

การพยากรณ์การเลือกกลุ่มคณะวิชาในการศึกษาต่อในระดับปริญญาตรี ของมหาวิทยาลัยรังสิต โดยใช้เทคนิคการทำเหมืองข้อมูล*

Prediction of faculty group selection for further study
In bachelor degree by using data mining techniques

ฐิตินันท์ อนุสรณ์ (Thitinan Anusorn)**

สมชาย เล็กเจริญ (Somchai Lekcharoen)***

*วิทยานิพนธ์หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ วิทยาลัยนวัตกรรมการจัดการและเทคโนโลยีสารสนเทศ มหาวิทยาลัยรังสิต

**นักศึกษาปริญญาโท, หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ วิทยาลัยนวัตกรรมการจัดการและเทคโนโลยีสารสนเทศ มหาวิทยาลัยรังสิต, E-mail: thitinan.a@rsu.ac.th

***ผู้ช่วยศาสตราจารย์, ผู้อำนวยการหลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสื่อสังคม วิทยาลัยนวัตกรรมการจัดการและเทคโนโลยีสารสนเทศ มหาวิทยาลัยรังสิต, E-mail: somchai.l@rsu.ac.th

ได้รับบทความ: 4 พ.ย. 62 / แก้ไขปรับปรุง: 20 พ.ค. 63 / อนุมัติให้ตีพิมพ์: 4 มิ.ย. 63 / เผยแพร่ออนไลน์: 22 มิ.ย. 63

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบแบบจำลองการพยากรณ์ระหว่าง ต้นไม้การตัดสินใจ ในอ็ฟเบย์และการวิเคราะห์การถดถอยทางโลจิสติกและพยากรณ์การเลือกกลุ่มคณะวิชาในการศึกษาต่อระดับปริญญาตรี มหาวิทยาลัยรังสิต การวิจัยครั้งนี้ใช้โปรแกรม RapidMiner Studio เป็นเครื่องมือช่วยสำหรับการพยากรณ์การเลือกกลุ่มคณะวิชาในการศึกษาต่อมหาวิทยาลัยรังสิตและเปรียบเทียบแบบจำลองการพยากรณ์ ก่อนที่จะทำการพยากรณ์การเลือกกลุ่มคณะวิชาในการศึกษาต่อ นั้น ผู้วิจัยได้ทำการเปรียบเทียบแบบจำลองการพยากรณ์ด้วย ONE-WAY ANOVA แล้วพบว่าแบบจำลองทั้ง 3 แบบ มีระดับนัยสำคัญทางสถิติ = 0.92 แสดงว่าแบบจำลองการพยากรณ์ทั้ง 3 แบบ ไม่แตกต่างกันอย่างมีนัยสำคัญที่ระดับ 0.05

ผู้วิจัยจึงเลือกใช้ ต้นไม้การตัดสินใจเป็นแบบจำลองการพยากรณ์ เพื่อวิเคราะห์หาปัจจัย โดยใช้กลุ่มตัวอย่าง เป็นนักศึกษามหาวิทยาลัยรังสิตที่เข้าศึกษาต่อใน ปีการศึกษา 2559-2562 จำนวน 2,540 คน ผลจากการ วิเคราะห์พบว่า กลุ่มตัวอย่างที่เลือกเรียนกลุ่มคณะสายวิทยาศาสตร์ จะเป็นกลุ่มที่เรียนมัธยมศึกษาตอน ปลายในแผนการเรียน วิทยาศาสตร์-คณิตศาสตร์ มีเกรดเฉลี่ยมากกว่า 3.76 และบิดามีรายได้เฉลี่ยต่อปีเกิน 300,000 บาทต่อปี และกลุ่มตัวอย่างที่เลือกเรียนกลุ่มคณะสายศิลปะ จะเป็นกลุ่มที่เรียนมัธยมศึกษาตอน ปลายในแผนการเรียน ศิลป์-ภาษา และมีเกรดเฉลี่ยมากกว่า 1.41

การทำงานวิจัยในครั้งนี้สามารถนำข้อมูลที่ได้จากการวิเคราะห์นำไปใช้วางแผนในการแนะแนวการ เลือกกลุ่มคณะวิชาเพื่อศึกษาต่อในมหาวิทยาลัยรังสิตซึ่งจะช่วยให้ผู้เข้าศึกษาต่อตัดสินใจได้ง่ายขึ้นหรือนำ ข้อมูลที่ได้จากการวิเคราะห์มาประกอบการการตัดสินใจของผู้บริหารในการวางแผนทำการตลาด และการแนะแนวการศึกษาต่อไป

คำสำคัญ: ต้นไม้การตัดสินใจ ในอีฟเบย์ การวิเคราะห์การถดถอยทางโลจิสติก เหมือนข้อมูล

Abstract

This research is to compare the models between decision tree, naive bayes and logistic regression as well as the forecast of selected faculty for Rangsit University. RapidMiner Studio is used for this research as the instrument to help for the selection and compare model before forecast the faculty selection. The researcher has compared the model using ONE-WAY ANOVA. The result of three models are quite the same, the statistical significance is 0.92 which show that the these three models are not different with statistical significance at level 0.05. This research, the researcher selected decision tree model for factors analysis. The sampling is students of Rangsit University year 2016 - 2019 with quantity 2,540 people. The sampling analysis shows that the science selected group is the high school student, math-science plan with grade

average at 3.76 and family yearly incomes is more than THB 300,000. Art selected group is the high school student, art-language plan with grade average more than 1.41.

The data from this research can be used for the guidance faculty for the expected new entry of Rangsit University, as well as for the easier way of student selection. Moreover, the data can be used for management marketing planning for the education guidance as well

Keywords: Decision tree, Naïve Bayes, Logistic Regression, data mining

บทนำ

ในปัจจุบันประเทศไทยมีจำนวนสถาบันอุดมศึกษาที่อยู่ภายใต้การดูแลของสำนักงานคณะกรรมการการอุดมศึกษา กระทรวงศึกษาธิการ จำนวน 172 แห่ง โดยแบ่งเป็นสถาบันอุดมศึกษาของรัฐในระบบราชการ จำนวน 65 แห่ง สถาบันอุดมศึกษาในกำกับของรัฐบาล 15 แห่ง สถาบันอุดมศึกษาเอกชน จำนวน 71 แห่ง วิทยาลัยชุมชนจำนวน 21 แห่ง นอกจากนี้ยังมีสถาบันที่จัดการศึกษาโดยมหาวิทยาลัย, วิทยาลัยหรือบัณฑิตวิทยาลัย ที่ไม่ได้สังกัดสำนักงานคณะกรรมการการอุดมศึกษาอีกจำนวน 9 แห่ง (สมาคมสภามหาวิทยาลัย ประเทศไทย, 2562) เนื่องจากในปัจจุบันจำนวนสถาบันอุดมศึกษาที่มีอยู่ในประเทศไทยนั้นมีจำนวนมากซึ่งเป็นปัจจัยให้เกิดการแข่งขันทางการศึกษาสูงและอัตราการเกิดของคนไทยยังมีแนวโน้มที่ลดลง โดยข้อมูลจากสำนักงานสถิติแห่งชาติ พบว่ามีอัตราการเกิดย้อนหลังใน 5 ปี มีแนวโน้มลดลง คือ ในปี พ.ศ. 2557 มีอัตราการเกิดอยู่ที่ 776,370 คน ปี พ.ศ. 2558 จำนวน 736,352 คน ปีพ.ศ. 2559 จำนวน 704,058 คน ปี พ.ศ. 2560 จำนวน 702,755 คน และปี พ.ศ. 2561 จำนวน 666,109 คน (สำนักงานสถิติแห่งชาติ, 2562) จากปัจจัยดังกล่าวจึงส่งผลให้มีจำนวนผู้ที่เข้าศึกษาต่อในระดับปริญญาตรีในอนาคตลดลงตามไปด้วย มหาวิทยาลัยรังสิต เป็นสถาบันศึกษาอุดมศึกษาเอกชนแห่งหนึ่งในประเทศไทย ซึ่งจะได้รับผลกระทบจากจำนวนผู้ที่เข้าศึกษาต่อในระดับอุดมศึกษาที่มีแนวโน้มลดลง และปัจจุบันมีนักศึกษาจำนวนหนึ่งที่ลาออกกลางคันเนื่องจากเลือกกลุ่มคณะไม่ตรงกับความต้องการของตนเอง และพื้นฐานความรู้ไม่เพียงพอ

ดังนั้นผู้วิจัยได้นำการพยากรณ์การเลือกกลุ่มคณะในการ เข้าศึกษาต่อระดับอุดมศึกษา ด้วยเทคนิคการทำเหมืองข้อมูลมาวิเคราะห์เพื่อค้นหารูปแบบและความสัมพันธ์จากข้อมูลจำนวนมากโดยอาศัย

หลักการทางสถิติ และการเรียนรู้ด้วยเครื่องจักร(machine learning)กระบวนการทำงานของเหมืองข้อมูลแบบ Classification (หนึ่งหทัย ชัยอากร, 2560)

จากศึกษางานวิจัยเกี่ยวกับการใช้เทคนิคการทำเหมืองข้อมูลในการจำแนกและคัดเลือกแขนงวิชา สำหรับนักศึกษาคณะเทคโนโลยีสารสนเทศ (จิราภา เลหาหวรรนนท์, รัชต์ ลิ้มสุทธิวันภูมิ, และบัณฑิต ฐานะโสภณ, 2558) ผู้วิจัยจึงมีแนวความคิดที่จะพยากรณ์การเลือกกลุ่มคณะวิชาที่จะศึกษาต่อใน มหาวิทยาลัยรังสิต เพื่อนำข้อมูลดังกล่าวไปทำการตลาดได้ตรงกลุ่มเป้าหมายมากยิ่งขึ้นและช่วยลดอัตราการตกออกของนักศึกษาในการวิจัยครั้งนี้จำเป็นที่จะต้องศึกษาถึงเทคนิควิธีการพยากรณ์ที่เหมาะสมที่สุด โดยมีกลุ่มตัวอย่างเป็นนักศึกษามหาวิทยาลัยรังสิตที่เข้าศึกษาต่อในปีการศึกษา 2559-2562 จำนวน 2,540 คน (สำนักงานทะเบียน มหาวิทยาลัยรังสิต, 2562)

วัตถุประสงค์

1. เพื่อเปรียบเทียบแบบจำลองการพยากรณ์ระหว่างต้นไม้การตัดสินใจ (Decision Tree) โนอีฟเบย์ (Naïve Bayes) และการวิเคราะห์การถดถอยทางโลจิสติก (Logistic Regression)
2. เพื่อพยากรณ์การเลือกกลุ่มคณะวิชาในการศึกษาต่อในระดับอุดมศึกษาในมหาวิทยาลัยรังสิต โดยใช้เทคนิคการทำเหมืองข้อมูล

ขอบเขตของการวิจัย

1. ประชากรและกลุ่มตัวอย่าง

กลุ่มตัวอย่างที่ใช้ในการวิจัยครั้งนี้เป็นนักศึกษาที่เข้าศึกษาต่อในปีการศึกษา 2559-2562 ที่เลือกเรียนในกลุ่มคณะวิชาวิทยาศาสตร์ และกลุ่มคณะวิชาศิลปะของมหาวิทยาลัยรังสิต จำนวน 2,540 คน โดยแบ่งเป็นนักศึกษาที่เลือกเรียนในกลุ่มคณะวิชาศิลปะ จำนวน 1,164 คน และนักศึกษาที่เลือกเรียนในกลุ่มคณะวิชาวิทยาศาสตร์จำนวน 1,376 คน (สำนักงานทะเบียน มหาวิทยาลัยรังสิต, 2562) เพื่อพยากรณ์การเลือกกลุ่มคณะวิชาในการศึกษาต่อมหาวิทยาลัยรังสิตด้วยเทคนิคการทำเหมืองข้อมูล

2. ตัวแปรที่ศึกษา

- 2.1 แผนการเรียนในระดับชั้นมัธยมศึกษาตอนปลาย(PROGRAM)
- 2.2 เกรดเฉลี่ยรวมในระดับชั้นมัธยมศึกษาตอนปลาย(UGRAD_GPA)

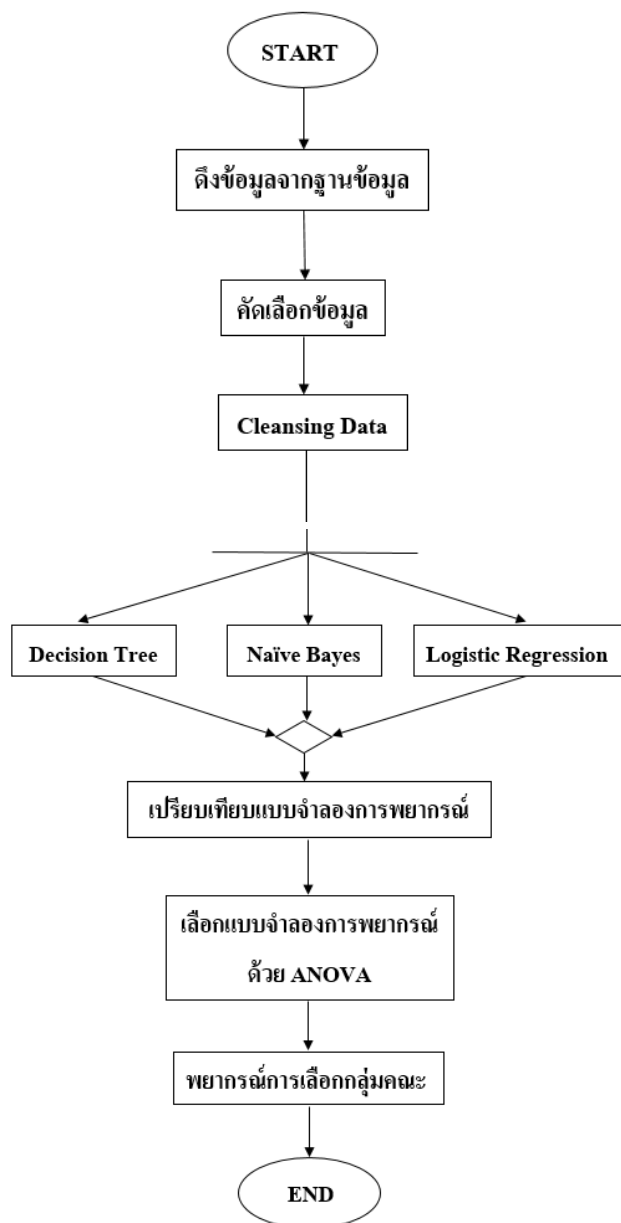
- 2.3 รายได้เฉลี่ยต่อปีของบิดา(FATH_INCOME)
- 2.4 รายได้เฉลี่ยต่อปีของมารดา(MATH_INCOME)
- 2.5 กลุ่มคณะวิชาที่เลือกศึกษาต่อระดับอุดมศึกษา(GROUP_TYPE)

อุปกรณ์และวิธีการ/วิธีการดำเนินการวิจัย

1. **ข้อมูลกลุ่มตัวอย่าง** การเตรียมข้อมูลกลุ่มตัวอย่างที่ใช้ในการวิจัยครั้งนี้เป็นนักศึกษาของมหาวิทยาลัยวังสิตที่เข้าศึกษาในปีการศึกษา 2559-2562 จำนวน 2,540 คน (สำนักงานทะเบียน มหาวิทยาลัยวังสิต, 2562) โดยนำข้อมูล ที่ได้มาผ่านกระบวนการตรวจสอบและแก้ไขข้อมูลที่ไม่ถูกต้องออกไปจากชุดข้อมูลเพื่อให้ได้ข้อมูลที่มีคุณภาพ ก่อนที่จะนำชุดข้อมูลนั้นไปประมวลผลเพื่อปัจจัยที่ส่งผลต่อการเลือกกลุ่มคณะวิชาในการศึกษาต่อในระดับอุดมศึกษา

2. **เทคนิคการทำเหมืองข้อมูล(Data Mining)** จิราภา เลาหะวรรณนท์, รชต ลิ้มสุทธิวันภูมิ, และ บัณฑิต ฐานะโสภณ (2558) กล่าวว่า เทคนิคการทำเหมืองข้อมูล(Data Mining) คือนำข้อมูลที่มีจำนวนมาก มาวิเคราะห์เพื่อสกัดสารสนเทศ รวมถึงรูปแบบความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลและสามารถนำผลสรุปที่เป็นประโยชน์ต่อการนำไปใช้ในการช่วยในการตัดสินใจที่เป็นประโยชน์ต่อองค์กร กระบวนการทำเหมืองข้อมูลสามารถทำได้หลายหลักการ เช่น เทคนิคการเรียนรู้ของเครื่อง(Machine Learning) สถิติ เป็นต้น

2.1 กระบวนการทำงานของ Knowledge Discovery in Database (KDD) หรือ Data Mining หมายถึง กระบวนการการค้นหาความรู้สารสนเทศจากข้อมูลที่มีอยู่จำนวนมาก เพื่อค้นหาความน่าสนใจของข้อมูลเหล่านั้น ซึ่งจะทำได้ข้อมูลที่เป็นเหตุเป็นผลที่สามารถเข้าใจได้ง่าย การนำเทคนิคเหมืองข้อมูลไปใช้ในการวิเคราะห์ข้อมูลนั้นจะมีวิธีที่แตกต่างกันออกไปขึ้นอยู่กับวัตถุประสงค์ของการนำไปใช้งาน



รูปที่ 1 กระบวนการทำเหมืองข้อมูล

ขั้นตอนการทำเหมืองข้อมูลดังรูปที่ 1 ซึ่งสามารถอธิบายได้ดังนี้

1. กระบวนการดึงข้อมูลจากฐานข้อมูลหลายๆแหล่งด้วยคำสั่ง SQL จากนั้นจัดทำข้อมูลให้อยู่ในรูปแบบไฟล์ .CSV เพื่อนำไปใช้ในการพยากรณ์ด้วยแบบจำลองการพยากรณ์ต่างๆ

2. กระบวนการคัดเลือกข้อมูลที่จะนำมาใช้ในการพยากรณ์ โดยคัดเลือกเฉพาะข้อมูลที่เกี่ยวข้องกับการพยากรณ์ เช่น เกรดเฉลี่ย รายได้ต่อปีของบิดาและมารดา เป็นต้น ส่วนข้อมูลที่ไม่เกี่ยวข้องจะไม่นำมาใช้ในการพยากรณ์ โดยจะใช้ Operators Select Attributes ใน RapidMiner Studio ในการคัดเลือกข้อมูลที่เกี่ยวข้องกับการพยากรณ์

3. Cleansing Data คือกระบวนการเตรียมข้อมูลสำหรับการทำเหมืองข้อมูล ในกระบวนการนี้จะมี การปรับคุณภาพของข้อมูลให้เหมาะสมกับนำไปทำเหมืองข้อมูล โดยจะตรวจสอบข้อมูลก่อนที่จะนำไปทำการ พยากรณ์หากพบข้อมูลเป็นค่า NULL จะทำการตัดข้อมูลนั้นออกหรือหากสามารถเติมข้อมูลที่ไม่มีสมบูรณ์ ลงไปได้ให้เติมข้อมูลลงไปให้ครบถ้วนและหากพบข้อมูลที่มีรูปแบบต่างต่างจากรูปแบบที่กำหนดจะทำการ แทนที่ข้อมูลนั้นให้อยู่ในรูปแบบที่กำหนด เช่น กำหนดให้เพศชายเท่ากับ 1 แต่มีข้อมูลบางระเบียนที่เพศชาย เท่ากับ M ก็จะทำให้แทนค่า 1 ในข้อมูลเท่ากับ M โดยจะใช้ Operators Replace Missing Value และ Replace ใน RapidMiner Studio ในการทำ Cleansing Data

4. กระบวนการเลือกแบบจำลองการพยากรณ์ โดยจะทำการเลือกแบบจำลองการพยากรณ์ ที่เหมาะสมกับการพยากรณ์ในการวิจัยครั้งนี้โดยเลือกแบบจำลองการพยากรณ์ 3 แบบ คือ ต้นไม้ การตัดสินใจ(Decision Tree) ในอ็ฟเบย์ (Naïve Bayes) และการวิเคราะห์ความถดถอยทางโลจิสติก (Logistic Regression) โดยแบบจำลองการพยากรณ์ทั้ง 3 แบบนี้เป็นแบบจำลองการพยากรณ์แบบ การจำแนกข้อมูล (Classification) ซึ่งเหมาะสมกับการพยากรณ์การเลือกกลุ่มคณะวิชาการศึกษาต่อ ในมหาวิทยาลัยรังสิต

5. กระบวนการเปรียบเทียบประสิทธิภาพของแบบจำลองการพยากรณ์ทั้ง 3 แบบ ได้แก่ คือ ต้นไม้ การตัดสินใจ(Decision Tree) ในอ็ฟเบย์ (Naïve Bayes) และการวิเคราะห์ความถดถอยทางโลจิสติก (Logistic Regression) ว่าแบบจำลองใดมีค่าความถูกต้อง (Accuracy) ที่มากที่สุด

6. กระบวนการเลือกแบบจำลองการพยากรณ์ที่เหมาะสมกับการพยากรณ์การเลือกกลุ่มคณะวิชาใน การศึกษาในมหาวิทยาลัยรังสิตโดยสถิติ ANOVA ในการเลือกแบบจำลองการพยากรณ์

7. กระบวนการพยากรณ์การเลือกกลุ่มคณะวิชาในการศึกษาต่อมหาวิทยาลัยรังสิต คือหลังจากที่ใช้ สถิติ ANOVA ในการเลือกแบบจำลองแล้วก็จำแนกแบบจำลองที่เลือกมาพยากรณ์การเลือกกลุ่มคณะวิชาใน การศึกษาต่อมหาวิทยาลัยรังสิต

3. **ต้นไม้การตัดสินใจ(Decision Tree)** กำธร ศรีอุดม และ วรรัตน์ รุ่งวรวุฒิ (2556) กล่าวว่า ต้นไม้การตัดสินใจ (Decision Tree) เป็นอีกหนึ่งอัลกอริทึมที่ได้รับความนิยมกันอย่างแพร่หลายเนื่องจากสามารถเข้าใจได้ง่ายและเป็นเครื่องมือที่ช่วยให้วิเคราะห์เหตุการณ์ หรือสถานการณ์เพื่อการตัดสินใจได้อย่างเป็นระบบและรวดเร็ว จากการศึกษาวิจัยยังพบว่าผู้วิจัยนิยมใช้ต้นไม้การตัดสินใจในการทำวิจัย เช่น การวิจัย เรื่องระบบสนับสนุนการตัดสินใจสำหรับผู้ให้บริการเว็บไซต์การท่องเที่ยว โดยใช้ต้นไม้การตัดสินใจ ผลงานวิจัยของ กำธร ศรีอุดม และ วรรัตน์ รุ่งวรวุฒิ เป็นต้น โดยต้นไม้การตัดสินใจจะมีการแยกข้อมูลออกเป็นกลุ่มโดยจะทำการเลือกแอททริบิวต์ที่มีค่าความสัมพันธ์กับคลาสให้อยู่บนสุด (Root Node) และถัดลงมาจะเป็นโหนดราก (child node) สามารถคำนวณได้ดังรูปสมการ (เอกสิทธิ์ พัทธวงศ์ศักดิ์, 2557)

$IG(\text{parent}, \text{child}) = \text{entropy}(\text{parent}) - [p(c_1) \times \text{entropy}(c_1) + p(c_2) \times \text{entropy}(c_2) + \dots]$
โดยที่ $\text{entropy}(c_1) = -p(c_1) \log p(c_1)$ และ $p(c_1)$ คือ ค่าความน่าจะเป็นของ c_1

Entropy ค่าคาดคะเนของข้อมูลเป็นค่าที่แยกสามารถคำนวณได้ดังรูปสมการ

$$\text{Entropy}(t_i) = 1 - \sum_{t=1}^N [p(t_i)] \log_2 p(t_i)$$

4. **ในอีฟเบย์(Naïve Bayes)** วิรัช วิรัชกุล (2557) กล่าวว่า ในอีฟเบย์ (Naïve Bayes) เป็นอัลกอริทึมที่เข้าใจง่ายและสามารถจำแนกข้อมูลได้อย่างมีประสิทธิภาพเหมาะสมต่อการพยากรณ์ คลาสของชุดข้อมูล นอกจากนี้อัลกอริทึมในอีฟเบย์ยังสามารถทำงานได้รวดเร็วหากมีการจัดกลุ่มของตัวแปร อินพุต โดยการวิเคราะห์หาความสัมพันธ์ระหว่างตัวแปรเพื่อนำไปใช้สร้างเงื่อนไขความน่าจะเป็นสำหรับแต่ละความสัมพันธ์ กำหนดให้ความน่าจะเป็นของข้อมูลที่จะเป็นกลุ่ม V_i สำหรับข้อมูลที่มีแอททริบิวต์ทั้งหมด n ตัว มีรูปแบบสมการดังนี้

$$P = \{a_1, a_2, \dots, a_n\} \text{ หรือใช้สัญลักษณ์ว่า } P(a_1, a_2, \dots, a_n | V_j) \quad (1)$$

$$\text{คือ } P(a_1, a_2, \dots, a_n | V_j) = \prod_{i=1}^n P(a_i | V_j)$$

โดยที่ \prod หมายถึง ผลคูณของค่า $P(a_i | V_j)$ ทั้งหมด

i คือ $1, 2, 3, \dots, n$

j คือ $1, 2, 3, \dots, n$

การนำวิธีการเรียนรู้แบบอย่างง่ายไปใช้ มีวิธีการดังต่อไปนี้ คือ

1. หาค่าความน่าจะเป็นของค่าที่พบในแต่ละกลุ่มโดยนำค่า $P(a_1, a_2, \dots, a_n | V_j)$ จากสมการที่ 1 มาคูณกับค่าความน่าจะเป็นของกลุ่มนั้นๆ คือ $P(V_j)$ ได้เท่ากับ V_{NB}
2. นำค่าที่ได้ มาเปรียบเทียบกัน กลุ่มที่มีค่าความน่าจะเป็นสูงสุด คือ คำตอบ ดังนั้นเราจะได้ว่าวิธีการจำแนกประเภทแบบเบย์อย่างง่าย ดังสมการที่ 2

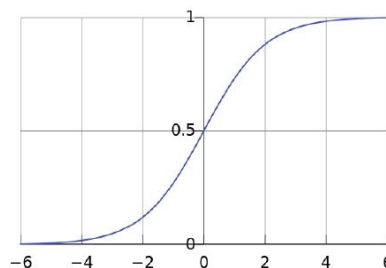
นำค่าที่ได้มาเปรียบเทียบกับกลุ่มที่มีค่าความน่าจะเป็นสูงสุด รูปแบบสมการ

$$V_{NB} = \arg_{v_j} \max P(V_j) \times \prod_{i=1}^n P(a_i | V_j) \quad (2)$$

5. การวิเคราะห์ความถดถอยทางโลจิสติก (Logistic Regression) Zhua, Uwaldemudiaa and Fengb (2019) การวิเคราะห์ความถดถอยทางโลจิสติก (Logistic Regression) เป็นการวิเคราะห์สถิติเชิงคุณภาพเพื่อหาความสัมพันธ์ระหว่างตัวแปรอิสระกับตัวแปรตามและพยากรณ์โอกาสที่จะเกิดเหตุการณ์ต่างๆ เหมาะสมกับการพยากรณ์ชุดข้อมูลที่มีตัวแปรตามเป็น binary เช่น ป่วย/ไม่ป่วย หรือ ใช่/ไม่ใช่ เป็นต้น ซึ่งผลลัพธ์ที่เกิดขึ้นเป็นได้เพียง 2 ค่า คือ 0 และ 1 หากตัวแปรต้นมีค่าน้อย ค่าของตัวแปรตามจะมีค่าเท่ากับ 0 และหากตัวแปรต้นมีค่ามาก ค่าของตัวแปรตามจะมีค่าเท่ากับ 1 เส้นโค้งที่แทนความสัมพันธ์นี้ได้จากฟังก์ชัน (ชณัฐดาภรณ์ เย็นประเสริฐ, 2557)

$$F(X) = \frac{1}{1 + \exp(-x)}$$

ฟังก์ชันโลจิสติก (Logistic Function) ตัวแปรตามที่ได้จากฟังก์ชันนี้จะมีค่าอยู่ระหว่าง 0 และ 1



รูปที่ 2 ฟังก์ชันโลจิสติก (ที่มา: ชณัฐดาภรณ์ เย็นประเสริฐ, 2557)

ในการวิเคราะห์การถดถอยทางโลจิสติกกรณีที่มีตัวแปรอิสระเพียงตัวเดียว สามารถเขียนสมการได้ดังนี้

$$F(X) = \frac{1}{1 + \exp(-x)}$$

$$\text{or Prob(event)} = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X)}}$$

เมื่อ β_0 และ β_1 เป็นค่าสัมประสิทธิ์ที่ประมาณได้จากข้อมูล

X เป็นตัวแปรอิสระ

e มีค่าประมาณ 2.718

สมการข้างต้นสามารถเขียนใหม่ได้เป็น

$$\text{Prob(event)} = \frac{e^Z}{1 + e^Z}$$

$$\text{or Prob(event)} = \frac{1^{e^Z}}{1 + e^{-Z}}$$

$$\text{โดย } Z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots \beta_p X_p$$

และโอกาสที่จะไม่เกิดเหตุการณ์สามารถประมาณได้จากสมการ

$$\text{Prob(noevent)} = 1 - \text{Prob(event)}$$

6. การวิเคราะห์ความแปรปรวน (Analysis of Variance: ANOVA) เป็นการวิเคราะห์ข้อมูลที่ใช้ในการทดสอบสมมติฐานของความแตกต่างค่าเฉลี่ยของประชากรมากกว่า 2 กลุ่ม โดยทำการทดสอบสมมติฐานเพียงครั้งเดียว หลักการของการวิเคราะห์ความแปรปรวนจะแบ่งความแปรปรวนของข้อมูลออกไปตามสาเหตุ ที่ทำให้ข้อมูลแตกต่างกัน คือความแปรปรวนภายในกลุ่มและความแปรปรวนระหว่างกลุ่ม โดยที่ (วิเคราะห์ความแปรปรวน, 2562)

$$\text{ความแปรปรวนทั้งหมด} = \text{ความแปรปรวนภายในกลุ่ม} + \text{ความแปรปรวนระหว่างกลุ่ม}$$

ลักษณะของข้อมูลที่ใช้ในการวิเคราะห์ความแปรปรวนทางเดียว ดังตารางที่ 1

ตารางที่ 1 ลักษณะของข้อมูลที่ใช้ในการวิเคราะห์ความแปรปรวนทางเดียว

	ทรีทเมนต์(treatment)					
	1	2	3	...	k	
	x_{11}	x_{21}	x_{31}	...	x_{k1}	
	x_{12}	x_{22}	x_{32}	...	x_{k2}	
	x_{13}	x_{23}	x_{33}	...	x_{k3}	
	:	:	:	...	:	
	x_{1n1}	x_{2n2}	x_{3n3}	...	x_{knk}	
รวม	T_1	T_2	T_3	...	T_k	T
ค่าเฉลี่ย	\bar{x}_1	\bar{x}_2	\bar{x}_3	...	\bar{x}_k	\bar{x}

- เมื่อ x_{ij} แทนข้อมูลของทรีทเมนต์ที่ i หน่วยทดลองที่ j
 $i = 1, 2, 3, \dots, k$ และ $j = 1, 2, 3, \dots, n_i$
 T_i แทนผลรวมของข้อมูลทรีทเมนต์ที่ i
 T แทนผลรวมข้อมูลทั้งหมด
 \bar{x}_i แทนค่าเฉลี่ยของข้อมูลทรีทเมนต์ที่ i
 \bar{x} แทนค่าเฉลี่ยของข้อมูลทั้งหมด
 k แทนจำนวนทรีทเมนต์
 n แทนจำนวนข้อมูลทั้งหมด เท่ากับ $n_1 + n_2 + n_3 + \dots + n_k$

6.1 ความแปรปรวนระหว่างกลุ่ม (Between Groups Sum of Square) เป็นการพิจารณาความแปรปรวนที่เกิดจากค่าเฉลี่ยของตัวอย่างในแต่ละกลุ่มแตกต่างจากค่าเฉลี่ยรวม โดยที่

$$SSB = \sum_{i=1}^k n_i (x_i - \bar{x})^2$$

6.2 ความแปรปรวนภายในกลุ่ม (Within Group Sum of Square) เป็นการพิจารณาความแปรปรวนที่เกิดภายในกลุ่มแต่ละกลุ่มซึ่งไม่ทราบสาเหตุความแปรปรวนที่เกิดขึ้น โดยที่

$$SSE = \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2$$

6.3 ความแปรปรวนรวม Total Sum of Square) เป็นการพิจารณาความแปรปรวนที่เกิดจากค่าสังเกตแต่ละค่าแตกต่างจากค่าเฉลี่ยรวม โดยที่

$$SST = \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2 \text{ และ } SST = SSB + SSE$$

การวิจัยนี้ ผู้วิจัยใช้ ANOVA เปรียบเทียบประสิทธิภาพของแบบจำลองการพยากรณ์ทั้ง 3 แบบ เนื่องจาก ANOVA สามารถเปรียบเทียบแบบจำลองการพยากรณ์ได้ตั้งแต่ 2 แบบจำลองขึ้นไป จึงมีความเหมาะสมที่จะนำมาใช้ในการเปรียบเทียบประสิทธิภาพของแบบจำลองการพยากรณ์ในการวิจัยครั้งนี้ โดยมีเงื่อนไขของการวิเคราะห์ความแปรปรวนในการทดสอบสมมติฐานเกี่ยวกับค่าเฉลี่ยของประชากร k กลุ่ม ด้วยเทคนิคการวิเคราะห์ความแปรปรวน คือ

1. ประชากร k กลุ่มมีการแจกแจงแบบปกติ
2. ความแปรปรวนของแต่ละประชากรเท่ากัน
3. ตัวอย่างสุ่มจากแต่ละประชากรเป็นอิสระต่อกัน

7. ค่าความแม่นยำ (Accuracy)

Subasi, Ahmed and Alickovic (2018) การวัดค่าความแม่นยำ (Accuracy) เพื่อเปรียบเทียบแบบจำลองการพยากรณ์เพื่อหาแบบจำลองที่เหมาะสม การหาค่าประสิทธิภาพในการทำงานได้ดังรูปแบบสมการ

$$Accuracy = \frac{TP + TN}{TN + TP + FN + FP} \times 100 \%$$

TP คือ ค่าที่ทำนายถูกต้องเชิงบวก

TN คือ ค่าที่ทำนายถูกต้องเชิงลบ

FN คือ ค่าที่ทำนายผิดพลาดเชิงบวก

FP คือ ค่าที่ทำนายผิดพลาดเชิงลบ

8. ค่าความเที่ยงตรง (Precision)

การวัดความเที่ยงตรง (Precision) คือการวัดค่าการพยากรณ์ว่าจริงได้ถูกต้องของแบบจำลองการพยากรณ์สามารถหาค่าความเที่ยงตรงได้ดังรูปแบบสมการ (Jaafaria, Zennerb and Thai Phamc, 2018)

$$Precision = \frac{TP}{TP + FP}$$

TP คือ ค่าที่ทำนายถูกต้องเชิงบวก

FP คือ ค่าที่ทำนายผิดพลาดเชิงลบ

9. ค่าความระลึก (Recall)

การวัดค่าความระลึกได้ (Recall) คือการวัดค่าการพยากรณ์ได้ว่าจริงของแบบจำลองการพยากรณ์ เป็นอัตราส่วนเท่าไรของจริงทั้งหมดสามารถหาค่าความระลึกได้ดังรูปแบบสมการ (Jaafaria, Zennerb and Thai Phamc, 2018)

$$Recall = \frac{TP}{TP + FN}$$

TP คือ ค่าที่ทำนายถูกต้องเชิงบวก

FN คือ ค่าที่ทำนายผิดพลาดเชิงบวก

10. เครื่องมือ (Tools)

ในการวิจัยครั้งนี้ผู้วิจัยใช้ Microsoft SQL Server Management Studio 2014 ในการดึงข้อมูลจากฐานข้อมูล และแปลงข้อมูลให้อยู่ในรูปแบบไฟล์ .CSV ในการพยากรณ์การเลือกกลุ่มคณะวิชาในการศึกษาต่อมหาวิทยาลัยวังสิตและเปรียบเทียบแบบจำลองการพยากรณ์โดยใช้โปรแกรม RapidMiner Studio เป็นเครื่องมือในการทำในการทดลองแบบจำลองการพยากรณ์และเปรียบเทียบแต่ละแบบจำลอง

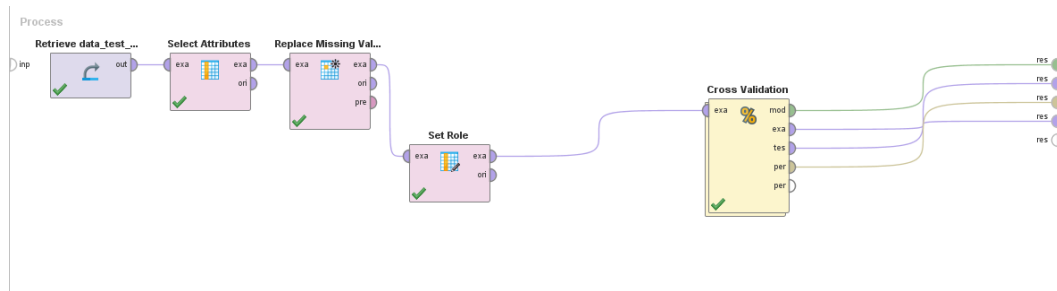
ผลการทดลอง

จากผลการวิจัยสามารถสรุปผลการวิจัยได้เป็น 3 ตอน ดังนี้

ตอนที่ 1 การเปรียบเทียบแบบจำลองการพยากรณ์ต้นไม้การตัดสินใจ (Decision Tree) ในอ็ฟเบย์ (Naïve Bayes) และการวิเคราะห์ความถดถอยทางโลจิสติก (Logistic Regression)

จากการเลือกแบบจำลองการพยากรณ์ 3 แบบ เพื่อหาค่า Accuracy Precision และ Recall ในแต่ละแบบจำลองการพยากรณ์ซึ่งจะประกอบด้วย Operators ดังต่อไปนี้ Retrieve Data คือการดึงข้อมูลที่อยู่ในรูปแบบไฟล์ .CSV ที่เตรียมไว้เพื่อใช้ในการพยากรณ์ด้วยแบบจำลองการพยากรณ์ต่างๆ Select Attributes คือการเลือกข้อมูลที่เกี่ยวข้องกับการพยากรณ์ Replace Missing Values คือการแทนที่ข้อมูลที่ไม่สมบูรณ์ Set Role คือการเลือก Attributes ที่ต้องการใช้เป็นคลาสในการพยากรณ์ และ Cross Validation คือการทดสอบประสิทธิภาพของแต่ละแบบจำลองการพยากรณ์โดย Operators Cross Validation นี้จะมี Sub Process ของแต่ละแบบจำลองการพยากรณ์ โดยใช้ RapidMiner Studio เป็นเครื่องช่วยในการพยากรณ์

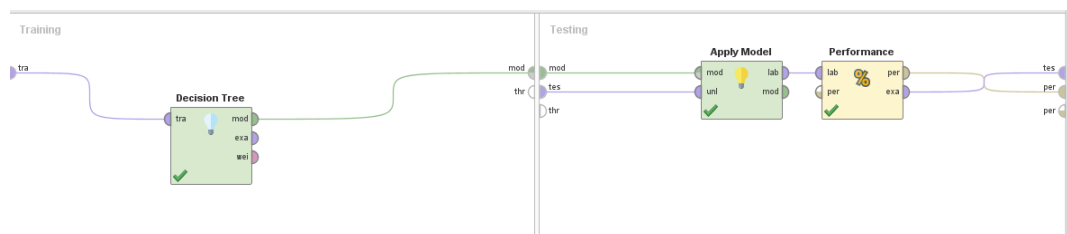
กระบวนการหลักของการพยากรณ์ด้วยแบบจำลองการพยากรณ์ ดังรูปที่ 3



รูปที่ 3 กระบวนการหลักของการพยากรณ์ด้วยแบบจำลองการพยากรณ์ต่างๆ

1.1 กระบวนการสร้างแบบจำลองการพยากรณ์โดยใช้ต้นไม้การตัดสินใจ (Decision Tree (J48)) ซึ่งมีจะประกอบด้วย Operators ดังต่อไปนี้ Decision Tree การพยากรณ์ด้วยแบบจำลองการพยากรณ์ต้นไม้การตัดสินใจ Apply Model คือการนำแบบจำลองการพยากรณ์ที่สร้างได้ไปใช้ในการพยากรณ์เพื่อหาคำตอบให้ข้อมูลที่ยังไม่ทราบคลาส Performance คือการวัดประสิทธิภาพของแบบจำลองการพยากรณ์ เช่น ค่า Accuracy Class Recall Precision Recall เป็นต้น

กระบวนการพยากรณ์ด้วยแบบจำลองการพยากรณ์ต้นไม้การตัดสินใจ(Decision Tree) ดังรูปที่ 4



รูปที่ 4 ขั้นตอนการพยากรณ์ด้วยแบบจำลองการพยากรณ์ด้วยต้นไม้การตัดสินใจ(Decision Tree)

หลังจากการพยากรณ์ด้วยแบบจำลองการพยากรณ์ต้นไม้การตัดสินใจ(Decision Tree) ได้ผลดังตารางที่ 2

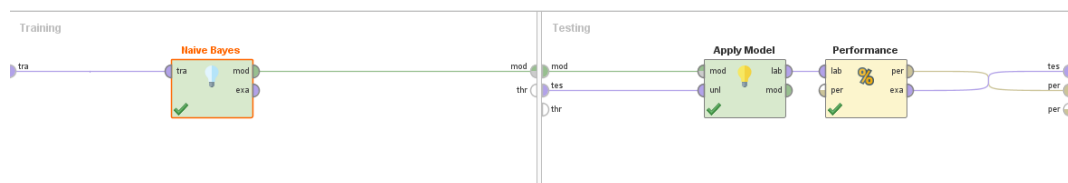
ตารางที่ 2 ผลการพยากรณ์ด้วยแบบจำลองการพยากรณ์ต้นไม้การตัดสินใจ(Decision Tree)

Model	Accuracy	Class Recall (Art)	Class Recall (Science)	Class Precision (Art)	Class Precision (Science)
Decision Tree	75.79%	62.54%	86.99%	80.26%	73.30%

แบบจำลองการพยากรณ์ต้นไม้การตัดสินใจ (Decision tree) มีค่าความแม่นยำ (Accuracy) 75.79 % ความถูกต้องของการพยากรณ์การเลือกเรียนกลุ่มคณะวิทยาศาสตร์ เท่ากับ 62.54 % ความถูกต้องของการพยากรณ์การเลือกเรียนกลุ่มคณะวิทยาศาสตร์ เท่ากับ 86.99 %

1.2 กระบวนการสร้างแบบจำลองการพยากรณ์โดยใช้ในอืฟเบย์ (Naïve Bayes) ซึ่งมีจะประกอบด้วย Operators ดังต่อไปนี้ Naïve Bayes การพยากรณ์ด้วยแบบจำลองการพยากรณ์ในอืฟเบย์ Apply Model คือการนำแบบจำลองการพยากรณ์ที่สร้างได้ไปใช้ในการพยากรณ์เพื่อหาคำตอบให้ข้อมูลที่ยังไม่ทราบคลาส Performance คือการวัดประสิทธิภาพของแบบจำลองการพยากรณ์ เช่น ค่า Accuracy Class Recall Precision Recall เป็นต้น

กระบวนการพยากรณ์ด้วยแบบจำลองการพยากรณ์ในอืฟเบย์ (Naïve Bayes) ดังรูปที่ 5



รูปที่ 5 ขั้นตอนการพยากรณ์ด้วยแบบจำลองการพยากรณ์ด้วยในอืฟเบย์ (Naïve Bayes)

หลังจากการพยากรณ์ด้วยแบบจำลองการพยากรณ์ในอีฟเบย์ (Naïve Bayes) ได้ผลดังตารางที่ 3

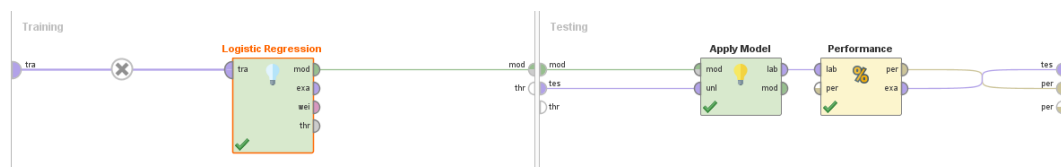
ตารางที่ 3 ผลการพยากรณ์ด้วยแบบจำลองการพยากรณ์ในอีฟเบย์ (Naïve Bayes)

Model	Accuracy	Class Recall (Art)	Class Recall (Science)	Class Precision (Art)	Class Precision (Science)
Naïve Bayes	75.75%	65.29%	84.59%	78.19%	74.23%

แบบจำลองการพยากรณ์ในอีฟเบย์ (Naïve Bayes) มีค่าความแม่นยำ (Accuracy) 75.75 % ความถูกต้องของการพยากรณ์การเลือกเรียนกลุ่มคณะวิชาศิลปะ เท่ากับ 65.29 % ความถูกต้องของการพยากรณ์การเลือกเรียนกลุ่มคณะวิทยาศาสตร์ เท่ากับ 84.59 %

1.3 กระบวนการสร้างแบบจำลองการพยากรณ์โดยใช้การวิเคราะห์ความถดถอยทางโลจิสติก (Logistic Regression) ซึ่งจะประกอบด้วย Operators ดังต่อไปนี้ Logistic Regression การพยากรณ์ด้วยแบบจำลองการพยากรณ์การวิเคราะห์ความถดถอยทางโลจิสติก Apply Model คือการนำแบบจำลองการพยากรณ์ที่สร้างได้ไปใช้ในการพยากรณ์เพื่อหาคำตอบให้ข้อมูลที่ยังไม่ทราบคลาส Performance คือการวัดประสิทธิภาพของแบบจำลองการพยากรณ์ เช่น ค่า Accuracy Class Recall Precision Recall เป็นต้น

กระบวนการพยากรณ์ด้วยแบบจำลองการพยากรณ์การวิเคราะห์ความถดถอยทางโลจิสติก (Logistic Regression) ดังรูปที่ 6



รูปที่ 6 ขั้นตอนการพยากรณ์ด้วยแบบจำลองการพยากรณ์ด้วยการวิเคราะห์ความถดถอยทางโลจิสติก (Logistic Regression)

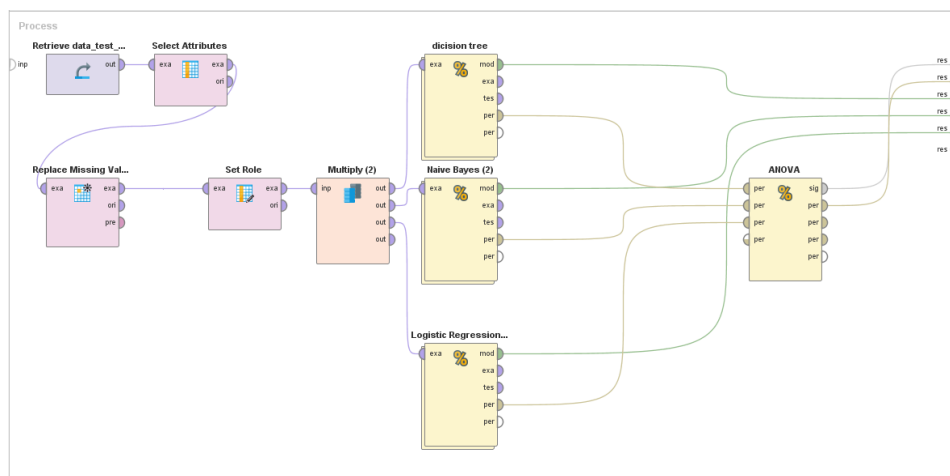
หลังจากการพยากรณ์ด้วยแบบจำลองการพยากรณ์การวิเคราะห์ความถดถอยทางโลจิสติก (Logistic Regression) ได้ดังตารางที่ 4

ตารางที่ 4 ผลการพยากรณ์ด้วยแบบจำลองการพยากรณ์ต้นไม้การตัดสินใจ

Model	Accuracy	Class Recall (Art)	Class Recall (Science)	Class Precision (Art)	Class Precision (Science)
Logistic Regression	76.26%	62.89%	87.57%	81.06%	73.61%

แบบจำลองการพยากรณ์การวิเคราะห์ความถดถอยทางโลจิสติก (Logistic Regression) มีความแม่นยำ (Accuracy) 76.26 % ความถูกต้องของการพยากรณ์การเลือกเรียนกลุ่มคณะวิชาศิลปะ เท่ากับ 62.89 % ความถูกต้องของการพยากรณ์การเลือกเรียนกลุ่มคณะวิทยาศาสตร์ เท่ากับ 87.57 %

ตอนที่ 2 เป็นการเปรียบเทียบประสิทธิภาพของแบบจำลองการพยากรณ์ต้นไม้การตัดสินใจ (Decision Tree) ในอีฟเบย์ (Naïve Bayes) และการวิเคราะห์ความถดถอยทางโลจิสติก (Logistic Regression) โดยใช้สถิติ ANOVA ในการเปรียบเทียบดังรูปที่ 7



รูปที่ 7 การเปรียบเทียบแบบจำลองการพยากรณ์ด้วย RapidMiner

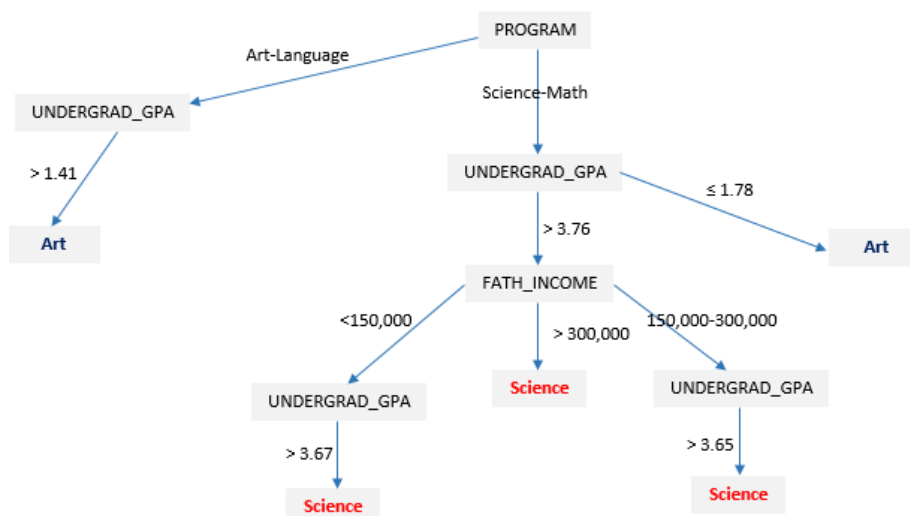
การเปรียบเทียบแบบจำลองการพยากรณ์ทั้ง 3 แบบ โดยใช้สถิติ ANOVA ทำการเปรียบเทียบแบบจำลองการพยากรณ์ได้ผลดังตารางที่ 5

ตารางที่ 5 ผลการเปรียบเทียบแบบจำลองการพยากรณ์ด้วย ANOVA

Source	Square Sums	DF	Mean Squares	F	Prob
Between	0.012	2	0.006	0.263	0.771
Residuals	0.013	27	0.072		
Total	0.013	29			

จากตารางที่ 5 ผลการเปรียบเทียบแบบจำลองการพยากรณ์ด้วยสถิติ ANOVA พบว่าแบบจำลองการพยากรณ์ต้นไม้การตัดสินใจ(Decision Tree) ในอีมพีเบย์(Naïve Bayes) และการวิเคราะห์ความถดถอยทางโลจิสติก (Logistic Regression) มีค่า F เท่ากับ 0.263 และค่า Prob เท่ากับ 0.771 ไม่แตกต่างกันอย่างมีนัยสำคัญที่ระดับ 0.05 แสดงว่าแบบจำลองการพยากรณ์ทั้ง 3 เทคนิค ไม่แตกต่างกันอย่างมีนัยสำคัญ ดังนั้นจึงสามารถเลือกใช้แบบจำลองการพยากรณ์ใดก็ได้ในการทำวิจัย

ตอนที่ 3 การพยากรณ์การเลือกกลุ่มคณะวิชาในการศึกษาต่อระดับปริญญาตรี โดยใช้ต้นไม้การตัดสินใจ (Decision Tree)



รูปที่ 8 ผลการพยากรณ์การเลือกคณะด้วยต้นไม้การตัดสินใจ(Decision Tree)

จากการพยากรณ์การเลือกกลุ่มคณะวิชาในการศึกษาต่อระดับอุดมศึกษาด้วยแบบจำลองการพยากรณ์ต้นไม้การตัดสินใจ(Decision Tree) ผลพบว่า

1. IF PROMGRAM = "Science-Math" AND UDERGRADE_GPA = "3.76" AND FATH_INCOM = ">300,000" THEN GROUP_TYPE = "Science"
2. IF PROMGRAM = "Science-Math" AND UDERGRADE_GPA = "3.67" AND FATH_INCOME = "<150,000" THEN GROUP_TYPE = "Science"
3. IF PROMGRAM = "Science-Math" AND UDERGRADE_GPA = "3.65" AND FATH_INCOME = "150,000-300,000" THEN GROUP_TYPE = "Science"
4. IF PROMGRAM = "Science-Math" AND UDERGRADE_GPA = "1.78" THEN GROUP_TYPE = "Art"
5. IF PROMGRAM = "Art-Language" AND UDERGRADE_GPA = "1.41" THEN GROUP_TYPE = "Art"

สรุปการวิเคราะห์ผลการวิจัย

จากการเปรียบเทียบแบบจำลองการพยากรณ์ต้นไม้การตัดสินใจ (Decision Tree) ในอืฟเบย์ (Naïve Bayes) และการวิเคราะห์ความถดถอยทางโลจิสติก (Logistic Regression) แบบพบว่ามีค่า F เท่ากับ 0.263 และค่า Prob เท่ากับ 0.771 ซึ่งไม่แตกต่างกันอย่างมีนัยสำคัญที่ระดับ 0.05 ในการวิจัยครั้งนี้ ผู้วิจัยเลือกแบบจำลองการพยากรณ์ต้นไม้การตัดสินใจ (Decision Tree) เพื่อใช้ในการพยากรณ์การเลือกกลุ่มคณะวิชาในการศึกษาต่อมหาวิทยาลัยรังสิตและจากการพยากรณ์การเลือกคณะในการศึกษาต่อมหาวิทยาลัยรังสิตพบว่ากลุ่มตัวอย่างที่เลือกเรียนกลุ่มคณะวิชาศิลปะ คือกลุ่มตัวอย่างที่เรียนในแผนการเรียนวิทยาศาสตร์-คณิตศาสตร์ หรือแผนการเรียนศิลป์-ภาษา และมีเกรดเฉลี่ยในระดับมัธยมศึกษาตอนปลายอยู่ที่ระหว่าง 1.14-1.78 ส่วนกลุ่มตัวอย่างที่เลือกเรียนกลุ่มคณะวิชาวิทยาศาสตร์ คือกลุ่มตัวอย่างที่เรียนในแผนการเรียนวิทยาศาสตร์-คณิตศาสตร์ และมีเกรดเฉลี่ยในระดับมัธยมศึกษาตอนปลายอยู่ที่ระหว่าง 3.65 – 3.76

อภิปรายผลการวิจัย

จากผลการเลือกกลุ่มคณะวิชาในการศึกษาต่อมหาวิทยาลัยรังสิตโดยใช้กลุ่มตัวอย่างเป็นนักศึกษาที่เข้าศึกษาในปีการศึกษา 2559-2562 จำนวน 2,540 คน และเปรียบเทียบแบบจำลองการพยากรณ์ การเลือกกลุ่มคณะวิชาในการศึกษาต่อมหาวิทยาลัยรังสิตมีประเด็นในการอภิปรายได้ดังนี้

1. จากการพยากรณ์การเลือกกลุ่มคณะวิชาในการศึกษาต่อมหาวิทยาลัยรังสิต พบว่า

1.1 การวิเคราะห์ด้วยต้นไม้การตัดสินใจ (Decision Tree) พบว่า ถ้ากลุ่มตัวอย่างที่เลือกเรียนกลุ่มคณะวิชาวิทยาศาสตร์ จะเป็นกลุ่มที่เรียนมัธยมศึกษาตอนปลายในแผนการเรียน วิทยาศาสตร์-คณิตศาสตร์ มีเกรดเฉลี่ยมากกว่า 3.76 และบิดามีรายได้เฉลี่ยต่อปีมากกว่า 300,000 บาทต่อปี

1.2 ถ้ากลุ่มตัวอย่างที่เลือกเรียนกลุ่มคณะวิชาวิทยาศาสตร์ จะเป็นกลุ่มที่เรียนมัธยมศึกษาตอนปลายในแผนการเรียน วิทยาศาสตร์-คณิตศาสตร์ มีเกรดเฉลี่ยมากกว่า 3.67 และบิดามีรายได้เฉลี่ยต่อปีน้อยกว่า 150,000 บาทต่อปี

1.3 ถ้ากลุ่มตัวอย่างที่เลือกเรียนกลุ่มคณะวิชาวิทยาศาสตร์ จะเป็นกลุ่มที่เรียนชั้นมัธยมศึกษาตอนปลายในแผนการเรียน วิทยาศาสตร์-คณิตศาสตร์ มีเกรดเฉลี่ยมากกว่า 3.65 และบิดามีรายได้เฉลี่ย 150,000-300,000 บาทต่อปี

1.4 ถ้ากลุ่มตัวอย่างที่เลือกเรียนกลุ่มคณะวิชาศิลปะ จะเป็นกลุ่มที่เรียนมัธยมศึกษาตอนปลายในแผนการเรียน วิทยาศาสตร์-คณิตศาสตร์ และมีเกรดเฉลี่ยน้อยกว่าหรือเท่ากับ 1-78

1.5 ถ้ากลุ่มตัวอย่างที่เลือกเรียนกลุ่มคณะวิชาศิลปะ จะเป็นกลุ่มที่เรียนมัธยมศึกษา ตอนปลายในแผนการเรียน ศิลป์-ภาษา และมีเกรดเฉลี่ยมากกว่า 1.41

จากผลการวิเคราะห์ พบว่า ปัจจัยที่ส่งผลต่อการเลือกกลุ่มคณะวิชาเพื่อศึกษาต่อระดับปริญญาตรี ในมหาวิทยาลัยรังสิต นอกจากแผนการเรียนในมัธยมศึกษาตอนปลายแล้ว เกรดเฉลี่ย และรายได้ของผู้ปกครองยังส่งผลต่อการเลือกกลุ่มคณะวิชาเพื่อศึกษาต่อในระดับปริญญาตรีอีกด้วย

2. จากการวิจัยเปรียบเทียบประสิทธิภาพและความถูกต้องของแบบจำลองการพยากรณ์ทั้ง 3 แบบ ซึ่งแบบจำลอง การวิเคราะห์ความถดถอยทางโลจิสติก (Logistic Regression) มีค่าความแม่นยำอยู่ที่ร้อยละ 76.26% แบบจำลองต้นไม้การตัดสินใจ (Decision Tree) อยู่ที่ร้อยละ 75.79% และแบบจำลองในอีฟเบย์ (Naive Bayes) มีความแม่นยำอยู่ที่ร้อยละ 75.75% พบว่า ไม่แตกต่างกันอย่างมีนัยสำคัญที่ระดับ

0.05 ดังนั้น ในการพยากรณ์จะเลือกใช้แบบจำลองการพยากรณ์ใดก็ได้ ซึ่งการวิจัยครั้งนี้ผู้วิจัยเลือกแบบจำลองการพยากรณ์ต้นไม้การตัดสินใจ (Decision Tree) ในการพยากรณ์เพื่อหาปัจจัยที่ส่งต่อการเลือกกลุ่มคณะวิชาการศึกษาต่อในระดับปริญญาตรีในมหาวิทยาลัยรังสิต เพราะแบบจำลองต้นไม้การตัดสินใจเป็นอัลกอริทึมที่ได้รับความนิยมกันอย่างแพร่หลายและสามารถเข้าใจได้ง่าย

ข้อเสนอแนะ

1. การวิจัยในอนาคต ควรมีการวิเคราะห์แยกกลุ่มคณะวิชาตามการประกันคุณภาพการศึกษาภายใน เพื่อให้ได้ข้อมูลเชิงลึกประกอบการตัดสินใจของผู้บริหารระดับมหาวิทยาลัย รวมทั้งเพื่อเป็นข้อมูลการทำการตลาดของฝ่ายที่เกี่ยวข้องกับการแนะแนว
2. เนื่องจากมหาวิทยาลัยรังสิตเปิดโอกาสให้ผู้ที่จะศึกษาต่อในระดับปริญญาตรีที่จบการศึกษามัธยมศึกษาตอนปลายในสายศิลป์ สามารถเลือกศึกษาต่อในกลุ่มคณะวิชาวิทยาศาสตร์ได้ เช่น วิทยาลัยการแพทย์แผนตะวันออก คณะทันตมาตรศาสตร์ เป็นต้น ดังนั้น ทางมหาวิทยาลัยสามารถนำผลการวิจัยไปใช้ประกอบการแนะแนวได้อีกทางเลือกหนึ่ง
3. สามารถนำผลการวิเคราะห์ใช้เป็นแนวทางในการเลือกกลุ่มคณะวิชาการศึกษาต่อระดับปริญญาตรี ในมหาวิทยาลัยรังสิตนำไปแนะนำนักเรียนที่จะศึกษาต่อในระดับปริญญาตรี เช่น นักเรียนที่จบการศึกษามัธยมศึกษาตอนปลายในแผนการเรียนวิทยาศาสตร์-คณิตศาสตร์ ที่มีเกรดเฉลี่ยต่ำควรแนะนำให้เลือกศึกษาต่อระดับปริญญาตรีในกลุ่มคณะวิชาศิลปะ จะทำให้ผู้ที่ศึกษาต่อระดับปริญญาตรีมีโอกาสที่จบการศึกษาในระดับปริญญาตรีสูงขึ้น และยังช่วยลดอัตราการตกออกของนักศึกษาที่เลือกเรียนไม่ตรงกับความถนัดของตนเองและมีความรู้พื้นฐานไม่เพียงพอ

- การวิเคราะห์ความแปรปรวน. (2562). สืบค้น 25 กันยายน 2562. จาก: <http://pws.npru.ac.th/chalida/data/files/บทที่%209การวิเคราะห์ความแปรปรวน.pdf>
- กำธร ศรีอุดม และวราวัณ รุ่งวรวิมล. (2556). ระบบสนับสนุนการตัดสินใจสำหรับผู้ให้บริการเว็บเซอร์วิสการท่องเที่ยวโดยใช้ต้นไม้การตัดสินใจ. (วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ) ขอนแก่น: มหาวิทยาลัย ขอนแก่น.
- จิราภา เลหาะวรรณ, รชต์ ลิ้มสุทธิวันภูมิ และ บัณฑิต ฐานะโสภณ. (2558). การใช้เทคนิคการทำเหมืองข้อมูลในการจำแนกและคัดเลือกแขนงวิชาสำหรับนักศึกษาคณะเทคโนโลยีสารสนเทศ. วารสารเทคโนโลยีสารสนเทศลาดกระบัง, 4(2).
- ชญ์ฐดาภรณ์ เย็นประเสริฐ. (2557). การเปรียบเทียบความแม่นยำการพยากรณ์สถานะการชำระหนี้ของลูกค้า โดยใช้เทคนิคการถดถอยโลจิสติก นาอ์ฟเบย์ และต้นไม้ตัดสินใจ. (วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ). กรุงเทพฯ: มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ.
- วิชา วิรัชกุล. (2557). การเปรียบเทียบประสิทธิภาพระหว่างเทคนิค Decision Tree C4.5 , k-NN และ Naive Bayes เพื่อใช้พยากรณ์ผลสอบมาตรฐานด้านเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏนครปฐม (วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ). กรุงเทพฯ: มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ.
- สมาคมสยามมหาวิทยาลัย (ประเทศไทย). (2562). สถาบันอุดมศึกษาในสังกัด สำนักงานคณะกรรมการการอุดมศึกษา. สืบค้น 7 ตุลาคม 2562, จาก: <https://tagbth.wordpress.com/>.
- สำนักงานทะเบียน มหาวิทยาลัยรังสิต. (2562). ข้อมูลนักศึกษามหาวิทยาลัยรังสิต ปีการศึกษา 2562. ปทุมธานี: มหาวิทยาลัยรังสิต.
- สำนักงานสถิติแห่งชาติ. (2562). จำนวนการเกิดจากการทะเบียน จำแนกตามเพศ ภาค และจังหวัด พ.ศ. 2551-2561. สืบค้น 7 ตุลาคม 2562, จาก: <http://statbbi.nso.go.th/staticreport/page/sector/th/01.aspx>.
- หนึ่งพิศ ชัยอากร. (2562). การวิเคราะห์ข้อมูลด้วยเทคนิคดาต้าไมนิง. สืบค้น 25 กันยายน 2562, จาก <http://www.erpmju.ac.th/articleDetail.aspx?qid=551>.
- เอกสิทธิ์ พัทธวงศ์ศักดิ์. (2557). ขั้นตอนการสร้างโมเดล Decision Tree สืบค้น 12 ตุลาคม 2562, จาก <http://dataminingtrend.com/2014/decision-tree-model/>.

- Jaafari, A., Zenner, E. K., & Pham, B. T. (2018). Wildfire spatial pattern analysis in the Zagros Mountains, Iran: A comparative study of decision tree based classifiers. *Ecological Informatics*, 43(2018), 200-211.
- Subasi, A., Ahmed, A., & Alickovic, E. (2018). Effect of Flash Stimulation for Migraine Detection Using DecisionTree Classifiers. *Procedia Computer Science*, 140(2018), 223–229.
- Zhu, C., Idemudia, C. U., & Feng, W. (2019). Improved logistic regression model for diabetes prediction by integrating PCA and K-means techniques. *Informatics in Medicine Unlocked*, 17 (2019). Retrieved October 7, 2019, from ScienceDirect database.