การใช้เทคนิคเหมืองข้อมูลเพื่อคัดกรองบุคคลที่มีแนวโน้มสัมฤทธิผล ในการศึกษาระดับปริญญาตรี สาขาวิชาเทคโนโลยีการศึกษา

ธนะวัฒน์ วรรณประภา

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมคอมพิวเตอร์และโทรคมนาคม
วิทยาลัยนวัตกรรมด้านเทคโนโลยีและวิศวกรรมศาสตร์
มหาวิทยาลัยธุรกิจบัณฑิตย์
พ.ศ. 2564

USING DATA MINING TECHNIQUES FOR SCREENING PEOPLE SUCCESSFULLY IN UNDERGRADUATE STUDIES OF EDUCATIONAL TECHNOLOGY

THANAWAT WANNAPRAPHA

A Thesis Submitted in Partial Fulfillment of the Requirements

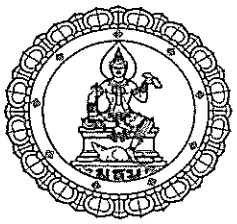
for the Degree of Master of Engineering

Department of Computer and Telecommunication Engineering

College of Innovative Technology and Engineering

Dhurakij Pundit University

2021



ใบรับรองวิทยานิพนธ์

วิทยาลัยนวัตกรรมด้านเทคโนโลยีและวิศวกรรมศาสตร์

มหาวิทยาลัยธุรกิจบัณฑิตย์

ปริญญา วิศวกรรมศาสตรมหาบัณฑิต

หัวข้อวิทยานิพนธ์ การใช้เทคนิคเหมืองข้อมูลเพื่อคัดกรองบุคคลที่มีแนวโน้มสัมฤทธิผลในการศึกษาระดับปริญญาตรี สาขาวิชาเทคโนโลยีการศึกษา

เสนอโดย นายธนะวัฒน์ วรรณประภา

สาขาวิชา วิศวกรรมคอมพิวเตอร์และโทรคมนาคม

อาจารย์ที่ปรึกษาวิทยานิพนธ์ อาจารย์ ดร.ชัยพร เขมะภาตะพันธ์

ได้พิจารณาเห็นชอบโดยคณะกรรมการสอบวิทยานิพนธ์แล้ว

.....ประธานกรรมการ

(รองศาสตราจารย์ ดร.ลัญฉกร วุฒิสิทธิกุลกิจ)

.....กรรมการและอาจารย์ที่ปรึกษาวิทยานิพนธ์

(อาจารย์ ดร.ชัยพร เขมะภาตะพันธ์)

.....กรรมการ

(อาจารย์ ดร.ธนัญ จารุวิทยโกวิท)

.....กรรมการ

(ผู้ช่วยศาสตราจารย์ ดร.มัชฌิกา อ่องแตง)

วิทยาลัยนวัตกรรมด้านเทคโนโลยีและวิศวกรรมศาสตร์รับรองแล้ว

.....คณบดีวิทยาลัยนวัตกรรมด้านเทคโนโลยีและวิศวกรรมศาสตร์

(อาจารย์ ดร.ชัยพร เขมะภาตะพันธ์)

วันที่ ....1..... เดือน ....ก.ค..... พ.ศ. ....2564....

หัวข้อวิทยานิพนธ์	การใช้เทคนิคเหมืองข้อมูลเพื่อคัดกรองบุคคลที่มีแนวโน้มสัมฤทธิผลในการศึกษาระดับปริญญาตรี สาขาวิชาเทคโนโลยีการศึกษา
ชื่อผู้เขียน	นายธนะวัฒน์ วรรณประภา
อาจารย์ที่ปรึกษา	ดร.ชัยพร เขมะภาตะพันธ์
สาขาวิชา	วิศวกรรมคอมพิวเตอร์และโทรคมนาคม
ปีการศึกษา	2563

บทคัดย่อ

งานวิจัยนี้นำเทคนิคการทำเหมืองข้อมูลมาใช้สำหรับการคัดกรองบุคคลที่มีแนวโน้มสัมฤทธิผลสำหรับการเรียนระดับปริญญาตรี สาขาวิชาเทคโนโลยีการศึกษา โดยมีวัตถุประสงค์การวิจัย 1) เพื่อสร้างตัวแบบสำหรับการคัดกรองบุคคลที่มีแนวโน้มสัมฤทธิผลในการศึกษาระดับปริญญาตรี สาขาวิชาเทคโนโลยีการศึกษา 2) เพื่อเปรียบเทียบประสิทธิภาพตัวแบบจำแนกตามร้อยละจำนวนการฝึกสอนและจำนวนข้อมูลฝึกสอนที่แตกต่างกัน ด้วยเทคนิค Decision Tree, Support Vector Machine และ Deep Learning เก็บข้อมูลจากบัณฑิตที่สำเร็จการศึกษาระหว่างปีการศึกษา 2554-2558 จากระบบกองทะเบียนและประมวลผลการศึกษา จำนวน 738 คน ซึ่งถูกใช้เป็นชุดข้อมูลที่ใช้ในการศึกษา โดยได้แบ่งชุดข้อมูลสำหรับการสอนและทดสอบด้วยอัตราส่วน 70:30 75:25 และ 80:20 นอกจากนี้ได้ใช้วิธีการวิเคราะห์ความถดถอยเพื่อคัดกรองจำนวนแอททริบิวต์ตั้งต้นที่มีทั้งหมด 30 แอททริบิวต์ให้ลดลงเหลือเท่าที่จำเป็นด้วยวิธีการ Backward Selection และ Stepwise Selection

ผลการศึกษาพบว่าการวิเคราะห์การถดถอยแบบ Stepwise selection ที่มีการแบ่งกลุ่มการสอนและการทดสอบที่ 80:20 ให้ผลค่าความถูกต้องดีที่สุด โดยตัวแบบที่ใช้วิธีการ Deep Learning ให้ค่าความถูกต้องดีกว่าตัวแบบที่ใช้วิธีการ Decision Tree และ Support Vector Machine ซึ่งมีค่าความถูกต้องเท่ากับ 85.94, 84.46 และ 80.81 ตามลำดับ ทั้งนี้แอททริบิวต์ที่ถูกคัดกรองและส่งผลต่อการวิเคราะห์ของตัวแบบมากที่สุดได้แก่ เกรดเฉลี่ยสะสมมัธยมศึกษาตอนปลายของผู้เรียนในสาขาวิชาเทคโนโลยีการศึกษา และการรับเข้าของนิสิตที่เป็นการรับตรงรอบที่ 5

คำสำคัญ: การทำเหมืองข้อมูล, การจำแนก, แบบจำลอง, Decision Tree, Deep Learning

Thesis Title	USING DATA MINING TECHNIQUES FOR SCREENING PEOPLE SUCCESSFULLY UNDERGRADUATE STUDIES OF EDUCATIONAL TECHNOLOGY
Author	Thanawat Wannaprapha
Thesis Advisor	Dr. Chaiyaporn Khemapatapan
Department	Computer and Telecommunication Engineering
Academic	2020

ABSTRACT

This research applies data mining techniques for screening people who will be successfully in undergraduate studies of the department of Educational Technology. The objectives of this work are 1) to create a model. 2) to study the efficiency of the models, for data mining, with percentage of number training and number of samples by the Decision Tree, Support Vector Machine and Deep Learning.The data are collected from graduated students who graduated during the 2011-2015 academic years. There are, in total, 738 graduated students used as a data set in the study. The data set has been divide for training and test processes with the ratios: 70:30, 75:25 and 80:20. In addition, regression analyses, backward selection and stepwise selection, are applied in order to filter and reduce attributes from a total of 30 to the required attributes.

The studied results found that regression analysis using stepwise selection with training to testing ratio of 80:20 provided best accuracy. A model using deep learning technique has the accuracy better than models using decision tree and support vector machine which are 85.94, 84.46 and 80,81, respectively. Moreover, filtered attributes, which have mostly impact for analysis are High school cumulative GPA and admission of students who are direct admission in the 5th round.

Keywords: Data Mining, Classification, Model, Decision Tree, Deep Learning

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จได้ด้วยความกรุณาเป็นอย่างยิ่งจาก ดร.ชัยพร เขมะภาตะพันธ์ อาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่คอยให้คำแนะนำ ช่วยเหลือตลอดระยะเวลาการทำวิทยานิพนธ์ ผู้วิจัยขอกราบขอบพระคุณเป็นอย่างสูงไว้ ณ โอกาสนี้

ผู้วิจัยขอกราบขอบพระคุณ รองศาสตราจารย์ ดร.ลัญฉกร วุฒิสิทธิกุลกิจ ที่กรุณาให้เกียรติเป็นประธานสอบวิทยานิพนธ์ ขอบพระคุณผู้ช่วยศาสตราจารย์ ดร.มัชฌิมา อ่องแตง และ ดร.ธนัญ จารุวิทยโกวิท เป็นกรรมการสอบวิทยานิพนธ์

ขอบคุณเจ้าหน้าที่มหาวิทยาลัยธุรกิจบัณฑิตย์ทุกคนที่ให้ความสะดวก และประสานงานสำหรับการทำวิทยานิพนธ์ให้ผู้วิจัยด้วยดีตลอดมา ทำให้การจัดทำวิทยานิพนธ์ของผู้วิจัยครั้งนี้สำเร็จลุล่วงไปด้วยดี

ท้ายนี้ ขอกราบขอบพระคุณ คุณพ่อ คุณแม่และครอบครัว ที่คอยเป็นกำลังใจและให้การสนับสนุนผู้วิจัยในทุกๆ ด้าน

ธนะวัฒน์ วรรณประภา

สารบัญ

หน้า

บทคัดย่อภาษาไทย..... ฆ

บทคัดย่อภาษาอังกฤษ..... ง

กิตติกรรมประกาศ..... จ

สารบัญ..... ฉ

สารบัญตาราง..... ซ

สารบัญภาพ..... ญ

บทที่

1 บทนำ..... 1

1.1 ความเป็นมาของปัญหา..... 1

1.2 วัตถุประสงค์การวิจัย..... 3

1.3 ขอบเขตการวิจัย..... 3

1.4 ประโยชน์ที่ได้รับ..... 3

1.5 แผนการดำเนินงาน..... 3

1.6 การตอบรับการตีพิมพ์..... 4

1.7 นิยามศัพท์เฉพาะ..... 5

2 แนวคิด ทฤษฎีและงานวิจัยที่เกี่ยวข้อง..... 6

2.1 เหมืองข้อมูล..... 6

2.2 เทคนิค Decision Tree..... 10

2.3 เทคนิค Support Vector Machine..... 12

2.4 เทคนิค Deep Learning..... 13

2.5 การวิเคราะห์การถดถอย (Regression Analysis)..... 14

2.6 RapidMiner Studio..... 16

2.7 งานวิจัยที่เกี่ยวข้อง..... 17

3 วิธีดำเนินการวิจัย..... 21

3.1 การเลือกข้อมูล (Data Selection)..... 21

3.2 การกลั่นกรองข้อมูล (Data Cleansing)..... 31

3.3 การแปลงข้อมูล (Data Transformation)..... 32

3.4 การสร้างตัวแบบทดสอบและเปรียบเทียบประสิทธิภาพ..... 33

สารบัญ (ต่อ)

บทที่	หน้า
4 ผลการวิจัย	40
4.1 ผลการทดสอบประสิทธิภาพของตัวแบบ	42
4.2 ผลการทดสอบด้วยเทคนิค Decision Tree	42
4.3 ผลการทดสอบด้วยเทคนิค Support Vector Machine	47
4.4 ผลการทดสอบด้วยเทคนิค Deep Learning	48
4.5 สรุปผลการทดลอง	49
5 สรุปผลการวิจัยและข้อเสนอแนะ	55
5.1 สรุปผลการทดลอง	55
5.2 ปัญหาและข้อเสนอแนะ	57
บรรณานุกรม	58
ประวัติผู้วิจัย	62

สารบัญตาราง

ตารางที่ **หน้า**

1.1 แผนการดำเนินงาน..... 4

3.1 การกำหนดตัวแปร..... 24

3.2 Model Summary ของโปรแกรม SPSS วิธีการ Backward..... 27

3.3 Model Summary ของโปรแกรม SPSS วิธีการ Stepwise..... 30

3.4 ตัวแปรสำหรับการประมวลผลข้อมูลด้วยเทคนิคเหมืองข้อมูล..... 33

4.1 จำนวนกลุ่มตัวอย่างสำหรับการเก็บข้อมูล จำแนกตามเพศ..... 40

4.2 จำนวนกลุ่มตัวอย่างที่สำเร็จการศึกษาสาขาวิชาเทคโนโลยีการศึกษา จำแนกตามปีการศึกษาที่เข้าเรียน และประเภทนิสิต..... 41

4.3 สัดส่วนการ Train และการ Test..... 42

4.4 ผลการทดสอบโมเดลด้วยเทคนิค Decision Tree นำข้อมูลเข้าแบบ All Attribute..... 43

4.5 ผลการทดสอบโมเดลด้วยเทคนิค Decision Tree นำข้อมูลเข้าแบบ Backward และ Stepwise..... 44

4.6 ผลการทดสอบโมเดลด้วยเทคนิค Support Vector Machine..... 47

4.7 ผลการทดสอบโมเดลด้วยเทคนิค Deep Learning นำข้อมูลเข้าแบบ All Attribute..... 48

4.8 ผลการทดสอบโมเดลด้วยเทคนิค Deep Learning นำข้อมูลเข้าแบบ Backward..... 48

4.9 ผลการทดสอบโมเดลด้วยเทคนิค Deep Learning นำข้อมูลเข้าแบบ Stepwise..... 49

4.10 ภาพรวมของ Accuracy จำแนกตามเทคนิคเหมืองข้อมูลและจำนวน Attribute..... 49

4.11 เปรียบเทียบความถูกต้องของผลการทดสอบโมเดลด้วยเทคนิคที่ต่างกัน จำแนกตามสัดส่วนการ Train และนำข้อมูลเข้าแบบ All Attribute..... 50

สารบัญตาราง (ต่อ)

ตารางที่ **หน้า**

4.12 เปรียบเทียบความถูกต้องของผลการทดสอบโมเดลด้วยเทคนิคที่ต่างกัน จำแนกตามสัดส่วนการ Train และนำข้อมูลเข้าแบบ Backward……... 51
4.13 เปรียบเทียบความถูกต้องของผลการทดสอบโมเดลด้วยเทคนิคที่ต่างกัน จำแนกตามสัดส่วนการ Train และนำข้อมูลเข้าแบบ Stepwise……. 52
4.14 ค่าเฉลี่ยร้อยละความถูกต้อง (Accuracy) จำแนกตามเทคนิคเหมืองข้อมูล 53
5.1 เปรียบเทียบความถูกต้องของผลการทดสอบโมเดลด้วยเทคนิคที่ต่างกัน จำแนกตามสัดส่วนการ Train และนำข้อมูลเข้าแบบ Stepwise..... 55
5.2 ค่าเฉลี่ยร้อยละความถูกต้อง (Accuracy) จำแนกตามเทคนิคเหมืองข้อมูล.. 56

สารบัญภาพ

ภาพที่ **หน้า**

2.1 การแปลงข้อมูลเป็นความรู้เพื่อช่วยในการตัดสินใจ..... 6

2.2 ตัวอย่างชุดข้อมูลสอน และต้นไม้ตัดสินใจที่สร้างได้จากชุดข้อมูลสอน.... 11

2.3 SVM Model..... 13

2.4 Deep Learning Model..... 14

2.5 เว็บไซต์ RapidMiner Studio..... 16

3.1 การกำหนดตัวแปรในโปรแกรม SPSS..... 25

3.2 ข้อมูลในไฟล์ Excel เป็นลักษณะรูปแบบตัวอักษร..... 26

3.3 ข้อมูลในไฟล์ Excel แทนค่าด้วยรูปแบบตัวเลข..... 26

3.4 ข้อมูลจากไฟล์ Excel มุมมองแบบ Data View..... 27

3.5 Attribute หลังการกรองด้วย Regression..... 31

3.6 การกลั่นกรองข้อมูล..... 32

3.7 การนำเข้าข้อมูลของโปรแกรม RapidMiner Studio..... 35

3.8 การกำหนดประเภทของข้อมูลและกำหนดบทบาทของ Attribute ในโปรแกรม Rapid Miner Studio..... 36

3.9 การกำหนด Validation การ Train 0.70, 0.75 และ 0.80..... 37

3.10 Process ของ Decision Tree..... 37

3.11 Process ของ Support Vector Machine..... 38

3.12 Process ของ Deep Learning..... 39

4.1 ผลการทดสอบโมเดลด้วยเทคนิค Decision Tree นำข้อมูลเข้าแบบ All Attribute..... 43

4.2 ผลการทดสอบโมเดลด้วยเทคนิค Decision Tree นำข้อมูลเข้าแบบ Backward และ Stepwise..... 44

4.3 แบบจำลองที่ได้จากการสร้างด้วยเทคนิค Decision Tree นำข้อมูลเข้าแบบ All Attribute..... 45

4.4 แบบจำลองที่ได้จากการสร้างด้วยเทคนิค Decision Tree นำข้อมูลเข้าแบบ Backward และ Stepwise..... 46

4.5 ผลการทดสอบโมเดลด้วยเทคนิค Support Vector Machine..... 47

สารบัญภาพ (ต่อ)

ภาพที่ **หน้า**

4.6 เปรียบเทียบความถูกต้องของผลการทดสอบโมเดลด้วยเทคนิคที่ต่างกัน จำแนกตามสัดส่วนการ Train และนำข้อมูลเข้าแบบ All Attribute..... 50

4.7 เปรียบเทียบความถูกต้องของผลการทดสอบโมเดลด้วยเทคนิคที่ต่างกัน จำแนกตามสัดส่วนการ Train และนำข้อมูลเข้าแบบ Backward..... 51

4.8 เปรียบเทียบความถูกต้องของผลการทดสอบโมเดลด้วยเทคนิคที่ต่างกัน จำแนกตามสัดส่วนการ Train และนำข้อมูลเข้าแบบ Stepwise..... 52

5.1 เปรียบเทียบความถูกต้องของผลการทดสอบโมเดลด้วยเทคนิคที่ต่างกัน จำแนกตามสัดส่วนการ Train และนำข้อมูลเข้าแบบ Stepwise..... 56

บทที่ 1
บทนำ

1.1 ความเป็นมาของปัญหา

ความก้าวหน้าของวิทยาการและเทคโนโลยีทำให้เกิดการเปลี่ยนแปลงทางสังคม เศรษฐกิจและการเมือง การเปลี่ยนสิ่งเหล่านี้เกิดขึ้นอยู่ตลอดเวลา ในลักษณะที่เป็นพลวัต (Dynamic) ส่งผลให้หลาย ๆ ประเทศทั่วโลกมีการแข่งขันและปรับกระบวนทัศน์ใหม่ (New Paradigm) เพื่อให้ทันกับการเปลี่ยนแปลงของสังคมโลกมากขึ้น กลยุทธ์หนึ่งของหลายประเทศคือการพัฒนาและให้ความสำคัญกับการส่งเสริมการศึกษาให้กับประชาชนเป็นลำดับต้น ๆ ในการพัฒนาประเทศ หลักสูตรให้ความสำคัญในประเด็นดังกล่าวโดยกำหนดวัตถุประสงค์เพื่อผลิตบัณฑิตให้ประกอบวิชาชีพเทคโนโลยีการศึกษา และอาชีพอิสระอื่น ๆ ที่เกี่ยวข้อง เมื่อสำเร็จการศึกษาจากหลักสูตรนี้ บัณฑิตจะสามารถเข้าทำงานในหน่วยงานต่าง ๆ ทั้งในภาครัฐและภาคเอกชน ผลิตบัณฑิตเป็นผู้มีความรู้ ความสามารถ และทักษะทางด้านเทคโนโลยีการศึกษา มีความใฝ่รู้ มีความชาญฉลาดในการนำความรู้ด้านเทคโนโลยีการศึกษาไปใช้ในงานต่าง ๆ ได้ และผลิตบัณฑิตให้มีโลกทัศน์ที่กว้างไกล รู้เท่าทันเทคโนโลยี ยึดมั่นในหลักเหตุผล มีบุคลิกภาพ ทัศนคติ และค่านิยมที่ดี ตามแบบอย่างของวัฒนธรรมไทย [10]

สาขาวิชาเทคโนโลยีการศึกษา เป็นสาขาวิชาที่บูรณาการทางด้านความรู้และทักษะหลอมรวมเข้าด้วยกัน เพื่อพัฒนานิสิตให้มีความสามารถทางด้านการออกแบบพัฒนาการใช้ การจัดการและการประเมินทางด้านนวัตกรรมและเทคโนโลยีเพื่อการแก้ปัญหาทางการเรียนการสอนอย่างเป็นระบบ ด้วยยุคสมัยที่เปลี่ยนแปลงไปนิสิตที่สำเร็จการศึกษาในสาขาวิชานี้มีการประกอบอาชีพทางด้านการผลิตสื่อวิดีทัศน์เพื่อความบันเทิง ผลิตสื่อสิ่งพิมพ์ในโรงพิมพ์หรือทำงานในส่วนของแผนกพัฒนาทรัพยากรมนุษย์ของบริษัท แต่ทั้งนี้นิสิตก็ยังคงใช้ทักษะทางด้านเทคโนโลยีการศึกษาในการพัฒนางาน เช่น ทักษะการถ่ายภาพ ทักษะการถ่ายทำและการตัดต่อวิดีทัศน์ด้วยโปรแกรมคอมพิวเตอร์ รวมทั้งทักษะทางด้านโปรแกรมกราฟิก ทักษะการจัดการฝึกอบรม และทักษะใช้แหล่งทรัพยากรต่าง ๆ ให้เกิดประโยชน์สูงสุดในการพัฒนางานในหน้าที่

กระบวนการคัดเลือกบุคคลเพื่อศึกษาต่อในระดับปริญญาตรี สาขาวิชาเทคโนโลยีการศึกษาที่ผ่านมามีจำนวนผู้ที่สนใจเพื่อเข้าศึกษาต่อในสาขาวิชาดังกล่าวเป็นจำนวนมาก การคัดกรองและการคัดเลือกบุคคลเพื่อเข้าศึกษามีความสำคัญเป็นอย่างมาก หากสามารถคัดกรองและคัดเลือกบุคคลที่มีความพร้อมทั้งความรู้ และความพร้อมทางด้านทักษะ จะทำให้การพัฒนาผู้เรียนเป็นไปตามวัตถุประสงค์ของหลักสูตรดังกล่าวข้างต้น รวมทั้งการต่อยอด สร้างประสบการณ์การเรียนรู้และพัฒนาศักยภาพของผู้เรียนได้เต็มตามศักยภาพในตัวของผู้เรียน กระบวนการคัดเลือกแบบเดิมมี 2 ขั้นตอนคือ 1) การทดสอบความรู้ และ 2) การสัมภาษณ์ ซึ่งอาจจะเป็นกระบวนการที่ไม่สามารถคัดกรองและคัดเลือกบุคคลได้เป็นไปตามศักยภาพที่ทางสาขาวิชาต้องการ เพราะบางปีการศึกษามีผู้เรียนที่ไม่พึ่งประสงค์ ไม่เป็นไปตามความต้องการของหลักสูตรสามารถผ่านกระบวนการคัดเลือกและเข้ามาเรียนได้ และเมื่อเข้ามาเรียนได้บางส่วนไม่สามารถสำเร็จการศึกษาตามที่หลักสูตรกำหนด บางส่วนออกกลางคัน ถ้าสามารถคัดกรองผู้เรียนส่วนนี้ได้ตั้งแต่ต้นจะทำให้หลักสูตรได้ผู้เรียนที่มีคุณภาพที่ดีขึ้น

จากการศึกษาแนวทางการแก้ไขปัญหาดังกล่าวข้างต้น มีหลากหลายวิธีการที่นำเทคนิคการทำเหมืองข้อมูลมาใช้ เช่น การใช้เทคนิคเหมืองข้อมูลในการเลือกกลุ่มสาขาวิชาที่เหมาะสมสำหรับการศึกษาต่อระดับปริญญาตรี [17] การวิเคราะห์ปัจจัยที่มีผลต่อการเลือกสาขาวิชาของนักศึกษาระดับปริญญาตรี คณะครุศาสตร์ โดยใช้เทคนิคการทำเหมืองข้อมูล [11] การใช้เทคนิคเหมืองข้อมูลในการเลือกกลุ่มสาขาวิชาที่เหมาะสม สำหรับการศึกษาต่อระดับปริญญาตรี [1] เป็นการนำเทคนิคเหมืองข้อมูล (Data Mining) มาสร้างตัวแบบ (Model) เพื่อช่วยให้การคัดกรองหรือการจำแนกประเภททำได้ดีขึ้นมีประสิทธิภาพประสิทธิผลดีกว่ากระบวนการเดิมที่เคยปฏิบัติกันมา

ผู้วิจัยเห็นประโยชน์ของการนำเทคนิคเหมืองข้อมูลที่สามารถนำมาคัดกรองบุคคลเพื่อเข้าการศึกษาระดับปริญญาตรี ถ้าสาขาวิชาเทคโนโลยีการศึกษา มีระบบการคัดกรองที่แม่นยำและสามารถตัดสินใจเพื่อคัดเลือกบุคคลเข้าศึกษาต่อระดับปริญญาตรีจะช่วยให้กระบวนการเดิมคัดกรองบุคคลได้มีคุณภาพมากขึ้น หรือได้บุคคลที่มีความเหมาะสมสอดคล้องกับความต้องการของหลักสูตรเข้าศึกษาต่อในระดับปริญญาตรีได้

1.2 วัตถุประสงค์การวิจัย

1. เพื่อสร้างตัวแบบสำหรับการคัดกรองบุคคลที่มีแนวโน้มสัมฤทธิผลในการศึกษาระดับปริญญาตรี สาขาวิชาเทคโนโลยีการศึกษา

2. เพื่อเปรียบเทียบประสิทธิภาพของตัวแบบจำแนกตามร้อยละจำนวนการฝึกสอนและจำนวนข้อมูลฝึกสอนที่แตกต่างกัน ด้วยเทคนิค Decision Tree, Support Vector Machine และ Deep Learning

1.3 ขอบเขตการวิจัย

1. ขอบเขตด้านข้อมูลที่นำมาสร้างตัวแบบเป็นข้อมูลผู้สำเร็จการศึกษา สาขาวิชาเทคโนโลยีการศึกษา จากระบบทะเบียนและสถิติจำนวน 738 คน ระหว่างปีการศึกษา 2554 – 2558 โดยเป็นผู้เรียนที่อยู่ภายใต้หลักสูตรเดียวกัน

2. ขอบเขตด้านเทคนิคการสร้างตัวแบบ

- Decision Tree
- Support Vector Machine
- Deep Learning

1.4 ประโยชน์ที่ได้รับ

1. ได้ตัวแบบสำหรับการคัดกรองบุคคลที่มีแนวโน้มสัมฤทธิผลในการศึกษาระดับปริญญาตรี สาขาวิชาเทคโนโลยีการศึกษา

2. สามารถจำแนกหรือคัดกรองบุคคลที่มีแนวโน้มสัมฤทธิผลในการศึกษาต่อระดับปริญญาตรี สาขาวิชาเทคโนโลยีการศึกษาได้ และมีความถูกต้องมากกว่าร้อยละ 80

3. ผู้บริหารหน่วยงานสามารถนำตัวแบบไปประยุกต์ใช้กับกระบวนการคัดเลือกเดิมได้อย่างมีประสิทธิภาพที่ดีขึ้น

1.5 แผนการดำเนินงาน

ผู้วิจัยกำหนดแผนการดำเนินงานวิจัยดังตารางที่ 1.1

ตารางที่ 1.1 แผนการดำเนินงาน

รายการดำเนินงาน	ระยะเวลา (เดือน)											
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
1. รวบรวมข้อมูลและศึกษาข้อมูลที่ได้จากระบบทะเบียนและสถิติ												
2. ศึกษาการทำงานของเทคนิคการทำเหมืองข้อมูล												
3. วิเคราะห์และเลือกข้อมูล (Data selection) ทำชุดข้อมูลในรูปแบบของไฟล์ Excel												
4. กลั่นกรองข้อมูล (Data cleansing)												
5. แปลงข้อมูล (Data transformation)												
6. สร้างตัวแบบ ทดสอบและเปรียบเทียบประสิทธิภาพ												
7. วิเคราะห์ข้อมูลและสรุปผลการวิจัย												
8. รวบรวมข้อมูลและจัดทำเล่มวิทยานิพนธ์												

1.6 การตอบรับการตีพิมพ์

งานวิจัยนี้ได้รับการตอบรับการตีพิมพ์ในการประชุมวิชาการระดับชาติ วิทยาศาสตร์และเทคโนโลยีครั้งที่ 5 (The 5th National Conference on Science and Technology) ปี 2564 นำเสนอเมื่อวันที่ 15 มกราคม พ.ศ. 2564 ภายใต้ชื่อหัวข้อว่า “การคัดกรองบุคคลที่มีแนวโน้มสัมฤทธิผลสำหรับการเรียนระดับปริญญาตรี สาขาวิชาเทคโนโลยีการศึกษา โดยใช้เทคนิคเหมืองข้อมูล SCREENING PEOPLE WHO ARE SUCCESSFUL FOR UNDERGRADUATE STUDIES DEPARTMENT OF EDUCATIONAL TECHNOLOGY BY DATA MINING TECHNIQUES”

1.7 นิยามศัพท์เฉพาะ

1. การคัดกรองบุคคล หมายถึง การจำแนกบุคคลที่มีแนวโน้มความสามารถเข้าศึกษาต่อระดับปริญญาตรี สาขาวิชาเทคโนโลยีการศึกษา โดยการศึกษาจากแอททริบิวต์ต่าง ๆ ที่สะท้อนถึงความสามารถเข้าศึกษาต่อตามที่หลักสูตรคาดหวัง

2. สัมฤทธิผล หมายถึง ความสำเร็จของผู้เรียนในระดับปริญญาตรี สาขาวิชาเทคโนโลยีการศึกษา ที่พิจารณาจากผลการเรียนเฉลี่ยสะสมระหว่าง 3.00 – 4.00 กำหนดเป็นสัมฤทธิผลดี และผลการเรียนเฉลี่ยสะสมระหว่าง 2.00 – 2.99 กำหนดเป็นสัมฤทธิผลไม่ดี

3. ประสิทธิภาพ หมายถึง อัลกอริทึมหรือเทคนิคของเหมืองข้อมูลที่ทำหน้าที่จำแนกข้อมูลที่ผู้วิจัยรวบรวมขึ้นด้วยเทคนิค Decision Tree, Support Vector Machine และ Deep Learning

4. เหมืองข้อมูล หมายถึง กระบวนการคัดกรองพฤติกรรมที่สนใจจากข้อมูลที่มีเป็นจำนวนมาก เพื่อค้นหารูปแบบและความสัมพันธ์ที่มีอยู่จริงแต่ถูกซ่อนอยู่ในข้อมูลจำนวนมากนั้น

5. Accuracy หมายถึง ค่าร้อยละความถูกต้องที่ได้จากการประเมินประสิทธิภาพการจำแนกประเภทข้อมูลโดยรวมทุกคลาสของโมเดล

6. Correlation หมายถึงค่าสหสัมพันธ์ เป็นการพิจารณาทิศทางความสัมพันธ์ระหว่างตัวแปร 2 ตัว จะมีค่าอยู่ระหว่าง -1.0 ถึง +1.0 ซึ่งหากมีค่าใกล้ -1.0 หมายความว่าตัวแปรทั้งสองตัวมีความสัมพันธ์กันอย่างมากในเชิงตรงกันข้าม หากมีค่าใกล้ +1.0 หมายความว่า ตัวแปรทั้งสองมีความสัมพันธ์กันโดยตรงอย่างมาก และหากมีค่าเป็น 0 หมายความว่า ตัวแปรทั้งสองตัวไม่มีความสัมพันธ์ต่อกัน

7. Specificity หมายถึง ค่าทางสถิติที่แสดงความแม่นยำของการทำนายข้อมูลที่อยู่ในคลาส Negative โดยหาจากอัตราส่วนของการทำนายข้อมูลที่อยู่ในคลาส Negative ได้ถูกต้องเทียบกับข้อมูลจริงทั้งหมดของคลาส Negative

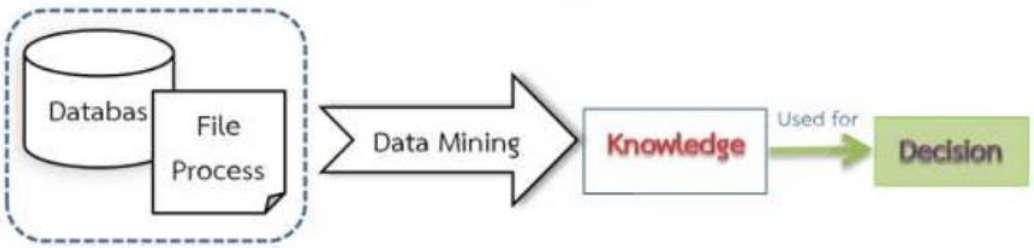
8. Cross-Entropy หมายถึง ความไม่เป็นระเบียบของข้อมูลหรือความแตกต่างระหว่างการแจกแจงของข้อมูล เป็นการจัดกลุ่มของข้อมูลว่าไปอย่างที่ควรจะเป็นมากน้อยเพียงใด ถ้ามีค่าเท่ากับ 1 แสดงว่าข้อมูลมีความแตกต่างกันสูง ถ้ามีค่าเท่ากับ 0 แสดงว่าข้อมูลมีความเหมือนกันสูง ค่า Cross-Entropy เป็นค่าพิจารณาสำหรับการตัดสินใจว่าจะใช้ Model ใด

บทที่ 2
แนวคิด ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

การวิจัยการใช้เหมืองข้อมูลเพื่อคัดกรองบุคคลที่มีแนวโน้มสัมฤทธิผลในการศึกษาระดับปริญญาตรี สาขาวิชาเทคโนโลยีการศึกษา ผู้วิจัยศึกษาการทำเหมืองข้อมูล เทคนิค Decision Tree เทคนิค Support Vector Machine เทคนิค Deep Learning การวิเคราะห์การถดถอย และงานวิจัยที่เกี่ยวข้อง โดยมีรายละเอียดดังนี้

2.1 เหมืองข้อมูล

เหมืองข้อมูล (Data Mining) คือ การนำเทคนิคการเรียนรู้ของเครื่อง (Machine Learning)วิธีการทางสถิติ (Statistical Methods) วิธีทางปัญญาประดิษฐ์ (Artificial Intelligence) หรือวิธีอื่นๆ มาท การวิเคราะห์และสกัดความรู้จากข้อมูลที่จัดเก็บไว้ในฐานข้อมูลหรือจัดเก็บไว้ในรูปอื่นๆ โดยจุดประสงค์ของการทำเหมืองข้อมูล คือ การวิเคราะห์แนวโน้ม ความสัมพันธ์ กฎหรือรูปแบบของข้อมูล ซึ่งเป็นความรู้ที่ถูกซ่อนอยู่ภายใต้ข้อมูลขนาดใหญ่ และนำสารสนเทศที่ได้มาใช้เพื่อช่วยในการวางแผนการตัดสินใจในการบริหารหรือแก้ปัญหาด้านต่างๆ ซึ่งถือได้ว่าเป็นเครื่องมือที่ช่วยเพิ่มคุณค่าให้กับข้อมูลที่มีอยู่เหมืองข้อมูลจะสามารถแก้ปัญหาได้บางปัญหาตามเทคนิควิธีการที่เลือกใช้ได้เท่านั้นโดยประโยชน์หลักของเหมืองข้อมูลคือการค้นหาความรู้ที่ซ่อนอยู่ในฐานข้อมูลเพื่อให้ได้ซึ่งความรู้มาช่วยประกอบในการตัดสินใจรายละเอียดการแปลข้อมูลมาเป็นความรู้เพื่อช่วยในการประกอบการตัดสินใจ ดังภาพที่ 2.1



ภาพที่ 2.1 การแปลงข้อมูลเป็นความรู้เพื่อช่วยในการตัดสินใจ [16]

ประเภทของการทำเหมืองข้อมูล [26, 27]

1. การจำแนกประเภทข้อมูล (Classification) เป็นการทำเหมืองข้อมูลที่มีจุดประสงค์ที่สร้างตัวแบบจำแนกประเภทข้อมูลจากข้อมูลเรียนรู้ที่มีการจำแนกประเภทแล้ว และใช้ตัวแบบที่สร้างได้จำแนกข้อมูลใหม่ที่ไม่ทราบประเภท

2. การจับกลุ่มข้อมูล (Clustering) เป็นการแบ่งข้อมูลออกเป็นกลุ่ม โดยข้อมูลที่อยู่ในกลุ่มเดียวกันจะมีลักษณะที่คล้ายคลึงกัน และข้อมูลที่อยู่ต่างกลุ่มจะมีลักษณะที่แตกต่างกัน ซึ่งการจับกลุ่มข้อมูลดังกล่าวต้องอาศัยมาตรวัดความเหมือนหรือแตกต่างระหว่างข้อมูลใด ๆ เพื่อจัดข้อมูลที่มีลักษณะคล้ายคลึงกันเข้าอยู่ในกลุ่มเดียวกัน

3. การสร้างกฎความสัมพันธ์ (Association Rule) เป็นการทำเหมืองข้อมูลที่ต้องการหากฎที่บอกความสัมพันธ์ระหว่างเหตุการณ์ที่มักเกิดขึ้นพร้อม ๆ กันอยู่เสมอ ข้อมูลที่จะทำการวิเคราะห์อาจเป็นรายการสินค้าที่ลูกค้าสั่งซื้อในแต่ละครั้ง กฎความสัมพันธ์ที่สร้างได้จะระบุถึงความสัมพันธ์ว่าเมื่อพบเหตุการณ์หนึ่งหรือหลายเหตุการณ์เกิดขึ้น จะมีโอกาสสูงที่เหตุการณ์อีกอย่างหนึ่งหรืออีกหลายเหตุการณ์จะเกิดขึ้นด้วย

4. การคาดคะเน (Estimation) มีลักษณะคล้ายกับการจำแนกประเภทข้อมูล คือ มีการสร้างตัวแบบเพื่อคาดคะเนค่าของตัวแปรเป้าประสงค์จากค่าตัวแปรอื่น ๆ ข้อแตกต่าง คือ ตัวแปรที่จะถูกคาดคะเนจะมีชนิดเป็นตัวเลข แทนที่จะเป็นประเภทข้อมูล เช่น การประมาณค่าปริมาณการขายสินค้าช่วงวันหยุด เป็นต้น

5. การค้นหาข้อมูลที่มีความผิดปกติ (Anomaly Detection) เป็นการค้นหาข้อมูลที่มีลักษณะผิดปกติ หรือความแตกต่างจากข้อมูลส่วนใหญ่ (Outlier Data) ข้อมูลผิดปกติเหล่านี้จะมีจำนวนน้อยปะปนอยู่ในข้อมูลส่วนใหญ่ ซึ่งอาจส่งผลให้การวิเคราะห์ข้อมูลมีความถูกต้องลดน้อยลง

ขั้นตอนการทำเหมืองข้อมูล [13]

การทำเหมืองข้อมูลมีขั้นตอนดังต่อไปนี้

1. ทำความเข้าใจปัญหา (Problem Understanding) ประกอบด้วย

1.1 กำหนดวัตถุประสงค์ (Determine Objective) คือการกำหนดวัตถุประสงค์ว่าการทำเหมืองข้อมูลครั้งนี้ต้องการที่จะแก้ปัญหาใด เช่น เหมืองข้อมูลเพื่อต้องการพยากรณ์การเพิ่มขึ้นของจำนวนผู้สมัครคัดเลือกบุคคลเข้าศึกษาในระดับปริญญาตรี เป็นต้น

1.2 กำหนดเกณฑ์ความสำเร็จ (Define Success Criteria) คือ การกำหนดเกณฑ์วัดความสำเร็จในการทำเหมืองข้อมูล ซึ่งอาจเป็นได้ทั้งความสำเร็จในด้านรูปธรรม เช่น สามารถเพิ่มจำนวนผู้สมัครในปีการศึกษานี้ได้ ร้อยละ 5 จากแผนการรับเข้าศึกษา และความสำเร็จใน

ด้านนามธรรม เช่น สามารถค้นพบความรู้ใหม่จากข้อมูลการรับสมัครเข้าศึกษาในระดับปริญญาตรี เป็นต้น

1.3 ประเมินสถานการณ์ (Assess Situation) คือ การประเมินสถานการณ์ในด้านต่างๆ เช่น ความรู้พื้นฐานในเรื่องที่จะทำเหมืองข้อมูล จำนวนข้อมูลมีเพียงพอหรือไม่ และผลประโยชน์จากการทำเหมืองข้อมูลจะคุ้มค่ากับต้นทุนที่เสียไปหรือไม่ เป็นต้น

1.4 กำหนดเป้าหมายในเชิงการทำเหมืองข้อมูล (Determine Data Mining Goals) คือ การตั้งเป้าหมายในเชิงการทำเหมืองข้อมูล ซึ่งต่างไปจากเป้าหมายหลักในการแก้ปัญหา เช่น เป้าหมายหลักคือต้องการเพิ่มจำนวนผู้สมัครคัดเลือกบุคคลเข้าศึกษาในระดับปริญญาตรี เป้าหมายในเชิงการทำเหมือง การทำเหมืองข้อมูล คือการหาลักษณะของผู้สมัครที่มีแนวโน้มต้องการจะสมัครเข้าศึกษา

1.5 วางแผนการทำเหมืองข้อมูล (Produce Project Plan) คือ การวางแผนการทำเหมืองข้อมูลว่าจะเก็บข้อมูลอย่างไร ใช้เทคนิคหรือวิธีการใดในการทำเหมืองข้อมูล

2. ทำความเข้าใจข้อมูล (Data Understanding) ประกอบไปด้วยกระบวนการย่อย ดังนี้

2.1 เก็บรวบรวมข้อมูล (Collect Initial Data)

2.2 กำหนดคุณสมบัติของข้อมูลที่เก็บมาได้ (Define Success Criteria)

2.3 อธิบายข้อมูล (Describe Data)

2.4 ตรวจสอบข้อมูลเบื้องต้น (Explore Data) คือ ตรวจสอบข้อมูลทั้งความสมบูรณ์และความถูกต้องของข้อมูล

2.5 ปรับปรุงข้อมูลให้มีคุณภาพ (Verify Data Quality)

3. การเตรียมข้อมูล (Data Preparation) ประกอบด้วยกระบวนการย่อย ดังนี้

3.1 การคัดเลือกข้อมูลที่จะนำมาใช้ (Select Data)

3.2 การทำความสะอาดข้อมูล (Clean Data) เป็นกระบวนการเตรียมข้อมูลให้เหมาะสมที่สุดเพื่อนำไปใช้ในขั้นตอนต่อไป ซึ่งมีวิธีการต่างๆ หลายวิธี ได้แก่

3.2.1 การแก้ไขข้อมูลให้ถูกต้องสมบูรณ์ เช่นการแก้ไขค่าว่างของข้อมูล (Missing Value) โดยใส่ค่า 0

3.2.2 ปรับเปลี่ยนข้อมูลให้มีค่าเหมาะสมในการตัดสินใจ เช่น ข้อมูลที่มีค่า “Coke” และ “Pepsi” อาจเปลี่ยนค่าให้เป็น “น้ำอัดลม”

3.2.3 เลือกข้อมูลเฉพาะที่สนใจ เช่น ต้องการหาลักษณะลูกค้าที่ซื้อรถ Sport ไม่ควรนำรายชื่อพนักงานขายเข้ามาเกี่ยวข้อง

3.2.4 คอลัมน์ที่มีค่าสำหรับทุกแถวเป็นค่าเดียวกัน เช่น “สัญชาติไทย” หรือคอลัมน์ที่มีค่าที่ไม่ซ้ำกันเลย เช่น “หมายเลขสมาชิก” ไม่ควรนำมาใช้เนื่องจากไม่สามารถบอกรูปแบบของข้อมูลได้

3.3 เปลี่ยนแปลงรูปแบบข้อมูล (Transform Data) การเตรียมพร้อมข้อมูลสำหรับการวิเคราะห์ตัวแปรที่ถูกเลือกมาแต่ละตัวนั้นจะต้องถูกทำความเข้าใจว่าตัวแปรแต่ละตัวหมายความว่าอย่างไร ประกอบด้วยอะไร มีค าอธิบายชัดเจนเกี่ยวกับข้อมูล ค่าที่เป็นไปได้ แหล่งกำหนดข้อมูล รูปแบบ และลักษณะอื่นๆ

4. การสร้างตัวแบบ (Modeling) ประกอบด้วยกระบวนการย่อย ดังนี้

4.1 การเลือกเทคนิค หรือวิธีการที่เหมาะสมในการทำเหมืองข้อมูล (Select Modeling Technique)

4.2 กำหนดรูปแบบการทดสอบผลลัพธ์ (Generate Test Design)

4.3 สร้างแบบจำลองตามอัลกอริทึมที่เลือก (Build a Model)

4.4 ทดสอบตัวแบบ (Assess the Model) ทดสอบตัวแบบที่ได้มานั้นว่ามีความถูกต้องและน่าเชื่อถือเพียงใด

5. การประเมินผล (Evaluation) การประเมินผลอาจจะประเมินตัวแบบที่สร้างขึ้นด้วยการทดลองนำไปใช้กับสถานการณ์จริงหรือกับสถานการณ์ที่จำลองขึ้น เพื่อดูว่าตัวแบบนี้ได้ผลหรือไม่เพียงใด และผิดพลาดตรงไหนถ้าหากผิดพลาดก็ทำการแก้ไขในกระบวนการก่อนหน้าก่อนที่จะนำไปใช้งานจริงประกอบด้วยกระบวนการย่อย ดังนี้

5.1 ประเมินตัวแบบที่สร้างขึ้น (Evaluate Results) การประเมินตัวแบบที่สร้างขึ้นด้วยการทดลองนำไปใช้กับสถานการณ์จริง หรือจำลองขึ้นเพื่อดูว่าตัวแบบที่สร้างได้ผลหรือไม่เพียงใด

5.2 ทบทวนตัวผลการทดสอบตัวแบบ (Review Process)

5.3 พิจารณาขั้นตอนการดำเนินงานกระบวนการที่ผ่านมาว่ามีตรงไหนที่ผิดพลาดอย่างไรก่อนจะนำไปใช้จริง (Determine Next Steps)

6. นำไปใช้และตรวจสอบ (Deployment)

การนำไปใช้และตรวจสอบผลว่าบรรลุเป้าหมายที่ตั้งไว้เพียงใด ประกอบด้วยกระบวนการย่อย ดังนี้

6.1 วางแผนสำหรับการติดตั้งใช้งาน (Plan the Deployment)

6.2 ติดตามให้การช่วยเหลือระหว่างการใช้งาน (Monitor and Maintain)

6.3 ทำรายงานสรุปผลการดำเนินการทั้งหมด (Final Report)

รูปแบบการทำเหมืองข้อมูล [14]

การทำเหมืองข้อมูลสามารถจำแนกตามจุดประสงค์ของการทำเหมืองข้อมูลเป็น 2 ลักษณะ ดังนี้

1. การสร้างตัวแบบสำหรับการบรรยาย (Descriptive Data Mining) เป็นการทำเหมืองข้อมูลที่มีจุดประสงค์ที่จะค้นหา Pattern หรือลักษณะของข้อมูลภายในชุดข้อมูลที่กำหนดให้ เช่น ลักษณะการรวมตัวเป็นกลุ่มของข้อมูล (Clustering) การปรากฏบ่อย ๆ ของรายการข้อมูล (Frequent Patterns) ความสัมพันธ์ของรายการข้อมูลที่มักปรากฏพร้อม ๆ กัน (Association Rules) ข้อมูลที่มีลักษณะผิดแปลกจากข้อมูลส่วนใหญ่ในชุดข้อมูล (Data Anomaly) เป็นต้น การทำเหมืองข้อมูลประเภทนี้ไม่ต้องอาศัยชุดข้อมูลเรียนรู้ในการค้นหารูปแบบที่ต้องการ จึงจัดเป็นการเรียนรู้ที่ไม่ต้องการแนะนำ (Unsupervised Learning)

2. การสร้างตัวแบบสำหรับการทำนาย (Predictive Data Mining) เป็นการทำเหมืองข้อมูลที่มีจุดประสงค์ที่จะหาค่าหรือทำนายค่าของตัวแปรเป้าประสงค์ จากตัวแปรอิสระ การทำเหมืองข้อมูลประเภทนี้จะใช้ข้อมูลชุดเรียนรู้ (Training Data) ในการสร้างตัวแบบที่จะทำนายค่าของตัวแปรเป้าประสงค์ การทำเหมืองข้อมูลในลักษณะนี้จัดเป็นการเรียนรู้ที่ได้รับการแนะนำ (Supervised Learning) เช่น การจำแนกประเภทข้อมูล (Classification) การคาดคะเน (Estimation) เป็นต้น

2.2 เทคนิค Decision Tree

[4] เดช ธรรมศิริ และคณะ ได้กล่าวในงานวิจัยว่าเทคนิคต้นไม้ตัดสินใจ (Decision Tree) เป็นการนำข้อมูลมาสร้างตัวแบบการพยากรณ์ในรูปของต้นไม้ช่วยตัดสินใจ ซึ่งต้นไม้ช่วยตัดสินใจนั้นทำงานแบบ Supervised Learning คือ สามารถสร้างตัวแบบการจัดหมวดหมู่ได้จากกลุ่มตัวอย่างของข้อมูลที่ได้กำหนดก่อนล่วงหน้าที่เรียกว่า Training Set ได้อัตโนมัติและสามารถพยากรณ์กลุ่มรายการที่ยังไม่เคยนำมาจัดหมวดหมู่ได้ด้วย

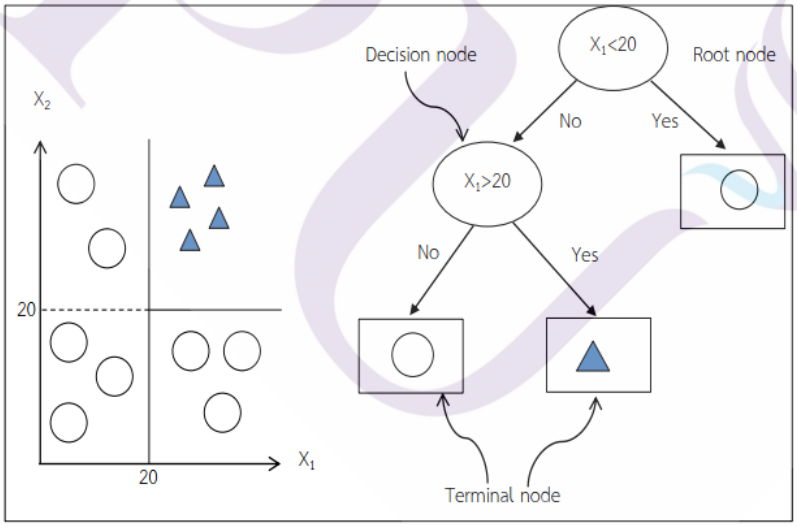
[6] บุญมา เพ่งช่วน กล่าวในงานวิจัยว่าต้นไม้ตัดสินใจมีลักษณะคล้ายต้นไม้ประกอบด้วยโหนดภายใน (Internal node) จะแสดงคุณลักษณะ (Attribute) ของข้อมูลโดยที่จุดเริ่มต้นของต้นไม้เรียกว่า โหนดราก แต่ละกิ่งแสดงค่าของคุณลักษณะของแต่ละโหนด และลีฟโหนดแสดงกลุ่ม (Class) ซึ่งเป็นผลลัพธ์ที่สามารถแยกแยะได้

[7] บุญเสริม กิจศิริกุล กล่าวว่าต้นไม้ตัดสินใจเป็นวิธีการเรียนรู้ที่นิยมใช้มากแบบหนึ่งในการเรียนรู้ของเครื่องการเรียนรู้แบบนี้เป็นการเรียนรู้โดยการแยกแยะ (Classification) ข้อมูลออกเป็นกลุ่ม (Class) ต่างๆ โดยใช้คุณสมบัติ (Attribute) ของข้อมูลในการแยกแยะต้นไม้ตัดสินใจ

ที่ได้จากการเรียนรู้ทำให้ทราบว่าคุณสมบัติใดของข้อมูลที่เป็นตัวกำหนดแยกแยะ และคุณสมบัติแต่ละตัวของข้อมูลมีความสำคัญมากน้อยต่างกันอย่างไรซึ่งเป็นประโยชน์ช่วยให้ผู้ใช้สามารถวิเคราะห์ข้อมูลและตัดสินใจได้ถูกต้องยิ่งขึ้น

[2] ชุติมา อุตมะมุณีย์ และประสงค์ ปราณีตพลกรัง กล่าวว่าต้นไม้ตัดสินใจเป็นกราฟต้นไม้ที่ประกอบด้วยจุดยอดและด้าน โดยจุดยอดภายในจะแทนลักษณะประจำที่นำมาใช้แยกกลุ่มข้อมูลด้านจะแทนค่า หรือผลของการทดสอบ ใบจะแทนค่าของคลาสหรือการกระจายของคลาส

[1] ฉัตรเกล้า เจริญผล กล่าวว่าต้นไม้ตัดสินใจเป็นเทคนิคของการเรียนรู้ (Learning Algorithm) ทำการสร้างตัวแบบที่ใช้สำหรับคาดคะเนหรือทำนายเหตุการณ์ล่วงหน้า เป็นผลลัพธ์ที่ได้จากการตัดสินใจจากต้นไม้จะไม่ซับซ้อน โดยจะมีการแตกแขนงจากโหนดราก (Root) สู่ใบ (Leaf) และมีกิ่งก้าน (Branch) แตกออกไปตามเงื่อนไขหรือข้อมูลที่คาดคะเนว่าจะเกิดขึ้นจากการศึกษาสรุปได้ว่าเทคนิคต้นไม้ตัดสินใจเป็นโมเดลที่มีรูปแบบที่ได้รับความนิยม โครงสร้างต้นไม้ตัดสินใจเป็นแบบลำดับชั้นโดยมีการตัดสินใจ ซึ่งประกอบด้วย โหนดที่ใช้ในการตัดสินใจ (Decision Node) และโหนดใบ (Leaf node or Terminal node) ซึ่งแต่ละโหนดตัดสินใจนั้นจะมีการสร้างฟังก์ชันที่เอาไว้สำหรับทดสอบทางเลือก f_m (x) จากการป้อนข้อมูลเข้า (Input) จะทดสอบตามทางเลือกไปเรื่อยๆ ไปจนถึง Terminal node จะได้คำตอบในที่สุด ดังภาพ 2.2



ภาพที่ 2.2 ตัวอย่างชุดข้อมูลสอนและต้นไม้ตัดสินใจที่สร้างได้จากชุดข้อมูลสอน [1]

ขั้นตอนการสร้างต้นไม้ตัดสินใจ

ในการสร้างแผนภาพต้นไม้ตัดสินใจนั้นจะเริ่มจากการที่มีข้อมูลอยู่นำข้อมูลมาแบ่งออกเป็นโครงสร้างต้นไม้ โดยที่แผนภาพนั้นจะเป็นโครงสร้างที่มีกฎต่าง ๆ เกิดขึ้นตามเป้าหมายของการใช้งานในแต่ละงาน โดยเมื่อได้โครงสร้างต้นไม้แล้วจะสามารถนำโครงสร้างที่ได้นั้นไปใช้งานกับข้อมูลอื่น ๆ ได้โดยนำข้อมูลนั้นผ่านกฎการตัดสินใจ (Decision Rules) เพื่อให้ได้โครงสร้างต้นไม้ตามความต้องการในขั้นตอนการสร้างแผนภาพต้นไม้ [19] มีดังต่อไปนี้

1. หาโหนดราก (Root Node) โดยการหาคุณลักษณะ (Attribute) ที่มีความน่าเชื่อถือมากที่สุดโดยใช้อัลกอริทึม เช่น CART, ID3, C4.5 เป็นต้น

2. เมื่อได้โหนดรากแล้วให้ทำการหาค่าความเป็นไปได้ของโหนดรากนั้นออกมาเป็นกิ่งของต้นไม้ตัดสินใจ

3. แบ่งข้อมูลตามกิ่งของต้นไม้ตัดสินใจ

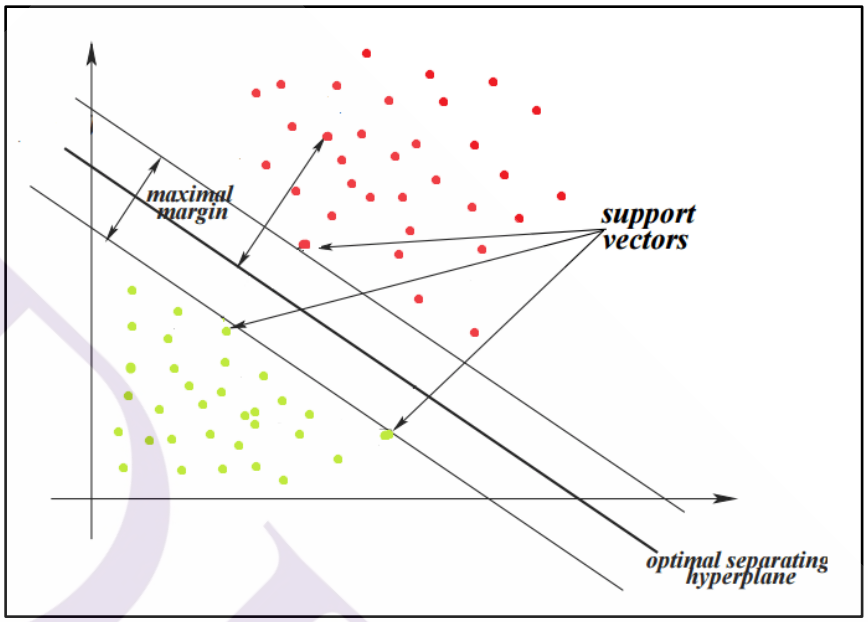
4. นำข้อมูลแต่ละกลุ่มที่ได้แบ่งแล้วมาทำซ้ำตั้งแต่ขั้นตอนเริ่มต้น และต้นไม้ตัดสินใจจะสิ้นสุดที่โหนดใบ (Leaf Node)

2.3 เทคนิค Support Vector Machine

เป็นอัลกอริทึมสำหรับในการคัดแยกที่มีการนำมาใช้กันอย่างกว้างขวางในด้านการประมวลผลภาพดิจิตอล หลักการของ SVM คือการให้ Input ที่ใช้ฝึกเป็นเวคเตอร์ในสเปซ N มิติ เช่นถ้าในกรณีของ 2 มิติ และ 3 มิติ จะเป็นจุดที่อยู่ในระนาบ xy และสเปซ xyz ตามลำดับ จากนั้นทำการสร้างไฮเปอร์เพลน (Hyperplane) ที่จะแยกกลุ่มของเวคเตอร์อินพุทออกเป็นประเภทต่าง ๆ ในกรณีที่เป็น 2 มิติ และ 3 มิติ ไฮเปอร์เพลน คือเส้นตรงและระนาบตามลำดับ ข้อเด่นของ SVM จะทำการเก็บแมพ (Map) เวคเตอร์ในสเปซ Input ให้เข้าสู่ Feature Space โดยใช้ฟังก์ชั่นหรือเรียกว่าเคอร์นัล (Kernel) ชนิดต่าง ๆ เช่น โพลีโนเมียล (Polynomial) เรเดียล (Radial) เป็นต้น ใน Feature Space ดังกล่าวเวคเตอร์อินพุท สามารถแยกประเภทได้โดยไฮเปอร์เพลน

นอกจากนั้น Support Vector Machine เป็นตัวแบบที่ใช้ในการระบุตัวบุคคลหรือ Object โดย SVM จะทำการแบ่งชั้นของข้อมูลด้วยระนาบหลายมิติ จากข้อมูล 2 กลุ่ม โดยตัวแบบของ SVM เกี่ยวข้องกับเครือข่ายประสาทเทียม ซึ่งตัวแบบของ SVM ใช้ Sigmoid Kernel Function มีค่าเท่ากันทั้ง 2 เลเยอร์เป็นตัวแบบที่ใช้ในการระบุตัวบุคคล โดย SVM จะทำการแบ่งชั้นของข้อมูลด้วยระนาบหลายมิติ จากข้อมูล 2 กลุ่มชุดข้อมูล โดยตัวแบบของ SVM เกี่ยวข้องกับเครือข่ายประสาทเทียม ซึ่งโดยอันที่จริงแล้วตัวแบบของ SVM ใช้ Sigmoid Kernel Function ซึ่งมีค่าเท่ากัน

ทั้ง 2 เลเยอร์ ตัวแบบของ SVM มีความคล้ายคลึงกับเพอร์เซฟตรอนซึ่งเป็นข่ายงานประสาทเทียมแบบง่ายมีหน่วยเดียวที่จำลองลักษณะของเซลล์ประสาท ด้วยการใช้ Kernel Function

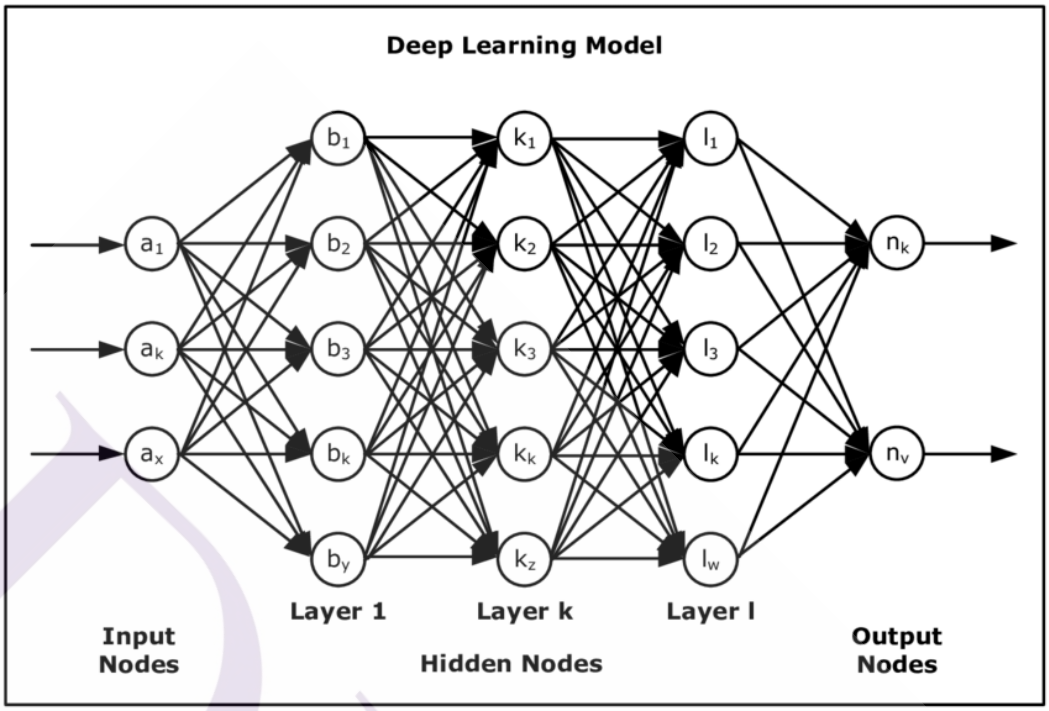


ภาพที่ 2.3 SVM Model

ที่มา: https://www.marktechpost.com/2021/03/25/introduction-to-support-vector-machines-svms/

2.4 เทคนิค Deep Learning

เทคนิค Deep Learning หรือการเรียนรู้เชิงลึก เป็นรูปแบบย่อยของ Machine Learning ที่มีลักษณะต่อยอดจากเทคนิคโครงข่ายประสาทเทียม (Artificial Neural Network: ANN) เป็นการจำลองการทำงานของสมองมนุษย์ โดยให้คอมพิวเตอร์รู้จักคิดและจดจำได้แบบเดียวกับโครงข่ายประสาทของมนุษย์ โดยจำลองว่า Input แต่ละตัวมีน้ำหนัก (Weight) อยู่ และเซลล์แต่ละหน่วยจะมีตัวกำหนด (Threshold) ว่าน้ำหนักรวมของ Input มากขนาดเพียงใด ถึงจะส่ง Output ต่อไปได้ การทำงานของ Deep Learning เป็นการใช้หลักการของ Artificial Neural Network มาใช้ในระดับที่ลึกและมีจำนวน Layer ที่ซับซ้อนกว่า



ภาพที่ 2.4 Deep Learning Model

ที่มา: https://www.researchgate.net/figure/Artificial-Neural Network-Deep-Learning-model_fig8_313408173

2.5 การวิเคราะห์การถดถอย (Regression Analysis)

[5] งานวิจัยที่มีวัตถุประสงค์ในการพยากรณ์หรือทำนายตัวแปรที่ต้องการศึกษา โดยมีตัวแปรตามที่เป็นตัวแปรต่อเนื่องเพียงหนึ่งตัว เรียกว่าการวิเคราะห์การถดถอย (Regression Analysis) โดยเรียกตัวแปรตามที่ต้องการศึกษาว่า ตัวแปรเกณฑ์ (Criterion Variable) และเรียกตัวแปรที่ใช้พยากรณ์ว่า ตัวแปรพยากรณ์ หรือตัวแปรทำนาย (Predictor Variable) กรณีใช้ตัวแปรพยากรณ์เพียงตัวเดียวจะเรียกว่า การวิเคราะห์การถดถอยอย่างง่าย (Simple Regression) ส่วนกรณีที่มีตัวแปรพยากรณ์มากกว่า 1 ตัว จะเรียกว่า การวิเคราะห์การถดถอยพหุคูณ (Multiple Regression)

จากข้อมูลที่ผู้วิจัยศึกษาเพื่อคัดกรองบุคคลที่มีแนวโน้มสัมฤทธิผลในการศึกษาระดับปริญญาตรี สาขาวิชาเทคโนโลยีการศึกษานั้น มีตัวแปรพยากรณ์ (Predictor Variable) จำนวน 29 ตัว หรือ 29 Attribute และตัวแปรเกณฑ์ (Criterion Variable) เพียง 1 ตัวหรือ Attribute ที่จะกำหนดเป็น Label คือ ผลการเรียนเฉลี่ยสะสมของผู้สำเร็จการศึกษา

[5] วิธีการวิเคราะห์การถดถอย มีดังนี้

1. วิธี Enter จะใช้ตัวแปรทำนายที่ศึกษาเข้าไปในสมการทำนายทุกตัว ถึงแม้ว่าตัวแปรทำนายบางตัวจะพยากรณ์ตัวแปรเกณฑ์ได้หรือไม่ก็ตาม วิธีนี้มักจะใช้ในกรณีที่ต้องการทราบว่าตัวแปรแต่ละตัวที่ทำการศึกษาจะสามารถทำนายตัวแปรเกณฑ์ได้หรือไม่ มากน้อยเพียงใด

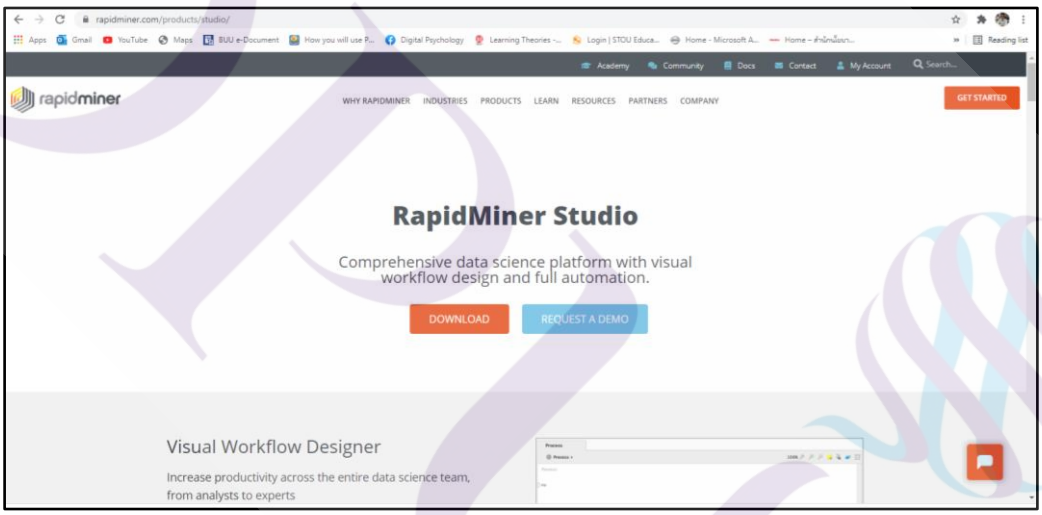
2. วิธี Forward เป็นวิธีการที่ต้องการได้โมเดลประหยัด คือจะคัดเลือกเฉพาะตัวแปรทำนายที่สามารถพยากรณ์ตัวแปรตามได้เท่านั้น โดยจะคัดเลือกตัวแปรพยากรณ์เข้ามาในสมการทีละตัว ขั้นแรกจะคัดเลือกตัวแปรทำนายที่มีค่าสัมประสิทธิ์สหสัมพันธ์กับตัวแปรเกณฑ์สูงสุดเข้ามาก่อน จากนั้นจึงคัดเลือกตัวแปรที่สำคัญรองลงมา โดยพิจารณาจากค่าสัมประสิทธิ์สหสมัมพันธ์บางส่วน (Partial Correlation) นั่นคือพิจารณาความสัมพันธ์ระหว่างตัวแปรทำนายที่เหลือแต่ละตัวกับตัวแปรเกณฑ์ โดยควบคุมอิทธิพลอันเนื่องมาจากตัวแปรทำนายที่เข้าสมการก่อนหน้า

3. วิธี Backward เป็นวิธีที่พยายามจะคัดเลือกตัวแปรที่ดีที่สุดในการทำนาย แต่เป็นวิธีที่ตรงข้ามกับวิธี Forward นั่นคือ ในตอนแรกจะนำตัวแปรทำนายทุกตัวเข้ามาในสมการ (เหมือนวิธี Enter) จากนั้นจะดำเนินการพิจารณาตัวแปรทำนายที่มีค่าสัมประสิทธิ์สหสมัมพันธ์บางส่วน (Partial Correlation) กับตัวแปรเกณฑ์ โดยควบคุมอิทธิพลของตัวแปรทำนายอื่น ๆ ซึ่งมีค่าต่ำสุดออกจากสมการ แล้วจึงดำเนินการทดสอบว่าค่าสัมประสิทธิ์การพยากรณ์ R^2 ลดลงอย่างมีนัยสำคัญหรือไม่ ถ้าพบว่าลดลงอย่างไม่มีนัยสำคัญแสดงว่าตัวแปรดังกล่าวไม่ได้มีส่วนทำให้การพยากรณ์ตัวแปรเกณฑ์เพิ่มขึ้นเลย จากนั้นจึงดำเนินการขจัดตัวแปรทำนายที่มีความสำคัญน้อยรองลงมาออกไปอีก ซึ่งการขจัดตัวแปรทำนายจะสิ้นสุดลง เมื่อพบว่ามีผลทำให้ค่า R^2 ลดลงอย่างมีนัยสำคัญทางสถิติ นั่นแสดงว่าตัวแปรดังกล่าวมีความสำคัญต่อการทำนายตัวแปรตาม

4. วิธี Stepwise เป็นวิธีที่มีความเหมาะสมในการพิจารณาคัดเลือกตัวแปรทำนายที่ดีที่สุด ซึ่งลำดับขั้นตอนจะคล้ายกับวิธี Forward เพียงแต่ว่าการวิเคราะห์ด้วย Stepwise นั้น จะทำการทดสอบตัวแปรทำนายที่เข้าสมการไปแล้วทุกครั้งที่มีการนำตัวแปรใหม่เข้าในสมการ นั่นหมายความว่า ตัวแปรบางตัวเข้าไปในสมการแล้ว ก็สามารถถูกขจัดออกจากสมการได้หากพบว่าไม่มีความสำคัญในการทำนาย ซึ่งวิธี Forward จะไม่ได้ทดสอบในส่วนนี้

2.6 โปรแกรม RapidMiner Studio

RapidMiner Studio เป็นโปรแกรมสำเร็จรูปที่ช่วยให้การทำเหมืองข้อมูลได้สะดวกมากขึ้น ทั้งนี้ข้อมูลที่นำมาวิเคราะห์จำเป็นจะต้องกรองข้อมูล (Data Cleansing) และตรวจสอบความสมบูรณ์ของข้อมูลให้เป็นที่เรียบร้อยแล้ว นับได้ว่าขั้นตอนการเตรียมข้อมูลก่อนการวิเคราะห์มีความสำคัญเป็นอย่างยิ่ง ถ้าไม่มีการกรองข้อมูล เมื่อนำเข้าโปรแกรม RapidMiner จะทำให้ตัวแบบที่ได้ (Model) มีความถูกต้องน้อย สามารถโหลดโปรแกรมได้ที่ https://rapidminer.com/ [25] RapidMiner เป็นการนำปัญญาประดิษฐ์มาใช้ในองค์กรผ่านแพลตฟอร์มวิทยาศาสตร์ข้อมูลแบบเปิดและขยายได้ สร้างขึ้นเพื่อการวิเคราะห์ข้อมูลโดยเฉพาะ RapidMiner รวมกระบวนการเกี่ยวกับการวิเคราะห์ข้อมูลตั้งแต่การเตรียมข้อมูลไปจนถึงการเรียนรู้ของเครื่อง (Machine Learning) ไปจนถึงการปรับใช้โมเดลทำนาย มีผู้เชี่ยวชาญด้านการวิเคราะห์ จำนวนมากกว่า 700,000 คนใช้ RapidMiner เพื่อเพิ่มรายได้ ลดต้นทุน และหลีกเลี่ยงความเสี่ยงต่าง ๆ



ภาพที่ 2.5 เว็บไซต์ RapidMiner Studio

RapidMiner Studio มีคุณสมบัติดังนี้ [25]

1. ทำงานแบบ Visual Workflow Designer รวมทั้งมีระบบ Wisdom of Crowds เป็นการให้คำแนะนำเชิงรุกในทุกขั้นตอนเพื่อช่วยเหลือผู้เริ่มต้นใช้งานโปรแกรมใหม่ ๆ

2. สามารถเชื่อมต่อกับทุกแหล่งข้อมูล (Connect to Any Data Source) โปรแกรมสามารถทำงานกับข้อมูลไม่ว่าจะอยู่ที่ใด สร้างการเชื่อมต่อแบบชี้และคลิกไปยังฐานข้อมูล คลังข้อมูลขององค์กร (Data Warehouses) ข้อมูลที่มีการจัดเก็บบนคลาวด์และโซเชียลมีเดียได้

3. สามารถประมวลผลแบบอัตโนมัติ (Automated In-Database Processing) สามารถเตรียมข้อมูลและประมวลผลได้ในฐานข้อมูล สร้างการสืบค้นและการดึงข้อมูลโดยไม่ต้องเขียน SQL ที่ซับซ้อน และรองรับ MySQL, PostgreSQL, and Google Big Query

4. มีความสามารถแสดงผลข้อมูลแบบ Data Visualization & Exploration

5. สามารถเตรียมข้อมูลได้อย่างง่าย (Data Prep & Blending)

6. ทำงานแบบ Visual & Automated Machine Learning คือสามารถสร้างโมเดล Machine Learning ที่มีประสิทธิภาพได้อย่างรวดเร็ว และมีกระบวนการตรวจสอบโมเดล (Model Validation) คือทำความเข้าใจประสิทธิภาพของโมเดลก่อนที่จะนำไปใช้งานจริง

7. ทำงานร่วมกับโปรแกรม R และ Python ได้

2.7 งานวิจัยที่เกี่ยวข้อง

[11] รัชฎา เทพประสิทธิ์ และจรัญ แสนราช ดำเนินการวิจัยเรื่อง การวิเคราะห์ปัจจัยที่มีผลต่อการเลือกสาขาวิชาของนักศึกษาระดับปริญญาตรี คณะครุศาสตร์ โดยใช้เทคนิคการทำเหมืองข้อมูล จากข้อมูลพื้นฐานนักศึกษาของงานส่งเสริมวิชาการและงานทะเบียน ในช่วงปีการศึกษา พ.ศ. 2556-2560 จำนวนนักศึกษา 3,867 คน มาวิเคราะห์ตัวแปรโดยใช้กระบวนการเหมืองข้อมูลด้วยกฎการจำแนกเทคนิคต้นไม้ตัดสินใจ ผลวิจัยพบว่า ปัจจัยที่มีผลต่อการเลือกสาขาวิชาคณะครุศาสตร์ 15 สาขาวิชา ได้แก่ แผนการเรียนก่อนเข้าศึกษา และเพศ จากตัวแปรทั้งหมด 9 ตัวแปร และการประเมินวัดประสิทธิภาพของโมเดลวัดค่าความแม่นยำ ได้ค่า 72.5 % ซึ่งถือว่าเป็นการทดสอบที่เชื่อถือได้

[15] สุวิมล สิทธิชาติ ดำเนินงานวิจัยนี้เพื่อศึกษาคุณลักษณะของนักศึกษาคณะวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา เชียงใหม่ ที่มีผลต่อการเรียนแคลคูลัส 1 สำหรับวิศวกรด้วยเทคนิคเหมืองข้อมูล เพื่อจำแนกนักศึกษาออกเป็น 2 กลุ่ม คือกลุ่มเสี่ยงที่จะไม่ผ่านรายวิชา และกลุ่มที่ไม่มีความเสี่ยง มีการเก็บข้อมูลคุณลักษณะที่มีผลต่อการเรียน ดังนี้ พฤติกรรมด้านความรู้และความคิด คุณลักษณะทางจิตใจ และคุณภาพทางการเรียนการสอนของนักศึกษา 453 คน แล้วดำเนินการด้วย (CRISP-DM: Cross Industry Standard Process for Data Mining) ใช้เทคนิคโครงข่ายประสาทเทียมและต้นไม้ตัดสินใจ ด้วย Filter Ranker Method คำนวนค่าน้ำหนักด้วย Chi-Square และ Gain Ratio จากการศึกษาพบว่า ความแม่นยำในการจำแนกข้อมูลจาก 50 คุณลักษณะนั้นวิธี ANN มีค่า 71.52% และ Decision Tree มีค่า 66.23% หลังจากทำการคัดเลือกคุณลักษณะแสดงให้เห็นว่าการจำแนกข้อมูลด้วยวิธี ANN จาก 5 คุณลักษณะแรกที่ได้จากการคัดเลือกโดยวิธี Filter Ranker Method ที่คำนวนค่าน้ำหนักด้วย Chi-Square ทำให้ได้ค่าความ

ถูกต้องมีค่าสูงสุดด้วย ANN เท่ากับ 80.13% และการจำแนกข้อมูลด้วย Decision Tree มีค่าความถูกต้องสูงสุดเท่ากับ 75.83%

[17] อนันต์ ปินะเต ทำการวิจัยเพื่อการใช้เทคนิคเหมืองข้อมูลในการเลือกกลุ่มสาขาวิชาที่เหมาะสมสำหรับการศึกษาต่อระดับปริญญาตรี หาแบบจำลองที่มีประสิทธิภาพสูงสุด มาพัฒนาเป็นระบบการพยากรณ์การทดสอบเลือกกลุ่มสาขาวิชา เพื่อใช้ในการประชาสัมพันธ์ให้ผู้สมัครได้ทดลองเลือกกลุ่มสาขาวิชาก่อนการเลือกสมัครในสาขาวิชาจริงเพื่อให้ผู้สมัครทราบถึงทักษะความรู้ของตนเองต่อกลุ่มสาขาวิชาที่จะเลือกสมัคร จากผลการทดลองพบว่า เทคนิคต้นไม้ตัดสินใจ ด้วยวิธี C4.5 มีประสิทธิภาพของแบบจำลองสูงสุด ซึ่งมีค่าความถูกต้อง โดยแยกตามกลุ่มข้อมูลทดลองดังนี้ กลุ่มมนุษยศาสตร์และสังคมศาสตร์ ร้อยละ 82.22 กลุ่มวิทยาศาสตร์เทคโนโลยี ร้อยละ 88.26 และกลุ่มวิทยาศาสตร์สุขภาพ ร้อยละ 87.73

[3] ณัฐริน เจริญเกียรติบวร ทำวิจัยเรื่องการใช้เทคนิคเหมืองข้อมูลเพื่อช่วยในการแนะแนวการศึกษาต่อระดับอุดมศึกษา การวิจัยครั้งนี้เป็นการนำผลการเรียนรายวิชาในระดับชั้นมัธยมศึกษาตอนปลายและเทคนิคการสุ่มตัวอย่าง โดยการเลือกตัวอย่างแบบมีระบบมาประยุกต์ใช้ในการทำเหมืองข้อมูล (Data Mining) เพื่อค้นหารูปแบบความสัมพันธ์หรือกฎที่ซ้อนอยู่และนำความสัมพันธ์เหล่านั้นมาแนะแนวการศึกษาต่อในระดับชั้นอุดมศึกษา การเลือกตัวอย่างใช้การแบ่งตามสัดส่วนข้อมูล Train : Validation 60:40, 65:35, 70:30 และ 75:25 ตามลำดับ ตัวแบบที่พัฒนาขึ้นวัดโดยใช้ค่าเฉลี่ยร้อยละความถูกต้องของตัวแบบที่สร้างขึ้นเมื่อใช้กับข้อมูล Validation จากการทำซ้ำ 10 ครั้ง พบว่า มีค่าเฉลี่ยความถูกต้องเป็น 82.32% , 81.87% , 91.78% และ 92.37% ตามลำดับ โดยมีค่าเฉลี่ยความถูกต้องรวมเป็น 82.08% เมื่อสัดส่วน Training set มากขึ้น

[9] พิสุทธ อ่อนเจริญ ได้ศึกษาการทำนายแนวโน้มของตลาดหุ้นด้วยตัวแบบการเรียนรู้เชิงลึกที่เพิ่มประสิทธิภาพร่วมกับข้อมูลเชิงตัวเลขและข้อมูลเชิงตัวอักษร ผู้วิจัยนำข้อมูลทั้งสองประเภทเพื่อทำนายแนวโน้มของตลาดหุ้น ซึ่งแบบจำลองนี้ประกอบไปด้วยนิวรอลเน็ตเวิร์กแบบคอนโวลูชัน (Convolutional Neural Network) และหน่วยความจำระยะสั้นแบบยาว (Long-Short Term Memory) โดยใช้ข้อมูลรับเข้าเป็นเหตุการณ์ฝังตัวซึ่งสกัดได้จากหัวข้อข่าว ราคาในอดีตและตัวชี้วัดทางเทคนิคซึ่งสร้างจากข้อมูลของราคาในอดีต รวมทั้งได้ทำการนำเสนอฟังก์ชันวัตถุประสงค์ชนิดใหม่ ที่สามารถช่วยเพิ่มประสิทธิภาพในด้านผลตอบแทนเฉลี่ยต่อปีที่ได้จากการจำลองการซื้อขายโดยการนำเอาค่าชาร์ปเรโชซึ่งเป็นตัวชี้วัดผลตอบแทนเมื่อเทียบกับความเสี่ยงมาใช้ร่วมกับค่าครอสเอนโทรปี

[18] อรทิพย์ เลื่อยงาม และชัยพร เขมะภาตะพันธ์ ทำวิจัยเรื่อง การจัดประเภทเอกสารด้วยวิธีเอสวีเอ็ม เพื่อการป้องกันเอกสารรั่วไหล เป็นการออกแบบระบบการแยกประเภทเอกสารสำคัญออกจากเอกสารทั่วไปเพื่อนำไปเป็นอินพุตของระบบการป้องกันข้อมูลรั่วไหล DLP พบว่าการแยกประเภทเอกสารโดยใช้ข้อมูลชุดฝึกสอน และข้อมูลชุดทดสอบทั้งสองประเภท คือเอกสารความลับ (1) และเอกสารทั่วไป (-1) สามารถแยกประเภทเอกสารตามที่ได้กำหนดประเภทเอกสารไว้ ซึ่งได้ผลเป็นที่น่าพอใจ

[12] วุฒิชัย กำจรกิตติคุณ และชัยพร เขมะภาตะพันธ์ ทำวิจัยเรื่อง การวิเคราะห์ความผิดปกติข้อมูลการใช้ไฟฟ้าด้วยนาอีฟเบย์ เป็นการคัดกรองความผิดปกติของข้อมูลการใช้ไฟฟ้ารายใหญ่ โดยนำสถิติการใช้ไฟฟ้ามาวิเคราะห์รูปแบบการใช้กำหลังไฟฟ้ารายเดือนควบคู่กับความสัมพันธ์ของพลังงานที่ใช้ในแต่ละเฟส พบว่า จำนวนครั้งในการตรวจสอบที่ไม่จำเป็นลดลงเมื่อเทียบกับการดำเนินการแบบเดิมที่ต้องตรวจสอบทุกกรณี และสร้างความน่าเชื่อถือให้กับระบบไฟฟ้าอีกด้วย

[23] Hanan ทำวิจัยเรื่อง การใช้เทคนิคการขุดข้อมูลเพื่อทำนายผลงานของนักศึกษาเพื่อสนับสนุนการตัดสินใจในระบบการรับเข้ามหาวิทยาลัย (Using Data Mining Techniques to Predict Student Performance to Support Decision Making in University Admission Systems) การศึกษานี้มุ่งเน้นไปที่วิธีการสนับสนุนมหาวิทยาลัยในการตัดสินใจรับสมัครโดยใช้เทคนิคการทำเหมืองข้อมูลเพื่อทำนายผลการเรียนของผู้สมัครในมหาวิทยาลัย ใช้ชุดข้อมูลของนักศึกษา 2,039 คนที่ลงทะเบียนเรียนในวิทยาลัยวิทยาศาสตร์คอมพิวเตอร์และสารสนเทศของมหาวิทยาลัยของรัฐในซาอุดิอาระเบียตั้งแต่ปี 2559 ถึง 2562 ผลการศึกษาแสดงให้เห็นว่าผู้สมัครสามารถคาดการณ์ผลการเรียนในระดับมหาวิทยาลัยก่อนเข้าเรียนได้โดยพิจารณาจากเกณฑ์ก่อนการรับเข้าเรียน (เกรดเฉลี่ยมัธยมปลายคะแนนการทดสอบการรับเข้าเรียนของนักวิชาการและคะแนนการทดสอบความถนัดทั่วไป) ผลการวิจัยยังแสดงให้เห็นว่าคะแนนการทดสอบผลสัมฤทธิ์ทางการเรียนของนักวิชาการเป็นเกณฑ์ก่อนการรับเข้าเรียนที่ทำนายผลการเรียนของนักเรียนในอนาคตได้แม่นยำที่สุด ดังนั้นควรกำหนดคะแนนนี้ให้มีน้ำหนักมากขึ้นในระบบการรับสมัคร นอกจากนี้ยังพบว่าเทคนิค Artificial Neural Network มีความแม่นยำ 79.22% เทคนิค Decision Trees มีความแม่นยำ 75.91%เทคนิค Support Vector Machines มีความแม่นยำ 75.28% และเทคนิค Naive Bayes มีความแม่นยำ 73.61%

[22] Ferda ทำวิจัยเกี่ยวกับการทำเหมืองข้อมูลสำหรับการทำนายผลการปฏิบัติงานของผู้เรียนในการศึกษา (Data Mining for Student Performance Prediction in Education) โดยใช้เทคนิคการทำเหมืองข้อมูล ดังนี้ 1) Decision Tree 2) Random Forest และ 3) Naive Bayes ทำนาย

แนวโน้มการปฏิบัติงานของนักเรียนเป็นสิ่งสำคัญมากในการปรับปรุงทักษะการสอนของพวกเขา โดยนำเทคนิคเหมืองข้อมูลมาใช้กับชุดข้อมูล 2 ชุด คือ บทเรียนคณิตศาสตร์และบทเรียนภาษาโปรตุเกส จำนวน 33 Attributes ผลการวิจัยพบว่า เทคนิค Decision Tree ได้ค่า Accuracy = 73.42% เทคนิค Random Forest ได้ค่า Accuracy = 71.14% และเทคนิค Naive Bayes ได้ค่า Accuracy = 70.38%

[24] Khasanah & Harwati ทำวิจัยเรื่อง การศึกษาเปรียบเทียบเพื่อทำนายผลการเรียนของนักศึกษาโดยใช้เทคนิคการทำเหมืองข้อมูลทางการศึกษา (A Comparative Study to Predict Student's Performance Using Educational Data Mining Techniques) เปรียบเทียบเทคนิค Bayesian Network กับเทคนิค Decision Tree จำนวน 12 Attributes ศึกษากับนักศึกษาในมหาวิทยาลัยจำนวน 178 คน โดยแบ่งเป็น 3 Scenario ดังนี้ 1) วิเคราะห์ทุก Attribute 2) วิเคราะห์ 9 Attribute และ 3) วิเคราะห์ 8 Attribute ผลการวิจัยพบว่า 1) การวิเคราะห์ 8 Attribute ด้วยเทคนิค Bayesian Network มีค่า Accuracy = 98.08% สูงที่สุด เมื่อเปรียบเทียบกับการวิเคราะห์ทุก Attribute เทคนิค Bayesian Network มีค่า Accuracy = 95.19% 2) การเปรียบเทียบระหว่างเทคนิค Bayesian Network กับเทคนิค Decision Tree เทคนิค Bayesian Network ได้ผลที่สูงกว่า

บทที่ 3
วิธีดำเนินการวิจัย

งานวิจัยนี้ได้นำเทคนิคเหมืองข้อมูลมาใช้เพื่อการคัดกรองบุคคลที่มีแนวโน้มสัมฤทธิผลในการศึกษาระดับปริญญาตรี สาขาวิชาเทคโนโลยีการศึกษา โดยมีขั้นตอนดำเนินการวิจัยดังนี้

3.1 การเลือกข้อมูล (Data Selection)

ข้อมูลที่นำมาศึกษาเพื่อการวิเคราะห์หาคุณลักษณะเพื่อการคัดกรองบุคคลที่มีแนวโน้มสัมฤทธิผลในการศึกษาระดับปริญญาตรี สาขาวิชาเทคโนโลยีการศึกษา ได้จากระบบของกองทะเบียนและประมวลผล ระหว่างปีการศึกษา 2554-2558 จำนวน 738 คน คัดเลือกคุณลักษณะ (Attribute) สำหรับการศึกษาได้จำนวน 30 Attribute ดังนี้

1. รหัสนิสิต
2. ระดับคะแนนเฉลี่ยสะสมของนิสิตสาขาวิชาเทคโนโลยีการศึกษา ระดับปริญญาตรี
3. กลุ่มเลือด A ของผู้เรียนในสาขาวิชาเทคโนโลยีการศึกษา ระดับปริญญาตรี
4. กลุ่มเลือด B ของผู้เรียนในสาขาวิชาเทคโนโลยีการศึกษา ระดับปริญญาตรี
5. กลุ่มเลือด AB ของผู้เรียนในสาขาวิชาเทคโนโลยีการศึกษา ระดับปริญญาตรี
6. กลุ่มเลือด O ของผู้เรียนในสาขาวิชาเทคโนโลยีการศึกษา ระดับปริญญาตรี
7. ภูมิลำเนา: กลุ่มจังหวัดภาคเหนือ ของผู้เรียนในสาขาวิชาเทคโนโลยีการศึกษา ระดับปริญญาตรี
8. ภูมิลำเนา: กลุ่มจังหวัดภาคตะวันออกเฉียงเหนือ ของผู้เรียนในสาขาวิชาเทคโนโลยีการศึกษา ระดับปริญญาตรี
9. ภูมิลำเนา: กลุ่มจังหวัดภาคตะวันออก ของผู้เรียนในสาขาวิชาเทคโนโลยีการศึกษา ระดับปริญญาตรี
10. ภูมิลำเนา: กลุ่มจังหวัดภาคใต้ ของผู้เรียนในสาขาวิชาเทคโนโลยีการศึกษา ระดับปริญญาตรี

11. ภูมิลำเนา: กลุ่มจังหวัดภาคตะวันตก ของผู้เรียนในสาขาวิชาเทคโนโลยีการศึกษา ระดับปริญญาตรี

12. ภูมิลำเนา: กลุ่มจังหวัดภาคกลาง ของผู้เรียนในสาขาวิชาเทคโนโลยีการศึกษา ระดับปริญญาตรี

13. เกรดเฉลี่ยสะสมมัธยมศึกษาตอนปลาย ของผู้เรียนในสาขาวิชาเทคโนโลยีการศึกษา ระดับปริญญาตรี

14. วิธีการรับเข้าของนิสิต: รับตรงรอบ 1 (โครงการพิเศษ)

15. วิธีการรับเข้าของนิสิต: รับตรงรอบ 2 (12 จังหวัด)

16. วิธีการรับเข้าของนิสิต: รับตรงรอบ 3 (ทั่วประเทศ)

17. วิธีการรับเข้าของนิสิต: Admission กลาง ทปอ.

18. วิธีการรับเข้าของนิสิต: รับตรงรอบ 5 (รอบพิเศษ)

19. รายได้โดยประมาณต่อปีของบิดา

20. รายได้โดยประมาณต่อปีของมารดา

21. บิดาไม่มีอาชีพ

22. อาชีพของบิดา: รับราชการ

23. อาชีพของบิดา: ค้าขาย

24. อาชีพของบิดา: เกษตรกร

25. อาชีพของบิดา: ธุรกิจส่วนตัว

26. มารดาไม่มีอาชีพ

27. อาชีพของมารดา: รับราชการ

28. อาชีพของมารดา: ค้าขาย

29. อาชีพของมารดา: เกษตรกร

30. อาชีพของมารดา: ธุรกิจส่วนตัว

คุณลักษณะ (Attribute) ดังกล่าวข้างต้น ผู้วิจัยดำเนินการจัดประชุมระดมความคิดเห็นของคณะกรรมการบริหารหลักสูตร สาขาวิชาเทคโนโลยีการศึกษา จำนวน 5 คน เพื่อคัดเลือก Attribute ที่จะกำหนดเป็น Label สำหรับการวิเคราะห์ข้อมูล ค้นหาตัวแบบคัดกรองบุคคลที่มีแนวโน้มสัมฤทธิผลในการศึกษาระดับปริญญาตรี สาขาวิชาเทคโนโลยีการศึกษา จากการประชุมคณะกรรมการบริหารหลักสูตร ได้กำหนด ผลการเรียนเฉลี่ยสะสมของผู้สำเร็จการศึกษาเป็น Label เป็นตัวชี้วัดของผลสำเร็จการเรียนรู้ของนิสิตสาขาวิชาเทคโนโลยีการศึกษา โดยกำหนดช่วงของคะแนนผลการเรียนเฉลี่ยสะสม เป็น 2 ระดับ ดังนี้

1. ผลการเรียนเฉลี่ยสะสมระหว่าง 3.00 – 4.00 กำหนดเป็นสัมฤทธิผลดี แทนค่าด้วย Good

2. ผลการเรียนเฉลี่ยสะสมระหว่าง 2.00 – 2.99 กำหนดเป็นสัมฤทธิผลไม่ดี แทนค่าด้วย Bad

การคัดเลือกคัดเลือกตัวแปรหรือคุณลักษณะ (Attribute)

จำนวนคุณลักษณะ (Attribute) ทั้งสิ้น 30 Attribute ผู้วิจัยได้ใช้วิธีการคัดเลือก Attribute โดยการวิเคราะห์การถดถอย (Regression) แบบการถดถอยพหุคูณ (Multiple Regression) ผู้วิจัยเลือกใช้วิธีการคัดเลือกตัวแปรหรือ Attribute 2 แบบคือ แบบถอยหลัง (Backward Selection) และแบบลำดับขั้น (Stepwise Selection)

1. การเลือกแบบถอยหลัง (Backward Selection) วิธีการนี้เป็นการนำตัวแปรทำนายทั้งหมดเข้าสมการ จากนั้นค่อย ๆ ขจัดตัวแปรทำนายออกทีละตัว เพื่อให้ได้สมการที่มีสัมประสิทธิ์การทำนายสูงสุด

2. การเลือกแบบลำดับขั้น (Stepwise Selection) วิธีการนี้เป็นการผสมผสานระหว่างวิธีการคัดเลือกตัวแปรแบบก้าวหน้า (Forward Selection) และวิธีการคัดเลือกแบบถอยหลัง (Backward Selection) ในขั้นแรกจะเลือกตัวแปรทำนายที่มีสหสัมพันธ์กับตัวแปรตามสูงที่สุดเข้าสมการก่อน จากนั้นก็จะทดสอบตัวแปรที่ไม่ได้อยู่ในสมการว่าจะมีตัวทำนายตัวใดบ้างมีสิทธิ์เข้ามาอยู่ในสมการด้วย Forward Selection ขณะเดียวกันก็จะทดสอบตัวแปรที่อยู่ในสมการด้วยว่าตัวแปรทำนายที่อยู่ในสมการตัวแปรใดมีโอกาสที่จะถูกขจัดออกจากสมการด้วย Backward Selection ทำจนกระทั่งไม่มีตัวแปรใดที่ถูกคัดออกจากสมการ และไม่มีตัวแปรใดที่จะถูกนำเข้าสมการ กระบวนการจะยุติและได้สมการถดถอยที่มีประสิทธิภาพการทำนายสูงสุด

ผู้วิจัยสร้างตารางตัวแปรและดำเนินการใช้โปรแกรม IBM SPSS Statistics (ผู้วิจัยใช้ Free Trial Version) ดังนี้

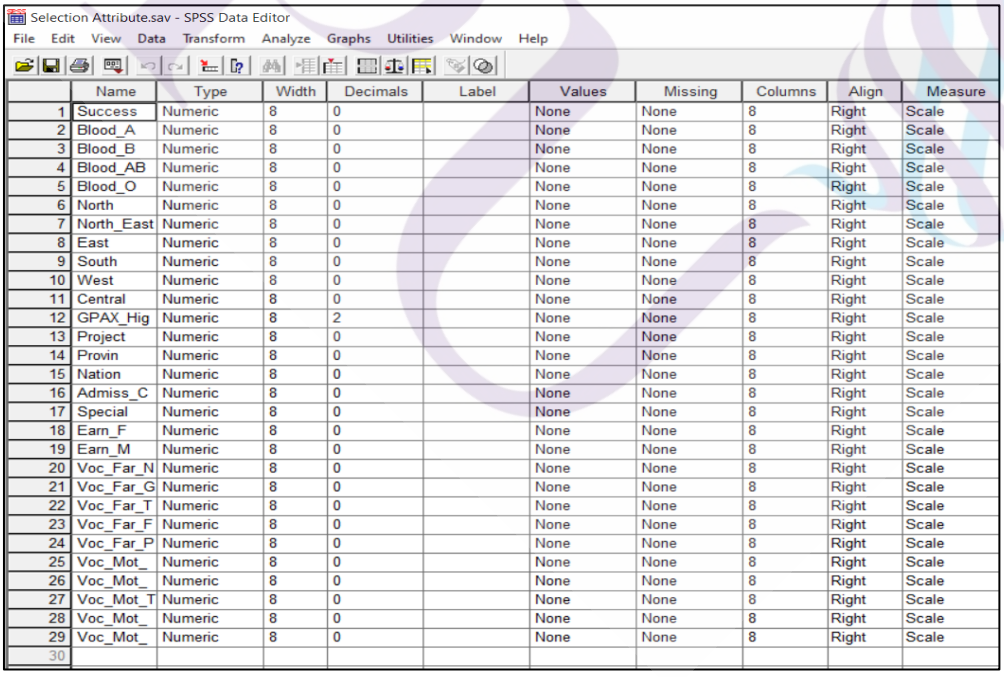
ตารางที่ 3.1 การกำหนดตัวแปร

Attribute	ความหมาย
ID	รหัสนิสิต
Success	ระดับสัมฤทธิผลสำหรับการเรียนของนิสิตสาขาวิชาเทคโนโลยีการศึกษา
Blood_A	กลุ่มเลือด A ของผู้เรียนในสาขาวิชาเทคโนโลยีการศึกษา
Blood_B	กลุ่มเลือด B ของผู้เรียนในสาขาวิชาเทคโนโลยีการศึกษา
Blood_AB	กลุ่มเลือด AB ของผู้เรียนในสาขาวิชาเทคโนโลยีการศึกษา
Blood_O	กลุ่มเลือด O ของผู้เรียนในสาขาวิชาเทคโนโลยีการศึกษา
North	ภูมิลำเนา: กลุ่มจังหวัดภาคเหนือ ของผู้เรียนในสาขาวิชาเทคโนโลยีการศึกษา
North_East	ภูมิลำเนา: กลุ่มจังหวัดภาคตะวันออกเฉียงเหนือ ของผู้เรียนในสาขาวิชาเทคโนโลยีการศึกษา
East	ภูมิลำเนา: กลุ่มจังหวัดภาคตะวันออก ของผู้เรียนในสาขาวิชาเทคโนโลยีการศึกษา
South	ภูมิลำเนา: กลุ่มจังหวัดภาคใต้ ของผู้เรียนในสาขาวิชาเทคโนโลยีการศึกษา
West	ภูมิลำเนา: กลุ่มจังหวัดภาคตะวันตก ของผู้เรียนในสาขาวิชาเทคโนโลยีการศึกษา
Central	ภูมิลำเนา: กลุ่มจังหวัดภาคกลาง ของผู้เรียนในสาขาวิชาเทคโนโลยีการศึกษา
GPAX_High	เกรดเฉลี่ยสะสมมัธยมศึกษาตอนปลาย ของผู้เรียนในสาขาวิชาเทคโนโลยีการศึกษา
Project	วิธีการรับเข้าของนิสิต: รับตรงรอบ 1 (โครงการพิเศษ)
Provin	วิธีการรับเข้าของนิสิต: รับตรงรอบ 2 (12 จังหวัด)
Nation	วิธีการรับเข้าของนิสิต: รับตรงรอบ 3 (ทั่วประเทศ)
Admiss_C	วิธีการรับเข้าของนิสิต: Admission กลาง ทปอ.
Special	วิธีการรับเข้าของนิสิต: รับตรงรอบ 5 (รอบพิเศษ)
Earn_F	รายได้โดยประมาณต่อปีของบิดา
Earn_M	รายได้โดยประมาณต่อปีของมารดา

ตารางที่ 3.1 การกำหนดตัวแปร (ต่อ)

Attribute	ความหมาย
Voc_Far_No	บิดาไม่มีอาชีพ
Voc_Far_Gov	อาชีพของบิดา: รับราชการ
Voc_Far_Trade	อาชีพของบิดา: ค้าขาย
Voc_Far_Farm	อาชีพของบิดา: เกษตรกร
Voc_Far_Person	อาชีพของบิดา: ธุรกิจส่วนตัว
Voc_Mot_No	มารดาไม่มีอาชีพ
Voc_Mot_Gov	อาชีพของมารดา: รับราชการ
Voc_Mot_Trade	อาชีพของมารดา: ค้าขาย
Voc_Mot_Farm	อาชีพของมารดา: เกษตรกร
Voc_Mot_Person	อาชีพของมารดา: ธุรกิจส่วนตัว

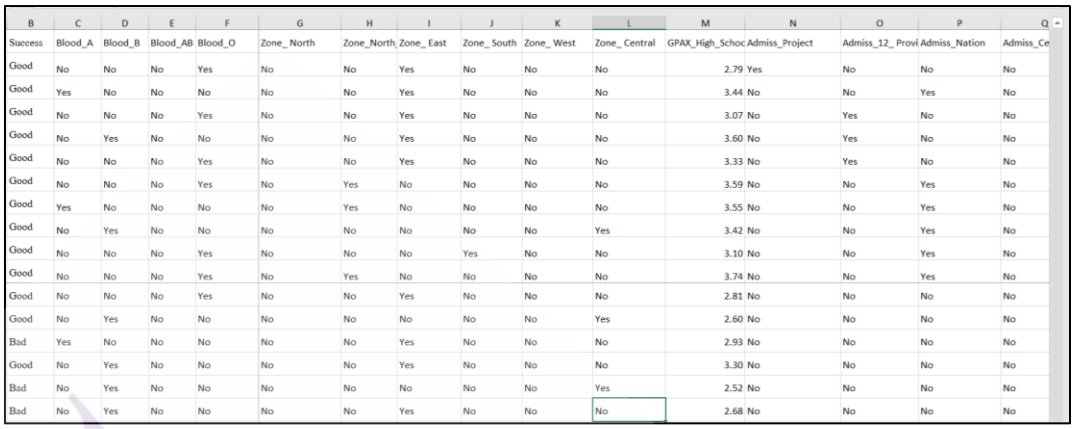
ผู้วิจัยกำหนดตัวแปรในโปรแกรม SPSS ดังภาพที่ 3.1


Selection Attribute.sav - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	Success	Numeric	8	0		None	None	8	Right	Scale
2	Blood_A	Numeric	8	0		None	None	8	Right	Scale
3	Blood_B	Numeric	8	0		None	None	8	Right	Scale
4	Blood_AB	Numeric	8	0		None	None	8	Right	Scale
5	Blood_O	Numeric	8	0		None	None	8	Right	Scale
6	North	Numeric	8	0		None	None	8	Right	Scale
7	North_East	Numeric	8	0		None	None	8	Right	Scale
8	East	Numeric	8	0		None	None	8	Right	Scale
9	South	Numeric	8	0		None	None	8	Right	Scale
10	West	Numeric	8	0		None	None	8	Right	Scale
11	Central	Numeric	8	0		None	None	8	Right	Scale
12	GPAX_Hig	Numeric	8	2		None	None	8	Right	Scale
13	Project	Numeric	8	0		None	None	8	Right	Scale
14	Provin	Numeric	8	0		None	None	8	Right	Scale
15	Nation	Numeric	8	0		None	None	8	Right	Scale
16	Admiss_C	Numeric	8	0		None	None	8	Right	Scale
17	Special	Numeric	8	0		None	None	8	Right	Scale
18	Earn_F	Numeric	8	0		None	None	8	Right	Scale
19	Earn_M	Numeric	8	0		None	None	8	Right	Scale
20	Voc_Far_N	Numeric	8	0		None	None	8	Right	Scale
21	Voc_Far_G	Numeric	8	0		None	None	8	Right	Scale
22	Voc_Far_T	Numeric	8	0		None	None	8	Right	Scale
23	Voc_Far_F	Numeric	8	0		None	None	8	Right	Scale
24	Voc_Far_P	Numeric	8	0		None	None	8	Right	Scale
25	Voc_Mot_	Numeric	8	0		None	None	8	Right	Scale
26	Voc_Mot_	Numeric	8	0		None	None	8	Right	Scale
27	Voc_Mot_T	Numeric	8	0		None	None	8	Right	Scale
28	Voc_Mot_	Numeric	8	0		None	None	8	Right	Scale
29	Voc_Mot_	Numeric	8	0		None	None	8	Right	Scale
30										

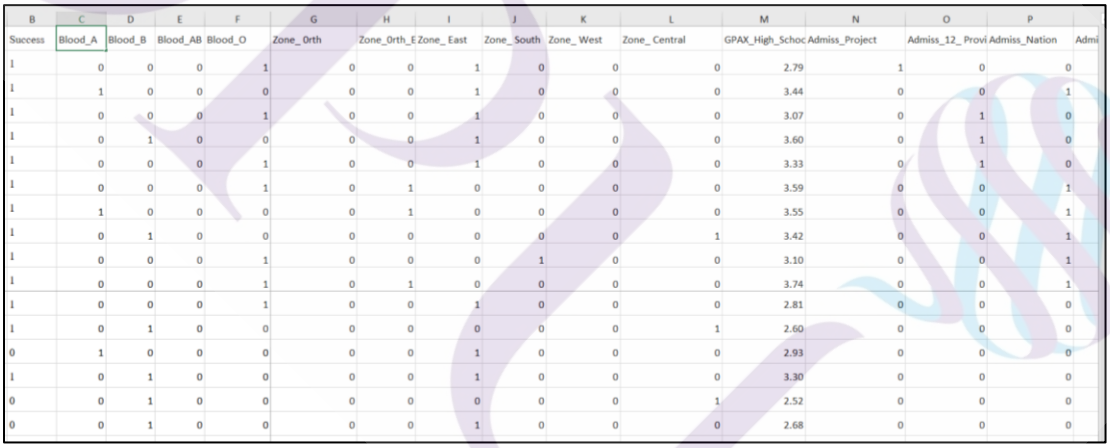
ภาพที่ 3.1 การกำหนดตัวแปรในโปรแกรม SPSS



B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
Success	Blood_A	Blood_B	Blood_AB	Blood_O	Zone_ North	Zone_North	Zone_ East	Zone_ South	Zone_ West	Zone_ Central	GPAX_High_Schoc	Admiss_Project	Admiss_12_ Provi	Admiss_Nation	Admiss_Ce
Good	No	No	No	Yes	No	No	Yes	No	No	No	2.79	Yes	No	No	No
Good	Yes	No	No	No	No	No	Yes	No	No	No	3.44	No	No	Yes	No
Good	No	No	No	Yes	No	No	Yes	No	No	No	3.07	No	Yes	No	No
Good	No	Yes	No	No	No	No	Yes	No	No	No	3.60	No	Yes	No	No
Good	No	No	No	Yes	No	No	Yes	No	No	No	3.33	No	Yes	No	No
Good	No	No	No	Yes	No	Yes	No	No	No	No	3.59	No	No	Yes	No
Good	Yes	No	No	No	No	Yes	No	No	No	No	3.55	No	No	Yes	No
Good	No	Yes	No	No	No	No	No	No	No	Yes	3.42	No	No	Yes	No
Good	No	No	No	Yes	No	No	No	Yes	No	No	3.10	No	No	Yes	No
Good	No	No	No	Yes	No	Yes	No	No	No	No	3.74	No	No	Yes	No
Good	No	No	No	Yes	No	No	Yes	No	No	No	2.81	No	No	No	No
Good	No	Yes	No	No	No	No	No	No	No	Yes	2.60	No	No	No	No
Bad	Yes	No	No	No	No	No	Yes	No	No	No	2.93	No	No	No	No
Good	No	Yes	No	No	No	No	Yes	No	No	No	3.30	No	No	No	No
Bad	No	Yes	No	No	No	No	No	No	No	Yes	2.52	No	No	No	No
Bad	No	Yes	No	No	No	No	Yes	No	No	No	2.68	No	No	No	No

ภาพที่ 3.2 ข้อมูลในไฟล์ Excel เป็นลักษณะรูปแบบตัวอักษร

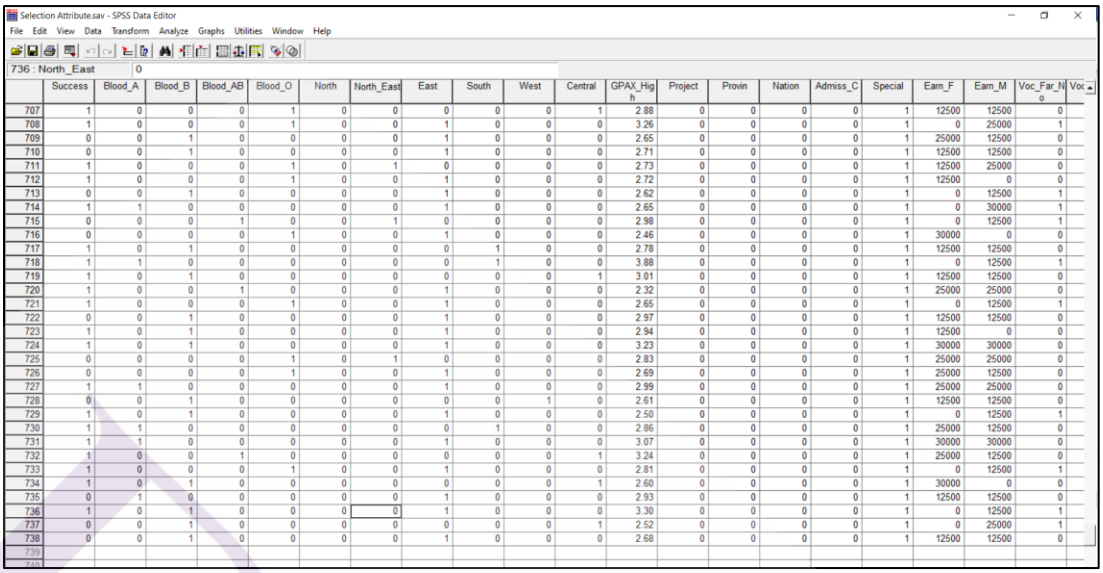
ผู้วิจัยแปลงข้อมูลที่เป็นตัวอักษรในไฟล์ Excel ให้อยู่ในรูปแบบของตัวเลข นั่นคือ ค่า (Values) ของ ตัวแปรหรือ Attribute ที่เป็น Yes แทนค่าเป็น 1 และที่เป็น No แทนค่าเป็น 0 ดังภาพที่ 3.3 สำหรับโปรแกรม Microsoft Excel เป็น Microsoft Campus License



B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	
Success	Blood_A	Blood_B	Blood_AB	Blood_O	Zone_ 0rth	Zone_0rth_E	Zone_ East	Zone_ South	Zone_ West	Zone_ Central	GPAX_High_Schoc	Admiss_Project	Admiss_12_ Provi	Admiss_Nation	Admi
1	0	0	0	1	0	0	1	0	0	0	2.79	1	0	0	
1	1	0	0	0	0	0	1	0	0	0	3.44	0	0	1	
1	0	0	0	1	0	0	1	0	0	0	3.07	0	1	0	
1	0	1	0	0	0	0	1	0	0	0	3.60	0	1	0	
1	0	0	0	1	0	0	1	0	0	0	3.33	0	1	0	
1	0	0	0	1	0	1	0	0	0	0	3.59	0	0	1	
1	1	0	0	0	0	1	0	0	0	0	3.55	0	0	1	
1	0	1	0	0	0	0	0	0	0	1	3.42	0	0	1	
1	0	0	0	1	0	0	0	1	0	0	3.10	0	0	1	
1	0	0	0	1	0	1	0	0	0	0	3.74	0	0	1	
1	0	0	0	1	0	0	1	0	0	0	2.81	0	0	0	
1	0	1	0	0	0	0	0	0	0	1	2.60	0	0	0	
0	1	0	0	0	0	0	1	0	0	0	2.93	0	0	0	
1	0	1	0	0	0	0	1	0	0	0	3.30	0	0	0	
0	0	1	0	0	0	0	0	0	0	1	2.52	0	0	0	
0	0	1	0	0	0	0	1	0	0	0	2.68	0	0	0	

ภาพที่ 3.3 ข้อมูลในไฟล์ Excel แทนค่าด้วยรูปแบบตัวเลข

ผู้วิจัยนำข้อมูลจากไฟล์ Excel มาวางในส่วนของ Data View ดังภาพที่ 3.4 จากนั้นผู้วิจัยดำเนินการวิเคราะห์ข้อมูล เพื่อคัดเลือกตัวแปรที่จะสร้างแบบจำลองต่อไป โดยเลือกคำสั่ง Analyze-Regression-Linear เลือกตัวแปร Success เข้าในช่อง Dependent ตัวแปรที่เหลือทั้งหมดเลือกเข้าช่อง Independent เลือก Method เป็น Backward ได้ผลดังตารางที่ 3.2 และแบบ Stepwise ได้ผลดังตารางที่ 3.3



ภาพที่ 3.4 ข้อมูลจากไฟล์ Excel มุมมองแบบ Data View

ตารางที่ 3.2 Model Summary ของโปรแกรม SPSS วิธีการ Backward

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.408(a)	.167	.136	.354	.167	5.475	26	711	.000
2	.408(b)	.167	.138	.354	.000	.000	1	711	.996
3	.408(c)	.167	.139	.354	.000	.012	1	712	.914
4	.408(d)	.167	.140	.354	.000	.027	1	713	.869
5	.408(e)	.167	.141	.353	.000	.064	1	714	.801
6	.408(f)	.166	.142	.353	.000	.239	1	715	.625
7	.408(g)	.166	.143	.353	.000	.265	1	716	.607
8	.407(h)	.166	.144	.353	.000	.146	1	717	.703
9	.407(i)	.166	.145	.353	.000	.252	1	718	.616
10	.407(j)	.165	.146	.352	.000	.296	1	719	.587
11	.406(k)	.165	.146	.352	.000	.301	1	720	.583
12	.406(l)	.165	.147	.352	.000	.216	1	721	.642

ตารางที่ 3.2 Model Summary ของโปรแกรม SPSS วิธีการ Backward (ต่อ)

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
13	.405(m)	.164	.148	.352	-.001	.849	1	722	.357
14	.405(n)	.164	.149	.352	.000	.068	1	723	.794
15	.404(o)	.163	.150	.352	.000	.167	1	724	.683
16	.403(p)	.163	.150	.351	-.001	.620	1	725	.431
17	.403(q)	.162	.150	.351	-.001	.614	1	726	.434
18	.400(r)	.160	.150	.352	-.002	1.738	1	727	.188
19	.398(s)	.158	.149	.352	-.002	1.505	1	728	.220
20	.395(t)	.156	.148	.352	-.002	1.900	1	729	.168
21	.392(u)	.154	.147	.352	-.002	2.070	1	730	.151
22	.389(v)	.151	.145	.352	-.003	2.225	1	731	.136

a Predictors: (Constant), Voc_Mot_Person, Blood_O, Project, Earn_F, North, South, West, North_East, Voc_Far_Farm, GPAX_High, Admiss_C, Earn_M, Blood_AB, Central, Provin, Voc_Far_Trade, Nation, Blood_A, Voc_Far_Person, Voc_Mot_Gov, Voc_Mot_Farm, Voc_Far_No, Voc_Mot_No, East, Voc_Mot_Trade, Voc_Far_Gov

b Predictors: (Constant), Voc_Mot_Person, Blood_O, Project, Earn_F, North, South, West, North_East, Voc_Far_Farm, GPAX_High, Admiss_C, Earn_M, Blood_AB, Provin, Voc_Far_Trade, Nation, Blood_A, Voc_Far_Person, Voc_Mot_Gov, Voc_Mot_Farm, Voc_Far_No, Voc_Mot_No, East, Voc_Mot_Trade, Voc_Far_Gov

c Predictors: (Constant), Voc_Mot_Person, Blood_O, Project, Earn_F, North, South, West, North_East, GPAX_High, Admiss_C, Earn_M, Blood_AB, Provin, Voc_Far_Trade, Nation, Blood_A, Voc_Far_Person, Voc_Mot_Gov, Voc_Mot_Farm, Voc_Far_No, Voc_Mot_No, East, Voc_Mot_Trade, Voc_Far_Gov

d Predictors: (Constant), Voc_Mot_Person, Project, Earn_F, North, South, West, North_East, GPAX_High, Admiss_C, Earn_M, Blood_AB, Provin, Voc_Far_Trade, Nation, Blood_A, Voc_Far_Person, Voc_Mot_Gov, Voc_Mot_Farm, Voc_Far_No, Voc_Mot_No, East, Voc_Mot_Trade, Voc_Far_Gov

e Predictors: (Constant), Voc_Mot_Person, Project, Earn_F, North, South, West, North_East, GPAX_High,

Admiss_C, Earn_M, Blood_AB, Provin, Voc_Far_Trade, Nation, Voc_Far_Person, Voc_Mot_Gov, Voc_Mot_Farm, Voc_Far_No, Voc_Mot_No, East, Voc_Mot_Trade, Voc_Far_Gov
f Predictors: (Constant), Voc_Mot_Person, Project, Earn_F, South, West, North_East, GPAX_High, Admiss_C, Earn_M, Blood_AB, Provin, Voc_Far_Trade, Nation, Voc_Far_Person, Voc_Mot_Gov, Voc_Mot_Farm, Voc_Far_No, Voc_Mot_No, East, Voc_Mot_Trade, Voc_Far_Gov
g Predictors: (Constant), Voc_Mot_Person, Project, Earn_F, South, West, North_East, GPAX_High, Admiss_C, Earn_M, Blood_AB, Provin, Voc_Far_Trade, Nation, Voc_Mot_Gov, Voc_Mot_Farm, Voc_Far_No, Voc_Mot_No, East, Voc_Mot_Trade, Voc_Far_Gov
h Predictors: (Constant), Voc_Mot_Person, Project, Earn_F, South, West, North_East, GPAX_High, Admiss_C, Earn_M, Blood_AB, Provin, Voc_Far_Trade, Nation, Voc_Mot_Gov, Voc_Mot_Farm, Voc_Mot_No, East, Voc_Mot_Trade, Voc_Far_Gov
i Predictors: (Constant), Voc_Mot_Person, Project, Earn_F, South, West, North_East, GPAX_High, Admiss_C, Earn_M, Blood_AB, Provin, Nation, Voc_Mot_Gov, Voc_Mot_Farm, Voc_Mot_No, East, Voc_Mot_Trade, Voc_Far_Gov
j Predictors: (Constant), Voc_Mot_Person, Project, Earn_F, South, North_East, GPAX_High, Admiss_C, Earn_M, Blood_AB, Provin, Nation, Voc_Mot_Gov, Voc_Mot_Farm, Voc_Mot_No, East, Voc_Mot_Trade, Voc_Far_Gov
k Predictors: (Constant), Voc_Mot_Person, Project, Earn_F, South, North_East, GPAX_High, Admiss_C, Earn_M, Blood_AB, Provin, Nation, Voc_Mot_Gov, Voc_Mot_Farm, Voc_Mot_No, Voc_Mot_Trade, Voc_Far_Gov
l Predictors: (Constant), Voc_Mot_Person, Project, Earn_F, South, GPAX_High, Admiss_C, Earn_M, Blood_AB, Provin, Nation, Voc_Mot_Gov, Voc_Mot_Farm, Voc_Mot_No, Voc_Mot_Trade, Voc_Far_Gov
m Predictors: (Constant), Voc_Mot_Person, Project, Earn_F, South, GPAX_High, Admiss_C, Earn_M, Blood_AB, Provin, Nation, Voc_Mot_Gov, Voc_Mot_Farm, Voc_Mot_No, Voc_Far_Gov
n Predictors: (Constant), Voc_Mot_Person, Project, Earn_F, South, GPAX_High, Admiss_C, Earn_M, Blood_AB, Provin, Nation, Voc_Mot_Farm, Voc_Mot_No, Voc_Far_Gov
o Predictors: (Constant), Voc_Mot_Person, Project, Earn_F, South, GPAX_High, Admiss_C, Earn_M, Blood_AB, Provin, Nation, Voc_Mot_Farm, Voc_Far_Gov
p Predictors: (Constant), Voc_Mot_Person, Project, Earn_F, South, GPAX_High, Admiss_C, Earn_M, Blood_AB, Provin, Nation, Voc_Far_Gov
q Predictors: (Constant), Project, Earn_F, South, GPAX_High, Admiss_C, Earn_M, Blood_AB, Provin, Nation, Voc_Far_Gov
r Predictors: (Constant), Project, Earn_F, South, GPAX_High, Admiss_C, Earn_M, Provin, Nation,

Voc_Far_Gov
s Predictors: (Constant), Earn_F, South, GPAX_High, Admiss_C, Earn_M, Provin, Nation, Voc_Far_Gov
t Predictors: (Constant), Earn_F, GPAX_High, Admiss_C, Earn_M, Provin, Nation, Voc_Far_Gov
u Predictors: (Constant), Earn_F, GPAX_High, Admiss_C, Earn_M, Provin, Voc_Far_Gov
v Predictors: (Constant), Earn_F, GPAX_High, Admiss_C, Earn_M, Voc_Far_Gov

จากตารางที่ 3.2 Model Summary ของโปรแกรม SPSS วิธีการ Backward พบว่าสามารถสร้างสมการทำนายได้ถึง 22 รูปแบบ สมการทำนายที่ใช้ Attribute สูงสุดสำหรับการทำนายใช้ถึง 26 ตัว และสมการทำนายที่ใช้ Attribute น้อยที่สุดจำนวน 5 Attribute ก็สามารถทำนายได้แล้ว ซึ่งได้แก่ Earn_F (รายได้โดยประมาณต่อปีของบิดา), GPAX_High (เกรดเฉลี่ยสะสมมัธยมศึกษาตอนปลาย ของผู้เรียนในสาขาวิชาเทคโนโลยีการศึกษา), Admiss_C (วิธีการรับเข้าของนิสิต: Admission กลาง ทปอ.), Earn_M (รายได้โดยประมาณต่อปีของมารดา) และ Voc_Far_Gov (อาชีพของบิดา: รับราชการ)

ตารางที่ 3.3 Model Summary ของโปรแกรม SPSS วิธีการ Stepwise

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.365(a)	.133	.132	.355	.133	112.922	1	736	.000
2	.380(b)	.144	.142	.353	.011	9.511	1	735	.002
a Predictors: (Constant), GPAX_High									
b Predictors: (Constant), GPAX_High, Special									

จากตารางที่ 3.3 Model Summary ของโปรแกรม SPSS วิธีการ Stepwise พบว่าสามารถสร้างสมการทำนายได้ถึง 2 รูปแบบ เมื่อพิจารณาค่า R Square สูงที่สุด พบว่าสมการทำนายที่สามารถทำนายได้ดีคือ Model ที่ 2 มี Attribute ประกอบด้วย GPAX_High (เกรดเฉลี่ยสะสมมัธยมศึกษาตอนปลาย ของผู้เรียนในสาขาวิชาเทคโนโลยีการศึกษา) และ Special (วิธีการรับเข้าของนิสิต: รับตรงรอบ 5- รอบพิเศษ)

สรุปผลลัพธ์การคัดเลือกตัวแปรหรือ Attribute

1. แบบ Backward Selection มีจำนวนตัวแปรที่มีอิทธิผลต่อการสร้างตัวแบบ จำนวน 5 Attribute ดังนี้

1.1 เกรดเฉลี่ยสะสมมัธยมศึกษาตอนปลาย ของผู้เรียนในสาขาวิชาเทคโนโลยีการศึกษา

1.2 วิธีการรับเข้าของนิสิต: Admission กลาง ทปอ.

1.3 รายได้โดยประมาณต่อปีของบิดา

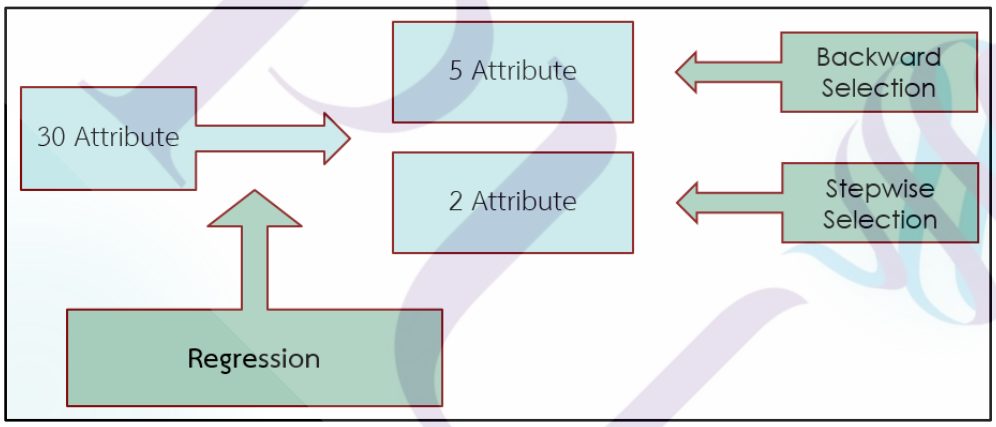
1.4 รายได้โดยประมาณต่อปีของมารดา

1.5 อาชีพของบิดา: รับราชการ

2. แบบ Stepwise Selection มีจำนวนตัวแปรที่มีอิทธิผลต่อการสร้างตัวแบบ จำนวน 2 Attribute ดังนี้

2.1 เกรดเฉลี่ยสะสมมัธยมศึกษาตอนปลาย ของผู้เรียนในสาขาวิชาเทคโนโลยีการศึกษา

2.2 วิธีการรับเข้าของนิสิต: รับตรงรอบ 5 (รอบพิเศษ)



ภาพที่ 3.5 Attribute หลังการกรองด้วยวิธี Regression

3.2 การกลั่นกรองข้อมูล (Data Cleansing)

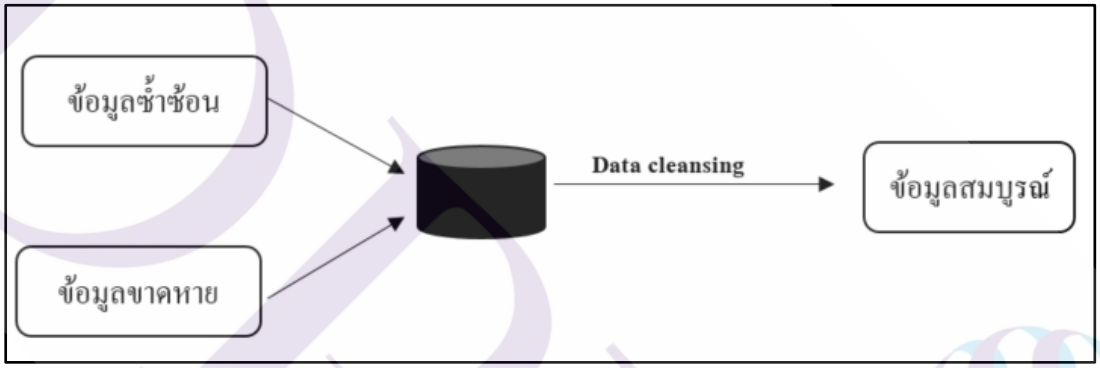
ผู้วิจัยดำเนินการตรวจสอบความถูกต้อง ความครบถ้วน และความเหมาะสมของข้อมูลที่ได้จากระบบของกองทะเบียนและประมวลผล มหาวิทยาลัยบูรพา เนื่องจากอาจจะมีข้อมูลบางส่วนที่ไม่มีความสมบูรณ์ เช่น ข้อมูลมีความซ้ำซ้อนกันหรือมีข้อมูลที่ขาดหายไป เป็นต้น

[16] ลักษณะข้อมูลที่จะทำการประมวลผลหรือการวิเคราะห์นั้นอาจมีข้อมูลที่มีสิ่งรบกวนปรากฏอยู่ในข้อมูลซึ่งมีหลากหลายสาเหตุ เช่น

1. ข้อมูลไม่สมบูรณ์ (Incomplete Data) เช่น ค่าของคุณลักษณะบางอย่างขาดหายไป (Missing Value) ขาดคุณลักษณะที่น่าสนใจหรือขาดรายละเอียดของข้อมูล เป็นต้น

2. ข้อมูลถูกรบกวน (Noisy Data) เช่น ข้อมูลมีค่าผิดพลาด (Error) หรือมีค่าผิดปกติ (Outliers) สาเหตุความผิดพลาด อาจจะเกิดจากอุปกณ์เก็บรวบรวมข้อมูลทำหน้าที่ผิดพลาด เกิดจากปัญหาการบันทึกหรือป้อนค่าข้อมูลและปัญหาการส่งข้อมูล (Data Transmission) ผิดพลาด เป็นต้น

3. ข้อมูลไม่สอดคล้อง (Inconsistent Data) เช่น ข้อมูลเดียวกันแต่ตั้งชื่อต่างกัน หรือใช้ค่าแทนข้อมูลที่ต่างกัน



ภาพที่ 3.6 การกลั่นกรองข้อมูล [16]

3.3 การแปลงข้อมูล (Data Transformation)

หลังจากได้ทำการกลั่นกรองข้อมูล เพื่อให้ได้ข้อมูลที่มีความสมบูรณ์สำหรับการประมวลผลต่อไป ทั้งนี้ข้อมูลที่นำมาวิเคราะห์ไม่ว่าจะเป็นตัวเลข ข้อมูลเป็นช่วง หรือข้อมูลที่เป็นข้อความ จะต้องทำการแปลงหรือแทนค่าให้อยู่ในรูปแบบมาตรฐานเดียวกัน เพื่อให้ง่ายต่อการวิเคราะห์

การแทนค่าข้อมูลที่ถูกต้องจะส่งให้การประมวลผลของคอมพิวเตอร์ทำได้อย่างรวดเร็ว มีความถูกต้องแม่นยำในการวิเคราะห์ข้อมูล ซึ่งการทดลองจะต้องแปลงข้อมูลในภาษามนุษย์ให้เป็นรูปแบบที่คอมพิวเตอร์เข้าใจ

ตารางที่ 3.4 ตัวแปรสำหรับการประมวลผลข้อมูลด้วยเทคนิคเหมืองข้อมูล

ตัวแปร	ความหมายและการแทนค่า
ID	เลขรหัสประจำตัวนิสิต
Success	ระดับสัมฤทธิผลสำหรับการเรียนของนิสิตสาขาวิชาเทคโนโลยีการศึกษา แบ่งตามคะแนนผลการเรียนเฉลี่ยสะสม 2 ระดับ ดังนี้ 1. ผลการเรียนเฉลี่ยสะสมระหว่าง 3.00 – 4.00 กำหนดเป็น Good 2. ผลการเรียนเฉลี่ยสะสมระหว่าง 2.00 – 2.99 กำหนดเป็น Bad
GPAX_High_School	เกรดเฉลี่ยสะสมมัธยมศึกษาตอนปลาย ของผู้เรียนในสาขาวิชาเทคโนโลยีการศึกษา
Admiss_Central	วิธีการรับเข้าของนิสิต: Admission กลาง ทปอ. Yes = ใช่ No = ไม่ใช่
Admiss_Special	วิธีการรับเข้าของนิสิต: รับตรงรอบ 5 (รอบพิเศษ) Yes = ใช่ No = ไม่ใช่
Earn_Farther	รายได้โดยประมาณต่อปีของบิดา
Earn_Mother	รายได้โดยประมาณต่อปีของมารดา
Voc_Far_Gov	อาชีพของบิดา: รับราชการ Yes = ใช่ No = ไม่ใช่

3.4 การสร้างตัวแบบทดสอบและเปรียบเทียบประสิทธิภาพ

การสร้างตัวแบบ (Model) ผู้วิจัยเลือกใช้เทคนิควิธี หรืออัลกอริทึมสำหรับการทดลอง ดังนี้

1. เทคนิค Decision Tree

จากการศึกษางานวิจัยที่เกี่ยวข้อง พบว่างานวิจัยส่วนใหญ่นำเทคนิคต้นไม้ตัดสินใจมาใช้สำหรับการวิเคราะห์และสร้างตัวแบบซึ่งผลการทดลองที่ได้มีความถูกต้องสูงในการสร้างสมการทำนาย อีกทั้งข้อมูลหรือคุณลักษณะที่ผู้วิจัยนำมาวิเคราะห์เป็นข้อมูลที่มีลักษณะของการใช้เทคนิค

ต้นไม้ตัดสินใจได้ สำหรับการคัดแยก การจำแนก หรือการคัดกรองคุณลักษณะเพื่อวิเคราะห์ดูว่าคุณลักษณะใดที่มีความเหมาะสมในการคัดเลือกบุคคลเข้าศึกษาต่อได้อย่างมีประสิทธิภาพ

2. เทคนิค Support Vector Machine

เทคนิค Support Vector Machine เป็นอัลกอริทึมสำหรับการคัดแยกที่มีการนำมาใช้กันอย่างกว้างขวางในด้านการประมวลผลภาพดิจิตอล หลักการของ SVM คือการให้ Input ที่ใช้ฝึกเป็นเวคเตอร์ในสเปซ N มิติ เช่นถ้าในกรณีของ 2 มิติ และ 3 มิติ จะเป็นจุดที่อยู่ในระนาบ xy และสเปซ xyz ตามลำดับ จากนั้นทำการสร้างไฮเปอร์เพลน (Hyperplane) ที่จะแยกกลุ่มของเวคเตอร์อินพุทออกเป็นประเภทต่าง ๆ ในกรณีที่เป็น 2 มิติ และ 3 มิติ ไฮเปอร์เพลน คือเส้นตรงและระนาบตามลำดับ ข้อเด่นของ SVM จะทำการเก็บแมพ (Map) เวคเตอร์ในสเปซ Input ให้เข้าสู่ Feature Space โดยใช้ฟังก์ชั่นหรือเรียกว่าเคอร์นัล (kernel) ชนิดต่าง ๆ เช่น โพลีโนเมียล (Polynomial) เรเดียล (Radial) เป็นต้น นอกจากนั้น Support Vector Machine เป็นตัวแบบที่ใช้ในการระบุตัวบุคคลหรือ Objectโดย SVM จะทำการแบ่งชั้นของข้อมูลด้วยระนาบหลายมิติ จากข้อมูล 2 กลุ่ม จากหลักการของ Support Vector Machine น่าจะสามารถนำมาคัดแยก การจำแนก หรือการคัดกรองคุณลักษณะเพื่อวิเคราะห์ดูว่าคุณลักษณะใดที่มีความเหมาะสมในการคัดเลือกบุคคลเข้าศึกษาต่อได้อย่างมีประสิทธิภาพ

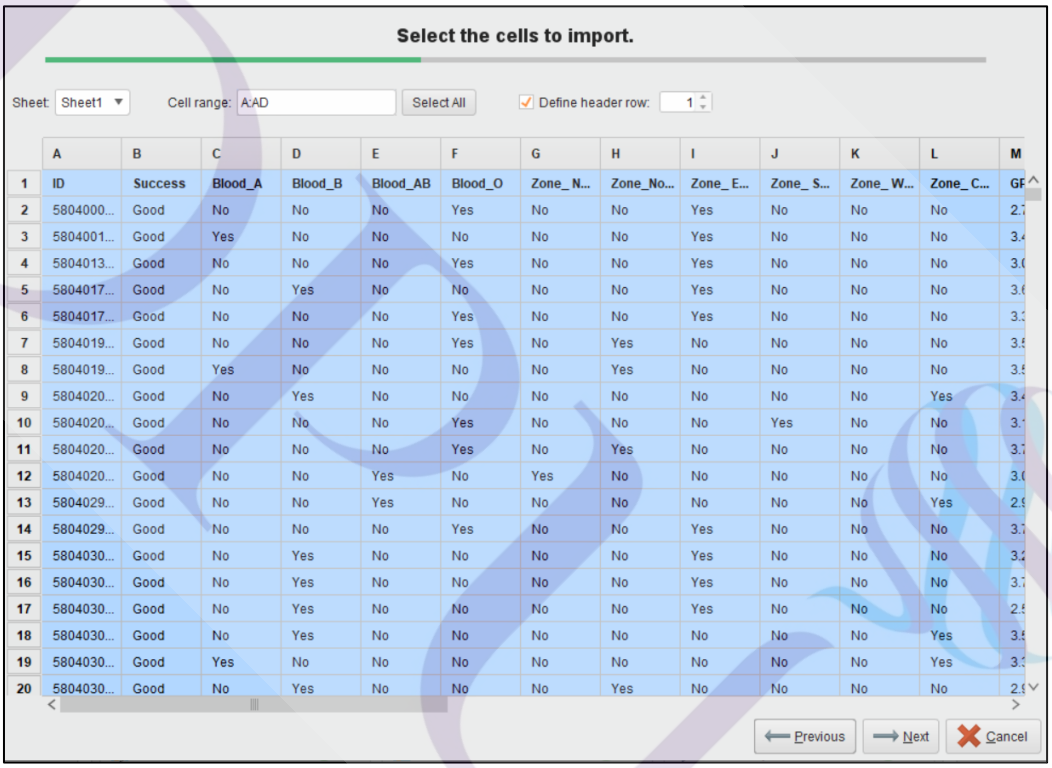
3. เทคนิค Deep Learning

เทคนิค Deep Learning เป็นรูปแบบที่มีลักษณะต่อยอดจากเทคนิคโครงข่ายประสาทเทียม (Artificial Neural Network: ANN) เป็นการจำลองการทำงานของสมองมนุษย์ โดยให้คอมพิวเตอร์รู้จักคิดและจดจำได้แบบเดียวกับโครงข่ายประสาทของมนุษย์ โดยจำลองว่า Input แต่ละตัวมีน้ำหนัก (Weight) อยู่ และเซลล์แต่ละหน่วยจะมีตัวกำหนด (Threshold) ว่าน้ำหนักรวมของ Input มากเพียงใด ถึงจะส่ง Output ต่อไปได้ การทำงานของ Deep Learning เป็นการใช้หลักการของ Artificial Neural Network มาใช้ในระดับที่ลึกและมีจำนวน Layer ที่ซับซ้อนกว่า โดยเทคนิคการเรียนรู้เชิงลึก (Deep Learning) เป็นขั้นตอนการเรียนรู้ที่จะกระทำผ่าน Neural Network ชุดข้อมูลสำหรับ Train จะถูกส่งไปสู่ Neural Network ข้อมูลขาเข้าแต่ละข้อมูลจะเข้าไปสู่ Neuron และถูกคูณด้วยน้ำหนัก (Weight) ผลลัพธ์จากการคูณจะนำไปสู่ Layer ต่อไปและกลายเป็นข้อมูลขาเข้าของ Layer ต่อไปด้วย กระบวนการนี้ถูกทำซ้ำในแต่ละ Layer ในโครงข่าย Layer สุดท้ายถูกเรียกว่า Layer ขาออก (Output Layer) มันจะคอยช่วยสนับสนุนค่าที่จะเกิดขึ้นจริงจากการศึกษา Attribute ที่เลือกสำหรับการคัดกรองบุคคลที่มีแนวโน้มสัมฤทธิผลในการศึกษาระดับปริญญาตรี สาขาวิชาเทคโนโลยีการศึกษา สามารถใช้เทคนิค Deep Learning เพื่อการวิเคราะห์ได้

ผู้วิจัยนำข้อมูลที่ผ่านกระบวนการคัดเลือก และกระบวนการ Cleansing นำเข้าโปรแกรม RapidMiner Studio (ผู้วิจัยใช้ e-mail สถาบันในการลงทะเบียนการใช้งานโปรแกรม) เพื่อทำการประมวลผลของข้อมูล โดยข้อมูลชุดฝึกสอนหรือเรียนรู้ (Train set) และชุดทดสอบ (Test) เป็นบัณฑิตที่สำเร็จการศึกษาระหว่างปีการศึกษา 2554-2558 จำนวน 738 คนโดยดำเนินการดังนี้

การสร้างตัวแบบด้วยเทคนิค Decision Tree

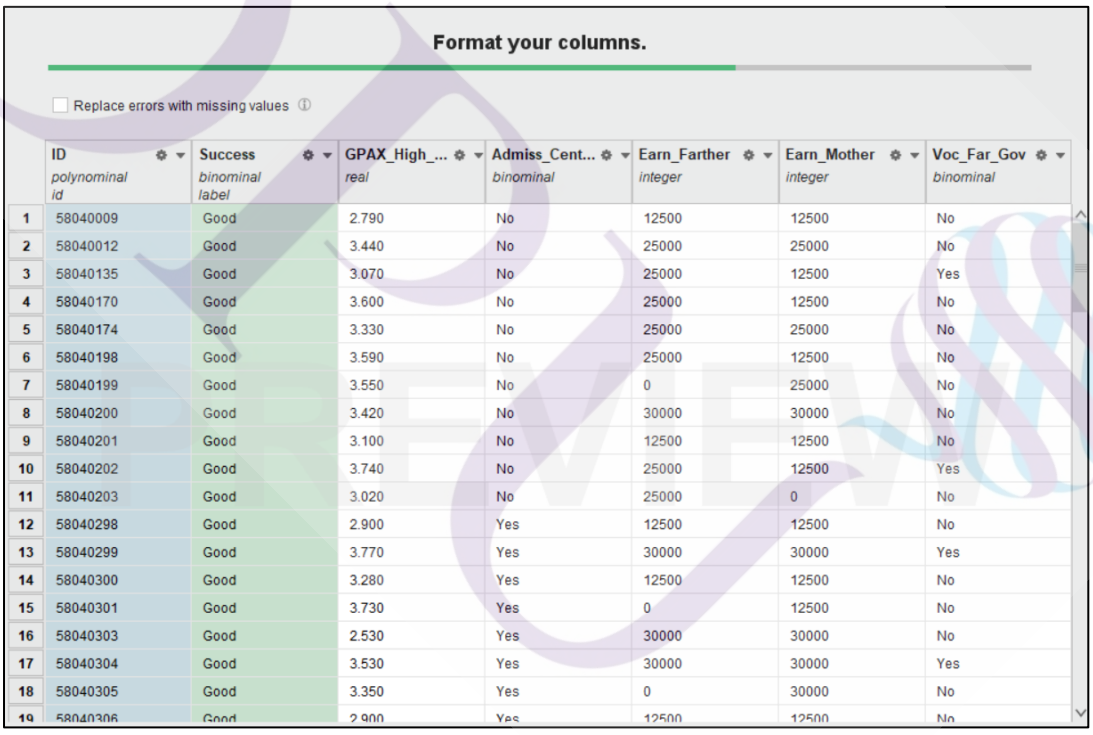
1. ผู้วิจัยเลือกข้อมูลเพื่อการนำเข้าข้อมูลสำหรับการวิเคราะห์ (Select the cells to import) ดังภาพที่ 3.7



ภาพที่ 3.7 การนำเข้าข้อมูลของโปรแกรม RapidMiner Studio

2. ผู้วิจัยกำหนดประเภทของข้อมูลและกำหนดบทบาทของ Attribute ดังนี้

Attribute	Change Type	Change Role
ID	Polynominal	ID
Success	Binominal	Label
GPAX_High_School	Real	-
Admiss_Central	Binominal	-
Admiss_Special	Binominal	-
Earn_Farther	Integer	-
Earn_Mother	Integer	-
Voc_Far_Gov	Binominal	-



Format your columns.

Replace errors with missing values

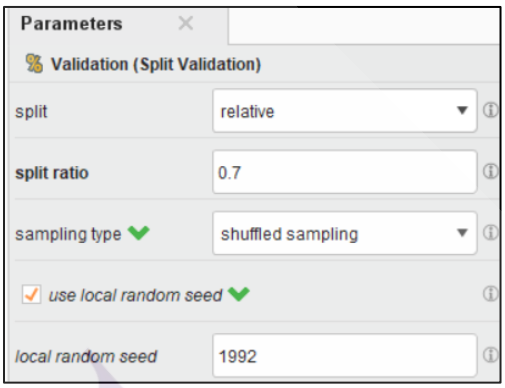
	ID *polynominal* *id*	Success *binominal* *label*	GPAX_High_... *real*	Admiss_Cent... *binominal*	Earn_Farther *integer*	Earn_Mother *integer*	Voc_Far_Gov *binominal*
1	58040009	Good	2.790	No	12500	12500	No
2	58040012	Good	3.440	No	25000	25000	No
3	58040135	Good	3.070	No	25000	12500	Yes
4	58040170	Good	3.600	No	25000	12500	No
5	58040174	Good	3.330	No	25000	25000	No
6	58040198	Good	3.590	No	25000	12500	No
7	58040199	Good	3.550	No	0	25000	No
8	58040200	Good	3.420	No	30000	30000	No
9	58040201	Good	3.100	No	12500	12500	No
10	58040202	Good	3.740	No	25000	12500	Yes
11	58040203	Good	3.020	No	25000	0	No
12	58040298	Good	2.900	Yes	12500	12500	No
13	58040299	Good	3.770	Yes	30000	30000	Yes
14	58040300	Good	3.280	Yes	12500	12500	No
15	58040301	Good	3.730	Yes	0	12500	No
16	58040303	Good	2.530	Yes	30000	30000	No
17	58040304	Good	3.530	Yes	30000	30000	Yes
18	58040305	Good	3.350	Yes	0	30000	No
19	58040306	Good	2.900	Yes	12500	12500	No

ภาพที่ 3.8 การกำหนดประเภทของข้อมูลและกำหนดบทบาทของ Attribute ในโปรแกรม Rapid Miner Studio

3. ผู้วิจัยกำหนด Operator ดังนี้

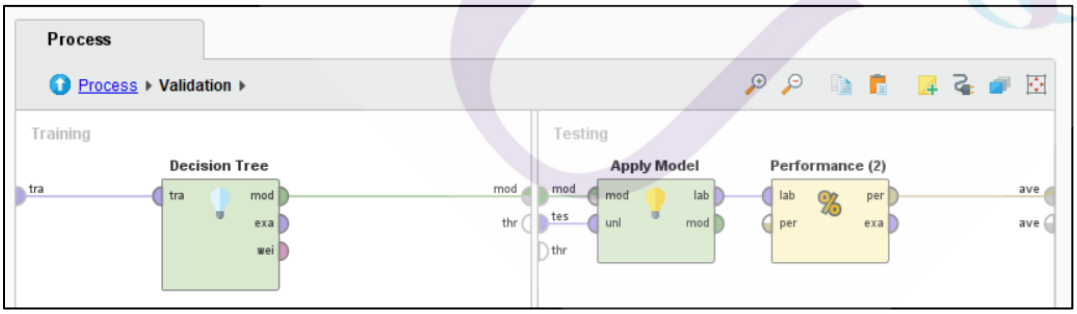
3.1 เลือก Retrieve Data เพื่อนำชุดข้อมูลที่สร้างขึ้นสำหรับการสร้างโมเดล

3.2 เลือก Validation เพื่อกำหนดสัดส่วนของการ Train และการ Test ในส่วนของ Split Ratio โดยผู้วิจัยกำหนดค่าสำหรับการ Train ตั้งแต่ 0.70, 0.75 และ 0.80 ตามลำดับ และใช้งานฟังก์ชั่น Use Local Random Seed โดยกำหนดเป็นค่า 1992 ดังภาพที่ 3.9



ภาพที่ 3.9 การกำหนด Validation การ Train 0.70, 0.75 และ 0.80

3.3 กำหนดค่า Operator ของ Validation ภายในจะแบ่งออกเป็น Training และ Testing โดยเลือก Operator Decision Tree ลงในด้านของ Training และเลือก Operator Apply Model และ Performance classification ในด้านของ Testing จากนั้นลากเส้นเชื่อมโยง Operator ดังภาพที่ 3.10



ภาพที่ 3.10 Process ของ Decision Tree

การสร้างตัวแบบด้วยเทคนิค Support Vector Machine

1. ผู้วิจัยเลือกข้อมูลเพื่อการนำเข้าข้อมูลสำหรับการวิเคราะห์ (Select the cells to import) เหมือนกับเทคนิค Decision Tree

2. ผู้วิจัยกำหนดประเภทของข้อมูลและกำหนดบทบาทของ Attribute ดังนี้

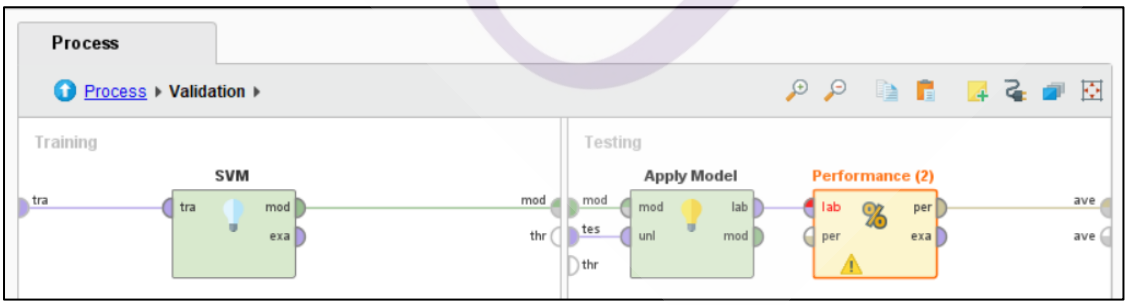
Attribute	Change Type	Change Role
Success	Binominal	Label
GPAX_High_School	Real	-
Earn_Farther	Integer	-
Earn_Mother	Integer	-

3. ผู้วิจัยกำหนด Operator สำหรับ Support Vector Machine ดังนี้

3.1 เลือก Retrieve Data เพื่อนำชุดข้อมูลที่สร้างขึ้นสำหรับการสร้างโมเดล

3.2 เลือก Validation เพื่อกำหนดสัดส่วนของการ Train และการ Test ในส่วนของ Split ratio โดยผู้วิจัยกำหนดค่าสำหรับการ Train ตั้งแต่ 0.70, 0.75 และ 0.80 ตามลำดับ และใช้งานฟังก์ชั่น use local random seed โดยกำหนดเป็นค่า 1992

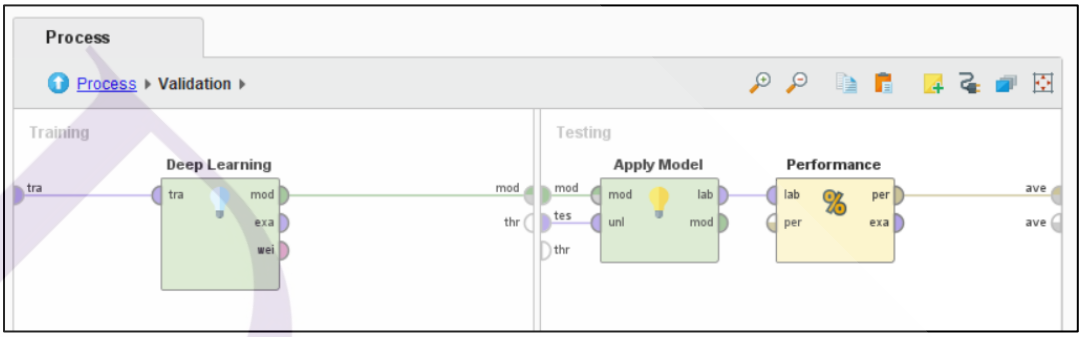
3.3 กำหนดค่า Operator ของ Validation ซึ่งภายในจะแบ่งออกเป็น Training และ Testing โดยเลือก Operator Support Vector Machine ลงในด้านของ Training และเลือก Operator Apply Model และ Performance Binominal Classification ในด้านของ Testing จากนั้นลากเส้นเชื่อมโยง Operator ดังภาพที่ 3.11



ภาพที่ 3.11 Process ของ Support Vector Machine

การสร้างตัวแบบด้วยเทคนิค Deep Learning

ขั้นตอนการนำข้อมูลเข้าสำหรับการวิเคราะห์เหมือนกับการสร้างโมเดลด้วยเทคนิค Decision Tree ดังภาพที่ 3.12 และผู้วิจัยกำหนดค่าใน Deep Learning เป็น 5 Hidden Layer แต่ละ Layer มี 50 Node



ภาพที่ 3.12 Process ของ Deep Learning

บทที่ 4
ผลการวิจัย

งานวิจัยนี้ได้นำเทคนิคการทำเหมืองข้อมูล เพื่อสร้างตัวแบบสำหรับการคัดกรองบุคคลที่มีแนวโน้มสัมฤทธิผลในการศึกษาระดับปริญญาตรี สาขาวิชาเทคโนโลยีการศึกษา และเพื่อเปรียบเทียบประสิทธิภาพของตัวแบบจำแนกตามร้อยละจำนวนการฝึกสอนและจำนวนข้อมูลฝึกสอนที่แตกต่างกัน ด้วยเทคนิค Decision Tree, Support Vector Machine และ Deep Learning

ตารางที่ 4.1 จำนวนกลุ่มตัวอย่างสำหรับการเก็บข้อมูล จำแนกตามเพศ

เพศ	จำนวน	ร้อยละ
ชาย	268	36.31
หญิง	470	63.69
รวม	**738**	**100**

จากตารางที่ 4.1 จำนวนกลุ่มตัวอย่างสำหรับการเก็บข้อมูลเป็นบัณฑิตที่สำเร็จการศึกษาระหว่างปีการศึกษา 2554-2558 จำนวน 738 คน จำแนกเป็นเพศชาย จำนวน 268 คน คิดเป็นร้อยละ 36.31 และเป็นเพศหญิงจำนวน 470 คน คิดเป็นร้อยละ 63.69

ตารางที่ 4.2 จำนวนกลุ่มตัวอย่างที่สำเร็จการศึกษาสาขาวิชาเทคโนโลยีการศึกษา จำแนกตามปีการศึกษาที่เข้าเรียนและประเภทนิสิต

ปีการศึกษาที่เข้าเรียน	ประเภท	จำนวน
2558	ภาคปกติ	21
	ภาคพิเศษ	115
	รวม	**136**
2557	ภาคปกติ	53
	ภาคพิเศษ	66
	รวม	**119**
2556	ภาคปกติ	39
	ภาคพิเศษ	77
	รวม	**116**
2555	ภาคปกติ	43
	ภาคพิเศษ	154
	รวม	**197**
2554	ภาคปกติ	87
	ภาคพิเศษ	83
	รวม	**170**
รวม	**ภาคปกติ**	**243**
	ภาคพิเศษ	**495**
	ทั้งหมด	**738**

จากตารางที่ 4.2 จำนวนกลุ่มตัวอย่างที่สำเร็จการศึกษาสาขาวิชาเทคโนโลยีการศึกษา จำแนกตามปีการศึกษาที่เข้าเรียน และประเภทนิสิตพบว่านิสิตที่เข้าเรียนปีการศึกษา 2555 มีจำนวนสูงที่สุด คือ 197 คน และระหว่างปีการศึกษา 2554-2558 เป็นนิสิตภาคพิเศษ จำนวน 495 คน

4.1 ผลการทดสอบประสิทธิภาพของตัวแบบ

การทดสอบประสิทธิภาพของตัวแบบ ผู้วิจัยดำเนินการ 3 เทคนิค ดังนี้ Decision Tree, Support Vector Machine และ Deep Learning ผู้วิจัยกำหนด Label เป็น 2 ระดับคือ Good และ Bad จากผลการเรียนเฉลี่ยสะสม (GPAX) ของนิสิตสาขาวิชาเทคโนโลยีการศึกษา จำแนกได้ดังนี้

1. ผลการเรียนเฉลี่ยสะสมระหว่าง 3.00 – 4.00 กำหนดเป็นสัมฤทธิผลดี แทนค่าด้วย Good

2. ผลการเรียนเฉลี่ยสะสมระหว่าง 2.00 – 2.99 กำหนดเป็นสัมฤทธิผลไม่ดี แทนค่าด้วย Bad

ผู้วิจัยดำเนินการแบ่งจำนวนกลุ่มตัวอย่างเพื่อการ Train ดังตารางที่ 4.3 โดยสัดส่วนนี้นำไปทดสอบกับเทคนิค Decision Tree, Support Vector Machine และ Deep Learning

ตารางที่ 4.3 สัดส่วนการ Train และการ Test

จำนวนข้อมูล	**สัดส่วน**		
	70%	**75%**	**80%**
ชุดการ Train	517	554	590
ชุดการ Test	221	184	148

จากตารางที่ 4.3 จำนวนกลุ่มตัวอย่าง 738 คน นำมาแบ่งสัดส่วนการ Train และการ Test เพื่อทดสอบตัวแบบตามเทคนิคการทำเหมืองข้อมูลทั้ง 3 เทคนิค คือ Decision Tree, Support Vector Machine และ Deep Learning แบ่งกลุ่มการฝึกสอน (Training set) ดังสัดส่วนต่อไปนี้ 70%, 75% และ 80% ได้จำนวนกลุ่มตัวอย่างดังนี้ 517, 554 และ 590 คน ตามลำดับ

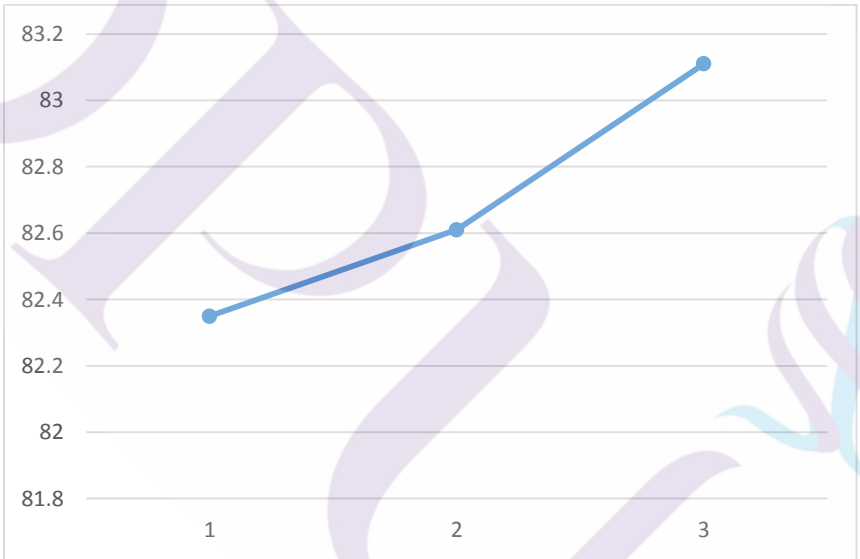
4.2 ผลการทดสอบด้วยเทคนิค Decision Tree

ผู้วิจัยนำข้อมูลทุก Attribute และข้อมูลที่มีการคัดเลือก Attribute ทั้งแบบ Backward Selection และ Stepwise Selection เข้าทดสอบตัวแบบด้วยเทคนิค Decision Tree และ Support Vector Machine พบว่า ข้อมูลที่มีการคัดเลือก Attribute จะมีค่าความถูกต้อง (Accuracy) เท่ากันทั้ง 2 เทคนิค ดังนี้

ตารางที่ 4.4 ผลการทดสอบโมเดลด้วยเทคนิค Decision Tree นำข้อมูลเข้าแบบ All Attribute

Train	70%	75%	80%
Accuracy	82.35	82.61	83.11
Classification error	17.65	17.39	16.89
Correlation	0.115	0.128	0.186
Cross-entropy	∞	∞	∞

จากตารางที่ 4.4 โมเดลการเรียนรู้ที่มีสัดส่วนของการ Train แตกต่างกัน พบว่าค่าความถูกต้องมีแนวโน้มสูงขึ้น (82.35%, 82.61% และ 83.11% ตามลำดับ) และ Attribute มีความสัมพันธ์ซึ่งกันและกัน



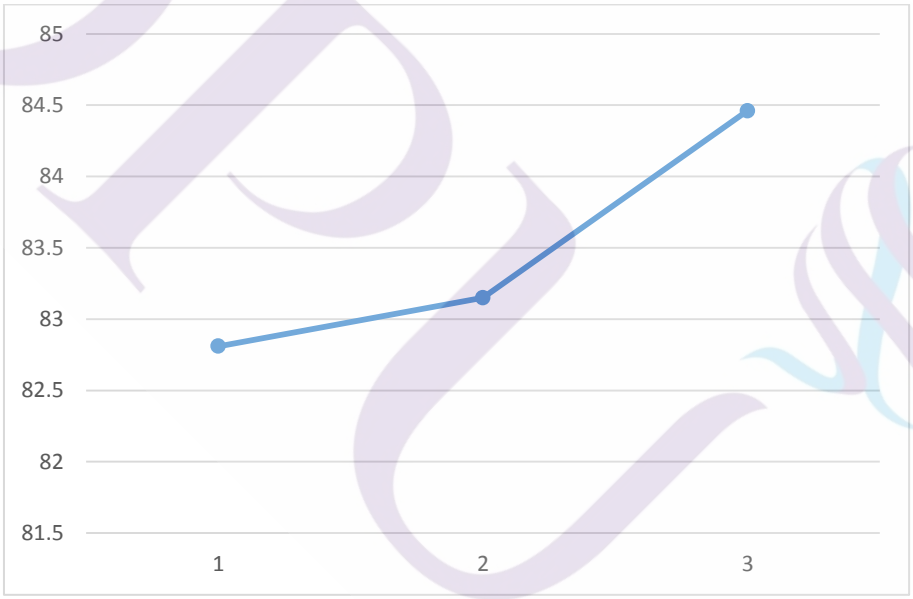
ภาพที่ 4.1 ผลการทดสอบโมเดลด้วยเทคนิค Decision Tree นำข้อมูลเข้าแบบ All Attribute

จากภาพที่ 4.1 ผลการทดสอบโมเดลด้วยเทคนิค Decision Tree นำข้อมูลเข้าแบบ All Attribute พบว่า ค่าความถูกต้องมีแนวโน้มสูงขึ้น

ตารางที่ 4.5 ผลการทดสอบโมเดลด้วยเทคนิค Decision Tree นำข้อมูลเข้าแบบ Backward และ Stepwise

Train	70%	75%	80%
Accuracy	82.81	83.15	84.46
Classification error	17.19	16.85	15.54
Correlation	0.187	0.209	0.307
Cross-entropy	0.664	0.656	∞

จากตารางที่ 4.5 โมเดลการเรียนรู้ที่มีสัดส่วนของการ Train แตกต่างกัน พบว่าค่าความถูกต้องมีแนวโน้มสูงขึ้น ดังนี้ 82.81%, 83.15% และ 84.46% ตามลำดับ และAttribute มีความสัมพันธ์ซึ่งกันและกัน โดยค่าที่ได้สูงกว่าการนำข้อมูลเข้าแบบ All Attribute

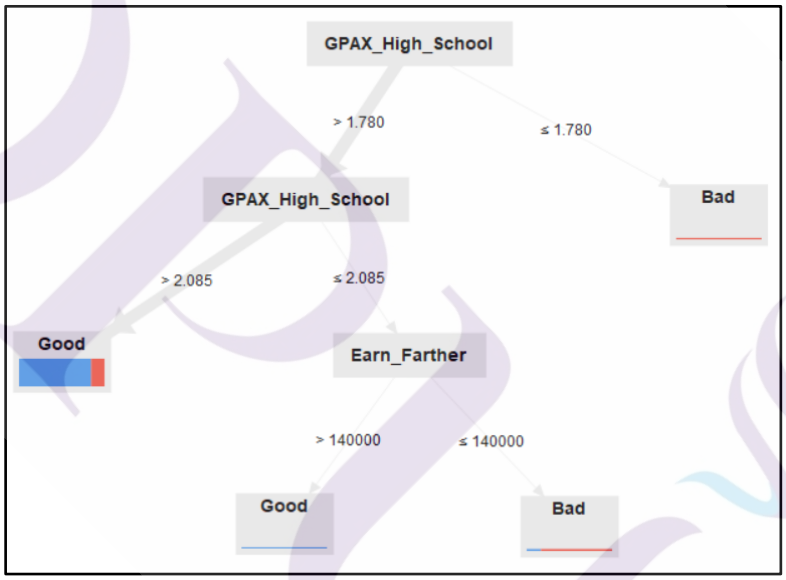


ภาพที่ 4.2 ผลการทดสอบโมเดลด้วยเทคนิค Decision Tree นำข้อมูลเข้าแบบ Backward และ Stepwise

จากภาพที่ 4.2 ค่าร้อยละความถูกต้อง (Accuracy) จะมีแนวโน้มสูงขึ้นมากกว่าร้อยละ 80 สามารถนำโมเดลที่เกิดขึ้นไปใช้ในภาคสนามได้

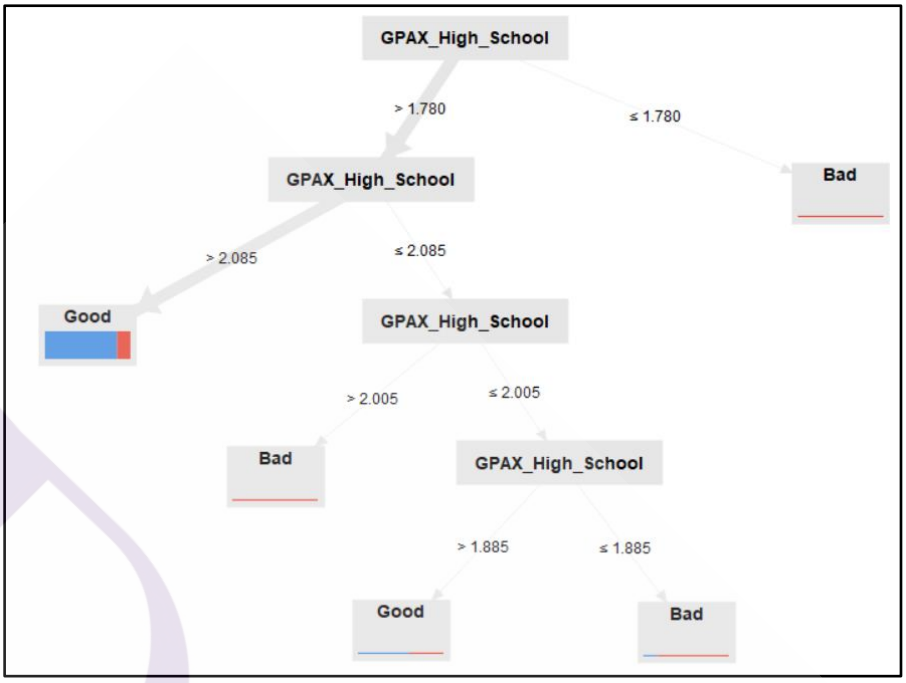
ความแตกต่างระหว่างค่าร้อยละความถูกต้อง (Accuracy) ในแต่ละสัดส่วนการ Train เนื่องจากข้อมูลที่นำมาวิเคราะห์เป็นบัณฑิตที่สำเร็จการศึกษาแล้ว โดยเข้าเรียนระหว่างปีการศึกษา 2554-2558 ซึ่งมีความแตกต่างกันในประเด็นด้านการเรียนการสอนระหว่างชั้นปีของผู้เรียน เพราะแต่ละวิชาจัดผู้สอนไม่เหมือนกันในแต่ละปีการศึกษา มีผลต่อการให้เกรดของผู้สอนในแต่ละชั้นปีด้วย ทำให้เกรดเฉลี่ยสะสมของผู้เรียนระหว่างชั้นปีมีความแตกต่างกัน อีกประเด็นหนึ่งคือด้านพัฒนาการเรียนรู้ของแต่ละบุคคลมีความแตกต่างกัน รวมทั้ง Cognitive Style ที่แต่ละคนมีความแตกต่างกัน

แบบจำลองที่สร้างจากเทคนิค Decision Tree



ภาพที่ 4.3 แบบจำลองที่ได้จากการสร้างด้วยเทคนิค Decision Tree นำข้อมูลเข้าแบบ All Attribute

จากภาพที่ 4.3 พบว่า Attribute ที่มีบทบาทต่อผู้เรียนที่จะมีสัมฤทธิผลที่ดีสำหรับการเรียนระดับปริญญาตรี สาขาวิชาเทคโนโลยีการศึกษา คือ เกรดเฉลี่ยสะสมมัธยมศึกษาตอนปลาย โดยผู้เรียนจะต้องมีผลการเรียนเฉลี่ยสะสมอยู่ระหว่าง 1.78 - 2.08 กรณีน้อยกว่าหรือเท่ากับ 1.78 จะมีสัมฤทธิผลที่ไม่ดีสำหรับการเรียนในสาขาวิชาเทคโนโลยีการศึกษา และถ้ารายได้โดยประมาณต่อปีของบิดามากกว่า 140,000 บาท แต่ผลการเรียนเฉลี่ยสะสมอยู่ระหว่าง 1.78 - 2.08 ผู้เรียนจะมีสัมฤทธิผลที่ดีต่อการเรียนในสาขาวิชาเทคโนโลยีการศึกษาได้



ภาพที่ 4.4 แบบจำลองที่ได้จากการสร้างด้วยเทคนิค Decision Tree นำข้อมูลเข้าแบบ Backward และ Stepwise

จากภาพที่ 4.4 พบว่า Attribute ที่มีบทบาทต่อผู้เรียนที่จะมีสัมฤทธิผลที่ดีสำหรับการเรียนระดับปริญญาตรี สาขาวิชาเทคโนโลยีการศึกษา คือ เกรดเฉลี่ยสะสมมัธยมศึกษาตอนปลาย เช่นกัน โดยผู้เรียนจะต้องมีผลการเรียนเฉลี่ยสะสมอยู่ระหว่าง 1.78 - 2.08 กรณีน้อยกว่าหรือเท่ากับ 1.78 จะมีสัมฤทธิผลที่ไม่ดีสำหรับการเรียนในสาขาวิชาเทคโนโลยีการศึกษา

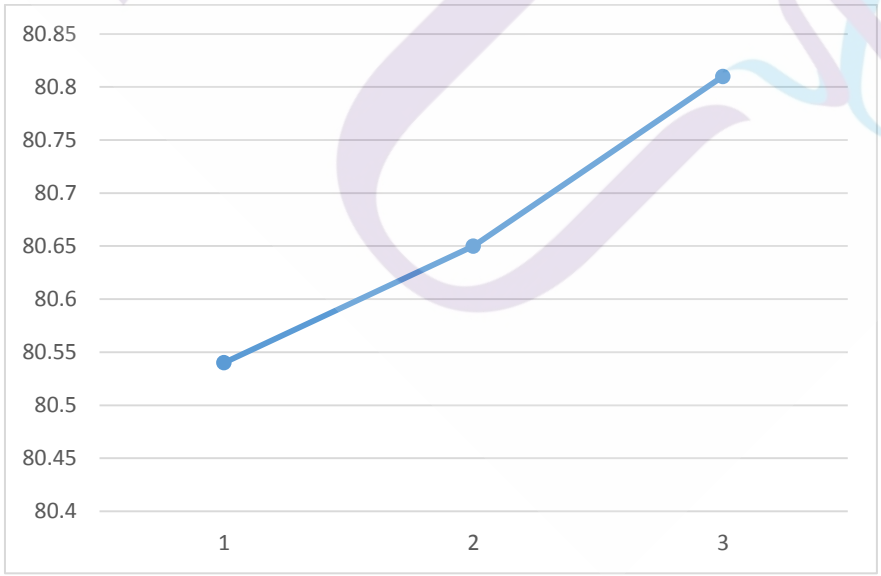
4.3 ผลการทดสอบด้วยเทคนิค Support Vector Machine

จากการนำข้อมูลเข้าวิเคราะห์ ทุก Attribute และการกรองข้อมูลทั้งแบบ Backward Selection และ Stepwise Selection ด้วยเทคนิค Support Vector Machine พบว่าได้ผลลัพธ์เหมือนกัน ดังตารางที่ 4.12

ตารางที่ 4.6 ผลการทดสอบโมเดลด้วยเทคนิค Support Vector Machine

Train	70%	75%	80%
Accuracy	80.54	80.65	80.81
Classification error	19.46	19.35	19.19
Specificity	100	100	100

จากตารางที่ 4.6 โมเดลการเรียนรู้ที่มีการฝึกสอน (Training set) แตกต่างกัน ด้วยเทคนิค Support Vector Machine พบว่า ค่าความถูกต้องมีแนวโน้มสูงขึ้น ดังนี้ 80.54%, 80.65% และ 80.81% ตามลำดับ แต่ทุกระดับการฝึกสอนมีค่าความจำเพาะ (Specificity) เท่ากับ 100% เป็นความแม่นยำของการทำนายข้อมูลที่อยู่ในคลาส Negative คำนวณได้จากอัตราส่วนของการทำนายข้อมูลที่อยู่ในคลาส Negative ได้ถูกต้องเทียบกับข้อมูลจริงทั้งหมดของคลาส Negative



ภาพที่ 4.5 ผลการทดสอบโมเดลด้วยเทคนิค Support Vector Machine

จากภาพที่ 4.5 ค่าร้อยละความถูกต้อง (Accuracy) จะมีแนวโน้มเพิ่มสูงขึ้นค่าที่ได้อยู่ในระดับร้อยละ 80 สามารถนำไปใช้ในภาคสนามได้

4.4 ผลการทดสอบด้วยเทคนิค Deep Learning

จากการนำข้อมูลเข้าวิเคราะห์ ทุก Attribute และการกรองข้อมูลทั้งแบบ Backward Selection และ Stepwise Selection ด้วยเทคนิค Deep Learning ได้ผลลัพท์ดังนี้

ตารางที่ 4.7 ผลการทดสอบโมเดลด้วยเทคนิค Deep Learning นำข้อมูลเข้าแบบ All Attribute

Train	70%	75%	80%
Accuracy	83.26	77.17	79.73
Classification error	16.74	22.83	20.27
Correlation	0.430	0.364	0.458
Cross-entropy	0.564	0.548	0.512

จากตารางที่ 4.7 โมเดลการเรียนรู้ที่มีการฝึกสอน (Training set) แตกต่างกันด้วยเทคนิค Deep Learning พบว่า ค่าความถูกต้องมีแนวโน้มขึ้นๆ ลงๆ ดังนี้ 83.26%, 77.17% และ 79.73% ตามลำดับ และ Attribute มีความสัมพันธ์ซึ่งกันและกันค่อนข้างสูง แสดงให้เห็นได้ว่า Attribute แต่ละตัวไม่มีอิสระต่อกัน

ตารางที่ 4.8 ผลการทดสอบโมเดลด้วยเทคนิค Deep Learning นำข้อมูลเข้าแบบ Backward

Train	70%	75%	80%
Accuracy	74.66	81.52	83.78
Classification error	25.34	18.48	16.22
Correlation	0.382	0.403	0.474
Cross-entropy	0.548	0.522	0.518

จากตารางที่ 4.8 โมเดลการเรียนรู้ที่มีการฝึกสอน (Training set) แตกต่างกันด้วยเทคนิค Deep Learning พบว่า ค่าความถูกต้องมีแนวโน้มขึ้นๆ ลงๆ ดังนี้ 74.66%, 81.52% และ 83.78% ตามลำดับ และ Attribute มีความสัมพันธ์ซึ่งกันและกันค่อนข้างสูง แสดงให้เห็นได้ว่า Attribute แต่ละตัวไม่มีอิสระต่อกัน

ตารางที่ 4.9 ผลการทดสอบโมเดลด้วยเทคนิค Deep Learning นำข้อมูลเข้าแบบ Stepwise

Train	**70%**	**75%**	**80%**
Accuracy	79.58	83.65	85.94
Classification error	20.42	16.35	14.06
Correlation	0.287	0.412	0.445
Cross-entropy	0.551	0.477	0.466

จากตารางที่ 4.9 โมเดลการเรียนรู้ที่มีการฝึกสอน (Training set) แตกต่างกันด้วยเทคนิค Deep Learning พบว่า ค่าความถูกต้องมีแนวโน้มขึ้นๆ ลงๆ ดังนี้ 79.58%, 83.65% และ 85.94% ตามลำดับ และ Attribute มีความสัมพันธ์ซึ่งกันและกันค่อนข้างสูง แสดงให้เห็นได้ว่า Attribute แต่ละตัวไม่มีอิสระต่อกัน

4.5 สรุปผลการทดลอง

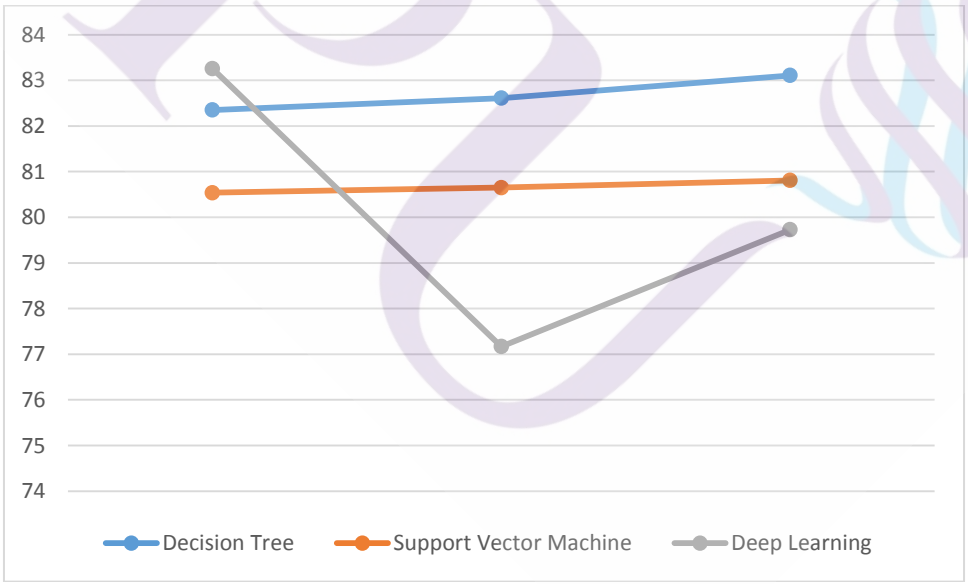
ตารางที่ 4.10 ภาพรวมของ Accuracy จำแนกตามเทคนิคเหมืองข้อมูลและจำนวน Attribute

เทคนิค	**Accuracy**								
	All Attribute			**5 Attribute**			**2 Attribute**		
	70%	**75%**	**80%**	**70%**	**75%**	**80%**	**70%**	**75%**	**80%**
Decision Tree	82.35	82.61	83.11	82.81	83.15	84.46	82.81	83.15	84.46
SVM	80.54	80.65	80.81	80.54	80.65	80.81	80.54	80.65	80.81
Deep Learning	83.26	77.17	79.73	74.66	81.52	83.78	79.58	83.65	85.94

จากตารางที่ 4.10 ภาพรวมของ Accuracy จำแนกตามเทคนิคเหมืองข้อมูลและจำนวน Attribute พบว่า ค่าความถูกต้องที่มีการกรอง Attribute แบบ Stepwise Selection มีค่าสูงกว่าการกรอง Attribute แบบ Backward Selection และแบบ All Attribute เนื่องจากวิธีการกรองแบบ Stepwise มีการขจัด Attribute ขยะที่ไม่มีบทบาทต่อการจำแนกข้อมูลออก เหลือเฉพาะ Attribute ที่มีบทบาทต่อการจำแนกข้อมูลเท่านั้น การกรอง Attribute แบบ Backward Selection จะมี Attribute ขยะบางส่วนทำให้ค่าความถูกต้อง (Accuracy) ไม่สูงมากนัก

ตารางที่ 4.11 เปรียบเทียบความถูกต้องของผลการทดสอบโมเดลด้วยเทคนิคที่ต่างกัน จำแนกตามสัดส่วนการ Train และนำข้อมูลเข้าแบบ All Attribute

เทคนิค	ร้อยละความถูกต้อง (Accuracy) N=738 คน		
	Train 70%	Train 75%	Train 80%
Decision Tree	82.35	82.61	83.11
Support Vector Machine	80.54	80.65	80.81
Deep Learning	83.26	77.17	79.73

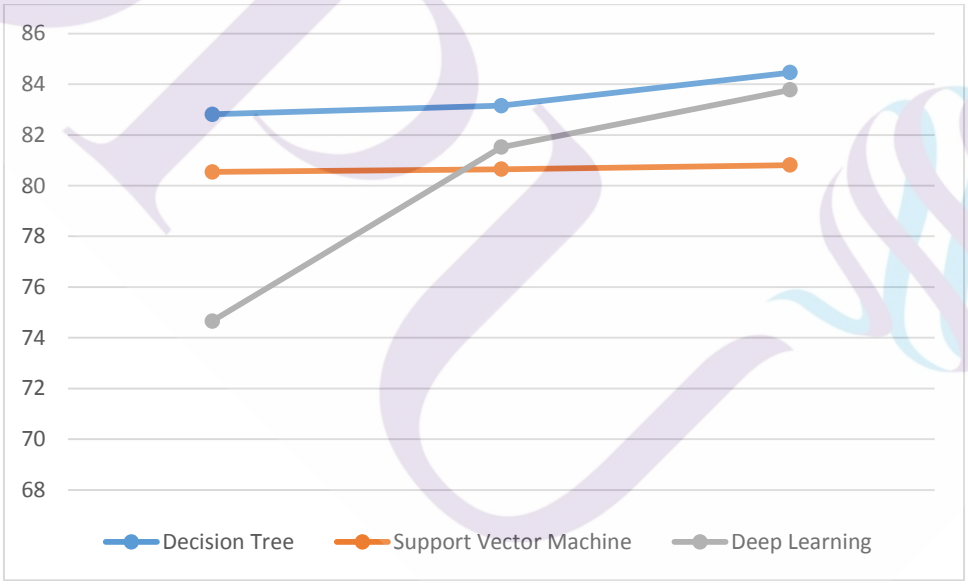


ภาพที่ 4.6 เปรียบเทียบความถูกต้องของผลการทดสอบโมเดลด้วยเทคนิคที่ต่างกัน จำแนกตามสัดส่วนการ Train และนำข้อมูลเข้าแบบ All Attribute

จากตารางที่ 4.11 และภาพที่ 4.6 พบว่า ร้อยละความถูกต้องในเทคนิค Decision Tree มีความถูกต้องสูงกว่า Support Vector Machine และ Deep Learning ตามลำดับ ค่าร้อยละความถูกต้องของเทคนิค Deep Learning มีแนวโน้มขึ้นๆ ลงๆ เนื่องจากมี Attribute ที่ไม่พึ่งประสงค์หรือมีความซ้ำซ้อนกันของ Attribute ทำให้การคำนวณออกมาได้ค่าร้อยละความถูกต้องขึ้นๆ ลงๆ

ตารางที่ 4.12 เปรียบเทียบความถูกต้องของผลการทดสอบโมเดลด้วยเทคนิคที่ต่างกัน จำแนกตามสัดส่วนการ Train และนำข้อมูลเข้าแบบ Backward

เทคนิค	ร้อยละความถูกต้อง (Accuracy) N=738 คน		
	Train 70%	Train 75%	Train 80%
Decision Tree	82.81	83.15	84.46
Support Vector Machine	80.54	80.65	80.81
Deep Learning	74.66	81.52	83.78

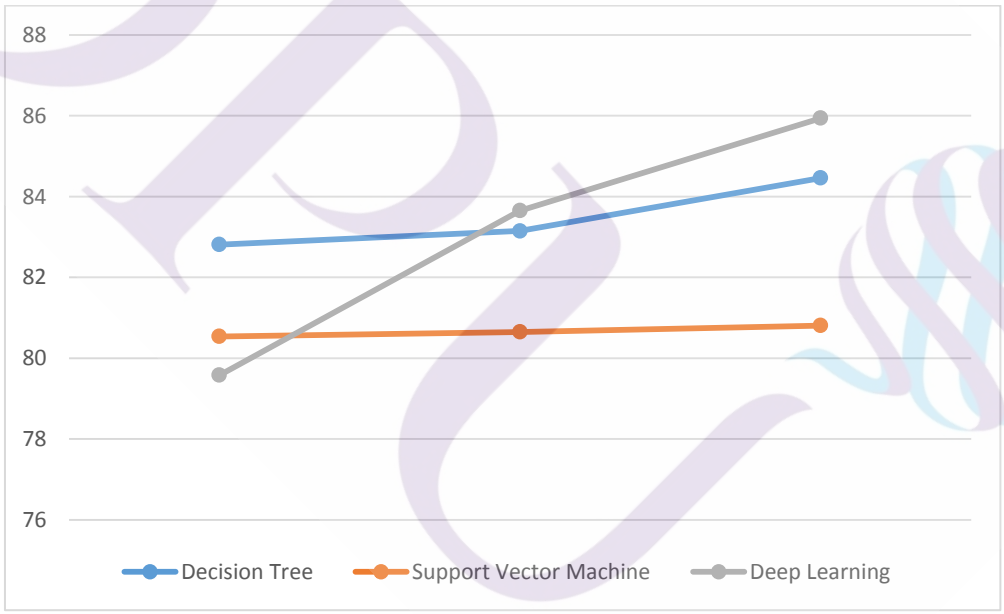


ภาพที่ 4.7 เปรียบเทียบความถูกต้องของผลการทดสอบโมเดลด้วยเทคนิคที่ต่างกัน จำแนกตามสัดส่วนการ Train และนำข้อมูลเข้าแบบ Backward

จากตารางที่ 4.12 และภาพที่ 4.7 พบว่า ร้อยละความถูกต้องในเทคนิค Decision Tree มีความถูกต้องสูงกว่า Deep Learning และ Support Vector Machine ตามลำดับ ซึ่งมีค่าร้อยละความถูกต้องมากกว่าร้อยละ 80

ตารางที่ 4.13 เปรียบเทียบความถูกต้องของผลการทดสอบโมเดลด้วยเทคนิคที่ต่างกัน จำแนกตามสัดส่วนการ Train และนำข้อมูลเข้าแบบ Stepwise

เทคนิค	**ร้อยละความถูกต้อง (Accuracy) N=738 คน**		
	Train 70%	Train 75%	Train 80%
Decision Tree	82.81	83.15	84.46
Support Vector Machine	80.54	80.65	80.81
Deep Learning	79.58	83.65	85.94



ภาพที่ 4.8 เปรียบเทียบความถูกต้องของผลการทดสอบโมเดลด้วยเทคนิคที่ต่างกัน จำแนกตามสัดส่วนการ Train และนำข้อมูลเข้าแบบ Stepwise

จากตารางที่ 4.13 และภาพที่ 4.8 พบว่า ร้อยละความถูกต้องในเทคนิค Deep Learning มีความถูกต้องสูงกว่า Decision Tree และ Support Vector Machine ตามลำดับ มีค่าร้อยละความถูกต้องมากกว่า 80%

ตารางที่ 4.14 ค่าเฉลี่ยร้อยละความถูกต้อง (Accuracy) จำแนกตามเทคนิคเหมืองข้อมูล

เทคนิค	ค่าเฉลี่ยร้อยละความถูกต้อง (Accuracy)
Decision Tree	83.47
Support Vector Machine	82.46
Deep Learning	83.06

จากตารางที่ 4.14 พบว่า ค่าเฉลี่ยร้อยละความถูกต้อง (Accuracy) ของเทคนิค Decision Tree และเทคนิค Deep Learning มีค่าเฉลี่ยร้อยละความถูกต้องใกล้เคียงกัน คือ 83.47% และ 83.06% ตามลำดับ รองลงมาเป็นเทคนิค Support Vector Machine = 82.46%

เมื่อพิจารณาจากผลการวิเคราะห์และการทดสอบตัวแบบสามารถสรุปได้ว่า

1. สามารถคัดเลือกตัวแปรเพื่อการวิเคราะห์ข้อมูล เพื่อคัดกรองบุคคลที่มีแนวโน้มสัมฤทธิผลในการศึกษาระดับปริญญาตรี สาขาวิชาเทคโนโลยีการศึกษา นั้นสามารถใช้เทคนิคการเลือกตัวแปรหรือแอททริบิวต์ด้วย Stepwise Selection เพื่อประหยัดเวลาสำหรับการประมวลผลและประหยัดเวลาในการเก็บรวบรวมข้อมูล เนื่องจาก Stepwise Selection สามารถสกัดตัวแปรที่ไม่เกี่ยวข้อง เหลือตัวแปรเพียง 2 ตัวก็สามารถสร้างโมเดลการทำนายได้ดีเท่ากับการคัดเลือกตัวแปรด้วย Backward Selection ซึ่งต้องวิเคราะห์ถึง 5 ตัวแปร กระบวนการคัดเลือก Attribute ดังกล่าวสอดคล้องกับ [16] งานวิจัยเรื่องการใช้เทคนิคการทำเหมืองข้อมูลเพื่อพยากรณ์ผลการเรียนของนักเรียน โรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน ที่ดำเนินการคัดเลือกคุณลักษณะ (Feature Selection) โดยใช้วิธี Correlation-based Feature Selection (CFS) และวิธี Information Gain (IG) และสอดคล้องกับ [20] งานวิจัยการเปรียบเทียบวิธีการคัดเลือกคุณลักษณะที่สำคัญในการปรับปรุงการพยากรณ์มะเร็งเต้านม โดยใช้วิธีการคัดเลือกคุณลักษณะจากเทคนิคต่าง ๆ จำนวน 7 เทคนิค ได้แก่ 1) เทคนิค Correlation Based Feature Selection 2) เทคนิค Information Gain (IG) 3) เทคนิค Gain Ratio (GR) 4) เทคนิค Chi-Square 5) เทคนิค Forward Selection 6) เทคนิค Backward Elimination และ 7) เทคนิค Evolutionary Selection พบว่า จากจำนวน

คุณลักษณะของข้อมูลทั้งหมด 30 คุณลักษณะ เทคนิค Evolutionary Selection สามารถลดคุณลักษณะที่สำคัญเหลือเพียง 16 คุณลักษณะ ให้ผลการวัดค่าความถูกต้องในการพยากรณ์ได้ดีถึง 95.26%

2. ผลการทดสอบโมเดลตามตารางที่ 4.12 และ 4.13 พบว่าค่าร้อยละความถูกต้องของเทคนิค Decision Tree และ Deep Learning มีแนวโน้มที่เพิ่มสูงขึ้นตามสัดส่วนการเรียนรู้ (Learning Set) ซึ่งสอดคล้องกับการวิจัยของ ณัฐริน เจริญเกียรติบวร ทำวิจัยเรื่องการใช้เทคนิคเหมืองข้อมูลเพื่อช่วยในการแนะแนวการศึกษาต่อระดับอุดมศึกษา ใช้การแบ่งตามสัดส่วนข้อมูล Train : Validation เท่ากับ 60:40, 65:35, 70:30 และ 75:25 ตามลำดับ พบว่าค่าความถูกต้องเป็น 82.32%, 81.87%, 91.78% และ 92.37% ตามลำดับ [3]

3. เทคนิคการทำเหมืองข้อมูล ทั้ง 3 เทคนิค คือ เทคนิค Decision Tree, Support Vector Machine และ Deep Learning สามารถเลือกเทคนิคใดเทคนิคหนึ่ง เพื่อคัดกรองบุคคลที่มีแนวโน้มสัมฤทธิผลในการศึกษาระดับปริญญาตรี สาขาวิชาเทคโนโลยีการศึกษาได้ เนื่องจากได้ค่าร้อยละความถูกต้อง (Accuracy) สูงกว่าร้อยละ 80

บทที่ 5
สรุปผลการวิจัยและข้อเสนอแนะ

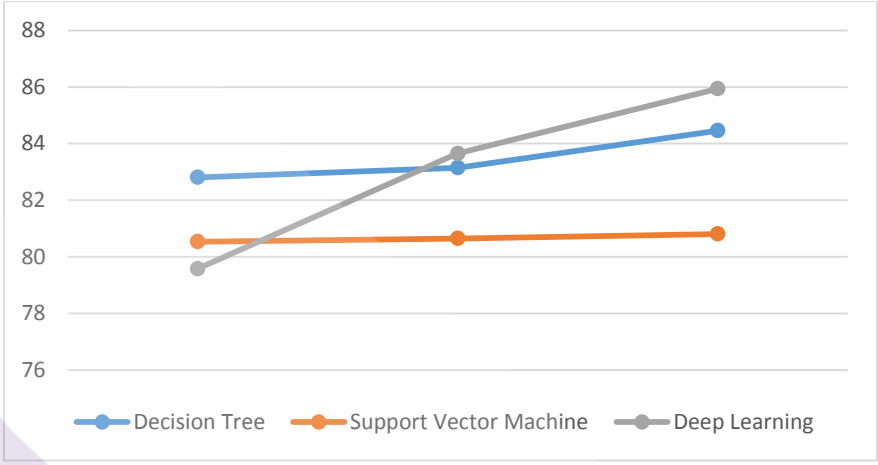
งานวิจัยนี้ได้นำเทคนิคการทำเหมืองข้อมูล เพื่อสร้างตัวแบบสำหรับการคัดกรองบุคคลที่มีแนวโน้มสัมฤทธิผลสำหรับการเรียนระดับปริญญาตรี สาขาวิชาเทคโนโลยีการศึกษา และเปรียบเทียบประสิทธิภาพของเทคนิคที่เหมะสมในการจำแนกข้อมูลโดยใช้เทคนิคการจำแนกข้อมูล Decision Tree, Support Vector Machine และ Deep Learning

5.1 สรุปผลการทดลอง

จากการดำเนินการวิจัยเพื่อสร้างตัวแบบสำหรับการคัดกรองบุคคลฯ สามารถจำแนกประเภทบุคคลได้จริงและได้ผลลัพธ์ความถูกต้องของการจำแนกตามเทคนิคที่ใช้งานดังนี้

ตารางที่ 5.1 เปรียบเทียบความถูกต้องของผลการทดสอบโมเดลด้วยเทคนิคที่ต่างกัน จำแนกตามสัดส่วนการ Train และนำข้อมูลเข้าแบบ Stepwise

เทคนิค	ร้อยละความถูกต้อง (Accuracy) N=738 คน		
	Train 70%	Train 75%	Train 80%
Decision Tree	82.81	83.15	84.46
Support Vector Machine	80.54	80.65	80.81
Deep Learning	79.58	83.65	85.94



ภาพที่ 5.1 เปรียบเทียบความถูกต้องของผลการทดสอบโมเดลด้วยเทคนิคที่ต่างกัน จำแนกตามสัดส่วนการ Train และนำข้อมูลเข้าแบบ Stepwise

จากตารางที่ 5.1 และภาพที่ 5.1 พบว่า ร้อยละความถูกต้องในเทคนิค Deep Learning มีความถูกต้องสูงกว่า Decision Tree และ Support Vector Machine ตามลำดับ มีค่าร้อยละความถูกต้องมากกว่า 80% สรุปได้ว่า เทคนิค Deep Learning สามารถจำแนกบุคคลได้ดีเทคนิค Decision Tree และ Support Vector Machine เมื่อมีการคัดกรอง Attribute แบบ Stepwise โดยเฉลี่ยทุกเทคนิคมีค่าร้อยละความถูกต้องร้อยละ 80 สามารถเลือกนำเทคนิคดังกล่าวไปใช้ในภาคสนามได้

ตารางที่ 5.2 ค่าเฉลี่ยร้อยละความถูกต้อง (Accuracy) จำแนกตามเทคนิคเหมืองข้อมูล

เทคนิค	ค่าเฉลี่ยร้อยละความถูกต้อง (Accuracy)
Decision Tree	83.47
Support Vector Machine	82.46
Deep Learning	83.06

ตารางที่ 5.2 พบว่า ค่าเฉลี่ยร้อยละความถูกต้อง (Accuracy) ของเทคนิค Decision Tree และเทคนิค Deep Learning มีค่าเฉลี่ยร้อยละความถูกต้องใกล้เคียงกัน คือ 83.47% และ 83.06% ตามลำดับ รองลงมาเป็นเทคนิค Support Vector Machine = 82.46%

ข้อค้นพบใหม่จากการวิจัยครั้งนี้

1. การดำเนินการทำเหมืองข้อมูลควรมีการใช้กระบวนการทางสถิติเพื่อการกรองหรือสกัดตัวแปรที่ไม่เกี่ยวข้องกับการสร้างตัวแบบโดยใช้การวิเคราะห์การถดถอย (Regression) แบบการถดถอยพหุคูณ (Multiple Regression) ผู้วิจัยเลือกใช้วิธีการคัดเลือกตัวแปรหรือ Attribute 2 แบบคือ แบบถอยหลัง (Backward Selection) และแบบลำดับขั้น (Stepwise Selection) เพื่อประหยัดเวลาสำหรับการเก็บรวบรวมข้อมูล และเพื่อประหยัดทรัพยากรด้านหน่วยความจำและการประมวลผลที่ไม่ต้องไปประมวลข้อมูลส่วนที่ไม่เกี่ยวข้อง

2. ประโยชน์สำหรับผู้บริหารหน่วยงาน คือ ไม่จำเป็นจะต้องพิจารณาตัวแปรมากมาย สามารถพิจารณาตัวแปรเพียง 2 ตัว ที่นำไปประกอบการตัดสินใจได้บนความถูกต้องที่ดีกว่าการพิจารณาตัวแปร 30 ตัว โดยการคัดกรองตัวแปรที่ไม่เกี่ยวข้องออกไปจากการวิเคราะห์

5.2 ปัญหาและข้อเสนอแนะ

1. ปัญหาที่พบในงานวิจัย

1.1 การเก็บข้อมูลที่จะนำมากำหนดเป็น Attribute อาจจะไม่สมบูรณ์เนื่องจากเป็นฐานข้อมูลของมหาวิทยาลัย อาจจะไม่ตรงกับความต้องการของผู้วิจัยมากนัก และเป็นข้อมูลที่ผู้เรียนได้สำเร็จการศึกษาจากมหาวิทยาลัยไปแล้วไม่สามารถติดตามผู้เรียน เพื่อเก็บข้อมูลบางประเด็นได้จึงใช้ข้อมูลที่มีอยู่ในระบบสำหรับการวิเคราะห์ข้อมูล

1.2 ความแตกต่างของข้อมูลในแต่ละปีการศึกษา เช่น ความแตกต่างด้านการเรียนรู้ของผู้เรียน (Learning Style) การให้เกรดแต่ละรายวิชา ผู้สอนแต่ละวิชามีความแตกต่างกัน ส่งผลทำให้คะแนนเฉลี่ยสะสมแต่ละปีการศึกษาแตกต่างกันไปด้วย จะส่งผลต่อการเรียนรู้หรือการ Trainning ของ Machine

1.3 ความแตกต่างของข้อมูลระหว่างผู้เรียนภาคปกติและภาคพิเศษทั้งด้านจำนวนของผู้เรียนและพฤติกรรมการเรียนรู้ที่แตกต่างกันส่งผลต่อการได้รับเกรดที่มีความแตกต่างกัน

2. ข้อเสนอแนะ

2.1 การวิจัยครั้งต่อไปควรมีการเก็บข้อมูลด้านทักษะของบุคคล (Soft Skills) รวมด้วยแต่จะต้องมีการวางแผนการเก็บข้อมูลผู้เรียนตั้งแต่ปีที่ 1 จนนิสิตสำเร็จการศึกษา

2.2 เพิ่มจำนวนกลุ่มตัวอย่างเพื่อให้มีประสิทธิภาพของ Machine Learning ดีมากยิ่งขึ้น

บรรณานุกรม

บรรณานุกรม

ภาษาไทย

[1] ฉัตรเกล้า เจริญผล. (2557). การใช้เทคนิคเหมืองข้อมูลในการเลือกกลุ่มสาขาวิชาที่เหมาะสมสำหรับการศึกษาต่อระดับปริญญาตรี. *วารสารวิทยาศาสตร์และเทคโนโลยีมหาวิทยาลัยมหาสารคาม*, *33*(6), 648-656.

[2] ชุติมา อุตมะมุณีย์ และประสงค์ ปราณีตพลกรัง. (2553). การพัฒนาตัวแบบระบบสนับสนุนการตัดสินใจ แบบอัตโนมัติออนไลน์สำหรับการเลือกสาขาวิชาเรียนของนักศึกษาระดับอุดมศึกษา. *Journal of Information Science and Technology*, *1*(2), 39-48.

[3] ณัฐริน เจริญเกียรติบวร. (2550). *การใช้เทคนิคเหมืองข้อมูลเพื่อช่วยในการแนะแนวการศึกษาต่อระดับอุดมศึกษา*. วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์. กรุงเทพฯ: มหาวิทยาลัยศิลปากร.

[4] เดช ธรรมศิริ และพยุง มีสัจ. (2554). การจำแนกข้อมูลด้วยวิธีแบบร่วมกันตัดสินใจจากพื้นฐานของเทคนิคต้นไม้ตัดสินใจ เทคนิคโครงข่ายประสาทเทียม และเทคนิคซัพพอร์ตเวกเตอร์แมชชีนร่วมกับการเลือกตัวแทนที่เหมาะสมด้วยขั้นตอนวิเชิงพันธุกรรม. *วารสารวิชาการพระจอมเกล้าพระนครเหนือ*, *21*(2), 293-303.

[5] ทรงศักดิ์ ภูสีอ่อน. (2556). *การประยุกต์ใช้ SPSS วิเคราะห์ข้อมูลงานวิจัย* (พิมพ์ครั้งที่ 6). มหาสารคาม: ตักสิลาการพิมพ์.

[6] บุญมา เพ่งซ่วน. (2548). *การใช้เทคนิคเหมืองข้อมูลเพื่อพัฒนาระบบสนับสนุนการตัดสินใจด้านการผลิตบัณฑิตระดับปริญญาตรี*. วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์. กรุงเทพฯ: มหาวิทยาลัยศิลปากร.

[7] บุญเสริม กิจศิริกุล. (2556). *เอกสารคำสอนวิชา 2110654 ปัญญาประดิษฐ์ (Artificial Intelligence)*. กรุงเทพฯ: ภาควิชาวิศวกรรมคอมพิวเตอร์ จุฬาลงกรณ์มหาวิทยาลัย.

[8] ปริญญา สงวนสัตย์. (2562). *Artificial Intelligence with Machine Learning, AI สร้างได้ด้วยแมชชีนเลิร์นนิ่ง*. นนทบุรี: ไอดีซี พรีเมียร์.

[9] พิศุทธ อ่อนเจริญ. (2561). การทำนายแนวโน้มของตลาดหุ้นด้วยแบบจำลองการเรียนรู้เชิงลึกที่เพิ่มประสิทธิภาพร่วมกับข้อมูลเชิงตัวเลขและข้อมูลเชิงตัวอักษร. วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาศาสตร์คอมพิวเตอร์. กรุงเทพฯ: จุฬาลงกรณ์มหาวิทยาลัย.

[10] ภาควิชานวัตกรรมและเทคโนโลยีการศึกษา. (2559). *มคอ.2 หลักสูตรการศึกษาบัณฑิต สาขาวิชาเทคโนโลยีการศึกษา หลักสูตรปรับปรุง พ.ศ. 2559*. ชลบุรี: คณะศึกษาศาสตร์ มหาวิทยาลัยบูรพา.

[11] รัชฎา เทพประสิทธิ์ และจรัญ แสนราช. (2563). การวิเคราะห์ปัจจัยที่มีผลต่อการเลือกสาขาวิชาของนักศึกษาระดับปริญญาตรี คณะครุศาสตร์ โดยใช้เทคนิคการทำเหมืองข้อมูล. *วารสารบัณฑิตศึกษา มหาวิทยาลัยราชภัฏวไลยอลงกรณ์ในพระบรมราชูปถัมภ์*, *14*(1), 134-146.

[12] วุฒิชัย กำจรกิตติคุณ และชัยพร เขมะภาตะพันธ์. (2559). การวิเคราะห์ความผิดปกติข้อมูลการใช้ไฟฟ้าด้วยนาอีฟเบย์. การประชุมวิชาการระดับประเทศด้านเทคโนโลยีสารสนเทศ (NCIT) ครั้งที่ 8, หน้า 403-408.

[13] สายชล สินสมบูรณ์ทอง. (2560). *การทำเหมืองข้อมูล เล่ม 1 : การค้นหาความรู้จากข้อมูล* (พิมพ์ครั้งที่ 2). กรุงเทพฯ: จามจุรีโปรดักส์.

[14] สุรพงค์ เอื้อวัฒนามงคล. (2561). *การทำเหมืองข้อมูล* (พิมพ์ครั้งที่ 2). กรุงเทพฯ: สำนักพิมพ์สถาบันบัณฑิตพัฒนบริหารศาสตร์.

[15] สุวิมล สิทธิชาติ. (2560). การวิเคราะห์คุณลักษณะพื้นฐานทางการศึกษาด้วยเทคนิคเหมืองข้อมูล. *วารสารเทคโนโลยีสารสนเทศ*, *13*(2), 20-28.

[16] เสกสรรค์ วิลัยลักษณ์, วิภา เจริญภัณฑารักษ์ และดวงดาว วิชาดากุล. (2558). การใช้เทคนิคการทำเหมืองข้อมูลเพื่อพยากรณ์ผลการเรียนของนักเรียน โรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน ศูนย์วิจัยและพัฒนาการศึกษา. *Veridian E-Journal, Science and Technology Silpakorn University, 2*(2), 1-17.

[17] อนันต์ ปินะเต. (2557). *การใช้เทคนิคเหมืองข้อมูลในการเลือกกลุ่มสาขาวิชาที่เหมาะสมสำหรับการศึกษาต่อระดับปริญญาตรี*. วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ. มหาสารคาม: มหาวิทยาลัยมหาสารคาม.

[18] อรทิพย์ เลื่อยงาม และชัยพร เขมะภาตะพันธ์. (2554). การจัดประเภทเอกสารด้วยวิธีเอสวีเอ็มเพื่อการป้องกันเอกสารรั่วไหล. Proceedings การประชุมวิชาการ “นเรศวรวิจัย” ครั้งที่ 7 กลุ่มวิทยาศาสตร์และเทคโนโลยี, หน้า 3-12.

[19] อังคณา พิจารโชติ. (2552). *ระบบสนับสนุนการตัดสินใจสำหรับวิเคราะห์ปัจจัยเสี่ยงการเป็นโรคเบาหวาน โดยใช้เทคนิคดาต้าไมน์นิ่ง*. การศึกษาอิสระปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ. ขอนแก่น: มหาวิทยาลัยขอนแก่น.

[20] อัจจิมา มณฑาพันธุ์. (2562). *รายงานการวิจัยเรื่อง การเปรียบเทียบวิธีการคัดเลือกคุณลักษณะที่สำคัญในการปรับปรุงการพยากรณ์มะเร็งเต้านม*. มหาวิทยาลัยศรีปทุม: กรุงเทพฯ.

[21] เอกสิทธิ์ พัชรวงศ์ศักดา. (2563). *Practical Data Mining with RapidMiner Studio 9*. กรุงเทพฯ: เอเซีย ดิจิตอลการพิมพ์.

ภาษาต่างประเทศ

[22] Ferda, U. (2019). Data Mining for Student Performance Prediction in Education. *IntechOpen*. DOI: http://dx.doi.org/10.5772/intechopen.91449 Available from: https: / / www. intechopen. com/ books/ data-mining-methods-applications-and-systems/data-mining-for-student-performance-prediction-in-education

[23] Hanan, M. (2020). Using Data Mining Techniques to Predict Student Performance to Support Decision Making in University Admission Systems. *IEEE Access. 8* DOI: 10.1109/ ACCESS.2020.2981905 Available from: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9042216

[24] Khasanah, A.U., and Harwati. (2017). A Comparative Study to Predict Student's Performance Using Educational Data Mining Techniques. *Materials Science and Engineering.* doi: 10. 1088/ 1757-899X/ 215/ 1/ 012036. Available from: https://iopscience.iop.org/article/10.1088/1757-899X/215/1/012036/pdf

[25] RapidMiner. (2021). *About RapidMiner.* Available from: https://rapidminer.com/us/

[26] Tan, P.N., Steinbach, M., and Kumer, V. (2005). *Introduction to Data Mining*. Pearson.

[27] Witten, Ian H., Frank, Eibe, H., and Mark, A. (2011). *Data Mining: Practical Machine Learning Tools and Techniques* (3 ed.), Elsevier.

ประวัติผู้วิจัย

ชื่อ-นามสกุล	นายธนะวัฒน์ วรรณประภา
ประวัติการศึกษา	- การศึกษาบัณฑิต สาขาวิชาเทคโนโลยีทางการศึกษา มหาวิทยาลัยบูรพา พ.ศ. 2545 - การศึกษามหาบัณฑิต สาขาวิชาเทคโนโลยีทางการศึกษา มหาวิทยาลัยบูรพา พ.ศ. 2547 - ปรัชญาดุษฎีบัณฑิต สาขาวิชาเทคโนโลยีการศึกษา มหาวิทยาลัยบูรพา พ.ศ. 2558 - วิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมคอมพิวเตอร์และโทรคมนาคม มหาวิทยาลัยธุรกิจบัณฑิตย์ พ.ศ. 2564
ตำแหน่งงานปัจจุบัน	อาจารย์ประจำภาควิชานวัตกรรมและเทคโนโลยีการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยบูรพา