

ระบบพยากรณ์เพื่อสนับสนุนการตัดสินใจศึกษาต่อหลังสำเร็จมัธยมศึกษาตอนต้น
โดยใช้อัลกอริธึมป่าสุ่ม
A PREDICTIVE SYSTEM TO SUPPORT DECISION-MAKING ON FURTHER
EDUCATION AFTER JUNIOR HIGH SCHOOL USING THE RANDOM
FOREST ALGORITHM

วรเทพ วันกาล^{1*} ณัฐพร เห็นเจริญเลิศ² และ วุฒิชัย ร่มสายหยุด³
Worathape Wankan^{1*}, Nuttaporn Hencharoenlert² and Walisa Romsaiyud³

สังกัด (สาขาวิชาวิทยาศาสตร์และเทคโนโลยี แขนงวิชาเทคโนโลยีสารสนเทศและการสื่อสาร มหาวิทยาลัยสุโขทัยธรรมาธิราช)^{1*,2,3}

*Corresponding author. E-mail: worathape.wan@gmail.com

บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์เพื่อ 1) สร้างแบบจำลองพยากรณ์เพื่อสนับสนุนการตัดสินใจศึกษาต่อหลังสำเร็จมัธยมศึกษาตอนต้น โดยใช้อัลกอริธึมป่าสุ่ม 2) ประเมินประสิทธิภาพของแบบจำลองพยากรณ์เพื่อสนับสนุนการตัดสินใจศึกษาต่อหลังสำเร็จมัธยมศึกษาตอนต้น และ 3) พัฒนาระบบพยากรณ์เพื่อสนับสนุนการตัดสินใจศึกษาต่อหลังสำเร็จมัธยมศึกษาตอนต้น

ผู้วิจัยใช้ชุดข้อมูลผลสัมฤทธิ์ทางการเรียนของนักเรียนโรงเรียนขยายโอกาสทางการศึกษาชั้นมัธยมศึกษาปีที่ 3 ในสังกัดสำนักงานเขตพื้นที่การศึกษาประถมศึกษานครสวรรค์ เขต 2 ตั้งแต่ปีการศึกษา 2558 - 2560 ได้ถูกนำมาวิเคราะห์โดยนำกลุ่มตัวอย่างข้อมูลของผลสัมฤทธิ์ทางการเรียน จำนวน 1,030 เรคคอร์ด ซึ่งประกอบไปด้วย 8 กลุ่มสาระวิชา 7 คุณลักษณะ นำมาสร้างแบบจำลองทดสอบประสิทธิภาพ โดยใช้อัลกอริธึม 3 อัลกอริธึม ได้แก่ 1) อัลกอริธึม C4.5(J48) 2) อัลกอริธึมต้นไม้สุ่ม และ 3) อัลกอริธึมป่าสุ่ม เพื่อทำการเปรียบเทียบประสิทธิภาพแบบจำลองพยากรณ์ และใช้การเลือกคุณลักษณะแบบ InfoGainAttributeEval ทดสอบประสิทธิภาพด้วยวิธี 10-fold cross-validation โดยใช้โปรแกรม Weka แล้วนำผลการทดสอบที่ได้มาพัฒนาระบบพยากรณ์เพื่อสนับสนุนการตัดสินใจศึกษาต่อหลังสำเร็จมัธยมศึกษาตอนต้นใช้งานผ่านเว็บเบราว์เซอร์

ผลการวิจัยนี้พบว่าอัลกอริธึมป่าสุ่มมีค่าประสิทธิภาพมากที่สุดค่าความถูกต้องร้อยละ 73.20 ค่าความแม่นยำเท่ากับร้อยละ 73.0 ค่าความครบถ้วนเท่ากับร้อยละ 73.0 ค่าประสิทธิภาพโดยรวมเท่ากับร้อยละ 73.0

คำสำคัญ: ระบบพยากรณ์ทางการเรียน ระบบสนับสนุนการตัดสินใจ อัลกอริธึมป่าสุ่ม กระบวนการคัดเลือกคุณลักษณะ

Abstract

The objectives of this study were: 1) to create the predictive system to support decision-making on further education after junior high school using the random forest algorithm, 2) to evaluate the efficiency of the predictive system to support decision-making on further education after junior high school, and 3) to develop the predictive system to support decision-making on further education after junior high school.

The researcher used the academic achievement data set of secondary school year 3 on educational opportunity expansion schools in Nakhonsawan Primary Education Service Area Office 2 in 2015-2017. The data were analyzed by education achievement scores of the sample of 1,030



records from 8 core topics and 7 features to create the efficiency testing model and using 3 algorithms for training and testing data 1) C4.5 (j48), 2) Random Tree, and 3) Random Forest to compare the efficiency of the testing model. Furthermore, using the InfoGainAttributeEval Method and for efficiency, testing using a 10-fold cross-validation model with Weka Program.

The testing result was used to develop the predictive system to support decision-making on further education after junior high school on web browsers. The finding found that the Random Forest Algorithm was the highest efficiency: Accuracy was 73.20%, Recall was 73.0%, Precision was 73.0%, and the F-measure was 73.0%.

Keywords: Predictive System, Support Decision-making, Random Forest Algorithm, Feature Selection

บทนำ

การศึกษาที่มีความสำคัญต่อการพัฒนาคนในสังคมให้มีความรู้ เป็นการวางรากฐานของความเจริญก้าวหน้าของประเทศ การจัดการศึกษาในประเทศไทยมีระบบการศึกษาขั้นพื้นฐาน ซึ่งจัดไม่น้อยกว่าสิบสองปีก่อนระดับอุดมศึกษา ปัจจุบันการศึกษาในระบบได้แบ่งออกเป็น 2 ระดับ คือ 1) การศึกษาขั้นพื้นฐาน และ 2) การศึกษาระดับอุดมศึกษา ในการศึกษาขั้นพื้นฐานเมื่อผู้เรียนจบชั้นมัธยมศึกษาปีที่ 3 จะต้องตัดสินใจเลือกศึกษาต่อในระดับมัธยมศึกษาตอนปลาย หรือการศึกษาต่อประเภทอาชีวศึกษา ในการศึกษาต่อแต่ละประเภทการศึกษาอาจดูได้จากระดับคะแนนหรือผลสัมฤทธิ์ทางการเรียนที่ผ่านมาของนักเรียนแต่ละคน ว่าในแต่ละวิชาได้ระดับคะแนนผลการเรียนอยู่ในระดับใด ถ้าได้ระดับผลการเรียนในวิชานั้นสูง แสดงว่านักเรียนมีความรู้ความเข้าใจในการเรียนวิชานั้นมาก แต่ถ้าระดับคะแนนผลการเรียนในวิชานั้นต่ำ แสดงว่านักเรียนมีความรู้ความเข้าใจในการเรียนวิชานั้นน้อย เพื่อเลือกแนวทางการศึกษาต่อ ในส่วนการศึกษาต่อประเภทอาชีวศึกษา เป็นอีกทางเลือกสำหรับผู้จบชั้นมัธยมศึกษาปีที่ 3 ซึ่งมุ่งศึกษาและฝึกฝนวิชาชีพสาขาใดสาขาหนึ่งโดยเฉพาะเพื่อจบออกมาประกอบอาชีพต่าง ๆ

โรงเรียนขยายโอกาสทางการศึกษาในสังกัดสำนักงานเขตพื้นที่การศึกษาประถมศึกษานครสวรรค์ เขต 2 ในแต่ละปีมีนักเรียนที่จบการศึกษาชั้นมัธยมศึกษาปีที่ 3 ไม่มีความชัดเจนในเป้าหมายของการศึกษาต่อ เนื่องจากโรงเรียนขาดบุคลากรครูแนะแนวซึ่งเป็นปัญหาต้นทางของเรื่อง นักเรียนจำนวนมากขาดการแนะแนวในด้านการศึกษาต่อ ส่วนใหญ่พบว่าเลือกศึกษาต่อตามเพื่อน ตามคำแนะนำของผู้ปกครอง ทำให้ไม่ประสบความสำเร็จในการเรียน เรียนไม่จบหลักสูตร เปลี่ยนสายการศึกษาต่อ และบางคนต้องเลิกศึกษาต่อทั้งที่ยังไม่จบการศึกษา เนื่องจากการตัดสินใจในการศึกษาต่อของนักเรียนยังขาดความรู้ ความเข้าใจเกี่ยวกับประเภทการศึกษาต่อ ทำให้ต้องเสียเวลาไปกับการเรียนที่ไม่มีเป้าหมาย

กระบวนการเหมืองข้อมูล (Crisp-DM) (เอกลีที พัทธวงค์ศักดิ์, 2561) เป็นวิธีการในการวิเคราะห์ข้อมูลปริมาณมากมีนักวิจัยได้นำเอาเทคนิคเหมืองข้อมูลมาใช้ในการพยากรณ์ ดังเช่นสำราญ วานนท์ (2561) ทำการศึกษาเทคนิคพยากรณ์ อาชีพสำหรับนักศึกษาในระดับปริญญาตรีสาขาคอมพิวเตอร์โดยใช้เทคนิคเหมืองข้อมูล และวัชรินทร์ จิตต์สกุล (2560) ได้ทำการวิเคราะห์การจำแนกข้อความด้วยการเปรียบเทียบความเสถียรของอัลกอริธึมซึ่งพบว่าอัลกอริธึม Random Forest มีค่าความเสถียรการจำแนกได้มากที่สุด ในขั้นตอนการเลือกคุณลักษณะ (feature selection) เป็นขั้นตอนสำคัญของการทำเหมืองข้อมูล อย่างที่ อธิษฐ์ คุณสุข และ จาริ ทองคำ (2562, 139) ได้ใช้กระบวนการคัดเลือกคุณลักษณะสำหรับเพิ่มประสิทธิภาพในการจำแนกความคิดเห็นของลูกค้า ร่วมกับการทดสอบแบบ 10-fold cross validation พบว่าให้ความถูกต้องที่ 89.08%

ผู้วิจัยจึงเล็งเห็นความสำคัญของการให้คำปรึกษาแนะแนวการศึกษาต่อให้กับนักเรียน จึงมีแนวความคิดในการนำผลสัมฤทธิ์ทางการเรียนของนักเรียนรายบุคคลในแต่ละภาคเรียนมาผ่านขั้นตอนกระบวนการทำเหมืองข้อมูล

ซึ่งใช้วิธีการคัดเลือกคุณลักษณะ แล้วนำมาเปรียบเทียบสร้างตัวแบบพยากรณ์ด้วย 3 อัลกอริธึม คือ 1) อัลกอริธึม C4.5(J48) (อรนุช พันโท, 2557) 2) อัลกอริธึมต้นไม้สุ่ม และ 3) อัลกอริธึมป่าสุ่ม (ประเมษฐ์ ชันวานนท์และคณะ, 2560) เพื่อวิเคราะห์ประสิทธิภาพให้ได้ตัวแบบพยากรณ์ที่ดีที่สุด สร้างตัวแบบพยากรณ์เพื่อสนับสนุนการตัดสินใจศึกษาต่อ หลังสำเร็จมัธยมศึกษาตอนต้น ในการพัฒนาระบบพยากรณ์ที่มีประสิทธิภาพในการแนะแนวการศึกษาต่อของนักเรียน เพื่อใช้เป็นแนวทางในการตัดสินใจเลือกประเภทในการศึกษาต่อ

วัตถุประสงค์ของการวิจัย

1. เพื่อสร้างแบบจำลองพยากรณ์เพื่อสนับสนุนการตัดสินใจศึกษาต่อหลังสำเร็จมัธยมศึกษาตอนต้น โดยใช้อัลกอริธึมป่าสุ่ม
2. เพื่อประเมินประสิทธิภาพของแบบจำลองพยากรณ์เพื่อสนับสนุนการตัดสินใจศึกษาต่อหลังสำเร็จมัธยมศึกษาตอนต้น
3. เพื่อพัฒนาระบบพยากรณ์เพื่อสนับสนุนการตัดสินใจศึกษาต่อหลังสำเร็จมัธยมศึกษาตอนต้น

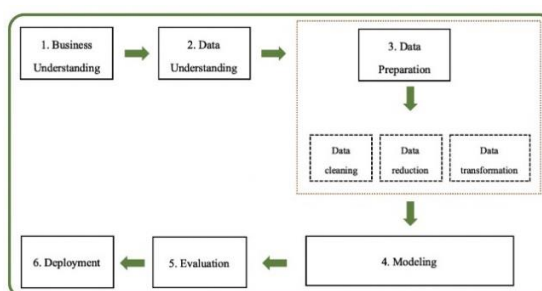
กรอบแนวคิดการวิจัย

ผู้วิจัยมุ่งเน้นการพยากรณ์จากผลสัมฤทธิ์ทางการเรียนของนักเรียนชั้นมัธยมศึกษาปีที่ 3 ในสังกัดสำนักงานเขตพื้นที่การศึกษาประถมศึกษานครสวรรค์ เขต 2 จังหวัดนครสวรรค์ ย้อนหลัง 3 ปี ตั้งแต่ปีการศึกษา 2558 - 2560 ใช้กลุ่มตัวอย่างข้อมูลของผลสัมฤทธิ์ทางการเรียน 8 กลุ่มสาระวิชาย้อนหลัง 6 ภาคเรียน (2558 - 2560) ในการศึกษาต่อประเภทสามัญศึกษาหรือประเภทอาชีวศึกษา ในการให้คำแนะนำของครูแนะแนว มีตัวแปรที่ศึกษา ดังนี้

1. แหล่งข้อมูลที่ใช้เป็นตัวแปรต้น คือ ชุดข้อมูลผลสัมฤทธิ์ทางการเรียนของนักเรียนชั้นมัธยมศึกษาปีที่ 3 สำนักงานเขตพื้นที่การศึกษาประถมศึกษานครสวรรค์ เขต 2 จำนวน 1,160 คน
2. ตัวแปรตาม คือ ผลการพยากรณ์ของนักเรียนแต่ละคนเพื่อสนับสนุนการตัดสินใจในการศึกษาต่อตามความถนัด

วิธีดำเนินการวิจัย

นำชุดข้อมูลผลสัมฤทธิ์ทางการเรียนของนักเรียนระดับชั้นมัธยมศึกษาปีที่ 3 ในสังกัดสำนักงานเขตพื้นที่การศึกษาประถมศึกษานครสวรรค์ เขต 2 ระหว่างปีการศึกษา 2558 – 2560 ซึ่งจะจัดอยู่ในรูปของไฟล์ .csv จำนวน 1,160 เรคคอร์ด 21 คุณลักษณะ มาผ่านขั้นตอนกระบวนการทำเหมืองข้อมูล เพื่อให้ได้ผลที่มีประสิทธิภาพ ซึ่งได้ชุดข้อมูลที่ผ่านกระบวนการ จำนวน 1,030 เรคคอร์ด จำนวน 11 คุณลักษณะ โดยใช้เครื่องมือในการประมวลผล คือ โปรแกรม WEKA ในการสร้างตัวแบบพยากรณ์ (Model) และทดสอบอัลกอริธึมตามแนวทางกระบวนการทำเหมืองข้อมูล ประกอบด้วย 6 ขั้นตอน ดังภาพที่ 1

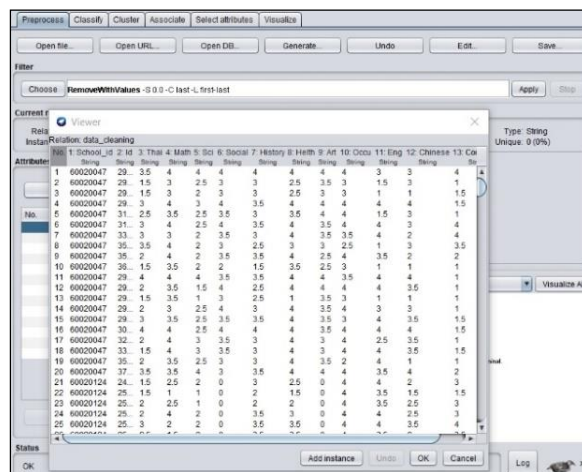


ภาพที่ 1 แสดงขั้นตอนการดำเนินการวิจัย

ขั้นตอนที่ 1 ทำความเข้าใจกับโจทย์ที่ต้องทำ เป็นขั้นตอนของการศึกษาปัญหาแนวทางเป้าหมายการศึกษาของนักเรียนโรงเรียนขยายโอกาสทางการศึกษาในสังกัดสำนักงานเขตพื้นที่การศึกษาประถมศึกษา นครสวรรค์ เขต 2 ที่สำเร็จการศึกษาชั้นมัธยมศึกษาปีที่ 3

ขั้นตอนที่ 2 ทำความเข้าใจกับข้อมูลที่ใช้ในการวิจัย นำข้อมูลผลสัมฤทธิ์ทางการเรียนจำนวน 6 ภาคเรียน ระหว่างปีการศึกษา 2558 - 2560 ของนักเรียนชั้นมัธยมศึกษาปีที่ 3 ในสังกัดสำนักงานเขตพื้นที่การศึกษา ประถมศึกษานครสวรรค์ เขต 2 จังหวัดนครสวรรค์ จำนวน 1,160 คน (ข้อมูลบริหารจัดการสถานศึกษาจากสำนักงานเขตพื้นที่การศึกษาประถมศึกษานครสวรรค์ เขต 2 จังหวัดนครสวรรค์ ปีการศึกษา 2560)

ขั้นตอนที่ 3 การเตรียมข้อมูล นำข้อมูลที่ได้อยู่ในรูปแบบของไฟล์ Excel จำนวน 1,160 เรคคอร์ด ซึ่งยังไม่สามารถนำมาวิเคราะห์ข้อมูลได้ เพื่อให้ได้ผลการทดลองที่มีประสิทธิภาพข้อมูลจะผ่านกระบวนการ 3 ขั้นตอนคือ 1) การกลั่นกรองข้อมูล (data cleaning) เพื่อเป็นการตรวจสอบความถูกต้องของข้อมูล ข้อมูลหากเกิดข้อผิดพลาด ข้อมูลไม่สอดคล้อง ข้อมูลขาดหาย จะได้ทำการปรับค่าข้อมูลเพื่อนำไปใช้ในการวิเคราะห์ด้วยเทคนิคเหมืองข้อมูล ในการตรวจสอบข้อมูล data set นี้จะใช้ด้วยโปรแกรม WEKA ทำการลบค่าของข้อมูลที่ขาดหายไป (Remove with value) คงเหลือชุดข้อมูลที่นำไปใช้ จำนวน 1,030 เรคคอร์ด



No.	School ID	Thai	Math	Sci	Social	History	Health	Art	Occu	Eng	Chinese	Co
1	60020047	29	2.5	4	4	4	4	4	4	3	3	4
2	60020047	29	1.5	3	2.5	3	3	2.5	3.5	3	1.5	3
3	60020047	29	1.5	3	2	3	3	2.5	3	3	1	1.5
4	60020047	29	3	4	3	4	3.5	4	4	4	4	1.5
5	60020047	31	2.5	3.5	2.5	3.5	3	3.5	4	4	1.5	3
6	60020047	31	3	4	2.5	4	3.5	4	3.5	4	4	3
7	60020047	33	3	3	2	3.5	3	4	3.5	3.5	4	2
8	60020047	35	3.5	4	2	3	2.5	3	3	2.5	1	3
9	60020047	35	2	4	2	3.5	3.5	4	2.5	4	3.5	2
10	60020047	36	1.5	3.5	2	2	1.5	3.5	2.5	3	1	1
11	60020047	29	4	4	4	3.5	3.5	4	4	3.5	4	4
12	60020047	29	2	3.5	1.5	4	2.5	4	4	4	4	3.5
13	60020047	29	1.5	3.5	1	3	2.5	4	3.5	3	1	1
14	60020047	29	2	3	2.5	4	3	4	3.5	4	3	3
15	60020047	29	3	3.5	2.5	3.5	3.5	4	3.5	3	4	3.5
16	60020047	33	4	4	2.5	4	4	3.5	4	4	4	1.5
17	60020047	32	4	3	3.5	3	4	3	4	4	2.5	3.5
18	60020047	33	1.5	4	3	3.5	3	4	3	4	4	3.5
19	60020047	35	2	3.5	2.5	2	3	4	3.5	2	4	1
20	60020047	37	3.5	3.5	4	3	3.5	4	4	4	3.5	4
21	60020124	24	1.5	2.5	2	0	3	2.5	0	4	4	2
22	60020124	25	1.5	1	1	0	2	1.5	0	4	3.5	1.5
23	60020124	25	2	2.5	1	0	2	2	0	4	3.5	2.5
24	60020124	25	2	4	2	0	3.5	3	0	4	4	2.5
25	60020124	25	3	2	2	0	3.5	3.5	0	4	4	3.5

ภาพที่ 2 ตัดข้อมูลที่มีความผิดปกติและข้อมูลขาดหายไป

2) การลดขนาดข้อมูล (data reduction) เนื่องจากข้อมูลมีปริมาณคุณลักษณะมากไป เพื่อให้ได้ข้อมูลที่สมบูรณ์ถูกต้อง สะดวกต่อการนำข้อมูลไปดำเนินการวิเคราะห์ โดยลดจำนวนของคุณลักษณะที่ไม่เกี่ยวข้องกับข้อมูลนักเรียนและผลสัมฤทธิ์บางตัวออกไป คงเหลือคุณลักษณะที่จำเป็นในการนำไปใช้สร้างแบบพยากรณ์ จำนวน 11 คุณลักษณะ ได้แก่ 1) เพศ 2) รายได้ครอบครัวต่อเดือน 3) เกรดเฉลี่ยกลุ่มสาระภาษาไทย 4) เกรดเฉลี่ยกลุ่มสาระคณิตศาสตร์ 5) เกรดเฉลี่ยกลุ่มสาระวิทยาศาสตร์ 6) เกรดเฉลี่ยกลุ่มสาระภาษาต่างประเทศ 7) เกรดเฉลี่ยกลุ่มสาระสังคมศึกษา ศาสนา และวัฒนธรรม 8) เกรดเฉลี่ยกลุ่มสาระสุขศึกษาพลศึกษา 9) เกรดเฉลี่ยกลุ่มสาระศิลปะ 10) เกรดเฉลี่ยกลุ่มสาระการงานอาชีพและเทคโนโลยี และ 11) ผลการเรียนเฉลี่ยตลอดหลักสูตรการศึกษา ดังภาพที่ 3

Gender	Income	Thai	Math	Sci	Foreign	Social	Health	Art	Career	GPA
ช	5,800	2.54	2.60	2.71	2.60	3.12	3.50	3.17	2.81	2.83
ช	6,500	1.81	1.77	1.79	1.30	2.35	3.33	2.17	2.41	2.04
ช	3,500	2.52	2.42	2.83	1.98	3.48	3.75	3.58	3.34	2.91
ช	8,000	2.73	2.54	2.67	1.93	3.20	3.75	3.50	3.47	2.89
ญ	3,500	2.23	2.56	2.81	1.45	3.20	3.58	3.42	3.28	2.74
ญ	4,500	1.48	1.08	1.69	1.38	2.05	2.83	2.00	2.03	1.74
ญ	7,500	1.46	1.25	1.90	1.38	1.98	3.25	2.17	2.31	1.85
ญ	7,100	3.40	3.50	3.83	2.58	3.77	3.75	3.75	3.75	3.53
ญ	5,000	2.67	2.13	2.90	1.80	3.57	3.83	3.08	3.25	2.85
ญ	8,000	2.96	2.83	3.46	2.58	3.45	3.42	3.33	3.41	3.16
ญ	2,500	2.29	2.06	2.27	1.45	2.70	3.83	2.92	2.84	2.44
ญ	3,850	2.44	2.35	2.54	1.73	2.83	3.25	2.75	2.84	2.55
ญ	5,000	1.75	1.60	2.00	1.38	2.53	3.17	2.33	2.63	2.10
ญ	9,000	1.88	1.31	1.83	1.28	2.52	3.17	2.25	2.63	2.03
ญ	10,000	1.65	1.27	2.08	1.53	2.52	3.00	2.67	2.63	2.07
ช	4,000	1.85	1.73	1.88	2.00	2.32	2.83	2.67	2.38	2.13
ช	10,000	2.10	1.65	2.29	1.73	2.95	2.92	2.58	3.00	2.36
ช	9,000	3.56	2.08	2.77	2.58	2.87	3.33	3.50	3.44	2.94
ช	10,000	3.85	3.56	3.77	3.75	3.93	3.92	4.00	3.69	3.80
ช	2,500	3.02	2.60	2.83	2.10	3.20	3.83	3.25	3.41	2.97

ภาพที่ 3 การลดจำนวนของคุณลักษณะที่ไม่เกี่ยวข้องกับข้อมูลนักเรียนและผลสัมฤทธิ์

3) การแปลงข้อมูล (data transformation) ทำการคำนวณหาค่าผลสัมฤทธิ์ทางการเรียนแต่ละกลุ่มสาระการเรียนรู้ของรายวิชาในแต่ละภาคเรียนด้วยโปรแกรม Microsoft Excel 2013 จากนั้นแปลงข้อมูลที่มีความหมายใกล้เคียงกันมาอยู่ในระดับเดียวกันเป็นค่าไม่ต่อเนื่อง (Discretization) โดยแบ่งค่าของข้อมูลออกเป็นช่วงย่อย ๆ จะเป็นการช่วยลดการประมวลผลในการทำเหมืองข้อมูล และลดการซ้ำซ้อนการกระจายของข้อมูล ดังตารางที่ 1

ตารางที่ 1 การแปลงข้อมูล

คุณลักษณะ	รายละเอียด	การแปลงข้อมูล
Gender	เพศ	Male = ผู้ชาย Female = ผู้หญิง
Income	รายได้ครอบครัวต่อเดือน	High = รายได้ต่อเดือนตั้งแต่ 33,301 บาท Medium = รายได้ตั้งแต่ 11,101 - 33,300 บาท Low = รายได้น้อยกว่า 11,100 บาท
Thai	เกรดเฉลี่ยกลุ่มสาระภาษาไทย	High = ผลการเรียนเฉลี่ย 3.01 - 4.00 Medium = ผลการเรียนเฉลี่ย 2.01 - 3.00 Low = ผลการเรียนเฉลี่ย 1.00 - 2.00
Math	เกรดเฉลี่ยกลุ่มสาระคณิตศาสตร์	High = ผลการเรียนเฉลี่ย 3.01 - 4.00 Medium = ผลการเรียนเฉลี่ย 2.01 - 3.00 Low = ผลการเรียนเฉลี่ย 1.00 - 2.00
Sci	เกรดเฉลี่ยกลุ่มสาระวิทยาศาสตร์	High = ผลการเรียนเฉลี่ย 3.01 - 4.00 Medium = ผลการเรียนเฉลี่ย 2.01 - 3.00 Low = ผลการเรียนเฉลี่ย 1.00 - 2.00
Foreign	เกรดเฉลี่ยกลุ่มสาระภาษาต่างประเทศ	High = ผลการเรียนเฉลี่ย 3.01 - 4.00 Medium = ผลการเรียนเฉลี่ย 2.01 - 3.00 Low = ผลการเรียนเฉลี่ย 1.00 - 2.00
Social	เกรดเฉลี่ยกลุ่มสาระสังคมศึกษา และวัฒนธรรม	High = ผลการเรียนเฉลี่ย 3.01 - 4.00 Medium = ผลการเรียนเฉลี่ย 2.01 - 3.00 Low = ผลการเรียนเฉลี่ย 1.00 - 2.00



ตารางที่ 1 การแปลงข้อมูล (ต่อ)

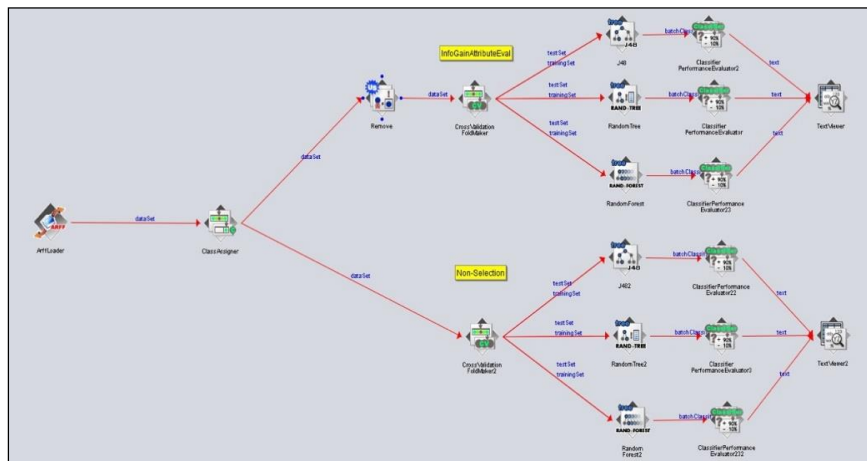
คุณลักษณะ	รายละเอียด	การแปลงข้อมูล
Health	เกรดเฉลี่ยกลุ่มสาระสุขภาพพลศึกษา	High = ผลการเรียนเฉลี่ย 3.01 - 4.00 Medium = ผลการเรียนเฉลี่ย 2.01 - 3.00 Low = ผลการเรียนเฉลี่ย 1.00 - 2.00
Art	เกรดเฉลี่ยกลุ่มสาระศิลปะ	High = ผลการเรียนเฉลี่ย 3.01 - 4.00 Medium = ผลการเรียนเฉลี่ย 2.01 - 3.00 Low = ผลการเรียนเฉลี่ย 1.00 - 2.00
Career	เกรดเฉลี่ยกลุ่มสาระการงานอาชีพและเทคโนโลยี	High = ผลการเรียนเฉลี่ย 3.01 - 4.00 Medium = ผลการเรียนเฉลี่ย 2.01 - 3.00 Low = ผลการเรียนเฉลี่ย 1.00 - 2.00
GPA	เกรดเฉลี่ยสะสม 6 เทอม	Excellent = ผลการเรียนเฉลี่ย 3.51 - 4.00 Very Good = ผลการเรียนเฉลี่ย 3.01 - 3.50 Good = ผลการเรียนเฉลี่ย 2.51 - 3.00 Above average = ผลการเรียนเฉลี่ย 2.01 - 2.50 Average = ผลการเรียนเฉลี่ย 1.51 - 2.00 Fair = ผลการเรียนเฉลี่ย 1.00 - 1.50
Course	แผนการศึกษาต่อ	BEC = ประเภทสามัญศึกษา VEC = ประเภทอาชีวศึกษา

ขั้นตอนที่ 4 การสร้างตัวแบบพยากรณ์ ใช้การคัดเลือกแบบ feature selection วิธีแบบไม่เลือกคุณลักษณะ (Non-Selection) และวิธีแบบเลือกคุณลักษณะ (InfoGainAttributeEval) โดยวัด Information Gain (IG) (Xu & Jiang, 2015) เป็นตัววัดความสัมพันธ์ของคุณลักษณะให้กับคลาส โดยเลือกค่าเรียงลำดับที่มีค่าระหว่าง 0.1 – 0.03 ซึ่งมีจำนวน 7 คุณลักษณะ ได้แก่ 1) GPA 2) Thai 3) Social 4) Math 5) Sci 6) Foreign และ 7) Career โดยมีจำนวน 2 เป้าหมายหรือผลลัพธ์ (Class) คือ 1) สายสามัญศึกษา (Bec) และ 2) สายอาชีวศึกษา (Vec) เพื่อนำไปสร้างตัวแบบพยากรณ์

ตารางที่ 2 คุณลักษณะที่ใช้ในการพยากรณ์

คุณลักษณะ	การเลือกคุณลักษณะ	
	ไม่เลือกคุณลักษณะ	เลือกคุณลักษณะ
Gender	✓	
Income	✓	
Thai	✓	✓
Math	✓	✓
Sci	✓	✓
Foreign	✓	✓
Social	✓	✓
Health	✓	
Art	✓	
Career	✓	✓
GPA	✓	✓

สร้างตัวแบบพยากรณ์และทำการทดสอบประสิทธิภาพ แบบ 10-fold cross validation ซึ่งเป็นการแบ่งข้อมูลที่ใช้ในการทดสอบแบ่งออกเป็นชุด ข้อมูลสอน 9 ชุด และ 1 ชุด เป็นชุดทดสอบประสิทธิภาพของตัวแบบพยากรณ์ ทำแบบนี้วนไปจนครบจำนวนชุดที่แบ่งไว้ เพื่อสร้างและทดสอบหาความน่าเชื่อถือ โดยนำชุดข้อมูลจำนวน 1,030 เรคคอร์ด ร่วมกับการคัดเลือกคุณลักษณะ แบบวิธีไม่เลือกคุณลักษณะและวิธีเลือกคุณลักษณะด้วยวิธี InfoGainAttributeEval และทำการเปรียบเทียบด้วย 3 อัลกอริทึม คือ 1) อัลกอริทึม C4.5(J48) 2) อัลกอริทึมต้นไม้สุ่ม Random Tree และ 3) อัลกอริทึมป่าสุ่ม Random Forest เพื่อวิเคราะห์วัดประสิทธิภาพในการจำแนกของตัวแบบพยากรณ์ที่ถูกต้องที่สุด



ภาพที่ 4 แสดงการวิเคราะห์ประสิทธิภาพด้วย KnowledgeFlow

ขั้นตอนที่ 5 การประเมินผล ทำการทดสอบประสิทธิภาพของตัวแบบพยากรณ์ เพื่อวัดประสิทธิภาพค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าครบถ้วน (Recall) และค่าประสิทธิภาพโดยรวม (F-Measure) จากตาราง confusion matrix เพื่อการประเมินผลลัพธ์การทำนาย ดังภาพที่ 5

Confusion Matrix

	Actually Positive (1)	Actually Negative (0)
Predicted Positive (1)	True Positives (TPs)	False Positives (FPs)
Predicted Negative (0)	False Negatives (FNs)	True Negatives (TNs)

ภาพที่ 5 ตาราง confusion matrix

ที่มา : <https://glassboxmedicine.com/2019/02/17/measuring-performance-the-confusion-matrix>

- True Positive (TP) คือ สิ่งที่โปรแกรมทำนายว่าจริงและสิ่งที่เกิดขึ้นเป็นจริง
- True Negative (TN) คือ สิ่งที่โปรแกรมทำนายว่าไม่จริงและสิ่งที่เกิดขึ้นไม่จริง
- False Positive (FP) คือ สิ่งที่โปรแกรมทำนายว่าจริงแต่สิ่งที่เกิดขึ้นไม่จริง
- False Negative (FN) คือ สิ่งที่ทำนายว่าไม่จริงแต่สิ่งที่เกิดขึ้นเป็นจริง

การประชุมสัมมนาวิชาการนำเสนอผลงานวิจัยระดับชาติเครือข่ายบัณฑิตศึกษา มหาวิทยาลัยราชภัฏภาคเหนือ ครั้งที่ 22 "ชีวิตวิถีใหม่ของการศึกษาในสังคมดิจิทัล"



ค่าความถูกต้อง (Accuracy)

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \quad (1)$$

ค่าความแม่นยำ (Precision)

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

ค่าครบถ้วน (Recall)

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

ค่าประสิทธิภาพโดยรวม (F-Measure)

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

```

=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      754      73.2039 %
Incorrectly Classified Instances    276      26.7961 %
Kappa statistic                     0.4642
Mean absolute error                 0.3388
Root mean squared error             0.4287
Relative absolute error             67.7952 %
Root relative squared error        85.76 %
Total Number of Instances         1030

=== Detailed Accuracy By Class ===
                TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                0.744   0.279   0.718     0.744   0.731     0.465   0.801    0.776    Vec
                0.721   0.256   0.746     0.721   0.733     0.465   0.801    0.786    Bec
Weighted Avg.   0.732   0.267   0.733     0.732   0.732     0.465   0.801    0.781

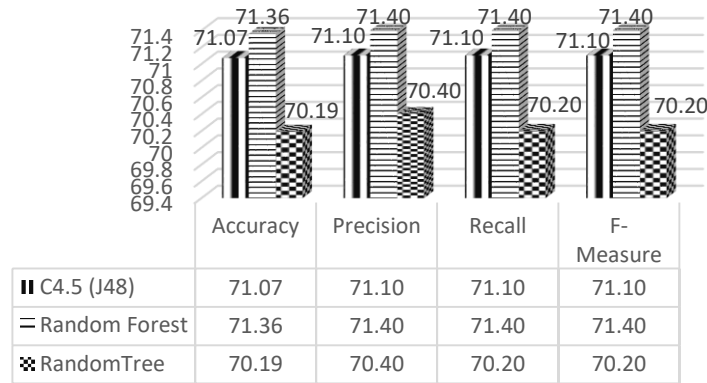
=== Confusion Matrix ===
  a  b  <-- classified as
375 129 |  a = Vec
147 379 |  b = Bec

```

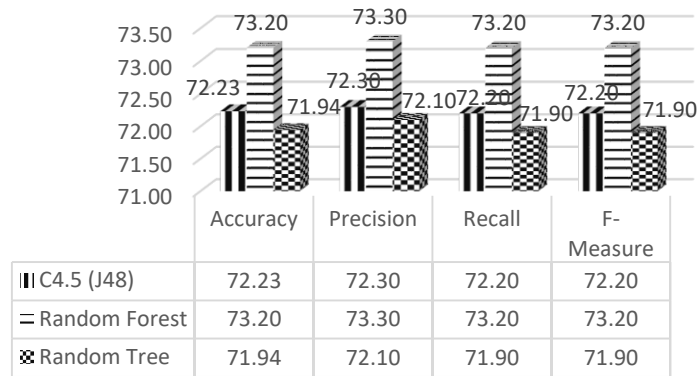
ภาพที่ 6 ผลลัพธ์แบบวิธี InfoGainAttributeEval ของตัวแบบพยากรณ์แบบ RandomForest

ตารางที่ 3 แสดงผลการเปรียบเทียบ

เลือกคุณลักษณะ	เทคนิค	Accuracy	Precision	Recall	F-Measure
แบบไม่เลือกคุณลักษณะ (Non-Selection)	C4.5 (J48)	71.07	71.10	71.10	71.10
	Random Forest	71.36	71.40	71.40	71.40
	Random Tree	70.19	70.40	70.20	70.20
แบบวิธี InfoGainAttributeEval	C4.5 (J48)	72.23	72.30	72.20	72.20
	Random Forest	73.20	73.30	73.20	73.20
	Random Tree	71.94	72.10	71.90	71.90



ภาพที่ 7 กราฟแสดงผลลัพธ์แบบไม่เลือกคุณลักษณะ (Non-Selection)



ภาพที่ 8 กราฟแสดงผลลัพธ์การเลือกคุณลักษณะแบบ InfoGainAttributeEval

จะเห็นว่าการเลือกคุณลักษณะแบบวิธี InfoGainAttributeEval โดยอัลกอริธึม Random Forest ให้ประสิทธิภาพในการพยากรณ์ถูกต้องมากกว่า มีค่า Accuracy 73.20 ค่า Precision 73.30 ค่า Recall 73.20 และค่า F-Measure 73.20

Dataset	(1) trees.J48	(2) trees.RandomForest	(3) trees.RandomTree
data_set (100)	70.83	71.49	70.30
	(✓/ /*)	(0/1/0)	(0/1/0)

ภาพที่ 9 แสดงผลทดสอบแบบประสิทธิภาพทั้ง อัลกอริธึม 3

ขั้นตอนที่ 6 การนำไปใช้งาน เมื่อได้ตัวแบบพยากรณ์ด้วยวิธีการเลือกคุณลักษณะแบบวิธี InfoGainAttributeEval จากขั้นตอนการทำเหมืองข้อมูลด้วยโปรแกรม WEKA ที่มีประสิทธิภาพในการพยากรณ์ถูกต้องมากที่สุด ผู้วิจัยได้พัฒนาระบบพยากรณ์เพื่อสนับสนุนการตัดสินใจศึกษาต่อหลังสำเร็จมัธยมศึกษาตอนต้น โดยใช้อัลกอริธึมป่าสุ่ม ในสังกัดสำนักงานเขตพื้นที่การศึกษาประถมศึกษา นครสวรรค์ เขต 2 ด้วยโปรแกรม Dreamweaver CS6 ร่วมกับภาษา PHP และโปรแกรม Appserv จำลอง web server ในการพัฒนาเว็บไซต์เพื่อให้ผู้เรียนสามารถใช้งานผ่านเว็บเบราว์เซอร์ ในการพยากรณ์แนวทางในการศึกษาต่อระดับสายสามัญ (Basic Education Commission : Bec) และสายอาชีวศึกษา (Vocational Education Commission : Vec)

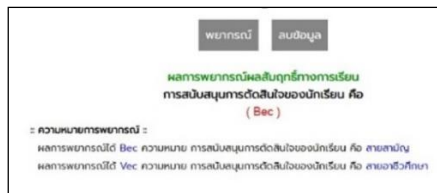
การประชุมสัมมนาวิชาการนำเสนอผลงานวิจัยระดับชาติเครือข่ายบัณฑิตศึกษา มหาวิทยาลัยราชภัฏภาคเหนือ ครั้งที่ 22 "ชีวิตวิถีใหม่ของการศึกษาในสังคมดิจิทัล"



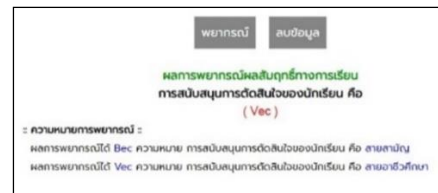
พยากรณ์ผลสัมฤทธิ์ทางการเรียน		
สำหรับ: วิทยาลัยนานาชาติสวนสุนันทา		
ผลสัมฤทธิ์ทางการเรียน 6 ภาคเรียนต่อเนื่อง :		
<input type="radio"/> เกรด 100 - 200	<input checked="" type="radio"/> เกรดเฉลี่ยวิทยาศาสตร์ :	<input type="radio"/> เกรด 301 - 400
	<input checked="" type="radio"/> เกรด 2.01 - 3.00	
<input type="radio"/> เกรด 100 - 200	<input checked="" type="radio"/> เกรดเฉลี่ยวิทยาศาสตร์ :	<input type="radio"/> เกรด 301 - 400
	<input checked="" type="radio"/> เกรด 2.01 - 3.00	
<input type="radio"/> เกรด 100 - 200	<input checked="" type="radio"/> เกรดเฉลี่ยวิทยาศาสตร์ :	<input type="radio"/> เกรด 301 - 400
	<input checked="" type="radio"/> เกรด 2.01 - 3.00	
<input type="radio"/> เกรด 100 - 200	<input checked="" type="radio"/> เกรดเฉลี่ยวิทยาศาสตร์ต่างประเทศ :	<input type="radio"/> เกรด 301 - 400
	<input checked="" type="radio"/> เกรด 2.01 - 3.00	
<input type="radio"/> เกรด 100 - 200	<input checked="" type="radio"/> เกรดเฉลี่ยวิทยาศาสตร์ต่างประเทศ :	<input type="radio"/> เกรด 301 - 400
	<input checked="" type="radio"/> เกรด 2.01 - 3.00	
<input type="radio"/> เกรด 100 - 200	<input type="radio"/> เกรดเฉลี่ยสังคมศึกษา :	<input checked="" type="radio"/> เกรด 301 - 400
	<input type="radio"/> เกรด 2.01 - 3.00	
<input type="radio"/> เกรด 100 - 200	<input checked="" type="radio"/> เกรดเฉลี่ยวิชาการงานอาชีพและเทคโนโลยี :	<input type="radio"/> เกรด 301 - 400
	<input checked="" type="radio"/> เกรด 2.01 - 3.00	
<input type="radio"/> เกรด 100 - 150	<input type="radio"/> เกรดเฉลี่ยทั้งหมด (GPA) :	<input type="radio"/> เกรด 201 - 250
<input checked="" type="radio"/> เกรด 251 - 300	<input type="radio"/> เกรด 151 - 2.00	<input type="radio"/> เกรด 351 - 400
	<input type="radio"/> เกรด 301 - 3.50	

ภาพที่ 10 แสดงตัวอย่างหน้าพยากรณ์ผลสัมฤทธิ์ของระบบ

การใช้งานระบบพยากรณ์ผู้ใช้งานกรอกข้อมูลเกรดเฉลี่ย จำนวน 7 วิชา ซึ่งประกอบไปด้วย 1) เกรดเฉลี่ยกลุ่มสาระภาษาไทย 2) เกรดเฉลี่ยกลุ่มสาระคณิตศาสตร์ 3) เกรดเฉลี่ยกลุ่มสาระวิทยาศาสตร์ 4) เกรดเฉลี่ยกลุ่มสาระภาษาต่างประเทศ 5) เกรดเฉลี่ยกลุ่มสาระสังคมศึกษา 6) เกรดเฉลี่ยกลุ่มสาระการงานอาชีพและเทคโนโลยี และ 7) ผลการเรียนเฉลี่ยตลอดหลักสูตร เมื่อผู้ใช้งานกรอกรายละเอียดข้อมูลผลการเรียนตามหัวข้อต่างๆ แล้วทำการคลิกปุ่มพยากรณ์ ระบบก็จะทำการประมวลผลข้อมูลเพื่อแสดงผลแนวทางในการศึกษาต่อของนักเรียนอยู่ในระดับสายสามัญ หรือสายอาชีวศึกษา



สายสามัญ



สายอาชีวศึกษา

ภาพที่ 11 แสดงผลการพยากรณ์การสนับสนุนการตัดสินใจ

ในส่วนผู้ดูแลระบบ จะประกอบไปด้วยการเพิ่มข้อมูลการเทรน (train) ของระบบพยากรณ์ ซึ่งในส่วนเมนูหน้านี้ผู้ดูแลระบบสามารถเพิ่มข้อมูลที่เป็นปัจจุบันของผลสัมฤทธิ์ทางการเรียนของนักเรียนในปีการศึกษาปัจจุบัน เพื่อให้ระบบพยากรณ์เพื่อสนับสนุนการตัดสินใจศึกษาต่อหลังสำเร็จมัธยมศึกษาตอนต้น โดยใช้อัลกอริธึมปาล์มมีความเป็นปัจจุบันและถูกต้องแม่นยำ



ภาพที่ 12 แสดงระบบการจัดการข้อมูลผู้ดูแลระบบ

อภิปรายผลการวิจัย

จากการศึกษาวิจัยครั้งนี้ได้นำการคัดเลือกคุณลักษณะแบบวิธี InfoGainAttributeEval และวิธีแบบไม่คัดเลือกคุณลักษณะร่วมกับอัลกอริทึม C4.5(J48) อัลกอริทึม Random Tree และ อัลกอริทึม Random Forest ของเทคนิคต้นไม้ตัดสินใจ เพื่อหาค่าความถูกต้องจากการพยากรณ์ผลสัมฤทธิ์ทางการเรียนซึ่งผลการวิจัยพบว่า การคัดเลือกคุณลักษณะวิธี InfoGainAttributeEval กับตัวแบบพยากรณ์ที่สร้างขึ้นด้วยอัลกอริทึม Random Forest มีค่าความถูกต้อง (Accuracy) ร้อยละ 73.20 สูงกว่าอัลกอริทึมอื่น โดยใช้เทคนิคเหมืองข้อมูลในการสร้างตัวแบบพยากรณ์เช่นเดียวกับ สำราญ วานนท์ (2561) ที่ศึกษาเทคนิคพยากรณ์อาชีพสำหรับนักศึกษาระดับปริญญาตรีสาขา คอมพิวเตอร์โดยใช้เทคนิคเหมืองข้อมูล ซึ่งพบว่าเทคนิคการจำแนกข้อมูลด้วยวิธี Random Forest ให้ความถูกต้อง ในการจำแนกประเภทข้อมูลสูงสุด ผู้วิจัยเลือกใช้อัลกอริทึม Random Forest ไปทดสอบกับชุดข้อมูลจำนวน 1,030 เรคคอร์ด ซึ่งสอดคล้องกับงานวิจัยของ กฤษณ์ บรรณะชัยศิริสุข (2558) ได้ทำการศึกษาเครื่องมือท่องเว็บไซต์บน อุปกรณ์เคลื่อนที่สำหรับผู้พิการทางสายตาโดยใช้การตรวจหาเนื้อหาหลัก ซึ่งผลของการวิจัยพบว่า การจำแนกแบบ ป่าสุ่ม ร่วมกับการทดสอบแบบ 10-fold cross validation มีผลลัพธ์ความถูกต้องมากที่สุด โดยการคัดเลือก คุณลักษณะด้วยวิธี InfoGainAttributeEval เช่นเดียวกับ วันวิสาข์ ชนะประเสริฐ (2559) ได้ทำการประยุกต์ใช้ เทคนิคเหมืองข้อมูลเพื่อแนะนำอาชีพสำหรับนักศึกษาระดับปริญญาตรี คณะโบราณคดี มหาวิทยาลัยศิลปากร โดยการ คัดเลือกคุณลักษณะด้วยวิธี InfoGainAttributeEval ให้ความถูกต้องมากที่สุด มาพัฒนาเป็นระบบพยากรณ์เพื่อ สนับสนุนการตัดสินใจศึกษาต่อหลังสำเร็จมัธยมศึกษาตอนต้น โดยใช้อัลกอริทึมป่าสุ่ม

ระบบพยากรณ์เพื่อสนับสนุนการตัดสินใจศึกษาต่อหลังสำเร็จมัธยมศึกษาตอนต้น สามารถนำไปใช้ ประกอบการแนะนำแนวทางในการวางแผนศึกษาต่อของนักเรียนชั้นมัธยมศึกษาปีที่ 3 ได้จริง ระบบพยากรณ์นี้อาจ มีข้อจำกัดหากมีการปรับเปลี่ยนหลักสูตรของรายวิชาในอนาคตทำให้ข้อมูลการเทรนของระบบพยากรณ์เกิดความ คลาดเคลื่อน ผู้จัดทำต้องพัฒนา ปรับปรุง แก้ไข เพื่อสามารถใช้งานได้ดียิ่งขึ้น

ข้อเสนอแนะ

1. ข้อเสนอแนะทั่วไป

ครูแนะแนวหรือครูที่ปรึกษานักเรียนระดับชั้นมัธยมศึกษาปีที่ 3 สามารถนำเอาระบบพยากรณ์เพื่อ สนับสนุนการตัดสินใจศึกษาต่อหลังสำเร็จมัธยมศึกษาตอนต้น โดยใช้อัลกอริทึมป่าสุ่ม ไปใช้ประกอบการแนะนำ แนวทางในการศึกษาต่อของนักเรียนได้

2. ข้อเสนอแนะในการวิจัยครั้งต่อไป

2.1 ในการศึกษาวิจัยครั้งต่อไปควรนำไปพัฒนาเป็นโมบายแอปพลิเคชันเพื่อให้นักเรียนเข้าถึงระบบ พยากรณ์ได้สะดวกยิ่งขึ้น



2.2 ควรมีการนำเอาเทคนิคเหมือนข้อมูลหรืออัลกอริธึมอื่นๆ เช่น Naive Bayes, k-Nearest Neighbo, Support Vector Machines มาเปรียบเทียบในระบบพยากรณ์ผลสัมฤทธิ์ทางการเรียนได้

2.3 ในการศึกษาวิจัยครั้งต่อไปควรมีเอากิจกรรมพัฒนาผู้เรียนมาใช้ในการวิเคราะห์ผลการเรียนของนักเรียน

กิตติกรรมประกาศ

งานวิจัยฉบับนี้สำเร็จลุล่วงไปได้ด้วยความกรุณาและความอนุเคราะห์จากรองศาสตราจารย์ อนุรักษ์ เห็นเจริญเลิศ อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก รองศาสตราจารย์ ดร.วฤชาญ ร่มสายหยุด อาจารย์ที่ปรึกษาร่วม ซึ่งได้เสียสละเวลาและให้คำแนะนำปรึกษา ให้ข้อเสนอแนะต่างๆ ตลอดจนแนะนำแนวทางในการศึกษาค้นคว้า งานวิจัยและข้อมูลต่างๆ ที่นำมาปรับปรุงแก้ไข จนทำให้งานวิจัยสำเร็จลุล่วง

เอกสารอ้างอิง

- กฤษฎี บรณณะชัยศิริสุข. (2558). เครื่องมือทอ้งเว็บไซต์บนอุปกรณ์เคลื่อนที่สำหรับผู้พิการทางสายตาโดยใช้การตรวจหาเนื้อหาลูก. วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต (สาขาวิทยาศาสตร์คอมพิวเตอร์). จุฬาลงกรณ์มหาวิทยาลัย, กรุงเทพฯ.
- ฉิยทธ์ คุณสุข, และจารี ทองคำ. (2562). กระบวนการคัดเลือกคุณลักษณะสำหรับเพิ่มประสิทธิภาพในการจำแนกความคิดเห็นของลูกค้า. *วารสาร มทร อีสาน ฉบับวิทยาศาสตร์และเทคโนโลยี*, 13(1), 129-143.
- ปรเมษฐ์ ธีนวนานท์และคณะ. (2560). การประยุกต์ใช้โมเดลการเรียนรู้แบบรวมกลุ่มเพื่อพยากรณ์แนวโน้มของราคาหลักทรัพย์ในตลาดหลักทรัพย์แห่งประเทศไทย. *วารสารวิทยาการและเทคโนโลยีสารสนเทศ*, 7(1), 12-21.
- วัชรวิวัฒน์ จิตต์สกุล. (2560). การวิเคราะห์การจำแนกข้อความด้วยการเปรียบเทียบความเสถียรของอัลกอริทึม. *วารสารวิจัยและพัฒนามหาวิทยาลัยราชภัฏบ้านสมเด็จเจ้าพระยา*, (9), 19-31.
- วันวิสาข์ ชนะประเสริฐ. (2559). การประยุกต์ใช้เทคนิคเหมือนข้อมูลเพื่อแนะนำอาชีพสำหรับนักศึกษาปริญญาตรี คณะโบราณคดี มหาวิทยาลัยศิลปากร. วิทยานิพนธ์ศิลปศาสตรมหาบัณฑิต (สาขาวิชาสารสนเทศศาสตร์ เพื่อการศึกษา). มหาวิทยาลัยศิลปากร, กรุงเทพฯ.
- สำราญ วานนท์. (2561). การศึกษาเทคนิคพยากรณ์อาชีพสำหรับนักศึกษาระดับปริญญาตรีสาขาคอมพิวเตอร์โดยใช้เทคนิคเหมือนข้อมูล. *วารสารวิชาการการจัดการเทคโนโลยีสารสนเทศและนวัตกรรม*, 5(1), 164-171.
- อรนุช พันโท. (2557). การเปรียบเทียบประสิทธิภาพการจำแนกรูปแบบการเรียนรู้ VARK ด้วยเทคนิคเหมือนข้อมูล. *วารสารเทคโนโลยีอุตสาหกรรม มหาวิทยาลัยราชภัฏอุบลราชธานี*, 4(1), 1-11.
- เอกสิทธิ์ พังรวงศ์ศักดิ์. (2561). กระบวนการวิเคราะห์ข้อมูลด้วย CRISP-DM และตัวอย่างการประยุกต์ใช้ทางด้านการศึกษา. สืบค้นจาก <https://datacube.ai/crisp-dm/>
- Glassboxmedicine. (2019). *Measuring Performance: The Confusion Matrix*. Retrieved from <https://glassboxmedicine.com/2019/02/17/measuring-performance-the-confusion-matrix/>
- Mishra, A. K., & Ratha, B., K. (2016). RathaStudy of Random Tree and Random Forest Data Mining Algorithms for Microarray Data Analysis. *International Journal on Advanced Electrical and Computer Engineering*, 3(4), 5-7.
- Xu, J., & Jiang, H. (2015). An Improved Information Gain Feature Selection Algorithm for SVM TextClassifier. *2015 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery*, 273-276.