



การวิเคราะห์การตีมูลค่าของประชาชนในจังหวัดหนองคาย
โดยเทคนิคเหมืองข้อมูลและการถดถอยโลจิสติก

จักรพงษ์ มานะดี

เสนอต่อมหาวิทยาลัยมหาสารคาม เพื่อเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการจัดการสถิติ
กุมภาพันธ์ 2557

ลิขสิทธิ์เป็นของมหาวิทยาลัยมหาสารคาม



การวิเคราะห์การดื่มสุราของประชาชนในจังหวัดหนองคาย
โดยเทคนิคเหมืองข้อมูลและการถดถอยโลจิสติก

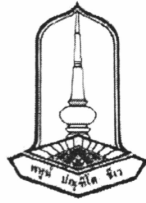
จักรพงษ์ มานะดี

เสนอต่อมหาวิทยาลัยมหาสารคาม เพื่อเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการจัดการสถิติ

กุมภาพันธ์ 2557

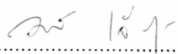
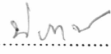
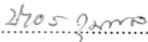
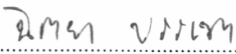
ลิขสิทธิ์เป็นของมหาวิทยาลัยมหาสารคาม



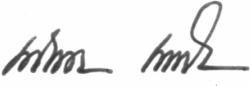


คณะกรรมการสอบวิทยานิพนธ์ ได้พิจารณาวิทยานิพนธ์ของนายจักรพงษ์ มานะดี แล้วเห็นสมควรรับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการจัดการสถิติ ของมหาวิทยาลัยมหาสารคาม

คณะกรรมการสอบวิทยานิพนธ์

 (อาจารย์ ดร.นริสา วงศ์พนารักษ์)	ประธานกรรมการ (กรรมการบัณฑิตศึกษาประจำคณะ)
 (อาจารย์ ดร.มนชยา เจียงประดิษฐ์)	กรรมการ (ประธานกรรมการควบคุมวิทยานิพนธ์)
 (ผศ.ดร.นิภาพร ชุตินันต์)	กรรมการ (กรรมการควบคุมวิทยานิพนธ์)
 (ผศ.ดร.บังอร กุมพล)	กรรมการ (กรรมการบัณฑิตศึกษาประจำคณะ)
 (อาจารย์ ดร.นิตยา บรรเทา)	กรรมการ (ผู้ทรงคุณวุฒิ)

มหาวิทยาลัยอุบลราชธานีได้รับวิทยานิพนธ์ฉบับนี้ เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการจัดการสถิติ ของมหาวิทยาลัยมหาสารคาม


.....
(รศ.เทียนศักดิ์ เมฆพรรณโอภาส)
ผู้รักษาราชการแทนคณบดีคณะวิทยาศาสตร์


.....
(รศ.ดร.สุนันท์ สายกระสุน)
ผู้รักษาราชการคณบดีบัณฑิตวิทยาลัย
วันที่ 26 เดือน ก.พ. พ.ศ. 2557



กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จได้ด้วยความรู้และความช่วยเหลือจากท่านอาจารย์ ดร. มนชยา เจียงประดิษฐ์ ประธานกรรมการควบคุมวิทยานิพนธ์ ผู้ช่วยศาสตราจารย์ ดร. นิภาพร ชุติมันต์ กรรมการ และโดยเฉพาะอย่างยิ่งท่านผู้ช่วยศาสตราจารย์ ดร. บังอร กุมพล ที่ได้สละเวลาชี้แนวทางในการศึกษาวิจัยและให้ข้อเสนอแนะในการแก้ไขข้อบกพร่อง และขอขอบคุณครู อาจารย์จากอดีตจนถึงปัจจุบันทุกๆ ท่านที่ช่วยสั่งสอนให้มีวิชาความรู้ตลอดเวลาที่ผ่านมา

ขอขอบพระคุณ อาจารย์ ดร. นริสา วงศ์พนารักษ์ ประธานกรรมการสอบวิทยานิพนธ์ และอาจารย์ ดร. นิตยา บรรเทา จากมหาวิทยาลัยราชภัฏมหาสารคาม กรรมการผู้ทรงคุณวุฒิภายนอก ที่กรุณาสละเวลาและให้ข้อเสนอแนะที่เป็นประโยชน์ยิ่ง ซึ่งผู้วิจัยรู้สึกเป็นเกียรติอย่างยิ่ง

ขอกราบขอบพระคุณและขอรำลึกในพระคุณของ คุณพ่อทวีชัย คุณแม่กรรณิการ์ มานะดี ครอบครัว ญาติพี่น้อง รวมถึงข้าราชการ พนักงานจากสำนักงานสถิติจังหวัดหนองคายและเพื่อนๆ ร่วมสาขาวิชา ที่คอยให้กำลังใจสนับสนุนมาโดยตลอด ผู้วิจัยขอขอบพระคุณไว้ ณ โอกาสนี้ด้วย

จักรพงษ์ มานะดี



ชื่อเรื่อง	การวิเคราะห์การติ่มสุราของประชาชนในจังหวัดหนองคาย โดยเทคนิค เหมือนข้อมูลและการถดถอยโลจิสติก		
ผู้วิจัย	นายจักรพงษ์ มานะดี		
ปริญญา	วิทยาศาสตรมหาบัณฑิต	สาขาวิชา	วิทยาการจัดการสถิติ
กรรมการควบคุม	อาจารย์ ดร. มนชยา เจียงประดิษฐ์ ผู้ช่วยศาสตราจารย์ ดร.นิภาพร ชุตินันต์		
มหาวิทยาลัย	มหาวิทยาลัยมหาสารคาม	ปีที่พิมพ์	2557

บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์ เพื่อสร้างรูปแบบที่เหมาะสมในการพยากรณ์กลุ่มของพฤติกรรมของ
การติ่มสุราของประชากรในจังหวัดหนองคายและเปรียบเทียบประสิทธิภาพการพยากรณ์ของเทคนิค
การทำเหมือนข้อมูลและเทคนิคการวิเคราะห์การถดถอยโลจิสติก โดยเทคนิคการทำเหมือนข้อมูลใช้
แบบจำลองของต้นไม้ตัดสินใจที่มีอัลกอริทึม C4.5 (J48) และแบบจำลองโครงข่ายประสาทเทียมที่มี
ลักษณะโครงสร้างแบบมัลติเลเยอร์เพอร์เซพตรอน (Multilayer Perceptron) ที่มีการเรียนรู้แบบ
แบ็คพรอพาเกชัน อัลกอริทึม (Back-propagation Algorithm) และเทคนิคการวิเคราะห์การถดถอย
โลจิสติก ใช้การวิเคราะห์การถดถอยโลจิสติกแบบหลายกลุ่ม (Multinomial Logistic Regression
Analysis) จากข้อมูลการสำรวจพฤติกรรมการสูบบุหรี่และติ่มสุราของประชากร พ.ศ. 2554 ซึ่ง
ดำเนินการสำรวจโดยสำนักงานสถิติแห่งชาติ จำนวน 605 ราย โดยใช้ค่าความถูกต้องของการพยากรณ์
MAPE และ RMSE เป็นเกณฑ์ในการวัดความแม่นยำของการพยากรณ์ ผลการศึกษาพบว่าแบบจำลอง
โครงข่ายประสาทเทียม มีความแม่นยำที่สุดในการพยากรณ์ และสามารถแบ่งกลุ่มการติ่มสุราออกได้เป็น
3 กลุ่ม ดังนี้ คือ กลุ่มไม่เคยติ่มสุรา กลุ่มติ่มเล็กน้อยหรือนาน ๆ ครั้ง และกลุ่มติ่มสม่ำเสมอ

คำสำคัญ : การติ่มสุรา; การทำเหมือนข้อมูล; การถดถอยโลจิสติก



TITLE The analysis of alcoholic consumption of the people in Nong Khai province by data mining techniques and logistic regression analysis.

AUTHOR Mr. Jakrapong Manadee

DEGREE Master Degree of Science **MAJOR** Management Science Statistics

COMMITTEE Monchaya Chiangpradit, Ph.D.
Asst. Prof. Nipapron Chutiman, Ph.D.

UNIVERSITY Maharakham University **YEAR** 2014

ABSTRACT

The purposes of this research were to establish an appropriate model for predicting the behavior of a group of people drinking alcohol in Nong Khai and to compare the performance of forecasting techniques with data mining techniques and logistic regression analysis. Data mining models, by decision tree with the algorithm C4.5 (J48) and the artificial neural networks (ANN) with the multilayer perceptron with the back propagation algorithm and logistic regression analysis with multinomial logistic regression analysis were analyzed. The secondary data of the survey data as smoking habits and alcohol consumption, population 2011 survey from NATIONAL STATISTICAL OFFICE, with 605 cases, were studies. The percentage of correction, MAPE and RMSE were the criteria. The ANN showed the best performance for forecasting. As a result, three groups of the alcoholic consumption were never drink group, drink little or infrequently group and uniform drinking group.

Keywords : Drinking Alcohol; Data mining; Logistic regression



สารบัญ

	หน้า
กิตติกรรมประกาศ	ก
บทคัดย่อภาษาไทย	ข
บทคัดย่อภาษาอังกฤษ	ค
สารบัญตาราง	ฉ
สารบัญภาพประกอบ	ช
บทที่ 1 บทนำ	1
1.1 ภูมิหลัง	1
1.2 ความมุ่งหมายของการวิจัย	4
1.3 ความสำคัญของการวิจัย	4
1.4 ขอบเขตของการวิจัย	4
1.5 นิยามศัพท์เฉพาะ	4
บทที่ 2 แนวคิด ทฤษฎี และงานวิจัยที่เกี่ยวข้อง	6
2.1 ความรู้เกี่ยวกับการดื่มสุราหรือเครื่องดื่มแอลกอฮอล์	6
2.2 การทำเหมืองข้อมูล	13
2.3 การวิเคราะห์การถดถอยโลจิสติกแบบหลายกลุ่ม (Multinomial Logistic Regression)	28
2.4 งานวิจัยที่เกี่ยวข้อง	29
บทที่ 3 วิธีดำเนินการวิจัย	35
3.1 ขั้นตอนในการดำเนินการศึกษาวิจัย	35
3.2 กรอบแนวคิดในการวิจัย	37
3.3 เครื่องมือที่ใช้ในการวิจัย	38
3.4 การเก็บรวบรวมข้อมูล	38
3.5 การจัดกระทำและการวิเคราะห์ข้อมูล	38
3.6 การเปรียบเทียบประสิทธิภาพแบบจำลอง	47
บทที่ 4 ผลการวิจัย	48
4.1 สัญลักษณ์ในการนำเสนอ	48
4.2 ผลการทดสอบแบบจำลองการทำเหมืองข้อมูล	49
4.3 ผลการทดสอบแบบจำลองการถดถอยโลจิสติก	59
4.4 การเปรียบเทียบประสิทธิภาพแบบจำลอง	62
บทที่ 5 สรุปผล อภิปรายผล และข้อเสนอแนะ	64
5.1 สรุปผล	64
5.2 อภิปรายผล	66
5.3 ข้อเสนอแนะ	67



	หน้า
เอกสารอ้างอิง	69
ภาคผนวก	74
ภาคผนวก ก รายละเอียดของเทคนิคการทำเหมืองข้อมูล	75
ภาคผนวก ข รายละเอียดแบบจำลองการถดถอยโลจิสติก	105
ประวัติย่อผู้วิจัย	110



สารบัญตาราง

		หน้า
ตาราง 2.1	ชนิดของเครื่องดื่มแอลกอฮอล์และปริมาณแอลกอฮอล์ (ดีกรี)	10
ตาราง 2.2	ปริมาณแอลกอฮอล์ที่บริโภคต่อวันที่ดื่มกับระดับความเสี่ยงแบบเฉียบพลัน จำแนกตามเพศ	11
ตาราง 3.1	ค่าของตัวแปรที่ทำการแปลงรูปแบบข้อมูล	39
ตาราง 4.1	การเปรียบเทียบประสิทธิภาพแบบจำลองต้นไม้ตัดสินใจ	50
ตาราง 4.2	ประสิทธิภาพแบบจำลองต้นไม้ตัดสินใจตามแบบจำลอง TR4X4 และ TR9X4	54
ตาราง 4.3	ประสิทธิภาพแบบจำลองต้นไม้ตัดสินใจตามแบบจำลอง TR9X4	54
ตาราง 4.4	สัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรอิสระและตัวแปรตาม	57
ตาราง 4.5	การวัดประสิทธิภาพการฝึกสอนและการทดสอบแบบจำลองโดยใช้โครงข่าย ประสาทเทียม	58
ตาราง 4.6	ประสิทธิภาพแบบจำลองโครงข่ายประสาทเทียม ตามแบบจำลอง ANN1	59
ตาราง 4.7	วิเคราะห์ความเหมาะสมของแบบจำลองการถดถอยโลจิสติกของตัวแปรที่ส่งผลต่อ ต่อการดื่มสุราของคนในจังหวัดหนองคาย	59
ตาราง 4.8	ค่า Odds Ratio จากการวิเคราะห์การถดถอยโลจิสติกของตัวแปรที่ส่งผลต่อ กลุ่มการดื่มสุราของประชาชนในจังหวัดหนองคาย	60
ตาราง 4.9	ประสิทธิภาพของแบบจำลองการถดถอยโลจิสติก	62



สารบัญภาพประกอบ

	หน้า
ภาพประกอบ 2.1 Model ของ Neuron ในสมองมนุษย์	19
ภาพประกอบ 2.2 Model ของ Neuron ในคอมพิวเตอร์	19
ภาพประกอบ 2.3 การแยกแยะระหว่างสี่เหลี่ยมและสามเหลี่ยม	20
ภาพประกอบ 2.4 โครงสร้างวงจร Neural Network	21
ภาพประกอบ 2.5 รูปแบบ Back-propagation Neural Network	21
ภาพประกอบ 2.6 อนุกรมวิธานของโครงข่ายประสาทเทียม	23
ภาพประกอบ 2.7 การเรียนรู้แบบมีการสอน (Supervised Learning)	23
ภาพประกอบ 2.8 การเรียนรู้แบบไม่มีการสอน Unsupervised Learning	24
ภาพประกอบ 2.9 สถาปัตยกรรมของ Feedforward Network	24
ภาพประกอบ 2.10 สถาปัตยกรรมของ Feedback Network	25
ภาพประกอบ 2.11 สถาปัตยกรรมของเพอร์เซพตรอนชั้นเดียว (Single-layer Perceptron)	26
ภาพประกอบ 2.12 โครงสร้างของ Perceptrons	27
ภาพประกอบ 3.1 ขั้นตอนในการดำเนินการศึกษาวิจัย	36
ภาพประกอบ 3.2 กรอบแนวคิดในการวิจัย	37
ภาพประกอบ 4.1 ขนาดของแบบจำลองต้นไม้ตัดสินใจ	51
ภาพประกอบ 4.2 ร้อยละของความแม่นยำ (Precision) ของแบบจำลองต้นไม้ตัดสินใจ	52
ภาพประกอบ 4.3 ร้อยละของความระลึก (Recall) ของแบบจำลองต้นไม้ตัดสินใจ	52
ภาพประกอบ 4.4 ร้อยละของค่าความเหวี่ยง (F-Measure) ของแบบจำลองต้นไม้ตัดสินใจ	53
ภาพประกอบ 4.5 ร้อยละความถูกต้องในการจำแนกกลุ่ม (Correctly Classified Instances) ของแบบจำลองต้นไม้ตัดสินใจ	53
ภาพประกอบ 4.6 ต้นไม้จำลองตัดสินใจ ตาม Model TR9X4	55



บทที่ 1

บทนำ

1.1 ภูมิหลัง

ปัจจุบันการดื่มสุราหรือเครื่องดื่มที่มีแอลกอฮอล์เป็นเรื่องปกติของคนในสังคม โดยประเทศไทย มีแนวโน้มการบริโภคเครื่องดื่มแอลกอฮอล์เพิ่มมากขึ้น โดยจากข้อมูลองค์การอนามัยโลก พบว่าปัจจุบันมีประชากรทั่วโลกดื่มเครื่องดื่มแอลกอฮอล์ประมาณ 2,000 คน หรือ 1 ใน 3 ของประชากรโลก สำหรับคนไทย พบว่า ดื่มเครื่องดื่มแอลกอฮอล์สูงเป็นอันดับที่ 40 ของโลก และอันดับ 3 ของเอเชีย โดยผู้ที่มีอายุ 15 ปีขึ้นไป ดื่มเครื่องดื่มแอลกอฮอล์ 16.2 ล้านคน หรือเฉลี่ยดื่มคนละ 58 ลิตรต่อปี สูงกว่าค่าเฉลี่ยโลก 9 เท่าตัว (สำนักงานคณะกรรมการพัฒนาการเศรษฐกิจและสังคมแห่งชาติ, 2555) อีกทั้งค่านิยมและวัฒนธรรมไทยให้ความสำคัญกับเครื่องดื่มแอลกอฮอล์ว่าเป็นเครื่องดื่มที่ใช้เพื่อเข้าสังคม ช่วยสร้างความสนุกสนาน ผ่อนคลายและลดความเครียด และสามารถหาซื้อได้ง่าย รัฐบาลเองก็ส่งเสริมสนับสนุนให้มีการผลิตและซื้อขายกันอย่างเสรี ไม่ผิดกฎหมาย ผู้ผลิตเองก็มีการพัฒนากลยุทธ์ในรูปแบบต่างๆ โฆษณาสินค้าในสื่อประเภทต่างๆ ไม่ว่าจะเป็นสื่อสิ่งพิมพ์ วิทยุ โทรทัศน์ ป้ายโฆษณาขนาดใหญ่ ที่บุคคลทุกระดับตั้งแต่ได้ระดับประถมศึกษาที่ยังไม่บรรลุนิติภาวะ กลุ่มเยาวชน กลุ่มวัยรุ่น จนถึงวัยผู้ใหญ่ที่สามารถรู้ผิดชอบชั่วดีแล้วก็ได้รับชม ซึ่งก็เป็นส่วนหนึ่งที่ส่งเสริมให้ผู้ชมสิ่งเหล่านี้มองว่าเป็นสิ่งที่ดีเพิ่มความสนุกสนาน (อมรรภรณ์ ผางแก้ว, 2555) เพราะทำไมมีแค่ผู้ใหญ่ที่ดื่มสุรา แต่ยังมีทั้งเด็กวัยรุ่น และเยาวชนที่ดื่มสุราเพิ่มขึ้นเป็นจำนวนมากและเป็นปัญหาที่ใหญ่ในสังคม ทั้งยังทวีความรุนแรงเพิ่มมากขึ้นทุกวัน ซึ่งจากสาเหตุของการดื่มอาจส่งผลทำให้เกิดผลกระทบทั้งทางด้านร่างกาย จิตใจ สังคม และผู้ที่ดื่มเครื่องดื่มแอลกอฮอล์เป็นประจำจะทำให้เกิดการติดแอลกอฮอล์ได้และอาจทำให้เกิดพฤติกรรมที่ไม่พึงประสงค์หรือที่ไม่ได้คาดหมายเกิดขึ้นได้ จากความรุนแรงของปัญหาและผลกระทบที่เกิดจากพฤติกรรมกรรมการดื่มสุราหรือเครื่องดื่มแอลกอฮอล์จะทำให้เกิดโรคตับแข็ง มะเร็งตับหรือหัวใจ ฯลฯ แล้วอันตรายที่หลายคนจะไม่นึกถึงไม่ได้คือผลต่อโรคทางจิตประสาท ซึ่งส่งผลกระทบต่อร่างกายอย่างมหาศาลได้ (กรมสุขภาพจิต, 2548) ส่วนผลกระทบต่อครอบครัวและสังคม อาจทำให้เกิดปัญหาทะเลาะวิวาท ปัญหาความรุนแรง หย่าร้าง เกิดขึ้นในครอบครัวหรือชุมชนสังคมที่อยู่ และผู้ติดสุราต้องเสียค่าใช้จ่ายในเรื่องสุรามากกว่าค่าใช้จ่ายในด้านอื่นๆ ขาดประสิทธิภาพในการทำงานทำให้รายได้จากการทำงานลดลง มีปัญหาเรื่องรายได้ในครอบครัว นอกจากนี้ผู้ที่ติดสุรายังขาดความรับผิดชอบในบทบาทของ พ่อ แม่ เด็กหรือลูกหลานถูกทอดทิ้ง ขาดความอบอุ่นในครอบครัว ซึ่งเป็นสาเหตุให้เด็กมีพฤติกรรมไม่ดี เกเร เป็นปัญหาของสังคม เพราะได้เห็นตัวอย่างที่ไม่เหมาะสมของผู้ติดสุราทำให้บุตรหลานและเยาวชนเห็น จากการศึกษาสถิติความรุนแรงในครอบครัวจากหนังสือพิมพ์ปี 2550-2551 โดยมูลนิธิเพื่อนหญิง พบว่า ในช่วงสองปีมีจำนวนการนำเสนอข่าวความรุนแรงในครอบครัวทั้งสิ้น 994 ข่าว โดยปี 2550 มี 518 ข่าวและลดลงเป็น 426 ข่าว ในปี 2551 ขณะที่สัดส่วนของข่าวความรุนแรงในครอบครัวที่มีเครื่องดื่มแอลกอฮอล์เป็นปัจจัยกระตุ้นกลับมีจำนวนมากขึ้น จากร้อยละ 18.9 ในปี 2550 เป็นร้อยละ 21.6 ในปี 2551 ซึ่งเป็นข่าวการฆ่ากัน



ตายในครอบครัวมากที่สุดถึงร้อยละ 53.3 รองลงมา คือ ข่าวกการฆ่าตัวตายและการทำร้ายกัน ตามลำดับ (ศูนย์วิจัยปัญหาสุรา, 2552) และข้อมูลจากสำนักงานสถิติแห่งชาติ ระบุว่า ประชากรอายุ 15 ปีขึ้นไป จำนวน 53.9 ล้านคน เป็นผู้ดื่มสุรา 17.0 ล้านคน (ร้อยละ 31.5) โดยผู้ชายมีอัตราการดื่มสูงกว่าผู้หญิง ประมาณ 5 เท่า และกลุ่มวัยทำงาน (25-59 ปี) มีอัตราการดื่มสูงกว่ากลุ่มอื่น คือร้อยละ 37.3 และมีข้อสังเกตว่ากลุ่มผู้สูงอายุของไทยยังมีการดื่มสุราสูงถึงร้อยละ 16.6 สำหรับอายุเฉลี่ยที่เริ่มดื่มสุรา โดยรวมจะเริ่มดื่มที่อายุ 20.3 ปี ผู้ชายจะเริ่มดื่มเร็วกว่าผู้หญิง คือ อายุ 19.4 ปี และ 24.9 ปี อายุเฉลี่ยของผู้เริ่มดื่มสุรา 20.3 ปี โดยกลุ่มเยาวชน (15-24 ปี) มีอายุเฉลี่ย ที่เริ่มดื่มสุรา 17.2 ปี ในขณะที่กลุ่มวัยทำงาน (25-59 ปี) มีอายุเฉลี่ยที่เริ่มดื่มสุรา 20.6 ปี (สำนักงานสถิติแห่งชาติ, 2554)

จากการสำรวจระดับชาติ 2551 ของกรมสุขภาพจิต พบว่า ความชุกของ ความผิดปกติพฤติกรรมดื่มสุรา (Alcohol Use Disorders) ของประชากรไทยอายุ 15-59 ปีสูงถึงร้อยละ 10.9 (95%CI, 10.1-11.7) หรือประมาณ 5.3 ล้านคน โดยผู้ชายพบมากกว่าผู้หญิงเกือบ 9 เท่า ประชากรในภาคเหนือพบความชุกของความผิดปกติพฤติกรรมดื่มสุราสูงถึงร้อยละ 13.3 รองลงมาเป็นภาคตะวันออกเฉียงเหนือ (ร้อยละ 11.5) ส่วนภาคใต้มีความชุกของความผิดปกติของพฤติกรรมดื่มสุราต่ำที่สุดคือ ร้อยละ 8.8 กลุ่มผู้ที่มีความผิดปกติของพฤติกรรมดื่มสุรามีอายุระหว่าง 25 - 34 ปี มากที่สุดคือ ร้อยละ 32 ขณะที่รองลงมาคือ อายุระหว่าง 15 - 24 ปี คือ ร้อยละ 23 ความผิดปกติ ของพฤติกรรมดื่มสุรา (Alcohol Use Disorders) ประกอบด้วย 2 กลุ่มโรค ได้แก่ ภาวะดื่มสุรา แบบอันตราย (Alcohol Abuse) และ ภาวะติดสุรา (Alcohol Dependence) จากผลการสำรวจพบว่า ความชุกของภาวะการดื่มสุราแบบอันตราย (Alcohol Abuse) ร้อยละ 4.2 (95%CI, 3.8-4.8) หรือประมาณ 2.1 ล้านคน และความชุกของภาวะติดสุรา (Alcohol dependence) เป็นร้อยละ 6.6 (95%CI, 6.0-7.2) หรือประมาณ 3.2 ล้านคน โดยความชุกของทั้งภาวะการดื่มสุราแบบอันตราย (Alcohol Abuse) และ ภาวะติดสุรา (Alcohol Dependence) พบในผู้ชายมากกว่าผู้หญิง และ อายุเฉลี่ยของภาวะการดื่มสุราแบบอันตราย (Alcohol Abuse) ต่ำกว่า ภาวะติดสุรา (Alcohol Dependence) ($32.25 + 0.62$ และ $34.97 + 0.51$ ปี ตามลำดับ) เมื่อวิเคราะห์ในรายละเอียดของสถานภาพสมรส พบว่า ส่วนใหญ่สถานภาพสมรสคู่มือโอกาสเสี่ยงสูงต่อการดื่มแบบผิดปกติสอดคล้องกัน ทั้งสองปัญหา (คือ ร้อยละ 55 ของผู้ที่ได้รับการประเมินว่าเป็น ภาวะการดื่มสุราแบบอันตราย (Alcohol Abuse) และร้อยละ 53 ของผู้ที่ได้รับการประเมินว่าเป็น ภาวะติดสุรา (Alcohol Dependence) มีสถานภาพสมรส) เมื่อเปรียบเทียบระหว่างระดับการศึกษาพบว่ากลุ่ม ภาวะติดสุรา (Alcohol Dependence) มีการศึกษาระดับประถมศึกษาหรือต่ำกว่าคิดเป็นร้อยละ 53 ขณะที่มีการศึกษามัธยมศึกษาถึงอนุปริญญา ร้อยละ 47 ในขณะที่กลุ่มภาวะการดื่มสุราแบบอันตราย (Alcohol Abuse) มีระดับการศึกษามัธยมศึกษาถึงอนุปริญญา มากกว่ากลุ่มการศึกษาระดับประถมศึกษาหรือต่ำกว่า คือ ร้อยละ 55 และร้อยละ 43 ตามลำดับ ส่วนกลุ่มอาชีพสามลำดับแรกของทั้งสองภาวะเหมือนกัน คือ รับจ้าง เกษตรกรรม และ ค่าขายธุรกิจ คิดเป็นร้อยละ 45, 26 และ 12 ของภาวะดื่มสุราแบบอันตราย (Alcohol Abuse) และ คิดเป็นร้อยละ 47, 24 และ 12 ของภาวะติดสุรา (Alcohol Dependence) ตามลำดับ ที่น่าสนใจ คือ กลุ่มนักเรียนนักศึกษาที่ถือว่าเป็นเยาวชน พบความชุกถึงร้อยละ 5 ของกลุ่ม Alcohol Dependence หรือประมาณ 1.3 แสนคน (ศูนย์วิจัยปัญหาสุรา, 2553)

จากการศึกษาเรื่องอุบัติการณ์การดื่มเครื่องดื่มแอลกอฮอล์ของเด็กและเยาวชนในสถานศึกษา ภาคตะวันออกเฉียงเหนือของนิรุจน์ อุทธา และคณะ พบว่า อุบัติการณ์การดื่มเครื่องดื่มแอลกอฮอล์



ของเด็กและเยาวชนในสถานศึกษาภาคตะวันออกเฉียงเหนือ มีถึงร้อยละ 24.5 ในการนี้เป็นเพศชาย ร้อยละ 33.6 เพศหญิง ร้อยละ 15.9 เป็นนักเรียนสายสามัญ ร้อยละ 14.3 และเป็นนักศึกษาสายอาชีพ ร้อยละ 39.6 อาศัยอยู่ในเขตเมือง ร้อยละ 26.0 เขตชนบท ร้อยละ 22.4 แยกเป็นรายจังหวัด คือ สกลนคร ร้อยละ 44.3 ขอนแก่น ร้อยละ 32.3 ยโสธร ร้อยละ 27.9 หนองคาย ร้อยละ 27.9 ชัยภูมิ ร้อยละ 26.0 นครราชสีมา ร้อยละ 25.2 นครพนม ร้อยละ 25.0 สุรินทร์ ร้อยละ 22.5 ศรีสะเกษ ร้อยละ 22.0 ร้อยเอ็ด ร้อยละ 21.9 มหาสารคาม ร้อยละ 21.8 อุบลราชธานี ร้อยละ 20.0 กาฬสินธุ์ ร้อยละ 17.8 และอำนาจเจริญ ร้อยละ 17.6 ตามลำดับ สำหรับแบบแผนการตีมนั้น คือ เคยตีเครื่องตีแอลกอฮอล์ ร้อยละ 43.3 อายุครั้งแรกที่เคยตี คือ 15 ปี และเคยมีประสบการณ์ซื้อเครื่องตีแอลกอฮอล์ เมื่ออายุน้อยกว่า 18 ปี ถึงร้อยละ 52.4 พบว่า บุคคลที่มักจะใช้ไปซื้อได้แก่ ญาติผู้ใหญ่ ร้อยละ 34.9 และบิดา ร้อยละ 11.6 อีกทั้งสถานที่ไปซื้อครั้งแรกคือ ร้านขายของชำใกล้บ้าน ร้อยละ 66.9 และรู้สึกเฉย ๆ เมื่อไปซื้อครั้งแรก ร้อยละ 53.7 (นิรุจน์ อุทธา และคณะ, 2548)

จังหวัดหนองคายซึ่งเป็น 1 ใน 20 จังหวัดของภาคตะวันออกเฉียงเหนือมีพฤติกรรมการบริโภคสุราของประชาชนเป็นไปในทิศทางเดียวกันที่กล่าวข้างต้นและมีปริมาณเพิ่มมากขึ้น สังเกตได้จากรายงานสถานการณ์การบริโภคเครื่องดื่มแอลกอฮอล์รายจังหวัด พ.ศ. 2550 พบว่า จังหวัดหนองคาย ความชุกของผู้ดื่มเครื่องดื่มแอลกอฮอล์ อายุ 20 ปีขึ้นไป มีร้อยละ 37.4 ความชุกของผู้ดื่มเครื่องดื่มแอลกอฮอล์ อายุ 15-19 ปี มีร้อยละ 37.4 โดยที่ความชุกของผู้ดื่มเครื่องดื่มแอลกอฮอล์ในรอบ 12 เดือน ช่วงอายุ 20 -40 ปี มีร้อยละ 34.6 ของประชากรทั้งหมด โดยเพศชายมีส่วนที่มากกว่าเพศหญิง ส่วนประเภทเครื่องดื่มที่นิยมดื่ม 5 อันดับแรก คือ สุราขาว/สุรากลั่นชุมชน ร้อยละ 68.83 รองลงมา คือ เบียร์ ร้อยละ 38.14 สุรายี่ห้อไทย ร้อยละ 11.26 สุรายี่ห้อต่างประเทศ ร้อยละ 1.72 ยาดองเหล้า ร้อยละ 0.77 ตามลำดับ และค่าใช้จ่ายในการซื้อเครื่องดื่มแอลกอฮอล์โดยเฉลี่ยอยู่ที่ 236 บาทต่อเดือน (ศูนย์วิจัยปัญหาสุรา, 2550) และจากมูลค่าของรายได้จากการจัดเก็บเงินภาษีในส่วนของภาษีสุราและเบียร์ ที่จัดเก็บในพื้นที่จังหวัดหนองคาย มีจำนวนเพิ่มขึ้นอย่างต่อเนื่อง โดยในปี พ.ศ.2550 จัดเก็บได้ 893.487 ล้านบาท ปี พ.ศ.2551 จัดเก็บได้ 1,191.334 ล้านบาท และปี พ.ศ.2552 จัดเก็บได้ 1,368.949 ล้านบาท (สำนักงานสรรพสามิตพื้นที่หนองคาย, 2553)

สำนักงานสาธารณสุขจังหวัดหนองคาย ได้ทำการศึกษาเก็บข้อมูลเกี่ยวกับสถานการณ์และพฤติกรรมการบริโภคเครื่องดื่มแอลกอฮอล์ของประชาชนจังหวัดหนองคาย พบว่า ประชาชนอายุ 15-74 ปีขึ้นไป ร้อยละ 18.3 (104,130 คน) ดื่มสุราหรือเครื่องดื่มมีแอลกอฮอล์เป็นประจำ และร้อยละ 4.97 (29,387 คน) ดื่มนาน ๆ ครั้ง ส่วนประเภทของเครื่องดื่มแอลกอฮอล์ที่ประชากรนิยมดื่มมากที่สุด ได้แก่ เบียร์และสุราขาว คิดเป็นร้อยละ 33.0 และ 32.3 ตามลำดับ ขณะที่ผลสำรวจจากนักเรียนมัธยมศึกษาในหนองคาย พบว่า นักเรียนชายและหญิงที่ดื่มสุราซึ่งมีอายุ 14-16 ปี และระบุว่าเริ่มดื่มตั้งแต่อายุระหว่าง 13-14 ปี ด้วยอัตราส่วนสูงที่สุดประมาณร้อยละ 44 และ 46 ตามลำดับ ส่วนนักเรียนชายและหญิงที่อายุ 17-20 ปี ระบุว่าเริ่มดื่มเมื่ออายุ 15-16 ปี สูงสุดด้วยอัตราใกล้เคียงกันคือประมาณร้อยละ 43 และ 40 ตามลำดับ จึงมีแนวโน้มว่านักเรียนเริ่มหัดดื่มเครื่องดื่มแอลกอฮอล์เมื่ออายุน้อยลงเรื่อย ๆ (สำนักงานสาธารณสุขจังหวัดหนองคาย, 2553)

จากข้อมูลที่กล่าวมาในข้างต้น อาจก่อให้เกิดปัญหาที่มีสาเหตุมาจากการบริโภคสุราตามมาในอนาคตไม่ว่าจะเป็นปัญหาด้านสุขภาพของผู้ดื่มเองหรือปัญหาด้านสังคม เศรษฐกิจ อาชญากรรม ฯลฯ ทำให้ชุมชนหรือจังหวัดไม่สามารถพัฒนาได้อย่างเต็มศักยภาพของประชาชนผู้วิจิตรจึงมีความสนใจศึกษา



ถึงการดื่มสุราของประชาชน เพื่อพยากรณ์กลุ่มของการดื่มสุราของประชากรในจังหวัดหนองคาย เพื่อนำผลการศึกษาที่ได้ เสนอต่อผู้กำกับงานด้านการดูแลและควบคุม การวางแผน และจัดกิจกรรมต่างๆ ในการลดการดื่มสุราในจังหวัดหนองคาย

1.2 ความมุ่งหมายของการวิจัย

1.2.1 สร้างรูปแบบที่เหมาะสมในการพยากรณ์กลุ่มของการดื่มสุราของประชากรในจังหวัดหนองคาย

1.2.2 เพื่อเปรียบเทียบประสิทธิภาพการพยากรณ์ของเทคนิคการทำเหมืองข้อมูลและเทคนิคการถดถอยโลจิสติก

1.3 ความสำคัญของการวิจัย

จากการวิจัยครั้งนี้ ทำให้ทราบ รูปแบบและเทคนิคที่เหมาะสมในการพยากรณ์กลุ่มของการดื่มสุราของประชากรในจังหวัดหนองคาย เพื่อใช้เป็นแนวทางให้หน่วยงานหรือผู้ที่เกี่ยวข้องนำไปใช้ในการแก้ไขปัญหาที่เกิดจากการดื่มสุราต่อไป

1.4 ขอบเขตของการวิจัย

1.4.1 การวิจัยครั้งนี้ใช้ข้อมูลทุติยภูมิ (Secondary Data) ที่ได้จากการสำรวจพฤติกรรม การสูบบุหรี่และดื่มสุราของประชากร พ.ศ. 2554 สำรวจโดยสำนักงานสถิติแห่งชาติ โดยเป็นบุคคลที่มีอายุ 15 ปีขึ้นไป จำนวน 605 ราย

1.4.2 เทคนิคการวิเคราะห์ข้อมูล 2 เทคนิค คือ

1.4.2.1 เทคนิคการทำเหมืองข้อมูล คือ การวิเคราะห์ต้นไม้การตัดสินใจโดยใช้อัลกอริทึม C4.5 และโครงข่ายประสาทเทียมการเรียนรู้แบบแพร่กลับ (Back-propagation Artificial Neural Network)

1.4.2.2 เทคนิคการถดถอยโลจิสติก คือ ใช้วิธีการวิเคราะห์การถดถอยโลจิสติกแบบหลายกลุ่ม (Multinomial Logistic Regression Analysis)

1.5 นิยามศัพท์เฉพาะ

1.5.1 ประชาชน หมายถึง บุคคลที่มีอายุตั้งแต่ 15 ปีขึ้นไป ทั้งชายและหญิง ที่อาศัยอยู่ในจังหวัดหนองคาย

1.5.2 สุราหรือเครื่องดื่มแอลกอฮอล์ หมายถึง เครื่องดื่มที่มีส่วนผสมของแอลกอฮอล์ เช่น สุราแม่โขง แสงโสม หงษ์ทอง สุรา 28 ดิกรี สุรา 40 ดิกรี สุราขาว กระแช่ สาโท บรั่นดี ไวน์ วิสกี้ เบียร์ เป็นต้น



1.5.3 ระดับการศึกษาสูงสุดที่เรียนจบ หมายถึง ระดับการศึกษาสูงสุดที่สำเร็จการศึกษา (ไม่รวมวิชาชีพพระยาชั้นที่ไม่สอนวิชาสามัญ เช่น ตัดผม ตัดเสื้อ ซ่อมวิทยุ เป็นต้น)

1.5.4 สถานภาพสมรส หมายถึง ความผูกพันระหว่างชายกับหญิงในการเป็นสามีภรรยา

1.5.5 อาชีพ หมายถึง ประเภทหรือชนิดของงานที่บุคคลนั้นทำอยู่

1.5.6 สถานภาพการทำงาน หมายถึง สถานะของบุคคลที่ทำงานในสถานที่ทำงานหรือธุรกิจ

1.5.7 รายได้ หมายถึง รายได้สุทธิโดยเฉลี่ยต่อเดือน

1.5.8 ปริมาณการดื่ม หมายถึง ปริมาณการดื่มสุราหรือเครื่องดื่มแอลกอฮอล์ ในระหว่าง 12 เดือน (ระหว่างเดือนมีนาคม 2553-กุมภาพันธ์ 2554)

1.5.9 คะแนนผลกระทบปัญหาสุขภาพที่มีผลต่อการทำงาน หมายถึง คะแนนที่ผู้ตอบ สัมภาษณ์ เห็นว่ามีผลกระทบต่องาน

1.5.10 คะแนนผลกระทบปัญหาสุขภาพที่มีผลต่อการทำกิจวัตรประจำวัน หมายถึง คะแนนที่ ผู้ตอบสัมภาษณ์เห็นว่าผลกระทบต่อการทำกิจวัตรประจำวัน

1.5.11 กลุ่มของการดื่มสุรา หมายถึง ความถี่ของการดื่มเครื่องดื่มแอลกอฮอล์ แบ่งออกเป็น 3 กลุ่ม คือ

1.5.11.1 กลุ่มที่ไม่เคยดื่มสุรา หมายถึง ผู้ที่ไม่เคยดื่มสุราเลยหรือเคยดื่มแต่ไม่ดื่มใน 12 เดือนที่ผ่านมา

1.5.11.2 กลุ่มที่ดื่มเล็กน้อยหรือนาน ๆ ครั้ง หมายถึง ผู้ที่ดื่มสุราตั้งแต่ 1-3 วันต่อเดือน หรือ น้อยกว่า

1.5.11.3 กลุ่มที่ดื่มสม่ำเสมอ หมายถึง ผู้ที่ดื่มสุราทุกวัน ดื่มเกือบทุกวัน (5-6 วันต่อ สัปดาห์) วันเว้นวัน (3-4 วันต่อสัปดาห์) หรือ ดื่มทุกสัปดาห์ (1-2 วันต่อสัปดาห์) (สำนักงานสถิติ แห่งชาติ, 2554)



บทที่ 2

แนวคิด ทฤษฎี และงานวิจัยที่เกี่ยวข้อง

ในการวิเคราะห์การดื่มสุราของประชาชนในจังหวัดหนองคาย โดยเทคนิคเหมืองข้อมูลและการถดถอยโลจิสติก ผู้ศึกษาค้นคว้าได้ศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องโดยนำเสนอเนื้อหาตามลำดับดังนี้

2.1 ความรู้เกี่ยวกับการดื่มสุราหรือเครื่องดื่มแอลกอฮอล์

2.2 การทำเหมืองข้อมูล

2.2.1 การวิเคราะห์ ต้นไม้การตัดสินใจ (Decision Tree)

2.2.2 โครงข่ายประสาทเทียม (Artificial Neural Network)

2.3 การวิเคราะห์ความถดถอยโลจิสติกแบบหลายกลุ่ม (Multinomial Logistic Regression Analysis)

2.4 งานวิจัยที่เกี่ยวข้อง

2.4.1 งานวิจัยภายในประเทศ

2.4.2 งานวิจัยต่างประเทศ

2.1 ความรู้เกี่ยวกับการดื่มสุราหรือเครื่องดื่มแอลกอฮอล์

ความหมายของสุราหรือเหล้า

ราชบัณฑิตยสถาน (2542) ได้ให้ความหมายของคำว่า “สุรา” มาจากภาษาบาลีและสันสกฤต แปลว่า เหล้าหรือน้ำเมาที่กลั่นแล้วหรือหมักแล้ว (ราชบัณฑิตยสถาน, 2542)

แอลกอฮอล์ คือ สารอินทรีย์ชนิดหนึ่ง ลักษณะเป็นของเหลวใส กลิ่นฉุน ระเหยง่าย มีจุดเดือด 78.6 องศาเซลเซียส ชื่อเต็ม คือ เอธิลแอลกอฮอล์ (Ethyl Alcohol) แต่มักเรียกสั้น ๆ ว่าเป็นแอลกอฮอล์ (Alcohol) โดยปกติมักเกิดจากการหมักสารประเภทแป้ง หรือน้ำตาลผสมยีสต์ ซึ่งมักเรียกว่าแป้งเชื้อหรือเชื้อหมัก เป็นองค์ประกอบสำคัญของสุรา และเมรัยทุกชนิด เมื่อดื่มเข้าไปจะเกิดอาการมึนเมา(ราชบัณฑิตยสถาน, 2542)

คำนิยามของสุรา ตามพระราชบัญญัติสุรา พ.ศ. 2493

สุรา หมายความว่า วัตถุทั้งหลายหรือของผสมที่มีแอลกอฮอล์ ดื่มกินได้เช่นเดียวกับน้ำสุรา หรือซึ่งดื่มกินไม่ได้แต่เมื่อได้ผสมกับน้ำหรือของเหลวอย่างอื่นแล้ว สามารถดื่มกินได้เช่นเดียวกับสุรา

สุราแช่ หมายความว่า สุราที่ไม่ได้กลั่น และให้หมายรวมถึงสุราแช่ที่ได้ผสมกับสุรากลั่นแล้ว แต่ยังมี ความแรงแอลกอฮอล์ไม่เกินสิบห้าดีกรีด้วย

สุรากลั่น หมายความว่า สุราที่ได้กลั่นแล้ว และให้หมายความว่าสุรากลั่นที่ได้ผสมกับสุราแช่แล้ว แต่ยังมี ความแรงแอลกอฮอล์เกินกว่าสิบห้าดีกรีด้วย

สุรา หรือเหล้า หรือแอลกอฮอล์ หมายถึง สารเสพติดชนิดหนึ่งที่เกิดจากการหมักผลไม้ ผักหรือเมล็ดพืชชนิดต่างๆ จนได้เป็นเครื่องดื่มที่สามารถใช้ดื่มได้และออกฤทธิ์ทำให้เกิดอาการมึนเมา



ถ้าดื่มอย่างรวดเร็วในปริมาณมากๆ จะไปกดทางเดินหายใจทำให้เข้าขั้นโคม่าและเสียชีวิตได้ในที่สุด (ศิริภรณ์ ชัยศิริ, 2551)

สุรา (liquor or spirit) หมายถึง น้ำเมาที่ได้จากการกลั่นสารบางประเภท อาทิ เอทิลแอลกอฮอล์ และเมรัย คือ น้ำเมาที่เกิดจากการหมักหรือแช่ให้เกิดสารบางประเภท เมื่อดื่มแล้ว สารนั้นจะออกฤทธิ์กดระบบประสาทส่วนกลาง หากดื่มไม่มากอาจรู้สึกผ่อนคลายเนื่องจากสารกดจิตได้ สำนึกที่คอยควบคุมตนเองทำให้กล้าแสดงออกมากขึ้น แต่เมื่อดื่มมากขึ้นก็จะกดสมองบริเวณอื่น ๆ ทำให้เสียการทรงตัว พูดไม่ชัด จนแม้กระทั่งหมดสติในที่สุด ทั้งสุราและเมรัยเรียกโดยภาษาปากว่า “เหล้า” (วิกิพีเดีย สารานุกรมเสรี, 2555)

องค์การอนามัยโลก (WHO, 2001) ได้ให้ความหมายของเครื่องดื่มแอลกอฮอล์ว่าเป็น ของเหลว ที่มีส่วนผสมของแอลกอฮอล์ชนิดดื่มหรือ เอทานอลใช้เพื่อการดื่ม ซึ่งส่วนใหญ่ได้มาจากการหมัก หรือได้จากการหมักแล้วกลั่น

สุรา หมายถึง เหล้า เบียร์ สุราขาว เหล้าปั่น สุราพื้นบ้าน เช่น สาโท สบาย ไวน์ เหล้ายา ดอง เหล้าขาวผสมน้ำแดง (สุทธิรักษ์ ไชยรักษ์ และคณะ, 2553)

จากที่กล่าวมาสรุปได้ว่า สุราหรือเครื่องดื่มแอลกอฮอล์ เป็นของเหลวชนิดหนึ่งที่มีส่วนผสมของแอลกอฮอล์หรือเอทานอล เกิดจากการหมักหรือกลั่น มีกลิ่นฉุน ดื่มแล้วทำให้เกิดอาการมึนเมา ส่งผลกระทบต่อร่างกายและพฤติกรรมที่แสดงออก

2.1.1 ประเภทของเครื่องดื่มแอลกอฮอล์

ตามพระราชบัญญัติภาษีสุรา พ.ศ. 2493 แบ่งประเภทของเครื่องดื่มแอลกอฮอล์ตามการจัดเก็บภาษีเครื่องดื่มแอลกอฮอล์ของกรมสรรพสามิต เพื่อใช้ในการจัดเก็บภาษีเครื่องดื่มแอลกอฮอล์ แบ่งออกเป็น 2 ประเภท ใหญ่ ๆ คือ (ไพรัตน์ อันอินทร์, 2552)

2.1.1.1 สุรากลิ่น คือ เครื่องดื่มแอลกอฮอล์ที่เกิดจากการเอาน้ำสา (แป้งหรือน้ำตาลหมักด้วยยีสต์จนเกิดมีแอลกอฮอล์) มากลิ่นอาจตามด้วยการเติมสี ปรงกลิ่น แต่งรส ด้วยสารปรุงแต่งอื่นๆ ลงไป ดิกรีแอลกอฮอล์หรือเก็บบ่มต่อไป เพื่อให้ถูกใจผู้บริโภคสุรากลิ่นของไทยส่วนมากที่มีจำหน่าย คือ สุราขาว (เหล้าโรง) สุราผสม สุราผสมพิเศษ สุราปรุงพิเศษ (เหล้าเหลียง) เครื่องดื่มแอลกอฮอล์ไทยนิยมใช้สารปรุงแต่ง เช่น น้ำเชื้อ (Essene หรือ Flavor) น้ำยาสกัด (Extract) และสีผสมลงไป

2.1.1.2 สุราไม่กลั่น (สุราแช่) คือ เครื่องดื่มแอลกอฮอล์ที่ได้จากการหมักสาให้เกิดเป็นน้ำเมาแต่ไม่มีการกลั่นมักจะหมักสาจากเมล็ดพืช แป้งจากพืช ผลไม้ และน้ำตาลจากพืช เช่น เบียร์ไวน์ กระแช่ น้ำขาวน้ำตาลเมา เป็นต้น สุราไม่กลั่นของไทยที่มีจำหน่ายส่วนใหญ่ คือ เบียร์ ไวน์ เบียร์ไทยจะมีดิกรีแอลกอฮอล์สูงและแรงกว่าเบียร์ต่างประเทศ ส่วนน้ำขาว กระแช่ น้ำตาลเมาชาวบ้านมักจะทำขึ้นเอง ซึ่งผิดกฎหมายจึงไม่มีจำหน่าย เหตุที่ชาวบ้านทำขึ้นเองนั้น อันเนื่องมาจากทำจากวัตถุดิบพืชธรรมชาติจริงๆ เช่น ข้าว น้ำตาลสดจากต้นตาล น้ำตาลจากยอดมะพร้าว ผลที่ได้คือ รสชาติอร่อยหวาน บริโภคแล้วรู้สึกสบายไม่อึดอัด หรือไม่มีอาการเมาค้างอย่างเครื่องดื่มแอลกอฮอล์ทั่วไป

2.1.2 ชนิดของเครื่องดื่มแอลกอฮอล์

ชนิดของเครื่องดื่มที่มีส่วนผสมแอลกอฮอล์ที่มีจำหน่ายในประเทศไทย แบ่งชนิดของเครื่องดื่มที่มีส่วนผสมของแอลกอฮอล์ออกเป็น ดังนี้ (สำนักงานสำรวจสุขภาพประชาชนไทย, 2551)



2.1.2.1 น้ำผลไม้ผสมแอลกอฮอล์ จัดเป็นเครื่องดื่มที่มีแอลกอฮอล์ จัดอยู่ในจำพวก ค็อกเทล ฟันช์ ซึ่งในปัจจุบันได้มีวางขายตามท้องตลาดทั่วไปซึ่งมีหลายยี่ห้อ ได้แก่ สเปย์ไวน์คูเลอร์ Cruiser, Nite, Bacardi, Smirnoff Premium ซึ่งเครื่องดื่มแอลกอฮอล์ที่กล่าวมานี้จะมีปริมาณในการ บรรจุ เท่ากับ 300 ml 5 ดิกรี นอกจากนี้ก็ยังมียี่ห้อ Smirnoff Black มีปริมาณบรรจุเท่ากับ 300 ml เช่นกันแต่จะมีดีกรีมากกว่าชนิดอื่นๆ คือ 7 ดิกรี

2.1.2.2 ไวน์ (Wine) ไวน์สามารถแบ่งออกได้เป็น 4 ชนิด ตามปริมาณของแอลกอฮอล์ ความหวาน คาร์บอนไดออกไซด์ สี และการปรุงแต่ง คือ

1) ไวน์ตั้งโต๊ะ (Table Wines, Still Wines หรือ Natural Wines) ทำมาจากการ หมักน้ำองุ่นโดยเป็นหมักตามธรรมชาติ บางครั้งอาจมีการเติมน้ำตาลหรือยีสต์ด้วยก็ได้ สามารถแบ่งไวน์ ตามสีของไวน์ได้ 3 ชนิด คือ สีเหลืองหรือสีทอง(White) สีแดงเข้มถึงม่วง(Red) และสีชมพูอ่อน (Rose) ซึ่งมีรสชาติหลากหลายตั้งแต่หวาน เช่น Aleatico และไม่หวานเรียกว่า Dry ไม่ขม เช่น Gamay Grenache, Grignolino และ Tavel ปริมาณแอลกอฮอล์ไวน์มีตั้งแต่ ร้อยละ 7-15 ส่วนมากจะมีปริมาณ แอลกอฮอล์อยู่ประมาณร้อยละ 12 ไวน์ตั้งโต๊ะ (Table Wines) ส่วนใหญ่ได้แก่ เหล้าองุ่น

2) ไวน์แรง (Fortified Wines, Dessert Wines หรือ Aperitif Wines) เป็นเหล้า องุ่นที่มีปริมาณแอลกอฮอล์ร้อยละ 17-23 โดยมีการเติมแอลกอฮอล์ระหว่างการผลิต มักเติมบรันดีที่ทำ จากองุ่น มีสีต่างๆ เช่น ขาวอำพัน แดงเจิดจ้า และแดงคล้ำ ไวน์ชนิดนี้นิยมดื่มก่อนหรือหลังรับประทานอาหาร ไวน์ที่นิยมดื่มกันมากที่สุด คือ Sherry ซึ่งเป็นไวน์ชนิดแรง มีคุณสมบัติที่เป็นเอกลักษณ์เฉพาะ จากแหล่งที่ปลูกและพันธุ์องุ่น รวมทั้งกระบวนการผลิตที่เฉพาะส่วนใหญ่ผลิตในประเทศสเปน มีรสชาติ ตั้งแต่หวานมากจนไปถึงไม่หวาน

3) ไวน์ซ่าหรือเหล้าองุ่นอัดลม (Sparkling Wines) เป็นไวน์ที่มีแก๊ส จึงทำให้มีรสซ่า มีสีต่างๆ ได้แก่ ขาว ชมพูและแดง Sparkling Wines ใช้กรรมวิธีการหมักไวน์ซ่าเป็นครั้งที่สองภายใน ขวดและเก็บรักษาแก๊สนี้ไว้ จึงทำให้เกิดรสซ่า ซึ่งเป็นที่นิยมกันมาก มีการจดลิขสิทธิ์ไวน์ชื่อ “Champagne” ของฝรั่งเศส ไวน์ที่ผลิตด้วยกรรมวิธีคล้ายคลึงกันจะใช้คำว่า “Sparkling Wines” แชมเปญนิยมดื่มเพื่อแสดงความยินดีต่อกัน มักจะเสิร์ฟโดยแช่เย็นจัด

4) ไวน์ปรุงแต่ง (Flavored Wines) มีหลากหลายชนิด บางชนิดผสมโดยผู้บริโภค เองและบางชนิดผสมโดยผู้ผลิต เช่น Glogg เป็นพันธุ์รสเผ็ดผลิตในประเทศสวีเดน ซึ่งมักจะใช้ไวน์แดง เครื่องเทศอัลมอนด์และลูกเกดผสมกัน เวอร์มูท (Vermouth) เป็นไวน์ไม่มีสีแดงมีการเติมสมุนไพร เพื่อให้รสชาติและกลิ่นที่เปลี่ยนไป ส่วน Aperitifs หมายถึง เหล้าองุ่นที่ดื่มก่อนอาหารจะมีการเติม ควินินและส่วนผสมอื่นๆ เพื่อทำให้รสชาติเข้มข้น และมักมีรสหวาน เช่น ดูปอนเนต (Doubonnet) ของ ฝรั่งเศสและแคมพารี (Campari) ของอิตาลี แอลกอฮอล์ที่ดื่มหลังอาหาร เช่นว่า Digestifs อาจใช้ Sherry ก็ได้ แต่โดยทั่วไปนิยมใช้บรันดีหรือ Liqueurs มากกว่า

2.1.2.3 เบียร์ คือ เครื่องดื่มที่ได้จากการหมักข้าวมอลท์ (Malt) มีแรงแอลกอฮอล์ ประมาณ 4-6 ดิกรี เบียร์ที่มีจำหน่ายในประเทศไทยมีหลายยี่ห้อ เช่น ลีโอ เรดฮอสส์ อาซา ช้าง ไทเกอร์ สิงห์ ไฮเนเก้น

2.1.2.4 เหล้าแดง (Whisky) หรือ สุรากลั่นที่ทำจากข้าวชนิดใดชนิดหนึ่ง หรือ หลากหลายชนิด ก็ได้ โดยนำมาหมักแล้วกลั่นให้มีดีกรีสูงขึ้น จากนั้นนำไปเก็บบ่มในถังไม้โอ๊กเพื่อให้ได้สี กลิ่น รสที่ดีขึ้น และก่อนจะนำมาบรรจุขวด บางชนิดอาจมีการปรุงแต่งสี กลิ่น รสชาติอีกครั้งหนึ่ง โดยมี



ปริมาณแอลกอฮอล์ประมาณ 40-50 ดีกรี เหล้าแดงที่มีจำหน่ายในประเทศไทย เช่น แม่โขง มังกรทอง หงส์ทอง แสงโสม Crown99 Mater Blend

2.1.2.5 เหล้าขาว หรือสุราขาว คือสุรากลั่นที่ได้จากน้ำสำ ที่หมักด้วยวัตถุดิบประเภท แป้ง น้ำตาล และ/หรือกากน้ำตาล ที่มีเปอร์เซ็นต์แอลกอฮอล์ไม่เกิน 80 % (ดีกรี) ในการบรรจุจำหน่าย ต้องมีการปรุงแต่งให้มีเปอร์เซ็นต์แอลกอฮอล์ไม่เกิน 40 %(ดีกรี) สุราขาวที่มีจำหน่ายในประเทศไทย เช่น เหล้าขาวเสือด้า เขียวเงิน Varintip

2.1.2.6 บรันดี (Brandy) บรันดีเป็นเหล้าที่นิยมกันมาก ได้จากการหมักองุ่นให้เป็นไวน์ แล้วจึงนำมากลั่นเป็นบรันดีอีกที จากนั้นนำไปเก็บบ่มให้ได้ สี กลิ่น รสชาติที่ดี บรันดีที่มีขายตาม ห้างตลาด สามารถแบ่งได้เป็น 3 ประเภท คือ

1) บรันดีพื้นเมือง (Domestic Brandy) เช่น Regency Brandy, German Brandy

2) บรันดีมาตรฐาน (Regular Brandy) ส่วนมากเป็นบรันดีที่นำเข้ามาจาก ต่างประเทศ

3) บรันดีเกรดสูง (Premium Brandy) เป็นบรันดีราคาแพงที่เก็บบ่มไว้นานถึงไม่โอ๊ค เป็นเวลานาน โดยระบุคุณภาพเป็นอักษรย่อ หรือชื่อพิเศษ เช่น คอนยัค (Cognac) หรือ Armagnac

2.1.2.7 น้ำตาลเมา เป็นเครื่องดื่มที่มีแอลกอฮอล์โดยผ่านกระบวนการหมักน้ำตาลที่ได้ จากการปาดจั่นมะพร้าว จั่นต้นตาล งวงตาล งวงต้นระกำ งวงสละ และงวงน้ำอ้อย โดยมีการเติม สมุนไพรที่มีคุณสมบัติเป็นแทนนิน เช่น เปลือกไม้เคี่ยม พะยอม ตะเคียน เปลือกมะเกลือ ลงไปใน กระบองน้ำตาล มีแอลกอฮอล์ประมาณ 3-4 ดีกรี

2.1.2.8 กระแช่ คือ การเปลี่ยนน้ำตาลในน้ำตาลเมาให้เป็นแอลกอฮอล์ด้วยการใช้เชื้อ ยีสต์จากเชื้อหมัก หรือที่เรียกว่า “แป้งใหญ่” ทำหน้าที่เปลี่ยนน้ำตาลให้เป็นแอลกอฮอล์และเปลี่ยน น้ำตาลเมาให้เป็นกระแช่

2.1.2.9 สาโท เป็นเหล้าแช่ประเภทไวน์ ทำจากข้าวเหนียวเปล่าหมักสำจนได้ปริมาณ แอลกอฮอล์ประมาณ 7-8 ดีกรี สาโทที่มีจำหน่ายในท้องตลาดทั่วไป ได้แก่ เจ้าพระยาสาโท สุราแช่ พื้นเมือง กรูปรี สาโท สาโทซ่า มีปริมาณบรรจุ 640 ml 6.7 ดีกรี

2.1.2.10 น้ำขาว อุ เป็นน้ำเมาชนิดหนึ่งที่ใช้ข้าวเหนียวหรือปลายข้าวเหนียว นำข้าว เหนียวแช่น้ำสะอาด พร้อมนำเกลบิ้งให้สุก ใส่กระดังฝิ่งให้แห้งทิ้งไว้ให้เย็น บดลูกแป้งให้ละเอียดพรม ลงบน ข้าวเหนียวให้ทั่ว ใส่กระบุงหมักให้เกิดแอลกอฮอล์หนึ่งคืน นำออกมาฝิ่งให้แห้งก่อนบรรจุลงไห กัดให้แน่น ปิดปากไหด้วยใบตองแห้งกับขี้เถ้าให้สนิท ทิ้งไว้ในอุณหภูมิปกติ ประมาณ 7-15 วัน จะเกิด แอลกอฮอล์ 10 ดีกรี จึงจะรับประทานได้ บรรจุลงไหขนาดบรรจุ 2 ลิตร 3 ลิตร 4 ลิตร และ 5 ลิตร

2.1.2.11 สาเก เป็นเครื่องดื่มประจำชาติของญี่ปุ่น ที่ได้จากการหมักข้าวโดยมีปริมาณ แอลกอฮอล์ประมาณ 15-17 ดีกรี สาเกแบ่งออกเป็น 3 ประเภทใหญ่ๆ คือ

1) Ginjo-shu เป็นสาเกชั้นหนึ่งทั้งในรสชาติและกลิ่น ในกระบวนการหมักจะมี การเติมแอลกอฮอล์ลงในข้าวที่มีหัวเชื้อและน้ำ โดยใช้ข้าวประมาณ 60% หรือต่ำกว่า

2) Junmai-shu เป็นสาเกที่ผลิตจากขบวนการหมักของข้าว, ข้าวที่มีหัวเชื้อและน้ำ โดยใช้ข้าวประมาณ 70% หรือต่ำกว่า



3) Honjozo-shu เป็นสาเกที่ผลิตจากขบวนการหมักจะมีการเติมแอลกอฮอล์ลงไป
ในข้าวและข้าวที่มีหัวเชื้อและน้ำ โดยใช้ประมาณ 70% หรือต่ำกว่า มีขบวนการหมักที่แตกต่างจาก
Ginjo – shu

ตาราง 2.1 ชนิดของเครื่องดื่มแอลกอฮอล์และปริมาณแอลกอฮอล์ (ดีกรี)

ชนิดเครื่องดื่ม	ปริมาณแอลกอฮอล์(ดีกรี)
รัม	60
วิสกีฝรั่ง	42 - 53
บรันดี วิสกี	45
ยีน	44
สุราพิเศษ - แสงโสม	40
วิสกีไทย - ธาราไก่แดง ไก่ขาว กวางทอง	40
แมโขง	35
สุราขาว	35
สุราจีน - เชียงขุน	30
แชมเปญ	16
ไวน์หวาน	15
ไวน์แดง	6 - 14
ไวน์ขาว	8
สุราหมักไทย - น้ำขาว	8 - 10
เบียร์	4 - 6

ที่มา : กฤษฎ์ ขุนลิก (2555)

2.1.3 ความถี่ของการดื่มเครื่องดื่มแอลกอฮอล์

ความถี่ของการดื่มเครื่องดื่มแอลกอฮอล์ หมายถึง จำนวนครั้งของการดื่มเครื่องดื่ม
แอลกอฮอล์ในระยะเวลาหนึ่งๆ เช่น ชุดความถี่ของการดื่มใน 12 เดือนที่มักใช้สอบถามผู้บริโภค ได้แก่
ดื่มทุกวัน, ดื่มเกือบทุกวัน (หรือดื่มวันเว้นวัน หรือ ดื่ม 5 - 6 วัน และ ดื่ม 3 - 4 วันต่อสัปดาห์ หรือ ดื่ม
3 - 6 วันต่อสัปดาห์), ดื่ม 1 - 2 วันต่อสัปดาห์ (หมายถึงดื่มเป็นประจำทุกสัปดาห์), ดื่ม 1 - 3 วันต่อ
เดือน (หมายถึงดื่มเป็นประจำทุกเดือน แต่ไม่ทุกสัปดาห์), ดื่ม 3 - 11 วันต่อปี (หมายถึงดื่มหลายครั้งใน
หนึ่งปี แต่ไม่ได้เป็นประจำทุกเดือน), และ ดื่ม 1 - 2 วันต่อปี (ดื่มน้อยครั้งมากในหนึ่งปี) ชุดความถี่ของ
การดื่มใน 30 วัน ได้แก่ ดื่ม 1 - 2 ครั้ง, ดื่ม 3 - 5 ครั้ง, ดื่ม 6 - 9 ครั้ง, ดื่ม 10 - 19 ครั้ง, และ ดื่ม
มากกว่า 20 ครั้งขึ้นไป หรืออาจเป็นการสอบถามความถี่ของการดื่มแบบคำถามปลายเปิดให้ผู้ตอบ
ตอบแบบอิสระแล้วนักวิจัยนำมาคำนวณหาค่าเฉลี่ยหรือจัดหมวดหมู่ของความถี่เองภายหลัง (ศุณย์วิทย์
ปัญหาสุรา, 2553)



2.1.4 ความเสี่ยงจากการดื่มเครื่องดื่มแอลกอฮอล์

ความเสี่ยงจากการดื่มเครื่องดื่มแอลกอฮอล์ หมายถึง การดื่มเครื่องดื่มแอลกอฮอล์ในระดับที่มีความเป็นไปได้สูงก่อให้เกิดปัญหาจากการดื่มเครื่องดื่มแอลกอฮอล์ ซึ่งมักแบ่งออกเป็น ความเสี่ยงต่อการเกิดปัญหาแบบเฉียบพลัน(High Risk Drinking for Acute Problems) และความเสี่ยงต่อการเกิดปัญหาแบบเรื้อรัง (High Risk Drinking for Chronic Harms) ซึ่งมีวิธีการวัดประเมินระดับความเสี่ยงหลายวิธี เช่น การวัดความเสี่ยงแบบเฉียบพลันนั้น องค์การอนามัยโลก ระบุถึงเกณฑ์วัดที่นิยมใช้สองวิธี ได้แก่ การวัดจำนวนหรือสัดส่วนของผู้ดื่มหรือของประชากรโดยรวมที่มีการดื่มในลักษณะที่เสี่ยง เช่น ดื่มเกินห้าดื่มมาตรฐาน (Standard Drink) ในหนึ่งครั้งที่ดื่มอย่างน้อยหนึ่งครั้งในหนึ่งเดือนที่ผ่านมา หรือ ดื่มอย่างน้อยหกดื่มมาตรฐานในหนึ่งครั้งที่ดื่มอย่างน้อยสัปดาห์ละหนึ่งครั้งในรอบหนึ่งปีที่ผ่านมา เป็นต้น และการเทียบระดับความเสี่ยงจากปริมาณการบริโภคเครื่องดื่มแอลกอฮอล์ โดยวิธีนี้จะต้องสอบถามเพื่อคำนวณวัดปริมาณการบริโภคเครื่องดื่มแอลกอฮอล์เฉลี่ยต่อวันที่ดื่ม (เฉลี่ยเฉพาะวันที่ดื่มเท่านั้น) ก่อน แล้วนำมาเทียบระดับกับมาตรฐานความเสี่ยงแบบเฉียบพลันที่องค์การอนามัยโลกแนะนำไว้ ดังตารางที่ 2.2

ตาราง 2.2 ปริมาณแอลกอฮอล์ที่บริโภคต่อวันที่ดื่มกับระดับความเสี่ยงแบบเฉียบพลัน จำแนกตามเพศ

ระดับความเสี่ยงแบบเฉียบพลัน	เพศชาย	เพศหญิง
-ความเสี่ยงในระดับเริ่มต้น (Low Risk)	1 - 40 กรัมต่อวันดื่ม	1 - 20 กรัมต่อวันดื่ม
-ความเสี่ยงระดับปานกลาง (Medium Risk)	41 - 60 กรัมต่อวันดื่ม	21 - 40 กรัมต่อวันดื่ม
-ความเสี่ยงระดับสูง (High Risk)	61 - 100 กรัมต่อวันดื่ม	41 - 60 กรัมต่อวันดื่ม
-ความเสี่ยงระดับสูงมาก (Very High Risk)	> 101 กรัมต่อวันดื่ม	> 61 กรัมต่อวันดื่ม

ที่มา : International Guide for Monitoring Alcohol Consumption and Related Harm (n.d.)

จากการสำรวจสภาวะสุขภาพอนามัยของประชาชนไทยโดยการตรวจร่างกายครั้งที่ 3 พ.ศ. 2546-47 ใช้ข้อมูลปริมาณการบริโภคเครื่องดื่มแอลกอฮอล์มาใช้จัดกลุ่มความเสี่ยงของการบริโภค โดยใช้ดัชนีสำคัญในการจัดกลุ่มด้วยกันสามดัชนี คือ

1. พฤติกรรมการดื่มประจำ (Regular) หรือดื่มบ่อยๆ ครั้ง (Occasional) โดยใช้เกณฑ์การดื่มมากกว่าหรือเท่ากับหนึ่งครั้งต่อสัปดาห์ขึ้นไปถือเป็นการดื่มประจำ
2. ปริมาณการบริโภคเครื่องดื่มแอลกอฮอล์สะสมที่มากเกินไป (Excessive) ซึ่งถือได้ว่าเป็นการประเมินความเสี่ยงแบบเรื้อรังของการบริโภคเครื่องดื่มแอลกอฮอล์ โดยใช้เกณฑ์ที่ถือได้ว่าเป็นการดื่มปริมาณสะสมที่มากเกินไป คือ ปริมาณการดื่มที่มากกว่า 14 หน่วยดื่มมาตรฐานขึ้นไปในหนึ่งสัปดาห์สำหรับผู้ชาย และมากกว่า 7 หน่วยดื่มมาตรฐานขึ้นไปต่อสัปดาห์สำหรับผู้หญิง (หมายเหตุ - คิดที่ 10 กรัมแอลกอฮอล์บริสุทธิ์ (Ethanol) เป็นหนึ่งหน่วยดื่มมาตรฐาน)
3. ปริมาณการบริโภคเครื่องดื่มแอลกอฮอล์ต่อครั้งที่มากเกินไป หรือ ที่เรียกว่าเม้าห้วนน้ำ (Binge) ซึ่งถือได้ว่าเป็นการประเมินความเสี่ยงแบบเฉียบพลันของการบริโภคเครื่องดื่มแอลกอฮอล์ โดยใช้เกณฑ์ที่ถือเป็นการดื่มต่อครั้งที่มากเกินไป หรือ ดื่มแบบเม้าห้วนน้ำ (Binge) คือ ปริมาณการดื่ม



ตั้งแต่ 5 หน่วยดื่มมาตรฐานต่อวันดื่มขึ้นไปสำหรับผู้ชาย และ ตั้งแต่ 4 หน่วยดื่มมาตรฐานต่อวันดื่มขึ้นไปสำหรับผู้หญิง

กลุ่มความเสี่ยงสามารถแบ่งออกเป็น 6 กลุ่ม ได้แก่

1. ดื่มประจำ มาก และเมา (Regular Excessive with Binge Drinker)
2. ดื่มประจำ มาก แต่ไม่เมา (Regular Excessive without Binge Drinker)
3. ดื่มประจำ ไม่มาก แต่เมา (Regular Non-Excessive with Binge Drinker)
4. ดื่มประจำ ไม่มาก และไม่เมา (Regular Non-Excessive without Binge Drinker

หรือ Regular Moderate Drinker)

5. ดื่มนานๆ ครั้ง แต่เมา (Occasional Binge Drinker) และ
6. ดื่มนานๆ ครั้ง และไม่เมา (Occasional Non-Binge Drinker)

(ศูนย์วิจัยปัญหาสุรา, 2553)

2.1.5 ผลกระทบของเครื่องดื่มแอลกอฮอล์

การดื่มแอลกอฮอล์มีผลกระทบต่อร่างกายคือ แอลกอฮอล์สามารถออกฤทธิ์ได้นานถึง 24-48 ชั่วโมง เมื่อเข้าสู่ร่างกายจะถูกเมตาโบไลต์ ที่ตับ โดยเอนไซม์แอลกอฮอล์ดีไฮโดรจีเนส (Alcohol Dehydrogenase) กลายเป็นอะเซตัลดีไฮด์ (Acetaldehyde) แล้วถูกเปลี่ยนเป็น อะซีเตท (Acetate) โดยเอนไซม์อะเซตัลดีไฮด์ ซึ่งหลังดื่มจะส่งผลกระทบต่อระบบต่างๆ ของร่างกาย ดังต่อไปนี้

2.1.5.1 ผลต่อระบบประสาทสมองส่วนกลาง แอลกอฮอล์มีฤทธิ์กดประสาทส่วนที่ควบคุม การตัดสินใจ สติ การยั้งคิด ทำให้เกิดอาการพูดอ้อแอ้ กล้ามเนื้อทำงานไม่ประสานกันเสียการทรงตัว มีความเคลิบเคลิ้ม การตัดสินใจผิดพลาด ถ้าหากดื่มปริมาณมากจะกดศูนย์ควบคุมการหายใจ และการไหลเวียนเลือด

2.1.5.2 ผลต่อระบบหัวใจ หลอดเลือด และแอลกอฮอล์มีฤทธิ์กดการบีบตัวของกล้ามเนื้อหัวใจทำให้เส้นเลือดขยายตัว ความดันโลหิตต่ำลงเล็กน้อย เส้นเลือดที่ผิวหนังขยายตัว ทำให้มีอาการใบหน้า ลำคอแดง รู้สึกร้อน ถ้าดื่มในปริมาณมาก อุณหภูมิของร่างกายจะลดต่ำลง เพราะเส้นเลือดส่วนปลายขยาย ทำให้ร่างกายสูญเสียความร้อน

2.1.5.3 ผลต่อระบบทางเดินอาหาร แอลกอฮอล์มีฤทธิ์ทำให้มีการหลั่งน้ำย่อยมากขึ้น ส่งผลให้เกิดการระคายเคืองเยื่อทางเดินอาหาร ทำให้เกิดแผลในกระเพาะอาหาร อาเจียน เกิดอาการแทรกซ้อนอื่นๆ เช่น สูญเสียเลือด ขาดโปรตีน วิตามินบี 1 (Thiamine) เกิดอาการชาตามปลายมือปลายเท้า

2.1.5.4 ผลต่อดับ ผู้ที่ดื่มแอลกอฮอล์มานานจะเกิดความผิดปกติของการเผาผลาญที่ตับการผลิตน้ำตาลลดลง ทำให้น้ำตาลในเลือดต่ำ ร่างกายมีกรดคีโตน (Ketoacidosis) และมีไขมันสะสมที่เซลล์พาราเรโนมาของตับ (Liver parenchyma or Fatty liver) ทำให้เกิดอาการตับอักเสบกลายเป็นตับแข็งและมะเร็งตับได้

2.1.5.5 ผลต่อระบบต่อมไร้ท่อ การตกค้างของแอลกอฮอล์ยังมีผลต่อระบบต่อมไร้ท่อทำให้เกิดความไม่สมดุลของเกลือแร่ในร่างกาย ความไม่สมดุลของสเตียรอยด์ฮอร์โมนเกิดอาการเต้านมโต ลูกอัณฑะเหี่ยว สติปัญญาเสื่อมถอยและอารมณ์แปรปรวน (ไพร์ตัน อันอินทร์, 2552)



2.2 การทำเหมืองข้อมูล (Data Mining)

การทำเหมืองข้อมูล (การทำเหมืองข้อมูล CSC662 Data Mining, Data Warehouse and Visualization, 2555: เว็บไซต์) เป็นศาสตร์ที่มีการผสมผสานระหว่างทฤษฎีสถิติและเทคโนโลยีคอมพิวเตอร์เข้าด้วยกัน แนวคิดของการทำเหมืองข้อมูล คือ การวิเคราะห์ข้อมูลจากฐานข้อมูลขนาดใหญ่ เพื่อให้ได้มาซึ่งความรู้ หรือสิ่งที่สำคัญ ซึ่งจะอยู่ในรูปแบบของฐานความรู้เพื่อนำมาใช้ในการวิเคราะห์ หรือทำนายสิ่งต่างๆ ที่จะเกิดขึ้น โดยมีกระบวนการหลักดังนี้

1. Pre-Processing เป็นกระบวนการในการเตรียมข้อมูลให้เหมาะสม และให้อยู่ในรูปแบบที่สามารถนำไปใช้งานได้ โดยต้องพิจารณาจากความต้องการของผู้ใช้ว่าต้องการวิเคราะห์ข้อมูลในส่วนใด รวมถึงเทคนิคของการทำเหมืองข้อมูลที่จะเลือกใช้งาน

2. Data Mining เป็นกระบวนการในการเลือกเทคนิคที่เหมาะสมกับงานที่ต้องการ ซึ่งสามารถทำการเลือกเทคนิคมากกว่าหนึ่งเทคนิคมาใช้งาน ซึ่งเทคนิคนี้เป็นการประมวลผลข้อมูลที่ผ่านมากระบวนการ Pre-Processing มาแล้วเพื่อดึงความรู้ที่สำคัญออกมา โดยผลลัพธ์ที่ได้ในขั้นตอนนี้ คือ ฐานความรู้ ซึ่งมีอยู่หลายรูปแบบ อาทิเช่น

Association Rule เป็นการค้นหากฎความสัมพันธ์ของข้อมูล ซึ่งกฎที่ได้อาจจะนำไปใช้ในการวิเคราะห์หรือทำนายผล

Classification and Prediction เป็นการจำแนกกลุ่มข้อมูล และทำการสร้างแบบจำลอง (Model) จากข้อมูลที่มีอยู่ เมื่อได้แบบจำลองออกมาแล้วสามารถนำไปทำนายผลข้อมูลชุดใหม่ได้

Cluster Analysis เป็นการจัดกลุ่มของข้อมูลที่มีคุณลักษณะคล้ายกัน

3. Post-Processing คือ ขั้นตอนในการนำฐานความรู้ที่ได้จากขั้นตอน Data Mining มาทดสอบและวิเคราะห์ดูว่าผลลัพธ์ที่ได้ออกมาถูกต้องตรงตามความต้องการหรือไม่ ซึ่งอาจจะต้องมีการวนกลับไปทำงานที่ขั้นตอน Data Mining อีกครั้ง จนกว่าจะได้รับความรู้หรือสิ่งที่สนใจออกมาตรงตามความต้องการ

2.2.1 เทคนิคของต้นไม้การตัดสินใจ (decision tree)

ต้นไม้การตัดสินใจ (การเรียนรู้ต้นไม้ตัดสินใจ, 2555: เว็บไซต์) เป็นวิธีหนึ่งที่จะประมาณฟังก์ชันที่มีค่าไม่ต่อเนื่อง (Discrete-value Function) ด้วยแผนผังต้นไม้ อาจประกอบด้วยเซตของกฎต่างๆ แบบ ถ้า-แล้ว (If-then) เพื่อให้สามารถอ่านแล้วเข้าใจการตัดสินใจของต้นไม้ได้

ลักษณะของต้นไม้การตัดสินใจ

ต้นไม้การตัดสินใจจะทำการจัดกลุ่ม (Classify) ชุดข้อมูลนำเข้าในแต่ละกรณี (Instance) แต่ละบัพ (Node) ของต้นไม้การตัดสินใจ คือ ตัวแปร (Attribute) ต่างๆ ของชุดข้อมูล เช่น หากต้องการตัดสินใจว่า จะไปเล่นกีฬาหรือไม่ก็จะมีตัวแปรต้นที่จะต้องพิจารณา คือ ทศนิยมภาพ ลม ความชื้น อุณหภูมิ เป็นต้น และมีตัวแปรตามซึ่งเป็นผลลัพธ์จากต้นไม้คือการตัดสินใจว่าจะไปเล่นกีฬาหรือไม่ ซึ่งแต่ละตัวแปรนั้น ก็จะมีค่าของตัวเอง (Value) เกิดเป็นชุดของตัวแปร-ค่าของตัวแปร (Attribute-value Pair) เช่น ทศนิยมภาพเป็นตัวแปร ก็อาจมีค่าได้เป็น ฝนตก แดดออก หรือการตัดสินใจว่าจะไปเล่นกีฬาหรือไม่นั้นก็อาจมีค่าได้เป็นใช่ กับ ไม่ใช่ เป็นต้น การทำนายประเภทด้วยต้นไม้ตัดสินใจ จะเริ่มจากบัพราก โดยทดสอบค่าตัวแปรของบัพ แล้วจึงตามกิ่งของต้นไม้ที่กำหนดค่า เพื่อไป



ยังบัปถูกัดไป การทดสอบนี้จะกระทำไปจนกระทั่งเจอบัปไบซึ่งจะแสดงผลการทำนาย ตัวอย่างต้นไม้ตัดสินใจใช้ทำนายว่าจะเล่นกีฬาหรือไม่ โดยพิจารณาจากลักษณะอากาศของวันนั้น โดยวัตถุที่ต้องการทำนายประเภท ประกอบด้วยลักษณะหรือตัวแปร 3 ตัว ได้แก่ ทัศนียภาพ ความชื้น และ กระแสลม ดังนั้น ถ้ากำหนดวันวันหนึ่งมีคุณลักษณะแสดงเป็นเวกเตอร์ เช่น [สภาพอากาศ=แดดออก, ความชื้น=สูง] การทำนายว่าจะเล่นกีฬาหรือไม่ จะเริ่มจากบัปแรก โดยทดสอบค่าตัวแปร “สภาพอากาศ” ซึ่งมีค่าเท่ากับ “แดดออก” จึงไปทดสอบค่าตัวแปร “ความชื้น” ในบัปถัดไป ทำให้ได้ประเภทของวันนี้ คือ “ไม่เล่นกีฬา”

ปัญหาที่เหมาะสมสำหรับต้นไม้การตัดสินใจ

เนื่องจากต้นไม้การตัดสินใจเป็นต้นไม้ที่แต่ละกิ่งที่ออกมาจากบัปแทนค่าของข้อมูลที่เป็นไปได้ในบัปนั้น เนื่องจากต้นไม้มีจำนวนกิ่งที่จำกัด ดังนั้นค่าของตัวแปรที่เป็นไปได้จึงต้องจำกัดด้วย จึงต้องมีจำนวนตัวแปรที่จำกัด และนอกจากนั้นยังบังคับว่าค่าของตัวแปรนั้นต้องไม่ต่อเนื่องด้วย โดยข้อมูลที่ เข้ามานั้นอาจมีความผิดพลาดได้บ้าง โดยต้นไม้การตัดสินใจจะมีกระบวนการที่จะไม่นำความผิดพลาดนั้นมาพิจารณาเรียกว่า การตัดแต่งกิ่ง (Post-pruning)

ขั้นตอนวิธีการสร้างต้นไม้การตัดสินใจ

ในปัจจุบันนี้มีการพัฒนาขั้นตอนวิธี (Algorithm) ในการฝึกฝน (Training) ต้นไม้การตัดสินใจมากมาย ซึ่งส่วนมากมาจากวิธีพื้นฐานวิธีหนึ่งซึ่งเป็นการค้นหาแบบละโมภ (Greedy Search) จากบนลงล่าง (Top-down) ชื่อว่า ID3 ซึ่งถูกพัฒนาโดย John Ross Quinlan ในปี 1986

เอนโทรปี (Entropy)

ID3 นั้นสร้างต้นไม้การตัดสินใจจากบนลงล่างด้วยการถามว่าลักษณะใด (ขอใช้คำว่า ลักษณะแทนตัวแปรต้น) ควรจะเป็นรากของต้นไม้การตัดสินใจต้นนี้ และถามซ้ำๆ ไปเรื่อยๆ เพื่อหาต้นไม้ทั้งต้นด้วยการเขียนโปรแกรมด้วยความสัมพันธ์แบบเวียนเกิด (Recursion) โดยในการเลือกว่าลักษณะใดดีที่สุดนั้นดูจากค่าของลักษณะเรียกว่าเกนความรู้ (Information gain) ก่อนที่จะรู้จักเกนความรู้จะต้องนิยามค่าหนึ่งที่ใช้บอกความไม่บริสุทธิ์ของข้อมูลก่อน เรียกว่าเอนโทรปี (Entropy) โดยนิยามเอนโทรปีของต้นไม้การตัดสินใจในตัวในเซตของตัวอย่าง S คือ $E(S)$ ดังนี้

$$E(S) = - \sum_{j=1}^n \rho_s(j) \log_2 \rho_s(j)$$

เมื่อ S คือ ตัวอย่างที่ประกอบด้วยชุดของตัวแปรต้นและตัวแปรตามหลายๆกรณี

$\rho_s(j)$ คือ อัตราส่วนของกรณีใน S ที่ตัวแปรตามหรือผลลัพธ์มีค่า j

โดยสำหรับต้นไม้การตัดสินใจที่มีผลลัพธ์เป็นแค่เพียงค่าตรรกะ (Boolean) ใช้กับไม่ใช่เหมือนกับที่ยกมาตอนต้นของบทความนั้น จะมีเอนโทรปีคือ

$$E(S) = -\rho_{yes} \log_2(\rho_{yes}) - \rho_{no} \log_2(\rho_{no})$$



เมื่อพิจารณาเอนโทรปีแล้วจะเห็นว่า เอนโทรปีจะมีค่าอยู่ระหว่าง 0 กับ 1 โดยจะมีค่าเป็นศูนย์เมื่อทุกๆกรณีมีผลลัพธ์เพียงแบบเดียว เช่น ใช่ทั้งหมด หรือ ไม่ใช่ทั้งหมด และจะมีค่ามากขึ้นเมื่อเริ่มมีค่าที่แตกต่างกันมากขึ้น หรือจะพูดอีกนัยหนึ่งก็คือ เอนโทรปีจะมีค่ามากขึ้นหากข้อมูลไม่บริสุทธิ์ และจะตัดสินใจได้ว่าผลลัพธ์จะเป็นอะไรเมื่อเอนโทรปีเป็น 0 เท่านั้น

เกนความรู้ (Information Gain)

ซึ่งจากการนิยามเอนโทรปีข้างต้น ทำให้เราสามารถนิยามลักษณะของตัวแปรต้นที่ดีได้ โดย ตัวแปร A จะเป็นตัวแปรต้นที่ดีก็ต่อเมื่อ หากว่าแบ่งข้อมูลตัวอย่าง (Example) ออกเป็นชุดๆ มีจำนวนชุดตามจำนวนค่าของ A ที่เป็นไปได้ เพื่อให้แต่ละกรณี (Instance) ในชุดนั้นมีค่า A เพียงค่าเดียว และค่าเฉลี่ยของเอนโทรปีของชุดข้อมูลที่ถูกแบ่งออก (Partition) มานั้นต่ำที่สุด เรียกค่าคาดหวังของการลดลงของเอนโทรปีหลังจากข้อมูลถูกแบ่งด้วย A ว่าเกนความรู้ของ A นิยามโดย

$$Gain(S, A) = E(S) - \sum_{v=valu(A)} \frac{|S_v|}{|S|} E(S_v)$$

เมื่อ S คือ ตัวอย่างที่ประกอบด้วยชุดของตัวแปรต้นและตัวแปรตามหลายๆกรณี
 E คือ เอนโทรปีของตัวอย่าง
 A คือ ตัวแปรต้นที่พิจารณา
 value (A) คือ เซตของค่าของ A ที่เป็นไปได้
 Sv คือ ตัวอย่างที่ A มีค่า v ทั้งหมด

จะเห็นว่าหากเกนความรู้ของ A ยิ่งมากแสดงว่า หลังจากแบ่งตัวอย่าง S ด้วย A แล้วในแต่ละชุดที่แบ่งได้จะมี Entropy เข้าใกล้ศูนย์มากยิ่งขึ้น ทำให้ใกล้ที่จะตัดสินใจได้มากขึ้น เกนความรู้จึงเป็นค่าที่ดีที่จะบอกความดีของตัวแปรต้นที่นำมาพิจารณา

การใช้ ID3 สอนต้นไม้การตัดสินใจ

เมื่อเราสามารถบอกความน่าเชื่อถือของตัวแปรต้นได้ จึงสามารถนำไปช่วยในการหาต้นไม้ การตัดสินใจด้วย ID3 ได้โดยมีกระบวนการดังนี้

1. นำตัวแปรต้นที่ยังไม่ถูกนำมาใช้ทั้งหมดมาหาเกนความรู้
2. เลือกตัวที่มีเกนสูงที่สุด
3. สร้างต้นไม้ที่มีบัพรากเป็นของตัวแปรต้นตัวนั้น

นำมาเขียนเป็นรหัสเทียม (Pseudo Code) ได้ดังนี้ โดยยกตัวอย่างมาเฉพาะกรณีที่ตัวแปรตาม มีแค่เพียงใช่กับไม่ใช่เท่านั้น เมื่อ

Examples คือตัวอย่างที่นำมาสอน

Target_Attribute คือตัวแปรตาม

Attribute คือตัวแปรต้น

ID3 (Examples, Target_Attribute, Attributes) มีขั้นตอนดังนี้

1. สร้างบัพซึ่งเป็นรากเปล่าๆสำหรับต้นไม้
2. ถ้าทุกตัวอย่างในต้นไม้มีค่าผลลัพธ์ของตัวแปรตามเป็นใช่
 return รากที่มีค่า + (ใช่)



3. ถ้าทุกตัวอย่างในต้นไม้มิมีค่าผลลัพธ์ของตัวแปรตามเป็นไม่ใช่
return รากที่มีค่า - (ไม่ใช่)
4. ถ้าเซตของ Attribute เป็นเซตว่าง
return รากที่มีค่าเป็นค่าของ Target_Attribute ที่มีจำนวนมากที่สุดใน

Examples

5. ถ้ามีฉะนั้น เริ่ม
 - 5.1 A = ตัวแปรต้นที่มีค่าของเกณฑ์ความรู้สูงสุด
ให้รากที่ค่าเป็น A
สำหรับแต่ละค่าที่เป็นไปได้, v_i , ของ A,
 - 5.1.1 สร้างกิ่งต่อจากรากที่จะตัดสินใจมาทางนี้เมื่อ $A = v_i$
 - 5.1.2 สร้าง Examples(v_i) เป็นสับเซตของ Example ที่ A มีค่า v_i
 - 5.1.3 ถ้า Examples(v_i) เป็นเซตว่าง
 - 5.1.4 ต่อกิ่งนี้ด้วยบัพที่มีใบมีค่าเป็นค่าของ Target_Attribute ที่มีจำนวน

มากที่สุด ใน Examples

- 5.1.5 มีฉะนั้น ต่อกิ่งนี้ด้วย ID3 (Examples(v_i), Target_Attribute,

Attributes – {A})

5.2 จบ

5.3 Return ราก

เทคนิคต้นไม้ตัดสินใจ (Decision Tree Classifiers (C4.5))

อัลกอริทึมหนึ่งที่ใช้ในการสร้างแบบจำลองต้นไม้ตัดสินใจคืออัลกอริทึม C4.5 ถูกพัฒนาขึ้นมาโดย J. Ross Quinlan โดยใช้หลักการเลือกข้อมูลตามลำดับของค่า Information Gain สูงที่สุดเป็น Node เริ่มต้น และ Node ถัดไปจะมีค่า Information Gain ลดหลั่นกันไป สูตรที่ใช้ในการคำนวณหาค่า Information Gain ต้องเริ่มจากการหาค่า Entropy

Entropy :

$$Entropy(s) = \sum_{i=1}^c - p_i \log_2 p_i$$

โดย S คือ Attribute ที่นำมาวัดค่า Entropy

P_i คือ สัดส่วนของจำนวนสมาชิกของกลุ่ม i กับจำนวนสมาชิกทั้งหมดของกลุ่มตัวอย่าง

C คือ จำนวนกลุ่มที่ใช้ในการจำแนกข้อมูล

Information Gain :

$$Gain(S, A) = Entropy(S) - \sum_{v = value(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$



โดย A คือ Attribute A
 $|S_v|$ คือ สมาชิกของ Attribute A ที่มีค่า v
 $|S|$ คือ จำนวนสมาชิกของกลุ่มตัวอย่าง

ปัญหา Over-Fitting เป็นปัญหาหนึ่งที่เกิดขึ้นในการสร้างแบบจำลองต้นไม้ตัดสินใจ ลักษณะปัญหานี้คือแบบจำลองที่ได้มี Node และ Branch ที่มีความลึก ทำให้แบบจำลองซับซ้อนเกินความจำเป็น หรือการที่จำนวนของข้อมูลฝึกมีน้อยเกินกว่าที่จะจัดกลุ่มได้ตามต้องการ ซึ่งอัลกอริทึม C4.5 ได้ใช้เทคนิค Pruning เพื่อลดขนาดและความซับซ้อนของแบบจำลองต้นไม้ตัดสินใจ

การตัดกิ่งต้นไม้ตัดสินใจ

ในขณะที่กำลังสร้างต้นไม้ตัดสินใจในแต่ละกิ่งอาจการสร้างอย่างผิดพลาด อันเนื่องมาจากข้อมูลฝึกที่มีข้อมูลรบกวน (Noise) ซึ่งเกิดจากการบันทึก ข้อมูลผิดพลาดหรือความผิดพลาดที่เกิดจากระบบเอง หรือในชุดข้อมูลอาจจะมีข้อมูลที่ผิดปกติจากข้อมูลส่วนใหญ่ (Outlier) ปะปนมาด้วย การตัดกิ่งต้นไม้ตัดสินใจเป็นเทคนิคที่ใช้สำหรับแก้ปัญหานี้ และจะช่วยลดการเกิดปัญหาการเจาะจงโมเดลกับข้อมูลมากเกินไป (Overfitting) ได้ (Cohen and Jensen, 1997) โดยปัญหานี้ทำให้ได้โครงสร้างต้นไม้ที่สามารถจำแนกข้อมูลได้ดีกับชุดข้อมูลที่ใช้สร้างต้นไม้ตัดสินใจ แต่เมื่อนำไปใช้กับข้อมูลใหม่ประสิทธิภาพในการจำแนกกลุ่มข้อมูลจะลดลง การตัดกิ่งต้นไม้ตัดสินใจจะใช้ค่าทางสถิติในการตัดกิ่งที่มีความน่าเชื่อถือน้อยที่สุดออกไป เพื่อให้ต้นไม้ใหม่ที่ได้สามารถทำงานได้รวดเร็วขึ้นและยังเป็นการปรับปรุงขีดความสามารถของต้นไม้ในการทำนายข้อมูลใหม่ ๆ ได้แม่นยำมากยิ่งขึ้นอีกด้วย โดยการตัดกิ่งต้นไม้ตัดสินใจที่เป็นที่นิยมมีอยู่ 2 ประเภท ดังต่อไปนี้

1. การตัดกิ่งขณะที่เรียนรู้ (Pre-pruning)

เป็นการตัดกิ่งต้นไม้หรือหยุดการแตกกิ่งในขั้นตอนการสร้างต้นไม้ตัดสินใจโดยการทำให้โหนดที่ถูกตัดนั้นเปลี่ยนเป็นใบ และให้ใบนั้นแสดงกลุ่มที่มีจำนวนข้อมูลสนับสนุนหรือมีความน่าจะเป็นที่ข้อมูลจะอยู่ในกลุ่มนั้นมากที่สุด (Breslow and Aha, 1997)

ในขณะทำการสร้างต้นไม้ตัดสินใจนั้น ต้องมีการคำนวณ หรือวัดค่าทางสถิติที่สำคัญต่างๆ เพื่อใช้ประเมินว่า ควรที่จะสร้างหรือแตกกิ่งของต้นไม้หรือไม่อย่างไรถ้าค่าที่วัดได้ไม่ถึงจุดที่กำหนดไว้ก็จะถือว่าโหนดนั้นไม่สมควรที่จะทำการแตกกิ่งต่อไปซึ่งเป็นการยากที่จะกำหนดว่าค่าที่ใช้เป็นเกณฑ์เหล่านั้นควรจะมีค่าเป็นเท่าไร ถ้ากำหนดค่านั้นสูงเกินไปก็จะทำให้ได้ต้นไม้ที่มีความซับซ้อน แต่ถ้ากำหนดค่าต่ำเกินไปก็จะทำให้ได้ต้นไม้ที่มีขนาดเล็กจนไม่สามารถนำไปใช้งานได้

2. การตัดกิ่งหลังการเรียนรู้ (Post-pruning)

เป็นการตัดกิ่งของต้นไม้ตัดสินใจที่ถูกสร้างขึ้นสมบูรณ์แล้วโดยใช้การวัดค่าความซับซ้อนของแต่ละโหนด หลังจากทำการตัดกิ่งของต้นไม้แล้วโหนดที่อยู่ล่างสุดที่ไม่ได้ถูกตัดจะถูกเปลี่ยนไปเป็นใบและจะแสดงกลุ่มที่มีจำนวนข้อมูลสนับสนุนมากที่สุด (Breslow and Aha, 1997) สำหรับทุกๆ โหนดที่ไม่ใช่ใบของต้นไม้จะมีการคำนวณค่าอัตราความผิดพลาดที่คาดหวังไว้ซึ่งเป็นค่าที่แสดงถึงความผิดพลาด ที่จะเกิดขึ้น ถ้าโหนดของต้นไม้ย่อนั้นถูกตัดออกไปโดยที่ค่าความผิดพลาดของโหนดที่ไม่ได้ถูกตัดจะถูกคำนวณโดยใช้ค่าผลรวมความผิดพลาดของแต่ละกิ่งและให้ค่าน้ำหนักตามสัดส่วนของกิ่งนั้นๆ ถ้าการตัดโหนดนั้นนำไปสู่การเกิดความผิดพลาดที่สูงขึ้น โหนดของต้นไม้ย่อนั้นก็จะต้องยังคงไว้แต่ถ้าการตัดนั้นทำให้ได้ค่าความผิดพลาดเป็นที่ยอมรับได้โหนดนั้นก็จะถูกตัดออกไป หลังจากทำ



การตัดกิ่งต้นไม้ตัดสินใจแล้วจะต้องทำการวัดค่าความแม่นยำ (Accuracy) ของต้นไม้ที่ทำการตัดกิ่งแล้วด้วยโดยที่ต้นไม้ที่ให้ค่าความผิดพลาดน้อยที่สุดจะถูกเลือก นอกจากการตัดกิ่งต้นไม้ตัดสินใจโดยอาศัยการวัดค่าความผิดพลาดที่จะเกิดขึ้นแล้วยังมีเทคนิคอื่นๆ เช่น การใช้ค่าการเข้ารหัส (Encode) ในการพิจารณา ตัดกิ่งของต้นไม้โดยใช้หลักการของ Minimum Description Length (MDL) ด้วย เป็นต้น (Quinlan and Rivest, 1989)

เทคนิคการตัดกิ่งโดยใช้ค่าความผิดพลาด (Error-based pruning)

Error-based pruning (EBP) วิธีการตัดกิ่งต้นไม้ตัดสินใจนี้ ได้นำมาใช้ในอัลกอริทึม สำหรับสร้างต้นไม้ตัดสินใจที่ชื่อ C4.5 (Quinlan, 1993) โดยพัฒนารูปแบบมาจากวิธี Pessimistic error pruning ปรับปรุงการคำนวณค่าอัตราความผิดพลาดที่คาดว่าจะเกิดขึ้น วิธีนี้ใช้ชุดข้อมูลฝึกสำหรับสร้างและตัดกิ่งต้นไม้ตัดสินใจ โดยไม่ต้องใช้ชุดข้อมูลที่แยกออกต่างหากสำหรับการตัดกิ่งโดยเฉพาะ

ขั้นตอนการทำงานจะตรวจสอบโหนดจากล่างสุดขึ้นไปยังรากของต้นไม้โดยใช้วิธีการเข้าถึงโหนดแบบ Post-order Traversal สำหรับแต่ละโหนด t มีทางเลือก 2 วิธีในการปรับปรุงต้นไม้ตัดสินใจคือ ตัดกิ่งต้นไม้ T ตรงตำแหน่งของโหนด t แล้วเปลี่ยนเป็นโหนดใบ ที่มีกลุ่มของข้อมูลเป็น กลุ่มหลักของกลุ่มตัวอย่างจากการแบ่งกลุ่มของชุดข้อมูลฝึก ที่โหนดนั้น หรือตัดกิ่ง T_t ที่เป็นต้นไม้ย่อยของ T_t ที่มีผลรวมของจำนวนตัวอย่างมากที่สุดขึ้นมาแทนที่ตรงตำแหน่งของโหนด t โดยที่ไม่ทำให้ค่าความผิดพลาดเมื่อปรับปรุงต้นไม้ตัดสินใจแล้วมีค่าเพิ่มขึ้น (Esposito et al., 1999)

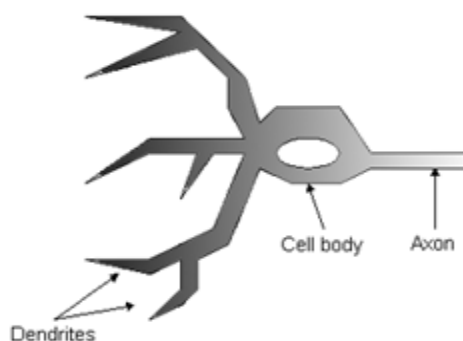
การตัดกิ่งด้วยวิธี EBP ใช้การคำนวณช่วงความเชื่อมั่น (Confidence Level) เพื่อลดความลำเอียงที่เกิดจากการใช้ชุดข้อมูลฝึก คำนวณค่าความผิดพลาดเพียงชุดข้อมูลเดียว โดยกำหนดให้รูปแบบการแจกแจงทวินามประมาณได้เป็นการแจกแจงปกติ (Normal Distribution) สำหรับชุดข้อมูลที่มีกลุ่มตัวอย่างขนาดใหญ่ ทำให้ผลของการตัดกิ่งต้นไม้ตัดสินใจวิธีนี้จะให้ค่าความถูกต้องที่น้อยลงเมื่อใช้กับชุดข้อมูลที่มีกลุ่มตัวอย่างจำนวนน้อยกว่า 100 ตัวอย่าง (Frank, 2000) การตัดกิ่งด้วยวิธี EBP มีขั้นตอนการทำงาน 2 วิธีในการปรับปรุงต้นไม้ตัดสินใจคือ ตัดกิ่งต้นไม้ T ตรงตำแหน่งของโหนด t แล้วเปลี่ยนเป็นโหนดใบ ที่มีกลุ่มของข้อมูลเป็นกลุ่มหลักของกลุ่มตัวอย่างจากการแบ่งกลุ่มของชุดข้อมูลฝึกที่โหนดนั้น หรือตัดกิ่ง T_t ที่เป็นต้นไม้ย่อยของ T_t ที่มีผลรวมของจำนวนตัวอย่างมากที่สุดขึ้นมาแทนที่ตรงตำแหน่งของโหนด t โดยที่ไม่ทำให้ค่าความผิดพลาดเมื่อปรับปรุงต้นไม้ตัดสินใจแล้วมีค่าเพิ่มขึ้น ดังนั้นค่าความซับซ้อนเชิงคำนวณของการตัดกิ่งด้วยวิธี EBP จึงมีค่าเท่ากับ $O(n(\log n)^2)$ เมื่อ n เป็นจำนวนโหนดของต้นไม้ตัดสินใจ (Witten and Frank, 2005)

2.2.2 โครงข่ายประสาทเทียม (Artificial Neural Network)

โครงข่ายประสาทเทียม (โครงข่ายประสาทเทียม Artificial Neural Network, 2555: เว็บไซต์) หรือที่มักจะเรียกสั้น ๆ ว่า ข่ายงานประสาท (Neural Network หรือ Neural Net) คือ โมเดลทางคณิตศาสตร์ สำหรับประมวลผลสารสนเทศด้วยการคำนวณแบบคอนเนชันนิสต์ (Connectionist) เพื่อจำลองการทำงานของเครือข่ายประสาทในสมองมนุษย์ ด้วยวัตถุประสงค์ที่จะสร้างเครื่องมือ ซึ่งมีความสามารถในการเรียนรู้การจดจำแบบรูป (Pattern Recognition) และการอุปमानความรู้ (Knowledge Deduction) เช่นเดียวกับความสามารถที่มีในสมองมนุษย์ แนวคิดเริ่มต้นของเทคนิคนี้ได้มาจากการศึกษาข่ายงานไฟฟ้าชีวภาพ (Bioelectric Network) ในสมอง ซึ่งประกอบด้วย เซลล์ประสาท หรือ “นิวรอน” (Neurons) และ จุดประสานประสาท (Synapses) แต่ละเซลล์ประสาท



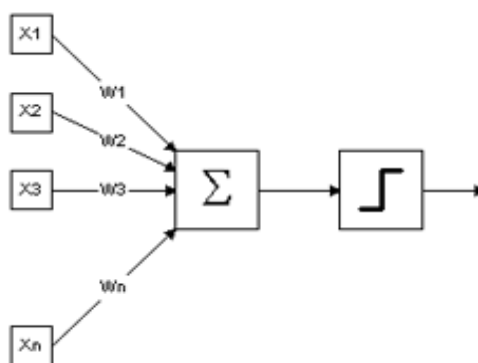
ประกอบด้วยปลายในการรับกระแสประสาท เรียกว่า “เดนไดรต์” (Dendrite) ซึ่งเป็น Input และ ปลายในการส่งกระแสประสาทเรียกว่า “แอกซอน” (Axon) ซึ่งเป็นเหมือนผลลัพธ์ (Output) ของเซลล์ เซลล์เหล่านี้ทำงานด้วยปฏิกิริยาไฟฟ้าเคมี เมื่อมีการกระตุ้นด้วยสิ่งเร้าภายนอกหรือกระตุ้นด้วยเซลล์ด้วยกัน กระแสประสาทจะวิ่งผ่าน Dendrite เข้าสู่นิวเคลียสซึ่งจะเป็นตัวตัดสินใจว่าต้องกระตุ้นเซลล์อื่น ๆ ต่อหรือไม่ ถ้ากระแสประสาทแรงพอ นิวเคลียสก็จะกระตุ้นเซลล์อื่น ๆ ต่อไปผ่านทาง Axon ของมันตาม โมเดลนี้ช่วยงานประสาทเกิดจากการเชื่อมต่อระหว่างเซลล์ประสาท จนเป็นเครือข่ายที่ทำงานร่วมกัน



ภาพประกอบ 2.1 Model ของ Neuron ในสมองมนุษย์

โครงสร้าง

นักวิจัยส่วนใหญ่ในปัจจุบันเห็นตรงกันว่าข่ายงานประสาทเทียมมีโครงสร้างแตกต่างจากข่ายงานในสมอง แต่ก็ยังเหมือนสมอง ในแง่ที่ว่าข่ายงานประสาทเทียม คือการรวมกลุ่มแบบขนานของหน่วยประมวลผลย่อย ๆ และการเชื่อมต่อนี้เป็นส่วนสำคัญที่ทำให้เกิดสติปัญญาของข่ายงาน เมื่อพิจารณาขนาดแล้วสมองมีขนาดใหญ่กว่าข่ายงานประสาทเทียมอย่างมาก รวมทั้งเซลล์ประสาทยังมีความซับซ้อนกว่าหน่วยย่อยของข่ายงาน อย่างไรก็ตามหน้าที่สำคัญของสมอง เช่น การเรียนรู้ยังคงสามารถถูกจำลองขึ้นอย่างง่ายด้วยโครงข่ายประสาทนี้



ภาพประกอบ 2.2 Model ของ Neuron ในคอมพิวเตอร์



หลักการ

สำหรับในคอมพิวเตอร์ Neurons ประกอบด้วยสิ่งที่ป้อนเข้า (Input) และ Output เหมือนกัน โดยจำลองให้ Input แต่ละอันมี Weight เป็นตัวกำหนดน้ำหนักของ Input โดย Neuron แต่ละหน่วยจะมีค่า Threshold เป็นตัวกำหนดว่าน้ำหนักรวมของ Input ต้องมากขนาดไหนจึงจะสามารถส่ง Output ไปยัง Neurons ตัวอื่นได้ เมื่อนำ Neuron แต่ละหน่วยมาต่อกันให้ทำงานร่วมกัน การทำงานนี้ในทางตรรกะแล้วก็จะเหมือนกับปฏิกิริยาเคมีที่เกิดขึ้นในสมอง เพียงแต่ในคอมพิวเตอร์ทุกอย่างเป็นตัวเลขเท่านั้นเอง

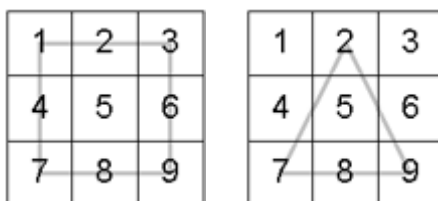
การทำงาน

การทำงานของ Neural Networks คือเมื่อมี Input เข้ามายัง Network ก็เอา input มาคูณกับ Weight ของแต่ละขา ผลที่ได้จาก Input ทุก ๆ ขาของ Neuron จะเอามารวมกันแล้วก็เอามาเทียบกับ Threshold ที่กำหนดไว้ ถ้าผลรวมมีค่ามากกว่า Threshold แล้ว Neuron ก็จะส่ง Output ออกไป Output นี้ก็จะถูกส่งไปยัง Input ของ Neuron อื่น ๆ ที่เชื่อมกันใน Network ถ้าค่าน้อยกว่า Threshold ก็จะไม่เกิด Output เขียนออกมาได้ดังนี้

if (sum(input * weight) > threshold) then output

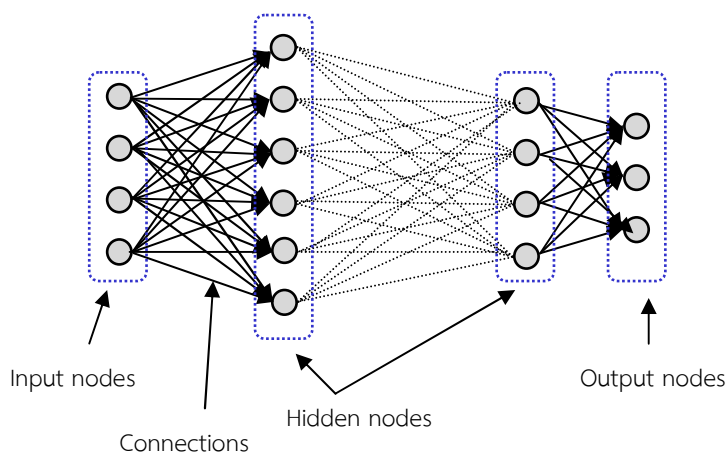
สิ่งสำคัญคือเราต้องทราบค่า Weight และ Threshold สำหรับสิ่งที่เราต้องการเพื่อให้คอมพิวเตอร์รู้จัก ซึ่งเป็นค่าที่ไม่แน่นอน แต่สามารถกำหนดให้คอมพิวเตอร์ปรับค่าเหล่านั้นได้โดยการสอนให้มันรู้จัก Pattern ของสิ่งที่เราต้องการให้มันรู้จัก เรียกว่า “Back Propagation” ซึ่งเป็นกระบวนการย้อนกลับของการรู้จัก ในการฝึก Feed-forward Neural Networks จะมีการใช้อัลกอริทึมแบบ Back-Propagation เพื่อใช้ในการปรับปรุงน้ำหนักคะแนนของเครือข่าย (Network Weight) หลังจากใส่รูปแบบข้อมูลสำหรับฝึกให้แก่เครือข่ายในแต่ละครั้งแล้ว ค่าที่ได้รับ (Output) จากเครือข่ายจะถูกนำไปเปรียบเทียบกับผลที่คาดหวัง แล้วทำการคำนวณหาค่าความผิดพลาด ซึ่งค่าความผิดพลาดนี้จะถูกส่งกลับเข้าสู่เครือข่ายเพื่อใช้แก้ไขค่าน้ำหนักคะแนนต่อไป

อย่างเช่นจะรู้จักรูปสามเหลี่ยม กับรูปสี่เหลี่ยม เราอาจแบ่ง Input เป็น 9 ตัวคือเป็นตาราง 3x3 ถ้าวาดรูปสี่เหลี่ยมหรือสามเหลี่ยมให้เต็มกรอบ 3x3 พอดี สี่เหลี่ยมจะมีส่วนของขอบอยู่ในช่อง 1,2,3,4,6,7,8,9 ก็สมมติให้น้ำหนักตรงช่องเหล่านี้มีค่ามาก ๆ ถ้ามีเส้นขีดผ่านก็เอามาคูณกับน้ำหนักแล้วก็เอามารวมกัน ตั้งค่าให้พอเหมาะก็จะสามารถแยกแยะระหว่างสี่เหลี่ยมกับสามเหลี่ยมได้ ซึ่งนี่คือหลักการของ Neural Network



ภาพประกอบ 2.3 การแยกแยะระหว่างสี่เหลี่ยมและสามเหลี่ยม





ภาพประกอบ 2.4 โครงสร้างวงจร Neural Network

Output ของแต่ละ Node

$$y_i = f(w_i^1 x_1 + w_i^2 x_2 + w_i^3 x_3 + \dots + w_i^m x_m)$$

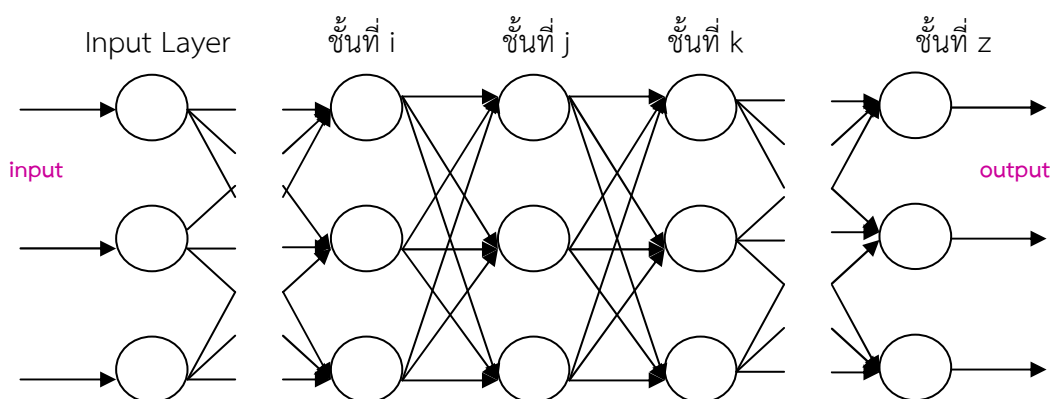
$$= f\left(\sum_j w_i^j x_j\right)$$

เมื่อ x_i = input จากโหนดอื่นๆ

w_{ij} = น้ำหนัก (weight) ของแต่ละแขน (connection)

Back Propagation Algorithm

Back-propagation เป็นอัลกอริทึมที่ใช้ในการเรียนรู้ของเครือข่ายประสาทวิธีหนึ่งที่นิยมใช้ใน Multilayer Perceptron เพื่อปรับค่าน้ำหนักในเส้นเชื่อมต่อระหว่างโหนด (Node) ให้เหมาะสม โดยการปรับค่านี้นี้จะขึ้นกับความแตกต่างของค่า Output ที่คำนวณได้กับค่า Output ที่ต้องการ พิจารณารูปต่อไปนี้ประกอบ



ภาพประกอบ 2.5 รูปแบบ Back-propagation Neural Network



ขั้นตอนของ Back-propagation Algorithm มีดังนี้

1. กำหนดค่าอัตราเร็วในการเรียนรู้ (Rate Parameter : r)
2. สำหรับแต่ละตัวอย่าง Input ให้ทำตามขั้นตอนต่อไปนี้จนกว่าได้ระดับ

Performance ที่ต้องการ

คำนวณหาค่า Output โดยใช้ค่าน้ำหนักเริ่มต้นซึ่งอาจได้จากการสุ่ม

คำนวณหาค่า β : แทนประโยชน์ที่จะได้รับสำหรับการเปลี่ยนค่า output ของ

แต่ละโหนด

ในชั้น output (Output Layer)

$$\beta_z = d_z - o_z$$

เมื่อ d_z = ค่า output ที่ต้องการ

o_z = ค่า output ที่คำนวณได้

- ในชั้นซ่อน (Hidden Layer)

$$\beta_j = \sum w_{j \rightarrow k} o_k (1 - o_k) \beta_k$$

เมื่อ $w_{j \rightarrow k}$ = น้ำหนักของเส้นเชื่อมระหว่างชั้นที่ j กับ k

คำนวณค่าน้ำหนักที่เปลี่ยนแปลงไปสำหรับในทุกน้ำหนัก ด้วยสมการต่อไปนี้

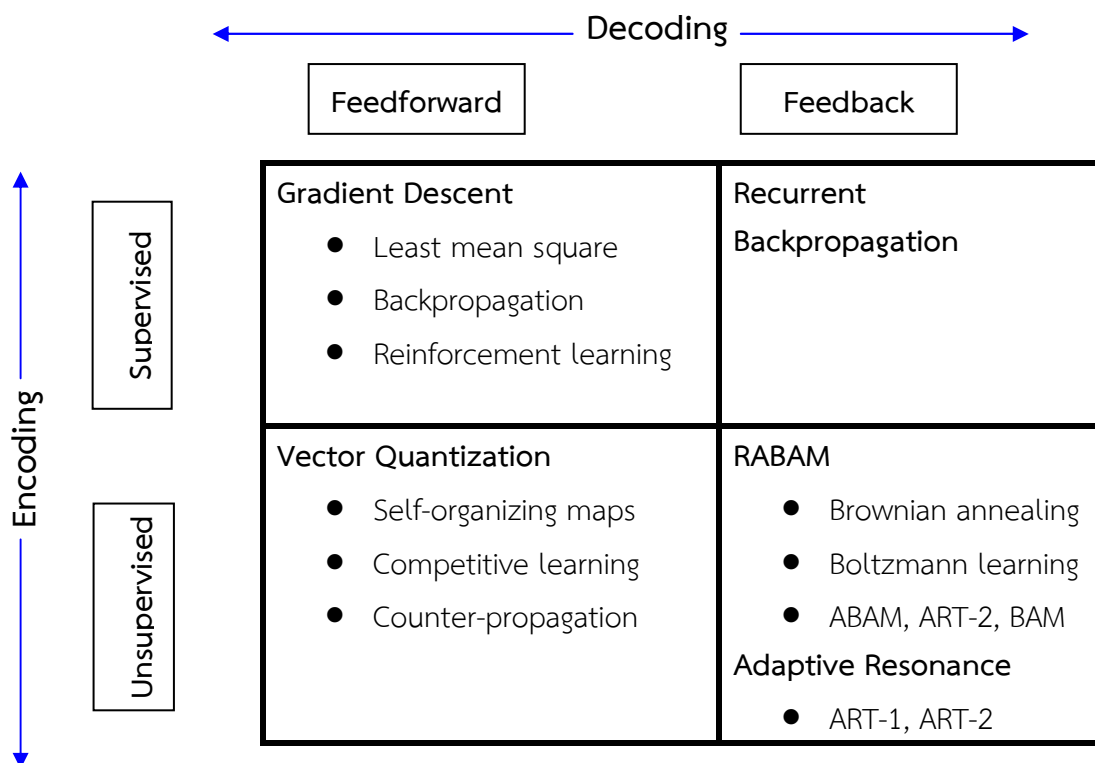
$$\Delta w_{i \rightarrow j} = r o_i o_j (1 - o_j) \beta_j$$

เพิ่มค่าน้ำหนักที่เปลี่ยนแปลง สำหรับตัวอย่าง Input ทั้งหมด และเปลี่ยนค่า

น้ำหนัก



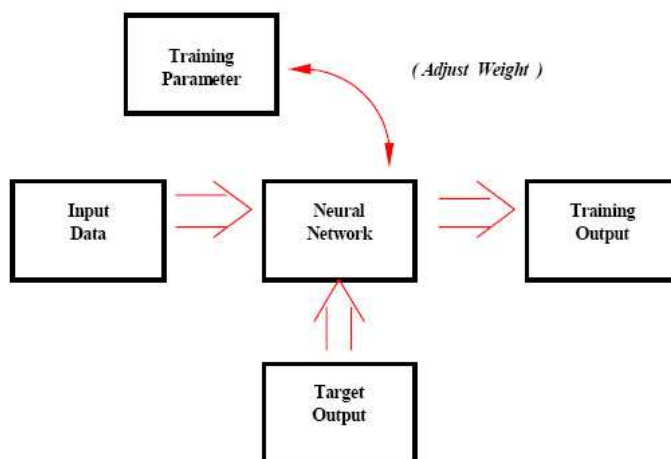
Neural Network Taxonomy



ภาพประกอบ 2.6 อนุกรมวิธานของโครงข่ายประสาทเทียม

การเรียนรู้สำหรับ Neural Network

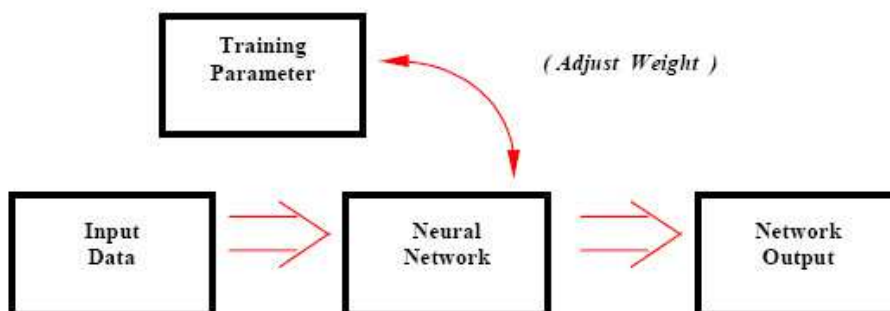
1. Supervised Learning การเรียนแบบมีการสอนเป็นการเรียนแบบที่มีการตรวจคำตอบเพื่อให้วงจรข่ายปรับตัว ชุดข้อมูลที่ใช้สอนวงจรข่ายจะมีคำตอบไว้คอยตรวจสอบว่าวงจรข่ายให้คำตอบที่ถูกต้องหรือไม่ ถ้าตอบไม่ถูก วงจรข่ายก็จะปรับตัวเองเพื่อให้ได้คำตอบที่ดีขึ้น (เปรียบเทียบกับคนเหมือนกับการสอนนักเรียนโดยมีครูผู้สอนคอยแนะนำ)



ภาพประกอบ 2.7 การเรียนรู้แบบมีการสอน (Supervised Learning)



2. Unsupervised Learning การเรียนแบบไม่มีการสอนเป็นการเรียนแบบไม่มีผู้แนะนำ ไม่มีการตรวจคำตอบว่าถูกหรือผิด วงจรข่ายจะจัดเรียงโครงสร้างด้วยตัวเองตามลักษณะของข้อมูล ผลลัพธ์ที่ได้ วงจรข่ายจะสามารถจัดหมวดหมู่ของข้อมูลได้ (เปรียบเทียบกับคน เช่น การที่เราสามารถแยกแยะพันธุ์พืช พันธุ์สัตว์ตามลักษณะรูปร่างของมันได้เองโดยไม่มีใครสอน)

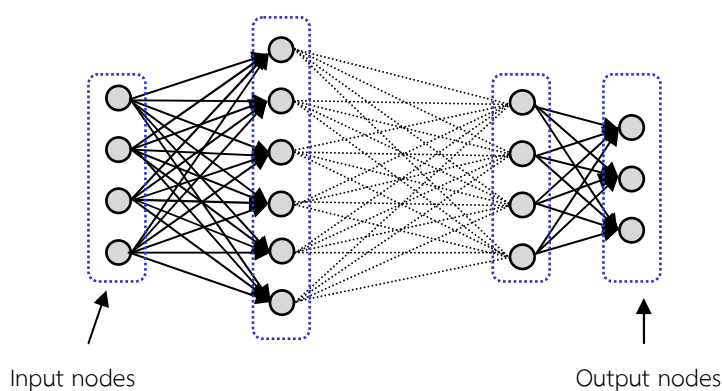


ภาพประกอบ 2.8 การเรียนรู้แบบไม่มีการสอน Unsupervised Learning

Network Architecture

1. Feedforward Network

ข้อมูลที่มีประมวลผลในวงจรข่ายจะถูกส่งไปในทิศทางเดียวจาก Input Nodes ส่งต่อมาเรื่อย ๆ จนถึง Output Nodes โดยไม่มีการย้อนกลับของข้อมูล หรือแม้แต่ Nodes ใน Layer เดียวกันก็ไม่มีการเชื่อมต่อกัน

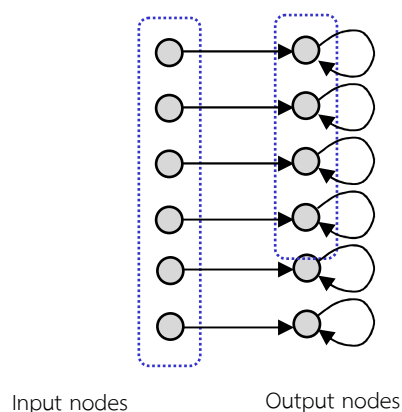


ภาพประกอบ 2.9 สถาปัตยกรรมของ Feedforward Network

2. Feedback network

ข้อมูลที่มีประมวลผลในวงจรข่าย จะมีการป้อนกลับเข้าไปยังวงจรข่ายหลาย ๆ ครั้ง จนกระทั่งได้คำตอบออกมา (บางที่เรียกว่า Recurrent network)





ภาพประกอบ 2.10 สถาปัตยกรรมของ Feedback Network

3. Network Layer

พื้นฐานสามัญที่สำคัญของ Artificial Neural Network ประกอบไปด้วย 3 ส่วน หรือ 3 layer ได้แก่ ชั้นของ input units ที่ถูกเชื่อมต่อกับชั้นของ hidden units ซึ่งเชื่อมต่อกับชั้นของ output units

การทำงานของ input unit จะทำหน้าที่แทนส่วนของข้อมูลดิบ ที่จะถูกป้อนเข้าสู่เครือข่าย

การทำงานของแต่ละ hidden units จะถูกกำหนด โดยการทำงานของ input units และ คำนวณน้ำหนักบนความสัมพันธ์ระหว่าง input units และ hidden units

พฤติกรรมการทำงานของ output units จะขึ้นอยู่กับการทำงานของ hidden units และ คำนวณน้ำหนักระหว่าง hidden units และ output units

ประเภทของเครือข่ายนี้เป็นที่น่าสนใจ เพราะเราสามารถกำหนดการแทนค่าให้แก่ input units ได้อย่างอิสระ คำนวณน้ำหนักระหว่าง input units และ hidden units จะถูกกำหนดเมื่อ hidden unit กำลังทำงาน ฉะนั้นเวลาที่แก้ไขค่าน้ำหนัก hidden units จะสามารถเลือกว่าอะไรคือค่าที่เราแทนเข้ามา

Architecture of Layer สามารถจำแนกสถาปัตยกรรมของชั้น (layer) ออกเป็น 2 ประเภทคือ Single-layer และ Multi-layer

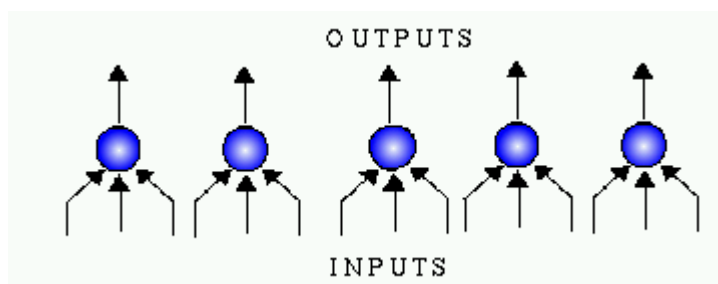
1) Single-layer perceptron เครือข่ายประสาทที่ประกอบด้วยชั้นเพียงชั้นเดียว จำนวน Input nodes ขึ้นอยู่กับจำนวน components ของ Input data และ Activation function ขึ้นอยู่กับลักษณะข้อมูลของ output เช่น ถ้า output ที่ต้องการเป็น “ใช่” หรือ “ไม่ใช่” เราจะต้องใช้ Threshold function

$$f(x) = \begin{cases} 1 & \text{if } x \geq T \\ 0 & \text{if } x < T \end{cases} \quad T = \text{Threshold}$$



หรือถ้า Output เป็นค่าตัวเลขที่ต่อเนื่อง เราต้องใช้ continuous function เช่น Sigmoid function

$$f(x) = \frac{1}{1 + e^{-ax}}$$



ภาพประกอบ 2.11 สถาปัตยกรรมของเพอร์เซพตรอนชั้นเดียว (Single-layer Perceptron)

2) Multi-layer Perceptron เครือข่ายประสาทจะประกอบด้วยหลายชั้น โดยในแต่ละชั้นจะประกอบด้วยโหนด (Nodes) หรือเปรียบได้กับตัวเซลล์ประสาท(neurons) ค่าน้ำหนักของเส้นที่เชื่อมต่อระหว่างโหนดของแต่ละชั้น (เมทริก W), ค่า bias vector (b) และค่า output vector (a) โดย m เป็นตัวเลขบอกลำดับชั้นกำกับไว้ด้านบน เมื่อ p เป็น input vector การคำนวณค่า output สำหรับเครือข่ายประสาทที่มี M ชั้นจะเป็นดังสมการ

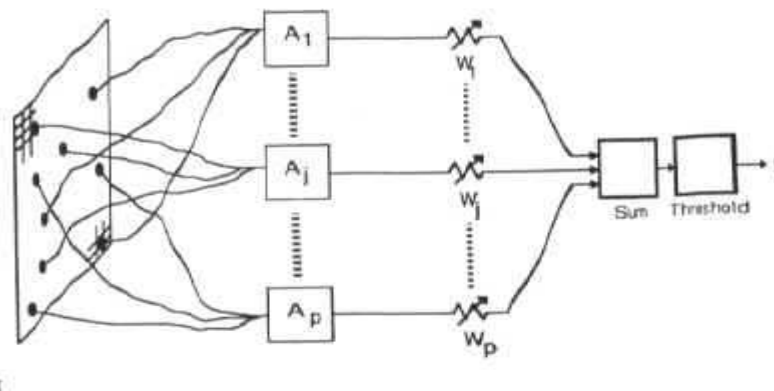
$$a^{m+1} = f^{m+1}(W^{m+1} a^m + b^{m+1})$$

$$\begin{aligned} \text{เมื่อ } m &= 0, 2, \dots, M-1 \\ a^0 &= p \\ a &= a^m \\ f &= \text{transfer function} \end{aligned}$$

4. Perceptrons

ในยุค 60s งานส่วนใหญ่ของข่ายงานได้รับการวิพากษ์วิจารณ์ในหัวข้อเรื่อง Perceptrons ซึ่งค้นพบโดย Frank Rosenblatt โดย Perceptron ซึ่งกลายเป็น MCP model (Neuron with Weighted Inputs) พร้อมกับส่วนต่อเติม จากรูปในส่วน A1,A2,Aj,Ap เรียกว่า Association Units การทำงานเพื่อคัดเลือกสิ่งที่แตกต่างออกมาจากรูปภาพที่รับเข้าไป โดย Perceptrons สามารถคัดลอกความคิดพื้นฐานภายในของสัตว์เลี้ยงลูกด้วยนม หลัก ๆ แล้วจะใช้ใน รูปแบบ Recognition และสามารถขยายให้มีความสามารถสูงกวานี้





ภาพประกอบ 2.12 โครงสร้างของ Perceptrons

ในปี ค.ศ. 1969 Minsky และ Papert ได้เขียนหนังสืออธิบายเกี่ยวกับขอบเขตของ Single-layer Perceptrons ผลกระทบที่ได้รับจากหนังสือเล่มนั้นร้ายแรง เป็นเหตุให้นักวิจัยสาขา Neural Network สูญเสียผลประโยชน์ เนื่องจากหนังสือสามารถถ่ายทอดออกมาได้ดี และแสดงข้อมูลในเชิงคำนวณว่า Single-layer Perceptrons ไม่สามารถที่จะสร้างรูปแบบการจดจำพื้นฐาน (Basic Pattern Recognition Operation) ได้ เช่น การกำหนดความคล้ายคลึงของรูปร่าง หรือกำหนดว่ารูปร่างใดสัมพันธ์กันหรือไม่ แต่สิ่งที่นักวิจัยไม่รู้จนกระทั่งยุค 80s คือ การได้รับการฝึกฝนที่ถูกต้อง ซึ่ง Multi-layer Perceptrons สามารถดำเนินการแก้ไขสิ่งเหล่านี้ได้

การประยุกต์ใช้งาน Neural Network

แบบข่ายงานระบบประสาท (Neural Network) เนื่องจากความสามารถในการจำลองพฤติกรรมทางกายภาพของระบบที่มีความซับซ้อนจากข้อมูลที่ป้อนให้เรียนรู้ การประยุกต์ใช้ข่ายงานระบบประสาทจึงเป็นทางเลือกใหม่ในการควบคุม ซึ่งมีผู้นำมาประยุกต์ใช้งานหลายประเภท ได้แก่

1. งานการจดจำรูปแบบที่มีความไม่แน่นอน เช่น ลายมือ ลายเซ็นต์ ตัวอักษร
รูปหน้า
2. งานการประมาณค่าฟังก์ชันหรือการประมาณความสัมพันธ์ (มี Inputs และ Outputs แต่ไม่ทราบว่า Inputs กับ Outputs มีความสัมพันธ์กันอย่างไร)
3. งานที่สิ่งแวดล้อมเปลี่ยนแปลงอยู่เสมอ (วงจรข่าย Neurons สามารถปรับตัวเองได้)
4. งานจัดหมวดหมู่และแยกแยะสิ่งของ
5. งานทำนาย เช่น พยากรณ์อากาศ พยากรณ์หุ้น
6. การประยุกต์ใช้ข่ายงานระบบประสาทควบคุมกระบวนการทางเคมีโดยวิธี
พยากรณ์แบบจำลอง (Model Predictive Control)
7. การประยุกต์ใช้ข่ายงานระบบประสาทแบบแพร่กระจายกลับในการทำนาย
พลังงานความร้อน ที่สะสมอยู่ในตัวอาคาร
8. การใช้ข่ายงานระบบประสาทในการหาโซโครเมตริกชาร์ท การประยุกต์ใช้ข่าย
งานระบบประสาทควบคุมระบบ HVAC



2.3 การวิเคราะห์การถดถอยโลจิสติกแบบหลายกลุ่ม (Multinomial Logistic Regression Analysis)

การวิเคราะห์การถดถอยโลจิสติก (กัลยา วานิชย์บัญชา, 2550) เป็นเทคนิคการวิเคราะห์สมการถดถอยโลจิสติก เพื่อศึกษาความสัมพันธ์ระหว่างตัวแปรตามและตัวแปรอิสระ และนำสมการถดถอยที่ได้ไปประมาณหรือพยากรณ์ค่าตัวแปรตาม เมื่อกำหนดค่าตัวแปรอิสระจำนวนกลุ่มของตัวแปรตามมี 2 กลุ่ม จะเรียกว่า การวิเคราะห์การถดถอยโลจิสติก 2 กลุ่ม (Binary Logistic Regression) จำนวนกลุ่มของตัวแปรตามมีมากกว่า 2 กลุ่มจะเรียกว่า การวิเคราะห์การถดถอยโลจิสติกแบบหลายกลุ่ม (Multinomial Logistic Regression) สำหรับการวิเคราะห์การถดถอยโลจิสติกแบบหลายกลุ่ม สมการพยากรณ์ที่ได้จากตัวแบบการวิเคราะห์จะเป็นสมการแสดงความน่าจะเป็นของการเกิดเหตุการณ์ที่สนใจ (Probability of Event)

เมื่อตัวแปรตาม (Y) เป็นตัวแปรเชิงกลุ่มที่มีค่ามากกว่า 2 ค่า เช่น ถ้าตัวแปรตาม Y หมายถึงระดับการมีส่วนร่วมในการพัฒนาชุมชน โดยแบ่งเป็น 4 ระดับ คือ

$$Y = \begin{cases} 0 ; \text{ไม่มีส่วนร่วมเลย} \\ 1 ; \text{มีส่วนร่วมบ้างเล็กน้อย} \\ 2 ; \text{มีส่วนร่วมปานกลาง} \\ 3 ; \text{มีส่วนร่วมมาก} \end{cases}$$

โดยตัวแปรอิสระอาจเป็น เพศ, ระดับการศึกษา, รายได้, อาชีพ, ศาสนา, อายุ, สถานภาพสมรส, ระยะเวลาที่อยู่ในครอบครัว เป็นต้น ถ้าสามารถสร้าง Model แสดงความสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระ จะทำให้ทราบว่าตัวแปรอิสระ หรือตัวแปรต้นตัวใดบ้างที่มีความสัมพันธ์หรือมีอิทธิพลต่อระดับการมีส่วนร่วมในการพัฒนาชุมชนของคนในชุมชนนั้นๆ และยังสามารถใช้ Model ดังกล่าวในการพยากรณ์คนใหม่ว่า จะมีส่วนร่วมในการพัฒนาชุมชนมากน้อยอย่างไร โดยทราบค่าตัวแปรอิสระของคนนั้น เช่น เพศ, อายุ, ระดับการศึกษา เป็นต้น

เงื่อนไข

1. ตัวแปรอิสระบางตัวอาจเป็นตัวแปรเชิงปริมาณ บางตัวอาจเป็นตัวแปรเชิงกลุ่ม
2. ตัวแปรตามต้องเป็นตัวแปรเชิงกลุ่มที่มีมากกว่า 2 ค่า
3. Odds ratio ของตัวแปรเชิงกลุ่ม 2 ค่า จะต้องเป็นอิสระกับค่าอื่นๆ

Logit Model

กรณีตัวแปรตาม (Y) มีค่า 2 ค่า จะเขียน (Binary Logistic Model) เป็น

$$\log \left[\frac{P(\text{การเกิดเหตุการณ์})}{1 - P(\text{การเกิดเหตุการณ์})} \right] = b_0 + b_1x_1 + \dots + b_px_p$$

โดยจะเรียก

$$\log \left[\frac{P(\text{การเกิดเหตุการณ์})}{1 - P(\text{การเกิดเหตุการณ์})} \right] \text{ หรือ } \log(\text{Odds}) \text{ ว่า } \text{logit}$$



แต่เมื่อ ตัวแปรตาม (Y) มีค่ามากกว่า 2 ค่า เช่น มี j ค่า ; $j > 2$ จะได้ logit จำนวน j-1 ค่า โดยที่แต่ละค่าจะเปรียบเทียบกับ Baseline Category Logit เช่น ถ้าให้ Baseline Category เป็นค่าคงที่ j จะได้ว่า logit ของ Category ที่ i จะเป็น

$$\log \left[\frac{P(\text{Category } i)}{1 - P(\text{Category } j)} \right] = b_{i0} + b_{i1}x_1 + \dots + b_{ip}x_p$$

จึงทำให้มีสัมประสิทธิ์ $b_{i0}, b_{i1}, \dots, b_{ip}$ สำหรับ Category ที่ i

สำหรับ Baseline Category ควรจะมีค่า $b_0 = b_1 = \dots = b_p = 0$

และในการตรวจสอบความเหมาะสมของตัวแบบ คือ การทดสอบสมมติฐานหลักที่ว่า β_i ทุกตัวมีค่าเท่ากับศูนย์พร้อมกันหรือไม่ ซึ่งใช้ Chi-square และถ้าได้ผลสรุปว่า β_i เป็นศูนย์ไม่ทุกตัว แสดงว่าตัวแบบนั้นมีความเหมาะสม การทดสอบค่าสัมประสิทธิ์ที่เป็นพารามิเตอร์มีค่าเป็นศูนย์หรือไม่ โดยใช้สถิติ

$$\text{Wald_Statistic} = \left(\frac{\hat{\beta}_i}{\text{Standard_Error_of_Estimate}} \right)^2$$

ตัวแปรหุ่น (Dummy Variable)

เป็นการสร้างตัวแปรโดยกำหนดให้ตัวแปรนั้นมี 2 ค่า คือ 0 และ 1 หรือเรียกอีกอย่างหนึ่งว่า ตัวแปรทวินาม (Dichotomous Variable) เช่น สร้างตัวแปรเพศ ให้เป็นตัวแปรหุ่น (Dummy Variable) หรือตัวแปรทวินาม โดยกำหนดให้ เพศชายมีค่าเป็น 1 ส่วนเพศหญิงมีค่าเป็น 0 (หรือเรียกอีกอย่างหนึ่งว่าเป็นเพศชาย มีค่าเท่ากับ หรือได้คะแนนเท่ากับ 1 ส่วนที่ไม่ใช่เป็นชายไม่มีค่า หรือไม่ได้คะแนน ซึ่ง ค่า 1 กับค่า 0 จะกลายเป็นตัวเลขที่คำนวณได้ ซึ่งตัวแปรที่จะนำมาสร้างตัวแปรหุ่น จะต้องเป็นตัวแปรที่บรรจุ ข้อมูลเชิงคุณภาพ (มาตรวัดเป็นนามบัญญัติ หรืออันดับ)

2.4 งานวิจัยที่เกี่ยวข้อง

2.4.1 งานวิจัยในประเทศ

บุญเสริม หุตะแพทย์ และคณะ (2552) ศึกษาการวิเคราะห์ความเกี่ยวข้องของการดื่มสุรากับความรุนแรงในการทำงาน ผลการศึกษาพบว่า การดื่มสุรามีความเกี่ยวข้องกับการกระทำความรุนแรงในการทำงาน 3 ประการ คือ การดื่มสุราระดับสูงทำให้เกิดการกระทำความรุนแรง การนำสุรามาอ้างเป็นเหตุผลเพื่อปิดความรับผิดชอบ และการทำให้ผู้ถูกระทำขาดโอกาสในการปกป้องตนเองจากการถูกละเมิดสิทธิ ลักษณะการกระทำความรุนแรงทั้งหมดเป็นการคุกคามและล่วงละเมิดทางเพศ ที่มีระดับความรุนแรงน้อยถึงขั้นรุนแรงมาก แนวทางให้ความช่วยเหลือและแก้ไขปัญหาดำเนินการโดยการให้คำแนะนำปรึกษา การฟื้นฟูเยียวยา และการให้คำแนะนำด้านกฎหมายและการดำเนินคดี ส่วนการป้องกันความเสี่ยงจากความรุนแรงและลดความเสี่ยงควรดำเนินการจัดให้มีหน่วยงานกลางที่ทำหน้าที่



รับเรื่องร้องทุกข์ร้องเรียนอย่างเป็นทางการ กำหนดขั้นตอนการประเมินความเสี่ยงภัยจากความรุนแรง และ การคุกคามทางเพศที่ชัดเจน รวมทั้งกำหนดแนวทางการป้องกัน และแก้ไขปัญหาและความเสี่ยง ภัยในการทำงานให้ทันสมัยอยู่เสมอ

ไพรัตน์ อ้นอินทร์ (2552) ศึกษาพฤติกรรมการดื่มเครื่องดื่มแอลกอฮอล์ของวัยรุ่นที่ส่งผลกระทบต่อสุขภาพและครอบครัว ผลการศึกษาพบว่า ปัจจัยที่ทำให้วัยรุ่นดื่มเครื่องดื่มแอลกอฮอล์ มี 3 ประการ คือ ปัจจัยนำ พบว่า วัยรุ่นขาดความเชื่อมั่นในการปฏิเสธการดื่ม มีทัศนคติและความคาดหวังเชิงบวกต่อผลจากการดื่ม ปัจจัยเอื้อ พบว่า ร้านค้าจำหน่ายเครื่องดื่มแอลกอฮอล์กระจายทั่วไปและ ฝ่าฝืนกฎหมาย องค์กรหลักในชุมชนไม่ควบคุมอย่างจริงจัง ปัจจัยเสริม ได้แก่ แรงกระตุ้นจากกลุ่มเพื่อน รุ่นพี่ แบบอย่างการดื่มจากผู้หลักผู้ใหญ่ ค่านิยมในชุมชน และทัศนคติของผู้ปกครองที่ยอมรับการดื่มของ วัยรุ่นเป็นเรื่องปกติ โดยเฉพาะในช่วงเทศกาล ผลกระทบที่เกิดจากการดื่มของวัยรุ่นด้านสุขภาพ การศึกษานี้ไม่พบผลกระทบเฉียบพลัน แต่พบว่าเป็นต้นเหตุนำไปสู่ความคึกคะนอง การบาดเจ็บจากการ ทะเลาะวิวาทมากที่สุด ผลกระทบด้านครอบครัว พบว่า ผู้ปกครองมีความห่วงใย ทุกข์ใจ น้อยใจ เกี่ยวกับพฤติกรรมการดื่มที่มากเกินไปของวัยรุ่น และการที่ต้องตกอยู่ในภาวะจำยอม เกี่ยวกับการ สูญเสียเงินโดยไม่จำเป็นจากการก่อคดีของวัยรุ่น

สัจจากจ จอมโนนเขวา (2552) ศึกษาการพัฒนาแบบจำลองการเลือกรูปแบบการเดินทางและแบบจำลองระยะทางการเดินทาง ภายในเขตเทศบาลนครราชสีมา พบว่า การพัฒนา แบบจำลองการเลือกรูปแบบการเดินทางโดยใช้แบบจำลองที่มีเลือกมากกว่าสองทาง (Multinomial Logit) พบว่าแบบจำลองมีร้อยละความถูกต้อง 77.57 โดยตัวแปรที่มีอิทธิพลต่อการเลือกรูปแบบการเดินทางได้แก่จำนวนยานพาหนะ สถานะภาพในบ้าน รายได้ของบุคคล และจำนวนเที่ยวการเดินทาง ส่วนแบบจำลองการเลือกรูปแบบการเดินทางโดยโครงข่ายประสาทเทียมการเรียนรู้แบบ LVQ มีร้อยละความถูกต้อง 71.00 เมื่อพิจารณาแบบจำลองระยะการเดินทางรวมโดยการวิเคราะห์ความถดถอย พบว่าแบบจำลองสำหรับครัวเรือน มีค่าร้อยละความคลาดเคลื่อนสัมบูรณ์เฉลี่ยเท่ากับ 7.48 ตัวแปรที่มี นัยสำคัญประกอบด้วย จำนวนคนในบ้านที่ทำงาน จำนวนคนในครัวเรือนที่กำลังศึกษา จำนวนรถยนต์ที่ ครัวเรือนมีครอบครอง และระยะทางจากที่ตั้งครัวเรือนถึงศูนย์กลางทางเศรษฐกิจของเมือง ส่วน แบบจำลองระยะการเดินทางรวมโดยใช้โครงข่ายประสาทเทียมการเรียนรู้แบบแพร่กลับ พบว่า มีค่าร้อยละความคลาดเคลื่อนสัมบูรณ์เฉลี่ยเท่ากับ 27.16 ผลลัพธ์จากทั้งสองแบบจำลองแสดงให้เห็นว่าการ พัฒนาแบบจำลองโดย Multinomial Logit และการวิเคราะห์ความถดถอยให้ผลการพยากรณ์ที่ดีกว่า โครงข่ายประสาทเทียม

ดุขฎี อายุวัฒน์ และวณิชชา ณรงค์ชัย (2553) ศึกษาการประเมินผลการบังคับใช้กฎหมาย ตามพระราชบัญญัติควบคุมเครื่องดื่มแอลกอฮอล์ พ.ศ. 2551 ในเขตภาคตะวันออกเฉียงเหนือ พบว่า ในแนวทางการเพิ่มประสิทธิผลของการบังคับใช้กฎหมายตามพระราชบัญญัติควบคุมเครื่องดื่ม แอลกอฮอล์ พ.ศ. 2551 ในการควบคุม ป้องกันและแก้ไขปัญหาสุราของประเทศไทยนั้น ควรมีการ เสริมสร้างความรู้ความเข้าใจเกี่ยวกับกฎหมายตามพระราชบัญญัติควบคุมเครื่องดื่มแอลกอฮอล์ พ.ศ. 2551 ตลอดจนพัฒนากระบวนการให้เจ้าพนักงานตามกฎหมาย มีความมั่นใจในการกำกับ ควบคุม เครื่องดื่มแอลกอฮอล์

รภัทรกร เพชรสุข (2553) ศึกษาแบบความสัมพันธ์เชิงสาเหตุของพฤติกรรมการบริโภค เครื่องดื่มแอลกอฮอล์ของนักเรียนผู้ใช้บริการหน่วยงานอนามัยโรงเรียน ผลการศึกษาพบว่า เจตคติ



การคล้อยตามกลุ่มอ้างอิง และการรับรู้ความสามารถในการควบคุมพฤติกรรมกรรมการบริโภคเครื่องดื่มที่มีแอลกอฮอล์ มีอิทธิพลทางตรงต่อความตั้งใจกระทำพฤติกรรมกรรมการบริโภคเครื่องดื่มที่มีแอลกอฮอล์ได้ร้อยละ 0.68, 0.72 และ 0.69 ตามลำดับ อย่างมีนัยสำคัญทางสถิติที่ระดับ .01

ณัฐชนันท์ กิ่งมณี (2554) ศึกษาการรับรู้โฆษณาตรงและโฆษณาแฝงสินค้าเครื่องดื่มแอลกอฮอล์ทางโทรทัศน์ในกลุ่มนักศึกษาในระดับอุดมศึกษา พบว่า มีเพียงการใช้ภาพลักษณ์ การให้ข้อคิดในเชิงสร้างสรรค์สังคม โดยมิได้มีภาพของผลิตภัณฑ์ หรือการชักชวนให้ซื้อสินค้าเครื่องดื่มแอลกอฮอล์ภายในเนื้อหาของโฆษณา ขณะที่การศึกษาการรับรู้โฆษณาตรงและโฆษณาแฝงเครื่องดื่มแอลกอฮอล์ในกลุ่มประชากรนักศึกษา พบว่า กลุ่มประชากรนักศึกษา มีระดับการรับรู้ และการจดจำโฆษณาเครื่องดื่มแอลกอฮอล์ผ่านสื่อโทรทัศน์ในระดับสูงมาก แต่ในด้านพฤติกรรมกรรมการบริโภคเครื่องดื่มแอลกอฮอล์ที่มีสาเหตุจากสื่อโฆษณาทางโทรทัศน์อยู่ในระดับปานกลาง

มันทนา ขอนดอก (2554) ศึกษาปัจจัยที่มีความสัมพันธ์กับการดื่มเครื่องดื่มแอลกอฮอล์ในระยะแรกของนักเรียนชายมัธยมศึกษาตอนต้น อำเภอบางน้ำเปรี้ยว จังหวัดฉะเชิงเทรา พบว่า กลุ่มตัวอย่างมีความชุกของการดื่มเครื่องดื่มแอลกอฮอล์ในระยะแรก ร้อยละ 63.2 อายุเฉลี่ยที่เริ่มดื่มครั้งแรกคือ 12.3 ปี ปัจจัยป้องกันและปัจจัยเสี่ยงที่มีความสัมพันธ์กับการดื่มเครื่องดื่มแอลกอฮอล์ในระยะแรกอย่างมีนัยสำคัญทางสถิติ 0.05

อุทัยทิพย์ จันทรพิณ (2554) ศึกษาปัจจัยที่มีอิทธิพลต่อการดื่มเครื่องดื่มแอลกอฮอล์แบบผิดปกติของ วัยรุ่นในจังหวัดขอนแก่น พบว่า ปัจจัยที่มีอิทธิพลต่อการดื่มเครื่องดื่มแอลกอฮอล์แบบผิดปกติของทั้งวัยรุ่นหญิงและชายในจังหวัดขอนแก่นคือ การแสวงหาความท้าทาย ความโน้มเอียงที่จะดื่ม บรรทัดฐานทางสังคมด้านการดื่มเครื่องดื่มแอลกอฮอล์ ค่านิยมด้านการดื่มเครื่องดื่มแอลกอฮอล์ การสับสนหวั่นไหว และสมรรถนะแห่งตนในการปฏิเสธการดื่มเครื่องดื่มแอลกอฮอล์ ปัจจัยที่แตกต่างกันคือ ความกังวลในการเข้าสังคม และทัศนคติต่อการดื่มเครื่องดื่มแอลกอฮอล์ มีอิทธิพลเฉพาะต่อการดื่มเครื่องดื่มแอลกอฮอล์แบบผิดปกติของวัยรุ่นชาย ความรู้ด้านการดื่มเครื่องดื่มแอลกอฮอล์ การคาดหวังผลในการดื่มเครื่องดื่มแอลกอฮอล์ทางที่ดี การมีเพื่อนดื่มเครื่องดื่มแอลกอฮอล์ มีอิทธิพลเฉพาะต่อการดื่มเครื่องดื่มแอลกอฮอล์แบบผิดปกติของวัยรุ่นหญิง

กฤษณ์ ขุนลึก (2555) ทำการศึกษารูปแบบการส่งเสริมการปฏิบัติตามมาตรการควบคุมเครื่องดื่มแอลกอฮอล์ของผู้จำหน่ายในจังหวัดหนองคาย ผลการศึกษา พบว่า

1. แบบจำลองความสัมพันธ์เชิงสาเหตุ มีความกลมกลืนกับข้อมูลเชิงประจักษ์ โดยตัวแปรเชิงสาเหตุส่งผลโดยตรงต่อตัวแปรการปฏิบัติตามมาตรการควบคุมเครื่องดื่มแอลกอฮอล์ของผู้จำหน่าย ในจังหวัดหนองคาย ได้แก่ ปัจจัยด้านความต้องการขายสินค้า ($\beta = -1.11$) ปัจจัยด้านการตอบสนองความต้องการของผู้ซื้อ ($\beta = -1.01$) ปัจจัยด้านการยอมรับการเปลี่ยนแปลง ($\beta = 0.71$) และปัจจัยด้านค่านิยมเกี่ยวกับเครื่องดื่มแอลกอฮอล์ในเทศกาลบุญต่าง ๆ ($\beta = -0.21$)

2. ผลการประชุมปฏิบัติการ และการพิจารณาของผู้เชี่ยวชาญ ทำให้ได้รูปแบบการส่งเสริมการปฏิบัติตามมาตรการควบคุมเครื่องดื่มแอลกอฮอล์ของผู้จำหน่าย ประกอบด้วย 9 กิจกรรม ได้แก่

- 1) การแต่งตั้งคณะกรรมการกำหนดนโยบายระดับอำเภอ/ตำบล
- 2) การประกาศเป็นนโยบายสาธารณะ
- 3) การประชุมเชิงปฏิบัติการพัฒนาศักยภาพของผู้จำหน่ายเพื่อสังคมอยู่เย็นเป็นสุข



4) ร้านค้าตัวอย่าง

5) การประชาสัมพันธ์ประเด็นสำคัญ ๆ ของมาตรการควบคุมเครื่องดื่มแอลกอฮอล์ให้ประชาชนได้รับทราบ

6) การสร้างแกนนำในการส่งเสริมการปฏิบัติตามมาตรการควบคุมเครื่องดื่มแอลกอฮอล์ในหมู่บ้าน

7) การรณรงค์ให้ความรู้เรื่องมาตรการควบคุมเครื่องดื่มแอลกอฮอล์ในโรงเรียน

8) หมู่บ้านงานบุญปลอดเหล้า

9) การตรวจเยี่ยมร้านค้าโดยคณะเจ้าหน้าที่

3. ผลการทดลองรูปแบบการส่งเสริมการปฏิบัติตามมาตรการควบคุมเครื่องดื่มแอลกอฮอล์ของผู้จำหน่าย ในจังหวัดหนองคาย ที่สร้างขึ้นมีผลทำให้การปฏิบัติตามมาตรการควบคุมเครื่องดื่มแอลกอฮอล์ของผู้จำหน่าย เพิ่มขึ้นมากกว่าก่อนการทดลองอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ทั้งการทดสอบโดยรวม (Multivariate Test) และการทดสอบทีละตัวแปร (Univariate Test)

เกรียงไกร พึ่งเชื้อ (2555) ทำการศึกษาปัจจัยทำนายพฤติกรรมการบริโภคเครื่องดื่มแอลกอฮอล์ของเยาวชนในจังหวัดนครปฐม พบว่า 1) พฤติกรรมการบริโภคเครื่องดื่มแอลกอฮอล์ของเยาวชนในจังหวัดนครปฐมอยู่ในระดับเสี่ยงต่ำหรือตึมน้อยครั้งและตึมปริมาณน้อยในแต่ละครั้งที่ตึม ร้อยละ 37.7 และมีความผิดปกติจากการบริโภคเครื่องดื่มแอลกอฮอล์หรือเครื่องดื่มแอลกอฮอล์ในระดับก่อให้เกิดผลกระทบต่อสุขภาพและการดำเนินชีวิตประจำวัน ร้อยละ 23.3 2) ประสพการณ์การตึมเครื่องดื่มแอลกอฮอล์ เพศชาย การรับข่าวสารการโฆษณาเครื่องดื่มแอลกอฮอล์ และทัศนคติต่อการบริโภคเครื่องดื่มแอลกอฮอล์ สามารถร่วมกันพยากรณ์พฤติกรรมการบริโภคเครื่องดื่มแอลกอฮอล์ของเยาวชนในจังหวัดนครปฐม ได้ร้อยละ 44.4 อย่างมีนัยสำคัญที่ระดับ .001 ส่วนปัจจัยการรับข่าวสาร การรณรงค์เพื่อลดการบริโภคเครื่องดื่มแอลกอฮอล์ การรับรู้การบริโภคเครื่องดื่มแอลกอฮอล์ของบุคคลใกล้ชิด และการรับรู้ผลกระทบที่เกิดขึ้นกับตนเองจากการบริโภคเครื่องดื่มแอลกอฮอล์ของบุคคลใกล้ชิด ไม่รวมทำนายพฤติกรรมการบริโภคเครื่องดื่มแอลกอฮอล์ของเยาวชนในจังหวัดนครปฐม การศึกษาพบว่า การรณรงค์เพื่อลดการบริโภคเครื่องดื่มแอลกอฮอล์ยังไม่สามารถร่วมกันทำนายพฤติกรรมการบริโภคเครื่องดื่มแอลกอฮอล์ของเยาวชนได้ หน่วยงานที่เกี่ยวข้องจึงควรพัฒนารูปแบบการนำเสนอเนื้อหาข่าวสาร รวมถึงช่องทางสื่อสารที่ทำให้เห็นผลเชิงประจักษ์ถึงผลกระทบจากการบริโภคเครื่องดื่มแอลกอฮอล์

ประณต คำนิม และดวงเดือน แซ่ตั้ง (2555) ได้ทำการศึกษาปัจจัยเชิงสาเหตุที่ส่งผลต่อพฤติกรรมการตึมเครื่องดื่มแอลกอฮอล์ของนิสิตมหาวิทยาลัย : กรณีสึกษามหาวิทยาลัยรัฐแห่งหนึ่งพบว่า

1. เพศ สถานที่เรียน และประสพการณ์จูงใจต่อการตึมเครื่องดื่มแอลกอฮอล์ สามารถร่วมกันทำนายพฤติกรรมการตึมเครื่องดื่มแอลกอฮอล์ได้ร้อยละ 27.4 ที่ระดับนัยสำคัญทางสถิติ .01 โดยที่ความน่าจะเป็นที่นิสิตชายจะตึมเครื่องดื่มแอลกอฮอล์เป็น 2.98 เท่าของนิสิตหญิง ความน่าจะเป็นของนิสิตซึ่งเรียนที่องครักษ์จะตึมเครื่องดื่มแอลกอฮอล์เป็น 2.01 เท่าของนิสิตซึ่งเรียนที่ประสานมิตร และถ้าปัจจัยด้านสิ่งแวดล้อมดีเพิ่มขึ้น 1 หน่วย จะทำให้ความน่าจะเป็นของการตึมเครื่องดื่มแอลกอฮอล์เป็น 0.95 เท่าของความน่าจะเป็นในการตึมครั้งก่อน



2. นิสิตที่ดื่มเครื่องดื่มแอลกอฮอล์มีบุคลิกภาพแบบพึ่งพาสูงกว่านิสิตที่ไม่ดื่มเครื่องดื่มแอลกอฮอล์อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่ไม่พบความแตกต่างในเรื่องความไม่มั่นคงทางอารมณ์ระหว่างสองกลุ่มนี้

3. นิสิตที่ไม่ดื่มเครื่องดื่มแอลกอฮอล์มีความสัมพันธ์กับบิดามารดาดีกว่านิสิตที่ดื่มเครื่องดื่มแอลกอฮอล์ที่ระดับนัยสำคัญทางสถิติ .05 แต่ไม่พบความแตกต่างในด้านความสัมพันธ์ระหว่างบิดากับมารดาทั้งสองกลุ่มนี้

4. นิสิตที่ดื่มเครื่องดื่มแอลกอฮอล์มีประสบการณ์ตั้งใจต่อการดื่มเครื่องดื่มแอลกอฮอล์มากกว่านิสิตที่ไม่ดื่มเครื่องดื่มแอลกอฮอล์อย่างมีนัยสำคัญทางสถิติที่ระดับ .001

2.4.2 งานวิจัยต่างประเทศ

Christopher และคณะ (2004) ศึกษาพฤติกรรมการบริโภคเครื่องดื่มแอลกอฮอล์ในนักศึกษาของมหาวิทยาลัยในภาคเหนือของประเทศไทย พบว่า เจตคติต่อการดื่มเครื่องดื่มแอลกอฮอล์สามารถทำนายความตั้งใจที่จะบริโภคเครื่องดื่มแอลกอฮอล์ได้ร้อยละ 16.0

Chalder และคณะ (2006) ศึกษาพฤติกรรมการบริโภคเครื่องดื่มแอลกอฮอล์และแรงจูงใจในการดื่มของวัยรุ่นที่พ่อแม่มีปัญหาจากการดื่มเครื่องดื่มแอลกอฮอล์ โดยทำการศึกษาในกลุ่มตัวอย่างในรัฐ South Wales จำนวน 1,744 คน พบว่า วัยรุ่นที่พ่อแม่มีปัญหาจากการดื่มแอลกอฮอล์เกือบ 1 ใน 5 ของกลุ่มตัวอย่างและพบว่าเป็นกลุ่มที่ดื่มถี่ ดื่มหนัก และดื่มลำพังเพียงคนเดียวมากกว่ากลุ่มวัยรุ่นที่พ่อแม่ไม่มีปัญหาจากการดื่มเครื่องดื่มแอลกอฮอล์ พ่อแม่ที่มีปัญหาจากการดื่มเครื่องดื่มแอลกอฮอล์สัมพันธ์กับแรงจูงใจภายใน (เลียนแบบ) ที่จะดื่มแอลกอฮอล์ของวัยรุ่น นอกจากนี้ยังพบว่าแรงจูงใจภายในที่จะดื่มเครื่องดื่มแอลกอฮอล์กับพ่อแม่ที่มีปัญหาจากการดื่มเครื่องดื่มแอลกอฮอล์สามารถร่วมกันทำนายพฤติกรรมการดื่มเครื่องดื่มแอลกอฮอล์และความถี่ในการดื่มได้

Duncan et al. (2006) ได้ศึกษาเรื่องการใช้เครื่องดื่มแอลกอฮอล์ในกลุ่มอายุ 9-16 ปี กลุ่มตัวอย่างจำนวน 405 คน เป็นผู้หญิง ร้อยละ 48.4 การศึกษานี้ครอบคลุมถึงการใช้เครื่องดื่มแอลกอฮอล์ในครอบครัว ปัญหาจากเครื่องดื่มแอลกอฮอล์ในครอบครัว มีเพื่อนที่เบี่ยงเบน มีเพื่อนที่ใช้เครื่องดื่มแอลกอฮอล์ และการสนับสนุนหรือการชักชวนให้ดื่มแอลกอฮอล์จากเพื่อน ผลการศึกษาพบว่า สัดส่วนของการใช้เครื่องดื่มแอลกอฮอล์เพิ่มขึ้นในกลุ่มที่ศึกษาตั้งแต่อายุ 9-16 ปี โดยมีนัยสำคัญทางสถิติที่พบมีทั้งที่ขัดแย้งและไปในทิศทางเดียวกัน ในกลุ่มผู้หญิงและกลุ่มที่มีผู้ปกครองใช้เครื่องดื่มแอลกอฮอล์ในระดับสูงมีความสัมพันธ์อย่างมากกับอัตราการเริ่มดื่มเครื่องดื่มแอลกอฮอล์และในทางกลับกันกลุ่มที่มีสนับสนุนหรือการชักชวนให้ดื่มเครื่องดื่มแอลกอฮอล์กับมีความสัมพันธ์ในระดับต่ำกับอัตราการเริ่มดื่มเครื่องดื่มแอลกอฮอล์ นอกจากนี้ยังพบว่า การมีเพื่อนที่เบี่ยงเบนและการมีเพื่อนที่สนับสนุนหรือการชักชวนให้ดื่มเครื่องดื่มแอลกอฮอล์มีความสัมพันธ์กับการเพิ่มอัตราการเครื่องดื่มแอลกอฮอล์ จากอายุ 9-16 ปี

Michael (2006) ศึกษาความสัมพันธ์ระหว่างเจตคติและการคล้อยตามกลุ่มอ้างอิงกับความตั้งใจที่จะบริโภคเครื่องดื่มแอลกอฮอล์ในผู้บริโภคเครื่องดื่มแอลกอฮอล์อายุมากกว่า 18 ปีขึ้นไป จำนวน 48 คน ในออสเตรเลียได้ พบว่าเจตคติต่อการบริโภคเครื่องดื่มแอลกอฮอล์สามารถทำนายความตั้งใจที่จะบริโภคเครื่องดื่มแอลกอฮอล์ได้ร้อยละ 43.0 และการคล้อยตามกลุ่มอ้างอิงสามารถทำนายความตั้งใจที่จะดื่มเครื่องดื่มแอลกอฮอล์ได้ร้อยละ 53.0



Mukta และ Usha (2009) ได้ทำการศึกษาเปรียบเทียบผลการวิจัยที่ใช้โครงข่ายประสาทเทียมกับเทคนิคการวิเคราะห์ ผลทางสถิติโดยทั่วไป ซึ่งได้แก่ การวิเคราะห์ความถดถอยพหุคูณ การวิเคราะห์ความถดถอยโลจิสติก และการจำแนกกลุ่มด้วย Discriminate Analysis จากจำนวนงานวิจัยทั้งหมด 96 งานวิจัย ปรากฏว่า โครงข่ายประสาทเทียมให้ผลการพยากรณ์ที่ดีกว่าจำนวน 58 เปอร์เซนต์ โครงข่ายประสาทเทียมกับเทคนิคการวิเคราะห์ผลทางสถิติโดยทั่วไป ให้ผลการพยากรณ์เหมือนกันจำนวน 24 เปอร์เซนต์ และเทคนิคการวิเคราะห์ผลทางสถิติโดยทั่วไป ให้ผลการพยากรณ์ที่ดีกว่าจำนวน 18 เปอร์เซนต์

Steingold (2009) ศึกษาปัจจัยทำนายความตั้งใจสูบบุหรี่และบริโภคเครื่องดื่มที่มีแอลกอฮอล์ในนักเรียนแอฟริกาตอนใต้ พบว่า เจตคติต่อการบริโภคเครื่องดื่มที่มีแอลกอฮอล์เป็นตัวทำนายความตั้งใจที่จะบริโภคเครื่องดื่มที่มีแอลกอฮอล์ร้อยละ 7.70 คล้อยตามกลุ่มอ้างอิงเป็นตัวทำนายความตั้งใจที่จะดื่มเครื่องดื่มแอลกอฮอล์ร้อยละ 37.10 และการรับรู้ความสามารถในการควบคุมพฤติกรรมสามารถทำนายความตั้งใจที่จะบริโภคเครื่องดื่มแอลกอฮอล์ร้อยละ 10.83



บทที่ 3

วิธีดำเนินการวิจัย

การวิจัย เรื่องการวิเคราะห์การตีมูลค่าของประชาชนในจังหวัดหนองคาย โดยเทคนิคการทำเหมืองข้อมูลและการถดถอยโลจิสติก ได้กำหนดวิธีดำเนินการวิจัยตามขั้นตอนดังต่อไปนี้

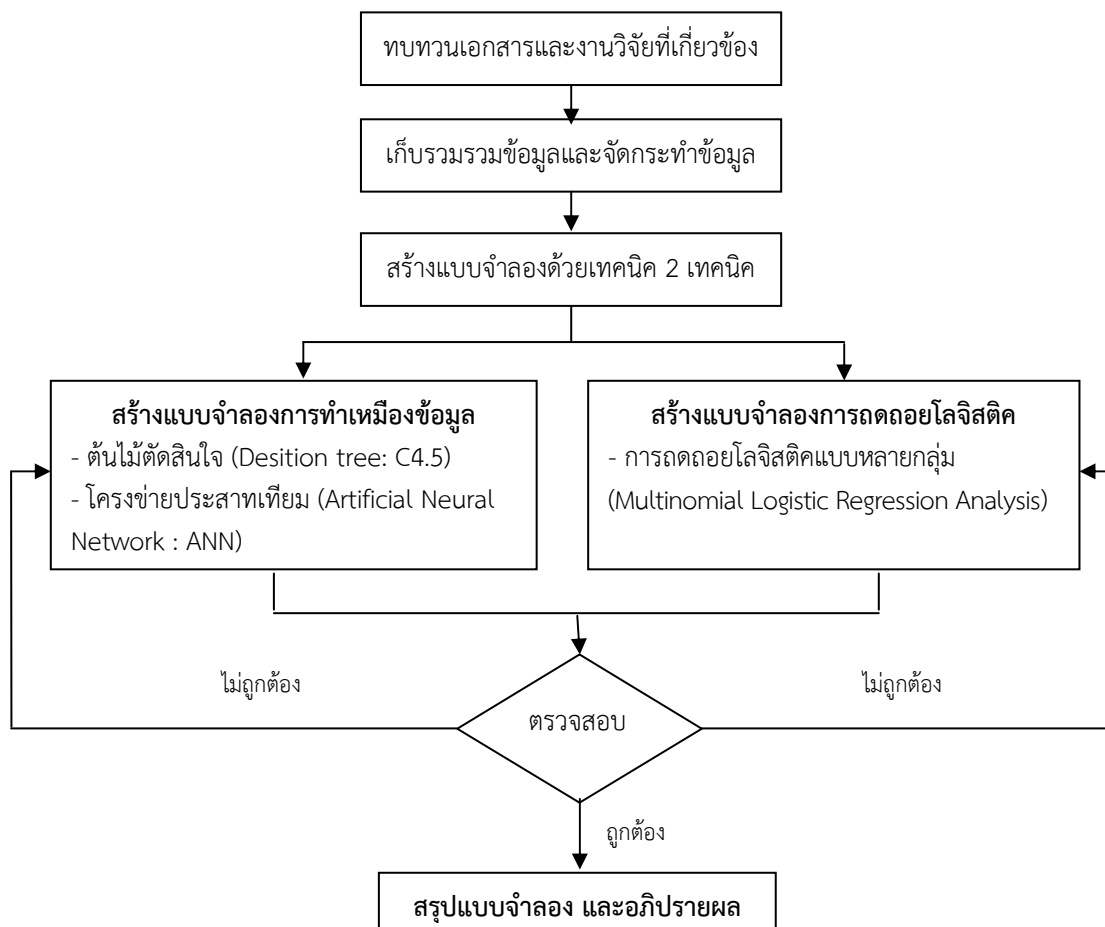
- 3.1 ขั้นตอนในการดำเนินการศึกษาวิจัย
- 3.2 กรอบแนวคิดในการวิจัย
- 3.3 เครื่องมือที่ใช้ในการวิจัย
- 3.4 การเก็บรวบรวมข้อมูล
- 3.5 การจัดกระทำและการวิเคราะห์ข้อมูล
- 3.6 การเปรียบเทียบประสิทธิภาพของแบบจำลอง

3.1 ขั้นตอนในการดำเนินการศึกษาวิจัย

งานวิจัยนี้มีขั้นตอนการดำเนินการดังแสดงดังภาพประกอบ 3.1 ดังนี้

1. ทบทวนเอกสารและงานวิจัยที่เกี่ยวข้องกับการวิเคราะห์เกี่ยวกับตีมูลค่าของประชาชน รวมถึงวิธีการต่าง ๆ ในการสำรวจข้อมูลและวิเคราะห์ข้อมูล
2. รวบรวมข้อมูลที่ใช้ในการวิจัย โดยการวิจัยครั้งนี้ใช้เป็นข้อมูลทุติยภูมิ (Secondary Data) ที่ได้จากการสำรวจพฤติกรรมการสูบบุหรี่และตีมูลค่าของประชากร พ.ศ. 2554 สำรวจโดยสำนักงานสถิติแห่งชาติ
3. จัดเตรียมข้อมูลให้เหมาะสมกับการสร้างแบบจำลอง โดยใช้โปรแกรมคอมพิวเตอร์ที่เหมาะสมในการจัดเตรียมข้อมูล
4. สร้างแบบจำลองการวิเคราะห์เกี่ยวกับตีมูลค่าของประชาชนในจังหวัดหนองคายด้วยเทคนิคการทำเหมืองข้อมูล และเทคนิคการถดถอยโลจิสติก
5. สรุปผลการวิจัยและจัดทำข้อเสนอแนะ



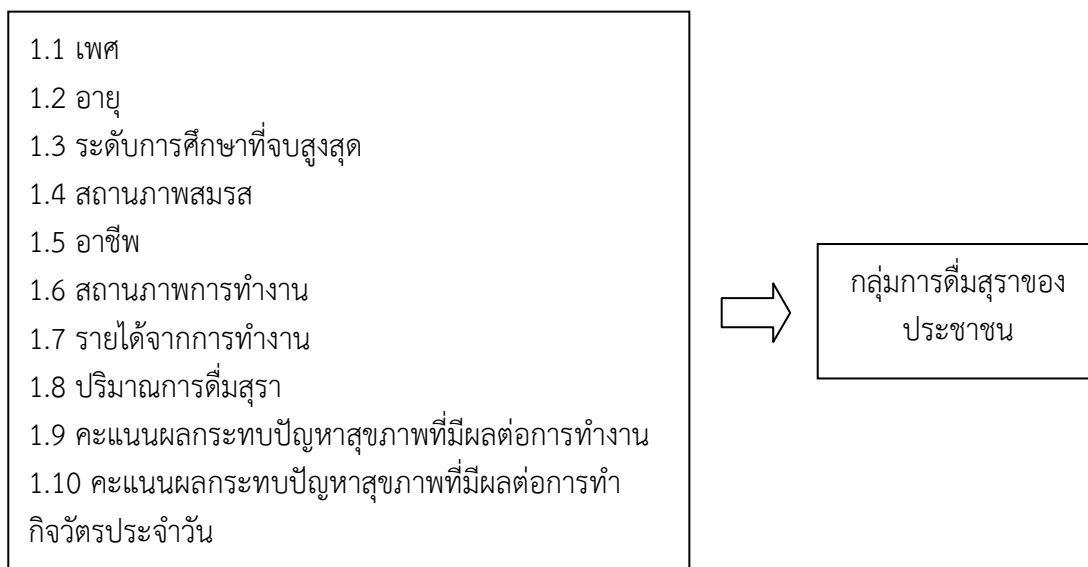


ภาพประกอบ 3.1 ขั้นตอนในการดำเนินการศึกษาวิจัย



3.2 กรอบแนวคิดในการวิจัย

จากการศึกษาทฤษฎี แนวคิด และงานวิจัยที่เกี่ยวข้อง ผู้วิจัยได้กำหนดกรอบแนวคิดในการวิจัย ดังนี้



ภาพประกอบ 3.2 กรอบแนวคิดในการวิจัย

ตัวแปรในการวิจัยมีจำนวนทั้งสิ้น 11 ตัวแปร จำแนกเป็นตัวแปรอิสระ 10 ตัว ตัวแปรตาม 1 ตัว ดังนี้

1. ตัวแปรอิสระ

เพศ

อายุ

ระดับการศึกษาที่จบสูงสุด

สถานภาพสมรส

อาชีพ

สถานภาพการทำงาน

รายได้จากการทำงาน

ปริมาณการดื่มสุรา

คะแนนผลกระทบปัญหาสุขภาพที่มีผลต่อการทำงาน

คะแนนผลกระทบปัญหาสุขภาพที่มีผลต่อการทำกิจกรรมประจำวัน

2. ตัวแปรตาม

การดื่มสุรา ใน 12 เดือนที่แล้ว แบ่งออกเป็น 3 กลุ่ม คือ

1) กลุ่มที่ไม่เคยดื่มสุรา

2) กลุ่มที่ดื่มเล็กน้อยหรือนาน ๆ ครั้ง

3) กลุ่มที่ดื่มสม่ำเสมอ



3.3 เครื่องมือที่ใช้ในการวิจัย

ในการวิจัยครั้งนี้ใช้โปรแกรมคอมพิวเตอร์สำเร็จรูปที่เหมาะสมกับการสร้างแบบจำลอง ด้วยวิธีการต่าง ๆ ซึ่งได้แก่ การสร้างแบบจำลองการวิเคราะห์การถดถอยของประชากรในจังหวัดหนองคาย ด้วยเทคนิคการทำเหมืองข้อมูล โดยวิธีต้นไม้ตัดสินใจ และวิธีโครงข่ายประสาทเทียม การเรียนรู้แบบแพร่กลับ กับเทคนิคการถดถอยโลจิสติก

3.4 การเก็บรวบรวมข้อมูล

การวิจัยครั้งนี้ใช้ข้อมูลทุติยภูมิ (Secondary Data) ที่ได้จากการสำรวจพฤติกรรม การสูบบุหรี่ และดื่มสุราของประชากร พ.ศ. 2554 สำรวจโดยสำนักงานสถิติแห่งชาติ โดยผู้วิจัยได้คัดเลือกข้อมูลที่ใช้ในการวิจัยทั้งสิ้นจำนวน 605 ราย จากการเก็บรวบรวมข้อมูลโดยเจ้าหน้าที่ของสำนักงานสถิติจังหวัดหนองคาย ไปสัมภาษณ์ข้อมูลครัวเรือนตัวอย่าง โดยสัมภาษณ์หัวหน้าครัวเรือนหรือสมาชิกในครัวเรือนในเรื่องของพฤติกรรมการสูบบุหรี่และดื่มสุรา ในระหว่างวันที่ 1 – 12 ของเดือน มีนาคม – พฤษภาคม พ.ศ. 2554

3.5 การจัดกระทำและการวิเคราะห์ข้อมูล

3.5.1 การทดสอบการสร้างแบบจำลองการทำเหมืองข้อมูล

3.5.1.1 การทดสอบการสร้างแบบจำลองต้นไม้ตัดสินใจ

การสร้างแบบจำลองต้นไม้ตัดสินใจ ด้วยอัลกอริทึม C4.5 (J48) มีขั้นตอนการดำเนินการดังต่อไปนี้

1. การเตรียมข้อมูลสำหรับการสร้างแบบจำลอง

จากชุดข้อมูลจำนวน 605 ราย ที่ได้ทำการคัดเลือกจากการสำรวจพฤติกรรม การสูบบุหรี่และดื่มสุราของประชากร พ.ศ. 2554 สำรวจโดยสำนักงานสถิติแห่งชาติ โดยทำการแปลงข้อมูลที่เลือกมาให้อยู่ในรูปแบบที่เหมาะสมสำหรับการนำไปใช้วิเคราะห์ตามอัลกอริทึม หลังจากได้ทำการ Cleaning ข้อมูลเป็นที่เรียบร้อยแล้วนั้น ผู้วิจัยได้ทำการแปลงรูปแบบของข้อมูลเพื่อให้เหมาะสมสำหรับการสร้างตัวแบบ ดังตารางที่ 3.1



ตาราง 3.1 ค่าของตัวแปรที่ทำการแปลงรูปแบบข้อมูล

ชื่อตัวแปร	ค่าข้อมูลเดิม	ค่าข้อมูลใหม่
เพศ (A4)	1 = ชาย 2 = หญิง	ใช้ค่าข้อมูลเดิม
อายุ (A5_RE)	Numeric	1 = 15 – 24 ปี 2 = 25 – 59 ปี 3 = 60 ปีขึ้นไป
ระดับการศึกษาที่จบสูงสุด (A6_RE)	1=ไม่เคยเรียน 2=ต่ำกว่าประถมศึกษา 3=ประถมศึกษา 4= มัธยมศึกษาตอนต้น 5=มัธยมศึกษาตอนปลาย สายสามัญ 6=มัธยมศึกษาตอนปลาย สายอาชีพ 7=ปวส./ปวท./ อนุปริญญา 8=ปริญญาตรี 9= สูงกว่าปริญญาตรี	-รหัส 1,2,3 แปลงค่าเป็น 1 =ประถมศึกษาหรือต่ำกว่า -รหัส 5 แปลงค่าเป็น 2=มัธยมศึกษาตอนปลายสายสามัญ -รหัส 6,7 แปลงค่าเป็น 3=มัธยมศึกษาตอนปลายสายสามัญ, ปวส./ปวท./อนุปริญญา -รหัส 8,9 แปลงค่าเป็น 4=ปริญญาตรีและสูงกว่า
สถานภาพสมรส (A7_RE)	1=โสด 2=สมรส 3=หม้าย 4=หย่า 5=แยกกันอยู่ 6=เคยสมรสแต่ไม่ทราบ สถานภาพ	-รหัส 1,3,4,5,6 แปลงค่าเป็น 1= โสด -รหัส 2 ใช้ค่าเดิม



ตาราง 3.1 (ต่อ)

ชื่อตัวแปร	ค่าข้อมูลเดิม	ค่าข้อมูลใหม่
อาชีพ (A8_RE)	1=เกษตรและประมง 2=ช่างเทคนิค ช่างฝีมือ เจ้าหน้าที่เทคนิค 3=ทำงานภาคบริการและ ผู้จำหน่ายสินค้า 4= กลุ่มผู้บริหาร ข้าราชการ และ ผู้ประกอบการวิชาชีพต่างๆ 5=ไม่ทำงาน	ใช้ค่าเดิม
สถานภาพการทำงาน (A10_RE)	1= นายจ้าง 2= ผู้ประกอบธุรกิจของ ตนเอง 3= ช่วยธุรกิจในครัวเรือน 4= ลูกจ้างรัฐบาล 5= ลูกจ้างรัฐวิสาหกิจ 6= ลูกจ้างเอกชน Missing = ไม่ทำงาน	-รหัส 1,2 แปลงค่าเป็น 1= นายจ้างและผู้ประกอบธุรกิจ ตนเอง -รหัส 3 แปลงค่าเป็น 2= ช่วยธุรกิจในครัวเรือน -รหัส 3,4 แปลงค่าเป็น 3=ลูกจ้างรัฐบาลและรัฐวิสาหกิจ -รหัส 6 แปลงค่าเป็น 4= ลูกจ้างเอกชน -ค่า Missing แปลงค่าเป็น 5=ไม่ทำงาน
รายได้จากการทำงาน (Inc)	Numeric	1 = < 2,000 บาทต่อเดือน 2 = 2,000-4,999 บาทต่อเดือน 3 = 5,000-9,999 บาทต่อเดือน 4 = 10,000-19,999 บาทต่อเดือน 5 = 20,000 บาทขึ้นไปต่อเดือน



ตาราง 3.1 (ต่อ)

ชื่อตัวแปร	ค่าข้อมูลเดิม	ค่าข้อมูลใหม่
ปริมาณการดื่มสุรา (B11_31_L)	Numeric	1 = 0.01–14.99 ลิตรต่อเดือน 2 = 15.00–29.99 ลิตรต่อเดือน 3 = 30 ลิตรขึ้นไปต่อเดือน 4 = ไม่เคยดื่ม
คะแนนผลกระทบปัญหา สุขภาพที่มีผลต่อการทำงาน (A42)	Numeric	ใช้ค่าข้อมูลเดิม
คะแนนผลกระทบปัญหา สุขภาพที่มีผลต่อการทำ กิจวัตรประจำวัน (A43)	Numeric	ใช้ค่าข้อมูลเดิม
การดื่มสุรา (B1_RE)	1= ไม่เคยดื่มเลยในชีวิตนี้ 2=เคยดื่ม แต่ไม่ดื่มใน 12 เดือนที่แล้ว 3=ทุกวัน (7 วัน/สัปดาห์) 4=เกือบทุกวัน (5-6 วัน/ สัปดาห์) 5=วันเว้นวัน (3-4 วัน/ สัปดาห์) 6=ทุกสัปดาห์ (1-2 วัน/ สัปดาห์) 7= นานๆ ครั้ง 8-11 วัน/ปี 8= นานๆ ครั้ง 4-7 วัน/ปี 9=นานๆ ครั้ง 1-3 วัน/ปี	-รหัส 3,4,5,6 แปลงค่าเป็น 1 = กลุ่มที่ดื่มสม่ำเสมอ -รหัส 7,8,9 แปลงค่าเป็น 2 = กลุ่มที่ดื่มเล็กน้อยหรือ นานๆ ครั้ง -รหัส 1,2 แปลงค่าเป็น 3 = กลุ่มที่ไม่เคยดื่มสุรา

จากนั้นนำข้อมูลมาเลือกอย่างสุ่มโดยใช้โปรแกรมสำเร็จรูปทางสถิติในการเลือก
จำนวน 5 ชุด โดยแบ่งข้อมูลออกเป็น 2 ส่วน คือ ส่วนแรกร้อยละ 80 ของจำนวนข้อมูลทั้งหมดใช้ในทุกชุด
ของการฝึกฝน และส่วนที่ 2 ร้อยละ 20 ให้ในชุดของการทดสอบ โดยผู้วิจัยได้นำชุดข้อมูลที่เตรียมไว้



จำนวน 5 ชุดนี้มาทดสอบการจำแนกข้อมูลด้วยโปรแกรม Weka 3.6.3 (Waikato Environment for Knowledge Analysis) พัฒนาโดย University of Waikato New Zealand เป็นโปรแกรมที่ใช้ในการทำ Data Mining และในโปรแกรมนี้นี้มีฟังก์ชันการสร้างแบบจำลองต้นไม้ตัดสินใจที่ใช้อัลกอริทึม C4.5 หรือที่รู้จักกันในชื่อ J48 ได้จำนวน 10 รูปแบบ เพื่อคัดเลือกรูปแบบเงื่อนไขการจำแนกกลุ่มการตัดสินใจที่ให้ค่าความถูกต้องสูงที่สุด โดยกำหนดให้กลุ่มการตัดสินใจเป็นแอตทริบิวต์เป้าหมาย ในการแบ่งข้อมูลสำหรับเรียนรู้ (Training Data) และข้อมูลสำหรับทดสอบ (Testing Data) ผู้วิจัยได้กำหนดเป็น 4 ชุด คือ

X1 คือ การกำหนดจำนวนข้อมูล Training Data และ Testing Data เป็น 50/50

X2 คือ การกำหนดจำนวนข้อมูล Training Data และ Testing Data เป็น 60/40

X3 คือ การกำหนดจำนวนข้อมูล Training Data และ Testing Data เป็น 80/20

X4 คือ การกำหนดจำนวนข้อมูล Training Data และ Testing Data เป็น 90/10

สำหรับขั้นตอนในการสร้างต้นไม้ตัดสินใจ มีการกำหนดค่าพารามิเตอร์ ดังนี้

1. Confidences Factor (CF) คือค่าความเชื่อมั่นที่ใช้สำหรับการเลือกวิธีการตัดกิ่งต้นไม้แบบ Pruning แบบ Error Base Pruning ซึ่งเทคนิคนี้เหมาะกับการสร้างต้นไม้ที่มีจำนวนข้อมูลไม่มาก โดยค่า CF คำนวณจาก Error Rate ได้จากการทดสอบกับ Test Set การเพิ่มค่า CF นี้จะทำให้ได้แบบจำลองต้นไม้ตัดสินใจที่มีลักษณะเฉพาะ แต่ถ้าลดค่า CF ลง จะทำให้แบบจำลองต้นไม้ตัดสินใจที่มีลักษณะทั่วไปมากขึ้น งานวิจัยนี้กำหนดค่า Confidences Factor (CF) = 0.25

2 Minimum Number Pruning (M) คือจำนวนข้อมูลที่น้อยที่สุดต่อ 1 Leaf ซึ่งต้องกำหนดเป็นจำนวนเต็มที่มีค่ามากกว่า 0 ในงานวิจัยนี้กำหนดให้มีค่าเท่ากับ 2

3. Fold คือ จำนวนข้อมูลที่ใช้ในการตัดกิ่ง ในงานวิจัยนี้กำหนดเท่ากับ 3

4. Binary Splits (Binary) คือ การสร้างต้นไม้แบบ Binary Tree เพื่อลดขนาดของต้นไม้ไม่ให้ซับซ้อนเกินไป ตัวเลือกสำหรับพารามิเตอร์นี้ คือ True, False ในงานวิจัยนี้กำหนดให้มีค่าทั้ง 2 ค่า

2. การทดสอบประสิทธิภาพในการสร้างแบบจำลองต้นไม้ตัดสินใจ จะทำการวัดประสิทธิภาพจาก

1. ขนาดของต้นไม้ จำนวน Node และ Leaf
2. ร้อยละความถูกต้องในการจำแนกกลุ่ม

$$\text{Correction} = \frac{\text{จำนวนข้อมูลฝึกฝนที่จำแนกได้}}{\text{จำนวนข้อมูลฝึกฝนทั้งหมด}} \times 100$$

3. ค่าความระลึก (Recall) คำนวณจากค่าของข้อมูลที่มีผลลัพธ์ถูกต้องโดยพิจารณาจากข้อมูลของผลลัพธ์เดียวกัน



$$\text{Recall} = \frac{\text{จำนวนข้อมูลที่จำแนกได้ถูกต้องของกลุ่มนั้น}}{\text{จำนวนข้อมูลที่อยู่ในกลุ่มนั้นจริง}}$$

4. ค่าความแม่นยำ (Precision) คำนวณจากค่าของข้อมูลที่มีผลลัพธ์ถูกต้องโดยพิจารณาจากจำนวนข้อมูลทั้งหมดที่ถูกจำแนกมีผลลัพธ์เดียวกัน

$$\text{Precision} = \frac{\text{จำนวนข้อมูลที่จำแนกได้ถูกต้องของกลุ่มนั้น}}{\text{จำนวนข้อมูลที่ถูกจำแนกว่าเป็นกลุ่มนั้นทั้งหมด}}$$

5. ค่าความเหวี่ยง (F-Measure) คำนวณจากค่าความแม่นยำ (P) และค่าความระลึก (R)

$$F(R, P) = \frac{2PR}{P + R}$$

3.5.1.2 ผลการสร้างแบบจำลองโครงข่ายประสาทเทียม (Artificial Neuron Network: ANN)

การสร้างแบบจำลองโครงข่ายประสาทเทียม (Artificial Neuron Network: ANN) มีขั้นตอนการดำเนินการดังต่อไปนี้

1. การศึกษาสัมประสิทธิ์สหสัมพันธ์และคัดเลือกตัวแปร

ในการศึกษาครั้งนี้ได้วิเคราะห์ความสัมพันธ์ระหว่างตัวแปรอิสระทั้ง 10 ตัวแปร คือ เพศ อายุ ระดับการศึกษาที่จบสูงสุด สถานภาพสมรส อาชีพ สถานภาพการทำงาน รายได้จากการทำงาน ปริมาณการดื่มสุรา คะแนนผลกระทบปัญหาสุขภาพที่มีผลต่อการทำงานและคะแนนผลกระทบปัญหาสุขภาพที่มีผลต่อการทำกิจวัตรประจำวัน กับตัวแปรตาม คือ กลุ่มการดื่มสุรา โดยใช้การหาความสัมพันธ์ระหว่างตัวแปรอิสระกับตัวแปรตาม 2 วิธี คือ

1) สัมประสิทธิ์ แบบ Cramer's V ทำการหาความสัมพันธ์ระหว่างตัวแปรอิสระ คือ เพศ ระดับการศึกษาที่จบสูงสุด สถานภาพสมรส อาชีพ สถานภาพการทำงาน กับ ตัวแปรตาม คือ กลุ่มการดื่มสุรา โดยใช้สูตร

$$V = \sqrt{\frac{x^2}{n(q - 1)}}$$

เมื่อ q = จำนวนแถวอน (r) หรือแถวตั้ง (c) ที่มีค่าน้อยที่สุด

n = จำนวนตัวอย่างทั้งหมด

โดยค่าของ V จะมีค่าระหว่าง 0 ถึง 1



2) สัมประสิทธิ์ แบบ Eta ทำการหาความสัมพันธ์ระหว่างตัวแปรอิสระ คือ รายได้จากการทำงาน ปริมาณการดื่มสุรา คะแนนผลกระทบปัญหาสุขภาพที่มีผลต่อการทำงาน คะแนนผลกระทบปัญหาสุขภาพที่มีผลต่อการทำกิจกรรมประจำวัน กับ ตัวแปรตาม คือ กลุ่มการดื่มสุรา โดยใช้สูตร

$$Eta = 1 - \frac{\sum_{i=1}^n (Y_i - \bar{Y}_k)^2}{\sum_{i=1}^n (Y_i - \bar{Y}_T)^2}$$

เมื่อ $(Y_i - \bar{Y}_k)^2$ = คะแนนเบี่ยงเบนจากค่าเฉลี่ยในแต่ละกลุ่ม
 $(Y_i - \bar{Y}_T)^2$ = คะแนนเบี่ยงเบนจากค่าเฉลี่ยรวม

2. กำหนดชุดข้อมูลนำเข้า (Input) ปัจจัยต่าง ๆ ที่มีอิทธิพลต่อการพยากรณ์ของโครงข่ายประสาทเทียม ในการพิจารณาการนำเข้าข้อมูลของวิธีโครงข่ายประสาทเทียมประกอบด้วย neuron input จำนวน 10 neuron โดยแต่ละ neuron ใช้สำหรับสัญญาณ input ต่างๆ สำหรับข้อมูล input จำนวน 605 ราย แยกเป็นใช้ในการเรียนรู้จะเลือกข้อมูลร้อยละ 90 หรือจำนวน 545 รายและข้อมูลที่ใช้ในการทดสอบโครงข่ายประสาทจะเลือกข้อมูลร้อยละ 10 หรือจำนวน 60 ราย และมีการกำหนดชุดข้อมูลของตัวแปรอิสระออกเป็น 3 ชุด คือ

ANN1 คือ กลุ่มตัวแปรอิสระที่ 1 ประกอบด้วยตัวแปรอิสระทั้งหมด 10 ตัว

ANN2 คือ กลุ่มตัวแปรอิสระที่ 2 ประกอบด้วยตัวแปรอิสระ มีค่าความสัมพันธ์ตั้งแต่ 0.50 ขึ้นไป

ANN3 คือ กลุ่มตัวแปรอิสระที่ 3 ประกอบด้วยตัวแปรอิสระ มีค่าความสัมพันธ์ต่ำกว่า 0.50

3. การออกแบบสถาปัตยกรรมโครงข่าย ในการกำหนดโครงสร้างของเครือข่ายประสาทเทียมไม่มีวิธีการที่แน่นอน โดยทั่วไปทำโดยวิธีการสุ่มค่าพารามิเตอร์ต่างๆ อันได้แก่ ค่าอัตราการเรียนรู้ ซึ่งเป็นค่าที่ควบคุมความคงตัว (Stability) ของค่าถ่วงน้ำหนักและความเร็วในการปรับค่าถ่วงน้ำหนักให้เข้าสู่ค่าที่สมดุล โดยอัตราการเรียนรู้มีค่าอยู่ระหว่าง 0 ถึง 1 โดยค่าถ่วงน้ำหนักที่เหมาะสมจะกำหนดให้อัตราการเรียนรู้มีค่าน้อยๆ อยู่ในช่วง 0.1 - 0.3 (อัญชลี วานิชทวีวัฒน์, 2540), ค่าโมเมนตัม เป็นค่าที่มีอิทธิพลต่อการเกิดค่า Error ของโครงข่าย โดยค่าถ่วงน้ำหนักไม่ควรเป็นค่าที่ใหญ่มากนัก โดยทั่วไปค่าถ่วงน้ำหนักเริ่มต้นจะสุ่มค่าระหว่าง -0.5 ถึง 0.5 (หรือ ระหว่าง -1 ถึง 1 ตามความเหมาะสม) (สุรยุทธ์ ปรีชา, 2541) และจำนวนครั้งที่ทำการฝึกฝนถ้ากำหนดให้จำนวนรอบในการฝึกฝนมีค่าสูงเกินไป ไม่ได้ทำให้โครงข่ายประสาทเทียมเรียนรู้ได้ดีมากขึ้น ควรประมาณการจำนวนรอบการสอนที่น้อยที่สุดที่จำเป็นสำหรับแก้ปัญหาหนึ่งๆ ที่ทำให้โครงข่ายเกิดการเรียนรู้ (ศุภัทรา สุนทรภักย์, 2539) จำนวนรอบการฝึกฝนอาจใช้ได้ตั้งแต่ 20,000 ถึง 100,000 รอบ ขึ้นอยู่กับลักษณะของโครงข่าย (ทัตดาว แบนเนียน, 2545)

งานวิจัยครั้งนี้ ผู้วิจัยได้ทำการสร้างและทดสอบตัวแบบการพยากรณ์ โดยใช้โปรแกรม Weka 3.6.6 ซึ่งมีการกำหนดค่าพารามิเตอร์ต่างๆ คือ จำนวนโหนดในชั้นซ่อน 9 ค่าอัตรา



การเรียนรู้ 0.2 ค่าโมเมนตัม 0.3 จำนวนรอบในการเรียนรู้ 50,000 รอบ โดยเป็นโครงข่ายประสาทเทียมแบบชั้นเดียว มีข้อมูลนำเข้า 545 ราย ส่งผ่านด้วยฟังก์ชันถ่ายโอน Log-Sigmoid และส่งถ่ายข้อมูลด้วยกระตุ้นแบบผลรวมเชิงเส้น Pure-Linear สู่ชั้นผลลัพธ์ คือกลุ่มการดื่มสุรา

4. ประสิทธิภาพของแบบจำลองโครงข่ายประสาทเทียม จะทำการวัดประสิทธิภาพจากการวัดความคลาดเคลื่อนระหว่างผลการทำนายและข้อมูลจริงจากชุดทดสอบ คือ ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Square Error: RMSE) และค่าร้อยละสัมบูรณ์ความคลาดเคลื่อนเฉลี่ย (Mean Absolute Percentage Error: MAPE)

1) ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Square Error: RMSE) คำนวณได้ดังสมการ

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (Predict_i - Actual_i)^2}$$

2) ค่าร้อยละสัมบูรณ์ความคลาดเคลื่อนเฉลี่ย (Mean Absolute Percentage Error: MAPE) คำนวณได้ดังสมการ

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{Predict_i - Actual_i}{Actual_i} \right| \times 100$$

โดยที่ Actual = ข้อมูลจริง

Predict = ข้อมูลที่ได้จากการพยากรณ์

N = จำนวนชุดข้อมูลที่ใช้ในการคำนวณ

3) ร้อยละความถูกต้องในการจำแนกกลุ่ม

$$\text{Correction} = \frac{\text{จำนวนข้อมูลฝึกฝนที่จำแนกได้}}{\text{จำนวนข้อมูลฝึกฝนทั้งหมด}} \times 100$$

3.5.2 การทดสอบแบบจำลองการถดถอยโลจิสติก

3.5.2.1 ตัวแปรที่เกี่ยวข้อง

ตัวแปรที่นำมาพิจารณาในการสร้างแบบจำลองการวิเคราะห์การดื่มสุราของประชาชนในจังหวัดหนองคาย โดยใช้การถดถอยโลจิสติกแบบหลายกลุ่ม (Multinomial Logistic Regression Analysis) โดยตัวแปรตาม คือ กลุ่มของการดื่มสุรา ส่วนตัวแปรอิสระใช้ตัวแปรทั้งหมด 10 ตัว คือ เพศ อายุ ระดับการศึกษาที่จบสูงสุด สถานภาพสมรส อาชีพ สถานภาพการทำงาน รายได้จากการทำงาน



ปริมาณการดื่มสุรา คะแนนผลกระทบปัญหาสุขภาพที่มีผลต่อการทำงานและคะแนนผลกระทบปัญหาสุขภาพที่มีผลต่อการทำกิจกรรมประจำวัน มาทำการวิเคราะห์เพื่อหาสมการในการพยากรณ์

3.5.2.2 โครงสร้างแบบจำลอง

สำหรับแบบจำลองการวิเคราะห์การดื่มสุราของประชาชนในจังหวัดหนองคายโดยใช้การถดถอยโลจิสติกแบบหลายกลุ่ม (Multinomial Logistic Regression Analysis) ในการสร้างแบบจำลอง โดยตัวแปรอิสระที่นำมาใช้เป็นองค์ประกอบของแบบจำลองซึ่งจะปรากฏอยู่ในฟังก์ชัน ดังนี้

$$G(1) = \beta_0 + \beta_{A4}X_{A4} + \beta_{A5_re}X_{A5_re} + \dots + \beta_{B43}X_{B43}$$

$$G(2) = \beta_0 + \beta_{A4}X_{A4} + \beta_{A5_re}X_{A5_re} + \dots + \beta_{B43}X_{B43}$$

โดยที่ $G(1)$	คือ สมการพยากรณ์กลุ่มการดื่มสม่ำเสมอเทียบกับกลุ่มไม่เคยดื่มสุรา
$G(2)$	คือ สมการพยากรณ์กลุ่มการดื่มเล็กน้อยหรือนานๆครั้งเทียบกับกลุ่มไม่เคยดื่มสุรา
$\beta_0, \beta_{A4}, \beta_{A5_RE}, \dots, \beta_{A43}$	คือ สัมประสิทธิ์ที่แสดงอิทธิพลของตัวแปรตัวที่ i ที่มีต่อกลุ่มการดื่มสุรา
$X_{A4}, X_{A5_RE}, \dots, X_{A43}$	คือ ตัวแปรตัวที่ i ที่มีอิทธิพลต่อกลุ่มการดื่มสุรา

แบบจำลองวิเคราะห์ความน่าจะเป็นที่พยากรณ์กลุ่มของรายใหม่ที่จะถูกจำแนกกลุ่มต่าง ๆ เป็นแบบจำลองที่อยู่ในรูปแบบจำลอง Logit ดังสมการ

$$P(1) = \frac{e^{G_1}}{e^{G_1} + e^{G_2} + e^0}$$

$$P(2) = \frac{e^{G_2}}{e^{G_1} + e^{G_2} + e^0}$$

3.5.2.3 การทดสอบประสิทธิภาพในการสร้างแบบจำลอง

ในการทดสอบประสิทธิภาพในการสร้างแบบจำลองการถดถอยโลจิสติก จะทำการวัดประสิทธิภาพด้วยวิธีการหาความแม่นยำ (Accuracy : A) ดังแสดงในสมการ

$$A = \frac{TP + TN}{TP + TN + FP + FN}$$



โดยที่	A	=	ค่าความแม่นยำ
	TP	=	ค่าการพยากรณ์ถูกทางบวก
	TN	=	ค่าการพยากรณ์ถูกทางลบ
	FP	=	ค่าการพยากรณ์ผิดทางบวก
	FN	=	ค่าการพยากรณ์ผิดทางลบ

3.6 การเปรียบเทียบประสิทธิภาพแบบจำลอง

เนื่องจากในแต่ละแบบจำลองได้ทำการพัฒนาโดยวิธีการที่แตกต่างกัน จึงต้องมีการเปรียบเทียบประสิทธิภาพแบบจำลองว่าแบบจำลองใดให้ผลการพยากรณ์ที่ดีกว่า สำหรับการเปรียบเทียบประสิทธิภาพแบบจำลองการวิเคราะห์การดื่มน้ำสุราของประชาชนในจังหวัดหนองคาย โดยเทคนิคการทำเหมืองข้อมูลและการถดถอยโลจิสติกนั้น ใช้ร้อยละของความถูกต้องในการพยากรณ์เป็นตัวเปรียบเทียบ

$$\text{ร้อยละของการพยากรณ์ถูก} = \frac{\text{จำนวนข้อมูลทดสอบที่จำแนกได้ถูกต้อง}}{\text{จำนวนข้อมูลทดสอบทั้งหมด}} \times 100$$



บทที่ 3

วิธีดำเนินการวิจัย

การวิจัย เรื่องการวิเคราะห์การตีมูลค่าของประชาชนในจังหวัดหนองคาย โดยเทคนิคการทำเหมืองข้อมูลและการถดถอยโลจิสติก ได้กำหนดวิธีดำเนินการวิจัยตามขั้นตอนดังต่อไปนี้

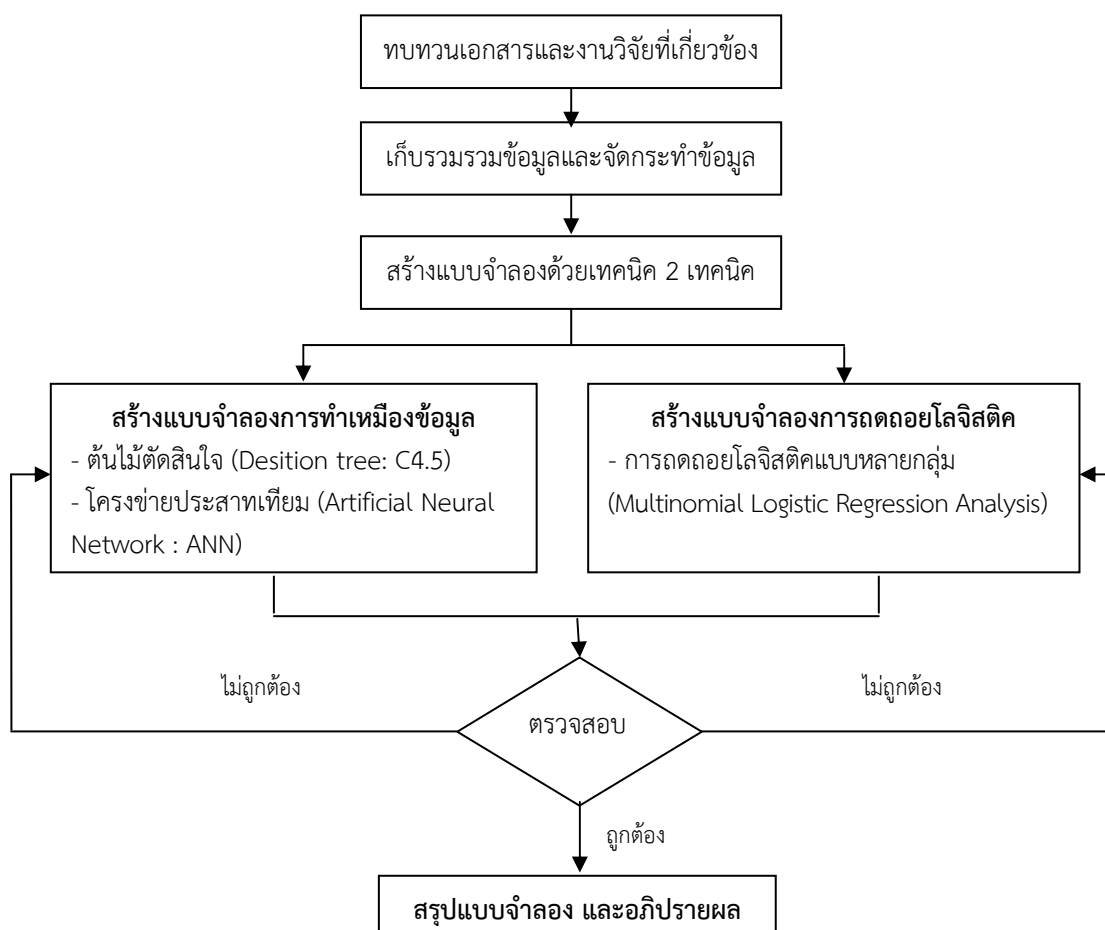
- 3.1 ขั้นตอนในการดำเนินการศึกษาวิจัย
- 3.2 กรอบแนวคิดในการวิจัย
- 3.3 เครื่องมือที่ใช้ในการวิจัย
- 3.4 การเก็บรวบรวมข้อมูล
- 3.5 การจัดกระทำและการวิเคราะห์ข้อมูล
- 3.6 การเปรียบเทียบประสิทธิภาพของแบบจำลอง

3.1 ขั้นตอนในการดำเนินการศึกษาวิจัย

งานวิจัยนี้มีขั้นตอนการดำเนินการดังแสดงดังภาพประกอบ 3.1 ดังนี้

1. ทบทวนเอกสารและงานวิจัยที่เกี่ยวข้องกับการวิเคราะห์เกี่ยวกับตีมูลค่าของประชาชน รวมถึงวิธีการต่าง ๆ ในการสำรวจข้อมูลและวิเคราะห์ข้อมูล
2. รวบรวมข้อมูลที่ใช้ในการวิจัย โดยการวิจัยครั้งนี้ใช้เป็นข้อมูลทุติยภูมิ (Secondary Data) ที่ได้จากการสำรวจพฤติกรรมการสูบบุหรี่และตีมูลค่าของประชากร พ.ศ. 2554 สำรวจโดยสำนักงานสถิติแห่งชาติ
3. จัดเตรียมข้อมูลให้เหมาะสมกับการสร้างแบบจำลอง โดยใช้โปรแกรมคอมพิวเตอร์ที่เหมาะสมในการจัดเตรียมข้อมูล
4. สร้างแบบจำลองการวิเคราะห์เกี่ยวกับตีมูลค่าของประชาชนในจังหวัดหนองคายด้วยเทคนิคการทำเหมืองข้อมูล และเทคนิคการถดถอยโลจิสติก
5. สรุปผลการวิจัยและจัดทำข้อเสนอแนะ



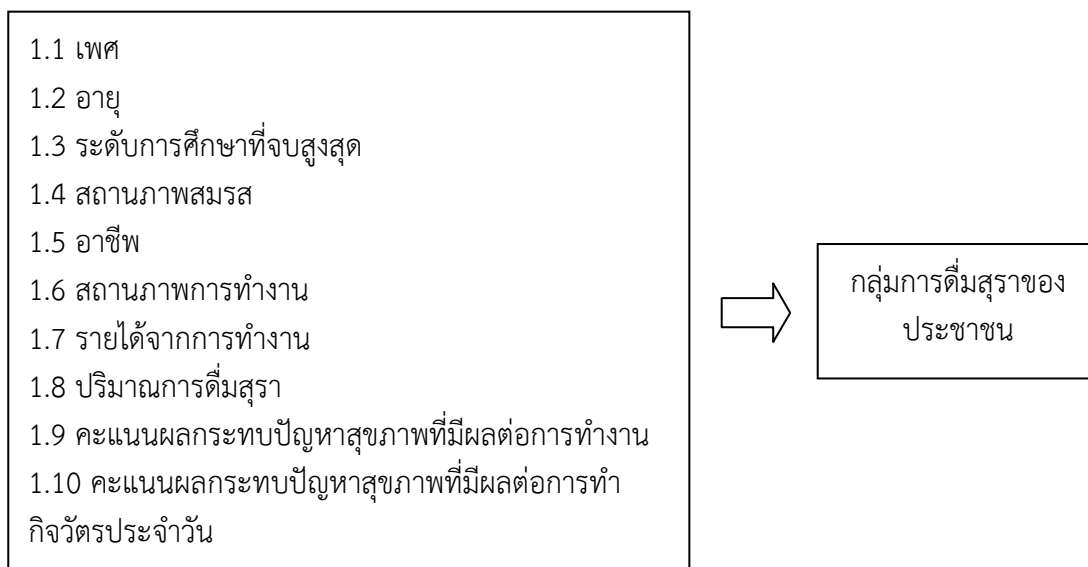


ภาพประกอบ 3.1 ขั้นตอนในการดำเนินการศึกษาวิจัย



3.2 กรอบแนวคิดในการวิจัย

จากการศึกษาทฤษฎี แนวคิด และงานวิจัยที่เกี่ยวข้อง ผู้วิจัยได้กำหนดกรอบแนวคิดในการวิจัย ดังนี้



ภาพประกอบ 3.2 กรอบแนวคิดในการวิจัย

ตัวแปรในการวิจัยมีจำนวนทั้งสิ้น 11 ตัวแปร จำแนกเป็นตัวแปรอิสระ 10 ตัว ตัวแปรตาม 1 ตัว ดังนี้

1. ตัวแปรอิสระ

เพศ

อายุ

ระดับการศึกษาที่จบสูงสุด

สถานภาพสมรส

อาชีพ

สถานภาพการทำงาน

รายได้จากการทำงาน

ปริมาณการดื่มสุรา

คะแนนผลกระทบปัญหาสุขภาพที่มีผลต่อการทำงาน

คะแนนผลกระทบปัญหาสุขภาพที่มีผลต่อการทำกิจกรรมประจำวัน

2. ตัวแปรตาม

การดื่มสุรา ใน 12 เดือนที่แล้ว แบ่งออกเป็น 3 กลุ่ม คือ

1) กลุ่มที่ไม่เคยดื่มสุรา

2) กลุ่มที่ดื่มเล็กน้อยหรือนาน ๆ ครั้ง

3) กลุ่มที่ดื่มสม่ำเสมอ



3.3 เครื่องมือที่ใช้ในการวิจัย

ในการวิจัยครั้งนี้ใช้โปรแกรมคอมพิวเตอร์สำเร็จรูปที่เหมาะสมกับการสร้างแบบจำลอง ด้วยวิธีการต่าง ๆ ซึ่งได้แก่ การสร้างแบบจำลองการวิเคราะห์การถดถอยของประชากรในจังหวัดหนองคาย ด้วยเทคนิคการทำเหมืองข้อมูล โดยวิธีต้นไม้ตัดสินใจ และวิธีโครงข่ายประสาทเทียม การเรียนรู้แบบแพร่กลับ กับเทคนิคการถดถอยโลจิสติก

3.4 การเก็บรวบรวมข้อมูล

การวิจัยครั้งนี้ใช้ข้อมูลทุติยภูมิ (Secondary Data) ที่ได้จากการสำรวจพฤติกรรม การสูบบุหรี่ และดื่มสุราของประชากร พ.ศ. 2554 สำรวจโดยสำนักงานสถิติแห่งชาติ โดยผู้วิจัยได้คัดเลือกข้อมูลที่ใช้ในการวิจัยทั้งสิ้นจำนวน 605 ราย จากการเก็บรวบรวมข้อมูลโดยเจ้าหน้าที่ของสำนักงานสถิติจังหวัดหนองคาย ไปสัมภาษณ์ข้อมูลครัวเรือนตัวอย่าง โดยสัมภาษณ์หัวหน้าครัวเรือนหรือสมาชิกในครัวเรือนในเรื่องของพฤติกรรมการสูบบุหรี่และดื่มสุรา ในระหว่างวันที่ 1 – 12 ของเดือน มีนาคม – พฤษภาคม พ.ศ. 2554

3.5 การจัดกระทำและการวิเคราะห์ข้อมูล

3.5.1 การทดสอบการสร้างแบบจำลองการทำเหมืองข้อมูล

3.5.1.1 การทดสอบการสร้างแบบจำลองต้นไม้ตัดสินใจ

การสร้างแบบจำลองต้นไม้ตัดสินใจ ด้วยอัลกอริทึม C4.5 (J48) มีขั้นตอนการดำเนินการดังต่อไปนี้

1. การเตรียมข้อมูลสำหรับการสร้างแบบจำลอง

จากชุดข้อมูลจำนวน 605 ราย ที่ได้ทำการคัดเลือกจากการสำรวจพฤติกรรม การสูบบุหรี่และดื่มสุราของประชากร พ.ศ. 2554 สำรวจโดยสำนักงานสถิติแห่งชาติ โดยทำการแปลงข้อมูลที่เลือกมาให้อยู่ในรูปแบบที่เหมาะสมสำหรับการนำไปใช้วิเคราะห์ตามอัลกอริทึม หลังจากได้ทำการ Cleaning ข้อมูลเป็นที่เรียบร้อยแล้วนั้น ผู้วิจัยได้ทำการแปลงรูปแบบของข้อมูลเพื่อให้เหมาะสมสำหรับการสร้างตัวแบบ ดังตารางที่ 3.1



ตาราง 3.1 ค่าของตัวแปรที่ทำการแปลงรูปแบบข้อมูล

ชื่อตัวแปร	ค่าข้อมูลเดิม	ค่าข้อมูลใหม่
เพศ (A4)	1 = ชาย 2 = หญิง	ใช้ค่าข้อมูลเดิม
อายุ (A5_RE)	Numeric	1 = 15 – 24 ปี 2 = 25 – 59 ปี 3 = 60 ปีขึ้นไป
ระดับการศึกษาที่จบสูงสุด (A6_RE)	1=ไม่เคยเรียน 2=ต่ำกว่าประถมศึกษา 3=ประถมศึกษา 4= มัธยมศึกษาตอนต้น 5=มัธยมศึกษาตอนปลาย สายสามัญ 6=มัธยมศึกษาตอนปลาย สายอาชีพ 7=ปวส./ปวท./ อนุปริญญา 8=ปริญญาตรี 9= สูงกว่าปริญญาตรี	-รหัส 1,2,3 แปลงค่าเป็น 1 =ประถมศึกษาหรือต่ำกว่า -รหัส 5 แปลงค่าเป็น 2=มัธยมศึกษาตอนปลายสายสามัญ -รหัส 6,7 แปลงค่าเป็น 3=มัธยมศึกษาตอนปลายสายสามัญ, ปวส./ปวท./อนุปริญญา -รหัส 8,9 แปลงค่าเป็น 4=ปริญญาตรีและสูงกว่า
สถานภาพสมรส (A7_RE)	1=โสด 2=สมรส 3=หม้าย 4=หย่า 5=แยกกันอยู่ 6=เคยสมรสแต่ไม่ทราบ สถานภาพ	-รหัส 1,3,4,5,6 แปลงค่าเป็น 1= โสด -รหัส 2 ใช้ค่าเดิม



ตาราง 3.1 (ต่อ)

ชื่อตัวแปร	ค่าข้อมูลเดิม	ค่าข้อมูลใหม่
อาชีพ (A8_RE)	1=เกษตรและประมง 2=ช่างเทคนิค ช่างฝีมือ เจ้าหน้าที่เทคนิค 3=ทำงานภาคบริการและ ผู้จำหน่ายสินค้า 4= กลุ่มผู้บริหาร ข้าราชการ และ ผู้ประกอบการวิชาชีพต่างๆ 5=ไม่ทำงาน	ใช้ค่าเดิม
สถานภาพการทำงาน (A10_RE)	1= นายจ้าง 2= ผู้ประกอบธุรกิจของ ตนเอง 3= ช่วยธุรกิจในครัวเรือน 4= ลูกจ้างรัฐบาล 5= ลูกจ้างรัฐวิสาหกิจ 6= ลูกจ้างเอกชน Missing = ไม่ทำงาน	-รหัส 1,2 แปลงค่าเป็น 1= นายจ้างและผู้ประกอบธุรกิจ ตนเอง -รหัส 3 แปลงค่าเป็น 2= ช่วยธุรกิจในครัวเรือน -รหัส 3,4 แปลงค่าเป็น 3=ลูกจ้างรัฐบาลและรัฐวิสาหกิจ -รหัส 6 แปลงค่าเป็น 4= ลูกจ้างเอกชน -ค่า Missing แปลงค่าเป็น 5=ไม่ทำงาน
รายได้จากการทำงาน (Inc)	Numeric	1 = < 2,000 บาทต่อเดือน 2 = 2,000-4,999 บาทต่อเดือน 3 = 5,000-9,999 บาทต่อเดือน 4 = 10,000-19,999 บาทต่อเดือน 5 = 20,000 บาทขึ้นไปต่อเดือน



ตาราง 3.1 (ต่อ)

ชื่อตัวแปร	ค่าข้อมูลเดิม	ค่าข้อมูลใหม่
ปริมาณการดื่มสุรา (B11_31_L)	Numeric	1 = 0.01–14.99 ลิตรต่อเดือน 2 = 15.00–29.99 ลิตรต่อเดือน 3 = 30 ลิตรขึ้นไปต่อเดือน 4 = ไม่เคยดื่ม
คะแนนผลกระทบปัญหา สุขภาพที่มีผลต่อการทำงาน (A42)	Numeric	ใช้ค่าข้อมูลเดิม
คะแนนผลกระทบปัญหา สุขภาพที่มีผลต่อการทำ กิจวัตรประจำวัน (A43)	Numeric	ใช้ค่าข้อมูลเดิม
การดื่มสุรา (B1_RE)	1= ไม่เคยดื่มเลยในชีวิตนี้ 2=เคยดื่ม แต่ไม่ดื่มใน 12 เดือนที่แล้ว 3=ทุกวัน (7 วัน/สัปดาห์) 4=เกือบทุกวัน (5-6 วัน/ สัปดาห์) 5=วันเว้นวัน (3-4 วัน/ สัปดาห์) 6=ทุกสัปดาห์ (1-2 วัน/ สัปดาห์) 7= นานๆ ครั้ง 8-11 วัน/ปี 8= นานๆ ครั้ง 4-7 วัน/ปี 9=นานๆ ครั้ง 1-3 วัน/ปี	-รหัส 3,4,5,6 แปลงค่าเป็น 1 = กลุ่มที่ดื่มสม่ำเสมอ -รหัส 7,8,9 แปลงค่าเป็น 2 = กลุ่มที่ดื่มเล็กน้อยหรือ นานๆ ครั้ง -รหัส 1,2 แปลงค่าเป็น 3 = กลุ่มที่ไม่เคยดื่มสุรา

จากนั้นนำข้อมูลมาเลือกอย่างสุ่มโดยใช้โปรแกรมสำเร็จรูปทางสถิติในการเลือก
จำนวน 5 ชุด โดยแบ่งข้อมูลออกเป็น 2 ส่วน คือ ส่วนแรกร้อยละ 80 ของจำนวนข้อมูลทั้งหมดใช้ในทุกชุด
ของการฝึกฝน และส่วนที่ 2 ร้อยละ 20 ให้ในชุดของการทดสอบ โดยผู้วิจัยได้นำชุดข้อมูลที่เตรียมไว้



จำนวน 5 ชุดนี้มาทดสอบการจำแนกข้อมูลด้วยโปรแกรม Weka 3.6.3 (Waikato Environment for Knowledge Analysis) พัฒนาโดย University of Waikato New Zealand เป็นโปรแกรมที่ใช้ในการทำ Data Mining และในโปรแกรมนี้นี้มีฟังก์ชันการสร้างแบบจำลองต้นไม้ตัดสินใจที่ใช้อัลกอริทึม C4.5 หรือที่รู้จักกันในชื่อ J48 ได้จำนวน 10 รูปแบบ เพื่อคัดเลือกรูปแบบเงื่อนไขการจำแนกกลุ่มการตัดสินใจที่ให้ค่าความถูกต้องสูงที่สุด โดยกำหนดให้กลุ่มการตัดสินใจเป็นแอตทริบิวต์เป้าหมาย ในการแบ่งข้อมูลสำหรับเรียนรู้ (Training Data) และข้อมูลสำหรับทดสอบ (Testing Data) ผู้วิจัยได้กำหนดเป็น 4 ชุด คือ

X1 คือ การกำหนดจำนวนข้อมูล Training Data และ Testing Data เป็น 50/50

X2 คือ การกำหนดจำนวนข้อมูล Training Data และ Testing Data เป็น 60/40

X3 คือ การกำหนดจำนวนข้อมูล Training Data และ Testing Data เป็น 80/20

X4 คือ การกำหนดจำนวนข้อมูล Training Data และ Testing Data เป็น 90/10

สำหรับขั้นตอนในการสร้างต้นไม้ตัดสินใจ มีการกำหนดค่าพารามิเตอร์ ดังนี้

1. Confidences Factor (CF) คือค่าความเชื่อมั่นที่ใช้สำหรับการเลือกวิธีการตัดกิ่งต้นไม้แบบ Pruning แบบ Error Base Pruning ซึ่งเทคนิคนี้เหมาะกับการสร้างต้นไม้ที่มีจำนวนข้อมูลไม่มาก โดยค่า CF คำนวณจาก Error Rate ได้จากการทดสอบกับ Test Set การเพิ่มค่า CF นี้จะทำให้ได้แบบจำลองต้นไม้ตัดสินใจที่มีลักษณะเฉพาะ แต่ถ้าลดค่า CF ลง จะทำให้แบบจำลองต้นไม้ตัดสินใจที่มีลักษณะทั่วไปมากขึ้น งานวิจัยนี้กำหนดค่า Confidences Factor (CF) = 0.25

2 Minimum Number Pruning (M) คือจำนวนข้อมูลที่น้อยที่สุดต่อ 1 Leaf ซึ่งต้องกำหนดเป็นจำนวนเต็มที่มีค่ามากกว่า 0 ในงานวิจัยนี้กำหนดให้มีค่าเท่ากับ 2

3. Fold คือ จำนวนข้อมูลที่ใช้ในการตัดกิ่ง ในงานวิจัยนี้กำหนดเท่ากับ 3

4. Binary Splits (Binary) คือ การสร้างต้นไม้แบบ Binary Tree เพื่อลดขนาดของต้นไม้ไม่ให้ซับซ้อนเกินไป ตัวเลือกสำหรับพารามิเตอร์นี้ คือ True, False ในงานวิจัยนี้กำหนดให้มีค่าทั้ง 2 ค่า

2. การทดสอบประสิทธิภาพในการสร้างแบบจำลองต้นไม้ตัดสินใจ จะทำการวัดประสิทธิภาพจาก

1. ขนาดของต้นไม้ จำนวน Node และ Leaf
2. ร้อยละความถูกต้องในการจำแนกกลุ่ม

$$\text{Correction} = \frac{\text{จำนวนข้อมูลฝึกฝนที่จำแนกได้}}{\text{จำนวนข้อมูลฝึกฝนทั้งหมด}} \times 100$$

3. ค่าความระลึก (Recall) คำนวณจากค่าของข้อมูลที่มีผลลัพธ์ถูกต้องโดยพิจารณาจากข้อมูลของผลลัพธ์เดียวกัน



$$\text{Recall} = \frac{\text{จำนวนข้อมูลที่จำแนกได้ถูกต้องของกลุ่มนั้น}}{\text{จำนวนข้อมูลที่อยู่ในกลุ่มนั้นจริง}}$$

4. ค่าความแม่นยำ (Precision) คำนวณจากค่าของข้อมูลที่มีผลลัพธ์ถูกต้องโดยพิจารณาจากจำนวนข้อมูลทั้งหมดที่ถูกจำแนกมีผลลัพธ์เดียวกัน

$$\text{Precision} = \frac{\text{จำนวนข้อมูลที่จำแนกได้ถูกต้องของกลุ่มนั้น}}{\text{จำนวนข้อมูลที่ถูกจำแนกว่าเป็นกลุ่มนั้นทั้งหมด}}$$

5. ค่าความเหวี่ยง (F-Measure) คำนวณจากค่าความแม่นยำ (P) และค่าความระลึก (R)

$$F(R, P) = \frac{2PR}{P + R}$$

3.5.1.2 ผลการสร้างแบบจำลองโครงข่ายประสาทเทียม (Artificial Neuron Network: ANN)

การสร้างแบบจำลองโครงข่ายประสาทเทียม (Artificial Neuron Network: ANN) มีขั้นตอนการดำเนินการดังต่อไปนี้

1. การศึกษาสัมประสิทธิ์สหสัมพันธ์และคัดเลือกตัวแปร

ในการศึกษาครั้งนี้ได้วิเคราะห์ความสัมพันธ์ระหว่างตัวแปรอิสระทั้ง 10 ตัวแปร คือ เพศ อายุ ระดับการศึกษาที่จบสูงสุด สถานภาพสมรส อาชีพ สถานภาพการทำงาน รายได้จากการทำงาน ปริมาณการดื่มสุรา คะแนนผลกระทบปัญหาสุขภาพที่มีผลต่อการทำงานและคะแนนผลกระทบปัญหาสุขภาพที่มีผลต่อการทำกิจวัตรประจำวัน กับตัวแปรตาม คือ กลุ่มการดื่มสุรา โดยใช้การหาความสัมพันธ์ระหว่างตัวแปรอิสระกับตัวแปรตาม 2 วิธี คือ

1) สัมประสิทธิ์ แบบ Cramer's V ทำการหาความสัมพันธ์ระหว่างตัวแปรอิสระ คือ เพศ ระดับการศึกษาที่จบสูงสุด สถานภาพสมรส อาชีพ สถานภาพการทำงาน กับ ตัวแปรตาม คือ กลุ่มการดื่มสุรา โดยใช้สูตร

$$V = \sqrt{\frac{x^2}{n(q - 1)}}$$

เมื่อ q = จำนวนแถวอน (r) หรือแถวตั้ง (c) ที่มีค่าน้อยที่สุด

n = จำนวนตัวอย่างทั้งหมด

โดยค่าของ V จะมีค่าระหว่าง 0 ถึง 1



2) สัมประสิทธิ์ แบบ Eta ทำการหาความสัมพันธ์ระหว่างตัวแปรอิสระ คือ รายได้จากการทำงาน ปริมาณการดื่มสุรา คะแนนผลกระทบปัญหาสุขภาพที่มีผลต่อการทำงาน คะแนนผลกระทบปัญหาสุขภาพที่มีผลต่อการทำกิจวัตรประจำวัน กับ ตัวแปรตาม คือ กลุ่มการดื่มสุรา โดยใช้สูตร

$$Eta = 1 - \frac{\sum_{i=1}^n (Y_i - \bar{Y}_k)^2}{\sum_{i=1}^n (Y_i - \bar{Y}_T)^2}$$

เมื่อ $(Y_i - \bar{Y}_k)^2$ = คะแนนเบี่ยงเบนจากค่าเฉลี่ยในแต่ละกลุ่ม
 $(Y_i - \bar{Y}_T)^2$ = คะแนนเบี่ยงเบนจากค่าเฉลี่ยรวม

2. กำหนดชุดข้อมูลนำเข้า (Input) ปัจจัยต่าง ๆ ที่มีอิทธิพลต่อการพยากรณ์ของโครงข่ายประสาทเทียม ในการพิจารณาการนำเข้าข้อมูลของวิธีโครงข่ายประสาทเทียมประกอบด้วย neuron input จำนวน 10 neuron โดยแต่ละ neuron ใช้สำหรับสัญญาณ input ต่างๆ สำหรับข้อมูล input จำนวน 605 ราย แยกเป็นใช้ในการเรียนรู้จะเลือกข้อมูลร้อยละ 90 หรือจำนวน 545 รายและข้อมูลที่ใช้ในการทดสอบโครงข่ายประสาทจะเลือกข้อมูลร้อยละ 10 หรือจำนวน 60 ราย และมีการกำหนดชุดข้อมูลของตัวแปรอิสระออกเป็น 3 ชุด คือ

ANN1 คือ กลุ่มตัวแปรอิสระที่ 1 ประกอบด้วยตัวแปรอิสระทั้งหมด 10 ตัว

ANN2 คือ กลุ่มตัวแปรอิสระที่ 2 ประกอบด้วยตัวแปรอิสระ มีค่าความสัมพันธ์ตั้งแต่ 0.50 ขึ้นไป

ANN3 คือ กลุ่มตัวแปรอิสระที่ 3 ประกอบด้วยตัวแปรอิสระ มีค่าความสัมพันธ์ต่ำกว่า 0.50

3. การออกแบบสถาปัตยกรรมโครงข่าย ในการกำหนดโครงสร้างของเครือข่ายประสาทเทียมไม่มีวิธีการที่แน่นอน โดยทั่วไปทำโดยวิธีการสุ่มค่าพารามิเตอร์ต่างๆ อันได้แก่ ค่าอัตราการเรียนรู้ ซึ่งเป็นค่าที่ควบคุมความคงตัว (Stability) ของค่าถ่วงน้ำหนักและความเร็วในการปรับค่าถ่วงน้ำหนักให้เข้าสู่ค่าที่สมดุล โดยอัตราการเรียนรู้มีค่าอยู่ระหว่าง 0 ถึง 1 โดยค่าถ่วงน้ำหนักที่เหมาะสมจะกำหนดให้อัตราการเรียนรู้มีค่าน้อยๆ อยู่ในช่วง 0.1 - 0.3 (อัญชลี วานิชทวีวัฒน์, 2540), ค่าโมเมนตัม เป็นค่าที่มีอิทธิพลต่อการเกิดค่า Error ของโครงข่าย โดยค่าถ่วงน้ำหนักไม่ควรเป็นค่าที่ใหญ่มากนัก โดยทั่วไปค่าถ่วงน้ำหนักเริ่มต้นจะสุ่มค่าระหว่าง -0.5 ถึง 0.5 (หรือ ระหว่าง -1 ถึง 1 ตามความเหมาะสม) (สุรยุทธ์ ปรีชา, 2541) และจำนวนครั้งที่ทำการฝึกฝนถ้ากำหนดให้จำนวนรอบในการฝึกฝนมีค่าสูงเกินไป ไม่ได้ทำให้โครงข่ายประสาทเทียมเรียนรู้ได้ดีมากขึ้น ควรประมาณการจำนวนรอบการสอนที่น้อยที่สุดที่จำเป็นสำหรับแก้ปัญหาหนึ่งๆ ที่ทำให้โครงข่ายเกิดการเรียนรู้ (ศุภัทรา สุนทรภักย์, 2539) จำนวนรอบการฝึกฝนอาจใช้ได้ตั้งแต่ 20,000 ถึง 100,000 รอบ ขึ้นอยู่กับลักษณะของโครงข่าย (ทัตดาว แบนเนียน, 2545)

งานวิจัยครั้งนี้ ผู้วิจัยได้ทำการสร้างและทดสอบตัวแบบการพยากรณ์ โดยใช้โปรแกรม Weka 3.6.6 ซึ่งมีการกำหนดค่าพารามิเตอร์ต่างๆ คือ จำนวนโหนดในชั้นซ่อน 9 ค่าอัตรา



การเรียนรู้ 0.2 ค่าโมเมนตัม 0.3 จำนวนรอบในการเรียนรู้ 50,000 รอบ โดยเป็นโครงข่ายประสาทเทียมแบบชั้นเดียว มีข้อมูลนำเข้า 545 ราย ส่งผ่านด้วยฟังก์ชันถ่ายโอน Log-Sigmoid และส่งถ่ายข้อมูลด้วยกระตุ้นแบบผลรวมเชิงเส้น Pure-Linear สู่ชั้นผลลัพธ์ คือกลุ่มการดื่มสุรา

4. ประสิทธิภาพของแบบจำลองโครงข่ายประสาทเทียม จะทำการวัดประสิทธิภาพจากการวัดความคลาดเคลื่อนระหว่างผลการทำนายและข้อมูลจริงจากชุดทดสอบ คือ ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Square Error: RMSE) และค่าร้อยละสัมบูรณ์ความคลาดเคลื่อนเฉลี่ย (Mean Absolute Percentage Error: MAPE)

1) ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Square Error: RMSE) คำนวณได้ดังสมการ

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (Predict_i - Actual_i)^2}$$

2) ค่าร้อยละสัมบูรณ์ความคลาดเคลื่อนเฉลี่ย (Mean Absolute Percentage Error: MAPE) คำนวณได้ดังสมการ

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{Predict_i - Actual_i}{Actual_i} \right| \times 100$$

โดยที่ Actual = ข้อมูลจริง

Predict = ข้อมูลที่ได้จากการพยากรณ์

N = จำนวนชุดข้อมูลที่ใช้ในการคำนวณ

3) ร้อยละความถูกต้องในการจำแนกกลุ่ม

$$Correction = \frac{\text{จำนวนข้อมูลฝึกฝนที่จำแนกได้}}{\text{จำนวนข้อมูลฝึกฝนทั้งหมด}} \times 100$$

3.5.2 การทดสอบแบบจำลองการถดถอยโลจิสติก

3.5.2.1 ตัวแปรที่เกี่ยวข้อง

ตัวแปรที่นำมาพิจารณาในการสร้างแบบจำลองการวิเคราะห์การดื่มสุราของประชาชนในจังหวัดหนองคาย โดยใช้การถดถอยโลจิสติกแบบหลายกลุ่ม (Multinomial Logistic Regression Analysis) โดยตัวแปรตาม คือ กลุ่มของการดื่มสุรา ส่วนตัวแปรอิสระใช้ตัวแปรทั้งหมด 10 ตัว คือ เพศ อายุ ระดับการศึกษาที่จบสูงสุด สถานภาพสมรส อาชีพ สถานภาพการทำงาน รายได้จากการทำงาน



ปริมาณการดื่มสุรา คะแนนผลกระทบปัญหาสุขภาพที่มีผลต่อการทำงานและคะแนนผลกระทบปัญหาสุขภาพที่มีผลต่อการทำกิจกรรมประจำวัน มาทำการวิเคราะห์เพื่อหาสมการในการพยากรณ์

3.5.2.2 โครงสร้างแบบจำลอง

สำหรับแบบจำลองการวิเคราะห์การดื่มสุราของประชาชนในจังหวัดหนองคายโดยใช้การถดถอยโลจิสติกแบบหลายกลุ่ม (Multinomial Logistic Regression Analysis) ในการสร้างแบบจำลอง โดยตัวแปรอิสระที่นำมาใช้เป็นองค์ประกอบของแบบจำลองซึ่งจะปรากฏอยู่ในฟังก์ชัน ดังนี้

$$G(1) = \beta_0 + \beta_{A4}X_{A4} + \beta_{A5_re}X_{A5_re} + \dots + \beta_{B43}X_{B43}$$

$$G(2) = \beta_0 + \beta_{A4}X_{A4} + \beta_{A5_re}X_{A5_re} + \dots + \beta_{B43}X_{B43}$$

โดยที่ $G(1)$	คือ สมการพยากรณ์กลุ่มการดื่มสม่ำเสมอเทียบกับกลุ่มไม่เคยดื่มสุรา
$G(2)$	คือ สมการพยากรณ์กลุ่มการดื่มเล็กน้อยหรือนานๆครั้งเทียบกับกลุ่มไม่เคยดื่มสุรา
$\beta_0, \beta_{A4}, \beta_{A5_RE}, \dots, \beta_{A43}$	คือ สัมประสิทธิ์ที่แสดงอิทธิพลของตัวแปรตัวที่ i ที่มีต่อกลุ่มการดื่มสุรา
$X_{A4}, X_{A5_RE}, \dots, X_{A43}$	คือ ตัวแปรตัวที่ i ที่มีอิทธิพลต่อกลุ่มการดื่มสุรา

แบบจำลองวิเคราะห์ความน่าจะเป็นที่พยากรณ์กลุ่มของรายใหม่ที่จะถูกจำแนกกลุ่มต่าง ๆ เป็นแบบจำลองที่อยู่ในรูปแบบจำลอง Logit ดังสมการ

$$P(1) = \frac{e^{G_1}}{e^{G_1} + e^{G_2} + e^0}$$

$$P(2) = \frac{e^{G_2}}{e^{G_1} + e^{G_2} + e^0}$$

3.5.2.3 การทดสอบประสิทธิภาพในการสร้างแบบจำลอง

ในการทดสอบประสิทธิภาพในการสร้างแบบจำลองการถดถอยโลจิสติก จะทำการวัดประสิทธิภาพด้วยวิธีการหาความแม่นยำ (Accuracy : A) ดังแสดงในสมการ

$$A = \frac{TP + TN}{TP + TN + FP + FN}$$



โดยที่	A	=	ค่าความแม่นยำ
	TP	=	ค่าการพยากรณ์ถูกทางบวก
	TN	=	ค่าการพยากรณ์ถูกทางลบ
	FP	=	ค่าการพยากรณ์ผิดทางบวก
	FN	=	ค่าการพยากรณ์ผิดทางลบ

3.6 การเปรียบเทียบประสิทธิภาพแบบจำลอง

เนื่องจากในแต่ละแบบจำลองได้ทำการพัฒนาโดยวิธีการที่ต่างกัน จึงต้องมีการเปรียบเทียบประสิทธิภาพแบบจำลองว่าแบบจำลองใดให้ผลการพยากรณ์ที่ดีกว่า สำหรับการเปรียบเทียบประสิทธิภาพแบบจำลองการวิเคราะห์การดื่มสุราของประชาชนในจังหวัดหนองคาย โดยเทคนิคการทำเหมืองข้อมูลและการถดถอยโลจิสติกนั้น ใช้ร้อยละของความถูกต้องในการพยากรณ์เป็นตัวเปรียบเทียบ

$$\text{ร้อยละของการพยากรณ์ถูก} = \frac{\text{จำนวนข้อมูลทดสอบที่จำแนกได้ถูกต้อง}}{\text{จำนวนข้อมูลทดสอบทั้งหมด}} \times 100$$



บทที่ 4

ผลการศึกษา

ผลการวิเคราะห์การตีสุราของประชาชนในจังหวัดหนองคาย โดยเทคนิคเหมืองข้อมูลและการถดถอยโลจิสติก ซึ่งมีลำดับขั้นตอนในการเสนอผลการวิเคราะห์ข้อมูล ดังนี้

- 4.1 สัญลักษณ์ในการนำเสนอ
- 4.2 ผลการทดสอบแบบจำลองการทำเหมืองข้อมูล
- 4.3 ผลการทดสอบแบบจำลองการถดถอยโลจิสติก
- 4.4 การเปรียบเทียบประสิทธิภาพแบบจำลอง

4.1 สัญลักษณ์ในการนำเสนอ

4.1.1 แบบจำลองการทำเหมืองข้อมูล

4.1.1.1 แบบจำลองต้นไม้ตัดสินใจ (Decision Tree)

ในการสร้างแบบจำลองต้นไม้ตัดสินใจที่ใช้พยากรณ์กลุ่มการตีสุราของประชาชนในจังหวัดหนองคาย มีสัญลักษณ์ในการนำเสนอและการกำหนดรายละเอียดในงานวิจัย ดังนี้

TR1 คือ การสร้างแบบจำลองโดยใช้เทคนิค EBP, ไม่กำหนดให้เป็นแบบ Binary, $CF=0.25$ จำนวน Leaf= 2, จำนวน fold ในการ pruning = 3 จากการสุ่มข้อมูลครั้งที่ 1

TR2 คือ การสร้างแบบจำลองโดยใช้เทคนิค EBP, ไม่กำหนดให้เป็นแบบ Binary, $CF=0.25$ จำนวน Leaf= 2, จำนวน fold ในการ pruning = 3 จากการสุ่มข้อมูลครั้งที่ 2

TR3 คือ การสร้างแบบจำลองโดยใช้เทคนิค EBP, ไม่กำหนดให้เป็นแบบ Binary, $CF=0.25$ จำนวน Leaf= 2, จำนวน fold ในการ pruning = 3 จากการสุ่มข้อมูลครั้งที่ 3

TR4 คือ การสร้างแบบจำลองโดยใช้เทคนิค EBP, ไม่กำหนดให้เป็นแบบ Binary, $CF=0.25$ จำนวน Leaf= 2, จำนวน fold ในการ pruning = 3 จากการสุ่มข้อมูลครั้งที่ 4

TR5 คือ การสร้างแบบจำลองโดยใช้เทคนิค EBP, ไม่กำหนดให้เป็นแบบ Binary, $CF=0.25$ จำนวน Leaf= 2, จำนวน fold ในการ pruning = 3 จากการสุ่มข้อมูลครั้งที่ 5

TR6 คือ การสร้างแบบจำลองโดยใช้เทคนิค EBP, กำหนดให้เป็นแบบ Binary, $CF=0.25$ จำนวน Leaf= 2, จำนวน fold ในการ pruning = 3 จากการสุ่มข้อมูลครั้งที่ 1

TR7 คือ การสร้างแบบจำลองโดยใช้เทคนิค EBP, กำหนดให้เป็นแบบ Binary, $CF=0.25$ จำนวน Leaf= 2, จำนวน fold ในการ pruning = 3 จากการสุ่มข้อมูลครั้งที่ 2

TR8 คือ การสร้างแบบจำลองโดยใช้เทคนิค EBP, กำหนดให้เป็นแบบ Binary, $CF=0.25$ จำนวน Leaf= 2, จำนวน fold ในการ pruning = 3 จากการสุ่มข้อมูลครั้งที่ 3

TR9 คือ การสร้างแบบจำลองโดยใช้เทคนิค EBP, กำหนดให้เป็นแบบ Binary, $CF=0.25$ จำนวน Leaf= 2, จำนวน fold ในการ pruning = 3 จากการสุ่มข้อมูลครั้งที่ 4



TR10 คือ การสร้างแบบจำลองโดยใช้เทคนิค EBP, กำหนดให้เป็นแบบ Binary, $CF=0.25$ จำนวน Leaf= 2, จำนวน fold ในการ pruning = 3 จากการสุ่มข้อมูลครั้งที่ 5

X1 คือ การกำหนดจำนวนข้อมูล Training Data และ Testing Data เป็น 50/50

X2 คือ การกำหนดจำนวนข้อมูล Training Data และ Testing Data เป็น 60/40

X3 คือ การกำหนดจำนวนข้อมูล Training Data และ Testing Data เป็น 80/20

X4 คือ การกำหนดจำนวนข้อมูล Training Data และ Testing Data เป็น 90/10

4.1.1.2 แบบจำลองโครงข่ายประสาทเทียม (Artificial Neuron Network: ANN)

ANN1 คือ กลุ่มตัวแปรอิสระที่ 1 ประกอบด้วยตัวแปรอิสระทั้งหมด 10 ตัว โดยไม่คำนึงถึงความสัมพันธ์

ANN2 คือ กลุ่มตัวแปรอิสระที่ 2 ประกอบด้วยตัวแปรอิสระ 3 ตัว มีค่าความสัมพันธ์ตั้งแต่ 0.50 ขึ้นไป

ANN3 คือ กลุ่มตัวแปรอิสระที่ 3 ประกอบด้วยตัวแปรอิสระ 7 ตัว มีค่าความสัมพันธ์ต่ำกว่า 0.50

4.1.2 แบบจำลองโลจิสติกรีเกรสชัน (Logistic Regression)

G(1) สมการพยากรณ์กลุ่มการดื่มสุร่าเสมอเทียบกับกลุ่มไม่เคยดื่มสุร่า

G(2) สมการพยากรณ์กลุ่มการดื่มเล็กน้อยหรือนานๆครั้งเทียบกับกลุ่มไม่เคยดื่มสุร่า

4.2 ผลการทดสอบแบบจำลองการทำเหมืองข้อมูล

4.2.1 ผลการทดสอบการสร้างแบบจำลองต้นไม้ตัดสินใจ

การสร้างแบบจำลองต้นไม้ตัดสินใจที่อาศัยโปรแกรม Weka 3.6.6 โดยใช้อัลกอริทึม C4.5(J48) เลือกวิธีการ Pruning แบบ Error Base Pruning มีกำหนดพารามิเตอร์ดังนี้

1. Confidences Factor (CF) = 0.25

2. Minimum Number Pruning (M) = 2

3. fold = 3

ผลการสร้างแบบจำลองต้นไม้ตัดสินใจที่ใช้ในการพยากรณ์การดื่มสุร่าของประชาชนในจังหวัดหนองคาย มีดังนี้



ตาราง 4.1 การเปรียบเทียบประสิทธิภาพแบบจำลองต้นไม้ตัดสินใจ

Model	Train/ Test	Leaves	Size	Precision	Recall	F-Measure	Correction
TR1	X1	13	18	89.30	89.70	89.10	89.71
	X2	13	18	88.50	89.00	88.10	88.99
	X3	13	18	88.50	87.20	86.80	87.16
	X4	13	18	88.90	88.90	88.90	88.89
TR2	X1	21	28	87.00	87.50	86.00	87.50
	X2	21	28	87.20	86.70	85.10	86.70
	X3	21	28	82.30	82.60	82.30	82.57
	X4	21	28	80.50	79.60	79.50	79.63
TR3	X1	23	31	87.70	87.10	84.90	87.13
	X2	23	31	87.40	87.60	85.70	87.62
	X3	23	31	81.40	84.40	81.70	84.40
	X4	23	31	87.30	87.00	85.40	87.04
TR4	X1	26	35	89.10	89.30	89.00	89.34
	X2	26	35	89.50	88.90	89.60	89.91
	X3	26	35	90.50	90.80	90.60	90.83
	X4	26	35	92.60	92.60	92.60	92.59
TR5	X1	11	14	87.90	87.10	85.90	87.13
	X2	11	14	89.80	87.20	85.90	87.16
	X3	11	14	89.70	89.00	88.10	88.99
	X4	11	14	82.50	83.30	82.50	83.33
TR6	X1	18	35	89.40	89.70	89.40	89.71
	X2	18	35	87.10	87.20	87.10	87.16
	X3	18	35	84.50	83.50	82.70	83.49
	X4	18	35	85.20	85.20	85.20	85.19
TR7	X1	13	25	88.30	88.60	88.50	88.60
	X2	13	25	90.40	90.40	90.10	90.37
	X3	13	25	86.90	87.20	86.90	87.16
	X4	13	25	80.50	79.60	79.50	79.63



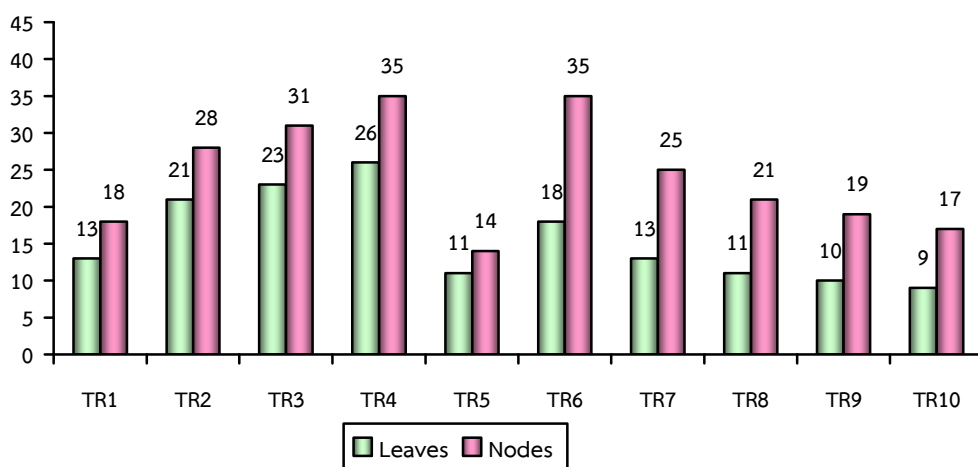
ตาราง 4.1 (ต่อ)

Model	Train/ Test	Leaves	Size	Precision	Recall	F-Measure	correction
TR8	X1	11	21	83.00	84.60	83.00	84.56
	X2	11	21	87.90	87.60	85.20	87.62
	X3	11	21	81.50	83.50	81.90	83.49
	X4	11	21	79.20	81.50	79.90	81.48
TR9	X1	10	19	88.70	89.00	88.60	88.97
	X2	10	19	87.50	88.10	87.60	88.07
	X3	10	19	88.50	89.00	88.50	88.99
	X4	10	19	92.60	92.60	92.60	92.59
TR10	X1	9	17	88.80	87.10	85.40	87.13
	X2	9	17	85.50	85.30	84.70	85.32
	X3	9	17	85.90	86.20	85.30	86.24
	X4	9	17	87.00	87.00	87.00	87.04

จากผลการสร้างแบบจำลองต้นไม้ตัดสินใจที่ใช้ในการพยากรณ์การดื่มสุราของประชาชน
ในจังหวัดหนองคาย จากตาราง พิจารณาได้ ดังนี้

1. ขนาดของต้นไม้ (Size) พิจารณาจากจำนวนโหนดและใบของต้นไม้ ผลดังภาพ

จำนวนโหนด

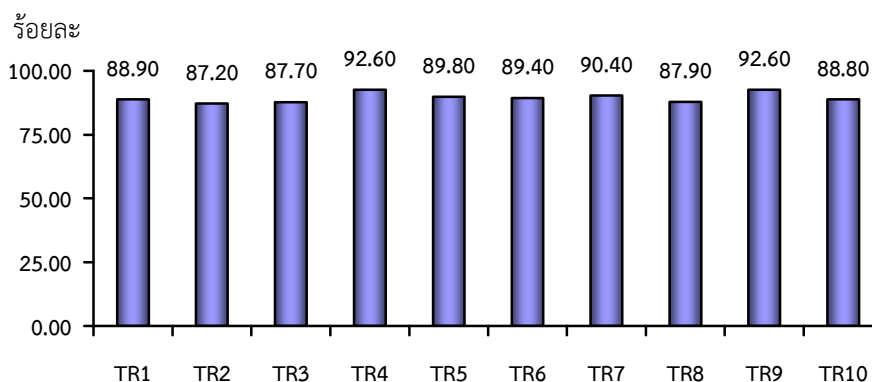


ภาพประกอบ 4.1 ขนาดของแบบจำลองต้นไม้ตัดสินใจ



จากภาพที่ 4.1 พบว่า ขนาดของแบบจำลองต้นไม้ TR4 และ TR6 มีขนาดใหญ่ที่สุด คือ 35 Nodes รองลงมา คือ TR3 จำนวน 31 Nodes TR2 จำนวน 28 Nodes TR7 จำนวน 25 Nodes TR2 จำนวน 21 Nodes TR9 จำนวน 19 Nodes TR1 จำนวน 18 Nodes TR10 จำนวน 17 Nodes และ TR5 มีขนาดเล็กที่สุด คือ 14 Nodes ตามลำดับ

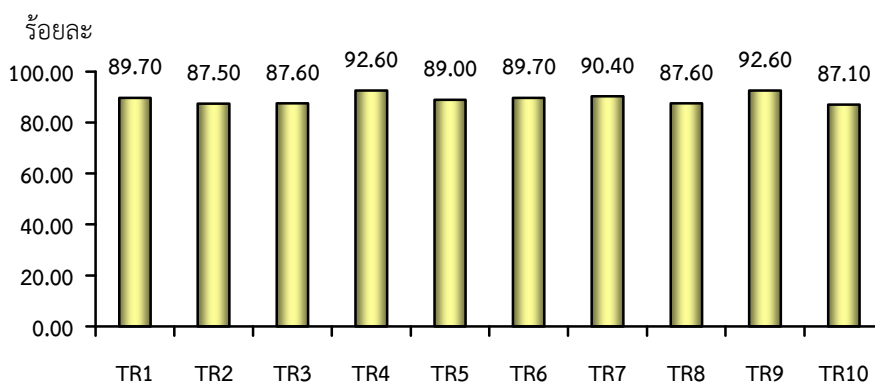
2. ความแม่นยำ (Precision) ค่าที่คำนวณจากค่าของข้อมูลที่มีผลลัพธ์ถูกต้องโดยพิจารณาจากจำนวนข้อมูลทั้งหมดที่ถูกจำแนกมีผลลัพธ์เดียวกัน ดังแผนภาพต่อไปนี้



ภาพประกอบ 4.2 ร้อยละของความแม่นยำ (Precision) ของแบบจำลองต้นไม้ตัดสินใจ

ความถูกต้องในการจำแนกกลุ่ม (Precision) จากภาพ พบว่า แบบจำลอง TR4 และแบบจำลอง TR9 มีความถูกต้องแม่นยำที่สุด คือ ร้อยละ 92.6 รองลงมา คือ แบบจำลอง TR5 ร้อยละ 89.8 แบบจำลอง TR6 ร้อยละ 89.4 แบบจำลอง TR1 ร้อยละ 88.9 ตามลำดับ

3. ค่าความระลึก (Recall) คือ ค่าที่คำนวณจากค่าของข้อมูลที่มีผลลัพธ์ถูกต้องโดยพิจารณาจากข้อมูลของผลลัพธ์เดียวกัน ดังภาพต่อไปนี้

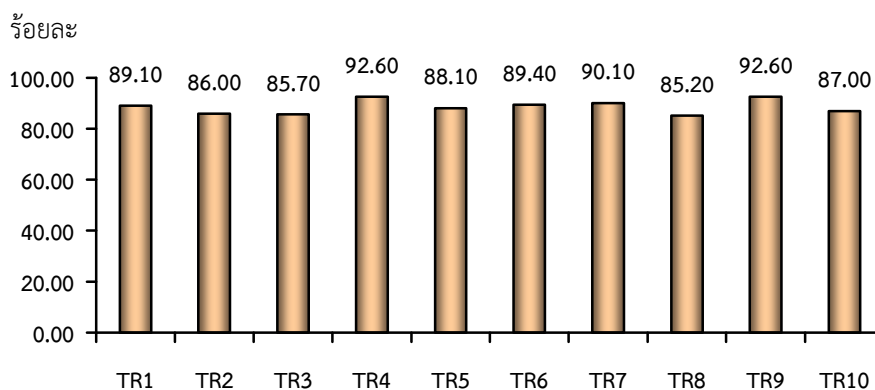


ภาพประกอบ 4.3 ร้อยละของความระลึก (Recall) ของแบบจำลองต้นไม้ตัดสินใจ



จากภาพที่ 4.3 พบว่า ความระลึก (Recall) ของแบบจำลองที่ TR4 และแบบจำลอง TR9 มีความถูกต้องแม่นยำที่สุด คือ ร้อยละ 92.6 รองลงมา คือ แบบจำลอง TR7 ร้อยละ 90.4 แบบจำลอง TR1 และแบบจำลอง TR6 ร้อยละ 89.7 แบบจำลอง TR5 ร้อยละ 89.0 ตามลำดับ

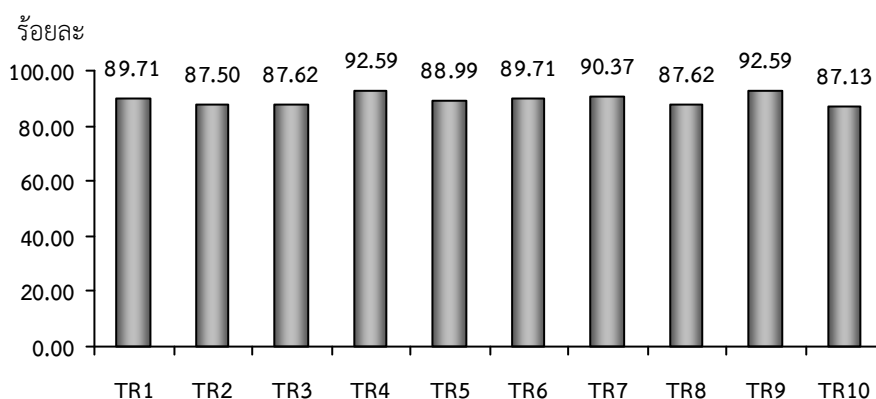
4. ค่าความเหวี่ยง (F-Measure) คือ ค่าที่คำนวณจากค่าความแม่นยำและค่าความระลึกได้ผลดังภาพต่อไปนี้



ภาพประกอบ 4.4 ร้อยละของค่าความเหวี่ยง (F-Measure) ของแบบจำลองต้นไม้ตัดสินใจ

ค่าความเหวี่ยง (F-Measure) จากภาพที่ 4.4 พบว่า แบบจำลองที่ TR4 และแบบจำลอง TR9 มีความถูกต้องแม่นยำที่สุด คือ ร้อยละ 92.6 รองลงมา คือ แบบจำลอง TR7 ร้อยละ 90.4 แบบจำลอง TR1 และแบบจำลอง TR6 ร้อยละ 89.7 แบบจำลอง TR5 ร้อยละ 89.0 ตามลำดับ

5. ค่าความถูกต้องในการจำแนกกลุ่ม (Correctly Classified Instances) ได้ผลดังภาพ



ภาพประกอบ 4.5 ร้อยละความถูกต้องในการจำแนกกลุ่ม (Correctly Classified Instances) ของแบบจำลองต้นไม้ตัดสินใจ



ความถูกต้องในการจำแนกกลุ่ม (Correctly Classified Instances) จากภาพที่ 4.5 พบว่าแบบจำลองที่ TR4 และแบบจำลอง TR9 มีความถูกต้องที่สุด คือ ร้อยละ 92.59 รองลงมา คือ แบบจำลอง TR7 ร้อยละ 90.37 แบบจำลอง TR1 และแบบจำลอง TR6 ร้อยละ 89.71 แบบจำลอง TR5 ร้อยละ 88.99 ตามลำดับ

6. ผลการพิจารณาเลือกแบบจำลองต้นไม้ตัดสินใจ

จากตาราง 4.1 และภาพที่ 4.1-4.5 พบว่า แบบจำลองต้นไม้ตัดสินใจแบบจำลองที่ TR4X4 และแบบจำลอง TR9X4 มีค่าความถูกต้องในการจำแนกกลุ่ม (Correctly Classified Instances) สูงที่สุด คือ ร้อยละ 92.59 ความแม่นยำ (Precision) ค่าความระลึก (Recall) ค่าความเหวี่ยง (F-Measure) เท่ากัน คือ ร้อยละ 92.60 ส่วนขนาดต้นไม้ไม่มีค่าแตกต่างกัน คือ แบบจำลองที่ TR4X4 มีขนาดต้นไม้ คือมีโหนดจำนวน 26 โหนด และใบจำนวน 35 ใบ ส่วนแบบจำลองที่ TR9X4 มีขนาดต้นไม้ คือมีโหนดจำนวน 10 โหนด และใบจำนวน 19 ใบ ดังตาราง

ตาราง 4.2 ประสิทธิภาพแบบจำลองต้นไม้ตัดสินใจตามแบบจำลอง TR4X4 และ TR9X4

Model	Train/ Test	Leaves	Size	Precision	Recall	F-Measure	correction
TR4	X4	26	35	92.60	92.60	92.60	92.59
TR9	X4	10	19	92.60	92.60	92.60	92.59

เมื่อพิจารณาแบบจำลองทั้ง 2 แบบ ผู้วิจัยพิจารณาเลือกแบบจำลอง TR9X4 เป็นแบบจำลองที่ดีที่สุด เนื่องจากมีจำนวน ขนาดของต้นไม้ที่เล็กกว่า และได้นำข้อมูล ชุด Test จำนวน 60 recode เข้าทดสอบ ได้ผลการทดสอบดังตาราง

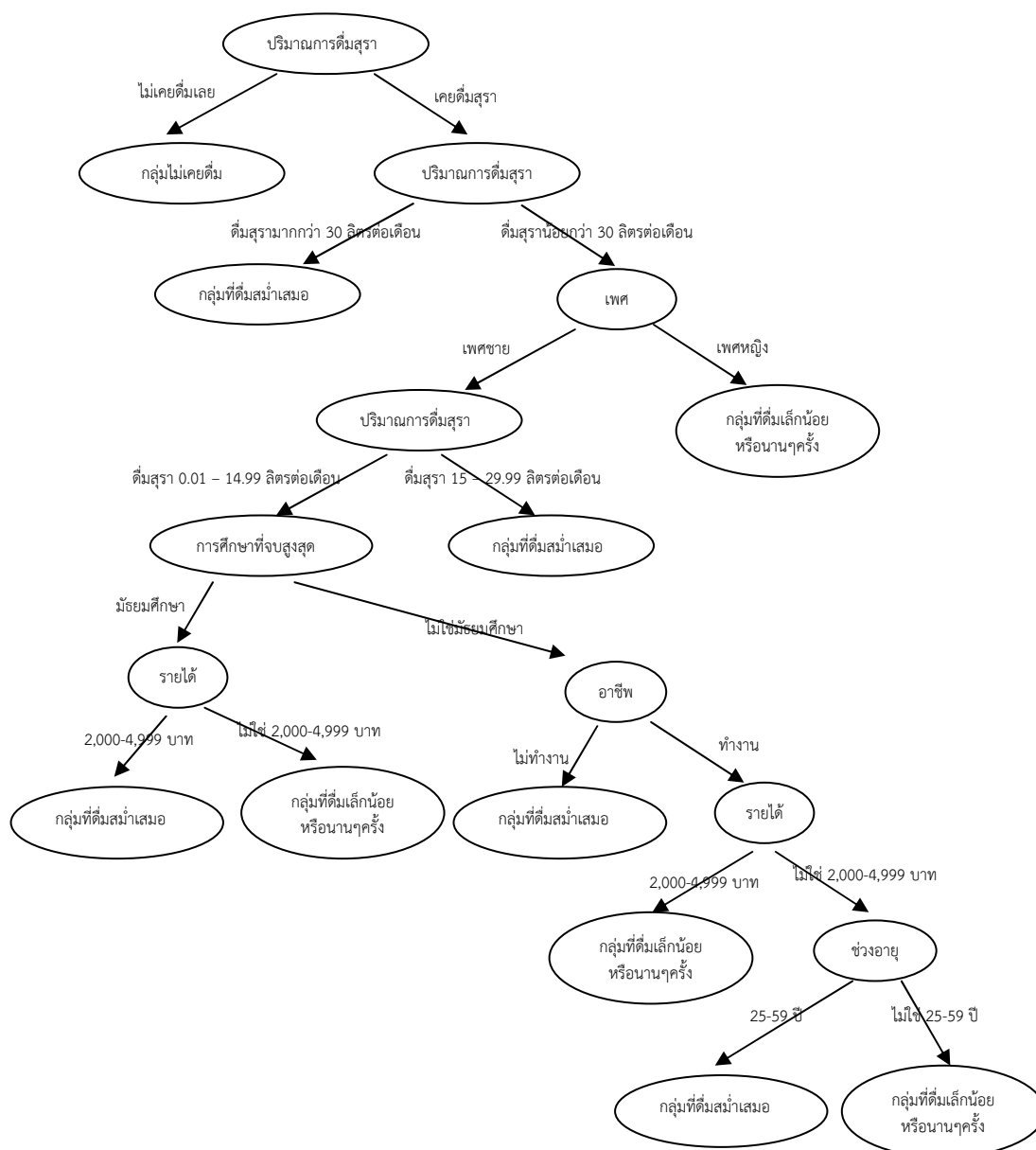
ตาราง 4.3 ประสิทธิภาพแบบจำลองต้นไม้ตัดสินใจตามแบบจำลอง TR9X4

Model	Train/ Test	Leaves	Size	Precision	Recall	F-Measure	correction
TR9	X4	10	19	100.0	100.0	100.0	100.0

จากแบบจำลอง แบบจำลอง TR9X4 นำมาสร้างแผนภาพของต้นไม้ตัดสินใจ

ดังภาพ





ภาพประกอบ 4.6 ต้นไม้จำลองตัดสินใจ ตาม Model TR9X4

จากภาพที่ 4.6 แบบจำลองต้นไม้ตัดสินใจ ตาม Model TR9X4 พบว่า ตัวแปรที่ถูกนำมาพิจารณาในการสร้างต้นไม้ตัดสินใจ มีทั้ง 6 ตัวแปร ได้แก่ เพศ อายุ ระดับการศึกษาที่จบสูงสุด อาชีพ รายได้จากการทำงาน ปริมาณการดื่มสุราต่อเดือน โดยสามารถสร้างกฎได้ ดังนี้

กฎการจำแนกกลุ่มการดื่มสุราของประชาชนในจังหวัดหนองคาย มีดังนี้

1. กลุ่มไม่เคยดื่มสุรา

1.1 ผู้ที่ไม่เคยดื่มสุราเลยหรือไม่เคยดื่มใน 12 เดือน



2. กลุ่มที่ดื่มเล็กน้อยหรือนานๆครั้ง

- 2.1. เป็นเพศหญิง ที่มีการดื่มสุราต่อเดือนน้อยกว่า 30.00 ลิตรต่อเดือน”
- 2.2. เป็นเพศชายที่จบการศึกษาสูงสุดระดับมัธยมศึกษาสายสามัญ มีรายได้ที่ไม่ใช่ 2,000 – 4,999 บาท และดื่มสุราต่อเดือนน้อยกว่า 15.00 ลิตรต่อเดือน
- 2.3. เป็นเพศชายที่จบการศึกษาสูงสุดนอกเหนือจากมัธยมศึกษาสายสามัญ ทำงานโดยมีรายได้ 2,000 – 4,999 บาท และดื่มสุราต่อเดือนน้อยกว่า 15.00 ลิตรต่อเดือน
- 2.4. เป็นเพศชาย ที่มีอายุนอกเหนือจาก 25-29 ปี จบการศึกษาสูงสุดนอกเหนือจากมัธยมศึกษาสายสามัญ ทำงานโดยมีรายได้ที่ไม่ใช่ 2,000 – 4,999 บาท และดื่มสุราต่อเดือนน้อยกว่า 15.00 ลิตรต่อเดือน

3. กลุ่มที่ดื่มสม่ำเสมอ

- 3.1. เป็นผู้ที่ดื่มสุราต่อเดือนมากกว่า 30.00 ลิตรต่อเดือน
- 3.2. เป็นเพศชายที่ดื่มสุราต่อเดือนระหว่าง 15.00 – 29.99 ลิตรต่อเดือน
- 3.3. เป็นผู้ชายที่จบการศึกษาสูงสุดระดับมัธยมศึกษาสายสามัญ มีรายได้ 2,000 – 4,999 บาท และดื่มสุราต่อเดือนน้อยกว่า 15.00 ลิตรต่อเดือน
- 3.4. เป็นผู้ชายที่จบการศึกษาสูงสุดนอกเหนือจากมัธยมศึกษาสายสามัญ ไม่ทำงานและดื่มสุราต่อเดือนน้อยกว่า 15.00 ลิตรต่อเดือน
- 3.5. เป็นผู้ชายที่มีอายุ 25-29 ปี จบการศึกษาสูงสุดนอกเหนือจากมัธยมศึกษาสายสามัญ มีรายได้ที่ไม่ใช่ 2,000 – 4,999 บาท และดื่มสุราต่อเดือนน้อยกว่า 15.00 ลิตรต่อเดือน

4.2.2 ผลการสร้างแบบจำลองโครงข่ายประสาทเทียม (Artificial Neuron Network: ANN)

การวิเคราะห์การดื่มสุราของประชาชนในจังหวัดหนองคาย เมื่อทำการหาแบบจำลองที่ดีที่สุดจากแบบจำลองโครงข่ายประสาทเทียม มีดังนี้

1. การวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรอิสระทั้ง 10 ตัวแปร คือ เพศ อายุ ระดับการศึกษาที่จบสูงสุด สถานภาพสมรส อาชีพ สถานภาพการทำงาน รายได้จากการทำงาน ปริมาณการดื่มสุรา คะแนนผลกระทบปัญหาสุขภาพที่มีผลต่อการทำงานและคะแนนผลกระทบปัญหาสุขภาพที่มีผลต่อการทำกิจวัตรประจำวัน กับตัวแปรตาม คือ กลุ่มการดื่มสุรา โดยใช้โปรแกรมสำเร็จรูปทางสถิติ ซึ่งได้ค่าความสัมพันธ์ดังตาราง 4.4



ตาราง 4.4 สัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรอิสระและตัวแปรตาม

ตัวแปรอิสระ	ความสัมพันธ์ กับกลุ่มการดื่มสุรา (Correlation)
เพศ*	0.55
อายุ**	0.16
ระดับการศึกษาที่จบสูงสุด*	0.16
สถานภาพสมรส*	0.07
อาชีพ*	0.20
สถานภาพการทำงาน*	0.20
รายได้จากการทำงาน**	0.58
ปริมาณการดื่มสุรา**	0.96
คะแนนผลกระทบปัญหาสุขภาพที่มีผลต่อการทำงาน**	0.20
คะแนนผลกระทบปัญหาสุขภาพที่มีผลต่อการทำกิจวัตรประจำวัน**	0.16

* ทดสอบความสัมพันธ์โดย Cramer's V

** ทดสอบความสัมพันธ์โดย Eta

ในการพิจารณาการนำเข้าข้อมูลของวิธีโครงข่ายประสาทเทียมประกอบด้วยนิรอนอินพุตจำนวน 10 นิรอน โดยแต่ละนิรอนใช้สำหรับสัญญาณอินพุตต่างๆ สำหรับข้อมูลอินพุตที่ใช้ในการเรียนรู้จะเลือกข้อมูล 545 รายและข้อมูลที่ใช้ในการทดสอบโครงข่ายประสาทจะเลือกข้อมูล 60 ราย

ANN1 คือ กลุ่มตัวแปรอิสระที่ 1 ประกอบด้วยตัวแปรอิสระทั้งหมด 10 ตัว คือ เพศ อายุ ระดับการศึกษาที่จบสูงสุด สถานภาพสมรส อาชีพ สถานภาพการทำงาน รายได้จากการทำงาน ปริมาณการดื่มสุรา คะแนนผลกระทบปัญหาสุขภาพที่มีผลต่อการทำงานและคะแนนผลกระทบปัญหาสุขภาพที่มีผลต่อการทำกิจวัตรประจำวัน โดยไม่คำนึงถึงความสัมพันธ์

ANN2 คือ กลุ่มตัวแปรอิสระที่ 2 ประกอบด้วยตัวแปรอิสระ 3 ตัว คือ เพศ รายได้จากการทำงานและปริมาณการดื่มสุรา มีค่าความสัมพันธ์ตั้งแต่ 0.50 ขึ้นไป

ANN3 คือ กลุ่มตัวแปรอิสระที่ 3 ประกอบด้วยตัวแปรอิสระ 7 ตัว คือ อายุ ระดับการศึกษาที่จบสูงสุด สถานภาพสมรส อาชีพ สถานภาพการทำงาน คะแนนผลกระทบปัญหาสุขภาพที่มีผลต่อการทำงานและคะแนนผลกระทบปัญหาสุขภาพที่มีผลต่อการทำกิจวัตรประจำวัน มีค่าความสัมพันธ์ต่ำกว่า 0.50

2. ผลการทดสอบตัวแบบโครงข่ายประสาทเทียมแบบ Multi-Layer Perceptron (MLP) จากการศึกษาการวิจัยครั้งนี้ผู้วิจัยได้ทำการทดสอบตัวแบบการพยากรณ์ โดยใช้โปรแกรม Weka 3.6.6 ซึ่งมีการกำหนดค่าพารามิเตอร์ต่างๆ ดังนี้



จำนวนโหนดในชั้นซ่อน 9

ค่าอัตราการเรียนรู้ 0.2

ค่าโมเมนตัม 0.3

จำนวนรอบในการเรียนรู้ 50,000 รอบ

ตัวแบบที่ 1 (ANN1) ใช้ตัวแบบ คือ 10:9:1 ชั้นนำเข้า 10 โหนด ชั้นซ่อน 9 โหนด และชั้นแสดงผล 1 โหนด ในชั้นซ่อนใช้ฟังก์ชันการแปลงค่าคือ ฟังก์ชันซิกมอยด์ (Sigmoid Function = f) ส่วนชั้นแสดงผลใช้ฟังก์ชันกระตุ้นแบบผลรวมเชิงเส้น

ตัวแบบที่ 2 (ANN2) ใช้ตัวแบบ คือ 3:9:1 ชั้นนำเข้า 3 โหนด ชั้นซ่อน 9 โหนด และชั้นแสดงผล 1 โหนด ในชั้นซ่อนใช้ฟังก์ชันการแปลงค่าคือ ฟังก์ชันซิกมอยด์ (Sigmoid Function = f) ส่วนชั้นแสดงผลใช้ ฟังก์ชันกระตุ้นแบบผลรวมเชิงเส้น

ตัวแบบที่ 3 (ANN3) ใช้ตัวแบบ คือ 7:9:1 ชั้นนำเข้า 7 โหนด ชั้นซ่อน 9 โหนด และชั้นแสดงผล 1 โหนด ในชั้นซ่อนใช้ฟังก์ชันการแปลงค่าคือ ฟังก์ชันซิกมอยด์ (Sigmoid Function = f) ส่วนชั้นแสดงผลใช้ ฟังก์ชันกระตุ้นแบบผลรวมเชิงเส้น

จากนั้นดำเนินการหาตัวแบบโดยนำค่าเหล่านี้ไปใช้ในการพยากรณ์หาตัวแบบด้วยวิธี MLP และแสดงผลการทดสอบตัวแบบในตาราง 4.5 สามารถสร้างตัวแบบที่เหมาะสมได้ 3 ตัวแบบ ดังนี้

ตาราง 4.5 การวัดประสิทธิภาพการฝึกสอนและการทดสอบแบบจำลองโดยใช้โครงข่ายประสาทเทียม

แบบจำลอง	การฝึกฝน		การทดสอบ	
	MAPE	RMSE	MAPE	RMSE
ANN1 : 10-9-1	0.0199	0.1012	0.0007	0.0011
ANN2 : 3-9-1	0.3830	0.4285	0.0611	0.1415
ANN3 : 7-9-1	0.6725	0.7338	0.7365	0.8207

จากตาราง 4.5 พบว่า ตัวแบบ ANN-1 มีประสิทธิภาพในการพยากรณ์สูงสุด โดยพิจารณาค่าความคลาดเคลื่อนของการทดสอบในชุดการฝึกฝน MAPE เท่ากับ 0.0199 RMSE เท่ากับ 0.1012 ซึ่งประกอบด้วยตัวแปรอิสระ จำนวน 10 คือ ประกอบด้วยตัวแปรอิสระทั้งหมด 10 ตัว คือ เพศ อายุ ระดับการศึกษาที่จบสูงสุด สถานภาพสมรส อาชีพ สถานภาพการทำงาน รายได้จากการทำงาน ปริมาณการดื่มสุรา คะแนนผลกระทบปัญหาสุขภาพที่มีผลต่อการทำงานและคะแนนผลกระทบปัญหาสุขภาพที่มีผลต่อการทำกิจวัตรประจำวัน เมื่อลดจำนวนตัวแปรอิสระลงโดยแบ่งตามความสัมพันธ์ในแบบจำลองที่ ANN2 และ ANN3 พบว่า ค่าความคลาดเคลื่อน MAPE และ RMSE มีค่าสูงกว่าตัวแบบ ANN-1



3. ประสิทธิภาพของแบบจำลองโครงข่ายประสาทเทียม

แบบจำลองที่มีประสิทธิภาพสูงสุดคือ แบบจำลอง ANN1 (10-9-1) ที่มีค่าความคลาดเคลื่อนน้อยที่สุด คือ ในชุดการฝึกฝน MAPE เท่ากับ 0.0199 RMSE เท่ากับ 0.1012 และชุดการทดสอบ คือ ในชุดการฝึกฝน MAPE เท่ากับ 0.0007 RMSE เท่ากับ 0.0011 และเมื่อนำไปหาค่าความถูกต้องของการพยากรณ์ พบว่า มีความถูกต้องของการพยากรณ์ ร้อยละ 97.79 ได้ผลดังตาราง

ตาราง 4.6 ประสิทธิภาพแบบจำลองโครงข่ายประสาทเทียม ตามแบบจำลอง ANN1

Model	Precision	Recall	F-Measure	correction
ANN1 : 10-9-1	97.80	97.80	97.80	97.79

4.3 ผลการทดสอบแบบจำลองการถดถอยโลจิสติก

การวิเคราะห์การติ่มสุราของประชาชนในจังหวัดหนองคาย เมื่อทำการหาแบบจำลองที่ดีที่สุดจากแบบจำลองการถดถอยโลจิสติก โดยใช้ตัวแปรอิสระทั้งหมด 10 ตัว คือ เพศ อายุ ระดับการศึกษาที่จบสูงสุด สถานภาพสมรส อาชีพ สถานภาพการทำงาน รายได้จากการทำงาน ปริมาณการติ่มสุรา คะแนนผลกระทบปัญหาสุขภาพที่มีผลต่อการทำงานและคะแนนผลกระทบปัญหาสุขภาพที่มีผลต่อการทำกิจวัตรประจำวัน มาทำการวิเคราะห์เพื่อหาสมการในการพยากรณ์ ผลมีดังนี้

1. ความเหมาะสมของแบบจำลอง เมื่อนำข้อมูลมาวิเคราะห์โดยใช้การถดถอยโลจิสติกแบบหลายกลุ่ม ได้ผลดังตาราง

ตาราง 4.7 วิเคราะห์ความเหมาะสมของแบบจำลองการถดถอยโลจิสติกของตัวแปรที่ส่งผลกระทบต่อกลุ่มการติ่มสุราของคนในจังหวัดหนองคาย

Model	ค่าสถิติ
-2 Log likelihood	620.033
χ^2	290.709
df	36
test P value	.000
n	605
R^2	.382

จากตาราง 4.7 พบว่า จากการทดสอบโดยใช้สถิติทดสอบ Chi – square ซึ่งได้ค่าของทดสอบ เท่ากับ .000 จึงสรุปได้ว่า ตัวแปรอิสระทั้งหมด 10 ตัวแปร คือ เพศ อายุ ระดับการศึกษาที่จบ



สูงสุด สถานภาพสมรส อาชีพ สถานภาพการทำงาน รายได้จากการทำงาน ปริมาณการดื่มสุรา คะแนนผลกระทบปัญหาสุขภาพที่มีผลต่อการทำงานและคะแนนผลกระทบปัญหาสุขภาพที่มีผลต่อการทำกิจกรรมประจำวัน เป็นตัวแปรที่ส่งผลกระทบต่อพยากรณ์กลุ่มการดื่มสุราของประชาชนในจังหวัดหนองคาย

2. สมการของแบบจำลองการถดถอยโลจิสติก

เมื่อทดสอบพบว่าแบบจำลองมีความเหมาะสมแล้ว จึงนำมาสร้างสมการในการพยากรณ์ ได้ผลดังตาราง

ตาราง 4.8 ค่า Odds Ratio จากการวิเคราะห์การถดถอยโลจิสติกของตัวแปรที่ส่งผลต่อกลุ่มการดื่มสุราของประชาชนในจังหวัดหนองคาย

ตัวแปร	Odds Ratio
กลุ่มที่ดื่มสม่ำเสมอ	
Constant	-5.143*
รายได้จากการทำงาน	.000
ปริมาณการดื่มแอลกอฮอล์	.061*
คะแนนผลกระทบปัญหาสุขภาพที่มีผลต่อการทำงาน	.857
คะแนนผลกระทบปัญหาสุขภาพที่มีผลต่อการทำกิจกรรมประจำวัน	1.604
เพศ (อ้างอิง:หญิง)	
ชาย	2.564*
อายุ (อ้างอิง:60 ปีขึ้นไป)	
15 - 24 ปี	-.555
25 - 59 ปี	1.125*
ระดับการศึกษาที่จบสูงสุด (อ้างอิง:ปริญญาตรีหรือสูงกว่า)	
ประถมศึกษาและต่ำกว่า	.724
มัธยมศึกษาสายสามัญ	1.000
มัธยมศึกษาสายอาชีพ(ปวช. ปวส. อนุปริญญา)	1.946
สถานภาพสมรส (อ้างอิง:สมรส)	
โสด หม้าย หย่า แยกกันอยู่ เคยสมรส	.179
อาชีพ (อ้างอิง:ไม่ทำงาน)	
เกษตรกรและประมง	.398
ช่างเทคนิค ช่างฝีมือ เจ้าหน้าที่เทคนิค	.000
ทำงานภาคบริการและผู้จำหน่ายสินค้า	-.204
ผู้บริหารข้าราชการและผู้ประกอบวิชาชีพต่างๆ	.687



ตาราง 4.8 (ต่อ)

ตัวแปร	Odds Ratio
สถานภาพการทำงาน (อ้างอิง:ไม่ได้ทำงานเชิงธุรกิจ)	
นายจ้างและผู้ประกอบธุรกิจตนเอง	-.485
ผู้ช่วยธุรกิจในครัวเรือน	-.228
ลูกจ้างรัฐบาลและรัฐวิสาหกิจ	.598
ลูกจ้างเอกชน	-
กลุ่มที่ดื่มเล็กน้อยหรือนานๆ ครั้ง	
Constant	-4.348*
รายได้จากการทำงาน	.000
ปริมาณการดื่มแอลกอฮอล์	.063*
คะแนนผลกระทบปัญหาสุขภาพที่มีผลต่อการทำงาน	.777
คะแนนผลกระทบปัญหาสุขภาพที่มีผลต่อการทำกิจวัตรประจำวัน	3.908
เพศ(อ้างอิง:หญิง)	
ชาย	1.813*
อายุ (อ้างอิง:60 ปีขึ้นไป)	
15 - 24 ปี	-.224
25 – 59 ปี	1.043*
ผู้ช่วยธุรกิจในครัวเรือน	-.644
ลูกจ้างรัฐบาลและรัฐวิสาหกิจ	-.386
ลูกจ้างเอกชน	-

กลุ่มอ้างอิง: กลุ่มที่ไม่เคยดื่มสุรา * $p < 0.05$

สมการ มีดังนี้

$$G(1) = -5.143 + 0.061(B11_31_L) + 2.564(A4) + 1.125(A5_re=2)$$

$$G(2) = -4.348 + 0.063(B11_31_L) + 1.813(A4) + 1.043(A5_re=2)$$

โดยที่ G(1) คือ สมการพยากรณ์กลุ่มการดื่มสม่ำเสมอเทียบกับกลุ่มไม่เคยดื่มสุรา
 G(2) คือ สมการพยากรณ์กลุ่มการดื่มเล็กน้อยหรือนานๆ ครั้งเทียบกับกลุ่มไม่เคยดื่มสุรา
 B11_31_L คือ ปริมาณการดื่มสุรา
 A4 คือ เพศ
 A5_re=2 คือ อายุ 25 – 59 ปี



3. ประสิทธิภาพของแบบจำลองการถดถอยโลจิสติก

ประสิทธิภาพของแบบจำลองการถดถอยโลจิสติกแบบหลายกลุ่ม โดยใช้ตัวแปรอิสระทั้งหมด 10 ตัว คือ เพศ อายุ ระดับการศึกษาที่จบสูงสุด สถานภาพสมรส อาชีพ สถานภาพการทำงาน รายได้จากการทำงาน ปริมาณการดื่มสุรา คะแนนผลกระทบปัญหาสุขภาพที่มีผลต่อการทำงานและคะแนนผลกระทบปัญหาสุขภาพที่มีผลต่อการทำกิจวัตรประจำวัน เป็นตัวแปรที่ส่งผลการพยากรณ์กลุ่ม การดื่มสุราของประชาชนในจังหวัดหนองคาย พบว่า มีความถูกต้องของการพยากรณ์ ร้อยละ 80.80 ได้ผลดังตาราง

ตาราง 4.9 ประสิทธิภาพของแบบจำลองการถดถอยโลจิสติก

ค่าสังเกต (Observed)	ค่าพยากรณ์ (Predicted)			
	กลุ่มที่ดื่มสม่ำเสมอ	กลุ่มที่ดื่มเล็กน้อยหรือนาน ๆ ครั้ง	กลุ่มที่ไม่เคยดื่มสุรา	ร้อยละความถูกต้อง
กลุ่มที่ดื่มสม่ำเสมอ	29	25	32	33.70%
กลุ่มที่ดื่มเล็กน้อยหรือนาน ๆ ครั้ง	9	15	49	20.50%
กลุ่มที่ไม่เคยดื่มสุรา	1	-	445	99.80%
ความถูกต้องโดยรวม	6.40%	6.60%	86.90%	80.80%

4.4 การเปรียบเทียบประสิทธิภาพแบบจำลอง

การศึกษาครั้งนี้ผู้วิจัยทำการเปรียบเทียบผลการพยากรณ์จากแบบจำลองทั้งสามวิธีโดยการวิเคราะห์ด้วยค่าร้อยละของความถูกต้องในการพยากรณ์ ผลดังตาราง 4.9

ตาราง 4.9 การเปรียบเทียบประสิทธิภาพแบบจำลองของเทคนิคการทำเหมืองข้อมูลและเทคนิคการถดถอยโลจิสติก

แบบจำลอง	ร้อยละของการพยากรณ์ถูกต้อง
เทคนิคการทำเหมืองข้อมูล	
ต้นไม้ตัดสินใจ (TR9X4)	92.59
โครงข่ายประสาทเทียม (ANN1)	97.80
เทคนิคการถดถอยโลจิสติก	
การถดถอยโลจิสติกแบบหลายกลุ่ม	80.80



จากตาราง 4.9 สรุปผลการทดสอบแบบจำลองได้ว่า การพยากรณ์กลุ่มการดื่มสุราของประชาชนในจังหวัดหนองคาย ด้วยข้อมูลการสำรวจพฤติกรรมการสูบบุหรี่และดื่มสุราของประชากร พ.ศ. 2554 ซึ่งดำเนินการสำรวจโดยสำนักงานสถิติแห่งชาติ การสร้างแบบจำลองด้วยเทคนิคการทำเหมืองข้อมูลด้วยต้นไม้ตัดสินใจมีความถูกต้องของการพยากรณ์ ร้อยละ 92.59 วิธีโครงข่ายประสาทเทียม ANN1 ร้อยละ 97.80 และเทคนิคการถดถอยโลจิสติกด้วยวิธีการถดถอยโลจิสติกแบบหลายกลุ่ม ร้อยละ 80.80 จึงสรุปได้ว่าการสร้างแบบจำลองโดยเทคนิคการทำเหมืองข้อมูลด้วยวิธีโครงข่ายประสาทเทียม ANN1 มีความถูกต้องมากที่สุด



บทที่ 5

สรุปผล อภิปรายผล และข้อเสนอแนะ

จากการวิเคราะห์การดื่มสุราของประชาชนในจังหวัดหนองคาย โดยเทคนิคเหมืองข้อมูลและการถดถอยโลจิสติก โดยเทคนิคการทำเหมืองข้อมูลใช้แบบจำลองของต้นไม้ตัดสินใจที่มีอัลกอริทึม C4.5(J48) และแบบจำลองโครงข่ายประสาทเทียม และเทคนิคการวิเคราะห์การถดถอยโลจิสติก ใช้การวิเคราะห์การถดถอยโลจิสติกแบบหลายกลุ่ม โดยใช้ข้อมูลการสำรวจพฤติกรรมการสูบบุหรี่และดื่มสุราของประชากร พ.ศ. 2554 ซึ่งดำเนินการสำรวจโดยสำนักงานสถิติแห่งชาติ มาทำการทดลองหาแบบจำลองที่เหมาะสม โดยใช้โปรแกรมสำเร็จรูป แล้วนำแบบจำลองที่ได้นั้นมาสร้างเป็นตัวแบบเพื่อทำการพยากรณ์จากการศึกษา สามารถสรุปผลการศึกษา อภิปรายผล และข้อเสนอแนะ ได้ดังนี้

5.1 สรุปผล

5.2 อภิปรายผล

5.3 ข้อเสนอแนะ

5.1 สรุปผล

5.1.1 ความมุ่งหมายของการวิจัย

การศึกษานี้มีความมุ่งหมายหลักเพื่อสร้างและเปรียบเทียบหารูปแบบที่เหมาะสมในการพยากรณ์กลุ่มของการดื่มสุราของประชากรในจังหวัดหนองคาย โดยใช้เทคนิคของเทคนิคการทำเหมืองข้อมูลและการถดถอยโลจิสติก

5.1.2 วิธีการวิจัย

ผู้วิจัยใช้ข้อมูลทุติยภูมิ (Secondary Data) ที่ได้จากการสำรวจพฤติกรรมการสูบบุหรี่และดื่มสุราของประชากร พ.ศ. 2554 สำรวจโดยสำนักงานสถิติแห่งชาติ จำนวน 605 ราย โดยมีตัวแปรทั้งหมด 11 ตัว แยกเป็นตัวแปรอิสระ จำนวน 10 ตัวแปร คือ เพศ อายุ ระดับการศึกษาที่จบสูงสุด สถานภาพสมรส อาชีพ สถานภาพการทำงาน รายได้จากการทำงาน ปริมาณการดื่มสุรา คะแนนผลกระทบปัญหาสุขภาพที่มีผลต่อการทำงาน คะแนนผลกระทบปัญหาสุขภาพที่มีผลต่อการทำกิจวัตรประจำวัน และตัวแปรตาม คือ การดื่มสุรา ใน 12 เดือนที่แล้ว แบ่งออกเป็น 3 กลุ่ม คือ กลุ่มที่ไม่เคยดื่มสุรา กลุ่มที่ดื่มเล็กน้อยหรือนาน ๆ ครั้ง และกลุ่มที่ดื่มสม่ำเสมอ ผู้วิจัยทำการวิเคราะห์ข้อมูลด้วยโปรแกรมคอมพิวเตอร์สำเร็จรูป โดยใช้การวิเคราะห์การทำเหมืองข้อมูลด้วยเทคนิคของต้นไม้ตัดสินใจและโครงข่ายประสาทเทียม และการถดถอยโลจิสติกแบบหลายกลุ่ม

5.1.3 การทดสอบการทำเหมืองข้อมูล

5.1.3.1 การทดสอบการสร้างแบบจำลองต้นไม้ตัดสินใจ โดยใช้ข้อมูลตัวแปรอิสระ จำนวน 10 ตัวแปร คือ เพศ อายุ ระดับการศึกษาที่จบสูงสุด สถานภาพสมรส อาชีพ สถานภาพการทำงาน รายได้จากการทำงาน ปริมาณการดื่มสุรา คะแนนผลกระทบปัญหาสุขภาพที่มีผลต่อการทำงาน คะแนนผลกระทบปัญหาสุขภาพที่มีผลต่อการทำกิจวัตรประจำวัน และตัวแปรตาม คือ การดื่มสุรา ใน



12 เดือนที่แล้ว พบว่า แบบจำลองต้นไม้ตัดสินใจที่ใช้อัลกอริทึม C4.5(J48) แบบจำลอง TR9X4 โดยใช้เทคนิค EBP, กำหนดให้เป็นแบบ Binary, $CF=0.25$ จำนวน Leaf= 2, จำนวน fold ในการ pruning = 3 ให้ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) ค่าความเที่ยง (F-Measure) เท่ากัน คือ ร้อยละ 92.60 เป็นแบบจำลองที่ดีที่สุดจากจำนวนแบบจำลองทั้งหมดที่สร้างขึ้น โดยมีค่าความถูกต้องของการพยากรณ์กลุ่มการดื่มสูงถึง ร้อยละ 92.59

5.1.3.2 การทดสอบแบบจำลองโครงข่ายประสาทเทียมที่ใช้ในการศึกษาคั้งนี้ คือ แบบโครงข่ายประสาทเทียมชนิดโครงข่ายแบบ Single Layer และการเรียนรู้แบบแพร่กระจายย้อนกลับ โดยใช้ในสร้างแบบจำลองการวิเคราะห์การดื่มสุราของประชากรในจังหวัดหนองคาย จำนวน 3 แบบจำลอง จากผลการเรียนรู้ พบว่า แบบจำลองที่มีประสิทธิภาพสูงสุดคือ แบบจำลอง ANN1 (10-9-1) ที่มีจำนวนโหนดในชั้นซ่อน 9 ค่าอัตราการเรียนรู้ 0.2 ค่าโมเมนตัม 0.3 จำนวนรอบในการเรียนรู้ 50,000 รอบ มีค่าความคลาดเคลื่อนน้อยที่สุด คือ ในชุดการฝึกฝน MAPE เท่ากับ 0.0199 RMSE เท่ากับ 0.1012 และชุดการทดสอบ คือ ในชุดการฝึกฝน MAPE เท่ากับ 0.0007 RMSE เท่ากับ 0.0011 และเมื่อนำไปหาค่าความถูกต้องของการพยากรณ์ พบว่า มีความถูกต้องของการพยากรณ์กลุ่มการดื่มสุรา ร้อยละ 97.79

5.1.4 ผลการทดสอบแบบจำลองการถดถอยโลจิสติก จากการหาแบบจำลองที่ดีที่สุดจากแบบจำลองการถดถอยโลจิสติก โดยใช้ตัวแปรอิสระทั้งหมด 10 ตัว คือ เพศ อายุ ระดับการศึกษาที่จบสูงสุด สถานภาพสมรส อาชีพ สถานภาพการทำงาน รายได้จากการทำงาน ปริมาณการดื่มสุรา คะแนนผลกระทบปัญหาสุขภาพที่มีผลต่อการทำงานและคะแนนผลกระทบปัญหาสุขภาพที่มีผลต่อการทำกิจกรรมประจำวัน มาทำการวิเคราะห์เพื่อหาสมการในการพยากรณ์ พบว่า ตัวแปรอิสระที่เข้าในสมการมีจำนวน 3 ตัวแปร คือ ปริมาณการดื่ม เพศ และระดับการศึกษา ได้สมการในการพยากรณ์ ดังนี้

กลุ่มการดื่มสุราสม่ำเสมอเมื่อเทียบกับกลุ่มที่ไม่ดื่ม คือ

$$G(1) = -5.143 + 0.061(B11_31_L) + 2.564(A4) + 1.125(A5_re=2)$$

กลุ่มการดื่มสุราเล็กน้อยหรือนานๆครั้งเมื่อเทียบกับกลุ่มที่ไม่ดื่ม คือ

$$G(2) = -4.348 + 0.063(B11_31_L) + 1.813(A4) + 1.043(A5_re=2)$$

โดยที่ B11_31_L คือ ปริมาณการดื่มสุรา
A4 คือ เพศ
A5_re=2 คือ อายุ 25 – 59 ปี

เมื่อนำแบบจำลองการถดถอยโลจิสติกที่สร้าง โดยใช้ตัวแปรอิสระทั้งหมด 10 ตัว มาหาค่าความถูกต้องในการพยากรณ์กลุ่ม พบว่า มีค่าความถูกต้องในการพยากรณ์กลุ่มการดื่มสุรา ร้อยละ 80.80



5.1.5 การเปรียบเทียบประสิทธิภาพแบบจำลอง พบว่า การพยากรณ์กลุ่มการดื่มสุราของประชาชนในจังหวัดหนองคาย ด้วยข้อมูลการสำรวจพฤติกรรมการสูบบุหรี่และดื่มสุราของประชากร พ.ศ. 2554 ซึ่งดำเนินการสำรวจโดยสำนักงานสถิติแห่งชาติ แบบจำลองโดยเทคนิคการทำเหมืองข้อมูลด้วยวิธีโครงข่ายประสาทเทียม ANN1 มีความถูกต้องมากที่สุด คือ ร้อยละ 97.80 รองลงมา คือ สร้างแบบจำลองด้วยเทคนิคการทำเหมืองข้อมูลด้วยต้นไม้ตัดสินใจมีความถูกต้องของการพยากรณ์ ร้อยละ 92.59 และเทคนิคการถดถอยโลจิสติกด้วยวิธีการถดถอยโลจิสติกแบบหลายกลุ่ม ร้อยละ 80.8 ตามลำดับ

5.2 อภิปรายผล

จากการวิเคราะห์การดื่มสุราของประชาชนในจังหวัดหนองคาย โดยเทคนิคเหมืองข้อมูลและการถดถอยโลจิสติก พบว่า

แบบจำลองโดยเทคนิคการทำเหมืองข้อมูลด้วยวิธีโครงข่ายประสาทเทียม มีความถูกต้องในการพยากรณ์มากที่สุด ด้วยแบบจำนวน ANN1 ใช้ตัวแบบ คือ 10:9:1 ชั้นนำเข้า 10 โหนด ชั้นซ่อน 9 โหนด และชั้นแสดงผล 1 โหนด ในชั้นซ่อนใช้ฟังก์ชันการแปลงค่าคือ ฟังก์ชันซิกมอยด์ (Sigmoid Function = f) ส่วนชั้นแสดงผลใช้ฟังก์ชันกระตุ้นแบบผลรวมเชิงเส้น เนื่องจากข้อมูลที่นำมาใช้ในการพยากรณ์ส่วนมากเป็นข้อมูลในเชิงคุณภาพและเป็นข้อมูลที่ค่อนข้างเป็นแบบแผน ซึ่งเหมาะสมกับเทคนิคการทำเหมืองข้อมูลอีกทั้งมีจำนวนรอบในการวิเคราะห์ที่สูงจึงทำให้มีความถูกต้องมากขึ้น โดยมีความถูกต้องของการพยากรณ์กลุ่มการดื่มสุรา ร้อยละ 97.79 สอดคล้องกับการศึกษาของภรณ์ยา อามฤครัตน์ และพวง มีสัจ (2553) ได้ทำการศึกษการเปรียบเทียบประสิทธิภาพการจำแนกกลุ่มข้อมูลของโรคลมร้อนด้วยวิธีการทางเครือข่ายประสาทเทียม โดยได้แบบจำลองที่ให้ค่าพยากรณ์ที่มีความถูกต้องสูงที่สุด ร้อยละ 97.67

สำหรับแบบจำลองต้นไม้ตัดสินใจ จากการทดสอบการสร้างแบบจำลองต้นไม้ตัดสินใจที่ใช้อัลกอริทึม C4.5(J48) แบบจำลอง TR9X4 โดยใช้เทคนิค EBP, กำหนดให้เป็นแบบ Binary, $CF=0.25$ จำนวน Leaf= 2, จำนวน fold ในการ pruning = 3 ให้ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) ค่าความเหวี่ยง (F-Measure) เท่ากัน คือ ร้อยละ 92.60 เป็นแบบจำลองที่ดีที่สุดจากจำนวนแบบจำลองทั้งหมดที่สร้างขึ้น จากแบบจำลองต้นไม้ตัดสินใจที่ได้ออกมานั้นชี้ให้เห็นว่าทางที่สั้นที่สุดที่จะทำให้การจำแนกกลุ่มการดื่มสุราของประชาชนในจังหวัดหนองคายได้ถูกต้องแม่นยำนั้นอยู่กับปริมาณการดื่มเป็นหลัก นอกจากนี้ยังมีปัจจัยอื่น เช่น เพศ ช่วงอายุ รายได้ อาชีพ โดยมีค่าความถูกต้องของการพยากรณ์กลุ่มการดื่มสูงถึง ร้อยละ 92.59 ซึ่งสอดคล้องกับการศึกษาของ ธรรมพร หาญผจญศึก และนันทน์ภัส สุจิมา (2554) ที่ทำการพัฒนาโปรแกรมสำเร็จรูปเพื่อสนับสนุนการตัดสินใจในการประกอบการพาณิชย์อิเล็กทรอนิกส์ เพื่อสร้างแบบจำลองต้นไม้ตัดสินใจ ซึ่งเลือกใช้อัลกอริทึม C4.5(J48) และเลือกวิธีการ Pruning แบบ Error Base Pruning แบบจำลองต้นไม้ตัดสินใจที่มีร้อยละความถูกต้องในการจำแนกกลุ่มสูงสุด คือ ร้อยละ 95.83 มีจำนวน Leaf = 21 ขนาดของต้นไม้ = 39 ค่า Precision = 0.965 Recall = 0.958 F-Measure = 0.958 และร้อยละความถูกต้องในการทำนาย



ผลสูงสุด คือ ร้อยละ 65.00 ซึ่งเป็นแบบจำลองที่ใช้พารามิเตอร์ $CF = 0.25$, $M = 1$ และ $CF = 0.5$, $M = 1$

ส่วนแบบจำลองการถดถอยโลจิสติกแบบหลายกลุ่มนั้น ให้ผลประสิทธิภาพน้อยที่สุดในแบบจำลองที่ทำการศึกษานี้ เนื่องจากข้อมูลที่นำมาใช้ในการวิเคราะห์ส่วนใหญ่เป็นข้อมูลเชิงคุณภาพและเมื่อนำมาหาความสัมพันธ์ พบว่า มีความสัมพันธ์ระหว่างตัวแปรอิสระและตัวแปรตามมีค่าค่อนข้างต่ำ จึงทำให้แบบจำลองที่ได้มีประสิทธิภาพไม่ดีเท่าที่ควร ประกอบกับข้อมูลที่นำมาวิเคราะห์ในกลุ่มที่ตีพิมพ์เล็กน้อยหรือนานๆ ครั้ง และกลุ่มที่ตีพิมพ์สม่ำเสมออาจมีจำนวนน้อยเกินไป จึงทำให้การวิเคราะห์อาจไม่ดีเท่าที่ควร

การเปรียบเทียบประสิทธิภาพแบบจำลอง พบว่า การพยากรณ์กลุ่มการดื่มสุราของประชาชนในจังหวัดหนองคาย ด้วยแบบจำลองโดยเทคนิคการทำเหมืองข้อมูลด้วยวิธีโครงข่ายประสาทเทียม ANN1 มีความถูกต้องมากที่สุด มากกว่าแบบจำลองด้วยเทคนิคการทำเหมืองข้อมูลด้วยต้นไม้ตัดสินใจ และแบบจำลองที่สร้างด้วยเทคนิคการถดถอยโลจิสติกด้วยวิธีการถดถอยโลจิสติกแบบหลายกลุ่ม ซึ่งสอดคล้องกับการศึกษาของ สักการะ จูชูและธนพล เจนสุทธิเวชกุล (2553) ที่ทำการศึกษเปรียบเทียบการพยากรณ์ปริมาณการยืมหนังสือห้องสมุดโรงเรียนนาบอน ด้วยการวิเคราะห์การถดถอยและโครงข่ายประสาทเทียม แบบจำลองโครงข่ายประสาทเทียมมีค่าพยากรณ์ใกล้เคียงกับค่าจริงมากกว่าแบบจำลองการถดถอย และสอดคล้องกับการศึกษาของ Mukta and Kumar (2009) ที่ได้ศึกษาเปรียบเทียบจากจำนวนงานวิจัยทั้งหมด 96 งานวิจัย โครงข่ายประสาทเทียมให้ผลการพยากรณ์ที่ดีกว่าจำนวน 58 เปอร์เซ็นต์ โครงข่ายประสาทเทียมกับเทคนิคการวิเคราะห์ผลทางสถิติโดยทั่วไป ให้ผลการพยากรณ์เหมือนกันจำนวน 24 เปอร์เซ็นต์ และเทคนิคการวิเคราะห์ผลทางสถิติโดยทั่วไปให้ผลการพยากรณ์ที่ดีกว่าจำนวน 18 เปอร์เซ็นต์

5.3 ข้อเสนอแนะ

5.3.1 ข้อเสนอแนะจากการวิจัยครั้งนี้

5.3.1.1 จากผลการทดสอบแบบจำลองโดยเทคนิคการทำเหมืองข้อมูลด้วยวิธีโครงข่ายประสาทเทียม ANN1 (10:9:1) มีความถูกต้องในการพยากรณ์มากที่สุด สามารถนำไปใช้ประโยชน์ในการพยากรณ์กลุ่มของผู้ดื่มสุราในจังหวัดหนองคายต่อไป

5.3.1.2 การวิเคราะห์ของแบบจำลองที่ใช้เทคนิคการทำเหมืองข้อมูลด้วยวิธีต้นไม้ตัดสินใจ และโครงข่ายประสาทเทียม ควรมีการกำหนดค่าพารามิเตอร์ของวิธีนั้นๆ ที่หลากหลาย และทำการทดสอบเพื่อหาประสิทธิภาพที่ดีที่สุด

5.3.1.3 การวิเคราะห์ด้วยแบบจำลองการถดถอยโลจิสติก ผลการทดสอบมีประสิทธิภาพไม่ดีเท่าที่ควร เนื่องจากข้อมูลที่นำมาใช้ในการสร้างแบบจำลองของกลุ่มผู้ดื่มสุราสม่ำเสมอและกลุ่มผู้ดื่มสุราเล็กน้อยหรือนานๆ ครั้ง อาจมีจำนวนที่น้อยเกินไป



5.3.2 ข้อเสนอแนะสำหรับการวิจัยต่อไป

5.3.2.1 ควรมีการวิเคราะห์เพื่อหาปัจจัยนอกเหนือจากที่ใช้ในการวิจัยนี้ เพื่อนำมาใช้ในการสร้างแบบจำลองให้มีความถูกต้องมากยิ่งขึ้น เนื่องจากในงานวิจัยครั้งนี้ ตัวแปรอิสระมีความสัมพันธ์กับตัวแปรตามค่อนข้างน้อย

5.3.2.2 ควรมีการเพิ่มข้อมูลในการวิเคราะห์มากขึ้นโดยเฉพาะกลุ่มของกลุ่มผู้ดื่มสุราสม่ำเสมอและกลุ่มผู้ดื่มสุราเล็กน้อยหรือนานๆ ครั้ง ให้มากกว่านี้

5.3.2.3 ควรนำเทคนิคการทำเหมืองข้อมูลและการวิเคราะห์ถดถอยโลจิสติก ไปใช้ในการพยากรณ์ในด้านต่างๆ นอกเหนือจากงานวิจัยครั้งนี้ เพราะเป็นเทคนิคที่ไม่ค่อยยุ่งยากและง่ายต่อการนำไปใช้



เอกสารอ้างอิง



เอกสารอ้างอิง

การทำเหมืองข้อมูล *CSC662 Data Mining, Data Warehouse and Visualization*. [ออนไลน์].

ได้จาก: <http://pioneer.netsew.chula.ac.th/~skrung/csc662/> [สืบค้นเมื่อ 23 สิงหาคม 2555].

การเรียนรู้ต้นไม้ตัดสินใจ. [ออนไลน์]. ได้จาก: <http://th.wikipedia.org/wiki/การเรียนรู้ต้นไม้ตัดสินใจ> [สืบค้นเมื่อ 24 สิงหาคม 2555].

กรมสุขภาพจิต (2548) ผลกระทบเนื่องมาจากการดื่มแอลกอฮอล์. [ออนไลน์]. ได้จาก:

<http://www.dmh.go.th/news/view.asp?id=966> [สืบค้นเมื่อ 30 สิงหาคม 2555].

กฤษณ์ ขุนลึก (2555) รูปแบบการส่งเสริมการปฏิบัติตามมาตรการควบคุมเครื่องดื่มแอลกอฮอล์ของผู้จำหน่าย ในจังหวัดหนองคาย. วิทยานิพนธ์ปริญญาคุษฎ์บัณฑิต. มหาวิทยาลัยราชภัฏมหาสารคาม.

กัลยา วานิชย์บัญชา (2550) การวิเคราะห์ข้อมูลหลายตัวแปร. พิมพ์ครั้งที่ 3. กรุงเทพฯ, ธรรมสาร.

————— (2551) การวิเคราะห์สถิติขั้นสูงด้วย *SPSS for Windows*. พิมพ์ครั้งที่ 6. กรุงเทพฯ, ธรรมสาร.

เกรียงไกร พึ่งเชื้อ (2555) ปัจจัยทำนายพฤติกรรมการบริโภคเครื่องดื่มแอลกอฮอล์ของเยาวชนในจังหวัดนครปฐม. วิทยานิพนธ์ปริญญามหาบัณฑิต. มหาวิทยาลัยเกษตรศาสตร์.

โครงข่ายประสาทเทียม *Artificial Neural Network*. [ออนไลน์]. ได้จาก: <http://alaska.reru.ac.th/text/NN.pdf> [สืบค้นเมื่อ 20 สิงหาคม 2555].

ณัฐนันท์ กิ่งมณี (2554) การศึกษาการรับรู้โฆษณาตรงและโฆษณาแฝงสินค้าเครื่องดื่มแอลกอฮอล์ทางโทรทัศน์ในกลุ่มนักศึกษาระดับอุดมศึกษา. วิทยานิพนธ์ปริญญามหาบัณฑิต. มหาวิทยาลัยรังสิต.

คุษฎ์ อายุวัฒน์ และวณิชชา ณรงค์ชัย (2553) โครงการการประเมินผลการบังคับใช้กฎหมายตามพระราชบัญญัติควบคุมเครื่องดื่มแอลกอฮอล์ พ.ศ. 2551 ในเขตภาคตะวันออกเฉียงเหนือ (จังหวัดอุดรธานีและ จังหวัดขอนแก่น). กรุงเทพฯ, ศูนย์วิจัยปัญหาสุรา.

หัตตดาว แบนเนียน (2545) การเปรียบเทียบการพยากรณ์ด้วยวิธีโครงข่ายประสาทเทียมและวิธีของบ็อกซ์และเจนกินส์ : กรณีศึกษา อัตราแลกเปลี่ยนเงินตราต่างประเทศ. วิทยานิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต. มหาวิทยาลัยเกษตรศาสตร์.

เทพินทร์ พิชรานุกรักษ์ (2541) พฤติกรรมการบริโภคสุรา : ทบทวนองค์ความรู้ สถานการณ์ และปัจจัยที่เกี่ยวข้องภายใต้โครงการสืบสานวัฒนธรรมไทยสู่สุขภาพที่ยั่งยืน. นนทบุรี, สำนักพัฒนาวิชาการแพทย์ กระทรวงสาธารณสุข.

นฤพนธ์ ว่องประชาณุกุล (2548) วิธีที่เหมาะสมสำหรับการตัดกิ่งต้นไม้ตัดสินใจของการทำเหมืองข้อมูลทางด้านวิทยาศาสตร์. วิทยานิพนธ์ปริญญามหาบัณฑิต, มหาวิทยาลัยเทคโนโลยีสุรนารี

นิรุจน์ อุทธา และคณะ (2548) อุบัติการณ์และแบบแผนการดื่มเครื่องดื่มแอลกอฮอล์ของเด็กเยาวชนในสถานศึกษา ภาคตะวันออกเฉียงเหนือ ภายหลังนโยบายคุ้มครองเด็ก. ขอนแก่น, ศูนย์บำบัดยาเสพติดขอนแก่น สถาบันธัญญารักษ์.



- บุญเสริม หุตะแพทย์ (2552) การวิเคราะห์ความเกี่ยวข้องของการดื่มสุรากับความรุนแรงในการทำงาน. นนทบุรี, ศูนย์สหวิทยาการชุมชนศึกษา (ศูนย์วิจัยเฉพาะทาง) สถาบันวิจัยและพัฒนา มหาวิทยาลัยสุโขทัยธรรมาธิราช.
- ธรรมพร หาญผจญศึก และนันท์นภัส สุจิมา (2554) การพัฒนาโปรแกรมสำเร็จรูปเพื่อสนับสนุนการตัดสินใจในการประกอบการพาณิชย์อิเล็กทรอนิกส์. เชียงใหม่, มหาวิทยาลัยฟาร์อีสเทอร์น.
- ไพรัตน์ อันอินทร์ (2552) พฤติกรรมการดื่มเครื่องดื่มแอลกอฮอล์ของวัยรุ่นที่ส่งผลกระทบต่อสุขภาพและครอบครัว. กรุงเทพฯ, สำนักงานกองทุนสนับสนุนการสร้างเสริมสุขภาพ สถาบันวิจัยระบบสาธารณสุขและศูนย์วิจัยปัญหาสุรา กรมสุขภาพจิต.
- ภรณ์ยา อามฤตรัตน์ และพยุ่ง มีสัจ (2553) การเปรียบเทียบประสิทธิภาพการจำแนกกลุ่มข้อมูลของโรคมาร่วมด้วยวิธีการทางเครือข่ายประสาทเทียม. [ออนไลน์]. ได้จาก: http://202.44.34.144/nccitedoc/admin/nccit_files/NCCIT-2011040251.pdf [สืบค้นเมื่อ 1 ตุลาคม 2556].
- มันทนา ขอนดอก (2554) ปัจจัยที่มีความสัมพันธ์กับการดื่มเครื่องดื่มแอลกอฮอล์ในระยะแรกของนักเรียนชายมัธยมศึกษาตอนต้น อำเภอบางน้ำเปรี้ยว จังหวัดฉะเชิงเทรา. วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ. มหาวิทยาลัยบูรพา.
- รภัทรกร เพชรสุข (2553) รูปแบบความสัมพันธ์เชิงสาเหตุของพฤติกรรมการบริโภคเครื่องดื่มแอลกอฮอล์ของนักเรียนผู้ใช้บริการหน่วยงานอนามัยโรงเรียน. วิทยานิพนธ์ปรัชญาดุษฎีบัณฑิต. มหาวิทยาลัยคริสเตียน.
- ราชบัณฑิตยสถาน พจนานุกรม ฉบับราชบัณฑิตยสถาน พ.ศ. 2542. (2546) กรุงเทพฯ, ราชบัณฑิตยสถาน.
- วิกิพีเดีย สารานุกรมเสรี (2555) สุรา. [ออนไลน์] ได้จาก: <http://th.wikipedia.org/wiki/%E0%B8%AA%E0%B8%B8%E0%B8%A3%E0%B8%B2> [สืบค้นเมื่อ 19 กันยายน 2555].
- ศิริภรณ์ ชัยศิริ (2551) ผลของโปรแกรมการวางแผนจำหน่ายผู้ป่วยที่มีความผิดปกติเกี่ยวกับการใช้สุราในโรงพยาบาลจิตเวชนครราชสีมาราชชนินทร์. วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ. มหาวิทยาลัยราชภัฏนครราชสีมา.
- ศุภัทรา สุนทรภักย์ (2539) การพัฒนาเครื่องมือซอฟต์แวร์ในการจำลองการทำงานของนิวรอลเน็ตเวิร์คโดยใช้แบบจำลองแบ็กพรอปเพกชัน. วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ. จุฬาลงกรณ์มหาวิทยาลัย.
- ศูนย์วิจัยปัญหาสุรา (2550) รายงานสถานการณ์การบริโภคเครื่องดื่มแอลกอฮอล์รายจังหวัด พ.ศ.2550. นนทบุรี, เดอะ กราฟิโก ซิสเต็มส์.
- (2552) รายงานสถานการณ์สุราประจำปี พ.ศ. 2552. กรุงเทพฯ, พิมพ์ดีการพิมพ์.
- (2553) รายงานสถานการณ์สุราประจำปี พ.ศ. 2553. กรุงเทพฯ, ศูนย์วิจัยปัญหาสุรา.
- สุทธิรักษ์ ไชยรักษ์ และคณะ (2553) พฤติกรรมการดื่มสุราของนักศึกษามหาวิทยาลัยเทคโนโลยีสุรนารี ปีการศึกษา 2551. นครราชสีมา, มหาวิทยาลัยเทคโนโลยีสุรนารี.
- สุรยุทธ ปรัชญา (2541) การรู้จำอักษรไทยโดยโครงข่ายประสาทเทียมแบบแพร่กลับ. วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ. มหาวิทยาลัยเชียงใหม่.



- สักการะ จู่ซู่และฉนพล เจนสุทธิเวชกุล (2553) เปรียบเทียบการพยากรณ์ปริมาณการยืมหนังสือ ห้องสมุดโรงเรียนนาบอน ด้วยการวิเคราะห์การถดถอยและโครงข่ายประสาทเทียม. [ออนไลน์]. ได้จาก: http://202.44.34.144/nccitedoc/admin/nccit_files/NCCIT-20113103013754.pdf [สืบค้นเมื่อ 1 ตุลาคม 2556]
- สังจากจาง จอมโนนเขวา (2552) การพัฒนาแบบจำลองการเลือกรูปแบบการเดินทางและแบบจำลองระยะทางการเดินทาง ภายในเขตเทศบาลนครราชสีมา. วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ. มหาวิทยาลัยเทคโนโลยีสุรนารี.
- สำนักงานสถิติแห่งชาติ (2555) สรุปข้อมูลที่สำคัญการสำรวจพฤติกรรมและการดื่มสุราของประชากร พ.ศ. 2554. กรุงเทพฯ, เท็กซ์ แอน เจอร์นัล พับลิเคชัน.
- สำนักงานสรรพสามิตจังหวัดหนองคาย (2552) ข้อมูลการจัดเก็บเงินภาษีในพื้นที่จังหวัดหนองคาย. หนองคาย, สำนักงานสรรพสามิตจังหวัดหนองคาย.
- สำนักงานสาธารณสุขจังหวัดหนองคาย (2553) ข้อมูลเกี่ยวกับสถานการณ์และพฤติกรรมบริโภคเครื่องดื่มแอลกอฮอล์ของประชาชนจังหวัดหนองคาย. หนองคาย, สำนักงานสาธารณสุขจังหวัดหนองคาย.
- สำนักงานสำรวจสุขภาพประชาชนไทย (2551) คู่มือการสำรวจสุขภาพประชากรไทยโดยการตรวจร่างกายครั้งที่ 4 พ.ศ. 2551. นนทบุรี, สถาบันวิจัยระบบสาธารณสุข กระทรวงสาธารณสุข.
- สำนักงานคณะกรรมการพัฒนาการเศรษฐกิจและสังคมแห่งชาติ (2555) สดช. แถลงสภาพสังคมไทย สถิติเด็ก “ดื่มสุรานัก-ติดเฟซบุ๊ก-คุณแม่วัยใสฟ่ง-เสพยาไอซ์กระจาย”. [ออนไลน์]. ได้จาก: <http://thaipublica.org/2012/02/nesdb-announces-thai-society/> [สืบค้นเมื่อ 19 กันยายน 2555].
- อมราภรณ์ ผางแก้ว (2555) บทบาทของครอบครัวต่อการบำบัดรักษาผู้ติดสุรา. [ออนไลน์]. ได้จาก: <http://www.klb.dmh.go.th/index.php?m=article&op=detail&gr=&researchId=1730> [สืบค้นเมื่อ 15 กันยายน 2555].
- อัญชลี วานิชวีวัฒน์ (2540) การจดจำอักขระภาษาไทยตัวพิมพ์โดยโครงข่ายประสาทเทียมแบบเซลล์ออร์แกนไนซิงแมปซ. วิทยานิพนธ์ปริญญาโทวิศวกรรมศาสตรมหาบัณฑิต. สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง.
- อุทัยทิพย์ จันทรเพ็ญ (2554) ปัจจัยที่มีอิทธิพลต่อการดื่มเครื่องดื่มแอลกอฮอล์แบบผิดปกติของวัยรุ่นในจังหวัดขอนแก่น. วิทยานิพนธ์ปริญญาโทปรัชญาดุษฎีบัณฑิต. มหาวิทยาลัยขอนแก่น.
- Chalder M, Elgar FJ and Bennett P (2006) Drinking and motivation to drink among adolescent children of parents with alcohol problems. *Alcohol and Alcoholism*, 41(1), 107-113.
- Christopher JA (2004) *Beyond attitudinal ambivalence: effects of belief homogeneity on – attitude – intention - behavior relations*. United Kingdom, University of Sheffield.
- Duncan SC, Dauncan TE and Strycker LA (2006) Alcohol use from ages 9 to 16 : a cohortsequential latent growth model. *Drug and Alcohol Dependence*, 81(1), 71-81.



- Michael H (2006) Drink Choice: Factors Influencing Intention the Drink Wine.
International Journal of Wine Maketing. 2006, 52-56.
- Mukta P and Kumar UA (2009) Review Neural networks and statistical techniques:
 A review of applications. *Expert Systems with Applications*. 36, 2–17.
- Quinlan JR (1993) *C4.5: Programs for Machine Learning*. California, Morgan Kaufmann.
- Steingold L (2009) *Using the theory of the planned behavior to predict tobacco and alcohol use in South African students*. Cap Town, University of Cap Town.
- Witten IH and Frank E (2005) *Data Mining: Practical Machine Learning Tools and Techniques*. 2nd edition. California, Morgan Kaufmann.
- World Health Organization (2001) The alcohol use disorders identification test.
 [Online] Available from: http://www.who.int/substance_abuse/msbalcinuries.pdf [Cited June 10, 2010].
- (2001) International guide for monitoring alcohol consumption and related harm. [Online] Available from: <http://www.who.int/> [Cited June 10, 2010].



ภาคผนวก



ภาคผนวก ก
รายละเอียดของเทคนิคการทำเหมืองข้อมูล



แบบจำลองต้นไม้ตัดสินใจ

=== Run information ===

Scheme:weka.classifiers.trees.J48 -C 0.25 -B -M 2

Relation: R4_train-weka.filters.unsupervised.attribute.NumericToNominal-Rfirst-last

Instances: 545

Attributes: 11

a4

a5_re

a6_re

a7_re

a8_re

a10_re

inc

b11_31_r

b42

b43

b1_re

Test mode:split 90.0% train, remainder test

=== Classifier model (full training set) ===



J48 pruned tree

```

b11_31_r = 4: 3 (405.0/6.0)
b11_31_r != 4
| b11_31_r = 3: 1 (43.0/8.0)
| b11_31_r != 3
| | a4 = 1
| | | b11_31_r = 1
| | | | a6_re = 2
| | | | | inc = 2: 1 (4.0/1.0)
| | | | | inc != 2: 2 (8.0/1.0)
| | | | | a6_re != 2
| | | | | a8_re = 5: 1 (5.0/1.0)
| | | | | a8_re != 5
| | | | | | inc = 2: 2 (8.0/1.0)
| | | | | | inc != 2
| | | | | | | a5_re = 2: 1 (38.0/13.0)
| | | | | | | a5_re != 2: 2 (3.0)
| | | b11_31_r != 1: 1 (12.0/6.0)
| | a4 != 1: 2 (19.0/5.0)

```

Number of Leaves : 10

Size of the tree : 19

Time taken to build model: 0.02 seconds

=== Evaluation on test split ===

=== Summary ===

Correctly Classified Instances	50	92.5926 %
Incorrectly Classified Instances	4	7.4074 %
Kappa statistic	0.7985	
Mean absolute error	0.0737	
Root mean squared error	0.1902	
Relative absolute error	27.411 %	
Root relative squared error	53.973 %	
Total Number of Instances	54	



=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.75	0.043	0.75	0.75	0.75	0.981	1
	0.5	0.04	0.5	0.5	0.5	0.965	2
	1	0	1	1	1	1	3
Weighted Avg.	0.926	0.009	0.926	0.926	0.926	0.995	

=== Confusion Matrix ===

a b c <-- classified as

6 2 0 | a = 1

2 2 0 | b = 2

0 0 42 | c = 3

=== Run information ===

Scheme:weka.classifiers.trees.J48 -C 0.25 -B -M 2

Relation: R4_test-weka.filters.unsupervised.attribute.NumericToNominal-Rfirst-last

Instances: 60

Attributes: 11

a4

a5_re

a6_re

a7_re

a8_re

a10_re

inc

b11_31_r

b42

b43

b1_re

Test mode:split 90.0% train, remainder test



=== Classifier model (full training set) ===

J48 pruned tree

b11_31_r = 4: 3 (46.0)

b11_31_r != 4

| inc = 4: 1 (2.0)

| inc != 4

| | b11_31_r = 2: 1 (2.0)

| | b11_31_r != 2: 2 (10.0/1.0)

Number of Leaves : 4

Size of the tree : 7

Time taken to build model: 0.01 seconds

=== Evaluation on test split ===

=== Summary ===

Correctly Classified Instances	6	100	%
Incorrectly Classified Instances	0	0	%
Kappa statistic	1		
Mean absolute error	0.0123		
Root mean squared error	0.037		
Relative absolute error	5.1491 %		
Root relative squared error	11.7465 %		
Total Number of Instances	6		



=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0	0	0	0	0	?	1
	1	0	1	1	1	1	2
	1	0	1	1	1	1	3
Weighted Avg.	1	0	1	1	1	1	

=== Confusion Matrix ===

a b c <-- classified as

0 0 0 | a = 1

0 1 0 | b = 2

0 0 5 | c = 3



แบบจำลองโครงข่ายประสาทเทียม

=== Run information ===

Scheme:weka.classifiers.functions.MultilayerPerceptron -L 0.3 -M 0.2 -N 50000 -V 0 -S 0 -
E 20 -H 9 -G -R

Relation: R4_train-weka.filters.unsupervised.attribute.NumericToNominal-Rfirst-last

Instances: 545

Attributes: 11

a4

a5_re

a6_re

a7_re

a8_re

a10_re

inc

b11_31_r

b42

b43

b1_re

Test mode:evaluate on training data

=== Classifier model (full training set) ===

Sigmoid Node 0

Inputs Weights

Threshold 0.9198103559446239

Node 3 -20.59075414540555

Node 4 -2.5515267652399074

Node 5 -0.4099842369718232

Node 6 -8.699245878155141

Node 7 10.398257101308875

Node 8 -11.7164857891756

Node 9 9.239375035639588

Node 10 7.813911863953875

Node 11 0.3887242921459552



Sigmoid Node 1

Inputs Weights

Threshold -0.916828060803706

Node 3 18.01327695863768

Node 4 2.717564764686102

Node 5 -7.194299080550307

Node 6 9.0903547475181

Node 7 -10.79052021608874

Node 8 -2.56366631292875

Node 9 -9.778155116500933

Node 10 -8.204163542391978

Node 11 -5.647210671284129

Sigmoid Node 2

Inputs Weights

Threshold -7.829812873992227

Node 3 -4.659748220288009

Node 4 6.3082413743827885

Node 5 7.986541579613255

Node 6 0.30847875803291425

Node 7 -4.7435832572469705

Node 8 12.770832429919464

Node 9 2.2515818488130446

Node 10 0.5579477680948728

Node 11 7.459720363956233

Sigmoid Node 3

Inputs Weights

Threshold 0.7345048167512036

Attrib a4 3.926092375496757

Attrib a5_re=1 -0.7387272836805395

Attrib a5_re=2 -3.1844671631141535

Attrib a5_re=3 3.135076303921286

Attrib a6_re=1 -0.4800635023939014

Attrib a6_re=2 -4.610223349271398

Attrib a6_re=3 1.6047150011432276

Attrib a6_re=4 2.07213125193357

Attrib a7_re 2.390521831628715

Attrib a8_re=1 -3.231015145440249



Attrib a8_re=2 0.5594725567160821
 Attrib a8_re=3 0.45418523285184875
 Attrib a8_re=4 0.12540592434908784
 Attrib a8_re=5 -0.03484271588840253
 Attrib a10_re=1 -1.1736454783957502
 Attrib a10_re=2 1.6250687589140567
 Attrib a10_re=3 -1.4829021436891483
 Attrib a10_re=4 -1.105560350187495
 Attrib a10_re=5 0.009635909637959345
 Attrib inc=1 -2.0658540649557753
 Attrib inc=2 -6.155859622460793
 Attrib inc=3 1.069432618886939
 Attrib inc=4 3.205973828333641
 Attrib inc=5 1.8010624707414562
 Attrib b11_31_r=1 -5.327376340023291
 Attrib b11_31_r=2 0.3695410420392065
 Attrib b11_31_r=3 -8.852348059627921
 Attrib b11_31_r=4 12.352099801732528
 Attrib b42=0 0.8459864879981007
 Attrib b42=2 -2.201967397871615
 Attrib b42=3 -0.8932736094763798
 Attrib b42=5 -0.772944984660951
 Attrib b42=7 0.9571340396362144
 Attrib b43=0 -1.1171625883506657
 Attrib b43=2 1.5924582540154637
 Attrib b43=3 -0.9188138916458175
 Attrib b43=5 -1.0746166636259973

Sigmoid Node 4

Inputs Weights

Threshold -1.3982444928854225
 Attrib a4 0.00719022769267131
 Attrib a5_re=1 1.391099141060777
 Attrib a5_re=2 -0.3005643332135948
 Attrib a5_re=3 0.2750225411995211
 Attrib a6_re=1 3.36360115663891
 Attrib a6_re=2 -3.464981962862713
 Attrib a6_re=3 1.4529111409364373



Attrib a6_re=4 1.352330340530494
 Attrib a7_re 5.761200222632432
 Attrib a8_re=1 2.5866635248903367
 Attrib a8_re=2 -5.478559356734486
 Attrib a8_re=3 3.2572037901775506
 Attrib a8_re=4 1.799642756863347
 Attrib a8_re=5 1.8969128105566933
 Attrib a10_re=1 4.936924142441566
 Attrib a10_re=2 -0.13111842219475361
 Attrib a10_re=3 1.5005106641066859
 Attrib a10_re=4 -4.030738979563938
 Attrib a10_re=5 1.890196854447633
 Attrib inc=1 2.8573907849438513
 Attrib inc=2 -1.6063334500626991
 Attrib inc=3 6.690566526342832
 Attrib inc=4 -1.9163040989930205
 Attrib inc=5 -1.8583078703989313
 Attrib b11_31_r=1 -0.978155761947932
 Attrib b11_31_r=2 0.2344452310636724
 Attrib b11_31_r=3 -3.370020234353922
 Attrib b11_31_r=4 6.841606679893258
 Attrib b42=0 -0.28948588152918114
 Attrib b42=2 1.3355150713116202
 Attrib b42=3 0.11417806107595584
 Attrib b42=5 1.3217420623499978
 Attrib b42=7 1.6471619022086008
 Attrib b43=0 -0.9680149860085886
 Attrib b43=2 1.3715295392399114
 Attrib b43=3 1.0367753869217593
 Attrib b43=5 1.2783003534987427

Sigmoid Node 5

Inputs Weights
 Threshold -0.25653561766883876
 Attrib a4 2.976093736998645
 Attrib a5_re=1 0.35491270248370893
 Attrib a5_re=2 -0.8335954735659259
 Attrib a5_re=3 0.641310639641721



Attrib a6_re=1 0.7531622965598992
 Attrib a6_re=2 -1.94670350338615
 Attrib a6_re=3 0.43581977862020344
 Attrib a6_re=4 1.217535308335095
 Attrib a7_re -0.7327491475478901
 Attrib a8_re=1 0.8034902753656485
 Attrib a8_re=2 -1.900841236426679
 Attrib a8_re=3 0.7380471070740892
 Attrib a8_re=4 -0.02997896358332964
 Attrib a8_re=5 1.1047211633357137
 Attrib a10_re=1 1.3349710739551959
 Attrib a10_re=2 -1.1284797465294323
 Attrib a10_re=3 0.38449118114445335
 Attrib a10_re=4 -0.9309184336034199
 Attrib a10_re=5 1.1306706326593432
 Attrib inc=1 -1.3466996398191224
 Attrib inc=2 -0.5184167903393653
 Attrib inc=3 1.554866425576792
 Attrib inc=4 0.9655652607868229
 Attrib inc=5 0.07648356815983962
 Attrib b11_31_r=1 -4.5765366188094685
 Attrib b11_31_r=2 -0.829038926753594
 Attrib b11_31_r=3 -1.137109468313233
 Attrib b11_31_r=4 7.077179719817816
 Attrib b42=0 -0.12601293801223948
 Attrib b42=2 0.2254662247819472
 Attrib b42=3 0.24506091051143647
 Attrib b42=5 0.2371648874532782
 Attrib b42=7 0.1686992971504423
 Attrib b43=0 -0.08775792643827574
 Attrib b43=2 0.1622616506331607
 Attrib b43=3 0.21458736460722241
 Attrib b43=5 0.15698983722501125

Sigmoid Node 6

Inputs Weights

Threshold -0.9751298535550382

Attrib a4 2.548497774664271



Attrib a5_re=1 -3.1444728071592762
 Attrib a5_re=2 1.136121295148214
 Attrib a5_re=3 3.0110384820501945
 Attrib a6_re=1 2.2368977506024383
 Attrib a6_re=2 2.0192900990009868
 Attrib a6_re=3 -4.0276506827640866
 Attrib a6_re=4 1.77951566665129
 Attrib a7_re 1.1271792542439698
 Attrib a8_re=1 -3.7492822553583878
 Attrib a8_re=2 0.856408283194038
 Attrib a8_re=3 1.3122831588462214
 Attrib a8_re=4 2.9794896328829656
 Attrib a8_re=5 1.5128320942586797
 Attrib a10_re=1 7.014200758994607
 Attrib a10_re=2 0.4709299387820958
 Attrib a10_re=3 -6.646272930986811
 Attrib a10_re=4 0.5989880303374894
 Attrib a10_re=5 1.5081436059679072
 Attrib inc=1 -0.5393083994165929
 Attrib inc=2 -5.703170372084659
 Attrib inc=3 8.565060107481015
 Attrib inc=4 0.9761625709408671
 Attrib inc=5 -0.27507132768151704
 Attrib b11_31_r=1 4.6651386378382815
 Attrib b11_31_r=2 -3.142224013283354
 Attrib b11_31_r=3 -8.81722370780902
 Attrib b11_31_r=4 9.367852318268458
 Attrib b42=0 3.4074555046255077
 Attrib b42=2 0.03418609363296846
 Attrib b42=3 -2.4473767563830338
 Attrib b42=5 0.9027987019392044
 Attrib b42=7 1.0180001802017085
 Attrib b43=0 3.269591224987801
 Attrib b43=2 1.0395587017490968
 Attrib b43=3 -0.1347059403361648
 Attrib b43=5 -2.26453833826915



Sigmoid Node 7

Inputs Weights

Threshold 0.2943943394683707

Attrib a4 2.759503730817293

Attrib a5_re=1 -0.3694797991774564

Attrib a5_re=2 -5.685487948792933

Attrib a5_re=3 5.701431299829968

Attrib a6_re=1 3.878114132862237

Attrib a6_re=2 -8.332569163956654

Attrib a6_re=3 2.3653436444068348

Attrib a6_re=4 1.5154837544195947

Attrib a7_re 6.685996775006959

Attrib a8_re=1 -1.0917280702255228

Attrib a8_re=2 -3.350452628050172

Attrib a8_re=3 -0.8537242153015769

Attrib a8_re=4 1.3059755600006742

Attrib a8_re=5 3.043064974808234

Attrib a10_re=1 4.46167255263795

Attrib a10_re=2 0.9826213054557014

Attrib a10_re=3 -6.311946977829952

Attrib a10_re=4 -3.183010416433495

Attrib a10_re=5 3.059250456601634

Attrib inc=1 2.75610247302087

Attrib inc=2 -11.671575709395567

Attrib inc=3 6.199240377785616

Attrib inc=4 -0.49316500753283293

Attrib inc=5 2.226308152200405

Attrib b11_31_r=1 -0.35578374001707724

Attrib b11_31_r=2 2.300166297220218

Attrib b11_31_r=3 -5.214516843290196

Attrib b11_31_r=4 2.676375842591322

Attrib b42=0 0.14987153689570537

Attrib b42=2 -0.02594722463921914

Attrib b42=3 -0.3251812541587887

Attrib b42=5 -0.38437701337537744

Attrib b42=7 -0.44413464283042187

Attrib b43=0 -1.2081814279150402



Attrib b43=2 -0.5331974358742896

Attrib b43=3 -0.3062326642010422

Attrib b43=5 1.4391957245489244

Sigmoid Node 8

Inputs Weights

Threshold -0.15580152276028142

Attrib a4 4.8116049031995844

Attrib a5_re=1 1.4972406867438375

Attrib a5_re=2 -1.6966074049908562

Attrib a5_re=3 0.37008439676183547

Attrib a6_re=1 0.24988868587441965

Attrib a6_re=2 -2.988763186355651

Attrib a6_re=3 1.6690956806598112

Attrib a6_re=4 1.4542177455319643

Attrib a7_re 1.6258371746747056

Attrib a8_re=1 -1.2853052600853005

Attrib a8_re=2 0.09225066403853485

Attrib a8_re=3 2.2826265046393637

Attrib a8_re=4 0.30418875057704986

Attrib a8_re=5 -0.6965452510528612

Attrib a10_re=1 2.615995414540802

Attrib a10_re=2 -0.5780949183429552

Attrib a10_re=3 -1.0230824782288566

Attrib a10_re=4 0.21971215898538923

Attrib a10_re=5 -0.7385653742957253

Attrib inc=1 3.818412812758846

Attrib inc=2 -4.291475445141174

Attrib inc=3 0.9532133037090412

Attrib inc=4 -0.1776644099608855

Attrib inc=5 0.2624132385601448

Attrib b11_31_r=1 -2.8085077860396814

Attrib b11_31_r=2 -0.37392767603312377

Attrib b11_31_r=3 -3.302844168531331

Attrib b11_31_r=4 6.971729587827041

Attrib b42=0 0.030026217545133637

Attrib b42=2 0.1444251833290665

Attrib b42=3 0.10107243752889362



Attrib b42=5 0.1267688949539743
 Attrib b42=7 0.09581030998625324
 Attrib b43=0 -0.030694213277332343
 Attrib b43=2 0.14781887528431414
 Attrib b43=3 0.16694647633941936
 Attrib b43=5 0.13125585764139164

Sigmoid Node 9

Inputs Weights
 Threshold 0.20980891801954735
 Attrib a4 1.1243029602528496
 Attrib a5_re=1 -0.3451665683478581
 Attrib a5_re=2 -1.8988780964255814
 Attrib a5_re=3 1.98089418115434
 Attrib a6_re=1 2.6330954592809306
 Attrib a6_re=2 -2.7354865241296356
 Attrib a6_re=3 -0.015516363636112043
 Attrib a6_re=4 -0.2314465300770374
 Attrib a7_re -3.942270377529578
 Attrib a8_re=1 0.12515246807315367
 Attrib a8_re=2 -1.514646312631013
 Attrib a8_re=3 -0.6984431555036558
 Attrib a8_re=4 -0.9911804502699433
 Attrib a8_re=5 2.5820492914963817
 Attrib a10_re=1 -1.55292562562461
 Attrib a10_re=2 -1.6533174081230353
 Attrib a10_re=3 0.23301429453742076
 Attrib a10_re=4 -0.11910828793199896
 Attrib a10_re=5 2.5033102376028338
 Attrib inc=1 0.33706837674526163
 Attrib inc=2 -2.5778735164059254
 Attrib inc=3 1.5860552907122831
 Attrib inc=4 -0.6850054632018381
 Attrib inc=5 0.7735961769850493
 Attrib b11_31_r=1 -4.805305746678996
 Attrib b11_31_r=2 -1.6117595682374302
 Attrib b11_31_r=3 -2.883166086978787
 Attrib b11_31_r=4 8.952300246551474



Attrib b42=0 -0.09311341818174722
 Attrib b42=2 -0.13893339590356185
 Attrib b42=3 0.28038216034365787
 Attrib b42=5 -0.26669099942468283
 Attrib b42=7 -0.24755844938950158
 Attrib b43=0 0.34412824857549856
 Attrib b43=2 -0.2288865151808815
 Attrib b43=3 -0.17716767964224706
 Attrib b43=5 -0.28753561407311246

Sigmoid Node 10

Inputs Weights
 Threshold 0.626363606116305
 Attrib a4 4.747551556959136
 Attrib a5_re=1 0.0666350293394306
 Attrib a5_re=2 0.6986172723941398
 Attrib a5_re=3 -1.438801304245288
 Attrib a6_re=1 -4.099803607777902
 Attrib a6_re=2 0.054343027949907974
 Attrib a6_re=3 1.1598406209583931
 Attrib a6_re=4 1.5061873591534811
 Attrib a7_re -2.9479314750455683
 Attrib a8_re=1 -3.190299279668939
 Attrib a8_re=2 2.5700786322971436
 Attrib a8_re=3 2.7692614459478535
 Attrib a8_re=4 0.14308170435809128
 Attrib a8_re=5 -4.192471799945346
 Attrib a10_re=1 -0.41334266822480126
 Attrib a10_re=2 2.429557952122681
 Attrib a10_re=3 2.156400993022655
 Attrib a10_re=4 -1.9818434832116283
 Attrib a10_re=5 -4.146933743383845
 Attrib inc=1 -0.021733316575060533
 Attrib inc=2 3.925501805158764
 Attrib inc=3 -2.9246613019943615
 Attrib inc=4 -2.5125620063549254
 Attrib inc=5 -0.5328363562899855
 Attrib b11_31_r=1 -7.0503407763606685



Attrib b11_31_r=2 2.460989763953164
 Attrib b11_31_r=3 -3.204445675424654
 Attrib b11_31_r=4 6.457503385022711
 Attrib b42=0 0.07890964446539203
 Attrib b42=2 -0.038273551886256786
 Attrib b42=3 -0.6418352933837382
 Attrib b42=5 -0.5473990821935015
 Attrib b42=7 -0.817106938106888
 Attrib b43=0 0.8293192194339044
 Attrib b43=2 -0.8559150383696859
 Attrib b43=3 -0.7209832786516247
 Attrib b43=5 -0.5538642887961993

Sigmoid Node 11

Inputs Weights
 Threshold -0.2695862509674588
 Attrib a4 2.8941982627689256
 Attrib a5_re=1 0.4096730065078945
 Attrib a5_re=2 -0.7876734949493992
 Attrib a5_re=3 0.6920525442083402
 Attrib a6_re=1 0.7245552052086129
 Attrib a6_re=2 -1.9474841522552457
 Attrib a6_re=3 0.413545311981586
 Attrib a6_re=4 1.1716440863131135
 Attrib a7_re -0.7332821244816063
 Attrib a8_re=1 0.7595481496989325
 Attrib a8_re=2 -1.8302511351282402
 Attrib a8_re=3 0.6780457464549202
 Attrib a8_re=4 -0.018557642945391344
 Attrib a8_re=5 1.1032842789516137
 Attrib a10_re=1 1.2600518954754094
 Attrib a10_re=2 -1.1132165914059968
 Attrib a10_re=3 0.44190347809880104
 Attrib a10_re=4 -0.9167128917193234
 Attrib a10_re=5 1.074936849063814
 Attrib inc=1 -1.2568646492702835
 Attrib inc=2 -0.529318240969968
 Attrib inc=3 1.594227005582474



Attrib inc=4 0.8958642778331617
 Attrib inc=5 0.0733512756504215
 Attrib b11_31_r=1 -4.532948478706089
 Attrib b11_31_r=2 -0.822296556269419
 Attrib b11_31_r=3 -1.0840946156303721
 Attrib b11_31_r=4 6.929873103687901
 Attrib b42=0 -0.11590540937183212
 Attrib b42=2 0.24031035722229538
 Attrib b42=3 0.25259591559118194
 Attrib b42=5 0.1995269138004495
 Attrib b42=7 0.10683332615368345
 Attrib b43=0 -0.0765396789652646
 Attrib b43=2 0.13383853608304494
 Attrib b43=3 0.20610078352239794
 Attrib b43=5 0.1762286699527412

Class 1

Input

Node 0

Class 2

Input

Node 1

Class 3

Input

Node 2

Time taken to build model: 565.6 seconds



=== Evaluation on training set ===

=== Summary ===

Correctly Classified Instances	533	97.7982 %
Incorrectly Classified Instances	12	2.2018 %
Kappa statistic	0.9478	
Mean absolute error	0.0199	
Root mean squared error	0.1012	
Relative absolute error	7.0093 %	
Root relative squared error	26.8794 %	
Total Number of Instances	545	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.926	0.011	0.938	0.926	0.932	0.989	1
	0.906	0.008	0.935	0.906	0.921	0.998	2
	1	0.021	0.993	1	0.996	0.997	3
Weighted Avg.	0.978	0.018	0.978	0.978	0.978	0.996	

=== Confusion Matrix ===

a b c <-- classified as

75 4 2 | a = 1

5 58 1 | b = 2

0 0 400 | c = 3



=== Run information ===

Scheme:weka.classifiers.functions.MultilayerPerceptron -L 0.3 -M 0.2 -N 50000 -V 0 -S 0 -
E 20 -H 9 -G -R

Relation: R4_test-weka.filters.unsupervised.attribute.NumericToNominal-Rfirst-last

Instances: 60

Attributes: 11

a4
a5_re
a6_re
a7_re
a8_re
a10_re
inc
b11_31_r
b42
b43
b1_re

Test mode:evaluate on training data

=== Classifier model (full training set) ===

Sigmoid Node 0

Inputs Weights

Threshold -0.7096859150018983

Node 3 -1.3661690634131778

Node 4 -6.008757230722381

Node 5 3.326623962804185

Node 6 -1.356375786288323

Node 7 -1.4113052020296006

Node 8 -1.3684024438757527

Node 9 4.2298048693817245

Node 10 -1.5958067804660063

Node 11 -1.9737040136602768

Sigmoid Node 1

Inputs Weights

Threshold 2.583453954320043



Node 3 -1.3357068772792335
 Node 4 4.419736343210017
 Node 5 -4.2472402230060275
 Node 6 -1.601906366125365
 Node 7 -0.9815549698159793
 Node 8 -1.2480268719749326
 Node 9 -5.152408443324045
 Node 10 -1.1360225035692382
 Node 11 -0.7211722790046623

Sigmoid Node 2

Inputs Weights
 Threshold -6.7819687616784625
 Node 3 2.552514296277092
 Node 4 -0.23051077337196815
 Node 5 -0.05376834996815166
 Node 6 2.950101177467769
 Node 7 2.1044342202540816
 Node 8 2.437359404773849
 Node 9 0.2012066473703068
 Node 10 2.6716565247118633
 Node 11 2.470140883909841

Sigmoid Node 3

Inputs Weights
 Threshold -0.13487793507918827
 Attrib a4 0.7762491358175122
 Attrib a5_re=1 -0.005634220791339901
 Attrib a5_re=2 -0.12258015860829129
 Attrib a5_re=3 0.1455205148096365
 Attrib a6_re=1 -0.052012442518862036
 Attrib a6_re=2 0.3328838247741469
 Attrib a6_re=3 -0.19524372507433274
 Attrib a6_re=4 0.2458696919261364
 Attrib a7_re 0.0429335233973615
 Attrib a8_re=1 -0.267188054188525
 Attrib a8_re=2 0.033962991706443395
 Attrib a8_re=3 0.3504244397554104
 Attrib a8_re=4 -0.02756097519368279



Attrib a8_re=5 0.13150133507083042
 Attrib a10_re=1 0.008940265319379425
 Attrib a10_re=2 0.10748235299173638
 Attrib a10_re=3 0.02240028788129612
 Attrib a10_re=4 0.029763524902886668
 Attrib a10_re=5 0.1462248354866796
 Attrib inc=1 0.10114635022839057
 Attrib inc=2 -0.008998921657208772
 Attrib inc=3 -0.014766344241393926
 Attrib inc=4 0.04953966684911945
 Attrib inc=5 0.20034669781289333
 Attrib b11_31_r=1 -0.8802587811748223
 Attrib b11_31_r=2 -0.26680449793378486
 Attrib b11_31_r=3 -0.805204328478195
 Attrib b11_31_r=4 2.1561137665824988
 Attrib b42 0.033486931761999884
 Attrib b43 0.011908782901326666

Sigmoid Node 4

Inputs Weights
 Threshold 0.07677167957731879
 Attrib a4 1.0044582676412481
 Attrib a5_re=1 0.6475588945225564
 Attrib a5_re=2 -0.9501442555169731
 Attrib a5_re=3 0.17686096853205188
 Attrib a6_re=1 0.3180236513582627
 Attrib a6_re=2 0.5561931327632925
 Attrib a6_re=3 -1.0149989524374512
 Attrib a6_re=4 0.005144446382946082
 Attrib a7_re 0.17611063929802606
 Attrib a8_re=1 -0.6556616565676511
 Attrib a8_re=2 -0.40489236131777534
 Attrib a8_re=3 0.8321930120520752
 Attrib a8_re=4 0.1464067590750961
 Attrib a8_re=5 -0.15154260475175207
 Attrib a10_re=1 0.1096464781585544
 Attrib a10_re=2 -0.5375617183668159
 Attrib a10_re=3 0.13140166271649123



Attrib a10_re=4 0.2468644614245882
 Attrib a10_re=5 -0.17539887067015897
 Attrib inc=1 0.39880477980133033
 Attrib inc=2 0.442521628947813
 Attrib inc=3 -0.008194761578361884
 Attrib inc=4 -1.1137459199886308
 Attrib inc=5 0.01848395305148563
 Attrib b11_31_r=1 0.9360376917076113
 Attrib b11_31_r=2 -2.0173132917902454
 Attrib b11_31_r=3 -0.06775531037782963
 Attrib b11_31_r=4 0.9179306425180904
 Attrib b42 -2.8612105049124508
 Attrib b43 -0.033472141554108756

Sigmoid Node 5

Inputs Weights
 Threshold 0.20691217989180885
 Attrib a4 -0.10179842496531569
 Attrib a5_re=1 -0.7178262535959099
 Attrib a5_re=2 0.6706209547991135
 Attrib a5_re=3 -0.12750381363828012
 Attrib a6_re=1 -0.41510964828892394
 Attrib a6_re=2 -0.3717605879507091
 Attrib a6_re=3 0.4879616190833415
 Attrib a6_re=4 -0.16021765391613618
 Attrib a7_re 0.07015599534499214
 Attrib a8_re=1 0.09896112315074078
 Attrib a8_re=2 0.09832878985922973
 Attrib a8_re=3 -0.7091520185805344
 Attrib a8_re=4 -0.44328455824411195
 Attrib a8_re=5 0.18855481399513627
 Attrib a10_re=1 -0.16841897211000084
 Attrib a10_re=2 0.21024089266619736
 Attrib a10_re=3 -0.4830272026952873
 Attrib a10_re=4 -0.48042972477719226
 Attrib a10_re=5 0.1270307779198642
 Attrib inc=1 -0.548628820444708
 Attrib inc=2 -0.6715894227532201



Attrib inc=3 -0.2453391495657106
 Attrib inc=4 0.8882854752935021
 Attrib inc=5 -0.13528818918241742
 Attrib b11_31_r=1 -1.8736727344064374
 Attrib b11_31_r=2 1.0028246708178519
 Attrib b11_31_r=3 -1.1589443870583709
 Attrib b11_31_r=4 1.4701877764496998
 Attrib b42 2.434473491666665
 Attrib b43 0.01594686951090471

Sigmoid Node 6

Inputs Weights
 Threshold -0.0991771183068823
 Attrib a4 0.7835519780635144
 Attrib a5_re=1 0.025271822163252602
 Attrib a5_re=2 -0.056603291845277924
 Attrib a5_re=3 0.16653197283076573
 Attrib a6_re=1 -0.06841392982166919
 Attrib a6_re=2 0.29775096957858577
 Attrib a6_re=3 -0.21981896710008433
 Attrib a6_re=4 0.2240092279293999
 Attrib a7_re 0.04137544377737173
 Attrib a8_re=1 -0.27464645125877774
 Attrib a8_re=2 0.07916332860279633
 Attrib a8_re=3 0.3614029754903284
 Attrib a8_re=4 -0.01634749947742783
 Attrib a8_re=5 0.1981590235067132
 Attrib a10_re=1 -0.004316093156368928
 Attrib a10_re=2 0.15376513463784194
 Attrib a10_re=3 -0.013227961137718609
 Attrib a10_re=4 0.028348470602985818
 Attrib a10_re=5 0.15121927057238868
 Attrib inc=1 0.06609241065244618
 Attrib inc=2 -0.03276146846140676
 Attrib inc=3 -0.022933930095443002
 Attrib inc=4 0.06850575758940299
 Attrib inc=5 0.25550216137062204
 Attrib b11_31_r=1 -0.9848553186769412



Attrib b11_31_r=2 -0.25074873801645764
 Attrib b11_31_r=3 -0.8619669301683581
 Attrib b11_31_r=4 2.326763318833864
 Attrib b42 0.08744594795637681
 Attrib b43 -0.04696606932834255

Sigmoid Node 7

Inputs Weights
 Threshold -0.07036636169959301
 Attrib a4 0.7273922277283829
 Attrib a5_re=1 0.07646589292776296
 Attrib a5_re=2 -0.11977191000227105
 Attrib a5_re=3 0.10977622531327878
 Attrib a6_re=1 -0.019519136652038575
 Attrib a6_re=2 0.285201207791193
 Attrib a6_re=3 -0.2987763935003911
 Attrib a6_re=4 0.19267804252518986
 Attrib a7_re 0.063427421497507
 Attrib a8_re=1 -0.26900448091929186
 Attrib a8_re=2 0.053088063636490414
 Attrib a8_re=3 0.34548043063927425
 Attrib a8_re=4 0.03444649775250517
 Attrib a8_re=5 0.15080600039327763
 Attrib a10_re=1 0.0029546376294025726
 Attrib a10_re=2 0.1192511747180326
 Attrib a10_re=3 0.027408311264237838
 Attrib a10_re=4 0.03946399805378656
 Attrib a10_re=5 0.1370342814730205
 Attrib inc=1 0.08231854520686624
 Attrib inc=2 0.004720424968995843
 Attrib inc=3 -0.016411285896393495
 Attrib inc=4 0.01684599552348077
 Attrib inc=5 0.23946409854838777
 Attrib b11_31_r=1 -0.6812901554317995
 Attrib b11_31_r=2 -0.3580919473236045
 Attrib b11_31_r=3 -0.6980568889768484
 Attrib b11_31_r=4 1.9497169840039306
 Attrib b42 0.030449480966121663



Attrib b43 0.022879267506905346

Sigmoid Node 8

Inputs Weights

Threshold -0.07825472927630302

Attrib a4 0.7904608269024327

Attrib a5_re=1 -0.0013247440360720127

Attrib a5_re=2 -0.11484850451254987

Attrib a5_re=3 0.12180033964971258

Attrib a6_re=1 -0.0894987217254055

Attrib a6_re=2 0.34020458376685925

Attrib a6_re=3 -0.24712404676885008

Attrib a6_re=4 0.1762609855805382

Attrib a7_re 0.07807871349506951

Attrib a8_re=1 -0.3127654929837577

Attrib a8_re=2 0.0072039720889477245

Attrib a8_re=3 0.371597535591229

Attrib a8_re=4 0.015722631506171914

Attrib a8_re=5 0.17827546125328259

Attrib a10_re=1 -0.013470457529768182

Attrib a10_re=2 0.10472648598103193

Attrib a10_re=3 -0.0031879978764385573

Attrib a10_re=4 0.045328952418213136

Attrib a10_re=5 0.11977049058415808

Attrib inc=1 0.12151187525845727

Attrib inc=2 0.007200089827026077

Attrib inc=3 0.002513112191980879

Attrib inc=4 0.03320964304529114

Attrib inc=5 0.20912203286658707

Attrib b11_31_r=1 -0.8303021851760567

Attrib b11_31_r=2 -0.24709732503597737

Attrib b11_31_r=3 -0.8127064664794678

Attrib b11_31_r=4 2.081871261479063

Attrib b42 0.0982596813514893

Attrib b43 -0.0023757695994812353

Sigmoid Node 9

Inputs Weights

Threshold 0.30617609568060994



Attrib a4 -0.10547986487512379
 Attrib a5_re=1 -0.8310024165149473
 Attrib a5_re=2 0.6696939488527816
 Attrib a5_re=3 -0.18075242905211503
 Attrib a6_re=1 -0.45438413747086026
 Attrib a6_re=2 -0.4370854534183025
 Attrib a6_re=3 0.5397409725725556
 Attrib a6_re=4 -0.254077644098746
 Attrib a7_re 0.06342544849479972
 Attrib a8_re=1 0.12252401905298557
 Attrib a8_re=2 0.1294218238060753
 Attrib a8_re=3 -0.7531292491846107
 Attrib a8_re=4 -0.5119894652988667
 Attrib a8_re=5 0.15803442468220344
 Attrib a10_re=1 -0.18914159708093325
 Attrib a10_re=2 0.2026542920278074
 Attrib a10_re=3 -0.4649066697302845
 Attrib a10_re=4 -0.5482768924891231
 Attrib a10_re=5 0.17719837456638324
 Attrib inc=1 -0.5993748717064482
 Attrib inc=2 -0.733647967025466
 Attrib inc=3 -0.2876894091839348
 Attrib inc=4 1.012295148625865
 Attrib inc=5 -0.21543546836750596
 Attrib b11_31_r=1 -2.121308951787239
 Attrib b11_31_r=2 1.1248717167028923
 Attrib b11_31_r=3 -1.3470735545697172
 Attrib b11_31_r=4 1.7705751065376671
 Attrib b42 2.764644500550137
 Attrib b43 -0.03177518077905532

Sigmoid Node 10

Inputs Weights
 Threshold -0.05779303057140363
 Attrib a4 0.8371402543635698
 Attrib a5_re=1 0.0737797871698359
 Attrib a5_re=2 -0.14533149251968763
 Attrib a5_re=3 0.14954709599342209



Attrib a6_re=1 -0.0369423200040363
 Attrib a6_re=2 0.3360286746160609
 Attrib a6_re=3 -0.2998740693546585
 Attrib a6_re=4 0.2191948104736547
 Attrib a7_re 0.04944570258079287
 Attrib a8_re=1 -0.3053686390497116
 Attrib a8_re=2 0.07380416964976977
 Attrib a8_re=3 0.3641447786343619
 Attrib a8_re=4 0.03901837732515178
 Attrib a8_re=5 0.12984816343479552
 Attrib a10_re=1 0.008832035250172284
 Attrib a10_re=2 0.1246294333904485
 Attrib a10_re=3 0.00487157299477935
 Attrib a10_re=4 0.031018153326723842
 Attrib a10_re=5 0.15317803222907644
 Attrib inc=1 0.07683273618994381
 Attrib inc=2 -0.012114567950970889
 Attrib inc=3 -0.007187555610711064
 Attrib inc=4 -0.006895203027449852
 Attrib inc=5 0.22551891681645506
 Attrib b11_31_r=1 -0.7847245029452065
 Attrib b11_31_r=2 -0.35642512237982954
 Attrib b11_31_r=3 -0.7559107527232178
 Attrib b11_31_r=4 2.187440983757975
 Attrib b42 0.020228744046921633
 Attrib b43 0.021303339082304765

Sigmoid Node 11

Inputs Weights
 Threshold -0.07845393182054727
 Attrib a4 0.8296989678544707
 Attrib a5_re=1 0.15371475500405143
 Attrib a5_re=2 -0.19376803664880823
 Attrib a5_re=3 0.1471103750319288
 Attrib a6_re=1 -0.014992513415435938
 Attrib a6_re=2 0.3434212645115564
 Attrib a6_re=3 -0.3452046767595124
 Attrib a6_re=4 0.21701650598225616



Attrib a7_re 0.10928247047010167
 Attrib a8_re=1 -0.34371928946271363
 Attrib a8_re=2 0.011072903181099665
 Attrib a8_re=3 0.42205365533234795
 Attrib a8_re=4 0.03356445296235303
 Attrib a8_re=5 0.15614352952954172
 Attrib a10_re=1 0.010214420624209946
 Attrib a10_re=2 0.049291539356507263
 Attrib a10_re=3 0.08427971145047358
 Attrib a10_re=4 0.03790762285258712
 Attrib a10_re=5 0.06449176939278257
 Attrib inc=1 0.16119489606989873
 Attrib inc=2 0.01958847304078606
 Attrib inc=3 -0.001930182895554143
 Attrib inc=4 -0.09863975625147198
 Attrib inc=5 0.18836003499675852
 Attrib b11_31_r=1 -0.556292572748468
 Attrib b11_31_r=2 -0.511545384796393
 Attrib b11_31_r=3 -0.6789286969405044
 Attrib b11_31_r=4 2.059317759001257
 Attrib b42 -0.16455154706859698
 Attrib b43 -0.012053049723794618

Class 1

Input

Node 0

Class 2

Input

Node 1

Class 3

Input

Node 2

Time taken to build model: 74.51 seconds



=== Evaluation on training set ===

=== Summary ===

Correctly Classified Instances	60	100	%
Incorrectly Classified Instances	0	0	%
Kappa statistic	1		
Mean absolute error	0.0007		
Root mean squared error	0.0011		
Relative absolute error	0.2713	%	
Root relative squared error	0.3034	%	
Total Number of Instances	60		

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	1	0	1	1	1	1	1
	1	0	1	1	1	1	2
	1	0	1	1	1	1	3
Weighted Avg.	1	0	1	1	1	1	

=== Confusion Matrix ===

a b c <-- classified as

5 0 0 | a = 1

0 9 0 | b = 2

0 0 46 | c = 3



ภาคผนวก ข
รายละเอียดแบบจำลองการถดถอยโลจิสติก



Model Fitting Information

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	910.742			
Final	620.033	290.709	36	.000

Likelihood Ratio Tests

Effect	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	620.033(a)	.000	0	.
RE_INC	654.838(b)	34.805	2	.000
B11_31_L	623.974(c)	3.941	2	.139
B42	20849314.367(b)	20848694.334	2	.000
B43	77166.929(b)	76546.896	2	.000
A4	287.515(c)	.	2	.
A5_RE	731.281(b)	111.249	4	.000
A6_RE	744.741(b)	124.709	6	.000
A7_RE	282.834(c)	.	2	.
A8_RE	633.972(b)	13.939	6	.030
A10_RE	70643764.721(b)	70643144.688	6	.000

The chi-square statistic is the difference in -2 log-likelihoods between the final model and a reduced model. The reduced model is formed by omitting an effect from the final model. The null hypothesis is that all parameters of that effect are 0.

a This reduced model is equivalent to the final model because omitting the effect does not increase the degrees of freedom.

b The log-likelihood value cannot be further increased after maximum number of step-halving.

c There is possibly a quasi-complete separation in the data. Either the maximum likelihood estimates do not exist or some parameter estimates are infinite.



Parameter Estimates

กลุ่มการตีพิมพ์ สมาคมน	กลุ่มการตีพิมพ์(a)	B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp(B)	
								Lower Bound	Upper Bound
กลุ่มที่ตีพิมพ์ สมาคมน	Intercept	-5.143	.966	28.344	1	.000			
	RE_INC	.000	.000	.397	1	.529	1.000	1.000	1.000
	B11_31_L	.061	.015	16.487	1	.000	1.063	1.032	1.095
	B42	.857	.740	1.342	1	.247	2.357	.553	10.054
	B43	1.604	.000	.	1	.	4.973	4.973	4.973
	[A4=1]	2.564	.404	40.329	1	.000	12.991	5.888	28.664
	[A4=2]	0(b)	.	.	0	.	.	.	.
	[A5_RE=1]	-.555	1.010	.302	1	.583	.574	.079	4.159
	[A5_RE=2]	1.125	.472	5.685	1	.017	3.079	1.222	7.761
	[A5_RE=3]	0(b)	.	.	0	.	.	.	.
	[A6_RE=1]	.724	.820	.778	1	.378	2.062	.413	10.293
	[A6_RE=2]	1.000	.830	1.451	1	.228	2.718	.534	13.835
	[A6_RE=3]	1.946	.997	3.811	1	.051	7.000	.992	49.385
	[A6_RE=4]	0(b)	.	.	0	.	.	.	.
	[A7_RE=1]	.179	.370	.233	1	.629	1.196	.579	2.471
	[A7_RE=2]	0(b)	.	.	0	.	.	.	.
	[A8_RE=1]	.398	.719	.306	1	.580	1.489	.363	6.099
	[A8_RE=2]	.000	.716	.000	1	1.000	1.000	.246	4.070
	[A8_RE=3]	-.204	.665	.094	1	.759	.815	.222	3.001
	[A8_RE=4]	.687	1.175	.342	1	.558	1.988	.199	19.882
	[A8_RE=5]	0(b)	.	.	0	.	.	.	.
	[A10_RE=1]	-.485	.551	.777	1	.378	.615	.209	1.811
	[A10_RE=2]	-.228	.668	.116	1	.733	.796	.215	2.952
	[A10_RE=3]	.598	.883	.458	1	.498	1.818	.322	10.273
	[A10_RE=4]	0(b)	.	.	0	.	.	.	.
	[A10_RE=5]	0(b)	.	.	0	.	.	.	.



กลุ่มการดื่มสุรา(a)		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp(B)	
								Lower Bound	Upper Bound
กลุ่มที่ดื่มเล็กน้อยหรือนาน ๆ ครั้ง	Intercept	-4.348	.875	24.672	1	.000			
	RE_INC	.000	.000	.258	1	.612	1.000	1.000	1.000
	B11_31_L	.063	.015	17.737	1	.000	1.065	1.034	1.097
	B42	.777	.740	1.103	1	.294	2.175	.510	9.273
	B43	3.908	.000	.	1	.	49.804	49.804	49.804
	[A4=1]	1.813	.338	28.772	1	.000	6.128	3.160	11.884
	[A4=2]	0(b)	.	.	0	.	.	.	.
	[A5_RE=1]	-.224	1.032	.047	1	.829	.800	.106	6.047
	[A5_RE=2]	1.043	.470	4.927	1	.026	2.839	1.130	7.131
	[A5_RE=3]	0(b)	.	.	0	.	.	.	.
	[A6_RE=1]	.339	.739	.210	1	.646	1.403	.330	5.966
	[A6_RE=2]	.683	.755	.817	1	.366	1.979	.450	8.698
	[A6_RE=3]	1.370	.962	2.028	1	.154	3.935	.597	25.933
	[A6_RE=4]	0(b)	.	.	0	.	.	.	.
	[A7_RE=1]	-.250	.372	.450	1	.502	.779	.376	1.615
	[A7_RE=2]	0(b)	.	.	0	.	.	.	.
	[A8_RE=1]	.808	.693	1.360	1	.244	2.244	.577	8.733
	[A8_RE=2]	.661	.693	.911	1	.340	1.937	.498	7.535
	[A8_RE=3]	.653	.622	1.101	1	.294	1.921	.568	6.501
	[A8_RE=4]	1.963	1.102	3.172	1	.075	7.118	.821	61.712
	[A8_RE=5]	0(b)	.	.	0	.	.	.	.
	[A10_RE=1]	-.641	.500	1.642	1	.200	.527	.198	1.404
	[A10_RE=2]	-.644	.623	1.068	1	.301	.525	.155	1.782
	[A10_RE=3]	-.386	.855	.204	1	.651	.680	.127	3.629
	[A10_RE=4]	0(b)	.	.	0	.	.	.	.
	[A10_RE=5]	0(b)	.	.	0	.	.	.	.

a The reference category is: กลุ่มที่ไม่เคยดื่มสุรา.

b This parameter is set to zero because it is redundant.



Classification

Observed	Predicted			
	กลุ่มที่ดื่ม สม่ำเสมอ	กลุ่มที่ดื่มเล็กน้อย หรือนาน ๆ ครั้ง	กลุ่มที่ไม่เคยดื่ม สุรา	Percent Correct
กลุ่มที่ดื่มสม่ำเสมอ	29	25	32	33.7%
กลุ่มที่ดื่มเล็กน้อยหรือ นาน ๆ ครั้ง	9	15	49	20.5%
กลุ่มที่ไม่เคยดื่มสุรา	1	0	445	99.8%
Overall Percentage	6.4%	6.6%	86.9%	80.8%



ประวัติย่อผู้วิจัย



ประวัติย่อผู้วิจัย

ชื่อ นามสกุล	นายจักรพงษ์ มานะดี
วัน เดือน ปีเกิด	วันที่ 12 มีนาคม พ.ศ. 2522
สถานที่เกิด	จังหวัดหนองคาย ประเทศไทย
ประวัติการศึกษา	พ.ศ. 2540 ระดับมัธยมศึกษาปีที่ 6 โรงเรียนชุมพลโพธิ์พิสัย อำเภอโพธิ์พิสัย จังหวัดหนองคาย พ.ศ. 2544 ปริญญาวิทยาศาสตรบัณฑิต (วท.บ.) สาขาวิชาสถิติประยุกต์ สถาบันราชภัฏอุดรธานี พ.ศ. 2557 ปริญญาวิทยาศาสตรมหาบัณฑิต (วท.ม.) สาขาวิชาวิทยาการจัดการสถิติ มหาวิทยาลัยมหาสารคาม
ตำแหน่ง สถานที่ทำงาน	นักวิชาการสถิติชำนาญการ สำนักงานสถิติจังหวัดหนองคาย ศาลากลางจังหวัด จังหวัดหนองคาย 43000
ที่อยู่ที่สามารถติดต่อได้	บ้านเลขที่ 258 หมู่ที่ 15 บ้านโพธิ์เงิน ตำบลชุมช้าง อำเภอโพธิ์พิสัย จังหวัดหนองคาย 43120

