การประยุกต์ใช้ Machine Learning ทำนายผลการเรียนวิชา Web Database ของนิสิตสาขาเทคโนโลยีการศึกษา คณะศึกษาศาตร์ มหาวิทยาลัยบูรพา

วีระพันธ์ พานิชย์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมคอมพิวเตอร์และโทรคมนาคม
วิทยาลัยนวัตกรรมด้านเทคโนโลยีและวิศวกรรมศาสตร์
มหาวิทยาลัยธุรกิจบัณฑิตย์
พ.ศ. 2564

APPLY OF MACHINE LEARNING TO GRADE PREDICTION IN WEB DATABASE COURSE OF EDUCATIONAL TECHNOLOGY STUDENTS AT FACULTY OF EDUCATION BURAPHA UNIVERSITY

WEERAPUN PANICH

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 การประยุกต์ใช้ Machine Learning ทำนายผลการเรียนวิชา Web Database ของนิสิตสาขาเทคโนโลยีการศึกษา คณะศึกษาศาตร์ มหาวิทยาลัยบูรพา

เสนอโดย นายวีระพันธ์ พานิชย์

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 2 เดือน ก.ค. พ.ศ. 2564

หัวข้อวิทยานิพนธ์	การประยุกต์ใช้ Machine Learning ทำนายผลการเรียนวิชา Web Database ของนิสิตสาขาเทคโนโลยีการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยบูรพา
ชื่อผู้เขียน	วีระพันธ์ พานิชย์
อาจารย์ที่ปรึกษา	ดร.ชัยพร เขมะภาตะพันธ์
สาขาวิชา	วิศวกรรมคอมพิวเตอร์และโทรคมนาคม
ปีการศึกษา	2563

บทคัดย่อ

การวิจัยครั้งนี้นำเสนอโมเดลการทำนายผลการเรียนวิชา Web Database ที่มีประสิทธิภาพดีที่สุด เพื่อนำโมเดลไปทำนายผลการเรียนวิชา Web Database โดยผลการทำนายจะเป็นข้อมูลที่สำคัญในการวางแผนจัดการเรียนการสอนให้เหมาะสมกับกลุ่มผู้เรียนที่มีความรู้ทักษะ ความสามารถพื้นฐานด้านคอมพิวเตอร์แตกต่างกัน ผู้วิจัยดำเนินการสังเคราะห์ Attribute จากเอกสารงานวิจัยที่เกี่ยวข้องและคัดกรอง Attribute ด้วยการวิเคราะห์ความถดถอยพหุคูณ (Multiple Regression) แบบลำดับขั้น (Stepwise Selection) ได้ Attribute จำนวน 6 Attribute ประกอบด้วย 1) Sub_soc (เกรดเฉลี่ยกลุ่มวิชาสังคม) 2)PHP (เคยเรียนภาษา PHP) 3)Computer (คอมพิวเตอร์ส่วนตัว) 4)Sub_create (เกรดเฉลี่ยกลุ่มวิชาความสร้างสรรค์) 5)Identity (เกรดเฉลี่ยกลุ่มวิชาอัตลักษณ์มหาวิทยาลัย) และ 6)Attitude (ความรู้สึกต่อการเรียนวิชาด้านคอมพิวเตอร์) ทำการทดสอบและเปรียบเทียบประสิทธิภาพโมเดลระหว่างวิธี Decision Tree, Naïve Bayes , Neural Network และSupport Vector Machine โดยใช้โปรแกรม RapidMiner Studio ข้อมูลนำเข้าโมเดลจำนวน 271 คน แบ่งเป็นกลุ่มเรียนรู้ (Train Set) 80 % และกลุ่มทดสอบ (Test set) 20 % ผลการวิจัยพบว่าประสิทธิภาพทำนายด้วยวิธี Neural Network มีค่าความถูกต้อง (Accuracy) มากที่สุด 88.89 %รองลงมาวิธี Support Vector Machine 85.19 % ลำดับที่สามวิธี Decision Tree 79.63 % ลำดับสุดท้ายวิธี Naïve Bayes 74.07 % และ ผลการปรับค่าพารามิเตอร์ของโมเดล Neural Network ที่เหมาะสมที่สุด คือ Training Cycle = 200, Hidden layer = 1, Hidden layer sizes = 22, Learning rate = 0.4 และ Momentum = 0.3 ซึ่งส่งผลให้ประสิทธิภาพการทำนายของโมเดล Neural Network เพิ่มขึ้น จากค่าความถูกต้อง (Accuracy) 88.89% เป็น 92.59 % (เพิ่มขึ้น 3.7 %)

คำสำคัญ: ประสิทธิภาพการทำนาย, Decision Tree, Naïve Bayes, Neural Network, Support Vector Machine

Thesis Title	APPLE OF MACHINE LEARNING TO GRADE PREDICTION IN WEB DATABASE COURSE OF EDUCATIONAL TECHNOLOGY STUDENTS AT FACULLTY OF EDUCATION BURAPHA UNIVERSITY
Author	Weerapun Panich
Thesis Advisor	Dr. Chaiyaporn Khemapatapan
Department	Computer and Telecommunications Engineering
Academic Year	2020

ABSTRACT

This research will present the prediction models for Web Database course which has the best efficiency. The predicted results will be used to make a learning management strategies that suitable to each groups of students that has various knowledge, skills as well as basic computer skills. We synthesis all attributes from many related research articles. The attribute selection will be done by the multiple regression with the stepwise selection. There are 6 attributes such that 1) Sub_soc (GPAX for social studies) 2) PHP programming 3) own their PC 4) Sub_create (GPAX of creativity subjects) 5) GPAX of Burapha University Identity subjects and 6) the attitude to computer related subjects. The predicted models are Decision Tree, Naïve Bayes , Neural Network and Support Vector Machine via the RapidMiner Studio with 271 total records. We set 80% of the record as the training set and 20% for the test set. The best efficiency model is the Neural Network which has highest accuracy score about 88.89%, followed by the Support Vector Machine has 85.19%, Decision Tree gives 79.63 % and Naïve Bayes with 74.0 %. The suitable parameters for the Neural Network are set as the Training Cycle is 200, Hidden layer is 1, Hidden layer sizes are 22, Learning rate is 0.4 and Momentum is 0.3. This gives more higher accuracy from 88.89% to 92.59% or 3.7% increases.

Keywords: Performance prediction, Decision Tree, Naive Bayes , Neural network, Support Vector Machine

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จได้ด้วยความกรุณาจากท่านอาจารย์ ดร.ชัยพร เขมะภาตะพันธ์ อาจารย์ ที่ปรึกษาวิทยานิพนธ์ ที่ได้กรุณาแนะนำแก้ไขข้อบกพร่องให้คำปรึกษา ให้ความช่วยเหลือตลอดระยะเวลาในการทำวิทยานิพนธ์ ผู้วิจัยขอกราบขอบพระคุณเป็นอย่างสูงไว้ ณ โอกาสนี้

ผู้วิจัยขอกราบขอบพระคุณ รองศาสตราจารย์ ดร.ลัญฉกร วุฒิสิทธิกุลกิจ ที่กรุณาให้เกียรติเป็นประธานสอบวิทยานิพนธ์ โดยมีผู้ช่วยศาตราจารย์ ดร.มัชฌิกา อ่องแตง และ ดร.ธนัญ จารุวิทยโกวิท เป็นกรรมการในการสอบวิทยานิพนธ์ ซึ่งได้กรุณาให้ข้อเสนอแนะการทำวิทยานิพนธ์ฉบับนี้ให้ถูกต้องสมบูรณ์ยิ่งขึ้น และขอขอบคุณเจ้าหน้าที่มหาวิทยาลัยธุรกิจบัณฑิตย์ทุกท่านที่ให้ความสะดวก และประสานงานในการทำวิทยานิพนธ์ให้ผู้วิจัยตลอดมา ทำให้การจัดทำวิทยานิพนธ์ของผู้วิจัยครั้งนี้สำเร็จลุล่วงไปด้วยดี

ท้ายนี้ ขอกราบขอบพระคุณบิดา มารดา และครู อาจารย์ที่ให้ความรัก ความเมตตาให้การสั่งสอนอบรม และเป็นกำลังใจแก่ผู้วิจัยด้วยดีตลอดมา ทำให้ผู้วิจัยทำวิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงด้วยดี จึงขอกราบขอบพระคุณเป็นอย่างสูงไว้ ณ โอกาสนี้

วีระพันธ์ พานิชย์

สารบัญ

หน้า

บทคัดย่อภาษาไทย…….. ค

บทคัดย่อภาษาอังกฤษ……..…. ง

กิตติกรรมประกาศ.....…….…. จ

สารบัญ…….…… ฉ

สารบัญตาราง…… ซ

สารบัญภาพ…….. ฌ

บทที่

1 บทนำ……..…….…… 1

1.1 ความเป็นมาของปัญหาวิจัย……..……..... 1

1.2 วัตถุประสงค์การวิจัย……..……... 4

1.3 ขอบเขตการวิจัย.....……..……... 4

1.4 ประโยชน์ของงานวิจัย.....……..……... 5

1.5 นิยามศัพท์เฉพาะ.....……..……... 6

1.6 แผนการดำเนินงานวิจัย.....……..……... 6

1.7 การตอบรับการตีพิมพ์.....……..……... 7

2 เอกสารและงานวิจัยที่เกี่ยวข้อง……..…… 8

2.1 ทักษะพื้นฐานการเขียนโปรแกรมคอมพิวเตอร์……..…… 8

2.2 การทำเหมืองข้อมูล (Data Mining)....……..……... 17

2.3 Decision Tree....……..…….. 22

2.4 Naïve Bayes....……..…….. 24

2.5 Support Vector Machine……..…….. 25

2.6. Neural Network .....……..…….. 28

2.7. การวิเคราะห์ความถดถอย (Regression Analysis).....…….. 33

2.8 โปรแกรม Rapid Miner Studio.....…….. 33

2.9 งานวิจัยที่เกี่ยวข้อง .....…….. 37

สารบัญ (ต่อ)

บทที่	หน้า
3 วิธีการดำเนินการวิจัย	41
3.1 ขั้นตอนที่ 1 การเลือกข้อมูล (Selection)	42
3.2 ขั้นตอนที่ 2 การเตรียมข้อมูล (Preprocessing)	43
3.3 ขั้นตอนที่ 3 การเปลี่ยนรูปข้อมูล (Transformation)	44
3.4 ขั้นตอนที่ 4 การทำเหมืองข้อมูล (Data Mining)	52
3.5 ขั้นตอนที่ 5 การแปรผลและการประเมินผล (Interpretation Evaluation)	61
4 ผลการวิจัย	65
4.1 ตอนที่ 1 ผลการวิเคราะห์วิเคราะห์ประสิทธิภาพโมเดลโดยใช้ข้อมูลชุดที่ 1	66
4.2 ตอนที่ 2 ผลการวิเคราะห์วิเคราะห์ประสิทธิภาพโมเดลโดยใช้ข้อมูลชุดที่ 2	67
4.3 ตอนที่ 3 การเปรียบผลการวิเคราะห์ประสิทธิภาพโมเดลด้วยข้อมูลชุดที่ 1 กับข้อมูลชุดที่ 2	69
4.4 ตอนที่ 4 การปรับค่าพารามิเตอร์ Neural Network และการเปรียบเทียบประสิทธิภาพการทำนาย	70
4.5 ตอนที่ 5 สรุปผลการวิเคราะห์ และขั้นตอนการนำโมเดลที่มีประสิทธิภาพมากที่สุดไปใช้	76
5 สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ	81
5.1 สรุปผลวิจัย	81
5.2 อภิปรายผล	82
5.3 ข้อเสนอแนะ	83
บรรณานุกรม	85
ประวัติผู้เขียน	91

สารบัญตาราง

ตารางที่ **หน้า**

1.1 แผนดำเนินงานวิจัย..... 6

3.1 แสดงขั้นตอนการดำเนินการวิจัย..... 41

3.2 ตัวอย่างข้อมูลจากแบบสอบถามและจากฐานข้อมูล..... 43

3.3 รูปแบบข้อมูล และการกำหนด Key code ข้อมูล..... 44

3.4 แสดงตัวอย่างรูปแบบข้อมูล และ Code ข้อมูล..... 46

3.5 ผลการวิเคราะห์ความสัมพันธ์ของ Attribute และ Label 49

3.6 แสดงค่าสัมประสิทธิ์การถดถอยในแบบคะแนนมาตรฐาน..... 50

3.7 แสดงชนิดข้อมูลของแต่ละ Attribute 54

4.1 การเปรียบเทียบประสิทธิภาพความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision)และ ค่าความระลึก (Recall) ชุดข้อมูลที่ 1 จำนวน 22 Attribute..... 66

4.2 การเปรียบเทียบประสิทธิภาพความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision)และ ค่าความระลึก (Recall) ชุดข้อมูลที่ 2 จำนวน 6 Attribute..... 67

4.3 ค่าความถูกต้อง (Accuracy) การวิเคราะห์ด้วยข้อมูลชุดที่ 1 และข้อมูลชุดที่ 2..... 69

4.4 ผลการทดลองปรับค่า Training Cycle ของโมเดล Neural Network..... 71

4.5 ผลการทดลองปรับจำนวน Hidden layer ของโมเดล Neural Network..... 72

4.6 ผลการทดลองปรับจำนวน Hidden layer sizes ของโมเดล Neural Network..... 73

4.7 ผลการทดลองปรับค่า Learning rate ของโมเดล Neural Network..... 74

4.8 ผลการทดลองปรับค่า Momentum ของโมเดล Neural Network..... 75

4.9 การทำ Key code ข้อมูลสำหรับทำนายผลการเรียน..... 77

สารบัญภาพ

ภาพที่ **หน้า**

2.1 แสดงการทำงานของภาษา PHP และ MySQL 15

2.2 แสดงกระบวนการการค้นหาความรู้ในการทำเหมืองข้อมูล..... 17

2.3 แสดงส่วนประกอบของต้นไม้ตัดสินใจ..... 22

2.4 ตัวอย่างต้นไม้ตัดสินใจ 23

2.5 เส้นตรงที่ใช้แบ่งกลุ่มข้อมูล (Hyper plane)..... 26

2.6 เปรียบเทียบค่า Margin เส้นสีส้มมีระยะมากที่สุด..... 27

2.7 ตัวอย่างการสร้าง Kernels จาก 2D เป็น 3D..... 27

2.8 โครงสร้างการทำงานของโครงข่ายประสาทเทียม..... 28

2.9 โครงสร้างโครงข่ายประสาทเทียมแบบชั้นเดียว..... 30

2.10 โครงสร้างโครงข่ายประสาทเทียมแบบหลายชั้น..... 30

2.11 โครงสร้างโครงข่ายประสาทเทียมอย่างง่าย..... 32

2.12 เว็บไซต์ Rapid Miner Studio และหน้าเวปไซด์ ดาวโหลดโปรแกรม..... 36

2.13 โปรแกรม Rapid Miner Studio 36

3.1 แสดงการใช้เมนู “Analyze” เมนูรอง “Regression” และเมนูย่อย “Linear”..... 47

3.2 การกำหนดตัวแปรตาม(Dependent)ตัวแปรต้น (Independent) การคัดเลือกแบบวิเคราะห์ความถดถอยพหุคูณ Stepwise Selection 47

3.3 การเลือกสถิติในการวิเคราะห์ข้อมูล..... 48

3.4 การตั้งค่าเกณฑ์ในการนำเข้าตัวแปรทำนายหรือขจัดตัวแปรทำนาย..... 48

3.5 โปรแกรม Rapid Miner Studio และ เมนู Import data..... 53

3.6 การตั้งค่าชนิดข้อมูลแต่ละ Attribute 53

3.7 แสดงชนิดข้อมูลของแต่ละ Attribute 55

3.8 การสร้างโมเดลแบบ Multiply 55

3.9 การตั้งค่าพารามิเตอร์ Operator Validation 56

3.10 การเปิด Operator Validation 56

3.11 โมเดล Decision Tree..... 57

3.12 การตั้งค่าพารามิเตอร์ Operator Deception tree , Operator Performance..... 57

สารบัญภาพ (ต่อ)

ภาพที่ **หน้า**

3.13 การ Run โมเดล Deception tree..... 58

3.14 โมเดล Naïve Bayes 58

3.15 ค่าพารามิเตอร์ Operator Performance โมเดล Naïve Bayes ……..……..... 59

3.16 โมเดล Support Vector Machine …..……..... 59

3.17 การตั้งค่าพารามิเตอร์ Operator Support Vector Machine, Operator Performance..... 60

3.18 โมเดล Neural Network 60

3.19 การตั้งค่าพารามิเตอร์ Neural Network..... 61

3.20 แสดงผลค่าความถูกต้อง (Accuracy) 62

3.21 แสดงผลค่าความแม่นยำ (Precision) เฉลี่ย..... 62

3.22 แสดงผลค่าความระลึก (Recall) เฉลี่ย..... 63

4.1 การเปรียบเทียบวิธี Naïve Bayes , Support Vector Machine, Neural Network และ Decision Tree โมเดล 22 Attribute ……..……..... 66

4.2 การเปรียบเทียบวิธี Naïve Bayes , Support Vector Machine, Neural Network และ Decision Tree โมเดล 6 Attribute ……..….....…… 68

4.3 การเปรียบเทียบค่าความถูกต้อง (Accuracy) โมเดล ข้อมูลชุดที่ 1 และชุดที่ 2..... 69

4.4 ประสิทธิภาพการทำนาย โมเดล Neural Network…….....…. 76

4.5 โมเดลทำนายผลการเรียนด้วยวิธี Neural Network ที่มีค่าความถูกต้องมากที่สุด... 78

4.6 การสร้างโมเดลการทำนายผลการเรียนด้วยวิธี Neural Network…….. 78

4.7 การนำข้อมูลใหม่เข้าโมเดลและการปรับโมเดลเพื่อการทำนายผลการเรียน…..... 79

4.8 ผลการทำนายผลการเรียนและค่าความเชื่อมั่นการจำแนก ด้วยวิธี Neural Network…… 80

บทที่ 1
บทนำ

1.1 ความเป็นมาของปัญหาวิจัย

ปัจจุบันการนำคอมพิวเตอร์มาใช้ในการจัดการเกี่ยวกับฐานข้อมูล (Database) มีความสำคัญต่อการดำเนินงานขององค์กร โดยเฉพาะอย่างยิ่งในองค์กรที่มีขนาดใหญ่ เช่น องค์กรทางธุรกิจงานทางด้านวิศวกรรมการแพทย์การกฎหมายหรือการศึกษา ทั้งนี้เนื่องจากระบบฐานข้อมูลมีกระบวนการจัดเก็บ ประมวลผลและการค้นหาข้อมูลได้รวดเร็ว ถูกต้องแม่นยำ ทำให้การดำเนินการขององค์กรมีประสิทธิภาพขึ้น ระบบฐานข้อมูล (Database System) คือ การจัดเก็บข้อมูลอย่างเป็นระบบ ซึ่งผู้ใช้สามารถเรียกใช้ข้อมูลในลักษณะต่าง ๆ เช่น การเพิ่มข้อมูล (Add Data) การแทรกข้อมูล (Insert Data) การเรียกใช้ข้อมูล (Retrieve Data) การแก้ไขและลบข้อมูล (Update & Delete Data) ตลอดจนการเคลื่อนย้ายข้อมูล (Move Data) ไปตามที่กำหนดสามารถนำไปประมวลผลต่อยอดเป็นข้อมูลสถิติที่สำคัญให้แก่องค์กร

ภาควิชานวัตกรรมและเทคโนโลยีการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยบูรพาได้ตระหนักถึงความสำคัญของระบบฐานข้อมูลในการใช้งานองค์กรต่าง ๆ จึงได้เปิดการสอนรายวิชา Web Database ให้กับนิสิตระดับปริญญาตรีสาขาเทคโนโลยีการศึกษา โดยจัดให้นิสิตเรียนในชั้นปีที่ 3 วัตถุประสงค์ของรายวิชาเพื่อให้นิสิตมีความรู้และทักษะในด้านโครงสร้างฐานข้อมูล การวิเคราะห์ การออกแบบฐานข้อมูล วิธีการจัดการข้อมูลบนเว็บ การประมวลผล และแสดงผลข้อมูล การเขียนโปรแกรมคอมพิวเตอร์ด้วยภาษา PHP และ MySQL เพื่อสร้างฐานข้อมูล และการจัดการฐานข้อมูล การเรียนในรายวิชา Web Database อาจารย์ประจำวิชาดำเนินการสอนเนื้อหาเชิงทฤษฎีและการปฏิบัติจริงการเพื่อให้นิสิตเกิดความรู้ที่สามารถออกแบบและจัดการงานระบบฐานข้อมูลได้สมบูรณ์ โดยนิสิตต้องลงมือปฏิบัติการเขียนโปรแกรมที่เกี่ยวข้องกับระบบจัดการฐานข้อมูลด้วยภาษา PHP และ MySQL ตามแนวทางในการเขียนโปรแกรม คอมพิวเตอร์ซึ่งจะมีหลักเกณฑ์ในการเขียนและการออกแบบโปรแกรมโดยแบ่งเป็น 7 ขั้นตอน ประกอบด้วย 1) ขั้นตอนการวิเคราะห์ปัญหา (Analysis the problem) 2) ขั้นตอนการออกแบบโปรแกรม (Design a program) 3) ขั้นตอนการเขียนโปรแกรม (Coding) 4) ขั้นตอนการตรวจสอบข้อผิดพลาดของโปรแกรม (Testing and Debugging) 5) ขั้นตอนการทดสอบความถูกต้องของโปรแกรม (Testing

and Validating) 6) ขั้นตอนการ ทำเอกสารประกอบโปรแกรม (Documentation) และ 7) ขั้นตอนการบำรุงรักษาโปรแกรม (Program maintenance) และจากการศึกษาปัจจัยที่ส่งผลต่อการประสบผลสำเร็จ ทางด้านการเรียนคอมพิวเตอร์ พบว่า ผลการเรียน (Grade) [1][2] เป็นปัจจัยที่สำคัญที่บ่งบอกถึงประสิทธิภาพการเรียนรู้เป็นอย่างไร 2) ประสบการณ์การเขียนโปรแกรม (Previous computer programming experiences) ส่งผลให้เกิดความสำเร็จในการเรียนรู้ 3) พื้นฐานทางคณิตศาสตร์ (Mathematics background) [3][4][5] ผู้ที่มีความรู้พื้นฐานคณิตศาสตร์ดีจะมีทักษะในการคิด และเข้าใจเรื่องของโครงสร้างและแนวทาง การดำเนินงานซึ่งเป็นพื้นฐานสำคัญของการเขียนโปรแกรมที่ดี จากการศึกษาปัจจัยพื้นฐานของการเรียนวิชาด้านคอมพิวเตอร์ที่ส่งผลต่อผลสัมฤทธิ์ทางการเรียนดังกล่าว พบว่าถ้ามีการรวบรวมข้อมูลพื้นฐานของผู้เรียน ประวัติการศึกษาผลการเรียนในกลุ่มวิชาต่าง ๆ สามารถนำข้อมูลมาวิเคราะห์เพื่อทำนายผลการเรียนที่จะเกิดขึ้นในอนาคต และผลการวิเคราะห์สามารถนำมาเป็นข้อมูลในการออกแบบการเรียนการสอนให้มีประสิทธิภาพต่อไปได้

การนำกระบวนการ Data Mining โดยใช้เทคนิค Machine Learning มาช่วยในทำนายเพื่อการวางแผนให้ตรงกับความต้องการใช้งาน เป็นการนำความรู้จากข้อมูลที่มีอยู่เพื่อประโยชน์ในการทำนายข้อมูลใหม่ที่จะเกิดขึ้นในอนาคต ซึ่งการทำเหมืองข้อมูลคือการกระทำกับข้อมูลจำนวนมากเพื่อค้นหารูปแบบและความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น ในปัจจุบันการทำเหมืองข้อมูลได้ถูกนำมาไปประยุกต์ใช้ในงานหลายประเภท ทั้งในด้านธุรกิจที่ช่วยในการตัดสินใจของผู้บริหาร ในด้านวิทยาศาสตร์และการแพทย์รวมทั้งในด้านเศรษฐกิจและสังคม ดังเช่นงานวิจัยเรื่องการวิเคราะห์ความผิดปกติข้อมูลการใช้ไฟฟ้าด้วยนาอีฟเบย์ [6] ได้ให้ข้อสรุปว่า การใช้นาอีฟเบย์มาช่วยในการวิเคราะห์ความผิดปกติจากข้อมูลการใช้ไฟฟ้านั้น สามารถลดค่าใช้จ่ายในการเดินทางและค่าแรงปฏิบัติงานของเจ้าหน้าที่ลงได้ นอกจากความผิดปกติประเภทกำลังไฟฟ้าไม่สมดุลที่ได้ทดลองแล้วนั้น ยังสามารถใช้แนวคิดนี้สำหรับแก้ปัญหาความผิดปกติในเรื่องอื่น ๆ ซึ่งหากกำหนดเงื่อนไข (Attributes) ที่สอดคล้องและมีผลกระทบต่อการวิเคราะห์แล้ว สามารถสร้างเป็นระบบตรวจจับความผิดปกติจากข้อมูลการใช้ไฟฟ้า อาจทราบถึงพฤติกรรมการละเมิดการใช้ไฟฟ้าได้อีกด้วย และงานวิจัยเรื่อง ประสิทธิภาพการจำแนกข้อมูลการเลือกอาชีพโดยอัตโนมัติด้วยเทคนิคเหมืองข้อมูล [7] วัตถุประสงค์ เพื่อเปรียบเทียบประสิทธิภาพในการจำแนกข้อมูลการเลือกอาชีพของนิสิตระดับปริญญาตรีหลังสำเร็จการศึกษา โดยใช้วิธี Decision Tree , Neural Network และ Naïve Bayesian Learning ผลจากการศึกษาพบว่าประสิทธิภาพในการจำแนกข้อมูลแบบ Decision Tree มีประสิทธิภาพในการจำแนกสูงสุดด้วยค่าเฉลี่ย 80.62% และปัจจัยสำคัญที่ทำให้การเลือกอาชีพ ตรงหรือไม่ตรงกับสาขา มี 4 ปัจจัย คือ สาขาวิชาที่เรียน เกรดเฉลี่ยเฉพาะวิชา สาขา เพศ และ

เกรดเฉลี่ยรวม ซึ่งผลการทดลองนี้สามารถนำไปประยุกต์ใช้กับคณะหรือหน่วยงานที่เกี่ยวข้อง เพื่อวางแผนพัฒนาโครงสร้างหลักสูตรหรือวางแผนการศึกษาให้กับนิสิตได้ และงานวิจัยเรื่องการประเมินความแม่นยำของเครือข่ายแบบเบย์สำหรับการตรวจจับรูปแบบการเรียนรู้ของนักเรียน [8] ได้ทำการวัดประสิทธิภาพความแม่นยำของข่ายงานเบย์ (Bayesian Network) ในการตรวจสอบรูปแบบการเรียนของนักศึกษาสาขาวิชาวิศวกรรมคอมพิวเตอร์ในรายวิชาปัญญาประดิษฐ์โดยทำการเปรียบเทียบรูปแบบการเรียนที่ได้จากข่ายงานเบย์ (Bayesian Network) กับแบบสอบถาม โดยที่มีการแบ่งรูปแบบการเรียนออกเป็น 3 ประเภท คือการรับรู้ (Perception) การประมวลผล (Processing) และความเข้าใจ (Understanding) ผลการทดลองพบว่าประสิทธิภาพของข่ายงานเบย์ (Bayesian Network) ให้ความถูกต้องของผลการทดลองอยู่ในระดับที่สูง

งานวิจัยนี้ผู้วิจัยได้ใช้กระบวนการทำเหมืองข้อมูล (Data Mining) โดยใช้เทคนิค Machine Learning เพื่อทำนายผลการเรียนวิชา Web Database โดยผลการเรียนแบ่งออกเป็น 2 กลุ่ม คือ กลุ่มเกรดพอใช้ ประกอบด้วย เกรด D, D+, C, C+ และ กลุ่มค่าเกรดดี ประกอบด้วย เกรด B, B+, A ข้อมูลที่นำมาวิเคราะห์ได้จากแบบสอบถามร่วมกับข้อมูลจากระบบฐานข้อมูลผลการเรียนของนิสิตสาขาเทคโนโลยีการศึกษาที่ผ่านการเรียนวิชา Web Database ปีการศึกษา 2561 -2563 จำนวน 271 ระเบียน (Record) นำมาผ่านการคัดเลือกตัวแปรที่มีอิทธิพลต่อผลการเรียนวิชา Web Database ด้วยวิธีการวิเคราะห์การถดถอยพหุคูณ (Multiple Regression) แบบการเลือกตัวแปรอิสระแบบขั้นตอน (Stepwise Regression) จากนั้นผู้วิจัยนำตัวแปรที่คัดเลือกแล้วเข้าสู่กระบวนการทำเหมืองข้อมูลด้วยโปรแกรม Rapid Miner Studio โดยสร้างโมเดล 4 วิธีการ คือ Decision Tree, Naïve Bayes , Neural network และ Support Vector Machine และนำผลการวิเคราะห์ 4 วิธีการมาเปรียบเทียบค่าความถูกต้อง (Accuracy) ผลของการเปรียบเทียบประสิทธิภาพการทำนายในงานวิจัยครั้งนี้จะได้โมเดลการทำนายผลที่ดีที่สุด สามารถนำไปใช้ในทำนายผลการเรียนวิชา Web Database ของนิสิตกลุ่มใหม่ที่ลงทะเบียนเรียนวิชา Web Database ได้ ทั้งนี้ผลการทำนายจะเป็นข้อมูลให้อาจารย์ที่ปรึกษาหรืออาจารย์ผู้สอนได้ใช้ประโยชน์ในการวางแผนจัดการเรียนการสอนต่อนิสิตเป็นรายบุคคล และต่อกลุ่มเรียนได้ตรงกลุ่มเป้าหมายมากขึ้น อันจะส่งผลให้นิสิตมีทักษะในการเขียนโปรแกรมคอมพิวเตอร์และสร้างฐานข้อมูลด้วยภาษา PHP และ MySQL ได้ตามวัตถุประสงค์การเรียนรู้ และเป็นการลดความเสี่ยงของนิสิตที่จะมีผลการเรียนที่ไม่ผ่านเกณฑ์

1.2 วัตถุประสงค์การวิจัย

เพื่อเปรียบเทียบประสิทธิภาพของโมเดลการทำนายผลการเรียนวิชา Web Database ระหว่างวิธี Decision Tree, Naïve Bayes, Neural Network และ Support Vector Machine

1.3 ขอบเขตการวิจัย

1.3.1 ขอบเขตด้านข้อมูล

ข้อมูลที่นำมาสร้างเป็น Attribute ใช้ข้อมูลของนิสิตสาขาวิชาเทคโนโลยีการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยบูรพา ที่ผ่านการเรียนวิชา Web Database ปีการศึกษา 2561 -2563 จำนวน 271 ระเบียน ประกอบด้วย Attribute ต่อไปนี้

1) เพศ
2) เคยเรียนภาษา HTML 3
3) เคยเรียนภาษา C หรือ C++ 4
4) เคยเรียนภาษา PHP
5) เกรดเฉลี่ยวิชาคณิตศาสตร์ ม.6
6) เกรดเฉลี่ยจบการศึกษาระดับชั้น ม. 6
7) แผนการเรียนมัธยมปลาย
8) ประเภทการรับเข้ามหาวิทยาลัย
9) ความรู้สึกต่อการเรียนวิชาด้านคอมพิวเตอร์
10) คอมพิวเตอร์ส่วนตัว
11) การทบทวนบทเรียน
12) เกรดเฉลี่ยกลุ่มวิชาภาษาอังกฤษ
13) เกรดเฉลี่ยกลุ่มวิชาภาษาอื่น ๆ
14) เกรดเฉลี่ยกลุ่มวิชาอัตลักษณ์มหาวิทยาลัย
15) เกรดเฉลี่ยกลุ่มวิชาสังคม
16) เกรดเฉลี่ยกลุ่มวิชาความสร้างสรรค์
17) เกรดเฉลี่ยกลุ่มวิชาเทคโนโลยีสารสนเทศ
18) ผลการเรียนวิชา Web database

1.3.2 ขอบเขตด้านอัลกอริทึม

1) Decision Tree

2) Naïve Bayes

3) Neural Network

4) Support Vector Machine

1.3.3 ขอบเขตด้านเครื่องมือในการวิจัย

1.3.3.1 ฮาร์ดแวร์ (Hardware) สำหรับดำเนินการวิจัย มีดังนี้

1) หน่วยประมวลผลกลาง (Central Processing Unit: CPU) Intel Core i5 หรือดีกว่า

2) หน่วยความจำหลัก (Random Access Memory: RAM) มีความจุอย่างน้อย 4 กิกะไบต์

3) ฮาร์ดดิสก์ (Hard Disk) ความจุอย่างน้อย 500 กิกะไบต์

1.3.3.2 ซอฟต์แวร์ (Software) สำหรับดำเนินการวิจัยมีดังนี้

1) ระบบปฏิบัติการวินโดว์ 10 (Windows 10)

2) โปรแกรมสำเร็จรูป Rapid Miner Studio

3) โปรแกรมสำเร็จรูป SPSS for Window

1.4 ประโยชน์ของงานวิจัย

1.4.1 อาจารย์ผู้สอนวิชาด้านคอมพิวเตอร์ ได้ข้อมูลพื้นฐานเพื่อนำไปออกแบบการสอนให้สอดคล้องกับความรู้ทักษะพื้นฐานของนิสิต

1.4.2 นิสิตปริญญาตรีสาขาวิชาเทคโนโลยีการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยบูรพา มีผลสัมฤทธิ์ทางการเรียนวิชา Web Database เป็นไปตามวัตถุประสงค์การเรียนรู้

14.3 บัณฑิตสาขาเทคโนโลยีการศึกษามีทักษะในการพัฒนาฐานข้อมูลบนเวปให้กับองค์กรต่าง ๆ ได้อย่างมีประสิทธิภาพ

1.5 นิยามศัพท์เฉพาะ

1.5.1 Web Database หมายถึง ฐานข้อมูลบนเว็บแอปพลิเคชั่น เป็นฐานข้อมูลที่ออกแบบมาเพื่อจัดการและเข้าถึงผ่านทางอินเทอร์เน็ต ผู้ให้บริการเว็บไซต์สามารถจัดการรวบรวมข้อมูลและนำเสนอผลการวิเคราะห์ข้อมูลบนเว็บแอปพลิเคชัน

1.5.2 การทำเหมืองข้อมูล (Data Mining) หมายถึง การสกัดสารสนเทศ (Information) หรือความรู้ (Knowledge) จากข้อมูลดิบ (Raw Data) โดยสารสนเทศหรือความรู้ดังกล่าวอยู่ในรูปของความสัมพันธ์ (Relationship) แบบรูป (Pattern) หรือมโนทัศน์ (Concept) ในเชิงลึกที่ไม่สามารถมองออกได้ชัดโดยวิธีการประมวลผลข้อมูลแบบพื้นฐาน (Non-trivial)

1.5.3 Machine Learning หมายถึง ศาสตร์ที่ว่าด้วยการศึกษาและสร้างอัลกอริทึมที่สามารถเรียนรู้ข้อมูล และทำนายข้อมูลได้ ในงานวิจัยนี้ผู้วิจัยใช้วิธี Decision Tree, Naïve Bayes Neural Network และ Support Vector Machine

1.6 แผนการดำเนินงาน

ตารางที่ 1.1 แผนดำเนินงานวิจัย

รายการดำเนินการ	เดือน					
	1	2	3	4	5	6
1. วิเคราะห์ปัจจัยที่ส่งผลต่อการเรียนวิชา Web Database	←→					
2. ศึกษารูปแบบข้อมูลการจากฐานข้อมูลนิสิตและแบบสอบถาม		←→				
3. ศึกษาการทำงานการของอัลกอริทึมต่าง ๆ			←→			
4. จัดทำชุดข้อมูลสำหรับทดสอบ			←→			
5. ทดสอบสิทธิภาพโมเดล				←→		
6. วิเคราะห์และสรุปผลการวิจัย					←→	
7. รวบรวมข้อมูลและจัดทำวิทยานิพนธ์						←→

1.7 การตอบรับการตีพิมพ์

หัวข้อ “การเปรียบเทียบประสิทธิภาพการทำนายผลการเรียนวิชา Web database ของนิสิตสาขาเทคโนโลยีการศึกษา คณะศึกษาศาตร์ มหาวิทยาลัยบูรพา ด้วยเทคนิคเหมืองข้อมูล (Data mining)

COMPARISON OF THE PERFORMANCE PREDICTION IN WEB DATABASE COURSE FOR EDUCATIONAL TECHNOLOGY STUDENTS AT FACULTY OF EDUCATION BURAPHA UNIVERSITY USING DATA MINING TECHNIQUES” การประชุมวิชาการระดับชาติวิทยาศาสตร์และเทคโนโลยี ครั้งที่ 5 (The 5th National Conference on Science and Technology) ปี 2564 นำเสนอวันที่ 15 มกราคม 2564

บทที่ 2
เอกสารและงานวิจัยที่เกี่ยวข้อง

การวิจัยเรื่อง การประยุกต์ใช้ Machine Learning ทำนายผลการเรียนวิชา Web Database ของนิสิตสาขาเทคโนโลยีการศึกษา คณะศึกษาศาตร์ มหาวิทยาลัยบูรพา ผู้วิจัยดำเนินการศึกษาเอกสารงานวิจัยที่เกี่ยวข้อง ประกอบด้วย 1) ทักษะพื้นฐานการเขียนโปรแกรมคอมพิวเตอร์ คอมพิวเตอร์ 2) การทำเหมืองข้อมูล(Data Mining) 3) Decision Tree 4) Naïve Bayes 5) Neural Network 6) Support Vector Machine 7) การวิเคราะห์ความถดถอย (Regression Analysis) 8) โปรแกรม Rapid Miner Studio และ 9) งานวิจัยที่เกี่ยวข้อง
ดังรายละเอียดต่อไปนี้

2.1 ทักษะพื้นฐานการเขียนโปรแกรมคอมพิวเตอร์

การเขียนโปรแกรมคอมพิวเตอร์ (Computer programming) [9] หรือการเขียนโปรแกรม (Programming) หรือ การเขียนโค้ด (Coding) เป็นขั้นตอนการเขียนทดสอบและดูแลซอร์สโค้ดของโปรแกรมคอมพิวเตอร์ ซึ่งซอร์สโค้ดนั้นจะเขียนด้วยภาษาโปรแกรมขั้นตอนการเขียนโปรแกรมต้องการความรู้ในหลายด้านด้วยกันเกี่ยวกับโปรแกรมที่ต้องการจะเขียนและขั้นตอนวิธีที่จะใช้ ซึ่งในวิศวกรรมซอฟต์แวร์นั้นการเขียนโปรแกรมถือเป็นเพียงขั้นหนึ่งในวงจรชีวิตของการพัฒนา Software การเขียนโปรแกรมจะได้มาซึ่งซอร์สโค้ดของโปรแกรมนั้น ๆ โดยปกติแล้วจะอยู่ในรูปแบบของข้อความธรรมดา ซึ่งไม่สามารถนำไปใช้งานได้ จะต้องผ่านการคอมไพล์ตัวซอร์สโค้ดนั้นให้เป็นภาษาเครื่อง (Machine Language) เสียก่อนจึงจะได้เป็นโปรแกรมที่พร้อมใช้งานการเขียนโปรแกรมถือว่าเป็นการผสมผสานกันระหว่างศาสตร์ของศิลปะ วิทยาศาสตร์ คณิตศาสตร์ และวิศวกรรม เข้าด้วยกัน

ในการศึกษาทักษะพื้นฐานการเขียนโปรแกรมคอมพิวเตอร์ ผู้วิจัยได้ศึกษาเอกสารงานวิจัยเพื่อสังเคราะห์เป็น Atribute นำเข้าโมเดลทำนายผลการเรียนวิชา Web Database ดังนี้ การศึกษางานวิจัย เรื่องการพัฒนาระบบการสอนเพื่อพัฒนาความสามารถในการเขียนโปรแกรมคอมพิวเตอร์สำหรับนักเรียนระดับชั้นมัธยมศึกษาตอนต้น [9] ซึ่งได้สรุปผลการสัมภาษณ์ ดร.ดวงแก้ว สวามิภักดิ์ คณะวิทยาศาสตร์ มหาวิทยาลัยธรรมศาสตร์ ในประเด็นนักศึกษาที่มาเรียน

สาขาศาสตร์คอมพิวเตอร์ ควรมีพื้นฐานที่ดีด้านใดบ้าง สรุปได้ว่า *"นักเรียนควรมีพื้นฐาน ทางคณิตศาสตร์ที่ดี และภาษาอังกฤษก็ช่วยได้มาก แต่ไม่ได้หมายความว่า เด็กที่มีพื้นฐานไม่ดีจะไม่ประสบความสำเร็จ นักเรียนกลุ่มนี้จะต้องเรียนหนักขึ้นและใช้ความพยายามมากขึ้น"* ประเด็นบัณฑิตด้านคอมพิวเตอร์ควรมีความสามารถอย่างไรบ้าง จึงจะพัฒนาเป็นนักคอมพิวเตอร์มืออาชีพได้ สรุปได้ว่า *"อย่างน้อยควรพัฒนาโปรแกรมได้ เราเน้นเสมอว่า CS ไม่ใช่การเขียนโปรแกรมแต่การพัฒนาโปรแกรมในที่นี้รวมถึงการ Apply ศาสตร์ด้านอื่น เช่น Data Structures, System Software, Operating Systems เข้าไปด้วยแต่ที่สำคัญที่สุดคือความสามารถในการพัฒนาตนเองโดยการไม่หยุดอยู่กับที่คอยติดตามเทคโนโลยีใหม่ ๆ อยู่เสมอ"* ประเด็นอะไรเป็นเคล็ดลับในการเรียนคอมพิวเตอร์ให้ประสบความสำเร็จสรุปได้ว่า *"การตื่นตัวในวงการคอมพิวเตอร์ ติดตามข่าวสารอยู่ตลอดเวลา ต้องบ้าคอมพิวเตอร์พร้อมที่จะนั่งอยู่กับมันเป็นคืนโดยไม่มีใครบังคับขยันอ่าน และขยัน Discuss กับคนอื่น ๆ"* ประเด็นนักศึกษาคอมพิวเตอร์ที่เป็นชายกับหญิงมีความแตกต่างกันในด้านการเรียนและ ความสามารถหรือไม่ สรุปได้ว่า *"ความแตกต่างระหว่างชาย / หญิง ก็คงแตกต่างกันเช่นเดียวกับสาขาอื่น ๆ คือโดยทั่วไปผู้หญิงจะละเอียดอ่อนและมีความตั้งใจกว่า ในขณะที่ผู้ชายจะคล่องตัวกว่าก็คงสรุป ไม่ได้ว่าเพศใดดีกว่าก็คงต้องอยู่ที่งานว่าเป็นประเภทใด"* การสัมภาษณ์ รศ. ยืน ภู่วรวรรณ ได้เสนอแนะว่า *"นิสิตที่จะเรียน CS ได้ ควรมีพื้นฐานทางด้านคณิตศาสตร์ที่ดี และมีความคิดเชิงระบบ ชอบงานทางด้านวิศวกรรมหรืองานสร้างสรรค์ มีมนุษย์สัมพันธ์ดี สามารถทำงานร่วมกับผู้อื่นได้"*

งานวิจัยเรื่อง การพัฒนากรอบงานความสามารถด้านการเขียนโปรแกรม [10] ซึ่งได้ค้นพบข้อสรุปจากงานวิจัยและมีข้อเสนอแนะถึงวิธีการประเมินความถนัดและทักษะการเขียนโปรแกรมคอมพิวเตอร์ไว้ว่า การกรอกแบบสอบถาม (Questionnaires) ก่อนเรียนว่าได้มีการเตรียมตัวก่อนเรียน หรือรู้เรื่องที่จะเรียนมากน้อยแค่ไหนบ้าง เพื่อที่จะได้เป็นข้อมูลประกอบในการวิเคราะห์ทักษะได้อย่างชัดเจนมากขึ้น และจำเป็นที่จะต้องมีแบบทดสอบก่อนที่จะเริ่มเรียนการเขียนโปรแกรม (Pretest) เพื่อที่จะได้รู้ว่าพื้นฐานการเขียนโปรแกรมของผู้เรียนการเขียนโปรแกรมเป็นอย่างไรบ้าง เนื่องจากในการเรียนการเขียนโปรแกรมนั้นจะมีความหลากหลายทางด้านทักษะและความรู้ในการเขียนโปรแกรมอยู่มาก ผู้เขียนบางคนเคยเขียนโปรแกรมมาเป็นระยะเวลานานต่างจากผู้เขียนบางคนที่ไม่เคยเขียนโปรแกรมมาเลยซึ่งข้อมูลในส่วนนี้จะมาช่วยในการประเมินทักษะ และยังบ่งบอกความก้าวหน้า หรือประสิทธิผลที่ได้จากการเรียนเขียนโปรแกรมได้อีกด้วย ซึ่งในเรื่องของแบบทดสอบก่อนเรียน และแบบสอบถามหลังเรียนเป็นเรื่องที่น่าสนใจ และควรทำเป็นอย่างยิ่ง เพราะจะสามารถบ่งบอกทักษะและพัฒนาการทางด้านการเรียนได้เป็นอย่างดี

งานวิจัยเรื่อง ตัวแปรทักษะพื้นฐานการเขียนโปรแกรมคอมพิวเตอร์และคณิตศาสตร์ [11] ได้ทำการศึกษาความสัมพันธ์ระหว่างการเขียนโปรแกรมและประสบการณ์ทางคณิตศาสตร์ กลุ่มตัวอย่างประกอบด้วย นักเรียนจำนวน 46 คน (อายุระหว่าง 9-17 ปี) ในระหว่างการเข้าค่ายอบรมคอมพิวเตอร์ช่วงปิดภาคเรียน ระดับของการเรียนเขียนโปรแกรมคอมพิวเตอร์กำหนดจากผลการทดสอบวัดระดับที่แบ่งเป็นระดับชั้นต้น ชั้นกลาง และชั้นสูง ประสบการณ์ทางคณิตศาสตร์กำหนดจากเกณฑ์การผ่านการเรียนวิชาพีชคณิต ตัวแปรของทักษะพื้นฐานได้จากคะแนนแบบทดสอบวัดทักษะการศึกษาในครั้งนี้จะบอกถึงลักษณะความสำคัญเชิงนัยระหว่างประสบการณ์การเรียนเขียนโปรแกรมคอมพิวเตอร์ และประสบการณ์ทางคณิตศาสตร์มีผลกับตัวแปรของทักษะพื้นฐาน โดยค่าความสัมพันธ์ของการเรียนเขียนโปรแกรมคอมพิวเตอร์มีค่าที่น่าเชื่อถือมากกว่าประสบการณ์ทางคณิตศาสตร์ ที่มีผลต่อตัวแปรของทักษะพื้นฐาน

จากการศึกษาปัจจัยที่ส่งผลต่อการประสบผลสำเร็จ ทางด้านการเรียนคอมพิวเตอร์ พบว่า ผลการเรียน (Grade) [1][2] เป็นปัจจัยที่สำคัญที่บ่งบอกถึงประสิทธิภาพการเรียนรู้เป็นอย่างไร 2) ประสบการณ์การเขียนโปรแกรม (Previous computer programming experiences) ส่งผลให้เกิดความสำเร็จในการเรียนรู้ 3) พื้นฐานทางคณิตศาสตร์ (Mathematics background)) [3][4][5] ผู้ที่มีความรู้พื้นฐานคณิตศาสตร์ดีจะมีทักษะในการคิดและเข้าใจเรื่องของโครงสร้างและแนวทางการดำเนินงานซึ่งเป็นพื้นฐานสำคัญของการเขียนโปรแกรมที่ดี

สรุปจากการศึกษาเอกสาร งานวิจัยที่เกี่ยวข้องกับทักษะพื้นฐานการเขียนโปรแกรมคอมพิวเตอร์ ผู้วิจัยได้สังเคราะห์เลือก Attribute ที่เกี่ยวข้องกับทักษะพื้นฐานการเขียนโปรแกรมคอมพิวเตอร์ 18 Attribute ได้แก่ 1) เพศ 2) เคยเรียนภาษา HTML 3) เคยเรียนภาษา C หรือ C++ 4) เคยเรียนภาษา PHP 5) เกรดเฉลี่ยวิชาคณิตศาสต์ ม.6 6) เกรดเฉลี่ยจบการศึกษาระดับชั้น ม. 6 7) แผนการเรียนมัธยมปลาย 8) ประเภทการรับเข้ามหาวิทยาลัย 9) ความรู้สึกต่อการเรียนวิชาด้านคอมพิวเตอร์ 10) คอมพิวเตอร์ส่วนตัว 11) การทบทวนบทเรียน 12) เกรดเฉลี่ยกลุ่มวิชาภาษาอังกฤษ 13) เกรดเฉลี่ยกลุ่มวิชาภาษาอื่น ๆ 14) เกรดเฉลี่ยกลุ่มวิชาอัตลักษณ์มหาวิทยาลัย 15) เกรดเฉลี่ยกลุ่มวิชาสังคม 16) เกรดเฉลี่ยกลุ่มวิชาความสร้างสรรค์ 17) เกรดเฉลี่ยกลุ่มวิชาเทคโนโลยีสารสนเทศ 18) ผลการเรียนวิชา Web Database

2.1.1 การเขียนโปรแกรมภาษา PHP

PHP แต่เดิมย่อมาจาก Personal Home Page แต่ต่อมาก็เปลี่ยนเป็นย่อมาจาก PHP Hypertext Preprocessor ในปัจจุบัน Website ต่าง ๆ ได้มีการพัฒนาในด้านต่าง ๆ อย่างรวดเร็ว อาทิ เช่น เรื่องของความสวยงามและแปลกใหม่ การบริการข่าวสารข้อมูลที่ทันสมัยเป็นสื่อกลางในการติดต่อ และสิ่งหนึ่งที่กำลังได้รับความนิยมเป็นอย่างมากซึ่งถือได้ว่าเป็นการปฏิวัติรูปแบบการขาย

ของก็คือ E-commerce ซึ่งเจ้าของสินค้าไม่จำเป็นต้องมีร้านค้าจริงและไม่จำเป็นต้องจ้างคนขายของ อีกต่อไปร้านค้าและตัวสินค้านั้นจะไปปรากฏอยู่บน Website และการซื้อขายก็เกิดขึ้นบนโลกของ Internet แล้ว PHP ช่วยเราให้เป็นการพัฒนา Website และความสามารถที่โดดเด่นอีกประการหนึ่งของ PHP นั้นคือ database enabled web page ทำให้เอกสารของ HTML สามารถที่จะเชื่อมต่อกับระบบฐานข้อมูล (database)ได้อย่างมีประสิทธิภาพและรวดเร็ว จึงทำให้ความต้องการในเรื่องการจัดรายการสินค้าและรับรายการสั่งของตลอดจนการจัดเก็บข้อมูลต่าง ๆ ที่สำคัญผ่านทาง Internet เป็นไปได้

PHP เป็นภาษาจำพวก scripting language คำสั่งต่าง ๆ จะเก็บอยู่ในไฟล์ที่เรียกว่า ภาษาสคริปต์ (Script) และเวลาใช้งานต้องอาศัยตัวแปลชุดคำสั่ง ตัวอย่างของภาษาสคริปต์ก็ เช่น JavaScript, Perl เป็นต้น ลักษณะของ PHP ที่แตกต่างจากภาษาสคริปต์แบบอื่น คือ PHP ได้รับการพัฒนาและออกแบบมาเพื่อใช้งานในการสร้างเอกสารแบบ HTML โดยสามารถสอดแทรกหรือแก้ไขเนื้อหาได้โดยอัตโนมัติ ดังนั้นจึงกล่าวว่า PHP เป็นภาษาที่เรียกว่า server-side หรือ HTML-embedded scripting language เป็นเครื่องมือที่สำคัญชนิดหนึ่งที่ช่วยให้เราสามารถสร้างเอกสารแบบ Dynamic HTML ได้อย่างมีประสิทธิภาพและมีลูกเล่นมากขึ้น

เนื่องจากว่า PHP ไม่ได้เป็นส่วนหนึ่งของตัว Web Server ดังนั้นถ้าจะใช้ PHP ก็จะต้องดูก่อนว่า Web server นั้นสามารถใช้สคริปต์ PHP ได้หรือไม่ ยกตัวอย่างเช่น PHP สามารถใช้ได้กับ Apache Web Server และ Personal Web Server สำหรับระบบปฏิบัติการ Windows 95/98/NT ในกรณีของ Apache เราสามารถใช้ PHP ได้ 2 รูปแบบคือ ในลักษณะของ CGI และ Apache Module ความแตกต่างอยู่ตรงที่ว่า ถ้าใช้ PHP เป็นแบบโมดูล PHP จะเป็นส่วนหนึ่งของ Apache หรือเป็นส่วนขยายในการทำงานนั่นเอง ซึ่งจะทำงานได้เร็วกว่าแบบที่เป็น CGI เพราะถ้าเป็น CGI แล้วตัวแปลชุดคำสั่งของ PHP ถือว่าเป็นแค่โปรแกรมภายนอก ซึ่ง Apache จะต้องเรียกขึ้นมาทำงานทุกครั้งที่ต้องการใช้ PHP ดังนั้น ถ้ามองในเรื่องของประสิทธิภาพในการทำงาน การใช้ PHP แบบที่เป็นโมดูลหนึ่งของ Apache จะทำงานได้มีประสิทธิภาพมากกว่า PHP เป็น Complier หรือ Interpreter PHP เป็น Interpreter ประมวลผลการทำงาน โดยแปลความหมายที่ละบรรทัดข้อดีของ Interpreter คือ เป็นการ Open source โปรแกรมที่ Open source จะพัฒนาอย่างรวดเร็วเนื่องจากเกิดการ Copy แก้ไข ตลอดจนพัฒนาขึ้นมาใหม่ตามแนวทางตัวอย่าง การที่มีต้นแบบหรือต้นฉบับจะทำให้ไม่เสียเวลาเขียนใหม่ นำไปปรับแก้ไขก็สามารถใช้ได้ ตัวอย่าง Open source เช่น Linux JavaScript Perl PHP ASP เป็นต้น

ภาษาสคริปต์ (Scripting Language)

โฮมเพจจำเป็นต้องหาอะไรแปลกใหม่ มาลงในเว็บไซต์เสมอ หากมีเพียงแค่ HTML เว็บไซต์ก็ ไม่มีการตอบโต้อะไรกลับมา ภาษาสคริปต์ สำหรับเว็บไซต์นั้นเขียนได้หลายภาษา และมีรูปแบบการเขียนอยู่ด้วยกันสองแบบ คือ

1. Client-Side Scripting เป็นการเขียนโปรแกรมภาษาสคริปต์ ให้ทำงานบน Web Browser โดยเขียนโปรแกรมแทรกหรือแฝง (Embed) เข้าไปเป็นส่วนหนึ่งของเอกสาร HTML โปรแกรมภาษาสคริปต์ประเภทนี้ ได้แก่ JavaScript, VBScript

2. Server-Side Scripting เป็นการเขียนโปรแกรมภาษาสคริปต์ ให้ทำงานบน Webserver โดย Web Browser จะเป็นเพียงแค่ตัวที่แสดงผลการทำงานเท่านั้น โปรแกรมที่ทำงานบนWeb Server เหล่านี้เราเรียกว่า CGI Script ซึ่งสามารถเขียนได้หลายภาษาด้วยกัน เช่น Perl, Pascal, VB เป็นต้น และ โปรแกรมภาษาสคริปต์ประเภทนี้ ได้แก่ JSP, ASP, PHP

3. หลักการทำงานของ PHP- ฝั่ง Client จะทำการร้องขอหรือเรียกใช้งานไฟล์ PHP ที่เก็บในเครื่อง Server

- ฝั่ง Server จะทำการประมวลผลไฟล์ PHP ตามที่Client ทำการร้องขอ
- ทำการประมวลผลไฟล์ PHP
- เป็นการติดต่อกับฐานข้อมูล และนำข้อมูลในฐานข้อมูลมาใช้ร่วมกับการประมวลผล

ความสามารถของภาษา PHP

เนื่องจากภาษา PHP เป็นภาษาสคริปต์ที่มีการประมวลผล และแปลคำสั่งประเภท Server-Side Scripting ในการเขียนคำสั่ง (Statement) หรือการทำงานจึงคล้ายกับภาษา Perl หรือภาษา C และสามารถทำงานร่วมกับภาษา HTML ได้เป็นอย่างดี การทำงานของภาษา PHP จึงมีรายละเอียด ดังต่อไปนี้ [12]

1. ภาษา PHP เป็นภาษาที่เป็นลักษณะโอเพนซอร์ส (Open source) ซึ่งนั้นหมายความว่าผู้ใช้งานสามารถที่จะดาวน์โหลด และติดตั้งได้เว็บเซิร์ฟเวอร์ (Web Server) โดยไม่เสียค่าใช้จ่าย

2. ภาษา PHP เป็นภาษาประเภท Server-Side Scripting มีการประมวลผลและแปลคำสั่งในฝั่งของเซิร์ฟเวอร์แล้วแปลงเป็นเอกสารในรูปแบบของภาษา HTML แล้วส่งผลลัพธ์ที่ได้กลับไปยังเว็บบราวเซอร์ในฝั่ง Client

3. ภาษา PHP สามารถทำงานได้บนระบบปฏิบัติการ (Operating System) ที่หลากหลาย ตัวอย่างเช่น Unix, Windows หรือ Mac OS เป็นต้น

4. ภาษา PHP สามารถทำงานร่วมกับเว็บบราวเซอร์ เช่น Personal Web Server (PWS), Apache, และ Internet Information Service (IIS) เป็นต้น

5. สามารถเขียนโปรแกรมเชิงวัตถุ (Object Oriented Programming) ได้

6. สามารถทำงานร่วมกับระบบการจัดการฐานข้อมูลได้อย่างหลากหลาย เช่น MySQL, Fire Pro, Solid, และ Front Base เป็นต้น

7. ภาษา PHP สามารถทำงานร่วมกับโปรโตคอลได้ (Protocol) ชนิดต่าง ๆ ได้ เช่น HTTP (Hypertext Transfer Protocol), IMAP หรือ LDAP เป็นต้น

หลักการทำงานของภาษา PHP

ภาษา PHP นั้น จะทำงานอยู่ในฝั่งเว็บเซิร์ฟเวอร์ ดังนั้น ผู้ที่เข้าใช้บริการจะไม่สามารถเห็นสคริปต์ของภาษา PHP ที่เครื่องคอมพิวเตอร์ในฝั่งไคลเอ็นท์ เพราะระหว่างที่ใช้งานเว็บบราวเซอร์เรียกชมเว็บไซต์นั้น ในฝั่งของเว็บเซิร์ฟเวอร์จะทำการประมวลผลภาษา PHP และส่งผลลัพธ์ที่ได้ เป็นภาษา HTML หลังจากนั้นจะส่งHTML ที่ได้จากการประมวลผลแล้วนั้นมายังเว็บบราวเซอร์ของไคลเอ็นท์ แล้วแสดงผลให้เห็นเป็นรูปร่างของเว็บไซต์และเมื่อดูซอร์สโค้ด (Source Code) ในฝั่งไคลเอ็นท์ จะเห็นเฉพาะ HTML tags ของภาษา HTML เพียงอย่างเดียว ไม่เห็นสคริปต์ PHP สามารถสรุปขั้นตอนและหลักการทำงานระหว่างไคลเอ็นท์ และเว็บเซิร์ฟเวอร์ได้ดังต่อไปนี้ [12]

1. เว็บเซิร์ฟเวอร์ คือ โปรแกรมที่ทำงานอยู่บนเครื่องฝั่งเซิร์ฟเวอร์ และทำหน้าที่ในการรับคำสั่งจากการร้องขอของไคลเอ็นท์ โดยผู้ใช้จะเรียกชมหน้าเว็บเพจได้โดยใช้โปรโตคอล HTTP ผ่านทางเว็บบราวเซอร์ และประมวลผลการทำงานจากการร้องขอดังกล่าว แล้วส่งข้อมูลกลับไปยังเครื่องของไคลเอ็นท์ที่ร้องขอ

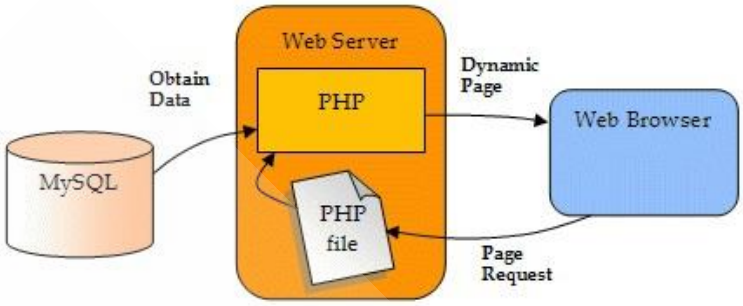
สรุปคือ เว็บเซิร์ฟเวอร์จะคอยให้บริการแก่ ไคลเอ็นท์ ที่ร้องขอข้อมูลเข้ามาโดยผ่านเว็บบราวเซอร์ หรือจากการร้องขอผ่านโปรโตคอล HTT โดยเว็บเพจที่เขียนก็สามารถเขียนได้หลายภาษา หลายรูปแบบ การเขียนสคริปต์ PHP ระบบจะสามารถทำงานได้นั้นจำเป็นต้องมีเว็บเซิร์ฟเวอร์ เป็นตัวเรียกใช้งาน Engine ของภาษา PHP ดังนั้นถ้าต้องการให้คอมพิวเตอร์ประมวลผลและแปลภาษา PHP ได้นั้น จำเป็นที่จะต้องมีเว็บเซิร์ฟเวอร์ หรือกรณีศึกษานี้สามารถจำลองเครื่องคอมพิวเตอร์ ธรรมดาให้ทำหน้าที่เป็นเว็บเซิร์ฟเวอร์ รวมถึงรูปแบบและวิธีการเข้าถึงข้อมูลจะเรียกใช้ผ่านบริการจากโปรโตคอล HTTP เพื่อให้ผู้ใช้สามารถอ่านข้อมูลภาพหรือข้อมูลอื่น ๆ ตามต้องการ

2. หลักการทำงานของเว็บเซิร์ฟเวอร์ เป็นการให้บริการผ่านระบบอินเทอร์เน็ต เช่นเดียวกันกับบริการอื่น ๆ ซึ่งอยู่ในรูปของไคลเอ็นท์-เซิร์ฟเวอร์ (Client-Server) โดยมีโปรแกรมเว็บไคลเอ็นท์ (Web Client) ที่ทำหน้าที่ในการร้องขอบริการ และมีโปรแกรมเว็บเซิร์ฟเวอร์ทำหน้าที่การให้บริการโปรแกรมเว็บไคลเอ็นท์ อาจจะกล่าวได้ว่า โปรแกรมเว็บบราวเซอร์ในเครื่องของผู้ใช้ สำหรับโปรแกรมเว็บเซิร์ฟเวอร์ นั้นจะถูกติดตั้งไว้ในเครื่องของผู้ให้บริการเว็บไซต์ การติดติอระหว่างโปรแกรมเว็บบราวเซอร์กับโปรแกรมเว็บเซิร์ฟเวอร์ จะกระทำผ่านโปรโตคอล HTTP

3. กลไกการทำงานของการร้องขอเว็บธรรมดาในการร้องขอเว็บธรรมดา โดยปกติจะมีนามสกุลที่เป็น .htm หรือ .html เมื่อใช้เว็บไคลเอ็นท์ เปิดดูหรือเรียกดูเว็บเพจใด เว็บเซิร์ฟเวอร์ก็จะส่งเว็บเพจนั้น ๆ กลับมายังเว็บไคลเอ็นท์ หลังจากนั้นเว็บไคลเอ็นทจะแสดงผลตามคำสั่งของ HTML

4. กลไกการทำงานของการร้องขอเว็บที่ใช้ภาษา PHP จะมีการประมวลผล และแสดงผลเป็นแบบไดนามิกเว็บเพจ สามารถติดต่อกับผู้ใช้ฐานข้อมูลคำนวณ และประมวลผลตามเงื่อนไขในรูปแบบต่าง ๆ ได้

ในการประมวลผลภาษา PHP เริ่มจากการที่เว็บไคล์เอ็นท์ขอร้องไฟล์ที่ใช้ภาษา PHP ในการพัฒนาร่วมกับ HTML มายังเซิร์ฟเวอร์เว็บเซิร์ฟเวอร์จะเรียกใช้ Engine ของภาษา PHP หรือตัวแปลคำสั่งของภาษา PHP กรณีที่มีการใช้คำสั่งในการเชื่อมต่อฐานข้อมูลระบบก็จะติดต่อไปยังฐานข้อมูลตามการร้องขอ มีการตรวจสอบสิทธิ์การเข้าถึงฐานข้อมูล และอื่น ๆ ตามลำาดับขั้นตอน อีกทั้งภาษา PHP ยังมีความยืดหยุ่นในการเขียนสคริปต์ มีฟังก์ชันการทำงานที่หลากหลาย สามารถประยุกต์ใช้งานได้ง่าย สำหรับการติดต่อกับผู้ใช้งานระบบ รวมถึงการมีความสามารถที่จะเชื่อมต่อฐานข้อมูลที่หลากหลาย ซึ่งฐานข้อมูลที่นิยมมากที่สุดก็คือ ฐานข้อมูล MySQL



ภาพที่ 2.1 แสดงการทำงานของภาษา PHP และ MySQL

ที่มา: https://qph.fs.quoracdn.net/main-qimg-6bae3154eeea1925989945c6c65e7ba2

MySQL คือ โปรแกรมระบบจัดการฐานข้อมูล ที่พัฒนาโดยบริษัท MySQL AB มีหน้าที่เก็บข้อมูลอย่างเป็นระบบ รองรับคำสั่ง SQL เป็นเครื่องมือสำหรับเก็บข้อมูลที่ต้องใช้ร่วมกับเครื่องมือหรือโปรแกรมอื่นอย่างบูรณาการ เพื่อให้ได้ระบบงานที่รองรับความต้องการของผู้ใช้ เช่นทำงานร่วมกับเครื่องบริการเว็บ (Web Server) เพื่อให้บริการแก่ภาษาสคริปต์ที่ทำงานฝั่งเครื่องบริการ (Server-Side Script) เช่น ภาษา php ภาษา aps.net หรือภาษา jsp เป็นต้น หรือทำงานร่วมกับโปรแกรมประยุกต์ (Application Program) เช่น ภาษา Visual Basic , ภาษา JAVA หรือภาษา C# เป็นต้น โปรแกรมถูกออกแบบให้สามารถทำงานได้บนระบบปฏิบัติการที่หลากหลาย และเป็นระบบฐานข้อมูลโอเพนซอร์ส (Open Source) ที่ถูกนำไปใช้งานมากที่สุด

MySQL: มายเอสคิวแอล เป็นระบบจัดการฐานข้อมูลโดยใช้ภาษา SQL. แม้ว่า MySQL เป็นซอฟต์แวร์โอเพนซอร์ส แต่แตกต่างจากซอฟต์แวร์โอเพนซอร์สทั่วไป โดยมีการพัฒนาภายใต้บริษัท MySQL AB ในประเทศสวีเดน โดยจัดการ MySQL ทั้งในแบบที่ให้ใช้ฟรี และแบบที่ใช้ในเชิงธุรกิจ MySQL สร้างขึ้นโดยชาวสวีเดน 2 คน และชาวฟินแลนด์ชื่อ David Axmark, Allan Larsson และ Michael "Monty" Widenius. ปัจจุบันบริษัทซันไมโครซิสเต็มส์ (Sun Microsystems, Inc.) เป็นเจ้าของ MySQL AB

ความสามารถและการทำงานของโปรแกรมภาษา MySQL

1. MySQL ถือเป็นระบบจัดการฐานข้อมูล (Data Base Management System (DBMS) ฐานข้อมูลมีลักษณะเป็นโครงสร้างของการเก็บรวบรวมข้อมูล การที่จะเพิ่มเติมเข้าถึงหรือประมวลผลข้อมูลที่เก็บในฐานข้อมูลจำเป็นจะต้องอาศัยระบบจัดการฐานข้อมูล ซึ่งจะทำหน้าที่เป็น

ตัวกลางในการจัดการกับข้อมูลในฐานข้อมูลทั้งสำหรับการใช้งานเฉพาะ และรองรับการทำงานของแอพลิเคชั่นอื่น ๆ ที่ต้องการใช้งานข้อมูลในฐานข้อมูล เพื่อให้ได้รับความสะดวกในการจัดการกับข้อมูลจำนวนมาก MySQL ทำหน้าที่เป็นทั้งตัวฐานข้อมูลและระบบจัดการฐานข้อมูล

2. MySQL เป็นระบบจัดการฐานข้อมูลแบบ relational โดยฐานข้อมูลนี้จะทำการเก็บข้อมูลทั้งหมดในรูปแบบของตารางแทนการเก็บข้อมูลทั้งหมดลงในไฟล์ เพียงไฟล์เดียว ทำให้ทำงานได้รวดเร็วและมีความยืดหยุ่น นอกจากนั้น แต่ละตารางที่เก็บข้อมูลสามารถเชื่อมโยงเข้าหากันทำให้สามารถรวมหรือจัดกลุ่มข้อมูลได้ตามต้องการ โดยอาศัยภาษา SQL ที่เป็นส่วนหนึ่งของโปรแกรม MySQL ซึ่งเป็นภาษามาตรฐานในการเข้าถึงฐานข้อมูล

3. MySQL เป็น Open Source น ผู้ใช้งาน MySQL ทุกคนสามารถใช้งานและปรับแต่งการทำงานได้ตามต้องการ สามารถดาวน์โหลดโปรแกรม MySQL ได้จากอินเทอร์เน็ตและนำมาใช้งานโดยไม่มีค่าใช้จ่ายใด ๆ

นอกจากนั้น MySQL ถูกออกแบบและพัฒนาขึ้นมาเพื่อทำหน้าที่ให้บริการรองรับการจัดการกับฐานข้อมูลขนาดใหญ่ ซึ่งการพัฒนายังคงดำเนินอยู่อย่างต่อเนื่อง ส่งผลให้มีฟังก์ชันการทำงานใหม่ ๆ ที่อำนวยความสะดวกแก่ผู้ใช้งานเพิ่มขึ้นอยู่ตลอดเวลา รวมไปถึงการปรับปรุงด้านความต่อเนื่อง ความเร็วในการทำงาน และความปลอดภัย ทำให้ MySQL เหมาะสมต่อการนำไปใช้งานเพื่อเข้าถึงฐานข้อมูลบนเครือข่ายอินเทอร์เน็ต

ฐานข้อมูลที่ใช้ในการพัฒนา MySQL เป็นฐานข้อมูลแบบโอเพนซอร์ส ที่ได้รับความนิยมในการใช้งานสูงสุดโปรแกรมหนึ่งบนเครื่องให้บริการ มีความสามารถในการจัดการกับฐานข้อมูลด้วยภาษา SQL (Structures Query Language) อย่างมีประสิทธิภาพมีความรวดเร็วในการทำงานรองรับการทำงานจากผู้ใช้หลายคน โดยหน้าที่ความสามารถและการทำงานของโปรแกรม MySQL ถือเป็นระบบจัดการฐานข้อมูล (DBMS) มีลักษณะเป็นโครงสร้างของการเก็บข้อมูลการที่จะเพิ่มเติมเข้าถึง หรือประมวลผลข้อมูลที่เก็บในฐานข้อมูลจำเป็นจะต้องอาศัยระบบจัดการฐานข้อมูล ซึ่งจะทำหน้าที่เป็นตัวกลางในการจัดการกับข้อมูลในฐานข้อมูลทั้งสำหรับการใช้งานเฉพาะ และรองรับการทำงานของแอพพลิเคชั่นอื่น ๆ ที่ต้องการใช้งานข้อมูลในฐานข้อมูล เพื่อให้ได้รับความสะดวกในการจัดการกับข้อมูลจำนวนมาก

MySQL ทำหน้าที่เป็นทั้งตัวฐานข้อมูล และระบบจัดการฐานข้อมูล มีลักษณะเป็นระบบฐานข้อมูลเชิงสัมพันธ์ โดยทำการเก็บข้อมูลทั้งหมดในรูปแบบของตารางแทนการเก็บข้อมูลทั้งหมดลงในไฟล์เพียงไฟล์เดียว ทำให้ทำงานได้รวดเร็วและมีความยืดหยุ่น นอกจากนั้นแต่ละตารางเก็บข้อมูลสามารถเชื่อมโยงหากัน ทำให้สามารถจัดกลุ่มข้อมูลได้ตามต้องการ โดยอาศัยภาษา SQL ที่เป็นส่วนหนึ่งของโปรแกรม

2.2 การทำเหมืองข้อมูล (Data Mining)

การทำเหมืองข้อมูล (Data Mining) มีผู้ให้คำนิยามไว้เป็นจำนวนมาก แต่สามารถสรุปหลักสำคัญได้ว่า การทำเหมืองข้อมูลเป็นกระบวนการของการกลั่นกรองสารสนเทศ (Information) ที่ซ่อนอยู่ในฐานข้อมูลขนาดใหญ่เพื่อทำนายแนวโน้มและพฤติกรรม โดยอาศัยข้อมูลในอดีตและเพื่อใช้สารสนเทศเหล่านี้ในการสนับสนุนการตัดสินใจทางธุรกิจ [13]

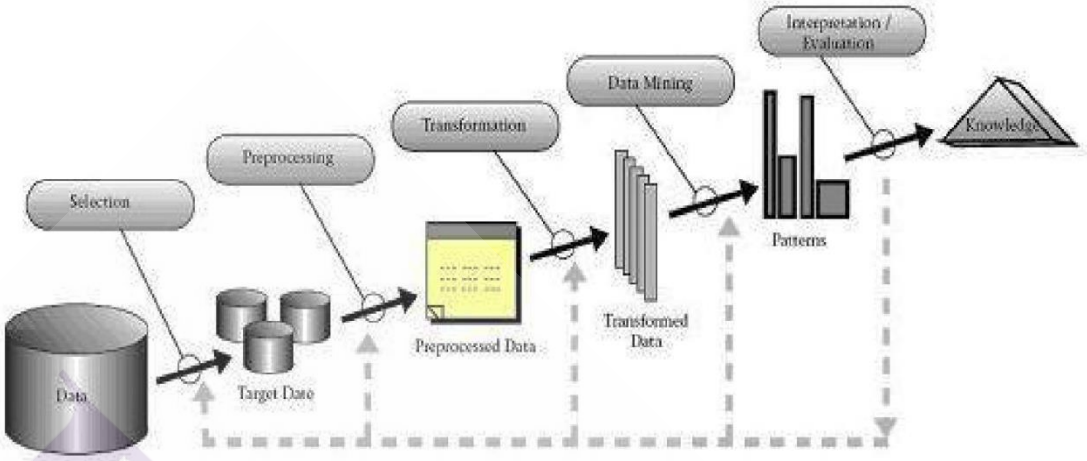
จากคำจำกัดความการทำเหมืองข้อมูล อาจหมายถึงการที่ผู้ใช้เข้าถึง ตรวจสอบ วิเคราะห์และสังเคราะห์ข้อมูลอย่างละเอียด โดยการสังเคราะห์ดังกล่าวอาจเป็นการเรียนรู้ข้อมูลในอดีตหรือข้อมูลในปัจจุบัน ผลลัพธ์ที่ได้ออกมาต้องมีลักษณะของข้อมูลที่เป็นข้อมูลที่ไม่รู้ (Unknown data) โดยเป็นสารสนเทศที่สมเหตุสมผล (Valid) และสามารถนำไปใช้ได้ (Actionable) โดยเป็นข้อมูลที่นำมาจากฐานข้อมูลขนาดใหญ่

ข้อมูลแบบ Unknown เป็นข้อมูลที่ไม่รู้มาก่อนและไม่มีความชัดเจน ไม่สามารถตั้งสมมติฐานล่วงหน้าได้ว่าควรเป็นแบบใด เช่น ข้อมูลการซื้อสินค้าของลูกค้าไม่แน่ชัดเจ้าของห้างสรรพสินค้าแห่งหนึ่งค้น พบว่า พฤติกรรมของผู้บริโภคที่เป็นพ่อบ้านมักจะซื้อสินค้าเบียร์และผ้าอ้อมในวันศุกร์ตอนเย็น สิ่งนี้เป็น แนวทางในการตัดสินใจให้เจ้าของกิจการควรจะเตรียมสินค้าไว้จำหน่ายซึ่งในขณะเดียวกันห้างสรรพสินค้าคู่แข่งอาจจะไม่รู้เรื่องนี้

ข้อมูลแบบ Valid เป็นข้อมูลที่ได้จากการใช้เทคนิคของการทำเหมืองข้อมูล โดยจะค้นพบสิ่งที่น่าสนใจตลอดเวลา แต่ต้องมีการพิจารณาด้วยว่าสิ่งนั้น Valid หรือไม่ เช่น ผู้ใช้มักพบว่ามีความสัมพันธ์ของการซื้อของสองสิ่งเสมอ เมื่อจำนวนความหลากหลายของสินค้ามากขึ้น แต่ไม่ได้หมายความว่าจะต้องให้ห้างสรรพสินค้าเก็บสินค้ามากขึ้น เพราะข้อมูลที่ได้อาจเกิดความคลาดเคลื่อน เพราะฉะนั้นจะต้องทำการ Validation และ Checking ความถูกต้องของข้อมูลและวิเคราะห์ความถูกต้องอีกครั้ง

ข้อมูลแบบ Actionable เป็นข้อมูลที่จะต้องถูกแปลงออกมาและนำมาใช้ในการตัดสินใจเพื่อความได้เปรียบในเชิงธุรกิจ บางครั้งข้อมูลที่ค้นพบเป็นสิ่งที่คู่แข่งได้ทำไปแล้ว ซึ่งต้องมีวิจารณญาณในการใช้ ซึ่งบางทีข้อมูลที่ได้อาจจะไม่มีประโยชน์อะไร

การทำเหมืองข้อมูล เป็นขั้นตอนหนึ่งในกระบวนการค้นหาความรู้ในฐานข้อมูล ซึ่งประกอบไปด้วยขั้นตอน ดังภาพที่ 2.2 [14]



ภาพที่ 2.2 แสดงกระบวนการการค้นหาความรู้ในการทำเหมืองข้อมูล

ที่มา: https://slidetodoc.com/presentation_image_h/1f82d86e05ab9a7b6319c148d8bb3179/image-12.jpg

จากภาพที่ 2.2 แสดงการทำเหมืองข้อมูลประกอบด้วยขั้นตอนทั้งหมด 5 ขั้นตอน ดังต่อไปนี้

1. การเลือกข้อมูล (Selection) เป็นการเลือกหรือแบ่งข้อมูลตามเงื่อนไขที่กำหนด เช่น เลือกพนักงานที่มีรถยนต์เป็นของตนเอง

2. การเตรียมข้อมูล (Preprocessing) เป็นขั้นตอนในการเตรียมข้อมูลโดยการแยกข้อมูลที่ไม่มีค่าข้อมูลที่ทำการบันทึกผิดข้อมูลที่มีความซ้ำซ้อนหรือไม่สอดคล้องกันออกไปและทำการรวบรวมข้อมูลที่ต้องการซึ่งได้มาจากหลายๆ ฐานข้อมูลเข้าไว้ด้วยกัน

3. การเปลี่ยนรูปข้อมูล (Transformation) เป็นขั้นตอนการเปลี่ยนรูปหรือรวบรวมข้อมูลให้อยู่ในรูปแบบที่เหมาะสมสำหรับการทำเหมืองข้อมูล เช่น ในรูปผลรวม หรือ ผลสรุป

4. การทำเหมืองข้อมูล (Data Mining) เป็นขั้นตอนที่สำคัญเพื่อทำการกลั่นกรองรูปแบบของข้อมูล (Data Pattern) จากข้อมูลดิบที่มีในขั้นตอนนี้มีการนำเทคนิคต่าง ๆ เพื่อเข้ามาช่วยในการดึงรูปแบบ (Pattern) ที่ซ่อนอยู่ในข้อมูลออกมา

5. การแปรผลและการประเมินผล (Interpretation Evaluation) เป็นขั้นตอนที่นำรูปแบบ (Pattern) ที่ได้จากการทำเหมืองข้อมูล แปลให้อยู่ในรูปของความรู้ (Knowledge) และนำเสนอต่อผู้ใช้เพื่อใช้สนับสนุนการตัดสินใจ

2.2.1 ลักษณะการทำงานของการทำเหมืองข้อมูล

การทำเหมืองข้อมูล (Data Mining) [14] เป็นกระบวนการวิเคราะห์ข้อมูลโดยอัตโนมัติ ซึ่งระบบจะทำการเรียนรู้เพื่อสร้างรูปแบบของตนเอง สามารถแบ่งลักษณะการทำงานของการทำเหมืองข้อมูลออกเป็น 3 ประเภท ได้แก่ กระบวนการค้นหา (Discovery) การวางรูปแบบการคาดการณ์ (Predictive Modeling) และการวิเคราะห์แบบพิสูจน์หลักฐาน (Forensic Analysis)

1. กระบวนการค้นหา (Discovery) เป็นกระบวนการของการสำรวจในฐานข้อมูลเพื่อที่จะทำการค้นหารูปแบบที่ซ่อนอยู่ โดยปราศจากการกำหนดความคิด หรือสมมติฐานของรูปแบบที่ควรจะเป็นมาก่อน กล่าวอีกนัยหนึ่ง คือโปรแกรมจะทำการเรียนรู้เพื่อทำงานค้นหาว่ารูปแบบที่น่าสนใจคืออะไร โดยไม่ได้มีการคิดคำถามที่เกี่ยวข้องมาก่อนความสมบูรณ์ของรูปแบบที่ถูกแสดงและค้นพบรวมถึงคุณภาพของสารสนเทศที่ได้ออกมาจะเป็นสิ่งกำหนดความสามารถและนำไปใช้ประโยชน์ได้

2. ตัวแบบการคาดการณ์ (Predictive Modeling) เป็นรูปแบบที่ถูกค้นพบจากฐานข้อมูล เพื่อใช้ทำนายอนาคต ซึ่งวิธีนี้จะยอมรับข้อมูลนำเข้าบางแถวที่มีข้อมูลไม่ครบ และระบบจะทำการคาดคะเนค่าของข้อมูลที่ไม่ทราบเหล่านั้นบนพื้นฐานของรูปแบบที่ถูกค้นพบจากฐานข้อมูลก่อนหน้านี้ กล่าวคือกระบวนการค้นหา (Discovery) ทำการค้นหารูปแบบในข้อมูลส่วนตัวแบบการคาดการณ์จะนำรูปแบบที่ได้มาใช้ประโยชน์ สำหรับการคาดคะเนค่าสำหรับข้อมูลใหม่

3. การวิเคราะห์แบบพิสูจน์หลักฐาน (Forensic analysis) เป็นกระบวนการของการนำรูปแบบที่กรองออกมาเพื่อใช้ประโยชน์ในการค้นหาข้อมูลที่ผิดหลักหรือผิดปกติ โดยการค้นหาข้อมูลที่ไม่ปกตินี้ในขั้นแรกจะต้องหาสิ่งที่เป็นมาตรฐานและสิ่งที่เป็นแบบแผนว่าคืออะไร จากนั้นจะทำการตรวจหาข้อมูลที่ผิดปกติไปจากข้อมูลมาตรฐานที่ได้กำหนดไว้ในตอนแรก

2.2.3 หลักการในการทำเหมืองข้อมูล

การทำเหมืองข้อมูลนั้น มีหลักการในการทำงานหลัก ๆ อยู่ 3 ประการด้วยกัน คือ การจำแนกประเภท(Classification) การพยากรณ์(Prediction) และการแบ่งกลุ่มข้อมูล (Clustering) [14]

1. การจำแนกประเภท (Classification) เป็นการจำแนกประเภทของข้อมูล (Categorical Class) วิธีการนี้มีขั้นตอนการทำงาน 2 ขั้น ตอนคือ การเรียนรู้ (Learning) เพื่อให้ได้มาซึ่งฐานความรู้ (Knowledge Base) เพื่อนำความรู้ที่ได้มาใช้ต่อไปในขั้นตอนของการวิเคราะห์และพยากรณ์ ซึ่งในแต่ละกลุ่มจะมีคุณสมบัติเฉพาะของตัวเองเรียกว่าตัวจำแนกข้อมูล (Classifier) และตัวจำแนกข้อมูลเหล่านี้สามารถแยกแยะข้อมูลที่เข้ามาใหม่ได้เทคนิคสำหรับกลไกการเรียนรู้

(Machine Learning) สามารถแบ่งออกได้เป็น 2 ลักษณะ ด้วยกันคือ การนำข้อมูลที่มีอยู่มาสร้างตัวแบบการเรียนรู้เพื่อจำแนกประเภทของข้อมูล โดยใช้ตัวจำแนกข้อมูล (Classifier) เมื่อมีข้อมูลใหม่เข้ามาจะสามารถแยกแยะข้อมูลใหม่เหล่านั้นได้อย่างอัตโนมัติ การเรียนรู้ลักษณะนี้เรียกว่าการเรียนรู้แบบมีการสอน (Supervised Learning) และอีกรูปแบบหนึ่งของเทคนิคกลไกการเรียนรู้คือการจัดกลุ่ม ข้อมูลจากข้อมูลที่มีอยู่โดยไม่ต้องอาศัยข้อมูลเดิมในการสร้างตัวแบบเพื่อการเรียนรู้ก่อนโดยจะเรียกเทคนิคในลักษณะนี้ว่าการเรียนรู้แบบไม่มีการสอน (Unsupervised Learning)

2. การพยากรณ์ (Prediction) เป็นขั้นตอนของการใช้ฐานความรู้ที่ได้ (Knowledge Base) ในการวิเคราะห์และพยากรณ์ถึงสิ่งที่สนใจหรือสิ่งที่ยังไม่รู้ หรือสิ่งที่ก่อให้เกิดความผิดพลาดที่อาจเกิดขึ้นได้ และจะเป็นขั้นตอนของการทดสอบตัวแบบที่สร้างขึ้นว่า มีประสิทธิภาพมากน้อยเพียงใด โดยสามารถดูได้จากค่าความถูกต้องของการใช้ตัวแบบในการพยากรณ์

3. การแบ่งกลุ่มข้อมูล (Clustering) การแบ่งกลุ่ม คือ การรวมกลุ่มกันของข้อมูลที่มีลักษณะเหมือนกัน รูปแบบและแนวโน้มที่เหมือนกัน โดยเริ่มจากการหาตัวแทนของกลุ่มจากนั้นทำการเปรียบเทียบข้อมูลกับตัวแทนของแต่ละกลุ่ม ถ้าข้อมูลคล้ายคลึงกับตัวแทนของกลุ่มใดก็จะถูกจัดให้อยู่ในกลุ่มนั้น

2.2.4 Machine Learning (ML)

Machine Learning (ML) คือศาสตร์ที่ว่าด้วย การศึกษาและสร้างอัลกอริทึมที่สามารถเรียนรู้ข้อมูล และทำนายข้อมูลได้ [15] โดยอัลกอริทึมอาศัยการเรียนรู้จากโมเดลของข้อมูลนำเข้าเป็นตัวอย่างทำให้คอมพิวเตอร์สามารถเรียนรู้เข้าใจ และ Machine จะสามารถทำนายข้อมูลได้ ทั้งนี้ [15] ได้ให้นิยามของ Machine Learning ไว้ว่า เราจะเรียกคอมพิวเตอร์โปรแกรมว่าได้เรียนรู้จากประสบการณ์ E เพื่อทำงาน T ได้โดยมีประสิทธิผล P เมื่อโปรแกรมนั้นสามารถทำงาน I ที่วัดผลด้วย P แล้วพัฒนาขึ้นจากประสบการณ์ E ซึ่งคำนิยามที่มีชื่อเสียงและเป็นที่ยอมรับมากในแง่มุมของการดำเนินการมากกว่าความรู้สึก ทำให้เข้าใจความหมายของ Machine Learning ในแง่มุมที่ว่า Machine จะสามารถทำงานที่มนุษย์ทำได้หรือไม่ มากกว่าแค่ Machine คิดได้หรือไม่ ปัจจุบัน Machine Learning ถูกแบ่งอย่างกว้างๆ ออกเป็น 3 ประเภท โดย [16] ได้อธิบายตามลักษณะของข้อมูลนำเข้า หรือ ข้อมูลฝึก ได้แก่ การเรียนรู้จากข้อมูลแบบมีโครงสร้าง (Supervised Learning) เป็น การเรียนรู้แบบมีผู้สอน หรือจากข้อมูลตัวอย่างในอดีตที่เฉลยผลลัพธ์ที่ควรจะเป็นแสดงเป็นลาเบล (Label)ไว้ นำมาสอน Machine ให้ค้นหาความสัมพันธ์ และสร้างกฎทั่วไปไว้เพื่อทำนายว่าข้อมูลนำเข้าแบบนี้ แล้วจะทำให้ได้ข้อมูลส่งออกแบบใด เช่น Machine Learning ในเรื่องของการสอบตก และสอบผ่านของนักเรียน โดยการนำเข้าตารางข้อมูลฝึกที่มีชื่อนักเรียน คะแนน และข้อมูลเฉลยที่มีป้ายชื่อเพื่อให้ Machine ได้เรียนรู้ว่าคะแนนประมาณเท่าไรคือสอบผ่าน

หรือคะแนนต่ำประมาณเท่าไรคือสอบตก ทำให้เมื่อมีข้อมูลนำเข้าใหม่ Machine ก็จะสามารถทำนายได้ว่านักเรียนคนไหนบ้างที่สอบผ่านและมีที่สอบตกเท่าใด เป็นต้น

การเรียนรู้จากข้อมูลแบบไม่มีโครงสร้าง (Unsupervised Learning) เป็นการเรียนรู้แบบไม่มีผู้สอนหรือไม่มีข้อมูลเฉลยผลลัพธ์ใดๆ ให้ Machine ได้เรียนรู้ ซึ่ง Machine ต้องหาโครงสร้างของข้อมูลนำเข้าเอง โดยบอกแค่ความต้องการข้อมูลแบบใด เช่นต้องการแบ่งกลุ่มข้อมูล (Clustering) แต่ไม่สามารถระบุผลลัพธ์ที่จะได้ Machine จะทำการแบ่งกลุ่มข้อมูลตามเงื่อนไขที่เราระบุเท่านั้น

การเรียนรู้ในรูปแบบที่ใกล้เคียงกับการเรียนรู้ของมนุษย์มากที่สุด การเรียนรู้แบบเสริมแรง (Reinforcement Learning) คือ Machine มีปฏิสัมพันธ์กับสิ่งแวดล้อมที่เปลี่ยนแปลงตลอดเวลา โดย Machine จะต้องทำงานบางอย่างเอง โดยที่ไม่มี “ผู้สอน” คอยบอกว่าวิธีการแบบใดจะทำให้ถึงเป้าหมาย หรือวิธีการที่ทำอยู่นั้นเข้าใกล้เป้าหมายแล้วหรือไม่ ตัวอย่างเช่น การเรียนรู้เพื่อเล่นเกมแทนมนุษย์ หรือการที่รถยนต์สามารถเคลื่อนที่เอง หรือขับขี่โดยอัตโนมัติ โดย Machine จะเรียนรู้จากการกระทำหรือ Action เมื่อ Machine ทำ Action ตามลำดับต่างๆ โมเดลจะให้คะแนนดี และได้รางวัล (Reward) Machine จะเริ่มจดจำลำดับการทำ Action ที่ทำแล้วได้คะแนนดี และพยายามจะทำ Action นั้นเรื่อย ๆ การที่ Machine ได้รางวัลเท่ากับเป็นการสนับสนุนให้ Machine ทำ Action นั้นซ้ำ ๆ ยิ่ง Action นั้นได้รางวัลมาก Machine ก็จะทำแบบนั้นบ่อย ๆ สิ่งที่สำคัญของการเรียนรู้แบบเสริมแรงจึงคือการสร้างโมเดลการให้คะแนนที่ดีนั่นเอง

จากการศึกษาการทำเหมืองข้อมูล (Data mining) และ Machine Learning ผู้วิจัยได้แนวทางการทำเหมืองข้อมูลในงานวิจัยครั้งนี้ โดยดำเนินการ 5 ขั้นตอน ได้แก่ 1) การเลือกข้อมูล (Selection 2) การเตรียมข้อมูล (Preprocessing) 3)การเปลี่ยนรูปข้อมูล (Transformation) 4) การทำเหมืองข้อมูล (Data Mining) และ 5) การแปรผลและการประเมินผล (Interpretation Evaluation) และการใช้เทคนิค Machine Learning ในงานวิจัยในครั้งนี้เป็นประเภท การเรียนรู้จากข้อมูลแบบมีโครงสร้าง (Supervised Learning) เป็นการเรียนรู้แบบมีผู้สอน หรือจากข้อมูลตัวอย่างในอดีตที่เฉลยผลลัพธ์ที่ควรจะเป็นแสดงเป็นลาเบล (Label)ไว้ นำมาสอน Machine ให้ค้นหาความสัมพันธ์และสร้างกฎทั่วไปไว้ เพื่อทำนายว่าข้อมูลนำเข้าแบบนี้แล้วจะทำให้ได้ข้อมูลส่งออกแบบใด

2.3 Decision Tree

Decision Tree คือ แบบจำลองทางคณิตศาสตร์เพื่อการหาทางเลือกที่ดีที่สุด โดยการนำข้อมูลมาสร้างแบบจำลองการพยากรณ์ในรูปแบบของโครงสร้างต้นไม้ ซึ่งมีการเรียนรู้ข้อมูลแบบมีผู้สอน (Supervised Learning) สามารถสร้างแบบจำลองการจัดหมวดหมู่ (Clustering) ได้จากกลุ่มตัวอย่างของข้อมูลที่กำหนดไว้ล่วงหน้า (Training set) ได้โดยอัตโนมัติ และสามารถพยากรณ์กลุ่มของรายการที่ยังไม่เคยนำมาจัดหมวดหมู่ได้อีกด้วย

โดยปกติมักประกอบด้วยกฎในรูปแบบ “ถ้า เงื่อนไข แล้ว ผลลัพธ์” เช่น

“If Income = High and Married = No THEN Risk = Poor”

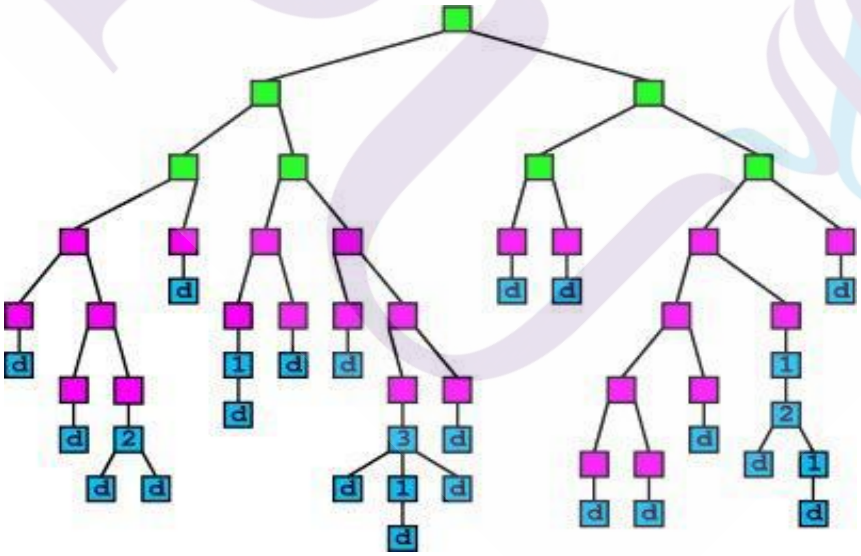
“If Income = High and Married = Yes THEN Risk = Good” [17]

2.3.1 ส่วนประกอบของต้นไม้ตัดสินใจ ประกอบด้วย [18]

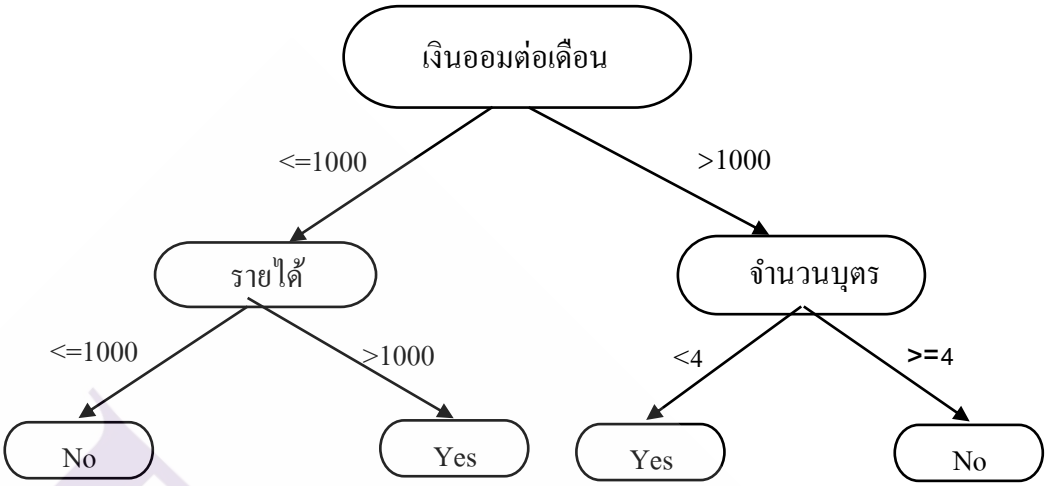
1) โหนด (Node) คือ คุณสมบัติต่าง ๆ เป็นจุดที่แยกข้อมูลว่าจะให้ไปในทิศทางใด ซึ่งโหนดที่อยู่สูงสุดเรียกว่า โหนดราก (Root Node)

2) กิ่ง (Branch) คือ คุณสมบัติของคุณสมบัติในโหนดที่แตกออกมา โดยจำนวนของกิ่งจะเท่ากับคุณสมบัติของโหนด

3) ใบ (Leaf) คือ กลุ่มของผลลัพธ์ในการแยกแยะข้อมูล โดยสามารถแสดงส่วนประกอบของต้นไม้ตัดสินใจ ดังภาพที่ 2.3



ภาพที่ 2.3 แสดงส่วนประกอบของต้นไม้ตัดสินใจ



ภาพที่ 2.4 ตัวอย่างต้นไม้ตัดสินใจ [19]

จากภาพที่ 2.4 Decition tree แสดงความสัมพันธ์ของปัจจัยที่มีผลต่อการออมเงิน > 1000 บาท หรือ ออมเงิน <= 1000 บาท โดยโหนดราก (Root Node) เป็นเงินออมต่อเดือน โดยมี 2 กิ่ง (Branch) คือ กิ่งรายได้ และ กิ่งจำนวนบุตร และมีใบ (Leaf) คือ กลุ่มของผลลัพธ์ในการแยกแยะข้อมูล คือ Yes และ No

2.3.2 การสร้างต้นไม้ตัดสินใจ หลักการพื้นฐานของการสร้างต้นไม้ตัดสินใจเป็นการสร้างในลักษณะจากบนลงล่าง (Top-Down) คือเริ่มจากการสร้างรากของต้นไม้ก่อนแล้วจึงแตกกิ่งไปจนถึงใบ โดยแสดงขั้นตอนการสร้างต้นไม้ตัดสินใจได้ดังนี้ [20]

1) ต้นไม้เริ่มต้นโดยมีโหนดเพียงโหนดเดียวแสดงถึงชุดข้อมูลฝึก (Training Set)

2) ถ้าข้อมูลทั้งหมดอยู่ในกลุ่มเดียวกันแล้ว ให้โหนดนั้นเป็นใบและตั้งชื่อแยกตามกลุ่มของข้อมูลนั้น

3) ถ้าในโหนดมีข้อมูลหลายกลุ่มปะปนอยู่ จะต้องวัดค่าเกน (Gain) ของแต่ละแอททริบิวต์เพื่อที่จะใช้เป็นเกณฑ์ในการคัดเลือกแอททริบิวต์ ที่มีความสามารถในการแบ่งแยกข้อมูลออกเป็นกลุ่มต่าง ๆ ได้ดีที่สุด โดยแอททริบิวต์ที่มีค่าเกนมากที่สุดจะถูกเลือกให้เป็นตัวทดสอบหรือแอททริบิวต์ ใช้ในการตัดสินใจ โดยแสดงในรูปของโหนดบนต้นไม้

4) กิ่งของต้นไม้ ถูกสร้างขึ้นจากค่าต่าง ๆ ที่เป็นไปได้ของโหนดทดสอบ และข้อมูลจะถูกแบ่งออกตามกิ่งต่าง ๆ ที่สร้างขึ้น

5) ทำการวนซ้ำเพื่อหาแอททริบิวต์ที่มีค่าเกนมากที่สุด สำหรับข้อมูลที่ถูกแบ่งแยกออกมาในแต่ละกิ่งเพื่อนำแอททริบิวต์นี้มาสร้างเป็นโหนดตัดสินใจต่อไป โดยที่แอททริบิวต์ที่ถูกเลือกมาเป็นโหนดแล้วจะไม่ถูกเลือกมาอีก สำหรับโหนดในระดับต่อ ๆ ไป

6) ทำการวนซ้ำเพื่อแบ่งข้อมูลและแตกกิ่งของต้นไม้ไปเรื่อย ๆ โดยการวนซ้ำจะสิ้นสุดก็ต่อเมื่อเงื่อนไขข้อใดข้อหนึ่งต่อไปนี้เป็นจริง

2.4 Naïve Bayes [21]

Naïve Bayes เป็น Supervised Learning ประเภท Classification รูปแบบหนึ่งที่ได้รับความนิยมเป็นอย่างมากในปัจจุบัน เนื่องจากสามารถเทรนโมเดลโดยใช้จำนวนชุดของ Training data ไม่มาก แต่ได้ความแม่นยำในระดับที่น่าพอใจ รวมถึงอิมพลีเมนต์ง่าย หลักการของวิธีการนี้จะใช้การคำนวณความน่าจะเป็นแบบมีเงื่อนไขที่เรียกว่า Conditional Probability ซึ่งแสดงดังสมการที่ (1)

$$P(A|\mathrm{B}) = \frac{P(A \cap B)}{P(B)} \qquad \text{.....(1)}$$

โดย $P(A|B)$ คือ ค่า Conditional Probability หรือค่าความน่าจะเป็นที่เกิดเหตุการณ์ B ขึ้นก่อนและจะมีเหตุการณ์ A ตามมา

$P(A \cap B)$ คือ ค่า Joint Probability หรือค่าความน่าจะเป็นที่เหตุการณ์ A และเหตุการณ์ B เกิดขึ้นร่วมกัน

$P(B)$ คือ ค่าความน่าจะเป็นที่เหตุการณ์ B เกิดขึ้นในลักษณะเดียวกันเราจะเขียน $P(B|A)$ หรือค่าความน่าจะเป็นที่เหตุการณ์ A เกิดขึ้นก่อนและเหตุการณ์ B เกิดขึ้นตามมาทีหลังได้เป็นสมการที่ (2)

$$P(B|\mathrm{A}) = \frac{P(B \cap A)}{P(A)} \qquad \text{…….....(2)}$$

จากทั้งสองสมการจะเห็นว่ามีค่า $P(A \cap B)$ ที่เหมือนกันอยู่ ดังนั้นเราสามารถเขียนสมการของ $P(A \cap B)$ ได้เป็นดังสมการที่ (3)

$$P(A \cap B) = P(A|\mathrm{B}) \times \mathrm{P(B)} = \mathrm{P(B|A)} \times \mathrm{P(A)}$$

$$P(B|\mathrm{A}) = \frac{P(A|B) \times P(B)}{P(A)} \quad \text{.....(3)}$$

เมื่อนำทฤษฎีของเบย์มาใช้ในงานทางด้าน Data Mining มักจะเปลี่ยนสัญลักษณ์ B เป็น C โดยให้ A คือ แอตทริบิวต์ (Attribute) และ C คือ คลาส (Class) ดังสมการที่ (4)

$$P(C|\mathrm{A}) = \frac{P(A|C) \times P(C)}{P(A)} \quad \text{.....(4)}$$

โดย Posterior probability หรือ P(C|A) คือ ค่าความน่าจะเป็นที่ข้อมูลที่มีแอตทริบิวต์เป็น A จะมีคลาส C

Likelihood หรือ P(A|C) คือ ค่าความน่าจะเป็นที่ข้อมูล Training data ที่มีคลาส C และมีแอตทริบิวต์ A โดยที่ $A = a_1 \cap a_2 \ldots \cap a_m$ โดยที่ m คือจานวนแอตทริบิวต์ใน Training data Prior probability หรือ P(C) คือ ค่าความน่าจะเป็นของคลาส C

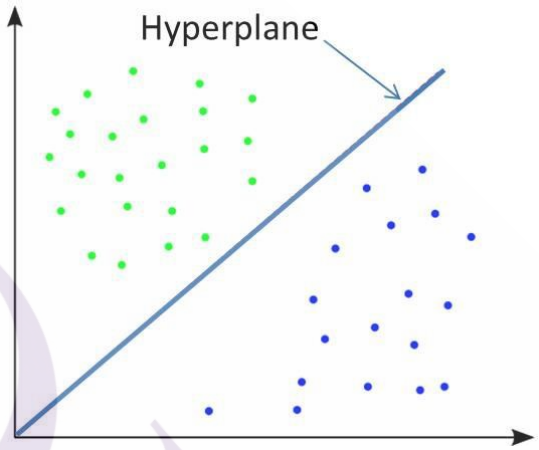
แต่การที่แอตทริบิวต์ $A = a_1 \cap a_2 \ldots \cap a_m$ ที่เกิดขึ้นใน training data อาจจะมีจำนวนน้อยมากหรือไม่มีรูปแบบของแอตทริบิวต์แบบนี้เกิดขึ้นเลย ดังนั้นจึงได้ใช้หลักการที่ว่าแต่ละแอตทริบิวต์เป็นอิสระต่อกัน ทำให้สามารถเปลี่ยนสมการ P(A|C) ได้เป็นสมการที่ (5)

$$P(A|C) = P(a_1|\mathrm{C}) \times \mathrm{P}(a_2|\mathrm{C}) \times \ldots \times \mathrm{P}(a_m|\mathrm{C}) \quad \text{.....(5)}$$

2.5 Support Vector Machine

Support Vector Machine [22] เป็นการใช้สมการเส้นตรงเพื่อแบ่งเขตข้อมูล 2 กลุ่มออกจากกัน ซึ่งหลักการของ Support Vector Machine คือการหาสัมประสิทธิ์ของสมการ เพื่อสร้างเส้นจำแนกประเภทข้อมูลในขั้นตอนการเรียนรู้ และเลือกเส้นจำแนกประเภทข้อมูลที่เหมาะสมที่สุด โดยพยายามให้ระยะห่างระหว่างขอบเขตของทั้ง 2 กลุ่มมีระยะห่างมากที่สุด เพื่อลดความผิดพลาดในการจำแนกประเภท เนื่องจากถ้าระยะห่างยิ่งมากความผิดพลาดในการจำแนกประเภทก็จะมีโอกาสเกิดขึ้นน้อยลง

แนวความคิดของ Support Vector Machine เกิดจากการที่นำค่าของกลุ่มข้อมูลมาวางลงในฟีเจอร์สเปซ (Feature Space) จากนั้นจึงหาเส้นที่ใช้แบ่งข้อมูลทั้งสองออกจากกันโดยจะสร้างเส้นแบ่ง (Hyperplane) ที่เป็นเส้นตรงขึ้นมา และเพื่อให้ทราบว่าเส้นตรงที่แบ่งสองกลุ่มออกจากกันนั้น เส้นตรงใดเป็นเส้นที่ดีที่สุด

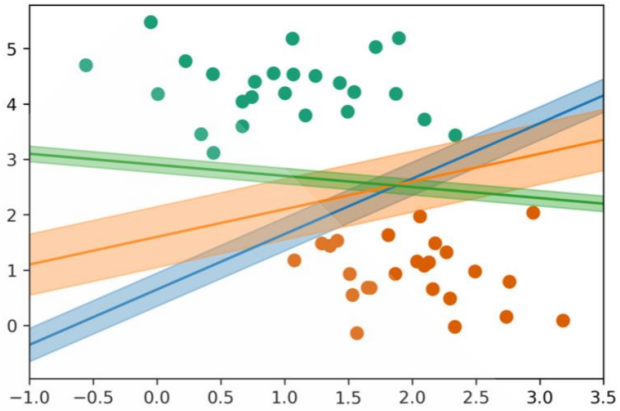


ภาพที่ 2.5 เส้นตรงที่ใช้แบ่งกลุ่มข้อมูล (Hyperplane)

ที่มา: https://miro.medium.com/max/1090/1*MrQ8LT-Zazt3kWgca7oKRg.jpeg

Max-Margin and Support Vectors

การแบ่งข้อมูลสามารถแบ่งได้หลายเส้นแต่จะเลือกเส้นที่มี Margin มากที่สุด คือ เส้นที่มีระยะแบ่งกว้างที่สุด เช่น ภาพที่ 2.5 เส้นสีส้มมีระยะมากที่สุด หาก Margin แคบไปขยับข้อมูลเดียวอาจจะทำให้ข้ามไปอีกฝั่งหนึ่งได้เลยทำให้มีโอกาส Overfit สูง ดังนั้น เราจะเลือก Margin ที่มาค่ามาก ทำให้ เกิด Over fit น้อย หรือเรียกกว่า Soft Margin

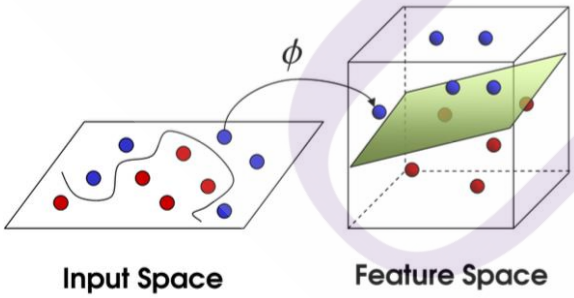


ภาพที่ 2.6 เปรียบเทียบค่า Margin เส้นสีส้มมีระยะมากที่สุด

ที่มา: https://miro.medium.com/max/700/1*28t84GRRb_QuuJKJ7_uCbw.png

Kernels

หากข้อมูลไม่สามารถแบ่งกลุ่มได้ด้วยเส้นตรง (Linear) จึงได้มีวิธีการ Kernels ที่เป็น non-linear เข้ามาแก้ไขปัญหาที่เกิดขึ้น โดยวิธีการคือ สร้างมิติขึ้นมาจากเดิม 2D เป็น 3D แล้วลากเส้นตัดผ่านตรงกลางจะทำให้สามารถแบ่งข้อมูลออกไปกลุ่มได้ ดังภาพที่ 2.7



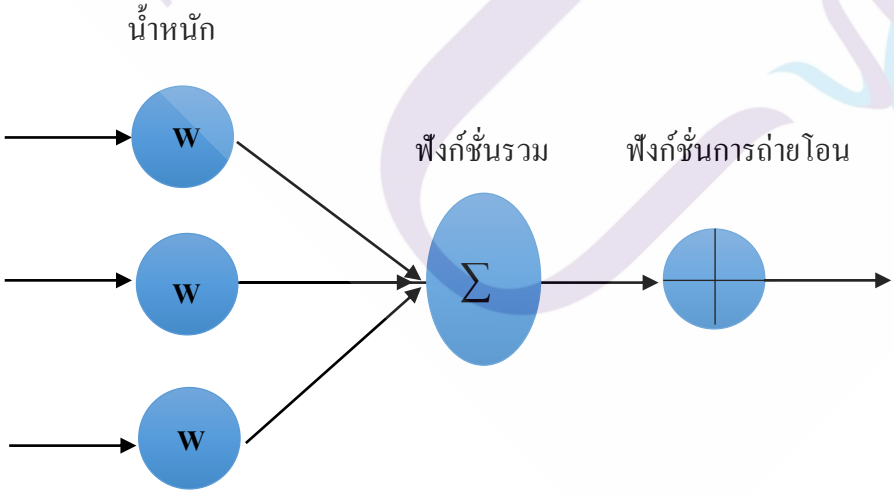
ภาพที่ 2.7 ตัวอย่างการสร้างเส้น Kernels จาก 2D เป็น 3D

ที่มา: https://miro.medium.com/max/624/1*LSbiL794r7wcZ8PK0KoO2g.png

2.6. Neural Network

Neural Network ถูกพัฒนาขึ้นเพื่อเลียนแบบความสามารถของระบบประสาททางชีวภาพ โครงสร้างของข่ายงานระบบประสาทมีด้วยกันหลายโครงสร้าง แต่ทุก ๆ โครงสร้างก็มีข้อดีที่สำคัญร่วมกัน เช่น ลักษณะที่สำคัญที่สุดของข่ายงานระบบประสาท คือ ความสามารถในการประมาณค่าฟังก์ชั่นต่อเนื่องแบบไม่เป็นเชิงเส้น (Nonlinear Continuous Function) ในระดับความแม่นที่ต้องการได้ ด้วยคุณลักษณะนี้ ข่ายงานระบบประสาทจึงถูกนำไปใช้ในการหาแบบจำลองของระบบไม่เป็นเชิงเส้นเพื่อนำไปประยุกต์ใช้ในการสังเคราะห์หาตัวควบคุมต่อไป [23]

โครงสร้างประสาทเทียมมีคุณลักษณะคล้ายกับการส่งผ่านสัญญาณประสาทในสมองของมนุษย์ มีความสามารถในการรวบรวมความรู้ (Knowledge) โดยผ่านกระบวนการเรียนรู้ (Learning Process) โดยที่ความรู้เหล่านั้นจะจัดเก็บในโครงข่ายในรูปแบบค่าน้ำหนัก (Weight) ซึ่งสามารถปรับเปลี่ยนค่าได้เมื่อมีการเรียนรู้สิ่งใหม่ ๆ เข้าไป โดยการประมวลผลต่าง ๆ จะเกิดขึ้นในหน่วยประมวลผลย่อย เรียกว่า โหนด (Node) เป็นการจำลองลักษณะการทำงานมาจากเซลล์การส่งสัญญาณ (Signal) ระหว่างโหนดที่เชื่อมต่อกัน (Connection) จำลองมาจากการเชื่อมต่อของเดนไดร์ทและแอคซอนในระบบประสาทของมนุษย์ ภายในโหนดจะมีฟังก์ชันกำหนดสัญญาณส่งออกที่ฟังก์ชันการถ่ายโอน (Transfer Function) ซึ่งทำหน้าที่เปรียบเสมือนกระบวนการทำงานในเซลล์ [24] ดังภาพที่ 2.8



ภาพที่ 2.8 โครงสร้างการทำงานของโครงข่ายประสาทเทียม

โครงข่ายประสาทเทียม ประกอบด้วยองค์ประกอบ 5 องค์ประกอบ ดังนี้

1) ข้อมูลป้อนเข้า (Input) เป็นข้อมูลตัวเลข หากเป็นข้อมูลเชิงคุณภาพต้องแปลงให้อยู่ในรูปปริมาณที่โครงข่ายประสาทเทียมยอมรับได้

2) ข้อมูลส่งออก (Output) คือผลลัพธ์ที่เกิดขึ้นจากกระบวนการเรียนรู้ของโครงข่ายประสาทเทียม

3) ค่าน้ำหนัก (Weights) คือสิ่งที่ได้จากการเรียนรู้ของโครงข่ายประสาททียม ค่านี้จะถูกเก็บเป็นทักษะเพื่อใช้ในการจดจำข้อมูลอื่นๆ ที่อยู่ในรูปแบบเดียวกัน

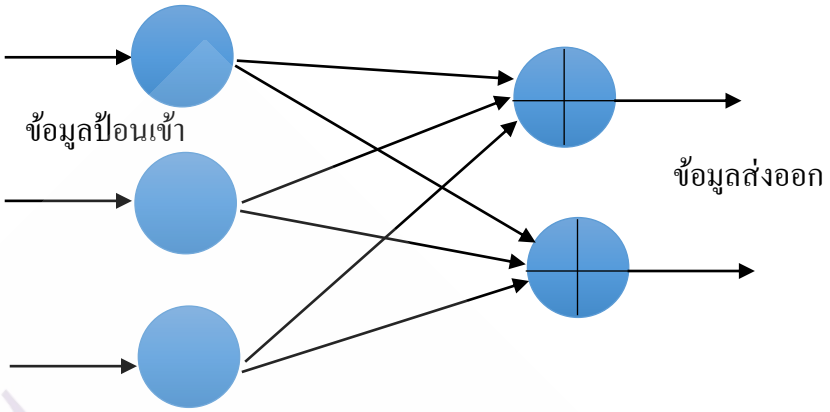
4) ฟังก์ชั่นรวม (Combination Function) เป็นการรวมผลของข้อมูลป้อนเข้า (Input) โดยขึ้นกับค่าน้ำหนัก (Weights) ของข้อมูลป้อนเข้า (Input) แต่ละตัว

5) ฟังก์ชันการถ่ายโอน (Transfer Function) เป็นการคำนวณการจำลองการทำงานของโครงข่ายประสาทเทียม โดยคำนวณค่าข้อมูลส่งออก (Output) ที่ได้จากฟังก์ชั่นรวม (Combination Function) ก่อนหน้านั้น [25]

โครงข่ายประสาทเทียมมีเซลล์ประสาทเทียม หรือโหนดจำนวนมากเชื่อมต่อกัน ซึ่งการเชื่อมต่อแบ่งออกเป็นกลุ่มย่อย เรียกว่า ชั้น (Layer) ชั้นแรก เป็นชั้นนำข้อมูลเข้า เรียกว่า ชั้นรับข้อมูลป้อนเข้า (Input layer) ส่วนชั้นสุดท้ายเรียกว่า ชั้นส่งข้อมูลออก (Output Layer) และชั้นที่อยู่ระหว่างชั้นรับข้อมูลป้อนเข้าและชั้นส่งข้อมูลออก เรียกว่า ชั้นแฝง (Hidden Layer) โดยทั่วไปชั้นแฝงอาจมีมากกว่า 1 ชั้นก็ได้ ด้วยเหตุนี้ จึงสามารถแบ่งประเภทของโครงสร้างประสาทเทียมตามจำนวนชั้นของโครงข่ายแบบกว้าง ๆ ได้ 2 แบบ ได้แก่ โครงข่ายแบบชั้นเดียว (Single Layer) และโครงข่ายแบบหลายชั้น (Multi Layer) [24]

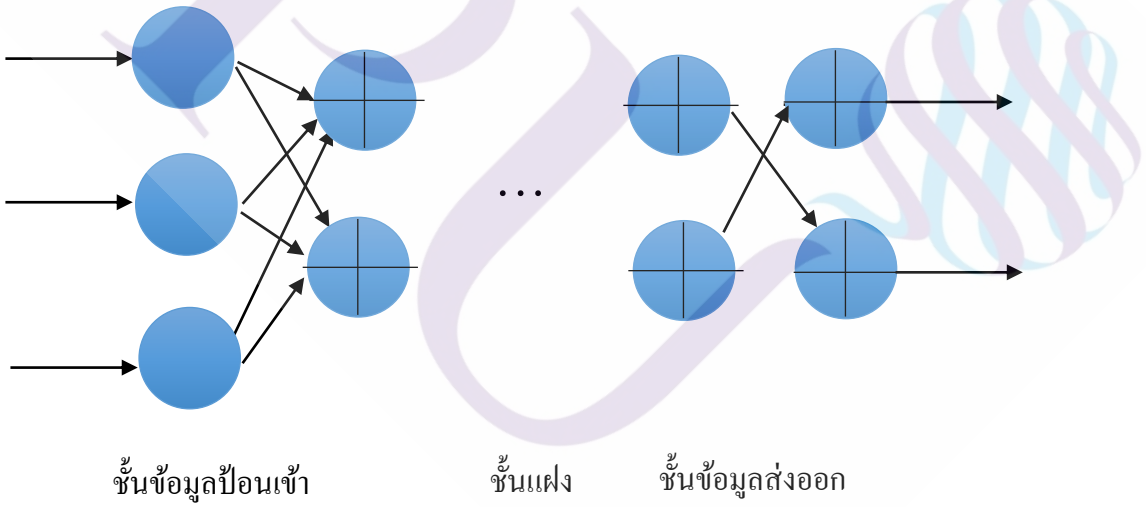
1) โครงข่ายแบบชั้นเดียว เป็นโครงข่ายประสาทเทียมอย่างง่ายที่มีเพียงชั้นรับข้อมูลป้อนเข้าและชั้นส่งข้อมูลออกเท่านั้น โหนดในชั้นรับข้อมูลป้อนเข้าทำหน้าที่รับข้อมูลเข้า(Input value) แล้วส่งข้อมูลผ่านเส้นเชื่อมโยงต่าง ๆ ไปให้โหนดในชั้นส่งข้อมูลออก ความเข้มของสัญญาณ หรือปริมาณข้อมูลที่นำเข้าสู่โหนดในชั้นส่งข้อมูลออก จะขึ้นอยู่กับค่าน้ำหนักที่อยู่บนเส้นเชื่อมโยง

โหนดในชั้นส่งข้อมูลออก จะนำข้อมูลที่ได้รับมาคำนวณโดยใช้ฟังก์ชันทางคณิตศาสตร์ที่เรียกว่า ฟังก์ชันการถ่ายโอน (Transfer Function) ที่เหมาะสมกับปัญหา แล้วส่งผลลัพธ์ที่ได้ออกมาเป็นข้อมูลส่งออก โดยลักษณะโครงข่ายแบบชั้นเดียว ดังภาพที่ 2.9



ภาพที่ 2.9 โครงสร้างโครงข่ายประสาทเทียมแบบชั้นเดียว

2) โครงข่ายแบบหลายชั้น เป็นโครงข่ายที่มีชั้นแฝง (Hidden Layer) ตั้งแต่ 1 ชั้นขึ้นไป โครงข่ายแบบหลายชั้นจะใช้ในกรณีที่ปัญหามีความซับซ้อน ซึ่งโครงข่ายแบบชั้นเดียวไม่สามารถแก้ปัญหาได้ จึงเพิ่มจำนวนโหนดที่มีการคำนวณ หรือชั้นแฝง (Hidden Layer) ให้กับโครงข่าย ตัวอย่างโครงข่ายแบบหลายชั้น โดยลักษณะโครงข่ายแบบหลายชั้น ดังภาพที่ 2.10



ภาพที่ 2.10 โครงสร้างโครงข่ายประสาทเทียมแบบหลายชั้น

ประโยชน์จากการประยุกต์ใช้โครงข่ายประสาทเทียม มีจำนวน 3 ประการ [26] คือ

1) ความทนจากการที่สามารถออกแบบให้มีจำนวนโหนดที่มากมายได้ และหากโหนดใดโหนดหนึ่งถูกทำลายก็จะไม่ทำให้เน็ตเวิร์คหยุดทำงานไป

2) ความสามารถในการปรับตัวให้เข้ากับสิ่งแวดล้อมใหม่

3) ความสามารถในการบ่งชี้ลักษณะทั่วไป ซึ่งเป็นผลมาจากความสามารถในการปรับตัวได้ดี จึงทำให้โครงข่ายประสาทเทียมสามารถจัดการกับข้อมูลที่ไม่สมบูรณ์ได้เป็นอย่างดี และข้อได้เปรียบของโครงข่ายประสาทเทียม คือ

แบบจำลองสามารถเรียนรู้ตัวแปรป้อนเข้า (Input) ได้ดีไม่ว่าตัวแปรนั้นจะเป็นตัวแปรอันตรภาค (Interval Variables) ตัวแปรนามบัญญัติ (Norminal Variables) หรือตัวแปรแบบทวิลักษณ์ หรือ ตัวแปรฐานสอง (Binary Variables) โดยที่ตัวแปรป้อนเข้า (Input) ไม่มีการกระจายตัวแบบ Normal สำหรับตัวแปรอันตรภาค (Interval Variables) และความสัมพันธ์ระหว่างตัวแปรป้อนเข้า (Input) และตัวแปรเป้าหมายไม่เป็นเชิงเส้นตรง [25]

โครงข่ายประสาทเทียมสำหรับการประมาณค่าและการทำนาย [40]

(Neural Networks for Estimation and Prediction)

เนื่องจากโครงข่ายประสาทเทียมให้ผลลัพธ์แบบต่อเนื่อง โครงข่ายประสาทเทียมอาจจะใช้สำหกรับการประมาณค่าและการทำนาย ตัวอย่างเช่น สมมติว่าเราสนใจในการทำนายราคาของสต็อกสินค้าชนิดนึ่งในอีก 3 เดือน ข้างหน้า เราลงรหัสราคาโดยใช้คะแนนปกติมาตรฐานเดียวกันน้อยที่สุด-มากที่สุด ตามข้างบนนี้ โดยโครงข่ายประสาทเทียมมีค่าผลลัธ์อยู่ระหว่าง 0 และ 1 จะต้องแปลงคะแนนปกติมาตรฐานเดียวกันน้อยที่สุด-มากที่สุดเพื่อทำให้ผลลัพธ์โครงข่ายประสาทเทียมสามารถเข้าใจสเกลราคาของสต็อกสินค้า โดยทั่วไปใช้การทำให้อยู่ในรูปปกติมาตรฐานเดียวกันอีกครั้งหนึ่ง (denormalization) โดยมีสูตรดังนี้

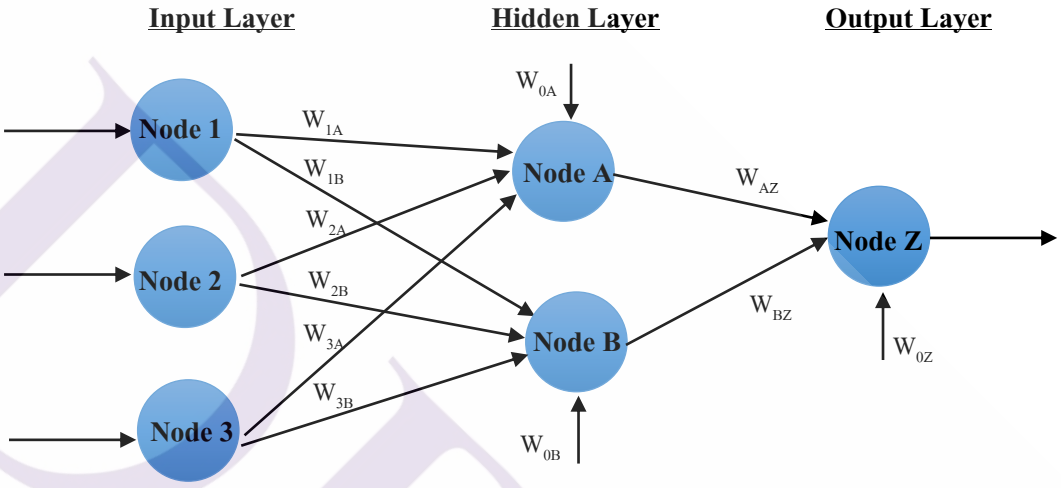
การทำนาย = ผลลัพธ์ (พิสัยของข้อมูล) + ค่าน้อยที่สุด

โดยที่ผลลัพธ์แทนผลลัพธ์โครงข่ายประสาทเทียมมีพิสัยในช่วง (0, 1) พิสัยของข้อมูลแทนพิสัยของค่าคุณลักษณะเริ่มต้นบนสเกลที่ไม่ได้อยู่ในรูปปกติมาตรฐาน และค่าน้อยที่สุดแทนค่าคุณลักษณะน้อยที่สุดบนสเกลที่ไม่ได้อยู่ในรูปปกติมาตรฐาน ตัวอย่างเช่น สมมติว่าราคาของสต็อกสินค้ามีพิสัยจาก 20 ถึง 30 ดอลล่าร์ และผลลัพธ์โครงข่ายประสาทเทียมเป็น 0.69 ดังนั้นราคาของสต็อกสินค้าที่ทำนายได้ในอีก 3 เดือน ข้างหน้า คือ

การทำนาย = ผลลัพธ์ (พิสัยของข้อมูล) + ค่าน้อยที่สุด

= 0.69 (10 ดอลล่าร์) + 20 ดอลล่าร์ = 26.90 ดอลล่าร์

ตัวอย่างของโครงข่ายประสาทเทียมอย่างง่าย (Simple Example of a Neural Network)



ภาพที่ 2.11 โครงสร้างโครงข่ายประสาทเทียมอย่างง่าย [40]

จากการตรวจสอบโครงข่ายประสาทเทียมอย่างง่ายดังแสดงในภาพที่ 2.11 โครงข่ายประสาทเทียมประกอบด้วย 3 ประเภท คือ โครงข่ายแบบชั้น (layered network) โครงข่ายแบบไปข้างหน้า (feed forward network) และโครงข่ายแบบเชื่อมต่อกันอย่างสมบูรณ์ (completely connected network) ของเซลล์ประสาท โครงข่ายแบบไปข้างหน้าเป็นโครงข่ายประสาทเทียมที่มีการจำกัดทิศทางการเคลื่อนที่เป็นแบบทางเดียวหรือเคลื่อนที่ไปข้างหน้าและไม่ได้เคลื่อนที่เป็นวงหรือวงกลมโครงข่ายประสาทเทียม ประกอบด้วย 2 ชั้น (layer) หรือมากกว่า 2 ชั้น โครงข่ายประสาทเทียมส่วนใหญ่ประกอบด้วย 3 ชั้น คือ ชั้นข้อมูลเข้า (input layer) ชั้นซ่อน (hidden layer) และชั้นข้อมูลออกหรือชั้นผลลัธ์ (output layer) โดยอาจจะมีชั้นซ่อนมากกว่า 1 ชั้น แม้ว่าโครงข่ายประสาทเทียมส่วนใหญ่มีชั้นซ้อนเพียง 1 ชั้นซึ่งเพียงพอสำหรับวัตถุประสงค์โดยส่วนใหญ่ โครงข่ายแบบเชื่อมต่อกันอย่างสมบูรณ์เป็นโครงข่ายประสาทเทียมที่โหนดทุกโหนดในชั้นที่กำหนดเชื่อมต่อกันทุกโหนดกับชั้นถัดไป แม้ว่าจะไม่ได้เชื่อมต่อกับโหนดอื่นใดในขณะในชั้นเดียวกัน การเชื่อมต่อกันระหว่างโหนดมีการถ่วงน้ำหนัก (ได้แก่ W_{1A}) ที่สัมพันธ์กันในขณะเริ่มต้นการถ่วงน้ำหนักถูกจัดอย่างสุ่มโดยมีค่าอยู่ระหว่าง 0 และ 1

โดยทั่วไปจำนวนของโหนดข้อมูลเข้าขึ้นอยู่กับจำนวนและชนิดของคุณลักษณะในชุดข้อมูล จำนวนของชั้นซ่อนและจำนวนของโหนดในชั้นซ่อนขึ้นอยู่กับผู้ใช้งานเป็นผู้กำหนด จำนวนของโหนดในชั้นข้อมูลออกอาจจะมีมากกว่า 1 โหนด ซึ่งขึ้นอยู่กับงานในการจำแนกกลุ่ม

คำถามก็คือควรจะมีโหนดในชั้นซ่อนจำนวนมากน้อยเท่าไร เนื่องจากโหนดที่เพิ่มขึ้นในชั้นซ่อนให้กำลัง (power) และความคล่องตัว(flexibility) ของโครงข่ายประสาทเทียมเพิ่มขึ้นสำหรับกำหนดรูปแบบที่ซับซ้อน เราอาจจะทดสอบเพื่อให้มีโหนดจำนวนมากในชั้นซ่อน แต่ชั้นซ่อนขนาดใหญ่เกินไปจะนำไปสู่ความซับซ้อนของตัวแบบมากเกินความจำเป็น (over fitting) ถ้าเกิดความซับซ้อนของตัวแบบมากเกินความจำเป็น สิ่งหนึ่งอาจจะพิจารณาให้ลดจำนวนของโหนดในชั้นซ่อน ในทางตรงกันข้ามถ้าความถูกต้องในการฝึกหัดต่ำจนไม่สามารถยอมรับได้ สิ่งหนึ่งอาจจะพิจารณาให้เพิ่มจำนวนของโหนดในชั้นซ่อน

ชั้นข้อมูลเข้าอนุญาตให้ข้อมูลเข้าผ่าน ได้แก่ ค่าคุณลักษณะ (attribute) ค่คุณลักษณะเหล่านี้จะผ่านไปตามชั้นซ่อนโดยปราศการประมวลผลข้อมูลเพิ่มเติม ดังนั้นโหนดในชั้นข้อมูลเข้าไม่ได้แบ่งโครงสร้างของโหนดในเชิงรายละเอียดเหมือนกับที่โหนดในชั้นซ่อนและโหนดในชั้นข้อมูลออกแบ่งโครงสร้างของโหนดในเชิงรายละเอียด

2.7 การวิเคราะห์ความถดถอย (Regression Analysis) [27]

การวิเคราะห์การถดถอยเป็นกระบวนการทางสถิติเพื่อให้ได้สมการถดถอยสำหรับทำนายปรากฏการณ์ต่าง ๆ สิ่งที่ถูกทำนายเรียกว่าตัวแปรเกณฑ์หรือตัวแปรตาม ตัวแปรทำนายคือตัวแปรอิสระ ในกระบวนการนี้ ตัวแปรตามหรือตัวแปรเกณฑ์จะมีเพียงตัวเดียว ส่วนตัวแปรอิสระหรือตัวแปรทำนายจะมีกี่ตัวก็ได้ ถ้าหากมีตัวเดียว จะเรียกว่าการถดถอยอย่างง่าย (Simple Regression) หากมีตั้งแต่ 2 ตัวขึ้นไป จะเรียกว่า การถดถอยพหุคูณ (Multiple Regression)

วิธีการคัดเลือกตัวแปร

วิธีการคัดเลือกตัวแปรเข้าสมการเพื่อให้สมการสามารถทำนายตัวแปรเกณฑ์ได้สูงสุดมีวิธีการคัดเลือกตัวแปรดังนี้

2.7.1 วิธีการเลือกแบบก้าวหน้า (Forward Selection)

วิธีการนี้จะเป็นการเลือกตัวแปรทำนายที่มีสหสัมพันธ์กับตัวแปรตามสูงที่สุดเข้าสมการก่อนส่วนตัวแปรที่เหลือจะมีการคำนวณค่าสัมประสิทธิ์การถดถอยในรูปคะแนนมาตรฐาน ค่า t-test ทดสอบนัยสำคัญของค่าสัมประสิทธิ์การถดถอยในรูปคะแนนมาตรฐาน และค่าสหสัมพันธ์แบบแยกส่วน (partial correlation) โดยเป็นความสัมพันธ์เฉพาะตัวแปรที่เหลือตัวนั้นกับตัวแปรตามโดยขจัดอิทธิพลของตัวแปรอื่น ๆ ออก ถ้าตัวแปรใดมีค่าสัมประสิทธิ์การถดถอยมีนัยสำคัญทาง

สถิติก็จะนำเข้าสมการต่อไป จะทำแบบนี้จนกระทั่งตัวแปรที่เหลืออยู่ไม่มีนัยสำคัญทางสถิติ ก็จะหยุดการคัดเลือกและได้สมการที่มีสัมประสิทธิ์การทำนายสูงสุด

2.7.2 วิธีการเลือกแบบถอยหลัง (Backward Selection)

วิธีการนี้เป็นการนำตัวแปรทำนายทั้งหมดเข้าสมการ จากนั้นก็จะค่อย ๆ ขจัดตัวแปรทำนายออกทีละตัว โดยพิจารณาจากค่าสัมประสิทธิ์การถดถอยของตัวแปรทำนายที่อยู่ในสมการ หากทดสอบแล้วพบว่าไม่มีนัยสำคัญทางสถิติ ก็จะขจัดออกจากสมการ ถ้ามีหลายตัวแปรที่ไม่มีนัยสำคัญทางสถิติ โปรแกรมจะเลือกตัวแปรที่มีค่าสถิติ t-test ต่ำสุด ค่านัยสำคัญสำคัญสูงสุดออกจากสมการ แล้วดำเนินการทดสอบตัวแปรที่เหลืออยู่ในสมการต่อไป จนกระทั่งตัวแปรทำนายแต่ละตัวมีนัยสำคัญทางสถิติ ก็จะหยุดการคัดเลือก และได้สมการการทดสอบที่มีสัมประสิทธิ์การทำนายสูงสุด

2.7.3 การคัดเลือกแบบลำดับขั้น (Stepwise Selection)

การคัดเลือกแบบนี้เป็นการผสมผสานระหว่างวิธีการคัดเลือกตัวแปรทำนายทั้งสองวิธีที่กล่าวมาแล้วเข้าด้วยกัน ในขั้นแรกจะเลือกตัวแปรทำนายที่มีสหสัมพันธ์กับตัวแปรตามสูงที่สุดเข้าสมการก่อน จากนั้นก็จะทดสอบตัวแปรที่ไม่ได้อยู่ในสมการว่าจะมีตัวทำนายตัวใดบ้างมีสิทธิ์เข้ามาอยู่ในสมการด้วยวิธีการคัดเลือกแบบก้าวหน้า (Forward Selection) และขณะเดียวกันก็จะทดสอบตัวแปรทีอยู่ในสมการด้วยว่าตัวแปรทำนายที่อยู่ในสมการตัวแปรใดมีโอกาสที่จะถูกขจัดออกจากสมการด้วยวิธีการคัดเลือกแบบถอยหลัง (Backward Selection) โดยจะกระทำการคัดเลือกผสมทั้งสองวิธีนี้ในทุกขั้นตอนจนกระทั่งไม่มีตัวแปรใดที่ถูกคัดออกจากสมการ และไม่มีตัวแปรใดที่จะถูกนำเข้าสมการ กระบวนการก็จะยุติและได้สมการถดถอยที่มีสัมประสิทธิ์การทำนายสูงสุด

ในการวิจัยครั้งนี้ผู้วิจัยใช้วิธีการวิธีการคัดเลือกตัวแปรแบบการคัดเลือกแบบลำดับขั้น (Stepwise Selection) ซึ่งมีความเหมาะสมและมีประสิทธิภาพต่องานวิจัย สอดคล้องกับ [28] ได้เสนอข้อแตกต่างจากผลการวิเคราะห์ข้อมูล พบว่าสมการพยากรณ์ที่วิเคราะห์ด้วยวิธี Enter จะได้สมการที่แตกต่างจากการวิเคราะห์ในอีก 3 วิธี คือ ภายในสมการพยากรณ์จะมีตัวแปรพยากรณ์อยู่ครบทุกตัว ทำให้มองเห็นตัวแปรทั้งที่พยากรณ์ได้และไม่ได้ ซึ่งวิธีการนี้จะมีประโยชน์กับการวิเคราะห์ข้อมูลในลักษณะที่ผู้วิจัยต้องการให้ตัวแปรทุกตัวอยู่ในสมการครบ เช่น ในการวิเคราะห์เส้นทางแบบ PAQ ที่ผู้วิจัยคำนึงถึงลำดับการพยากรณ์ของตัวแปร หรือเส้นทางการส่งผ่านอิทธิพลระหว่างตัวแปรพยากรณ์แต่ละตัว โดยสมการพยากรณ์ในอีก 3 วิธี ที่เหลือ คือ Forward, Backward และ Stepwise จะเป็นการวิเคราะห์ที่ได้โมเดลประหยัด และ ผลที่ได้ในสมการพยากรณ์ของทั้ง 3 วิธี ได้ผลที่เหมือนกัน เนื่องจากในวิธี Forward และ Backward เป็นการคัดเลือกตัวแปรพยากรณ์ที่ดีที่สุดที่สามารถพยากรณ์ตัวแปรตามได้เช่นเดียวกันทั้งสองวิธี ต่างกันเพียงขั้นตอนการนำเข้าและ

ถอยตัวแปรออกแต่โดยหลักการอื่น ๆ แล้วมีวัตถุประสงค์เหมือนกัน ส่งผลให้ 2 วิธีนี้ได้ผลลัพธ์ในสมการที่มีค่าเหมือนกัน ส่วนวิธี Stepwise เนื่องจากในการนำเสนอการวิเคราะห์ข้อมูลโมเดลที่ได้ไม่มีการถอยตัวแปรออกหลังจากมีการนำตัวแปรพยากรณ์เข้าในแต่ละรูปแบบทำให้โมเดลท้ายสุดที่ได้ผลออกมามีค่าเท่ากับวิธี Forward และ Backward แต่โดยหลักการวิธี Stepwise ถือเป็นวิธีการคัดเลือกตัวแปรเข้าสู่สมการถดถอยพหุคูณแบบโมเดลประหยัดที่ดีที่สุดเนื่องจากว่า

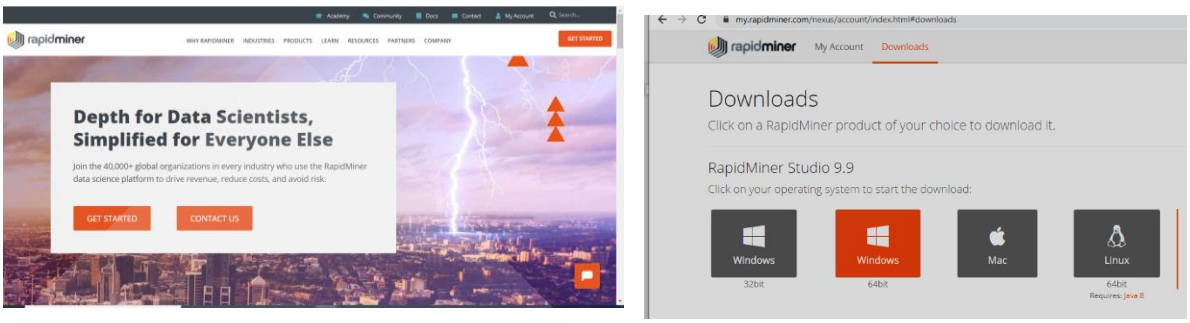
- วิธี Forward เป็นการนำตัวแปรเข้าวิเคราะห์ทีละตัวและมีการตรวจสอบว่าตัวแปรนั้นจะอยู่ในรูปแบบต่อไปหรือไม่ ถ้าอยู่ก็จะทำการวิเคราะห์ขั้นต่อไปเรื่อย ๆ จนกระทั่งไม่มีตัวแปรพยากรณ์เหลือ วิธีการเพิ่มตัวแปรก็จะสิ้นสุด ถึงแม้ว่าวิธีการนี้จะทำให้ได้โมเดลประหยัดก็จริง แต่ในขั้นตอนการนำตัวแปรพยากรณ์ เข้าสมการจะพิจารณาเฉพาะตัวแปรที่เข้าไปใหม่ว่าสามารถพยากรณ์ตัวแปรเกณฑ์ได้เพิ่มมากขึ้นหรือไม่ ไม่ได้ตรวจสอบผลกระทบที่เกิดเนื่องจากตัวแปรพยากรณ์ตัวใหม่ที่เข้าไปในสมการว่าส่งผลอย่างไรกับตัวแปร ที่อยู่ในรูปแบบก่อนหน้านี้แล้ว

- วิธี backward เป็นการนำตัวแปรเข้าวิเคราะห์พร้อมกันทั้งหมดทีเดียว ต่อจากนั้นจะเป็นการคัดเลือกตัวแปรออกทีละตัว แล้วทดสอบว่าค่า R2 ลดลงอย่างมีนัยสำคัญทางสถิติหรือไม่ ถ้าไม่มีนัยสำคัญ แสดงว่าตัวแปรดังกล่าวสามารถขจัดออกจากสมการได้ ซึ่งวิธีการนี้จะทำให้ได้โมเดลประหยัดก็จริง แต่ผู้วิเคราะห์จะไม่ทราบว่าตัวแปรพยากรณ์ที่เหลืออยู่ในสมการแต่ละตัวนั้นสามารถอธิบายความแปรปรวนของตัวแปรตามได้เท่าไร บอกได้เพียงแต่ว่าตัวแปรพยากรณ์ชุดนั้นร่วมกันอธิบายความแปรปรวนได้เท่าใด

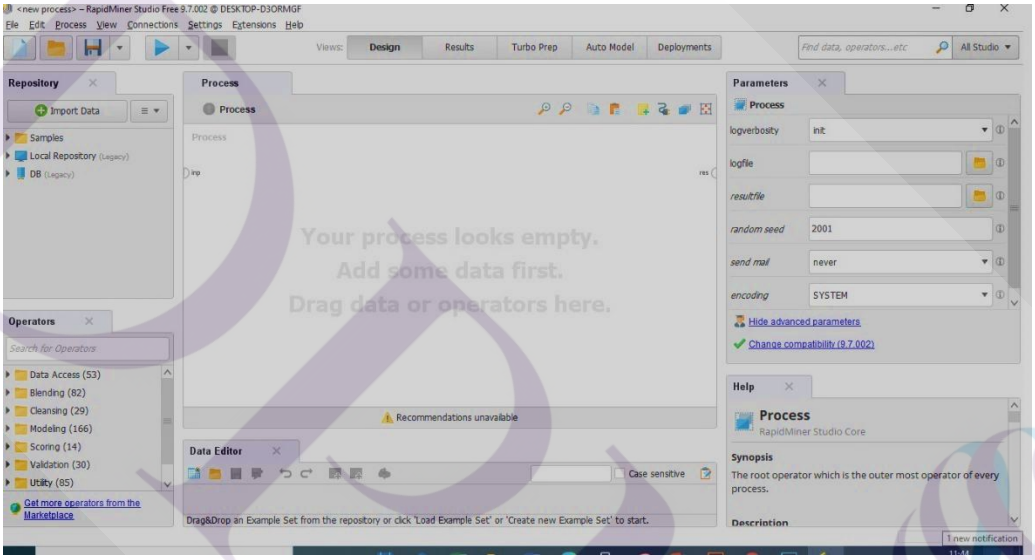
- วิธี Stepwise เป็นวิธีที่มีความเหมาะสมในการพิจารณาคัดเลือกตัวแปรพยากรณ์ที่ดีที่สุด ซึ่งลำดับขั้นตอนจะคล้ายกับวิธี Forward เพียงแต่ว่าการวิเคราะห์ด้วย Stepwise นั้นจะทำการทดสอบตัวแปรพยากรณ์ที่เข้าสมการไปแล้วทุกครั้งที่มีการนำตัวแปรใหม่เข้าในสมการว่าส่งผลกระทบต่อตัวแปรบางตัวที่อยู่ในสมการก่อนหน้าแล้ว โดยตัวแปรที่อยู่ก่อนสามารถถูกขจัดออกจากสมการได้ หากพบว่าไม่มีความสำคัญในการพยากรณ์ซึ่งวิธี Forward ไม่ได้ทดสอบในส่วนนี้

2.8 โปรแกรม Rapid Miner Studio [29]

Rapid Miner Studio แรกเริ่มพัฒนาขึ้นจากบริษัทที่ชื่อว่า Rapid-I ในประเทศเยอรมนี และ ปี พ.ศ. 2556 ได้รับทุนจากนักลงทุนในประเทศสหรัฐอเมริกาจึงเปลี่ยนชื่อบริษัทจาก Rapid-I เป็น Rapid Miner แทนและย้ายสำนักงานใหญ่มาอยู่ประเทศสหรัฐอเมริกา ผู้ต้องการใช้งานสามารถดาวน์โหลด Software Rapid Miner Studio version ปัจจุบัน 9.9 ได้จากเว็บไซต์ apidminer.com แสดงดังภาพที่ 2.12



ภาพที่ 2.12 เว็บไซต์ Rapid Miner Studio และหน้าเวปไซต์ ดาวโหลดโปรแกรม



ภาพที่ 2.13 โปรแกรม Rapid Miner Studio

ความสามารถและการทำงานของ Rapid Miner Studio version Education [29] เป็น Software ที่สามารถใช้งานฟรีทำให้สามารถประหยัดค่าใช้จ่าย กรณีต้องการใช้งานเพื่อทดสอบการวิเคราะห์ข้อมูล การทำความสะอาดข้อมูล (Data Cleansing) และการรายงานข้อมูลในรูปแบบต่าง ๆ เช่น ตาราง แผนภูมิเป็นต้น ความสามารถและการทำงานของ Software Rapid Miner Studio ดังนี้

1. สามารถนำเข้าข้อมูลได้หลายลักษณะ เช่น การเชื่อมโยงจากฐานข้อมูลโดยตรงไฟล์ Excel และไฟล์ CSV เป็นต้น

2. สามารถเขียนไฟล์ให้อยู่ในรูปแบบของ Excel และ CSV

3. สามารถแสดงข้อมูลในกราฟแบบต่าง ๆ เช่น Scatter Plot และ Time Series

4. สามารถแปลงข้อมูลจากฐานข้อมูล Relation Database ให้เป็นฐานข้อมูล Transaction Database

5. สามารถหากฎความสัมพันธ์ (Association Rules) การสร้างจาก Item ที่เกิดขึ้นบ่อย ๆ โดยเทคนิคการหากฎความสัมพันธ์ ได้แก่ Apriori และ FP Growth

6. สามารถแบ่งกลุ่มข้อมูล (Clustering) การแบ่งกลุ่มข้อมูล โดยข้อมูลที่มีลักษณะคล้าย ๆ กัน อยู่กลุ่มเดียวกัน และ ข้อมูลที่อยู่คนละกลุ่มจะมีลักษณะที่แตกต่างกันมาก ๆ เทคนิคการแบ่งกลุ่มข้อมูล ได้แก่ K-Means, Agglomerative Clustering และ DB Scan

7. สามารถจำแนกประเภทข้อมูล (Classification) การนำข้อมูลเดิมที่มีคำตอบที่สนใจหรือ คลาส (Class) มาสร้างเป็นโมเดล (Model) เพื่อหาคำตอบให้กับข้อมูลใหม่ (Unseen Data) โดยคลาสคำตอบเป็นประเภท (Nominal) เช่น ฝนตกหรือไม่ตก หรือ Spam Email หรือ Normal Email เป็นต้น เทคนิคการจำแนกประเภทข้อมูล ได้แก่ Linear Regression, Naïve Bayes, Decision Tree, K-Nearest Neighbors, Neural Networks และ Support Vector Machines

8. สามารถเปรียบเทียบประสิทธิภาพของโมเดลจากเทคนิค Classification ต่าง ๆ และ T-test

9. สามารถทำ Text Mining และ Image Mining

2.9 งานวิจัยที่เกี่ยวข้อง

[30] งานวิจัยเรื่อง การประยุกต์เทคนิคเหมืองข้อมูลค้นหาลักษณะนิสัย ของผู้ประกอบอาชีพด้านคอมพิวเตอร์ มีวัตถุประสงค์เพื่อสร้างและเปรียบเทียบประสิทธิภาพตัวแบบความสัมพันธ์ของข้อมูลจากเทคนิค Neural Networks, Naïve Bayes และ Decision Tree เพื่อค้นหาลักษณะนิสัยของผู้ประกอบอาชีพด้านคอมพิวเตอร์ตามทฤษฎีการเลือกอาชีพของฮอลแลนด์โดยพิจารณาจากตัวแบบจากค่าความถูกต้อง (Accuracy) และความแม่นยำ (Precision) ดำเนินการเก็บรวบรวมข้อมูลจากผู้ประกอบอาชีพด้านคอมพิวเตอร์ ใช้การสุ่มตัวอย่างจากจำนวนผู้ที่สำเร็จการศึกษาระดับปริญญาตรี หลักสูตรวิศวกรรมคอมพิวเตอร์ วิทยาการคอมพิวเตอร์และวิทยาศาสตร์คอมพิวเตอร์ กลุ่มตัวอย่าง 395 คน หลังจากรวบรวมข้อมูล แล้วจัดเตรียมข้อมูลให้ อยู่ในรูปแบบไฟล์ CSV จากนั้นวิเคราะห์ด้วยวิธี5 Fold-Cross-Validation แล้วสร้างตัวแบบทั้ง 3 เทคนิค จากนั้นวัดค่าประสิทธิภาพและค่าความแม่นยำของตัวแบบ ผลการศึกษา พบว่า ตัวแบบที่มีประสิทธิภาพมากที่สุด คือ ตัวแบบที่ได้จากเทคนิค Naïve Bayes ที่ค่าความถูกต้อง (Accuracy) เท่ากับ 79.64% (± 3.59 %) และ ค่าความแม่นยำ (Precision) เท่ากับ 74.04% (± 7.86%) รองลงมา คือ เทคนิค Neural Network มีค่าความถูกต้องเท่ากับ 77.82% (± 5.06 %) และค่าความแม่นยำ 73.60% (± 9.07%) และ

เทคนิค Decision Tree มีค่าความถูกต้องเท่ากับ 54.69% (±1.66%) และค่าความแม่นยำเท่ากับ 38.72% (± 4.62%) ตามลำดับ

[31] งานวิจัยเรื่อง การพัฒนาระบบสนับสนุนการตัดสินใจในการเลือกสาขาการเรียนต่อระดับประกาศนียบัตรวิชาชีพ สังกัดอาชีวศึกษาจังหวัดพระนครศรีอยุธยา ผลการศึกษาพบว่าการเปรียบเทียบประสิทธิภาพของโมเดลการจำแนกข้อมูลทั้ง 3 ได้แก่ K-nearest Neighbors, Decision Tree และ Rule based ซึ่งผลการเปรียบเทียบประสิทธิภาพโมเดลที่ได้ จากเทคนิค K-nearest Neighbors (K=6) มีความถูกต้องมากที่สุด 76.62% ซึ่งได้นำไปพัฒนาระบบสนับสนุนการตัดสินใจ การพัฒนาระบบตัดสินใจในการเลือกสาขาการเรียนนั้นได้ใช้ภาษา HTML, PHP โดยมีโครงสร้างจากตัวช่วยในการจัดการเว็บไซต์คือ Bootstrap (Bootstrap Front End Framework) แสดงผลในรูปแบบ Web Application จากนั้นได้ทำการออกแบบในส่วนของแบบทดสอบเลือกสาขาการเรียนเมื่อผู้ใช้ทำการทดสอบเสร็จแล้ว ระบบจะแสดงผลลัพธ์จากการทดสอบโดยใช้บริการสร้างแผนภูมิจาก Google (Google Chart API) และใช้ตัวแบบที่ได้จากการเก็บรวบรวมข้อมูลมาเป็นตัวตัดสินใจในการเลือกสาขาการเรียน โดยมีการแนะนำวิทยาลัยในจังหวัดพระนครศรีอยุธยาที่เหมาะสมกับนักศึกษา

[32] งานวิจัยเรื่อง การใช้เทคนิคเหมืองข้อมูลในการประเมินความรู้ และหาความถนัดเพื่อพัฒนาศักยภาพของนักศึกษา โดยงานวิจัย นี้นำผลการทำแบบทดสอบประเมินความรู้ก่อนเรียนและแบบทดสอบวัด ความถนัดของกลุ่มตัวอย่างมาวิเคราะห์โดยใช้เทคนิคการทำเหมืองข้อมูลผลการวิเคราะห์แบบประเมินความรู้ก่อนเรียน ผลลัพธ์ที่ได้สามารถจำแนกนักศึกษา โดยใช้เทคนิคการจัดกลุ่มด้วย K-mean Algorithm กำหนดการจัดกลุ่มนักศึกษาได้ เป็น 2 กลุ่ม กลุ่มนักศึกษาที่ด้อย (W) และกลุ่มนักศึกษาที่เด่น (S) จากการทำนายของชุดข้อมูลสอนให้ความถูกต้องเฉลี่ย 80.90 ผลการวิเคราะห์แบบทดสอบความถนัด ผลลัพธ์ที่ได้จำแนกผู้ตอบแบบสอบถามออกเป็นสาขาวิชา 4 สาขา ได้แก่ สาขาการเขียนโปรแกรม (P) สาขาเครือข่ายคอมพิวเตอร์ (N) สาขาฐานข้อมูล (D) และสาขาการผลิต Software (S) สามารถสร้างเป็น Decision Tree ทำให้ได้กฎความสัมพันธ์ของต้นแบบ 7 กฎ จากการทำนายของชุดข้อมูลสอนให้ร้อยละ ความถูกต้องเฉลี่ย 79.42 ผลจากงานวิจัยนี้สามารถนำแบบรูปที่ได้ไปใช้ในการทำนายคุณสมบัติของนักศึกษา เพื่อเป็นข้อมูลให้อาจารย์ที่ปรึกษาหรืออาจารย์ผู้สอนได้ใช้ประโยชน์ในการดูแลนักศึกษาได้ตรงกลุ่มเป้าหมายมากขึ้น ลดความเสี่ยงของนักศึกษาที่จะมีผลการเรียนที่ไม่ผ่านเกณฑ์ และได้แบบทดสอบวัดความถนัดเฉพาะสาขาทางวิทยาการ คอมพิวเตอร์ เพื่อพัฒนาศักยภาพของนักศึกษาตามความถนัดมีโอกาสได้งานทำหลังจบการศึกษามากขึ้น

[7] งานวิจัยเรื่อง ประสิทธิภาพการจำแนกข้อมูลการเลือกอาชีพโดยอัตโนมัติด้วยเทคนิคเหมืองข้อมูล งานนวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพในการจำแนกข้อมูลการเลือกอาชีพของนิสิตระดับปริญญาตรีหลังสำเร็จการศึกษา โดยในงานวิจัยนี้ได้ใช้ชุดข้อมูลภาวะการมีงานทำของบัณฑิต และข้อมูลระเบียนประวัติของนิสิตระดับปริญญาตรีหลังสำเร็จการศึกษา คณะวิทยาการสารสนเทศ มหาวิทยาลัยมหาสารคาม ระหว่างปีพ.ศ. 2550-2554 จำนวน 12 คุณลักษณะ และ 2,515 ระเบียน ซึ่งได้นำเทคนิค Decision Tree , เทคนิค Neural Network และเทคนิค Naïve Bayesian Learning มาทำการเปรียบเทียบประสิทธิภาพ ผลจากการศึกษาพบว่าประสิทธิภาพในการจำแนกข้อมูลแบบ Decision Tree มีประสิทธิภาพในการจำแนกสูงสุดด้วยค่าเฉลี่ย 80.62% และปัจจัยสำคัญที่ทำให้การเลือกอาชีพ ตรงหรือไม่ตรงกับสาขา มี 4 ปัจจัย คือ สาขาวิชาที่เรียน เกรดเฉลี่ยเฉพาะวิชา สาขา เพศ และเกรดเฉลี่ยรวม ซึ่งผลการทดลองนี้สามารถนำไปประยุกต์ใช้กับคณะหรือหน่วยงานที่เกี่ยวข้อง เพื่อวางแผนพัฒนาโครงสร้างหลักสูตรหรือวางแผนการศึกษาให้กับนิสิตได้

[33] งานวิจัยเรื่อง การค้นหาความรู้คุณลักษณะสำคัญของนักศึกษาที่มีต่อผลสัมฤทธิ์ทางการเรียนในกลุ่มโปรแกรมคอมพิวเตอร์ มหาวิทยาลัยราชภัฏนครปฐม ด้วยเทคนิคเหมืองข้อมูล การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อค้นหาความรู้คุณลักษณะสำคัญของนักศึกษาที่มีต่อผลสัมฤทธิ์ทางการเรียนในกลุ่มโปรแกรมคอมพิวเตอร์ มหาวิทยาลัยราชภัฏนครปฐม ด้วยเทคนิคเหมืองข้อมูลเพื่อสร้างโมเดลหรือต้นแบบ เพื่อการพัฒนาระบบสนับสนุนการตัดสินใจการเข้าศึกษาสำหรับผู้สนใจเรียนในสาขาวิชาของกลุ่มโปรแกรมคอมพิวเตอร์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏนครปฐม โดยได้ทำการศึกษาข้อมูลจากนักศึกษาที่ศึกษาอยู่ในสาขาคอมพิวเตอร์ชั้นปีที่ 3 และ 4 ในปัจจุบัน โดยใช้ระดับผลการเรียนเฉลี่ย ณ ปัจจุบันเป็นเกณฑ์ในการจำแนกกลุ่มนักศึกษา ในการศึกษาผู้วิจัยได้ทำการทดลองกับเทคนิคทางเหมืองข้อมูล 2 เทคนิค ได้แก่ เทคนิค Decision Tree J48 (C4.5) และเทคนิค Bayesian Network โดยแบ่งการทดลองออกเป็น 2 ลักษณะ คือ การทดลองกับชุดข้อมูลทั้งหมดทุก Attribute และการทดลองด้วยการเลือกคุณลักษณะด้วยเทคนิค SNR (Signal to Noise Ratio) เพื่อเพิ่มประสิทธิภาพในการทำงานของอัลกอริทึมซึ่งจากผลการทดลองทั้งสองแบบยืนยันว่า เทคนิค Decision Tree J48 (C4.5) ให้ประสิทธิภาพในการทำงานที่ดีกว่าเมื่อเทียบกับเทคนิค Bayesian Network

[34] งานวิจัยเรื่อง การจำแนกประเภทผู้ใช้ไฟฟ้าด้วยนาอีฟเบย์และนิวรัลเน็ทเวิร์ก โดยใช้วิธี Naïve Bayes และ Neural Networks สำหรับจำแนกผู้ใช้ไฟฟ้าออกตามประเภทกิจการที่ได้กำหนดไว้ จากผลการทดสอบพบว่าวิธีการ Naïve Bayes ให้ความถูกต้อง 81.25 % และ Neural Networks ให้ความถูกต้อง 87.5 % ผลการวิจัยส่งผลให้หน่วยงานการไฟฟ้าส่วนภูมิภาค (กฟภ.)

และการไฟฟ้านครหลวง (กฟน.) สามารถจัดสรรรูปแบบการให้บริการไฟฟ้าที่เหมาะสมกบกับกิจการแต่ละประเภท ทั้งยังสามารถวิเคราะห์พฤติกรรมการบริโภคพลังงานของผู้ใช้ไฟฟ้าที่ไม่สอดคล้องกับประเภทกิจการที่ระบุไว้ตอนเริ่มต้น

[8] งานวิจัยเรื่อง การวัดประสิทธิภาพความแม่นยำของวิธี Bayesian Network ในการตรวจสอบรูปแบบการเรียนของนักศึกษาสาขาวิชาวิศวกรรมคอมพิวเตอร์ในรายวิชาปัญญาประดิษฐ์ โดยทำการเปรียบเทียบรูปแบบการเรียนที่ได้จากวิธี Bayesian Network กับแบบสอบถาม โดยที่มีการแบ่งรูปแบบการเรียนออกเป็น 3 ประเภท คือการรับรู้ (Perception) การประมวลผล (Processing) และความเข้าใจ (Understanding) ผลการทดลองพบว่าประสิทธิภาพของ วิธี Bayesian Network ให้ความถูกต้องของผลการทดลองอยู่ในระดับที่สูง

[35] งานวิจัยเรื่อง การสกัดข้อมูลจากรายงานทางพยาธิวิทยาภายในโรงพยาบาล (Information extraction from pathology reports in a hospital setting) นำการทำเหมืองข้อความมาประยุกต์ใช้ในกระบวนการทำงานของโรงพยาบาลรอยัลเมลเบิร์น เพื่อสกัดหาข้อมูลที่เป็นประโยชน์จากรายงานทางพยาธิวิทยาและลดการใช้ความรู้จากผู้เชี่ยวชาญลง โดยใช้ตัวจำแนกประเภทชนิดต่าง ๆ ที่ได้รับความนิยมในการทำเหมืองข้อมูลมาเปรียบเทียบประสิทธิภาพที่ได้จากการจำแนกประเภทข้อมูลชุดเดียวกัน ซึ่งผลลัพธ์ที่ได้คือตัวจำแนกประเภทวิธี Naïve Bayesian Learning จะเหมาะสำหรับข้อมูลที่เป็นค่าไม่ต่อเนื่อง ส่วนตัวจำแนกประเภทวิธี Support Vector Machine จะเหมาะสำหรับข้อมูลตัวเลข

บทที่ 3
วิธีการดำเนินการวิจัย

การวิจัยเรื่อง การประยุกต์ใช้ Machine Learning ทำนายผลการเรียนวิชา Web Database ของนิสิตสาขาเทคโนโลยีการศึกษา คณะศึกษาศาตร์ มหาวิทยาลัยบูรพา มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของโมเดลการทำนายผลการเรียนวิชา Web Database ระหว่างวิธี Decision Tree, Naïve Bayes, Neural Network และ Support Vector Machine การดำเนินการวิจัยมี 5 ขั้นตอน ดังตารางที่ 3.1

ตารางที่ 3.1 แสดงขั้นตอนการดำเนินการวิจัย

ขั้นตอน	การดำเนินการ
ขั้นตอนที่ 1 การเลือกข้อมูล (Selection)	- การกำหนด Attribute - การเก็บข้อมูลด้วยแบบสอบถาม - การเก็บข้อมูลจากฐานข้อมูลนิสิต
ขั้นตอนที่ 2 การเตรียมข้อมูล (Preprocessing)	- แยกข้อมูลที่ไม่มีค่าข้อมูล ข้อมูลซ้ำซ้อน หรือทำการบันทึกผิดออก - รวบรวมข้อมูลเข้าไว้ด้วยกัน
ขั้นตอนที่ 3 การเปลี่ยนรูปข้อมูล (Transformation)	- เปลี่ยนรูปข้อมูลให้อยู่ในรูปแบบที่สอดคล้องกับ Machine Learning - วิเคราะห์ความถดถอยพหุคูณ (Multiple Regression) แบบการคัดเลือกแบบลำดับขั้น (Stepwise Selection) ด้วยโปรแกรม SPSS for Window
ขั้นตอนที่ 4 การทำเหมืองข้อมูล (Data Mining)	- สร้างโมเดลทำนายผลการเรียนด้วยวิธี Decision Tree, Naïve Bayes, Neural Network และ Support Vector Machine ด้วยโปรแกรม RapidMiner Studio
ขั้นตอนที่ 5 การแปรผลและการประเมินผล (Interpretation Evaluation)	- เปรียบเทียบประสิทธิภาพการทำนายผลการเรียน - แปลผลอยู่ในรูปของความรู้ (Knowledge) - การนำโมเดลที่มีค่าความถูกต้องมากที่สุดไปใช้

รายละเอียดการดำเนินการแต่ละขั้นตอน ดังนี้

3.1 ขั้นตอนที่ 1 การเลือกข้อมูล (Selection)

จากการศึกษาเอกสารงานวิจัยที่เกี่ยวผู้วิจัยดำเนินสังเคราะห์เลือก Attribute เบื้องต้นที่เกี่ยวข้องต่อผลการเรียนวิชา Web Database จำนวน 18 Attribute ได้แก่ 1) เพศ 2) ประเภทการรับเข้ามหาวิทยาลัย 3) แผนการเรียนมัธยมปลาย 4) เกรดเฉลี่ยจบการศึกษาระดับชั้น ม. 6 5) เกรดเฉลี่ยวิชาคณิตศาสตร์ ม.6 6) เคยเรียนภาษา HTML 7) เคยเรียนภาษา C หรือ C++ 8) เคยเรียนภาษา PHP 9) คอมพิวเตอร์ส่วนตัว 10) ความรู้สึกต่อการเรียนวิชาด้านคอมพิวเตอร์ 11) การทบทวนบทเรียน 12) เกรดเฉลี่ยกลุ่มวิชาภาษาอังกฤษ 13) เกรดเฉลี่ยกลุ่มวิชาภาษาอื่น ๆ 14) เกรดเฉลี่ยกลุ่มวิชาอัตลักษณ์มหาวิทยาลัย 15) เกรดเฉลี่ยกลุ่มวิชาสังคม 16) เกรดเฉลี่ยกลุ่มวิชาความสร้างสรรค์ 17) เกรดเฉลี่ยกลุ่มวิชาเทคโนโลยีสารสนเทศ 18) ผลการเรียนวิชา Web database

จากการกำหนด Attribute ที่เกี่ยวข้อง จำนวน 18 Attribute ผู้วิจัยดำเนินเก็บข้อมูล 2 วิธีการตามลักษณะและแหล่งของข้อมูล ได้แก่ การเก็บข้อมูลด้วยแบบสอบถาม จำนวน 7 Attribute และการเก็บข้อมูลจากฐานข้อมูลทะเบียนและวัดผล จำนวน 11 Attribute รายละเอียดดังนี้

1. การเก็บข้อมูลด้วยแบบสอบถาม ผู้วิจัยสร้างแบบสอบถามด้วย Google Forms ประเด็นคำถามจาก Attribute จำนวน 7 Attribute ได้แก่

1) ก่อนเรียนวิชา Web Database เคยเรียนภาษา HTML หรือไม่

2) ก่อนเรียนวิชา Web Database เคยเรียนภาษา C หรือ C++ หรือไม่

3) ก่อนเรียนวิชา Web Database เคยเรียนภาษา PHP หรือไม่

4) เกรดเฉลี่ยวิชาคณิตศาสตร์ ม.6

5) ความรู้สึกชอบหรือไม่ชอบต่อการเรียนวิชาด้านคอมพิวเตอร์

6) มีคอมพิวเตอร์ส่วนตัวหรือไม่

7) หลังจากเรียนวิชา Web Database ในแต่ละครั้งมีการทบทวนความรู้หรือไม่

ผู้วิจัยส่งแบบสอบถามออนไลน์ Google Forms ให้กับนิสิตสาขาวิชาเทคโนโลยีการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยบูรพา ที่ผ่านการเรียนวิชา Web Database ปีการศึกษา 2561 -2563 จำนวน 271 ระเบียน และจัดเก็บผลการตอบแบบสอบถามในรูปแบบเอกสารอิเล็กทรอนิกส์ Excel เพื่อนำไปสู่การดำเนินการวิจัยในขั้นต่อไป

2. การเก็บข้อมูลจากฐานข้อมูลทะเบียนและวัดผลจำนวน 11 Attribute ได้แก่

1) เพศ

2) เกรดเฉลี่ยจบการศึกษาระดับชั้น ม. 6

3) แผนการเรียนมัธยมปลาย

4) ประเภทการรับเข้ามหาวิทยาลัย

5) เกรดเฉลี่ยกลุ่มวิชาภาษาอังกฤษ

6) เกรดเฉลี่ยกลุ่มวิชาภาษาอื่น ๆ

7) เกรดเฉลี่ยกลุ่มวิชาอัตลักษณ์มหาวิทยาลัย

8) เกรดเฉลี่ยกลุ่มวิชาสังคม

9) เกรดเฉลี่ยกลุ่มวิชาความสร้างสรรค์

10) เกรดเฉลี่ยกลุ่มวิชาเทคโนโลยีสารสนเทศ

11) เกรดวิชา Web Database

ผู้วิจัยใช้ข้อมูลนิสิตสาขาวิชาเทคโนโลยีการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยบูรพา ที่ผ่านการเรียนวิชา Web Database ปีการศึกษา 2561 -2563 จำนวน 271 ระเบียบ โดยรวบรวมข้อมูลเป็นรายบุคคลและจัดเก็บรูปแบบเอกสารอิเล็กทรอนิกส์ Excel เพื่อนำไปสู่การดำเนินการวิจัยในขั้นต่อไป

3.2 ขั้นตอนที่ 2 การเตรียมข้อมูล (Preprocessing)

ผู้วิจัยรวบรวมข้อมูลเอกสารอิเล็กทรอนิกส์ Excel จากการเก็บข้อมูลด้วยแบบสอบถามจำนวน 7 Attribute และการเก็บข้อมูลจากฐานข้อมูลทะเบียนและวัดผล จำนวน 11 Attribute รวมเข้าเป็นเอกสารอิเล็กทรอนิกส์ Excel ไฟล์เดียวกันเพื่อตรวจสอบความสมบูรณ์ของข้อมูลโดยการแยกข้อมูลที่มีการบันทึกผิดหรือการบันทึกซ้ำซ้อนออกไป

ตารางที่ 3.2 ตัวอย่างข้อมูลจากแบบสอบถามและจากฐานข้อมูล

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	ID	Sex	admission	MajorHS	GPAHS	GPAMath	HTML	CCPP	PHP	Computer	attitude	Review	sub_eng	sub_lang	identity	sub_soc	sub_create	sub_ict	GWBD
2	xxxxxxx	female	direct	art_lang	3.4	3	yes	yes	no	yes	yes	sometim	2.5	3.5	4	4	2.75	2.5	3.5
3	xxxxxxx	female	direct	art_lang	3.26	3.5	yes	no	yes	yes	yes	sometim	2	2.5	3	4	2	2.5	4
4	xxxxxxx	female	direct	art_cal	2.67	2	yes	yes	no	yes	yes	sometim	1.25	1	4	4	2	2	3.5
5	.....	.....	....	.....	.....	.....	.....	....	....	.....	...	....	...	...	...	...	...	.....	....
6	xxxxxxx	female	direct	art_lang	3.28	2.5	yes	yes	yes	yes	yes	sometim	2	3	3.75	4	3.75	2.5	3
7	xxxxxxx	female	direct	art_cal	2.93	2.5	yes	no	yes	yes	yes	sometim	1.17	3	3.5	3.86	3.5	1	4
8																			

3.3 ขั้นตอนที่ 3 การเปลี่ยนรูปข้อมูล (Transformation)

3.3.1 การกำหนด Key code ของข้อมูล

ผู้วิจัยกำหนด Key code ข้อมูล และเปลี่ยนรูปข้อมูลให้อยู่ในรูปแบบที่เหมาะสมสำหรับ Machin Learning วิธี Decision Tree, Naïve Bayes, Neural Network และ Support Vector Machine โดยมี Attribute ที่ต้องเปลี่ยนรูปข้อมูล จำนวน 3 Attribute ได้แก่ 1) Admission (ประเภทการรับเข้ามหาวิทยาลัย) แยกเป็น 4 Attribute 2) Major HS (แผนการเรียนมัธยมปลาย) แยกเป็น 3 Attribute 3) GWBD (เกรดวิชา Web Database) ซึ่งใช้เป็น Label กำหนดให้มี 2 ค่า คือ กลุ่มเกรดพอใช้ (D, D+, C, C+) และ กลุ่มเกรดดี (B, B+, A)

เหตุผลในการแบ่งเกรดออกเป็น 2 กลุ่ม เนื่องจากผู้วิจัยต้องการนำโมเดลจากงานวิจัยนี้ไปใช้ในการทำนายผลการเรียนของนิสิตในรุ่นต่อไป ซึ่งผลจากการทำนายแบ่งเป็น 2 กลุ่มจะทำให้อาจารย์ผู้สอนวางแผนการสำหรับสอนกลุ่มที่คาดว่าจะมีผลการเรียนพอใช้ให้มีผลการเรียนอยู่ในระดับดีขึ้นได้ ในส่วนกลุ่มนิสิตที่คาดว่าจะมีผลเรียนดีจะทำให้อาจารย์สามารถวางแผนการสอนที่เหมาะสมสำหรับรองรับความแตกต่างของนิสิตทั้งสองกลุ่มได้ และนอกจากนั้นอาจารย์ผู้สอนสามารถสร้างการทำงานเป็นทีมของนิสิตโดยใช้กลยุทธ์การสอนเพื่อนช่วยเพื่อน ในแต่ละทีมมีการคละกันของสมาชิกที่ผลการเรียนดีและพอใช้ จะสามารถยกระดับผลการเรียนนิสิตที่คาดว่าจะอยู่ในเกรดพอใช้ให้เปลี่ยนอยู่ในกลุ่มเกรดดีได้

ตารางที่ 3.3 รูปแบบข้อมูล และการกำหนด Key code ข้อมูล

No.	Attribute	Detail	Key code
1	Sex	เพศ	0 = ชาย 1 = หญิง
2	Admission1	โครงการเพชรตะวันออก	0 = ไม่ใช่ 1 = ใช่
3	Admission2	Admission ส่วนกลาง	0 = ไม่ใช่ 1 = ใช่
4	Admission3	สอบเข้า	0 = ไม่ใช่ 1 = ใช่
5	Admission4	รับตรง	0 = ไม่ใช่ 1 = ใช่

No.	Attribute	Detail	Key code
6	MajorHS1	แผนการเรียนมัธยมปลาย กลุ่ม ศิลป์/คำนวณ	0 = ไม่ใช่ 1 = ใช่
7	MajorHS2	แผนการเรียนมัธยมปลาย กลุ่ม ศิลป์/ภาษา	0 = ไม่ใช่ 1 = ใช่
8	MajorHS3	แผนการเรียนมัธยมปลาย กลุ่ม คณิต/วิทย์	0 = ไม่ใช่ 1 = ใช่
9	GPAHS	เกรดเฉลี่ยจบการศึกษาระดับชั้น ม. 6	ใช้ค่าเกรดตามจริง
10	GPA Math	เกรดเฉลี่ยวิชาคณิตศาสตร์ ม.6	ใช้ค่าเกรดตามจริง
11	HTML	เคยเรียนภาษา HTML	0 = ไม่เคย 1 = เคย
12	CCPP	เคยเรียนภาษา C หรือ C++	0 = ไม่เคย 1 = เคย
13	PHP	เคยเรียนภาษา PHP	0 = ไม่เคย 1 = เคย
14	Computer	คอมพิวเตอร์ส่วนตัว	0 = ไม่มี 1 = มี
15	Attitude	ความรู้สึกต่อการเรียนวิชาด้านคอมพิวเตอร์	0 = ไม่ชอบ 1 = ชอบ
16	Review	การทบทวนบทเรียน	0 = ไม่เคย 1 = เคย
17	Sub_eng	เกรดเฉลี่ยกลุ่มวิชาภาษาอังกฤษ	ใช้ค่าเกรดตามจริง
18	Sub_lang	เกรดเฉลี่ยกลุ่มวิชาภาษาอื่น ๆ	ใช้ค่าเกรดตามจริง
19	Identity	เกรดเฉลี่ยกลุ่มวิชาอัตลักษณ์มหาวิทยาลัย	ใช้ค่าเกรดตามจริง
20	Sub_soc	เกรดเฉลี่ยกลุ่มวิชาสังคม	ใช้ค่าเกรดตามจริง
21	Sub_create	เกรดเฉลี่ยกลุ่มวิชาความสร้างสรรค์	ใช้ค่าเกรดตามจริง
22	Sub_ict	เกรดเฉลี่ยกลุ่มวิชาเทคโนโลยีสารสนเทศ	ใช้ค่าเกรดตามจริง
23	GWBD	เกรดวิชา Web Database	0 = พอใช้ (D, D+, C, C+) 1 = ดี (B, B+, A)

จากตารางที่ 3.3 แสดงรูปแบบของข้อมูลในแต่ละ Attribute พบว่า ลักษณะข้อมูลมี 2 ประเภทคือข้อมูที่มี 2 ค่า (Binary number) คือ 0 กับ 1 และข้อมูลที่เป็นจำนวนจริง จากการกำหนดรูปแบบข้อมูลดังกล่าวจึงมีความเหมาะสม ต่อการนำเข้าโมเดลเพื่อทำนายผลการเรียนด้วยวิธี Decision Tree, Naïve Bayes, Neural Network และ Support Vector Machine โดยทุกวิธีสามารถใช้ชุดข้อมูลเดียวกัน ตัวอย่างรูปแบบข้อมูลดังตารางที่ 3.4

ตารางที่ 3.4 แสดงตัวอย่างรูปแบบข้อมูล และ Code ข้อมูล

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
1	ID	Sex	Admission1	Admission2	Admission3	Admission4	MajorHS1	MajorHS2	MajorHS3	GPAHS	GPAMath	HTML	C C++	PHP	Computer	Attitude	Review	Sub_eng	Sub_lang	Identity	Sub_soc	Sub_create	Sub_ict	GWDB
2	####	1	0	0	0	1	0	0	1	3.26	2.5	1	0	0	1	1	0	1.63	2.5	4	3.86	4	2	1
3	####	0	0	1	0	0	0	0	1	3.02	2.5	1	0	0	1	1	1	1.25	2.5	3.75	3.86	3.5	2	1
4	####	0	0	1	0	0	0	1	0	3.15	2.5	0	0	1	1	1	1	1.25	2.5	4	3.86	3.75	2	1
5	####	0	0	0	0	1	0	0	1	2.75	3	1	1	1	1	1	1	2.33	3	4	3.86	3.25	3	1
6	####	1	0	0	0	1	0	0	1	3.03	2.5	0	1	0	1	1	1	2.83	3	4	3.71	3.75	2.5	1
7	####	1	1	0	0	0	0	0	1	3.64	3.5	1	0	0	1	1	0	1.88	3	4	4	4	3	1
8	####	1	1	0	0	0	0	0	1	3.84	3.5	1	0	0	1	1	0	3	4	4	4	3.5	3.5	1
9	####	0	1	0	0	0	0	0	1	3.61	3.5	1	1	0	1	1	1	3.5	4	4	4	3.5	4	1
10	####	1	1	0	0	0	0	1	0	3.32	2.5	1	0	0	1	0	1	1.88	2.5	4	3.86	3.75	2.5	1
11	####	0	0	0	0	1	0	0	1	3.43	3.5	0	1	0	1	1	1	2.5	3.5	4	3.57	3.75	3.5	1
12	####	1	0	0	0	1	0	1	0	2.56	2.5	0	1	0	1	1	0	2.33	2	3.75	3.14	2	2.5	0
13	####	0	0	0	0	1	1	0	0	2.86	2.5	0	1	0	1	1	1	2.5	3	3.75	3.86	3.5	2.5	1
14	####	1	0	0	0	1	1	0	0	3.07	2.5	1	0	0	1	1	1	2.67	3.5	3.5	3.29	3.25	3	1
15	####	0	0	0	0	1	1	0	0	2.95	3	0	1	0	1	1	1	2.38	3	4	3.71	3.25	3	1
16	####	1	0	0	0	1	0	0	1	2.83	2.5	0	1	0	1	0	1	2.33	3	3.75	3.43	3.5	2.5	0
17	####	1	1	0	0	0	1	0	0	3.65	3.5	1	0	0	1	1	0	3.33	3.5	3.75	4	3.75	4	1
18	####	0	0	0	0	1	1	0	0	2.82	3	1	1	1	1	1	1	2.67	3	4	3.71	3.75	2.5	1
19	####	1	0	0	0	1	0	1	0	3.03	2.5	0	1	0	1	1	1	1.13	2.5	3.5	3.43	3.5	1.5	1
20	####	1	0	0	0	1	0	1	0	3.1	2.5	1	0	0	1	1	1	1.63	2.5	3.5	3.57	3.25	1.5	1
21	####	0	0	0	0	1	0	1	0	3.02	2.5	1	0	0	0	0	1	2.17	3	4	4	3.5	3	0
22	####	0	0	0	0	1	0	0	1	2.97	2.5	0	1	0	1	1	1	1.75	3	4	4	3.5	3.5	1
23	####	1	1	0	0	0	0	0	1	3.41	3.5	1	0	0	1	1	1	2.67	4	3.75	3.71	3.5	3	1
24	####	1	0	0	0	1	0	1	0	3.42	3	1	0	0	0	1	1	2	3	4	3.71	3.75	3	1

datafinal Sheet1

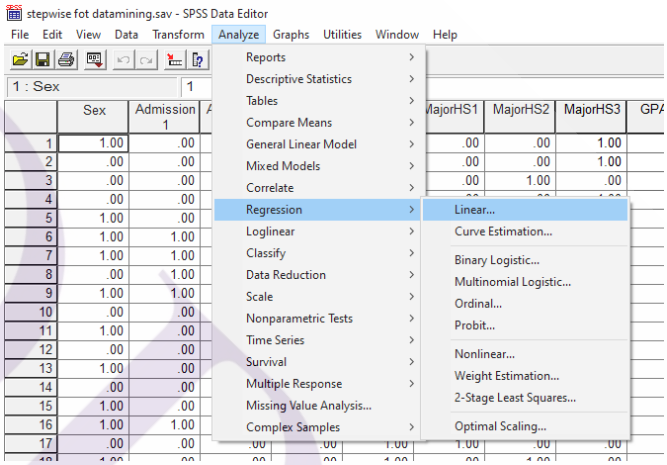
จากตารางที่ 3.4 จะพบว่าหลังการดำเนินการรวบรวมข้อมูลและแปรรูปข้อมูลเรียบร้อยได้ข้อมูลทั้งหมด 22 Attribute สำหรับนำเข้าโมเดลทำนายผลการเรียน โดยมี Label คือ ผลการเรียนวิชา Web Database แบ่งเป็น 2 ค่า คือ กลุ่มเกรดพอใช้ Code = 0 และ กลุ่มเกรดดี Code = 1 ผู้วิจัยกำหนดให้ข้อมูลชุดนี้ คือ ข้อมูลชุดที่ 1

3.3.2 การวิเคราะห์ความพันธ์ระหว่าง Attribute และ Label

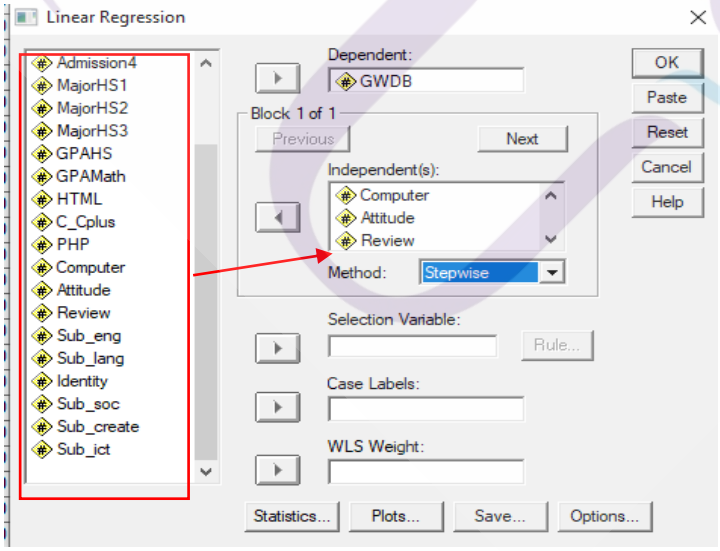
หลังจากการดำเนินการรวบรวมข้อมูลและแปรรูปข้อมูลเรียบร้อยแล้ว ผู้วิจัยดำเนินการวิเคราะห์ความพันธ์ระหว่าง Attribute และ Label วัตถุประสงค์การวิเคราะห์เพื่อคัดกรอง Attribute ที่ส่งผลหรือมีอิทธิพลต่อผลการเรียนวิชา Web Database อย่างมีนัยสำคัญทางสถิติ (F-test) ที่ .05

เมื่อได้ Attribute ที่คัดกรองแล้วผู้วิจัยเข้าโมเดลทำนายผลการเรียน เพื่อเปรียบเทียบประสิทธิภาพการทำนายกับโมเดลที่ใช้ข้อมูลโดยไม่ได้คัดกรอง Attribute

ในการคัดกรอง Attribute ผู้วิจัยใช้วิธีการวิเคราะห์ความถดถอยพหุคูณ (Multiple Regression) แบบการคัดเลือกแบบลำดับขั้น (Stepwise Selection) ด้วยโปรแกรม SPSS for Window เป็นการวิเคราะห์ว่าตัวแปรอิสระ (Attribute) ใดที่ส่งผลหรือมีอิทธิพลต่อตัวแปรตาม (Label) ในการวิเคราะห์ด้วยโปรแกรม SPSS for Window มีขั้นตอนดังภาพที่ 3.1 – 3.4

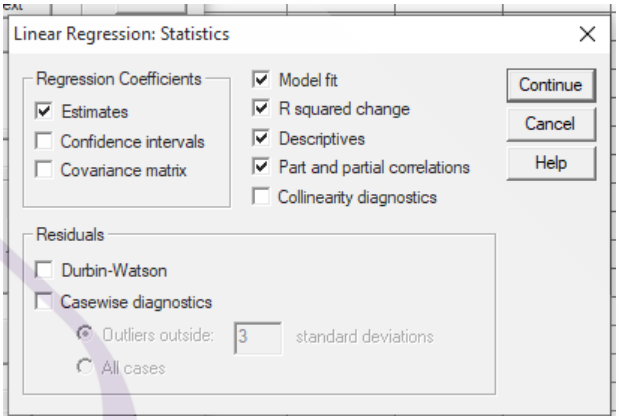


ภาพที่ 3.1 แสดงการใช้เมนู “Analyze” เมนูรอง “Regression” และเมนูย่อย “Linear”



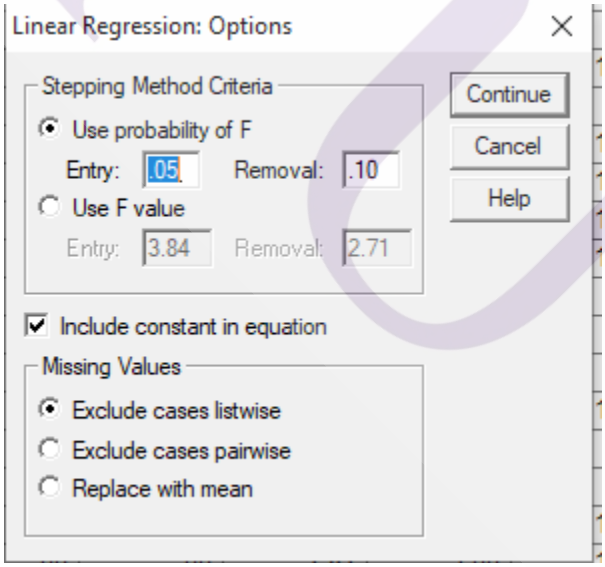
ภาพที่ 3.2 การตั้งค่าหน้า เมนู Linear Regression

จากภาพที่ 3.2 กำหนดตัวแปรตาม (Dependent) เป็น GWDB และนำ Attribute ที่เหลือทั้งหมดเข้าช่องตัวแปรต้น (Independent) การคัดเลือกแบบวิเคราะห์ความถดถอยพหุคูณ Stepwise Selection



ภาพที่ 3.3 การเลือกสถิติในการวิเคราะห์ข้อมูล

จากภาพที่ 3.3 การเลือกสถิติในการวิเคราะห์ข้อมูล โดยตั้งค่า Regression Coefficient เป็น Estimate และเลือกค่า Model fit, R squared chang, Descriptive, Part and partial correlation



ภาพที่ 3.4 การตั้งค่าเกณฑ์ในการนำเข้าตัวแปรทำนายหรือขจัดตัวแปรทำนาย

จากภาพที่ 3.4 กำหนดค่า Use probability of F , Entry =.05 และ Remove = .10 เพื่อให้คัดกรอง Attibute วิธีการคัดเลือกตัวแปรอิสระเข้าสมการ ใช้วิธี Stepwise เกณฑ์ในการคัดเลือกตัวแปรอิสระเข้าสมการโดยใช้การทดสอบการมีนัยสำคัญทางสถิติ (F-test) ที่ระดับน้อยกว่าหรือเท่ากับ 0.05 เกณฑ์ในการคัดเลือกตัวแปรอิสระออกจากสมการโดยใช้การทดสอบการมีนัยสำคัญทางสถิติ (F-test) ที่ระดับมากกว่าหรือเท่ากับ 0.100

ผลการวิเคราะห์ความสัมพันธ์ของ Attribute และ Label ดังตารางที่ 3.5

ตารางที่ 3.5 ผลการวิเคราะห์ความสัมพันธ์ของ Attribute และ Label

Model Summary

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	.449(a)	.201	.198	.41211	.201	67.736	1	269	.000
2	.539(b)	.291	.285	.38907	.089	33.793	1	268	.000
3	.564(c)	.318	.311	.38214	.028	10.804	1	267	.001
4	.584(d)	.341	.332	.37626	.023	9.413	1	266	.002
5	.596(e)	.355	.343	.37304	.014	5.623	1	265	.018
6	.605(f)	.366	.351	.37062	.011	4.460	1	264	.036

a Predictors: (Constant), Sub_soc

b Predictors: (Constant), Sub_soc, PHP

c Predictors: (Constant), Sub_soc, PHP, Computer

d Predictors: (Constant), Sub_soc, PHP, Computer, Sub_create

e Predictors: (Constant), Sub_soc, PHP, Computer, Sub_create, Identity

f Predictors: (Constant), Sub_soc, PHP, Computer, Sub_create, Identity, Attitude

จากตารางที่ 3.5 พิจารณาจากค่า R [36] ค่าสัมประสิทธิ์สหสัมพันธ์พหุคูณ (Multiple Correlation) เป็นค่าที่แสดงถึงระดับความสัมพันธ์ระหว่างกลุ่มของตัวแปรอิสระทั้งหมดในสมการกับตัวแปรตาม ค่าเข้าใกล้ 1 แสดงว่ากลุ่มของตัวแปรอิสระมีความสัมพันธ์กับตัวแปรตามมาก

จากจำนวน Attribute ที่นำเข้าโปรแกรม SPSS เพื่อวิเคราะห์ความสัมพันธ์ จำนวน 22 Attribute โปรแกรมได้คัดกรอง Attribute ที่มีอิทธิพลต่อผลการเรียนวิชา Web Database เหลือ 6 Attribute และโปรแกรมได้สร้างโมเดลความสัมพันธ์ 6 โมเดล พบว่า โมเดลที่ 6 มีค่า R (ค่าสัมประสิทธิ์สหสัมพันธ์พหุคูณ) มากที่สุด 0.605 (60.50 %) โดยโมเดลที่ 6 มี Attribute ประกอบด้วย 1) Sub_soc (เกรดเฉลี่ยกลุ่มวิชาสังคม), 2) PHP (เคยเรียนภาษา PHP), 3) Computer (คอมพิวเตอร์ส่วนตัว), 4) Sub_create (เกรดเฉลี่ยกลุ่มวิชาความสร้างสรรค์), 5) Identity (เกรดเฉลี่ยกลุ่มวิชาอัตลักษณ์มหาวิทยาลัย), และ 6) Attitude (ความรู้สึกต่อการเรียนวิชาด้านคอมพิวเตอร์)

ตารางที่ 3.6 แสดงค่าสัมประสิทธิ์การถดถอยในแบบคะแนนมาตรฐาน

Coefficients(a)									
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations		
		B	Std. Error	Beta			Zero-order	Partial	Part
6	(Constant)	-1.809	.263		-6.873	.000			
	Sub_soc	.312	.052	.323	6.050	.000	.449	.349	.297
	PHP	.251	.048	.263	5.233	.000	.354	.307	.256
	Computer	.301	.093	.163	3.258	.001	.211	.197	.160
	Sub_create	.132	.046	.150	2.881	.004	.310	.175	.141
	Identity	.123	.053	.120	2.340	.020	.224	.143	.115
	Attitude	.178	.084	.106	2.112	.036	.187	.129	.104
a Dependent Variable: GWDB									

จากตารางที่ 3.6 จากค่า Beta [37] ที่สามารถบอกได้ว่าตัวแปรอิสระได้มีผลหรืออิทธิพลต่อตัวแปรตามมากหรือน้อยกว่ากัน ถ้า Beta ของตัวแปรอิสระใดมีค่ามาก แสดงว่าตัวแปรอิสระนั้นจะมีอิทธิพลต่อตัวแปรตามมาก ดังนั้น Attribute ที่มีอิทธิพลต่อผลการเรียนวิชา Web Database จาก Beta มากไปน้อย ได้แก่ Sub_soc (เกรดเฉลี่ยกลุ่มวิชาสังคม) Beta = 0.323 , PHP (เคยเรียนภาษา PHP) ค่า Beta =0.263, Computer (คอมพิวเตอร์ส่วนตัว) Beta = .163, Sub_create (เกรดเฉลี่ยกลุ่มวิชาความสร้างสรรค์) Beta =0.150, Identity (เกรดเฉลี่ยกลุ่มวิชาอัตลักษณ์มหาวิทยาลัย) Beta = Beta =0.120, และ Attitude (ความรู้สึกต่อการเรียนวิชาด้านคอมพิวเตอร์) Beta =0.106

สรุปขั้นตอนการเปลี่ยนรูปข้อมูล (Transformation) ผู้วิจัยได้ข้อมูล จำนวน 2 ชุดนำเข้าโมเดลเพื่อเปรียบเทียบประสิทธิภาพการทำนายผลการเรียนวิชา Web Database ดังนี้

ข้อมูลชุดที่ 1 เป็น Attribute ที่ได้จากการสังเคราะห์เอกสารงานวิจัยที่เกี่ยวข้องจำนวน 22 Attribute ประกอบด้วย

1) Sex (เพศ)

2) Admission1 (ประเภท โครงการเพชรตะวันออก)

3) Admission2 (ประเภท Admission ส่วนกลาง)

4) Admission3 (ประเภท สอบเข้า)

5) Admission4 (ประเภท รับตรง)

6) MajorHS1 (แผนการเรียนมัธยมปลาย กลุ่ม ศิลป์/คำนวณ)

7) MajorHS2 (แผนการเรียนมัธยมปลาย กลุ่ม ศิลป์/ภาษา)

8) MajorHS3 (แผนการเรียนมัธยมปลาย คณิต/วิทย์)

9) GPAHS (เกรดเฉลี่ยจบการศึกษาระดับชั้น ม. 6)

10) GPA Math (เกรดเฉลี่ยวิชาคณิตศาสตร์ ม.6)

11) HTML (เคยเรียนภาษา HTML)

12) CC++ (เคยเรียนภาษา C หรือ C++)

13) PHP (เคยเรียนภาษา PHP)

14) Computer (คอมพิวเตอร์ส่วนตัว)

15) Attitude (ความรู้สึกต่อการเรียนวิชาด้านคอมพิวเตอร์)

16) Review (การทบทวนบทเรียน)

17) Sub_eng (เกรดเฉลี่ยกลุ่มวิชาภาษาอังกฤษ)

18) Sub_lang (เกรดเฉลี่ยกลุ่มวิชาภาษาอื่น ๆ)

19) Identity (เกรดเฉลี่ยกลุ่มวิชาอัตลักษณ์มหาวิทยาลัย)

20) Sub_soc (เกรดเฉลี่ยกลุ่มวิชาสังคม)

21) Sub_create (เกรดเฉลี่ยกลุ่มวิชาความสร้างสรรค์)

22) Sub_ict (เกรดเฉลี่ยกลุ่มวิชาเทคโนโลยีสารสนเทศ)

ข้อมูลชุดที่ 2 เป็น Attribute ที่ได้จากการคัดกรองด้วยการวิเคราะห์ความถดถอยพหุคูณ (Multiple Regression) แบบการคัดเลือกแบบลำดับขั้น (Stepwise Selection) จำนวน 6 Attribute ประกอบด้วย

1) Sub_soc (เกรดเฉลี่ยกลุ่มวิชาสังคม)

2) PHP (เคยเรียนภาษา PHP)

3) Computer (คอมพิวเตอร์ส่วนตัว) Attitude

4) Sub_create (เกรดเฉลี่ยกลุ่มวิชาความสร้างสรรค์)

5) Identity (เกรดเฉลี่ยกลุ่มวิชาอัตลักษณ์มหาวิทยาลัย)

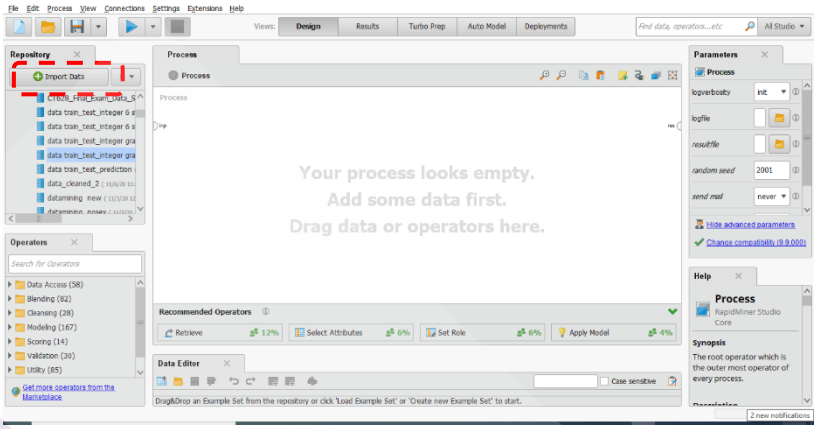
6) Attitude (ความรู้สึกต่อการเรียนวิชาด้านคอมพิวเตอร์)

3.4 ขั้นตอนที่ 4 การทำเหมืองข้อมูล (Data Mining)

3.4.1 ผู้วิจัยทำการสร้างโมเดลทำนายผลการเรียนโดยใช้โปรแกรม Rapid Miner Studio ทำการทดสอบประสิทธิภาพโมเดล โดยใช้วิธี Decision Tree, Naïve Bayes , Neural Network และ Support Vector Machine จากนั้นทำการเปรียบเทียบค่าความถูกต้อง (Accuracy), ค่าความแม่นยำ (Precision) และค่าความระลึก (Recall) ของแต่ละชุดข้อมูล และแต่วิธี เพื่อหาโมเดลทำนายที่มีประสิทธิภาพความถูกต้อง (Accuracy) ดีที่สุด

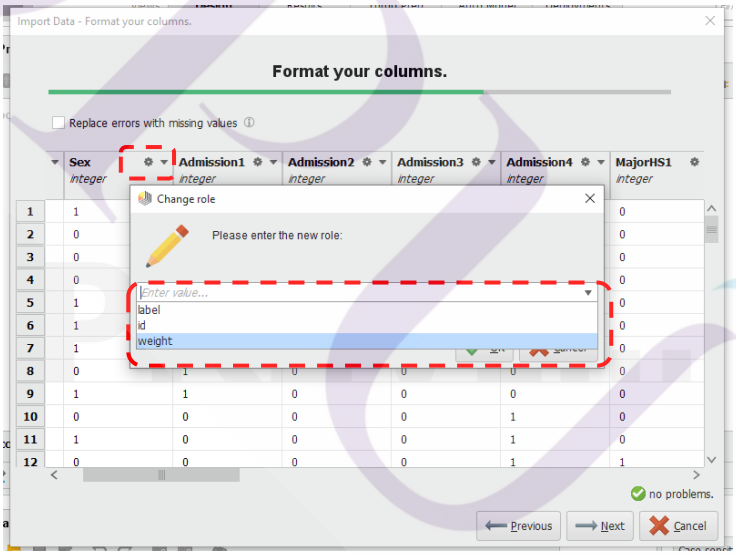
3.4.2 การนำข้อมูลชุดที่ 1, 2 เข้าโมเดลทำนายผลการเรียน ในแต่ละชุดข้อมูล ผู้วิจัยกำหนดให้ GWDB (ผลการเรียนวิชา Web Database) เป็น Label มี 2 ค่า (Binorminal) คือ กลุ่มเกรดพอใช้ Code = 0 และ กลุ่มเกรดดี Code = 1 ขั้นตอนการนำชุดข้อมูลเข้าโปรแกรม Rapid Miner Studio ดังภาพที่ 3.5 ดังนี้

เปิดโปรแกรม Rapid Miner Studio และใช้เมนู Import data เพื่อนำข้อมูลเข้าโปรแกรม



ภาพที่ 3.5 โปรแกรม Rapid Miner Studio และ เมนู Import data

หลังจากนำข้อมูลเข้าโปรแกรม ให้ดำเนินตั้งค่าชนิดของข้อมูลในแต่ละ Attribute ตัวอย่างการตั้งค่าชนิดข้อมูล ดังภาพที่ 3.6



ภาพที่ 3.6 การตั้งค่าชนิดข้อมูลแต่ละ Attribute

ในงานวิจัยนี้ได้ตั้งค่าชนิดข้อมูลของแต่ละ Attribute ดังตาราง 3.7

ตารางที่ 3.7 แสดงชนิดข้อมูลของแต่ละ Attribute

No.	Attribute	Type	Role
1	ID	polynominal	id
1	Sex	integer	
2	Admission1	integer	
3	Admission2	integer	
4	Admission3	integer	
5	Admission4	integer	
6	MajorHS1	integer	
7	MajorHS2	integer	
8	MajorHS3	integer	
9	GPAHS	real	
10	GPA Math	real	
11	HTML	integer	
12	CCPP	integer	
13	PHP	integer	
14	Computer	integer	
15	Attitude	integer	
16	Review	integer	
17	Sub_eng	real	
18	Sub_lang	real	
19	Identity	real	
20	Sub_soc	real	
21	Sub_create	real	
22	Sub_ict	real	
23	GWBD	binominal	Label

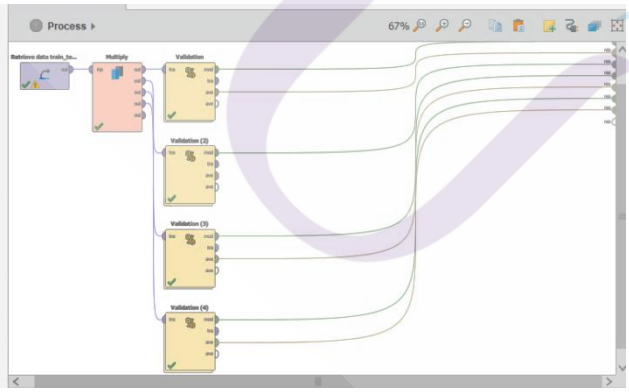
เมื่อตั้งค่าชนิดข้อมูลทั้งหมดเรียบร้อยแล้วจะได้ผลลัพธ์ดังภาพที่ 3.8

GWDB	Sex	Admission1	Admission2	Admission3	Admission4	MajorHS1	MajorHS2	MajorHS3	GPAHS	GPAMath	HTML
1	1	0	0	0	1	0	0	1	3.260	2.500	1
1	0	0	1	0	0	0	0	1	3.020	2.500	1
1	0	0	1	0	0	0	1	0	3.150	2.500	0
1	0	0	0	0	1	0	0	1	2.750	3	1
1	1	0	0	0	1	0	0	1	3.030	2.500	0
1	1	1	0	0	0	0	0	1	3.640	3.500	1
1	1	1	0	0	0	0	0	1	3.840	3.500	1
1	0	1	0	0	0	0	0	1	3.610	3.500	1
1	1	1	0	0	0	0	1	0	3.320	2.500	1
1	0	0	0	0	1	0	0	1	3.430	3.500	0
0	1	0	0	0	1	0	1	0	2.560	2.500	0
1	0	0	0	0	1	1	0	0	2.860	2.500	0
1	1	0	0	0	1	1	0	0	3.070	2.500	1
1	0	0	0	0	1	1	0	0	2.950	3	0

ภาพที่ 3.7 แสดงชนิดข้อมูลของแต่ละ Attribute

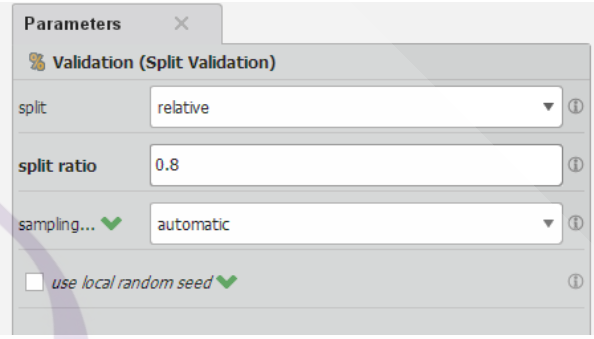
3.4.4 ขั้นตอนการสร้างโมเดลด้วยโปรแกรม Rapid Miner Studio ดังนี้

ขั้นที่ 1 เลือก Data set และ เลือก Operator Multiply เพื่อใช้ข้อมูลชุดเดียวกันเข้าโมเดลทำนายผลด้วยวิธี Decision Tree, Naïve Bayes, Support Vector Machine, Neural Network และเลือก Operator Split Validation เพื่อแบ่งกลุ่ม Train และ กลุ่ม Test ของแต่ละวิธี จากนั้นลากเส้นเชื่อมโยง Operator ต่าง ๆ เข้าด้วยกันเพื่อทำให้เกิด Process ดังภาพที่ 3.8



ภาพที่ 3.8 การสร้างโมเดลแบบ Multiply

ขั้นที่ 2 ตั้งค่าพารามิเตอร์ Operator Split Validation งานวิจัยนี้แบ่งข้อมูลออกเป็นกลุ่ม Train=80 % และ กลุ่ม Test= 20% ดังนั้นการตั้งค่าในแต่ละ Operator Validation มีดังนี้ split = relative, split ratio = 0.8 (Train set 80 %, Test set 20 %) และเลือกกลุ่มตัวอย่างแบบ Automatic ดังภาพที่ 3.9



ภาพที่ 3.9 การตั้งค่าพารามิเตอร์ Operator Validation

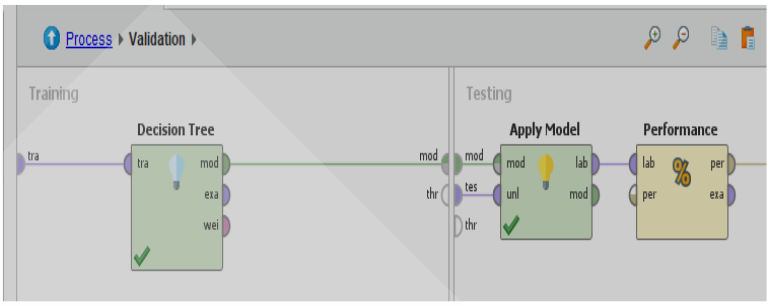
ขั้นที่ 3 เลือก Operator สำหรับสร้างโมเดลทำนาย

การสร้างโมเดล Decision Tree

1) เปิด Operator Validation ดังภาพที่ 3.10

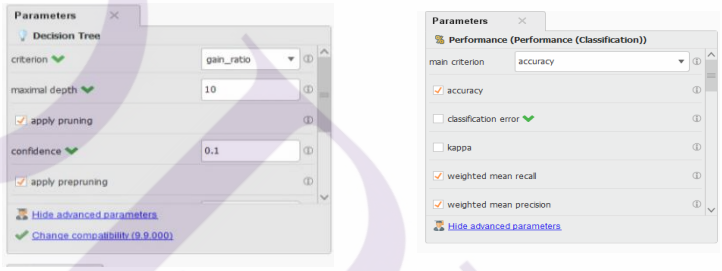
ภาพที่ 3.10 การเปิด Operator Validation

2) หลังจากเปิด Operator Validation จะพบหน้าต่าง Traing และ Teasing ในหน้าต่าง Traing เลือก Operator Decition tree และในหน้าต่าง Teasing วาง Operator Apply Model และ Operator Performance จากนั้น ลากเส้นเชื่อมโยง operator ต่าง ๆ เข้าด้วยกันเพื่อทำให้เกิด Process ดังภาพที่ 3.11



ภาพที่ 3.11 โมเดล Decision Tree

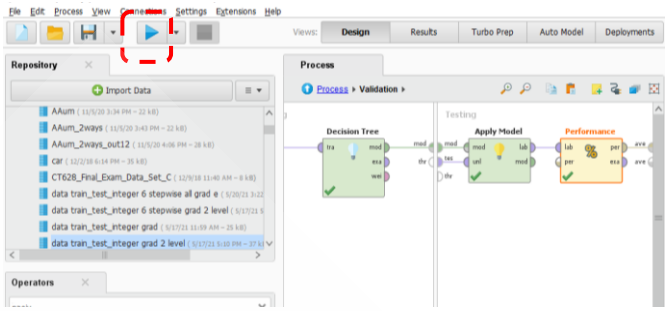
3) การตั้งค่าพารามิเตอร์ Operator Decition tree ค่าพารามิเตอร์ Operator Performance ดังภาพที่ 3.12



ภาพที่ 3.12 การตั้งค่าพารามิเตอร์ Operator Deception tree , Operator Performance

จากภาพที่ 3.12 งานวิจัยนี้ตั้งค่า deception tree ดังนี้ criterion = gain_ratio, maximal dept = 10, confidence = 0.1 และตั้งค่า Performance ดังนี้ main criterion = accuracy และเลือกผลทำนายประกอบด้วย ค่า accuracy , weighted mean recall, weighted mean precision

ขั้นที่ 4 Run โมเดล เพื่อดูค่าความถูกต้อง (Accuracy), ค่าความแม่นยำ (Precision) และค่าความระลึก (Recall) การ Run โมเดลดังภาพที่ 3.13

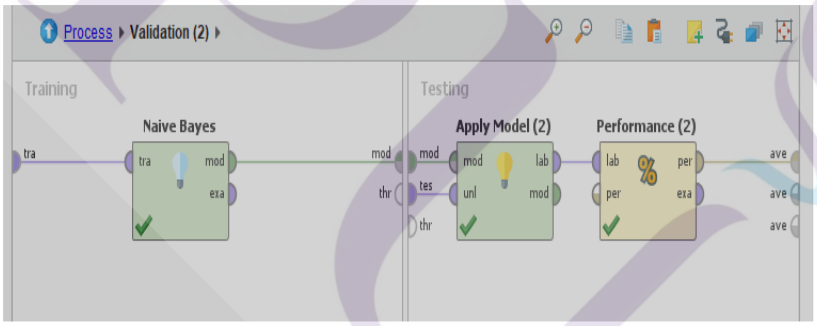


ภาพที่ 3.13 การ Run โมเดล deception tree

รายละเอียดวิธีดูผลลัพธ์ประสิทธิภาพโมเดล แสดงในหัวข้อที่ 3.5 การแปรผลและการประเมินผล (Interpretation Evaluation) หน้า 60

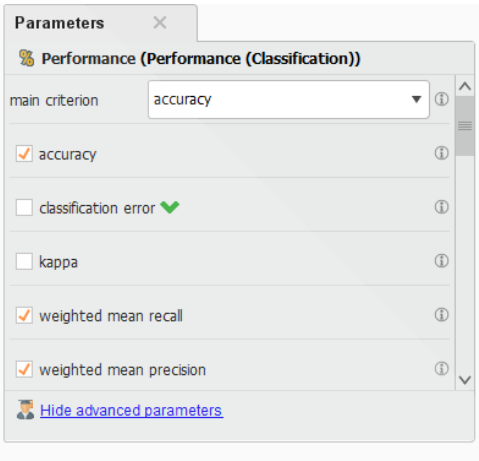
การสร้างโมเดล Naïve Bayes

1) เปิด Operator Validation (2) หน้าต่าง Traing เลือก Operator Naïve Bayes และหน้าต่าง Tesing วาง Operator Apply Model และ Operator Performance ลากเส้นเชื่อมโยง Operator ต่าง ๆ เข้าด้วยกันเพื่อทำให้เกิด Process ดังภาพที่ 3.14



ภาพที่ 3.14 โมเดล Naïve Bayes

การตั้งค่าพารามิเตอร์ Operator Performance ดังภาพที่ 3.15

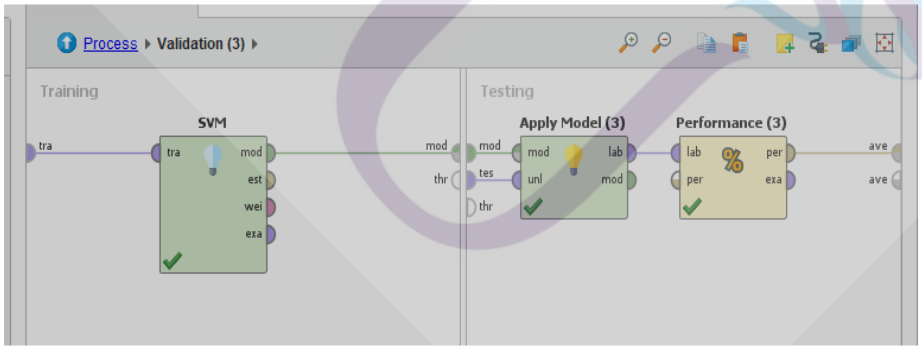


ภาพที่ 3.15 ค่าพารามิเตอร์ Operator Performance โมเดล Naïve Bayes

จากภาพที่ 3.15 งานวิจัยนี้ตั้งค่า Performance ดังนี้ main criterion = accuracy และเลือกผลทำนายประกอบด้วยค่า accuracy , weighted mean recall, weighted mean precision

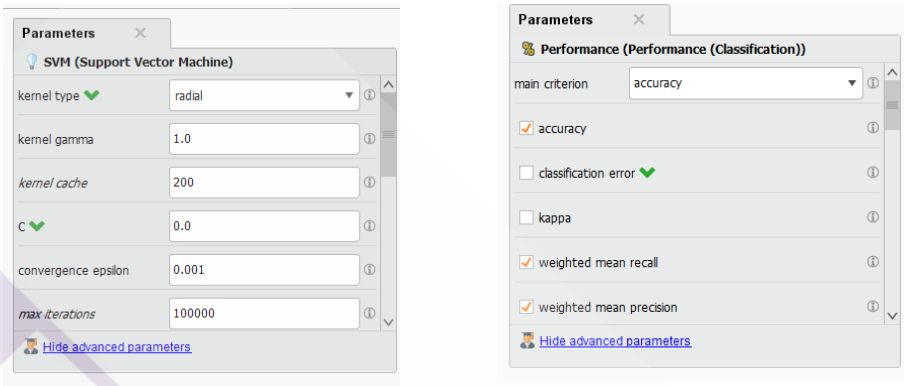
การสร้างโมเดล Support Vector Machine

1) เปิด Operator Validation (3) หน้าต่าง Traing เลือก Operator Support Vector Machine และลากเส้นเชื่อมโยง Operator ต่าง ๆ เข้าด้วยกันเพื่อทำให้เกิด Process ดังภาพที่ 3.15



ภาพที่ 3.16 โมเดล Support Vector Machine

2) การตั้งค่าพารามิเตอร์ Operator Support Vector Machine , Operator Performance และ ดังภาพที่ 3.17

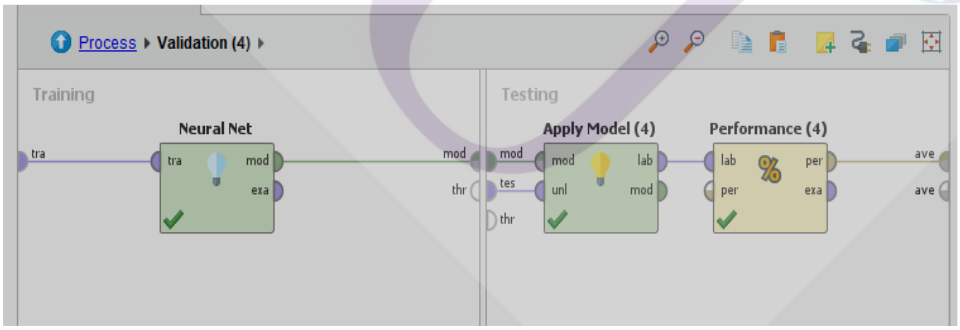


ภาพที่ 3.17 การตั้งค่าพารามิเตอร์ Operator Support Vector Machine, Operator Performance

จากภาพที่ 3.17 งานวิจัยนี้ตั้งค่า Support Vector Machine ดังนี้ kernel type = radial, kernal gamma = 1.0, kernal cache = 200, convergence epsilon = 0.001, ค่า maxiterations = 100000 และตั้งค่า Performance ดังนี้ main criterion = accuracy และเลือกผลทำนายประกอบด้วย ค่า accuracy , weighted mean recall, weighted mean precision

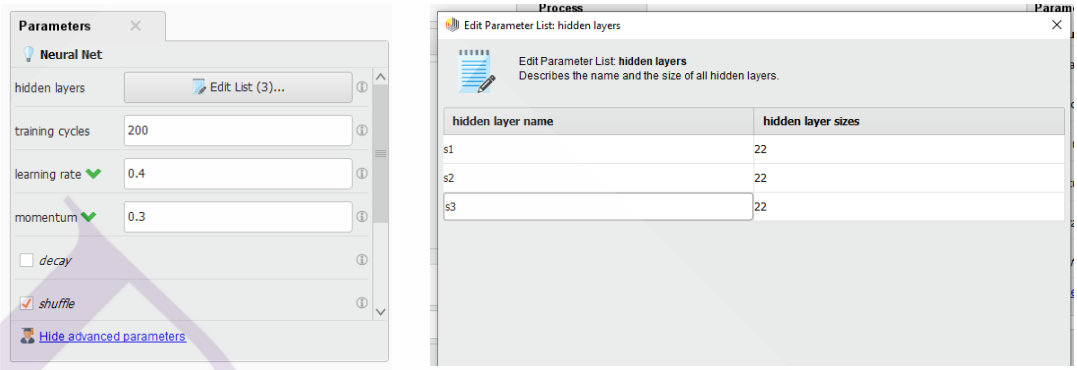
การสร้างโมเดล Neural Network

1) เปิด Operator Validation (4) หน้าต่าง Traing เลือก Operator Neural Network และลากเส้นเชื่อมโยง Operator ต่าง ๆ เข้าด้วยกันเพื่อทำให้เกิด Process



ภาพที่ 3.18 โมเดล Neural Network

2) การตั้งค่าพารามิเตอร์ Operator Neural Network, Operator Performance ดังภาพที่ 3.18



ภาพที่ 3.19 การตั้งค่าพารามิเตอร์ Neural Network

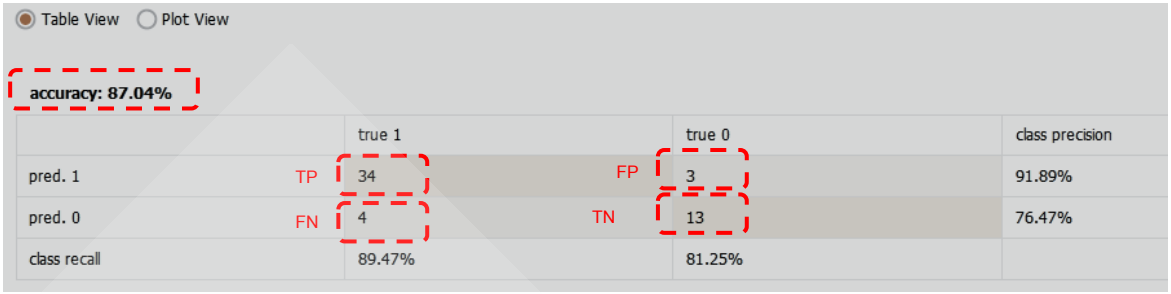
ในงานวิจัยนี้ตั้งค่า training Cycle = 200, learning rate =0.4, momentum = 0.3 , สร้าง 3 hidden layer, ตั้งค่า hidden layer sizes = 22 ส่วนการตั้งค่า Operator Performance ดังภาพที่ 3.13 จากนั้นทำการ Run โมเดล

3.5 ขั้นที่ 5 การแปรผลและการประเมินผล (Interpretation Evaluation)

หลังจากดำเนินการสร้างโมเดลทำนายผลการเรียนด้วยข้อมูลชุดที่ 1 และ ชุดที่ 2 ผู้วิจัยนำเสนอผลการทดสอบประสิทธิภาพการทำนายวิธี Decision Tree, Naïve Bayes , Neural Network และ Support Vector Machine โดยให้ความสำคัญการเปรียบเทียบค่าความถูกต้อง (Accuracy) ส่วนค่าความแม่นยำ (Precision) และค่าความระลึก (Recall) นำเสนอเพื่อประกอบผลการทำนาย

ตัวอย่างผลการทดสอบประสิทธิภาพโมเดล ด้วยโปรแกรม Rapid Miner Studio ดังนี้

การแสดงผลค่าความถูกต้อง (Accuracy) ดังภาพที่ 3.20



ภาพที่ 3.20 แสดงผลค่าความถูกต้อง (Accuracy)

จากภาพที่ 3.20 Accuracy คือ การคำนวณจำนวนคำตอบที่ถูกต้องเปรียบเทียบกับจำนวนคำตอบทั้งหมดที่นำไปให้โมเดลทำนาย [39]

สูตรการคำนวณ Accuracy = (TP + TN) / (TP + TN + FP + FN) [39]

โดยที่ TP = True positive; FP = False positive; TN = True negative;

FN = False negative

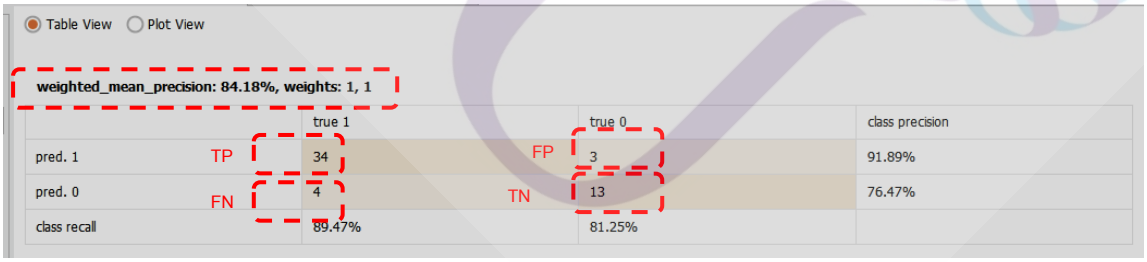
จากภาพที่ 3.20 กรณีตัวอย่างจำนวนคำตอบทั้งหมด 54 คำตอบ แยกเป็น TP = 34; FP = 3; TN = 13; FN = 4 การหาค่า Accuracy ดังนี้

Accuracy = (34 + 13) / (34+ 13 + 3 + 4)

= 47/ 54

= 0.87037 หรือ 87.04 %

การแสดงผลค่าแม่นยำ (Precision) เฉลี่ย ดังภาพที่ 3.21



ภาพที่ 3.21 แสดงผลค่าความแม่นยำ (Precision) เฉลี่ย [38]

จากภาพที่ 3.21 ค่าความแม่นยำ (Precision) หรือ ค่าที่โมเดลทำทายเป็นคลาสที่กำลังพิจารณาถูก / ค่าที่โมเดลทำนายว่าเป็นคลาสที่กำลังพิจารณาทั้งถูกและผิด [39]

ทำนายเป็น 1 Precision = TP/(TP+FP)

= 34 / (34+3)

= 34/37

= 0.9189 หรือ 91.89 %

ทำนายเป็น 0 Precision = TN/(TN+FN)

= 13 / (13+4)

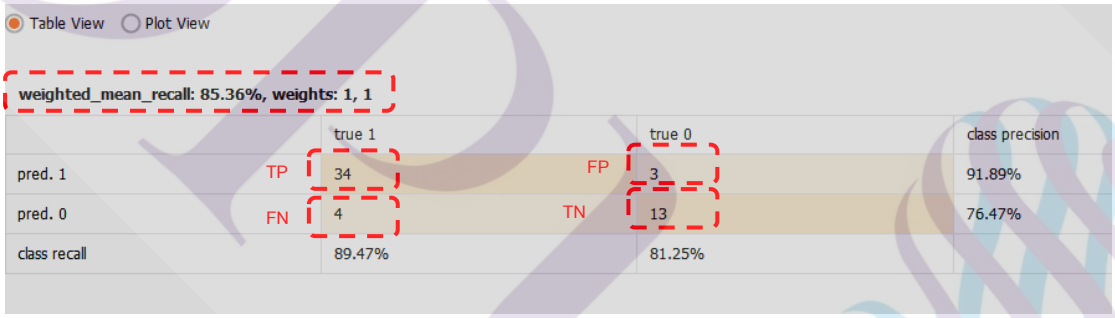
= 13/17

= 0.7647 หรือ 76.47 %

Mean precition = (91.89 +76.47)/2

= 84.18 %

การแสดงผลค่าความระลึก (Recall) เฉลี่ย ดังภาพที่ 3.22



ภาพที่ 3.22 แสดงผลค่าความระลึก (Recall) เฉลี่ย [38]

จากภาพที่ 3.22 ค่าความระลึก (Recall) คือ ค่าที่โมเดลทำนายเป็นคลาสที่กำลังพิจารณาถูก /ค่าเหตุการณ์จริงเป็นคลาสที่กำลังพิจารณาทั้งถูกและผิด [39]

ทำนายเป็น 1 Recall = TP/(TP+FN)

= 34 / (34+4)

= 34 / 38

= 0.8947 หรือ 89.47 %

ทำนายเป็น 0 Recall = TN/(TN+FP)

= 13 / (13+3)

= 13/16

= 0.8125 หรือ 81.25 %

Mean precition = (89.47 + 81.25) / 2

= 85.36

รายละเอียดผลการเปรียบเทียบประสิทธิภาพการทำนายผลการเรียน จากการสร้างโมเดลโดยใช้วิธี Decision Tree, Naïve Bayes , Neural Network และ Support Vector Machine และผลการเปรียบเทียบประสิทธิภาพการทำนายผลการเรียน โดยใช้ข้อมูลชุดที่ 1 และ ข้อมูลชุดที่ 2 ผู้วิจัยนำเสนอในบทที่ 4 ต่อไป

บทที่ 4
ผลการวิจัย

การวิจัยเรื่อง การประยุกต์ใช้ Machine Learning ทำนายผลการเรียนวิชา Web Database ของนิสิตสาขาเทคโนโลยีการศึกษา คณะศึกษาศาตร์ มหาวิทยาลัยบูรพา มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของโมเดลการทำนายผลการเรียนวิชา Web Database ระหว่างวิธี Decision Tree, Naïve Bayes, Neural Network และ Support Vector Machine กำหนดให้ GWDB (ผลการเรียนวิชา Web Database) เป็น Label มี 2 ค่า (Binominal) คือ กลุ่มเกรดพอใช้ แทนด้วย Key code 0 และ กลุ่มเกรดดี แทนด้วย Key code 1 ผู้วิจัยเสนอผลการวิเคราะห์ข้อมูลแบ่งเป็น 5 ตอน ดังนี้

ตอนที่ 1 ผลการวิเคราะห์วิเคราะห์ประสิทธิภาพโมเดลโดยใช้ข้อมูลชุดที่ 1 Attribute ที่ได้จากการสังเคราะห์เอกสารงานวิจัยที่เกี่ยวข้องจำนวน 22 Attribute

ตอนที่ 2 ผลการวิเคราะห์วิเคราะห์ประสิทธิภาพโมเดลโดยใช้ข้อมูลชุดที่ 2 Attribute ที่ได้จากการคัดกรองด้วยการวิเคราะห์ความถดถอยพหุคูณ (Multiple Regression) แบบการคัดเลือกลำดับขั้น (Stepwise Selection) จำนวน 6 Attribute

ตอนที่ 3 การเปรียบผลการวิเคราะห์ประสิทธิภาพโมเดลด้วยข้อมูลชุดที่ 1 กับข้อมูลชุดที่ 2

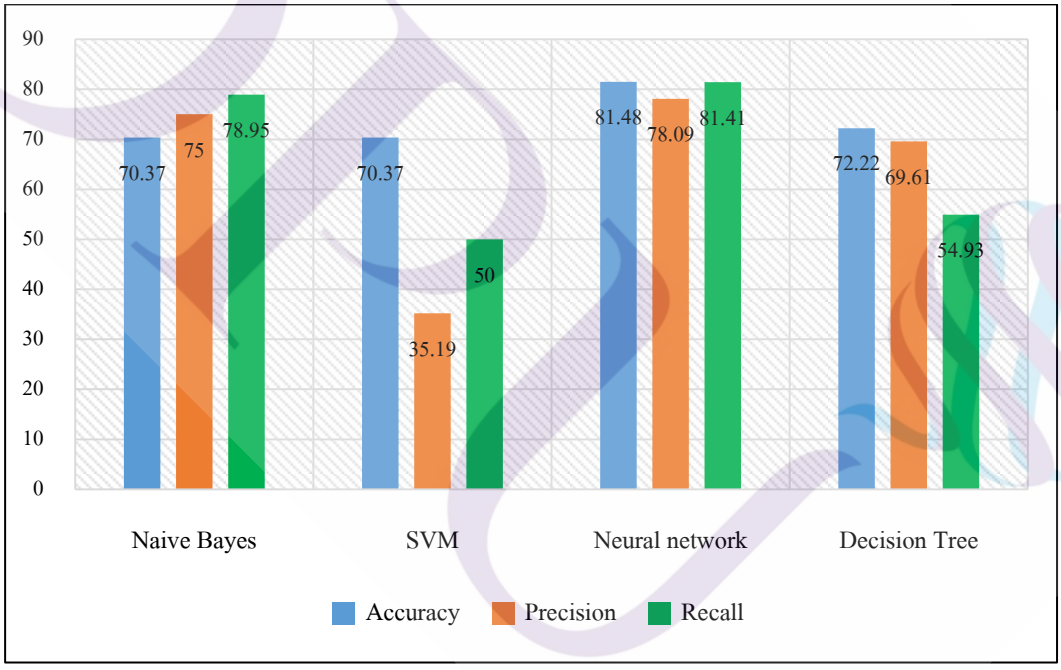
ตอนที่ 4 การปรับค่าพารามิเตอร์ Neural Network และการเปรียบเทียบประสิทธิภาพการทำนาย

ตอนที่ 5 สรุปผลการวิเคราะห์ และขั้นตอนการนำโมเดลจากการวิจัยไปใช้รายละเอียดแต่ละตอนมีดังนี้

4.1 ตอนที่ 1 ผลการวิเคราะห์วิเคราะห์ประสิทธิภาพโมเดลโดยใช้ข้อมูลชุดที่ 1 Attribute ที่ได้จากการสังเคราะห์เอกสารงานวิจัยที่เกี่ยวข้อง จำนวน 22 Attribute

ตารางที่ 4.1 การเปรียบเทียบประสิทธิภาพความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) เฉลี่ย และ ค่าความระลึก (Recall) เฉลี่ย ชุดข้อมูลที่ 1 จำนวน 22 Attribute

Algorithms	Accuracy (%)	Mean_Precision (%)	Mean_Recall (%)
Naïve Bayes	70.37	75.00	78.95
Support Vector Machine	70.37	35.19	50.00
Neural Network	81.48	78.09	81.41
Decision Tree	72.22	69.61	54.93



ภาพที่ 4.1 การเปรียบเทียบประสิทธิภาพวิธี Naïve Bayes , Support Vector Machine, Neural Network และ Decision Tree ชุดข้อมูลที่ 1 จำนวน 22 Attribute

จากตารางที่ 4.1 และ ภาพที่ 4.1 สรุปได้ว่า การนำชุดข้อมูลที่ 1 จำนวน 22 Attribute ที่ได้จากการสังเคราะห์เอกสารและงานวิจัยที่เกี่ยวข้อง มาสร้างโมเดลเพื่อทำนายผลการเรียนวิชา Web Database พบว่า ประสิทธิภาพทำนายวิธี Neural Network มีค่าความถูกต้อง (Accuracy) มากที่สุด 81.48 % รองลงมาวิธี Decision Tree 72.22 % ลำดับสุดท้ายวิธี Naïve Bayes และ Support Vector Machine มีค่าความถูกต้อง (Accuracy) เท่ากัน 70.37 %

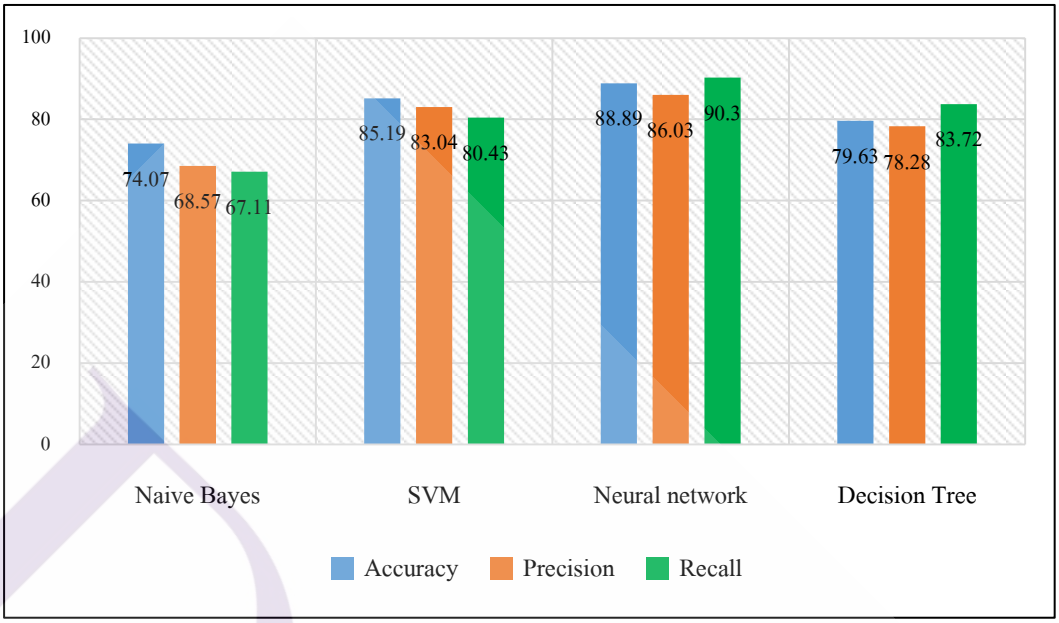
ด้านการเปรียบเทียบ ค่าความแม่นยำ (Precision) ฉลี่ย พบว่าวิธี Neural Network มากที่สุด 78.09 % รองลงมาวิธี Naïve Bayes 75.00 % อันดับที่สาม Decision Tree 69.61 % และ ลำดับสุดท้ายวิธี Support Vector Machine 35.19 %

ด้านการเปรียบเทียบ ค่าความระลึก (Recall) เฉลี่ย พบว่าวิธี Neural Network มากที่สุด 81.41% รองลงมาวิธี Naïve Bayes 78.95 % อันดับที่สาม Decision Tree 54.93 % และ ลำดับสุดท้ายวิธี Support Vector Machine 50.00 %

4.2 ตอนที่ 2 ผลการวิเคราะห์วิเคราะห์ประสิทธิภาพโมเดลโดยใช้ข้อมูลชุดที่ 2 Attribute ที่ได้จากการคัดกรองด้วยการวิเคราะห์ความถดถอยพหุคูณ (Multiple Regression) แบบการคัดเลือกลำดับขั้น (Stepwise Selection) จำนวน 6 Attribute

ตารางที่ 4.2 การเปรียบเทียบประสิทธิภาพความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) และ ค่าความระลึก (Recall) ชุดข้อมูลที่ 2 จำนวน 6 Attribute

Algorithms	Accuracy (%)	Mean_Precision (%)	Mean_Recall (%) (%)
Naïve Bayes	74.07	68.57	67.11
Support Vector Machine	85.19	83.04	80.43
Neural Network	88.89	86.03	90.30
Decision Tree	79.63	78.28	83.72



ภาพที่ 4.2 การเปรียบเทียบประสิทธิภาพวิธี Naïve Bayes , Support Vector Machine, Neural Network และ Decision Tree ชุดข้อมูลที่ 2 จำนวน 6 Attribute

จากตารางที่ 4.2 และ ภาพที่ 4.2 สรุปได้ว่า การนำชุดข้อมูลที่ 2 จำนวน 6 Attribute ที่ได้จากการคัดกรองด้วยการวิเคราะห์ความถดถอยพหุคูณ (Multiple Regression) แบบการคัดเลือกลำดับขั้น (Stepwise Selection) มาสร้างโมเดลเพื่อทำนายผลการเรียนวิชา Web Database พบว่าประสิทธิภาพทำนายวิธี Neural Network มีค่าความถูกต้อง (Accuracy) มากที่สุด 88.89 % รองลงมาวิธี Support Vector Machine 85.19 % ลำดับที่สามวิธี Decision Tree 79.63 % ลำดับสุดท้ายวิธี Naïve Bayes 74.07 %

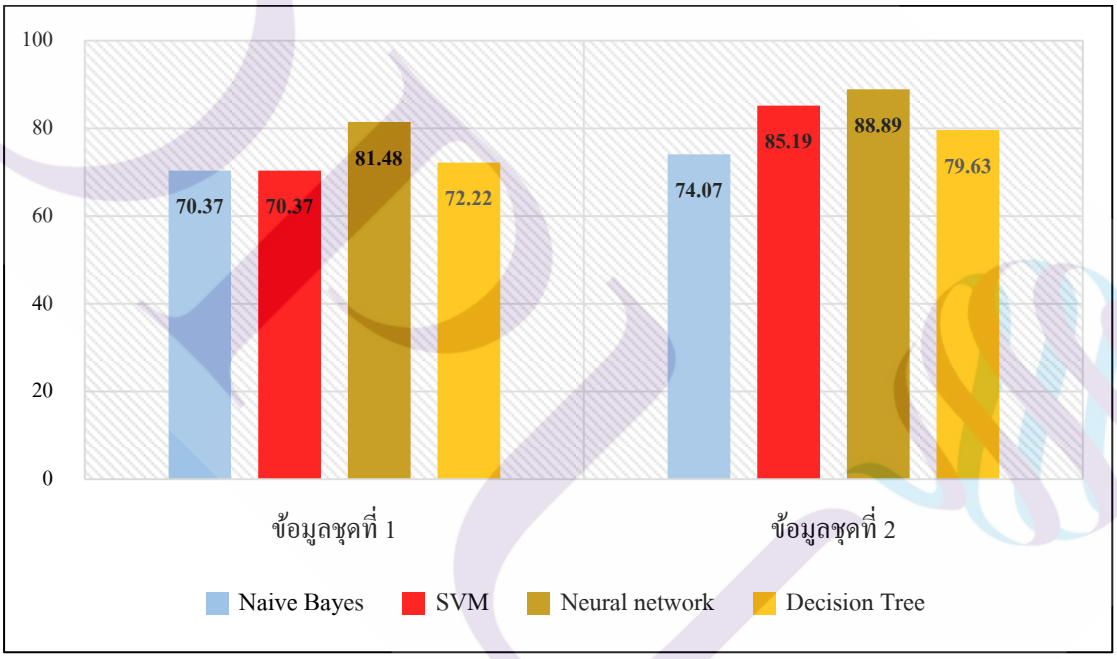
ด้านการเปรียบเทียบ ค่าความแม่นยำ (Precision) เฉลี่ย พบว่าวิธี Neural Network มากที่สุด 86.03 % รองลงมาวิธี Support Vector Machine 83.04 % อันดับที่สาม Decision Tree 78.28 % และลำดับสุดท้ายวิธี Naïve Bayes 68.57 %

ด้านการเปรียบเทียบ ค่าความระลึก (Recall) เฉลี่ย พบว่าวิธี Neural Network มากที่สุด 90.30 % รองลงมาวิธี Decision Tree 83.72 % อันดับที่สามวิธี Support Vector Machine 80.43 % และลำดับสุดท้ายวิธี Naïve Bayes 67.11 %

4.3 ตอนที่ 3 การเปรียบผลการวิเคราะห์ประสิทธิภาพโมเดลด้วยข้อมูลชุดที่ 1 กับข้อมูลชุดที่ 2

ตารางที่ 4.3 ค่าความถูกต้อง (Accuracy) การวิเคราะห์ด้วยข้อมูลชุดที่ 1 และข้อมูลชุดที่ 2

Algorithms	ค่าความถูกต้อง (Accuracy) ข้อมูลชุดที่ 1	ค่าความถูกต้อง (Accuracy) ข้อมูลชุดที่ 2
Naïve Bayes	70.37	74.07
Support Vector Machine	70.37	85.19
Neural Network	81.48	88.89
Decision Tree	72.22	79.63



ภาพที่ 4.3 การเปรียบเทียบค่าความถูกต้อง (Accuracy) วิธี Naïve Bayes , Support Vector Machine, Neural Network และ Decision Tree ข้อมูลชุดที่ 1 และชุดที่ 3

จากตารางที่ 4.3 และ ภาพที่ 4.3 สรุปได้ว่า จากการเปรียบเทียบค่าความถูกต้อง (Accuracy) ของการใช้ข้อมูลชุดที่ 1 และข้อมูลชุดที่ 2 ทำนายผลการเรียน พบว่าโมเดลทำนายโดยใช้ข้อมูลชุดที่ 2 มีค่าความถูกต้อง (Accuracy) เพิ่มขึ้นทุกวิธี พิจารณารายวิธีพบว่า วิธี Naïve Bayes เพิ่มขึ้น 3.7 % , Support Vector Machine เพิ่มขึ้น 14.82 %, Neural Network เพิ่มขึ้น 7.41 %,

และ Decision Tree เพิ่มขึ้น 7.41 %, และสรุปได้ว่า โมเดลการทำนายผลการเรียนวิชา Web Database ด้วยวิธี Neural Network นำข้อมูล 6 Attribute เข้าโมเดลทำนาย มีค่าความถูกต้องมากที่สุด (Accuracy) 88.89 %

4.4 ตอนที่ 4 การปรับค่าพารามิเตอร์ Neural Network และการเปรียบเทียบประสิทธิภาพการทำนาย

จากการที่ผู้วิจัยได้กำหนดค่าพารามิเตอร์โมเดล Neural Network เบื้องต้น ได้แก่ Training Cycle = 200, Learning rate =0.4, Momentum = 0.3 , จำนวน 3 Hidden layer, ค่า Hidden layer sizes = 22 จากผลการวิจัยพบว่าโมเดล Neural Network มีค่าความถูกต้อง (Accuracy) มากที่สุด = 88.89 %

เพื่อให้ได้โมเดล Neural Network ที่เหมาะสมที่สุด ผู้วิจัยได้ดำเนินการทดลองปรับค่าพารามิเตอร์ของโมเดล Neural Network โดยทดลองปรับค่าพารามิเตอร์ให้มีค่าต่ำกว่า และสูงกว่าค่าพารามิเตอร์ที่ผู้วิจัยได้กำหนดขึ้นในเบื้องต้น ค่าพารามิเตอร์โมเดล Neural Network ที่ทดลองปรับ ได้แก่จำนวน Training Cycle, จำนวน Hidden layer, จำนวน Hidden layer sizes, ค่า Learning rate และ ค่า Momentum

Training Cycle หมายถึง จำนวนรอบในการเรียนรู้

Hidden layer หมายถึง เป็นชั้นที่อยู่กลางระหว่าง Input Layer กับ Output Layer ซึ่งจะมีผลอย่างมากต่อประสิทธิภาพในการเรียนรู้ของโมเดล

Hidden layer sizes หมายถึง จำนวนโหนด (Node) ที่อยู่ใน Hidden layer

Learning rate หมายถึง ปริมาณที่ต้องเปลี่ยนน้ำหนักในแต่ละ step ว่ามากน้อยเพียงใด

Momentum หมายถึง การเพิ่มสัดส่วนของน้ำหนักในครั้งก่อนหน้าเพื่อ update ค่าปัจจุบัน เพื่อป้องกันค่าสูงสุดในบริเวณใดๆ (ไม่ใช่ค่าสูงสุดจริงๆ) และ ความต่อเนื่องของการหาค่าที่ดีที่สุด

ผลจากการปรับค่าพารามิเตอร์ของโมเดล Neural Network แสดงดังตารางที่ 4.4 - 4.7

ตารางที่ 4.4 ผลการทดลองปรับค่า Training Cycle ของโมเดล Neural Network

Training Cycle	Hidden layer	Hidden layer sizes	Learning rate	Momentum	Accuracy
50	3	22	0.4	0.3	85.19 %
100	3	22	0.4	0.3	85.19 %
200	3	22	0.4	0.3	88.89 %
300	3	22	0.4	0.3	87.04 %
400	3	22	0.4	0.3	87.04 %
500	3	22	0.4	0.3	88.89 %
600	3	22	0.4	0.4	87.04 %
700	3	22	0.4	0.4	87.04 %
1000	3	22	0.4	0.4	87.04 %

ตารางที่ 4.4 พบว่า เมื่อปรับจำนวน Training Cycle ของโมเดล Neural Network น้อยกว่าจำนวน Training Cycle ที่ผู้วิจัยกำหนดขึ้นในเบื้องต้น (Training Cycle =200) โดยปรับจำนวน Training Cycle = 100 และ 50 พบว่า ประสิทธิภาพการทำนาย ค่าความถูกต้อง (Accuracy) ลดลง 3.7 %

เมื่อปรับจำนวน Training Cycle มากกว่าจำนวน Training Cycle ที่ผู้วิจัยกำหนดขึ้นในเบื้องต้น (Training Cycle =200) โดยปรับจำนวน Training Cycle = 300, 400, 500, 600,700 และ 1000 พบว่าประสิทธิภาพการทำนายของ Training Cycle = 300, 400,600,700 และ 1000 ค่าความถูกต้อง (Accuracy) ลดลง 1.85 % ส่วน Training Cycle = 500 พบว่าค่าความถูกต้อง (Accuracy) เท่ากับการตั้งค่าเบื้องต้น (Training Cycle =200) ค่าความถูกต้อง (Accuracy) = 88.89%

ดังนั้นจำนวน Training Cycle ที่เหมาะสมสำหรับโมเดลทำนายผลการเรียนวิชา Web Database คือ Training Cycle = 200 เนื่องจากมีความคุ้มค่าด้านพลังงานและระยะเวลาที่ใช้ในการวิเคราะห์ข้อมูล

ตารางที่ 4.5 ผลการทดลองปรับจำนวน Hidden layer ของโมเดล Neural Network

Training Cycle	Hidden layer	Hidden layer sizes	Learning rate	Momentum	Accuracy
200	1	22	0.4	0.3	92.59 %
200	2	22	0.4	0.3	87.04 %
200	3	22	0.4	0.3	88.89 %
200	4	22	0.4	0.3	70.37 %
200	5	22	0.4	0.3	70.37 %
200	6	22	0.4	0.3	70.37 %
200	10	22	0.4	0.4	70.37 %

ตารางที่ 4.5 พบว่า เมื่อลดจำนวน Hidden layer ของโมเดล Neural Network น้อยกว่าจำนวน Hidden layer ที่ผู้วิจัยกำหนดขึ้นในเบื้องต้น (Hidden layer = 3) โดยปรับจำนวน Hidden layer ลดลงครั้งละ 1 พบว่าประสิทธิภาพการทำนายของโมเดล Hidden layer ค่าความถูกต้อง (Accuracy) เพิ่มขึ้น 3.71% ส่วนประสิทธิภาพการทำนายของโมเดล 2 Hidden layer ลดลง 1.85%

เมื่อเพิ่มจำนวน Hidden layer ของโมเดล Neural Network มากกว่าจำนวน Hidden layer ที่ผู้วิจัยกำหนดขึ้นในเบื้องต้น (Hidden layer = 3) โดยปรับจำนวน Hidden layer เพิ่มครั้งละ 1 พบว่าประสิทธิภาพการทำนายของโมเดล Hidden layer = 4, 5, 6 ค่าความถูกต้อง (Accuracy) ลดลง 3.71 % และ ประสิทธิภาพการทำนายของโมเดล 10 Hidden layer ลดลง 3.71 % เท่ากัน

ดังนั้น จำนวน Hidden layer ที่เหมาะสมสำหรับโมเดลทำนายผลการเรียนวิชา Web Database คือ 1 Hidden layer และ ผู้วิจัยได้ใช้โมเดล 1 Hidden layer ในการทดลองปรับ Hidden layer sizes, Learning rate และ Momentum ในขั้นตอนต่อไป

ตารางที่ 4.6 ผลการทดลองปรับจำนวน Hidden layer sizes ของโมเดล Neural Network

Training Cycle	Hidden layer	Hidden layer sizes	Learning rate	Momentum	Accuracy
200	1	19	0.4	0.3	87.04 %
200	1	20	0.4	0.3	87.04 %
200	1	21	0.4	0.3	87.04 %
200	1	22	0.4	0.3	92.59 %
200	1	23	0.4	0.3	90.74 %
200	1	24	0.4	0.3	87.04 %
200	1	25	0.4	0.4	87.04 %

ตารางที่ 4.6 พบว่า เมื่อลดจำนวน Hidden layer sizes ของโมเดล Neural Network น้อยกว่าจำนวน Hidden layer sizes ที่ผู้วิจัยกำหนดขึ้นในเบื้องต้น (Hidden layer sizes = 22) โดยปรับจำนวน Hidden layer sizes ลดลงครั้งละ 1 พบว่าประสิทธิภาพการทำนายของโมเดล Hidden layer sizes = 21, 22, 23 ค่าความถูกต้อง (Accuracy) ลดลง 5.55 % เท่ากัน

เมื่อเพิ่มจำนวน Hidden layer sizes ของโมเดล Neural Network มากกว่าจำนวน Hidden layer ที่ผู้วิจัยกำหนดขึ้นในเบื้องต้น (Hidden layer sizes = 22) โดยปรับจำนวน Hidden layer sizes เพิ่มครั้งละ 1 พบว่าประสิทธิภาพการทำนายของโมเดล Hidden layer sizes = 23 ค่าความถูกต้อง (Accuracy) ลดลง 1.85 % และ ประสิทธิภาพการทำนายของโมเดล Hidden layer sizes = 24, 25 ลดลง 5.55 % เท่ากัน

ดังนั้น จำนวน Hidden layer sizes ที่เหมาะสมสำหรับโมเดลทำนายผลการเรียนวิชา Web Database คือ Hidden layer sizes = 22 และ ผู้วิจัยได้ใช้โมเดล 1 Hidden layer, 22 Hidden ในการทดลองปรับ Learning rate และ Momentum ในขั้นตอนต่อไป

ตารางที่ 4.7 ผลการทดลองปรับค่า Learning rate ของโมเดล Neural Network

Training Cycle	Hidden layer	Hidden layer sizes	Learning rate	Momentum	Accuracy
200	1	22	0.1	0.3	87.04 %
200	1	22	0.2	0.3	87.04 %
200	1	22	0.3	0.3	88.89 %
200	1	22	0.4	0.3	92.59 %
200	1	22	0.5	0.3	92.59 %
200	1	22	0.6	0.3	92.59 %
200	1	23	0.7	0.3	92.59 %
200	1	24	0.9	0.3	92.59 %
200	1	25	1.0	0.3	92.59 %

ตารางที่ 4.7 พบว่า เมื่อลดค่า Learning rate ของโมเดล Neural Network น้อยกว่าค่า Learning rate ที่ผู้วิจัยกำหนดขึ้นในเบื้องต้น (Learning rate = 0.4) โดยปรับค่า Learning rate ลดลงครั้งละ 0.1 พบว่า Learning rate = 0.3 ประสิทธิภาพการทำนายลดลง 3.7 % และ Learning rate = 0.2, 0.1 ลดลง 5.55 %

เมื่อเพิ่มค่า Learning rate ของโมเดล Neural Network มากกว่าค่า Learning rate ที่ผู้วิจัยกำหนดขึ้นในเบื้องต้น (Learning rate = 0.4) โดยปรับค่า Learning rate เพิ่มครั้งละ 0.1 พบว่าประสิทธิภาพการทำนายของโมเดล Learning rate = 0.5, 0.6, 0.7, 0.8, 0.9 และ 1.0 มีประสิทธิภาพการทำนาย เท่ากัน กับโมเดลค่า Learning rate ที่ผู้วิจัยกำหนดขึ้นในเบื้องต้น

ดังนั้น ในงานวิจัยครั้งนี้ค่า Learning rate ที่เหมาะสมสำหรับโมเดลทำนายผลการเรียนวิชา Web Database คือ 0.4 – 1.00 และ ผู้วิจัยได้ใช้โมเดล 1 Hidden layer, 22 Hidden, Learning rate = 0.4 ในการทดลองปรับค่า Momentum ในขั้นตอนต่อไป

ตารางที่ 4.8 ผลการทดลองปรับค่า Momentum ของโมเดล Neural Network

Training Cycle	Hidden layer	Hidden layer sizes	Learning rate	Momentum	Accuracy
200	1	22	0.4	0.1	88.89 %
200	1	22	0.4	0.2	90.74 %
200	1	22	0.4	0.3	92.59 %
200	1	22	0.4	0.4	92.59 %
200	1	22	0.4	0.5	92.59 %
200	1	22	0.4	0.6	92.59 %
200	1	22	0.4	0.7	92.59 %
200	1	22	0.4	0.8	92.59 %
200	1	22	0.4	0.9	90.74 %
200	1	22	0.4	1.0	70.37 %

ตารางที่ 4.8 พบว่า เมื่อลดค่า Momentum ของโมเดล Neural Network น้อยกว่าค่า Momentum ที่ผู้วิจัยกำหนดขึ้นในเบื้องต้น (Momentum = 0.3) โดยปรับ Momentum ลดลงครั้งละ 0.1 พบว่าโมเดลค่า Momentum = 0.2 ประสิทธิภาพการทำนายลดลง 1.85 % และโมเดลค่า Momentum = 0.1 ลดลง 3.7 %

เมื่อเพิ่มค่า Momentum ของโมเดล Neural Network มากกว่า ค่า Momentum ที่ผู้วิจัยกำหนดขึ้นในเบื้องต้น (Momentum = 0.3) โดยปรับค่า Momentum เพิ่มครั้งละ 0.1 พบว่าประสิทธิภาพการทำนายของโมเดลค่า Momentum = 0.4, 0.5, 0.6, 0.7, และ 0.8 ประสิทธิภาพการทำนาย เท่ากัน กับโมเดล ค่า Momentum ที่ผู้วิจัยกำหนดขึ้นในเบื้องต้น 1 ส่วนประสิทธิภาพการทำนายโมเดล ค่า Momentum = 0.9 มีค่าความถูกต้องลดลง 1.85 % และโมเดล Momentum = 1.0 มีค่าความถูกต้องลดลง 22.22 %

ดังนั้นในงานวิจัยครั้งนี้ค่า Momentum ที่เหมาะสมสำหรับโมเดลทำนายผลการเรียนวิชา Web Database คือ 0.3 – 0.8 โดยผู้วิจัยเลือกค่า Momentum น้อยที่สุด คือ 0.3 ในการปรับค่าโมเดล

จากผลการปรับค่าพารามิเตอร์ของโมเดล Neural Network ทำให้ผู้วิจัยได้ค่าพารามิเตอร์ของโมเดล Neural Network ที่เหมาะสมที่สุด คือ Training Cycle = 200, Hidden layer = 1, Hidden layer sizes = 22, Learning rate = 0.4 และ Momentum = 0.3 โดยค่าพารามิเตอร์ที่ปรับใหม่แตกต่างจากค่าพารามิเตอร์ที่ผู้วิจัยกำหนดขึ้นในเบื้องต้น คือ จำนวน Hidden layer ลดลงจาก 3 Hidden layer เหลือ 1 Hidden layer ซึ่งส่งผลให้โมเดล Neural Network มีประสิทธิภาพการทำนายเพิ่มขึ้น จากค่าความถูกต้อง (Accuracy) 88.89 % เป็น 92.59 % (เพิ่มขึ้น 3.7 %) ดังภาพที่ 4.4

accuracy: 92.59%

	true 1	true 0	class precision
pred. 1	35	1	97.22%
pred. 0	3	15	83.33%
class recall	92.11%	93.75%	

ภาพที่ 4.4 ประสิทธิภาพการทำนาย โมเดล Neural Network

จากภาพที่ 4.4 พบว่า โมเดล Neural Network มีมีค่าความถูกต้อง (Accuracy) 92.59 % ค่าความแม่นยำ (Precision) เฉลี่ย 90.28 % และ ค่าความระลึก (Recall) เฉลี่ย 92.93 %

4.5 ตอนที่ 5 สรุปผลการวิเคราะห์ และขั้นตอนการนำโมเดลที่มีประสิทธิภาพมากที่สุดไปใช้

4.5.1 ผลการวิจัยพบว่า ในการเตรียมข้อมูล (Attribute) เพื่อนำเข้าโมเดล ทำนายผลการเรียนวิชา Web Database ของนิสิตสาขาเทคโนโลยีการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยบูรพา โดยแบ่งผลการเรียนวิชา Web Database เป็น 2 ค่า (Binominal) คือ กลุ่มเกรดพอใช้ และ กลุ่มเกรดดี สรุปได้ว่า Attribute ที่ส่งผลต่อค่าความถูกต้องของการทำนายมากที่สุด ประกอบด้วยข้อมูล 6 Attribute ได้แก่

1) Sub_soc (เกรดเฉลี่ยกลุ่มวิชาสังคม)
2) PHP (เคยเรียนภาษา PHP)
3) Computer (คอมพิวเตอร์ส่วนตัว) Attitude
4) Sub_create (เกรดเฉลี่ยกลุ่มวิชาความสร้างสรรค์)

5) Identity (เกรดเฉลี่ยกลุ่มวิชาอัตลักษณ์มหาวิทยาลัย)

6) Attitude (ความรู้สึกต่อการเรียนวิชาด้านคอมพิวเตอร์)

4.5.2 ประสิทธิภาพของโมเดลการทำนายผลการเรียนวิชา Web Database ผลการวิจัยสรุปได้ว่าวิธี Neural Network เป็นวิธีการที่นำไปสร้างโมเดลทำนายผลการเรียนวิชา Web database โดยใช้ข้อมูล 6 Attribute ให้ค่าความถูกต้อง (Accuracy) มากที่สุด 92.59 %

4.5.3 การนำโมเดลดีที่สุดจากการวิจัย ได้แก่วิธี Neural Network ไปใช้จริงในการทำนายผลการเรียนวิชา Web Database ของนิสิตสาขาเทคโนโลยีการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยบูรพา มีขั้นตอนการดำเนินการดังนี้

4.5.3.1 เก็บข้อมูลของนิสิตเพื่อการทำผลการเรียน จำนวน 6 Attribute ประกอบ

1) Sub_soc (เกรดเฉลี่ยกลุ่มวิชาสังคม)

2) PHP (เคยเรียนภาษา PHP)

3) Computer (คอมพิวเตอร์ส่วนตัว) Attitude

4) Sub_create (เกรดเฉลี่ยกลุ่มวิชาความสร้างสรรค์)

5) Identity (เกรดเฉลี่ยกลุ่มวิชาอัตลักษณ์มหาวิทยาลัย)

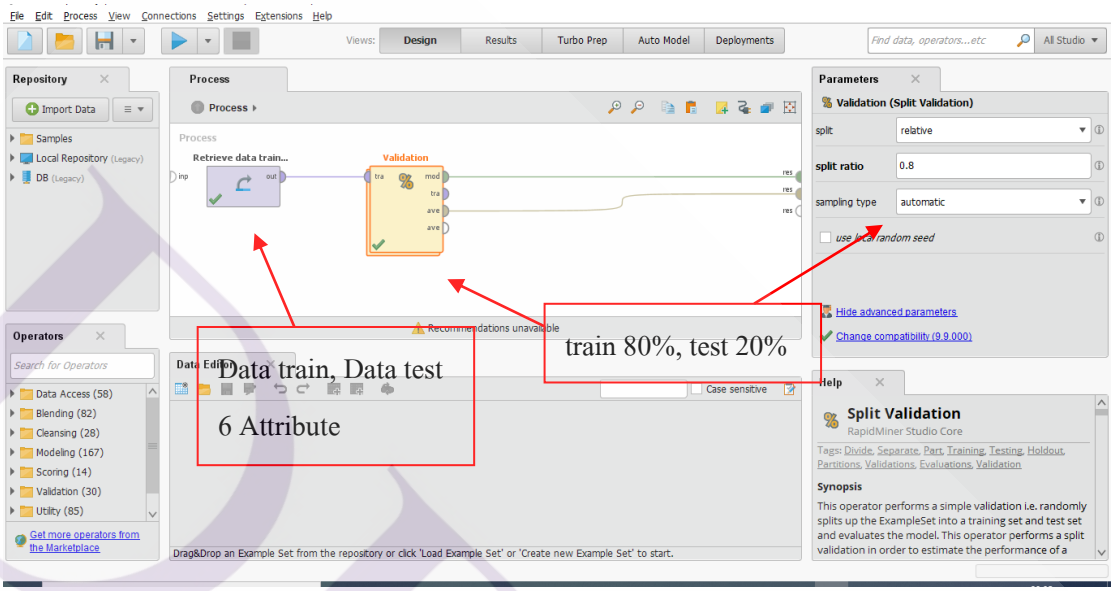
6) Attitude (ความรู้สึกต่อการเรียนวิชาด้านคอมพิวเตอร์)

4.5.3.2 เปลี่ยนรูปข้อมูล (Transformation) แปลง Key code ของข้อมูลบันทึกไฟล์เป็น Excel โดย Key code ตามตาราง 4.9 ดังนี้

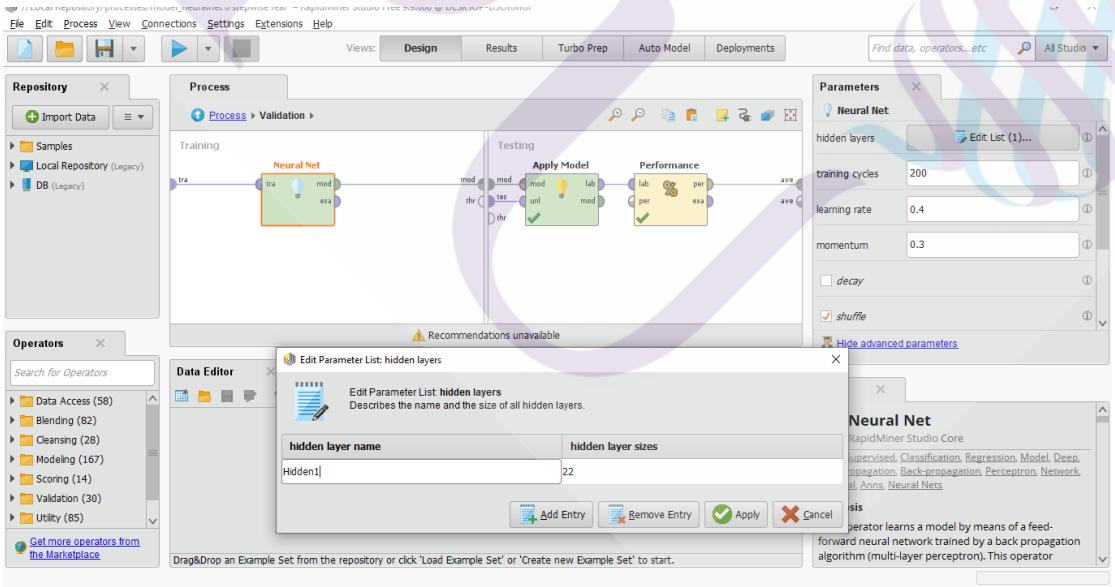
ตารางที่ 4.9 การทำ Key code ข้อมูลสำหรับทำนายผลการเรียน

ลำดับที่	ชื่อ Attribute	Detail	Key code
1	PHP	เคยเรียนภาษา PHP	0 = ไม่เคย 1 = เคย
2	Computer	คอมพิวเตอร์ส่วนตัว	0 = ไม่มี 1 = มี
3	Attitude	ความรู้สึกต่อการเรียนวิชาด้านคอมพิวเตอร์	0 = ไม่ชอบ 1 = ชอบ
4	Identity	เกรดเฉลี่ยกลุ่มวิชาอัตลักษณ์มหาวิทยาลัย	ใช้ค่าเกรดตามจริง
5	Sub_soc	เกรดเฉลี่ยกลุ่มวิชาสังคม	ใช้ค่าเกรดตามจริง
6	Sub_create	เกรดเฉลี่ยกลุ่มวิชาความสร้างสรรค์	ใช้ค่าเกรดตามจริง

4.5.3.3 เปิดโปรแกรม RapidMiner Studio และเปิด Process โมเดลการวิเคราะห์ด้วยวิธี Neural Network ที่มีค่าความถูกต้องมากที่สุด เพื่อจะใช้ทำนายผลการเรียน ตัวอย่างโมเดลทำนายผลการเรียนด้วยวิธี Neural Network ดังภาพที่ 4.5



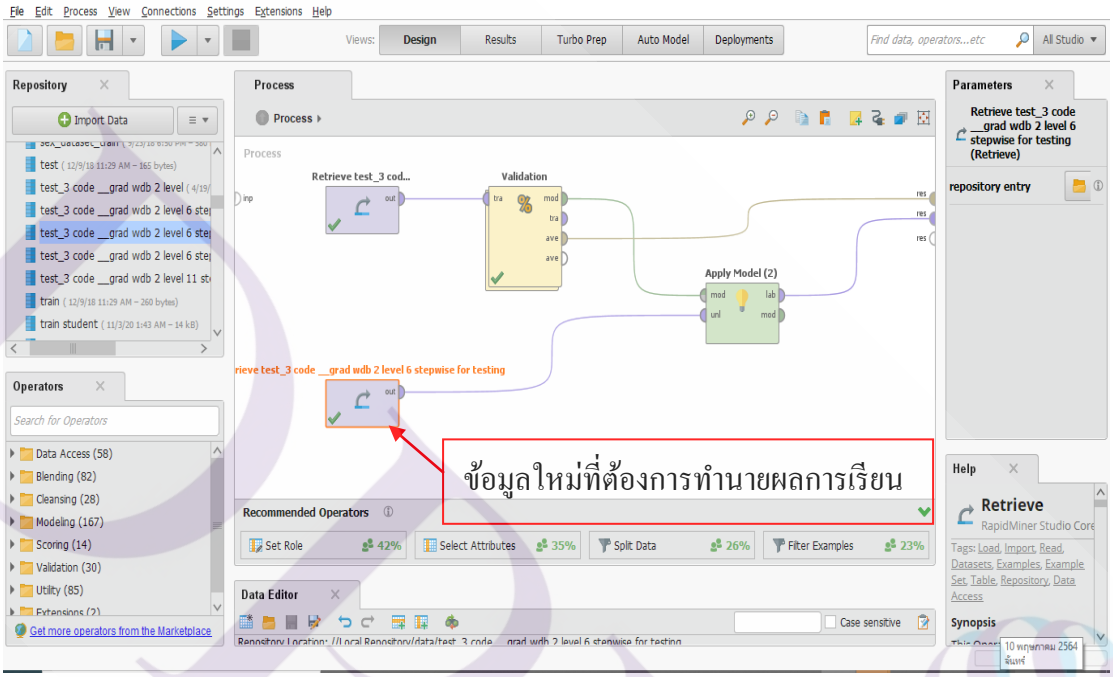
ภาพที่ 4.5 โมเดลทำนายผลการเรียนด้วยวิธี Neural Network ที่มีค่าความถูกต้องมากที่สุด



ภาพที่ 4.6 การสร้างโมเดลการทำนายผลการเรียนด้วยวิธี Neural Network

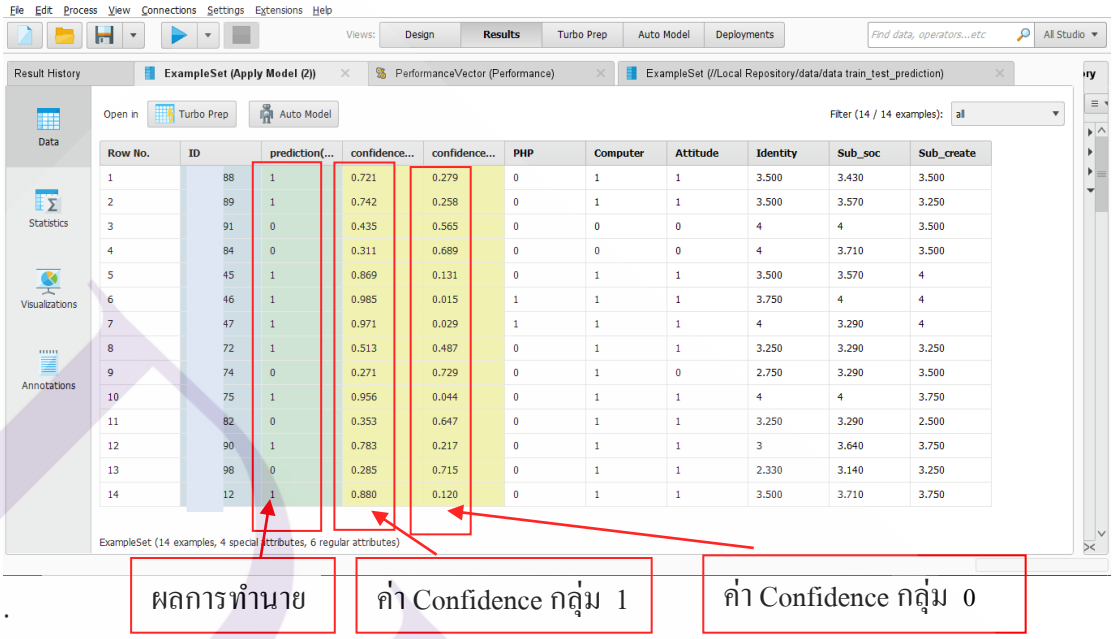
จากภาพที่ 4.6 การปรับค่าพารามิเตอร์ของโมเดล Neural Network ได้แก่ Training Cycle = 200, Hidden layer = 1, Hidden layer sizes = 22, Learning rate = 0.4, Momentum = 0.3

4.5.3.4 นำเข้าข้อมูลใหม่ที่ต้องการทำนายผลการเรียน เข้าโมเดลและปรับโมเดล ดังภาพที่ 4.7



ภาพที่ 4.7 การนำข้อมูลใหม่เข้าโมเดลและการปรับโมเดลเพื่อการทำนายผลการเรียน

4.5.3.5 Run โมเดลและตรวจสอบผลการทำนาย ดังภาพที่ 4.8



.

ภาพที่ 4.8 ผลการทำนาย และ ค่า Confidence การทำนายด้วยโมเดล Neural Network

จากภาพที่ 4.8 ผลจากการใช้โมเดลการทำนายผลการเรียนด้วยวิธี Neural Network จะแสดงผลการทำนายเป็นรายบุคคลใน Attribute ชื่อ prediction ผลทำนาย = 0 หมายถึง นิสิตรายนี้อยู่ในกลุ่มเกรดพอใช้ ผลทำนาย = 1 หมายถึง นิสิตรายนี้อยู่ในกลุ่มเกรดดี

ค่า Confidence รายบุคคล จะแบ่งเป็นค่า Confidence เกรดพอใช้ และค่า Confidence เกรดดี โมเดลจะทำนายว่านิสิตอยู่กลุ่มใด ดูจากค่า Confidence ที่มากกว่า

บทที่ 5
สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

การวิจัยเรื่อง การประยุกต์ใช้ Machine Learning ทำนายผลการเรียนวิชา Web Database ของนิสิตสาขาเทคโนโลยีการศึกษา คณะศึกษาศาตร์ มหาวิทยาลัยบูรพา มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของโมเดลการทำนายผลการเรียนวิชา Web database ระหว่างวิธี Decision Tree, Naïve Bayes, Neural Network และ Support Vector Machine สามารถสรุปผลวิจัยได้ดังนี้

5.1 สรุปผลวิจัย

5.1.1 ขั้นตอนแรกของการวิจัย ผู้วิจัยได้ศึกษาเอกสารงานวิจัยที่เกี่ยวข้องกับคุณลักษณะที่ส่งผลต่อผลการเรียนวิชาด้านการเขียนโปรแกรมคอมพิวเตอร์ และได้สังเคราะห์ Attribute ที่เกี่ยวข้อง จำนวน 17 Attribute ได้แก่ 1) เพศ 2) ประเภทการรับเข้ามหาวิทยาลัย 3) แผนการเรียนมัธยมปลาย 4) เกรดเฉลี่ยจบการศึกษาระดับชั้น ม. 6 5) เกรดเฉลี่ยวิชาคณิตศาสตร์ ม.6 6) เคยเรียนภาษา HTML 7) เคยเรียนภาษา C หรือ C++ 8) เคยเรียนภาษา PHP 9) คอมพิวเตอร์ส่วนตัว 10) ความรู้สึกต่อการเรียนวิชาด้านคอมพิวเตอร์ 11) การทบทวนบทเรียน 12) เกรดเฉลี่ยกลุ่มวิชาภาษาอังกฤษ 13) เกรดเฉลี่ยกลุ่มวิชาภาษาอื่น ๆ 14) เกรดเฉลี่ยกลุ่มวิชาอัตลักษณ์มหาวิทยาลัย 15) เกรดเฉลี่ยกลุ่มวิชาสังคม 16) เกรดเฉลี่ยกลุ่มวิชาความสร้างสรรค์ 17) เกรดเฉลี่ยกลุ่มวิชาเทคโนโลยีสารสนเทศ

5.1.2 ผู้วิจัยได้ดำเนินการวิเคราะห์ความสัมพันธ์ระหว่าง (Attribute) กับ (Label) คือ ผลการเรียนวิชา Web Database แบ่งเป็น 2 ค่า คือ กลุ่มเกรดพอใช้ และกลุ่มเกรดดี โดยใช้วิธีการวิเคราะห์ความถดถอยพหุคูณ (Multiple Regression) แบบการคัดเลือกแบบลำดับขั้น (Stepwise Selection) ด้วยโปรแกรม SPSS for Window พบว่า Attribute ที่มีอิทธิพลต่อ Label ประกอบด้วยข้อมูล 6 Attribute ได้แก่ 1) Sub_soc (เกรดเฉลี่ยกลุ่มวิชาสังคม) 2)PHP (เคยเรียนภาษา PHP) 3) Computer (คอมพิวเตอร์ส่วนตัว) 4) Sub_create (เกรดเฉลี่ยกลุ่มวิชาความสร้างสรรค์) 5) Identity (เกรดเฉลี่ยกลุ่มวิชาอัตลักษณ์มหาวิทยาลัย) และ 6) Attitude (ความรู้สึกต่อการเรียนวิชาด้านคอมพิวเตอร์)

5.1.3 ผลการวิจัย การเปรียบเทียบค่าความถูกต้อง (Accuracy) ของการใช้ข้อมูลชุดที่ 1 และข้อมูลชุดที่ 2 ทำนายผลการเรียน พบว่า โมเดลทำนายโดยใช้ข้อมูลชุดที่ 2 ที่มีข้อมูล 6 Attribute มีค่าความถูกต้อง (Accuracy) เพิ่มขึ้นทุกโมเดล

5.1.4 ผลการวิจัยการนำชุดข้อมูลที่ 2 จำนวน 6 Attribute ที่ได้จากการคัดกรองด้วยการวิเคราะห์ความถดถอยพหุคูณ (Multiple Regression) แบบการคัดเลือกลำดับขั้น(Stepwise Selection) มาสร้างโมเดลเพื่อทำนายผลการเรียนวิชา Web Database พบว่า ประสิทธิภาพทำนายวิธี Neural Network มีค่าความถูกต้อง (Accuracy) มากที่สุด 88.89 %

5.1.5 ผลการปรับค่าพารามิเตอร์ของโมเดล Neural Network ที่เหมาะสมที่สุด คือ Training Cycle = 200, Hidden layer = 1, Hidden layer sizes = 22, Learning rate = 0.4 และ Momentum = 0.3 โดยค่าพารามิเตอร์ที่ปรับใหม่แตกต่างจากค่าพารามิเตอร์ที่ผู้วิจัยกำหนดขึ้นในเบื้องต้น คือ จำนวน Hidden layer ลดลงจาก 3 Hidden layer เหลือ 1 Hidden layer ซึ่งส่งผลให้โมเดล Neural Network มีประสิทธิภาพการทำนายเพิ่มขึ้น จากค่าความถูกต้อง (Accuracy) 88.89 % เป็น 92.59 % (เพิ่มขึ้น 3.7 %)

5.2 อภิปรายผล

ผู้วิจัยนำเสนอการอภิปรายผลการวิจัยใน 2 ประเด็นดังนี้

ในงานวิจัยครั้งนี้พบว่า การทำนายผลการเรียนวิชา วิชา Web Database สามารถทำนายผลโดยใช้ข้อมูล 6 Attribute ได้แก่ 1) ประสบการณ์เคยเรียนภาษา PHP มาก่อน ซึ่งสอดคล้องกับเนื้อหาการเรียนวิชา Web Database ที่มีการเรียนเขียนโปรแกรมเพื่อเชื่อมฐานข้อมูลบนเวปไซต์ ดังนั้นถ้านิสิตเคยเรียน PHP มีโอกาสที่ผลการเรียนจะอยู่ในระดับดี แต่ถ้านิสิตไม่เคยเรียนภาษา PHP มาก่อน มีโอกาสที่ผลการเรียนจะอยู่ในระดับพอใช้ 2) เกรดเฉลี่ยกลุ่มวิชาสังคม 3) เกรดเฉลี่ยกลุ่มวิชาความสร้างสรรค์ 4) เกรดเฉลี่ยกลุ่มวิชาอัตลักษณ์มหาวิทยาลัย โดยผลการเรียนทั้ง 3 กลุ่มวิชาสามารถสะท้อนความรู้ ความสามารถ และความรับผิดชอบ ในการเรียนรู้ของนิสิต โดยพบว่า นิสิตที่มีผลการเรียนดีทั้งสามกลุ่มวิชา มีโอกาสที่ผลการเรียนวิชา Web Database จะอยู่ในระดับดี แต่ถ้านิสิตที่มีผลการเรียนระดับต่ำทั้งสามกลุ่มวิชา มีโอกาสที่ผลการเรียนวิชา Web Database จะอยู่ในระดับพอใช้เช่นเดียวกัน 5) การมีคอมพิวเตอร์ส่วนตัว การที่นิสิตมีหรือไม่มีคอมพิวเตอร์ส่วนตัวอาจเป็นปัจจัยที่มีความสำคัญต่อผลการเรียนด้านคอมพิวเตอร์ เนื่องจากการเรียนในห้องปฏิบัติการโดยใช้คอมพิวเตอร์ของสถาบันอาจไม่สามารถเรียนหรือส่งงานได้ตามเวลา เพราะการเรียนด้านการเขียนโปรแกรมคอมพิวเตอร์ผู้เรียนต้องทบทวนและฝึกเขียนด้วยตนเองอย่างสม่ำเสมอ และต้องขยายความรู้จากในห้องเรียน ดังนั้นถ้านิสิตที่มีคอมพิวเตอร์ส่วนตัว มีโอกาสที่

ผลการเรียนวิชา Web Database จะอยู่ในระดับดี แต่ถ้านิสิตไม่มีคอมพิวเตอร์ส่วนตัว มีโอกาสที่ผลการเรียนวิชา Web Database จะอยู่ในระดับพอใช้ 6) ความรู้สึกต่อการเรียนวิชาด้านคอมพิวเตอร์ การที่นิสิตมีความรู้สึกดี หรือมีความพึงพอใจต่อการเรียนวิชาใด ๆ ย่อมทำให้เกิดแรงจูงใจ ความสนใจ ความรับผิดชอบ ความพยายามในการเรียนวิชานั้น ซึ่งความพยายามที่มาจากความรู้สึกดีหรือพึงพอใจจะเป็นความพยายามที่มั่นคง และอดทน ผลความรู้สึกดีหรือพึงพอใจต่อการเรียนวิชาด้านคอมพิวเตอร์เช่นเดียวกัน อาจทำให้นิสิตมีผลการเรียนวิชา Web Database อยู่ในระดับดี แต่ถ้านิสิตมีความรู้สึกไม่ดีหรือไม่พึงพอใจต่อการเรียนวิชาด้านคอมพิวเตอร์ อาจทำให้นิสิตมีผลการเรียนวิชา Web Database อยู่ในระดับพอใช้ ซึ่งจะส่งผลต่อการเรียนวิชาด้านคอมพิวเตอร์อื่น ๆ ต่อไป

ในงานวิจัยครั้งนี้พบว่า ประสิทธิภาพของโมเดลการทำนายผลการเรียนวิชา Web database โดยใช้ข้อมูล 6 Attribute เหมือนกันพบว่า วิธี Neural Network มีค่าความถูกต้อง (Accuracy) มากที่สุด ทั้งนี้เนื่องจากการเตรียมข้อมูลมีความเหมาะสมต่อข้อมูลนำเข้าโมเดลทั้ง 6 Attribute นอกจากนั้นผู้วิจัยได้แปลงรูปข้อมูล และกำหนด Key code ข้อมูลให้สอดคล้องกับการรับข้อมูลเข้าโมเดลเพื่อการเรียนรู้ (Train) และทดสอบ (Test) ทั้ง 4 วิธี คือ Decision Tree, Naïve Bayes, Neural Network และ Support Vector Machine แต่การแปลงรูปข้อมูลในงานวิจัยในครั้งนี้อาจสอดคล้องกับกับวิธี Neural Network มากที่สุด เนื่องจากวิธี Neural Network เป็นแบบจำลองสามารถเรียนรู้ตัวแปรป้อนเข้า (Input) ได้ดีไม่ว่าตัวแปรนั้นจะเป็นตัวแปรอันตรภาค (Interval Variables) ตัวแปรนามบัญญัติ (Norminal Variables) หรือตัวแปรแบบทวิลักษณ์ หรือ ตัวแปรฐานสอง (Binary Variables) โดยที่ตัวแปรป้อนเข้า (Input) ไม่มีการกระจายตัวแบบ Normal สำหรับตัวแปรอันตรภาค (Interval Variables) และความสัมพันธ์ระหว่างตัวแปรป้อนเข้า (Input) และตัวแปรเป้าหมายไม่เป็นเชิงเส้นตรง [25]

5.3 ข้อเสนอแนะ

5.3.1 ข้อเสนอแนะจากการทำวิจัย

จากงานวิจัยพบว่าพบว่าวิธีการ Neural Network มีค่าความถูกต้อง (Accuracy) สูงที่สุด ดังนั้นผู้สอนวิชาการเขียนโปรแกรมคอมพิวเตอร์ ควรนำวิธีการ Neural Network ไปใช้ในทำนายผลการเรียนวิชาด้านการเขียนโปรแกรมคอมพิวเตอร์ โดยปรับ Attribute ให้สอดคล้องกับรายวิชาที่สอน ผลการทำนายจะเป็นข้อมูลสำหรับวางแผนการจัดเรียนการสอนให้มีคุณภาพและมีประสิทธิภาพต่อไป

5.3.2 ข้อเสนอการทำวิจัยครั้งต่อไป

5.3.2.1 งานวิจัยนี้กำหนดแอทริบิวข้อมูลเป็นบางส่วนเท่านั้น การพัฒนางานวิจัยครั้งต่อไปขั้นตอนแรกควรเพิ่ม Attribute ที่เกี่ยวข้องกับผู้เรียนให้มากขึ้น และใช้วิธีการวิเคราะห์หาความสัมพันธ์ระหว่างตัวแปร โดยใช้วิธีการวิเคราะห์ความถดถอยพหุคูณ (Multiple Regression) เพื่อหา Attribute ที่ส่งผลต่อค่าความถูกต้อง (Accuracy) ให้สูงขึ้น

5.3.2.2 ควรมีการวิจัยและพัฒนาโปรแกรมเพื่อการทำนายผลการเรียนวิชาอื่นๆ โดยใช้กฎความสัมพันธ์ที่ได้จากการทำเหมืองข้อมูล

5.3.2.2 ควรมีการวิจัยและพัฒนาโปรแกรมเพื่อการทำนายผลการเรียน หรือจำแนกข้อมูลอื่นๆ โดยประยุกต์ ระบบ AI ระบบ IOT เป็นเครื่องมือเก็บข้อมูลแต่ละ Attribute และประมวลผล Real time บน Cloud

บรรณานุกรม

บรรณานุกรม

ภาษาไทย

จิตตภู พูลวัน. (2550). *ระบบวิเคราะห์โรคทั่วไปโดยใช้การจำแนกข้อมูลบนอินเทอร์เน็ต.* ภาควิชาเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ.

จงสวัสดิ์ จงวัฒน์ผล. (2561). *การวิเคราะห์ข้อมูลทางธุรกิจสาหรับการตัดสินใจเชิงกลยุทธ์.* กรุงเทพมหานคร: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

ฉัตรศิริ ปิยะพิมลสิทธิ์. (2548). *การใช้ SPSS เพื่อการวิเคราะห์ข้อมูล.* กรุงเทพฯ : มหาวิทยาลัยเกษตรศาสตร์.

ฉัตรศิริ ปิยะมิพลสิธิ์. (2548). *การใช้ SPSS ในการวิเคราะห์ข้อมูล.* ภาควิชาการประเมินผลและการวิจัย. คณะศึกษาศาสตร์. สงขลา: มหาวิทยาลัยทักษิณ.

ชัชชฎา วันดี, จิรัฏฐา ภูบุญอบ และ ฉัตรเกล้า เจริญผล. (2556). ประสิทธิภาพการจำแนกข้อมูลการเลือกอาชีพโดยอัตโนมัติด้วยเทคนิคเหมืองข้อมูล. *วารสารวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยมหาสารคาม ฉบับพิเศษ การประชุมทางวิชาการมหาวิทยาลัยมหาสารคาม* ครั้งที่ 9 : 263-270.

ชัยวัฒน์ ฉวีวรรณ. (2560) *การพัฒนากรอบงานความสามารถด้านการเขียนโปรแกรมจากความถนัดและทักษะ*. วิทยานิพนธ์ วิทยาศาสตรมหาบัณฑิต (วท.ม.), จุฬาลงกรณ์มหาวิทยาลัย.

ชนะวงศ์ คงสอน และณัฐวุฒิ วิสุทธิพิเนตร. (2559). *การพัฒนาระบบสนับสนุนการตัดสินใจในการเลือกสาขาการเรียนต่อระดับประกาศนียบัตรวิชาชีพ สังกัดอาชีวศึกษาจังหวัดพระนครศรีอยุธยา.* ปริญญานิพนธ์วิทยาศาสตร์บัณฑิต. สาขาเทคโนโลยีสารสนเทศ, มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ ศูนย์หันตรา.

ธนาวุฒิ ประกอบผล. (2552). โครงข่ายประสาทเทียม = Artificial neural networks. *วารสาร มฉก. วิชาการ. ปีที่ 12 ฉบับที่ 24 (ม.ค.-มิ.ย. 2552) หน้า 72-87.*

ประยูรศรี บุตรแสนคม. (2555). การคัดเลือกตัวแปรพยากรณ์เข้าในสมการถดถอยพหุคูณ. *วารสารการวัดผลการศึกษา ปีที่ 17 ฉบับที่ 1 กรกฎาคม 2555 มหาวิทยาลัยมหาสารคาม.*

บรรณานุกรม (ต่อ)

ภาษาไทย

ประเทือง วงษ์ทอง, วาฤทธิ์ กันแก้ว อภิชชญา ขวัญแก้ว, และ ยุวดี โฉมแดง. (2563). *การประยุกต์เทคนิคเหมืองข้อมูลค้นหาลักษณะนิสัย ของผู้ประกอบอาชีพด้านคอมพิวเตอร์.* มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ.

ปริญญา น้อยดอนไพร. (2556). *การเขียนโปรแกรมบนเว็บด้วย PHP ร่วมกับฐานข้อมูล MySQL.* มหาวิทยาลัยราชภัฏสุราษฎร์ธานี: คณะวิทยาศาสตร์และเทคโนโลยี.

พยูน พาณิชย์กุล. (2548*). การพัฒนาระบบดาต้าไมน์นิ่งโดยใช้ Decision Tree" โครงการพัฒนาระบบงาน.* ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศแขนงวิทยาการสารสนเทศสถาบันเทคโนโลยีพระจอมเกล้าคุณทหารลาดกระบัง.

พรพล ธรรมรงค์รัตน์.(2552). *การจำแนกประเภทเว็บเพจโดยวิธีการลดขนาดลักษณะเฉพาะและซัพพอร์ตเวกเตอร์แมชชีน.* มหาวิทยาลัยสงขลานครินทร์, สงขลา.

วรุณย์พันธุ์ สุขสมมโน. (2542). *การปรับจูนตัวเองของตัวควบคุมพีไอดีโดยใช้ข่ายงานระบบประสาท* . จุฬาลงกรณ์มหาวิทยาลัย, กรุงเทพฯ.

วุฒิชัย กำจรกิตติคุณ และ ชัยพร เขมะภาตะพันธ์. (2559. การวิเคราะห์ความผิดปกติข้อมูลการใช้ไฟฟ้าด้วยนาอีฟเบย์. *การประชุมวิชาการระดับประเทศทางด้านเทคโนโลยีสารสนเทศครั้งที่ 8,* สืบค้นจาก https://cite.dpu.ac.th/research.html

วุฒิชัย กำจรกิตติคุณ. (2561). *การจำแนกประเภทผู้ใช้ไฟฟ้าด้วยนาอีฟเบย์และนิวรัลเน็ทเวิร์ก* วิทยานิพนธ์ปริญญาโท สาขาวิชาวิศวกรรมคอมพิวเตอร์และโทรคมนาคม. วิทยาลัยนวัตกรรมด้านเทคโนโลยีและวิศวกรรมศาสตร์ มหาวิทยาลัยธุรกิจบัณฑิตย์.

ศุภชัย ประคองศิลป์. (2551*). การออกแบบและพัฒนาระบบสนับสนุนการตัดสินใจในการอนุมัติลูกบ้านเข้าโครงการโดยใช้เทคนิคต้นไม้ตัดสินใจ กรณีศึกษา มูลนิธิที่อยู่อาศัยเพื่อมนุษยชาติ.* ปัญหาพิเศษปริญญาวิทยาศาสตร์มหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศคณะเทคโนโลยีสารสนเทศ บัณฑิตวิทยาลัย สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ.

สายชล สินสมบูรณ์ทอง. (2560). *การทำเหมืองข้อมูลเล่ม 1 การค้นหาความรู้จากข้อมูล Data mining 1 : Discovering Knowledge in Data.* พิมพ์ครั้งที่ 2. กรุงเทพฯ: จามจุรีโปรดักส์.

บรรณานุกรม (ต่อ)

ภาษาไทย

สุทิน ชนะบุญ. (2560). *สถิติและการวิเคราะห์ข้อมูลในงานวิจัยด้านสุขภาพเบื้องต้น. ขอนแก่น.* สำนักงานสาธารณสุขจังหวัดขอนแก่น.

สุธีรา วงศ์อนันทรัพย์ , ต้องใจ แย้มผกา และ อรวรรณ มูสิกะ. (2558). การใช้เทคนิคเหมืองข้อมูลในการประเมินความรู้ และหาความถนัดเพื่อพัฒนาศักยภาพของนักศึกษา.*วารสารสังคมศาสตร์ ปีที่ 5 ฉบับที่ 1 มกราคม - มิถุนายน 2559.*

อาจณรงค์ มโนสุทธิฤทธิ์. (2556). *การพัฒนาระบบการสอนเพื่อพัฒนาความสามารถในการเขียนโปรแกรมคอมพิวเตอร์สำหรับนักเรียนระดับชั้นมัธยมศึกษาตอนต้น* . วิทยานิพนธ์ปริญญาเอก สาขาเทคโนโลยีการศึกษา. คณะศึกษาศาสตร์ มหาวิทยาลัยบูรพา.

เอกสิทธิ์ พัชรวงศ์ศักดา. (2557). *Introduction to Data Mining with Rapid Miner Studio 7.* กรุงเทพฯ: ห้างหุ้นส่วนสามัญ ดาต้าคิวบ์.

ใกล้รุ่ง เฮงพระพรหม , สุพจน์ เฮงพระพรหม, และ สุวิมล มรรควิบูลย์ชัย. (2557). การค้นหาความรู้คุณลักษณะสำคัญของนักศึกษาที่มีต่อผลสัมฤทธิ์ทางการเรียนในกลุ่มโปรแกรมคอมพิวเตอร์ มหาวิทยาลัยราชภัฏนครปฐม ด้วยเทคนิคเหมืองข้อมูล. *วารสารวิชาการราชภัฏตะวันตก, ปีที่ 9, ฉบับที่ 1, กรกฎาคม – ธันวาคม 2557, หน้า 71 – 80.*

ภาษาต่างประเทศ

Edin, O., Mirza, S. (2012) : *Data Mining Approach for Predicting Student Performance*, Economic Review: Journal of Economics and Business, ISSN 1512-8962, University of Tuzla, Faculty of Economics, Tuzla, Vol. 10, Iss. 1, pp. 3-12

Erdogan, Y., Aydin, E., & Kabaca, T. (2007). *Identifying Predictors of Programming Achievement*. in Conf. 6th WSEAS International Conference on Education and Educational Technology, Venice, Italy, Nov. 2007, pp. 600-115

บรรณานุกรม (ต่อ)

ภาษาต่างประเทศ

Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P.,& Uthurusamy, R. (1996). *Advances in Knowledge Discovery and Data Mining*. Menlo Park, Calif.: AAAI Press.

Garcia, P., Amandi A., SchiaffinoS. and Campo M. (2007). *Evaluating Bayesiannetworks' precisionfor detectionstudents' learningstyles.* Computer & Education:794-808.

Graham, P. (2003). "*Hackers and Painters"* สืบค้นจาก http://www.paulgraham.com/hp.html

Han, J. & Kamber, M. (2001). *Data Mining: Concepts and Techniques*. CA: Morgan Kaufmann, San Francisco.

Martinez, D. & Y. Li. (2011). *Information extraction from pathology reports in a hospital setting*, in Proceedings of the 20th ACM international conference on Information and knowledge management. 2011, ACM: Glasgow, Scotland, UK. p.1877-1882.

Mitchell, T. (1997). *Machine Learning*, McGraw Hill.

Pacharawongsakda, E. (2018). *Naïve Bayes introduction.* [cited 2018 Dec 20]; Available from: http://dataminingtrend.com/2014/naive-bayes/

Rapid Miner Studio Core. (2021). *Performance Classification.* สืบค้นจาก https://docs.rapidminer.com/latest/studio/operators/validation/performance/predictive/performance_classification.html

Shahiria, A. M., Husaina,W.,& Rashida, N. A., (2015). "*A Review on Predicting Student's Performance using Data Mining Techniques,"* in The Third Information Systems International Conference, 2015.

Shalev-Shwartz, S., & Ben-David, S. (2009). *Understanding Machine Learning*.doi: 10.1017/cbo9781107298019.

Trippi, R. & E. Turban (eds).(1992). *Neural Networks in Financing and Investing:Applying Artificial Intelligence to Improve Real-World Performance*, Chicago: Probus Publishing Co.

บรรณานุกรม (ต่อ)

ภาษาต่างประเทศ

Wilson, B. C., & Shrock. S. (2001). *Contributing to Success in an Introductory Computer Science Course: A Study of Twelve Factors*, in Proceedings of thirty-second SIGCSE technical symposium on Computer Science Education, New York, 2001, pp. 184-18.

Wilson, B. C. (2002). A Study of Factors Promoting Success in Computer Science Including Gender Differences, *Computer Science Education, vol. 12, no. 1-2, pp. 141- 164, 2002.*

Wikipedia.(2021). *Confusion matrix* . สืบค้นจาก https://en.wikipedia.org/wiki/Confusion_matrix

Witten, I. H.,& Frank, E. (2000) *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations.* Morgan Kaufmann Publishers, 2000.

ประวัติผู้เขียน

ชื่อ – นามสกุล	นายวีระพันธ์ พานิชย์
ประวัติการศึกษา	ครุศาสตรบัณฑิต สาขาเทคโนโลยีและนวัตกรรมการศึกษา สถาบันราชภัฏจันทรเกษม พ.ศ. 2539 การศึกษามหาบัณฑิต สาขาเทคโนโลยีทาง การศึกษา มหาวิทยาลัยบูรพา พ.ศ. 2549 ปรัชญาดุษฎีบัณฑิต สาขาเทคโนโลยีการศึกษา มหาวิทยาลัยบูรพา พ.ศ. 2554 วิศวกรรมศาสตรมหาบัณฑิต สาขาวิศวกรรมคอมพิวเตอร์และโทรคมนาคม มหาวิทยาลัยธุรกิจบัณฑิตย์ พ.ศ. 2564
ตำแหน่งงานปัจจุบัน	อาจารย์ประจำภาควิชานวัตกรรมและเทคโนโลยีการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยบูรพา