hall, L. O. and Kegelmayer, W. P. 2002. SMOTE: Synthetic Minority Over- Sampling Technique. *Journal of Artificial Intelligent Research*. 321-357.
8. Alls, O. F., Karakurt, E. and Melli, P. 2000. Data mining for database marketing at Garanti Bank. *WIT Transactions on Information and Communication Technologies*. 25: 93-107.
9. Hong, H., Liu, J., Bui, D. T., Pradhan, B., Acharya, T. D., Pham, B.T. and Ahmad, B.B. 2018. Landslide susceptibility mapping using J48 decision tree with AdaBoost, Bagging and Rotation forest ensembles in the Guangchang area (China). *Catena*. 163: 399-413.
10. González, C., Mira-McWilliams, J. and Juárez, I. 2015. Important variable assessment and electricity price forecasting based on regression tree models: classification and regression trees, bagging and random forests. *IET Generation, Transmission & Distribution*. 9 (11): 1120-1128.

11. เอกสิทธิ์ พัชรวงศ์ศักดา. 2557. *การวิเคราะห์ข้อมูลด้วยเทคนิคดาต้า ไมน์นิง เบื้องต้น*. กรุงเทพมหานคร: ดาต้าคิวบ์.

12. Breiman, L. 2001. Random forests. *Machine Learning*. 45(1): 5-32.

13. Breiman, L. 1996. Bagging predictors. *Machine Learning*. 24(2): 123-140.

14. Mohammed, J., Meira Wagner, J.R. 2014. *Data Mining and Analysis Fundamental Concepts and Algorithms*. Cambridge University Press. New York. USA.

15. ภรัณยา ปาลวิสุทธิ์. 2559. การเพิ่มประสิทธิภาพเทคนิคต้นไม้ตัดสินใจบนชุดข้อมูลที่ไม่สมดุล โดยวิธีการสุ่มเพิ่มตัวอย่างกลุ่มน้อยสำหรับข้อมูลการเป็นโรคติดอินเทอร์เน็ต. *วารสารเทคโนโลยีสารสนเทศ*. 12(1): 54-63.