ระบบพยากรณ์เพื่อสนับสนุนการตัดสินใจศึกษาต่อ หลังสำเร็จมัธยมศึกษาตอนต้น โดยใช้อัลกอริธึมป่าสุ่ม

นายวรเทพ วันกาล

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
แขนงวิชาเทคโนโลยีสารสนเทศและการสื่อสาร สาขาวิชาวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยสุโขทัยธรรมาธิราช
พ.ศ. 2564

A Predictive System to Support Decision-making on Further Education After Junior High School Using the Random Forest Algorithm

Mr. Worathape Wankan

A Thesis Submitted in Partial Fulfillment of the Requirements for

the Degree of Master of Information and Communication Technology

School of Science and Technology

Sukhothai Thammathirat Open University

2021

หัวข้อวิทยานิพนธ์ ระบบพยากรณ์เพื่อสนับสนุนการตัดสินใจศึกษาต่อหลังสำเร็จมัธยมศึกษาตอนต้น โดยใช้อัลกอริธึมป่าสุ่ม

ชื่อและนามสกุล นายวรเทพ วันกาล

แขนงวิชา เทคโนโลยีสารสนเทศและการสื่อสาร

สาขาวิชา วิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยสุโขทัยธรรมาธิราช

อาจารย์ที่ปรึกษา 1. รองศาสตราจารย์ ณัฐพร เห็นเจริญเลิศ

2. รองศาสตราจารย์ ดร.วฤษาย์ ร่มสายหยุด

วิทยานิพนธ์นี้ ได้รับความเห็นชอบให้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรระดับปริญญาโท เมื่อวันที่ 31 พฤษภาคม 2565

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ

(ผู้ช่วยศาสตราจารย์ ดร.วราภรณ์ วิยานนท์)

..... กรรมการ

(รองศาสตราจารย์ ณัฐพร เห็นเจริญเลิศ)

..... กรรมการ

(รองศาสตราจารย์ ดร.วฤษาย์ ร่มสายหยุด)

..... ประธานกรรมการบัณฑิตศึกษา

(รองศาสตราจารย์ ดร.นราธิป ศรีราม)

ชื่อวิทยานิพนธ์ ระบบพยากรณ์เพื่อสนับสนุนการตัดสินใจศึกษาต่อหลังสำเร็จมัธยมศึกษาตอนต้น โดยใช้อัลกอริธึมป่าสุ่ม

ผู้วิจัย นายวรเทพ วันกาล **รหัสนักศึกษา** 2599600471 ปริญญา วิทยาศาสตรมหาบัณฑิต (เทคโนโลยีสารสนเทศและการสื่อสาร) **อาจารย์ที่ปรึกษา** (1) รองศาสตราจารย์ ณัฐพร เห็นเจริญเลิศ (2) รองศาสตราจารย์ ดร.วฤษาย์ ร่มสายหยุด **ปีการศึกษา** 2564

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 และใช้การเลือกคุณลักษะแบบ InfoGainAttributeEval ทดสอบประสิทธิภาพด้วยวิธี 10-fold cross-validation โดยใช้โปรแกรม Weka แล้วนำผลการทดสอบที่ได้มาพัฒนาระบบพยากรณ์เพื่อสนับสนุนการตัดสินใจศึกษาต่อหลังสำเร็จมัธยมศึกษาตอนต้นใช้งานผ่านเว็บเบราว์เซอร์

ผลการวิจัยนี้พบว่าอัลกอริธึมป่าสุ่มมีค่าประสิทธิภาพมากที่สุดค่าความถูกต้องร้อยละ 73.20 ค่าความแม่นยำเท่ากับร้อยละ 73.0 ค่าความครบถ้วนเท่ากับร้อยละ 73.0 ค่าประสิทธิภาพโดยรวมเท่ากับร้อยละ 73.0

คำสำคัญ ระบบพยากรณ์ สนับสนุนการตัดสินใจ อัลกอริธึมป่าสุ่ม กระบวนการคัดเลือกคุณลักษณะ

Thesis title: A Predictive System to Support Decision-making on Further Education After Junior High School Using the Random Forest Algorithm
Researcher: Worathape Wankan; **ID:** 2599600471; **Degree:** Master of Science (Information and Communication Technology); **Thesis advisors:** (1) Nuttaporn Hencharoenlert, Associate Professor; (2) Dr.Walisa Romsaiyud, Associate Professor; **Academic year:** 2021

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 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 3rd 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 for efficiency, testing using a 10-fold cross-validation model with Weka Program. Afterward, 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 Process

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้ สำเร็จลุล่วงไปได้ด้วยความกรุณาและความอนุเคราะห์จากรองศาสตราจารย์ ณัฐพร เห็นเจริญเลิศ อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก รองศาสตราจารย์ ดร.วฤษาย์ ร่มสายหยุด อาจารย์ที่ปรึกษาร่วม ซึ่งได้เสียสละเวลาและให้คำแนะนำปรึกษา ให้ข้อเสนอแนะต่างๆ ในการจัดทำวิทยานิพนธ์ตลอดจนแนะนำแนวทางในการศึกษาค้นคว้างานวิจัยและข้อมูลต่างๆ ที่นำมาปรับปรุงแก้ไข จนทำให้วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วง ผู้วิจัยรู้สึกซาบซึ้งในความกรุณาของอาจารย์และกราบขอบพระคุณเป็นอย่างสูงมา ณ โอกาสนี้

ขอขอบคุณสำนักงานเขตพื้นที่การศึกษาประถมศึกษานครสวรรค์เขต 2 ที่ให้ความอนุเคราะห์ข้อมูลในการวิเคราะห์ในการทำวิทยานิพนธ์ รวมทั้งบุคลากร เจ้าหน้าที่และเพื่อนร่วมงานทุกท่านจากโรงเรียนบ้านปางสุดและโรงเรียนนารีรัตน์จังหวัดแพร่ที่ให้ความช่วยเหลือแนะนำ สนับสนุน และคอยให้คำปรึกษาตลอดระยะเวลาการศึกษาในครั้งนี้

สุดท้ายนี้ ขอกราบขอบพระคุณบิดา มารดา ที่ได้มอบโอกาสทางการศึกษา และเป็นกำลังใจต่อข้าพเจ้าตลอดมา ขอขอบพระคุณคณาจารย์ประจำสาขาวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยสุโขทัยธรรมาธิราช เพื่อนนักศึกษาตลอดจนผู้มีส่วนร่วมทุกท่านที่ทำให้วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปไกด้ด้วยดี ผู้วิจัยหวังว่าวิทยานิพนธ์ฉบับนี้คงเป็นประโยชน์ไม่มากก็น้อยสำหรับผู้ที่สนใจจะศึกษาต่อไป หากผิดพลาดประการใด ผู้วิจัยขออภัยเป็นอย่างสูงมา ณ ที่นี้

นายวรเทพ วันกาล

พฤษภาคม 2565

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ง
บทคัดย่อภาษาอังกฤษ ..... จ
กิตติกรรมประกาศ ..... ฉ
สารบัญตาราง ..... ฌ
สารบัญภาพ ..... ญ
บทที่ 1 บทนำ 1
ความเป็นมาและความสำคัญของปัญหา 1
วัตถุประสงค์การวิจัย 3
ขอบเขตของงานวิจัย 3
นิยามศัพท์เฉพาะ 3
ประโยชน์ที่คาดว่าจะได้รับ 4
บทที่ 2 วรรณกรรมที่เกี่ยวข้อง 5
การศึกษาขั้นพื้นฐาน 5
การพยากรณ์ 8
เทคนิคเหมืองข้อมูล 9
เทคนิคการจำแนกกลุ่มข้อมูล 11
ซอฟต์แวร์ 18
งานวิจัยที่เกี่ยวข้อง 19
บทที่ 3 วิธีดำเนินการวิจัย 24
ชุดข้อมูล 24
เครื่องมือที่ใช้ในการวิจัย 25
ขั้นตอนการดำเนินการวิจัย 25
บทที่ 4 ผลการดำเนินงานวิจัย 29
ทำความเข้าใจกับโจทย์ที่ต้องทำ 29
ทำความเข้าใจกับข้อมูลที่ใช้ในการวิจัย 30
การเตรียมข้อมูล 32
การสร้างตัวแบบพยากรณ์ 37

สารบัญ (ต่อ)

หน้า

การประเมินผล 46
การนำไปใช้งาน 50
บทที่ 5 สรุปการวิจัย อภิปรายผล และข้อเสนอแนะ 59
สรุปการวิจัย 59
อภิปรายผล 60
ข้อเสนอแนะ 61
บรรณานุกรม 62
ภาคผนวก 67
ก การติดตั้งโปรแกรมจำลองเว็บเซิฟเวอร์ 68
ข ชุดข้อมูล (data set) 73
ค คู่มือการใช้งานระบบพยากรณ์เพื่อสนับสนุนการตัดสินใจ 76
ประวัติผู้วิจัย 81

สารบัญตาราง

หน้า

ตารางที่ 4.1 การแปลงข้อมูล 35
ตารางที่ 4.2 แสดงคุณลักษณะที่จะนำไปวิเคราะห์ 41
ตารางที่ 4.3 แสดงผลการเปรียบเทียบ 47
ตารางที่ 4.4 Data Dictionary ของตาราง tbl_data 51
ตารางที่ 4.5 Data Dictionary ของตาราง tbl_admin 51

สารบัญภาพ

หน้า

ภาพที่ 2.1 เส้นทางการศึกษาต่อประเภทสามัญศึกษา 6
ภาพที่ 2.2 เส้นทางการศึกษาต่อประเภทประเภทอาชีวศึกษา 7
ภาพที่ 2.3 กระบวนการการพยากรณ์ 8
ภาพที่ 2.4 ขั้นตอนกระบวนการ CRISP-DM 9
ภาพที่ 2.5 การแบ่งข้อมูลแบบ K-fold-10-fold cross validation 10
ภาพที่ 2.6 ตาราง confusion matrix 11
ภาพที่ 2.7 การทำงานพื้นฐานของการจำแนกข้อมูล 12
ภาพที่ 2.8 ตัวอย่างต้นไม้ตัดสินใจสำหรับการจำแนกคุณลักษณะของลูกค้าที่ทาการซื้อคอมพิวเตอร์ 13
ภาพที่ 2.9 โครงสร้างการจำแนกแบบต้นไม้สุ่ม (random tree) 14
ภาพที่ 2.10 โครงสร้างการจำแนกแบบป่าสุ่ม (random forest) 15
ภาพที่ 2.11 การเปรียบเทียบแนวความคิดของการคัดเลือกคุณสมบัติ (a) วิธีฟิลเตอร์ (b) วิธีแรปเปอร์ และ (c) วิธีฝังตัว 16
ภาพที่ 2.12 หน้าจอหลักโปรแกรม Weka 18
ภาพที่ 2.13 หน้าส่วนประกอบเมนู Explorer โปรแกรม Weka 19
ภาพที่ 3.1 แสดงขั้นตอนการการดำเนินการวิจัย 26
ภาพที่ 3.2 ขั้นตอกนการเตรียมข้อมูล 27
ภาพที่ 4.1 กระบวนการตรวจสอบ สำรวจข้อมูลเบื้องต้น Exploratory Data Analysis 30
ภาพที่ 4.2 แสดงข้อมูลที่นำมาใช้ในการศึกษา 31
ภาพที่ 4.3 การตรวจสอบความถูกต้องของข้อมูล 32
ภาพที่ 4.4 ตัดข้อมูลที่มีความผิดปกติและข้อมูลขาดหายไป 33
ภาพที่ 4.5 การลดจำนวนของคุณลักษณะที่ไม่เกี่ยวข้องกับข้อมูลนักเรียนและผลสัมฤทธิ์ 34
ภาพที่ 4.6 การคำนวณหาค่าผลสัมฤทธิ์แต่ละกลุ่มสาระการเรียนรู้ 35
ภาพที่ 4.7 การแปลงข้อมูลให้อยู่ในรูปแบบที่พร้อมนำไปวิเคราะห์ 37
ภาพที่ 4.8 ขั้นตอนการคัดเลือกคุณลักษณะหรือฟีเจอร์ 38
ภาพที่ 4.9 รูปแบบการสร้างตัวแบบพยากรณ์ 39
ภาพที่ 4.10 ไม่เลือกคุณลักษณะของชุดข้อมูล 39

สารบัญภาพ (ต่อ)

หน้า

ภาพที่ 4.11 การคัดเลือกคุณลักษณะแบบวิธี InfoGainAttributeEval 40
ภาพที่ 4.12 แสดงการวิเคราะห์ประสิทธิภาพด้วย KnowledgeFlow 42
ภาพที่ 4.13 ผลลัพธ์แบบไม่เลือกคุณลักษณะ (Non-Selection) ของโมเดลแบบ C4.5 (J48) 43
ภาพที่ 4.14 ผลลัพธ์แบบไม่เลือกคุณลักษณะ (Non-Selection) ของโมเดลแบบ RandomForest 43
ภาพที่ 4.15 ผลลัพธ์แบบไม่เลือกคุณลักษณะ (Non-Selection) ของโมเดลแบบ RandomTree 44
ภาพที่ 4.16 ผลลัพธ์แบบวิธี InfoGainAttributeEval ของโมเดลแบบ ของโมเดลแบบ C4.5 (J48) 44
ภาพที่ 4.17 ผลลัพธ์แบบวิธี InfoGainAttributeEval ของโมเดลแบบ RandomForest 45
ภาพที่ 4.18 ผลลัพธ์แบบวิธี InfoGainAttributeEval ของโมเดลแบบ RandomTree 45
ภาพที่ 4.19 กราฟแสดงผลลัพธ์แบบไม่เลือกคุณลักษณะ (Non-Selection) 47
ภาพที่ 4.20 กราฟแสดงผลลัพธ์การเลือกคุณลักษณะแบบ InfoGainAttributeEval 48
ภาพที่ 4.21 แสดงผลทดสอบแบบจำลองทั้ง 3 อัลกอริธึม 48
ภาพที่ 4.22 แสดงผลทดสอบแบบประสิทธิภาพทั้ง 3 อัลกอริธึม 49
ภาพที่ 4.23 แสดงผลทดสอบโมเดลในการพยากรณ์ 50
ภาพที่ 4.24 ฐานข้อมูล tbl_data 52
ภาพที่ 4.25 ฐานข้อมูล tbl_admin 52
ภาพที่ 4.26 การนำเข้าชุดข้อมูลในฐานข้อมูล 53
ภาพที่ 4.27 library weka.jar และไฟล์โมเดล 53
ภาพที่ 4.28 การแปลงไฟล์จาก .csv เป็น .arff ของตัวระบบ 54
ภาพที่ 4.29 แสดงตัวอย่างหน้าแรกของระบบ 55
ภาพที่ 4.30 แสดงตัวอย่างหน้าพยากรณ์ผลสัมฤทธิ์ของระบบ 55
ภาพที่ 4.31 แสดงการกรอกรายละเอียดข้อมูลผลการเรียน 56
ภาพที่ 4.32 แสดงผลการพยากรณ์การสนับสนุนการตัดสินใจในสายสามัญ 56
ภาพที่ 4.33 แสดงผลการพยากรณ์การสนับสนุนการตัดสินใจในสายอาชีวศึกษา 57
ภาพที่ 4.34 แสดงตัวอย่างผู้พัฒนาระบบ 57
ภาพที่ 4.35 แสดงหน้าผู้ดูแลระบบ 58
ภาพที่ 4.36 แสดงระบบการจัดการข้อมูลผู้ดูแลระบบ 58

บทที่ 1
บทนำ

1. ความเป็นมาและความสำคัญของปัญหา

การศึกษามีความสำคัญต่อการพัฒนาคนในสังคมให้มีความรู้ เป็นการวางรากฐานของความเจริญก้าวหน้าของประเทศ การจัดการ การศึกษาในประเทศไทยมีระบบการศึกษาขั้นพื้นฐานซึ่งจัดไว้ไม่น้อยกว่าสิบสองปี ก่อนระดับอุดมศึกษา ปัจจุบันการศึกษาในระบบได้แบ่งออกเป็น 2 ระดับ คือ 1) การศึกษาขั้นพื้นฐานซึ่งหมายถึงการศึกษาก่อนระดับอุดมศึกษาแบ่งการศึกษาออกเป็น 3 ระดับ คือ 1.1) ระดับก่อนประถมศึกษา เป็นการจัดการศึกษาให้กับเด็กที่มีอายุ 3 – 6 ปี 1.2) ระดับประถมศึกษา เป็นการจัดการศึกษาระดับประถมศึกษาปีที่ 1 ถึง ประถมศึกษาปีที่ 6 1.3) ระดับมัธยมศึกษาซึ่งแบ่งออกเป็น 2 ระดับมัธยมศึกษาตอนต้นและมัธยมศึกษาตอนปลาย มัธยมศึกษาตอนต้นใช้เวลาเรียน 3 ปี มัธยมศึกษาตอนปลายใช้เวลาเรียน 3 ปี มัธยมศึกษาตอนปลายแบ่งเป็น 2 ประเภท คือ ประเภทสามัญศึกษา เป็นการศึกษาระดับชั้นมัธยมศึกษาปีที่ 4 – 6 และประเภทอาชีวศึกษา เป็นการจัดการศึกษาเพื่อพัฒนาความรู้และทักษะในการประกอบอาชีพ หรือหลักสูตรประกาศนียบัตรวิชาชีพ (ปวช.) 2) การศึกษาระดับอุดมศึกษาแบ่งออกเป็น 2 ระดับ คือ ระดับต่ำกว่าปริญญาหรือหลักสูตรประกาศนียบัตรวิชาชั้นสูง (ปวส.) และระดับปริญญา ในการศึกษาขั้นพื้นฐานเมื่อผู้เรียนสำเร็จการศึกษาจากระดับมัธยมศึกษาปีที่ 3 แล้วตัดสินใจเลือกที่จะศึกษาต่อในระดับมัธยมศึกษาตอนปลาย ในการศึกษาต่อแต่ละประเภทการศึกษาอาจดูได้จากระดับคะแนนหรือผลสัมฤทธิ์ทางการเรียนที่ผ่านมาของผู้เรียนแต่ละคน ว่าในแต่ละวิชาของผู้เรียนนั้นได้ระดับคะแนนผลการเรียนอยู่ในระดับใด ถ้าได้ระดับผลการเรียนในวิชานั้นสูง แสดงว่าผู้เรียนมีความรู้ ความสนใจ และความเข้าใจในการเรียนวิชานั้นมาก แต่ถ้าระดับคะแนนผลการเรียนในวิชานั้นต่ำ แสดงว่าผู้เรียนมีความรู้ ความสนใจ และความเข้าใจในการเรียนวิชานั้นน้อย ในส่วนการศึกษาต่อประเภทอาชีวศึกษา ซึ่งเป็นอีกทางเลือกสำหรับผู้เรียนที่สำเร็จการศึกษาระดับชั้นมัธยมศึกษาปีที่ 3 ซึ่งมุ่งศึกษาและฝึกฝนวิชาชีพสาขาใดสาขาหนึ่งโดยเฉพาะ เพื่อจบออกมาประกอบอาชีพในสาขาต่าง ๆ หลักสูตรที่เปิดสอนสายอาชีพ มี 3 ระดับ คือ 1) หลักสูตรระดับประกาศนียบัตรวิชาชีพ (ปวช.) ใช้เวลาศึกษาต่อจากชั้นมัธยมศึกษาปีที่ 3 เป็นระยะเวลา 3 ปี 2) หลักสูตรระดับประกาศนียบัตรวิชาชีพชั้นสูง (ปวส.) ใช้เวลาศึกษาต่อจากระดับประกาศนียบัตร

วิชาชีพ (ปวช.) หรือจากระดับชั้นมัธยมศึกษาปีที่ 6 เป็นระยะเวลา 2 ปี และ 3) หลักสูตรระดับวิชาชีพระยะสั้น ใช้เวลาศึกษาตามที่แต่ละสาขาวิชากำหนด

โรงเรียนขยายโอกาสทางการศึกษาในสังกัดสำนักงานเขตพื้นที่การศึกษาประถมศึกษานครสวรรค์เขต 2 ซึ่งส่วนใหญ่จะตั้งอยู่ในพื้นที่ห่างไกลเป็นโรงเรียนที่เปิดสอนตั้งแต่ระดับชั้นอนุบาล ถึง ชั้นมัธยมศึกษาปีที่ 3 และชั้นมัธยมศึกษาปีที่ 6 ซึ่งในแต่ละปี มีนักเรียนที่จบการศึกษาชั้นมัธยมศึกษาปีที่ 3 แล้วไม่มีความชัดเจนในเป้าหมายของการศึกษาต่อ เนื่องจากโรงเรียนขาดบุคลากรครูแนะแนว ซึ่งเป็นปัญหาต้นทางของเรื่องนี้ นักเรียนจำนวนมากขาดการแนะแนวในด้านการศึกษาต่อ จึงเป็นส่วนหนึ่งที่ทำให้นักเรียนประสบปัญหาเรื่องการค้นพบตัวเองช้าว่าอยากเรียนอะไร หรือปรับตัวไม่ทันกับประเภทการเรียนที่เลือกไปแล้ว ทำให้ต้องเสียเวลาไปกับการเรียนที่ไม่มีเป้าหมาย บทบาทครูแนะแนวจึงถือเป็นบทบาทที่สำคัญมาก เพราะครูกลุ่มนี้จะสามารถเข้าถึงปัญหาที่นักเรียนประสบอยู่ในมิติที่ใกล้ชิดกว่าครูผู้สอนวิชาอื่น ๆ นักเรียนที่ได้รับการแนะแนวจากครูแนะแนวมาแล้ว จะสามารถตัดสินใจได้ในระดับหนึ่งว่าสนใจเรียนอะไร ซึ่งในการศึกษาต่อของนักเรียนที่ผ่านมาส่วนใหญ่พบว่าเลือกศึกษาต่อตามเพื่อน หรือตามคำแนะนำของผู้ปกครอง ทำให้นักเรียนกลุ่มนี้ไม่ประสบความสำเร็จในการเรียน หรือมีผลการเรียนต่ำ และปัญหาที่ตามมาคือการเรียนไม่จบหลักสูตร เปลี่ยนสายการศึกษาต่อ และนักเรียนบางคนต้องเลิกศึกษาต่อทั้งที่ยังไม่จบหลักสูตร หนึ่งในสาเหตุของปัญหาที่เกิดขึ้นคือ นักเรียนยังไม่รู้ถึงความสามารถ ความต้องการของตัวเองก่อนที่จะในการศึกษาต่อ และยังขาดความรู้ ความเข้าใจเกี่ยวกับประเภทการศึกษาต่อ อีกทั้งไม่ได้รับการแนะแนวหรือข้อมูลประกอบการวิเคราะห์อย่างเพียงพอเพื่อใช้ในการตัดสินใจ ทำให้นักเรียนเลือกศึกษาต่อไม่ตรงกับความรู้ ความสามารถและความถนัดในด้านการเรียนของตนเอง ทำให้เสียเวลาและสูญเสียทรัพยากรหลาย ๆ อย่าง ดังนั้นในการตัดสินใจเลือกเรียนประเภทการศึกษาต่อนั้น ควรเลือกให้เหมาะสมกับตัวนักเรียน โดยพิจารณาจากเจตคติของนักเรียนที่มีต่อการเรียนประเภทนั้น ๆ เช่น สังคม ความเป็นอยู่ ชุมชน และสภาพแวดล้อมที่เอื้อต่อการศึกษาต่อของนักเรียน การแนะแนวจึงสำคัญเพื่อส่งเสริมให้นักเรียนที่สำเร็จการศึกษาจากชั้นมัธยมศึกษาปีที่ 3 ได้ศึกษาต่อตามความถนัด ความสนใจ และความสามารถ ของตัวเอง เพื่อนำความรู้ไปใช้เป็นพื้นฐานของการศึกษาต่อในระดับที่สูงขึ้น หรือนำไปใช้ในการประกอบอาชีพตลอดจนสามารถนำความรู้ที่ได้จากการศึกษาไปพัฒนาตามศักยภาพของตน เพื่อเป็นทรัพยากรมนุษย์ที่มีคุณภาพในสังคม

จากปัญหาดังกล่าวผู้วิจัยจึงเล็งเห็นความสำคัญของการให้คำปรึกษาแนะแนวการศึกษาต่อให้กับนักเรียน จึงมีแนวความคิดในการนำผลสัมฤทธิ์ทางการเรียนของนักเรียนรายบุคคลในแต่ละภาคเรียนมาจัดทำเหมืองข้อมูล (Data Mining) ซึ่งใช้วิธีการคัดเลือกคุณลักษณะแล้วนำมาเปรียบเทียบสร้างตัวแบบพยากรณ์ด้วย 3 อัลกอริธึม คือ 1) อัลกอริธึม C4.5(J48) 2) อัลกอริธึม

ต้นไม้สุ่ม และ 3) อัลกอริธึมป่าสุ่ม เพื่อวิเคราะห์ประสิทธิภาพให้ได้ตัวแบบพยากรณ์ที่ดีที่สุด สร้างตัวแบบพยากรณ์เพื่อสนับสนุนการตัดสินใจในการศึกษาต่อตามความถนัดของนักเรียนชั้นมัธยมศึกษาปีที่ 3 ในการพัฒนาระบบพยากรณ์ที่มีประสิทธิภาพในการแนะแนวการศึกษาต่อของนักเรียน เพื่อใช้เป็นแนวทางในการตัดสินใจเลือกประเภทในการศึกษาต่อ

2. วัตถุประสงค์การวิจัย

2.1 เพื่อสร้างแบบจำลองพยากรณ์เพื่อสนับสนุนการตัดสินใจศึกษาต่อหลังสำเร็จมัธยมศึกษาตอนต้น โดยใช้อัลกอริธึมป่าสุ่ม

2.2 เพื่อประเมินประสิทธิภาพของแบบจำลองพยากรณ์เพื่อสนับสนุนการตัดสินใจศึกษาต่อหลังสำเร็จมัธยมศึกษาตอนต้น

2.3 เพื่อพัฒนาระบบพยากรณ์เพื่อสนับสนุนการตัดสินใจศึกษาต่อหลังสำเร็จมัธยมศึกษาตอนต้น

3. ขอบเขตของการวิจัย

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

3.1 แหล่งข้อมูลที่ใช้เป็นตัวแปรต้น คือ ชุดข้อมูลผลสัมฤทธิ์ทางการเรียนของนักเรียนชั้นมัธยมศึกษาปีที่ 3 สำนักงานเขตพื้นที่การศึกษาประถมศึกษานครสวรรค์ เขต 2 จำนวน 1,160 คน

3.2 ตัวแปรตาม คือ ผลการพยากรณ์ของนักเรียนแต่ละคนเพื่อสนับสนุนการตัดสินใจในการศึกษาต่อตามความถนัด

4. นิยามศัพท์เฉพาะ

4.1 อัลกอริธึมป่าสุ่ม (Random Forest Algorithm) หมายถึง การสร้างโมเดลด้วยเทคนิคต้นไม้ตัดสินใจ (Decision Tree) หลาย ๆ โมเดล ตั้งแต่ 10 โมเดลขึ้นไป โดยแต่ละโมเดลจะ

มีชุดของข้อมูลที่ไม่เหมือนกันโมเดลแต่ละตัวจะเรียนรู้อย่างเป็นอิสระต่อกัน เอาผลการตัดสินใจของโมเดลที่ทำนายแล้วจากเสียงส่วนมาก (Vote) ว่า Class ไหนที่ถูกเลือกมากที่สุดมาใช้ในการตัดสินใจ

4.2 การพยากรณ์ (Prediction) หมายถึง การทำนายเหตุการณ์ในอนาคต โดยนำข้อมูลในอดีตมาใช้พยากรณ์ผ่านการใช้หลักการทางคณิตศาสตร์

4.3 การศึกษาประเภทสามัญศึกษา หมายถึง การจัดการเรียนการสอนอย่างมีระบบรูปแบบขั้นตอนตรงตามวัตถุประสงค์ตามหลักสูตรการศึกษา ซึ่งประกอบด้วยการจัดการเรียนการสอนตั้งแต่ชั้นก่อนประถมศึกษา ชั้นประถมศึกษา ชั้นมัธยมศึกษาตอนต้น และชั้นมัธยมศึกษาตอนปลาย

4.4 การศึกษาประเภทอาชีวศึกษา หมายถึง สถานศึกษาที่จัดการเรียนการสอนในหลักสูตรการศึกษาที่ได้รับการรับรองจากกระทรวงศึกษาธิการเปิดสอนในระดับชั้นประกาศนียบัตรวิชาชีพ (ปวช.) และ ระดับประกาศนียบัตรวิชาชีพชั้นสูง (ปวส.) หรืออนุปริญญา

4.5 โรงเรียนขยายโอกาสทางการศึกษา หมายถึงโรงเรียนสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน (สพฐ.) และสังกัดการปกครองส่วนท้องถิ่น (อปท.) ซึ่งเปิดการเรียนการสอนตั้งแต่ระดับอนุบาล ถึงระดับมัธยมศึกษาตอนต้นหรือบางที่เปิดถึงระดับชั้นมัธยมศึกษาตอนปลาย ซึ่งอยู่ในพื้นที่ห่างไกลที่เยาวชนไม่สามารถเดินทางไปเรียนโรงเรียนระดับมัธยมศึกษาได้

5. ประโยชน์ที่คาดว่าจะได้รับ

5.1 ได้แบบจำลองพยากรณ์ผลสัมฤทธิ์ทางการเรียนเพื่อสนับสนุนการตัดสินใจการศึกษาต่อหลังจบชั้นมัธยมศึกษาปีที่ 3

5.2 นักเรียนได้ข้อมูลสำหรับวางแผนการศึกษาต่อที่ได้จากการพยากรณ์ข้อมูลผลสัมฤทธิ์ทางการเรียนเพื่อสนับสนุนการตัดสินใจ ครูแนะแนวใช้วางแผนการให้คำปรึกษาการศึกษาต่อให้กับนักเรียน

5.3 ได้ระบบพยากรณ์ที่มีประสิทธิภาพที่มีความถูกต้อง ความแม่นยำและน่าเชื่อถือทำให้สามารถสนับสนุนการตัดสินใจการศึกษาต่อหลังจบชั้นมัธยมศึกษาปีที่ 3ได้

บทที่ 2
วรรณกรรมที่เกี่ยวข้อง

การศึกษาวิจัยระบบพยากรณ์เพื่อสนับสนุนการตัดสินใจศึกษาต่อหลังสำเร็จมัธยมศึกษาตอนต้น โดยใช้อัลกอริธึมป่าสุ่ม ในครั้งนี้ผู้วิจัยได้ทำการศึกษาค้นคว้าเอกสารข้อมูลแหล่งสารสนเทศและงานวิจัยที่เกี่ยวข้องเพื่อนำไปประยุกต์ใช้ ดังหัวข้อต่อไปนี้

1. การศึกษาขั้นพื้นฐาน (Basic Education)
2. การพยากรณ์ (Prediction)
3. เหมืองข้อมูล (Data mining)
4. เทคนิคการจำแนกข้อมูล (Classification)
5. ซอฟต์แวร์ Weka (Waikato Environment for Knowledge Analysis)
6. งานวิจัยที่เกี่ยวข้อง

1. การศึกษาขั้นพื้นฐาน (Basic Education)

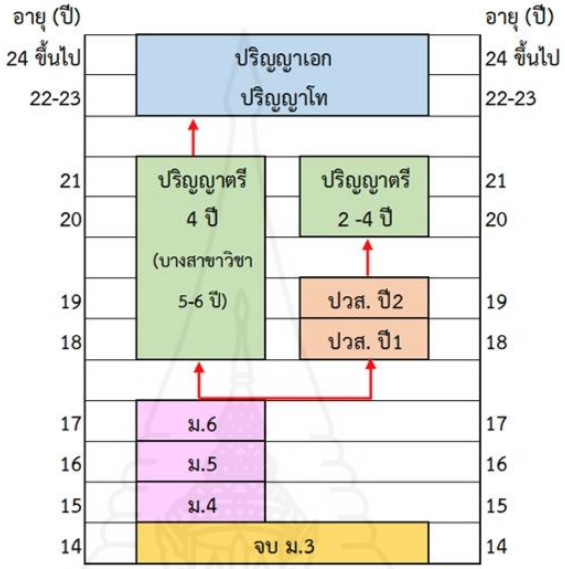
ระบบการศึกษาไทยตามที่กำหนดไว้ในพระราชบัญญัติการศึกษาแห่งชาติ พ.ศ.2542 แก้ไขเพิ่มเติม (ฉบับที่ 2) 2545 แบ่งการศึกษาเป็น 2 ระดับ คือ 1) ระดับการศึกษาขั้นพื้นฐาน คือ การจัดการศึกษาไม่น้อยกว่า 12 ปี หรือการศึกษาก่อนระดับอุดมศึกษา ซึ่งแบ่งออกเป็นก่อนประถมศึกษา ประถมศึกษา และมัธยมศึกษา 2) ระดับการศึกษาอุดมศึกษา หรือหลังการศึกษาขั้นพื้นฐานซึ่งจะแบ่งออกเป็นระดับต่ำกว่าปริญญา และ ปริญญา

การศึกษาก่อนประถมศึกษา หมายถึง การศึกษาในประเภทศูนย์พัฒนาเด็กเล็ก และชั้นเด็กเล็ก โดยมีระยะเวลาเรียน 1 ปี และ การจัดการศึกษาประเภทอนุบาล มี 2 หลักสูตร คือ อนุบาลหลักสูตร 2 ปี และ หลักสูตร 3 ปี

การศึกษาระดับประถมศึกษา หมายถึง การศึกษาที่มุ่งให้ผู้เรียนมีความรู้ความสามารถขั้นพื้นฐาน ในเวลาเรียน 6 ปี

การศึกษาระดับมัธยมศึกษา หมายถึง การศึกษาหลังระดับประถมศึกษา ซึ่งแบ่งออกเป็น 2 ช่วงตอน คือ ระดับมัธยมศึกษาตอนต้น ตั้งแต่ชั้นมัธยมศึกษาปีที่ 1

ถึงชั้นมัธยมศึกษาปีที่ 3 และ ระดับมัธยมศึกษาตอนปลาย การศึกษาต่อในชั้นมัธยมศึกษาตอนปลาย แบ่งเป็น 2 ประเภท คือ 1) ประเภทสามัญศึกษา คือ ชั้นมัธยมศึกษาปีที่ 4 ถึง ชั้นมัธยมศึกษาปีที่ 6 ดังภาพที่ 2.1

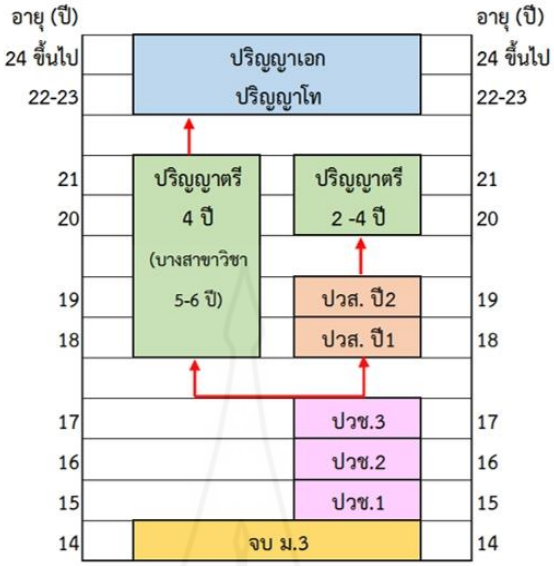


ภาพที่ 2.1 เส้นทางการศึกษาต่อประเภทสามัญศึกษา

ที่มา : https://sites.google.com/site/senthangkarsuksatxhlangcbm3

จากภาพที่ 2.1 แสดงเส้นทางการศึกษาต่อประเภทสามัญศึกษา สำหรับผู้เรียนที่สำเร็จการศึกษาชั้นมัธยมศึกษาปีที่ 3 ที่ตัดสินใจเลือกศึกษาต่อในประเภทสามัญศึกษา

มัธยมศึกษาตอนปลายประเภทสามัญเป็นการจัดการศึกษาเพื่อเป็นพื้นฐานในการศึกษาต่อในระดับอุดมศึกษา 2) ประเภทอาชีวศึกษา คือ หลักสูตรประกาศนียบัตรวิชาชีพ (ปวช.) เป็นการจัดการศึกษาเพื่อพัฒนาความรู้และทักษะในการประกอบอาชีพ หรือ ศึกษาต่อในระดับประกาศนียบัตรวิชาชีพชั้นสูง (ปวส.) ต่อไป ดังภาพที่ 2.2



ภาพที่ 2.2 เส้นทางการศึกษาต่อประเภทประเภทอาชีวศึกษา

ที่มา : https://sites.google.com/site/senthangkarsuksatxhlangcbm3

จากภาพที่ 2.2 แสดงเส้นทางการศึกษาต่อประเภทประเภทอาชีวศึกษา สำหรับผู้เรียนที่สำเร็จการศึกษาชั้นมัธยมศึกษาปีที่ 3 ที่ตัดสินใจเลือกศึกษาต่อในประเภทอาชีวศึกษา

การจัดการศึกษามี 3 รูปแบบ คือ การศึกษาในระบบ การศึกษานอกระบบ และการศึกษาตามอัธยาศัย

1.1 การศึกษาในระบบ หมายถึง การศึกษาที่กำหนดจุดมุ่งหมาย วิธีการ หลักสูตร ระยะเวลาการศึกษารวมทั้งการวัดและการประเมินผล ซึ่งเป็นเงื่อนไขของการสำเร็จการศึกษาที่แน่นอน

1.2 การศึกษานอกระบบ หมายถึง การศึกษาที่มีความยืดหยุ่นของการจัดรูปแบบวิธีการจัดการศึกษา ระยะเวลา รวมทั้งการวัดและประเมินผล ซึ่งเป็นเงื่อนไขของการสำเร็จการศึกษา โดยมีเนื้อหาที่สอดคล้องกับสภาพปัญหาและความต้องการของกลุ่มบุคคล

1.3 การศึกษาตามอัธยาศัย เป็นการศึกษาที่ให้ผู้เรียนมีโอกาสได้เรียนรู้ด้วยตนเองตามความสนใจ ความพร้อม และโอกาส โดยศึกษาจากบุคคล ประสบการณ์ สังคม สภาพแวดล้อม สื่อหรือแหล่งความรู้อื่นๆ

2. การพยากรณ์ (Prediction)

การพยากรณ์ หมายถึงการคาดการณ์เหตุการณ์ในอนาคต แนวโน้มของสิ่งที่สนใจที่จะเกิดขึ้นในอนาคต เพื่อนำมาเป็นสารสนเทศประกอบการตัดสินใจ การพยากรณ์จะต้องดำเนินการเป็นส่วนแรกสุดที่จะต้องวางแผน หรือ เตรียมการก่อนที่จะเริ่มทำอะไรเพื่อความถูกต้องแม่นยำในการตัดสินใจ โดยอาศัยเทคนิคการพยากรณ์ต่างๆ เพื่อเพิ่มประสิทธิภาพของการดำเนินงาน (รุ่งนภา ศรีประโค, 2556) ดังภาพที่ 2.3

ภาพที่ 2.3 กระบวนการการพยากรณ์

การพยากรณ์เป็นการทำนายหรือการคาดคะเนเหตุการณ์ในอนาคต ดังนั้นการพยากรณ์กับความเป็นจริงนั้นอาจจะเหมือนกันหรือไม่เหมือนกันก็ได้ การพยากรณ์ที่ดีต้องมีผลที่ใกล้เคียงกับความเป็นจริงดังนั้นเทคนิคหรือวิธีการพยากรณ์จึงเป็นเรื่องจำเป็นที่จะต้องนำมาทำความเข้าใจกัน เทคนิคของการพยากรณ์นั้นแบ่งออกเป็น 2 ประเภท ตามลักษณะวิธีการพยากรณ์ที่อาศัยข้อมูลในอดีตที่เก็บรวบรวมไว้อยู่มาใช้ประกอบการสร้างรูปแบบการพยากรณ์ได้ดังนี้ (รุ่งนภา ศรีประโค, 2556)

2.1 เทคนิคการพยากรณ์แบบไม่เป็นทางการ (Informal Forecasting Technique)

วิธีพยากรณ์แบบนี้จะใช้ประสบการณ์และดุลพินิจของผู้ทำการพยากรณ์ ผู้พยากรณ์จะเป็นผู้ที่มีความรู้ความเชี่ยวชาญในเรื่องนั้นๆ อย่างดีจึงสามารถคาดคะเนสิ่งต่างๆ ได้

2.2 เทคนิคการพยากรณ์แบบเป็นทางการ (Formal Forecasting Technique)

การพยากรณ์วิธีนี้จะต้องอาศัยข้อมูลมาสนับสนุนและใช้ความรู้ทางสถิติ คณิตศาสตร์เข้ามาช่วยในการวิเคราะห์ข้อมูล ผู้พยากรณ์จะเป็นใครก็ได้ที่เข้าใจวิธี และขั้นตอนในการนำข้อมูลมาวิเคราะห์เทคนิควิธีการพยากรณ์แบบเป็นทางการนี้ยังแบ่งออกเป็น 2 ประเภท คือ การพยากรณ์เชิงปริมาณ (Quantitative Forecasting) และการพยากรณ์เชิงคุณภาพ (Qualitative Forecasting)

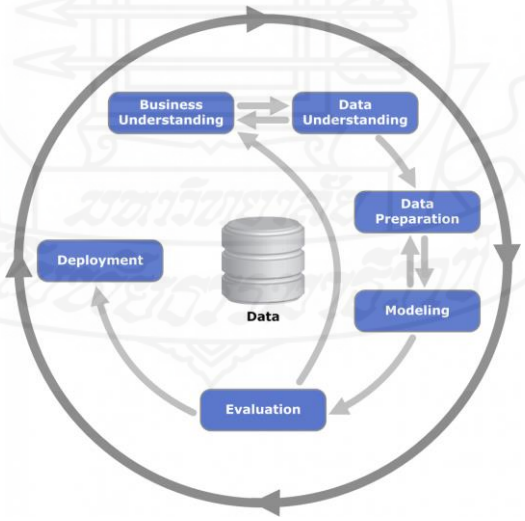
2.2.1 การพยากรณ์เชิงปริมาณ (Quantitative Forecasting) เป็นการพยากรณ์ที่ใช้ข้อมูลเชิงปริมาณ (ตัวเลข) อาศัยข้อมูลในอดีตที่มีอยู่ในการวิเคราะห์อย่างเป็นระบบ จากหลักการ

ทางคณิตศาสตร์หรือทางสถิติ การพยากรณ์รูปแบบที่นิยม คือ วิธีการวิเคราะห์ทางอนุกรมเวลา (Time Series Analysis) และวิธีทางความเป็นเหตุเป็นผล (Casual or Explanatory Method)

2.2.2 การพยากรณ์เชิงคุณภาพ (Qualitative Forecasting) เป็นการพยากรณ์ที่ใช้ผู้ที่มีประสบการณ์ ความรู้ ความสามารถ เป็นผู้พยากรณ์ ไม่ใช้ตัวแบบคณิตศาสตร์ ไม่มีการเก็บรวบรวมข้อมูลในอดีต จึงตรวจสอบความแม่นยำของการพยากรณ์ได้ยากกว่าการพยากรณ์เชิงปริมาณ ทั้งนี้ขึ้นอยู่กับความสามารถและประสบการณ์ของผู้พยากรณ์ เช่น การสอบถามกลุ่มผู้เชี่ยวชาญ (Delphi Method) การพยากรณ์โดยกลุ่มผู้บริหาร (Jury of executive opinion) พนักงานขายทำการพยากรณ์ (Sale force composite) และการสำรวจ (Consumer market survey) เป็นต้น

3. เทคนิคเหมืองข้อมูล (Data Mining)

เหมืองข้อมูล คือ การค้นหาสิ่งที่มีประโยชน์จากฐานข้อมูลที่มีขนาดใหญ่ โดยทำการวิเคราะห์ แยกแยะประเภท จำแนกรูปแบบและความสัมพันธ์ของข้อมูล มีเทคนิคต่างๆ หลากหลายวิธี กระบวนการทำเหมืองข้อมูลมีมาตรฐานในการวิเคราะห์ข้อมูลที่เรียนว่า “Cross-Industry Standard Process for Data Mining” หรือเรียกย่อๆว่า “Crisp-DM” จะประกอบด้วย 6 ขั้นตอน (Shearer, C., 2000: 13–22) ดังภาพที่ 2.4



ภาพที่ 2.4 ขั้นตอนกระบวนการ CRISP-DM

ที่มา : https://en.wikipedia.org/wiki/Cross_Industry_Standard_Process_for_Data_Mining

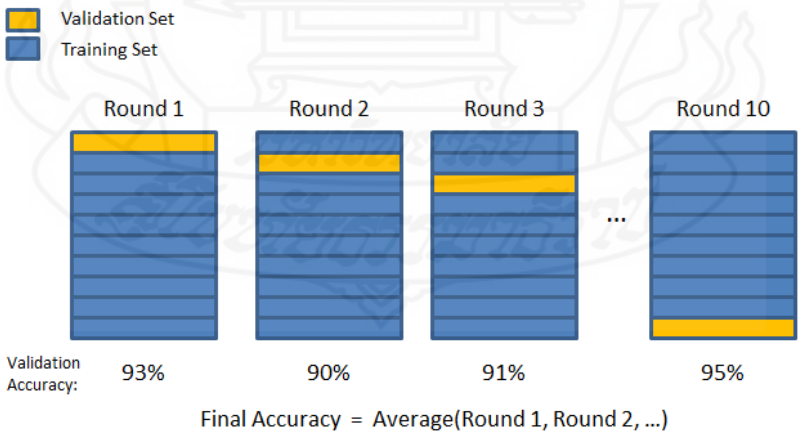
3.1 การวิเคราะห์ปัญหาและโอกาสทางธุรกิจ (Business Understanding) เป็นขั้นตอนที่เน้นไปที่การเข้าใจเกี่ยวกับปัญหาในรูปของโจทย์เพื่อนำไปใช้ในการวิเคราะห์ข้อมูลทาง data mining

3.2 การทำความเข้าใจข้อมูล (Data Understanding) เป็นการเก็บรวบรวมข้อมูล ตรวจสอบข้อมูลที่ได้รวบรวมมาเพื่อดูความถูกต้องของข้อมูลมาใช้ในการวิเคราะห์

3.3 การเตรียมข้อมูล (Data Preparation) เป็นขั้นตอนที่ทำการแปลงข้อมูลที่ได้เก็บรวบรวมมาให้กลายเป็นข้อมูลที่ถูกต้องหรือการเติมเต็มข้อมูลที่ขาดหายไปให้สามารถนำไปวิเคราะห์ได้เป็นขั้นตอนที่ต้องใช้เวลานาน

3.4 การสร้างโมเดล (Modeling) เป็นขั้นตอนของการวิเคราะห์ข้อมูลด้วยเทคนิคเหมืองข้อมูล เพื่อสร้างโมเดลให้ได้ผลลัพธ์ที่ดีที่สุด

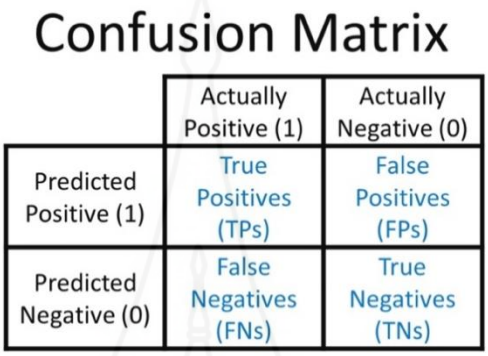
3.5 การวัดประสิทธิภาพ (Evaluation) เป็นการวัดประสิทธิภาพของผลลัพธ์ว่าตรงกับวัตถุประสงค์ที่ได้ตั้งไว้หรือไม่ มีความน่าเชื่อถือมากน้อยเพียงได้ ของโมเดลที่สร้างไว้ในการสร้างแบบจำลองวิจัยในครั้งนี้จะทดสอบประสิทธิภาพของแบบจำลอง (Model) แบบ 10-fold cross validation ซึ่งเป็นการแบ่งข้อมูลที่ใช้ในการทดสอบออกเป็นชุดๆ เพื่อทดสอบแบบจำลองหาความน่าเชื่อถือแบบ 10-fold cross validation แบ่งข้อมูลฝึกสอนเป็น 9 ชุด และ 1 ชุดเป็นข้อมูลทดสอบประสิทธิภาพของโมเดล ทำแบบวนรอบไปเรื่อยๆ จนครบจำนวนชุดที่แบ่งไว้ ตัวอย่างการแบ่งชุดข้อมูล ดังภาพที่ 2.5



ภาพที่ 2.5 การแบ่งข้อมูลแบบ K-fold-10-fold cross validation

ที่มา : Ofer D. (2015, pp.18)

ก่อนนำแบบจำลองไปใช้งานจริงเราจะต้องทราบประสิทธิภาพของแบบจำลองก่อน โดยทั่วไปตัววัดที่นิยมนำมาใช้ในงานวิจัย คือ การวัดค่าประสิทธิภาพค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าครบถ้วน (Recall) และค่าประสิทธิภาพโดยรวม (F-Measure) จากตาราง confusion matrix เป็นการประเมินผลลัพธ์การทำนาย ดังภาพที่ 2.6



ภาพที่ 2.6 ตาราง confusion matrix

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

จากภาพที่ 2.6 ค่าในตารางประกอบไปด้วย

True Positive (TP) คือ สิ่งที่โปรแกรมทำนายว่าจริงและสิ่งที่เกิดขึ้นเป็นจริง

True Negative (TN) คือ สิ่งที่โปรแกรมทำนายว่าไม่จริงและสิ่งที่เกิดขึ้นไม่จริง

False Positive (FP) คือ สิ่งที่โปรแกรมทำนายว่าจริงแต่สิ่งที่เกิดขึ้นไม่จริง

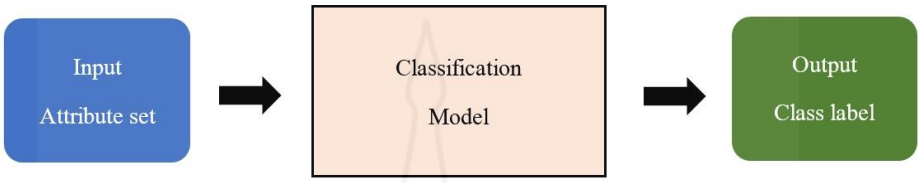
False Negative (FN) คือ สิ่งที่ทำนายว่าไม่จริงแต่สิ่งที่เกิดขึ้นเป็นจริง

3.6 การนำโมเดลที่พัฒนาไปใช้งาน (Deployment) เป็นการนำเอาโมเดลที่สร้างไว้จากการวิเคราะห์ข้อมูลไปใช้จริง

4. เทคนิคการจำแนกประเภทข้อมูล (Classification)

เทคนิคการจำแนกประเภทข้อมูล คือ เทคนิคการทำเหมืองข้อมูล โดยค้นหาลักษณะสำคัญของชุดข้อมูลที่ต้องการ เพื่อสร้างโมเดลเชื่อมโยงความสัมพันธ์ในการระบุประเภท (class) หรือกลุ่มจากคุณสมบัติต่างๆ ของชุดข้อมูล เป็นการจำแนกประเภทที่สามารถทราบค่าเป้าหมาย (target) สามารถระบุประเภทของข้อมูลที่ต้องการจำแนกไว้ก่อนได้

หลักการทำงานพื้นฐานของการจำแนกข้อมูล เริ่มจากการนำเข้าชุดข้อมูล (attribute set) โดยการใช้เทคนิคจำแนกประเภทเพื่อสร้างโมเดล (classification model) โดยนำโมเดลที่ได้มาระบุประเภท (class label) ของข้อมูลดังภาพที่ 2.7 ในการวิจัยนี้ได้ใช้เทคนิคการจำแนกข้อมูล ดังนี้



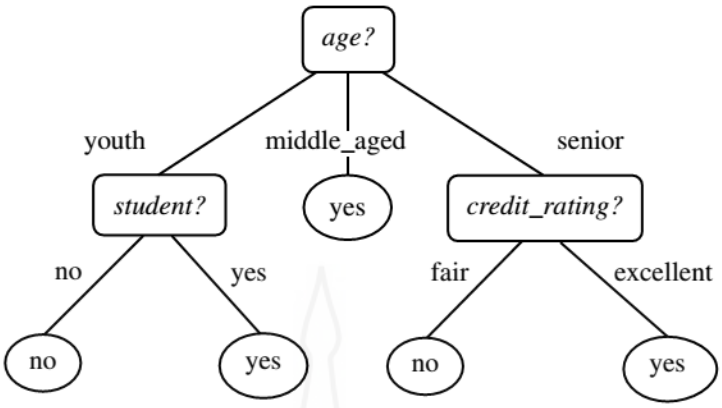
ภาพที่ 2.7 การทำงานพื้นฐานของการจำแนกข้อมูล

ที่มา : Tan, Steinbach. and Kumar (2005, pp.146)

4.1 อัลกอริธึมการตัดสินใจแบบโครงสร้างต้นไม้ (Decision Tree)

อัลกอริธึมการตัดสินใจแบบโครงสร้างต้นไม้ เป็นหนึ่งในเทคนิคของ Classification เป็นวิธีการแบ่งประเภทหรือแยกหมวดหมู่ข้อมูล ซึ่ง Classification นั้นเป็นเทคนิคหนึ่งของเหมืองข้อมูล (Data Mining) เป็นอัลกอริธึมที่อาศัยโครงสร้างของต้นไม้ ประกอบด้วยโหนดรากมีการแตกกิ่งก้านแทนคุณลักษณะของข้อมูล และใบแทนการแบ่งประเภทของข้อมูลซึ่งได้รับความนิยมในการจำแนกประเภท ส่วนประกอบของต้นไม้ตัดสินใจประกอบไปด้วยโหนด (node) คือจุดแยกข้อมูลของคุณลักษณะจะไปในทิศทางใดซึ่งโหนดบนสุดเรียกว่า โหนดราก (root node) กิ่ง (branch) คือ คุณสมบัติของโหนดที่แตกออกมาจะเท่ากับคุณสมบัติของโหนด ใบ (leaf) ส่วนที่อยู่ล่างสุดของต้นไม้ตัดสินใจคือค่าที่เป็นไปได้จากการทดสอบชุดข้อมูลนั้นๆ เป้าหมาย (class) คือผลลัพธ์ของการทำนาย (สุรวัชร ศรีเปารยะ, 2560)

อัลกอริธึมการเรียนรู้แบบ C4.5(J48) เป็นอัลกอริธึมที่ใช้ในการสร้างต้นไม้ตัดสินใจ (Decision Tree) ซึ่งได้รับการพัฒนาเพิ่มเติมมาจากอัลกอริธึม ID 3 โดย C4.5(J48) จะสร้างรูปแบบต้นไม้ตัดสินใจมาจากชุดข้อมูลที่ใช้หลักการของ Information Entropy (Quinlan, 1992) ในการเลือกโหนดนั้นจะใช้ค่าความถูกต้องของคุณลักษณะโดยแต่ละคุณลักษณะเพื่อใช้แยกเส้นทางตัดสินใจจะพิจารณาจากค่า Normalized Information Gain ในการทำงานของอัลกอริธึม C4.5(J48) จะมีลักษณะการทำงานแบบวนซ้ำในการเลือกคุณลักษณะที่เหมาะสมที่จะเป็นโหนดตัดสินใจไปเรื่อยๆ จนถึงโหนดสุดท้ายที่จะเป็นเส้นทางของใบ ดังภาพที่ 2.8

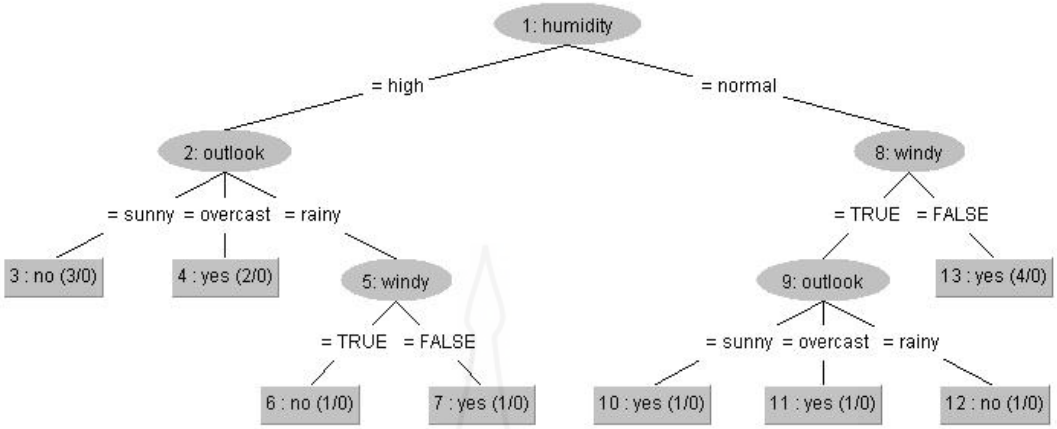


ภาพที่ 2.8 ตัวอย่างต้นไม้ตัดสินใจสำหรับการจำแนกคุณลักษณะของลูกค้าที่ทำการซื้อคอมพิวเตอร์
ที่มา : Han, J. and Kamber M. (2006)

จากภาพที่ 2.8 ตัวอย่างต้นไม้ตัดสินใจสำหรับการจำแนกคุณลักษณะของลูกค้าที่ทำการซื้อคอมพิวเตอร์ ประกอบไปด้วย student และ credit rating ที่เป็นโหนด (node) เป็นตัวทดสอบ yes และ no เป็นใบ (leaf) และโหนดที่บนสุดโหนดราก (root node) คือ age จากตัวอย่างได้แสดงแบบจำลองในการตัดสินใจในการซื้อคอมพิวเตอร์จะเห็นว่ากลุ่มที่เป็น “middle aged” หรือ “youth” ที่เป็น student หรือ “senior” ที่มี “credit rating” ที่เป็น excellent จะตัดสินใจซื้อคอมพิวเตอร์

4.2 อัลกอริธึมการจำแนกข้อมูลต้นไม้สุ่ม (Random Tree)

อัลกอริธึมต้นไม้สุ่ม (Random Tree) คือ เทคนิคที่ใช้ในการจำแนกหมวดหมู่เช่นเดียวกับ C4.5(J48) โดยมีหลักการสร้างต้นไม้จากการสุ่มต้นไม้หลายๆ แบบ ในแต่ละโหนดแล้วเลือก มาประมวลผลโดยใช้การตัดแต่งกิ่ง (Pruning) และเนื่องจากจำนวนของต้นไม้เพิ่มขึ้นอย่างรวดเร็ว ซึ่งยากแก่การแก้ปัญหาการสร้างต้นไม้ที่เป็นไปได้ทั้งหมด ส่วนประกอบสามารถสร้างชุดสุ่มของ ต้นไม้ออกมาจากการกระจายชุดต่างๆ ของต้นไม้ Random Tree เป็นการสุ่มวาดที่สุ่มจากชุดของต้นไม้ ที่เป็นไปได้ (Weka. Available: http://www.cs.waikato.ac.nz/ml, 2016) ดังภาพที่ 2.9



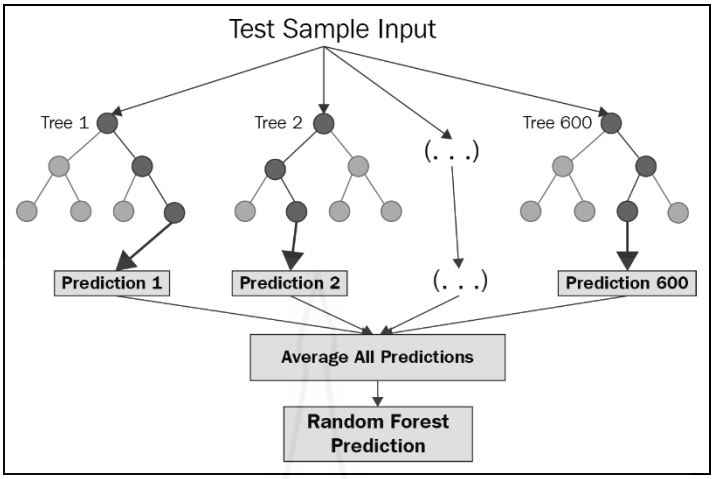
ภาพที่ 2.9 โครงสร้างการจำแนกแบบต้นไม้สุ่ม (random tree)

ที่มา : โครงสร้างการจำแนกแบบต้นไม้สุ่มจากโปรแกรม WEKA

จากภาพที่ 2.9 เป็นโครงสร้างการจำแนกแบบต้นไม้สุ่มแสดงให้เห็นการปรับปรุงประสิทธิภาพของต้นไม้ตัดสินใจ เกิดจากวิธีการสุ่มชุดข้อมูลที่อยู่ภายในโครงสร้างต้นไม้ โดยมี humidity เป็นโหนดราก และสุ่มโหนดที่มีความเป็นไปได้ outlook และ windy เพื่อไปสู่หรือผลลัพธ์ (Class) “no” หรือ “yes”

4.3 อัลกอริธึมการจำแนกข้อมูลป่าสุ่ม (Random Forest)

คือการสร้างโมเดลด้วยเทคนิคต้นไม้ตัดสินใจ (Decision Tree) หลาย ๆ โมเดล ตั้งแต่ 10 โมเดลขึ้นไป โดยแต่ละโมเดลจะมีชุดของข้อมูลที่ไม่เหมือนกันโมเดลแต่ละตัวจะเรียนรู้อย่างเป็นอิสระต่อกัน เอาผลการตัดสินใจของโมเดลที่ทำนายแล้วจากเสียงส่วนมาก (Vote) ว่า Class ไหนที่ถูกเลือกมากที่สุดมาใช้ในการตัดสินใจ ข้อดีของเทคนิคการจำแนกข้อมูลด้วยป่าสุ่มคือให้ผลการทำนายที่แม่นยำและเกิดปัญหา Overfitting น้อย (ปรเมษฐ์ ธันวานนท์, ชัยกร ยิ่งเสร, และ วรพล พงษ์เพ็ชร, 2560) ดังภาพที่ 2.10



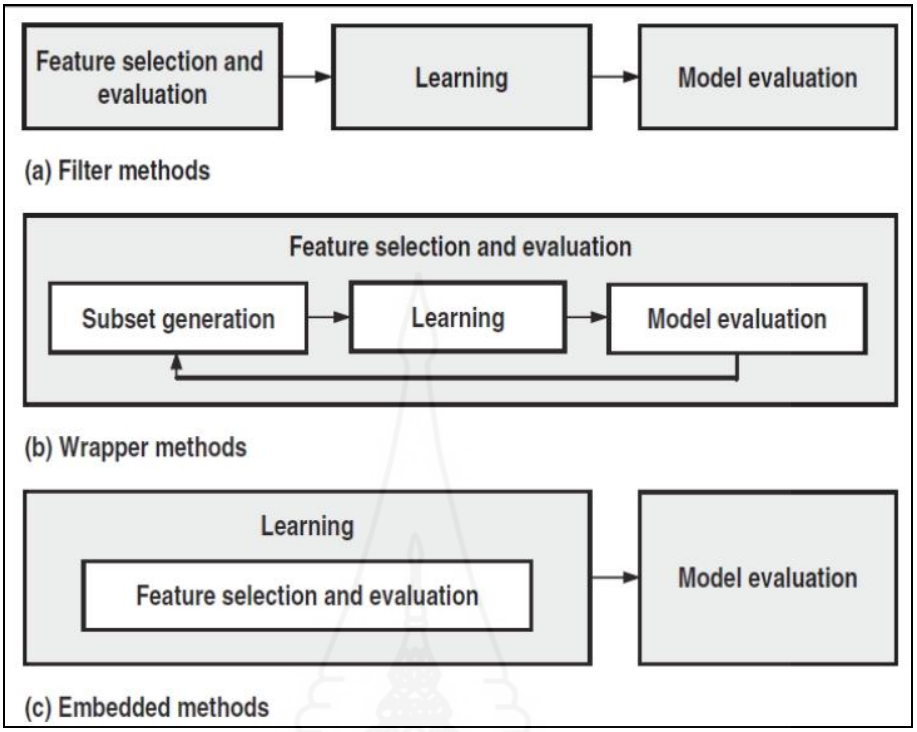
ภาพที่ 2.10 โครงสร้างการจำแนกแบบป่าสุ่ม (random forest)

ที่มา : https://medium.com/swlh/random-forest-and-its-implementation-71824ced454f

จากภาพที่ 2.10 เป็นการสร้างต้นไม้ตัดสินใจหลายๆ ต้น เพื่อพยากรณ์ชุดข้อมูลแล้วนำผลการตัดสินใจหรือผลลัพธ์ของต้นไม้แต่ละต้นมาโหวตว่าผลลัพธ์ (class) ไหนถูกเลือกมากที่สุดจะถูกนำมาใช้

4.4 การเลือกคุณลักษณะ (Feature Selection)

เป็นการคัดเลือกจำนวนคุณลักษณะที่เหมาะสมและลดจำนวนของคุณลักษณะที่ไม่เกี่ยวข้องออกทำให้ได้คุณลักษณะที่มีความสำคัญเกี่ยวข้องกับเป้าหมาย วิธีการคัดเลือกคุณลักษณะแบ่งออกเป็น 3 วิธี ได้แก่ 1) วิธีฟิลเตอร์ (Filter method) 2) วิธีแรปเปอร์ (Wrapper method) 3) วิธีฝังตัว (Embed method) แสดงการเปรียบเทียบแนวคิดของทั้ง 3 วิธี (Hilario, M. and Kalousis, A, 2008) ดังภาพที่ 2.11



ภาพที่ 2.11 การเปรียบเทียบแนวความคิดของการคัดเลือกคุณสมบัติ (a) วิธีฟิลเตอร์ (b) วิธีแรปเปอร์ และ (c) วิธีฝังตัว

ที่มา : Hilario, M. and Kalousis, A. (2008)

วิธีฟิลเตอร์ (Filter Method) เป็นการคัดเลือกตัวแปรโดยประเมินความสำคัญของตัวแปรด้วยการพิจารณาคุณสมบัติที่แท้จริงของข้อมูล ซึ่งวิธีฟิลเตอร์จะคำนวณค่าความสำคัญของตัวแปรจากดัชนีวัดความสำคัญและเลือกตัวแปรที่ให้ค่าดัชนีสูง เทคนิคในการคำนวณค่าน้ำหนักของคุณลักษณะมีหลายวิธี เช่น ค่าสารสนเทศร่วม (Mutual Information : MI) Information Gain (IG) Chi-Square (Chi2) Fisher Score, Gini index และ Relief เป็นต้น ข้อดีของวิธีฟิลเตอร์ คือ สามารถปรับขนาดของชุดข้อมูลที่มีหลายมิติได้ง่ายรวดเร็วและเป็นอิสระกับวิธีการจำแนกประเภท ทำให้วิธีนี้จะต้องดำเนินการเพียงครั้งเดียวจึงสามารถประเมินตัวแปรที่แตกต่างกันได้ (Saeys, Inza, and Larrañaga, 2007)

ในงานวิจัยนี้ผู้วิจัยได้เลือกคุณลักษณะพื้นฐานที่ได้รับความนิยมในงานทางเหมืองข้อมูลคือ ค่าการเพิ่มสารสนเทศ information Gain (Xu, J. and Jiang, H, 2015) เป็นวิธีที่ใช้เทคนิคการจัดเรียงลำดับของคุณลักษณะโดยคำนวณหาค่าน้ำหนักความสัมพันธ์ระหว่างฟีเจอร์และคลาส

ต่างๆ ด้วยค่า Information Gain เป็นการประเมินเพื่อใช้ในการแบ่งข้อมูลด้วยการคำนวณค่า Gain แต่ละมิติข้อมูลถ้าข้อมูลใดมีค่า Gain สูงสุดหรือมีค่ามากที่สุดจะใช้สำหรับเป็นโหนดราก จะถูกเลือกให้เป็นกลุ่มย่อยที่มีอำนาจจำแนก ดังสมการที่ 1 แสดงการคำนวณวัดความสามารถในการจำแนกประเภทข้อมูลภายในกลุ่มในทฤษฎีสารสนเทศหรือค่า Entropy ผลของค่า Entropy น้อยย่อมดี และคำนวณค่า Gain ดังสมการที่ 2

$$\text{Entropy}(p) = \sum_{i=0}^{c-1} p(j|t) log_2 p(j|t) \quad (1)$$

โดยที่ $\sum_i$ คือ ผลรวมของความน่าจะเป็นของค่า j ที่เกิดในคลาส t

$p(j|t)$ คือค่าความถี่ที่มีความสัมพันธ์ของกลุ่ม j กับโหนด t

$$Gain = Entropy(p) - \left(\sum_{i=1}^{k} \frac{ni}{n} Entropy(i)\right) \quad (2)$$

โดยที่ Entropy (p) คือ ค่า Entropy ของตัว Root

$\sum_{i=1}^{k} \frac{ni}{n} Entropy\ (i)$ คือ ค่า Entropy ในแต่ละโหนดย่อย

วิธีแรปเปอร์ (Wrapper method) เป็นการคัดเลือกตัวแปรด้วยการสร้างตัวแบบทำนายขึ้นมาจากเซตตัวแปรที่กำหนดไว้และวัดความสำคัญของเซตย่อยของตัวแปร โดยเลือกเซตย่อยที่มีความแม่นยำในการจำแนกประเภทข้อมูลสูง หรือใช้ความแม่นยำในการจำแนกประเภทข้อมูลเมื่อใช้เซตย่อยนั้นๆ ในการเป็นดัชนีวัดความสำคัญของเซตย่อย (Saeys, Inza, and Larranaga, 2007; Yu, L. and Liu, H. 2004)

วิธีฝังตัว (Embed method) เป็นวิธีการคัดเลือกตัวแปรที่ใช้กระบวนการเรียนรู้ของตัวแบบการทำนาย โดยทำการเลือกเซตที่มีตัวแปรที่เหมาะสมในแต่ละขั้นตอนวิธี พร้อมไปกับการสร้างตัวแบบสำหรับการจำแนกประเภทสำหรับการทำนาย (Janecek, 2009, pp.41-43; Saeys, Inza, and Larranaga, 2007)

5. ซอฟต์แวร์ Weka (Waikato Environment for Knowledge Analysis)

Weka ย่อมาจาก Waikato Environment for Knowledge Analysis ซึ่งเป็นโปรแกรมที่สามารถใช้งานได้ฟรีสามารถดาวน์โหลดได้จากเว็บไซต์ ลักษณะโปรแกรมได้ถูกพัฒนาขึ้นโดยใช้ภาษาจาวา ซึ่งเขียนมาโดยเน้นกับงานทางด้านการเรียนรู้ด้วยเครื่อง (Machine Learning) และการทำเหมืองข้อมูล (Data Mining) โปรแกรมจะประกอบไปด้วยโมดูลย่อย ๆ ไว้ใช้ในการจัดการข้อมูล และเป็นโปรแกรมที่สามารถใช้ Graphic User Interface (GUI) และใช้คำสั่งในการให้ซอฟต์แวร์ประมวลผล ดังภาพที่ 2.12

ภาพที่ 2.12 หน้าจอหลักโปรแกรม Weka

5.1 โปรแกรมหลักของซอฟต์แวร์ Weka

5.1.1 Explorer เป็นส่วนที่ใช้ฟังก์ชั่นการทำงานต่างๆ ในลักษณะ GUI

5.1.2 Experimenter เป็นส่วนที่ออกแบบการทดลองและการทดสอบผล

5.1.3 Knowledge Flow เป็นส่วนของการนำเทคนิคต่างๆ มาเรียงต่อกันเพื่อวิเคราะห์

5.1.4 Workbench ใช้ปรับแต่งค่าข้อมูลที่นำมาทดลอง

5.1.5 Simple CLI เป็นส่วนที่ใช้รับคำสั่งการทำงานผ่านการพิมพ์

5.2 ส่วนประกอบเมนูหลักในการทำเหมืองข้อมูล Explorer ดังภาพที่ 2.13

5.2.1 Preprocess การเตรียมข้อมูล

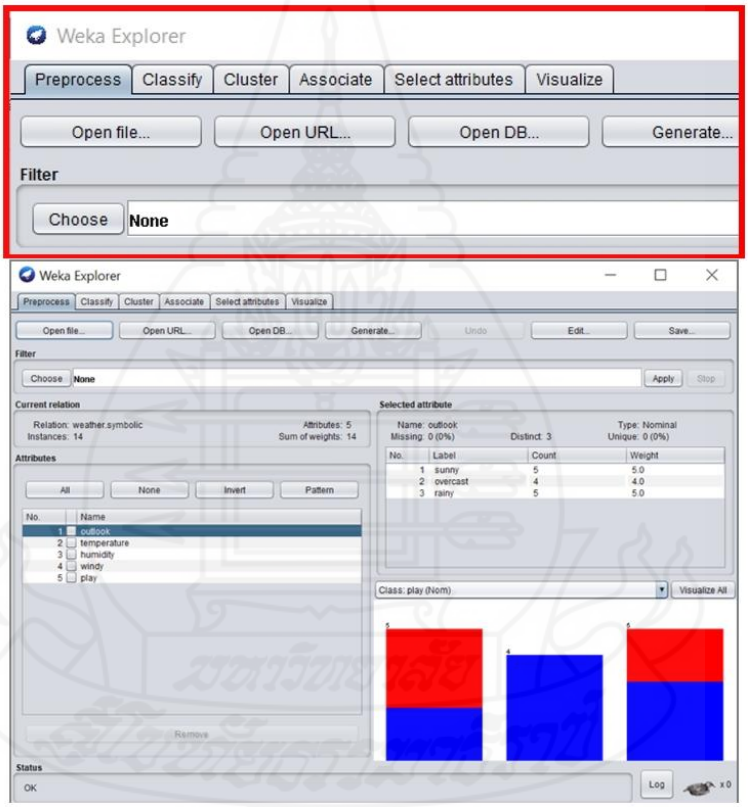
5.2.2 Classify รวมโมดูลการทำเหมืองข้อมูลแบบจัดจำแนกประเภท

5.2.3 Cluster รวมโมดูลการทำเหมืองข้อมูลแบบการเกาะกลุ่ม

5.2.4 Associate รวมโมดูลการทำเหมืองข้อมูลแบบกฎเชื่อมโยง

5.2.5 Select attributes รวมโมดูลสำหรับการวิเคราะห์ความเกี่ยวพันของลักษณะประจำ

5.2.6 Visualize นำเสนอข้อมูลด้วยภาพนามธรรมสองมิติ



ภาพที่ 2.13 หน้าส่วนประกอบเมนู Explorer โปรแกรม Weka

6. งานวิจัยที่เกี่ยวข้อง

รุจิรา ธรรมสมบัติ (2554) ได้ทำการศึกษาและสร้างโมเดลที่ถูกสร้างจาก อัลกอริธึมต้นไม้ตัดสินใจ ID3 และอัลกอริธึม C4.5 (J48) ผลการศึกษาพบว่า อัลกอริทึม ID3 มีค่าความถูกต้อง

มากกว่า C4.5 (J48) โดยมีค่าความถูกต้อง (Correctly Classified Instances) เมื่อทดสอบกับชุดกลุ่มข้อมูลสำหรับการเรียนรู้ (Training Data) จำนวน 1,000 ชุดเท่ากับ 92.3% และเมื่อนำอัลกอริธึมต้นไม้ตัดสินใจ ID3 มาทดสอบกับชุดข้อมูลทดสอบ (Testing Data) จำนวน 500 ชุด ผลการศึกษาพบว่า มีค่าความถูกต้องเท่ากับ 92.2% และเมื่อพิจารณาค่า Confusion Matrix พบว่าผลการทำนายจากโมเดลมีจำนวนข้อมูลค่าจริง กับจำนวนข้อมูลจากการทำนายของโมเดลมีผลลัพธ์ตรงกันได้ค่าเฉลี่ยรวมเท่ากับ 83.06% ซึ่งเป็นค่าเฉลี่ยที่อยู่ในระดับค่อนข้างสูง สามารถนำโมเดลที่ได้ไปพัฒนาระบบต่อไป

เสกสรรค์ วิลัยลักษณ์ (2558) ได้ศึกษาเรื่อง การใช้เทคนิคการทำเหมืองข้อมูลเพื่อพยากรณ์ผลการเรียนของนักเรียน โดยมีวัตถุประสงค์เพื่อพัฒนาคลังข้อมูล และสร้างตัวแบบพยากรณ์ผลการเรียนของนักเรียนโรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน ศูนย์วิจัยและพัฒนาการศึกษา โดยใช้ข้อมูลนักเรียนระดับมัธยมศึกษาระหว่างปีการศึกษา 2548 – 2556 เพื่อพัฒนาคลังข้อมูลโดยใช้โครงสร้างแบบสโนว์เฟลก สกีมา และนำเสนอรายงาน โดยใช้ข้อมูลของนักเรียนชั้นมัธยมศึกษาปีที่ 4 ระหว่าง ปีการศึกษา 2553 – 2556 จำนวน 525 ระเบียน ประกอบด้วย 16 คุณลักษณะ นำมาสร้างตัวแบบพยากรณ์ผลการเรียน โดยใช้กระบวนการคัดเลือกคุณลักษณะแบบ Correlation-based Feature Selection (CFS) และวิธี Information Gain (IG) ใช้เทคนิคเหมืองข้อมูลแบบโครงข่ายประสาทเทียม แบบมัลติเลเยอร์เพอร์เซ็ปตรอน (MLP) ซัพพอร์ตเวกเตอร์แมชชีน (SVM) และต้นไม้ตัดสินใจ (Decision Tree) มาสร้างตัวแบบพยากรณ์และเปรียบเทียบทดสอบประสิทธิภาพแบบ 10-fold Cross Validation ผลการวิจัยพบว่า การทำเหมืองข้อมูลสร้างตัวแบบพยากรณ์ ชุดข้อมูลแบบไม่จัดกลุ่มโดยวิธีคัดเลือกคุณลักษณะด้วยวิธี Correlation-based Feature Selection (CFS) ร่วมกับโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน ให้ค่าความถูกต้องสูงที่สุดที่ร้อยละ 94.48 และมีค่ารากที่สองของความคลาดเคลื่อนน้อยที่สุดที่ 0.1880 ที่จะนำไปสร้างระบบพยากรณ์ผลการเรียนของนักเรียน

ฐิติมา ช่วงชัย (2559) ได้ศึกษาการวิเคราะห์หารูปแบบการเรียนรู้โดยใช้เหมืองข้อมูลของนักศึกษาต่อการจัดทำปริญญานิพนธ์ โดยการวิเคราะห์ข้อมูลด้วยการทำเหมืองข้อมูลจากแบบสอบถาม เลือกวิธีวิเคราะห์ข้อมูลด้วยรูปแบบ Rule Based Classification ด้วยวิธี Decision Table, Jrip และ PART และรูปแบบ Decision Tree Classification ด้วยวิธี LMT, J48 และ Random Tree มาวิเคราะห์ตัวแปรเพศและชั้นปีเพื่อใช้เป็นแนวทางการแนะนำนักศึกษาสำหรับการเลือกหัวข้อในการจัดทำปริญญานิพนธ์ให้เหมาะกับตนเองด้วยวิธี Random Tree ผลการวิเคราะห์พบว่าส่วนใหญ่มีรูปแบบการเรียนรู้เป็นเอกนัย โดยรูปแบบของ Decision Tree ด้วยวิธี Random Tree ให้ค่าความถูกต้อง 100% ทำนายรูปแบบการเรียนรู้เป็นแบบเอกนัย และรูปแบบของ Rule Based

ด้วยวิธี PART ให้ค่าความถูกต้องสูงสุด 84.12% เพศหญิงจะมีรูปแบบการเรียนรู้แบบเอกนัย ส่วนเพศชายมีการเรียนรู้แบบอเนกนัย เช่นกันในชั้นปีที่ 3 ซึ่งเป็นชั้นปีที่ต้องจัดทำปริญญานิพนธ์

กฤษฎิ์ บรรณะชัยศิริสุข (2558) ได้ทำการศึกษาเครื่องมือท่องเว็บไซต์บนอุปกรณ์เคลื่อนที่สำหรับผู้พิการทางสายตาโดยใช้การตรวจหาเนื้อหาหลัก มีวัตถุประสงค์เพื่อพัฒนาเครื่องมือสำหรับผู้พิการทางสายตาในการเข้าถึงเนื้อเว็บไซต์ผ่านเครื่องมือสื่อสารระบบไอโอเอส โดยใช้การค้นหาเนื้อหาหลักของหน้าเว็บไซต์ด้วยวิธีการทดสอบเพื่อเปรียบเทียบการค้นหาเนื้อหาหลักด้วย 3 วิธี การจำแนกแบบป่าสุ่ม การจำแนกแบบต้นไม้ตัดสินใจ และการจำแนกแบบโครงข่ายประสาทเทียม ร่วมกับการทดสอบแบบ 10-fold cross validation ผลการศึกษาพบว่า การทดสอบตัวจำแนกที่สร้างด้วยวิธีป่าแบบสุ่มด้วยการใช้วิธี 10-fold cross validation มีผลลัพธ์ความถูกต้องในการจำแนกระหว่างส่วนประกอบอื่นๆ กับส่วนประกอบที่เป็นเนื้อหาหลักอยู่ที่ 87.72% ซึ่งมีความถูกต้องมากกว่า การจำแนกแบบต้นไม้ตัดสินใจ และการจำแนกแบบโครงข่ายประสาทเทียม

วันวิสาข์ ชนะประเสริฐ (2559) ได้ทำการประยุกต์ใช้เทคนิคเหมืองข้อมูลเพื่อแนะนำอาชีพสำหรับนักศึกษาปริญญาตรี คณะโบราณคดี มหาวิทยาลัยศิลปากร โดยมีวัตถุประสงค์เพื่อคัดเลือกคุณลักษณะที่เหมาะสมของตัวแปร นำมาใช้ในการวิเคราะห์หาเทคนิคเหมืองข้อมูลเพื่อสร้างแบบจำลองที่มีประสิทธิภาพ ใช้ทำนายผลสำหรับแนะแนวทางประกอบอาชีพนักศึกษาปริญญาตรี คณะโบราณคดี มหาวิทยาลัยศิลปากร ผลการวิจัยพบว่าประสิทธิภาพแบบจำลองที่ดีที่สุดโดยการคัดเลือกคุณลักษณะด้วยวิธี InfoGainAttributeEval ร่วมกับเทคนิค Neural Network มีค่าความถูกต้องของการจำแนกข้อมูลมีความถูกต้องสูงที่สุด

รัชพล กลัดชื่น (2560) การเปรียบเทียบประสิทธิภาพอัลกอริธึมและการคัดเลือกคุณลักษณะที่เหมาะสมเพื่อการทำนายผลสัมฤทธิ์ทางการเรียนของนักศึกษาระดับอาชีวศึกษา โดยมีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของอัลกอริธึมในการทำนายและคุณลักษณะที่มีต่อผลสัมฤทธิ์ทางการเรียนของนักศึกษาระดับอาชีวศึกษา โดยทำการศึกษาข้อมูลนักศึกษาระดับประกาศนียบัตรวิชาชีพ จำนวน 5,100 ระเบียน ของปีการศึกษา 2550 - 2559 จำนวน 9 สาขาวิชา 27 คุณลักษณะ โดยใช้เทคนิคการจำแนกข้อมูล 3 เทคนิค ได้แก่ 1) Decision Tree : J48graft 2) Naïve Bayes และ 3) Rule Induction มาทำการประเมินเปรียบเทียบประสิทธิภาพของตัวแบบการทำนายใช้คุณลักษณะทั้งหมดกับการเลือกคุณลักษณะแบบ forward select มาทดสอบประสิทธิภาพด้วยวิธี 10-fold cross validation ผลการศึกษาพบว่าเทคนิค Decision Tree : J48graft ด้วยวิธีการเลือกคุณลักษณะแบบ Forward Selection และ การเลือกคุณลักษณะทั้งหมด มีค่าความถูกต้องเท่ากับ 83.08% และ 81.71% และเมื่อทดสอบด้วยวิธี T-Test พบว่าการทดสอบทั้งสองแบบมีความ

แตกต่างกันที่ระดับ .05 จากผลการเปรียบเทียบประสิทธิภาพแล้วนำเทคนิค Decision Tree : J48graft ไปใช้ในการทำนายผลสัมฤทธิ์ทางการเรียน

พิชัย ระเวงวัน (2560) ได้ทำการศึกษาเรื่อง โมเดลเพื่อการพยากรณ์สถานภาพทางการศึกษาของนักศึกษาโดยมีวัตถุประสงค์เพื่อ ศึกษาปัจจัยที่มีผลต่อสถานภาพการศึกษาของนักศึกษา เพื่อทำการสร้างโมเดลเพื่อพยากรณ์สถานภาพทางการศึกษา ของนักศึกษาในสถาบันอุดมศึกษาเอกชนที่ตั้งในภาคตะวันออกเฉียงเหนือ โดยการประยุกต์ใช้วิธีการเหมืองข้อมูลและได้รวมรวมข้อมูลของนักศึกษาระดับปริญญาตรีที่เข้าศึกษาระหว่างปีการศึกษา 2551 – 2553 จำนวนทั้งสิ้น 2,272 คน มาใช้ในการสร้างโมเดลเพื่อทำนายสถานภาพของนักศึกษา ด้วยเทคนิคต้นไม้ตัดสินใจเบย์เซี่ยนบีลิฟเน็ตเวิร์ค และโลจิสติกรีเกรสชั่น ผลการศึกษาพบว่าเมื่อทดสอบการพยากรณ์โมเดลที่สร้างขึ้นโดยใช้ข้อมูลนักศึกษาที่เข้าศึกษาในปี 2554 ผลการพยากรณ์ของโมเดลต้นไม้ตัดสินใจวิธีการเรียนรู้แบบเบย์เซี่ยนเน็ตเวิร์ค และวิธีการถดถอยโลจิสติค มีค่าความถูกต้อง 82.85 %, 78.50 % , 78.98 % ตามลำดับ โมเดลที่สร้างด้วยวิธีต้นไม้ตัดสินใจให้ค่าความถูกต้องดีที่สุด

วัชรีวรรณ จิตต์สกุล (2560) ได้ทำการศึกษาเรื่อง การวิเคราะห์การจำแนกข้อความด้วยการเปรียบเทียบความเสถียรของอัลกอริทึม โดยการวิจัยครั้งนี้ได้ศึกษาหาความเสถียรของอัลกอริธึมเพื่อการจำแนก ผลการศึกษาพบว่าต้นไม้ตัดสินใจโดยมีอัลกอริธึม Random Forest มีค่าความเสถียรการจำแนกได้มากกว่าอัลกอริธึมอื่น มีค่าเฉลี่ย ROC > 0.80 และผลต่างของค่ากลางความถูกต้องเท่ากับ 0 โดยการทดสอบวิธี 10-fold cross validation และวิธี Test Dataset สามารถนำไปประยุกต์ใช้การสนับสนุนการตัดสินใจหรือการแนะนำได้

Korivi (2016) ได้ทำการศึกษาเรื่อง “Identifying Poverty-Driven Need By Augmenting Census And Community Survey Data” งานวิจัยนี้มีวัตถุประสงค์เป้าหมายหลักคือการสร้างแบบจำลองเพื่อประเมินความต้องการพื้นฐานตามตัวแปรทางเศรษฐกิจและสังคมโดยรวมที่บ่งชี้ถึงความยากจนแบบสัมบูรณ์และแบบสัมพัทธ์ โดยชุดข้อมูลที่ใช้ในงานวิจัยนำมาจากชุดข้อมูล American Community Survey (ACS) มาสร้างแบบจำลองเพื่อประเมินความต้องการและระบุระดับความยากจน โดยขั้นตอนเตรียมข้อมูลมีการทำ Missing Values, Feature Selection จากนั้นทำการเปรียบเทียบประสิทธิภาพของอัลกอริธึมต่าง ๆ คือ Random Forest, Linear Support Vector Machine และ Logistic Regression โดยใช้หลักการ Confusion Matrix เพื่อเปรียบเทียบประสิทธิภาพ ผลการศึกษาพบว่าอัลกอริทึม Random Forest มีประสิทธิภาพสูงที่สุด

Rincon (2019) ได้ทำการศึกษาเรื่อง “Quarterly Multidimensional Poverty Predictions in Mexico using Machine Learning Algorithms” งานวิจัยนี้มีวัตถุประสงค์คือสร้างแบบจำลองเพื่อประเมินอัตราความยากจนในประเทศเม็กซิโกเป็นรายไตรมาส โดยชุดข้อมูลที่ใช้ในงานวิจัยนำมา

จากชุดข้อมูลรายได้ของครัวเรือนและการใช้จ่าย ตั้งแต่ปี 2008 ถึงปี 2016 ของประเทศเม็กซิโก โดยการจำแนกว่าครอบครัวนี้จนหรือไม่โดยการสร้างแบบจำลองตามกระบวนการ Machine Learning โดยใช้การเปรียบเทียบแต่ละอัลกอริทึม 3 แบบคือ Linear Discriminant Analysis (LDA), Random Forest (RF) and Support Vector Machines (SVM) และทำการวัดประสิทธิภาพด้วยวิธี 10-Fold Cross-Validationและ Confusion Matrix ผลการศึกษาพบว่าการเปรียบเทียบประสิทธิภาพของแบบจำลองดังกล่าวซึ่งค่า Accuracy, Recall, Specificity, Precision, NPV และ F1-Score ประสิทธิภาพโดยรวมของ Random Forest (RF) ได้อัตราเฉลี่ยดีที่สุดสำหรับชุดข้อมูล ตั้งแต่ปี 2008 ถึงปี 2016

ผู้วิจัยได้ศึกษาแนวคิด ทฤษฎี และผลงานจากงานวิจัยข้างต้น เป็นการศึกษาที่เกี่ยวข้องกับการศึกษาและเทคนิคเหมืองข้อมูลมาประยุกต์ใช้ โดยการให้ความสำคัญกับการคัดเลือกคุณลักษณะและการเลือกใช้อัลกอริธึมแบบต่างๆ ผู้วิจัยจึงประยุกต์ใช้การนำเอาอัลกอริธึม C4.5(J48) 1) อัลกอริทึม C4.5(J48) 2) อัลกอริธึมต้นไม้สุ่ม และ 3) อัลกอริธึมป่าสุ่ม โดยเลือกใช้วิธีการคัดเลือกคุณลักษณะแบบไม่เลือกคุณลักษณะมาใช้ทดสอบโมเดลระบบพยากรณ์ผลสัมฤทธิ์ทางการเรียนเพื่อสนับสนุนการตัดสินใจการศึกษาต่อหลังจบชั้นมัธยมศึกษาปีที่ 3 โดยใช้โปรแกรม WEKA มาเป็นเครื่องมือในการวิเคราะห์ข้อมูลและทดสอบ ซึ่งงานวิจัยนี้จะใช้ตัววัดประสิทธิภาพได้แก่ ค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าครบถ้วน (Recall) และค่าประสิทธิภาพโดยรวม (F-Measure) เพื่อหาต้นแบบพยากรณ์ที่ดีที่สุดเพื่อนำไปใช้ในการพยากรณ์ผลสัมฤทธิ์ทางการเรียนเพื่อสนับสนุนการตัดสินใจการศึกษาต่อหลังจบชั้นมัธยมศึกษาตอนต้นในการเลือกศึกษาต่อในระดับสามัญศึกษา หรือระดับอาชีวศึกษาต่อไปผ่านทางเว็บไซต์ต่อไป

บทที่ 3
วิธีดำเนินการวิจัย

การดำเนินงานวิจัยระบบพยากรณ์เพื่อสนับสนุนการตัดสินใจศึกษาต่อหลังสำเร็จมัธยมศึกษาตอนต้น โดยใช้อัลกอริธึมป่าสุ่ม โดยการใช้วิธีการคัดเลือกคุณลักษณะแบบไม่เลือกคุณลักษณะ และการคัดเลือกแบบวิธี InfoGainAttributeEval แล้วทำการเปรียบเทียบสร้างตัวแบบพยากรณ์ด้วย 3 อัลกอริธึม คือ 1) อัลกอริธึม C4.5(J48) 2) อัลกอริธึมต้นไม้สุ่ม และ 3) อัลกอริธึมป่าสุ่ม ผู้วิจัยได้กำหนดรายละเอียดการดำเนินงาน ดังนี้

1. ชุดข้อมูล
2. เครื่องมือที่ใช้ในการวิจัย
3. ขั้นตอนการดำเนินการวิจัย

1. ชุดข้อมูล

ชุดข้อมูลที่จะนำมาวิเคราะห์เป็นชุดข้อมูลผลสัมฤทธิ์ทางการเรียนของนักเรียนระดับชั้นมัธยมศึกษาปีที่ 3 ในสังกัดสำนักงานเขตพื้นที่การศึกษาประถมศึกษานครสวรรค์ เขต 2 ระหว่างปีการศึกษา 2558 – 2560 ซึ่งจะจัดอยู่ในรูปของไฟล์ .csv ประกอบไปด้วยคุณลักษณะ 21 คุณลักษณะ ได้แก่ 1) รหัสโรงเรียน 2) เลขบัตรประชาชน 3) รหัสนักเรียน 4) เพศ 5) ชื่อนักเรียน 6) วันเกิด 7) หมู่โลหิต 8) ชื่อบิดา 9) ชื่อมารดา 10) รายได้ครอบครัวต่อเดือน 11) ที่อยู่ 12) เกรดเฉลี่ยกลุ่มสาระภาษาไทย 13) เกรดเฉลี่ยกลุ่มสาระคณิตศาสตร์ 14) เกรดเฉลี่ยกลุ่มสาระวิทยาศาสตร์ 15) เกรดเฉลี่ยกลุ่มสาระภาษาต่างประเทศ 16) เกรดเฉลี่ยกลุ่มสาระสังคมศึกษาศาสนาและวัฒนธรรม 17) เกรดเฉลี่ยกลุ่มสาระสุขศึกษาพลศึกษา 18) เกรดเฉลี่ยกลุ่มสาระศิลปะ 19) เกรดเฉลี่ยกลุ่มสาระการงานอาชีพและเทคโนโลยี 20) กิจกรรมพัฒนาผู้เรียน และ 21) ผลการเรียนเฉลี่ยตลอดหลักสูตรการศึกษา และเป้าหมาย จำนวน 2 เป้าหมาย หรือผลลัพธ์ (Class) คือ 1) สายสามัญศึกษา และ 2) สายอาชีวศึกษา

2. เครื่องมือที่ใช้ในการวิจัย

ในการดำเนินการวิจัยครั้งนี้เลือกใช้เครื่องมือและซอฟต์แวร์ที่ใช้ในการประมวลผล คือ โปรแกรม WEKA เป็นซอฟต์แวร์ที่ใช้ในการทำดาต้าไมน์นิง ที่นำมาใช้ในการวิเคราะห์ข้อมูล โดยมีเครื่องมือที่ในการสร้างโมเดล และมีอัลกอริธึมหรือเทคนิคการทำดาต้าไมน์นิงให้ใช้งานเป็นจำนวนมาก

2.1 ฮาร์ดแวร์

เครื่องคอมพิวเตอร์ รายละเอียดคุณสมบัติ

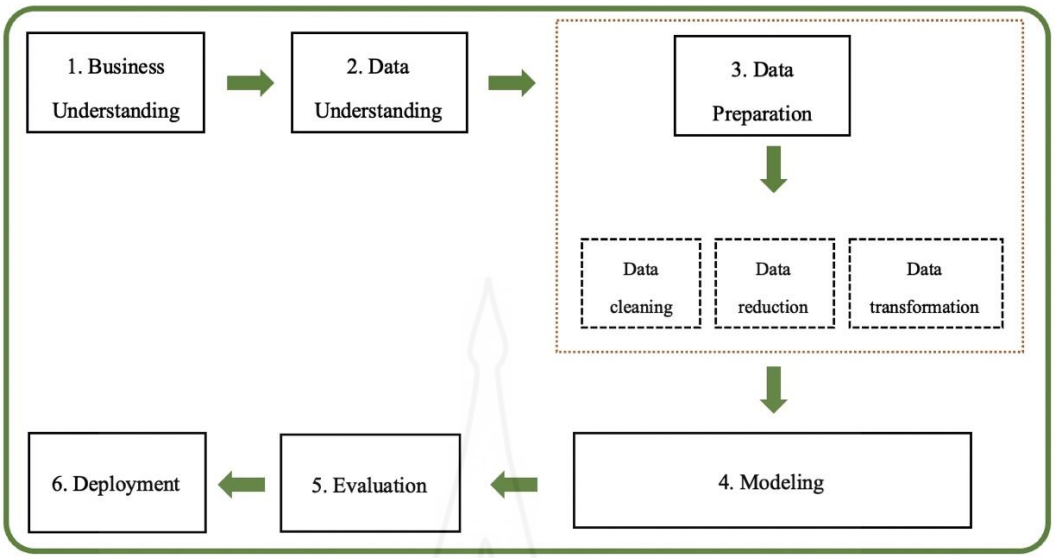
1) หน่วยประมวลกลาง (CPU) ความเร็ว 2.4 GHz
2) หน่วยความจำหลัก (RAM) ความจุ 8 GB
3) อุปกรณ์บันทึกข้อมูล (Hard disk) ความจุ 500 GB

2.2 ซอฟต์แวร์

1) ระบบปฏิบัติการไมโครซอฟท์วินโดวส์ Windows 10
2) โปรแกรมเวก้า (WEKA)
3) โปรแกรม Google Chrome 60 ขึ้นไป
4) โปรแกรม Appserv จำลอง web server
5) โปรแกรมสร้างเว็บเพจ Dreamweaver CS6

3. ขั้นตอนการดำเนินการวิจัย

ผู้วิจัยได้ดำเนินงานวิจัยสำหรับระบบพยากรณ์เพื่อสนับสนุนการตัดสินใจการศึกษาต่อหลังจบชั้นมัธยมศึกษาตอนต้น ทำการจัดเตรียมข้อมูลที่ใช้ในการสร้างแบบจำลอง (Model) และทดสอบอัลกอริธึมตามแนวทางกระบวนการทำเหมืองข้อมูล (Crisp-DM) จะประกอบด้วย 6 ขั้นตอน ดังภาพที่ 3.1



ภาพที่ 3.1 แสดงขั้นตอนการการดำเนินการวิจัย

3.1 ทำความเข้าใจกับโจทย์ที่ต้องทำ (Business Understanding)

ผู้วิจัยได้ทำการศึกษาปัญหาแนวทางเป้าหมายของการศึกษาของนักเรียนโรงเรียนขยายโอกาสทางการศึกษาในสังกัดสำนักงานเขตพื้นที่การศึกษาประถมศึกษานครสวรรค์เขต 2 ที่สำเร็จการศึกษาชั้นมัธยมศึกษาปีที่ 3

3.2 ทำความเข้าใจกับข้อมูลที่ใช้ในการวิจัย (Data Understanding)

ข้อมูลผลสัมฤทธิ์ทางการเรียนจำนวน 6 ภาคเรียน ระหว่างปีการศึกษา 2558 - 2560 ของนักเรียนชั้นมัธยมศึกษาปีที่ 3 ในสังกัดสำนักงานเขตพื้นที่การศึกษาประถมศึกษานครสวรรค์ เขต 2 จังหวัดนครสวรรค์ จำนวน 1,160 คน ในรูปแบบของไฟล์ Excel (ข้อมูลบริหารจัดการสถานศึกษาจากสำนักงานเขตพื้นที่การศึกษาประถมศึกษานครสวรรค์ เขต 2 จังหวัดนครสวรรค์ ปีการศึกษา 2560)

3.3 การเตรียมข้อมูล (Data Preparation)

นำข้อมูลที่ได้อยู่ในรูปแบบของไฟล์ Excel จำนวน 1,160 เรคคอร์ด ซึ่งยังไม่สามารถนำมาวิเคราะห์ด้วยวิธีการทำเหมืองข้อมูลได้ ผู้วิจัยได้มีการเตรียมข้อมูลก่อนนำไปใช้ในการทำเหมืองข้อมูลเพื่อให้ได้ผลการทดลองที่มีประสิทธิภาพ โดยขั้นตอนนี้ประกอบด้วยขั้นตอน 3 ขั้นตอนคือ การกรองข้อมูล การแปลงข้อมูล และการลดขนาดข้อมูล ดังภาพที่ 3.2

ภาพที่ 3.2 ขั้นตอกนการเตรียมข้อมูล

3.3.1 การกลั่นกรองข้อมูล (data cleaning) ข้อมูลที่จัดเก็บอาจมีความผิดปกติต่างๆ ข้อมูลบางคุณลักษณะขาดหายไป เช่น ในบางโรงเรียนมีข้อมูลผลสัมฤทธิ์ของนักเรียนไม่ครบทั้ง 6 เทอม ซึ่งอาจเกิดจากที่นักเรียนย้ายสถานศึกษาระหว่างเรียน เป็นต้น ผู้วิจัยได้ทำการ Data cleaning เพื่อเป็นการตรวจสอบความถูกต้องของข้อมูล หากเกิดข้อผิดพลาด ข้อมูลไม่สอดคล้อง ข้อมูลขาดหาย จะได้ทำการปรับค่าข้อมูลเพื่อนำไปใช้ในการวิเคราะห์ศึกษาด้วยเทคนิคเหมืองข้อมูล ในการตรวจสอบข้อมูล data set นี้จะใช้ด้วยโปรแกรม WEKA ทำการลบค่าของข้อมูลที่ขาดหายออกไป (Remove with value) เหลือชุดข้อมูลที่นำไปใช้ทดสอบ จำนวน 1,030 เรคคอร์ด

3.3.2 การลดขนาดข้อมูล (data reduction) นำข้อมูลปี 2558 – 2560 มาทำการลดขนาดเนื่องจากข้อมูลมีปริมาณมีคุณลักษณะ มากไป เพื่อให้ได้ข้อมูลที่มีความสมบูรณ์ถูกต้องสะดวกต่อการนำข้อมูลไปดำเนินการวิเคราะห์ โดยลดจำนวนของคุณลักษณะมีที่ไม่เกี่ยวข้องกับข้อมูลนักเรียนและผลสัมฤทธิ์ของนักเรียนอยู่บางตัวออกไป เนื่องจากจะใช้ข้อมูลผลสัมฤทธิ์ทางการเรียนและข้อมูลที่เกี่ยวข้องกับนักเรียนในการเรียน บางคุณลักษณะไม่มีส่วนเกี่ยวข้องกับผลสัมฤทธิ์ทางการเรียนจึงตัดออกไป ได้แก่ 1) รหัสโรงเรียน 2) เลขบัตรประชาชน 3) รหัสนักเรียน 4) ชื่อนักเรียน 5) วันเกิด 6) หมู่โลหิต 7) ชื่อบิดา 8) ชื่อมารดา 9) ที่อยู่ 10) กิจกรรมพัฒนาผู้เรียน นำไฟล์ข้อมูลให้อยู่ในรูปแบบของไฟล์ .arff แล้วจึงนำไปสู่โปรแกรม WEKA โดยการคัดเลือกแบบ feature selection เพื่อลดจำนวนคุณลักษณะที่ไม่เกี่ยวข้องออก ทำให้ได้คุณลักษณะที่มีความสำคัญเกี่ยวข้องกับเป้าหมายเหมาะสมที่จะนำไปสร้างแบบจำลอง ทำให้การประมวลผลมีความรวดเร็วและถูกต้องแม่นยำ

3.3.3 การแปลงข้อมูล (data transformation) ทำการแปลงข้อมูลที่มีความหมายใกล้เคียงกันมาอยู่ในระดับเดียวกันเป็นค่าไม่ต่อเนื่อง (Discretization) เป็นการแบ่งออกเป็นกลุ่มย่อยของข้อมูลที่มีรายละเอียดต่ำลงโดยแบ่งค่าของข้อมูลออกเป็นช่วงย่อยๆ จะเป็นการช่วยลดการประมวลผลในการทำเหมืองข้อมูล และลดการซ้ำซ้อนการกระจายของข้อมูล เหลือเพียง

จำนวนข้อมูล 11 คุณลักษณะ และเป้าหมาย (Class) จำนวน 2 เป้าหมาย คือ สายสามัญศึกษา (Bec) สายอาชีวศึกษา (Vec) เพื่อสามารถนำไปใช้ทดสอบกับอัลกอริธึมที่รองรับ

3.4 การสร้างตัวแบบพยากรณ์ (Modeling)

การสร้างตัวแบบพยากรณ์ (Modeling) ของระบบพยากรณ์เพื่อสนับสนุนการตัดสินใจศึกษาต่อหลังสำเร็จมัธยมศึกษาตอนต้น โดยใช้อัลกอริธึมป่าสุ่ม ในสังกัดสำนักงานเขตพื้นที่การศึกษาประถมศึกษานครสวรรค์ เขต 2 จังหวัดนครสวรรค์ ระหว่างปีการศึกษา 2558 - 2560 จำนวน 1,030 คน 11 คุณลักษณะ จากขั้นตอนการเตรียมข้อมูล ใช้วิธีการคัดเลือกคุณลักษณะแบบไม่เลือกคุณลักษณะ และการคัดเลือกแบบวิธี InfoGainAttributeEval ทดสอบประสิทธิภาพของแบบจำลอง (Model) แบบ 10-fold cross validation ทำการเปรียบเทียบสร้างตัวแบบพยากรณ์ด้วย 3 อัลกอริธึม คือ 1) อัลกอริธึม C4.5(J48) 2) อัลกอริธึมต้นไม้สุ่ม Random Tree และ 3) อัลกอริธึมป่าสุ่ม Random Forest เพื่อวิเคราะห์วัดประสิทธิภาพในการจำแนกของตัวแบบพยากรณ์ที่ถูกต้องที่สุดจากค่าความถูกต้องของการจำแนกข้อมูล (Accuracy) ค่าความแม่นยำ (Precision) ค่าครบถ้วน (Recall) และค่าประสิทธิภาพโดยรวม (F-Measure)

3.5 การประเมินผล (Evaluation)

ข้อมูลที่ได้จากข้อมูลผลสัมฤทธิ์ทางการเรียนของนักเรียนชั้นมัธยมศึกษาปีที่ 3 ในสังกัดสำนักงานเขตพื้นที่การศึกษาประถมศึกษานครสวรรค์ เขต 2 จังหวัดนครสวรรค์ ระหว่างปีการศึกษา 2558 - 2560 จำนวน 1,030 เรคคอร์ด ผ่านการเตรียมข้อมูลเพื่อนำไปสร้างแบบจำลองทดสอบประสิทธิภาพของแบบจำลอง (Model) แบบ 10-fold cross validation เป็นการแบ่งข้อมูลที่ใช้ในการทดสอบออกเป็นชุดๆ ละเท่ากันเพื่อสร้างและทดสอบแบบจำลองหาความน่าเชื่อถือ โดยที่ข้อมูลแต่ละชุดจะมีข้อมูลส่วนหนึ่งใช้เป็นตัวทดสอบประสิทธิภาพของแบบจำลอง ผู้วิจัยใช้การประเมินประสิทธิภาพของตัวแบบพยากรณ์จากค่าความถูกต้องของการจำแนกข้อมูล (Accuracy) ค่าความแม่นยำ (Precision) ค่าครบถ้วน (Recall) และค่าประสิทธิภาพโดยรวม (F-Measure) ของผลต่างระหว่างอัลกอริธึม

3.6 การนำไปใช้งาน (Deployment)

เมื่อได้แบบจำลองพยากรณ์จากการวิเคราะห์วัดประสิทธิภาพแล้วนำมาพัฒนาเป็นระบบพยากรณ์เพื่อสนับสนุนการตัดสินใจศึกษาต่อหลังสำเร็จมัธยมศึกษาตอนต้น โดยใช้อัลกอริธึมป่าสุ่ม ใช้โปรแกรม Dreamweaver CS6 สร้างพัฒนาเป็นเว็บไซต์และโปรแกรม Appserv จำลองเป็น web server ในการใช้งานผ่านทางเว็บไซต์ เพื่อเป็นแนวทางในการศึกษาต่อในระดับสายสามัญศึกษาและสายอาชีวศึกษา

บทที่ 4
ผลการดำเนินงานวิจัย

งานวิจัยระบบพยากรณ์เพื่อสนับสนุนการตัดสินใจศึกษาต่อหลังสำเร็จมัธยมศึกษาตอนต้น โดยใช้อัลกอริธึมป่าสุ่ม ในสังกัดสำนักงานเขตพื้นที่การศึกษาประถมศึกษานครสวรรค์เขต 2 ปีการศึกษา 2558 - 2560 ผู้วิจัยได้นำมาวิเคราะห์วัดประสิทธิภาพพัฒนาเป็นระบบพยากรณ์ผลสัมฤทธิ์ทางการเรียน ตามแนวทางกระบวนการทำเหมืองข้อมูล (Crisp-DM) จะประกอบด้วย 6 ขั้นตอน ได้แก่

ตอนที่ 1 ทำความเข้าใจกับโจทย์ที่ต้องทำ (Business Understanding)
ตอนที่ 2 ทำความเข้าใจกับข้อมูลที่ใช้ในการวิจัย (Data Understanding)
ตอนที่ 3 การเตรียมข้อมูล (Data Preparation)
ตอนที่ 4 การสร้างตัวแบบพยากรณ์ (Modeling)
ตอนที่ 5 การประเมินผล (Evaluation)
ตอนที่ 6 การนำไปใช้งาน (Deployment)

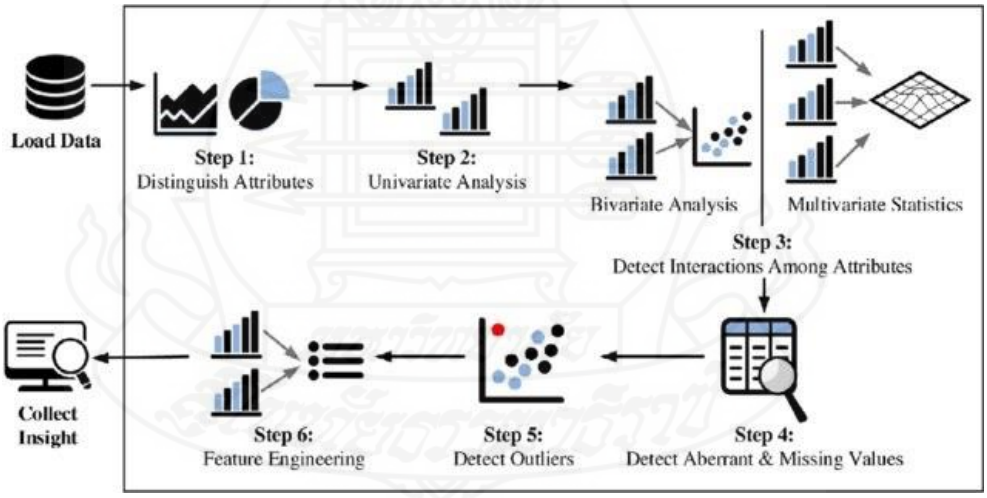
ตอนที่ 1 ทำความเข้าใจกับโจทย์ที่ต้องทำ (Business Understanding)

ทำการศึกษางานวิจัยที่เกี่ยวข้องและรวบรวมปัญหาเป้าหมายของการศึกษาต่อของนักเรียนโรงเรียนขยายโอกาสทางการศึกษาในสังกัดสำนักงานเขตพื้นที่การศึกษาประถมศึกษานครสวรรค์เขต 2 ที่สำเร็จการศึกษาชั้นมัธยมศึกษาตอนต้น ซึ่งพบว่าโรงเรียนขาดบุคลากรครูแนะแนว ซึ่งเป็นปัญหาต้นทางของเรื่องนี้ นักเรียนจำนวนมากขาดการแนะแนวด้านการศึกษาต่อ จึงเป็นส่วนหนึ่งที่ทำให้นักเรียนประสบปัญหาเรื่องการค้นพบตัวเองช้าว่าอยากเรียนอะไร หรือปรับตัวไม่ทันกับการเรียนที่เลือกไปแล้ว ทำให้ต้องเสียเวลาไปกับการเรียนที่ไม่มีเป้าหมายในการศึกษาต่อของนักเรียนที่ผ่านมาส่วนใหญ่พบว่าเลือกศึกษาต่อตามเพื่อน ตามคำแนะนำของผู้ปกครอง ทำให้ไม่ประสบความสำเร็จในการเรียน มีผลการเรียนต่ำ เกิดปัญหาตามมาคือการเรียนไม่จบหลักสูตร เปลี่ยนสายการศึกษาต่อ และนักเรียนบางคนต้องเลิกศึกษาต่อทั้งที่ยังไม่จบการศึกษา หากมีเครื่องมือที่ช่วยแนะนำแนวทางสนับสนุนการตัดสินใจให้กับนักเรียนได้จะทำให้การวางแผนใน

การเลือกเรียนประเภทการศึกษาต่อในระดับสายสามัญศึกษาและสายอาชีวศึกษาหลังจบชั้นมัธยมศึกษาตอนต้น ที่มีความชัดเจน

ตอนที่ 2 ทำความเข้าใจกับข้อมูลที่ใช้ในการวิจัย (Data Understanding)

ทำการรวบรวมข้อมูลผลสัมฤทธิ์ทางการศึกษาของนักเรียนโรงเรียนขยายโอกาสทางการศึกษาในสังกัดสำนักงานเขตพื้นที่การศึกษาประถมศึกษานครสวรรค์เขต 2 ที่สำเร็จการศึกษาชั้นมัธยมศึกษาตอนต้น ตั้งแต่ปีการศึกษา 2558 - 2560 ใช้กลุ่มตัวอย่างข้อมูลของผลสัมฤทธิ์ทางการเรียน 8 กลุ่มสาระวิชาย้อนหลัง 6 ภาคเรียน (2558 - 2560) ในการศึกษาต่อประเภทสามัญศึกษา และประเภทอาชีวศึกษา เพื่อนำไปวิเคราะห์ด้วยเทคนิคเหมืองข้อมูล จำนวน 1,160 เรคคอร์ด จำนวน 21 คุณลักษณะ และเป้าหมายจำนวน 2 เป้าหมาย หรือในส่วนผลลัพธ์ (Class) คือ 1) สายสามัญศึกษา และ 2) สายอาชีวศึกษา โดยใช้กระบวนการตรวจสอบ สำรวจข้อมูลเบื้องต้น Exploratory Data Analysis (EDA) เพื่อค้นหารูปแบบและข้อมูลเชิงลึกก่อนการนำไปใช้งาน ประกอบด้วย 6 ขั้นตอน ดังภาพที่ 4.1



ภาพที่ 4.1 กระบวนการตรวจสอบ สำรวจข้อมูลเบื้องต้น Exploratory Data Analysis
ที่มา : https://www.glurgeek.com/education/datawranglingwithpandas

จากภาพที่ 4.1 คือ 6 ขั้นตอนกระบวนการตรวจสอบ สำรวจข้อมูลเบื้องต้นซึ่งประกอบด้วย

1) การเลือกข้อมูล (Distinguish Attributes) การนำชุดข้อมูลมาทำเป็นคอลัมน์ เพื่อให้เห็นถึงความแตกต่างของข้อมูลแยกออกเป็นคุณลักษณะต่างๆ เช่น เพศ วิชา เป็นต้น

2) การวิเคราะห์ข้อมูลตัวแปรเดียว (Univariate Analysis) เป็นการวิเคราะห์เชิงสถิติ ที่ทำให้เห็นความสำคัญของแต่ละคุณลักษณะ เช่น ค่าเฉลี่ย ผลรวม เป็นต้น

3) การวิเคราะห์ข้อมูลมากกว่า 1 ตัวแปร (Bi/Multivariate Analysis) การวิเคราะห์ข้อมูลมากกว่า 1 ตัวแปร เพื่อให้เห็นถึงความสัมพันธ์ขั้นต้น เช่น การหาความสัมพันธ์

4) ค้นหาสิ่งผิดพลาดในชุดข้อมูล (Detect Aberrant and Missing Values) เป็นการหาสิ่งผิดปกติในชุดข้อมูล และข้อมูลที่ขาดหายไป

5) การวิเคราะห์ข้อมูลที่ผิดปกติ (Detect Outlier) การวิเคราะห์หาข้อมูลที่ผิดปกติไปจากค่ากลาง

6) การสร้างคุณลักษณะที่จะนำไปวิเคราะห์เชิงลึก (Feature Engineering) การสร้างตัวแปรที่จะนำไปวิเคราะห์เชิงลึก เช่น การแบ่งช่วงผลการเรียนเป็น 3 ช่วง เป็นต้น

ข้อมูลที่นำมาใช้ในการศึกษา ซึ่งอยู่ในรูปแบบ Excel ดังภาพที่ 4.2

ลำดับ	รหัสโรงเรีย	รหัสนักเรียน	2558 ชั้นมัธยมศึกษาปีที่	ท21101 ภ	ค21101 คณ	ว21101 วิท	ส21101 สั
1	60020008	'1515	NULL	NULL	3	4	2.5	2.5
2	60020008	'1516	NULL	NULL	2	2	1.5	2
3	60020008	'1518	NULL	NULL	2.5	2.5	2.5	3.5
4	60020008	'1519	NULL	NULL	3	3	2.5	2.5
5	60020008	'1520	NULL	NULL	2.5	2.5	3	2.5
6	60020008	'1521	NULL	NULL	1.5	2	1	2
7	60020008	'1523	NULL	NULL	2	2	1.5	2
8	60020008	'1572	NULL	NULL	3.5	2.5	4	4
9	60020008	'1632	NULL	NULL	3	2	3	4
10	60020008	'1636	NULL	NULL	3	3	3	3.5
11	60020008	'1832	NULL	NULL	3	1.5	2	2.5
12	60020008	'1930	NULL	NULL	2.5	2	2	3
13	60020008	'1942	NULL	NULL	2.5	1.5	2	1.5
14	60020008	'1944	NULL	NULL	2	2	1.5	2.5
15	60020008	'1946	NULL	NULL	1.5	2	1.5	2.5
16	60020008	'1996	NULL	NULL	1.5	1.5	2	2
17	60020008	'1997	NULL	NULL	1.5	1.5	2	2.5
18	60020008	'2085	NULL	NULL	3.5	1.5	3	2.5
19	60020008	'1528	NULL	NULL	4	3.5	4	4

ภาพที่ 4.2 แสดงข้อมูลที่นำมาใช้ในการศึกษา

ตอนที่ 3 การเตรียมข้อมูล (Data Preparation)

นำชุดข้อมูลที่ได้จากการเก็บรวบรวมมาทำการเตรียมข้อมูลผลสัมฤทธิ์ทางการเรียน จำนวน 6 ภาคเรียน (2558 - 2560) จำนวน 1,160 เรคคอร์ด จำนวน 21 คุณลักษณะ และเป้าหมาย จำนวน 2 เป้าหมาย หรือในส่วนผลลัพธ์ (Class) คือ สายสามัญศึกษา (Bec) สายอาชีวศึกษา (Vec) ซึ่งข้อมูลดังกล่าวยังไม่สามารถนำไปวิเคราะห์ได้ อาจมีส่วนของข้อมูลที่ผิดปกติ เพื่อลดความผิดพลาดผู้วิจัยได้ดำเนินการดังนี้

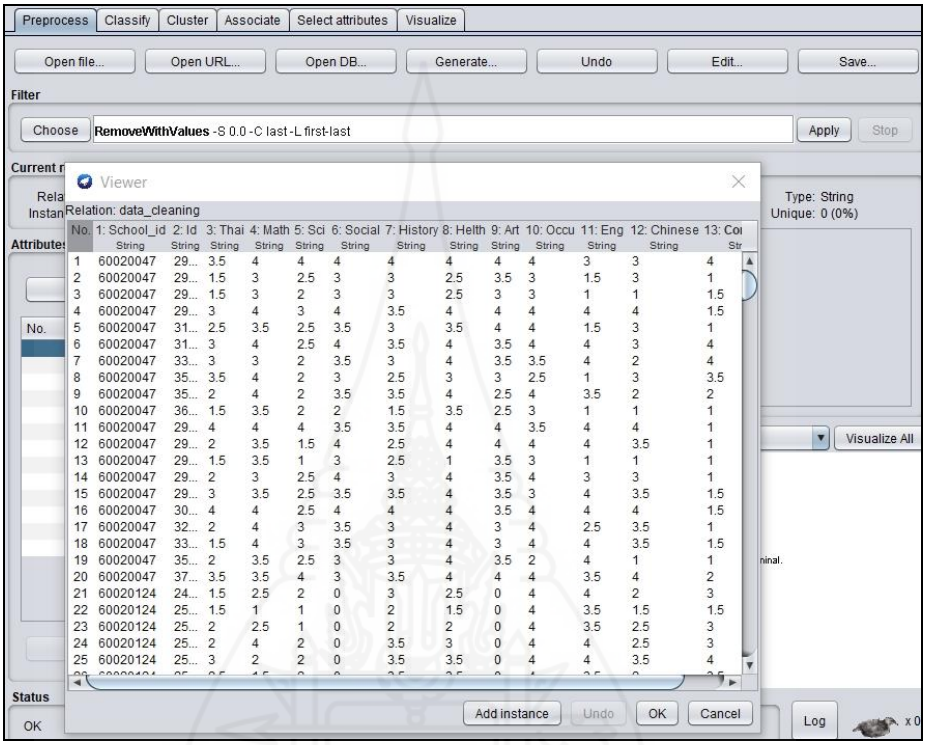
3.1 การกลั่นกรองข้อมูล (data cleaning)

เป็นการตรวจสอบความถูกต้องข้อมูลที่จัดเก็บอาจมีความผิดปกติต่างๆ ข้อมูลบางคุณลักษณะขาดหายไป เช่น ในบางโรงเรียนมีข้อมูลผลสัมฤทธิ์ของนักเรียนไม่ครบทั้ง 6 ภาคเรียน ซึ่งอาจเกิดจากที่นักเรียนย้ายสถานศึกษาระหว่างเรียน เป็นต้น ผู้วิจัยได้ทำการ Data cleaning เพื่อเป็นการตรวจสอบความถูกต้องของข้อมูล หากเกิดข้อผิดพลาด ข้อมูลไม่สอดคล้อง ข้อมูลขาดหายในบางคุณลักษณะ เช่น ผลการเรียนเป็น ร หรือ ไม่มีคะแนนผลการเรียน ซึ่งอาจเกิดจากกรณีที่นักเรียนย้ายสถานศึกษาไปยังเขตพื้นที่การศึกษาอื่นหรือการบันทึกข้อมูลลงในระบบที่ไม่ถูกต้อง เป็นต้น ทำการลบข้อมูลที่ผิดปกติออกไป เพื่อนำข้อมูลที่ถูกต้องไปใช้ในการวิเคราะห์ศึกษาด้วยเทคนิคเหมืองข้อมูล ดังภาพที่ 4.3

รหัสโรงเรีย	รหัสนักเรียน	2558	ชั้นมัธยมศึกษาปีที่	ท21101	ภ	ค21101	ค	ว21101	วิท	ส21101	สั	ส21102	ปร	พ21101	สุ
60020047	'2902	NULL	NULL	3.5		4		4		4		4		4	
60020047	'2904	NULL	NULL	1.5		3		2.5		3		3		2.5	
60020047	'2918	NULL	NULL	1.5		3		2		3		3		2.5	
60020047	'2961	NULL	NULL	3		4		3		4		3.5		4	
60020047	'3109	NULL	NULL	2.5		3.5		2.5		3.5		3		3.5	
60020047	'3146	NULL	NULL	3		4		2.5		4		3.5		4	
60020047	'3326	NULL	NULL	3		3		2		3.5		3		4	
60020047	'3589	NULL	NULL	3.5		4		2		3		2.5		3	
60020047	'3592	NULL	NULL	2		4		2		3.5		3.5		4	
60020047	'3603	NULL	NULL	1.5		3.5		2				1.5		3.5	
60020047	'2907	NULL	NULL	4		4		4		3.5		3.5		4	
60020047	'2908	NULL	NULL	2		3.5		1.5		4		2.5		4	
60020047	'2910	NULL	NULL	1.5		3.5		ร		3		2.5		1	
60020047	'2911	NULL	NULL	2		3		2.5		4		3		4	
60020047	'2912	NULL	NULL	3		3.5		2.5		3.5		3.5		4	
60020047	'3055	NULL	NULL	4		4		2.5		4		4		4	
60020047	'3277	NULL	NULL	2		4		3		3.5		3		4	
60020047	'3322	NULL	NULL			4		3		3.5		3		4	
60020047	'3585	NULL	NULL	2		3.5		2.5		3		3		4	
60020047	'3723	NULL	NULL	3.5		3.5		4		3		3.5		4	

ภาพที่ 4.3 การตรวจสอบความถูกต้องของข้อมูล

ในการตรวจสอบข้อมูล data set นี้จะใช้ด้วยโปรแกรม WEKA ทำการลบค่าว่างในส่วนของช่องที่ไม่มีผลการเรียนของนักเรียนในบางรายวิชาในชุดข้อมูลของคุณลักษณะที่ขาดหายออกไป (Remove Missing Value) คงเหลือข้อมูล จำนวน 1,030 เรคคอร์ด ดังภาพที่ 4.4



ภาพที่ 4.4 ตัดข้อมูลที่มีความผิดปกติและข้อมูลขาดหายไป

3.2 การลดขนาดข้อมูล (data reduction)

นำข้อชุดข้อมูลปี 2558 – 2560 มาทำการลดขนาดเนื่องจากข้อมูลมีปริมาณคุณลักษณะมากเกินไป เพื่อให้ได้ข้อมูลที่มีความสมบูรณ์ถูกต้อง สะดวกต่อการนำข้อมูลไปดำเนินการวิเคราะห์ โดยลดจำนวนของคุณลักษณะมีที่ไม่เกี่ยวข้องกับข้อมูลนักเรียนและผลสัมฤทธิ์ของนักเรียนอยู่บางตัวออก จำนวน 10 คุณลักษณะ ได้แก่ 1) รหัสโรงเรียน 2) เลขบัตรประชาชน 3) รหัสนักเรียน 4) ชื่อนักเรียน 5) วันเกิด 6) หมู่โลหิต 7) ชื่อบิดา 8) ชื่อมารดา 9) ที่อยู่ และ 10) กิจกรรมพัฒนาผู้เรียน คงเหลือคุณลักษณะที่จำเป็นในการนำไปใช้สร้างแบบจำลอง จำนวน 11 คุณลักษณะได้แก่ 1) เพศ 2) รายได้ครอบครัวต่อเดือน 3) เกรดเฉลี่ยกลุ่มสาระภาษาไทย 4) เกรดเฉลี่ยกลุ่มสาระคณิตศาสตร์ 5) เกรดเฉลี่ยกลุ่มสาระวิทยาศาสตร์ 6) เกรดเฉลี่ยกลุ่มสาระภาษาต่างประเทศ 7) เกรดเฉลี่ยกลุ่มสาระสังคมศึกษาศาสนาและวัฒนธรรม 8) เกรดเฉลี่ยกลุ่มสาระสุขศึกษาพลศึกษา

9) เกรดเฉลี่ยกลุ่มสาระศิลปะ 10) เกรดเฉลี่ยกลุ่มสาระการงานอาชีพและเทคโนโลยี และ 11) ผลการเรียนเฉลี่ยตลอดหลักสูตรการศึกษา ดังภาพที่ 4.5

Gender	Income	Thai	Math	Sci	Foreign	Social	Helth	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 4.5 การลดจำนวนของคุณลักษณะที่ไม่เกี่ยวข้องกับข้อมูลนักเรียนและผลสัมฤทธิ์

3.3 การแปลงข้อมูล (data transformation)

ทำการคำนวณหาค่าผลสัมฤทธิ์ทางการเรียนแต่ละกลุ่มสาระการเรียนรู้ของรายวิชาในแต่ละภาคเรียนด้วยโปรแกรม Microsoft Excel 2013 ดังภาพที่ 4.6 ซึ่งสามารถหาได้จาก ผลการเรียนทุกวิชาในกลุ่มสาระการเรียนรู้นั้นๆ ในรายวิชาพื้นฐาน คือ ผลการเรียน x จำนวนหน่วยกิต และนำผลรวมทั้งหมดมาหารด้วยจำนวนหน่วยกิตทั้งหมดของกลุ่มสาระการเรียนรู้

$$\text{ผลสัมฤทธิ์รายวิชา} = \text{ผลการเรียน X จำนวนหน่วยกิต} \quad (3)$$

$$\text{ผลสัมฤทธิ์ทุกวิชาในกลุ่มสาระการเรียนรู้} = \frac{\text{ผลสัมฤทธิ์รายวิชา} + \text{ผลสัมฤทธิ์รายวิชา} + ..n}{\text{จำนวนหน่วยกิตทั้งหมดของกลุ่มสาระการเรียนรู้}} \quad (4)$$

โดย n คือ จำนวนผลสัมฤทธิ์รายวิชาที่เรียนในภาคเรียนนั้นๆ

การคิดผลสัมฤทธิ์ประจำภาคเรียน GPA

$$\text{คะแนนผลสัมฤทธิ์ประจำภาคเรียน } GPA = \frac{\text{ค่าคะแนนที่ได้จากผลการเรียน}}{\text{จำนวนหน่วยกิตทั้งหมดในภาคเรียน}} \quad (5)$$

Thai	Math	Sci	Foreign	Social	Health	Art	Career	GPA
2.17	2.50	2.50	1.61	2.48	3.25	2.89	2.75	2.49
3.00	2.83	3.00	2.55	3.05	3.33	3.04	3.08	2.97
2.50	1.98	2.08	1.68	2.59	2.08	2.14	2.06	2.17
2.67	2.89	2.25	2.34	2.73	3.17	2.68	2.31	2.61
3.50	2.96	2.67	2.55	3.14	3.17	3.04	2.83	2.98
3.17	3.15	2.75	2.58	3.36	3.42	3.32	3.06	3.10
2.50	2.70	2.83	2.34	2.84	3.00	2.71	2.73	2.71
2.83	2.83	2.33	2.50	2.95	3.00	2.68	2.85	2.76
3.67	2.89	2.50	2.63	3.30	3.42	3.14	3.15	3.09
3.50	3.74	3.17	2.97	3.21	3.67	3.57	3.21	3.36
3.25	3.28	2.67	2.76	3.29	3.17	2.79	3.13	3.07
3.42	3.67	3.17	2.61	3.39	3.33	3.64	2.94	3.26
3.67	3.67	4.00	3.42	3.66	3.58	3.50	3.42	3.62
3.33	3.61	3.67	3.84	3.50	3.75	3.61	3.33	3.55
1.40	1.42	1.08	1.15	1.23	1.58	1.96	1.67	1.43
0.85	1.17	0.83	1.02	0.98	1.96	1.58	1.54	1.20
3.52	2.50	2.75	2.71	3.03	3.54	3.33	3.29	3.10

ภาพที่ 4.6 การคำนวณหาค่าผลสัมฤทธิ์แต่ละกลุ่มสาระการเรียนรู้

ทำการแปลงข้อมูลที่มีความหมายใกล้เคียงกันมาอยู่ในระดับเดียวกันเป็นค่าไม่ต่อเนื่อง (Discretization) เป็นการแบ่งออกเป็นกลุ่มย่อยของข้อมูลที่มีรายละเอียดต่ำลง โดยแบ่งค่าของข้อมูลออกเป็นช่วงย่อยๆ จะเป็นการช่วยลดการประมวลผลในการทำเหมืองข้อมูล และลดการซ้ำซ้อนการกระจายของข้อมูล เพื่อสามารถนำไปใช้ทดสอบกับอัลกอริธึมที่รองรับ ดังตารางที่ 4.1

ตารางที่ 4.1 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

ตารางที่ 4.1 (ต่อ)

คุณลักษณะ	รายละเอียด	การแปลงข้อมูล
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	เกรดเฉลี่ยกลุ่มสาระสังคมศึกษาศาสนาและวัฒนธรรม	High = ผลการเรียนเฉลี่ย 3.01 - 4.00 Medium = ผลการเรียนเฉลี่ย 2.01 - 3.00 Low = ผลการเรียนเฉลี่ย 1.00 - 2.00
Health	เกรดเฉลี่ยกลุ่มสาระสุขศึกษาพลศึกษา	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 ภาคเรียน	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

ในการแปลงข้อมูลที่จะนำไปใช้วิเคราะห์ ผู้วิจัยได้ศึกษางานวิจัยของ เสกสรร วิลัยลักษณ์ (2558:75) ในการใช้เทคนิคการทำเหมืองข้อมูลเพื่อพยากรณ์ผลการเรียนของนักเรียนโรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน ในการแปลงข้อมูลผลการเรียนเฉลี่ยมีความละเอียดออกเป็น 3 กลุ่ม ได้แก่

High คือ ผลการเรียนเฉลี่ยของนักเรียนตั้งแต่ 3.01 – 4.00

Medium คือ ผลการเรียนเฉลี่ยของนักเรียนตั้งแต่ 2.01 – 3.00

Low คือ ผลการเรียนเฉลี่ยของนักเรียนตั้งแต่ 1.00 – 2.00

เป็นการแปลงข้อมูลให้เหมาะสมเพื่อที่จะทำการวิเคราะห์ข้อมูลได้อย่างถูกต้องสำหรับการทำเหมืองข้อมูลด้วยโปรแกรม WEKA โดยใช้ฟังก์ชั่นในโปรแกรม Microsoft Excel 2013 ในการแปลงข้อมูลผลการเรียน นำไฟล์ข้อมูลให้อยู่ในรูปแบบของไฟล์ .arff แล้วจึงนำไปสู่โปรแกรม WEKA ดังภาพที่ 4.7

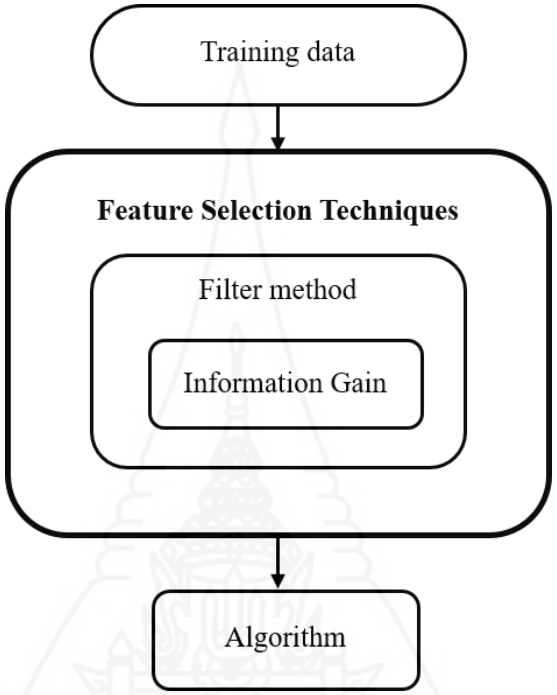
Gender	Income	Thai	Math	Sci	Foreign	Social	Helth	Art	Career	GPA	Course
Male	Low	Medium	Medium	Medium	Medium	Hight	Hight	Hight	Medium	Medium	Vec
Male	Low	Low	Low	Low	Low	Medium	Hight	Medium	Medium	Medium	Bec
Male	Low	Medium	Medium	Medium	Low	Hight	Hight	Hight	Hight	Medium	Vec
Male	Low	Medium	Medium	Medium	Low	Hight	Hight	Hight	Hight	Medium	Vec
Male	Low	Low	Low	Medium	Low	Medium	Medium	Medium	Medium	Medium	Bec
Male	Low	Low	Low	Low	Low	Medium	Medium	Medium	Medium	Medium	Bec
Male	Low	Medium	Low	Medium	Low	Medium	Medium	Medium	Medium	Medium	Bec
Male	Low	Hight	Medium	Medium	Medium	Medium	Hight	Hight	Hight	Medium	Vec
Female	Low	Hight	Hight	Hight	Hight	Hight	Hight	Hight	Hight	Hight	Bec
Female	Low	Hight	Medium	Medium	Medium	Hight	Hight	Hight	Hight	Medium	Vec
Female	Medium	Medium	Medium	Hight	Hight	Hight	Hight	Hight	Hight	Hight	Bec
Female	Low	Medium	Low	Medium	Low	Medium	Hight	Hight	Medium	Medium	Vec
Female	Low	Hight	Hight	Hight	Hight	Hight	Hight	Hight	Hight	Hight	Bec
Female	Low	Hight	Hight	Hight	Hight	Hight	Hight	Hight	Hight	Hight	Bec
Female	Low	Hight	Hight	Hight	Hight	Hight	Hight	Hight	Hight	Hight	Bec
Female	Low	Medium	Medium	Medium	Low	Hight	Hight	Hight	Hight	Medium	Vec
Female	Low	Medium	Medium	Hight	Hight	Hight	Hight	Hight	Hight	Hight	Bec
Female	Medium	Hight	Medium	Hight	Medium	Hight	Hight	Medium	Hight	Hight	Bec
Female	Medium	Low	Low	Medium	Medium	Medium	Medium	Medium	Medium	Medium	Bec
Male	Medium	Hight	Hight	Medium	Medium	Hight	Hight	Hight	Medium	Hight	Bec
Male	Medium	Medium	Hight	Medium	Medium	Medium	Medium	Medium	Medium	Medium	Vec
Male	Low	Hight	Hight	Hight	Hight	Hight	Hight	Hight	Hight	Hight	Bec

ภาพที่ 4.7 การแปลงข้อมูลให้อยู่ในรูปแบบที่พร้อมนำไปวิเคราะห์

ตอนที่ 4 การสร้างตัวแบบพยากรณ์ (Modeling)

ชุดข้อมูลที่ผ่านการเตรียมข้อมูลจากทั้งหมด 1,160 เรคคอร์ด 21 คุณลักษณะ ลดเหลือข้อมูลที่ใช้ในงานวิจัยจำนวน 1,030 เรคคอร์ด 11 คุณลักษณะ นำมาสร้างระบบพยากรณ์ใช้วิธีแบบไม่เลือกคุณลักษณะ และการคัดเลือกคุณลักษณะแบบ InfoGainAttributeEval ในการคัดเลือกคุณลักษณะสำคัญมาใช้งาน (feature selection) เป็นวิธีการลดมิติของข้อมูลทำให้ได้คุณลักษณะที่มี

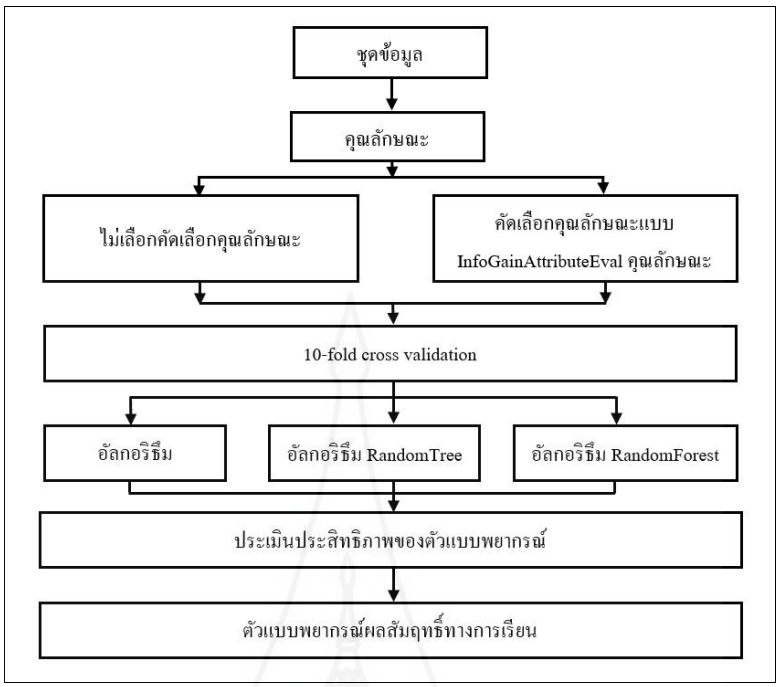
ความสำคัญเกี่ยวข้องกับเป้าหมายเหมาะสมที่จะนำไปสร้างแบบจำลอง ทำให้การประมวลผลมีความรวดเร็วและถูกต้องแม่นยำ โดยจะเลือกเซตของคุณลักษณะที่เหมาะสม กระบวนการเลือกคุณลักษณะมีความสำคัญเกี่ยวข้องกับเป้าหมายซึ่งมีขั้นตอนการคัดเลือกคุณลักษณะ ดังภาพที่ 4.8



ภาพที่ 4.8 ขั้นตอนการคัดเลือกคุณลักษณะหรือฟีเจอร์

จากภาพที่ 4.8 ขั้นตอนการคัดเลือกคุณลักษณะหรือฟีเจอร์ผู้วิจัยได้เลือกคุณลักษณะแบบวิธีฟิลเตอร์ (Filter Method) เป็นวิธีที่ใช้เทคนิคการจัดเรียงลำดับของคุณลักษณะโดยคำนวณหาค่าน้ำหนักความสัมพันธ์ระหว่างฟีเจอร์และคลาสต่างๆ ด้วยค่า Information Gain เป็นค่าที่ใช้ในการวัดคุณลักษณะข้อมูลที่ต้องการมากที่สุดที่เหมาะสมในการนำไปใช้ โดยการหาค่า Gain แต่ละมิติข้อมูล ซึ่งมิติข้อมูลที่มีค่า Gain สูงสุดจะถูกนำมาคำนวณหาค่า Information Gain คือการวัดค่า Entropy ก่อนที่จะมีการแบ่งข้อมูลออกตามคลาสและหลังการแบ่งว่ามีประสิทธิภาพดีขึ้นหรือไม่ ถ้ามีประสิทธิภาพดีขึ้นค่า Information Gain จะมีค่าสูง วิธีฟิลเตอร์นี้สามารถปรับขนาดของชุดข้อมูลที่มีหลายมิติได้ง่ายรวดเร็วและเป็นอิสระกับวิธีการจำแนกประเภท

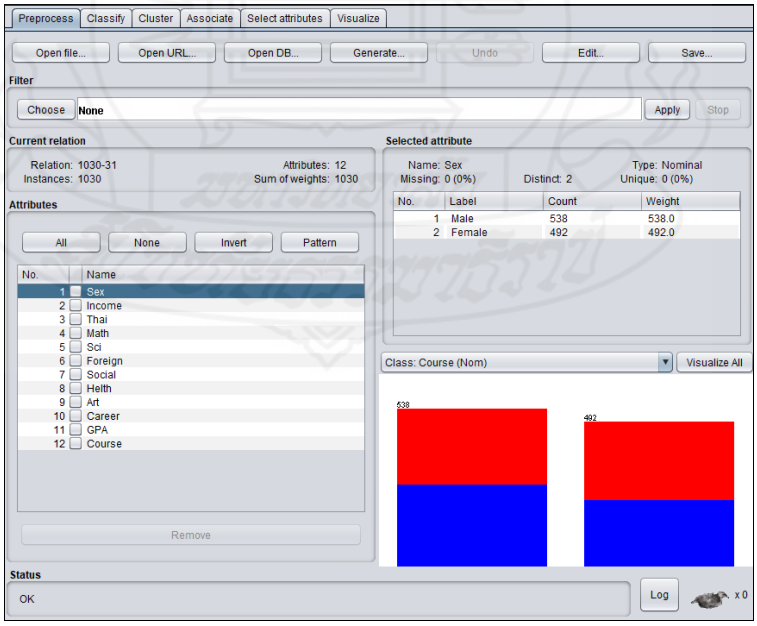
ทำการเปรียบเทียบสร้างตัวแบบพยากรณ์ด้วยอัลกอริธึม 1) อัลกอริธึม C4.5(J48) 2) อัลกอริธึม Random Tree และ 3) อัลกอริธึม Random Forest มาวิเคราะห์วัดประสิทธิภาพพัฒนาเป็นระบบพยากรณ์ผลสัมฤทธิ์ทางการเรียน ดังภาพที่ 4.9



ภาพที่ 4.9 รูปแบบการสร้างตัวแบบพยากรณ์

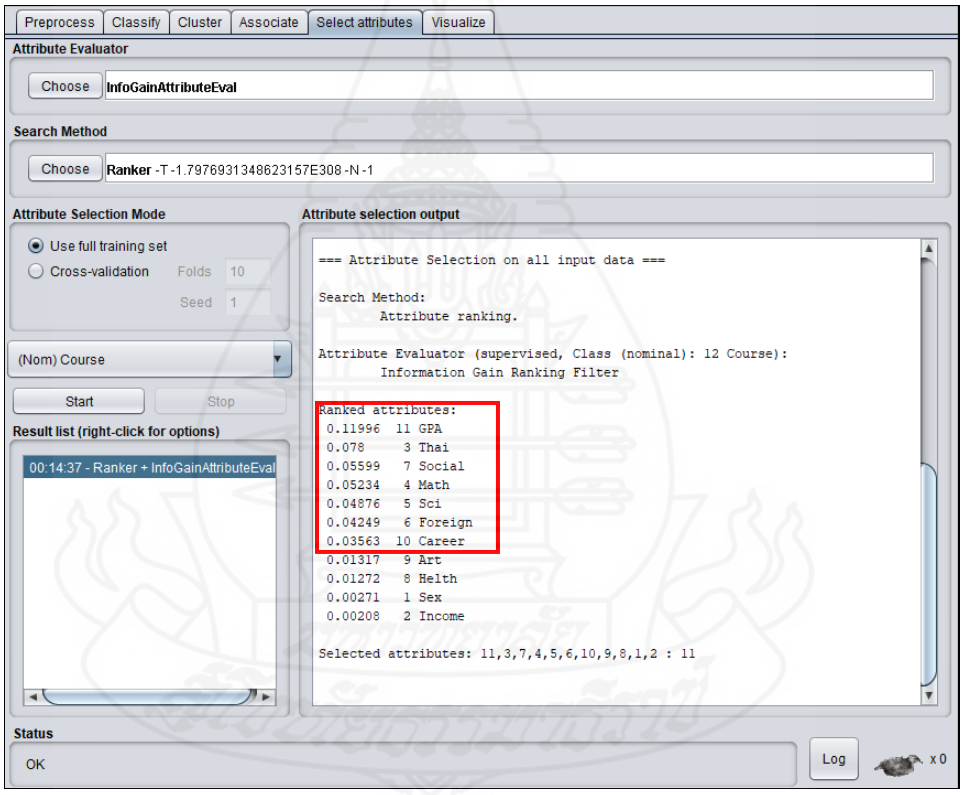
4.1 การคัดเลือกคุณลักษณะ

1) การคัดเลือกคุณลักษณะแบบไม่เลือกคุณลักษณะ (Non-Selection)



ภาพที่ 4.10 ไม่เลือกคุณลักษณะของชุดข้อมูล

2) การคัดเลือกคุณลักษณะแบบวิธี InfoGainAttributeEval จากการคัดเลือกคุณลักษณะจำนวน 11 คุณลักษณะของชุดข้อมูล แบบ InfoGainAttributeEval จะเห็นว่าคุณลักษณะ GPA มีความสัมพันธ์สูงสุดกับคลาสเอาต์พุต นอกจากนี้ยังมีคุณลักษณะหลายอย่างที่มีความสัมพันธ์กัน ผู้วิจัยได้ศึกษางานวิจัยของ เสกสรร วิลัยลักษณ์ (2558:75) ในการใช้เทคนิคการทำเหมืองข้อมูลเพื่อพยากรณ์ผลการเรียนของนักเรียนโรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน โดยจะเลือกคุณลักษณะจากการเรียงลำดับสำคัญไว้ตาม Ranked จากการคำนวณค่า IG ของทุกแอตทริบิวต์ ซึ่งได้เลือกคุณลักษณะที่มีค่าระหว่าง 0.1 – 0.03 ซึ่งมีจำนวน 7 คุณลักษณะ ได้แก่ 1) GPA 2) Thai 3) Social 4) Math 5) Sci 6) Foreign และ 7) Career เพื่อนำไปวิเคราะห์ดังภาพที่ 4.11



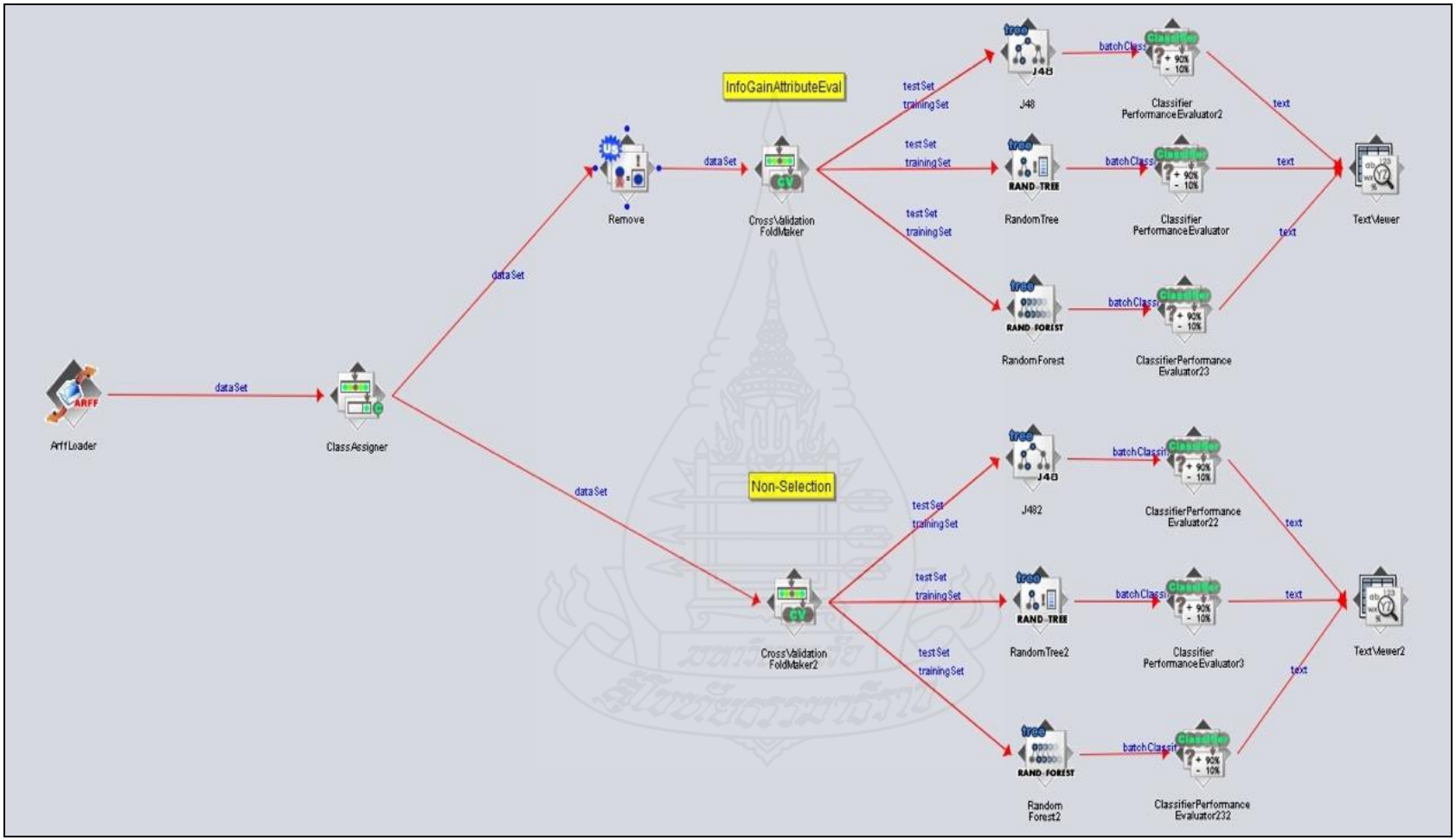
ภาพที่ 4.11 การคัดเลือกคุณลักษณะแบบวิธี InfoGainAttributeEval

ผลการเลือกคุณลักษณะ (feature selection) แบบเรียงตามลำดับเป็นอิสระต่อกันโดยวัด Information Gain (IG) เป็นตัววัดความสัมพันธ์ของคุณลักษณะให้กับคลาส ดังตารางที่ 4.2

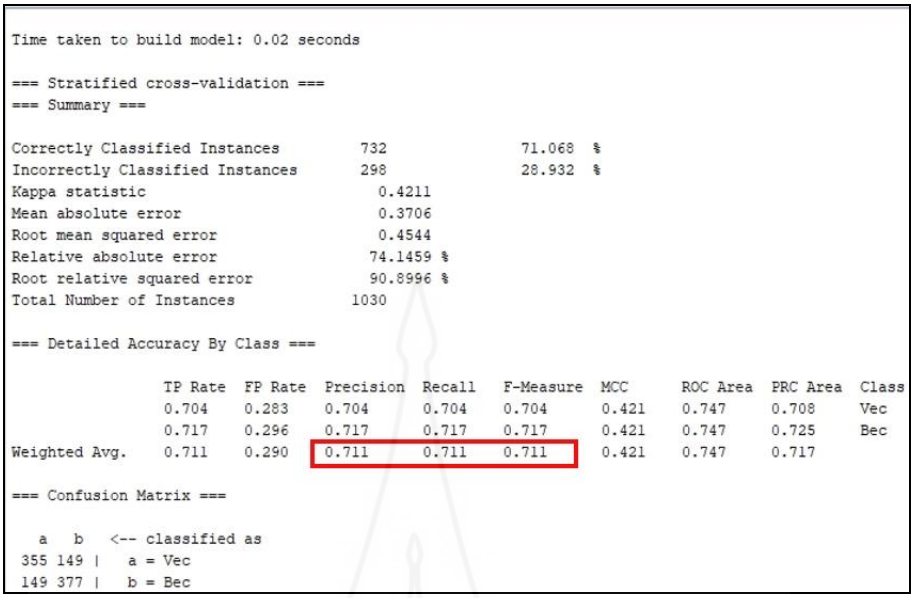
ตารางที่ 4.2 แสดงคุณลักษณะที่จะนำไปวิเคราะห์

คุณลักษณะ	การเลือกคุณลักษณะ	
	ไม่เลือกคุณลักษณะ	InfoGainAttributeEval
Sex	✓	
Income	✓	
Thai	✓	✓
Math	✓	✓
Sci	✓	✓
Foreign	✓	✓
Social	✓	✓
Health	✓	
Art	✓	
Career	✓	✓
GPA	✓	✓

นำข้อมูลที่ได้จากการคัดเลือกคุณลักษณะไปทดสอบประสิทธิภาพของโมเดลใช้อัลกอริธึมที่เลือกไว้ โดยนำชุดข้อมูลจำนวน 1,030 เรคคอร์ด ร่วมกับวิธีแบบไม่เลือกคุณลักษณะและวิธีเลือกคุณลักษณะด้วยวิธี InfoGainAttributeEval ทดสอบประสิทธิภาพของแบบจำลอง แบบ 10-fold cross validation แบ่งชุดข้อมูลสอน 9 ชุด และ 1 ชุดเป็นชุดทดสอบประสิทธิภาพของโมเดล ทำแบบนี้วนไปจนครบจำนวนชุดที่แบ่งไว้ มาทำการเปรียบเทียบประสิทธิภาพกับตัวแบบพยากรณ์ด้วยอัลกอริธึม 1) อัลกอริธึม C4.5(J48) 2) อัลกอริธึมต้นไม้สุ่ม และ 3) อัลกอริธึมป่าสุ่ม เพื่อวิเคราะห์หาประสิทธิภาพในการจำแนกของตัวแบบพยากรณ์ที่ถูกต้องที่สุดจากค่าของการจำแนกข้อมูล (Accuracy) ค่าความแม่นยำ (Precision) ค่าครบถ้วน (Recall) และค่าประสิทธิภาพโดยรวม (F-Measure) ดังภาพที่ 4.12



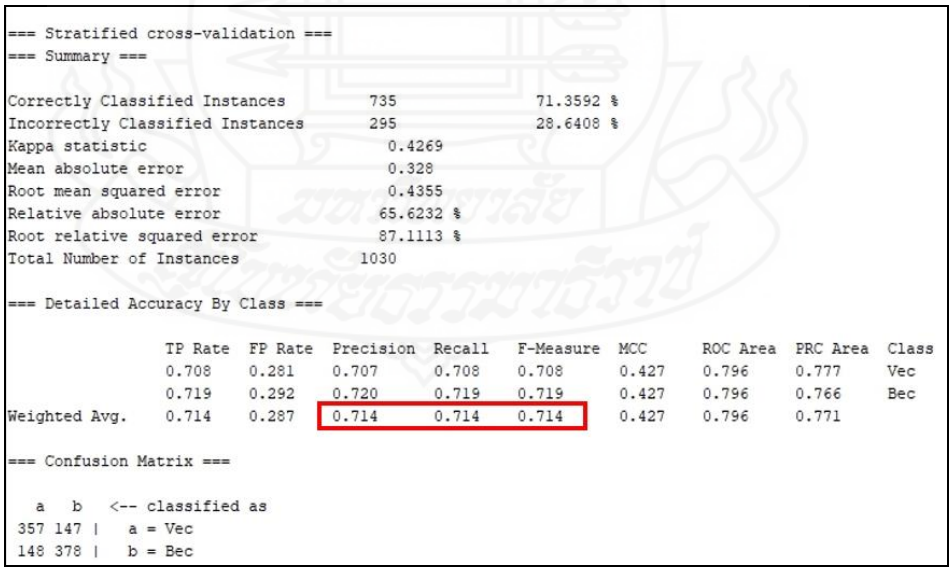
ภาพที่ 4.12 แสดงการวิเคราะห์ประสิทธิภาพด้วย KnowledgeFlow



```
Time taken to build model: 0.02 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         732               71.068  %
Incorrectly Classified Instances       298               28.932  %
Kappa statistic                          0.4211
Mean absolute error                      0.3706
Root mean squared error                  0.4544
Relative absolute error                 74.1459 %
Root relative squared error             90.8996 %
Total Number of Instances             1030

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.704    0.283    0.704      0.704    0.704      0.421    0.747     0.708     Vec
                 0.717    0.296    0.717      0.717    0.717      0.421    0.747     0.725     Bec
Weighted Avg.    0.711    0.290    0.711      0.711    0.711      0.421    0.747     0.717

=== Confusion Matrix ===

   a   b   <-- classified as
 355 149 |   a = Vec
 149 377 |   b = Bec
```

ภาพที่ 4.13 ผลลัพธ์แบบไม่เลือกคุณลักษณะ (Non-Selection) ของโมเดลแบบ C4.5 (J48)

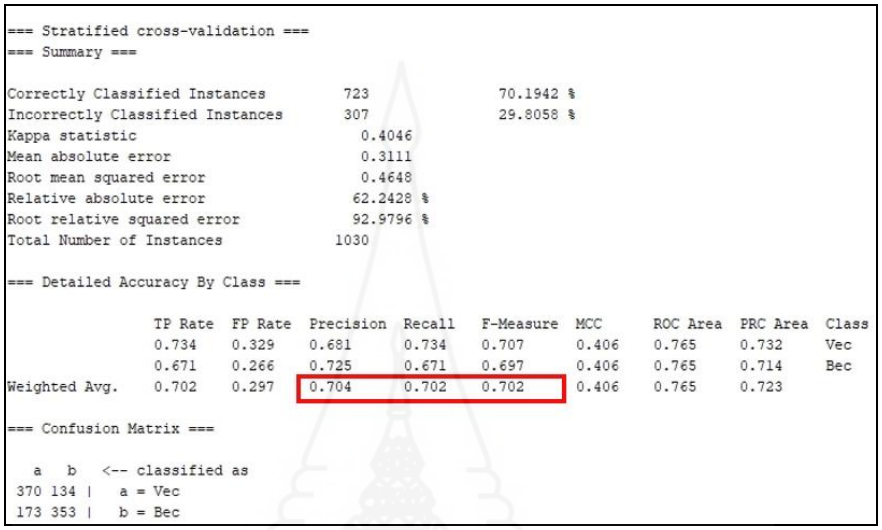
จากภาพที่ 4.13 ผลลัพธ์แบบไม่เลือกคุณลักษณะ (Non-Selection) ของโมเดลแบบ C4.5 (J48) พบว่ามีประสิทธิภาพค่าความถูกต้อง (correctly classified instances) ร้อยละ 71.07 ค่าความแม่นยำ ร้อยละ 71.10 ค่าครบถ้วน ร้อยละ 71.10 และค่าประสิทธิภาพโดยรวม ร้อยละ 71.10



```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         735               71.3592 %
Incorrectly Classified Instances       295               28.6408 %
Kappa statistic                          0.4269
Mean absolute error                      0.328
Root mean squared error                  0.4355
Relative absolute error                 65.6232 %
Root relative squared error             87.1113 %
Total Number of Instances             1030

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.708    0.281    0.707      0.708    0.708      0.427    0.796     0.777     Vec
                 0.719    0.292    0.720      0.719    0.719      0.427    0.796     0.766     Bec
Weighted Avg.    0.714    0.287    0.714      0.714    0.714      0.427    0.796     0.771

=== Confusion Matrix ===

   a   b   <-- classified as
 357 147 |   a = Vec
 148 378 |   b = Bec
```

ภาพที่ 4.14 ผลลัพธ์แบบไม่เลือกคุณลักษณะ (Non-Selection) ของโมเดลแบบ Random Forest

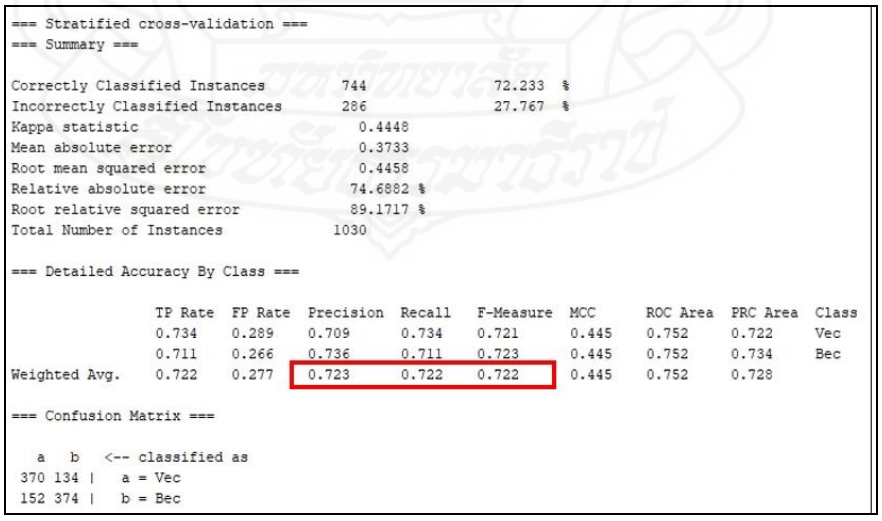
ภาพที่ 4.14 ผลลัพธ์แบบไม่เลือกคุณลักษณะ (Non-Selection) ของโมเดลแบบ Random Forest พบว่ามีประสิทธิภาพค่าความถูกต้อง ร้อยละ 71.36 ค่าความแม่นยำ ร้อยละ 71.40 ค่าครบถ้วน ร้อยละ 71.40 และค่าประสิทธิภาพโดยรวม ร้อยละ 71.40



```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         723               70.1942 %
Incorrectly Classified Instances       307               29.8058 %
Kappa statistic                          0.4046
Mean absolute error                      0.3111
Root mean squared error                  0.4648
Relative absolute error                 62.2428 %
Root relative squared error             92.9796 %
Total Number of Instances             1030

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure  MCC      ROC Area  PRC Area  Class
               0.734     0.329     0.681       0.734   0.707      0.406    0.765     0.732     Vec
               0.671     0.266     0.725       0.671   0.697      0.406    0.765     0.714     Bec
Weighted Avg.  0.702     0.297     0.704       0.702   0.702      0.406    0.765     0.723

=== Confusion Matrix ===

   a   b   <-- classified as
 370 134 |   a = Vec
 173 353 |   b = Bec
```

ภาพที่ 4.15 ผลลัพธ์แบบไม่เลือกคุณลักษณะ (Non-Selection) ของโมเดลแบบ Random Tree

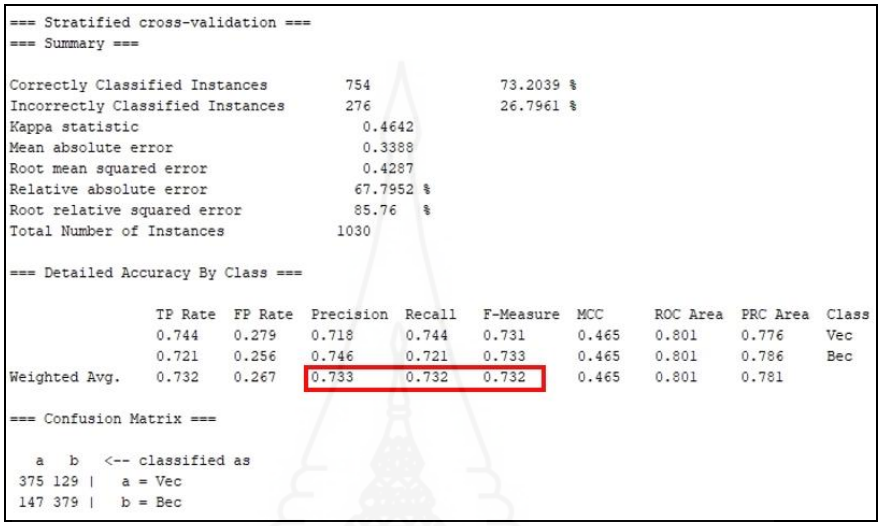
ภาพที่ 4.15 ผลลัพธ์แบบไม่เลือกคุณลักษณะ (Non-Selection) ของโมเดลแบบ Random Tree พบว่ามีประสิทธิภาพค่าความถูกต้อง ร้อยละ 70.19 ค่าความแม่นยำ ร้อยละ 70.40 ค่าครบถ้วน ร้อยละ 70.20 และค่าประสิทธิภาพโดยรวม ร้อยละ 70.20



```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         744               72.233  %
Incorrectly Classified Instances       286               27.767  %
Kappa statistic                          0.4448
Mean absolute error                      0.3733
Root mean squared error                  0.4458
Relative absolute error                 74.6882 %
Root relative squared error             89.1717 %
Total Number of Instances             1030

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure  MCC      ROC Area  PRC Area  Class
               0.734     0.289     0.709       0.734   0.721      0.445    0.752     0.722     Vec
               0.711     0.266     0.736       0.711   0.723      0.445    0.752     0.734     Bec
Weighted Avg.  0.722     0.277     0.723       0.722   0.722      0.445    0.752     0.728

=== Confusion Matrix ===

   a   b   <-- classified as
 370 134 |   a = Vec
 152 374 |   b = Bec
```

ภาพที่ 4.16 ผลลัพธ์แบบวิธี InfoGainAttributeEval ของโมเดลแบบ ของโมเดลแบบ C4.5 (J48)

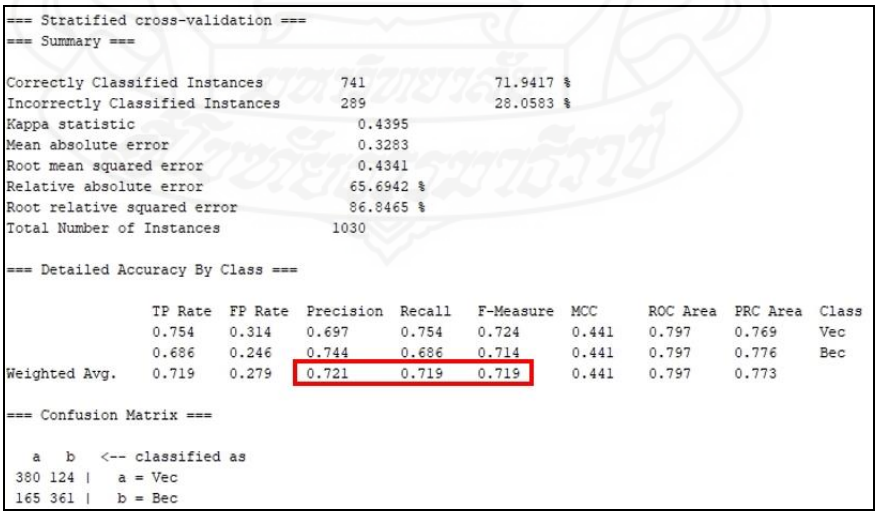
ภาพที่ 4.16 ผลลัพธ์แบบวิธี InfoGainAttributeEval ของโมเดลแบบ ของโมเดลแบบ C4.5 (J48) พบว่ามีประสิทธิภาพค่าความถูกต้อง ร้อยละ 72.23 ค่าความแม่นยำ ร้อยละ 72.30 ค่าครบถ้วน ร้อยละ 72.20 และค่าประสิทธิภาพโดยรวม ร้อยละ 72.20



```
=== 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 4.17 ผลลัพธ์แบบวิธี InfoGainAttributeEval ของโมเดลแบบ Random Forest

ภาพที่ 4.17 ผลลัพธ์แบบวิธี InfoGainAttributeEval ของโมเดลแบบ Random Forest พบว่ามีประสิทธิภาพค่าความถูกต้อง ร้อยละ 73.20 ค่าความแม่นยำ ร้อยละ 73.30 ค่าครบถ้วน ร้อยละ 73.20 และค่าประสิทธิภาพโดยรวม ร้อยละ 73.20



```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         741               71.9417 %
Incorrectly Classified Instances       289               28.0583 %
Kappa statistic                          0.4395
Mean absolute error                      0.3283
Root mean squared error                  0.4341
Relative absolute error                 65.6942 %
Root relative squared error             86.8465 %
Total Number of Instances             1030

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.754    0.314    0.697      0.754    0.724      0.441    0.797     0.769     Vec
                 0.686    0.246    0.744      0.686    0.714      0.441    0.797     0.776     Bec
Weighted Avg.    0.719    0.279    0.721      0.719    0.719      0.441    0.797     0.773

=== Confusion Matrix ===

   a   b   <-- classified as
 380 124 |   a = Vec
 165 361 |   b = Bec
```

ภาพที่ 4.18 ผลลัพธ์แบบวิธี InfoGainAttributeEval ของโมเดลแบบ RandomTree

ภาพที่ 4.18 ผลลัพธ์แบบวิธี InfoGainAttributeEval ของโมเดลแบบ Random Tree พบว่ามีประสิทธิภาพค่าความถูกต้อง ร้อยละ 71.94 ค่าความแม่นยำ ร้อยละ 72.10 ค่าครบถ้วน ร้อยละ 71.90 และค่าประสิทธิภาพโดยรวม ร้อยละ 71.90

ตอนที่ 5 การประเมินผล (Evaluation)

ทำการทดสอบประสิทธิภาพของแบบจำลอง (Model) แบบ 10-fold cross validation ซึ่งเป็นการแบ่งข้อมูลที่ใช้ในการทดสอบออกเป็นชุดๆ ละเพื่อทดสอบแบบจำลองหาความน่าเชื่อถือแบบ 10-fold cross validation แบ่งข้อมูลฝึกสอนเป็น 9 ชุด และ 1 ชุดเป็นข้อมูลทดสอบ เพื่อวัดประสิทธิภาพของค่าความถูกต้อง (Accuracy) ค่าความแม่ยำ (Precision) ค่าครบถ้วน (Recall) และค่าประสิทธิภาพโดยรวม (F-Measure)

5.1 ค่าความถูกต้อง (Accuracy) เป็นการวัดความถูกต้องของแบบจำลอง

True Positive (TP) คือ สิ่งที่โปรแกรมทำนายว่าจริงและสิ่งที่เกิดขึ้นเป็นจริง

True Negative (TN) คือ สิ่งที่โปรแกรมทำนายว่าไม่จริงและสิ่งที่เกิดขึ้นไม่จริง

False Positive (FP) คือ สิ่งที่โปรแกรมทำนายว่าจริงแต่สิ่งที่เกิดขึ้นไม่จริง

False Negative (FN) คือ สิ่งที่ทำนายว่าไม่จริงแต่สิ่งที่เกิดขึ้นเป็นจริง

$$\text{Accuracy} = \frac{TP+TN}{\text{TP+TN+FP+FN}} \times 100 \qquad (6)$$

5.2 ค่าความแม่นยำ (Precision) เป็นการวัดความแม่นยำของข้อมูล

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \qquad (7)$$

5.3 ค่าครบถ้วน (Recall) เป็นการวัดความถูกต้องของ Model

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \qquad (8)$$

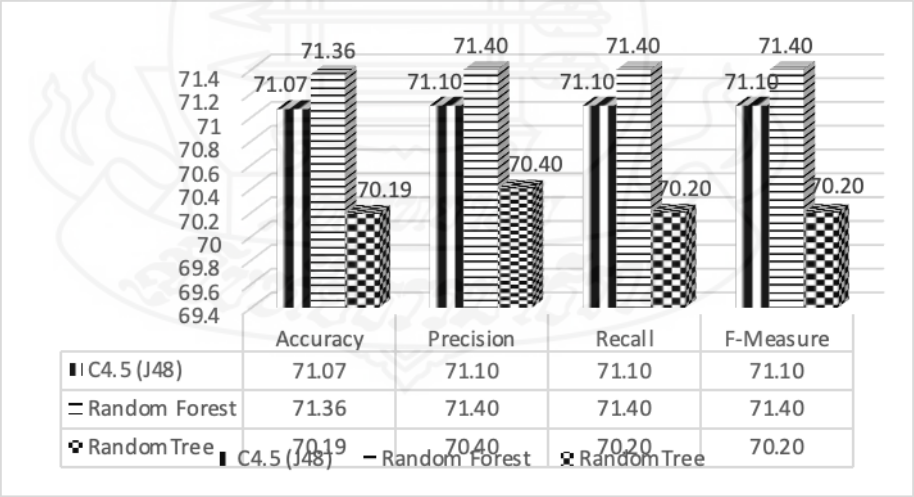
5.4 ค่าประสิทธิภาพโดยรวม (F-Measure) เป็นการหาค่าประสิทธิภาพโดยรวมนำสองค่า (Precision และ Recall) มาพิจารณาร่วมกันเพื่อความแม่นยำ

$$F - Measure = \frac{2xPrecisionxRecall}{Precision+Recall} \quad (9)$$

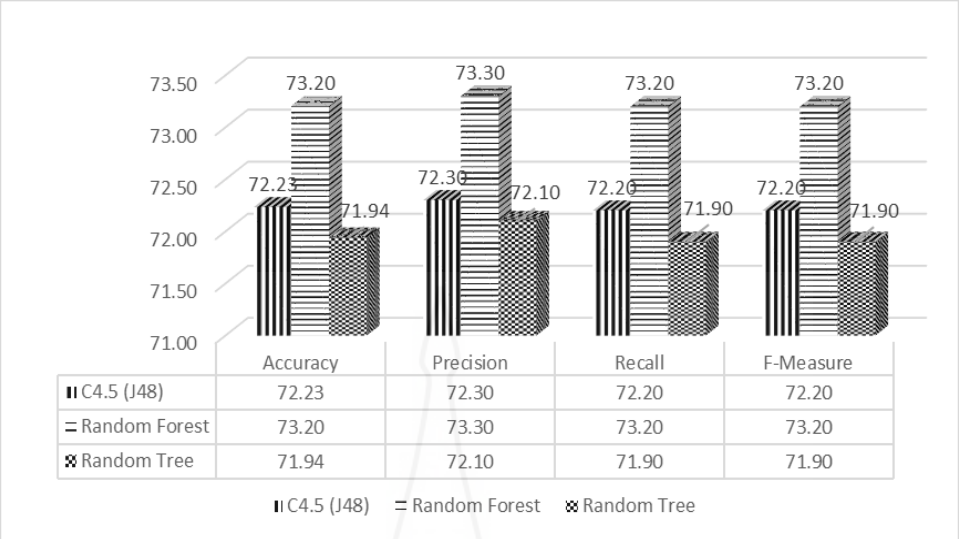
ผลการประเมินประสิทธิภาพของตัวแบบพยากรณ์ระบบพยากรณ์เพื่อสนับสนุนการตัดสินใจศึกษาต่อหลังสำเร็จมัธยมศึกษาตอนต้น โดยการนำค่าวัดประสิทธิภาพได้แก่ค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าครบถ้วน (Recall) และค่าประสิทธิภาพโดยรวม (F-Measure) ของแต่ละอัลกอริธึมมาเปรียบเทียบกัน ดังตารางที่ 4.3

ตารางที่ 4.3 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

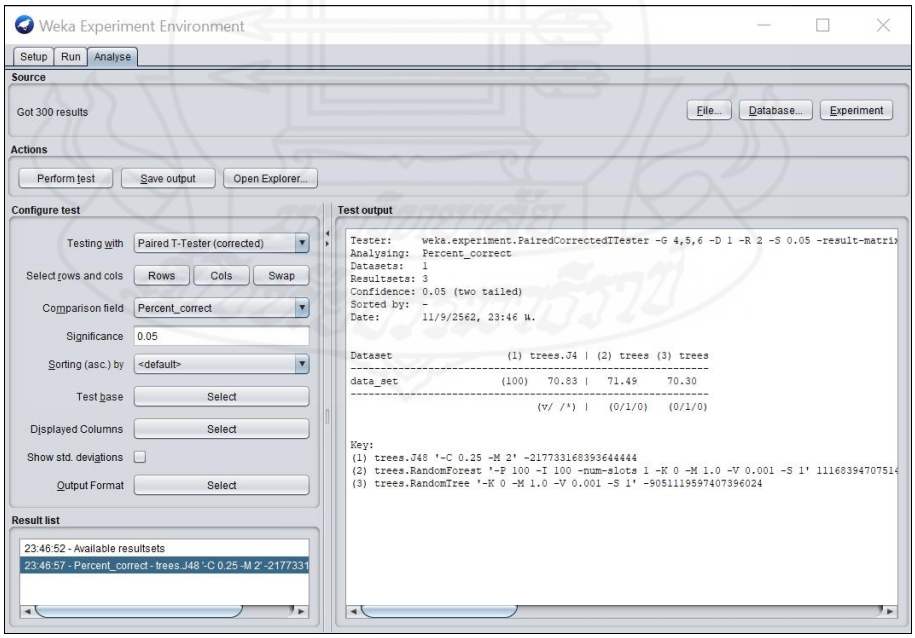


ภาพที่ 4.19 กราฟแสดงผลลัพธ์แบบไม่เลือกคุณลักษณะ (Non-Selection)



ภาพที่ 4.20 กราฟแสดงผลลัพธ์การเลือกคุณลักษณะแบบ InfoGainAttributeEval

จากภาพที่ 4.20 จะเห็นว่าการเลือกคุณลักษณะแบบวิธี InfoGainAttributeEval โดยอัลกอริธึม Random Forest ให้ประสิทธิภาพในการพยากรณ์ถูกต้องมากกว่า มีค่า Accuracy 73.20 ค่า Precision 73.30 ค่า Recall 73.20 และค่า F-Measure 73.20 ด้วยจำนวน 7 คุณลักษณะ คือ 1) GPA 2) Thai 3) Social 4) Math 5) Sci 6) Foreign และ 7) Career



ภาพที่ 4.21 แสดงผลทดสอบแบบจำลองทั้ง 3 อัลกอริธึม

จากภาพที่ 4.21 นำผลลัพธ์มาขยายดังภาพที่ 4.22 จะเห็นว่าผลการทดสอบประสิทธิภาพของแบบจำลองทั้ง 3 อัลกอริธึมจากชุดข้อมูลที่นำมาทดสอบโดยแสดงเป็นเปอร์เซ็นความถูกต้องของ อัลกอริธึม C4.5(J48) 70.83% ค่า V หมายถึงผลลัพธ์เจาะจงดี ส่วน * คือผลลัพธ์ที่ต่ำกว่าเส้นฐาน อัลกอริธึมป่าสุ่ม 71.49% มีค่าสถิติดีกว่าค่าพื้นฐานที่กำหนดโดยอัลกอริธึม C4.5(J48) มีค่าพื้นฐานปัจจุบันที่ 0.5 จากชุดข้อมูลที่ใช้ในการทดสอบ (v/ /*) และอัลกอริธึมต้นไม้สุ่ม 70.30% มีค่าสถิติในการทดสอบ (0/1/0) ที่ดีกว่าค่าพื้นฐานที่กำหนดโดยอัลกอริธึม C4.5(J48) จากชุดข้อมูลที่ใช้ในการทดสอบ (0/1/0) จะเห็นว่าอัลกอริธึมป่าสุ่มมีผลลัพธ์สูงกว่าเส้นฐานและถูกต้อง มากกว่าอัลกอริธึมอื่น ดังภาพที่ 4.22

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

ภาพที่ 4.22 แสดงผลทดสอบแบบประสิทธิภาพทั้ง 3 อัลกอริธึม

นำอัลกอริธึมป่าสุ่มโดยวิธีเลือกคุณลักษณะแบบวิธี InfoGainAttributeEval ที่ให้ประสิทธิภาพในการพยากรณ์ถูกต้องมาทดสอบโมเดลในการพยากรณ์โดยใช้ชุดข้อมูลจำนวน 10 เรคคอร์ด โดยมีคลาสเป้าหมายที่ไม่ทราบค่าเพื่อวัดประสิทธิภาพในการพยากรณ์ผลลัพธ์คือโมเดลสามารถพยากรณ์ได้ถูกต้อง ดังภาพที่ 4.23

```
Time taken to build model: 0.15 seconds

=== Predictions on test set ===

    inst#     actual  predicted error prediction
        1        1:?      2:Bec       1
        2        1:?      2:Bec       1
        3        1:?      2:Bec       1
        4        1:?      1:Vec       1
        5        1:?      2:Bec       0.99
        6        1:?      1:Vec       1
        7        1:?      2:Bec       0.9
        8        1:?      2:Bec       1
        9        1:?      2:Bec       1
       10        1:?      1:Vec       1

=== Evaluation on test set ===

Time taken to test model on supplied test set: 0.01 seconds

=== Summary ===

Total Number of Instances               0
Ignored Class Unknown Instances                10
```

ภาพที่ 4.23 แสดงผลทดสอบโมเดลในการพยากรณ์

จากภาพที่ 4.23 แสดงผลทดสอบโมเดลในการพยากรณ์ actual คือค่าจริงที่ป้อนเข้ามาให้โมเดลทำนาย predicted คือผลที่โมเดลพยากรณ์ได้ว่าเป็น 1:Vec หรือ 2:Bec ส่วน prediction คือค่าความถูกต้องของที่โมเดลพยากรณ์ซึ่ง 1 หมายถึง 100 เปอร์เซ็นต์

ตอนที่ 6 การนำไปใช้งาน (Deployment)

เมื่อได้ตัวแบบพยากรณ์ด้วยการเลือกคุณลักษณะแบบวิธี InfoGainAttributeEval จากขั้นตอนการทำเหมืองข้อมูลด้วยโปรแกรม WEKA ที่เหมาะสมแล้ว มีผลลัพธ์หรือคลาส (Class) คือแผนการศึกษาต่อเป็นแนวทางในการสนับสนุนการตัดสินใจให้ผู้เรียน ผู้วิจัยได้พัฒนาระบบพยากรณ์เพื่อสนับสนุนการตัดสินใจศึกษาต่อ หลังสำเร็จมัธยมศึกษาตอนต้น โดยใช้อัลกอริธึมป่าสุ่ม ในสังกัดสำนักงานเขตพื้นที่การศึกษาประถมศึกษานครสวรรค์ เขต 2 ด้วยโปรแกรม Dreamweaver CS6 ร่วมกับภาษา PHP และโปรแกรม Appserv จำลอง web server ในการพัฒนาเว็บไซต์เพื่อให้ผู้เรียนสามารถใช้งานผ่านเว็บเบราว์เซอร์ ในการพยากรณ์แนวทางในการศึกษาต่อระดับสายสามัญ (Basic Education Commission : Bec) และสายอาชีวศึกษา (Vocational Education Commission : Vec) โดยแบ่งออกเป็น 2 ส่วน คือ 1) ผู้ใช้งานทั่วไป 2) ผู้ดูแลระบบ

6.1 การออกแบบฐานข้อมูล

ฐานข้อมูลในการเก็บข้อมูลที่เกี่ยวกับระบบพยากรณ์เพื่อสนับสนุนการตัดสินใจศึกษาต่อหลังสำเร็จมัธยมศึกษาตอนต้น โดยใช้อัลกอริธึมป่าสุ่ม มี 2 ตารางได้แก่ tbl_data ใช้สำหรับเก็บข้อมูลคุณลักษณะของชุดข้อมูลทดสอบ และ tbl_admin ใช้สำหรับเก็บข้อมูลผู้ดูแลระบบ ตารางแสดงในรูปแบบ Data Dictionary ดังตารางที่ 4.4 และตารางข้อมูลผู้ดูและระบบที่ 4.5

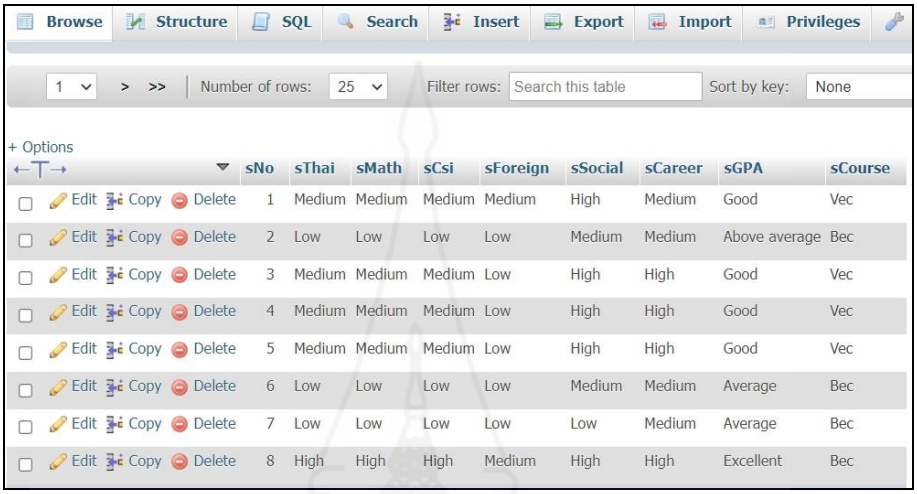
ตารางที่ 4.4 Data Dictionary ของตาราง tbl_data

ลำดับ	ชื่อฟิลด์	PK	ชนิดข้อมูล	คำอธิบาย
1	sNo	PK	int(11)	แสดงลำดับแบบ auto_increment
2	sThai		varchar(255)	ผลสัมฤทธิ์ของกลุ่มสาระวิชาภาษาไทย
3	sMath		varchar(255)	ผลสัมฤทธิ์ของกลุ่มสาระวิชาคณิตศาสตร์
4	sSci		varchar(255)	ผลสัมฤทธิ์ของกลุ่มสาระวิชาวิทยาศาสตร์
5	sForeign		varchar(255)	ผลสัมฤทธิ์ของกลุ่มสาระวิชาภาษาต่างประเทศ
6	sSocial		varchar(255)	ผลสัมฤทธิ์ของกลุ่มสาระวิชาสังคมศึกษาศาสนาและวัฒนธรรม
7	sCareer		varchar(255)	ผลสัมฤทธิ์ของกลุ่มสาระวิชาการงานอาชีพและเทคโนโลยี
8	sGPA		varchar(255)	ผลการเรียนเฉลี่ยตลอดหลักสูตร
9	sCourse		varchar(255)	แผนการศึกษาต่อ

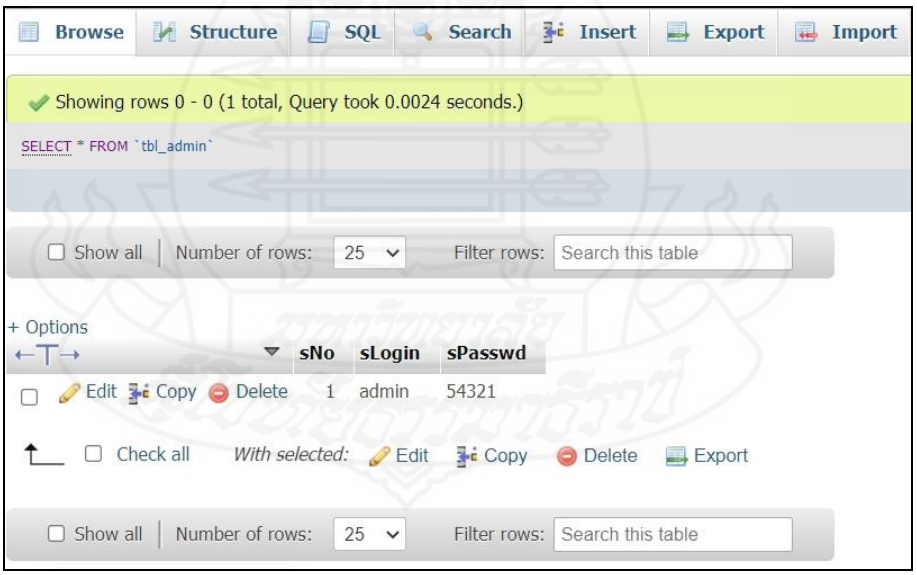
ตารางที่ 4.5 Data Dictionary ของตาราง tbl_admin

ลำดับ	ชื่อฟิลด์	PK	ชนิดข้อมูล	คำอธิบาย
1	sNo	PK	int(11)	แสดงลำดับแบบ auto_increment
2	sLogin		varchar(255)	ชื่อผู้ใช้งาน
3	sPassword		varchar(255)	รหัสผ่านผู้ใช้งาน

เมื่อได้ Data Dictionary ของตารางทั้ง 2 ตารางแล้วก็ดำเนินการจัดการข้อมูลให้อยู่ในรูปแบบของฐานข้อมูล ผู้วิจัยได้ใช้ โปรแกรม Appserv ซึ่งเป็นโปรแกรมช่วยในการจัดการฐานข้อมูล ดังภาพที่ 4.24



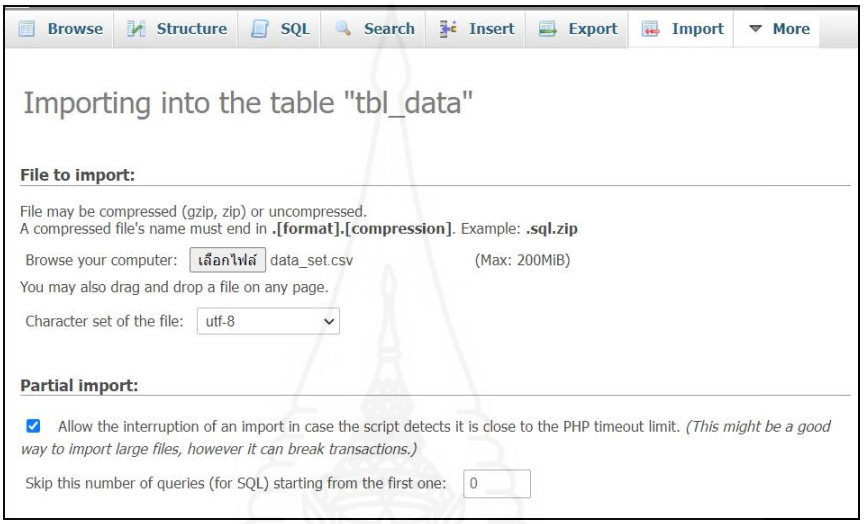
ภาพที่ 4.24 ฐานข้อมูล tbl_data



ภาพที่ 4.25 ฐานข้อมูล tbl_admin

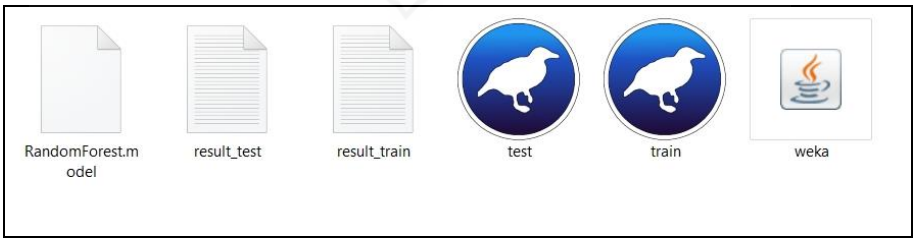
6.2 การออกแบบกระบวนการ

ในการออกแบบกระบวนการของระบบพยากรณ์เพื่อสนับสนุนการตัดสินใจศึกษาต่อหลังสำเร็จมัธยมศึกษาตอนต้น โดยใช้อัลกอริธึมป่าสุ่ม ผู้วิจัยได้ทำการแปลงชุดข้อมูลจาก .arff ให้อยู่ในรูปแบบของไฟล์ .csv เพื่อนำเข้าสู่ฐานข้อมูล ดังภาพที่ 4.26



ภาพที่ 4.26 การนำเข้าชุดข้อมูลในฐานข้อมูล

นำ library weka.jar และไฟล์โมเดลอัลกอริธึมป่าสุ่มที่ได้จากชุดข้อมูลที่ได้ผ่านการทดสอบประสิทธิภาพจากโปรแกรม WEKA มาอยู่ในโฟลเดอร์เดียวกันกับเว็บไซต์ระบบพยากรณ์ โดยมีรูปแบบการเรียกใช้ WEKA ผ่านตัว PHP มีการรับค่าจากผู้ใช้งานมาใช้ในการวิเคราะห์แปลงเป็น .csv โดยปรับเปลี่ยนข้อมูลให้อยู่ในรูปแบบที่เหมาะกับอัลกอริธึมป่าสุ่ม แล้ทำการเรียก WEKA มาแปลงไฟล์จาก .csv เป็น .arff เพื่อจะนำไปทดสอบกับโมเดลอัลกอริธึมป่าสุ่มที่ได้สร้างไว้ ทำการประมวลผลออกมา ดังภาพที่ 4.27



ภาพที่ 4.27 library weka.jar และไฟล์โมเดล

```
@relation RandomForest

@attribute sThai {Medium,Low,High}
@attribute sMath {Medium,Low,High}
@attribute sCsi {Medium,Low,High}
@attribute sForeign {Medium,Low,High}
@attribute sSocial {High,Medium,Low}
@attribute sCareer {Medium,High,Low}
@attribute sGPA {Good,Above_average,Average,Excellent,Very_Good,Fair}
@attribute sCourse {Vec,Bec}

@data
Medium,Medium,Medium,Medium,High,Medium,Good,Vec
Low,Low,Low,Low,Medium,Medium,Above_average,Bec
Medium,Medium,Medium,Low,High,High,Good,Vec
Medium,Medium,Medium,Low,High,High,Good,Vec
Medium,Medium,Medium,Low,High,High,Good,Vec
```

ภาพที่ 4.28 การแปลงไฟล์จาก .csv เป็น .arff ของตัวระบบ

6.3 การออกแบบส่วนต่อประสานผู้ใช้

ในการออกแบบส่วนต่อประสานผู้ใช้ เพื่อใช้งานผ่าน web browser ประกอบด้วย 2 ส่วน ดังนี้

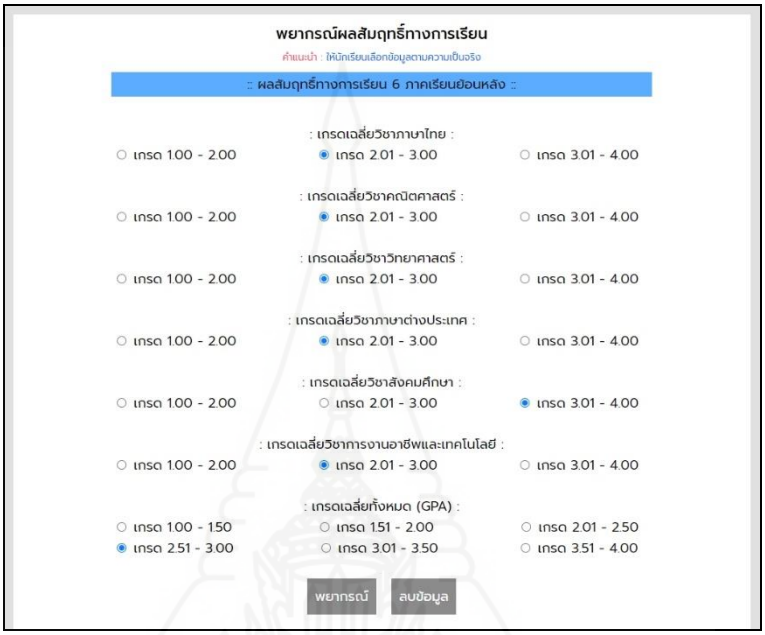
ส่วนที่ 1 หน้าจอหลักของผู้ใช้งานทั่วไป จะประกอบไปด้วยเมนู หน้าแรก พยากรณ์ผลสัมฤทธิ์ ผู้ดูแลระบบ ผู้พัฒนา ซึ่งในส่วนเมนูหน้าแรกจะกล่าวถึงที่มาและความสำคัญของการวิจัยระบบพยากรณ์เพื่อสนับสนุนการตัดสินใจศึกษาต่อ หลังสำเร็จมัธยมศึกษาตอนต้น โดยใช้อัลกอริธึมป่าสุ่ม ดังภาพที่ 4.29

ภาพที่ 4.29 แสดงตัวอย่างหน้าแรกของระบบ

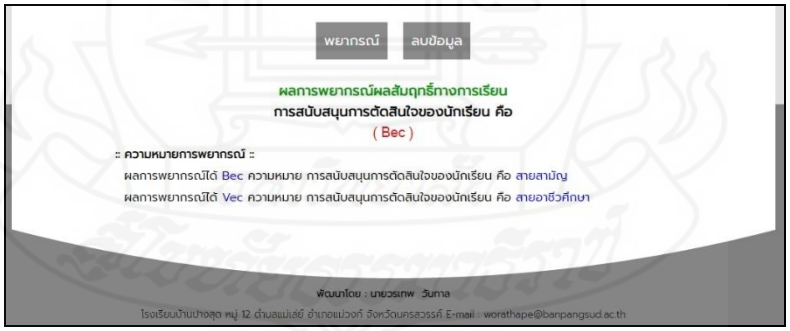
เมนูพยากรณ์ผลสัมฤทธิ์ เพื่อให้ผู้ใช้งานใส่ข้อมูลประกอบไปด้วยคุณลักษณะจำนวน 7 คุณลักษณะ คือ 1) เกรดเฉลี่ยกลุ่มสาระภาษาไทย 2) เกรดเฉลี่ยกลุ่มสาระคณิตศาสตร์ 3) เกรดเฉลี่ยกลุ่มสาระวิทยาศาสตร์ 4) เกรดเฉลี่ยกลุ่มสาระภาษาต่างประเทศ 5) เกรดเฉลี่ยกลุ่มสาระสังคมศึกษา 6) เกรดเฉลี่ยกลุ่มสาระการงานอาชีพและเทคโนโลยี 7) ผลการเรียนเฉลี่ยตลอดหลักสูตร ดังรูปที่ 4.30

ภาพที่ 4.30 แสดงตัวอย่างหน้าพยากรณ์ผลสัมฤทธิ์ของระบบ

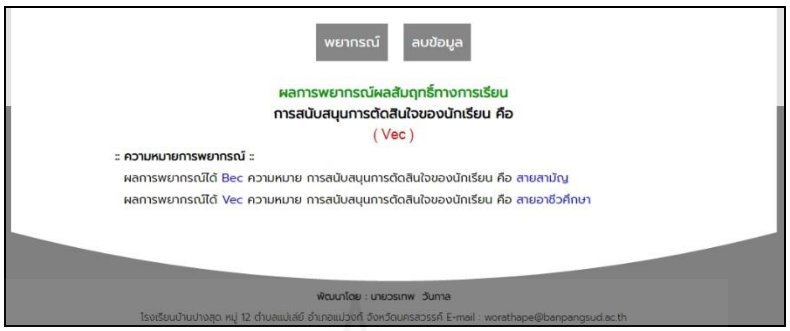
เมื่อผู้ใช้งานกรอกรายละเอียดข้อมูลผลการเรียนตามหัวข้อต่างๆ แล้วทำการคลิกปุ่มพยากรณ์ ระบบก็จะทำการประมวลผลข้อมูลเพื่อแสดงผลแนวทางในการศึกษาต่อของนักเรียนอยู่ในระดับสายสามัญ หรือสายอาชีวศึกษา



ภาพที่ 4.31 แสดงการกรอกรายละเอียดข้อมูลผลการเรียน



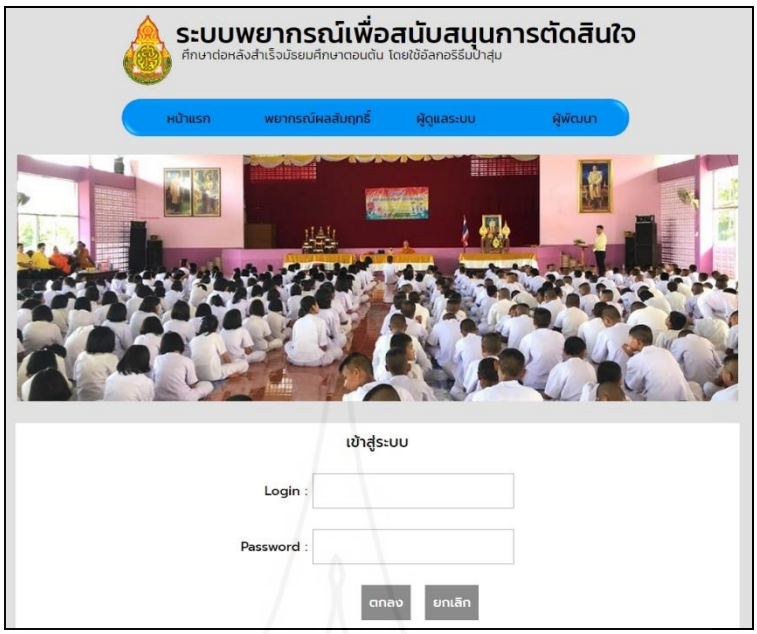
ภาพที่ 4.32 แสดงผลการพยากรณ์การสนับสนุนการตัดสินใจในสายสามัญ



ภาพที่ 4.33 แสดงผลการพยากรณ์การสนับสนุนการตัดสินใจในสายอาชีวศึกษา

ภาพที่ 4.34 แสดงตัวอย่างผู้พัฒนาระบบ

ส่วนที่ 2 ผู้ดูแลระบบ จะประกอบไปด้วยการเพิ่มข้อมูลการเทรน (train) ของแบบจำลอง ซึ่งในส่วนเมนูหน้านี้ผู้ดูแลระบบสามารถเพิ่มข้อมูลที่เป็นปัจจุบันของผลสัมฤทธิ์ทางการเรียนของนักเรียนในปีการศึกษาปัจจุบัน เพื่อให้ระบบพยากรณ์เพื่อสนับสนุนการตัดสินใจศึกษาต่อหลังสำเร็จมัธยมศึกษาตอนต้น โดยใช้อัลกอริธึมป่าสุ่มมีความเป็นปัจจุบันและถูกต้องแม่นยำ โดยการ login เข้าสู่ระบบ ดังภาพที่ 4.35



ภาพที่ 4.35 แสดงหน้าผู้ดูแลระบบ

ในส่วนผู้ดูแลระบบสามารถเพิ่มข้อมูลที่มีนามสกุลไฟล์ .csv ที่เป็นปัจจุบันของผลสัมฤทธิ์ทางการเรียนของนักเรียนในปีการศึกษาปัจจุบันเข้าสู่ฐานข้อมูล ดังภาพที่ 4.36

ภาพที่ 4.36 แสดงระบบการจัดการข้อมูลผู้ดูแลระบบ

บทที่ 5
สรุปการวิจัย อภิปรายผล และข้อเสนอแนะ

งานวิจัยเรื่องระบบพยากรณ์เพื่อสนับสนุนการตัดสินใจศึกษาต่อหลังสำเร็จมัธยมศึกษาตอนต้น โดยใช้อัลกอริธึมป่าสุ่ม โดยนำผลสัมฤทธิ์ทางการเรียนของนักเรียน จำนวน 6 ภาคเรียน 8 กลุ่มสาระการเรียนรู้ มาวิเคราะห์ด้วยการคัดเลือกคุณลักษณะกับเทคนิคเหมืองข้อมูลที่เหมาะสมด้วยอัลกอริธึม C4.5(J48) อัลกอริธึมต้นไม้สุ่ม และอัลกอริธึมป่าสุ่ม เพื่อเปรียบเทียบประสิทธิภาพ

ข้อมูลผู้วิจัยได้ใช้ข้อมูลผลสัมฤทธิ์ทางการเรียนของนักเรียนชั้นมัธยมศึกษาปีที่ 3 ในสังกัดสำนักงานเขตพื้นที่การศึกษาประถมศึกษานครสวรรค์ เขต 2 ระหว่างปีการศึกษา 2558 – 2560 จำนวน 1,030 เรคคอร์ด 11 คุณลักษณะ ใช้วิธีการคัดเลือกคุณลักษณะแบบวิธี InfoGainAttributeEval ทดสอบประสิทธิภาพของแบบจำลองแบบ 10-fold cross validation โดยอัลกอริธึมป่าสุ่ม ด้วยโปรแกรม WEKA จากนั้นนำมาพัฒนาเป็นระบบพยากรณ์เพื่อสนับสนุนการตัดสินใจศึกษาต่อหลังสำเร็จมัธยมศึกษาตอนต้น โดยใช้อัลกอริธึมป่าสุ่ม ใช้โปรแกรม Dreamweaver CS6 พัฒนาเป็นเว็บไซต์ เพื่อให้ผู้ใช้สามารถใช้งานผ่านเว็บเบราว์เซอร์

1. สรุปการวิจัย

1.1 จากการสร้างแบบจำลองพยากรณ์เพื่อสนับสนุนการตัดสินใจศึกษาต่อ หลังสำเร็จมัธยมศึกษาตอนต้น โดยใช้อัลกอริธึมป่าสุ่ม โดยนำข้อมูลผลสัมฤทธิ์ทางการเรียนของนักเรียนชั้นมัธยมศึกษาปีที่ 3 ในสังกัดสำนักงานเขตพื้นที่การศึกษาประถมศึกษานครสวรรค์ เขต 2 ตั้งแต่ปีการศึกษา 2558 - 2560 ใช้กลุ่มตัวอย่างข้อมูลของผลสัมฤทธิ์ทางการเรียน 8 กลุ่มสาระวิชาย้อนหลัง 6 ภาคเรียน (2558 - 2560) จำนวน 1,030 เรคคอร์ด ผลการวิเคราะห์ข้อมูลด้วยเทคนิคเหมืองข้อมูลจากการคัดเลือกคุณลักษณะพบว่าแบบการคัดเลือกแบบวิธี InfoGainAttributeEval จากจำนวน 21 คุณลักษณะ เหลือ 7 คุณลักษณะ ได้แก่ 1) เกรดเฉลี่ยกลุ่มสาระภาษาไทย 2) เกรดเฉลี่ยกลุ่มสาระคณิตศาสตร์ 3) เกรดเฉลี่ยกลุ่มสาระวิทยาศาสตร์ 4) เกรดเฉลี่ยกลุ่มสาระภาษาต่างประเทศ 5) เกรดเฉลี่ยกลุ่มสาระสังคมศึกษาศาสนาและวัฒนธรรม 6) เกรดเฉลี่ยกลุ่มสาระการงานอาชีพและเทคโนโลยี และ 7) เกรดเฉลี่ยสะสม 6 ภาคเรียน ทำการทดสอบประสิทธิภาพของแบบจำลอง

แบบ 10-fold cross validation ร่วมกับอัลกอริธึม Random Forest พบว่าประสิทธิภาพค่าความถูกต้อง ร้อยละ 73.20 สูงสุดเมื่อเทียบกับอัลกอริธึมอื่น

1.2 ผลการประเมินประสิทธิภาพระบบพยากรณ์เพื่อสนับสนุนการตัดสินใจศึกษาต่อหลังสำเร็จมัธยมศึกษาตอนต้น โดยใช้อัลกอริธึมป่าสุ่ม พบว่าอัลกอริธึมป่าสุ่ม มีประสิทธิภาพค่าความถูกต้อง (Accuracy) ร้อยละ 73.20 ค่าความแม่นยำ (Precision) ร้อยละ 73.30 ค่าครบถ้วน (Recall) ร้อยละ 73.20 ค่าประสิทธิภาพโดยรวม (F-Measure) ร้อยละ 73.20

1.3 ผู้วิจัยได้พัฒนาระบบพยากรณ์เพื่อสนับสนุนการตัดสินใจศึกษาต่อหลังสำเร็จมัธยมศึกษาตอนต้น โดยใช้อัลกอริธึมป่าสุ่ม โดยใช้งานผ่านเว็บเบราว์เซอร์ เพื่อให้นักเรียนและครูที่ปรึกษาหรือครูแนะแนวใช้ในการพยากรณ์แนวทางสนับสนุนการตัดสินใจในการศึกษาต่อหลังสำเร็จมัธยมศึกษาปีที่ 3

2. อภิปรายผล

ในการวิจัยครั้งนี้ได้นำการตามวัตถุประสงค์ของการวิจัยคือ สร้างแบบจำลองพยากรณ์เพื่อสนับสนุนการตัดสินใจศึกษาต่อหลังสำเร็จมัธยมศึกษาตอนต้น โดยใช้อัลกอริธึมป่าสุ่ม ด้วยการใช้ชุดข้อมูลผลสัมฤทธิ์ทางการเรียนของนักเรียนชั้นมัธยมศึกษาปีที่ 3 ย้อนหลังจำนวน 6 ภาคเรียน โดยการคัดเลือกคุณลักษณะแบบวิธี InfoGainAttributeEval และแบบไม่คัดเลือกคุณลักษณะร่วมกับอัลกอริธึม C4.5(J48) อัลกอริธึม Random Tree และ อัลกอริธึม Random Forest แบบ 10-fold cross validation เพื่อประเมินประสิทธิภาพของแบบจำลองพยากรณ์ผลสัมฤทธิ์ทางการเรียน ซึ่งผลการวิจัยพบว่าแบบคัดเลือกคุณลักษณะแบบวิธี InfoGainAttributeEval ของแบบจำลองที่สร้างขึ้นด้วยอัลกอริธึม Random Forest มีค่าความถูกต้อง (Accuracy) ร้อยละ 73.20 ซึ่งมากกว่าแบบจำลองที่สร้างด้วยอัลกอริธึม C4.5(J48) ซึ่งมีค่าความถูกต้อง 72.23 และแบบจำลองที่สร้างด้วยอัลกอริธึม Random Tree ซึ่งมีค่าความถูกต้อง 71.94 เนื่องจากอัลกอริธึม Random Forest มีกระบวนการนำชุดของข้อมูลที่ไม่เหมือนกัน โมเดลแต่ละตัวจะเรียนรู้อย่างเป็นอิสระต่อกัน แล้วเอาผลการตัดสินใจของโมเดลที่ทำนายจากเสียงส่วนมากมาใช้ในการตัดสินใจ ผลการวิจัยครั้งนี้สอดคล้องกับงานวิจัยที่ใช้การพยากรณ์ด้วยเทคนิคเหมืองข้อมูลเช่นเดียวกับ เสกสรรค์ วิลัยลักษณ์ (2558) ใช้เทคนิคการทำเหมืองข้อมูลเพื่อพยากรณ์ผลการเรียนของนักเรียนเพื่อพัฒนาคลังข้อมูลและสร้างตัวแบบพยากรณ์ผลการเรียนของนักเรียนโรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกาแพงแสน ผู้วิจัยจึงเลือกใช้อัลกอริธึม Random Forest ไปทดสอบกับชุดข้อมูลจำนวน 1,030 เรคคอร์ด ซึ่งสอดคล้องกับงานวิจัยของ กฤษฎิ์ บรรณะชัยศิริสุข (2558) ได้ทำการศึกษา

เครื่องมือท่องเว็บไซต์บนอุปกรณ์เคลื่อนที่สำหรับผู้พิการทางสายตาโดยใช้การตรวจหาเนื้อหาหลัก ผลการศึกษาพบว่า ตัวจำแนกที่สร้างด้วยวิธีป่าสุ่มด้วยการใช้วิธี 10-fold cross validation มีความถูกต้องสูงสุด และวันวิสาข์ ชนะประเสริฐ (2559) ได้ทำการประยุกต์ใช้เทคนิคเหมืองข้อมูลเพื่อแนะนำอาชีพสำหรับนักศึกษาปริญญาตรี คณะโบราณคดี มหาวิทยาลัยศิลปากร ผลการวิจัยพบว่าประสิทธิภาพแบบจำลองที่ดีที่สุดโดยการคัดเลือกคุณลักษณะแบบวิธี InfoGainAttributeEval มีค่าความถูกต้องของการจำแนกข้อมูลสูงที่สุด และ วัชรีวรรณ จิตต์สกุล (2560) ทำการวิเคราะห์การจำแนกข้อความด้วยการเปรียบเทียบความเสถียรของอัลกอริธึม โดยการศึกษาหาความเสถียรของอัลกอริธึมเพื่อการจำแนก ผลการศึกษาพบว่าอัลกอริธึมป่าสุ่ม มีค่าความเสถียรการจำแนกได้มากกว่าอัลกอริธึมอื่นๆ ทำการทดสอบประสิทธิภาพของแบบจำลอง (Model) แบบ 10-fold cross validation ผลการทดสอบมีค่าความถูกต้อง เท่ากับ 73.20 % และผลที่พยากรณ์ไม่ถูกต้อง เท่ากับ 27.76% จากผลการทดลองผู้วิจัยจึงเลือกใช้แบบจำลองที่สร้างขึ้นโดยใช้อัลกอริธึม Random Forest และการคัดเลือกคุณลักษณะแบบวิธี InfoGainAttributeEval มาพัฒนาเป็นระบบพยากรณ์เพื่อสนับสนุนการตัดสินใจศึกษาต่อหลังสำเร็จมัธยมศึกษาตอนต้น โดยใช้อัลกอริธึมป่าสุ่ม โรงเรียนบ้านปางสุด สังกัดสำนักงานเขตพื้นที่การศึกษาประถมศึกษานครสวรรค์ เขต 2 จังหวัดนครสวรรค์

3. ข้อเสนอแนะ

3.1 ข้อเสนอแนะในการนำผลการวิจัยไปใช้

ครูแนะแนวหรือครูที่ปรึกษานักเรียนระดับชั้นมัธยมศึกษาปีที่ 3 สามารถนำเอาระบบพยากรณ์เพื่อสนับสนุนการตัดสินใจศึกษาต่อหลังสำเร็จมัธยมศึกษาตอนต้น โดยใช้อัลกอริธึมป่าสุ่ม ในการวิจัยครั้งนี้ไปใช้ประกอบการแนะนำแนวทางในการศึกษาต่อของนักเรียนได้

3.2 ข้อเสนอแนะในการวิจัยครั้งต่อไป

3.2.1 ในการศึกษาวิจัยครั้งต่อไปควรนำไปพัฒนาเป็นโมบายแอปพลิเคชันเพื่อให้นักเรียนเข้าถึงระบบพยากรณ์ได้สะดวกยิ่งขึ้น

3.2.2 ควรมีการนำเอาเทคนิคเหมืองข้อมูลหรืออัลกอริธึมอื่นๆ เช่น Naive Bayes, k-Nearest Neighbo, Support Vector Machines มาเปรียบเทียบในระบบพยากรณ์ผลสัมฤทธิ์ทางการเรียนได้

3.2.3 ในการศึกษาวิจัยครั้งต่อไปควรนำเอากิจกรรมพัฒนาผู้เรียนมาใช้ในการวิเคราะห์ผลการเรียนของนักเรียน

บรรณานุกรม

บรรณานุกรม

กชกร ณ นครพนม. (2558). “การทำเหมืองข้อมูลด้วยการจำแนกประเภทและการพยากรณ์” ใน *ประมวลสาระชุดวิชาคลังข้อมูล เหมืองข้อมูล และธุรกิจอัจฉริยะ หน่วยที่ 9-1* . นนทบุรี: มหาวิทยาลัยสุโขทัยธรรมาธิราช.

กฤษฎิ์ บรรณะชัยศิริสุข (2558). *เครื่องมือท่องเว็บไซต์บนอุปกรณ์เคลื่อนที่สำหรับผู้พิการทางสายตาโดยใช้การตรวจหาเนื้อหาหลัก* (วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต). จุฬาลงกรณ์มหาวิทยาลัย, กรุงเทพฯ.

กานต์สินี เจริญกิจวัชรชัย (2561). *ปัญญาประดิษฐ์กับการพยากรณ์ค่าจ้างแรงงานไทย* (ค้นคว้าอิสระวิทยาศาสตรมหาบัณฑิต). มหาวิทยาลัยธรรมศาสตร์, กรุงเทพฯ.

กิตติศักดิ์ สุมามาลย์ (2555). *การคัดกรองสุขภาพเบื้องต้นโดยใช้เทคนิคเหมืองข้อมูล* (สารนิพนธ์วิทยาศาสตรมหาบัณฑิต). มหาวิทยาลัยธุรกิจบัณฑิตย์, กรุงเทพฯ.

โกเมศ อัมพวัน (2559, 23 มีนาคม). การจำแนกประเภทและการทำนายข้อมูล. *การทำเหมืองข้อมูล* น. 1-11 สืบค้นจาก https://staff.informatics.buu.ac.th/~komate/886464/%5B6%5D-Classification.pdf

ขจีจิตร ยุระตา. (ม.ป.ป). *เส้นทางการศึกษาต่อหลังจบ ม 3.* เข้าถึงได้จาก https://www.glurgeek.com/education/datawranglingwithpandas/

จามรกุล หล่าเกียรติกุล (2558). *เหมืองข้อมูลเบื้องต้น* (พิมพ์ครั้งที่ 1). กรุงเทพฯ: แคเน็กซ์อินเตอร์คอร์ปอเรชั่น.

ฐิติมา ช่วงชัย (2559). การวิเคราะห์หารูปแบบการเรียนรู้โดยใช้เหมืองข้อมูลของนักศึกษาต่อการจัดทำปริญญานิพนธ์ *วารสารบัณฑิตศึกษามหาวิทยาลัยราชภัฏวไลยอลงกรณ์ในพระบรมราชูปถัมภ์* 2559(2), 53-62.

ทัศนีย์ เพียรทาดี (2558). *การพยากรณ์ผลสัมฤทธิ์ทางการเรียนด้วยเทคนิคเหมืองข้อมูล* (วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต). วิทยาลัยอาชีวศึกษาเทคนิคบริหารธุรกิจกรุงเทพ, กรุงเทพฯ.

ทิพย์หทัย ทองธรรมชาติ (2554). *การพยากรณ์ผลสัมฤทธิ์ทางการเรียนด้วยเทคนิคเหมืองข้อมูล* (วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต). มหาวิทยาลัยเชียงใหม่, เชียงใหม่.

ปรเมษฐ์ ธันวานนท์และคณะ (2560). การประยุกต์ใช้โมเดลการเรียนรู้แบบรวมกลุ่มเพื่อพยากรณ์แนวโน้มของราคาหลักทรัพย์ในตลาดหลักทรัพย์แห่งประเทศไทย *วารสารวิทยาการและเทคโนโลยีสารสนเทศ* 2560(7), 12-21.

พิชัย ระเวงวัน (2560). โมเดลเพื่อการพยากรณ์สถานภาพทางการศึกษาของนักศึกษา *การประชุมวิชาการเสนอผลงานวิจัยบัณฑิตศึกษา ระดับชาติและนานาชาติ มหาวิทยาลัยขอนแก่น* 2560, 273-283.

ภูริพัทร์ ทองคำ (2559). *อัลกอริทึมแบบรวมสำหรับการเลือกคุณสมบัติของข้อมูล* (วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต). มหาวิทยาลัยธรรมศาสตร์, กรุงเทพฯ.

รัชพล กลัดชื่น. (2561). การเปรียบเทียบประสิทธิภาพอัลกอริทึมและการคัดเลือกคุณลักษณะที่เหมาะสม เพื่อการทำนายผลสัมฤทธิ์ทางการเรียนของนักศึกษาระดับอาชีวศึกษา *วารสารวิจัยมหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี* 2561(17), 1-10.

รุ่งนภา ศรีประโค (2556). *การลดปริมาณการขาดแคลนสินค้าโดยใช้เทคนิคการพยากรณ์ กรณีศึกษา บริษัท ไอเซโล (ประเทศไทย) จำกัด* (วิทยานิพนธ์วิศวกรรมศาสตรมหาบัณฑิต). มหาวิทยาลัยธุรกิจบัณฑิตย์, กรุงเทพฯ.

รุจิรา ธรรมสมบัติ (2555). *ระบบสนับสนุนการตัดสินใจในการเลือกใช้แพคเกจอินเทอร์เน็ตมือถือโดยใช้ต้นไม้ตัดสินใจ* (รายงานการวิจัย). วิทยาลัยลัยราชพฤกษ์, กรุงเทพฯ.

วัชรีวรรณ จิตต์สกุล. (2560). การวิเคราะห์การจำแนกข้อความด้วยการเปรียบเทียบความเสถียรของอัลกอริทึม *วารสารวิจัยและพัฒนามหาวิทยาลัยราชภัฏบ้านสมเด็จเจ้าพระยา* 2560(9), 19-31.

วันวิสาข์ ชนะประเสริฐ (2559). *การประยุกต์ใช้เทคนิคเหมืองข้อมูลเพื่อแนะนำอาชีพสำหรับนักศึกษาปริญญาตรี คณะโบราณคดี มหาวิทยาลัยศิลปากร* (วิทยานิพนธ์ศิลปศาสตรมหาบัณฑิต). มหาวิทยาลัยศิลปากร, กรุงเทพฯ.

วิภา เจริญภัณฑารักษ์. (2558). “หลักการพื้นฐานของการทำเหมืองข้อมูล” ใน *ประมวลสาระชุดวิชาคลังข้อมูล เหมืองข้อมูล และธุรกิจอัจฉริยะ หน่วยที่ 8-1* . นนทบุรี : มหาวิทยาลัยสุโขทัยธรรมาธิราช.

สุรวัชร ศรีเปารยะ (2560). การเปรียบเทียบประสิทธิภาพวิธีการจำแนกกลุ่มการเป็นโรคไตเรื้อรังกรณีศึกษาโรงพยาบาลแห่งหนึ่งในประเทศอินเดีย *วารสารวิทยาศาสตร์และเทคโนโลยี* 2560(5), 841-853.

เสกสรรค์ วิลัยลักษณ์. (2558). *การใช้เทคนิคการทำเหมืองข้อมูลเพื่อพยากรณ์ผลการเรียนของนักเรียน โรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน ศูนย์วิจัยและพัฒนาการศึกษา* (วิทยานิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต). มหาวิทยาลัยสุโขทัยธรรมาธิราช, นนทบุรี.

เอกสิทธิ์ พัชรวงค์ศักดิ์ดา. (2557). *การวิเคราะห์ข้อมูลด้วยเทคนิค ดาต้า ไมน์นิ่ง เบื้องต้น*. (พิมพ์ครั้งที่ 2). กรุงเทพฯ: เอเชีย ดิจิตอลการพิมพ์.

Aubrey, M. M. (2017). Application of Data Mining techniques to identify significant patterns in the Grade 12 results of the Free State Department of Education. (Master's thesis). Stellenbosch University.

Chakure, A. (2019). *Random Forest Regression*. Retrieved from https://medium.com/swlh/random-forest-and-its-implementation-71824ced454f

Glassboxmedicine. (2019). *Measuring Performance: The Confusion Matrix*. Retrieved from https://glassboxmedicine.com/2019/02/17/measuring-performance-the-confusion-matrix/

GlurGeek. (2021). *Python Data Wrangling with Pandas for Data Analysis*. Retrieved from https://www.glurgeek.com/education/datawranglingwithpandas

Han, J., Kamber, M., and Pei, J. (2006). *Data Mining: Concepts and Techniques*. 2nd ed. San Franccisco: Morgan Kaufmann Publishers is an imprint of Elsevier.

Hilario, M., & Kalousis, A. (2008). Approaches to dimensionality reduction in proteomic biomarker studies. *Briefings in Bioinformatics*, 9(2), 102-118.

Janecek, A. (2009). *Efficient feature reduction and classification methods: applications in drug discovery and email categorization*. Doctoral dissertation, University of Vienna.

Karlsson, I. (2017*). Order in the random forest*. Retrieved from https://www.diva-portal.org/smash/get/diva2:1090364/FULLTEXT01.pdf

Kevin, P. (2018). *Latent class analysis and random forest ensemble to identify ay-risk students in higher education.* (Master's thesis). San Diego State University.

Korivi, K. (2016). *Identifying Poverty-Driven Need By Augmenting Census And Community Survey Data.* (Master of Science). Kansas state university, Kansas.

Louppe, G. (2014). *Understanding Random Forests from theory to practice*. Retrieved from https://arxiv.org/pdf/1407.7502.pdf

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.*

Ofer D. (2015). *Machine Learning for Protein Function* (Master's thesis). Jerusalem University, Jerusalem.

Quinlan, J.R. (1992). Learning with Continuous Classes. *Proceedings of Australian Joint Conference on Artificial Intelligence*, 343-348.

Rincon, R. (2019). *Quarterly Multidimensional Poverty Predictions in Mexico using Machine Learning Algorithms.* Retrieved form https://sistemas.colmex.mx/Reportes/LACEALAMES/LACEA-LAMES2019_paper_754.pdf

Saeys, Y., Inza, I., & Larranaga, P. (2007). A review of feature selection techniques in bioinformatics. *Bioinformatics*, 23(19), 2507–2517.

Shearer, C. (2000). The CRISP-DM Model: The New Blueprint for Data Mining. *Journal of Data Warehousing*, 3(5), 13-22.

Tan, P., Steinbach., M., and Kumar., V. (2005). *Introduction to data Mining*. 2nd ed. New York: Pearson Education, 146.

Wikipedia. (n.d.). *Cross-industry standard process for data mining*. Retrieved from https://en.wikipedia.org/wiki/Cross_Industry_Standard_Process_for_Data_Mining

Xu, J., and 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.

Yu, L., and Liu, H. (2004). Efficient feature selection via analysis of relevance and redundancy. *Journal of Machine Learning Research*, 5,1205–1224.

ภาคผนวก

ภาคผนวก ก

การติดตั้งโปรแกรมจำลองเว็บเซิฟเวอร์

การติดตั้งโปรแกรมจำลองเว็บเซิฟเวอร์

เครื่องมือที่ใช้ในการจัดเก็บข้อมูล คือ โปรแกรม AppServ ซึ่งมีขั้นตอนการติดตั้งโปรแกรมดังนี้

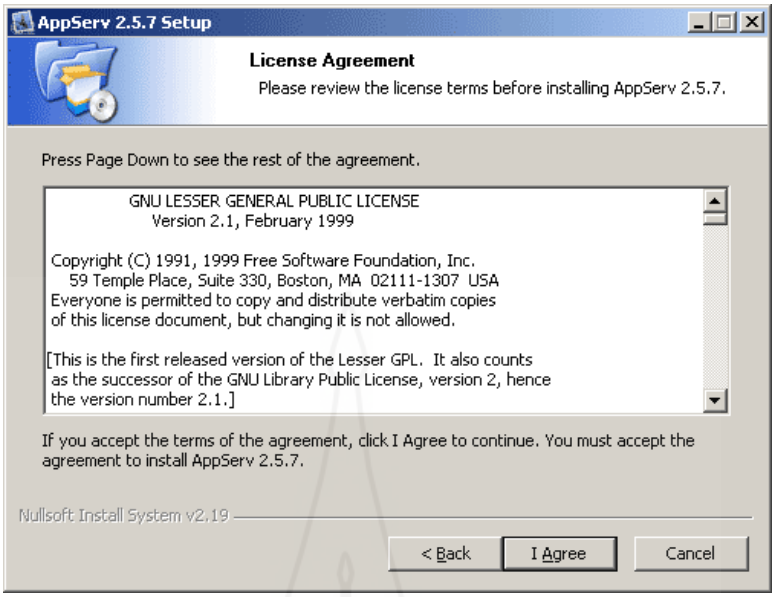
1. ทำการดาวน์โหลดโปรแกรม AppServ จากเว็บไซต์แล้วทำการติดตั้ง

ภาพที่ 1.1 ไฟล์ติดตั้งโปรแกรม AppServ

2. เริ่มติดตั้งโปรแกรม โปรแกรมจะแสดงหน้าหลักของโปรแกรมขึ้นมาให้ทำการเลือก “Next”

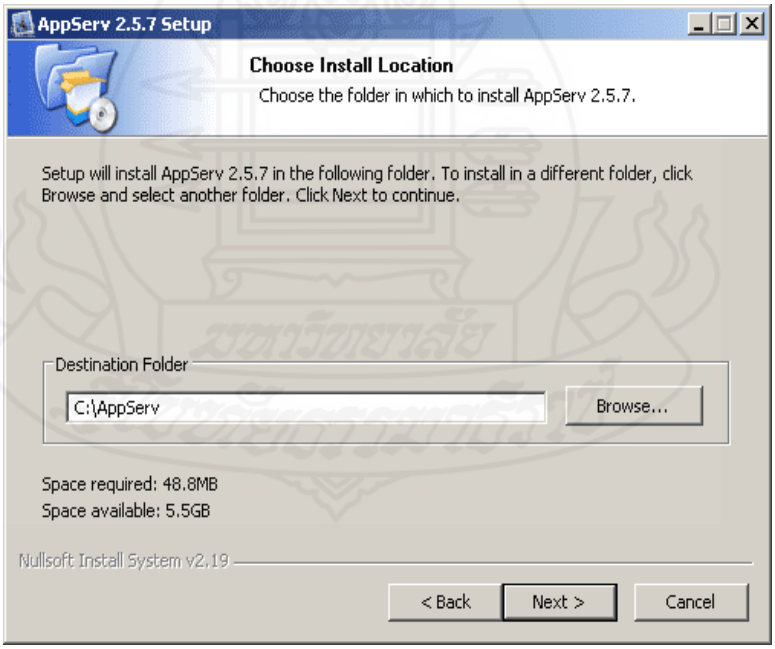
ภาพที่ 1.2 แสดงหน้าหลักในการติดตั้งโปรแกรม AppServ

3. เข้าสู่ขั้นตอนเงื่อนไขและรายละเอียดการใช้งานโปรแกรม AppServ หากยอมรับเงื่อนไขให้กด “I Agree” เพื่อเข้าสู่การติดตั้งในขั้นต่อไป



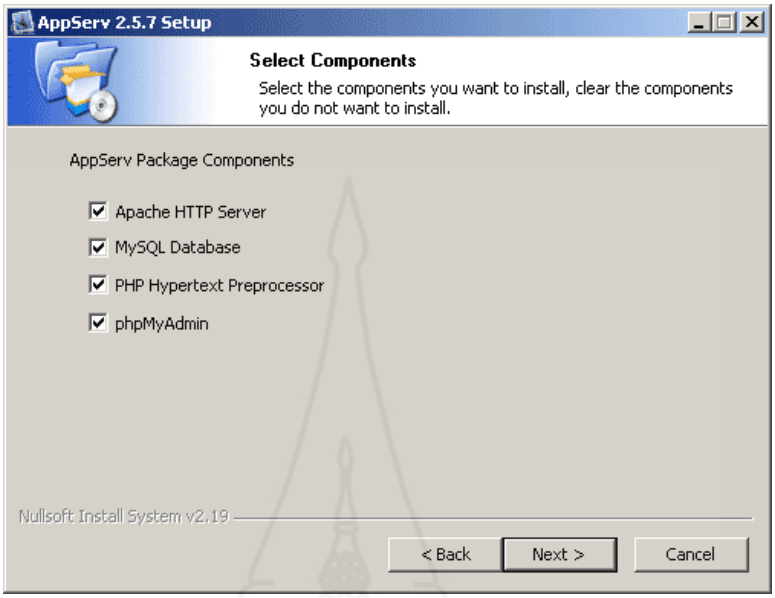
ภาพที่ 1.3 เงื่อนไขและรายละเอียดการใช้งานโปรแกรม AppServ

4. กำหนดที่อยู่ปลายทางที่ต้องการติดตั้งของโปรแกรม จากนั้นคลิก “Next”



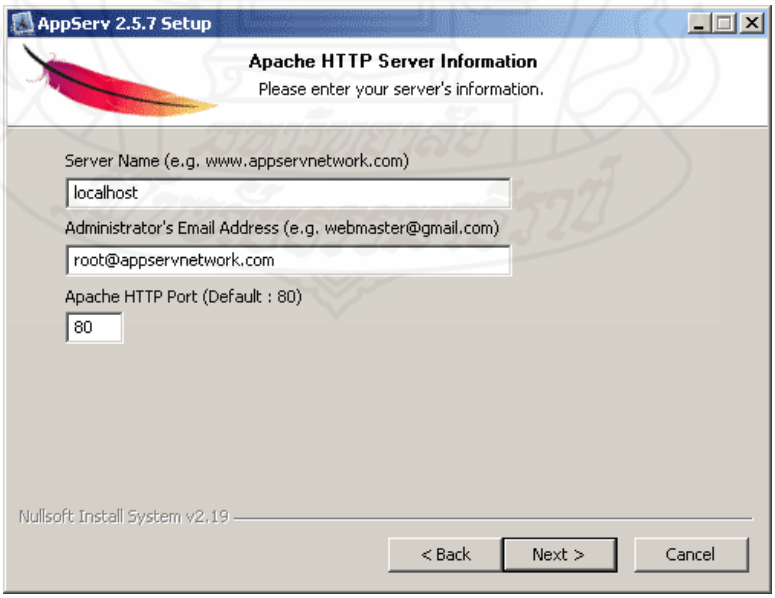
ภาพที่ 1.4 กำหนดที่อยู่ปลายทางของโปรแกรม

5. เลือก Package Components ที่ต้องการติดตั้งทั้งหมด จากนั้นคลิก “Next”



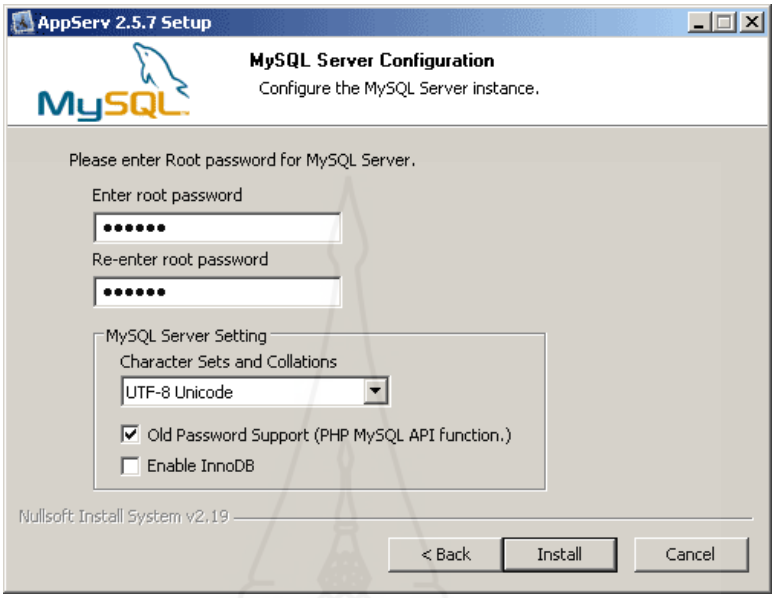
ภาพที่ 1.5 หน้าจอเลือกโปรแกรมที่ต้องการติดตั้ง

6. กำหนดค่าของ Apache Web Server ประกอบด้วย 3 ส่วน คือ 1) Server Name คือช่องสำหรับป้อนข้อมูลชื่อ Web Server 2) Admin Email คือ ช่องสำหรับป้อนอีเมล์ผู้ดูแลระบบ 3) HTTP Port คือช่องสำหรับระบุ Port ที่จะเรียกใช้งาน Apache Web Server



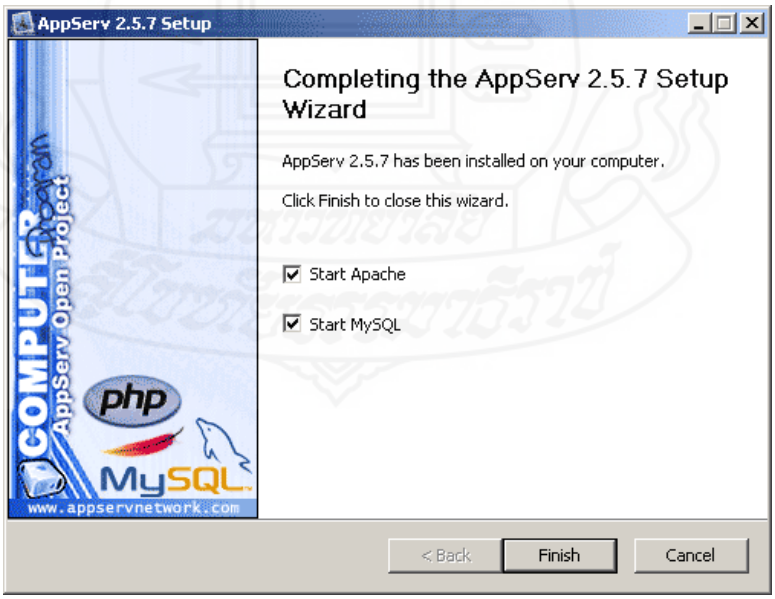
ภาพที่ 1.6 การกำหนดค่าของ Apache Web Server

7. กำหนดค่ารหัสผ่านการเข้าใช้งานฐานข้อมูลและกำหนดค่าภาษาที่ใช้ในการจัดเก็บฐานข้อมูล



ภาพที่ 1.7 หน้ากำหนดรหัสผ่านการเข้าใช้งานฐานข้อมูล

8. เสร็จสิ้นขั้นตอนการติดตั้งโปรแกรม AppServ จากนั้นคลิก “Finish”



ภาพที่ 1.8 หน้าจอการติดตั้งโปรแกรมเสร็จสิ้น

ภาคผนวก ข

ชุดข้อมูล (data set)

ชุดข้อมูล (data set)

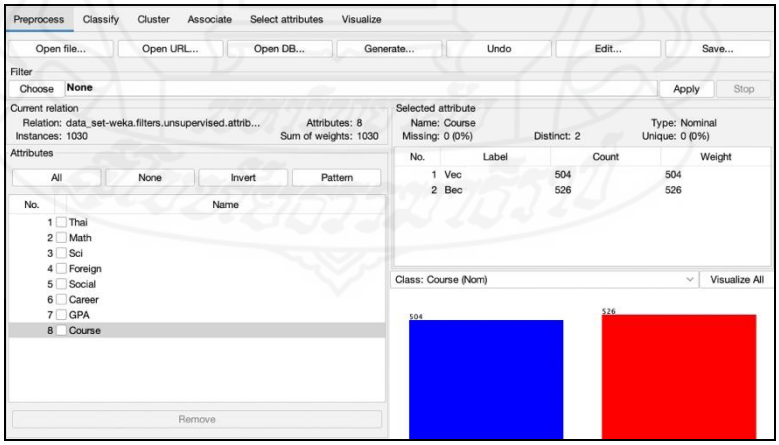
1. เอาชุดข้อมูลที่ผ่านกระบวนการการเตรียมข้อมูล (Data Preparation) จำนวน 1,030 เรคคอร์ด 11 คุณลักษณะ ทำการแปลงชุดข้อมูลจากไฟล์ .csv เป็นไฟล์ .arff เพื่อใช้กับโปรแกรม WEKA

Relation: data_set

No.	1: Thai Nominal	2: Math Nominal	3: Sci Nominal	4: Foreign Nominal	5: Social Nominal	6: Career Nominal	7: GPA Nominal	8: **Course** Nominal
1	Medium	Medium	Medium	Medium	High	Medium	Good	Vec
2	Medium	Medium	Medium	Low	High	High	Good	Vec
3	Medium	Medium	Medium	Low	High	High	Good	Vec
4	Medium	Medium	Medium	Low	High	High	Good	Vec
5	Medium	Medium	Medium	Low	High	High	Good	Vec
6	Medium	Medium	Medium	Low	Medium	Medium	Above average	Vec
7	Medium	Medium	Medium	Low	Medium	Medium	Good	Vec
8	Medium	Medium	Medium	Low	Medium	Medium	Good	Vec
9	Medium	Medium	Medium	Low	Medium	Medium	Above average	Vec
10	Medium	Medium	High	Medium	Medium	Medium	Above average	Vec
11	Medium	Medium	Medium	Low	Medium	High	Good	Vec
12	Medium	Medium	High	Medium	Medium	Medium	Above average	Vec
13	Medium	Medium	Medium	Low	Medium	Medium	Above average	Vec
14	Medium	Medium	Medium	Low	Medium	Medium	Above average	Vec
15	Medium	Medium	Low	Medium	Medium	High	Good	Vec
16	Medium	Medium	Medium	Medium	Medium	Medium	Good	Vec
17	Medium	Medium	Medium	High	High	Medium	Good	Vec
18	Medium	Medium	Medium	Medium	Medium	Medium	Good	Vec
19	Medium	Medium	Medium	Medium	Medium	Medium	Good	Vec

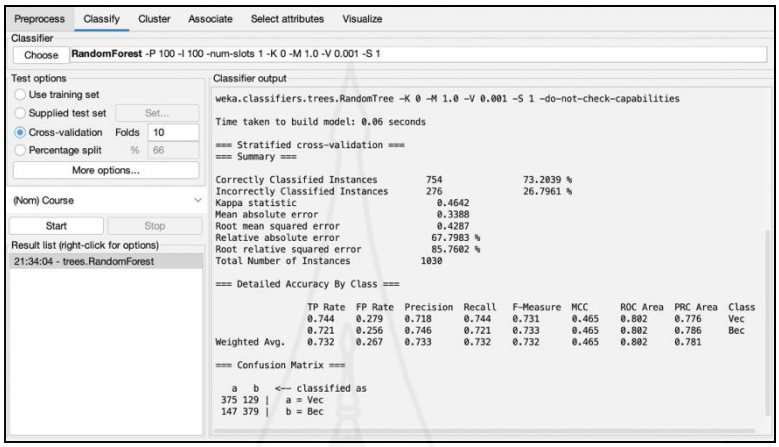
ภาพที่ 2.1 แสดงชุดข้อมูล

2. นำชุดข้อมูลเข้าสู่การสร้างตัวแบบพยากรณ์ (Modeling) ด้วยโปรแกรม WEKA



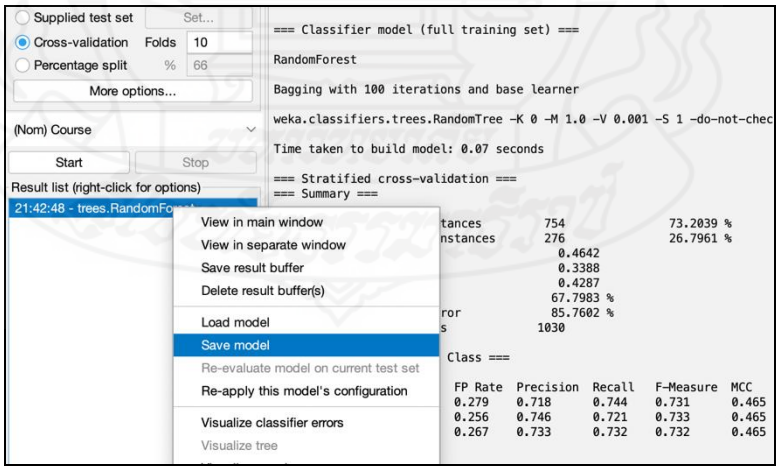
ภาพที่ 2.2 แสดงหน้าจอการนำชุดข้อมูลเข้าโปรแกรม WEKA

3. นำชุดข้อมูลไปวิเคราะห์คลิกที่แท็บ “classify” โดยคลิกที่ “choose” เลือกอัลกอริธึมป่าสุ่ม RandomForest และเลือกแบบทดสอบประสิทธิภาพของแบบจำลองแบบ 10-fold cross validation คลิก “start” โปรแกรมจะทำนายและแสดงผลลัพธ์ในส่วน classifier output



ภาพที่ 2.3 หน้าจอแสดงการวิเคราะห์ข้อมูล

4. เมื่อทำการสร้างโมเดล Classification ด้วยใช้อัลกอริธึมป่าสุ่ม RandomForest เสร็จเรียบร้อยแล้ว ทำการบันทึกโมเดลเพื่อนำไปใช้ในการพัฒนาระบบพยากรณ์เพื่อสนับสนุนการตัดสินใจศึกษาต่อหลังสำเร็จมัธยมศึกษาตอนต้น โดยใช้อัลกอริธึมป่าสุ่ม



ภาพที่ 2.4 แสดงการบันทึกโมเดล

ภาคผนวก ค

คู่มือการใช้งานระบบพยากรณ์เพื่อสนับสนุนการตัดสินใจ

คู่มือการใช้งานระบบพยากรณ์เพื่อสนับสนุนการตัดสินใจ

วิธีการใช้งานระบบพยากรณ์เพื่อสนับสนุนการตัดสินใจศึกษาต่อหลังสำเร็จมัธยมศึกษาตอนต้น โดยใช้อัลกอริธึมป่าสุ่ม ดังนี้

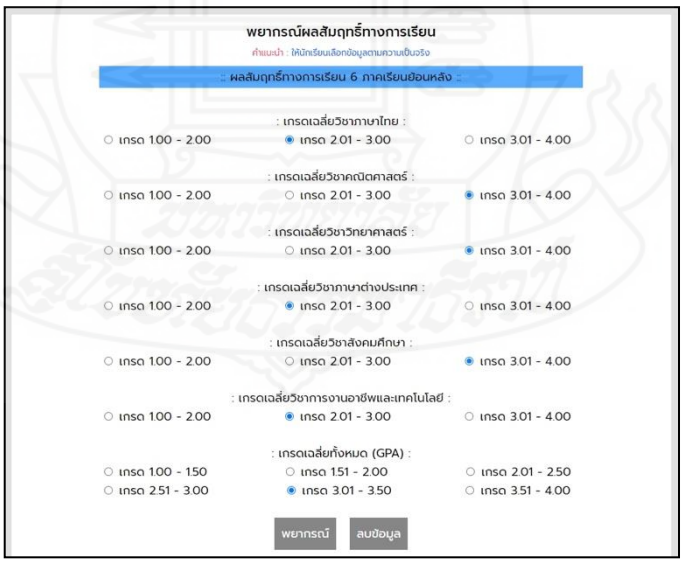
1. เข้าสู่หน้าเว็บของระบบพยากรณ์เพื่อสนับสนุนการตัดสินใจศึกษาต่อหลังสำเร็จมัธยมศึกษาตอนต้น โดยใช้อัลกอริธึมป่าสุ่ม ผ่านทางเว็บเบราว์เซอร์

ภาพที่ 3.1 แสดงหน้าแรกของระบบพยากรณ์ผลสัมฤทธิ์ทางการเรียน

2. ผู้ใช้เข้าสู่หน้าจอพยากรณ์ผลสัมฤทธิ์เพื่อเลือกข้อมูลของผู้ใช้งานในการพยากรณ์ผลสัมฤทธิ์ทางการเรียนใน 6 ภาคเรียนที่ผ่านมา

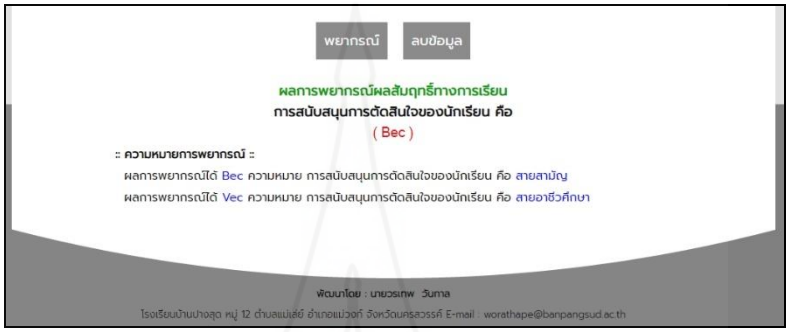
ภาพที่ 3.2 แสดงหน้าการเลือกข้อมูลที่ใช้ในการพยากรณ์

3. หลังจากผู้ใช้เลือกข้อมูลในการพยากรณ์ผลสัมฤทธิ์ทางการเรียนครบถ้วนแล้ว ให้คลิกปุ่ม “พยากรณ์” เพื่อแสดงผลลัพธ์แนวทางในการศึกษาต่อของนักเรียนอยู่ในระดับสายสามัญหรือสายอาชีวศึกษา หากต้องการลบข้อมูลหรือพยากรณ์ผลสัมฤทธิ์ทางการเรียนใหม่ให้คลิกปุ่ม “ลบข้อมูล” ดังภาพที่ 1.3

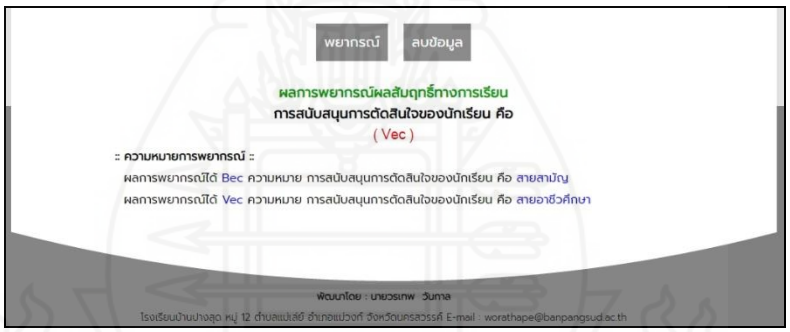


ภาพที่ 3.3 เลือกข้อมูลที่ใช้ในการพยากรณ์

4. เมื่อผู้ใช้คลิกปุ่ม “พยากรณ์” ระบบจะแสดงผลลัพธ์การพยากรณ์แนวทางในการศึกษาต่อของผู้ใช้งาน หากผลพยากรณ์ได้ “Bec” ความหมายคือ “การสนับสนุนการตัดสินใจของนักเรียน คือ สายสามัญ (Basic Education Commission : Bec)” และ “Vec” ความหมายคือ “การสนับสนุนการตัดสินใจของนักเรียน คือ สายอาชีวศึกษา (Vocational Education Commission : Vec)”



ภาพที่ 3.4 แสดงผลการพยากรณ์ของระบบการสนับสนุนการตัดสินใจในสายสามัญ



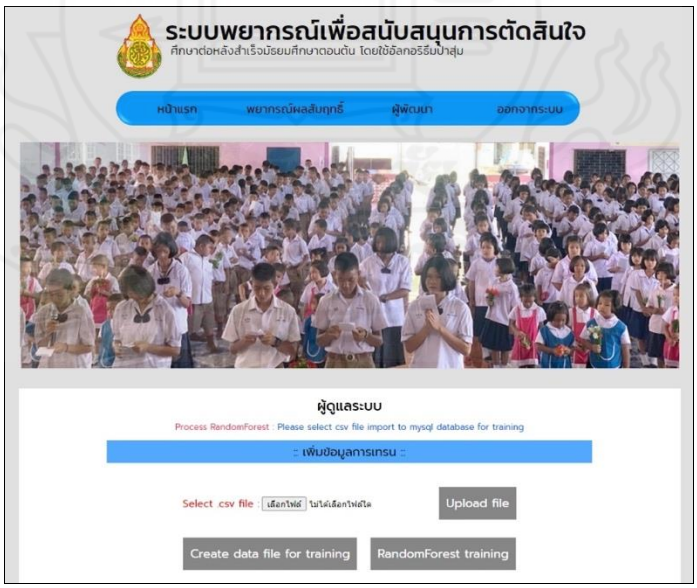
ภาพที่ 3.5 แสดงผลการพยากรณ์ของระบบการสนับสนุนการตัดสินใจในสายอาชีวศึกษา

วิธีการใช้งานระบบพยากรณ์เพื่อสนับสนุนการตัดสินใจศึกษาต่อหลังสำเร็จมัธยมศึกษาตอนต้น โดยใช้อัลกอริธึมป่าสุ่ม ส่วนของผู้ดูแลระบบมี ดังนี้

5. เข้าสู่หน้าเว็บของผู้ดูแลระบบพยากรณ์เพื่อสนับสนุนการตัดสินใจศึกษาต่อหลังสำเร็จมัธยมศึกษาตอนต้น โดยใช้อัลกอริธึมป่าสุ่ม ผ่านทางเว็บเบราว์เซอร์ ผู้ดูแลระบบเข้าสู่ระบบด้วยชื่อผู้ใช้งานและรหัสผ่านที่กำหนดไว้

ภาพที่ 3.6 หน้าผู้ดูแลระบบ

ผู้ดูแลระบบสามารถเพิ่มข้อมูลผลสัมฤทธิ์ทางการเรียนของนักเรียนในปีการศึกษาปัจจุบันเข้าสู่ฐานข้อมูลใช้ไฟล์ที่มีนามสกุลเป็น .csv ในการอัพโหลดเข้าสู่ระบบ จากนั้นทำการคลิกปุ่ม Create data file for training แล้วทำการคลิกปุ่ม Random Forrest training เป็นขั้นตอนสุดท้าย ดังภาพที่ 1.7



ภาพที่ 3.7 ระบบการจัดการข้อมูลผู้ดูแลระบบ

ประวัติผู้วิจัย

ชื่อ	นายวรเทพ วันกาล
วัน เดือน ปีเกิด	11 เมษายน 2526
สถานที่เกิด	อำเภอลอง จังหวัดแพร่
ประวัติการศึกษา	วิศวกรรมศาสตรบัณฑิต สาขาวิชาวิศวกรรมมัลติมีเดียและระบบอินเทอร์เน็ต มหาวิทยาลัยกรุงเทพ พ.ศ. 2552
สถานที่ทำงาน	โรงเรียนนารีรัตน์จังหวัดแพร่
ตำแหน่ง	ปฏิบัติงานด้านการเรียนการสอน