การวิเคราะห์ปัจจัยที่ก่อให้เกิดโรคเบาหวานด้วยเทคนิคการทำเหมืองข้อมูล

การศึกษาค้นคว้าอิสระ

ของ

กฤตกนก ศรีพิมพ์สอ

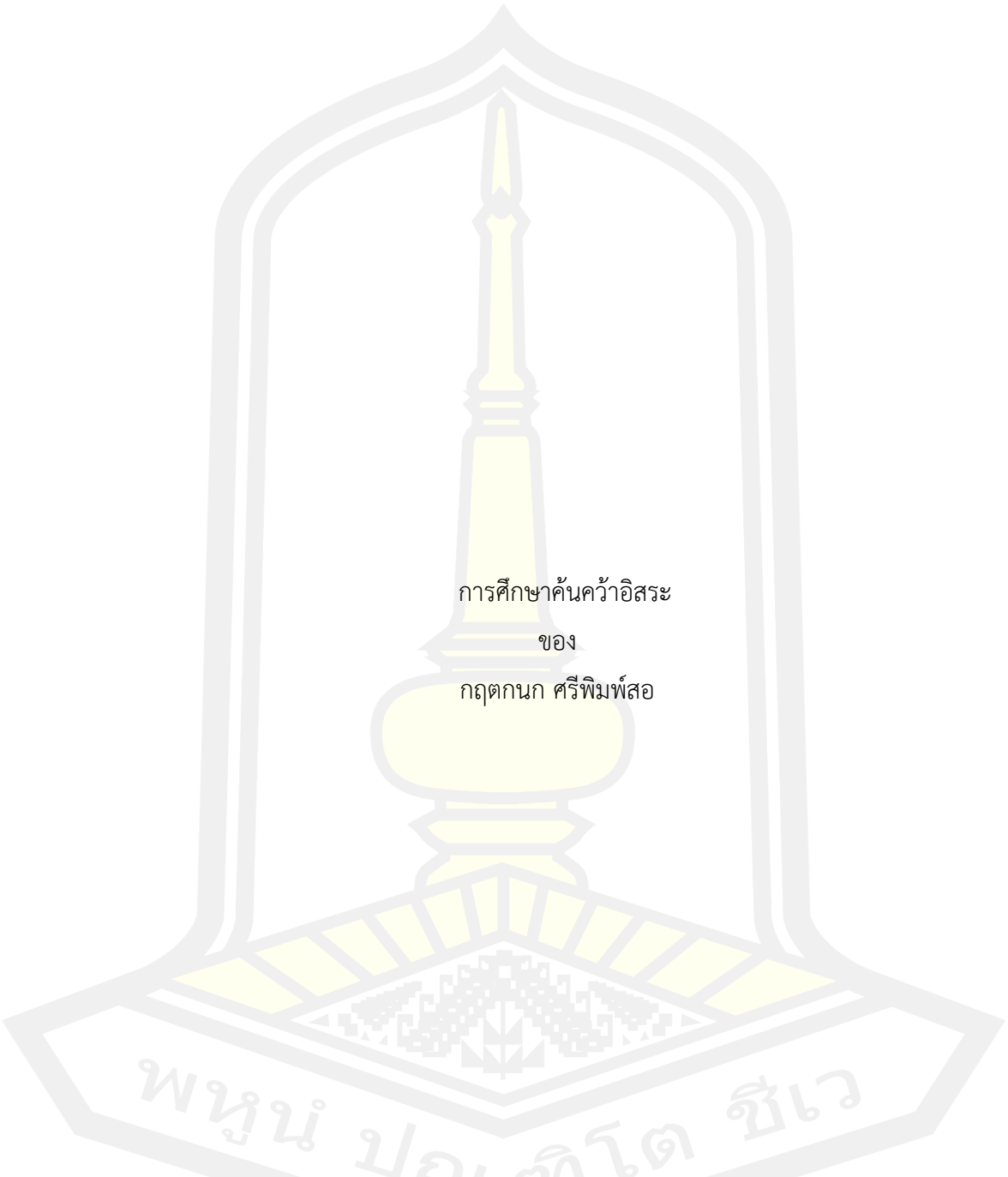
เสนอต่อมหาวิทยาลัยมหาสารคาม เพื่อเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

ปริญญาบริหารธุรกิจมหาบัณฑิต สาขาวิชาบริหารธุรกิจและนวัตกรรมดิจิทัล

กุมภาพันธ์ 2566

ลิขสิทธิ์เป็นของมหาวิทยาลัยมหาสารคาม

การวิเคราะห์ปัจจัยที่ก่อให้เกิดโรคเบาหวานด้วยเทคนิคการทำเหมืองข้อมูล



การศึกษาค้นคว้าอิสระ

ของ

กฤตกนก ศรีพิมพ์สอ

เสนอต่อมหาวิทยาลัยมหาสารคาม เพื่อเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

ปริญญาบริหารธุรกิจมหาบัณฑิต สาขาวิชาบริหารธุรกิจและนวัตกรรมดิจิทัล

กุมภาพันธ์ 2566

ลิขสิทธิ์เป็นของมหาวิทยาลัยมหาสารคาม

Analysis of Factors Affecting Diabetes Mellitus by Using Data Mining Techniques

Kitkanog Sripimsor

A Independent Study Submitted in Partial Fulfillment of Requirements

for Master of Business Administration (Business Administration and Digital Innovation)

February 2023

Copyright of Mahasarakham University

คณะกรรมการสอบการศึกษาค้นคว้าอิสระ ได้พิจารณาการศึกษาค้นคว้าอิสระของนายกฤตกนก ศรีพิมพ์สอ แล้วเห็นสมควรรับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาบริหารธุรกิจมหาบัณฑิต สาขาวิชาบริหารธุรกิจและนวัตกรรมดิจิทัล ของมหาวิทยาลัยมหาสารคาม

คณะกรรมการสอบวิทยานิพนธ์

.....	ประธานกรรมการ
(ดร. การันต์ กิจระการ)	
.....	อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(รศ. ดร. กิตติพล วิแสง)	
.....	กรรมการ
(ดร. พงศธร ตันตระบัณฑิตย์)	
.....	กรรมการผู้ทรงคุณวุฒิภายนอก
(รศ. ดร. ธีระวัฒน์ เจริญราษฎร์)	

มหาวิทยาลัยอนุมัติให้รับวิทยานิพนธ์ฉบับนี้ เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญา บริหารธุรกิจมหาบัณฑิต สาขาวิชาบริหารธุรกิจและนวัตกรรมดิจิทัล ของมหาวิทยาลัยมหาสารคาม

.....	.....
(อ. ดร. ชลธิชา ธรรมวิญญู)	(รศ. ดร. กริสน์ ชัยมูล)
คณบดีคณะการบัญชีและการจัดการ	คณบดีบัณฑิตวิทยาลัย

ชื่อเรื่อง การวิเคราะห์ปัจจัยที่ก่อให้เกิดโรคเบาหวานด้วยเทคนิคการทำเหมืองข้อมูล
ผู้วิจัย กฤตกนก ศรีพิมพ์สอ
อาจารย์ที่ปรึกษา รองศาสตราจารย์ ดร. กิตติพล วิแสง
ปริญญา บริหารธุรกิจมหาบัณฑิต **สาขาวิชา** บริหารธุรกิจและนวัตกรรมดิจิทัล
มหาวิทยาลัย มหาวิทยาลัยมหาสารคาม **ปีที่พิมพ์** 2566

บทคัดย่อ

งานวิจัยนี้ได้ประยุกต์ใช้เทคนิคการทำเหมืองข้อมูลเพื่อพยากรณ์ผู้ป่วยโรคเบาหวานด้วยเทคนิคเพอร์เซ็ปตรอนแบบหลายชั้น (Multilayer perceptron) เทคนิคนาอีฟ เบย์ (Naïve bayes) เทคนิคโลจิสติก (Logistic Regression) เทคนิคป่าสุ่ม (Random forest) และเทคนิคต้นไม้ตัดสินใจ (Decision Trees) เพื่อสร้างแบบจำลองในการพยากรณ์ผู้ป่วยและนำค่าวัดประสิทธิภาพของการจำแนกประเภทข้อมูลมาเปรียบเทียบ โดยค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) ค่าความถ่วงดุล (F-measure) และการดูเส้นกราฟ ROC (Receiver operating characteristic; ROC) ที่ให้ค่ามากที่สุด ผลของการวิจัยพบว่า เทคนิคป่าสุ่ม (Random Forest) ให้ค่าความถูกต้องมากที่สุดที่อยู่ที่ 99.75% มีค่าความแม่นยำ (Precision) ในแบบที่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 1.000% และในแบบที่ไม่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 0.997% ที่ค่าความระลึก (Recall) ในแบบที่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 0.990% และในแบบที่ไม่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 1.000% ค่าความถ่วงดุล (F-measure) ในแบบที่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 0.995% และในแบบที่ไม่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 0.998% และการดูเส้นกราฟ ROC ในแบบที่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 1.000% และในแบบที่ไม่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 1.000% สามารถนำผลลัพธ์ที่ได้ไปใช้ในการประกอบการรักษาผู้ป่วยโรคเบาหวานต่อไปในอนาคต

คำสำคัญ : โรคเบาหวาน, การทำเหมืองข้อมูล, การควบคุมระดับน้ำตาลในเลือด, ผู้สูงอายุ, ความแม่นยำ

TITLE Analysis of Factors Affecting Diabetes Mellitus by Using Data Mining Techniques

AUTHOR Kitkanog Sripimsor

ADVISORS Associate Professor Kittipol Wisaeng , Ph.D.

DEGREE Master of Business Administration **MAJOR** Business Administration and Digital Innovation

UNIVERSITY Mahasarakham University **YEAR** 2023

ABSTRACT

This study used data mining techniques to predict diabetes patients by employing multilayer perceptrons, Nave Bayes, logistic regression, random forests, and decision trees to generate a patient forecasting model and compare the performance measures of data classification with a confusion matrix containing accuracy, precision, recall, and balance. (F-measure) and ROC (Receiver Operating Characteristic; ROC) curve that provides the greatest benefit. According to the findings of the study, the random forest technique provides the highest accuracy (99.75% with precision). The recall value for the risk of diabetes was 0.990%, while the risk of not having diabetes was 0.997%. The F-measure of the diabetes-risk model was 0.995%, while the F-measure of the no-risk model was 0.995%. 0.998%, and the ROC curve was 1.000% for both the diabetes and non-diabetes risks. In the future, the findings can be used to predict diabetes patients.

Keyword : Diabetes, Data Mining, Blood Sugar Control, Elderly, Accuracy

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้ สำเร็จลงได้ด้วยอาศัยผู้ที่มีส่วนเกี่ยวข้องเป็นจำนวนมากที่ช่วยสนับสนุนส่งเสริมให้ผู้วิจัยสามารถทำวิทยานิพนธ์ฉบับนี้สำเร็จได้ด้วยดี ในการนี้ผู้วิจัยขอขอบคุณผู้มีส่วนเกี่ยวข้องดังต่อไปนี้

ขอขอบคุณรองศาสตราจารย์ ดร.กิตติพล วิแสง ที่เป็นที่ปรึกษาในการทำงานวิจัยในครั้งนี้ ขอขอบคุณผู้ทรงคุณวุฒิที่ตรวจสอบความเที่ยงตรงของเครื่องมือแบบสอบถาม ขอขอบคุณ ผู้เขียนตำรา หนังสือ เอกสารทางวิชาการ และเว็บไซต์ต่างๆ ที่ผู้วิจัยได้นำมาอ้างอิง รวมทั้งกลุ่มตัวอย่างที่ให้ความร่วมมือเป็นอย่างดีในการตอบแบบสอบถาม ขอขอบคุณเพื่อนทุกคนที่ให้กำลังใจและให้ความช่วยเหลือทำให้การศึกษาวิจัยนี้สำเร็จได้ด้วยดี

กฤตกนก ศรีพิมพ์สอ

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ง

บทคัดย่อภาษาอังกฤษ ..... จ

กิตติกรรมประกาศ ..... ฉ

สารบัญ ..... ช

สารบัญตาราง ..... ฌ

สารบัญภาพประกอบ ..... ฎ

บทที่ 1 บทนำ 1

1.1 ภูมิหลัง 1

1.2 ความมุ่งหมายของการวิจัย 3

1.3 การใช้เครื่องมือการทำเหมืองข้อมูล 3

1.4 กรอบแนวคิดที่ใช้ในการวิจัย 4

1.5 สมมุติฐานของการวิจัย 5

1.6 ขอบเขตของการวิจัย 5

1.7 การวิเคราะห์ข้อมูล 6

1.8 ความสำคัญของการวิจัย 7

1.9 นิยามศัพท์ 7

บทที่ 2 วรรณกรรมและผลงานวิจัยที่เกี่ยวข้อง 9

2.1 ข้อมูลทั่วไปของโรคเบาหวาน 9

2.2 ปัจจัยที่ส่งผลต่อการก่อให้เกิดโรคเบาหวาน 11

2.3 เครื่องมือการคัดกรองโรคเบาหวาน 12

2.4 แนวคิดเทคนิคการทำเหมืองข้อมูล 16

2.5 งานวิจัยที่เกี่ยวข้อง 24
บทที่ 3 วิธีดำเนินการวิจัย 27
3.1 ประชากรและกลุ่มตัวอย่าง 27
3.2 เครื่องมือที่ใช้ในการเก็บรวบรวมข้อมูล 29
3.3 การสร้างและพัฒนาเครื่องมือ 30
3.4 การเก็บรวบรวมข้อมูลและการจัดการข้อมูล 31
3.5 ขั้นตอนการเก็บรวบรวมข้อมูล 31
3.6 การจัดกระทำกับข้อมูลและการวิเคราะห์ข้อมูล 32
3.7 การเตรียมข้อมูล (Data Preparation) 34
3.8 วิธีดำเนินการวิจัย 35
3.9 ขั้นตอนการวิเคราะห์ข้อมูล 36
บทที่ 4 การทดสอบและอภิปรายผล 40
4.1 ข้อมูลที่ใช้ในการทดสอบ 40
4.2 ขั้นตอนการเตรียมข้อมูล 43
4.3 การทดสอบและการคัดเลือกอัลกอริทึม 58
4.4 การวัดประสิทธิภาพของเครื่องมือ 58
4.5 วิธีการดำเนินการวิจัย 62
บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ 67
ข้อเสนอแนะ 67
บรรณานุกรม 69
ประวัติผู้เขียน 78

สารบัญตาราง

หน้า

ตาราง 1 ผลลัพธ์ของแต่ละเครื่องมือการพยากรณ์โรคเบาหวาน 13

ตาราง 2 ผลลัพธ์การประมวลผลของอัลกอลิทึม 13

ตาราง 3 ผลลัพธ์ของความแม่นยำของอัลกอลิทึม 14

ตาราง 4 ตัวอย่างการนำเทคนิคเหมืองข้อมูลมาใช้ในทางการแพทย์ 14

ตาราง *5* การประยุกต์เทคนิคเหมืองข้อมูลในทางการแพทย์ 24

ตาราง 6 ข้อมูลกลุ่มผู้ป่วยที่เข้ารับการคัดกรองโรคเบาหวานและกลุ่มผู้ป่วยที่เป็นโรคเบาหวาน ... 28

ตาราง 7 คุณลักษณะและรายละเอียดของข้อมูล 32

ตาราง 8 ปัจจัยสภาวะความเสี่ยงสำคัญต่อการเกิดโรคเบาหวาน 34

ตาราง 9 การนำข้อมูลปัจจัยและสภาวะความเสี่ยงของโรคเบาหวานแทนค่าด้วยตัวเลขก่อนทำ ... 36

ตาราง 10 ตัวอย่างการแทนค่าข้อมูลปัจจัยเสี่ยงของโรคเบาหวานด้วยตัวเลข 36

ตาราง 11 ข้อมูลปัจจัยเสี่ยงชุดที่ 1 37

ตาราง 12 ข้อมูลปัจจัยเสี่ยงชุดที่ 2 เพิ่มตัวแปรความยาวเส้นรอบเอวตาม การสูบบุหรี่ การดื่มสุรา ตามคำแนะนำจากแพทย์ผู้เชี่ยวชาญ 37

ตาราง 13 ชุดข้อมูลนำเข้าและชุดข้อมูลเป้าหมายจากข้อมูลชุดที่ 1 และข้อมูลชุดที่ 2 ที่ถูกแปลงค่าให้อยู่ในรูปแบบเดียวกันก่อนการสอนการเรียนรู้ 38

ตาราง 14 อธิบายผลการแปลงค่าแต่ละตัวแปรดังนี้ 38

ตาราง 15 ตารางแสดงข้อมูลทั่วไปของผู้ป่วย 40

ตาราง 16 แสดงข้อมูลพฤติกรรมการรับประทานอาหาร 42

ตาราง 17 แสดงข้อมูลพฤติกรรมด้านการออกกำลังกาย 42

ตาราง 18 แสดงข้อมูลพฤติกรรมด้านการรับประทานยา 42

ตาราง 19 แสดงข้อมูลพฤติกรรมด้านการสูบบุหรี่ 42

ตาราง 20 แสดงข้อมูลด้านการต้านทานการเกิดโรค 43

ตาราง 21 แสดงข้อมูลด้านการต้านทานความชุกของโรค 43
ตาราง 22 ประเภทของข้อมูลที่ใช้ในการวิเคราะห์โรคเบาหวาน 2 ด้าน 45
ตาราง 23 แสดงการแบ่งกลุ่มข้อมูลในประเภทที่ 2 เพื่อการพยากรณ์โรคเบาหวาน 45
ตาราง 24 ข้อมูลเพศของผู้ตอบแบบสอบถาม 45
ตาราง 25 ข้อมูลอายุของผู้ตอบแบบสอบถาม 46
ตาราง 26 ข้อมูลระดับการศึกษาของผู้ตอบแบบสอบถาม 47
ตาราง 27 ข้อมูลนับถือศาสนาของผู้ตอบแบบสอบถาม 48
ตาราง 28 ข้อมูลสถานภาพของผู้ตอบแบบสอบถาม 49
ตาราง 29 ข้อมูลดัชนีมวลกายของผู้ตอบแบบสอบถาม 50
ตาราง 30 ข้อมูลความยาวเส้นรอบเอวของผู้ตอบแบบสอบถาม 51
ตาราง 31 ข้อมูลประวัติเป็นความดันเลือดสูงของผู้ตอบแบบสอบถาม 52
ตาราง 32 ข้อมูลประวัติการดื่มสุราของผู้ตอบแบบสอบถาม 53
ตาราง 33 ข้อมูลประวัติการสูบบุหรี่ของผู้ตอบแบบสอบถาม 54
ตาราง 34 ข้อมูลครอบครัวหรือญาติสนิทมีประวัติการเป็นเบาหวานของผู้ตอบแบบสอบถาม 55
ตาราง 35 ข้อมูลส่วนสูงของผู้ตอบแบบสอบถาม 56
ตาราง 36 ข้อมูลน้ำหนักของผู้ตอบแบบสอบถาม 57
ตาราง 37 ตารางเมทริกซ์ความสับสน (Confusion Matrix) 59
ตาราง 38 ตารางแสดงค่าการอ่านผลของ AUC 60
ตาราง 39 ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยวิธีเพอร์เซ็ปตรอนหลายชั้น 62
ตาราง 40 ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยวิธีนาอีฟ เบย์ 63
ตาราง 41 ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยวิธีโลจิสติกพหุกลุ่ม 64
ตาราง 42 ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยวิธีเทคนิคป่าสุ่ม 64
ตาราง 43 ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ 65
ตาราง 44 แสดงผลลัพธ์เปรียบเทียบทั้ง 5 อัลกอริทึม 66

สารบัญภาพประกอบ

หน้า

ภาพประกอบ 1 กรอบแนวคิดในการวิจัย 4
ภาพประกอบ 2 ขั้นตอนการรวบรวม นำเข้าข้อมูล วิเคราะห์ข้อมูลและเปรียบเทียบประสิทธิภาพ 5
ภาพประกอบ 3 แสดงโครงสร้าง Decision Tree 23
ภาพประกอบ 4 แผนภูมิแสดงจำนวนผู้ป่วยเข้ารับการคัดกรองโรคเบาหวานและผู้ป่วยที่เป็นโรคเบาหวาน 29
ภาพประกอบ 5 ตัวอย่างของการปรับโครงสร้างข้อมูลเพื่อลดความซ้ำซ้อนของข้อมูลและจัดลำดับความสำคัญของข้อมูล 44
ภาพประกอบ 6 แผนภูมิข้อมูลเพศของผู้ตอบแบบสอบถาม 46
ภาพประกอบ 7 แผนภูมิข้อมูลอายุของผู้ตอบแบบสอบถาม 47
ภาพประกอบ 8 แผนภูมิข้อมูลระดับการศึกษาของผู้ตอบแบบสอบถาม 48
ภาพประกอบ 9 แผนภูมิข้อมูลนับถือศาสนาของผู้ตอบแบบสอบถาม 49
ภาพประกอบ 10 แผนภูมิข้อมูลสถานภาพของผู้ตอบแบบสอบถาม 50
ภาพประกอบ 11 แผนภูมิข้อมูลดัชนีมวลกายของผู้ตอบแบบสอบถาม 51
ภาพประกอบ 12 แผนภูมิข้อมูลความยาวเส้นรอบเอวของผู้ตอบแบบสอบถาม 52
ภาพประกอบ 13 แผนภูมิข้อมูลประวัติเป็นความดันเลือดสูงของผู้ตอบแบบสอบถาม 53
ภาพประกอบ 14 แผนภูมิข้อมูลประวัติการดื่มสุราของผู้ตอบแบบสอบถาม 54
ภาพประกอบ 15 แผนภูมิข้อมูลประวัติการสูบบุหรี่ของผู้ตอบแบบสอบถาม 55
ภาพประกอบ 16 แผนภูมิข้อมูลครอบครัวหรือญาติสนิทมีประวัติการเป็นเบาหวานของผู้ตอบแบบสอบถาม 56
ภาพประกอบ 17 แผนภูมิข้อมูลส่วนสูงของผู้ตอบแบบสอบถาม 57
ภาพประกอบ 18 แผนภูมิข้อมูลน้ำหนักของผู้ตอบแบบสอบถาม 58

ภาพประกอบ 19 แสดงกราฟ ROC และพื้นที่ใต้กราฟเพื่อใช้ในการวัดประสิทธิภาพของโมเดลการจำแนก ที่ AUC=0.8597 61

ภาพประกอบ 20 แสดงกราฟ ROC และพื้นที่ใต้กราฟเพื่อใช้ในการวัดประสิทธิภาพของโมเดลการจำแนก ที่ AUC=0.9998 61

บทที่ 1
บทนำ

1.1 ภูมิหลัง

เบาหวานเป็นโรคเรื้อรังที่เป็นปัญหาสำคัญทางด้านสาธารณสุขของโลก มีการประเมินว่ามีผู้ป่วยเบาหวาน รวม 425 ล้านรายทั่วโลก แบ่งเป็นกลุ่มผู้สูงอายุ 65 ปีขึ้นไป จำนวน 98 ล้านราย และช่วงอายุ 20-64 ปี จำนวน 327 ล้านราย และคาดการณ์ว่าในปี พ.ศ. 2588 จะมีผู้ป่วยเบาหวาน จำนวน 629 ล้านรายทั่วโลก โดยแบ่งเป็นผู้สูงอายุ 65 ปีขึ้นไป จำนวน 191 ล้านราย หรือมีอัตราการเพิ่มอยู่ที่ร้อยละ 94.8 และในช่วงอายุ 20-64 ปี จำนวน 438 ล้านราย หรือมีอัตราการเพิ่มอยู่ร้อยละ 33.9 ((International Diabetes Federation) สมาพันธ์เบาหวานนานาชาติ, 2560)

โรคเบาหวานเป็นโรคเรื้อรังที่เกิดจากความผิดปกติของร่างกายที่มีการผลิตฮอร์โมนอินซูลินไม่เพียงพอหรือร่างกายไม่สามารถนำน้ำตาลไปใช้งานได้อย่างมีประสิทธิภาพ ทำให้ระดับน้ำตาลในเลือดสูง ซึ่งการเกิดภาวะน้ำตาลในเลือดสูงเป็นระยะเวลานาน ส่งผลให้อวัยวะเสื่อมสมรรถภาพ และทำงานล้มเหลว เป็นเหตุให้เกิดภาวะแทรกซ้อนที่ตา ไต หลอดเลือดหัวใจและหลอดเลือดสมอง รวมถึงเป็นแผลง่ายหายยาก ชาปลายมือปลายเท้า (กองโรคไม่ติดต่อ กรมควบคุมโรค กระทรวงสาธารณสุข, 2564) ส่งผลกระทบต่อสุขภาพในระยะยาวและเป็นสาเหตุการตายอันดับ 10 ของวัยผู้ใหญ่ ผู้ป่วยเบาหวานมีอัตราการเพิ่มขึ้นอย่างต่อเนื่องทุกปี อันเนื่องจากวิถีชีวิตที่เปลี่ยนแปลงไปและพฤติกรรมการบริโภคที่ไม่ดีต่อสุขภาพ ผู้ป่วยโรคเบาหวานที่ไม่สามารถควบคุมระดับน้ำตาลในเลือดได้จะทำให้เกิดภาวะแทรกซ้อนของระบบหัวใจและหลอดเลือด ส่งผลต่อการเสียชีวิตและทุพพลภาพ เกิดภาระค่าใช้จ่ายด้านการรักษาพยาบาลที่เพิ่มขึ้น (ฉวีวรรณ ศรีดาวเรือง และจิราพร วรวงศ์, 2565)

สาเหตุที่แท้จริงของโรคที่เกิดขึ้นยังไม่ทราบแน่ชัด แต่จากการศึกษามีปัจจัยหลายอย่างที่เกี่ยวข้องทั้งสาเหตุจากพันธุกรรม สิ่งแวดล้อมและพฤติกรรมสุขภาพ สำหรับด้านพฤติกรรมสุขภาพพบว่าสาเหตุสำคัญมาจากพฤติกรรมที่ไม่เหมาะสม เช่น พฤติกรรมการกินที่เปลี่ยนไปโดยมักรับประทานอาหารที่มีรสหวาน มัน เค็ม ขณะที่มีส่วนประกอบของผัก ผลไม้น้อย และพฤติกรรมที่มีการเคลื่อนไหวร่างกายน้อยลง ขาดการออกกำลังกาย ภาวะ น้ำหนักเกิน ภาวะอ้วนที่เพิ่มขึ้น การมีความเครียดเรื้อรังรวมทั้งการบริโภค เครื่องดื่มแอลกอฮอล์ปริมาณเพิ่มขึ้น การสูบบุหรี่ ล้วนเป็นสาเหตุของการเกิดโรคทั้งสิ้น ดังนั้นการลดพฤติกรรมเสี่ยงและปรับพฤติกรรมในด้านการออกกำลังกาย การรับประทานอาหาร การจัดการด้านอารมณ์และจิตใจ และการหลีกเลี่ยงปัจจัยเสี่ยงทางสุขภาพในการดำเนินชีวิต นอกจากจะช่วยป้องกันการเกิดโรคเบาหวานแล้ว ยังช่วยให้ลดภาระค่าใช้จ่าย ลดความทุกข์ทรมานของผู้ป่วยที่เป็นโรคอย่างยาวนาน และส่งเสริมการมีคุณภาพชีวิตที่ดีขึ้นด้วย (จรรยา สุนทรปกาสิต และขณัฏฐา เพียงแก้ว, 2561)

สำหรับประเทศไทยพบอุบัติการณ์โรคเบาหวานมีแนวโน้มเพิ่มขึ้นอย่างต่อเนื่อง มีผู้ป่วยรายใหม่เพิ่มขึ้นประมาณ 3 แสนคนต่อปี และมีผู้ป่วยโรคเบาหวานอยู่ในระบบทะเบียน 3.2 ล้านคน ของกระทรวงสาธารณสุข ก่อให้เกิดการสูญเสียค่าใช้จ่ายในการรักษาด้านสาธารณสุขอย่างมหาศาล เฉพาะเบาหวานเพียงโรคเดียวทำให้สูญเสียค่าใช้จ่ายในการรักษาเฉลี่ยสูงถึง 47,596 ล้านบาทต่อปี

และหากรวมอีก 3 โรค คือ โรคความดันโลหิตสูง โรคหัวใจ และโรคหลอดเลือดสมอง ทำให้ภาครัฐสูญเสียงบประมาณในการรักษารวมกันสูงถึง 302,367 ล้านบาทต่อปี โรคเบาหวานยังคงเป็นสาเหตุหลักที่ก่อให้เกิดโรคอื่นๆ ในกลุ่มโรค NCDs อีกมากมาย อาทิ โรคหัวใจ โรคหลอดเลือดสมอง โรคความดันโลหิตสูง และโรคไต ฯลฯ

จากสถานการณ์โรคเบาหวานในจังหวัดมหาสารคาม ตั้งแต่ปี พ.ศ. 2556 - 2561 มีจำนวนผู้ป่วยโรคเบาหวานเพิ่มมากขึ้น จากปี 2554 มีจำนวนในผู้ป่วยเบาหวาน 9,409 คน เมื่อเวลาผ่านไป 5 ปีพบว่ามีจำนวนผู้ป่วยเบาหวานเพิ่มขึ้นเป็น 13,431 คน ในปี พ.ศ. 2558 (กองโรคไม่ติดต่อ กรมควบคุมโรค กระทรวงสาธารณสุข, 2564) สำหรับอำเภอกันทรวิชัย จังหวัดมหาสารคาม มีจำนวน 183 หมู่บ้าน 28,888 ครัวเรือน มีจำนวนประชากร 86,654 คน แบ่งเป็นชาย จำนวน 40,406 คน และหญิง จำนวน 45,748 คน แบ่งการปกครองเป็น 10 ตำบล มีสถานพยาบาล 17 แห่ง คือ โรงพยาบาลกันทรวิชัย 1 แห่ง สำนักงานสาธารณสุขอำเภอ 1 แห่ง ศูนย์บริการทางการแพทย์ 1 แห่ง และโรงพยาบาลส่งเสริมสุขภาพตำบลในเครือข่าย 14 แห่ง (กระทรวงสาธารณสุข กรมอนามัย, 2565) ปี พ.ศ. 2560 - 2563 อัตราป่วยด้วยโรคเบาหวาน 1,033.9, 2,141.8, 3,548.5, และ 3,751.7 ต่อประชากรแสนคน ตามลำดับ พบอัตราความชุกของโรคเบาหวานในประชากรที่มีอายุตั้งแต่ 35 ปีขึ้นไป ร้อยละ 4.90, 4.96, 5.44 และ 7.30 ตามลำดับ (สถิติสุขภาพคคนไทย, 2563) ซึ่งจะพบว่ามีผู้ป่วยเบาหวานรายใหม่เพิ่มขึ้นทุกปี

จากการทบทวนวรรณกรรมพบว่า ความรู้เรื่องโรคเบาหวานมีความสัมพันธ์ทางบวกกับพฤติกรรมการจัดการตนเองและสามารถทำนายพฤติกรรมการจัดการตนเองได้ในกลุ่มผู้สูงอายุที่มีภาวะเสี่ยงต่อการเกิดโรคหัวใจและหลอดเลือด การรับรู้สมรรถนะแห่งตนเป็นความเชื่อมั่นของบุคคลในความสามารถของตนเองเกี่ยวกับการปฏิบัติตนที่มีความสัมพันธ์ทางบวกกับพฤติกรรมการจัดการตนเองและสามารถทำนายพฤติกรรมการจัดการตนเองได้ การกำกับตนเองซึ่งเป็นกระบวนการที่บุคคลใช้ในการปรับเปลี่ยนพฤติกรรมให้ผสมผสานกับชีวิตประจำวันหรือวิถีชีวิต มีความสัมพันธ์ทางบวกกับการปรับเปลี่ยนพฤติกรรมสุขภาพของผู้ป่วยเบาหวานชนิดที่ 2 ทั้งที่ใช้อินซูลินและไม่ใช้อินซูลิน และการได้รับการสนับสนุนจากบุคคลใกล้ชิด แพทย์ พยาบาลและทีมสุขภาพในกลุ่มผู้สูงอายุโรคเบาหวานชนิดที่ 2 มีความสัมพันธ์กับการปฏิบัติตัวของผู้ป่วยเบาหวาน (กมลพร สิริคุตจตุพร, 2560)

ในการศึกษาครั้งนี้ ผู้วิจัยมีความสนใจที่จะศึกษาประสิทธิภาพเครื่องมือพยากรณ์โรคเบาหวาน และปัจจัยที่มีผลต่อการเกิดโรคเบาหวาน เพื่อรับรู้ถึงความเสี่ยงของผู้ป่วยที่สงสัยป่วยเป็นโรคเบาหวาน ซึ่งปัจจัยคัดสรรที่นำมาศึกษาประกอบด้วย เพศ อายุ น้ำหนัก ดัชนีมวลกาย ความยาวเส้นรอบเอว ประวัติเป็นความดันเลือดสูง ครอบครัวหรือญาติสนิทมีประวัติการเป็นเบาหวาน ประวัติการดื่มสุรา ประวัติการสูบบุหรี่ เพื่อเป็นแนวทางการพัฒนาบริการคลินิกเบาหวานและช่วยให้รับรู้ความชุกของเบาหวานเบื้องต้น ลดค่าใช้จ่ายในการตรวจตรวจสุขภาพหรือค่าใช้จ่ายในการเจาะเลือด ลดความรุนแรงและภาวะแทรกซ้อนของโรคเบาหวานในผู้ป่วยเบาหวานต่อไป

ในการวิจัยครั้งนี้จะกล่าวถึงผู้ป่วยเบาหวานจากกลุ่มเสี่ยงเบาหวาน อำเภอกันทรวิชัย จังหวัดมหาสารคาม เนื่องจากมีการตรวจสอบเกณฑ์สถิติผู้ป่วยเบาหวานรายใหม่จากกลุ่มเสี่ยง

เบาหวาน มีเกณฑ์ที่เกินมาตรฐาน ตามเกณฑ์มาตรฐานที่ตั้งไว้อัตราผู้ป่วยเบาหวานรายใหม่จากกลุ่มเสี่ยงเบาหวาน มีเกณฑ์มาตรฐานไม่เกิน ร้อยละ 3.05 (กระทรวงสาธารณสุข กรมอนามัย, 2561)

ผลที่คาดว่าจะได้รับจากการศึกษาค้นคว้าอิสระครั้งนี้ จะทำให้พยาบาลและบุคลากรทางด้านสาธารณสุข ได้รับรู้ถึงข้อมูลความเสี่ยงเบื้องต้นที่จะเกิดโรคเบาหวาน และนำปัจจัยเสี่ยงที่มีผลต่อการเกิดโรคเบาหวาน ไปทำแบบสอบถามหรือแบบประเมินความเสี่ยงที่จะก่อให้เกิดโรคเบาหวานแทนการเจาะเลือดตรวจเลือดผู้ป่วยในอนาคต ผู้ที่มีปัจจัยเสี่ยงต่อการเกิดโรคเบาหวานจะได้ปรับเปลี่ยนพฤติกรรมการใช้ชีวิตได้อย่างตรงเป้าหมาย และพัฒนาเป็นระบบผู้เชี่ยวชาญในการวินิจฉัยโรคเบาหวานต่อไป

1.2 ความมุ่งหมายของการวิจัย

1. เพื่อศึกษาปัจจัยเสี่ยงที่ส่งผลต่อการก่อให้เกิดโรคเบาหวาน
2. เพื่อหาคุณลักษณะที่มีผลต่อการเป็นโรคเบาหวานของประชาชน
3. เพื่อศึกษาเครื่องมือเหมืองข้อมูลที่ใช้ในการพยากรณ์ที่ก่อให้เกิดโรคเบาหวาน
3. เพื่อวัดประสิทธิภาพความแม่นยำของโปรแกรมในการพยากรณ์

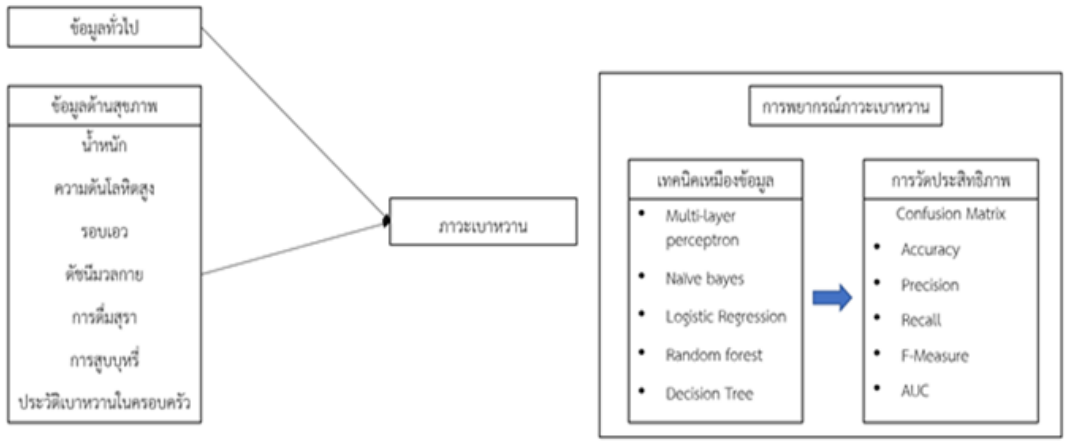
1.3 การใช้เครื่องมือการทำเหมืองข้อมูล

ในปัจจุบันการนำเทคนิคเหมืองข้อมูลมาใช้วิเคราะห์ข้อมูลในทางการแพทย์มีมากขึ้น การประยุกต์ใช้เทคนิคการทำเหมืองข้อมูลกับงานทางการแพทย์เพื่อการวิเคราะห์ข้อมูลสาเหตุของความเจ็บป่วยการเกิดโรคอาการของโรคต่างๆ และแนวทางในการรักษาวิเคราะห์ปัจจัยที่ส่งผลต่อความสำเร็จในการรักษาโรค ผลลัพธ์ที่ได้คือเทคนิคการทำเหมืองข้อมูลสามารถตอบปัญหาและค้นพบความรู้ใหม่จากฐานข้อมูลที่มีอยู่มากมายและซับซ้อนได้ดี โดยเฉพาะเทคนิค Classification ซึ่งเป็นวิธีหนึ่งในการทำเหมืองข้อมูลที่นิยมนำมาใช้จำแนกทำนายผลกลุ่มประชากรจากพฤติกรรมได้อย่างแม่นยำและมีความถูกต้องในระดับสูง (วัจนา ขาวฟ้า และอรศิริ ศิลาสัย, 2561)

จากข้อมูลข้างต้น หากโรงพยาบาลสามารถนำเทคนิคเหมืองข้อมูลมาใช้เพื่อการวิเคราะห์หรือทำนายผลทางการแพทย์ที่อยู่ในโรงพยาบาลได้ในด้านต่างๆ จะส่งผลให้เกิดองค์ความรู้ที่สามารถนำไปใช้ต่อยอดได้หลากหลายมากขึ้น ทั้งนี้ผู้วิจัยจึงเล็งเห็นประโยชน์ในการนำเทคนิคเหมืองข้อมูลมาใช้เพื่อพยากรณ์โรคเบาหวานที่เป็นปัญหาทางสุขภาพร่างกายที่พบได้บ่อยในผู้ป่วย โดยการนำเทคนิคเหมืองข้อมูลมาใช้ในการวิเคราะห์จะช่วยลดภาระ เวลา ค่าใช่จ่าย ของบุคลากรทางการแพทย์ในการคัดกรองโรคเบาหวาน และสามารถนำไปใช้เพื่อการตัดสินใจทางการให้คำปรึกษาและช่วยเหลือผู้ป่วยได้ตั้งแต่ต้น

1.4 กรอบแนวคิดที่ใช้ในการวิจัย

โรคเบาหวานมีสาเหตุมาจากการบกพร่องของฮอร์โมนอินซูลิน ผู้ที่เป็นเบาหวานจะพบว่าตับอ่อนผลิตอินซูลินได้น้อยหรือผลิตไม่ได้เลยหรือผลิตได้ปกติ แต่ประสิทธิภาพของอินซูลินลดลง เช่น ที่พบในคนอ้วน ซึ่งเรียกว่าภาวะดื้อต่ออินซูลิน (Insulin Resistance) เมื่อขาดอินซูลินหรืออินซูลินทำหน้าที่ไม่ได้ น้ำตาลในเลือดจึงเข้าสู่เซลล์ต่างๆ ได้น้อยกว่าปกติ จึงทำให้เกิดการคั่งของน้ำตาลในเลือด และน้ำตาลก็จะถูกขับออกมาทางปัสสาวะ นี่จึงเป็นสาเหตุว่าทำไมเราถึงเรียกโรคนี้ว่า เบาหวาน ผู้ป่วยเบาหวานที่เป็นมาก (มีระดับน้ำตาลในเลือดสูงมาก) มักจะมีอาการปัสสาวะบ่อยและมาก เพราะน้ำตาลที่ออกมาทางไตจะดึงเอาน้ำออกมาด้วย จึงทำให้มีปัสสาวะออกมามากกว่าปกติ เมื่อผู้ป่วยถ่ายปัสสาวะมากก็จะทำให้รู้สึกกระหายน้ำจนต้องคอยดื่มน้ำบ่อยๆ และเนื่องจากร่างกายของผู้ป่วยเบาหวานจะไม่สามารถนำน้ำตาลมาเผาผลาญเป็นพลังงานได้ ร่างกายจึงหันมาเผาผลาญกล้ามเนื้อและไขมันแทน จึงทำให้ร่างกายผอม กล้ามเนื้ออ่อนลีบ ไม่มีไขมัน อ่อนเปลี้ยเพลียแรง นอกจากนี้ การมีน้ำตาลในเลือดสูงเป็นเวลานานๆ ยังทำให้อวัยวะต่างๆ เกิดการเปลี่ยนแปลงผิดปกติ และนำมาซึ่งภาวะแทรกซ้อนได้มากมาย เบาหวานเป็นโรคที่พบได้สูงในคนทุกอายุและทั้งสองเพศ และพบได้สูงขึ้นตามอายุที่มากขึ้น ในบ้านเราพบคนเป็นโรคเบาหวานประมาณ 4-6% ของประชากรทั่วไป 7.1% ของคนไทยช่วงอายุ 20-79 ปี และ 9.6% ของคนไทยที่มีอายุ 35 ปีขึ้นไป และทางสหพันธ์เบาหวานนานาชาติมีการคาดการณ์ว่าจำนวนผู้ป่วยเบาหวานทั่วโลกจะเพิ่มขึ้นจาก 415 ล้านคน ในปี พ.ศ. 2558 เป็น 642 ล้านคน ในปี พ.ศ. 2583 (ศ.เกียรติคุณ พญ.พวงทอง ไกรพิบูลย์). [เว็ปไซต์]. เข้าถึงได้จาก : haamor.com. [22 พ.ย. 2017].



ภาพประกอบ 1 กรอบแนวคิดในการวิจัย

การวิจัยเรื่องการวิเคราะห์ปัจจัยที่ก่อให้เกิดโรคเบาหวานด้วยเทคนิคการทำเหมืองข้อมูล ผู้วิจัยได้ใช้กรอบแนวความคิดในการวิจัย ดังนี้

1. ประชากรและกลุ่มตัวอย่าง

2. เครื่องมือที่ใช้ในการศึกษาค้นคว้าอิสระ

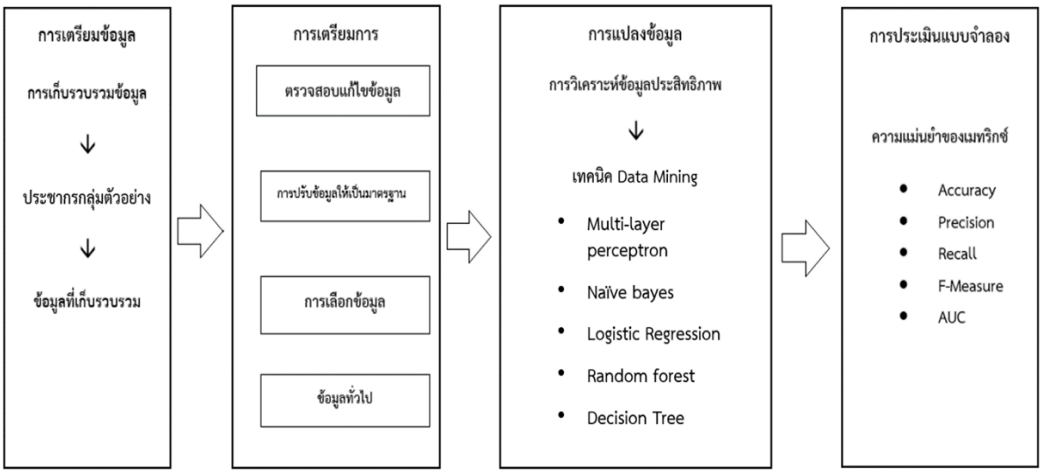
2.1 โปรแกรม

2.2 แบบสอบถาม

3. การเก็บรวบรวมข้อมูล

3.1 สอบถามหรือเก็บข้อมูล

3.2 นำข้อมูลที่ได้มาตรวจสอบความสมบูรณ์ และทำการวิเคราะห์ทางสถิติต่อไป ลำดับขั้นตอนสำหรับเก็บรวมรวมและนำเข้าข้อมูลก่อนเข้าสู่การประมวลผลด้วยเทคนิคเหมืองข้อมูล และทำการเปรียบเทียบประสิทธิภาพของแต่ละเทคนิค ดังภาพประกอบ 2



ภาพประกอบ 2 ขั้นตอนการรวบรวม นำเข้าข้อมูล วิเคราะห์ข้อมูลและเปรียบเทียบประสิทธิภาพ

1.5 สมมุติฐานของการวิจัย

1. ระบบการพยากรณ์โรคเบาหวานมีประสิทธิภาพในการพยากรณ์แตกต่างกันหรือไม่

2. ปัจจัยที่ส่งผลต่อการเกิดโรคเบาหวานมีปัจจัยที่ส่งผลแตกต่างกันหรือไม่

3. ตัวแปรคัดสรร (อายุผู้ป่วย ค่าดัชนีมวลกาย เพศ ประวัติความดันโลหิต ประวัติเบาหวานในพ่อแม่ พี่น้อง ความยาวรอบเอว การสูบบุหรี่ และการดื่มสุรา) สามารถร่วมกันทำนายสาเหตุของการเกิดโรคเบาหวานได้

1.6 ขอบเขตของการวิจัย

1. ประชากรและกลุ่มตัวอย่าง

ประชากร (Population) ที่ใช้ในการวิจัย ได้แก่กลุ่มผู้ป่วยที่มีความเสี่ยงการเกิดโรคเบาหวาน ที่มีอายุระหว่าง 35 - 60 ปี (สถิติสุขภาพคนไทย, 2563) พื้นที่ที่ใช้ในการวิจัย ได้แก่ คลินิกเบาหวานโรงพยาบาลกันทรวิชัย จังหวัดมหาสารคาม ระยะเวลาที่ใช้ในการเก็บรวบรวมข้อมูล วันที่ 1 ธันวาคม พ.ศ. 2563 ถึง 31 มีนาคม 2564

2. ตัวแปรที่ใช้ในการวิจัย

การวิจัยเป็นการวิจัยเชิงสำรวจ ข้อมูลนำเข้าที่ใช้ในการศึกษาเป็นข้อมูลของกลุ่มผู้ป่วยที่มีความเสี่ยงการเกิดโรคเบาหวาน ที่มีอายุระหว่าง 35-60 ปี ในโรงพยาบาลกันทรวิชัย จังหวัดมหาสารคาม จำนวนของกลุ่มตัวอย่างทั้งหมด 400 คน จากการคำนวณด้วยตารางสำเร็จรูปของ ทาโร ยามาเน่ (Taro Yamane) ที่กำหนดให้มีสัดส่วนของลักษณะที่สนใจในประชากร เท่ากับ 0.5 ระดับความคลาดเคลื่อนที่ยอมรับได้ 5% และระดับความเชื่อมั่น 95% เครื่องมือที่ใช้ในการเก็บข้อมูลเป็นการสอบถามผ่านแบบสอบถามที่มี 2 ส่วน คือแบบสอบถามข้อมูลทั่วไป และแบบสอบถามเกี่ยวกับพฤติกรรมความเสี่ยงที่ก่อให้เกิดโรคเบาหวาน โดยเทคนิคที่ใช้ในการศึกษาครั้งนี้คือเทคนิคการจำแนกข้อมูล (Classification) ในการจำแนกกลุ่มข้อมูลด้วยคุณลักษณะโรคเบาหวาน ทั้ง 5 ระดับ โดย “ไม่เคยปฏิบัติ” ให้ 1 คะแนน จนถึง “ปฏิบัติเป็นประจำ” ให้ 5 คะแนน คะแนนมาก หมายถึงมีการปฏิบัติพฤติกรรมการป้องกันโรคเบาหวานอยู่ในระดับสูง คือน้อยมาก น้อย ปานกลาง มาก และมากที่สุด ซึ่งเป็นวิธีที่เหมาะกับการสร้างตัวแบบจำลองสำหรับการพยากรณ์ค่าข้อมูล (Predictive Modeling) ในอนาคตจากกลุ่มตัวอย่างที่มี วิธีการคัดกรองโรคเบาหวาน การทบทวนวรรณกรรมพบว่ามีหลายสาเหตุที่เป็นความเสี่ยงที่จะก่อให้เกิดโรคเบาหวานได้ โดยเครื่องมือที่ใช้ในการเก็บรวบรวมข้อมูลจะอยู่ในรูปแบบแบบสอบถาม (ดารณี ทองสัมฤทธิ์ และคณะ, 2560) แบบสอบถามเกี่ยวกับความเสี่ยงที่ก่อให้เกิดโรคเบาหวาน แบ่งออกเป็น 2 ส่วน ดังนี้ ส่วนที่ 1 ข้อมูลทั่วไป ประกอบด้วยข้อมูลเกี่ยวกับ เพศ อายุ การศึกษา การนับถือศาสนา สถานภาพ อาชีพ ดัชนีมวลกาย ความยาวเส้นรอบเอว ความดันโลหิต การดื่มสุรา การสูบบุหรี่ ส่วนสูง น้ำหนัก และพันธุกรรมด้านการเป็นเบาหวาน จำนวนทั้งสิ้น 13 ข้อ มีลักษณะคำตอบเป็นแบบการปฏิบัติพฤติกรรมเสี่ยงที่ก่อให้เกิดโรคเบาหวาน ส่วนที่ 2 แบบสอบถามเกี่ยวกับพฤติกรรมความเสี่ยงที่ก่อให้เกิดโรคเบาหวาน โดยเทคนิคที่ใช้ในการศึกษาครั้งนี้คือเทคนิคการจำแนกข้อมูล (Classification) ในการจำแนกกลุ่มข้อมูลด้วยคุณลักษณะโรคเบาหวาน ทั้ง 5 ระดับ โดย “ไม่เคยปฏิบัติ” ให้ 1 คะแนน จนถึง “ปฏิบัติเป็นประจำ” ให้ 5 คะแนน คะแนนมาก หมายถึง มีการปฏิบัติพฤติกรรมเสี่ยงที่ก่อให้เกิดโรคเบาหวาน อยู่ในระดับสูง คือ น้อยมาก น้อย ปานกลาง มาก และมากที่สุด จำนวนทั้งสิ้น 32 ข้อ

1.7 การวิเคราะห์ข้อมูล

นำข้อมูลของผู้ป่วยโรคเบาหวานที่ถูกจัดเก็บไว้ในฐานข้อมูล (Database) และนำแนวคิดเทคนิคการทำเหมืองข้อมูล (Data Mining) โดยใช้กระบวนการมาตรฐานการทำเหมืองข้อมูล (CRISP-DM) มาช่วยในการวิเคราะห์ข้อมูลการคัดเลือกคุณสมบัติ (Feature Selection) เพื่อให้ทราบถึงคุณลักษณะที่มีผลต่อการมีความเสี่ยงที่จะก่อให้เกิดโรคเบาหวาน โดยเลือกใช้เทคนิควิธีการจำแนกประเภทกลุ่มข้อมูล (Classification) มาพัฒนาตัวแบบ (Model) ที่จะช่วยสนับสนุนการตัดสินใจว่ามีความเสี่ยงจะเกิดโรคเบาหวานด้วยเทคนิควิธีต้นไม้ตัดสินใจ (Decision Tree) การจำแนกประเภทข้อมูลแบบเบย์อย่างง่าย (Naïve Bayes) โครงข่ายประสาท (Neural Network) และซัพพอร์ท

เวกเตอร์แมชชีน (Support Vector Machine: SVM) จากนั้นทำการเปรียบเทียบประสิทธิภาพแบบจำลอง โดยดูจากค่าความถูกต้อง (Accuracy) ค่าความแม่นยำของแบบจำลอง (Precision) ค่าความระลึกของแบบจำลอง (Recall) และค่าความเหวี่ยงของแบบจำลอง (F-Measure) เพื่อหาแบบจำลองที่ดีที่สุดสำหรับการนำไปใช้เป็นข้อมูลในการทำนายความเสี่ยงที่จะก่อให้เกิดโรคเบาหวานต่อไป

1.8 ความสำคัญของการวิจัย

โรงพยาบาลกันทรวิชัย อำเภอกันทรวิชัย จังหวัดหมาสารคามมีการรับผู้ป่วยโรคเบาหวานจำนวนมาก ผลการวิจัยครั้งนี้จะทำให้พยาบาลและบุคลากรทางด้านสาธารณสุข ได้รับรู้ถึงข้อมูลความเสี่ยงที่เกิดโรคเบาหวานเบื้องต้น จากเครื่องมือการพยากรณ์ในประชากรอายุ 35 ปีขึ้นไป ในเขตอำเภอกันทรวิชัย จังหวัดมหาสารคาม ซึ่งจะเป็นประโยชน์ในการนำไปวางแผนกลยุทธ์ เพื่อแก้ไขปัญหาโรคเรื้อรังหรือจัดทำโครงการส่งเสริมสุขภาพและป้องกันโรคความดันโลหิตสูงและโรคเบาหวานต่อไป

1.9 นิยามศัพท์

1. โรคเบาหวาน (Diabetes) หมายถึง โรคที่เกิดจากความผิดปกติของการทำงานของฮอร์โมนที่ชื่อว่า อินสุลิน (Insulin) ซึ่งโดยปกติแล้วร่างกายของคนเราจำเป็นต้องมีอินสุลิน เพื่อนำน้ำตาลในกระแสเลือดไปเลี้ยงอวัยวะต่างๆ ของร่างกายโดยเฉพาะสมองและกล้ามเนื้อ ในภาวะที่อินสุลินมีความผิดปกติ ไม่ว่าจะเป็นการลดลงของปริมาณอินสุลินในร่างกาย หรือการที่อวัยวะต่างๆ ของร่างกายตอบสนองต่ออินสุลินลดลง หรือที่เรียกว่า ภาวะดื้ออินสุลิน จะทำให้ร่างกายไม่สามารถนำน้ำตาลที่อยู่ในกระแสเลือดไปใช้ได้อย่างเต็มประสิทธิภาพ ทำให้มีปริมาณน้ำตาลคงเหลือในกระแสเลือดมากกว่าปกติ สาเหตุของโรคเบาหวานเกิดจากการที่ระดับน้ำตาลในกระแสเลือดสูงมากขึ้นถึงระดับหนึ่ง จนทำให้ไตดูดกลับน้ำตาลได้ไม่หมด ซึ่งปกติไตจะมีหน้าที่ดูดกลับน้ำตาลจากสารที่ถูกกรองจากหน่วยไตไปใช้ ส่งผลให้มีน้ำตาลรั่วออกมากับปัสสาวะ หากปล่อยให้เกิดภาวะเช่นนี้ไปนานๆ โดยไม่ได้รับการรักษาอย่างถูกวิธี จะทำให้เกิดภาวะแทรกซ้อนที่ร้ายแรงตามมาในที่สุด (พญ.ศศิภัสช์ ช้อนทอง, 2564)

2. การทำเหมืองข้อมูล (Data Mining) หมายถึง การทำเหมืองข้อมูลคือเทคนิคที่ใช้คอมพิวเตอร์ช่วยในการวิเคราะห์เพื่อประมวลผลและสำรวจชุดข้อมูลขนาดใหญ่ เมื่อใช้เครื่องมือและวิธีการทำเหมืองข้อมูล องค์กรสามารถค้นพบรูปแบบและความสัมพันธ์ที่ซ่อนอยู่ในข้อมูลของตน การทำเหมืองข้อมูลแปลงข้อมูลดิบเป็นความรู้เชิงปฏิบัติ บริษัทใช้ความรู้นี้ในการแก้ไขปัญหา วิเคราะห์ผลกระทบในอนาคตของการตัดสินใจทางธุรกิจ และเพิ่มขอบเขตกำไรของบริษัท การทำเหมืองข้อมูลเป็นการเรียกชื่อที่ผิดเพราะเป้าหมายของการทำเหมืองข้อมูลไม่ใช่การดึงหรือทำเหมืองข้อมูลเอง แต่กลับกลายเป็นการมีข้อมูลจำนวนมากอยู่แล้ว และการทำเหมืองข้อมูลดึงความหมายหรือความรู้ที่มีค่าจากข้อมูลนั้น กระบวนการทั่วไปของการเก็บรวบรวมข้อมูล การเก็บ การวิเคราะห์ และการขุด การเก็บรวบรวมข้อมูลกำลังรวบรวมข้อมูลจากแหล่งต่างๆ เช่น ผลตอบรับจากลูกค้า การชำระเงิน และ

ใบสั่งซื้อ การเก็บคลังข้อมูลเป็นกระบวนการเก็บข้อมูลของฐานข้อมูลขนาดใหญ่หรือคลังข้อมูล การวิเคราะห์ข้อมูลคือการประมวลผล เก็บ และวิเคราะห์ข้อมูลเพิ่มเติมโดยใช้ซอฟต์แวร์และอัลกอริธึมที่ซับซ้อน การทำเหมืองข้อมูลเป็นสาขาหนึ่งของการวิเคราะห์ข้อมูลหรือกลยุทธ์การวิเคราะห์ที่ใช้เพื่อค้นหารูปแบบที่ซ่อนอยู่หรือไม่รู้จักก่อนหน้านี้ในข้อมูล (Amazon Web Services (AWS), 2565)

3. การควบคุมระดับน้ำตาลในเลือด หมายถึง เป้าหมายหลักในการดูแลควบคุมระดับน้ำตาลในเลือดให้อยู่ในระดับใกล้เคียงกับปกติมากที่สุด โดยควบคุมระดับน้ำตาลในเลือดก่อนทานอาหารให้ได้ประมาณ 90 - 130 มก./ดล. ระดับน้ำตาลในเลือดก่อนอาหารเช้า ควรอยู่ประมาณ 90 - 110 มก./ดล.ส่วนหลังอาหารควรต่ำกว่า 180มก./ดล หากควบคุมระดับน้ำตาลตั้งแต่ระยะเริ่มต้นของโรคให้ดี ก็จะช่วยชะลอความเสื่อมของโรค และช่วยให้การควบคุมระดับน้ำตาลในเลือดง่ายขึ้นเท่านั้น จุดมุ่งหมายของการควบคุมระดับน้ำตาลในเลือด คือ มีค่าน้ำตาลสะสม (HbA1c) ให้น้อยกว่า 6.5% จะทำให้มีประโยชน์ต่อการลดความเสี่ยงการเกิดภาวะแทรกซ้อนจากหลอดเลือดเล็ก และหลอดเลือดใหญ่ ควรตรวจค่าน้ำตาลสะสม (HbA1c) ทุก 3-6 เดือน (โรงพยาบาลเปาโล พหลโยธิน, 2564)

4. ผู้สูงอายุ หมายถึง ผู้ที่มีอายุตั้งแต่ 60 ปีขึ้นไป ในวัยนี้จะมีการเปลี่ยนแปลงหลายด้าน ทั้งทางร่างกาย ทางสมอง ทางอารมณ์และทางสังคม จึงเป็นวัยที่คนส่วนมากกลัว ดังนั้นการได้เข้าใจถึงการเปลี่ยนแปลงต่างๆ ดังกล่าวก็จะช่วยให้ผู้สูงอายุได้ปรับตัวได้ดียิ่งขึ้นรวมทั้งบุตรหลานหรือผู้ใกล้ชิดมีส่วนสำคัญมากในการช่วยให้ท่านปรับตัวได้อย่างมีความสุข การเปลี่ยนแปลงในด้านต่างๆ ของผู้สูงอายุ การเปลี่ยนแปลงทางร่างกาย เริ่มอ่อนแอลงและเชื่องช้า ร่างกายเสื่อมโทรม บางรายมือเท้าสั่นการทรงตัวไม่ดี ปรับตัวกับสภาพดินฟ้าอากาศได้ยาก ทำให้เจ็บป่วยได้ง่าย การเปลี่ยนแปลงทางด้านสมอง หลงลืมง่าย ความจำไม่ดี แต่เหตุผลและความสุขุมรอบคอบอาจจะยังคงที่ การเปลี่ยนแปลงทางด้านอารมณ์ อารมณ์ไม่คงที่ อ่อนไหวง่าย มักเอาแต่ใจตัวเอง โกรธง่าย ชอบบ่น แต่บางรายใจดี ทั้งนี้ขึ้นอยู่กับสภาพเศรษฐกิจในครอบครัวด้วย การเปลี่ยนแปลงทางสังคม ส่วนมากก็มักจะฝากจิตใจไว้กับวัดและยึดศาสนาเป็นที่พึ่ง บางรายอาจต้องหาเลี้ยงชีพอยู่หรือเลี้ยงดูลูกหลานในบ้าน ซึ่งก็ช่วยให้มีความสุขเพลิดเพลิน (คณะแพทยศาสตร์โรงพยาบาลรามาธิบดี มหาวิทยาลัยมหิดล, 2561)

5. ความแม่นยำ (Accuracy) หมายถึง การวัดได้ใกล้เคียงกับค่าการตรวจมาตรฐานสูงสุด (Gold Standard) หรือค่าที่ตั้งใจจะวัดหรือค่าจริงหรือไม่ อีกนัยหนึ่งกล่าวได้ว่า ความแม่นยำ คือ ความถูกต้องของค่าที่วัดได้ เป็นความใกล้เคียงกับค่าจริงหรือใกล้เคียงกับค่าจากเครื่องมือมาตรฐาน ยกตัวอย่างเช่น หากมีการวัด น้ำหนักของนักมวยก่อนขึ้นชกจริง ซึ่งสามารถวัดค่าน้ำหนักได้ 70.1 kg เมื่อเทียบกับนำหนักจริงของนักมวย ซึ่งมีค่าน้ำหนักที่ 70 kg สามารถกล่าวได้ว่าค่าที่ได้จากการวัดจากเครื่องชั่งน้ำหนักของฝ่ายการจัดการแข่งขัน มีความแม่นยำหรือความถูกต้อง (กิตติคุณ แสงนิล และประสพชัย พสุนนท์, 2561)

บทที่ 2
วรรณกรรมและผลงานวิจัยที่เกี่ยวข้อง

การวิจัย เรื่องการวิเคราะห์ปัจจัยที่ก่อให้เกิดโรคเบาหวานด้วยวิธีการเหมืองข้อมูล ผู้วิจัยได้ศึกษาจากเอกสารงานวิจัยที่เกี่ยวข้องเพื่อใช้เป็นแนวทางในการศึกษาวิจัย ดังนี้

1. ข้อมูลทั่วไปของโรคเบาหวาน
2. ปัจจัยที่ส่งผลต่อการก่อให้เกิดโรคเบาหวาน
3. เครื่องมือการคัดกรองโรคเบาหวาน
4. แนวคิดเทคนิคการทำเหมืองข้อมูล
5. งานวิจัยที่เกี่ยวข้อง

2.1 ข้อมูลทั่วไปของโรคเบาหวาน

โรคเบาหวาน

โรคเบาหวาน (Diabetes Mellitus : DM) คือภาวะที่ร่างกายมีระดับน้ำตาลในเลือดสูงกว่าปกติ (Hyperglycemia) ซึ่งเป็นผลมาจากความผิดปกติทางเมตะบอลิซึม เนื่องจากการขาดฮอร์โมนอินซูลินหรือร่างกายไม่ตอบสนองต่อฮอร์โมนอินซูลิน ซึ่งฮอร์โมนอินซูลินเป็นฮอร์โมนสำคัญของร่างกายสร้างและหลั่งจากเบต้าเซลล์ของตับอ่อน และมีหน้าที่ในการพาน้ำตาลกลูโคสเข้าสู่เนื้อเยื่อต่างๆ ของร่างกาย เพื่อเผาพลาญพลังงานในการดำเนินชีวิต ถ้าขาดฮอร์โมนอินซูลินหรือการออกฤทธิ์ของฮอร์โมนอินซูลินบกพร่อง ร่างกายก็ใช้น้ำตาลไม่ได้ จึงทำให้มีน้ำตาลในเลือดสูงและมีอาการต่างๆ ของโรคเบาหวาน ซึ่งระดับนำตาลที่สูงขึ้นนี้เป็นตัวการสำคัญที่ทำให้เกิดปัญหาต่างๆ ตามมา ที่สำคัญคือเป็นตัวเร่งให้เกิดการเสื่อมของหลอดเลือดแดงทั่วร่างกาย ทั้งหลอดเลือดแดงที่มาเลี้ยงสมอง หัวใจ ตา ไต แขน ขา รวมทั้งหลอดเลือดแดงเล็กๆ ที่เลี้ยงปลายประสาทอีกด้วย ทำให้เกิดการตีบตันของหลอดเลือดแดง (สุภาภรณ์ ตันตินันทตระกูล, 2555) มีผู้ที่เสี่ยงต่อการเป็นโรคเบาหวาน คือมีค่าระดับน้ำตาลอยู่ในช่วง 100-125 mg% ที่ผ่านตามเกณฑ์การคัดเลือกกลุ่มตัวอย่างเข้าร่วมวิจัย และข้อมูลมีความสมบูรณ์จำนวน ทั้งสิ้น 227 คน โดยร้อยละ 85.5 เป็นเพศหญิง อยู่ในวัยผู้ใหญ่ตอนปลาย ร้อยละ 44.5 มีอายุเฉลี่ย 64.41 (SD = 42) ส่วนใหญ่มีสถานภาพสมรสคู่ ร้อยละ 56.8 สำเร็จการศึกษาชั้นประถมศึกษาและเป็นแม่บ้าน คิดเป็นร้อยละ 69.6 และ 37.4 ตามลำดับ กลุ่มตัวอย่างร้อยละ 44.1 มีญาติสายตรงเป็นเบาหวาน มากกว่าครึ่งของกลุ่มตัวอย่าง (ร้อยละ 71.8) มีน้ำหนักมากกว่าเกณฑ์มาตรฐาน คือ มีภาวะน้ำหนักเกินร้อยละ 21.6 และมีภาวะอ้วนร้อยละ 50.2 เมื่อพิจารณาความถี่ของการปฏิบัติพฤติกรรมในแต่ละด้าน พบว่ากลุ่มตัวอย่างมีพฤติกรรมการรับประทานอาหารที่ไม่เหมาะสมและปฏิบัติเป็นประจำได้แก่ การรับประทานอาหารประเภทไขมัน ผลไม้รสหวานจัด อาหารที่มีกะทิ ขนมหวาน และแป้ง โดยมีกลุ่มตัวอย่างที่รับประทานอาหารไขมันต่ำและที่รับประทานผักต่าง ๆ เป็นประจำ พฤติกรรมด้านการออกกำลังกาย พบว่าในชีวิตประจำวันมีกลุ่มตัวอย่างบางกลุ่มเท่านั้นที่ออกกำลังกายเป็นประจำและออกกำลังกายบางครั้ง (เขมารดี มาสิงบุญนิ และคณะ, 2562)

โรคเบาหวานเป็นโรคเรื้อรังที่เป็นปัญหาสาธารณสุขทั้งระดับโลกและระดับประเทศที่สามารถเกิดขึ้นได้กับทุกเพศทุกวัย รวมทั้งความชุกของกลุ่มเสี่ยงเบาหวานในเด็กก็มีแนวโน้มสูงขึ้นมากเช่นเดียวกัน จึงควรส่งเสริมให้มีการตรวจคัดกรองความเสี่ยงโรคเบาหวานในกลุ่มเด็กอายุ ต่ำกว่า 15 ปี ที่เป็นกลุ่มเสี่ยงต่อการเป็นโรคเบาหวานสูง และส่งเสริมการปฏิบัติตัวของเด็กและบุคคลในครอบครัวเพื่อให้เกิดการเปลี่ยนแปลงพฤติกรรมเพื่อลดอ้วนลดพุง ซึ่งจะลดปัจจัยเสี่ยงต่อโรคเบาหวานและเพื่อเป็นแนวทางในการป้องกันการเกิดโรคเบาหวานในเด็กได้ชัดเจนยิ่งขึ้น (จิตตรา ตันเจริญ และคณะ, 2561) ความเสี่ยงต่อการเกิดโรค เบาหวานชนิดที่ 2 ของผู้หญิงวัยกลางคน ผู้หญิงวัยกลางคนที่มีความเสี่ยงต่อการเกิดโรคจำนวน 19 คน ผลการศึกษาพบว่าในผู้ที่สมัครใจเข้ารับการสำรวจปัจจัยเสี่ยงเป็นผู้หญิงวัยกลางคนที่อายุ 40-60 ปีจำนวน 42 คน มีปัจจัยเสี่ยงต่อการเกิดโรคเบาหวานชนิดที่ 2 ได้แก่ ค่าดัชนีมวลกายตั้งแต่ 23 กก./ม2 ขึ้นไป ร้อยละ 88.09 เส้นรอบเอวตั้งแต่ 80 ซม.ขึ้นไป ร้อยละ 85.71 อายุตั้งแต่ 45 ปีขึ้นไป ร้อยละ 78.57 และมีญาติสายตรงเป็นโรคเบาหวาน ร้อยละ 21.43 ผู้หญิงกลุ่มนี้ รับรู้ว่าภาวะเสี่ยงข้างต้นเกี่ยวข้องกับเงื่อนไขการมีญาติเป็นโรคเบาหวานและภาวะอ้วนซึ่งประการหลังเป็นปัจจัยเสี่ยงที่ตนเองสามารถปรับเปลี่ยนได้เพราะเกี่ยวข้องกับวิถีชีวิต ได้แก่ การมีชีวิตที่ไม่ค่อยออกแรง ไม่ชอบออกกำลังกาย จำเป็นต้องรับประทานอาหารกากใยน้อยตามความชอบของครอบครัว (จิราภรณ์ ชิณโสม และ วิพร เสนารักษ์, 2554)

ปจจุบันวัยรุนมีพฤติกรรมเสี่ยงตอการเกิดโรคเรื้อรังที่มีผลกระทบตอภาวะสุขภาพ ผลการวิจัยพบวากลุมตัวอย่าง มีประวัติบุคคลในครอบครัวป่วยเป็นโรคความดันโลหิตสูงและโรคเบาหวาน ร้อยละ 17 มีค่าดัชนีมวลกายเกินมาตรฐาน ร้อยละ 82 มีความรูเกี่ยวกับโรคความดันโลหิตสูงและเบาหวานอยูในระดับต่ำมีเพียง ร้อยละ 2 เทานั้น ที่มีความรู้อยู่ในระดับสูงด้านพฤติกรรมเสี่ยงตอโรคพบวา ร้อยละ 33 บริโภคอาหาร หวาน มัน เค็ม เป็นประจำสวนร้อยละ 20 รับประทานผักผลไม้เป็นประจำมีเพียง ร้อยละ 24 ออกกำลังกายหรือเล่นกีฬาเป็นประจำ ร้อยละ 15 มีพฤติกรรมจัดการความเครียดที่ไม่เหมาะสม เชนสูบบุหรี่ดื่มเครื่องดื่มที่มีแอลกอฮอล์หรือใชยานอนหลับ/ยาระงับประสาท (กิตติศักดิ์ พรมดาว และคณะ, 2556) ผู้ป่วยเบาหวานชนิดที่ 2 จำนวน 1,177 ราย โดยมีความชุก 7.80% (95% Cl, 7.11-8.47) ความชุกของภาวะแทรกซ้อนจากโรคเบาหวานที่เกิดจากการมองเห็นไม่ดี โปรตีนในปัสสาวะ ครีเอตินิน ที่เพิ่มขึ้น หลอดเลือดหัวใจตีบ กล้ามเนื้อหัวใจตาย โรคหลอดเลือดสมองและแผลที่เท้าเพิ่มขึ้นในกลุ่มผู้ป่วยเบาหวาน ผู้ป่วยเบาหวานสูงอายุได้รับการบริการที่ดีจากผู้ป่วยเบาหวาน รวมทั้งการตรวจตาเป็นประจำ การจัดกระดูกประจำปี และการเข้ารับการตรวจของแพทย์ อย่างไรก็ตาม ความเข้าใจเกี่ยวกับการจัดการโรคเบาหวานในแต่ละวันนั้นไม่ดี โดยมีความชุกของความบกพร่องทางสติปัญญา (22.5%) ในกลุ่มผู้ป่วยเบาหวาน (Hewitt & Jonathan 2007)

การจำแนกประเภทของเบาหวาน โดยแบ่งประเภทของเบาหวานเป็น 3 ประเภท

1) เบาหวานชนิดที่ 1 เกิดจากเบต้าเซลล์ของตับอ่อนถูกทำลายจากปฏิกิริยาทางภูมิคุ้มกัน ทำให้ไม่ สามารถผลิตอินซูลินได้อย่างสมบูรณ์ ต้องได้รับการรักษาด้วยอินซูลินตลอดชีวิต พบได้ตลอดทุกช่วงอายุ แต่พบมากในเด็กและผู้ใหญ่ตอนต้น (กฤชฐา จีระวงศ์พานิช และคณะ, 2564)

2) เบาหวานชนิดที่ 2 เกิดจากร่างกายไม่สามารถใช้อินซูลินได้อย่างมีประสิทธิภาพ และความบกพร่องของการหลั่งสารอินซูลินในร่างกายทำให้ร่างกายมีภาวะน้ำตาลในเลือดสูงกว่าปกติ ระดับน้ำตาลในเลือดที่สูงเป็นระยะเวลานาน ทำให้เกิดการเปลี่ยนแปลงของหลอดเลือดแดงทั่วร่างกาย จะส่งผลให้เกิดภาวะแทรกซ้อนต่อเท้า ไต ตา ระบบประสาท หัวใจ และหลอดเลือดสมอง อันเป็นสาเหตุทำให้เกิดอาการป่วยและตายก่อนวัยอันสมควร (ฐิติภัทร จันเกษม และ ปัทมา สุพรรณกุล, 2562) 3) เบาหวานขณะตั้งครรภ์ เกิดจากการเปลี่ยนแปลงทางด้านร่างกายและความผิดปกติของการเผาผลาญคาร์โบไฮเดรตทำให้น้ำตาลในกระแสเลือดสูง ส่งผลผลกระทบต่อภาวะสุขภาพของหญิงตั้งครรภ์และทารกในครรภ์ เช่น ความดันโลหิตสูง ติดเชื้อ คลอดยาก ตกเลือดหลังคลอด แท้ง ทารกเสียชีวิตในครรภ์ และทารกพิการแต่กำเนิด (วลัยลักษณ์ สุวรรณภักดี และคณะ, 2562)

การศึกษาครั้งนี้ผู้ศึกษาสนใจศึกษาเบาหวานชนิดที่ 2 เนื่องจากเป็นเบาหวานที่พบได้บ่อยในโรงพยาบาลกันทรวิชัย อำเภอกันทรวิชัย จังหวัดมหาสารคาม อันจะทำให้ทราบถึงพฤติกรรมของผู้ป่วยเบาหวานในแต่ละด้าน เพื่อนำไปสู่การรับรู้อาการเบื้องต้น และรับการรักษาให้มีความเหมาะสมกับผู้ป่วยต่อไป

2.2 ปัจจัยที่ส่งผลต่อการก่อให้เกิดโรคเบาหวาน

ปัจจุบันสาเหตุของการเกิดเบาหวานชนิดที่ 2 ยังไม่เป็นที่ทราบแน่ชัด แต่พบว่าเป็น ภาวะที่มีความเกี่ยวข้องกับปัจจัยทางด้านพฤติกรรม พันธุกรรม และสภาพแวดล้อมเป็นอย่างมาก จากการศึกษาพบว่ามีปัจจัยหลากหลายที่มีความสัมพันธ์ต่อการเกิดเบาหวานชนิดที่ 2 ได้แก่

1. อายุ เป็นปัจจัยหนึ่งที่มีความสัมพันธ์กับการเกิดโรคเบาหวาน นอกจากการเปลี่ยนแปลงตามวัย ความสามารถการทำกิจวัตรประจำวันลดลง การทำงานของสมองผิดปกติ พลัดตกหกล้ม ซึมเศร้า กลั้นปัสสาวะไม่ได้ รวมทั้งสมรรถภาพทางเพศเสื่อมในผู้ชายแล้วยังมีปัจจัยอื่นๆ ที่ทำให้มีโอกาสเป็นเบาหวานมากขึ้น สำหรับผู้สูงอายุการสังเคราะห์และการหลั่งฮอร์โมนอินซูลินลดลง (นพนัฐ จำปาเทศ และคณะ, 2561)

2. เพศ เป็นปัจจัยเสี่ยงสำคัญที่ทำให้ความชุกของโรคเบาหวานเพิ่มขึ้นทุกปี โดยเฉพาะในเขตเมือง คือ ความอ้วนและน้ำหนักตัวเกิน ผู้หญิงไทยมีสัดส่วนการเป็นโรคเบาหวานสูงกว่าเพศชายอยู่ที่ ร้อยละ 9.8 และ 7.9 ตามลำดับ โดยภาคตะวันออกเฉียงเหนือมีความชุกสูงที่สุด ตามด้วยภาคกลาง ภาคเหนือ และภาคใต้ ที่น่าสนใจคือ ร้อยละ 43.1 ของผู้ที่เป็นเบาหวานไม่ทราบว่าตนเองเป็นเบาหวานมาก่อน เท่ากับว่าภาวะโรคเบาหวานในปัจจุบันมีคนที่เป็นเบาหวานแล้วมากกว่า 4 ล้านคน ทั้งยังพบกลุ่มเสี่ยงต่อเบาหวานอีกถึง 7.7 ล้านคน ซึ่งภายในร้อยละ 5-10 ต่อปีจะป่วยเป็นเบาหวาน (เอชโฟกัส เจาะลึกระบบสุขภาพ สำนักข่าว Hfocus เจาะลึกระบบสุขภาพ และ มูลนิธิภิวัฒน์สาธารณสุขไทย เลขที่ 7 ถ.อธิบดี ต.หมากแข้ง อ.เมือง จ.อุดรธานี 41000, 2561)

3. ค่าดัชนีมวลกาย ผู้ป่วยที่มีดัชนีมวลกายปกติ มีค่าดัชนีมวลกายที่ 18.5-22.9 กิโลกรัมต่อตารางเมตร และกลุ่มอ้วน คือข้อมูลที่ได้จากกลุ่มผู้ที่มีภาวะอ้วนตั้งแต่อ้วนถึงอ้วนมาก มีค่าดัชนีมวลกายตั้งแต่ 25.0 กิโลกรัมต่อตารางเมตร ทำให้ระดับไขมันในเลือดเพิ่มขึ้น และเพิ่มโอกาสเป็นโรคเบาหวาน (ปิยะมาศ เอมอิ่มอนันต์, 2562)

4. ค่าความดันโลหิตสูง ระดับความดันโลหิตที่เหมาะสมสำหรับผู้ป่วยโรคความดันโลหิตสูง คือ 120-130/70-79 มม.ปรอท โดยการลด SBP ลงมาต่ำกว่า 130 มม.ปรอท อาจสามารถลดการเกิดโรคทางระบบหัวใจและหลอดเลือดได้ แต่การลด SBP จนต่ำกว่า 120 มม.ปรอท อาจสัมพันธ์กับการเพิ่มความเสี่ยงต่อการเกิดโรคเบาหวาน โรคหัวใจและหลอดเลือดมากขึ้น (สมาคมความดันโลหิตสูงแห่งประเทศไทย, 2562)

5. ประวัติเบาหวานในพ่อแม่ พี่น้อง โรคเบาหวานที่เกิดจากความผิดปกติทางพันธุกรรม โรคเบาหวานที่เกิดจากโรคของตับอ่อน จากความผิดปกติของต่อมไร้ท่อ จากยา จากการติดเชื้อ จากปฏิกิริยาภูมิคุ้มกัน หรือโรคเบาหวานที่พบร่วมกับกลุ่มอาการต่างๆ ผู้ป่วยจะมีลักษณะจำเพาะของโรคหรือกลุ่มอาการนั้นๆ หรือมีอาการและอาการแสดงของโรคที่ทำให้เกิดเบาหวาน (สมาคมโรคเบาหวานแห่งประเทศไทย, 2560)

6. เส้นรอบเอว ภาวะน้ำหนักเกินก่อให้เกิดผลกระทบทั้งด้าน ร่างกาย จิตใจ สังคม และเศรษฐกิจในด้านร่างกายผู้ที่มีน้ำหนักเกิน เพิ่มความเสี่ยงของการเจ็บป่วยจากกระดูกและข้อ อักเสบ (osteoarthritis) เพิ่มความดันโลหิตสูง เพิ่มระดับไขมันในเลือด เพิ่มโอกาสเป็นโรคเบาหวาน โดยรอบเอวที่ เพิ่มขึ้นทุกๆ 5 เซนติเมตร จะเพิ่มโอกาสเกิดโรคเบาหวาน ได้ 3-5 เท่า อีกทั้งการเพิ่มของน้ำหนักตัว ดัชนีมวลกาย ปริมาณไขมันในช่องท้อง และปริมาณไขมันในเลือดสูง จะส่งผลให้เกิดปัจจัยเสี่ยงเกี่ยวกับโรคหัวใจและหลอดเลือด (ปราณี จันธิมา และสมเกียรติ ศรีธาราธิคุณ, 2560)

7. การสูบบุหรี่ การสูบบุหรี่เป็นปัจจัยเสี่ยงทางพฤติกรรมและเป็นสาเหตุหนึ่งในการทำให้เกิดโรคเรื้อรังที่ไม่ติดต่อ รวมทั้งการเกิดโรคเบาหวาน ผู้ที่สูบบุหรี่จะมีความเสี่ยงต่อการเกิดโรคเบาหวานมากกว่าผู้ที่ไม่สูบบุหรี่ ผู้ป่วยเบาหวานที่มีพฤติกรรมสูบบุหรี่หากเลิกสูบบุหรี่ไม่ได้จะทำให้เกิดภาวะแทรกซ้อนเพิ่มขึ้น จากการศึกษาความสัมพันธ์ระหว่างการสูบบุหรี่กับการเกิดโรคแทรกซ้อนของหลอดเลือดที่หลอดเลือดเล็กและหลอดเลือดใหญ่ (อารยา หาอุปละ และน้ำเพชร สายบัวทอง, 2561)

8. การดื่มสุรา เป็นปัจจัยหนึ่งที่มีความสัมพันธ์กับการเกิดเบาหวาน จากการดื่มสุราเป็นเหตุให้มีแนวโน้มการเกิดโรคเรื้อรังเพิ่มขึ้น โดยพบว่าความชุกของโรคเรื้อรังที่เกี่ยวข้องกับระบบหัวใจหลอดเลือด มีภาวะความดันโลหิตสูง และระดับไขมันในเลือดสูง พบในเพศชายและเพศหญิง เป็นปัจจัยเสี่ยงที่จะก่อให้เกิดโรคเบาหวาน (ประการ เข้มแข็ง และคณะ, 2559)

จากการทบทวนวรรณกรรมที่เกี่ยวข้อง พบว่าปัจจัยปัญหาด้านสุขภาพ ปัญหาด้านอายุ เพศ ดัชนีมวลกาย ความดันโลหิตสูง ประวัติเบาหวานในพ่อแม่ พี่น้อง รอบเอว การสูบบุหรี่ รวมทั้งการดื่มสุรา มีความสัมพันธ์เชิงบวกอย่างมีนัยสำคัญทางสถิติกับโรคเบาหวาน

2.3 เครื่องมือการคัดกรองโรคเบาหวาน

การจำแนกผลลัพธ์และการปฏิบัติการพยาบาลตามระบบสากล วิเคราะห์ข้อมูลด้วยสถิติ ผลการวิจัยพบว่า ระดับคะแนนเฉลี่ยคุณภาพการบันทึกก่อนการทดลองในคะแนนรวมและแต่ละขั้นตอนของกระบวนการพยาบาลอยู่ในระดับต่ำมากเป็นส่วนใหญ่ และหลังทดลองสูงขึ้นเป็นระดับต่ำและปานกลาง ผลการวิจัยครั้งนี้เสนอแนะว่าพยาบาลควรได้รับการฝึกฝนทางคลินิกเพิ่มมากขึ้นพร้อม

กับการนิเทศทางคลินิกจากหัวหน้าหอผู้ป่วย รวมถึงควรใช้ระบบการกำหนดข้อวินิจฉัยการพยาบาล การจำแนกผลลัพธ์ และการปฏิบัติการพยาบาลที่เป็นระบบสากล (รัชฎาพร สุวรรณรัตน์ และคณะ, 2562) แนะนำควรมีระบบคัดกรองโรคเบาหวานในผู้ป่วยทั่วไปด้วยแบบคัดกรองที่มีประสิทธิภาพ ซึ่งสามารถนำไปใช้คัดกรองในสถานพยาบาลหรือชุมชนแต่ละสถานที่ได้ อย่างไรก็ตามเครื่องมือเหล่านี้ใช้แค่เพียงการคัดกรอง ไม่ควรใช้เพื่อการวินิจฉัยโรค

Vigneswari และคณะ (2562) กล่าวถึงการพยากรณ์ภาวะเบาหวานโดยใช้เครื่องมือเหมืองข้อมูล ได้แก่ วิธี Random Forest, Decision Tree C4.5, Random Tree, REPTree และ Logistic Model Tree (LMT) ได้วิเคราะห์ความถูกต้องของแต่ละเครื่องมือได้ ดังตาราง 1

ตาราง 1 ผลลัพธ์ของแต่ละเครื่องมือการพยากรณ์โรคเบาหวาน

เทคนิค	TPR	FPR	Precision	Recall	ROC
Random Forest	0.785	0.312	0.78	0.785	0.838
C4.5	0.762	0.342	0.756	0.762	0.796
Random Tree	0.724	0.309	0.741	0.724	0.708
REPTree	0.755	0.32	0.755	0.755	0.813
LMT	0.793	0.322	0.787	0.793	0.864

Abbas และคณะ (2562) กล่าวถึงการพยากรณ์ภาวะเบาหวานโดยใช้เครื่องมือซัพพอร์ตเวคเตอร์แมชชีน ผลลัพธ์ของการวิจัยมีความแม่นยำเท่ากับ 84.1% และมีค่าความระลึก เท่ากับ 81.1% ดังตาราง 2

ตาราง 2 ผลลัพธ์การประมวลผลของอัลกอลิทึม

เทคนิค	ระยะเวลาในการประมวลผล (วินาที)
Random Forest	0.04
C4.5	0.16
Random Tree	0.03
REPTree	0.16
LMT	0.49

Friedman และคณะ (2558) กล่าวถึงการใช้เครือข่ายเบย์ (Bayes Network) สำหรับการพยากรณ์ภาวะเบาหวานเปรียบเทียบได้ดังตาราง 3

ตาราง 3 ผลลัพธ์ของความแม่นยำของอัลกอลิทึม

เทคนิค	ค่าความแม่นยำ
Proposed Byes Network	72.3%
Naïve Bayes Network	71.5%

มีการนำเครื่องมือเพื่อใช้ทดสอบภาวะเสี่ยงที่ก่อให้เกิดโรคเบาหวานในประเทศไทย โดยรายละเอียดดังตาราง 4

ตาราง 4 ตัวอย่างการนำเทคนิคเหมืองข้อมูลมาใช้ในทางการแพทย์

ผู้วิจัย	งานวิจัย	เครื่องมือ (แบบสอบถาม)
อุมากร ใจยั่งยืน และคณะ, (2565)	ปัจจัยที่มีผลต่อพฤติกรรมการป้องกันโรคเบาหวานชนิดไม่พึ่งอินซูลินในกลุ่มเสี่ยง	ข้อมูลส่วนบุคคล, แบบสอบถามพฤติกรรมการป้องกันโรคเบาหวาน, แบบทดสอบความรู้เกี่ยวกับโรคเบาหวาน, และแบบสอบถามการรับรู้สมรรถนะแห่งตน
วรรณภาภรณ์ จงกลาง และนาฏนภา หีบแก้ว ปัดชา สุวรรณ, (2564)	ปัจจัยที่มีความสัมพันธ์กับกลุ่มเสี่ยงโรคเบาหวานชนิดที่ 2 ของกลุ่มวัยทำงาน อำเภอโนนไทย จังหวัดนครราชสีมา	ข้อมูลส่วนบุคคล, ทักษะความรู้ความเข้าใจ และความรอบรู้ด้านสุขภาพ
สมลักษณ์ เทพสุริยานนท์ และคณะ, (2564)	ปัจจัยทำนายความเสี่ยงต่อภาวะฉุกเฉินจากโรคเบาหวานของผู้สูงอายุโรคเบาหวานในชุมชน	แบบบันทึกขอมูลความเสี่ยงต่อภาวะฉุกเฉินจากโรคเบาหวานของผู้สูงอายุโรคเบาหวาน และแบบประเมินความเสี่ยงต่อภาวะฉุกเฉินจากโรคเบาหวานของผู้สูงอายุโรคเบาหวาน

ตาราง 4 ตัวอย่างการนำเทคนิคเหมืองข้อมูลมาใช้ในทางการแพทย์ (ต่อ)

เขมารดี มาสิงบุญนิ และคณะ, (2562)	ปัจจัยทำนายพฤติกรรมป้องกันโรคเบาหวานในวัยรุ่นที่มีภาวะเสี่ยงต่อโรคเบาหวานในจังหวัดสมุทรปราการ	แบบสอบถามข้อมูลทั่วไป และแบบสอบถามข้อมูลด้านสุขภาพ
ทิพวรรณ นมเนย และวรวรรณ ทิพย์วารีรมย์, (2561)	ปัจจัยที่มีอิทธิพลตอพฤติกรรมการดูแลตนเองของประชาชนกลุ่มเสี่ยงโรคเบาหวานชาวไทยทรงดำ จังหวัดสุโขทัย	ข้อมูลทั่วไปของประชาชนกลุ่มเสี่ยงโรคเบาหวาน, แบบสอบถามเกี่ยวกับโครงสรางทางสังคมและวัฒนธรรม และแบบสอบถามเกี่ยวกับพฤติกรรมการดูแลตนเองของประชาชนกลุ่มเสี่ยงโรคเบาหวาน
สายฝน ม่วงคุ้ม และคณะ, (2560)	ปัจจัยที่มีความสัมพันธ์กับพฤติกรรมป้องกันการเกิดโรคเบาหวานในกลุ่มเสี่ยงต่อการเป็นโรคเบาหวาน	แบบสอบถามข้อมูลทั่วไป, แบบประเมินการรับรู้ความเสี่ยงต่อการเป็นโรคเบาหวาน, แบบประเมินความรู้เกี่ยวกับโรคเบาหวาน, แบบประเมินการรับรู้สมรรถนะของตนเองในการปรับเปลี่ยนพฤติกรรมสุขภาพ และแบบประเมินพฤติกรรมการป้องกันการเกิดโรคเบาหวาน
ชลลดา ไชยกุลวัฒนา และคณะ, (2559)	ปัจจัยทำนายพฤติกรรมสุขภาพของประชำกรกลุ่มเสี่ยงสูงต่อโรคเบาหวาน จังหวัดพะเยา	แบบสอบถามข้อมูลทั่วไป, แบบสอบถามด้านปัจจัยนำ, แบบสอบถามด้านปัจจัยเอื้อ, แบบสอบถามด้านปัจจัยเสริม และพฤติกรรมสุขภาพของกลุ่มเสี่ยงโรคเบาหวาน

ตาราง 4 ตัวอย่างการนำเทคนิคเหมืองข้อมูลมาใช้ในทางการแพทย์ (ต่อ)

สมใจ จางวาง และคณะ, (2559)	ปัจจัยที่มีความสัมพันธ์กับพฤติกรรมการดูแลตนเองเพื่อป้องกันโรคเบาหวานและความดันโลหิตสูงของประชาชนกลุ่มเสี่ยง	ข้อมูลทั่วไป, ความรู้เกี่ยวกับโรคเบาหวานและโรคความดันโลหิตสูง, การรับรู้โอกาสเสี่ยงของการเป็นโรค และพฤติกรรมในการดูแลตนเองเพื่อป้องกันโรคเบาหวานและโรคความดันโลหิตสูง

2.4 แนวคิดเทคนิคการทำเหมืองข้อมูล

เหมืองข้อมูล (Data Mining)

เทคนิคดาต้าไมน์นิงเป็นกระบวนการทางคอมพิวเตอร์ที่ใช้ในการวิเคราะห์ข้อมูลขนาดใหญ่ เป็นวิธีการค้นหาข้อมูลสำคัญเพื่อนำไปใช้ประโยชน์ เทคนิคนี้เป็นที่ยอมรับและถูกใช้อย่างแพร่หลายในหลากหลายวงการ สำหรับข้อมูลทางด้านวิทยาศาสตร์สุขภาพก็มีงานวิจัยมากมายที่ใช้เทคนิคนี้ในการค้นหาคำตอบ หรือข้อสรุปที่เป็นประโยชน์ เทคนิคดาต้า ไมน์นิงถูกนำมาใช้ในการหากฎความสัมพันธ์ (Association Rules Discovery) การแบ่งกลุ่มข้อมูล (Clustering) และการจำแนกประเภทของข้อมูล (Classification) ได้อย่างมีประสิทธิภาพ เทคนิคดาต้า ไมน์นิง หรือ การทำเหมืองข้อมูล คือเทคนิคในการวิเคราะห์ข้อมูลแบบหนึ่ง ซึ่งได้รับความนิยมนำมาใช้ ในการวิเคราะห์ข้อมูลขนาดใหญ่อย่างแพร่หลาย โดยข้อมูลที่ถูกนำมาวิเคราะห์นั้นไม่จำกัดประเภท ดังนั้นเราจึงเห็นผลจากการวิเคราะห์ข้อมูลโดยเทคนิคดาต้า ไมน์นิงถูกนำมาใช้กับข้อมูลทางการตลาด ข้อมูลเกี่ยวกับการการศึกษา ข้อมูลทางบัญชี ข้อมูลที่เกี่ยวข้องกับคอมพิวเตอร์ หรือแม้แต่ข้อมูลด้านสุขภาพ (อ.ภญ.สิริกัลยา เบ็ญจวรรณ์, 2559) ข้อมูลที่เหมือนกันมีความแตกต่างกันอย่างมาก บางส่วนอยู่ในรูปของข้อความ บางส่วนอยู่ในรูปแบบของภาพ เสียง หรือกลุ่มอื่นๆ แต่ข้อมูลสำหรับการตรวจสอบโดยกลยุทธ์การทำเหมืองข้อมูลนั้น ข้อมูลสามารถแบ่งตามโครงสร้างของข้อมูลออกเป็นสองกลุ่ม คือ ข้อมูลที่จัดโดยเฉพาะ (Organized Information) และข้อมูลที่ไม่มีโครงสร้าง (Unstructured Data)

ข้อมูลที่มีการจัดระเบียบ (Organized Information) เป็นข้อมูลที่จัดอยู่ในรูปร่างที่คิดไม่ถึงโดยการรวบรวมจุดประสงค์หรือการจัดเรียงข้อมูลที่เตรียมไว้ล่วงหน้า เพื่อใช้ประโยชน์ข้อมูลที่จัดระเบียบเหล่านี้อาจถูกจัดเก็บในบันทึกที่เกินความคาดหมายหรือในการจัดระเบียบฐานข้อมูลประเภท Excel หรืออยู่ในรูปแบบฐานข้อมูล (Database) ตัวอย่างเช่น ข้อมูลร้านยาที่แสดงในตารางที่ 1 ประกอบด้วย 5 แถว และ 5 คอลัมน์ซึ่งมีคำศัพท์ทางคอมพิวเตอร์เฉพาะเพื่อใช้เรียกข้อมูลแต่ละแถวว่า Sample หรือ Instance โดยแถวแต่ละแถวนั้นแทนข้อมูลตัวอย่างหนึ่งตัวอย่าง คอลัมน์มีชื่อเรียกเฉพาะว่า Attribute หรือ Feature ซึ่งคอลัมน์หนึ่งๆ นั้นแทนกลุ่มหรือประเภทข้อมูลเดียวกันของทุกตัวอย่างจะพูดถึงการทดสอบข้อมูลหนึ่งรายการคอลัมน์เรียกว่าคุณภาพหรือไฮไลต์ โดยที่คอลัมน์จะกล่าวถึงการรวบรวมหรือการจัดเรียงข้อมูลแบบเดียวกันของการทดสอบทั้งหมด การ

วิเคราะห์ข้อมูลด้วยเทคนิคดาต้า ไมน์นิงสามารถแยกได้เป็น 2 ประเภทหลักๆ ได้แก่ เทคนิคการเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) และเทคนิคการเรียนรู้แบบมีผู้สอน (Supervised Learning) ซึ่งมีความแตกต่างกันใน เทคนิคการเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) เป็นการพิจารณาเพื่อค้นหาความสัมพันธ์หรือความเกี่ยวข้องของข้อมูลกลุ่มย่อยจากข้อมูลทั้งหมด ซึ่งเทคนิคนี้ไม่ต้องการแบบแผนของข้อมูลเป็นตัวกำหนดกรอบของการวิเคราะห์ผล สามารถแยกออกได้เป็น 2 เทคนิคย่อย คือ การค้นหากฎความสัมพันธ์ (Association Rule Discovery) และการแบ่งกลุ่มข้อมูล (Clustering) เทคนิคการเรียนรู้แบบมีผู้สอน (Supervised Learning) เป็นการคาดการณ์ หรือทำนายผลของสิ่งที่สนใจจากข้อมูลเดิมที่มีอยู่ เทคนิคนี้ต้องการแบบแผนการวิเคราะห์ข้อมูลเป็นหลักในการทำงาน ดังนั้น โมเดลหรือสมการจะถูกสร้างขึ้นจากข้อมูลเดิม เพื่อใช้เป็นแบบแผนในการวิเคราะห์ข้อมูลต่อไป เทคนิคนี้สามารถแยกออกได้เป็น 2 เทคนิคย่อย คือการจำแนกประเภทข้อมูล (Classification) และการประมาณค่าของข้อมูล (Regression)

ขั้นตอนการทำเหมืองข้อมูลที่สามารถใช้ในการตรวจสอบข้อมูลได้แก่ การหากฎความสัมพันธ์ (Association Rules Discovery) การแบ่งกลุ่มข้อมูล (Clustering) และการจำแนกประเภทข้อมูล (Classification) มีหลักการทำงานในแต่ละส่วน ดังนี้

1) การหากฎความสัมพันธ์ (Association Rules Discovery) คือการค้นหาความสัมพันธ์ระหว่างชุดข้อมูลตั้งแต่สองชุดขึ้นไปจากชุดข้อมูลที่มีอยู่มากมาย การเชื่อมต่อเหล่านี้มีความน่าสนใจเป็นพิเศษ ตัวอย่างที่มักถูกหยิบยกขึ้นมาเพื่ออธิบายถึงกฎความสัมพันธ์คือพฤติกรรมของผู้ซื้อที่ Wal-Mart พิจารณา ซึ่งอาจเป็นบริษัทค้าปลีกที่ขยายตัวภายในรัฐที่เข้าร่วมและหลายประเทศทั่วโลก เห็นด้วยกับการศึกษาหนึ่งครั้งในเย็นวันศุกร์จะมีลูกค้ามารวมตัวกัน ซึ่งอาจจะเป็นผู้ชายเข้ามาซื้อของพร้อมกัน 2 ชิ้นอย่างต่อเนื่อง นั่นคือ เบียร์สดและผ้าอ้อม โดยสินค้าทั้ง 2 ชิ้นวางอยู่หลายมุมของร้านค้าสำนักงาน และไม่สามารถเกี่ยวข้องได้เลย แต่ลองคิดให้ลึกซึ้งยิ่งขึ้น พบว่าลูกค้าเหล่านี้เป็นพ่อที่มีลูกเล็กๆ และไม่ได้ออกไปเที่ยวกับเพื่อนในคืนวันศุกร์ ดังนั้นควรซื้อเบียร์ไว้ดื่มในช่วงท้ายสัปดาห์ เมื่อรู้เช่นนั้น แผนกร้านค้าก็นำพัสดุของทั้งสองประเภทมาวางไว้ในตำแหน่งที่มองเห็นได้ง่าย เพื่อขยายโอกาสที่ลูกค้าจะซื้อของให้มากขึ้น

2) การแบ่งกลุ่มข้อมูล (Clustering) อาจเป็นกลยุทธ์ในการวิเคราะห์ข้อมูลโดยแยกชุดข้อมูลออกเป็นกลุ่มตามลักษณะบางอย่างที่เหมือนกันหรือเปรียบเทียบกับข้อมูลเหล่านั้นเป็นตัวเรียงลำดับ และจัดระเบียบข้อมูลเฉพาะเป็นกลุ่มๆ ลักษณะของข้อมูลที่สามารถรวบรวมได้มักจะเป็นเวกเตอร์ ซึ่งเป็นข้อมูลเชิงคณิตศาสตร์ ที่มีทั้งขนาดและทิศทาง กลวิธีที่นิยมใช้ในการแยกส่วนข้อมูลได้แก่การหาระยะห่างระหว่างข้อมูลการแบ่งกลุ่มข้อมูลด้วย วิธี K-Means และการแบ่งกลุ่มโดย K-Means Agglomerative Clustering วิธีการวิเคราะห์แต่ละกลยุทธ์มีความซับซ้อน ยกตัวอย่างการเก็บรวบรวมข้อมูลพฤติกรรมของผู้บริโภคในห้างสรรพสินค้าแห่งหนึ่ง พวกเขาได้เก็บรวบรวมข้อมูลการซื้อของลูกค้าผ่านการใช้บัตรคลับการ์ด และทำการแยกกลุ่มลูกค้าเพื่อจัดส่งข้อมูลโปรโมชันที่ตรงกับพฤติกรรมการใช้จ่ายของลูกค้าแต่ละราย การทำเช่นนี้เป็นการขยายไปสู่การขยายดีลเฉพาะกลุ่มลูกค้าและทำกำไรให้กับห้างสรรพสินค้า คลังสินค้าในการจัดหาสินค้าแต่ละประเภทให้เพียงพอต่อความต้องการของลูกค้า แล้วยังส่งผลเชิงบวกด้านทัศนคติของลูกค้าที่มีต้องห้างสรรพสินค้าอีกด้วย

เนื่องจากลูกค้ารู้สึกว่าได้รับการดูแลจากห้างสรรพสินค้าและได้รับสิทธิพิเศษที่ตรงตามความต้องการอย่างต่อเนื่อง

3) การจำแนกประเภทข้อมูล (Classification) คือ การนำข้อมูลที่มีอยู่มาสร้างเป็นรูปเป็นรูปแบบสำหรับจัดการ เพื่อใช้ศึกษากรอบในการจดจำเรียกว่า การสร้างเงื่อนไขหรือแบบจำลอง (Model) เพื่อให้มีความสามารถในการคาดการณ์หรือคาดการณ์สิ่งที่สนใจได้อย่างแม่นยำ มีกลยุทธ์การสร้างแบบจำลองที่แพร่หลายมากมายได้แก่ วิธี Decision Tree วิธี Naive Bayes วิธี K-Nearest Neighbors และวิธี Neural Network ยกตัวอย่างจำแนกประเภทของข้อมูล เช่นการพยากรณ์อากาศหรือทำนายสภาพอากาศตามตัวเลขภูมิอากาศ มักเกิดจากการรวบรวมข้อมูลตามภูมิภาคที่กำหนดในแต่ละวัน เป็นเวลานานหรือการเก็บข้อมูลหลายปี แล้วมีการใช้รูปแบบการแสดงความน่าจะเป็นของสภาพอากาศตามองค์ประกอบที่เกี่ยวข้อง ไม่ว่าจะเป็นปริมาณเมฆ อุณหภูมิ ความชื้นในอากาศ ความเร็วลม ฯลฯ เมื่อเตรียมโมเดลที่แน่นอนสำหรับการคาดการณ์สภาพอากาศแล้ว สามารถนำข้อมูลองค์ประกอบที่เกี่ยวข้องในช่วงเวลาหนึ่งๆ มาจัดทำเพื่อคาดการณ์สภาพอากาศที่จะเกิดขึ้นในอนาคตได้

วัตถุประสงค์ในการใช้เหมืองข้อมูล

1. เพื่อการค้นพบองค์ความรู้ใหม่ในฐานข้อมูล (Knowledge Discovery In Databases)
2. เพื่อการสกัดองค์ความรู้ที่ซ่อนเร้นอยู่ (Knowledge Extraction)
3. เพื่อจัดการกับข้อมูลในอดีต (Data Archeology)
4. เพื่อสำรวจข้อมูล (Data Exploration)
5. เพื่อค้นหา Pattern ของข้อมูลที่ซ่อนอยู่ (Data Pattern Processing)
6. เพื่อใช้ขุดเจาะข้อมูล (Data Dredging)
7. เพื่อเก็บเกี่ยวผลประโยชน์ให้ได้มาซึ่งสารสนเทศที่มีประโยชน์

ขั้นตอนการทำเหมืองข้อมูล

1. Data Cleaning เป็นขั้นตอนสำหรับการคัดข้อมูลที่ไม่เกี่ยวข้องออกไป
2. Data Integration เป็นขั้นตอนการรวมข้อมูลที่มีหลายแหล่งให้เป็นข้อมูลชุดเดียวกัน
3. Data Selection เป็นขั้นตอนการดึงข้อมูลสำหรับการวิเคราะห์จากแหล่งที่บันทึกไว้
4. Data Transformation เป็นขั้นตอนการแปลงข้อมูลให้เหมาะสมสำหรับการใช้งาน
5. Data Mining เป็นขั้นตอนการค้นหารูปแบบที่เป็นประโยชน์จากข้อมูลที่มีอยู่
6. Pattern Evaluation เป็นขั้นตอนการประเมินรูปแบบที่ได้จากการทำเหมืองข้อมูล
7. Knowledge Representation เป็นขั้นตอนการนำเสนอความรู้ที่ค้นพบ

เทคนิคการทำเหมืองข้อมูลในงานวิจัย

1. เทคนิคเพอร์เซ็ปตรอนหลายชั้น (Multi-layer Perceptron)

วิบล ญึก และจารี ทองคำ (2561) กล่าวถึง Multi-Layer Perceptron (MLP) ว่าเป็นเทคนิคที่มีรูปแบบเครือข่ายประสาทเทียม (Artificial Neural Network: ANN) แบบ Feedforward ซึ่งจะจับคู่ระหว่างชุดข้อมูลนำเข้าและชุดของ ผลลัพธ์ที่เหมาะสม เทคนิคนี้ประกอบด้วย โหนดหลายชั้นในกราฟกำกับโดยแต่ละชั้นจะเชื่อมต่อกับโหนดหนึ่งต่อไปอีกโหนด แต่ละโหนดจะเป็นเซลล์ประสาทหรือส่วนการประมวลผลที่มีฟังก์ชันการเปิดใช้งานแบบไม่เชิงเส้น MLP ได้ทั้งการจำแนกและวิเคราะห์ข้อมูลการถดถอยเชิงเส้นตรง

สายชล สินสมบูรณ์ทอง (2563) กล่าวว่าวิธีโครงข่ายประสาทเทียมใช้อัลกอริทึมชนิดเพอร์เซปตรอนหลายชั้น (Multilayer Perceptron) กำหนดค่าอัตราการเรียนรู้ (Learning Rate) เป็น 0.1 ค่าโมเมนตัม (Momentum) เป็น 0.9 จำนวนรอบการสอน (Training Time) 20,000 รอบ อัลกอริทึมของวิธีโครงข่ายประสาทเทียมชนิดเพอร์เซปตรอนหลายชั้นที่มีชั้นซ่อน (Hidden Layer) 1 ชั้น แม้ว่าโครงสร้างโครงข่ายประสาทเทียมที่ซับซ้อนสามารถมีชั้นซ่อนมากกว่า 1 ชั้น แต่ในทางปฏิบัตินั้นการกำหนดชั้นซ่อน 1 ชั้น ก็เพียงพอต่อการวิเคราะห์ข้อมูล

อนุวัฒน์ เปพาทย์ และคณะ (2565) กล่าวว่าโครงข่ายประสาทเทียมเพอร เซ็ปตรอนหลายชั้นเป็นโครงข่ายประสาทเทียมแบบ Feed Forward ซึ่งเป็นแบบจำลองทางคณิตศาสตร์ที่ใชอย่างแพรหลายและได้รับความสนใจนำมาศึกษาวิจัยอย่างมากในหลากหลายสาขาวิชา พัฒนาขึ้นเพื่อจำลองการทำงานของโครงข่ายประสาทในสมองมนุษย์ ประกอบด้วยเซลล์ประสาทเทียมหรือโหนดจำนวนมาก เชื่อมต่อกัน ซึ่งการเชื่อมตอแบงออกเป็นกลุ่มย่อยเรียกวา ชั้น (Layer) ชั้นแรกเป็นชั้นรับขอมูลปอนเขา (Input Layer) ส่วนชั้นสุดท้ายเรียกวา ชั้นส่งขอมูลออก (Output Layer) และชั้นที่อยู่ระหว่างชั้นรับขอมูลปอนเข้าและชั้นส่งข้อมูลออกเรียกวา ชั้นซอน (Hidden Layer) ซึ่งโดยทั่วไปชั้นซอนอาจมีมากกวา 1 ชั้นก็ได้ ดวยเหตุนี้ จึงสามารถแบ่งประเภทของโครงข่ายประสาทเทียมตามจำนวนชั้นของโครงข่ายแบบกว้างๆ ได้ 2 แบบ ได้แกโครงข่ายแบบชั้นเดียว (Single Layer) และ โครงข่ายแบบหลายชั้น (Multilayer)

จากการศึกษาพบว่าเทคนิคเพอร์เซ็ปตรอนแบบหลายชั้น (Multilayer Perceptron) เป็นการคำนวณเพอร์เซ็ปตรอนหลาย ชั้นสามารถคำนวณแบบง่ายๆ โดยใช้ XOR ค่าที่ใช้ในการจัดเรียงแต่ละจุดของหน่วยที่ได้รับ รวมค่าของช่องว่างการแลกเปลี่ยน ถึงจุดคุ้มทุนกับ 1 ค่าความนิยมต่อหน่วยผลตอบแทนของระดับ 1 จะอยู่ในโอกาสที่ค่าประมาณที่คำนวณได้ในกรณีที่ค่าประมาณที่คำนวณได้นั้นมีความสำคัญมากกว่าการแลกเปลี่ยน หน่วยผลตอบแทนของเลเยอร์ 1 จะเป็น 1 การคำนวณแบบหลายชั้นนี้ ผลตอบแทนของระดับ 1 จะเป็นอินพุตของระดับ 2 และผลตอบแทนของระดับ 2 จะเป็นอินพุตของระดับ 3 ตามลำดับ

2. เทคนิคนาอีฟ เบย์ (Naïve bayes)

จุฑาทิพย์ ทิพย์พูล และนิเวศ จิระวิชิตชัย (2559) กล่าวว่าวิธีการของนาอีฟเบย์เป็นวิธีการเรียนรู้ที่ใช้หลักของความน่าจะเป็นตามกฎทฤษฎีของนาอีฟเบย์เข้ามาช่วยในการเรียนรู้เพื่อหาสมมติฐานหนึ่งๆ ร่วมกับข้อมูลการเรียนรู้แบบนาอีฟเบย์ อาศัยหลักการของการคำนวณความน่าจะเป็นของแต่ละ สมมติฐาน โดยการเรียนรู้แบบนาอีฟเบย์เป็นการเรียนรู้เพิ่มได้ เนื่องจากตัวอย่างใหม่ที่ได้มาถูกนำมาปรับเปลี่ยนการแจกแจง ซึ่งมีผลต่อการเพิ่มหรือลดความน่าจะเป็น ทำให้มีการเรียนรู้ที่เปลี่ยนไป

พิพัฒน์ ลิ้มประไพพงษ์ และสมชาย เล็กเจริญ (2560) กล่าวว่านาอีฟเบย์ถูกนำมาใช้อย่างแพร่หลายในงานจำแนกหมวดหมู่เอกสาร มีประสิทธิภาพการจำแนกที่ดี ซึ่งอาศัยหลักความน่าจะเป็นในการทำนายผลลัพธ์ โดยการวิเคราะห์ความสัมพันธ์ระหวางตัวแปรเพื่อใช้ในการสร้างเงื่อนไขความน่าจะเป็นสำหรับแต่ละความสัมพันธ์

อนัตต์ชัย ชุติภาสเจริญ และดร.จรัญ แสนราช (2561) กล่าวถึงนาอีฟเบย์ว่านาอีฟเบย์ คือเครื่องจักรเรียนรู้ที่อาศัยหลักการความน่าจะเป็นตามทฤษฎีของเบย์ (Bayes Theorem) ซึ่งมีอัลกอริทึมที่ไม่ซับซ้อน เป็นขั้นตอนวิธีในการจำแนกข้อมูลโดยการเรียนรู้ปัญหาที่เกิดขึ้น เพื่อนำมาสร้างเงื่อนไขการจำแนกข้อมูลใหม่ หลักการของนาอีฟเบย์ใช้การคำนวณหาความน่าจะเป็นในการทำนายผลเป็นเทคนิคในการแกปัญหาแบบจำแนกประเภทที่สามารถคาดการณ์ผลลัพธ์ได้จะทำการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปร เพื่อใช้ในการสร้างเงื่อนไขความน่าจะเป็นสำหรับแต่ละความสัมพันธ์เหมาะกับกรณีของเซตตัวอย่างที่มีจำนวนมากและคุณสมบัติ (Attribute) ของตัวอย่างไม่ขึ้นต่อกัน

จากการศึกษาพบว่าเทคนิคนาอีฟ เบย์ (Naïve bayes) เป็นกลุ่มของเทคนิคจำแนกประเภทที่รวดเร็วและเรียบง่าย ซึ่งมักจะเหมาะสำหรับชุดข้อมูลที่มีมิติสูงมาก เนื่องจากมีความรวดเร็วและมี พารามิเตอร์ที่ปรับแต่งได้น้อยมาก จึงมีประโยชน์อย่างมากในฐานะพื้นฐานที่รวดเร็วสำหรับปัญหาการจำแนกประเภท

3. เทคนิคโลจิสติกส์ (Logistic Regression)

สุรวัชร ศรีเปารยะ และสายชล สินสมบูรณ์ทอง (2560) กล่าวว่าวิธีการถดถอยโลจิสติก (Logistic Regression) ใช้เพื่อหาความสัมพันธ์ระหว่างตัวแปรอิสระและตัวแปรตาม โดยตัวแปรตามมีเพียงสองค่า คือ 0 และ 1 หากตัวแปรอิสระมีค่าน้อย ค่าของตัวแปรตามจะมีค่าเท่ากับ 0 และหากค่าตัวแปรอิสระมีค่ามาก ค่าของตัวแปรตามจะมีค่าเท่ากับ 1

กัลยา บุญหล้า และวรวุฒิ มหาโพธิ์ (2565) กล่าวว่า วิธีการถดถอยโลจิสติก เป็นการวิเคราะห์เพื่อทำนายโอกาสที่เหตุการณ์ที่สนใจจะเกิดขึ้นและสมการถดถอยโลจิสติกที่ดีจะต้องประกอบด้วยตัวแปรอิสระที่เหมาะสมที่จะทำให้ค่าทำนายโอกาสที่จะเกิดนั้นใกล้เคียงกับความเป็นจริง ที่ตัวแปรตามเป็นข้อมูลเชิงคุณภาพมีค่าได้เพียง 2 ค่า ส่วนตัวแปรอิสระอาจจะเป็นข้อมูลเชิงปริมาณหรือข้อมูลเชิงคุณภาพ จะใช้ประมาณโอกาสที่จะเกิดเหตุการณ์ ผลของการวิเคราะห์ขึ้นอยู่กับปัจจัยที่สำคัญ คือการใช้แบบจำลองที่เหมาะสมในการเลือกปัจจัยที่สำคัญ การแบ่งกลุ่มย่อยและการแบ่งกลุ่มย่อยและจำนวนตัวอย่างในแต่ละกลุ่มย่อยของปัจจัยนั้น โดยมีเป้าหมายหลักเพื่ออธิบายปัจจัยที่สำคัญ การแบ่งกลุ่มย่อยและจำนวนตัวอย่างในแต่ละกลุ่มย่อยของปัจจัยนั้น

(ปริยานุช ประเสริฐสิริกุล และคณะ (2565) กล่าวว่าวิธีการถดถอยโลจิสติกเป็นการอาศัยหลักการความน่าจะเป็นของผลลัพธ์การทำนายว่าจะเกิดขึ้นหรือไม่โดยขึ้นอยู่กับคุณลักษณะ สมการของ Logistic Regression เป็นฟังก์ชัน Sigmoid ที่มีลักษณะเป็นเส้นโค้งรูป S โดยมีค่าระหว่าง 0 ถึง 1

จากการศึกษาพบว่าเทคนิคโลจิสติก (Logistic Regression) เป็นเทคนิคการวิเคราะห์ความสัมพันธ์ของตัวแปรในรูปของการทำนายโอกาสหรือความน่าจะเป็น ของเหตุการณ์หรือการไม่เกิดขึ้นของเหตุการณ์ที่สนใจ สำหรับกรณีที่ตัวแปรรองอาจเป็นตัวแปรเชิงหมวดหมู่ และตัวแปรอิสระ

อาจเป็นตัวแปรเชิงปริมาณหรือตัวแปรพวงก็ได้ เป็นการวิเคราะห์ความถดถอยแบบเส้นตรงทั่วไป ใช้ในการพยากรณ์ผลลัพธ์ของ 2 ตัวแปร เช่น Yes กับ No และ 0 กับ 1 แต่เนื่องจากตัวแปรตาม (Dependent Variable) มีค่าเพียง 2 อย่างเท่านั้น จึงไม่สามารถสร้างแบบจำลอง (Model) ได้สำหรับการวิเคราะห์แบบ Logic Regression ดังนั้นแทนที่จะทำการพยากรณ์โดยอาศัยเพียงค่าของตัวแปรตามที่ได้เราจะสร้าง Model โดยอาศัย Algorithm ของความน่าจะเป็นของการเกิดเหตุการณ์ เราเรียก Algorithm นี้ว่า Log Odds หรือ Logic Transformation ตามสมการ 1

$$\text{อัตราส่วนความน่าจะเป็น} = \frac{\text{ความน่าจะเป็นที่จะเกิดเหตุการณ์}}{\text{ความน่าจะเป็นที่จะไม่เกิดเหตุการณ์}} \quad (1)$$

4. เทคนิคแรนดอมฟอเรส (Random forest)

ธนัท จรณะสมบูรณ์ (2561) กล่าวว่า เทคนิคแรนดอมฟอเรส เป็นแบบจำลองอีกการทำนายที่ใช้พื้นฐานจากต้นไม้ตัดสินใจ (Decision Tree) เป็นการทำนายแบบชุดของ Decision Tree หลายๆ ต้น (Ensemble of Decision Trees) โดยสร้างจากการสุ่มข้อมูลตัวอย่างแบบเลือกแล้วใส่กลับ (random sampling with replacement) เพื่อนำมาสร้างเป็นแบบจำลองต้นไม้โดยแต่ละต้นมีลักษณะที่ไม่ซ้ำกัน โดยแต่ละแบบจำลองจะมีการทำนายผล ซึ่งผลจากการทำนายของต้นไม้แต่ละต้นจะทำการโหวตเลือกผลการทำนายที่ได้รับการโหวตมากที่สุด

อัครพล พลูสวัสดิ์ และจรัญ แสนราช (2562) กล่าวว่าเทคนิคแรนดอมฟอเรส เป็นชุดของการจำแนกประเภทแบบไม่ตัดแต่งกิ่งหรือต้นไม้ถดถอย ซึ่งถูกสร้างจากการนำข้อมูลฝึกสอนไปสุ่มตัวอย่างข้อมูลและคุณลักษณะข้อมูลแล้วนำมาสร้างเป็นต้นไม้ตัดสินใจ ซึ่งมีตัวอย่างส่วนหนึ่งที่ไม่ถูกเลือกเรียกข้อมูลส่วนนี้ว่า Out-of-Bag จะถูกนำมาในการทดสอบต้นไม้ตัดสินใจ

สุภัสสรา สมเจตนา และจารี ทองคำ (2564) กล่าวว่าเทคนิคแรนดอมฟอเรส เป็นเทคนิคการสุ่มเลือกใช้ข้อมูลและคุณลักษณะต้นไม้ตัดสินใจ ซึ่งถูกสร้างจากการนำข้อมูลไปสุ่มเลือกตัวอย่างแบบเลือกแล้วใส่กลับ แล้วนำมาสร้างเป็นต้นไม้ ซึ่งจะมีตัวอย่างส่วนหนึ่งที่ไม่ถูกเลือก ซึ่งข้อมูลส่วนนี้เรียกว่า Out-of-Bag (OOB) จะถูกนำมาใช้ในการทดสอบ วิธีการดังกล่าวนี้เรียกว่า Bagging ผลลัพธ์ที่ได้อย่างอิสระจาก Decision Tree ในแต่ละต้นถูกนำมาคิดเป็นผลการโหวต ผลโหวตที่มากที่สุดจะใช้ระบุสถานะของคลาสเทคนิค Random Forest ไม่จำเป็นต้องมีข้อมูลทดสอบเพื่อประมาณความผิดพลาดเพราะข้อมูล OOB นั้นถูกนำมาใช้ทดสอบ Decision Tree นั้นแล้ว

จากการศึกษาพบว่าเทคนิคแรนดอมฟอเรส ถูกสร้างขึ้นมาเพื่อปัญหาการจำแนกและสมการถดถอยมีลักษณะคล้ายเทคนิคต้นไม้การตัดสินใจอันเนื่องมาจากการรวมกันของต้นไม้อันเกิดจากการฝึกข้อมูลของบรรจุถุง (Bagging) หรือ บูตสแตรป (Bootstrap) ความแตกต่างหลักระหว่างเทคนิคต้นไม้การตัดสินใจและเทคนิคป่าสุ่มคือการสร้างโหนดรูทและการแยกโหนดจะทำแบบสุ่มในภายหลัง พบว่าเทคนิคป่าสุ่มจะใช้วิธีการบรรจุถุงเพื่อสร้างการคาดการณ์ที่จำเป็นบรรจุถุงจะมีส่วนช่วยในการสุ่มชุดข้อมูลฝึกจำนวนหนึ่งจะประกอบด้วยการสังเกตและคุณลักษณะที่ใช้ในการทำนายเพื่อผลิตผลลัพธ์ในรูปแบบต้นไม้การตัดสินใจโดยขึ้นอยู่กับชุดข้อมูลฝึกที่ป้อนเข้าไปในเทคนิคป่าสุ่มซึ่ง

จะเรียงลำดับผลลัพธ์ที่สูงที่สุดจะเป็นผลลัพธ์สุดท้ายในเทคนิคป่าสุ่มของเทคนิคการจำแนกประเภทใช้วิธีการทั้งมวล (Ensemble) เพื่อให้ได้ผลลัพธ์ชุดข้อมูลฝึกจะถูกป้อนเพื่อฝึกแบบต้นไม้การตัดสินใจต่างๆ ซึ่งข้อมูลชุดนี้ประกอบด้วยการสังเกตและคุณลักษณะที่จะถูกเลือกแบบสุ่มในระหว่างการแยกโหนด ระบบป่าฝนอาศัยต้นไม้ตัดสินใจต่างๆ แผนผังการตัดสินใจ ทั้งหมดประกอบด้วยโหนดการตัดสินใจ โหนดปลายสุดและโหนดราก โหนดปลายสุดของแผนผังต้นไม้แต่ละต้นเป็นผลลัพธ์สุดท้ายที่สร้างโดยแผนผังการตัดสินใจการเลือกผลลัพธ์สุดท้าย

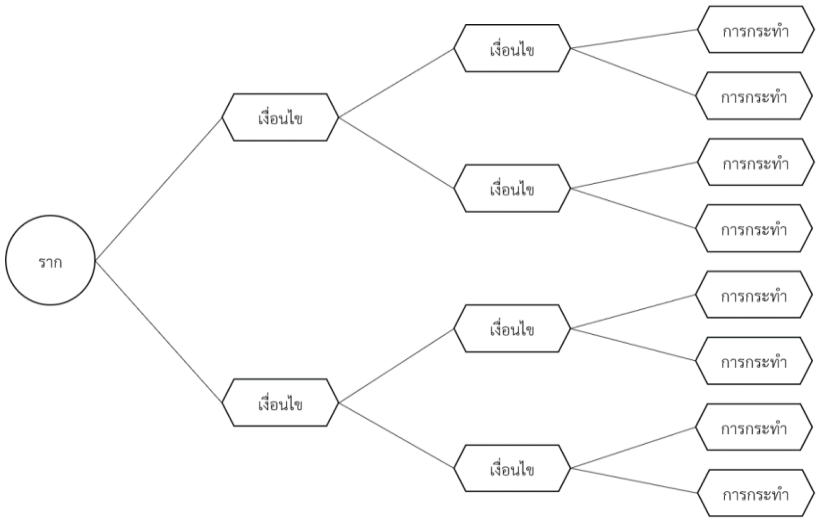
5. เทคนิคต้นไม้ตัดสินใจ (Decision Trees)

วิชญ์วิสิฐ เกสรสิทธิ์ และคณะ (2561) กล่าวว่าต้นไม้ตัดสินใจเป็นกฎการตัดสินใจเพื่อหาทางเลือกที่ดีที่สุด กฎการทำนายโดยใช้เทคนิคต้นไม้การตัดสินใจจะจำแนกการกลับมารักษาซ้ำของผู้ป่วยที่มีร้อยละของผลบวกเท็จ ต่ำกว่าร้อยละผลลบเท็จ และมีร้อยละของการจำแนกกลุ่มได้ถูกต้องสูงกว่า ผลการวิจัยพบว่าการแก้ปัญหาข้อมูลไม่สมดุลด้วยวิธีสังเคราะห์ข้อมูลใหม่จะมีประสิทธิภาพดีที่สุดในการจำแนกข้อมูล

ภูมิพัฒน์ ดวงกลาง และรัจนา เครือแก้ว (2562) กล่าวว่าต้นไม้ คือเทคนิคการสร้างแบบจำลองที่ใช้ช่วยในการตัดสินใจ โดยแบบจำลองจะอยู่ในรูปแบบโครงสร้างต้นไม้ (Tree) โดยมีใบ (Leaf) เป็นสิ่งที่อยู่ล่างสุดบ่งบอกถึงชุดข้อมูลคำตอบ (Class) ซึ่งเป็นผลลัพธ์สุดท้ายที่ผ่าน การทดสอบเงื่อนไขตามคุณลักษณะของแต่ละโหนด (Node) ตามเส้นทางของกิ่งต้นไม้ (Branch)

อัชฌาพร กว้างสวาสดิ์ และคณะ (2562) กล่าวว่าต้นไม้ตัดสินใจเป็นแบบจำลองทางคณิตศาสตร์ เพื่อหาทางเลือกที่ดีที่สุด โดยการนำเอาข้อมูลมาสร้างแบบจำลองการพยากรณ์ในรูปแบบโครงสร้างแบบต้นไม้ โดยจะมีการเรียนรู้ข้อมูลแบบมีผู้สอน (Supervised Learning) สามารถสร้างแบบจำลองการจัดหมวดหมู่ (Clustering) ได้จากกลุ่มตัวอย่างที่มีการกำหนดข้อมูลไว้ล่วงหน้า (Training Set) ได้โดยอัตโนมัติและสามารถพยากรณ์กลุ่มของข้อมูลที่ยังไม่เคยนำมาจัดหมวดหมู่ได้อีกด้วย

จากการศึกษาพบว่าเทคนิคต้นไม้การตัดสินใจ (Decision Tree) สร้างการจำแนกประเภท หรือแบบจำลองการถดถอยในรูปแบบของโครงสร้างต้นไม้ โดยจะแบ่งชุดข้อมูลออกเป็นชุดย่อยที่เล็ก ลงและเล็กลง เป็นการนำข้อมูลมาสร้างแบบจำลองเพื่อการตัดสินใจแบบ Tree ซึ่งการที่จะตัดสินใจทำอะไรอย่างใดอย่างหนึ่งขึ้นอยู่กับเงื่อนไขที่มี โดยเงื่อนไขก็จะประกอบไปด้วยตัวแปรต่างๆ ที่เป็นปัจจัยในการตัดสินใจ Decision Tree จะประกอบด้วย ราก (Root) เป็นจุดเริ่มต้นของลำดับการตัดสินใจ ถัดมาจะเป็น Node ซึ่งเป็นกิ่งก้านสาขาของต้นไม้ คือเงื่อนไขของระบบ สุดท้ายจะเป็น Leaf ซึ่งจะหมายถึงการกระทำหรือการปฏิบัติตามเงื่อนไข ดังภาพประกอบ 3



ภาพประกอบ 3 แสดงโครงสร้าง Decision Tree

จากตัวอย่างโครงสร้าง Decision Tree ในข้างต้นจะเห็นว่าการเลือกใช้เทคนิค Decision Tree เงื่อนไข ต้องมีมากกว่า 1 เงื่อนไข และการกระทำที่ถูกเลือกจะขึ้นอยู่กับเงื่อนไขที่ผ่านมาด้วย ข้อจำกัดของ Decision Tree การแบ่งกลุ่มแบบ Decision Tree กรณีเป็นข้อมูลที่มีค่าต่อเนื่อง เช่น ข้อมูลรายได้ ข้อมูลราคา ต้องทำการแปลงให้อยู่ในช่วงหรือตัดเป็นกลุ่มก่อน เมื่อ Algorithm เลือกว่าจะใช้ค่าไหนเป็นตัวแบ่งกลุ่มแล้วก็จะไม่สนใจค่าอื่นที่อาจมีความสำคัญเช่นเดียวกัน การจัดการกับข้อมูลที่ไม่ทราบค่า อาจมีผลกระทบกับผลลัพธ์ของ Decision Tree ที่มีระดับชั้นมากเกินไป จะทำให้ข้อมูลที่ผ่าน Node แตกออกเป็นชิ้นเล็กชิ้นน้อย ซึ่งข้อมูลเหล่านั้น จะไม่มีประโยชน์ในการนำมาใช้ทำการวิเคราะห์ ปัญหาเรื่อง Over fitting / Overtraining เกิดจากการที่แบบจำลองได้เรียนรู้เข้าไปถึงรายละเอียดของข้อมูล มากเกินไปจะทำให้เกิด Node ที่เป็นส่วนเฉพาะเจาะจงกับกลุ่มข้อมูลที่ใช้ในการเรียนรู้ ซึ่งจะต้องหาวิธีการในการตัดกิ่งนี้ออกไป

ในปัจจุบันด้านการแพทย์ได้นำ Data Mining มาช่วยวิเคราะห์อาการของคนไข้วิเคราะห์การจ่ายยาพยากรณ์แนวโน้มการเกิดโรคระบาด การประยุคใช้เหมือนข้อมูลทางการแพทย์ (รุ่งโรจน์ บุญมา และนิเวศ จิระวิชิตชัย, 2562) วัตถุประสงค์ของงานวิจัยนี้คือการสร้างแบบจำลองการจำแนกประเภทผู้ป่วยโรคเบาหวานโดยใช้เทคนิคเหมืองข้อมูลและการเลือกคุณลักษณะจากความสัมพันธ์ของข้อมูลและทำการเปรียบเทียบประสิทธิภาพของแบบจำลองของเทคนิคเหมืองข้อมูล 4 ประเภท ได้แก่ เนอีฟเบย์, เคเนียเรสเนเบอร์, ต้นไม้ตัดสินใจ และซัพพอร์ทเวกเตอร์แมชชีน จากการทดลองพบว่า ซัพพอร์ทเวกเตอร์แมชชีน มีประสิทธิภาพการทำนายสูงสุดคิดเป็นร้อยละ 76.95% สามารถนำผลที่ได้จากงานวิจัยนี้ไปประยุกต์ใช้ในการคัดกรองและสร้างระบบสนับสนุนการตัดสินใจในส่วนของแนวทางการรักษาของแพทย์ต่อไป

2.5 งานวิจัยที่เกี่ยวข้อง

มีการประยุกต์เทคนิคเหมืองข้อมูลในทางการแพทย์สามารถสรุปได้ดังตาราง 5

ตาราง 5 การประยุกต์เทคนิคเหมืองข้อมูลในทางการแพทย์

ปีที่ตีพิมพ์	ผู้วิจัย	งานวิจัย	เทคนิคเหมืองข้อมูล
2565	นพรัตน์ นนท์ศิริ พิศณุ ชัยจิตวณิชกุล กริช สมกันธา	การจำแนกข้อมูลเพื่อวินิจฉัยความเสี่ยงการเป็นโรคเบาหวานโดยใช้เทคนิควิธีแบบร่วมกันตัดสินใจและวิธีเลือกคุณลักษณะเด่นไปข้างหน้า	Decision Tree, Random Forest, NaïveBayes, Support Vector Machine และ K-Nearest Neighbor
2564	เพชรรัตน์ ม่วงน้อย จักรพันธ์ พลาผล ภรัณยา ปาลวิสุทธิ์	ตัวแบบประเมินภาวะความเสี่ยงการเป็นโรคซึมเศร้าของนักศึกษาด้วยเทคนิคเหมืองข้อมูล	Decision tree
2564	นงเยาว์ ในอรุณ	การเปรียบเทียบประสิทธิภาพของแบบจำลองการทำนายความเสี่ยงโรคหัวใจและหลอดเลือดโดยใช้อัลกอริทึมเหมืองข้อมูล	Neural Network, Random Forest, NaïveBayes, K-Nearest Neighbors และ Decision Tree
2562	รุ่งโรจน์ บุญมา นิเวศ จิระวิชิตชัย	การจำแนกประเภทผู้ป่วยโรคเบาหวานโดยใช้เทคนิคเหมืองข้อมูลและการเลือกคุณลักษณะจากความสัมพันธ์ของข้อมูล	Decision tree, Support Vector Machine, NaïveBayes และ K-Nearest Neighbors
2562	ตัวช่วย ดวงแก้ว สายัญ สายยศ	การวิเคราะห์ปัจจัยที่มีผลต่อพัฒนาการล่าช้าตามช่วงอายุในเด็กปฐมวัยด้วยเทคนิคเหมืองข้อมูล	Decision tree, K-Nearest Neighbors และ NaïveBayes

ตาราง 5 การประยุกต์เทคนิคเหมืองข้อมูลในทางการแพทย์ (ต่อ)

2562	ณัฐวดี หงส์บุญมี ประภาสิริ ตรีพาณิชกุล	การเปรียบเทียบประสิทธิภาพการจำแนกข้อมูลเพื่อวิเคราะห์ปัจจัยความเสี่ยงที่ส่งผลต่อการเกิดโรคไฮเปอร์ไทรอยด์ด้วยเทคนิคเหมืองข้อมูล	Bayes' Theorem, Decision tree และ K-fold cross validation
2561	วนิดา พงษ์สงวน ทิพยา ถินสูงเนิน มาโนช ถินสูงเนิน	การพัฒนาแบบจำลองปัจจัยที่มีผลต่อการเป็นโรคเบาหวานด้วยเทคนิคต้นไม้ตัดสินใจ	Back propagation, Neural Networks, Radial Basis Function Network และ NaïveBayes
2561	จารี ทองคำ วาทินี สุขมาก ภีมพศ สุขมาก	การเปรียบเทียบประสิทธิภาพของเทคนิค Apriori และ FP-Growth ในการสร้างกฎความสัมพันธ์ของโรคมะเร็งต่อมลูกหมาก	Apriori และ FP-Growth
2560	ณัฐพล แสนคำ ธนากร ปุรารัมย์ ทิพวัลย์ แสนคำ	ระบบสนับสนุนทางการแพทย์สำหรับคัดกรองผู้ป่วยโรคไตเรื้อรังโดยใช้เทคนิคเหมืองข้อมูล	Decision tree และ Random Forest
2558	เบญจภัค จงหมื่นไวย์	การเปรียบเทียบปัจจัยโรคประจำตัวผู้สูงอายุโดยใช้อัลกอริทึมการจัดกลุ่ม J48 และ NaiveBayes กรณีศึกษาสาธารณสุขโพธิ์กลางนครราชสีมา	Decision Tree (J48) และ NaïveBayes

จากการทบทวนวรรณกรรมข้างต้น สามารถสรุปกลุ่มการทำเหมืองข้อมูลผู้ป่วยโรคเบาหวาน ได้ดังนี้ การวิจัยในประเทศ เช่น นักวิจัยใช้การทำเหมืองข้อมูลในการเปรียบเทียบอัลกอริทึมการจำแนกแบบมีผู้สอนและไม่มีผู้สอน โดยใช้วิธี Logistic Regression, Random Forest, K-Nearest Neighbor, Support Vector Machine และ Decision Tree เพื่อวัดประสิทธิภาพ โดยใช้ข้อมูลผู้ป่วยโรคเบาหวานในประเทศไทย มีนักวิจัยที่ใช้การทำเหมืองข้อมูลในการพยากรณ์และจำแนกผู้ป่วย โรคเบาหวาน โดยนำเทคนิคการวิเคราะห์องค์ประกอบพื้นฐานและแบบจำลองการจำแนกผสมในการ ค้นหาจำนวนขั้นต่ำของของ Attribute เพื่อปรับปรุงให้มีความเร็วมากขึ้น ผ่านเทคนิคจำแนกข้อมูลเพื่อ ระบุข้อดีและข้อเสีย อีกทั้งประเมินประสิทธิภาพของ Algorithm ที่มีอยู่จะดำเนินการเพื่อกำหนดแนวทางที่ดีที่สุด

วิทยานิพนธ์เล่มนี้ผู้วิจัยมีวัตถุประสงค์ในการพัฒนาและวัดประสิทธิภาพแบบจำลองการพยากรณ์ผู้ป่วยโรคเบาหวาน เพื่อนำไปใช้ในการจำแนกผู้ป่วยโรคเบาหวาน เทคนิคการทำเหมืองข้อมูล (Data Mining) จะถูกนำมาประยุกต์ใช้ในกระบวนการจัดการกับข้อมูลจำนวนมาก เพื่อค้นหารูปแบบ แนวทาง และความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น มาสร้างโมเดลสำหรับการพยากรณ์โอกาส (สาเหตุและปัจจัยที่ส่งผลต่อการเป็นโรคเบาหวาน) ข้อมูลขนาดใหญ่ที่ผู้วิจัยนำใช้ในการทำเหมืองข้อมูล คือจำนวนผู้ป่วยโรคเบาหวานจากระบบฐานข้อมูลของโรงพยาบาลกันทรวิชัย อำเภอกันทรวิชัย จังหวัด มหาสารคาม ที่เข้ารับบริการการรักษาในช่วงระยะเวลา 1 ปีย้อนหลัง ตั้งแต่วันที่ 1 สิงหาคม 2563 ถึง วันที่ 30 ธันวาคม 2564 มาใช้ในการวิเคราะห์ในครั้งนี้ ผลลัพธ์ที่ได้จากการวิเคราะห์ สามารถนำไปวางแผนและจัดการการรักษาผู้ป่วยโรคเบาหวานในโรงพยาบาลกันทรวิชัย ให้มีประสิทธิภาพมากขึ้น ทั้งนี้ยังสามารถนำมาใช้ประกอบการการรักษาแพทย์เพื่อให้รองรับจำนวนผู้ป่วยที่จะเพิ่มขึ้นในอนาคตได้

บทที่ 3
วิธีดำเนินการวิจัย

การวิจัยครั้งนี้เป็นการศึกษาคนคว้าหาข้อเท็จ จริงที่ปรากฏอยู่ มีอยู่ ทั้งนี้เพื่อบรรยายสถานภาพที่ปรากฏอยู่ มีอยู่นั้น เปรียบเทียบกับสถานภาพที่ปรากฏอยู่ในลักษณะต่างๆ จากงานวิจัยที่เกี่ยวข้องในบทที่ 2 สามารถนำไปใช้ในการวิเคราะห์ปัจจัยเสี่ยงที่มีผลต่อการเกิดโรคเบาหวาน มีหลากหลายปัจจัยที่จะทำให้เกิดโรคเบาหวาน โดยบทที่ 3 นี้ จะอธิบายถึงวิธีขั้นตอนการพยากรณ์โรคเบาหวานด้วยเทคนิคการทำเหมืองข้อมูล โดยการคัดเลือกตัวแบบที่เหมาะสมในการพยากรณ์ปัจจัยเสี่ยงที่จะก่อให้เกิดโรคเบาหวาน ดังนี้

1. ประชากรและกลุ่มตัวอย่าง
2. เครื่องมือที่ใช้ในการเก็บรวบรวมข้อมูล
3. การเก็บรวบรวมข้อมูลและการจัดการข้อมูล
4. วิธีดำเนินการวิจัย

3.1 ประชากรและกลุ่มตัวอย่าง

ประชากรเป้าหมาย

เป็นผู้ที่เป็นเบาหวาน โดยมีแพทย์ผู้เชี่ยวชาญวินิจฉัยว่าเป็นโรคเบาหวาน ในคลินิกเบาหวาน โรงพยาบาลกันทรวิชัย จังหวัดมหาสารคาม จากข้อมูลเวชระเบียนผู้ป่วย พบว่ามีจำนวนผู้ป่วยโรคเบาหวานของคลินิกเบาหวาน โรงพยาบาลกันทรวิชัย จังหวัดมหาสารคาม จำนวนทั้งสิ้น 9,841 คน ในปี พ.ศ. 2563 (สถิติสุขภาพคนไทย, 2563)

ประชากรตัวอย่าง

เป็นผู้ป่วยที่แพทย์ผู้เชี่ยวชาญวินิจฉัยว่าเป็นโรคเบาหวานและผู้ป่วยที่เข้ารับการคัดกรองโรคเบาหวาน ตามข้อมูลเวชระเบียนผู้ป่วย โรงพยาบาลกันทรวิชัย จังหวัดมหาสารคาม ในปี พ.ศ. 2563 สามารถคำนวณหาขนาดของกลุ่มตัวอย่างในการทำวิจัยจากตารางสำเร็จรูปของ ทาโร่ ยามาเน่ (Taro Yamane) ที่กำหนดให้มีสัดส่วนของลักษณะที่สนใจในประชากร เท่ากับ 0.5 ระดับความคลาดเคลื่อนที่ยอมรับได้ 5% และระดับความเชื่อมั่น 95% มีจำนวน 400 คน

กลุ่มตัวอย่าง

กลุ่มตัวอย่างในการศึกษาครั้งนี้ คือผู้ป่วยโรคเบาหวานในคลินิกเบาหวาน โรงพยาบาลกันทรวิชัย จังหวัดมหาสารคาม คือระหว่างเดือนสิงหาคม พ.ศ. 2563 ถึงเดือนธันวาคม พ.ศ. 2564 โดยกำหนดคุณสมบัติ ดังนี้

1. เกณฑ์คัดเข้าของกลุ่มตัวอย่าง (Inclusion criteria)
   1.1 ผู้ป่วยที่มีแพทย์ผู้เชี่ยวชาญวินิจฉัยว่าเป็นโรคเบาหวาน
   1.2 กลุ่มประชากรตัวอย่าง

1.3 มีอายุตั้งแต่ 35-60 ปี

1.4 สามารถอ่าน ฟัง เขียนภาษาไทยได้

1.5 ไม่มีการเจ็บป่วยรุนแรงที่เป็นอุปสรรคต่อการเข้าร่วมการตอบแบบสอบถาม

1.6 ยินยอมเข้าร่วมการวิจัย

2. เกณฑ์การคัดออกของกลุ่มตัวอย่าง (Exclusion criteria)

เกณฑ์การคัดออกของกลุ่มตัวอย่างสำหรับการวิจัยครั้งนี้ คือผู้เป็นเบาหวานไม่สะดวกเข้าร่วมในการตอบแบบสอบถาม หรือตอบแบบสอบถามได้ไม่ครบถ้วนตามที่กำหนดไว้

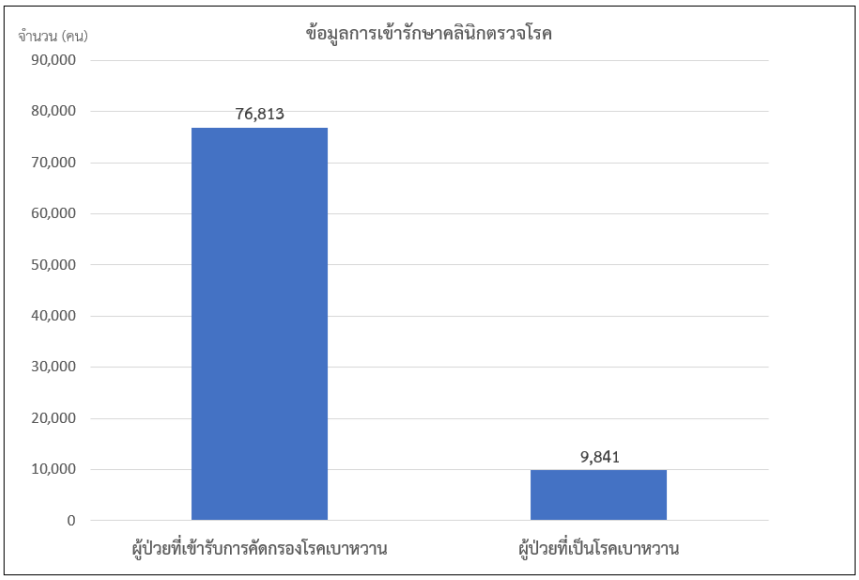
การสุ่มตัวอย่าง

การสุ่มตัวอย่างใช้ความน่าจะเป็นและเป็นการสุ่มแบบไม่ใส่กลับคืน กระบวนการสุ่มกลุ่มตัวอย่าง ใช้การสุ่มตัวอย่างจากการคัดกรองโรคเบาหวานจากประชากรใน อำเภอกันทรวิชัย จังหวัดมหาสารคาม ได้กลุ่มตัวอย่างผู้ป่วยที่เข้ารับการคัดกรองโรคเบาหวาน จำนวน 300 คน และกลุ่มผู้ป่วยที่เป็นโรคเบาหวาน จำนวน 100 คน กระบวนการสุ่มตัวอย่างใช้การสุ่มตัวอย่างแบบแบ่งชั้นภูมิประกอบด้วย 2 ชั้นภูมิ (Stratum) โดยใช้เกณฑ์ที่จำแนกความแตกต่างในชั้นที่หนึ่ง คือแบ่งตามผู้ป่วยที่เข้ารับการคัดกรองโรคเบาหวานและเกณฑ์ที่ใช้ในชั้นที่สองคือผู้ป่วยที่แพทย์ผู้เชี่ยวชาญวินิจฉัยว่าเป็นโรคเบาหวาน ในชั้นภูมิที่ 1 ผู้วิจัยใช้การสุ่มตัวอย่างโดยใช้สัดส่วนที่เท่ากัน ดังนั้นจากขนาดตัวอย่าง 400 คน จากประชากรทั้งสิ้น 86,654 คน จำนวนกลุ่มตัวอย่างแสดงดังตาราง 6

ตาราง 6 ข้อมูลกลุ่มผู้ป่วยที่เข้ารับการคัดกรองโรคเบาหวานและกลุ่มผู้ป่วยที่เป็นโรคเบาหวาน

ห้องตรวจ	ประชากร	กลุ่มตัวอย่าง	ร้อยละ
ผู้ป่วยที่เข้ารับการคัดกรองโรคเบาหวาน	**76,813**	**300**	**88.65**
ผู้ป่วยที่เป็นโรคเบาหวาน	**9,841**	**100**	**11.35**
รวม	**86,654**	**400**	**100.0**

ตาราง 6 แสดงจำนวนป่วยที่เข้ารับการรักษาและคัดกรองโรคเบาหวานในคลินิกห้องตรวจโรคของโรงพยาบาลกันทรวิชัย อำเภอกันทรวิชัย จังหวัดมหาสารคาม ข้อมูลประวัติกลุ่มตัวอย่างผู้ป่วยที่เข้ารับการคัดกรองโรคเบาหวาน จำนวน 300 คน คิดเป็นร้อยละ 88.65 และข้อมูลผู้ป่วยที่เป็นโรคเบาหวานโดยมีแพทย์ผู้เชี่ยวชาญวินิจฉัยว่าเป็นโรคเบาหวาน จำนวน 100 คน คิดเป็นร้อยละ 11.35 ดังแสดงในภาพประกอบ 4



ภาพประกอบ 4 แผนภูมิแสดงจำนวนผู้ป่วยเข้ารับการคัดกรองโรคเบาหวานและผู้ป่วยที่เป็นโรคเบาหวาน

3.2 เครื่องมือที่ใช้ในการเก็บรวบรวมข้อมูล

ส่วนที่ 1 แบบสอบถามข้อมูลทั่วไป

ผู้วิจัยได้พัฒนาเครื่องมือจากการทบทวนวรรณกรรมที่เกี่ยวข้อง ส่วนที่ 1 ข้อมูลทั่วไป ประกอบด้วยข้อมูลเกี่ยวกับ เพศ อายุ การศึกษา การนับถือศาสนา สถานภาพ อาชีพ ดัชนีมวลกาย ความยาวเส้นรอบเอว ความดันโลหิต การดื่มสุรา การสูบบุหรี่ ส่วนสูง น้ำหนัก และพันธุกรรมด้านการเป็นเบาหวาน

ส่วนที่ 2 แบบสอบถามเกี่ยวกับพฤติกรรมความเสี่ยงที่ก่อให้เกิดโรคเบาหวาน

โดยเทคนิคที่ใช้ในการศึกษาครั้งนี้คือเทคนิคการจำแนกข้อมูล (Classification) ในการจำแนกกลุ่มข้อมูลด้วยคุณลักษณะโรคเบาหวาน ทั้ง 5 ระดับ โดย “ไม่เคยปฏิบัติ” ให้ 1 คะแนน จนถึง “ปฏิบัติเป็นประจำ” ให้ 5 คะแนน คะแนนมาก หมายถึง มีการปฏิบัติพฤติกรรมเสี่ยงก่อให้เกิดโรคเบาหวาน อยู่ในระดับสูง คือ น้อยมาก น้อย ปานกลาง มาก และมากที่สุด จำนวนทั้งสิ้น 32 ข้อ

เครื่องมือที่ใช้ในการเก็บรวบรวมข้อมูลจะอยู่ในรูปแบบแบบสอบถาม (ดารณี ทองสัมฤทธิ์ และคณะ, 2560) แบบสอบถามเกี่ยวกับความเสี่ยงที่ก่อให้เกิดโรคเบาหวาน แบ่งออกเป็น 2 ส่วน ดังนี้ ส่วนที่ 1 ข้อมูลทั่วไป ประกอบด้วยข้อมูลเกี่ยวกับ เพศ อายุ การศึกษา การนับถือศาสนา สถานภาพ อาชีพ ดัชนีมวลกาย ความยาวเส้นรอบเอว ความดันโลหิต การดื่มเครื่องดื่มแอลกอฮอล์ การสูบบุหรี่ ส่วนสูง น้ำหนัก และพันธุกรรมด้านการเป็นเบาหวาน จำนวนทั้งสิ้น 13 ข้อ มีลักษณะคำตอบเป็นแบบการปฏิบัติพฤติกรรมเสี่ยงที่ก่อให้เกิดโรคเบาหวาน ส่วนที่ 2 แบบสอบถามเกี่ยวกับพฤติกรรมความเสี่ยงที่ก่อให้เกิดโรคเบาหวาน โดยเทคนิคที่ใช้ในการศึกษาครั้งนี้คือเทคนิคการจำแนกข้อมูล (Classification) ในการจำแนกกลุ่มข้อมูลด้วยคุณลักษณะโรคเบาหวาน ทั้ง 5 ระดับ โดย “ไม่เคยปฏิบัติ” ให้ 1 คะแนน จนถึง “ปฏิบัติเป็นประจำ” ให้ 5 คะแนน คะแนนมาก หมายถึง มีการ

ปฏิบัติพฤติกรรมเสี่ยงที่ก่อให้เกิดโรคเบาหวาน อยู่ในระดับสูง คือ น้อยมาก น้อย ปานกลาง มาก และมากที่สุด จำนวนทั้งสิ้น 32 ข้อ การตอบแบบสอบถามฉบับนี้ ข้อมูลที่ผู้ป่วยตอบแบบสอบถามจะเก็บเป็นความลับ โดยจะนำเสนอผลงานวิจัยในภาพรวมเท่านั้น อาสาสมัครสามารถถอนตัวระหว่างการวิจัยโดยไม่กระทบต่อมาตรฐานการรักษาพยาบาลที่จะได้รับ

3.3 การสร้างและพัฒนาเครื่องมือ

ในการสร้างและพัฒนาเครื่องมือ ผู้วิจัยได้ดำเนินงานตามลำดับ ดังนี้

1. ศึกษาเอกสารงานวิจัยที่เกี่ยวข้องกับปัจจัยที่ส่งผลให้เกิดโรคเบาหวานและเครื่องมือเหมืองข้อมูล เพื่อเป็นแนวทางในการสร้างแบบสอบถาม

2. นำผลของการศึกษาตามข้อ 1 มากำหนดโครงร่างแบบสอบถามตามประเด็นความสำคัญ โดยพิจารณาเนื้อหาให้สอดคล้องกับกรอบแนวคิด ความมุ่งหมาย และสมมุติฐานในการวิจัย

3. นำแบบสอบถามที่สร้างขึ้นตามกรอบแนวคิดเสนออาจารย์ที่ปรึกษาวิทยานิพนธ์ เพื่อพิจารณาความเหมาะสม ความถูกต้องของการใช้ภาษาและครอบคลุมเนื้อหาของงานวิจัย เพื่อนำมาปรับปรุงแก้ไขตามที่อาจารย์ที่ปรึกษาวิทยานิพนธ์แนะนำ

4. ปรับปรุงแก้ไขแบบสอบถามตามที่อาจารย์ที่ปรึกษาวิทยานิพนธ์แนะนำแล้วนำเสนอต่อผู้เชี่ยวชาญ เพื่อตรวจสอบความถูกต้องและครอบคลุมเนื้อหาของการวิจัย ผู้เชี่ยวชาญ 3 ท่าน ประกอบด้วย

4.1 รองศาสตราจารย์ ดร.ธีระวัฒน์ เจริญราษฎร์ ผู้ทรงคุณวุฒิภายนอก

4.2 ผู้ช่วยศาสตราจารย์ ดร.นิติพงษ์ ส่งศรีโรจน์

4.3 อาจารย์ ดร.พีรวัฒน์ ไชยล้อม

5. ปรับปรุงแก้ไขตามข้อเสนอแนะของผู้เชี่ยวชาญ แล้วนำเสนออาจารย์ที่ปรึกษาวิทยานิพนธ์พิจารณาอีกครั้ง

6. การตรวจสอบคุณภาพของเครื่องมือ

6.1 นำแบบสอบถามไปทดลองใช้ (Try - Out) กับผู้ป่วยภายในโรงพยาบาลกันทรวิชัย จำนวน 40 คนแรก

6.2 การหาค่าอำนาจจำแนกเป็นรายข้อ (Discriminant Power) ของแต่ละด้าน โดยใช้เทคนิค Item - total Correlation

6.3 การหาค่าความเชื่อมั่นของเครื่องมือ (Reliability Test) ของแบบสอบถามเป็นรายด้าน โดยใช้วิธีหาค่าสัมประสิทธิ์แอลฟา (Alpha Coefficient) ตามวิธีของครอนบาค (Cronbach)

7. นำผลที่ได้รับจากการตรวจสอบคุณภาพของแบบสอบถามเสนออาจารย์ที่ปรึกษาวิทยานิพนธ์ เพื่อปรับปรุงแก้ไขตามคำแนะนำ

8. ส่งเอกสารเพื่อขอเสนอโครงการวิจัยที่เข้าข่ายการพิจารณาแบบเร่งรัด (Expedited review) เพื่อขอรับการพิจารณาจริยธรรมการวิจัยในคน

9. จัดทำแบบสอบถามฉบับสมบูรณ์ เพื่อนำไปใช้ในการเก็บรวบรวมข้อมูลจากกลุ่มตัวอย่าง

3.4 การเก็บรวบรวมข้อมูลและการจัดการข้อมูล

ผู้วิจัยได้ชี้แจ้งวัตถุประสงค์ความสำคัญของการศึกษาสิทธิของผู้ป่วย เพื่อเก็บข้อมูลจากการคัดกรองความเสี่ยงต่อโรคเบาหวานรายบุคคล ผู้วิจัยได้สร้างแบบสอบถามเพื่อเป็นการเก็บข้อมูลปัจจัยเสี่ยงที่มีอยู่ คือ อายุ เพศ ค่าดัชนีมวลกายค่าความดันโลหิตสูง และประวัติเบาหวานในพ่อแม่ พี่น้อง เพื่อเปรียบเทียบปัจจัยเสี่ยงต่อโรคเบาหวาน และทางแพทย์ผู้เชี่ยวชาญได้เสนอให้เก็บปัจจัยเสี่ยงเพิ่มอีกหนึ่งปัจจัยคือ เส้นรอบเอว เพราะประชาชนไทยเริ่มเป็นโรคอ้วนมากขึ้น เนื่องจากความยาวของเส้นรอบเอวก็เป็นปัจจัยหนึ่งที่ทำให้เกิดโรคเบาหวาน และเสนอให้เก็บปัจจัย การสูบบุหรี่ และการดื่มสุราเข้าไปด้วย โดยประชากรกลุ่มตัวอย่างได้จากผู้ป่วยที่มารับบริการที่คลินิกเบาหวาน โรงพยาบาลกันทรวิชัย จังหวัดมหาสารคาม ตามทะเบียนผู้ป่วยเบาหวานในเดือนสิงหาคม 2563 ถึงเดือนธันวาคม 2564 จำนวน 400 คน

3.5 ขั้นตอนการเก็บรวบรวมข้อมูล

1. เก็บรวบรวมข้อมูลโดยแบบสอบถาม ลักษณะของแบบสอบถามมีลักษณะที่เกี่ยวข้องกับ อายุ เพศ ค่าดัชนีมวลกายค่าความดันโลหิตสูง ประวัติเบาหวานในพ่อแม่ พี่น้อง เส้นรอบเอว การสูบบุหรี่ และการดื่มสุรา ในกลุ่มประชากรตัวอย่าง จำนวน 100 คน
2. เก็บรวบรวมข้อมูลโดยประวัติผู้ป่วยที่มีลักษณะเกี่ยวข้องกับ อายุ เพศ ค่าดัชนีมวลกาย ค่าความดันโลหิตสูง และประวัติเบาหวานในพ่อแม่ พี่น้อง เส้นรอบเอว การสูบบุหรี่ การดื่มสุรา และมีแพทย์ผู้เชี่ยวชาญวินิจฉัยว่าเป็นโรคเบาหวาน จำนวน 300 คน
3. วัดประสิทธิภาพของเครื่องมือจากข้อมูลที่พบในแบบสอบถามเทียบกับข้อมูลที่แพทย์วินิจฉัยว่าเป็นโรคเบาหวาน
4. ผู้วิจัยดำเนินการเข้าเก็บข้อมูลในคลิกนิคเบาหวานโรงพยาบาลกันทรวิชัย โดยมีกระบวนการในการชี้แจงและขอความยินยอมในกลุ่มอาสาสมัครตามขั้นตอนของการทำวิจัยในคน จำนวนทั้งสิ้น 400 คน โดยใช้วิธีในการสุ่มแบบไม่ใส่กลับคืน
5. ระยะเวลาในการดำเนินการระหว่าง สิงหาคม 2563 - ธันวาคม 2564 ตามทะเบียนผู้ป่วยเบาหวานโรงพยาบาลกันทรวิชัย จังหวัดมหาสารคาม
6. ตรวจสอบความสมบูรณ์ของแบบสอบถามที่ได้รับการตอบกลับ
7. ดำเนินการเก็บรวบรวมข้อมูลจากแบบสอบถามที่ได้รับเพื่อนำมาวิเคราะห์ข้อมูลและแปลผลต่อไป

3.6 การจัดกระทำกับข้อมูลและการวิเคราะห์ข้อมูล

3.6.1 การนำเข้าข้อมูล (Data Input)

การศึกษาครั้งนี้เพื่อรวบรวมข้อมูลมาวิเคราะห์ และสร้างโมเดลแบบจำลองการพยากรณ์ระดับภาวะที่ก่อให้เกิดโรคเบาหวาน ข้อมูลสำหรับวิเคราะห์ภาวะที่ก่อให้เกิดโรคเบาหวานมีรายละเอียดดังตารางที่ 7

ตาราง 7 คุณลักษณะและรายละเอียดของข้อมูล

ตัวแปร	ความหมายของข้อมูล	ค่าข้อมูล
A1	เพศ	
A2	อายุ	
A3	ระดับการศึกษา	
A4	นับถือศาสนา	
A5	สถานภาพ	
A6	ดัชนีมวลกาย	
A7	ความยาวเส้นรอบเอว	
A8	ประวัติเป็นความดันเลือดสูง	
A9	ประวัติการดื่มสุรา	
A10	ประวัติการสูบบุหรี่	
A11	ครอบครัวหรือญาติสนิทมีประวัติการเป็นเบาหวาน	
A12	ส่วนสูง	
A13	น้ำหนัก	
B1	ท่านรับประทานอาหารตรงเวลาครบ 3 มื้อทุกวัน	ระดับคะแนน 1-5
B2	ท่านดื่มเครื่องดื่มที่มีส่วนผสมของน้ำตาล 20% เป็นประจำ	ระดับคะแนน 1-5
B3	ท่านไม่เคยดื่มเครื่องดื่มแอลกอฮอล์	ระดับคะแนน 1-5
B4	ท่านเคยดื่มเครื่องดื่มแอลกอฮอล์ แต่ปัจจุบันท่านไม่ดื่มเครื่องดื่มแอลกอฮอล์	ระดับคะแนน 1-5
B5	ท่านรับประทานอาหารที่มีไขมันเป็นส่วนประกอบประจำ เช่น เนื้อสัตว์ติดมัน อาหารที่มีส่วนผสมของกะทิ อาหารทอด เป็นต้น	ระดับคะแนน 1-5
C1	ท่านออกกำลังกายเป็นประจำหรือไม่	ระดับคะแนน 1-5
C2	ท่านเริ่มต้นการออกกำลังกายทีละน้อย ตามกำลังของตนเอง แล้วค่อยๆเพิ่มขึ้น	ระดับคะแนน 1-5

ตาราง 7 คุณลักษณะและรายละเอียดของข้อมูล (ต่อ)

C3	ท่านใช้เวลาออกกำลังกายแต่ละครั้งนานหรือไม่	ระดับคะแนน 1-5
D1	ท่านรับประทานยาเบาหวานตามแพทย์สั่งหรือไม่	ระดับคะแนน 1-5
D2	ท่านลืมรับประทานยาเบาหวานตามที่แพทย์กำหนดหรือไม่	ระดับคะแนน 1-5
D3	ท่านไปรับการตรวจเลือด วัดความดันตามที่แพทย์สั่ง	ระดับคะแนน 1-5
D4	ท่านปฏิบัติตัวตามคำแนะนำของแพทย์หรือเจ้าหน้าที่เมื่อท่านเป็นโรคเบาหวาน	ระดับคะแนน 1-5
E1	ท่านสูบบุหรี่เป็นประจำหรือไม่	ระดับคะแนน 1-5
E2	ท่านเคยสูบบุหรี่หรือไม่	ระดับคะแนน 1-5
F1	ท่านดูแลสุขภาพตนเองด้วยการตรวจสุขภาพเป็นประจำ	ระดับคะแนน 1-5
F2	ท่านมีการวัดความดันโลหิตสม่ำเสมออย่างน้อยทุก 3 เดือน	ระดับคะแนน 1-5
F3	ท่านไปเจาะเลือดตรวจระดับน้ำตาล ไขมันในเลือด และความผิดปกติของหัวใจเป็นระยะตามที่แพทย์แนะนำ	ระดับคะแนน 1-5
F4	ท่านนำความรู้ความเข้าใจมาใช้ในการปฏิบัติตนด้านสุขภาพเพื่อป้องกันโรคเบาหวาน	ระดับคะแนน 1-5
F5	ท่านออกกำลังกาย พักผ่อนให้เพียงพอ ไม่สูบบุหรี่ และกินอาหารครบห้าหมู่เพื่อไม่ให้เกิดโรคติดต่อใด ๆ	ระดับคะแนน 1-5
G1	ท่านกินอาหารโดยไม่ใช้น้ำมันและไม่ใส่ผงปรุงรส	ระดับคะแนน 1-5
G2	ท่านกินอาหารแต่ละมื้อโดยไม่มีการเติมปรุงรสเพิ่ม	ระดับคะแนน 1-5
G3	ท่านหมั่นศึกษาความรู้เกี่ยวกับโรคเบาหวานและการปฏิบัติตนจากสื่อต่าง ๆ	ระดับคะแนน 1-5
G4	ท่านทำกิจกรรมที่ทำให้เสียเหงื่อหรือไม่ เช่น รดน้ำต้นไม้ ปลูกต้นไม้ ทำงานบ้าน เป็นต้น	ระดับคะแนน 1-5
G5	ท่านทำกิจกรรมในยาวว่างเพื่อให้รู้สึกผ่อนคลาย เช่น ฟังเพลง ดูโทรทัศน์ อ่านหนังสือที่ชอบ เป็นต้น	ระดับคะแนน 1-5
G6	ท่านได้รับกำลังใจและใส่ใจจากครอบครัวเมื่อฉันเกิดความเครียด	ระดับคะแนน 1-5
G7	ท่านพูดคุยปรกษาปัญหากับบุคคลใกล้ชิดที่ไว้วางใจได้หรือบุคคลที่ฉันยอมรับเมื่อท่านไม่สบายใจ	ระดับคะแนน 1-5
G8	ท่านรับคำแนะนำจากครอบครัว เพื่อน และชุมชนเกี่ยวกับการทำกิจกรรมป้องกันโรคเบาหวาน	ระดับคะแนน 1-5
G9	ท่านให้ความร่วมมือกับผู้ให้บริการสาธารณสุขเพื่อป้องกันหรือลดการเกิดโรคเบาหวาน	ระดับคะแนน 1-5

3.7 การเตรียมข้อมูล (Data Preparation)

3.7.1. การทำความสะอาดข้อมูล (Data Understanding)

ในโครงสร้างการวินิจฉัยโรคเบาหวานนั้น โดยทั่วไปจะทำการวินิจฉัย 3 วิธี คือ พฤติกรรมเสี่ยงที่จะเกิด หรือมีอาการแสดงออกอย่างเห็นได้ชัดทางคลินิกและวิธีวินิจฉัยสุดท้ายก็คือผลจากการตรวจเลือด ซึ่งหากรู้ผลจากการตรวจเลือดแล้วก็อาจจะเพียงพอในการวินิจฉัยแล้ว แต่เนื่องจากการเจาะเลือดหรือตรวจเลือดนั้น ต้องมีค่าใช้จ่ายที่สูง และไม่สามารถตรวจสอบหรือวินิจฉัยได้อย่างครอบคลุม และพฤติกรรมเสี่ยงจะเก็บข้อมูลของปัจจัยเสี่ยงได้ยาก เนื่องจากพฤติกรรมแต่ละบุคคลแตกต่างกันทำให้ยากต่อการเก็บข้อมูลเป็นฐานความรู้ ผู้ศึกษาจึงหันมาสนใจในการวินิจฉัยที่มีอาการแสดงออกทางคลินิกมาใช้วิเคราะห์เพียงวิธีเดียว โดยแพทย์ผู้เชี่ยวชาญได้สรุปปัจจัยและความเสี่ยงสำคัญที่จะก่อให้เกิดโรคเบาหวานที่มีนัยสำคัญทางสถิติตามตารางดังต่อไปนี้

ตาราง 8 ปัจจัยสภาวะความเสี่ยงสำคัญต่อการเกิดโรคเบาหวาน

ปัจจัยเสี่ยง	ข้อมูล	สภาวะความเสี่ยงต่อเบาหวาน
อายุ(ปี)	<44	ความเสี่ยงน้อย
	45-49	ความเสี่ยงปานกลาง
	>=50	ความเสี่ยงสูง
เพศ	ผู้หญิง	ความเสี่ยงน้อย
	ผู้ชาย	ความเสี่ยงสูง
BMI (kg/m2)	<23	ความเสี่ยงน้อย
	23-27.5	ความเสี่ยงปานกลาง
	>27.5	ความเสี่ยงสูง
ความดันโลหิตสูง	ไม่มี	ความเสี่ยงน้อย
	มี	ความเสี่ยงสูง
มีประวัติเบาหวานในพ่อแม่ พี่น้อง	ไม่มี	ความเสี่ยงน้อย
	มี	ความเสี่ยงสูง

3.7.2 การคัดเลือกตัวแบบ

วิธีการคัดเลือกตัวแบบทำได้โดยการนำเอาข้อมูลปัจจัยเสี่ยงเข้าสอนการเรียนรู้ด้วยตัวแบบในโปรแกรม ซึ่ง ประสิทธิภาพของแต่ละตัวแบบ วัดได้จากค่า Mean Absolute Error (MAE) ค่า Root Mean Squared Error (RMSE) และค่าความถูกต้องของการจำแนกความเป็นสมาชิกของข้อมูลที่ได้จากตัวแบบ นำมาเปรียบเทียบกับผลการวินิจฉัยทางการแพทย์ ดังนี้

1. MAE คือ ค่าสัมบูรณ์ของค่าคลาดเคลื่อนเฉลี่ยของชุดข้อมูลดังสมการ 2

$$MAE = \frac{\sum_{i=0}^{n}|e_i|}{\mathbf{n}} \quad (2)$$

โดยที่ ei คือ ผลต่างระหว่างค่าข้อมูลจริงและค่าพยากรณ์
n คือ จำนวนข้อมูลในการพยากรณ์

การวัดประสิทธิภาพด้วยวิธีนี้จะทำให้ทราบถึงค่าความผิดพลาดของการประมาณค่าที่ชัดเจน

2. RMSE คือ ค่าความคลาดเคลื่อนเฉลี่ยกำลังสอง ดังสมการ 3

$$RMSE = \sqrt{\frac{1}{n}}\sum_{i=1}^{n}[r-t]^2 \quad (3)$$

โดยที่ n คือ จำนวนข้อมูลนำเข้าทั้งหมด
r คือ ค่าของข้อมูลจริงที่ได้จากคำนวณ
t คือ ค่าข้อมูลที่ได้จากการพยากรณ์

3. ผลการวินิจฉัยทางการแพทย์ (ที่เป็นเบาหวานและไม่เป็นเบาหวาน) จากจำนวนผู้มารับการคัดกรองเบาหวานทั้งหมด 400 คน ผลการวินิจฉัยจากแพทย์ คือเป็นโรคเบาหวาน จำนวน 300 คน และไม่เป็นโรคเบาหวาน จำนวน 100 คน การที่นำ 2 ลักษณะนี้มาเปรียบเทียบกัน เพื่อวัดประสิทธิภาพการทำนายของเครื่องมือจากข้อมูลที่พบในแบบสอบถามเทียบกับข้อมูลที่แพทย์วินิจฉัยว่าเป็นโรคเบาหวาน

3.8 วิธีดำเนินการวิจัย

3.8.1 ขั้นตอนการดำเนินงาน

1. แบ่งข้อมูลปัจจัยเสี่ยงเป็น 2 ชุดย่อย คือ

(1) ชุดที่ 1 ปัจจัยเสี่ยงประกอบไปด้วย อายุ เพศ BMI ความดันโลหิตสูง และประวัติเบาหวานในพ่อแม่ พี่น้อง

(2) ชุดที่ 2 เพิ่มตัวแปรความยาวเส้นรอบเอว ตามคำแนะนำจากแพทย์ ผู้เชี่ยวชาญ

2. นำข้อมูลปัจจัยเสี่ยงชุดที่ 1 เข้าสอนการเรียนรู้กับ 5 ตัวแบบ คือ

(1) Multi-layer Perceptron

(2) Naive Bayes

(3) Logistic Regression

(4) Random Forest

(5) Decision Tree

การนำข้อมูลปัจจัยเสี่ยงชุดที่ 1 เข้าสอนการเรียนรู้กับ 5 ตัวแบบ เพื่อคัดเลือกตัวแบบที่เหมาะสมที่สุดในการวิเคราะห์ปัจจัยเสี่ยงของโรคเบาหวาน โดยผลลัพธ์ของข้อมูลนำเข้า คือผลการวินิจฉัยทางการแพทย์ (เป็นเบาหวานและไม่เป็นเบาหวาน)

ตาราง 9 การนำข้อมูลปัจจัยและสภาวะความเสี่ยงของโรคเบาหวานแทนค่าด้วยตัวเลขก่อนทำการสอนการเรียนรู้

รายละเอียด	ข้อมูล	ตัวแปร
อายุ(ปี)	<44	0
	45-49	1
	>=50	2
เพศ	ผู้หญิง	0
	ผู้ชาย	1
BMI (kg/m2)	<23	1
	23-27.5	2
	>27.5	3
ความดันโลหิตสูง	มี	0
	ไม่มี	1
มีประวัติเบาหวานในพ่อแม่ พี่น้อง	มี	0
	ไม่มี	1

3.9 ขั้นตอนการวิเคราะห์ข้อมูล

ขั้นตอนที่ 1 นำข้อมูลปัจจัยเสี่ยงของโรคเบาหวานในตารางที่ 10 คอลัมน์แรกมาทำการแทนค่าตัวแปรก่อนสอนการเรียนรู้ด้วยกับแบบทั้งหมด

ตาราง 10 ตัวอย่างการแทนค่าข้อมูลปัจจัยเสี่ยงของโรคเบาหวานด้วยตัวเลข

SEX	BMI	AGE	HT	FMD
1	22	39	0	0
0	26	46	1	0
1	28	57	1	1

ขั้นตอนที่ 2 กำหนดและสร้างชุดข้อมูลปัจจัยเสี่ยงได้ 2 ชุดย่อย ตามตารางที่ 11 และ 12

ตาราง 11 ข้อมูลปัจจัยเสี่ยงชุดที่ 1

ชุดข้อมูล	**การเก็บข้อมูล**	**ปัจจัยเสี่ยง**
1	**ผู้ป่วยคลิกนิคเบาหวาน จำนวน 400 คน**	**AGE, BMI, SEX, HT, FDM**

AGE คือ อายุผู้ป่วย
BMI คือ ค่าดัชนีมวลกาย
SEX คือ เพศ
HT คือ ประวัติความดันโลหิต
FDM คือ ประวัติเบาหวานในพ่อแม่ พี่น้อง

ตาราง 12 ข้อมูลปัจจัยเสี่ยงชุดที่ 2 เพิ่มตัวแปรความยาวเส้นรอบเอวตาม การสูบบุหรี่ การดื่มสุรา ตามคำแนะนำจากแพทย์ผู้เชี่ยวชาญ

ชุดข้อมูล	การเก็บข้อมูล	ปัจจัยเสี่ยง
2	ผู้ป่วยคลินิกเบาหวาน จำนวน 400 คน	**AGE, BMI, SEX, HT, FDM, OR, SM, D**

OR คือ ความยาวรอบเอว
SM คือ การสูบบุหรี่
D คือ การดื่มสุรา

ขั้นตอนที่ 3 ทำการสอนการเรียนรู้และทำการทดสอบชุดข้อมูลแต่ละชุด

(1) แปลงข้อมูลนำเข้า (Input) ให้อยู่ในรูปแบบตัวเลขเพื่อให้คอมพิวเตอร์เข้าใจและประมวลผลได้

(2) ในขั้นตอนการเทรนหรือการสอนการเรียนรู้ นอกจากข้อมูลนำเข้าแล้วยังต้องมีชุดข้อมูลที่เป็นเป้าหมายป้อนเข้าไปด้วย เพื่อที่จะให้โครงข่ายเรียนรู้ค่าเป้าหมายที่ต้องการ ดังนี้

ตาราง 13 ชุดข้อมูลนำเข้าและชุดข้อมูลเป้าหมายจากข้อมูลชุดที่ 1 และข้อมูลชุดที่ 2 ที่ถูกแปลงค่าให้อยู่ในรูปแบบเดียวกันก่อนการสอนการเรียนรู้

Sex	BMI	AGE	HT	FMD	OR	SM	D	Target
1	22	39	0	0	0	1	1	0
0	26	46	1	0	1	0	1	1
1	28	57	1	1	1	1	0	1

ตาราง 14 อธิบายผลการแปลงค่าแต่ละตัวแปรดังนี้

รายละเอียด	ข้อมูล	ตัวแปร
SEX	ผู้หญิง	0
	ผู้ชาย	1
BMI	<23	1
	23-27.5	2
	>27.5	3
FMD	ไม่มี	0
	มี	1
AGE	>30	0
	<44	1
	45-49	2
	>=50	3
HT	ไม่มี	0
	มี	1
OR	ความยาวเส้นรอบเอว >=90 ซม.(ผู้ชาย) และ >=80 ซม. (ผู้หญิง)	0
	ความยาวเส้นรอบเอว <=90 ซม.(ผู้ชาย) และ <=80 ซม. (ผู้หญิง)	1
SM	ไม่สูบบุหรี่	0
	สูบบุหรี่	1
D	ไม่ดื่มสุรา	0
	ดื่มสุรา	1
TARGET	ไม่ป่วยเป็นโรคเบาหวาน	0
	ป่วยเป็นโรคเบาหวาน	1

การจัดลำดับความสำคัญปัจจัยเสี่ยงของโรคเบาหวาน

1. ขั้นตอนการจัดลำดับความสำคัญของปัจจัยเสี่ยง

1.1 การจัดลำดับความสำคัญปัจจัยเสี่ยงของโรคเบาหวานต้องเริ่มด้วยการเอาข้อมูลปัจจัยเสี่ยงทุกรายการซึ่งจะประกอดด้วย เพศ ดัชนีมวลกาย ประวัติเบาหวานในพ่อแม่ พี่น้อง อายุผู้ป่วย ประวัติความดันโลหิตสูง ความยาวเส้นรอบเอว การสูบบุหรี่ และการดื่มสุรา เข้าสอนการเรียนรู้ด้วยตัวแบบทั้ง 5 ตัวแบบ โดยเป้าหมายของแต่ละรายการของข้อมูลนำเข้าคือ

ผลการวินิจฉัยทางการแพทย์

2. การตรวจสอบความถูกต้องย้อนกลับ

2.1 ข้อมูลปัจจัยเสี่ยงซึ่งประกอบด้วย AGE, BMI, SEX, HT, FDM, OR, SM และ D จะถูกมาทดสอบความรู้ย้อนกลับเพื่อตรวจสอบลำดับความสำคัญของปัจจัยเสี่ยง ด้วยตัวแบบทั้งหมด โดยจะลดข้อมูลที่นำเข้าทีละตัวแปร

2.2 ผลการทดสอบความถูกต้องย้อนกลับจะวัดได้จากค่า MAE, และค่า RMSE

บทที่ 4
การทดสอบและอภิปรายผล

การวัดประสิทธิภาพของระบบจะทดสอบประสิทธิภาพจากค่าความถูกต้อง ค่าความแม่นยำของข้อมูล (Precision) ค่าความระลึกการวัดความถูกต้องของโมเดล (Recall) ค่าความถ่วงดุล (F-Measure) และค่าพื้นที่ใต้กราฟ (Receiver Operating Characteristic; ROC) ในการพยากรณ์โรคเบาหวานในผู้ป่วย โรงพยาบาลกันทรวิชัย เพื่อจำแนกภาวะไม่มีความเสี่ยงและมีความเสี่ยง โดยเปรียบเทียบอัลกอริทึมในจำการจำแนกทั้งหมด 5 อัลกอริทึม คือเพอร์เซ็ปตรอนแบบหลายชั้น (Multilayer perceptron) นาอีฟ เบย์ (Naïve bayes) โลจิสติก (Logistic Regression) เทคนิคป่าสุ่ม (Random forest) และต้นไม้ตัดสินใจ (Decision Trees) ซึ่งเนื้อหาในบทนี้จะประกอบด้วยข้อมูลที่ใช้ในการทดสอบประสิทธิภาพการพยากรณ์โรคเบาหวาน โดยแบ่งเป็นการทดสอบประสิทธิภาพการจำแนกด้วยค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) ค่าความถ่วงดุล (F-measure) และค่าเส้นกราฟ ROC (Receiver operating characteristic; ROC) ซึ่งทั้งหมดจะถูกนำมาอธิบายผลลัพธ์ได้ตามข้อมูลด้านล่างนี้

4.1 ข้อมูลที่ใช้ในการทดสอบ

การทดสอบการพยากรณ์ในการพยากรณ์โรคเบาหวานจะใช้ข้อมูลที่ได้จากแบบสอบถามจากผู้ป่วยในโรงพยาบาลกันทรวิชัย ซึ่งประกอบด้วยข้อมูลจำนวน 400 คน ซึ่งประกอบด้วยข้อมูลทั่วไป ข้อมูลพฤติกรรมความเสี่ยงที่ก่อให้เกิดโรคเบาหวาน และแบบทดสอบข้อมูลด้านการต้านทานความชุกของโรค เพื่อจำแนกข้อมูลออกเป็นกลุ่มที่ไม่มีความเสี่ยงและกลุ่มที่มีความเสี่ยงข้อมูลดังกล่าวได้ ดังตาราง 15

ตาราง 15 ตารางแสดงข้อมูลทั่วไปของผู้ป่วย

ตัวแปร	ความหมายของข้อมูล	ค่าข้อมูล
A1	เพศ	0 หมายถึง หญิง 1 หมายถึง ชาย
A2	อายุ	0 หมายถึง น้อยกว่า 30 ปี 1 หมายถึง 30 – 40 ปี 2 หมายถึง 41 – 50 ปี 3 หมายถึง มากกว่า 50 ปี

ตาราง 15 ตารางแสดงข้อมูลทั่วไปของผู้ป่วย (ต่อ)

A3	ระดับการศึกษา	0 หมายถึง ต่ำกว่าปริญญาตรี 1 หมายถึง ปริญญาตรี 2 หมายถึง ปริญญาตรีขึ้นไป
A4	นับถือศาสนา	0 หมายถึง พุทธ 1 หมายถึง คริสต์ 2 หมายถึง อิสลาม
A5	สถานภาพ	0 หมายถึง โสด 1 หมายถึง สมรสและคู่สมรสอาศัยอยู่ด้วยกัน 2 หมายถึง หม้าย 3 หมายถึง หย่า/แยก 4 หมายถึง สมรสแต่คู่สมรสไม่ได้อาศัยอยู่ด้วยกัน
A6	ดัชนีมวลกาย	1 หมายถึง น้อยกว่า 23 2 หมายถึง 23 น้อยกว่า 27.5 3 หมายถึง มากกว่าหรือเท่ากับ 27.5
A7	ความยาวเส้นรอบเอว	0 หมายถึง ผู้ชาย น้อยกว่า 90 ซม. 1 หมายถึง ผู้หญิง น้อยกว่า 80 ซม. 2 หมายถึง ผู้ชาย มากกว่าหรือเท่ากับ 90 ซม. 3 หมายถึง ผู้หญิง มากกว่าหรือเท่ากับ 80 ซม.
A8	ประวัติเป็นความดันเลือดสูง	0 หมายถึง ไม่มีความดันเลือดสูง 1 หมายถึง เป็นความดันเลือดสูง
A9	ประวัติการดื่มสุรา	0 หมายถึง ดื่มสุรา 1 หมายถึง ไม่ดื่มสุรา
A10	ประวัติการสูบบุหรี่	0 หมายถึง สูบบุหรี่ 1 หมายถึง ไม่สูบบุหรี่
A11	ครอบครัวหรือญาติสนิทมีประวัติการเป็นเบาหวาน	0 หมายถึง ไม่มีประวัติ 1 หมายถึง มีประวัติ
A12	ส่วนสูง	0 หมายถึง น้อยกว่า 150 เซนติเมตร 1 หมายถึง 150 - 160 เซนติเมตร 2 หมายถึง 161 - 180 เซนติเมตร 3 หมายถึง มากกว่า180 เซนติเมตร
A13	น้ำหนัก	1 หมายถึง น้อยกว่า 45 กิโลกรัม 2 หมายถึง 45 – 60 กิโลกรัม 3 หมายถึง 61 – 70 กิโลกรัม 4 หมายถึง มากกว่า 70 กิโลกรัม

ตาราง 16 แสดงข้อมูลพฤติกรรมการรับประทานอาหาร

ตัวแปร	ความหมายของข้อมูล	ค่าข้อมูล
B1	ท่านรับประทานผักสด หรือผักลวก เป็นประจำ	1, 2, 3, 4, และ 5
B2	ท่านดื่มเครื่องดื่มที่มีส่วนผสมของน้ำตาล 20% เป็นประจำ	1, 2, 3, 4, และ 5
B3	ท่านไม่เคยดื่มเครื่องดื่มแอลกอฮอล์	1, 2, 3, 4, และ 5
B4	ท่านเคยดื่มเครื่องดื่มแอลกอฮอล์ แต่ปัจจุบันท่านไม่ดื่มเครื่องดื่มแอลกอฮอล์	1, 2, 3, 4, และ 5
B5	ท่านรับประทานอาหารที่มีไขมันเป็นส่วนประกอบประจำ เช่น เนื้อสัตว์ติดมัน อาหารที่มีส่วนผสมของกะทิ อาหารทอด เป็นต้น	1, 2, 3, 4, และ 5

ตาราง 17 แสดงข้อมูลพฤติกรรมด้านการออกกำลังกาย

ตัวแปร	ความหมายของข้อมูล	ค่าข้อมูล
C1	ท่านออกกำลังกายเป็นประจำหรือไม่	1, 2, 3, 4, และ 5
C2	ท่านเริ่มต้นการออกกำลังกายทีละน้อย ตามกำลังของตนเอง แล้วค่อยๆเพิ่มขึ้น	1, 2, 3, 4, และ 5
C3	ท่านใช้เวลาออกกำลังกายแต่ละครั้งนานหรือไม่	1, 2, 3, 4, และ 5

ตาราง 18 แสดงข้อมูลพฤติกรรมด้านการรับประทานยา

ตัวแปร	ความหมายของข้อมูล	ค่าข้อมูล
D1	ท่ารับประทานยาเบาหวานตามแพทย์สั่งหรือไม่	1, 2, 3, 4, และ 5
D2	ท่านลืมรับประทานยาเบาหวานตามที่แพทย์กำหนดหรือไม่	1, 2, 3, 4, และ 5
D3	ท่านไปรับการตรวจเลือด วัดความดันตามที่แพทย์สั่ง	1, 2, 3, 4, และ 5
D4	ท่านปฏิบัติตัวตามคำแนะนำของแพทย์หรือเจ้าหน้าที่เมื่อท่านเป็นโรคเบาหวาน	1, 2, 3, 4, และ 5

ตาราง 19 แสดงข้อมูลพฤติกรรมด้านการสูบบุหรี่

ตัวแปร	ความหมายของข้อมูล	ค่าข้อมูล
E1	ท่านสูบบุหรี่เป็นประจำหรือไม่	1, 2, 3, 4, และ 5
E2	ท่านเคยสูบบุหรี่หรือไม่	1, 2, 3, 4, และ 5

ตาราง 20 แสดงข้อมูลด้านการต้านทานการเกิดโรค

ตัวแปร	ความหมายของข้อมูล	ค่าข้อมูล
F1	ท่านดูแลสุขภาพตนเองด้วยการตรวจสุขภาพเป็นประจำ	1, 2, 3, 4, และ 5
F2	ท่านมีการวัดความดันโลหิตสม่ำเสมออย่างน้อยทุก 3 เดือน	1, 2, 3, 4, และ 5
F3	ท่านไปเจาะเลือดตรวจระดับน้ำตาล ไขมันในเลือด และความผิดปกติของหัวใจเป็นระยะตามที่แพทย์แนะนำ	1, 2, 3, 4, และ 5
F4	ท่านนำความรู้ความเข้าใจมาใช้ในการปฏิบัติตนด้านสุขภาพเพื่อป้องกันโรคเบาหวาน	1, 2, 3, 4, และ 5
F5	ท่านออกกำลังกาย พักผ่อนให้เพียงพอ ไม่สูบบุหรี่ และกินอาหารครบห้าหมู่เพื่อไม่ให้เกิดโรคติดต่อใดๆ	1, 2, 3, 4, และ 5

ตาราง 21 แสดงข้อมูลด้านการต้านทานความชุกของโรค

ตัวแปร	ความหมายของข้อมูล	ค่าข้อมูล
G1	ท่านกินอาหารโดยไม่ใช้น้ำมันและไม่ใส่ผงปรุงรส	1, 2, 3, 4, และ 5
G2	ท่านกินอาหารแต่ละมื้อโดยไม่มีการเติมปรุงรสเพิ่ม	1, 2, 3, 4, และ 5
G3	ท่านหมั่นศึกษาความรู้เกี่ยวกับโรคเบาหวานและการปฏิบัติตนจากสื่อต่างๆ	1, 2, 3, 4, และ 5
G4	ท่านทำกิจกรรมที่ทำให้เสียเหงื่อหรือไม่ เช่น รดน้ำต้นไม้ ปลูกต้นไม้ ทำงานบ้าน เป็นต้น	1, 2, 3, 4, และ 5
G5	ท่านทำกิจกรรมในยาวว่างเพื่อให้รู้สึกผ่อนคลาย เช่น ฟังเพลง ดูโทรทัศน์ อ่านหนังสือที่ชอบเป็นต้น	1, 2, 3, 4, และ 5
G6	ท่านได้รับกำลังใจและใส่ใจจากครอบครัวเมื่อฉันเกิดความเครียด	1, 2, 3, 4, และ 5
G7	ท่านพูดคุยปรึกษาปัญหากับบุคคลใกล้ชิดที่ไว้วางใจได้หรือบุคคลที่ฉันยอมรับเมื่อท่านไม่สบายใจ	1, 2, 3, 4, และ 5
G8	ท่านรับคำแนะนำจากครอบครัว เพื่อน และชุมชนเกี่ยวกับการทำกิจกรรมป้องกันโรคเบาหวาน	1, 2, 3, 4, และ 5
G9	ท่านให้ความร่วมมือกับผู้ให้บริการสาธารณสุขเพื่อป้องกันหรือลดการเกิดโรคเบาหวาน	1, 2, 3, 4, และ 5

4.2 ขั้นตอนการเตรียมข้อมูล

การเตรียมข้อมูลเพื่อใช้ในการพยากรณ์โรคเบาหวานจะสามารถจำแนกได้เป็น 2 ขั้นตอน ซึ่ง จะแสดงรายละเอียดในแต่ละขั้นตอนดังนี้

4.2.1 การเก็บรวบรวมข้อมูล

การเก็บรวบรวมข้อมูลได้จากการสุ่มตัวอย่างผู้ป่วยโรงพยาบาลกันทรวิชัย พบว่ามีจำนวนผู้ป่วยที่ตอบแบบสอบถามออนไลน์ทั้งสิ้น 475 คน และผู้วิจัยได้นำข้อมูลดังกล่าวมาผ่านการวิเคราะห์และตรวจสอบความสมบูรณ์ของข้อมูลพบว่ามีข้อมูลที่ถูกต้อง ครบถ้วนและสามารถนำไปใช้ในการพยากรณ์โรคเบาหวานได้จริงจำนวนทั้งสิ้น 400 คน

4.2.2 การเตรียมข้อมูล

พิจารณาข้อมูลจากขั้นตอนที่ 4.2.1 พบว่ามีข้อมูลที่ไม่ถูกต้องครบถ้วน จำนวน 75 คน ซึ่ง ผู้วิจัยได้ตัดข้อมูลส่วนนี้ออกตามกระบวนการตรวจสอบและการแก้ไข (หรือลบ) รายการข้อมูลที่ไม่ ถูกต้องออกไปจากชุดข้อมูล (Data cleansing) เพื่อให้การจำแนกข้อมูลถูกต้องมากที่สุด ทำให้เหลือ จำนวนข้อมูลที่สามารถนำมาใช้ในการวิเคราะห์ได้จริงจำนวนทั้งสิ้น 400 คน จากนั้นนำข้อมูลที่ได้ มาทำให้ข้อมูลอยู่ในโครงสร้างหรือรูปแบบบรรทัดฐานเดียวกัน (Normalization) เพื่อลดความซ้ำซ้อนของข้อมูล ดังภาพประกอบที่ 5

ลำดับที่	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	B1	B2	B3	B4	B5	C1	C2	C3	D1	D2	D3	D4	E1	E2	F1	F2	F3	F4	F5	G1	G2	G3	G4	G5	G6	G7	G8	G9
1	0	0	0	1	1	3	1	1	0	0	1	2	4	2	2	4	5	4	1	5	1	4	3	3	3	3	5	5	2	5	2	4	2	4	1	5	1	4	1	3	5
2	1	1	0	1	0	2	0	0	0	1	1	1	3	1	2	1	5	5	3	3	2	4	3	2	5	2	5	3	4	2	3	4	3	2	1	5	1	1	3	4	3
3	0	1	2	0	4	1	3	1	0	1	0	2	2	5	3	5	3	4	3	3	4	1	3	1	5	5	2	1	2	1	2	3	2	5	3	4	3	5	2	3	4
4	0	2	2	0	3	1	1	0	1	0	1	1	4	4	3	5	1	3	4	4	1	5	4	5	5	3	3	4	1	3	2	5	5	1	5	4	2	3	3	3	1
5	0	2	2	2	0	2	3	1	0	0	0	2	3	3	5	3	4	1	2	4	2	3	4	3	3	2	5	4	1	2	2	4	1	2	4	2	1	5	3	2	4
6	0	1	0	1	2	1	1	1	0	1	0	1	4	3	5	1	4	1	5	5	1	1	1	3	1	5	2	5	4	5	4	2	1	3	2	1	5	1	4	2	1
7	1	2	1	0	0	2	2	1	1	0	1	1	3	4	1	5	4	2	2	4	5	4	5	3	2	5	4	1	1	5	5	3	2	5	4	3	5	3	3	5	3
8	1	2	1	1	3	1	0	1	1	1	1	2	2	5	1	3	4	3	2	3	1	4	1	1	3	4	4	5	1	2	1	1	2	5	2	4	4	2	2	5	1
9	1	0	0	2	3	1	2	1	1	1	0	2	4	1	3	4	4	4	2	2	1	3	3	5	3	4	4	4	4	4	5	5	2	3	3	2	1	2	4	1	2
10	1	2	2	0	0	2	0	0	1	0	0	1	2	4	1	1	1	1	1	1	5	3	1	2	3	1	1	1	3	5	3	3	5	2	5	1	4	5	2	5	5
11	1	2	1	2	0	2	0	0	0	1	0	2	1	1	3	1	3	3	3	1	3	2	5	1	2	5	4	4	3	3	1	1	4	1	5	4	2	5	1	3	1
12	0	1	0	0	2	1	1	0	1	1	0	0	4	5	5	1	1	5	4	2	4	1	1	2	3	1	3	2	2	4	4	4	4	1	3	5	5	4	2	3	1
13	1	3	2	1	3	1	0	1	0	1	1	2	2	4	4	4	1	1	2	5	1	5	4	4	5	1	2	5	4	1	2	3	2	2	4	1	4	4	2	3	4
14	1	3	0	0	1	3	2	0	0	1	0	2	4	4	5	4	2	1	1	3	1	5	5	4	5	2	2	1	2	5	5	3	4	5	1	2	3	4	5	4	5
15	1	2	2	0	4	1	0	1	1	0	1	1	2	2	2	4	1	3	3	2	2	4	4	3	4	5	5	1	2	2	5	2	3	3	1	2	1	3	4	5	1
16	0	1	2	2	1	3	1	0	0	0	0	1	3	2	1	5	4	3	4	4	2	3	5	4	5	5	4	4	4	4	5	1	1	2	2	3	2	1	3	1	1
17	1	0	0	0	2	1	0	1	0	1	0	2	4	2	5	4	3	1	4	3	4	1	1	5	4	5	5	2	4	4	4	2	4	4	2	1	4	3	1	5	3
18	1	3	2	1	1	3	0	1	1	0	1	2	3	1	4	5	5	2	1	3	3	4	1	1	1	2	4	5	3	3	5	2	1	3	4	4	4	3	2	4	2
19	1	3	2	2	2	1	0	1	0	0	0	2	3	5	4	5	3	3	4	1	5	3	3	1	1	2	2	2	1	2	5	3	5	5	3	2	5	3	1	1	5
20	1	3	1	2	1	3	0	1	0	1	1	2	2	1	1	2	5	4	1	4	4	2	1	4	3	5	3	5	4	2	3	1	1	5	4	2	5	5	1	2	2
21	0	1	2	2	4	1	3	0	1	0	0	2	4	5	1	5	2	2	5	3	2	5	5	1	3	3	2	5	5	2	1	3	3	3	4	4	5	3	2	5	4
22	0	3	2	2	3	1	1	1	0	0	1	0	2	4	5	2	3	1	1	4	1	4	3	3	1	2	1	1	3	1	5	3	2	4	4	2	4	5	1	3	2
23	1	3	0	2	1	3	0	0	0	0	0	0	1	2	2	3	4	1	1	1	5	2	5	4	4	5	3	5	4	3	1	1	1	3	2	3	1	3	3	1	2
24	0	1	0	2	2	1	1	0	0	1	0	2	4	2	3	2	2	1	1	3	4	2	3	3	4	1	2	2	2	3	1	3	1	1	2	2	1	2	4	5	1
25	1	2	2	1	2	1	0	0	0	1	0	2	4	3	5	4	1	4	3	4	5	3	3	2	4	2	5	5	5	1	3	5	1	1	4	1	4	1	2	5	3
26	0	1	1	2	2	1	1	1	0	0	0	2	4	4	4	5	1	4	4	2	4	1	3	5	2	4	3	4	2	2	4	4	2	2	2	4	5	3	3	1	5
27	0	2	2	0	0	2	1	1	1	1	0	1	4	1	2	1	4	1	2	3	4	2	1	4	4	3	5	3	3	3	4	4	5	1	1	3	3	4	4	4	1
28	1	3	2	2	1	3	0	1	0	0	1	2	2	4	4	1	4	3	1	3	1	3	1	5	1	3	2	5	3	5	1	5	4	3	1	3	2	5	3	2	2

ภาพประกอบ 5 ตัวอย่างของการปรับโครงสร้างข้อมูลเพื่อลดความซ้ำซ้อนของข้อมูลและจัดลำดับความสำคัญของข้อมูล

จากภาพประกอบ 5 พบว่าข้อมูลสามารถจัดลำดับความสำคัญของข้อมูลได้โดยตัวอย่างของการปรับโครงสร้างข้อมูลเพื่อลดความซ้ำซ้อนของข้อมูลและจัดลำดับความสำคัญของข้อมูลจะสามารถนำแบ่งกลุ่มข้อมูลได้ทั้งหมด 2 ประเภท ดังที่แสดงในตาราง 22

ตาราง 22 ประเภทของข้อมูลที่ใช้ในการวิเคราะห์โรคเบาหวาน 2 ด้าน

ประเภทของตัวแปร	จำนวนข้อมูล (ตัวแปร)
ประเภทที่ 1 ข้อมูลส่วนบุคคล	**13**
ประเภทที่ 2 ข้อมูลพฤติกรรมความเสี่ยงจำนวน 6 ด้าน	**28**
พฤติกรรมการรับประทานอาหาร	**5**
พฤติกรรมด้านการออกกำลังกายท่านออกกำลังกายของแพทย์	**3**
พฤติกรรมด้านการรับประทานยา	**4**
พฤติกรรมด้านการสูบบุหรี่	**2**
ด้านการต้านทานการเกิดโรค	**5**
ด้านการต้านทานความชุกของโรค	**9**

พิจารณาจากข้อมูลในตาราง 22 จะสามารถนำจำนวนข้อมูลในแต่ละประเภท (ตัวแปร) เพื่อไปจัด กลุ่มเพื่อหาความเชื่อมโยงหรือสอดคล้องกับการพยากรณ์โรคเบาหวาน โดยจะนำข้อมูลในประเภทที่ 2 ไปแบ่งกลุ่มเพื่อหาความสำคัญของแต่ละตัวแปรโดยสามารถแสดงการแบ่งกลุ่มข้อมูลได้ดังตาราง 23

ตาราง 23 แสดงการแบ่งกลุ่มข้อมูลในประเภทที่ 2 เพื่อการพยากรณ์โรคเบาหวาน

ประเภทของตัวแปร	จำนวนข้อมูลในแต่ละประเภท (ตัวแปร)
ประเภทที่ 2 ข้อมูลพฤติกรรมความเสี่ยง	**5**
ประเภทที่ 3 ข้อมูลการต้านทานความชุกของโรค	**9**

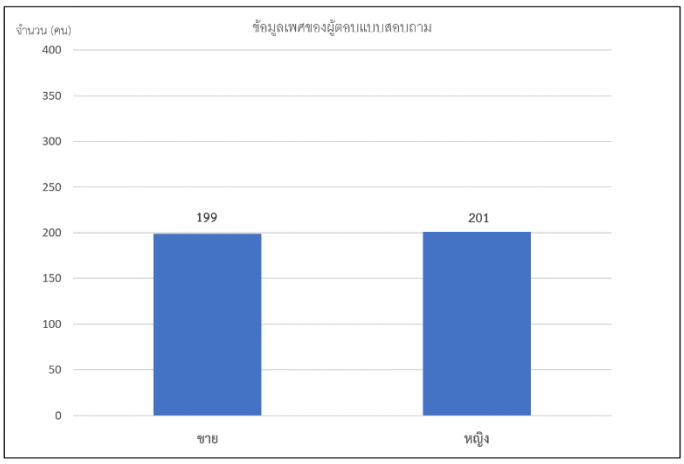
พิจารณาจากข้อมูลในตาราง 23 การเก็บข้อมูลในแต่ละปัจจัยจะมีความแตกต่างกันทำให้ยากต่อการสร้างเป็นฐานข้อมูลความรู้ ดังนั้นผู้วิจัยจึงนำประเภทของแต่ละตัวแปรที่แสดงออกทางพฤติกรรมมาใช้ในการวิเคราะห์ข้อมูล ทั้งหมด 2 ประเภท (ยกเว้นประเภทที่ 1 ข้อมูลส่วนบุคคล) ข้อมูลแต่ละด้านสามารถนำมากำหนดค่าตัวแปรความหมายของข้อมูลค่าประจำตัวแปรและแสดงจำนวนข้อมูลของผู้ตอบแบบสอบถามได้ดังตาราง 24

ตาราง 24 ข้อมูลเพศของผู้ตอบแบบสอบถาม

ความหมายของข้อมูล	จำนวน	ร้อยละ
ชาย	199	49.8
หญิง	201	50.3

รวม		100

ตาราง 24 แสดงถึงข้อมูลของผู้ป่วยโรคเบาหวานจำนวน 400 คน แบ่งเป็นผู้ป่วยที่เป็นเพศชาย จำนวน 199 คน คิดเป็นร้อยละ 49.8 และผู้ป่วยที่เป็นเพศหญิง จำนวน 201 คน คิดเป็นร้อยละ 50.3 ดังแสดงในภาพประกอบ 6

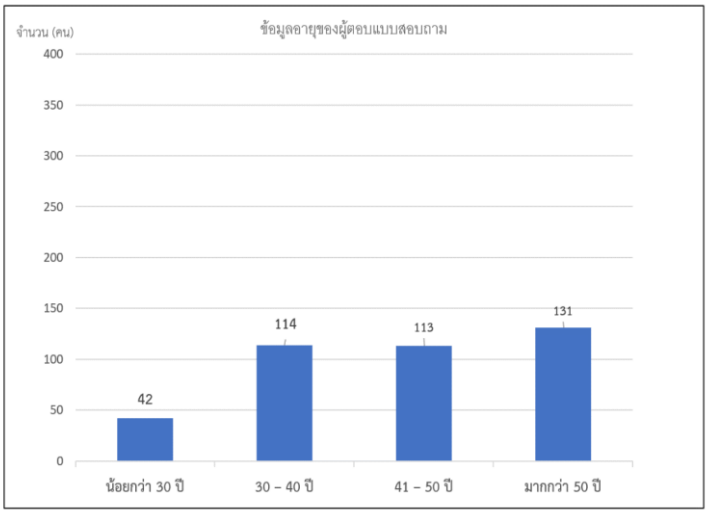


ภาพประกอบ 6 แผนภูมิข้อมูลเพศของผู้ตอบแบบสอบถาม

ตาราง 25 ข้อมูลอายุของผู้ตอบแบบสอบถาม

ความหมายของข้อมูล	จำนวน	ร้อยละ
อายุ		
น้อยกว่า 30 ปี	**42**	**10.5**
31 – 40 ปี	**114**	**28.5**
41 – 50 ปี	**113**	**28.3**
มากกว่า 50 ปี	**131**	**32.8**
รวม		**100**

ตาราง 25 แสดงถึงข้อมูลของผู้ป่วยโรคเบาหวานจำนวน 400 คน แบ่งเป็นอายุของผู้ป่วยที่น้อยกว่า 30 ปี จำนวน 42 คน คิดเป็นร้อยละ 10.5 ผู้ป่วยที่อายุ 31-40 ปี จำนวน 114 คน คิดเป็นร้อยละ 28.5 ผู้ป่วยที่อายุ 41 – 50 ปี จำนวน 113 คน คิดเป็นร้อยละ 28.3 และผู้ป่วยที่อายุมากกว่า ปี จำนวน 131 คน คิดเป็นร้อยละ 32.8 ดังแสดงในภาพประกอบ 7

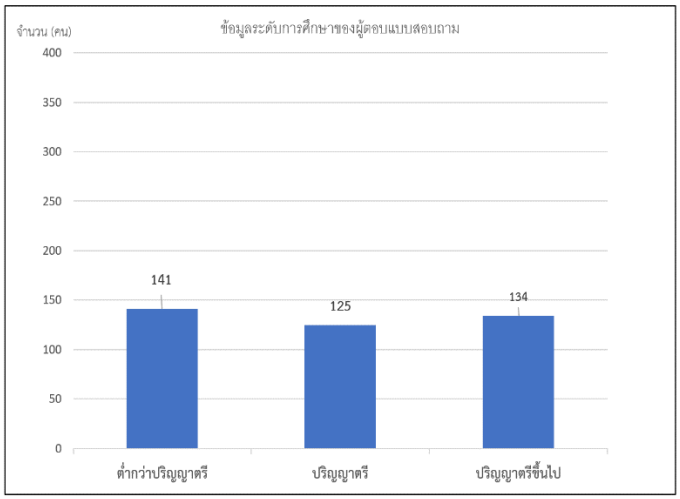


ภาพประกอบ 7 แผนภูมิข้อมูลอายุของผู้ตอบแบบสอบถาม

ตาราง 26 ข้อมูลระดับการศึกษาของผู้ตอบแบบสอบถาม

ความหมายของข้อมูล	จำนวน	ร้อยละ
ระดับการศึกษา		
ต่ำกว่าปริญญาตรี	**141**	**35.3**
ปริญญาตรี	**125**	**31.3**
ปริญญาตรีขึ้นไป	**134**	**33.5**
รวม		**100**

ตาราง 26 แสดงถึงข้อมูลของผู้ป่วยโรคเบาหวานจำนวน 400 คน แบ่งเป็นผู้ป่วยที่ระดับการศึกษาต่ำกว่าปริญญาตรี จำนวน 141 คน คิดเป็นร้อยละ 35.3 ผู้ป่วยที่ระดับการศึกษาปริญญาตรี จำนวน 125 คน คิดเป็นร้อยละ 31.3 และผู้ป่วยที่ระดับการศึกษาปริญญาตรีขึ้นไป จำนวน 134 คน คิดเป็นร้อยละ 33.5 ดังแสดงในภาพประกอบ 8

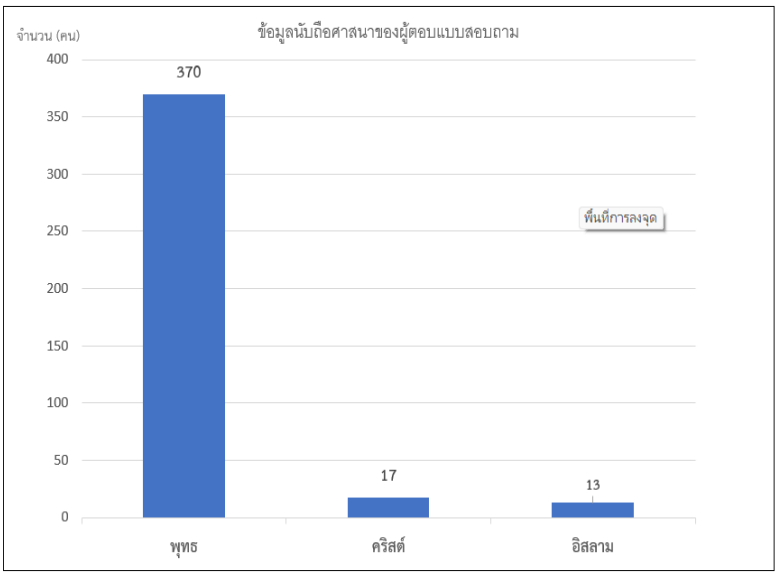


ภาพประกอบ 8 แผนภูมิข้อมูลระดับการศึกษาของผู้ตอบแบบสอบถาม

ตาราง 27 ข้อมูลนับถือศาสนาของผู้ตอบแบบสอบถาม

ความหมายของข้อมูล	จำนวน	ร้อยละ
นับถือศาสนา		
พุทธ	370	92.5
คริสต์	17	4.3
อิสลาม	13	3.3
รวม		100

ตาราง 27 แสดงถึงข้อมูลของผู้ป่วยโรคเบาหวานจำนวน 400 คน แบ่งเป็นผู้ป่วยที่นับถือศาสนาพุทธ จำนวน 370 คน คิดเป็นร้อยละ 92.5 ผู้ป่วยที่นับถือศาสนาคริสต์ จำนวน 17 คน คิดเป็นร้อยละ 4.3 และผู้ป่วยที่นับถือศาสนาอิสลาม จำนวน 13 คน คิดเป็นร้อยละ 3.3 ดังแสดงในภาพประกอบ 9

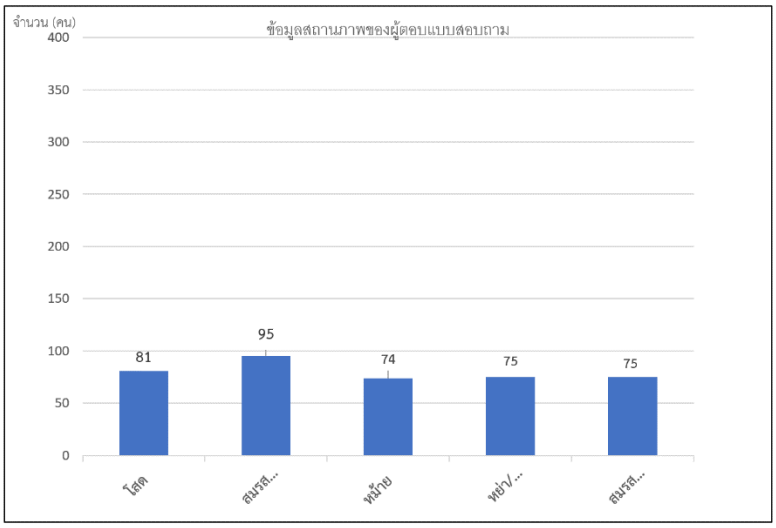


ภาพประกอบ 9 แผนภูมิข้อมูลนับถือศาสนาของผู้ตอบแบบสอบถาม

ตาราง 28 ข้อมูลสถานภาพของผู้ตอบแบบสอบถาม

ความหมายของข้อมูล	จำนวน	ร้อยละ
สถานภาพ		
โสด	**81**	**20.3**
สมรสและคู่สมรสอาศัยอยู่ด้วยกัน	**95**	**23.8**
หม้าย	**74**	**18.5**
หย่า/แยก	**75**	**18.8**
สมรสแต่คู่สมรสไม่ได้อาศัยอยู่ด้วยกัน	**75**	**18.8**
รวม		**100**

ตาราง 28 แสดงถึงข้อมูลของผู้ป่วยโรคเบาหวานจำนวน 400 คน แบ่งเป็นผู้ป่วยสถานภาพโสด จำนวน 81 คน คิดเป็นร้อยละ 20.3 ผู้ป่วยสถานภาพสมรสและคู่สมรสอาศัยอยู่ด้วยกัน จำนวน 95 คน คิดเป็นร้อยละ 23.8 ผู้ป่วยสถานภาพหม้าย จำนวน 74 คน คิดเป็นร้อยละ 18.5 ผู้ป่วยสถานภาพหย่า/แยก จำนวน 75 คน คิดเป็นร้อยละ 18.8 และผู้ป่วยสถานภาพสมรสแต่คู่สมรสไม่ได้อาศัยอยู่ด้วยกัน จำนวน 75 คน คิดเป็นร้อยละ 18.8 ดังแสดงในภาพประกอบ 10

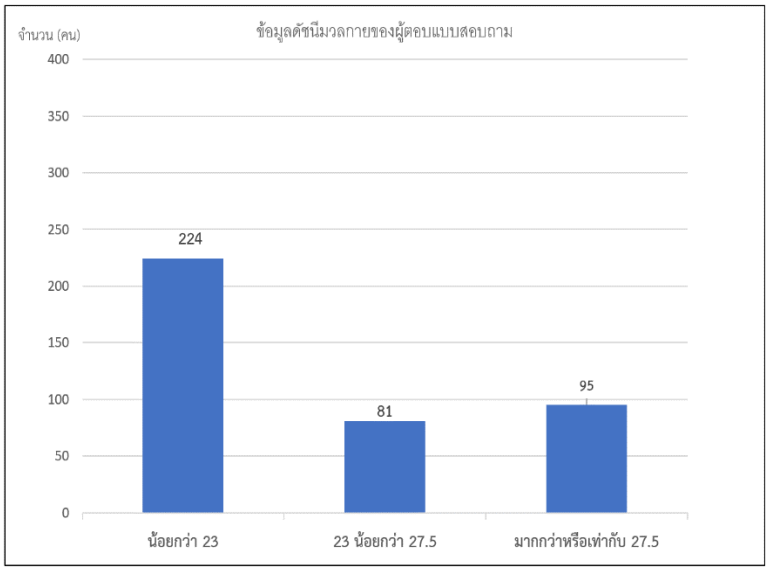


ภาพประกอบ 10 แผนภูมิข้อมูลสถานภาพของผู้ตอบแบบสอบถาม

ตาราง 29 ข้อมูลดัชนีมวลกายของผู้ตอบแบบสอบถาม

ความหมายของข้อมูล	**จำนวน**	**ร้อยละ**
ดัชนีมวลกาย		
น้อยกว่า 23	224	56.0
23 น้อยกว่า 27.5	81	20.3
มากกว่าหรือเท่ากับ 27.5	95	23.8
รวม		100

ตาราง 29 แสดงถึงข้อมูลของผู้ป่วยโรคเบาหวานจำนวน 400 คน แบ่งเป็นผู้ป่วยที่ดัชนีมวลกายน้อยกว่า 23 จำนวน 224 คน คิดเป็นร้อยละ 56.0 ผู้ป่วยที่ดัชนีมวลกาย 23 น้อยกว่า 27.5 จำนวน 81 คน คิดเป็นร้อยละ 20.3 และผู้ป่วยที่ดัชนีมวลกายมากกว่าหรือเท่ากับ 27.5จำนวน 95 คน คิดเป็นร้อยละ 23.8 ดังแสดงในภาพประกอบ 11

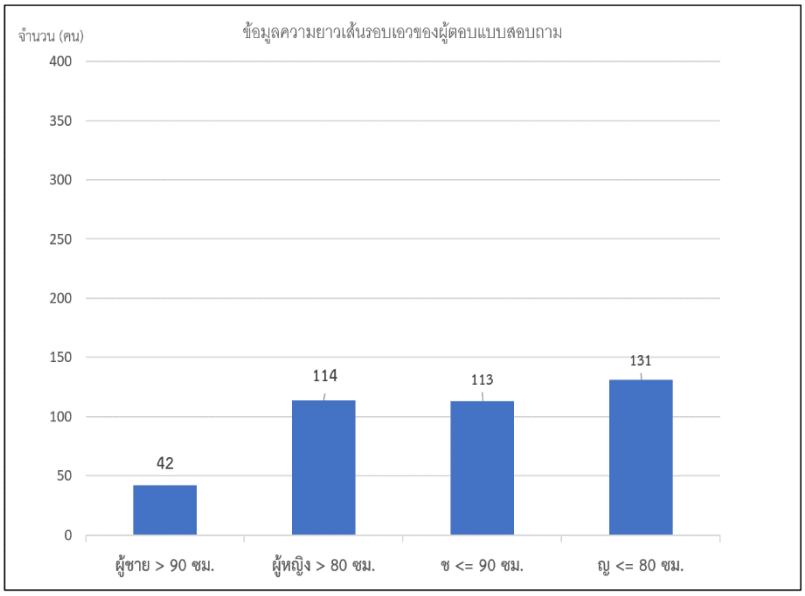


ภาพประกอบ 11 แผนภูมิข้อมูลดัชนีมวลกายของผู้ตอบแบบสอบถาม

ตาราง 30 ข้อมูลความยาวเส้นรอบเอวของผู้ตอบแบบสอบถาม

ความหมายของข้อมูล	จำนวน	ร้อยละ
ความยาวเส้นรอบเอว		
ผู้ชาย น้อยกว่า 90 ซม.	**142**	**35.5**
ผู้หญิง น้อยกว่า 80 ซม.	**140**	**35.0**
ผู้ชาย มากกว่าหรือเท่ากับ 90 ซม.	**62**	**15.5**
ผู้หญิง มากกว่าหรือเท่ากับ 80 ซม.	**56**	**14.0**
รวม		**100**

ตาราง 30 แสดงถึงข้อมูลของผู้ป่วยโรคเบาหวานจำนวน 400 คน แบ่งเป็นความยาวเส้นรอบเอวของผู้ป่วยที่ผู้ชาย น้อยกว่า 90 ซม. จำนวน 142 คน คิดเป็นร้อยละ 35.5 ความยาวเส้นรอบเอวของผู้ป่วยที่ผู้หญิงน้อยกว่า 80 ซม. จำนวน 140 คน คิดเป็นร้อยละ 35.0 ความยาวเส้นรอบเอวของผู้ป่วยที่ผู้ชาย มากกว่าหรือเท่ากับ 90 ซม.จำนวน 62 คน คิดเป็นร้อยละ 15.5 และความยาวเส้นรอบเอวของผู้ป่วยผู้หญิง มากกว่าหรือเท่ากับ 80 ซม.รวมจำนวน 56 คน คิดเป็นร้อยละ 14.0 ดังแสดงในภาพประกอบ 12

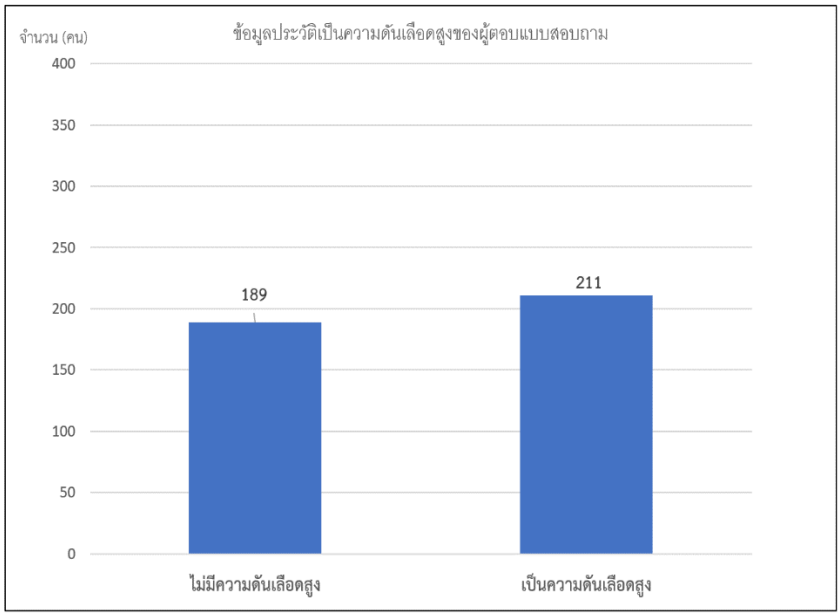


ภาพประกอบ 12 แผนภูมิข้อมูลความยาวเส้นรอบเอวของผู้ตอบแบบสอบถาม

ตาราง 31 ข้อมูลประวัติเป็นความดันเลือดสูงของผู้ตอบแบบสอบถาม

ความหมายของข้อมูล	จำนวน	ร้อยละ
ประวัติเป็นความดันเลือดสูง		
ไม่มีความดันเลือดสูง	**189**	**47.3**
เป็นความดันเลือดสูง	**211**	**52.8**
รวม		**100**

ตาราง 31 แสดงถึงข้อมูลของผู้ป่วยโรคเบาหวานจำนวน 400 คน แบ่งเป็นผู้ป่วยที่ไม่มีประวัติเป็นความดันเลือดสูง จำนวน 189 คน คิดเป็นร้อยละ 47.3 และผู้ป่วยที่มีประวัติเป็นความดันเลือดสูง จำนวน 211 คน คิดเป็นร้อยละ 52.8 ดังแสดงในภาพประกอบ 13

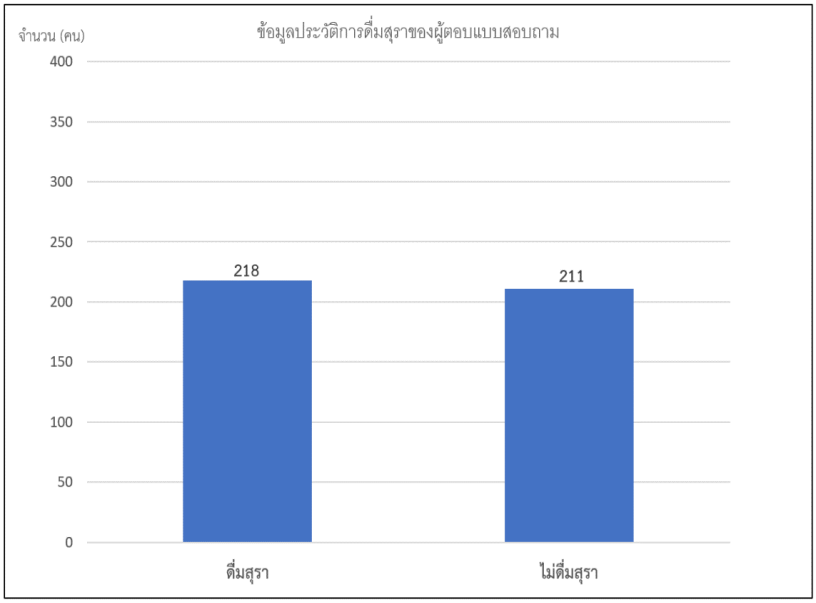


ภาพประกอบ 13 แผนภูมิข้อมูลประวัติเป็นความดันเลือดสูงของผู้ตอบแบบสอบถาม

ตาราง 32 ข้อมูลประวัติการดื่มสุราของผู้ตอบแบบสอบถาม

ความหมายของข้อมูล	จำนวน	ร้อยละ
ประวัติการดื่มสุรา		
ดื่มสุรา	**218**	**54.5**
ไม่ดื่มสุรา	**182**	**45.5**
รวม		**100**

ตาราง 32 แสดงถึงข้อมูลของผู้ป่วยโรคเบาหวานจำนวน 400 คน แบ่งเป็นผู้ป่วยที่ไม่มีดื่มสุรา จำนวน 218 คน คิดเป็นร้อยละ 54.5 และผู้ป่วยที่ไม่ดื่มสุรา จำนวน 182 คน คิดเป็นร้อยละ 45.5 ดังแสดงในภาพประกอบ 14

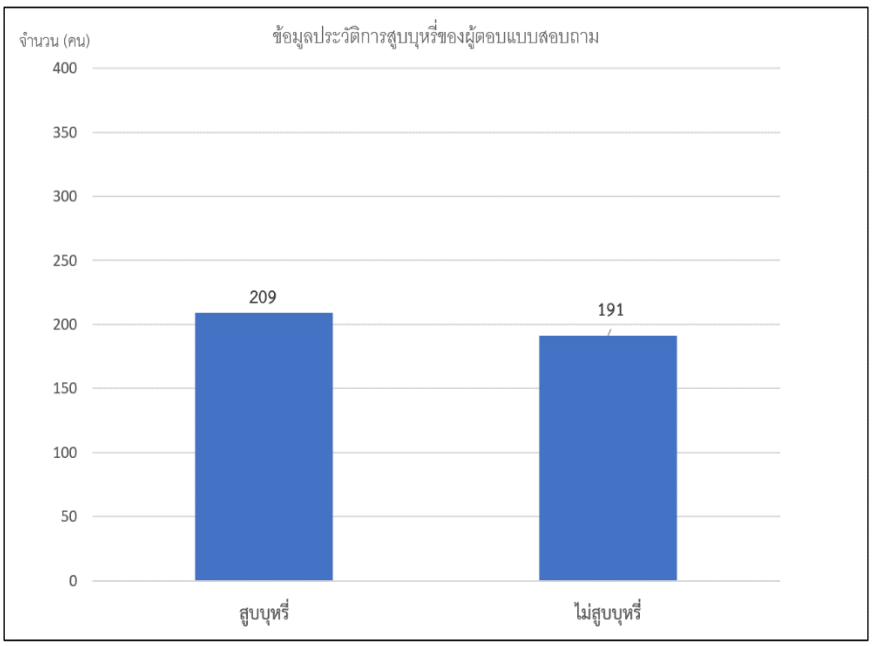


ภาพประกอบ 14 แผนภูมิข้อมูลประวัติการดื่มสุราของผู้ตอบแบบสอบถาม

ตาราง 33 ข้อมูลประวัติการสูบบุหรี่ของผู้ตอบแบบสอบถาม

ความหมายของข้อมูล	จำนวน	ร้อยละ
ประวัติการสูบบุหรี่		
สูบบุหรี่	**209**	**52.3**
ไม่สูบบุหรี่	**191**	**47.8**
รวม		**100**

ตาราง 33 แสดงถึงข้อมูลของผู้ป่วยโรคเบาหวานจำนวน 400 คน แบ่งเป็นผู้ป่วยที่สูบบุหรี่ จำนวน 209 คน คิดเป็นร้อยละ 52.3 และผู้ป่วยที่ไม่สูบบุหรี่ จำนวน 191 คน คิดเป็นร้อยละ 47.8 ดังแสดงในภาพประกอบ 15

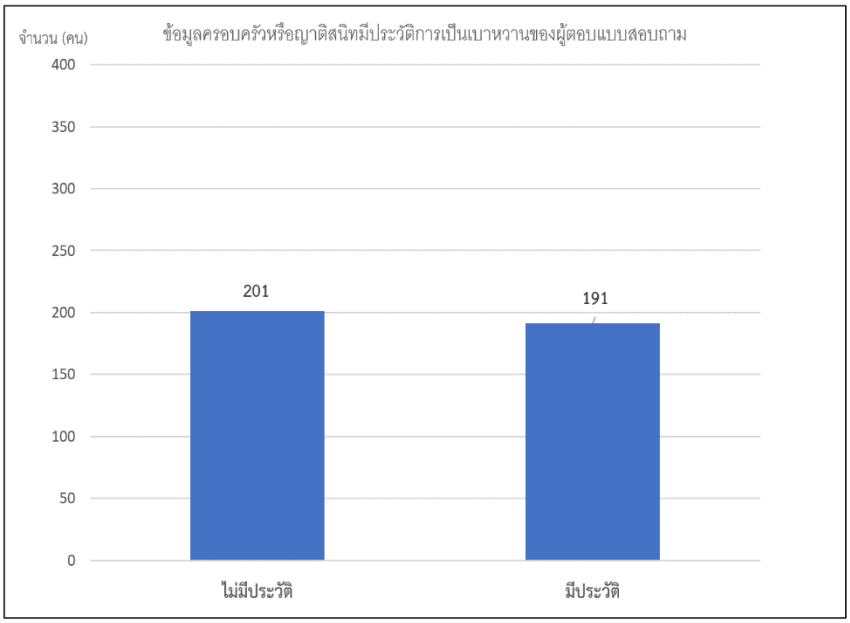


ภาพประกอบ 15 แผนภูมิข้อมูลประวัติการสูบบุหรี่ของผู้ตอบแบบสอบถาม

ตาราง 34 ข้อมูลครอบครัวหรือญาติสนิทมีประวัติการเป็นเบาหวานของผู้ตอบแบบสอบถาม

ความหมายของข้อมูล	จำนวน	ร้อยละ
ครอบครัวหรือญาติสนิทมีประวัติการเป็นเบาหวาน		
ไม่มีประวัติ	**201**	**50.3**
มีประวัติ	**199**	**49.8**
รวม		**100**

ตาราง 34 แสดงถึงข้อมูลของผู้ป่วยโรคเบาหวานจำนวน 400 คน แบ่งเป็นผู้ป่วยที่มีประวัติการเป็นเบาหวาน จำนวน 201 คน คิดเป็นร้อยละ 50.3 และผู้ป่วยที่ไม่มีประวัติการเป็นเบาหวาน จำนวน 199 คน คิดเป็นร้อยละ 49.8 ดังแสดงในภาพประกอบ 16

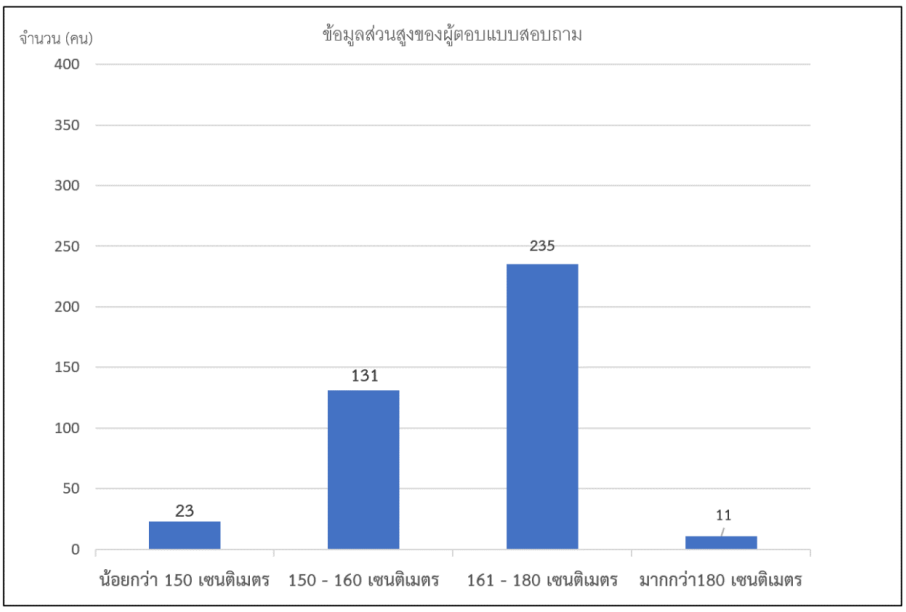


ภาพประกอบ 16 แผนภูมิข้อมูลครอบครัวหรือญาติสนิทมีประวัติการเป็นเบาหวานของผู้ตอบแบบสอบถาม

ตาราง 35 ข้อมูลส่วนสูงของผู้ตอบแบบสอบถาม

ความหมายของข้อมูล	จำนวน	ร้อยละ
ส่วนสูง		
น้อยกว่า 150 เซนติเมตร	**23**	**5.8**
150 - 160 เซนติเมตร	**131**	**32.8**
161 - 180 เซนติเมตร	**235**	**58.8**
มากกว่า180 เซนติเมตร	**11**	**2.8**
รวม		**100**

ตาราง 35 แสดงถึงข้อมูลของผู้ป่วยโรคเบาหวานจำนวน 400 คน แบ่งเป็นผู้ป่วยที่มีส่วนสูงน้อยกว่า 150 เซนติเมตร มีจำนวน 23 คน คิดเป็นร้อยละ 5.8 ผู้ป่วยที่มีส่วนสูง 150 – 160 เซนติเมตร มีจำนวน 131 คน คิดเป็นร้อยละ 32.8 ผู้ป่วยที่มีส่วนสูง161 - 180 เซนติเมตร มีจำนวน 235 คน คิดเป็นร้อยละ 58.8 และผู้ป่วยที่มีมากกว่า180 เซนติเมตร มีจำนวน 11 คน คิดเป็นร้อยละ 2.8 ดังแสดงในภาพประกอบ 17

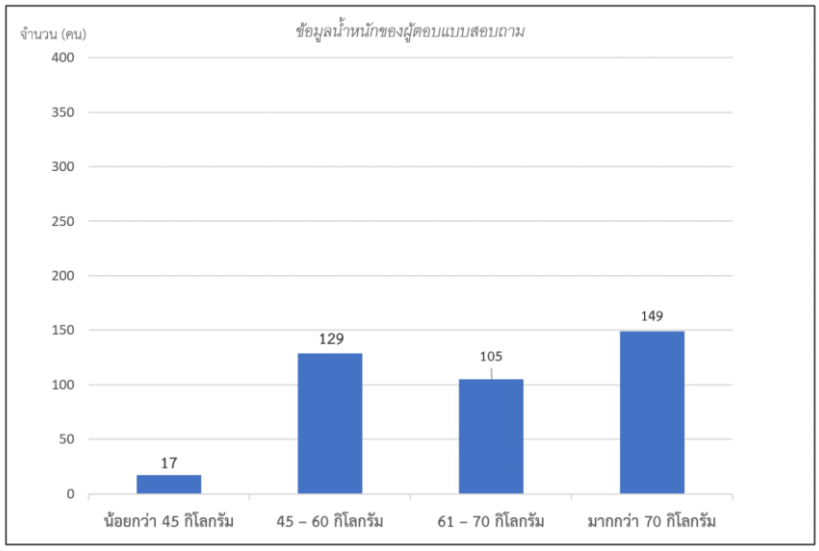


ภาพประกอบ 17 แผนภูมิข้อมูลส่วนสูงของผู้ตอบแบบสอบถาม

ตาราง 36 ข้อมูลน้ำหนักของผู้ตอบแบบสอบถาม

ความหมายของข้อมูล	จำนวน	ร้อยละ
น้ำหนัก		
น้อยกว่า 45 กิโลกรัม	17	4.3
45 – 60 กิโลกรัม	129	32.3
61 – 70 กิโลกรัม	105	26.3
มากกว่า 70 กิโลกรัม	149	37.3
รวม		100

ตาราง 36 แสดงถึงข้อมูลของผู้ป่วยโรคเบาหวานจำนวน 400 คน แบ่งเป็นผู้ป่วยที่มีน้ำหนักน้อยกว่า 45 กิโลกรัม มีจำนวน 17 คน คิดเป็นร้อยละ 4.3 ผู้ป่วยที่มีน้ำหนัก 45 – 60 กิโลกรัม มีจำนวน 129 คน คิดเป็นร้อยละ 32.3 ผู้ป่วยที่มีน้ำหนัก 61 – 70 กิโลกรัม มีจำนวน 105 คน คิดเป็นร้อยละ 26.3 และผู้ป่วยที่มีน้ำหนักมากกว่า 70 กิโลกรัม มีจำนวน 149 คน คิดเป็นร้อยละ 37.3 ดังแสดงในภาพประกอบ 18



ภาพประกอบ 18 แผนภูมิข้อมูลน้ำหนักของผู้ตอบแบบสอบถาม

4.3 การทดสอบและการคัดเลือกอัลกอริทึม

การทดสอบและการคัดเลือกอัลกอริทึมจะสามารถวัดประสิทธิภาพได้จากค่าความถูกต้อง (Accuracy) ค่าความแม่นยำของข้อมูล (Precision) ค่าความระลึกการวัดความถูกต้องของโมเดล (Recall) ค่าความถ่วงดุล (F-Measure) และค่าพื้นที่ใต้กราฟ (Receiver Operating Characteristic; ROC) ในการพยากรณ์การเกิดโรคเบาหวาน เพื่อจำแนกภาวะไม่มีความเสี่ยงและมีความเสี่ยง โดยมีรายละเอียดของวิธีการวัดประสิทธิภาพดังนี้

4.4 การวัดประสิทธิภาพของเครื่องมือ

การวัดประเมินผลมีความสำคัญเนื่องจากนำมาใช้ในการวิเคราะห์ประสิทธิภาพการ ทำงานของอัลกอริทึม สำหรับการวัดผลของอัลกอริทึมที่ใช้ในการจำแนกข้อมูล ตัวชี้วัดการ ประเมินผลสามารถที่จะพัฒนาจากเมทริกซ์ความสับสน (Confusion matrix) ซึ่งเป็นตาราง crosstabs ขนาด 2*2 โดยมีการกำหนดให้แนวนอน คือ ค่า Actual result และแนวตั้งคือ prediction result ดังตาราง 37

ตาราง 37 ตารางเมทริกซ์ความสับสน (Confusion Matrix)

Confusion matrix		Predicted result	
		Negative	Positive
Actual result	Negative	TN	FP
	Positive	FN	TP

โดยมีคำจำกัดความดังนี้

1. True Positive คือผลของการพยากรณ์พบว่ามีความเสี่ยงต่อโรคเบาหวานและผลที่เกิดขึ้นจริงพบว่ามีความเสี่ยงต่อโรคเบาหวาน

2. True Negative คือผลของการพยากรณ์พบว่าไม่มีความเสี่ยงต่อโรคเบาหวานและผลที่เกิดขึ้นจริงพบว่าไม่มีความเสี่ยงต่อโรคเบาหวาน

3. False Positive คือผลของการพยากรณ์พบว่ามีความเสี่ยงต่อโรคเบาหวานแต่พบว่าผลที่เกิดขึ้นจริงไม่มีความเสี่ยงโรคเบาหวาน

4. False Negative คือเครื่องมือมีการพยากรณ์ว่าไม่มีความเสี่ยงต่อโรคเบาหวานแต่พบว่าผลที่เกิดขึ้นจริงคือมีความเสี่ยงต่อโรคเบาหวาน

5. Condition Positive (P) คือจำนวนของผลที่แท้จริงที่มีความเสี่ยงต่อโรคเบาหวานทั้งหมดที่อยู่ในข้อมูล = TP + FN

6. Condition Negative (N) คือจำนวนของผลที่แท้จริงที่ไม่มีความเสี่ยงต่อโรคเบาหวานทั้งหมดที่อยู่ในข้อมูล = FP + TN

จากตาราง 20 สามารถนำมาคำนวณค่าความถูกต้องค่าความแม่นยำค่าระลึกค่าความถ่วงดุลเส้นกราฟ ROC ได้ดังนี้

1. ค่าความถูกต้อง (accuracy) เป็นการแสดงผลการวัดที่ได้มีความถูกต้องดังสมการ 4

$$accuracy = \frac{TP+TN}{TP+FN+FP+TN} \quad (4)$$

โดยค่าความถูกต้อง จะมีค่าอยู่ระหว่าง 0 – 100 ยิ่งเข้าใกล้ 100 แปลว่าแบบจำลองทำนายผลได้ดีมาก

2. ค่าความแม่นยำ (precision) คือความน่าจะเป็นที่แบบจำลองจะทำนายได้ถูกต้องจากการทำนายทั้งหมด ดังสมการ 5

$$precision = \left(\frac{TP}{TP+FP}\right) \quad (5)$$

3. ค่าความระลึก (recall) หรือค่าความไว (sensitivity) คือความน่าจะเป็นที่แบบจำลองสามารถทำนายได้ถูกต้อง เทียบกับจำนวนของผลที่แท้จริงที่เป็นใช่ทั้งหมดดังสมการ 6

$$\text{recall} = \left(\frac{TP}{TP+FP}\right) \quad (6)$$

4. ค่าความถ่วงดุล (F-Measure) คือค่าเฉลี่ยแบบ Harmonic mean ระหว่าง ค่าความแม่นยำ และค่าความระลึก ที่วัดความสามารถของโมเดล โดยไม่ต้องเลือกระหว่าง Precision, Recall เพราะทำการหาค่าเฉลี่ยให้แล้ว ดังสมการ 7

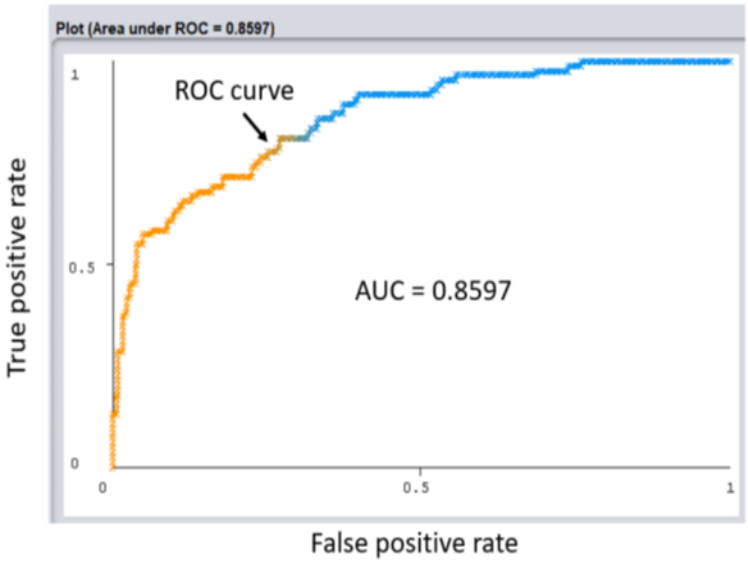
$$\text{F} = \left(\frac{2^{*}(\text{precision}^{*}\text{recall})}{precision+\text{recall}}\right) \quad (7)$$

5. เส้นโค้ง ROC (Receiver Operating Characteristic; ROC curve) และค่าพื้นที่ใต้กราฟ (Area Under Curve; AUC) เส้นโค้ง ROC เป็นเส้นโค้งที่พล็อตกราฟระหว่างค่า sensitivity และ 1-specificity โดยที่แกน x คือ false positive rate หรือ 1-specificity แกน y คือ true positive rate หรือ ค่าความไว (sensitivity) หรือ ค่าระลึก (recall) ตาราง 38

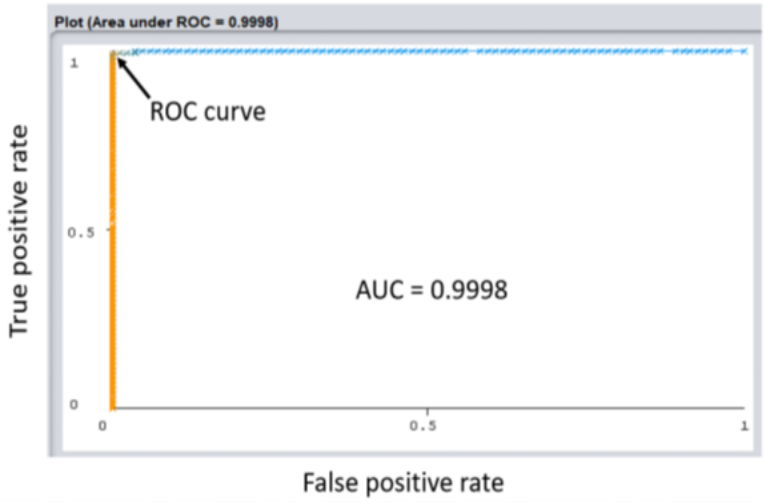
ตาราง 38 ตารางแสดงค่าการอ่านผลของ AUC

ค่าที่ได้	การอ่านผล
0.9-1.0	ดีมาก
0.8-0.9	ดี
0.7-0.8	ปกติ
0.6-0.7	ต่ำ
0.5-0.6	ผิดพลาด
<0.5	เป็นการทดสอบที่ไม่มีประสิทธิภาพ

จากตาราง 38 ยิ่งค่า AUC เข้าใกล้หนึ่งยิ่งแสดงถึงประสิทธิภาพการทดสอบที่มีความแม่นยำที่มาก แสดงถึงความน่าจะเป็นเครื่องมือที่ใช้ในการทดสอบจะให้ผลถูกต้องตามความเป็นจริง แสดงดังภาพประกอบ 19 และ 20



ภาพประกอบ 19 แสดงกราฟ ROC และพื้นที่ใต้กราฟเพื่อใช้ในการวัดประสิทธิภาพของโมเดลการ จำแนก ที่ AUC=0.8597



ภาพประกอบ 20 แสดงกราฟ ROC และพื้นที่ใต้กราฟเพื่อใช้ในการวัดประสิทธิภาพของโมเดลการ จำแนก ที่ AUC=0.9998

6. True Positive Rate หรือค่าความไว (Sensitivity) หรือค่าระลึก (Recall) เครื่องมือที่มี sensitivity สูง จะเป็นเครื่องมือที่เหมาะสมสำหรับคัดกรองโรค (screening test) หรือ อาจกล่าวได้ว่าสำหรับเป็นเครื่องมือ rule out โรคที่สงสัย เมื่อผลตรวจเป็นลบ เนื่องจากเมื่อ sensitivity สูงจะทำให้ค่า False negative rate (FNR) ต่ำ ดังนั้น เมื่อผลการทดสอบออกมาเป็นลบ ก็มักจะเป็นลบจริง ๆ หรือกล่าวได้ว่าโอกาสเป็นโรคจึงต่ำมากจริง ๆ วิธีการคำนวณดังสมการ 8

$$sensitivity = \frac{TP}{TP+TN} \quad (8)$$

7. True Negative Rate หรือมีค่าความจำเพาะ (Specificity) เครื่องมือ ที่มี specificity สูง จะเป็นเครื่องมือที่เหมาะสมสำหรับวินิจฉัยโรค (diagnostic test) หรืออาจกล่าวได้ว่าสำหรับเป็นเครื่องมือ rule in โรคที่สงสัย เมื่อผลตรวจเป็นบวก เนื่องจากเมื่อ specificity สูงจะทำให้ค่า False positive rate (FPR) ต่ำ ดังนั้น เมื่อผลการทดสอบออกมาเป็นบวก มักจะบวกจริง หรือกล่าวได้ว่าโอกาสเป็นโรคจริงจึงสูงมาก วิธีการคำนวณดังสมการ 9

$$specificity = \frac{TN}{FP+TN} \quad (9)$$

4.5 วิธีการดำเนินการวิจัย

ขั้นตอนในการดำเนินการวิจัยในการทดสอบประสิทธิภาพการพยากรณ์โรคเบาหวาน ของทั้งอัลกอริทึม 5 จะสามารถได้โดยการนำข้อมูลชุดเดียวกันนำเข้าตัวแบบทดสอบเพื่อให้สามารถตอบข้อสมมุติฐานนำข้อมูลป้อนเข้าที่ผ่านขั้นตอนการปรับโครงสร้างข้อมูลเรียบร้อยแล้วเข้าทดสอบกับ 5 อัลกอริทึม 1) เพอร์เซ็ปตรอนแบบหลายชั้น 2) นาอีฟ เบย์ 3) โลจิสติก 4) แรนดอมฟอเรส 5) ต้นไม้ตัดสินใจ

4.5.1 วิธีเพอร์เซ็ปตรอนหลายชั้น

เมื่อนำข้อมูลป้อนเข้าจำนวน 400 รายการ และทำการวิเคราะห์ด้วยวิธีเพอร์เซ็ปตรอนหลายชั้น ทำให้ได้ผลลัพธ์ดังที่แสดงในตาราง 39

ตาราง 39 ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยวิธีเพอร์เซ็ปตรอนหลายชั้น

ความถูกต้อง	โรคเบาหวาน	ความแม่นยำ	ค่าระลึก	ค่าถ่วงดุล	เส้นกราฟ ROC
98.5 %	ไม่มีความเสี่ยง	0.980	1.000	0.990	0.945
	มีความเสี่ยง	1.000	0.940	0.969	0.963

พิจารณาข้อมูลจากตาราง 39 วิเคราะห์ข้อมูลด้วยวิธีเพอร์เซ็ปตรอนนหลายชั้น ผลลัพธ์ที่ได้จากอัลกอริทึมให้ค่าความถูกต้องต่อการเกิดโรคเบาหวานเท่ากับ 98.5 มีค่าความแม่นยำจากการพยากรณ์ในรูปแบบที่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 1.000 และพยากรณ์ในรูปแบบที่ไม่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 0.980 มีค่าพยากรณ์ความระลึกในรูปแบบที่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 0.940 และพยากรณ์ในรูปแบบที่ไม่มีความเสี่ยงต่อการเกิด

โรคเบาหวานเท่ากับ 1.000 ค่าถ่วงดุลพยากรณ์ออกมาในรูปแบบที่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 0.969 และพยากรณ์ในรูปแบบที่ไม่มีความเสี่ยงการเกิดโรคเบาหวานมีค่าเท่ากับ 0.990 นอกจากนี้ยังพยากรณ์ค่า ROC ผ่านอัลกอริทึมในรูปแบบที่มีความเสี่ยงต่อการเกิดโรคเบาหวานให้ค่าเท่ากับ 0.963 และพยากรณ์ในรูปแบบที่ไม่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 0.945 แสดงให้เห็นประสิทธิภาพในการทดสอบการพยากรณ์มีความแม่นยำที่ดีมาก

4.5.2 วิธีนาอีฟ เบย์

เป็นหนึ่งในอัลกอริทึมที่อาศัยหลักการความน่าจะเป็นตามทฤษฎีของเบย์ (BayesTheorem) ซึ่งมีอัลกอริทึมที่ไม่ซับซ้อน โดยใช้ขั้นตอนในการจำแนกข้อมูลเพื่อการเรียนรู้ปัญหาที่เกิดขึ้น แล้วนำมาสร้างเงื่อนไขการจำแนกข้อมูลใหม่ โดยใช้หลักการในการคำนวณหาค่าน่าจะเป็นในการทำนายผล เป็นอัลกอริทึมในการแก้ปัญหาแบบจำแนกประเภทที่สามารถคาดการณ์ผลลัพธ์ได้และทำการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปร เพื่อใช้ในการสร้างเงื่อนไขความจะเป็นสำหรับแต่ละความสัมพันธ์ เมื่อนำข้อมูลป้อนเข้าจำนวน 400 รายการ และทำการวิเคราะห์ด้วยวิธีวิธีนาอีฟ เบย์ ทำให้ได้ผลลัพธ์ดังที่แสดงในตาราง 40

ตาราง 40 ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยวิธีนาอีฟ เบย์

ความถูกต้อง	โรคเบาหวาน	ความแม่นยำ	ค่าระลึก	ค่าถ่วงดุล	เส้นกราฟ ROC
78 %	ไม่มีความเสี่ยง	0.785	0.973	0.869	0.775
	มีความเสี่ยง	0.714	0.200	0.313	0.755

พิจารณาข้อมูลจากตาราง 40 วิเคราะห์ข้อมูลด้วยวิธีนาอีฟ เบย์ ผลลัพธ์ที่ได้จากอัลกอริทึมให้ค่าความถูกต้องต่อการเกิดโรคเบาหวานเท่ากับ 78 มีค่าความแม่นยำจากการพยากรณ์ในรูปแบบที่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 0.714 และพยากรณ์ในรูปแบบที่ไม่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 0.758 มีค่าพยากรณ์ความระลึกในรูปแบบที่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 0.200 และพยากรณ์ในรูปแบบที่ไม่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 0.973 ค่าถ่วงดุลพยากรณ์ออกมาในรูปแบบที่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 0.313 และพยากรณ์ในรูปแบบที่ไม่มีความเสี่ยงการเกิดโรคเบาหวานมีค่าเท่ากับ 0.869 นอกจากนี้ยังพยากรณ์ค่า ROC ผ่านอัลกอริทึมในรูปแบบที่มีความเสี่ยงต่อการเกิดโรคเบาหวานให้ค่าเท่ากับ 0.755 และพยากรณ์ในรูปแบบที่ไม่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 0.755 แสดงให้เห็นประสิทธิภาพในการทดสอบการพยากรณ์มีความแม่นยำที่ดี

4.5.3 วิธีโลจิสติกพหุกลุ่ม

ทำการนำเข้าข้อมูลจำนวน 400 รายการ ทำการวิเคราะห์ด้วยวิธีโลจิสติกพหุกลุ่ม ทำให้ได้ผลลัพธ์ดังที่แสดงในตาราง 41

ตาราง 41 ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยวิธีโลจิสติกพหุกลุ่ม

ความถูกต้อง	โรคเบาหวาน	ความแม่นยำ	ค่าระลึก	ค่าถ่วงดุล	เส้นกราฟ ROC
77.25 %	ไม่มีความเสี่ยง	0.783	0.963	0.864	0.733
	มีความเสี่ยง	0.645	0.200	0.305	0.733

พิจารณาข้อมูลจากตาราง 41 วิเคราะห์ข้อมูลด้วยวิธีวิธีโลจิสติกพหุกลุ่ม ผลลัพธ์ที่ได้จากอัลกอริทึมให้ค่าความถูกต้องต่อการเกิดโรคเบาหวานเท่ากับ 77.25 มีค่าความแม่นยำจากการพยากรณ์ในรูปแบบที่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 0.645 และพยากรณ์ในรูปแบบที่ไม่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 0.783 มีค่าพยากรณ์ความระลึกในรูปแบบที่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 0.200 และพยากรณ์ในรูปแบบที่ไม่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 0.963 ค่าถ่วงดุลพยากรณ์ออกมาในรูปแบบที่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 0.305 และพยากรณ์ในรูปแบบที่ไม่มีความเสี่ยงการเกิดโรคเบาหวานมีค่าเท่ากับ 0.864 นอกจากนี้ยังพยากรณ์ค่า ROC ผ่านอัลกอริทึมในรูปแบบที่มีความเสี่ยงต่อการเกิดโรคเบาหวานให้ค่าเท่ากับ 0.733 และพยากรณ์ในรูปแบบที่ไม่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 0.733 แสดงให้เห็นประสิทธิภาพในการทดสอบการพยากรณ์มีความแม่นยำที่ดี

4.5.4 วิธีเทคนิคป่าสุ่ม

นำข้อมูลป้อนเข้าจำนวน 400 รายการ และทำการวิเคราะห์ด้วยวิธีเทคนิคป่าสุ่ม ทำให้ได้ผลลัพธ์ดังที่แสดงในตาราง 42

ตาราง 42 ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยวิธีเทคนิคป่าสุ่ม

ความถูกต้อง	โรคเบาหวาน	ความแม่นยำ	ค่าระลึก	ค่าถ่วงดุล	เส้นกราฟ ROC
99.75 %	ไม่มีความเสี่ยง	0.997	1.000	0.998	1.000
	มีความเสี่ยง	1.000	0.990	0.995	1.000

พิจารณาข้อมูลจากตาราง 42 วิเคราะห์ข้อมูลด้วยวิธีแรนเทคนิคป่าสุ่ม ผลลัพธ์ที่ได้จากอัลกอริทึมให้ค่าความถูกต้องต่อการเกิดโรคเบาหวานเท่ากับ 99.75 มีค่าความแม่นยำจากการพยากรณ์ในรูปแบบที่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 1.000 และพยากรณ์ในรูปแบบที่ไม่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 0.997 มีค่าพยากรณ์ความระลึกในรูปแบบที่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 0.990 และพยากรณ์ในรูปแบบที่ไม่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 1.000 ค่าถ่วงดุลพยากรณ์ออกมาในรูปแบบที่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 0.995 และพยากรณ์ในรูปแบบที่ไม่มีความเสี่ยงการเกิดโรคเบาหวานมีค่าเท่ากับ 0.998 นอกจากนี้ยังพยากรณ์ค่า ROC ผ่านอัลกอริทึมในรูปแบบที่มีความเสี่ยงต่อการเกิดโรคเบาหวานให้ค่าเท่ากับ 1.000 และพยากรณ์ในรูปแบบที่ไม่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 1.000 แสดงให้เห็นประสิทธิภาพในการทดสอบการพยากรณ์มีความแม่นยำที่ดีมาก

4.5.5 วิธีต้นไม้ตัดสินใจ

ทำการนำเข้าข้อมูลจำนวน 400 รายการ ทำการวิเคราะห์ด้วยวิธีต้นไม้ตัดสินใจ ทำให้ได้ผลลัพธ์ดังที่แสดงในตาราง 43

ตาราง 43 ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ

ความถูกต้อง	โรคเบาหวาน	ความแม่นยำ	ค่าระลึก	ค่าถ่วงดุล	เส้นกราฟ ROC
94.75 %	ไม่มีความเสี่ยง	0.946	0.987	0.966	0.972
	มีความเสี่ยง	0.954	0.830	0.888	0.972

พิจารณาข้อมูลจากตาราง 43 วิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ ผลลัพธ์ที่ได้จากอัลกอริทึมให้ค่าความถูกต้องต่อการเกิดโรคเบาหวานเท่ากับ 94.75 มีค่าความแม่นยำจากการพยากรณ์ในรูปแบบที่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 0.954 และพยากรณ์ในรูปแบบที่ไม่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 0.946 มีค่าพยากรณ์ความระลึกในรูปแบบที่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 0.830 และพยากรณ์ในรูปแบบที่ไม่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 0.987 ค่าถ่วงดุลพยากรณ์ออกมาในรูปแบบที่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 0.888 และพยากรณ์ในรูปแบบที่ไม่มีความเสี่ยงการเกิดโรคเบาหวานมีค่าเท่ากับ 0.966 นอกจากนี้ยังพยากรณ์ค่า ROC ผ่านอัลกอริทึมในรูปแบบที่มีความเสี่ยงต่อการเกิดโรคเบาหวานให้ค่าเท่ากับ 0.972 และพยากรณ์ในรูปแบบที่ไม่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 0.972 แสดงให้เห็นประสิทธิภาพในการทดสอบการพยากรณ์มีความแม่นยำที่ดีมาก ผลความถูกต้องจาก 5 อัลกอริทึ่มสามารถแสดงได้ในตาราง 44

ตาราง 44 แสดงผลลัพธ์เปรียบเทียบทั้ง 5 อัลกอริทึม

อัลกอริที่ม	ความถูกต้อง	ความแม่นยำ	ค่าระลึก	F-measure	ROC
เพอร์เซ็ปตรอนหลายชั้น	98.50	0.985	0.985	0.985	0.952
นาอีฟ เบย์	78.00	0.767	0.780	0.730	0.755
โลจิสติกพหุกลุ่ม	77.25	0.749	0.773	0.724	0.733
แรนดอมฟลอเรส	99.75	0.998	0.998	0.997	1.000
ต้นไม้ตัดสินใจ	94.75	0.948	0.948	0.946	0.972

พิจารณาข้อมูลในตาราง 44 ผลลัพธ์ที่ได้จากทั้ง 5 อัลกอริทึมพบว่าวิธีเทคนิคป่าสุ่ม ให้ค่าความถูกต้องในการพยากรณ์มากที่สุด ที่ 99.75 ค่า ความแม่นยำที่ 0.998 ค่าระลึกที่ 0.998 ค่า F-measure ที่ 0.997 และเส้นกราฟ ROC ที่ 1.000
ดังนั้นจึงสรุปได้ว่าวิธีเทคนิคป่าสุ่ม มีประสิทธิภาพในการพยากรณ์การเกิดโรคเบาหวานได้ดีที่สุด

บทที่ 5
สรุปผลการวิจัยและข้อเสนอแนะ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาปัจจัยเสี่ยงที่ส่งผลต่อการก่อให้เกิดโรคเบาหวานและเพื่อการพยากรณ์การเกิดโรคเบาหวาน โดยใช้เครื่องมือเป็นเหมืองข้อมูลในการพยากรณ์และวัดประสิทธิภาพความแม่นยำของโปรแกรมในการพยากรณ์ เพื่อเป็นแนวทางในการช่วยในการคัดกรองผู้ป่วยโรคเบาหวานของ โรงพยาบาลกันทรวิชัย อำเภอกันทรวิชัย จังหวัดมหาสารคาม ช่วยลดเวลา ลดค่าใช้จ่ายในการตรวจหาโรคเบาหวาน และทำให้มีความน่าเชื่อถือขึ้นให้มีประสิทธิภาพต่อไป ผู้วิจัยใช้วิธีการวิจัยเชิงสำรวจ (Survey Research) โดยเก็บรวมรวมข้อมูลจาก ประชาชนและผู้ป่วยที่เป็นโรคเบาหวาน โรงพยาบาลกันทรวิชัย อำเภอกันทรวิชัย จังหวัดมหาสารคาม ปีพุทธศักราช 2565 โดยเครื่องที่ใช้เก็บรวบรวมข้อมูลเป็นแบบสอบถาม ซึ่งได้ข้อมูลมาทั้งหมด 400 ชุด โดยแบบสอบถามที่ใช้แบ่งออกเป็น 3 ตอน ดังนี้

ตอนที่ 1 แบบสอบถามสภาพทั่วไป จำนวน 13 ข้อ

ตอนที่ 2 แบบสอบถามเกี่ยวกับพฤติกรรมความเสี่ยงที่ก่อให้เกิดโรคเบาหวาน จำนวน 28 ข้อ

ตอนที่ 3 ข้อเสนอแนะอื่นๆ

ในปัจจุบันการใช้เทคนิคเหมืองข้อมูลนั้น มีอิทธิผลต่อการตัดสินใจและใช้จัดการกับข้อมูลที่ซับซ้อนได้อย่างแม่นยำและน่าเชื่อถือหรือในการพยากรณ์ก็เช่นกัน ดังนั้นจึงนำเทคนิคเหมืองข้อมูลมาพยากรณ์โรคเบาหวานและใช้เครื่องมือเป็นแบบสอบถามเพื่อเป็นข้อมูลในการวัดเกณฑ์ความเสี่ยงต่างๆ ที่จะก่อให้เกิดโรคเบาหวาน ทางโรงพยาบาลกันทรวิชัยจะเป็นประโยชน์มากในการรับรู้ความเสี่ยงล่วงหน้าที่จะเกิดขึ้น ทั้งเรื่องการประหยัดเวลา การประหยัดค่าใช้จ่าย รวมถึงการรู้ถึงความเสี่ยงที่จะเกิดขึ้น ผลการศึกษาวิจัยเรื่องการวิเคราะห์ปัจจัยที่ก่อให้เกิดโรคเบาหวานด้วยเทคนิคเหมืองข้อมูล สามารถทำนายผลได้แม่นยำโดยใช้เครื่องมือ แรนดอมฟอร์เรท มากกว่าเครื่องมืออื่นๆ

ข้อเสนอแนะ

ผลการวิจัยจากการศึกษาครั้งนี้ สามารถนำไปใช้ได้ดังนี้

1. เป็นแนวทางในการใช้เพื่อพยากรณ์ภาวะเบาหวานสำหรับประชาชนทั่วไปเบื้องต้น เพื่อให้ประชาชนสามารถคาดการณ์ภาวะเบาหวานที่จะเกิดขึ้นกับตนเอง ผ่านแบบสอบถาม เพื่อเป็นการคัดกรองเบื้องต้นด้วยตนเอง

2. เป็นแนวทางในการประยุกต์เทคนิคเหมืองข้อมูลมาใช้เพื่อการพยากรณ์ภาวะเบาหวาน

ข้อเสนอแนะในการทำวิจัยครั้งต่อไป

1. การวิจัยนี้เป็นการวิจัยที่รวบรวมข้อมูลเพียงประชากรที่อยู่ในจังหวัดมหาสารคาม ไม่ได้เป็นตัวแทนของประชากรในประเทศ

2. การวิจัยนี้ใช้เครื่องมือเหมืองข้อมูลที่เปรียบเทียบอัลกอริทึมจำนวน 5 อัลกอริทึม ซึ่งอาจไม่ครอบคลุมเครื่องมือที่มีอยู่ในปัจจุบันทั้งหมด

บรรณานุกรม

บรรณานุกรม

กมลพร สิริคุตจตุพร, วิราพรรณ วิโรจน์รัตน์, & นารีรัตน์ จิตรมนตรี. (2560). ปัจจัยทำนายพฤติกรรมการจัดการตนเองของผู้สูงอายุโรคเบาหวานชนิดที่ 2;Factors Predicting Self-Management Behaviour of Elderly Type 2 Diabetes Patients. *วารสารสภาการพยาบาล*, *32*(1), 81–93. https://he02.tci-thaijo.org/index.php/TJONC/article/view/82615

กระทรวงสาธารณสุข กรมอนามัย. (2561). *อัตราผู้ป่วยเบาหวานรายใหม่จากกลุ่มเสี่ยงเบาหวาน เขตฯ 07*. http://healthkpi.moph.go.th/kpi2/kpi/index/?id=957&kpi_year=2561&lv=2&z=07

กระทรวงสาธารณสุข กรมอนามัย. (2565). *ข้อมูลประชากร*. https://dashboard.anamai.moph.go.th/population/pop-all/changwat?year=2017&cw=44

กฤชฐา จีระวงศ์พานิช, จามจุรี เวียงนาค, หยาดฝน ดิษบงค์, & อดุลย์ คร้ามสมบุญ. (2564). การศึกษาเบื้องต้นการดูแลผู้ป่วยเบาหวาน ชนิดที่ 1 ในผู้ใหญ่ตอนต้น ณ โรงพยาบาลมหาวิทยาลัยบูรพา. *บูรพาเวชสาร*, *8*(1), 41–55. https://he01.tci-thaijo.org/index.php/BJmed/article/view/250475

กองโรคไม่ติดต่อ กรมควบคุมโรค กระทรวงสาธารณสุข. (2564). *กองโรคไม่ติดต่อ กรมควบคุมโรค กระทรวงสาธารณสุข*. http://www.thaincd.com/2016/mission/documents.php?tid=32&gid=1-020

กัลยา บุญหล้า, & วรวุฒิ มหาโพธิ์. (2565). การจำแนกคุณภาพน้ำในประเทศไทยโดยใช้เทคนิคเหมืองข้อมูล. *วารสารวิทยาศาสตร์และเทคโนโลยี*, *30*(5), 29–30. https://doi.org/10.14456/tstj.2022.49

กิตติคุณ แสงนิล, & ประสพชัย พสุนนท์. (2561). ความน่าเชื่อถือ ความถูกต้อง ความแม่นยำ และความเที่ยงตรง. *วิทยาศาสตร์และเทคโนโลยีมหาวิทยาลัยศิลปากร Veridian E-Journal*, *5*(6), 5–5.

กิตติศักดิ์ พรมดาว, กมลทิพย์ เทียมทัด, & สุปราณี การพึ่งตน. (2556). ความรู้และพฤติกรรมเสี่ยงต่อการเกิดโรคความดันโลหิตสูงและโรคเบาหวานของนักเรียนชั้นมัธยมศึกษาตอนปลาย อำเภอเมือง จังหวัดอุบลราชธานี. *วารสารสุขภาพและการศึกษาพยาบาล*, *19*(1), 19–30. https://he02.tci-thaijo.org/index.php/Jolbcnm/article/view/10594

เขมารดี มาสิงบุญนิ, ภาวรรณ สามารถกิจ, & แก้วโบราณ, ส. (2562). ปัจจัยทำนายพฤติกรรมป้องกันโรคเบาหวานในวัยรุ่นที่มีภาวะเสี่ยงต่อโรคเบาหวาน ในจังหวัดสมุทรปราการ Factors Predicting Diabetes Prevention Behaviors among Adolescents at Risk For Type II

Diabetes in Samutprakan. *Journal of Nursing and Health Care*, *37*(2), 218–227. https://he01.tci-thaijo.org/index.php/jnat-ned/article/view/200878

คณะแพทยศาสตร์โรงพยาบาลรามาธิบดี มหาวิทยาลัยมหิดล. (2561). *วัยสูงอายุ | คณะแพทยศาสตร์โรงพยาบาลรามาธิบดี มหาวิทยาลัยมหิดล*. https://www.rama.mahidol.ac.th/ramamental/generalknowledge/general/06162014-1444

จรรยา สุนทรปกาสิต, & ขณัฏฐา เพียงแก้ว. (2561). การพัฒนารูปแบบการส่งเสริมสุขภาพเชิงรุกเพื่อป้องกันการเกิดโรคเบาหวานรายใหม่ใน. *วารสารวิชาการสำ นักงานสาธารณสุขจังหวัดมหาสารคาม*, *2*(3), 3–3.

จารี ทองคำ, วาทินี สุขมาก, & ภีมพศ สุขมาก. (2561). การเปรียบเทียบประสิทธิภาพของเทคนิค Apriori และ FP-Growth ในการสร้างกฎความสัมพันธ์ของโรคมะเร็งต่อมลูกหมาก Performance Comparison of Apriori and FP-Growth Techniques in Generating Association Rules to Prostate Cancer. *วารสารวิทยาการสารสนเทศและเทคโนโลยีประยุกต์* , *1*(2).

จิตตรา ตันเจริญ, อรอนงค์ บุรีเลิศ, นพรัตน์ ส่งเสริม, & ญาณิฐา แพงประโคน. (2561). ความชุกของความเสี่ยงต่อโรคเบาหวานในนักเรียนชั้นประถมศึกษา เขตสุขภาพที่ 10. *วารสาร UBRU เพื่อการวิจัยทางสาธารณสุข*, *7*(2), 18–25. https://he02.tci-thaijo.org/index.php/ubruphjou/article/view/170890

จิราภรณ์ ชิณโสม, & วิพร เสนารักษ์. (2554). สถานการณ์ความเสี่ยงต่อการเกิดโรคเบาหวานชนิดที่ 2 ของผู้หญิงอีสาน. *วารสารสมาคมพยาบาลฯ สาขาภาคตะวันออกเฉียงเหนือ*, *29*(1), 31–40. https://he01.tci-thaijo.org/index.php/jnat-ned/article/view/5899

จุฑาทิพย์ ทิพย์พูล, & นิเวศ จิระวิชิตชัย. (2559). การจำแนกจดหมายอิเล็กทรอนิกส์ที่เป็นสแปมโดยใช้เทคนิคเหมืองข้อมูล. *วารสาร มทร. วิทยาศาสตร์และเทคโนโลยี* , *6*(1), 104–105.

ฉวีวรรณ ศรีดาวเรือง, & จิราพร วรวงศ์. (2565). ความสัมพันธ์ระหว่างความรอบรู้ด้านสุขภาพกับพฤติกรรมสุขภาพตามหลัก 3อ. 2ส. ของผู้สูงอายุที่ป่วยเป็นโรคเบาหวานชนิดที่ 2. *วารสารศูนย์อนามัยที่ 9*, *9*(2), 456–456.

ชลลดา ไชยกุลวัฒนา, วิชานีย์ ใจมาลัย, & ประกายดาว สุทธิ. (2559). ปัจจัยทำนายพฤติกรรมสุขภาพของประชากรกลุ่มเสี่ยงสูงต่อ โรคเบาหวาน จังหวัดพะเยา. *วารการพยาบาลและการดูแลสุขภาพ* , *34*(4), 115–122. https://he01.tci-thaijo.org/index.php/jnat-ned/article/view/76735

ฐิติภัทร จันเกษม, & ปัทมา สุพรรณกุล. (2562). ปัจจัยพยากรณ์ระดับน้ำตาลในเลือดของผู้ป่วยโรคเบาหวานชนิดที่ 2. *วารสารวิชาการมหาวิทยาลัยอีสเทิร์นเอเชีย ฉบับวิทยาศาสตร์และ*

เทคโนโลยี, *13*(1).

ณัฐพล แสนคำ, ธนากร ปุรารัมย์, & ทิพวัลย์ แสนคำ. (2560). ระบบสนับสนุนทางการแพทย์สำหรับคัดกรองผู้ป่วยโรคไตเรื้อรังโดยใช้เทคนิคเหมืองข้อมูล. *วารสารวิชาการ โรงเรียนนายร้อยพระจุลจอมเกล้า*, *15*(1), 161–170. https://ph01.tci-thaijo.org/index.php/crma-journal/article/view/243110

ณัฐวดี หงษ์บุญมี, & ประภาสิริ ตรีพาณิชกุล. (2562). การเปรียบเทียบประสิทธิภาพการจำแนกข้อมูลเพื่อวิเคราะห์ปัจจัยความเสี่ยงที่ส่งผลต่อการเกิดโรคไฮเปอร์ไทรอยด์ด้วยเทคนิคเหมืองข้อมูล. *วารสารวิทยาศาสตร์และเทคโนโลยีสารสนเทศ*, *9*(1).

ดารณี ทองสัมฤทธิ์, กนกวรรณ บริสุทธิ์, & เยาวลักษณ์มีบุญมาก. (2560). ผลของโปรแกรมการปรับเปลี่ยนพฤติกรรมสุขภาพต่อความสามารถในการดูแลตนเองในกลุ่มเสี่ยงเบาหวาน ตำบลวัดเพลง อำเภอวัดเพลง จังหวัดราชบุรี. *วารสารวิทยาลัยพยาบาลพระปกเกล้า*, *28*(1), 26–37. https://he01.tci-thaijo.org/index.php/pnc/article/view/117837

ตัวช่วย ดวงแก้ว, & สายัญ สายยศ. (2562). การวิเคราะห์ปัจจัยที่ส่งผลต่อพัฒนาการล่าช้าตามอายุของเด็กปฐมวัยด้วยเทคนิคเหมืองข้อมูล. *วารสารวิทยาศาสตร์และเทคโนโลยีสารสนเทศ (JIST)*, *9*(2).

ทิพวรรณ นมเนย วรวรรณ ทิพย์วารีรมย์. (2561). ปัจจัยที่มีอิทธิพลต่อพฤติกรรมการดูแลตนเองของประชาชน กลุ่มเสี่ยงโรคเบาหวานชาวไทยทรงดำ จังหวัดสุโขทัย. *วารสารพยาบาลและสุขภาพ*, *12*(2), 164–173. https://he01.tci-thaijo.org/index.php/NurseNu/article/view/171915

ธนัท จรณะสมบูรณ์. (2561). การทำนายการซื้อซ้ำของผู้ซื้อโดยใช้เทคนิคการเรียนรู้ของเครื่องจักร. *บัณฑิตวิทยาลัยมหาวิทยาลัยศรีนครินทรวิโรฒ*, 4–5.

นงเยาว์ ในอรุณ. (2564). การเปรียบเทียบประสิทธิภาพของแบบจำลองการทำนายความเสี่ยงโรคหัวใจ และหลอดเลือดโดยใช้อัลกอริทึมเหมืองข้อมูล Efficiency comparison of cardiovascular risk prediction models using data mining algorithms ผู้ ช่ วย ศาสตราจารย์ , สาขาวิ ชาเทคโนโลยี สารสนเทศ คณะวิ ทยาศาสตร์ และเทคโนโลยี มหาวิทยาลั ยราชภั ฏพระนครศรี อยุ ธยา, จั งหวั ดพระนครศรี อยุ ธยา 13000. *วารสารวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยมหาสารคาม*, *40*(2).

นพนัฐ จำปาเทศ, ละเอียด แจ่มจันทร์, & พัชราภรณ์ ฝ่ายหมื่นไวย์. (2561). บทบาทพยาบาลในการประยุกต์ใช้ทฤษฎีการรับรู้ค. *วารสารมหาวิทยาลัยหัวเฉียวเฉลิมพระเกียรติ วิชาการ 153*, *21*(42), 155–156. https://webcache.googleusercontent.com/search?q=cache:XiltAVnWTaAJ:https://he01.tci-thaijo.org/index.php/HCUJOURNAL/article/download/146787/108183/&cd=7&hl=

th&ct=clnk&gl=th

นพรัตน์ นนท์ศิริ, พิศณุ ชัยจิตวณิชกุล, & กริช สมกันธา. (2565). การจำแนกข้อมูลเพื่อวินิจฉัยความเสี่ยงการเป็นโรคเบาหวานโดยใช้เทคนิค วิธีแบบร่วมกันตัดสินใจและวิธีเลือกคุณลักษณะเด่นไปข้างหน้า. *วารสารวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏอุดรธานี*, *10*(2), 107–122. https://ph01.tci-thaijo.org/index.php/scudru/article/view/245642

เบญจภัค จงหมื่นไวย์. (2558). การเปรียบเทียบปัจจัยโรคประจำตัวผู้สูงอายุโดยใช้อัลกอริทึมการจัดกลุ่ม J48 และNaiveBayes กรณีศึกษาสาธารณสุขโพธิ์กลางนครราชสีมา. *วารสารโครงงานวิทยาการคอมพิวเตอร์และเทคโนโลยีสารสนเทศ* , *1*(2).

ประการ เข้มแข็ง, นันทยา อ่อนคง, & มณีรัตน์ วงศ์พุ่ม. (2559). การศึกษาเปรียบเทียบความรอบรู้ด้านสุขภาพและพฤติกรรมสุขภาพตามหลัก 3อ.2ส. ในกลุ่มเสี่ยง โรคเบาหวาน ความดันโลหิตสูง ในเขตเมืองและเขตชนบท จังหวัดอุตรดิตถ์. *วารสารวิชาการป้องกันควบคุมโรค สคร.2 พิษณุโลก*, *4*(1), 22–22. https://he01.tci-thaijo.org/index.php/dpcphs/article/view/252075

ปราณี จันธิมา, & สมเกียรติ ศรีธาราธิคุณ. (2560). ผลของการสนับสนุนการจัดการตนเองต่อการปรับเปลี่ยนพฤติกรรมสุขภาพ ในกลุ่มเสี่ยงโรคอ้วนลงพุง. *วารสารพยาบาล มหาวิทยาลัยเชียงใหม่*, *44*(2), 162–171. https://he02.tci-thaijo.org/index.php/cmunursing/article/view/97857

ปริยานุช ประเสริฐสิริกุล, ศิริสรรพ เหล่าหะเกียรติ, เรืองศักดิ์ ตระกูลพุทธิรักษ์, & ศศิวิมล สุขพัฒน. (2565). การเปรียบเทียบประสิทธิภาพแบบจำลองการทำนายผลการเรียนของนิสิตที่ใช้งานระบบการจัดการเรียนรู้ออนไลน์ด้วยการเรียนรู้ของเครื่อง | วารสารวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏบุรีรัมย์ (ออนไลน์). *วารสารวิทยาศาสตร์และเทคโนโลยีมหาวิทยาลัยราชภัฏบุรีรัมย์*, *6*(1), 81–82. https://ph02.tci-thaijo.org/index.php/scibru/article/view/245527

ปิยะมาศ เอมอิ่มอนันต์, & Aemaimanan, P. (2562). ความสัมพันธ์ของดัชนีมวลกายกับจำนวนฟันปกติที่ใช้งานได้ของผู้ป่ วยโรคเบาหวานชนิดที่ 2 The Association Between Body Mass Index and the Number of Healthy Teeth of Type 2 Diabetes Mellitus. *ทันตแพทยสมาคมแห่งประเทศไทยในพระบรมราชูปถัมภ์*, *69*(4). https://doi.org/10.14456/jdat.2019.56

พญ.ศศิภัสช์ ช้อนทอง, & โรงพยาบาลเมดพาร์ค. (2564). *โรคเบาหวาน (Diabetes): สาเหตุ อาการและวิธีรักษา | MedPark Hospital*. https://www.medparkhospital.com/content/diabetes-mellitus

พิพัฒน์ ลิ้มประไพพงษ์, & สมชาย เล็กเจริญ. (2560). ศึกษาปัจจัยที่ส่งผลต่อคะแนนการประเมินคุณภาพภายในระดับคณะ ของมหาวิทยาลัยรังสิต โดยใช้เทคนิคเหมืองข้อมูล. *การประชุม*

วิชาการระดับชาติ มหาวิทยาลัยรังสิต ประจำปี ๒๕๖๐ (RSU National Research Conference 2017), 109–110.

เพชรรัตน์ ม่วงน้อย, จักรพันธ์ พลาผล, ภรัณยา ปาลวิสุทธิ์, & สารสนเทศ, ส. ช. (2564). ตัวแบบประเมินภาวะความเสี่ยงการเป็นโรคซึมเศร้าของนักศึกษาด้วยเทคนิคเหมืองข้อมูล. *วารสารวิชาการการประยุกต์ใช้เทคโนโลยีสารสนเทศ*, *7*(1).

ภูมิพัฒน์ ดวงกลาง, & รัจนา เครือแก้ว. (2562). แบบจำลองการทำนายแบบอากาศยานจากข้อมูลเป้าหมายไม่ทราบฝ่ายอัตโนมัติ. *วารสารวิทยาศาสตร์และเทคโนโลยีนายเรืออากาศ*, *15*, 3–4.

รัชฎาพร สุวรรณรัตน์, พวงรัตน์ บุญญานุรักษ์, ณัฐนิชา เหลืองอ่อน, & สุชาดา สมบูรณ์. (2562). ผลของการฝึกฝนทางคลินิกในการใช้ระบบการจำแนกข้อวินิจฉัย ผลลัพธ์ และการปฏิบัติการพยาบาลต่อคุณภาพการบันทึกทางการพยาบาล. *วารสารคณะพยาบาลศาสตร์ มหาวิทยาลัยบูรพา*, *27*(4), 53–65. https://he02.tci-thaijo.org/index.php/Nubuu/article/view/232594

รุ่งโรจน์ บุญมา, & นิเวศ จิระวิชิตชัย. (2562). การจำแนกประเภทผู้ป่วยโรคเบาหวานโดยใช้เทคนิคเหมืองข้อมูล และการเลือกคุณลักษณะจากความสัมพันธ์ของข้อมูล. *PKRU SciTech Journal*, *3*(2), 11–19. https://ph01.tci-thaijo.org/index.php/pkruscitech/article/view/201722

โรงพยาบาลเปาโล พหลโยธิน. (2564). *การควบคุมระดับน้ำตาลในเลือด | โรงพยาบาลเปาโล - Paolo Hospital*. https://www.paolohospital.com/th-TH/center/Article/Details/การควบคุมระดับน้ำตาลในเลือด

วนิดา พงษ์สงวน, ทิพยา ถิ่นสูงเนิน, & มาโนช ถิ่นสูงเนิน. (2561). การพัฒนาแบบจำลองสำหรับผู้ป่วยโรคเบาหวานโดยใช้เทคนิคต้นไม้ตัดสินใจ. *วารสารวิทยาศาสตณ์และเทคโนโลยี มหาวิทยาลัยราชภัฏมหาสารคาม*, *1*(1), 1–8. https://ph02.tci-thaijo.org/index.php/jstrmu/article/view/245737

วรรณภาภรณ์ จงกลาง, & นาฏนภา หีบแก้ว ปัดชาสุวรรณ. (2564). ปัจจัยที่มีความสัมพันธ์กับกลุ่มเสี่ยงโรคเบาหวานชนิดที่ 2 ของกลุ่มวัยทำงาน อำเภอโนนไทย จังหวัดนครราชสีมา THE RISK FACTOR OF TYPE 2 DIABETES IN WORKING POPULATIONS,NON-THAI DISTRICT NAKHON-RATCHASIMA PROVINCE. *วารสารวิจัยและพัฒนาระบบสุขภาพ*, *14*(3), 71–72.

วลัยลักษณ์ สุวรรณภักดี, มลิวัลย์ บุตรดำ, ทัศณีย์ หนูนารถ, & เบญจวรรณ ละหุการ. (2562). ภาวะเบาหวานขณะตั้งครรภ์. *ราชาวดีสาร วิทยาลัยพยาบาลบรมราชชนนี สุรินทร์*, *9*(2), 100–113. https://he01.tci-thaijo.org/index.php/bcnsurin/article/view/179414

วัจนา ขาวฟ้า, & อรศิริ ศิลาสัย. (2561). การประยุกต์ใช้เทคนิคการทำเหมืองข้อมูลในการตัดสินใจ. *คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยสวนดุสิต*. http://comsci.sci.dusit.ac.th/wp-content/uploads/2018/07/บทความวิชาการ_การประยุกต์ใช้เทคนิคการทำเหมืองข้อมูลใน

การตัดสินใจ.pdf

วิชญ์วิสิฐ เกสรสิทธิ์, ดร.วิชิต หล่อจีระชุณห์กลุ, & ดร.จิราวัลย์ จิตรถเวช. (2561). การแก้ปัญหาข้อมูลไม่สมดุลของข้อมูลสำหรับการจำแนกผู้ป่วยโรคเบาหวาน. *วารสารวิจัย มข. (ฉบับบัณฑิตศึกษา)*, *18*(3), 16–17.

วิบล ญึก, & จารี ทองคำ. (2561). การเปรียบเทียบเทคนิคอนุกรมเวลาเพื่อพยากรณ์ราคาทองและราคาน้ำมัน. *วารสาร มทร.อีสาน ฉบับวิทยาศาสตร์และเทคโนโลยี*, *11*(2), 157–158.

สมาพันธ์เบาหวานนานาชาติ. (2560). *สมาพันธ์เบาหวานนานาชาติ ประเมินทั่วโลกมีผู้ป่วยเบาหวาน 425 ล้านราย | Hfocus.org เจาะลึกระบบสุขภาพ*. https://www.hfocus.org/content/2019/11/18054

สถิติสุขภาพคนไทย. (2563). *ความชุกของโรคเบาหวาน*. https://www.hiso.or.th/thaihealthstat/area/index.php?ma=1&pf=01818101&tm=2&tp=3302

สมใจ จางวาง, เทพกร พิทยภินัน, & ชูติพัฒนะ, น. (2559). ปัจจัยที่มีความสัมพันธ์กับพฤติกรรมการดูแลตนเองเพื่อป้องกันโรคเบาหวานและความดันโลหิตสูงของประชาชนกลุ่มเสี่ยง. In *วารสารเครือข่ายวิทยาลัยพยาบาลและการสาธารณสุขภาคใต้* (Vol. 3, Issue 1, pp. 110–128).

สมลักษณ์ เทพสุริยานนท์, อมรรัตน์ นธะสนธิ์, สายันต์ จันทรดี, จารุณี บุญมั่น, บุญแต่ง โอชาพันธ์, มนชนก โลมรัตน์, & ประพิศพักตร์ พูลศรี. (2564). ปัจจัยทำนายความเสี่ยงต่อภาวะฉุกเฉินจากโรคเบาหวานของผู้สูงอายุโรคเบาหวานในชุมชน. *วารสารพยาบาลและสุขภาพ*, *12*(4). https://doi.org/10.1371/JOURNAL

สมาคมความดันโลหิตสูงแห่งประเทศไทย. (2562). *แแนวทางการรักษาโรคความดันโลหิตสูง ในเวชปฏิบัติทั่วไป พ.ศ. 2562*.

สมาคมโรคเบาหวานแห่งประเทศไทย. (2560). *แนวทางการพัฒนา เพื่อมาตรฐานคลินิกเบาหวาน*. https://www.dmthai.org/new/index.php/sara-khwam-ru/bukhlakr-thangkar-phaethy/cpg/diabetes-care-2024

สายชล สินสมบูรณ์ทอง. (2563). การเปรียบเทียบประสิทธิภาพในการทำนายผลความไม่สมดุลของข้อมูลในการจำแนกด้วยเทคนิคการทำเหมืองข้อมูล. *วารสารวิทยาศาสตร์และเทคโนโลยี*, *28*(3), 387–388. https://doi.org/10.14456/tstj.2020.31

สายฝน ม่วงคุ้ม, สุวรรณี มหากายนันท์, & เขมารดี มาสิงบุญ. (2560). ปัจจัยที่มีความสัมพันธ์กับพฤติกรรมป้องกันการเกิดโรคเบาหวาน ในกลุ่มเสี่ยงต่อการเป็นโรคเบาหวาน. *วารสารพยาบาลกระทรวงสาธารณสุข*, *27*(2), 214–227. https://he02.tci-thaijo.org/index.php/tnaph/article/view/97735

สุภัสสรา สมเจตนา, & จารี ทองคำ. (2564). การเปรียบเทียบประสิทธิภาพเทคนิคเหมืองข้อมูลในการสร้างแบบจำลองเพื่อจำแนกความคิดเห็นของผู้ปกครองต่อการใช้สมาร์ทโฟนของบุตร. *วารสารวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยอุบลราชธานี* , *23*(1), 26–27.

สุภาภรณ์ ตันตินันทตระกูล. (2555). เบาหวาน....ถ้ารู้ทัน...ก็ไร้เสี่ยง. *วารสารวิชาการมหาวิทยาลัยการกีฬาแห่งชาติ*, *4*(2), 125–136. https://he02.tci-thaijo.org/index.php/TNSUJournal/article/view/254484

สุรวัชร ศรีเปารยะ, & สายชล สินสมบูรณ์ทอง. (2560). การเปรียบเทียบประสิทธิภาพวิธีการจำแนกกลุ่มการเป็นโรคไตเรื้อรัง : กรณีศึกษาโรงพยาบาลแห่งหนึ่งในประเทศอินเดีย. *วารสารวิทยาศาสตร์และเทคโนโลยี*, *25*(5), 845–846. https://li01.tci-thaijo.org/index.php/tstj/article/view/85101

อนัตต์ชัย ชุติภาสเจริญ, & ดร.จรัญ แสนราช. (2561). การเปรียบเทียบประสิทธิภาพของอัลกอริทึมและการคัดเลือกคุณลักษณะที่เหมาะสมเพื่อการพยากรณ์โอกาศความสำเร็จในการโอนเงินข้ามประเทศของบุคคลทั่ว. *วารสารวิจัย มหาวิทยาลัยขอนแก่น (ฉบับบัณฑิตศึกษา) สาขามนุษยศาสตร์และสังคมศาสตร์*, *6*(3), 105–106.

อนุวัฒน์ เปพาทย์, วงกต ศรีอุไร, & ณัฏฐ์ ดิษเจริญ. (2565). การพยากรณ์การออกกลางคันของนักศึกษามหาวิทยาลัยจากการปรับปรุงด้วยการคัดเลือกคุณลักษณะร่วมกับวิธีโครงข่ายประสาทเทียมเพอร์เซ็ปตรอนหลายชั้น. *วารสารวิทยาศาสตร์และวิทยาศาสตร์ศึกษา (JSSE)*, *5*(1), 42–43. https://doi.org/10.14456/JSSE.2022.4

อัครพล พลูสวัสดิ์, & จรัญ แสนราช. (2562). การศึกษาเทคนิคพยากรณ์การได้รับปัจจัยพื้นฐานนักเรียนยากจนของนักเรียน โรงเรียนวัดพระขาว (ประชานุเคราะห์) ด้วยเทคนิคเหมืองข้อมูล. *วารสารวิทยาศาสตร์ แห่งมหาวิทยาลัยราชภัฏเพชรบุรี* , *16*(2), 6–7.

อัชฌาพร กว้างสวาสดิ์, เพียงฤทัย หนูสวัสดิ์, วราลี คงเหมาะ, ปวีณา ทิพยากุลรักษ์, & บุษกร สังขนันท์. (2562). ระบบการทำนายระดับความเครียดโดยใช้แผนผังการตัดสินใจ. *Rattanakosin Journal of Science and Technology*, *1*(2), 13–26. https://ph02.tci-thaijo.org/index.php/RJST/article/view/239865

อารยา หาอุปละ, & น้ำเพชร สายบัวทอง. (2561). การควบคุมระดับน้ำตาลในเลือดในผู้ป่วยเบาหวานกับผู้สูบบุหรี่และผู้ที่เคยสูบบุหรี่ | วารสารโรงพยาบาลมหาสารคาม. *วารสารโรงพยาบาลมหาสารคาม*. https://he02.tci-thaijo.org/index.php/MKHJ/article/view/195483

อุมากร ใจยั่งยืน, สุภาภรณ์ วรอรุณ, & สาวิตรี ศิริผลวุฒิชัย. (2565). ปัจจัยที่มีผลต่อพฤติกรรมการป้องกันโรคเบาหวานชนิดไม่พึ่งอินซูลินในกลุ่มเสี่ยง. *วารสารกองการพยาบาล*, *49*(1), 1–13. https://he01.tci-thaijo.org/index.php/JND/article/view/256659

เอชโฟกัส เจาะลึกระบบสุขภาพ สำนักข่าว Hfocus เจาะลึกระบบสุขภาพ, & มูลนิธิภิวัฒน์สาธารณสุขไทย เลขที่ 7 ถ.อธิบดี ต.หมากแข้ง อ.เมือง จ.อุดรธานี 41000. (2561). *ประเทศไทยพบเบาหวานในผู้หญิงมากกว่าผู้ชาย | Hfocus.org เจาะลึกระบบสุขภาพ*. https://www.hfocus.org/content/2018/02/15478

Abbas, H., Alic, L., Rios, M., Abdul-Ghani, M., & Qaraqe, K. (2019). Predicting diabetes in healthy population through machine learning. *Proceedings - IEEE Symposium on Computer-Based Medical Systems*, *2019-June*, 567–570. https://doi.org/10.1109/CBMS.2019.00117

Amazon Web Services (AWS). (2565). *การทำเหมืองข้อมูลคืออะไร คำอธิบายการทำเหมืองข้อมูล - AWS*. https://aws.amazon.com/th/what-is/data-mining/

Friedman, N., Linial, M., Nachman, I., & Pe'er, D. (2000). Using Bayesian networks to analyze expression data. *Journal of Computational Biology*, *7*(3–4), 601–620. https://doi.org/10.1089/106652700750050961

Vigneswari, D., Kumar, N. K., Ganesh Raj, V., Gugan, A., & Vikash, S. R. (2019). Machine Learning Tree Classifiers in Predicting Diabetes Mellitus. *2019 5th International Conference on Advanced Computing and Communication Systems, ICACCS 2019*, 84–87. https://doi.org/10.1109/ICACCS.2019.8728388

ประวัติผู้เขียน

ชื่อ	นายกฤตกนก ศรีพิมพ์สอ
วันเกิด	22 สิงหาคม 2537
สถานที่เกิด	มหาสารคาม
สถานที่อยู่ปัจจุบัน	4/15 หมู่ 14 ตำบลเกิ้ง อำเภอเมือง จังหวัดมหาสารคาม
ตำแหน่งหน้าที่การงาน	เจ้าหน้าที่บริหารงานทั่วไป
สถานที่ทำงานปัจจุบัน	คณะการบัญชีและการจัดการ มหาวิทยาลัยมหาสารคาม ตำบลขามเรียง อำเภอกันทรวิชัย จังหวัดมหาสารคาม
ประวัติการศึกษา	พ.ศ. 2562 วิศวกรรมศาสตรบัณฑิต วศ.บ. สาขาวิชาวิศวกรรมเมคาทรอนิกส์ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยมหาสารคาม พ.ศ. 2566 บริหารธุรกิจมหาบัณฑิต บธ.ม. สาขาวิชาบริหาธุรกิจและนวัตกรรมดิจิทัล คณะการบัญชีและการจัดการ มหาวิทยาลัยมหาสารคาม