แบบจำลองเพื่อพยากรณ์การเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุ ด้วยเทคนิคเหมืองข้อมูล: กรณีศึกษาโรงพยาบาลนาหว้า จังหวัดนครพนม

วิทยานิพนธ์

ของ

ศรัณยู ชูทองรัตนะ

เสนอต่อมหาวิทยาลัยราชภัฏสกลนคร เพื่อเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการสารสนเทศและเทคโนโลยี
มีนาคม 2565

ลิขสิทธิ์เป็นของมหาวิทยาลัยราชภัฏสกลนคร

แบบจำลองเพื่อพยากรณ์การเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุ ด้วยเทคนิคเหมืองข้อมูล: กรณีศึกษาโรงพยาบาลนาหว้า จังหวัดนครพนม

วิทยานิพนธ์

ของ

ศรัณยู ชูทองรัตนะ

เสนอต่อมหาวิทยาลัยราชภัฏสกลนคร เพื่อเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการสารสนเทศและเทคโนโลยี
มีนาคม 2565

ลิขสิทธิ์เป็นของมหาวิทยาลัยราชภัฏสกลนคร

PREDICTION MODELS FOR CHRONIC NON-COMMUNICABLE DISEASES IN THE ELDERLY USING DATA MINING TECHNIQUES: A CASE STUDY OF NAWA HOSPITAL, NAKHON PHANOM PROVINCE

BY

SARANYU CHUTHONGRATTANA

A Thesis Submitted in Partial Fulfillment of Requirements for

The Master of Education Degree in Information Science and Technology

Sakon Nakhon Rajabhat University

March 2022

All Rights Reserved by Sakon Nakhon Rajabhat University

ใบรับรองวิทยานิพนธ์

มหาวิทยาลัยราชภัฏสกลนคร

ปริญญาวิทยาศาสตรมหาบัณฑิต

สาขาวิชาวิทยาการสารสนเทศและเทคโนโลยี

ชื่อเรื่องวิทยานิพนธ์ แบบจำลองเพื่อพยากรณ์การเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุด้วยเทคนิคเหมืองข้อมูล: กรณีศึกษาโรงพยาบาลนาหว้า จังหวัดนครพนม

ชื่อผู้ทำวิทยานิพนธ์ ศรัณยู ชูทองรัตนะ

คณะกรรมการสอบวิทยานิพนธ์

ศรัญญา แซ่ตั้ง ..... ประธานกรรมการสอบ
(ผู้ช่วยศาสตราจารย์ ดร.ศรัญญา แซ่ตั้ง) และผู้ทรงคุณวุฒิ

สุทิศา ..... กรรมการสอบและประธานที่ปรึกษาวิทยานิพนธ์
(ดร.สุทิศา ซองเหล็กนอก)

..... กรรมการสอบ
(ผู้ช่วยศาสตราจารย์ ดร.สมบูรณ์ ชาวชายโขง) แต่งตั้งเพิ่มเติม

สุธาสินี คุปตะบุตร ..... กรรมการสอบและกรรมการที่ปรึกษาวิทยานิพนธ์
(ดร.สุธาสินี คุปตะบุตร)

..... กรรมการสอบ
(รองศาสตราจารย์ ดร.จำนง วงษ์ชาชม) แต่งตั้งเพิ่มเติม

คณะกรรมการบริหารหลักสูตรรับรองแล้ว

.....
(ผู้ช่วยศาสตราจารย์ ดร.สมบูรณ์ ชาวชายโขง)
ประธานหลักสูตร
มหาวิทยาลัยราชภัฏสกลนคร

บัณฑิตวิทยาลัยรับรองแล้ว

.....
(รองศาสตราจารย์ ดร.ศิกานต์ เพียรธัญญกรณ์)
ผู้อำนวยการบัณฑิตวิทยาลัย
มหาวิทยาลัยราชภัฏสกลนคร

เมื่อวันที่ 5 เดือน มีนาคม พ.ศ. 2565

ลิขสิทธิ์เป็นของมหาวิทยาลัยราชภัฏสกลนคร

ประกาศคุณูปการ

วิทยานิพนธ์ฉบับนี้สำเร็จได้ด้วยความกรุณาและความช่วยเหลือเป็นอย่างดียิ่งจาก ดร.สุทิศา ซองเหล็กนอก ประธานที่ปรึกษาวิทยานิพนธ์ และ ดร.สุธาสินี คุปตะบุตร กรรมการที่ปรึกษาวิทยานิพนธ์ ที่ให้ความกรุณาเอาใจใส่ ดูแล ให้คำปรึกษา คำแนะนำ ตลอดจนให้ความช่วยเหลือในการตรวจสอบ และปรับปรุงแก้ไขข้อบกพร่องต่าง ๆ ส่งผลให้วิทยานิพนธ์สำเร็จลุล่วงด้วยดี ผู้วิจัยรู้สึกซาบซึ้งในความเมตตา และขอขอบคุณมา ณ โอกาสนี้

ขอขอบคุณโรงพยาบาลนาหว้า จังหวัดนครพนม ที่ให้ข้อมูลในการทำวิจัยและขอขอบคุณนายแพทย์กิตติเชษฐ์ ธีรกุลพงศ์เวช ผู้อำนวยการโรงพยาบาลนาหว้า นายแพทย์ (ด้านเวชกรรมป้องกัน) สำนักงานสาธารณสุขจังหวัดนครพนม นางสาวกรรณิการ์ อุดมกัน นักวิชาการสาธารณสุขชำนาญการ นางอุไรลักษณ์ คำหา พยาบาลวิชาชีพชำนาญการ นางอรรชร สุนทรารักษ์ พยาบาลวิชาชีพชำนาญการ นางสาววิภาสินี วะเศษสร้อย พยาบาลวิชาชีพชำนาญการ นางพูล สุวัฒนเสนีย์ พยาบาลวิชาชีพชำนาญการ และ นางสาวทัศนีย์ สิทธิ พยาบาลวิชาชีพชำนาญการ ที่ให้ความอนุเคราะห์ประเมินประสิทธิภาพแบบจำลองในการวิจัยครั้งนี้ ตลอดจนให้คำแนะนำให้ข้อคิดเห็นที่เป็นประโยชน์ต่อผู้วิจัย

และที่สำคัญที่สุด ขอขอบคุณบิดา มารดา และสมาชิกในครอบครัวทุกคนที่ให้การสนับสนุนและช่วยเหลือในทุก ๆ ด้าน คอยดูแลและให้กำลังใจแก่ผู้วิจัยตลอดมา

คุณค่าและประโยชน์ของวิทยานิพนธ์ฉบับนี้ ขอมอบเป็นเครื่องบูชาพระคุณบิดา มารดา ครูอาจารย์ ที่ได้อบรมสั่งสอนและสนับสนุนการศึกษาของข้าพเจ้าตั้งแต่ต้น จนปัจจุบัน

ศรัณยู ชูทองรัตนะ

ชื่อเรื่อง แบบจำลองเพื่อพยากรณ์การเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุ ด้วยเทคนิคเหมืองข้อมูล: กรณีศึกษาโรงพยาบาลนาหว้า จังหวัดนครพนม

ผู้วิจัย ศรัณยู ชูทองรัตนะ

กรรมการที่ปรึกษา ดร.สุทิศา ซองเหล็กนอก

ดร.สุธาสินี คุปตะบุตร

ปริญญา วท.ม. (วิทยาการสารสนเทศและเทคโนโลยี)

สถาบัน มหาวิทยาลัยราชภัฏสกลนคร

ปีที่พิมพ์ 2565

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อสร้างและเปรียบเทียบแบบจำลองพยากรณ์การเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุ ได้แก่ โรคเบาหวาน โรคความดันโลหิตสูง และโรคหัวใจขาดเลือด ด้วยเทคนิคเหมืองข้อมูล ได้แก่ ต้นไม้ตัดสินใจ (Decision Tree) เพื่อนบ้านใกล้ที่สุด (K–Nearest Neighbors) และนาอีฟเบย์ (Naive Bayes) ข้อมูลที่ใช้ศึกษาคือข้อมูลผู้สูงอายุที่รับบริการจากโรงพยาบาลนาหว้า จังหวัดนครพนม ตั้งแต่เดือนตุลาคม พ.ศ. 2554 ถึง เดือนกันยายน พ.ศ. 2562 จำนวน 262,655 ระเบียน

ผลการวิจัยพบว่า (1) การสร้างแบบจำลองพยากรณ์การเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุ แบบจำลองพยากรณ์การเกิดโรคเบาหวานที่มีประสิทธิภาพสูงสุดคือแบบจำลองด้วยเทคนิคต้นไม้ตัดสินใจมีค่าความถูกต้อง 0.952 ค่าความแม่นยำ 0.952 ค่าความระลึก 0.952 แบบจำลองพยากรณ์การเกิดโรคความดันโลหิตสูงที่มีประสิทธิภาพสูงสุดคือแบบจำลองด้วยเทคนิคต้นไม้ตัดสินใจ และเพื่อนบ้านใกล้ที่สุดมีค่าความถูกต้อง 0.921 ค่าความแม่นยำ 0.922 ค่าความระลึก 0.922 และแบบจำลองพยากรณ์การเกิดโรคหัวใจขาดเลือดที่มีประสิทธิภาพสูงสุดคือแบบจำลองด้วยเทคนิคต้นไม้ตัดสินใจ และเพื่อนบ้านใกล้ที่สุดมีค่าความถูกต้อง 0.739 ค่าความแม่นยำ 0.728 ค่าความระลึก 0.739 การวิเคราะห์ปัจจัยที่มีผลต่อการเกิดโรคเบาหวานด้วยเทคนิคต้นไม้ตัดสินใจคือ น้ำตาลในเลือด ไขมันในเลือด และความดันโลหิตตามลำดับ ปัจจัยที่มีผลต่อการเกิดโรคความดัน

โลหิตสูงและโรคหัวใจขาดเลือดด้วยเทคนิคต้นไม้ตัดสินใจคือ ไขมันในเลือด น้ำตาลในเลือด และความดันโลหิตตามลำดับ (2) ผลการเปรียบเทียบแบบจำลองพยากรณ์การเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุได้แก่ โรคเบาหวาน โรคความดันโลหิตสูง และโรคหัวใจขาดเลือด เทคนิคต้นไม้ตัดสินใจมีประสิทธิภาพสูงสุด และเมื่อนำแบบจำลองพยากรณ์การเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุให้ผู้เชี่ยวชาญจำนวน 7 คน ประชุมกลุ่มเพื่อประเมินประสิทธิภาพของแบบจำลองกับการวินิจฉัยโรคมีค่าเฉลี่ยรวมอยู่ในระดับมาก ($\overline{X}$=4.43)

คำสำคัญ เหมืองข้อมูล ผู้สูงอายุ โรคไม่ติดต่อเรื้อรัง แบบจำลองเพื่อพยากรณ์

TITLE Prediction Models for Chronic Non–communicable Diseases in the Elderly Using Data Mining Techniques: A Case Study of Nawa Hospital, Nakhon Phanom Province

AUTHOR Saranyu Chuthongrattana

ADVISORS Dr. Sutisa Songleknok

Dr. Suthasinee Kuptabut

DEGREE M.Sc. (Information Science and Technology)

INSTITUTION Sakon Nakhon Rajabhat University

YEAR 2022

ABSTRACT

The purposes of this study were to create and to compare the prediction models for non–communicable diseases in the elderly by using data mining techniques. The investigated diseases included diabetes, hypertension and Ischemic heart disease while the mining techniques employed consisted of the Decision Tree, K–nearest neighbors and Naïve Bayes. The information of 262,655 patients' records who received treatments and services from Nawa Hospital, Nakhon Phanom Province from October 2011 to September 2019, had been examined and analyzed.

The study revealed the following results:

1. Among the created models for chronic non–communicable diseases in the elderly, the model for predicting diabetes using Decision Tree was the most efficient; it had 0.952 accuracy, 0.952 precision, 0.952 recall. Regarding the models to forecast hypertension, the ones using Decision Tree and K–nearest neighbors were the most efficient; it had 0.921 accuracy, 0.922 precision, 0.922 recall. Of the models to foretell Ischemic heart disease, the models which adopted the Decision Tree and K–nearest neighbors were the most efficient; it had 0.739 accuracy, 0.728 precision, 0.739 recall. Employing Decision Tree to predict diabetes, the patients' blood sugar, blood lipids, and blood pressure, respectively, were analyzed as the factors affecting the disease. Likewise, to foretell hypertension and Ischemic heart disease using

Decision Tree technique, the patients' blood lipids, blood sugar, and blood pressure were also analyzed as the factors impacting these diseases.

2. When comparing the created models for predicting the elderly's non-communicable diseases—diabetes, hypertension, and Ischemic heart disease, using data mining techniques, it was found that the Decision Tree technique was the most efficient. After the models for predicting the elderly's non-communicable diseases were assessed by 7 specialists, it was indicated that these prediction models' efficiency, as a whole, was at the high level ($\overline{X}$=4.43*)*.

Keywords Data mining, elderly, chronic non-communicable diseases, prediction model

สารบัญ

บทที่ **หน้า**

1 บทนำ 1
ภูมิหลัง 1
คำถามการวิจัย 4
ความมุ่งหมายของการวิจัย 4
ความสำคัญของการวิจัย 4
ขอบเขตของการวิจัย 5
กรอบแนวคิดการวิจัย 5
นิยามศัพท์เฉพาะ 8

2 เอกสารและงานวิจัยที่เกี่ยวข้อง 9
ผู้สูงอายุกับโรคไม่ติดต่อเรื้อรัง 10
ความหมายของผู้สูงอายุ 10
ปัญหาสุขภาพของผู้สูงอายุ 13
ความหมายของโรคเรื้อรัง 14
ประเภทของโรคเรื้อรัง 15
โรคไม่ติดต่อเรื้อรังในผู้สูงอายุ 20
สถานการณ์ปัญหาของโรคไม่ติดต่อเรื้อรังในผู้สูงอายุของประเทศไทย 22
ปัจจัยที่มีผลต่อการเกิดโรคไม่ติดต่อเรื้อรัง 27
เหมืองข้อมูล 32
ความหมายของเหมืองข้อมูล 32
กระบวนการเหมืองข้อมูล 33
ประเภทของข้อมูลที่ใช้ในเหมืองข้อมูล 37
เทคนิคเหมืองข้อมูล 38
เหมืองข้อมูลกับสารสนเทศทางการแพทย์ 40
การสร้างแบบจำลองด้วยเทคนิคเหมืองข้อมูล 41
ตัววัดประสิทธิภาพของแบบจำลอง 61

สารบัญ (ต่อ)

บทที่ **หน้า**

การแบ่งข้อมูลเพื่อใช้ในการวัดประสิทธิภาพ..... 62
โรงพยาบาลนาหว้า 65
โครงสร้างองค์กร วิสัยทัศน์ พันธกิจ ของโรงพยาบาลนาหว้า 65
นโยบายของโรงพยาบาลนาหว้า 71
โรคไม่ติดต่อเรื้อรังในผู้สูงอายุของโรงพยาบาลนาหว้า..... 71
การจัดเก็บข้อมูลของโรงพยาบาลนาหว้า 73
ตารางข้อมูลที่เกี่ยวข้อง 75
งานวิจัยที่เกี่ยวข้อง 76
ภายในประเทศ 76
ต่างประเทศ 82

3 วิธีการดำเนินการวิจัย 87
ระยะที่ 1 วิเคราะห์และเตรียมข้อมูล 87
ข้อมูล 87
เครื่องมือที่ใช้ในการวิจัย 87
วิธีการดำเนินการวิจัย 87
สถิติและการวิเคราะห์ข้อมูล 88
ระยะที่ 2 สร้างแบบจำลอง 88
ข้อมูล 88
เครื่องมือที่ใช้ในการวิจัย 89
วิธีการดำเนินการวิจัย 89
สถิติและการวิเคราะห์ข้อมูล 89
ระยะที่ 3 ประเมินประสิทธิภาพแบบจำลองโดยผู้เชี่ยวชาญ 89
วิธีการจัดประชุมกลุ่ม 89
เครื่องมือที่ใช้ในการวิจัย 90
วิธีการดำเนินการวิจัย 90

สารบัญ (ต่อ)

บทที่ **หน้า**

สถิติและการวิเคราะห์ข้อมูล 91

4 ผลการวิเคราะห์ข้อมูล 92
การสร้างแบบจำลองพยากรณ์การเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุ .. 92
การสร้างแบบจำลองพยากรณ์การเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุ โรคเบาหวาน 92
การสร้างแบบจำลองพยากรณ์การเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุ โรคความดันโลหิตสูง 105
การสร้างแบบจำลองพยากรณ์การเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุ โรคหัวใจขาดเลือด 120
การเปรียบเทียบประสิทธิภาพ 134
การเปรียบเทียบประสิทธิภาพแบบจำลองโรคเบาหวาน 134
การเปรียบเทียบประสิทธิภาพแบบจำลองโรคความดันโลหิตสูง 135
การเปรียบเทียบประสิทธิภาพแบบจำลองโรคหัวใจขาดเลือด 136
การประเมินประสิทธิภาพแบบจำลองโดยผู้เชี่ยวชาญ 139

5 สรุปผล อภิปรายผลและข้อเสนอแนะ 143
ความมุ่งหมายของการวิจัย 143
วิธีการดำเนินการวิจัย 143
สรุปผลการวิจัย 147
อภิปรายผลการวิจัย 149
ข้อเสนอแนะ 150

บรรณานุกรม 151
ภาคผนวก 158
ประวัติย่อของผู้วิจัย 200

บัญชีตาราง

ตาราง **หน้า**

1	การวิเคราะห์เกณฑ์อายุของผู้สูงอายุ	12
2	โรคไม่ติดต่อเรื้อรังที่พบบ่อยในผู้สูงอายุ.....	21
3	ร้อยละผู้ที่ได้รับการตรวจความดันโลหิตสูงภายใน 1 ปีที่ผ่านมา (พ.ศ. 2553 และ พ.ศ. 2558).....	25
4	การประเมินค่าดัชนีมวลกาย	30
5	การแปลค่าระดับความดันโลหิต	30
6	การแปลค่าระดับน้ำตาล	31
7	การแปลค่าระดับไตรกลีเซอไรด์	32
8	ข้อมูลที่เก็บสภาพภูมิอากาศ 14 วันย้อนหลัง	42
9	แบบจำลองนาอีฟเบย์ จากข้อมูลตาราง 8	58
10	ตัวอย่างเทคนิคเพื่อนบ้านใกล้ที่สุด	60
11	คอนฟิวชันเมทริกซ์ของข้อมูลที่มี 2 คลาส คือ yes และ no	61
12	แผนยุทธศาสตร์โรงพยาบาลนาหว้า ปีงบ 2560 – 2564	69
13	ข้อมูลผู้ป่วยเบาหวานโรงพยาบาลนาหว้า	72
14	ข้อมูลผู้ป่วยความดันโลหิตสูงโรงพยาบาลนาหว้า.....	72
15	ข้อมูลผู้ป่วยหัวใจขาดเลือดโรงพยาบาลนาหว้า.....	73
16	ข้อมูลที่เกี่ยวข้องกับการพยากรณ์การเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุ	76
17	การจัดกลุ่มแทนค่าข้อมูลโรคเบาหวาน	95
18	คอนฟิวชันเมทริกซ์ของแบบจำลองโรคเบาหวานของเทคนิคต้นไม้ตัดสินใจ	97
19	อธิบายกฎแบบจำลองโรคเบาหวานของเทคนิคต้นไม้ตัดสินใจ	99
20	คอนฟิวชันเมทริกซ์ของแบบจำลองโรคเบาหวานของเทคนิคเพื่อนบ้านใกล้ที่สุด.....	101
21	ตัวอย่างแบบจำลองโรคเบาหวานของเทคนิคเพื่อนบ้านใกล้ที่สุด	102

บัญชีตาราง (ต่อ)

ตาราง **หน้า**

22 คอนฟิวชันเมทริกซ์ของแบบจำลองโรคเบาหวาน ของเทคนิคนาอีฟเบย์ 103
23 แบบจำลองโรคเบาหวานของเทคนิคนาอีฟเบย์ 104
24 การจัดกลุ่มแทนค่าข้อมูลโรคความดันโลหิตสูง 108
25 คอนฟิวชันเมทริกซ์ของแบบจำลองโรคความดันโลหิตสูง ของเทคนิคต้นไม้ตัดสินใจ 110
26 อธิบายกฎแบบจำลองโรคความดันโลหิตสูง ของเทคนิคต้นไม้ตัดสินใจ 112
27 คอนฟิวชันเมทริกซ์ของแบบจำลองโรคความดันโลหิตสูง ของเทคนิคเพื่อนบ้านใกล้ที่สุด 115
28 ตัวอย่างแบบจำลองโรคความดันโลหิตสูง ของเทคนิคเพื่อนบ้านใกล้ที่สุด..... 116
29 คอนฟิวชันเมทริกซ์ของแบบจำลองโรคความดันโลหิตสูง ของเทคนิคนาอีฟเบย์ 118
30 แบบจำลองโรคความดันโลหิตสูงของเทคนิคนาอีฟเบย์ 119
31 การจัดกลุ่มแทนค่าข้อมูลโรคหัวใจขาดเลือด 123
32 คอนฟิวชันเมทริกซ์ของแบบจำลองโรคหัวใจขาดเลือด ของเทคนิคต้นไม้ตัดสินใจ 125
33 อธิบายกฎแบบจำลองโรคหัวใจขาดเลือดของเทคนิคต้นไม้ตัดสินใจ ... 127
34 คอนฟิวชันเมทริกซ์ของแบบจำลองโรคหัวใจขาดเลือด ของเทคนิคเพื่อนบ้านใกล้ที่สุด 129
35 ตัวอย่างแบบจำลองโรคหัวใจขาดเลือดของเทคนิคเพื่อนบ้านใกล้ที่สุด 130
36 คอนฟิวชันเมทริกซ์ของแบบจำลองโรคหัวใจขาดเลือด ของเทคนิคนาอีฟเบย์ 131
37 แบบจำลองโรคหัวใจขาดเลือดของเทคนิคนาอีฟเบย์ 132
38 ผลประสิทธิภาพของแบบจำลองโรคเบาหวาน 134

บัญชีตาราง (ต่อ)

ตาราง **หน้า**

39 ผลประสิทธิภาพของแบบจำลองโรคความดันโลหิตสูง 135

40 ผลประสิทธิภาพของแบบจำลองโรคหัวใจขาดเลือด 136

41 ข้อมูลผู้เชี่ยวชาญ 139

42 สรุปผลประสิทธิภาพของแบบจำลองโดยผู้เชี่ยวชาญด้านสาธารณสุข 140

บัญชีภาพประกอบ

ภาพประกอบ **หน้า**

1	กรอบแนวคิดเบื้องต้นในการวิจัย	7
2	แนวโน้มการเกิดโรคหลอดเลือดหัวใจและจำแนกกลุ่มอายุ.....	27
3	ปัจจัยที่มีผลต่อการเกิดโรคไม่ติดต่อเรื้อรัง	29
4	ขั้นตอนการค้นหาความรู้ในฐานข้อมูล.....	33
5	ขั้นตอนคริสป์–ดีเอ็ม.....	36
6	วิธีการคำนวณค่าเกนความรู้ของแอตทริบิวต์ Outlook	43
7	วิธีการคำนวณค่าเกนความรู้ของแอตทริบิวต์ Temperature	45
8	วิธีการคำนวณค่าเกนความรู้ของแอตทริบิวต์ Humidity	46
9	วิธีการคำนวณค่าเกนความรู้ของแอตทริบิวต์ Windy	47
10	วิธีการคำนวณค่าเกนความรู้ของแอตทริบิวต์ Outlook = Sunny และ Temperature	49
11	วิธีการคำนวณค่าเกนความรู้ของแอตทริบิวต์ Outlook = Sunny และ Humidity	51
12	วิธีการคำนวณค่าเกนความรู้ของแอตทริบิวต์ Outlook = Sunny และ Windy	52
13	วิธีการคำนวณค่าเกนความรู้ของแอตทริบิวต์ Outlook = Rainy และ Temperature	54
14	วิธีการคำนวณค่าเกนความรู้ของแอตทริบิวต์ Outlook = Rainy และ Windy	55
15	แบบจำลองต้นไม้ตัดสินใจที่สร้างจากข้อมูลในตาราง 8	57
16	ตัวอย่างการแบ่งข้อมูลแบบ 5–Fold Cross–validation	64
17	โครงสร้างองค์กรโรงพยาบาลนาหว้า	68
18	ตัวอย่างข้อมูลผู้รับการรักษาและผลการเป็นโรคเบาหวาน	93
19	ตัวอย่างข้อมูลโรคเบาหวานที่ผ่านการแปลงค่า	96
20	ผลลัพธ์ประสิทธิภาพแบบจำลองโรคเบาหวานของเทคนิคต้นไม้ตัดสินใจ	97

บัญชีภาพประกอบ (ต่อ)

ภาพประกอบ **หน้า**

21 แบบจำลองโรคเบาหวานของเทคนิคต้นไม้ตัดสินใจ 98
22 ผลลัพธ์ประสิทธิภาพแบบจำลองโรคเบาหวานของเทคนิค
เพื่อนบ้านใกล้ที่สุด 100
23 ผลลัพธ์ประสิทธิภาพแบบจำลองโรคเบาหวานของเทคนิคนาอีฟเบย์ 103
24 ตัวอย่างข้อมูลผู้รับการรักษาและผลการเป็นโรคความดันโลหิตสูง 106
25 ตัวอย่างข้อมูลโรคความดันโลหิตสูงที่ผ่านการแปลงค่า 109
26 ผลลัพธ์ประสิทธิภาพแบบจำลองโรคความดันโลหิตสูง
ของเทคนิคต้นไม้ตัดสินใจ 110
27 แบบจำลองโรคความดันโลหิตสูงของเทคนิคต้นไม้ตัดสินใจ 111
28 ผลลัพธ์ประสิทธิภาพแบบจำลองโรคความดันโลหิตสูง
ของเทคนิคเพื่อนบ้านใกล้ที่สุด 115
29 ผลลัพธ์ประสิทธิภาพแบบจำลองโรคความดันโลหิตสูง
ของเทคนิคนาอีฟเบย์ 117
30 ตัวอย่างข้อมูลผู้รับการรักษาและผลการเป็นโรคหัวใจขาดเลือด ... 121
31 ตัวอย่างข้อมูลโรคหัวใจขาดเลือดที่ผ่านการแปลงค่า 124
32 ผลลัพธ์ประสิทธิภาพแบบจำลองโรคหัวใจขาดเลือด
ของเทคนิคต้นไม้ตัดสินใจ 125
33 แบบจำลองโรคหัวใจขาดเลือดของเทคนิคต้นไม้ตัดสินใจ 126
34 ผลลัพธ์ประสิทธิภาพแบบจำลองโรคหัวใจขาดเลือด
ของเทคนิคเพื่อนบ้านใกล้ที่สุด 128
35 ผลลัพธ์ประสิทธิภาพแบบจำลองโรคหัวใจขาดเลือด
ของเทคนิคนาอีฟเบย์ 131
36 เทียบประสิทธิภาพของแบบจำลองโรคเบาหวาน 137
37 เทียบประสิทธิภาพของแบบจำลองโรคความดันโลหิตสูง..... 138
38 เทียบประสิทธิภาพของแบบจำลองโรคหัวใจขาดเลือด 138
39 ผลการประเมินประสิทธิภาพแบบจำลองโดยผู้เชี่ยวชาญด้านสาธารณสุข 141

บทที่ 1

บทนำ

ภูมิหลัง

ประเทศไทยเผชิญกับการเข้าสู่สังคมผู้สูงอายุ ส่งผลกระทบด้านเศรษฐกิจ สังคม และสุขภาพ ได้แก่ ปัญหาการขาดแคลนแรงงานในตลาด ปัญหาผู้สูงอายุถูกทอดทิ้ง ปัญหาทางด้านสภาพจิตใจ และปัญหาสุขภาพร่างกาย จากปัญหาดังกล่าวทำให้บริบททางสังคมเปลี่ยนแปลงไป เมื่อผู้สูงอายุมีมากขึ้นวัยกลางคนหรือวัยทำงานต้องออกไปทำงานเพื่อหารายได้ให้เพียงพอต่อการเลี้ยงครอบครัวมากขึ้นโดยละเลยการดูแลรักษาสุขภาพ ทำให้เกิดการสะสมของโรคในระยะยาวกลายเป็นผู้สูงอายุที่มีปัญหาด้านสุขภาพจากพฤติกรรมการใช้ชีวิต และเมื่อผู้สูงอายุมีอายุที่ยืนยาวขึ้นทำให้เกิดการดูแลระยะยาว (Long–Term Care) ซึ่งสูญเสียทั้งด้านอัตรากำลังและค่าใช้จ่ายในครอบครัว จึงเป็นวาระแห่งชาติที่รัฐบาลไทยได้ตระหนักและให้ความสำคัญ โดยให้หน่วยงานที่เกี่ยวข้องเตรียมและจัดทำแผนงานรองรับ เช่น แผนผู้สูงอายุแห่งชาติ ฉบับที่ 2 (พ.ศ. 2545 – 2564) แผนงานบูรณาการพัฒนาศักยภาพคนตามช่วงวัย แผนงานบูรณาการสร้างความเสมอภาคเพื่อรองรับสังคมผู้สูงอายุ เป็นต้น

ผู้สูงอายุ (Elderly) หมายถึงบุคคลซึ่งมีอายุ 60 ปีบริบูรณ์ขึ้นไป (พระราชบัญญัติผู้สูงอายุ, 2546) โดยแบ่งระดับการเข้าสู่สังคมผู้สูงอายุ เป็น 3 ระดับ ได้แก่ ระดับการก้าวเข้าสู่สังคมผู้สูงอายุ ระดับสังคมผู้สูงอายุโดยสมบูรณ์ และระดับสังคมผู้สูงอายุอย่างเต็มที่ ดังนั้นการเข้าสู่สังคมผู้สูงอายุจึงต้องวางแผนการใช้ชีวิตในบั้นปลาย การลงทุน การปรับตัวทางสังคมและจิตใจ รวมทั้งการจัดกิจกรรมเผยแพร่ความรู้ทางด้านสุขภาพอนามัยของผู้สูงอายุ เพราะปัญหาด้านสุขภาพร่างกายของผู้สูงอายุเสี่ยงสูงต่อการเกิดโรคเรื้อรัง (Chronic Disease) โดยโรคเรื้อรังที่พบมากในผู้สูงอายุจะเป็นโรคเบาหวาน โรคความดันโลหิตสูง โรคหัวใจและหลอดเลือด โรคไต เป็นต้น (กลุ่มยุทธศาสตร์และแผนงาน สำนักโรคไม่ติดต่อ, 2559)

โรคเรื้อรัง หมายถึงภาวะเจ็บป่วยเรื้อรัง เมื่อเป็นแล้วจะมีอาการหรือต้องรักษาต่อเนื่องเป็นระยะเวลานาน และรักษาไม่หาย ได้แก่โรคไม่ติดเชื้อ หรือเรียกว่าโรคไม่ติดต่อ

เรื้อรัง เช่น โรคเบาหวาน ความดันโลหิตสูง อัมพาต โรคหัวใจ โรคหืด หลอดลม อักเสบเรื้อรัง ถุงลมพอง ตับแข็ง มะเร็ง เป็นต้น (รุ่งทิวา ชอบชื่น, 2556) จากรายงานผลสำเร็จการดำเนินงานด้านสุขภาพตามนโยบายและยุทธศาสตร์กระทรวงสาธารณสุข ประจำปีงบประมาณ พ.ศ. 2558 พบว่า ในปี พ.ศ. 2557 อัตราการตายจากโรคไม่ติดต่อเรื้อรังมีแนวโน้มเพิ่มขึ้น เมื่อเปรียบเทียบกับปี พ.ศ. 2553 อาทิ โรคมะเร็งทุกชนิด จากร้อยละ 91.2 เป็นร้อยละ 107.9 โรคความดันเลือดสูงและโรคหลอดเลือดสมอง จากร้อยละ 31.1 เป็นร้อยละ 49.6 โรคหัวใจ จากร้อยละ 28.9 เป็นร้อยละ 38.5 และโรคเบาหวาน จากร้อยละ 10.8 เป็นร้อยละ 17.5 จากข้อมูลดังกล่าว กระทรวงสาธารณสุขจึงได้มียุทธศาสตร์ที่จะพัฒนาด้านการสร้างเสริมสุขภาพและป้องกันโรค โดยนำเทคโนโลยีสารสนเทศมารวบรวมจัดเก็บข้อมูลการบริการด้านสุขภาพของประชาชนเพื่อใช้ในการวางแผนงาน การสนับสนุนการตัดสินใจ และการจัดทำสถิติด้านสุขภาพให้เหมาะสมและเป็นประโยชน์ต่อการทำงานของกระทรวง และก่อให้เกิดประโยชน์สูงสุดกับประชาชน เช่น ข้อมูลคนไข้ ข้อมูลเวชภัณฑ์ เป็นต้น เกิดเป็นข้อมูลขนาดใหญ่ หรือที่เรียกว่า อภิมหาข้อมูล (Big Data) โดยเทคโนโลยีในการวิเคราะห์ข้อมูลขนาดใหญ่และหาความรู้ที่สำคัญเพื่อนำไปประกอบการตัดสินใจคือ การทำเหมืองข้อมูล (Data Mining) (สายชล สินสมบูรณ์ทอง, 2560; เอกสิทธิ์ พัชรวงศ์ศักดา, 2557)

เหมืองข้อมูล หมายถึงกระบวนการสกัด หรือค้นหาสารสนเทศ เพื่อให้ได้ความรู้ หรือสารสนเทศบางมุมที่ซ่อนเร้นอยู่ในฐานข้อมูลขนาดใหญ่ โดยอาศัยข้อมูลในอดีตเป็นจำนวนมาก เพื่อนำความรู้ที่ได้ไปช่วยในการวิเคราะห์ และประกอบการตัดสินใจของผู้บริหาร ในปัจจุบันเหมืองข้อมูลได้ถูกนำไปประยุกต์ใช้ในงานหลายประเภทและมีการนำมาใช้ประโยชน์ในองค์กรมากขึ้น ซึ่งรวมถึงสถานพยาบาลด้วย ข้อมูลจำนวนมากทั้งที่เป็นข้อมูลบริการและข้อมูลทางคลินิก สามารถจัดเก็บขึ้นเป็นคลังข้อมูลโดยเลือกใช้โมเดลการจัดทำเหมืองข้อมูลและเทคนิคที่เหมาะสม เช่น เทคนิคต้นไม้ตัดสินใจ การแบ่งกลุ่ม การจัดกลุ่ม การถดถอย ความสัมพันธ์ การพยากรณ์ การตรวจจับการเอนเอียง เป็นต้น สร้างแบบจำลองความสัมพันธ์ที่เป็นผลลัพธ์จากเหมืองข้อมูลมาสรุปหาความรู้ใหม่ที่เกิดขึ้นกับองค์กร ทั้งในด้านธุรกิจที่ช่วยในการตัดสินใจของผู้บริหาร ในด้านวิทยาศาสตร์ รวมทั้งในด้านเศรษฐกิจสังคม และด้านการแพทย์ (อดุลย์ ยิ้มงาม, 2554; Roiger, R. J., and Geatz, M., 2003)

ปัจจุบันเหมืองข้อมูลถูกนำมาใช้ในงานทางสุขภาพหรือทางการแพทย์ ในการค้นหารูปแบบการรักษาโรคที่หลากหลาย การพยากรณ์และการป้องกันโรค โดยนักวิจัยเหมืองข้อมูลให้ความสนใจมากขึ้น เพราะแนวโน้มทางการแพทย์ที่เป็นส่วนบุคคลมากขึ้น (Personalized Medicine) ที่มุ่งไปที่การค้นหา การรักษาและมาตรการป้องกันส่วนบุคคล ซึ่งข้อมูลที่สำคัญในเหมืองข้อมูลมาจากความก้าวหน้าทางเทคโนโลยีและการจัดเก็บข้อมูลที่มีมากขึ้น โดย ปัณณวิชญ์ วงศ์วิวัฒนานนท์ (2553) แบ่งเหมืองข้อมูลทางการแพทย์ออกเป็น 4 ด้าน ดังนี้ 1) งานวินิจฉัยโรค เพื่อวิเคราะห์ว่าผู้ป่วยเจ็บป่วยจากการแพทย์อะไรบ้าง 2) การพยากรณ์โรค เพื่อคาดคะเนว่าผู้ป่วยจะหายเจ็บป่วยได้ดีขนาดไหน และโรคจะดำเนินไปอย่างไรตามระยะเวลา 3) การรักษาที่เหมาะสม เพื่อคาดคะเนผลการรักษาโรค 4) การทำความเข้าใจเกี่ยวกับกลไกโรค เพื่อให้เกิดแนวคิดหรือความรู้ใหม่เกี่ยวกับสาเหตุการเกิดโรค

จากการทบทวนวรรณกรรมเหมืองข้อมูลกับงานทางการแพทย์ พบงานวิจัยในด้านนี้ เช่น งานวิจัยเรื่องการสร้างโมเดลทำนายโอกาสการกลับมารักษาตัวซ้ำของผู้ป่วยโรคเบาหวาน โดยใช้เทคนิคเหมืองข้อมูลและเปรียบเทียบประสิทธิภาพของโมเดลที่สร้างจากเทคนิคต้นไม้ตัดสินใจ เทคนิคเพื่อนบ้านใกล้ที่สุดและเทคนิคนาอีฟเบย์ ผลการวิจัยพบว่า โมเดลที่สร้างจากเทคนิคต้นไม้ตัดสินใจให้ค่าความถูกต้องมากที่สุดคิดเป็นร้อยละ 85.50 (วิธวินท์ แสงมณี, วีระวุฒ รัตนเจริญเลิศ, ณัฏฐภพ โพธิรัชต์ และเพียงฤทัย หนูสวัสดิ์, 2560) และงานวิจัยเรื่องระบบสนับสนุนทางการแพทย์เพื่อวิเคราะห์โรคหลอดเลือดแดงโป่งพองโดยใช้เทคนิคเหมืองข้อมูล กรณีศึกษาการรักษาหลอดเลือดแดงโป่งพองผ่านสายสวน โดยใช้เทคนิคต้นไม้ตัดสินใจ เทคนิคนาอีฟเบย์ และเทคนิคโครงข่ายประสาทเทียม ซึ่งจากผลการทดลองของการศึกษาแสดงให้เห็นว่าโมเดลต้นไม้ตัดสินใจมีประสิทธิภาพค่าความถูกต้องสูงที่สุดมีค่าเท่ากับร้อยละ 98.10 (นพรัตน์ พจน์จิราภรณ์, 2557)

โรงพยาบาลนาหว้า ตำบลนาหว้า อำเภอนาหว้า จังหวัดนครพนม เป็นโรงพยาบาลชุมชน ขนาด 30 เตียง ให้บริการด้านสุขภาพกับผู้ป่วย โดยมุ่งเน้นการส่งเสริมป้องกัน รักษา และฟื้นฟูภาวะความเจ็บป่วย หรือโรคต่าง ๆ ทั้งทางร่างกายและทางจิตใจ โดยโรงพยาบาลนาหว้ามีนโยบายในการใช้ข้อมูลสารสนเทศทางแพทย์ในการแก้ปัญหาสาธารณสุขในพื้นที่ ซึ่งมีแพทย์และพยาบาลเป็นผู้ประสานงานให้ความสำคัญในการดูแลเฝ้าระวังกลุ่มผู้ป่วยโรคไม่ติดต่อเรื้อรัง และการพัฒนาแนวทางการดูแล บริหารจัดการ

ผู้ป่วยกลุ่มโรคไม่ติดต่อเรื้อรังอยู่เสมอโดยใช้ข้อมูลสารสนเทศจากฐานข้อมูลของโรงพยาบาล โดยมีการสร้างระบบรายงานตอบสนองต่อความต้องการของผู้ใช้งาน รวมไปถึงการเสนอรายงานข้อมูลให้ผู้บริหารอยู่ตลอด แต่สำหรับการค้นหาคัดกรองผู้ป่วยและการส่งเสริมป้องกันโรคให้กับประชาชนในพื้นที่ยังคงใช้การสำรวจโดยอาสาสมัครในชุมชน พบว่าใช้กำลังคนมาก และเสียเวลาในการคัดกรองผู้ป่วย

จากปัญหาของโรคไม่ติดต่อเรื้อรังในผู้สูงอายุและความสำคัญของเหมืองข้อมูล ผู้วิจัยจึงสนใจศึกษาการสร้างแบบจำลองเพื่อพยากรณ์การเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุด้วยเทคนิคเหมืองข้อมูล: กรณีศึกษาโรงพยาบาลนาหว้า จังหวัดนครพนม จากข้อมูลที่อยู่ในฐานข้อมูลของโรงพยาบาลนาหว้า โดยใช้โปรแกรมเวก้า (Weka) ในการวิเคราะห์ข้อมูล และสร้างแบบจำลองมาประยุกต์ใช้งานคัดกรองและค้นหาโรคไม่ติดต่อเรื้อรังในผู้สูงอายุ ตามแผนงานกระทรวงสาธารณสุขเรื่องการป้องกันควบคุมโรคไม่ติดต่อเรื้อรังและภัยสุขภาพของกลุ่มผู้สูงอายุ เพื่อวางแผนป้องกันการเกิดโรคไม่ติดต่อเรื้อรังของผู้สูงอายุในอนาคต

คำถามการวิจัย

แบบจำลองเพื่อพยากรณ์การเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุ ด้วยเทคนิคเหมืองข้อมูล: กรณีศึกษาโรงพยาบาลนาหว้า จังหวัดนครพนม เป็นอย่างไร

ความมุ่งหมายของการวิจัย

1. เพื่อสร้างแบบจำลองเพื่อพยากรณ์การเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุ
2. เพื่อเปรียบเทียบประสิทธิภาพของแบบจำลองการเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุ

ความสำคัญของการวิจัย

1. ได้แบบจำลองเพื่อพยากรณ์การเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุ ที่สามารถประยุกต์กับการคัดกรองโรคไม่ติดต่อเรื้อรังในผู้สูงอายุได้
2. ได้ปัจจัยในการทำเหมืองข้อมูลโรคไม่ติดต่อเรื้อรังในผู้สูงอายุ
3. ได้เทคนิคเหมืองข้อมูลที่เหมาะสมในการหาแบบจำลองเพื่อพยากรณ์การเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุ

ขอบเขตของการวิจัย

ขอบเขตด้านข้อมูล

ข้อมูลที่ใช้ในการวิจัยแบบจำลองเพื่อพยากรณ์การเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุด้วยเทคนิคเหมืองข้อมูล คือข้อมูลผู้ที่มีอายุ 60 ปีขึ้นไป ที่มารับการรักษาจากโรงพยาบาลนาหว้า อ.นาหว้า จ.นครพนม ตั้งแต่เดือนตุลาคม พ.ศ. 2554 ถึง เดือนกันยายน พ.ศ. 2562 จำนวน 262,655 ระเบียน

ขอบเขตด้านตัวแปร

ตัวแปรอิสระ ได้แก่ เพศ อายุ การสูบบุหรี่ ดัชนีมวลกาย ความดันโลหิต น้ำตาลในเลือด ไขมันในเลือด และโรคไม่ติดต่อเรื้อรังได้แก่ โรคเบาหวาน โรคความดันโลหิตสูง และโรคหัวใจขาดเลือด

ตัวแปรตาม ได้แก่ แบบจำลองเพื่อพยากรณ์การเกิดโรคโรคเบาหวานในผู้สูงอายุ แบบจำลองเพื่อพยากรณ์การเกิดโรคความดันโลหิตสูงในผู้สูงอายุ และแบบจำลองเพื่อพยากรณ์การเกิดโรคหัวใจขาดเลือดในผู้สูงอายุ

ขอบเขตด้านเทคนิค

การพัฒนาแบบจำลองเพื่อพยากรณ์การเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุด้วยเทคนิคเหมืองข้อมูล ใช้เทคนิคของการจำแนกประเภทข้อมูล (Classification) ได้แก่ เทคนิคต้นไม้ตัดสินใจ (Decision Tree) เทคนิคนาอีฟเบย์ (Naïve Bayes) และเทคนิคเพื่อนบ้านใกล้ที่สุด (K–Nearest Neighbors)

กรอบแนวคิดการวิจัย

กรอบแนวคิดการวิจัยจะใช้ขั้นตอนคริสป์–ดีเอ็ม (Cross–Industry Standard Process for Data Mining : CRISP–DM) ของบริษัทเดมเลอร์ไครสเลอร์ (DaimlerChrysler) เอสพีเอสเอส (SPSS) และเอ็นซีอาร์ (NCR) เป็นกระบวนการมาตรฐานการวิเคราะห์ข้อมูลด้านเหมืองข้อมูล (สายชล สินสมบูรณ์ทอง, 2560) มาสร้างแบบจำลองเพื่อพยากรณ์การเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุด้วยเทคนิคเหมืองข้อมูลตามขั้นตอน คริสป์–ดีเอ็ม 6 ขั้นตอนคือ

1. ขั้นตอนการทำความเข้าใจการวิจัยและธุรกิจ (Business Understanding)
2. ขั้นตอนการทำความเข้าใจข้อมูล (Data Understanding)

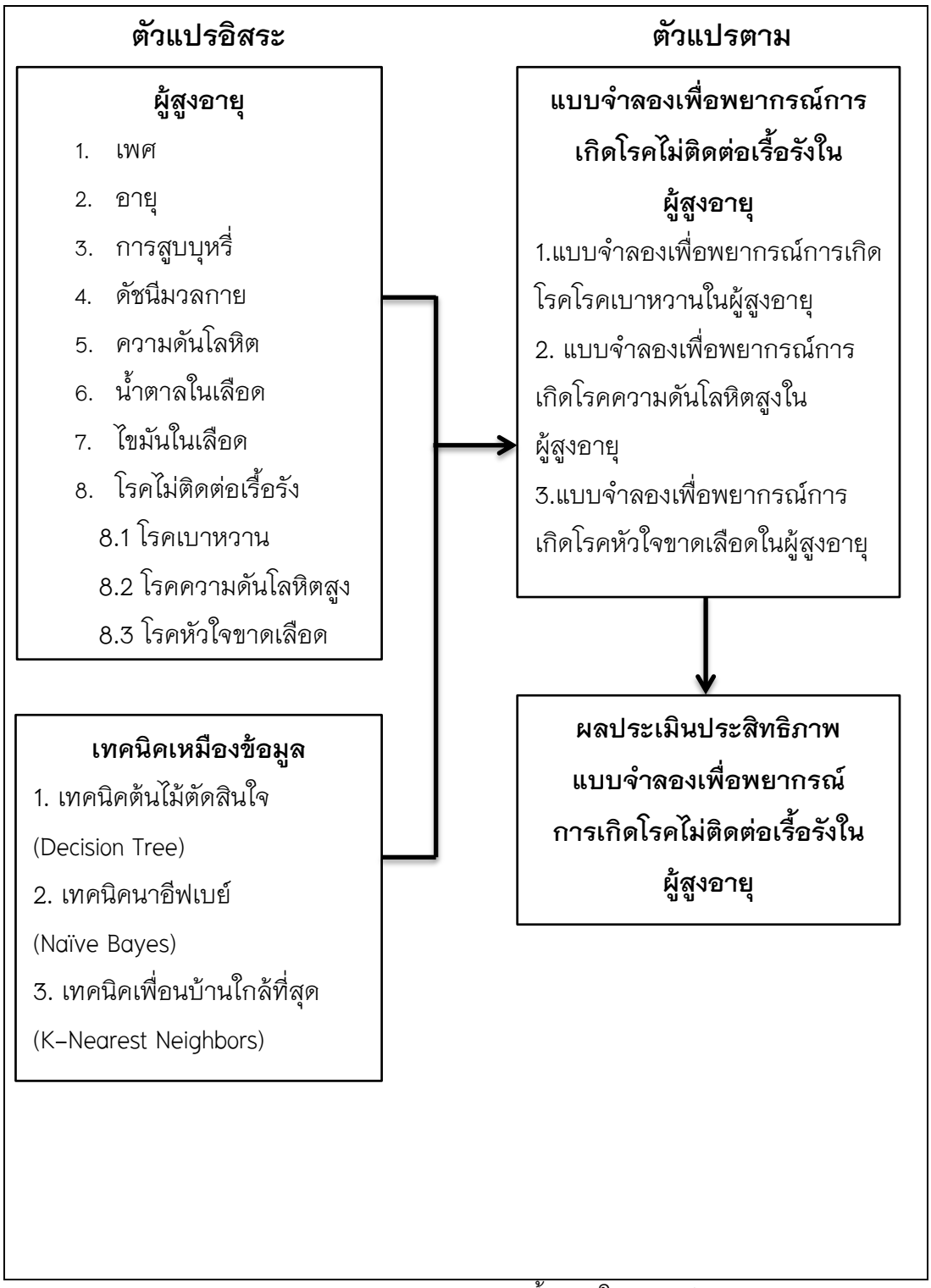
3. ขั้นตอนการจัดเตรียมข้อมูล (Data Preparation)

4. ขั้นตอนการสร้างตัวแบบ (Modeling)

5. ขั้นตอนการประเมินตัวแบบ (Evaluation)

6. ขั้นตอนการใช้งาน (Deployment)

ในการวิจัยครั้งนี้ผู้วิจัยดำเนินการเพียง 5 ขั้นตอน คือ 1) ขั้นตอนการทำความเข้าใจการวิจัยและธุรกิจ 2) ขั้นตอนการทำความเข้าใจข้อมูล 3) ขั้นตอนการจัดเตรียมข้อมูล 4) ขั้นตอนการสร้างตัวแบบ 5) ขั้นตอนการประเมินตัวแบบ และมีตัวแปรที่ใช้ในการวิจัยคือปัจจัยที่มีผลต่อการเกิดโรคไม่ติดต่อเรื้อรัง (กรมการแพทย์ กระทรวงสาธารณสุข, 2557) ได้แก่ เพศ อายุ การสูบบุหรี่ ดัชนีมวลกาย ความดันโลหิต น้ำตาลในเลือด ไขมันในเลือด และ 3 โรคไม่ติดต่อเรื้อรังคือ โรคเบาหวาน โรคความดันโลหิตสูง และโรคหัวใจขาดเลือด ดังภาพประกอบ 1



ภาพประกอบ 1 กรอบแนวคิดเบื้องต้นในการวิจัย

นิยามศัพท์เฉพาะ

1. ผู้สูงอายุ หมายถึง บุคคลที่มาใช้บริการโรงพยาบาลนาหว้า ที่มีอายุ ตั้งแต่ 60 ปี ขึ้นไป โดยแบ่งกลุ่มผู้สูงอายุได้ 3 กลุ่มได้แก่ กลุ่มที่ 1 ผู้สูงอายุวัยต้น (อายุ 60–69 ปี) เป็นช่วงที่ยังมีพลังช่วยเหลือตนเองได้กลุ่มที่ 2 ผู้สูงอายุวัยกลาง (อายุ 70–79 ปี) เริ่มมีอาการเจ็บป่วย ร่างกายเริ่มอ่อนแอ มีโรคประจำตัว หรือโรคเรื้อรัง กลุ่มที่ 3 ผู้สูงอายุวัยปลาย (อายุ 80 ปีขึ้นไป) เจ็บป่วยบ่อยขึ้น อวัยวะเสื่อมสภาพ อาจมีภาวะทุพพลภาพ

2. โรคไม่ติดต่อเรื้อรังในผู้สูงอายุ หมายถึงภาวะเจ็บป่วยเรื้อรัง เมื่อเป็นแล้วจะมีอาการหรือต้องรักษาต่อเนื่องเป็นระยะเวลานาน และรักษาไม่หายที่พบบ่อยในผู้สูงอายุ คือ 1) โรคโรคเบาหวาน 2) โรคความดันโลหิตสูง 3) โรคหัวใจขาดเลือด

3. เทคนิคเหมืองข้อมูล หมายถึง กระบวนการสกัด หรือ ค้นหาสารสนเทศเพื่อให้ได้ความรู้ หรือสารสนเทศบางมุมที่ซ่อนเร้นอยู่ในฐานข้อมูลขนาดใหญ่ โดยอาศัยข้อมูลในอดีตเป็นจำนวนมาก เพื่อนำความรู้ที่ได้ไปช่วยวิเคราะห์ และประกอบการตัดสินใจ ได้แก่ 1) เทคนิคต้นไม้ตัดสินใจ เป็นการคัดเลือกแอตทริบิวต์ที่มีความสัมพันธ์กับคลาสมากที่สุดขึ้นมาเป็นโหนดบนสุดของต้นไม้ (Root Node) หลังจากนั้นจะหาแอตทริบิวต์ถัดไปโดยจะใช้ตัววัดคือค่าเกนความรู้ (Information Gain) 2) เทคนิคนาอีฟเบย์ เป็นการใช้ทฤษฎีความน่าจะเป็นตามกฎของเบย์ (Bayes' Theorem) เพื่อหาว่าสมมติฐานใดน่าจะถูกต้องที่สุด โดยใช้ความรู้ก่อนหน้า (Prior Knowledge) 3) เทคนิคเพื่อนบ้านใกล้ที่สุด เป็นการเปรียบเทียบความคล้ายคลึงกันของข้อมูลที่สนใจกับข้อมูลอื่นว่ามีความคล้ายคลึงหรืออยู่ใกล้กับข้อมูลใดมากที่สุด k ตัว โดยใช้วิธีการวัดระยะทางแบบยุคลิด (Euclidean Distance)

4. แบบจำลองเพื่อพยากรณ์การเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุ หมายถึง สิ่งที่ได้รับการพัฒนาขึ้นเพื่ออธิบายหรือแสดงให้เห็นถึงองค์ประกอบที่สำคัญของการเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุ สามารถทำความเข้าใจการทำงานของระบบได้ง่ายกว่าการศึกษาโดยตรง

5. โรงพยาบาลนาหว้า ตำบลนาหว้า อำเภอนาหว้า จังหวัดนครพนม เป็นโรงพยาบาลชุมชน ขนาด 30 เตียง ให้บริการด้านสุขภาพให้กับผู้ป่วย โดยโรงพยาบาลนาหว้ามีนโยบายในการใช้ข้อมูลสารสนเทศทางแพทย์ในการแก้ปัญหาสาธารณสุขในพื้นที่ ซึ่งมีแพทย์และพยาบาลเป็นผู้ประสานงานให้ความสำคัญในการดูแลเฝ้าระวังกลุ่มผู้ป่วยโรคไม่ติดต่อเรื้อรัง และการพัฒนาแนวทางการดูแล บริหารจัดการผู้ป่วยกลุ่มโรคไม่ติดต่อเรื้อรังอยู่เสมอโดยใช้ข้อมูลสารสนเทศจากฐานข้อมูลของโรงพยาบาล

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

การศึกษาวิจัยเรื่อง แบบจำลองเพื่อพยากรณ์การเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุด้วยเทคนิคเหมืองข้อมูล: กรณีศึกษาโรงพยาบาลนาหว้า จังหวัดนครพนม ผู้วิจัยทบทวนวรรณกรรมที่เกี่ยวข้องซึ่งสามารถสรุปสาระสำคัญตามลำดับ ดังนี้

1. ผู้สูงอายุกับโรคไม่ติดต่อเรื้อรัง
    1.1 ความหมายของผู้สูงอายุ
    1.2 ปัญหาสุขภาพของผู้สูงอายุ
    1.3 ความหมายของโรคเรื้อรัง
    1.4 ประเภทของโรคเรื้อรัง
    1.5 โรคไม่ติดต่อเรื้อรังในผู้สูงอายุ
    1.6 สถานการณ์ปัญหาของโรคไม่ติดต่อเรื้อรังในผู้สูงอายุของประเทศไทย
    1.7 ปัจจัยที่มีผลต่อการเกิดโรคไม่ติดต่อเรื้อรัง
2. เหมืองข้อมูล
    2.1 ความหมายของเหมืองข้อมูล
    2.2 กระบวนการเหมืองข้อมูล
    2.3 ประเภทของข้อมูลที่ใช้ในเหมืองข้อมูล
    2.4 เทคนิคเหมืองข้อมูล
    2.5 เหมืองข้อมูลกับสารสนเทศทางการแพทย์
    2.6 การสร้างแบบจำลองด้วยเทคนิคเหมืองข้อมูล
    2.7 ตัววัดประสิทธิภาพของแบบจำลอง
    2.8 การแบ่งข้อมูลเพื่อใช้ในการวัดประสิทธิภาพ
3. โรงพยาบาลนาหว้า
    3.1 โครงสร้างองค์กร วิสัยทัศน์ พันธกิจ ของโรงพยาบาลนาหว้า
    3.2 นโยบายของโรงพยาบาลนาหว้า
    3.3 โรคไม่ติดต่อเรื้อรังในผู้สูงอายุของโรงพยาบาลนาหว้า
    3.4 การจัดเก็บข้อมูลของโรงพยาบาลนาหว้า

3.5 ตารางข้อมูลที่เกี่ยวข้อง

4. งานวิจัยที่เกี่ยวข้อง

4.1 ในประเทศ

4.2 ต่างประเทศ

1. ผู้สูงอายุกับโรคไม่ติดต่อเรื้อรัง

แนวคิดทฤษฎีเกี่ยวกับผู้สูงอายุกับโรคไม่ติดต่อเรื้อรังประกอบไปด้วยหัวข้อดังนี้

1.1 ความหมายของผู้สูงอายุ

ปัจจุบันประเทศไทยได้ให้ความหมายของผู้สูงอายุไว้คือผู้ที่มีอายุ 60 ปีบริบูรณ์ขึ้นไป ตามพระราชบัญญัติผู้สูงอายุฉบับปัจจุบัน พ.ศ. 2546 และใช้เป็นอายุที่ได้รับสิทธิจากราชการ เช่น อายุเกษียณของข้าราชการ หรืออายุที่เริ่มได้รับเบี้ยยังชีพ (กระทรวงการพัฒนาสังคมและความมั่นคงของมนุษย์, 2553) โดยแบ่งกลุ่มผู้สูงอายุได้ 3 กลุ่มได้แก่

กลุ่มที่ 1 ผู้สูงอายุวัยต้น (อายุ 60–69 ปี) เป็นช่วงที่ยังมีพลังช่วยเหลือตนเองได้

กลุ่มที่ 2 ผู้สูงอายุวัยกลาง (อายุ 70–79 ปี) เริ่มมีอาการเจ็บป่วย ร่างกายเริ่มอ่อนแอ มีโรคประจำตัว หรือโรคเรื้อรัง

กลุ่มที่ 3 ผู้สูงอายุวัยปลาย (อายุ 80 ปีขึ้นไป) เจ็บป่วยบ่อยขึ้น อวัยวะเสื่อมสภาพ อาจมีภาวะทุพพลภาพ

การกำหนดผู้สูงอายุที่ 60 ปีไม่ได้เป็นเกณฑ์หรือข้อตกลงที่เป็นสากล แต่มีความแตกต่างกันไปในแต่ละประเทศ เนื่องด้วยกระบวนการสูงอายุเป็นกระบวนการที่ความเป็นพลวัต ภายใต้เงื่อนไขเชิงโครงสร้างของแต่ละสังคมเป็นผู้ให้ความหมายการสูงอายุ ทำให้อายุเริ่มต้นที่ใช้ในการนิยามผู้สูงอายุ หรือ ระบุถึงการเข้าสู่ความสูงอายุของบุคคลที่ใช้อยู่ปัจจุบัน ไม่สามารถอธิบายถึงหลักการและเหตุผลที่ชัดเจน หรือเกณฑ์ข้อตกลงที่เป็นสากลได้ โดยทั่วไปพิจารณาการสูงอายุจากองค์ประกอบในด้านต่าง ๆ เช่น การสูงอายุตามวัยหรือตามปีปฏิทินที่ผ่านไป การเปลี่ยนแปลงทางชีววิทยาของสภาพร่างกายและความสามารถในการดำเนินกิจกรรมต่าง ๆ ในชีวิต หรือการเปลี่ยนแปลงของสภาพและบทบาททางสังคมของบุคคล มาตรฐานที่ใช้ในการกำหนดได้ส่วนใหญ่เป็นการกำหนดโดยใช้อายุตามปีปฏิทินของบุคคลที่เชื่อว่าสามารถแสดงถึงช่วงวัยของบุคคลที่มีลักษณะ

ของความสูงอายุซึ่งแตกต่างกันไปในแต่ละสังคม (รศรินทร์ เกรย์, อุมาภรณ์ ภัทรวาณิชย์, เฉลิมพล แจ่มจันทร์ และเรวดี สุวรรณนพเก้า, 2556)

องค์การสหประชาชาติ (United Nations: UN) ให้ความหมายผู้สูงอายุ (Elderly) หมายถึงบุคคลหรือกลุ่มประชากรที่มีอายุตามปีปฏิทิน (Calendar Age, Chronological Age) ตั้งแต่ 60 ปีขึ้นไป ซึ่งเป็นเกณฑ์อายุเริ่มต้นเหมือนกับองค์การอนามัยโลก (World Health Organization) ที่ใช้กำหนดอายุของผู้สูงอายุ ในกลุ่มประเทศพัฒนาแล้ว รวมถึงประเทศไทย บางประเทศกำหนดอายุของผู้สูงอายุแตกต่างจากองค์การสหประชาชาติ เช่น กลุ่มประเทศในภูมิภาคแอฟริกา กำหนดให้ผู้สูงอายุมีเกณฑ์อายุของบุคคล ตั้งแต่ 50 ปี หรือ 55 ปีขึ้นไป

นอกจากนี้ยังมีงานวิจัยได้อธิบาย “ลักษณะของความสูงอายุ” เกี่ยวกับความหมายและลักษณะของการสูงอายุในมุมมองต่าง ๆ โดยพิจารณาจากการเปลี่ยนแปลงทางด้านร่างกายและสุขภาพ ความสูงอายุจะถูกอธิบายในเชิงลบว่าเป็นปรากฏการณ์ที่เกิดขึ้นในช่วงวัยที่เป็นระยะสุดท้ายของชีวิตมนุษย์ ซึ่งมีลักษณะและพัฒนาการที่ตรงข้ามกับวัยเด็ก มีแต่ความเสื่อมโทรมและสึกหรอ ซึ่งดำเนินไปอย่างค่อยเป็นค่อยไป มีสภาพร่างกายที่ผิวหนังเหี่ยวย่น ผมหงอกขาว ฟันสั่นคลอน ด้านสุขภาพที่ค่อย ๆ เสื่อมสภาพ เรี่ยวแรงน้อยลง เหนื่อยง่าย ตาฝ้าฟาง หูตึง รับกลิ่นรสได้แย่กว่าแต่ก่อน ความจำเริ่มเสื่อม เจ็บป่วยง่ายแต่หายได้ช้า และด้านจิตใจที่ขาดความมั่นใจในตนเอง กังวลง่าย เหงา เศร้าและขี้น้อยใจ บางคนโมโหร้ายและชอบแยกตัว จากลักษณะดังกล่าวนี้ ให้นัยต่อคำว่า ผู้สูงอายุ หรือ ผู้ที่มีลักษณะของความสูงอายุ ถูกตีความในลักษณะ “ความไม่มีประโยชน์” และ “ความเสื่อมถอยด้านศักยภาพในการทำงาน และการพึ่งพาตนเอง” (สุชาดา ทวีสิทธิ์, 2553) นิยามผู้สูงอายุมีการเปลี่ยนแปลง ไปตามกาลเวลาหรือวัตถุประสงค์ของโครงการที่เกี่ยวข้อง โดยนักวิชาการทั้งในประเทศและต่างประเทศที่เป็นสังคมผู้สูงอายุ เสนอให้มีการเลื่อนนิยามผู้สูงอายุในประเทศของตนเองให้สูงขึ้น เช่น ประเทศญี่ปุ่น จาก 65 ปี เป็น 75 ปี และสำหรับประเทศไทยจาก 60 ปี เป็น 65 ปี จะมีผลกระทบในวงกว้างทั้งทางด้านเศรษฐกิจ จิตวิทยาสังคม และสุขภาพ (วรชัย ทองไทย, 2549; ปราโมทย์ ประสาทกุล และปัทมา ว่าพัฒนวงศ์, 2553) การกำหนดนิยามผู้สูงอายุทั้งในประเทศไทยและต่างประเทศ อยู่ภายใต้แนวคิดว่าวัยสูงอายุเป็นวัยที่เปราะบาง ควรได้รับสวัสดิการการดูแลจากภาครัฐเพราะประสิทธิภาพการทำงานลดลงจึงมักเป็นอายุ

เดียวกันกับอายุเกษียณ (สำนักงานคณะกรรมการข้าราชการพลเรือน, 2542; รศรินทร์ เกรย์, อุมาภรณ์ ภัทรวาณิชย์, เฉลิมพล แจ่มจันทร์ และเรวดี สุวรรณนพเก้า, 2556)

จากการสังเคราะห์ความหมายที่กล่าวถึงเกณฑ์อายุของผู้สูงอายุ ผู้วิจัยจึงจัดทำตาราง 1 ดังนี้

ตาราง 1 การวิเคราะห์เกณฑ์อายุของผู้สูงอายุ

ผู้ให้นิยาม	United Nations [UN] (n.d.)	World Health Organization [WHO] (n.d.)	พรบ. ผู้สูงอายุ (2546)	ผู้วิจัย
อายุของผู้สูงอายุ	ตั้งแต่ 60 ปีขึ้นไป	ในกลุ่มประเทศที่พัฒนาแล้วตั้งแต่ 65 ปีขึ้นไป ในกลุ่มประเทศกำลังพัฒนาตั้งแต่ 60 ปีขึ้นไป บางประเทศในแอฟฟริกาตั้งแต่ 50,55 ปีขึ้นไป	ตั้งแต่ 60 ปีขึ้นไป	ตั้งแต่ 60 ปีขึ้นไป

ผู้วิจัยสรุปความหมายของผู้สูงอายุในงานวิจัยฉบับนี้หมายถึง บุคคลที่มีอายุ 60 ปีบริบูรณ์ และแบ่งกลุ่มผู้สูงอายุได้ 3 กลุ่มได้แก่ ผู้สูงอายุวัยต้น (อายุ 60–69 ปี) ผู้สูงอายุวัยกลาง (อายุ 70–79 ปี) และผู้สูงอายุวัยปลาย (อายุ 80 ปีขึ้นไป) เนื่องจากเป็นงานวิจัยที่ใช้กับกลุ่มผู้สูงอายุในประเทศไทย ผู้วิจัยจึงยึดตามพระราชบัญญัติผู้สูงอายุ พ.ศ. 2546

1.2 ปัญหาสุขภาพของผู้สูงอายุ

ผู้สูงอายุในมุมมองทางด้านจิตวิทยาสังคม และสุขภาพภาวะสูงอายุทางสรีระวิทยาเป็นการเปลี่ยนแปลงที่เกิดขึ้นกับผู้สูงอายุทางร่างกาย เป็นการเปลี่ยนแปลงทางกายภาพที่จะปรากฏให้เห็นอย่างชัดเจนกับร่างกายของคนเมื่อวัยสูงขึ้น กระบวนการนี้มี 2 ลักษณะ คือการเปลี่ยนแปลงทางร่างกายและการเปลี่ยนแปลงทางสรีระวิทยา (สมศักดิ์ ศรีสันติสุข, 2539)

1.2.1 การเปลี่ยนแปลงทางร่างกาย เป็นการเปลี่ยนแปลงทางกายภาพและโครงสร้างของร่างกายที่ปรากฏให้เห็นอย่างชัดเจน เช่น ผิวหนังเหี่ยวย่น ตกกระ ผิวบาง เกิดบาดแผลได้ง่าย กล้ามเนื้อลดจำนวนลงทำให้ความแข็งแรงของกล้ามเนื้อลดลง กระดูกเปราะบาง กระดูผุ กระดูกข้ออักเสบ เป็นต้น

1.2.2 การเปลี่ยนแปลงทางสรีระวิทยาเป็นการเปลี่ยนแปลงคุณภาพของอวัยวะต่าง ๆ ในร่างกายที่เคยใช้งานได้ดี เช่น การใช้สายตา หู ลิ้น หย่อนสมรรถภาพลง จำนวนเซลล์สมองลดลงทำให้ความจำเสื่อมลงไป ระบบทางเดินหายใจ ทางเดินอาหาร ทางเดินปัสสาวะทำงานได้น้อยลง ทำให้เกิดอาการผิดปกติตามมา เช่น อาหารไม่ย่อย เป็นไข้หวัดได้ง่าย อั้นปัสสาวะไม่ได้ เป็นต้น

ผลการสำรวจสุขภาพของผู้สูงอายุไทยในงานวิจัย พบว่าปัญหาหลักของผู้สูงอายุคือ ปัญหาสุขภาพ ทั้งที่สุขภาพไม่สมบูรณ์ ไม่แข็งแรง และปัญหาการเจ็บป่วยด้วยโรคต่าง ๆ ซึ่งโรคที่มักพบในผู้สูงอายุทั่วไปได้แก่ ปวดหลัง ปวดเอว ไขข้ออักเสบ ความดันโลหิตสูง โรคกระเพาะ โรคหัวใจ ต้อกระจกตาต้อเนื้อตา โรคเกี่ยวกับหู อัมพาต อัมพฤกษ์ และจากการศึกษาของ ปรีชา อุปโยคิน และคณะ (2541) พบว่าผู้สูงอายุยอมรับว่าตนเองเป็นผู้สูงอายุ เพราะสังเกตและประเมินจากสุขภาพตัวเอง ยอมรับความบกพร่องของร่างกาย และปัญหาสุขภาพ โดยยกตัวอย่างภาวะเจ็บป่วยที่สำคัญคือ ปวดเมื่อยตามแขนขาไหล่ เข่า ปวดหลัง ปวดข้อ เวลาเดินทางไปไหนมาไหนรู้สึกเหนื่อยง่าย เดินไปได้ไม่ไกลมีความอ่อนเพลีย ไม่มีเรี่ยวแรง และเป็นลมหน้ามืดบ่อย ตามัวมองอะไรไม่ค่อยเห็น หูไม่ค่อยได้ยินเวลาใครพูดต้องใช้มือป้องหู ฟันหลุด ผมหงอก ผิวหนังเหี่ยวย่นเป็นสัญญาณที่บ่งบอกความชราภาพ รวมถึงการมีโรคประจำตัวซึ่งจะไม่ค่อยพบในวัยหนุ่มสาว แต่เมื่อมีอายุมากขึ้นและเป็นผู้สูงอายุ ส่วนใหญ่หรือแทบทุกคนจะต้องมีโรคใดโรคหนึ่ง เช่น โรคเบาหวาน โรคความดันโลหิตสูง วัณโรค โรคหัวใจ โรคระบบประสาท เป็นต้น

1.3 ความหมายของโรคเรื้อรัง

ตามความหมายที่กระทรวงสาธารณสุขใช้มี 2 ความหมาย คือภาวะเจ็บป่วย และด้านความผิดปกติ (โครงการสำรวจอนามัยและสวัสดิการ, 2546) ได้แก่

โรคเรื้อรัง หมายถึง ภาวะเจ็บป่วยเรื้อรัง ที่ไม่ติดต่อหรือภาวะที่เกิดจากความเสื่อมของร่างกายตามอายุ ส่วนใหญ่จะไม่มีสาเหตุที่แน่นอน มีปัจจัยเสี่ยงหลายอย่าง มีระยะเวลาฟักตัวของโรครวมถึงระยะเวลาของการดำเนินโรคที่ยาวนาน ไม่ติดต่อโดยทางสัมผัส สามารถทำให้เกิดความพิการหรือการทำงานที่ผิดปกติของร่างกาย และรักษาไม่หาย

โรคเรื้อรัง หมายถึง ภาวะที่มีความผิดปกติหรือเบี่ยงเบนไปจากปกติของร่างกาย เป็นภาวะที่เป็นอย่างถาวร ทำให้เกิดความพิการ เกิดจากความเปลี่ยนแปลงทางพยาธิสภาพที่ไม่สามารถรักษาให้กลับคืนเป็นปกติได้ ผู้ป่วยที่อยู่ในภาวะนี้ต้องการ การทำกายภาพบำบัดหรือการปรับตัวสำหรับภาวะนี้ นอกจากนั้นยังต้องการรักษาติดตามดูแลในระยะยาว รวมทั้งโรคประจำตัวที่สามารถรักษาหายได้ แต่ถ้ารักษาติดต่อกันนานเกิน 3 เดือนแล้วยังไม่หาย ยังต้องรักษาต่อไป ให้ถือว่าเป็นโรคเรื้อรัง

รุ่งทิวา ชอบชื่น (2556) กล่าวว่า โรคเรื้อรัง (Chronic Disease) หมายถึง โรคที่เมื่อเป็นแล้วจะมีอาการหรือต้องรักษาติดต่อกันนาน เป็นแรมเดือนแรมปีหรือตลอดชีวิต มักได้แก่โรคไม่ติดเชื้อ (Non–Infectious Disease) เป็นส่วนใหญ่ เช่น โรคเบาหวาน ความดันโลหิตสูง อัมพาต โรคหัวใจ โรคหืด หลอดลม อักเสบเรื้อรัง ถุงลมพอง ตับแข็ง มะเร็ง เป็นต้น

สำนักงานกองทุนสนับสนุนการสร้างเสริมสุขภาพ หรือ สสส. (2557) กล่าวว่า กลุ่มโรคเอ็นซีดี (Non–Communicable Diseases: NCDs) หรือ ชื่อภาษาไทยเรียกว่า กลุ่มโรคไม่ติดต่อเรื้อรัง เป็นชื่อเรียก กลุ่มโรคที่ไม่ได้มีสาเหตุมาจากการติดเชื้อ ไม่ได้เกิดจากเชื้อโรค ไม่สามารถติดต่อได้ผ่านการสัมผัส คลุกคลี หรือ ติดต่อ ผ่านตัวนำโรค (พาหะ) หรือสารคัดหลั่งต่าง ๆ หากแต่เกิดจากปัจจัยต่าง ๆ ภายในร่างกาย ซึ่งส่วนใหญ่เป็นผลจากวิธีการใช้ชีวิต ที่มีพฤติกรรมเสี่ยงได้แก่ ดื่มสุรา สูบบุหรี่ ขาดการออกกำลังกาย รับประทานอาหารหวานมันเค็มจัด และมีความเครียด

ผู้วิจัยจึงสรุปได้ว่า โรคเรื้อรังหมายถึง ภาวะเจ็บป่วยเรื้อรัง ที่เมื่อเป็นแล้วจะมีอาการหรือต้องรักษาติดต่อกันนาน และรักษาไม่หาย ซึ่งแบ่งออกเป็นโรคติดต่อเรื้อรัง และโรคไม่ติดต่อเรื้อรัง

1.4 ประเภทของโรคเรื้อรัง

ในปี พ.ศ. 2558 องค์การอนามัยโลก (World Health Organization, 2015) ได้จัดทำรายงานเกี่ยวกับโรคเรื้อรัง โดยมีการกำหนดโรคเรื้อรังหลักที่สำคัญ 4 โรคและโรคเรื้อรังอื่นที่อยู่ในเกณฑ์มี 35 โรค มีรายละเอียดดังนี้

1.4.1 มะเร็ง (Cancer) คือ โรคที่เกิดจากความผิดปกติของเซลล์ในอวัยวะต่าง ๆ ของร่างกาย โดยมีการเจริญเติบโตที่ผิดปกติเกิดเป็นก้อนเนื้อที่มีการลุกลามไปยังอวัยวะข้างเคียง หรือกระจายไปยังส่วนอื่น ๆ ของร่างกายได้ ผ่านทางระบบเลือด หรือระบบทางเดินน้ำเหลือง โรคมะเร็งมีหลากหลายชนิดขึ้นอยู่กับตำแหน่งที่เป็นจุดกำเนิดของโรค และชนิดของเซลล์มะเร็ง โดยองค์การอนามัยโลกได้จัดให้มะเร็งเป็นโรคเรื้อรังหลัก

1.4.2 โรคหัวใจและหลอดเลือด (Cardiovascular Diseases) ได้แก่ โรคหลอดเลือดสมอง (Cerebrovascular Disease) หลอดเลือดหัวใจล้มเหลว (Heart Failure) และโรคหัวใจขาดเลือด (Ischemic Cardiopathy) เป็นกลุ่มโรคเกี่ยวกับหัวใจหรือหลอดเลือด คือเป็นโรคที่มีผลต่อระบบหัวใจและหลอดเลือด รวมทั้งกลุ่มโรคหลอดเลือดเสี้ยงหัวใจ กลไกของโรคจะขึ้นอยู่กับเหตุ โรคหลอดเลือดเสี้ยงหัวใจ โรคหลอดเลือดสมอง และโรคหลอดเลือดแดงส่วนปลาย จะมีภาวะหลอดเลือดแดงแข็ง โดยองค์การอนามัยโลกได้จัดให้โรคหัวใจและหลอดเลือดเป็นโรคเรื้อรังหลัก

1.4.3 โรคระบบทางเดินหายใจเรื้อรัง (Chronic Respiratory Diseases) ได้แก่ โรคหอบหืด (Asthma) และโรคปอดอุดกั้นเรื้อรัง (Chronic Obstructive Pulmonary Disease: COPD) คือความบกพร่องอย่างเรื้อรังในการไหลผ่านของอากาศของระบบทางเดินหายใจ และมักจะแย่ลงเมื่อเวลาผ่านไป อาการที่พบบ่อยได้แก่ เหนื่อย ไอและมีเสมหะ โดยองค์การอนามัยโลกได้จัดให้โรคระบบทางเดินหายใจเรื้อรัง เป็นโรคเรื้อรังหลัก

1.4.4 โรคเบาหวาน (Diabetes Disease) เป็นกลุ่มโรคเกี่ยวกับการเผาผลาญอาหารซึ่งมีระดับน้ำตาลในเลือดสูงเป็นเวลานาน น้ำตาลในเลือดสูงก่อให้เกิดอาการปัสสาวะบ่อย กระหายน้ำและความหิวเพิ่มขึ้น หากไม่ได้รับการรักษาเบาหวานอาจก่อให้เกิดอาการแทรกซ้อนจำนวนมาก ภาวะแทรกซ้อนเฉียบพลัน โดยองค์การอนามัยโลกได้จัดให้โรคเบาหวานเป็นโรคเรื้อรังหลัก

1.4.5 ติดยา (Addiction) คืออาการที่ผู้เสพยาไม่สามารถควบคุมหรือบังคับตนเองจากอาการอยากยาได้

1.4.6 โรคอัลไซเมอร์ (Alzheimer's Disease) เป็นภาวะสมองเสื่อม

เกิดจากความเสื่อมถอยของการทำงานหรือโครงสร้างของเนื้อเยื่อของสมองซึ่งมักพบในผู้สูงอายุ

1.4.7 ภาวะหัวใจเต้นผิดจังหวะ (Atrial Fibrillation) ภาวะแทรกซ้อนของการมีคลื่นไฟฟ้าหัวใจเต้นผิดปกติ ภาวะแทรกซ้อนของโรคนี้มี 2 อย่างหลัก ๆ คือ การเกิดลิ่มเลือด และหัวใจวาย

1.4.8 โรคสมาธิสั้น (Attention Deficit Hyperactivity Disorder) เป็นภาวะที่ผู้ป่วยมีพฤติกรรมที่หุนหันพลันแล่น ไม่มีสมาธิ และซน

1.4.9 โรคภูมิคุ้มกันทำลายตนเอง (Autoimmune Disease) เป็นโรคที่ร่างกายมีการโจมตีตนเอง เนื่องจากระบบภูมิคุ้มกันของเรา สูญเสียความสามารถในการแยกแยะ

1.4.10 โรคอารมณ์แปรปรวนสองขั้ว หรือโรคไบโพลาร์ (Bipolar Disorder) คือ โรคที่ทำให้มีอารมณ์ซึมเศร้าในช่วงหนึ่ง และมีอารมณ์ดีผิดปกติในอีกช่วงหนึ่งสลับกันไป

1.4.11 ตาบอด (Blindness) คือ ภาวะที่ดวงตาไม่สามารถมองเห็นสิ่งต่าง ๆ หรือรับรู้ความแตกต่างระหว่างแสงสว่างและความมืดได้

1.4.12 อัมพาตสมองทุกประเภท (Cerebral Palsy All Types) คือ สมองพิการ นิยมเรียก โรคซีพี เป็นโรคในเด็ก ทำให้มีปัญหาการทรงตัวซึ่งแต่ละคนจะมีลักษณะที่แตกต่างกัน เช่น กล้ามเนื้ออ่อนแรง เกร็ง งุ่มง่าม เคลื่อนไหวช้า ทรงตัวได้ไม่ดี เด็กสมองพิการบางคนอาจมีความบกพร่องอื่นร่วมด้วย

1.4.13 ภาวะสเต็มเซลล์ใหม่ต้านร่างกายเรื้อรัง (Chronic Graft–Versus–Host Disease: GVHD) เป็นภาวะแทรกซ้อนที่พบได้บ่อยของการปลูกถ่ายอวัยวะ ส่วนใหญ่พบกับการปลูกถ่ายเซลล์ต้นกำเนิดหรือไขกระดูก แต่ก็พบได้กับการปลูกถ่ายเนื้อเยื่ออื่น ๆ เช่นกัน เกิดจากการที่เม็ดเลือดขาวชนิดทีเซลล์ (T–Cell) ในเนื้อเยื่อที่ปลูกถ่ายทำปฏิกิริยากับเนื้อเยื่อในร่างกายของผู้รับการปลูกถ่ายโดยมองว่าเป็นสิ่งแปลกปลอม

1.4.14 โรคตับอักเสบเรื้อรัง (Chronic Hepatitis) เป็นภาวะที่มีการอักเสบของตับและเกิดการทำลายของเซลล์ตับ ทำให้การทำหน้าที่ต่าง ๆ ของตับผิดปกติ ร่างกายอาจแสดงอาการเจ็บป่วยเล็กน้อยหรือไม่แสดงอาการเลยแต่มักจะนำไปสู่อาการดีซ่าน

1.4.15 โรคไตเรื้อรัง (Chronic Kidney Disease) คือภาวะที่ไตไม่สามารถทำงานได้อย่างมีประสิทธิภาพเป็นระยะยาว

1.4.16 โรคข้อเข่าเสื่อมเรื้อรังเช่นโรคข้อเข่าเสื่อมและโรคข้ออักเสบรูมาตอยด์ (Chronic Osteoarticular Diseases, Such As Osteoarthritis And Rheumatoid Arthritis) เป็นกลุ่มของภาวะที่เกิดการทำลายข้อต่อของร่างกาย ข้ออักเสบมีได้มากกว่าร้อยรูปแบบ โดยรูปแบบที่พบมากที่สุดคือข้อเสื่อม

1.4.17 อาการปวดเรื้อรัง (Chronic Pain Syndromes) หมายถึงอาการปวดที่เป็นอยู่นานกว่า 12 สัปดาห์ แม้จะได้รับการรักษาด้วยยาหรือการรักษาอื่น ๆ แล้วก็ตาม

1.4.18 โรคหูหนวกและความบกพร่องทางการได้ยิน (Deafness And Hearing Impairment) บุคคลที่บกพร่องหรือสูญเสียทางการ ได้ยินเป็นเหตุให้การรับฟังเสียงต่าง ๆ ได้ ไม่ชัดเจน ตั้งแต่ระดับรุนแรงถึงระดับน้อย

1.4.19 โรคการกินผิดปกติ (Eating Disorders) เป็นความผิดปกติทางจิตซึ่งนิยามจากนิสัยการกินที่มีผลกระทบทางลบต่อสุขภาพทางกายหรือจิตของบุคคล ได้แก่ ความผิดปกติของการรับประทานแบบตะกละ

1.4.20 โรคหนังยืดผิดปกติ (Ehlers–Danlos Syndrome) ภาวะที่เส้นเอ็นทั่วร่างกายหย่อนแต่กำเนิด เป็นกลุ่มในโรคทางพันธุกรรม ที่เกี่ยวกับความบกพร่องของชั้นคอนเนคทิฟทิชชู

1.4.21 เยื่อบุโพรงมดลูกเจริญผิดที่ (Endometriosis) คือการที่มีเซลล์คล้ายกับเยื่อบุโพรงมดลูกไปเจริญขึ้นที่อื่นนอกเหนือจากในโพรงมดลูกที่ควรอยู่

1.4.22 โรคลมบ้าหมู (Epilepsy) คือโรคที่เกิดจากความผิดปกติของระบบประสาทส่วนกลางซึ่งทำหน้าที่ในการควบคุมการทำงานของร่างกาย จนทำให้เกิดอาการชัก โรคลมชักสามารถเกิดขึ้นได้กับทุกเพศทุกวัย แต่ก็มักจะพบในผู้ป่วยเด็ก และผู้สูงอายุ

1.4.23 โรคปวดกล้ามเนื้อเรื้อรังทั่วทั้งตัว (Fibromyalgia) เป็นโรคเรื้อรังที่ส่งผลให้ผู้ป่วยรู้สึกเจ็บปวดตามกล้ามเนื้อร่วมกับมีอาการอ่อนเพลีย มีปัญหาด้านความจำและสมาธิ รวมทั้งอาจนอนไม่หลับหรือหลับไม่สนิท โรคนี้ไม่อาจรักษาให้หายขาด แพทย์จึงจะประคับประคองอาการโดยให้ผู้ป่วยรับประทานยา ออกกำลังกาย และปรับพฤติกรรมเพื่อลดความเครียดเป็นหลัก

1.4.24 โรคเอชไอวี/เอดส์ (HIV/AIDS) เป็นโรคของระบบภูมิคุ้มกันของมนุษย์ ซึ่งเกิดจากการติดเชื้อไวรัสเอชไอวี (Human Immunodeficiency Virus, HIV) ทำให้

ผู้ป่วยมีการทำงานของระบบภูมิคุ้มกันบกพร่อง เสี่ยงต่อการติดเชื้อฉวยโอกาสและการเกิดเนื้องอกบางชนิด

1.4.25 โรคฮันติงตัน (Huntington's Disease) คือโรคที่เกิดจากการเสื่อมของระบบประสาทส่วนกลาง/สมอง เกิดจากความผิดปกติทางพันธุกรรมอย่างหนึ่ง ส่งผลให้เซลล์ของสมองบริเวณเบซัลแกงเกลีย (Basal Ganglia) เกิดการเสื่อมและตายลง ส่งผลให้ผู้ป่วยมีอาการเคลื่อนไหวร่างกายผิดปกติแบบโคเรีย (Chorea) คือการเคลื่อนไหวผิดปกติที่ไม่เป็นจังหวะ ไม่สม่ำเสมอ เร็ว และไม่ซ้ำแบบเดิม

1.4.26 โรคความดันโลหิตสูง (Hypertension) เป็นโรคเรื้อรังชนิดหนึ่งที่ผู้ป่วยมีความดันเลือดในหลอดเลือดแดงสูงกว่าปกติตลอดเวลา ความดันเลือดประกอบด้วยสองค่า ได้แก่ ความดันช่วงหัวใจบีบและความดันช่วงหัวใจคลาย ซึ่งเป็นความดันสูงสุดและต่ำสุดในระบบหลอดเลือดแดงตามลำดับ ความดันช่วงหัวใจบีบเกิดเมื่อหัวใจห้องล่างซ้ายบีบตัวมากที่สุด ความดันช่วงหัวใจคลายเกิดเมื่อหัวใจห้องล่างซ้ายคลายตัวมากที่สุดก่อนการบีบตัวครั้งถัดไป

1.4.27 โรคลายม์ (Lyme Disease) เป็นโรคที่เกิดจากเชื้อแบคทีเรีย Borrelia Burgdorferi (Lyme Borreliosis) ที่แพร่โดยตัวเห็บ (Ixodes Ricinus Tick) ทำให้เกิดข้ออักเสบภายหลังการติดเชื้อ ผู้ป่วยที่ติดเชื้อแบคทีเรียชนิดนี้อาจมีอาการทางผิวหนัง ทางระบบประสาท หัวใจ ตาและอวัยวะอื่น ๆ

1.4.28 โรคปลอกประสาทอักเสบ (Multiple Sclerosis) เป็นโรคที่มีการอักเสบของปลอกประสาทที่หุ้มเส้นประสาทในส่วนของ สมอง เส้นประสาทตา และไขสันหลัง โดยปลอกประสาทมีหน้าที่สำคัญในการนำกระแสประสาท ซึ่งเมื่อมีการอักเสบของปลอกประสาทส่งผลให้การนำกระแสประสาทของอวัยวะนั้นผิดปกติ

1.4.29 โรคไข้สมองอักเสบจากการติดเชื้อ (Myalgic Encephalomyelitis) หรือโรคอ่อนเพลียเรื้อรัง (Hronic Fatigue Syndrome) เป็นภาวะที่ส่งผลกระทบต่อร่างกายในระยะยาวทำให้เกิดความเหนื่อยล้าหรืออ่อนเพลียอยู่อย่างถาวรที่ไม่ดีขึ้นด้วยการนอนหลับหรือพักผ่อน และส่งผลต่อการดำเนินชีวิตประจำวัน โดยทั่วไป โรคจะมีผลกระทบต่อระบบประสาท และระบบภูมิคุ้มกันของร่างกาย ในบางครั้งก็มีการวินิจฉัยอาการดังกล่าวเป็นกลุ่มอาการอ่อนล้าหลังติดเชื้อไวรัส

1.4.30 โรคลมหลับ (Narcolepsy) เป็นความผิดปกติด้านการนอนที่เกิดขึ้นอย่างเรื้อรัง ผู้ป่วยจะรู้สึกง่วงอย่างมากในช่วงกลางวันและมักหลับไปโดยไม่รู้ตัว

1.4.31 โรคอ้วน (Obesity) หมายถึง สภาวะร่างกายที่มีน้ำหนักตัวมากกว่าปกติ เกิดจากการสะสม ของไขมันใต้ผิวหนังมากกว่าปกติจนมีผลกระทบต่อสุขภาพ

1.4.32 โรคกระดูกพรุน (Osteoporosis) คือโรคที่ลดความแข็งแรงของกระดูกและเพิ่มความเสี่ยงที่จะทำให้เกิดกระดูกหัก โรคนี้เป็นสาเหตุหลักของการเกิดกระดูกหักในผู้สูงอายุ ซึ่งมักเกิดกับกระดูกไขสันหลัง, กระดูกแขนท่อนปลายและกระดูกสะโพก โรคกระดูกพรุนมักไม่แสดงอาการ ผู้ป่วยต้องอาศัยการตรวจกระดูกเท่านั้นจึงจะทราบ

1.4.33 โรคพาร์กินสัน (Parkinson's Disease) เป็นโรคที่เกิดจากกลุ่มของเซลล์เล็ก ๆ ในสมองทำงานผิดปกติ ทำให้ผู้ป่วยมีอาการสั่น เกร็ง เคลื่อนไหวช้า

1.4.34 โรคปริทันต์ (Periodontal Disease) หมายถึง โรคที่เกิดกับเนื้อเยื่อหรืออวัยวะรอบตัวฟัน ได้แก่ เหงือก เนื้อเยื่อปริทันต์ และกระดูกหุ้มรากฟัน ในระยะแรกที่เป็นโรคนี้จะไม่มีอาการ ต่อมาจะมีการทำลายอวัยวะเหล่านี้ ทำให้กระดูกหุ้มรากฟันละลาย และเหงือกร่น อาจมีการปวดบวม ฟันโยกและหลุดในที่สุด ชาวบ้านมักเรียกว่าโรครำมะนาด

1.4.35 กลุ่มอาการหัวใจเต้นเร็วระหว่างเปลี่ยนท่า (Postural Orthostatic tachycardia syndrome) เป็นภาวะที่อัตราการเต้นของหัวใจจะเร็วขึ้นอย่างน้อย 30 ครั้งต่อนาทีเมื่อยืนขึ้น แต่แม้ว่าอัตราการเต้นของหัวใจจะเพิ่มขึ้น แต่ความดันโลหิตจะยังอยู่ในระดับเดิม การที่มีอัตราการเต้นของหัวใจเพิ่มขึ้นนั้น หมายความว่า ระบบหัวใจ และหลอดเลือดกำลังพยายามควบคุมความดันโลหิตและการไหลเวียนของเลือดเข้าสู่สมอง

1.4.36 โรคเม็ดเลือดแดงรูปเคียว (Sickle Cell Anemia And Other Hemoglobin Disorders) คือโรคที่ถ่ายทอดทางพันธุกรรม ผู้ป่วยจะมีความผิดปกติของยีนที่เกี่ยวข้องกับการสร้างเม็ดเลือดแดง ทำให้เม็ดเลือดแดงมีรูปร่างผิดปกติ คือ เป็นรูปเคียว

1.4.37 ภาวะหยุดหายใจขณะหลับ (Sleep Apnea) เป็นภาวะความผิดปกติอย่างหนึ่งของการหายใจที่เกิดขึ้นระหว่างนอนหลับ เป็นอันตรายและอาจทำให้เกิดความผิดปกติอื่นตามมาจนถึงเสียชีวิตได้ พบได้ในทุกอายุ โดยมักจะพบมากขึ้นตามอายุที่มากขึ้น ผู้ชายมากกว่าผู้หญิง หรือกลุ่มคนที่น้ำหนักตัวมาก

1.4.38 โรคไทรอยด์ (Thyroid Disease) เป็นภาวะที่เกิดขึ้นเมื่อต่อมไทรอยด์ผลิตฮอร์โมนออกมามากเกินไป ต่อมไทรอยด์เป็นต่อมที่อยู่บริเวณคอ มีหน้าที่

สร้างฮอร์โมนที่เกี่ยวข้องกับกระบวนการสร้างใช้พลังงาน และกระบวนการอื่น ๆ เมื่อมีการผลิตฮอร์โมนออกมามากเกินไป จึงทำให้การทำงานของร่างกายหลาย ๆ ส่วนเกิดขึ้นอย่างรวดเร็วจนเกินไป

1.4.39 การใช้ยาสูบ (Tobacco Use and Related Conditions) เป็นภาวะติดยา

จากการทบทวนวรรณกรรมประเภทโรคเรื้อรังที่พบในคนทั่วไป หากวิเคราะห์ตามลักษณะของโรคโดยทั่วไปสามารถแบ่งได้เป็น 2 ประเภท คือ โรคติดต่อเรื้อรัง และ โรคไม่ติดต่อเรื้อรัง

ซึ่งกลุ่มโรคไม่ติดต่อเรื้อรัง (Non–Communicable Diseases) นั้นเป็นชื่อเรียกกลุ่มโรคที่ไม่ได้มีสาเหตุมาจากการติดเชื้อ ไม่ได้เกิดจากเชื้อโรค ไม่สามารถติดต่อได้ผ่านการสัมผัส คลุกคลี หรือ ติดต่อ ผ่านตัวนำโรค (พาหะ) หรือสารคัดหลั่งต่าง ๆ หากแต่เกิดจากปัจจัยต่าง ๆ ภายในร่างกาย ซึ่งส่วนใหญ่เป็นผลจากการใช้ชีวิตประจำวันที่ไม่เหมาะสม เช่น โรคเบาหวาน (Diabetes Mellitus) โรคหลอดเลือดสมองและหัวใจ (Cardiovascular & Cerebrovascular Diseases) โรคถุงลมโป่งพอง (Emphysema) โรคมะเร็ง (Cancer) โรคความดันโลหิตสูง (Hypertension) โรคอ้วนลงพุง (Obesity) เป็นต้น

1.5 โรคไม่ติดต่อเรื้อรังในผู้สูงอายุ

โรคไม่ติดต่อเรื้อรังในผู้สูงอายุ คือโรคไม่ติดต่อเรื้อรังที่พบบ่อยในผู้สูงอายุ ส่วนหนึ่งมาจากการเจ็บป่วยสะสมตั้งแต่วัยหนุ่มสาวหรือวัยทำงาน แต่ไม่ได้รับการรักษาหรือการดูแลอย่างต่อเนื่อง ทำให้เกิดอาการรุนแรงในวัยสูงอายุ หลายโรคเกิดจากพฤติกรรมที่ไม่เหมาะสม ทั้งการบริโภคอาหาร ดื่มเครื่องดื่มแอลกอฮอล์ สูบบุหรี่ ขาดการออกกำลังกาย และขาดการควบคุมอารมณ์ที่ดี จากการสังเคราะห์งานวิจัยและบทความที่มีการกล่าวถึงโรคไม่ติดต่อเรื้อรังที่พบบ่อยในผู้สูงอายุซึ่งส่วนใหญ่เป็นโรคไม่ติดต่อเรื้อรังดังตาราง 2

ตาราง 2 โรคไม่ติดต่อเรื้อรังที่พบบ่อยในผู้สูงอายุ

โรคไม่ติดต่อเรื้อรังที่พบบ่อยในผู้สูงอายุ	วรพรรณ เสนาณรงค์ (2557)	สสส. (2561)	รพ.พญาไท (2560)	เอกชัย เพียรศรีวัชรา (2557)	ศศิวิมล พงษ์สุพจน์ (2559)	ภานุวัฒน์ ปานเกตุ (2557)	ผู้วิจัย
ภาวะไขมันในเลือดสูง			/	/			
ความดันโลหิตสูง	/	/	/	/	/		/
โรคหัวใจขาดเลือด		/		/	/	/	/
โรคหลอดเลือดสมอง				/			
โรคข้อเข่าเสื่อม				/			
โรคเบาหวาน	/	/	/	/	/	/	/
โรคในช่องปาก				/			
โรคทางตา				/	/		
โรคระบบทางเดินอาหารและระบบปัสสาวะ				/	/		
โรคมะเร็ง				/		/	
สมองเสื่อม		/		/	/		
โรคไต			/		/		
โรคทางเดินหายใจเรื้อรัง						/	
อัมพาต	/						
โรคอ้วน			/				
โรคเก๊าท์					/		

จากตาราง 2 โรคไม่ติดต่อเรื้อรังที่พบบ่อยในผู้สูงอายุมีหลายโรค ผู้วิจัยจึงเลือกศึกษาและทำวิจัยในโรคไม่ติดต่อเรื้อรังที่พบบ่อยในผู้สูงอายุ คือ โรคเบาหวาน โรคความดันโลหิตสูง และโรคหัวใจขาดเลือด

1.6 สถานการณ์ปัญหาของโรคไม่ติดต่อเรื้อรังในผู้สูงอายุของประเทศไทย

จากการทบทวนวรรณกรรมโรคไม่ติดต่อเรื้อรังในประเทศไทยพบว่าโรคไม่ติดต่อเรื้อรังที่ถูกกล่าวถึงและให้ความสำคัญเป็นอันดับต้น ๆ ของประเทศไทยได้แก่ โรคเบาหวาน โรคความดันโลหิตสูง และโรคหัวใจขาดเลือด

1.6.1 โรคเบาหวาน

เบาหวาน (Diabetes Mellitus) เป็นโรคไม่ติดต่อเรื้อรังที่รักษาไม่หาย ต้องได้รับการดูแลรักษาตลอดชีวิต ทั้งนี้เกิดจากการที่ในเลือดมีน้ำตาลสูงกว่าปกติ พบได้สูงในคนทุกอายุ และทั้งสองเพศ แต่จะพบสูงขึ้นเมื่อสูงอายุ ปัจจุบันเนื่องจากจำนวนผู้สูงอายุมีมากขึ้นทั่วโลกที่รวมถึงประเทศไทย จึงพบเบาหวานสูงขึ้นเรื่อย ๆ มีการคาดการณ์จากสหพันธ์เบาหวานนานาชาติ (The International Diabetes Federation: IDF) จำนวนผู้ป่วยเบาหวานทั่วโลกจะเพิ่มขึ้นจาก 366 ล้านคนในปี ค.ศ. 2011 (พ.ศ. 2554) เป็น 552 ล้านคนในปี ค.ศ. 2030 (พวงทอง ไกรพิบูลย์, 2562)

ซึ่งปัจจัยเสี่ยงต่อการเกิดโรคเบาหวานมีดังนี้

1.6.1.1 โรคอ้วนและน้ำหนักตัวเกิน ซึ่งส่งผลให้เซลล์ต่าง ๆ ดื้อต่ออินซูลิน

1.6.1.2 ขาดการออกกำลังกาย เพราะการออกกำลังกายจะช่วยควบคุมน้ำหนัก และช่วยให้เซลล์ต่าง ๆ ไวต่อการนำน้ำตาลไปใช้ หรือช่วยการเผาผลาญน้ำตาลในเลือดได้ดี

1.6.1.3 พันธุกรรม เพราะพบว่า คนที่มีครอบครัวสายตรง (พ่อ แม่ พี่น้อง ท้องเดียวกัน) เป็นเบาหวานมีโอกาสเป็นเบาหวานสูงกว่าคนทั่วไป

1.6.1.4 เชื้อชาติ เพราะพบว่า คนบางเชื้อชาติเป็นเบาหวานสูงกว่า เช่น ในคนเอเชีย และในคนผิวดำ

1.6.1.5 อายุ เพราะพบว่ายิ่งอายุสูงขึ้น โอกาสเป็นเบาหวานยิ่งสูงขึ้น อาจจากการเสื่อมถอยของเซลล์ตับอ่อน หรือขาดการออกกำลังกายจากสุขภาพที่เสื่อมถอย

1.6.1.6 มีไขมันในเลือดสูง เพราะพบว่ายิ่งอายุสูงขึ้นไขมันสะสมมากขึ้น

1.6.1.7 มีความดันโลหิตสูง

รายงานจากการสำรวจสุขภาพประชาชนไทยโดยการตรวจร่างกายครั้งที่ 4 พ.ศ. 2551 – 2552 พบความชุกของโรคเบาหวานในประชากรไทยอายุ 15 ปีขึ้นไป คิดเป็นร้อยละ 6.9 ทั้งนี้ พบว่าหนึ่งในสามของผู้ที่เป็นเบาหวานไม่เคยได้รับการวินิจฉัยว่าเป็นเบาหวานมาก่อน และมีผู้ที่ได้รับการวินิจฉัยว่าเป็นเบาหวานแต่ไม่ได้รับการรักษาคิดเป็นร้อยละ 3.3 ผู้ป่วยเบาหวานที่สามารถควบคุมน้ำตาลให้อยู่ในเกณฑ์ที่เหมาะสมได้มีประมาณร้อยละ 40 อัตราการเข้าพักรักษาตัวในโรงพยาบาลของผู้ป่วยที่เป็นโรคเบาหวานมีแนวโน้มเพิ่มมากขึ้นอย่างต่อเนื่อง ตั้งแต่ ปี พ.ศ. 2541 – 2551 โดยพบว่ามีอัตราเพิ่มขึ้นจาก 175.7 เป็น 675.7 ต่อแสนประชากร และมีผู้เสียชีวิตจากโรคเบาหวานใน ปี พ.ศ. 2552 ประมาณ 7,019 คน หรือประมาณวันละ 19 คน

สหพันธ์เบาหวานนานาชาติได้คาดการณ์ว่าในปี พ.ศ.2583 ประเทศไทยจะมีผู้ป่วยเบาหวานสูงถึง 5.3 ล้านคน กระทรวงสาธารณสุขได้เล็งเห็นถึงความสำคัญดังกล่าว โดยระบุให้เบาหวานเป็นหนึ่งในโรคที่กระทรวงให้ความสำคัญในแง่ของการป้องกันและรักษา ตามแผนยุทธศาสตร์ชาติ ระยะ 20 ปี (ด้านสาธารณสุข) ซึ่งจัดทำโดยสำนักนโยบายและยุทธศาสตร์ในปี พ.ศ.2560 คณะรัฐมนตรีได้อนุมัติในหลักการของแผนยุทธศาสตร์การป้องกัน และควบคุมโรคไม่ติดต่อ 5 ปี (พ.ศ.2560–2564) เพื่อให้ทุกภาคส่วนมีความเป็นเอกภาพในการบริหารจัดการเพื่อสุขภาพของประชาชน บูรณาการการทำงานร่วมกัน และเน้นป้องกันโรคมากกว่าการรักษา โดยการสร้างเสริมสุขภาพ และคุณภาพชีวิตของประชาชนไทยตลอดช่วงชีวิตทุกกลุ่มวัย เพื่อการก้าวสู่สังคมสุขภาพที่ยั่งยืน

1.6.2 โรคความดันโลหิตสูง

โรคความดันโลหิตสูง หรือ ภาวะความดันโลหิตสูง (Hypertension ย่อว่า HT หรือ HTN, หรือ High blood pressure: HBP) เป็นโรคพบบ่อยมากอีกโรคหนึ่งในผู้ใหญ่ พบสูงถึงประมาณ 25% ของประชากรโลกที่เป็นผู้ใหญ่ทั้งหมด โดยพบในผู้ชายมากกว่าในผู้หญิง และพบได้สูงขึ้นในผู้สูงอายุ ในบางประเทศ เช่นในยุโรป พบโรคนี้ได้สูงถึง 30–45% ส่วนในเด็กพบโรคนี้ได้เช่นกัน แต่พบได้น้อยกว่าในผู้ใหญ่มาก (พวงทอง ไกรพิบูลย์, 2562)

ปัจจัยเสี่ยงสำคัญที่ทำให้เกิดโรคความดันโลหิตสูง ได้แก่

1.6.2.1 พันธุกรรม โอกาสมีความดันโลหิตสูง จะสูงขึ้นเมื่อมีคนในครอบครัวเป็นโรคนี้

1.6.2.2 โรคเบาหวาน เพราะก่อให้เกิด การอักเสบที่ต่อเนื่องเป็นการตีบแคบของหลอดเลือดต่าง ๆ รวมทั้งของไต

1.6.2.3 โรคอ้วนและน้ำหนักตัวเกิน เพราะเป็นสาเหตุสำคัญของโรคเบาหวาน และโรคหลอดเลือดต่าง ๆ ตีบจากภาวะไขมันเกาะผนังหลอดเลือด (โรคหลอดเลือดแดงแข็ง)

1.6.2.4 โรคไตเรื้อรัง เพราะจะส่งผลถึงการสร้างเอนไซม์ และฮอร์โมนที่ควบคุมความดันโลหิต

1.6.2.5 โรคนอนหลับแล้วหยุดหายใจ (Sleep apnea)

1.6.2.6 สูบบุหรี่ เพราะสารพิษในควันบุหรี่ส่งผลให้เกิด การอักเสบ ตีบตัน ของหลอดเลือดต่าง ๆ รวมทั้งหลอดเลือดไต และหลอดเลือดหัวใจ

1.6.2.7 การติดสุรา ซึ่งยังไม่ทราบชัดเจนถึงกลไกว่าทำไมดื่มสุราแล้วจึงเป็นปัจจัยเสี่ยงต่อการเกิดความดันโลหิตสูง แต่การศึกษาต่าง ๆ ให้ผลตรงกันว่า คนที่ติดสุรา จะส่งผลให้หัวใจเต้นแรงกว่าปกติ และมีโอกาสเป็นโรคความดันโลหิตสูง ประมาณ 50%ของผู้ติดสุราทั้งหมด

1.6.2.8 กินอาหารเค็มสม่ำเสมอต่อเนื่อง

1.6.2.9 ขาดการออกกำลังกาย เพราะเป็นปัจจัยเสี่ยงต่อโรคอ้วนและโรคเบาหวาน

1.6.2.10 ผลข้างเคียงจากยาบางชนิด เช่น ยาในกลุ่มสเตียรอยด์

รายงานจากการสำรวจสุขภาพประชาชนไทยโดยการตรวจร่างกายครั้งที่ 4 พ.ศ.2551 – 2552 พบความชุกของโรคความดันโลหิตสูงในประชากรไทยอายุ 15 ปีขึ้นไปร้อยละ 21.4 โดยพบว่าร้อยละ 60 ในชาย และ ร้อยละ 40 ในหญิงไม่เคยได้รับการวินิจฉัยมาก่อน และ ร้อยละ 8–9 ได้รับการวินิจฉัยแต่ไม่ได้รับการรักษาทั้งนี้ จากการศึกษาในปี 2555 พบว่าร้อยละของผู้ป่วยที่สามารถควบคุมความดันโลหิตให้อยู่ในเกณฑ์ที่เหมาะสมประมาณร้อยละ 60 การที่ผู้ป่วยส่วนใหญ่ไม่สามารถควบคุมภาวะความดันโลหิต สูงให้อยู่ในเกณฑ์ ที่เหมาะสม ส่งผลให้เกิดภาวะแทรกซ้อนต่าง ๆ ตามมา ข้อมูล

จากสำนักนโยบายและยุทธศาสตร์ กระทรวงสาธารณสุข พบว่าใน ปี พ.ศ. 2551 อัตราการเข้าพักรักษาตัวในโรงพยาบาลของผู้ป่วยที่เป็นโรคความดันโลหิตสูงมีแนวโน้มเพิ่มมากขึ้นอย่างต่อเนื่อง ตั้งแต่ ปี พ.ศ. 2541 – 2551 โดยพบว่า มีอัตราเพิ่มขึ้นจาก 169.6 เป็น 760.5 ต่อแสนประชากร

จากข้อมูลรายงานผลการดำเนินงานประจำปี พ.ศ. 2558 ของสำนักโรคไม่ติดต่อ พบว่าร้อยละของผู้ที่ได้รับการตรวจความดันโลหิตสูงมีแนวโน้มที่เพิ่มขึ้นจากปี พ.ศ. 2553 ถึง ปี พ.ศ. 2558 ซึ่งกลุ่มผู้สูงอายุจะมีมากกว่ากลุ่มอายุอื่น ดังตาราง 3

ตาราง 3 ร้อยละผู้ที่ได้รับการตรวจความดันโลหิตสูงภายใน 1 ปีที่ผ่านมา (พ.ศ. 2553 และ พ.ศ. 2558)

กลุ่มอายุ	พ.ศ. 2553			พ.ศ. 2558		
	ชาย	หญิง	รวม	ชาย	หญิง	รวม
15–24	41.8	50.8	46.2	46.1	52.3	48.9
25–34	54.9	67.7	61.3	57.7	73.5	66
35–44	66.6	78.3	72.7	69.1	83	76.4
45–54	74.2	81	77.8	79.6	88.4	84.3
55–64	80.3	87.9	84.4	86.4	91.9	89.3
65–74	83.9	88.3	86.4	90	92.4	91.3
75–79	–	–	–	88	91.3	89.8
รวม	62.6	73.2	68	68.5	79.4	74.1

1.6.3 โรคหัวใจขาดเลือด

โรคหัวใจขาดเลือด (Ischemic Heart Disease) อยู่ในกลุ่มโรคหัวใจและหลอดเลือด (Cardiovascular Diseases; CVDs) เป็นกลุ่มโรคความผิดปกติของหัวใจและหลอดเลือด เป็นสาเหตุของการเสียชีวิตอันดับหนึ่งทั่วโลก โดยในปี 2008 พบว่า จำนวนผู้เสียชีวิตจากโรคหลอดเลือดหัวใจมีประมาณ 7.3 ล้านคน และโรคหลอดเลือดสมองประมาณ 6.2 ล้านคน (Geneva: World Health Organization, 2011)

สมภพ พระธานี (2562) ระบุว่าปัจจัยเสี่ยงให้เกิดโรคหัวใจขาดเลือดที่สำคัญ คือ

1.6.3.1 สูบบุหรี่

1.6.3.2 โรคเบาหวาน

1.6.3.3 โรคไขมันในเลือดสูง

1.6.3.4 โรคความดันเลือดสูง

1.6.3.5 บุคคลที่ทำงานมีความเครียดสูง

1.6.3.6 ผู้มีประวัติโรคหัวใจในครอบครัว

1.6.3.7 ผู้สูงอายุ (ผู้ชายมากกว่า 55 ปี, ผู้หญิงมากกว่า 65 ปี)

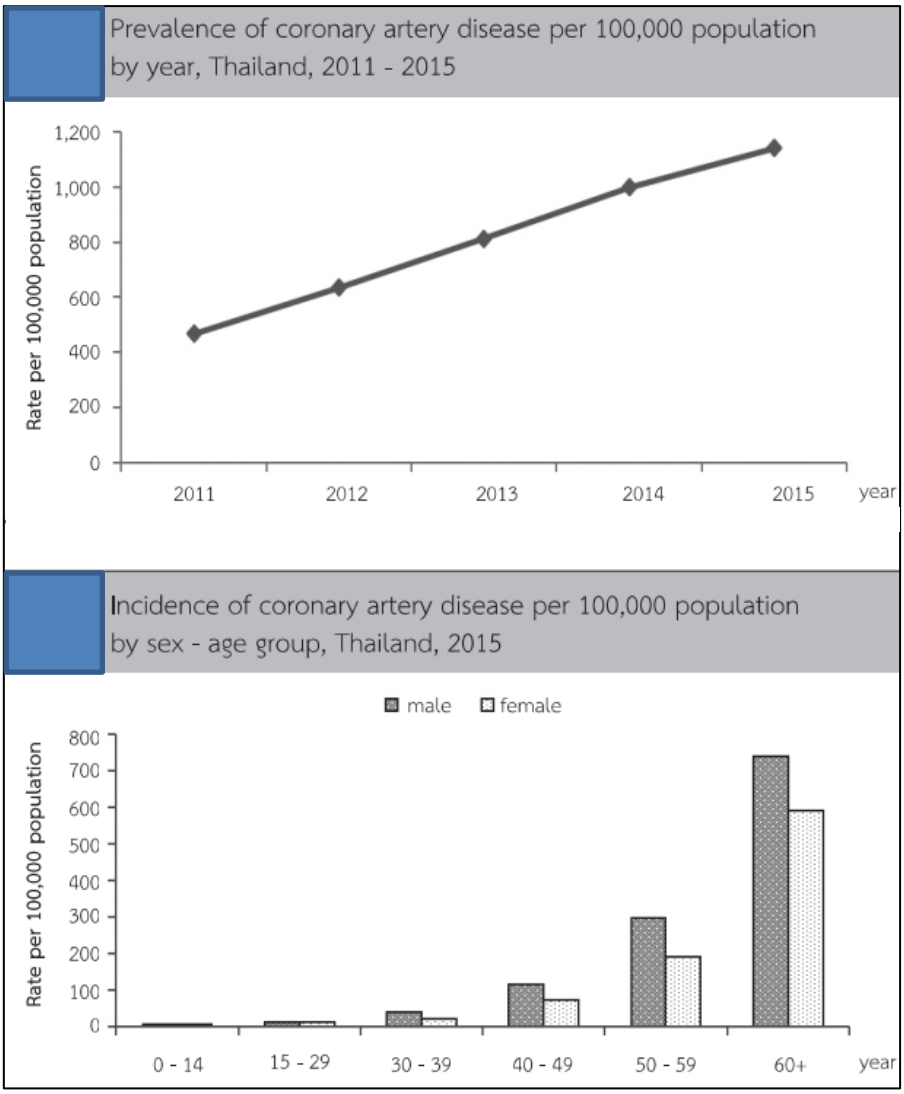
1.6.3.8 โรคอ้วน (ดัชนีมวลกายมากกว่า 30)

1.6.3.9 โรคไตเรื้อรัง

1.6.3.10 บุคคลที่ขาดการออกกำลังกาย

1.6.3.11 มีโรคเครียด

จากการสรุปรายงานการเฝ้าระวังโรคประจำปี พ.ศ. 2558 สถานการณ์โรคหลอดเลือดหัวใจในประเทศไทย จากข้อมูลศูนย์เทคโนโลยีสารสนเทศและการสื่อสาร สำนักงานปลัดกระทรวงสาธารณสุข พบว่ามีรายงานผู้ป่วยโรคหลอดเลือดหัวใจรายใหม่ในปีจำนวน 98,148 ราย คิดเป็น 150.01 ต่อประชากรแสนคนซึ่งสูงขึ้นทุกปีและพบว่าผู้ป่วยอายุมากกว่า 60 ปี มีการเกิดโรคสูงที่สุด รองลงมาคือกลุ่ม 50 – 59 ปี และ 40 – 49 ปี ตามลำดับ ดังภาพประกอบ 2



ภาพประกอบ 2 แนวโน้มการเกิดโรคหลอดเลือดหัวใจและจำแนกกลุ่มอายุ

1.7 ปัจจัยที่มีผลต่อการเกิดโรคไม่ติดต่อเรื้อรัง

โรคเรื้อไม่ติดต่อรังเป็นสาเหตุของการเจ็บป่วยและการเสียชีวิตของวัยผู้ใหญ่ทั่วโลก ทั้งนี้ โรคไม่ติดต่อเรื้อรังที่สำคัญ ได้แก่ โรคเบาหวาน โรคความดันโลหิตสูง โรคหัวใจและหลอดเลือด โรคมะเร็ง และโรคระบบทางเดินหายใจเรื้อรัง (โรคปอดอุดกั้นเรื้อรังและโรคหอบหืด) โรคต่าง ๆ ดังกล่าว มักมีสาเหตุจากปัจจัยเสี่ยงร่วมหนึ่งปัจจัยหรือมากกว่าได้แก่ การสูบบุหรี่ การบริโภคอาหารที่ไม่เหมาะสมตามหลักโภชนาการ การขาดการออกกำลังกาย และความเครียด เป็นต้น (กระทรวงสาธารณสุข, 2557)

สถาบันวิจัยและประเมินเทคโนโลยีทางการแพทย์ (2557) ได้ทบทวนวรรณกรรมสถานการณ์ปัจจุบัน และรูปแบบการบริการด้านโรคไม่ติดต่อเรื้อรัง พบว่า ปัจจัยที่มีผลต่อการเกิดโรคไม่ติดต่อเรื้อรังแบ่งเป็น ปัจจัยที่ไม่สามารถเปลี่ยนแปลงได้ (Non-modifiable Risk Factors) เช่น อายุ เพศ และพันธุกรรม และปัจจัยที่เปลี่ยนแปลงได้ (Modifiable Risk Factors) เช่น การสูบบุหรี่ การรับประทานอาหารที่ไม่เหมาะสม การขาดการออกกำลังกาย และการดื่มแอลกอฮอล์ ซึ่งปัจจัยเหล่านี้ ทำให้เกิดภาวะน้ำหนักเกินและโรคอ้วน ความดันโลหิตสูง น้ำตาลในเลือดสูง และภาวะไขมันในเลือดผิดปกติ ซึ่งนำ ไปสู่การเกิดโรคต่าง ๆ (WHO, 2005)

การสูบบุหรี่ เป็นปัจจัยหนึ่งที่สามารถเปลี่ยนแปลงได้ ซึ่งเป็นปัจจัยเสี่ยงที่สำคัญของการทำงานของระบบต่าง ๆ ภายในร่างกาย เช่น ระบบไหลเวียนโลหิต มีโอกาสเสี่ยงต่อโรคหัวใจและหลอดเลือดในสมอง มีผลต่อระบบทางเดินหายใจ ซึ่งเป็นสาเหตุสำคัญของการเกิดภาวะหลอดลมอักเสบเรื้อรัง โรคถุงลมโป่งพอง การอักเสบของเนื้อเยื่อภายในปอด และต่อระบบประสาท (สหดล ปุญญาถาวร, 2550, หน้า 10-11) ซึ่งการวิจัยทางการแพทย์พบว่า ในบุหรี่มีสารเคมีมากกว่า 7,000 ชนิด และเป็นสารพิษมากกว่า 250 ชนิด เช่น ทาร์ นิโคติน ฟอร์มาดิไฮด์ แอมโมเนีย สารหนู เป็นต้น อีกทั้งสารก่อมะเร็งมากกว่า 70 ชนิด (ประกิต วาทีสาธกกิจ, 2555)

ภาพประกอบ 3 ปัจจัยที่มีผลต่อการเกิดโรคไม่ติดต่อเรื้อรัง
(กรมการแพทย์ กระทรวงสาธารณสุข, 2557)

1.7.1 ปัจจัยเสี่ยงขั้นกลาง (Intermediate Risk Factors)

ปัจจัยบ่งชี้ที่นำไปสู่การวิเคราะห์การเกิดโรคไม่ติดต่อเรื้อรัง ซึ่งอาศัยการดูค่าดังนี้

1.7.1.1 ค่าดัชนีมวลกาย (BMI: Body Mass Index) เป็นค่าที่อาศัยความสัมพันธ์ระหว่างน้ำหนักตัวและส่วนสูงมาเป็นตัวชี้วัดสภาวะของร่างกายว่ามีความสมดุลของน้ำหนักตัวต่อส่วนสูงอยู่ในเกณฑ์ที่เหมาะสมหรือไม่ ตัวเลขดัชนีมวลกาย ถูกใช้เป็นเครื่องมือคัดกรองเพื่อระบุผู้ที่อาจมีน้ำหนักเกินหรือเป็นโรคอ้วน ซึ่งค่าดัชนีมวลกายสามารถคำนวณได้โดย นำน้ำหนักตัว (หน่วยเป็นกิโลกรัม) หารด้วย ส่วนสูงยกกำลังสอง (หน่วยเป็นเมตร) (กรมอนามัย, 2543) (ACSM, 2001) โดยสามารถแปลผลค่าดัชนีมวลกายดังตาราง 4

ตาราง 4 การประเมินค่าดัชนีมวลกาย

BMI มาตรฐานสากล(ยุโรป)	BMI มาตรฐานอาเซียน(เอเชีย)	การแปรผล
< 18.5	< 18.5	น้อยกว่ามาตรฐาน
18.5–24.9	18.5–22.9	ปกติ
25–29.9	23–24.9	อ้วนระดับ 1
30–34.9	25–29.9	อ้วนระดับ 2
35–39.9	มากกว่าหรือเท่ากับ 30	อ้วนระดับ 3
มากกว่าหรือเท่ากับ 40	–	อ้วนระดับ 4

1.7.1.2 ความดันโลหิต หรือ ความดันเลือด (Blood Pressure) คือ ความดันในหลอดเลือดเมื่อหัวใจบีบตัวสูบฉีดเลือดเข้าสู่หลอดเลือด ซึ่งเรียกว่า ความดันโลหิตซีสโตลิค (Systolic Blood Pressure) และเมื่อหัวใจพักคลายตัว ซึ่งเรียกว่า ความดันโลหิตไดแอสโตลิค (Diastolic Blood Pressure) ดังนั้น การรายงานผลความดันโลหิต จึงประกอบด้วยตัวเลข 2 ตัวเสมอ โดยจะบันทึกความดันซีสโตลิกเป็นตัวแรก หรือ ตัวบน ส่วนความดันไดแอสโตลิกจะบันทึกเป็นตัวตาม หรือ ตัวล่าง เช่น วัดความดันโลหิตได้ 120/80 หมายความว่า ความดันซีสโตลิค คือ 120 ส่วนความดันไดแอสโตลิค คือ 80 ค่าของความดันโลหิตสามารถแปลผล (WHO, 2003) ได้ดังนี้

ตาราง 5 การแปลค่าระดับความดันโลหิต

ค่าบน (ความดันซีสโตลิก)	ค่าล่าง (ความดันไดแอสโตลิก)	ระดับความดันโลหิต
160 ขึ้นไป	100 ขึ้นไป	ระดับอันตราย
140–159	90–99	สูงมาก
121–139	80–89	ค่อนข้างสูง
น้อยกว่าหรือเท่ากับ 120	น้อยกว่าหรือเท่ากับ 80	ปกติ

ความดันโลหิตจึงจัดเป็นหนึ่งในสัญญาณชีพที่สำคัญ (ความดันโลหิต อัตราการหายใจ ชีพจร และอุณหภูมิของร่างกาย) ซึ่งสามารถบอกถึงสุขภาพ และ

โรคต่าง ๆ ได้ โดยเฉพาะเป็นความ สำคัญเบื้องต้นที่บอกถึง โรคความดันโลหิตสูง การทำงานของหัวใจ และโรคหัวใจ (Madhur, M. et al., 2014)

1.7.1.3 ระดับน้ำตาลในเลือด (Fasting Plasma Glucose: FPG) หรือ (Fasting Blood Sugar: FBS) ใช้วัดระดับน้ำตาล (กลูโคส) ในเลือดหรือพลาสม่าโดยเจาะเลือดหลังจากงดอาหาร 2 ถึง 14 ชั่วโมง อาจอนุญาตให้ดื่มน้ำได้แต่งดเครื่องดื่มทุกชนิด เป็นการตรวจที่นิยมใช้ในการคัดกรองของโรคเบาหวานที่ไม่มีหรือพร่องอินซูลินแล้วมีระดับกลูโคสสูงอย่างถาวร เป็นวิธีที่ใช้มากในโรงพยาบาลเพราะสะดวกและมีความแม่นยำในการวินิจฉัยว่าผู้ป่วยเป็นโรคเบาหวาน เมื่อระดับน้ำตาลในเลือด สูงกว่า 126 mg/dl จากการเจาะเลือดทั้ง 2 ครั้ง (ศูนย์รวมคำแนะนำและข้อมูลผู้เชี่ยวชาญด้านสุขภาพระดับสากล, 2559) โดยสามารถแปลค่าระดับน้ำตาลดังตาราง 6

ตาราง 6 การแปลค่าระดับน้ำตาล

ระดับน้ำตาลก่อนอาหาร (mg/dl)	แปลระดับน้ำตาล
ตั้งแต่ 126 ขึ้นไป	เป็นเบาหวาน
100 – 125	ภาวะเตือนเบาหวาน
70 – 99	ภาวะปกติ

1.7.1.4 ไขมันในเลือดผิดปกติ หมายถึงร่างกายมีไขมันในกระแสเลือดสูง เช่น ระดับคอเลสเตอรอล (Cholesterol) หรือ ไตรกลีเซอไรด์ (Triglyceride) ไขมันสูงจะเพิ่มความเสี่ยงต่อการเกิดโรคกล้ามเนื้อหัวใจขาดเลือด โรคอัมพาต เนื่องจากไขมันสูงจะตกตะกอนที่ผิวของผนังหลอดเลือดที่เรียกว่าคราบไขมันหรือพลัค (Plaque) ซึ่งจะทำให้หลอดเลือดตีบ หรือคราบอาจจะหลุดลอยไปอุดหลอดเลือดทำให้เกิดโรค (Fredrickson DS, Lees RS, 1965; JAMA, 2001) โดยสามารถแปลค่าระดับไตรกลีเซอไรด์ดังตาราง 7

ตาราง 7 การแปลค่าระดับไตรกลีเซอไรด์

ค่าไตรกลีเซอไรด์ (mg/dl)	ระดับไตรกลีเซอไรด์
น้อยกว่า 150	ปกติ
150 – 199	สูงก้ำกึ่ง
200 – 499	สูง
มากกว่า 499	สูงมาก

จากการศึกษาปัจจัยเสี่ยงของการเกิดโรคไม่ติดต่อเรื้อรัง ทำให้ผู้วิจัยมีแนวคิดในการนำข้อมูลผู้สูงอายุจากประวัติการรับบริการได้แก่ อายุ เพศ การสูบบุหรี่ ดัชนีมวลกาย ความดันโลหิต น้ำตาลในเลือด และไขมันในเลือด ให้สามารถอยู่ในรูปแบบที่นำไปประยุกต์ใช้ในเหมืองข้อมูลได้

2. เหมืองข้อมูล

แนวคิดทฤษฎีเกี่ยวกับเหมืองข้อมูลประกอบไปด้วยหัวข้อต่าง ๆ ดังนี้

2.1 ความหมายของเหมืองข้อมูล

อดุลย์ ยิ้มงาม (2554) ให้ความหมายของเหมืองข้อมูลคือ กระบวนการที่กระทำกับข้อมูลจำนวนมากเพื่อค้นหารูปแบบและความสัมพันธ์ ที่ซ่อนอยู่ในชุดข้อมูลนั้น ในปัจจุบันการทำเหมืองข้อมูลได้ถูกนำไปประยุกต์ใช้ในงานหลายประเภท ทั้งในด้านธุรกิจที่ช่วยในการตัดสินใจของผู้บริหาร ในด้านวิทยาศาสตร์และการแพทย์รวมทั้งในด้านเศรษฐกิจและสังคม

David Hand, Heikki Mannila และ Padhraic Smyth (2001) ให้ความหมายของเหมืองข้อมูลคือ การวิเคราะห์ข้อมูลจำนวนมากเพื่อหาความสัมพันธ์และการสรุปผลของข้อมูลซึ่งสามารถเข้าใจได้และเป็นประโยชน์ต่อผู้ทำการเก็บรวบรวมข้อมูล

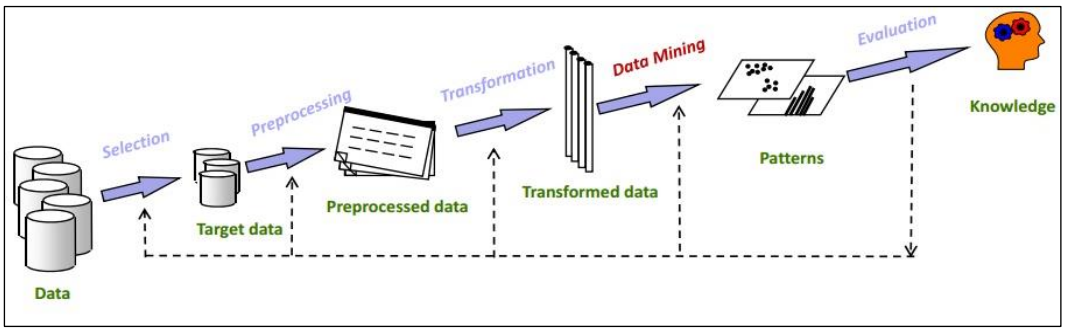
Peter Cabena, Pablo Hadjinian, Rolf Stadler, Jaap Verhees และ Alessandro Zanasi (1998) ให้ความหมายของเหมืองข้อมูลคือ การนำวิธีการจากการเรียนรู้ของเครื่อง การรู้จำรูปแบบ สถิติ และฐานข้อมูล เพื่อสกัดข้อมูลที่มีประโยชน์จากฐานข้อมูลขนาดใหญ่ได้

จากความหมายของเหมืองข้อมูลที่ได้ศึกษาและรวบรวมมานั้นผู้วิจัยสังเคราะห์และสรุปความหมายของเหมืองข้อมูล คือ กระบวนการสกัด หรือ ค้นหาสารสนเทศ เพื่อให้ได้ความรู้ หรือสารสนเทศบางมุมที่ซ่อนเร้นอยู่ในฐานข้อมูลขนาดใหญ่ โดยอาศัยข้อมูลในอดีตเป็นจำนวนมาก เพื่อนำความรู้ที่ได้ไปช่วยในการวิเคราะห์ และประกอบการตัดสินใจของผู้บริหารในธุรกิจด้านต่าง ๆ ในปัจจุบันเหมืองข้อมูลได้ถูกนำไปประยุกต์ใช้ในงานหลายประเภททั้งในด้านธุรกิจที่ช่วยในการตัดสินใจของผู้บริหาร ในด้านวิทยาศาสตร์และการแพทย์รวมทั้งในด้านเศรษฐกิจและสังคม

2.2 กระบวนการเหมืองข้อมูล

จากการศึกษากระบวนการเหมืองข้อมูล ผู้วิจัยพบว่ามีกระบวนการเหมืองข้อมูลมีรูปแบบตามความเหมาะสมของงาน อันได้แก่ กระบวนการค้นหาความรู้ในฐานข้อมูล และ กระบวนการคริสป์-ดีเอ็ม ซึ่งมีขั้นตอนการดำเนินการแตกต่างกันดังต่อไปนี้

2.2.1 การค้นหาความรู้ในฐานข้อมูล (Knowledge Discovery in Databases: KDD) เป็นกระบวนการต้นแบบในการค้นหาความรู้ในฐานข้อมูล จากข้อมูลจำนวนมหาศาลโดยอัตโนมัติ โดยใช้ขั้นตอนวิธีจากวิชาสถิติ การเรียนรู้ของเครื่อง และการรู้จำแบบ โดยประกอบไปด้วยขั้นตอนดังนี้ (Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P., 1996)



ภาพประกอบ 4 ขั้นตอนการค้นหาความรู้ในฐานข้อมูล

2.2.1.1 คัดเลือกข้อมูล (Selection) เป็นการคัดเลือกข้อมูลที่เกี่ยวข้องกับเป้าหมาย หรือวัตถุประสงค์ที่ได้ตั้งไว้ เพื่อให้ผลการวิเคราะห์ข้อมูลหรือผลของการทำกระบวนการเหมืองข้อมูลสามารถนำไปใช้ในการประกอบการตัดสินใจได้ตามความต้องการ โดยสามารถเลือกจากแหล่งข้อมูลต่าง ๆ ทั้งที่อยู่ภายใน และภายนอกองค์กร

ซึ่งแหล่งข้อมูลหลักของเหมืองข้อมูล คือ คลังข้อมูล (Data Warehouse) หรือฐานข้อมูล (Database)

2.2.1.2 ขั้นตอนก่อนกระบวนการ (Preprocessing) เป็นขั้นตอนการปรับข้อมูลที่ได้คัดเลือกมานั้น ให้มีคุณภาพมากยิ่งขึ้น เนื่องจากการคัดเลือกข้อมูลจากหลากหลายแหล่งข้อมูล อาจจะทำให้มีรูปแบบ (Format) ในการเก็บข้อมูลไม่เหมือนกัน ส่งผลให้เกิดการผิดพลาดในการทำกระบวนการเหมืองข้อมูลได้ ซึ่งการปรับในขั้นตอนนี้ประกอบไปด้วยกระบวนการต่าง ๆ ดังนี้

1) ทำความสะอาดข้อมูล (Data Cleaning) เป็นกระบวนการทำความสะอาดข้อมูล โดยการเติมข้อมูลที่ขาดหายไป (Missing) ในบาง Field รวมทั้งเลือกข้อมูลที่จำเป็นต่อเหมืองข้อมูล เช่น ถ้าต้องการที่จะวิเคราะห์ความต้องการของลูกค้า ควรจะเลือกข้อมูลที่สำคัญต่อการวิเคราะห์ ซึ่งอาจจะตัดข้อมูลในส่วนรายละเอียดที่ไม่จำเป็น โดยทั่วไปจะเป็นลักษณะข้อมูลที่ระบุความเป็นหนึ่งของสิ่งต่าง ๆ เช่น เลขที่บัตรประชาชน เบอร์โทรศัพท์ บ้านเลขที่ เป็นต้น นอกจากนี้ยังจัดการกับความผิดปกติของข้อมูล และจัดการกับข้อมูลที่ไม่สอดคล้อง เช่น อายุ ไม่ตรงกับปีที่เกิด, วันที่มีเดือนที่ 13, เพศหญิงแต่มีคำนำหน้าว่า นาย เป็นต้น

2) การรวมกันของข้อมูล (Data Integration) เป็นกระบวนการรวมข้อมูล โดยต้องมีการตรวจสอบก่อนว่าการรวบรวมข้อมูลจากแหล่งข้อมูลต่าง ๆ นั้น มีรูปแบบในการจัดเก็บที่เหมือนกันหรือต่างกัน เช่น ข้อมูลที่มีรูปแบบการจัดเก็บต่างกัน หรืออาจจะมีหน่วยวัดที่ต่างกัน สัญลักษณ์ที่ใช้ไม่เหมือนกัน การกำหนดประเภทของข้อมูลที่ต่างกัน รวมไปถึงลดความซ้ำซ้อนของข้อมูลด้วย

3) การลดทอนข้อมูล (Data Reduction) เป็นกระบวนการในการลดขนาดของข้อมูล ซึ่งสามารถทำได้ 2 แนวทางคือ การลดขนาดข้อมูล (Data Size Reduction) ใช้แถว (Row) เป็นหลักในการลดข้อมูล และ การลดมิติข้อมูล (Dimensionality Reduction) ใช้ตัวแปร (Attribute) เป็นหลักในการลดข้อมูล เนื่องจากในเทคนิคบางอย่างของกระบวนการเหมืองข้อมูล จะต้องอาศัยข้อมูลในเรื่องต่าง ๆ เป็นจำนวนที่เท่า ๆ กัน เพื่อความเที่ยงตรง และความถูกต้องของผลเหมืองข้อมูล

2.2.1.3 แปลงข้อมูล (Transformation) เป็นขั้นตอนในการแปลงข้อมูลเพื่อให้ใช้กับแบบจำลองต่าง ๆ ของกระบวนการเหมืองข้อมูลได้ ซึ่งวิธีการแปลงข้อมูลในรูปแบบต่าง ๆ ดังนี้

1) นอร์มัลไลเซชัน (Normalization) คือการแปลงข้อมูลให้อยู่ในช่วงค่าที่กำหนด จัดได้ว่าเป็นการลดระยะห่างของข้อมูล เนื่องจากข้อมูลชนิดเดียวกัน แต่มีระยะห่างมากเกินไป ซึ่งทำให้มีผลต่อการวิเคราะห์ข้อมูลในบางแบบจำลองของเหมืองข้อมูล เช่น ข้อมูลที่มีค่าช่วง 0–1,000,000 แปลงให้เป็นช่วง 0–20 เป็นต้น

2) การแบ่งค่า (Discretization) คือกระบวนการแปลงข้อมูลที่เป็นจำนวนจริงต่อเนื่อง ให้เป็นช่วงข้อมูล จัดได้ว่าเป็นการกำหนดกลุ่มให้ข้อมูล ลดการกระจายของข้อมูล เช่น กำหนดกลุ่มให้คะแนนของนักศึกษา อายุของลูกค้า แปลงให้เป็นช่วงข้อมูล เช่น ช่วงอายุ 20–25 ปี, ช่วง 26–30 ปี, ช่วง 31–35 ปี ตามลำดับ

3) กำหนดรหัส (1 Of N Coding) คือกระบวนการแปลงข้อมูลจากตัวอักษรให้เป็นตัวเลข ที่ไม่ได้บอกถึงลำดับความมากน้อย เช่น การกำหนดรหัสให้กับข้อมูล เป็นต้น

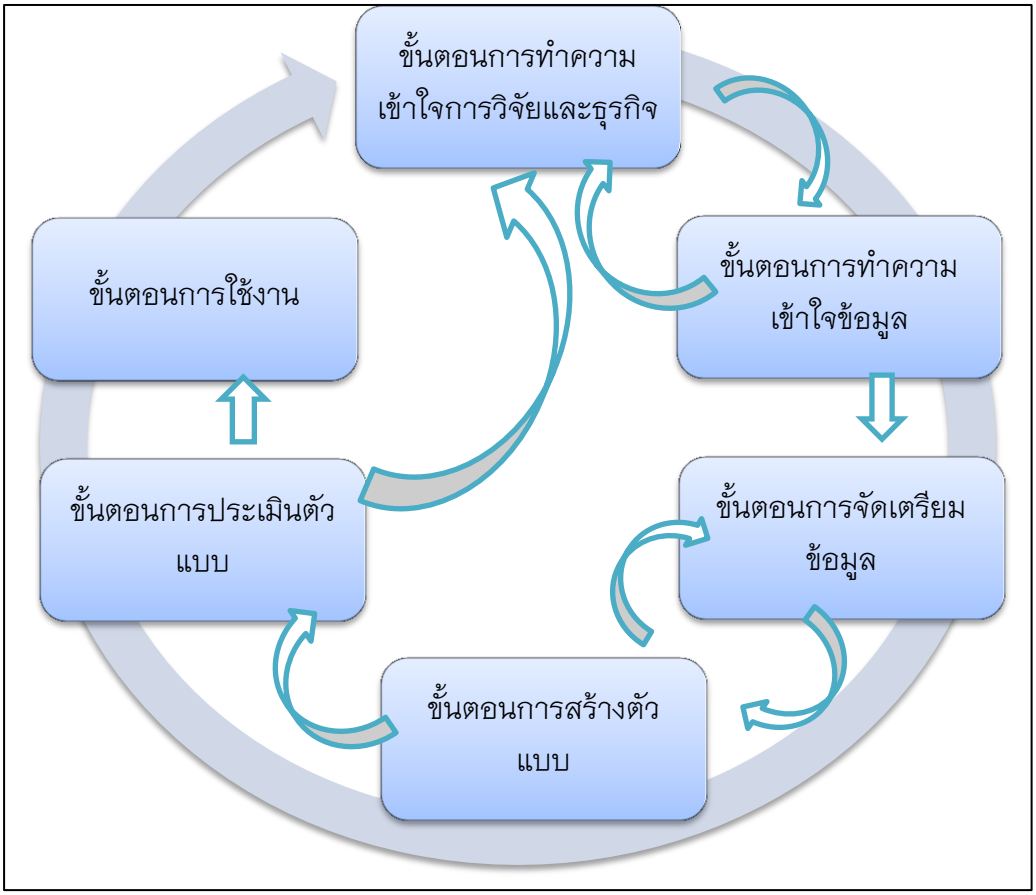
4) ลงสรุปเป็นนัยทั่วไป (Generalization) คือกระบวนการแปลงข้อมูลที่มีรายละเอียดมากเกินไป ให้มีความกระชับมากยิ่งขึ้น เช่น ที่อยู่ของลูกค้า อาจจะทำการแปลงข้อมูลให้กระชับขึ้นเป็น อยู่ในกทม. หรืออยู่ต่างจังหวัด เป็นต้น

2.2.1.4 เหมืองข้อมูล (Data Mining) เป็นขั้นตอนเหมืองข้อมูล โดยมีเทคนิค หรือการทำงาน (Operation) ในเหมืองข้อมูลหลายรูปแบบ เช่น การจัดกลุ่มข้อมูล (Database Segmentation/Clustering), การจำแนกข้อมูล (Data Classification) และความสัมพันธ์ (Link Analysis/Association) เป็นต้น ซึ่งแต่ละเทคนิคนั้น จะมีอัลกอริทึม (Algorithm) ให้เลือกใช้ ซึ่งจะกล่าวถึงในหัวข้อถัดไป เรื่อง เทคนิคเหมืองข้อมูล

2.2.1.5 ประเมินผล (Interpretation/Evaluation/Analysis Of Result and Knowledge Presentation) เป็นขั้นตอนสุดท้ายสำหรับนักวิเคราะห์ข้อมูล ที่จะต้องเก็บผลลัพธ์ของกระบวนการเหมืองข้อมูลมาทำการสรุปความหมายหรือตีความหมายของผลลัพธ์ที่ได้ออกมาเป็นข้อมูลความรู้ใหม่ (Knowledge) ที่สามารถนำไปเป็นสารสนเทศที่ช่วยในการตัดสินใจของผู้บริหารต่อไปได้

2.2.2 คริสป์–ดีเอ็ม (CRISP–DM) หรือ “Cross–Industry Standard Process for Data Mining” เป็นกระบวนการมาตรฐานในการวิเคราะห์ข้อมูลด้านเหมืองข้อมูล ถูกพัฒนาขึ้นในปี ค.ศ. 1996 โดยความร่วมมือกันของ 3 บริษัท คือ เดมเลอร์ไครสเลอร์ (DaimlerChrysler) เอสพีเอสเอส (SPSS) และเอ็นซีอาร์ (NCR) ซึ่งถูกพัฒนามาจากกระบวนการการค้นหาความรู้ในฐานข้อมูล เป็นกระบวนการมาตรฐานที่เป็น

ประโยชน์สำหรับทำให้เหมืองข้อมูล เหมาะสมในการแก้ปัญหาโดยทั่วไปของธุรกิจหรือการวิจัย (Chapman, P. and et al., 2000) โดยประกอบด้วย 6 ขั้นตอนได้แก่



ภาพประกอบ 5 ขั้นตอนคริสป์-ดีเอ็ม

2.2.2.1 ขั้นตอนการทำความเข้าใจการวิจัยและธุรกิจ (Business Understanding) เป็นขั้นตอนแรกในกระบวนการคริสป์-ดีเอ็มซึ่งเน้นไปที่การเข้าใจปัญหาและแปลงปัญหาที่ได้ให้อยู่รูปโจทย์ของการวิเคราะห์ข้อมูลทำเหมืองข้อมูล พร้อมทั้งวางแผนในการดำเนินการคร่าว ๆ

2.2.2.2 ขั้นตอนการทำความเข้าใจข้อมูล (Data Understanding) เป็นขั้นตอนเริ่มเก็บรวบรวมข้อมูล หลังจากนั้นจะเป็นการตรวจสอบข้อมูลที่ได้ทำการรวบรวมมา เพื่อดูความถูกต้องของข้อมูล และพิจารณาว่าจะใช้ข้อมูลทั้งหมดหรือจำเป็นต้องเลือกข้อมูลบางส่วนมาใช้ในการวิเคราะห์

2.2.2.3 ขั้นตอนการจัดเตรียมข้อมูล (Data Preparation) เป็นขั้นตอนที่ทำการแปลงข้อมูลที่ได้ทำการเก็บรวบรวมมา (Raw Data) ให้กลายเป็นข้อมูลที่สามารถนำไปวิเคราะห์ในขั้นถัดไปได้ โดยการแปลงข้อมูลนี้อาจจะต้องมีการทำให้ข้อมูลถูกต้อง (Data Cleaning) เช่น การแปลงข้อมูลให้อยู่ในช่วง (Scale) เดียวกัน หรือการเติมข้อมูลที่ขาดหายไป เป็นต้น โดยขั้นตอนนี้จะเป็นขั้นตอนที่ใช้เวลามากที่สุดของกระบวนการคริสป์–ดีเอ็ม

2.2.2.4 ขั้นตอนการสร้างตัวแบบ (Modeling) เป็นขั้นตอนการวิเคราะห์ข้อมูลด้วยเทคนิคทางเหมืองข้อมูล ซึ่งในขั้นตอนนี้หลายเทคนิคจะถูกนำมาใช้เพื่อให้ได้คำตอบที่ดีที่สุด ดังนั้นในบางครั้งอาจจะต้องมีการย้อนกลับไปที่ขั้นตอนที่ 3 ขั้นตอนการจัดเตรียมข้อมูล เพื่อแปลงข้อมูลบางส่วนให้เหมาะสมกับแต่ละเทคนิค

2.2.2.5 ขั้นตอนการประเมินตัวแบบ (Evaluation) เป็นขั้นตอนที่ได้ผลการวิเคราะห์ข้อมูลด้วยเทคนิคเหมืองข้อมูลแล้ว แต่ก่อนที่จะนำผลลัพธ์ที่ได้ไปใช้งานต่อไปจะต้องมีการวัดประสิทธิภาพของผลลัพธ์ที่ได้ ว่าตรงกับวัตถุประสงค์ที่ได้ตั้งไว้ในขั้นตอนแรก หรือมีความน่าเชื่อถือมากน้อยเพียงใด ซึ่งอาจย้อนกลับไปยังขั้นตอนก่อนหน้าเพื่อเปลี่ยนแปลงแก้ไขให้ได้ผลลัพธ์ตามที่ต้องการได้

2.2.2.6 ขั้นตอนการใช้งาน (Deployment) ในกระบวนการทำงานของคริสป์–ดีเอ็มไม่ได้หยุดเพียงแค่ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยเทคนิคเหมืองข้อมูลเท่านั้น แม้ว่าผลลัพธ์ที่ได้จะแสดงถึงองค์ความรู้ที่เป็นประโยชน์ แต่จะต้องนำความรู้เหล่านี้ไปใช้จริงในองค์กร หรือบริษัท เช่น การสร้างรายงานเพื่อให้ผู้บริหารหรือนักการตลาดเข้าใจง่ายและสามารถนำไปออกโปรโมชั่นได้ เป็นต้น

จากกระบวนการดังกล่าว ผู้วิจัยเลือกใช้ กระบวนการคริสป์–ดีเอ็ม เนื่องจากเป็นกระบวนการที่ถูกพัฒนามาจากกระบวนการค้นหาความรู้ในฐานข้อมูล ซึ่งจะมีมาตรฐานเหมาะสมในการแก้ปัญหาทั่วไป มีความยืดหยุ่นและสอดคล้องกับงานในปัจจุบัน และใช้เพียง 5 ขั้นตอน เนื่องจาก ขั้นตอนที่ 6 ขั้นตอนการใช้งานจะเป็นการนำไปใช้ขององค์กรโรงพยาบาล แพทย์และพยาบาลกับผู้ป่วยต่อไปซึ่งจะเป็นวิชาชีพเฉพาะ

2.3 ประเภทของข้อมูลที่ใช้ในกระบวนการเหมืองข้อมูล

สายชล สินสมบูรณ์ทอง (2560) ได้แบ่งประเภทของข้อมูลที่สามารถทำเหมืองข้อมูลได้ 4 ประเภทดังนี้

2.3.1 ระบบฐานข้อมูลเชิงสัมพันธ์ (Relational Database) เป็นฐานข้อมูลที่จัดเก็บอยู่ในรูปแบบของตาราง โดยในแต่ละตารางจะประกอบไปด้วยแถวและคอลัมน์ ความสัมพันธ์ของข้อมูลทั้งหมดสามารถแสดงได้โดยความสัมพันธ์ระหว่างข้อมูล (Entity–Relationship Model: ER Model)

2.3.2 คลังข้อมูล (Data Warehouses) เป็นการเก็บรวบรวมข้อมูลจากหลายแหล่งมาเก็บไว้ในรูปแบบเดียวกันและรวบรวมไว้ในที่เดียวกัน

2.3.3 ฐานข้อมูลการดำเนินการธุรกรรม (Transactional Database) ประกอบด้วยข้อมูลที่แต่ละธุรกรรม (Transaction) แทนด้วยเหตุการณ์ในขณะใดขณะหนึ่ง เช่น ใบเสร็จรับเงิน จะเก็บข้อมูลในรูปแบบ ชื่อลูกค้าและรายการสินค้าที่ลูกค้ารายนั้นซื้อ เป็นต้น

2.3.4 ฐานข้อมูลรูปแบบอื่น (Advanced Database) เป็นฐานข้อมูลที่จัดเก็บในรูปแบบอื่น เช่น ข้อมูลแบบเชิงวัตถุ (Object–Oriented) ข้อมูลที่เป็นไฟล์ข้อความ (Text File) ข้อมูลมัลติมีเดีย ข้อมูลในรูปของเว็บ เป็นต้น

2.4 เทคนิคเหมืองข้อมูล

เทคนิคเหมืองข้อมูล ในปัจจุบันมีหลายรูปแบบ โดยผู้วิจัยได้ทำการศึกษาค้นคว้า เกี่ยวกับเทคนิคเหมืองข้อมูลจากแหล่งข้อมูลต่าง ๆ ซึ่งเหมืองข้อมูลที่ได้รับความนิยม มี 2 เทคนิคใหญ่ ๆ (Giudici, 2004) ดังนี้

2.4.1 การเรียนรู้แบบไม่มีผู้สอน (Descriptive Techniques หรือ Unsupervised Learning) เป็นวิธีสำหรับอธิบายลักษณะของการจัดกลุ่มข้อมูล (Cluster) ในกรณีที่ไม่ทราบตัวแปรเป้าหมาย (Target Variable) และไม่ทราบจำนวนกลุ่มข้อมูล แต่ต้องการจัดข้อมูลลงในกลุ่ม ให้ข้อมูลในกลุ่มเดียวกันมีลักษณะคล้ายคลึงกันหรือมีความสัมพันธ์กันมากที่สุด โดยผู้ศึกษาไม่จำเป็นต้องมีความรู้เกี่ยวกับกลุ่มข้อมูลล่วงหน้า เทคนิคนี้สามารถใช้วิธีการวิเคราะห์ได้ 3 วิธีดังนี้

2.4.1.1 กฎความสัมพันธ์ (Association Rules) เป็นกฎของการค้นหากลุ่มความสัมพันธ์ของข้อมูลหรือสินค้า บริการที่ลูกค้าคนหนึ่งมีแนวโน้มจะซื้อหรือมาใช้บริการ ณ เวลาเดียวกันหรือต่างเวลากัน เช่นต้องการตรวจสอบหรือศึกษาเบื้องต้นว่าสินค้าชนิดใดควรจะขายควบคู่กัน โดยที่รูปแบบของกฎ (Rules) จะเป็นไปในทิศทางเดียวจากซ้ายไปขวาเท่านั้น ตัวอย่างในร้านอาหารแห่งหนึ่งอาจพบว่าลูกค้าสั่งเครื่องดื่มประเภทน้ำเปล่าและเป๊ปซี่ดังนี้

น้ำเปล่า –> เป๊ปซี่ มีความสัมพันธ์กันสูง

แต่เป๊ปซี่ –> น้ำเปล่า มีความสัมพันธ์ต่ำกว่า

ทั้งด้านซ้ายและด้านขวาของกฎความสัมพันธ์ อาจจะมีมากกว่าหนึ่งรายการ (Items) ก็ได้เช่น

น้ำเปล่า –> เป๊ปซี่, ข้าวโพดคั่ว, อ้อย

น้ำเปล่า –> แฟนต้า, ข้าวโพดคั่ว, อ้อย เป็นต้น

น้ำเปล่า –> เป๊ปซี่, มิรินด้า, แฟนต้า เป็นต้น

ในการพิจารณาว่าจะใช้กฎความสัมพันธ์ใด ผู้ศึกษาควรดูรายการสินค้าประกอบกับการอ่านค่าสนับสนุน (Support) ความเชื่อมั่น (Confidence) และค่าลิฟท์ (Lift) โดยค่าสนับสนุน เป็นค่าที่บอกถึงความถี่ในการเกิดความสัมพันธ์มากน้อยแค่ไหน ค่าความเชื่อมั่น เป็นค่าที่บอกถึงโอกาสที่เกิดความสัมพันธ์ที่ปรากฏมากน้อยแค่ไหน ดังนั้นในการพิจารณาเลือกกฎความสัมพันธ์ ควรเลือกกฎที่ให้ค่าสนับสนุน และค่าความเชื่อมั่นที่สูง ส่วนค่าลิฟท์ เป็นค่าวัดประสิทธิภาพของการเกิดกฎความสัมพันธ์มีค่าอยู่ระหว่าง 1 ถึง –1 ซึ่งเครื่องเป็นตัวบอกทิศทางความสัมพันธ์ ถ้าค่าลิฟท์มีค่าเท่ากับ 1 หมายความว่ากฎความสัมพันธ์ระหว่างกฎทางซ้ายกับกฎทางขวาเป็นอิสระกัน ซึ่งเป็นข้อที่ควรระมัดระวังของการวิเคราะห์ความสัมพันธ์ (Association Rule) ไม่ควรใช้ข้อมูลตัววัดตัวใดตัวหนึ่งเป็นหลัก ควรพิจารณาจากค่าวัดทั้งสามจะสามารถทำนายกฎความสัมพันธ์ของข้อมูลทั้งหมดที่ซ่อนอยู่ในฐานข้อมูลขนาดใหญ่ได้ และสามารถประยุกต์กฎที่ได้ไปใช้ในการจำแนกข้อมูลทั้งหมดที่ซ่อนอยู่ในฐานข้อมูลขนาดใหญ่ได้และสามารถประยุกต์กฎที่ได้ไปใช้ในการจำแนกข้อมูลสินค้าตามลักษณะความต้องการของลูกค้าให้อยู่ในกลุ่มเดียวกันได้ และจากการที่ต้องพิจารณาค่าวัดทั้งสามไปพร้อม ๆ กันทำให้วิธีนี้มีระดับความถูกต้องสูงโดยแสดงค่าเป็นร้อยละ (Percentage) (Berry, M.A. and Linoff, G.S., 2000)

2.4.1.2 การวิเคราะห์การจัดกลุ่ม (Cluster Analysis) เป็นการจัดกลุ่มข้อมูลจากข้อมูลที่มีอยู่โดยไม่มีการกำหนดกลุ่มให้แต่ละค่าสังเกตล่วงหน้า เทคนิคนี้จะทำหน้าที่ในการหาจำนวนกลุ่มที่เหมาะสมให้แก่ชุดข้อมูล ค่าสังเกตที่มีความคล้ายคลึงกันจะถูกจัดอยู่ในกลุ่มเดียวกัน การวิเคราะห์การจัดกลุ่มสามารถจัดกลุ่มได้จากตัวแปรเป้าหมาย หรือจากตัวแปรนำเข้าก็ได้

2.4.1.3 การพยากรณ์ด้วยอนุกรมเวลา (Time Series) เป็นการพยากรณ์ค่าในอนาคตด้วยอนุกรมเวลา โดยอาศัยข้อมูลในอดีต จากอิทธิพลของแนวโน้ม

(Trend) ฤดูกาล (Seasonal) วัฏจักร (Cycle) และเหตุการณ์ผิดปรกติ (Irregular) โดยเทคนิคนี้ค้นหาตัวแปรเป้าหมายจากข้อมูลนำเข้าที่เป็นลักษณะข้อมูลอนุกรมเวลา

2.4.2 การเรียนรู้แบบมีผู้สอน (Predictive Techniques หรือ Supervised Learning) เป็นการอธิบายความสัมพันธ์ของข้อมูลที่มีอยู่ด้วยการสร้างแบบแผนการเรียนรู้มีตัวแปรเป้าหมายอย่างชัดเจน กล่าวคือ ผู้วิเคราะห์จะทราบค่าของตัวแปรเป้าหมายอยู่ก่อนแล้ว เพื่อใช้ในการคำนวณค่าสัมประสิทธิ์ต่าง ๆ ของตัวแปรนำเข้า ตัวแบบที่ได้ จะถูกใช้ในการพยากรณ์ค่าตัวแปรเป้าหมายของค่าสังเกตใหม่ (New Observation) ส่วนวิธีวิเคราะห์ที่นำมาใช้จะขึ้นกับลักษณะของตัวแปร และมีการแบ่งข้อมูล (Partition Data) อย่างสุ่มออกเป็น 3 ส่วน คือ 1) ข้อมูลฝึกสอน (Training Data) เป็นส่วนแรกสำหรับใช้ในการสร้างตัวแบบ ปกติข้อมูลในส่วนนี้จะมีขนาดใหญ่สุด เช่น ร้อยละ 70 ของข้อมูลทั้งหมด 2) ข้อมูลตรวจสอบ (Validation Data) เป็นข้อมูลส่วนที่สองมีขนาดเล็กลงมา เช่น ร้อยละ 15 ของข้อมูลทั้งหมด ใช้สำหรับคัดเลือกตัวแบบ คือหลังจากที่ใช้ข้อมูลฝึกสอนสร้างตัวแบบที่เป็นไปได้หลายตัวแบบ จะใช้ข้อมูลรอบที่สอบทำการเปรียบเทียบประสิทธิภาพของตัวแบบ โดยตัวแบบที่มีค่าความผิดพลาด (Error) น้อยที่สุดจะเป็นตัวแบบที่ต้องการ และ 3) ข้อมูลทดสอบ (Test Data) เป็นข้อมูลในส่วนสุดท้ายที่เหลืออีก ร้อยละ 15 ของข้อมูลทั้งหมด ส่วนมากจะมีขนาดเท่ากับข้อมูลที่ตรวจสอบ ใช้สำหรับหาค่าความคลาดเคลื่อนของตัวแบบ แล้วจึงนำข้อมูลทั้งสามส่วนมาศึกษาวิเคราะห์ด้วยเทคนิคพยากรณ์ที่นิยมคือ ต้นไม้ตัดสินใจ (Decision Tree) และ โครงข่ายประสาทเทียม (Neural Networks) เป็นต้น (Roiger, R. J., and Geatz, M., 2003; Witten, I.H. and Frank, E., 2005)

ข้อแตกต่างระหว่างการเรียนรู้แบบมีผู้สอน และการเรียนรู้แบบไม่มีผู้สอน คือ โครงสร้างของตัวแบบ (Model Structure) ในการเรียนรู้แบบมีผู้สอนจะมีความสัมพันธ์ระหว่างตัวแปรนำเข้า (Input Variables) และตัวแปรเป้าหมาย ถูกกำหนดขึ้นและมีการสร้างข้อมูลฝึกสอนเพื่อพัฒนาตัวแบบการจำแนกและการคาดการณ์ข้อมูล แต่ในการเรียนรู้แบบไม่มีผู้สอน ไม่มีการกำหนดตัวแปรใดให้เป็นตัวแปรเป้าหมาย (Berry, M.A. and Linoff, G.S., 2000) สำหรับงานวิจัยในครั้งนี้ผู้วิจัยจะใช้เทคนิคเหมืองข้อมูลแบบการเรียนรู้แบบมีผู้สอน พัฒนาแบบจำลองเพื่อพยากรณ์โรคไม่ติดต่อเรื้อรังในผู้สูงอายุ

2.5 เหมืองข้อมูลกับสารสนเทศทางการแพทย์

ปัจจุบันเหมืองข้อมูลกับสารสนเทศทางสุขภาพหรือทางการแพทย์ (Health/Medical Informatics) นักวิจัยในเหมืองข้อมูลเริ่มให้ความสนใจมากขึ้น

โดยเหตุผลสำคัญคือแนวโน้มทางการแพทย์ที่เป็นส่วนบุคคลมากขึ้น (Personalized Medicine) ซึ่งเหมืองข้อมูลจะมุ่งเน้นไปที่การค้นหาโรค การรักษาและมาตรการป้องกัน ซึ่งข้อมูลที่สำคัญในเหมืองข้อมูลมาจากความก้าวหน้าทางเทคโนโลยีและการจัดเก็บข้อมูลที่มีมากขึ้น โดยแบ่งเหมืองข้อมูลทางการแพทย์ออกเป็น 4 ด้านดังนี้
(ปัณณวิชญ์ วงศ์วิวัฒนานนท์, 2553)

2.5.1 งานวินิจฉัยโรค เพื่อวิเคราะห์ว่าผู้ป่วยเจ็บป่วยจากการแพทย์อะไรบ้าง เช่นการวินิจฉัยผู้ป่วยมะเร็งปอดและช่องปากระยะแรกเป็นเรื่องยากหากกระทำโดยวิธีทางการแพทย์ตามปกติ การใช้ข้อมูลพันธุศาสตร์ช่วยได้มากในการวินิจฉัยที่รวดเร็วขึ้น รวมทั้งความแม่นยำในการวินิจฉัย

2.5.2 การพยากรณ์โรค เพื่อคาดคะเนว่าผู้ป่วยจะหายเจ็บป่วยได้ดีขนาดไหน และโรคจะดำเนินไปอย่างไรตามระยะเวลา เช่น การใช้ดัชนีชี้วัดทางชีวภาพ (Biomarker) ในการคาดคะเนว่าอวัยวะที่ปลูกถ่ายสามารถทนอยู่ในร่างผู้รับได้นานเท่าใด

2.5.3 การรักษาที่เหมาะสม เพื่อคาดคะเนผลการรักษาโรค เช่น การใช้ดัชนีชี้วัดทางชีวภาพ ในการคาดคะเนว่าการรักษาทางเคมีบำบัดได้ผลอย่างไร

2.5.4 การทำความเข้าใจเกี่ยวกับกลไกโรค เพื่อให้เกิดแนวคิดหรือความรู้ใหม่เกี่ยวกับสาเหตุการเกิดโรค เช่น การวิจัยเกี่ยวกับการถ่ายโอนสัญญาณ (Signaling Pathway) ในระหว่างการติดเชื้อไวรัส (Virus) เป็นต้น

จากเหมืองข้อมูลในปัจจุบันนั้น ได้มีการนำมาใช้ประโยชน์ในองค์กรกันมากขึ้น รวมถึงสถานพยาบาล โดยข้อมูลจำนวนมหาศาลทั้งที่เป็นข้อมูลบริการและข้อมูลทางคลินิก สามารถจัดเก็บขึ้นเป็นคลังข้อมูล ซึ่งสามารถเลือกใช้เทคนิคเหมืองข้อมูลที่เหมาะสม เช่น เทคนิคต้นไม้ตัดสินใจ การแบ่งกลุ่ม การจัดกลุ่ม การถดถอย ความสัมพันธ์ การพยากรณ์ การตรวจจับการเอนเอียง เป็นต้น แล้วนำเอารูปแบบความสัมพันธ์ที่เป็นผลลัพธ์จากเหมืองข้อมูลมาสรุปหาความรู้ใหม่ ๆ ให้เกิดขึ้นในองค์กร

2.6 การสร้างแบบจำลองด้วยเทคนิคเหมืองข้อมูล

จากการทบทวนวรรณกรรมการใช้เทคนิคเหมืองข้อมูลมีหลายวิธี แต่วิธีที่นิยมใช้และใกล้เคียงกับงานวิจัยฉบับนี้มากที่สุดได้แก่ 3 เทคนิคดังนี้

2.6.1 เทคนิคต้นไม้ตัดสินใจ (Decision Tree) จะทำการคัดเลือกแอตทริบิวต์ที่มีความสัมพันธ์กับคลาสมากที่สุดขึ้นมาเป็นโหนดบนสุดของต้นไม้ (Root Node)

หลังจากนั้นก็จะหาแอตทริบิวต์ถัดไปเรื่อย ๆ ในการหาความสัมพันธ์ของแอตทริบิวต์นี้จะใช้ตัววัด ที่เรียกว่าค่าเกนความรู้ (Information Gain: IG) ค่านี้คำนวณได้จากสมการดังนี้

IG (Parent, Child) = Entropy (Parent) – [$p(c_1) \times$ Entropy (c_1) + $p(c_2) \times$ Entropy (c_2) + ...]

โดยที่ Entropy(c_1) = $-p(c_1) \log p(c_1)$ และ $p(c_1)$ คือ ค่าความน่าจะเป็นของ c_1 ซึ่งค่าเอนโทรปี (Entropy) จะใช้ในการวัดความแตกต่างกันของข้อมูล ถ้าข้อมูลมีความแตกต่างกันน้อยค่าเอนโทรปีจะมีค่าต่ำ แต่ถ้าข้อมูลมีความแตกต่างกันมากค่าเอนโทรปีจะมีค่าสูง ดังนั้นถ้าข้อมูลเอนโทรปีของโหนดลูก (Child) สามารถแบ่งแยกข้อมูลได้ดีจะมีค่าเอนโทรปีต่ำและจะทำให้ค่าเกนความรู้มีค่าสูงเมื่อเทียบกับโหนดบน (Parent) ในขั้นตอนการสร้างแบบจำลองต้นไม้ตัดสินใจจะคำนวนค่าเกนความรู้ของแต่ละแอตทริบิวต์เทียบกับคลาสเพื่อหาแอตทริบิวต์ที่มีค่าเกนความรู้มากที่สุดมาเป็นโหนดบนสุดของต้นไม้

ตัวอย่างข้อมูลในตาราง 8 ซึ่งเป็นข้อมูลที่เก็บสภาพภูมิอากาศ 14 วันย้อนหลังเพื่อดูว่าจะมีการจัดแข่งขันกีฬาหรือไม่ ดังนี้

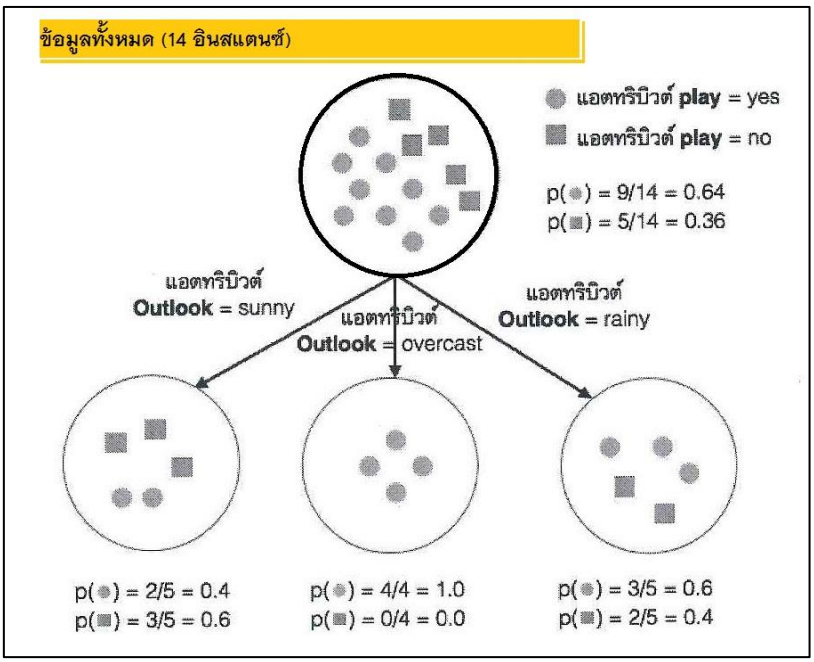
ตาราง 8 ข้อมูลที่เก็บสภาพภูมิอากาศ 14 วันย้อนหลัง

No	Outlook	Temperature	Humidity	Windy	Play
1	Sunny	Hot	High	FALSE	no
2	Sunny	Hot	High	TRUE	no
3	Overcast	Hot	High	FALSE	yes
4	Rainy	Mild	High	FALSE	yes
5	Rainy	Cool	Normal	FALSE	yes
6	Rainy	Cool	Normal	TRUE	no
7	Overcast	Cool	Normal	TRUE	yes
8	Sunny	Mild	High	FALSE	no
9	Sunny	Mild	Normal	FALSE	yes
10	Rainy	Mild	Normal	FALSE	yes
11	Sunny	Mild	Normal	TRUE	yes
12	Overcast	Mild	High	TRUE	yes
13	Overcast	Hot	Normal	FALSE	yes
14	Rainy	Mild	High	TRUE	no

จากข้อมูลในตาราง 8 มี 5 แอตทริบิวต์ ได้แก่ Outlook แสดงสภาพภูมิอากาศ ประกอบด้วย 3 ค่า คือ Sunny, Overcast และ Rainny Temperature แสดงอุณหภูมิ ประกอบด้วย 3 ค่า คือ Hot, Mild และ Cool Humidity แสดงค่าความชื้นในอากาศ ประกอบด้วย 2 ค่า คือ High และ Normal Windy แสดงว่าเป็นวันที่ลมแรงหรือไม่ ประกอบด้วย 2 ค่า คือ TRUE และ FALSE Play แสดงการจัดแข่งขันกีฬา ซึ่งเป็นคลาส ประกอบด้วย 2 ค่า คือ yes และ no

เมื่อได้ข้อมูลดังกล่าวจะทำการคำนวณค่าแต่ละแอตทริบิวต์เทียบกับคลาสเพื่อหาแอตทริบิวต์ที่มีค่าเกนความรู้ มากที่สุดมาเป็นโหนดบนสุดของต้นไม้ตัดสินใจ ดังนี้

2.6.1.1 คำนวณค่าเกนความรู้ ของแอตทริบิวต์ Outlook มีวิธีการคำนวณ ดังนี้



ภาพประกอบ 6 วิธีการคำนวณค่าเกนความรู้ ของแอตทริบิวต์ Outlook (เอกสิทธิ์ พัชรวงศ์ศักดา, 2557)

จากภาพประกอบ 6 พบว่าข้อมูลคลาสในแอตทริบิวต์ Outlook ทั้งหมด 14 จำนวนได้แก่ คลาสคำตอบ yes = 9 คลาสคำตอบ no = 5 และสามารถแบ่งแอตทริบิวต์ Outlook ออกเป็น 3 กลุ่ม ได้แก่ แอตทริบิวต์ Outlook = sunny ซึ่งมีคลาสคำตอบ yes = 2 และคลาสคำตอบ no = 3 แอตทริบิวต์ Outlook = overcast ซึ่งมีคลาสคำตอบ yes = 4

และคลาสคำตอบ no = 0 และ แอตทริบิวต์ Outlook = rainy ซึ่งมีคลาสคำตอบ yes = 3 และคลาสคำตอบ no = 2 จากนั้นจึงนำไปคำนวณตามสูตรได้ดังนี้

Entropy (Parent) = $-p(yes) \times \log_2 p(yes) - p(no) \times \log_2 p(no)$

$= -[0.64 \times \log_2(0.64) + 0.36 \times \log_2(0.36)]$

$= -[0.64 \times -0.64 + 0.36 \times -1.47]$

$= 0.94$

Entropy (Outlook = Sunny) = $-p(yes) \times \log_2 p(yes) - p(no) \times \log_2 p(no)$

$= -[0.4 \times \log_2(0.4) + 0.6 \times \log_2(0.6)]$

$= -[0.4 \times -1.32 + 0.6 \times -0.74]$

$= 0.97$

Entropy (Outlook = Overcast) = $-p(yes) \times \log_2 p(yes) - p(no) \times \log_2 p(no)$

$= -[1.0 \times \log_2(1.0) + 0 \times \log_2(0)]$

$= -[1.0 \times 0 + 0 \times 1]$

$= 0$

Entropy (Outlook = Rainy) = $-p(yes) \times \log_2 p(yes) - p(no) \times \log_2 p(no)$

$= -[0.6 \times \log_2(0.6) + 0.4 \times \log_2(0.4)]$

$= -[0.6 \times -0.74 + 0.4 \times -1.32]$

$= 0.97$

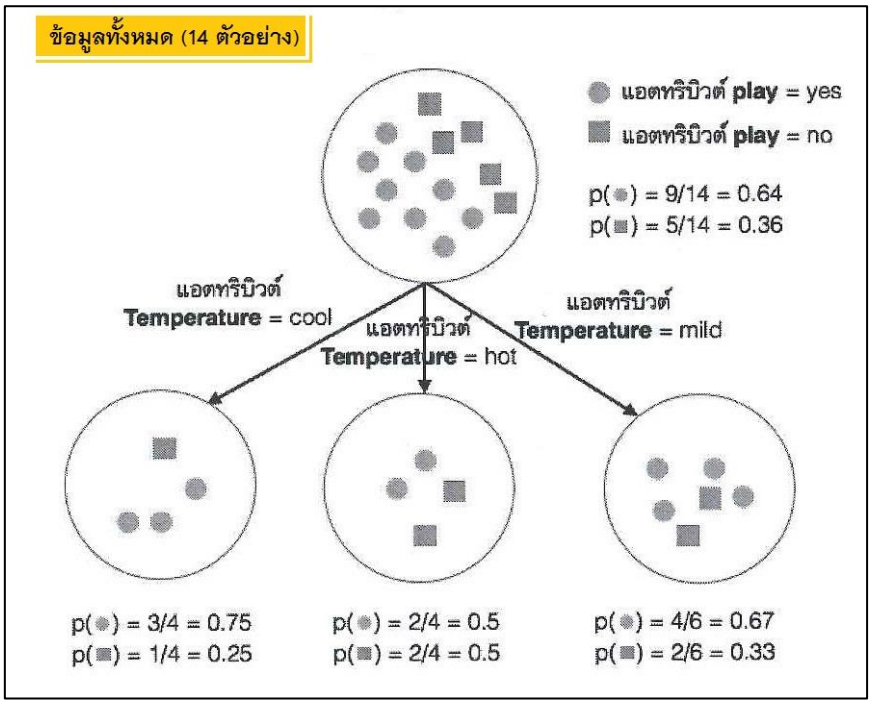
IG(Parent, Child) = Entropy (Parent) – [p(Outlook = Sunny) × Entropy (Outlook = Sunny)+p(Outlook = Overcast) × Entropy (Outlook = Overcast)+p(Outlook = Rainy) × Entropy (Outlook = Rainy)]

$= 0.94 - [0.35 \times 0.97 + 0.30 \times 0 + 0.35 \times 0.97]$

$= 0.96 - 0.68$

$= 0.28$

2.6.1.2 คำนวณค่าเกนความรู้ ของแอตทริบิวต์ Temperature มีวิธีการคำนวณดังนี้



ภาพประกอบ 7 วิธีการคำนวณค่าเกนความรู้ ของแอตทริบิวต์ Temperature (เอกสิทธิ์ พัชรวงศ์ศักดา, 2557)

จากภาพประกอบ 7 พบว่าข้อมูลคลาสในแอตทริบิวต์ Temperature ทั้งหมด 14 จำนวนได้แก่ คลาสคำตอบ yes = 9 คลาสคำตอบ no = 5 และสามารถแบ่งแอตทริบิวต์ Temperature ออกเป็น 3 กลุ่ม ได้แก่ แอตทริบิวต์ Temperature = cool ซึ่งมีคลาสคำตอบ yes = 3 และคลาสคำตอบ no = 1 แอตทริบิวต์ Temperature = hot ซึ่งมีคลาสคำตอบ yes = 2 และคลาสคำตอบ no = 2 และ แอตทริบิวต์ Temperature = mild ซึ่งมีคลาสคำตอบ yes = 4 และคลาสคำตอบ no = 2 จากนั้นจึงนำไปคำนวณตามสูตรได้ดังนี้

Entropy (Temperature = Cool) = $-p(yes) \times \log_2 p(yes) - p(no) \times \log_2 p(no)$

$= -[0.75 \times \log_2(0.75) + 0.25 \times \log_2(0.25)]$

$= 0.81$

Entropy (Temperature = Hot) = $-p(yes) \times \log_2 p(yes) - p(no) \times \log_2 p(no)$

$= -[0.5 \times \log_2(0.5) + 0.5 \times \log_2(0.5)]$

$= 1$

Entropy (Temperature = Mild) = $-p(yes) \times \log_2 p(yes) - p(no) \times \log_2 p(no)$

$= -[0.67 \times \log_2(0.67) + 0.33 \times \log_2(0.33)]$

$= 0.91$

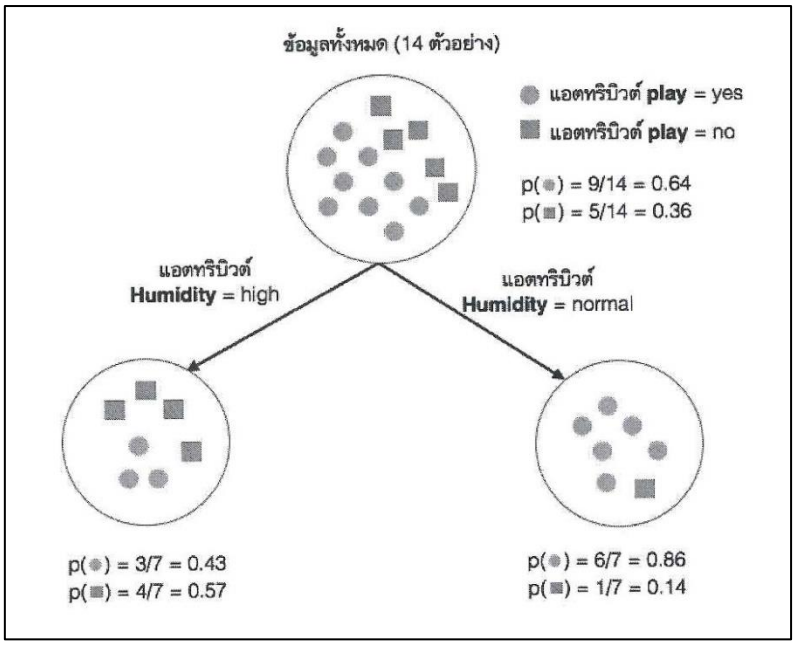
IG(Parent, Child) = Entropy (Parent) – [p(Temperature = Cool) × Entropy (Temperature = Cool)+p(Temperature = Hot) × Entropy (Temperature = Hot)+p(Temperature = Mild) × Entropy (Temperature = Mild)]

$= 0.94 - [0.29 \times 0.81 + 0.29 \times 1 + 0.42 \times 0.91]$

$= 0.96 - 0.91$

$= 0.05$

2.6.1.3 คำนวณค่าเกนความรู้ ของแอตทริบิวต์ Humidity มีวิธีการคำนวณดังนี้



ภาพประกอบ 8 วิธีการคำนวณค่าเกนความรู้ ของแอตทริบิวต์ Humidity

(เอกสิทธิ์ พัชรวงศ์ศักดา, 2557)

จากภาพประกอบ 8 พบว่าข้อมูลคลาสในแอตทริบิวต์ Humidity ทั้งหมด 14 จำนวนได้แก่ คลาสคำตอบ yes = 9 คลาสคำตอบ no = 5 และสามารถแบ่งแอตทริบิวต์

Humidity ออกเป็น 2 กลุ่ม ได้แก่ แอตทริบิวต์ Humidity = high ซึ่งมีคลาสคำตอบ yes = 3 และคลาสคำตอบ no = 4 และ แอตทริบิวต์ Humidity = normal ซึ่งมีคลาสคำตอบ yes = 6 และคลาสคำตอบ no = 1 จากนั้นจึงนำไปคำนวณตามสูตรได้ดังนี้

Entropy (Humidity = High) = $-p(yes) \times \log_2 p(yes) - p(no) \times \log_2 p(no)$

$= -[0.43 \times \log_2 (0.43) + 0.57 \times \log_2 (0.57)]$

$= 0.99$

Entropy (Humidity = Normal) = $-p(yes) \times \log_2 p(yes) - p(no) \times \log_2 p(no)$

$= -[0.86 \times \log_2 (0.86) + 0.14 \times \log_2 (0.14)]$

$= 0.58$

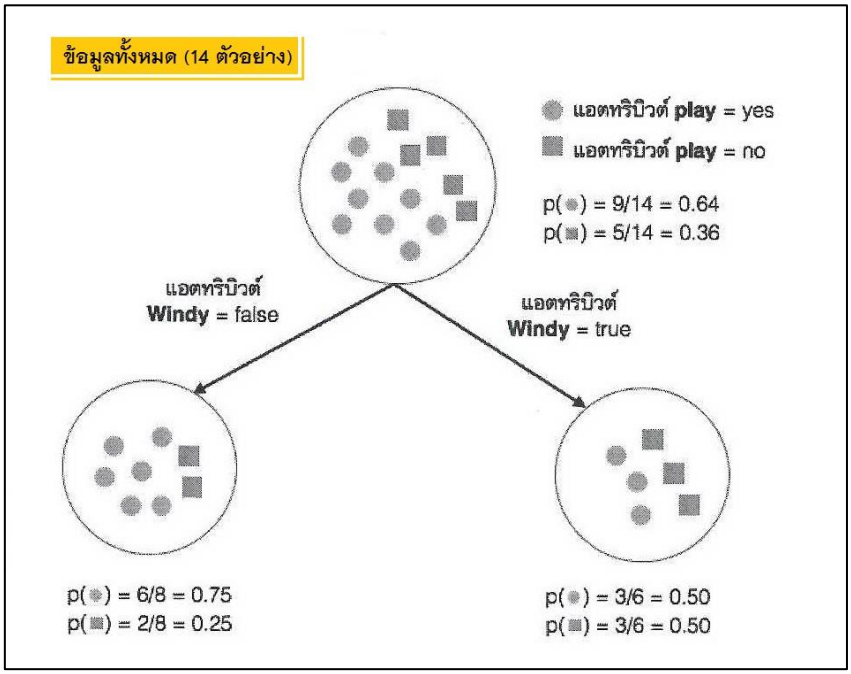
IG(Parent, Child) = Entropy (Parent) – [p(Humidity = Cool) × Entropy (Humidity = Cool)+p(Humidity = Hot) × Entropy (Humidity = Hot)]

$= 0.94-[0.5 \times 0.99 + 0.5 \times 0.58]$

$= 0.96 - 0.79$

$= 0.17$

2.6.1.4 คำนวณค่าเกนความรู้ ของแอตทริบิวต์ Windy มีวิธีการคำนวณดังนี้



ภาพประกอบ 9 วิธีการคำนวณค่าเกนความรู้ ของแอตทริบิวต์ Windy

(เอกสิทธิ์ พัชรวงศ์ศักดา, 2557)

จากภาพประกอบ 9 พบว่าข้อมูลคลาสในแอตทริบิวต์ Windy ทั้งหมด 14 จำนวนได้แก่ คลาสคำตอบ yes = 9 คลาสคำตอบ no = 5 และสามารถแบ่งแอตทริบิวต์ Windy ออกเป็น 2 กลุ่ม ได้แก่ แอตทริบิวต์ Windy = false ซึ่งมีคลาสคำตอบ yes = 6 และคลาสคำตอบ no = 2 และ แอตทริบิวต์ Windy = true ซึ่งมีคลาสคำตอบ yes = 3 และคลาสคำตอบ no = 3 จากนั้นจึงนำไปคำนวณตามสูตรได้ดังนี้

Entropy (Windy = FALSE) = $-p(yes) \times \log_2 p(yes) - p(no) \times \log_2 p(no)$

$= -[0.75 \times \log_2 (0.75) + 0.25 \times \log_2 (0.25)]$

$= 0.81$

Entropy (Windy = TRUE) = $-p(yes) \times \log_2 p(yes) - p(no) \times \log_2 p(no)$

$= -[0.5 \times \log_2 (0.5) + 0.5 \times \log_2 (0.5)]$

$= 1$

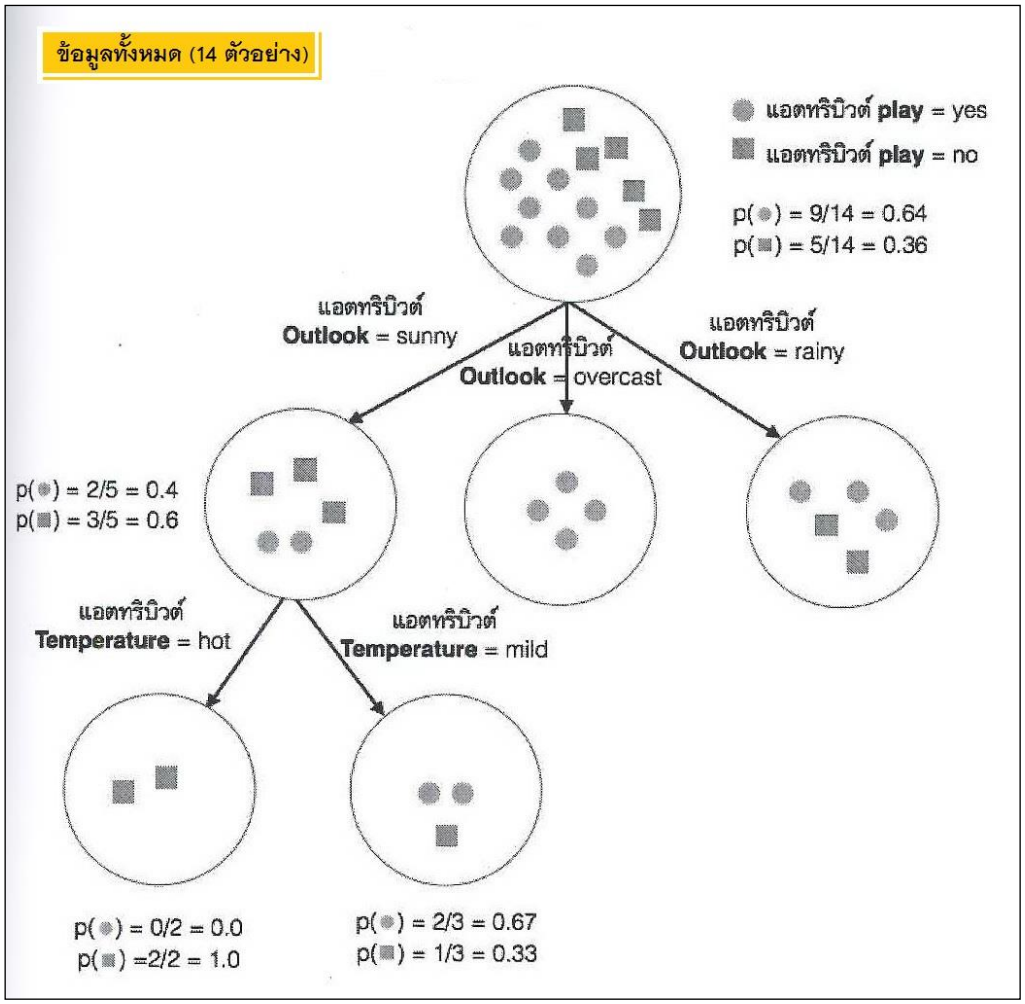
IG(Parent, Child) = Entropy (Parent) – [p(Windy = FALSE) × Entropy (Windy = FALSE)+p(Windy = TRUE) × Entropy (Windy = TRUE)]

$= 0.94-[0.57 \times 0.81 + 0.43 \times 1]$

$= 0.94 - 0.89$

$= 0.05$

2.6.1.5 จากการคำนวณค่าเกนความรู้ ของทุกแอตทริบิวต์พบว่าค่าเกนความรู้ ของแอตทริบิวต์ Outlook มีค่ามากที่สุด (0.28) ดังนั้นจึงเลือกแอตทริบิวต์ Outlook ขึ้นมาเป็นโหนดบนสุด ณ ขั้นตอนนี้จะเห็นว่าข้อมูลที่อยู่ในโหนดที่แอตทริบิวต์ Outlook = Overcast มีคลาสเดียวกันคือ Play = yes ดังนั้นในโหนดนี้จึงไม่ต้องทำการแตกกิ่งต่อไปอีก แต่โหนดอื่น จะต้องทำการแตกกิ่งออกไปจนข้อมูลในแต่ละโหนดมีคลาสคำตอบเดียวกัน โดยจะเริ่มจากการแตกกิ่งออกจากโหนดที่มีแอตทริบิวต์ Outlook = Sunny โดยเริ่มจากนำโหนดที่มีแอตทริบิวต์เป็น Temperature มาต่อก่อนและคำนวณค่าเกนความรู้ ของ Temperature มีวิธีการคำนวณดังนี้



ภาพประกอบ 10 วิธีการคำนวณค่าเกนความรู้ ของแอตทริบิวต์ Outlook = Sunny และ Temperature (เอกสิทธิ์ พัชรวงศ์ศักดา, 2557)

จากภาพประกอบ 10 พบว่าข้อมูลคลาสในแอตทริบิวต์ Temperature เมื่อแอตทริบิวต์ Outlook = Sunny ทั้งหมด 5 จำนวนได้แก่ คลาสคำตอบ yes = 2 คลาสคำตอบ no = 3 และสามารถแบ่งแอตทริบิวต์ Temperature ออกเป็น 2 กลุ่ม ได้แก่ แอตทริบิวต์ Temperature = hot ซึ่งมีคลาสคำตอบ yes = 0 และคลาสคำตอบ no = 2 และ แอตทริบิวต์ Temperature = mild ซึ่งมีคลาสคำตอบ yes = 2 และคลาสคำตอบ no = 1 จากนั้นจึงนำไปคำนวณตามสูตรได้ดังนี้

Entropy (Parent) = $-p(yes) \times \log_2 p(yes) - p(no) \times \log_2 p(no)$

$= -[0.4 \times \log_2 (0.4) + 0.6 \times \log_2 (0.6)]$

$= 0.97$

Entropy (Humidity = High) = $-p(yes) \times \log_2 p(yes) - p(no) \times \log_2 p(no)$

$= -[0.0 \times \log_2 (0.0) + 1.0 \times \log_2 (1.0)]$

$= 0$

Entropy (Temperature = Mild และ Outlook = Sunny)

$= -p(yes) \times \log_2 p(yes) - p(no) \times \log_2 p(no)$

$= -[0.67 \times \log_2 (0.67) + 0.33 \times \log_2 (0.33)]$

$= 0.91$

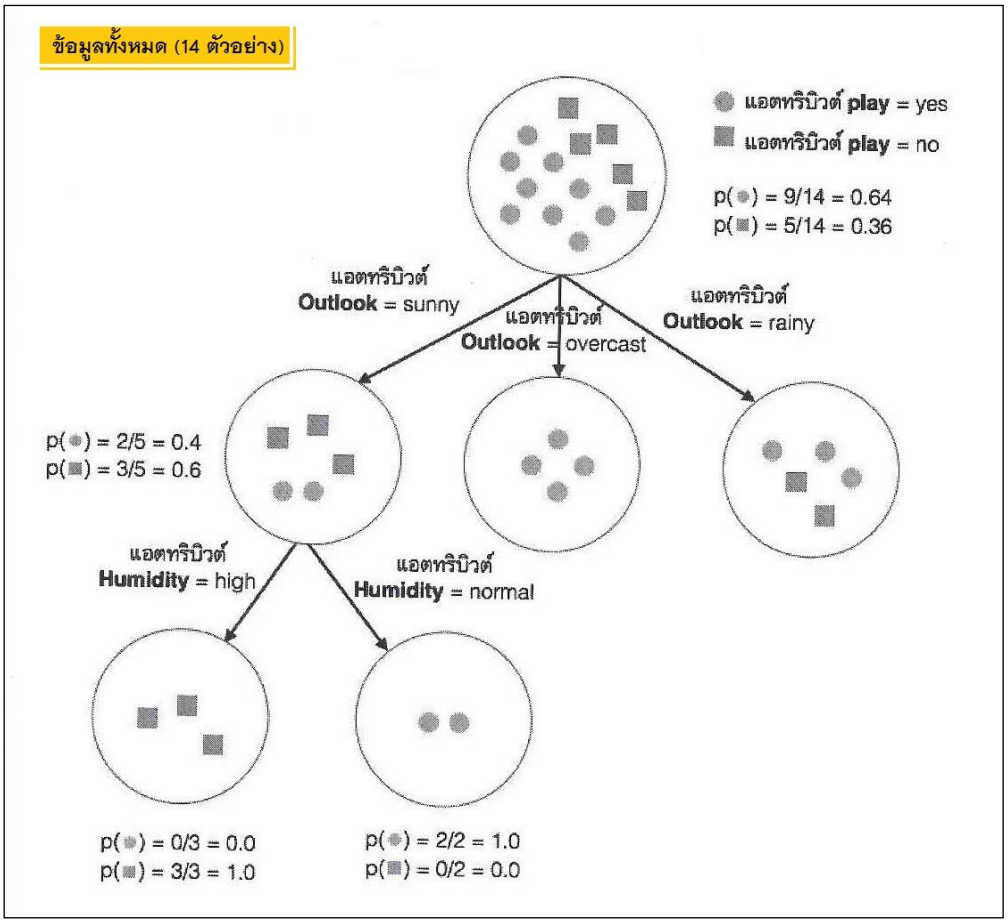
IG(Parent, Child) = Entropy (Parent) – [p(Temperature = Hot) × Entropy (Temperature = Hot)+p(Temperature = Mild) × Entropy (Temperature = Mild)]

$= 0.97-[0.4 \times 0.0 + 0.60 \times 0.91]$

$= 0.97 - 0.55$

$= 0.42$

2.6.1.6 คำนวณค่าเกนความรู้ ของทุกแอตทริบิวต์ Humidity เมื่อ Outlook=Sunny มีวิธีการคำนวณดังนี้



ภาพประกอบ 11 วิธีการคำนวณค่าเกนความรู้ ของแอตทริบิวต์ Outlook = Sunny และ Humidity (เอกสิทธิ์ พัชรวงศ์ศักดา, 2557)

จากภาพประกอบ 11 พบว่าข้อมูลคลาสในแอตทริบิวต์ Humidity เมื่อแอตทริบิวต์ Outlook = Sunny ทั้งหมด 5 จำนวนได้แก่ คลาสคำตอบ yes = 2 คลาสคำตอบ no = 3 และสามารถแบ่งแอตทริบิวต์ Humidity ออกเป็น 2 กลุ่ม ได้แก่ แอตทริบิวต์ Humidity = high ซึ่งมีคลาสคำตอบ yes = 0 และคลาสคำตอบ no = 3 และแอตทริบิวต์ Humidity = normal ซึ่งมีคลาสคำตอบ yes = 2 และคลาสคำตอบ no = 0 จากนั้นจึงนำไปคำนวณตามสูตรได้ดังนี้

Entropy (Humidity = High และ Outlook=Sunny)

$= -p(yes) \times \log_2 p(yes) - p(no) \times \log_2 p(no)$

$= -[0.0 \times \log_2 (0.0) + 1.0 \times \log_2 (1.0)]$

$= 0.0$

Entropy (Humidity = Normal และ Outlook=Sunny)

= $-p(yes) \times \log_2 p(yes) - p(no) \times \log_2 p(no)$

= $-[1.0 \times \log_2 (1.0) + 0.0 \times \log_2 (0.0)]$

= 0.0

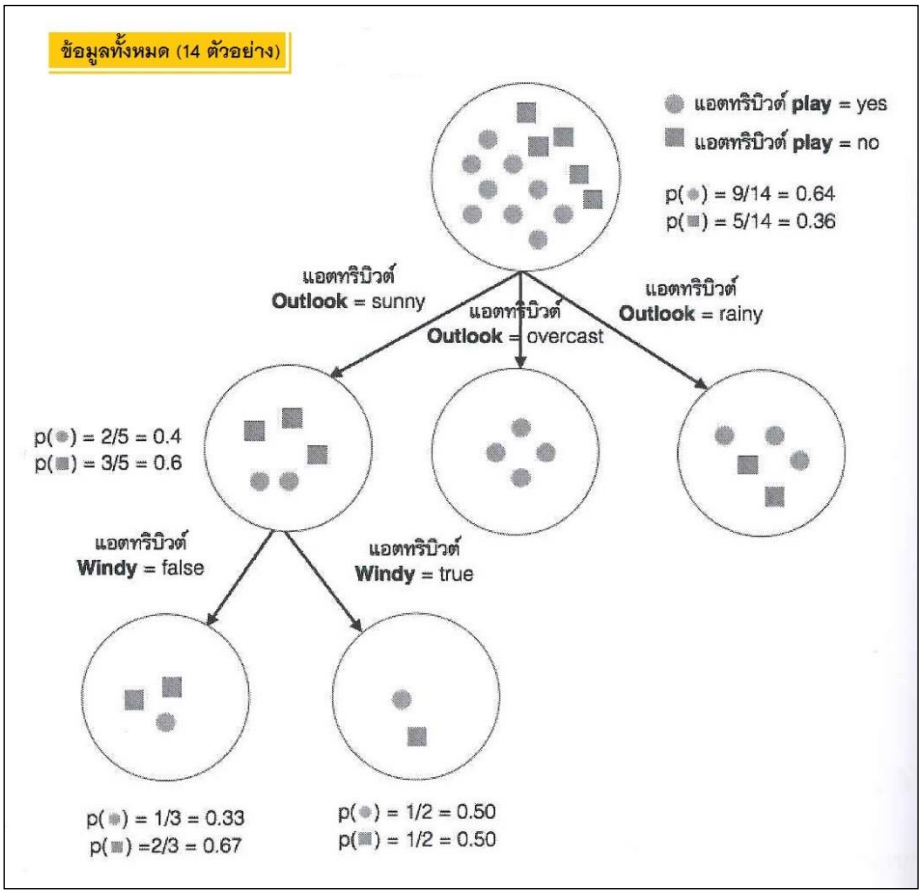
IG(Parent, Child) = Entropy (Parent) – [p(Humidity = High) × Entropy (Humidity = High)+p(Humidity = Normal) × Entropy (Humidity = Normal)]

= 0.97–[0.60 × 0.0 + 0.40 × 0.0]

= 0.97 – 0.0

= 0.97

2.6.1.7 คำนวณค่าเกนความรู้ ของทุกแอตทริบิวต์ Windy เมื่อ Outlook=Sunny มีวิธีการคำนวณดังนี้



ภาพประกอบ 12 วิธีการคำนวณค่าเกนความรู้ ของแอตทริบิวต์ Outlook = Sunny และ Windy (เอกสิทธิ์ พัชรวงศ์ศักดา, 2557)

จากภาพประกอบ 12 พบว่าข้อมูลคลาสในแอตทริบิวต์ Windy เมื่อแอตทริบิวต์ Outlook = Sunny ทั้งหมด 5 จำนวนได้แก่ คลาสคำตอบ yes = 2 คลาสคำตอบ no = 3 และสามารถแบ่งแอตทริบิวต์ Windy ออกเป็น 2 กลุ่ม ได้แก่ แอตทริบิวต์ Windy = false ซึ่งมีคลาสคำตอบ yes = 1 และคลาสคำตอบ no = 2 และ แอตทริบิวต์ Windy = true ซึ่งมีคลาสคำตอบ yes = 1 และคลาสคำตอบ no = 1 จากนั้นจึงนำไปคำนวณตามสูตรได้ดังนี้

Entropy (Windy = False และ Outlook=Sunny)

$= -p(yes) \times \log_2 p(yes) - p(no) \times \log_2 p(no)$

$= -[0.33 \times \log_2 (0.33) + 0.67 \times \log_2 (0.67)]$

$= 0.91$

Entropy (Windy = True และ Outlook=Sunny)

$= -p(yes) \times \log_2 p(yes) - p(no) \times \log_2 p(no)$

$= -[0.50 \times \log_2 (0.50) + 0.50 \times \log_2 (0.50)]$

$= 1.00$

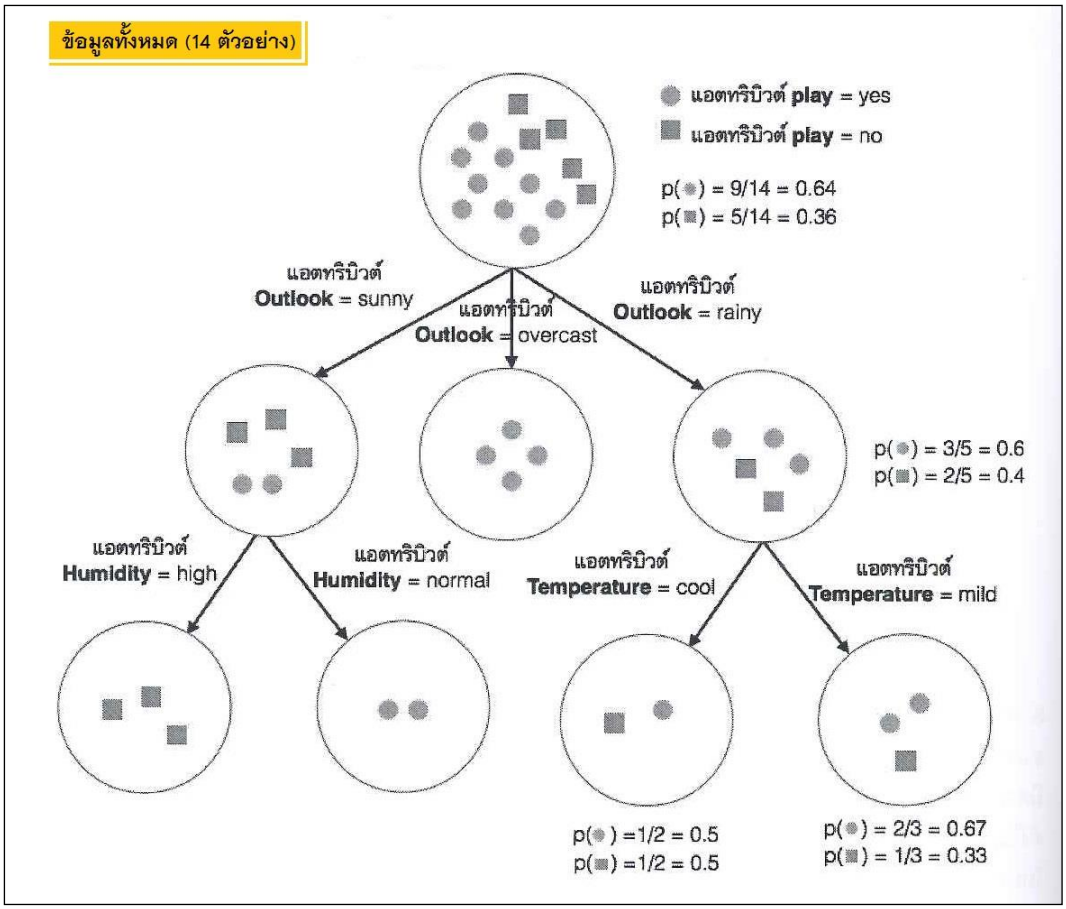
IG(Parent, Child) = Entropy (Parent) – [p(Windy = False) × Entropy (Windy = False)+p(Windy = True) × Entropy (Windy = True)]

$= 0.97-[0.40 \times 0.91 + 0.60 \times 1.0]$

$= 0.97 - 0.96$

$= 0.01$

2.6.1.8 จากการคำนวณค่าเกนความรู้ ของแอตทริบิวต์ Temperature, Humidity และ Windy เมื่อ Outlook = Sunny จะเห็นได้ว่าค่าเกนความรู้ ของแอตทริบิวต์ Humidity มีค่ามากที่สุด คือ 0.97 ดังนั้นจึงเลือกแอตทริบิวต์ Humidity ขึ้นมาเป็นโหนดที่ต่อกับ Outlook = Sunny จากนั้นจะเหลือ Outlook = Rainy ที่ต้องหาโหนดอื่นมาต่อเพิ่ม โดยเริ่มจาก Temperature มาต่อและคำนวณหาค่าเกนความรู้ ของ Temperature ดังนี้



ภาพประกอบ 13 วิธีการคำนวณค่าเกนความรู้ ของแอตทริบิวต์ Outlook = Rainy และ Temperature (เอกสิทธิ์ พัชรวงศ์ศักดา, 2557)

จากภาพประกอบ 13 พบว่าข้อมูลคลาสในแอตทริบิวต์ Temperature เมื่อแอตทริบิวต์ Outlook = Rainy ทั้งหมด 5 จำนวนได้แก่ คลาสคำตอบ yes = 3 คลาสคำตอบ no = 2 และสามารถแบ่งแอตทริบิวต์ Temperature ออกเป็น 2 กลุ่ม ได้แก่ แอตทริบิวต์ Temperature = cool ซึ่งมีคลาสคำตอบ yes = 1 และคลาสคำตอบ no = 1 และ แอตทริบิวต์ Temperature = mild ซึ่งมีคลาสคำตอบ yes = 2 และคลาสคำตอบ no = 1 จากนั้นจึงนำไปคำนวณตามสูตรได้ดังนี้

Entropy (Parent) = $-p(yes) \times \log_2 p(yes) - p(no) \times \log_2 p(no)$

$= -[0.6 \times \log_2 (0.6) + 0.4 \times \log_2 (0.4)]$

$= 0.97$

Entropy (Temperature = Cool และ Outlook= Rainy)

= –p(yes) × $\log_2$ p(yes) – p(no) × $\log_2$ p(no)

= –[0.5 × $\log_2$ (0.5) + 0.5× $\log_2$ (0.5)]

= 1.0

Entropy (Temperature = Mild และ Outlook= Rainy)

= –p(yes) × $\log_2$ p(yes) – p(no) × $\log_2$ p(no)

= –[0.67 × $\log_2$ (0.67) + 0.33× $\log_2$ (0.33)]

= 0.91

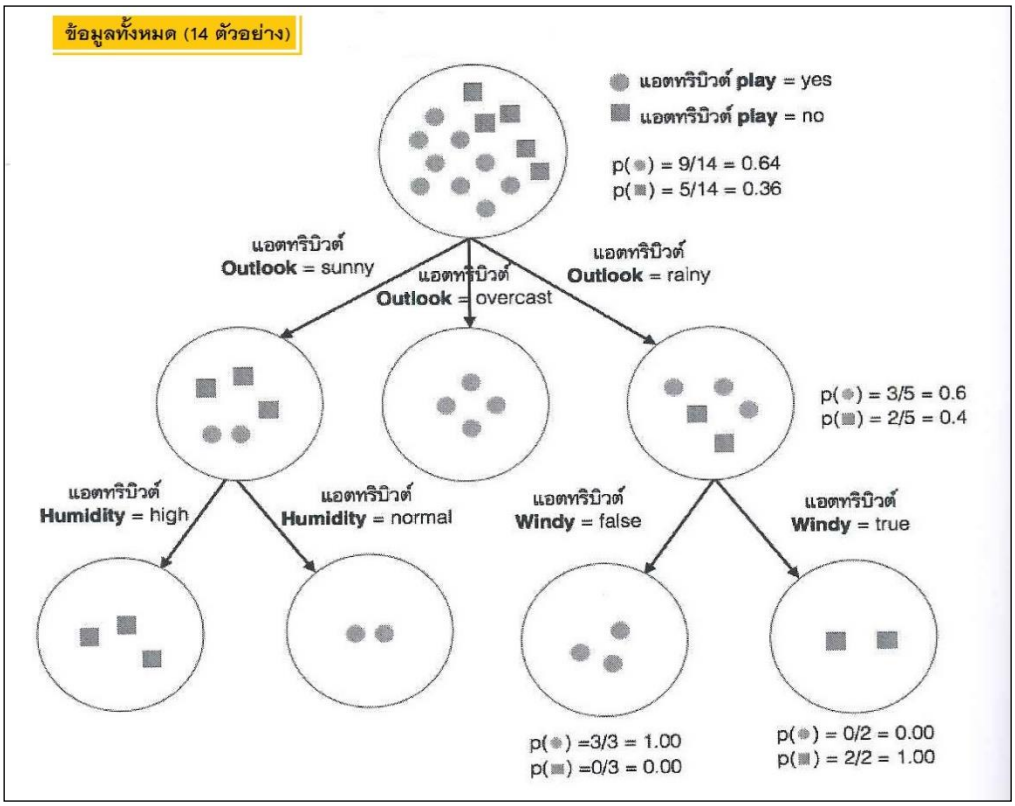
IG(Parent, Child) = Entropy (Parent) – [p(Temperature = Cool) × Entropy (Temperature = Cool)+p(Temperature = Mild) × Entropy (Temperature = Mild)]

= 0.97–[0.40 × 1.0 + 0.60 × 0.91]

= 0.97 – 0.95

= 0.02

2.6.1.9 คำนวณค่าเกนความรู้ ของแอตทริบิวต์ Windy เมื่อ Outlook=Rainy



ภาพประกอบ 14 วิธีการคำนวณค่าเกนความรู้ ของแอตทริบิวต์ Outlook = Rainy และ Windy (เอกสิทธิ์ พัชรวงศ์ศักดา, 2557)

จากภาพประกอบ 14 พบว่าข้อมูลคลาสในแอตทริบิวต์ Windy เมื่อแอตทริบิวต์ Outlook = Rainy ทั้งหมด 5 จำนวนได้แก่ คลาสคำตอบ yes = 3 คลาสคำตอบ no = 2 และสามารถแบ่งแอตทริบิวต์ Windy ออกเป็น 2 กลุ่ม ได้แก่ แอตทริบิวต์ Windy = false ซึ่งมีคลาสคำตอบ yes = 3 และคลาสคำตอบ no = 0 และ แอตทริบิวต์ Windy = true ซึ่งมีคลาสคำตอบ yes = 0 และคลาสคำตอบ no = 2 จากนั้นจึงนำไปคำนวณตามสูตรได้ดังนี้

Entropy (Windy = False และ Outlook= Rainy)

$= -p(yes) \times \log_2 p(yes) - p(no) \times \log_2 p(no)$

$= -[1.0 \times \log_2 (1.0) + 0.0 \times \log_2 (0.0)]$

$= 0.00$

Entropy (Windy = True และ Outlook= Rainy)

$= -p(yes) \times \log_2 p(yes) - p(no) \times \log_2 p(no)$

$= -[0.0 \times \log_2 (0.0) + 1.0 \times \log_2 (1.0)]$

$= 0.00$

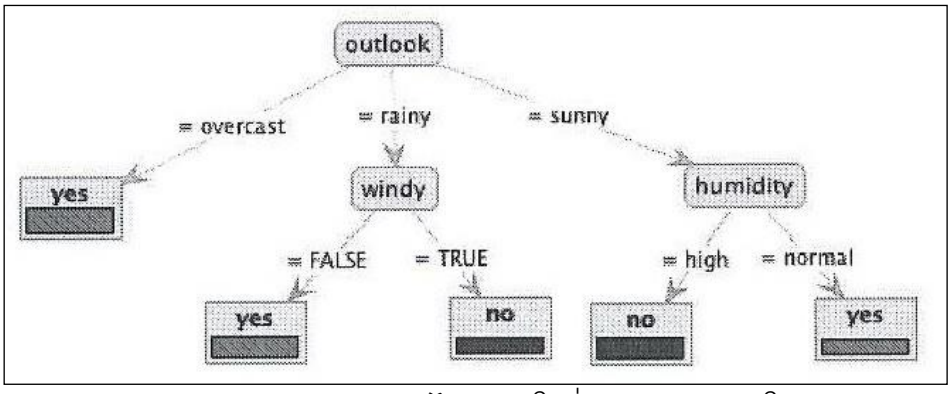
IG(Parent, Child) = Entropy (Parent) – [p(Windy = False) × Entropy (Windy = False)+p(Windy = True) × Entropy (Windy = True)]

$= 0.97-[0.60 \times 0.0 + 0.40 \times 0.0]$

$= 0.97 - 0.00$

$= 0.97$

จากการคำนวณค่าเกนความรู้ ของแอตทริบิวต์ Temperature และ Windy เมื่อ Outlook = Rainy จะเห็นได้ว่าค่าเกนความรู้ ของแอตทริบิวต์ Windy มีค่ามากที่สุด IG = 0.97 ดังนั้นจึงเลือกแอตทริบิวต์ Windy ขึ้นมาเป็นโหนดที่ต่อกับ Outlook = Rainy และพบว่าข้อมูลในโหนดลูกสุดท้าย (Leaf Node) ทุกโหนดสามารถทำการตอบคลาสได้อย่างชัดเจนแล้วจึงได้แบบจำลองต้นไม้ตัดสินใจ ดังนี้



ภาพประกอบ 15 แบบจำลองต้นไม้ตัดสินใจที่สร้างจากข้อมูลในตาราง 8 (เอกสิทธิ์ พัชรวงศ์ศักดา, 2557)

จากภาพประกอบ 15 แบบจำลองต้นไม้ตัดสินใจสามารถอธิบายได้ดังนี้

1) ถ้าวันไหนที่สภาพอากาศเป็นแบบ Outlook มีค่าเป็น Overcast แล้วผลคือจะมีการจัดแข่งขันกีฬา

2) ถ้าวันไหนที่สภาพอากาศเป็นแบบ Outlook มีค่าเป็น Rainy และสภาพอากาศเป็นแบบ Windy มีค่าเป็น False แล้วผลคือจะมีการจัดแข่งขันกีฬา

3) ถ้าวันไหนที่สภาพอากาศเป็นแบบ Outlook มีค่าเป็น Rainy และสภาพอากาศเป็นแบบ Windy มีค่าเป็น True แล้วผลคือจะไม่มีการจัดแข่งขันกีฬา

4) ถ้าวันไหนที่สภาพอากาศเป็นแบบ Outlook มีค่าเป็น Sunny และสภาพอากาศเป็นแบบ Humidity มีค่าเป็น high แล้วผลคือจะไม่มีการจัดแข่งขันกีฬา

5) ถ้าวันไหนที่สภาพอากาศเป็นแบบ Outlook มีค่าเป็น Sunny และสภาพอากาศเป็นแบบ Humidity มีค่าเป็น normal แล้วผลคือจะมีการจัดแข่งขันกีฬา

จากการสร้างแบบจำลองต้นไม้ตัดสินใจ มีข้อดีได้แก่ เป็นแบบจำลองที่เข้าใจง่าย โดยอาจจะนำมาสร้างเป็นกฎถ้า – แล้ว (IF–THEN) ได้ ซึ่งสามารถแปลความจากแบบจำลองได้ และแบบจำลองที่สร้างได้คัดเลือกแอตทริบิวต์ที่มีความสัมพันธ์กับคลาสคำตอบ ดังนั้นจึงอาจจะไม่ใช้ทุกแอตทริบิวต์ในข้อมูลเทรนนิ่ง (เอกสิทธิ์ พัชรวงศ์ศักดา, 2557)

2.6.2 เทคนิคนาอีฟเบย์ (Naïve Bayes) เป็นอีกวิธีในการสร้างแบบจำลอง โดยใช้ทฤษฎีความน่าจะเป็นตามกฎของเบย์ (Bayes' Theorem) เพื่อหาว่าสมมติฐานใดน่าจะถูกต้องที่สุด โดยใช้ความรู้ก่อนหน้า (Prior Knowledge) ได้แก่ ความน่าจะเป็นก่อนหน้าสำหรับสมมติฐานหนึ่ง ๆ ร่วมกับข้อมูล เช่น ความน่าจะเป็นที่สังเกตได้สำหรับสมมติฐานหนึ่ง ๆ เพื่อหาสมมติฐานที่ดีที่สุด โดยมีสมการดังนี้

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

P(B) คือ ความน่าจะเป็นก่อนหน้าของสมมติฐาน B หรือคือความน่าจะเป็นของคลาส B (Prior Probability)

P(A) คือ ความน่าจะเป็นก่อนหน้าของชุดข้อมูลตัวอย่าง A

P(B|A) คือ ความน่าจะเป็นของ B ขึ้นต่อ A (Posterior Probability) หรือคือความน่าจะเป็นที่ข้อมูลที่มีแอตทริบิวต์เป็น A จะมีคลาส B

P(A|B) คือ ความน่าจะเป็นของ A ขึ้นต่อ B (Likelihood) หรือคือความน่าจะเป็นที่ข้อมูลเทรนนิ่ง ดาต้าที่มีคลาส B และมีแอตทริบิวต์ A โดยที่ $A = a_1 \cap a_2 \ldots \cap a_M$ และ M คือจำนวนแอตทริบิวต์ในเทรนนิ่งดาต้า

ตาราง 9 แบบจำลองนาอีฟเบย์ จากข้อมูลตาราง 8

แอตทริบิวต์	คลาส Play = yes	คลาส Play = no
Outlook = Sunny	0.22	0.60
Outlook = Overcast	0.45	0.00
Outlook = Rainy	0.33	0.40
Temperature = Hot	0.22	0.40
Temperature = Mild	0.45	0.40
Temperature = Cool	0.33	0.20
Humidity = High	0.33	0.80
Humidity = Normal	0.67	0.20
Windy = True	0.33	0.60
Windy = False	0.67	0.40

จากข้อมูลสภาพอากาศตาราง 8 ตัวอย่างในวันปัจจุบันซึ่งยังไม่ทราบคลาส คำตอบ คือ Outlook = Sunny, Temperature = Hot, Humidity = High, Windy = False และ Play = ? สามารถคำนวณได้ดังนี้

P(Play=yes|A) = P(Outlook = Sunny | Play = yes) x P(Temperature = Hot | Play = yes) x P(Humidity = High | Play = yes) x P(Windy = False | Play = yes) x P(Play = yes)

= 0.22 x 0.22 x 0.33 x 0.67 x 0.64

= 0.0068

P(Play=no|A) = P(Outlook = Sunny | Play = no) x P(Temperature = Hot | Play = no) x P(Humidity = High | Play = no) x P(Windy = False | Play = no) x P(Play = no)

= 0.60 x 0.40 x 0.80 x 0.40 x 0.36

= 0.0276

ดังนั้นข้อมูล Outlook = Sunny, Temperature = Hot, Humidity = High, Windy = False จะทำนายว่ามีคลาส Play = no เนื่องจากค่า P(Play=no|A) = 0.0276 ซึ่งมีค่ามากกว่า P(Play=yes|A) = 0.0068

2.6.3 เทคนิคเพื่อนบ้านใกล้ที่สุด (K–Nearest Neighbors) คือ วิธีการในการจัดแบ่งคลาส เทคนิคนี้จะตัดสินใจว่าคลาสใดที่จะแทนเงื่อนไขหรือกรณีใหม่ ๆ ได้บ้าง โดยการตรวจสอบจำนวนบางจำนวนของกรณีหรือเงื่อนไขที่เหมือนกันหรือใกล้เคียงกันมากที่สุด โดยจะหาผลรวมของจำนวนเงื่อนไข หรือกรณีต่าง ๆ สำหรับแต่ละคลาส และกำหนดเงื่อนไขใหม่ ให้คลาสที่เหมือนกันกับคลาสที่ใกล้เคียงกันมากที่สุดหลักการทำงานของเทคนิคเพื่อนบ้านใกล้ที่สุดคล้ายกับการแบ่งข้อมูลคือทำการวัดระยะห่างระหว่างข้อมูลที่ต้องการทำนายกับข้อมูลที่อยู่ใกล้เคียงเป็นจำนวน K ตัว และคำตอบที่ได้คือคลาสที่พบมากที่สุดของข้อมูลที่เป็นเพื่อนบ้านทั้ง K ตัว ในเทคนิคนี้มักจะใช้วิธีการวัดระยะห่างแบบยุคลิด (Euclidean Distance) ซึ่งเกิดจากรากที่สองของผลต่างระหว่างแอตทริบิวต์ต่าง ๆ ยกกำลังสอง ดังสมการ $D_{Euclidean} = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_L - y_L)^2}$

โดยที่ x_1 คือแอตทริบิวต์ที่ 1 ของข้อมูลจุดที่ 1 และ y_1 คือแอตทริบิวต์ที่ 1 ของจุดที่ 2 โดยข้อมูล x และ y มีจำนวนแอตทริบิวต์เท่ากับ L ซึ่งวิธีเพื่อนบ้านใกล้ที่สุดไม่มีการสร้างแบบจำลองแต่จะใช้ข้อมูลเทรนนิ่งดาต้าทั้งหมดมาช่วยในการทำนาย

ทำให้วิธีนี้ถูกจัดอยู่ในกลุ่มของเลซี่โมเดล (Lazy Model) (เอกสิทธิ์ พัชรวงศ์ศักดา, 2557)

ตาราง 10 ตัวอย่างเทคนิคเพื่อนบ้านใกล้ที่สุด

Data (ชุดข้อมูล)	x_1 (แอตทริบิวต์ที่ 1)	x_2 (แอตทริบิวต์ที่ 2)	Y (คลาสคำตอบ)	Distance (ระยะห่าง)	Nearest Sign (คำตอบที่ใกล้)
D1	6	1	+	3.6	+
	Distance $D1 = \sqrt{(3-6)^2 + (3-1)^2}$ = 3.6				
D2	5	2	+	2.2	+
	Distance $D2 = \sqrt{(3-5)^2 + (3-2)^2}$ = 2.2				
D3	1	4	–	2.2	–
	Distance $D3 = \sqrt{(3-1)^2 + (3-4)^2}$ = 2.2				
D4	2	5	–	2.2	–
	Distance $D4 = \sqrt{(3-2)^2 + (3-5)^2}$ = 2.2				
D5	7	4	–	4.1	
	Distance $D5 = \sqrt{(3-7)^2 + (3-4)^2}$ = 4.1				
D6	8	1	+	5.8	
	Distance $D6 = \sqrt{(3-8)^2 + (3-1)^2}$ = 5.8				
D7	6	10	–	7.6	
	Distance $D7 = \sqrt{(3-6)^2 + (3-10)^2}$ = 7.6				
D8	6	6	+	4.2	
	Distance $D8 = \sqrt{(3-6)^2 + (3-6)^2}$ = 4.2				
D9	5	1	–	2.8	–
	Distance $D9 = \sqrt{(3-5)^2 + (3-1)^2}$ = 2.8				
ข้อมูลที่จะทำนาย	3	3	?		
กำหนดค่า	K = 5				

จากตาราง 10 กำหนดให้ K = 5 ดังนั้น ระยะทางที่ใกล้กับจุด (3,3) มากที่สุด 5 ลำดับ คือ D1(+), D2(+), D3(–), D4(–), D9(–) จากระยะทางที่ใกล้ที่สุดทั้ง 5 ให้สังเกตว่ากลุ่มคลาสคำตอบที่มีจำนวนมากที่สุด ปรากฏว่าคือ คลาสคำตอบ (–) ดังนั้น จึงกำหนดกลุ่มคลาสคำตอบให้กับจุด (3,3) คือ คลาสคำตอบ (–)

2.7 ตัววัดประสิทธิภาพของแบบจำลอง

การวัดประสิทธิภาพของแบบจำลองนั้นจะใช้คอนฟิวชันเมทริกซ์ (Confusion Matrix) เป็นเครื่องมือในการอธิบายและคำนวณผลการทำนายของแบบจำลอง ซึ่งจะแสดงตารางแบบจัตุรัสโดยมีจำนวนแถวเท่ากับจำนวนคอลัมน์และเท่ากับจำนวนคลาส ซึ่งมีคลาสคำตอบอยู่ 2 ค่า คือ yes และ no โดยข้อมูลด้านคอลัมน์คือ คลาสที่อยู่ในข้อมูลเทรนนิ่งดาต้า (Actual) และข้อมูลในแนวแถวคือ คลาสที่แบบจำลองทำนายมาได้ (Predicted)

ตาราง 11 คอนฟิวชันเมทริกซ์ของข้อมูลที่มี 2 คลาส คือ yes และ no

Predicted / Actual	yes	no
yes	TP	FP
no	FN	TN

ความหมายของคอนฟิวชันเมทริกซ์จะอธิบายได้ดังนี้

ความถูกต้องเชิงบวก (True Positive: TP) คือจำนวนข้อมูลที่ทำนายถูกว่าเป็นคลาสซึ่งกำลังสนใจอยู่

ความถูกต้องเชิงลบ (True Negative: TN) คือจำนวนข้อมูลที่ทำนายถูกว่าเป็นคลาสซึ่งไม่ได้สนใจอยู่

ความผิดพลาดเชิงบวก (False Positive: FP) คือจำนวนข้อมูลที่ทำนายผิดมาเป็นคลาสซึ่งกำลังสนใจอยู่

ความผิดพลาดเชิงลบ (False Negative: FN) คือจำนวนข้อมูลที่ทำนายผิดมาเป็นคลาสซึ่งไม่ได้สนใจอยู่

การนำแบบจำลองไปใช้งานจริงได้ จะต้องทราบประสิทธิภาพของแบบจำลอง โดยเอกสิทธิ์ พัชรวงศ์ศักดา (2557) ได้แบ่งตัววัดที่นิยมใช้ในงานวิจัยและการทำงานอยู่ 4 ค่า คือ

1) การวัดความแม่นยำของแบบจำลอง โดยพิจารณาแยกทีละคลาส หรือค่าความแม่นยำ (Precision)

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

2) การวัดความถูกต้องของแบบจำลอง โดยพิจารณาแยกทีละคลาสหรือค่าความระลึก (Recall)

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

3) การวัดค่า Precision และ Recall พร้อมกันของแบบจำลอง โดยพิจารณาแยกทีละคลาส หรือค่าความถ่วงดุล (F–measure)

$$\text{F–measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

4) การวัดค่าความถูกต้องของแบบจำลอง โดยพิจารณารวมทุกคลาส (Accuracy)

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

2.8 การแบ่งข้อมูลเพื่อใช้ในการวัดประสิทธิภาพ

การวัดประสิทธิภาพของแบบจำลองได้ จำเป็นต้องมีการแบ่งข้อมูลออกเป็น 2 ส่วน โดยส่วนที่ 1 ใช้เพื่อสร้างแบบจำลองและส่วนที่ 2 ให้แบบจำลองทำนายคลาสคำตอบออกมา ซึ่งการแบ่งข้อมูลเพื่อทำการทดสอบ เอกสิทธิ์ พัชรวงศ์ศักดา (2557) ได้แบ่งเป็น 3 วิธีการ ดังนี้

2.8.1 วิธีทดสอบความสอดคล้องของตนเอง (Self Consistency Test / Use Training Set)

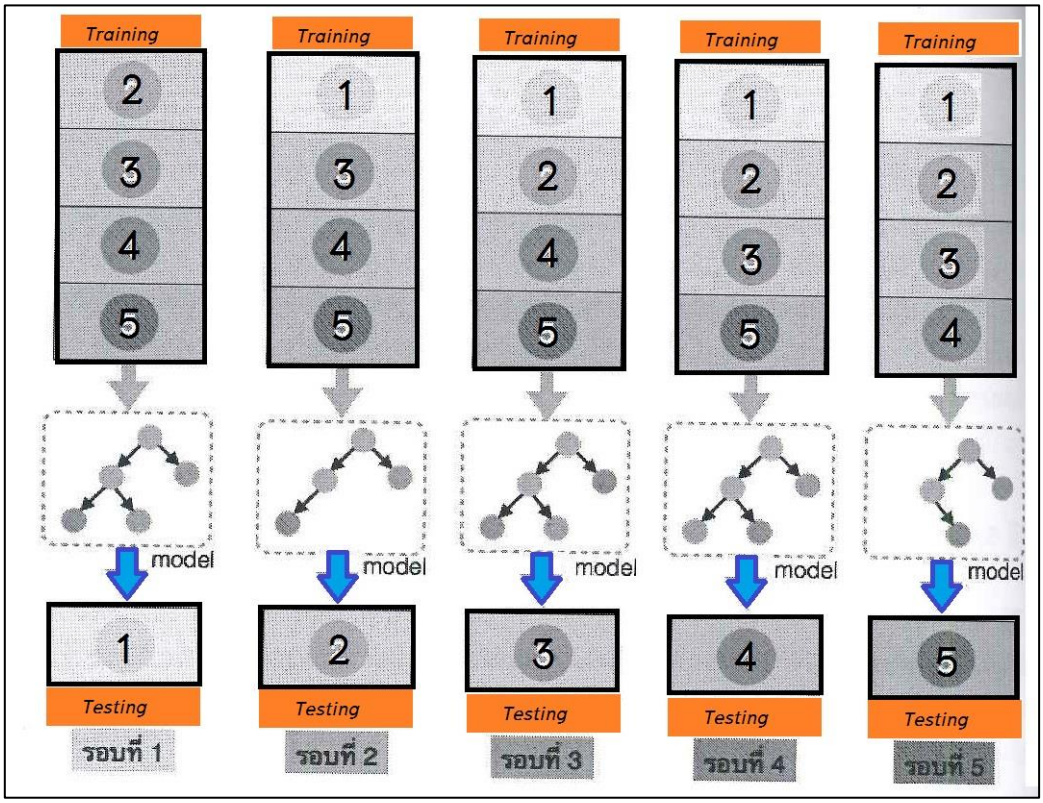
เป็นวิธีการที่ง่ายที่สุด คือข้อมูลที่ใช้ในการสร้างแบบจำลองและข้อมูลที่ใช้ในการทดสอบแบบจำลองเป็นข้อมูลชุดเดียวกัน กระบวนการนี้เริ่มจากสร้างแบบจำลองด้วยข้อมูลเทรนนิ่งดาต้า หลังจากนั้นนำแบบจำลองที่สร้างได้มาทำนายข้อมูลเทรนนิ่งดาต้าชุดเดิม การวัดประสิทธิภาพด้วยวิธีนี้จะให้ผลการวัดประสิทธิภาพที่มีค่าสูงมาก เนื่องจากเป็นข้อมูลชุดเดิมที่ระบบได้ทำการเรียนรู้มาแล้ว แต่ผลการวัดที่ได้ไม่เหมาะที่จะนำไปรายงานในงานวิจัย ซึ่งวิธีการนี้เหมาะสำหรับใช้ในการทดสอบประสิทธิภาพเพื่อดูแนวโน้มของแบบจำลองที่สร้างขึ้น ถ้าได้ผลการวัดที่น้อยแสดงว่าแบบจำลองไม่เหมาะสมกับข้อมูลจึงไม่ควรจะนำไปทดสอบด้วยวิธีการแบ่งข้อมูลแบบอื่น

2.8.2 วิธีแยกทดสอบ (Split Test)

วิธีเป็นการแบ่งข้อมูลด้วยการสุ่มออกเป็น 2 ส่วน เช่น ร้อยละ 70 ต่อ ร้อยละ 30 หรือ ร้อยละ 80 ต่อ ร้อยละ 20 โดยข้อมูลส่วนที่หนึ่ง ร้อยละ 70 หรือ ร้อยละ 80 ใช้ในการสร้างแบบจำลองและข้อมูลส่วนที่สอง ร้อยละ 30 หรือ ร้อยละ 20 ใช้ในการทดสอบประสิทธิภาพของแบบจำลอง แต่การทดสอบแบบแยกทดสอบ นี้ทำการสุ่มข้อมูลเพียงครั้งเดียว ซึ่งในบางครั้งการสุ่มข้อมูลที่ใช้ในการทดสอบที่มีลักษณะคล้ายกับข้อมูลที่ใช้สร้างแบบจำลองทำให้ผลการวัดประสิทธิภาพออกมาดี ในทางตรงข้ามถ้าการสุ่มข้อมูลที่ใช้ในการทดสอบที่มีลักษณะแตกต่างกับข้อมูลที่ใช้สร้างแบบจำลองมากทำให้ผลการวัดประสิทธิภาพออกมาแย่ ดังนั้นวิธีแยกทดสอบ จึงต้องทำการสุ่มหลายครั้ง แต่ข้อดีของวิธีการแยกทดสอบคือใช้เวลาในการสร้างแบบจำลองน้อย ซึ่งเหมาะกับชุดข้อมูลที่มีขนาดใหญ่มาก

2.8.3 วิธีทดสอบครอสวาลิเดชั่น (Cross-validation Test)

วิธีนี้เป็นวิธีที่นิยมใช้ในการทดสอบประสิทธิภาพของแบบจำลอง เนื่องจากผลที่ได้มีความน่าเชื่อถือ (เอกสิทธิ์ พัชรวงศ์ศักดา, 2557) การวัดประสิทธิภาพด้วยวิธีครอสวาลิเดชั่น นี้จะทำการแบ่งข้อมูลออกเป็นหลายส่วน (มักแสดงด้วยค่า K) เช่น 5-Fold Cross-validation คือ ทำการแบ่งข้อมูลออกเป็น 5 ส่วนโดยที่แต่ละส่วนมีจำนวนข้อมูลเท่ากัน หรือ 10-Fold Cross-validation คือ ทำการแบ่งข้อมูลออกเป็น 10 ส่วนโดยที่แต่ละส่วนมีจำนวนข้อมูลเท่ากัน หลังจากนั้นข้อมูลหนึ่งส่วนจะใช้เป็นตัวทดสอบประสิทธิภาพของแบบจำลอง ทำวนไปจนครบจำนวนที่แบ่งไว้



ภาพประกอบ 16 ตัวอย่างการแบ่งข้อมูลแบบ 5-Fold Cross-validation

(เอกสิทธิ์ พัชรวงศ์ศักดา, 2557)

จากภาพประกอบ 16 แบ่งข้อมูลเทรนนิ่งดาต้าออกเป็น 5 ส่วนที่มีจำนวนเท่ากัน หลังจากนั้นทำการทดสอบประสิทธิภาพของแบบจำลอง 5 ครั้ง ดังนี้

รอบที่ 1 ใช้ข้อมูลส่วนที่ 2, 3, 4 และ 5 สร้างแบบจำลอง และใช้แบบจำลองทำนายข้อมูลส่วนที่ 1

รอบที่ 2 ใช้ข้อมูลส่วนที่ 1, 3, 4 และ 5 สร้างแบบจำลอง และใช้แบบจำลองทำนายข้อมูลส่วนที่ 2

รอบที่ 3 ใช้ข้อมูลส่วนที่ 1, 2, 4 และ 5 สร้างแบบจำลอง และใช้แบบจำลองทำนายข้อมูลส่วนที่ 3

รอบที่ 4 ใช้ข้อมูลส่วนที่ 1, 2, 3 และ 5 สร้างแบบจำลอง และใช้แบบจำลองทำนายข้อมูลส่วนที่ 4

รอบที่ 5 ใช้ข้อมูลส่วนที่ 1, 2, 3 และ 4 สร้างแบบจำลอง และใช้แบบจำลองทำนายข้อมูลส่วนที่ 5

ซึ่งข้อมูลทุกชุด จะได้เป็นตัวทดสอบประสิทธิภาพของแบบจำลอง โดยการทดสอบประสิทธิภาพของแบบจำลองแต่ละรอบจะได้จำนวน TP, TN, FP, FN ใส่ลงไปในตารางคอนฟิวชันเมทริกซ์และบวกเพิ่มเข้าไป สุดท้ายจะได้ตารางคอนฟิวชันเมทริกซ์ที่เป็นค่ารวมทั้งหมดหลังจากนั้นจึงทำการคำนวณค่าความแม่นยำ ค่าความระลึก ค่าความถ่วงดุล และ ค่าความถูกต้องต่อไป

ในงานวิจัยนี้การทดสอบประสิทธิภาพของแบบจำลองเลือกใช้วิธีทดสอบครอสวาลิเดชั่นโดยแบ่งข้อมูลออกเป็น 10 ส่วน (10-Fold Cross-validation) เพราะมีความน่าเชื่อถือ และเป็นที่นิยมใช้ในงานวิจัย (เอกสิทธิ์ พัชรวงศ์ศักดา, 2557)

3. โรงพยาบาลนาหว้า

โรงพยาบาลนาหว้า คือโรงพยาบาลชุมชน ขนาด 30 เตียง ตั้งประจำอยู่ที่ตำบลนาหว้า อำเภอนาหว้า จังหวัดนครพนม ซึ่งเป็นสถานที่สำหรับให้บริการด้านสุขภาพให้กับผู้ป่วย โดยจะมุ่งเน้นการส่งเสริม ป้องกัน รักษา และฟื้นฟูภาวะความเจ็บป่วย หรือโรคต่าง ๆ ทั้งทางร่างกายและทางจิตใจ

3.1 โครงสร้างองค์กร วิสัยทัศน์ พันธกิจ ของโรงพยาบาลนาหว้า

โครงสร้างองค์กรของโรงพยาบาลนาหว้า แบ่งออกเป็น 8 กลุ่ม คือ

3.1.1 กลุ่มการพยาบาล คือกลุ่มงานที่มีหน้าที่รับผิดชอบเกี่ยวกับการให้บริการด้านการพยาบาล ติดตาม และส่งต่อผู้ป่วย พัฒนาระบบ รูปแบบและวิธีการด้านการพยาบาล ซึ่งแบ่งงานออกดังนี้

3.1.1.1 งานผู้ป่วยนอกและงานเปล

3.1.1.2 งานอุบัติเหตุและฉุกเฉิน

3.1.1.3 งานผู้ป่วยใน

3.1.1.4 งานห้องคลอด

3.1.1.5 งานควบคุมการติดเชื้อ, จ่ายกลาง

3.1.1.6 งานห้องผ่าตัด

3.1.2 กลุ่มงานเวชศาสตร์ครอบครัวและบริการปฐมภูมิ คือกลุ่มงานที่ให้บริการด้านการเสริมสร้าง ป้องกัน รักษาโรคเบื้องต้นในระดับปฐมภูมิ และฟื้นฟูสุขภาพ การเฝ้าระวังและการควบคุมโรคติดต่อ ตลอดจนการส่งเสริมการมีส่วนร่วมของชุมชน เพื่อให้ผู้ป่วยและครอบครัวมีความรู้ ความเข้าใจในการดูแลสุขภาพได้ถูกต้องเหมาะสม

ซึ่งแบ่งงานออกดังนี้

3.1.2.1 งานเวชปฏิบัติ (โรคเรื้อรัง)

3.1.2.2 งานบริการปฐมภูมิ

3.1.2.3 งานส่งเสริมสุขภาพ

3.1.2.4 งานควบคุมโรค

3.1.2.5 งานสุขภาพจิตและจิตเวช

3.1.2.6 งานบำบัดยาเสพติด

3.1.2.7 งานสุขภาพภาคประชาชน

3.1.2.8 งานดูแลต่อเนื่องที่บ้าน

3.1.2.9 งานกายภาพบำบัด

3.1.2.10 งานการแพทย์แผนไทยและการแพทย์ทางเลือก

3.1.3 กลุ่มงานการจัดการ คือกลุ่มงานบริหารทั่วไปเป็นงานที่เกี่ยวข้องกับการจัดระบบบริหารองค์กร ให้บริการและบริหารงานอื่น ๆ ซึ่งแบ่งงานออกดังนี้

3.1.3.1 งานธุรการ, งานประชาสัมพันธ์และงานทรัพยากรบุคคล

3.1.3.2 งานการเงินและบัญชี

3.1.3.3 งานควบคุมภายในและตรวจสอบภายใน

3.1.3.4 งานพัสดุ, งานอาคารสถานที่, งานซ่อมบำรุง และสาธารณูปโภค

3.1.3.5 งานยานพาหนะ

3.1.3.6 งานรักษาความปลอดภัย

3.1.3.7 งานสวนสนาม, ระบบบำบัด

3.1.4 กลุ่มงานการแพทย์ คือกลุ่มที่มีหน้าที่ทำการวินิจฉัยโรคและรักษาผู้ป่วย ได้แก่ งานเวชศาสตร์ฉุกเฉินและงานเวชกรรมทั่วไป

3.1.5 กลุ่มงานเภสัชกรรมและคุ้มครองผู้บริโภค คือกลุ่มงานที่ให้การคุ้มครองผู้บริโภคด้านสาธารณสุข โดยจัดให้มียาและผลิตภัณฑ์ที่เกี่ยวข้องกับสุขภาพที่มีคุณภาพกระจายไปสู่ผู้บริโภคอย่างทั่วถึงโดยใช้กลไกการบริหาร รวมทั้งให้ความรู้และข้อมูลที่ถูกต้องแก่ประชาชนเพื่อให้สามารถเลือกใช้ยาและผลิตภัณฑ์ที่เกี่ยวข้องได้โดยเหมาะสม ซึ่งแบ่งงานออกดังนี้

3.1.5.1 งานเภสัชกรรม

3.1.5.2 งานคุ้มครองผู้บริโภค

3.1.6 กลุ่มงานทันตกรรม คือกลุ่มงานที่ทำหน้าที่ ตรวจ วินิจฉัย บำบัดรักษา ฟื้นฟูสภาพ ส่งเสริมและป้องกันทางทันตกรรม

3.1.7 กลุ่มงานเทคนิคการแพทย์และรังสีวิทยา เป็นกลุ่มงานที่ให้บริการตรวจวิเคราะห์ทางห้องปฏิบัติการ ที่ถูกต้อง แม่นยำ ทันเวลา ตามมาตรฐานวิชาชีพ เพื่อประกอบการวินิจฉัย ซึ่งแบ่งงานออกดังนี้

3.1.7.1 งานเทคนิคการแพทย์

3.1.7.2 งานรังสีวิทยา

3.1.8 กลุ่มงานประกันสุขภาพ ยุทธศาสตร์และสารสนเทศทางการแพทย์ เป็นกลุ่มงานที่ให้บริการเกี่ยวกับสิทธิการรักษาพยาบาลของประชาชนตามโครงการหลักประกันสุขภาพถ้วนหน้าเป็นแหล่งข้อมูลเรื่องสิทธิประโยชน์ในการรักษาพยาบาลที่ถูกต้องและทันสมัย และรักษาผลประโยชน์ของหน่วยงานที่พึงได้ ตลอดจนเป็นกลุ่มงานที่มีบทบาทในการจัดทำแผนมุ่งเน้นผลสัมฤทธิ์ และแผนงานด้านสุขภาพอื่น ๆ โดยใช้ข้อมูลการวิเคราะห์สภาวะสุขภาพจากสารสนเทศทางการแพทย์เพื่อให้ทุกหน่วยงานมีการวางแผนการปฏิบัติงานให้เป็นไปตามทิศทางและนโยบาย ซึ่งแบ่งงานออกดังนี้

3.1.8.1 งานเวชสถิติ

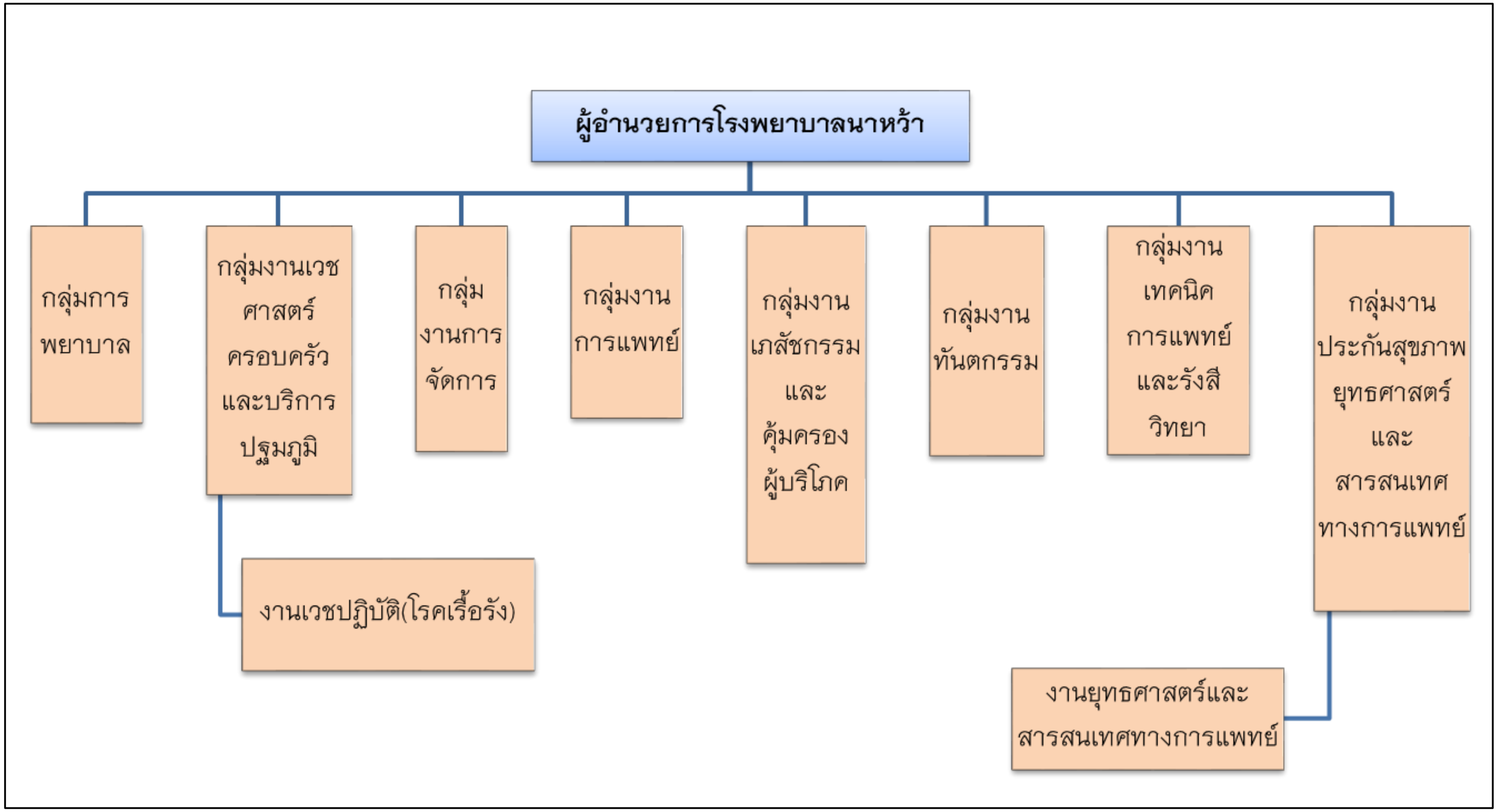
3.1.8.2 งานบริการเวชระเบียน

3.1.8.3 งานประกันสุขภาพ

3.1.8.4 งานยุทธศาสตร์และสารสนเทศทางการแพทย์

3.1.8.5 งานธุรกรรมจัดเก็บรายได้

ซึ่งแต่ละกลุ่มงานอยู่ภายใต้การบริหารของผู้อำนวยการโรงพยาบาลนาหว้า โดยแสดงรายละเอียดดังภาพประกอบ 17



ภาพประกอบ 17 โครงสร้างองค์กรโรงพยาบาลนาหว้า

วิสัยทัศน์ ของโรงพยาบาลนาหว้าคือ “เป็นโรงพยาบาลที่มีคุณภาพมาตรฐาน สิ่งแวดล้อมดี ผู้รับบริการประทับใจ”

พันธกิจของโรงพยาบาลนาหว้า ได้แก่

1. จัดบริการสุขภาพที่ได้มาตรฐานและมีคุณภาพ
2. พัฒนาการมีส่วนร่วมกับภาคีเครือข่าย ในการสร้างเสริมสุขภาพและจัดบริการสุขภาพ
3. พัฒนาระบบบริหารจัดการและกระบวนงานที่สำคัญ
4. พัฒนาบุคลากรให้มีความสุขในการทำงาน สุขภาพดีและมีสมรรถนะ

ตาราง 12 แผนยุทธศาสตร์โรงพยาบาลนาหว้า ปีงบ 2560 – 2564

พันธกิจ	ประเด็นยุทธศาสตร์	เป้าประสงค์
1. จัดบริการสุขภาพที่ได้มาตรฐานและมีคุณภาพ	1. การจัดบริการสุขภาพที่มีคุณภาพมาตรฐาน	1. โรงพยาบาลมีการจัดบริการสุขภาพที่ได้มาตรฐาน มีคุณภาพ ตอบสนองต่อปัญหาสุขภาพและความต้องการของประชาชนในพื้นที่
	2. การเฝ้าระวังโรคที่เป็นปัญหาสำคัญของพื้นที่	2. โรงพยาบาลมีการเฝ้าระวังโรคและภัยคุกคามที่สำคัญมีประสิทธิภาพ
2. พัฒนาการมีส่วนร่วมกับภาคีเครือข่าย ในการสร้างเสริมสุขภาพและจัดบริการสุขภาพ	3. การจัดการสุขภาพชุมชนสร้างเสริมสุขภาพประชาชนโดยมีส่วนร่วมของภาคีเครือข่าย	3. มีการจัดการสุขภาพชุมชน โดยการมีส่วนร่วมของภาคีเครือข่าย
3. พัฒนาระบบบริหารจัดการและกระบวนงานที่สำคัญ	4. ระบบจัดการความรู้และระบบสารสนเทศที่เอื้อการบริการและการบริหารงาน	4. เป็นโรงพยาบาลที่มีการจัดการสารสนเทศที่ได้มาตรฐาน
		5. โรงพยาบาลมีระบบการจัดการความรู้ที่เอื้อต่อการพัฒนาคุณภาพงาน
	5. ระบบการบริหารจัดการมีประสิทธิภาพ	6. โรงพยาบาลมีระบบการบริหารจัดการงานสนับสนุนบริการที่มีประสิทธิภาพ

ตาราง 12 (ต่อ)

พันธกิจ	ประเด็นยุทธศาสตร์	เป้าประสงค์
3. พัฒนาระบบบริหารจัดการและกระบวนงานที่สำคัญ	6. โรงพยาบาลมีระบบการบริหารงานยุทธศาสตร์อย่างถั่วถึง	7. เป็นโรงพยาบาลมีการบริหารยุทธศาสตร์สุขภาพเชื่อมโยงกันอย่างเป็นระบบ มีประสิทธิภาพ
	7. ระบบการบริหารการเงินที่มีประสิทธิภาพ	8. โรงพยาบาลที่มีระบบการเงินการคลังมีเสถียรภาพและประสิทธิภาพ
	8. ระบบโครงสร้างสิ่งแวดล้อมเอื้อต่อการรักษาและการทำงาน	9. เป็นหน่วยบริการที่มีระบบป้องกันการติดเชื้อและการจัดการสิ่งแวดล้อมที่ได้มาตรฐาน
	9. บุคลากรมีความผาสุกในการทำงาน มีคุณธรรม และมีสุขภาพที่ดี	10. บุคลากรมีความผาสุก มีจริยธรรมในการปฏิบัติงาน มีสวัสดิการที่เอื้อต่อการปฏิบัติงาน
4. พัฒนาบุคลากรให้มีความสุขในการทำงาน สุขภาพดีและมีสมรรถนะ	10. บุคลากรมีความผาสุกในการทำงาน มีคุณธรรมและมีสุขภาพที่ดี 11.บุคลากรมีสมรรถนะสอดคล้องตามภารกิจ	11. บุคลากรมีสุขภาพดี เหมาะสมกับงาน
		12. เป็นหน่วยบริการที่มีระบบการจัดการทรัพยากรบุคคลและเสริมสร้างสมรรถนะเหมาะสมกับงาน

3.2 นโยบายของโรงพยาบาลนาหว้า

3.2.1 หลักการสำคัญของนโยบาย

3.2.1.1 มุ่งเน้นประโยชน์ที่เกิดกับประชาชนและผู้มารับบริการให้ได้รับบริการที่มีมาตรฐาน ปลอดภัย มีความประทับใจ

3.2.1.2 มุ่งเน้นการพัฒนาคุณภาพโรงพยาบาลที่ต่อเนื่องและยั่งยืน

3.2.1.3 มุ่งเน้นการมีส่วนร่วมของทุกหน่วยในโรงพยาบาลและภาคส่วนต่างมีส่วนร่วมในการดูแลสุขภาพ

3.2.1.4 มุ่งเน้นการพัฒนาบุคลากรให้ทางานมีประสิทธิภาพมีความสุขและความภาคภูมิใจในงาน

3.2.1.5 มุ่งเน้นการบริหารทรัพยากรการเงินการคลังที่มีประสิทธิภาพคุ้มค่า

3.2.2 นโยบาย

3.2.2.1 พัฒนางานสาธารณสุขตามแนวพระราชดาริและโครงการเฉลิมพระเกียรติ

3.2.2.2 พัฒนาการดำเนินงานสุขภาพของในเครือข่าย คปสอ.นาหว้าและภาคีสุขภาพ ส่วนต่าง ๆ เพื่อเพิ่มการเข้าถึงบริการสุขภาพของประชาชน เพิ่มประสิทธิภาพการบริหารจัดการทรัพยากรร่วมกัน

3.2.2.3 พัฒนาสนับสนุนการมีส่วนร่วมในการดำเนินงานด้านสุขภาพของภาคส่วนต่าง ๆ

3.2.2.4 พัฒนาระบบบริการในโรงพยาบาลให้ผ่านการรับรองมาตรฐานคุณภาพ บริการด้วยความเป็นมิตร บริหารจัดการทรัพยากรอย่างมีประสิทธิภาพด้วยธรรมมาภิบาล

3.3 โรคไม่ติดต่อเรื้อรังในผู้สูงอายุของโรงพยาบาลนาหว้า

จากการจัดลำดับความสำคัญของปัญหาสาธารณสุขในพื้นที่อำเภอนาหว้าโดยใช้ข้อมูลการอัตราการป่วย อัตราการตายของประชากร พร้อมนำเกณฑ์การจัดลำดับความสำคัญของปัญหาสาธารณสุข โดยใช้เกณฑ์ดังนี้คือ ขนาดของปัญหา ความรุนแรงของปัญหา ความยากง่ายในการจัดการปัญหา ความตระหนักหรือการมีส่วนร่วมของผู้ที่เกี่ยวข้องของปัญหา และปัญหานั้นเป็นจุดเน้นของนโยบายของจังหวัด/หน่วยงาน ผลการวิเคราะห์ความสำคัญของปัญหาสาธารณสุขในพื้นที่พบว่า การเจ็บป่วยด้วยโรคเบาหวาน

โรคความดันโลหิตสูง มีคะแนนการลำดับความสำคัญของปัญหาสูงสุด (คณะกรรมการประสานงานสาธารณสุขระดับอำเภอนาหว้า, 2560)

ตาราง 13 ข้อมูลผู้ป่วยเบาหวานโรงพยาบาลนาหว้า

ผู้ป่วยเบาหวาน ปีงบประมาณ	ต่ำกว่า 50 ปี (ราย)	50-59 ปี (ราย)	กลุ่มผู้สูงอายุ		
			60-69 ปี (ราย)	70-79 ปี (ราย)	80 ปี ขึ้นไป (ราย)
2555	153	423	604	419	169
2556	190	462	633	436	154
2557	198	490	651	455	152
2558	234	526	679	446	142
2559	278	550	677	421	127

ตาราง 14 ข้อมูลผู้ป่วยความดันโลหิตสูงโรงพยาบาลนาหว้า

ผู้ป่วยความดันโลหิตสูง ปีงบประมาณ	ต่ำกว่า 50 ปี (ราย)	50-59 ปี (ราย)	กลุ่มผู้สูงอายุ		
			60-69 ปี (ราย)	70-79 ปี (ราย)	80 ปี ขึ้นไป (ราย)
2555	209	622	875	755	462
2556	220	605	845	733	407
2557	235	644	896	757	388
2558	272	730	967	784	393
2559	311	765	981	766	350

ตาราง 15 ข้อมูลผู้ป่วยหัวใจขาดเลือดโรงพยาบาลนาหว้า

ผู้ป่วยหัวใจขาดเลือด ปีงบประมาณ	ต่ำกว่า 50 ปี (ราย)	50–59 ปี (ราย)	กลุ่มผู้สูงอายุ		
			60–69 ปี (ราย)	70–79 ปี (ราย)	80 ปี ขึ้นไป (ราย)
2555	25	45	80	55	21
2556	24	43	71	64	20
2557	25	35	72	57	18
2558	16	31	58	50	16
2559	20	24	52	55	22

จากข้อมูลผู้ป่วยเบาหวาน ความดันโลหิตสูงและโรคหัวใจขาดเลือด 5 ปี ย้อนหลัง (ข้อมูลจากฐานข้อมูลโรงพยาบาลนาหว้า, 2560) อัตราผู้ป่วยด้วยโรคเบาหวานและโรคความดันโลหิตสูง มีแนวโน้มเพิ่มสูงขึ้น แต่โรคหัวใจขาดเลือดมีแนวโน้มต่ำลง ส่วนกลุ่มอายุที่พบมากที่สุดของโรคเบาหวาน โรคความดันโลหิตสูง และโรคหัวใจขาดเลือด คือกลุ่มผู้สูงอายุ 60 ปีขึ้นไป รองลงมาคือ กลุ่มอายุ 50–59 ปี และ กลุ่มต่ำกว่า 50 ปี ตามลำดับ

3.4 การจัดเก็บข้อมูลของโรงพยาบาลนาหว้า

ระบบฐานข้อมูลของโรงพยาบาลนาหว้า จัดทำขึ้นเพื่อจัดการดำเนินงานและจัดเก็บข้อมูล โดยการนำโปรแกรมฮอสเอกซ์พี (HOSxP) มาใช้สำหรับสถานพยาบาล สถานีอนามัยและโรงพยาบาล เพื่อเก็บข้อมูลผู้ป่วยและช่วยบุคลากรทางการแพทย์ในการให้บริการผู้ป่วย มีลักษณะเป็นเฮชไอเอส (Hospital Information System: HIS) ถูกพัฒนาขึ้นเมื่อปี พ.ศ. 2542 และเป็นโปรแกรมที่สามารถเชื่อมโยงโครงสร้างพื้นฐานด้านเครือข่ายทั้งสำนักงานสาธารณสุขจังหวัด โรงพยาบาลและสถานีอนามัย เพื่อยกระดับคุณภาพชีวิตของคนไทยและสังคมไทย ปัจจุบันโปรแกรมฮอสเอกซ์พี ถูกนำไปใช้ในโรงพยาบาลมากกว่า 500 แห่งทั่วประเทศไทย ทั้งโรงพยาบาลรัฐบาลและเอกชน
การทำงานของระบบได้แก่

3.4.1 ระบบประชาสัมพันธ์

3.4.2 ระบบเวชระเบียน

3.4.3 ระบบตรวจสอบสิทธิ

3.4.4 ระบบซักประวัติ

3.4.5 ระบบนัดหมาย

3.4.6 ระบบห้องทำงานแพทย์

3.4.7 ระบบงานห้องฉุกเฉิน

3.4.8 ระบบคลินิกพิเศษ

3.4.9 ระบบคัดกรองกลุ่มเสี่ยงเรื้อรัง

3.4.10 ระบบทันตกรรม

3.4.11 ระบบชันสูตร

3.4.12 ระบบรังสีรักษา

3.4.13 ระบบเวชศาสตร์ฟื้นฟู

3.4.14 ระบบแพทย์แผนไทย

3.4.15 ระบบเภสัชกรรม

3.4.16 ระบบการเงิน

3.4.17 ระบบห้องผ่าตัดและวิสัญญี

3.4.18 ระบบ Admission Center

3.4.19 ระบบผู้ป่วยใน

3.4.20 ระบบห้องคลอด

3.4.21 ระบบงานโภชนาการ

3.4.22 ระบบส่งเสริมสุขภาพและระบบ One Stop Service

3.4.23 ระบบงานสำรองข้อมูล

3.4.24 ระบบงานรายงาน

3.4.25 ระบบงานผู้ดูแลระบบ

3.4.26 ระบบส่งออกข้อมูล

ในการทำงานของแต่ละระบบงาน จะถูกบันทึกและจัดเก็บลงในฐานข้อมูลโรงพยาบาล ซึ่งโปรแกรมฮอสเอกซ์พีถูกออกแบบให้ทำงานโดยใช้คำสั่งเอสคิวแอล (SQL) จึงสามารถทำงานได้กับระบบจัดการฐานข้อมูล (Data Base Management System)

หลายตัวขึ้นอยู่กับความต้องการของผู้ใช้ เช่น MySQL, Microsoft SQL Server, PostgreSQL, Interbase/Firebird

จากความสามารถของระบบฐานข้อมูลโรงพยาบาลนาหว้า ดังที่ได้กล่าวไว้ข้างต้น ช่วยให้การทำงานของผู้ใช้ระบบเป็นไปด้วยความสะดวก ทั้งการสืบค้นข้อมูล การบันทึก และการแก้ไขมูลต่าง ๆ ลงในฐานข้อมูล ซึ่งมีแพทย์และพยาบาลเป็นผู้ประสานงานให้ความสำคัญในการดูแลเฝ้าระวังกลุ่มผู้ป่วยโรคไม่ติดต่อเรื้อรัง และการพัฒนาแนวทางการดูแล บริหารจัดการผู้ป่วยกลุ่มโรคไม่ติดต่อเรื้อรังอยู่เสมอโดยใช้ข้อมูลสารสนเทศจากฐานข้อมูลของโรงพยาบาล โดยมีการสร้างระบบรายงานตอบสนองต่อความต้องการของผู้ใช้งาน รวมไปถึงการเสนอรายงานข้อมูลให้ผู้บริหารอยู่ตลอด แต่สำหรับการค้นหาคัดกรองผู้ป่วยและการส่งเสริมป้องกันโรคให้กับประชาชนในพื้นที่ยังคงใช้ระบบคัดกรองจากการสำรวจซึ่งใช้เวลาและกำลังคนมาก อีกทั้งยังไม่มีการนำเอาข้อมูลที่อยู่ในระบบมาใช้ในเหมืองข้อมูลด้านการคัดกรองให้รวดเร็วยิ่งขึ้น (กรรณิการ์ อุดมกัน, สัมภาษณ์, 1 ตุลาคม 2561)

3.5 ตารางข้อมูลที่เกี่ยวข้อง

เนื่องจากระบบฐานข้อมูลของโรงพยาบาลนาหว้า มีตารางที่เกี่ยวข้องกับระบบงานในทุกระบบงานจำนวนมากกว่า 1,000 ตารางทั้งนี้ขึ้นอยู่กับการปรับปรุงรุ่นโปรแกรมฮอสเอกซ์พี ซึ่งจะเพิ่มหรือลดตารางตามการปรับปรุงรุ่น (กรรณิการ์ อุดมกัน, สัมภาษณ์, 1 ตุลาคม 2561)

จากการศึกษาสภาพระบบสารสนเทศของโรงพยาบาลนาหว้า มีระบบงานที่เกี่ยวข้องกับโรคไม่ติดต่อเรื้อรังในผู้สูงอายุ คือระบบเวชระเบียน ระบบซักประวัติ ระบบห้องทำงานแพทย์ และระบบชันสูตร ซึ่งระบบเหล่านี้มีข้อมูลที่ปัจจัยที่เกี่ยวข้องคือ เพศ อายุ การสูบบุหรี่ ดัชนีมวลกาย ความดันโลหิต น้ำตาลในเลือด ไขมันในเลือด และโรคไม่ติดต่อเรื้อรัง คือโรคเบาหวาน โรคความดันโลหิตสูง และโรคหัวใจขาดเลือด โดยมีตารางที่เกี่ยวข้องดังนี้

3.5.1 ตารางผู้ป่วย (Patient) เป็นตารางที่ใช้เก็บข้อมูลพื้นฐานของผู้มารับบริการในโรงพยาบาล

3.5.2 ตารางการรับบริการผู้ป่วยนอก (Vn_stat) เป็นตารางที่ใช้เก็บข้อมูลการรับบริการผู้ป่วยนอกที่มารับบริการที่โรงพยาบาล

3.5.3 ตารางการซักประวัติ (Opdscreen) เป็นตารางที่ใช้เก็บข้อมูลการซักประวัติ

3.5.4 ตารางการสั่งตรวจทางห้องปฏิบัติการ (Lab_order) เป็นตารางที่ใช้เก็บข้อมูลการสั่งตรวจและผลทางห้องปฏิบัติการ

3.5.5 ตารางการให้รหัสการวินิจฉัยผู้ป่วยนอก (Ovstdiag) เป็นตารางที่ใช้เก็บข้อมูลการให้รหัสการวินิจฉัยผู้ป่วยนอก

จากตารางข้างต้น ข้อมูลในตารางข้อมูลและปัจจัยที่ทำให้เกิดโรคที่ได้จากการทบทวนวรรณกรรม มีตัวแปรที่เกี่ยวข้องและจำเป็นในการพยากรณ์การเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุ ดังตาราง 16

ตาราง 16 ข้อมูลที่เกี่ยวข้องกับการพยากรณ์การเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุ

ชื่อตาราง	ชื่อฟิลด์	ชนิดข้อมูล	ความหมาย
patient	sex	char(1)	รหัสเพศ
vn_stat	age_y	smallint(6)	อายุ (ปี)
opdscreen	bmi	double(15)	ค่าดัชนีมวลกาย
opdscreen	smoking_type	int(11)	รหัสการสูบบุหรี่
opdscreen	bps	double(15)	ค่าความดันโลหิตซีสโตลิค
opdscreen	bpd	double(15)	ค่าความดันโลหิตไดแอสโตลิค
lab_order	lab_order_result	varchar(255)	ผลการตรวจทางห้องปฏิบัติการ
ovstdiag	icd10	varchar(9)	รหัสการวินิจฉัยโรค

4. งานวิจัยที่เกี่ยวข้อง

การศึกษางานวิจัยที่เกี่ยวข้องแบ่งเป็นงานวิจัยที่เกี่ยวข้องภายในประเทศและต่างประเทศ ดังนี้

4.1 ภายในประเทศ

วิธวินท์ แสงมณี, วีระวุฒ รัตนเจริญเลิศ, ณัฏฐภพ โพธิรัชต์ และเพียงฤทัย หนูสวัสดิ์ (2560) ทำวิทยานิพนธ์เรื่อง การสร้างโมเดลทำนายโอกาสการกลับมารักษาตัวซ้ำของผู้ป่วยโรคเบาหวาน โดยใช้เทคนิคเหมืองข้อมูล โดยมีวัตถุประสงค์เพื่อ วิเคราะห์ปัจจัยที่ส่งผลต่อโอกาสการกลับมารักษาซ้ำของผู้ป่วยโรคเบาหวาน,สร้างโมเดลสำหรับ

ทำนายโอกาสการกลับเข้ามารักษาตัวซ้ำของผู้ป่วยโรคเบาหวานโดยใช้เทคนิคเหมืองข้อมูล และเปรียบเทียบประสิทธิภาพของโมเดลที่สร้างจากเทคนิคต้นไม้ตัดสินใจ เทคนิคนาอีฟเบย์ และเทคนิคเพื่อนบ้านใกล้ที่สุด ซึ่งใช้กลุ่มข้อมูลตัวอย่างจากผู้ป่วยที่เป็นโรคเบาหวานในโรงพยาบาล ของสหรัฐอเมริกาตั้งแต่ปี พ.ศ. 2542 – 2551 เป็นเวลา 10 ปี จำนวน 101,766 แถว โดยใช้ขั้นตอนอ้างอิงตามวิธีพัฒนาระบบ (Cross Industry Standard Process for Data Mining: CRISP DM) โดยมีปัจจัยนำเข้าใช้ในการวิเคราะห์จำนวน 14 แอตทริบิวต์ ได้แก่ รหัสประจำตัวผู้ป่วย, เพศ, อายุ, น้ำหนัก, ประเภทของการเข้ารับการรักษา, ประเภทของการเข้ารับการรักษาต่อ, ประวัติการรักษาของผู้ป่วย, ระยะเวลาในการเข้ารับการรักษา, จำนวนครั้งในการเข้ารับการรักษาในห้องปฏิบัติการ, จำนวนครั้งในการเข้ารับการรักษาของผู้ป่วยนอกในปีก่อนหน้า, จำนวนครั้งในการเข้ารับการรักษาของผู้ป่วยในในปีก่อนหน้า, จำนวนครั้งในการเข้ารับการรักษาของฉุกเฉินของผู้ป่วย, จำนวนยาที่ได้รับ, จำนวนขั้นตอนของการรักษาและ โอกาสการกลับมารักษาตัวของผู้ป่วย ผลการดำเนินงานวิจัยพบว่า โมเดลที่สร้างจากเทคนิคต้นไม้ตัดสินใจ ให้ค่าความถูกต้องมากที่สุดคิดเป็นร้อยละ 85.50 ตัวอย่างการแปลผลข้อมูล เช่น ผู้ป่วยอายุ 0 – 50 เป็นเพศหญิง ช่วงน้ำหนัก 75 – 100 เข้ารับการรักษาแบบฉุกเฉิน มีกระบวนการรักษาคือ ออกจากโรงพยาบาลและไปรักษาโรงพยาบาลใกล้บ้าน และเป็นผู้ป่วยที่เข้ารับการรักษาแบบฉุกเฉิน มีผลการทำนายคือ YES หมายถึง มีโอกาสกลับเข้ามารักษาตัวซ้ำ ผู้ป่วยอายุ 81 – 100 เป็นเพศชาย ช่วงน้ำหนัก 0 – 50 เข้ารับการรักษาแบบฉุกเฉิน มีกระบวนการรักษา คือ รับการรักษาตัวต่อที่บ้าน และมีประวัติการรักษาตัวแบบฉุกเฉิน มีผลทำนายออกมาคือ NO หมายถึง ไม่ต้องกลับมารักษาตัวซ้ำอีก เป็นต้น จากผลงานวิจัยสามารนำความรู้ที่ได้จากการสร้างโมเดลไปเป็นแม่แบบในการพัฒนาระบบสนับสนุนการตัดสินใจ เพื่อใช้ในทางการแพทย์สมัยใหม่นำโมเดลไปประยุกต์ในการสร้างระบบช่วยสนับสนุนการตัดสินใจของแนวทางการรักษาของแพทย์และผู้ป่วยต่อไป

ทัศนนท์ ฐานประเสริฐกุล และกมล เกียรติเรืองกมลา (2558) ทำวิทยานิพนธ์เรื่อง การศึกษาสาเหตุการเกิดโรคความดันโลหิตสูง กรณีศึกษาโรงพยาบาลภาครัฐแห่งหนึ่งในประเทศไทยโดยมีวัตถุประสงค์เพื่อ ศึกษาแนวทางในการหาปัจจัยที่เป็นเหตุนำไปสู่การเกิดโรคความดันโลหิตสูง และเพื่อศึกษาความสัมพันธ์ของข้อมูลในการพัฒนาตัวแบบเพื่อใช้พยากรณ์โรค จากการวิเคราะห์กลุ่มตัวอย่างผู้ป่วยที่เป็นโรคความดันโลหิตสูงของโรงพยาบาลภาครัฐแห่งหนึ่ง จำนวน 140,538 ราย ด้วยเทคนิคเหมืองข้อมูล

ตามกรอบ คริสป์–ดีเอ็ม วิเคราะห์ร่วมกันของ 2 ขั้นตอนคือ การสร้างแบบจำลองการแบ่งกลุ่ม (Clustering) และการวิเคราะห์การถดถอยโลจิสติกทวิ (Binary Logistic Regression) ผลที่ได้จากงานวิจัยนี้คือ หลังจากทำการแบ่งกลุ่ม สามารถจำแนกกลุ่มโรคความดันโลหิตสูงได้เป็น 8 กลุ่มพร้อมกับทดสอบโมเดล พบว่าทั้ง 8 กลุ่ม มีโรคร่วมกับความดันโลหิตสูงที่สามารถพยากรณ์ได้อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 เช่น ภาวะไขมันในเลือดสูงเกินไม่ระบุรายละเอียด, เบาหวานชนิดที่ไม่ต้องพึ่งอินซูลินไม่มีภาวะแทรกซ้อน, โรคหัวใจจากหลอดเลือดแดงแข็ง และไตวายเรื้อรัง เป็นต้น ซึ่งการหาความสัมพันธ์ของการเกิดโรค ทำให้ทราบถึงปัจจัยใดที่ส่งผลให้เกิดโรคความดันโลหิตสูง

ชฎิภกรน์ ทรายหมอ (2557) ทำวิทยานิพนธ์เรื่อง การพัฒนาแบบจำลองพยากรณ์ภาวะแทรกซ้อนของโรคอื่นในผู้ป่วยโรคเบาหวาน โดยใช้เทคนิคเหมืองข้อมูล: กรณีโรงพยาบาลปทุมธานี โดยมีวัตถุประสงค์เพื่อ สร้างแบบจำลองพยากรณ์ภาวะแทรกซ้อนของโรคอื่นในผู้ป่วยโรคเบาหวาน และทดสอบแบบจำลองโดยใช้กับกลุ่มงานอายุรกรรม ห้องตรวจผู้ป่วยเบาหวาน โรงพยาบาลปทุมธานี วิธีการดำเนินการวิจัยโดยใช้ประวัติผลการตรวจรักษาผู้ป่วยเบาหวาน โรงพยาบาลปทุมธานี จำนวน 17,043 คน สร้างแบบจำลองพยากรณ์ภาวะแทรกซ้อน โดยการหาความสัมพันธ์ของข้อมูลระหว่างอัลกอริทึมเอพีออรี่ (Apriori Algorithm) และ อัลกอริทึมเอฟพี–กโรธ (FPGrowth Algorithm) ใช้โปรแกรมเวก้าวิเคราะห์ข้อมูลพยากรณ์ภาวะแทรกซ้อนโรคอื่นหลังจากที่ผู้ป่วยเป็นโรคเบาหวาน ผลการพยากรณ์ความสัมพันธ์ภาวะแทรกซ้อนของโรคเบาหวานหลังจากที่ผู้ป่วยเป็นโรคเบาหวาน พบว่าผู้ป่วยที่เป็นโรคเบาหวานชนิดที่ต้องพึ่งอินซูลิน ร่วมกับภาวะแทรกซ้อนทางตา แล้วจะเป็นโรคแทรกซ้อนดังนี้ โรคจอตาจากโรคเบาหวานชนิดไม่งอกขยายหรือเอ็นพีดีอาร์ คิดเป็นร้อยละ 65 โรคต้อกระจกในผู้สูงอายุ คิดเป็นร้อยละ 46 โรคจอตาจากโรคเบาหวานชนิดงอกขยายหรือพีดีอาร์ คิดเป็นร้อย 44

นพรัตน์ พจน์จิราภรณ์ (2557) ทำงานวิจัย เรื่องระบบสนับสนุนทางการแพทย์เพื่อวิเคราะห์โรคหลอดเลือดแดง โป่งพองโดยใช้เทคนิคเหมืองข้อมูล กรณีศึกษาการรักษาหลอดเลือดแดงโป่งพองผ่านสายสวน เพื่อหาความสัมพันธ์ระหว่างข้อมูลของผู้ป่วย ประวัติการเจ็บป่วยในอดีต และโรคหลอดเลือดแดงโป่งพอง วัตถุประสงค์ของการวิจัย คือ 1. เพื่อพัฒนาระบบสนับสนุนทางการแพทย์เพื่อวิเคราะห์โรคหลอดเลือดแดงใหญ่โป่งพอง 2. เพื่อประเมินประสิทธิภาพของอัลกอริทึมเพื่อวิเคราะห์โรคหลอดเลือดแดงใหญ่โป่งพอง 3. เพื่อประเมินประสิทธิภาพของระบบสนับสนุนทางการแพทย์เพื่อวิเคราะห์โรค

หลอดเลือดแดงใหญ่โป่งพอง ที่พัฒนาขึ้น 4. เพื่อประเมินความพึงพอใจของผู้ใช้งานระบบสนับสนุนทางการแพทย์เพื่อวิเคราะห์โรคหลอดเลือดแดงใหญ่โป่งพอง ที่พัฒนาขึ้น โดยมีการใช้ข้อมูลมาจากการตรวจคัดกรองผู้ป่วยโรคหลอดเลือดแดงโป่งพอง และผู้ป่วยที่เป็นโรคหลอดเลือดแดงโป่งพองที่ผ่านการรักษาผ่านสายสวน ตั้งแต่วันที่ 1 มกราคม 2557 ถึง 31 ธันวาคม 2557 โดยใช้ทฤษฎี เหมืองข้อมูลกับสารสนเทศทางการแพทย์ (Data mining in medical informatics) ทฤษฎีต้นไม้ตัดสินใจ (Decision Tree) นาอีฟเบย์ (Naïve Bayes) เครือข่ายประสาทเทียม (Neuron Network) การวัดประสิทธิภาพของโมเดลการจำแนกข้อมูล ซึ่งจากผลการทดลองของการศึกษาครั้งนี้ แสดงให้เห็นว่าโมเดลต้นไม้ตัดสินใจ มีประสิทธิภาพค่าความถูกต้องสูงที่สุดมีค่าเท่ากับ 98.106% และมีการพยากรณ์ที่เที่ยงตรงจากการพิจารณาค่าความคลาดเคลื่อนระหว่างค่าจริงและค่าที่พยากรณ์ได้ (Root Mean Squared Error: RMSE) เท่ากับ 0.110 หลังจากการพัฒนาแอพลิเคชันเว็บโดยใช้ J48 ประเมินประสิทธิภาพในการพยากรณ์ของระบบพบว่าระบบที่พัฒนาขึ้นสามารถพยากรณ์โรคได้ถูกต้อง 87.57% และผลการสำรวจที่ประสิทธิภาพโดยรวม พบว่าความพึงพอใจเฉลี่ยอยู่ที่ 4.45 ที่ส่วนเบี่ยงเบนมาตรฐานที่ 0.56

ธีรพงษ์ สังข์ศรี (2557) ทำงานวิจัยเรื่อง การวิเคราะห์พฤติกรรมสำหรับการเลือกสมัครสาขาวิชาเรียนและการเปรียบเทียบตัวแบบพยากรณ์จำนวนนักศึกษาใหม่โดยใช้เทคนิคเหมืองข้อมูล งานวิจัยนี้มีวัตถุประสงค์เพื่อวิเคราะห์พฤติกรรมสำหรับการเลือกสมัครสาขาวิชาเรียนของผู้เข้าศึกษาต่อโดยใช้เทคนิคการจัด กลุ่มข้อมูล (Clustering) และเทคนิคการหากฎความสัมพันธ์ของข้อมูล (Association Rule Mining: ARM) ในการจัดกลุ่มข้อมูลพฤติกรรมการเลือกสมัครสาขาวิชาของผู้สมัครศึกษาต่อและค้นหากฎความสัมพันธ์ของแต่ละกลุ่มพฤติกรรมผู้สมัคร เพื่อวิเคราะห์ปัจจัยที่ส่งผลต่อการเลือกสมัครเรียนในสาขาวิชาต่าง ๆ ในแต่ละกลุ่มพฤติกรรม และการเปรียบเทียบตัวแบบการพยากรณ์สำหรับทำนายจำนวนของนักศึกษาใหม่ ซึ่งมีการเปรียบเทียบประสิทธิภาพระหว่างตัวแบบพยากรณ์ที่ถูกพัฒนาขึ้นด้วยเทคนิคต้นไม้ช่วยตัดสินใจ (Decision Tree) กับเทคนิคโครงข่ายประสาทเทียม (Artificial Neural Network: ANN) โดยเลือกศึกษาข้อมูลของผู้สมัครเข้าศึกษาต่อ ประจำปีการศึกษา 2552–2555 มหาวิทยาลัยราชภัฏนครราชสีมา เป็นกรณีศึกษา โดยผลการวิจัยพบว่าการเปรียบเทียบตัวแบบพยากรณ์จำนวนนักศึกษาใหม่ โดยตัวแบบพยากรณ์ที่ถูกสร้างขึ้นด้วยเทคนิคต้นไม้ช่วยตัดสินใจ (Decision Tree) มีค่าความถูกต้องเท่ากับ 93.76% และตัวแบบที่ถูกสร้างขึ้น ด้วยเทคนิค

โครงข่ายประสาทเทียม (Artificial Neural Network) แบบเพอร์เซ็บตรอนหลายชั้น (Multi-layer Perceptron) มีค่าความถูก ต้องเท่ากับ 93.60%

เสกสรรค์ วิลัยลักษณ์ (2558) ทำวิทยานิพนธ์ เรื่อง การใช้เทคนิคเหมืองข้อมูลเพื่อพยากรณ์ผลการเรียนของนักเรียน โรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน ศูนย์วิจัยและพัฒนาการศึกษา มีวัตถุประสงค์เพื่อพัฒนาคลังข้อมูล และสร้างตัวแบบพยากรณ์ผลการเรียนของนักเรียน โรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกาแพงแสน ศูนย์วิจัยและพัฒนาการศึกษา โดยใช้ข้อมูลนักเรียนระดับมัธยมศึกษาระหว่างปีการศึกษา 2548 – 2556 เพื่อพัฒนาคลังข้อมูลโดยใช้โครงสร้างแบบสโนว์เฟลกสกีมาและนำเสนอรายงาน จากนั้นใช้ข้อมูลนักเรียนระดับมัธยมศึกษาปีที่ 4 ระหว่าง ปีการศึกษา 2553 – 2556 จำนวน 525 ระเบียน ประกอบด้วย 16 คุณลักษณะ มาสร้างตัวแบบพยากรณ์ผลการเรียนโดยใช้ชุดข้อมูล 2 แบบ คือ ข้อมูลแบบไม่จัดกลุ่ม (Original Data) และข้อมูลแบบจัดกลุ่ม (Cluster Data) จากนั้นนำไปผ่านกระบวนการคัดเลือกคุณลักษณะ (Feature Selection) ซึ่งใช้วิธีซีเอฟเอส (CFS) และวิธีหาค่าเกนความรู้ (IG) แล้วใช้เทคนิคเหมืองข้อมูลแบบโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน (MLP) ซัพพอร์ตเวกเตอร์แมชชีน (SVM) และต้นไม้ตัดสินใจ (Decision Tree) มาสร้างตัวแบบพยากรณ์และเปรียบเทียบตัวแบบ ด้วยการทดสอบประสิทธิภาพแบบ 10 คลอสวาลิเดชั่น (10-fold Cross Validation) ผลการวิจัยพบว่า คลังข้อมูลนักเรียนระดับมัธยมศึกษาที่ใช้งานโดยผู้บริหาร หัวหน้ากลุ่มสาระการเรียนรู้ และอาจารย์ประจำชั้นมีความพึงพอใจการใช้งานคลังข้อมูลอยู่ในระดับดี และในเหมืองข้อมูล พยากรณ์ พบว่าชุดข้อมูลแบบไม่จัดกลุ่มนำมาคัดเลือกคุณลักษณะด้วยวิธีซีเอฟเอส ร่วมกับเทคนิคโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน ให้ ค่าความถูกต้องสูงที่สุดที่ร้อยละ 94.48 และมีค่ารากที่สองของความคลาดเคลื่อนน้อยที่สุดที่ 0.1880 เหมาะสมสำหรับการสร้างระบบพยากรณ์ผลการเรียนของนักเรียน

วัฒนพัชร์ เก้าลิ้ม (2557) ทำงานวิจัยเรื่องการค้นหาคำสำคัญในเอกสารภาษาไทยโดยใช้เทคนิคการค้นหารูปแบบความสัมพันธ์ในเหมืองข้อมูล มีวัตถุประสงค์เพื่อการพัฒนาระบบสกัดคำสำคัญในบทความข่าวภาษาไทย โดยใช้เทคนิครูปแบบลำดับ (Sequential Pattern Mining) ร่วมกับอัลกอริทึมการจำแนกข้อมูลโดยใช้ต้นไม้ตัดสินใจ ซึ่งผู้วิจัยทำการเปรียบเทียบประสิทธิภาพของเทคนิคการหาคำสำคัญวิธีการอื่น ได้แก่ การใช้โครงข่ายประสาทเทียม การหาค่า Word Co-occurrence และการหาค่า TFIDF ผลการ

ทดลองแสดงให้เห็นแนวโน้มที่ว่า หากบทความของเอกสารที่เป็นบทความสั้น ๆ เช่น บทความของข่าวสารบนเว็บไซต์ การใช้เทคนิค Sequential Pattern Mining ร่วมกับอัลกอริทึมการจำแนกข้อมูลโดยใช้ต้นไม้ตัดสินใจให้ผลลัพธ์เป็นที่น่าพอใจและใกล้เคียงกับวิธีอื่น ๆ ข้างต้น

ชนิดา เสือเปีย (2557) ทำวิทยานิพนธ์ เรื่องการศึกษาสาเหตุการเกิดโรคเบาหวานชนิดที่ 2 ด้วยเทคนิคเหมืองข้อมูล กรณีศึกษาโรงพยาบาลรัฐบาลแห่งหนึ่ง โดยมีวัตถุประสงค์เพื่อศึกษาสาเหตุที่เป็นเหตุนำไปสู่การเกิดโรคเบาหวานชนิดที่ 2 และเพื่อศึกษาความสัมพันธ์ของข้อมูลในการนำไปพัฒนาเป็นตัวแบบ เพื่อใช้พยากรณ์โรคเบาหวานชนิดที่ 2 ซึ่งจะใช้ข้อมูลผู้ป่วยที่เป็นโรคเบาหวานชนิดที่ 2 ที่เข้ารับการรักษาที่โรงพยาบาลรัฐแห่งหนึ่งในเขตกรุงเทพมหานคร ระหว่างเดือนมกราคม 2552 – เดือนตุลาคม 2557 รวม 5 ปี ผู้ป่วยจำนวน 46,000 ราย งานวิจัยนี้ใช้เทคนิคเหมืองข้อมูล (Data Mining) ตามกรอบคริสป์–ดีเอ็ม ผู้วิจัยได้เริ่มจากการสร้างแบบจำลองการแบ่งกลุ่ม (Clustering) ด้วยวิธี Simple K–Means เพื่อใช้ในการจัดกลุ่มผู้ป่วยโรคเบาหวานชนิดที่ 2 จากนั้นนำผลที่ได้มาวิเคราะห์การถดถอยโลจิสติกทวิ (Binary Logistic Regression Analysis) เพื่อพยากรณ์โอกาสที่จะนำไปสู่พฤติกรรมของผู้ป่วยโรคเบาหวานชนิดที่ 2 ผลจากการวิจัยพบว่าลักษณะของผู้ป่วยโรคเบาหวานชนิดที่ 2 ของโรงพยาบาลรัฐแห่งหนึ่งสามารถแบ่งออกได้เป็น 8 กลุ่ม โดยผลลัพธ์ที่ได้จากแบบจำลองการแบ่งกลุ่ม ในแต่ละกลุ่มจะถูกนำไปวิเคราะห์ เพื่อใช้พยากรณ์การเกิดโรค

ศศิมา มณฑาสุวรรณ (2557) ทำวิทยานิพนธ์เรื่อง การพัฒนาระบบค้นหารหัส ICD–10 สำหรับงานเวชระเบียน โดยมีวัตถุประสงค์ เพื่อพัฒนาระบบค้นหารหัส ICD–10 สำหรับงานเวชระเบียน และประเมินการใช้งานระบบค้นหารหัส ICD–10 สำหรับงานเวชระเบียน ซึ่งระบบการจำแนกรหัส ICD–10 ถูกสร้างโดยโปรแกรม WEKA เลือกใช้อัลกอริทึม Naïve Bay นำมาประยุกต์ใช้ร่วมกับอัลกอริทึม Apiori เพื่อให้ได้ผลลัพธ์การค้นหาที่ดีขึ้น และจากผลการวิจัยพบว่า การเปรียบเทียบค่าความแม่นยำ อัลกอริทึม C4.5 ให้ค่าความแม่นยำร้อยละ 71.64 อัลกอริทึม Naïve Bay ให้ค่าความแม่นยำร้อยละ 79.82 และผลการทดสอบระบบ ได้ค่าความถูกต้อง ร้อยละ 86

4.2 ต่างประเทศ

Shubpreet Kaur (2015) ทำวิจัยเรื่องแนวโน้มในอนาคตของเหมืองข้อมูลในการทำนายโรคต่าง ๆ ในระบบการรักษาพยาบาลทางการแพทย์ ศึกษาการประยุกต์ใช้ในด้านการแพทย์และสาธารณสุข นำไปสู่การค้นพบความรู้ในฐานข้อมูล ทางด้านการแพทย์และการดูแลสุขภาพสำหรับการตัดสินใจทางคลินิกที่มีศักยภาพ เพื่อใช้ในการปรับปรุงคุณภาพการรักษาในโรงพยาบาลและเพิ่มอัตราการรอดชีวิตของผู้ป่วย การวินิจฉัยโรคเป็นหนึ่งในเครื่องมือที่ทำให้เหมืองข้อมูลประสบความสำเร็จ เหมืองข้อมูลมีความสำคัญกับการสำรวจความรู้จากฐานข้อมูลอย่างเป็นระบบของเทคนิคการค้นพบความรู้ในฐานข้อมูล เทคนิคเหมืองข้อมูลที่ใช้ในการวิจัยทางการแพทย์ในปัจจุบัน มีการอภิปรายเกี่ยวกับการวินิจฉัยโรคและการค้นพบรูปแบบการดูแลสุขภาพที่ซ่อนอยู่จากฐานข้อมูลที่เกี่ยวข้อง นอกจากนี้ยังมีการใช้เหมืองข้อมูลเพื่อค้นหาความสัมพันธ์ระหว่างเงื่อนไขสุขภาพกับโรค และยังกล่าวถึงเครื่องมือที่สามารถใช้สำหรับการประมวลผลและจัดหมวดหมู่ข้อมูล บทความนี้สรุปบทความด้านเทคนิคต่าง ๆ เกี่ยวกับการวินิจฉัยทางการแพทย์และการพยากรณ์โรค มุ่งเน้นไปที่การวิจัยในปัจจุบันที่กำลังดำเนินการโดยใช้เทคนิคเหมืองข้อมูลเพื่อปรับปรุงกระบวนการพยากรณ์โรค ให้แนวโน้มในอนาคตของเทคนิคการค้นหาความรู้ในฐานข้อมูลโดยใช้เครื่องมือเหมืองข้อมูลสำหรับการดูแลสุขภาพ นอกจากนี้ยังให้ประเด็นสำคัญและความท้าทายเกี่ยวกับเหมืองข้อมูลและการดูแลสุขภาพโดยทั่วไป การวิจัยพบว่ามีการใช้งานเหมืองข้อมูลจำนวนมากรวมถึงการวิเคราะห์ศูนย์ดูแลสุขภาพเพื่อการกำหนดนโยบายด้านสุขภาพที่ดีขึ้นการตรวจหาการระบาดของโรคและการเสียชีวิตจากโรงพยาบาลที่สามารถป้องกันได้

Thanigaivel Rajan และ Ramesh Kumar (2015) ทำวิจัยเรื่องทบทวนระบบทำนายโรคหัวใจโดยใช้เทคนิคเหมืองข้อมูลที่ใช้คอมพิวเตอร์ในการวิเคราะห์ชุดข้อมูลจำนวนมากและจากนั้นการแยกความหมายของข้อมูล คาดการณ์แนวโน้มในอนาคต ตัดสินใจเชิงรุกและขับเคลื่อนด้วยความรู้ เหมืองข้อมูลสามารถตอบคำถามที่มักใช้เวลานานในการแก้ไข จากจำนวนข้อมูลที่สร้างขึ้นเป็นจำนวนมาก การทำนายของโรคหัวใจมีความซับซ้อนมากจากการประมวลผลและวิเคราะห์ด้วยวิธีการแบบเดิม ดังนั้นเหมืองข้อมูลให้วิธีการและเทคโนโลยีในการแปลงข้อมูลเหล่านี้จากกองข้อมูลให้เป็นข้อมูลที่เป็นประโยชน์สำหรับการตัดสินใจ โดยการใช้เทคนิคเหมืองข้อมูลนั้นใช้เวลาน้อยและให้การทำนายของโรคด้วยความแม่นยำมากขึ้น ในบทความนี้จะสำรวจเอกสารที่แตกต่างกันซึ่ง

หนึ่งหรือมากกว่าหนึ่งขั้นตอนของเหมืองข้อมูลที่ใช้ในการทำนายโรคหัวใจ ผลจากการใช้เครือข่ายประสาทเทียมมีค่าเกือบ 100% ในกระดาษแผ่นเดียวและ เพื่อการทำนายโดยใช้ขั้นตอนวิธีเหมืองข้อมูลให้ผลลัพธ์ที่มีประสิทธิภาพ การใช้เทคนิคเหมืองข้อมูลการรักษาโรคหัวใจสามารถให้ผลการทำงานที่เชื่อถือได้

Babu Lakshmanan (2015) ทำวิจัยเรื่องเหมืองข้อมูลด้วยแผนภูมิการตัดสินใจในการประเมินรูปแบบประสิทธิผลของการบำบัดวัณโรคในปอด: เทคนิคการจัดกลุ่มและจำแนก เพื่อศึกษาเหมืองข้อมูลเป็นกระบวนการที่ช่วยในการเปิดเผยรูปแบบข้อมูลที่น่าสนใจในข้อมูลจำนวนมาก ขั้นตอนนี้กลายเป็นกิจกรรมที่เคยมีขึ้นในทุกด้านของการวิจัยด้านวิทยาศาสตร์การแพทย์โดยเฉพาะอย่างยิ่งในด้านการดูแลสุขภาพ เหมืองข้อมูลทำให้เกิดรูปแบบการซ่อนข้อมูลที่มีประโยชน์จากฐานข้อมูลขนาดใหญ่ ในเอกสารฉบับนี้ได้เสนอวิธีการสำหรับการเปิดรับและการจำแนกตามโปรแกรมเพื่อประเมินรูปแบบของประสิทธิภาพในการรักษาผู้ป่วยวัณโรคปอด (Pulmonary Tuberculosis หรือ PTB) วัณโรคเป็นโรคที่เกิดจาก Mycobacterium ซึ่งแพร่กระจายผ่านทางอากาศและกระทบร่างกายภูมิคุ้มกันต่ำได้อย่างง่ายดาย วิธีการของเราขึ้นอยู่กับการจำแนกกลุ่มและจำแนกประเภทที่ใช้จำแนกตามประเภทของการทดสอบความไวต่อยา (DST) ได้แก่ ความไวต่อยาทั้งหมดและความต้านทานต่อยาใดชนิดหนึ่ง อายุและน้ำหนักเป็นปัจจัยสำคัญที่ส่งผลต่อผู้ป่วยโรคปอดอุดกั้นเรื้อรังการใช้กลุ่มข้อมูลสองขั้นตอน (TSC) ใช้ในการจัดกลุ่มข้อมูลเป็นกลุ่มต่าง ๆ และกำหนดระดับตามอายุและน้ำหนักนอกจากนี้ยังมีการเปรียบเทียบระหว่างกลุ่มอายุและกลุ่มอายุ น้ำหนัก. ต่อมาได้มีการจัดทำอัลกอริทึมการจำแนกประเภทต่าง ๆ กันหลายชุดเพื่อสร้างแบบจำลองการจำแนกประเภทสุดท้ายตามโครงสร้างการตัดสินใจพร้อมกับวิธีการตรวจสอบแบบ K-fold cross ประสิทธิภาพการรักษาที่ดีที่สุดคือ 97.9% ตามรูปแบบที่กำหนดจากการจำแนกประเภทและการถดถอยต้นไม้ (CART) แนวทางที่เสนอจะช่วยให้แพทย์ในขั้นตอนการวางแผนการรักษาของตนสำหรับประเภทต่าง ๆ (ผ่านต้นไม้ตัดสินใจ) เพื่อค้นหาความสัมพันธ์ที่ซ่อนอยู่ในข้อมูล

Priyavadana V. (2015) ทำวิจัยเรื่องการศึกษาเปรียบเทียบการใช้ข้อมูลเหมืองในการวินิจฉัยโรค เพื่อศึกษาเหมืองข้อมูลได้กลายเป็นเทคนิคในการสกัดและค้นพบความรู้ใหม่ ๆ ในข้อมูลที่มีนัยสำคัญในคลังข้อมูลขนาดใหญ่เพื่อให้สามารถตัดสินใจและกำหนดกลยุทธ์ได้ดีขึ้น เป็นกระบวนการที่มีประสิทธิภาพและรวดเร็วในการสืบค้นข้อมูลการค้นคืนและการวิเคราะห์ความหมายจากมุมมองที่ต่างกัน เหมืองข้อมูลช่วยในการ

นำเสนอและสรุปข้อมูลที่ดึงออกมาจากคลังข้อมูลได้อย่างมีนัยสำคัญซึ่งสามารถนำมาใช้เพื่อเพิ่มประสิทธิภาพทางธุรกิจเพิ่มรายได้และลดค่าใช้จ่าย เป็นกระบวนการที่ดีที่สุดในการค้นหารูปแบบหรือความสัมพันธ์ระหว่างฐานข้อมูลเชิงสัมพันธ์ เหมืองข้อมูลได้เผยถึงความคุ้นเคยทางด้านการแพทย์และการดูแลสุขภาพสำหรับการตัดสินใจทางคลินิกที่มีศักยภาพในการปรับปรุงคุณภาพการรักษาในโรงพยาบาลและเพิ่มอัตราการรอดชีวิตของผู้ป่วย การวินิจฉัยโรคเป็นหนึ่งในแอพพลิเคชันที่เครื่องมือทำเหมืองข้อมูลกำลังสร้างผลลัพธ์ที่ประสบความสำเร็จ บทความนี้สรุปเทคนิคบางอย่างเกี่ยวกับการวินิจฉัยทางการแพทย์และการพยากรณ์โรค นอกจากนี้ยังได้มุ่งเน้นไปที่การวิจัยในปัจจุบันที่กำลังดำเนินการโดยใช้เทคนิคเหมืองข้อมูลเพื่อปรับปรุงกระบวนการพยากรณ์โรค

Dinh Thuan Nguyen (2013) ทำวิจัยเรื่องแนวทางเหมืองข้อมูลในการดูแลสุขภาพ: อัลกอริธึม K-means ที่ปรับปรุงใหม่ ปัจจุบันการประยุกต์ใช้ข้อมูลเหมืองแร่ในอุตสาหกรรมการดูแลสุขภาพเป็นสิ่งที่จำเป็น เหมืองข้อมูลนำชุดเครื่องมือและเทคนิคต่าง ๆ มาประยุกต์ใช้เพื่อค้นหารูปแบบที่ซ่อนอยู่ซึ่งจะทำให้บุคลากรทางการแพทย์เป็นแหล่งความรู้เพิ่มเติมในการตัดสินใจ รายละเอียดเพิ่มเติมการจัดกลุ่มผู้ป่วยที่มีสถานะเดียวกันช่วยในการค้นพบโรคใหม่ ๆ แต่จำนวนกลุ่มที่เหมาะสมมักไม่ชัดเจน บทความนี้จะตรวจสอบวิธีการที่มีอยู่สำหรับการเลือกจำนวนกลุ่มสำหรับอัลกอริทึม จากนั้นจะนำเสนออัลกอริธึมที่ได้รับการปรับปรุงเพื่อเรียนรู้ขณะจัดกลุ่ม สุดท้ายประเมินอัลกอริทึมนำไปใช้กับชุดข้อมูลของผู้ป่วยและผลลัพธ์แสดงให้เห็นถึงประสิทธิภาพ

Swapnil D. Raut (2013) ทำวิจัยเรื่องแนวทางเหมืองข้อมูลในการประยุกต์ใช้ในการดูแลสุขภาพ ในโลกปัจจุบันการดูแลสุขภาพถือเป็นปัจจัยสำคัญที่สุดที่มีผลต่อชีวิตมนุษย์ การจัดการฐานข้อมูลการดูแลสุขภาพเป็นเรื่องที่ท้าทายที่สุดในยุคนี้ เหมืองข้อมูลได้รับการใช้อย่างกว้างขวางและกว้างขวางโดยหลายองค์กรที่เกี่ยวข้องกับการดูแลสุขภาพ ในด้านการดูแลสุขภาพความต้องการข้อมูลเหมืองแร่เพิ่มขึ้นอย่างรวดเร็ว มีขั้นตอนวิธีเหมืองข้อมูลต่าง ๆ ที่ใช้ในฐานข้อมูลด้านการดูแลสุขภาพ บทความนี้นำเสนอภาพรวมของขั้นตอนวิธีเหมืองข้อมูลประเภทต่าง ๆ เช่น K-means และ D-stream Algorithm รวมถึงการศึกษาเปรียบเทียบขั้นตอนวิธีเหมืองข้อมูลเหล่านี้ จากการศึกษาพบว่าประสิทธิภาพและข้อจำกัดของขั้นตอนวิธีเหมืองข้อมูลเหล่านี้

Israa Ahmed Zriqat, Ahmad Mousa Altamimi, Mohammad Azzeh (2017) ทำวิจัยเรื่องการศึกษาเปรียบเทียบการทำนายโรคหัวใจ โดยใช้วิธีการจำแนกประเภทการทำ

เหมืองข้อมูล การปรับปรุงความแม่นยำในการตรวจหาโรคหัวใจได้รับการตรวจสอบโดยนักวิจัยหลายคน การปรับปรุงดังกล่าวเกิดจากค่าใช้จ่ายด้านการดูแลสุขภาพและการวินิจฉัยที่ผิดพลาด ด้วยเหตุนี้ จึงได้มีการเสนอวิธีการต่าง ๆ เพื่อวิเคราะห์ปัจจัยของโรคที่มีเป้าหมายเพื่อลดภาระของแพทย์ ลดต้นทุนทางการแพทย์และลดข้อผิดพลาด ในบทความนี้ มีจุดประสงค์เพื่อการพัฒนาระบบสนับสนุนการตัดสินใจทางการแพทย์อย่างชาญฉลาดที่มีประสิทธิภาพโดยใช้เทคนิคการทำเหมืองข้อมูล ในบริบทนี้ มีการใช้อัลกอริธึมการจำแนกประเภทการทำเหมืองข้อมูลจำนวน 5 ชุด พร้อมด้วยชุดข้อมูลขนาดใหญ่ ประเมินและวิเคราะห์ปัจจัยเสี่ยงที่เกี่ยวข้องกับโรคหัวใจในเชิงสถิติ เพื่อเปรียบเทียบประสิทธิภาพของตัวแยกประเภทที่นำไปใช้ ได้แก่เทคนิคนาอีฟเบย์ เทคนิคต้นไม้ตัดสินใจ เทคนิคดิสคริมิแนนต์ เทคนิคการสุ่มป่าไม้และเทคนิคซัพพอร์ตเวกเตอร์แมทชีน ตัวแยกประเภทที่เลือกได้ถูกนำมาใช้โดยใช้เครื่องมือแมตแล็บ (Matlab) กับชุดข้อมูลสองชุด ผลการทดลองแสดงให้เห็นว่าอัลกอริธึมการจำแนกประเภททั้งหมดเป็นแบบคาดการณ์และสามารถให้คำตอบที่ค่อนข้างถูกต้อง ซึ่งเทคนิคต้นไม้ตัดสินใจมีประสิทธิภาพดีกว่า ให้ความแม่นยำร้อยละ 99.0 ตามด้วยเทคนิคการสุ่มป่าไม้ ที่เป็นเช่นนี้เพราะทั้งสองเทคนิคมีกลไกที่ค่อนข้างเหมือนกัน

S. Senthil and B. Ayshwarya (2018) ทำวิจัยเรื่องการทำนายมะเร็งปอดโดยใช้เทคนิคที่มีคุณสมบัติเหมาะสมที่สุด โดยงานวิจัยนี้มีวัตถุประสงค์เพื่อหาวิธีการทำนายมะเร็งปอดการทำนาย การเปรียบเทียบประสิทธิภาพของอัลกอริทึม ได้แก่ เทคนิคเพื่อนบ้านใกล้ที่สุด เทคนิคนาอีฟเบย์ เทคนิคซัพพอร์ตเวกเตอร์แมทชีน เทคนิคโครงข่ายประสาทเทียม และ การเพิ่มประสิทธิภาพจับกลุ่มอนุภาคเทคนิคโครงข่ายประสาทเทียม (NN–Particle Swarm Optimization) จากผลการวิจัยพบว่าการเพิ่มประสิทธิภาพจับกลุ่มอนุภาคเทคนิคโครงข่ายประสาทเทียม ดีที่สุดมีค่าความถูกต้องร้อยละ 97.8

A. Rahman Khan,M. Rahman,J. Us Salehin,S. Islam and F. Rabbi. (2021) ทำวิจัยเรื่อง เทคนิคการขุดข้อมูลอย่างมีประสิทธิภาพสำหรับโรคหัวใจการทำนายและการวิเคราะห์เปรียบเทียบของอัลกอริทึมการจำแนกประเภท งานวิจัยนี้มีวัตถุประสงค์เพื่อวิเคราะห์เปรียบเทียบการทำนายโรคหัวใจโดยใช้เทคนิคนาอีฟเบย์ เทคนิคต้นไม้ตัดสินใจ เทคนิคเครือข่ายประสาทเทียมและเทคนิคการสุ่มป่าไม้ การวิจัยใช้ชุดข้อมูลดิบและเครื่องมือเวก้า สำหรับการวิเคราะห์ประสิทธิภาพ งานวิจัยนี้พบว่าเทคนิคดีที่สุดเป็นเทคนิคการสุ่มป่าไม้ โดยให้ค่าความถูกต้องร้อยละ 97.7049

จากการศึกษางานวิจัยเหมืองข้อมูลทางด้านการแพทย์ ผู้วิจัยจึงมีแนวคิดในการนำเทคนิคเหมืองข้อมูลมาช่วยแก้ไขปัญหาการคัดกรองโรคไม่ติดต่อเรื้อรังในผู้สูงอายุ ซึ่งคาดว่าจะช่วยลดปัญหาสุขภาพด้านโรคไม่ติดต่อเรื้อรังของผู้สูงอายุ และจากงานวิจัยที่เกี่ยวข้องในลักษณะที่คล้ายกันพบว่ามีการนำเทคนิคเหมืองข้อมูลหลากหลายเทคนิคมาใช้ในงานวิจัยทางการแพทย์ที่แตกต่างกัน ดังนั้นผู้วิจัยจะใช้เทคนิคเหมืองข้อมูลที่นิยมใช้มากที่สุดเพื่อสร้างแบบจำลองและใช้พยากรณ์ในทางการแพทย์ ซึ่งได้แก่ เทคนิคต้นไม้ตัดสินใจ เทคนิคนาอีฟเบย์ และ เทคนิคเพื่อนบ้านใกล้ที่สุด มาสร้างแบบจำลองและเปรียบเทียบตัวแบบ ด้วยการทดสอบประสิทธิภาพแบบครอสวาลิเดชั่นโดยแบ่งข้อมูลเป็น 10 ส่วน (10-Fold Cross-validation) ซึ่งจะทำให้ผลการวิจัยมีความน่าเชื่อถือมากขึ้น

บทที่ 3

วิธีการดำเนินการวิจัย

งานวิจัยเรื่อง แบบจำลองเพื่อพยากรณ์การเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุด้วยเทคนิคเหมืองข้อมูล: กรณีศึกษาโรงพยาบาลนาหว้า จังหวัดนครพนม มีการดำเนินการ 3 ระยะ โดยมีรายละเอียดดังนี้

ระยะที่ 1 วิเคราะห์และเตรียมข้อมูล

เป็นขั้นตอนศึกษาทำความเข้าใจกับเนื้อหาข้อมูลและเตรียมข้อมูลสำหรับการนำไปทำเหมืองข้อมูล โดยอ้างอิงจากกระบวนเหมืองข้อมูลของคริสป์-ดีเอ็ม (CRISP-DM) ประกอบด้วย 3 ขั้นคือ 1) ขั้นตอนการทำความเข้าใจการวิจัยและธุรกิจ 2) ขั้นตอนการทำความเข้าใจข้อมูล และ 3) ขั้นตอนการจัดเตรียมข้อมูล

1. ข้อมูล

ผู้วิจัยทำหนังสือขอความอนุเคราะห์ข้อมูลจากโรงพยาบาลนาหว้า จังหวัดนครพนม คือข้อมูลผู้รับบริการที่มีอายุ 60 ปีขึ้นไป และมารับการรักษาที่โรงพยาบาลนาหว้า จังหวัดนครพนม ตั้งแต่เดือนตุลาคม พ.ศ. 2554 ถึง เดือนกันยายน พ.ศ. 2562 จำนวน 262,655 ระเบียน

โดยผู้วิจัยได้เสนอโครงการวิจัยเพื่อพิจารณาจริยธรรมการวิจัยในมนุษย์กับคณะกรรมการจริยธรรมการวิจัยในมนุษย์ มหาวิทยาลัยราชภัฏสกลนครเรียบร้อยแล้ว เลขที่ใบรับรอง ๐๔๗/๒๕๖๓

2. เครื่องมือที่ใช้ในการวิจัย

โปรแกรมไมโครซอฟท์เอ็กเซล (Microsoft Excel) ใช้ในการเตรียมและจัดการข้อมูลให้เป็นรูปแบบไฟล์สำหรับเก็บข้อมูลในรูปแบบตาราง ใช้เครื่องหมายจุลภาค หรือ คอมม่า (,) ในการแบ่ง (Comma Separated Value) เพื่อนำเข้าโปรแกรมเวก้า (Weka)

3. วิธีการดำเนินการวิจัย

1) ขั้นตอนการทำความเข้าใจการวิจัยและธุรกิจ

ผู้วิจัยทำการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องกับโรคไม่ติดต่อเรื้อรังในผู้สูงอายุ 3 โรคได้แก่ โรคเบาหวาน โรคความดันโลหิตสูง และโรคหัวใจขาดเลือด การพยากรณ์การเกิดโรคด้วยเทคนิคเหมืองข้อมูล 3 เทคนิค ได้แก่ เทคนิคต้นไม้ตัดสินใจ เทคนิคนาอีฟเบย์ และเทคนิคเพื่อนบ้านใกล้ที่สุด

2) ขั้นตอนการทำความเข้าใจข้อมูล

ผู้วิจัยทำความเข้าใจข้อมูลและตัวแปรที่ก่อให้เกิดโรคเบาหวาน จำนวน 7 ตัวแปร ได้แก่ เพศ อายุ การสูบบุหรี่ ดัชนีมวลกาย ความดันโลหิต น้ำตาลในเลือด ไขมันในเลือด และตัวแปรผลลัพธ์ ได้แก่ เป็นหรือไม่เป็นโรค

3) ขั้นตอนการจัดเตรียมข้อมูล

นำข้อมูลในขั้นตอนที่ 2 มาทำความสะอาด (Cleansing) ด้วยโปรแกรมไมโครซอฟท์เอ็กเซล โดยตรวจสอบระเบียนข้อมูลที่มีค่าไม่ครบทุกฟิลด์หรือระเบียนที่มีค่าผิดปกติ (Outliner) เช่น อายุ 300 ปี น้ำหนัก 251 กิโลกรัม ผู้วิจัยแก้ไขโดยลบระเบียนที่มีค่าผิดปกติจำนวน 11 ระเบียน ดังนั้นเหลือระเบียบที่ใช้สร้างแบบจำลองจำนวน 262,644 ระเบียน จากนั้นทำการแปลงข้อมูลให้อยู่ในรูปแบบพร้อมใช้ในการวิเคราะห์

4. สถิติและการวิเคราะห์ข้อมูล

ผู้วิจัยใช้สถิติในการวิเคราะห์ข้อมูลโดยอ้างอิงจากกระบวนเหมืองข้อมูลของคริสป์–ดีเอ็ม (CRISP–DM) ในขั้นตอนการจัดเตรียมข้อมูล ได้แก่ การแปลงข้อมูลและทำให้ข้อมูลถูกต้อง (Data Cleaning) เช่น การแปลงข้อมูลให้อยู่ในช่วง (Scale) เดียวกัน และการตัดข้อมูลที่ขาดหายไป

ระยะที่ 2 สร้างแบบจำลอง

เป็นขั้นตอนสร้างแบบจำลองทั้ง 3 โรคคือโรคเบาหวาน โรคความดันโลหิตสูง และโรคหัวใจขาดเลือด ด้วยโปรแกรมเวก้า ด้วยเทคนิคเหมืองข้อมูล 3 เทคนิคคือ ต้นไม้ตัดสินใจ เพื่อนบ้านใกล้ที่สุด และนาอีฟเบย์

1. ข้อมูล

ข้อมูลที่ผ่านกระบวนการระยะที่ 1 จำนวน 262,644 ระเบียน ในรูปแบบไฟล์สำหรับเก็บข้อมูลในรูปแบบตาราง ใช้เครื่องหมายจุลภาค หรือคอมม่า (,) ในการแบ่ง (Comma Separated Value)

2. เครื่องมือที่ใช้ในการวิจัย

โปรแกรมเวก้า (Weka) ใช้ในการนำเข้าข้อมูลและสร้างแบบจำลอง

3. วิธีการดำเนินการวิจัย

ผู้วิจัยนำข้อมูลที่ผ่านกระบวนการระยะที่ 1 ไปสร้างแบบจำลองทั้ง 3 โรคคือ โรคเบาหวาน โรคความดันโลหิตสูง และโรคหัวใจขาดเลือด ด้วยโปรแกรมเวก้า ด้วยเทคนิคเหมืองข้อมูล 3 เทคนิคคือ ต้นไม้ตัดสินใจ เพื่อนบ้านใกล้ที่สุด และนาอีฟเบย์

ในการวัดประสิทธิภาพแบบจำลอง ผู้วิจัยได้ทำการทดลองตามหลักการแบ่งข้อมูลออกเป็นชุดเรียนรู้และชุดทดสอบ โดยการแบ่งข้อมูลออกเป็น 10 ส่วนเท่าๆ กัน ใช้ 9 ส่วนเป็นข้อมูลสอน และ 1 ส่วน เป็นข้อมูลทดสอบ ทำงานสลับกันจนครบ 10 รอบ (10 Fold Cross-Validation) การประเมินประสิทธิภาพใช้ตัววัดจากค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าความถ่วงดุล (F-Measure)

4. สถิติและการวิเคราะห์ข้อมูล

ผู้วิจัยใช้สถิติในการวิเคราะห์ข้อมูลของแบบจำลองทางเหมืองข้อมูลโดยวัดจากค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าความถ่วงดุล (F-Measure)

ระยะที่ 3 ประเมินประสิทธิภาพแบบจำลองโดยผู้เชี่ยวชาญ

เป็นขั้นตอนนำแบบจำลองแบบจำลองจากเทคนิคที่มีประสิทธิภาพดีที่สุดของโรคเบาหวาน โรคความดันโลหิตสูง และโรคหัวใจขาดเลือด ให้ผู้เชี่ยวชาญด้านสาธารณสุขประเมินประสิทธิภาพของแบบจำลอง ด้วยวิธีประชุมกลุ่ม (Focus Group) โดยมีรายละเอียดดังนี้

1. วิธีการจัดประชุมกลุ่ม

ผู้วิจัยนำแบบจำลองจากระยะที่ 2 มาเปรียบเทียบประสิทธิภาพการพยากรณ์การเกิดโรคเบาหวาน โรคความดันโลหิตสูง และโรคหัวใจขาดเลือด จากเทคนิค 3 เทคนิค แล้วเลือกค่าประสิทธิภาพดีที่สุด โดยอ้างอิงค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าความถ่วงดุล (F-Measure) ให้ผู้เชี่ยวชาญด้านสาธารณสุขที่มีประสบการณ์การทำงานด้านสาธารณสุขไม่น้อยกว่า 10 ปี หรือมีตำแหน่งชำนาญการขึ้นไป จำนวน 7 คน ประเมินประสิทธิภาพของแบบจำลองด้วยวิธีประชุมกลุ่ม (Focus Group)

2. เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการวิจัยระยะที่ 3 มีดังนี้

1) แบบจำลองจากเทคนิคที่มีประสิทธิภาพดีที่สุด โดยวัดจากค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าความถ่วงดุล (F–Measure)

2) ประเด็นการสนทนากลุ่ม (Focus group) แบ่งเป็น 5 ประเด็นคือ ความครบถ้วนของปัจจัยที่ใช้ในการพยากรณ์ ความถูกต้องของโมเดล ความน่าเชื่อถือของผลการทำนาย ความสามารถนำไปประกอบการตัดสินใจ ประโยชน์ของแบบจำลองการพยากรณ์การเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุ

โดยลักษณะเครื่องมือเป็นแบบประเมินประสิทธิภาพแบบจำลองสำหรับผู้เชี่ยวชาญ เป็นแบบมาตราส่วนประมาณค่า (Rating Scale) ตามแบบของลิเคอริท (Likert) แสดงทัศนะห้าช่วงคือ มากที่สุด มาก ปานกลาง น้อย และน้อยที่สุด โดยมีเกณฑ์ให้คะแนนดังนี้

มากที่สุด	หมายความว่า	มีความพึงพอใจมากที่สุด
มาก	หมายความว่า	มีความพึงพอใจมาก
ปานกลาง	หมายความว่า	มีความพึงพอใจปานกลาง
น้อย	หมายความว่า	มีความพึงพอใจน้อย
น้อยที่สุด	หมายความว่า	ไม่มีความพึงพอใจ

3. วิธีการดำเนินการวิจัย

1) ผู้วิจัยติดต่อผู้เชี่ยวชาญจำนวน 7 คน

2) ผู้วิจัยประสานและส่งรายชื่อผู้เชี่ยวชาญให้เจ้าหน้าที่สำนักงานบัณฑิตวิทยาลัยเพื่อจัดทำหนังสือเชิญผู้เชี่ยวชาญ

3) ผู้วิจัยนำส่งหนังสือเชิญให้กับผู้เชี่ยวชาญจำนวน 7 คน

4) ผู้วิจัยขอใช้และจัดเตรียมห้องสนทนากลุ่ม โดยใช้ห้องประชุมพุทธรักษา โรงพยาบาลนาหว้า จังหวัดนครพนม

5) ผู้วิจัยดำเนินการจัดกิจกรรมสนทนากลุ่ม โดยผู้วิจัยนำเสนอแบบจำลองจากเทคนิคที่มีประสิทธิภาพดีที่สุดและแจกแบบประเมินประสิทธิภาพแบบจำลองจากประเด็นการสนทนากลุ่ม 5 ประเด็น

6) ถอดเทปบันทึกการสนทนากลุ่ม

7) วิเคราะห์และบันทึกข้อเสนอแนะ

8) เขียนผลการวิจัย

4. สถิติและการวิเคราะห์ข้อมูล

ค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐาน ในการแปลผลคะแนนเฉลี่ยของผู้ประเมินประสิทธิภาพแบบจำลองมีเกณฑ์ ดังนี้ (บุญชม ศรีสะอาด, 2556)

ค่าเฉลี่ย	4.51 – 5.00	อยู่ในระดับมากที่สุด
ค่าเฉลี่ย	3.51 – 4.50	อยู่ในระดับมาก
ค่าเฉลี่ย	2.51 – 3.50	อยู่ในระดับปานกลาง
ค่าเฉลี่ย	1.51 – 2.50	อยู่ในระดับน้อย
ค่าเฉลี่ย	1.00 – 1.50	อยู่ในระดับน้อยที่สุด

บทที่ 4

ผลการวิเคราะห์ข้อมูล

งานวิจัยเรื่อง แบบจำลองเพื่อพยากรณ์การเกิดโรคเรื้อรังในผู้สูงอายุด้วยเทคนิคเหมืองข้อมูล: กรณีศึกษาโรงพยาบาลนาหว้า จังหวัดนครพนม มีจุดประสงค์การวิจัยคือ 1) เพื่อสร้างแบบจำลองเพื่อพยากรณ์การเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุ และ 2) เพื่อเปรียบเทียบประสิทธิภาพของแบบจำลองการเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุ

การสร้างแบบจำลองพยากรณ์การเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุ

ในการสร้างแบบจำลองพยากรณ์การเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุ ผู้วิจัยได้เตรียมข้อมูลเพื่อพร้อมใช้งานสำหรับการสร้างแบบจำลองของโรค 3 โรค ได้แก่ โรคเบาหวาน โรคความดันโลหิตสูง และโรคหัวใจขาดเลือด ด้วยเทคนิคเหมืองข้อมูล 3 เทคนิค ได้แก่ ต้นไม้ตัดสินใจ (Decision Tree) เพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbors) และนาอีฟเบย์ (Naïve Bayes) เพื่อหาว่าเทคนิคใดเหมาะสมกับการพยาการณ์การเกิดโรคไม่ติดต่อเรื้อรังแต่ละโรค โดยใช้โปรแกรมเวก้า (Weka) และวัดประสิทธิภาพด้วยวิธีแบ่งข้อมูลออกเป็นชุดเรียนรู้และชุดทดสอบ โดยการแบ่งข้อมูลออกเป็น 10 ส่วนเท่า ๆ กัน ใช้ 9 ส่วนเป็นข้อมูลสอน และ 1 ส่วน เป็นข้อมูลทดสอบ ทำงานสลับกันจนครบ 10 รอบ (10 Fold Cross-Validation) การประเมินประสิทธิภาพใช้ตัววัดจากค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าความถ่วงดุล (F-Measure)

1. การสร้างแบบจำลองพยากรณ์การเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุโรคเบาหวาน

1.1 การเตรียมข้อมูลโรคเบาหวาน

1) ข้อมูลที่ใช้ ผู้วิจัยนำข้อมูลผู้สูงอายุที่รับบริการรักษา จากโรงพยาบาลนาหว้า อ.นาหว้า จ.นครพนม ตั้งแต่เดือนตุลาคม พ.ศ. 2554 ถึง เดือนกันยายน พ.ศ. 2562 ประกอบไปด้วยข้อมูล เพศ อายุ การสูบบุหรี่ ดัชนีมวลกาย

ความดันโลหิต น้ำตาลในเลือด และไขมันในเลือด และข้อมูลผลการเป็นโรคไม่ติดต่อเรื้อรัง คือ โรคเบาหวาน ดังภาพประกอบ 18

	A	B	C	D	E	F	G	H	I
1	**Gender**	**Age**	**Bmi**	**Smoke**	**SBp**	**DBp**	**Fpg**	**Tg**	**DM**
2	2	72	5	1	130	80	130	198	E119
3	2	72	5	1	121	81	127	170	E119
4	2	72	5	1	119	75	112	144	E119
5	2	71	5	1	121	81	131	199	E119
6	2	71	5	1	132	81	119	201	E119
7	2	71	5	1	150	92	140	181	E119
8	2	71	5	1	166	89	127	157	E119
9	2	71	5	1	139	80	129	163	E119
10	1	79	4	1	110	65	91	138	-
11	1	79	2	1	115	78	92	146	-
12	1	79	2	1	118	67	99	120	-
13	1	79	2	1	121	81	87	111	-
14	1	79	2	1	130	88	86	147	-
15	1	79	2	1	110	67	108	137	-

ภาพประกอบ 18 ตัวอย่างข้อมูลผู้รับการรักษาและผลการเป็นโรคเบาหวาน

2) การแปลงข้อมูลผู้รับการรักษาโรคเบาหวานด้วยโปรแกรมไมโครซอฟต์เอกเซล (Microsoft Excel) ดังนี้

เพศ ข้อมูลมีค่าคือ 1 หมายถึงเพศชาย และ 2 หมายถึงเพศหญิง ผู้วิจัยแทนค่าเพศ 1 ด้วย M และ เพศ 2 ด้วย F

อายุ ผู้วิจัยมีการจัดกลุ่มใหม่ 3 กลุ่มดังนี้

กลุ่มที่ 1 ผู้สูงอายุวัยต้น (อายุ 60–69 ปี) แทนด้วย 1

กลุ่มที่ 2 ผู้สูงอายุวัยกลาง (อายุ 70–79 ปี) แทนด้วย 2

กลุ่มที่ 3 ผู้สูงอายุวัยปลาย (อายุ 80 ปีขึ้นไป) แทนด้วย 3

การสูบบุหรี่ ข้อมูลมีค่าคือ 1 หมายถึงไม่สูบบุหรี่ และ 2 หมายถึงสูบบุหรี่ ผู้วิจัยแทนค่าการสูบบุหรี่ 1 ด้วย N และ การสูบบุหรี่ 2 ด้วย Y

ดัชนีมวลกาย ผู้วิจัยมีการจัดกลุ่มใหม่ 5 กลุ่มดังนี้

กลุ่มที่ 1 คือค่า BMI น้อยกว่า 18.5 แทนด้วย 1

กลุ่มที่ 2 คือค่า BMI 18.5 ถึง 22.9 แทนด้วย 2

กลุ่มที่ 3 คือค่า BMI 23 ถึง 24.9 แทนด้วย 3

กลุ่มที่ 4 คือค่า BMI 25 ถึง 29.9 แทนด้วย 4

กลุ่มที่ 5 คือค่า BMI ตั้งแต่ 30 ขึ้นไป แทนด้วย 5

ความดันโลหิต ผู้วิจัยมีการจัดกลุ่มใหม่ 4 กลุ่มดังนี้

กลุ่มที่ 1 ค่าความดันซีสโตลิก (Systolic) น้อยกว่า 120 และค่าความดันไดแอสโตลิก (Diastolic) น้อยกว่า 80 แทนด้วย 1

กลุ่มที่ 2 ค่าความดันซีสโตลิก 120 ถึง 139 หรือ ค่าความดันไดแอสโตลิก 80 ถึง 89 แทนด้วย 2

กลุ่มที่ 3 ค่าความดันซีสโตลิก 140 ถึง 159 หรือ ค่าความดันไดแอสโตลิก 90 ถึง 99 แทนด้วย 3

กลุ่มที่ 4 ค่าความดันซีสโตลิกตั้งแต่ 160 ขึ้นไป หรือ ค่าความดันไดแอสโตลิกตั้งแต่ 100 ขึ้นไป แทนด้วย 4

น้ำตาลในเลือด ผู้วิจัยมีการจัดกลุ่มใหม่ 3 กลุ่มดังนี้

กลุ่มที่ 1 คือระดับน้ำตาลน้อยกว่า 110 แทนด้วย 1

กลุ่มที่ 2 คือระดับน้ำตาล 110 ถึง 125 แทนด้วย 2

กลุ่มที่ 3 คือระดับน้ำตาลตั้งแต่ 126 ขึ้นไป แทนด้วย 3

ไตรกลีเซอไรด์ ผู้วิจัยมีการจัดกลุ่มใหม่ 4 กลุ่มดังนี้

กลุ่มที่ 1 ค่าไตรกลีเซอไรด์ น้อยกว่า 150 แทนด้วย 1

กลุ่มที่ 2 ค่าไตรกลีเซอไรด์ 150 ถึง 199 แทนด้วย 2

กลุ่มที่ 3 ค่าไตรกลีเซอไรด์ 200 ถึง 499 แทนด้วย 3

กลุ่มที่ 4 ค่าไตรกลีเซอไรด์ ตั้งแต่ 500 ขึ้นไป แทนด้วย 4

โรคเบาหวาน ข้อมูลมีค่าตั้งแต่ E10 ถึง E14 หมายความว่าเป็นโรคเบาหวาน และ – หมายความว่าไม่เป็นโรคเบาหวาน ผู้วิจัยแทนค่าโรคเบาหวาน ตั้งแต่ E10 ถึง E14 ด้วย Y และ แทนค่าโรคเบาหวาน – ด้วย N

ผลจากการเตรียมข้อมูลสรุปภาพรวมเป็นตารางการจัดกลุ่มแทนค่าข้อมูลโรคเบาหวาน ดังแสดงในตาราง 17

ตาราง 17 การจัดกลุ่มแทนค่าข้อมูลโรคเบาหวาน

ชื่อฟิลด์	ความหมาย	ข้อมูลดิบ	การแทนค่า
Gender	เพศ	1	M
		2	F
Age	อายุ	60–69	1
		70–79	2
		>= 80	3
Bmi	ดัชนีมวลกาย	0–18.5	1
		18.5–22.9	2
		23–24.9	3
		25–29.9	4
		>= 30	5
Smoke	การสูบบุหรี่	1	N
		2	Y
Bp	ความดันโลหิต (ความดันซีสโตลิก และ ความดันไดแอสโตลิก)	SBP < 120 and DBP < 80	1
		SBP 120–139 or DBP 80–89	2
		SBP 140–159 or DBP 90–99	3
		SBP >= 160 or DBP >= 100	4
FPG	น้ำตาลในเลือด	< 110	1
		110 – 125	2
		>= 126	3
Tg	ไตรกลีเซอไรด์	< 150	1
		150 – 199	2
		200 – 499	3
		>= 500	4
Dm	ผลลัพธ์โรคเบาหวาน	E10 – E14	Y
		-	N

จากข้อมูลตาราง 17 แสดงตัวอย่างการแทนค่าข้อมูลได้ดังภาพประกอบ 19

	A	B	C	D	E	F	G	H
1	Gender	Age	Bmi	Smoke	Bp	FPG	Tg	Dm
262613	M	2	1	N	3	1	2	N
262614	F	2	3	N	2	3	2	Y
262615	F	2	2	N	2	3	2	Y
262616	F	1	5	N	2	1	1	N
262617	F	1	1	N	1	1	1	N
262618	F	1	3	N	1	1	1	N
262619	F	1	3	N	1	1	1	N
262620	F	1	3	N	1	1	1	N
262621	F	3	2	N	2	1	1	N
262622	F	3	2	N	2	1	1	N
262623	F	3	2	N	1	1	1	N
262624	F	3	2	N	2	1	1	N
262625	M	1	2	N	1	1	1	N
262626	M	1	2	N	1	1	1	N

ภาพประกอบ 19 ตัวอย่างข้อมูลโรคเบาหวานที่ผ่านการแปลงค่า

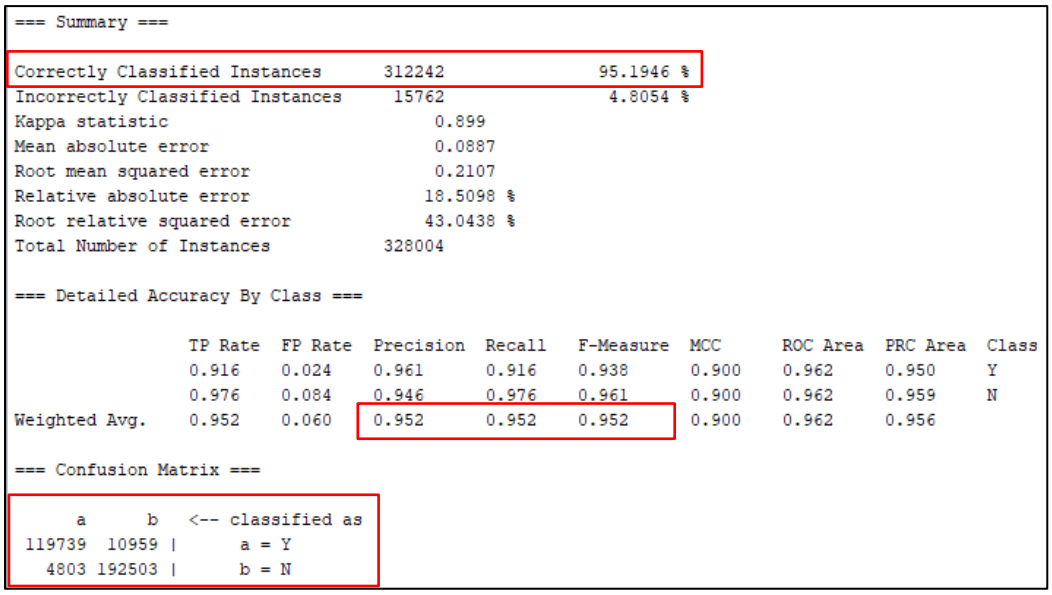
จากภาพประกอบ 19 ข้อมูลแถวที่ 262613 คอลัมน์ A ฟิลด์ Gender มีค่าเท่ากับ M หมายถึงเพศชาย คอลัมน์ B ฟิลด์ Age มีค่าเท่ากับ 2 หมายถึง ผู้สูงอายุที่มีอายุอยู่ระหว่าง 70 ถึง 79 คอลัมน์ C ฟิลด์ Bmi มีค่าเท่ากับ 2 หมายถึง ค่าดัชนีมวลกายระหว่างค่า 18.5 ถึง 22.9 คอลัมน์ D ฟิลด์ Smoke มีค่าเท่ากับ N หมายถึง ไม่สูบ คอลัมน์ E ฟิลด์ Bp มีค่าเท่ากับ 3 หมายถึง ค่าความดันโลหิตที่ค่า SBP อยู่ระหว่าง 140 ถึง 159 หรือ ค่า DBP อยู่ระหว่าง 90 ถึง 99 คอลัมน์ F ฟิลด์ FPG มีค่าเท่ากับ 1 หมายถึง ระดับน้ำตาลในเลือดมีค่าน้อยกว่า 110 คอลัมน์ G ฟิลด์ Tg มีค่าเท่ากับ 2 หมายถึง ระดับไขมันในเลือดอยู่ระหว่าง 150 ถึง 199 คอลัมน์ H ฟิลด์ Dm มีค่าเท่ากับ N หมายถึง ผลวินิจฉัยไม่เป็นโรคเบาหวาน

เมื่อทำการแปลงข้อมูลแล้วให้ทำการบันทึกเป็นนามสกุล CSV จากนั้นผู้วิจัยนำข้อมูลมาวิเคราะห์ด้วยโปรแกรมเวก้า

1.2 การสร้างแบบจำลองการพยากรณ์โรคเบาหวานสร้างแบบจำลองจาก 3 เทคนิค

1) ต้นไม้ตัดสินใจ

การสร้างแบบจำลองการพยากรณ์โรคเบาหวานด้วยเทคนิคต้นไม้ตัดสินใจพบว่ามีค่าความถูกต้อง (Accuracy) 95.1946% ค่าความแม่นยำ (Precision) 0.952 ค่าความระลึก (Recall) 0.952 และค่าความถ่วงดุล (F-Measure) 0.952 ดังภาพประกอบ 20



```
=== Summary ===

Correctly Classified Instances      312242               95.1946 %
Incorrectly Classified Instances     15762                4.8054 %
Kappa statistic                          0.899
Mean absolute error                      0.0887
Root mean squared error                  0.2107
Relative absolute error                 18.5098 %
Root relative squared error             43.0438 %
Total Number of Instances           328004

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                 0.916    0.024    0.961      0.916   0.938      0.900    0.962     0.950     Y
                 0.976    0.084    0.946      0.976   0.961      0.900    0.962     0.959     N
Weighted Avg.    0.952    0.060    0.952      0.952   0.952      0.900    0.962     0.956

=== Confusion Matrix ===

     a      b   <-- classified as
 119739  10959 |      a = Y
   4803 192503 |      b = N
```

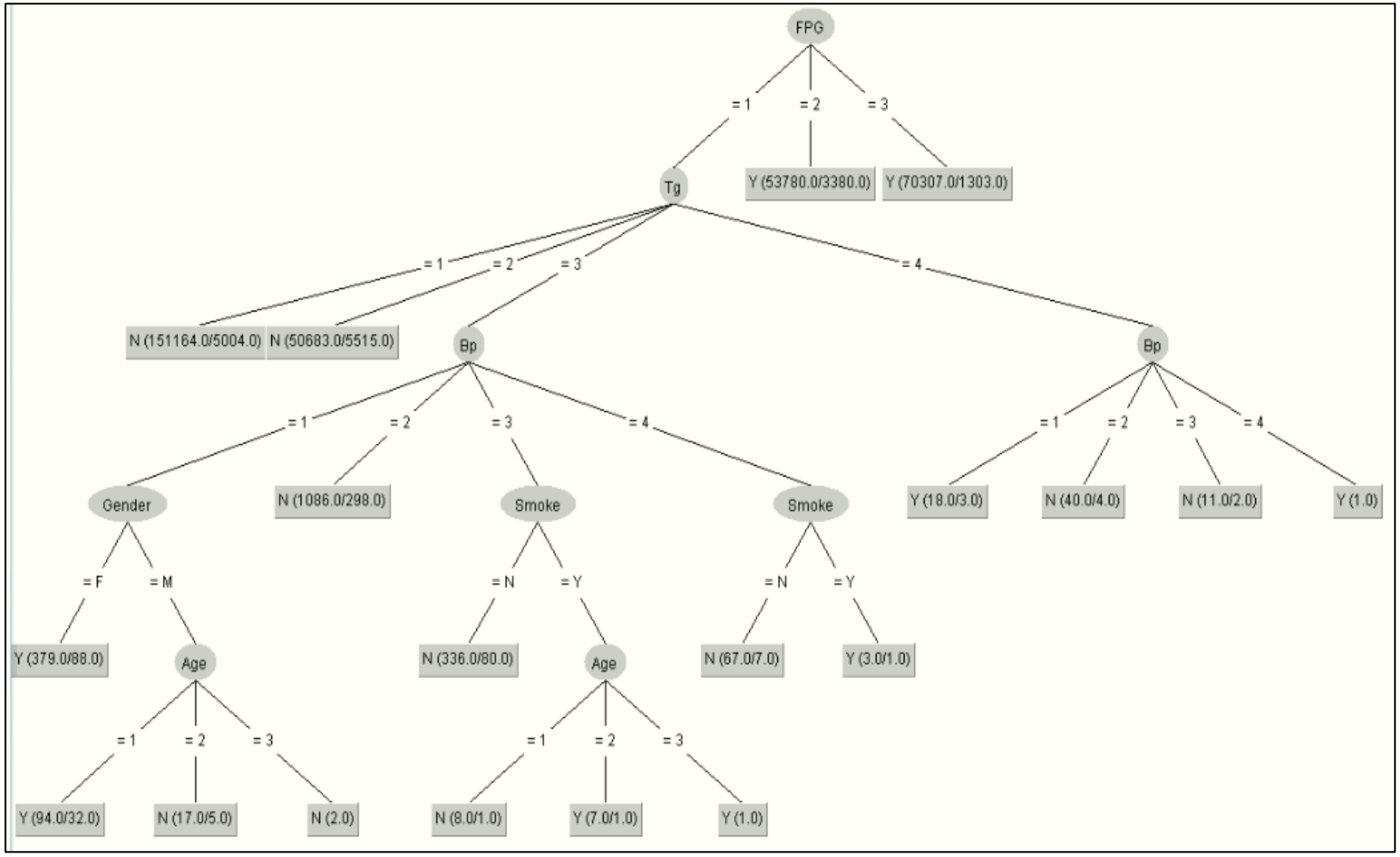
ภาพประกอบ 20 ผลลัพธ์ประสิทธิภาพแบบจำลองโรคเบาหวานของเทคนิคต้นไม้ตัดสินใจ

และจากภาพประกอบ 20 สามารถเขียนเป็นคอนฟิวชันเมทริกซ์ (Confusion Matrix) ได้ดังนี้

ตาราง 18 คอนฟิวชันเมทริกซ์ของแบบจำลองโรคเบาหวานของเทคนิคต้นไม้ตัดสินใจ

Predicted / Actual	Y	N
Y	119,739	10,959
N	4,803	192,503

จากตาราง 18 คอนฟิวชันเมทริกซ์ของแบบจำลองโรคเบาหวานของเทคนิคต้นไม้ตัดสินใจ ผลการทำนายได้ค่าความถูกต้องเชิงบวก (True Positive: TP) หรือ ทำนายว่าเป็นโรคเบาหวานและระเบียนผู้ป่วยป่วยเป็นโรคเบาหวานจริง คือ 119,739 ระเบียน ได้ค่าความถูกต้องเชิงลบ (True Negative: TN) หรือ ทำนายว่าไม่เป็นโรคเบาหวานและผู้ป่วยไม่ได้ป่วยเป็นโรคเบาหวานถูกต้องมี 192,503 ระเบียน ความผิดพลาดเชิงบวก (False Positive: FP) หรือ ผลการทำนายว่าเป็นโรคเบาหวาน แต่ผู้ป่วยไม่ได้ป่วยเป็นโรคเบาหวานมี 10,959 ระเบียน และความผิดพลาดเชิงลบ (False Negative: FN) หรือผลการทำนายว่าไม่เป็นโรคเบาหวาน แต่ผู้ป่วยป่วยเป็นโรคเบาหวานมี 4,803 ระเบียน



ภาพประกอบ 21 แบบจำลองโรคเบาหวานของเทคนิคต้นไม้ตัดสินใจ

จากภาพประกอบ 21 เป็นภาพจำลองต้นไม้ตัดสินใจของโรคเบาหวาน โดยการพิจารณาว่าผู้ป่วยเป็นโรคเบาหวานมีระดับการพิจารณา อยู่ 6 ระดับ คือ FPG Tg Bp Gender Smoke และ Age ซึ่งอธิบายกฎความสัมพันธ์ของแบบจำลอง ดังตาราง 19

ตาราง 19 อธิบายกฎแบบจำลองโรคเบาหวานของเทคนิคต้นไม้ตัดสินใจ

กฎความสัมพันธ์ของแบบจำลองโรคเบาหวาน (Rule)	**ผลลัพธ์ (Result)**
1. FPG = 1 and Tg = 1	N
2. FPG = 1 and Tg = 2	N
3. FPG = 1 and Tg = 3 and Bp = 1 and Gender = F	Y
4. FPG = 1 and Tg = 3 and Bp = 1 and Gender = M and Age = 1	Y
5. FPG = 1 and Tg = 3 and Bp = 1 and Gender = M and Age = 2	N
6. FPG = 1 and Tg = 3 and Bp = 1 and Gender = M and Age = 3	N
7. FPG = 1 and Tg = 3 and Bp = 2	N
8. FPG = 1 and Tg = 3 and Bp = 3 and Smoke = N	N
9. FPG = 1 and Tg = 3 and Bp = 3 and Smoke = Y and Age = 1	N
10. FPG = 1 and Tg = 3 and Bp = 3 and Smoke = Y and Age = 2	Y
11. FPG = 1 and Tg = 3 and Bp = 3 and Smoke = Y and Age = 3	Y
12. FPG = 1 and Tg = 3 and Bp = 4 and Smoke = N	N
13. FPG = 1 and Tg = 3 and Bp = 4 and Smoke = Y	Y
14. FPG = 1 and Tg = 4 and Bp = 1	Y
15. FPG = 1 and Tg = 4 and Bp = 2	N
16. FPG = 1 and Tg = 4 and Bp = 3	N
17. FPG = 1 and Tg = 4 and Bp = 4	Y
18. FPG = 2	Y
19. FPG = 3	Y

ตัวอย่าง กฎข้อที่ 1 ถ้าค่าน้ำตาลในเลือดน้อยกว่า 110 ค่าไขมันในเลือดน้อยกว่าหรือเท่ากับ 199 แล้วสถานะไม่เป็นโรคเบาหวาน

กฎข้อที่ 2 ถ้าค่าน้ำตาลในเลือดน้อยกว่า 110 ค่าไขมันในเลือด 200 ถึง 499 ค่าความดันโลหิตค่าความดันซีสโตลิกน้อยกว่า 120 ค่าความดันไดแอสโตลิกน้อยกว่า 80 และเป็นเพศหญิง แล้วสถานะเป็นโรคเบาหวาน

กฎข้อที่ 3 เมื่อค่าน้ำตาลในเลือดน้อยกว่า 110 ค่าไขมันในเลือด 200 ถึง 499 ค่าความดันโลหิตค่าความดันซีสโตลิกน้อยกว่า 120 ค่าความดันไดแอสโตลิกน้อยกว่า 80 เป็นเพศชาย และ อายุ 60 ถึง 69 ปี แล้วสถานะเป็นโรคเบาหวาน

กฎข้อที่ 4 เมื่อค่าน้ำตาลในเลือดน้อยกว่า 110 ค่าไขมันในเลือด 200 ถึง 499 ค่าความดันโลหิตค่าความดันซีสโตลิกน้อยกว่า 120 ค่าความดันไดแอสโตลิกน้อยกว่า 80 เป็นเพศชาย และอายุ 70 ปีขึ้นไป แล้วสถานะไม่เป็นโรคเบาหวาน เป็นต้น

2) เพื่อนบ้านใกล้ที่สุด

การสร้างแบบจำลองการพยากรณ์โรคเบาหวานด้วยเทคนิคเพื่อนบ้านใกล้ที่สุดพบว่ามีค่าความถูกต้อง (Accuracy) 95.1903% ค่าความแม่นยำ (Precision) 0.952 ค่าความระลึก (Recall) 0.952 และค่าความถ่วงดุล (F-Measure) 0.952 ดังภาพประกอบ 22

```
=== Summary ===

Correctly Classified Instances      312228               95.1903 %
Incorrectly Classified Instances     15776                4.8097 %
Kappa statistic                          0.8988
Mean absolute error                      0.0847
Root mean squared error                  0.2069
Relative absolute error                 17.6638 %
Root relative squared error             42.2646 %
Total Number of Instances           328004

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                 0.915    0.024      0.963      0.915     0.938      0.900    0.971     0.966     Y
                 0.976    0.085      0.945      0.976     0.961      0.900    0.971     0.973     N
Weighted Avg.    0.952    0.061      0.952      0.952     0.952      0.900    0.971     0.970

=== Confusion Matrix ===

     a      b   <-- classified as
 119570  11128 |      a = Y
   4648 192658 |      b = N
```

ภาพประกอบ 22 ผลลัพธ์ประสิทธิภาพแบบจำลองโรคเบาหวานของเทคนิคเพื่อนบ้านใกล้ที่สุด

และจากภาพประกอบ 22 สามารถเขียนเป็นคอนฟิวชันเมทริกซ์ (Confusion Matrix) ได้ดังนี้

ตาราง 20 คอนฟิวชันเมทริกซ์ของแบบจำลองโรคเบาหวานของเทคนิคเพื่อนบ้านใกล้ที่สุด

Predicted / Actual	Y	N
Y	119570	11128
N	4648	192658

จากตาราง 20 คอนฟิวชันเมทริกซ์ของแบบจำลองโรคเบาหวานของเทคนิคเพื่อนบ้านใกล้ที่สุด ได้ค่าความถูกต้องเชิงบวก (True Positive: TP) หรือ ทำนายว่าเป็นโรคเบาหวานและระเบียนผู้ป่วยป่วยเป็นโรคเบาหวานจริง คือ 119,739 ระเบียน ได้ค่าความถูกต้องเชิงลบ (True Negative: TN) หรือ ทำนายว่าไม่เป็นโรคเบาหวานและผู้ป่วยไม่ได้ป่วยเป็นโรคเบาหวานถูกต้องมี 119,570 ระเบียน ความผิดพลาดเชิงบวก (False Positive: FP) หรือ ผลการทำนายว่าเป็นโรคเบาหวาน แต่ผู้ป่วยไม่ได้ป่วยเป็นโรคเบาหวานมี 11,128 ระเบียน และความผิดพลาดเชิงลบ (False Negative: FN) หรือผลการทำนายว่าไม่เป็นโรคเบาหวาน แต่ผู้ป่วยป่วยเป็นโรคเบาหวานมี 4,648 ระเบียน

ตาราง 21 ตัวอย่างแบบจำลองโรคเบาหวานของเทคนิคเพื่อนบ้านใกล้ที่สุด

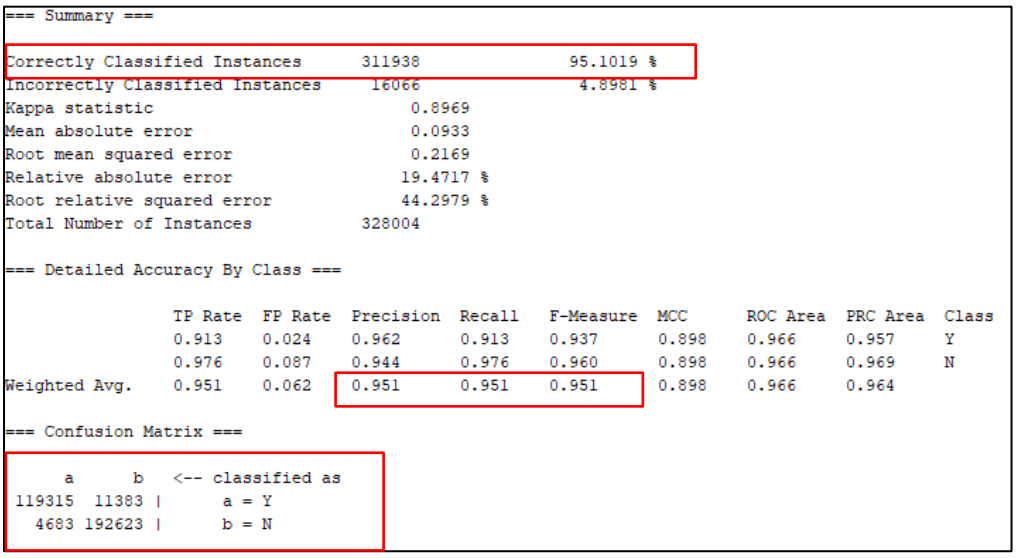
Data	Gender	Age	Bmi	Smoke	Bp	FPG	Tg	Distance	DM	Nearest Sign
D1	0	1	1	0	1	1	2	2.82	–	–
Distance $\boldsymbol{D1} = \sqrt{(\mathbf{1}-\mathbf{0})^2+(\mathbf{2}-\mathbf{1})^2+(\mathbf{2}-\mathbf{1})^2+(\mathbf{0}-\mathbf{0})^2+(\mathbf{3}-\mathbf{1})^2+(\mathbf{2}-\mathbf{1})^2+(\mathbf{2}-\mathbf{2})^2}$ = 2.82										
D2	1	2	1	0	1	2	4	3	+	
Distance $\boldsymbol{D2} = \sqrt{(\mathbf{1}-\mathbf{1})^2+(\mathbf{2}-\mathbf{2})^2+(\mathbf{2}-\mathbf{1})^2+(\mathbf{0}-\mathbf{0})^2+(\mathbf{3}-\mathbf{1})^2+(\mathbf{2}-\mathbf{2})^2+(\mathbf{2}-\mathbf{4})^2}$ = 3										
D3	1	1	1	0	1	1	3	2.82	+	+
Distance $\boldsymbol{D3} = \sqrt{(\mathbf{1}-\mathbf{1})^2+(\mathbf{2}-\mathbf{1})^2+(\mathbf{2}-\mathbf{1})^2+(\mathbf{0}-\mathbf{0})^2+(\mathbf{3}-\mathbf{1})^2+(\mathbf{2}-\mathbf{1})^2+(\mathbf{2}-\mathbf{3})^2}$ = 2.82										
D4	0	1	1	1	1	2	1	3	+	
Distance $\boldsymbol{D4} = \sqrt{(\mathbf{1}-\mathbf{0})^2+(\mathbf{2}-\mathbf{1})^2+(\mathbf{2}-\mathbf{1})^2+(\mathbf{0}-\mathbf{1})^2+(\mathbf{3}-\mathbf{1})^2+(\mathbf{2}-\mathbf{2})^2+(\mathbf{2}-\mathbf{1})^2}$ = 3										
D5	1	2	2	0	2	1	1	1.73	–	–
Distance $\boldsymbol{D5} = \sqrt{(\mathbf{1}-\mathbf{1})^2+(\mathbf{2}-\mathbf{2})^2+(\mathbf{2}-\mathbf{2})^2+(\mathbf{0}-\mathbf{0})^2+(\mathbf{3}-\mathbf{2})^2+(\mathbf{2}-\mathbf{1})^2+(\mathbf{2}-\mathbf{1})^2}$ = 1.73										
D6	0	1	4	0	2	1	1	3	–	
Distance $\boldsymbol{D6} = \sqrt{(\mathbf{1}-\mathbf{0})^2+(\mathbf{2}-\mathbf{1})^2+(\mathbf{2}-\mathbf{4})^2+(\mathbf{0}-\mathbf{0})^2+(\mathbf{3}-\mathbf{2})^2+(\mathbf{2}-\mathbf{1})^2+(\mathbf{2}-\mathbf{1})^2}$ = 3										
K=3	1	2	2	0	3	2	2		?	

จากตาราง 21 แบบจำลองโรคเบาหวานของเทคนิคเพื่อนบ้านใกล้ที่สุด ของตัวอย่างข้อมูลบางส่วนกำหนดให้ คลาส DM = Y แทนด้วย + และคลาส DM = N แทนด้วย –, Gender = F แทนด้วย 0 และ Gender = M แทนด้วย 1, Smoke = N แทนด้วย 0 และ Smoke = Y แทนด้วย 1 จะใช้วิธีการวัดระยะทางแบบยุคลิด (Euclidean Distance) โดยกำหนด K = 3 พบว่า คลาส DM = – ปรากฏมากที่สุด ดังนั้นจึงสรุปได้ว่าข้อมูล Gender = M, Age = 2, Bmi = 2, Smoke = N, Bp = 3, FPG = 2 และ Tg = 2 จะทำนายว่ามีคลาส DM = N

3) นาอีฟเบย์

การสร้างแบบจำลองการพยากรณ์โรคเบาหวานด้วยเทคนิคนาอีฟเบย์พบว่ามีค่าความถูกต้อง (Accuracy) 95.1019% ค่าความแม่นยำ (Precision) 0.951 ค่าความระลึก (Recall) 0.951 และค่าความถ่วงดุล (F–Measure) 0.951

ดังภาพประกอบ 23



```
=== Summary ===

Correctly Classified Instances      311938               95.1019 %
Incorrectly Classified Instances     16066                4.8981 %
Kappa statistic                          0.8969
Mean absolute error                      0.0933
Root mean squared error                  0.2169
Relative absolute error                 19.4717 %
Root relative squared error             44.2979 %
Total Number of Instances           328004

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                 0.913    0.024     0.962      0.913     0.937     0.898    0.966     0.957     Y
                 0.976    0.087     0.944      0.976     0.960     0.898    0.966     0.969     N
Weighted Avg.    0.951    0.062     0.951      0.951     0.951     0.898    0.966     0.964

=== Confusion Matrix ===

     a      b   <-- classified as
 119315  11383 |      a = Y
   4683 192623 |      b = N
```

ภาพประกอบ 23 ผลลัพธ์ประสิทธิภาพแบบจำลองโรคเบาหวานของเทคนิคนาอีฟเบย์

และจากภาพประกอบ 23 สามารถเขียนเป็นคอนฟิวชันเมทริกซ์ (Confusion Matrix) ได้ดังนี้

ตาราง 22 คอนฟิวชันเมทริกซ์ของแบบจำลองโรคเบาหวานของเทคนิคนาอีฟเบย์

Predicted / Actual	Y	N
Y	119315	11383
N	4683	192623

จากตาราง 22 คอนฟิวชันเมทริกซ์ของแบบจำลองโรคเบาหวานของเทคนิคนาอีฟเบย์ ได้ค่าความถูกต้องเชิงบวก (True Positive: TP) หรือ ทำนายว่าเป็นโรคเบาหวานและระเบียนผู้ป่วยป่วยเป็นโรคเบาหวานจริง คือ 119,315 ระเบียน ได้ค่าความถูกต้องเชิงลบ (True Negative: TN) หรือ ทำนายว่าไม่เป็นโรคเบาหวานและผู้ป่วยไม่ได้ป่วยเป็นโรคเบาหวานถูกต้องมี 192,623 ระเบียน ความผิดพลาดเชิงบวก (False Positive: FP) หรือผลการทำนายว่าเป็นโรคเบาหวาน แต่ผู้ป่วยไม่ได้ป่วยเป็นโรคเบาหวานมี 11,383 ระเบียน และความผิดพลาดเชิงลบ (False Negative: FN) หรือผลการทำนายว่าไม่เป็นโรคเบาหวานแต่ผู้ป่วยป่วยเป็นโรคเบาหวานมี 4,683 ระเบียน

ตาราง 23 แบบจำลองโรคเบาหวานของเทคนิคนาอีฟเบย์

แอตทริบิวต์	คลาส DM = Y	คลาส DM = N	ความน่าจะเป็น คลาส DM = Y	ความน่าจะเป็น คลาส DM = N
Gender = F	94101	108402	0.72	0.55
Gender = M	36599	88906	0.28	0.45
Age = 1	80279	103081	0.61	0.52
Age = 2	41687	70107	0.32	0.36
Age = 3	8735	24121	0.07	0.12
Bmi = 1	11823	43285	0.09	0.22
Bmi = 2	48789	78540	0.37	0.40
Bmi = 3	25974	31033	0.20	0.16
Bmi = 4	34919	36526	0.27	0.18
Bmi = 5	9198	7927	0.07	0.04
Smoke = N	125166	188723	0.96	0.96
Smoke = Y	5534	8585	0.04	0.04
Bp = 1	42792	80588	0.33	0.41
Bp = 2	68750	92648	0.52	0.47
Bp = 3	17128	20030	0.13	0.10
Bp = 4	2032	4044	0.02	0.02
Fpg = 1	11295	192624	0.09	0.98
Fpg = 2	50401	3381	0.38	0.02
Fpg = 3	69005	1304	0.53	0.01
Tg = 1	42709	148660	0.33	0.75
Tg = 2	81597	46364	0.62	0.24
Tg = 3	5933	2187	0.05	0.01
Tg = 4	463	99	0.00	0.00

จากตาราง 23 แบบจำลองโรคเบาหวานของเทคนิคนาอีฟเบย์ สามารถนำค่าไปหาความน่าจะเป็นโดยใช้ทฤษฎีของเบย์ (Bayes Theorem) ตามสมการ

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

เช่น Gender = F, Age = 1, Bmi = 1, Smoke = N, Bp = 1, Fpg = 1, Tg = 1 และ DM = ? สามารถคำนวณได้ดังนี้

$$\begin{aligned}
P(DM = Y|A) &= P(Gender = F \mid DM = Y) \times P(Age = 1 \mid DM = Y) \\
&\quad \times P(Bmi = 1 \mid DM = Y) \times P(Smoke = N \mid DM = Y) \\
&\quad \times P(Bp = 1 \mid DM = Y) \times P(Fpg = 1 \mid DM = Y) \\
&\quad \times P(Tg = 1 \mid DM = Y) \\
&= 0.72 \times 0.61 \times 0.09 \times 0.96 \times 0.33 \times 0.09 \times 0.33 \\
&= 0.00037
\end{aligned}$$

$$\begin{aligned}
P(DM = N|A) &= P(Gender = F \mid DM = N) \times P(Age = 1 \mid DM = N) \\
&\quad \times P(Bmi = 1 \mid DM = N) \times P(Smoke = N \mid DM = N) \\
&\quad \times P(Bp = 1 \mid DM = N) \times P(Fpg = 1 \mid DM = N) \\
&\quad \times P(Tg = 1 \mid DM = N) \\
&= 0.55 \times 0.52 \times 0.22 \times 0.96 \times 0.41 \times 0.98 \times 0.75 \\
&= 0.01820
\end{aligned}$$

ดังนั้นข้อมูล Gender = F, Age = 1, Bmi = 1, Smoke = N, Bp = 1, Fpg = 1 และ Tg = 1 จะทำนายว่ามีคลาส DM = N เนื่องจากค่า P(DM = N|A) = 0.01820 ซึ่งมีค่ามากกว่า P(DM = Y|A) = 0.00037

2. การสร้างแบบจำลองพยากรณ์การเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุโรคความดันโลหิตสูง

2.1 การเตรียมข้อมูลโรคความดันโลหิตสูง

1) ข้อมูลที่ใช้ ผู้วิจัยนำข้อมูลผู้สูงอายุที่รับบริการรักษา จากโรงพยาบาลนาหว้า อ.นาหว้า จ.นครพนม ตั้งแต่เดือนตุลาคม พ.ศ. 2554 ถึง เดือนกันยายน พ.ศ. 2562 ประกอบไปด้วยข้อมูล เพศ อายุ การสูบบุหรี่ ดัชนีมวลกาย ความดันโลหิต น้ำตาลในเลือด และไขมันในเลือด และข้อมูลผลการเป็นโรคไม่ติดต่อเรื้อรัง คือ โรคความดันโลหิตสูง ดังภาพประกอบ 24

	A	B	C	D	E	F	G	H	I
1	**Gender**	**Age**	**Bmi**	**Smoke**	**SBp**	**DBp**	**Fpg**	**Tg**	**HT**
2	2	72	5	1	130	80	130	198	I10
3	2	72	5	1	121	81	127	170	-
4	2	72	5	1	119	75	112	144	-
5	2	71	5	1	121	81	131	199	-
6	2	71	5	1	132	81	119	201	-
7	2	71	5	1	150	92	140	181	-
8	2	71	5	1	166	89	127	157	I10
9	2	71	5	1	139	80	129	163	I10
10	1	79	4	1	110	65	91	138	-
11	1	79	2	1	115	78	92	146	-
12	1	79	2	1	118	67	99	120	-
13	1	79	2	1	121	81	87	111	-
14	1	79	2	1	130	88	86	147	-
15	1	79	2	1	110	67	108	137	-

ภาพประกอบ 24 ตัวอย่างข้อมูลผู้รับการรักษาและผลการเป็นโรคความดันโลหิตสูง

2) การแปลงข้อมูลผู้รับการรักษาโรคความดันโลหิตสูงด้วยโปรแกรมไมโครซอฟต์เอกเซล (Microsoft Excel) ดังนี้

เพศ ข้อมูลมีค่าคือ 1 หมายถึงเพศชาย และ 2 หมายถึงเพศหญิง ผู้วิจัยแทนค่าเพศ 1 ด้วย M และ เพศ 2 ด้วย F

อายุ ผู้วิจัยมีการจัดกลุ่มใหม่ 3 กลุ่มดังนี้

กลุ่มที่ 1 ผู้สูงอายุวัยต้น (อายุ 60–69 ปี) แทนด้วย 1

กลุ่มที่ 2 ผู้สูงอายุวัยกลาง (อายุ 70–79 ปี) แทนด้วย 2

กลุ่มที่ 3 ผู้สูงอายุวัยปลาย (อายุ 80 ปีขึ้นไป) แทนด้วย 3

การสูบบุหรี่ ข้อมูลมีค่าคือ 1 หมายถึงไม่สูบบุหรี่ และ 2 หมายถึงสูบบุหรี่ ผู้วิจัยแทนค่าการสูบบุหรี่ 1 ด้วย N และ การสูบบุหรี่ 2 ด้วย Y

ดัชนีมวลกาย ผู้วิจัยมีการจัดกลุ่มใหม่ 5 กลุ่มดังนี้

กลุ่มที่ 1 คือค่า BMI น้อยกว่า 18.5 แทนด้วย 1

กลุ่มที่ 2 คือค่า BMI 18.5 ถึง 22.9 แทนด้วย 2

กลุ่มที่ 3 คือค่า BMI 23 ถึง 24.9 แทนด้วย 3

กลุ่มที่ 4 คือค่า BMI 25 ถึง 29.9 แทนด้วย 4

กลุ่มที่ 5 คือค่า BMI ตั้งแต่ 30 ขึ้นไป แทนด้วย 5

ความดันโลหิต ผู้วิจัยมีการจัดกลุ่มใหม่ 4 กลุ่มดังนี้

กลุ่มที่ 1 ค่าความดันซีสโตลิก (Systolic) น้อยกว่า 120 และค่าความดันไดแอสโตลิก (Diastolic) น้อยกว่า 80 แทนด้วย 1

กลุ่มที่ 2 ค่าความดันซีสโตลิก 120 ถึง 139 หรือ ค่าความดันไดแอสโตลิก 80 ถึง 89 แทนด้วย 2

กลุ่มที่ 3 ค่าความดันซีสโตลิก 140 ถึง 159 หรือ ค่าความดันไดแอสโตลิก 90 ถึง 99 แทนด้วย 3

กลุ่มที่ 4 ค่าความดันซีสโตลิกตั้งแต่ 160 ขึ้นไป หรือ ค่าความดันไดแอสโตลิกตั้งแต่ 100 ขึ้นไป แทนด้วย 4

น้ำตาลในเลือด ผู้วิจัยมีการจัดกลุ่มใหม่ 3 กลุ่มดังนี้

กลุ่มที่ 1 คือระดับน้ำตาลน้อยกว่า 110 แทนด้วย 1

กลุ่มที่ 2 คือระดับน้ำตาล 110 ถึง 125 แทนด้วย 2

กลุ่มที่ 3 คือระดับน้ำตาลตั้งแต่ 126 ขึ้นไป แทนด้วย 3

ไตรกลีเซอไรด์ ผู้วิจัยมีการจัดกลุ่มใหม่ 4 กลุ่มดังนี้

กลุ่มที่ 1 ค่าไตรกลีเซอไรด์ น้อยกว่า 150 แทนด้วย 1

กลุ่มที่ 2 ค่าไตรกลีเซอไรด์ 150 ถึง 199 แทนด้วย 2

กลุ่มที่ 3 ค่าไตรกลีเซอไรด์ 200 ถึง 499 แทนด้วย 3

กลุ่มที่ 4 ค่าไตรกลีเซอไรด์ ตั้งแต่ 500 ขึ้นไป แทนด้วย 4

โรคความดันโลหิตสูง ข้อมูลมีค่าตั้งแต่ I10 ถึง I15 หมายความว่าเป็นโรคความดันโลหิตสูง และ – หมายความว่าไม่เป็นโรคความดันโลหิตสูง ผู้วิจัยแทนค่าโรคความดันโลหิตสูง ตั้งแต่ I10 ถึง I15 ด้วย Y และ แทนค่าโรคความดันโลหิตสูง – ด้วย N

ผลจากการเตรียมข้อมูลสรุปภาพรวมเป็นตารางการจัดกลุ่มแทนค่าข้อมูลโรคความดันโลหิตสูง ดังแสดงในตาราง 24

ตาราง 24 การจัดกลุ่มแทนค่าข้อมูลโรคความดันโลหิตสูง

ชื่อฟิลด์	ความหมาย	ข้อมูลดิบ	การแทนค่า
Gender	เพศ	1	M
		2	F
Age	อายุ	60–69	1
		70–79	2
		>= 80	3
Bmi	ดัชนีมวลกาย	0–18.5	1
		18.5–22.9	2
		23–24.9	3
		25–29.9	4
		>= 30	5
Smoke	การสูบบุหรี่	1	N
		2	Y
Bp	ความดันโลหิต (ความดันซีสโตลิก และ ความดันไดแอสโตลิก)	SBP < 120 and DBP < 80	1
		SBP 120–139 or DBP 80–89	2
		SBP 140–159 or DBP 90–99	3
		SBP >= 160 or DBP >= 100	4
FPG	น้ำตาลในเลือด	< 110	1
		110 – 125	2
		>= 126	3
Tg	ไตรกลีเซอไรด์	< 150	1
		150 – 199	2
		200 – 499	3
		>= 500	4
Ht	ผลลัพธ์โรคความดันโลหิตสูง	I10 – I15	Y
		-	N

จากข้อมูลตาราง 24 แสดงตัวอย่างการแทนค่าข้อมูลได้ดังภาพประกอบ 25

	A	B	C	D	E	F	G	H
1	Gender	Age	Bmi	Smoke	Bp	Fpg	Tg	HT
2	F	2	5	N	2	3	2	Y
3	F	2	5	N	2	3	2	N
4	F	2	5	N	1	2	1	N
5	F	2	5	N	2	3	2	N
6	F	2	5	N	2	2	3	N
7	F	2	5	N	3	3	2	N
8	F	2	5	N	4	3	2	Y
9	F	2	5	N	2	3	2	Y
10	M	2	4	N	1	1	1	N
11	M	2	2	N	1	1	1	N
12	M	2	2	N	1	1	1	N
13	M	2	2	N	2	1	1	N
14	M	2	2	N	2	1	1	N
15	M	2	2	N	1	1	1	N

ภาพประกอบ 25 ตัวอย่างข้อมูลโรคความดันโลหิตสูงที่ผ่านการแปลงค่า

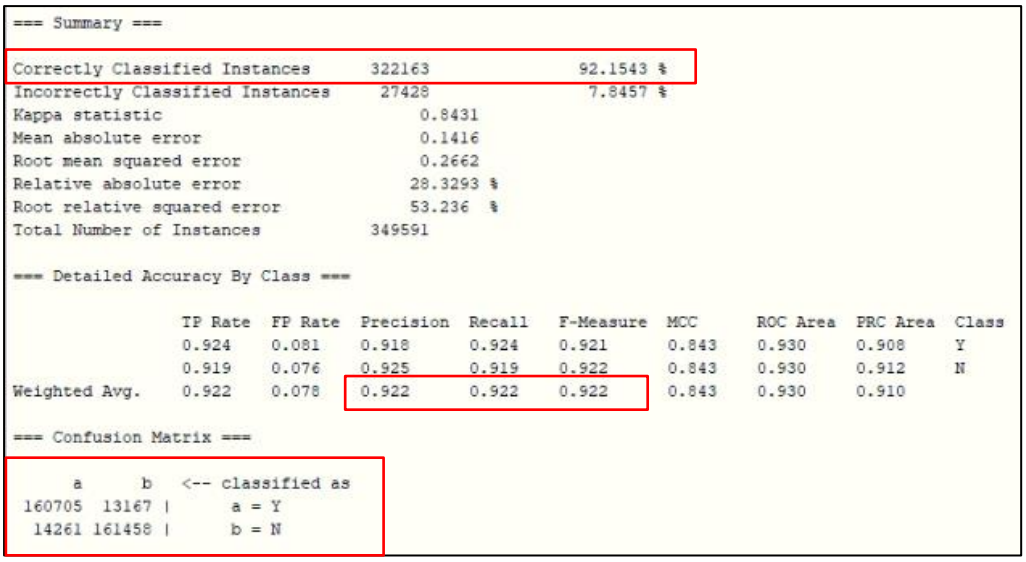
จากภาพประกอบ 25 ข้อมูลแถวที่ 2 คอลัมน์ A ฟิลด์ Gender มีค่าเท่ากับ F หมายถึงเพศหญิง คอลัมน์ B ฟิลด์ Age มีค่าเท่ากับ 2 หมายถึง ผู้สูงอายุที่มีอายุอยู่ระหว่าง 70 ถึง 79 คอลัมน์ C ฟิลด์ Bmi มีค่าเท่ากับ 5 หมายถึง ค่าดัชนีมวลกายค่ามากกว่าหรือเท่ากับ 30 คอลัมน์ D ฟิลด์ Smoke มีค่าเท่ากับ N หมายถึง ไม่สูบ คอลัมน์ E ฟิลด์ Bp มีค่าเท่ากับ 2 หมายถึง ค่าความดันโลหิตที่ค่า SBP อยู่ระหว่าง 120 ถึง 139 หรือ ค่า DBP อยู่ระหว่าง 80 ถึง 89 คอลัมน์ F ฟิลด์ FPG มีค่าเท่ากับ 3 หมายถึง ระดับน้ำตาลในเลือดมีค่ามากกว่าหรือเท่ากับ 126 คอลัมน์ G ฟิลด์ Tg มีค่าเท่ากับ 2 หมายถึง ระดับไขมันในเลือดอยู่ระหว่าง 150 ถึง 199 คอลัมน์ H ฟิลด์ HT มีค่าเท่ากับ Y หมายถึง ผลวินิจฉัยเป็นโรคความดันโลหิตสูง

เมื่อทำการแปลงข้อมูลแล้วให้ทำการบันทึกเป็นนามสกุล CSV จากนั้นผู้วิจัยนำข้อมูลมาวิเคราะห์ด้วยโปรแกรมเวก้า

2.2 การสร้างแบบจำลองการพยากรณ์โรคความดันโลหิตสูงสร้างแบบจำลองจาก 3 เทคนิค

1) ต้นไม้ตัดสินใจ

การสร้างแบบจำลองการพยากรณ์โรคความดันโลหิตสูงด้วยเทคนิคต้นไม้ตัดสินใจพบว่ามีค่าความถูกต้อง (Accuracy) 92.1543% ค่าความแม่นยำ (Precision) 0.922 ค่าความระลึก (Recall) 0.922 และค่าความถ่วงดุล (F-Measure) 0.922 ดังภาพประกอบ 26



```
=== Summary ===

Correctly Classified Instances      322163               92.1543 %
Incorrectly Classified Instances     27428                7.8457 %
Kappa statistic                          0.8431
Mean absolute error                      0.1416
Root mean squared error                  0.2662
Relative absolute error                 28.3293 %
Root relative squared error             53.236  %
Total Number of Instances           349591

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.924    0.081    0.918      0.924    0.921      0.843    0.930     0.908     Y
                 0.919    0.076    0.925      0.919    0.922      0.843    0.930     0.912     N
Weighted Avg.    0.922    0.078    0.922      0.922    0.922      0.843    0.930     0.910

=== Confusion Matrix ===

     a      b   <-- classified as
 160705  13167 |      a = Y
  14261 161458 |      b = N
```

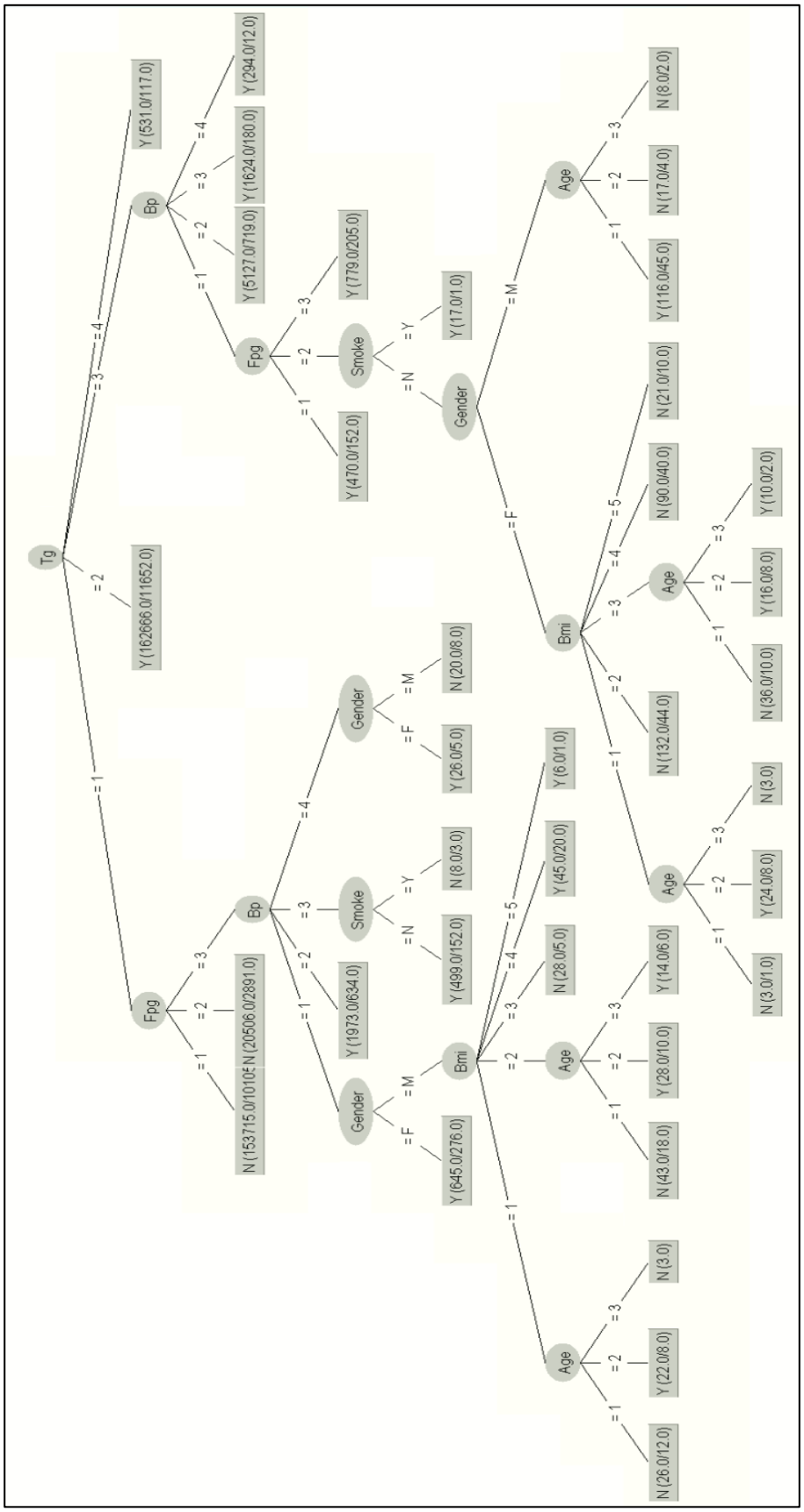
ภาพประกอบ 26 ผลลัพธ์ประสิทธิภาพแบบจำลองโรคความดันโลหิตสูงของเทคนิคต้นไม้ตัดสินใจ

และจากภาพประกอบ 26 สามารถเขียนเป็นคอนฟิวชันเมทริกซ์ (Confusion Matrix) ได้ดังนี้

ตาราง 25 คอนฟิวชันเมทริกซ์ของแบบจำลองโรคความดันโลหิตสูงของเทคนิคต้นไม้ตัดสินใจ

Predicted / Actual	Y	N
Y	160705	13167
N	14261	161458

จากตาราง 25 คอนฟิวชันเมทริกซ์ของแบบจำลองโรคความดันโลหิตสูงของเทคนิคต้นไม้ตัดสินใจ ได้ค่าความถูกต้องเชิงบวก (True Positive: TP) หรือ ทำนายว่าเป็นโรคความดันโลหิตสูงและระเบียนผู้ป่วยป่วยเป็นโรคความดันโลหิตสูงจริง คือ 160,705 ระเบียน ได้ค่าความถูกต้องเชิงลบ (True Negative: TN) หรือ ทำนายว่าไม่เป็นโรคความดันโลหิตสูง และผู้ป่วยไม่ได้ป่วยเป็นโรคความดันโลหิตสูงถูกต้องมี 161,458 ระเบียน ความผิดพลาดเชิงบวก (False Positive: FP) หรือ ผลการทำนายว่าเป็นโรคความดันโลหิตสูง แต่ผู้ป่วยไม่ได้ป่วยเป็นโรคความดันโลหิตสูงมี 13,167 ระเบียน และความผิดพลาดเชิงลบ (False Negative: FN) หรือผลการทำนายว่าไม่เป็นโรคความดันโลหิตสูง แต่ผู้ป่วยป่วยเป็นโรคความดันโลหิตสูงมี 14,261 ระเบียน



ภาพประกอบ 27 แบบจำลองโรคความดันโลหิตสูงของเทคนิคต้นไม้ตัดสินใจ

จากภาพประกอบ 27 เป็นภาพจำลองต้นไม้ตัดสินใจของโรคความดันโลหิตสูง โดยการพิจารณาว่าผู้ป่วยเป็นโรคความดันโลหิตสูง มีระดับการพิจารณา อยู่ 7 ระดับ คือ Tg Bp Fpg Smoke Gender Bmi และ Age ซึ่งอธิบายกฎความสัมพันธ์ของแบบจำลอง ดังตาราง 26

ตาราง 26 อธิบายกฎแบบจำลองโรคความดันโลหิตสูงของเทคนิคต้นไม้ตัดสินใจ

กฎความสัมพันธ์ของแบบจำลองโรคความดันโลหิตสูง (Rule)	ผลลัพธ์ (Result)
1. Tg = 1 and Fpg = 1	N
2. Tg = 1 and Fpg = 2	N
3. Tg = 1 and Fpg = 3 and Bp = 1 and Gender = F	N
4. Tg = 1 and Fpg = 3 and Bp = 1 and Gender = M and BMI = 1 and Age = 1	N
5. Tg = 1 and Fpg = 3 and Bp = 1 and Gender = M and BMI = 1 and Age = 2	Y
6. Tg = 1 and Fpg = 3 and Bp = 1 and Gender = M and BMI = 1 and Age = 3	N
7. Tg = 1 and Fpg = 3 and Bp = 1 and Gender = M and BMI = 2 and Age = 1	N
8. Tg = 1 and Fpg = 3 and Bp = 1 and Gender = M and BMI = 2 and Age = 2	Y
9. Tg = 1 and Fpg = 3 and Bp = 1 and Gender = M and BMI = 2 and Age = 3	Y
10. Tg = 1 and Fpg = 3 and Bp = 1 and Gender = M and BMI = 3	N
11. Tg = 1 and Fpg = 3 and Bp = 1 and Gender = M and BMI = 4	Y
12. Tg = 1 and Fpg = 3 and Bp = 1 and Gender = M and BMI = 5	Y
13. Tg = 1 and Fpg = 3 and Bp = 2	Y
14. Tg = 1 and Fpg = 3 and Bp = 3 and Smoke = N	Y
15. Tg = 1 and Fpg = 3 and Bp = 3 and Smoke = Y	Y

ตาราง 26 (ต่อ)

กฎความสัมพันธ์ของแบบจำลองโรคความดันโลหิตสูง (Rule)	ผลลัพธ์ (Result)
16. Tg = 1 and Fpg = 3 and Bp = 4 and Gender = F	Y
17. Tg = 1 and Fpg = 3 and Bp = 4 and Gender = M	N
18. Tg = 2	Y
19. Tg = 3 and Bp = 1 and Fpg = 1	Y
20. Tg = 3 and Bp = 1 and Fpg = 2 and Smoke = N and Gender = F and Bmi = 1 and Age = 1	N
21. Tg = 3 and Bp = 1 and Fpg = 2 and Smoke = N and Gender = F and Bmi = 1 and Age = 2	Y
22. Tg = 3 and Bp = 1 and Fpg = 2 and Smoke = N and Gender = F and Bmi = 1 and Age = 3	N
23. Tg = 3 and Bp = 1 and Fpg = 2 and Smoke = N and Gender = F and Bmi = 2	N
24. Tg = 3 and Bp = 1 and Fpg = 2 and Smoke = N and Gender = F and Bmi = 3 and Age = 1	N
25. Tg = 3 and Bp = 1 and Fpg = 2 and Smoke = N and Gender = F and Bmi = 3 and Age = 2	Y
26. Tg = 3 and Bp = 1 and Fpg = 2 and Smoke = N and Gender = F and Bmi = 3 and Age = 3	Y
27. Tg = 3 and Bp = 1 and Fpg = 2 and Smoke = N and Gender = F and Bmi = 4	N
28. Tg = 3 and Bp = 1 and Fpg = 2 and Smoke = N and Gender = F and Bmi = 5	N
29. Tg = 3 and Bp = 1 and Fpg = 2 and Smoke = N and Gender = M and Age = 1	Y
30. Tg = 3 and Bp = 1 and Fpg = 2 and Smoke = N and Gender = M and Age = 2	N

ตาราง 26 (ต่อ)

กฎความสัมพันธ์ของแบบจำลองโรคความดันโลหิตสูง (Rule)	ผลลัพธ์ (Result)
31. Tg = 3 and Bp = 1 and Fpg = 2 and Smoke = N and Gender = M and Age = 3	N
32. Tg = 3 and Bp = 1 and Fpg = 2 and Smoke = Y	Y
33. Tg = 3 and Bp = 1 and Fpg = 3	Y
34. Tg = 3 and Bp = 2	Y
35. Tg = 3 and Bp = 3	Y
36. Tg = 3 and Bp = 4	Y
37. Tg = 4	Y

ตัวอย่าง กฎข้อที่ 1 ถ้าค่าไขมันในเลือดน้อยกว่า 200 และค่าน้ำตาลในเลือดน้อยกว่า 126 แล้วสถานะไม่เป็นโรคความดันโลหิตสูง

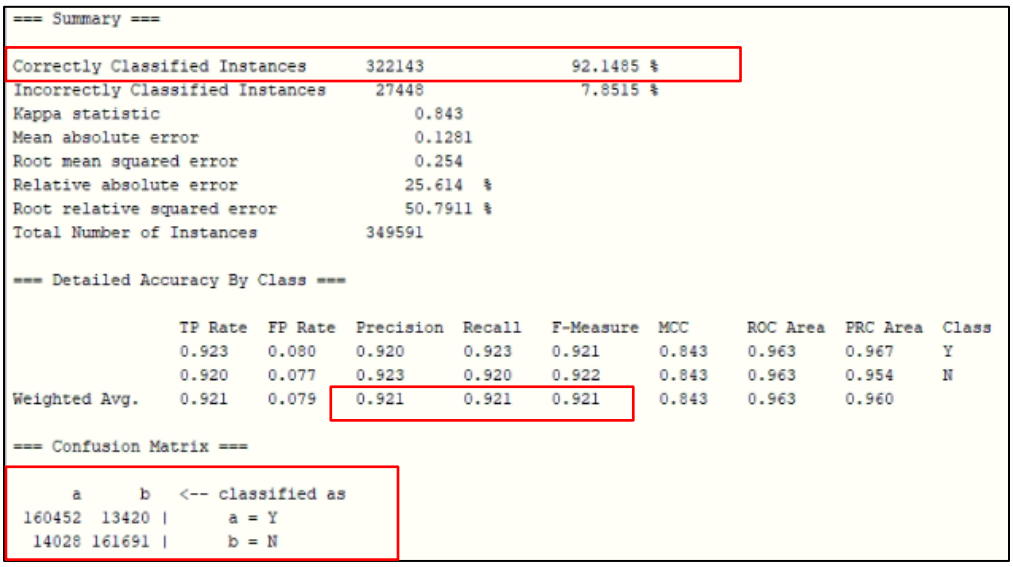
กฎข้อที่ 2 ถ้าค่าไขมันในเลือดน้อยกว่า 200 ค่าน้ำตาลในเลือดมากกว่าหรือเท่ากับ 126 ค่าความดันโลหิตค่าความดันซีสโตลิกน้อยกว่า 120 ค่าความดันไดแอสโตลิกน้อยกว่า 80 และเป็นเพศหญิง แล้วสถานะไม่เป็นโรคความดันโลหิตสูง

กฎข้อที่ 3 ถ้าค่าไขมันในเลือดน้อยกว่า 200 ค่าน้ำตาลในเลือดมากกว่าหรือเท่ากับ 126 ค่าความดันโลหิตค่าความดันซีสโตลิกน้อยกว่า 120 ค่าความดันไดแอสโตลิกน้อยกว่า 80 เป็นเพศชาย ดัชนีมวลกายน้อยกว่าหรือเท่ากับ 18.5 อายุอยู่ระหว่าง 60 ถึง 69 ปี แล้วสถานะไม่เป็นโรคความดันโลหิตสูง

กฎข้อที่ 4 ถ้าค่าไขมันในเลือดน้อยกว่า 200 ค่าน้ำตาลในเลือดมากกว่าหรือเท่ากับ 126 ค่าความดันโลหิตค่าความดันซีสโตลิกน้อยกว่า 120 ค่าความดันไดแอสโตลิกน้อยกว่า 80 เป็นเพศชาย ดัชนีมวลกายน้อยกว่าหรือเท่ากับ 18.5 อายุอยู่ระหว่าง 70 ถึง 79 ปี แล้วสถานะเป็นโรคความดันโลหิตสูง เป็นต้น

2) เพื่อนบ้านใกล้ที่สุด

การสร้างแบบจำลองการพยากรณ์โรคความดันโลหิตสูงด้วยเทคนิคเพื่อนบ้านใกล้ที่สุดพบว่ามีค่าความถูกต้อง (Accuracy) 92.1485% ค่าความแม่นยำ (Precision) 0.921 ค่าความระลึก (Recall) 0.921 และค่าความถ่วงดุล (F-Measure) 0.921 ดังภาพประกอบ 28



```
=== Summary ===

Correctly Classified Instances      322143               92.1485 %
Incorrectly Classified Instances     27448                7.8515 %
Kappa statistic                          0.843
Mean absolute error                      0.1281
Root mean squared error                  0.254
Relative absolute error                 25.614  %
Root relative squared error             50.7911 %
Total Number of Instances           349591

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.923    0.080    0.920      0.923    0.921      0.843    0.963     0.967     Y
                 0.920    0.077    0.923      0.920    0.922      0.843    0.963     0.954     N
Weighted Avg.    0.921    0.079    0.921      0.921    0.921      0.843    0.963     0.960

=== Confusion Matrix ===

     a      b   <-- classified as
 160452  13420 |      a = Y
  14028 161691 |      b = N
```

ภาพประกอบ 28 ผลลัพธ์ประสิทธิภาพแบบจำลองโรคความดันโลหิตสูงของเทคนิคเพื่อนบ้านใกล้ที่สุด

และจากภาพประกอบ 28 สามารถเขียนเป็นคอนฟิวชันเมทริกซ์ (Confusion Matrix) ได้ดังนี้

ตาราง 27 คอนฟิวชันเมทริกซ์ของแบบจำลองโรคความดันโลหิตสูงของเทคนิคเพื่อนบ้านใกล้ที่สุด

Predicted / Actual	Y	N
Y	160452	13420
N	14028	161691

จากตาราง 27 คอนฟิวชันเมทริกซ์ของแบบจำลองโรคความดันโลหิตสูงของเทคนิคเพื่อนบ้านใกล้ที่สุด ได้ค่าความถูกต้องเชิงบวก (True Positive: TP) หรือ ทำนายว่าเป็นโรคความดันโลหิตสูงและระเบียนผู้ป่วยป่วยเป็นโรคความดันโลหิตสูงจริง คือ 160,452 ระเบียน ได้ค่าความถูกต้องเชิงลบ (True Negative: TN) หรือ ทำนายว่าไม่เป็นโรคความดันโลหิตสูงและผู้ป่วยไม่ได้ป่วยเป็นโรคความดันโลหิตสูงถูกต้องมี 161,691 ระเบียน ความผิดพลาดเชิงบวก (False Positive: FP) หรือ ผลการทำนายว่าเป็นโรคความดันโลหิตสูง แต่ผู้ป่วยไม่ได้ป่วยเป็นโรคความดันโลหิตสูงมี 13,420 ระเบียน และความผิดพลาดเชิงลบ (False Negative: FN) หรือผลการทำนายว่าไม่เป็นโรคความดันโลหิตสูง แต่ผู้ป่วยป่วยเป็นโรคความดันโลหิตสูงมี 14,028 ระเบียน

ตาราง 28 ตัวอย่างแบบจำลองโรคความดันโลหิตสูงของเทคนิคเพื่อนบ้านใกล้ที่สุด

Data	Gender	Age	Bmi	Smoke	Bp	FPG	Tg	Distance	HT	Nearest Sign
D1	0	2	5	0	4	3	2	3.46	+	
Distance $D1 = \sqrt{(1-0)^2 + (2-2)^2 + (2-5)^2 + (0-0)^2 + (3-4)^2 + (2-3)^2 + (2-2)^2}$ = 3.46										
D2	1	2	4	0	1	1	1	3.16	–	
Distance $D2 = \sqrt{(1-1)^2 + (2-2)^2 + (2-4)^2 + (0-0)^2 + (3-1)^2 + (2-1)^2 + (2-1)^2}$ = 3.16										
D3	0	1	1	0	1	1	3	3	+	
Distance $D3 = \sqrt{(1-0)^2 + (2-1)^2 + (2-1)^2 + (0-0)^2 + (3-1)^2 + (2-1)^2 + (2-3)^2}$ = 3										
D4	1	1	2	1	2	1	1	2.23	–	–
Distance $D4 = \sqrt{(1-1)^2 + (2-1)^2 + (2-2)^2 + (0-1)^2 + (3-2)^2 + (2-1)^2 + (2-1)^2}$ = 2.23										
D5	1	3	2	0	1	1	1	2.64	–	–
Distance $D5 = \sqrt{(1-1)^2 + (2-3)^2 + (2-2)^2 + (0-0)^2 + (3-1)^2 + (2-1)^2 + (2-1)^2}$ = 2.64										
D6	1	2	3	0	3	3	2	1.41	+	+
Distance $D6 = \sqrt{(1-1)^2 + (2-2)^2 + (2-3)^2 + (0-0)^2 + (3-3)^2 + (2-3)^2 + (2-2)^2}$ = 1.41										
K=3	1	2	2	0	3	2	2		?	

จากตาราง 28 แบบจำลองโรคความดันโลหิตสูงของเทคนิคเพื่อนบ้านใกล้ที่สุดของตัวอย่างข้อมูลบางส่วนกำหนดให้ คลาส HT = Y แทนด้วย + และคลาส HT = N

แทนด้วย –, Gender = F แทนด้วย 0 และ Gender = M แทนด้วย 1, Smoke = N แทนด้วย 0 และ Smoke = Y แทนด้วย 1 จะใช้วิธีการวัดระยะทางแบบยุคลิด (Euclidean Distance) โดยกำหนด K = 3 พบว่า คลาส HT = – ปรากฏมากที่สุด ดังนั้นจึงสรุปได้ว่าข้อมูล Gender = M, Age = 2, Bmi = 2, Smoke = N, Bp = 3, FPG = 2 และ Tg = 2 จะทำนายว่ามีคลาส HT = N

3) นาอีฟเบย์

การสร้างแบบจำลองการพยากรณ์โรคความดันโลหิตสูงด้วยเทคนิคนาอีฟเบย์พบว่ามีค่าความถูกต้อง (Accuracy) 91.8525% ค่าความแม่นยำ (Precision) 0.919 ค่าความระลึก (Recall) 0.919 และค่าความถ่วงดุล (F–Measure) 0.919 ดังภาพประกอบ 29

```
=== Summary ===

Correctly Classified Instances      321108               91.8525 %
Incorrectly Classified Instances     28483                8.1475 %
Kappa statistic                          0.837
Mean absolute error                      0.1387
Root mean squared error                  0.2756
Relative absolute error                 27.7447 %
Root relative squared error             55.1201 %
Total Number of Instances           349591

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.913    0.076    0.923      0.913    0.918      0.837    0.921     0.863     Y
                 0.924    0.087    0.914      0.924    0.919      0.837    0.921     0.931     N
Weighted Avg.    0.919    0.082    0.919      0.919    0.919      0.837    0.921     0.897

=== Confusion Matrix ===

      a      b   <-- classified as
 158684  15188 |      a = Y
  13295 162424 |      b = N
```

ภาพประกอบ 29 ผลลัพธ์ประสิทธิภาพแบบจำลองโรคความดันโลหิตสูงของเทคนิคนาอีฟเบย์

และจากภาพประกอบ 29 สามารถเขียนเป็นคอนฟิวชันเมทริกซ์ (Confusion Matrix) ได้ดังนี้

ตาราง 29 คอนฟิวชันเมทริกซ์ของแบบจำลองโรคความดันโลหิตสูงของเทคนิคนาอีฟเบย์

Predicted / Actual	Y	N
Y	158684	15188
N	13295	162424

จากตาราง 29 คอนฟิวชันเมทริกซ์ของแบบจำลองโรคความดันโลหิตสูงของเทคนิคนาอีฟเบย์ ได้ค่าความถูกต้องเชิงบวก (True Positive: TP) หรือ ทำนายว่าเป็นโรคความดันโลหิตสูงและระเบียนผู้ป่วยป่วยเป็นโรคความดันโลหิตสูงจริง คือ 158,684 ระเบียน ได้ค่าความถูกต้องเชิงลบ (True Negative: TN) หรือ ทำนายว่าไม่เป็นโรคความดันโลหิตสูงและผู้ป่วยไม่ได้ป่วยเป็นโรคความดันโลหิตสูงถูกต้องมี 162,424 ระเบียน ความผิดพลาดเชิงบวก (False Positive: FP) หรือ ผลการทำนายว่าเป็นโรคความดันโลหิตสูง แต่ผู้ป่วยไม่ได้ป่วยเป็นโรคความดันโลหิตสูงมี 15,188 ระเบียน และความผิดพลาดเชิงลบ (False Negative: FN) หรือผลการทำนายว่าไม่เป็นโรคความดันโลหิตสูง แต่ผู้ป่วยป่วยเป็นโรคความดันโลหิตสูงมี 13,295 ระเบียน

ตาราง 30 แบบจำลองโรคความดันโลหิตสูงของเทคนิคนาอีฟเบย์

แอตทริบิวต์	คลาส HT = Y	คลาส HT = N	ความน่าจะเป็น คลาส HT=Y	ความน่าจะเป็น คลาส HT=N
Gender = F	109078	100914	0.63	0.57
Gender = M	64796	74807	0.37	0.43
Age = 1	91885	97393	0.53	0.55
Age = 2	63843	59019	0.37	0.34
Age = 3	18147	19310	0.10	0.11
Bmi = 1	22903	37744	0.13	0.21
Bmi = 2	64326	70774	0.37	0.40
Bmi = 3	32888	27568	0.19	0.16
Bmi = 4	42515	32720	0.24	0.19
Bmi = 5	11245	6918	0.07	0.04
Smoke = N	171715	172067	0.99	0.98
Smoke = Y	2159	3654	0.01	0.02
Bp = 1	35989	83909	0.21	0.48
Bp = 2	109430	72372	0.63	0.41
Bp = 3	23936	16623	0.14	0.09
Bp = 4	4521	2819	0.02	0.02
Fpg = 1	108126	144209	0.62	0.82
Fpg = 2	20158	18502	0.12	0.11
Fpg = 3	45591	13011	0.26	0.07
Tg = 1	15189	162420	0.09	0.92
Tg = 2	151015	11653	0.87	0.07
Tg = 3	7257	1532	0.04	0.01
Tg = 4	415	118	0.00	0.00

จากตาราง 30 แบบจำลองโรคความดันโลหิตสูงของเทคนิคนาอีฟเบย์ สามารถนำค่าไปหาความน่าจะเป็นโดยใช้ทฤษฎีของเบย์ (Bayes Theorem) ตามสมการ

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

เช่น Gender = F, Age = 1, Bmi = 1, Smoke = N, Bp = 1, Fpg = 1, Tg = 1 และ HT = ? สามารถคำนวณได้ดังนี้

P(HT = Y|A) = P(Gender = F | HT = Y) x P(Age = 1 | HT = Y)
x P(Bmi = 1 | HT = Y) x P(Smoke = N | HT = Y)
x P(Bp = 1 | HT = Y) x P(Fpg = 1 | HT = Y)
x P(Tg = 1 | HT = Y)
= 0.63 x 0.53 x 0.13 x 0.99 x 0.21 x 0.62 x 0.09
= 0.00050

P(HT = N|A) = P(Gender = F | HT = N) x P(Age = 1 | HT = N)
x P(Bmi = 1 | HT = N) x P(Smoke = N | HT = N)
x P(Bp = 1 | HT = N) x P(Fpg = 1 | HT = N)
x P(Tg = 1 | HT = N)
= 0.57 x 0.55 x 0.21 x 0.98 x 0.48 x 0.82 x 0.92
= 0.02336

ดังนั้นข้อมูล Gender = F, Age = 1, Bmi = 1, Smoke = N, Bp = 1, Fpg = 1 และ Tg = 1 จะทำนายว่ามีคลาส HT = N เนื่องจากค่า P(HT = N|A) = 0.02336 ซึ่งมีค่ามากกว่า P(HT = Y|A) = 0.00050

3. การสร้างแบบจำลองพยากรณ์การเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุโรคหัวใจขาดเลือด

3.1 การเตรียมข้อมูลโรคหัวใจขาดเลือด

1) ข้อมูลที่ใช้ ผู้วิจัยนำข้อมูลผู้สูงอายุที่รับบริการรักษาจากโรงพยาบาลนาหว้า อ.นาหว้า จ.นครพนม ตั้งแต่เดือนตุลาคม พ.ศ. 2554 ถึงเดือนกันยายน พ.ศ. 2562 ประกอบไปด้วยข้อมูล เพศ อายุ การสูบบุหรี่ ดัชนีมวลกาย ความดันโลหิต น้ำตาลในเลือด และไขมันในเลือด และข้อมูลผลการเป็นโรคไม่ติดต่อเรื้อรัง คือ โรคหัวใจขาดเลือด ดังภาพประกอบ 30

	A	B	C	D	E	F	G	H	I
1	**Gender**	**Age**	**Bmi**	**Smoke**	**SBp**	**DBp**	**Fpg**	**Tg**	**Heart**
2	2	72	5	1	130	80	130	198	I259
3	2	72	5	1	121	81	127	170	-
4	2	72	5	1	119	75	112	144	-
5	2	71	5	1	121	81	131	199	-
6	2	71	5	1	132	81	119	201	I259
7	2	71	5	1	150	92	140	181	-
8	2	71	5	1	166	89	127	157	-
9	2	71	5	1	139	80	129	163	-
10	1	79	4	1	110	65	91	138	-
11	1	79	2	1	115	78	92	146	-
12	1	79	2	1	118	67	99	120	-
13	1	79	2	1	121	81	87	111	-
14	1	79	2	1	130	88	86	147	-
15	1	79	2	1	110	67	108	137	-

ภาพประกอบ 30 ตัวอย่างข้อมูลผู้รับการรักษาและผลการเป็นโรคหัวใจขาดเลือด

2) การแปลงข้อมูลผู้รับการรักษาโรคความดันโลหิตสูงด้วยโปรแกรมไมโครซอฟต์เอกเซล (Microsoft Excel) ดังนี้

เพศ ข้อมูลมีค่าคือ 1 หมายถึงเพศชาย และ 2 หมายถึงเพศหญิง ผู้วิจัยแทนค่าเพศ 1 ด้วย M และ เพศ 2 ด้วย F

อายุ ผู้วิจัยมีการจัดกลุ่มใหม่ 3 กลุ่มดังนี้

กลุ่มที่ 1 ผู้สูงอายุวัยต้น (อายุ 60–69 ปี) แทนด้วย 1

กลุ่มที่ 2 ผู้สูงอายุวัยกลาง (อายุ 70–79 ปี) แทนด้วย 2

กลุ่มที่ 3 ผู้สูงอายุวัยปลาย (อายุ 80 ปีขึ้นไป) แทนด้วย 3

การสูบบุหรี่ ข้อมูลมีค่าคือ 1 หมายถึงไม่สูบบุหรี่ และ 2 หมายถึงสูบบุหรี่ ผู้วิจัยแทนค่าการสูบบุหรี่ 1 ด้วย N และ การสูบบุหรี่ 2 ด้วย Y

ดัชนีมวลกาย ผู้วิจัยมีการจัดกลุ่มใหม่ 5 กลุ่มดังนี้

กลุ่มที่ 1 คือค่า BMI น้อยกว่า 18.5 แทนด้วย 1

กลุ่มที่ 2 คือค่า BMI 18.5 ถึง 22.9 แทนด้วย 2

กลุ่มที่ 3 คือค่า BMI 23 ถึง 24.9 แทนด้วย 3

กลุ่มที่ 4 คือค่า BMI 25 ถึง 29.9 แทนด้วย 4

กลุ่มที่ 5 คือค่า BMI ตั้งแต่ 30 ขึ้นไป แทนด้วย 5

ความดันโลหิต ผู้วิจัยมีการจัดกลุ่มใหม่ 4 กลุ่มดังนี้

กลุ่มที่ 1 ค่าความดันซีสโตลิก (Systolic) น้อยกว่า 120 และค่าความดันไดแอสโตลิก (Diastolic) น้อยกว่า 80 แทนด้วย 1

กลุ่มที่ 2 ค่าความดันซีสโตลิก 120 ถึง 139 หรือ ค่าความดันไดแอสโตลิก 80 ถึง 89 แทนด้วย 2

กลุ่มที่ 3 ค่าความดันซีสโตลิก 140 ถึง 159 หรือ ค่าความดันไดแอสโตลิก 90 ถึง 99 แทนด้วย 3

กลุ่มที่ 4 ค่าความดันซีสโตลิกตั้งแต่ 160 ขึ้นไป หรือ ค่าความดันไดแอสโตลิกตั้งแต่ 100 ขึ้นไป แทนด้วย 4

น้ำตาลในเลือด ผู้วิจัยมีการจัดกลุ่มใหม่ 3 กลุ่มดังนี้

กลุ่มที่ 1 คือระดับน้ำตาลน้อยกว่า 110 แทนด้วย 1

กลุ่มที่ 2 คือระดับน้ำตาล 110 ถึง 125 แทนด้วย 2

กลุ่มที่ 3 คือระดับน้ำตาลตั้งแต่ 126 ขึ้นไป แทนด้วย 3

ไตรกลีเซอไรด์ ผู้วิจัยมีการจัดกลุ่มใหม่ 4 กลุ่มดังนี้

กลุ่มที่ 1 ค่าไตรกลีเซอไรด์ น้อยกว่า 150 แทนด้วย 1

กลุ่มที่ 2 ค่าไตรกลีเซอไรด์ 150 ถึง 199 แทนด้วย 2

กลุ่มที่ 3 ค่าไตรกลีเซอไรด์ 200 ถึง 499 แทนด้วย 3

กลุ่มที่ 4 ค่าไตรกลีเซอไรด์ ตั้งแต่ 500 ขึ้นไป แทนด้วย 4

โรคหัวใจขาดเลือด ข้อมูลมีค่าตั้งแต่ I20 ถึง I25 หมายความว่าเป็นโรคหัวใจขาดเลือด และ – หมายความว่าไม่เป็นโรคหัวใจขาดเลือด ผู้วิจัยแทนค่าโรคหัวใจขาดเลือด ตั้งแต่ I20 ถึง I25 ด้วย Y และ แทนค่าโรคหัวใจขาดเลือด – ด้วย N

ผลจากการเตรียมข้อมูลสรุปภาพรวมเป็นตารางการจัดกลุ่มแทนค่าข้อมูลโรคหัวใจขาดเลือด ดังแสดงในตาราง 31

ตาราง 31 การจัดกลุ่มแทนค่าข้อมูลโรคหัวใจขาดเลือด

ชื่อฟิลด์	ความหมาย	ข้อมูลดิบ	การแทนค่า
Gender	เพศ	1	M
		2	F
Age	อายุ	60–69	1
		70–79	2
		>= 80	3
Bmi	ดัชนีมวลกาย	0–18.5	1
		18.5–22.9	2
		23–24.9	3
		25–29.9	4
		>= 30	5
Smoke	การสูบบุหรี่	1	N
		2	Y
Bp	ความดันโลหิต (ความดันซีสโตลิก และ ความดันไดแอสโตลิก)	SBP < 120 and DBP < 80	1
		SBP 120–139 or DBP 80–89	2
		SBP 140–159 or DBP 90–99	3
		SBP >= 160 or DBP >= 100	4
FPG	น้ำตาลในเลือด	< 110	1
		110 – 125	2
		>= 126	3
Tg	ไตรกลีเซอไรด์	< 150	1
		150 – 199	2
		200 – 499	3
		>= 500	4
Heart	ผลลัพธ์โรคหัวใจขาดเลือด	I20 – I25	Y
		-	N

จากข้อมูลตาราง 31 แสดงตัวอย่างการแทนค่าข้อมูลได้ดังภาพประกอบ 31

	A	B	C	D	E	F	G	H
1	Gender	Age	Bmi	Smoke	Bp	Fpg	Tg	Heart
258407	F	2	2	N	1	3	2	N
258408	F	2	2	N	1	2	1	N
258409	M	1	2	N	2	1	1	Y
258410	M	1	2	N	2	1	1	N
258411	M	1	2	N	1	1	1	N
258412	M	1	2	N	2	1	1	Y
258413	M	1	2	N	2	1	1	Y
258414	M	1	2	N	2	1	1	Y
258415	M	1	1	Y	2	1	1	N
258416	M	1	2	N	1	1	1	N
258417	M	1	2	N	1	1	1	N
258418	M	1	2	N	1	1	1	N

ภาพประกอบ 31 ตัวอย่างข้อมูลโรคหัวใจขาดเลือดที่ผ่านการแปลงค่า

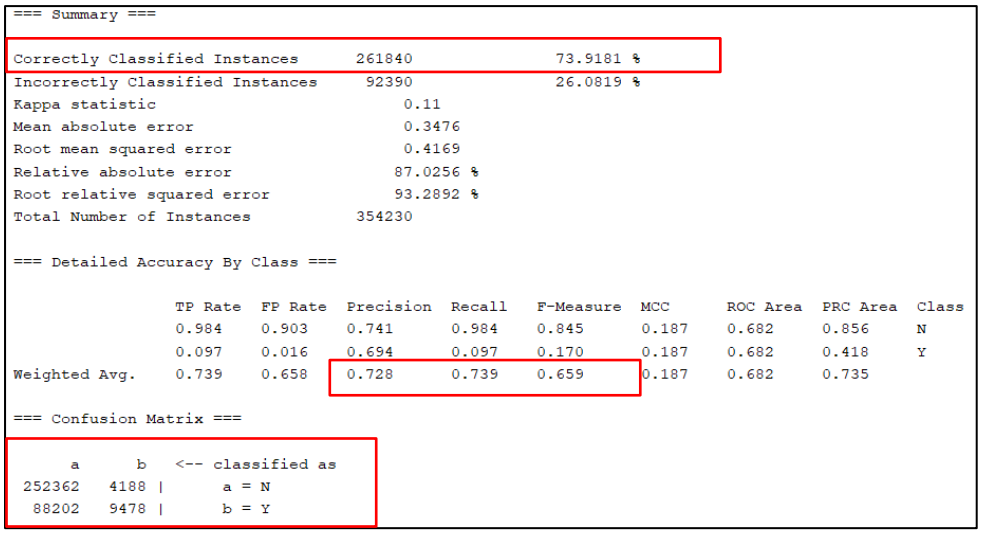
จากภาพประกอบ 31 ข้อมูลแถวที่ 258407 คอลัมน์ A ฟิลด์ Gender มีค่าเท่ากับ F หมายถึงเพศหญิง คอลัมน์ B ฟิลด์ Age มีค่าเท่ากับ 2 หมายถึง ผู้สูงอายุที่มีอายุอยู่ระหว่าง 70 ถึง 79 คอลัมน์ C ฟิลด์ Bmi มีค่าเท่ากับ 2 หมายถึง ค่าดัชนีมวลกายค่าระหว่าง 18.5 ถึง 22.9 คอลัมน์ D ฟิลด์ Smoke มีค่าเท่ากับ N หมายถึง ไม่สูบ คอลัมน์ E ฟิลด์ Bp มีค่าเท่ากับ 1 หมายถึง ค่าความดันโลหิตที่ค่า SBP น้อยกว่า 120 และค่า DBP น้อยกว่า 80 คอลัมน์ F ฟิลด์ FPG มีค่าเท่ากับ 3 หมายถึง ระดับน้ำตาลในเลือดมีค่ามากกว่าหรือเท่ากับ 126 คอลัมน์ G ฟิลด์ Tg มีค่าเท่ากับ 2 หมายถึง ระดับไขมันในเลือดอยู่ระหว่าง 150 ถึง 199 คอลัมน์ H ฟิลด์ Heart มีค่าเท่ากับ Y หมายถึง ผลวินิจฉัยเป็นโรคหัวใจขาดเลือด

เมื่อทำการแปลงข้อมูลแล้วให้ทำการบันทึกเป็นนามสกุล CSV จากนั้นผู้วิจัยนำข้อมูลมาวิเคราะห์ด้วยโปรแกรมเวก้า

3.2 การสร้างแบบจำลองการพยากรณ์โรคหัวใจขาดเลือดสร้างแบบจำลองจาก 3 เทคนิค

1) ต้นไม้ตัดสินใจ

การสร้างแบบจำลองการพยากรณ์โรคหัวใจขาดเลือดด้วยเทคนิคต้นไม้ตัดสินใจพบว่ามีค่าความถูกต้อง (Accuracy) 73.9181% ค่าความแม่นยำ (Precision) 0.728 ค่าความระลึก (Recall) 0.739 และค่าความถ่วงดุล (F–Measure) 0.659 ดังภาพประกอบ 32



```
=== Summary ===

Correctly Classified Instances      261840               73.9181 %
Incorrectly Classified Instances     92390               26.0819 %
Kappa statistic                          0.11
Mean absolute error                      0.3476
Root mean squared error                  0.4169
Relative absolute error                 87.0256 %
Root relative squared error             93.2892 %
Total Number of Instances           354230

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.984    0.903    0.741      0.984    0.845      0.187    0.682     0.856     N
                 0.097    0.016    0.694      0.097    0.170      0.187    0.682     0.418     Y
Weighted Avg.    0.739    0.658    0.728      0.739    0.659      0.187    0.682     0.735

=== Confusion Matrix ===

      a      b   <-- classified as
 252362   4188 |      a = N
  88202   9478 |      b = Y
```

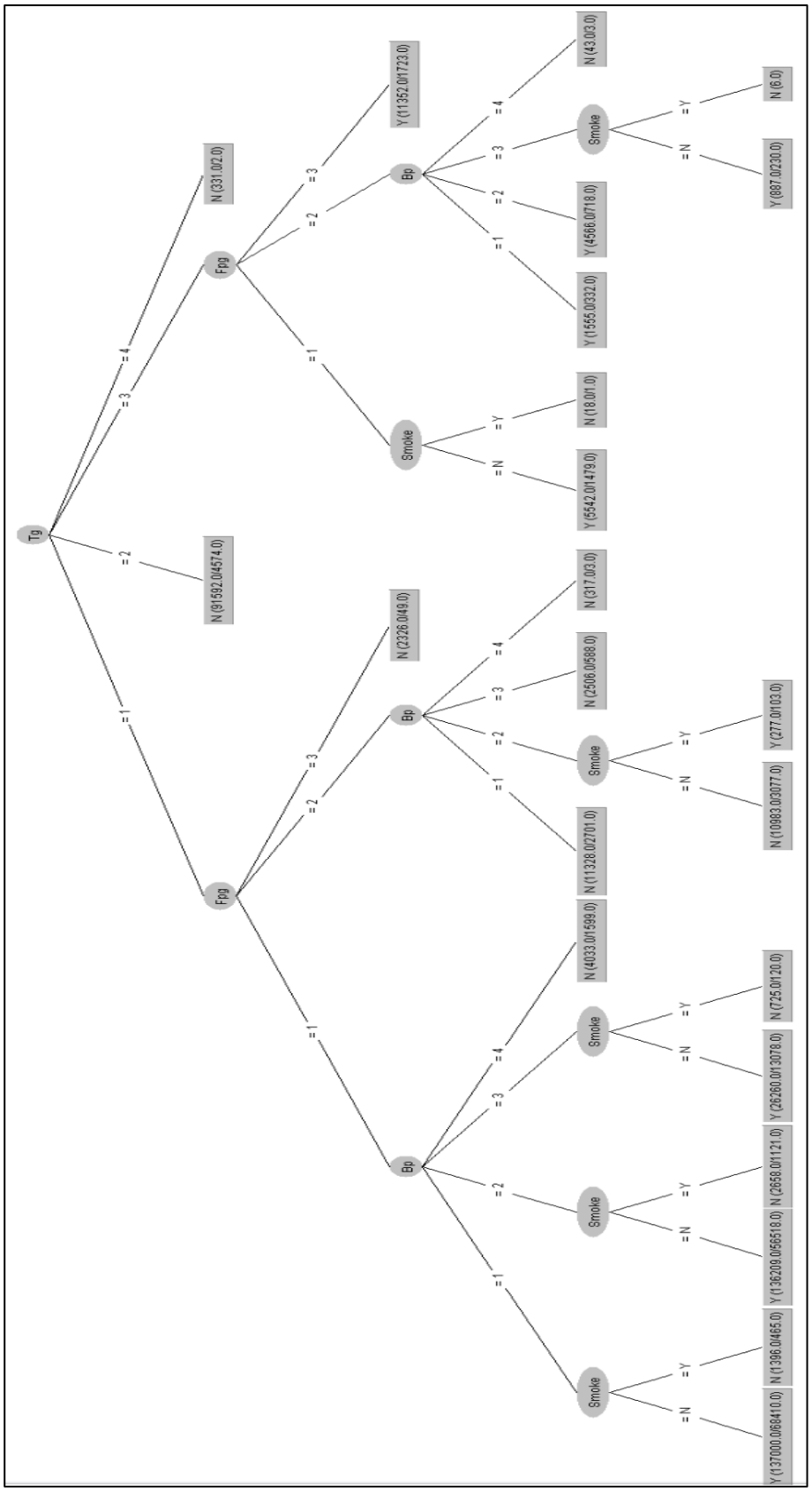
ภาพประกอบ 32 ผลลัพธ์ประสิทธิภาพแบบจำลองโรคหัวใจขาดเลือดของเทคนิคต้นไม้ตัดสินใจ

และจากภาพประกอบ 32 สามารถเขียนเป็นคอนฟิวชันเมทริกซ์ (Confusion Matrix) ได้ดังนี้

ตาราง 32 คอนฟิวชันเมทริกซ์ของแบบจำลองโรคหัวใจขาดเลือดของเทคนิคต้นไม้ตัดสินใจ

Predicted / Actual	Y	N
Y	9478	88202
N	4188	252362

จากตาราง 32 คอนฟิวชันเมทริกซ์ของแบบจำลองโรคหัวใจขาดเลือดของเทคนิคต้นไม้ตัดสินใจ ได้ค่าความถูกต้องเชิงบวก (True Positive: TP) หรือ ทำนายว่าเป็นโรคหัวใจขาดเลือดและระเบียนผู้ป่วยป่วยเป็นโรคหัวใจขาดเลือดจริง คือ 9,478 ระเบียน ได้ค่าความถูกต้องเชิงลบ (True Negative: TN) หรือ ทำนายว่าไม่เป็นโรคหัวใจขาดเลือดและผู้ป่วยไม่ได้ป่วยเป็นโรคหัวใจขาดเลือดถูกต้องมี 252,362 ระเบียน ความผิดพลาดเชิงบวก (False Positive: FP) หรือ ผลการทำนายว่าเป็นโรคหัวใจขาดเลือดแต่ผู้ป่วยไม่ได้ป่วยเป็นโรคหัวใจขาดเลือดมี 88,202 ระเบียน และความผิดพลาดเชิงลบ (False Negative: FN) หรือผลการทำนายว่าไม่เป็นโรคหัวใจขาดเลือด แต่ผู้ป่วยป่วยเป็นโรคหัวใจขาดเลือดมี 4,188 ระเบียน



ภาพประกอบ 33 แบบจำลองโครห์ริโซขาดเลือดของเทคนิคต้นไม้ตัดสินใจ

จากภาพประกอบ 33 เป็นภาพจำลองต้นไม้ตัดสินใจของโรคหัวใจขาดเลือดโดยการพิจารณาว่าผู้ป่วยเป็นโรคหัวใจขาดเลือด มีระดับการพิจารณา อยู่ 4 ระดับ คือ Tg Fpg Bp และ Smoke ซึ่งอธิบายกฎความสัมพันธ์ของแบบจำลอง ดังตาราง 33

ตาราง 33 อธิบายกฎแบบจำลองโรคหัวใจขาดเลือดของเทคนิคต้นไม้ตัดสินใจ

กฎของแบบจำลองโรคหัวใจขาดเลือดของเทคนิคต้นไม้ตัดสินใจ (Rule)	ผลลัพธ์ (Result)
1. Tg = 1 and Fpg = 1 and Bp = 1 and Smoke = N	N
2. Tg = 1 and Fpg = 1 and Bp = 1 and Smoke = Y	Y
3. Tg = 1 and Fpg = 1 and Bp = 2 and Smoke = N	Y
4. Tg = 1 and Fpg = 1 and Bp = 2 and Smoke = Y	N
5. Tg = 1 and Fpg = 1 and Bp = 3 and Smoke = N	Y
6. Tg = 1 and Fpg = 1 and Bp = 3 and Smoke = Y	N
7. Tg = 1 and Fpg = 1 and Bp = 4	N
8. Tg = 1 and Fpg = 2 and Bp = 1	N
9. Tg = 1 and Fpg = 2 and Bp = 2 and Smoke = N	N
10. Tg = 1 and Fpg = 2 and Bp = 2 and Smoke = Y	Y
11. Tg = 1 and Fpg = 2 and Bp = 3	N
12. Tg = 1 and Fpg = 2 and Bp = 4	N
13. Tg = 1 and Fpg = 3	N
14. Tg = 2	N
15. Tg = 3 and Fpg = 1 and Smoke = N	Y
16. Tg = 3 and Fpg = 1 and Smoke = Y	N
17. Tg = 3 and Fpg = 2 and Bp = 1	Y
18. Tg = 3 and Fpg = 2 and Bp = 2	Y
19. Tg = 3 and Fpg = 2 and Bp = 3 and Smoke = N	Y
20. Tg = 3 and Fpg = 2 and Bp = 3 and Smoke = Y	N
21. Tg = 3 and Fpg = 2 and Bp = 4	N
22. Tg = 3 and Fpg = 3	Y
23. Tg = 4	N

ตัวอย่าง กฎข้อที่ 1 ถ้าค่าไขมันในเลือดน้อยกว่าหรือเท่ากับ 199 ค่าน้ำตาลในเลือดน้อยกว่า 110 ค่าความดันโลหิตค่าความดันซีสโตลิกน้อยกว่า 120 ค่าความดันไดแอสโตลิกน้อยกว่า 80 และไม่สูบบุหรี่ แล้วสถานะไม่เป็นโรคหัวใจขาดเลือด

กฎข้อที่ 2 ถ้าค่าไขมันในเลือดน้อยกว่าหรือเท่ากับ 199 ค่าน้ำตาลในเลือดน้อยกว่า 110 ค่าความดันโลหิตค่าความดันซีสโตลิกน้อยกว่า 120 ค่าความดันไดแอสโตลิกน้อยกว่า 80 และสูบบุหรี่ แล้วสถานะเป็นโรคหัวใจขาดเลือด

กฎข้อที่ 3 ถ้าค่าไขมันในเลือดน้อยกว่าหรือเท่ากับ 199 ค่าน้ำตาลในเลือดน้อยกว่า 110 ค่าความดันโลหิตค่าความดันซีสโตลิก 120 ถึง 139 หรือ ค่าความดันไดแอสโตลิก 80 ถึง 89 และไม่สูบบุหรี่ แล้วสถานะเป็นโรคหัวใจขาดเลือด

กฎข้อที่ 4 ถ้าค่าไขมันในเลือดน้อยกว่าหรือเท่ากับ 199 ค่าน้ำตาลในเลือดน้อยกว่า 110 ค่าความดันโลหิตค่าความดันซีสโตลิก 120 ถึง 139 หรือ ค่าความดันไดแอสโตลิก 80 ถึง 89 และสูบบุหรี่ แล้วสถานะไม่เป็นโรคหัวใจขาดเลือด เป็นต้น

2) เพื่อนบ้านใกล้ที่สุด

การสร้างแบบจำลองการพยากรณ์โรคหัวใจขาดเลือดด้วยเทคนิคเพื่อนบ้านใกล้ที่สุดพบว่ามีค่าความถูกต้อง (Accuracy) 73.9189% ค่าความแม่นยำ (Precision) 0.728 ค่าความระลึก (Recall) 0.739 และค่าความถ่วงดุล (F-Measure) 0.659 ดังภาพประกอบ 34

```
=== Summary ===

Correctly Classified Instances      261843               73.9189 %
Incorrectly Classified Instances     92387               26.0811 %
Kappa statistic                          0.11
Mean absolute error                      0.3374
Root mean squared error                  0.4107
Relative absolute error                 84.4631 %
Root relative squared error             91.9097 %
Total Number of Instances           354230

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                 0.984    0.903    0.741      0.984   0.845      0.187    0.732     0.888     N
                 0.097    0.016    0.694      0.097   0.170      0.187    0.732     0.470     Y
Weighted Avg.    0.739    0.659    0.728      0.739   0.659      0.187    0.732     0.773

=== Confusion Matrix ===

      a      b   <-- classified as
 252370   4180 |      a = N
  88207   9473 |      b = Y
```

ภาพประกอบ 34 ผลลัพธ์ประสิทธิภาพแบบจำลองโรคหัวใจขาดเลือดของเทคนิคเพื่อนบ้านใกล้ที่สุด

และจากภาพประกอบ 34 สามารถเขียนเป็นคอนฟิวชันเมทริกซ์ (Confusion Matrix) ได้ดังนี้

ตาราง 34 คอนฟิวชันเมทริกซ์ของแบบจำลองโรคหัวใจขาดเลือดของเทคนิคเพื่อนบ้านใกล้ที่สุด

Predicted / Actual	Y	N
Y	9473	88207
N	4180	252370

จากตาราง 34 คอนฟิวชันเมทริกซ์ของแบบจำลองโรคหัวใจขาดเลือดของเทคนิคเพื่อนบ้านใกล้ที่สุด ได้ค่าความถูกต้องเชิงบวก (True Positive: TP) หรือ ทำนายว่าเป็นโรคหัวใจขาดเลือดและระเบียนผู้ป่วยป่วยเป็นโรคหัวใจขาดเลือดจริง คือ 9,473 ระเบียน ได้ค่าความถูกต้องเชิงลบ (True Negative: TN) หรือ ทำนายว่าไม่เป็นโรคหัวใจขาดเลือดและผู้ป่วยไม่ได้ป่วยเป็นโรคหัวใจขาดเลือดถูกต้องมี 252,370 ระเบียน ความผิดพลาดเชิงบวก (False Positive: FP) หรือ ผลการทำนายว่าเป็นโรคหัวใจขาดเลือดแต่ผู้ป่วยไม่ได้ป่วยเป็นโรคหัวใจขาดเลือดมี 88,207 ระเบียน และความผิดพลาดเชิงลบ (False Negative: FN) หรือผลการทำนายว่าไม่เป็นโรคหัวใจขาดเลือด แต่ผู้ป่วยป่วยเป็นโรคหัวใจขาดเลือดมี 4,180 ระเบียน

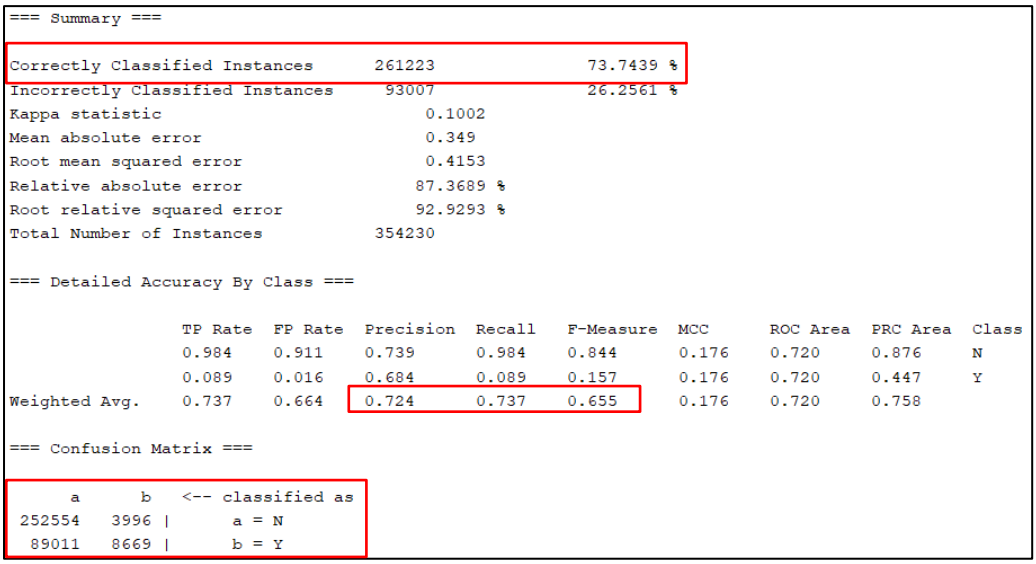
ตาราง 35 ตัวอย่างแบบจำลองโรคหัวใจขาดเลือดของเทคนิคเพื่อนบ้านใกล้ที่สุด

Data	Gender	Age	Bmi	Smoke	Bp	FPG	Tg	Distance	Heart	Nearest Sign
D1	1	2	4	0	1	1	1	3.16	–	
Distance $D1 = \sqrt{(1-1)^2 + (2-2)^2 + (2-4)^2 + (0-0)^2 + (3-1)^2 + (2-1)^2 + (2-1)^2}$ = 3.16										
D2	0	2	5	0	2	3	2	3.46	–	
Distance $D2 = \sqrt{(1-0)^2 + (2-2)^2 + (2-5)^2 + (0-0)^2 + (3-2)^2 + (2-3)^2 + (2-2)^2}$ = 3.46										
D3	1	1	2	0	3	1	1	1.73	–	–
Distance $D3 = \sqrt{(1-1)^2 + (2-1)^2 + (2-2)^2 + (0-0)^2 + (3-3)^2 + (2-1)^2 + (2-1)^2}$ = 1.73										
D4	1	1	4	0	2	1	1	2.82	+	
Distance $D4 = \sqrt{(1-1)^2 + (2-1)^2 + (2-4)^2 + (0-0)^2 + (3-2)^2 + (2-1)^2 + (2-1)^2}$ = 2.82										
D5	0	2	2	0	2	1	2	1.73	–	–
Distance $D5 = \sqrt{(1-0)^2 + (2-2)^2 + (2-2)^2 + (0-0)^2 + (3-2)^2 + (2-1)^2 + (2-2)^2}$ = 1.73										
D6	1	1	2	0	1	3	1	2.64	+	+
Distance $D6 = \sqrt{(1-1)^2 + (2-1)^2 + (2-2)^2 + (0-0)^2 + (3-1)^2 + (2-3)^2 + (2-1)^2}$ = 2.64										
K=3	1	2	2	0	3	2	2		?	

จากตาราง 35 แบบจำลองโรคหัวใจขาดเลือดของเทคนิคเพื่อนบ้านใกล้ที่สุด ของตัวอย่างข้อมูลบางส่วนกำหนดให้ คลาส Heart = Y แทนด้วย + และคลาส Heart = N แทนด้วย –, Gender = F แทนด้วย 0 และ Gender = M แทนด้วย 1, Smoke = N แทนด้วย 0 และ Smoke = Y แทนด้วย 1 จะใช้วิธีการวัดระยะทางแบบยุคลิด (Euclidean Distance) โดยกำหนด K = 3 พบว่า คลาส Heart = – ปรากฏมากที่สุด ดังนั้นจึงสรุปได้ว่าข้อมูล Gender = M, Age = 2, Bmi = 2, Smoke = N, Bp = 3, FPG = 2 และ Tg = 2 จะทำนายว่ามีคลาส Heart = N

3) นาอีฟเบย์

การสร้างแบบจำลองการพยากรณ์โรคหัวใจขาดเลือดด้วยเทคนิคนาอีฟเบย์พบว่ามีค่าความถูกต้อง (Accuracy) 73.7439% ค่าความแม่นยำ (Precision) 0.724 ค่าความระลึก (Recall) 0.737 และค่าความถ่วงดุล (F–Measure) 0.655 ดังภาพประกอบ 35



```
=== Summary ===

Correctly Classified Instances      261223               73.7439 %
Incorrectly Classified Instances     93007               26.2561 %
Kappa statistic                          0.1002
Mean absolute error                      0.349
Root mean squared error                  0.4153
Relative absolute error                 87.3689 %
Root relative squared error             92.9293 %
Total Number of Instances           354230

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.984    0.911    0.739      0.984    0.844      0.176    0.720     0.876     N
                 0.089    0.016    0.684      0.089    0.157      0.176    0.720     0.447     Y
Weighted Avg.    0.737    0.664    0.724      0.737    0.655      0.176    0.720     0.758

=== Confusion Matrix ===

      a      b   <-- classified as
 252554   3996 |      a = N
  89011   8669 |      b = Y
```

ภาพประกอบ 35 ผลลัพธ์ประสิทธิภาพแบบจำลองโรคหัวใจขาดเลือดของเทคนิคนาอีฟเบย์

และจากภาพประกอบ 35 สามารถเขียนเป็นคอนฟิวชันเมทริกซ์ (Confusion Matrix) ได้ดังนี้

ตาราง 36 คอนฟิวชันเมทริกซ์ของแบบจำลองโรคหัวใจขาดเลือดของเทคนิคนาอีฟเบย์

Predicted / Actual	Y	N
Y	8669	89011
N	3996	252554

จากตาราง 36 คอนฟิวชันเมทริกซ์ของแบบจำลองโรคหัวใจขาดเลือดของเทคนิคนาอีฟเบย์ ได้ค่าความถูกต้องเชิงบวก (True Positive: TP) หรือ ทำนายว่าเป็นโรคหัวใจขาดเลือดและระเบียนผู้ป่วยป่วยเป็นโรคหัวใจขาดเลือดจริง คือ 8,669 ระเบียน ได้ค่าความถูกต้องเชิงลบ (True Negative: TN) หรือ ทำนายว่าไม่เป็นโรคหัวใจขาดเลือดและผู้ป่วยไม่ได้ป่วยเป็นโรคหัวใจขาดเลือดถูกต้องมี 252,554 ระเบียน ความผิดพลาดเชิงบวก (False Positive: FP) หรือ ผลการทำนายว่าเป็นโรคหัวใจขาดเลือดแต่ผู้ป่วยไม่ได้ป่วยเป็นโรคหัวใจขาดเลือดมี 89,011 ระเบียน และความผิดพลาดเชิงลบ (False Negative: FN) หรือผลการทำนายว่าไม่เป็นโรคหัวใจขาดเลือด แต่ผู้ป่วยป่วยเป็นโรคหัวใจขาดเลือดมี 3,996 ระเบียน

ตาราง 37 แบบจำลองโรคหัวใจขาดเลือดของเทคนิคนาอีฟเบย์

แอตทริบิวต์	คลาส Heart = Y	คลาส Heart = N	ความน่าจะเป็นคลาส Heart = Y	ความน่าจะเป็นคลาส Heart = N
Gender = F	152636	45057	0.59	0.46
Gender = M	103916	52625	0.41	0.54
Age = 1	140384	45368	0.55	0.46
Age = 2	88623	37217	0.34	0.38
Age = 3	27546	15098	0.11	0.16
Bmi = 1	47839	21546	0.19	0.22
Bmi = 2	100595	37999	0.39	0.39
Bmi = 3	42878	18068	0.17	0.19
Bmi = 4	52855	17631	0.20	0.18
Bmi = 5	12388	2441	0.05	0.02
Smoke = N	251921	96433	0.98	0.99
Smoke = Y	4631	1249	0.02	0.01
Bp = 1	99438	39212	0.39	0.40
Bp = 2	123943	48993	0.48	0.50
Bp = 3	28127	8326	0.11	0.09
Bp = 4	5046	1153	0.02	0.01
Fpg = 1	192968	84849	0.75	0.87
Fpg = 2	28185	6436	0.11	0.07
Fpg = 3	35400	6398	0.14	0.06
Tg = 1	164659	85681	0.64	0.88
Tg = 2	87019	2287	0.34	0.02
Tg = 3	4546	9713	0.02	0.10
Tg = 4	330	3	0.00	0.00

จากตาราง 37 แบบจำลองโรคหัวใจขาดเลือดของเทคนิคนาอีฟเบย์ สามารถนำค่าไปหาความน่าจะเป็นโดยใช้ทฤษฎีของเบย์ (Bayes Theorem) ตามสมการ

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

เช่น Gender = F, Age = 1, Bmi = 1, Smoke = N, Bp = 1, Fpg = 1, Tg = 1 และ Heart = ? สามารถคำนวณได้ดังนี้

P(Heart = Y|A) = P(Gender = F | Heart = Y) x P(Age = 1 | Heart = Y)
x P(Bmi = 1 | Heart = Y) x P(Smoke = N | Heart = Y)
x P(Bp = 1 | Heart = Y) x P(Fpg = 1 | Heart = Y)
x P(Tg = 1 | Heart = Y)
= 0.59 x 0.55 x 0.19 x 0.98 x 0.39 x 0.75 x 0.64
= 0.01131

P(Heart = N|A) = P(Gender = F | Heart = N) x P(Age = 1 | Heart = N)
x P(Bmi = 1 | Heart = N) x P(Smoke = N | Heart = N)
x P(Bp = 1 | Heart = N) x P(Fpg = 1 | Heart = N)
x P(Tg = 1 | Heart = N)
= 0.46 x 0.46 x 0.22 x 0.99 x 0.40 x 0.87 x 0.88
= 0.01411

ดังนั้นข้อมูล Gender = F, Age = 1, Bmi = 1, Smoke = N, Bp = 1, Fpg = 1 และ Tg = 1 จะทำนายว่ามีคลาส Heart = N เนื่องจากค่า P(Heart = N|A) = 0.01411 ซึ่งมีค่ามากกว่า P(Heart = Y|A) = 0.01131

การเปรียบเทียบประสิทธิภาพ

1. การเปรียบเทียบประสิทธิภาพแบบจำลองโรคเบาหวาน

จากการศึกษาและเปรียบเทียบประสิทธิภาพของแบบจำลองโรคเบาหวานด้วยเทคนิค 3 เทคนิค ได้ผลประสิทธิภาพของแบบจำลอง ค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าความถ่วงดุล (F–Measure) ดังตาราง 38

ตาราง 38 ผลประสิทธิภาพของแบบจำลองโรคเบาหวาน

เทคนิคที่ใช้	ค่าความถูกต้อง(Accuracy)	ค่าความแม่นยำ (Precision)	ค่าความระลึก (Recall)	ค่าความถ่วงดุล (F–Measure)
เทคนิคนาอีฟเบย์	0.950	0.951	0.951	0.950
เทคนิค เพื่อนบ้านใกล้ที่สุด	0.951	0.952	0.952	0.952
เทคนิคต้นไม้ตัดสินใจ	0.952	0.952	0.952	0.952

จากตาราง 38 ผลประสิทธิภาพแบบจำลอง 3 เทคนิคของโรคเบาหวานพบว่า ค่าความถูกต้อง (Accuracy) ของเทคนิคต้นไม้ตัดสินใจ (ค่าความถูกต้อง 0.952) มีค่าสูงสุดและมีค่ามากกว่า เทคนิคเพื่อนบ้านใกล้ที่สุด (ค่าความถูกต้อง 0.951) และเทคนิคนาอีฟเบย์ (ค่าความถูกต้อง 0.950) ตามลำดับ

ค่าความแม่นยำ (Precision) ของเทคนิคต้นไม้ตัดสินใจ (ค่าความแม่นยำ 0.952) และเทคนิคเพื่อนบ้านใกล้ที่สุด (ค่าความแม่นยำ 0.952) มีค่าสูงสุดและมีค่ามากกว่าเทคนิคนาอีฟเบย์ (ค่าความแม่นยำ 0.951)

ค่าความระลึก (Recall) ของเทคนิคต้นไม้ตัดสินใจ (ค่าความระลึก 0.952) และเทคนิคเพื่อนบ้านใกล้ที่สุด (ค่าความระลึก 0.952) มีค่าสูงสุดและมีค่ามากกว่าเทคนิคนาอีฟเบย์ (ค่าความระลึก 0.951)

ค่าความถ่วงดุล (F–Measure) ของเทคนิคต้นไม้ตัดสินใจ (ค่าความถ่วงดุล 0.952) และเทคนิคเพื่อนบ้านใกล้ที่สุด(ค่าความถ่วงดุล 0.952) มีค่าสูงสุดและมีค่ามากกว่าเทคนิคนาอีฟเบย์ (ค่าความถ่วงดุล 0.950)

จากการเปรียบเทียบค่าการทำนายในแบบจำลองโรคเบาหวาน พบว่ามีค่าความถูกต้อง ค่าความแม่นยำ ค่าระลึกและค่าความถ่วงดุลสูงกว่าเทคนิคเพื่อนบ้านใกล้ที่สุดและเทคนิคนาอีฟเบย์

2. การเปรียบเทียบประสิทธิภาพแบบจำลองโรคความดันโลหิตสูง

จากการศึกษาและเปรียบเทียบประสิทธิภาพของแบบจำลองโรคความดันโลหิตสูงด้วยเทคนิค 3 เทคนิค ได้ผลประสิทธิภาพของแบบจำลอง ค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าความถ่วงดุล (F–Measure) ดังตาราง 39

ตาราง 39 ผลประสิทธิภาพของแบบจำลองโรคความดันโลหิตสูง

เทคนิคที่ใช้	ค่าความถูกต้อง(Accuracy)	ค่าความแม่นยำ (Precision)	ค่าความระลึก (Recall)	ค่าความถ่วงดุล (F–Measure)
เทคนิคนาอีฟเบย์	0.919	0.919	0.919	0.919
เทคนิค เพื่อนบ้านใกล้ที่สุด	0.921	0.922	0.922	0.922
เทคนิคต้นไม้ตัดสินใจ	0.921	0.922	0.922	0.922

จากตาราง 39 ผลประสิทธิภาพแบบจำลอง 3 เทคนิคของโรคความดันโลหิตสูง พบว่า ค่าความถูกต้อง (Accuracy) ของเทคนิคต้นไม้ตัดสินใจ (ค่าความถูกต้อง 0.921) เท่ากันกับเทคนิคเพื่อนบ้านใกล้ที่สุด (ค่าความถูกต้อง 0.921) มีค่าสูงสุด และมีค่ามากกว่าเทคนิคนาอีฟเบย์ (ค่าความถูกต้อง 0.919)

ค่าความแม่นยำ (Precision) ของเทคนิคต้นไม้ตัดสินใจ (ค่าความแม่นยำ 0.922) และเทคนิคเพื่อนบ้านใกล้ที่สุด (ค่าความแม่นยำ 0.922) มีค่าสูงสุด และมีค่ามากกว่าเทคนิคนาอีฟเบย์ (ค่าความแม่นยำ 0.919)

ค่าความระลึก (Recall) ของเทคนิคต้นไม้ตัดสินใจ (ค่าความระลึก 0.922) และเทคนิคเพื่อนบ้านใกล้ที่สุด (ค่าความระลึก 0.922) มีค่าสูงสุด และมีค่ามากกว่าเทคนิคนาอีฟเบย์ (ค่าความระลึก 0.919)

ค่าความถ่วงดุล (F–Measure) ของเทคนิคต้นไม้ตัดสินใจ (ค่าความถ่วงดุล 0.922) และเทคนิคเพื่อนบ้านใกล้ที่สุด(ค่าความถ่วงดุล 0.922) มีค่าสูงสุด และมีค่ามากกว่าเทคนิคนาอีฟเบย์ (ค่าความถ่วงดุล 0.919)

จากการเปรียบเทียบค่าการทำนายในแบบจำลองโรคความดันโลหิตสูง พบว่ามีค่าความถูกต้อง ค่าความแม่นยำ ค่าระลึกและค่าความถ่วงดุลสูงกว่าเทคนิคเพื่อนบ้านใกล้ที่สุดและเทคนิคนาอีฟเบย์

3. การเปรียบเทียบประสิทธิภาพแบบจำลองโรคหัวใจขาดเลือด

จากการศึกษาและเปรียบเทียบประสิทธิภาพของแบบจำลองโรคหัวใจขาดเลือดด้วยเทคนิค 3 เทคนิค ได้ผลประสิทธิภาพของแบบจำลอง ค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าความถ่วงดุล (F–Measure) ดังตาราง 40

ตาราง 40 ผลประสิทธิภาพของแบบจำลองโรคหัวใจขาดเลือด

เทคนิคที่ใช้	ค่าความถูกต้อง(Accuracy)	ค่าความแม่นยำ (Precision)	ค่าความระลึก (Recall)	ค่าความถ่วงดุล (F–Measure)
เทคนิคนาอีฟเบย์	0.737	0.724	0.737	0.655
เทคนิค เพื่อนบ้านใกล้ที่สุด	0.739	0.728	0.739	0.659
เทคนิคต้นไม้ตัดสินใจ	0.739	0.728	0.739	0.659

จากตาราง 40 ผลประสิทธิภาพแบบจำลอง 3 เทคนิคของโรคความดันโลหิตสูง พบว่า ค่าความถูกต้อง (Accuracy) ของเทคนิคต้นไม้ตัดสินใจ (ค่าความถูกต้อง 0.739) เท่ากันกับเทคนิคเพื่อนบ้านใกล้ที่สุด (ค่าความถูกต้อง 0.739) มีค่าสูงสุด และมีค่ามากกว่าเทคนิคนาอีฟเบย์ (ค่าความถูกต้อง 0.737)

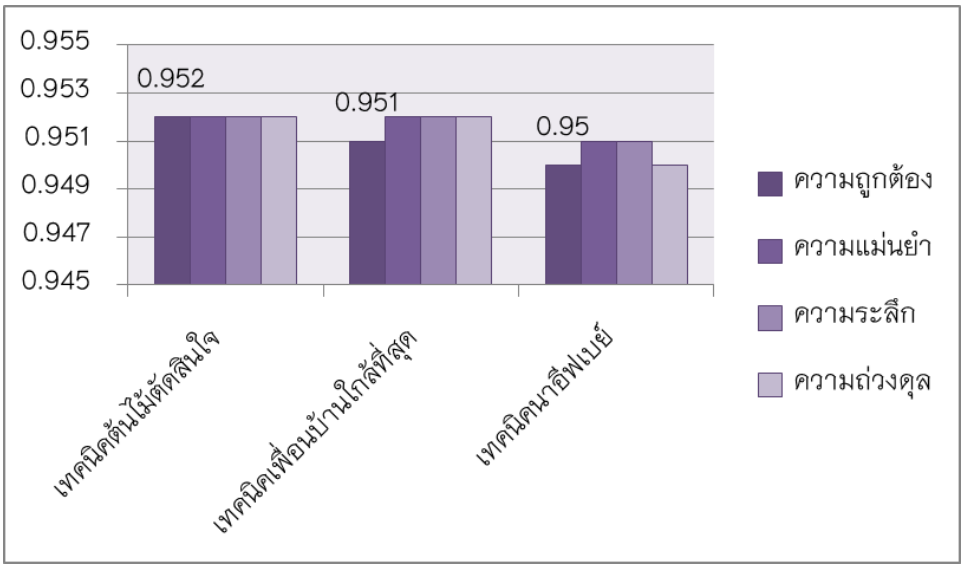
ค่าความแม่นยำ (Precision) ของเทคนิคต้นไม้ตัดสินใจ (ค่าความแม่นยำ 0.728) เท่ากันกับเทคนิคเพื่อนบ้านใกล้ที่สุด (ค่าความแม่นยำ 0.728) มีค่าสูงสุด และมีค่ามากกว่าเทคนิคนาอีฟเบย์ (ค่าความแม่นยำ 0.724)

ค่าความระลึก (Recall) ของเทคนิคต้นไม้ตัดสินใจ (ค่าความระลึก 0.739) เท่ากันกับเทคนิคเพื่อนบ้านใกล้ที่สุด (ค่าความระลึก 0.739) มีค่าสูงสุด และมีค่ามากกว่าเทคนิคนาอีฟเบย์ (ค่าความระลึก 0.737)

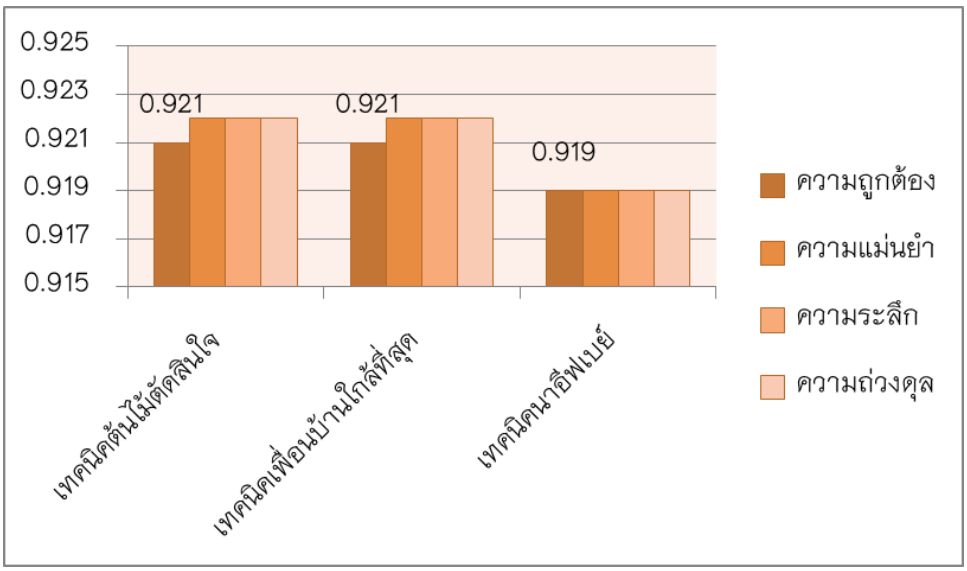
ค่าความถ่วงดุล (F–Measure) ของเทคนิคต้นไม้ตัดสินใจ (ค่าความถ่วงดุล 0.659) เท่ากันกับเทคนิคเพื่อนบ้านใกล้ที่สุด (ค่าความถ่วงดุล 0.659) มีค่าสูงสุด และมีค่ามากกว่าเทคนิคนาอีฟเบย์ (ค่าความถ่วงดุล 0.655)

จากการเปรียบเทียบค่าการทำนายในแบบจำลองโรคหัวใจขาดเลือด พบว่า มีค่าความถูกต้อง ค่าความแม่นยำ ค่าระลึกและค่าความถ่วงดุลสูงกว่าเทคนิคเพื่อนบ้านใกล้ที่สุดและเทคนิคนาอีฟเบย์

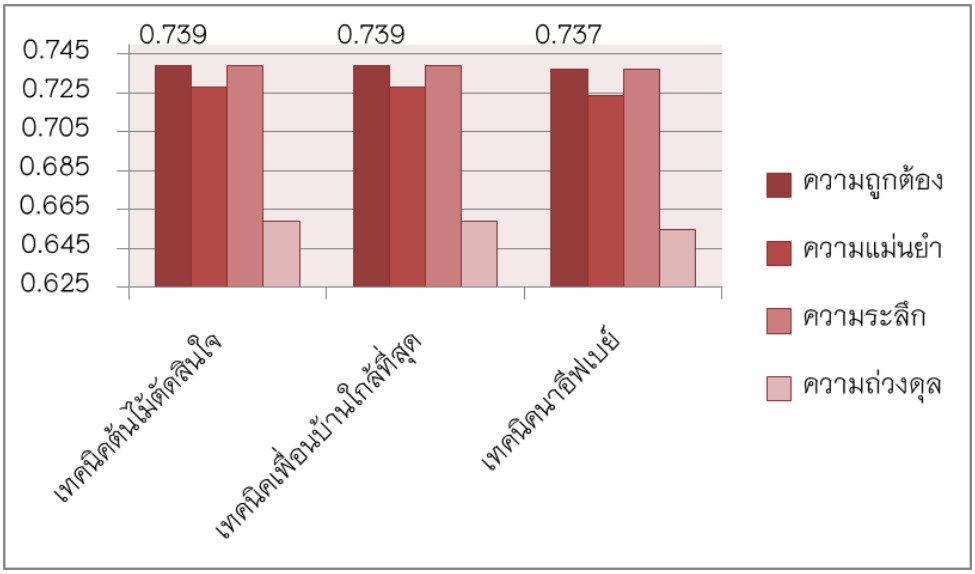
นำประสิทธิภาพของแบบจำลองทั้ง 3 โรคมาเปรียบเทียบกันดังภาพประกอบ 36 ภาพประกอบ 37 และภาพประกอบ 38



ภาพประกอบ 36 เทียบประสิทธิภาพของแบบจำลองโรคเบาหวาน



ภาพประกอบ 37 เทียบประสิทธิภาพของแบบจำลองโรคความดันโลหิตสูง



ภาพประกอบ 38 เทียบประสิทธิภาพของแบบจำลองโรคหัวใจขาดเลือด

จากภาพประกอบ 36 ภาพประกอบ 37 และภาพประกอบ 38 หากพิจารณาในภาพรวมการพยากรณ์โรค 3 โรคได้แก่ โรคเบาหวาน โรคความดันโลหิตสูง และโรคหัวใจขาดเลือด ด้วยเทคนิคเหมืองข้อมูล 3 เทคนิคได้แก่ เทคนิคต้นไม้ตัดสินใจ เทคนิคเพื่อนบ้านใกล้ที่สุด และเทคนิคนาอีฟเบย์ พบว่าเทคนิคที่ดีที่สุดคือ เทคนิคต้นไม้ตัดสินใจ เพราะมีค่าความถูกต้อง ค่าความแม่นยำ ค่าระลึก และค่าความถ่วงดุลสูงสุด

การประเมินประสิทธิภาพแบบจำลองโดยผู้เชี่ยวชาญ

ผู้วิจัยนำแบบจำลองต้นไม้การตัดสินใจเพื่อพยากรณ์การเกิดโรคเรื้อรังในผู้สูงอายุประกอบด้วย โรคเบาหวาน โรคความดันโลหิตสูง และโรคหัวใจขาดเลือด ที่พัฒนาด้วยเทคนิคเหมืองข้อมูลให้ผู้เชี่ยวชาญด้านสาธารณสุขประเมินประสิทธิภาพของแบบจำลอง ด้วยเทคนิคการประชุมกลุ่ม(Focus group) จำนวน 7 คน เมื่อวันที่ 9 เดือนมีนาคม พ.ศ. 2564 ณ ห้องประชุมพุทธรักษา โรงพยาบาลนาหว้า จังหวัดนครพนม ซึ่งมีรายชื่อดังตาราง 41

ตาราง 41 ข้อมูลผู้เชี่ยวชาญ

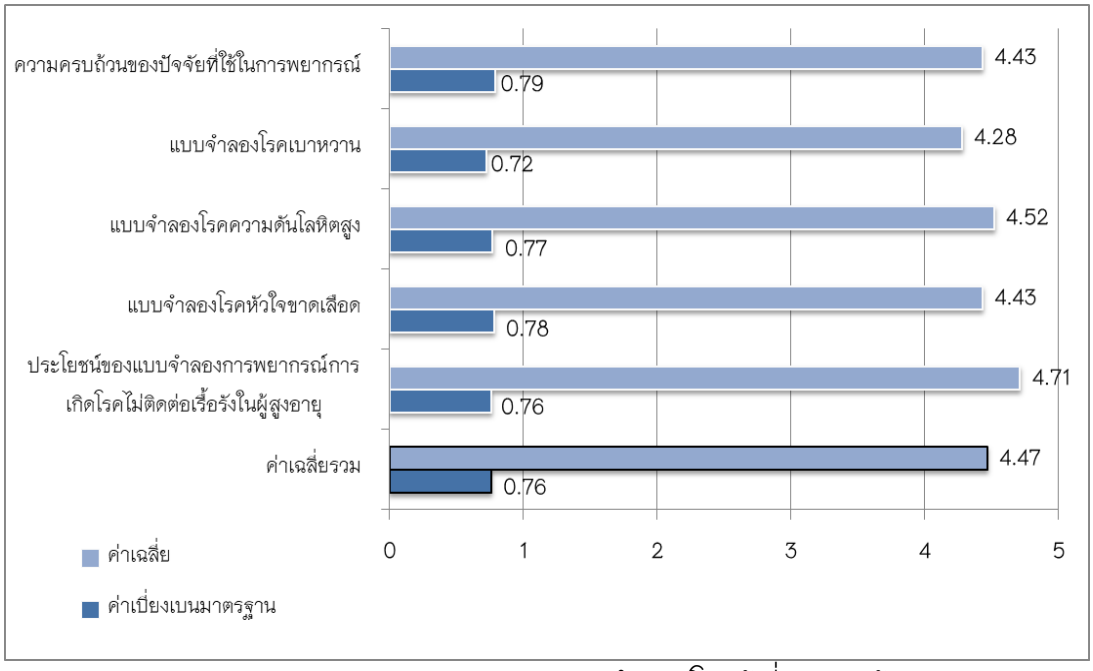
ชื่อ–สกุล	ตำแหน่ง/สังกัด
นายแพทย์กิตติเชษฐ์ ธีรกุลพงศ์เวช	นายแพทย์ชำนาญการ หน่วยงาน โรงพยาบาลนาหว้า
นางสาวกรรณิการ์ อุดมกัน	นักวิชาการสาธารณสุขชำนาญการ หน่วยงาน โรงพยาบาลนาหว้า
นางอุไรลักษณ์ คำหา	พยาบาลวิชาชีพชำนาญการ หน่วยงาน โรงพยาบาลนาหว้า
นางอรรชร สุนทรารักษ์	พยาบาลวิชาชีพชำนาญการ หน่วยงาน โรงพยาบาลนาหว้า
นางสาววิภาสินี วะเศษสร้อย	พยาบาลวิชาชีพชำนาญการ หน่วยงาน โรงพยาบาลนาหว้า
นางพูล สุวัฒนเสนีย์	พยาบาลวิชาชีพชำนาญการ หน่วยงาน โรงพยาบาลนาหว้า
นางสาวทัศนีย์ สิทธิ	พยาบาลวิชาชีพชำนาญการ หน่วยงาน โรงพยาบาลนาหว้า

ผลการประเมินประสิทธิภาพแบบจำลองเพื่อพยากรณ์การเกิดโรค 3 โรคโดยผู้เชี่ยวชาญ มีรายละเอียดดังตาราง 42

ตาราง 42 สรุปผลประสิทธิภาพของแบบจำลองโดยผู้เชี่ยวชาญด้านสาธารณสุข

ประเด็นความคิดเห็น	ค่าเฉลี่ย	ค่าเบี่ยงเบนมาตรฐาน	ความหมายของระดับ
1. ความครบถ้วนของปัจจัยที่ใช้ในการพยากรณ์	4.43	0.79	มาก
2. แบบจำลองโรคเบาหวาน			
2.1 ความถูกต้องของโมเดล	4.43	0.79	มาก
2.2 ความน่าเชื่อถือของผลการทำนาย	4.00	0.58	มาก
2.3 ความสามารถนำไปประกอบการตัดสินใจ	4.43	0.79	มาก
ค่าเฉลี่ยรวมแบบจำลองโรคเบาหวาน	4.28	0.72	มาก
3. แบบจำลองโรคความดันโลหิตสูง			
3.1 ความถูกต้องของโมเดล	4.71	0.76	มากที่สุด
3.2 ความน่าเชื่อถือของผลการทำนาย	4.29	0.76	มาก
3.3 ความสามารถนำไปประกอบการตัดสินใจ	4.57	0.79	มากที่สุด
ค่าเฉลี่ยรวมแบบจำลองโรคความดันโลหิตสูง	4.52	0.77	มากที่สุด
4. แบบจำลองโรคหัวใจขาดเลือด			
4.1 ความถูกต้องของโมเดล	4.57	0.79	มากที่สุด
4.2 ความน่าเชื่อถือของผลการทำนาย	4.29	0.76	มาก
4.3 ความสามารถนำไปประกอบการตัดสินใจ	4.43	0.79	มาก
ค่าเฉลี่ยรวมแบบจำลองโรคหัวใจขาดเลือด	4.43	0.78	มาก
5. ประโยชน์ของแบบจำลองการพยากรณ์การเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุ	4.71	0.76	มากที่สุด
ค่าเฉลี่ยรวม	4.47	0.76	มาก

จากตาราง 42 ผลการประเมินประสิทธิภาพแบบจำลองเพื่อการพยากรณ์การเกิดโรคพบว่า ความครบถ้วนของปัจจัยที่ใช้ในการพยากรณ์ มีค่าเฉลี่ยเท่ากับ 4.43 (อยู่ในระดับมาก) แบบจำลองโรคเบาหวานมีค่าความถูกต้องของแบบจำลอง ความน่าเชื่อถือของผลการทำนาย และความสามารถนำไปประกอบการตัดสินใจ มีค่าเฉลี่ยเท่ากับ 4.43 4.40 และ 4.43 ตามลำดับ และมีค่าเฉลี่ยรวมเท่ากับ 4.28 (อยู่ในระดับมาก) แบบจำลองโรคความดันโลหิตสูงมีค่าความถูกต้องของแบบจำลอง ความน่าเชื่อถือของผลการทำนาย และความสามารถนำไปประกอบการตัดสินใจ มีค่าเฉลี่ยเท่ากับ 4.71 4.29 และ 4.57 ตามลำดับ และมีค่าเฉลี่ยรวมเท่ากับ 4.52 (อยู่ในระดับมากที่สุด) แบบจำลองโรคหัวใจขาดเลือดมีค่าความถูกต้องของแบบจำลอง ความน่าเชื่อถือของผลการทำนาย และความสามารถนำไปประกอบการตัดสินใจ มีค่าเฉลี่ยเท่ากับ 4.57 4.29 และ 4.43 ตามลำดับ และมีค่าเฉลี่ยรวมเท่ากับ 4.43 (อยู่ในระดับมาก) ประโยชน์ของแบบจำลองการพยากรณ์การเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุมีค่าเฉลี่ยเท่ากับ 4.71 (อยู่ในระดับมากที่สุด) มีค่าเฉลี่ยรวมในทุกด้านประเด็นเท่ากับ 4.47 (อยู่ในระดับมาก) ดังภาพประกอบ 39



ภาพประกอบ 39 ผลการประเมินประสิทธิภาพแบบจำลองโดยผู้เชี่ยวชาญด้านสาธารณสุข

นอกจากนี้ผู้เชี่ยวชาญมีข้อเสนอแนะเพิ่มเติมเกี่ยวกับแบบจำลองเพื่อการพยากรณ์การเกิดโรค ดังนี้

1) ทบทวนข้อมูลนำเข้า เช่น ผู้ที่เป็นโรคแล้วได้รับการรักษาในระยะเวลาหนึ่งอาจไม่นำเข้า เพราะมีปัจจัยการรักษาอื่นที่เกี่ยวข้อง เช่นการรับประทานยา ซึ่งจะทำให้ผลจากค่าตัวแปรแตกต่างจากที่ควรจะเป็น

2) การทบทวนปัจจัยเสี่ยง ควรแยกโรคและปัจจัยเฉพาะโรคเพื่อการพยากรณ์ที่แม่นยำขึ้น เช่น การใช้ค่าตัวแปรไขมันชนิดไม่ดี LDL (Low–Density Lipoprotein) แทนการใช้ค่าตัวแปรไขมันไตรกลีเซอไรด์ ในกลุ่มผู้ป่วยเบาหวานและความดันโลหิต เป็นต้น เพื่อให้แบบพยากรณ์แม่นยำยิ่งขึ้น

3) สามารถพัฒนาการวิจัยให้สามารถขยายผลใช้กับการพยากรณ์โรคอื่น เช่น โรคติดต่อและโรคไม่ติดต่ออื่น กลุ่มอายุอื่น หรือ การวิเคราะห์หาปัจจัยที่ทำให้น้ำตาลในเลือดสูงในกลุ่มผู้ป่วยเบาหวาน เพื่อนำมาพัฒนางานต่อไป

บทที่ 5

สรุปผล อภิปรายผลและข้อเสนอแนะ

งานวิจัยเรื่อง แบบจำลองเพื่อพยากรณ์การเกิดโรคเรื้อรังในผู้สูงอายุด้วยเทคนิคเหมืองข้อมูล: กรณีศึกษาโรงพยาบาลนาหว้า จังหวัดนครพนม ผู้วิจัยได้ดำเนินการวิจัย สรุปผลการวิจัย อภิปรายผลการวิจัย และนำเสนอข้อเสนอแนะดังนี้

1. ความมุ่งหมายของการวิจัย
2. วิธีการดำเนินการวิจัย
3. สรุปผลการวิจัย
4. อภิปรายผลการวิจัย
5. ข้อเสนอแนะ

ความมุ่งหมายของการวิจัย

ในการวิจัยครั้งนี้ผู้วิจัยได้กำหนดความมุ่งหมายของการวิจัย ได้ดังนี้

1) เพื่อสร้างแบบจำลองเพื่อพยากรณ์การเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุ

2) เพื่อเปรียบเทียบประสิทธิภาพของแบบจำลองการเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุ

วิธีการดำเนินการวิจัย

งานวิจัยเรื่อง แบบจำลองเพื่อพยากรณ์การเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุด้วยเทคนิคเหมืองข้อมูล: กรณีศึกษาโรงพยาบาลนาหว้า จังหวัดนครพนม มีการดำเนินการ 3 ระยะดังนี้

1. ระยะที่ 1 วิเคราะห์และเตรียมข้อมูล

เป็นขั้นตอนศึกษาทำความเข้าใจกับเนื้อหาข้อมูลและเตรียมข้อมูลสำหรับการนำไปทำเหมืองข้อมูล โดยอ้างอิงจากกระบวนเหมืองข้อมูลของคริสป์–ดีเอ็ม (CRISP–DM) ประกอบด้วย 3 ขั้นคือ 1) ขั้นตอนการทำความเข้าใจการวิจัยและธุรกิจ 2) ขั้นตอนการทำความเข้าใจข้อมูล และ 3) ขั้นตอนการจัดเตรียมข้อมูล

1.1 ข้อมูล

ผู้วิจัยทำหนังสือขอความอนุเคราะห์ข้อมูลจากโรงพยาบาลนาหว้า จังหวัดนครพนม คือข้อมูลผู้รับบริการที่มีอายุ 60 ปีขึ้นไป และมารับการรักษาที่โรงพยาบาลนาหว้า จังหวัดนครพนม ตั้งแต่เดือนตุลาคม พ.ศ. 2554 ถึง เดือนกันยายน พ.ศ. 2562 จำนวน 262,655 ระเบียน

โดยผู้วิจัยได้เสนอและผ่านการพิจารณาจริยธรรมการวิจัยในมนุษย์กับคณะกรรมการจริยธรรมการวิจัยในมนุษย์ มหาวิทยาลัยราชภัฏสกลนคร โดยได้รับเลขที่ใบรับรอง ๐๔๗/๒๕๖๓

1.2 เครื่องมือที่ใช้ในการวิจัย

โปรแกรมไมโครซอฟท์เอ็กเซล (Microsoft Excel) ใช้ในการเตรียมและจัดการข้อมูลให้เป็นรูปแบบไฟล์สำหรับเก็บข้อมูลในรูปแบบตาราง ใช้เครื่องหมายจุลภาคหรือคอมม่า (,) ในการแบ่ง (Comma Separated Value) เพื่อนำเข้าโปรแกรมเวก้า (Weka)

1.3 วิธีการดำเนินการวิจัย

1) ขั้นตอนการทำความเข้าใจการวิจัยและธุรกิจ

ผู้วิจัยทำการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องกับโรคไม่ติดต่อเรื้อรังในผู้สูงอายุ 3 โรคได้แก่ โรคเบาหวาน โรคความดันโลหิตสูง และโรคหัวใจขาดเลือด การพยากรณ์การเกิดโรคด้วยเทคนิคเหมืองข้อมูล 3 เทคนิค ได้แก่ เทคนิคต้นไม้ตัดสินใจ เทคนิคนาอีฟเบย์ และเทคนิคเพื่อนบ้านใกล้ที่สุด

2) ขั้นตอนการทำความเข้าใจข้อมูล

ผู้วิจัยทำความเข้าใจข้อมูลและตัวแปรที่ก่อให้เกิดโรคเบาหวาน จำนวน 7 ตัวแปร ได้แก่ เพศ อายุ การสูบบุหรี่ ดัชนีมวลกาย ความดันโลหิต น้ำตาลในเลือด ไขมันในเลือด และตัวแปรผลลัพธ์ ได้แก่ เป็นหรือไม่เป็นโรค

3) ขั้นตอนการจัดเตรียมข้อมูล

นำข้อมูลในขั้นตอนที่ 2 มาทำความสะอาด (Cleansing) ด้วยโปรแกรมไมโครซอฟท์เอ็กเซล โดยตรวจสอบระเบียนข้อมูลที่มีค่าไม่ครบทุกฟิลด์หรือระเบียนที่มีค่าผิดปกติ (Outliner) เช่น อายุ 300 ปี น้ำหนัก 251 กิโลกรัม ผู้วิจัยแก้ไขโดยลบระเบียนที่มีค่าผิดปกติจำนวน 11 ระเบียน ดังนั้นเหลือระเบียบที่ใช้สร้างแบบจำลองจำนวน 262,644 ระเบียน จากนั้นทำการแปลงข้อมูลให้อยู่ในรูปแบบพร้อมใช้ในการวิเคราะห์

1.4 สถิติและการวิเคราะห์ข้อมูล

ผู้วิจัยใช้การวิเคราะห์ข้อมูลโดยอ้างอิงจากกระบวนเหมืองข้อมูลของคริสป์–ดีเอ็ม (CRISP–DM) ในขั้นตอนการจัดเตรียมข้อมูล ได้แก่ การแปลงข้อมูลและทำให้ข้อมูลถูกต้อง (Data Cleaning) คือ การแปลงข้อมูลให้อยู่ในช่วง (Scale) เดียวกันและการตัดข้อมูลที่ขาดหายไป

2. ระยะที่ 2 สร้างแบบจำลอง

เป็นขั้นตอนสร้างแบบจำลองทั้ง 3 โรคคือโรคเบาหวาน โรคความดันโลหิตสูง และโรคหัวใจขาดเลือด ด้วยโปรแกรมเวก้า ด้วยเทคนิคเหมืองข้อมูล 3 เทคนิคคือ ต้นไม้ตัดสินใจ เพื่อนบ้านใกล้ที่สุด และนาอีฟเบย์

2.1 ข้อมูล

ข้อมูลที่ผ่านกระบวนการระยะที่ 1 จำนวน 262,644 ระเบียน ในรูปแบบไฟล์สำหรับเก็บข้อมูลในรูปแบบตาราง ใช้เครื่องหมายจุลภาค หรือคอมม่า (,) ในการแบ่ง (Comma Separated Value)

2.2 เครื่องมือที่ใช้ในการวิจัย

โปรแกรมเวก้า (Weka) ใช้ในการนำเข้าข้อมูลและสร้างแบบจำลอง

2.3 วิธีการดำเนินการวิจัย

ผู้วิจัยนำข้อมูลที่ผ่านกระบวนการระยะที่ 1 ไปสร้างแบบจำลองทั้ง 3 โรค คือโรคเบาหวาน โรคความดันโลหิตสูง และโรคหัวใจขาดเลือด ด้วยโปรแกรมเวก้า ด้วยเทคนิคเหมืองข้อมูล 3 เทคนิคคือ ต้นไม้ตัดสินใจ เพื่อนบ้านใกล้ที่สุด และนาอีฟเบย์

ในการวัดประสิทธิภาพแบบจำลอง ผู้วิจัยได้ทำการทดลองตามหลักการแบ่งข้อมูลออกเป็นชุดเรียนรู้และชุดทดสอบ โดยการแบ่งข้อมูลออกเป็น 10 ส่วนเท่า ๆ กัน ใช้ 9 ส่วนเป็นข้อมูลสอน และ 1 ส่วน เป็นข้อมูลทดสอบ ทำงานสลับกันจนครบ 10 รอบ

(10 Fold Cross–Validation) การประเมินประสิทธิภาพใช้ตัววัดจากค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าความถ่วงดุล (F–Measure)

2.4 สถิติและการวิเคราะห์ข้อมูล

ผู้วิจัยใช้สถิติในการวิเคราะห์ข้อมูลของแบบจำลองทางเหมืองข้อมูล โดยวัดจากค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าความถ่วงดุล (F–Measure)

3. ระยะที่ 3 ประเมินประสิทธิภาพแบบจำลองโดยผู้เชี่ยวชาญ

เป็นขั้นตอนนำแบบจำลองแบบจำลองจากเทคนิคที่มีประสิทธิภาพดีที่สุดของโรคเบาหวาน โรคความดันโลหิตสูง และโรคหัวใจขาดเลือด ให้ผู้เชี่ยวชาญด้านสาธารณสุข ประเมินประสิทธิภาพของแบบจำลอง ด้วยวิธีประชุมกลุ่ม (Focus Group) โดยมีรายละเอียดดังนี้

3.1 วิธีการจัดประชุมกลุ่มย่อย

ผู้วิจัยนำแบบจำลองจากระยะที่ 2 มาเปรียบเทียบประสิทธิภาพการพยากรณ์การเกิดโรคเบาหวาน โรคความดันโลหิตสูง และโรคหัวใจขาดเลือด จากเทคนิค 3 เทคนิค แล้วเลือกค่าประสิทธิภาพดีที่สุด โดยอ้างอิงค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าความถ่วงดุล (F–Measure) ให้ผู้เชี่ยวชาญด้านสาธารณสุขที่มีประสบการณ์การทำงานด้านสาธารณสุขไม่น้อยกว่า 10 ปี หรือมีตำแหน่งชำนาญการขึ้นไป จำนวน 7 คน ประเมินประสิทธิภาพของแบบจำลองด้วยวิธีประชุมกลุ่ม (Focus Group)

3.2 เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการวิจัยระยะที่ 3 มีดังนี้

1) แบบจำลองจากเทคนิคที่มีประสิทธิภาพดีที่สุด โดยวัดจากค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าความถ่วงดุล (F–Measure)

2) ประเด็นการสนทนากลุ่ม (Focus group) แบ่งเป็น 5 ประเด็นคือ ความครบถ้วนของปัจจัยที่ใช้ในการพยากรณ์ ความถูกต้องของโมเดล ความน่าเชื่อถือของผลการทำนาย ความสามารถนำไปประกอบการตัดสินใจ และประโยชน์ของแบบจำลองการพยากรณ์การเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุ

3.3 วิธีการดำเนินการวิจัย

1) ผู้วิจัยติดต่อผู้เชี่ยวชาญจำนวน 7 คน

2) ผู้วิจัยประสานและส่งรายชื่อผู้เชี่ยวชาญให้เจ้าหน้าที่สำนักงานบัณฑิตวิทยาลัยเพื่อจัดทำหนังสือเชิญผู้เชี่ยวชาญ

3) ผู้วิจัยนำส่งหนังสือเชิญให้กับผู้เชี่ยวชาญจำนวน 7 คน

4) ผู้วิจัยขอใช้และจัดเตรียมห้องสนทนากลุ่ม โดยใช้ห้องประชุมพุทธรักษา โรงพยาบาลนาหว้า จังหวัดนครพนม

5) ผู้วิจัยดำเนินการจัดกิจกรรมสนทนากลุ่ม โดยผู้วิจัยนำเสนอแบบจำลองจากเทคนิคที่มีประสิทธิภาพดีที่สุดและแจกแบบประเมินประสิทธิภาพแบบจำลองจากประเด็นการสนทนากลุ่ม 5 ประเด็น

6) ถอดเทปบันทึกการสนทนากลุ่ม

7) วิเคราะห์และบันทึกข้อเสนอแนะ

8) เขียนผลการวิจัย

3.4 สถิติและการวิเคราะห์ข้อมูล

ค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐาน

สรุปผลการวิจัย

1) สรุปผลการสร้างแบบจำลองในข้อมูลผู้สูงอายุ โรงพยาบาลนาหว้า จังหวัดนครพนม เพื่อพยากรณ์การเกิดโรคไม่ติดต่อเรื้อรัง 3 โรค ด้วยเทคนิคเหมืองข้อมูล 3 เทคนิคได้ดังนี้

โรคเบาหวาน การวัดประสิทธิภาพแบบจำลองที่พัฒนาได้แบ่งเป็น 4 ค่าดังนี้

ค่าความถูกต้อง เทคนิคต้นไม้ตัดสินใจมีค่าเท่ากับ 0.952 มากกว่าเทคนิคเพื่อนบ้านใกล้ที่สุดมีค่าเท่ากับ 0.951 และเทคนิคนาอีฟเบย์มีค่าเท่ากับ 0.950

ค่าความแม่นยำ เทคนิคต้นไม้ตัดสินใจและเทคนิคเพื่อนบ้านใกล้ที่สุดมีค่าเท่ากับ 0.952 มากกว่าเทคนิคนาอีฟเบย์มีค่าเท่ากับ 0.951

ค่าความระลึก เทคนิคต้นไม้ตัดสินใจและเทคนิคเพื่อนบ้านใกล้ที่สุดมีค่าเท่ากับ 0.952 มากกว่าเทคนิคนาอีฟเบย์มีค่า 0.951

ค่าความถ่วงดุล เทคนิคต้นไม้ตัดสินใจและเทคนิคเพื่อนบ้านใกล้ที่สุดมีค่าเท่ากับ 0.952 มากกว่าเทคนิคนาอีฟเบย์มีค่าเท่ากับ 0.950

โรคความดันโลหิตสูง การวัดประสิทธิภาพแบบจำลองที่พัฒนาได้แบ่งเป็น 4 ค่าดังนี้

ค่าความถูกต้อง เทคนิคต้นไม้ตัดสินใจและเทคนิคเพื่อนบ้านใกล้ที่สุด มีค่าเท่ากับ 0.921 มากกว่าเทคนิคนาอีฟเบย์มีค่าเท่ากับ 0.919

ค่าความแม่นยำ เทคนิคต้นไม้ตัดสินใจและเทคนิคเพื่อนบ้านใกล้ที่สุด มีค่าเท่ากับ 0.922 มากกว่าเทคนิคนาอีฟเบย์มีค่าเท่ากับ 0.919

ค่าความระลึก เทคนิคต้นไม้ตัดสินใจและเทคนิคเพื่อนบ้านใกล้ที่สุด มีค่าเท่ากับ 0.922 มากกว่าเทคนิคนาอีฟเบย์มีค่า 0.919

ค่าความถ่วงดุล เทคนิคต้นไม้ตัดสินใจและเทคนิคเพื่อนบ้านใกล้ที่สุด มีค่าเท่ากับ 0.922 มากกว่าเทคนิคนาอีฟเบย์มีค่าเท่ากับ 0.919

โรคหัวใจขาดเลือด การวัดประสิทธิภาพแบบจำลองที่พัฒนาได้แบ่งเป็น 4 ค่าดังนี้

ค่าความถูกต้อง เทคนิคต้นไม้ตัดสินใจและเทคนิคเพื่อนบ้านใกล้ที่สุด มีค่าเท่ากับ 0.739 มากกว่าเทคนิคนาอีฟเบย์มีค่าเท่ากับ 0.737

ค่าความแม่นยำ เทคนิคต้นไม้ตัดสินใจและเทคนิคเพื่อนบ้านใกล้ที่สุด มีค่าเท่ากับ 0.728 มากกว่าเทคนิคนาอีฟเบย์มีค่าเท่ากับ 0.724

ค่าความระลึก เทคนิคต้นไม้ตัดสินใจและเทคนิคเพื่อนบ้านใกล้ที่สุด มีค่าเท่ากับ 0.739 มากกว่าเทคนิคนาอีฟเบย์มีค่า 0.737

ค่าความถ่วงดุล เทคนิคต้นไม้ตัดสินใจและเทคนิคเพื่อนบ้านใกล้ที่สุด มีค่าเท่ากับ 0.659 มากกว่าเทคนิคนาอีฟเบย์มีค่าเท่ากับ 0.655

2) สรุปผลการเปรียบเทียบประสิทธิภาพของแบบจำลองการเกิดโรคไม่ติดต่อเรื้อรัง 3 โรคกับข้อมูลผู้สูงอายุ โรงพยาบาลนาหว้า จังหวัดนครพนม ด้วยเทคนิคต้นไม้ตัดสินใจ เทคนิคเพื่อนบ้านใกล้ที่สุดและเทคนิคนาอีฟเบย์ พบว่า เทคนิคต้นไม้ตัดสินใจมีประสิทธิภาพดีที่สุดเพราะมีค่าความถูกต้อง ค่าความแม่นยำ ค่าความระลึก และค่าความถ่วงดุลสูงสุด

อภิปรายผลการวิจัย

1) เทคนิคต้นไม้ตัดสินใจสามารถพยากรณ์โรคเบาหวาน โรคความดันโลหิตสูง และโรคหัวใจขาดเลือด มีค่าความถูกต้องสูง มีความความแม่นยำสูง มีค่าความระลึกสูง และมีค่าความถ่วงดุลสูง หากเทียบกับเทคนิคเพื่อนบ้านใกล้ที่สุดและเทคนิคนาอีฟเบย์ สอดคล้องกับงานวิจัยของวิธวินท์ แสงมณี, วีระวุฒ รัตนเจริญเลิศ, ณัฏฐภพ โพธิรัชต์ และเพียงฤทัย หนูสวัสดิ์ (2560) ทำวิทยานิพนธ์เรื่อง การสร้างโมเดลทำนายโอกาสการกลับมารักษาตัวซ้ำของผู้ป่วยโรคเบาหวาน โดยใช้เทคนิคเหมืองข้อมูลและเปรียบเทียบประสิทธิภาพของแบบจำลองที่สร้างจากเทคนิคต้นไม้ตัดสินใจ เทคนิคเพื่อนบ้านใกล้ที่สุด และเทคนิคนาอีฟเบย์ ผลการวิจัยพบว่า แบบจำลองที่สร้างจากเทคนิคต้นไม้ตัดสินใจให้ค่าความถูกต้องมากที่สุดคิดเป็นร้อยละ 85.50 งานวิจัยของนพรัตน์ พจน์จิราภรณ์ (2557) ทำงานวิจัยเรื่องระบบสนับสนุนทางการแพทย์เพื่อวิเคราะห์โรคหลอดเลือดแดง โป่งพอง โดยใช้เทคนิคเหมืองข้อมูล กรณีศึกษาการรักษาหลอดเลือดแดงโป่งพองผ่านสายสวน โดยใช้ทฤษฎีต้นไม้ตัดสินใจ นาอีฟเบย์ เครือข่ายประสาทเทียม ซึ่งจากผลการทดลองของการศึกษาแสดงให้เห็นว่าแบบจำลองต้นไม้ตัดสินใจ มีประสิทธิภาพค่าความถูกต้องสูงที่สุด มีค่าเท่ากับร้อยละ 98.10 และ งานวิจัยของ Israa Ahmed Zriqat Ahmad Mousa Altamimi และ Mohammad Azzeh (2560) ทำงานวิจัยเรื่องการศึกษาเปรียบเทียบการทำนายโรคหัวใจ โดยใช้วิธีการจำแนกประเภทการทำเหมืองข้อมูล โดยเปรียบเทียบประสิทธิภาพการทำนายโรคหัวใจของ 5 เทคนิคจำแนกได้แก่ เทคนิคนาอีฟเบย์ เทคนิคต้นไม้ตัดสินใจ เทคนิคดิสครีมิแนนต์ เทคนิคการสุ่มป่าไม้และเทคนิคซัพพอร์ตเวกเตอร์แมทชีน ผลการวิจัยพบว่าเทคนิคต้นไม้ตัดสินใจมีประสิทธิภาพดีที่สุดมีค่าเท่ากับร้อยละ 99.00 จากงานวิจัยที่ผลสอดคล้องทั้ง 3 งานพบว่ามีการนำปัจจัยที่ส่งผลต่อการเกิดโรคมาใช้ในการพยากรณ์ โดยปัจจัยจะมีความสำคัญไม่เท่ากันและเทคนิคต้นไม้ตัดสินใจมีค่าความถูกต้องสูงที่สุด

2) ตัวแปรที่นำมาใช้ในการพยากรณ์สามารถส่งผลต่อค่าความถูกต้องในการพยากรณ์ พบว่าโรคเบาหวาน ให้ค่าความถูกต้อง (0.952) สูงที่สุดหากเทียบกับโรคความดันโลหิตสูง (0.921) และโรคหัวใจขาดเลือด (0.739) เนื่องจากโรคเบาหวานมีตัวแปรสำคัญที่ส่งผลต่อการพยากรณ์ คือ ค่าน้ำตาลในเลือด (FPG) ซึ่งสอดคล้องกับ ศูนย์รวมคำแนะนำและข้อมูลผู้เชี่ยวชาญด้านสุขภาพระดับสากล (2559) และพวงทอง ไกรพิบูลย์ (2562) ระบุว่าผู้ป่วยโรคเบาหวาน จะมีระดับน้ำตาลในเลือดสูง

ข้อเสนอแนะ

จากการนำเสนอผลการพยากรณ์ และการประชุมกลุ่มที่ห้องประชุมพุทธรักษา โรงพยาบาลนาหว้า จังหวัดนครพนม ผู้เชี่ยวชาญเสนอแนะเพิ่มเติมเกี่ยวกับประสิทธิภาพในการทำงานคือ ให้นำตัวแปรไขมันชนิดไม่ดี LDL (Low–Density Lipoprotein) มาใช้ในการพยากรณ์ร่วมด้วยสำหรับแบบจำลองโรคเบาหวาน เนื่องจากผู้ป่วยเบาหวานมักมีปัญหาไขมันในเลือดร่วมด้วย เป็นลักษณะของความผิดปกติของไขมันที่เกี่ยวข้องกับการดื้ออินซูลิน และโรงพยาบาลควรมีการลงข้อมูลให้ครบถ้วนสมบูรณ์ คือ เพศ อายุ น้ำหนัก ส่วนสูง ค่าความดันโลหิต ผลตรวจทางห้องปฏิบัติการ การให้รหัสวินิจฉัย และควรมีการเก็บข้อมูลปัจจัยอื่นที่เกี่ยวข้องเพิ่มเติมในลักษณะเขตข้อมูล (Field) ได้แก่ เก็บข้อมูลการออกกำลังกาย และกรรมพันธุ์ครอบครัว

ผู้วิจัยได้เสนอแนวทางการวิจัยในครั้งต่อไปไว้ดังนี้

1) หากจะพัฒนาหรือต่อยอดงานวิจัยนี้เพื่อให้มีความแม่นยำมากขึ้น ควรเพิ่มปัจจัยที่ส่งผลต่อการเกิดโรคไม่ติดต่อเรื้อรังให้ครอบคลุม ได้แก่ ไขมันชนิดไม่ดี LDL (Low–Density Lipoprotein) ข้อมูลออกกำลังกาย และกรรมพันธุ์ครอบครัว

2) พัฒนาเป็นแอปพลิเคชัน (Application) การเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุในรูปแบบของโรงพยาบาล เพื่อง่ายต่อการใช้งาน พยากรณ์โรคไม่ติดต่อเรื้อรังในผู้สูงอายุเพื่อเฝ้าระวัง แก้ไขปัญหาด้านสุขภาพ และในรูปแบบการให้บริการกับประชาชน ประเมินความเสี่ยงของการเกิดโรคด้วยตนเอง พร้อมให้คำแนะนำเบื้องต้น

บรรณานุกรม

บรรณานุกรม

กระทรวงสาธารณสุข (2557). *ยุทธศาสตร์ ตัวชี้วัด และแนวทางการจัดเก็บข้อมูล กระทรวงสาธารณสุข ปีงบประมาณ พ.ศ. 2558*. นนทบุรี: สำนักนโยบายและยุทธศาสตร์ กระทรวงสาธารณสุข.

กลุ่มยุทธศาสตร์และแผนงาน สำนักโรคไม่ติดต่อ. (2559). *รายงานประจำปี 2558*. กรุงเทพฯ: สำนักงานกิจการโรงพิมพ์องค์การสงเคราะห์ทหารผ่านศึกในพระบรมราชูปถัมภ์.

กิตติศักดิ์ สุมามาลย์. (2555). *การคัดกรองสุขภาพเบื้องต้นโดยใช้เทคนิคเหมืองข้อมูล*. สารนิพนธ์ วท.ม. กรุงเทพฯ: มหาวิทยาลัยธุรกิจบัณฑิตย์.

ชฎิภกรน์ ทรายหมอ. (2557). *การพัฒนาแบบจำลองพยากรณ์ภาวะแทรกซ้อนของโรคอื่นในผู้ป่วยโรคเบาหวาน โดยใช้เทคนิคการทำเหมืองข้อมูล : กรณีโรงพยาบาลปทุมธานี*. วิทยานิพนธ์ วท.ม. นนทบุรี: มหาวิทยาลัยสุโขทัยธรรมาธิราช.

ชนิดา เสือเปีย. (2557). *การศึกษาสาเหตุการเกิดโรคเบาหวานชนิดที่ 2 ด้วยเทคนิคการทำเหมืองข้อมูล: กรณีศึกษาโรงพยาบาลรัฐบาลแห่งหนึ่ง*. วิทยานิพนธ์ วท.ม. กรุงเทพฯ: มหาวิทยาลัยธรรมศาสตร์.

ณัฐดนัย สิงห์คลีวรรณ. (2555). *ระบบสนับสนุนการตัดสินใจ การจัดการเครื่องมือแพทย์ในโรงพยาบาล*. ดุษฎีนิพนธ์ วท.ด. กรุงเทพฯ: มหาวิทยาลัยราชภัฏบ้านสมเด็จเจ้าพระยา.

ทัศนนท์ ฐานประเสริฐกุล และ กมล เกียรติเรืองกมลา. (2558). *การศึกษาสาเหตุการเกิดโรคความดันโลหิตสูง กรณีศึกษาโรงพยาบาลภาครัฐแห่งหนึ่งในประเทศไทย*. วิทยานิพนธ์ ค.อ.บ. กรุงเทพฯ : มหาวิทยาลัยธรรมศาสตร์.

ธีรพงษ์ สังข์ศรี. (2557). *การวิเคราะห์พฤติกรรมสำหรับการเลือกสมัครสาขาวิชาเรียนและการเปรียบเทียบตัวแบบพยากรณ์จำนวนนักศึกษาใหม่โดยใช้เทคนิคเหมืองข้อมูล*. ใน การประชุมทางวิชาการระดับชาติด้านคอมพิวเตอร์และเทคโนโลยีสารสนเทศครั้งที่ 10 (หน้า. 963–968). กรุงเทพฯ: มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ.

นพรัตน์ พจน์จิราภรณ์. (2557). *ระบบสนับสนุนทางการแพทย์เพื่อวิเคราะห์โรคหลอดเลือดแดงโป่งพองโดยใช้เทคนิคเหมืองข้อมูล กรณีศึกษาการรักษาหลอดเลือดแดงโป่งพองผ่านสายสวน.* ปัญหาพิเศษ วท.ม. กรุงเทพฯ : มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ.

นันทกรณ์ หาไชยอินทร์. (2554). *การจำแนกและคาดการณ์กลุ่มอาการผู้ป่วยที่มาโรงพยาบาลครั้งแรกด้วยเทคนิคเหมืองข้อมูล กรณีศึกษาโรงพยาบาลศรีนครินทร์ คณะแพทยศาสตร์ มหาวิทยาลัยขอนแก่น.* วิทยานิพนธ์ วศ.ม. กรุงเทพฯ: มหาวิทยาลัยเกษตรศาสตร์.

บุญชม ศรีสะอาด. (2556). *วิธีการทางสถิติสำหรับการวิจัย เล่ม 1.* (พิมพ์ครั้งที่ 5). กรุงเทพฯ: สุวีริยาสาส์น.

ประกิต วาทีสาธกกิจ. (2555). ยกเครื่องกฎหมายควบคุมยาสูบเพื่อเด็กและสังคมไทยปลอดบุหรี่. *Smart วารสารเพื่อคนรุ่นใหม่ไม่สูบบุหรี่.* 13, 151.

ปราโมทย์ ประสาทกุล และปัทมา ว่าพัฒนวงศ์. (2553). *นิยามผู้สูงอายุด้วยช่วงชีวิตข้างหน้า.* นครปฐม: สำนักพิมพ์ประชากรและสังคม.

ปรีชา อุปโยคิน, สุรีย์ กาญจนวงศ์, วีณา ศิริสุข และมัลลิกา มัติโก. (2541). *ไม้ใกล้ฝั่ง: สถานภาพและบทบาทผู้สูงอายุไทย.* กรุงเทพฯ: เจริญดีการพิมพ์.

ปัณณวิชญ์ วงศ์วิวัฒนานนท์. (2559). *การทำเหมืองข้อมูลกับสารสนเทศทางการแพทย์.* เข้าถึงได้จาก http://dmbj.ejnal.com/e-journal/showdetail/?show_detail=T&art_id=1232. 1 ตุลาคม 2562

ผู้เชี่ยวชาญด้านสุขภาพระดับสากล. (2559).*เบาหวาน (Diabetes mellitus/DM).* เข้าถึงได้จาก https://thailandonlinehospital.com/th/disease/view/31. 20 มีนาคม 2563

พวงทอง ไกรพิบูลย์. (2562). *เบาหวาน (Diabetes mellitus) และ ความดันโลหิตสูง (Hypertension).* เข้าถึงได้จาก http://haamor.com/th/. 30 กรกฎาคม 2563

มณีรัตน์ ภารนันท์. (2554). *ประสิทธิภาพการประมวลผลเหมืองข้อมูลบนระบบปฏิบัติการกลุ่มเมฆแบบส่วนตัว กรณีศึกษา การวินิจฉัยภาวะแทรกซ้อนผู้ป่วยโรคเบาหวาน โรงพยาบาลศิริราช.* วิทยานิพนธ์ วท.ม. กรุงเทพฯ: มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ.

รศรินทร์ เกรย์, อุมาภรณ์ ภัทรวาณิชย์, เฉลิมพล แจ่มจันทร์ และเรวดี สุวรรณนพเก้า. (2556). *มโนทัศน์ใหม่ของนิยามผู้สูงอายุ : มุมมองเชิงจิตวิทยาสังคม และสุขภาพ.* นครปฐม: สถาบันวิจัยประชากรและสังคม มหาวิทยาลัยมหิดล

รุ่งทิวา ชอบชื่น. (2556). Nursing care in pressure sore. *Srinagarind Med J 2013*, 28, 41–46.

รุ่งฤดี บุษดาจันทร์ และทศนัย ชุ่มวัฒนะ. (2560). การทำนายความเสี่ยงในการเกิดภาวะแทรกซ้อนโรคความดันโลหิตสูงของผู้ป่วยโรคเบาหวาน โดยใช้เทคนิคการทำเหมืองข้อมูล. *วารสารสมาคมเวชสารสนเทศไทย*, 3(1), 1–8.

เรวดี ศักดิ์ดุลยธรรม. (2553). *การใช้เทคนิคดาต้าไมน์นิงในการวิเคราะห์ปัจจัยที่มีผลต่อความสำเร็จในการรักษา โรคนิ้วล็อกในแบบต่างๆ ของคณะแพทยศาสตร์วชิรพยาบาล มหาวิทยาลัยกรุงเทพมหานคร.* งานวิจัย. นนทบุรี: มหาวิทยาลัยราชพฤกษ์.

วรชัย ทองไทย. (2549). *อายุเริ่มต้นของผู้สูงอายุไทย.* ใน การประชุมวิชาการประชาคมศาสตร์แห่งชาติ 2549, สมาคมนักประชากรไทย.

วัฒนพัชร์ เก้าลิ้ม. (2557). *การค้นหาคำสำคัญในเอกสารภาษาไทยโดยใช้เทคนิคการค้นหารูปแบบความสัมพันธ์ในเหมืองข้อมูล.* วิทยานิพนธ์ วท.ม. นครปฐม : มหาวิทยาลัยศิลปากร.

วิธวินท์ แสงมณี, วีระวุฒ รัตนเจริญเลิศ, ณัฏฐภพ โพธิรัชต์ และเพียงฤทัย หนูสวัสดิ์, (2560). *การสร้างโมเดลทำนายโอกาสการกลับมารักษาตัวซ้ำของผู้ป่วยโรคเบาหวาน โดยใช้เทคนิคการทำเหมืองข้อมูล.* วิทยานิพนธ์ วท.ม. นครปฐม: มหาวิทยาลัยเทคโนโลยีราชมงคลรัตนโกสินทร์.

ศศิมา มณฑาสุวรรณ. (2557).*การพัฒนาระบบค้นหารหัส ICD–10 สำหรับงานเวชระเบียน.* วิทยานิพนธ์ ศศ.ม. นครปฐม: มหาวิทยาลัยศิลปากร

สมเกียรติ โพธิสัตย์, เนติมา คูนีย์, รัชนีบูลย์ อุดมชัยรัตน์, พรทิพย์ ปรีชาไชยวิทย์, สุรีพร คนละเอียด, ศุภลักษณ์ มิรัตนไพร และเกตุแก้ว สายน้ำเย็น. (2557). *การทบทวนวรรณกรรม : สถานการณ์ปัจจุบันและรูปแบบการบริการด้านโรคไม่ติดต่อเรื้อรัง.* นนทบุรี: สถาบันวิจัยและประเมินเทคโนโลยีทางการแพทย์ กรมการแพทย์ กระทรวงสาธารณสุข.

สมภพ พระธานี. (2562). *เจ็บหน้าอกจากหัวใจขาดเลือด กล้ามเนื้อหัวใจขาดเลือด อาการปวดเค้นหัวใจ (Angina Pectoris)*. เข้าถึงได้จาก http://haamor.com/th/. 21 มีนาคม 2563

สมศักดิ์ ศรีสันติสุข. (2539). *สังคมวิทยาภาวะสูงอายุ : ความเป็นจริงและการคาดการณ์ในสังคมไทย*. กรุงเทพมหานคร : โรงพิมพ์จุฬาลงกรณ์มหาวิทยาลัย.

สหดล ปุญญาถาวร. (2550). จุลสารเครือข่ายวิชาชีพแพทย์ในการบริโภคยาสูบ. 1(1), 10–11.

สายชล สินสมบูรณ์ทอง. (2560). *การทำเหมืองข้อมูล เล่ม 1 : การค้นหาความรู้จากข้อมูล พิมพ์ครั้งที่ 2 ฉบับปรับปรุง*. กรุงเทพฯ: จามจุรีโปรดักส์

สำนักงานคณะกรรมการข้าราชการพลเรือน. (2542). รายงานการวิจัย เรื่อง ระบบการเกษียณอายุตามลักษณะงาน. กรุงเทพฯ: สำนักงานคณะกรรมการข้าราชการ

สุชาดา ทวีสิทธิ์. (2553). *ประชากรและสังคม 2553: คุณค่าผู้สูงอายุในสายตาสังคมไทย*. นครปฐม: สำนักพิมพ์ประชากรและสังคม.

เสกสรรค์ วิลัยลักษณ์. (2558). *การใช้เทคนิคการทำเหมืองข้อมูลเพื่อพยากรณ์ผลการเรียนของนักเรียนโรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน ศูนย์วิจัยและพัฒนาการศึกษา*. วิทยานิพนธ์ วท.ม. นนทบุรี: มหาวิทยาลัยสุโขทัยธรรมาธิราช.

อดุลย์ ยิ้มงาม. (2554). *การทำเหมืองข้อมูล (Data Mining)*. เข้าถึงได้จาก http://compcenter.bu.ac.th/news-information/data-mining. 21 มีนาคม 2563

เอกสิทธิ์ พัชรวงศ์ศักดา. (2557). *การวิเคราะห์ข้อมูลด้วยเทคนิคดาต้า ไมน์นิ่ง เบื้องต้น*. กรุงเทพฯ: เอเชีย ดิจิตอลการพิมพ์ จำกัด.

A. Rahman Khan,M. Rahman,J. Us Salehin,S. Islam and F. Rabbi. (2021). *Efficient Data Mining Techniques for Heart Disease Prediction and Comparative Analysis of Classification Algorithms*. Noakhali: Noakhali Science and Technology University.

Anthony, O.C. (2010). Managing Behavior and Emotional Issues in Older People. *The Social Science*, 5(5), 401–413.

Berry, M.A. and Linoff, G.S. (2000), Mastering Data Mining: The Art and Science of Customer Relationship Management. *Industrial Management & Data Systems*, 100(5), 245–246.

Cabena, P., Hadjinian, P., Stadler, R., Verhees, J. and Zanasi, A. (1998). *Discovering Data Mining : From Concept to Implementation*. Upper Saddle River. New Jersey: Prentice Hall.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinart, T., Shearer, C., and Wirth, R. (2000). *CRISP–DM Step by Step Data Mining Guide*. Retrieved from http://www.crisp–dm.org/. 10 August 2019.

Cowgill, D. O. (1986). *Aging Around the World*. Belmont, Ca.: Wadsworth.

Fayyad, U., Piatetsky–Shapiro, G., & Smyth, P. (1996). *From Data Mining to Knowledge Discovery in Databases*. Retrieved from https://doi.org/10.1609/aimag.v17i3.1230. 20 June 2018.

Han, J.,Kamber, M. & Pei, J. (2011). *Data mining concepts and techniques (3rd ed.)*. United States of America: Morgan Kaufman Publishers.

Hand, D., Mannila, H., and Smyth, P. (2001). *Principles of Data Mining*. Retrieved from https://docplayer.net/5895744–Principles–of–data–mining–by–david–hand–heikki–mannila–and–padhraic–smyth–the–mit–press–2001–546–pages.html. 21 June 2018.

Hinohara, S. (2006). *Living Long, Living Good*. Tokyo: IBC Publishing.

Huff Post. (2012). *The Elderly vs. The Middle Age: Who Is A Senior Citizen, Who Is Middle Aged And Why? (February 5, 2012)*. Retrieved from http://www.huffingtonpost.com/2012/05/02/elderly–senior–citizens–middle–age–aged_n_1471176.html. 11 August 2019.

Israa Ahmed Zriqat, Ahmad Mousa Altamimi and Mohammad Azzeh. (2017). *A Comparative Study for Predicting Heart Diseases Using Data Mining Classification Methods*. Amman: Applied Science Private University

Orimo, H., et, al. (2006). Reviewing the definition of “elderly”. *Japan Geriatrics Society*, 6, 149–158.

R. Thanigaivel and K Ramesh Kumar. (2015). *Data Mining Association Rules for Heart Disease Prediction System*. Tamil Nadu: Hindustan University.

Roiger, R. J., and Geatz, M. (2003). *Data mining: A tutorial–based primer*. Boston: Addison Wesley.

S. Senthil and B. Ayshwarya. (2018). *Lung Cancer Prediction using Feed Forward Back Propagation Neural Networks with Optimal Features*. Karnataka: Reva University.

Shubpreet Kaur and R. K. Bawa. (2015). *Future Trends of Data Mining in Predicting the Various Diseasesin Medical Healthcare System*. Patiala: Punjabi University.

Uotinen, V. (2005). *I am Old and I feel: Subjective Age in Finish Adults*. Jyväskylä: Jyväskylä University Printing House.

Witten, I.H. and Frank, E. (2005) *Data mining: Practical machine learning tools and techniques.* 2nd Edition. Burlington: Morgan Kaufmann Publisher.

World Health Organization (WHO). (n.d). *Definition of an older or elderly person: Proposed Working Definition of an Older Person in Africa for the MDS Project*. Retrieved from http://www.who.int/healthinfo/survey/ageingdefnolder/en/index.html. 20 August 2018.

ภาคผนวก

ตารางข้อมูลของตารางผู้ป่วยโรงพยาบาลนาหว้า

ลำดับ	ชื่อฟิลด์	ชนิด	ความยาว	คีย์หลัก
1	hos_guid	varchar	38	PK
2	hn	varchar	9	–
3	pname	varchar	15	–
4	fname	varchar	30	–
5	lname	varchar	30	–
6	occupation	varchar	4	–
7	citizenship	char	3	–
8	birthday	date	0	–
9	addrpart	varchar	50	–
10	moopart	char	3	–
11	tmbpart	char	2	–
12	amppart	char	2	–
13	chwpart	char	2	–
14	bloodgrp	varchar	20	–
15	clinic	varchar	100	–
16	deathday	date	0	–
17	drugallergy	varchar	250	–
18	familyno	int	11	–
19	fathername	varchar	50	–
20	firstday	date	0	–
21	hometel	varchar	20	–
22	informaddr	varchar	200	–
23	informname	varchar	50	–
24	informrelation	varchar	50	–
25	informtel	varchar	20	–
26	marrystatus	char	1	–

ตารางข้อมูลของตารางผู้ป่วยโรงพยาบาลนาหว้า (ต่อ)

ลำดับ	ชื่อฟิลด์	ชนิด	ความยาว	คีย์หลัก
27	mathername	varchar	50	–
28	hn_int	int	11	–
29	nationality	char	3	–
30	opdlocation	varchar	50	–
31	pttype	char	2	–
32	religion	char	2	–
33	sex	char	1	–
34	spsname	varchar	50	–
35	truebirthday	char	1	–
36	workaddr	varchar	50	–
37	worktel	varchar	20	–
38	hcode	varchar	5	–
39	cid	varchar	13	–
40	hid	int	11	–
41	educate	char	1	–
42	family_status	char	1	–
43	labor_type	char	1	–
44	last_update	datetime	0	–
45	type_area	char	1	–
46	road	varchar	50	–
47	father_cid	varchar	13	–
48	mother_cid	varchar	13	–
49	couple_cid	varchar	13	–
50	person_type	char	2	–
51	private_doctor_name	varchar	75	–
52	legal_action	char	1	–

ตารางข้อมูลของตารางผู้ป่วยโรงพยาบาลนาหว้า (ต่อ)

ลำดับ	ชื่อฟิลด์	ชนิด	ความยาว	คีย์หลัก
53	death_code504	char	2	–
54	death_diag	varchar	6	–
55	node_id	char	1	–
56	admit	char	1	–
57	midname	varchar	25	–
58	po_code	varchar	5	–
59	fatherlname	varchar	30	–
60	motherlname	varchar	30	–
61	spslname	varchar	30	–
62	country	char	2	–
63	email	varchar	50	–
64	birthtime	time	0	–
65	mother_hn	varchar	9	–
66	last_visit	date	0	–
67	death	char	1	–
68	height	int	11	–
69	inregion	char	1	–
70	reg_time	time	0	–
71	oldcode	varchar	50	–
72	lang	char	2	–
73	gov_chronic_id	varchar	10	–
74	in_cups	char	1	–
75	patient_type_id	tinyint	4	–
76	addr_soi	varchar	100	–
77	work_addr	varchar	230	–
78	father_hn	varchar	9	–
79	alias_name	varchar	100	–

ตารางข้อมูลของตารางผู้ป่วยโรงพยาบาลนาหว้า (ต่อ)

ลำดับ	ชื่อฟิลด์	ชนิด	ความยาว	คีย์หลัก
80	destroyed	char	1	–
81	old_addr	varchar	250	–
82	fname_soundex	varchar	50	–
83	lname_soundex	varchar	50	–
84	bloodgroup_rh	varchar	5	–
85	passport_no	varchar	25	–
86	addressid	varchar	6	–
87	mobile_phone_number	varchar	20	–
88	anonymous_person	char	1	–
89	ec_fname	varchar	50	–
90	ec_lname	varchar	50	–
91	hospital_department_id	int	11	–
92	membercard_no	varchar	15	–
93	ec_relation_type_id	int	11	–
94	patient_color_id	int	11	–
95	number_of_relatives	int	11	–
96	birth_order	int	11	–
97	person_labor_type_id	int	11	–
98	is_card_destroy	char	1	–
99	card_destroy_date	date	0	–

ตารางข้อมูลของตารางการรับบริการผู้ป่วยนอกโรงพยาบาลนาหว้า

ลำดับ	ชื่อฟิลด์	ชนิด	ความยาว	คีย์หลัก
1	vn	varchar	13	PK
2	hn	varchar	9	–
3	pdx	varchar	6	–
4	gr504	smallint	6	–
5	lastvisit	smallint	6	–
6	accident_code	varchar	6	–
7	dx_doctor	varchar	7	–
8	dx0	varchar	6	–
9	dx1	varchar	6	–
10	dx2	varchar	6	–
11	dx3	varchar	6	–
12	dx4	varchar	6	–
13	dx5	varchar	6	–
14	sex	char	1	–
15	age_y	smallint	6	–
16	age_m	smallint	6	–
17	age_d	smallint	6	–
18	aid	varchar	6	–
19	moopart	varchar	5	–
20	count_in_month	smallint	6	–
21	count_in_year	smallint	6	–
22	pttype	char	2	–
23	income	double	15	–
24	paid_money	double	15	–
25	remain_money	double	15	–
26	uc_money	double	15	–

ตารางข้อมูลของตารางการรับบริการผู้ป่วยนอกโรงพยาบาลนาหว้า (ต่อ)

ลำดับ	ชื่อฟิลด์	ชนิด	ความยาว	คีย์หลัก
27	item_money	double	15	–
28	dba	tinyint	4	–
29	spclty	char	2	–
30	vstdate	date	0	–
31	op0	varchar	6	–
32	op1	varchar	6	–
33	op2	varchar	6	–
34	op3	varchar	6	–
35	op4	varchar	6	–
36	op5	varchar	6	–
37	rcp_no	varchar	10	–
38	print_count	tinyint	4	–
39	print_done	char	1	–
40	pttype_in_region	char	1	–
41	pttype_in_chwpart	char	1	–
42	pcode	char	2	–
43	hcode	varchar	5	–
44	inc01	double	15	–
45	inc02	double	15	–
46	inc03	double	15	–
47	inc04	double	15	–
48	inc05	double	15	–
49	inc06	double	15	–
50	inc07	double	15	–
51	inc08	double	15	–
52	inc09	double	15	–

ตารางข้อมูลของตารางการรับบริการผู้ป่วยนอกโรงพยาบาลนาหว้า (ต่อ)

ลำดับ	ชื่อฟิลด์	ชนิด	ความยาว	คีย์หลัก
53	inc10	double	15	–
54	inc11	double	15	–
55	inc12	double	15	–
56	inc13	double	15	–
57	inc14	double	15	–
58	inc15	double	15	–
59	inc16	double	15	–
60	hospmain	varchar	5	–
61	hospsub	varchar	5	–
62	pttypeno	varchar	20	–
63	pttype_expire	date	0	–
64	cid	varchar	13	–
65	main_pdx	char	3	–
66	inc17	double	15	–
67	inc_drug	double	15	–
68	inc_nondrug	double	15	–
69	pt_subtype	tinyint	4	–
70	rcpno_list	varchar	25	–
71	ym	varchar	7	–
72	node_id	char	1	–
73	ill_visit	char	1	–
74	count_in_day	tinyint	4	–
75	pttype_begin	date	0	–
76	lastvisit_hour	int	11	–
77	rcpt_money	double	15	–

ตารางข้อมูลของตารางการรับบริการผู้ป่วยนอกโรงพยาบาลนาหว้า (ต่อ)

ลำดับ	ชื่อฟิลด์	ชนิด	ความยาว	คีย์หลัก
78	discount_money	double	15	–
79	old_diagnosis	char	1	–
80	debt_id_list	varchar	50	–
81	vn_guid	varchar	38	–
82	lastvisit_vn	varchar	13	–
83	hos_guid	varchar	38	–
84	rx_license_no	varchar	15	–

ตารางข้อมูลของตารางการซักประวัติโรงพยาบาลนาหว้า

ลำดับ	ชื่อฟิลด์	ชนิด	ความยาว	คีย์หลัก
1	hos_guid	varchar	38	PK
2	vn	varchar	13	–
3	hn	varchar	9	–
4	vstdate	date	0	–
5	vsttime	time	0	–
6	begintime	time	0	–
7	outtime	time	0	–
8	endtime	time	0	–
9	bpd	double	15	–
10	bps	double	15	–
11	bw	double	15	–
12	cc	text	0	–
13	hr	double	15	–
14	pe	text	0	–
15	pulse	double	15	–

ตารางข้อมูลของตารางการซักประวัติโรงพยาบาลนาหว้า (ต่อ)

ลำดับ	ชื่อฟิลด์	ชนิด	ความยาว	คีย์หลัก
16	temperature	double	15	–
17	note	text	0	–
18	rr	double	15	–
19	cc_begin_date	date	0	–
20	cc_cause_of_visit	varchar	250	–
21	cc_sign	varchar	250	–
22	cc_duration	varchar	150	–
23	cc_position	varchar	250	–
24	cc_note	text	0	–
25	his_begin_date	date	0	–
26	his_frequency	varchar	250	–
27	his_severity	varchar	250	–
28	his_cause	varchar	250	–
29	his_expand	varchar	250	–
30	his_cause_increase	varchar	250	–
31	his_cause_decrease	varchar	250	–
32	his_related_sign	varchar	250	–
33	height	int	11	–
34	screen_dep	char	3	–
35	waiting	char	1	–
36	fbs	double	15	–
37	help1	char	1	–
38	help2	char	1	–
39	help3	char	1	–
40	help4	char	1	–
41	help1_time	time	0	–

ตารางข้อมูลของตารางการซักประวัติโรงพยาบาลนาหว้า (ต่อ)

ลำดับ	ชื่อฟิลด์	ชนิด	ความยาว	คีย์หลัก
42	help1_bps	int	11	–
43	help1_bpd	int	11	–
44	help2_time	time	0	–
45	help2_temp	double	15	–
46	help3_icode	varchar	100	–
47	help3_time	time	0	–
48	help3_qty	tinyint	4	–
49	help4_note	text	0	–
50	advice1	char	1	–
51	advice2	char	1	–
52	advice3	char	1	–
53	advice4	char	1	–
54	advice5	char	1	–
55	advice6	char	1	–
56	advice7	char	1	–
57	cradle	char	1	–
58	pe_ga	char	1	–
59	pe_heent	char	1	–
60	pe_heart	char	1	–
61	pe_lung	char	1	–
62	pe_ab	char	1	–
63	pe_ext	char	1	–
64	pe_neuro	char	1	–
65	pe_ga_text	varchar	250	–
66	pe_heent_text	varchar	250	–
67	pe_heart_text	varchar	250	–
68	pe_lung_text	varchar	250	–

ตารางข้อมูลของตารางการซักประวัติโรงพยาบาลนาหว้า (ต่อ)

ลำดับ	ชื่อฟิลด์	ชนิด	ความยาว	คีย์หลัก
69	pe_ab_text	varchar	250	–
70	pe_neuro_text	varchar	250	–
71	pe_ext_text	varchar	250	–
72	bmi	double	15	–
73	tg	double	15	–
74	hdl	double	15	–
75	glucurine	char	1	–
76	blank1	char	1	–
77	bun	double	15	–
78	creatinine	double	15	–
79	ua	double	15	–
80	hba1c	double	15	–
81	riskdm	char	1	–
82	skin_color	varchar	20	–
83	found_amphetamine	char	1	–
84	pregnancy	char	1	–
85	advice7_note	varchar	250	–
86	checkup	char	1	–
87	er_note	text	0	–
88	found_allergy	char	1	–
89	hpi	text	0	–
90	pmh	text	0	–
91	fh	text	0	–
92	sh	text	0	–
93	ros	text	0	–
94	tc	double	15	–

ตารางข้อมูลของตารางการซักประวัติโรงพยาบาลนาหว้า (ต่อ)

ลำดับ	ชื่อฟิลด์	ชนิด	ความยาว	คีย์หลัก
95	ldl	double	15	–
96	ast	double	15	–
97	alt	double	15	–
98	symptom	varchar	200	–
99	walk_id	int	11	–
100	peak_flow	int	11	–
101	cholesterol	double	15	–
102	waist	double	15	–
103	advice8	char	1	–
104	breast_feeding	char	1	–
105	cradle_lie	char	1	–
106	pain_score	int	11	–
107	pefr	int	11	–
108	opdscreen_patient_type_id	int	11	–
109	creatinine_kidney_percent	double	15	–
110	sodium	double	15	–
111	chloride	double	15	–
112	potassium	double	15	–
113	tco2	double	15	–
114	smoking_type_id	int	11	–
115	drinking_type_id	int	11	–
116	pulse_regulation_type_id	int	11	–
117	spo2	double	15	–
118	urine_albumin	double	15	–
119	urine_creatinine	double	15	–
120	pefr_percent	double	15	–
121	macro_albumin	int	11	–

ตารางข้อมูลของตารางการซักประวัติโรงพยาบาลนาหว้า (ต่อ)

ลำดับ	ชื่อฟิลด์	ชนิด	ความยาว	คีย์หลัก
122	micro_albumin	int	11	–
123	egfr	double	15	–
124	hb	double	15	–
125	upcr	double	15	–
126	bicarb	double	15	–
127	phosphate	double	15	–
128	pth	double	15	–
129	pe_gy	char	1	–
130	pe_gy_text	varchar	250	–
131	pe_gu	char	1	–
132	pe_gu_text	varchar	250	–
133	pe_gi	char	1	–
134	pe_gi_text	varchar	250	–
135	bsa	double	15	–
136	pe_head	char	1	–
137	pe_head_text	varchar	250	–
138	pe_skin	char	1	–
139	pe_skin_text	varchar	250	–
140	g6pd	char	1	–
141	pe_rtf	longtext	0	–
142	o2sat	int	11	–
143	pe_pv	char	1	–
144	pe_pv_text	varchar	250	–
145	pe_pr	char	1	–
146	pe_pr_text	varchar	250	–

ตารางข้อมูลของตารางการซักประวัติโรงพยาบาลนาหว้า (ต่อ)

ลำดับ	ชื่อฟิลด์	ชนิด	ความยาว	คีย์หลัก
147	pe_gen	char	1	–
148	pe_gen_text	varchar	250	–
149	pre_pain_score	int	11	–
150	post_pain_score	int	11	–
151	head_cricumference	double	15	–
152	fev1_percent	double	15	–
153	pe_rtf_blob	longblob	0	–
154	bp_stable	char	1	–
155	pe_chest	char	1	–
156	pe_chest_text	varchar	150	–
157	lmp_date	date	0	–
158	opdscreen_bp_loc_type_id	int	11	–
159	menstrual_cycle_type_id	int	11	–
160	adherence_percent	double	15	–
161	fev1_fevc	double	15	–
162	vaccine_screen_type_id	int	11	–
163	development_screen_type_id	int	11	–
164	ambu	char	1	–

ตารางข้อมูลของตารางการสั่งตรวจทางห้องปฏิบัติการโรงพยาบาลนาหว้า

ลำดับ	ชื่อฟิลด์	ชนิด	ความยาว	คีย์หลัก
1	lab_order_number	int	11	PK
2	lab_items_code	int	11	PK
3	lab_order_result	varchar	250	–
4	lab_order_remark	varchar	250	–
5	staff	varchar	50	–
6	confirm	char	1	–
7	lab_items_name_ref	varchar	250	–
8	lab_items_normal_value_ref	varchar	150	–
9	specimen_code	int	11	–
10	lab_items_sub_group_code	int	11	–
11	order_type	char	1	–
12	item_cost	double	15	–
13	hos_guid	varchar	38	–
14	staff_lock_result	char	1	–
15	laborder_date	date	0	–
16	abnormal_result	char	1	–
17	hos_guid_ext	varchar	64	–
18	check_key	varchar	50	–
19	lab_result_status	int	11	–
20	update_datetime	datetime	0	–
21	lab_hist_compare_type_id	int	11	–
22	printed	char	1	–
23	printed_datetime	datetime	0	–
24	critical_result	char	1	–
25	check_key_a	varchar	50	–

ตารางข้อมูลของตารางการให้รหัสการวินิจฉัยผู้ป่วยนอกโรงพยาบาลนาหว้า

ลำดับ	ชื่อฟิลด์	ชนิด	ความยาว	คีย์หลัก
1	ovst_diag_id	int	11	PK
2	vn	varchar	13	–
3	icd10	varchar	9	–
4	hn	varchar	9	–
5	vstdate	date	0	–
6	vsttime	time	0	–
7	diagtype	char	2	–
8	icd103	char	3	–
9	hcode	varchar	5	–
10	doctor	varchar	6	–
11	episode	tinyint	4	–
12	ext_code	char	2	–
13	hos_guid	varchar	38	–
14	dep_flag	char	2	–
15	ovst_oper_type	int	11	–
16	staff	varchar	15	–
17	dx_guid	varchar	38	–
18	lock_dx	char	1	–
19	dx_code_note	varchar	100	–
20	ovstdiag_severe_type_id	int	11	–
21	diag_no	int	11	–
22	update_datetime	datetime	0	–
23	confirm	char	1	–
24	confirm_staff	varchar	25	–
25	opi_guid	varchar	38	–

บันทึกข้อความ

ส่วนราชการ โรงพยาบาลนาหว้า อ.นาหว้า จ.นครพนม ๔๘๑๘๐ โทรศัพท์ ๐ ๔๒๕๙ ๗๐๒๑-๒ ต่อ ๑๑๗

ที่ ๐๐๓๒.๓๐๑.๖/๑๕๓ **วันที่** ๕ ตุลาคม ๒๕๖๓

เรื่อง ขออนุญาตข้อมูลทำวิจัย

เรียน ผู้อำนวยการโรงพยาบาลนาหว้า

เนื่องด้วยกระผมนายศรัณยู ชูทองรัตนะ ตำแหน่งนักวิชาการคอมพิวเตอร์ ปฏิบัติงานที่กลุ่มงานประกันสุขภาพ ยุทธศาสตร์และสารสนเทศทางการแพทย์ โรงพยาบาลนาหว้า ซึ่งปัจจุบันอยู่ระหว่างดำเนินการวิจัย เรื่อง แบบจำลองเพื่อพยากรณ์การเกิดโรคเรื้อรังในผู้สูงอายุด้วยเทคนิคเหมืองข้อมูล กรณีศึกษา โรงพยาบาลนาหว้า (Models Forecasting Of Chronic Disease In The Elderly By Using Data Mining Techniques Case Study Nawa Hospital) โดยมีวัตถุประสงค์เพื่อสร้างแบบจำลองเพื่อพยากรณ์การเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุ ตลอดจนเพื่อศึกษาข้อมูลและเทคนิคเหมืองข้อมูลที่เหมาะสมสำหรับการสร้างแบบจำลองเพื่อพยากรณ์การเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุ โดยผู้วิจัยจะขอความอนุเคราะห์ในการเข้าถึงสารสนเทศของโรงพยาบาลนาหว้า ซึ่งจะมีรายละเอียดข้อมูลตามเอกสารที่แนบมาด้วย

จึงเรียนมาเพื่อพิจารณาอนุญาต

นายศรัณยู ชูทองรัตนะ
นักวิชาการคอมพิวเตอร์

อนุญาต

นายกิตติเชษฐ์ ธีรกุลพงศ์เวช
นายแพทย์ชำนาญการ รักษาการในตำแหน่ง
ผู้อำนวยการโรงพยาบาลนาหว้า

โรงพยาบาลนาหว้า จังหวัดนครพนม

บันทึกข้อความ

ส่วนราชการ สถาบันวิจัยและพัฒนา สำนักงานคณะกรรมการจริยธรรมการวิจัยในมนุษย์ โทร. IP Phone ๒๖๓

ที่ อว. ๐๖๒๑.๐๘/๘๗๘ **วันที่** ๑๒ พฤศจิกายน ๒๕๖๓

เรื่อง การแจ้งผลการพิจารณารับรองจริยธรรมการวิจัยในมนุษย์

เรียน นายศรัณยู ชูทองรัตนะ

ตามที่ ท่านได้ยื่นเสนอโครงการวิจัยเพื่อขอรับการยกเว้นการพิจารณาจริยธรรมการวิจัยในมนุษย์ เรื่อง (ภาษาไทย) แบบจำลองเพื่อพยากรณ์การเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุ ด้วยเทคนิคเหมืองข้อมูล กรณีศึกษาโรงพยาบาลนาหว้า (ภาษาอังกฤษ) Models forecasting of chronic disease in the elderly by using data mining techniques case study Nawa Hospital หมายเลขสำคัญโครงการ HE ๖๓ - ๐๕๓ นั้น

คณะกรรมการจริยธรรมการวิจัยในมนุษย์ มหาวิทยาลัยราชภัฏสกลนคร ได้พิจารณาแล้วเห็นว่าเป็นโครงการวิจัยที่เข้าข่ายการรับรองด้านจริยธรรมการวิจัยในมนุษย์แบบยกเว้น เนื่องจากเป็นโครงการวิจัยประเมินระบบงาน โครงการวิจัยบริหารจัดการ และโครงการวิจัยสถาบัน ซึ่งเป็นโครงการวิจัยที่ได้รับอนุญาตจากหัวหน้าผู้รับผิดชอบหน่วยงานหรือองค์กร และมีจุดประสงค์เพื่อพัฒนาระบบหรือนโยบาย โดยการศึกษา ทดสอบ หรือ ประเมินระบบการให้บริการ ประโยชน์ต่อสาธารณชน วิธีการรับผลประโยชน์หรือบริการ รวมทั้งความเป็นไปได้หรือทางเลือกของโครงการวิจัย และเป็นโครงการวิจัยที่ใช้ข้อมูลทุติยภูมิ ซึ่งจะต้องได้รับอนุญาตจากหัวหน้าผู้รับผิดชอบหน่วยงานหรือองค์กร หรือองค์กรที่เป็นเจ้าของข้อมูล

จึงเรียนมาเพื่อโปรดทราบ

สุธาสินี คุปตะบุตร

(นางสาวสุธาสินี คุปตะบุตร)
ประธานกรรมการจริยธรรมการวิจัยในมนุษย์
มหาวิทยาลัยราชภัฏสกลนคร

เลขที่ใบรับรอง ๐๔๗/๒๕๖๓

คณะกรรมการจริยธรรมการวิจัยในมนุษย์ มหาวิทยาลัยราชภัฏสกลนคร

ขอรับรองว่าโครงการวิจัยเรื่อง : แบบจำลองเพื่อพยากรณ์การเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุ ด้วยเทคนิคเหมืองข้อมูล กรณีศึกษาโรงพยาบาลนาหว้า (Models forecasting of chronic disease in the elderly by using data mining techniques case study Nawa Hospital)

หัวหน้าโครงการวิจัย : นายศรัณยู ชูทองรัตนะ

หน่วยงานต้นสังกัด : สาขาวิชาวิทยาการสารสนเทศและเทคโนโลยี

เอกสารที่รับรอง

๑. แบบฟอร์มขอเสนอโครงการวิจัย
๒. แบบฟอร์มโครงการวิจัย
๓. เครื่องมือวิจัย
๔. เอกสารคำชี้แจงสำหรับอาสาสมัคร

ได้ผ่านการรับรองจริยธรรมการวิจัยในมนุษย์แบบยกเว้น จากคณะกรรมการจริยธรรมการวิจัยในมนุษย์ มหาวิทยาลัยราชภัฏสกลนคร โดยยึดหลักพื้นฐานของหลักจริยธรรมการวิจัยสากล

รับรอง ณ วันที่ ๑๒ เดือน พฤศจิกายน พ.ศ. ๒๕๖๓

สุธาสินี คุปตะบุตร

(นางสาวสุธาสินี คุปตะบุตร)
ประธานกรรมการจริยธรรมการวิจัยในมนุษย์
มหาวิทยาลัยราชภัฏสกลนคร

วันหมดอายุ วันที่ ๑๒ เดือน พฤศจิกายน พ.ศ. ๒๕๖๔

สำนักงานคณะกรรมการจริยธรรมการวิจัยในมนุษย์ มหาวิทยาลัยราชภัฏสกลนคร ที่อยู่ สถาบันวิจัยและพัฒนา มหาวิทยาลัยราชภัฏสกลนคร เลขที่ ๖๘๐ ถ.นิตโย ต.ธาตุเชิงชุม อ.เมืองสกลนคร จ.สกลนคร รหัสไปรษณีย์ ๔๗๐๐๐ โทรศัพท์ ๐๔๒-๗๗๐-๑๕๔

เอกสารชี้แจงอาสาสมัครที่เข้าร่วมโครงการวิจัย (สำหรับตอบแบบสอบถาม)

เรียน ผู้ตอบแบบสอบถามทุกท่าน

เนื่องด้วยข้าพเจ้า นายศรัณยู ชูทองรัตนะ นักศึกษาปริญญาโท สาขาวิชาวิทยาการสารสนเทศและเทคโนโลยี คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏสกลนคร กำลังดำเนินการวิจัย เรื่อง “แบบจำลองเพื่อพยากรณ์การเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุ ด้วยเทคนิคเหมืองข้อมูล กรณีศึกษาโรงพยาบาลนาหว้า” โดยมีวัตถุประสงค์ของการวิจัย 1) เพื่อสร้างแบบจำลองเพื่อพยากรณ์การเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุ 2) เพื่อเปรียบเทียบประสิทธิภาพของแบบจำลองการเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุ ท่านอาจจะไม่ได้รับประโยชน์โดยตรงจากการเข้าร่วมโครงการวิจัยนี้ แต่ข้อมูลที่ได้จะมีประโยชน์ต่อการพยากรณ์การเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุ ซึ่งสามารถประยุกต์ใช้เพื่อคัดกรองโรคไม่ติดต่อเรื้อรังในผู้สูงอายุของหน่วยงานทางสาธารณสุขได้

หากท่านตัดสินใจเข้าร่วมการวิจัยแล้ว ผู้วิจัยจะขอให้ท่านตอบแบบสอบถามในประเด็นระดับความพึงพอใจ /การนำไปใช้ประโยชน์ของแบบจำลอง ซึ่งประกอบด้วยคำถาม มีทั้งหมด 3 ส่วน จำนวน 10 ข้อ) โดยใช้เวลาในการตอบแบบสอบถามประมาณ 15 นาที และจะขอรับแบบสอบถามคืนโดยผู้วิจัยจะไปเก็บด้วยตนเอง

หากท่านรู้สึกอึดอัด หรือรู้สึกไม่สบายใจกับบางคำถาม ท่านมีสิทธิ์ที่จะไม่ตอบคำถามเหล่านั้นได้ รวมถึงท่านมีสิทธิ์ถอนตัวออกจากโครงการนี้เมื่อใดก็ได้ โดยไม่ต้องแจ้งให้ทราบล่วงหน้า และการไม่เข้าร่วมวิจัยหรือถอนตัวออกจากโครงการวิจัยนี้ จะไม่มีผลกระทบใดๆ ต่อการปฏิบัติงานของท่านแต่ประการใด ข้อมูลในการตอบแบบสอบถามของท่านจะถูกเก็บรักษาไว้ ไม่เปิดเผยต่อสาธารณะเป็นรายบุคคล แต่จะรายงานผลการวิจัยในภาพรวมเท่านั้น และจะดำเนินการทำลายข้อมูลที่เกี่ยวข้องภายหลังเสร็จสิ้นการวิจัย

การวิจัยครั้งนี้ท่านจะไม่ได้รับค่าตอบแทนและไม่เสียค่าใช้จ่ายใดๆ ทั้งสิ้น

หากท่านมีข้อสงสัยเกี่ยวกับงานวิจัย โปรดติดต่อได้ที่ นายศรัณยู ชูทองรัตนะ คณะวิทยาศาสตร์ สาขาวิชาวิทยาการสารสนเทศและเทคโนโลยี โทรศัพท์ 09-1061-6308

หากท่านได้รับการปฏิบัติไม่ตรงตามที่ระบุไว้หรือต้องการทราบสิทธิของท่านขณะเข้าร่วมการวิจัยนี้ สามารถติดต่อได้ที่ “สำนักงานคณะกรรมการจริยธรรมการวิจัยในมนุษย์ มหาวิทยาลัยราชภัฏสกลนคร สถาบันวิจัยและพัฒนา ชั้น 4 อาคาร 10 มหาวิทยาลัยราชภัฏสกลนคร” โทร. 0-4297-0154 ,09-7179-6445 เบอร์ภายใน 261

ขอขอบพระคุณอย่างสูง

.....

(นายศรัณยู ชูทองรัตนะ)

ที่ อว ๐๖๒๑.๑๒/ว ๒๙๓

มหาวิทยาลัยราชภัฏสกลนคร
๖๘๐ ถ.นิตโย ต.ธาตุเชิงชุม
อ.เมือง จ.สกลนคร ๔๗๐๐๐

๘ มีนาคม ๒๕๖๔

เรื่อง ขอความอนุเคราะห์เป็นผู้เชี่ยวชาญเสนอแนวทางพัฒนา

เรียน นายแพทย์กิตติเชษฐ์ ธีรกุลพงศ์เวช

สิ่งที่ส่งมาด้วย เครื่องมือการวิจัย

ด้วย นายศรัณยู ชูทองรัตนะ รหัสประจำตัวนักศึกษา ๕๙๔๒๕๑๑๗๑๐๒ ซึ่งเป็นนักศึกษาปริญญาโท สาขาวิชาวิทยาการสารสนเทศและเทคโนโลยี มหาวิทยาลัยราชภัฏสกลนคร กำลังศึกษาและทำวิทยานิพนธ์เรื่อง **"แบบจำลองเพื่อพยากรณ์การเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุด้วยเทคนิคเหมืองข้อมูล กรณีศึกษาโรงพยาบาลนาหว้า"** ซึ่งเป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรมหาบัณฑิต (วท.ม.) โดยมี ดร.สุทิศา ชองเหล็กนอก เป็นประธานที่ปรึกษาวิทยานิพนธ์ และ ดร.สุธาสินี คุปตะบุตร เป็นกรรมการที่ปรึกษาวิทยานิพนธ์

มหาวิทยาลัยราชภัฏสกลนคร ได้พิจารณาแล้วเห็นว่าท่านเป็นผู้มีความรู้ความสามารถและเชี่ยวชาญในเรื่องดังกล่าว จึงขอความอนุเคราะห์ท่านเป็นผู้เชี่ยวชาญเสนอแนวทางพัฒนาแบบจำลองเพื่อพยากรณ์การเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุด้วยเทคนิคเหมืองข้อมูล กรณีศึกษาโรงพยาบาลนาหว้า **ในวันที่ ๙ มีนาคม พ.ศ. ๒๕๖๔ เวลา ๑๓.๓๐ น. ณ ห้องประชุมพุทธรักษา โรงพยาบาลนาหว้า** โดยเครื่องมือที่ใช้ในการวิจัยครั้งนี้ ได้ผ่านการรับรองจริยธรรมการวิจัยในมนุษย์แบบยกเว้น จากคณะกรรมการจริยธรรมการวิจัยในมนุษย์ มหาวิทยาลัยราชภัฏสกลนคร เลขที่ใบรับรอง ๐๔๗/๒๕๖๓ ลงวันที่ ๑๒ พฤศจิกายน พ.ศ. ๒๕๖๓

จึงเรียนมาเพื่อโปรดพิจารณา หวังเป็นอย่างยิ่งว่าคงได้รับความอนุเคราะห์จากท่านด้วยดี และขอขอบคุณมา ณ โอกาสนี้

ขอแสดงความนับถือ

(รองศาสตราจารย์ ดร.ศิกานต์ เพียรธัญญกรณ์)
ผู้อำนวยการบัณฑิตวิทยาลัย ปฏิบัติราชการแทน
อธิการบดีมหาวิทยาลัยราชภัฏสกลนคร

บัณฑิตวิทยาลัย
โทร. ๐ ๔๓๙๗ ๐๒๒๙
โทรสาร ๐ ๔๒๙๗ ๐๐๓๒
ผู้ประสานงาน นายศรันยู ชูทองรัตนะ โทรศัพท์เคลื่อนที่ ๐๙ ๑๐๖๑ ๖๓๐๘

"อยู่สกล รักสกล ทำเพื่อสกลนคร"

ที่ อว ๐๖๒๑.๑๒/ว ๒๙๓

มหาวิทยาลัยราชภัฏสกลนคร
๖๘๐ ถ.นิตโย ต.ธาตุเชิงชุม
อ.เมือง จ.สกลนคร ๔๗๐๐๐

๘ มีนาคม ๒๕๖๔

เรื่อง ขอความอนุเคราะห์เป็นผู้เชี่ยวชาญเสนอแนวทางพัฒนา

เรียน นางสาวกรรณิการ์ อุดมกัน

สิ่งที่ส่งมาด้วย เครื่องมือการวิจัย

ด้วย นายศรัณยู ชูทองรัตนะ รหัสประจำตัวนักศึกษา ๕๙๔๒๕๑๐๗๑๐๒ ซึ่งเป็นนักศึกษาปริญญาโท สาขาวิชาวิทยาการสารสนเทศและเทคโนโลยี มหาวิทยาลัยราชภัฏสกลนคร กำลังศึกษาและทำวิทยานิพนธ์เรื่อง **"แบบจำลองเพื่อพยากรณ์การเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุด้วยเทคนิคเหมืองข้อมูล กรณีศึกษาโรงพยาบาลนาหว้า"** ซึ่งเป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรมหาบัณฑิต (วท.ม.) โดยมี ดร.สุทิศา ซองเหล็กนอก เป็นประธานที่ปรึกษาวิทยานิพนธ์ และ ดร.สุธาสินี คุปตะบุตร เป็นกรรมการที่ปรึกษาวิทยานิพนธ์

มหาวิทยาลัยราชภัฏสกลนคร ได้พิจารณาแล้วเห็นว่าท่านเป็นผู้มีความรู้ความสามารถและเชี่ยวชาญในเรื่องดังกล่าว จึงขอความอนุเคราะห์ท่านเป็นผู้เชี่ยวชาญเสนอแนวทางพัฒนาแบบจำลองเพื่อพยากรณ์การเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุด้วยเทคนิคเหมืองข้อมูล กรณีศึกษาโรงพยาบาลนาหว้า **ในวันที่ ๙ มีนาคม พ.ศ. ๒๕๖๔ เวลา ๑๓.๓๐ น. ณ ห้องประชุมพุทธรักษา โรงพยาบาลนาหว้า** โดยเครื่องมือที่ใช้ในการวิจัยครั้งนี้ ได้ผ่านการรับรองจริยธรรมการวิจัยในมนุษย์แบบยกเว้น จากคณะกรรมการจริยธรรมการวิจัยในมนุษย์ มหาวิทยาลัยราชภัฏสกลนคร เลขที่ใบรับรอง ๐๔๗/๒๕๖๓ ลงวันที่ ๑๒ พฤศจิกายน พ.ศ. ๒๕๖๓

จึงเรียนมาเพื่อโปรดพิจารณา หวังเป็นอย่างยิ่งว่าคงได้รับความอนุเคราะห์จากท่านด้วยดี และขอขอบคุณมา ณ โอกาสนี้

ขอแสดงความนับถือ

(รองศาสตราจารย์ ดร.ศิกานต์ เพียรธัญญกรณ์)
ผู้อำนวยการบัณฑิตวิทยาลัย ปฏิบัติราชการแทน
อธิการบดีมหาวิทยาลัยราชภัฏสกลนคร

บัณฑิตวิทยาลัย
โทร. ๐ ๔๓๙๗ ๐๒๒๙
โทรสาร ๐ ๔๒๙๗ ๐๐๓๒
ผู้ประสานงาน นายศรันยู ชูทองรัตนะ โทรศัพท์เคลื่อนที่ ๐๙ ๑๐๖๑ ๖๓๐๘

"อยู่สกล รักสกล ทำเพื่อสกลนคร"

ที่ อว ๐๖๒๑.๑๒/ว ๒๙๓

มหาวิทยาลัยราชภัฏสกลนคร
๖๘๐ ถ.นิตโย ต.ธาตุเชิงชุม
อ.เมือง จ.สกลนคร ๔๗๐๐๐

๘ มีนาคม ๒๕๖๔

เรื่อง ขอความอนุเคราะห์เป็นผู้เชี่ยวชาญเสนอแนวทางพัฒนา

เรียน นางอุไรลักษณ์ คำหา

สิ่งที่ส่งมาด้วย เครื่องมือการวิจัย

ด้วย นายศรัณยู ชูทองรัตนะ รหัสประจำตัวนักศึกษา ๕๙๔๒๕๑๑๗๑๐๒ ซึ่งเป็นนักศึกษาปริญญาโท สาขาวิชาวิทยาการสารสนเทศและเทคโนโลยี มหาวิทยาลัยราชภัฏสกลนคร กำลังศึกษาและทำวิทยานิพนธ์เรื่อง **"แบบจำลองเพื่อพยากรณ์การเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุด้วยเทคนิคเหมืองข้อมูล กรณีศึกษาโรงพยาบาลนาหว้า"** ซึ่งเป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรมหาบัณฑิต (วท.ม.) โดยมี ดร.สุทิศา ชองเหล็กนอก เป็นประธานที่ปรึกษาวิทยานิพนธ์ และ ดร.สุธาสินี คุปตะบุตร เป็นกรรมการที่ปรึกษาวิทยานิพนธ์

มหาวิทยาลัยราชภัฏสกลนคร ได้พิจารณาแล้วเห็นว่าท่านเป็นผู้มีความรู้ความสามารถและเชี่ยวชาญในเรื่องดังกล่าว จึงขอความอนุเคราะห์ท่านเป็นผู้เชี่ยวชาญเสนอแนวทางพัฒนาแบบจำลองเพื่อพยากรณ์การเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุด้วยเทคนิคเหมืองข้อมูล กรณีศึกษาโรงพยาบาลนาหว้า **ในวันที่ ๙ มีนาคม พ.ศ. ๒๕๖๔ เวลา ๑๓.๓๐ น. ณ ห้องประชุมพุทธรักษา โรงพยาบาลนาหว้า** โดยเครื่องมือที่ใช้ในการวิจัยครั้งนี้ ได้ผ่านการรับรองจริยธรรมการวิจัยในมนุษย์แบบยกเว้น จากคณะกรรมการจริยธรรมการวิจัยในมนุษย์ มหาวิทยาลัยราชภัฏสกลนคร เลขที่ใบรับรอง ๐๔๗/๒๕๖๓ ลงวันที่ ๑๒ พฤศจิกายน พ.ศ. ๒๕๖๓

จึงเรียนมาเพื่อโปรดพิจารณา หวังเป็นอย่างยิ่งว่าคงได้รับความอนุเคราะห์จากท่านด้วยดี และขอขอบคุณมา ณ โอกาสนี้

ขอแสดงความนับถือ

(รองศาสตราจารย์ ดร.ศิกานต์ เพียรธัญญกรณ์)
ผู้อำนวยการบัณฑิตวิทยาลัย ปฏิบัติราชการแทน
อธิการบดีมหาวิทยาลัยราชภัฏสกลนคร

บัณฑิตวิทยาลัย
โทร. ๐ ๔๓๙๗ ๐๒๒๙
โทรสาร ๐ ๔๒๙๗ ๐๐๓๒
ผู้ประสานงาน นายศรันยู ชูทองรัตนะ โทรศัพท์เคลื่อนที่ ๐๙ ๑๐๖๑ ๖๓๐๘

"อยู่สกล รักสกล ทำเพื่อสกลนคร"

ที่ อว ๐๖๒๑.๑๒/ว ๒๙๓

มหาวิทยาลัยราชภัฏสกลนคร
๖๘๐ ถ.นิตโย ต.ธาตุเชิงชุม
อ.เมือง จ.สกลนคร ๔๗๐๐๐

๘ มีนาคม ๒๕๖๔

เรื่อง ขอความอนุเคราะห์เป็นผู้เชี่ยวชาญเสนอแนวทางพัฒนา

เรียน นางอรรชร สุนทรารักษ์

สิ่งที่ส่งมาด้วย เครื่องมือการวิจัย

ด้วย นายศรัณยู ชูทองรัตนะ รหัสประจำตัวนักศึกษา ๕๙๔๒๕๑๑๗๑๐๒ ซึ่งเป็นนักศึกษาปริญญาโท สาขาวิชาวิทยาการสารสนเทศและเทคโนโลยี มหาวิทยาลัยราชภัฏสกลนคร กำลังศึกษาและทำวิทยานิพนธ์เรื่อง **"แบบจำลองเพื่อพยากรณ์การเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุด้วยเทคนิคเหมืองข้อมูล กรณีศึกษาโรงพยาบาลนาหว้า"** ซึ่งเป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรมหาบัณฑิต (วท.ม.) โดยมี ดร.สุทิศา ชองเหล็กนอก เป็นประธานที่ปรึกษาวิทยานิพนธ์ และ ดร.สุธาสินี คุปตะบุตร เป็นกรรมการที่ปรึกษาวิทยานิพนธ์

มหาวิทยาลัยราชภัฏสกลนคร ได้พิจารณาแล้วเห็นว่าท่านเป็นผู้มีความรู้ความสามารถและเชี่ยวชาญในเรื่องดังกล่าว จึงขอความอนุเคราะห์ท่านเป็นผู้เชี่ยวชาญเสนอแนวทางพัฒนาแบบจำลองเพื่อพยากรณ์การเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุด้วยเทคนิคเหมืองข้อมูล กรณีศึกษาโรงพยาบาลนาหว้า **ในวันที่ ๙ มีนาคม พ.ศ. ๒๕๖๔ เวลา ๑๓.๓๐ น. ณ ห้องประชุมพุทธรักษา โรงพยาบาลนาหว้า** โดยเครื่องมือที่ใช้ในการวิจัยครั้งนี้ ได้ผ่านการรับรองจริยธรรมการวิจัยในมนุษย์แบบยกเว้น จากคณะกรรมการจริยธรรมการวิจัยในมนุษย์ มหาวิทยาลัยราชภัฏสกลนคร เลขที่ใบรับรอง ๐๔๗/๒๕๖๓ ลงวันที่ ๑๒ พฤศจิกายน พ.ศ. ๒๕๖๓

จึงเรียนมาเพื่อโปรดพิจารณา หวังเป็นอย่างยิ่งว่าคงได้รับความอนุเคราะห์จากท่านด้วยดี และขอขอบคุณมา ณ โอกาสนี้

ขอแสดงความนับถือ

(รองศาสตราจารย์ ดร.ศิกานต์ เพียรธัญญกรณ์)
ผู้อำนวยการบัณฑิตวิทยาลัย ปฏิบัติราชการแทน
อธิการบดีมหาวิทยาลัยราชภัฏสกลนคร

บัณฑิตวิทยาลัย
โทร. ๐ ๔๓๙๗ ๐๒๒๙
โทรสาร ๐ ๔๒๙๗ ๐๐๓๒
ผู้ประสานงาน นายศรัณยู ชูทองรัตนะ โทรศัพท์เคลื่อนที่ ๐๙ ๑๐๖๑ ๖๓๐๘

"อยู่สกล รักสกล ทำเพื่อสกลนคร"

ที่ อว ๐๖๒๑.๑๒/ว ๒๙๓

มหาวิทยาลัยราชภัฏสกลนคร
๖๘๐ ถ.นิตโย ต.ธาตุเชิงชุม
อ.เมือง จ.สกลนคร ๔๗๐๐๐

๘ มีนาคม ๒๕๖๔

เรื่อง ขอความอนุเคราะห์เป็นผู้เชี่ยวชาญเสนอแนวทางพัฒนา

เรียน นางสาววิภาสินี วะเศษสร้อย

สิ่งที่ส่งมาด้วย เครื่องมือการวิจัย

ด้วย นายศรัณยู ชูทองรัตนะ รหัสประจำตัวนักศึกษา ๕๙๔๒๕๑๑๗๑๐๒ ซึ่งเป็นนักศึกษาปริญญาโท สาขาวิชาวิทยาการสารสนเทศและเทคโนโลยี มหาวิทยาลัยราชภัฏสกลนคร กำลังศึกษาและทำวิทยานิพนธ์เรื่อง **“แบบจำลองเพื่อพยากรณ์การเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุด้วยเทคนิคเหมืองข้อมูล กรณีศึกษาโรงพยาบาลนาหว้า”** ซึ่งเป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรมหาบัณฑิต (วท.ม.) โดยมี ดร.สุทิศา ชองเหล็กนอก เป็นประธานที่ปรึกษาวิทยานิพนธ์ และ ดร.สุธาสินี คุปตะบุตร เป็นกรรมการที่ปรึกษาวิทยานิพนธ์

มหาวิทยาลัยราชภัฏสกลนคร ได้พิจารณาแล้วเห็นว่าท่านเป็นผู้มีความรู้ความสามารถและเชี่ยวชาญในเรื่องดังกล่าว จึงขอความอนุเคราะห์ท่านเป็นผู้เชี่ยวชาญเสนอแนวทางพัฒนาแบบจำลองเพื่อพยากรณ์การเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุด้วยเทคนิคเหมืองข้อมูล กรณีศึกษาโรงพยาบาลนาหว้า **ในวันที่ ๙ มีนาคม พ.ศ. ๒๕๖๔ เวลา ๑๓.๓๐ น. ณ ห้องประชุมพุทธรักษา โรงพยาบาลนาหว้า** โดยเครื่องมือที่ใช้ในการวิจัยครั้งนี้ ได้ผ่านการรับรองจริยธรรมการวิจัยในมนุษย์แบบยกเว้น จากคณะกรรมการจริยธรรมการวิจัยในมนุษย์ มหาวิทยาลัยราชภัฏสกลนคร เลขที่ใบรับรอง ๐๔๗/๒๕๖๓ ลงวันที่ ๑๒ พฤศจิกายน พ.ศ. ๒๕๖๓

จึงเรียนมาเพื่อโปรดพิจารณา หวังเป็นอย่างยิ่งว่าคงได้รับความอนุเคราะห์จากท่านด้วยดี และขอขอบคุณมา ณ โอกาสนี้

ขอแสดงความนับถือ

(รองศาสตราจารย์ ดร.ศิกานต์ เพียรธัญญกรณ์)
ผู้อำนวยการบัณฑิตวิทยาลัย ปฏิบัติราชการแทน
อธิการบดีมหาวิทยาลัยราชภัฏสกลนคร

บัณฑิตวิทยาลัย
โทร. ๐ ๔๓๙๗ ๐๒๒๙
โทรสาร ๐ ๔๒๙๗ ๐๐๓๒
ผู้ประสานงาน นายศรัณยู ชูทองรัตนะ โทรศัพท์เคลื่อนที่ ๐๙ ๑๐๖๑ ๖๓๐๘

“อยู่สกล รักสกล ทำเพื่อสกลนคร”

ที่ อว ๐๖๒๑.๑๒/ว ๒๙๓

มหาวิทยาลัยราชภัฏสกลนคร
๖๘๐ ถ.นิตโย ต.ธาตุเชิงชุม
อ.เมือง จ.สกลนคร ๔๗๐๐๐

๘ มีนาคม ๒๕๖๔

เรื่อง ขอความอนุเคราะห์เป็นผู้เชี่ยวชาญเสนอแนวทางพัฒนา

เรียน นางพูล สุวัฒนเสนีย์

สิ่งที่ส่งมาด้วย เครื่องมือการวิจัย

ด้วย นายศรัณยู ชูทองรัตนะ รหัสประจำตัวนักศึกษา ๕๙๔๒๕๑๑๗๑๐๒ ซึ่งเป็นนักศึกษาปริญญาโท สาขาวิชาวิทยาการสารสนเทศและเทคโนโลยี มหาวิทยาลัยราชภัฏสกลนคร กำลังศึกษาและทำวิทยานิพนธ์เรื่อง **"แบบจำลองเพื่อพยากรณ์การเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุด้วยเทคนิคเหมืองข้อมูล กรณีศึกษาโรงพยาบาลนาหว้า"** ซึ่งเป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรมหาบัณฑิต (วท.ม.) โดยมี ดร.สุทิศา ชองเหล็กนอก เป็นประธานที่ปรึกษาวิทยานิพนธ์ และ ดร.สุธาสินี คุปตะบุตร เป็นกรรมการที่ปรึกษาวิทยานิพนธ์

มหาวิทยาลัยราชภัฏสกลนคร ได้พิจารณาแล้วเห็นว่าท่านเป็นผู้มีความรู้ความสามารถและเชี่ยวชาญในเรื่องดังกล่าว จึงขอความอนุเคราะห์ท่านเป็นผู้เชี่ยวชาญเสนอแนวทางพัฒนาแบบจำลองเพื่อพยากรณ์การเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุด้วยเทคนิคเหมืองข้อมูล กรณีศึกษาโรงพยาบาลนาหว้า **ในวันที่ ๙ มีนาคม พ.ศ. ๒๕๖๔ เวลา ๑๓.๓๐ น. ณ ห้องประชุมพุทธรักษา โรงพยาบาลนาหว้า** โดยเครื่องมือที่ใช้ในการวิจัยครั้งนี้ ได้ผ่านการรับรองจริยธรรมการวิจัยในมนุษย์แบบยกเว้น จากคณะกรรมการจริยธรรมการวิจัยในมนุษย์ มหาวิทยาลัยราชภัฏสกลนคร เลขที่ใบรับรอง ๐๔๗/๒๕๖๓ ลงวันที่ ๑๒ พฤศจิกายน พ.ศ. ๒๕๖๓

จึงเรียนมาเพื่อโปรดพิจารณา หวังเป็นอย่างยิ่งว่าคงได้รับความอนุเคราะห์จากท่านด้วยดี และขอขอบคุณมา ณ โอกาสนี้

ขอแสดงความนับถือ

(รองศาสตราจารย์ ดร.ศิกานต์ เพียรธัญญกรณ์)
ผู้อำนวยการบัณฑิตวิทยาลัย ปฏิบัติราชการแทน
อธิการบดีมหาวิทยาลัยราชภัฏสกลนคร

บัณฑิตวิทยาลัย
โทร. ๐ ๔๓๙๗ ๐๒๒๙
โทรสาร ๐ ๔๒๙๗ ๐๐๓๒
ผู้ประสานงาน นายศรันยู ชูทองรัตนะ โทรศัพท์เคลื่อนที่ ๐๙ ๑๐๖๑ ๖๓๐๘

"อยู่สกล รักสกล ทำเพื่อสกลนคร"

ที่ อว ๐๖๒๑.๑๒/ว ๒๙๓

มหาวิทยาลัยราชภัฏสกลนคร
๖๘๐ ถ.นิตโย ต.ธาตุเชิงชุม
อ.เมือง จ.สกลนคร ๔๗๐๐๐

๘ มีนาคม ๒๕๖๔

เรื่อง ขอความอนุเคราะห์เป็นผู้เชี่ยวชาญเสนอแนวทางพัฒนา

เรียน นางสาวทัศนีย์ สิทธิ

สิ่งที่ส่งมาด้วย เครื่องมือการวิจัย

ด้วย นายศรัณยู ชูทองรัตนะ รหัสประจำตัวนักศึกษา ๕๙๔๒๕๑๑๗๑๐๒ ซึ่งเป็นนักศึกษาปริญญาโท สาขาวิชาวิทยาการสารสนเทศและเทคโนโลยี มหาวิทยาลัยราชภัฏสกลนคร กำลังศึกษาและทำวิทยานิพนธ์เรื่อง **"แบบจำลองเพื่อพยากรณ์การเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุด้วยเทคนิคเหมืองข้อมูล กรณีศึกษาโรงพยาบาลนาหว้า"** ซึ่งเป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรมหาบัณฑิต (วท.ม.) โดยมี ดร.สุทิศา ชองเหล็กนอก เป็นประธานที่ปรึกษาวิทยานิพนธ์ และ ดร.สุธาสินี คุปตะบุตร เป็นกรรมการที่ปรึกษาวิทยานิพนธ์

มหาวิทยาลัยราชภัฏสกลนคร ได้พิจารณาแล้วเห็นว่าท่านเป็นผู้มีความรู้ความสามารถและเชี่ยวชาญในเรื่องดังกล่าว จึงขอความอนุเคราะห์ท่านเป็นผู้เชี่ยวชาญเสนอแนวทางพัฒนาแบบจำลองเพื่อพยากรณ์การเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุด้วยเทคนิคเหมืองข้อมูล กรณีศึกษาโรงพยาบาลนาหว้า **ในวันที่ ๙ มีนาคม พ.ศ. ๒๕๖๔ เวลา ๑๓.๓๐ น. ณ ห้องประชุมพุทธรักษา โรงพยาบาลนาหว้า** โดยเครื่องมือที่ใช้ในการวิจัยครั้งนี้ ได้ผ่านการรับรองจริยธรรมการวิจัยในมนุษย์แบบยกเว้น จากคณะกรรมการจริยธรรมการวิจัยในมนุษย์ มหาวิทยาลัยราชภัฏสกลนคร เลขที่ใบรับรอง ๐๔๗/๒๕๖๓ ลงวันที่ ๑๒ พฤศจิกายน พ.ศ. ๒๕๖๓

จึงเรียนมาเพื่อโปรดพิจารณา หวังเป็นอย่างยิ่งว่าคงได้รับความอนุเคราะห์จากท่านด้วยดี และขอขอบคุณมา ณ โอกาสนี้

ขอแสดงความนับถือ

(รองศาสตราจารย์ ดร.ศิกานต์ เพียรธัญญกรณ์)
ผู้อำนวยการบัณฑิตวิทยาลัย ปฏิบัติราชการแทน
อธิการบดีมหาวิทยาลัยราชภัฏสกลนคร

บัณฑิตวิทยาลัย
โทร. ๐ ๔๓๙๗ ๐๒๒๙
โทรสาร ๐ ๔๒๙๗ ๐๐๓๒
ผู้ประสานงาน นายศรัณยู ชูทองรัตนะ โทรศัพท์เคลื่อนที่ ๐๙ ๑๐๖๑ ๖๓๐๘

"อยู่สกล รักสกล ทำเพื่อสกลนคร"

แบบประเมินแบบจำลองเพื่อพยากรณ์การเกิดโรคไม่ติดต่อเรื้อรัง ในผู้สูงอายุด้วยเทคนิคเหมืองข้อมูลโดยผู้เชี่ยวชาญ

คำชี้แจง

1.แบบประเมินนี้ใช้ประกอบการบันทึกผลประเมินของผู้เชี่ยวชาญ สำหรับประชุมกลุ่ม (Focus Group) ผู้เชี่ยวชาญในการประเมินแบบจำลองเพื่อพยากรณ์การเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุด้วยเทคนิคเหมืองข้อมูล

2.ผลการประเมินจะถูกนำไปใช้ในการเขียนผลการวิจัยในบทที่ 4 เมื่อเขียนผลการวิจัยเสร็จแล้ว ข้อมูลในแบบประเมินนี้จะถูกทำลายทันที

ตอนที่ 1 ข้อมูลผู้เชี่ยวชาญ

ชื่อ–สกุล .....

ตำแหน่ง .....ระดับ.....

ชื่อหน่วยงานที่สังกัด.....

ตอนที่ 2 การประเมินระดับความคิดเห็นที่มีต่อแบบจำลองการพยากรณ์การเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุ

คำชี้แจง ทำเครื่องหมาย ✓ ลงในช่องตามระดับความคิดเห็นของท่าน

ประเด็นความคิดเห็น	ระดับความคิดเห็น				
	มากที่สุด 5	มาก 4	ปานกลาง 3	น้อย 2	น้อยที่สุด 1
1. ความครบถ้วนของปัจจัยที่ใช้ในการพยากรณ์					
2. แบบจำลองโรคเบาหวาน					
2.1 ความถูกต้องของโมเดล					
2.2 ความน่าเชื่อถือของผลการทำนาย					
2.3 ความสามารถนำไปประกอบการตัดสินใจ					
3. แบบจำลองโรคความดันโลหิตสูง					
3.1 ความถูกต้องของโมเดล					
3.2 ความน่าเชื่อถือของผลการทำนาย					
3.3 ความสามารถนำไปประกอบการตัดสินใจ					
4. แบบจำลองโรคหัวใจขาดเลือด					
4.1 ความถูกต้องของโมเดล					
4.2 ความน่าเชื่อถือของผลการทำนาย					
4.3 ความสามารถนำไปประกอบการตัดสินใจ					
5. ประโยชน์ของแบบจำลองการพยากรณ์การเกิดโรคไม่ติดต่อเรื้อรังในผู้สูงอายุ					

ตอนที่ 3 ข้อเสนอแนะอื่นๆ

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

ลงชื่อ ..... ผู้เชี่ยวชาญ

งานประชุมวิชาการระดับชาตินนทรีอีสาน ครั้งที่ 9 (ออนไลน์)
มหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตเฉลิมพระเกียรติ จังหวัดสกลนคร

การเปรียบเทียบประสิทธิภาพการพยากรณ์การเกิดโรคเบาหวานในผู้สูงอายุ ด้วยเทคนิคเหมืองข้อมูล กรณีศึกษาโรงพยาบาลนาหว้า

An Efficiency Comparison of Prediction for Diabetes Mellitus in the Elderly People Using Data Mining Techniques Case Study of Nawa Hospital

ศรัณยู ชูทองรัตนะ[1*] สุทิศา ซองเหล็กนอก[2] และสุธาสินี คุปตะบุตร[3]
Saranyu Chuthongrattana[1*] Sutisa Songleknok[2] and Suthasinee Kuptabut[3]

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อสร้างและเปรียบเทียบแบบจำลองพยากรณ์การเกิดโรคเบาหวานในผู้สูงอายุด้วยการใช้เทคนิคเหมืองข้อมูล 3 เทคนิคได้แก่ ต้นไม้ตัดสินใจ เพื่อนบ้านใกล้ที่สุด และนาอีฟเบย์ โดยทำการศึกษากับข้อมูลผู้สูงอายุที่รับบริการ โรงพยาบาลนาหว้า จังหวัดนครพนม ตั้งแต่เดือนตุลาคม พ.ศ. 2554 ถึง เดือนกันยายน พ.ศ. 2562 จำนวน 262,655 ระเบียน

ผลการสร้างแบบจำลองด้วยเทคนิคเหมืองข้อมูล 3 เทคนิคพบว่าเทคนิคต้นไม้ตัดสินใจ มีค่าความถูกต้อง 0.952 มากกว่าค่าความถูกต้องของเทคนิคเพื่อนบ้านใกล้ที่สุด (ค่าความถูกต้อง 0.951) และ เทคนิคนาอีฟเบย์ (ค่าความถูกต้อง 0.950) และนำแบบจำลองต้นไม้ตัดสินใจให้ผู้เชี่ยวชาญประเมินประสิทธิภาพมีค่าเฉลี่ยรวม 4.43 (อยู่ในระดับมาก)

คำสำคัญ: เหมืองข้อมูล ผู้สูงอายุ เบาหวาน แบบจำลอง

[1] สาขาวิชาวิทยาการสารสนเทศและเทคโนโลยี คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏสกลนคร
[1] Information Science and Technology, Faculty of Science and Technology, Sakon Nakhon Rajabhat University
[2,3] สาขาวิชาคอมพิวเตอร์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏสกลนคร
[2,3] Computer Program, Faculty of Science and Technology, Sakon Nakhon Rajabhat University
* Corresponding author; e-mail address: saranyu.f@gmail.com

Abstract

The purpose of this research was to create and comparison a predictive model of diabetes mellitus in the elderly by using 3 data mining techniques, Decision Tree K-Nearest Neighbors and Naive Bayes by studying the elderly service Nawa Hospital Nakhon Phanom Province from October 2011 to September 2019, 262,655 records

The results of modeling by 3 data mining techniques showed that the Decision Tree technique it has an accuracy of 0.952 greater than the accuracy of the K-Nearest Neighbors technique (Accuracy 0.951) and Naive Bayes technique (Accuracy 0.950) respectively, And Decision Tree model was evaluated by experts with a total average efficiency of 4.43 (At a high level)

Keywords: Data Mining, Elderly, Diabetes Mellitus, Model

คำนำ

ปัจจุบันประเทศไทยเข้าสู่สังคมผู้สูงอายุส่งผลให้ภาพรวมของประเทศต้องจัดสรรงบประมาณทางด้านสุขภาพรวมถึงสวัสดิการต่างๆ ที่เกี่ยวข้องกับผู้สูงอายุ เพื่อเตรียมการและป้องกันปัญหาสุขภาพและความเสี่ยงสูงต่อการเกิดโรคเรื้อรัง (Chronic disease) เช่น โรคเบาหวาน โรคความดัน ทำให้รัฐบาลต้องจัดสรรงบประมาณค่ารักษาพยาบาลเพิ่มขึ้น ขณะเดียวกันผู้สูงบางรายไม่มีรายได้หรือมีรายได้ลดลง ทำให้การเข้าถึงบริการสุขภาพลดลง ส่งผลให้ความเป็นอยู่ของผู้สูงอายุมีความลำบาก กระทรวงสาธารณสุขจึงมียุทธศาสตร์ที่จะพัฒนาด้านการสร้างเสริมสุขภาพและป้องกันโรคเพื่อให้ประชาชนมีพฤติกรรมสุขภาพที่ถูกต้องเหมาะสม เพื่อให้การดูแลทั่วถึง กระทรวงสาธารณสุขจึงนำเทคโนโลยีสารสนเทศมารวบรวม จัดเก็บข้อมูลการบริการด้านสุขภาพของประชาชน (กลุ่มยุทธศาสตร์และแผนงาน สำนักโรคไม่ติดต่อ, 2559) เพื่อใช้ในการวางแผนงาน การสนับสนุนการตัดสินใจ และการจัดทำสถิติด้านสุขภาพให้เหมาะสมและเป็นประโยชน์ต่อการทำงานของกระทรวง และก่อให้เกิดประโยชน์สูงสุดกับประชาชน เช่น ข้อมูลคนไข้ ข้อมูลเวชภัณฑ์ เป็นต้น เกิดเป็นข้อมูลขนาดใหญ่ หรือที่เรียกว่า อภิมหาข้อมูล (Big data) โดยเทคโนโลยีในการวิเคราะห์ข้อมูลขนาดใหญ่และหาความรู้ที่สำคัญเพื่อนำไปประกอบการตัดสินใจคือ การทำเหมืองข้อมูล (Data mining) (สายชล, 2560; เอกสิทธิ์, 2557)

จากการทบทวนวรรณกรรมที่เกี่ยวข้องกับเหมืองข้อมูลพบว่าปัจจุบันนักวิชาการที่ทำงานเกี่ยวข้องกับข้อมูลสาธารณสุขได้นำเทคนิคเหมืองข้อมูลมาใช้เพื่อหาปัจจัย สร้างแบบจำลอง คาดการณ์หรือพยากรณ์จากข้อมูลที่มีอยู่เช่น การสร้างโมเดลทำนายการกลับมารักษาซ้ำ การวิเคราะห์ปัจจัยที่มีผลต่อการเกิดโรค หรือพยากรณ์ความสัมพันธ์ปัจจัยที่แทรกซ้อนก่อให้เกิดโรคเพื่อใช้เป็นแนวทางในการรักษาของแพทย์และหน่วยงานอื่นนำเทคนิคเหมืองข้อมูลไปทำงานกับข้อมูลของหน่วยงาน เช่น มหาวิทยาลัยใช้เหมืองข้อมูลในการพยากรณ์ผลการเรียนของนักเรียน การท่องเที่ยวใช้เทคนิคเหมืองข้อมูลพัฒนาฐานข้อมูลแนะนำแหล่งท่องเที่ยวทาง

ธรรมชาติจากความต้องการและพฤติกรรมของนักท่องเที่ยว นำไปสู่การพัฒนาระบบสนับสนุนการตัดสินใจในการปฏิบัติงาน หรือวางแผนการทำงานในอนาคต (วิธวินท์ และคณะ, 2560; เสกสรรค์, 2558; ชฎิภกรณ์, 2557; เกรียงศักดิ์, 2561) ซึ่งเทคนิคเหมือนข้อมูลที่ใช้มีหลายเทคนิค เช่น เทคนิคต้นไม้ตัดสินใจ เป็นการคัดเลือกแอตทริบิวต์ที่มีความสัมพันธ์กับคลาสมากที่สุดขึ้นมาเป็นโหนดบนสุดของต้นไม้ (Root node) หลังจากนั้นจะหาแอตทริบิวต์ถัดไปโดยจะใช้ตัววัดคือค่าเกนความรู้ (Information gain) เทคนิคเพื่อนบ้านใกล้ที่สุด เป็นการจัดกลุ่ม เปรียบเทียบความคล้ายคลึงกันของข้อมูลที่สนใจกับข้อมูลอื่นว่ามีความคล้ายคลึงหรืออยู่ใกล้กับข้อมูลใดมากที่สุด k ตัว โดยจะใช้วิธีการวัดระยะทางแบบยุคลิด (Euclidean distance) เทคนิคนาอีฟเบย์เป็นการใช้ทฤษฎีความน่าจะเป็นตามกฎของเบย์ (Bayes' theorem) เพื่อหาว่าสมมติฐานใดน่าจะถูกต้องที่สุด โดยใช้ความรู้ก่อนหน้า (Prior knowledge) เทคนิคโครงข่ายประสาทเทียมเป็นการใช้โมเดลทางคณิตศาสตร์สำหรับประมวลผลสารสนเทศด้วยการคำนวณแบบคอนเนคชันนิสต์ (Connectionist) เพื่อจำลองการทำงานของเครือข่ายประสาทในสมองมนุษย์โดยสามารถเรียนรู้การจดจำรูปแบบ (Pattern recognition) และการสร้างความรู้ใหม่ (Knowledge extraction) และเทคนิคซัพพอร์ตเวกเตอร์แมทชีน (Support vector machine) เป็นการใช้หลักการหาสัมประสิทธิ์ของสมการเพื่อสร้างเส้นแบ่งแยกกลุ่มข้อมูลที่ถูกป้อนเข้าสู่กระบวนการสอนให้ระบบเรียนรู้ โดยเน้นไปยังเส้นแบ่งแยกแยะกลุ่มข้อมูล

ผู้วิจัยจึงสนใจที่จะศึกษาและสร้างแบบจำลองเพื่อพยากรณ์การเกิดโรคด้วยเทคนิคเหมืองข้อมูล โดยเลือกศึกษาโรคเบาหวาน ซึ่งเป็นโรคที่พบมากในผู้สูงอายุ (สำนักงานกองทุนสนับสนุนการสร้างเสริมสุขภาพ, 2561) โดยใช้ข้อมูลจากฐานข้อมูลของโรงพยาบาลนาหว้า ด้วยเทคนิคเหมืองข้อมูล 3 เทคนิคคือ เทคนิคต้นไม้ตัดสินใจ เทคนิคเพื่อนบ้านใกล้ที่สุด และเทคนิคนาอีฟเบย์ เพื่อวิเคราะห์หาปัจจัย ความสัมพันธ์ สร้างและเปรียบเทียบแบบจำลองโรคเบาหวานในผู้สูงอายุ สำหรับหาความเสี่ยงการเป็นโรคเบาหวานของผู้รับบริการทั่วไป

เครื่องมือและวิธีการ

เครื่องมือ

1. โปรแกรมไมโครซอฟท์เอ็กเซล (Microsoft excel) ใช้ในการเตรียมข้อมูล
2. โปรแกรมเวก้า (Weka) ใช้ในการพยากรณ์โรคเบาหวาน
3. แบบจำลองจากเทคนิคที่มีประสิทธิภาพดีที่สุด โดยวัดจากค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าความถ่วงดุล (F-Measure)

ข้อมูล

ข้อมูลของผู้รับบริการที่มีอายุ 60 ปีขึ้นไป ที่มารับการรักษาตั้งแต่เดือนตุลาคม พ.ศ. 2554 ถึงเดือนกันยายน พ.ศ. 2562 จำนวน 262,655 ระเบียน

วิธีการ

งานวิจัยนี้ผู้วิจัยได้เสนอโครงการวิจัยเพื่อพิจารณาจริยธรรมการวิจัยในมนุษย์กับคณะกรรมการจริยธรรมการวิจัยในมนุษย์ มหาวิทยาลัยราชภัฏสกลนครเรียบร้อยแล้ว เลขที่ใบรับรอง ๐๔๗/๒๕๖๓

วิธีการทำงาน ผู้วิจัยแบ่งเป็น 3 ระยะดังนี้

ระยะที่ 1 วิเคราะห์และเตรียมข้อมูล

ผู้วิจัยศึกษาทำความเข้าใจกับเนื้อหาข้อมูลและเตรียมข้อมูลสำหรับการนำไปทำเหมืองข้อมูล โดยอ้างอิงจากกระบวนเหมืองข้อมูลของคริสป์-ดีเอ็ม (CRISP-DM) ประกอบด้วย 3 ขั้นคือ 1) ขั้นตอนการทำความเข้าใจการวิจัย 2) ขั้นตอนการทำความเข้าใจข้อมูล และ 3) ขั้นตอนการจัดเตรียมข้อมูล

การเก็บรวบรวมข้อมูลในระยะที่ 1 วิเคราะห์และเตรียมข้อมูลมีขั้นตอนดังนี้

1. ขั้นตอนการทำความเข้าใจการวิจัย

ผู้วิจัยทำการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องกับโรคเบาหวาน และการพยากรณ์ในการเกิดโรคเบาหวานด้วยเทคนิคเหมืองข้อมูล 3 เทคนิค ได้แก่ เทคนิคต้นไม้ตัดสินใจ เทคนิคนาอีฟเบย์ และเทคนิคเพื่อนบ้านใกล้ที่สุด

2. ขั้นตอนการทำความเข้าใจข้อมูล

ผู้วิจัยทำความเข้าใจข้อมูลและตัวแปรที่ก่อให้เกิดโรคเบาหวาน จำนวน 7 ตัวแปร ได้แก่ เพศ อายุ การสูบบุหรี่ ดัชนีมวลกาย ความดันโลหิต น้ำตาลในเลือด ไขมันในเลือด (สมเกียรติ และคณะฯ, 2557) และตัวแปรผลลัพธ์ ได้แก่ เป็นหรือไม่เป็นโรคเบาหวาน

3. ขั้นตอนการจัดการกับข้อมูล

นำข้อมูลในขั้นตอนที่ 2 มาทำความสะอาด (Cleansing) ด้วยโปรแกรมไมโครซอฟท์เอ็กเซล โดยตรวจสอบระเบียนข้อมูลที่มีค่าไม่ครบทุกฟิลด์หรือระเบียนที่มีค่าผิดปกติ (Outliner) เช่น อายุ 300 ปี น้ำหนัก 251 กิโลกรัม ดังนั้นผู้วิจัยลบระเบียนที่มีค่าผิดปกติจำนวน 11 ระเบียน ดังนั้นเหลือระเบียบที่ใช้สร้างแบบจำลองจำนวน 262,644 ระเบียน จากนั้นทำการแปลงข้อมูลให้อยู่ในรูปแบบพร้อมใช้ในการวิเคราะห์

ระยะที่ 2 สร้างแบบจำลอง

ผู้วิจัยนำข้อมูลที่ผ่านกระบวนการระยะที่ 1 ไปสร้างแบบจำลองด้วยโปรแกรมเวก้า ด้วยเทคนิคเหมืองข้อมูล 3 เทคนิคคือ ต้นไม้ตัดสินใจ เพื่อนบ้านใกล้ที่สุด และนาอีฟเบย์ ในการวัดประสิทธิภาพแบบจำลอง ผู้วิจัยได้ทำการทดลองตามหลักการแบ่งข้อมูลออกเป็นชุดเรียนรู้และชุดทดสอบ โดยการแบ่งข้อมูลออกเป็น 10 ส่วนเท่าๆ กัน ใช้ 9 ส่วนเป็นข้อมูลสอน และ 1 ส่วน เป็นข้อมูลทดสอบ ทำงานสลับกันจนครบ 10 รอบ (10 Fold cross-validation) การประเมินประสิทธิภาพใช้ตัววัดจากค่าความถูกต้อง ค่าความแม่นยำ ค่าความระลึกและค่าความถ่วงดุล

ระยะที่ 3 ประเมินประสิทธิภาพแบบจำลองโดยผู้เชี่ยวชาญ

ผู้วิจัยนำแบบจำลองมาเปรียบเทียบประสิทธิภาพการพยากรณ์การเกิดโรคเบาหวาน แล้วเลือกค่าประสิทธิภาพดีที่สุดให้ผู้เชี่ยวชาญด้านสาธารณสุขที่มีประสบการณ์การทำงานด้านสาธารณสุขไม่น้อยกว่า 10 ปี หรือมีตำแหน่งชำนาญการขึ้นไป จำนวน 7 คน ประเมินประสิทธิภาพของแบบจำลอง ด้วยวิธีประชุมกลุ่ม (Focus group)

สถิติที่ใช้ในการวิจัย

ค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐาน

ผลและวิจารณ์ผลการทดลอง

1. ผลจากการวิเคราะห์และเตรียมข้อมูล

ผลจากการวิเคราะห์และเตรียมข้อมูลจากฐานข้อมูลผู้มารับบริการอายุ 60 ปีขึ้นไป ของโรงพยาบาลนาหว้ามีการแทนค่าข้อมูลดังนี้ (Table 1)

1.1 เพศ ข้อมูลมีค่าคือ 1 หมายถึงเพศชาย และ 2 หมายถึงเพศหญิง ผู้วิจัยแทนค่าเพศ 1 ด้วย M และ เพศ 2 ด้วย F

1.2 อายุ ผู้วิจัยมีการจัดกลุ่มใหม่ 3 กลุ่มดังนี้

กลุ่มที่ 1 ผู้สูงอายุวัยต้น (อายุ 60-69 ปี) แทนด้วย 1

กลุ่มที่ 2 ผู้สูงอายุวัยกลาง (อายุ 70-79 ปี) แทนด้วย 2

กลุ่มที่ 3 ผู้สูงอายุวัยปลาย (อายุ 80 ปีขึ้นไป) แทนด้วย 3

1.3 การสูบบุหรี่ ข้อมูลมีค่าคือ 1 หมายถึงไม่สูบบุหรี่ และ 2 หมายถึงสูบบุหรี่ ผู้วิจัยแทนค่าการสูบบุหรี่ 1 ด้วย N และ การสูบบุหรี่ 2 ด้วย Y

1.4 ดัชนีมวลกาย ผู้วิจัยมีการจัดกลุ่มใหม่ 5 กลุ่มดังนี้

กลุ่มที่ 1 คือค่า BMI น้อยกว่า 18.5 แทนด้วย 1

กลุ่มที่ 2 คือค่า BMI 18.5 ถึง 22.9 แทนด้วย 2

กลุ่มที่ 3 คือค่า BMI 23 ถึง 24.9 แทนด้วย 3

กลุ่มที่ 4 คือค่า BMI 25 ถึง 29.9 แทนด้วย 4

กลุ่มที่ 5 คือค่า BMI ตั้งแต่ 30 ขึ้นไป แทนด้วย 5

1.5 ความดันโลหิต ผู้วิจัยมีการจัดกลุ่มใหม่ 4 กลุ่มดังนี้

กลุ่มที่ 1 ค่าความดันซีสโตลิก (Systolic) น้อยกว่า 120 และค่าความดันไดแอสโตลิก (Diastolic) น้อยกว่า 80 แทนด้วย 1

กลุ่มที่ 2 ค่าความดันซีสโตลิก 120 ถึง 139 หรือ ค่าความดันไดแอสโตลิก 80 ถึง 89 แทนด้วย 2

กลุ่มที่ 3 ค่าความดันซีสโตลิก 140 ถึง 159 หรือ ค่าความดันไดแอสโตลิก 90 ถึง 99 แทนด้วย 3

กลุ่มที่ 4 ค่าความดันซีสโตลิกตั้งแต่ 160 ขึ้นไป หรือ ค่าความดันไดแอสโตลิกตั้งแต่ 100 ขึ้นไป แทนด้วย 4

1.6 น้ำตาลในเลือด ผู้วิจัยมีการจัดกลุ่มใหม่ 3 กลุ่มดังนี้

กลุ่มที่ 1 คือระดับน้ำตาลน้อยกว่า 110 แทนด้วย 1

กลุ่มที่ 2 คือระดับน้ำตาล 110 ถึง 125 แทนด้วย 2

กลุ่มที่ 3 คือระดับน้ำตาลตั้งแต่ 126 ขึ้นไป แทนด้วย 3

1.7 ระดับไขมันในเลือดค่าไตรกลีเซอไรด์ ผู้วิจัยมีการจัดกลุ่มใหม่ 4 กลุ่มดังนี้

กลุ่มที่ 1 ค่าไตรกลีเซอไรด์ น้อยกว่า 150 แทนด้วย 1

กลุ่มที่ 2 ค่าไตรกลีเซอไรด์ 150 ถึง 199 แทนด้วย 2

กลุ่มที่ 3 ค่าไตรกลีเซอไรด์ 200 ถึง 499 แทนด้วย 3

กลุ่มที่ 4 ค่าไตรกลีเซอไรด์ ตั้งแต่ 500 ขึ้นไป แทนด้วย 4

1.8 โรคเบาหวาน ข้อมูลมีค่าตั้งแต่ E10 ถึง E14 หมายความว่าเป็นโรคเบาหวาน และ Null หมายความว่าไม่เป็นโรคเบาหวาน ผู้วิจัยแทนค่าโรคเบาหวาน ตั้งแต่ E10 ถึง E14 ด้วย Y และ แทนค่าโรคเบาหวาน Null ด้วย N

Table 1 Represent Data

Field Name	Meaning	Raw Data	Representation
Gender	Gender , Sex	1	M
		2	F
Age	Age	60-69	1
		70-79	2
		>= 80	3
Bmi	Body Mass Index	0-18.5	1
		18.5-22.9	2
		23-24.9	3
		25-29.9	4
		>= 30	5
Smoke	Smoking	1	N
		2	Y
Bp	blood pressure SBP (Systolic) / DBP (Diastolic)	SBP < 120 and DBP < 80	1
		SBP 120-139 or DBP 80-89	2
		SBP 140-159 or DBP 90-99	3
		SBP >= 160 or DBP >= 100	4
FPG	Blood Sugar	< 110	1

Field Name	Meaning	Raw Data	Representation
		110 – 125	2
		>= 126	3
Tg	Triglycerides	< 150	1
		150 – 199	2
		200 – 499	3
		>= 500	4
Dm	Results Of Diabetes Mellitus	E10 – E14	Y
		null	N

2. ผลการเปรียบเทียบประสิทธิภาพแบบจำลองการพยากรณ์การเกิดโรคเบาหวาน

ผลการเปรียบเทียบประสิทธิภาพแบบจำลองการพยากรณ์การเกิดโรคเบาหวาน ผลการวัดประสิทธิภาพของแบบจำลองด้วยเทคนิคต้นไม้ตัดสินใจ เพื่อนบ้านใกล้ที่สุด และนาอีฟเบย์ด้วยค่าความถูกต้อง ค่าความแม่นยำ ค่าความระลึก และค่าความถ่วงดุล ได้ดังนี้ (Table 2)

Table 2 the Efficacy Results of the Diabetes Mellitus Model

Diabetes Mellitus Model	Accuracy	Precision	Recall	F-Measure
Naive Bayes	0.950	0.951	0.951	0.950
K-Nearest Neighbors	0.951	0.952	0.952	0.952
Decision Tree	0.952	0.952	0.952	0.952

จาก Table 2 อธิบายได้ว่า การวัดประสิทธิภาพในด้านความถูกต้องเทคนิคต้นไม้ตัดสินใจมีค่าเท่ากับ 0.952 มากกว่าเทคนิคเพื่อนบ้านใกล้ที่สุดมีค่าเท่ากับ 0.951 และเทคนิคนาอีฟเบย์มีค่าเท่ากับ 0.950 ด้านความแม่นยำเทคนิคต้นไม้ตัดสินใจและเทคนิคเพื่อนบ้านใกล้ที่สุดมีค่าเท่ากับ 0.952 มากกว่าเทคนิคนาอีฟเบย์มีค่าเท่ากับ 0.951 ด้านความระลึกเทคนิคต้นไม้ตัดสินใจและเทคนิคเพื่อนบ้านใกล้ที่สุดมีค่าเท่ากับ 0.952 มากกว่าเทคนิคนาอีฟเบย์มีค่า 0.951 ด้านความถ่วงดุลเทคนิคต้นไม้ตัดสินใจและเทคนิคเพื่อนบ้านใกล้ที่สุดมีค่าเท่ากับ 0.952 มากกว่าเทคนิคนาอีฟเบย์มีค่าเท่ากับ 0.950

ดังนั้นผู้วิจัยจึงเลือกเทคนิคต้นไม้ตัดสินใจเพราะมีค่าความถูกต้อง ค่าความแม่นยำ ค่าระลึกและค่าความถ่วงดุลสูงกว่าเทคนิคเพื่อนบ้านใกล้ที่สุดและเทคนิคนาอีฟเบย์

ผู้วิจัยนำเทคนิคต้นไม้ตัดสินใจไปสร้างแบบจำลองในการพยากรณ์โรคเบาหวานสำหรับกลุ่มผู้สูงอายุ แสดงดัง Figure 1

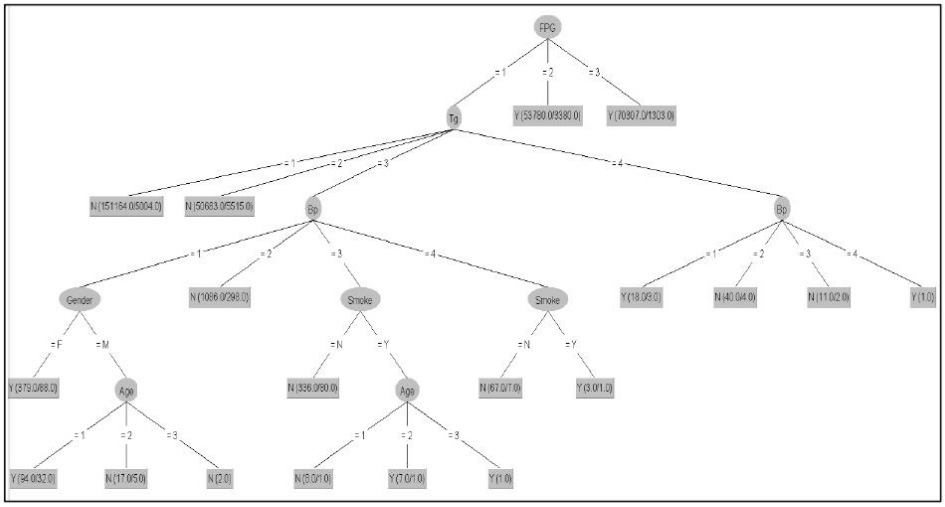


Figure 1 Decision Tree Model of Diabetes Mellitus

จาก Figure 1 อธิบายเป็นกฎการพยากรณ์ได้ 19 กฎ ดัง Table 3

Table 3 Explain the Relationship Rules of the Diabetes Mellitus Model

(Rule)	(Status)
1. FPG = 1 and Tg = 1	N
2. FPG = 1 and Tg = 2	N
3. FPG = 1 and Tg = 3 and Bp = 1 and Gender = F	Y
4. FPG = 1 and Tg = 3 and Bp = 1 and Gender = M and Age = 1	Y
5. FPG = 1 and Tg = 3 and Bp = 1 and Gender = M and Age = 2	N
6. FPG = 1 and Tg = 3 and Bp = 1 and Gender = M and Age = 3	N
7. FPG = 1 and Tg = 3 and Bp = 2	N
8. FPG = 1 and Tg = 3 and Bp = 3 and Smoke = N	N
9. FPG = 1 and Tg = 3 and Bp = 3 and Smoke = Y and Age = 1	N
10. FPG = 1 and Tg = 3 and Bp = 3 and Smoke = Y and Age = 2	Y
11. FPG = 1 and Tg = 3 and Bp = 3 and Smoke = Y and Age = 3	Y
12. FPG = 1 and Tg = 3 and Bp = 4 and Smoke = N	N
13. FPG = 1 and Tg = 3 and Bp = 4 and Smoke = Y	Y
14. FPG = 1 and Tg = 4 and Bp = 1	Y
15. FPG = 1 and Tg = 4 and Bp = 2	N

(Rule)	(Status)
16. FPG = 1 and Tg = 4 and Bp = 3	N
17. FPG = 1 and Tg = 4 and Bp = 4	Y
18. FPG = 2	Y
19. FPG = 3	Y

จาก Table 3 ผู้วิจัยวิเคราะห์และสรุปกฎได้ดังนี้

กฎข้อที่ 1 ถ้าค่าน้ำตาลในเลือดน้อยกว่า 110 ค่าไขมันในเลือดน้อยกว่าหรือเท่ากับ 199 แล้วสถานะไม่เป็นโรคเบาหวาน

กฎข้อที่ 2 ถ้าค่าน้ำตาลในเลือดน้อยกว่า 110 ค่าไขมันในเลือด 200 ถึง 499 ค่าความดันโลหิตค่าความดันซีสโตลิกน้อยกว่า 120 ค่าความดันไดแอสโตลิกน้อยกว่า 80 และเป็นเพศหญิง แล้วสถานะเป็นโรคเบาหวาน

กฎข้อที่ 3 เมื่อค่าน้ำตาลในเลือดน้อยกว่า 110 ค่าไขมันในเลือด 200 ถึง 499 ค่าความดันโลหิตค่าความดันซีสโตลิกน้อยกว่า 120 ค่าความดันไดแอสโตลิกน้อยกว่า 80 เป็นเพศชาย และ อายุ 60 ถึง 69 ปี แล้วสถานะเป็นโรคเบาหวาน

กฎข้อที่ 4 เมื่อค่าน้ำตาลในเลือดน้อยกว่า 110 ค่าไขมันในเลือด 200 ถึง 499 ค่าความดันโลหิตค่าความดันซีสโตลิกน้อยกว่า 120 ค่าความดันไดแอสโตลิกน้อยกว่า 80 เป็นเพศชาย และอายุ 70 ปีขึ้นไป แล้วสถานะไม่เป็นโรคเบาหวาน

กฎข้อที่ 5 เมื่อค่าน้ำตาลในเลือดน้อยกว่า 110 ค่าไขมันในเลือด 200 ถึง 499 และค่าความดันโลหิตค่าความดันซีสโตลิก 120 ถึง 139 หรือ ค่าความดันไดแอสโตลิก 80 ถึง 89 แล้วสถานะไม่เป็นโรคเบาหวาน

กฎข้อที่ 6 เมื่อค่าน้ำตาลในเลือดน้อยกว่า 110 ค่าไขมันในเลือด 200 ถึง 499 ค่าความดันโลหิตค่าความดันซีสโตลิก 140 ถึง 159 หรือ ค่าความดันไดแอสโตลิก 90 ถึง 99 และไม่สูบบุหรี่ แล้วสถานะไม่เป็นโรคเบาหวาน

กฎข้อที่ 7 เมื่อค่าน้ำตาลในเลือดน้อยกว่า 110 ค่าไขมันในเลือด 200 ถึง 499 ค่าความดันโลหิตค่าความดันซีสโตลิก 140 ถึง 159 หรือ ค่าความดันไดแอสโตลิก 90 ถึง 99 สูบบุหรี่ และอายุ 60 ถึง 69 ปี แล้วสถานะไม่เป็นโรคเบาหวาน

กฎข้อที่ 8 เมื่อค่าน้ำตาลในเลือดน้อยกว่า 110 ค่าไขมันในเลือด 200 ถึง 499 ค่าความดันโลหิตค่าความดันซีสโตลิก 140 ถึง 159 หรือ ค่าความดันไดแอสโตลิก 90 ถึง 99 สูบบุหรี่ และอายุมากกว่า 69 ปี แล้วสถานะเป็นโรคเบาหวาน

กฎข้อที่ 9 เมื่อค่าน้ำตาลในเลือดน้อยกว่า 110 ค่าไขมันในเลือด 200 ถึง 499 ค่าความดันโลหิตค่าความดันซีสโตลิก 140 ถึง 159 หรือ ค่าความดันไดแอสโตลิก 90 ถึง 99 สูบบุหรี่ และอายุมากกว่า 69 ปี แล้วสถานะเป็นโรคเบาหวาน

กฎข้อที่ 10 เมื่อค่าน้ำตาลในเลือดน้อยกว่า 110 ค่าไขมันในเลือด 200 ถึง 499 ค่าความดันโลหิตค่าความดันซีสโตลิก 160 ขึ้นไป หรือ ค่าความดันไดแอสโตลิก 100 ขึ้นไป แล้วสถานะไม่เป็นโรคเบาหวาน

กฎข้อที่ 11 เมื่อค่าน้ำตาลในเลือดน้อยกว่า 110 ค่าไขมันในเลือด 200 ถึง 499 ค่า ค่าความดันโลหิตค่าความดันซีสโตลิก 140 ถึง 159 หรือ ค่าความดันไดแอสโตลิก 90 ถึง 99 สูบบุหรี่ แล้วสถานะเป็นโรคเบาหวาน

กฎข้อที่ 12 เมื่อค่าน้ำตาลในเลือดน้อยกว่า 110 ค่าไขมันในเลือดมากกว่าหรือเท่ากับ 500 ค่าความดันโลหิตค่าความดันซีสโตลิกน้อยกว่า 120 ค่าความดันไดแอสโตลิกน้อยกว่า 80 หรือ ค่าความดันโลหิตค่าความดันซีสโตลิก 160 ขึ้นไป หรือ ค่าความดันไดแอสโตลิก 100 ขึ้นไป แล้วสถานะเป็นโรคเบาหวาน

กฎข้อที่ 13 เมื่อค่าน้ำตาลในเลือดน้อยกว่า 110 ค่าไขมันในเลือดมากกว่าหรือเท่ากับ 500 และค่าความดันโลหิตค่าความดันซีสโตลิก 120 ถึง 139 หรือ ค่าความดันไดแอสโตลิก 80 ถึง 89 หรือค่าความดันโลหิตค่าความดันซีสโตลิก 140 ถึง 159 หรือ ค่าความดันไดแอสโตลิก 90 ถึง 99 แล้วสถานะไม่เป็นโรคเบาหวาน

กฎข้อที่ 14 เมื่อค่าน้ำตาลในเลือดมากกว่าหรือเท่ากับ 110 แล้วสถานะเป็นโรคเบาหวาน

3. การประเมินประสิทธิภาพของแบบจำลองต้นไม้การตัดสินใจโดยผู้เชี่ยวชาญ

ผู้วิจัยนำแบบจำลองต้นไม้การตัดสินใจเพื่อพยากรณ์การเกิดโรคโรคเบาหวานในผู้สูงอายุให้ผู้เชี่ยวชาญด้านสาธารณสุขจำนวน 7 คน ประเมินประสิทธิภาพแบบจำลองกับการใช้ประกอบการทำงานของบุคลากรสาธารณสุข ด้วยการประชุมกลุ่ม ณ หอประชุมพุทธรักษา โรงพยาบาลนาหว้า จังหวัดนครพนม โดยเป็นการประเมินความครบถ้วนของปัจจัยที่ใช้ในการพยากรณ์และประเมินประสิทธิภาพของแบบจำลอง กับการปฏิบัติงาน ผลการประเมินแสดงดัง Table 3

Table 3 Summary of Model Efficiency Results by Public Health Experts

Opinion Issues	Average	Standard Deviation	Interpret
1. The Completeness of the Factors Used in Forecasting	4.43	0.79	High
2. Diabetes Mellitus Model			
2.1 Modeling Accuracy	4.43	0.79	High
2.2 Reliability of Prediction Results	4.00	0.58	High
2.3 Ability to Make Decisions	4.43	0.79	High
Mean Total Diabetes Mellitus Model	**4.32**	**0.74**	**High**

ผลการประเมินประสิทธิภาพแบบจำลองโดยผู้เชี่ยวชาญด้านสาธารณสุขพบว่าความครบถ้วนของปัจจัยที่ใช้ในการพยากรณ์ มีค่าเฉลี่ยเท่ากับ 4.43 ค่าความถูกต้องของแบบจำลอง ความน่าเชื่อถือของผลการทำนายและความสามารถนำไปประกอบการตัดสินใจ มีค่าเฉลี่ยเท่ากับ 4.43 4.40 และ 4.43 ตามลำดับ และมีค่าเฉลี่ยรวมเท่ากับ 4.32 (อยู่ในระดับมาก)

สรุปผล

ผลจากการเปรียบเทียบแบบจำลองพยากรณ์การเกิดโรคเบาหวานในผู้สูงอายุด้วยเทคนิคต้นไม้ตัดสินใจเพื่อนบ้านใกล้ที่สุด และนาอีฟเบย์ พบว่าเทคนิคต้นไม้ตัดสินใจมีค่าความถูกต้อง ค่าความแม่นยำ ค่าความระลึกและค่าความถ่วงดุลมีค่าเท่ากับ 0.952 สูงกว่าเทคนิคเพื่อนบ้านใกล้ที่สุดและเทคนิคนาอีฟเบย์

ผลจากการนำแบบจำลองพยากรณ์การเกิดโรคเบาหวานในผู้สูงอายุด้วยเทคนิคต้นไม้นำเสนอและประชุมกลุ่มกับผู้เชี่ยวชาญด้านสาธารณสุขมีค่าเฉลี่ยรวมของประสิทธิภาพแบบจำลองเท่ากับ 4.32 (อยู่ในระดับมาก)

อภิปรายผล

จากการสร้างและเปรียบเทียบแบบจำลองโรคเบาหวานในผู้สูงอายุ พบว่าเทคนิคต้นไม้ตัดสินใจมีประสิทธิภาพดีกว่าเทคนิคเพื่อนบ้านใกล้ที่สุดและเทคนิคนาอีฟเบย์ มีค่าเท่ากับร้อยละ 95.20 สอดคล้องกับงานวิจัยของวิธวินท์ และคณะ (2560) ทำวิทยานิพนธ์เรื่องการสร้างโมเดลทำนายโอกาสการกลับมารักษาตัวซ้ำของผู้ป่วยโรคเบาหวาน โดยใช้เทคนิคเหมืองข้อมูลและเปรียบเทียบประสิทธิภาพของโมเดลที่สร้างจากเทคนิคต้นไม้ตัดสินใจ เทคนิคเพื่อนบ้านใกล้ที่สุดและเทคนิคนาอีฟเบย์ ผลการวิจัยพบว่า โมเดลที่สร้างจากเทคนิคต้นไม้ตัดสินใจให้ค่าความถูกต้องมากที่สุดคิดเป็นร้อยละ 85.50 และงานวิจัยของนพรัตน์ (2557) ทำงานวิจัยเรื่องระบบสนับสนุนทางการแพทย์เพื่อวิเคราะห์โรคหลอดเลือดแดง โป่งพองโดยใช้เทคนิคเหมืองข้อมูลกรณีศึกษาการรักษาหลอดเลือดแดงโป่งพองผ่านสายสวน โดยใช้เทคนิคต้นไม้ตัดสินใจ เทคนิคนาอีฟเบย์ และเทคนิคโครงข่ายประสาทเทียม ซึ่งจากผลการทดลองของการศึกษาแสดงให้เห็นว่าโมเดลต้นไม้ตัดสินใจมีประสิทธิภาพค่าความถูกต้องสูงที่สุดมีค่าเท่ากับร้อยละ 98.10

กิตติกรรมประกาศ

การวิจัยนี้ผู้วิจัยขอขอบคุณโรงพยาบาลนาหว้า จังหวัดนครพนม ที่ให้ข้อมูลในการทำวิจัย และผู้เชี่ยวชาญด้านสาธารณสุข ที่ให้ความอนุเคราะห์ประเมินประสิทธิภาพแบบจำลอง

เอกสารอ้างอิง

กลุ่มยุทธศาสตร์และแผนงาน สำนักโรคไม่ติดต่อ. 2559. **รายงานประจำปี 2558**. โรงพิมพ์องค์การสงเคราะห์ทหารผ่านศึกในพระบรมราชูปถัมภ์. กรุงเทพฯ

เกรียงศักดิ์ รักภักดี. 2561. **การใช้เทคนิคเหมืองข้อมูลเพื่อพัฒนาฐานข้อมูลแนะนำแหล่งท่องเที่ยวทางธรรมชาติในจังหวัดอุบลราชธานี จากความต้องการและพฤติกรรมของนักท่องเที่ยว**. วิทยานิพนธ์, มหาวิทยาลัยราชภัฏอุบลราชธานี

ชฎิภกรน์ ทรายหมอ. 2557. **การพัฒนาแบบจำลองพยากรณ์ภาวะแทรกซ้อนของโรคอื่นในผู้ป่วยโรคเบาหวาน โดยใช้เทคนิคการทำเหมืองข้อมูล : กรณีโรงพยาบาลปทุมธานี**. วิทยานิพนธ์, มหาวิทยาลัยสุโขทัยธรรมาธิราช

นพรัตน์ พจน์จิราภรณ์. 2557. **ระบบสนับสนุนทางการแพทย์เพื่อวิเคราะห์โรคหลอดเลือดแดงโป่งพองโดยใช้เทคนิคเหมืองข้อมูล กรณีศึกษาการรักษาหลอดเลือดแดงโป่งพองผ่านสายสวน**. ปัญหาพิเศษ. มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

วิธวินท์ แสงมณี,วีระวุฒ รัตนเจริญเลิศ,ณัฏฐภพ โพธิรัชต์ และเพียงฤทัย หนูสวัสดิ์. 2560. **การสร้างโมเดลทำนายโอกาสการกลับมารักษาตัวซ้ำของผู้ป่วยโรคเบาหวาน โดยใช้เทคนิคการทำเหมืองข้อมูล**. วิทยานิพนธ์, มหาวิทยาลัยเทคโนโลยีราชมงคลรัตนโกสินทร์

สายชล สินสมบูรณ์ทอง. 2560. **การทำเหมืองข้อมูล เล่ม 1 : การค้นหาความรู้จากข้อมูล พิมพ์ครั้งที่ 2 ฉบับปรับปรุง.** จามจุรีโปรดักส์, กรุงเทพฯ

สำนักงานกองทุนสนับสนุนการสร้างเสริมสุขภาพ, 2561. **รายงานประจำปี 2561 สำนักงานกองทุนสนับสนุนการสร้างเสริมสุขภาพ (สสส.)**. สำนักงานกองทุนสนับสนุนการสร้างเสริมสุขภาพ, กรุงเทพฯ

สมเกียรติ โพธิสัตย์,เนติมา คูนีย์,รัชนีบูลย์ อุดมชัยรัตน์,พรทิพย์ ปรีชาไชยวิทย์,สุรีพร คนละเอียด,ศุภลักษณ์ มิรัตนไพร และเกตุแก้ว สายน้ำเย็น. 2557. **การทบทวนวรรณกรรม : สถานการณ์ปัจจุบันและรูปแบบการบริการด้านโรคไม่ติดต่อเรื้อรัง**. สถาบันวิจัยและประเมินเทคโนโลยีทางการแพทย์ กรมการแพทย์ กระทรวงสาธารณสุข

เสกสรรค์ วิลัยลักษณ์. 2558. **การใช้เทคนิคการทำเหมืองข้อมูลเพื่อพยากรณ์ผลการเรียนของนักเรียนโรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน ศูนย์วิจัยและพัฒนาการศึกษา**. วิทยานิพนธ์, มหาวิทยาลัยสุโขทัยธรรมาธิราช

เอกสิทธิ์ พัชรวงศ์ศักดา. 2557. **การวิเคราะห์ข้อมูลด้วยเทคนิคดาต้าไมน์นิ่ง เบื้องต้น**. เอเชีย ดิจิตอลการพิมพ์ จำกัด, กรุงเทพฯ

ประวัติย่อของผู้วิจัย

ประวัติย่อของผู้วิจัย

ชื่อ–สกุล	ศรัณยู ชูทองรัตนะ
วันเดือนปีเกิด	วันที่ 6 เดือน กรกฎาคม พ.ศ. 2532
สถานที่อยู่ปัจจุบัน	บ้านเลขที่ 207 หมู่ 4 ถนนโชคเจริญ ตำบลนาหว้า อำเภอนาหว้า จังหวัดนครพนม
ตำแหน่งปัจจุบัน	นักวิชาการคอมพิวเตอร์ปฏิบัติการ
สถานที่ทำงาน	โรงพยาบาลนาหว้า อำเภอนาหว้า จังหวัดนครพนม
ประวัติการศึกษา	
พ.ศ. 2544	ชั้นประถมศึกษาปีที่ 6 โรงเรียนบ้านนาหว้า อำเภอนาหว้า จังหวันครพนม
พ.ศ. 2550	ชั้นมัธยมศึกษาปีที่ 6 โรงเรียนสกลราชวิทยานุกูล อำเภอเมือง จังหวัดสกลนคร
พ.ศ. 2554	ปริญญาวิทยาศาสตรบัณฑิต (วท.บ.) สาขาวิชาเทคโนโลยีสารสนเทศ มหาวิทยาลัยเกษตรศาสตร์
ประวัติการทำงาน	
พ.ศ. 2555–2565	นักวิชาการคอมพิวเตอร์ปฏิบัติการ โรงพยาบาลนาหว้า อำเภอนาหว้า จังหวัดนครพนม