

การวิเคราะห์การดื่มสุราของประชาชนในจังหวัดหนองคาย โดยเทคนิคเหมืองข้อมูลและการถดถอยโลจิสติก

The Analysis of Alcoholic Consumption for the People in Nong Khai Province by Data Mining Techniques and Logistic Regression Analysis

จักรพงษ์ มานะดี¹, มนชยา เจียงประดิษฐ์², นิภาพร ชุติมันต์³

Jakrapong Manadee¹, Monchaya Chiangpradit², Nipapron Chutiman³

Received: 12 January 2014 ; Accepted: 31 March 2014

บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์ เพื่อสร้างรูปแบบที่เหมาะสมในการพยากรณ์กลุ่มของพฤติกรรมของการดื่มสุราของประชากรในจังหวัดหนองคายและเปรียบเทียบประสิทธิภาพการพยากรณ์ของเทคนิคการทำเหมืองข้อมูลโดยแบบจำลองโครงข่ายประสาทเทียมและเทคนิคการวิเคราะห์การถดถอยโลจิสติกแบบหลายกลุ่ม จากข้อมูลการสำรวจพฤติกรรมการสูบบุหรี่และดื่มสุราของประชากร พ.ศ. 2554 ซึ่งดำเนินการสำรวจโดยสำนักงานสถิติแห่งชาติ จำนวน 605 ราย โดยใช้ร้อยละของความถูกต้องของการ เป็นเกณฑ์ในการวัดความแม่นยำของการพยากรณ์ ผลการศึกษาพบว่ากลุ่มการดื่มสุรา 3 กลุ่ม คือ กลุ่มไม่เคยดื่มสุรา กลุ่มดื่มเล็กน้อย หรือนาน ๆ ครั้ง และกลุ่มดื่มสม่ำเสมอ โดยแบบจำลองโครงข่ายประสาทเทียม มีความแม่นยำ ในการพยากรณ์มากกว่าคิดเป็น ร้อยละ 97.79

คำสำคัญ: การดื่มสุรา การทำเหมืองข้อมูล การถดถอยโลจิสติก

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 by artificial neural networks (ANN) and multinomial logistic regression. The secondary data of the survey were forecasting with 97.79% of correction.

Keywords: Drinking Alcohol; Data Mining; Logistic Regression

บทนำ

ปัจจุบันการดื่มสุราหรือเครื่องดื่มที่มีแอลกอฮอล์เป็นเรื่องปกติของคนในสังคม โดยประเทศไทยมีแนวโน้มการบริโภคเครื่องดื่มแอลกอฮอล์เพิ่มมากขึ้นโดย จากข้อมูลองค์การอนามัยโลก พบว่า ปัจจุบันมีประชากรทั่วโลกดื่มเครื่องดื่มแอลกอฮอล์ ประมาณ 2,000 ล้านคน หรือ 1 ใน 3 ของประชากรโลก สำหรับคนไทย พบว่า ดื่มเครื่องดื่มแอลกอฮอล์สูงเป็นอันดับที่ 40 ของโลก และอันดับ 3 ของเอเชีย โดยผู้ที่มีอายุ 15 ปี

ขึ้นไป ดื่มเครื่องดื่มแอลกอฮอล์ 16.2 ล้านคน หรือเฉลี่ยดื่มคนละ 58 ลิตรต่อปี สูงกว่าค่าเฉลี่ยโลก 9 เท่าตัว¹ และข้อมูลจากสำนักงานสถิติแห่งชาติ ระบุว่า ประชากรอายุ 15 ปีขึ้นไป จำนวน 53.9 ล้านคน เป็นผู้ดื่มสุรา 17.0 ล้านคน (ร้อยละ 31.5) โดยผู้ชายมีอัตราการดื่มสูงกว่าผู้หญิงประมาณ 5 เท่า และกลุ่มวัยทำงาน (25-59 ปี) มีอัตราการดื่มสูงกว่ากลุ่มอื่น คือ ร้อยละ 37.3 และมีข้อสังเกตว่ากลุ่มผู้สูงอายุของไทยยังมีการดื่มสุราสูงถึงร้อยละ 16.6 สำหรับอายุเฉลี่ยที่เริ่มดื่มสุราโดย

¹ นิสิตระดับปริญญาโท, ² อาจารย์, ³ ผู้ช่วยศาสตราจารย์ คณะวิทยาศาสตร์ มหาวิทยาลัยมหาสารคาม, อำเภอกันทรวิชัย จังหวัดมหาสารคาม 44150,

¹ Master student, ² Lecturer, ³ Assist.Prof., Faculty of Science, Mahasarakham University, Kantharawichai District, Maha Sarakham 44150

รวมจะเริ่มดื่มที่อายุ 20.3 ปี ผู้ชายจะเริ่มดื่มเร็วกว่าผู้หญิง คือ อายุ 19.4 ปี และ 24.9 ปี โดยกลุ่มเยาวชน (15-24 ปี) มีอายุเฉลี่ยที่เริ่มดื่มสุรา 17.2 ปี ในขณะที่ กลุ่มวัยทำงาน (25-59 ปี) มีอายุเฉลี่ยที่เริ่มดื่มสุรา 20.6 ปี² และจากเรื่องอุบัติเหตุที่เกิดขึ้นจากการดื่มเครื่องดื่มแอลกอฮอล์ ของเด็กและเยาวชน ในสถานศึกษาของภาคตะวันออกเฉียงเหนือ พบว่า มีถึงร้อยละ 24.5 โดยจังหวัดพบมากที่สุด คือ สกลนคร ร้อยละ 44.3 ขอนแก่น ร้อยละ 32.3 ยโสธร ร้อยละ 27.9 หนองคาย ร้อยละ 27.9 ชัยภูมิ ร้อยละ 26.0 ตามลำดับ เมื่อพิจารณาจากกลุ่มของการดื่มสุราโดยใช้ความถี่ในการดื่ม คือ ผู้ที่ดื่มสม่ำเสมอ กลุ่มที่ดื่มเล็กน้อยหรือนาน ๆ ครั้ง พบว่า กลุ่มของเล็กน้อยหรือนาน ๆ ครั้ง มีสัดส่วนสูงถึง ร้อยละ 55.8 ของผู้ดื่มทั้งหมด³

สำหรับจังหวัดหนองคายการบริโภคสุราหรือเครื่องดื่มที่มีแอลกอฮอล์ของประชาชนเป็นไปในทิศทางเดียวกันที่กล่าวข้างต้นและมีปริมาณเพิ่มมากขึ้น จากรายงานสถานการณ์การบริโภคเครื่องดื่มแอลกอฮอล์รายจังหวัด พ.ศ. 2550 พบว่า จังหวัดหนองคาย ความชุกของผู้ดื่มเครื่องดื่มแอลกอฮอล์ อายุ 20 ปีขึ้นไป มีร้อยละ 37.4 ความชุกของผู้ดื่มเครื่องดื่มแอลกอฮอล์ อายุ 15-19 ปี มีร้อยละ 15.9 เครื่องดื่มที่นิยมดื่ม 5 อันดับแรก คือ สุราขาวหรือสุรากลั่นชุมชน ร้อยละ 68.83 เบียร์ ร้อยละ 38.14 สุรายี่ห้อไทย ร้อยละ 11.26 สุรายี่ห้อต่างประเทศ ร้อยละ 1.72 ยาแดงเหล้า ร้อยละ 0.77 ตามลำดับ และค่าใช้จ่าย ในการซื้อเครื่องดื่มแอลกอฮอล์โดยเฉลี่ยอยู่ที่ 236 บาทต่อเดือน⁴ 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 ล้านบาท⁵ นอกจากนี้สำนักงานสาธารณสุขจังหวัดหนองคาย ได้ทำการศึกษากับข้อมูลเกี่ยวกับสถานการณ์และพฤติกรรมการบริโภคเครื่องดื่มแอลกอฮอล์ของประชาชนจังหวัดหนองคาย พบว่า ประชาชนอายุ 15 -74 ปีขึ้นไป ร้อยละ 18.3 (104,130 คน) ดื่มสุราหรือเครื่องดื่มมีแอลกอฮอล์เป็นประจำ และร้อยละ 4.97 (29,387 คน) ดื่มนาน ๆ ครั้ง ส่วนประเภทของเครื่องดื่มแอลกอฮอล์ที่ประชาชนนิยมดื่มมากที่สุด ได้แก่ เบียร์และสุราขาว คิดเป็นร้อยละ 33.0 และ 32.3 ตามลำดับ ขณะที่ผลสำรวจจากนักเรียนมัธยมศึกษาในหนองคาย พบว่า นักเรียนชายและหญิงที่ดื่มสุราซึ่งมีอายุ 14-16 ปี และระบุว่าเริ่มดื่มตั้งแต่อายุระหว่าง 13-14ปี ด้วยอัตราส่วนสูงที่สุดประมาณร้อยละ 44 และ 46 ตามลำดับ ส่วนนักเรียนชายและหญิงที่อายุ 17-20 ปี ระบุว่าเริ่มดื่มเมื่ออายุ 15-16 ปี สูงสุดด้วย

อัตราใกล้เคียงกันคือประมาณร้อยละ 43 และ 40 ตามลำดับ จึงมีแนวโน้มว่านักเรียนเริ่มหันมาดื่มเครื่องดื่มแอลกอฮอล์เมื่ออายุน้อยลงเรื่อยๆ⁶

จากสถานการณ์วิจัยในปัจจุบันที่เกี่ยวข้องกับการดื่มสุราส่วนใหญ่เป็นงานวิจัยในรูปแบบการหาปัจจัยที่ส่งผลต่อการดื่มสุราและผลกระทบในด้านต่างๆ ที่เกิดจากการดื่มสุรา โดยยังขาดงานวิจัยที่ใช้ในการพยากรณ์เกี่ยวกับการดื่มสุรา^{7,8} ซึ่งเทคนิคหรือวิธีการในการพยากรณ์เมื่อตัวแปรเป็นเชิงกลุ่มที่นิยมใช้ส่วนใหญ่ได้แก่เทคนิคโครงข่ายประสาทเทียม (Artificial Neural Networks : ANN)⁷ และเทคนิคการถดถอย โลจิสติก (Logistic Regression Analysis) และมีการนำไปประยุกต์ใช้ในการแก้ไขปัญหาต่างๆ มากมาย เช่น ด้านการสื่อสารใช้กรองอีเมลล์ (E-mail) ขยะ ด้านการแพทย์ใช้ศึกษาปัจจัยที่มีผลต่อการทำแท้ง⁸ เป็นต้น

ดังนั้นการวิจัยในครั้งนี้จึงมีวัตถุประสงค์ที่จะหาเทคนิคที่เหมาะสมในการวิเคราะห์กลุ่มการดื่มสุราของประชาชนในจังหวัดหนองคาย เพื่อเป็นข้อมูลประกอบการตัดสินใจในการกำหนดนโยบาย การกำกับดูแลและควบคุมการวางแผน และจัดกิจกรรมต่างๆ ของหน่วยงานที่เกี่ยวข้องในการลดการดื่มสุราและผลเกี่ยวเนื่องอื่นๆ ที่จะตามมาจากการดื่มสุราของประชาชนในจังหวัดหนองคาย

วิธีการดำเนินการวิจัย

1. ข้อมูลที่ใช้ในการวิจัย

ข้อมูลสำหรับการศึกษาวิจัยในครั้งนี้ได้มาจากการสำรวจพฤติกรรมการสูบบุหรี่และดื่มสุราของประชากร พ.ศ. 2554 สำรวจโดยสำนักงานสถิติแห่งชาติ โดยเป็นบุคคลที่มีอายุ 15 ปีขึ้นไป จำนวน 605 ราย จากการเก็บรวบรวมข้อมูลโดยเจ้าหน้าที่ของสำนักงานสถิติจังหวัดหนองคาย ระหว่างวันที่ 1-12 ของเดือน มีนาคม ถึงพฤษภาคม พ.ศ. 2554²

2. ตัวแปรที่ใช้ในการศึกษา

สำหรับตัวแปรที่ศึกษาในการวิจัยครั้งนี้มีจำนวนทั้งสิ้น 11 ตัวแปร จำแนกเป็นตัวแปรอิสระ 10 ตัว ได้แก่ 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 ตามเกณฑ์การจำแนกกลุ่มของศูนย์วิจัยปัญหาการดื่มสุรา

3. การวิเคราะห์ข้อมูล

วิธีการที่เลือกใช้ในงานวิจัยครั้งนี้คือ การทำเหมืองข้อมูลและการวิเคราะห์การถดถอยโลจิสติกแบบหลายกลุ่ม ซึ่งมีรายละเอียดดังนี้

3.1 เทคนิคการทำเหมืองข้อมูล (Data Mining) โดยใช้แบบจำลองโครงข่ายประสาทเทียม (Artificial Neuron Network: ANN)

ขั้นที่ 1

การพิจารณาการดื่มสุราของประชาชนในจังหวัดหนองคาย จากข้อมูลจำนวน 605 ราย ได้แบ่งข้อมูลออกเป็น 2 ส่วน คือ ชุดการเรียนรู้และชุดทดสอบ ในอัตราส่วนที่มีการปรับเปลี่ยนไป 2 ชุด คือ 60:40 และ 90:10 นอกจากนี้ยังได้ทำการวิเคราะห์ข้อมูลด้วยการแบ่งด้วย K-Fold Validation ซึ่งทำซ้ำ 10 รอบ วัดด้วยค่าสัมบูรณ์ความคลาดเคลื่อนเฉลี่ย (Mean Absolute Percentage Error: MAPE) ที่ต่ำสุด (Table 1) ได้ชุดข้อมูลที่ดีที่สุดคือ 90:10 หรือจำนวน 545 รายและข้อมูลที่ใช้ในการทดสอบโครงข่ายประสาทจะเลือกข้อมูลร้อยละ 10 หรือจำนวน 60 ราย

Table 1 The efficiency of data

data (training:testing)	MAPE
60:40	0.0905
90:10	0.0854
K-Fold Validation	0.0918

ขั้นที่ 2

การคัดเลือกตัวแปรเพื่อสร้างชุดข้อมูลของตัวแปรอิสระ ดำเนินการโดยวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรอิสระทั้ง 10 ตัวแปรกับตัวแปรตาม คือ กลุ่มการดื่มสุรา โดยการกำหนดชุดข้อมูลนำเข้าของปัจจัยต่าง ๆ ที่มีอิทธิพลต่อการพยากรณ์ของโครงข่ายประสาทเทียม โดยกำหนดชุดข้อมูลของตัวแปรอิสระออกเป็น 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

โดยทำการหาความสัมพันธ์ระหว่างตัวแปรอิสระกับตัวแปรตาม 2 วิธี ดังนี้

1. สัมประสิทธิ์ แบบ Cramer's V ทำการหาความสัมพันธ์ระหว่างตัวแปรอิสระ คือ เพศ ระดับการศึกษาที่จบ

สูงสุด สถานภาพสมรส อาชีพ สถานภาพการทำงาน กับตัวแปรตาม คือ กลุ่มการดื่มสุรา โดยใช้สูตร

$$v = \sqrt{\frac{\chi^2}{n(q-1)}} \tag{1}$$

เมื่อ q จำนวนแถวนอน (r) หรือแถวตั้ง (c) ที่มีค่าน้อยที่สุด

n จำนวนตัวอย่างทั้งหมด

โดยค่าของ V จะมีค่าระหว่าง 0 ถึง 1

2. สัมประสิทธิ์ แบบ Eta ทำการหาความสัมพันธ์

ระหว่างตัวแปรอิสระ คือรายได้จากการทำงาน ปริมาณการดื่มสุรา คะแนนผลกระทบปัญหาสุขภาพที่มีผลต่อการทำงาน คะแนนผลกระทบปัญหาสุขภาพที่มีผลต่อการทำงานประจำวัน กับ ตัวแปรตาม คือ กลุ่มการดื่มสุรา โดยใช้สูตร

$$\text{Eta} = 1 - \frac{\sum_{i=1}^n (X_i - \bar{X}_k)^2}{\sum_{i=1}^n (X_i - \bar{X}_T)^2} \tag{2}$$

เมื่อ $(X_i - \bar{X}_k)^2$ คะแนนเบี่ยงเบนจากค่าเฉลี่ยในแต่ละกลุ่ม $(X_i - \bar{X}_T)^2$ คะแนนเบี่ยงเบนจากค่าเฉลี่ยรวม

ขั้นที่ 3

สำหรับการออกแบบสถาปัตยกรรมโครงข่ายประสาทเทียม ในงานวิจัยครั้งนี้ ผู้วิจัยได้ทำการสร้างและทดสอบตัวแบบการพยากรณ์ โดยใช้โปรแกรม Weka 3.6.6 ซึ่งมีการกำหนดค่าพารามิเตอร์ต่างๆ คือ จำนวนโหนดในชั้นซ่อน 9 ค่าอัตราการเรียนรู้ 0.2 ค่าโมเมนตัม 0.3 จำนวนรอบในการเรียนรู้ 50,000 รอบ^{10,11,12} โดยเป็นโครงข่ายประสาทเทียมแบบชั้นเดียว มีข้อมูลนำเข้า 545 ราย ส่งผ่านด้วยฟังก์ชันถ่ายโอนและส่งถ่ายข้อมูลด้วยกระตุ้นแบบผลรวมเชิงเส้น สู่ชั้นผลลัพธ์คือกลุ่มการดื่มสุรา โดยสถาปัตยกรรมชุดข้อมูลของตัวแปรอิสระ 3 ชุด มีดังนี้

ตัวแบบที่ 1 (ANN1) ประกอบด้วยตัวแปรอิสระทั้งหมด 10 ตัว ใช้ตัวแบบ คือ 10:9:1 ชั้นนำเข้า 10 โหนด ชั้นซ่อน 9 โหนด และชั้นแสดงผล 1 โหนด

ตัวแบบที่ 2 (ANN2) ประกอบด้วยตัวแปรอิสระ มีค่าความสัมพันธ์ตั้งแต่ 0.50 ขึ้นไป ใช้ตัวแบบ คือ 3:9:1 ชั้นนำเข้า 3 โหนด ชั้นซ่อน 9 โหนด และชั้นแสดงผล 1 โหนด

ตัวแบบที่ 3 (ANN3) ประกอบด้วยตัวแปรอิสระ มีค่าความสัมพันธ์ต่ำกว่า 0.50 ใช้ตัวแบบ คือ 7:9:1 ชั้นนำเข้า 7 โหนด ชั้นซ่อน 9 โหนด และชั้นแสดงผล 1 โหนด โดยทั้ง 3 ตัวแบบในชั้นซ่อนใช้ฟังก์ชันการแปลงค่าคือ ฟังก์ชันซิกมอยด์ $f = \frac{1}{1 + e^{-x}}$ และชั้นแสดงผลใช้ฟังก์ชันกระตุ้นแบบผลรวมเชิงเส้น

ขั้นที่ 4

การตรวจสอบประสิทธิภาพของโครงข่ายวัดจากความคลาดเคลื่อนระหว่างผลการทำนายและข้อมูลจริงจากชุดทดสอบ ดังนี้คือ

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

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Predict_i - Actual_i)^2} \quad (3)$$

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

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{Predict_i - Actual_i}{Actual_i} \right| \times 100 \quad (4)$$

โดยที่

Actual แทนค่าข้อมูลจริง

Predict แทนค่าข้อมูลที่ได้จากการพยากรณ์

n แทนจำนวนชุดข้อมูลที่ใช้ในการคำนวณ

3.2 เทคนิคการวิเคราะห์การถดถอยโลจิสติกแบบหลายกลุ่ม (Multinomial Logistic Regression Analysis)

ในการวิเคราะห์การถดถอยโลจิสติกเป็นวิธีการที่นิยมใช้อย่างแพร่หลาย¹³ ซึ่งเป็นวิธีการที่ใช้วิเคราะห์ในกรณีที่มีตัวแปรตามเป็นตัวแปรเชิงกลุ่มและความน่าจะเป็นที่จะเกิดค่าในตัวแปรตามถูกพยากรณ์ด้วยตัวแปรอิสระ ดังสมการ

$$P(Y) = \frac{1}{1 + e^{-(b_0 + b_1x_1 + b_2x_2 + \dots + b_jx_k)}} \quad (5)$$

เมื่อ P(Y) แทน ความน่าจะเป็นที่จะเกิดค่าในตัวแปรตาม
 x_i แทน ตัวแปรอิสระ; $i=1,2,3,\dots,k$

b_j แทน สัมประสิทธิ์การถดถอย; $j=0,1,2,\dots,k$

โดยที่ค่าสัมประสิทธิ์การถดถอย จะทำการประมาณค่าด้วยวิธีการความน่าจะเป็นสูงสุด¹⁴ ตัวแปรในการวิเคราะห์ใช้ตัวแปรอิสระทั้งหมด 10 ตัว คือ เพศ อายุ ระดับการศึกษาที่จบสูงสุด สถานภาพสมรส อาชีพ สถานภาพการทำงาน

รายได้จากการทำงาน ปริมาณการดื่มสุรา คะแนนผลกระทบปัญหาสุขภาพที่มีผลต่อการทำงานและคะแนนผลกระทบปัญหาสุขภาพที่มีผลต่อการทำงานประจำวัน และตัวแปรตาม คือ กลุ่มของการดื่มสุรา

4. ประสิทธิภาพของตัวแบบ

สำหรับการเปรียบเทียบประสิทธิภาพตัวแบบการวิเคราะห์การดื่มสุราของประชาชนในจังหวัดหนองคาย โดยเทคนิคการทำเหมืองข้อมูลและการถดถอยโลจิสติกนั้น ใช้ร้อยละของความถูกต้องในการพยากรณ์เป็นตัวเปรียบเทียบ ดังนี้คือ

$$Correction = \frac{\text{จำนวนข้อมูลที่จำแนกได้}}{\text{จำนวนข้อมูลทั้งหมด}} \times 100 \quad (6)$$

ผลการวิเคราะห์ข้อมูล

1. ผลการทดสอบแบบจำลองการทำเหมืองข้อมูลการวิเคราะห์หาความสัมพันธ์ระหว่างตัวแปรอิสระ และตัวแปรตาม (Table 2) เพื่อแบ่งตัวแบบได้ตัวแบบ ดังนี้

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

Table 2 The correlation coefficient between dependent variable and independent variables

Independent variables	Correlation
Sex*	0.55
Age**	0.16
Highest education*	0.16
Marital status*	0.07
Occupation*	0.20
Working*	0.20
Income**	0.58
Amount of alcohol drinking**	0.96
Health score**	0.20
Daily score**	0.16

* by Cramer's V coefficient

** by Eta coefficient

การทำงานและคะแนนผลกระทบปัญหาสุขภาพที่มีผลต่อการทำกิจวัตรประจำวัน

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

จากนั้นดำเนินการหาตัวแบบโดยนำค่าเหล่านี้ไปใช้ในการพยากรณ์หาตัวแบบด้วยวิธีของโครงข่ายประสาทเทียมเปอร์เซ็ปตรอน แบบหลายชั้น (Multilayer Perceptron: MLP) และแสดงผลการแบบจำลองโดยใช้โครงข่ายประสาทเทียมทดสอบตัวแบบใน Table 3

พบว่า ตัวแบบ ANN1 มีประสิทธิภาพในการพยากรณ์สูงสุด โดยพิจารณาค่าความคลาดเคลื่อนของการทดสอบในชุดการเรียนรู้ MAPE เท่ากับ 0.0199 RMSE เท่ากับ 0.1012 เมื่อลดจำนวนตัวแปรอิสระลงโดยแบ่งตามความสัมพันธ์ ในตัวแบบ ANN2 และ ANN3 พบว่า ค่าความคลาดเคลื่อน MAPE และ RMSE มีค่าสูงกว่าตัวแบบ ANN1 และเมื่อนำไปหาค่าความถูกต้องของการพยากรณ์ พบว่า มีความถูกต้องของการพยากรณ์ ร้อยละ 97.79 ได้ผลดัง Table 4

Table 3 The model efficiency of training and testing data set

Model	Training		Testing	
	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

Table 4 The ANN1's efficiency

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

2. ผลการทดสอบแบบจำลองการถดถอยโลจิสติกเมื่อนำข้อมูลมาวิเคราะห์โดยใช้การถดถอยโลจิสติกแบบหลายกลุ่มจาก Table 5 พบว่า เมื่อทดสอบโดยใช้สถิติทดสอบ Chi - square พบว่า มีตัวแปรอย่างน้อย 1 ตัว ที่ส่งผลกระทบต่อพยากรณ์กลุ่มการดื่มสุราของประชาชนในจังหวัดหนองคาย (P-value<0.000) และสามารถสร้างสมการในการพยากรณ์กลุ่มในการดื่มสุรา ได้ดังนี้คือ

$$G(1) = -5.143 + 0.061(\text{quantity}) + 2.564(\text{sex})+1.125(\text{age}=2)$$

$$G(2) = -4.348 + 0.063(\text{quantity}) + 1.813(\text{sex}) + 1.043(\text{age}=2)$$

โดยที่

G(1) คือ สมการพยากรณ์กลุ่มการดื่มสุร่าเสมอเทียบกับกลุ่มไม่เคยดื่มสุรา

G(2) คือ สมการพยากรณ์กลุ่มการดื่มสุร่าเล็กน้อยหรือนานๆครั้งเทียบกับกลุ่มไม่เคยดื่มสุรา

quantity คือ ปริมาณการดื่มสุรา

sex คือ เพศ

age = 2 คือ อายุ 25 – 59 ปี

จากสมการดังกล่าว พบว่า มีความถูกต้องของการพยากรณ์ ร้อยละ 80.8 ดังแสดงใน Table 6

4. การเปรียบเทียบประสิทธิภาพของตัวแบบ

การศึกษาค้นคว้าครั้งนี้ผู้วิจัยทำการเปรียบเทียบผลการพยากรณ์จากแบบจำลองทั้งสองวิธีโดยการวิเคราะห์ด้วยค่าร้อยละของความถูกต้องในการพยากรณ์ จากตาราง 4 และ ตาราง 6 พบว่า การสร้างแบบจำลองโดยเทคนิคการทำเหมืองข้อมูลด้วยวิธีโครงข่ายประสาทเทียม ANN1 มีความถูกต้องมากกว่า

วิจารณ์ผลการวิจัย

จากการวิเคราะห์การดื่มสุราของประชาชนในจังหวัดหนองคายโดยเทคนิคเหมืองข้อมูลและการถดถอยโลจิสติก พบว่าแบบจำลองโดยเทคนิคการทำเหมืองข้อมูลด้วยวิธีโครงข่ายประสาทเทียม มีความถูกต้องในการพยากรณ์มากที่สุด ด้วยแบบจำนวน ANN1 ใช้ ตัวแบบ คือ 10:9:1 เนื่องจากข้อมูลที่นำมาใช้ในการพยากรณ์ส่วนมากเป็นข้อมูลในเชิงคุณภาพและเป็นข้อมูลที่เป็นค่อนข้างเป็นแบบแผน ซึ่งเหมาะสมกับเทคนิคการทำเหมืองข้อมูลอีกทั้งมีจำนวนรอบในการวิเคราะห์ที่สูงจึงทำให้มีความถูกต้องมากขึ้น โดยมีความถูกต้องของการพยากรณ์กลุ่มการดื่มสุรา ร้อยละ 97.79 สอดคล้องกับการศึกษาของภรณ์ยา อ่ามฤครัตน์ และพยุง มีสีจ¹⁵ ได้ทำการศึกษาการเปรียบเทียบประสิทธิภาพการจำแนก

กลุ่มข้อมูลของโรคลมร้อนด้วยวิธีการทางเครือข่ายประสาทเทียม โดยได้แบบจำลองที่ให้ค่าพยากรณ์ที่มีความถูกต้องสูงที่สุด ร้อยละ 97.67 และสอดคล้องกับ สักการะ จูชู่และชนพล เจนสุทธิเวชกุล¹⁶ ที่ทำการศึกษเปรียบเทียบการพยากรณ์ปริมาณการยืมหนังสือห้องสมุดโรงเรียนนาบอน ด้วยการวิเคราะห์การถดถอยและโครงข่ายประสาทเทียม แบบจำลองโครงข่ายประสาทเทียมมีค่าพยากรณ์ใกล้เคียงกับค่าจริงมากกว่าแบบจำลองการถดถอย และสอดคล้องกับการศึกษาของ Mukta Paliwal and Usha A.Kumar¹⁷ ที่ได้ศึกษาเปรียบเทียบจากจำนวนงานวิจัยทั้งหมด 96 งานวิจัย โครงข่ายประสาทเทียมให้ผลการพยากรณ์ที่ดีกว่าจำนวน 58 เปอร์เซ็นต์ โครงข่ายประสาทเทียมกับเทคนิคการวิเคราะห์ผลทางสถิติโดยทั่วไป ให้ผลการพยากรณ์เหมือนกันจำนวน 24 เปอร์เซ็นต์ และเทคนิคการวิเคราะห์ผลทางสถิติโดยทั่วไปให้ผลการพยากรณ์ที่ดีกว่าจำนวน 18 เปอร์เซ็นต์

ส่วนตัวแบบการถดถอยโลจิสติกแบบหลายกลุ่มนั้น ให้ผลประสิทธิภาพน้อยกว่าในแบบจำลองที่ทำการศึกษ เนื่องจากข้อมูลที่นำมาใช้ในการวิเคราะห์ส่วนใหญ่เป็นข้อมูลเชิงคุณภาพและเมื่อนำมาหาความสัมพันธ์พบว่า มีความสัมพันธ์ระหว่างตัวแปรอิสระและตัวแปรตามมีค่าค่อนข้างต่ำ จึงทำให้แบบจำลองที่ได้มีประสิทธิภาพไม่ดีเท่าที่ควร ประกอบกับข้อมูลที่นำมาวิเคราะห์ในกลุ่มที่ตีพิมพ์น้อยหรือนานๆครั้ง และกลุ่มที่ตีพิมพ์สม่ำเสมออาจมีจำนวนน้อยเกินไป จึงทำให้การวิเคราะห์อาจไม่ดีเท่าที่ควร

ข้อเสนอแนะ

จากผลการทดสอบแบบจำลองโดยเทคนิคการทำเหมืองข้อมูลด้วยวิธีโครงข่ายประสาทเทียม ANN1 (10:9:1) มีความถูกต้องในการพยากรณ์มากกว่าส่วนตัวแบบการถดถอยโลจิสติกแบบหลายกลุ่ม โดยตัวแปรอิสระทั้ง 10 ตัวแปร สามารถนำไปใช้ในการพยากรณ์กลุ่ม การตีพิมพ์ของประชาชนในจังหวัดหนองคายได้ โดยในการศึกษาครั้งต่อไป ควรนำตัวแปร เพศ อายุ ปริมาณการตีพิมพ์ รายได้จากการทำงาน มาศึกษาต่อในเรื่องของการตีพิมพ์เนื่องจากเป็นตัวแปรที่มีความสัมพันธ์ค่อนข้างสูง อาจส่งผลต่อกลุ่มการตีพิมพ์ได้ ดังนั้น จึงควรมีนโยบายรณรงค์ ประชาสัมพันธ์ให้ประชาชนได้ทราบถึงโทษของการตีพิมพ์ รวมทั้งหามาตรการในการป้องกัน สอดส่องดูแล ควบคุม และผลกระทบที่อาจเกิดขึ้นจากการตีพิมพ์ของนักตีพิมพ์หน้าใหม่ และบังคับใช้กฎหมายเพื่อควบคุม ตรวจสอบการผลิตหรือจำหน่ายสุราของผู้ประกอบการให้เป็นไปตามกฎหมายอย่างเคร่งครัด

เอกสารอ้างอิง

1. สำนักงานคณะกรรมการพัฒนาการเศรษฐกิจและสังคมแห่งชาติ. สศช.แถลงสภาพสังคมไทย สถิติเด็ก “ตีพิมพ์สุราหนัก-ติดเฟสบุ๊ก-คุณแม่วัยใสฟุ้ง-เสพยาไอซ์กระจาย”. [ออนไลน์]. [สืบค้นเมื่อวันที่ 19 กันยายน 2555]; ได้จาก <http://thaipublica.org/- 2012/02/nesdb-announces-thai-society/>
2. สำนักงานสถิติแห่งชาติ. สรุปข้อมูลที่สำคัญการสำรวจพฤติกรรม การสูบบุหรี่และการดื่มสุราของประชากร พ.ศ. 2554. กรุงเทพฯ. เท็กซ์ แอน เจอร์นัล พับลิเคชั่น. 2555.
3. นิรุจน์ อุทธา และคณะ. อุบัติการณ์และแบบแผนการดื่มเครื่องดื่มแอลกอฮอล์ของเด็กเยาวชนในสถานศึกษา ภาคตะวันออกเฉียงเหนือ ภายหลังนโยบายคุ้มครองเด็ก. [ออนไลน์]. [สืบค้นเมื่อวันที่ 1 ตุลาคม 2556]; ได้จาก <http://www.jvkk.go.th/research/qrrsearch.asp?code=0101880>
4. ศูนย์วิจัยปัญหาสุรา. รายงานสถานการณ์การบริโภคเครื่องดื่มแอลกอฮอล์รายจังหวัด พ.ศ. 2550. นนทบุรี: เดอะ กราฟิโก ซิสเต็มส์, 2550. ผ-70.
5. สำนักงานสรรพสามิตจังหวัดหนองคาย. ข้อมูลการจัดเก็บเงินภาษีในพื้นที่จังหวัดหนองคาย. หนองคาย: สำนักงานสรรพสามิตจังหวัดหนองคาย. 2552.
6. สำนักงานสาธารณสุขจังหวัดหนองคาย. ข้อมูลเกี่ยวกับสถานการณ์และพฤติกรรม การบริโภคเครื่องดื่มแอลกอฮอล์ของประชาชนจังหวัดหนองคาย, หนองคาย: สำนักงานสาธารณสุขจังหวัดหนองคาย. 2553.
7. กฤษณ์ ชุนลิก. รูปแบบการส่งเสริม การปฏิบัติตามมาตรการควบคุมเครื่องดื่มแอลกอฮอล์ของผู้จำหน่ายในจังหวัดหนองคาย. [วิทยานิพนธ์ปริญญาปรัชญาดุษฎีบัณฑิต]. มหาสารคาม: มหาวิทยาลัยราชภัฏมหาสารคาม, 2555
8. พงษ์เดช สารการ และคณะ. พฤติกรรมการดื่มและผลกระทบจากการดื่มเครื่องดื่มแอลกอฮอล์ของกลุ่มแรงงานภาคอุตสาหกรรม จังหวัดนครราชสีมา. [ออนไลน์]. [สืบค้นเมื่อวันที่ 20 มกราคม 2556];. <http://cas.or.th/index.php/research/read/108>
9. Kanlapat Mahipan, Nipaporn Chutiman, Bungon Kumphon. A Forecasting Model for Thailand's Unemployment Rate. Modern Applied Science. 2013;(7):10-16.
10. อัญชลี วานิชทวีวัฒน์. การจดจำอักษรภาษาไทยตัวพิมพ์โดยโครงข่ายประสาทเทียมแบบเซลฟือร์แกนในซิงแมปซ์. [วิทยานิพนธ์ปริญญาวิศวกรรมศาสตรมหา

- บัณฑิต]. กรุงเทพฯ: สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง, 2540.
11. สุรยุทธ์ ปรินญา. การรู้จำอักษรไทยโดยโครงข่ายประสาทเทียมแบบแพร่กลับ. [วิทยานิพนธ์ปริญญาวิศวกรรมศาสตรมหาบัณฑิต]. เชียงใหม่: มหาวิทยาลัยเชียงใหม่, 2541.
 12. ทัดดาว แนบเนียน. การเปรียบเทียบการพยากรณ์ด้วยวิธีโครงข่ายประสาทเทียมและวิธีของบ็อกซ์และเจนนินส์ : กรณีศึกษา อัตราแลกเปลี่ยนเงินตราต่างประเทศ. [วิทยานิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต]. กรุงเทพฯ: มหาวิทยาลัยเกษตรศาสตร์, 2545.
 13. Bagley.s.c.,White, H., & Golomb, B. A.. Logistic regression in the medical literature: standards for use and reporting, with particular attention to one medical domain. *Journal of Clinical Epidemiology*. 2001; 54(10):979-985.
 14. Field, A. *discovering statistics using SPSS* (third ed.). London: Sage. 2009.
 15. ภรณ์ยา อ่ามฤครัตน์ และพยุ่ง มีสัจ. การเปรียบเทียบประสิทธิภาพการจำแนกกลุ่มข้อมูลของโรคลมร้อนด้วยวิธีการทางเครือข่ายประสาทเทียม.[ออนไลน์].[สืบค้นเมื่อวันที่ 1 ตุลาคม 2556]; ได้จาก http://202.44.34.144/nccitedoc/admin/nccit_files/NCCIT-2011040251.pdf
 16. สักการะ จุฑูและธนพล เจนสุทธิเวชกุล. เปรียบเทียบการพยากรณ์ปริมาณการยืมหนังสือห้องสมุดโรงเรียนนาบอนด้วยการวิเคราะห์การถดถอยและโครงข่ายประสาทเทียม. [ออนไลน์].[สืบค้นเมื่อ 1 ตุลาคม 2556] ; ได้จาก http://202.44.34.144/nccitedoc/admin/nccit_files/NC-CIT-20113103013754.pdf
 17. Mukta, Paliwal. & Kumar, Usha A. "Review Neural networks and statistical techniques: A review of applications. *Expert Systems with Applications*. 2009 ; 36, 2–17.

Table 5 Odds ratio from logistic regression

Variable	Odds ratio of Uniform drinking	Odds ratio of Little or infrequently drinking
Constant	-5.143*	-4.348*
Income	0.000	0.000
Amount of alcohol drinking	0.061*	0.063*
Health score	0.857	0.777
Daily score	1.604	3.908
Sex (base on female)	2.564*	1.813*
Age (base on ≥ 60 yrs)		
15 - 24 yrs	-0.555	-0.224
25 – 59 yrs	1.125*	1.043*
Education (base on Bachelor degree or higher)		
Primary school and lower	0.724	0.339
Secondary school	1.000	0.683
Cerificate	1.946	1.370
Marital status (base on married)		
Single/ widow/ divorce/ separate	0.179	-0.250
Occupation (base on no working)		
Agriculture/Fishing	0.398	0.808

Table 5 Odds ratio from logistic regression (Cont.)

Variable	Odds ratio of Uniform drinking	Odds ratio of Little or infrequently drinking
Techicial	0.000	0.661
Service and distributor	-0.204	0.653
Government service	0.687	1.963
Working status (base on not involve in business)		
Employer	-0.485	-0.641
Unpaid household	-0.228	-0.644
Government and state enterprise employees	0.598	-0.386
Private employee	-	-

-2 Log likelihood =620.033, $c^2=290.709$, $df=36$, p - value < 0.000, $N=605$, $R^2=0.382$

Base on never drinking group

*significant level at 0.05

Table 6 The efficiency of logistic regression model

Observed	Predicted			Correction
	Uniform drinking	Little or infrequently drinking	Never drinking	
Uniform drinking	29	25	32	33.7%
Little or infrequently drinking	9	15	49	20.5%
Never drinking	1	-	445	99.8%
Correction	6.4%	6.6%	86.9%	80.8%