

การพัฒนาต้นแบบ Web Application เพื่อแนะนำการเลือกซื้อผลิตภัณฑ์ประกันชีวิต
โดยใช้เทคนิคโครงข่ายประสาทเทียม และต้นไม้ตัดสินใจ

พัชรนันท์ เดชธนบดี

TNI

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
บัณฑิตวิทยาลัย สาขาเทคโนโลยีสารสนเทศ
สถาบันเทคโนโลยีไทย-ญี่ปุ่น
ปีการศึกษา 2559

DESIGN OF FRAMEWORK FOR LIFE INSURANCE PRODUCT RECOMMENDATION BY
NEURAL NETWORK AND DECISION TREE

Pattanun Datetanabodee

TNI

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science Program in Information Technology

Graduate School

Thai-Nichi Institute of Technology

Academic Year 2016

หัวข้อวิทยานิพนธ์

การพัฒนาระบบ Web Application เพื่อแนะนำการเลือก
ซื้อผลิตภัณฑ์ประกันชีวิต โดยใช้เทคนิคโครงข่ายประสาท
เทียม และต้นไม้ตัดสินใจ

โดย

พัทธนันท์ เดชชนบดี

สาขาวิชา

เทคโนโลยีสารสนเทศ

อาจารย์ที่ปรึกษาวิทยานิพนธ์

ดร. ธงชัย แก้วกิริยา

บัณฑิตวิทยาลัย สถาบันเทคโนโลยีไทย-ญี่ปุ่น อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็น
ส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรบัณฑิต

.....คณบดีบัณฑิตวิทยาลัย

(รองศาสตราจารย์ ดร. พิชิต สุขเจริญพงษ์)

วันที่.....เดือน.....พ.ศ.....

คณะกรรมการสอบวิทยานิพนธ์

.....ประธานกรรมการ

(ผู้ช่วยศาสตราจารย์ ดร. ณัฐวี อดทนฤกษ์)

.....กรรมการ

(ดร. สรสมัยพร เจริญพิทย์)

.....กรรมการ

(ผู้ช่วยศาสตราจารย์ ดร. ประจักษ์ เนิตโณม)

.....อาจารย์ที่ปรึกษาวิทยานิพนธ์

(ดร. ธงชัย แก้วกิริยา)

พัชรนันท์ เดชธนบดี : การพัฒนาต้นแบบ Web Application เพื่อแนะนำการเลือกซื้อผลิตภัณฑ์ประกันชีวิต โดยใช้เทคนิคโครงข่ายประสาทเทียม และต้นไม้ตัดสินใจ อาจารย์ที่ปรึกษา: ดร. ธงชัย แก้วกิริยา, 90 หน้า.

การพัฒนาต้นแบบ Web Application เพื่อแนะนำการเลือกซื้อผลิตภัณฑ์ประกันชีวิต โดยใช้เทคนิคโครงข่ายประสาทเทียม และต้นไม้ตัดสินใจ มีวัตถุประสงค์ในการทำวิจัย ดังนี้ 1) เพื่อสังเคราะห์และประเมินแบบจำลองการแนะนำการเลือกซื้อผลิตภัณฑ์ประกันชีวิต 2) เพื่อวิเคราะห์รูปแบบการแนะนำการเลือกซื้อผลิตภัณฑ์ประกันชีวิต 3) เพื่อพัฒนาต้นแบบ Web Application ในการแนะนำการเลือกซื้อผลิตภัณฑ์ประกันชีวิต ซึ่งผลของการวิจัยมีดังนี้ ผลลัพธ์ที่ได้จากผลการสังเคราะห์และประเมินแบบจำลองการพัฒนาต้นแบบ Web Application เพื่อแนะนำการเลือกซื้อผลิตภัณฑ์ประกันชีวิต โดยใช้เทคนิคโครงข่ายประสาทเทียม และต้นไม้ตัดสินใจ ได้กรอบแนวคิด 3 โมดูล คือ 1) โมดูล Data Preparation คือโมดูลการเตรียมข้อมูลเพื่อนำสู่กระบวนการวิเคราะห์ 2) โมดูล Modeling คือโมดูลการวิเคราะห์ข้อมูลให้ได้ผลลัพธ์ที่ต้องการโดยใช้อัลกอริทึม Decision Tree และ Neural Network และวิเคราะห์เปรียบเทียบเปอร์เซ็นต์ความแม่นยำในการทำนาย 3) โมดูล Evaluation คือโมดูลส่วนของการทดสอบ ประเมินผล ซึ่งผลการประเมินความเหมาะสมของกรอบแนวคิดจากผู้เชี่ยวชาญอยู่ในระดับสูง เฉลี่ยอยู่ที่ 4.24/5 (S.D. = 0.04), และผลการวิเคราะห์รูปแบบการพัฒนาต้นแบบ Web Application เพื่อแนะนำการเลือกซื้อผลิตภัณฑ์ประกันชีวิต โดยใช้เทคนิคโครงข่ายประสาทเทียม และต้นไม้ตัดสินใจ มีค่าความถูกต้องและแม่นยำอยู่ที่ 91.73%, ส่วนของผลประเมินความพึงพอใจต้นแบบ Web Application การแนะนำการเลือกซื้อแบบประกันชีวิต จากผู้ใช้งานที่เป็นตัวแทนประกันชีวิต ให้ความพึงพอใจในด้านภาพรวมของการใช้งานในระดับสูงมาก เฉลี่ยที่ 4.80/5 (S.D. = 0.45) และความพึงพอใจในด้านความสะดวกการใช้งานเป็นอันดับที่สอง เฉลี่ยที่ 4.60/5 (S.D. = 0.55)

บัณฑิตวิทยาลัย

สาขาวิชา เทคโนโลยีสารสนเทศ

ปีการศึกษา 2559

ลายมือชื่อนักศึกษา

ลายมือชื่ออาจารย์ที่ปรึกษา

PATTANUN DATETANABODEE: DESIGN OF FRAMEWORK FOR LIFE INSURANCE
PRODUCT RECOMMENDATION BY NEURAL NETWORK AND DECISION TREE.
ADVISOR: DR. THONGCHAI KAEWKIRIYA, 90 PP.

This research's objective is the developments of a recommendation system for recommend insurance products are suited for the customer with hybrid data mining techniques by using neural network and decision tree. The objectives of this research are as follows: 1) to synthesize and evaluate the pattern of Web Application prototype for insurance product recommendation. 2) To analyze the patterns of Web Application prototype for insurance product recommendation. 3) To develop the Web Application prototype for insurance product recommendation. The conceptual framework is composed of three modules. (1) Data preparation module is a process of preparing information or data set to the data analysis process. (2) Modeling module is process of the techniques selection for data analysis. (3) Evaluation module is a process of testing, evaluation, and performing the test results. The result of evaluation from the experts are on the very suitable level with the average score 4.24/5 (S.D. = 0.04), the result of measurement precision and accuracy of using neural network and decision tree at 91.73%. The evaluation from end users – insurance agent 1) Satisfaction on the overall of web application prototype process is 4.80/5 (S.D. = 0.45). 2) Satisfaction on the web application prototype user friendly is 4.60/5 (S.D. = 0.55).

TNI

Graduate School

Field of Study Information Technology

Academic Year 2016

Student's Signature.....

Advisor's Signature.....

กิตติกรรมประกาศ

การทำวิทยานิพนธ์นี้จะสำเร็จลงได้นั้น ผู้วิจัยขอกราบขอบพระคุณในความกรุณาของ อาจารย์ดร. ธงชัย แก้วกิริยา ที่ได้สละเวลาในการให้คำแนะนำ คำปรึกษา และกำลังใจเต็มเปี่ยม ตลอดระยะเวลาในการทำวิทยานิพนธ์ฉบับนี้

ทั้งนี้ผู้วิจัยขอกราบขอบพระคุณคณะกรรมการสอบทุกท่าน ได้แก่ ผู้ช่วยศาสตราจารย์ ดร. ณัฐ วี อดุทธภูมิ อาจารย์ ดร. สรมย์พร เจริญพิทย์ และผู้ช่วยศาสตราจารย์ ดร. ประจักษ์ เฉิดโถม ที่ได้กรุณาให้ข้อเสนอแนะที่เป็นประโยชน์ และช่วยตรวจสอบข้อบกพร่องของวิทยานิพนธ์ฉบับนี้ด้วยความเอาใจใส่ จนทำให้วิทยานิพนธ์ฉบับนี้สำเร็จและมีความสมบูรณ์

อีกทั้งผู้วิจัยขอขอบคุณ คณะเทคโนโลยีสารสนเทศสถาบันเทคโนโลยีไทย-ญี่ปุ่นในการให้ทุนสนับสนุนสำหรับการนำเสนอบทความในงานประชุมระดับชาติ NCIT2016 ที่ได้สนับสนุนค่าใช้จ่ายในการลงทะเบียนเพื่อส่งบทความ เพื่อส่งเสริมการทำงานวิจัยให้มีความสมบูรณ์และมีประสิทธิภาพยิ่งขึ้น และขอขอบคุณทุกท่านในคณะเทคโนโลยีสารสนเทศสถาบันเทคโนโลยีไทย-ญี่ปุ่นที่ให้การสนับสนุน และกำลังใจความช่วยเหลือที่ผ่านมา

เหนือสิ่งอื่นใดผู้วิจัยขอกราบขอบพระคุณบิดามารดาของผู้วิจัยที่คอยให้กำลังใจและให้การสนับสนุนในทุกๆ ด้านอย่างที่สุดเสมอมา

ผู้วิจัยหวังเป็นอย่างยิ่งว่า วิทยานิพนธ์ฉบับนี้จะเกิดประโยชน์ต่อผู้ที่มีความสนใจ สามารถนำไปต่อยอด องค์ความรู้เพื่อประยุกต์ใช้งานกับระบบอื่นๆให้เกิดคุณค่าและสามารถใช้งานได้จริง

พัทธนันท์ เดชชนบดี

TNI

THAI - NICHII INSTITUTE OF TECHNOLOGY

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ง
บทคัดย่อภาษาอังกฤษ.....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญ.....	ช
สารบัญตาราง.....	ญ
สารบัญรูป.....	ฎ
บทที่	
1 บทนำ.....	1
1.1 สภาวะความเป็นมา และความสำคัญของปัญหา.....	1
1.2 วัตถุประสงค์ของการวิจัย	2
1.3 ขอบเขตของการวิจัย.....	3
1.4 ประโยชน์ที่คาดว่าจะได้รับ.....	4
1.5 นิยามศัพท์เฉพาะ	4
2 หลักการพื้นฐาน เอกสาร และงานวิจัยที่เกี่ยวข้อง	6
2.1 หลักการพื้นฐาน.....	6
2.2 งานวิจัยที่เกี่ยวข้อง.....	16
3 วิธีการดำเนินงาน	20
3.1 ขั้นที่ 1 การวิเคราะห์ข้อมูลปัญหางานวิจัยต่าง	21
3.2 ขั้นที่ 2 การศึกษาตัวแปรที่เกี่ยวข้อง	22
3.3 ขั้นที่ 3 การสังเคราะห์และวิเคราะห์แบบจำลองการพัฒนาต้นแบบ Web Application เพื่อแนะนำการเลือกซื้อผลิตภัณฑ์ประกันชีวิต โดยใช้เทคนิคโครงข่ายประสาทเทียม และต้นไม้ตัดสินใจ	23
3.4 ขั้นที่ 4 พัฒนาด้านแบบ Web Application เพื่อแนะนำการเลือกซื้อผลิตภัณฑ์ประกันชีวิต โดยใช้เทคนิคโครงข่ายประสาทเทียม และต้นไม้ตัดสินใจ	62

สารบัญ (ต่อ)

บทที่	หน้า
3	3.5 ขั้นที่ 5 การทดสอบระบบ 65
	3.6 ขั้นที่ 6 การสรุปผลการวิจัย..... 67
4	ผลของการวิจัย..... 69
	4.1 ผลการสังเคราะห์และประเมินรูปแบบจำลองการแนะนำการเลือกซื้อผลิตภัณฑ์ ประกันชีวิต โดยใช้เทคนิคโครงข่ายประสาทเทียม และต้นไม้ตัดสินใจ (วัตถุประสงค์ข้อ 1)..... 69
	4.2 ผลการวิเคราะห์รูปแบบการแนะนำการเลือกซื้อผลิตภัณฑ์ประกันชีวิต โดยใช้เทคนิคโครงข่ายประสาทเทียม และต้นไม้ตัดสินใจ (วัตถุประสงค์ข้อ 2)..... 70
	4.3 การพัฒนาต้นแบบ Web Application เพื่อแนะนำการเลือกซื้อผลิตภัณฑ์ ประกันชีวิต โดยใช้เทคนิคโครงข่ายประสาทเทียม และต้นไม้ตัดสินใจ (วัตถุประสงค์ข้อ 3)..... 73
	4.4 ผลการประเมินส่วนติดต่อผู้ใช้งาน (วัตถุประสงค์ข้อ 4)..... 76
5	บทสรุป อภิปราย และข้อเสนอแนะ 77
	5.1 สรุปผลการวิจัย..... 77
	5.2 อภิปรายผลการวิจัย 78
	5.3 ข้อเสนอแนะงานวิจัย 79
	5.4 ประโยชน์ที่จะได้จากการทำวิทยานิพนธ์..... 80
บรรณานุกรม	 81
ภาคผนวก	 85
	ภาคผนวก ก. แบบสำรวจที่ใช้ในงานวิจัย 86
ประวัติย่อผู้วิจัย	 90

สารบัญตาราง

ตาราง	หน้า
3.1 การแปรที่เกี่ยวข้องกับการซื้อประกันชีวิตของลูกค้า	23
3.2 ตัวอย่างข้อมูลผู้ซื้อประกันชีวิต กรณีศึกษาบริษัทประกันชีวิตแห่งหนึ่ง	29
3.3 แสดงค่า Information Gain (IG) ของแต่ละ Attribute.....	53
3.4 ผลลัพธ์การแบ่งกลุ่มข้อมูลด้วย K-means Clustering โดยใช้โปรแกรม Weka	55
3.5 ตัวอย่างข้อมูลเพื่อสร้างแบบจำลองการแนะนำการเลือกซื้อแบบประกัน	56
3.6 ค่า Weight ระหว่าง Input Node และ Hidden Node	56
3.7 ค่า Weight ระหว่าง Hidden Node และ Output Node.....	56
3.8 ตารางตัวอย่าง case ของชุดข้อมูลเพื่อทำการทดสอบ	57
3.9 Confusion Matrix สำหรับคำนวณ Precision & Recall.....	66
3.10 เภณฑ์การวัดประเมินความเหมาะสมของกรอบแนวคิดจากผู้เชี่ยวชาญ	68
4.1 ผลการประเมินความเหมาะสมของกรอบแนวคิดจากผู้เชี่ยวชาญ.....	70
4.2 เปรียบเทียบค่าความถูกต้องแม่นยำของการเลือกใช้อัลกอริทึม สำหรับการแนะนำการเลือก ซื้อผลิตภัณฑ์แบบประกัน	71
4.3 ผลความพึงพอใจหลังใช้งานต้นแบบ Web Application เพื่อแนะนำการเลือกซื้อ แบบประกัน	76

TNI

THAI - NICHI INSTITUTE OF TECHNOLOGY

สารบัญรูป

รูป	หน้า
2.1 Neural Network Structure	11
2.2 Decision Tree Structure	12
2.3 K-means Clustering.....	14
2.4 Limitation of K-means Clustering: Differing Density	15
2.5 Limitation of K-means Clustering: Non-globular Shapes	16
2.6 Limitation of K-means Clustering: Differing Sizes	16
3.1 ขั้นตอนการวิเคราะห์ข้อมูลปัญหานางวิจัย.....	20
3.2 ขั้นตอนการศึกษาตัวแปรที่เกี่ยวข้อง.....	22
3.3 ขั้นตอนการสังเคราะห์และวิเคราะห์แบบจำลอง.....	23
3.4 กรอบแนวคิดของงานวิจัย (Conceptual Framework)	26
3.5 ขั้นตอนการวิเคราะห์รูปแบบการแนะนำการเลือกซื้อผลิตภัณฑ์ประกันชีวิต	27
3.6 การคำนวณหาค่า Information Gain ของ Gender.....	32
3.7 การคำนวณหาค่า Information Gain ของ Age	34
3.8 การคำนวณหาค่า Information Gain ของ Status.....	36
3.9 การคำนวณหาค่า Information Gain ของ Income	38
3.10 การคำนวณหาค่า Information Gain ของ Province	40
3.11 การคำนวณหาค่า Information Gain ของ Height	42
3.12 การคำนวณหาค่า Information Gain ของ Weight	45
3.13 การคำนวณหาค่า Information Gain ของ Occupation.....	47

สารบัญรูป (ต่อ)

รูป	หน้า
3.14 การคำนวณหาค่า Information Gain ของ Occupation Group	49
3.15 การคำนวณหาค่า Information Gain ของ Occupation Level.....	52
3.16 ผลลัพธ์จากการคำนวณค่า Information Gain โดยใช้โปรแกรม Weka.....	54
3.17 แสดงการคำนวณค่าในโหนดของ Hidden Node	59
3.18 แสดงการคำนวณค่าในโหนดของ Output Node	61
3.19 Weka and PHP Integration Overview Structure	63
3.20 ออกแบบหน้าจอ Login	64
3.22 ออกแบบหน้าจอกรอกข้อมูลลูกค้า	64
3.22 ออกแบบหน้าจอแสดงผลการแนะนำแบบประกัน	65
3.23 ผลการแนะนำการเลือกซื้อแบบประกันชีวิตโดยใช้โปรแกรม Weka	66
4.1 รูปแบบโมเดลที่ได้จากวิเคราะห์ด้วย Neural Network.....	72
4.2 หน้าจอ Login เพื่อใช้งานระบบ.....	73
4.3 หน้าจอกรอกข้อมูลลูกค้า.....	74
4.4 หน้าจอแนะนำแบบประกันชีวิต.....	75
4.5 หน้าจอแสดง PDF รายละเอียดแบบประกัน.....	75

TNI

TAHITI - NICHI INSTITUTE OF TECHNOLOGY

บทที่ 1

บทนำ

1.1 สภาวะความเป็นมา และความสำคัญของปัญหา

Data Science หรือ Big Data [1] เป็นคำที่พบบ่อยและได้รับความสนใจมากในปัจจุบัน ด้วยเรื่องศาสตร์และเทคโนโลยีของการวิเคราะห์ข้อมูลที่ถูกพัฒนาให้มีประสิทธิภาพยิ่งขึ้นเรื่อยๆ ซึ่งในบริษัทหรือองค์กรธุรกิจใหญ่ๆหลายแห่งมีนักวิทยาศาสตร์ข้อมูล (data scientist) ที่ทำงานในองค์กรหรือบริษัทนั้นๆ มีหน้าที่ในการออกแบบการแสดงผลข้อมูล (visualization) และการวิเคราะห์ข้อมูลด้วยเทคนิค Data Mining หรือรวมไปถึงการใช้ Machine Learning Model เพื่อหาความหมายจากข้อมูลและแสดงผลการวิเคราะห์ข้อมูลนี้ให้แก่หน่วยงานอื่นๆในองค์กร เพื่อสร้างองค์ความรู้ประกอบการตัดสินใจทางธุรกิจ เสนอแนะการเปลี่ยนแปลงและการวางแผนวิธีการดำเนินธุรกิจให้มีประสิทธิภาพมากขึ้นโดยใช้ผลลัพธ์จากการวิเคราะห์ข้อมูล

ธุรกิจประกันเป็นหนึ่งในธุรกิจอันดับต้นๆที่มีการทำ Data Science หรือ Big Data เนื่องจากเป็นธุรกิจที่ต้องมีบุคลากรนักสถิติและนักคณิตศาสตร์การประกันภัยที่ทำหน้าที่ประเมินวิเคราะห์ และคาดการณ์ ผลกระทบทางด้านการเงินของความเสี่ยงและความไม่แน่นอนต่างๆในธุรกิจ [2] รวมถึงการเก็บรวบรวมข้อมูลประวัติของลูกค้าเพื่อการประกันไว้อย่างมหาศาล ดังนั้นธุรกิจประกันจึงให้ความสำคัญเป็นอย่างมากและมีการลงทุนสูงทางด้าน Data Science เพื่อนำองค์ความรู้ไปใช้ประโยชน์และวางแผนการดำเนินธุรกิจต่างๆ อาทิเช่น การนำข้อมูลเพื่อลดขั้นตอนการทำประกัน, การนำข้อมูลเพื่อพยากรณ์กลุ่มลูกค้าที่ปฏิเสธการทำประกัน และการนำข้อมูลเพื่อไปประเมินและวิเคราะห์ความเสี่ยงการฉ้อโกงประกัน เป็นต้น

สำหรับแนวโน้มภาคธุรกิจประกันชีวิตไทยในปัจจุบันและในอนาคตมีอัตราการเติบโตที่สูงขึ้น เนื่องจากประชากรไทยส่วนใหญ่ในปัจจุบันได้เล็งเห็นถึงประโยชน์และความสำคัญของการทำประกันชีวิต รวมถึงการได้รับการสนับสนุนส่งเสริมธุรกิจประกันชีวิตจากทางภาครัฐ ในการให้ความรู้ด้านประกันชีวิตแก่ประชาชนทุกกลุ่ม ผ่านช่องทางการสื่อสารที่หลากหลาย ดังนั้นการนำเสนอและออกผลิตภัณฑ์แบบประกันชีวิตที่เหมาะสมและตรงกับความต้องการของลูกค้าทุกกลุ่มในยุคปัจจุบันเป็นสิ่งสำคัญที่ภาคธุรกิจประกันชีวิตควรศึกษาและพัฒนาอย่างต่อเนื่อง เพื่อให้สามารถแข่งขันและเทียบเท่ากับบริษัทคู่แข่งในธุรกิจประกันชีวิตได้

จากการสำรวจ [3] วัตถุประสงค์หลักของการเลือกซื้อแบบประกันชีวิตของประชากรมีความหลากหลาย อาทิเช่น เพื่อสร้างนิสัยการออม, เพื่อการลงทุนในระยะยาว, เพื่อลดภาษี, เพื่อการบริหารความเสี่ยง และเพื่อความคุ้มครองชีวิตและทรัพย์สิน ซึ่งเป็นปัจจัยสำคัญที่ส่งผลต่อการเลือก

ซื้อแบบประกันชีวิตของผู้เอาประกัน รวมถึงข้อมูลประชากรลูกค้าไม่ว่าจะเป็นข้อมูลอาชีพ เพศ อายุ รายได้ ระดับการศึกษา ล้วนเป็นอีกหนึ่งปัจจัยในการเลือกซื้อแบบประกันของผู้เอาประกัน ดังนั้น การค้นพบกลุ่มเป้าหมายลูกค้าประกันชีวิต เป็นสิ่งที่มีค่าสำหรับองค์กรธุรกิจประกัน เพื่อให้องค์กรได้มีการนำเสนอและออกผลิตภัณฑ์แบบประกันให้ตรงกับความต้องการของลูกค้า และเพื่อให้ลูกค้าเองได้เลือกซื้อแบบประกันชีวิตที่เหมาะสมตรงกับความต้องการของตนเอง

จากที่กล่าวมาข้างต้น นักวิจัยต้องการนำเสนอการทำ Data Science สำหรับธุรกิจประกันชีวิต โดยเลือกใช้เทคนิคการทำเหมืองข้อมูล (Data Mining) มาใช้ในการวิเคราะห์การแนะนำการเลือกซื้อผลิตภัณฑ์ประกันชีวิตของลูกค้า ซึ่งเป็นเครื่องมือที่ช่วยให้หลายประเภทธุรกิจสามารถเข้าถึงกลุ่มเป้าหมายลูกค้าได้สำเร็จ เพื่อนำไปปรับเปลี่ยนกลยุทธ์ทางการตลาด และด้านการขายให้ทันสถานการณ์ในภาวะที่มีการแข่งขันสูง ดังนั้น งานวิจัยนี้จึงมีวัตถุประสงค์เพื่อการพัฒนาต้นแบบ Web Application เพื่อแนะนำการเลือกซื้อผลิตภัณฑ์ประกันชีวิต ด้วยการทำเหมืองข้อมูล โดยใช้เทคนิคอัลกอริทึมแบบผสม (Hybrid data mining technique) โดยเลือกใช้เทคนิคโครงข่ายประสาทเทียม (Neural Network) และเทคนิคต้นไม้ตัดสินใจ (Decision Tree) ในการวิเคราะห์ข้อมูลสถิติ และสกัดข้อมูลที่มีประโยชน์เพื่อใช้ในองค์กร ได้อย่างมีประสิทธิภาพมากที่สุด และเพื่อผลลัพธ์ในการพยากรณ์ที่แม่นยำ

1.2 วัตถุประสงค์ของการวิจัย

1.2.1 เพื่อสังเคราะห์และประเมินแบบจำลองการแนะนำการเลือกซื้อผลิตภัณฑ์ประกันชีวิต โดยใช้เทคนิคโครงข่ายประสาทเทียม และต้นไม้ตัดสินใจ

1.2.2 เพื่อวิเคราะห์รูปแบบการแนะนำการเลือกซื้อผลิตภัณฑ์ประกันชีวิต โดยใช้เทคนิคโครงข่ายประสาทเทียม และต้นไม้ตัดสินใจ

1.2.3 เพื่อพัฒนาต้นแบบ Web Application เพื่อแนะนำการเลือกซื้อผลิตภัณฑ์ประกันชีวิต โดยใช้เทคนิคโครงข่ายประสาทเทียม และต้นไม้ตัดสินใจ

1.3 ขอบเขตของการวิจัย

1.3.1 ขอบเขตด้านเนื้อหา

1.3.1.1 ประเภทของแบบประกันชีวิต แบ่งออกเป็น 4 ประเภท คือ

1. ประเภทแบบประกันคุ้มครองชีวิต (Protective) เป็นประเภทแบบประกันที่เน้นความคุ้มครองชีวิตสูง ทุนเอาประกันสูง ซึ่งระยะเวลาคุ้มครองมีทั้งแบบตลอดชีพ (Whole Life) และแบบชั่วระยะเวลา (Term Life) แบบประกันประเภทนี้ส่วนใหญ่จะไม่มีเงินคืน มีการชำระเบี้ยประกันและมีความคุ้มครองตามระยะเวลาของแบบประกันที่เลือก

2. ประเภทแบบประกันออมทรัพย์ (Saving or Endowment) เป็นประเภทแบบประกันชีวิตที่เน้นการออมเงิน ออมเงินเพื่อตามเป้าหมายที่ต้องการ ซึ่งแบบประกันประเภทนี้ค่าเบี้ยประกันจะสูงแต่จำนวนทุนเอาประกันต่ำ

3. ประเภทแบบบำนาญ (Annuity) เป็นประเภทแบบประกันกึ่งออมทรัพย์ แต่เน้นออมเพื่อการเกษียณ เป้าหมายคือหลังเกษียณมีเงินใช้จากแบบประกันจนถึงช่วงอายุที่กำหนด ในแบบประกันที่เลือก

4. ประเภทแบบควบการลงทุน (Unit Link) เป็นประเภทแบบประกันกึ่งคุ้มครองชีวิตกึ่งลงทุน เบี้ยประกันส่วนหนึ่งนำไปลงทุนในกองทุนรวม ที่สามารถจัดพอร์ตการลงทุนเองได้ และเบี้ยอีกส่วนเป็นค่าเอาประกันชีวิต ซึ่งมีผลตอบแทนสูงแต่ก็มีความเสี่ยงด้วยเช่นกัน

1.3.2 ขอบเขตด้านประชากรและกลุ่มตัวอย่าง

1.3.2.1 กลุ่มผู้เชี่ยวชาญ

1. ผู้เชี่ยวชาญที่ใช้ในการประเมินแบบจำลองที่ได้สังเคราะห์ขึ้น ซึ่งได้แก่ผู้เชี่ยวชาญด้านการประกันชีวิตที่มีประสบการณ์การทำงานด้านการประกันชีวิตของบริษัทประกันชีวิตแห่งหนึ่งจำนวน 5 ท่าน

1.3.2.2 กลุ่มข้อมูลตัวอย่าง

1. กลุ่มข้อมูลผู้เอาประกันชีวิต หรือผู้ที่ทำการซื้อแบบประกันชีวิตจากบริษัทประกันชีวิตแห่งหนึ่ง จำนวน 3,290 ราย ซึ่งมีข้อมูลประชากรศาสตร์ ได้แก่ เพศ, อายุ, สถานะ, รายได้ต่อปี และอาชีพ เป็นต้น และข้อมูลการถือครองผลิตภัณฑ์ประกันชีวิต ได้แก่ ซื้อแบบประกันชีวิต, ทุนเอาประกันชีวิต และเบี้ยประกันชีวิต

1.3.3 ขอบเขตด้านเทคนิคและวิธีการ

1.3.3.1 ส่วนของการวิเคราะห์การแนะนำการเลือกซื้อผลิตภัณฑ์ประกันชีวิตตามแบบจำลองที่ได้สังเคราะห์ขึ้นจากเทคนิคโครงข่ายประสาทเทียมและต้นไม้ตัดสินใจ ได้ทำการเริ่มต้นจากการคัดเลือก Attribute ที่สำคัญและทำการแบ่งกลุ่มข้อมูลลูกค้า หลังจากนั้นทำการสร้างรูปแบบ (Model) จากเทคนิคโครงข่ายประสาทเทียม เพื่อการแนะนำการเลือกซื้อผลิตภัณฑ์ประกันชีวิต

1.3.3.2 เครื่องมือที่ใช้ในการทดสอบ คือต้นแบบ Web Application การแนะนำการเลือกซื้อผลิตภัณฑ์ประกันชีวิต ที่ถูกพัฒนาขึ้นจากแบบจำลองด้วยการทำเหมืองข้อมูลโดยใช้เทคนิคโครงข่ายประสาทเทียมและต้นไม้ตัดสินใจ ซึ่งจะได้เป็นรูปแบบ (Model) เพื่อใช้ในการแนะนำแก่ผู้ใช้งานไม่ว่าจะเป็นตัวแทนขายประกันหรือช่องทางทางการขายอื่นๆ โดยจะใช้งานผ่าน Web Browser ในการแนะนำและนำเสนอผลิตภัณฑ์แบบประกันที่เหมาะสมแก่ลูกค้า

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1.4.1 ได้รู้จักแนวคิดกระบวนการสร้างพัฒนาด้านแบบตั้งแต่เริ่มต้น จนถึงสิ้นสุด ทำให้เกิดกระบวนการเรียนรู้และกระบวนการคิดที่เป็นแบบแผน

1.4.2 Web Application แนะนำการเลือกซื้อผลิตภัณฑ์ประกันชีวิต สามารถแนะนำและนำเสนอผลิตภัณฑ์ประกันชีวิตที่เหมาะสมให้กับลูกค้า โดยอ้างอิงจากประวัติการซื้อแบบประกันของลูกค้าจากฐานข้อมูลที่ได้เก็บรวบรวมมา

1.4.3 ได้ฝึกฝนวิธีการแก้ไขปัญหาและการนำสิ่งต่างๆที่เกี่ยวข้องมาประยุกต์ใช้ให้เกิดประโยชน์

1.4.4 เพื่อเป็นประโยชน์และเป็นแนวทางในการศึกษาแก่ผู้ที่สนใจในเรื่องการพัฒนาต้นแบบ Web Application การแนะนำผลิตภัณฑ์แบบประกัน ที่สามารถนำไปประยุกต์และต่อยอดองค์ความรู้ในด้านพัฒนาระบบต่างๆได้ในอนาคต

1.5 นิยามศัพท์เฉพาะ

1.5.1 **ผู้เอาประกันภัย (The Insured)** หมายถึง คู่สัญญาประกันภัยซึ่งมีหน้าที่เปิดเผยข้อความจริงต่อผู้รับประกันภัย ตลอดจนมีหน้าที่ชำระเบี้ยประกันภัย และเมื่อเกิดความเสียหายขึ้นในส่วนที่เอาประกันภัยไว้ ผู้เอาประกันภัยก็มีสิทธิในการเรียกร้องค่าสินไหมทดแทน ตามความเสียหายที่เกิดขึ้นจริง

1.5.2 **ผู้รับประกันภัย (Insurer) หรือบริษัทประกันภัย [4]** หมายถึง ผู้มีหน้าที่ชดเชยค่าสินไหมทดแทนเมื่อเกิดความความสูญเสียขึ้นกับผู้เอาประกันภัยตามเงื่อนไขที่ระบุไว้ในกรมธรรม์ประกันภัย โดยค่าสินไหมทดแทนนี้จะจ่ายให้กับผู้รับผลประโยชน์ที่ระบุในสัญญาหรือทายาทของผู้เอาประกันภัยตามลำดับ

1.5.3 **กรมธรรม์ประกันภัย (Policy)** หมายถึง สัญญาที่แสดงรายละเอียดเงื่อนไขข้อตกลง และความคุ้มครองต่างๆ ระหว่างผู้รับประกันภัยและผู้เอาประกันภัย

1.5.4 **เบี้ยประกันภัย (Premium)** หมายถึง เงินที่ผู้เอาประกันภัยจะต้องส่งให้กับผู้รับประกันภัยหรือบริษัทประกันภัยตามที่ระบุไว้ในกรมธรรม์ประกันภัยเพื่อซื้อความคุ้มครองเนื่องจากสัญญาประกันภัยเป็นสัญญาต่างตอบแทน โดยเบี้ยประกันภัยอาจจะจ่ายเป็นรายปี ราย6เดือน ราย3เดือน รายเดือน แล้วแต่ข้อตกลงที่ระบุไว้ในกรมธรรม์ประกันภัย

1.5.5 **ทุนเอาประกันภัย (Sum insured)** หมายถึง จำนวนเงินของค่าสินไหมทดแทนสูงสุดที่ผู้รับประกันภัยหรือบริษัทประกันภัยจะต้องจ่ายให้กับผู้เอาประกันภัยหรือผู้รับประโยชน์[4] โดยจำนวนเงินนั้นจะถูกจ่ายให้ตามรายละเอียดเงื่อนไข ข้อตกลง และความคุ้มครองที่ระบุไว้ในกรมธรรม์ประกันภัยเมื่อเกิดความสูญเสียขึ้น

1.5.6 แบบประกันชีวิต สามารถแบ่งออกได้เป็น 4 แบบหลัก [5] ดังนี้

1.5.6.1 แบบชั่วระยะเวลา คือ แบบประกันชีวิตที่มุ่งเน้นการคุ้มครองกรณีเสียชีวิตเพียงอย่างเดียว หากผู้เอาประกันชีวิตเสียชีวิตภายในระยะสัญญา บริษัทจะจ่ายจำนวนเงินเอาประกันให้แก่ผู้รับประโยชน์ แต่กรณีที่ผู้เอาประกันมีชีวิตอยู่จนครบสัญญา ผู้เอาประกันก็จะไม่ได้รับเงินคืนจากบริษัท เหมาะกับผู้ที่ต้องการความคุ้มครองสูง แต่ชำระเบี้ยประกันรายต่ำ

1.5.6.2 แบบตลอดชีพ คือ แบบประกันชีวิตที่มุ่งเน้นในเรื่องการคุ้มครองกรณีเสียชีวิต โดยให้ความคุ้มครองจนกว่าผู้เอาประกันจะเสียชีวิต หากผู้เอาประกันเสียชีวิตขณะที่กรมธรรม์มีผลบังคับ บริษัทจะจ่ายเงินให้แก่ผู้รับประโยชน์ตามจำนวนเงินเอาประกันราย เหมาะกับผู้ที่ต้องการความคุ้มครองการเสียชีวิต และต้องการความคุ้มครองที่ยาวนาน

1.5.6.3 แบบสะสมทรัพย์ คือ แบบประกันชีวิตที่ให้ความคุ้มครองและออมทรัพย์ โดยหากผู้เอาประกันเสียชีวิตลงภายในระยะเวลาที่กำหนดขณะที่กรมธรรม์ยังมีผลบังคับ บริษัทจะจ่ายเงินให้แก่ผู้รับประโยชน์ แต่หากผู้เอาประกันมีชีวิตอยู่เมื่อครบสัญญา บริษัทก็จะจ่ายจำนวนเงินเอาประกันรายให้แก่ผู้เอาประกันรายหรือเจ้าของกรมธรรม์ เหมาะกับผู้ที่ต้องการความคุ้มครองจากการประกันชีวิต ขณะเดียวกันก็ต้องการออมทรัพย์ด้วย

1.5.6.4 แบบได้เงินประจำ คือ แบบประกันชีวิตที่มุ่งเน้นการออมทรัพย์ โดยเมื่อผู้เอาประกันมีอายุถึงเกณฑ์หรือตามเงื่อนไขที่กำหนดไว้ในกรมธรรม์ บริษัทจะจ่ายเงินได้ประจำ หรือเงินบำนาญให้แก่ผู้เอาประกันราย เหมาะกับผู้ที่ต้องการมีเงินไว้ใช้ในยามที่เกษียณอายุแล้ว

1.5.7 การทำเหมืองข้อมูล (Data Mining) หมายถึง การค้นหาความสัมพันธ์และค้นหารูปแบบโมเดลของข้อมูลที่มีอยู่ในฐานข้อมูล ซึ่งความสัมพันธ์และรูปแบบโมเดลนั้นอยู่ภายใต้ข้อมูลจำนวนมาก การทำเหมืองข้อมูลจะทำการสำรวจและวิเคราะห์ข้อมูลให้อยู่ในรูปแบบที่เต็มไปด้วยความหมายและอยู่ในรูปของกฎความสัมพันธ์ [6] โดยความสัมพันธ์นี้แสดงให้เห็นถึงองค์ความรู้ที่มีค่า (value) และมีประโยชน์ในฐานข้อมูล ซึ่งในปัจจุบันองค์กรธุรกิจส่วนใหญ่พบกับปัญหาของข้อมูลที่มีจำนวนมากมหาศาลแต่ข้อมูลที่น่ามาใช้นั้นมีน้อย ดังนั้นการทำเหมืองข้อมูล จึงได้เป็นที่รู้จักและได้ถูกนำมาประยุกต์ใช้อย่างกว้างขวาง

บทที่ 2

หลักการพื้นฐาน เอกสาร และงานวิจัยที่เกี่ยวข้อง

2.1 หลักการพื้นฐาน

การศึกษาแนวคิด ทฤษฎี และงานวิจัยที่เกี่ยวข้อง เพื่อเป็นแนวทางในการศึกษางานวิจัยและค้นคว้าวิธีการพัฒนาต้นแบบ Web Application เพื่อแนะนำการเลือกซื้อผลิตภัณฑ์ประกันชีวิต โดยใช้เทคนิคโครงข่ายประสาทเทียม และต้นไม้ตัดสินใจ ได้มีการศึกษาทฤษฎี หรืองานวิจัยที่เกี่ยวข้องแบ่งเป็นหัวข้อดังต่อไปนี้

- 2.1.1 ทฤษฎีและหลักพื้นฐานเกี่ยวกับการประกันภัย
- 2.1.2 ปัจจัยที่ส่งผลต่อการเลือกทำประกันภัย
- 2.1.3 ทฤษฎีเกี่ยวกับ Data Mining (เหมืองข้อมูล)
- 2.1.4 การทำเหมืองข้อมูลโดยใช้เทคนิคโครงข่ายประสาทเทียม (Neural Network)
- 2.1.5 การทำเหมืองข้อมูลโดยใช้เทคนิคต้นไม้ตัดสินใจ (Decision Tree)
- 2.1.6 การทำเหมืองข้อมูลโดยใช้เทคนิคการจำแนกกลุ่ม (K-means Clustering)

2.1.1 ทฤษฎีและหลักพื้นฐานเกี่ยวกับการประกันภัย

2.1.1.1 ประเภทของภัย หรือสิ่งที่เป็นอันตราย [7] สามารถแบ่งภัยออกได้เป็น 3 ประเภท คือ

1. ภัยจากธรรมชาติ (Natural Perils) หมายถึง ภัยหรือเหตุการณ์ที่มนุษย์ไม่สามารถควบคุมเอาไว้ได้ เช่น ไฟไหม้ป่า น้ำป่าไหลหลาก พายุฝน ภูเขาไฟระเบิด เป็นต้น ซึ่งก่อให้เกิดผลกระทบและความเสียหายของมนุษย์และทรัพย์สิน

2. ภัยที่มนุษย์สร้างขึ้น (Human Perils) หมายถึง ภัยหรือเหตุการณ์ที่เกิดจากการกระทำของมนุษย์เอง เช่น อาชญากรรม การจลาจล ภัยสงคราม เป็นต้น

3. ภัยจากเศรษฐกิจ (Economic Perils) หมายถึง ภัยหรือเหตุการณ์ที่เกิดจากสภาพทางเศรษฐกิจ เช่น ปัญหาภาวะเงินฝืดเงินเฟ้อ ปัญหาการว่างงาน ปัญหาการตลาด เป็นต้น

2.1.1.2 วิธีการจัดการความเสี่ยง (Risk Management) คือ กระบวนการป้องกันและควบคุมโอกาสที่จะเกิดความเสียหาย ลดความเสียหายที่เกิดขึ้น หรือการหาช่องทางบรรเทาภาระทางการเงิน อันเนื่องมาจากความเสียหายนั้นๆ สามารถแบ่งออกเป็น 4 ประเภท คือ

1. การหลีกเลี่ยงความเสี่ยงภัย (Avoidance) คือการไม่ทำกิจกรรมหรือเข้าไปยุ่งเกี่ยวกับงานที่จะก่อให้เกิดความเสี่ยง เช่น กลัวจมน้ำหรือตกน้ำ ก็หลีกเลี่ยงโดยไม่โดยสารทางเรือ เป็นต้น

2. การลดความเสี่ยงภัย (Risk Reduction) คือการลดความรุนแรงและอัตราการเกิดภัย ซึ่งสามารถแบ่งออกเป็น 3 วิธี คือ

2.1 การป้องกันการเกิดความเสียหาย (Loss Prevention) กระทำหรือเตรียมการบางอย่างก่อนที่จะเกิดความเสียหาย

2.2 การควบคุมความเสียหาย (Loss Control) กระทำในขณะที่ หรือหลังจากเกิดความเสียหายแล้ว เพื่อควบคุมความรุนแรงให้ลดลง

2.3 การแยกทรัพย์สิน (Separation) กระทำหรือเตรียมการก่อนจะเกิดความเสียหายแก่ทรัพย์สิน

3. การรับความเสี่ยงภัยไว้เอง (Risk Retention) คือการยินยอมรับความเสียหายที่เกิดขึ้น ถึงแม้โดยเจตนาหรือไม่ตาม และรับภาระนี้ไว้ทั้งหมดที่ตัวเอง

4. การโอนความเสี่ยงภัย (Risk Transfer) คือวิธีการจัดการความเสี่ยงโดยโอนความเสี่ยงภัยทั้งหมดหรือบางส่วนให้กับบุคคลอื่น ซึ่งสามารถกระทำได้ 2 วิธี คือ

4.1 การโอนความเสี่ยงภัยไปให้บุคคลอื่นที่ไม่ใช่การประกันภัย (Non-Insurance Transfer) คือการโอนความเสี่ยงที่ไม่ใช่บริษัทประกันภัยโดยสัญญา

4.2 การโอนความเสี่ยงภัยในรูปของการประกันภัย (Insurance Transfer) คือ การโอนความเสี่ยงไปยังบริษัทประกันภัย โดยมีการทำสัญญาประกันภัย ที่มีการระบุไว้ในสัญญาว่าจะชดเชยค่าเสียหายที่เกิดขึ้นแก่ผู้เอาประกันภัย

2.1.1.3 ทฤษฎีที่เกี่ยวกับประกันภัย หรือหลักการในการพิจารณาการรับประกันภัย มีดังนี้

1. กฎความน่าจะเป็น (Theory of Probability) คือทฤษฎีพื้นฐานที่บริษัทประกันภัยนำไปใช้ในการประมาณการคำนวณเบี้ยประกันภัย ซึ่งจะเรียกเก็บจากผู้เอาประกันภัย โดยคำนึงถึงโอกาสที่จะเกิดเหตุภัย ย่อมมีความเสี่ยงที่จะเกิดภัยมาก อัตราค่าเบี้ยประกันจะสูงตามด้วย เพื่อความคุ้มค่ากับการลงทุนสำหรับความเสี่ยง

2. กฎว่าด้วยจำนวนมาก (Law of Large Number) คือทฤษฎีที่มีหลักการที่ว่า เมื่อมีจำนวนวัตถุที่ร่วมเสี่ยงภัยหรือเอาประกันมากขึ้น ค่าความเสียหายที่คาดว่าจะมีโอกาสเกิดขึ้น จะมีความถูกต้องแม่นยำมากขึ้น

3. กฎของการเฉลี่ย (Law of Average) หมายถึง กรณีถ้ามีกลุ่มผู้เสี่ยงภัยน้อย อัตราค่าเบี้ยประกันภัยจะสูง ถ้ากลุ่มผู้เสี่ยงภัยมีมาก อัตราค่าเบี้ยประกันจะต่ำ

2.1.1.4 หลักพื้นฐานเกี่ยวกับการประกันภัยประกอบไปด้วย 6 ประการ คือ

1. หลักส่วนได้เสียในเหตุประกันภัย (Insurance Interest) หมายถึง ผู้มีสิทธิเอาประกันภัย จะต้องเป็นผู้ที่มีส่วนได้เสียตามเหตุประกันภัย

1.1 กรณีประกันชีวิต ส่วนได้เสียในเหตุประกันภัย คือ ผู้เอาประกันที่มีความผูกพันทางกฎหมาย ความผูกพันทางครอบครัว หรือมีส่วนได้เสียอันเกิดจากการเป็นหุ้นส่วนกันที่ทำการกิจกรรมร่วมกัน

2. หลักสุจริตอย่างยิ่ง (Utmost Good Faith) หมายถึง การที่ผู้เอาประกันและผู้รับประกันมีความจริงใจสุจริตต่อกัน ในการเปิดเผยข้อเท็จจริง ไม่แกล้งความเท็จ ในขณะที่ร่วมกันทำสัญญา การปฏิบัติตามหลักสุจริตอย่างยิ่งมี 3 ประการ คือ

2.1 การเปิดเผยข้อความจริง (Representations)

2.2 การแกล้งข้อความเท็จ (Non - Misrepresentations)

2.3 การปฏิบัติตามรับรอง (Warranty)

3. หลักชดใช้ค่าสินไหมทดแทน (Indemnity) หมายถึง กรณีมีความเสียหายเกิดขึ้นอันเป็นผลที่ได้รับโดยตรงจากภัยที่รับประกันไว้ ผู้รับประกันภัยต้องชดใช้ค่าสินไหมทดแทนให้แก่ผู้ที่มีสิทธิรับค่าสินไหมทดแทนตามความเสียหายที่เกิดขึ้นจริง

4. หลักการรับช่วงสิทธิ (Subrogation) หมายถึง หลักการที่ผู้รับประกันภัยรับช่วงสิทธิจากผู้เอาประกันภัยในการเรียกร้องจากบุคคลภายนอกซึ่งเป็นผู้ก่อความเสียหายแก่วัตถุที่เอาประกันภัย ซึ่งความเสียหายนี้จะต้องเกิดขึ้นจากการกระทำของมนุษย์เท่านั้น

5. หลักการร่วมชดใช้ค่าสินไหมทดแทน (Contribution) หมายถึง หลักการเมื่อเกิดความเสียหาย ถึงแม้ผู้เอาประกันภัยจะทำประกันภัยเกินมูลค่าส่วนเสีย หรือทำประกันภัยกับผู้รับประกันภัยหลายราย ทำให้เกิดจำนวนเงินเกินมูลค่าส่วนเสียก็ตาม ผู้เอาประกันภัยจะได้รับค่าชดใช้สินไหมตามมูลค่าเสียหายที่เกิดขึ้นตามจริงเท่านั้น และเฉลี่ยค่าชดใช้ดังกล่าวตามอัตราส่วน

6. หลักสาเหตุใกล้เคียง (Proximate Cause) หมายถึง การชดใช้ค่าสินไหมอันเนื่องมาจากความเสียหายที่เกิดขึ้นจากสาเหตุใกล้เคียงกับภัย เช่น บ้านนาย ก เกิดภัยไฟไหม้ ทำให้ไฟลามอย่างต่อเนื่องมาที่บ้านนาย ข ทำให้บ้านนาย ข เกิดความเสียหาย กรณีนาย ข มีประกันอัคคีภัยคุ้มครองบ้าน บริษัทประกันภัยต้องจ่ายค่าสินไหมให้แก่ นาย ข ด้วย

2.1.2 ปัจจัยที่ส่งผลต่อการเลือกทำประกันภัย

จากงานวิจัยของชุดิญา [8] ได้ศึกษาปัจจัยที่มีอิทธิพลต่อพฤติกรรมการซื้อประกันสุขภาพของบริษัทเอไอเอ จำกัด มีการศึกษาขอบเขตด้านตัวแปรต้นและตัวแปรตาม ดังนี้

1. ตัวแปรต้น (Independent Variables) ที่ประกอบด้วย 1) ลักษณะทางประชากรศาสตร์ ได้แก่ เพศ อายุ สถานภาพ ระดับการศึกษา อาชีพ และรายได้ 2) ปัจจัยจิตวิทยา ได้แก่ แรงจูงใจ การรับรู้ การเรียนรู้ ความเชื่อและทัศนคติ และบุคลิกภาพ

2. ตัวแปรตาม (Dependent Variables) ได้แก่ รูปแบบกรมธรรม์สุขภาพที่เลือกซื้อ เหตุผลที่เลือก สถานที่ในการเลือกซื้อ จำนวนเงินที่เลือกซื้อ และผู้มีอิทธิพลในการตัดสินใจเลือกซื้อประกันสุขภาพ

โดยผู้วิจัยได้ใช้แบบสอบถามในการรวบรวมข้อมูลเพื่อนำไปวิเคราะห์ข้อมูลหาปัจจัยที่ส่งผลต่อการเลือกทำประกันสุขภาพต่อไป โดยชุดแบบสอบถามมีคำถาม 3 ส่วน คือ

ส่วนที่ 1 คำถามเกี่ยวกับลักษณะบุคคล ได้แก่ เพศ อายุ สถานภาพ ระดับ การศึกษา อาชีพ รายได้ สาเหตุที่ไม่ตัดสินใจซื้อประกันสุขภาพ และโรคประจำตัว

ส่วนที่ 2 คำถามเกี่ยวกับปัจจัยที่มีอิทธิพลต่อการเลือกซื้อ ได้แก่ แรงจูงใจ การรับรู้ การเรียนรู้ ความเชื่อและทัศนคติ และบุคลิกภาพ

ส่วนที่ 3 ข้อมูลพฤติกรรมการเลือกซื้อประกันสุขภาพ ได้แก่ ประเภทรูปแบบกรมธรรม์ที่เลือกซื้อ เหตุผลที่เลือกซื้อ สถานที่ในการเลือกซื้อ จำนวนเงินที่ซื้อ และผู้มีอิทธิพลในการตัดสินใจเลือกซื้อ

และทำการกำหนดค่าน้ำหนักจากคำตอบ โดยใช้หลักการของลิเคิร์ต (Likert) และนำค่าน้ำหนักมาหาค่าเฉลี่ย และแปลความหมายจากค่าเฉลี่ยทั้งหมด 5 ระดับ ซึ่งผลจากการทดสอบในส่วน of คำถามที่ 1 ได้คำตอบ ดังต่อไปนี้

1. ผู้บริโภคที่มีเพศต่างกัน มีการเลือกซื้อรูปแบบกรมธรรม์สุขภาพ เหตุผลที่เลือกซื้อ สถานที่ในการเลือกซื้อ จำนวนเงินที่ซื้อ และผู้มีอิทธิพลในการตัดสินใจเลือกซื้อ แตกต่างกัน

2. ผู้บริโภคที่มีอายุต่างกัน มีการเลือกซื้อรูปแบบกรมธรรม์สุขภาพไม่ต่างกัน แต่สำหรับเหตุผลที่เลือกซื้อ สถานที่ในการเลือกซื้อ จำนวนเงินที่ซื้อ และผู้มีอิทธิพลในการตัดสินใจเลือกซื้อ แตกต่างกัน

3. ผู้บริโภคที่มีสถานภาพต่างกัน มีการเลือกซื้อรูปแบบกรมธรรม์สุขภาพ เหตุผลที่เลือกซื้อ สถานที่ในการเลือกซื้อ จำนวนเงินที่ซื้อ และผู้มีอิทธิพลในการตัดสินใจเลือกซื้อแตกต่างกัน

4. ผู้บริโภคที่มีอาชีพต่างกัน มีการเลือกซื้อรูปแบบกรมธรรม์สุขภาพ เหตุผลที่เลือกซื้อ สถานที่ในการเลือกซื้อ จำนวนเงินที่ซื้อ และผู้มีอิทธิพลในการตัดสินใจเลือกซื้อ แตกต่างกัน

5. ผู้บริโภคที่มีรายได้ต่างกัน มีการเลือกซื้อรูปแบบกรมธรรม์สุขภาพ เหตุผลที่เลือกซื้อ สถานที่ในการเลือกซื้อ จำนวนเงินที่ซื้อ และผู้มีอิทธิพลในการตัดสินใจเลือกซื้อ แตกต่างกัน

2.1.3 ทฤษฎีเกี่ยวกับ Data Mining (เหมืองข้อมูล)

Data mining หรือการทำเหมืองข้อมูล [9] เป็นการวิเคราะห์ข้อมูลจำนวนมากมหาศาล (Big Data) เพื่อค้นหารูปแบบความสัมพันธ์ของข้อมูลที่ซ่อนอยู่ภายใน โดยการเลือกใช้เทคนิคต่างๆ ของ Data Mining ที่สามารถนำไปประยุกต์ใช้ได้จริงและมีความเหมาะสมกับข้อมูล ในการสกัด วิเคราะห์ข้อมูลเพื่อให้ได้องค์ความรู้ที่สามารถนำไปใช้ในการดำเนินธุรกิจ ประกอบการตัดสินใจ และสามารถทำนายอนาคตได้ ซึ่งเทคนิคต่างๆ ของ Data Mining มี 5 เทคนิคหลัก ดังนี้

1. Association Algorithm [10] การค้นหากฎความสัมพันธ์ของข้อมูล เพื่อนำไปวิเคราะห์และทำนายเหตุการณ์ต่างๆ เทคนิคนี้มาจากการวิเคราะห์ตระกร้าซื้อสินค้าของลูกค้า (Market Basket Analysis) ซึ่งเป็นการประเมินและวิเคราะห์การซื้อสินค้าของลูกค้า (Transaction) ตัวอย่างที่นำไปประยุกต์ใช้จริง ได้แก่ ระบบแนะนำหนังสือออนไลน์ของ Amazon ข้อมูลการสั่งซื้อหนังสือของลูกค้าจะถูกนำมาประมวลผลและวิเคราะห์เพื่อหาความสัมพันธ์ เช่น ลูกค้าที่ซื้อหนังสือเล่มหนึ่งมักจะซื้อหนังสือเล่มใดพร้อมกันด้วย ซึ่งความสัมพันธ์ของกระบวนการสามารถคาดการณ์ได้ว่าควรจะแนะนำหนังสือเล่มใดให้แก่ลูกค้า

2. Classification and Prediction เป็นกระบวนการแยกข้อมูลให้อยู่ในกลุ่มโดยใช้คุณสมบัติของข้อมูลหรือ Attribute เป็นตัวกำหนด ซึ่งรูปแบบ (Model) อาจแสดงอยู่ในรูปแบบต้นไม้ตัดสินใจ (Decision Tree) หรือโครงข่ายประสาทเทียม (Neural Network)

3. Clustering หรือ Segmentation เป็นกระบวนการจำแนกกลุ่มหรือตัวแปรของข้อมูลที่มีความคล้ายคลึงกัน (Similarity) รวมเอาไว้ด้วยกัน และกลุ่มที่อยู่ต่างกันจะมีลักษณะหรือตัวแปรที่แตกต่างกัน รวมถึงตัวแปรที่อยู่ในกลุ่มเดียวกันจะมีความสัมพันธ์มากกว่าตัวแปรที่อยู่ต่างกลุ่ม และตัวแปรที่อยู่ต่างกลุ่มจะมีความสัมพันธ์น้อยหรือไม่มีความสัมพันธ์เลย ซึ่งกระบวนการจำแนกกลุ่มนี้จะต้องเป็นการจำแนกหรือแบ่งกลุ่มย่อยๆ ตั้งแต่ 2 กลุ่มขึ้นไป

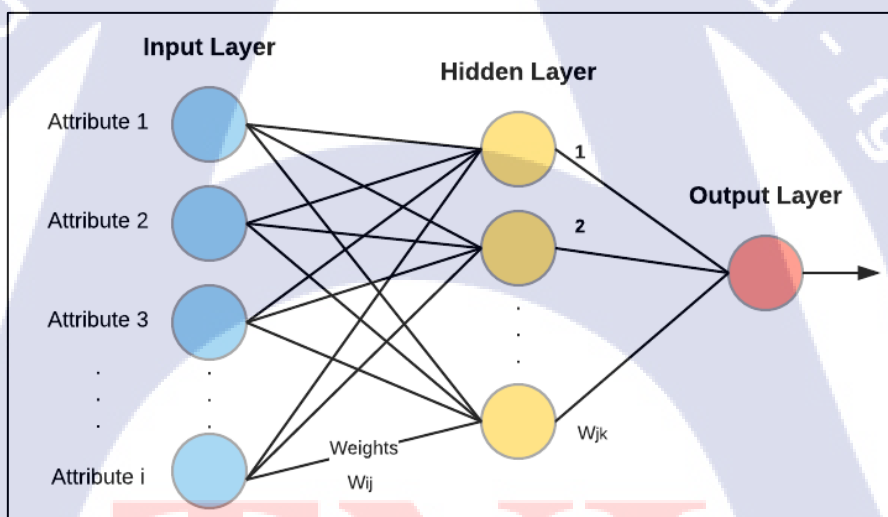
4. Deviation Detection เป็นเทคนิคที่ใช้ในการหาค่าที่ต่างจากมาตรฐาน หรือแสดงลักษณะของข้อมูลผิดปกติที่เกิดขึ้น ซึ่งเป็นข้อมูลที่แยกออกมาจากกลุ่มหรือผิดแตกต่างไปจากข้อมูลกลุ่มอื่นๆ ค่าผิดปกติสามารถเกิดขึ้นได้จากการบันทึกและเก็บรวบรวมข้อมูลที่คลาดเคลื่อนหรือกลุ่มตัวอย่างข้อมูลที่เก็บรวบรวมมามีความแตกต่างจากกลุ่มจริง ซึ่งควรมีการตรวจสอบความถูกต้องของข้อมูลก่อนนำไปวิเคราะห์ สามารถใช้เทคนิคนี้ในการตรวจสอบลายเซ็นหรือเครดิตปลอมได้

5. Link Analysis เป็นการหาความสัมพันธ์ของข้อมูลภายในกลุ่มเดียวกัน เพื่อบอกลักษณะที่จะเกิดขึ้นกับข้อมูลอีกชุดหนึ่ง ลักษณะการหาความสัมพันธ์แบ่งออกเป็น 3 กลุ่ม คือ การหาความสัมพันธ์ระหว่างข้อมูล (Association Discovery) การหาความสัมพันธ์ที่เป็นลำดับข้อมูล

(Sequential Pattern Discovery) และ การหาความสัมพันธ์กับช่วงเวลาใดๆ (Similar Time Sequence Discovery)

2.1.4 การทำเหมืองข้อมูลโดยใช้เทคนิคโครงข่ายประสาทเทียม (Neural Network)

โครงข่ายประสาทเทียม (Neural Network) เป็นหนึ่งในเทคนิคการทำเหมืองข้อมูล ที่เลียนแบบวิธีการเรียนรู้ของสมองมนุษย์เพื่อสร้างเครื่องมือที่มีความสามารถในการเรียนรู้การจดจำ รูปแบบและการสร้างความรู้ใหม่ โครงข่ายประสาทเทียมหรือ Neural Network นี้มีลักษณะเป็น โหนด (Node) ประกอบไปด้วย Input และ Output แต่ละโหนดจะเชื่อมต่อกันด้วยค่าน้ำหนัก (Weights) คือ สิ่งที่ได้จากการเรียนรู้ของโครงข่ายประสาทเทียม ดังรูปที่ 2.1 โดยทั่วไปแล้วประเภท ของการเรียนรู้ของโครงข่ายประสาทเทียมมี 2 ประเภท คือ การเรียนรู้แบบมีผู้สอน (Supervised Learning) และการเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning)



รูปที่ 2.1 Neural Network Structure

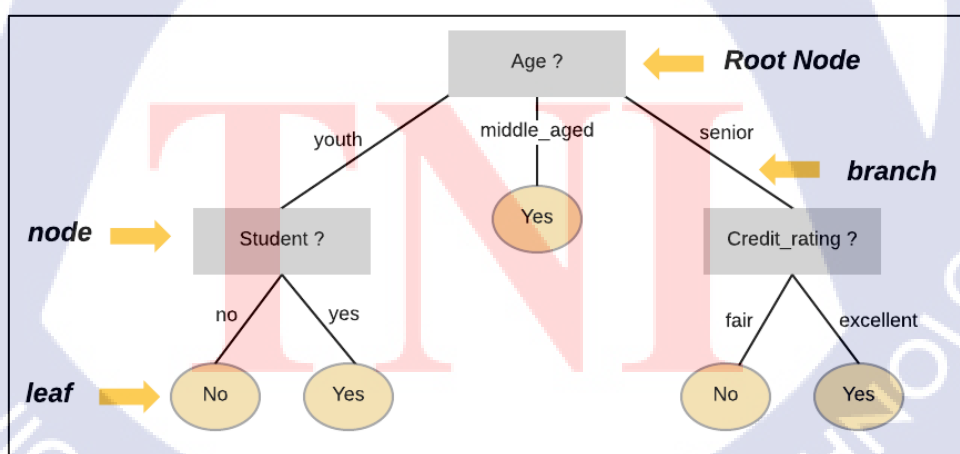
การเรียนรู้แบบมีผู้สอน (Supervised Learning) [11] หรือการเรียนรู้ที่มี “ผู้สอน” เป็นโปรแกรมที่สามารถเรียนรู้ได้ด้วยตัวเองจากชุดข้อมูลที่ใช้สอน (ข้อมูลที่มีผลลัพธ์) ชุดข้อมูลนี้จะประกอบไปด้วย กลุ่มข้อมูลตัวอย่าง แต่ละตัวอย่างจะมีคู่ของข้อมูล คือ ข้อมูลป้อนเข้า (Input) หรือ ข้อมูลที่ต้องการสอน และข้อมูลผลลัพธ์ที่ต้องการส่งออก (Output) เมื่อมีการนำข้อมูลในลักษณะเดียวกันมาเป็นข้อมูลป้อนเข้า โครงข่ายจะกำหนดค่าผลลัพธ์ที่เป็นเป้าหมายให้กับข้อมูลป้อนเข้าแต่ละตัว (x, y) , $x \in X, y \in Y$ ซึ่งสามารถเขียนอยู่ในรูปของฟังก์ชันได้ดังนี้ $f: X \rightarrow Y$

เพื่อให้ได้ข้อมูลผลลัพธ์ตรงตามเป้าหมาย โครงข่ายจะมีการคำนวณค่าผิดพลาดระหว่างค่าเป้าหมาย Y กับค่าผลลัพธ์ $f(x)$ ที่ได้ มาใช้ในการปรับค่าน้ำหนักเพื่อให้ค่าผลลัพธ์ใกล้เคียงกับเป้าหมายมากที่สุด กรณีนี้จะใช้กับโครงข่ายแบบหลายชั้น (MLP: Multilayer Perceptron) เพื่อแก้ปัญหาที่มีความซับซ้อน ตัวอย่างของโครงข่ายแบบหลายชั้น เช่น การแพร่ย้อนกลับ (Back Propagations) เป็นต้น การเรียนรู้แบบมีผู้สอนนี้แบ่งออกเป็น 2 ประเภท คือ การจำแนกข้อมูล (Classification) และการถดถอย (Regression)

การเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) [12] เป็นโปรแกรมที่สามารถเรียนรู้ได้ด้วยตัวเองโดยการนำข้อมูลป้อนเข้าอย่างต่อเนื่องเพียงอย่างเดียว ไม่มีการส่งค่าผลลัพธ์เป้าหมายให้กับข้อมูลป้อนเข้า ค่าน้ำหนักจะปรับตามกลุ่มข้อมูลป้อนเข้าที่มีรูปแบบคล้ายคลึงกัน ซึ่งวิธีวิเคราะห์การจัดกลุ่ม (Cluster Analysis Method) จะถูกใช้ในการเรียนรู้แบบไม่มีผู้สอน

2.1.5 ทฤษฎีการทำเหมืองข้อมูลโดยใช้เทคนิคต้นไม้ตัดสินใจ (Decision Tree)

ต้นไม้ตัดสินใจ (Decision Tree) [13] เป็นโครงสร้างข้อมูลชนิดเป็นลำดับชั้น (Hierarchy) ภายในต้นไม้ประกอบไปด้วยโหนด (Node) ซึ่งแต่ละโหนดจะมีคุณลักษณะ (Attribute) เป็นตัวทดสอบ กิ่งของต้นไม้ (Branch) แสดงถึงค่าที่เป็นไปได้ของคุณลักษณะที่ถูกเลือกทดสอบ และใบ (Leaf) เป็นสิ่งที่อยู่ล่างสุดของต้นไม้ตัดสินใจแสดงถึงกลุ่มของข้อมูล (Class) หรือก็คือผลลัพธ์ที่ได้จากการทำนาย โหนดที่อยู่บนสุดของต้นไม้เรียกว่าโหนดราก (Root Node) ดังแสดงตัวอย่าง รูปที่ 2.2



รูปที่ 2.2 Decision Tree Structure

การสร้างโมเดล Decision Tree ด้วยเทคนิค C4.5 - วิธีการสร้างต้นไม้ตัดสินใจด้วยวิธี C4.5 ได้ถูกพัฒนาขึ้นโดย Ross Quinlan, 1986 ซึ่งขั้นตอนวิธีนี้ใช้เพื่อสร้างต้นไม้ตัดสินใจสำหรับจัดแบ่งกลุ่มข้อมูล โดยจะทำการคัดเลือกคุณลักษณะ (Attribute) ที่มีความสัมพันธ์กับ class มากที่สุดขึ้นมาเป็นบัพโนหนดสุดของต้นไม้ตัดสินใจ (Root Node) ในการหาความสัมพันธ์ของคุณลักษณะหรือ attribute นี้จะใช้ตัววัด ที่เรียกว่า Information Gain (IG) คำนวณได้จากสมการ ดังนี้

$$\text{Gain}(S, A) = E(S) - \sum_{v=\text{value}(A)} \frac{|S_v|}{|S|} E(S_v) \quad (2.1)$$

เมื่อ

S คือ ตัวอย่างที่ประกอบด้วยชุดของตัวแปรและตัวแปรตามหลายๆกรณี

E คือ Entropy ของตัวอย่าง

A คือ ตัวแปรต้นที่พิจารณา

$\text{value}(A)$ คือ เซตของค่าของ A ที่เป็นไปได้

S_v คือ ตัวอย่างที่ A มีค่า v ทั้งหมด

โดยที่ Entropy คำนวณได้จาก

$$\text{entropy}(p_1, p_2, \dots, p_n) = -p_1 \log_2 p_1 - p_2 \log_2 p_2 \dots - p_n \log_2 p_n \quad (2.2)$$

จากการคำนวณเลือกคุณลักษณะ (Attribute) ที่มีค่า Information Gain มากที่สุด จะได้ผลลัพธ์จากการเรียนรู้ต้นไม้ตัดสินใจ และนำไปแปลงเป็นกฎเพื่อใช้งานต่อไป

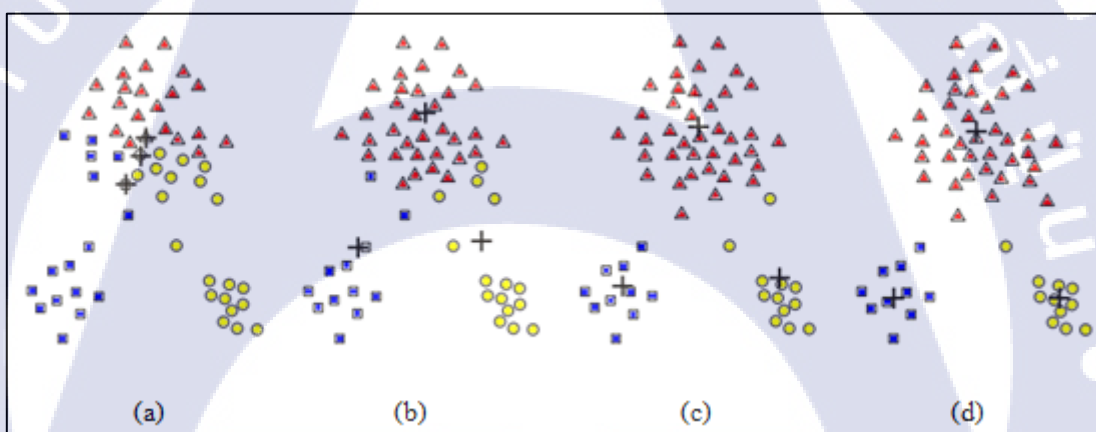
2.1.6 การทำเหมืองข้อมูลโดยใช้เทคนิคการจำแนกกลุ่ม (Clustering)

Cluster Analysis เป็นประเภทหนึ่งของการทำเหมืองข้อมูล (Data Mining) ซึ่งเป็นกระบวนการจัดกลุ่มวัตถุต่างๆได้อย่างเหมาะสม โดยวัตถุที่มีคล้ายคลึงกัน (Similarity) จะถูกจัดให้อยู่ในกลุ่มเดียวกัน ส่วนวัตถุที่มีความใกล้เคียงน้อยหรือไม่มีความคล้ายคลึง (Dissimilarity in distance) กันจะถูกจัดให้อยู่ในกลุ่มอื่น

Clustering เป็นการจัดกลุ่มซึ่งแตกต่างจากการแบ่งประเภทข้อมูล (Classification) โดยมีการแบ่งกลุ่มข้อมูลจากความคล้ายคลึงกันที่ไม่ได้กำหนดจำนวนกลุ่มก่อน ซึ่งเป็นการเรียนรู้แบบไม่มีผู้สอน (Unsupervised Classification) ทัวไปแล้ววิธีหรือเทคนิคที่นิยมสำหรับการแบ่งกลุ่มข้อมูล (Data Clustering) มี 2 เทคนิค คือ การแบ่งกลุ่มแบบลำดับขั้น

(Hierarchy) และการแบ่งกลุ่มออกเป็นส่วนๆ (K-means) ซึ่งแต่ละเทคนิคในการเลือกใช้งานจะขึ้นอยู่กับความเหมาะสมของจำนวนข้อมูลที่ต้องแบ่งกลุ่ม

การจัดกลุ่มด้วยเทคนิค K-means หรือเรียกอีกอย่างหนึ่งว่า การวิเคราะห์กลุ่มแบบไม่เป็นขั้นตอน (Nonhierarchical Cluster Analysis) จะมีการแบ่งส่วน (Partition) ของวัตถุออกเป็น K กลุ่ม เป็นเทคนิคการเรียนรู้แบบไม่มีผู้สอน แทนค่าแต่ละกลุ่มด้วยค่าเฉลี่ยของกลุ่ม ที่ใช้เป็นค่าจุดศูนย์กลาง (Centroid) ของกลุ่มในการวัดระยะทางของข้อมูลในกลุ่มเดียวกัน เทคนิค K-means นี้จะการทำงานหลายๆรอบ (Iteration) โดยในแต่ละรอบมีการจัดกลุ่มข้อมูลที่มีความคล้ายคลึงกันให้อยู่กลุ่มเดียวกัน มีการวัดระยะทางระหว่างค่าของข้อมูลและค่าศูนย์กลางของกลุ่ม ซึ่งเป็นวิธีการพิจารณาว่าข้อมูลใดที่มีความคล้ายคลึงกันและข้อมูลนั้นควรถูกเลือกให้จัดไว้อยู่ที่กลุ่มใด โดยดูจากค่าระยะทางที่น้อยที่สุด หลังจากนั้น ก็ทำการคำนวณค่าจุดศูนย์กลางของกลุ่มไปเรื่อยๆ จนกระทั่งค่าจุดศูนย์กลางของกลุ่มไม่เปลี่ยนแปลง หรือครบจำนวนรอบตามที่กำหนดเอาไว้ ซึ่งเทคนิคการจัดกลุ่มด้วย K-means มีขั้นตอนดังต่อไปนี้ ดังรูปที่ 2.3



รูปที่ 2.3 K-means Clustering [14]

ขั้นตอนการจัดกลุ่ม K-means แบ่งออกเป็น 4 ส่วนดังนี้

- (a) ขั้นตอนแรกกำหนดจำนวนกลุ่ม K และกำหนดค่าจุดศูนย์กลางเริ่มต้นซึ่งจากรูปภาพใช้สัญลักษณ์ + แทนจุดศูนย์กลางของแต่ละกลุ่ม โดยมีทั้งหมด 3 กลุ่ม
- (b) จัดวัตถุที่มีค่าจุดศูนย์กลางใกล้เคียงกับวัตถุนั้นๆมากที่สุด โดยคำนวณจากค่าระยะทางที่น้อยที่สุด
- (c) คำนวณค่าจุดศูนย์กลาง K ใหม่ ซึ่งจะได้ค่าจุดศูนย์กลางและความสัมพันธ์ระหว่างวัตถุที่มีการเปลี่ยนแปลง

(b) ทำซ้ำในข้อ (b) ไปเรื่อยๆ จนกระทั่งค่าจุดศูนย์กลางไม่เปลี่ยนแปลง

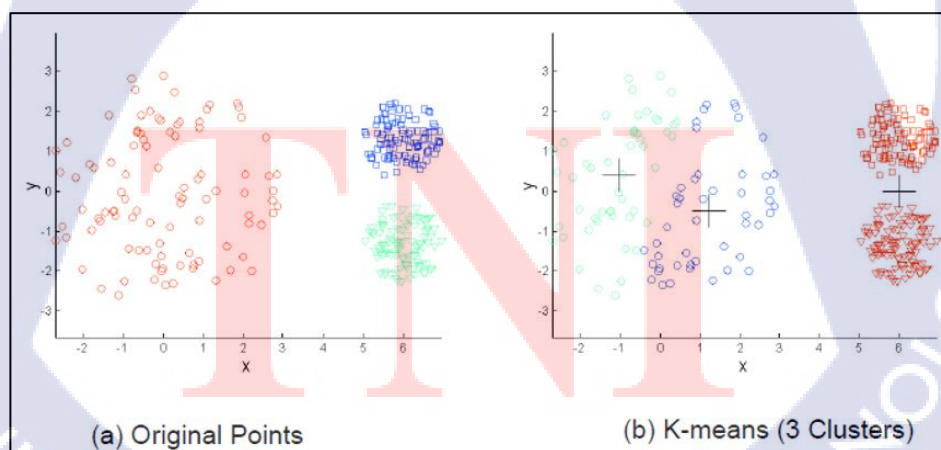
ในการคำนวณหาค่าระยะห่างระหว่างค่าจุดศูนย์กลางและวัตถุ วิธีที่นิยมใช้คือ Squared Euclidean (sqEuclidean) มีสมการ ดังนี้

$$\text{dist} = \sum_{j=1}^K (a_j - b_j)^2 \quad \text{หรือ} \quad \sqrt{\sum_{j=1}^K (a_j - b_j)^2} \quad (2.3)$$

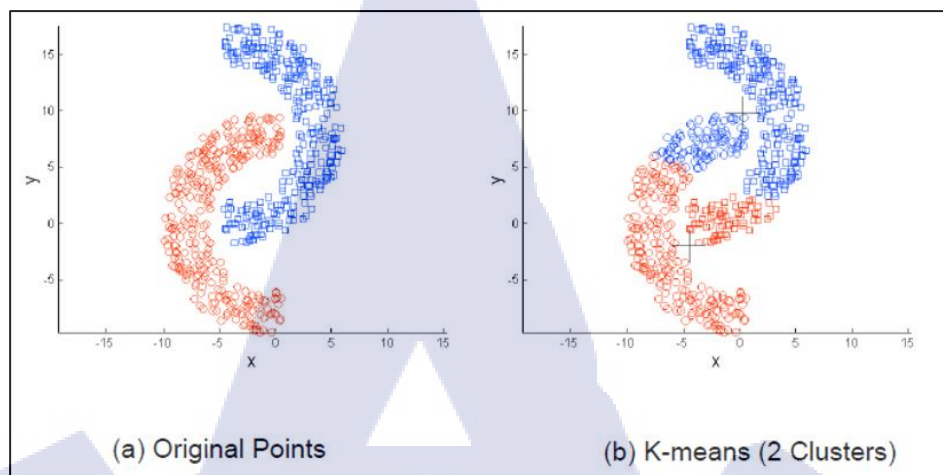
ข้อดีและข้อด้อยของการจัดกลุ่มด้วยเทคนิค K-means

1. การจัดกลุ่มด้วยเทคนิค K-means สามารถคำนวณได้รวดเร็วกว่าเทคนิคอื่นๆ เมื่อมีจำนวนข้อมูลมากและมีจำนวนกลุ่มของข้อมูลน้อย
2. ขั้นตอนการหาค่าจุดศูนย์กลางหรือค่าเฉลี่ยแบบวิธีของ K-means ได้สมาชิกที่อยู่ภายในกลุ่มหนาแน่นกว่าเทคนิคอื่น โดยเฉพาะถ้ามีการจัดกลุ่มเป็นรูปแบบวงกลม
3. การหาค่าจำนวนกลุ่ม K ที่เหมาะสมสามารถคาดเดาได้ยาก
4. ทำงานได้ไม่ดีถ้าการแบ่งกลุ่มข้อมูลไม่ได้เป็นรูปแบบวงกลม
5. ข้อจำกัดในเรื่องของความหนาแน่น รูปร่าง และขนาด ดังรูปที่ 2.4, 2.5 และ

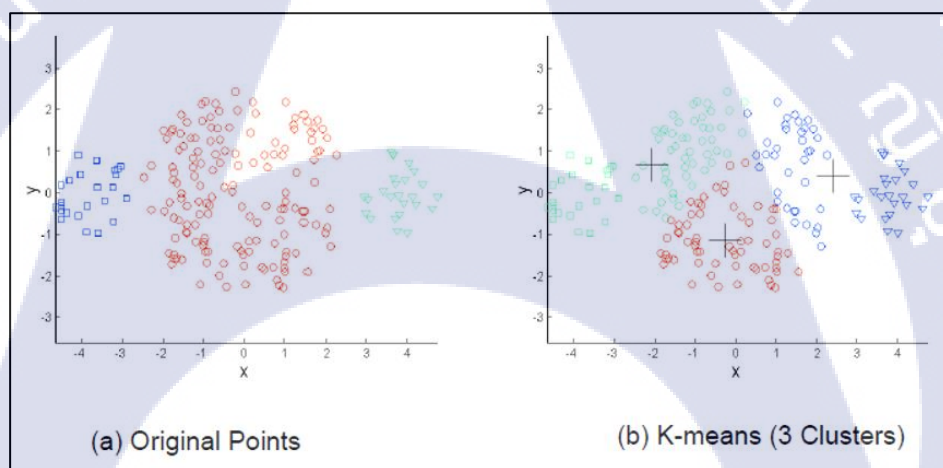
2.6



รูปที่ 2.4 Limitation of K-means Clustering: Differing Density [14]



รูปที่ 2.5 Limitation of K-means Clustering: Non-globular Shapes [14]



รูปที่ 2.6 Limitation of K-means Clustering: Differing Sizes [14]

2.2 งานวิจัยที่เกี่ยวข้อง

การศึกษางานวิจัยที่เกี่ยวข้องในการพัฒนาระบบ Web Application เพื่อแนะนำการเลือกซื้อผลิตภัณฑ์ประกันชีวิต โดยใช้เทคนิคโครงข่ายประสาทเทียม และต้นไม้ตัดสินใจมีข้อมูลดังนี้

1. งานวิจัยที่เกี่ยวข้องกับทางด้านการทำเหมืองข้อมูล

N. Suchrittham et al. [15] ได้สำรวจงานวิจัยของบุคคลอื่น ถึงความเป็นไปได้ในการใช้เทคนิค Data Mining สำหรับธุรกิจประกันชีวิต โดยได้นำเสนอ วิธีการทำเหมืองข้อมูลหรือ Data Mining เพื่อตอบเงื่อนไขเบื้องต้น 3 ประการในการกำหนดประเภทของแบบประกันชีวิตสำหรับลูกค้า

ของแต่ละบุคคล ประการแรก คือการแบ่งประเภทกลุ่มลูกค้า ประการที่สอง คือ ผลิตภัณฑ์ชนิดใดที่เหมาะสมตรงกับความต้องการของลูกค้า และประการสุดท้ายคือ อัตราความเสี่ยงใดของลูกค้าที่มีความสัมพันธ์กับจำนวนการทำประกัน โดยใช้ข้อมูลประชากรของลูกค้าที่ได้ทำสัญญาหรือกรรมธรรม์ฉบับใหม่ในการจำแนกแบ่งกลุ่มลูกค้า (Customer Segmentation) ซึ่งมีการนำเสนอเทคนิคต่างๆ ดังนี้

ประการแรก คือการแบ่งประเภทกลุ่มลูกค้า นักวิจัยได้นำเสนอ เทคนิคการจัดกลุ่ม (Clustering Techniques) คือการใช้เทคนิคเพื่อจำแนกกลุ่ม และค้นหาตำแหน่งของลูกค้าผู้มุ่งหวังในตลาดประกันชีวิตปัจจุบัน โดยการจัดกลุ่มลูกค้าผู้มุ่งหวังจะประกอบไปด้วยตัวแปรต่างๆ ได้แก่ เพศ อายุ รายได้ การศึกษา เป็นต้น ซึ่งลูกค้าผู้มุ่งหวังที่มีลักษณะเหมือนหรือคล้ายคลึงกันจะถูกจัดให้เข้าไปอยู่ในกลุ่มเดียวกัน

ประการที่สอง คือ ผลิตภัณฑ์ชนิดใดที่เหมาะสมตรงกับความต้องการของลูกค้า โดยนักวิจัยได้นำเสนอเทคนิคกฎความสัมพันธ์ (Association Rule Mining Techniques) ในการวิเคราะห์หาความสัมพันธ์ (Apriori Algorithm) และวิเคราะห์ตะกร้าการซื้อ (Market Basket Analysis) เพื่อหาทุกรูปแบบการเชื่อมโยงข้อมูลของข้อมูลลูกค้า เพื่อนำเสนอพฤติกรรมของลูกค้าว่าลักษณะของสินค้าหรือผลิตภัณฑ์ใดที่ถูกซื้อคู่ด้วยกันบ่อย

ประการที่สาม คือ อัตราความเสี่ยงใดของลูกค้าที่มีความสัมพันธ์กับจำนวนการทำประกัน โดยประเมินและกำหนดอัตราความเสี่ยงเพื่อคาดการณ์การสูญเสียจากการเกิดอุบัติเหตุ, การเรียกร้องค่าชดเชยสุขภาพ (Claim) หรือ การเรียกร้องค่าชดเชยเมื่อได้รับภัยพิบัติ เพื่อรองรับและสนับสนุนกระบวนการพิจารณาการรับประกัน (Underwriting Process) ซึ่งอัตราความเสี่ยงนี้จะประเมินจากสุขภาพของลูกค้าและลักษณะทางกายภาพที่จะมีผลต่ออัตราเบี้ยประกัน ซึ่งนักวิจัยได้นำเสนอ ProbE method เทคนิค ซึ่งเป็นเทคนิคการประเมินความน่าจะเป็น เป็นขั้นตอนการทำนายแบบจำลอง (Predictive-modeling Algorithm) ในการสร้างแบบจำลองคาดคะเนความเสี่ยงสำหรับการประกันความเสี่ยง เพื่อลดอัตราความเสี่ยงก่อนการรับประกันลูกค้าของบริษัท

ภคพันธ์ นิสิตศิริ [16] ได้ศึกษาปัจจัยที่มีผลต่อการตัดสินใจซื้อประกันชีวิตของพนักงานในองค์กรการไฟฟ้าฝ่ายผลิตโดยแบ่งผลการศึกษาเป็น 3 ส่วน ดังนี้ ข้อมูลทั่วไปของพนักงานในองค์กร ข้อมูลพฤติกรรมทางด้านสุขภาพ การออม การทำประกันชีวิตและข้อมูลปัจจัยที่มีผลต่อการตัดสินใจทำประกันชีวิต โดยใช้ข้อมูลจากแบบสอบถามจากกลุ่มตัวอย่าง 100 คนซึ่งได้สรุปผล พบว่าพนักงานส่วนใหญ่เป็นเพศชาย อายุระหว่าง 41-50 ปี มีสถานะภาพสมรส และมีระดับการศึกษาชั้นปริญญาตรีหรือเทียบเท่า มีรายได้ในครอบครัวต่อปีมากกว่า 1 ล้านบาท ครอบครัวของพนักงานส่วนใหญ่เคยเข้ารับการรักษาพยาบาลด้านสุขภาพคิดเป็นร้อยละ 93 ในด้านพฤติกรรมกรรมการออมพบว่าพนักงานมีการวางแผนการออมเพื่อเกษียณอายุทั้งหมด ในด้านการทำประกันชีวิต พบว่าพนักงานที่ได้ทำประกัน

ชีวิตไว้แล้วเป็นจำนวนมาก คิดเป็นร้อยละ 97 ในด้านข้อมูลปัจจัยที่มีผลต่อการตัดสินใจทำประกันชีวิตมากที่สุดคือประกันชีวิตที่สามารถนำไปลดหย่อนภาษีได้ รองลงมาคือ หลักประกันชีวิตสร้างความมั่นคงแก่ครอบครัว และอันดับสุดท้ายคือ ประกันชีวิตให้ความคุ้มครองที่หลากหลายทางด้านสวัสดิการเพิ่มเติมอื่นๆ เช่น ค่ารักษาพยาบาลและโรคร้ายแรงอื่นๆ

จันทร์พร ดาวลอย และเทพฤทธิ์ บัณฑิตวัฒนาวงศ์ [17] ได้นำเสนองานวิจัยการวิเคราะห์และทำนายพฤติกรรมการทำประกันชีวิตด้วยเทคนิคการทำเหมืองข้อมูล โดยเลือกใช้เทคนิคต้นไม้ตัดสินใจด้วยวิธี J48 และเลือกการใช้วิเคราะห์กลุ่มแบบไม่เป็นขั้นตอน โดยหลักการของ K-means Clustering มาช่วยจัดกลุ่มข้อมูลของลูกค้าที่ต้องการ ที่มีลักษณะเดียวกันไว้ด้วยกัน และเลือกรูปแบบที่มีประสิทธิภาพมากที่สุด ด้วยชุดข้อมูลการซื้อประกันภัยที่ใช้ในการวิจัยจำนวน 25,982 ระเบียบ ผลการทดลองแบบจำลองสำหรับการทำนายพฤติกรรมการทำประกันภัยของผู้บริโภคสามารถทำนายถูกต้อง 60.40%

กฤษณพร สุริยะบรรเทิง และกมล เกียรติเรืองกมล [18] นำเสนอการสร้างแบบจำลองการขายผลิตภัณฑ์ประกันภัยเพื่อนำเสนอผลิตภัณฑ์ใหม่สำหรับลูกค้าผู้สูงอายุที่ถือบัญชีออมทรัพย์ธนาคาร โดยใช้เทคนิคการทำเหมืองข้อมูล ด้วยวิธี Simple K-Means สำหรับการแบ่งกลุ่มลูกค้า (Clustering) และใช้ขั้นตอนวิธี Apriori สำหรับการหาความสัมพันธ์ของผลิตภัณฑ์ผลลัพธ์ที่ได้จากแบบจำลองการแบ่งกลุ่มและแบบจำลองการหาความสัมพันธ์ ฝ่ายการตลาดสามารถนำแบบจำลองการหาความสัมพันธ์มาใช้ได้ โดยการออกโปรโมชั่นให้กับลูกค้า ด้วยการดูความสัมพันธ์ของผลิตภัณฑ์ อีกทั้งสามารถนำมาใช้ในการขายสำหรับฝ่ายขาย โดยการดูว่าลูกค้าอยู่กลุ่มใด และสามารถขายผลิตภัณฑ์อะไร

2. งานวิจัยที่เกี่ยวข้องกับทางด้านการทำเหมืองข้อมูลโดยใช้เทคนิคผสม

C. Rohitash et al. [19] นำเสนอการใช้เทคนิคการทำเหมืองข้อมูลโดยใช้เทคนิค Neural Network และ Decision Tree เพื่อวิเคราะห์ข้อมูลการจำแนกประเภทของไวน์โดยใช้ข้อมูลหลังจากการวิเคราะห์ด้วย Neural Network (Trained Data) นำมาสร้างเป็น rules base ด้วยเทคนิค J48 Decision Tree และเพื่อเปรียบเทียบประสิทธิภาพความแม่นยำในการจำแนกประเภทของไวน์ระหว่างการใช้เทคนิค Neural Network และ Decision Tree พบว่าการ การใช้เทคนิค Neural Network สามารถจำแนกประเภทของไวน์ได้แม่นยำมากที่สุด แต่ข้อมูลที่ได้หลังจากวิเคราะห์ด้วยเทคนิคนี้ไม่สามารถอธิบายที่มาของคำตอบได้ และการใช้เทคนิค Decision Tree แบบเดี่ยวให้ผลลัพธ์ความแม่นยำได้ดีกว่าการวิเคราะห์ข้อมูลหลังจากใช้ Neural Network

D. Maryam and A. Marzieh [20] นำเสนอการใช้เทคนิคการทำเหมืองข้อมูลแบบผสม เพื่อวิเคราะห์รูปแบบการตอบสนองของลูกค้า (Customer Response Modeling) สำหรับการ

สื่อสารทางการตลาดโดยตรง (Direct Marketing) และทำการเปรียบเทียบการใช้เทคนิคแบบผสม (Hybrid Approach) และเทคนิคแบบสุ่ม (Sampling Approach) สำหรับขั้นตอนแรกเป็นการแบ่งกลุ่มข้อมูล ซึ่งจะใช้ข้อมูลเกี่ยวกับประวัติการซื้อสินค้าของลูกค้า (Historical Purchase Data) ในการสร้างรูปแบบนี้ และเพื่อแก้ปัญหาข้อมูลที่มีน้ำหนักไม่สมดุล (Imbalance Dataset) โดยใช้เทคนิค C&R Tree เพื่อจำแนกข้อมูล และกำหนดจำนวน 7 กลุ่มลูกค้าด้วยเทคนิค K-means หลังจากนั้นใช้เทคนิค Bagging Neural Network (BNN) ในการจำแนกจำนวนลูกค้าที่ตอบสนองและจำนวนที่ลูกค้าไม่ตอบสนองจากลูกค้าทั้งหมด 7 กลุ่ม ซึ่งจะได้เปอร์เซ็นต์การตอบสนองของลูกค้าในแต่ละกลุ่ม (Response Rate) พบว่าการใช้เทคนิคแบบผสมให้ผลลัพธ์ในด้านความแม่นยำสูงกว่าการใช้เทคนิค Bagging Neural Network (BNN) อย่างเดียว

F. Simon et al. [21] นำเสนอการใช้เทคนิคทำเหมืองข้อมูลแบบผสมโดยใช้เทคนิค Back-propagation Neural Network ในการคัดเลือกตัวแปรที่เกี่ยวข้องและสำคัญ และใช้ C4.5 Decision Tree ในการสร้าง rule base เพื่อสร้างระบบแนะนำมหาวิทยาลัยที่เหมาะสมแก่นักเรียนที่จะเริ่มเข้าเรียนศึกษาระดับปริญญาตรี (Recommender System of Admission to University: RSAU) โดยวิเคราะห์และทำนายข้อมูลจากประวัติข้อมูลส่วนตัวของนักศึกษามหาวิทยาลัยแห่งหนึ่ง จากการทดสอบพบว่าการใช้เทคนิคแบบผสม (Hybrid of Decision Tree & Neural Network) ให้ผลลัพธ์ในการแนะนำแม่นยำและมีประสิทธิภาพมากกว่าใช้เทคนิคแบบเดี่ยว (Single Decision Tree or Neural Network)

ข้อสรุป

จากงานวิจัยที่ได้ทำการศึกษาพบ ปัญหาส่วนใหญ่ที่พบคือในเรื่องของข้อจำกัดในด้านของข้อมูล ไม่ว่าจะเป็นวิธีการเตรียมข้อมูลให้สมบูรณ์ และจำนวนขนาดข้อมูลที่เพียงพอ ซึ่งมีผลต่อประสิทธิภาพและความแม่นยำในการทำเหมืองข้อมูล อีกทั้งในเรื่องของข้อจำกัดด้านเวลาในการเลือกวิธีการทำเหมืองข้อมูลที่เหมาะสม

สำหรับงานวิจัยของนักศึกษาได้เลือกใช้ข้อมูลจากแหล่งเดียวที่มีจำนวนขนาดข้อมูลเพียงพอ และเป็นข้อมูลที่ไม่ล้าสมัย รวมถึงการศึกษาและทำความเข้าใจถึงข้อเด่นและข้อด้อยของแต่ละเทคนิควิธีการทำเหมืองข้อมูล ซึ่งทำให้สามารถเลือกใช้วิธีการทำเหมืองข้อมูลได้อย่างถูกต้องเหมาะสม เพื่อการวิเคราะห์พยากรณ์ข้อมูลที่ประสิทธิภาพ แม่นยำ และรวดเร็ว พร้อมนำไปใช้เกิดประโยชน์ เนื่องจากโลกธุรกิจมีการแข่งขันสูงต้องปรับตัวให้เร็วทันต่อการเปลี่ยนแปลงจากภายนอก

บทที่ 3

วิธีการดำเนินงาน

จากการศึกษางานวิจัยเรื่อง วิธีการพัฒนาต้นแบบ Web Application เพื่อการแนะนำการเลือกซื้อผลิตภัณฑ์ประกันชีวิต โดยใช้เทคนิคโครงข่ายประสาทเทียม และต้นไม้ตัดสินใจ มีขั้นตอนกระบวนการต่อไปนี้

ขั้นที่ 1 การวิเคราะห์ข้อมูลปัญหางานวิจัย

ขั้นที่ 2 การศึกษาตัวแปรที่เกี่ยวข้อง

ขั้นที่ 3 การสังเคราะห์และวิเคราะห์แบบจำลองการพัฒนาต้นแบบ Web Application เพื่อแนะนำการเลือกซื้อผลิตภัณฑ์ประกันชีวิต โดยใช้เทคนิคโครงข่ายประสาทเทียม และต้นไม้ตัดสินใจ

ขั้นที่ 4 พัฒนาต้นแบบ Web Application เพื่อการแนะนำการเลือกซื้อผลิตภัณฑ์ประกันชีวิต โดยใช้เทคนิคโครงข่ายประสาทเทียม และต้นไม้ตัดสินใจ

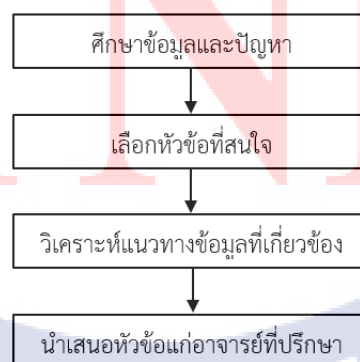
ขั้นที่ 5 การทดสอบระบบ

ขั้นที่ 6 การสรุปผลการวิจัย

โดยขั้นตอนการดำเนินงานวิจัยทั้ง 6 ขั้นตอน มีวิธีการและกระบวนการดำเนินงาน แบ่งเป็นรายละเอียดในแต่ละขั้นตอนดังนี้

3.1 ขั้นที่ 1 การวิเคราะห์ข้อมูลปัญหางานวิจัยต่างๆ

ผู้วิจัยได้ดำเนินการวิเคราะห์ข้อมูล เพื่อหาปัญหาทางด้านการแนะนำผลิตภัณฑ์แบบประกันชีวิตให้มีความเหมาะสมและตรงกับความต้องการของลูกค้า ดังต่อไปนี้



รูปที่ 3.1 ขั้นตอนการวิเคราะห์ข้อมูลปัญหางานวิจัย

3.1.1 ศึกษาข้อมูลและปัญหา

ศึกษาข้อมูลและปัญหา ซึ่งเป็นการศึกษาข้อมูลและเหตุการณ์ที่เกิดขึ้นจริง ณ ปัจจุบัน กับตัวแทนการขาย, ช่องทางการขายประกันชีวิต รวมถึงลูกค้าผู้ทำประกันชีวิต สำหรับตัวแทนและช่องทางการขายอื่นๆพบปัญหาการแนะนำและการนำเสนอผลิตภัณฑ์แบบประกันที่เหมาะสมและตรงกับความต้องการของลูกค้า และเพื่อลดปัญหาการที่ตัวแทนหรือช่องทางการขายอื่นๆนำเสนอผลิตภัณฑ์แบบประกันที่เกินกำลังความสามารถและเกินความจำเป็นต่อลูกค้า เพื่อผลประโยชน์ของลูกค้าผู้ทำประกันจะได้ทำประกันที่มีความเหมาะสมกับตนเอง มีความสามารถในการชำระเบี้ยประกัน และได้สิทธิประโยชน์สูงสุดตามที่ลูกค้าต้องการ รวมถึงเพื่อลดปัญหาการยกเลิกทำประกันของลูกค้า ซึ่งเป็นสาเหตุให้บริษัทประกันชีวิตขาดรายได้และสูญเสียลูกค้าไป

3.1.2 การเลือกหัวข้อปัญหาที่สนใจ

จากการศึกษาข้อมูลและปัญหา รวมถึงได้มีโอกาสทำงานในด้านธุรกิจประกันชีวิต จึงมีความสนใจที่จะพัฒนาในแง่ทัศนคติและภาพรวมของการทำประกันชีวิต ดังนั้น จึงตั้งใจทำงานวิจัยนี้เพื่อให้ความสำคัญของการทำประกันชีวิตระหว่างบริษัทประกันชีวิตและลูกค้าหรือผู้เอาประกัน รวมถึงเพื่อการพัฒนาต้นแบบ Web Application การแนะนำและวิธีการนำเสนอผลิตภัณฑ์แบบประกันได้อย่างมีอาชีพสำหรับธุรกิจประกันชีวิต

3.1.3 วิเคราะห์แนวทางข้อมูลที่เกี่ยวข้อง

ในการเลือกหัวข้อของปัญหาที่กล่าวมาข้างต้น ทำให้ผู้วิจัยเห็นถึงปัจจัยความสำคัญของการทำประกันชีวิตของลูกค้า ซึ่งหากลูกค้าได้เลือกทำประกันชีวิตที่ไม่มีความเหมาะสมกับตนเอง หรือทำประกันชีวิตโดยไม่มีวัตถุประสงค์ที่ไม่มีเหตุจำเป็น ก็จะเป็นสาเหตุให้ลูกค้ายกเลิกกรมธรรม์หรือเลิกทำประกันไปในที่สุด ซึ่งเป็นผลเสียต่อบริษัทประกันชีวิตที่ทำให้บริษัทขาดรายได้และเสียลูกค้าไป รวมถึงการที่ลูกค้ายกเลิกกรมธรรม์ก่อนวันครบสัญญา ทำให้ลูกค้าสูญเสียผลประโยชน์ ดังนั้นการพัฒนาต้นแบบ Web Application เพื่อการแนะนำการเลือกซื้อผลิตภัณฑ์ประกันชีวิต โดยใช้เทคนิคโครงข่ายประสาทเทียม และต้นไม้ตัดสินใจ เพื่อให้บริษัทประกันสามารถแนะนำและนำเสนอผลิตภัณฑ์แบบประกันที่เหมาะสมและตรงกับความต้องการของลูกค้าทุกกลุ่มได้

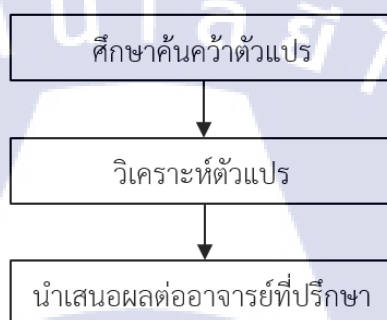
3.1.4 เสนอหัวข้อกับอาจารย์ที่ปรึกษา

หลังจากมีการวิเคราะห์ปัญหา ทฤษฎีการเรียนรู้ และการศึกษางานวิจัยที่เกี่ยวข้องต่างๆ ผู้วิจัยได้เลือกนำทฤษฎีเทคนิค Data Mining คือ เทคนิคต้นไม้ตัดสินใจ (Decisions Tree) และโครงข่ายประสาทเทียม (Neural Network) มาวิเคราะห์ข้อมูลร่วมกัน เพื่อผลลัพธ์ที่ถูกต้องและ

แม่นยำ จากนั้นจึงได้เสนอหัวข้อกับอาจารย์ที่ปรึกษา โดยใช้ชื่อหัวข้อว่า “การพัฒนาต้นแบบ Web Application เพื่อการแนะนำการเลือกซื้อผลิตภัณฑ์ประกันชีวิต โดยใช้เทคนิคโครงข่ายประสาทเทียม และต้นไม้ตัดสินใจ”

3.2 ขั้นที่ 2 การศึกษาตัวแปรที่เกี่ยวข้อง

ในกระบวนการศึกษาตัวแปรที่เกี่ยวข้องสำหรับการแนะนำผลิตภัณฑ์แบบประกันชีวิตแก่กลุ่มลูกค้าที่มีความสนใจและต้องการซื้อแบบประกันชีวิตได้อย่างเหมาะสมและตรงกับความต้องการของลูกค้า มีขั้นตอนการดำเนินการดังต่อไปนี้



รูปที่ 3.2 ขั้นตอนการศึกษาตัวแปรที่เกี่ยวข้องที่เกี่ยวข้อง

3.2.1 ศึกษาตัวแปรที่มีผลต่อการซื้อแบบประกันของลูกค้า

ผู้วิจัยได้ศึกษาถึงตัวแปรที่มีความเกี่ยวข้องกับการซื้อแบบประกันของลูกค้า โดยการศึกษาหาข้อมูล จากงานวิจัยท่านอื่น เอกสารต่างๆที่เกี่ยวข้อง และความเห็นชอบจากกลุ่มผู้เชี่ยวชาญ

3.2.2 วิเคราะห์ตัวแปรที่มีผลกระทบต่อการซื้อแบบประกันของลูกค้า

จากการศึกษาและวิเคราะห์ตัวแปรที่เกี่ยวข้องกับการซื้อแบบประกันของลูกค้า ซึ่งกระบวนการนี้ได้จากการศึกษาจากงานวิจัยท่านอื่นๆ ข้อมูลต่างๆ และความเห็นชอบจากกลุ่มผู้เชี่ยวชาญ พร้อมกับการนำเสนอผลที่ได้ศึกษามาต่ออาจารย์ที่ปรึกษา จนกว่าจะได้ตัวแปรที่เหมาะสม ซึ่งพบว่าตัวแปรที่เกี่ยวข้องมีดังต่อไปนี้ ตามตารางที่ 3.1

ตารางที่ 3.1 ตัวแปรที่เกี่ยวข้องกับการซื้อประกันชีวิตของลูกค้า

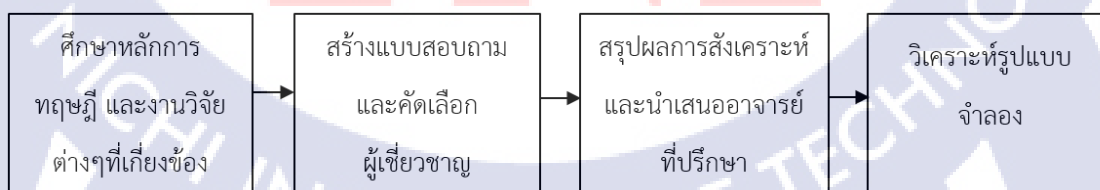
ตัวแปรต้น		ตัวแปรตาม
ปัจจัยพื้นฐาน	- เพศ (Gender) - อายุ (Age) - สถานะ (Status) - รายได้ต่อปี (Income) - จังหวัด (Province) - น้ำหนัก (Height) - ส่วนสูง (Weight) - อาชีพ (Occupation) - กลุ่มอาชีพ (Occupation Group) - ชั้นอาชีพ (Occupation Level)	- แบบประกันคุ้มครองชีวิต (Protective) - แบบประกันออมทรัพย์ (Saving) - แบบประกันลดหย่อนภาษี (Annuity) - แบบประกันควบการลงทุน (Unit Link)

3.2.2 นำเสนอผลต่อที่ปรึกษา

โดยผู้วิจัยได้นำเสนอผลการศึกษาและการวิเคราะห์ตัวแปรที่เป็นปัจจัยการซื้อแบบประกันชีวิต ต่ออาจารย์ที่ปรึกษา ก่อนจะนำเข้าสู่กระบวนการวิเคราะห์หรือการทำเหมืองข้อมูลต่อไป

3.3 ขั้นที่ 3 การสังเคราะห์และวิเคราะห์แบบจำลองการพัฒนาต้นแบบ Web Application เพื่อการแนะนำการเลือกซื้อแบบประกันชีวิต โดยใช้เทคนิคโครงข่ายประสาทเทียมและต้นไม้ตัดสินใจ

ผู้วิจัยได้ดำเนินการสังเคราะห์และวิเคราะห์แบบจำลองการพัฒนาต้นแบบ Web Application เพื่อการแนะนำการเลือกซื้อแบบประกันชีวิต โดยใช้เทคนิคโครงข่ายประสาทเทียมและต้นไม้ตัดสินใจ โดยมีขั้นตอนการดำเนินการวิจัยดังต่อไปนี้



รูปที่ 3.3 ขั้นตอนการสังเคราะห์และวิเคราะห์แบบจำลอง

3.3.1 ศึกษาหลักการ แนวคิด ทฤษฎี เอกสาร และงานวิจัยที่เกี่ยวข้อง

เป็นขั้นตอนในการศึกษาหลักการ แนวคิด ทฤษฎี เอกสาร และงานวิจัยที่เกี่ยวข้อง เพื่อแนวทางในการสังเคราะห์แบบจำลองการแนะนำการเลือกซื้อผลิตภัณฑ์ประกันชีวิต ซึ่งมีเนื้อหาที่ทำการศึกษาดังต่อไปนี้

- ทฤษฎีและหลักการพื้นฐานเกี่ยวกับการประกันภัย
- ปัจจัยที่ส่งผลต่อการเลือกทำประกันชีวิต
- ทฤษฎีเกี่ยวกับการทำเหมืองข้อมูล (Data Mining)
- การทำเหมืองข้อมูลโดยใช้เทคนิคโครงข่ายประสาทเทียม (Neural Network)
- การทำเหมืองข้อมูลโดยใช้เทคนิคต้นไม้ตัดสินใจ (Decision Tree)
- การทำเหมืองข้อมูลโดยใช้เทคนิคการจำแนกกลุ่ม (K-means Clustering)
- การศึกษางานวิจัยต่างๆที่เกี่ยวข้อง

หลังจากการศึกษาและวิเคราะห์ข้อมูลต่างๆ แล้ว ผู้วิจัยได้กำหนดกรอบแนวคิดเกี่ยวกับการสังเคราะห์แบบจำลองการแนะนำการเลือกซื้อผลิตภัณฑ์ประกันชีวิต ที่วิเคราะห์ด้วยเทคนิคเหมืองข้อมูล (Data Mining) ซึ่งประกอบด้วย 3 โมดูล ตามรูปที่ 3.4 มีรายละเอียดดังนี้

1. โมดูล Gathering Preparation เป็นขั้นตอนการเตรียมข้อมูลหรือ Data Set ก่อนนำไปวิเคราะห์ ข้อมูล สำหรับการทำวิจัยนี้จะเป็นข้อมูลจากกลุ่มตัวอย่างจำนวน 3,290 คน ที่ได้ซื้อผลิตภัณฑ์ประกันชีวิตของบริษัทประกันชีวิตแห่งหนึ่ง ซึ่งประกอบไปด้วยข้อมูลประชากรศาสตร์ของลูกค้า ดังนี้ อายุ, เพศ, อาชีพ, ขั้นอาชีพ, รายได้ต่อปี, แบบประกันหลักที่ซื้อ, ทุนประกันชีวิต, เบี้ยประกันชีวิต, งวดการชำระเบี้ย, วิธีการชำระเบี้ย, ระยะเวลาเอาประกัน, แบบประกันสัญญาเพิ่มเติม, ทุนประกันสัญญาเพิ่มเติม, เบี้ยประกันสัญญาเพิ่มเติม และระยะเวลาเอาประกันสัญญาเพิ่มเติม ซึ่งเป็นข้อมูลปัจจัยที่มีผลต่อการตัดสินใจทำประกันชีวิต เพื่อใช้ในการวิเคราะห์ข้อมูลหาผลลัพธ์ (Output) ในการแนะนำการเลือกซื้อแบบประกันชีวิต 4 ประเภทแบบประกันดังนี้ คือ แบบประกันคุ้มครองชีวิต (Protective), แบบประกันลดหย่อนภาษีหรือบำนาญ (Annuity), แบบประกันออมทรัพย์ (Saving), และแบบประกันควบการลงทุน (Unit Link)

2. โมดูล Modeling Process เป็นขั้นตอนการแปลงสภาพข้อมูล (Data Transformation) ให้อยู่ในรูปแบบที่พร้อมใช้งานและนำไปวิเคราะห์ข้อมูลด้วยเทคนิคการทำเหมืองข้อมูล (Data Mining Technique) ต่อไป

2.1 การแปลงสภาพข้อมูล (Data Transformation) เป็นการตรวจสอบความถูกต้องและความสมบูรณ์ของข้อมูล (Data Cleaning) เพื่อให้ได้ข้อมูลที่ครบถ้วนและเหมาะสมในการทำเหมืองข้อมูล เช่น การกำจัดข้อมูลในตารางที่เป็นค่าว่าง (Null) การปรับรูปแบบข้อมูลให้

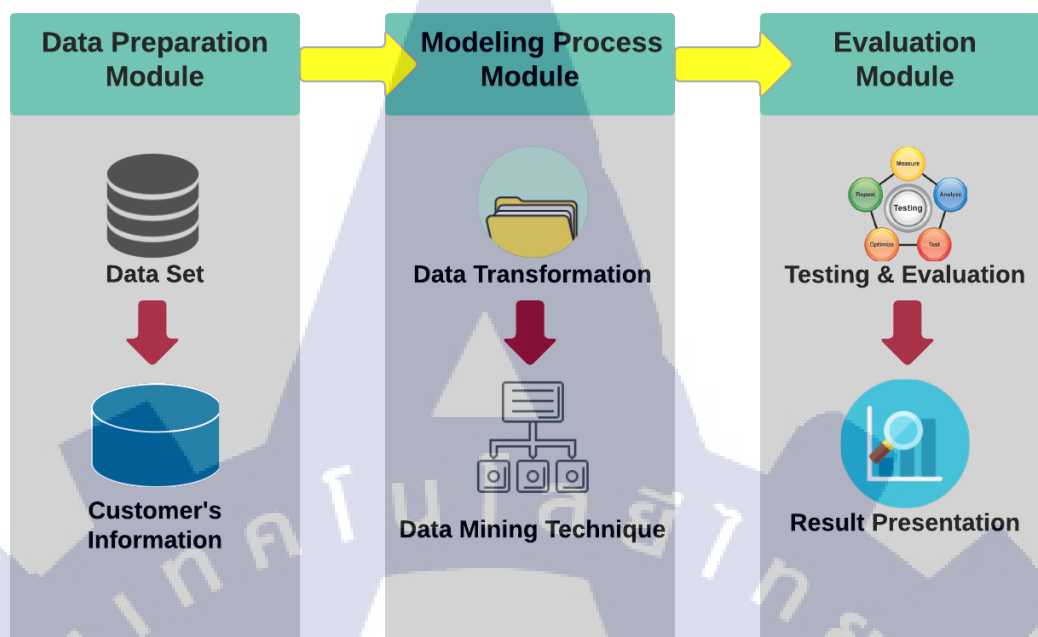
เหมือนกัน การจัดกลุ่มข้อมูลเพื่อลดการกระจาย (Binning Data) และการปรับรูปแบบของข้อมูลให้อยู่ในรูปแบบของตัวเลข เป็นต้น ซึ่งรวมถึงการแปรรูปของข้อมูลให้พร้อมใช้งาน อาจอยู่ในรูปของไฟล์ excel หรือ csv

2.2 การวิเคราะห์ข้อมูลด้วยเทคนิคการทำเหมืองข้อมูล (Data Mining Technique) เป็นขั้นตอนการนำเทคนิค Data Mining มาใช้ในการวิเคราะห์ข้อมูล หลังจากมีการเตรียมข้อมูลเป็นที่เรียบร้อยแล้ว จะทำการโหลดข้อมูลเพื่อสร้างรูปแบบโดยใช้เครื่องมือที่ชื่อว่าเวก้า (WEKA) Version 3.8 เพื่อช่วยในการวิเคราะห์และจำแนกข้อมูล สำหรับการสร้างรูปแบบการทำเหมืองข้อมูลของงานวิจัยนี้ จะเลือกใช้เทคนิคต้นไม้ตัดสินใจ (Decision Tree) ในการทำ feature selection เลือกใช้เทคนิค K-means Clustering ในการแบ่งกลุ่มลูกค้า และเทคนิคโครงข่ายประสาทเทียม (Neural Network) ในสร้างรูปแบบการแนะนำการเลือกซื้อผลิตภัณฑ์ประกันชีวิต

3. โมดูล Evaluation เป็นขั้นตอนการทดสอบและประเมินผลการวิเคราะห์รูปแบบการแนะนำการเลือกซื้อผลิตภัณฑ์แบบประกัน (Testing & Evaluation) และการพัฒนาต้นแบบ Web Application เพื่อแสดงผลการแนะนำการเลือกซื้อผลิตภัณฑ์แบบประกัน (Result Presentation)

3.1 การทดสอบและประเมินผลประสิทธิภาพของการวิเคราะห์รูปแบบการแนะนำการเลือกซื้อผลิตภัณฑ์แบบประกันชีวิต (Testing & Evaluation) เพื่อวัดผลเปอร์เซ็นต์ความถูกต้องและความแม่นยำในการแนะนำ โดยใช้วิธีการทดสอบโดยแบ่งชุดข้อมูล 20:80 (Evaluation on Training Set) โดยใช้ชุดข้อมูลจำนวนร้อยละ 80 สำหรับ Training และร้อยละ 20 สำหรับการทดสอบ รวมถึงทดสอบด้วยการคำนวณ Precision & Recall เพื่อการวัดผลและประเมินประสิทธิภาพของรูปแบบ

3.2 การพัฒนาต้นแบบ Web Application เพื่อแสดงผลการแนะนำการเลือกซื้อผลิตภัณฑ์แบบประกัน (Result Presentation) ที่สามารถใช้งานบน web browser เพื่อให้ผู้ใช้งานที่เป็นตัวแทนขายประกันชีวิต สามารถแนะนำและนำเสนอผลิตภัณฑ์แบบประกันให้แก่ลูกค้า



รูปที่ 3.4 กรอบแนวคิดของงานวิจัย (Conceptual Framework)

3.3.2 การสร้างแบบสอบถามและคัดเลือกผู้เชี่ยวชาญ

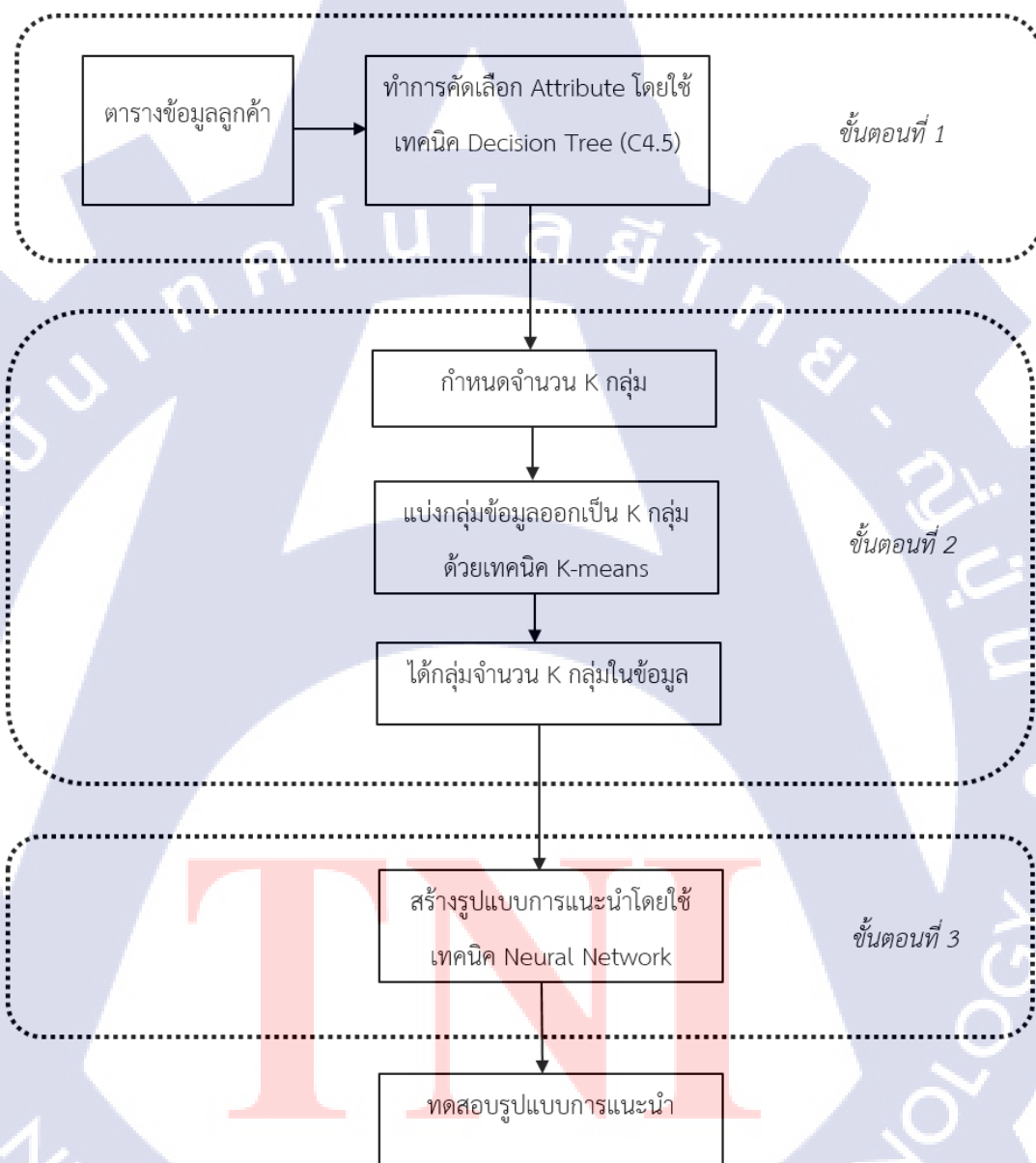
เป็นขั้นตอนการสร้างแบบสอบถามสำหรับงานวิจัยนี้ เพื่อประเมินผลกรอบแนวคิดหรือแบบจำลองที่ได้สังเคราะห์ขึ้น ซึ่งแบ่งผลระดับการประเมินออกเป็น 5 ระดับ และทำการคัดเลือกผู้เชี่ยวชาญจำนวน 5 ท่านที่มีความรู้และประสบการณ์เกี่ยวกับการทำประกันชีวิต ทำการประเมินความเหมาะสมของแบบจำลองนี้ โดยการทำแบบสอบถามที่ได้สร้างขึ้น

3.3.3 สรุปผลการสังเคราะห์และนำเสนอต่ออาจารย์ที่ปรึกษา

เป็นขั้นตอนการสรุปผลการประเมินกรอบแนวคิดหรือแบบจำลองจากผู้เชี่ยวชาญที่คัดเลือกทั้งหมดจำนวน 5 ท่าน และนำผลการประเมินนี้หาค่าเฉลี่ย (Mean) และส่วนเบี่ยงเบนมาตรฐาน (Standard Division) แล้วนำผลสรุปการประเมินนี้มานำเสนอต่ออาจารย์ที่ปรึกษา เพื่อให้อาจารย์ที่ปรึกษาทำการตรวจสอบและปรับปรุงแก้ไขให้มีความถูกต้องและมีความเหมาะสมมากยิ่งขึ้น

3.3.4 การวิเคราะห์แบบจำลองการแนะนำการเลือกซื้อผลิตภัณฑ์ประกันชีวิต

เป็นขั้นตอนการวิเคราะห์เพื่อหาแบบจำลองการแนะนำการเลือกซื้อผลิตภัณฑ์ประกัน โดยใช้เทคนิคการทำเหมืองข้อมูล (Data Mining Technique) มีขั้นตอนดังต่อไปนี้ ตามรูปที่ 3.5



รูปที่ 3.5 ขั้นตอนการวิเคราะห์แบบจำลองการแนะนำการเลือกซื้อผลิตภัณฑ์ประกันชีวิต

จากรูปที่ 3.5 นักวิจัยได้แบ่งขั้นตอนการวิเคราะห์เพื่อหาแบบจำลองการแนะนำการเลือกซื้อผลิตภัณฑ์แบบประกันออกเป็น 3 ขั้นตอน ได้แก่ ขั้นตอนแรกเป็นการคัดเลือก Attribute หรือการกรองและคัดเลือกคุณลักษณะที่สำคัญและมีความเกี่ยวข้องในการแสดงผลการแนะนำการเลือกซื้อผลิตภัณฑ์แบบประกัน โดยเลือกใช้เทคนิคต้นไม้ตัดสินใจในการทำสำหรับขั้นตอนนี้ เพราะสามารถรองรับการวิเคราะห์ข้อมูลได้ทุกประเภทของข้อมูลเป็นอย่างดีและให้ผลลัพธ์ที่มีประสิทธิภาพสูง ขั้นตอนที่สองเป็นการแบ่งกลุ่มข้อมูลหลังจากทำการคัดเลือก Attribute ไปแล้ว เป็นการแบ่งกลุ่มข้อมูลที่มีความคล้ายคลึงกันจัดให้อยู่กลุ่มเดียวกัน ซึ่งจะทำให้ข้อมูลเป็นระเบียบและทำให้ง่ายต่อการวิเคราะห์ในขั้นตอนต่อไป จึงเลือกใช้เทคนิค Clustering K-means สำหรับทำการแบ่งกลุ่มข้อมูลนี้ เพราะสามารถจับกลุ่มข้อมูลที่มีความคล้ายคลึงกันได้ดี รวมถึงมีการประมวลผลที่รวดเร็ว ขั้นตอนสุดท้าย เป็นการวิเคราะห์ข้อมูลเพื่อสร้างแบบจำลองการแนะนำการเลือกซื้อผลิตภัณฑ์แบบประกัน โดยเลือกใช้เทคนิคโครงข่ายประสาทเทียม ที่สามารถรองรับการวิเคราะห์ข้อมูลที่มีจำนวนมากและมีความซับซ้อน อีกทั้งให้ผลลัพธ์ที่มีความแม่นยำสูง ซึ่งแต่ละขั้นตอนมีรายละเอียด ดังต่อไปนี้

3.3.4.1 **ขั้นตอนที่ 1** เป็นขั้นตอนการนำตารางข้อมูลลูกค้าและทำการคัดเลือก Attribute (Feature Selection) โดยใช้เทคนิคต้นไม้ตัดสินใจ Decision Tree (C4.5) เพื่อหาความสัมพันธ์ระหว่าง Attribute กับ Class โดยการคำนวณค่า Information Gain (IG) หรือค่าน้ำหนักในการคัดเลือก Attribute ที่สำคัญมาวิเคราะห์ข้อมูล การคำนวณค่า Information Gain คำนวณได้จากสมการ ดังนี้

$$IG(\text{parent, child}) = \text{entropy}(\text{parent}) - [p(c1) * \text{entropy}(c1) + \dots] \quad (3.1)$$

โดยที่ $\text{entropy}(c1) = -p(c1) \log p(c1)$ ซึ่ง $p(c1)$ คือ ค่าความน่าจะเป็นของ $c1$ ซึ่งจะทำให้การคัดเลือก Attribute ที่มีค่า IG สูงเพื่อการวิเคราะห์ให้มีประสิทธิภาพ จากตัวอย่างวิธีการคำนวณเพื่อหาค่า IG จากตัวอย่างตารางข้อมูลผู้ซื้อประกันชีวิต ตามตารางที่ 3.2 มีดังต่อไปนี้

ตารางที่ 3.2 ตารางตัวอย่างข้อมูลผู้ซื้อประกันชีวิต กรณีศึกษาบริษัทประกันชีวิตแห่งหนึ่ง

Gender	Age	Status	Salary	Province	Height	Weight	Occupation	Occ_Group	Level	Plan
M	53	Married	360000	Surat Thani	170	65	Grocery	Trading	1	Protective
M	60	Single	0	Chiang Mai	165	62	Housemaid	No occupational	1	Protective
F	25	Single	360000	Nonthaburi	155	55	SellingFood	Trading	2	Saving
F	4	Single	0	Udon Thani	103	21	Student	No occupational	1	Saving
M	60	Widowed	90000	Udon Thani	166	75	Registrar	Police	1	Protective
F	49	Single	150000	Ang Thong	143	63	Nurse	Doctor-Hospital	1	Annuity
M	33	Married	350000	Nakhon Si Thammarat	160	65	Business Owner	Private Enterprise	1	Protective
F	32	Single	360000	Surin	157	49	Nurse	Doctor-Hospital	1	UnitLink
M	49	Married	600000 0	Chiang Mai	176	69	Business Owner	Private Enterprise	1	Saving
M	59	Married	100000 0	Surat Thani	165	69	Business Owner	Private Enterprise	1	UnitLink

จากตารางตัวอย่างข้อมูลลูกค้ามีทั้งหมด 10 คุณสมบัติ (Attributes) ดังต่อไปนี้

1. Gender (เพศ) มี 2 Instance ได้แก่

1.2 Gender = M

1.2 Gender = F

2. Age (อายุ) มี 8 Instance ได้แก่

2.1 Age = 53

2.2 Age = 60

2.3 Age = 25

2.4 Age = 4

2.5 Age = 49

2.6 Age = 33

2.7 Age = 32

2.8 Age = 59

3. Status (สถานะ) มี 3 Instance ได้แก่

3.1 Status = Married

3.2 Status = Single

3.3 Status = Widowed

4. Income (รายได้ต่อปี) มี 7 Instance ได้แก่

4.1 Income = 360,000

4.2 Income = 0

4.3 Income = 90,000

4.4 Income = 150,000

4.5 Income = 350,000

4.6 Income = 6,000,000

4.7 Income = 1,000,000

5. Province (จังหวัดที่อยู่) มี 7 Instance ได้แก่

5.1 Province = Surat Thani

5.2 Province = Chiang Mai

5.3 Province = Nonthaburi

5.4 Province = Udon Thani

5.5 Province = Nakhon Si Thammarat

5.6 Province = Ang Thong

5.7 Province = Surin

6. Height (ความสูง) มี 9 Instance ได้แก่

6.1 Height = 170

6.2 Height = 165

6.3 Height = 155

6.4 Height = 103

6.5 Height = 166

6.6 Height = 143

6.7 Height = 160

6.8 Height = 157

6.9 Height = 176

7. Weight (น้ำหนัก) มี 8 Instance ได้แก่

7.1 Weight = 65

7.2 Weight = 62

7.3 Weight = 55

7.4 Weight = 21

7.5 Weight = 75

7.6 Weight = 63

7.7 Weight = 49

7.8 Weight = 69

8. Occupation (อาชีพ) มี 7 Instance ได้แก่

8.1 Occupation = Grocery

8.2 Occupation = Housemaid

8.3 Occupation = SellingFood

8.4 Occupation = Student

8.5 Occupation = Registrar

8.6 Occupation = Nurse

8.7 Occupation = Business Owner

9. Occupation Group (กลุ่มอาชีพ) มี 5 Instance ได้แก่

9.1 Occupation Group = Trading

9.2 Occupation Group = No Occupation

9.3 Occupation Group = Police

9.4 Occupation Group = Doctor-Hospital

9.5 Occupation Group = Private Enterprise

10. Occupation Level (ชั้นอาชีพ) มี 2 Instance ได้แก่

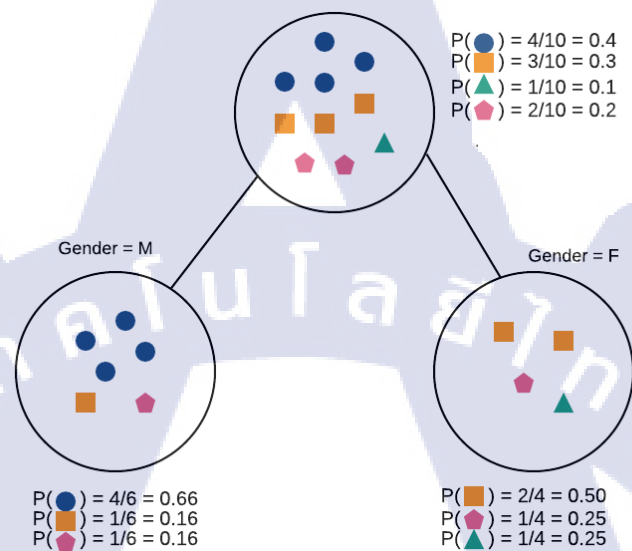
10.1 Occupation Level = 1

10.2 Occupation Level = 2

สำหรับ Attribute **Plan** จะถูกกำหนดให้เป็น Predefine Class

การคำนวณค่า Information Gain ของแต่ละ Attribute จากตัวอย่างตารางข้างต้น
มีวิธีการทำดังต่อไปนี้

1. คำนวณค่า IG ของ Gender มีดังต่อไปนี้



รูปที่ 3.6 การคำนวณหาค่า Information Gain ของ Gender

1.1. เริ่มต้นจากการคำนวณค่า Entropy ของ Class

$$\begin{aligned}
 \text{Entropy (Plan)} &= -p(1) * \log_2 p(1) - p(2) * \log_2 p(2) - \\
 &\quad p(3) * \log_2 p(3) - p(4) * \log_2 p(4) \\
 &= -[0.4 * -1.32 + 0.3 * -1.74 + 0.1 * \\
 &\quad -3.32 + 0.2 * -2.32] \\
 &= 1.846
 \end{aligned}$$

1.2. Entropy ของ Gender = M

$$\begin{aligned}
 \text{Entropy (Gender = M)} &= -[0.66 * \log_2 (0.66) + 0.16 * \log_2 (0.16) \\
 &\quad + 0.16 * \log_2 (0.16)] \\
 &= -[0.66 * -0.59 + 0.16 * -2.64 + 0.16 * \\
 &\quad -2.64] \\
 &= 1.23
 \end{aligned}$$

$$p(\text{Gender} = M) = 6/10 = 0.6$$

1.3. Entropy ของ Gender = F

$$\begin{aligned} \text{Entropy}(\text{Gender} = F) &= - [0.50 * \log_2(0.50) + 0.25 * \log_2(0.25) \\ &\quad + 0.25 * \log_2(0.25)] \\ &= - [0.50 * -1.00 + 0.25 * -2.00 + 0.25 * \\ &\quad -2.00] \\ &= 1.50 \end{aligned}$$

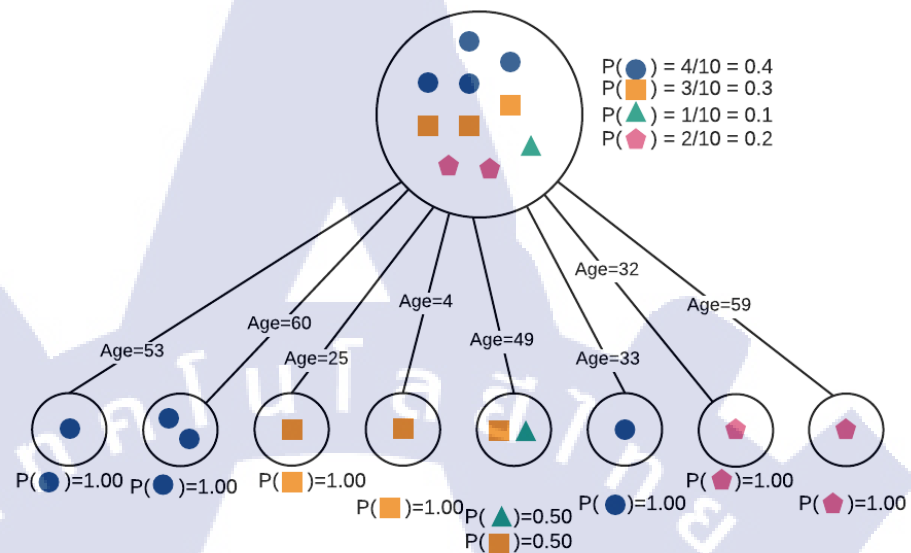
$$p(\text{Gender} = F) = 4/10 = 0.4$$

ดังนั้น

$$\begin{aligned} \text{IG}(\text{Plan}, \text{Gender}) &= \text{Entropy}(\text{Plan}) - [p(\text{Gender} = M) * \\ &\quad \text{Entropy}(\text{Gender} = M) + p(\text{Gender} = F) * \\ &\quad \text{Entropy}(\text{Gender} = F)] \\ &= 1.84 - [0.6 * 1.23 + 0.4 * 1.5] \\ &= 0.50 \end{aligned}$$

TNI

2. คำนวณค่า IG ของ Age มีดังต่อไปนี้



รูปที่ 3.7 การคำนวณหาค่า Information Gain ของ Age

2.1. Entropy ของ Age = 53

$$\begin{aligned} \text{Entropy (Age = 53)} &= - [1 * \log_2 (1)] \\ &= 0 \end{aligned}$$

$$p (\text{Age} = 53) = 1/10 = 0.1$$

2.2. Entropy ของ Age = 60

$$\begin{aligned} \text{Entropy (Age = 60)} &= - [1 * \log_2 (1)] \\ &= 0 \end{aligned}$$

$$p (\text{Age} = 60) = 1/10 = 0.1$$

2.3. Entropy ของ Age = 25

$$\begin{aligned} \text{Entropy (Age = 25)} &= - [1 * \log_2 (1)] \\ &= 0 \end{aligned}$$

$$p (\text{Age} = 25) = 1/10 = 0.1$$

2.4. Entropy ของ Age = 4

$$\begin{aligned}\text{Entropy (Age = 4)} &= - [1 * \log_2 (1)] \\ &= 0\end{aligned}$$

$$p (\text{Age} = 4) = 1/10 = 0.1$$

2.5. Entropy ของ Age = 49

$$\begin{aligned}\text{Entropy (Age = 49)} &= - [0.5 * \log_2 (0.5) + 0.5 * \log_2 (0.5)] \\ &= 1\end{aligned}$$

$$p (\text{Age} = 49) = 2/10 = 0.2$$

2.6. Entropy ของ Age = 33

$$\begin{aligned}\text{Entropy (Age = 33)} &= - [1 * \log_2 (1)] \\ &= 0\end{aligned}$$

$$p (\text{Age} = 33) = 1/10 = 0.1$$

2.7. Entropy ของ Age

$$\begin{aligned}\text{Entropy (Age = 32)} &= - [1 * \log_2 (1)] \\ &= 0\end{aligned}$$

$$p (\text{Age} = 32) = 1/10 = 0.1$$

2.8. Entropy ของ Age = 59

$$\begin{aligned}\text{Entropy (Age = 59)} &= - [1 * \log_2 (1)] \\ &= 0\end{aligned}$$

$$p (\text{Age} = 59) = 1/10 = 0.1$$

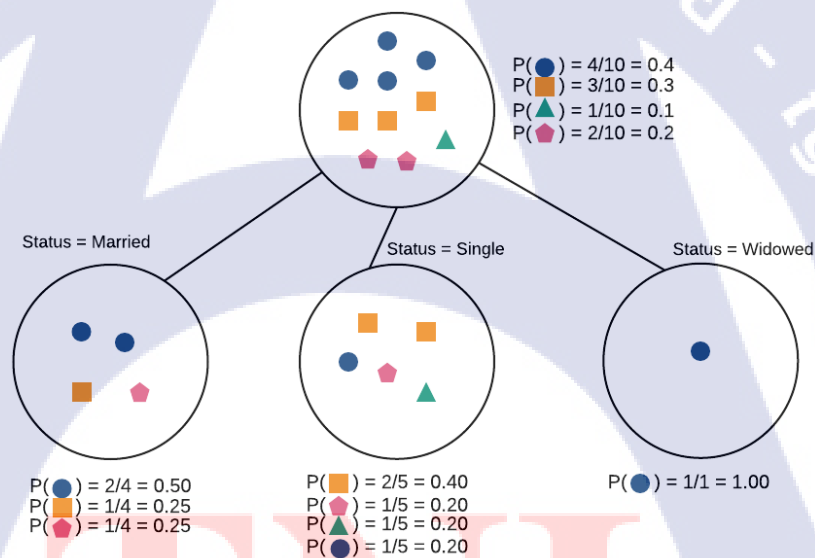
ดังนั้น

IG (Plan, Age)

$$\begin{aligned}&= \text{Entropy (Plan)} - [p (\text{Age} = 53) \\ &\quad * \text{Entropy (Age = 53)} + p (\text{Age} = 60) * \\ &\quad \text{Entropy (Age = 60)} + p (\text{Age} = 25) * \\ &\quad \text{Entropy (Age = 25)} \\ &\quad + p (\text{Age} = 4) * \text{Entropy (Age = 4)} +\end{aligned}$$

$$\begin{aligned}
 & p(\text{Age} = 49) * \text{Entropy}(\text{Age} = 49) + \\
 & p(\text{Age} = 33) * \text{Entropy}(\text{Age} = 33) + \\
 & p(\text{Age} = 32) * \text{Entropy}(\text{Age} = 32) + \\
 & p(\text{Age} = 59) * \text{Entropy}(\text{Age} = 59)] \\
 & = 1.84 - [0.1 * 0 + 0.2 * 0.5 + 0.1 * 0 + \\
 & 0.1 * 0 + 0.2 * 1 + 0.1 * 0 + 0.1 * 0 + 0.1 \\
 & * 0] \\
 & = 1.54
 \end{aligned}$$

3. คำนวณค่า IG ของ Status มีดังต่อไปนี้



รูปที่ 3.8 การคำนวณหาค่า Information Gain ของ Status

3.1. Entropy ของ Status = Married

$$\begin{aligned}
 \text{Entropy}(\text{Status} = \text{Married}) &= - [0.50 * \log_2(0.50) + 0.25 * \log_2 \\
 & (0.25) + 0.25 * \log_2(0.25)] \\
 &= - [0.50 * -1.00 + 0.25 * -2.00 + 0.25 * \\
 & -2.00]
 \end{aligned}$$

$$= 1.50$$

$$p(\text{Status} = \text{Married}) = 4/10 = 0.4$$

3.2. Entropy ของ Status = Single

$$\begin{aligned} \text{Entropy}(\text{Status} = \text{Single}) &= - [0.40 * \log_2(0.40) + 0.20 * \log_2(0.20) \\ &\quad + 0.20 * \log_2(0.20) + 0.20 * \log_2(0.20)] \\ &= - [0.40 * -1.32 + 0.20 * -2.32 + 0.20 * \\ &\quad -2.32 + 0.20 * -2.32] \\ &= 1.92 \end{aligned}$$

$$p(\text{Status} = \text{Single}) = 5/10 = 0.5$$

3.3. Entropy ของ Status = Widowed

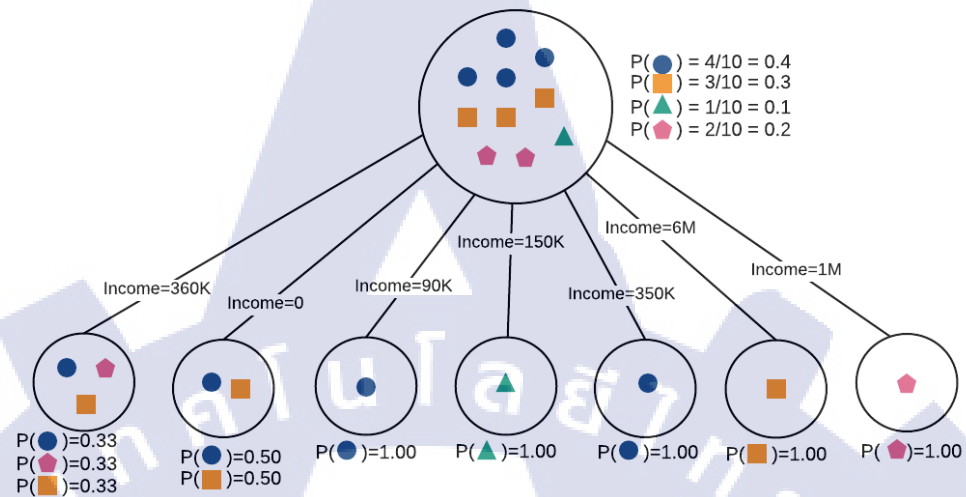
$$\begin{aligned} \text{Entropy}(\text{Status} = \text{Widowed}) &= - [1 * \log_2(1)] \\ &= 0 \end{aligned}$$

$$p(\text{Status} = \text{Widowed}) = 1/10 = 0.1$$

ดังนั้น

$$\begin{aligned} \text{IG}(\text{Plan}, \text{Status}) &= \text{Entropy}(\text{Plan}) - [p(\text{Status} = \text{Married}) * \\ &\quad \text{Entropy}(\text{Status} = \text{Married}) + p(\text{Status} = \\ &\quad \text{Single}) * \text{Entropy}(\text{Status} = \text{Single}) + p \\ &\quad (\text{Status} = \text{Widowed}) * \text{Entropy}(\text{Status} = \\ &\quad \text{Widowed})] \\ &= 1.84 - [0.4 * 1.50 + 0.5 * 1.92 + 0 * 0.1] \\ &= 0.28 \end{aligned}$$

4. คำนวณค่า IG ของ Income มีดังต่อไปนี้



รูปที่ 3.9 การคำนวณหาค่า Information Gain ของ Income

4.1. Entropy ของ Income = 360K

$$\begin{aligned}
 \text{Entropy (Income = 360K)} &= - [0.33 * \log_2 (0.33) + 0.33 * \log_2 (0.33) + 0.33 * \log_2 (0.33)] \\
 &= - [0.33 * -1.59 + 0.33 * -1.59 + 0.33 * -1.59] \\
 &= 1.57
 \end{aligned}$$

$$p(\text{Income} = 360K) = 3/10 = 0.3$$

4.2. Entropy ของ Income = 0

$$\begin{aligned}
 \text{Entropy (Income = 0)} &= - [0.50 * \log_2 (0.50) + 0.50 * \log_2 (0.50)] \\
 &= - [0.50 * -1 + 0.50 * -1] \\
 &= 1.00
 \end{aligned}$$

$$p(\text{Income} = 0) = 2/10 = 0.2$$

4.3. Entropy ของ Income = 90K

$$\begin{aligned}\text{Entropy (Income = 90K)} &= - [1 * \log_2 (1)] \\ &= 0\end{aligned}$$

$$p (\text{Income} = 90K) = 1/10 = 0.1$$

4.4. Entropy ของ Income = 150K

$$\begin{aligned}\text{Entropy (Income = 150K)} &= - [1 * \log_2 (1)] \\ &= 0\end{aligned}$$

$$P (\text{Income} = 150K) = 1/10 = 0.1$$

4.5. Entropy ของ Income = 350K

$$\begin{aligned}\text{Entropy (Income = 350K)} &= - [1 * \log_2 (1)] \\ &= 0\end{aligned}$$

$$p (\text{Income} = 350K) = 1/10 = 0.1$$

4.6. Entropy ของ Income = 6M

$$\begin{aligned}\text{Entropy (Income = 6M)} &= - [1 * \log_2 (1)] \\ &= 0\end{aligned}$$

$$p (\text{Income} = 6M) = 1/10 = 0.1$$

4.7. Entropy ของ Income = 1M

$$\begin{aligned}\text{Entropy (Income = 1M)} &= - [1 * \log_2 (1)] \\ &= 0\end{aligned}$$

$$p (\text{Income} = 1M) = 1/10 = 0.1$$

ดังนั้น

IG (Plan, Income)

$$\begin{aligned}&= \text{Entropy (Plan)} - [p (\text{Income}=360K) * \\ &\text{Entropy (Income}=360K) + p (\text{Income} = 0) \\ &* \text{Entropy (Income}=0) + p (\text{Income}=90K) \\ &* \text{Entropy (Income}=90K) + \\ &p (\text{Income}=150K) * \text{Entropy (Income}=150K)\end{aligned}$$

$$+ p(\text{Income}=350K) *$$

$$\text{Entropy}(\text{Income}=350K)$$

$$+ p(\text{Income}=6M) * \text{Entropy}(\text{Income}=6M)$$

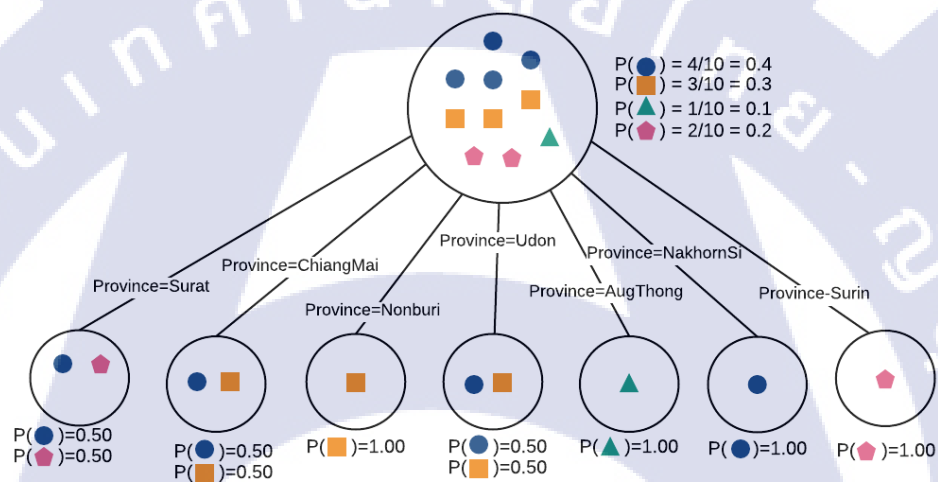
$$+ p(\text{Income}=1M) * \text{Entropy}(\text{Income}=1M)]$$

$$= 1.84 - [0.3 * 1.57 + 0.2 * 1 + 0.1 * 0 +$$

$$0.1 * 0 + 0.1 * 0 + 0.1 * 0 + 0.1 * 0]$$

$$= 1.16$$

5. คำนวณค่า IG ของ Province มีดังต่อไปนี้



รูปที่ 3.10 การคำนวณหาค่า Information Gain ของ Province

5.1. Entropy ของ Province = Surat

$$\text{Entropy}(\text{Province} = \text{Surat}) = - [0.50 * \log_2(0.50) + 0.50 * \log_2(0.50)]$$

$$= - [0.50 * -1 + 0.50 * -1]$$

$$= 1$$

$$p(\text{Province} = \text{Surat}) = 2/10 = 0.2$$

5.2. Entropy ของ Province = ChiangMai

$$\text{Entropy}(\text{Province} = \text{ChiangMai}) = - [0.50 * \log_2(0.50) + 0.50 * \log_2(0.50)]$$

$$= - [0.50 * -1 + 0.50 * -1]$$

$$= 1$$

$$= 1$$

$$p(\text{Province} = \text{ChiangMai}) = 2/10 = 0.2$$

5.3. Entropy ของ Province = Nonburi

$$\text{Entropy}(\text{Province} = \text{Nonburi}) = -[1 * \log_2(1)]$$

$$= 0$$

$$p(\text{Province} = \text{Nonburi}) = 1/10 = 0.1$$

5.4. Entropy ของ Province = Udon

$$\text{Entropy}(\text{Province} = \text{Udon}) = -[0.50 * \log_2(0.50) + 0.50 * \log_2(0.50)]$$

$$= -[0.50 * -1 + 0.50 * -1]$$

$$= 1$$

$$p(\text{Province} = \text{Udon}) = 2/10 = 0.2$$

5.5. Entropy ของ Province = AngThong

$$\text{Entropy}(\text{Province} = \text{AngThong}) = -[1 * \log_2(1)]$$

$$= 0$$

$$p(\text{Province} = \text{AngThong}) = 1/10 = 0.1$$

5.6. Entropy ของ Province = NaKhornSi

$$\text{Entropy}(\text{Province} = \text{NaKhornSi}) = -[1 * \log_2(1)]$$

$$= 0$$

$$p(\text{Province} = \text{NaKhornSi}) = 1/10 = 0.1$$

5.7. Entropy ของ Province = Surin

$$\text{Entropy}(\text{Province} = \text{Surin}) = -[1 * \log_2(1)]$$

$$= 0$$

$$p(\text{Province} = \text{Surin}) = 1/10 = 0.1$$

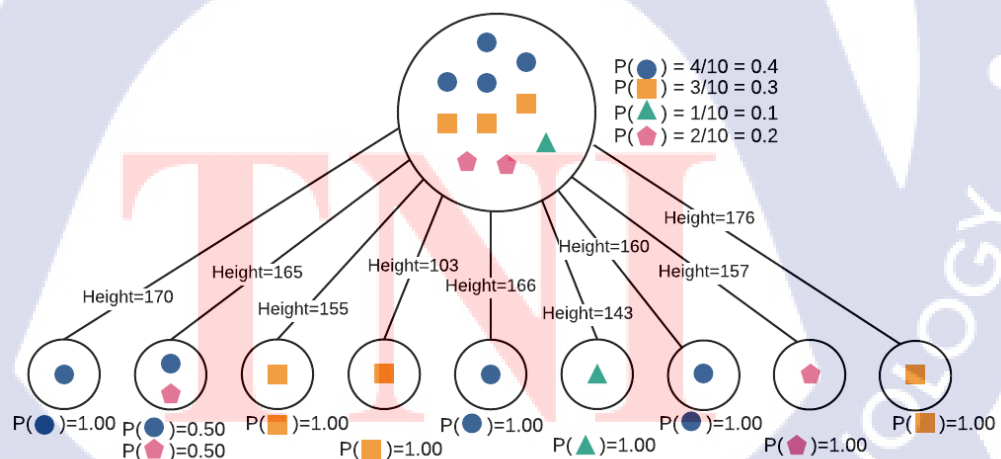
ดังนั้น

$$IG(\text{Plan}, \text{Province}) = \text{Entropy}(\text{Plan}) - [p(\text{Province} = \text{Surat}) * \text{Entropy}(\text{Province} = \text{Surat}) + p(\text{Province}$$

$$\text{Entropy}(\text{Province} = \text{Surat}) + p(\text{Province}$$

$$\begin{aligned}
 &= \text{ChiangMai} * \text{Entropy (Province =} \\
 &\text{ChiangMai)} + p(\text{Province = Nonburi}) * \\
 &\text{Entropy (Province = Nonburi)} + \\
 &p(\text{Province = Udon}) * \text{Entropy (Province} \\
 &= \text{Udon)} + p(\text{Province = AngThong}) * \\
 &\text{Entropy (Province = AngThong)} + \\
 &p(\text{Province = NaKhornSi}) * \\
 &\text{Entropy (Province = NaKhornSi)} + \\
 &p(\text{Province = Surin}) * \\
 &\text{Entropy (Province = Surin)} \\
 &= 1.84 - [0.2 * 1 + 0.2 * 1 + 0.1 * 0 + 0.2 * \\
 &1 + 0.1 * 0 + 0.1 * 0 + 0.1 * 0] \\
 &= 1.24
 \end{aligned}$$

6. คำนวณค่า IG ของ Height มีดังต่อไปนี้



รูปที่ 3.11 การคำนวณหาค่า Information Gain ของ Height

6.1. Entropy ของ Height = 170

$$\begin{aligned}\text{Entropy (Height = 170)} &= - [1 * \log_2 (1)] \\ &= 0\end{aligned}$$

$$p (\text{Height} = 170) = 1/10 = 0.1$$

6.2. Entropy ของ Height = 165

$$\begin{aligned}\text{Entropy (Height = 165)} &= - [0.5 * \log_2 (0.5) + 0.5 * \log_2 (0.5)] \\ &= - [0.50 * -1 + 0.50 * -1] \\ &= 1\end{aligned}$$

$$p (\text{Height} = 165) = 2/10 = 0.2$$

6.3. Entropy ของ Height = 155

$$\begin{aligned}\text{Entropy (Height = 155)} &= - [1 * \log_2 (1)] \\ &= 0\end{aligned}$$

$$p (\text{Height} = 155) = 1/10 = 0.1$$

6.4. Entropy ของ Height = 103

$$\begin{aligned}\text{Entropy (Height = 103)} &= - [1 * \log_2 (1)] \\ &= 0\end{aligned}$$

$$p (\text{Height} = 103) = 1/10 = 0.1$$

6.5. Entropy ของ Height = 166

$$\begin{aligned}\text{Entropy (Height = 166)} &= - [1 * \log_2 (1)] \\ &= 0\end{aligned}$$

$$p (\text{Height} = 166) = 1/10 = 0.1$$

6.6. Entropy ของ Height = 143

$$\begin{aligned}\text{Entropy (Height = 143)} &= - [1 * \log_2 (1)] \\ &= 0\end{aligned}$$

$$p (\text{Height} = 143) = 1/10 = 0.1$$

6.7. Entropy ของ Height = 160

$$\begin{aligned}\text{Entropy (Height = 160)} &= - [1 * \log_2 (1)] \\ &= 0\end{aligned}$$

$$p (\text{Height} = 160) = 1/10 = 0.1$$

6.8. Entropy ของ Height = 157

$$\begin{aligned}\text{Entropy (Height = 157)} &= - [1 * \log_2 (1)] \\ &= 0\end{aligned}$$

$$p (\text{Height} = 157) = 1/10 = 0.1$$

6.9. Entropy ของ Height = 176

$$\begin{aligned}\text{Entropy (Height = 176)} &= - [1 * \log_2 (1)] \\ &= 0\end{aligned}$$

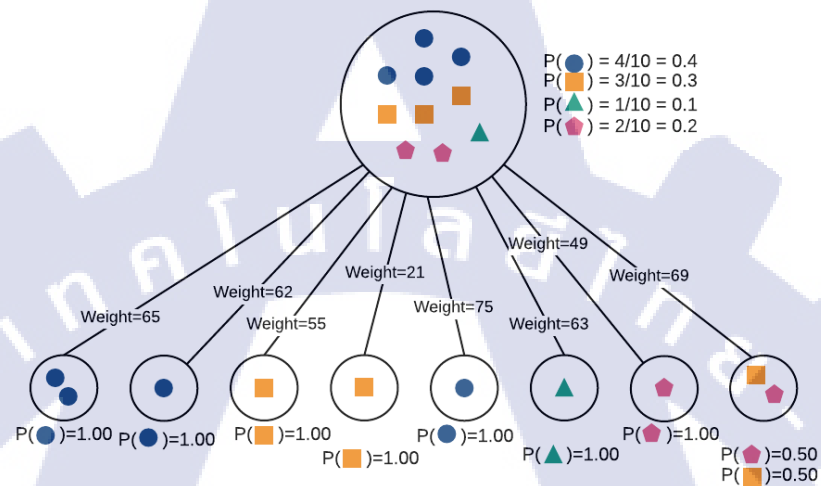
$$p (\text{Height} = 176) = 1/10 = 0.1$$

ดังนั้น

$$\begin{aligned}\text{IG (Plan, Height)} &= \text{Entropy (Plan)} - [p (\text{Height} = 170) * \\ &\text{Entropy (Height = 170)} + p (\text{Height} = 165) \\ &* \text{Entropy (Height = 165)} + p (\text{Height} = 155) * \text{Entropy (Height = 155)} + p (\text{Height} = 103) * \text{Entropy (Height = 103)} + \\ &p (\text{Height} = 166) * \text{Entropy (Height = 166)} \\ &+ p (\text{Height} = 143) * \text{Entropy (Height = 143)} + p (\text{Height} = 160) * \text{Entropy (Height = 160)} + p (\text{Height} = 157) * \text{Entropy (Height = 157)} + p (\text{Height} = 176) * \\ &\text{Entropy (Height = 176)}] \\ &= 1.84 - [0.1 * 0 + 0.2 * 1 + 0.1 * 0 + 0.1 * 0 + 0.1 * 0 + 0.1 * 0 + 0.1 * 0 + 0.1 * 0 + 0.1 * 0]\end{aligned}$$

$$+ 0.1 * 0] \\ = 1.64$$

7. คำนวณค่า IG ของ Weight มีดังต่อไปนี้



รูปที่ 3.12 การคำนวณหาค่า Information Gain ของ Weight

7.1. Entropy ของ Weight = 65

$$\begin{aligned} \text{Entropy (Weight = 65)} &= - [1 * \log_2 (1)] \\ &= 0 \end{aligned}$$

$$p (\text{Weight} = 65) = 2/10 = 0.2$$

7.2. Entropy ของ Weight = 62

$$\begin{aligned} \text{Entropy (Weight = 62)} &= - [1 * \log_2 (1)] \\ &= 0 \end{aligned}$$

$$p (\text{Weight} = 62) = 1/10 = 0.1$$

7.3. Entropy ของ Weight = 55

$$\begin{aligned} \text{Entropy (Weight = 55)} &= - [1 * \log_2 (1)] \\ &= 0 \end{aligned}$$

$$p (\text{Weight} = 55) = 1/10 = 0.1$$

7.4. Entropy ของ Weight = 21

$$\begin{aligned}\text{Entropy (Weight = 21)} &= - [1 * \log_2 (1)] \\ &= 0\end{aligned}$$

$$p (\text{Weight} = 21) = 1/10 = 0.1$$

7.5. Entropy ของ Weight = 75

$$\begin{aligned}\text{Entropy (Weight = 75)} &= - [1 * \log_2 (1)] \\ &= 0\end{aligned}$$

$$p (\text{Weight} = 75) = 1/10 = 0.1$$

7.6. Entropy ของ Weight = 63

$$\begin{aligned}\text{Entropy (Weight = 163)} &= - [1 * \log_2 (1)] \\ &= 0\end{aligned}$$

$$p (\text{Weight} = 63) = 1/10 = 0.1$$

7.7. Entropy ของ Weight = 49

$$\begin{aligned}\text{Entropy (Weight = 49)} &= - [1 * \log_2 (1)] \\ &= 0\end{aligned}$$

$$p (\text{Weight} = 49) = 1/10 = 0.1$$

7.8. Entropy ของ Weight = 69

$$\begin{aligned}\text{Entropy (Weight = 69)} &= - [0.5 * \log_2 (0.5) + 0.5 * \log_2 (0.5)] \\ &= - [0.50 * -1 + 0.50 * -1] \\ &= 1\end{aligned}$$

$$p (\text{Weight} = 69) = 2/10 = 0.2$$

ดังนั้น

$$\begin{aligned}\text{IG (Plan, Weight)} &= \text{Entropy (Plan)} - [p (\text{Weight} = 65) * \\ &\quad \text{Entropy (Weight = 65)} + p (\text{Weight} = 62) \\ &\quad * \text{Entropy (Weight = 62)} + p (\text{Weight} = 55) \\ &\quad * \text{Entropy (Weight = 55)} + p (\text{Weight} = 21)\end{aligned}$$

*Entropy (Weight = 21) + p (Weight = 75)

*Entropy (Weight = 75) + p (Weight = 63)

*Entropy (Weight = 63) + p (Weight = 49)

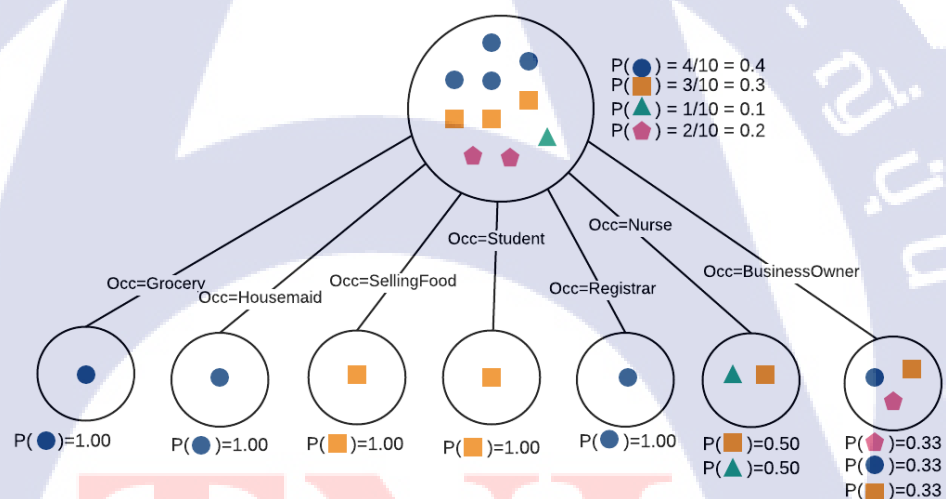
*Entropy (Weight = 49) + p (Weight = 69)

* Entropy (Weight = 69)]

$$= 1.84 - [0.2 * 0 + 0.1 * 0 + 0.1 * 0 + 0.1 * 0 + 0.1 * 0 + 0.1 * 0 + 0.1 * 0 + 0.2 * 1]$$

$$= 1.64$$

8. คำนวณค่า IG ของ Occupation มีดังต่อไปนี้



รูปที่ 3.13 การคำนวณหาค่า Information Gain ของ Occupation

8.1. Entropy ของ Occ = Grocery

$$\text{Entropy (Occ = Grocery)} = - [1 * \log_2 (1)]$$

$$= 0$$

$$p (\text{Occ = Grocery}) = 1/10 = 0.1$$

8.2. Entropy ของ Occ = Housemaid

$$\begin{aligned}\text{Entropy (Occ = Housemaid)} &= - [1 * \log_2 (1)] \\ &= 0\end{aligned}$$

$$p (\text{Occ} = \text{Housemaid}) = 1/10 = 0.1$$

8.3. Entropy ของ Occ = SellingFood

$$\begin{aligned}\text{Entropy (Occ = SellingFood)} &= - [1 * \log_2 (1)] \\ &= 0\end{aligned}$$

$$p (\text{Occ} = \text{SellingFood}) = 1/10 = 0.1$$

8.4. Entropy ของ Occ = Student

$$\begin{aligned}\text{Entropy (Occ = Student)} &= - [1 * \log_2 (1)] \\ &= 0\end{aligned}$$

$$p (\text{Occ} = \text{Student}) = 1/10 = 0.1$$

8.5. Entropy ของ Occ = Registrar

$$\begin{aligned}\text{Entropy (Occ = Registrar)} &= - [1 * \log_2 (1)] \\ &= 0\end{aligned}$$

$$p (\text{Occ} = \text{Registrar}) = 1/10 = 0.1$$

8.6. Entropy ของ Occ = Nurse

$$\begin{aligned}\text{Entropy (Occ = Nurse)} &= - [0.5 * \log_2 (0.5) + 0.5 * \log_2 (0.5)] \\ &= - [0.50 * -1 + 0.50 * -1] \\ &= 1\end{aligned}$$

$$p (\text{Occ} = \text{Nurse}) = 2/10 = 0.2$$

8.7. Entropy ของ Occ = BusinessOwner

$$\begin{aligned}\text{Entropy (Occ = BusinessOwner)} &= - [0.33 * \log_2 (0.33) + 0.33 * \log_2 \\ &\quad (0.33) + 0.33 * \log_2 (0.33)] \\ &= - [0.33 * -1.59 + 0.33 * -1.59 + 0.33 * \\ &\quad -1.59] \\ &= 1.57\end{aligned}$$

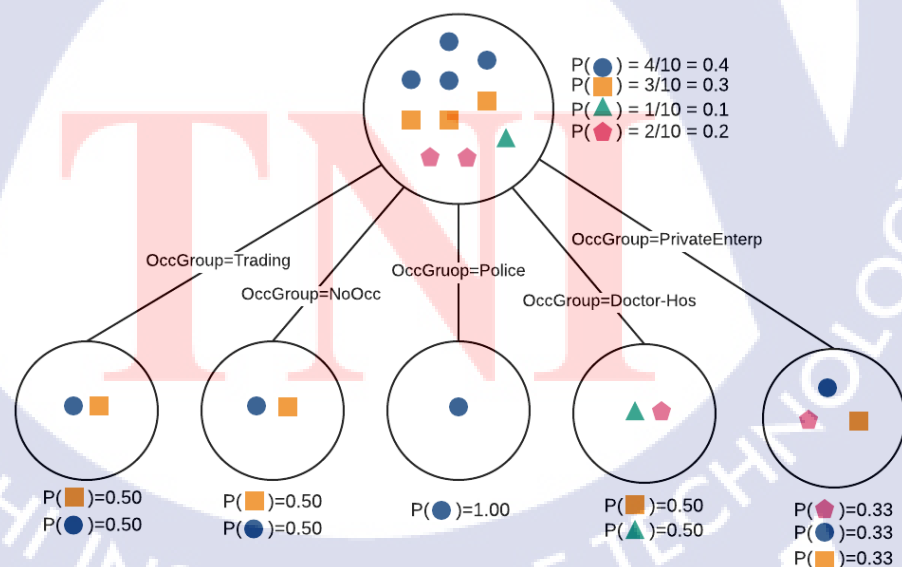
$$p (\text{Occ} = \text{BusinessOwner}) = 3/10 = 0.3$$

ดังนั้น

IG (Plan, Occ)

$$\begin{aligned}
 &= \text{Entropy (Plan)} - [p (\text{Occ} = \text{Grocery}) * \\
 &\text{Entropy (Occ} = \text{Grocery)} + p (\text{Occ} = \\
 &\text{Housemaid}) * \text{Entropy (Occ} = \text{Housemaid}) \\
 &+ p (\text{Occ} = \text{SellingFood}) * \text{Entropy (Occ} \\
 &= \text{SellingFood}) + p (\text{Occ} = \text{Student}) * \\
 &\text{Entropy (Occ} = \text{Student}) + p (\text{Occ} = \\
 &\text{Registrar}) * \text{Entropy (Occ} = \text{Registrar}) + \\
 &p (\text{Occ} = \text{Nurse}) * \text{Entropy (Occ} = \text{Nurse}) \\
 &+ p (\text{Occ} = \text{BusinessOwner}) * \\
 &\text{Entropy (Occ} = \text{BusinessOwner})] \\
 &= 1.84 - [0.1 * 0 + 0.1 * 0 + 0.1 * 0 + 0.1 * \\
 &0 + 0.1 * 0 + 0.2 * 1 + 0.3 * 1.57] \\
 &= 1.16
 \end{aligned}$$

9. คำนวณค่า IG ของ Occupation Group มีดังต่อไปนี้



รูปที่ 3.14 การคำนวณหาค่า Information Gain ของ Occupation Group

9.1. Entropy ของ OccGroup = Trading

$$\begin{aligned}
 \text{Entropy (OccGroup = Trading)} &= - [0.5 * \log_2 (0.5) + 0.5 * \log_2 (0.5)] \\
 &= - [0.50 * -1 + 0.50 * -1] \\
 &= 1
 \end{aligned}$$

$$p (\text{OccGroup} = \text{Trading}) = 2/10 = 0.2$$

9.2. Entropy ของ OccGroup = NoOcc

$$\begin{aligned}
 \text{Entropy (OccGroup = NoOcc)} &= - [0.5 * \log_2 (0.5) + 0.5 * \log_2 (0.5)] \\
 &= - [0.50 * -1 + 0.50 * -1] \\
 &= 1
 \end{aligned}$$

$$p (\text{OccGroup} = \text{NoOcc}) = 2/10 = 0.2$$

9.3. Entropy ของ OccGroup = Police

$$\begin{aligned}
 \text{Entropy (OccGroup = Police)} &= - [1 * \log_2 (1)] \\
 &= 0
 \end{aligned}$$

$$p (\text{OccGroup} = \text{Police}) = 1/10 = 0.1$$

9.4. Entropy ของ OccGroup = Doctor-Hos

$$\begin{aligned}
 \text{Entropy (OccGroup = Doctor-Hos)} &= - [0.5 * \log_2 (0.5) + 0.5 * \log_2 (0.5)] \\
 &= - [0.50 * -1 + 0.50 * -1] \\
 &= 1
 \end{aligned}$$

$$p (\text{OccGroup} = \text{Doctor-Hos}) = 2/10 = 0.2$$

9.5. Entropy ของ OccGroup = PrivateEnterp

$$\begin{aligned}
 \text{Entropy (OccGroup = PrivateEnterp)} &= - [0.33 * \log_2 (0.33) + 0.33 * \log_2 (0.33) + 0.33 * \log_2 (0.33)] \\
 &= - [0.33 * -1.59 + 0.33 * -1.59 + 0.33 * -1.59] \\
 &= 1.57
 \end{aligned}$$

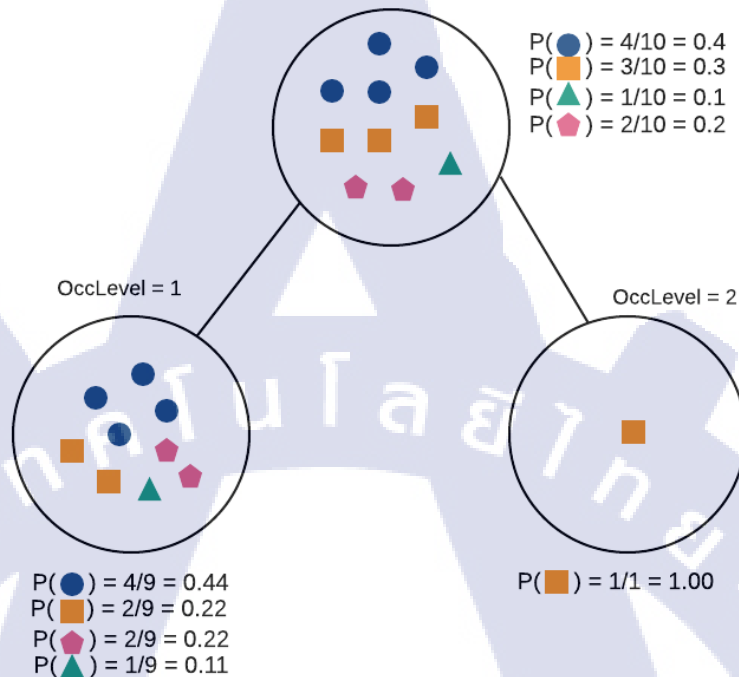
$$p(\text{OccGroup} = \text{PrivateEnterp}) = 3/10 = 0.3$$

ดังนั้น

$$\begin{aligned} \text{IG}(\text{Plan}, \text{OccGroup}) &= \text{Entropy}(\text{Plan}) - [p(\text{OccGroup} = \text{Trading}) * \text{Entropy}(\text{OccGroup} = \text{Trading}) \\ &+ p(\text{OccGroup} = \text{NoOcc}) * \text{Entropy}(\text{OccGroup} = \text{NoOcc}) + p(\text{OccGroup} = \text{Police}) * \text{Entropy}(\text{OccGroup} = \text{Police}) + \\ &p(\text{OccGroup} = \text{Doctor-Hos}) * \text{Entropy}(\text{OccGroup} = \text{Doctor-Hos}) + \\ &p(\text{OccGroup} = \text{PrivateEnterp}) * \text{Entropy}(\text{OccGroup} = \text{PrivateEnterp})] \\ &= 1.84 - [0.2 * 1 + 0.2 * 1 + 0.1 * 0 + 0.2 * 1 + 0.3 * 1.57] \\ &= 0.77 \end{aligned}$$

TNI

10. คำนวณค่า IG ของ Occupation Level มีดังต่อไปนี้



รูปที่ 3.15 การคำนวณหาค่า Information Gain ของ Occupation Level

10.1. Entropy ของ OccLevel = 1

$$\begin{aligned}
 \text{Entropy (OccLevel = 1)} &= - [0.44 * \log_2 (0.44) + 0.22 * \log_2 (0.22) \\
 &\quad + 0.22 * \log_2 (0.22) + 0.11 * \log_2 (0.11)] \\
 &= - [0.44 * -1.18 + 0.22 * -2.18 + 0.22 * - \\
 &\quad 2.18 + 0.22 * -2.18 + 0.11 * -3.18] \\
 &= 2.3078
 \end{aligned}$$

$$p(\text{OccLevel} = 1) = 9/10 = 0.9$$

10.2. Entropy ของ OccLevel = 2

$$\begin{aligned}
 \text{Entropy (OccLevel = 2)} &= - [1 * \log_2 (1)] \\
 &= 0
 \end{aligned}$$

$$p(\text{OccLevel} = 2) = 1/10 = 0.1$$

ดังนั้น

$$\begin{aligned}
 IG(\text{Plan}, \text{OccLevel}) &= \text{Entropy}(\text{Plan}) - [p(\text{OccLevel} = 1) * \\
 &\quad \text{Entropy}(\text{OccLevel} = 1) + p(\text{OccLevel} = \\
 &\quad 2) * \text{Entropy}(\text{OccLevel} = 2)] \\
 &= 1.84 - [0.9 * 2.3078 + 0.1 * 0] \\
 &= 0
 \end{aligned}$$

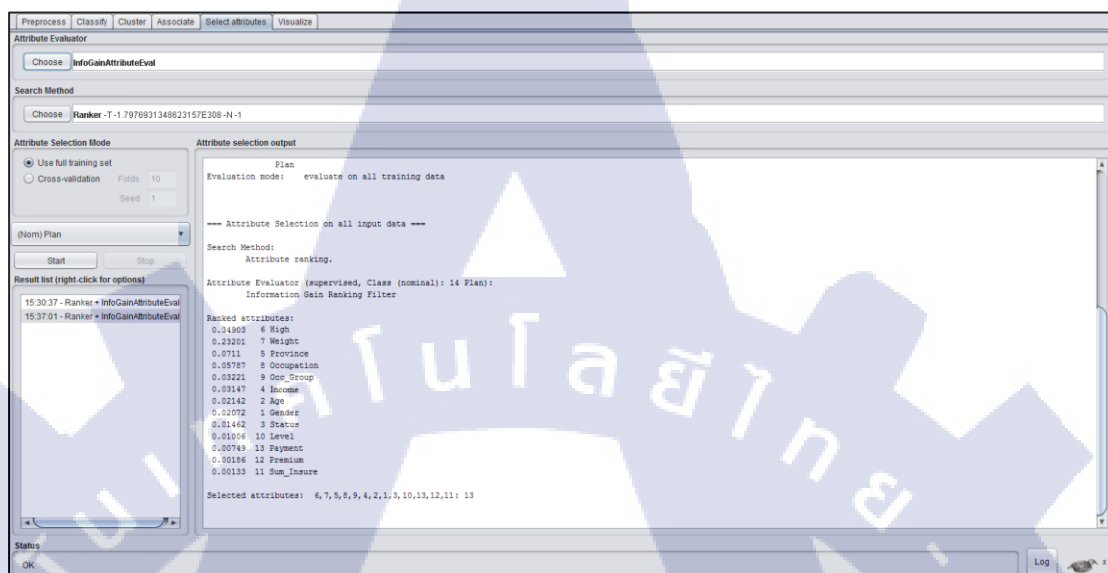
ค่า Information Gain (IG) ที่ได้จากการคำนวณของแต่ละ Attribute ต่างๆ แสดงดังตารางที่ 3.3 ด้านล่างนี้

ตารางที่ 3.3 ตารางแสดงค่า Information Gain (IG) ของแต่ละ Attribute

Attribute	Information Gain
Height	1.64
Weight	1.64
Age	1.54
Province	1.24
Occupation	1.16
Income	1.16
Occupation Group	0.77
Gender	0.50
Status	0.28
Occupation Level	0

จากตารางแสดงค่า Information Gain (IG) ที่ได้จากการคำนวณของแต่ละ Attribute ซึ่งจะเห็นได้ว่า Occupation Level มีค่า Information Gain น้อยที่สุด Status มีค่าน้อยสุดเป็นอันดับที่ 2 และ Gender เป็นอันดับที่ 3 ซึ่งจากชุดข้อมูลตัวอย่างนี้ สามารถสรุปได้ว่า Attribute ที่จะถูกคัดเลือกเพื่อนำไปวิเคราะห์ข้อมูลต่อไป ได้แก่ Height, Weight, Age, Province, Occupation และ Income เป็นต้น

จากงานวิจัยจำนวนข้อมูลที่นำมาวิเคราะห์มีจำนวนมาก ดังนั้นผู้วิจัย จึงใช้โปรแกรม Weka ในการช่วยคำนวณค่า Information Gain (IG) ซึ่งได้ผลลัพธ์ ดังรูปที่ 3.16



รูปที่ 3.16 ผลลัพธ์จากการคำนวณค่า Information Gain โดยใช้โปรแกรม Weka

จากรูปภาพที่ 3.16 แสดงผลลัพธ์การคำนวณค่า Information Gain โดยโปรแกรม Weka จะเห็นได้ว่า Attribute ที่มีค่า Information Gain มากที่สุด ได้แก่ Height ที่มีค่า IG = 0.35 และลำดับต่อมา ได้แก่ Weight ที่มีค่า IG = 0.23 และ Province ที่มีค่า IG = 0.0711 ส่วน Attribute ที่มีค่า Information Gain น้อยที่สุด คือ Sum_Insure ที่มีค่า IG = 0.00133, Premium ที่มีค่า IG = 0.00186 และ Payment = 0.00749 ตามลำดับ ซึ่ง Attribute ที่มีค่า IG น้อยนี้จะไม่ได้ถูกคัดเลือกเพื่อนำไปวิเคราะห์ข้อมูลต่อไป

3.3.4.2 ขั้นตอนที่ 2 เป็นขั้นตอนการแบ่งกลุ่มลูกค้า โดยใช้เทคนิคการจัดกลุ่ม Clustering K-means ในการแบ่งกลุ่มลูกค้าที่มีลักษณะคล้ายคลึงกันอยู่กลุ่มเดียวกัน ซึ่งจะได้ออกเป็นจำนวน K กลุ่มตามที่คุณวิจัยได้กำหนด แทนค่าแต่ละกลุ่มด้วยค่าเฉลี่ยของกลุ่ม ที่ใช้เป็นจุดศูนย์กลาง (Centroid) ของกลุ่มในการวัดระยะห่างของข้อมูลในกลุ่มเดียวกัน ซึ่งจากงานวิจัยได้กำหนดจำนวนกลุ่มออกเป็น 10 กลุ่ม แล้วนำข้อมูลเข้าสู่ระบบ Weka เพื่อทำการแบ่งกลุ่มข้อมูลต่อไป ซึ่งได้ผลลัพธ์ดังแสดงตารางที่ 3.4

ตารางที่ 3.4 ตารางผลลัพธ์การแบ่งกลุ่มข้อมูลด้วย K-means clustering โดยใช้โปรแกรม Weka

	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5	Cluster6	Cluster7	Cluster8	Cluster9	Cluster10
Gender	M	F	F	F	F	F	M	M	F	F
Age	30	65	1	26	41	29	61	38	33	39
Status	2	2	2	2	1	2	1	1	1	1
Income	150K	6M	0	150K	200K	300K	1.5M	0	276K	1M
Province	47	2	9	2	9	6	9	9	9	70
Height	180	154	79	158	157	162	176	173	149	162
Weight	85	52	47	58	50	61	69	60	60	60
Occ	1	30	32	51	66	51	164	122	66	16
OccGroup	1	25	2	8	4	8	25	46	4	8
OccLevel	1	1	1	2	1	2	1	1	1	1
Plan	Protective	Saving	Protective	Saving	Saving	Saving	UnitLink	Protective	Saving	Saving
Instance	17%	10%	29%	4%	9%	5%	3%	10%	7%	6%

ดังนั้นชุดข้อมูล (Data) ที่จะนำไปวิเคราะห์ขั้นตอนต่อไป จะมี Attribute เพิ่มอีก 1 Attribute คือ Cluster ที่ได้จากการ Classify ด้วยเทคนิค K-means Clustering

3.3.4.3 ขั้นตอนที่ 3 ขั้นตอนการสร้างแบบจำลองการแนะนำการเลือกซื้อผลิตภัณฑ์แบบประกันให้แก่ลูกค้า โดยใช้เทคนิคโครงข่ายประสาทเทียม (Neural Network) ในการสร้างแบบจำลอง ซึ่งมีกระบวนการ ดังต่อไปนี้

การคำนวณหาค่า Weight เพื่อสร้างแบบจำลองการแนะนำการเลือกซื้อผลิตภัณฑ์แบบประกันให้แก่ลูกค้า มีวิธีการคำนวณ ดังนี้

จากตัวอย่าง Case ที่กำหนด โดยมีตัวอย่างชุดข้อมูลตัวอย่างตามตารางที่ 3.5 หลังจากนั้นนำชุดตัวอย่างข้อมูลตามตารางที่ 3.5 เข้าโปรแกรม Weka เพื่อช่วยคำนวณหาค่า Wight ระหว่าง Input Node และ Hidden Node และคำนวณหาค่า Weight ระหว่าง Hidden Node และ Output Node ดังแสดงผลลัพธ์ตามตารางที่ 3.6 และ 3.7

ตารางที่ 3.5 ตารางตัวอย่างข้อมูลเพื่อสร้างแบบจำลองการแนะนำการเลือกซื้อผลิตภัณฑ์แบบประกัน

Gender	Age	Status	Income	Province	Height	Weight	Occ	OccGroup	OccLevel	Plan
2	27	1	180K	1	165	51	2	6	1	1
2	27	1	180K	1	165	51	2	6	1	1
1	54	2	1.5M	2	168	71	3	6	1	2
2	3	1	0	1	87	11	8	3	1	1
1	2	1	0	1	89	12	1	1	1	3
2	49	1	150K	3	143	63	4	4	1	3
1	32	2	200K	4	178	85	5	5	3	4
1	35	2	1.2M	5	180	70	6	2	2	2
1	40	2	3.6M	1	168	70	7	7	2	3

ตารางที่ 3.6 ตารางค่า Weight ระหว่าง Input Node และ Hidden Node

Input Node	Node 1	Node 2	Node 3	Node 4	Node 5
Gender	-1.18	-0.44	-0.52	-0.36	-0.18
Age	1.51	0.14	0.36	0.12	-0.04
Status	-1.74	0.02	-0.16	0.15	0.52
Income	-0.26	0.53	0.51	0.50	0.45
Province	0.19	-0.07	-0.11	-0.08	0.03
Height	-1.42	-0.25	-0.38	-0.11	0.40
Weight	0.98	0.30	0.38	0.18	0.05
Occupation	-0.03	-0.05	0.44	-0.01	-0.12
OccGroup	-0.22	0.51	0.55	0.41	0.36
OccLevel	1.04	0.95	1.07	0.85	0.36
Threshold	-0.38	-0.79	-0.70	-0.77	-0.58

ตารางที่ 3.7 ตารางค่า Weight ระหว่าง Hidden Node และ Output Node

Hidden Node	Output
Node 1	2.44
Node 2	0.70
Node 3	0.97

ตารางที่ 3.7 ตารางค่า Weight ระหว่าง Hidden Node และ Output Node (ต่อ)

Node 4	0.49
Node 5	-0.30
Threshold	-1.45

ขั้นตอนการทำงานของแบบจำลองเพื่อแนะนำแบบประกันชีวิตจากตัวอย่าง Case ของชุดข้อมูล ดังตารางที่ 3.8

ตารางที่ 3.8 ตารางตัวอย่าง case ของชุดข้อมูลเพื่อทำการทดสอบ

Gender	Age	Status	Income	Province	Height	Weight	Occ	OccGroup	OccLevel	Plan
2	49	1	150K	3	143	63	4	4	1	?

ขั้นตอนการทำงานของแบบจำลอง มีดังนี้

1. หาค่า Normalization Attribute (-1, 1) ของแต่ละ Attribute ดังต่อไปนี้จากสูตรด้านล่าง โดยที่ค่า max, min คือค่าที่มากที่สุดและค่าที่น้อยที่สุดของแต่ละ Attribute

$$\text{range} = (\max - \min)/2$$

$$\text{base} = (\max + \min)/2$$

$$\text{norm_attribute} = (\text{attribute} - \text{base})/\text{range}$$

1.1 ค่า Normalization Attribute ของ Gender

$$\text{range} = (2 - 1)/2 = 0.5$$

$$\text{base} = (2 + 1)/2 = 1.5$$

$$\text{norm_Gender} = (2 - 1.5)/0.5 = 1$$

1.2 ค่า Normalization Attribute ของ Age

$$\text{range} = (54 - 2)/2 = 26$$

$$\text{base} = (54 + 2)/2 = 28$$

$$\text{norm_Age} = (49 - 28)/26 = 0.81$$

1.3 ค่า Normalization Attribute ของ Gender

$$\text{range} = (2 - 1)/2 = 0.5$$

$$\text{base} = (2 + 1)/2 = 1.5$$

$$\text{norm_Gender} = (2 - 1.5)/0.5 = 1$$

1.4 ค่า Normalization Attribute ของ Status

$$\text{range} = (2 - 1)/2 = 0.5$$

$$\text{base} = (2 + 1)/2 = 1.5$$

$$\text{norm_Status} = (1 - 1.5)/0.5 = -1$$

1.5 ค่า Normalization Attribute ของ Income

$$\text{range} = (3,600,000 - 0)/2 = 1,800,000$$

$$\text{base} = (3,600,000 + 0)/2 = 1,800,000$$

$$\text{norm_Income} = (150,000 - 1,800,000)/1,800,000 = -0.17$$

1.6 ค่า Normalization Attribute ของ Province

$$\text{range} = (5 - 1)/2 = 2$$

$$\text{base} = (5 + 1)/2 = 3$$

$$\text{norm_Province} = (3 - 3)/2 = 0$$

1.7 ค่า Normalization Attribute ของ Height

$$\text{range} = (180 - 87)/2 = 46.5$$

$$\text{base} = (180 + 87)/2 = 133.5$$

$$\text{norm_Height} = (143 - 133.5)/46.5 = 0.20$$

1.8 ค่า Normalization Attribute ของ Weight

$$\text{range} = (85 - 11)/2 = 37$$

$$\text{base} = (85 + 11)/2 = 48$$

$$\text{norm_Weight} = (63 - 48)/37 = 0.41$$

1.9 ค่า Normalization Attribute ของ Occupation

$$\text{range} = (8 - 1)/2 = 3.5$$

$$\text{base} = (8 + 1)/2 = 4.5$$

$$\text{norm_Occ} = (4 - 4.5)/3.5 = -0.14$$

10. ค่า Normalization Attribute ของ Occupation Group

$$\text{range} = (7 - 1)/2 = 3$$

$$\text{base} = (7 + 1)/2 = 4$$

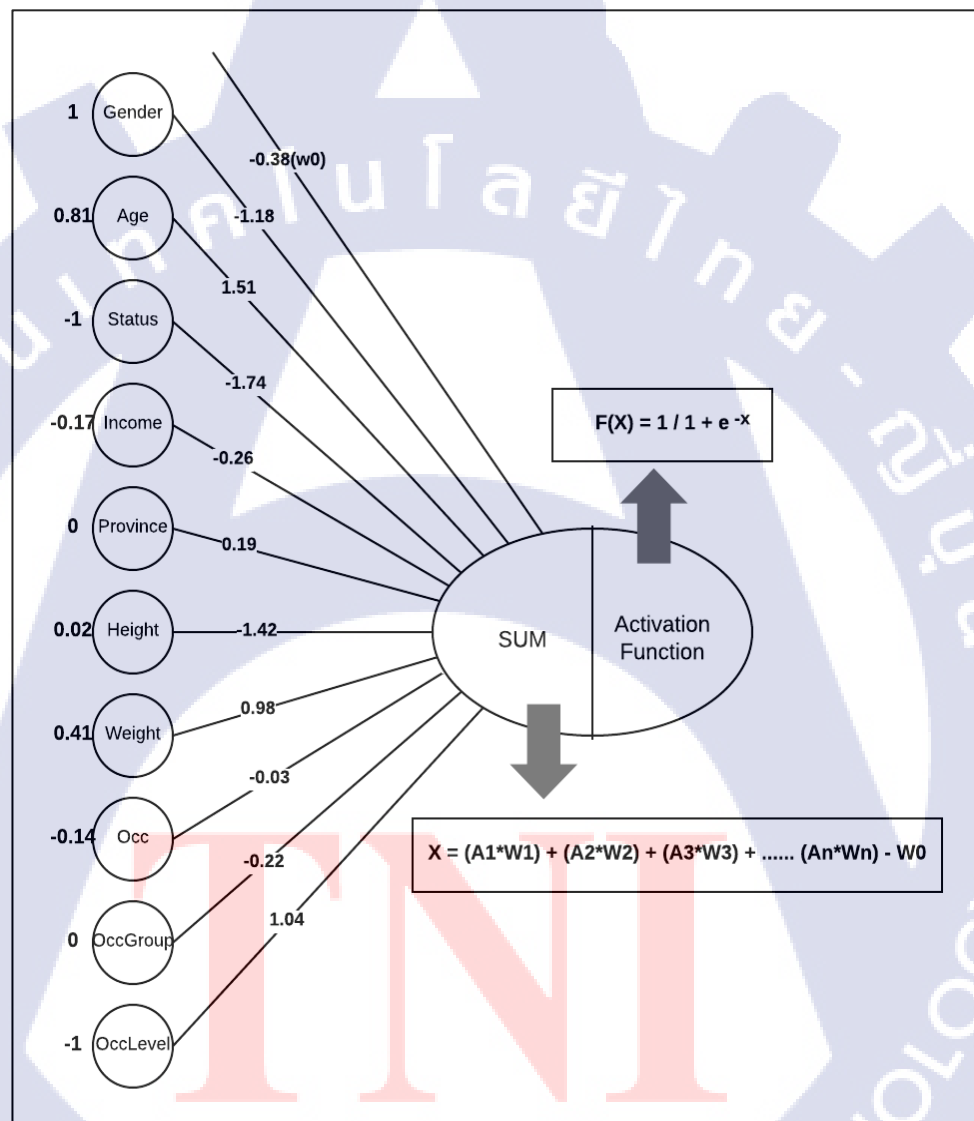
$$\text{norm_OccGroup} = (4 - 4)/3 = 0$$

11. ค่า Normalization Attribute ของ Occupation Level

$$\text{range} = (3 - 1)/2 = 1$$

$$\begin{aligned} \text{base} &= (3 + 1)/2 = 2 \\ \text{norm_OccLevel} &= (1 - 2)/1 = -1 \end{aligned}$$

2. เมื่อได้ค่า Normalization Attribute ของแต่ละ Attribute แล้ว คำนวณหาค่า Activate Function เพื่อคำนวณค่าในโหนดใน Hidden Node ตามรูปภาพที่ 3.17



รูปที่ 3.17 แสดงการคำนวณค่าในโหนดของ Hidden Node

จากตารางที่ 3.6 มีทั้งหมด 5 Hidden Node ซึ่งมีการคำนวณหาค่าในแต่ละโหนด
ดังต่อไปนี้

2.1 คำนวณหาค่าใน Hidden Node โหนดที่ 1

$$\begin{aligned}\text{Node 1} &= (1 * -1.18) + (0.81 * 1.51) + (-1 * -1.74) \\ &+ (-0.17 * -0.26) + (0 * 0.19) + (0.20 * - \\ &1.42) + (0.41 * 0.98) + \\ &(-0.14 * -0.03) + (0 * -0.22) + (-1 * 1.04) - \\ &0.38 \\ &= 0.5293 \\ F(\text{Node 1}) &= 1 / 1 + e^{-0.5293} \\ &= 0.71\end{aligned}$$

2.2 คำนวณหาค่าใน Hidden Node โหนดที่ 2

$$\begin{aligned}\text{Node 2} &= (1 * -0.44) + (0.81 * 0.14) + (-1 * -0.02) \\ &+ (-0.17 * 0.53) + (0 * -0.07) + (0.20 * - \\ &0.25) + (0.41 * 0.30) + (-0.14 * -0.05) + (0 \\ &* 0.51) + (-1 * 0.95) - 0.79 \\ &= -2.2063 \\ F(\text{Node 2}) &= 1 / 1 + e^{-(-2.2063)} \\ &= 0.1\end{aligned}$$

2.3 คำนวณหาค่าใน Hidden Node โหนดที่ 3

$$\begin{aligned}\text{Node 3} &= (1 * -0.52) + (0.81 * 0.36) + (-1 * -0.16) \\ &+ (-0.17 * 0.51) + (0 * -0.11) + (0.20 * - \\ &0.38) + (0.41 * 0.38) + (-0.14 * -0.04) + (0 \\ &* 0.55) + (-1 * 1.07) - 0.70 \\ &= -1.8509 \\ F(\text{Node 3}) &= 1 / 1 + e^{-(-1.8509)} \\ &= 0.13\end{aligned}$$

2.4 คำนวณหาค่าใน Hidden Node โหนดที่ 4

$$\begin{aligned}\text{Node 4} &= (1 * -0.36) + (0.81 * 0.12) + (-1 * -0.15) \\ &+ (-0.17 * 0.50) + (0 * -0.08) + (0.20 * - \\ &0.11) + (0.41 * 0.18) + (-0.14 * -0.01) + (0\end{aligned}$$

$$* 0.41) + (-1 * 0.85) - 0.77$$

$$= -2.0646$$

F (Node 4)

$$= 1 / 1 + e^{(-2.0646)}$$

$$= 0.11$$

2.5 คำนวณหาค่าใน Hidden Node โหนดที่ 5

Node 5

$$= (1 * -0.18) + (0.81 * -0.04) + (-1 * -0.52)$$

$$+ (-0.17 * 0.45) + (0 * 0.03) + (0.20 * 0.40)$$

$$+ (0.41 * 0.05) + (-0.14 * -0.12) + (0 *$$

$$0.36) + (-1 * 0.36) - 0.58$$

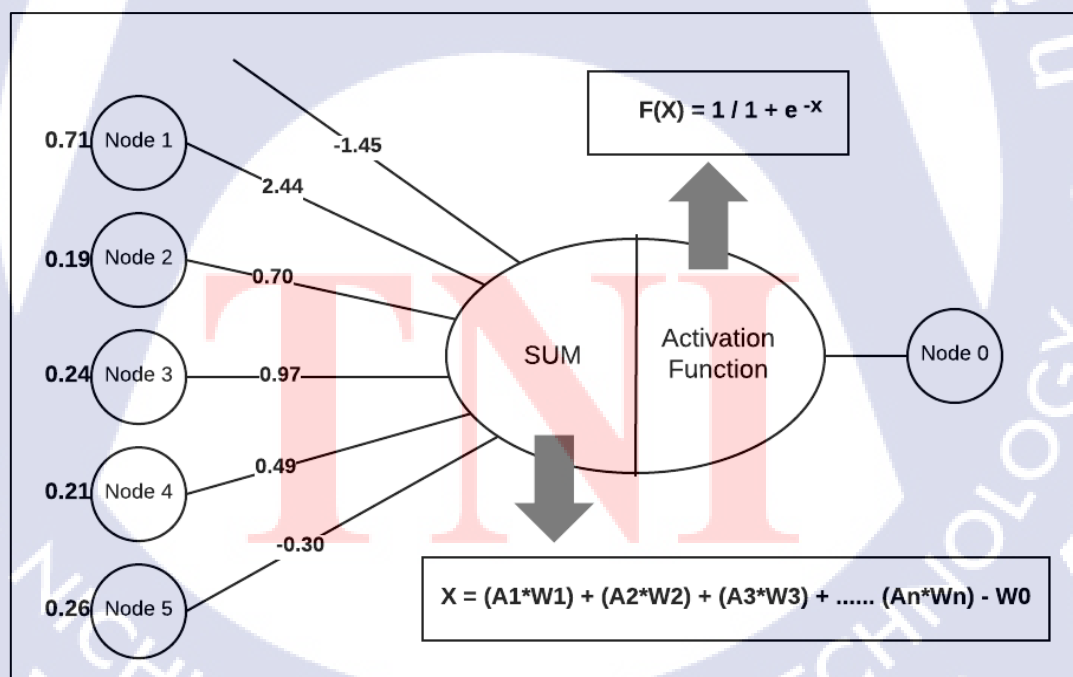
$$= -1.6316$$

F (Node 5)

$$= 1 / 1 + e^{(-1.6316)}$$

$$= 0.16$$

3. หลังจากได้ค่าโหนดใน Hidden Node แล้ว คำนวณหาค่าในโหนด Output Node ดังต่อไปนี้



รูปที่ 3.18 แสดงการคำนวณค่าในโหนดของ Output Node

คำนวณหาค่าใน Output Node ดังนี้

$$\begin{aligned}
 \text{Node 0} &= (0.71 * 2.44) + (0.19 * 0.70) + (0.24 * 0.97) + (0.21 * 0.49) + (0.26 * -0.03) - 1.45 \\
 &= 0.6731 \\
 F(\text{Node 0}) &= 1 / 1 + e^{-0.6731} \\
 &= 0.66
 \end{aligned}$$

4. หลังจากได้ค่าคำตอบแล้ว แปลงค่ากลับให้อยู่ในช่วงของข้อมูลใน Attribute

$$\begin{aligned}
 \text{range} &= (4 - 1)/2 = 1.5 \\
 \text{base} &= (4+1)/2 = 2.5 \\
 \text{Plan} &= (0.66 * 1.5) + 2.5 \\
 &= 3.49
 \end{aligned}$$

ซึ่งค่าคำตอบที่ถูกต้องจากชุดตัวข้อมูลสำหรับ case นี้ มีค่าเท่ากับ 3

ดังนั้น

$$\text{Error rate} \sim 0.49$$

3.4 ขั้นที่ 4 การพัฒนาต้นแบบการแนะนำการเลือกซื้อผลิตภัณฑ์แบบประกันชีวิต

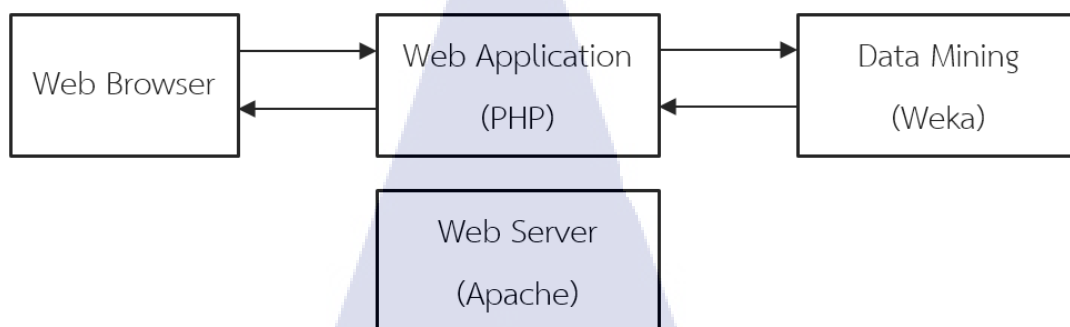
ในการพัฒนาต้นแบบ Web Application เพื่อการแนะนำการเลือกซื้อผลิตภัณฑ์ประกันชีวิต ผู้วิจัยได้ดำเนินการในการวิเคราะห์และพัฒนาระบบ ดังต่อไปนี้

3.4.1 วิเคราะห์และศึกษาความเป็นไปได้

ผู้วิจัยได้ศึกษาวิเคราะห์ถึงรูปแบบของการแนะนำการเลือกซื้อผลิตภัณฑ์ประกันชีวิต รวมถึงการศึกษความเป็นไปได้ในการพัฒนาต้นแบบและการทำงานของระบบ ซึ่งผู้วิจัยดำเนินการพัฒนาต้นแบบในรูปแบบของ Web Application ที่สามารถรองรับการใช้งานผ่าน Web Browser IE และ Google Chrome ที่สามารถใช้งานผ่านการเชื่อมต่ออินเทอร์เน็ตได้

3.4.2 การพัฒนาระบบ

ในการพัฒนาต้นแบบ Web Application เพื่อการแนะนำการเลือกซื้อผลิตภัณฑ์ประกันชีวิต ผู้วิจัยใช้โปรแกรมภาษา PHP ในการเขียนโปรแกรมพัฒนา Web Application เพื่อเชื่อมต่อกับโปรแกรม Weka มีโครงสร้างในการพัฒนาการเชื่อมต่อ ดังรูปต่อไปนี้

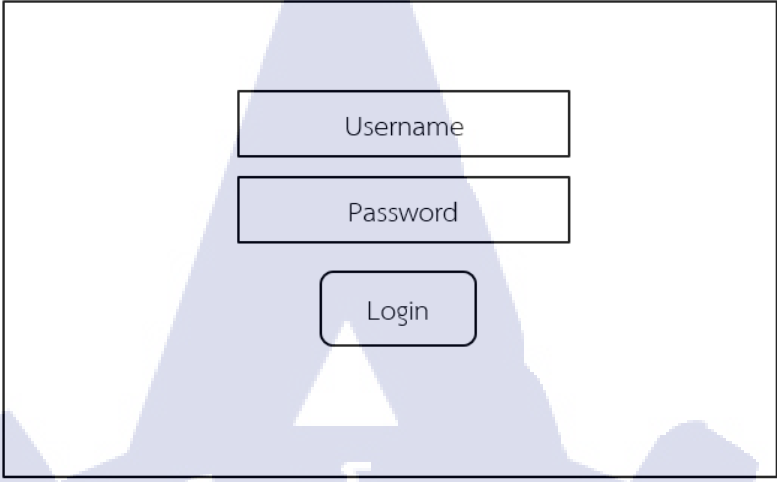


รูปที่ 3.19 Weka and PHP Integration Overview Structure

องค์ประกอบหลักในการพัฒนาการเชื่อมต่อระหว่าง Web Application กับโปรแกรม Weka คือ (1) Web server เช่น Apache (2) โปรแกรมภาษา PHP (3) Library สำหรับใช้ในการวิเคราะห์ข้อมูลด้วยเทคนิค Data Mining ในส่วนของการพัฒนาผู้วิจัยได้ทำการสร้างโพลเดอร์ weka ภายใต้โพลเดอร์ app server เพื่อใช้ในการเก็บ data และไฟล์ classification model ที่ได้จากการวิเคราะห์ข้อมูลด้วยเทคนิค Data mining และทำการเขียนคำสั่งด้วยภาษา PHP เพื่อเรียกใช้งานโปรแกรม Weka เพื่อแสดงผลลัพธ์ผ่านทางหน้าจอ

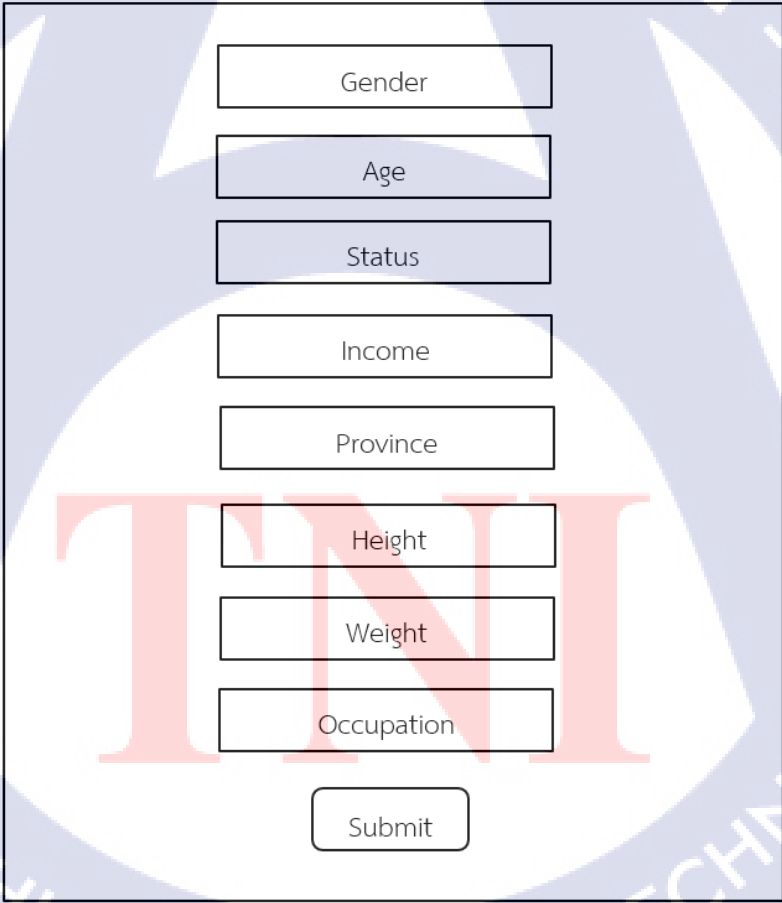
3.4.3 การออกแบบหน้าจอระบบการใช้งาน

ในการพัฒนาต้นแบบ Web Application เพื่อการแนะนำการเลือกซื้อผลิตภัณฑ์ประกันชีวิต ผู้วิจัยได้มีการออกแบบหน้าจอระบบสำหรับในส่วนของผู้ใช้งาน ให้มีการใช้งานและเข้าใจง่าย (User friendly) และเพื่อให้เป็นหน้าจอในการรับค่าข้อมูล รวมถึงเพื่อใช้หน้าจอนี้ในการแสดงผลลัพธ์ต่อผู้ใช้งาน โดยการออกแบบหน้าจอสำหรับการติดต่อกับผู้ใช้งาน (Client) เพื่อที่จะนำไปพัฒนาเป็นเว็บไซต์ มีการออกแบบหน้าจอ ดังต่อไปนี้



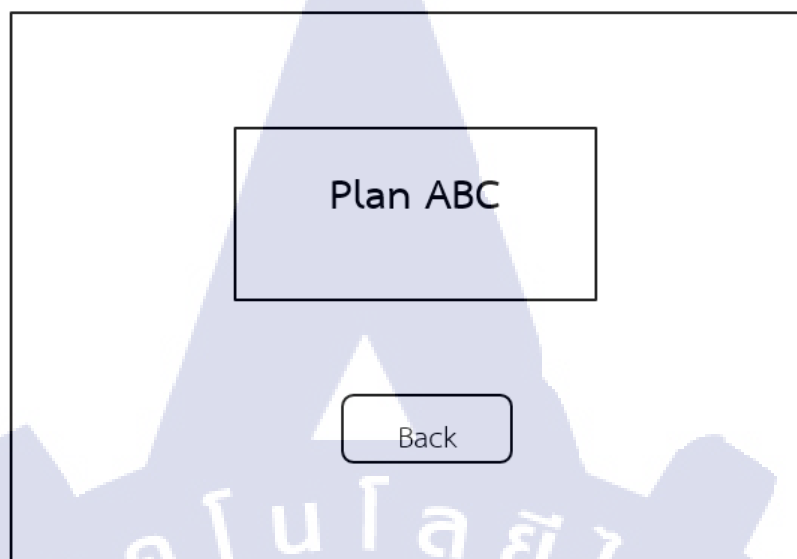
A diagram of a login form. It consists of three rectangular input fields stacked vertically, each with a light blue background and a thin black border. The first field is labeled "Username", the second "Password", and the third "Login". The "Login" field is a button with rounded corners. The entire form is enclosed in a black rectangular border.

รูปที่ 3.20 ออกแบบหน้าจอ Login



A diagram of a registration form. It consists of eight rectangular input fields stacked vertically, each with a light blue background and a thin black border. The fields are labeled "Gender", "Age", "Status", "Income", "Province", "Height", "Weight", and "Occupation". Below these fields is a "Submit" button with rounded corners. The entire form is enclosed in a black rectangular border.

รูปที่ 3.21 ออกแบบหน้าจอกรอกข้อมูลลูกค้า



รูปภาพที่ 3.22 ออกแบบหน้าจอแสดงผลการแนะนำแบบประกัน

3.5 ขั้นที่ 5 ทดสอบระบบ

สำหรับการทดสอบ เมื่อได้พัฒนาต้นแบบ Web Application เพื่อแนะนำการเลือกซื้อผลิตภัณฑ์แบบประกัน ด้วยการทำเหมืองข้อมูลโดยใช้เทคนิคโครงข่ายประสาทเทียม (Neural Network) และต้นไม้ตัดสินใจ (Decision Tree) เสร็จสมบูรณ์แล้ว ผู้ใช้งานสามารถทดลองเข้าใช้งาน Web Application เพื่อช่วยในการแนะนำและนำเสนอแบบประกันที่เหมาะสมให้แก่ลูกค้า ซึ่งในการพัฒนาและทดสอบเพื่อวัดประสิทธิภาพของรูปแบบการแนะนำจะให้ผลลัพธ์ คือ เปอร์เซ็นต์ความถูกต้องและความแม่นยำของรูปแบบการแนะนำการเลือกซื้อผลิตภัณฑ์แบบประกันชีวิตจากการทำเหมืองข้อมูล ซึ่งผู้วิจัยได้เลือกใช้วิธีการทดสอบ ดังนี้

1. การทดสอบโดยการแบ่งชุดข้อมูล 20:80 (Evaluation on training set) โดยใช้ชุดข้อมูลจำนวนร้อยละ 80 สำหรับ Training และร้อยละ 20 สำหรับการทดสอบ
2. การคำนวณ Precision & Recall เพื่อการวัดผลและประเมินประสิทธิภาพของรูปแบบ ซึ่งการคำนวณ Precision & Recall จะเป็นตัววัดประสิทธิภาพของโมเดลจากตาราง Confusion Matrix ที่ 3.8

ตารางที่ 3.9 Confusion Matrix สำหรับคำนวณ Precision & Recall

Predicted \ True	Yes	No
Yes	TP	FP
No	FN	TN
Recall		

โดยที่

True Positive (TP) คือ จำนวนที่แนะนำตรงกับข้อมูลจริงใน class ที่กำลังพิจารณา

True Negative (TN) คือ จำนวนที่แนะนำตรงกับข้อมูลจริงใน class ที่ไม่ได้กำลังพิจารณา

False Positive (FP) คือ จำนวนที่แนะนำผิดเป็น class ที่กำลังพิจารณา

False Negative (FN) คือ จำนวนที่แนะนำผิดเป็น class ที่ไม่ได้กำลังพิจารณา

ซึ่งสูตรการคำนวณ Precision และ Recall มีดังต่อไปนี้

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

ตัวอย่างการคำนวณหาค่า Precision & Recall จากรูปแบบการแนะนำการเลือกซื้อผลิตภัณฑ์แบบประกันชีวิต จากรูปที่ 3.23 ผลการแนะนำการเลือกซื้อผลิตภัณฑ์แบบประกันชีวิต จำนวนทั้งหมด 13 case โดยใช้โปรแกรม Weka

Result3.arff															
Relation: Dataset_clustered_predicted															
No.	1: Instance_number	2: Gender	3: Age	4: Status	5: Income	6: Province	7: Height	8: Weight	9: Occupation	10: OccGroup	11: OccLevel	12: prediction margin	13: predicted Plan	14: Plan	15: Cluster
	Numeric	Nominal	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Nominal	Nominal	Nominal
1	0.0	M	53.0	1.0	360000.0	1.0	170.0	65.0	1.0	1.0	1.0	0.999532	Protective	Protective	cluster0
2	1.0	M	60.0	2.0	0.0	2.0	165.0	62.0	2.0	2.0	1.0	0.994181	Protective	Protective	cluster0
3	2.0	F	57.0	1.0	0.0	3.0	152.0	65.0	1.0	1.0	1.0	0.999999	Protective	Protective	cluster2
4	3.0	F	67.0	3.0	0.0	3.0	152.0	69.0	3.0	2.0	1.0	0.996961	Protective	Protective	cluster2
5	93.0	F	21.0	1.0	0.0	3.0	150.0	48.0	29.0	8.0	2.0	0.852472	Saving	Saving	cluster5
6	94.0	F	21.0	1.0	0.0	3.0	150.0	48.0	29.0	8.0	2.0	0.852419	Saving	Saving	cluster5
7	95.0	F	21.0	1.0	0.0	3.0	150.0	48.0	29.0	8.0	2.0	0.852365	Saving	Saving	cluster5
8	344.0	F	1.0	2.0	0.0	9.0	88.0	10.0	31.0	2.0	1.0	-0.999548	Saving	Annuity	cluster8
9	514.0	M	36.0	2.0	0.0	9.0	180.0	80.0	44.0	19.0	2.0	-0.748787	Saving	Annuity	cluster7
10	607.0	M	16.0	2.0	0.0	2.0	171.0	64.0	32.0	2.0	1.0	-0.676504	Saving	Annuity	cluster0
11	2475.0	F	32.0	2.0	360000.0	27.0	157.0	49.0	9.0	6.0	1.0	-0.998189	Saving	UnitLink	cluster1
12	2544.0	M	59.0	1.0	1000000.0	1.0	165.0	69.0	16.0	8.0	1.0	-0.554967	Annuity	UnitLink	cluster6
13	2748.0	F	48.0	2.0	800000.0	69.0	158.0	60.0	16.0	8.0	1.0	-0.309879	Saving	UnitLink	cluster9

รูปที่ 3.23 ผลการแนะนำการเลือกซื้อแบบประกันชีวิตโดยใช้โปรแกรม Weka

การคำนวณ Precision & Recall จากผลการแนะนำ มีรายละเอียดดังต่อไปนี้

1. Precision & Recall สำหรับการพิจารณา Protective

$$\text{Precision} = \frac{4}{4} \times 100 = 100 \%$$

$$\text{Recall} = \frac{4}{4} \times 100 = 100\%$$

2. Precision & Recall สำหรับการพิจารณา Saving

$$\text{Precision} = \frac{3}{8} \times 100 = 37.5 \%$$

$$\text{Recall} = \frac{3}{8} \times 100 = 37.5 \%$$

3. Precision & Recall สำหรับการพิจารณา Annuity

$$\text{Precision} = \frac{0}{4} \times 100 = 0 \%$$

$$\text{Recall} = \frac{0}{4} \times 100 = 0 \%$$

4. Precision & Recall สำหรับการพิจารณา Unit Link

$$\text{Precision} = \frac{0}{3} \times 100 = 0 \%$$

$$\text{Recall} = \frac{0}{3} \times 100 = 0 \%$$

ผลการคำนวณ Precision & Recall สำหรับการแนะนำผลิตภัณฑ์ประกันที่เป็น Protective มีความถูกต้องและแม่นยำในการแนะนำเทียบกับ class ที่กำลังพิจารณาอยู่ที่ 100% การแนะนำผลิตภัณฑ์ประกันภัยที่เป็น Saving มีความถูกต้องและแม่นยำอยู่ที่ 37.5% และการแนะนำผลิตภัณฑ์ประกันภัยที่เป็น Annuity และ Unit Link มีความถูกต้องและแม่นยำอยู่ที่ 0%

3.6 ขั้นที่ 6 การสรุปผลการวิจัย

สำหรับในส่วนของการสรุปผลการวิจัย ผลที่ได้จากการวิเคราะห์และพัฒนาต้นแบบ Web Application เพื่อแนะนำการเลือกซื้อผลิตภัณฑ์แบบประกัน โดยใช้เทคนิคโครงข่ายประสาทเทียม และต้นไม้ตัดสินใจ รวมถึงความเหมาะสมและความเป็นไปได้ของกรอบแนวคิดของงานวิจัยนี้จาก ความเห็นของผู้เชี่ยวชาญ โดยพิจารณาจากค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน ซึ่งมีเกณฑ์การ ประเมินตามตารางที่ 3.10 ดังนี้

ตารางที่ 3.10 เกณฑ์การวัดประเมินความเหมาะสมของกรอบแนวคิดจากผู้เชี่ยวชาญ

ค่าเฉลี่ย	ความหมาย
4.50 – 5.00	เหมาะสมมากที่สุด
2.50 – 3.49	เหมาะสมมาก
2.50 – 3.49	เหมาะสมปานกลาง
1.50 – 2.49	เหมาะสมน้อย
1.00 – 1.49	เหมาะสมน้อยที่สุด

ในการหาความเหมาะสมของกรอบแนวคิดงานวิจัย พิจารณาค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานจาก สูตรการคำนวณหาค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน ดังนี้

การหาค่าเฉลี่ย $\bar{X} = \frac{\sum X}{N}$ เมื่อ $\bar{X}$ คือ ผลรวมคะแนนทั้งหมดในกลุ่ม,
 $\sum X$ คือ ผลรวมคะแนนทั้งหมดในกลุ่ม และ N คือ จำนวนผู้เชี่ยวชาญ

การหาส่วนเบี่ยงเบนมาตรฐาน

$$S.D. = \sqrt{\frac{N \sum X^2 - (\sum X)^2}{N(N-1)}} \quad (3.2)$$

เมื่อ

S.D. คือ ส่วนเบี่ยงเบนมาตรฐาน

$\sum X$ คือ ผลรวมของคะแนนในกลุ่ม

$\sum X^2$ คือ ผลรวมของคะแนนแต่ละตัวยกกำลังสอง

N คือ จำนวนผู้ผู้เชี่ยวชาญ

บทที่ 4

ผลของการวิจัย

ผลการวิเคราะห์งานวิจัยเรื่อง การพัฒนาต้นแบบ Web Application เพื่อแนะนำการเลือกซื้อผลิตภัณฑ์ประกันชีวิต โดยใช้เทคนิคโครงข่ายประสาทเทียม และต้นไม้ตัดสินใจโดยเปรียบเทียบกับวัตถุประสงค์ของงานวิจัย ดังนี้

1. ผลการสังเคราะห์และประเมินแบบจำลองการแนะนำการเลือกซื้อผลิตภัณฑ์ประกันชีวิต โดยใช้เทคนิคโครงข่ายประสาทเทียม และต้นไม้ตัดสินใจ
2. ผลการวิเคราะห์รูปแบบการแนะนำการเลือกซื้อผลิตภัณฑ์ประกันชีวิต โดยใช้เทคนิคโครงข่ายประสาทเทียม และต้นไม้ตัดสินใจ
3. การพัฒนาต้นแบบ Web Application เพื่อแนะนำการเลือกซื้อผลิตภัณฑ์ประกันชีวิต โดยใช้เทคนิคโครงข่ายประสาทเทียม และต้นไม้ตัดสินใจ
4. ผลการประเมินส่วนติดต่อผู้ใช้งาน

4.1 ผลการสังเคราะห์และประเมินแบบจำลองการแนะนำการเลือกซื้อผลิตภัณฑ์ประกันชีวิต โดยใช้เทคนิคโครงข่ายประสาทเทียม และต้นไม้ตัดสินใจ (วัตถุประสงค์ข้อ 1)

จากการนำเสนอแนวคิดการพัฒนาต้นแบบ Web Application เพื่อแนะนำการเลือกซื้อผลิตภัณฑ์ประกันชีวิต กรณีศึกษาบริษัทประกันชีวิตแห่งหนึ่ง โดยใช้เทคนิคโครงข่ายประสาทเทียม และต้นไม้ตัดสินใจ จากแบบจำลองการแนะนำการเลือกซื้อผลิตภัณฑ์แบบประกันที่เหมาะสมให้กับลูกค้า งานวิจัยนี้ได้มีการประเมินผลกรอบแนวคิดการวิจัย โดยการวัดผลการประเมินจากผู้เชี่ยวชาญ 5 ท่าน มีเกณฑ์ในการวัดประเมินผล ดังนี้

- ค่าเฉลี่ย 4.50 – 5.00 มีความหมายว่า เหมาะสมมากที่สุด
- ค่าเฉลี่ย 3.50 – 4.49 มีความหมายว่า เหมาะสมมาก
- ค่าเฉลี่ย 2.50 – 3.49 มีความหมายว่า เหมาะสมปานกลาง
- ค่าเฉลี่ย 1.50 – 2.49 มีความหมายว่า เหมาะสมน้อย
- ค่าเฉลี่ย 1.00 – 1.49 มีความหมายว่า เหมาะสมน้อยที่สุด

ผลประเมินที่ได้รับจากกลุ่มตัวอย่างผู้เชี่ยวชาญ ในการวิเคราะห์เปรียบเทียบค่าเฉลี่ย เพื่อวัดระดับความเหมาะสมของกรอบแนวคิดและแบบจำลองการพัฒนาต้นแบบ Web Application

เพื่อแนะนำการเลือกซื้อผลิตภัณฑ์ประกันชีวิต กรณีศึกษา บริษัทประกันชีวิตแห่งหนึ่ง โดยใช้เทคนิคโครงข่ายประสาทเทียมและต้นไม้ตัดสินใจ มีรายละเอียดจากผู้เชี่ยวชาญตามตารางที่ 4.1 ดังต่อไปนี้

ตารางที่ 4.1 ผลการประเมินความเหมาะสมของกรอบแนวคิดจากผู้เชี่ยวชาญ

หัวข้อการประเมิน	ค่าเฉลี่ย $\bar{X}$	S.D.	ระดับ
1. ความเหมาะสมของ Data Preparation Module	4.40	0.55	สูงมาก
2. ความเหมาะสมของ Data Transformation	3.60	0.55	สูง
3. ความเหมาะสมของ Data Mining Technique	4.20	0.45	สูงมาก
4. ความเหมาะสมของ Testing & Evaluation	4.40	0.55	สูงมาก
5. ความเหมาะสมของภาพรวมของ Model นี้	4.60	0.55	สูงมาก
Total	4.24	0.04	สูงมาก

จากตารางที่ 4.1 แสดงให้เห็นว่า ผลการประเมินความเหมาะสมของกรอบแนวคิดและแบบจำลองการพัฒนาต้นแบบ Web Application เพื่อการแนะนำการเลือกซื้อผลิตภัณฑ์ประกันชีวิต กรณีศึกษา บริษัทประกันชีวิตแห่งหนึ่ง โดยใช้เทคนิคโครงข่ายประสาทเทียมและต้นไม้ตัดสินใจ โดยผลการประเมินโดยรวมจากผู้เชี่ยวชาญมีความเหมาะสมอยู่ในระดับสูงมาก ซึ่งมีค่าเฉลี่ย $\bar{X}$ ที่ 4.24

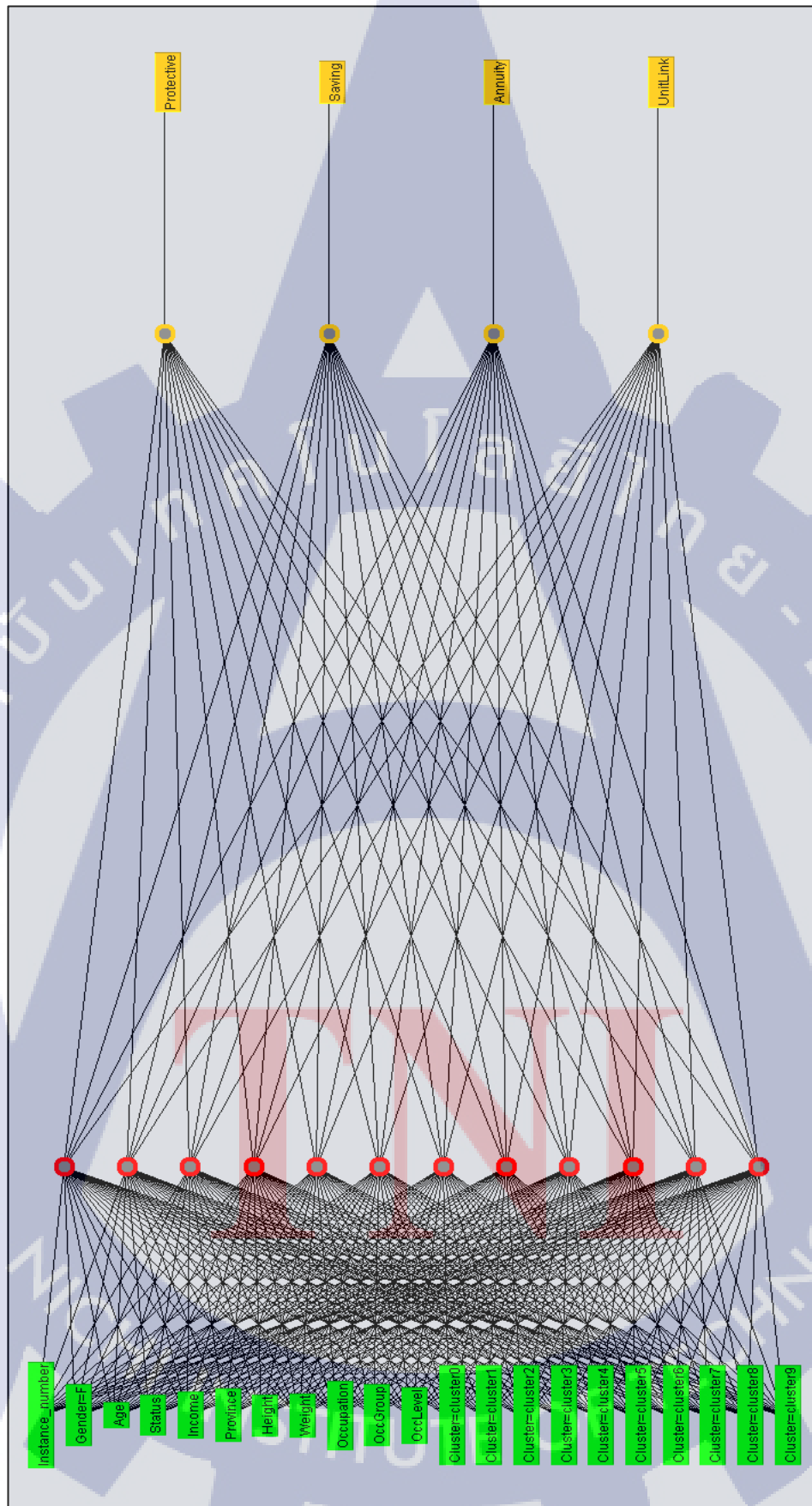
4.2 ผลการวิเคราะห์รูปแบบการแนะนำการเลือกซื้อผลิตภัณฑ์ประกันชีวิต โดยใช้เทคนิคโครงข่ายประสาทเทียม และต้นไม้ตัดสินใจ (วัตถุประสงค์ข้อ 2)

ผลการวิเคราะห์แบบจำลองการแนะนำการเลือกซื้อผลิตภัณฑ์ประกันชีวิต บริษัทประกันชีวิตแห่งหนึ่ง โดยใช้เทคนิคโครงข่ายประสาทเทียมและต้นไม้ตัดสินใจ ซึ่งงานวิจัยนี้ใช้วิธีการทดสอบโดยการแบ่งชุดข้อมูล (Evaluation on training set) โดยใช้ชุดข้อมูลจำนวนร้อยละ 80 สำหรับ Training และร้อยละ 20 สำหรับการทดสอบ เพื่อทดสอบและประเมินประสิทธิภาพของรูปแบบ โดยมีการเปรียบเทียบเปอร์เซ็นต์ความถูกต้องและความแม่นยำของการวิเคราะห์ข้อมูล (Correctly Classified Instances) จากจำนวนข้อมูลทั้งหมด 3,290 Instances จากการเลือกใช้อัลกอริทึม 3 วิธี แบ่งข้อมูลตามตารางที่ 4.2 ดังต่อไปนี้

ตารางที่ 4.2 เปรียบเทียบค่าความถูกต้องแม่นยำของการเลือกใช้อัลกอริทึม สำหรับการแนะนำการเลือกซื้อผลิตภัณฑ์แบบประกัน

ลำดับ	อัลกอริทึม	เปอร์เซ็นต์ความถูกต้องแม่นยำ (%)
1	Decision Tree & Neural Network	91.73%
2	Neural Network	63.80%
3	Decision Tree	53.98%

จากตารางที่ 4.2 แสดงผลเปรียบเทียบความถูกต้องแม่นยำของอัลกอริทึม แสดงให้เห็นว่าการเลือกใช้อัลกอริทึมโครงข่ายประสาทเทียม ร่วมกับอัลกอริทึมตัดไม้ตัดสินใจ ให้เปอร์เซ็นต์ความถูกต้องแม่นยำมากที่สุดที่ 91.73% ในขณะที่การเลือกใช้อัลกอริทึมเดียวมีเปอร์เซ็นต์ความถูกต้องแม่นยำน้อยกว่าที่ 63.80% และ 53.98% ตามลำดับ จึงเลือกใช้อัลกอริทึมแบบผสมในการวิเคราะห์รูปแบบการแนะนำการเลือกซื้อผลิตภัณฑ์แบบประกัน โดยนำข้อมูลเข้าโปรแกรม Weka เพื่อช่วยในการวิเคราะห์หารูปแบบโมเดล ผลลัพธ์รูปแบบโมเดลที่วิเคราะห์ได้ อยู่ในลักษณะของ Node ดังแสดงในรูปที่ 4.1



รูปที่ 4.1 รูปแบบโมเดลที่ได้จากการวิเคราะห์ด้วย Neural Network

4.3 การพัฒนาต้นแบบ Web Application เพื่อแนะนำการเลือกซื้อผลิตภัณฑ์ประกันชีวิต โดยใช้เทคนิคโครงข่ายประสาทเทียม และต้นไม้ตัดสินใจ (วัตถุประสงค์ข้อ 3)

ส่วนของการติดต่อผู้ใช้งาน ผู้ดำเนินงานวิจัยได้พัฒนาต้นแบบ Web Application เพื่อแนะนำการเลือกซื้อผลิตภัณฑ์ประกันชีวิต ที่สามารถใช้งานได้บน Web Browser โดยมีหน้าจอ (User Interface) สำหรับการใช้งาน ดังแสดงในรูปที่ 4.2 – 4.5



รูปที่ 4.2 หน้าจอ Login เพื่อเข้าใช้งานระบบ

ผู้ใช้งานซึ่งเป็นตัวแทนขายประกันชีวิต ที่มีรหัสตัวแทนหรือมีเลขที่ใบอนุญาต จึงจะสามารถ Login ขอเข้าใช้งานระบบได้ โดยกรอกข้อมูลลงในช่อง Username และ Password แล้วกดปุ่ม Login เพื่อเข้าสู่ระบบ

TNB INSURANCE

เพศ	อายุ	สถานภาพ	รายได้ต่อปี
หญิง ▾	26	โสด ▾	360,000
อาชีพ	จังหวัด	น้ำหนักร	ส่วนสูง
พนักงานบริษัท, พนักงาน, 1 ▾	กรุงเทพ ▾	46	165

ทำนายแบบประกัน →

รูปที่ 4.3 หน้าจอรกรอกข้อมูลลูกค้า

ผู้ใช้งานที่เป็นตัวแทนขายประกันชีวิต กรอกข้อมูลลูกค้าเบื้องต้น เพื่อแนะนำและนำเสนอผลิตภัณฑ์แบบประกันที่เหมาะสมกับลูกค้า โดยมีการกรอกข้อมูลลงในช่องกรอกดังนี้

- เลือก Drop down list เพศ
- กรอกอายุ (ตัวเลข)
- เลือก Drop down list สถานภาพ
- กรอกรายได้ต่อปี (ตัวเลข)
- เลือก Drop down list อาชีพ (ที่มีข้อมูลกลุ่มอาชีพและชั้นอาชีพ)
- เลือก Drop down list จังหวัดที่อยู่
- กรอกข้อมูลส่วนสูงน้ำหนักปัจจุบัน (ตัวเลข)

ต้องกรอกข้อมูลที่จำเป็นนี้ให้ครบถ้วน (Required Field) จึงจะสามารถกดปุ่มทำนายแบบประกัน

TNB INSURANCE	
รับทรัพย์ตลอดชีพ 90/7	
จุดเด่นของผลิตภัณฑ์	ชำระเบี้ย 7 ปี คุ้มครองถึงอายุ 90 ปี รับเงินคืนเมื่อครบสัญญา
ทุนประกัน	1,741,300 บาท
ระยะเวลาชำระเบี้ย	7 ปี
ระยะเวลาคุ้มครอง	อายุ 90 ปี
ความคุ้มครองสูงสุด	100%
เงินจ่ายคืน	2% ทุกๆ 2 ปี จนครบอายุ 89 ปี
เงินครบกำหนดสัญญา	100%
ผลประโยชน์	100%
ดูรายละเอียด	
← ย้อนกลับ	

รูปที่ 4.4 หน้าจอทำนายแบบประกันให้แก่ลูกค้า

TNB INSURANCE

×

ปิดหน้าต่าง

รับทรัพย์ตลอดชีพ 90/7

MTL FAMILY

จ่ายเบี้ยสั้น คุ้มครองยาว

พร้อมรับเงินจ่ายคืนตลอดอายุสัญญา

ชำระเบี้ยประกันภัยสั้นเพียง 7 ปี

แต่คุ้มครองยาวนานถึงครบอายุ 90 ปี

รับเงินจ่ายคืน

2%* ทุก 2 ปี ตั้งแต่สิ้นปีกรมธรรม์ที่ 2, 4, 6, 8, 10, 12, 14

2%* ทุกปี ตั้งแต่ปีกรมธรรม์ที่ 16 จนถึงครบอายุ 89 ปี

ได้รับความคุ้มครองชีวิตตลอดสัญญา 100%*

เบี้ยประกันภัยคงที่ตลอดอายุสัญญา

โครงการเมืองไทย

รับทรัพย์

ตลอดชีพ 90/7

←

ย้อนกลับ

รูปที่ 4.5 หน้าจอแสดง PDF รายละเอียดแบบประกัน

4.4 ผลการประเมินส่วนติดต่อผู้ใช้งาน (วัตถุประสงค์ข้อ 4)

ผลการประเมินต้นแบบ Web Application เพื่อการแนะนำการเลือกซื้อผลิตภัณฑ์แบบ ประกัน ได้ดำเนินการจากกลุ่มผู้ใช้งานที่เป็นตัวแทนขายประกันชีวิตจำนวน 5 ราย เพื่อวัดความพึงพอใจของการใช้งานระบบ และผลการแนะนำประเภทแบบประกันให้ลูกค้า ซึ่งมีการแสดงผลการ ประเมินระบบส่วนติดต่อผู้ใช้งาน ดังตารางที่ 4.3

ตารางที่ 4.3 แสดงผลความพึงพอใจหลังใช้งานต้นแบบ Web Application เพื่อแนะนำการเลือกซื้อ ผลิตภัณฑ์แบบประกัน

ลำดับ	รายการ	ค่าเฉลี่ย $\bar{x}$	S.D.	ระดับความพึงพอใจ
1	ด้านความเหมาะสมของต้นแบบ Web App	4.20	0.45	มากที่สุด
2	ด้านความสะดวกในการใช้งาน	4.60	0.55	มากที่สุด
3	ด้านความพึงพอใจในการประมวลผล	3.40	0.55	มาก
4	ด้านภาพรวมของการใช้ระบบงาน	4.80	0.45	มากที่สุด
	Total	4.25	0.06	มากที่สุด

ต้นแบบ Web Application เพื่อแนะนำการเลือกซื้อผลิตภัณฑ์แบบประกัน ตัวแทนขาย ประกันชีวิตให้ความสนใจและความพึงพอใจในด้านของภาพรวมของการใช้งานระบบในระดับมากที่สุด ที่ค่าเฉลี่ยที่ 4.80 เนื่องจากเป็นเครื่องมือช่วยการขายที่มีประโยชน์ ในลำดับถัดมาเป็นด้านความ สะดวกในการใช้งานที่ค่าเฉลี่ย 4.60 ซึ่งเป็นส่วนที่ตัวแทนขายประกันชีวิตได้มีเครื่องมือช่วยแนะนำ แบบประกันที่ลูกค้าต้องการ ลำดับที่สามคือ ด้านความเหมาะสมของต้นแบบ Web Application ที่มี ค่าเฉลี่ยอยู่ที่ 4.20 และลำดับสุดท้ายคือด้านความสำเร็จหลังทดสอบกับลูกค้า ค่าเฉลี่ยอยู่ที่ 3.40

บทที่ 5

บทสรุป อภิปราย และข้อเสนอแนะ

จากการศึกษางานวิจัยเรื่อง การพัฒนาต้นแบบ Web Application เพื่อแนะนำการเลือกซื้อผลิตภัณฑ์ประกันชีวิต โดยใช้เทคนิคโครงข่ายประสาทเทียม และต้นไม้ตัดสินใจ กรณีศึกษาบริษัทประกันชีวิตแห่งหนึ่ง โดยงานวิจัยนี้มีวัตถุประสงค์ เพื่อวิเคราะห์รูปแบบและพัฒนาต้นแบบ Web Application เพื่อแนะนำการเลือกซื้อแบบประกัน โดยใช้เทคนิคโครงข่ายประสาทเทียมและต้นไม้ตัดสินใจ ได้มีผลสรุปแบ่งออกเป็นหัวข้อต่างๆ ดังต่อไปนี้

1. สรุปผลการวิจัย
2. อภิปรายผลการวิจัย
3. ข้อเสนอแนะงานวิจัย
4. ประโยชน์ที่จะได้รับจากการทำวิทยานิพนธ์

5.1 สรุปผลการวิจัย

สรุปผลการวิจัยเรื่องการพัฒนาต้นแบบ Web Application เพื่อแนะนำการเลือกซื้อผลิตภัณฑ์ประกันชีวิต โดยใช้เทคนิคโครงข่ายประสาทเทียม และต้นไม้ตัดสินใจ โดยสรุปผล ดังรายละเอียดต่อไปนี้

5.1.1 สรุปผลการวิเคราะห์แบบจำลองการพัฒนาต้นแบบ Web Application เพื่อแนะนำการเลือกซื้อผลิตภัณฑ์ประกันชีวิต โดยใช้เทคนิคโครงข่ายประสาทเทียม และต้นไม้ตัดสินใจ

เริ่มต้นโดยการทำ Attribute Selection โดยใช้เทคนิค Decision Tree (ID3) เพื่อคำนวณค่า Information Gain (IG) ของแต่ละ Attribute โดยจะทำการคัดเลือก Attribute ที่มีค่า IG สูง เพื่อประสิทธิภาพในการวิเคราะห์ข้อมูล หลังจากนั้นนำข้อมูลที่ได้นำไปแบ่งกลุ่มข้อมูลด้วย K-means Clustering เมื่อจำแนกกลุ่มข้อมูลออกเป็น 10 กลุ่มแล้ว นำไปสร้างรูปแบบการแนะนำการเลือกซื้อผลิตภัณฑ์แบบประกันด้วยเทคนิคโครงข่ายประสาทเทียม ซึ่งจากการวิเคราะห์เปรียบเทียบเทคนิคหรือวิธีการเลือกใช้อัลกอริทึมระหว่าง อัลกอริทึมเดียว คือ Neural Network และ Decision Tree กับอัลกอริทึมผสม โดยใช้โปรแกรม Weka ในการวิเคราะห์ค่าความถูกต้องแม่นยำในการพยากรณ์ ซึ่งการเลือกใช้อัลกอริทึมแบบผสมนี้ให้ค่าความถูกต้องแม่นยำสูงกว่าเลือกใช้อัลกอริทึมเดียวที่ 91.73% แล้วนำโมเดลที่ได้จากการวิเคราะห์ไปพัฒนาต้นแบบ Web Application การแนะนำการเลือกซื้อผลิตภัณฑ์ประกัน

5.1.2 สรุปผลการสังเคราะห์และประเมินแบบจำลองการพัฒนาต้นแบบ Web Application เพื่อแนะนำการเลือกซื้อผลิตภัณฑ์ประกันชีวิต โดยใช้เทคนิคโครงข่ายประสาทเทียม และต้นไม้ตัดสินใจ

กรอบแนวคิดการทำแบบจำลองสำหรับการวิจัยนี้ประกอบไปด้วย 3 ส่วนคือ 1) Data Preparation Module คือโมดูลการเตรียมข้อมูลเพื่อนำสู่กระบวนการวิเคราะห์ 2) Modeling Module คือโมดูลการวิเคราะห์ข้อมูลให้ได้ผลลัพธ์ที่ต้องการ โดยใช้อัลกอริทึม Decision Tree และ Neural Network วิเคราะห์เปรียบเทียบเปอร์เซ็นต์ความแม่นยำในการแนะนำ 3) Evaluation Module คือโมดูลส่วนของการทดสอบ ประเมินผล ซึ่งผลของการวิจัยในครั้งนี้ จะสามารถนำไปประยุกต์ใช้งานเพื่อแนะนำการเลือกซื้อแบบประกันของลูกค้าได้อย่างเหมาะสมโดยผู้เชี่ยวชาญ 5 ท่านได้ประเมินความเหมาะสมของกรอบแนวคิดนี้อยู่ในระดับสูง เฉลี่ยที่ ($\bar{X} = 4.24$), (S.D. = 0.04)

5.1.3 สรุปผลการพัฒนาต้นแบบ Web Application เพื่อแนะนำการเลือกซื้อผลิตภัณฑ์ประกันชีวิต โดยใช้เทคนิคโครงข่ายประสาทเทียม และต้นไม้ตัดสินใจ

ผลประเมินต้นแบบ Web Application การแนะนำการเลือกซื้อแบบประกันชีวิต จากตัวแทนขายประกันชีวิตจำนวน 5 ราย ที่ได้ใช้ต้นแบบ Web Application ในการแนะนำแบบประกันที่ตรงกับความต้องการของลูกค้า ตัวแทนขายประกันชีวิตทั้ง 5 รายได้ประเมินความพึงพอใจหลังการใช้งานระบบ โดยให้ความพึงพอใจในด้านของภาพรวมของการทำงานของระบบในระดับมากที่สุด เฉลี่ยที่ ($\bar{X} = 4.80$), (S.D. = 0.45) โดยให้ความเห็นว่าเป็นเครื่องมือที่มีประโยชน์ สามารถช่วยการขาย และสร้างความน่าเชื่อถือให้แก่ผู้ขายประกัน ในลำดับที่สอง เป็นด้านความสะดวกในการใช้งาน ที่ค่าเฉลี่ย ($\bar{X} = 4.60$), (S.D. = 0.55)

5.2 อภิปรายผลการวิจัย

การวิเคราะห์แบบจำลองการพัฒนาต้นแบบ Web Application เพื่อแนะนำการเลือกซื้อผลิตภัณฑ์ประกันชีวิต โดยใช้เทคนิคอัลกอริทึมแบบผสม โดยใช้เทคนิคโครงข่ายประสาทเทียม (Neural Network) และเทคนิคต้นไม้ตัดสินใจ (Decision Tree) ในการวิเคราะห์ข้อมูล ผลจากการวิเคราะห์รูปแบบการแนะนำการเลือกซื้อแบบประกันในขั้นตอนสุดท้ายโดยเลือกใช้เทคนิคโครงข่ายประสาทเทียม ได้ผลออกมาเป็นจำนวน Node และค่าน้ำหนักระหว่างแต่ละ Node (Input Node, Hidden Node และ Output Node) เพื่อนำไปสร้างต้นแบบ Web Application การแนะนำการเลือกซื้อผลิตภัณฑ์ประกันชีวิต ที่มีค่าความถูกต้องและแม่นยำอยู่ที่ 91.73%

การเลือกใช้เทคนิค Neural Network เป็นขั้นตอนสุดท้ายในการวิเคราะห์ข้อมูล ถึงแม้ว่าผลลัพธ์การแนะนำการเลือกซื้อผลิตภัณฑ์ประกันชีวิตมีความถูกต้องและมีความแม่นยำสูง แต่ข้อเสียคือ ไม่สามารถทำความเข้าใจและทราบได้ว่าที่มาของผลลัพธ์นั้นมาได้อย่างไร ซึ่งต่างจากการใช้เทคนิค Decision Tree ที่ผลลัพธ์ออกมาในรูปแบบของ rule base ซึ่งสามารถทำความเข้าใจได้ง่ายว่าความเป็นมาของผลลัพธ์นั้นมาได้อย่างไร

อย่างไรก็ตามผลลัพธ์ในการแนะนำผลิตภัณฑ์ประกันชีวิตสำหรับงานวิจัยนี้ เป็นข้อเสนอแนะหรือกลุ่มของแบบประกันที่แนะนำเท่านั้น ซึ่งผลลัพธ์นี้ยังไม่สามารถนำไปแนะนำเพื่อนำเสนอขายได้จริง ดังนั้น ผลลัพธ์ที่ดีควรแนะนำเป็นชื่อแบบประกันชีวิตที่มีเหมาะสมกับลูกค้า

5.3 ข้อเสนอแนะ

5.3.1 การเตรียมพร้อมข้อมูลให้สมบูรณ์เพื่อนำไปวิเคราะห์ เป็นขั้นตอนที่ใช้เวลานานเนื่องจากข้อมูลที่ได้รับมา มีความผิดพลาดและไม่สมบูรณ์ เช่น ความซ้ำซ้อนของข้อมูล (Duplicated Data), ข้อมูลสูญหาย และการแปลงประเภทของข้อมูล (Data Type) ให้สามารถใช้งานได้กับโปรแกรมที่เลือกใช้ เช่น การแปลงข้อมูลให้อยู่ในรูปของตัวเลข (Categorical Data) และการแปลงข้อมูลให้อยู่ในรูปแบบ ARFF Format เป็นต้น

5.3.2 ผลการแนะนำยังมีความผิดพลาด เนื่องจากชุดข้อมูลที่ให้ระบบเรียนรู้ยังมีไม่เพียงพอ ตัวอย่างเช่น ข้อมูลที่ใช้ในงานวิจัยนี้ ส่วนใหญ่เป็นข้อมูลที่มีลูกค้าซื้อแบบประกันประเภท Protective และ Saving เป็นจำนวนมาก มีข้อมูลที่ลูกค้าซื้อแบบประกันประเภท Annuity และ Unit Link มีจำนวนอยู่น้อย เทียบเป็นร้อยละแค่ 15% ของจำนวนข้อมูลทั้งหมด ดังนั้นผลลัพธ์การแนะนำที่ถูกต้องที่ควรจะออกมาเป็น Annuity และ Unit Link ก็ลดน้อยลงไปด้วย

5.3.3 การพัฒนาระบบเพื่อต่อยอดจากงานวิจัยนี้ เพื่อนำไปใช้งานได้จริงในอนาคต สำหรับผู้ที่มีความสนใจสามารถนำแนวคิดการทำเหมืองข้อมูลนี้ไปประยุกต์ใช้กับองค์กรที่มีฐานข้อมูลลูกค้าที่ทำประกันชีวิตอยู่แล้ว ให้เกิดมูลค่ามากยิ่งขึ้น วิเคราะห์ข้อมูลลูกค้าเพื่อออกโปรโมชั่น สร้าง campaign หรือสร้างระบบเพื่อแนะนำผลิตภัณฑ์แบบประกันชีวิตตามเป้าหมายของลูกค้า หรือที่เรียกว่าระบบ Financial Needs Analysis ที่สามารถแนะนำแบบประกันตามเป้าหมาย (Life goal) และแนะนำแบบประกันตามความสามารถในการชำระเบี้ยของลูกค้าได้ เพื่อแก้ไขปัญหาการนำเสนอขายแบบประกันที่ไม่เหมาะสม และไม่ตรงกับความต้องการของลูกค้า เนื่องจากตัวแทนจะเน้นเลือกขายแบบประกันที่ได้รับค่าคอมมิชชั่นสูงให้กับลูกค้า โดยไม่สนใจว่าลูกค้ามีความต้องการและสามารถชำระค่าเบี้ยประกันได้หรือไม่

5.4 ประโยชน์ที่ได้รับจากการทำวิทยานิพนธ์

1. ผู้วิจัยได้มีการศึกษา วิเคราะห์ และเรียนรู้กระบวนการทำเหมืองข้อมูล ที่สามารถนำไปประยุกต์ใช้และช่วยเสริมการทำงานอยู่ในปัจจุบัน
2. ผู้วิจัยได้มีการฝึกค้นคว้าหาข้อมูล (Research) เกี่ยวกับทฤษฎีการทำเหมืองข้อมูล รวมถึงศึกษาค้นคว้าผลงานบุคคลอื่นๆที่เกี่ยวข้อง ซึ่งเป็นพื้นฐานที่สำคัญของการทำวิจัยและการสร้างสรรค์ผลงานอื่นๆ
3. ผู้วิจัยได้เรียนรู้และฝึกฝนการออกแบบระบบการทำงาน (Application Flow) และการพัฒนาต้นแบบ Web Application เพื่อแนะนำการเลือกซื้อผลิตภัณฑ์แบบประกันชีวิต



บรรณานุกรม

TNI

บรรณานุกรม

- [1] W. Chaisangmomgkon, “Machine Learning คืออะไร และต่างจาก Data Mining และ Data Science อย่างไร,” Big Data Experience Organization [Online]. Available: <http://bigdataexperience.org/2016/06/134>. [Accessed: September 28, 2016].
- [2] L. Bedgood, “Big Data Is Making a Splash in the Insurance Industry,” [Online]. Available: <http://datascienceth.com/biginsurancedata/>. [Accessed: September 28, 2016].
- [3] พิเชฐ เจริญณิทธิสิน, “How Important of Life Insurance,” Sawasdeeactuary [Online]. Available: http://www.sawasdeeactuary.com/art_lifeins.html. [Accessed: February 17, 2016]
- [4] บริษัท ซิลค์สแปน จำกัด, “ศัพท์น่ารู้เกี่ยวกับการประกันภัย,” Silkspan [Online]. Available: <http://www.silkspan.in.th/silkspan1/article/mbinsmain2.asp>. [Accessed: January 19, 2016]
- [5] Cymiz, “การแบ่งประเภทของประกันชีวิต,” Cymiz [Online]. Available: <http://www.cymiz.com/insurance/life-insurance-category3.html>. [Accessed: February 23, 2016]
- [6] J. Raethong, “Data Mining,” NextProject.net [Online]. Available: <http://www.nextproject.net/contents/?00043>. [Accessed: May 13, 2016]
- [7] บริษัท มาสเตอร์ทีม จำกัด, “หลักพื้นฐานของการประกันภัย,” MasterTeam [Online]. Available: <http://www.masterteam.co.th/insurist/main>. [Accessed: January 19, 2016]
- [8] ชุตินา จิรภฤตยากุล, “ปัจจัยที่มีอิทธิพลต่อพฤติกรรมการซื้อประกันสุขภาพของบริษัท เอไอเอ จำกัด ของผู้บริโภคในเขตกรุงเทพมหานคร,” สารนิพนธ์ บธ.ม. (การตลาด), มหาวิทยาลัยธุรกิจบัณฑิต, กรุงเทพฯ, 2558.
- [9] Automated, “Data Mining,” Automated Software [Online]. Available: <http://www.autosoft.in.th/data-warehouse/>. [Accessed: September 29, 2016]
- [10] Gintamaz, “Data Mining Technique,” Llemonz Blogspot [Online]. Available: http://llem0nz.blogspot.com/2012/03/data-mining_8699.html. [Accessed: September 29, 2016]

- [11] MathWorks, “*Supervised Learning Workflow and Algorithms*,” MathWorks, [Online]. Available: <http://www.mathwoks.com/help/stats/supervised-learning-Machine-learning-workflow-and-algorithms.html>. [Accessed: January 25, 2016]
- [12] O. Richard et al., *Pattern Classification Second Edition*, John Wiley & Sons, 2001.
- [13] C. Kaewchinporn, “A combination of decision tree learning and clustering for data classification,” *2011 Eighth International Joint Conference on Computer Science and Software Engineering, JCSSE*, Mahidol University, Salaya Campus, Nakhon Pathom, Thailand, May 11-13, 2011, pp. 16.
- [14] W. Buathong, “*Clustering K-means*,” [Online] . Available: <https://wipawanblog.wordpress.com/data-mining/>. [Accessed: January 19, 2016]
- [15] N. Sucharittham et al., “*Data Mining for Life Insurance Knowledge Extraction: A Survey*,” M.S. Thesis (Information Technology), Thammasat University, Bangkok, Thailand, 2011.
- [16] ภคนันท์ นิสิตศิริ, “ปัจจัยที่มีผลต่อการตัดสินใจซื้อประกันชีวิตของพนักงานในองค์กรการไฟฟ้าฝ่ายผลิตแห่งประเทศไทย สาขาแม่เมาะ จังหวัดลำปาง,” สารนิพนธ์ ศ.ม. (เทคโนโลยีสารสนเทศและการสื่อสาร), มหาวิทยาลัยเชียงใหม่, เชียงใหม่, 2558.
- [17] จันทร์พร ดาวลอย และ เทพฤทธิ์ บัณฑิตวัฒนาวงศ์, “การวิเคราะห์และทำนายพฤติกรรมการทำงานทำประกันชีวิต,” สารนิพนธ์ วท.ม. (เทคโนโลยีสารสนเทศและการสื่อสาร), มหาวิทยาลัยศรีปทุม, กรุงเทพฯ, 2557.
- [18] กฤษณพร สุริยะบรรเทิง และ กมล เกียรติเรืองกมล, “การสร้างแบบจำลองการขายผลิตภัณฑ์ประกันภัยให้กับลูกค้าผู้สูงอายุกลุ่มบัญชีออมทรัพย์โดยการทำเหมืองข้อมูล,” วิทยานิพนธ์ วท.ม. (การบริหารเทคโนโลยี), มหาวิทยาลัยธรรมศาสตร์, กรุงเทพฯ, 2557.
- [19] C. Rohitash et al., “The combination and comparison of neural networks with decision trees for wine classification,” *The 2007 International Conference on Data Mining, DMIN 2007*, Las Vegas, USA, June 25-28, 2007, pp. 13-15.
- [20] D. Maryam and A. Marzieh, “A hybrid data mining model to improve customer response modeling in direct marketing,” *Indian Journal of Computer Science and Engineering, IJCSE 2013*, Bay of Bengal, India, July 4-6, 2013, pp. 844-855.

- [21] F. Simon et al., "Applying a hybrid model of neural network and decision tree classifier for predicting university admission," *International Conference on Information Communications System, ICICS 2009*, Beijing, China, December 20-22, 2009, pp. 1-5.



ภาคผนวก

TNI

ภาคผนวก ก.
แบบสำรวจที่ใช้ในงานวิจัย

TNI

แบบสำรวจที่ใช้ในงานวิจัย

1. แบบประเมินแบบจำลองการพัฒนาต้นแบบ Web Application เพื่อแนะนำการเลือกซื้อผลิตภัณฑ์ประกันชีวิต โดยใช้เทคนิคโครงข่ายประสาทเทียม และต้นไม้ตัดสินใจ

แบบประเมินแบบจำลอง

เรื่อง ประเมินแบบจำลองการแนะนำการเลือกซื้อผลิตภัณฑ์ประกันชีวิต

คำชี้แจง

1. แบบสอบถามนี้มีวัตถุประสงค์เพื่อประเมินแบบจำลองการพัฒนาต้นแบบ Web Application เพื่อแนะนำการเลือกซื้อผลิตภัณฑ์ประกันชีวิต โดยใช้เทคนิคโครงข่ายประสาทเทียม และต้นไม้ตัดสินใจ
 2. แบบสอบถามนี้แบ่งเป็น 2 ส่วน ดังนี้
 - ส่วนที่ 1 ข้อมูลพื้นฐานของผู้ตอบแบบสอบถาม
 - ส่วนที่ 2 ข้อมูลรายละเอียดการประเมินแบบจำลอง
- โปรดกรอกข้อมูลให้ครบทุกส่วนและตรงกับความเป็นจริงมากที่สุด

ส่วนที่ 1 ข้อมูลพื้นฐานของผู้ตอบแบบสำรวจ

ชื่อผู้ประเมิน..... ตำแหน่ง.....

ความเชี่ยวชาญ มาก/ปานกลาง/น้อย

ส่วนที่ 2 ข้อมูลรายละเอียดการประเมินแบบจำลอง

รายละเอียดการประเมิน	ความเหมาะสมของแนวคิด				
	มากที่สุด	มาก	ปานกลาง	น้อย	น้อยที่สุด
1. ความเหมาะสมของ Data Preparation					
2. ความเหมาะสมของ Data Transformation					
3. ความเหมาะสมของ Data Mining Technique					
4. ความเหมาะสมของ Testing & Evaluation					
5. ความเหมาะสมของภาพรวมของ Model นี้					

ข้อเสนอแนะ

.....

.....

ขอขอบคุณทุกท่านที่สละเวลาตอบแบบสอบถาม

2. แบบสอบถามความพึงพอใจของผู้ใช้งานต้นแบบ Web Application เพื่อการแนะนำการเลือกซื้อผลิตภัณฑ์แบบประกันชีวิต ที่วิเคราะห์ด้วยเหมืองข้อมูล

แบบสอบถาม

เรื่อง ความพึงพอใจของผู้ใช้งานต้นแบบ Web Application เพื่อการแนะนำการเลือกซื้อผลิตภัณฑ์แบบประกันชีวิต ที่วิเคราะห์ด้วยเหมืองข้อมูล

คำชี้แจง

1. แบบสำรวจนี้จัดทำขึ้นเพื่อสอบถามความพึงพอใจของผู้ใช้งานต้นแบบ Web Application เพื่อการแนะนำการเลือกซื้อแบบประกันชีวิต ที่วิเคราะห์ด้วยเหมืองข้อมูล จึงใคร่ขอความร่วมมือจากท่านในการตอบแบบสอบถามตามความเป็นจริง พร้อมแสดงข้อเสนอแนะ (ถ้ามี) เพื่อเป็นประโยชน์ต่องานวิจัยต่อไป

2. แบบสอบถามนี้แบ่งเป็น 2 ส่วน

ส่วนที่ 1 ข้อมูลพื้นฐานของผู้ตอบแบบสอบถาม

ส่วนที่ 2 ความพึงพอใจในการใช้งานระบบ

โปรดกรอกข้อมูลให้ครบทุกส่วนและตรงกับความเป็นจริงมากที่สุด

ส่วนที่ 1 ข้อมูลพื้นฐานของผู้ตอบแบบสอบถาม

ชื่อ.....เพศ.....อายุ.....

ตัวแทนขายประกันชีวิตบริษัท.....สังกัด.....

TNI

THAI - NICHI INSTITUTE OF TECHNOLOGY

ส่วนที่ 2 ความพึงพอใจในการใช้งานระบบ

โปรดทำเครื่องหมาย ✓ ลงในช่องที่ท่านคิดว่าเหมาะสมที่สุด

รายการ	ระดับความคิดเห็น				
	มากที่สุด (5)	มาก (4)	ปานกลาง (3)	น้อย (2)	น้อยที่สุด (1)
1. ด้านความเหมาะสมของต้นแบบ Web Application					
2. ด้านความสะดวกในการใช้งาน					
3. ด้านความพึงพอใจในการประมวลผล					
4. ด้านภาพรวมของการใช้งานระบบ					

ข้อเสนอแนะ

.....

.....

.....

ผู้วิจัยขอขอบคุณในความร่วมมือของผู้ทดสอบระบบมา ณ โอกาสนี้

TNI

TAHITI - NICHI INSTITUTE OF TECHNOLOGY