การพยากรณ์และความสัมพันธ์ของการยอมรับการใช้เทคโนโลยีที่มีผลต่อพฤติกรรมการใช้แอปพลิเคชัน LAZADA ช้อปปิ้งบนสมาร์ทโฟนของประชากรในเขตกรุงเทพมหานครและปริมณฑล โดยใช้เทคนิคการทำเหมืองข้อมูล

โดย

วรพจน์ ลิลิตวัฒน์

มหาวิทยาลัยรังสิต Rangsit University

วิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตาม

หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ

วิทยาลัยนวัตกรรมดิจิทัลเทคโนโลยี

บัณฑิตวิทยาลัย มหาวิทยาลัยรังสิต

ปีการศึกษา 2565

PREDICTION AND ASSOCIATION RULE OF TECHNOLOGY ADOPTION AFFECTING LAZADA SHOPPING APPLICATION USAGE BEHAVIOR ON SMARTPHONES AMONG PEOPLE IN BANGKOK AND ITS VICINITIES BY USING DATA MINING TECHNIQUES

BY

WORAPHOT LILITWAT

มหาวิทยาลัยรังสิต Rangsit University

A THESIS SUBMITED IN PARTIAL FULFILLMENT

OF THE REQUIREMENTS FOR

THE DEGREE OF MASTER OF SCIENCE

IN INFORMATION TECHNOLOGY

COLLEGE OF DIGITAL INNOVATION TECHNOLOGY

GRADUATE SCHOOL, RANGSIT UNIVERSITY

ACADEMIC YEAR 2022

วิทยานิพนธ์เรื่อง

การพยากรณ์และความสัมพันธ์ของการยอมรับการใช้เทคโนโลยีที่มีผลต่อพฤติกรรมการใช้แอปพลิเคชัน LAZADA ช้อปปิ้งบนสมาร์ทโฟนของประชากรในเขตกรุงเทพมหานครและปริมณฑล โดยใช้เทคนิคการทำเหมืองข้อมูล

โดย
วรพจน์ ลิลิตวัฒน์

ได้รับการพิจารณาให้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ

มหาวิทยาลัยรังสิต
ปีการศึกษา 2565

รศ.ดร.ปริญญา สงวนสัตย์
ประธานกรรมการสอบ

รศ.ดร.เชฎฐเนติ ศรีสอ้าน
กรรมการ

ผศ.ดร.สมชาย เล็กเจริญ
กรรมการและอาจารย์ที่ปรึกษา

บัณฑิตวิทยาลัยรับรองแล้ว

(ผศ.ร.ต. หญิง ดร.วรรณี ศุขสาตร)
คณบดีบัณฑิตวิทยาลัย
20 ธันวาคม 2565

มหาวิทยาลัยรังสิต Rangsit University

Thesis entitled

PREDICTION AND ASSOCIATION RULE OF TECHNOLOGY ADOPTION AFFECTING LAZADA SHOPPING APPLICATION USAGE BEHAVIOR ON SMARTPHONES AMONG PEOPLE IN BANGKOK AND ITS VICINITIES BY USING DATA MINING TECHNIQUES

by

WORAPHOT LILITWAT

was submitted in partial fulfillment of the requirements
for the degree of Master of Science in Information Technology

Rangsit University
Academic Year 2022

Assoc.Prof. Parinya Sanguansat, Ph.D.
Examination Committee Chairperson

Assoc.Prof. Chetneti Srisa-an, Ph.D.
Member

Asst.Prof. Somchai Lekcharoen, Ph.D.
Member and Advisor

Approved by Graduate School

(Asst.Prof.Plt.Off. Vannee Sooksatra, D.Eng.)
Dean of Graduate School
December 20, 2022

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จสมบูรณ์ได้ด้วยดี ผู้วิจัยขอกราบขอบพระคุณ บุคคล และกลุ่มบุคคลต่าง ๆ ที่ได้กรุณาให้คำปรึกษา แนะนำ ช่วยเหลือ เป็นอย่างดี ทั้งในด้านวิชาการ และการดำเนินงานวิจัย ที่ช่วยอำนวยความสะดวกในการทำงานวิจัยนี้ อันได้แก่ ผศ.ดร.สมชาย เล็กเจริญ อาจารย์ที่ปรึกษาที่คอยช่วยเหลือด้านวิชาการตั้งแต่พื้นฐาน และการดำเนินงานวิจัย รวมถึงด้านอื่น ๆ อีกมากมาย อีกทั้งยังคอยช่วยตรวจทานวิทยานิพนธ์ฉบับนี้ให้มีความสมบูรณ์ถูกต้อง

ขอขอบคุณ พี่ น้องและเพื่อน รวมถึงมิตรสหายที่คอยให้คำปรึกษา คอยให้กำลังใจและมิตรภาพที่ดีมาโดยตลอด

ขอขอบพระคุณมหาวิทยาลัยรังสิตที่ให้ทุนสนับสนุนการศึกษาต่อในระดับปริญญาโท และขอขอบพระคุณครูบาอาจารย์ทุกท่านตั้งแต่อดีตจนถึงปัจจุบันที่ประสิทธ์ประสาทวิชาการความรู้ต่าง ๆ กราบขอบพระคุณ คุณพ่อวันชัย ลิลิตวัฒน์ และคุณแม่เสาวณีย์ ลิลิตวัฒน์ ที่เลี้ยงดู และให้ความอบอุ่นอยู่เคียงข้างคอยให้กำลังใจผู้วิจัยเสมอมา ขอขอบคุณ คุณอาริสา มินหะรีสุไรมาน ครอบครัวมินหะรีสุไรมาน และ ครอบครัว ลิลิตวัฒน์ ที่ได้มอบความรัก ความห่วงใย และคอยเป็นกำลังใจให้ผู้วิจัยโดยตลอด

วรพจน์ ลิลิตวัฒน์
ผู้วิจัย

มหาวิทยาลัยรังสิต Rangsit University

5908039 : วรพจน์ ลิลิตวัฒน์

ชื่อวิทยานิพนธ์ : การพยากรณ์และความสัมพันธ์ของการยอมรับการใช้เทคโนโลยีที่มีผลต่อพฤติกรรมการใช้แอปพลิเคชัน LAZADA ช้อปปิ้งบนสมาร์ทโฟนของประชากรในเขตกรุงเทพมหานครและปริมณฑล โดยใช้เทคนิคการทำเหมืองข้อมูล

หลักสูตร : วิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ

อาจารย์ที่ปรึกษา : ผศ.ดร.สมชาย เล็กเจริญ

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบแบบจำลองการพยากรณ์ระหว่างต้นไม้การตัดสินใจ (Decision Tree) โครงข่ายประสาทเทียม (Neural Network) และไนอีฟเบย์ (Naïve Bayes) และเพื่อหาความสัมพันธ์ของข้อคำถามการยอมรับการใช้เทคโนโลยีที่มีผลต่อพฤติกรรมการใช้แอปพลิชั่น LAZADA ช้อปปิ้งบนสมาร์ทโฟนของประชากรในเขตกรุงเทพมหานครและปริมณฑล โดยใช้เทคนิคการทำเหมืองข้อมูล และข้อมูลที่ใช้ในการวิจัยครั้งนี้ได้จากผู้ตอบแบบสอบถามจำนวนทั้งสิ้น 386 คน ได้มาโดยการสุ่มตัวอย่างแบบง่าย (Simple Random Sampling) จากผู้ที่มีประสบการณ์การใช้แอปพลิเคชัน LAZADA ช้อปปิ้งบนสมาร์ทโฟน การวิจัยครั้งนี้ใช้โปรแกรม Rapid Miner Studio เป็นเครื่องมือช่วยดำเนินการวิจัย สำหรับการพยากรณ์ และ การหาความสัมพันธ์ ของการยอมรับการใช้เทคโนโลยีที่มี ผลต่อพฤติกรรม การใช้แอปพลิเคชัน LAZADAช้อปปิ้งบนสมาร์ทโฟนของประชากรในเขตกรุงเทพมหานครและปริมณฑลและมีการเปรียบเทียบแบบจำลองการพยากรณ์ โดยใช้สถิติ ONE-WAY ANOVA การวิจัยครั้งนี้พบว่าแบบจำลองทั้ง 3 วิธี มีค่าสถิติไม่แตกต่างกันอย่างมีนัยสำคัญที่ระดับ 0.05 ดังนั้นผู้วิจัยจึงเลือกใช้ เทคนิคการจำแนกข้อมูลด้วยวิธีต้นไม้การตัดสินใจ เป็นแบบจำลองการพยากรณ์ เพื่อสร้างโมเดลการจำแนกข้อมูล นอกจากนี้ได้นำข้อมูลดังกล่าวไปวิเคราะห์เพื่อหาความสัมพันธ์ของชุดข้อมูลมาประกอบการตัดสินใจแนวทางการพัฒนาแอปพลิเคชันช้อปปิ้งบนสมาร์ทโฟนเพื่อตอบสนองความต้องการของผู้บริโภคให้เกิดประสิทธิภาพมากที่สุด

(วิทยานิพนธ์มีจำนวนทั้งสิ้น 95 หน้า)

คำสำคัญ: เหมืองข้อมูล, ต้นไม้การตัดสินใจ, เครือข่ายประสาทเทียม, ไนอีฟเบย์

ลายมือชื่อนักศึกษา ..... ลายมือชื่ออาจารย์ที่ปรึกษา .....

5908039 : Woraphot Lilitwat

Thesis Title : Prediction and Association Rule of Technology Adoption Affecting Lazada Shopping Application Usage Behavior on Smartphones Among People in Bangkok and Its Vicinities by Using Data Mining Techniques

Program : Master of Science in Information Technology

Thesis Advisor : Asst.Prof. Somchai Lekcharoen, Ph.D.

Abstract

The purposes of this research were to compare predicting models among decision tree, neural network, and naïve bayes, and to determine the relationship between technology adoption and behavior of using the LAZADA shopping application on smartphones. We included subjects in Bangkok and its vicinities who had experience using the LAZADA application. By simple random sampling method, 386 respondents were recruited and analyzed by Rapid Miner Program Studio as a research predictive tool. By using One-Way analysis of variance (ANOVA), we found that all three models had no significant statistical difference at value > 0.05 Thus, the decision tree method was adopted as a predicting model. The data set was also analyzed how to develop a shopping application on a smartphone to meet the needs of consumers more effectively.

(Total 95 pages)

Keywords: Data Mining, Decision Tree, Neural Network, Naive Bayes

มหาวิทยาลัยรังสิต Rangsit University

Student's Signature …..... Thesis Advisor's Signature .....

สารบัญ

		หน้า
กิตติกรรมประกาศ		**ก**
บทคัดย่อภาษาไทย		**ข**
บทคัดย่อภาษาอังกฤษ		**ค**
สารบัญ		**ง**
สารบัญตาราง		**ฉ**
สารบัญรูป		**ช**
บทที่ 1	**บทนำ**	**1**
	1.1 ความเป็นมาและความสำคัญของปัญหา	1
	1.2 วัตถุประสงค์ของการวิจัย	2
	1.3 ขอบเขตของการวิจัย	2
	1.4 ประโยชน์ที่คาดว่าจะได้รับ	3
บทที่ 2	**ทบทวนวรรณกรรมและทฤษฎีที่เกี่ยวข้อง**	**4**
	2.1 ทฤษฎีที่เกี่ยวข้อง	4
	2.2 งานวิจัยที่เกี่ยวข้อง	34
บทที่ 3	**วิธีดำเนินการวิจัย**	**36**
	3.1 ข้อมูลกลุ่มตัวอย่าง	36
	3.2 เทคนิคการทำเหมืองข้อมูล (Data Mining)	42
บทที่ 4	**ผลการวิจัย**	**46**
	4.1 ชุดข้อมูล	46
	4.2 การทดลองและการวัดผล	46
	4.3 ผลลัพธ์และการพิจารณา	87

มหาวิทยาลัยรังสิต Rangsit University

สารบัญ (ต่อ)

		หน้า
บทที่ 5	**สรุปผล**	**88**
	5.1 สรุปผล	88
	5.2 ข้อเสนอแนะ	89
บรรณานุกรม		**90**
ภาคผนวก	จดหมายตอบรับตีพิมพ์เผยแพร่บทความในวารสารรังสิตสารสนเทศ	**93**
ประวัติผู้วิจัย		**95**

มหาวิทยาลัยรังสิต Rangsit University

สารบัญตาราง

ตารางที่		หน้า
2.1	ตารางลักษณะของข้อมูลที่ใช้ในการวิเคราะห์ความแปรปรวนทางเดียว	23
3.1	ตัวอย่างข้อมูลที่จะนำไปเปรียบเทียบ	39
3.2	ตัวอย่างข้อมูลที่จะนำไปหาความสัมพันธ์	41
4.1	สรุปผลการพยากรณ์ด้วยแบบจำลองโดยใช้เทคนิคการจำแนกข้อมูลด้วยวิธีต้นไม้การตัดสินใจ	57
4.2	สรุปผลการพยากรณ์ด้วยแบบจำลองโดยใช้เทคนิคการจำแนกข้อมูลด้วยวิธีเครือข่ายประสาทเทียม	65
4.3	สรุปผลการพยากรณ์ด้วยแบบจำลองโดยใช้เทคนิคการจำแนกข้อมูลด้วยวิธีไนอีฟเบย์	75
4.4	สรุปผลการเปรียบเทียบแบบจำลองการพยากรณ์ด้วย ANOVA	77
4.5	สรุปผลการหาความสัมพันธ์ของข้อมูลด้วย Association Rules	86

มหาวิทยาลัยรังสิต Rangsit University

สารบัญรูป

รูปที่		หน้า
2.1	กรอบแนวคิดในการวิจัยตามทฤษฎีการยอมรับและการใช้เทคโนโลยี (Unified theory of acceptance and use of technology: UTAUT)	5
2.2	รูปแสดงระยะห่างยูคลิเดียน	17
2.3	รูปแสดงโครงสร้างของเครือข่ายประสาทเทียม	22
2.4	องค์ประกอบของโปรแกรม RapidMiner Studio 9	25
2.5	เมนูใน RapidMiner Studio 9	25
2.6	เครื่องมือ Repository	26
2.7	เครื่องมือ Operators	27
2.8	หน้าจอ Process	28
2.9	หน้าจอ Parameters	29
2.10	หน้าจอ Help	29
2.11	การใช้ Operators Read CSV	30
2.12	การใช้ Operators Set Role	30
2.13	การใช้ Operator Numerical to Binominal	31
2.14	การแสดงข้อมูลที่ได้ทำการดึงมา	31
2.15	การใช้ Operators Cross Validation	32
2.16	การเลือก Operators ภายใน Cross Validation	33
2.17	การเลือก Operations ภายใน Vote	33
2.18	ตัวอย่างการแสดงผลการพยากรณ์พนักงานที่จะลาออก ด้วย Software RapidMiner	34
3.1	ขั้นตอนทำการเปรียบเทียบเพื่อพยากรณ์	42
3.2	ขั้นตอนเพื่อหาความสัมพันธ์ของข้อมูล	44
4.1	Operators หลักที่ใช้	47
4.2	Operators Retrieve	47
4.3	การ Set Parameters Operators Retrieve	48
4.4	Operators Set Role	48

มหาวิทยาลัยรังสิต Rangsit University

สารบัญรูป (ต่อ)

รูปที่		หน้า
4.5	การ Set Parameters Operators Set Role	48
4.6	Operators Cross Validation	49
4.7	การ Set Parameters Operators Cross Validation	49
4.8	Operators หลัก สำหรับ Decision Tree	50
4.9	Operators Retrieve Decision	50
4.10	การ Set Parameters Operators Retrieve Decision	51
4.11	Operators Set Role Decision	51
4.12	การ Set Parameters Operators Set Role Decision	51
4.13	Operators Cross Validation Decision	52
4.14	การ Set Parameters Operators Cross Validation Decision	52
4.15	Operators ภายใน Operators Cross Validation Decision	53
4.16	Operators Decision Tree	53
4.17	การ Set Parameters Operators Decision Tree	54
4.18	Operators Apply Model	54
4.19	การ Set Parameters Operators Apply Model	55
4.20	Operators Performance	55
4.21	การ Set Parameters Operators Performance	56
4.22	ผลการพยากรณ์แบบ Table View โดยเทคนิคต้นไม้การตัดสินใจ	56
4.23	ผลการพยากรณ์แบบ Chart view โดยเทคนิคต้นไม้การตัดสินใจ	57
4.24	Operators หลักสำหรับ Neural Network	58
4.25	Operators Retrieve Neural	58
4.26	การ Set Parameters Retrieve Neural	59
4.27	Operators Set Role	59
4.28	การ Set Parameters Operators Set Role	59
4.29	Operators Cross Validation Neural	60
4.30	Set Parameters Cross Validation Neural	60

มหาวิทยาลัยรังสิต Rangsit University

สารบัญรูป (ต่อ)

รูปที่		หน้า
4.31	Operators ภายใน Operators Neural Network	61
4.32	Operators Neural Net	61
4.33	การ Set Parameters Operators Neural Net	62
4.34	Operators Apply Model	62
4.35	การ Set Parameters operators Apply Model	63
4.36	Operators Performance	63
4.37	การ Set Parameters Operators Performance	64
4.38	ผลการพยากรณ์แบบ Table View โดยเทคนิคเครือข่ายประสาทเทียม	64
4.39	ผลการพยากรณ์แบบ Chart View โดยเทคนิคเครือข่ายประสาทเทียม	65
4.40	Operators หลักสำหรับ Naïve Bayes	66
4.41	Operators Retrieve Naïve	66
4.42	Set Parameters Operators Retrieve Naïve	67
4.43	Operators Set Role	67
4.44	Set Parameters Operators Set Role	67
4.45	Operators Cross Validation Naïve	68
4.46	Set Parameters Operators Cross Validation Naïve	68
4.47	Operators ภายใน Operators Cross Validation Naïve	69
4.48	Operators Naïve Bayes	69
4.49	Set Parameters Operators Naïve Bayes	69
4.50	Operators Apply Model	70
4.51	Set Parameters Operators Apply Model	70
4.52	Operators Performance	70
4.53	Set Parameters Operators Performance	71
4.54	ผลการพยากรณ์แบบ Table View โดยเทคนิค ไนอีฟเบอร์	71
4.55	ผลการพยากรณ์หัวข้อ PE แบบ Chart View โดยเทคนิคไนอีฟเบอร์	72
4.56	ผลการพยากรณ์หัวข้อ EE แบบ Chart View โดยเทคนิคไนอีฟเบอร์	72

มหาวิทยาลัยรังสิต Rangsit University

สารบัญรูป (ต่อ)

รูปที่		หน้า
4.57	ผลการพยากรณ์หัวข้อ SI แบบ Chart View โดยเทคนิคไนอีฟเบอร์	73
4.58	ผลการพยากรณ์หัวข้อ FC แบบ Chart View โดยเทคนิคไนอีฟเบอร์	73
4.59	ผลการพยากรณ์หัวข้อ BI แบบ Chart View โดยเทคนิคไนอีฟเบอร์	74
4.60	ผลการพยากรณ์หัวข้อ BU แบบ Chart View โดยเทคนิคไนอีฟเบอร์	74
4.61	Operators หลักที่ใช้ในการเปรียบเทียบโดยเทคนิค ANOVA	75
4.62	Operators ANOVA	76
4.63	Set Parameters Operators ANOVA	76
4.64	ผลการเปรียบเทียบ 3 เทคนิคโดยใช้ ANOVA แสดงค่าแบบ Table View	76
4.65	การพยากรณ์โดยใช้ Decision Tree	77
4.66	Operators หลักที่ใช้ใน Association Rule	79
4.67	Operators Retrieve associate Support	80
4.68	Set Parameters Operators Retrieve Associate Support	80
4.69	Operators Select Attributes	80
4.70	Set Parameters Operators Select Attributes	81
4.71	การเลือก Parameters Attributes	81
4.72	Operators Numerical to Binominal	82
4.73	Set Parameters Operators Numerical to Binominal	82
4.74	Operators FP-Growth	83
4.75	Set Parameters Operators FP-Growth	84
4.76	Operators Create Association Rules	85
4.77	Set Parameters Operators Create Association Rules	85
4.78	แสดงความสัมพันธ์ของข้อมูล แบบ Table View โดยใช้เทคนิค Association Rules	86

มหาวิทยาลัยรังสิต Rangsit University

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ในปัจจุบันมีการแพร่ระบาดของ COVID-19 ทำให้ผู้บริโภคปรับเปลี่ยนพฤติกรรมการซื้อสินค้าทำให้การค้าออนไลน์เติบโต จากผลสำรวจจาก Future Shopper 2021 โดยบริษัท Wunderman Thompson เผยว่าคนไทย 94% เริ่มซื้อสินค้าออนไลน์ตั้งแต่ปี 2563 ขณะที่คนไทยอีก 90% ระบุว่าจะซื้อสินค้าผ่านทางช่องทางการค้าออนไลน์อีกต่อไป ในปี 2564 ข้อมูลจากกรมพัฒนาธุรกิจการค้าพบว่า บริษัท ลาซาด้า จำกัด แจ้งรายได้รวม 14,675 ล้านบาท เพิ่มขึ้น 46% กำไรสุทธิ 226 ล้านบาท เพิ่มขึ้น 105% โดยได้ครองเจ้าตลาดการค้าออนไลน์ และเทคโนโลยีของมือถือ สมาร์ทโฟน และแท็บเล็ตต่าง ๆ ได้ถูกพัฒนาและเติบโตอย่างรวดเร็ว ผู้บริโภคนิยมหันมาใช้งานอุปกรณ์เหล่านี้เป็นจำนวนมาก เนื่องจากมีขนาดเล็ก พกพาง่าย และสามารถใช้งานอินเทอร์เน็ตได้ นอกจากนี้ยังมีแอปพลิเคชันต่าง ๆ ให้ดาวน์โหลดมาใช้งานมากมาย ด้วยเหตุผลนี้ในการค้าออนไลน์ อุปกรณ์เหล่านี้ย่อมมีผลต่อการทำการตลาดอย่างแน่นอน เนื่องจากพฤติกรรมการใช้งานอินเทอร์เน็ตของผู้บริโภคจะอยู่ที่อุปกรณ์สมาร์ทโฟนและ แท็บเล็ตมากขึ้น ตามผลสำรวจ Insight & Digital Stat Thai 2022 (Nattapon, 2022) ผู้ครอบครองสมาร์ทโฟนเป็นจำนวน 96.2% ของประชากรโลก

จากการศึกษางานวิจัยเกี่ยวกับการพยากรณ์การเลือกกลุ่มคณะวิชาในการศึกษาต่อในระดับปริญญาตรี ของมหาวิทยาลัยรังสิต โดยใช้เทคนิคการทำเหมืองข้อมูล (ฐิตินันท์ อนุสรณ์ และสมชาย เล็กเจริญ, 2563) และ ทฤษฎี กฎความสัมพันธ์ (Association Rule) เพื่อหาความสัมพันธ์ระหว่างกลุ่มของ 2 กลุ่มขึ้นไป ผู้วิจัยจึงมีแนวคิดที่จะพยากรณ์และหาความสัมพันธ์ของการยอมรับการใช้เทคโนโลยีที่มีผลต่อพฤติกรรมการใช้แอปพลิเคชัน LAZADA ช้อปปิ้งบนสมาร์ทโฟนของประชากรในเขตกรุงเทพมหานครและปริมณฑล เพื่อนำข้อมูลดังกล่าวไปพัฒนาแอปพลิเคชันให้ตรงกับความต้องการของผู้บริโภคมากยิ่งขึ้น การวิจัยครั้งนี้จำเป็นที่จะต้องศึกษาถึงเทคนิควิธีการพยากรณ์ที่เหมาะสมกับข้อมูลมากที่สุด

มหาวิทยาลัยรังสิต Rangsit University

ดังนั้นผู้วิจัยจึงนำเอาข้อมูลทุติยภูมิของผู้ที่เคยใช้แอปพลิเคชันช้อปปิ้งบนสมาร์ทโฟน (LAZADA) นำมาวิเคราะห์ด้วยวิธีการทำเหมืองข้อมูล เพื่อทำการพยากรณ์ และค้นหารูปแบบความสัมพันธ์จากข้อมูล โดยอาศัยหลักการทางสถิติและการเรียนรู้ด้วยเครื่องจักร (Machine Learning) กระบวนการทำงานของเหมืองข้อมูลแบบการจำแนกข้อมูล (Classification) (หนึ่งหทัยชัยอาภร, 2560) โดยผู้วิจัยได้เลือกเทคนิคการจำแนกข้อมูลมา 3 แบบ คือ ต้นไม้การตัดสินใจ โครงข่ายประสาทเทียม และ ไนอีฟเบย์ เนื่องจากเป็นเทคนิคที่สามารถอธิบายให้เข้าใจในผลลัพธ์ที่ได้โดยง่าย จึงเหมาะต่อผู้ที่สนใจทั่วไป โดยประโยชน์ที่ได้รับจะสามารถนำไปเพิ่มประสิทธิภาพแอปพลิเคชันช้อปปิ้งบนสมาร์ทโฟนให้ตอบสนองความต้องการผู้ใช้งานให้เกิดประโยชน์สูงสุด ซึ่งจะทำให้เกิดการลดต้นทุนของการสร้างแอปพลิเคชัน และเป็นการลดต้นทุนของผู้ขายและผู้ซื้อด้วยเช่นกัน

1.2 วัตถุประสงค์ของการวิจัย

1.2.1 เพื่อเปรียบเทียบแบบจำลองการพยากรณ์ระหว่างต้นไม้การตัดสินใจ (Decision Tree) โครงข่ายประสาทเทียม (Neural Network) และไนอีฟเบย์ (Naïve Bayes) โดยทำนายการจำแนกคุณลักษณะข้อมูล ของการยอมรับการใช้เทคโนโลยีที่มีผลต่อพฤติกรรมการใช้แอปพลิเคชัน LAZADA ช้อปปิ้งบนสมาร์ทโฟนของประชากรในกรุงเทพมหานครและปริมณฑล ประกอบด้วย 6 ด้าน ซึ่งจะบอกได้ว่าด้านใดมีความสัมพันธ์กับด้านใด ซึ่งจะสามารถนำไปประยุกต์ใช้กับทางโฆษณาแอปพลิเคชัน LAZADA ได้

1.2.2 เพื่อหาความสัมพันธ์ของการยอมรับการใช้เทคโนโลยีที่มีผลต่อพฤติกรรมการใช้แอปพลิเคชัน LAZADA ช้อปปิ้งบนสมาร์ทโฟนของประชากรในเขตกรุงเทพมหานครและปริมณฑล โดยใช้เทคนิคการกฎความสัมพันธ์ Association Rules

1.3 ขอบเขตของการวิจัย

1.3.1 ดำเนินการสืบค้นและศึกษางานวิจัยที่เกี่ยวข้องของผู้ที่ใช้งานแอปพลิเคชัน LAZADA ช้อปปิ้งบนสมาร์ทโฟน ที่อาศัยในเขตกรุงเทพมหานครและปริมณฑล

มหาวิทยาลัยรังสิต Rangsit University

1.3.2 ดำเนินการสืบค้นและศึกษางานวิจัยการใช้เทคนิคการทำเหมืองข้อมูลในเชิงเปรียบเทียบ และการหาความสัมพันธ์ของชุดข้อมูล

1.3.3 ดำเนินการเพื่อเปรียบเทียบและหาความสัมพันธ์ของข้อมูลโดยใช้โปรแกรม Rapid miner Studio Version 9.10

1.3.4 วิเคราะห์และสรุปผลที่ได้จากการทดสอบ พร้อมทั้งปรับปรุงแก้ไขส่วนที่บกพร่อง

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1.4.1 ทำให้ทราบถึงลักษณะความแตกต่างกันส่งผลต่อพฤติกรรมการใช้สมาร์ทโฟนในการช้อปปิ้งสินค้า

1.4.2 ทำให้ทราบถึงความสัมพันธ์ของข้อมูลในด้านต่าง ๆ ที่จะนำไปพัฒนาแอปพลิเคชันให้มีประสิทธิ์ภาพสูงสุด

1.4.3 เพื่อเป็นแนวทางในการพัฒนาแอปพลิเคชันในด้านต่าง ๆ เพื่อตอบสนองผู้ใช้งานได้มากที่สุด

มหาวิทยาลัยรังสิต Rangsit University

บทที่ 2

ทบทวนวรรณกรรมและทฤษฎีที่เกี่ยวข้อง

การวิจัยการพยากรณ์และความสัมพันธ์ของการยอมรับการใช้เทคโนโลยีที่มีผลต่อพฤติกรรมการใช้แอปพลิเคชัน LAZADA ช้อปปิ้งบนสมาร์ทโฟนของประชากรในเขตกรุงเทพมหานครและปริมณฑล โดยใช้เทคนิคการทำเหมืองข้อมูล ผู้วิจัยได้รวบรวมสาระความรู้และงานวิจัยที่เกี่ยวข้องเพื่อเป็นแนวทางศึกษาดังนี้

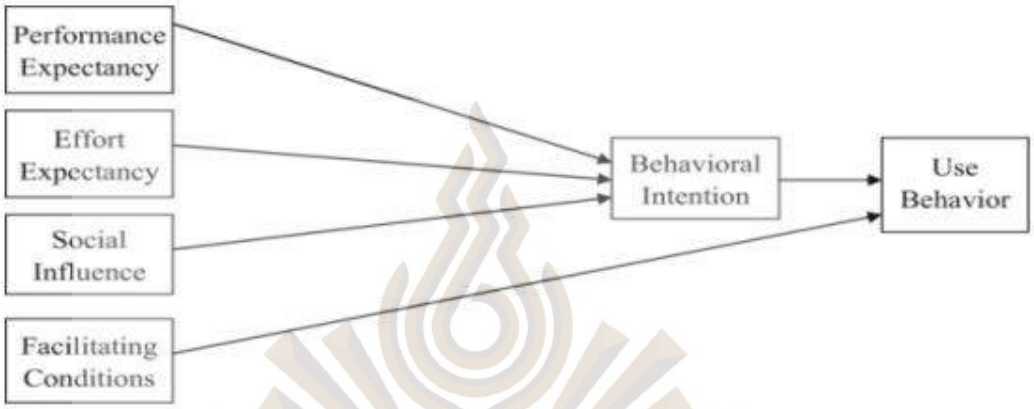
2.1 ทฤษฎีที่เกี่ยวข้อง

2.1.1 กรอบแนวคิดในการวิจัยตามทฤษฎีการยอมรับและการใช้เทคโนโลยี

Venkatesh, Morris, Davis and Davis (2003) ได้กล่าวไว้ว่า การยอมรับและการใช้เทคโนโลยีประกอบด้วย ความคาดหวังในประสิทธิภาพ (Performance Expectancy) ความคาดหวังในความพยายาม (Effort Expectancy) และอิทธิพลของสังคม (Social Influence) ส่วนสภาพสิ่งอำนวยความสะดวกในการใช้งานมีความสัมพันธ์โดยตรงต่อพฤติกรรมการใช้งาน การยอมรับและการใช้เทคโนโลยีเป็นแนวทางที่ช่วยให้ผู้ศึกษาสามารถวิเคราะห์พฤติกรรมมนุษย์ได้ ใช้เป็นทางเลือกสำหรับอธิบายความตั้งใจและพยากรณ์พฤติกรรมมนุษย์ในการยอมรับการใช้เทคโนโลยีสารสนเทศได้อย่างมีประสิทธิภาพยิ่งขึ้น (สิงหะ ฉวีสุข และสุนันทา วงศ์จตุรภัทร, 2555) โดยแบ่งออกเป็น 6 หัวข้อ ได้แก่

1) ความคาดหวังด้านประสิทธิภาพ (Perfomance Expectancy)
2) ความคาดหวังด้านพยายาม (Effort Expectancy)
3) อิทธิพลของสังคม (Social Influence)
4) สภาพสิ่งอำนวยความสะดวก (Facilitating Conditions)
5) พฤติกรรมความตั้งใจ (Behavioral Intention)
6) พฤติกรรมการใช้งาน (Use Behavior)

โดยความเกี่ยวข้องจะมีหัวข้อ Performance Expectancy (PE), Effort Expectancy (EE) และ Social Influence (SI) จะมีส่งผลในด้านหัวข้อ Behavioral Intention (BI) และจะส่งผลต่อไปยัง Use Behavior (UB) โดยตรงส่วนหัวข้อ Facilitating Conditions จะส่งผลโดยตรงกับหัวข้อ Use Behavior ดังรูปที่ 2.1



รูปที่ 2.1 กรอบแนวคิดในการวิจัยตามทฤษฎีการยอมรับและการใช้เทคโนโลยี
(Unified theory of acceptance and use of technology: UTAUT)
ที่มา: Venkatesh et al., 2003

2.1.1.1 การวิเคราะห์รูปแบบความสัมพันธ์เชิงสาเหตุแบบมีตัวแปรแฝง (Causal Structural Models with Latent Variable) นงลักษณ์ วิรัชชัย (2542) ได้กล่าวไว้ว่า ตัวอย่างในการทดสอบค่าความเหมาะสมควรจะอยู่ในอัตราส่วน 10-20 ต่อ 1 ตัวแปรหรือตัวอย่างอย่างน้อยที่สุดที่สามารถยอมรับได้ให้ดูจากค่าสถิติ Holster ที่ต้องมากกว่า 200 จึงจะถือว่าเป็นรูปแบบของความสัมพันธ์เชิงสาเหตุมีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์

2.1.1.2 สถิติเชิงบรรยาย (Descriptive Statistics) เป็นสถิตที่ต้องใช้ในการวิจัยเชิงปริมาณ สถิติที่นิยมใช้ คือ

1) การแจกแจงความถี่ (Frequency) เป็นการจัดหมวดหมู่ของข้อมูลให้เป็นระบบระเบียบเพื่อสะดวกในการแปลความหมาย โดยวิธีการเรียงลำดับข้อมูลจากมากไปหาน้อยหรือจากน้อยไปหามาก เพื่อแสดงว่าตัวเลขนั้น ๆ ซ้ำกัน หรือมีความถี่กี่ครั้ง

2) ร้อยละ (Percentage) ใช้สัญลักษณ์ % การคำนวณใช้สูตรดังนี้

$$\text{ร้อยละ } \frac{\text{จำนวนที่ต้องการเปรียบเทียบ}}{\text{จำนวนทั้งหมด}} \times 100 \quad (2\text{-}1)$$

3) ค่าเฉลี่ย (Mean) คือ ค่ากลางที่คำนวณได้จากการนำข้อมูลทั้งหมดมารวมกันแล้วหารด้วยจำนวนข้อมูลทั้งหมด มีสูตรดังนี้

$$\bar{X} \frac{\sum X}{N} \quad (2\text{-}2)$$

เมื่อ $\sum X$ แทน ผลรวมของข้อมูลทั้งหมด

N แทน จำนวนข้อมูลทั้งหมด

4) ส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation) คือ ค่าที่ใช้วัดการกระจายของข้อมูลแต่ละตัวว่าต่างไปจากค่าเฉลี่ยมากน้อยเพียงใด ใช้สัญลักษณ์ S.D. การคำนวณใช้สูตรดังนี้

$$S.D. = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}} \quad (2\text{-}3)$$

เมื่อ S.D. คือ ส่วนเบี่ยงเบนมาตรฐาน

X คือ ข้อมูล (ตัวที่ 1, 2, 3…, n)

$\bar{x}$ คือ ค่าเฉลี่ยเลขคณิต

n คือ จำนวนข้อมูลทั้งหมด

2.1.1.3 สถิติเชิงอนุมาน (Inferential Statistics) คือ สถิติที่ใช้ในการวิเคราะห์ความสัมพันธ์เชิงสาเหตุหาเส้นทางอิทธิพลในเชิงหาสาเหตุของตัวแปรหาขนาดอิทธิพล และทิศทางว่าเป็นอย่างไรในแนวคิดและทฤษฎีที่อ้างอิง มีการทดสอบความสอดคล้องกลมกลืมระหว่างโมเดลสมมุติฐานกับข้อมูลเชิงประจักษ์ ค่าสถิติ Chi-square P-value มากกว่า .05, CMIN/DF น้อยกว่า 2.00, GRI มากกว่า .90, AGFI มากกว่า .90 และ RMSEA น้อยกว่า .08 ซึ่งสอดคล้องกับสถิติวิเคราะห์โมเดลสมการโครงสร้าง (กริช แรงสูงเนิน, 2554) การสรุปอ้างอิงค่าสถิติต่าง ๆ ที่เกิดขึ้นในกลุ่มตัวอย่างที่ศึกษาของกลุ่มตัวอย่างนั้น

2.1.2 การทำเหมืองข้อมูล (Data Mining)

ความหมายของการทำเหมืองข้อมูล

การทำเหมืองข้อมูล จิราภา เลาหะวรนันท์, รชต ลิ้มสุทธิวันภูมิ, และบัณฑิต ฐานะโสภณ (2558) กล่าวว่า เทคนิคการทำเหมืองข้อมูล (Data Mining) คือ การนำข้อมูลจำนวนมากมาวิเคราะห์ เพื่อหาความสัมพันธ์ของข้อมูลที่ซ่อนอยู่ โดยทำการจำแนกประเภท รูปแบบ เชื่อมโยงข้อมูลที่มีความสัมพันธ์กัน และความน่าจะเป็นที่จะเกิดขึ้น เพื่อให้ได้องค์ความรู้ใหม่ ที่สามารถนำไปใช้ประกอบการตัดสินใจด้านต่าง ๆ

วิวัฒนาการของ การทำเหมืองข้อมูล

ปี 1960 Data Collection คือ การนำข้อมูลมาจัดเก็บอย่างเหมาะสมในอุปกรณ์ที่น่าเชื่อถือและป้องกันการสูญหายได้เป็นอย่างดี

ปี 1980 Data Access คือ การนำข้อมูลที่จัดเก็บมาสร้างความสัมพันธ์ต่อกันในข้อมูลเพื่อประโยชน์ในการนำไปวิเคราะห์ และการตัดสินใจอย่างมีคุณภาพ

ปี 1990 Data Warehouse & Decision Support คือ การรวบรวมข้อมูลมาจัดเก็บลงไปในฐานข้อมูลขนาดใหญ่โดยครอบคลุมทุกแง่ทุกมุมขององค์กร เพื่อช่วยสนับสนุนการตัดสินใจ

ปี 2000 การทำเหมืองข้อมูล คือ การนำข้อมูลจากฐานข้อมูลมาวิเคราะห์และประมวลผลโดยการสร้างแบบจำลองและความสัมพันธ์ทางสถิติจากคำจำกัดความของ การทำเหมืองข้อมูลหมายถึงการที่ผู้ใช้งานสามารถดึงข้อมูลโดยการสังเคราะห์และตรวจสอบข้อมูลอย่างละเอียด โดยการสังเคราะห์ดังกล่าวอาจเป็นการเรียนรู้ข้อมูลในอดีตหรือข้อมูลในปัจจุบัน ผลลัพธ์ที่ได้มาต้องมีลักษณะของข้อมูลที่เป็นข้อมูลแบบ Unknown, ข้อมูลแบบ Valid, และข้อมูลแบบ Actionable มาจากฐานข้อมูลขนาดใหญ่ซึ่งมาจากรายการ Transaction, ฐานข้อมูลของฝ่ายขาย, E-Mail เพื่อนำข้อมูลดังกล่าวไปใช้เป็นพื้นฐานในการประกอบการตัดสินใจ ในเชิงธุรกิจ ทำให้เข้าใจแนวโน้มและรูปแบบของตลาดข้อมูลแบบ Unknown ข้อมูลที่ถูกใช้จะต้องเป็นข้อมูลผู้ใช้งานไม่รู้มาก่อนและไม่ชัดเจนจนไม่สามารถตั้งสมมติฐานได้ล่วงหน้าว่าควรจะเป็นแบบใด ตัวอย่างเช่น เจ้าของห้างสรรพสินค้าแห่งหนึ่งเพิ่งจะค้นพบว่าพฤติกรรมของผู้บริโภคใหม่ที่เป็นพ่อบ้านมักจะซื้อสินค้าประเภทเบียร์และผ้าอ้อมในวันศุกร์ตอนเย็น ดังนั้นเพื่อเป็นสัญญาณให้เจ้าของกิจการควรจะเตรียม

สินค้าไว้เพื่อจำหน่ายในขณะเดียวกันห้างสรรพสินค้าคู่แข่งอาจจะไม่รู้เรื่องนี้ หรือตัวอย่างของเจ้าของร้านขายรถยนต์พบว่ารถขนาดใหญ่ราคาแพงมักจะมีผู้ซื้อเป็นผู้สูงอายุ ซึ่งเจ้าของร้านไม่รู้มาก่อนแต่ข้อมูลดังกล่าวไม่เป็นลักษณะ Unknown ตามสมมติฐานดังกล่าวมีอยู่ เพราะคนที่มีอายุมักจะมีฐานะที่ดีขึ้นเมื่อเทียบกับคนในวัยที่อายุน้อยกว่าข้อมูลแบบ Valid เมื่อผู้ใช้ได้เริ่มใช้เทคนิคการทำเหมืองข้อมูลจะค้นพบสิ่งที่น่าสนใจโดยต้องพิจารณาด้วยว่าสิ่งนั้น Valid หรือไม่ เช่น ผู้ใช้มักจะพบว่ามีความสัมพันธ์ของการซื้อของ 2 สิ่งเสมอ เมื่อจำนวนความหลากหลายของสินค้ามีมากขึ้น แต่ไม่ได้หมายความว่าจะต้องให้ห้างสรรพสินค้าเก็บสต๊อกสินค้ามากขึ้น เพราะข้อมูลที่ได้อาจเกิดความคลาดเคลื่อน เพราะฉะนั้นจะต้องทำ การValidation และ Checking ความถูกต้องของข้อมูลและวิเคราะห์ความถูกต้องอีกครั้งข้อมูลแบบ Actionable ข้อมูลจะต้องถูกแปลงออกมาและนำมาตัดสินใจให้เป็นความได้เปรียบเชิงธุรกิจในบางครั้ง ข้อมูลที่เราค้นพบเป็นสิ่งที่คู่แข่งได้ทำไปแล้วจึงต้องมีวิจารณญาณในการใช้ด้วย บางทีข้อมูลดังกล่าว อาจจะไม่มีประโยชน์ก็ได้

คำว่าการทำเหมืองข้อมูลนั้นมีความหมายแตกต่างกันใน 2 มุมมอง คือ ในมุมมองทางวิชาการและในมุมมองเชิงธุรกิจ ในมุมมองเชิงวิชาการนั้น นักวิจัยจะอ้างถึงกระบวนการทั้งหมดในการทำ Data Mining ว่า “Knowledge Discovery in Database (KDD)” และใช้คำว่า “การทำเหมืองข้อมูล” แทนขั้นตอนขั้นหนึ่งของกระบวนการที่เกี่ยวข้องกับการค้นหารูปแบบ ความสัมพันธ์ของข้อมูลเท่านั้น อย่างไรก็ตาม ในแง่มุมเชิงธุรกิจแล้ว จะใช้คำว่า “การทำเหมืองข้อมูล” แทนความหมายของขั้นตอนทั้งหมด เดิมงานค้นคว้าทางด้านการทำเหมืองข้อมูลนั้น มีการทำการค้นคว้ากันอยู่แล้วในหลาย ๆ สาขาวิชา แต่มีชื่อเรียกแตกต่างกันไปตามแต่ละด้าน นักวิจัยในด้านสถิติ (Statistics), ฐานข้อมูล (Database), Neural Networks, Pattern Recognition, Machine Learning, Econometrics และอีกหลาย ๆ ด้าน ต่างก็มีการค้นคว้าเกี่ยวกับปัญหาในลักษณะเดียวกันนี้ แต่ยังไม่ค่อยมีการใช้ประโยชน์ของการค้นคว้าของอีกฝ่ายหนึ่ง คือ ต่างฝ่ายต่างทำ การค้นคว้าของตนเอง ไม่ค่อยมีการแลกเปลี่ยนความรู้กันทำให้การค้นคว้าและการเผยแพร่ผลงานดำเนินไปอย่างไม่รวดเร็วเท่าที่ควร ต่อมาจึงมีการใช้ “การทำเหมืองข้อมูล” เป็นชื่อรวม ของวิธีการแก้ปัญหาในลักษณะนี้ ซึ่งทำให้การเผยแพร่ความรู้ในการแก้ปัญหาลักษณะนี้ทำได้รวดเร็วและสามารถอ้างอิงได้ สะดวกขึ้น

หลักการทั่วไปของ Knowledge Discovery in Database (KDD) และการทำเหมืองข้อมูล

KDD หมายถึงกระบวนการในการค้นหาลักษณะแฝงของข้อมูลที่อยู่ในกลุ่มข้อมูลจำนวนมาก ซึ่งมีขั้นตอนการทำ การทำเหมืองข้อมูล เป็นกระบวนการที่สำคัญในการค้นหาลักษณะที่

น่าสนใจของข้อมูลเหล่านี้ เช่น รูปแบบ ความสัมพันธ์ การเปลี่ยนแปลง โครงสร้างที่เด่นชัด หรือลักษณะที่ผิดปกติของข้อมูลจากข้อมูลจำนวนมาก ๆ ที่เก็บอยู่ในฐานข้อมูล หรือแหล่งที่เก็บข้อมูลอื่น ๆ ซึ่งวิธีการต่าง ๆ ที่นำมาใช้ในการทำ Mining นี้ก็มีวัตถุประสงค์ต่าง ๆ กันขึ้นอยู่กับผลลัพธ์ของ กระบวนการโดยรวมที่ต้องการ ดังนั้นจึงควรมีการนำ เสนอวิธีการที่หลากหลายสำหรับเป้าหมายที่แตกต่างกัน เพื่อให้ได้ผลลัพธ์ ที่เหมาะสมตามที่ต้องการ หลังจากนำ ไปใช้งานแล้ว และเนื่องจากความแพร่หลายของการจัดเก็บข้อมูลในลักษณะที่เป็นรูปแบบทางอิเล็กทรอนิกส์ และความต้องการในการเปลี่ยนข้อมูลเหล่านั้นให้เป็นข้อมูลที่มีประโยชน์ต่อการนำ ไปประยุกต์ ใช้ในงานด้านต่าง ๆ เช่น การวิเคราะห์ด้านการตลาด การบริหารธุรกิจ รวมถึงระบบที่ช่วยสนับสนุนการตัดสินใจ เป็นต้น ดังนั้นจึงทำ ให้การนำ การทำเหมืองข้อมูล มาใช้ได้รับความสนใจมากในช่วง 2-3 ปีที่ผ่านมา จากที่ได้กล่าวแล้วว่า การทำเหมืองข้อมูล เป็นขั้นตอนหนึ่งที่สำคัญในกระบวนการค้นหาลักษณะของข้อมูลแฝง ที่มีประโยชน์ในฐานข้อมูล (Knowledge Discovery in Database: KDD) ซึ่งโดยทั่วไปกระบวนการของ KDD นั้นประกอบด้วยขั้นตอนต่าง ๆ ดังนี้

1) การคัดเลือกข้อมูล (Data Selection) เป็นการบอกถึงแหล่งข้อมูลที่จะนำมาใช้ในการทำ Data Mining รวมถึงการนำข้อมูลที่ต้องการออกมาจากข้อมูล เพื่อทำการพิจารณาในขั้นตอนต่อไป

2) การกรองข้อมูล (Data Cleaning) เป็นกระบวนการที่ทำ ให้เกิดความมั่นใจในคุณภาพของข้อมูลที่จะนำมาใช้วิเคราะห์ว่าถูกต้อง โดยการนำ ข้อมูลที่ไม่ถูกต้องออก

3) การแปลงรูปแบบข้อมูล (Data Transformation) เป็นการแปลงข้อมูลที่เลือกมาให้อยู่ในรูปแบบที่เหมาะสม สำ หรับการนำ ไปใช้วิเคราะห์ตามอัลกอริทึม (Algorithm) และแบบจำลองที่ใช้ ในการทำ เหมืองข้อมูล ต่อไป

4) การทำ Mining ข้อมูล (การทำเหมืองข้อมูล) การใช้เทคนิคภายใน การทำเหมืองข้อมูล เพื่อทำการ Mine ข้อมูล โดยทั่วไป ประเภทของงานตามลักษณะของแบบจำ ลองที่ใช้ในการทำ การทำเหมืองข้อมูล นั้นสามารถแบ่งกลุ่มได้เป็น 2 ประเภทใหญ่ๆ คือ

4.1) Predictive การทำเหมืองข้อมูล คือ เป็นการคาดคะเนลักษณะหรืออประมาณค่าที่ชัดเจนของข้อมูลที่จะเกิดขึ้น โดยใช้พื้นฐานจากข้อมูลที่ผ่านมาในอดีต

4.2) Descriptive การทำเหมืองข้อมูล คือ เป็นการหาแบบจำลองเพื่ออธิบายลักษณะบางอย่างของข้อมูลที่มีอยู่ ซึ่งโดยส่วนมากจะเป็นลักษณะการแบ่งกลุ่มให้กับข้อมูล

5) การวิเคราะห์และประเมินผลลัพธ์ที่ได้ (Result Analysis and Evaluation) เป็นขั้นตอนการแปลความหมายและการประเมินผลลัพธ์ที่ได้ว่ามีความเหมาะสมหรือตรงกับวัตถุประสงค์ที่ต้องการหรือไม่ โดยทั่วไปควรมีการแสดงผลในรูปแบบที่สามารถเข้าใจได้โดยง่าย

ขั้นตอนการทำงานของ การทำเหมืองข้อมูล

1) การกำหนดวัตถุประสงค์ทางธุรกิจ (Problem Formulation)

การกำหนดวัตถุประสงค์ทางธุรกิจ คือ จะต้องเข้าใจปัญหาและความต้องการทางธุรกิจ เพื่อที่จะทำเป็นส่วนที่กำหนดว่าเมื่อไหร่ที่จะใช้ การทำเหมืองข้อมูล ในการแก้ปัญหาซึ่งในส่วนนี้จะประกอบไปด้วยการ วิเคราะห์ทางธุรกิจ และการวิเคราะห์เบื้องต้นว่าเรามีข้อมูลใดอยู่บ้างและต้องการอะไรจากข้อมูล ซึ่งขั้นตอนนี้จะสามารถมองถึง อัลกอริทึม และฐานข้อมูลที่สัมพันธ์กับวัตถุประสงค์ทางธุรกิจได้ การใช้งาน การทำเหมืองข้อมูล ให้ได้ประโยชน์สูงสุดจำเป็นต้องมีการกำหนดวัตถุประสงค์ที่ชัดเจน เช่น ต้องการเพื่อจะเพิ่มยอดการตอบรับการขายทางจดหมาย ขึ้นอยู่กับการระบุเป้าหมายว่า จะเพิ่มอัตราการตอบรับหรือเพิ่มมูลค่าการตอบรับซึ่ง จำเป็นที่จะต้องสร้าง Model ที่แตกต่างกัน วัตถุประสงค์ที่กำหนดขึ้นมาจะต้องมีการระบุวิธีการในการวัดผลลัพธ์ที่ได้จากโครงการ และรวมถึงต้นทุนที่สมเหตุสมผลด้วย

2) การคัดเลือกและการเตรียมข้อมูล (Data Selection and Preparation)

การเตรียมข้อมูล (Data Preparation) เป็นหัวใจของขั้นตอนในการทำทั้งหมด เป็นช่วงที่ใช้เวลามากที่สุดในขั้นตอนโดยปกติและต้องการเวลาประมาณ 60% ของเวลาทั้งหมดในการเตรียมข้อมูล ในขั้นตอนนี้อาจสามารถแบ่งออกได้เป็นขั้นตอนย่อยดังต่อไปนี้

การเลือกข้อมูล (Data Selection) จุดประสงค์ คือ การระบุแหล่งของข้อมูลที่มีและดึงเอาข้อมูลออกมาใช้สำหรับการวิเคราะห์เบื้องต้นในการเตรียมตัวสำหรับการที่จะทำการ Mining ในขั้นต่อ ๆ ไป การเลือกข้อมูลนั้นจะแตกต่างไปตามวัตถุประสงค์ของแต่ละธุรกิจ ที่ได้กำหนดไว้ตั้งแต่แรก และการเลือกข้อมูลก็ยังถูกกำหนดโดยลักษณะงานที่จะถูกนำมาใช้อีกด้วยตัวแปรที่ถูกเลือกมาแต่ละตัวแปรนั้นจะต้องถูกทำความเข้าใจว่าตัวแปรแต่ละตัวหมายความว่าอะไร ประกอบด้วยอะไร ไม่เพียงแต่คำจำกัดความทางธุรกิจเท่านั้น แต่จะต้องมีคำอธิบายอย่างชัดเจนเกี่ยวกับชนิดของข้อมูล, ค่าที่เป็นไปได้, แหล่งกำเนิดของข้อมูล, รูปแบบของข้อมูล และลักษณะอื่น ๆ จะมีตัวแปร 2 ชนิด คือ

ตัวแปรแบบ Categorical

1) Nominal Variable คือ Object ที่มันอ้างถึงแต่ไม่มีลำดับในค่าที่เป็นไปได้ (Possible Value) ตัวอย่างเช่น สถานการณ์แต่งงาน (โสด, แต่งงาน, หย่า, ไม่ทราบ), เพศ (ชาย, หญิง), ระดับการศึกษา (ปริญญาโท, ปริญญาตรี, ม.ปลาย, ปวช)

2) Ordinal Variable คือ ตัวแปรที่มีการลำดับสำหรับค่าที่เป็นไปได้ ตัวอย่างเช่น ลำดับของลูกค้า (ดี, ปานกลาง, ไม่ดี)

ตัวแปรแบบ Quantitative

ซึ่งมีการวัดความแตกต่างระหว่างค่าที่เป็นไปได้

1) Continuous (ค่าที่ต่อเนื่อง) เช่นรายได้, เฉลี่ยจำนวนครั้งที่ซื้อ, รายได้

2) Discrete (ค่าเป็นจำนวนเต็ม) เช่นจำนวนพนักงาน, เวลาปี (เดือน, ฤดู, ไตรมาส)

ตัวแปรของข้อมูลมีหลายตัวมากแต่ตัวแปรที่ถูกเลือกสำหรับทำการทำเหมืองข้อมูลนั้นถูกเรียกว่า “Active Variable” เพราะว่ามันจะถูกใช้สร้างความแตกต่างของกลุ่มย่อยต่าง ๆ และสามารถถูกนำมาพยากรณ์ได้ เมื่อได้ทำ การเลือกข้อมูลจะต้องพิจารณาอายุของข้อมูลด้วย เพราะว่าสถานการณ์ภายนอกเปลี่ยนแปลงตลอดเวลาซึ่งจะทำให้ประสิทธิภาพของการทำ Mining ลดลง ตัวอย่าง รสนิยมการใช้ชีวิต การเปลี่ยนงาน

3) การกลั่นกรองข้อมูล (Data Preprocessing)

จุดประสงค์ก็เพื่อทำ ให้มั่นใจว่าคุณภาพของข้อมูลที่ถูกเลือกนั้นเหมาะสม ข้อมูลที่สมบูรณ์เป็นเครื่องประกันว่าการทำเหมืองข้อมูลจะสำเร็จในขั้นตอนนี้เป็นขั้นตอนที่มีปัญหามากกว่า ในขั้นตอนของการเตรียมข้อมูล เพราะข้อมูลส่วนใหญ่ที่มีในองค์กรไม่ได้ถูกเตรียมมาเพื่องาน Mining โดยเฉพาะข้อมูลจะถูกนำมาจากแหล่งต่าง ๆ ถูกจัดเก็บไม่ดี ข้อมูลที่ถูกนำมาจากภายนอก แล้วนำมาเพื่อให้เข้ากับข้อมูลภายในที่มีอยู่ ปัญหาหลักของ Data คือ คุณภาพและความถูกต้องของข้อมูล (Data Integrity) ในขั้นตอนนี้ก่อนอื่นจะต้องทำการทบทวนโครงสร้างของข้อมูลใหม่ และวัดคุณภาพของมัน โดยวิธีทางสถิติ หรือสุ่มตัวอย่าง เครื่องมือที่ใช้ในการทำการกลั่นกรองข้อมูลมีดังต่อไปนี้

3.1) ค่าตัวแปรเป็นแบบ Categorical การแบ่งความถี่ของค่าจะเป็นวิธีที่ทำให้เกิดความเข้าใจใน Data Content เครื่องมือทางด้านกราฟฟิคจะเป็นตัวช่วยให้เห็นและกำหนดค่าที่หายไปได้

3.2) ตัวแปรแบบ Quantitative ตัวแปรประเภทนี้มักมีการใช้การวัดตัวอย่างเช่น ค่าสูงสุดค่าต่ำสุด ค่าเฉลี่ย ค่ากลาง ค่ามัธยฐาน และค่าอื่น ๆ ทางสถิติ เมื่อนำค่าเหล่านี้มาเข้าสูตรคำนวณก็จะบอกถึงค่าที่ไม่สมบูรณ์ หรือค่าที่มีปัญหาเครื่องมือทางกราฟฟิคอื่น ๆ เช่น Scatter Plots คือ รูป 2 มิติ ซึ่งแสดงความสัมพันธ์ระหว่างตัวแปร 2 ตัวแปรขึ้นไป หรือมากกว่า จากกราฟตัวอย่างจะเห็นได้ว่ามีการเปรียบเทียบรายได้กับอายุ จะเห็นได้ว่าจุดจะอยู่สูงขึ้นตามระดับของอายุ ทำให้เราพอที่จะทำนายได้ว่ารายได้ของ อาชีพนี้จะสูงขึ้นเมื่ออายุสูงขึ้น ส่วน Box Plot ถูกใช้ให้เป็นประโยชน์สำหรับเปรียบเทียบศูนย์กลาง (ค่าเฉลี่ย) หรือกระจาย (ค่าเบี่ยงเบน) ของตัวแปรตั้งแต่ 2 ตัวแปรขึ้นไป จากกราฟตัวอย่างตารางแสดง Data Element ของข้อมูลอธิบายถึงรายได้ของผู้ชายและผู้หญิง รูปสี่เหลี่ยมคือเรียกว่า Box และเส้นตั้ง 2 เส้นเรียกว่า Whisker จากความสูงของ Box พอจะสรุปได้ว่ารายได้ของผู้ชายสูงกว่าผู้หญิงระหว่างการทำขั้นตอนการกลั่นกรองข้อมูลจะมีปัญหาบ่อย ๆ ที่มักพบได้ ได้แก่ Noisy Data คือ ตัวแปรตัวหนึ่งหรือมากกว่ามีค่าซึ่งเกินกว่าค่าที่เราคาดไว้ ซึ่งอาจจะหมายถึงทั้งข้อดีและข้อเสียก็ได้ ในข้อดีก็คือ มันจะแสดงอย่างชัดเจนถึงโอกาสซึ่งเรากำลังมองหาอยู่ ในข้อเสีย คือมันอาจจะเป็นข้อมูลที่ไม่สมบูรณ์ สาเหตุที่เกิดขึ้นอาจจะมาจากความไม่รอบคอบของมนุษย์ ตัวอย่างเช่น Operator ใส่อายุให้คนเป็น 300 ปี หรือใส่ค่าของรายได้ เป็นติดลบ ค่าเหล่านี้ควรจะถูกแก้ไข หรือเอาออกจากการวิเคราะห์ ควรมีขั้นตอนการตรวจสอบข้อมูลก่อนนำมาใช้ค่าที่หายไป Missing Value คือ ค่าที่ไม่ได้แสดงในข้อมูลที่เราได้เลือกแล้ว หรือค่าที่ไม่สมบูรณ์ที่เราลบออกไป ระหว่างการทำ Noise Detection ค่าอาจจะหายไปเพราะเกิดจากความไม่รอบคอบของมนุษย์ เพราะว่าไม่มีข้อมูลนั้นระหว่างการทำ Input ข้อมูล การจัดการกับค่าที่หายไปนั้นสามารถจัดการได้ด้วยเทคนิคที่ต่างกัน

4) การสำรวจและตรวจสอบข้อมูล (Data Cleaning and Exploration)

เมื่อทำการเก็บข้อมูลเรียบร้อยแล้ว ขั้นตอนต่อไปที่ควรกระทำ ก็ คือการตรวจสอบข้อมูลเหตุที่ต้องทำ การตรวจสอบข้อมูลมี 2 ข้อ คือ ข้อแรก นักวิเคราะห์ควรมีความคุ้นเคยกับตัวข้อมูล ไม่ใช่รู้แต่ชื่อของ Attribute และความหมายของมันเท่านั้น แต่ต้องรู้ถึงเนื้อหา (Content) หรือความมุ่งหมายที่แท้จริงของข้อมูลด้วย ข้อสองอาจมีความผิดพลาดของการเก็บสะสมข้อมูลเกิดขึ้นในขณะที่ทำการรวบรวมข้อมูลจากฐานข้อมูลหลาย ๆ แหล่งเข้ามาเป็นหนึ่งเดียวเพื่อใช้ในการวิเคราะห์ ซึ่งนักวิเคราะห์ที่ดีจะต้องทำการตรวจสอบข้อมูลเหล่านี้ให้ถูกต้อง ตัวอย่างของความผิดพลาดที่เกิดขึ้น

ได้แก่ ความผิดพลาดในการเก็บข้อมูล จาก Attribute ที่ไม่ต้องการ ซึ่งเกิดจากความสับสนในการตั้งชื่อ Attribute นั้น (Mislabeling of Field) เช่น เราต้องการเก็บค่าของระดับการศึกษาของผู้สมัครเข้าศึกษาต่อ ซึ่งในความเป็นจริงถูกเก็บไว้ใน Attribute ที่ชื่อ “LEVEL_EDU” แต่ในฐานข้อมูลนั้นบังเอิญมี Attribute อีกตัวหนึ่งชื่อ “EDUCATION” ซึ่งเก็บระดับการศึกษาที่ผู้สมัครต้องการเข้าศึกษา ซึ่งถ้าเราไม่ได้ตรวจสอบความสัมพันธ์และความมุ่งหมายที่แท้จริงของแต่ละ Attribute แล้ว ก็อาจเกิดการสับสนโดยเก็บข้อมูลของ Attribute “EDUCATION” ไปแทนก็ได้ ซึ่งเมื่อนำข้อมูลที่ได้ไปทำ การทำเหมืองข้อมูล ผลลัพธ์ที่ได้ ก็จะผิดพลาดด้วย

5) การแปลงข้อมูล (Data Transformation)

ระหว่างขั้นตอนของการแปลงข้อมูลข้อมูลที่ได้กลั่นกรองแล้วจะถูกแปลงให้เป็นรูปแบบของข้อมูลที่พร้อมจะถูกวิเคราะห์ รูปแบบของข้อมูลที่พร้อมจะถูกวิเคราะห์ คือ รูปแบบของข้อมูลที่ไม่มีความขัดแย้ง ถูกจัดระเบียบมาอย่างเรียบร้อยกลั่นกรองมาจากแหล่งข้อมูลภายนอก และภายในขั้นตอนนี้เป็นขั้นตอนที่สำคัญมากเนื่องจากความถูกต้อง และสมบูรณ์ของผลลัพธ์สุดท้ายซึ่งขึ้นอยู่กับว่า นักวิเคราะห์ข้อมูลนั้นตัดสินใจกำหนดโครงสร้างและเสนอลักษณะของ Input อย่างไร ตัวอย่างเช่น หลักการรูปแบบของข้อมูลถูกกำหนด แล้วข้อมูลที่ถูกกลั่นกรองจะเหมาะสมกับรูปแบบเฉพาะสำหรับแต่ละ กรรมวิธีของการทำเหมืองข้อมูล ที่จะถูกใช้การแปลงข้อมูลยังรวมไปถึงการทำ Data Recording และ Data Format Conversion เช่น การแปลงวันที่ เป็นต้น ทางสถิติการทำการแปลงข้อมูลยังมีเทคนิคของ Data Reduction จุดประสงค์เพื่อที่จะลดตัวแปรสำหรับการทำการ Process โดยการนำเอาตัวแปรตั้งแต่ 2 ตัวขึ้นไปมารวมกันแล้วทำการ Process ข้อดีก็คือลดจำนวนของตัวแปรลงและยังสามารถจัดการได้ง่ายขึ้นอีกเทคนิคเรียกว่า Discretization โดยการแปลงตัวแปรแบบ Quantitative ให้เป็นแบบ Categorical โดยการแบ่งค่าของตัวแปรที่จะเป็น Input ให้เป็นช่วง ๆ เช่น การแปลงเงินเดือน อายุ อีกเทคนิคเรียกว่า One of N โดยการแปลงตัวแปรแบบ Categorical ให้เป็น Numeric ตัวอย่างเช่น ชนิดของรถ Ford, Lincoln, Nissan ให้เป็น 100, 010, 001 ปกติแบบนี้มักจะเป็น Input ของพวก Neural Network

6) การปรับแต่งข้อมูล (Data Engineering)

ขั้นตอนก่อนหน้านี้เป็นขั้นตอนของการสร้าง และการตรวจสอบความถูกต้องของข้อมูลที่จะนำมาใช้ แต่ในขั้นตอนนี้ที่เราต้องทำ คือการปรับแต่งฐานข้อมูล ซึ่งในขั้นตอนนี้จะมีปัญหาหลัก ๆ อยู่ 3 ข้อ คือ

6.1) ข้อมูลที่ได้อาจมี Attributes จำนวนมากที่สามารถใช้ประโยชน์ได้แต่ถูกละเลย การเลือกกลุ่มของ Attributes ที่จะใช้เป็นปัญหาที่สำคัญปัญหาหนึ่ง

6.2) ฐานข้อมูลที่ได้อาจมีจำนวนระเบียน (Record) มากเกินไปกว่าที่จะสามารถทำการวิเคราะห์ให้เสร็จลงได้ในเวลาที่เหมาะสม ซึ่งในกรณีนี้เราต้องทำการสุ่มข้อมูลตัวอย่างขึ้นมาใช้แทน

6.3) ข้อมูลบางอย่างอาจใช้ให้เกิดประโยชน์ได้ โดยการนำเสนอในรูปแบบของการวิเคราะห์แบบเฉพาะเจาะจง การทำ Data Engineering นั้นจะมีการทำซ้ำขึ้นมาหลาย ๆ ครั้ง เพื่อทดสอบ การใช้ Attribute ที่แตกต่าง, ขนาดของกลุ่มตัวอย่างที่ต่างกัน เช่น เราจะพยากรณ์อนาคตเมื่อเวลาผ่านไป 1, 2, 3, หรือ 4 เดือน เราอาจพยากรณ์ได้โดยใช้เพียง Attribute เป็นตัวพยากรณ์หรืออาจใช้ข้อมูลทุกอย่างที่เรามีเป็นตัวพยากรณ์ก็ได้ เป็นต้น

7) Visualization

เป็นการนำเสนอข้อมูลในรูปแบบกราฟฟิค การนำเสนอจะสามารถทำได้มากกว่า 2 มิติ ซึ่งจะสร้างความละเอียด ของการนำเสนอ และสร้างความเข้าใจให้มากขึ้น

8) Analysis

หลังจากเลือก Algorithm ที่เหมาะสมกับลักษณะของปัญหาแล้ว เราก็จะนำ Algorithm นั้นมาทำ การวิเคราะห์ข้อมูลในฐานข้อมูลที่เตรียมไว้ ซึ่งในบางครั้งขั้นตอนนี้จะถูกเรียกว่า “Data Mining” ในขณะที่จะเรียกกระบวนการทั้งหมดว่า “Knowledge Discovery in Databases” ผลลัพธ์ที่ได้จากขั้นตอนนี้จะเป็นรูปแบบของความสัมพันธ์ของข้อมูลที่จะนำมาใช้ ในการพยากรณ์ (Prediction) หรือวิเคราะห์ต่อไปนำข้อมูลที่จัดเตรียมไว้มาทำ การทำเหมืองข้อมูล ซึ่งมีการทำงานอยู่ 4 ชนิดด้วยกัน คือ

8.1) Data Segmentation เป็นกระบวนการแบ่ง Database ออกเป็นกลุ่มเพื่อให้ง่ายต่อกาวิเคราะห์ เช่นการแบ่งลูกค้าออก ตามรายได้ เพศ อายุ เป็นต้น

8.2) Predictive Modeling เป็นการสร้างแบบจำลองพยากรณ์ แบ่งเป็น 2 ลักษณะ คือ Classification เป็นการจัดกลุ่มให้กับแต่ละข้อมูลในฐานข้อมูล โดยมีการระบุค่า หรือลักษณะที่เป็นไปได้ของข้อมูล ภายในแต่ละกลุ่ม เช่น การแบ่งประเภทของกลุ่มของผู้ป่วยตามผลของการใช้ยารักษา เพื่อระบุรูปแบบการรักษาให้กับผู้ป่วยใหม่ ที่เข้ารับการรักษา เป็นต้น

8.3) Value Prediction เป็นการพยากรณ์ค่าที่เป็นไปได้ หรือการกระจายของค่าที่เป็นไปได้ของตัวแปรใด ๆ ในกลุ่มข้อมูล การพยากรณ์ค่าที่เป็นตัวเลข เช่น การพยากรณ์ภาษีที่จะเก็บได้ในปี เป็นต้น

8.4) Link Analysis (Associations) เป็นการหาความสัมพันธ์ของข้อมูลภายในกลุ่มข้อมูลเพื่อใช้ลักษณะของข้อมูลหนึ่ง ๆ ในการบอกลักษณะที่จะเกิดขึ้นกับข้อมูลอีกตัวหนึ่ง ซึ่งอาจจะเป็นการหาความสัมพันธ์ของข้อมูลในกลุ่มเดียวกัน เช่น การระบุว่าในกลุ่มของลูกค้าที่ซื้อนมนั้น จะมีลูกค้า 64% ที่ซื้อขนมปังด้วย หรืออาจจะเป็นการหาความสัมพันธ์ของ ตัวแปรระหว่างกลุ่มข้อมูลก็ได้ เช่น ในทุก ๆ ครั้งที่ดัชนีของตลาดหุ้นหนึ่งลดลง 5% ดัชนีของตลาดหุ้นอื่นจะเพิ่มขึ้น 13% ภายในช่วง 2-6 เดือนหลังจากนั้น เป็นต้น ซึ่งลักษณะของการหาความสัมพันธ์นั้นอาจแบ่งได้เป็น 3 กลุ่ม คือ การหาความสัมพันธ์ระหว่างข้อมูล (Association Discovery) การหาความสัมพันธ์ในลักษณะที่เป็นลำดับของข้อมูล (Sequential Pattern Discovery) และการหาความสัมพันธ์ของข้อมูลกับช่วงเวลาใด ๆ (Similar Time Sequence Discovery)

8.5) Deviation Detection เป็นเทคนิคที่ใช้ในการแสดงลักษณะของข้อมูลที่ผิดปกติ หรือผิดไปจากที่คาดไว้ โดยมีการแสดงผลอยู่ในลักษณะที่สามารถทำความเข้าใจและแปลความหมายได้ง่าย เช่น การใช้กราฟ เป็นต้น

9) Interprete

หลังจากที่การสร้าง Model แล้วจำ เป็นต้องตรวจสอบผลลัพธ์และตีความหมาย ความถูกต้องที่ตรวจออกมาได้นั้น เป็นชุดตัวอย่างที่ส่งเข้าไป ดังนั้นผลลัพธ์ที่ได้อาจ มีความปรวนแปรหากมีการนำ ไปใช้กับข้อมูลอื่น ๆ

10) Presentation

เป็นการแสดงผลการวิเคราะห์โดยอาศัยเครื่องมือที่มีความสามารถและเข้าใจง่าย การแสดงผลอาจจะอยู่ในรูปแบบของ กราฟ รายงาน ตาราง แผนที่หลายมิติ เป็นต้น

โดยเทคนิคการวิเคราะห์ข้อมูลด้วย Datamining สามารถแบ่งได้ 2 ประเภทหลัก (เอกสิทธิ์ พัชรวงศ์ศักดา, 2557) ดังนี้

1) การเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) ไม่จำเป็นต้องมีค่าเป้าหมายของแต่ละข้อมูลตัวอย่างในระหว่างการเรียนรู้ เป็นการเรียนรู้ได้ด้วยตนเอง โดยไม่ต้องมีค่าเป้าหมายของแต่ละข้อมูล ซึ่งวิธีการนี้จะเป็นการใส่ข้อมูลต่าง ๆ และกำหนดสิ่งที่ต้องการจากข้อมูลเหล่านี้ โดย

มหาวิทยาลัยรังสิต Rangsit University

ให้ระบบวิเคราะห์จากการจำแนกและสร้างแบบแผนจากข้อมูลที่ได้รับมา โดยที่นิยมใช้จะมี 3 อัลกอริทึม คือ

1.1) Association Algorithm หรือกฎความสัมพันธ์ (Association Rule) เป็นการค้นหาความสัมพันธ์ของข้อมูลที่ซ่อนอยู่ในข้อมูล โดยความสัมพันธ์สามารถอธิบายในรูปแบบกฎ (Rules) หรือรูปแบบของการเกิดร่วมกันของข้อมูลที่พบบ่อย ๆ (Frequency Pattern) (Dietrich, Heller and Yang, 2015) การใช้งานกฎความสัมพันธ์โดยทั่วไปจะใช้วิเคราะห์หาความสัมพันธ์ของข้อมูลที่การจัดเก็บข้อมูลที่เปลี่ยนแปลง (Transaction Database) (Olson & Lauhoff, 2019) โดยแต่ละเกณฑ์มีความหมายและคำนวณได้ดังนี้ค่าสนับสนุน (Support) คือ ค่าที่ใช้บ่งบอกถึงเหตุการณ์ A และ B มีความถี่ในการเกิดขึ้นมากน้อยเพียงใด คำนวณได้จากสมการ

$$\text{ค่าสนับสนุน}(A \rightarrow B) = \frac{\text{ค่าความถี่ของเหตุการณ์ที่มีทั้ง } A \text{ และ } B}{\text{จำนวนรายการทั้งหมด}} \quad (2\text{-}4)$$

1.1.1) ค่าความเชื่อมั่น (Confidence) คือ ค่าที่ใช้บ่งบอกความเชื่อมั่นของการเกิดเหตุการณ์ A และ B ว่ามีส่วนร่วมด้วยมากน้อยเพียงใด คำนวณได้จากสมการ

$$\text{ค่าความเชื่อมั่น}(A \rightarrow B) = \frac{\text{ค่าสนับสนุน}(A \rightarrow B)}{\text{ค่าสนับสนุน}(A)} \quad (2\text{-}5)$$

โดยที่ค่าสนับสนุน $(A \rightarrow B)$ สามารถเขียนแทนด้วยสมการ 1.1.1 และค่าสนับสนุน (A) คำนวณได้จากค่าความถี่ของเหตุการณ์ที่มี A หารด้วยจำนวนรายการทั้งหมด ซึ่งเขียนเป็นสมการคำนวณความเชื่อมั่น $(A \rightarrow B)$ ใหม่ได้ดังนี้

$$\text{ค่าความเชื่อมั่น}(A \rightarrow B) = \frac{\frac{\text{ค่าความถี่ของเหตุการณ์ที่มีทั้ง } \boldsymbol{A} \text{ และ } \boldsymbol{B}}{\text{จำนวนรายการทั้งหมด}}}{\frac{\text{ค่าความถี่ของเหตุการณ์ } \boldsymbol{A}}{\text{จำนวนรายการทั้งหมด}}} \quad (2\text{-}6)$$

จากสมการข้างต้นสามารถนำมาเขียนใหม่ได้ดังนี้

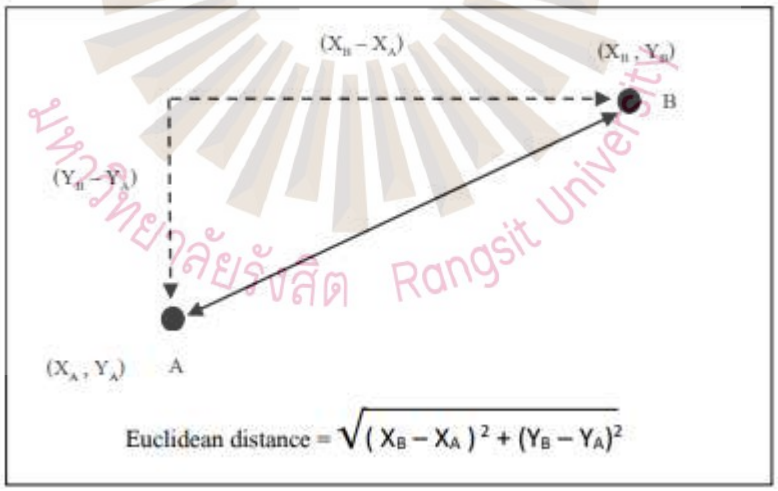
$$\text{ค่าความเชื่อมั่น}(A \rightarrow B) = \frac{\text{จำนวนรายการที่มีทั้ง } A \text{ และ } B}{\text{จำนวนรายการที่มี } A} \quad (2\text{-}7)$$

1.1.2) ค่าลิฟท์ (Lift) คือ ค่าที่ใช้บ่งบอกความสัมพันธ์ของเหตุการณ์ A และ B ว่ามีความสัมพันธ์กันมากน้อยเพียงใด โดยถ้าค่า Lift ที่คำนวณได้เท่ากับ 1 แสดงว่าเหตุการณ์ A และ B ไม่สัมพันธ์กัน แต่ถ้ามากกว่า 1 แสดงว่าเหตุการณ์ A และ B มีความสัมพันธ์กัน ซึ่งคำนวณได้โดยสมการดังนี้

$$\text{ค่าลิฟท์}(A \rightarrow B) = \frac{\text{ค่าสนับสนุน}(A \rightarrow B)}{\text{ค่าสนับสนุน}\ (A) \times \text{ค่าสนับสนุน}\ (B)} \qquad (2\text{-}8)$$

1.2) การแบ่งกลุ่ม (Clustering)

เทคนิคนี้จะเป็นการแบ่งกลุ่มข้อมูลที่มีลักษณะคล้ายคลึงกันให้อยู่รวมกัน โดยใช้วิธีการวัดความคล้ายคลึง (Similarity) หรือการวัดระยะห่าง (Distance) ระหว่างข้อมูลแต่ละตัวเพื่อใช้ในการจำแนกกลุ่มต่อไป ค่าการคำนวณวัดระยะห่างมีหลายประเภทเช่น Superman Rank Correlation, Person Correlation, Euclidean แต่ที่นิยมใช้กัน คือ ระยะห่างยูคลิเดียน (Euclidean Distance) ตามรูปที่ 2.2 (เอกสิทธิ์ พัชรวงศ์ศักดา, 2557) ส่วนเทคนิคที่นิยมในการแบ่งกลุ่มส่วนใหญ่ คือ DBSCAN–Density Based Spatial Clustering of Application with Noise และ K – Mean Agglomerative Clustering



รูปที่ 2.2 รูปแสดงระยะห่างยูคลิเดียน

1.3) Time Series Algorithm

Time Series Algorithm ยอดขายในอนาคตโดยคาดว่าจะมีลักษณะ เช่นเดียวกับยอดขายในปัจจุบันหรืออนาคต ยอดขายหรืออุปสงค์ในความเป็นจริงได้รับอิทธิพลจากแนวโน้ม

(Trend) ฤดูกาล (Seasonal) วัฏจักร (Cycle) และเหตุการณ์ผิดปกติ (Irregular Variation) การใช้อนุกรมเวลามี 3 วิธี คือ

1.3.1) การพยากรณ์อย่างง่าย (Naïve Forecast) เป็นการพยากรณ์ว่ายอดขายในอนาคตจะเท่ากับยอดขายปัจจุบัน

1.3.2) การหาค่าเฉลี่ยเคลื่อนที่ (Moving Average) เป็นการหาค่าเฉลี่ยของยอดขายโดยใช้ข้อมูล 3 ช่วงเวลาขึ้นไปในการคำนวณ เมื่อเวลาผ่านไป 1 ช่วงก็ใช้ข้อมูลใหม่มาเฉลี่ยแทนข้อมูลในช่วงเวลาไกลที่สุดซึ่งจะถูกตัดทิ้งไป

1.3.3) การปรับเรียบด้วยเอ็กซ์โปแนนเชียล (Exponential Smoothing) เป็นการหาค่าเฉลี่ยเคลื่อนที่แบบถ่วงน้ำหนักที่จัดค่าพยากรณ์ออกมาในรูปการใช้สมการคำนวณ ซึ่งจะใช้ค่าข้อมูลเริ่มต้น ค่าเดียวและถ่วงน้ำหนักโดยใช้สัมประสิทธิ์เชิงเรียบ ที่มีค่าอยู่ระหว่าง 0 ถึง 1.00

2) การเรียนรู้แบบมีผู้สอน (Supervised Learning) เทคนิคนี้จะนำข้อมูลในอดีตมาวิเคราะห์เพื่อสร้างโมเดล (Model) เพื่อพยากรณ์สิ่งที่จะเกิดขึ้น โดยใช้หลักการทางสถิติคณิตศาสตร์ เทคนิคนี้จะเป็นการจำแนกข้อมูล (Classification) และการประมาณค่าข้อมูล (Regression) ความแตกต่างของ 2 กระบวนการทำงาน คือ การค้นหาคำตอบในชุดกลุ่มข้อมูลที่ซ่อนอยู่ การพยากรณ์ข้อมูลแบบการวิเคราะห์ถดถอยจะเป็นลักษณะค่าที่เป็นตัวเลข ส่วนการจำแนกข้อมูลจะมีการทำนายค่าที่เป็นตัวอักษร เป็นต้น โดยการวิจัยนี้จะใช้เทคนิคการจำแนกประเภทข้อมูล (Classification)

2.1) การจำแนกประเภทข้อมูล (Classification)

การจำแนกประเภทข้อมูล คือ การนำข้อมูลที่มีในอดีตทำการการเรียนรู้และจดจำด้วยการสร้างแบบจำลองโมเดล (Model) ขึ้น เพื่อที่จะนำข้อมูลชุดใหม่มาพยากรณ์ค้นหาคำตอบ โดยกำหนดค่าข้อความที่เป็นหัวเรื่อง (Lebel) เพื่อให้ระบบสามารถจำแนกประเภทของข้อมูลว่ามีลักษณะเป็นอย่างไร ในการประยุกต์ใช้งานสามารถแบ่งออกเป็น 3 ขั้นตอน ดังนี้

1) ขั้นตอนการสร้างโมเดล เป็นขั้นตอนในการจัดเตรียมข้อมูล เพื่อสร้างข้อมูลให้เหมาะสมกับการที่จะนำไปใช้เพื่อทำการทดสอบ โดยทำการ Training จะได้ แอดทริบิวต์มา 2

ส่วน คือ แอทริบิวต์ทั่วไป กับ แอดทริบิวต์ที่น่าสนใจ และแต่ละแอดทริบิวต์จะต้องมีความสัมพันธ์กับแต่ละคำตอบ จึงจะทำให้โมเดลมีความน่าเชื่อถือและมีประสิทธิภาพ

2) ขั้นตอนการทดสอบความถูกต้องหรือประสิทธิภาพของโมเดลนี้จะสามารถใช้เทคนิคต่าง ๆ มาทำการเปรียบเทียบกันโดยในแต่ล่ะโมเดล จะต้องหาค่าความถูกต้อง (Accuracy) โดยจะพิจารณารวม Class ทุก Class เพื่อให้เกิดความสัมพันธ์กันของข้อมูลและได้คำตอบที่ถูกต้อง

3) ขั้นตอนการนำโมเดลไปใช้งานเพื่อพยากรณ์ทำนายข้อมูลใหม่ หลังจากการวัดประสิทธิภาพแต่ละโมเดลและได้คำตอบที่น่าเชื่อถือและถูกต้อง เราสามารถนำโมเดลนี้ไปทำนายข้อมูลที่เข้ามาใหม่เพื่อให้ได้คำตอบที่ถูกต้องแม่นยำมากยิ่งขึ้น เมื่อรู้ถึงขั้นตอนการจำแนกประเภทข้อมูลว่ามีลักษณะแบบใดนั้นขั้นตอนต่อไปคือการวัดประสิทธิภาพ ดังนี้

3.1) ค่าความแม่นยำ (Accuracy) การวัดค่าความแม่นยำ (Accuracy) Subasi, Ahmed and Alickovic (2018) ได้กล่าวไว้ว่า เพื่อการเปรียบเทียบแบบจำลองการพยากรณ์เพื่อจะหาแบบจำลองที่เหมาะสมนั้น การหาค่าประสิทธิภาพในการทำงานได้ดังสมการนี้

$$Accuracy = \frac{TP+TN}{TN+TP+FN+FP} \; x \; 100\% \qquad (2\text{-}9)$$

TP คือ ค่าที่ทำนายถูกต้องเชิงบวก

TN คือ ค่าที่ทำนายถูกต้องเชิงลบ

FN คือ ค่าที่ทำนายผิดพลาดเชิงบวก

FP คือ ค่าที่ทำนายผิดพลาดเชิงลบ

3.2) ค่าความเที่ยงตรง (Precision) เป็นการวัดค่าการพยากรณ์ว่าจริงได้ถูกต้องของจำลองการพยากรณ์สามารถหาได้ดังสมการนี้ (Jaafaria, Zennerb and Phame, 2018)

$$Precision = \frac{TP}{TP+FP} \qquad (2\text{-}10)$$

TP คือ ค่าที่ทำนายถูกต้องเชิงบวก

FP คือ ค่าที่ทำนายผิดพลาดเชิงลบ

3.3) ค่าความระลึก (Recall) เป็นการวัดค่าพยากรณ์ได้ว่าเป็นจริงของแบบจำลองการพยากรณ์อัตราส่วนเท่าไรของจริงทั้งหมด สามารถหาได้ดังสมการนี้ (Jaafaria, Zennerb and Phame, 2018)

$$Recall = \frac{TP}{TP+FN} \quad (2\text{-}11)$$

TP คือ ค่าที่ทำนายถูกต้องเชิงบวก

FN คือ ค่าที่ทำนายผิดพลาดเชิงบวก

2.2) อัลกอริทึมที่ใช้การจำแนกประเภทของข้อมูล

การแยกประเภทของข้อมูลเป็นเทคนิคของการทำเหมืองข้อมูล (Data Mining) โดยนำข้อมูลที่เป็นอดีตมาเป็นแบบจำลองเพื่อพยากรณ์ข้อมูลใหม่ในอนาคตด้วยการเรียนรู้ของเครื่อง (Machine Learning) โดยมีอัลกอริทึมเป็นเครื่องมือในการช่วยวิเคราะห์มีรายละเอียดดังนี้

1) เทคนิคการจำแนกข้อมูลโดยต้นไม้ตัดสินใจ (Decision Tree) ในช่วงปลายของยุค 1970 ได้มีนักวิจัยทางด้านการเรียนรู้ของเครื่อง (Machine Learning) คือ J. Ross Quinlan ได้คิดค้นอัลกอริทึมสำหรับสร้างต้นไม้ตัดสินใจที่มีชื่อว่า ID3 (Iterative Dichotomiser) ต่อมา เขาได้พัฒนาต่อยอด ID3 ไปเป็น C4.5 ซึ่งได้กลายมาเป็นอัลกอริทึมพื้นฐานที่ใช้สำหรับเปรียบเทียบ ประสิทธิภาพการทำงานของอัลกอริทึมต่าง ๆ ทางด้านการเรียนรู้แบบมีผู้สอน (Supervised Learning) ID3 และ C4.5 ได้ทำการประยุกต์ใช้วิธีการเชิงละโมภ (Greedy Approach) ในการสร้างต้นไม้ภายใต้ วิธีการแบบ “Top-Down Recursive Divide and Conquer” โดยทำการพิจารณาชุดข้อมูลสำหรับเรียนรู้ (Training Data, เซตของเรคคอร์ดของข้อมูลที่แต่ละเรคคอร์ดจะประกอบไปด้วยเซตของแอทริบิวต่าง ๆ และแอทริบิวที่บ่งบอกถึงหมวดหมู่ของข้อมูลเรคคอร์ดนั้น ๆ) ด้วยการแบ่งข้อมูลออกเป็นส่วนย่อย ๆ ในระหว่างกระบวนการสร้างต้นไม้ อนุวัฒน์ เปพาทย์ วงกต ศรีอุไร และณัฏฐ์ ดิษเจริญ (2565) กล่าวว่า ต้นไม้ตัดสินใจ เป็นเทคนิคหนึ่งที่ให้ผลลัพธ์และอธิบายความสัมพันธ์ได้ง่าย โดยใช้หลักการของ Information Gain(IG) หรือ Entropy Reduction เพื่อจำแนกโหนด(Node) โดยคัดเลือกคุณลักษณะที่

มีความสัมพันธ์กับคลาสมากที่สุดมาเป็นราก (Root Node) จากนั้นหาคุณลักษณะไปเรื่อย ๆ โดยในโหนดถัดไปจะมีค่า Gain ลดหลั่นกันไป ซึ่งแต่ละโหนดจะแสดงถึงการตัดสินใจบนชุดข้อมูลของคุณสมบัติต่าง ๆ ของกิ่งไม้ โดยข้อมูลชั้นล่างสุดของต้นไม้ตัดสินใจจะแสดงถึงกลุ่มของข้อมูล (Class) ในการหาความสัมพันธ์ของคุณลักษณะจะใช้ค่า IG ซึ่งคำนวณได้จาก

$$IG(parent, child) = Entropy(parent) - [p(c_1) \times Entropy(c_1) + p(c_2) \times Entropy(c_2) + \cdots]$$

โดยที่

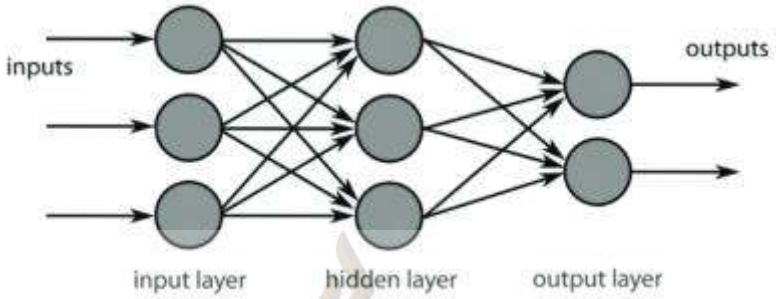
$Entropy(c)$ คือ $-p(c)\log p(c)$ (2-12)

$p(c_i)$ คือ ค่าความน่าจะเป็นของค่า c_i

c คือ คลาสเป้าหมาย (Class)

2) เทคนิคการจำแนกข้อมูลโดยเครือข่ายประสาทเทียม (Neural Network) ในปี พ.ศ. 2486 ได้ถือว่าเป็นปีที่กำหนดสาขาเครือข่ายประสาทเทียมในวงการวิทยาศาสตร์ (Garson, 1998) ที่สามารถทำงานร่วมกับคอมพิวเตอร์ได้ เครือข่ายประสาทเทียมเป็นโมเดลทางคณิตศาสตร์ที่จำลองการทำงานของสมองมนุษย์ โดยมีเส้นประสาทเชื่อมต่อกันหลายจุด และโยงใยจนเป็นระบบประสาทเทียมขึ้นมา การทำงานของเครือข่ายประสาทเทียมจะเป็นการนำข้อมูลเข้าไปกำหนดน้ำหนัก (Weight) ต่างละโหนด และทำการปรับเปลี่ยนค่าน้ำหนักด้วยการนำไปคูณกับข้อมูลที่นำเข้ามาของแต่ละขา ผลลัพธ์ที่ได้ของแต่ละขาข้อมูลนั้นจะนำมาเปรียบเทียบเกณฑ์การแบ่งค่า (Threshold) ที่กำหนดไว้หากมีค่ามากกว่าก็จะส่งข้อมูลผลลัพธ์ไปเป็นขาออกเพื่อนำไปยังเส้นประสาทอื่น และเปรียบเทียบจำแนกเข้าตามกลุ่ม และถ้ามีค่าน้อยกว่าการแบ่งค่าจะไม่ส่งผลลัพธ์ใดออกมา ข้อสำคัญคือต้องรู้ค่าปริมาณน้ำหนักของชุดข้อมูล และเกณฑ์การแบ่งค่าที่ต้องการเพื่อให้คอมพิวเตอร์เรียนรู้จดจำ ซึ่งค่าดังกล่าวเป็นค่าไม่แน่นอน แต่สามารถกำหนดคอมพิวเตอร์ให้ปรับเปลี่ยนค่าเหล่านั้น โดยการสอนให้คอมพิวเตอร์ทำการเรียนรู้และจดจำรูปแบบ การจดจำนี้นี้เรียกว่า อัลกอริทึมการแพร่ย้อนกลับ (Back Propagation) ซึ่ง นคร ละลอกน้ำ (2562) ได้กล่าวไว้ว่าสำหรับโครงสร้างของเครือข่ายประสาทเทียมจะประกอบด้วย Input และ Output เช่นกัน โดยแบ่งเป็นชั้นหรือ Layer ซึ่งจะมีชั้นคั่นตรงกลางคือ Hidden Layer โดยโครงสร้างเครือข่ายประสาทเทียมจะมีหน่วยย่อย เรียกว่า Perceptron ซึ่งเทียบเท่าได้กับเซลล์สมองของมนุษย์หนึ่งเซลล์โดยหลังจากใส่รูปแบบข้อมูลสำหรับการเรียนรู้ให้กับเครือข่ายในแต่ละครั้ง ค่าที่ได้รับจากเครือข่ายจะนำไปเปรียบเทียบกับผลที่คาดหวังและทำการคำนวณหาค่าความผิดพลาด ซึ่งค่าความผิดพลาดจะถูก

ส่งกลับไปสู่เครือข่ายเพื่อใช้ในการปรับปรุงค่าน้ำหนักคะแนนต่อไป อัลกอริทึมนี้เหมาะสมสำหรับการพยากรณ์อนาคต หรือการแยกกลุ่มข้อมูลที่เป็นชนิด ข้อมูลที่เป็นคลาสและข้อมูลที่เป็นตัวเลขได้ดี



รูปที่ 2.3 รูปแสดงโครงสร้างของเครือข่ายประสาทเทียม

ที่มา: Wikimedai Commons, 2020

3) เทคนิคการจำแนกข้อมูลโดยไนอีฟเบย์ (Naïve Bayes) นิเวศ จิระวิชิตชัย (2560) กล่าวว่า เป็นอัลกอริทึมที่เข้าใจง่ายและสามารถแบ่งประเภทข้อมูลได้อย่างมีประสิทธิภาพ เหมาะสมสำหรับการพยากรณ์คลาสของชุดข้อมูล และยังสามารถทำงานได้รวดเร็วหากทำการจัดกลุ่มของตัวแปรอินพุต โดยการวิเคราะห์ เพื่อหาความสัมพันธ์ระหว่างตัวแปรและนำไปใช้ในการสร้างเงื่อนไขความน่าจะเป็นสำหรับแต่ละความสัมพันธ์ กำหนดให้ความน่าจะเป็นของข้อมูลที่เป็นกลุ่ม V สำหรับข้อมูลที่มีแอททริบิวต์ทั้งหมด n ตัว มีรูปแบบสมการดังรูปด้านล่าง

$X = \{a_1, a_2, a_3, a_n\}$ หรือใช้สัญลักษณ์ว่า $P(a_1, a_2, \ldots, a_3 | v_j)$ (2-13)

คือ $P(a_1, a_2, \ldots, a_3 | v_j) = \prod_{i=1}^{n} P(a_i | v_j)$

$P(a_i | v_j)$

$P(v_j)$

V_{NB}

โดยที่ $\prod$ หมายถึง ผลคูณของค่า $P(a_i|V_j)$ ทั้งหมด $i = 1,2,3, \ldots, n$ และ $j = 1,2,3, \ldots, n$ ทำการหาค่าความน่าจะเป็นของคำที่พบในแต่ละกลุ่ม โดยนำค่า $P(a_1, a_2, \ldots, a_3 | v_j)$ จากสมการมาคูณกับค่าความน่าจะเป็นกลุ่มนั้น ๆ คือ $P(v_j)$ ได้เท่ากับ V_{NB} นำค่าที่ได้ มาเปรียบเทียบกับกัน กลุ่มที่มีค่าความน่าจะเป็นสูงสุด คือ คำตอบ ดังนั้นจะได้ว่า วิธีการจำแนกประเภทแบบเบย์อย่างง่าย ดังสมการ

$$V_{NB} = \arg max_{v_j \in V} P(V_j) x \prod_{i=1}^{n} P(a_i|V_j) \quad (2\text{-}14)$$

4) เทคนิควิเคราะห์ความแปรปรวน (Analysis of Variance: ANOVA) คือ การวิเคราะห์ข้อมูลที่ใช้การทดสอบสมมุติฐานของความแตกต่างของค่าเฉลี่ยของข้อมูลมากกว่า 2 กลุ่ม โดยทำการทดสอบสมมุติฐานเพียงครั้งเดียว หลักการของเทคนิควิเคราะห์ความแปรปรวนจะแบ่งความแปรปรวนของข้อมูลออกไปตามสาเหตุ ที่ทำให้ข้อมูลแตกต่างกัน คือความแปรปรวนภายในกลุ่มและระหว่างกลุ่ม โดยที่ (ปัทมา สมัครไทย, 2562)

ตารางที่ 2.1 ตารางลักษณะของข้อมูลที่ใช้ในการวิเคราะห์ความแปรปรวนทางเดียว

ความแปรปรวนทั้งหมด = ความแปรปรวนภายในกลุ่ม + ความแปรปรวนระหว่างกลุ่ม

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

เมื่อ X_{ij} แทนข้อมูลของทรีทเมนต์ที่ i หน่วยทดลองที่ j

i = 1,2,3,…,k และ j = 1,2,3,…n_i

T_i แทนผลรวมของข้อมูลทรีทเมนต์ที่ i

T แทนผลรวมข้อมูลทั้งหมด

$\bar{x}_i$ แทนค่าเฉลี่ยของข้อมูลทรีทเมนต์ที่ i

$\bar{x}$ แทนค่าเฉลี่ยของข้อมูลทั้งหมด

k แทนจำนวนทรีทเมนต์

n แทนจำนวนข้อมูลทั้งหมด เท่ากับ $n_1 + n_{2+} n_{3+\ldots+} n_k$

มหาวิทยาลัยรังสิต Rangsit University

4.1) ความแปรปรวนระหว่างกลุ่ม (Between Groups Sum of Square) เป็นการพิจารณาความแปรปรวนที่เกิดจากค่าเฉลี่ยของตัวอย่างในแต่ล่ะกลุ่มที่แตกต่างจากค่าเฉลี่ยรวม โดยที่

$$SSB = \sum_{i=1}^{k} n_i \, (x_i - \bar{x})^2 \qquad (2\text{-}15)$$

4.2) ความแปรปรวนภายในกลุ่ม (Within Group Sum of Square) เป็นการพิจารณาของความแปรปรวนที่เกิดขึ้นภายในกลุ่มแต่ล่ะกลุ่ม ซึ่งไม่ทราบถึงสาเหตุความแปรปรวนที่เกิดขึ้น โดยที่

$$SSE = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 \qquad (2\text{-}16)$$

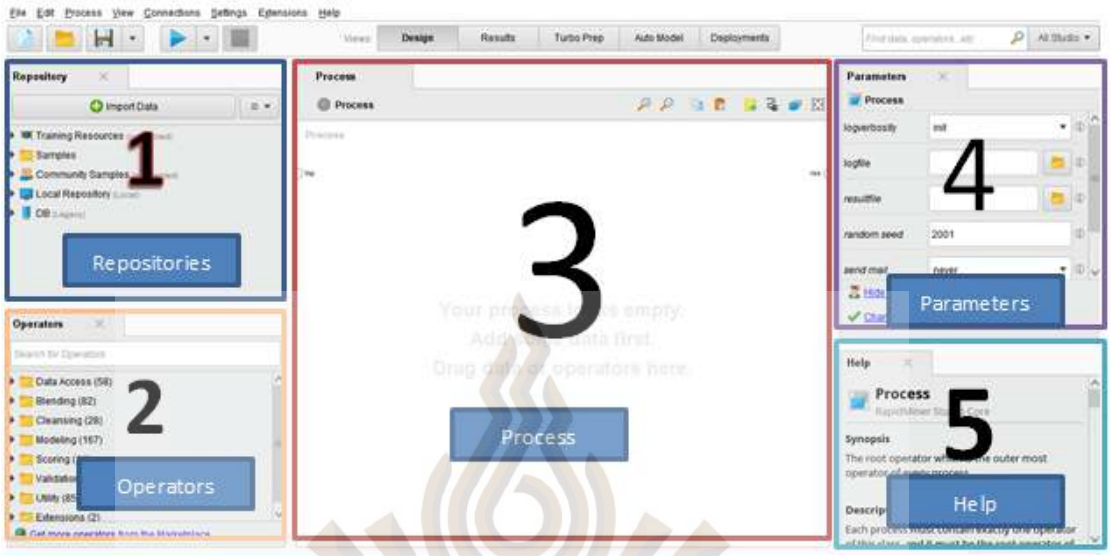
4.3) ความแปรปรวนรวม (Total Sum of Square) เป็นการพิจารณาความแปรปรวนที่เกิดจากค่าการสังเกตแต่ละค่าแตกต่างจากค่าเฉลี่ยรวม โดยที่

$$SST = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 \text{ และ } SST = SSB + SSE \qquad (2\text{-}17)$$

2.1.3 Software RapidMiner Studio

คือ ซอฟต์แวร์ Data Science ใช้สำหรับการเตรียมข้อมูล การเรียนรู้เครื่อง การเรียนรู้ลึก การทำเหมืองข้อความ และการวิเคราะห์การทำนาย (Predictive Analysis) เป็นซอฟต์แวร์ที่ช่วยในการจัดส่งข้อมูล และลดข้อผิดพลาดจนแทบจะไม่จำเป็นต้องเขียนโค้ดเพิ่ม แต่ที่ทำให้เป็นเครื่องมือที่เหล่า Data Scientist นิยมเลือกใช้เป็นเพราะว่าตัว Rapidminer มีขั้นตอนพร้อมสำหรับการทำ Data Mining (ชุดข้อมูล) และ Machine Learning ซึ่งรวมไปถึงการโหลดและการแปลงข้อมูล (ETL) การประมวลผลล่วงหน้าและการวาดภาพจากข้อมูล การวิเคราะห์เชิงพยากรณ์และการสร้างแบบจำลองทางสถิติ การประเมินผลและการปรับใช้ต่าง ๆ ล้วนเป็นสิ่งที่เหล่า Data Scientist จำเป็นต้องทำในการเข้าใจข้อมูลมากขึ้น RapidMiner ในตอนแรกใช้ชื่อบริษัท Rapid-I ก่อน และตั้งขึ้นเมื่อปี 2006 ในช่วงแรกบริษัทตั้งอยู่ที่ ประเทศเยอรมันนี ปี 2013 ได้เปลี่ยนชื่อบริษัทเป็น RapidMiner หลังจากได้รับเงินลงทุนจำนวน 5 ล้านเหรียญและย้ายบริษัท มาอยู่ที่บอสตันประเทศสหรัฐอเมริกาผลิตภัณฑ์หลักของบริษัท คือ RapidMiner Studio สามารถ Download Software ได้ที่ https://rapidminer.com

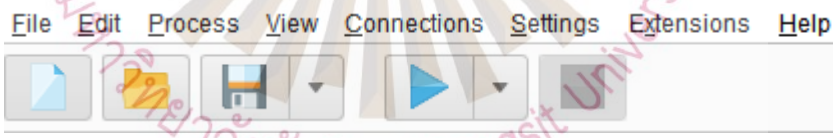
1) องค์ประกอบของโปรแกรม RapidMiner Studio 9 ดังรูปที่ 2.4



รูปที่ 2.4 องค์ประกอบของโปรแกรม RapidMiner Studio 9

ที่มา: นฤชล โรจนบุรานนท์, 2562

2) เมนูใน RapidMiner Studio 9 ดังรูปที่ 2.5



รูปที่ 2.5 เมนูใน RapidMiner Studio 9

ที่มา: นฤชล โรจนบุรานนท์, 2562

ใช้สร้าง Process ใหม่

โหลด Process เดิม

บันทึก Process เดิม หรือสามารถเลือกให้บันทึก Process เป็นชื่อใหม่

สั่งให้ Process ทำงาน Locally หรือ สามารถเลือกสั่งให้ Process ทำงาน Background

สั่งให้ Process หยุดทำงาน

3) Repository ดังรูปที่ 2.6

4) เป็นที่เก็บข้อมูลและ Process เพื่อใช้งานใน RapidMiner

5) ทำให้ไม่ต้องโหลดข้อมูลจาก File ใหม่ทุกครั้ง

รูปที่ 2.6 เครื่องมือ Repository

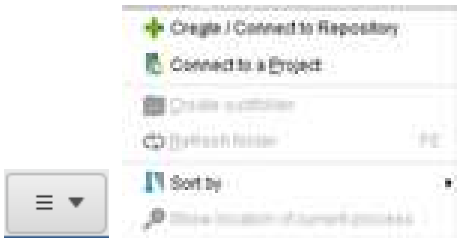
ที่มา: นฤชล โรจนบุรานนท์, 2562

6) องค์ประกอบในส่วน Repository

7) ส่วนที่ 1

8) Import Data เป็นเข้ามาเก็บไว้และใช้งานได้ตลอด ถ้าข้อมูลใน File ไม่เปลี่ยนแปลง ก็ไม่ต้องทำการ Update แต่ถ้าข้อมูลมีการ Update ต้องทำการโหลดข้อมูลใหม่

มหาวิทยาลัยรังสิต Rangsit University



9) เป็นการเลือกเมนเพิ่มเติมโดยสามารถสร้าง Repository ใหม่ สามารถเชื่อมต่อกับ Project ภายนอกได้ หรือสามารถสร้าง Folder ใหม่ได้

10) ส่วนที่ 2

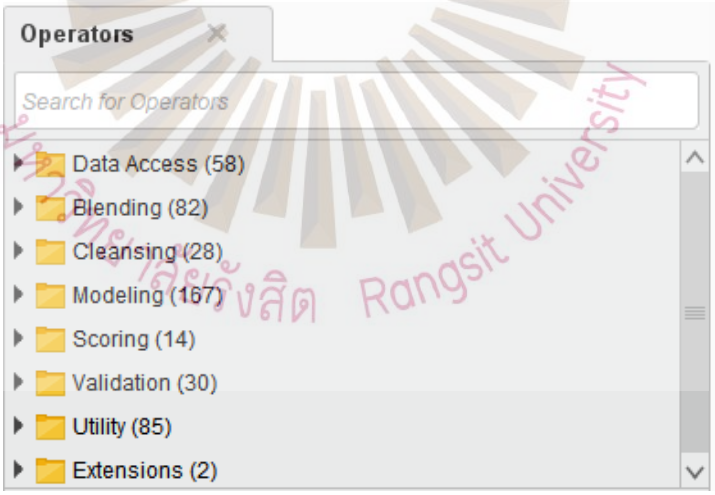
11) Training Resources เป็นการ Train Online ในหัวข้อที่เราสนใจ

12) Samples เป็นตัวอย่างข้อมูลและ Processes ต่าง ๆ

13) Community Samples เป็นการดึงข้อมูลตัวอย่างหรือ Processes ใน Community

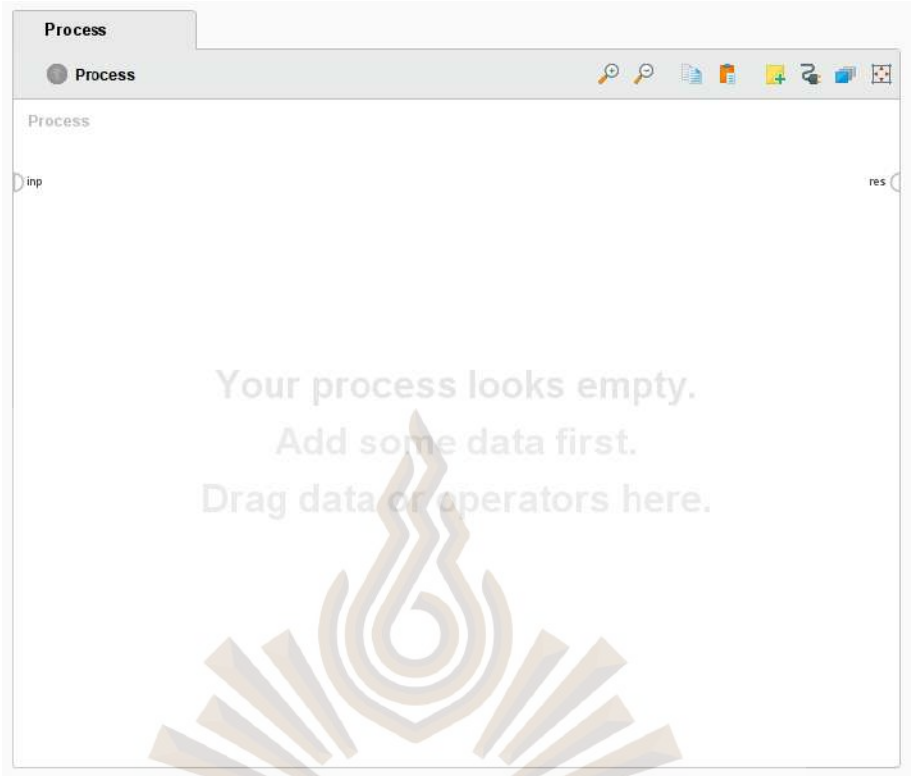
14) Local Repository เป็นข้อมูลที่เก็บอยู่ในแต่ละ Repository

15) Operators ซึ่งเป็นการอำนวยความสะดวกโดยการจัดหาหน้าที่การทำงาน โดยจะเป็นลักษณะ Block Diagram ทำให้ง่ายต่อการใช้งาน เช่น เมื่อต้องการจะอ่านไฟล์ Excel ก็สามารถเรียกใช้ตัวดำเนินการที่ชื่อ “Read Excel” เพื่อช่วยในการอ่านไฟล์ Excel ที่ต้องการเข้ามาในระบบงานของเรา จะเห็นว่าซอฟต์แวร์มีตัวดำเนินการให้ใช้งานมากมาย ดังรูปที่ 2.7



รูปที่ 2.7 เครื่องมือ Operators
ที่มา: นฤชล โรจนบุรานนท์, 2562

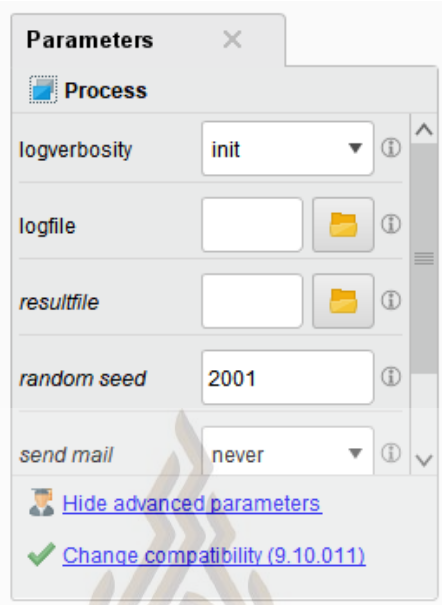
16) Process ซึ่งเป็นส่วนสำคัญมากในการออกแบบการไหลของระบบงาน โดยการลากตัว Operators ที่ต้องการมาวางแล้วทำการเชื่อมต่อกัน จนได้ระบบงานที่ต้องการ ดังรูปที่ 2.8



รูปที่ 2.8 หน้าจอ Process

ที่มา: นฤชล โรจนบุรานนท์, 2562

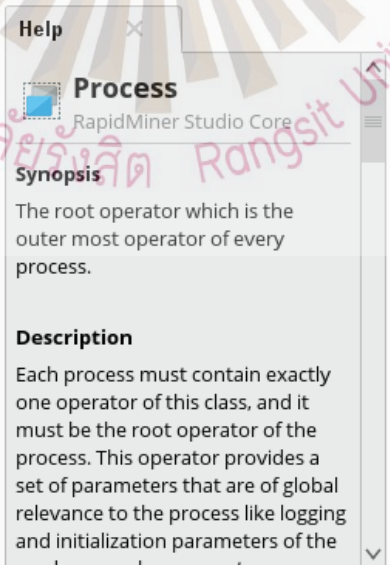
17) Parameters เมื่อมีการใช้ Operators บางครั้งอาจจะต้องกำหนดรายละเอียดเพิ่มเติมเพื่อต้องการให้ตัว Operators สามารถทำงานตามที่ผู้ใช้ต้องการ เช่น การอ่านไฟล์ Excel เข้ามาในระบบงาน ก็ต้องระบุชื่อไฟล์ Excel ซึ่งจะต้องระบุในส่วนของ Parameters ดังกล่าว ดังรูปที่ 2.9



รูปที่ 2.9 หน้าจอ Parameters

ที่มา: นฤชล โรจนบุรานนท์, 2562

18) Help เป็นการอธิบายองค์ประกอบ การทำงาน รูปแบบการใช้งาน และตัวอย่างการใช้งาน ของตัว Operators ดังรูปที่ 2.10



รูปที่ 2.10 หน้าจอ Help

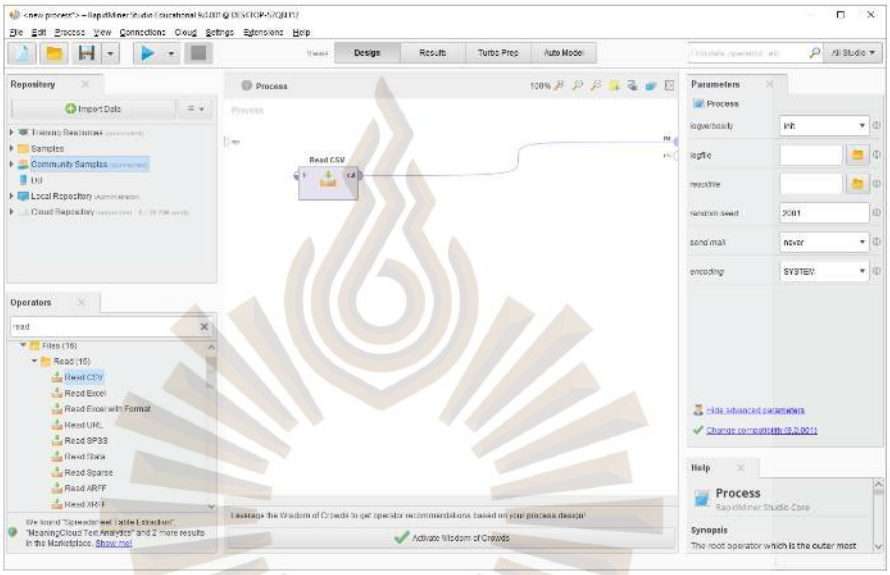
ที่มา: นฤชล โรจนบุรานนท์, 2562

มหาวิทยาลัยรังสิต Rangsit University

ตัวอย่างการใช้ Software Rapidminer ในการทำนายพนักงานที่จะลาออกด้วยเทคนิค Data Mining

ขั้นตอนที่ 1

ดึงข้อมูลโดยใช้ Operator “ Read CSV” ดังรูปที่ 2.11

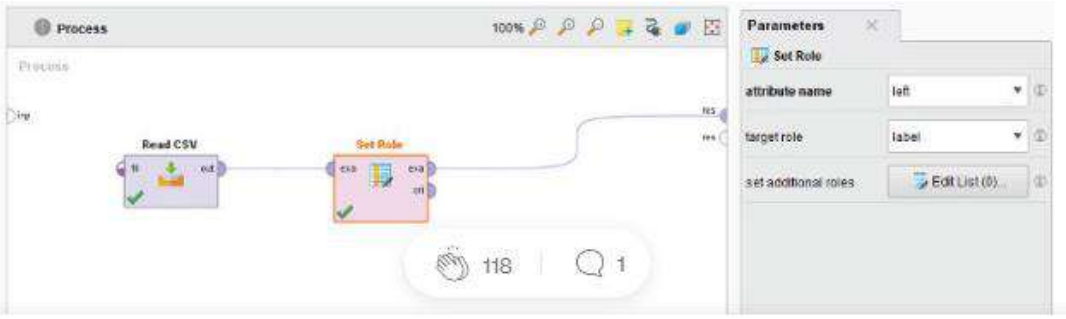


รูปที่ 2.11 การใช้ Operators Read CSV

ที่มา: Pimkote, 2018

ขั้นตอนที่ 2

ใช้ Operators Set Role เพื่อกำหนด Column Left ให้เป็นประเภท Label ดังรูปที่ 2.12



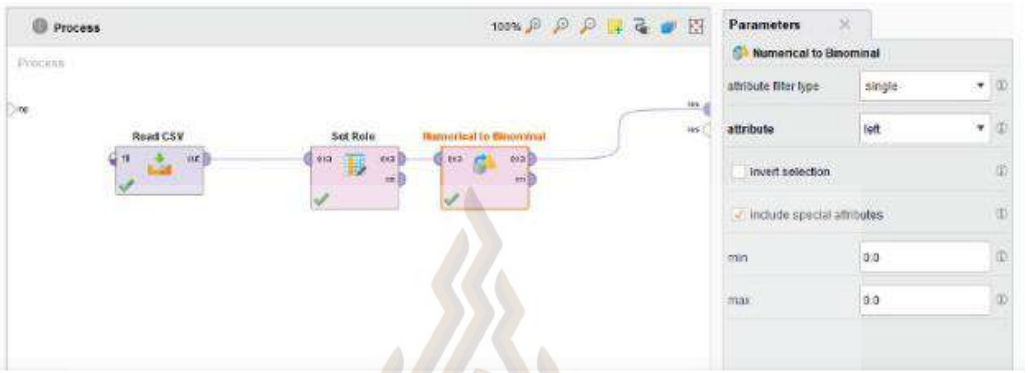
รูปที่ 2.12 การใช้ Operators Set Role

ที่มา: Pimkote, 2018

มหาวิทยาลัยรังสิต Rangsit University

ขั้นตอนที่ 3

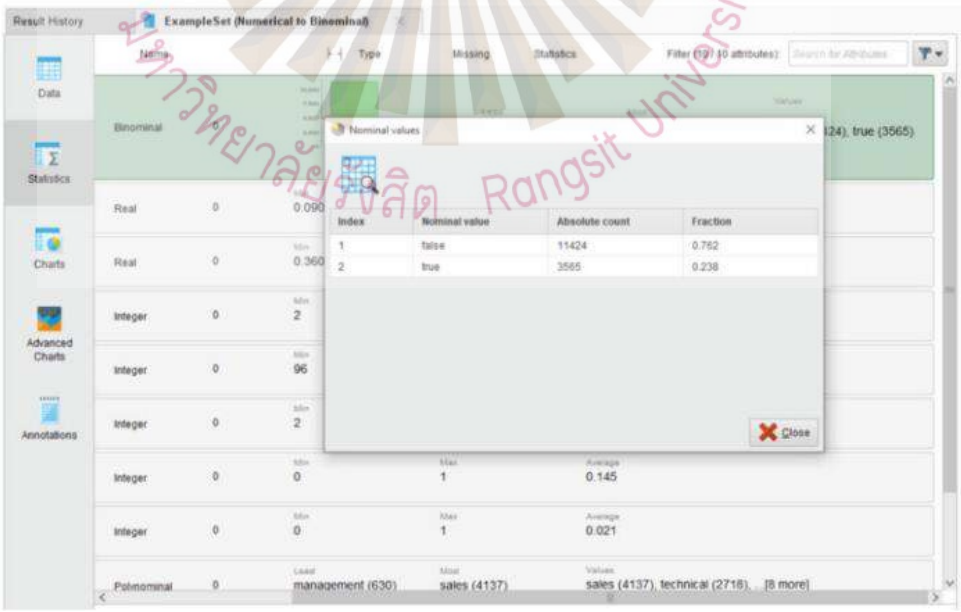
ใช้ Operator Numerical to Binominal กำหนดให้ Column Left เป็นประเภท Binominal หรือ True, False ดังรูปที่ 2.13



รูปที่ 2.13 การใช้ Operator Numerical to Binominal

ที่มา: Pimkote, 2018

เมื่อทำการรัน Process และดูที่ Statistics จะพบว่า ข้อมูลที่เป็นพนักงานลาออก หรือ False จะมีเพียง 3,565 ตัวอย่างหรือคิดเป็นร้อยละ 23.8 ดังรูปที่ 2.14

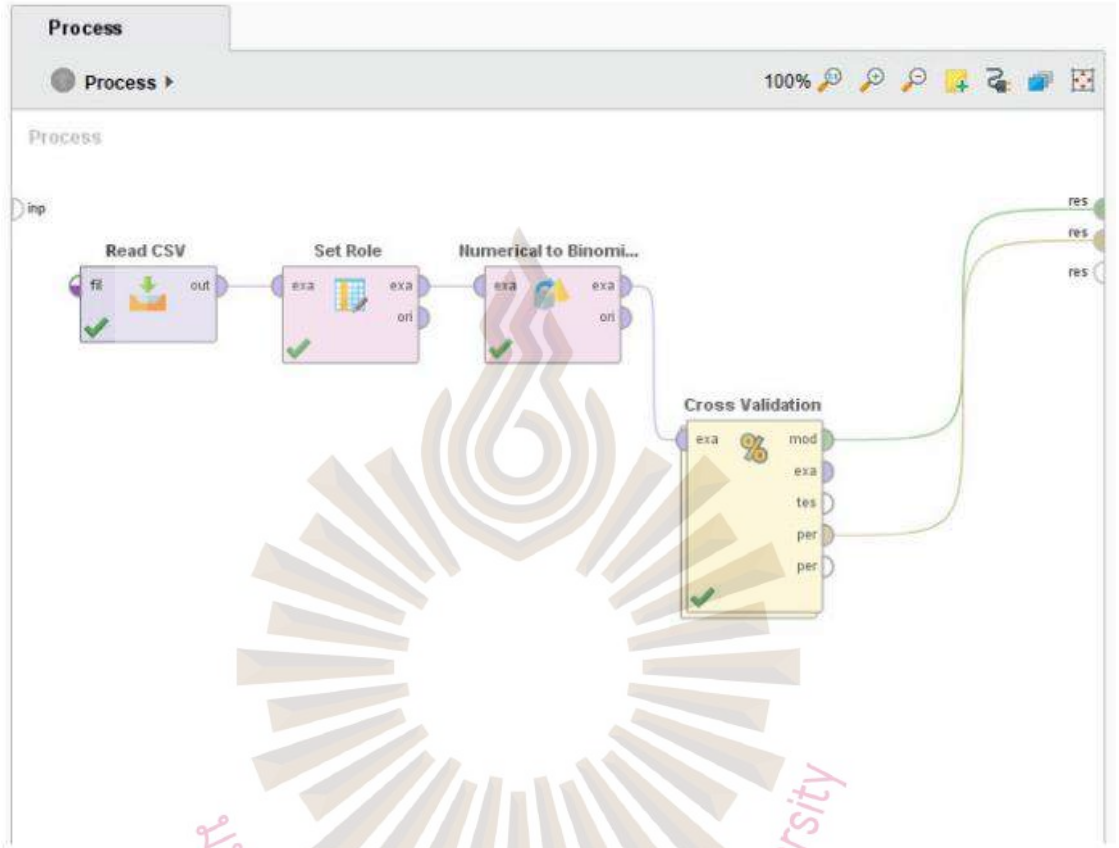


รูปที่ 2.14 การแสดงข้อมูลที่ได้ทำการดึงมา

ที่มา: Pimkote, 2018

ขั้นตอนที่ 4

เลือก Operators Cross Validation ดังรูปที่ 2.15

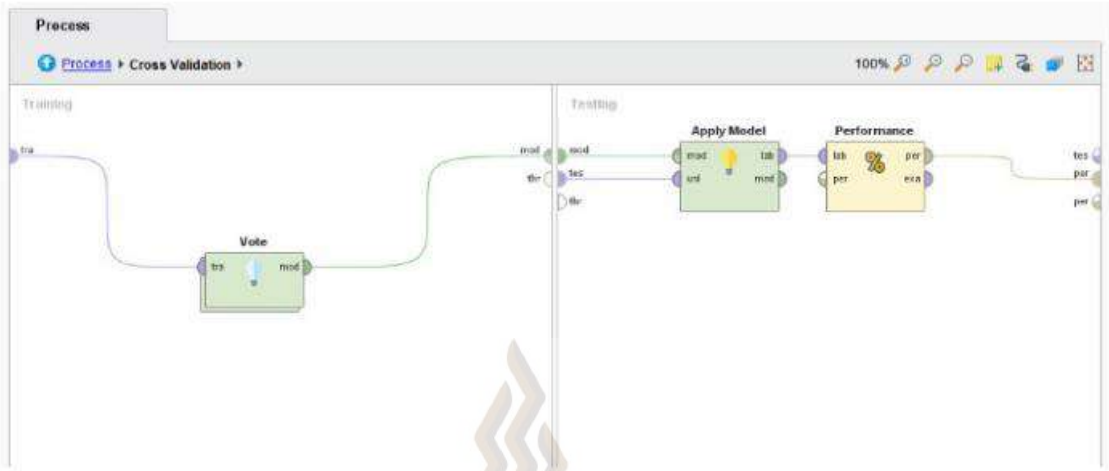


รูปที่ 2.15 การใช้ Operators Cross Validation

ที่มา: Pimkote, 2018

มหาวิทยาลัยรังสิต Rangsit University

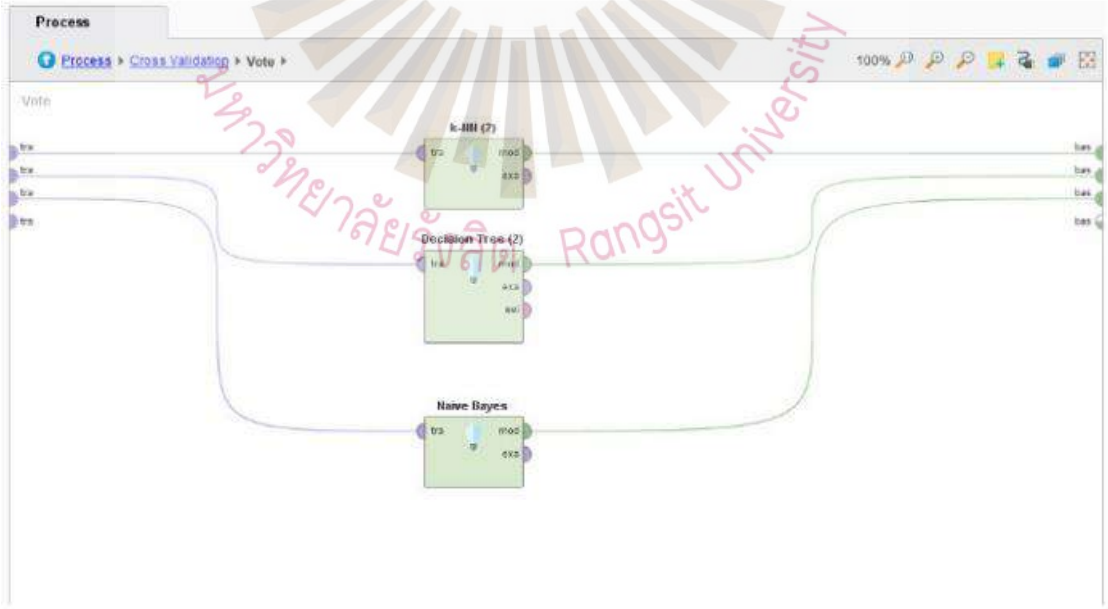
Double Click ที่ Operators Cross Validation และเลือก Operators ดังรูปที่ 2.16



รูปที่ 2.16 การเลือก Operators ภายใน Cross Validation

ที่มา: Pimkote, 2018

Double Click ที่ Operators Vote แล้วเลือก Model Operators ตามรูปที่ 2.17



รูปที่ 2.17 การเลือก Operations ภายใน Vote

ที่มา: Pimkote, 2018

มหาวิทยาลัยรังสิต Rangsit University

แล้วทำการ Run Process

ขั้นตอนที่ 5
ผลการพยากรณ์ ดังรูปที่ 2.18

Table View Plot View

accuracy: 97.19% +/- 0.24% (micro average: 97.19%)

	true false	true true	class precision
pred. false	11184	181	98.41%
pred. true	240	3384	93.38%
class recall	97.90%	94.92%	

รูปที่ 2.18 ตัวอย่างการแสดงผลการพยากรณ์พนักงานที่จะลาออก ด้วย Software RapidMiner
ที่มา: Pimkote, 2018

2.2 งานวิจัยที่เกี่ยวข้อง

ผู้วิจัยได้ศึกษางานวิจัยที่เกี่ยวข้องกับรูปแบบความสัมพันธ์เชิงสาเหตุของการยอมรับและการใช้เทคโนโลยีที่มีผลต่อพฤติกรรมการใช้แอปพลิเคชันช้อปปิ้งบนสมาร์โฟนของประชากรในเขตกรุงเทพมหานครและปริมณฑล (สุพจน์ อุ่นเรือน และสุมามาลย์ ปานคำ, 2560) และการทำเหมืองข้อมูล (จิราภา เลาหะวรนันท์ รชต ลิ้มสุทธิวันภูมิ และบัณฑิต ฐานะโสภณ, 2558) ได้มีงานวิจัยที่นำเทคนิคการวิเคราะห์ข้อมูลเพื่อมาพยากรณ์ เช่น การพยากรณ์การเลือกกลุ่มคณะวิชาในการศึกษาต่อในระดับปริญญาตรีของมหาวิทยาลัยรังสิต โดยใช้เทคนิคการทำเหมืองข้อมูล (ฐิตินันท์ อนุสรณ์ และสมชาย เล็กเจริญ, 2563) เมื่อผู้วิจัยได้เห็นถึงความสามารถในการทำเหมืองข้อมูลและวิเคราะห์โดยใช้เทคนิคต่าง ๆ เปรียบเทียบ ผู้วิจัยจึงได้ศึกษาเทคนิคที่เกี่ยวกับชุดข้อมูลที่ผู้วิจัยมีเป็นข้อมูลในลักษณะตัวเลข จึงได้ทำการศึกษาเทคนิคต้นไม้ตัดสินใจ (Decision Tree) (กำธร ศรีอุดม และวรารัตน์ รุ่งวรวุฒิ, 2556) เป็นอัลกอริทึมที่ได้รับความนิยมกันแพร่หลายเนื่องจากเป็นเทคนิคที่ใช้ในการจัดหมวดหมู่สามารถวิเคราะห์ผลลัพธ์ได้ผลที่ดีมาก ใช้ตัวจำแนกข้อมูลที่เหมาะสมสำหรับการจัดกลุ่ม (เดช ธรรมศิริ และพยุง มีสัจ, 2554) ในส่วนของงานวิจัยที่ใช้เทคนิคโครงข่ายประสาทเทียม (Neural Network) นั้น

(Rowley, Baluja and Kanade, 1998) ได้วิจัยเกี่ยวกับการตรวจจับใบหน้าด้วยเทคนิคเครือข่ายประสาทเทียมซึ่งใช้ตรวจจับใบหน้าได้ดีและมีประสิทธิภาพ ส่วนเทคนิคไนอีฟเบย์ (Naïve Bayes) (วิรชา วิรัชกุล , 2557) เป็นอัลกอริทึมที่เข้าใจง่ายและสามารถจำแนกข้อมูลได้อย่างรวดเร็ว และมีประสิทธิภาพเหมาะสมกับการพยากรณ์คลาสของชุดข้อมูล และใช้การวิเคราะห์ความแปรปรวน (Analysis of Variance: ANOVA) เพื่อเปรียบเทียบประสิทธิภาพของเทคนิคทั้ง 3 แบบ เพื่อให้ได้ผลลัพธ์ที่แม่นยำถูกต้องที่สุด พร้อมทั้งใช้ กฎความสัมพันธ์ (Association Rules) เป็นการค้นหาความสัมพันธ์ของข้อมูลที่ซ่อนอยู่ในข้อมูล โดยความสัมพันธ์สามารถอธิบายในรูปแบบกฎ (Rules) หรือรูปแบบของการเกิดร่วมกันของข้อมูลที่พบบ่อย ๆ (Frequency Pattern) (Dietrich, Heller and Yang., 2015)

มหาวิทยาลัยรังสิต Rangsit University

บทที่ 3

วิธีดำเนินการวิจัย

การวิจัยเรื่องการพยากรณ์และความสัมพันธ์ของการยอมรับการใช้เทคโนโลยีที่มีผลต่อพฤติกรรมการใช้แอปพลิเคชัน LAZADA ช้อปปิ้งบนสมาร์ทโฟนของประชากรในเขตกรุงเทพมหานครและปริมณฑล ซึ่งการวิจัยนี้เป็นการวิจัยเชิงเปรียบเทียบเทคนิคที่เหมาะสมกับชุดข้อมูล โดยมีสาระสำคัญในการดำเนินการตามลำดับ ดังนี้

3.1 ข้อมูลกลุ่มตัวอย่าง

ตัวอย่างข้อมูลในการวิจัยครั้งนี้เป็นข้อมูลทุติยภูมิซึ่งได้มาจากการวิจัยของสุพจน์ อุ่นเรือน และสุมามาลย์ ปานคำ (2560) จากเรื่องเทคโนโลยีที่มีผลต่อพฤติกรรมการใช้แอปพลิเคชัน LAZADA ช้อปปิ้งบนสมาร์ทโฟนของประชากรในเขตกรุงเทพมหานครและปริมณฑลจำนวน 386 คน โดยใช้เทคนิคการสุ่มอย่างง่าย และมีหัวข้อดังนี้

3.1.1 ความคาดหวังถึงประสิทธิภาพต่อการสื่อสาร (Performance Expectancy : PE) มี 6 หัวข้อย่อย

1) ท่านคิดว่าแอปพลิเคชันช้อปปิ้งบนสมาร์ทโฟน (LAZADA) มีประโยชน์ในการซื้อสินค้า (PE1)

2) ท่านคิดว่าแอปพลิเคชันช้อปปิ้งบนสมาร์ทโฟน (LAZADA) ช่วยให้เกิดความสะดวกในการซื้อสินค้า (PE2)

3) ท่านคิดว่าแอปพลิเคชันช้อปปิ้งบนสมาร์ทโฟน (LAZADA) ช่วยให้เกิดความรวดเร็วในการซื้อสินค้า (PE3)

4) ท่านคิดว่าแอปพลิเคชันช้อปปิ้งบนสมาร์ทโฟน (LAZADA) ช่วยประหยัดเวลาในการซื้อสินค้า (PE4)

5) ท่านคิดว่าแอปพลิเคชันช้อปปิ้งบนสมาร์ทโฟน (LAZADA) ช่วยให้ประหยัดค่าใช้จ่ายซื้อสินค้า (PE5)

มหาวิทยาลัยรังสิต Rangsit University

6) ท่านคิดว่าแอปพลิเคชันช้อปปิ้งบนสมาร์ทโพน (LAZADA) มีความปลอดภัยในการซื้อสินค้า (PE6)

3.1.2 ความคาดหวังในความพยายามการใช้งาน (Effort Expectancy : EE) มี 5 หัวข้อย่อย

1) ท่านคิดว่าแอปพลิเคชันช้อปปิ้งบนสมาร์ทโฟน (LAZADA) เป็นระบบที่เข้าใจง่าย (EE1)

2) ท่านคิดว่าแอปพลิเคชันช้อปปิ้งบนสมาร์ทโฟน (LAZADA) เป็นระบบที่ไม่ซับซ้อน (EE2)

3) ท่านคิดว่าแอปพลิเคชันช้อปปิ้งบนสมาร์ทโฟน (LAZADA) แสดงข้อมูลได้ถูกต้อง (EE3)

4) ท่านคิดว่าแอปพลิเคชันช้อปปิ้งบนสมาร์ทโฟน (LAZADA) สามารถเรียนรู้การใช้งานได้ง่าย (EE4)

5) ท่านคิดว่าการเพิ่มทักษะความชำนาญในการใช้แอปพลิเคชันช้อปปิ้งบนสมาร์ทโฟน (LAZADA) ทำได้ง่าย (EE5)

3.1.3 อิทธิพลของสังคม (Social Influence : SI) มี 5 หัวข้อย่อย

1) ท่านคิดว่าเพื่อนมีผลทำให้ท่านใช้แอปพลิเคชันช้อปปิ้งบนสมาร์ทโฟน (LAZADA) (SI1)

2) ท่านคิดว่าครอบครัวของท่านมีผลทำให้ท่านใช้แอปพลิเคชันช้อปปิ้งบนสมาร์ทโฟน (LAZADA) (SI2)

3) ท่านคิดว่าสื่อโฆษณามีผลทำให้ท่านใช้แอปพลิเคชันช้อปปิ้งบนสมาร์ทโฟน (LAZADA) (SI3)

4) ท่านคิดว่าการซื้อสินค้าผ่านแอปพลิเคชันช้อปปิ้งบนสมาร์ทโฟน (LAZADA) ช่วยทำให้ท่านก้าวทันเทคโนโลยี (SI4)

5) ท่านคิดว่าการใช้แอปพลิเคชันช้อปปิ้งบนสมาร์ทโฟน (LAZADA) ช่วยเสริมสร้างภาพลักษณ์ว่าเป็นคนทันสมัยและทันต่อเทคโนโลยี (SI6)

3.1.4 สภาพสิ่งอำนวยความสะดวกในการใช้งาน (Facilitating Conditions : FC) มี 5 หัวข้อย่อย

1) ท่านมีอุปกรณ์เชื่อมในการใช้งานแอปพลิเคชันช้อปปิ้งบนสมาร์โฟน (LAZADA) ได้ตลอดเวลา (FC1)

2) ท่านคิดว่าหน้าจอแสดงผลบนโทรศัพท์มือถือที่เล็กไม่เป็นอุปสรรคในการใช้งานแอปพลิเคชันช้อปปิ้งบนสมาร์โฟน (FC2)

3) ท่านมีความรู้ทางเทคโนโลยีมากพอที่จะใช้แอปพลิเคชันช้อปปิ้งบนสมาร์โฟน (LAZADA) ในการซื้อสินค้า (FC3)

4) ท่านคิดว่าความเร็วของระบบอินเทอร์เน็ตที่ท่านใช้มีผลในการใช้แอปพลิเคชันช้อปปิ้งบนสมาร์โฟน(FC4)

5) ท่านคิดว่าการซื้อสินค้าผ่านแอปพลิเคชันช้อปปิ้งบนสมาร์โฟน (LAZADA) ง่ายกว่าการซื้อสินค้าที่ห้างสรรพสินค้า (FC5)

3.1.5 พฤติกรรมความตั้งใจในการใช้งาน (Behavior Intention : BI) มี 4 หัวข้อย่อย

1) ท่านมีความตั้งใจในการใช้แอปพลิเคชันช้อปปิ้งบนสมาร์โฟน (LAZADA) ในการซื้อสินค้าแทนการไปตามห้างสรรพสินค้าและร้านค้า (BI1)

2) ท่านใช้แอปพลิเคชันช้อปปิ้งบนสมาร์โฟน (LAZADA) ในการซื้อสินค้าเป็นประจำอย่างต่อเนื่อง (BI2)

3) ท่านมีความตั้งใจว่าจะใช้งานแอปพลิเคชันช้อปปิ้งบนสมาร์โฟน (LAZADA) มากขึ้นในอนาคต (BI3)

4) เมื่อท่านเข้าถึงแอปพลิเคชันช้อปปิ้งบนสมาร์ทโฟน (LAZADA) ท่านมีความตั้งใจที่จะทำรายการใดรายการหนึ่ง (BI4)

5) ท่านมีความตั้งใจที่จะแนะนำแอปพลเคชันช้อปปิ้งบนสมาร์โฟน (LAZADA) ให้กับผู้อื่น (BI5)

3.1.6 พฤติกรรมการใช้งาน(Use Behavior : UB)

1) ท่านใช้แอปพลิเคชันช้อปปิ้งบนสมาร์โฟน (LAZADA) ในการซื้อสินค้าเพื่อความสะดวก (UB1)

2) ท่านใช้แอปพลิเคชันช้อปปิ้งบนสมาร์โฟน (LAZADA) ในการซื้อสินค้าเพื่อความรวดเร็ว (UB2)

3) ท่านใช้แอปพลิเคชันช้อปปิ้งบนสมาร์โฟน (LAZADA) ในการซื้อสินค้าเพื่อประหยัดเวลาในการเดินทางไปซื้อสินค้าตามห้างสรรพสินค้าและร้านค้า (UB3)

4) ท่านใช้แอปพลิเคชันช้อปปิ้งบนสมาร์โฟน (LAZADA) ในการซื้อสินค้าเป็นประจำ (UB4)

5) ท่านใช้แอปพลิเคชันช้อปปิ้งบนสมาร์โฟน (LAZADA) เพื่อซื้อสินค้าและการชำระค่าสินค้า (UB5)

6) ท่านใช้แอปพลิเคชันช้อปปิ้งบนสมาร์โฟน (LAZADA) เพื่อตรวจสอบราคาสินค้า (UB6)

ตัวอย่างข้อมูลทุติยภูมิที่ใช้ในการพยากรณ์เปรียบเทียบ ดังตารางที่ 3.1

ตารางที่ 3.1 ตัวอย่างข้อมูลที่จะนำไปเปรียบเทียบ

PE_total	EE_total	SI_total	FC_total	BI_total	BU_total	RESULT
4.83	4.8	4	4	3.6	4	N
5.83	5	5	5	3.2	4.33	N
4	4	4	4	3.6	4	N
5.5	5	5	5.8	2.8	2	N
4.17	3	4	4.6	1.8	2.17	N
4.83	5.4	4	4.2	3	3.33	N
6.17	5	5	3.8	2.8	1	N
4.67	5	3.2	4.2	2.6	3	N
5.5	5	5.2	4.8	3	3.33	N
3.33	3.4	3.4	4.2	3	3	N
4.5	5.2	3.4	4.8	1.8	1.33	N
3.67	4	4.2	4	2.2	2.5	N
5	4	4	3	2.2	1.33	N
4.67	4.8	4.8	4.8	2.6	3.33	N
1.5	2.6	1	4.4	1	1	N
4.67	3.6	4	4.8	3.6	3.83	N

มหาวิทยาลัยรังสิต Rangsit University

ตารางที่ 3.1 ตัวอย่างข้อมูลที่จะนำไปเปรียบเทียบ (ต่อ)

PE_total	EE_total	SI_total	FC_total	BI_total	BU_total	RESULT
3.83	4	4.6	3.8	1.4	4	N
4.5	4	3.4	2.6	2	2.67	N
5	5	5	5	2.4	1	N
1.83	2.8	1.4	3	1.2	2.17	N
1.83	2.8	1.4	3	1.2	2.17	N
2.5	2.8	2.6	3	2.8	1.5	N
4	4	4	4	3.6	4	N
6	5.8	3.8	5.8	3.8	3.33	N
5.67	5.6	5.4	5.8	4.6	5.17	N
4.67	4.4	4.8	4.8	4.4	4.67	N
5	5.4	2	6.8	2.6	5.5	N
5.33	5.4	3.8	4.6	3.2	4.17	N
5.83	6.4	6.6	5	5.8	1.83	N
5.83	6	5.8	6.6	4.6	5.17	N
5.5	5.4	4	5.2	3.2	1.5	N
2.67	3	3.2	3	3.2	3.17	N
3.67	4.6	4.4	5.8	3.6	3.83	N
2.33	2.4	4.8	6	4.8	4.33	N
1.83	2.6	2	2.4	2	1.67	N
4.17	3.8	4.2	4	4.4	4.33	N
6	4.6	2.2	6.2	2.4	3	N
3.67	3.6	3.8	3.6	3.6	3.5	N
4.33	5.2	3.4	4.2	2.8	3.83	N
3.17	5.4	1	4.6	3.6	4.67	N

มหาวิทยาลัยรังสิต Rangsit University

ตัวอย่างข้อมูลทุติยภูมิที่ใช้ในการหาความสัมพันธ์ ดังตารางที่ 3.2

ตารางที่ 3.2 ตัวอย่างข้อมูลที่จะนำไปหาความสัมพันธ์

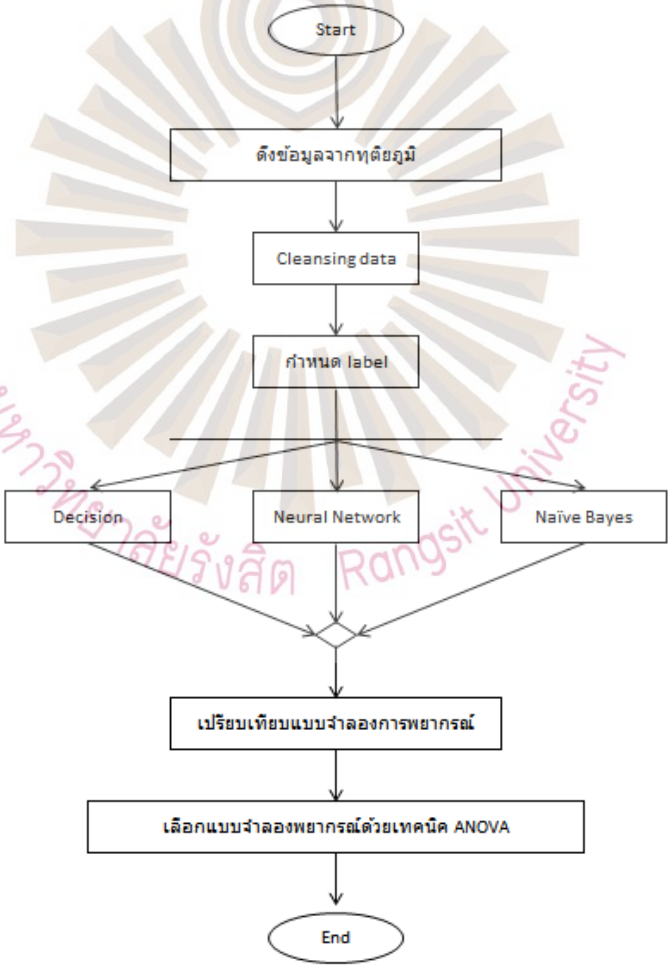
PE1	PE2	PE3	PE4	PE5	PE6	EE1	EE2	EE3	EE4	EE5	SI1	SI2	SI3	SI4	SI5
4	3	4	3	5	3	4	5	5	4	5	4	7	5	3	4
5	5	7	7	2	3	7	5	4	4	4	4	4	4	4	4
5	6	6	7	6	4	5	5	4	5	5	5	5	7	7	5
4	5	3	2	4	3	3	3	4	3	5	5	4	4	3	4
7	7	3	3	4	3	3	3	7	7	7	3	1	5	4	4
5	5	5	6	4	5	4	4	6	5	4	4	5	5	7	7
7	7	7	7	5	6	6	6	6	6	6	7	4	6	6	6
6	6	7	5	4	6	6	6	7	7	5	5	6	6	7	7
6	6	6	6	6	5	5	5	5	5	5	6	4	5	5	5
7	7	5	4	4	4	5	5	5	5	5	4	4	5	5	4
6	6	5	5	2	3	5	5	4	5	6	6	2	3	5	6
7	7	7	7	7	6	6	6	6	6	7	6	7	7	7	7
4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
6	6	6	6	4	5	5	5	5	5	5	6	4	5	5	5
3	4	7	5	2	4	3	4	2	3	3	4	4	5	3	4
5	7	7	5	5	7	5	5	7	7	5	3	1	4	2	5
5	5	6	5	4	4	5	5	6	6	5	5	3	4	4	4
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
7	7	5	7	4	3	4	4	4	5	7	5	3	7	7	7
5	5	4	4	3	3	4	4	4	4	4	4	4	3	5	4
7	7	7	7	4	5	5	5	3	7	5	5	3	7	5	5
5	5	5	5	4	4	5	5	5	5	5	3	3	3	4	3
7	7	7	5	3	4	4	4	6	6	5	4	4	5	6	7
3	4	4	4	3	2	3	3	3	4	4	4	2	3	4	4

มหาวิทยาลัยรังสิต Rangsit University

3.2 เทคนิคการทำเหมืองข้อมูล (Data Mining)

คือ การนำเอาข้อมูลที่มีจำนวนมากมาสกัดสารสนเทศเพื่อนำมาวิเคราะห์ถึงรูปแบบความสัมพันธ์ที่ซ่อนอยู่ในชุดรูปแบบ และสามารถนำข้อสรุปที่เป็นประโยชน์ มาช่วยในการตัดสินใจที่เป็นประโยชน์ต่อองค์กร โดยกระบวนการทำเหมืองข้อมูลมีด้วยกันหลายหลังการ เช่น เทคนิคการเรียนรู้ของเครื่อง (Machine Learning) ทางสถิติ เป็นต้น (จิราภา เลาหะวรนันท์ และคณะ, 2558)

วิธีดำเนินการเพื่อเปรียบเทียบแบบจำลองการพยากรณ์ระหว่างต้นไม้การตัดสินใจ (Decision Tree) โครงข่ายประสาทเทียม (Neural Network) และไนอีฟเบย์ (Naïve Bayes)



รูปที่ 3.1 ขั้นตอนทำการเปรียบเทียบเพื่อพยากรณ์

ขั้นตอนทำการเปรียบเทียบเพื่อพยากรณ์ดังรูปที่ 3.1 ซึ่งสามารถอธิบายได้ดังนี้

1) ขั้นตอนการเอาข้อมูลจากทุติยภูมิ ซึ่งได้จากการวิจัยสุพจน์ อุ่นเรือน และสุมามาลย์ ปานคำ (2560) จากเรื่อง เทคโนโลยีที่มีผลต่อพฤติกรรมการใช้แอปพลิเคชัน LAZADA ช้อปปิ้งบนสมาร์ทโฟนของประชากรในเขตกรุงเทพมหานครและปริมณฑล มาจัดทำในรูปของ CSV เพื่อให้เหมาะกับการพยากรณ์

2) ขั้นตอน Cleansing Data เพื่อเตรียมข้อมูลสำหรับการทำเหมืองข้อมูล เพื่อปรับข้อมูลให้เหมาะสมกับการนำไปใช้ในการทำเหมืองข้อมูล โดยการตรวจสอบความถูกต้องของข้อมูล ถ้าเจอความไม่ถูกต้องจะดำเนินการตัดข้อมูลหรือปรับปรุงข้อมูลให้ครบถ้วน

3) ขั้นตอนกำหนด Label เพื่อเตรียมกำหนดข้อมูลที่จะใช้ในการพยากรณ์ โดยจะใช้ Operators Set Role ใน RapidMiner Studio ในการกำหนด Label

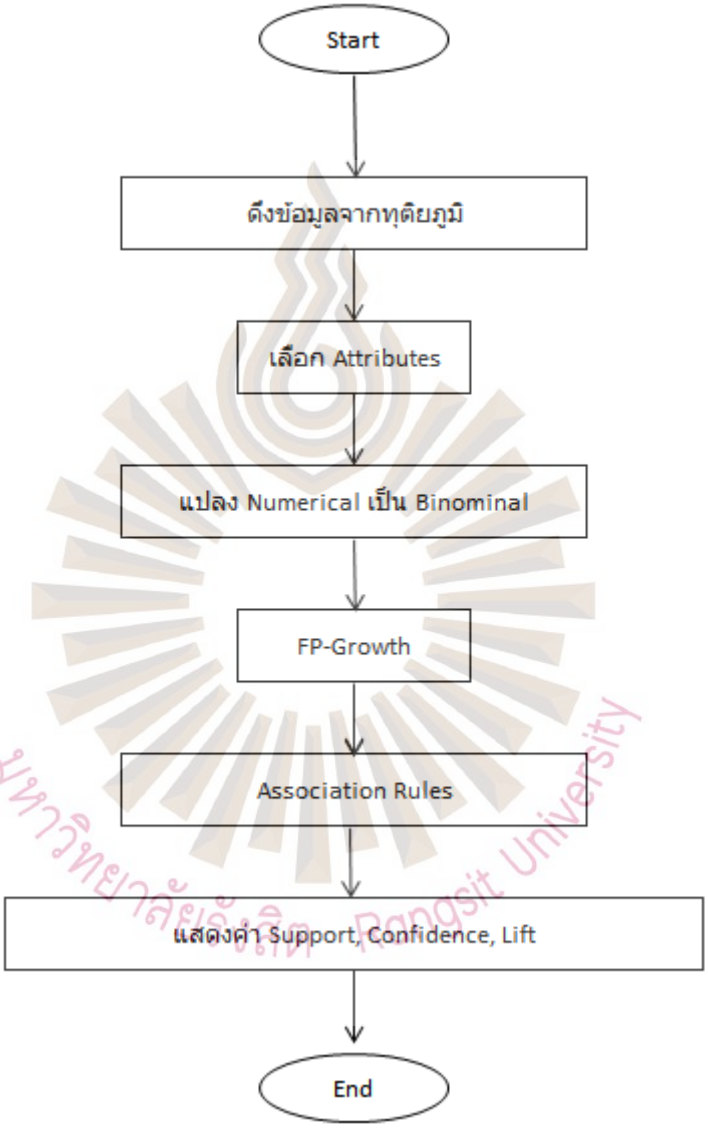
4) ขั้นตอนการเลือกแบบจำลองการพยากรณ์ ในการวิจัยครั้งนี้ได้เลือกแบบจำลองการพยากรณ์ 3 แบบที่เหมาะสมกับการวิจัยในครั้งนี้ ได้แก่เทคนิคต้นไม้การตัดสินใจ เทคนิคเครือข่ายประสาทเทียม และ เทคนิคไนอีฟเบย์ โดยแบบจำลองดังกล่าว เป็นแบบการจำแนกข้อมูล (Classification) ซึ่งเหมาะสมกับการพยากรณ์และความสัมพันธ์ของการยอมรับการใช้เทคโนโลยีที่มีผลต่อพฤติกรรมการใช้แอปพลิเคชัน LAZADA ช้อปปิ้งบนสมาร์ทโฟนของประชากรในเขตกรุงเทพมหานครและปริมณฑล

5) ขั้นตอนการเปรียบเทียบประสิทธิภาพของแบบจำลองทั้ง 3 แบบ คือ เทคนิคต้นไม้การตัดสินใจ เทคนิคเครือข่ายประสาทเทียม และ เทคนิคไนอีฟเบย์ หาค่าความถูกต้อง (Accuracy) ค่าความเที่ยงตรง (Precision) และค่าความระลึก (Recall) ที่ได้ค่ามากที่สุด

6) ขั้นตอนการเลือกแบบจำลองการพยากรณ์ที่เหมาะสมกับการพยากรณ์และความสัมพันธ์ของการยอมรับการใช้เทคโนโลยีที่มีผลต่อพฤติกรรมการใช้แอปพลิเคชัน LAZADA ช้อปปิ้งบนสมาร์ทโฟนของประชากรในเขตกรุงเทพมหานครและปริมณฑล โดยใช้สถิติ ANOVA ในการเปรียบเทียบแบบจำลอง

มหาวิทยาลัยรังสิต Rangsit University

วิธีการเพื่อหาความสัมพันธ์ของการยอมรับการใช้เทคโนโลยีที่มีผลต่อพฤติกรรมการใช้แอปพลิเคชัน LAZADA ช้อปปิ้งบนสมาร์ทโฟนของประชากรในเขตกรุงเทพมหานครและปริมณฑล โดยใช้เทคนิคการกฎความสัมพันธ์ Association Rules มีวิธีการดังรูปที่ 3.2



รูปที่ 3.2 ขั้นตอนเพื่อหาความสัมพันธ์ของข้อมูล

1) ขั้นตอนการเอาข้อมูลจากทุติยภูมิ ซึ่งได้จากการวิจัยสุพจน์ อุ่นเรือน และสุมามาลย์ ปานคำ (2560) จากเรื่อง เทคโนโลยีที่มีผลต่อพฤติกรรมการใช้แอปพลิเคชัน LAZADA ช้อปปิ้งบน

สมาร์ทโฟน ของประชากรในเขตกรุงเทพมหานครและปริมณฑล มาจัดทำในรูปของ CSV เพื่อให้เหมาะกับการพยากรณ์

2) ขั้นตอนเลือก Attributes ที่ต้องการหาความสัมพันธ์ของข้อมูล

3) ขั้นตอนแปลง Numerical เป็น Binominal กำหนดทุก Column ให้เป็น Binominal

4) ขั้นตอน FP-Growth นำข้อมูลที่แปลงเป็น Binominal มาทำการเข้าโมเดล FP-Growth โดยกำหนด Min Support

5) ขั้นตอน Association Rule เป็นการสร้างกฎตามเพื่อหาความสัมพันธ์ของข้อมูล โดยทำการกำหนดค่า min confidence

6) ขั้นตอนแสดงค่า Support, Confidence, Lift นำค่าดังกล่าวมาวิเคราะห์หาความสัมพันธ์ที่เกี่ยวข้องกับข้อมูลชุดดังกล่าว

มหาวิทยาลัยรังสิต Rangsit University

บทที่ 4

ผลการวิจัย

ชุดข้อมูลและการทดสอบและเกณฑ์การวัดประสิทธิภาพถูกนำมาใช้ในบทนี้โดยจะเริ่มจากชุดข้อมูล

4.1 ชุดข้อมูล

งานวิจัยนี้ได้ใช้ชุดข้อมูลทุติยภูมิ แบ่งออกเป็น 2 ประเภท คือ ชุดข้อมูลที่ใช้ในการเปรียบเทียบแบบจำลองการพยากรณ์ระหว่างต้นไม้การตัดสินใจ (Decision Tree) โครงข่ายประสาทเทียม (Neural Network) และไนอีฟเบย์ (Naïve Bayes) และ ชุดข้อมูลที่ใช้ในการหาความสัมพันธ์ของข้อคำถามการยอมรับการใช้เทคโนโลยีที่มีผลต่อพฤติกรรมการใช้แอปพลิเคชัน LAZADA ช้อปปิ้งบนสมาร์ทโฟนของประชากรในเขตกรุงเทพมหานครและปริมณฑล

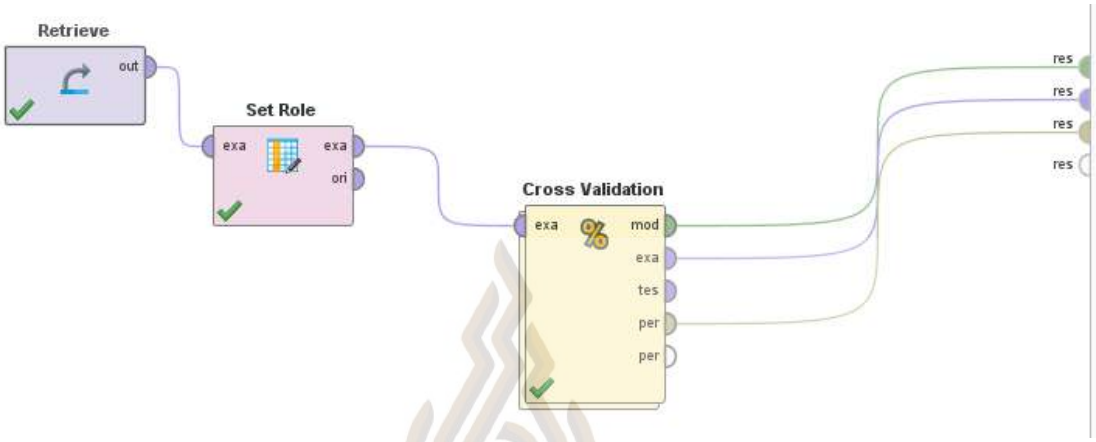
4.2 การทดลองและการวัดผล

จากผลการวิจัยสามารถสรุปผลการวิจัยได้เป็น 4 ตอน ดังนี้

ตอนที่ 1 การเปรียบเทียบแบบจำลองการพยากรณ์โดยใช้เทคนิคการจำแนกข้อมูลด้วยวิธีต้นไม้การตัดสินใจ เทคนิคการจำแนกข้อมูลด้วยวิธีเครือข่ายประสาทเทียม และเทคนิคการจำแนกข้อมูลด้วยวิธีไนอีฟเบย์ โดยใช้ข้อมูลจาก

1) ขั้นตอนการดำเนินการ จากการเลือกแบบจำลองพยากรณ์ทั้ง 3 แบบ เพื่อหาค่า Recall และ Accuracy Precision ในแต่ล่ะเทคนิคในการพยากรณ์ จะประกอบด้วย Operators ดังนี้ Retrieve Data เป็นการดึงข้อมูลที่อยู่ในรูปแบบไฟล์นามสกุล CSV ที่มีข้อมูลที่จะใช้ในการพยากรณ์ด้วยแบบจำลองโดยเทคนิคต่าง ๆ Set Role เป็นการเลือก Attributes ที่ต้องการใช้เป็น Class ในการพยากรณ์ และ Cross Validation เป็นการทดสอบประสิทธิภาพของแต่ละเทคนิคในการพยากรณ์ โดย Cross Validation จะมี Sub Process ของแต่ละเทคนิคการพยากรณ์โดยใช้โปรแกรม Rapid Miner

Studio เป็นเครื่องมือช่วยในการพยากรณ์ โดยการทดสอบแต่ละเทคนิคจะมี Operators หลัก ๆ ที่ใช้อยู่ ดังรูปที่ 4.1



รูปที่ 4.1 Operators หลักที่ใช้

รายละเอียด Operators

Retrieve ทำการ Set Parameters ดังรูปที่ 4.2 และ 4.3

Repository Entry ให้ทำการดึงข้อมูลมาจากฐานข้อมูลในโปรแกรม RapidMiner

รูปที่ 4.2 Operators Retrieve

มหาวิทยาลัยรังสิต Rangsit University

รูปที่ 4.3 การ Set Parameters Operators Retrieve

Set Role ทำการ Set Parameters ดังรูปที่ 4.4 และ 4.5

Attribute Name = Filed Result

Target Role = Label

รูปที่ 4.4 Operators Set Role

รูปที่ 4.5 การ Set Parameters Operators Set Role

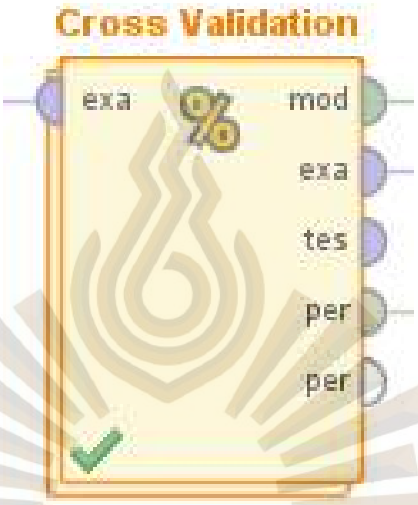
มหาวิทยาลัยรังสิต Rangsit University

Cross Validation ทำการ Set Parameters ดังรูปที่ 4.6 และ 4.7

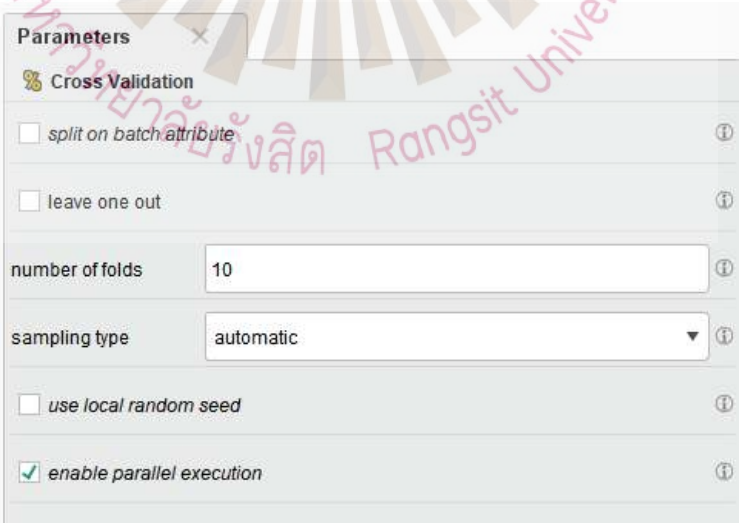
Number of Folds = 10

Sampling Type = Automatic

Enable Parallel Execution = Turn



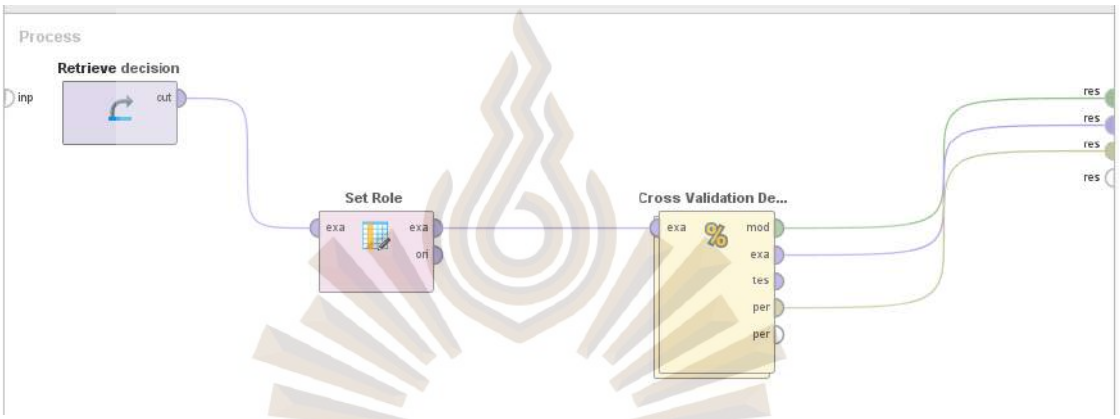
รูปที่ 4.6 Operators Cross Validation



รูปที่ 4.7 การ Set Parameters Operators Cross Validation

มหาวิทยาลัยรังสิต Rangsit University

2) กระบวนการสร้างแบบจำลองการพยากรณ์โดยใช้เทคนิคการจำแนกข้อมูลด้วยวิธีต้นไม้การตัดสินใจ จะประกอบไปด้วย Operators ดังต่อไปนี้ Decision Tree เป็นการจำลองการพยากรณ์โดยใช้เทคนิคการจำแนกข้อมูลด้วยวิธีต้นไม้การตัดสินใจ Apply Model เป็นการสร้างแบบจำลองการพยากรณ์ที่สร้างได้ไปใช้ในการพยากรณ์ เพื่อให้ได้คำตอบของข้อมูลที่ยังไม่ทราบคลาส Performance เป็นการวัดประสิทธิภาพของแบบจำลองของการพยากรณ์ จะได้ค่า Accuracy Class Recall Precision Recall เป็นต้น ดังรูปที่ 4.8



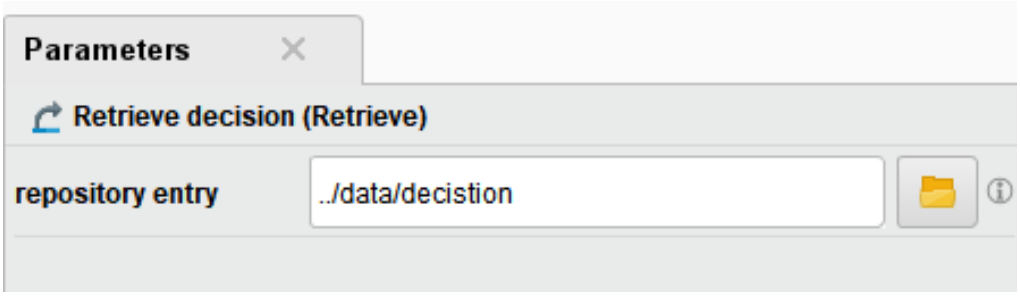
รูปที่ 4.8 Operators หลัก สำหรับ Decision Tree

รายละเอียด Operators

Retrieve ทำการ Set Parameters ดังรูปที่ 4.9 และ 4.10

Repository Entry ให้ทำการดึงข้อมูลมาจากฐานข้อมูลในโปรแกรม RapidMiner

รูปที่ 4.9 Operators Retrieve Decision

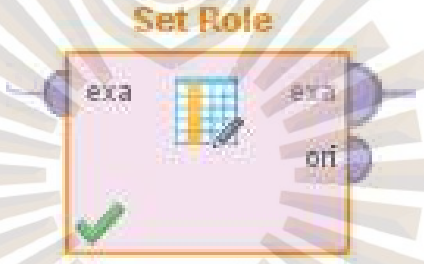


รูปที่ 4.10 การ Set Parameters Operators Retrieve Decision

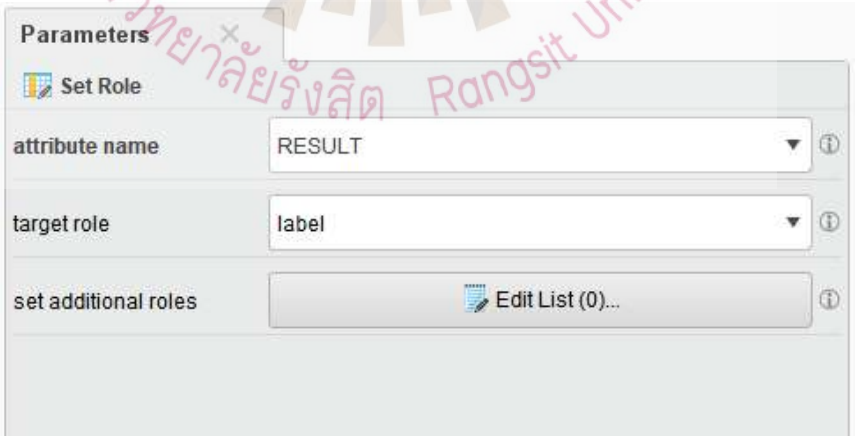
Set Role ทำการ Set Parameters ดังรูปที่ 4.11 และ 4.12

Attribute Name = Filed Result

Target Role = Label



รูปที่ 4.11 Operators Set Role Decision



รูปที่ 4.12 การ Set Parameters Operators Set Role Decision

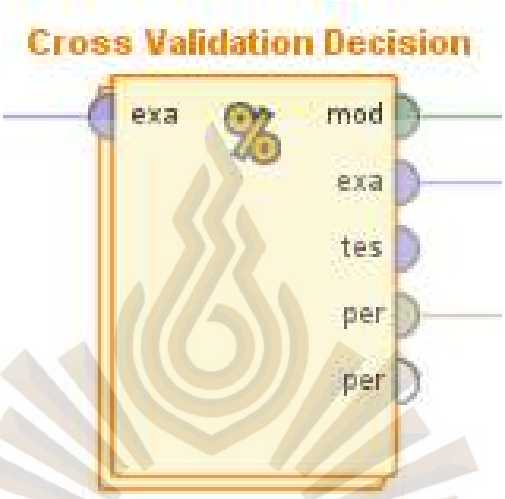
มหาวิทยาลัยรังสิต Rangsit University

Cross Validation Decision ทำการ Set Parameters ดังนี้

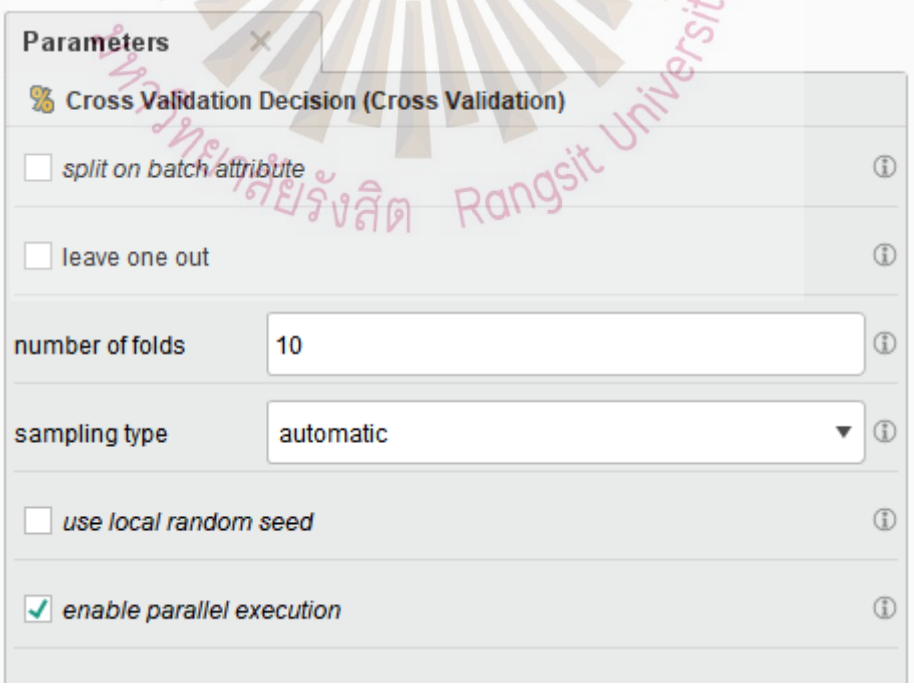
Number of Folds = 10

Sampling Type = Automatic

Enable Parallel Execution = Turn

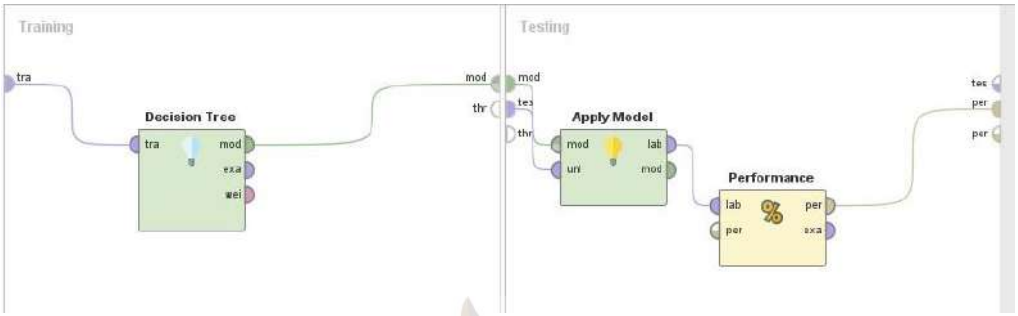


รูปที่ 4.13 Operators Cross Validation Decision



รูปที่ 4.14 การ Set Parameters Operators Cross Validation Decision

Operators ภายใน Cross Validation Decision



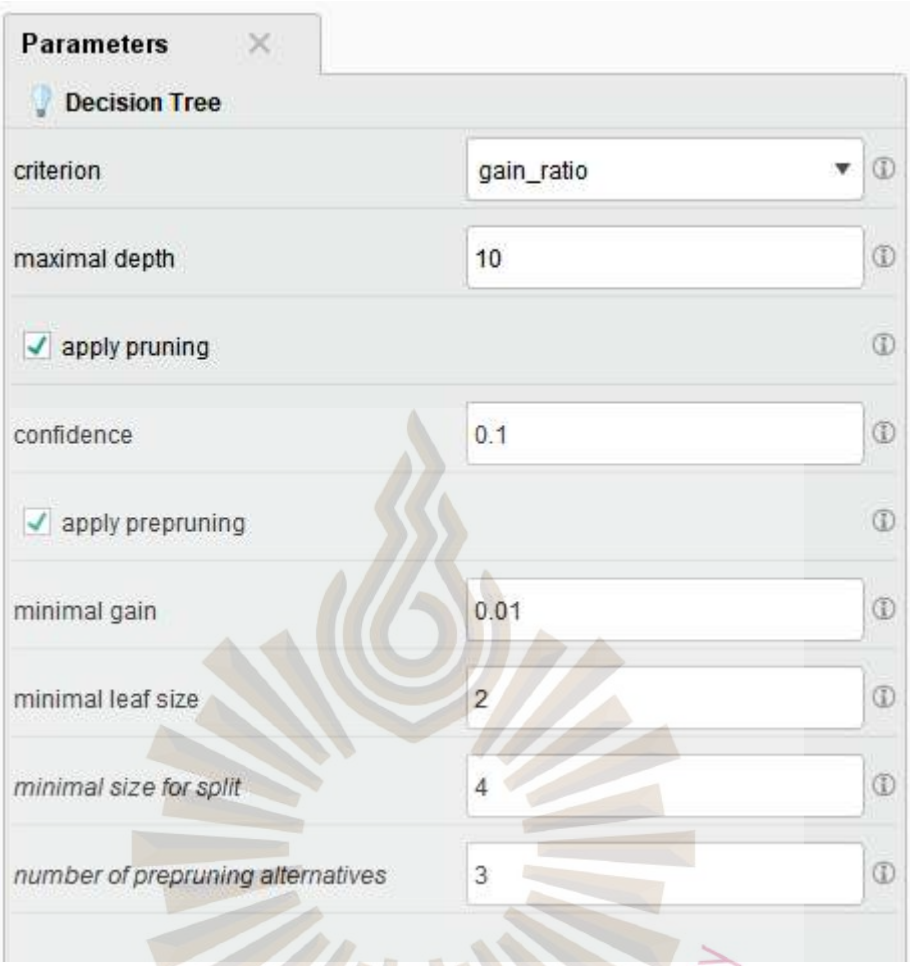
รูปที่ 4.15 Operators ภายใน Operators Cross Validation Decision

Decision Tree set parameters ดังรูปที่ 4.16 และ 4.17

Criterion	=	Gain_ratio
Maximal Depth	=	10
Apply Pruning	=	Turn
Confidence	=	0.1
○ Apply Prepruning	=	Turn
○ Minimal Gain	=	0.01
○ Minimal size for Split	=	4
○ Number of Prepruning Alternatives	=	3

รูปที่ 4.16 Operators Decision Tree

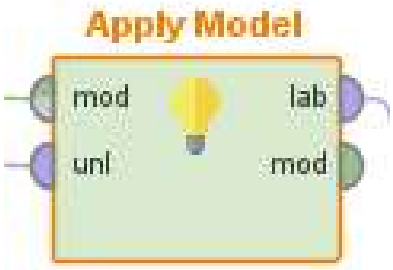
มหาวิทยาลัยรังสิต Rangsit University



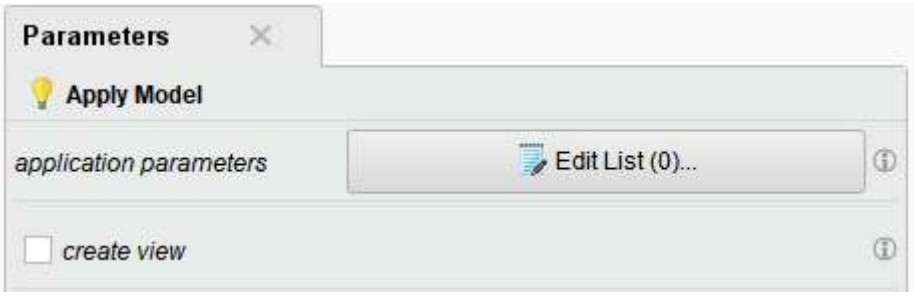
รูปที่ 4.17 การ Set Parameters Operators Decision Tree

Apply Model set parameters ดังรูปที่ 4.18 และ 4.19

No set



รูปที่ 4.18 Operators Apply Model



รูปที่ 4.19 การ Set Parameters Operators Apply Model

Performance Set Parameters ดังนี้

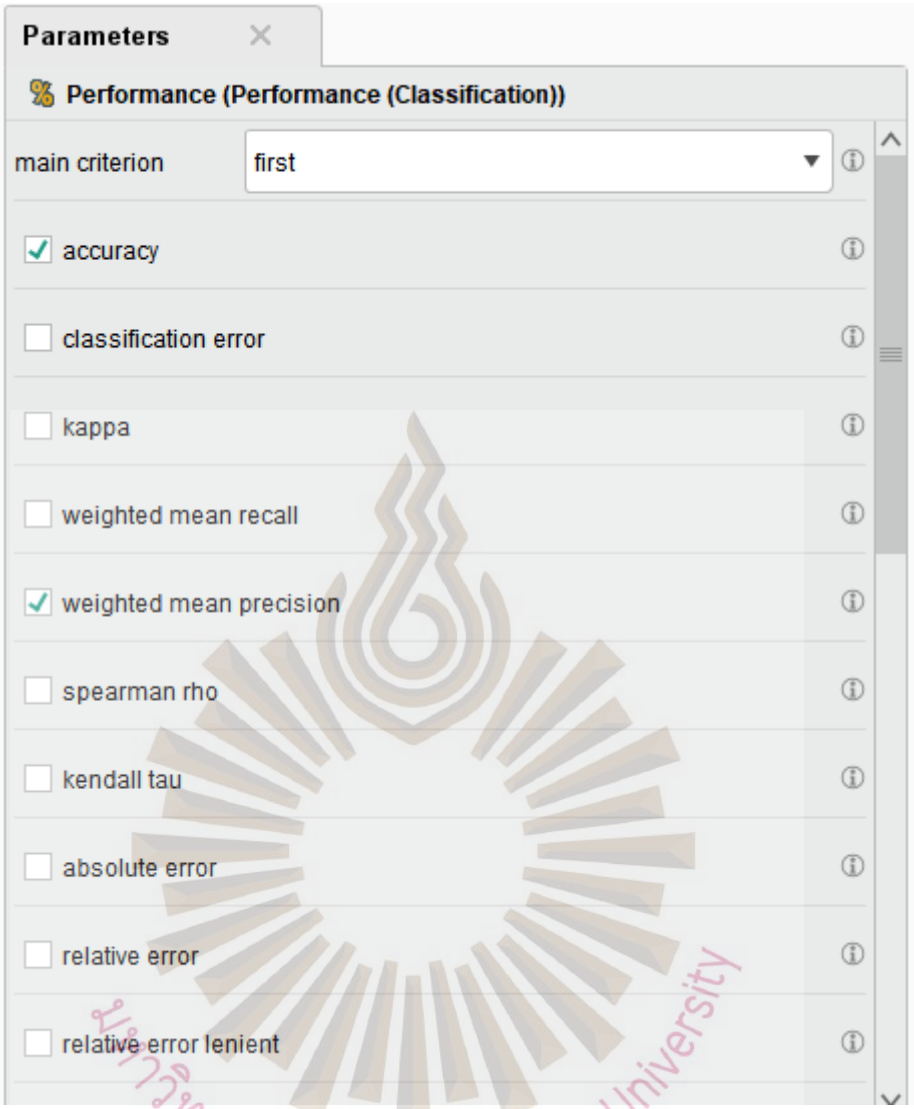
Main Criterion = First

Accuracy = Turn

Weighted Mean Precision = Turn

รูปที่ 4.20 Operators Performance

มหาวิทยาลัยรังสิต Rangsit University



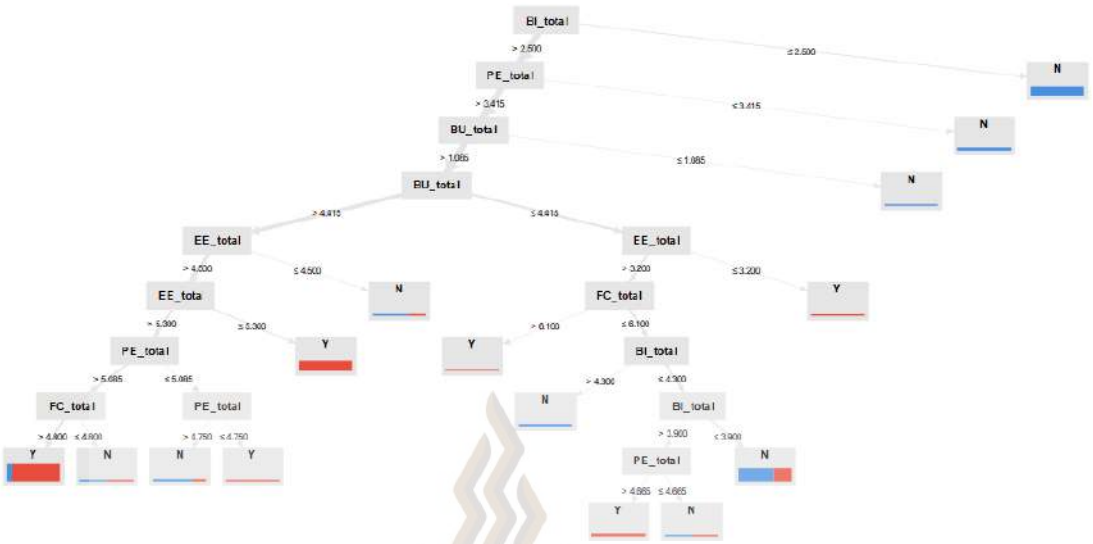
รูปที่ 4.21 การ Set Parameters Operators Performance

การแสดงผลการพยากรณ์โดยใช้เทคนิคการจำแนกข้อมูลด้วย เทคนิคต้นไม้ตัดสินใจ ได้ผลดังรูปที่ 4.22

accuracy: 74.29% +/- 12.56% (micro average: 74.21%)

	true N	true Y	class precision
pred. N	53	26	67.09%
pred. Y	15	65	81.25%
class recall	77.94%	71.43%	

รูปที่ 4.22 ผลการพยากรณ์แบบ Table View โดยเทคนิคต้นไม้การตัดสินใจ



รูปที่ 4.23 ผลการพยากรณ์แบบ Chart View โดยเทคนิคต้นไม้การตัดสินใจ

หลังจากทำการพยากรณ์ด้วยแบบจำลองโดยใช้เทคนิคการจำแนกข้อมูลด้วยวิธีต้นไม้การตัดสินใจ สรุปได้ผลดังตารางที่ 4.1

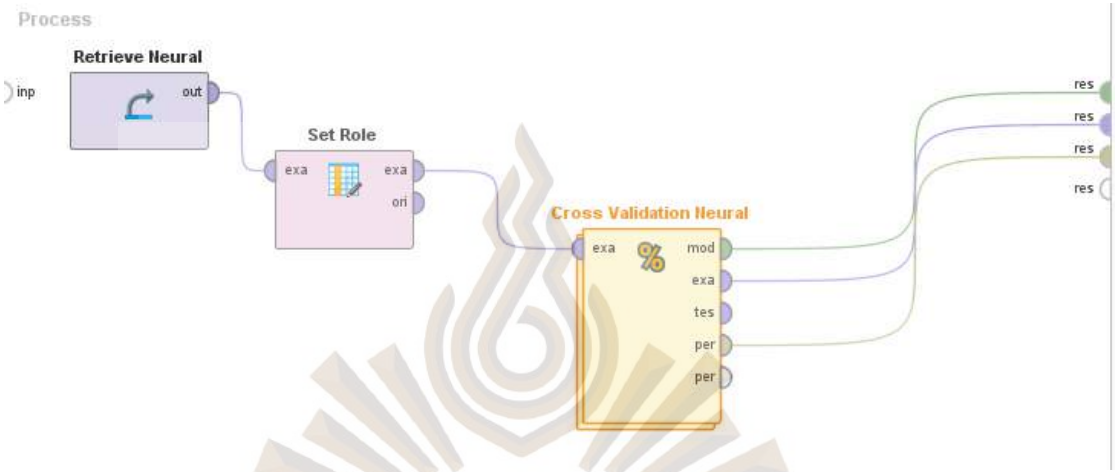
ตารางที่ 4.1 สรุปผลการพยากรณ์ด้วยแบบจำลองโดยใช้เทคนิคการจำแนกข้อมูลด้วยวิธีต้นไม้การตัดสินใจ

Model	Accuracy	Class Recall (N)	Class Recall (Y)	Class Precision (N)	Class Precision (Y)
Decision Tree	74.29 %	77.94 %	71.43 %	67.09 %	81.25 %

แบบจำลองการพยากรณ์โดยใช้เทคนิคการจำแนกข้อมูลด้วยวิธีต้นไม้การตัดสินใจ มีค่าความแม่นยำ (Accuracy) 74.29% ความถูกต้องของการพยากรณ์การเลือกไม่ใช้แอปพลิเคชันต่อเท่ากับ 77.94% ความถูกต้องของการพยากรณ์การเลือกใช้งานแอปพลิเคชันต่อ เท่ากับ 71.43%

3) กระบวนการสร้างแบบจำลองการพยากรณ์โดยใช้เทคนิคการจำแนกข้อมูลด้วยวิธีเครือข่ายประสาทเทียม จะประกอบไปด้วย Operators ดังต่อไปนี้ Neural Network เป็นการจำลอง

การพยากรณ์โดยใช้เทคนิคการจำแนกข้อมูลด้วยวิธีต้นไม้การตัดสินใจ Apply Model เป็นการสร้างแบบจำลองการพยากรณ์ที่สร้างได้ไปใช้ในการพยากรณ์ เพื่อให้ได้คำตอบของข้อมูลที่ยังไม่ทราบคลาส Performance เป็นการวัดประสิทธิภาพของแบบจำลองของการพยากรณ์ จะได้ค่า Accuracy Class Recall Precision Recall เป็นต้น ดังรูปที่ 4.24



รูปที่ 4.24 Operators หลักสำหรับ Neural Network

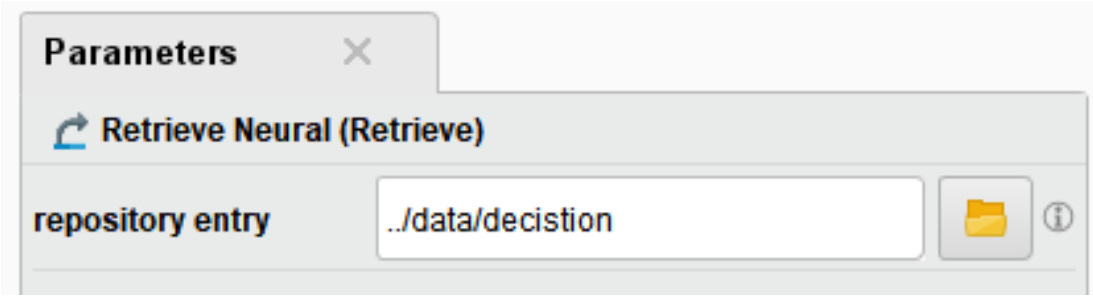
รายละเอียด Operators

- Retrieve Neural ทำการ Set Parameters ดังรูปที่ 4.25 และ 4.26

Repository Entry ให้ทำการดึงข้อมูลมาจากฐานข้อมูลในโปรแกรม RapidMiner

รูปที่ 4.25 Operators Retrieve Neural

มหาวิทยาลัยรังสิต Rangsit University



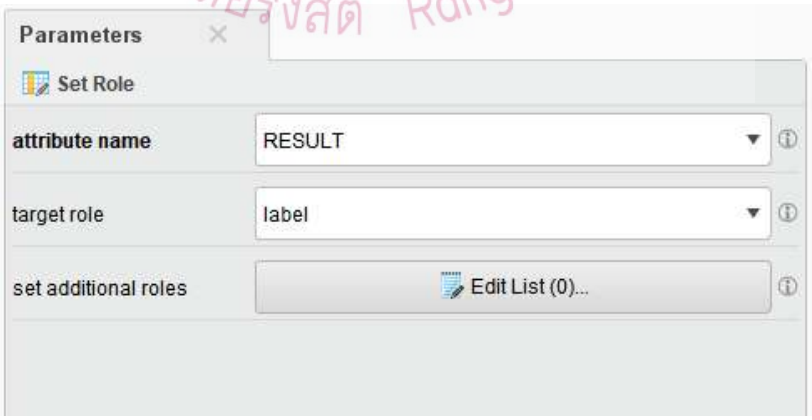
รูปที่ 4.26 การ Set Parameters Retrieve Neural

Set Role ทำการ Set Parameters ดังรูป 4.27 และ 4.28

Attribute Name = Filed Result

Target Role = Label

รูปที่ 4.27 Operators Set Role



รูปที่ 4.28 การ Set Parameters Operators Set Role

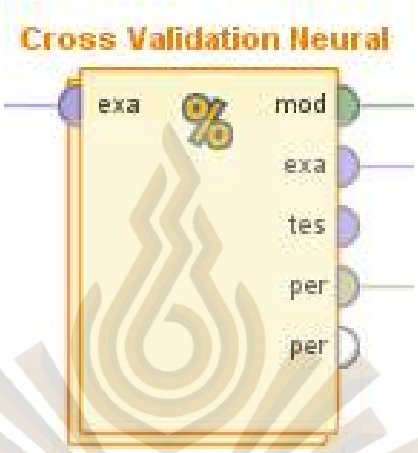
มหาวิทยาลัยรังสิต Rangsit University

Cross Validation Neural ทำการ Set Parameters ดังรูปที่ 4.29 และ 4.30

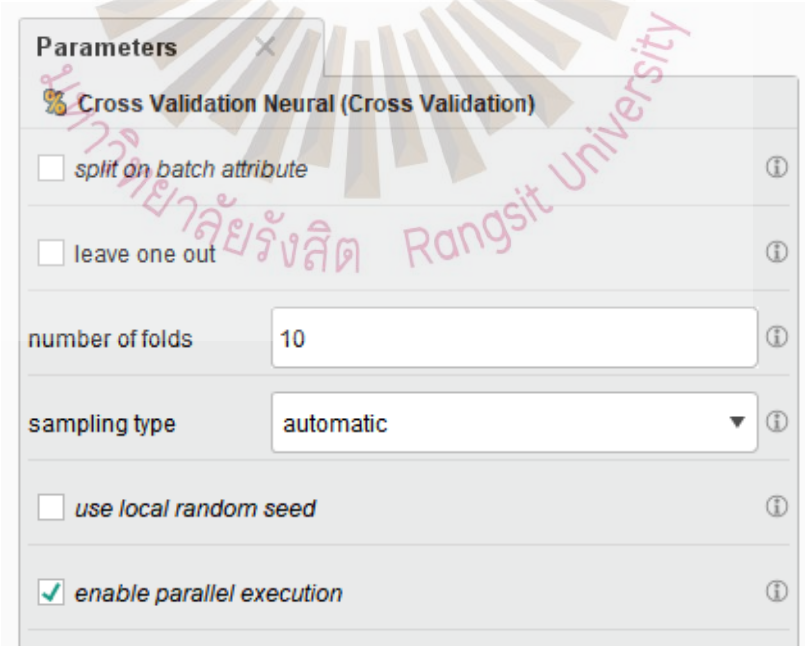
Number of Folds = 10

Sampling Type = Automatic

Enable Parallel Execution = Turn

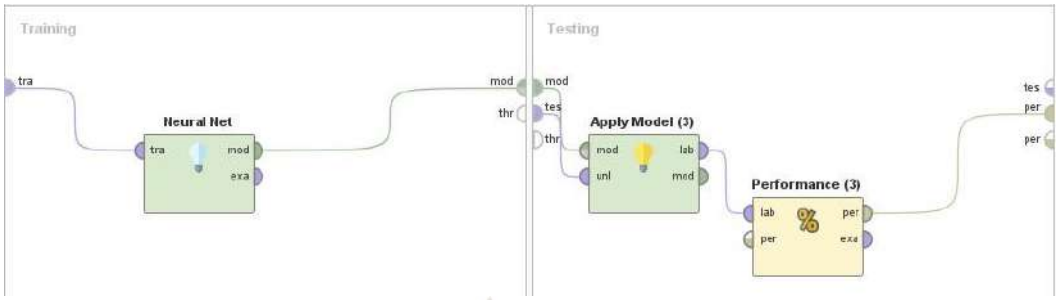


รูปที่ 4.29 Operators Cross Validation Neural



รูปที่ 4.30 Set Parameters Cross Validation Neural

ทำการ Set Operators ภายใน Operators Cross Validation Neural ดังรูปที่ 4.31



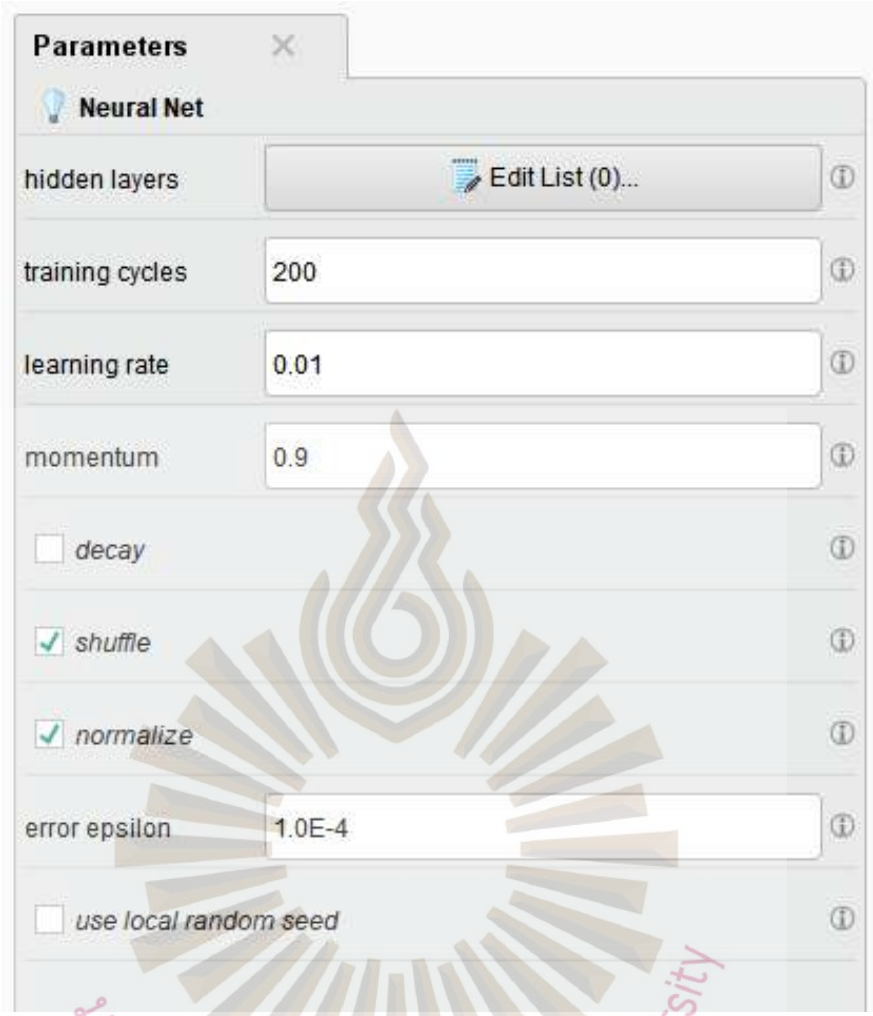
รูปที่ 4.31 Operators ภายใน Operators Neural Network

Neural Net set parameters ดังรูปที่ 4.32 และ 4.33

Training cycles	=	200
Learning rate	=	0.01
Momentum	=	0.9
Error epsilon	=	1.0E-4

รูปที่ 4.32 Operators Neural Net

มหาวิทยาลัยรังสิต Rangsit University

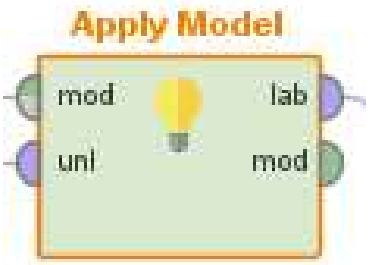


รูปที่ 4.33 การ Set Parameters Operators Neural Net

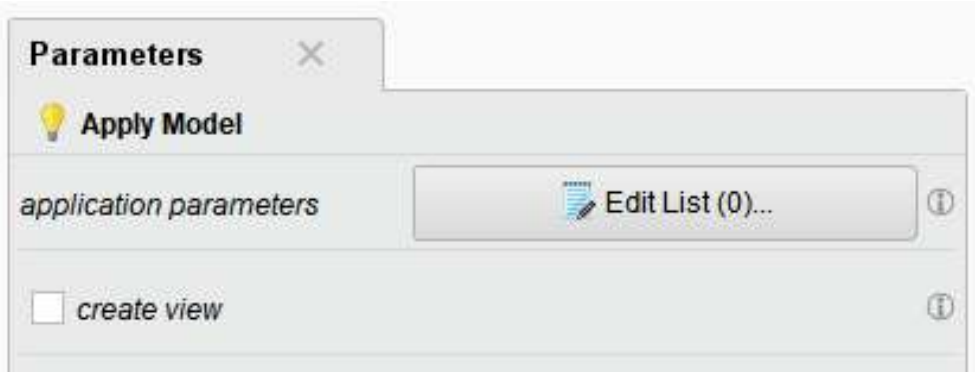
มหาวิทยาลัยรังสิต Rangsit University

- Apply Model set parameters ดังรูปที่ 4.34 และ 4.35

  ○ No set



รูปที่ 4.34 Operators Apply Model



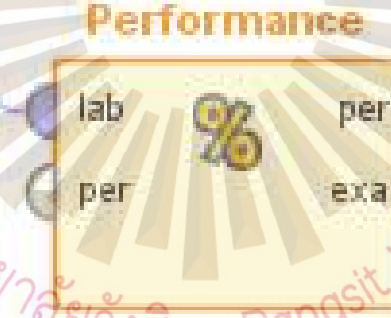
รูปที่ 4.35 การ Set Parameters operators Apply Model

Performance Set Parameters ดังรูปที่ 4.36 และ 4.37

Main Criterion = First

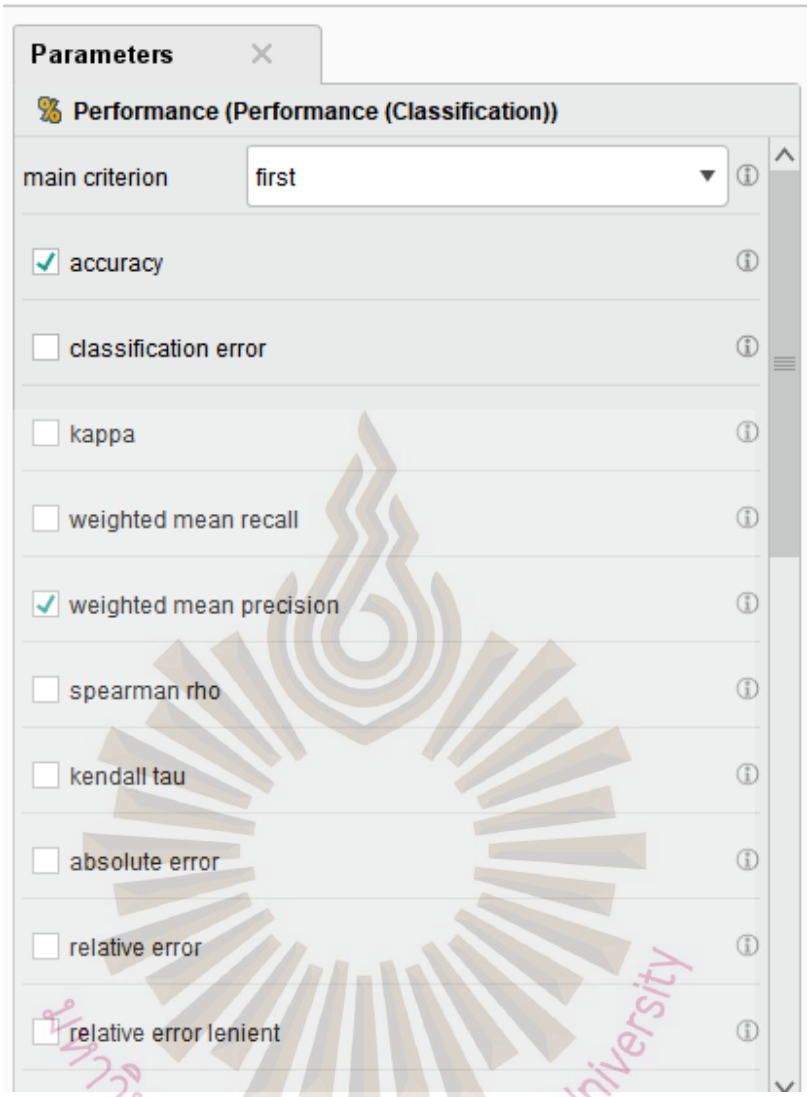
Accuracy = Turn

Weighted Mean Precision = Turn



รูปที่ 4.36 Operators Performance

มหาวิทยาลัยรังสิต Rangsit University



รูปที่ 4.37 การ Set Parameters Operators Performance

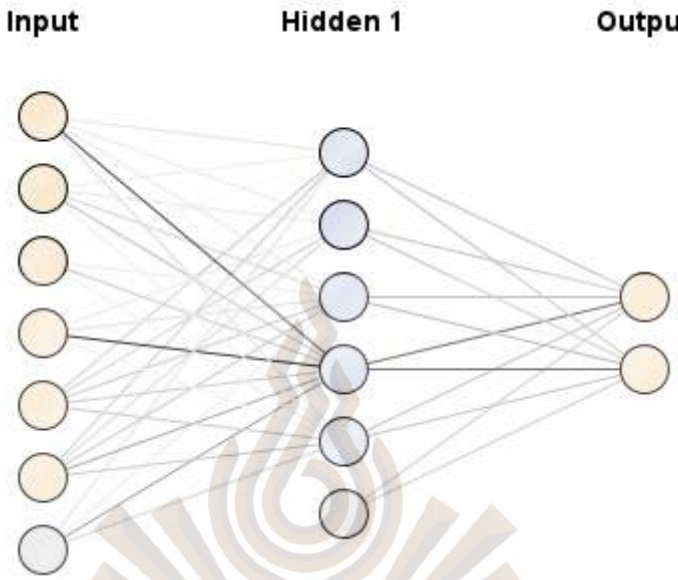
แสดงผลการทำนายโดยใช้เทคนิคเครือข่ายประสาทเทียม แบบ Table View ดังรูปที่ 4.38

accuracy: 78.67% +/- 12.51% (micro average: 78.62%)

	true N	true Y	class precision
pred. N	50	16	75.76%
pred. Y	18	75	80.65%
class recall	73.53%	82.42%	

รูปที่ 4.38 ผลการพยากรณ์แบบ Table View โดยเทคนิคเครือข่ายประสาทเทียม

แสดงผลการทำนายโดยใช้เทคนิคเครือข่ายประสาทเทียม แบบ Chart View ดังรูปที่ 4.39



รูปที่ 4.39 ผลการพยากรณ์แบบ Chart View โดยเทคนิคเครือข่ายประสาทเทียม

หลังจากทำการพยากรณ์ด้วยแบบจำลองโดยใช้เทคนิคการจำแนกข้อมูลด้วยวิธีเครือข่ายประสาทเทียม (Neural Network) สรุปได้ผลดังตารางที่ 4.2

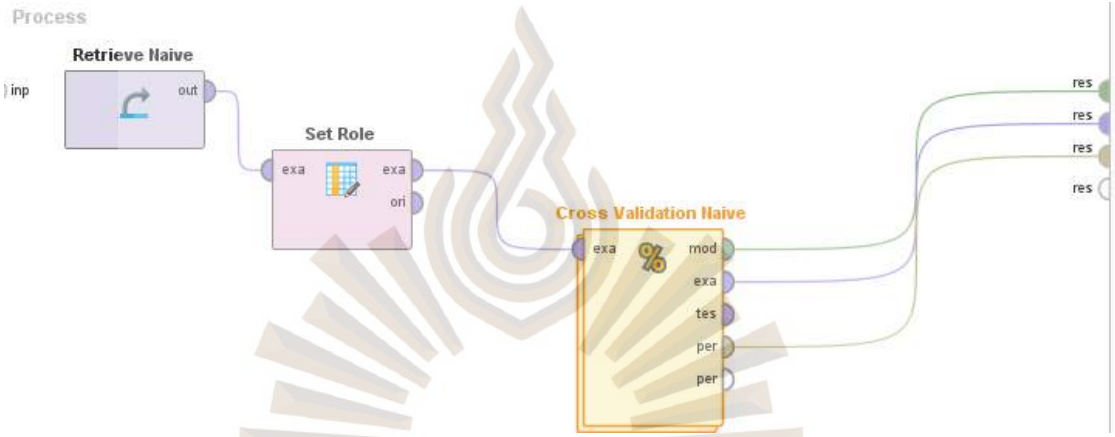
ตารางที่ 4.2 สรุปผลการพยากรณ์ด้วยแบบจำลองโดยใช้เทคนิคการจำแนกข้อมูลด้วยวิธีเครือข่ายประสาทเทียม

Model	Accuracy	Class Recall (N)	Class Recall (Y)	Class Precision (N)	Class Precision (Y)
Neural Network	78.67 %	73.53 %	82.42 %	75.76 %	80.65 %

แบบจำลองการพยากรณ์โดยใช้เทคนิคการจำแนกข้อมูลด้วยวิธีเครือข่ายประสาทเทียม (Neural Network) มีค่าความแม่นยำ (Accuracy) 78.67% ความถูกต้องของการพยากรณ์การเลือกไม่ใช้แอปพลิเคชันต่อ เท่ากับ 73.53% ความถูกต้องของการพยากรณ์การเลือกใช้งานแอปพลิเคชันต่อ เท่ากับ 82.42%

มหาวิทยาลัยรังสิต Rangsit University

4) กระบวนการสร้างแบบจำลองการพยากรณ์โดยใช้เทคนิคการจำแนกข้อมูลด้วยวิธีไนอีฟเบย์ จะประกอบไปด้วย Operators ดังต่อไปนี้ Naïve Bayes เป็นการจำลองการพยากรณ์โดยใช้เทคนิคการจำแนกข้อมูลด้วยวิธีต้นไม้การตัดสินใจ Apply Model เป็นการสร้างแบบจำลองการพยากรณ์ที่สร้างได้ไปใช้ในการพยากรณ์ เพื่อให้ได้คำตอบของข้อมูลที่ยังไม่ทราบคลาส Performance เป็นการวัดประสิทธิภาพของแบบจำลองของการพยากรณ์ จะได้ค่า Accuracy Class Recall Precision Recall ดังรูปที่ 4.40



รูปที่ 4.40 Operators หลักสำหรับ Naïve Bayes

รายละเอียด Operators

Retrieve Naïve ทำการ Set Parameters ดังรูปที่ 4.41 และ 4.42

Repository Entry ให้ทำการดึงข้อมูลมาจากฐานข้อมูลในโปรแกรม RapidMiner

รูปที่ 4.41 Operators Retrieve Naïve

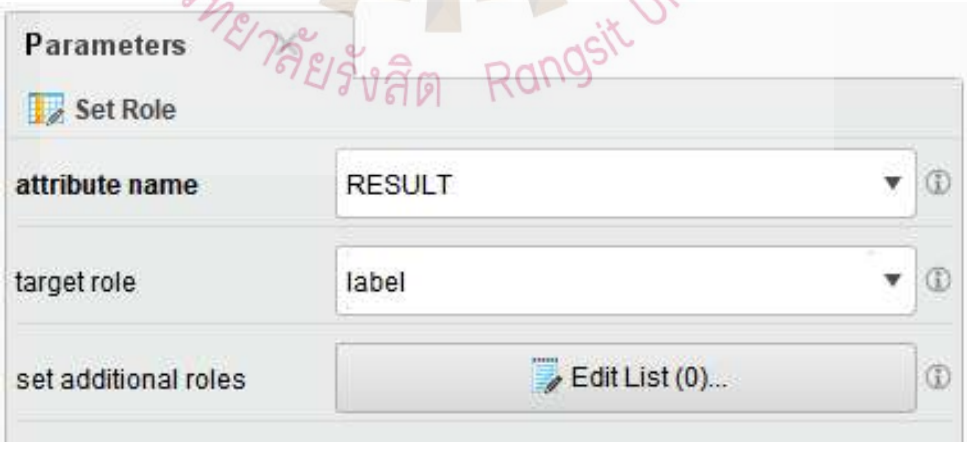
รูปที่ 4.42 Set Parameters Operators Retrieve Naïve

Set Role ทำการ Set Parameters ดังรูปที่ 4.43 และ 4.44

Attribute Name = Filed Result

Target Role = Label

รูปที่ 4.43 Operators Set Role



รูปที่ 4.44 Set Parameters Operators Set Role

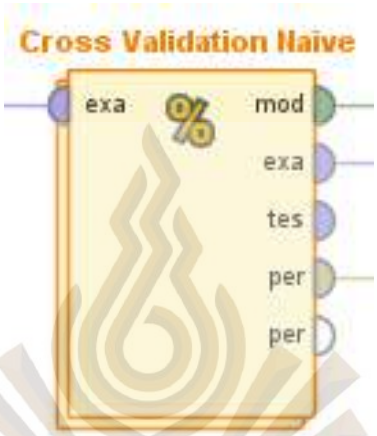
มหาวิทยาลัยรังสิต Rangsit University

Cross Validation Decision ทำการ Set Parameters ดังรูปที่ 4.45 และ 4.46

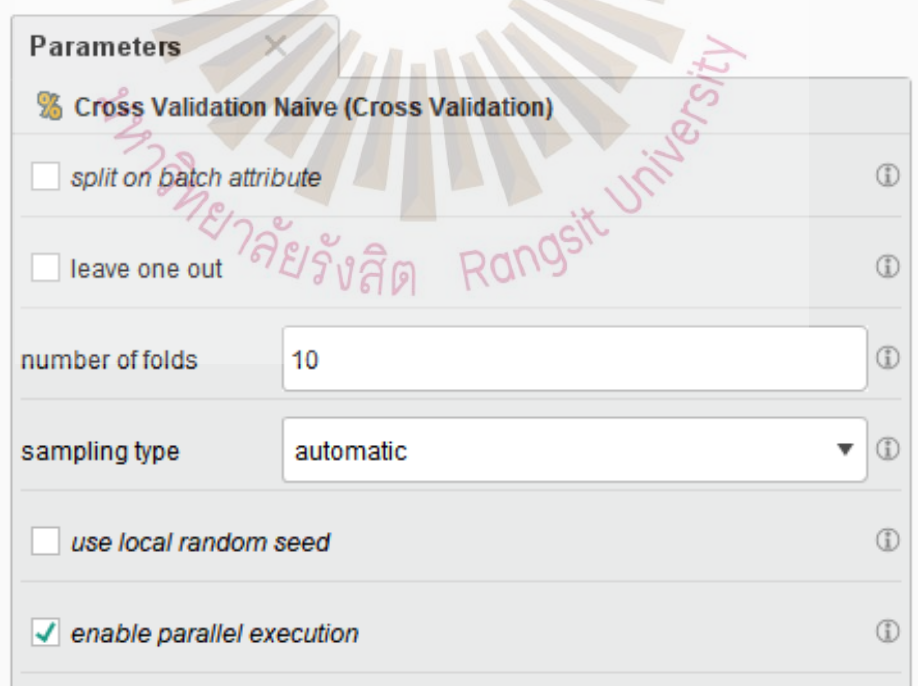
Number of Folds = 10

Sampling Type = Automatic

Enable Parallel Execution = Turn



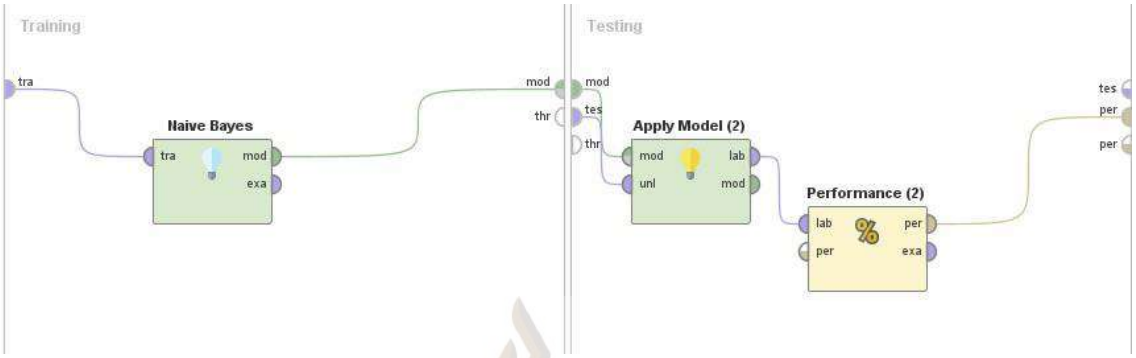
รูปที่ 4.45 Operators Cross Validation Naïve



รูปที่ 4.46 Set Parameters Operators Cross Validation Naïve

มหาวิทยาลัยรังสิต Rangsit University

Operators ภายใน Operators Cross Validation Naïve ดังรูปที่ 4.47

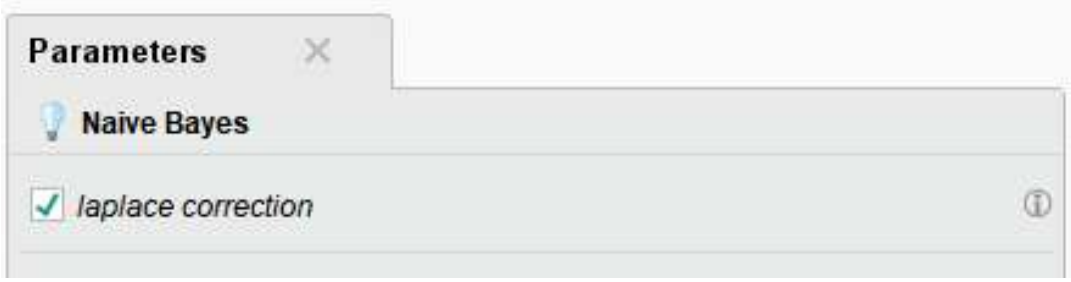


รูปที่ 4.47 Operators ภายใน Operators Cross Validation Naïve

Naïve Bayes Set Parameters ดังรูปที่ 4.48 และ 4.49

Laplace Correction = Ture

รูปที่ 4.48 Operators Naïve Bayes

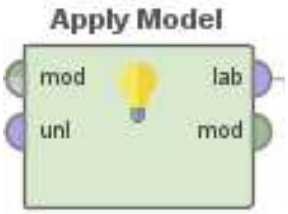


รูปที่ 4.49 Set Parameters Operators Naïve Bayes

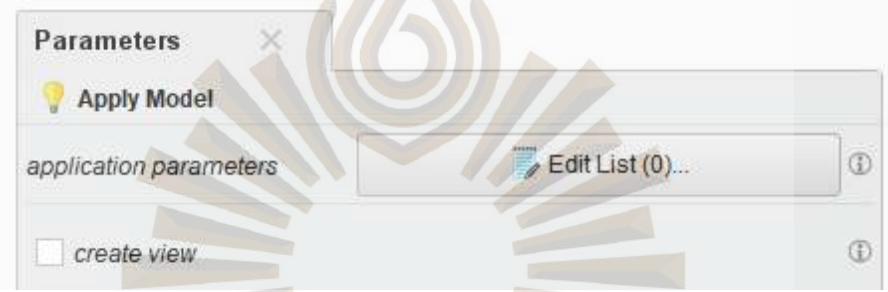
มหาวิทยาลัยรังสิต Rangsit University

Apply Model Set Parameters ดังรูปที่ 4.50 และ 4.51

No Set



รูปที่ 4.50 Operators Apply Model



รูปที่ 4.51 Set Parameters Operators Apply Model

Performance set parameters ดังนี้

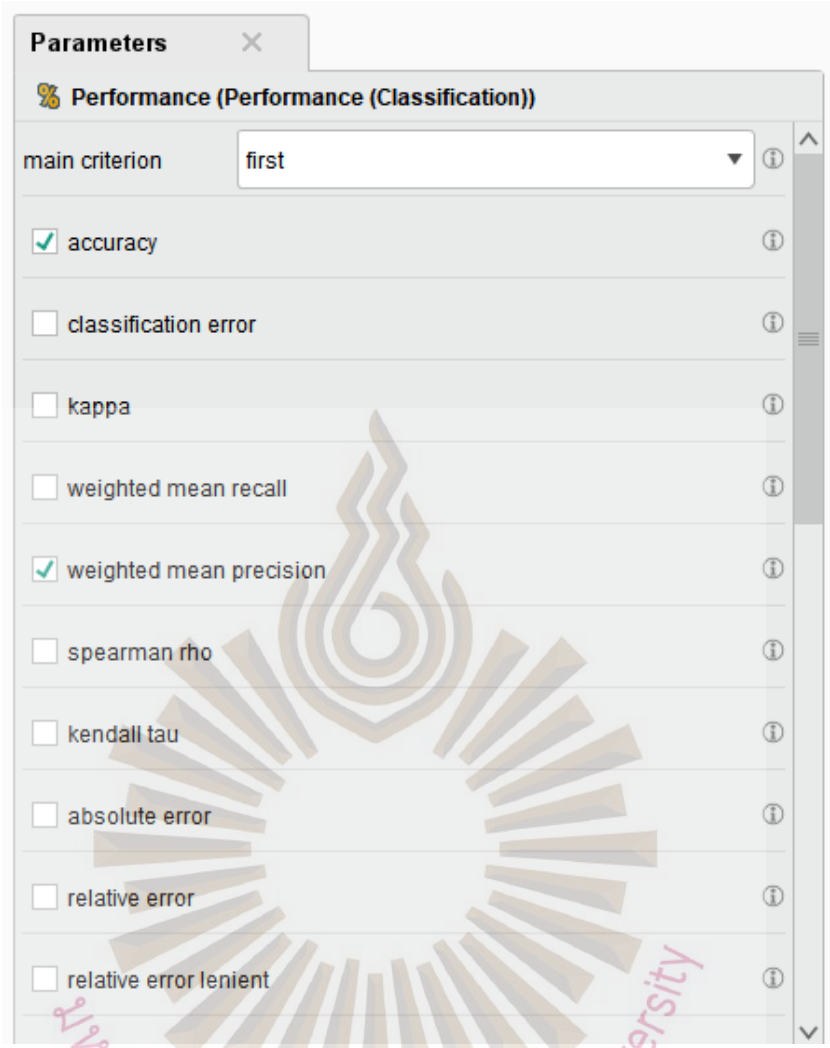
Main Criterion = First

Accuracy = Turn

Weighted Mean Precision = Turn

รูปที่ 4.52 Operators Performance

มหาวิทยาลัยรังสิต Rangsit University



รูปที่ 4.53 Set Parameters Operators Performance

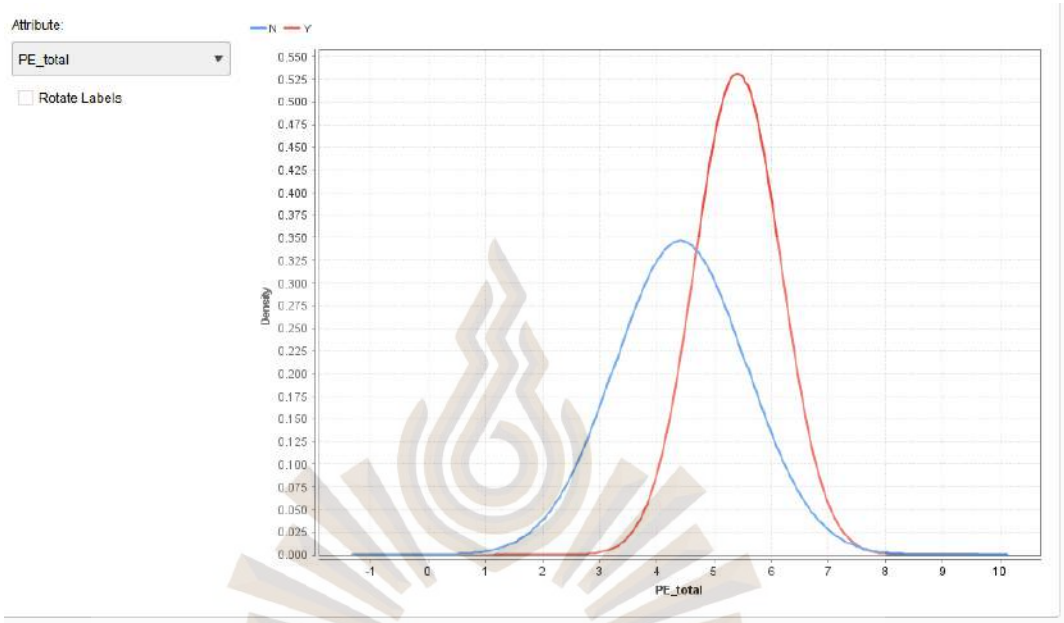
แสดงสรุปผลการพยากรณ์โดยใช้เทคนิคการจำแนกข้อมูลด้วยวิธีไนอีฟเบย์แบบ Table View

accuracy: 78.67% +/- 11.03% (micro average: 78.62%)

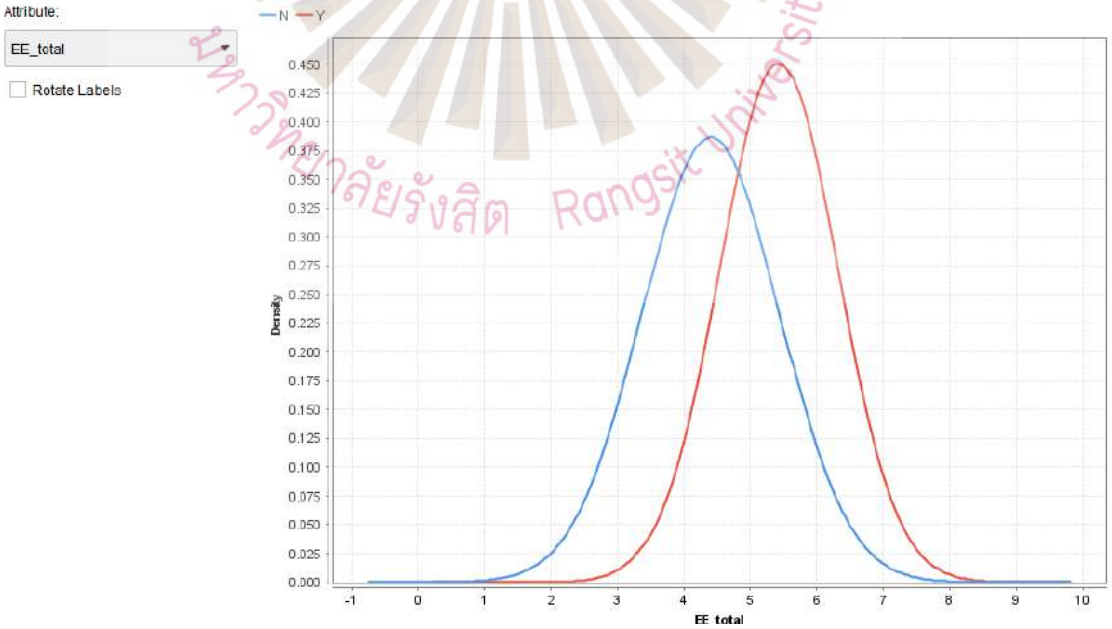
	true N	true Y	class precision
pred. N	51	17	75.00%
pred. Y	17	74	81.32%
class recall	75.00%	81.32%	

รูปที่ 4.54 ผลการพยากรณ์แบบ Table View โดยเทคนิค ไนอีฟเบอร์

แสดงผลการทำนายโดยใช้เทคนิคการจำแนกข้อมูลด้วยวิธีไนอีฟเบย์แบบ Chart View ดังรูปที่ 4.55

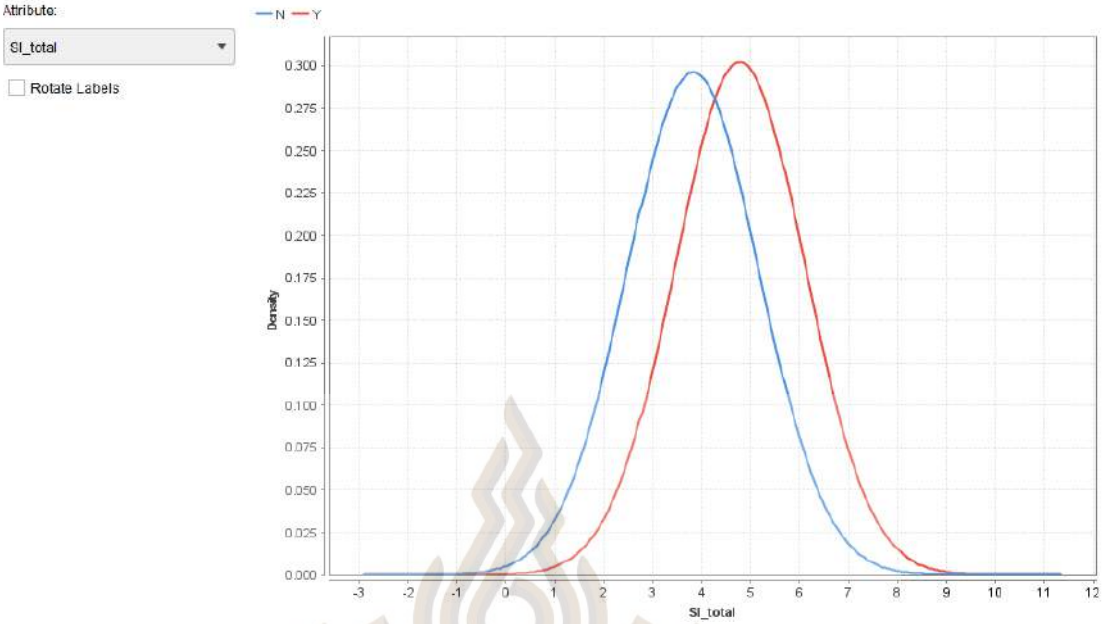


รูปที่ 4.55 ผลการพยากรณ์หัวข้อ PE แบบ Chart View โดยเทคนิคไนอีฟเบอร์

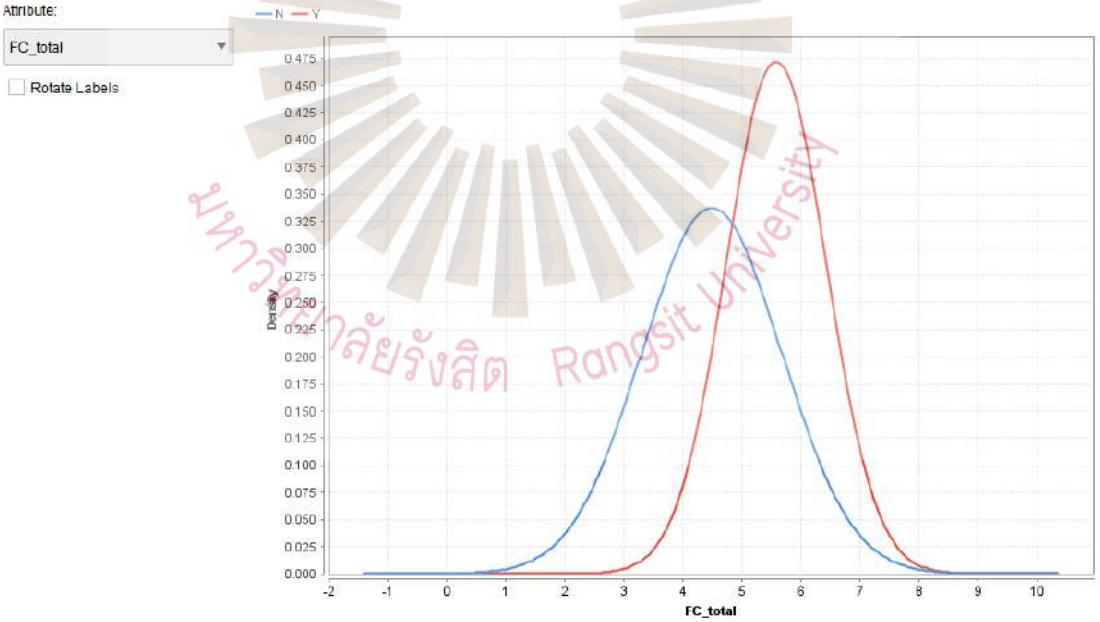


รูปที่ 4.56 ผลการพยากรณ์หัวข้อ EE แบบ Chart View โดยเทคนิคไนอีฟเบอร์

มหาวิทยาลัยรังสิต Rangsit University

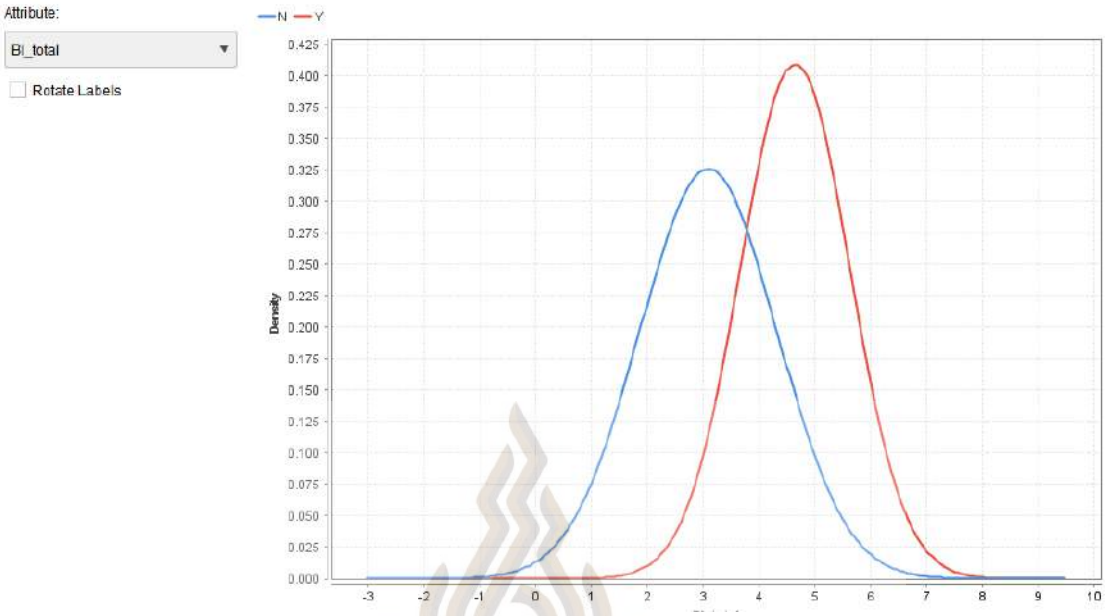


รูปที่ 4.57 ผลการพยากรณ์หัวข้อ SI แบบ Chart View โดยเทคนิคไนอีฟเบอร์

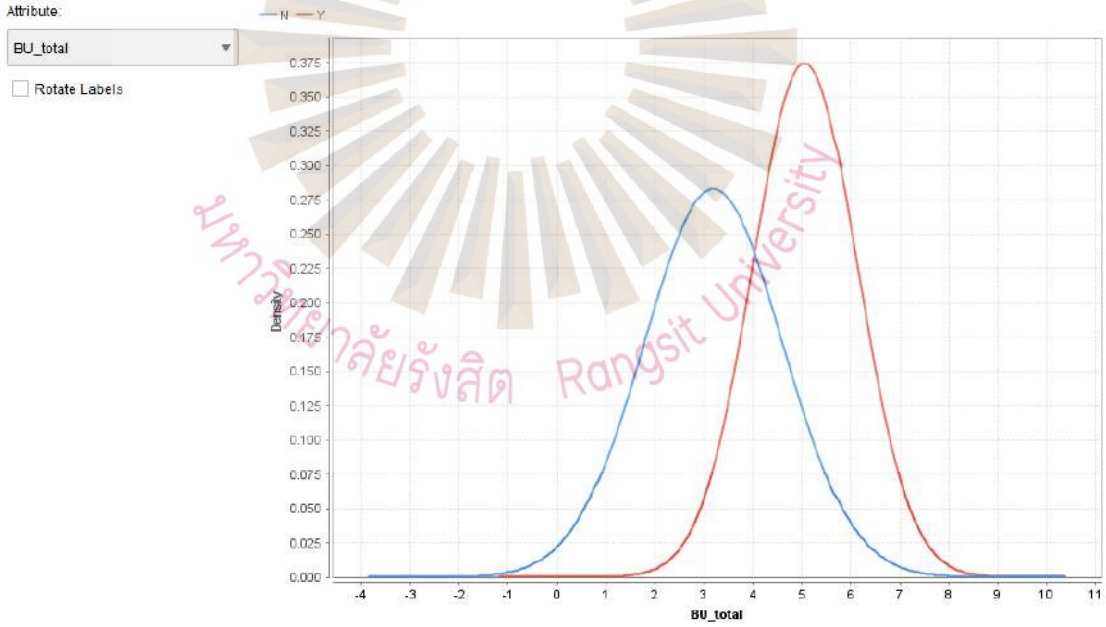


รูปที่ 4.58 ผลการพยากรณ์หัวข้อ FC แบบ Chart View โดยเทคนิคไนอีฟเบอร์

มหาวิทยาลัยรังสิต Rangsit University



รูปที่ 4.59 ผลการพยากรณ์หัวข้อ BI แบบ Chart View โดยเทคนิคไนอีฟเบอร์



รูปที่ 4.60 ผลการพยากรณ์หัวข้อ BU แบบ Chart View โดยเทคนิคไนอีฟเบอร์

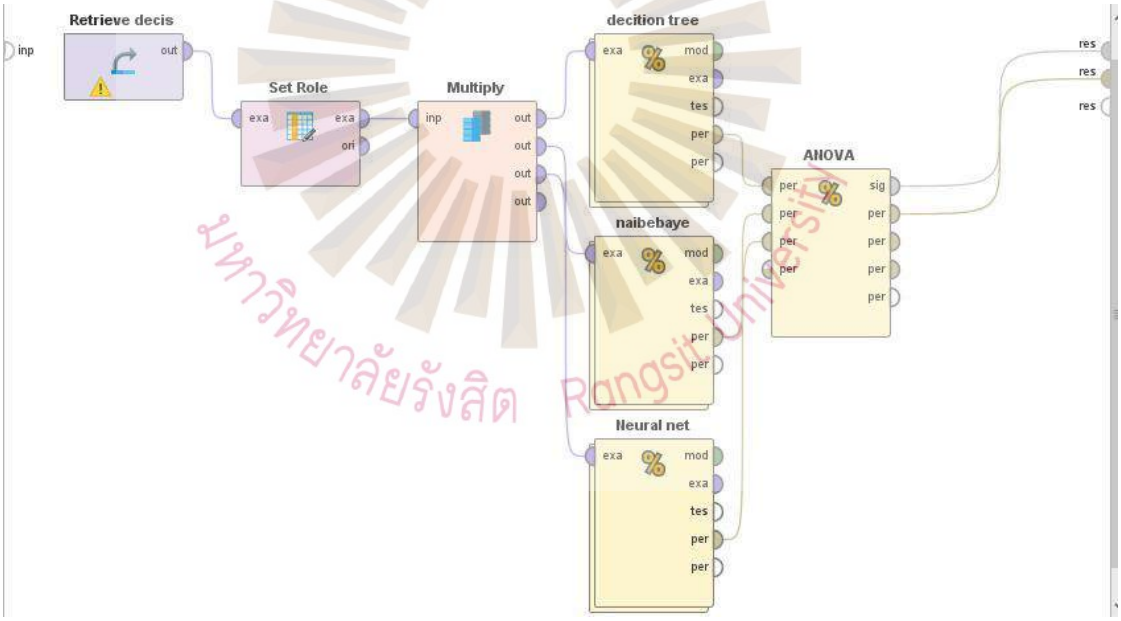
หลังจากทำการพยากรณ์ด้วยแบบจำลองโดยใช้เทคนิคการจำแนกข้อมูลด้วยวิธีไนอีฟเบย์ (Naïve Bayes) สรุปได้ผลดังตารางที่ 4.3

ตารางที่ 4.3 สรุปผลการพยากรณ์ด้วยแบบจำลองโดยใช้เทคนิคการจำแนกข้อมูลด้วยวิธีไนอีฟเบย์

Model	Accuracy	Class Recall (N)	Class Recall (Y)	Class Precision (N)	Class Precision (Y)
Neural Network	78.67 %	75.00 %	81.32 %	75.00 %	81.32 %

แบบจำลองการพยากรณ์โดยใช้เทคนิคการจำแนกข้อมูลด้วยวิธีไนอีฟเบย์ (Naïve Bayes) มีค่าความแม่นยำ (Accuracy) 78.67% ความถูกต้องของการพยากรณ์การเลือกไม่ใช้แอปพลิเคชันต่อ เท่ากับ 75.00% ความถูกต้องของการพยากรณ์การเลือกใช้งานแอปพลิเคชันต่อ เท่ากับ 81.32%

ตอนที่ 2 เป็นการใช้สถิติ ANOVA เพื่อทำการเปรียบเทียบประสิทธิภาพของแบบจำลองของแต่ล่ะเทคนิคการจำแนกข้อมูล ดังที่กล่าวมาข้างต้น ดังรูปที่ 4.61

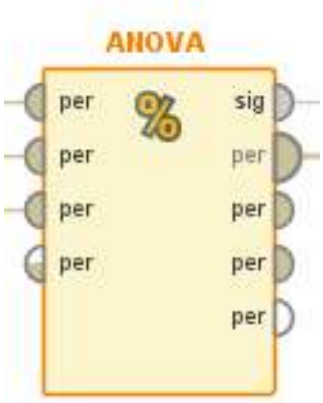


รูปที่ 4.61 Operators หลักที่ใช้ในการเปรียบเทียบโดยเทคนิค ANOVA

รายละเอียด Operaorts

ANOVA ทำการ Set Parameters ดังรูปที่ 4.62 และ 4.63

Alpha = 0.05



รูปที่ 4.62 Operators ANOVA

รูปที่ 4.63 Set Parameters Operators ANOVA

แสดงผลการเปรียบเทียบ 3 เทคนิคโดยใช้ ANOVA แสดงค่าแบบ Table View ดังรูปที่ 4.64

Source	Square Sums	DF	Mean Squares	F	Prob
Between	0.010	2	0.005	0.332	0.721
Residuals	0.414	27	0.015		
Total	0.424	29			

รูปที่ 4.64 ผลการเปรียบเทียบ 3 เทคนิคโดยใช้ ANOVA แสดงค่าแบบ Table View

การเปรียบเทียบแบบจำลองการพยากรณ์ทั้ง 3 แบบ และใช้สถิติ ANOVA ทำการเปรียบเทียบแบบจำลองการพยากรณ์ได้ผลดังตารางที่ 4.4

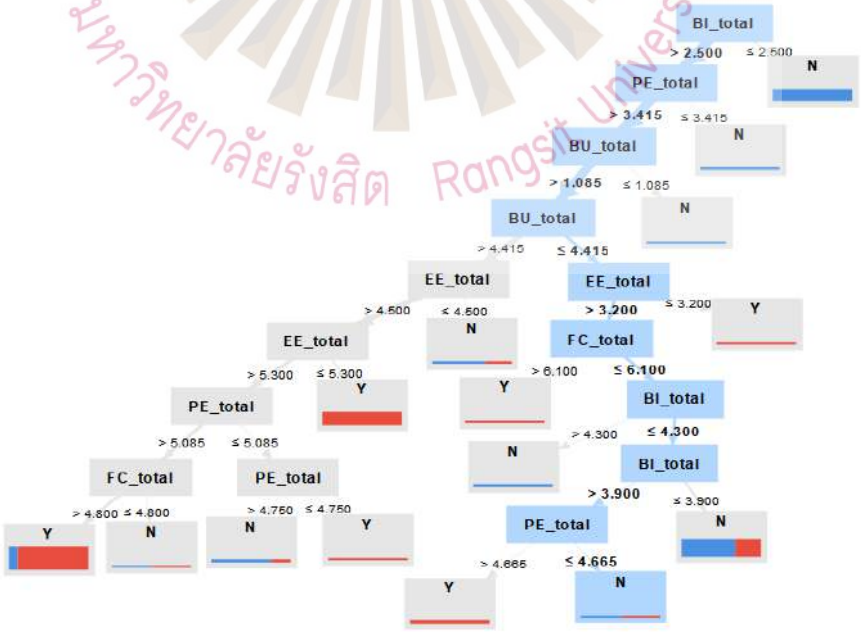
มหาวิทยาลัยรังสิต Rangsit University

ตารางที่ 4.4 สรุปผลการเปรียบเทียบแบบจำลองการพยากรณ์ด้วย ANOVA

Source	Square Sums	DF	Mean Squares	F	Prob
Between	0.010	2	0.005	0.332	0.721
Residuals	0.414	27	0.015		
Total	0.424	29			

จากตารางที่ 4.4 ผลการเปรียบเทียบแบบจำลองการพยากรณ์ด้วยสถิติ ANOVA พบว่า แบบจำลองการพยากรณ์โดยใช้เทคนิคต้นไม้การตัดสินใจ เครือข่ายประสาทเทียม และไนอีฟเบย์ มีค่า F เท่ากับ 0.332 และค่า Prob เท่ากับ 0.721 ซึ่งไม่แตกต่างกันอย่างมีนัยสำคัญที่ระดับ 0.05 แสดงว่า แบบจำลองทั้ง 3 แบบ ไม่แตกต่างกันมากอย่างมีนัยสำคัญ ดังนั้นผลสรุปได้ว่า สามารถเลือกใช้แบบจำลองการพยากรณ์แบบใดก็ได้ในการทำวิจัย ซึ่งผู้วิจัยได้เลือกเทคนิค Decision Tree เพราะสามารถอธิบายให้ผู้สนใจเข้าใจได้ชัดเจน

ตอนที่ 3 การพยากรณ์และความสัมพันธ์ของการยอมรับการใช้เทคโนโลยีที่มีผลต่อพฤติกรรมการใช้แอปพลิเคชัน LAZADA ช้อปปิ้งบนสมาร์ทโฟนของประชากรในเขตกรุงเทพมหานครและปริมณฑล โดยใช้เทคนิคการจำแนกข้อมูลด้วยวิธีต้นไม้การตัดสินใจ ดังรูปที่ 4.65



รูปที่ 4.65 การพยากรณ์โดยใช้ Decision Tree

จากรูปที่ 4.65 ซึ่งสามารถอธิบายได้ดังนี้

1) IF BI_Total <= 2.500 THEN Result = N

2) IF BI_Total > 2.500 AND PE_total <= 3.415 THEN Result = N

3) IF BI_Total > 2.500 AND PE_total > 3.415 AND BU_total <= 1.085 THEN Result = N

4) IF BI_Total > 2.500 AND PE_total > 3.415 AND BU_total between 1.086 to 4.415 AND EE_total <= 3.200 THEN Result = Y

5) IF BI_Total > 2.500 AND PE_total > 3.415 AND BU_total between 1.086 to 4.415 AND EE_total > 3.200 AND FC_total > 6.100 THEN Result = Y

6) IF BI_Total > 2.500 AND PE_total > 3.415 AND BU_total between 1.086 to 4.415 AND EE_total > 3.200 AND FC_total <= 6.100 AND BI_Total > 4.300 THEN Result = N

7) IF BI_Total > 2.500 AND PE_total > 3.415 AND BU_total between 1.086 to 4.415 AND EE_total > 3.200 AND FC_total <= 6.100 AND BI_Total between 3.901 to 4.300 AND PE_total > 4.665 THEN Result = Y

8) IF BI_Total > 2.5 AND PE_total > 3.415 AND BU_total > 4.415 AND EE_total <= 4.500 THEN Result = N

9) IF BI_Total > 2.5 AND PE_total > 3.415 AND BU_total > 4.415 AND EE_total between 4.501 to 5.300 THEN Result = Y

10) IF BI_Total > 2.5 AND PE_total > 3.415 AND BU_total > 4.415 AND EE_total between 4.501 to 5.300 AND PE_total > 5.085 AND FC_total > 4.800 THEN Result = Y

11) IF BI_Total > 2.5 AND PE_total > 3.415 AND BU_total > 4.415 AND EE_total between 4.501 to 5.300 AND PE_total > 5.085 AND FC_total <= 4.800 THEN Result = N

12) IF BI_Total > 2.5 AND PE_total > 3.415 AND BU_total > 4.415 AND EE_total between 4.501 to 5.300 AND PE_total between 4.750 to 5.085 THEN Result = N

13) IF BI_Total > 2.5 AND PE_total > 3.415 AND BU_total > 4.415 AND EE_total between 4.501 to 5.300 AND PE_total <= 4.501 THEN Result = Y

ผลการวิจัยการพยากรณ์และความสัมพันธ์ของการยอมรับการใช้เทคโนโลยีที่มีผลต่อพฤติกรรมการใช้แอปพลิเคชัน LAZADA ช้อปปิ้งบนสมาร์ทโฟนของประชากรในเขตกรุงเทพมหานครและปริมณฑล โดยใช้เทคนิคการจำแนกข้อมูลด้วยวิธีต้นไม้การตัดสินใจ พบว่า กลุ่มตัวอย่างที่ใช้งานแอปพลิเคชันช้อปปิ้งบนสมาร์ทโฟนต่อเนื่อง มีดังนี้

1) คะแนนพฤติกรรมความตั้งใจในการใช้งานมากกว่า 2.500 และ ความคาดหวังในประสิทธิภาพมากกว่า 3.415 และ พฤติกรรมการใช้งานอยู่ระหว่าง 1.086 ถึง 4.415 และ ความพยายามใช้งานน้อยกว่าหรือเท่ากับ 3.200

2) คะแนนพฤติกรรมความตั้งใจในการใช้งานมากกว่า 2.500 และ ความคาดหวังในประสิทธิภาพมากกว่า 3.415 และ พฤติกรรมการใช้งานอยู่ระหว่าง 1.086 ถึง 4.415 และ ความพยายามใช้งานมากกว่า 3.200 และสภาพสิ่งอำนวยความสะดวกในการใช้งานมากกว่า 6.100

3) คะแนนพฤติกรรมความตั้งใจในการใช้งานอยู่ระหว่าง 3.901 ถึง 4.300 และ ความคาดหวังในประสิทธิภาพมากกว่า 4.665 และ พฤติกรรมการใช้งานอยู่ระหว่าง 1.086 ถึง 4.415 และ ความพยายามใช้งานมากกว่า 3.200 และ สภาพสิ่งอำนวยความสะดวกในการใช้งานน้อยกว่า 6.1

4) คะแนนพฤติกรรมความตั้งใจในการใช้งานมากกว่า 2.500 และ ความคาดหวังประสิทธิภาพการใช้งานมากกว่า 3.415 และ พฤติกรรมการใช้งานมากกว่า 4.415 และ ความพยายามใช้งานอยู่ระหว่าง 4.501 ถึง 5.300

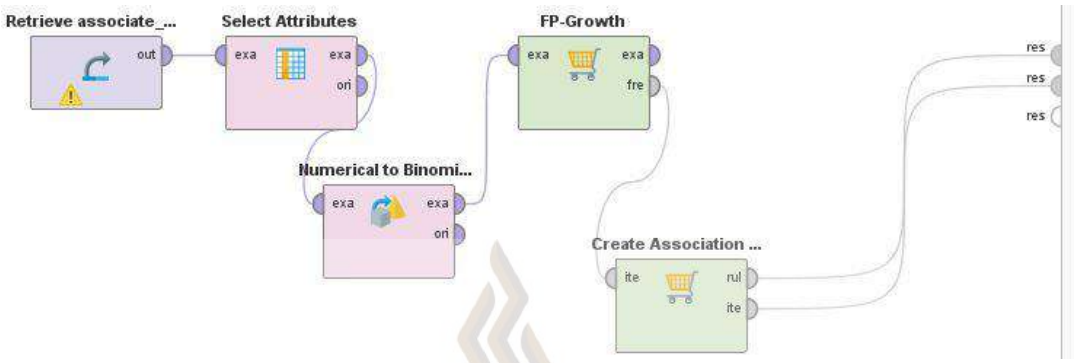
5) คะแนนพฤติกรรมความตั้งใจในการใช้งานมากกว่า 2.500 และ ความคาดหวังประสิทธิภาพการใช้งานอยู่ระหว่าง 3.416 ถึง 4.750 และ พฤติกรรมการใช้งานมากกว่า 4.415 และ ความพยายามใช้งานมากกว่า 5.300

คะแนนพฤติกรรมความตั้งใจในการใช้งานมากกว่า 2.500 และ ความคาดหวังประสิทธิภาพการใช้งานมากกว่า 5.085 และ พฤติกรรมการใช้งานมากกว่า 4.415 และ ความพยายามใช้งานมากกว่า 5.300 และ สภาพสิ่งอำนวยความสะดวกในการใช้งานมากกว่า 4.800

ตอนที่ 4 เพื่อหาความสัมพันธ์ของข้อคำถามการยอมรับการใช้เทคโนโลยีที่มีผลต่อพฤติกรรมการใช้แอปพลิเคชัน LAZADA ช้อปปิ้งบนสมาร์ทโฟนของประชากรในเขตกรุงเทพมหานครและปริมณฑล โดยใช้กฎความสัมพันธ์

ขั้นตอนการพยากรณ์และความสัมพันธ์ของการยอมรับการใช้เทคโนโลยีที่มีผลต่อพฤติกรรมการใช้แอปพลิเคชัน LAZADA ช้อปปิ้งบนสมาร์ทโฟนของประชากรในเขตกรุงเทพมหานคร เพื่อหาค่า Support Confidence และ Lift โดยใช้ Association Rules ในการหาค่าต่าง ๆ จะประกอบด้วย Operators ดังนี้ Retrieve Data เป็นการดึงข้อมูลที่อยู่ในรูปแบบไฟล์นามสกุล CSV Select Attributes เป็นการเลือก Attributes ที่ต้องการหาความสัมพันธ์ของข้อมูลในการวิจัยนี้ ผู้วิจัยเลือกหัวข้อสภาพสิ่งอำนวยความสะดวกในการใช้งาน (Facilitating Conditions : FC) มี 5 หัวข้อย่อย Numerical to Binominal กำหนด Column เป็นประเภท Binominal FP-Growth คือ การใช้

อัลกอริทึมในการคำนวณ Create Association Rules เป็นการสร้างกฎความสัมพันธ์ โดยใช้โปรแกรม Rapid MinerStudio เป็นเครื่องมือช่วย ดังรูปที่ 4.66



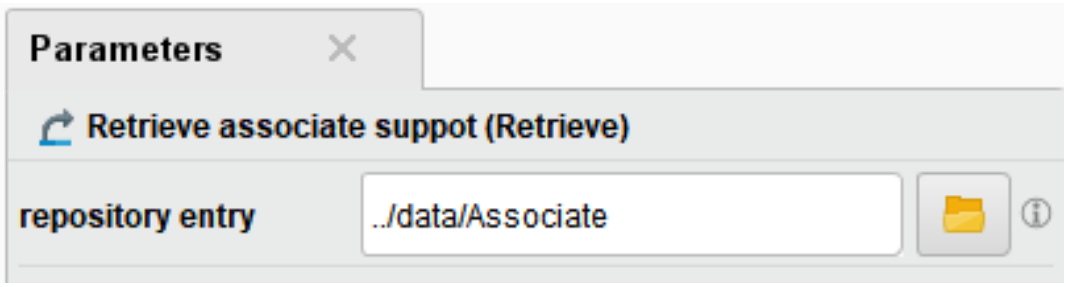
รูปที่ 4.66 Operators หลักที่ใช้ใน Association Rule

รายละเอียด Operators

Retrieve associate Suppot ทำการ Set Parameters ดังรูปที่ 4.67 และ 4.68

Repository entry = ดึง Data จากฐานข้อมูล Associate ใน Repository

รูปที่ 4.67 Operators Retrieve associate support



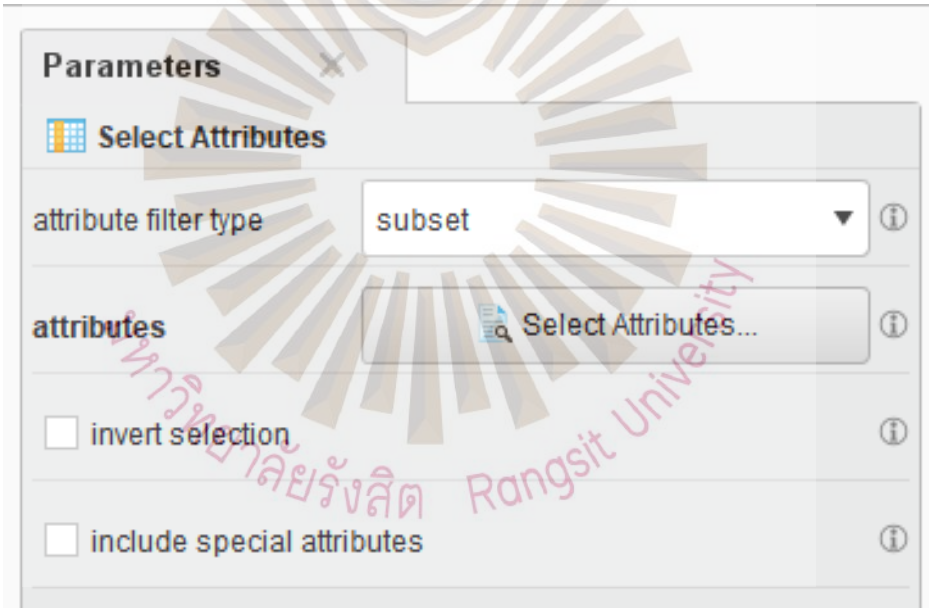
รูปที่ 4.68 Set Parameters Operators Retrieve Associate Support

Select Attributes ทำการ set Parameters ดังรูปที่ 4.69, 4.70 และ 4.71

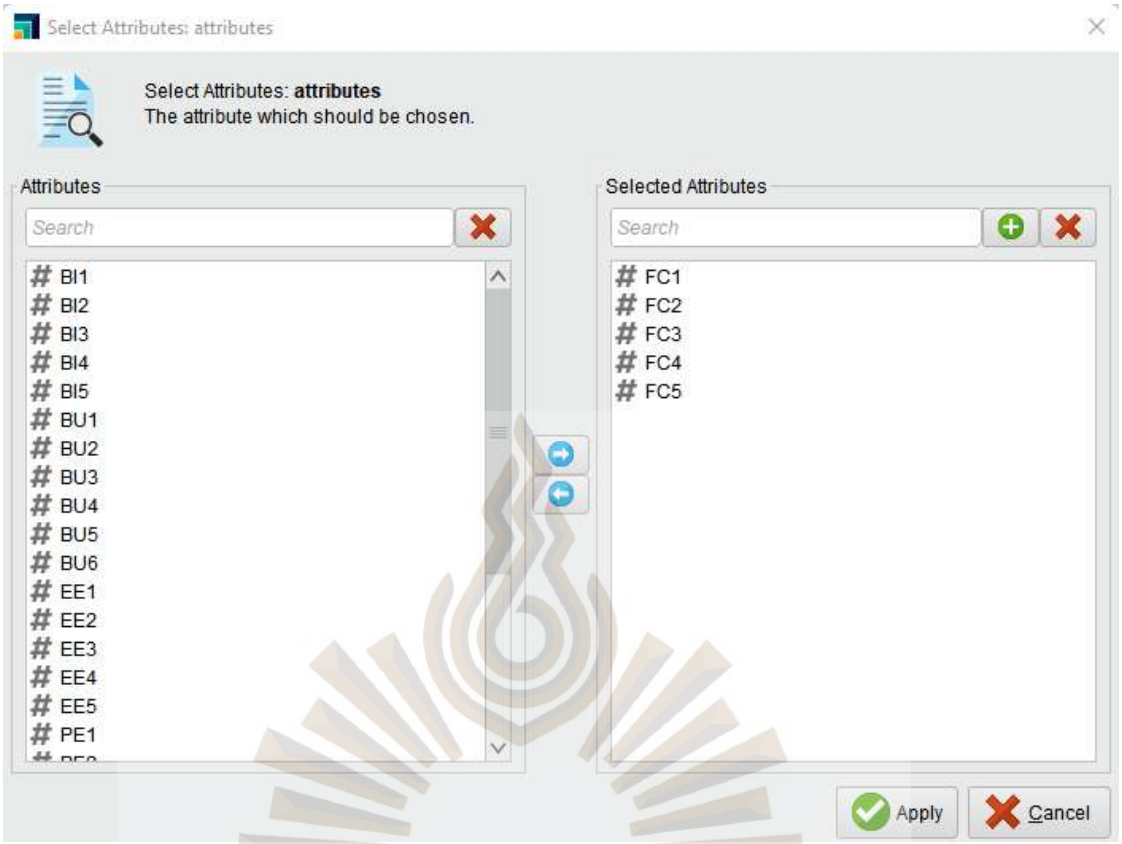
Attributes Filter Type = Subset

Attributes = FC1, FC2, FC3, FC4, FC5

รูปที่ 4.69 Operators Select Attributes



รูปที่ 4.70 Set Parameters Operators Select Attributes



รูปที่ 4.71 การเลือก Parameters Attributes

Numerical to Binominal ทำการ Set Parameters ดังรูปที่ 4.72 และ 4.73

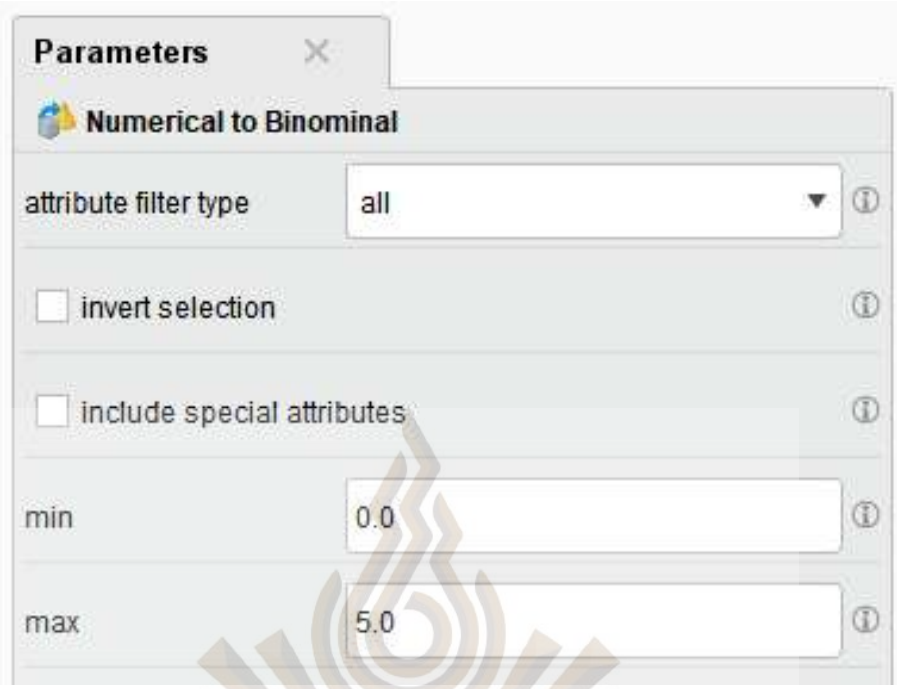
Attribute Filter Type = All

Min = 0.0

Max = 5.0

รูปที่ 4.72 Operators Numerical to Binominal

มหาวิทยาลัยรังสิต Rangsit University



รูปที่ 4.73 Set Parameters Operators Numerical to Binominal

FP-Growth ทำการ Set Parameters ดังรูปที่ 4.74 และ 4.75

Input Format = Items in Dummy Coded Colums

Min Requirement = Support

Min Support = 0.6

Min Items Per Itemset = 1

Max Items Per Itemset = 0

Max Number of Itemsets = 1000000

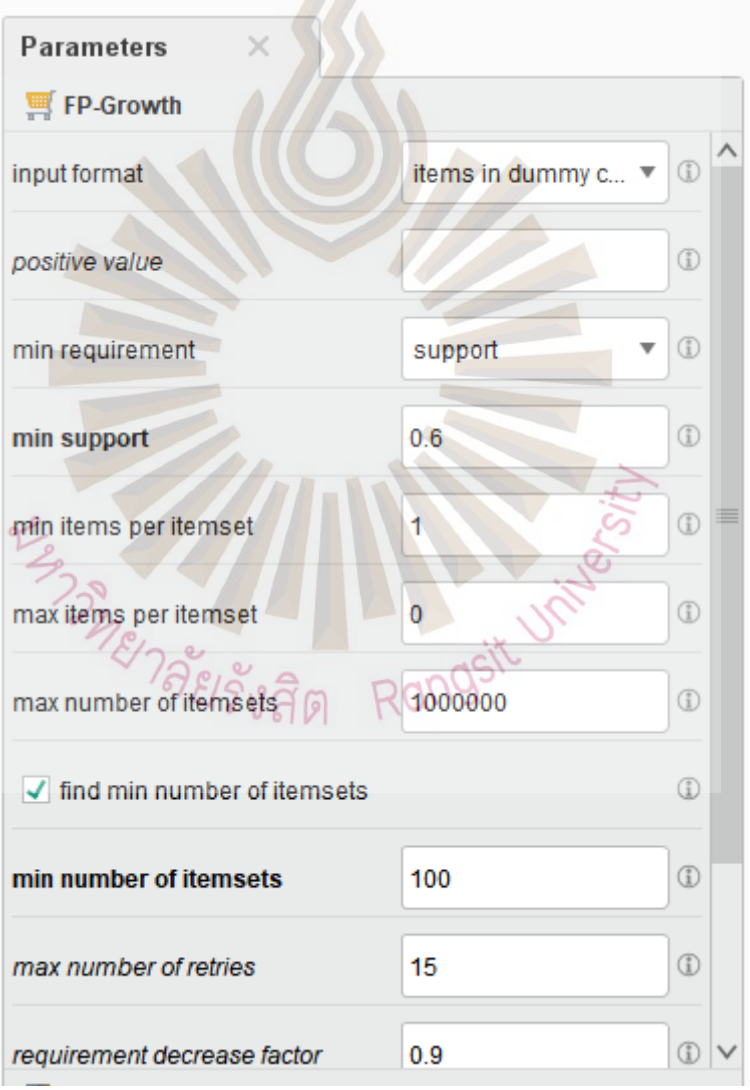
Find Min Number of Itemsets = True

Min Number of Itemsets = 100

Max Number of Retries = 15

Requirement Decrease Factor = 0.9

รูปที่ 4.74 Operators FP-Growth

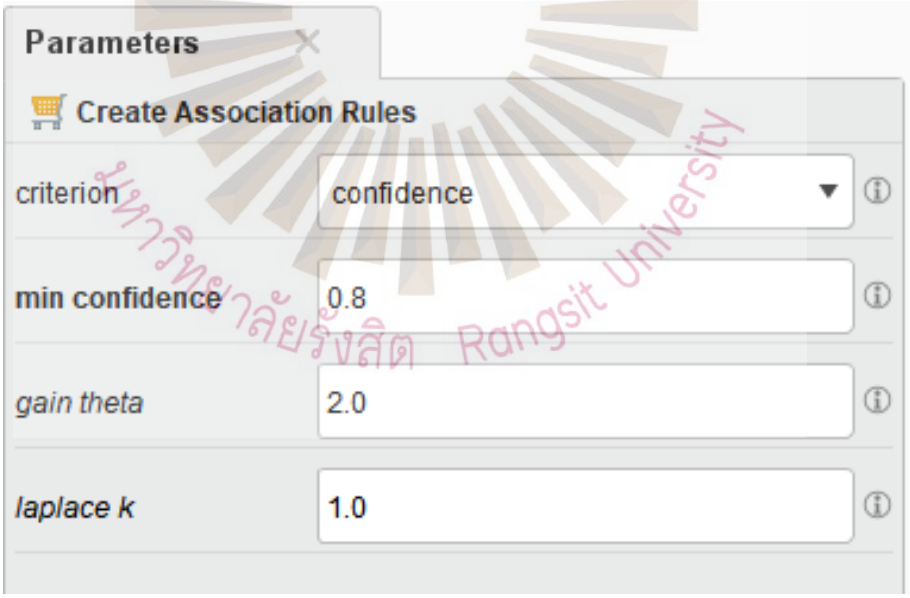


รูปที่ 4.75 Set Parameters Operators FP-Growth

Create Association Rules ทำการ Set Parameters ดังรูปที่ 4.76 และ 4.77

Criterion	= Confidence
Min confidence	= 0.8
Gain theta	= 2.0
Laplace k	= 1.0

รูปที่ 4.76 Operators Create Association Rules



รูปที่ 4.77 Set Parameters Operators Create Association Rules

แสดงความสัมพันธ์ของข้อมูล แบบ Table View โดยใช้เทคนิค Association Rules ดังรูปที่ 4.78

Premises	Conclusion	Support	Confiden... ↓	LaPlace	Gain	p-s	Lift	Conviction
FC5, FC2	FC1	0.140	0.885	0.984	-0.176	0.076	2.176	5.170
FC3, FC4, FC2	FC1	0.135	0.881	0.984	-0.171	0.073	2.167	5.000
FC4, FC2	FC1	0.166	0.877	0.980	-0.212	0.089	2.155	4.812
FC1, FC3, FC5	FC4	0.142	0.873	0.982	-0.184	0.079	2.262	4.835
FC5, FC2	FC4	0.132	0.836	0.978	-0.184	0.071	2.166	3.745
FC1, FC5	FC4	0.184	0.826	0.968	-0.262	0.098	2.139	3.520
FC3, FC4, FC5	FC1	0.142	0.821	0.974	-0.205	0.072	2.018	3.312
FC5, FC2	FC3	0.130	0.820	0.975	-0.187	0.066	2.055	3.333
FC4, FC5	FC1	0.184	0.816	0.966	-0.267	0.092	2.006	3.226
FC3, FC2	FC1	0.168	0.812	0.968	-0.246	0.084	1.998	3.164
FC1, FC4, FC2	FC3	0.135	0.812	0.973	-0.197	0.069	2.037	3.206

รูปที่ 4.78 แสดงความสัมพันธ์ของข้อมูล แบบ Table View โดยใช้เทคนิค Association Rules

ตารางที่ 4.5 สรุปผลการหาความสัมพันธ์ของข้อมูลด้วย Association Rules

Rule	Caption	Support	Confidence
1	ถ้ามีการเลือกหัวข้อ FC5 และ FC2 แล้วจะมีโอกาสให้เลือกหัวข้อ FC1	14.0%	88.5%
2	ถ้ามีการเลือกหัวข้อ FC3 และ FC4 และ FC2 แล้วจะมีโอกาสให้เลือกหัวข้อ FC1	13.5%	88.1%
3	ถ้ามีการเลือกหัวข้อ FC4 และ FC2 แล้วจะมีโอกาสให้เลือกหัวข้อ FC1	16.6%	87.7%
4	ถ้ามีการเลือกหัวข้อ FC1 และ FC3 และ FC5 แล้วจะมีโอกาสให้เลือกหัวข้อ FC4	14.2%	87.3%
5	ถ้ามีการเลือกหัวข้อ FC5 และ FC2 แล้วจะมีโอกาสให้เลือกหัวข้อ FC4	13.2%	83.6%
6	ถ้ามีการเลือกหัวข้อ FC1และ FC5 แล้วจะมีโอกาสให้เลือกหัวข้อ FC4	18.4 %	82.6%
7	ถ้ามีการเลือกหัวข้อ FC3 และ FC4 และ FC5 แล้วจะมีโอกาสให้เลือกหัวข้อ FC1	14.2%	82.1%
8	ถ้ามีการเลือกหัวข้อFC5 และ FC2 แล้วจะมีโอกาสให้เลือกหัวข้อ FC3	13.0%	82.0%
9	ถ้ามีการเลือกหัวข้อ FC4 และ FC5 แล้วจะมีโอกาสให้เลือกหัวข้อ FC1	18.4%	81.6%
10	ถ้ามีการเลือกหัวข้อ FC3 และ FC2 แล้วจะมีโอกาสให้เลือกหัวข้อ FC1	16.8%	81.3%
11	ถ้ามีการเลือกหัวข้อ FC1 และ FC4 และ FC2 แล้วจะมีโอกาสให้เลือกหัวข้อ FC3	13.5%	81.3%

ผลการศึกษาวิจัยเพื่อหาความสัมพันธ์ของการยอมรับการใช้เทคโนโลยีที่มีผลต่อพฤติกรรมการใช้แอปพลิเคชัน LAZADA ช้อปปิ้งบนสมาร์ทโฟนของประชากรในเขตกรุงเทพมหานครและปริมณฑล ผู้วิจัยได้ยกหัวข้อสภาพสิ่งอำนวยความสะดวกในการใช้งาน (Facilitating Conditions : FC) มาหาความสัมพันธ์ โดยได้กำหนดค่า Min Support 0.1 และ Min Confidence 0.8 เพื่อสนใจเฉพาะกฎความสัมพันธ์ที่มีความเชื่อมั่นที่มีค่าสูง และกฎความสัมพันธ์ที่ได้นั้นจะต้องมีการ ซ้ำกัน

ได้เรียงลำดับกฎความสัมพันธ์ค่าความเชื่อมั่น (Confidence) สูงสุดไปหาน้อยสุด ผลข้อมูลที่แสดงจากกฎการทดลอง ประกอบด้วย กฎด้านซ้าย (LHS), กฎด้านขวา (RHS), ค่าสนับสนุน (Support), ค่าความเชื่อมั่น (Confidence), และค่าความสอดคล้อง (Lift) พบว่าสามารถสร้างกฎความสัมพันธ์ของข้อมูลทั้งหมด 11 กฎ

4.3 ผลลัพธ์และการพิจารณา

จากการวิจัยเปรียบเทียบแบบจำลองการพยากรณ์ระหว่าง เทคนิคการจำแนกข้อมูลด้วยวิธีต้นไม้การตัดสินใจ เทคนิคการจำแนกข้อมูลด้วยวิธีเครือข่ายประสาทเทียม เทคนิคการจำแนกข้อมูลด้วยวิธีไนอีฟเบย์ ได้ผลการพยากรณ์ค่า F เท่ากับ 0.332 และ ค่า Prob เท่ากับ 0.771 ซึ่งไม่แตกต่างกันอย่างมีนัยสำคัญที่ระดับ 0.05 ดังนั้น ผู้วิจัยจึงทำการเลือกการพยากรณ์โดยใช้เทคนิคการจำแนกข้อมูลด้วยวิธีต้นไม้การตัดสินใจ เพื่อใช้ในการพยากรณ์และการหาความสัมพันธ์ของการยอมรับการใช้เทคโนโลยีที่มีผลต่อพฤติกรรมการใช้แอปพลิเคชัน LAZADA ช้อปปิ้งบนสมาร์ทโฟนของประชากรในเขตกรุงเทพมหานครและปริมณฑล ได้รูปแบบการพยากรณ์ที่ส่งผลให้ผู้บริโภคมีการยอมรับการใช้เทคโนโลยีที่มีผลต่อพฤติกรรมการใช้แอปพลิเคชัน LAZADA ช้อปปิ้งบนสมาร์ทโฟนของประชากรในเขตกรุงเทพมหานครและปริมณฑลจำนวน 6 รูปแบบ และเมื่อได้ผลการพยากรณ์ที่องค์ประกอบหลักแล้ว ซึ่งแต่ละองค์ประกอบนั้นจะมีองค์ประกอบย่อย จึงมาหาความสัมพันธ์แต่ละองค์ประกอบย่อย โดยใช้กฎความสัมพันธ์ ซึ่งสามารถนำไปใช้ในการพัฒนาแอปพลิเคชันที่เหมาะสม โดยใช้กฎความสัมพันธ์

มหาวิทยาลัยรังสิต Rangsit University

บทที่ 5

สรุปผล

5.1 สรุปผล

จากการพยากรณ์ โดยใช้กลุ่มตัวอย่างเป็นผู้มีประสบการณ์การใช้แอปพลิเคชันช้อปปิ้งบนสมาร์ทโฟน (LAZADA) จำนวน 386 คน อาศัยในเขตกรุงเทพมหานครและปริมณฑล มีประเด็นในการอภิปรายได้ดังนี้

5.1.1 จากการวิจัยเปรียบเทียบความถูกต้องและประสิทธิภาพของแบบจำลองพยากรณ์ทั้ง 3 เทคนิค ซึ่ง เทคนิคการจำแนกข้อมูลด้วยวิธีต้นไม้การตัดสินใจ มีความแม่นยำอยู่ที่ 74.29% เทคนิคการจำแนกข้อมูลด้วยวิธีเครือข่ายประสาทเทียม มีความแม่นยำอยู่ที่ 78.67% เทคนิคการจำแนกข้อมูลด้วยวิธีไนอีฟเบย์ มีความแม่นยำอยู่ที่ 78.67% ซึ่งไม่แตกต่างกันอย่างมีนัยสำคัญที่ระดับ 0.05 ดังนั้นผู้วิจัยจึงทำการเลือกการพยากรณ์โดยใช้วิธีต้นไม้การตัดสินใจ เพราะเป็นเทคนิคที่ใช้กันแพร่หลายและสามารถเข้าใจได้ง่าย พบว่า กลุ่มตัวอย่างที่ใช้งานต่อเนื่องมี ดังนี้

1) คะแนนพฤติกรรมความตั้งใจในการใช้งานมากกว่า 2.500 และความคาดหวังในประสิทธิภาพมากกว่า 3.415 และพฤติกรรมการใช้งานอยู่ระหว่าง 1.086 ถึง 4.415 และความพยายามใช้งานน้อยกว่าหรือเท่ากับ 3.200

2) คะแนนพฤติกรรมความตั้งใจในการใช้งานมากกว่า 2.500 และความคาดหวังในประสิทธิภาพมากกว่า 3.415 และพฤติกรรมการใช้งานอยู่ระหว่าง 1.086 ถึง 4.415 และความพยายามใช้งานมากกว่า 3.200 และสภาพสิ่งอำนวยความสะดวกในการใช้งานมากว่า 6.100

3) คะแนนพฤติกรรมความตั้งใจในการใช้งานอยู่ระหว่าง 3.901 ถึง 4.300 และความคาดหวังในประสิทธิภาพมากกว่า 4.665 และพฤติกรรมการใช้งานอยู่ระหว่าง 1.086 ถึง 4.415 และความพยายามใช้งานมากกว่า 3.200 และสภาพสิ่งอำนวยความสะดวกในการใช้งานน้อยกว่า 6.1

4) คะแนนพฤติกรรมความตั้งใจในการใช้งานมากกว่า 2.500 และความคาดหวังประสิทธิภาพการใช้งานมากกว่า 3.415 และพฤติกรรมการใช้งานมากกว่า 4.415 และความพยายามใช้งานอยู่ระหว่าง 4.501 ถึง 5.300

5) คะแนนพฤติกรรมความตั้งใจในการใช้งานมากกว่า 2.500 และความคาดหวังประสิทธิภาพการใช้งานอยู่ระหว่าง 3.416 ถึง 4.750 และ พฤติกรรมการใช้งานมากกว่า 4.415 และความพยายามใช้งานมากกว่า 5.300

6) คะแนนพฤติกรรมความตั้งใจในการใช้งานมากกว่า 2.500 และความคาดหวังประสิทธิภาพการใช้งานมากกว่า 5.085 และพฤติกรรมการใช้งานมากกว่า 4.415 และความพยายามใช้งานมากกว่า 5.300 และสภาพสิ่งอำนวยความสะดวกในการใช้งานมากกว่า 4.800

5.1.2 จากการวิจัยเพื่อหาความสัมพันธ์ของข้อมูล ในหัวข้อสภาพสิ่งอำนวยความสะดวกในการใช้งาน (Facilitating Conditions : FC) โดยใช้กฎความสัมพันธ์ พบว่า สิ่งที่ผู้ใช้งานให้ความสำคัญ คือ การซื้อสินค้าผ่านแอปพลิเคชันช้อปปิ้งบนสมาร์ทโฟน (LAZADA) ง่ายกว่าการซื้อสินค้าที่ห้างสรรพสินค้า และหน้าจอแสดงผลบนโทรศัพท์มือถือที่เล็กไม่เป็นอุปสรรคในการใช้งานแอปพลิเคชันช้อปปิ้งบนสมาร์ทโฟน แล้วจะมีการเลือก สามารถอุปกรณ์เชื่อมในการใช้งานแอปพลิเคชันช้อปปิ้งบนสมาร์ทโฟน(LAZADA)ได้ตลอดเวลา

5.2 ข้อเสนอแนะ

5.2.1 ผู้ประกอบการที่ขายสินค้าออนไลน์ในรูปแบบแอปพลิเคชัน ต้องพัฒนาแอปพลิเคชันในรูปแบบที่ตรงความต้องการของผู้บริโภคให้ได้มากที่สุด เพราะผลการวิจัยบ่งบอกถึงตัวแปรหลักของการใช้งานแอปพลิเคชัน คือ พฤติกรรมความตั้งใจในการใช้งาน

5.2.2 การวิจัยในอนาคต ควรมีการเก็บข้อมูลเชิงลึกถึงแรงจูงใจที่จะทำให้เกิดพฤติกรรมความตั้งใจในการใช้งานแอปพลิเคชันเพื่อจะได้นำมาประกอบการตัดสินใจของผู้บริหารเพื่อพัฒนาแอปพลิเคชันให้ตรงกับความต้องการของผู้บริโภคได้มากที่สุด

มหาวิทยาลัยรังสิต Rangsit University

บรรณานุกรม

กริช แรงสูงเนิน. (2554). *การวิเคราะห์ปัจจัยด้วย SPSS และ AMOS เพื่อการวิจัย*. กรุงเทพฯ: ซีเอ็ดยูเคชั่น.

จิราภา เลาหะวรนันท์, รชต ลิ้มสุทธิวันภูมิ, และบัณฑิต ฐานะโสภณ. (2558). การใช้เทคนิคการทำเหมือง ข้อมูลในการจำแนกและคัดเลือกแขนงวิชาสำหรับนักศึกษาคณะเทคโนโลยีสารสนเทศ. *วารสารเทคโนโลยีสารสนเทศลาดกระบัง, 4*(1), 1-9.

ชณัฐดาภรณ์ เย็นประเสริฐ. (2557). *การเปรียบเทียบความแม่นยำการพยากรณ์สถานะการชำระหนี้ของ ลูกหนี้โดยใช้เทคนิคการถดถอยโลจิสติก นาอีฟเบย์และต้นไม้ตัดสินใจ* (Unpublished Master's thesis). มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ, กรุงเทพมหานคร.

ฐิตินันท์ อนุสรณ์, และสมชาย เล็กเจริญ. (2563). *การพยากรณ์การเลือกกลุ่มคณะวิชาในการศึกษาต่อในระดับปริญญาตรีของมหาวิทยาลัยรังสิต โดยใช้เทคนิคการทำเหมืองข้อมูล* (Unpublished Master's thesis). มหาวิทยาลัยรังสิต, ปทุมธานี.

เดช ธรรมศิริ, และพยุง มีสัจ. (2554). การจำแนกข้อมูลด้วยเทคนิคซัพพอร์ตเวกเตอร์แมชชีนโดยการปรับพารามิเตอร์และเลือกคุณลักษณะที่เหมาะสมด้วยขั้นตอนวิธีเชิงพันธุกรรม. *วารสารวิชาการพระจอมเกล้าพระนครเหลือ, 21*(2), 293-303.

นคร ละลอกน้ำ. (2562). การใช้เทคนิคดาต้าไมน์นิงเพื่อการศึกษา. *วารสารการศึกษาและการพัฒนาสังคม*, *14*(2), 1-16.

นงลักษณ์ วิรัชชัย. (2542). *โมเดลลิสเรล : สถิติวิเคราะห์สำหรับการวิจัย* (พิมพ์ครั้งที่ 3). กรุงเทพฯ: สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

นฤชล โรจนบุรานนท์. (2562). *คู่มือการใช้งาน RapidMiner Studio 9 เบื้องต้น*. สืบค้นจาก https://cdn.gotoknow.org/assets/media/files/001/343/666/original_20190227023723.pdf?1551209843

นิเวศ จิระวิชิตชัย. (2560). แบบจำลองการตรวจสอบการทุจริตสำหรับข้อมูลที่ไม่สมดุลโดยใช้เทคนิคการลดมิติข้อมูลร่วมกับการเรียนรู้ของเครื่อง (Unpublished Master's thesis). มหาวิทยาลัยศรีปทุม, กรุงเทพฯ.

ปัทมา สมัครไทย. (2562). *การปรับปรุงกระบวนการขัดเพื่อเพิ่มผลิตผลของกระบวนการผลิตโดยการออกแบบการทดลอง* (Unpublished Independent Study). มหาวิทยาลัยธรรมศาสตร์, กรุงเทพฯ.

มหาวิทยาลัยรังสิต Rangsit University

บรรณานุกรม (ต่อ)

สิงหะ ฉวีสุข และสุนันทา วงศ์จตุรภัทร. (2555). ทฤษฎีการยอมรับการใช้เทคโนโลยี. *วารสารเทคโนโลยีสารสนเทศลาดกระบัง, 1*(1), 1-21.

สุพจน์ อุ่นเรือน และสุมามาลย์ ปานคำ. (2560). รูปแบบความสัมพันธ์เชิงสาเหตุของการยอมรับและการใช้เทคโนโลยีที่มีผลต่อพฤติกรรมการใช้แอปพลิเคชันช้อปปิ้งบนสมาร์ทโฟนของประชากรในเขตกรุงเทพมหานครและปริมณฑล. *การประชุมนำเสนอผลงานวิจัยระดับบัณฑิตศึกษา, 12,* 201-210

อนุวัฒน์ เปพาทย์, วงกต ศรีอุไร, และณัฏฐ์ ดิษเจริญ. (2565). การพยากรณ์การออกกลางคันของนักศึกษามหาวิทยาลัยจากการปรับปรุงด้วยการคัดเลือกคุณลักษณะร่วมกับวิธีโครงข่ายประสาทเทียมเพอร์เซ็ปตรอนหลายชั้น. *วารสารวิทยาศาสตร์และวิทยาศาสตร์ศึกษา (JSSE), 5*(1). 39-48.

Dietrich, D., Heller, B. and Yang, B. (2015). *Data Science & Big Data Analytics Discovering, Analyzing, Visualizing and Presenting Data*. (pp. 420). John Wiley & Sons, Inc.

Garson, G. D. (1998). *Neural Networks: An Introductory Guide for Social Scientists*. London: Sage Publications Ltd.

Jaafari, A., Zenner, E. K. and Pham, B. T. (2018). Wildfire spatial pattern analysis in the Zagros Mountains, Iran: A comparative study of decision tree based classifiers. *Ecological Informatics, 43*(1), 200-211.

Nattapon, M. (2022). *สรุป Insight & Digital Stat Thai 2022 พฤติกรรมการออนไลน์ของคนไทย We are social*. Retrieved from https://www.everydaymarketing.co/trend-insight/digital-stat-2022-thai-insight-and-online-behaviour-from-we-are-social/

Olson, D. L., & Lauhoff, G. (2019). Market Basket Analysis. In D. L. Olson & G. Lauhoff (Eds.), *Descriptive Data Mining* (pp. 31-44). Springer, Singapore.

Pimkote, P. (2018). *ทำนายพนักงานที่จะลาออกด้วยเทคนิค Data Mining แบบไม่ต้องเขียน Code*. Retrieved from https://medium.com/@pichitchai.pim/@pichitchai.pim/ทำนายพนักงานที่จะลาออกด้วยเทคนิค-data-mining-แบบไม่ต้องเขียน-code-d8e339e68f9a

Rowley, H. A., Baluja, S. and Kanade, T. (1998). Neural network-based face detection. *IEEE Transactions on pattern analysis and machine intelligence, 20*(1), 23-38.

บรรณานุกรม (ต่อ)

Subasi, A., Ahmed, A. and Alickovic, E. (2018). Effect of Flash Stimulation for Migraine Detection Using DecisionTree Classifiers. *Procedia Computer Science, 140*(1), 223–229.

Venkatesh, V., Morris, M. G., Davis, G. B., and Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS quarterly, 27*(3), 425-478.

Wikimedai Commons. (2020). *โครงสร้าง Layer ของโครงข่ายประสาทเทียมแบบหลายชั้น รูปโดย Chrislb*. Retrieved from https://commons.wikimedia.org/wiki/File:MultiLayerNeural NetworkBigger_english.png

มหาวิทยาลัยรังสิต Rangsit University

ภาคผนวก

จดหมายตอบรับตีพิมพ์เผยแพร่บทความในวารสารรังสิตสารสนเทศ

มหาวิทยาลัยรังสิต Rangsit University

มหาวิทยาลัยรังสิต
RANGSIT UNIVERSITY
มหาวิทยาลัยรังสิต
เมืองเอก ถ.พหลโยธิน
จ.ปทุมธานี 12000
Rangsit University
Muang-Ake, Paholyothin Rd.
Pathumthani 12000, Thailand
T. (66) 2997 2200-30
F. (66) 2791 5757
E. info@rsu.ac.th

ที่ มรส.สหส. ๖๒๐๐/๐๔๐๔ สำนักหอสมุด

๒๗ กันยายน ๒๕๖๕

เรื่อง การตีพิมพ์เผยแพร่บทความในวารสารรังสิตสารสนเทศ

เรียน นายวรพจน์ ลิลิตวัฒน์

ตามที่ท่านได้กรุณาส่งบทความเรื่อง **การพยากรณ์และความสัมพันธ์ของ การยอมรับการใช้เทคโนโลยีที่มีผลต่อพฤติกรรมการใช้แอปพลิเคชั่น** LAZADA **ข้อปิ้งบนสมาร์ทโฟนของประชากรในเขตกรุงเทพมหานครและปริมณฑล โดยใช้เทคนิคการทำเหมืองข้อมูล** (Prediction and Association Rule of Technology Adoption Affecting LAZADA Shopping App Usage Behavior on Smartphones among People in Bangkok and its Vicinities by Using Data Mining Techniques) เพื่อตีพิมพ์ลงในวารสารรังสิตสารสนเทศนั้น บทความของท่านได้รับการพิจารณาเข้าสู่การประเมินคุณภาพบทความเพื่อเผยแพร่ในวารสารรังสิตสารสนเทศประจำปีที่ ๒๘ ฉบับที่ ๒ เดือนกรกฎาคม – ธันวาคม ๒๕๖๕

ทางผู้จัดทำวารสาร ขอขอบพระคุณอย่างสูงที่ท่านได้กรุณาสละเวลาเขียนบทความอันเป็นประโยชน์และให้ความรู้แก่ผู้อ่านได้เป็นอย่างมาก และหวังเป็นอย่างยิ่งว่าคงจะได้รับความกรุณาจากท่านเขียนบทความเพื่อพิมพ์เผยแพร่ในโอกาสต่อไป

ขอแสดงความนับถือ

(ดร.มลิวัลย์ ประดิษฐ์ธีระ)
บรรณาธิการวารสารรังสิตสารสนเทศ

ผู้ประสานงาน : นางสาวรัตนาภรณ์ กาศโอสถ แผนกพัฒนา สำนักหอสมุด
โทร. ๐๒-๙๙๗-๒๒๐๐ ต่อ ๓๔๖๓

กระดาษรีไซเคิล 100% ผลิตจากกล่องเครื่องดื่มใช้แล้ว

ประวัติผู้วิจัย

ชื่อ	วรพจน์ ลิลิตวัฒน์
วัน เดือน ปีเกิด	8 เมษายน 2522
สถานที่เกิด	กรุงเทพมหานคร ประเทศไทย
ประวัติการศึกษา	มหาวิทยาลัยศรีปทุม ปริญญาบริหารธุรกิจบัณฑิต สาขาวิชาคอมพิวเตอร์ธุรกิจ, 2549 มหาวิทยาลัยรังสิต ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ, 2565
ที่อยู่ปัจจุบัน	22 ซอย 15 แยก 1 ถนนราษฎร์พัฒนา แขวงสะพานสูง เขตสะพานสูง จังหวัดกรุงเทพมหานคร 10250

มหาวิทยาลัยรังสิต Rangsit University