



การเปรียบเทียบตัวแบบการถดถอยลอจิสติกและเทคนิคเหมืองข้อมูลสำหรับพยากรณ์การ
เป็นโรคเบาหวาน



รัชনীวรรณ ไพศาลรเกียรติ

วิทยานิพนธ์เสนอบัณฑิตวิทยาลัย มหาวิทยาลัยนเรศวร
เพื่อเป็นส่วนหนึ่งของการศึกษา หลักสูตรวิทยาศาสตรมหาบัณฑิต
สาขาวิชาสถิติ
ปีการศึกษา 2564
ลิขสิทธิ์เป็นของมหาวิทยาลัยนเรศวร

การเปรียบเทียบตัวแบบการถอดยลจิสติกและเทคนิคเหมืองข้อมูลสำหรับพยากรณ์การ
เป็นโรคเบาหวาน



วิทยานิพนธ์เสนอบัณฑิตวิทยาลัย มหาวิทยาลัยนครสวรรค์
เพื่อเป็นส่วนหนึ่งของการศึกษา หลักสูตรวิทยาศาสตรมหาบัณฑิต
สาขาวิชาสถิติ
ปีการศึกษา 2564
ลิขสิทธิ์เป็นของมหาวิทยาลัยนครสวรรค์

วิทยานิพนธ์ เรื่อง "การเปรียบเทียบตัวแบบการถดถอยลอจิสติกและเทคนิคเหมืองข้อมูลสำหรับ
พยากรณ์การเป็นโรคเบาหวาน"
ของ รัชนีวรรณ ไพศาลวรเกียรติ
ได้รับการพิจารณาให้นับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาสถิติ

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการสอบวิทยานิพนธ์
(รองศาสตราจารย์ ดร.อัชฌา อระวีพร)

..... ประธานที่ปรึกษาวิทยานิพนธ์
(รองศาสตราจารย์ ดร.อนามัย นาอุดม)

..... กรรมการที่ปรึกษาวิทยานิพนธ์
(ผู้ช่วยศาสตราจารย์ ดร.จรัสศรี รุ่งรัตนอุบล)

..... กรรมการผู้ทรงคุณวุฒิภายใน
(ผู้ช่วยศาสตราจารย์ ดร.กัลยา บุญหล้า)

อนุมัติ

.....
(รองศาสตราจารย์ ดร.กรองกาญจน์ ชูทิพย์)

คณบดีบัณฑิตวิทยาลัย

ชื่อเรื่อง	การเปรียบเทียบตัวแบบการถอดยลจิสติกและเทคนิคเหมืองข้อมูล สำหรับพยากรณ์การเป็นโรคเบาหวาน
ผู้วิจัย	รัชนิวรรณ ไพศาลวรเกียรติ
ประธานที่ปรึกษา	รองศาสตราจารย์ ดร.อนามัย นาอุดม
กรรมการที่ปรึกษา	ผู้ช่วยศาสตราจารย์ ดร.จรัสศรี รุ่งรัตนอุบล
ประเภทสารนิพนธ์	วิทยานิพนธ์ วท.ม. สาขาวิชาสถิติ, มหาวิทยาลัยนเรศวร, 2564
คำสำคัญ	โรคเบาหวาน การถอดยลจิสติกทวิภาค ต้นไม้ตัดสินใจ โครงข่าย ประสาทเทียม ค่าความถูกต้อง ค่าความถูกต้องในการจำแนกแบบ สมดุล

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อวิเคราะห์ปัจจัยที่ส่งผลต่อการเป็นโรคเบาหวานและเปรียบเทียบตัวแบบพยากรณ์การเป็นโรคเบาหวานด้วยวิธีการถอดยลจิสติกทวิภาค ต้นไม้ตัดสินใจ ด้วยอัลกอริทึม J48 และ LMT และเทคนิคโครงข่ายประสาทเทียม โดยใช้ข้อมูลผู้ป่วยที่เข้ารับบริการในโรงพยาบาลมหาวิทยาลัยนเรศวร จำนวน 5,081 ชุด โดยแบ่งข้อมูลเป็นข้อมูลเรียนรู้ และข้อมูลทดสอบด้วยสัดส่วน 70:30 และ 80:20 ทำการเปรียบเทียบประสิทธิภาพตัวแบบพยากรณ์บนข้อมูลทดสอบด้วยค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล จากการศึกษาพบว่า ปัจจัยที่ส่งผลต่อการเป็นโรคเบาหวาน คือ ค่าความดันขณะหัวใจบีบตัว ค่าความดันขณะหัวใจคลายตัว อัตราการเต้นของหัวใจ น้ำหนัก ความสูง และระดับน้ำตาลในเลือด นอกจากนี้จากผลการศึกษาพบว่า เทคนิคโครงข่ายประสาทเทียมมีประสิทธิภาพในการพยากรณ์ดีที่สุดในทุกชุดข้อมูลทั้ง 2 แบบ โดยในชุดข้อมูลแบบที่ 1 (70:30) ให้ค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล เท่ากับ 81.7824% และ 73.9704% ตามลำดับ และในชุดข้อมูลแบบที่ 2 (80:20) ให้ค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล เท่ากับ 81.4159% และ 73.7482% ตามลำดับ โดยเทคนิคที่มีประสิทธิภาพรองลงมา คือ เทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม J48 และ LMT และเทคนิคการถอดยลจิสติกทวิภาค ตามลำดับ

จากการศึกษาพบว่าแม้เทคนิคการถอดยลจิสติกทวิภาคไม่ได้ให้ตัวแบบพยากรณ์ที่มีประสิทธิภาพเมื่อเปรียบเทียบกับตัวแบบพยากรณ์ทางเทคนิคเหมืองข้อมูล แต่การประยุกต์ใช้การ

คัดเลือกตัวแปรอิสระตามเทคนิคการถดถอยลอจิสติกทวิภาคสามารถนำมาปรับปรุงและเพิ่มประสิทธิภาพให้กับตัวแบบพยากรณ์ด้วยเทคนิคเหมืองข้อมูลได้



Title	A COMPARISON OF LOGISTIC REGRESSION MODEL AND DATA MINING TECHNIQUES FOR PREDICTING DIABETES
Author	RATCHANEewan PAISANWARAKIAT
Advisor	Associate Professor Anamai Na-udom, Ph.D.
Co-Advisor	Assistant Professor Jaratsri Rungrattanaubol, Ph.D.
Academic Paper	M.S. Thesis in Statistics - (Type A 2), Naresuan University, 2021
Keywords	Diabetes Binary Logistic Regression Decision Tree Artificial Neural Network Accuracy Balance Classification Accuracy

ABSTRACT

This research aims to analyze factors affecting diabetes and to compare the predictive model for predicting diabetes using a binary logistic regression model, a decision tree with algorithms J48, LMT and an artificial neural network. The dataset obtained from the patients who visited to Naresuan University Hospital, consists of 5,081 cases. The dataset is divided into the training and testing dataset with two ratios, 70:30 and 80:20, respectively. The performance of a predictive model is measured on the testing dataset with the accuracy and balance classification accuracy. The results showed that the factors affecting diabetes are systolic blood pressure, diastolic blood pressure, heart rate, body weight, height and fasting blood sugar. Moreover, the predictive model from artificial neural network performs best in both datasets with the accuracy and balanced classification accuracy 81.7824% and 73.9704% on the dataset (70:30) and 81.4159% and 73.7482% on the second dataset (80:20). The second-best performance and later are the decision tree with algorithm J48 and LMT and the binary logistic regression, respectively.

The results of this study indicate that, although the binary logistic regression model did not provide the most effective predictive model compared with those models from the data mining technique. However, the application of independent

variable selection based on the binary logistic regression can be used to improve and increase the accuracy of the predictive model with data mining techniques.



ประกาศคุณูปการ

ผู้วิจัยขอขอบพระคุณประธานที่ปรึกษาวิทยานิพนธ์ รองศาสตราจารย์ ดร.อนามัย นาอุดม เป็นอย่างสูงที่สละเวลาให้คำปรึกษาและให้คำแนะนำตลอดระยะเวลาการทำวิทยานิพนธ์ ขอขอบคุณ ผู้ช่วยศาสตราจารย์ ดร.จรัสศรี รุ่งรัตน์อุบล กรรมการที่ปรึกษาวิทยานิพนธ์ ที่ให้คำปรึกษาเกี่ยวกับการจัดการข้อมูลด้วยวิธีการทำเหมืองข้อมูล ขอขอบคุณ รองศาสตราจารย์ ดร.อัชฌา อระวีพร ประธาน กรรมการสอบวิทยานิพนธ์ และกรรมการผู้ทรงคุณวุฒิภายนอก ที่ให้คำแนะนำแก้ไขส่วนที่บกพร่องของ งานวิทยานิพนธ์เล่มนี้

ขอขอบคุณ ผู้ช่วยศาสตราจารย์ พญ.ศรินยา สัทธานนท์ อาจารย์ประจำภาควิชาอายุรศาสตร์ คณะแพทยศาสตร์ มหาวิทยาลัยนเรศวร ที่ให้คำแนะนำเกี่ยวกับข้อมูลทางด้านการแพทย์แก่ผู้วิจัย

ขอขอบคุณ คุณอรนิภา ชันสาคร นักวิชาการคอมพิวเตอร์ หน่วยงานเทคโนโลยีสารสนเทศ โรงพยาบาลมหาวิทยาลัยนเรศวร ที่ให้ความอนุเคราะห์ด้านข้อมูลผู้เข้ามาใช้บริการ รวมถึงเจ้าหน้าที่ทุกท่านที่ให้ความช่วยเหลือ

ผู้วิจัยขอกราบขอบพระคุณคุณพ่อชยันต์ ไพศาลวรเกียรติ คุณแม่แก้ว ไพศาลวรเกียรติ คุณแม่ กัญจนา แซ่โจ้ว ที่เป็นกำลังใจและแรงผลักดัน และขอบคุณเพื่อนร่วมรุ่นปริญญาโท และรุ่นพี่ปริญญาเอกทุกท่านที่ให้การสนับสนุนงานวิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยดี

ผู้วิจัยหวังเป็นอย่างยิ่งว่า งานวิจัยเรื่องการเปรียบเทียบตัวแบบการถดถอยลอจิสติกและเทคนิคเหมืองข้อมูลสำหรับพยากรณ์การเป็นโรคเบาหวาน จะเป็นประโยชน์ต่อประชาชน เจ้าหน้าที่ทางการแพทย์และบุคคลผู้สนใจที่จะนำงานวิจัยนี้ไปต่อยอดเพื่อพัฒนาองค์ความรู้ใหม่ ๆ ต่อไป

รัชนิวรรณ ไพศาลวรเกียรติ

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ค
บทคัดย่อภาษาอังกฤษ.....	จ
ประกาศคุุณูปการ.....	ช
สารบัญ.....	ซ
สารบัญตาราง.....	ณ
สารบัญภาพ.....	พ
บทที่ 1 บทนำ.....	1
1.1 ที่มาและความสำคัญ.....	1
1.2 จุดมุ่งหมายของการวิจัย.....	4
1.3 ขอบเขตการวิจัย.....	4
1.4 ประโยชน์ที่คาดว่าจะได้รับ.....	5
บทที่ 2 เอกสารและงานวิจัยที่เกี่ยวข้อง.....	7
2.1 การถดถอยลอจิสติก (Logistic Regression).....	7
2.1.1 วัตถุประสงค์ของการถดถอยลอจิสติกทวิภาค.....	7
2.1.2 ข้อตกลงเบื้องต้นของการถดถอยลอจิสติกทวิภาค.....	7
2.1.3 ตัวแบบการถดถอยลอจิสติกทวิภาค.....	8
2.1.4 วิธีการคัดเลือกตัวแปร (Variable Selection).....	9
2.1.5 การตรวจสอบความสัมพันธ์ระหว่างตัวแปรอิสระ.....	10
2.1.6 การตรวจสอบความเหมาะสมของตัวแบบการถดถอยลอจิสติกทวิภาค.....	10

1) การทดสอบอัตราส่วนภาวะน่าจะเป็น (Likelihood Ratio Test).....	11
2) เกณฑ์สารสนเทศของอะไคเกะ (Akaike's Information Criterion (AIC))	11
การทดสอบความมีนัยสำคัญของสัมประสิทธิ์ถดถอย (Regression Coefficient)	12
2.1.7 สถิติทดสอบความน่าเชื่อถือของตัวแบบการถดถอยลอจิสติกทวิภาค	12
1) สถิติทดสอบค็อกซ์และสเนล (Cox & Snell R^2).....	12
2) สถิติทดสอบนาเจลเคอร์คี (Nagelkerke R^2).....	13
2.2 ทฤษฎีการทำเหมืองข้อมูล (Data Mining).....	13
2.2.1 กระบวนการของการทำเหมืองข้อมูล.....	13
2.3 เทคนิคต้นไม้ตัดสินใจ	16
2.3.1 ขั้นตอนการสร้างต้นไม้ตัดสินใจ	17
2.3.2 อัลกอริทึม ID3 (Iterative Dichotomiser 3).....	17
2.3.3 อัลกอริทึม J48 หรือ C4.5.....	18
2.3.4 อัลกอริทึม LMT.....	18
2.4 เทคนิคโครงข่ายประสาทเทียม (Artificial Neural Network).....	23
2.4.1 ตัวอย่างการประยุกต์ตัวแบบโครงข่ายประสาทเทียมเพื่อการพยากรณ์.....	26
2.4.2 การหาน้ำหนักถ่วงด้วยการทำงานแบบส่งค่าย้อนกลับ.....	28
2.4.3 ตัวอย่างการหาน้ำหนักถ่วง	30
2.5 การเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์	35
2.5.1 ตัวอย่างการแสดงค่าการวัดประสิทธิภาพต่าง ๆ จากข้อมูลการพยากรณ์การ เป็นโรคเบาหวาน	36
2.6 งานวิจัยที่เกี่ยวข้อง	37
บทที่ 3 วิธีดำเนินการวิจัย	44

3.1 ข้อมูลประชากร และกลุ่มตัวอย่างในงานวิจัย	44
3.2 วิธีการดำเนินงานวิจัย.....	45
3.3 การสร้างตัวแบบพยากรณ์.....	50
3.3.1 การสร้างตัวแบบพยากรณ์ด้วยเทคนิคการถดถอยลอจิสติกทวิภาค	51
3.3.2 การสร้างตัวแบบพยากรณ์ด้วยเทคนิคต้นไม้ตัดสินใจ	51
1) อัลกอริทึม J48	52
2) อัลกอริทึม LMT (Logistic Model Tree).....	52
3.3.3 การสร้างตัวแบบพยากรณ์ด้วยเทคนิคโครงข่ายประสาทเทียม	52
3.4 เกณฑ์การเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์	53
3.5 เครื่องมือที่ใช้ในการวิจัย	53
บทที่ 4 ผลการวิจัย	55
.....	55
4.1 ผลการเตรียมข้อมูลเพื่อสร้างตัวแบบพยากรณ์	55
4.2 ผลการสร้างและพัฒนาตัวแบบพยากรณ์ชุดข้อมูลแบบที่ 1 (70 : 30)	56
4.2.1 เทคนิคการถดถอยลอจิสติกทวิภาค	56
ผลการสร้างตัวแบบพยากรณ์วิธีการเลือกแบบพื้นฐาน	56
ผลการสร้างตัวแบบพยากรณ์จากเทคนิคการถดถอยลอจิสติกทวิภาควิธีการ เลือกแบบพื้นฐาน	57
ผลการสร้างตัวแบบพยากรณ์วิธีการเลือกแบบไปข้างหน้า	58
ผลการสร้างตัวแบบพยากรณ์จากเทคนิคการถดถอยลอจิสติกทวิภาควิธีการ เลือกแบบไปข้างหน้า	60
ผลการสร้างตัวแบบพยากรณ์ด้วยวิธีการกำจัดแบบถอยหลัง	60

ผลการสร้างตัวแบบพยากรณ์จากเทคนิคการถดถอยลอจิสติกทวิภาควิธีการ กำจัดแบบถอยหลัง	62
ผลการสร้างตัวแบบพยากรณ์วิธีการถดถอยที่ละชั้น.....	63
ผลการสร้างตัวแบบพยากรณ์จากเทคนิคการถดถอยลอจิสติกทวิภาควิธีการ ถดถอยที่ละชั้น.....	65
ผลการเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์จากเทคนิคการถดถอยลอจ ิสติกทวิภาค ของชุดข้อมูลแบบที่ 1	67
4.2.2 เทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม J48.....	67
ผลการสร้างตัวแบบพยากรณ์จากเทคนิคต้นไม้ตัดสินใจ J48 กรณีตัวแปร อิสระทุกตัว.....	68
ผลการสร้างตัวแบบพยากรณ์จากเทคนิคต้นไม้ตัดสินใจ J48 วิธีการเลือกแบบ พื้นฐาน.....	69
ผลการสร้างตัวแบบพยากรณ์จากเทคนิคต้นไม้ตัดสินใจ J48 วิธีการเลือกแบบ ไปข้างหน้า.....	70
ผลการสร้างตัวแบบพยากรณ์จากเทคนิคต้นไม้ตัดสินใจ J48 วิธีการกำจัดแบบ ถอยหลัง.....	71
ผลการสร้างตัวแบบพยากรณ์จากเทคนิคต้นไม้ตัดสินใจ J48 วิธีการถดถอยที่ ละชั้น.....	72
ผลการเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์จากเทคนิคต้นไม้ตัดสินใจ ด้วยอัลกอริทึม J48 ของชุดข้อมูลแบบที่ 1	73
4.2.3 เทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม LMT.....	73
ผลการสร้างตัวแบบพยากรณ์จากเทคนิคต้นไม้ตัดสินใจ LMT กรณีตัวแปร อิสระทุกตัว.....	74

ผลการสร้างตัวแบบพยากรณ์จากเทคนิคต้นไม้ตัดสินใจ LMT วิธีการเลือก แบบพื้นฐาน	75
ผลการสร้างตัวแบบพยากรณ์จากเทคนิคต้นไม้ตัดสินใจ LMT วิธีการเลือก แบบไปข้างหน้า	76
ผลการสร้างตัวแบบพยากรณ์จากเทคนิคต้นไม้ตัดสินใจ LMT วิธีการกำจัด แบบถอยหลัง	77
ผลการสร้างตัวแบบพยากรณ์จากเทคนิคต้นไม้ตัดสินใจ LMT วิธีการถดถอยที่ ละชั้น.....	78
ผลการเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์จากเทคนิคต้นไม้ตัดสินใจ ด้วยอัลกอริทึม LMT ของชุดข้อมูลแบบที่ 1	79
4.2.4 เทคนิคโครงข่ายประสาทเทียม	79
ผลการสร้างตัวแบบพยากรณ์จากเทคนิคโครงข่ายประสาทเทียม กรณีตัวแปร อิสระทุกตัว.....	80
ผลการสร้างตัวแบบพยากรณ์จากเทคนิคโครงข่ายประสาทเทียมวิธีการเลือก แบบพื้นฐาน	81
ผลการสร้างตัวแบบพยากรณ์จากเทคนิคโครงข่ายประสาทเทียมวิธีการเลือก แบบไปข้างหน้า	82
ผลการสร้างตัวแบบพยากรณ์จากเทคนิคโครงข่ายประสาทเทียมวิธีการกำจัด แบบถอยหลัง	83
ผลการสร้างตัวแบบพยากรณ์จากเทคนิคโครงข่ายประสาทเทียมวิธีการ ถดถอยทีละชั้น.....	84
ผลการเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์จากเทคนิคโครงข่ายประสาท เทียม ของชุดข้อมูลแบบที่ 1	85

4.3 ผลการเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์ในชุดข้อมูลแบบที่ 1 (70:30)	86
4.4 ผลการสร้างและพัฒนาตัวแบบพยากรณ์ชุดข้อมูลแบบที่ 2 (80:20)	86
4.4.1 เทคนิคการถดถอยลอจิสติกทวิภาค	86
ผลการสร้างตัวแบบพยากรณ์วิธีการเลือกแบบพื้นฐาน	86
ผลการสร้างตัวแบบพยากรณ์จากเทคนิคการถดถอยลอจิสติกทวิภาควิธีการเลือกแบบพื้นฐาน	88
ผลการสร้างตัวแบบพยากรณ์วิธีการเลือกแบบไปข้างหน้า	89
ผลการสร้างตัวแบบพยากรณ์จากเทคนิคการถดถอยลอจิสติกทวิภาควิธีการเลือกแบบไปข้างหน้า	90
ผลการสร้างตัวแบบพยากรณ์ด้วยวิธีการกำจัดแบบถอยหลัง	91
ผลการสร้างตัวแบบพยากรณ์จากเทคนิคการถดถอยลอจิสติกทวิภาควิธีการกำจัดแบบถอยหลัง	93
ผลการสร้างตัวแบบพยากรณ์วิธีการถดถอยทีละชั้น	94
ผลการสร้างตัวแบบพยากรณ์จากเทคนิคการถดถอยลอจิสติกทวิภาควิธีการถดถอยทีละชั้น	95
ผลการเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์จากเทคนิคการถดถอยลอจิสติกทวิภาคของชุดข้อมูลแบบที่ 2	97
4.4.2 เทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม J48	98
ผลการสร้างตัวแบบพยากรณ์จากเทคนิคต้นไม้ตัดสินใจ J48 กรณีตัวแปรอิสระทุกตัว	98
ผลการสร้างตัวแบบพยากรณ์จากเทคนิคต้นไม้ตัดสินใจ J48 วิธีการเลือกแบบพื้นฐาน	99

ผลการสร้างตัวแบบพยากรณ์จากเทคนิคต้นไม้ตัดสินใจ J48 วิธีการเลือกแบบ ไปข้างหน้า.....	100
ผลการสร้างตัวแบบพยากรณ์จากเทคนิคต้นไม้ตัดสินใจ J48 วิธีการกำจัดแบบ ถอยหลัง.....	101
ผลการสร้างตัวแบบพยากรณ์จากเทคนิคต้นไม้ตัดสินใจ J48 วิธีการถดถอยที่ ละชั้น.....	102
ผลการเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์จากเทคนิคต้นไม้ตัดสินใจ ด้วยอัลกอริทึม J48 ของชุดข้อมูลแบบที่ 2.....	103
4.4.3 เทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม LMT.....	103
ผลการสร้างตัวแบบพยากรณ์จากเทคนิคต้นไม้ตัดสินใจ LMT กรณีตัวแปร อิสระทุกตัว.....	104
ผลการสร้างตัวแบบพยากรณ์จากเทคนิคต้นไม้ตัดสินใจ LMT วิธีการเลือก แบบพื้นฐาน.....	105
ผลการสร้างตัวแบบพยากรณ์จากเทคนิคต้นไม้ตัดสินใจ LMT วิธีการเลือก แบบไปข้างหน้า.....	106
ผลการสร้างตัวแบบพยากรณ์จากเทคนิคต้นไม้ตัดสินใจ LMT วิธีการกำจัด แบบถอยหลัง.....	107
ผลการสร้างตัวแบบพยากรณ์จากเทคนิคต้นไม้ตัดสินใจ LMT วิธีการถดถอยที่ ละชั้น.....	108
ผลการเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์จากเทคนิคต้นไม้ตัดสินใจ ด้วยอัลกอริทึม LMT ของชุดข้อมูลแบบที่ 2.....	109
4.4.4 เทคนิคโครงข่ายประสาทเทียม.....	109
ผลการสร้างตัวแบบพยากรณ์จากเทคนิคโครงข่ายประสาทเทียม กรณีตัวแปร อิสระทุกตัว.....	110

ผลการสร้างตัวแบบพยากรณ์จากเทคนิคโครงข่ายประสาทเทียมวิธีการเลือก แบบพื้นฐาน	111
ผลการสร้างตัวแบบพยากรณ์จากเทคนิคโครงข่ายประสาทเทียมวิธีการเลือก แบบไปข้างหน้า	112
ผลการสร้างตัวแบบพยากรณ์จากเทคนิคโครงข่ายประสาทเทียมวิธีการกำจัด แบบถอยหลัง	113
ผลการสร้างตัวแบบพยากรณ์จากเทคนิคโครงข่ายประสาทเทียมวิธีการ ถดถอยทีละชั้น.....	114
ผลการเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์จากเทคนิคโครงข่ายประสาท เทียม ของชุดข้อมูลแบบที่ 2	115
4.5 ผลการเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์ในชุดข้อมูลแบบที่ 2 (80:20)	116
4.6 ผลการเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์ทั้งสองชุดข้อมูล	116
บทที่ 5 สรุปผลการวิจัย	117
5.1 อภิปรายผล.....	120
5.2 ข้อเสนอแนะ.....	120
บรรณานุกรม	144
ประวัติผู้วิจัย	148

สารบัญตาราง

หน้า

ตาราง 1 แสดงอัตราความชุกของผู้ป่วยเบาหวานใน ปี พ.ศ. 2561 จำแนกตามกลุ่มอายุ..	2
ตาราง 2 แสดงอัตราความชุกของผู้ป่วยเบาหวาน ปี พ.ศ. 2561 จำแนกตามภูมิภาค	3
ตาราง 3 แสดงตัวอย่างข้อมูล.....	19
ตาราง 4 แสดงตัวอย่างข้อมูลผู้ป่วย โดยจำแนกผู้ป่วยออกเป็น 2 กลุ่ม	26
ตาราง 5 แสดงตัวอย่างข้อมูลผู้ป่วยที่ทำการแปลงข้อมูล.....	27
ตาราง 6 แสดงตัวอย่างข้อมูลผู้ป่วยที่แสดงผลลัพธ์จากตัวแบบพยากรณ์ด้วยเทคนิค โครงข่ายประสาทเทียม	28
ตาราง 7 แสดงผลลัพธ์ที่ได้จากการหาค่าฟังก์ชันค่าความผิดพลาดจากตัวแบบพยากรณ์ด้วยเทคนิค โครงข่ายประสาทเทียม	34
ตาราง 8 แสดงตัวอย่างตารางเมทริกซ์ความสับสน (Confusion Matrix).....	35
ตาราง 9 แสดงตัวอย่างตารางเมทริกซ์ความสับสนของการพยากรณ์การเป็นโรคเบาหวาน	36
ตาราง 10 แสดงผลสรุปเพื่อเปรียบเทียบงานวิจัยที่เกี่ยวข้องที่มีการประยุกต์ใช้ตัวแบบ พยากรณ์การเป็นโรคเบาหวาน	40
ตาราง 11 แสดงตัวแปรอิสระทั้งหมดของข้อมูลผู้ป่วยที่เข้ารับการรักษาในโรงพยาบาล มหาวิทยาลัยนเรศวร	45
ตาราง 12 แสดงรายละเอียดข้อมูลผู้ป่วยโรคเบาหวาน ตัวแปรอิสระ (X) ตัวแปรตาม (Y)	46
ตาราง 13 แสดงรายละเอียดข้อมูลผู้ป่วยจำนวน 5,081 ชุด โดยแยกเป็นกรณีผู้ป่วยที่ไม่ เป็นโรคเบาหวาน และผู้ป่วยที่เป็นโรคเบาหวาน.....	47
ตาราง 14 แสดงการแปลงข้อมูลให้อยู่การทำให้เป็นปรกติ	48

ตาราง 15 แสดงค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรอิสระและตัวแปรตาม	49
ตาราง 16 แสดงผลลัพธ์จากการสร้างตัวแบบพยากรณ์ด้วยเทคนิคการถดถอยลอจิสติก ทวิภาคีวิธีการเลือกแบบพื้นฐาน ของชุดข้อมูลแบบที่ 1	56
ตาราง 17 แสดงการตรวจสอบความสัมพันธ์ระหว่างตัวแปรอิสระวิธีการเลือกแบบพื้นฐาน ของชุดข้อมูลแบบที่ 1	57
ตาราง 18 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคการถดถอยลอจิสติก ทวิภาคีวิธีการเลือกแบบพื้นฐาน ของชุดข้อมูลเรียนรู้ 70 เปอร์เซ็นต์	57
ตาราง 19 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคการถดถอยลอจิสติก ทวิภาคีวิธีการเลือกแบบพื้นฐาน ของชุดข้อมูลทดสอบ 30 เปอร์เซ็นต์	58
ตาราง 20 แสดงผลการคัดเลือกตัวแปรอิสระวิธีการเลือกแบบไปข้างหน้า ของชุดข้อมูล แบบที่ 1	58
ตาราง 21 แสดงการตรวจสอบความสัมพันธ์ระหว่างตัวแปรอิสระวิธีการเลือกแบบไป ข้างหน้าของชุดข้อมูลแบบที่ 1	59
ตาราง 22 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคการถดถอยลอจิสติก ทวิภาคีวิธีการเลือกแบบไปข้างหน้า ของชุดข้อมูลเรียนรู้ 70 เปอร์เซ็นต์	60
ตาราง 23 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคการถดถอยลอจิสติก ทวิภาคีวิธีการเลือกแบบไปข้างหน้า ของชุดข้อมูลทดสอบ 30 เปอร์เซ็นต์	60
ตาราง 24 แสดงผลการคัดเลือกตัวแปรอิสระวิธีการกำจัดแบบถอยหลัง ของชุดข้อมูลแบบ ที่ 1	60
ตาราง 25 แสดงการตรวจสอบความสัมพันธ์ระหว่างตัวแปรอิสระวิธีการกำจัดแบบถอย หลัง ของชุดข้อมูลแบบที่ 1	62
ตาราง 26 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคการถดถอยลอจิสติก ทวิภาคีวิธีการกำจัดแบบถอยหลัง ของชุดข้อมูลเรียนรู้ 70 เปอร์เซ็นต์	62

ตาราง 27 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคการถดถอยลอจิสติกทวิภาควิธีการกำจัดแบบถอยหลัง ของชุดข้อมูลทดสอบ 30 เปอร์เซ็นต์.....	63
ตาราง 28 แสดงผลการคัดเลือกตัวแปรอิสระวิธีการถดถอยทีละชั้น ของชุดข้อมูลแบบที่ 1	63
ตาราง 29 แสดงการตรวจสอบความสัมพันธ์ระหว่างตัวแปรอิสระวิธีการถดถอยทีละชั้น ของชุดข้อมูลแบบที่ 1.....	64
ตาราง 30 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคการถดถอยลอจิสติกทวิภาควิธีการถดถอยทีละชั้น ของชุดข้อมูลเรียนรู้ 70 เปอร์เซ็นต์	65
ตาราง 31 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคการถดถอยลอจิสติกทวิภาควิธีการถดถอยทีละชั้น ของชุดข้อมูลทดสอบ 30 เปอร์เซ็นต์.....	65
ตาราง 32 แสดงผลการเปรียบเทียบการตรวจสอบความเหมาะสมของตัวแบบพยากรณ์ ด้วยเทคนิคการถดถอยลอจิสติกทวิภาค ของชุดข้อมูลแบบที่ 1.....	66
ตาราง 33 แสดงผลระดับความน่าเชื่อถือของตัวแบบพยากรณ์ด้วยเทคนิคการถดถอยลอจิสติกทวิภาค ของชุดข้อมูลแบบที่ 1	66
ตาราง 34 แสดงผลเปรียบเทียบประสิทธิภาพการพยากรณ์จากเทคนิคการถดถอยลอจิสติกทวิภาค ของชุดข้อมูลแบบที่ 1	67
ตาราง 35 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วย อัลกอริทึม J48 กรณีตัวแปรอิสระทุกตัว ของชุดข้อมูลเรียนรู้ 70 เปอร์เซ็นต์	68
ตาราง 36 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วย อัลกอริทึม J48 กรณีตัวแปรอิสระทุกตัว ของชุดข้อมูลทดสอบ 30 เปอร์เซ็นต์.....	68
ตาราง 37 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วย อัลกอริทึม J48 วิธีการเลือกแบบพื้นฐาน ของชุดข้อมูลเรียนรู้ 70 เปอร์เซ็นต์	69
ตาราง 38 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วย อัลกอริทึม J48 วิธีการเลือกแบบพื้นฐาน ของชุดข้อมูลทดสอบ 30 เปอร์เซ็นต์.....	69

ตาราง 39 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วย อัลกอริทึม J48 วิธีการเลือกแบบไปข้างหน้า ของชุดข้อมูลเรียนรู้ 70 เปอร์เซนต์	70
ตาราง 40 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วย อัลกอริทึม J48 วิธีการเลือกแบบไปข้างหน้า ของชุดข้อมูลทดสอบ 30 เปอร์เซนต์	70
ตาราง 41 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วย อัลกอริทึม J48 วิธีการกำจัดแบบถอยหลัง ของชุดข้อมูลเรียนรู้ 70 เปอร์เซนต์	71
ตาราง 42 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วย อัลกอริทึม J48 วิธีการกำจัดแบบถอยหลัง ของชุดข้อมูลทดสอบ 30 เปอร์เซนต์	71
ตาราง 43 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วย อัลกอริทึม J48 วิธีการถดถอยทีละชั้น ของชุดข้อมูลเรียนรู้ 70 เปอร์เซนต์	72
ตาราง 44 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วย อัลกอริทึม J48 วิธีการถดถอยทีละชั้น ของชุดข้อมูลทดสอบ 30 เปอร์เซนต์	72
ตาราง 45 แสดงผลการเปรียบเทียบประสิทธิภาพการพยากรณ์ ด้วยเทคนิคต้นไม้ตัดสินใจ ด้วยอัลกอริทึม J48 ของชุดข้อมูลแบบที่ 1	73
ตาราง 46 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วย อัลกอริทึม LMT กรณีตัวแปรอิสระทุกตัว ของชุดข้อมูลเรียนรู้ 70 เปอร์เซนต์	74
ตาราง 47 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วย อัลกอริทึม LMT กรณีตัวแปรอิสระทุกตัว ของชุดข้อมูลทดสอบ 30 เปอร์เซนต์	74
ตาราง 48 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วย อัลกอริทึม LMT วิธีการเลือกแบบพื้นฐาน ของชุดข้อมูลเรียนรู้ 70 เปอร์เซนต์	75
ตาราง 49 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วย อัลกอริทึม LMT วิธีการเลือกแบบพื้นฐาน ของชุดข้อมูลทดสอบ 30 เปอร์เซนต์	75
ตาราง 50 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วย อัลกอริทึม LMT วิธีการเลือกแบบไปข้างหน้า ของชุดข้อมูลเรียนรู้ 70 เปอร์เซนต์	76

ตาราง 51 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม LMT วิธีการเลือกแบบไปข้างหน้า ของชุดข้อมูลทดสอบ 30 เปอร์เซนต์.....	76
ตาราง 52 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม LMT วิธีการกำจัดแบบถอยหลัง ของชุดข้อมูลเรียนรู้ 70 เปอร์เซนต์.....	77
ตาราง 53 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม LMT วิธีการกำจัดแบบถอยหลัง ของชุดข้อมูลทดสอบ 30 เปอร์เซนต์.....	77
ตาราง 54 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม LMT วิธีการถดถอยทีละชั้น ของชุดข้อมูลเรียนรู้ 70 เปอร์เซนต์.....	78
ตาราง 55 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม LMT วิธีการถดถอยทีละชั้น ของชุดข้อมูลทดสอบ 30 เปอร์เซนต์.....	78
ตาราง 56 แสดงผลการเปรียบเทียบประสิทธิภาพการพยากรณ์ ด้วยเทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม LMT ของชุดข้อมูลแบบที่ 1.....	79
ตาราง 57 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคโครงข่ายประสาทเทียม กรณีตัวแปรอิสระทุกตัว ของชุดข้อมูลเรียนรู้ 70 เปอร์เซนต์.....	80
ตาราง 58 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคโครงข่ายประสาทเทียม กรณีตัวแปรอิสระทุกตัว ของชุดข้อมูลทดสอบ 30 เปอร์เซนต์.....	80
ตาราง 59 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคโครงข่ายประสาทเทียม วิธีการเลือกแบบพื้นฐาน ของชุดข้อมูลเรียนรู้ 70 เปอร์เซนต์.....	81
ตาราง 60 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคโครงข่ายประสาทเทียม วิธีการเลือกแบบพื้นฐาน ของชุดข้อมูลทดสอบ 30 เปอร์เซนต์.....	81
ตาราง 61 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคโครงข่ายประสาทเทียม วิธีการเลือกแบบไปข้างหน้า ของชุดข้อมูลเรียนรู้ 70 เปอร์เซนต์.....	82
ตาราง 62 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคโครงข่ายประสาทเทียม วิธีการเลือกแบบไปข้างหน้า ของชุดข้อมูลทดสอบ 30 เปอร์เซนต์.....	82

ตาราง 63 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคโครงข่ายประสาทเทียม วิธีการกำจัดแบบถอยหลัง ของชุดข้อมูลเรียนรู้ 70 เปอร์เซ็นต์.....	83
ตาราง 64 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคโครงข่ายประสาทเทียม วิธีการกำจัดแบบถอยหลัง ของชุดข้อมูลทดสอบ 30 เปอร์เซ็นต์.....	83
ตาราง 65 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคโครงข่ายประสาทเทียม วิธีการถดถอยทีละชั้น ของชุดข้อมูลเรียนรู้ 70 เปอร์เซ็นต์.....	84
ตาราง 66 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคโครงข่ายประสาทเทียม วิธีการถดถอยทีละชั้น ของชุดข้อมูลทดสอบ 30 เปอร์เซ็นต์.....	84
ตาราง 67 แสดงผลการเปรียบเทียบประสิทธิภาพการพยากรณ์ ด้วยเทคนิคโครงข่ายประสาทเทียม ของชุดข้อมูลแบบที่ 1.....	85
ตาราง 68 แสดงผลการเปรียบเทียบประสิทธิภาพตัวแบบพยากรณ์ด้วยเทคนิคต่าง ๆ ของชุดข้อมูลแบบที่ 1.....	86
ตาราง 69 แสดงผลลัพธ์จากการสร้างตัวแบบพยากรณ์ด้วยเทคนิคการถดถอยลอจิสติก ทวิภาควิธีการเลือกแบบพื้นฐาน ของชุดข้อมูลแบบที่ 2.....	86
ตาราง 70 แสดงการตรวจสอบความสัมพันธ์ระหว่างตัวแปรอิสระวิธีการเลือกแบบพื้นฐาน.....	87
ตาราง 71 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคการถดถอยลอจิสติกทวิภาควิธีการเลือกแบบพื้นฐาน ของชุดข้อมูลเรียนรู้ 80 เปอร์เซ็นต์.....	88
ตาราง 72 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคการถดถอยลอจิสติกทวิภาควิธีการเลือกแบบพื้นฐาน ของชุดข้อมูลทดสอบ 20 เปอร์เซ็นต์.....	88
ตาราง 73 แสดงผลการคัดเลือกตัวแปรอิสระวิธีการเลือกแบบไปข้างหน้า ของชุดข้อมูลแบบที่ 2.....	89
ตาราง 74 แสดงการตรวจสอบความสัมพันธ์ระหว่างตัวแปรอิสระวิธีการเลือกแบบไปข้างหน้า ของชุดข้อมูลแบบที่ 2.....	90

ตาราง 75 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคการถดถอยลอจิสติกทวิภาควิธีการเลือกแบบไปข้างหน้า ของชุดข้อมูลเรียนรู้ 80 เปอร์เซ็นต์.....	90
ตาราง 76 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคการถดถอยลอจิสติกทวิภาควิธีการเลือกแบบไปข้างหน้า ของชุดข้อมูลทดสอบ 20 เปอร์เซ็นต์	91
ตาราง 77 แสดงผลการคัดเลือกตัวแปรอิสระวิธีการกำจัดแบบถอยหลัง ของชุดข้อมูลแบบที่ 2	91
ตาราง 78 แสดงการตรวจสอบความสัมพันธ์ระหว่างตัวแปรอิสระด้วยวิธีการกำจัดแบบถอยหลัง	92
ตาราง 79 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคการถดถอยลอจิสติกทวิภาควิธีการกำจัดแบบถอยหลัง ของชุดข้อมูลเรียนรู้ 80 เปอร์เซ็นต์	93
ตาราง 80 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคการถดถอยลอจิสติกทวิภาควิธีการกำจัดแบบถอยหลัง ของชุดข้อมูลทดสอบ 20 เปอร์เซ็นต์.....	93
ตาราง 81 แสดงผลการคัดเลือกตัวแปรอิสระวิธีการถดถอยทีละชั้น ของชุดข้อมูลแบบที่ 2	94
ตาราง 82 แสดงการตรวจสอบความสัมพันธ์ระหว่างตัวแปรอิสระด้วยวิธีการถดถอยทีละชั้น	95
ตาราง 83 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคการถดถอยลอจิสติกทวิภาควิธีการถดถอยทีละชั้น ของชุดข้อมูลเรียนรู้ 80 เปอร์เซ็นต์	95
ตาราง 84 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคการถดถอยลอจิสติกทวิภาควิธีการถดถอยทีละชั้น ของชุดข้อมูลทดสอบ 20 เปอร์เซ็นต์.....	96
ตาราง 85 แสดงผลการเปรียบเทียบการตรวจสอบความเหมาะสมของตัวแบบพยากรณ์ด้วยเทคนิคการถดถอยลอจิสติกทวิภาค ของชุดข้อมูลแบบที่ 2.....	96
ตาราง 86 แสดงผลระดับความน่าเชื่อถือของตัวแบบพยากรณ์จากเทคนิคการถดถอยลอจิสติกทวิภาคของชุดข้อมูลแบบที่ 2.....	97

ตาราง 87 แสดงผลการเปรียบเทียบประสิทธิภาพจากเทคนิคการถดถอยลอจิสติกทวิภาค ของชุดข้อมูลแบบที่ 2.....	97
ตาราง 88 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม J48 กรณีตัวแปรอิสระทุกตัว ของชุดข้อมูลเรียนรู้ 80 เปอร์เซ็นต์	98
ตาราง 89 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม J48 กรณีตัวแปรอิสระทุกตัว ของชุดข้อมูลทดสอบ 20 เปอร์เซ็นต์.....	99
ตาราง 90 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม J48 วิธีการเลือกแบบพื้นฐาน ของชุดข้อมูลเรียนรู้ 80 เปอร์เซ็นต์.....	99
ตาราง 91 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม J48 วิธีการเลือกแบบพื้นฐาน ของชุดข้อมูลทดสอบ 20 เปอร์เซ็นต์	99
ตาราง 92 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม J48 วิธีการเลือกแบบไปข้างหน้า ของชุดข้อมูลเรียนรู้ 80 เปอร์เซ็นต์.....	100
ตาราง 93 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม J48 วิธีการเลือกแบบไปข้างหน้า ของชุดข้อมูลทดสอบ 20 เปอร์เซ็นต์.....	100
ตาราง 94 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม J48 วิธีการกำจัดแบบถอยหลัง ของชุดข้อมูลทดสอบ 80 เปอร์เซ็นต์	101
ตาราง 95 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม J48 วิธีการกำจัดแบบถอยหลัง ของชุดข้อมูลทดสอบ 20 เปอร์เซ็นต์	101
ตาราง 96 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม J48 วิธีการถดถอยทีละชั้น ของชุดข้อมูลทดสอบ 80 เปอร์เซ็นต์.....	102
ตาราง 97 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม J48 วิธีการถดถอยทีละชั้น ของชุดข้อมูลทดสอบ 20 เปอร์เซ็นต์	102
ตาราง 98 แสดงผลการเปรียบเทียบประสิทธิภาพด้วยเทคนิคต้นไม้ตัดสินใจอัลกอริทึม J48 ของชุดข้อมูลแบบที่ 2.....	103

ตาราง 99 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วย อัลกอริทึม LMT กรณีตัวแปรอิสระทุกตัว ของชุดข้อมูลเรียนรู้ 80 เปอร์เซ็นต์.....	104
ตาราง 100 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วย อัลกอริทึม LMT กรณีตัวแปรอิสระทุกตัว ของชุดข้อมูลทดสอบ 20 เปอร์เซ็นต์.....	104
ตาราง 101 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วย อัลกอริทึม LMT วิธีการเลือกแบบพื้นฐาน ของชุดข้อมูลเรียนรู้ 80 เปอร์เซ็นต์.....	105
ตาราง 102 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วย อัลกอริทึม LMT วิธีการเลือกแบบพื้นฐาน ของชุดข้อมูลทดสอบ 20 เปอร์เซ็นต์.....	105
ตาราง 103 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วย อัลกอริทึม LMT วิธีการเลือกแบบไปข้างหน้า ของชุดข้อมูลเรียนรู้ 80 เปอร์เซ็นต์.....	106
ตาราง 104 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วย อัลกอริทึม LMT วิธีการเลือกแบบไปข้างหน้า ของชุดข้อมูลทดสอบ 20 เปอร์เซ็นต์.....	106
ตาราง 105 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วย อัลกอริทึม LMT วิธีการกำจัดแบบถอยหลัง ของชุดข้อมูลเรียนรู้ 80 เปอร์เซ็นต์.....	107
ตาราง 106 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วย อัลกอริทึม LMT วิธีการกำจัดแบบถอยหลัง ของชุดข้อมูลทดสอบ 20 เปอร์เซ็นต์.....	107
ตาราง 107 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วย อัลกอริทึม LMT วิธีการถดถอยทีละชั้น ของชุดข้อมูลเรียนรู้ 80 เปอร์เซ็นต์.....	108
ตาราง 108 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วย อัลกอริทึม LMT วิธีการถดถอยทีละชั้น ของชุดข้อมูลทดสอบ 20 เปอร์เซ็นต์.....	108
ตาราง 109 แสดงผลการเปรียบเทียบประสิทธิภาพจากเทคนิคต้นไม้ตัดสินใจอัลกอริทึม LMT ของชุดข้อมูลแบบที่ 2.....	109
ตาราง 110 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคโครงข่ายประสาท เทียม กรณีตัวแปรอิสระทุกตัว ของชุดข้อมูลเรียนรู้ 80 เปอร์เซ็นต์.....	110

ตาราง 111 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคโครงข่ายประสาทเทียม กรณีตัวแปรอิสระทุกตัว ของชุดข้อมูลทดสอบ 20 เปอร์เซ็นต์.....	110
ตาราง 112 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคโครงข่ายประสาทเทียม วิธีการเลือกแบบพื้นฐาน ของชุดข้อมูลเรียนรู้ 80 เปอร์เซ็นต์.....	111
ตาราง 113 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคโครงข่ายประสาทเทียม วิธีการเลือกแบบพื้นฐาน ของชุดข้อมูลทดสอบ 20 เปอร์เซ็นต์.....	111
ตาราง 114 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคโครงข่ายประสาทเทียม วิธีการเลือกแบบไปข้างหน้า ของชุดข้อมูลเรียนรู้ 80 เปอร์เซ็นต์.....	112
ตาราง 115 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคโครงข่ายประสาทเทียม วิธีการเลือกแบบไปข้างหน้า ของชุดข้อมูลทดสอบ 20 เปอร์เซ็นต์.....	112
ตาราง 116 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคโครงข่ายประสาทเทียม วิธีการกำจัดแบบถอยหลัง ของชุดข้อมูลทดสอบ 20 เปอร์เซ็นต์.....	113
ตาราง 117 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคโครงข่ายประสาทเทียม วิธีการกำจัดแบบถอยหลัง ของชุดข้อมูลทดสอบ 20 เปอร์เซ็นต์.....	113
ตาราง 118 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคโครงข่ายประสาทเทียม วิธีการถดถอยทีละชั้น ของชุดข้อมูลเรียนรู้ 80 เปอร์เซ็นต์.....	114
ตาราง 119 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคโครงข่ายประสาทเทียมวิธีการถดถอยทีละชั้น ของชุดข้อมูลทดสอบ 20 เปอร์เซ็นต์.....	114
ตาราง 120 แสดงผลการเปรียบเทียบประสิทธิภาพการพยากรณ์ด้วยเทคนิคโครงข่ายประสาทเทียม ของชุดข้อมูลแบบที่ 2.....	115
ตาราง 121 แสดงผลการเปรียบเทียบประสิทธิภาพตัวแบบพยากรณ์ด้วยเทคนิคต่าง ๆ ของชุดข้อมูลแบบที่ 2.....	116
ตาราง 122 แสดงผลการเปรียบเทียบประสิทธิภาพตัวแบบพยากรณ์ที่ดีที่สุด.....	116
ตาราง 123 แสดงผลการคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกพหุภาค	118

ตาราง 124 แสดงผลการเปรียบเทียบประสิทธิภาพตัวแบบพยากรณ์ที่ดีที่สุด 118

ตาราง 125 แสดงเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์ที่ดีที่สุดในแต่ละชุดข้อมูล
 119



สารบัญภาพ

	หน้า
ภาพ 1 จำนวนผู้ป่วยโรคเบาหวานทั่วโลก ในปี พ.ศ. 2543 – 2562	2
ภาพ 2 กระบวนการของการทำเหมืองข้อมูล	14
ภาพ 3 ตัวอย่างต้นไม้ตัดสินใจ	16
ภาพ 4 เหตุการณ์การเลือกตัวแปรอิสระเป็นโหนดเริ่มต้น	22
ภาพ 5 การพิจารณาโหนดภายใน middle_aged	22
ภาพ 6 ต้นไม้ตัดสินใจ	23
ภาพ 7 เซลล์ประสาทจริงและแบบจำลองเซลล์ประสาทเทียม	24
ภาพ 8 โครงข่ายประสาทเทียมหลายชั้นแบบไปข้างหน้า	24
ภาพ 9 กระบวนการเรียนรู้ของโครงข่ายประสาทเทียม	25
ภาพ 10 ตัวอย่างโครงข่ายประสาทเทียมที่ได้จากโปรแกรมอาร์	27
ภาพ 11 แสดงโครงสร้างของตัวแบบโครงข่ายประสาทเทียมของข้อมูลชุดที่ 1	30
ภาพ 12 แสดงโครงสร้างของตัวแบบโครงข่ายประสาทเทียมของข้อมูลชุดที่ 2	32
ภาพ 13 การแบ่งชุดข้อมูลแบบที่ 1 ด้วยวิธีการสุ่มตัวอย่างแบบแบ่งชั้นภูมิด้วยสัดส่วน 70 : 30	50
ภาพ 14 การแบ่งชุดข้อมูลแบบที่ 2 ด้วยวิธีการสุ่มตัวอย่างแบบแบ่งชั้นภูมิด้วยสัดส่วน 80 : 20	51
ภาพ 15 แผนผังการสร้างและพัฒนาตัวแบบพยากรณ์ด้วยเทคนิคการถดถอยลอจิสติก ทวิภาค	53
ภาพ 16 แผนผังการสร้างและพัฒนาตัวแบบพยากรณ์ด้วยเทคนิคต้นไม้ตัดสินใจ	54

ภาพ 17 แผนผังการสร้างและพัฒนาตัวแบบพยากรณ์ด้วยเทคนิคโครงข่ายประสาทเทียม

.....54



บทที่ 1

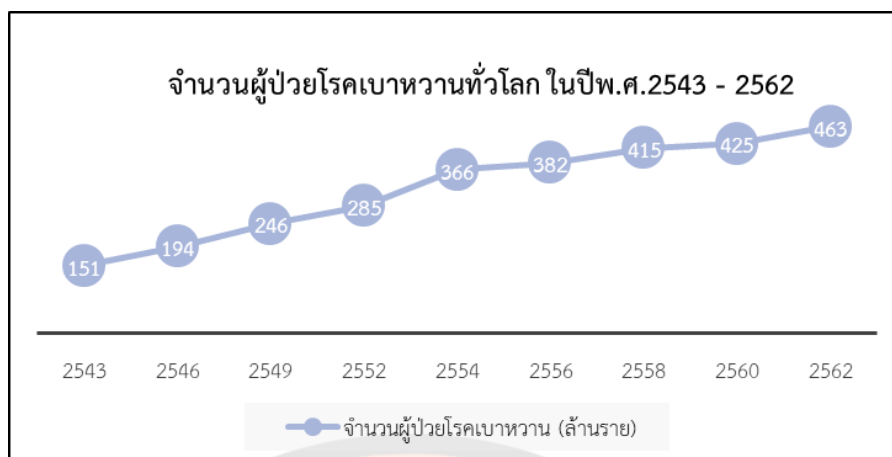
บทนำ

1.1 ที่มาและความสำคัญ

โรคเบาหวาน (Diabetes Mellitus: DM) เป็นภาวะที่ร่างกายมีน้ำตาลในเลือดสูงกว่าปกติ เนื่องจากการขาดฮอร์โมนอินซูลิน (Insulin) หรือไม่สามารผลิตอินซูลินได้เพียงพอ และไม่สามารถใช้อินซูลินได้อย่างมีประสิทธิภาพ อินซูลินเป็นฮอร์โมนที่ผลิตขึ้นจากเบต้าเซลล์ในตับอ่อน ทำหน้าที่ช่วยให้กลูโคสจากกระแสเลือดเข้าสู่เซลล์ของร่างกาย ซึ่งกลูโคสนั้นจะถูกเปลี่ยนเป็นพลังงาน อินซูลินยังจำเป็นสำหรับการเผาผลาญโปรตีนและไขมัน การขาดอินซูลินหรือไม่สามารถตอบสนองต่อเซลล์ได้จะทำให้ระดับน้ำตาลในเลือดสูง (Hyperglycaemia) หากปล่อยทิ้งไว้โดยไม่ได้รับการรักษา อาจทำให้เกิดความเสียหายต่ออวัยวะต่าง ๆ ของร่างกาย ซึ่งนำไปสู่ภาวะแทรกซ้อนทางสุขภาพและเป็นอันตรายถึงชีวิตได้ (*IDF Diabetes Atlas 9th Edition 2019, 2019*)

โรคเบาหวานแบ่งออกเป็น 4 ชนิด ได้แก่ โรคเบาหวานชนิดที่ 1 (Type 1 Diabetes Mellitus, T1DM) คือ การที่เบต้าเซลล์ถูกทำลายจากภูมิคุ้มกันของร่างกาย ส่วนใหญ่พบในเด็ก รูปร่างไม่อ้วน มีอาการปัสสาวะมาก กระหายน้ำมาก อ่อนเพลีย น้ำหนักลดลง อาจเกิดขึ้นได้อย่างรวดเร็วและรุนแรง ซึ่งบางกรณีพบภาวะเลือดเป็นกรดจากสารคีโตน (Ketoacidosis) โรคเบาหวานชนิดที่ 2 (Type 2 Diabetes Mellitus, T2DM) เป็นผลมาจากมีภาวะดื้อต่ออินซูลิน (Insulin Resistance) ร่วมกับความบกพร่องในการผลิตอินซูลินที่เหมาะสม (Relative Insulin Deficiency) ส่วนมากพบในคนสูงอายุ รูปร่างท้วมหรืออ้วน โรคเบาหวานชนิดที่ 3 คือ โรคเบาหวานขณะตั้งครรภ์ (Gestational Diabetes Mellitus, GDM) เกิดจากการมีภาวะดื้อต่ออินซูลินมากขึ้นระหว่างตั้งครรภ์ จากปัจจัยที่เกิดจากรกและตับอ่อนของมารดาไม่สามารถผลิตอินซูลินให้เพียงพอต่อความต้องการได้ ทำให้เกิดโรคเบาหวานชนิดนี้ และ โรคเบาหวานชนิดที่ 4 คือ โรคเบาหวานที่มีสาเหตุจำเพาะ (Specific Types of Diabetes Due to other Causes) เป็นโรคเบาหวานที่มีสาเหตุชัดเจน เช่น โรคเบาหวานที่เกิดจากความผิดปกติทางพันธุกรรม (Maturity - Onset Diabetes of the Young: MODY) (สมาคมโรคเบาหวานแห่งประเทศไทยในพระราชูปถัมภ์สมเด็จพระเทพรัตนราชสุดาฯ สยามบรมราชกุมารี, 2560)

จากข้อมูลสมาพันธ์เบาหวานนานาชาติ ในปี พ.ศ. 2562 มีการประเมินว่ามีผู้ป่วยเบาหวานทั่วโลกประมาณ 463 ล้านราย และคาดการณ์ในปี พ.ศ.2573 จำนวนผู้ป่วยเพิ่มสูงขึ้นถึง 573 ล้านราย และภายในปี พ.ศ. 2588 อาจเพิ่มสูงขึ้นถึง 700 ล้านราย ซึ่งจะเห็นว่าจำนวนผู้ป่วยโรคเบาหวานทั่วโลกเพิ่มขึ้นอย่างต่อเนื่อง ดังภาพ 1



ภาพ 1 จำนวนผู้ป่วยโรคเบาหวานทั่วโลก ในปี พ.ศ. 2543 – 2562

จากข้อมูลผู้ป่วยโรคเบาหวานในประเทศไทย ในปี พ.ศ. 2561 ซึ่งจัดเก็บโดยสำนักกระบาดวิทยา กรมควบคุมโรค กระทรวงสาธารณสุข จากข้อมูลทั่วประเทศ 77 จังหวัด พบว่าจำนวนผู้ป่วยเบาหวานที่ทั้งหมด 2,659,940 ราย แยกเป็น เพศชาย 943,775 ราย เพศหญิง 1,716,165 ราย คิดเป็นอัตราส่วนเพศชายต่อเพศหญิง เท่ากับ 1:1.8 เมื่อจำแนกตามช่วงอายุ พบว่ากลุ่มอายุที่พบการเป็นโรคเบาหวานมากที่สุด 3 อันดับแรก คือ อายุ 70 – 79 ปี 60 – 69 ปี และ 80 ปีขึ้นไป ตามลำดับ แสดงดังตาราง 1

ตาราง 1 แสดงอัตราความชุกของผู้ป่วยเบาหวานใน ปี พ.ศ. 2561 จำแนกตามกลุ่มอายุ

กลุ่มอายุ	อัตราความชุก (ต่อประชากรแสนราย)
น้อยกว่า 15 ปี	19.45
15 – 29 ปี	131.08
30 – 44 ปี	1,103.53
45 – 59 ปี	6,229.43
60 – 69 ปี	14,597.35
70 – 79 ปี	16,692.85
มากกว่า 80 ปี	13,341.85

เมื่อจำแนกตามภูมิภาค พบว่าภูมิภาคที่พบมากที่สุด คือ ภาคกลาง ภาคเหนือ ภาคใต้ และภาคตะวันออกเฉียงเหนือ ตามลำดับ แสดงดังตาราง 2

ตาราง 2 แสดงอัตราความชุกของผู้ป่วยเบาหวาน ปี พ.ศ. 2561 จำแนกตามภูมิภาค

ภาค	อัตราความชุก (ต่อประชากรแสนราย)
เหนือ	1,922.90
กลาง	2,154.89
ตะวันออกเฉียงเหนือ	1,625.37
ใต้	1,909.36

จากการสำรวจพฤติกรรมเสี่ยงการเป็นโรคเบาหวานด้วยข้อมูลปี พ.ศ. 2548 – 2558 พบว่า ปัจจัยที่ส่งผลต่อการเป็นโรคเบาหวาน คือการสูบบุหรี่ การดื่มแอลกอฮอล์ และภาวะน้ำหนักเกิน (ณัฐกุล ไชยสงคราม, 2561)

จากที่กล่าวมาโรคเบาหวานเป็นโรคที่พบแทบทุกภูมิภาคของประเทศไทย โดยจำนวนผู้ป่วยมีแนวโน้มเพิ่มขึ้นเป็นผลให้อัตราการเสียชีวิตจากผู้ป่วยโรคเบาหวานเพิ่มขึ้น ทำให้ผู้วิจัยสนใจที่จะศึกษาปัจจัยที่ส่งผลต่อการเป็นโรคเบาหวาน ตลอดจนนำเสนอแนวทางการพัฒนาตัวแบบเพื่อพยากรณ์การเป็นโรคเบาหวานด้วยวิธีการทางสถิติและวิธีทางเทคนิคเหมืองข้อมูล โดยตัวแบบที่มีประสิทธิภาพจะสามารถนำมาใช้คัดกรองผู้ป่วยโรคเบาหวานเบื้องต้นได้

ตัวแบบพยากรณ์การเป็นโรคเบาหวานที่ผู้วิจัยนำเสนอในงานวิจัยนี้จะเป็นการทำนายการเป็นโรคเบาหวานและไม่เป็นโรคเบาหวาน ซึ่งการทำงานในลักษณะนี้จะสอดคล้องกับหลักการทำงานของการถดถอยลอจิสติกทวิภาค ซึ่งเป็นตัวแบบพยากรณ์ทางสถิติที่เป็นการวิเคราะห์ที่มีตัวแปรตามเป็นตัวแปรเชิงคุณภาพที่มีเพียง 2 ค่า ซึ่งสอดคล้องกับงานวิจัยนี้ที่เป็นการพยากรณ์คำตอบเพียง 2 ค่าคือการเป็นและไม่เป็นโรคเบาหวาน การถดถอยลอจิสติกทวิภาคเป็นการวิเคราะห์ทางสถิติที่ได้รับค่านิยมในการนำมาใช้เพื่อการจำแนกประเภทข้อมูลเชิงคุณภาพ โดยใช้หลักการความน่าจะเป็นมาเพื่อพยากรณ์โอกาสที่จะเกิดเหตุการณ์ที่สนใจ และสามารถวิเคราะห์ได้ว่าตัวแปรอิสระใดเป็นปัจจัยที่ส่งผลให้เกิดโรคเบาหวาน (กัลยา วานิชย์บัญชา, 2555) (Montgomery et al., 2020)

นอกจากนี้ได้มีผู้วิจัยหลายท่านได้ประยุกต์ใช้เทคนิคเหมืองข้อมูล เพื่อศึกษาปัจจัยการเป็นโรคเบาหวาน และพัฒนาตัวแบบพยากรณ์การเป็นโรคเบาหวาน โดยได้มีการประยุกต์ใช้หลักการทำงานตามรูปแบบเหมืองข้อมูล เช่น การแบ่งชุดข้อมูล เป็นชุดข้อมูลเรียนรู้ และชุดข้อมูลทดสอบ การพัฒนาและทดสอบตัวแบบพยากรณ์ตามแนวทาง 10 Fold Cross – Validation การประเมินประสิทธิภาพตัวแบบพยากรณ์ด้วยค่าความถูกต้อง เป็นต้น โดยเทคนิคเหมืองข้อมูลที่ได้รับค่านิยมสำหรับใช้ในการพัฒนาตัวแบบพยากรณ์ คือเทคนิคต้นไม้ตัดสินใจ และ เทคนิคโครงข่ายประสาทเทียม ดังเช่น การเปรียบเทียบประสิทธิภาพการจำแนกผู้ป่วยโรคเบาหวานด้วยเทคนิคเหมืองข้อมูลบน ข้อมูล 620 ชุด พบว่าเทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม C4.5 มีค่าความถูกต้องสูงที่สุด

(Kumar & Umatejaswi, 2017) การสร้างตัวแบบเพื่อพยากรณ์การเป็นโรคเบาหวานบนข้อมูล 768 ชุด โดยใช้วิธีการคัดเลือกตัวแปรอิสระที่มีปัจจัยหลักที่ส่งผลให้เป็นโรคเบาหวาน ด้วยเทคนิคต้นไม้ตัดสินใจ พบว่าค่าความถูกต้อง คือ 79.08 เปอร์เซ็นต์ (Paul, 2017) และการสร้างตัวแบบพยากรณ์การเป็นโรคเบาหวาน โดยใช้เทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม C4.5 โดย แบ่งข้อมูลเป็นชุดข้อมูลเรียนรู้ 80 เปอร์เซ็นต์ และชุดข้อมูลทดสอบ 20 เปอร์เซ็นต์ พบว่าค่าความถูกต้อง คือ 72.08 เปอร์เซ็นต์ (Sunge et al., 2019) และนอกจากนี้ยังมีการศึกษาเปรียบเทียบประสิทธิภาพในการพยากรณ์การเป็นโรคเบาหวานบนข้อมูล 1,233 ชุด พบว่าเทคนิคโครงข่ายประสาทเทียมมีประสิทธิภาพในการพยากรณ์ดีที่สุด โดยให้ค่าความถูกต้อง เท่ากับ 95.94 (สายชล สันสมบุรณ์ทอง, 2561)

จากการศึกษางานวิจัยที่เกี่ยวข้องกับการพัฒนาตัวแบบพยากรณ์ด้วยเทคนิคเหมืองข้อมูล พบว่าเทคนิคต้นไม้ตัดสินใจ และเทคนิคโครงข่ายประสาทเทียม ให้ตัวแบบพยากรณ์ที่มีประสิทธิภาพ ดังนั้นผู้วิจัยจึงมีแนวคิดที่จะใช้เทคนิคเหมืองข้อมูล 2 เทคนิคนี้ในการพัฒนาตัวแบบพยากรณ์การเป็นโรคเบาหวาน

เนื่องจากข้อมูลของผู้ป่วยที่เข้ารับการรักษาในโรงพยาบาลมีเป็นจำนวนมาก ทำให้ผู้วิจัยสนใจศึกษาข้อมูลที่เกี่ยวข้องของผู้ป่วยโรคเบาหวาน เพื่อนำมาสร้างตัวแบบพยากรณ์การเป็นโรคเบาหวาน โดยใช้วิธีการทางสถิติและเทคนิคเหมืองข้อมูล ทำการเปรียบเทียบประสิทธิภาพโดยใช้ค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุลในการพยากรณ์

1.2 จุดมุ่งหมายของการวิจัย

1. เพื่อวิเคราะห์ปัจจัยที่ส่งผลต่อการเป็นโรคเบาหวาน โดยใช้การถดถอยลอจิสติกทวิภาค
2. เพื่อสร้างตัวแบบพยากรณ์การเป็นโรคเบาหวาน โดยใช้การถดถอยลอจิสติกทวิภาค

(Binary Logistic Regression) ต้นไม้ตัดสินใจ (Decision Tree) และ โครงข่ายประสาทเทียม (Artificial Neural Network)

3. เพื่อเปรียบเทียบประสิทธิภาพในการพยากรณ์การเป็นโรคเบาหวานของตัวแบบการถดถอยลอจิสติกทวิภาค ต้นไม้ตัดสินใจ และ โครงข่ายประสาทเทียม

1.3 ขอบเขตการวิจัย

1. ข้อมูลที่ใช้ในการวิจัยครั้งนี้ คือข้อมูลผู้ป่วยทั่วไปที่เข้ารับการรักษาในโรงพยาบาล มหาวิทยาลัยนเรศวร ตั้งแต่วันที่ 1 มกราคม 2563 – 31 มิถุนายน 2563 จำนวน 5,081 ชุด ตัวแปรอิสระที่ใช้ในการวิเคราะห์โอกาสการเป็นโรคเบาหวานมีจำนวน 13 ตัวแปร ได้แก่ เพศ (Gender) อายุ (Age) ค่าความดันขณะหัวใจบีบตัว (Systolic Blood Pressure) ค่าความดันขณะหัวใจคลายตัว

(Diastolic Blood Pressure) อัตราการเต้นของหัวใจ (Heart Rate) น้ำหนัก (Body Weight) ความสูง (Height) ดัชนีมวลกาย (BMI) ระดับน้ำตาลในเลือด (FBS) ระดับคอเลสเตอรอล ในเลือด (Cholesterol) ระดับไตรกลีเซอไรด์ในเลือด (Triglyceride) ระดับไขมันในเลือดที่มีความหนาแน่นสูง (HDL) และ ระดับไขมันในเลือดที่มีความหนาแน่นต่ำ (LDL) และตัวแปรตาม คือ การเป็นโรคเบาหวาน

2. สร้างตัวแบบพยากรณ์โดยใช้เทคนิคการถดถอยลอจิสติกทวิภาค เทคนิคต้นไม้ตัดสินใจ และ เทคนิคโครงข่ายประสาทเทียม ตามลำดับ โดยแบ่งออกเป็น ชุดข้อมูลเรียนรู้ (Training Data) 70 เปอร์เซ็นต์ และ 80 เปอร์เซ็นต์ ชุดข้อมูลทดสอบ (Testing Data) 20 เปอร์เซ็นต์ และ 30 เปอร์เซ็นต์ สำหรับเกณฑ์ที่ใช้ในการเปรียบเทียบประสิทธิภาพตัวแบบสำหรับการพยากรณ์ ได้แก่ ค่าความถูกต้อง (Accuracy) และค่าความถูกต้องในการจำแนกแบบสมดุล (Balance Classification Accuracy)

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1. ทราบถึงปัจจัยที่ส่งผลต่อโรคเบาหวาน เพื่อนำมาใช้ในการสร้างตัวแบบทางสถิติ
2. ได้ตัวแบบทางสถิติสำหรับการพยากรณ์โรคเบาหวานด้วยเทคนิคเหมืองข้อมูล
3. ได้ตัวแบบพยากรณ์ที่พัฒนามาจากการคัดเลือกตัวแปรอิสระจากเทคนิคการถดถอย ลอจิสติกทวิภาค

1.5 นิยามศัพท์เฉพาะ

ในการนำเสนอผลการวิจัยสำหรับงานวิจัยครั้งนี้ เพื่อให้เกิดความเข้าใจตรงกันในการแปลความหมาย ผู้วิจัยจึงใช้สัญลักษณ์และอักษรย่อในการวิจัยดังนี้

X_1	แทน	เพศ
X_2	แทน	อายุ
X_3	แทน	ค่าความดันขณะหัวใจบีบตัว
X_4	แทน	ค่าความดันขณะหัวใจคลายตัว
X_5	แทน	อัตราการเต้นของหัวใจ
X_6	แทน	น้ำหนัก
X_7	แทน	ความสูง
X_8	แทน	ดัชนีมวลกาย
X_9	แทน	ระดับน้ำตาลในเลือด

X_{10}	แทน	ระดับคอเลสเทอรอลในเลือด
X_{11}	แทน	ระดับไตรกลีเซอไรด์ในเลือด
X_{12}	แทน	ระดับไขมันในเลือดที่มีความหนาแน่นสูง
X_{13}	แทน	ระดับไขมันในเลือดที่มีความหนาแน่นต่ำ
β_i	แทน	ค่าสัมประสิทธิ์การถดถอยลอจิสติก
OR	แทน	อัตราส่วนออดส์ (Odds Ratio)
AIC	แทน	เกณฑ์สารสนเทศอะกะอิเกะ (Akaike's Information Criterion)
ACC	แทน	ค่าความถูกต้อง (Accuracy)
BCA	แทน	ค่าความถูกต้องในการจำแนกแบบสมดุล (Balance Classification Accuracy)
df	แทน	องศาเสรี (Degree of Freedom)
LL	แทน	ลอการิทึมของภาวะน่าจะเป็น (Log Likelihood)
R_{ij}^2	แทน	สัมประสิทธิ์สหสัมพันธ์พหุคูณ
SE	แทน	ความคลาดเคลื่อนมาตรฐาน (Standard Error)
TP	แทน	ค่าความถูกต้องเชิงบวก (True Positive)
TN	แทน	ค่าความถูกต้องเชิงลบ (True Negative)
FP	แทน	ค่าความผิดพลาดเชิงบวก (False Positive)
FN	แทน	ค่าความผิดพลาดเชิงลบ (False Negative)
$Recall$	แทน	ค่าระลึก
$Precision$	แทน	ค่าความแม่นยำ
$F - Measure$	แทน	ค่าประสิทธิภาพโดยรวม

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

การวิจัยครั้งนี้ มีวัตถุประสงค์เพื่อเปรียบเทียบค่าความถูกต้องของตัวแบบการถดถอยลอจิสติกทวิภาค และเทคนิคเหมืองข้อมูลสำหรับพยากรณ์การเป็นโรคเบาหวาน โดยใช้ข้อมูลผู้ป่วยที่เข้ามารับการรักษาในโรงพยาบาลมหาวิทยาลัยนเรศวร โดยมีทฤษฎีและงานวิจัยที่เกี่ยวข้องดังนี้

2.1 การถดถอยลอจิสติก (Logistic Regression)

การถดถอยลอจิสติกเป็นวิธีการทางสถิติที่ใช้ในการศึกษาเพื่อการพยากรณ์ความน่าจะเป็นของการเกิดเหตุการณ์ที่สนใจต่อความน่าจะเป็นของการไม่เกิดเหตุการณ์ที่สนใจ ซึ่งการถดถอยแบบลอจิสติกนี้ จะมีตัวแปรตามเป็นตัวแปรเชิงคุณภาพ แบ่งออกเป็น 2 ประเภท ได้แก่ การถดถอยลอจิสติกทวิภาค (Binary Logistic Regression) คือตัวแปรตามมีค่าเพียง 2 ค่า (Dichotomus Variable) และการถดถอยลอจิสติกพหุ (Multinomial Logistic Regression) คือตัวแปรตามมีค่ามากกว่า 2 ค่าขึ้นไป ในส่วนของตัวแปรอิสระ อาจจะเป็นตัวแปรเชิงปริมาณ หรือตัวแปรเชิงคุณภาพ หรืออาจจะเป็นทั้งตัวแปรเชิงปริมาณและตัวแปรเชิงคุณภาพก็ได้ (กัลยา วานิชย์บัญชา, 2555) ซึ่งในงานวิจัยนี้จะศึกษาเฉพาะตัวแบบการถดถอยลอจิสติกทวิภาค

2.1.1 วัตถุประสงค์ของการถดถอยลอจิสติกทวิภาค

1. เพื่อศึกษาระดับความสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระแต่ละตัว และศึกษาว่ามีตัวแปรอิสระใดบ้างที่สามารถอธิบายโอกาสที่จะทำให้เกิดเหตุการณ์ที่สนใจ และการไม่เกิดเหตุการณ์ที่สนใจ
2. เพื่อพยากรณ์โอกาสที่จะเกิดเหตุการณ์ที่สนใจ จากตัวแบบการถดถอยลอจิสติกทวิภาคที่เหมาะสม

2.1.2 ข้อตกลงเบื้องต้นของการถดถอยลอจิสติกทวิภาค

1. ตัวแปรตามเป็นข้อมูลเชิงคุณภาพ มีเพียง 2 ค่า
2. ลอจิท (Logit) หรือ ลอการิทึมของอัตราส่วนระหว่างความน่าจะเป็นของการเกิดเหตุการณ์ที่สนใจกับความน่าจะเป็นของการไม่เกิดเหตุการณ์ที่สนใจของตัวแปรตาม มีความสัมพันธ์เชิงเส้นกับตัวแปรอิสระ
3. ตัวแปรอิสระไม่ควรค่าที่มีอิทธิพล (Influential Value) เช่น ค่าสุดขีด (Extreme Value) หรือ ค่านอกเกณฑ์ (Outlier)
4. ตัวแปรอิสระไม่ควรมีความสัมพันธ์กันหรือไม่ควรเกิดปัญหาความสัมพันธ์เชิงเส้นพหุ (Multicollinearity) (Kassambara, 2018)

2.1.3 ตัวแบบการถดถอยลอจิสติกทวิภาค

การถดถอยลอจิสติกทวิภาค เป็นการศึกษาเพื่อพยากรณ์โอกาสที่จะเกิดเหตุการณ์ที่สนใจ โดยใช้รูปแบบความสัมพันธ์ที่สร้างจากตัวแปรอิสระ โดยตัวแปรตามของเหตุการณ์ที่สนใจ จะมีค่าเป็นไปได้เพียง 2 ค่า คือ 0 และ 1 เมื่อให้ Y_i แทนตัวแปรตามดังกล่าว จะได้ว่า Y_i มีการแจกแจงแบบแบร์นูลลี และ $P(Y_i)$ คือ ความน่าจะเป็นที่จะเกิดเหตุการณ์ ดังนั้นถ้าให้ความน่าจะเป็นของการเกิดเหตุการณ์ที่สนใจ $P(Y_i = 1)$ มีค่าเท่ากับ π_i แล้วความน่าจะเป็นของการไม่เกิดเหตุการณ์ที่สนใจ $P(Y_i = 0)$ จะมีค่าเท่ากับ $1 - \pi_i$ โดยที่ π_i อยู่ในช่วง $[0,1]$

พิจารณา $X_i' = [1, X_{i1}, X_{i2}, \dots, X_{ik}]$, $\beta' = [\beta_0, \beta_1, \beta_2, \dots, \beta_k]$ และ $E(y_i) = X_i' \beta = \pi_i$ ตัวแบบการถดถอยลอจิสติกจะมีรูปแบบดังต่อไปนี้ (Montgomery, Peck and Vining, 2012)

$$\begin{aligned} E(Y_i) &= \frac{\exp(X_i' \beta)}{1 + \exp(X_i' \beta)} \\ &= \frac{1}{1 + \exp(-X_i' \beta)} \end{aligned} \quad (1)$$

ซึ่งฟังก์ชันลอจิสติกในสมการ (1) สามารถเขียนให้อยู่ในรูปอัตราส่วนออดส์ (Odds Ratio) ซึ่งหมายถึง อัตราส่วนระหว่างโอกาสที่จะเกิดเหตุการณ์ที่สนใจ ($Y_i = 1$) ต่อโอกาสที่จะไม่เกิดเหตุการณ์ที่สนใจ ($Y_i = 0$) ได้ดังนี้

$$\text{Odds Ratio} = \frac{\pi_i}{1 - \pi_i} \quad (2)$$

สำหรับการวิเคราะห์การถดถอยจึงต้องมีการปรับให้ความสัมพันธ์อยู่ในรูปเชิงเส้น นั่นคือเขียนให้อยู่ในรูปลอการิทึมของออดส์ เรียกว่า ลอจิต ซึ่งสามารถเขียนความสัมพันธ์ให้อยู่ในรูปเชิงเส้นได้ดังนี้

$$\begin{aligned} \ln(\text{Odds}) = \text{logit} &= \ln\left(\frac{\pi_i}{1 - \pi_i}\right) \\ &= X_i' \beta \end{aligned} \quad (3)$$

เป้าหมายของการถดถอยลอจิสติกทวิภาค คือ การประมาณค่า $X_i' \beta$ และเนื่องจาก Y_i มีการแจกแจงแบบแบร์นูลลี ดังนั้นสำหรับตัวอย่างหน่วยที่ i มีฟังก์ชันความน่าจะเป็นคือ

$$f_i(y_i) = (\pi_i)^{y_i} (1 - \pi_i)^{1 - y_i}$$

เมื่อ $i = 1, 2, \dots, n$ และเนื่องจากข้อมูลตัวอย่าง n หน่วยมีความเป็นอิสระกัน ดังนั้นจะได้ฟังก์ชันภาวะน่าจะเป็น (Likelihood Function) คือ

$$\begin{aligned} L(y_1, y_2, \dots, y_n, \beta) &= \prod_{i=1}^n f_i(y_i) \\ &= \prod_{i=1}^n (\pi_i)^{y_i} (1 - \pi_i)^{1 - y_i} \end{aligned} \quad (4)$$

ซึ่งลอการิทึมธรรมชาติของฟังก์ชันภาวะน่าจะเป็นสูงสุด คือ

$$\begin{aligned}
 \ln L(y_1, y_2, \dots, y_n, \beta) &= \ln \prod_{i=1}^n (\pi_i)^{y_i} (1 - \pi_i)^{1-y_i} \\
 &= \sum_{i=1}^n \left[y_i \ln \left(\frac{\pi_i}{1 - \pi_i} \right) + \ln(1 - \pi_i) \right] \\
 &= \sum_{i=1}^n \left[y_i \ln \left(\frac{\pi_i}{1 - \pi_i} \right) \right] + \sum_{i=1}^n \ln(1 - \pi_i) \\
 &= \sum_{i=1}^n y_i X_i' \beta - \sum_{i=1}^n \ln[1 + \exp(X_i' \beta)]
 \end{aligned}$$

เนื่องจาก $\ln \left(\frac{\pi_i}{1 - \pi_i} \right) = X_i' \beta$ และ $1 - \pi_i = [1 + \exp(X_i' \beta)]^{-1}$ ดังนั้น

$$\ln L(y_1, y_2, \dots, y_n, \beta) = \sum_{i=1}^n \left[y_i (X_i' \beta) - \ln(1 + \exp(X_i' \beta)) \right] \quad (5)$$

จากที่กล่าวมาข้างต้น จะได้ว่า การประมาณค่า $X_i' \beta$ จะใช้วิธีภาวะน่าจะเป็นสูงสุด (Maximum Likelihood method) หรือประมาณค่า $X_i' \beta$ ที่ทำให้ $\ln L(y_1, y_2, \dots, y_n, \beta)$ มีค่ามากที่สุด โดยใช้เทคนิคการทำซ้ำ (Iteration techniques) ซึ่ง (Haberman, 1978) ได้ศึกษาโดยใช้วิธีการของนิวตัน-ราฟสัน (Newton-Raphson method) พบว่าเป็นวิธีที่มีประสิทธิภาพ ส่วนใหญ่จะใช้โปรแกรมสำเร็จรูปทางสถิติในการประมาณค่า (ภักศุภางค์ มาปรีดา, 2560)

2.1.4 วิธีการคัดเลือกตัวแปร (Variable Selection)

การถดถอยลอจิสติกทวิภาค เป็นการวิเคราะห์เพื่อพยากรณ์โอกาสที่จะเกิดเหตุการณ์ที่สนใจ โดยตัวแบบการถดถอยลอจิสติกจะต้องประกอบด้วยตัวแปรอิสระที่เหมาะสมที่จะทำให้ค่าพยากรณ์ที่จะเกิดขึ้นนั้นใกล้เคียงกับค่าจริง ในการเลือกตัวแปรอิสระมีวิธีเลือก 3 วิธี ดังนี้ (พจนานว สวัสดิ์, 2547)

1) วิธีการเลือกแบบไปข้างหน้า (Forward Selection) เป็นวิธีการเลือกตัวแปรอิสระที่มีความสัมพันธ์กับตัวแปรตามมากที่สุด เข้ามาอยู่ในตัวแบบทีละตัว โดยพิจารณาจากตัวแปรอิสระที่มีความสัมพันธ์มากที่สุดเข้ามาในตัวแบบเป็นอันดับแรก จากนั้นทำการทดสอบว่าตัวแปรที่เข้ามาสามารถพยากรณ์ตัวแปรตามได้เพิ่มขึ้นอย่างมีนัยสำคัญเชิงสถิติหรือไม่ ขั้นตอนต่อไปจะทำการคัดเลือกตัวแปรอิสระที่มีความสัมพันธ์รองลงมา จนกระทั่งไม่มีตัวแปรอิสระใดเข้ามาอยู่ในตัวแบบได้อีก วิธีการเลือกตัวแปรอิสระแบบไปข้างหน้าจะสิ้นสุดการทำงาน

2) วิธีการกำจัดแบบถอยหลัง (Backward Elimination) เป็นวิธีที่นำตัวแปรอิสระทุกตัวเข้าสู่ตัวแบบพร้อมกัน แล้วพิจารณาว่าควรจะนำตัวแปรอิสระตัวใดออกจากตัวแบบ โดยในแต่ละขั้นตอนจะตัดตัวแปรอิสระออกจากตัวแบบครั้งละตัว กรณีในการพิจารณาว่าควรตัดตัวแปรอิสระใดออกจากตัวแบบ คือการตัดตัวแปรอิสระที่ไม่มีความสัมพันธ์ต่อตัวแปรตามออก ทำเช่นนี้ไปจนกระทั่ง

เหลือตัวแปรอิสระที่สามารถอธิบายความผันแปร หรือมีความสัมพันธ์ต่อตัวแปรตามอย่างมีนัยสำคัญเชิงสถิติ

3) การถดถอยทีละขั้น (Stepwise Regression) เป็นวิธีการเลือกตัวแปรอิสระเข้าสู่ตัวแบบ โดยเริ่มจากการนำตัวแปรอิสระที่มีความสัมพันธ์กับตัวแปรตามเข้าสู่ตัวแบบทีละตัว จากนั้นทำการตัดตัวแปรอิสระที่มีความสัมพันธ์กับตัวแปรตามตัวที่สองเข้าไป ซึ่งวิธีการนี้คือ วิธีการเลือกแบบไปข้างหน้า หลังจากนั้นทำการคัดเลือกตัวแปรอิสระโดยวิธีการกำจัดแบบถดถอยหลัง เพื่อตรวจสอบว่าตัวแปรอิสระที่เข้าสู่ตัวแบบทั้งสองตัว ตัวแปรอิสระใดควรถูกคัดออกจากตัวแบบ ทำเช่นนี้ไปจนกระทั่งตัวแปรอิสระที่อยู่ในตัวแบบสามารถอธิบายความผันแปรของตัวแปรตามได้อย่างมีนัยสำคัญเชิงสถิติ

2.1.5 การตรวจสอบความสัมพันธ์ระหว่างตัวแปรอิสระ

ในการถดถอยลอจิสติกทวิภาค ตัวแปรอิสระไม่ควรมีความสัมพันธ์กัน หรือไม่ควรเกิดปัญหาความสัมพันธ์เชิงเส้นพหุ ซึ่งการตรวจสอบความสัมพันธ์ของตัวแปรอิสระ สามารถพิจารณาจากค่า Variance Inflation Factor (VIF) โดยมีหลักเกณฑ์ในการพิจารณา คือ ถ้าค่า VIF มากกว่า 10 แสดงว่าตัวแปรอิสระเหล่านั้น มีความสัมพันธ์กันในระดับมาก หรือ เรียกว่า ความสัมพันธ์เชิงเส้นพหุ (Multicollinearity) ซึ่งค่า VIF สามารถคำนวณได้จากสมการ (6) (Myers, 1986, 1990)

$$VIF = \frac{1}{1 - R_i^2} \quad (6)$$

หรือพิจารณาจากค่าความคลาดเคลื่อนยินยอม (Tolerance) ของตัวแปรอิสระ X_i เท่ากับ $1 - R_i^2$ มีค่าอยู่ระหว่าง 0 ถึง 1 เพื่อแสดงความสัมพันธ์ระหว่างตัวแปรอิสระ X_i กับตัวแปรอิสระตัวอื่น ๆ คือ ถ้าค่าความคลาดเคลื่อนยินยอม มีค่าเข้าใกล้ 1 แสดงว่าตัวแปรอิสระ X_i นั้นมีความสัมพันธ์กับตัวแปรอื่น ๆ น้อย แต่ถ้าค่าความคลาดเคลื่อนยินยอมมีค่าเข้าใกล้ศูนย์ แสดงว่าตัวแปรอิสระ X_i มีความสัมพันธ์กับตัวแปรอิสระอื่น ๆ มาก นั่นคือเกิดความสัมพันธ์เชิงเส้นพหุ ถ้าค่าต่ำกว่า 0.1 แสดงว่ามีปัญหาเกี่ยวกับความสัมพันธ์เชิงเส้นพหุขั้นรุนแรง

$$Tolerance = 1 - R_i^2 = \frac{1}{VIF} \quad (7)$$

โดยที่ R_i^2 แทน สัมประสิทธิ์การกำหนด (Coefficient of Determination) ของการถดถอยของตัวแปรอิสระ X_i

2.1.6 การตรวจสอบความเหมาะสมของตัวแบบการถดถอยลอจิสติกทวิภาค

การตรวจสอบความเหมาะสมของตัวแบบการถดถอยลอจิสติกทวิภาค มี 2 วิธี คือการทดสอบอัตราส่วนภาวะน่าจะเป็น (Likelihood Ratio Test) (ยูทธ ไกยวรรณ, 2555) และเกณฑ์สารสนเทศของอะกะอิเกะ (AIC)

1) การทดสอบอัตราส่วนภาวะน่าจะเป็น (Likelihood Ratio Test)

เนื่องจากฟังก์ชันภาวะน่าจะเป็น (L) มีค่าระหว่าง 0 ถึง 1 ดังนั้นจึงศึกษาค่า $-2LL$ ($-2 \text{ Log Likelihood}$) ซึ่งเป็นค่ามาจากล็อกภาวะน่าจะเป็น (Log Likelihood) ที่คูณด้วย -2 เพื่อต้องการให้ค่าที่ได้มีการแจกแจงที่มีลักษณะเป็นการแจกแจงไคกำลังสอง โดยจะใช้ค่า $-2LL$ วัดความเหมาะสมของตัวแบบการถดถอยลอจิสติกทวิภาค ถ้าตัวแบบการถดถอยลอจิสติกทวิภาคมีความเหมาะสม ค่า $-2LL$ จะมีค่าไม่มาก

สำหรับการทดสอบสมมติฐาน

$$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$$

$$H_1: \text{มี } \beta_j \neq 0 \text{ อย่างน้อย 1 ค่า } ; j = 1, 2, \dots, k$$

สถิติทดสอบอัตราส่วนภาวะน่าจะเป็น เป็นอัตราส่วนของค่าที่ทำให้ฟังก์ชันภาวะน่าจะเป็นสูงสุดเมื่อมีตัวแปรอิสระ p ตัว (L_1) กับค่าที่ทำให้ฟังก์ชันภาวะน่าจะเป็นสูงสุด เมื่อมีเฉพาะค่าคงที่ (L_0) มีค่ามากที่สุด

$$\begin{aligned} \text{สถิติทดสอบอัตราส่วนภาวะน่าจะเป็น} &= -2 \log \left(\frac{L_0}{L_1} \right) \\ &= -2 [\log(L_0) - \log(L_1)] \\ &= -2 [LL(0) - LL(1)] \end{aligned} \quad (8)$$

ค่าสถิติทดสอบอัตราส่วนภาวะน่าจะเป็นคือค่า $-2LL$ ที่เปลี่ยนแปลงไป ซึ่งมาจากผลต่างของค่า $-2LL(0)$ และ $-2LL(X_{i1}, X_{i2}, \dots, X_{ik})$ หมายถึง การเปลี่ยนแปลงของค่า $-2LL$ โดยที่ $-2LL(0)$ คือฟังก์ชันภาวะน่าจะเป็นสูงสุดที่มีเฉพาะค่าคงที่ในสมการ ถ้าผลต่างมีค่ามาก แสดงว่าเมื่อเพิ่มตัวแปรอิสระเข้าไปในสมการแล้ว จะทำให้ค่า $-2LL$ ลดลงอย่างมาก และจะปฏิเสธ H_0

การทดสอบตัวแบบโดยใช้วิธีอัตราส่วนภาวะน่าจะเป็น หรือ การทดสอบไคกำลังสอง (Chi-square test) เป็นค่าที่ใช้วัดการเปลี่ยนแปลงของความเหมาะสมของรูปแบบความสัมพันธ์ที่มีตัวแปรอิสระเทียบกับกรณีที่ไม่มีตัวแปรอิสระในสมการ

$$\begin{aligned} \chi^2 &= [-2LL(\beta_0)] - [-2LL(\beta_0, \beta_1, \dots, \beta_k)] \\ &= [-2LL(0)] - [-2LL(X_{i1}, X_{i2}, \dots, X_{ik})] \end{aligned} \quad (9)$$

โดยที่มีองศาเสรีเท่ากับจำนวนพารามิเตอร์ในตัวแบบ (p) จะเห็นได้ว่าเป็นสถิติทดสอบอัตราส่วนภาวะน่าจะเป็น ซึ่งเป็นค่าที่แสดงถึงผลต่างของค่าความผิดพลาดในการพยากรณ์ของตัวแบบถดถอยลอจิสติก ที่มีเฉพาะค่าคงที่ กับความผิดพลาดในการพยากรณ์ของตัวแบบถดถอยลอจิสติกที่มีตัวแปรอิสระ p ตัว

2) เกณฑ์สารสนเทศของอะกะอิเกะ (Akaike's Information Criterion (AIC))

เกณฑ์สารสนเทศของอะกะอิเกะถูกนำเสนอโดย (Akaike, 1973) เพื่อคัดเลือกตัวแบบที่ให้ค่าพยากรณ์แม่นยำที่สุด พิจารณาจากการประมาณค่าความคลาดเคลื่อนรวมกับข้อสนเทศ

(Information) ของค่าสังเกต และใช้แนวคิดจากการหาค่าต่ำสุดของข้อสนเทศด้วยหลักการคูลส์เบลิค-ไลท์เบอร์ (Kullback-Leibler Information) ข้อตกลงเบื้องต้น คือตัวประมาณได้มาจากวิธีภาวน่าจะเป็นสูงสุด และยังเป็นทางเลือกตัวแบบที่เหมาะสมภายใต้ตัวแบบที่มีการใช้ข้อมูลชุดเดียวกัน แต่มีจำนวนตัวแปรอิสระที่แตกต่างกัน ใช้อธิบายความสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระชุดหนึ่ง ตัวแบบที่เหมาะสมกว่าจะมีค่า AIC ต่ำกว่า (ภักสุภางค์ มาปรีดา, 2560)

$$AIC_k = -2\ln\hat{L}_k + 2p_k \quad (10)$$

โดยที่ \ln = ลอการิทึมธรรมชาติ

\hat{L}_k = ฟังก์ชันภาวน่าจะเป็นสูงสุด (Maximum Likelihood Function) ของตัวแบบที่ k

p_k = จำนวนตัวแปรอิสระของตัวแบบที่ k

การทดสอบความมีนัยสำคัญของสัมประสิทธิ์ถดถอย (Regression Coefficient)

ในการทดสอบสมมติฐาน

$$H_0: \beta_j = 0 \quad ; j = 1, 2, \dots, k$$

$$H_1: \beta_j \neq 0$$

การทดสอบวัตต์ (Wald test) โดยมีการแจกแจงแบบไคกำลังสองที่องศาเสรี (Degree of Freedom) เท่ากับ 1

$$Wald = \left[\frac{\beta_j}{SE(\beta_j)} \right]^2 \quad (11)$$

กรณีที่ยอมรับ H_0 แสดงว่าตัวแปรทำนาย i ไม่มีผลต่อการเปลี่ยนแปลงอัตราส่วนออกดี ดังนั้นจึงไม่มีผลต่อโอกาสการเกิดเหตุการณ์ที่สนใจ และถ้ากรณีปฏิเสธ H_0 และค่าสัมประสิทธิ์เป็นบวก แสดงว่าตัวแปรทำนายนั้นมีผลต่อการเพิ่มโอกาสการเกิดเหตุการณ์ที่สนใจ แต่ถ้าหากค่าสัมประสิทธิ์เป็นลบ แสดงว่าตัวแปรทำนายนั้นมีผลต่อการลดโอกาสการเกิดเหตุการณ์ที่สนใจ (ศิริชัย พงษ์วิชัย, 2549)

2.1.7 สถิติทดสอบความน่าเชื่อถือของตัวแบบการถดถอยลอจิสติกทวิภาค

ในตัวแบบถดถอยลอจิสติกทวิภาค การวัดระดับความสัมพันธ์ระหว่างตัวแปรตาม กับตัวแปรอิสระ จะใช้ค่าสัมประสิทธิ์การกำหนด (R^2) ซึ่งมีสถิติที่ทดสอบระดับความสัมพันธ์หลายค่า ได้แก่ ค็อกซ์และสเนล (Cox & Snell R^2) และ นาเจลเคอร์คิ (Nagelkerke R^2) แสดงรายละเอียดดังนี้ (ยุทธ ไกยวรรณ, 2555)

1) สถิติทดสอบค็อกซ์และสเนล (Cox & Snell R^2)

สถิติทดสอบค็อกซ์และสเนลเป็นการตรวจสอบความสอดคล้องของตัวแบบ หรือเปอร์เซ็นต์ที่สามารถอธิบายความผันแปรในการถดถอยลอจิสติกทวิภาค ซึ่งค็อกซ์และสเนล มีค่าน้อยกว่า 1 เสมอ

$$R_{cs}^2 = 1 - \frac{[LL(0)]^2}{[LL(1)]^2} \quad (12)$$

โดยที่ $LL(0)$ = ฟังก์ชันภาวะน่าจะเป็นของฟังก์ชันที่มีเฉพาะค่าคงที่

$LL(1)$ = ฟังก์ชันภาวะน่าจะเป็นของฟังก์ชันที่มีตัวแปรอิสระที่กำหนด

2) สถิติทดสอบนาเจลเคอร์คิ (Nagelkerke R^2)

สถิติทดสอบนาเจลเคอร์คิมีลักษณะเหมือนกับค็อกซ์และสเนล แต่จะมีค่ามากกว่าเสมอ ค่าสัมประสิทธิ์การกำหนด (R^2) จากค็อกซ์และสเนล และนาเจลเคอร์คิสามารถคิดเป็นเปอร์เซ็นต์ได้ จากการนำไปคูณด้วย 100 เป็นค่าเปอร์เซ็นต์ที่อธิบายความผันแปรในการถดถอยลอจิสติกทวิภาค

$$R_N^2 = \frac{R_{cs}^2}{1 - [LL(0)]^2} \quad (13)$$

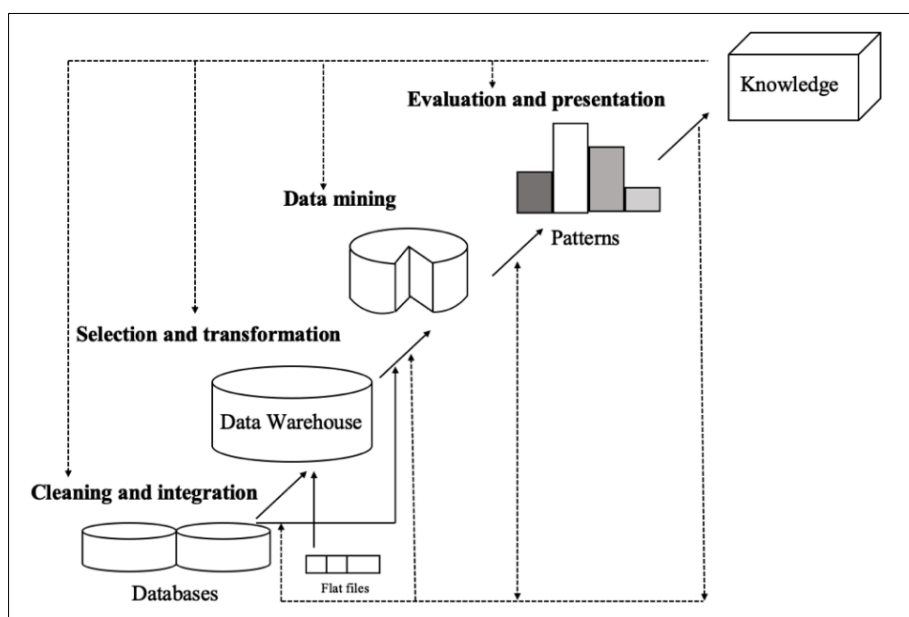
โดยที่ $LL(0)$ = ฟังก์ชันภาวะน่าจะเป็นของฟังก์ชันที่มีเฉพาะค่าคงที่

2.2 ทฤษฎีการทำเหมืองข้อมูล (Data Mining)

การทำเหมืองข้อมูล คือ กระบวนการวิเคราะห์ข้อมูลขนาดใหญ่ เพื่อค้นหารูปแบบและความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น ในปัจจุบันการทำเหมืองข้อมูลส่วนใหญ่จะสามารถนำไปใช้ในหลากหลายสาขา เช่น ด้านการแพทย์ เพื่อความรวดเร็วในการวินิจฉัยโรค อีกทั้งในด้านธุรกิจ เพื่อช่วยในการตัดสินใจของนักลงทุน การทำเหมืองข้อมูลเปรียบเสมือนวิวัฒนาการหนึ่งในการจัดเก็บและตีความหมายข้อมูล จากเดิมที่มีการจัดเก็บข้อมูลอย่างง่ายมาสู่การจัดเก็บในรูปแบบข้อมูลที่สามารถดึงข้อมูลสารสนเทศมาใช้ จนถึงการทำเหมืองข้อมูลที่สามารถค้นพบความรู้ที่ซ่อนอยู่ในข้อมูล

2.2.1 กระบวนการของการทำเหมืองข้อมูล

กระบวนการของการทำเหมืองข้อมูล เป็นกระบวนการในการค้นหาลักษณะแฝงของข้อมูล (Pattern) ที่ซ่อนอยู่ในฐานข้อมูลโดยมีขั้นตอนดังภาพ 2



ภาพ 2 กระบวนการของการทำเหมืองข้อมูล

ที่มา : (Han et al., 2012)

ขั้นตอนของการทำเหมืองข้อมูล มี 6 ขั้นตอนดังนี้

1. การกรองข้อมูล (Data Cleaning) เป็นขั้นตอนการคัดข้อมูลที่ไม่เกี่ยวข้องออก เนื่องจากข้อมูลมีความผิดปกติ เช่น ข้อมูลขาดหาย (Missing Value) ข้อมูลรบกวน (Noisy Data) ข้อมูลมีค่าผิดพลาด (Error) หรือมีค่าผิดปกติ (Outliers)
2. การรวบรวมข้อมูล (Data Integration) เป็นขั้นตอนการรวมข้อมูลจากหลายแหล่งให้มาอยู่ในข้อมูลชุดเดียวกัน เพื่อลดความซ้ำซ้อน และความไม่สอดคล้องของข้อมูล เช่น ข้อมูลที่มีอยู่ในคลังข้อมูล (Data Warehouse) ในรูปแบบของดาต้าคิวบ์ (Data Cube) และข้อมูลที่มีในรูปแบบฐานข้อมูลเชิงสัมพันธ์ (Relational Database) จำเป็นต้องทำการรวมข้อมูลให้เป็นข้อมูลชุดเดียวกัน
3. การคัดเลือกข้อมูล (Data Selection) เป็นขั้นตอนที่ระบุถึงแหล่งข้อมูลที่จะนำมาใช้ในการทำเหมืองข้อมูล รวมถึงการนำข้อมูลที่ต้องการออกจากฐานข้อมูล เพื่อสร้างกลุ่มข้อมูลที่จะใช้ในขั้นต่อไป
4. การแปลงข้อมูล (Data Transformation) เป็นขั้นตอนการแปลงข้อมูลให้อยู่ในรูปแบบอย่างง่าย สำหรับนำมาใช้ในการทำเหมืองข้อมูล เช่น การทำให้เป็นปกติ (Normalization)
5. การทำเหมืองข้อมูล (Data Mining) เป็นขั้นตอนการค้นหารูปแบบที่เป็นประโยชน์จากข้อมูลที่มี (Larose & Larose, 2014) โดยใช้เทคนิคต่าง ๆ ซึ่งสามารถแบ่งการวิเคราะห์เป็น 2 กลุ่มหลัก ได้แก่

5.1 การวิเคราะห์เชิงพยากรณ์ (Predictive Analytics)

การวิเคราะห์เชิงพยากรณ์เป็นกระบวนการการวิเคราะห์ข้อมูลที่มีอยู่ เพื่อพยากรณ์สิ่งที่จะเกิดขึ้นในอนาคตที่สอดคล้องกับการเรียนรู้แบบมีผู้สอน (Supervised Learning) คือ การเรียนรู้ที่ทราบเป้าหมายที่ชัดเจนที่ต้องการศึกษา โดยการนำข้อมูลที่มีอยู่มาใช้ในการพยากรณ์ข้อมูลอนาคตที่ไม่ทราบผลลัพธ์ เพื่อให้เกิดความผิดพลาดน้อยที่สุด ซึ่งแบ่งตามคุณสมบัติของข้อมูล 2 ลักษณะหลัก ดังนี้

5.1.1 การจำแนกประเภทข้อมูล (Classification) คือ กระบวนการในการค้นหาตัวแบบพยากรณ์ ที่มาจากชุดข้อมูลเรียนรู้ (Training Dataset) เพื่อนำมาใช้ในการพยากรณ์ชุดข้อมูลที่ไม่ทราบ หรือชุดข้อมูลทดสอบ (Testing Dataset) เพื่อประเมินประสิทธิภาพของตัวแบบพยากรณ์ ในชุดข้อมูลเรียนรู้ ใช้สำหรับพยากรณ์ที่มีผลลัพธ์ที่มีค่าไม่ต่อเนื่อง (Discrete Data) เทคนิคที่นิยมใช้คือ นาอิวเบย์ (Naïve Bayes) ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machines) ทฤษฎีเพื่อนบ้านใกล้ที่สุด k ตัว (K-Nearest Neighbor) ต้นไม้ตัดสินใจ และโครงข่ายประสาทเทียม เป็นต้น

5.1.2 การวิเคราะห์การถดถอย (Regression) คือ การวิเคราะห์การถดถอย เป็นวิธีทางสถิติที่นิยมใช้สำหรับการพยากรณ์ผลลัพธ์ที่มีค่าต่อเนื่อง (Continuous Data) กระบวนการสร้างตัวแบบพยากรณ์ทำเช่นเดียวกับการจำแนกประเภทข้อมูล เทคนิคที่นิยมใช้ คือ การวิเคราะห์ถดถอยเชิงเส้น (Linear Regression) และการวิเคราะห์ถดถอยแบบไม่เชิงเส้น (Nonlinear Regression) เป็นต้น

5.2 การวิเคราะห์เชิงบรรยาย (Descriptive Analytics)

เป็นกระบวนการเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) คือ การค้นหาลักษณะบางอย่างของข้อมูล ซึ่งไม่ทราบว่าข้อมูลนั้นจัดอยู่ในกลุ่มใด ใช้วิธีการวิเคราะห์ข้อมูลโดยอาศัยความเหมือน (Similarity) หรือความแตกต่าง (Dissimilarity) ของตัวแปรอิสระ เพื่อจัดข้อมูลหลัก เทคนิคที่นิยมใช้คือ เทคนิคการหาความสัมพันธ์ (Association) และเทคนิคการจัดกลุ่ม (Clustering)

5.2.1 การจัดกลุ่มข้อมูล (Clustering) คือการแบ่งข้อมูลออกเป็นกลุ่ม ข้อมูลที่มีลักษณะเหมือนกันหรือคล้ายคลึงกันจะถูกจัดให้อยู่ในกลุ่มเดียวกัน เทคนิคที่นิยมใช้ คือ การแบ่งกลุ่มข้อมูลแบบเคมีน (K-Means Clustering) และ การวิเคราะห์กลุ่มแบบขั้นตอน (Hierarchical Clustering) เป็นต้น

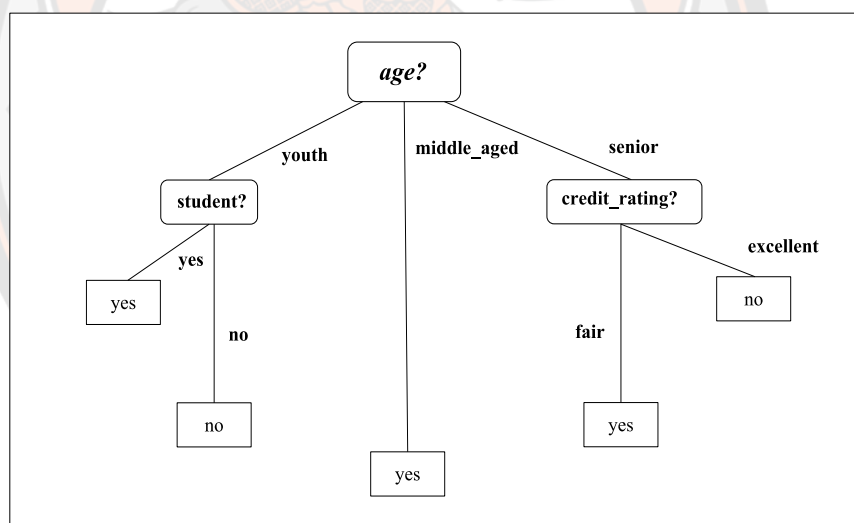
5.2.2 กฎความสัมพันธ์ (Association Rule) คือการค้นหาความสัมพันธ์ของข้อมูลในข้อมูลขนาดใหญ่ ที่เน้นความสัมพันธ์ของเหตุการณ์ในข้อมูลที่สามารถเกิดขึ้นด้วยกัน เทคนิคที่นิยมใช้คือ Apriori ในการหาความสัมพันธ์ (Association Rule Mining)

6. การประเมินผล (Pattern Evaluation) เป็นขั้นตอนที่นำเสนอองค์ความรู้ที่ได้ เพื่อวิเคราะห์ แปลความหมาย และประเมินผลว่าผลลัพธ์นั้นเหมาะสมหรือตรงวัตถุประสงค์หรือไม่

จากที่กล่าวข้างต้นว่า การทำเหมืองข้อมูลสามารถทำได้หลายรูปแบบ ในงานวิจัยเล่มนี้จะกล่าวถึงการทำเหมืองที่สัมพันธ์กับข้อมูลที่น่าสนใจ คือ วิธีการจำแนกประเภทของข้อมูลด้วยวิธีต้นไม้ตัดสินใจ และวิธีโครงข่ายประสาทเทียม

2.3 เทคนิคต้นไม้ตัดสินใจ

เทคนิคต้นไม้ตัดสินใจ เป็นเทคนิคที่ใช้การเรียนรู้แบบมีผู้สอน (Supervised Learning) และเป็นเทคนิคการจำแนกประเภทข้อมูลที่มีลักษณะคล้ายต้นไม้ ประกอบไปด้วย โหนดราก (Root Node) โหนดภายใน (Internal Node) กิ่ง (Branch) และ ใบ (Leaf) สำหรับการสร้างต้นไม้ตัดสินใจ จะขึ้นอยู่กับ การคำนวณค่าความสัมพันธ์ระหว่างตัวแปรอิสระกับกลุ่มคำตอบ (Class Label) ซึ่งหลักการพื้นฐานของการสร้างต้นไม้ตัดสินใจเป็นการสร้างในลักษณะจากบนลงล่าง (Top - Down) (สายชล สีนสมบูรณ์ทอง, 2560)



ภาพ 3 ตัวอย่างต้นไม้ตัดสินใจ

ที่มา : (Han et al., 2012)

จากภาพ 3 เป็นตัวอย่างของต้นไม้ตัดสินใจที่มีผลลัพธ์ คือ yes และ no โดยมีโหนดราก คือ age ซึ่งประกอบไปด้วยกิ่ง 3 กิ่ง คือ กิ่ง youth กิ่ง middle_aged และ กิ่ง senior ในส่วนของกิ่ง youth มีโหนดภายใน คือ student ประกอบไปด้วยกิ่ง 2 กิ่ง คือ yes และ no ในกิ่งมีใบที่เป็นคำตอบ คือ yes และ no ตามลำดับ กิ่ง middle_aged มีใบที่เป็นคำตอบคือ yes และส่วนสุดท้ายกิ่ง senior

มีโหนดภายในคือ `credit_rating` ประกอบไปด้วยกิ่ง 2 กิ่ง คือ `fair` และ `excellent` ในกิ่งมีใบที่เป็นคำตอบคือ `yes` และ `no` ตามลำดับ

2.3.1 ขั้นตอนการสร้างต้นไม้ตัดสินใจ

1. เริ่มต้นสร้างโหนดเพียงโหนดเดียวที่แสดงถึงชุดข้อมูล ถ้าภายในชุดข้อมูลมีตัวแปรตามเป็นตัวเดียวกันทั้งหมด ให้โหนดที่สร้างขึ้นมาเป็นใบ และกำหนดตัวแปรในใบด้วยตัวแปรตามนั้น
2. ถ้าภายในชุดข้อมูลมีหลายตัวแปร จะทำการเลือกตัวแปรอิสระที่มีความเหมาะสมที่สุดในการจำแนกชุดข้อมูล โดยวัดจากค่าเกน (Gain) ของแต่ละตัวแปรอิสระนั้น และให้โหนดที่สร้างขึ้นมาเป็นโหนดราก และกำหนดตัวแปรในโหนดรากด้วยตัวแปรอิสระนั้น
3. สร้างกิ่งออกมาจากโหนดรากด้วยค่าต่าง ๆ ที่เป็นไปได้ของโหนดรากนั้น และจำแนกชุดข้อมูลออกตามกิ่งต่าง ๆ ที่สร้างขึ้น
4. ทำวนซ้ำ เพื่อหาตัวแปรอิสระที่มีค่าเกนมากที่สุดของข้อมูลที่ถูกแบ่งออกมาในแต่ละกิ่ง เพื่อนำตัวแปรอิสระมาสร้างโหนดตัดสินใจต่อไป โดยที่ตัวแปรอิสระที่ถูกเลือกมาเป็นโหนดแล้ว จะไม่ถูกเลือกมาอีก
5. ทำซ้ำจนกว่าจะได้ใบครบทุกกิ่งของต้นไม้

2.3.2 อัลกอริทึม ID3 (Iterative Dichotomiser 3)

ID3 เป็นอัลกอริทึมแรกที่ใช้ในการสร้างต้นไม้ตัดสินใจ ถูกคิดค้นโดย John Ross Quinlan ซึ่งจะใช้แนวคิดที่ว่าตัวแปรอิสระตัวใดสามารถจำแนกชุดข้อมูลได้ดีที่สุด (Han et al., 2012) สำหรับการเลือกตัวแปรอิสระมาเป็นโหนดของต้นไม้ตัดสินใจ จะวัดค่าเกนซึ่งคำนวณได้จากเกนสารสนเทศ (Information gain) ดังสมการ (14)

$$Info(D) = - \sum_{i=1}^m p_i \log_2(p_i) \quad (14)$$

โดยที่ D คือ ชุดข้อมูลที่สนใจ

p_i คือ ความน่าจะเป็นของจำนวนของตัวแปรตาม i ต่อจำนวนตัวแปรตามทั้งหมด

i คือ กลุ่มของตัวแปรตาม ซึ่งมีทั้งหมด m กลุ่ม

m คือ จำนวนกลุ่มทั้งหมดของตัวแปรตาม

จากนั้นหาเกนสารสนเทศของตัวแปรอิสระแต่ละตัว ดังสมการ (15)

$$Info_A(D) = \sum_{j=1}^v \frac{|D_j|}{|D|} \times Info(D_j) \quad (15)$$

โดยที่ D คือ ชุดข้อมูลที่สนใจ

D_j คือ ตัวแปรอิสระตัวที่ j

j คือ กลุ่มของค่าในตัวแปรอิสระ ซึ่งมีทั้งหมด v กลุ่ม

v คือ จำนวนกลุ่มทั้งหมดของค่าในตัวแปรอิสระ

เมื่อได้เกณฑ์สารสนเทศของข้อมูลทั้งหมด และเกณฑ์สารสนเทศของตัวแปรอิสระแล้ว ขั้นตอนต่อไป คือ หาค่าเกณฑ์ของตัวแปรอิสระแต่ละตัว ดังสมการ (16) จากนั้นจึงเลือกตัวแปรอิสระที่มีค่าเกณฑ์สูงสุดเป็นตัวจำแนกชุดข้อมูล

$$Gain(A) = Info(D) - Info_A(D) \quad (16)$$

โดยที่ D คือ ชุดข้อมูลที่สนใจ

A คือ ตัวแปรอิสระที่สนใจ

2.3.3 อัลกอริทึม J48 หรือ C4.5

J48 เป็นอัลกอริทึมที่ใช้ในการสอนต้นไม้ตัดสินใจ ที่ถูกพัฒนาต่อมาจากอัลกอริทึม ID3 หลักการทำงานของอัลกอริทึม J48 จะใช้การคำนวณหาเกณฑ์สารสนเทศจากชุดข้อมูลทั้งหมด เช่นเดียวกับอัลกอริทึม ID3 เพื่อมาหาค่าเกณฑ์ของตัวแปรอิสระแต่ละตัว เลือกตัวแปรที่มีค่าเกณฑ์สูงสุดเป็นโหนดราก แล้วจึงแตกกิ่งไปจนถึงใบ แต่จะมีส่วนที่แตกต่างจาก ID3 ตรงที่มีการแก้ไขความเอนเอียงของค่าเกณฑ์ และทำการปรับค่าเกณฑ์โดยใช้ค่าสารสนเทศของการจำแนก (Split information) ของตัวแปรอิสระแต่ละตัว ดังสมการ (17) และเมื่อได้ค่าสารสนเทศของการจำแนกแล้ว สามารถคำนวณอัตราส่วนเกณฑ์ (Gain ratio) เพื่อลดความเอนเอียง ดังสมการ (18)

$$SplitInfo(A) = \sum_{j=1}^v \frac{|D_j|}{|D|} \times \log_2 \left(\frac{|D_j|}{|D|} \right) \quad (17)$$

$$GainRatio(A) = \frac{Gain(A)}{SplitInfo(A)} \quad (18)$$

โดยที่ A คือ ตัวแปรอิสระที่สนใจ

D_j คือ ตัวแปรอิสระตัวที่ j

j คือ กลุ่มของค่าในตัวแปรอิสระ ซึ่งมีทั้งหมด v กลุ่ม

v คือ จำนวนกลุ่มทั้งหมดของค่าในตัวแปรอิสระ

2.3.4 อัลกอริทึม LMT

อัลกอริทึม LMT เป็นอัลกอริทึมที่ได้รับความนิยมสำหรับการเรียนรู้แบบมีผู้สอน ใช้ในการจำแนกประเภทข้อมูล ซึ่งอัลกอริทึมนี้ประกอบไปด้วย 2 เทคนิค คือ เทคนิคต้นไม้ตัดสินใจ และเทคนิคการถดถอยลอจิสติก โดยกระบวนการของอัลกอริทึมนี้จะสร้างตัวแบบจากการถดถอยลอจิสติกและใช้อัลกอริทึมลอจิสตบัส (LogitBoost Algorithm) หลังจากสร้างตัวแบบจากการถดถอย

ลอจิสติกส์เสร็จแล้วก็จะใช้การตรวจสอบข้าม (Cross Validation) เพื่อกำหนดการทำซ้ำ และหยุดการ
สร้างตัวแบบพยากรณ์ สามารถเขียนได้ดังสมการ (19)

$$F_j(x) = \beta_0 + \sum_{i=1}^T \beta_i x_i \quad (19)$$

โดยที่ $F_j(x)$ คือ ค่าผลลัพธ์ที่สามารถจัดให้อยู่ในกลุ่มคำตอบได้

T คือ จำนวนตัวแปรอิสระทั้งหมด

$$p(j|x) = \frac{e^{F_j(x)}}{\sum_{k=1}^j e^{F_k(x)}} \quad , \sum_{k=1}^j F_k(x) = 0 \quad (20)$$

โดยที่ $p(j|x)$ คือ ความน่าจะเป็นก่อน (Prior Probability) มีผลรวมอยู่ในช่วง $[0,1]$ (Maulana & Defriani, 2020)

ตาราง 3 แสดงตัวอย่างข้อมูล

RID	age	income	student	credit_rating	Class: buys_computer
1	youth	high	no	fair	no
2	youth	high	no	excellent	no
3	middle_aged	high	no	fair	yes
4	senior	medium	no	fair	yes
5	senior	low	yes	fair	yes
6	senior	low	yes	excellent	no
7	middle_aged	low	yes	excellent	yes
8	youth	medium	no	fair	no
9	youth	low	yes	fair	yes
10	senior	medium	yes	fair	yes
11	youth	medium	yes	excellent	yes
12	middle_aged	medium	no	excellent	yes
13	middle_aged	high	yes	fair	yes
14	senior	medium	no	excellent	no

ที่มา : การสร้างต้นไม้ตัดสินใจโดยอัลกอริทึม J48 หรือ C4.5 (Han et al., 2012)

จากตาราง 3 แสดงตัวอย่างข้อมูลการตัดสินใจซื้อคอมพิวเตอร์ ซึ่งมีข้อมูลทั้งหมด 14 ชุด 4 ตัวแปรอิสระ และตัวแปรตาม ได้แก่ การตัดสินใจซื้อคอมพิวเตอร์ สามารถแสดงขั้นตอนการสร้างแบบจำลองต้นไม้ตัดสินใจ J48 หรือ C4.5 ได้ดังนี้

ขั้นตอนที่ 1 คำนวณค่าเอนโทรปีของข้อมูล โดยพิจารณาจากกลุ่มคำตอบของตัวแปรตาม จากสมการ (14)

$$Info(D) = -\frac{9}{14} \log_2\left(\frac{9}{14}\right) - \frac{5}{14} \log_2\left(\frac{5}{14}\right) = 0.940$$

ขั้นตอนที่ 2 คำนวณเอนโทรปีของข้อมูลของทุกตัวแปรอิสระ จากสมการ (15)

$$\begin{aligned} Info_{age}(D) &= \frac{5}{14} \times \left(-\frac{2}{5} \log_2\left(\frac{2}{5}\right) - \frac{3}{5} \log_2\left(\frac{3}{5}\right) \right) + \frac{4}{14} \times \left(-\frac{4}{4} \log_2\left(\frac{4}{4}\right) \right) \\ &\quad + \frac{5}{14} \times \left(-\frac{3}{5} \log_2\left(\frac{3}{5}\right) - \frac{2}{5} \log_2\left(\frac{2}{5}\right) \right) \\ &= 0.694 \end{aligned}$$

$$\begin{aligned} Info_{income}(D) &= \frac{4}{14} \times \left(-\frac{3}{4} \log_2\left(\frac{3}{4}\right) - \frac{1}{4} \log_2\left(\frac{1}{4}\right) \right) \\ &\quad + \frac{6}{14} \times \left(-\frac{4}{6} \log_2\left(\frac{4}{6}\right) - \frac{2}{6} \log_2\left(\frac{2}{6}\right) \right) \\ &\quad + \frac{4}{14} \times \left(-\frac{2}{4} \log_2\left(\frac{2}{4}\right) - \frac{2}{4} \log_2\left(\frac{2}{4}\right) \right) \\ &= 0.911 \end{aligned}$$

$$\begin{aligned} Info_{student}(D) &= \frac{7}{14} \times \left(-\frac{6}{7} \log_2\left(\frac{6}{7}\right) - \frac{1}{7} \log_2\left(\frac{1}{7}\right) \right) \\ &\quad + \frac{7}{14} \times \left(-\frac{3}{7} \log_2\left(\frac{3}{7}\right) - \frac{4}{7} \log_2\left(\frac{4}{7}\right) \right) \\ &= 0.787 \end{aligned}$$

$$\begin{aligned} Info_{credit_rating}(D) &= \frac{8}{14} \times \left(-\frac{6}{8} \log_2\left(\frac{6}{8}\right) - \frac{2}{8} \log_2\left(\frac{2}{8}\right) \right) \\ &\quad + \frac{6}{14} \times \left(-\frac{3}{6} \log_2\left(\frac{3}{6}\right) - \frac{3}{6} \log_2\left(\frac{3}{6}\right) \right) \\ &= 0.892 \end{aligned}$$

ขั้นตอนที่ 3 คำนวณค่าเอนโทรปีของทุกตัวแปรอิสระ จากสมการ (16)

$$Gain(age) = 0.940 - 0.694 = 0.246$$

$$Gain(income) = 0.940 - 0.911 = 0.029$$

$$Gain(student) = 0.940 - 0.787 = 0.151$$

$$Gain(credit) = 0.940 - 0.892 = 0.048$$

หลักการพิจารณาค่าเกณฑ์ คือ เลือกค่าเกณฑ์ที่มีค่าสูงสุดเป็นโหนดเริ่มต้นในการสร้างต้นไม้ตัดสินใจ จากขั้นตอนที่ 3 ค่าเกณฑ์สูงสุดคือ age ดังนั้นเราจะเลือก age เป็นโหนดเริ่มต้น หรือโหนดรากในการสร้างต้นไม้ตัดสินใจ ในส่วนของอัลกอริทึม ID3 จะมีหยุดการทำงานในขั้นตอนที่ 3 และทำซ้ำในกรณีเลือกโหนดถัดมา ส่วนอัลกอริทึม J48 จะเริ่มสร้างตั้งแต่ขั้นตอนที่ 1 - 3 และพิจารณาขั้นตอนต่อไป

ขั้นตอนที่ 4 คำนวณหาค่าสารสนเทศของการจำแนกของทุกตัวแปรอิสระจากสมการที่ (17)

$$\begin{aligned} \text{Split Info}(age) &= -\left(\frac{5}{14}\right) \times \log_2\left(\frac{5}{14}\right) - \left(\frac{4}{14}\right) \times \log_2\left(\frac{4}{14}\right) \\ &\quad - \left(\frac{5}{14}\right) \times \log_2\left(\frac{5}{14}\right) \\ &= 1.557 \end{aligned}$$

$$\begin{aligned} \text{Split Info}(income) &= -\left(\frac{4}{14}\right) \times \log_2\left(\frac{4}{14}\right) - \left(\frac{6}{14}\right) \times \log_2\left(\frac{6}{14}\right) \\ &\quad - \left(\frac{4}{14}\right) \times \log_2\left(\frac{4}{14}\right) \\ &= 1.564 \end{aligned}$$

$$\begin{aligned} \text{Split Info}(student) &= -\left(\frac{7}{14}\right) \times \log_2\left(\frac{7}{14}\right) - \left(\frac{7}{14}\right) \times \log_2\left(\frac{7}{14}\right) \\ &= 1.000 \end{aligned}$$

$$\begin{aligned} \text{Split Info}(credit_rating) &= -\left(\frac{8}{14}\right) \times \log_2\left(\frac{8}{14}\right) - \left(\frac{6}{14}\right) \times \log_2\left(\frac{6}{14}\right) \\ &= 0.985 \end{aligned}$$

ขั้นตอนที่ 5 คำนวณหาค่าอัตราส่วนเกณฑ์ของทุกตัวแปรอิสระจากสมการที่ (18)

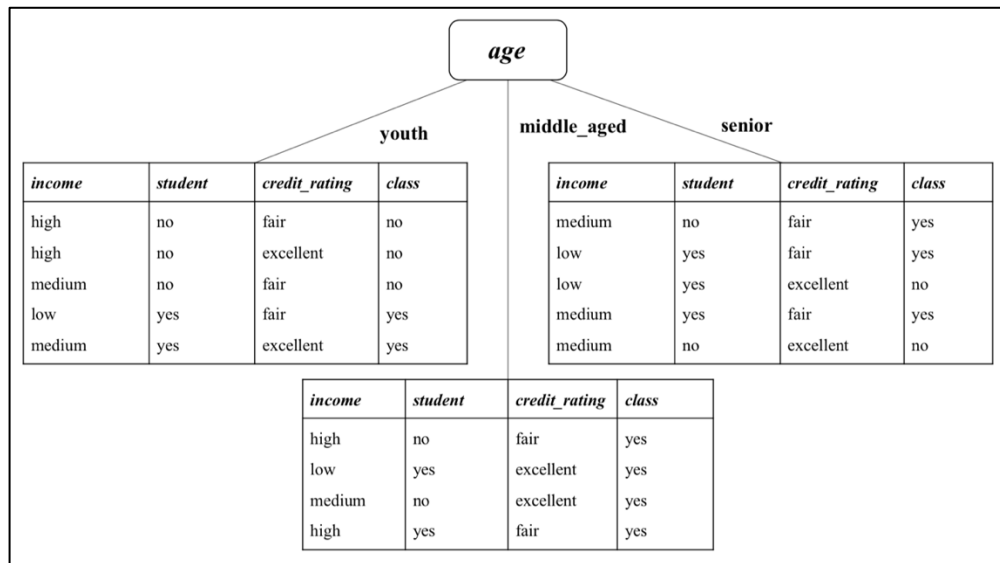
$$\text{GainRatio}(age) = \frac{0.246}{1.557} = 0.158$$

$$\text{GainRatio}(student) = \frac{0.029}{1.564} = 0.019$$

$$\text{GainRatio}(student) = \frac{0.151}{1.000} = 0.151$$

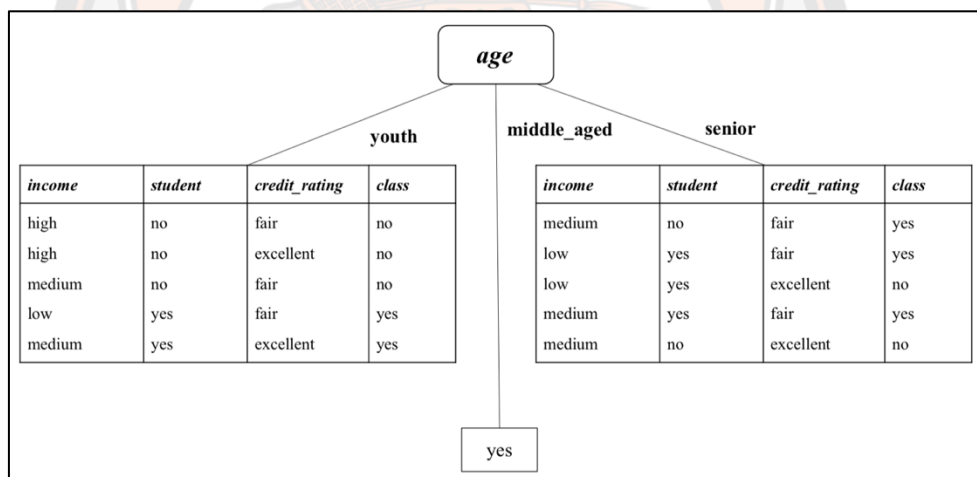
$$\text{GainRatio}(credit_rating) = \frac{0.048}{0.985} = 0.049$$

จากขั้นตอนที่ 5 ค่าอัตราส่วนเกณฑ์ที่มีค่ามากที่สุด คือ age ดังนั้นเราจะเลือก age เป็นโหนดเริ่มต้น หรือ โหนดราก ดังภาพ 4



ภาพ 4 เหตุการณ์การเลือกตัวแปรอิสระเป็นโหนดเริ่มต้น

จากภาพ 4 จะสังเกตเห็นได้ว่าโหนดภายใน middle_aged มีใบหนึ่งใบ คือ yes ดังภาพ 5



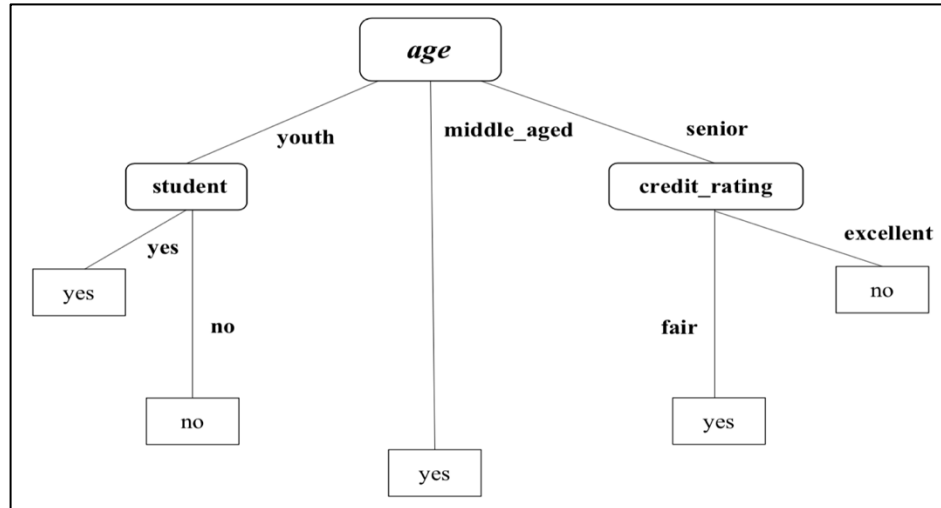
ภาพ 5 การพิจารณาโหนดภายใน middle_aged

จากภาพ 5 โหนดราก คือ age ประกอบไปด้วยโหนดภายใน คือ youth middle_aged และ senior โดยมีกิ่ง middle_aged ที่จะมีใบ คือ yes ในขณะที่กิ่งที่เหลือต้องมีการพิจารณาโหนดภายในต่อไป ขั้นตอนที่ 6 ทำซ้ำในขั้นตอนที่ 1 - 5 ในกรณีเลือกโหนดถัดมา และจะได้ต้นไม้ตัดสินใจดังภาพที่ 6 โดยมีข้อสังเกต คือ

ข้อสังเกตที่ 1 โหนดภายใน youth ตัวแปรอิสระ student ถูกเลือกเป็นโหนดภายในโดย กิ่งที่ได้คือ yes no และใบที่ได้ คือ yes และ no

ข้อสังเกตที่ 2 โหนดภายใน middle_aged ใบที่ได้ yes

ข้อสังเกตที่ 3 โหนดภายใน senior ตัวแปรอิสระ credit_rating ถูกเลือกเป็นโหนดภายใน โดยกิ่งที่ได้ คือ fair และ excellent ใบที่ได้ คือ yes และ no



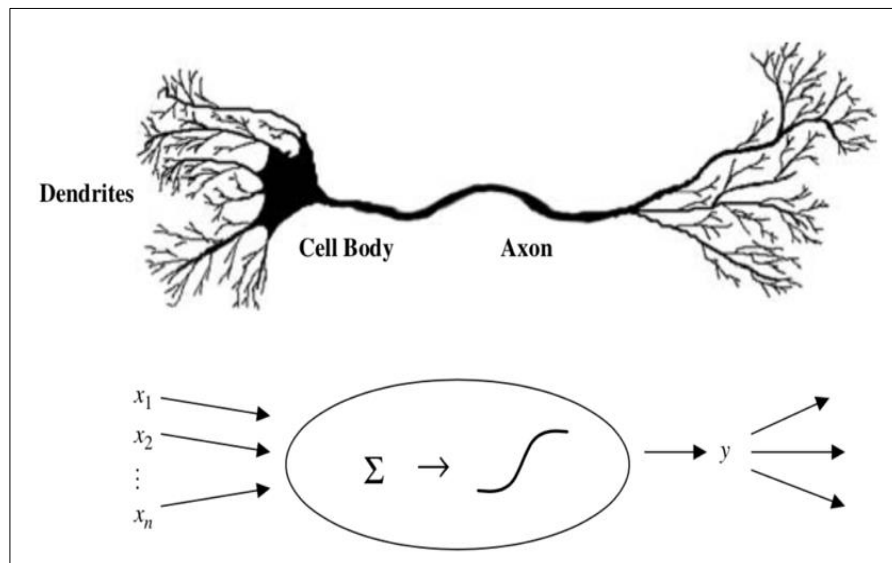
ภาพ 6 ต้นไม้ตัดสินใจ

จากต้นไม้ในภาพ 6 สามารถสร้างกฎการตัดสินใจได้ 5 กฎ รายละเอียดมีดังนี้

- | | |
|---|---------------------------|
| กฎที่ 1 ถ้า (age = youth) และ (student = yes) | แล้วซื้อคอมพิวเตอร์ = yes |
| กฎที่ 2 ถ้า (age = youth) และ (student = no) | แล้วซื้อคอมพิวเตอร์ = no |
| กฎที่ 3 ถ้า (age = middle) | แล้วซื้อคอมพิวเตอร์ = yes |
| กฎที่ 4 ถ้า (age = senior) และ (credit = fair) | แล้วซื้อคอมพิวเตอร์ = yes |
| กฎที่ 5 ถ้า (age = senior) และ (credit = excellent) | แล้วซื้อคอมพิวเตอร์ = no |

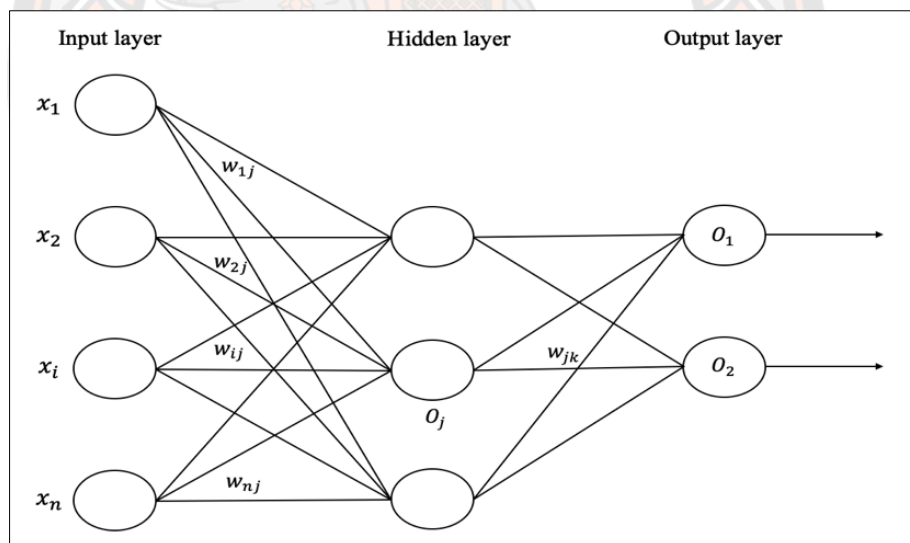
2.4 เทคนิคโครงข่ายประสาทเทียม (Artificial Neural Network)

โครงข่ายประสาทเทียม ถูกสร้างมาจากระบบการทำงานที่ซับซ้อนของเซลล์สมองมนุษย์ ประกอบด้วยเซลล์ที่เชื่อมต่อกัน หรือที่เรียกว่าเซลล์ประสาท (Neurons) เป็นจำนวนมาก ส่วนประกอบของเซลล์ประสาทที่สำคัญ ได้แก่ เดนไดรต์ (Dendrites) ทำหน้าที่รับข้อมูลจากเซลล์ประสาทอื่น ๆ ที่เชื่อมต่อกันเข้าสู่ตัวเซลล์ (Cell Body) เพื่อทำการรวมสัญญาณ (Summation) แล้วส่งข้อมูลไปยังเซลล์ประสาทอื่นผ่านทางแอกซอน (Axon) ดังภาพที่ 7



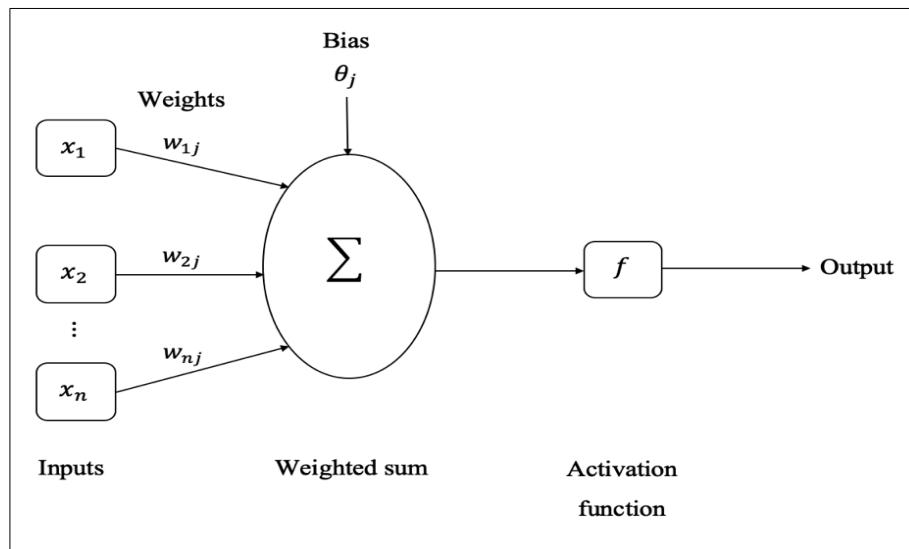
ภาพ 7 เซลล์ประสาทจริงและแบบจำลองเซลล์ประสาทเทียม

ที่มา: (Larose & Larose, 2014)



ภาพ 8 โครงข่ายประสาทเทียมหลายชั้นแบบไปข้างหน้า

จากภาพ 8 จะเห็นว่าโครงข่ายประสาทเทียมจะประกอบไปด้วย 3 ชั้น คือ ชั้นนำเข้าข้อมูล (Input Layer) ซึ่งประกอบไปด้วยเซลล์ประสาทที่มีหน้าที่รับและส่งข้อมูลไปยังชั้นถัดไป คือชั้นซ่อน (Hidden Layer) ในชั้นซ่อนนี้จะประกอบไปด้วยกี่ชั้นก็ได้ ขึ้นอยู่กับตัวแปรอิสระที่เราต้องการทดลอง เพื่อให้ได้ผลลัพธ์ที่ดีที่สุด และส่งออกมายังชั้นผลลัพธ์ (Output Layer) โดยกระบวนการการเรียนรู้ของโครงข่ายประสาทเทียม ดังภาพ 9



ภาพ 9 กระบวนการเรียนรู้ของโครงข่ายประสาทเทียม

จากภาพ 9 การหาผลลัพธ์ของโครงข่ายประสาทเทียมมี 5 องค์ประกอบดังนี้

1. นำเข้าข้อมูล (Input Layer) จากภาพ 9 คือ x_1, \dots, x_n
2. ค่าน้ำหนักถ่วง (Weights) ได้มาจากการเรียนรู้ของโครงข่ายประสาทเทียม แล้วทำการปรับค่าให้มีความเหมาะสม จากภาพ 9 คือ w_{1j}, \dots, w_{nj}
3. ฟังก์ชันผลรวม (Combination Function) ได้มาจากผลรวมของข้อมูลนำเข้า และค่าน้ำหนักถ่วงจากภาพ 9 โดยที่ Σ จะทำหน้าที่รวมผลคูณของตัวแปรอิสระและน้ำหนักถ่วงแต่ละตัวรวมทั้งความเอนเอียงหรือน้ำหนักถ่วง (Bias; b) ของแต่ละโหนดไว้ในฟังก์ชันผลรวม
4. ฟังก์ชันกระตุ้น (Activation Function) สำหรับการหาผลลัพธ์ของโครงข่ายประสาทเทียมจะใช้ฟังก์ชันซิกมอยด์ (Sigmoid Function) เป็นฟังก์ชันกระตุ้นจากภาพ 9 คือ ฟังก์ชัน f
5. ผลลัพธ์ (Output) หลังเสร็จสิ้นกระบวนการเรียนรู้ของโครงข่ายประสาทเทียมจะทำการปรับค่าต่าง ๆ ให้เหมาะสมกับชุดข้อมูลแล้วจะได้ผลลัพธ์ที่ดีที่สุด ซึ่งกระบวนการดังกล่าวสามารถเขียนให้อยู่ในรูปสมการ (21)

$$y = f\left(\sum_{i=1}^d x_i w_i + b\right) \quad (21)$$

โดยทั่วไปฟังก์ชันกระตุ้นที่ใช้ขึ้นอยู่กับลักษณะของผลลัพธ์ ซึ่งงานส่วนใหญ่มีผลลัพธ์เป็นค่าต่อเนื่อง จึงใช้ฟังก์ชันซิกมอยด์ ดังสมการ (22)

$$f(x) = \frac{1}{1 + e^{-x}} \quad (22)$$

สำหรับการพยากรณ์ค่าการจำแนกนั้นคำตอบจะขึ้นอยู่กับค่าที่ตั้งค่าเทรชโฮลด์ (Threshold) ที่เหมาะสม โดยทั่วไปถ้าค่าการจำแนกมี 2 ค่า ถ้าค่าการพยากรณ์มากกว่าหรือเท่ากับ 0.5 ผลลัพธ์ที่ได้ คือ 1 และถ้าน้อยกว่า 0.5 ผลลัพธ์ที่ได้ คือ 0

2.4.1 ตัวอย่างการประยุกต์ตัวแบบโครงข่ายประสาทเทียมเพื่อการพยากรณ์

จากตัวอย่างตาราง 4 ซึ่งข้อมูลที่ใช้เป็นตัวอย่างข้อมูลที่ได้จากผู้ป่วยที่เข้ามารับการรักษาในโรงพยาบาลมหาวิทยาลัยนเรศวร โดยจำแนกผู้ป่วยออกเป็น 2 กลุ่ม และมีตัวแปรอิสระที่ประกอบไปด้วย ดัชนีมวลกาย (Body Mass Index) และ ระดับน้ำตาลในเลือด (Fasting Blood Sugar) สำหรับตัวแปรตาม กรณีตอบ 0 คือไม่เป็นโรคเบาหวาน และ 1 คือเป็นโรคเบาหวาน จะมีขั้นตอนการทำงานดังต่อไปนี้

ตาราง 4 แสดงตัวอย่างข้อมูลผู้ป่วย โดยจำแนกผู้ป่วยออกเป็น 2 กลุ่ม

n	BMI (x_1)	FBS (x_2)	Outcome (y)
1	18.70	115	1
2	26.80	114	0
3	24.50	98	0
4	33.10	128	1
5	28.30	113	0

ขั้นตอนที่ 1 เริ่มต้นจากแปลงข้อมูลให้อยู่ในรูปการทำให้เป็นปกติ จากนั้นทำการแปลงข้อมูลให้อยู่ในรูปการทำให้เป็นปกติดังสมการ (23) เพื่อให้ได้ค่าอยู่ระหว่าง 0 ถึง 1 ดังตาราง 5

$$X^* = \frac{X - \min(X)}{\max(X) - \min(X)} \quad (23)$$

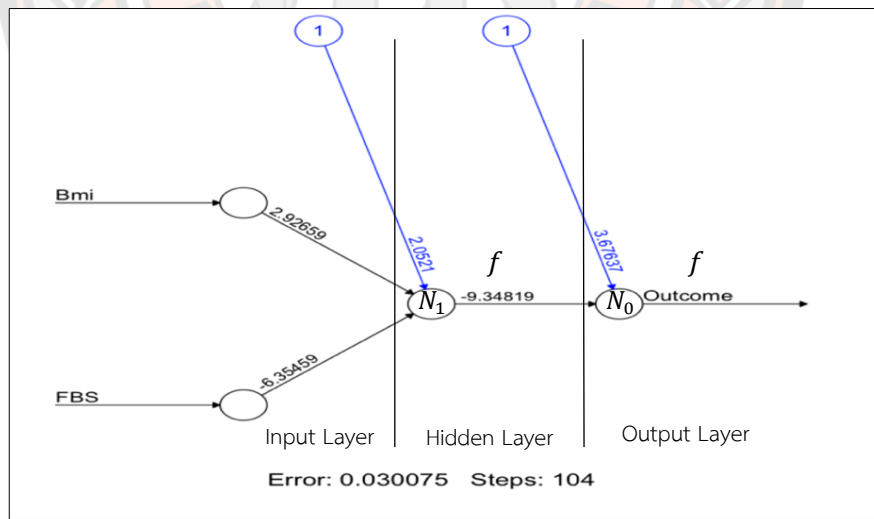
โดยที่

- X^* คือ ค่าที่ได้จากการแปลงค่าอยู่ในรูปการทำให้เป็นปกติ
- X คือ ค่าปัจจุบันที่นำมาแปลงค่าอยู่ในรูปการทำให้เป็นปกติ
- $\min(X)$ คือ ค่าข้อมูลที่มีค่าน้อยที่สุดในชุดข้อมูล
- $\max(X)$ คือ ค่าข้อมูลที่มีค่ามากที่สุดในชุดข้อมูล

ตาราง 5 แสดงตัวอย่างข้อมูลผู้ป่วยที่ทำการแปลงข้อมูล

n	BMI (x_1)	FBS (x_2)	Outcome (y)
1	$\frac{18.70 - 18.70}{33.10 - 18.70} = 0$	$\frac{115 - 98}{128 - 98} = 0.57$	1
2	$\frac{26.80 - 18.70}{33.10 - 18.70} = 0.56$	$\frac{114 - 98}{128 - 98} = 0.53$	0
3	$\frac{24.50 - 18.70}{33.10 - 18.70} = 0.40$	$\frac{98 - 98}{128 - 98} = 0$	0
4	$\frac{33.10 - 18.70}{33.10 - 18.70} = 1$	$\frac{128 - 98}{128 - 98} = 1$	1
5	$\frac{28.30 - 18.70}{33.10 - 18.70} = 0.67$	$\frac{113 - 98}{128 - 98} = 0.50$	0

ขั้นตอนที่ 2 นำข้อมูลในตาราง 5 ไปพัฒนาตัวแบบพยากรณ์ด้วยโปรแกรมอาร์ตังภาพ 10
 ขั้นตอนที่ 3 จากโปรแกรมอาร์ตังภาพ 10 จะได้อ่านค่าน้ำหนักของแต่ละเส้นเชื่อมระหว่างโหนด เมื่อนำ
 ข้อมูลหรือตัวแปรอิสระแต่ละตัวที่ได้จากการแปลงในขั้นตอนที่ 1 มาคูณกับค่าน้ำหนักและนำมารวม
 กับค่าเอนเอียงไว้ที่ฟังก์ชันผลรวม และส่งผลรวมที่ได้ไปที่ฟังก์ชันซิกมอยด์ เพื่อหาผลลัพธ์ดังสมการ
 (21)



ภาพ 10 ตัวอย่างโครงข่ายประสาทเทียมที่ได้จากโปรแกรมอาร์ตังภาพ

เมื่อนำข้อมูลชุดที่ 1 คือ $x_1 = 0$ และ $x_2 = 0.57$ มาหาผลลัพธ์ด้วยเทคนิคโครงข่าย
 ประสาทเทียมดังภาพ 10 โดยอ้างอิงจากสมการ (21) จะได้ว่า

$$\begin{aligned} \text{output}(N_1) &= f[0(2.92659) + 0.57(-6.35459) + 2.0521] \\ &= 0.1722 \end{aligned}$$

และ

$$\begin{aligned} \text{output}(N_0) &= \hat{y} = f[0.1722(-9.34819) + 3.67637] \\ &= 0.8876 \end{aligned}$$

ซึ่ง $\text{output}(N_0)$ มีค่ามากกว่า 0.5 ทำให้ค่าพยากรณ์ คือ 1

เมื่อนำข้อมูลชุดที่ 2 คือ $x_1 = 0.56$ และ $x_2 = 0.53$ โดย $y = 0$ มาพิจารณา

จะได้ว่า

$$\begin{aligned} \text{output}(N_1) &= f[0.56(2.92659) + 0.53(-6.35459) + 2.0521] \\ &= 0.5801 \end{aligned}$$

และ

$$\begin{aligned} \text{output}(N_0) &= \hat{y} = f[0.5801(-9.34819) + 3.67637] \\ &= 0.1485 \end{aligned}$$

หมายเหตุ f คือ ฟังก์ชันซิกมอยด์

เมื่อดำเนินการในลักษณะเดียวกันจนครบทุกชุดข้อมูลจะได้ผลลัพธ์ ดังตาราง 6

ตาราง 6 แสดงตัวอย่างข้อมูลผู้ป่วยที่แสดงผลลัพธ์จากตัวแบบพยากรณ์ด้วยเทคนิคโครงข่ายประสาทเทียม

n	BMI (x_1)	FBS (x_2)	Outcome (y)	Prediction	Outcome (\hat{y})
1	0	0.57	1	0.8876	1
2	0.56	0.53	0	0.1485	0
3	0.40	0	0	0.0049	0
4	1	1	1	0.8571	1
5	0.67	0.5	0	0.0550	0

จากตาราง 6 พบว่าผลลัพธ์ที่ได้จากตัวแบบพยากรณ์ด้วยเทคนิคโครงข่ายประสาทเทียม และผลลัพธ์จริงมีค่าตรงกัน

2.4.2 การหาน้ำหนักถ่วงด้วยการทำงานแบบส่งค่าย้อนกลับ

กระบวนการสำคัญของการสร้างตัวแบบพยากรณ์ด้วยเทคนิคโครงข่ายประสาทเทียม คือ การหาน้ำหนักถ่วงที่เชื่อมต่อกันระหว่างโหนด และปรับค่าความเอนเอียงของแต่ละโหนด เพื่อให้ผลลัพธ์ที่ได้จากตัวแบบพยากรณ์มีความคลาดเคลื่อนจากค่าจริงน้อยที่สุด ซึ่งวิธีที่นิยมใช้ในการหาน้ำหนักถ่วง คือการทำงานแบบการส่งค่าย้อนกลับ (Back Propagation)

หลักการหาน้ำหนักถ่วงด้วยการทำงานแบบการส่งค่าย้อนกลับ เริ่มจากการสุ่มค่าน้ำหนักถ่วง และความเอนเอียงทั้งหมดให้อยู่ในช่วง 0 ถึง 1 จากนั้นคำนวณหาผลลัพธ์ของข้อมูลชุดแรก เพื่อนำมาหาผลต่างจากค่าจริง (Prediction Error) ในสมการ (24)

$$\text{Prediction Error} = y - \hat{y} \quad (24)$$

จากสมการ (24) จะนำผลต่างที่ได้มาปรับค่าน้ำหนักถ่วงย้อนกลับจากชั้นผลลัพธ์ไปยังชั้นตัวแปรอิสระ ค่าน้ำหนักถ่วงที่ปรับใหม่จะถูกนำมาคำนวณผลลัพธ์ของข้อมูลชุดสอง แล้วหาผลต่างของผลตอบสนองกับค่าจริง

หลักการหาน้ำหนักถ่วงด้วยการทำงานแบบการส่งค่าย้อนกลับ จะมีการคำนวณเพื่อหาน้ำหนักถ่วงใหม่ทุก ๆ ชุดข้อมูล ทำเช่นนี้จนครบทุกชุดข้อมูลนับเป็น 1 รอบ (Epoch) ซึ่งการทำซ้ำหลาย ๆ รอบ จะทำให้ความคลาดเคลื่อนระหว่างค่าจริงกับค่าพยากรณ์ลดลงด้วยวิธีนี้ การเคลื่อนลงตามความชัน (Gradient Descent Method)

$$\nabla SSE(w) = \left[\frac{\partial SSE}{\partial w_1}, \frac{\partial SSE}{\partial w_2}, \dots, \frac{\partial SSE}{\partial w_n} \right] \quad (25)$$

การหาน้ำหนักถ่วงด้วยการทำงานแบบการส่งค่าย้อนกลับ มีพารามิเตอร์ที่เกี่ยวข้อง ได้แก่ อัตราการเรียนรู้ (Learning rate: η) และค่าโมเมนตัม (Momentum: α) ที่มีค่าอยู่ในช่วง [0,1] โดยการปรับน้ำหนักถ่วงสามารถคำนวณได้ (Larose, 2014) ดังสมการ (26)

$$w_{ij,new} = w_{ij,current} + \Delta w_{ij,current} + \alpha \Delta w_{ij,previous} \quad ; \quad \Delta w_{ij,current} = \eta \delta_j x_{ij} \quad (26)$$

เมื่อ

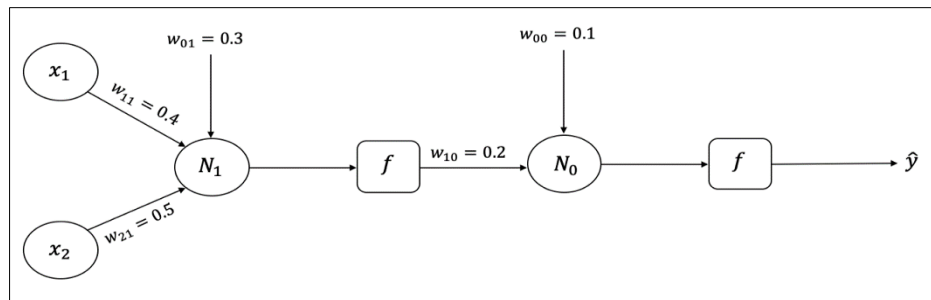
$w_{ij,new}$	คือ น้ำหนักถ่วงใหม่ของตัวแปรอิสระที่ i ไปยัง โหนด j
$w_{ij,current}$	คือ น้ำหนักถ่วงเดิมของตัวแปรอิสระที่ i ไปยัง โหนด j
η	คือ อัตราการเรียนรู้
x_{ij}	คือ ค่าตัวแปรอิสระที่ i ไปยัง โหนด j
α	คือ ค่าโมเมนตัม
$\Delta w_{ij,previous}$	คือ $\Delta w_{ij,current}$ ของจุดทดลองก่อนหน้า
δ_j	คือ ค่าความคลาดเคลื่อนของ โหนด j คำนวณได้ดังนี้
$\delta_j = \hat{y}(1 - \hat{y})(y^* - \hat{y})$	เมื่อ โหนด j อยู่ในชั้นผลตอบสนอง และ
$\delta_j = output_j(1 - output_j) \sum_{downstream} w_{jk} \delta_j$	เมื่อ โหนด j อยู่ในชั้น

ซ่อน และ โหนด k อยู่ในชั้นถัดไปทางขวา

2.4.3 ตัวอย่างการหาน้ำหนักถ่วง

การหาน้ำหนักถ่วง เริ่มต้นจากการกำหนดอัตราการเรียนรู้เท่ากับ 0.2 ค่าโมเมนตัมเท่ากับ 0.5 และค่าน้ำหนักถ่วงของ $w_{00}, w_{10}, w_{01}, w_{11}$ และ w_{21} เท่ากับ 0.1, 0.2, 0.3, 0.4 และ 0.5 ตามลำดับ และคำนวณหาน้ำหนักถ่วงที่เหมาะสมที่สุด โดยใช้สมการ (21) (22) และ (26) ตามลำดับ เมื่อใช้ข้อมูลในตาราง 5 มาคำนวณหาน้ำหนักถ่วงที่เหมาะสม สามารถทำได้ดังขั้นตอนต่อไปนี้

ข้อมูลชุดที่ 1 $x_1 = 0$ และ $x_2 = 0.57$ โดย $y = 1$



ภาพ 11 แสดงโครงสร้างของตัวแบบโครงข่ายประสาทเทียมของข้อมูลชุดที่ 1

จะได้ว่า

$$\begin{aligned} \text{output}(N_1) &= f[0(0.4) + 0.57(0.5) + 0.3] \\ &= 0.6422 \end{aligned}$$

และ

$$\begin{aligned} \text{output}(N_0) &= \hat{y} = f[0.6422(0.2) + 0.1] \\ &= 0.5569 \end{aligned}$$

ดังนั้น ความคลาดเคลื่อนของการพยากรณ์เท่ากับ

$$y - \hat{y} = 1 - 0.5569 = 0.4431$$

เนื่องจากโครงสร้างของตัวแบบโครงข่ายประสาทเทียม มีโหนดในชั้นผลลัพธ์เพียง 1 โหนดจะได้

$$\delta_0 = \hat{y}(1 - \hat{y})(y - \hat{y})$$

ในทำนองเดียวกัน โหนดในชั้นซ่อนจำนวน 1 โหนด คือ โหนด 1 และ โหนดในชั้นถัดไปทางขวามีเพียง 1 โหนด คือ โหนด 0 จะได้

$$\delta_1 = (\text{output}(N_0))(1 - \text{output}(N_0)) w_{10} \delta_0$$

ในการคำนวณเพื่อปรับค่าน้ำหนักถ่วงเริ่มต้นจากชั้นผลลัพธ์

เนื่องจาก

$$\delta_0 = \hat{y}(1 - \hat{y})(y - \hat{y})$$

จะได้

$$\delta_0 = 0.5569(1 - 0.5569)(1 - 0.5569) = 0.1093$$

และ

$$\Delta w_{00,current} = \eta \delta_0 x_{00}$$

$$\Delta w_{00,current} = (0.2)(0.1093)(1) = 0.0219$$

เนื่องจาก ข้อมูลชุดนี้เป็นข้อมูลชุดแรก
จะได้ว่า

$$\alpha \Delta w_{ij,previous} = 0$$

ดังนั้น ปรับค่า w_{00} ได้เป็น

$$w_{00,new} = w_{00,current} + \Delta w_{00,current} + \alpha \Delta w_{00,previous}$$

$$w_{00,new} = 0.1 + 0.0219 + 0 = 0.1219$$

พิจารณาน้ำหนักถ่วงในชั้นซ่อน

เนื่องจาก

$$\delta_1 = (\text{output}(N_1))(1 - \text{output}(N_1)) w_{10} \delta_0$$

จะได้

$$\delta_1 = 0.6422(1 - 0.6422)(0.2)(0.1093) = 0.0050$$

และ

$$\Delta w_{10,current} = \eta \delta_1 x_{10} \quad \text{เมื่อ } x_{10} = \text{output}(N_1)$$

$$\Delta w_{10,current} = 0.2(0.0050)(0.6422) = 0.0006$$

ดังนั้น ปรับค่า w_{10} ได้เป็น

$$w_{10,new} = w_{10,current} + \Delta w_{10,current} + \alpha \Delta w_{10,previous}$$

$$w_{10,new} = 0.2 + 0.0006 + 0 = 0.2006$$

ใช้ค่า $\delta_1 = 0.0050$ ในการปรับค่า w_{01} , w_{11} และ w_{21} ดังนี้

สำหรับ w_{01} จะได้ว่า

$$\Delta w_{01,current} = \eta \delta_1 x_0$$

$$\Delta w_{01,current} = 0.2(0.0050)(1) = 0.0010$$

ดังนั้น ปรับค่า w_{01} ได้เป็น

$$w_{01,new} = w_{01,current} + \Delta w_{01,current} + \alpha \Delta w_{01,previous}$$

$$w_{01,new} = 0.3 + 0.0010 + 0 = 0.3010$$

สำหรับ w_{11} จะได้ว่า

$$\Delta w_{11,current} = \eta \delta_1 x_1$$

$$\Delta w_{11,current} = 0.2(0.0050)(0) = 0$$

ดังนั้น ปรับค่า w_{11} ได้เป็น

$$\begin{aligned}w_{11,new} &= w_{11,current} + \Delta w_{11,current} + \alpha \Delta w_{11,previous} \\w_{11,new} &= 0.4 + 0 + 0 = 0.4\end{aligned}$$

สำหรับ w_{21} จะได้ว่า

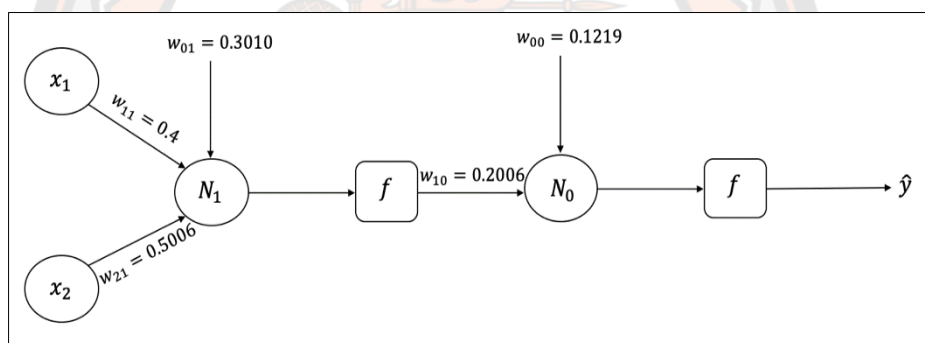
$$\begin{aligned}\Delta w_{21,current} &= \eta \delta_1 x_2 \\ \Delta w_{21,current} &= 0.2(0.0050)(0.57) = 0.0006\end{aligned}$$

ดังนั้น ปรับค่า w_{21} ได้เป็น

$$\begin{aligned}w_{21,new} &= w_{21,current} + \Delta w_{21,current} + \alpha \Delta w_{21,previous} \\w_{21,new} &= 0.5 + 0.0006 + 0 = 0.5006\end{aligned}$$

จะได้นำน้ำหนักถ่วงที่ปรับแล้วจากข้อมูลชุดที่ 1 คือ $w_{00} = 0.1219$, $w_{10} = 0.2006$, $w_{01} = 0.3010$, $w_{11} = 0.4$ และ $w_{21} = 0.5006$ แล้วนำไปใช้กับข้อมูลชุดที่ 2

ข้อมูลชุดที่ 2 $x_1 = 0.56$ และ $x_2 = 0.53$ โดย $y = 0$



ภาพ 12 แสดงโครงสร้างของตัวแบบโครงข่ายประสาทเทียมของข้อมูลชุดที่ 2

จะได้ว่า

$$\begin{aligned}output(N_1) &= f[0.56(0.4) + 0.53(0.5006) + 0.3010] \\ &= 0.6879\end{aligned}$$

และ

$$\begin{aligned}output(N_0) &= \hat{y} = f[0.6879(0.2006) + 0.1219] \\ &= 0.5646\end{aligned}$$

ดังนั้น ความคลาดเคลื่อนของการพยากรณ์ เท่ากับ

$$y - \hat{y} = 0 - 0.5646 = -0.5646$$

เนื่องจากโครงสร้างของตัวแบบโครงข่ายประสาทเทียม มีโหนดในชั้นผลลัพธ์เพียง 1 โหนดจะได้

$$\begin{aligned}\delta_0 &= \hat{y}(1 - \hat{y})(y - \hat{y}) \\ &= 0.5646(1 - 0.5646)(0 - 0.5646) = -0.1388\end{aligned}$$

และ

$$\begin{aligned}\Delta w_{00,current} &= \eta \delta_0 x_{00} \\ \Delta w_{00,current} &= (0.2)(-0.1388)(1) = -0.0278\end{aligned}$$

เนื่องจากเราได้กำหนดค่าโมเมนตัม เท่ากับ 0.5 จะได้

$$\alpha \Delta w_{00,previous} = 0.5(0.0219) = 0.0110$$

ดังนั้น ปรับค่า w_{00} ได้เป็น

$$\begin{aligned}w_{00,new} &= w_{00,current} + \Delta w_{00,current} + \alpha \Delta w_{00,previous} \\ w_{00,new} &= 0.1219 - 0.0278 + 0.0110 = 0.1051\end{aligned}$$

เมื่อพิจารณาน้ำหนักถ่วงในชั้นซ่อน
เนื่องจาก

$$\delta_1 = (\text{output}(N_1))(1 - \text{output}(N_1)) w_{10} \delta_0$$

จะได้

$$\delta_1 = 0.6879(1 - 0.6879)(0.2006)(-0.1388) = -0.0060$$

และ

$$\begin{aligned}\Delta w_{10,current} &= \eta \delta_1 x_{10} && \text{เมื่อ } x_{10} = \text{output}(N_1) \\ \Delta w_{10,current} &= 0.2(-0.0060)(0.6879) = -0.0008\end{aligned}$$

เนื่องจากเราได้กำหนดค่าโมเมนตัม เท่ากับ 0.5 จะได้

$$\alpha \Delta w_{10,previous} = 0.5(0.0006) = 0.0003$$

ดังนั้น ปรับค่า w_{10} ได้เป็น

$$\begin{aligned}w_{10,new} &= w_{10,current} + \Delta w_{10,current} + \alpha \Delta w_{10,previous} \\ w_{10,new} &= 0.2006 + (-0.0008) + 0.0003 = 0.2001\end{aligned}$$

ใช้ค่า $\delta_1 = -0.0060$ ในการปรับค่า w_{01}, w_{11} และ w_{21} ดังนี้

สำหรับ w_{01} จะได้ว่า

$$\begin{aligned}\Delta w_{01,current} &= \eta \delta_1 x_0 \\ \Delta w_{01,current} &= 0.2(-0.006)(1) = -0.0012\end{aligned}$$

เนื่องจากเราได้กำหนดค่าโมเมนตัม เท่ากับ 0.5 จะได้

$$\alpha \Delta w_{01,previous} = 0.5(0.0010) = 0.0005$$

ดังนั้น ปรับค่า w_{01} ได้เป็น

$$\begin{aligned}w_{01,new} &= w_{01,current} + \Delta w_{01,current} + \alpha \Delta w_{01,previous} \\ w_{01,new} &= 0.3010 + (-0.0012) + 0.0005 = 0.3003\end{aligned}$$

สำหรับ w_{11} จะได้ว่า

$$\Delta w_{11,current} = \eta \delta_1 x_1$$

$$\Delta w_{11,current} = 0.2(-0.0060)(0.56) = -0.0007$$

เนื่องจากเราได้กำหนดค่าโมเมนตัม เท่ากับ 0.5 จะได้

$$\alpha \Delta w_{11,previous} = 0.5(0) = 0$$

ดังนั้น ปรับค่า w_{11} ได้เป็น

$$w_{11,new} = w_{11,current} + \Delta w_{11,current} + \alpha \Delta w_{11,previous}$$

$$w_{11,new} = 0.4 + (-0.0007) + 0 = 0.3993$$

สำหรับ w_{21} จะได้ว่า

$$\Delta w_{21,current} = \eta \delta_1 x_2$$

$$\Delta w_{21,current} = 0.2(-0.0060)(0.53) = -0.0006$$

เนื่องจากเราได้กำหนดค่าโมเมนตัม เท่ากับ 0.5 จะได้

$$\alpha \Delta w_{21,previous} = 0.5(0.0006) = 0.0003$$

ดังนั้น ปรับค่า w_{21} ได้เป็น

$$w_{21,new} = w_{21,current} + \Delta w_{21,current} + \alpha \Delta w_{21,previous}$$

$$w_{21,new} = 0.5006 + (-0.0006) + 0.0003 = 0.5003$$

จะได้น้ำหนักถ่วงที่ปรับแล้วจากข้อมูลชุดที่ 2 คือ $w_{00} = 0.1051$, $w_{10} = 0.2001$, $w_{01} = 0.3003$, $w_{11} = 0.3993$ และ $w_{21} = 0.5003$ และนำน้ำหนักถ่วงใหม่ไปใช้กับข้อมูลชุดถัดไปจนครบทั้งหมด ผลลัพธ์ที่ได้แสดงดังตาราง 7

ตาราง 7 แสดงผลลัพธ์ที่ได้จากการหาน้ำหนักถ่วงจากตัวแบบพยากรณ์ด้วยเทคนิคโครงข่ายประสาทเทียม

n	BMI (x_1)	FBS (x_2)	Outcome (y)	Prediction	Outcome (\hat{y})
1	0	0.57	1	0.5489	1
2	0.56	0.53	0	0.5506	1
3	0.40	0	0	0.5470	1
4	1	1	1	0.5546	1
5	0.67	0.5	0	0.8151	1

จากตาราง 7 พบว่าผลลัพธ์ที่ได้จากการหาน้ำหนักถ่วงที่เหมาะสมและผลลัพธ์จริงมีความถูกต้องใน 3 ชุดข้อมูล สามารถสรุปได้น้ำหนักถ่วงนี้ไม่เหมาะสมกับตัวแบบพยากรณ์ด้วยเทคนิคโครงข่ายประสาทเทียม

2.5 การเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์

การเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์การเป็นโรคเบาหวาน ด้วยเทคนิคการถดถอยลอจิสติกทวิภาค เทคนิคต้นไม้ตัดสินใจ และ เทคนิคโครงข่ายประสาทเทียม ผู้วิจัยใช้ค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล เป็นเกณฑ์ที่ใช้ในการพิจารณาประสิทธิภาพตัวแบบสำหรับการพยากรณ์

ตาราง 8 แสดงตัวอย่างตารางเมทริกซ์ความสับสน (Confusion Matrix)

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)	
	Positive	Negative
Positive	<i>TP</i>	<i>FN</i>
Negative	<i>FP</i>	<i>TN</i>

โดยที่

ค่าความถูกต้องเชิงบวก (*TP*) คือ ผลลัพธ์การพยากรณ์เป็นจริง และผลลัพธ์จริงเป็นจริง
 ค่าความถูกต้องเชิงลบ (*TN*) คือ ผลลัพธ์การพยากรณ์เป็นเท็จ และผลลัพธ์จริงเป็นเท็จ
 ค่าความผิดพลาดเชิงบวก (*FP*) คือ ผลลัพธ์การพยากรณ์เป็นจริง แต่ผลลัพธ์จริงเป็นเท็จ
 ค่าความผิดพลาดเชิงลบ (*FN*) คือ ผลลัพธ์การพยากรณ์เป็นเท็จ แต่ผลลัพธ์จริงเป็นจริง
 ค่าความถูกต้อง (Accuracy) คือ ค่าที่แสดงจำนวนครั้งที่พยากรณ์ถูกต้องต่อจำนวนข้อมูล

ทั้งหมด ดังสมการ (27)

$$\text{ค่าความถูกต้อง} = \frac{TP + TN}{TP + FP + FN + TN} \quad (27)$$

ค่าระลึก (Recall) คือ จำนวนข้อมูลที่พยากรณ์ถูกต้องในกลุ่มนั้นต่อจำนวนข้อมูลเป็นจริงทั้งหมดในกลุ่มนั้น ดังสมการ (28)

$$\text{ค่าระลึก} = \frac{TP}{TP + FN} \quad (28)$$

ค่าความแม่นยำ (Precision) คือ จำนวนข้อมูลที่พยากรณ์ถูกต้องในกลุ่มนั้นต่อจำนวนข้อมูลที่พยากรณ์ว่าเป็นกลุ่มนั้นทั้งหมด ดังสมการ (29)

$$\text{ค่าความแม่นยำ} = \frac{TP}{TP + FP} \quad (29)$$

ค่าเฉลี่ยประสิทธิภาพโดยรวม (F-measure) คือ ค่าเฉลี่ยที่เกิดจากค่าระลึก และค่าความแม่นยำ ดังสมการ (30)

$$\text{ค่าเฉลี่ยประสิทธิภาพโดยรวม} = \frac{2 \times \text{Precision} \times \text{Recall}}{(\text{Precision} + \text{Recall})} \quad (30)$$

ค่าความถูกต้องในการจำแนกแบบสมดุล BCA (Balance Classification Accuracy) คือ ค่าที่แก้ไขความโน้มเอียงในการจำแนกข้อมูลของค่าความถูกต้อง เนื่องจากข้อมูลในแต่ละกลุ่มมี

ความไม่สมดุลกัน ทำให้ค่าความถูกต้องที่ได้มีความโน้มเอียงไปในกลุ่มที่มีข้อมูลมากกว่า ดังนั้นค่าความถูกต้องในการจำแนกแบบสมดุลจึงถูกนำมาใช้แก้ปัญหานี้ โดยการหาค่าเฉลี่ยการจำแนกค่าความถูกต้องในแต่ละกลุ่มดังสมการ (31) (Yang & Fong, 2011)

$$\text{ค่าความถูกต้องในการจำแนกแบบสมดุล} = \frac{1}{R} \sum_{i=1}^R \frac{\text{จำนวนครั้งที่ทำนายถูกในคลาส } i}{\text{จำนวนข้อมูลในคลาส } i} \quad (31)$$

โดยที่ R คือ จำนวนคลาสทั้งหมด

i คือ คลาสที่จะทำนาย

2.5.1 ตัวอย่างการแสดงค่าการวัดประสิทธิภาพต่าง ๆ จากข้อมูลการพยากรณ์การเป็นโรคเบาหวาน

ตาราง 9 แสดงตัวอย่างตารางเมตริกซ์ความสับสนของการพยากรณ์การเป็นโรคเบาหวาน

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)	
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน
เป็นโรคเบาหวาน	188	121
ไม่เป็นโรคเบาหวาน	157	1060

จากตาราง 9 ข้อมูลมีทั้งหมด 1,526 ชุด พบว่าการพยากรณ์ในกลุ่มที่เป็นโรคเบาหวานถูกต้อง 188 ชุด และการพยากรณ์ในกลุ่มที่เป็นไม่เป็นโรคเบาหวานถูกต้อง 1,060 ชุด

$$\begin{aligned} \text{ค่าความถูกต้อง} &= \frac{188+1,060}{188+157+121+1,060} \\ &= \frac{1,248}{1,526} \\ &= 0.817824 \end{aligned}$$

$$\begin{aligned} \text{ค่าความถูกต้องในการจำแนกแบบสมดุล} &= \frac{1}{2} \left(\left(\frac{188}{188+121} \right) + \left(\frac{1,060}{1,060+157} \right) \right) \\ &= 0.739705 \end{aligned}$$

จากตาราง 9 พบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุลคือ 0.817824 และ 0.739705 ตามลำดับ หรือ 81.7824 เปอร์เซ็นต์ และ 73.9705 เปอร์เซ็นต์ตามลำดับ

2.6 งานวิจัยที่เกี่ยวข้อง

สายชล สิ้นสมบุรณ์ทอง (2561) (สายชล สิ้นสมบุรณ์ทอง, 2561) (สายชล สิ้นสมบุรณ์ทอง, 2561) (สายชล สิ้นสมบุรณ์ทอง, 2561) ทำการศึกษาเปรียบเทียบประสิทธิภาพในการพยากรณ์การเป็นโรคเบาหวานของโรงพยาบาลแห่งหนึ่ง ข้อมูลตั้งแต่เดือนมกราคม - เมษายน พ.ศ. 2559 จำนวน 1,233 ชุด โดยใช้เทคนิคการจำแนกประเภทข้อมูล 6 เทคนิค ประกอบด้วย เทคนิคความใกล้เคียงกันมากที่สุดโดยใช้อัลกอริทึม IBk เทคนิคต้นไม้ตัดสินใจโดยใช้อัลกอริทึมชนิด J48 เทคนิคโครงข่ายประสาทเทียมโดยใช้อัลกอริทึมชนิดเพอร์เซปตรอนแบบหลายชั้น เทคนิคซัพพอร์ตเวกเตอร์แมชชีนโดยใช้อัลกอริทึม SMO ชนิดโพลีโนเมียลเคอร์เนล เทคนิคการถดถอยโลจิสติกทวิภาค และเทคนิคนาอิวเบย์ โดยพิจารณาค่าความถูกต้อง ค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) และค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) เป็นเกณฑ์ในการเปรียบเทียบประสิทธิภาพ จากผลการศึกษาพบว่าเทคนิคโครงข่ายประสาทเทียมมีประสิทธิภาพในการพยากรณ์ดีที่สุด โดยให้ค่าความถูกต้อง ค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย และค่าความคลาดเคลื่อนกำลังสองเฉลี่ย เท่ากับ 95.94 เปอร์เซ็นต์ 0.0491 และ 0.0396 ตามลำดับ

อรุณรักรักษ์ ต้นพานิช และคณะ (2562) ทำการวิเคราะห์ข้อมูล และเปรียบเทียบตัวแบบกระบวนการเรียนรู้ของเครื่อง (Machine Learning) ด้วยโปรแกรม RapidMiner Studio ข้อมูลที่ใช้ได้มาจากโรงพยาบาลส่งเสริมสุขภาพประจำตำบลท่าจีน จังหวัดสงขลา จำนวนทั้งสิ้น 300 ชุด โดยมีกลุ่มตัวอย่างจำนวน 2 กลุ่ม ได้แก่ กลุ่มผู้ป่วยโรคเบาหวานชนิดที่ 2 ที่มีภาวะซาปลายเท้า และผู้ป่วยปกติ เทคนิคที่ใช้ ได้แก่ เทคนิคซัพพอร์ตเวกเตอร์แมชชีน เทคนิคการเรียนรู้แบบอัตโนมัติด้วยการเลียนแบบการทำงานของโครงข่ายประสาทของมนุษย์ (Deep Learning) และ เทคนิคป่าสุ่ม (Random Forest) เกณฑ์ที่ใช้ในการเปรียบเทียบประสิทธิภาพ ได้แก่ ค่าความถูกต้อง ส่วนเบี่ยงเบนมาตรฐาน (SD) และระยะเวลาในการประมวลผล จากผลการศึกษาพบว่าเทคนิคซัพพอร์ตเวกเตอร์แมชชีนพยากรณ์ได้ผลลัพธ์ดีที่สุด โดยค่าความถูกต้อง ส่วนเบี่ยงเบนมาตรฐาน และระยะเวลาในการประมวลผล เท่ากับ 93.3 เปอร์เซ็นต์ 8.2 เปอร์เซ็นต์ และ 46 วินาที ตามลำดับ

Perveen et al. (2016) (Perveen et al., 2016) (Perveen et al., 2016) (Perveen et al., 2016) ศึกษาและเปรียบเทียบประสิทธิภาพด้วยเทคนิคแบบรวมกลุ่ม (Ensemble) เพื่อทำการจำแนกผู้ป่วยโรคเบาหวานได้อย่างมีประสิทธิภาพ ข้อมูลที่ใช้ได้มาจากเครือข่ายเฝ้าระวังการดูแลผู้ป่วยเบื้องต้นของประเทศแคนาดา (CPCSSN) ซึ่งแบ่งชุดข้อมูลเรียนรู้ 60 เปอร์เซ็นต์ ข้อมูลทดสอบ 40 เปอร์เซ็นต์ และทำการแบ่งกลุ่มอายุออกเป็น 3 กลุ่ม คือ 18 – 35 ปี 36 – 55 ปี และ อายุมากกว่า 55 ปีขึ้นไป เพื่อพิจารณาว่าเทคนิคใดเหมาะสมสำหรับกลุ่มอายุนั้น ๆ ซึ่งใช้เทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึมชนิด J48 สำหรับสร้างตัวแบบพยากรณ์ รวมทั้งใช้เทคนิคแบบรวมกลุ่มมาช่วยในการจำแนกผู้ป่วยโรคเบาหวานให้มีประสิทธิภาพมากขึ้น ซึ่งเทคนิคแบบรวมกลุ่มที่ใช้ คือ Adaptive

Boosting และ Bagging และใช้ค่าพื้นที่ใต้กราฟที่มีส่วนโค้ง (Area Under Receiver Operating Characteristic : AROC) สำหรับการเปรียบเทียบประสิทธิภาพ และพบว่าเทคนิคแบบรวมกลุ่มที่มีประสิทธิภาพมากที่สุด คือ Adaptive Boosting เหมาะสำหรับกลุ่มอายุ 18 – 35 ปี และ 36 – 55 ปี และเทคนิคแบบรวมกลุ่มที่มีประสิทธิภาพรองลงมา คือ เทคนิค Bagging เหมาะสำหรับอายุมากกว่า 55 ปีขึ้นไป

Kumar and Umatejaswi (2017) ศึกษาและเปรียบเทียบประสิทธิภาพการจำแนกผู้ป่วยโรคเบาหวานด้วยเทคนิคเหมืองข้อมูล สำหรับข้อมูลที่ใช้ คือ ข้อมูลผู้ป่วยโรคเบาหวานจำนวน 620 ชุด โดยแบ่งออกเป็น 3 ประเภท คือ เบาหวานประเภทที่ 1 จำนวน 115 ชุด ประเภทที่ 2 จำนวน 359 ชุด และเบาหวานขณะตั้งครรภ์ จำนวน 146 ชุด แต่ละประเภทแบ่งออกเป็น 3 ระดับ คือ ไม่รุนแรง ปานกลาง และรุนแรง และทำงานแบ่งข้อมูลแบบ 10 Fold Cross - Validation โดยเทคนิคที่ใช้คือ เทคนิคนาอิวเบย์ เทคนิค Randon Tree เทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม C4.5 และเทคนิค Simple Logistic เกณฑ์ที่ใช้ในการเปรียบเทียบประสิทธิภาพ ได้แก่ ค่าความแม่นยำ ค่าระลอก และค่าความถูกต้อง พบว่าเทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม C4.5 มีค่าความแม่นยำ ค่าระลอก และค่าความถูกต้องมากที่สุด 100 เปอร์เซ็นต์

Paul and Latha (2017) ศึกษาและสร้างตัวแบบที่สามารถพยากรณ์โอกาสการเป็นโรคเบาหวาน สำหรับข้อมูลที่ใช้ในการศึกษา คือ ข้อมูลผู้ป่วยโรคเบาหวานชาวพม่า ประเทศอินเดีย จำนวน 768 ชุด โดยใช้วิธีการคัดเลือกตัวแปรอิสระที่เป็นปัจจัยที่ส่งผลให้เป็นโรคเบาหวาน เทคนิคที่ใช้ คือ เทคนิคต้นไม้ตัดสินใจ พบว่าปัจจัยที่ทำให้ได้ตัวแบบพยากรณ์ที่เหมาะสม ได้แก่ จำนวนการตั้งครรภ์ ความดันขณะหัวใจคลายตัว ความหนาของชั้นไขมันใต้ท้องแขน ระดับน้ำตาลในเลือด และดัชนีมวลการ สำหรับเกณฑ์ที่ใช้ในการเปรียบเทียบประสิทธิภาพ ได้แก่ ค่าความถูกต้อง (Accuracy) ค่าความไว (Sensitivity) และค่าความจำเพาะ (Specificity) คือ 79.08 เปอร์เซ็นต์ 90.56 เปอร์เซ็นต์ และ 56.26 เปอร์เซ็นต์

Selvakumar et al. (2017) ศึกษาและจำแนกผู้ป่วยโรคเบาหวานโดยเทคนิคเหมืองข้อมูล ข้อมูลที่ใช้ในการศึกษามีจำนวน 100 ชุด เทคนิคที่ใช้ คือ เทคนิคการถดถอยลอจิสติกแบบทวิภาค เทคนิคโครงข่ายประสาทเทียมแบบหลายชั้น (Multilayer Perceptron) และ เทคนิคเพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbour) เกณฑ์ที่ใช้ในการเปรียบเทียบประสิทธิภาพ ได้แก่ ค่าความ ค่าความจำเพาะ และค่าความถูกต้อง และพบว่าเทคนิคเพื่อนบ้านใกล้ที่สุด มีค่าความไว ค่าจำเพาะ และค่าความถูกต้องมากที่สุด คือ 82.50 เปอร์เซ็นต์ 78.30 เปอร์เซ็นต์ และ 80 เปอร์เซ็นต์ตามลำดับ

Manimaran and Muthuraman (2017) (Manimaran & Muthuraman, 2017) (Manimaran & Muthuraman, 2017) (Manimaran & Muthuraman, 2017) ศึกษาการจำแนกประเภทข้อมูลและพยากรณ์การเป็นโรคเบาหวาน สำหรับข้อมูลที่ใช้ในการศึกษามีจำนวน 1024 ชุด

แบ่งเป็นชุดข้อมูลเรียนรู้ 70 เปอร์เซ็นต์ และชุดข้อมูลทดสอบ 30 เปอร์เซ็นต์ เทคนิคที่ใช้คือ เทคนิคโครงข่ายประสาทเทียมแบบหลายชั้น เทคนิคข่ายงานเบย์ (Bayesian networks) เทคนิค JRip เทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม C4.5 และ เทคนิค Fuzzy Lattice Reasoning เกณฑ์ที่ใช้ในการเปรียบเทียบประสิทธิภาพ ได้แก่ ค่าความถูกต้อง ค่าระลอก และอัตราความผิดพลาด (Error Rate) ผลการศึกษาพบว่าเทคนิค JRip และ เทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม C4.5 มีค่าความถูกต้องสูงสุด 86 เปอร์เซ็นต์ ในส่วนของเทคนิคข่ายงานเบย์ มีค่าระลอกสูงสุด 0.5 และอัตราความผิดพลาดอยู่ในช่วง 0.25 ถึง 0.35

Sisodia and Sisodia (2018) (Sisodia & Sisodia, 2018) (Sisodia & Sisodia, 2018) (Sisodia & Sisodia, 2018) ศึกษาและสร้างตัวแบบที่สามารถพยากรณ์โอกาสการเป็นโรคเบาหวานสำหรับข้อมูลที่ใช้ในการศึกษา คือ ข้อมูลผู้ป่วยโรคเบาหวานชาวพม่า ประเทศอินเดีย จำนวน 768 ชุด โดยแบ่งการทดสอบแบบ 10 Fold Cross-Validation เทคนิคที่ใช้คือ เทคนิคนาอิวเบย์ เทคนิคซัพพอร์ตเวกเตอร์แมชชีน และเทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม J48 เกณฑ์ที่ใช้ในการเปรียบเทียบประสิทธิภาพ ได้แก่ ค่าความแม่นยำ ค่าระลอก F-Measure ค่าความถูกต้อง และเส้นโค้งลักษณะเฉพาะดำเนินการตัวรับ (Receiver Operating Characteristic : ROC) จากการศึกษาพบว่าเทคนิคที่มีประสิทธิภาพมากที่สุด คือ เทคนิคนาอิวเบย์ มีค่าความแม่นยำ ค่าระลอก F-Measure ค่าความถูกต้อง และ ROC คือ 75.90 เปอร์เซ็นต์ 76.30 เปอร์เซ็นต์ 76.00 เปอร์เซ็นต์ และ 76.30 เปอร์เซ็นต์ 81.90 เปอร์เซ็นต์ตามลำดับ

Kumar, Singh and Singh (2019) ศึกษาและเปรียบเทียบประสิทธิภาพเกี่ยวกับการวินิจฉัยโรคเบาหวานด้วยวิธีการทำเหมืองข้อมูล สำหรับข้อมูลที่ใช้ในการศึกษา คือ ข้อมูลผู้ป่วยโรคเบาหวานชาวพม่า จำนวน 2000 ชุด เทคนิคที่ใช้ในการศึกษา ได้แก่ เทคนิคนาอิวเบย์ เทคนิคต้นไม้ตัดสินใจ เทคนิคป่าสุ่ม และเทคนิคเกรเดียนท์บูสติง (Gradient Boosting) โดยเทคนิคแต่ละแบบใช้วิธีการเรียงลำดับความสัมพันธ์ของตัวแปรอิสระที่แตกต่างกัน จากการศึกษาพบว่าเทคนิคป่าสุ่ม มีค่าความถูกต้องสูงสุดในชุดข้อมูลเรียนรู้ (Training Data) คือ 100 เปอร์เซ็นต์ และในชุดข้อมูลทดสอบ (Testing Data) คือ 99.50 เปอร์เซ็นต์

Sunge et al. (2019) ศึกษาและพัฒนาเทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม J48 เพื่อใช้ในการวินิจฉัยผู้ป่วยโรคเบาหวาน ข้อมูลที่ใช้ศึกษามีจำนวน 768 ชุด โดยทำการแบ่งชุดข้อมูลเรียนรู้ 80 เปอร์เซ็นต์ และชุดข้อมูลทดสอบ 20 เปอร์เซ็นต์ เกณฑ์ที่ใช้ในการเปรียบเทียบประสิทธิภาพ ได้แก่ ค่าความถูกต้อง ค่าระลอก ค่าความแม่นยำ ผลการศึกษาพบว่าค่าความถูกต้อง ค่าระลอก ค่าความแม่นยำ และ พื้นที่ใต้กราฟ (AUC) คือ 72.08 เปอร์เซ็นต์ 97 เปอร์เซ็นต์ 70.80 เปอร์เซ็นต์ และ 70.80 เปอร์เซ็นต์ตามลำดับ

ตาราง 10 แสดงผลสรุปเพื่อเปรียบเทียบงานวิจัยที่เกี่ยวข้องที่มีการประยุกต์ใช้ตัวแบบพยากรณ์การเป็นโรคเบาหวาน

งานวิจัย	ตัวแปรอิสระ	การแบ่งชุดข้อมูล	เทคนิคที่ใช้	เทคนิคที่ดีที่สุด
สายชล สิน สมบูรณ์ทอง (2018)	1. ค่าความดันขณะ หัวใจคลายตัว 2. ดัชนีมวลกาย 3. อายุ (ปี)	ข้อมูล 1,024 ชุด ชุดข้อมูลเรียนรู้ 70% ชุดข้อมูลทดสอบ ประสิทธิภาพ 20% ชุดข้อมูลทดสอบ 10%	1. KNN (lbc) 2. Decision Tree (J48) 3. ANN (Multi perceptron) 4. SVM (SMO) 5. Binary Logistic Regression 6. Naïve Baye	โครงข่าย ประสาทเทียม ค่าความถูกต้อง 95.94%
อรุณรักษ์ ตัน พานิช และ คณะ (2019)	1. เพศ 2. อายุ 3. น้ำหนัก 4. ส่วนสูง 5. ค่าความดันขณะ หัวใจบีบตัว 6. ดัชนีมวลกาย	ข้อมูล 300 ชุด	1. SVM 2. Deep Learning 3. Random Forest	SVM ค่าความถูกต้อง 93.3%

ตาราง 10 (ต่อ)

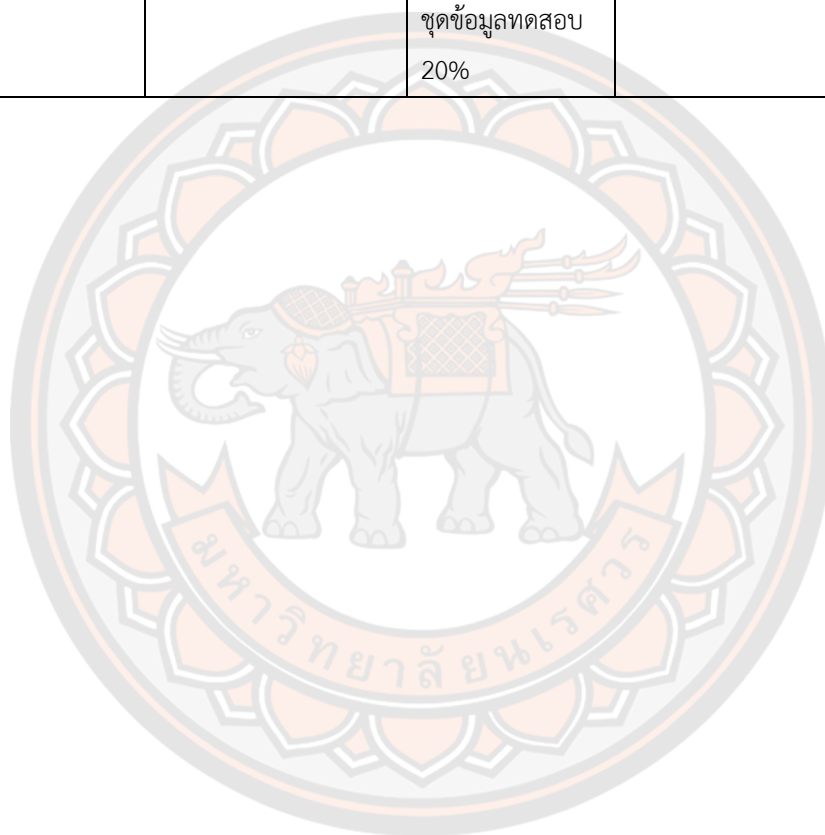
งานวิจัย	ตัวแปรอิสระ	การแบ่งชุดข้อมูล	เทคนิคที่ใช้	เทคนิคที่ดีที่สุด
Sajida Perveen, Muhammad Shanbaz, Aziz Guergachi , Karim Keshavjee (2016)	<ol style="list-style-type: none"> เพศ ค่าความดันขณะหัวใจบีบตัว ค่าความดันขณะหัวใจคลายตัว ระดับน้ำตาลในเลือด ระดับไตรกลีเซอไรด์ในเลือด ระดับไขมันในเลือดที่มีความหนาแน่นสูง ระดับไขมันในเลือดที่มีความหนาแน่นต่ำ ดัชนีมวลกาย 	<p>ข้อมูล 4,678 ชุด</p> <p>ข้อมูลเรียนรู้ 60% ข้อมูลทดสอบ 40%</p>	<ol style="list-style-type: none"> Decision Tree (J48) Decision Tree (Adaptive Boosting) Decision Tree (Bagging) 	<ul style="list-style-type: none"> ช่วงอายุ 18 – 35 ปี Adaptive Boosting ช่วงอายุ 36– 55 ปี Adaptive Boosting ช่วงอายุ มากกว่า 55 ปี Bagging
P.Suresh Kumar and V.Umatejaswi (2017)	<ol style="list-style-type: none"> เพศ ระดับน้ำตาลในเลือด ค่าความดันขณะหัวใจบีบตัว ค่าความดันขณะหัวใจคลายตัว ดัชนีมวลกาย อายุ 	<p>ข้อมูล 620 ชุด</p> <p>10 Fold Cross-validation</p>	<ol style="list-style-type: none"> Naïve Baye Random Forest Decision Tree (J48) Simple Logistic 	<p>Decision Tree (J48)</p> <p>ค่าความถูกต้อง 100%</p>
S.Paul and B.C.Latha (2017)	<ol style="list-style-type: none"> ค่าความดันขณะหัวใจคลายตัว ระดับน้ำตาลในเลือด ดัชนีมวลกาย อายุ 	<p>ข้อมูลเรียนรู้ 70% ข้อมูลทดสอบ 30%</p>	<ol style="list-style-type: none"> Decision Tree (CART) 	<p>Decision Tree (CART)</p> <p>ค่าความถูกต้อง 79.08%</p>

ตาราง 10 (ต่อ)

งานวิจัย	ตัวแปรอิสระ	การแบ่งชุดข้อมูล	เทคนิคที่ใช้	เทคนิคที่ดีที่สุด
S.Selvakumar, K.Senthamarai and S.GothaiNachiyar (2017)	1. ค่าความดันขณะ หัวใจคลายตัว 2. ระดับน้ำตาลใน เลือด 3. ดัชนีมวลกาย 4. อายุ	ข้อมูล จำนวน 100 ชุด จากเว็บ ไซด์	1. Binary Logistic Regression 2. ANN (Hyperbolic tangent) 3. KNN	KNN ค่าความถูกต้อง 80%
R.Manimaran and Dr. M. Vanitha (2017)	1. อายุ 2. เพศ	ข้อมูลจำนวน 1024 ชุด ชุดข้อมูลเรียนรู้ 712 ชุด ชุดข้อมูลทดสอบ 312 ชุด	1. ANN (Mutilayer Perceptron) 2. Baye Network 3. JRip to Ripper 4. ต้นไม้ตัดสินใจ (J48) 5. Fuzzy Lattice Reasoning	JRip to Ripper และ ต้นไม้ตัดสินใจ J48 โดยพิจารณา ค่าความถูกต้อง 86%
Deepti Sisodia, Dilip Singh Sisodia (2018)	1. ค่าความดันขณะ หัวใจคลายตัว 2. ระดับน้ำตาลในเลือด 3. ดัชนีมวลกาย 4. อายุ	ข้อมูล Pima Indians Diabetes จำนวน 768 ชุด 10 fold cross- validation	1. Naïve Baye 2. SVM 3. Decision Tree (J48)	Naïve Baye ค่าความถูกต้อง 76.30%
Chandan Kumar, Nanhay Singh, Jaspreet Singh (2019)	1. ค่าความดันขณะ หัวใจคลายตัว 2. ระดับน้ำตาลใน เลือด 3. ดัชนีมวลกาย 4. อายุ	ข้อมูลจาก 2000 ชุด ชุดข้อมูลเรียนรู้ 80% ชุดข้อมูลทดสอบ 20%	1. Naïve Baye 2. Random Forest 3. Decision Tree (J48) 4. Gradient Boosting	Random Forest ค่าความถูกต้อง 99.5%

ตาราง 10 (ต่อ)

งานวิจัย	ตัวแปรอิสระ	การแบ่งชุดข้อมูล	เทคนิคที่ใช้	เทคนิคที่ดีที่สุด
Sunge, A.S. and et al. (2019)	<ol style="list-style-type: none"> ค่าความดันขณะหัวใจคลายตัว ระดับน้ำตาลในเลือด ดัชนีมวลกาย อายุ 	ข้อมูล Pima Indians Diabetes จำนวน 768 ชุด ชุดข้อมูลเรียนรู้ 80% ชุดข้อมูลทดสอบ 20%	Decision Tree (J48)	Decision Tree (J48) ค่าความถูกต้อง 72.08%



บทที่ 3

วิธีดำเนินการวิจัย

ในบทนี้จะนำเสนอวิธีดำเนินการวิจัย เรื่องการเปรียบเทียบตัวแบบการถดถอยลอจิสติกและเทคนิคเหมืองข้อมูลสำหรับพยากรณ์การเป็นโรคเบาหวาน ข้อมูลที่ใช้ คือข้อมูลทั่วไปของผู้ป่วยที่เข้ารับการรักษาในโรงพยาบาลมหาวิทยาลัยนเรศวร เพื่อสร้างตัวแบบพยากรณ์การเป็นโรคเบาหวานโดยใช้โปรแกรมอาร์ (เวอร์ชัน 1.3.1073) และเพื่อเปรียบเทียบประสิทธิภาพของแต่ละเทคนิคที่เลือกใช้ มีขั้นตอนดังนี้

1. ข้อมูลประชากรและกลุ่มตัวอย่างที่ใช้ในการวิจัย
2. วิธีดำเนินการวิจัย
3. การสร้างตัวแบบพยากรณ์
 - 3.1 เทคนิคการถดถอยลอจิสติก
 - 3.2 เทคนิควิธีต้นไม้ตัดสินใจ
 - 3.3 เทคนิควิธีโครงข่ายประสาทเทียม
4. การประเมินประสิทธิภาพ
5. เครื่องมือที่ใช้ในการวิจัย

3.1 ข้อมูลประชากร และกลุ่มตัวอย่างในงานวิจัย

1. ประชากรที่ใช้ในการศึกษา (Population) คือ ผู้ป่วยที่เข้ารับการรักษาในโรงพยาบาลมหาวิทยาลัยนเรศวรตั้งแต่วันที่ 1 มกราคม 2563 – 30 มิถุนายน 2563 จำนวน 10,038 ชุด
2. เกณฑ์การคัดเข้า (Inclusion Criteria) คือ ข้อมูลของผู้ป่วยที่เข้ารับการรักษาในโรงพยาบาลมหาวิทยาลัยนเรศวร ประกอบไปด้วย 17 ตัวแปรอิสระ ได้แก่ เลขประจำตัวผู้ป่วย วันที่รับการรักษา เพศ อายุ ความดันขณะหัวใจบีบตัว ความดันขณะหัวใจคลายตัว อัตราการเต้นของหัวใจ น้ำหนัก ความสูง อุณหภูมิของร่างกาย ดัชนีมวลกาย ระดับน้ำตาลในเลือด ระดับคอเลสเตอรอลในเลือด ระดับไตรกลีเซอไรด์ในเลือด ระดับไขมันในเลือดที่มีความหนาแน่นสูง ระดับไขมันในเลือดที่มีความหนาแน่นต่ำ และระดับปริมาณไนโตรเจนในกระแสเลือด และตัวแปรตาม 1 ตัวแปร คือ การเป็นโรคเบาหวาน
3. เกณฑ์การคัดออก (Exclusion Criteria) ในที่นี้ข้อมูลเลขประจำตัวผู้ป่วย (patient_id) ที่ได้มาเป็นข้อมูลที่มีการเข้ารหัสจึงไม่สามารถระบุตัวตนของผู้ป่วยได้ วันที่เข้ามารับการรักษา (visit_date) และอุณหภูมิของร่างกาย (bt) เป็นปัจจัยที่ไม่เกี่ยวข้องกับการเป็นโรคเบาหวาน และพบว่าข้อมูลระดับปริมาณไนโตรเจนในกระแสเลือด (BUN) มีค่าสูญหายจำนวนมาก แสดงดังตาราง

- 11 ดังนั้นผู้วิจัยจึงตัดสินใจคัดตัวแปรอิสระเหล่านี้ ออก ทำให้ชุดข้อมูลสุดท้ายมีตัวแปรอิสระทั้งหมด 13 ตัวแปร เพื่อใช้สร้างตัวแบบพยากรณ์การเป็นโรคเบาหวาน

ตาราง 11 แสดงตัวแปรอิสระทั้งหมดของข้อมูลผู้ป่วยที่เข้ารับการรักษาในโรงพยาบาล มหาวิทยาลัยนเรศวร

patient_id	...	visit_date	...	bt	...	BUN
02510xxxxxxx	...	73-5-28	ไม่ตรวจ
02414xxxxxxx	...	58-7-6	ไม่ตรวจ
10314xxxxxxx	..	48-3-2	ไม่ตรวจ
12307xxxxxxx	...	70-7-21	ไม่ตรวจ
60113xxxxxxx	...	85-10-1	ไม่ตรวจ
91908xxxxxxx	...	49-4-16	14.4
12108xxxxxxx	...	45-10-30	...	36.5 C	...	34.1
12511xxxxxxx	...	53-0-8	20.6
42811xxxxxxx	...	80-0-30	...	36.5 C	...	26.9
83121xxxxxxx	...	66-3-4	...	36.2 C	...	23.5

3.2 วิธีการดำเนินงานวิจัย

การวิจัยครั้งนี้เป็นการศึกษาเพื่อเปรียบเทียบตัวแบบพยากรณ์การเป็นโรคเบาหวาน โดยเลือกสร้างตัวแบบพยากรณ์จาก 3 เทคนิค คือ เทคนิคการถดถอยลอจิสติกทวิภาค เทคนิคต้นไม้ตัดสินใจ และเทคนิคโครงข่ายประสาทเทียม และได้นำกระบวนการทำเหมืองข้อมูลมาเป็นแนวทางในการพัฒนาตัวแบบพยากรณ์และเปรียบเทียบประสิทธิภาพ โดยแบ่งเป็นขั้นตอนดังต่อไปนี้

3.2.1 การทำความเข้าใจกับข้อมูล (Data Understanding) จากการคัดกรองข้อมูลก่อนหน้านี้ทำให้ข้อมูลประกอบด้วยตัวแปรอิสระ 13 ตัวแปร ได้แก่ เพศ อายุ ค่าความดันขณะหัวใจบีบตัว ค่าความดันขณะหัวใจคลายตัว อัตราการเต้นของหัวใจ น้ำหนัก ความสูง ดัชนีมวลกาย ระดับน้ำตาลในเลือด ระดับคอเลสเตอรอลในเลือด ระดับไตรกลีเซอไรด์ในเลือด ระดับไขมันในเลือดที่มีความหนาแน่นสูง และ ระดับไขมันในเลือดที่มีความหนาแน่นต่ำ และตัวแปรตาม คือ การเป็นโรคเบาหวาน โดยตัวแปรอิสระแต่ละตัวมีความหมาย มาตรฐาน และหน่วยวัดดังตาราง 12

ตาราง 12 แสดงรายละเอียดข้อมูลผู้ป่วยโรคเบาหวาน ตัวแปรอิสระ (X) ตัวแปรตาม (Y)

ตัวแปร	ความหมาย	มาตรวัด	หน่วย
X_1	เพศ (Gender)	นามบัญญัติ	ชาย, หญิง (Male, Female)
X_2	อายุ (Age)	เชิงตัวเลข	ปี (year)
X_3	ค่าความดันขณะหัวใจบีบตัว (Systolic Blood Pressure)	เชิงตัวเลข	มิลลิเมตรปรอท (mmHg.)
X_4	ค่าความดันขณะหัวใจคลายตัว (Diastolic Blood Pressure)	เชิงตัวเลข	มิลลิเมตรปรอท (mmHg.)
X_5	อัตราการเต้นของหัวใจ (Heart Rate)	เชิงตัวเลข	มิลลิเมตรปรอท (mmHg.)
X_6	น้ำหนัก (Body Weight)	เชิงตัวเลข	กิโลกรัม (kg.)
X_7	ความสูง (Height)	เชิงตัวเลข	เซนติเมตร (cm.)
X_8	ดัชนีมวลกาย (Body Mass Index)	เชิงตัวเลข	กิโลกรัม/ตารางเมตร (kg./m ²)
X_9	ระดับน้ำตาลในเลือด (Fasting Blood Sugar)	เชิงตัวเลข	มิลลิกรัม/เดซิลิตร (mg/dL)
X_{10}	ระดับคอเลสเตอรอลในเลือด (Cholesterol)	เชิงตัวเลข	มิลลิกรัม/เดซิลิตร (mg/dL)
X_{11}	ระดับไตรกลีเซอไรด์ในเลือด (Triglyceride)	เชิงตัวเลข	มิลลิกรัม/เดซิลิตร (mg/dL)
X_{12}	ระดับไขมันในเลือดที่มีความหนาแน่นสูง (HDL)	เชิงตัวเลข	มิลลิกรัม/เดซิลิตร (mg/dL)
X_{13}	ระดับไขมันในเลือดที่มีความหนาแน่นต่ำ (LDL)	เชิงตัวเลข	มิลลิกรัม/เดซิลิตร (mg/dL)
Y	การเป็นโรคเบาหวาน	นามบัญญัติ	0 = ไม่เป็นโรคเบาหวาน 1 = เป็นโรคเบาหวาน

3.2.2 การเตรียมข้อมูล (Data Preprocessing) คือการทำให้ข้อมูลมีคุณภาพมากขึ้น โดยในขั้นนี้ ได้ดำเนินการดังนี้

1) การทำความสะอาดข้อมูล ในการเตรียมข้อมูลพบว่าข้อมูลบางส่วนที่มีความซ้ำซ้อน และสูญหาย ทำให้ผู้วิจัยตัดข้อมูลเหล่านี้ออก และเหลือข้อมูลทั้งสิ้น 8,994 ชุด ซึ่งแบ่งเป็นข้อมูลผู้ป่วยที่ไม่เป็นโรคเบาหวานจำนวน 7,847 ชุด และข้อมูลผู้ป่วยที่เป็นโรคเบาหวานจำนวน 1,147 ชุด

2) การทำข้อมูลให้สมดุล จากข้อมูลจำนวน 8,994 ชุด พบว่าอัตราส่วนของข้อมูลผู้ป่วยที่ไม่เป็นโรคเบาหวานต่ออัตราส่วนของข้อมูลผู้ป่วยที่เป็นโรคเบาหวาน คือ 6.8 : 1 ทำให้เกิดความไม่สมดุลของข้อมูลอย่างมาก ดังนั้นเพื่อลดอัตราความไม่สมดุลของข้อมูลผู้วิจัยใช้วิธีการเลือกตัวอย่างแบบใช้ความน่าจะเป็น (Probability Sampling) ซึ่งการเลือกตัวอย่างแบบใช้ความน่าจะเป็นมีหลายวิธี ในที่นี้ผู้วิจัยเลือกใช้วิธีการเลือกตัวอย่างสุ่มแบบง่ายแบบไม่คืนที่ (Simple Random Sampling without replacement) สำหรับใช้ในชุดข้อมูลผู้ป่วยที่ไม่เป็นโรคเบาหวาน โดยเลือกสุ่มมา 50 เปอร์เซ็นต์ของกลุ่ม เพื่อให้อัตราส่วนของข้อมูลผู้ป่วยที่ไม่เป็นโรคเบาหวานต่ออัตราส่วนของข้อมูลผู้ป่วยที่เป็นโรคเบาหวานมีความสมดุลมากขึ้น คือ 3.4 : 1 และทำให้ได้จำนวนข้อมูลผู้ป่วยที่ไม่เป็นโรคเบาหวาน 3,934 ชุด และข้อมูลผู้ป่วยที่เป็นโรคเบาหวาน 1,147 ชุด ดังนั้นเหลือข้อมูลสุดท้ายทั้งสิ้น 5,081 ชุด โดยข้อมูล 5,081 ชุดจะเป็นชุดข้อมูลที่ผู้วิจัยจะใช้ในการสร้างตัวแบบพยากรณ์ โดยรายละเอียดของข้อมูลเป็นดังตาราง 13

ตาราง 13 แสดงรายละเอียดข้อมูลผู้ป่วยจำนวน 5,081 ชุด โดยแยกเป็นกรณีผู้ป่วยที่ไม่เป็นโรคเบาหวาน และผู้ป่วยที่เป็นโรคเบาหวาน

ตัวแปร	ผู้ป่วยที่ไม่เป็นโรคเบาหวาน				ผู้ป่วยที่เป็นโรคเบาหวาน			
	จำนวน	Min	Max	Avg± SD	จำนวน	Min	Max	Avg± SD
เพศ	ชาย 1,713 หญิง 2,221	-	-	-	ชาย 464 หญิง 683	-	-	-
อายุ		16	100	75.10±6.07		17	96	62.00±11.84
ค่าความดันขณะหัวใจบีบตัว		68	212	141.72±19.39		95	230	138.87±18.73
ค่าความดันขณะหัวใจคลายตัว		28	134	70.86±12.34		7	111	72.72±12.40
อัตราการเต้นของหัวใจ		40	134	75.80±13.30		36	134	82.15±13.07
น้ำหนัก		29	150	60.25±11.81		36.4	165	68.93±15.10

ตาราง 13 (ต่อ)

ผู้ป่วยที่ไม่เป็นโรคเบาหวาน					ผู้ป่วยที่เป็นโรคเบาหวาน			
ตัวแปร	จำนวน	Min	Max	Avg± SD	จำนวน	Min	Max	Avg± SD
ความสูง		98	195	158.36±8.63	135	186		159.31±8.43
ดัชนีมวลกาย		13.06	93.71	23.99±4.24	15.6	67.8		27.10±5.18
ระดับน้ำตาลในเลือด		63	639	109.07±29.73	46	598		144.00±45.43
ระดับคอเลสเตอรอล ในเลือด		64	381	169.30±35.83	82	368		169.66±36.38
ระดับไตรกลีเซอไรด์ ในเลือด		31	399	132.11±60.63	37	393		145.73±69.01
ระดับไขมันในเลือดที่ มีความหนาแน่นสูง		21	143	51.98±14.24	19	112		49.63±12.36
ระดับไขมันในเลือดที่ มีความหนาแน่นต่ำ		13	291	90.90±30.90	14	261		90.89±33.12

3) การแปลงข้อมูล เป็นการแปลงข้อมูลเพื่อให้แต่ละตัวแปรอิสระมีค่าที่อยู่ในการทำให้เป็นปรกติ (Normalization) โดยในที่นี้ใช้การแปลงข้อมูลของแต่ละตัวแปรอิสระให้อยู่ในช่วง [0,1] เพื่อให้ตัวแปรอิสระแต่ละตัวมีความเท่าเทียมกัน

$$X^* = \frac{X - \min(X)}{\max(X) - \min(X)} \quad (32)$$

โดยที่

- X^* คือ ค่าที่ได้จากการแปลงค่าอยู่ในรูปการทำให้เป็นปรกติ
 X คือ ค่าปัจจุบันที่นำมาแปลงค่าอยู่ในรูปการทำให้เป็นปรกติ
 $\min(X)$ คือ ค่าข้อมูลที่มีค่าน้อยที่สุดในชุดข้อมูล
 $\max(X)$ คือ ค่าข้อมูลที่มีค่ามากที่สุดในชุดข้อมูล

ตาราง 14 แสดงการแปลงข้อมูลให้อยู่การทำให้เป็นปรกติ

เพศ	อายุ	ค่าความดัน ขณะหัวใจบีบ ตัว	ค่าความดัน ขณะหัวใจ คลายตัว	อัตราการเต้น ของหัวใจ	น้ำหนัก	ความ สูง	ดัชนีมวล กาย	ระดับ น้ำตาล ในเลือด	ระดับ คอเลสเตอรอลใน เลือด	ระดับไตรกลี เซอไรด์ใน เลือด	ระดับไขมัน ในเลือดที่มี ความ หนาแน่นสูง	ระดับไขมันใน เลือดที่มีความ หนาแน่นต่ำ	การเป็น โรคเบาหวาน
0	0.7143	0.7083	0.7097	0.1389	0.3810	0.6667	0.2882	0.3696	0.4476	0.4511	1.0000	0.2778	0
0	1.0000	0.6667	0.0323	0.1111	0.0952	0.3333	0.2213	0.5435	0.5619	0.8681	0.5000	0.3222	1
0	0.3571	0.2292	0.1613	0.8056	0.0952	0.6667	0.0724	0.7391	0.8571	0.2596	0.6250	0.9556	1
1	0.0000	0.2083	0.3226	0.0000	0.0952	0.8333	0.0078	0.6522	0.6476	0.0000	0.4583	0.8889	1
0	0.3810	0.0000	0.0323	0.1667	0.2619	1.0000	0.0604	0.1087	0.0000	0.4979	0.0000	0.0000	0
1	0.5476	1.0000	0.4516	0.4722	0.5714	0.8333	0.3462	0.0000	0.8476	1.0000	0.2083	0.6667	0
1	0.6667	0.5625	0.3871	0.1111	1.0000	0.3333	1.0000	0.4565	0.8476	0.3957	0.6667	0.8556	0
1	0.2857	0.4167	0.2258	1.0000	0.4524	0.0000	0.7561	0.5435	1.0000	0.8596	0.7917	0.7667	1
1	0.2143	0.0000	0.0000	0.7778	0.1190	0.3333	0.2416	0.7609	0.4857	0.2681	0.5417	0.5333	1
1	0.3333	0.7708	1.0000	0.7222	0.0000	0.6667	0.0000	1.0000	0.9619	0.3617	0.7083	1.0000	1

4) การวิเคราะห์ความสัมพันธ์ของตัวแปรอิสระกับตัวแปรตามด้วยสัมประสิทธิ์สหสัมพันธ์ทาวบิของเคนดัล (Kendall's Tau-b correlation coefficient) เพื่อใช้ในการพิจารณาข้อมูลเบื้องต้นว่าตัวแปรอิสระใดบ้างมีความสัมพันธ์กับตัวแปรตามแสดงดังตาราง 15 (สำราญ มีแจ้ง, 2546) ซึ่งเป็นการหาสัมประสิทธิ์สหสัมพันธ์แบบอันดับของเคนดอลส์ ดังสมการที่ (33)

$$\tau = \frac{2S}{n(n-1)} \quad (33)$$

โดยที่ τ = สัมประสิทธิ์สหสัมพันธ์แบบอันดับของเคนดอลส์

S = ความแตกต่างระหว่างผลบวกของจำนวนอันดับที่สอดคล้องกันกับผลบวกของอันดับที่ไม่สอดคล้องกัน

n = ขนาดตัวอย่าง

ตาราง 15 แสดงค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรอิสระและตัวแปรตาม

ตัวแปร	r	p - value
X_1 (เพศ)	0.026	0.063
X_2 (อายุ)	0.061	<0.001*
X_3 (ค่าความดันขณะหัวใจบีบตัว)	0.075	<0.001*
X_4 (ค่าความดันขณะหัวใจคลายตัว)	-0.039	0.001*
X_5 (อัตราการเต้นของหัวใจ)	0.116	<0.001*
X_6 (น้ำหนัก)	0.108	<0.001*
X_7 (ความสูง)	-0.034	0.003*
X_8 (ดัชนีมวลกาย)	0.144	<0.001*
X_9 (ระดับน้ำตาลในเลือด)	0.396	<0.001*
X_{10} (ระดับคอเลสเตอรอลในเลือด)	-0.108	<0.001*
X_{11} (ระดับไตรกลีเซอไรด์ในเลือด)	0.071	<0.001*
X_{12} (ระดับไขมันในเลือดที่มีความหนาแน่นสูง)	-0.094	<0.001*
X_{13} (ระดับไขมันในเลือดที่มีความหนาแน่นต่ำ)	-0.112	<0.001*

*มีนัยสำคัญเชิงสถิติ 0.05

จากตาราง 15 แสดงค่าความสัมพันธ์ของตัวแปรอิสระกับตัวแปรตาม ทดสอบด้วยสัมประสิทธิ์สหสัมพันธ์ทาวบิของเคนดัล พบว่าตัวแปรอิสระ ได้แก่ อายุ ค่าความดันขณะหัวใจบีบตัว ค่าความดันขณะหัวใจคลายตัว อัตราการเต้นของหัวใจ น้ำหนัก ความสูง ดัชนีมวลกาย ระดับน้ำตาลในเลือด ระดับคอเลสเตอรอลในเลือด ระดับไตรกลีเซอไรด์ในเลือด ระดับไขมันในเลือดที่มีความ

หนาแน่นสูง และ ระดับไขมันในเลือดที่มีความหนาแน่นต่ำ มีความสัมพันธ์กับตัวแปรตาม คือการเป็นโรคเบาหวาน ที่ระดับนัยสำคัญ 0.05

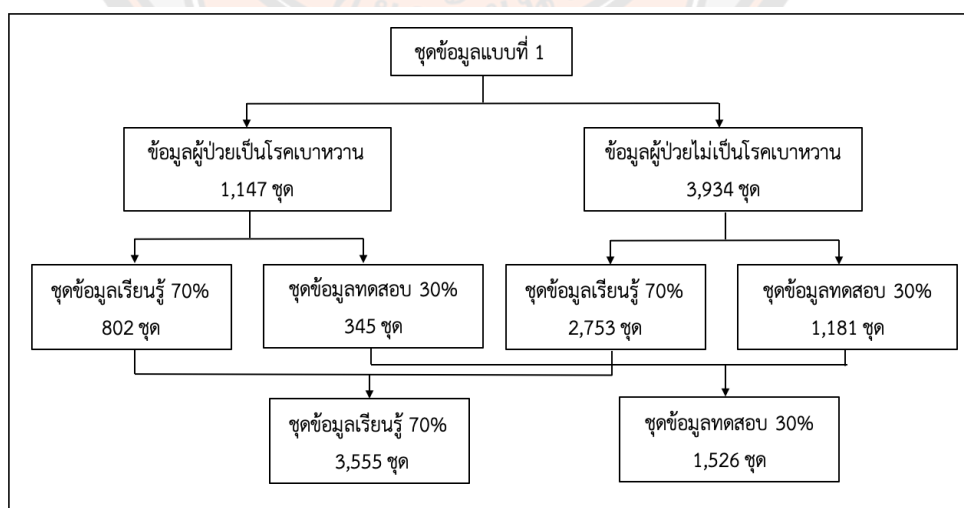
เมื่อพิจารณาที่ค่าสัมประสิทธิ์สหสัมพันธ์ (r) พบว่า ระดับน้ำตาลในเลือดเป็นปัจจัยหลักที่ส่งผลต่อการเป็นโรคเบาหวาน รองลงมาคือ ดัชนีมวลกาย และอัตราการเต้นของหัวใจตามลำดับ

3.3 การสร้างตัวแบบพยากรณ์

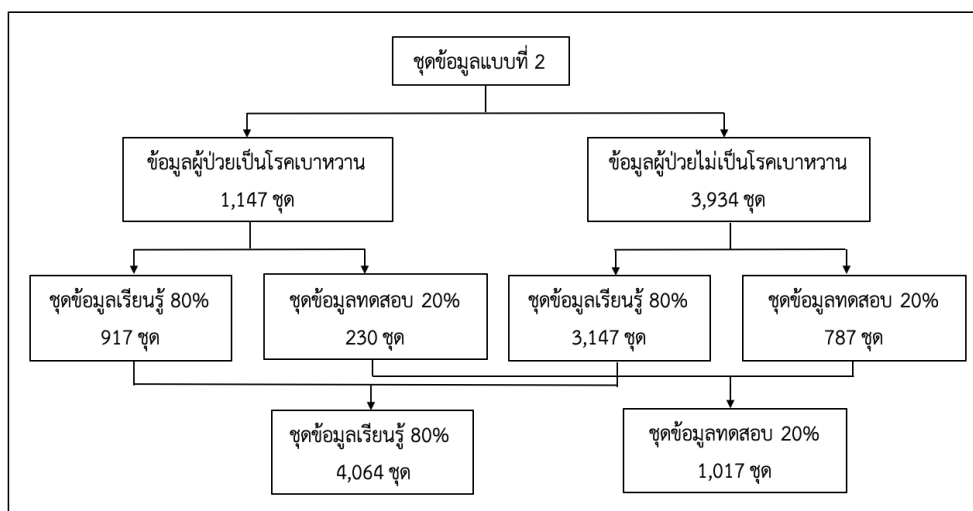
หลักการสร้างตัวแบบพยากรณ์เพื่อการเปรียบเทียบประสิทธิภาพนั้น จะต้องมีการแบ่งข้อมูลเป็นชุดเรียนรู้ และชุดทดสอบจากข้อมูลทั้งสิ้น 5,081 ชุด โดยหลักการ คือจะนำชุดเรียนรู้ไปสร้างตัวแบบพยากรณ์ และนำชุดทดสอบมาใช้วัดประสิทธิภาพของตัวแบบพยากรณ์

ในที่นี้ผู้วิจัยเลือกใช้การสุ่มเลือกข้อมูล ด้วยวิธีการสุ่มตัวอย่างแบบแบ่งชั้นภูมิ (Stratified Random Sampling) เพื่อให้มั่นใจว่าสัดส่วนของข้อมูลของการเป็นโรคเบาหวาน และการไม่เป็นโรคเบาหวานยังคงเท่าเดิมทั้งในชุดข้อมูลเรียนรู้และชุดข้อมูลทดสอบ โดยในที่นี้มีการแบ่งชุดข้อมูลเป็น 2 แบบ

1. ชุดข้อมูลแบบที่ 1 เป็นการสุ่มเลือกข้อมูล ด้วยวิธีการสุ่มตัวอย่างแบบแบ่งชั้นภูมิโดยเลือก 70 เปอร์เซ็นต์มาเป็นชุดข้อมูลเรียนรู้ และ 30 เปอร์เซ็นต์มาเป็นชุดข้อมูลทดสอบแสดงได้ดังภาพ 13
2. ชุดข้อมูลแบบที่ 2 เป็นการสุ่มเลือกข้อมูล ด้วยวิธีการสุ่มตัวอย่างแบบแบ่งชั้นภูมิโดยเลือก 80 เปอร์เซ็นต์มาเป็นชุดข้อมูลเรียนรู้ และ 20 เปอร์เซ็นต์มาเป็นชุดข้อมูลทดสอบแสดงได้ดังภาพ 14



ภาพ 13 การแบ่งชุดข้อมูลแบบที่ 1 ด้วยวิธีการสุ่มตัวอย่างแบบแบ่งชั้นภูมิด้วยสัดส่วน 70 : 30



ภาพ 14 การแบ่งชุดข้อมูลแบบที่ 2 ด้วยวิธีการสุ่มตัวอย่างแบบแบ่งชั้นภูมิด้วยสัดส่วน 80 : 20

โดยในการศึกษานี้เลือกใช้ 3 เทคนิค เพื่อพัฒนาตัวแบบพยากรณ์ ได้แก่

- 1) เทคนิคการถดถอยลอจิสติกทวิภาค
- 2) เทคนิคต้นไม้ตัดสินใจ
- 3) เทคนิคโครงข่ายประสาทเทียม

3.3.1 การสร้างตัวแบบพยากรณ์ด้วยเทคนิคการถดถอยลอจิสติกทวิภาค

ขั้นตอนที่ 1 หาอัตราส่วนออกัส หรืออัตราส่วนระหว่างโอกาสที่จะเกิดเหตุการณ์ที่สนใจต่อโอกาสไม่เกิดเหตุการณ์ที่สนใจ

ขั้นตอนที่ 2 ตรวจสอบความสัมพันธ์ระหว่างตัวแปรอิสระ

ขั้นตอนที่ 3 ตรวจสอบความเหมาะสมของตัวแบบการถดถอยลอจิสติก

ขั้นตอนที่ 4 ทดสอบความมีนัยสำคัญของสัมประสิทธิ์ถดถอย

ขั้นตอนที่ 5 ทดสอบความน่าเชื่อถือของตัวแบบพยากรณ์

ขั้นตอนที่ 6 คัดเลือกตัวแปรอิสระทั้งหมด 4 วิธี คือ วิธีการเลือกแบบพื้นฐาน วิธีการเลือกแบบไปข้างหน้า วิธีการกำจัดแบบถอยหลัง และวิธีการถดถอยทีละขั้น กำหนดระดับนัยสำคัญ 0.05

3.3.2 การสร้างตัวแบบพยากรณ์ด้วยเทคนิคต้นไม้ตัดสินใจ

สำหรับงานวิจัยนี้ผู้วิจัยกำหนดจำนวนตัวแปรอิสระเข้าสู่โครงข่าย โดยการคัดเลือกตัวแปรอิสระตามผลลัพธ์ที่ได้จากเทคนิคการถดถอยลอจิสติกทวิภาค และมีขั้นตอนการสร้างตัวแบบพยากรณ์ดังนี้

1) อัลกอริทึม J48

ขั้นตอนที่ 1 คำนวณค่าเกณฑ์สารสนเทศของตัวแปรตาม

ขั้นตอนที่ 2 คำนวณค่าเกณฑ์สารสนเทศของตัวแปรอิสระทั้งหมด

ขั้นที่ตอน 3 คำนวณค่าเกณฑ์ของตัวแปรอิสระแต่ละตัว และพิจารณาเลือกตัวแปรที่มีค่าเกณฑ์สูงสุดเป็นโหนดราก

2) อัลกอริทึม LMT (Logistic Model Tree)

ขั้นตอนที่ 1 คำนวณขั้นที่ 1 – 3 เช่นเดียวกับอัลกอริทึม J48 สำหรับการสร้างต้นไม้

ขั้นตอนที่ 2 ใช้อัลกอริทึม LogitBoost ในการแตกกิ่ง และ การตรวจสอบข้าม (Cross – Validation) เพื่อกำหนดการทำซ้ำและหยุดการสร้างตัวแบบพยากรณ์

การเลือกตัวแปรอิสระเพื่อใช้ในการพัฒนาตัวแบบพยากรณ์ เป็นปัจจัยสำคัญหนึ่งที่จะส่งผลต่อประสิทธิภาพของตัวแบบพยากรณ์ โดยในงานวิจัยนี้จะคัดเลือกตัวแปรอิสระตามผลลัพธ์ที่ได้จากการพัฒนาตัวแบบพยากรณ์การถดถอยลอจิสติกทวิภาค

3.3.3 การสร้างตัวแบบพยากรณ์ด้วยเทคนิคโครงข่ายประสาทเทียม

ตัวแบบพยากรณ์โครงข่ายประสาทเทียมมีหลักการสำคัญ คือ ต้องมีการออกแบบโครงสร้างโครงข่าย ซึ่งได้แก่ จำนวนตัวแปรเข้า จำนวนชั้นซ่อน จำนวนโหนด และจำนวนผลลัพธ์ โดยภายในโครงข่ายสามารถกำหนดฟังก์ชันกระตุ้นได้ตามความเหมาะสมของตัวแบบพยากรณ์ และสามารถกำหนดอัตราการเรียนรู้ โมเมนตัม และจำนวนรอบ เพื่อการเรียนรู้ของโครงข่าย สำหรับงานวิจัยนี้ผู้วิจัยกำหนดจำนวนตัวแปรอิสระเข้าสู่โครงข่าย โดยการคัดเลือกตัวแปรอิสระตามผลลัพธ์ที่ได้จากการเทคนิคการถดถอยลอจิสติกทวิภาค และกำหนดค่าพารามิเตอร์ของเทคนิคโครงข่ายประสาทเทียมดังนี้

- 1) ฟังก์ชันกระตุ้น คือฟังก์ชันซิกมอยด์
- 2) จำนวนชั้นซ่อน 1 ชั้น
- 3) จำนวนโหนดในชั้นซ่อน กำหนดเป็น 5, 6, 8, 9 และ 10
- 4) อัตราการเรียนรู้ คือ 0.1, 0.2 และ 0.3
- 5) ค่าโมเมนตัม คือ 0.1, 0.2 และ 0.3
- 6) จำนวนรอบ 100 รอบ

จากการกำหนดพารามิเตอร์ดังกล่าวข้างต้น ทำให้ต้องทำซ้ำ 45 ครั้ง ในการสร้างตัวแบบพยากรณ์ของแต่ละกรณีการเลือกตัวแปรอิสระ ตามการคัดเลือกตัวแปรอิสระตามผลลัพธ์ที่ได้จากการพัฒนาตัวแบบพยากรณ์จากเทคนิคการถดถอยลอจิสติกทวิภาค

3.4 เกณฑ์การเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์

ผู้วิจัยเลือกใช้เกณฑ์การพิจารณาประสิทธิภาพ 2 เกณฑ์ดังนี้

1. ค่าความถูกต้อง (Accuracy)
2. ค่าความถูกต้องในการจำแนกแบบสมดุล (Balance Classification Accuracy)

3.5 เครื่องมือที่ใช้ในการวิจัย

1. ด้านโปรแกรมคอมพิวเตอร์ หรือ ซอฟต์แวร์

1.1 โปรแกรม Microsoft Excel for Mac Version 16 เพื่อใช้ในการคัดกรองข้อมูล

1.2 โปรแกรม RStudio for Mac Version 4.0.2 เพื่อใช้ในการสร้างตัวแบบพยากรณ์

1.2.1 แพ็คเกจ blorr ใช้ในการสร้างตัวแบบพยากรณ์ด้วยเทคนิคการถดถอย

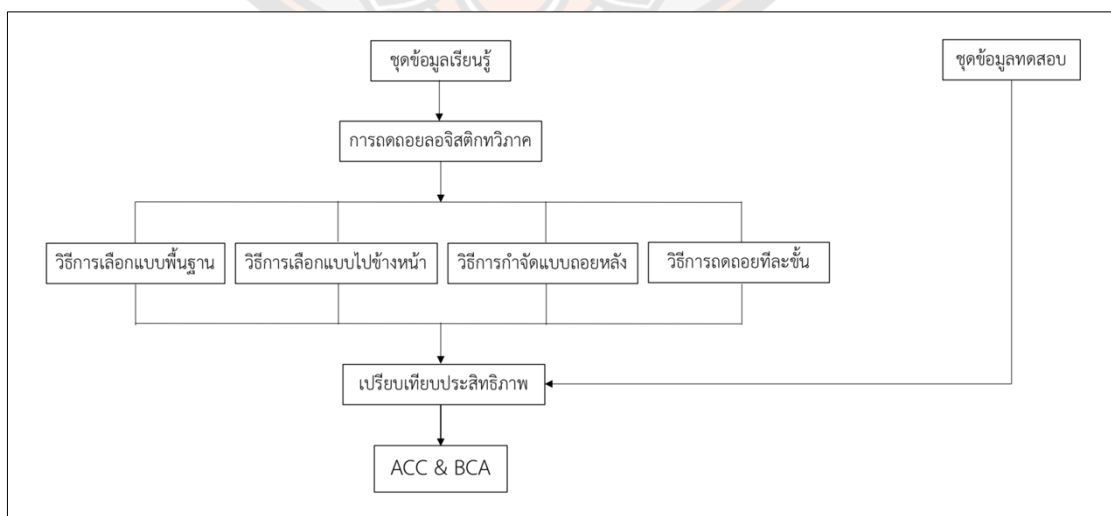
ลอจิสติก

1.2.2 แพ็คเกจ RWeka และ rJava ใช้ในการสร้างตัวแบบพยากรณ์ด้วยเทคนิคต้นไม้ตัดสินใจอัลกอริทึม J48 และอัลกอริทึม LMT

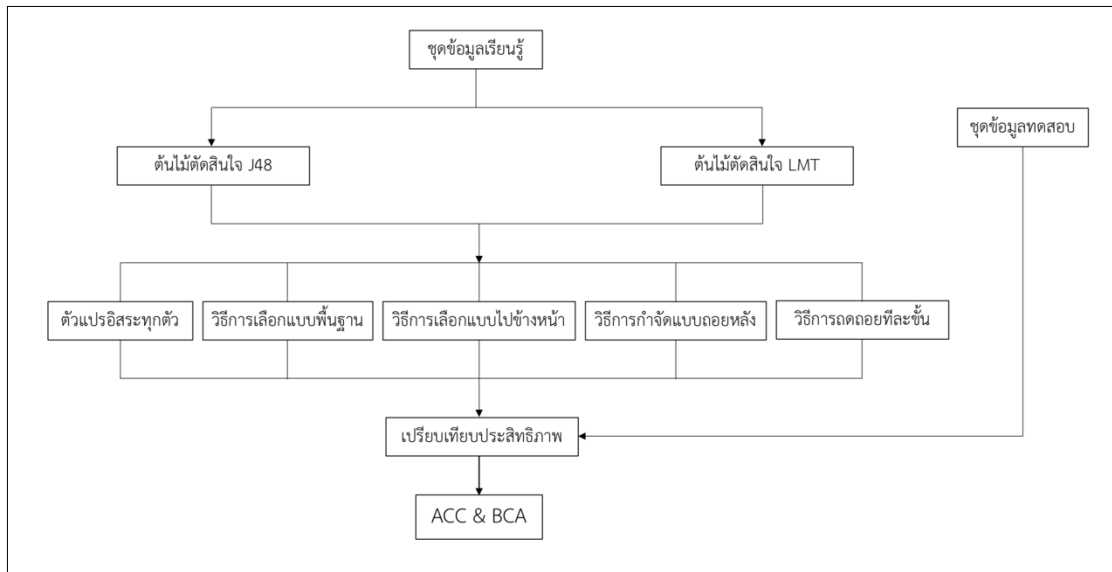
1.2.3 แพ็คเกจ ANN2 ใช้ในการสร้างตัวแบบพยากรณ์ด้วยเทคนิคโครงข่ายประสาทเทียม

1.2.4 แพ็คเกจอื่น ๆ คือ broom car caret dplyr mlbench MASS readr readxl และ tidyverse

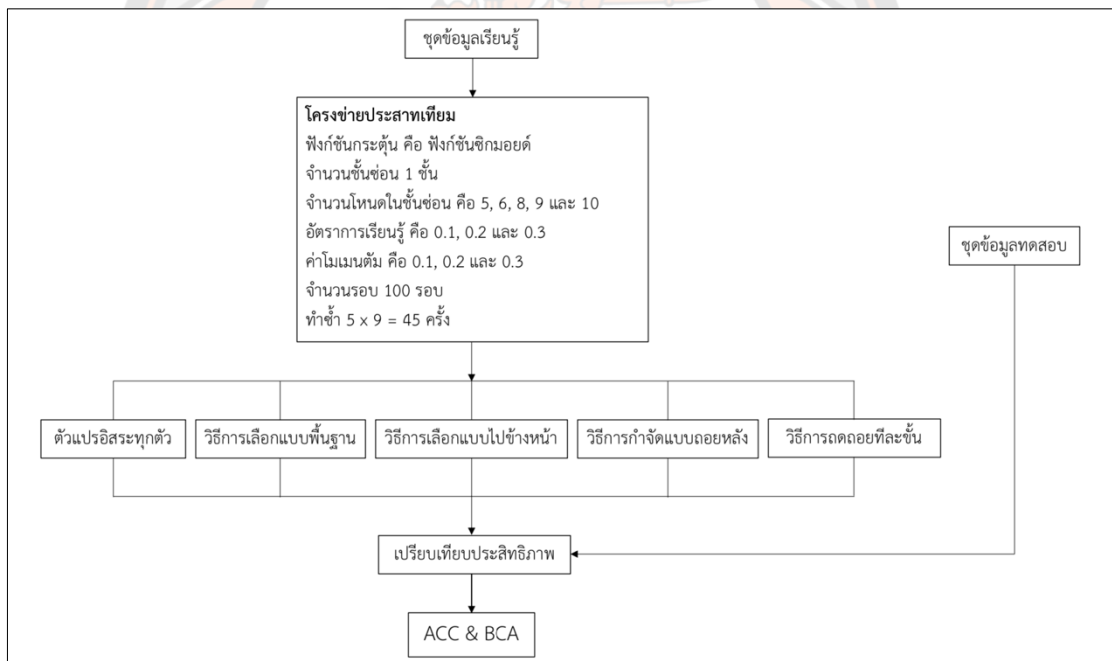
กระบวนการสร้างและพัฒนาตัวแบบพยากรณ์จากเทคนิคต่าง ๆ สามารถสรุปได้ดังผังงานในภาพ 15 – 17 โดยในที่นี้ออกแบบชุดข้อมูลเพื่อการพัฒนาตัวแบบพยากรณ์เป็น 2 แบบ คือชุดข้อมูลแบบที่ 1 และชุดข้อมูลแบบที่ 2



ภาพ 15 แผนผังการสร้างและพัฒนาตัวแบบพยากรณ์ด้วยเทคนิคการถดถอยลอจิสติกทวิภาค



ภาพ 16 แผนผังการสร้างและพัฒนาตัวแบบพยากรณ์ด้วยเทคนิคต้นไม้ตัดสินใจ



ภาพ 17 แผนผังการสร้างและพัฒนาตัวแบบพยากรณ์ด้วยเทคนิคโครงข่ายประสาทเทียม

บทที่ 4

ผลการวิจัย

ในบทนี้จะนำเสนอผลการวิจัยการสร้างและพัฒนาตัวแบบพยากรณ์การเป็นโรคเบาหวานให้มีประสิทธิภาพ โดยมีขั้นตอนดังนี้

1. ผลการเตรียมข้อมูลเพื่อสร้างตัวแบบพยากรณ์
2. ผลการสร้างและพัฒนาตัวแบบพยากรณ์ชุดข้อมูลแบบที่ 1 (70 : 30)
3. ผลการเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์ ชุดข้อมูลแบบที่ 1 (70 : 30)
4. ผลการสร้างและพัฒนาตัวแบบพยากรณ์ชุดข้อมูลแบบที่ 2 (80 : 20)
5. ผลการเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์ ชุดข้อมูลแบบที่ 2 (80 : 20)
6. ผลการเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์ทั้งสองชุดข้อมูล

4.1 ผลการเตรียมข้อมูลเพื่อสร้างตัวแบบพยากรณ์

จากข้อมูลผู้ป่วยทั่วไปที่เข้ามารับการรักษาในโรงพยาบาลมหาวิทยาลัยนเรศวร ตั้งแต่วันที่ 1 มกราคม 2563 – 31 มิถุนายน 2563 จำนวน 10,038 ชุด ทำการคัดเลือกข้อมูลและตัวแปรที่เกี่ยวข้อง เหลือจำนวน 5,081 ชุด ซึ่งประกอบด้วยตัวแปรอิสระที่ใช้ในการวิเคราะห์โอกาสการเป็นโรคเบาหวานมีจำนวน 13 ตัวแปร ได้แก่ เพศ อายุ ค่าความดันขณะหัวใจบีบตัว ค่าความดันขณะหัวใจคลายตัว อัตราการเต้นของหัวใจ น้ำหนัก ความสูง ดัชนีมวลกาย ระดับน้ำตาลในเลือด ระดับคอเลสเตอรอลในเลือด ระดับไตรกลีเซอไรด์ในเลือด ระดับไขมันในเลือดที่มีความหนาแน่นสูง และระดับไขมันในเลือดที่มีความหนาแน่นต่ำ และตัวแปรตาม คือ การเป็นโรคเบาหวาน

ผู้วิจัยเลือกใช้วิธีการทางสถิติ ได้แก่ การถดถอยลอจิสติกทวิภาค และเทคนิคเหมืองข้อมูล 2 เทคนิค ได้แก่ เทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม J48 และอัลกอริทึม LMT และเทคนิคโครงข่ายประสาทเทียม ซึ่งได้มีการปรับแต่งโครงข่าย ค่าอัตราการเรียนรู้ และค่าโมเมนตัม เพื่อให้ได้ตัวแบบพยากรณ์ที่มีประสิทธิภาพที่ดีที่สุด

ในงานวิจัยนี้ผู้วิจัยต้องการเปรียบเทียบประสิทธิภาพของเทคนิคการถดถอยลอจิสติกทวิภาค เทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม J48 และอัลกอริทึม LMT และเทคนิคโครงข่ายประสาทเทียม โดยการแบ่งชุดข้อมูลเรียนรู้ และข้อมูลทดสอบเป็น 2 แบบ คือ ชุดข้อมูลแบบที่ 1 ประกอบไปด้วย ชุดข้อมูลเรียนรู้ 70 เปอร์เซ็นต์ ชุดข้อมูลทดสอบ 30 เปอร์เซ็นต์ และชุดข้อมูลแบบที่ 2 ประกอบไปด้วย ข้อมูลเป็นชุดข้อมูลเรียนรู้ 80 เปอร์เซ็นต์ และชุดข้อมูลทดสอบ 20 เปอร์เซ็นต์

4.2 ผลการสร้างและพัฒนาตัวแบบพยากรณ์ชุดข้อมูลแบบที่ 1 (70 : 30)

4.2.1 เทคนิคการถดถอยลอจิสติกทวิภาค

ผู้วิจัยเริ่มสร้างตัวแบบพยากรณ์ในชุดข้อมูลแบบที่ 1 ซึ่งมีข้อมูลเรียนรู้ 70 เปอร์เซ็นต์ ชุดข้อมูลทดสอบ 30 เปอร์เซ็นต์ และใช้วิธีการคัดเลือกตัวแปรอิสระซึ่งประกอบไปด้วย วิธีการเลือกแบบพื้นฐาน วิธีการเลือกแบบไปข้างหน้า วิธีการกำจัดแบบถอยหลัง และ วิธีการถดถอยทีละขั้น กำหนดระดับนัยสำคัญ 0.05

ผลการสร้างตัวแบบพยากรณ์วิธีการเลือกแบบพื้นฐาน

ตาราง 16 แสดงผลลัพธ์จากการสร้างตัวแบบพยากรณ์ด้วยเทคนิคการถดถอยลอจิสติกทวิภาค วิธีการเลือกแบบพื้นฐาน ของชุดข้อมูลแบบที่ 1

ตัวแปรอิสระ	β	S.E.	Wald	OR	p
ระดับน้ำตาลในเลือด (X_0)	0.0309	0.0016	19.6072	1.0314	<0.001*
น้ำหนัก (X_6)	0.0265	0.0037	52.2087	1.0268	<0.001*
ความสูง (X_7)	-0.0271	0.0062	18.8269	0.9733	<0.001*
ค่าความดันขณะหัวใจคลายตัว (X_4)	-0.0304	0.0044	48.2910	0.9701	<0.001*
ค่าความดันขณะหัวใจบีบตัว (X_3)	0.0137	0.0027	25.3946	1.0138	0.0015*
อัตราการเต้นของหัวใจ (X_5)	0.0148	0.0036	17.1821	1.0149	<0.001*
ค่าคงที่ (Constant)	-3.2126	1.0140	10.0386	0.0403	0.0015*

$-2 \text{ Log likelihood} = 2977.583$ ($df = 7$)
 $AIC = 2991.583$
 $Cox \ \& \ Snell \ R^2 = 0.2056, \ Nagelkerke \ R^2 = 0.3134$

* มีนัยสำคัญเชิงสถิติ 0.05

จากตาราง 16 สามารถวิเคราะห์ได้ดังนี้ ตัวแปรอิสระที่ส่งผลให้เกิดโรคเบาหวาน ได้แก่ ระดับน้ำตาลในเลือด น้ำหนัก ความสูง ค่าความดันขณะหัวใจคลายตัว ค่าความดันขณะหัวใจบีบตัว และอัตราการเต้นของหัวใจอย่างมีนัยสำคัญเชิงสถิติ 0.05

จากค่าอัตราส่วนออดส์ ซึ่งจะเห็นว่าเมื่อระดับน้ำตาลในเลือด น้ำหนัก ค่าความดันขณะหัวใจบีบตัว และอัตราการเต้นของหัวใจ ถ้าค่าเหล่านี้เพิ่มขึ้น 1 หน่วยจะทำให้มีโอกาสในการเป็นโรคเบาหวานเพิ่มขึ้นประมาณ 1 เท่า ในทางตรงกันข้ามเมื่อความสูง และค่าความดันขณะหัวใจคลายตัว ถ้าค่าเหล่านี้เพิ่มขึ้น 1 หน่วย ทำให้มีโอกาสในการเป็นโรคเบาหวานลดลงประมาณ 0.01 เท่า โดยสามารถนำมาเขียนสมการตัวแบบพยากรณ์ได้ดังสมการ (34)

$$P(Y = 1) = \pi_i = \frac{e^{w_1}}{1 + e^{w_1}} \quad (34)$$

เมื่อ w_1 คือ $3.2126 + 0.0137X_3 - 0.0304X_4 + 0.0148X_5 + 0.0265X_6 - 0.0271X_7 - 0.0309X_9$ และเมื่อพิจารณาค่า $-2 \text{ Log likelihood}$ มีค่าเท่ากับ 2977.583 และค่า AIC เท่ากับ 2991.583 เมื่อพิจารณาความสามารถในการพยากรณ์ตัวแบบนี้ พบว่าสามารถอธิบายความผันแปรของการเป็นโรคเบาหวานได้ร้อยละ 20.56 จากนั้นผู้วิจัยทำการตรวจสอบความสัมพันธ์ของตัวแปรอิสระได้ผลลัพธ์ดังตาราง 17

ตาราง 17 แสดงการตรวจสอบความสัมพันธ์ระหว่างตัวแปรอิสระวิธีการเลือกแบบพื้นฐาน ของชุดข้อมูลแบบที่ 1

ตัวแปรอิสระ	Tolerance	VIF
ระดับน้ำตาลในเลือด (X_9)	0.9264	1.0795
น้ำหนัก (X_6)	0.6989	1.4307
ความสูง (X_7)	0.7216	1.3858
ค่าความดันขณะหัวใจคลายตัว (X_4)	0.7038	1.4209
ค่าความดันขณะหัวใจบีบตัว (X_3)	0.7303	1.3692
อัตราการเต้นของหัวใจ (X_5)	0.9417	1.0619

จากตาราง 17 พบว่าค่า Variance Inflation Factor (VIF) และค่าความคลาดเคลื่อน ยินยอม (Tolerance) ของตัวแปรอิสระไม่มีความสัมพันธ์กัน

ผลการสร้างตัวแบบพยากรณ์จากเทคนิคการถดถอยลอจิสติกทวิภาคีวิธีการเลือกแบบพื้นฐาน

สำหรับผลลัพธ์ของตัวแบบพยากรณ์การเป็นโรคเบาหวานจากเทคนิคการถดถอยลอจิสติกทวิภาคีวิธีการเลือกแบบพื้นฐานแสดงดังตาราง 18 และ 19 ตามลำดับ

ตาราง 18 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคการถดถอยลอจิสติกทวิภาคีวิธีการเลือกแบบพื้นฐาน ของชุดข้อมูลเรียนรู้ 70 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	257	545	$\frac{257}{(257 + 545)} = 32.0449$
ไม่เป็นโรคเบาหวาน	141	2612	$\frac{2612}{(2612 + 141)} = 94.8783$
ค่าความถูกต้องของตัวแบบพยากรณ์			80.7032

จากตาราง 18 พบว่าค่าความถูกต้อง คือ 80.7032 เปอร์เซ็นต์ และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 63.4616 เปอร์เซ็นต์

**ตาราง 19 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคการถดถอยลอจิสติก
ทวิภาคีวิธีการเลือกแบบพื้นฐาน ของชุดข้อมูลทดสอบ 30 เปอร์เซ็นต์**

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	102	243	$\frac{102}{(102 + 243)} = 29.5652$
ไม่เป็นโรคเบาหวาน	53	1128	$\frac{1128}{(1128 + 53)} = 95.5123$
ค่าความถูกต้องของตัวแบบพยากรณ์			80.6029

จากตาราง 19 พบว่าค่าความถูกต้อง คือ 80.6029 เปอร์เซ็นต์ และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 62.5387 เปอร์เซ็นต์

ผลการสร้างตัวแบบพยากรณ์วิธีการเลือกแบบไปข้างหน้า

ตาราง 20 แสดงผลการคัดเลือกตัวแปรอิสระวิธีการเลือกแบบไปข้างหน้า ของชุดข้อมูลแบบที่ 1

ตัวแปรอิสระ	β	<i>S.E.</i>	<i>Wald</i>	<i>OR</i>	<i>p</i>
ระดับน้ำตาลในเลือด (X_9)	0.0303	0.0016	375.4526	1.0308	<0.001*
ดัชนีมวลกาย (X_8)	0.0613	0.0093	43.5128	1.0632	<0.001*
ระดับคอเลสเตอรอลในเลือด (X_{10})	-0.0071	0.0013	28.2903	0.9929	<0.001*
ค่าความดันขณะหัวใจคลายตัว (X_4)	-0.0278	0.0043	42.7335	0.9725	<0.001*
ค่าความดันขณะหัวใจบีบตัว (X_3)	0.0135	0.0027	24.9446	1.0136	<0.001*
อัตราการเต้นของหัวใจ (X_5)	0.0160	0.0036	19.8602	1.0161	<0.001*
ค่าคงที่ (Constant)	-6.3103	0.5225	145.8329	0.0018	<0.001*

$$-2 \text{ Log likelihood} = 2957.365 \text{ (df} = 7)$$

$$\text{AIC} = 2971.365$$

$$\text{Cox \& Snell } R^2 = 0.2102, \text{ Nagelkerke } R^2 = 0.3202$$

* มีนัยสำคัญเชิงสถิติ 0.05

จากตาราง 20 สามารถวิเคราะห์ได้ดังนี้ ตัวแปรอิสระที่ส่งผลให้เกิดโรคเบาหวาน ได้แก่ ระดับน้ำตาลในเลือด ดัชนีมวลกาย ระดับคอเลสเตอรอลในเลือด ค่าความดันขณะหัวใจคลายตัว ค่าความดันขณะหัวใจบีบตัว และค่าอัตราการเต้นของหัวใจ อย่างมีนัยสำคัญเชิงสถิติ 0.05

จากค่าอัตราส่วนออดส์ ซึ่งจะเห็นว่าเมื่อระดับน้ำตาลในเลือด ดัชนีมวลกาย ค่าความดันขณะหัวใจบีบตัว และอัตราการเต้นของหัวใจ ถ้าค่าเหล่านี้เพิ่มขึ้น 1 หน่วยจะทำให้มีโอกาสในการเป็นโรคเบาหวานเพิ่มขึ้นประมาณ 1 เท่า ในทางตรงกันข้ามเมื่อระดับคอเลสเตอรอลในเลือด และค่าความดันขณะหัวใจคลายตัว ถ้าค่าเหล่านี้เพิ่มขึ้น 1 หน่วย ทำให้มีโอกาสในการเป็นโรคเบาหวานลดลงประมาณ 0.03 เท่า

โดยสามารถนำมาเขียนสมการตัวแบบพยากรณ์ได้ดังสมการ (35)

$$P(Y = 1) = \pi_i = \frac{e^{w_2}}{1 + e^{w_2}} \quad (35)$$

เมื่อ w_2 คือ $-6.3103 + 0.0135X_3 - 0.0278X_4 + 0.0160X_5 + 0.0613X_8 + 0.0303X_9 - 0.0071X_{10}$ และเมื่อพิจารณาค่า $-2 \text{ Log likelihood}$ มีค่าเท่ากับ 2957.365 และค่า AIC เท่ากับ 2971.365 เมื่อพิจารณาความสามารถในการพยากรณ์ตัวแบบนี้ พบว่าสามารถอธิบายความผันแปรของการเป็นโรคเบาหวานได้ร้อยละ 21.02 จากนั้นผู้วิจัยทำการตรวจสอบความสัมพันธ์ของตัวแปรอิสระได้ผลลัพธ์ดังตาราง 21

ตาราง 21 แสดงการตรวจสอบความสัมพันธ์ระหว่างตัวแปรอิสระวิธีการเลือกแบบไปข้างหน้าของชุดข้อมูลแบบที่ 1

ตัวแปรอิสระ	Tolerance	VIF
ระดับน้ำตาลในเลือด (X_9)	0.9237	1.0825
ดัชนีมวลกาย (X_8)	0.9381	1.0659
ระดับคอเลสเตอรอลในเลือด (X_{10})	0.9794	1.0210
ค่าความดันขณะหัวใจคลายตัว (X_4)	0.7329	1.3644
ค่าความดันขณะหัวใจบีบตัว (X_3)	0.7452	1.3419
อัตราการเต้นของหัวใจ (X_5)	0.9448	1.0585

จากตาราง 21 พบว่าค่า Variance Inflation Factor (VIF) และค่าความคลาดเคลื่อน ยินยอม (Tolerance) ของตัวแปรอิสระทุกตัวไม่มีความสัมพันธ์กัน

ผลการสร้างตัวแบบพยากรณ์จากเทคนิคการถดถอยลอจิสติกทวิภาคีวิธีการเลือกแบบไปข้างหน้า

สำหรับผลลัพธ์ของตัวแบบพยากรณ์การเป็นโรคเบาหวานจากเทคนิคการถดถอยลอจิสติกทวิภาคีวิธีการเลือกแบบไปข้างหน้า แสดงดังตาราง 22 และ 23 ตามลำดับ

ตาราง 22 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคการถดถอยลอจิสติกทวิภาคีวิธีการเลือกแบบไปข้างหน้า ของชุดข้อมูลเรียนรู้ 70 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	263	539	32.7930
ไม่เป็นโรคเบาหวาน	141	2612	94.8783
ค่าความถูกต้องของตัวแบบพยากรณ์			80.8720

จากตาราง 22 พบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 80.8720 เปอร์เซ็นต์ และ 63.8357 เปอร์เซ็นต์

ตาราง 23 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคการถดถอยลอจิสติกทวิภาคีวิธีการเลือกแบบไปข้างหน้า ของชุดข้อมูลทดสอบ 30 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	98	247	28.4058
ไม่เป็นโรคเบาหวาน	56	1125	95.2583
ค่าความถูกต้องของตัวแบบพยากรณ์			80.1442

จากตาราง 23 พบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 80.1442 เปอร์เซ็นต์ และ 61.8320 เปอร์เซ็นต์

ผลการสร้างตัวแบบพยากรณ์ด้วยวิธีการกำจัดแบบถอยหลัง

ตาราง 24 แสดงผลการคัดเลือกตัวแปรอิสระวิธีการกำจัดแบบถอยหลัง ของชุดข้อมูลแบบที่ 1

ตัวแปรอิสระ	β	<i>S.E.</i>	<i>Wald</i>	<i>OR</i>	<i>p</i>
ค่าความดันขณะหัวใจบีบตัว (X_3)	0.0131	0.0027	22.8531	1.0132	<0.001*
ค่าความดันขณะหัวใจคลายตัว (X_4)	-0.0283	0.0044	41.5012	0.9721	<0.001*

ตาราง 24 (ต่อ)

ตัวแปรอิสระ	β	<i>S.E.</i>	<i>Wald</i>	<i>OR</i>	<i>p</i>
อัตราการเต้นของหัวใจ (X_5)	0.0157	0.0036	18.9944	1.0158	<0.001*
น้ำหนัก(X_6)	0.0255	0.0037	46.9157	1.0258	<0.001*
ความสูง (X_7)	-0.0277	0.0063	19.4550	0.9726	<0.001*
ระดับน้ำตาลในเลือด (X_9)	0.0297	0.0016	356.9478	1.0301	<0.001*
ระดับไขมันในเลือดที่มีความหนาแน่นสูง (X_{12})	-0.0113	0.0038	9.1494	0.9887	0.0025*
ระดับไขมันในเลือดที่มีความหนาแน่นต่ำ (X_{13})	-0.0070	0.0015	21.7694	0.9931	<0.001*
ค่าคงที่ (Constant)	-1.7942	1.0599	2.8658	0.1663	0.0905

2 Log likelihood = 2943.963 (df = 9)
AIC = 2961.963
Cox & Snell R^2 = 0.2131, Nagelkerke R^2 = 0.3248

* มีนัยสำคัญเชิงสถิติ 0.05

จากตาราง 24 สามารถวิเคราะห์ได้ดังนี้ ตัวแปรอิสระที่ส่งผลให้เกิดโรคเบาหวาน ได้แก่ ค่าความดันขณะหัวใจบีบตัว ค่าความดันขณะหัวใจคลายตัว ค่าอัตราการเต้นของหัวใจ น้ำหนัก ความสูง ระดับน้ำตาลในเลือด ระดับไขมันในเลือดที่มีความหนาแน่นสูง และระดับไขมันในเลือดที่มีความหนาแน่นต่ำ อย่างมีนัยสำคัญเชิงสถิติ 0.05

จากค่าอัตราส่วนออดส์ ซึ่งจะเห็นว่าเมื่อค่าความดันขณะหัวใจบีบตัว ค่าอัตราการเต้นของหัวใจ น้ำหนัก และระดับน้ำตาลในเลือด ถ้าค่าเหล่านี้เพิ่มขึ้น 1 หน่วยจะทำให้มีโอกาสในการเป็นโรคเบาหวานเพิ่มขึ้นประมาณ 1 เท่า ในทางตรงกันข้ามเมื่อค่าความดันขณะหัวใจคลายตัว ความสูง ระดับไขมันในเลือดที่มีความหนาแน่นสูง และระดับไขมันในเลือดที่มีความหนาแน่นต่ำ ถ้าค่าเหล่านี้เพิ่มขึ้น 1 หน่วย ทำให้มีโอกาสในการเป็นโรคเบาหวานลดลงประมาณ 0.03 เท่า

โดยสามารถนำมาเขียนสมการตัวแบบพยากรณ์ได้ดังสมการ (36)

$$P(Y = 1) = \pi_i = \frac{e^{w_3}}{1 + e^{w_3}} \quad (36)$$

เมื่อ w_3 คือ $-1.7942 + 0.0131X_3 - 0.0283X_4 + 0.0157X_5 + 0.0255X_6 - 0.0277X_7 + 0.0297X_9 - 0.0113X_{12} - 0.0070X_{13}$

และเมื่อพิจารณาค่า $-2 \text{ Log likelihood}$ มีค่าเท่ากับ 2943.963 และค่า AIC เท่ากับ 2961.963 เมื่อพิจารณาความสามารถในการพยากรณ์ตัวแบบนี้ พบว่าสามารถอธิบายความผันแปรของการเป็น

โรคเบาหวานได้ร้อยละ 21.31 จากนั้นผู้วิจัยทำการตรวจสอบความสัมพันธ์ของตัวแปรอิสระได้ผลลัพธ์ดังตาราง 25

ตาราง 25 แสดงการตรวจสอบความสัมพันธ์ระหว่างตัวแปรอิสระวิธีการกำจัดแบบถอยหลัง ของชุดข้อมูลแบบที่ 1

ตัวแปรอิสระ	Tolerance	VIF
ค่าความดันขณะหัวใจบีบตัว (X_3)	0.7259	1.3777
ค่าความดันขณะหัวใจคลายตัว (X_4)	0.6958	1.4371
อัตราการเต้นของหัวใจ (X_5)	0.9405	1.0632
น้ำหนัก(X_6)	0.6722	1.4877
ความสูง (X_7)	0.7213	1.3865
ระดับน้ำตาลในเลือด (X_9)	0.9062	1.1035
ระดับไขมันในเลือดที่มีความหนาแน่นสูง (X_{12})	0.9172	1.0902
ระดับไขมันในเลือดที่มีความหนาแน่นต่ำ (X_{13})	0.9752	1.0254

จากตาราง 25 พบว่าค่า Variance Inflation Factor (VIF) และค่าความคลาดเคลื่อน ยินยอม (Tolerance) ของตัวแปรอิสระทุกตัวไม่มีความสัมพันธ์กัน

ผลการสร้างตัวแบบพยากรณ์จากเทคนิคการถดถอยลอจิสติกทวิภาคีวิธีการกำจัดแบบถอยหลัง

สำหรับผลลัพธ์ของตัวแบบพยากรณ์การเป็นโรคเบาหวานจากเทคนิคการถดถอยลอจิสติกทวิภาคีวิธีการกำจัดแบบถอยหลัง แสดงดังตาราง 26 และ 27 ตามลำดับ

ตาราง 26 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคการถดถอยลอจิสติกทวิภาคีวิธีการกำจัดแบบถอยหลัง ของชุดข้อมูลเรียนรู้ 70 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	265	537	33.0424
ไม่เป็นโรคเบาหวาน	143	2610	94.8057
ค่าความถูกต้องของตัวแบบพยากรณ์			80.8720

จากตาราง 26 พบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 80.8720 เปอร์เซ็นต์ และ 63.9240 เปอร์เซ็นต์

ตาราง 27 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคการถดถอยลอจิสติก
ทวิภาคีวิธีการกำจัดแบบถอยหลัง ของชุดข้อมูลทดสอบ 30 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	98	247	28.4058
ไม่เป็นโรคเบาหวาน	59	1122	95.0042
ค่าความถูกต้องของตัวแบบพยากรณ์			79.9476

จากตาราง 27 พบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 79.9476 เปอร์เซ็นต์ และ 61.7050 เปอร์เซ็นต์

ผลการสร้างตัวแบบพยากรณ์วิธีการถดถอยทีละขั้น

ตาราง 28 แสดงผลการคัดเลือกตัวแปรอิสระวิธีการถดถอยทีละขั้น ของชุดข้อมูลแบบที่ 1

ตัวแปรอิสระ	β	S.E.	Wald	OR	p
ระดับน้ำตาลในเลือด (X_9)	0.0303	0.0016	375.4526	1.0308	<0.001*
ดัชนีมวลกาย (X_8)	0.0613	0.0093	43.5128	1.0632	<0.001*
ระดับคอเลสเตอรอลในเลือด (X_{10})	-0.0071	0.0013	28.2903	0.9929	<0.001*
ค่าความดันขณะหัวใจคลายตัว (X_4)	-0.0278	0.0043	42.7335	0.9725	<0.001*
ค่าความดันขณะหัวใจบีบตัว (X_3)	0.0135	0.0027	24.9446	1.0136	<0.001*
อัตราการเต้นของหัวใจ (X_5)	0.0160	0.0036	19.8602	1.0161	<0.001*
ค่าคงที่ (Constant)	-6.3103	0.5225	145.8329	0.0018	<0.001*

$-2 \text{ Log likelihood} = 2957.365$ ($df = 7$)

AIC = 2971.365

Cox & Snell $R^2 = 0.2102$, Nagelkerke $R^2 = 0.3202$

* มีนัยสำคัญเชิงสถิติ 0.05

จากตาราง 28 สามารถวิเคราะห์ได้ดังนี้ ตัวแปรอิสระที่ส่งผลให้เกิดโรคเบาหวาน ได้แก่ ระดับน้ำตาลในเลือด ดัชนีมวลกาย ระดับคอเลสเตอรอลในเลือด ค่าความดันขณะหัวใจคลายตัว ค่าความดันขณะหัวใจบีบตัว และค่าอัตราการเต้นของหัวใจ อย่างมีนัยสำคัญเชิงสถิติ 0.05

จากค่าอัตราส่วนออดส์ ซึ่งจะเห็นว่าเมื่อระดับน้ำตาลในเลือด ดัชนีมวลกาย ค่าความดันขณะหัวใจบีบตัว และอัตราการเต้นของหัวใจ ถ้าค่าเหล่านี้เพิ่มขึ้น 1 หน่วยจะทำให้มีโอกาสในการ

เป็นโรคเบาหวานเพิ่มขึ้นประมาณ 1 เท่า ในทางตรงกันข้ามเมื่อระดับคอเลสเตอรอลในเลือด และค่าความดันขณะหัวใจคลายตัว ถ้าค่าเหล่านี้เพิ่มขึ้น 1 หน่วย ทำให้มีโอกาสในการเป็นโรคเบาหวานลดลงประมาณ 0.03 เท่า

โดยสามารถนำมาเขียนสมการตัวแบบพยากรณ์ได้ดังสมการ (37)

$$P(Y = 1) = \pi_i = \frac{e^{w_4}}{1 + e^{w_4}} \quad (37)$$

เมื่อ w_4 คือ $-6.3103 + 0.0135X_3 - 0.0278X_4 + 0.0160X_5 + 0.0613X_8 + 0.0303X_9 - 0.0071X_{10}$ และเมื่อพิจารณาว่า $-2 \text{ Log likelihood}$ มีค่าเท่ากับ 2957.365 และค่า AIC เท่ากับ 2971.365 เมื่อพิจารณาความสามารถในการพยากรณ์ตัวแบบนี้ พบว่าสามารถอธิบายความผันแปรของการเป็นโรคเบาหวานได้ร้อยละ 21.02 จากนั้นผู้วิจัยทำการตรวจสอบความสัมพันธ์ของตัวแปรอิสระได้ผลลัพธ์ดังตาราง 29

ตาราง 29 แสดงการตรวจสอบความสัมพันธ์ระหว่างตัวแปรอิสระวิธีการถดถอยทีละขั้น ของชุดข้อมูลแบบที่ 1

ตัวแปรอิสระ	Tolerance	VIF
ระดับน้ำตาลในเลือด (X_9)	0.9237	1.0825
ดัชนีมวลกาย (X_8)	0.9381	1.0659
ระดับคอเลสเตอรอลในเลือด (X_{10})	0.9794	1.0210
ค่าความดันขณะหัวใจคลายตัว (X_4)	0.7329	1.3644
ค่าความดันขณะหัวใจบีบตัว (X_3)	0.7452	1.3419
อัตราการเต้นของหัวใจ (X_5)	0.9448	1.0585

จากตาราง 29 พบว่าค่า Variance Inflation Factor (VIF) และค่าความคลาดเคลื่อน ยินยอม (Tolerance) ของตัวแปรอิสระทุกตัวไม่มีความสัมพันธ์กัน

ผลการสร้างตัวแบบพยากรณ์จากเทคนิคการถดถอยลอจิสติกทวิภาคีวิธีการถดถอยทีละชั้น

สำหรับผลลัพธ์ของตัวแบบพยากรณ์การเป็นโรคเบาหวานจากเทคนิคการถดถอยลอจิสติกทวิภาคีวิธีการถดถอยทีละชั้น แสดงดังตาราง 30 และ 31 ตามลำดับ

ตาราง 30 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคการถดถอยลอจิสติกทวิภาคีวิธีการถดถอยทีละชั้น ของชุดข้อมูลเรียนรู้ 70 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	263	539	32.7930
ไม่เป็นโรคเบาหวาน	141	2612	94.8783
ค่าความถูกต้องของตัวแบบพยากรณ์			80.8720

จากตาราง 30 พบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 80.8720 เปอร์เซ็นต์ และ 63.8357 เปอร์เซ็นต์

ตาราง 31 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคการถดถอยลอจิสติกทวิภาคีวิธีการถดถอยทีละชั้น ของชุดข้อมูลทดสอบ 30 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	98	247	28.4058
ไม่เป็นโรคเบาหวาน	56	1125	95.2583
ค่าความถูกต้องของตัวแบบพยากรณ์			80.1442

จากตาราง 31 พบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 80.1442 เปอร์เซ็นต์ และ 61.8320 เปอร์เซ็นต์

หลังจากนั้นผู้วิจัยทำการตรวจสอบความเหมาะสมของตัวแบบการถดถอยลอจิสติกทวิภาคีโดยใช้การทดสอบอัตราส่วนภาวะน่าจะเป็น ได้ผลลัพธ์ดังตาราง 32

ตาราง 32 แสดงผลการเปรียบเทียบการตรวจสอบความเหมาะสมของตัวแบบพยากรณ์ ด้วยเทคนิคการถดถอยลอจิสติกทวิภาค ของชุดข้อมูลแบบที่ 1

เทคนิคการถดถอยลอจิสติกทวิภาค	$-2LL$	AIC
วิธีการเลือกแบบพื้นฐาน	2977.583	2991.583
วิธีการเลือกแบบไปข้างหน้า	2957.365	2971.365
วิธีการกำจัดแบบถอยหลัง	2943.963	2961.963
วิธีการถดถอยทีละขั้น	2957.365	2971.365

จากตาราง 32 เทคนิคการถดถอยลอจิสติกทวิภาค พบว่ามีค่าลอการิทึมของภาวะน่าจะเป็น ($-2 \text{ Log likelihood}$) และค่าเกณฑ์สารสนเทศอะกะอิเกะ (AIC) เหมาะสมในวิธีการกำจัดแบบถอยหลัง ซึ่งเราจะพิจารณาระดับความน่าเชื่อถือของตัวแบบพยากรณ์ร่วมด้วยเพื่อให้ตัวแบบที่ได้มีความเหมาะสมมากที่สุด แสดงดังตาราง 33

ตาราง 33 แสดงผลระดับความน่าเชื่อถือของตัวแบบพยากรณ์ด้วยเทคนิคการถดถอยลอจิสติกทวิภาค ของชุดข้อมูลแบบที่ 1

เทคนิคการถดถอยลอจิสติกทวิภาค	$Cox \& Snell R^2$ (%)	$Nagelkerke R^2$ (%)
วิธีการเลือกแบบพื้นฐาน	20.5648	31.3374
วิธีการเลือกแบบไปข้างหน้า	21.0153	32.0239
วิธีการกำจัดแบบถอยหลัง	21.3125	32.4767
วิธีการถดถอยทีละขั้น	21.0153	32.0239

จากตาราง 33 เทคนิคการถดถอยลอจิสติกทวิภาควิธีการกำจัดแบบถอยหลัง พบว่าสถิติทดสอบความน่าเชื่อถือของตัวแบบพยากรณ์ $Cox \& Snell R^2$ ตัวแบบพยากรณ์สามารถอธิบายการเป็นโรคเบาหวานได้ร้อยละ 21.3125 และ $Nagelkerke R^2$ พบว่าตัวแบบพยากรณ์สามารถอธิบายการเป็นโรคเบาหวานได้ร้อยละ 32.4767 สามารถอธิบายการเป็นโรคเบาหวานได้ดีที่สุด

ผลการเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์จากเทคนิคการถดถอยลอจิสติก
ทวิภาค ของชุดข้อมูลแบบที่ 1

ตาราง 34 แสดงผลเปรียบเทียบประสิทธิภาพการพยากรณ์จากเทคนิคการถดถอยลอจิสติก
ทวิภาค ของชุดข้อมูลแบบที่ 1

เทคนิคการถดถอยลอจิสติกทวิภาค	ชุดเรียนรู้		ชุดทดสอบ	
	ACC	BCA	ACC	BCA
วิธีการเลือกแบบพื้นฐาน	80.7032	63.4616	80.6029	62.5387
วิธีการเลือกแบบไปข้างหน้า	80.8720	63.8357	80.1442	61.8320
วิธีการกำจัดแบบถอยหลัง	80.8720	63.9240	79.9476	61.7050
วิธีการถดถอยทีละชั้น	80.8720	63.8357	80.1442	61.8320

จากตาราง 34 เมื่อพิจารณาผลลัพธ์ในชุดข้อมูลทดสอบ พบว่าเทคนิคการถดถอยลอจิสติกทวิภาควิธีการเลือกแบบพื้นฐาน มีค่าความถูกต้องในชุดข้อมูลทดสอบ 80.6029 เปอร์เซ็นต์ ค่าความถูกต้องในการจำแนกแบบสมดุล 62.5387 เปอร์เซ็นต์

หลังจากผู้วิจัยเริ่มสร้างตัวแบบพยากรณ์ในชุดข้อมูลแบบที่ 1 ซึ่งมีข้อมูลเรียนรู้ 70 เปอร์เซ็นต์ ชุดข้อมูลทดสอบ 30 เปอร์เซ็นต์ ด้วยเทคนิคการถดถอยลอจิสติกทวิภาคแล้ว ต่อมาผู้วิจัยพิจารณาสร้างตัวแบบพยากรณ์ด้วยเทคนิคต้นไม้ตัดสินใจอัลกอริทึม J48 และอัลกอริทึม LMT ตามลำดับ

4.2.2 เทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม J48

ผู้วิจัยเลือกใช้ค่าพารามิเตอร์ดังต่อไปนี้ กำหนดช่วงความเชื่อมั่นที่ใช้ในการแตกกิ่งต้นไม้เท่ากับ 95% และจำนวนการแตกกิ่งเท่ากับ 2

สำหรับผลลัพธ์ของตัวแบบพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม J48 โดยมีวิธีการคัดเลือกตัวแปรอิสระดังนี้

1) กรณีตัวแปรอิสระทุกตัว ได้แก่ ตัวแปรอิสระ $X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9, X_{10}, X_{11}, X_{12}$

และ X_{13}

2) วิธีการเลือกแบบพื้นฐาน ได้แก่ ตัวแปรอิสระ X_3, X_4, X_5, X_6, X_7 และ X_9

3) วิธีการเลือกแบบไปข้างหน้า X_3, X_4, X_5, X_8, X_9 และ X_{10}

4) วิธีการกำจัดแบบถอยหลัง $X_3, X_4, X_5, X_6, X_7, X_9, X_{12}$ และ X_{13}

5) วิธีการถดถอยทีละชั้น X_3, X_4, X_5, X_8, X_9 และ X_{10}

ผลการสร้างตัวแบบพยากรณ์จากเทคนิคต้นไม้ตัดสินใจ J48 กรณีตัวแปรอิสระทุกตัว

สำหรับผลลัพธ์ของตัวแบบพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วย อัลกอริทึม J48 กรณีตัวแปรอิสระทุกตัวแสดงดังตาราง 35 และ 36 ตามลำดับ

ตาราง 35 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วย อัลกอริทึม J48 กรณีตัวแปรอิสระทุกตัว ของชุดข้อมูลเรียนรู้ 70 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	590	286	67.3516
ไม่เป็นโรคเบาหวาน	212	2467	92.0866
ค่าความถูกต้องของตัวแบบพยากรณ์			85.9916

จากตาราง 35 พบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 85.9916 เปอร์เซ็นต์ และ 79.7191 เปอร์เซ็นต์

ตาราง 36 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วย อัลกอริทึม J48 กรณีตัวแปรอิสระทุกตัว ของชุดข้อมูลทดสอบ 30 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	206	162	55.9783
ไม่เป็นโรคเบาหวาน	139	1019	87.9965
ค่าความถูกต้องของตัวแบบพยากรณ์			80.2752

จากตาราง 36 พบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 80.2752 เปอร์เซ็นต์ และ 71.9874 เปอร์เซ็นต์

ผลการสร้างตัวแบบพยากรณ์จากเทคนิคต้นไม้ตัดสินใจ J48 วิธีการเลือกแบบพื้นฐาน

สำหรับผลลัพธ์ของตัวแบบพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วย
อัลกอริทึม J48 วิธีการเลือกแบบพื้นฐานแสดงดังตาราง 37 และ 38 ตามลำดับ

**ตาราง 37 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วย
อัลกอริทึม J48 วิธีการเลือกแบบพื้นฐาน ของชุดข้อมูลเรียนรู้ 70 เปอร์เซ็นต์**

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	520	319	61.9785
ไม่เป็นโรคเบาหวาน	282	2434	89.6171
ค่าความถูกต้องของตัวแบบพยากรณ์			83.0942

จากตาราง 37 พบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ
83.0942 เปอร์เซ็นต์ และ 75.7978 เปอร์เซ็นต์

**ตาราง 38 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วย
อัลกอริทึม J48 วิธีการเลือกแบบพื้นฐาน ของชุดข้อมูลทดสอบ 30 เปอร์เซ็นต์**

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	205	140	59.4203
ไม่เป็นโรคเบาหวาน	140	1041	88.1456
ค่าความถูกต้องของตัวแบบพยากรณ์			81.6514

จากตาราง 38 พบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ
81.6514 เปอร์เซ็นต์ และ 73.7830 เปอร์เซ็นต์

ผลการสร้างตัวแบบพยากรณ์จากเทคนิคต้นไม้ตัดสินใจ J48 วิธีการเลือกแบบไปข้างหน้า

สำหรับผลลัพธ์ของตัวแบบพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม J48 วิธีการเลือกแบบไปข้างหน้าแสดงดังตาราง 39 และ 40 ตามลำดับ

ตาราง 39 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม J48 วิธีการเลือกแบบไปข้างหน้า ของชุดข้อมูลเรียนรู้ 70 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	499	303	62.2195
ไม่เป็นโรคเบาหวาน	303	2450	88.9938
ค่าความถูกต้องของตัวแบบพยากรณ์			82.9536

จากตาราง 39 พบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 82.9536 เปอร์เซ็นต์ และ 75.6066 เปอร์เซ็นต์

ตาราง 40 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม J48 วิธีการเลือกแบบไปข้างหน้า ของชุดข้อมูลทดสอบ 30 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	196	132	59.7561
ไม่เป็นโรคเบาหวาน	149	1049	87.5626
ค่าความถูกต้องของตัวแบบพยากรณ์			81.5859

จากตาราง 40 พบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 81.5859 เปอร์เซ็นต์ และ 73.6594 เปอร์เซ็นต์

ผลการสร้างตัวแบบพยากรณ์จากเทคนิคต้นไม้ตัดสินใจ J48 วิธีการกำจัดแบบถอยหลัง สำหรับผลลัพธ์ของตัวแบบพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วย อัลกอริทึม J48 วิธีการกำจัดแบบถอยหลังแสดงดังตาราง 41 และ 42 ตามลำดับ

ตาราง 41 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วย อัลกอริทึม J48 วิธีการกำจัดแบบถอยหลัง ของชุดข้อมูลเรียนรู้ 70 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	493	250	66.3526
ไม่เป็นโรคเบาหวาน	309	2503	89.0114
ค่าความถูกต้องของตัวแบบพยากรณ์			84.2757

จากตาราง 41 พบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 84.2757 เปอร์เซ็นต์ และ 77.6820 เปอร์เซ็นต์

ตาราง 42 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วย อัลกอริทึม J48 วิธีการกำจัดแบบถอยหลัง ของชุดข้อมูลทดสอบ 30 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	178	113	61.1684
ไม่เป็นโรคเบาหวาน	167	1068	86.4777
ค่าความถูกต้องของตัวแบบพยากรณ์			81.6514

จากตาราง 42 พบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 81.6514 เปอร์เซ็นต์ และ 73.8231 เปอร์เซ็นต์

ผลการสร้างตัวแบบพยากรณ์จากเทคนิคต้นไม้ตัดสินใจ J48 วิธีการถดถอยทีละชั้น

สำหรับผลลัพธ์ของตัวแบบพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วย อัลกอริทึม J48 วิธีการถดถอยทีละชั้นแสดงดังตาราง 43 และ 44 ตามลำดับ

ตาราง 43 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วย อัลกอริทึม J48 วิธีการถดถอยทีละชั้น ของชุดข้อมูลเรียนรู้ 70 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	499	303	62.2195
ไม่เป็นโรคเบาหวาน	303	2450	88.9938
ค่าความถูกต้องของตัวแบบพยากรณ์			82.9536

จากตาราง 43 พบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 82.9536 เปอร์เซ็นต์ และ 75.6066 เปอร์เซ็นต์

ตาราง 44 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วย อัลกอริทึม J48 วิธีการถดถอยทีละชั้น ของชุดข้อมูลทดสอบ 30 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	196	132	59.7561
ไม่เป็นโรคเบาหวาน	149	1049	87.5626
ค่าความถูกต้องของตัวแบบพยากรณ์			81.5859

จากตาราง 44 พบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 81.5859 เปอร์เซ็นต์ และ 73.6594 เปอร์เซ็นต์

ผลการเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์จากเทคนิคต้นไม้ตัดสินใจด้วย
อัลกอริทึม J48 ของชุดข้อมูลแบบที่ 1

ตาราง 45 แสดงผลการเปรียบเทียบประสิทธิภาพการพยากรณ์ ด้วยเทคนิคต้นไม้ตัดสินใจด้วย
อัลกอริทึม J48 ของชุดข้อมูลแบบที่ 1

เทคนิคต้นไม้ตัดสินใจ อัลกอริทึม J48	ชุดเรียนรู้		ชุดทดสอบ	
	ACC	BCA	ACC	BCA
กรณีตัวแปรอิสระทุกตัว	85.9916	79.7191	80.2752	71.9874
วิธีการเลือกแบบพื้นฐาน	83.0942	75.7978	81.6514	73.7830
วิธีการเลือกแบบไปข้างหน้า	82.9536	75.6066	81.5859	73.6594
วิธีการกำจัดแบบถอยหลัง	84.2757	77.6820	81.6514	73.8231
วิธีการถดถอยทีละขั้น	82.9536	75.6066	81.5859	73.6594

จากตาราง 45 พบว่าเทคนิคต้นไม้ตัดสินใจอัลกอริทึม J48 กรณีการคัดเลือกตัวแปรอิสระด้วยวิธีการกำจัดแบบถอยหลังในเทคนิคการถดถอยลอจิสติกทวิภาค ประกอบไปด้วยตัวแปรอิสระ ค่าความดันขณะหัวใจบีบตัว ค่าความดันขณะหัวใจคลายตัว ค่าอัตราการเต้นของหัวใจ น้ำหนัก ความสูง ระดับน้ำตาลในเลือด ระดับไขมันในเลือดที่มีความหนาแน่นสูง และระดับไขมันในเลือดที่มีความหนาแน่นต่ำ มีค่าความถูกต้องในชุดข้อมูลทดสอบที่ดีที่สุด 81.6514 เปอร์เซ็นต์ ค่าความถูกต้องในการจำแนกแบบสมดุล 73.8231 เปอร์เซ็นต์

4.2.3 เทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม LMT

ผู้วิจัยกำหนดจำนวนการทำซ้ำสำหรับอัลกอริทึมลอจิสติกบูส (LogitBoost) เท่ากับ -1 ซึ่งหมายถึง การทำซ้ำแบบการตรวจสอบข้าม (Cross-Validation) จำนวนตัวแปรอิสระขั้นต่ำในโหนดที่ถูกพิจารณา กำหนดค่าเท่ากับ 15

สำหรับผลลัพธ์ของตัวแบบพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม LMT โดยมีวิธีการคัดเลือกตัวแปรอิสระทั้ง 5 รูปแบบ กรณีตัวแปรอิสระทุกตัว วิธีการเลือกแบบพื้นฐาน วิธีการเลือกแบบไปข้างหน้า วิธีการกำจัดแบบถอยหลัง และวิธีการถดถอยทีละขั้น

ผลการสร้างตัวแบบพยากรณ์จากเทคนิคต้นไม้ตัดสินใจ LMT กรณีตัวแปรอิสระทุกตัว สำหรับผลลัพธ์ของตัวแบบพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วย อัลกอริทึม LMT กรณีตัวแปรอิสระทุกตัวแสดงดังตาราง 46 แล 47 ตามลำดับ

ตาราง 46 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วย อัลกอริทึม LMT กรณีตัวแปรอิสระทุกตัว ของชุดข้อมูลเรียนรู้ 70 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	506	263	65.7997
ไม่เป็นโรคเบาหวาน	296	2490	89.3754
ค่าความถูกต้องของตัวแบบพยากรณ์			84.2757

จากตาราง 46 พบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 84.2757 เปอร์เซ็นต์ และ 77.5876 เปอร์เซ็นต์

ตาราง 47 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วย อัลกอริทึม LMT กรณีตัวแปรอิสระทุกตัว ของชุดข้อมูลทดสอบ 30 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	188	127	59.6825
ไม่เป็นโรคเบาหวาน	157	1054	87.0355
ค่าความถูกต้องของตัวแบบพยากรณ์			81.3893

จากตาราง 47 มีพบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 81.3893 เปอร์เซ็นต์ และ 73.3590 เปอร์เซ็นต์

ผลการสร้างตัวแบบพยากรณ์จากเทคนิคต้นไม้ตัดสินใจ LMT วิธีการเลือกแบบพื้นฐาน สำหรับผลลัพธ์ของตัวแบบพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วย อัลกอริทึม LMT วิธีการเลือกแบบพื้นฐานแสดงดังตาราง 48 และ 49 ตามลำดับ

ตาราง 48 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วย อัลกอริทึม LMT วิธีการเลือกแบบพื้นฐาน ของชุดข้อมูลเรียนรู้ 70 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	483	289	62.5648
ไม่เป็นโรคเบาหวาน	319	2464	88.5375
ค่าความถูกต้องของตัวแบบพยากรณ์			82.8973

จากตาราง 48 พบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 82.8973 เปอร์เซ็นต์ และ 75.5512 เปอร์เซ็นต์

ตาราง 49 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วย อัลกอริทึม LMT วิธีการเลือกแบบพื้นฐาน ของชุดข้อมูลทดสอบ 30 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	188	125	60.0639
ไม่เป็นโรคเบาหวาน	157	1056	87.0569
ค่าความถูกต้องของตัวแบบพยากรณ์			81.5203

จากตาราง 49 มีพบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 81.5203 เปอร์เซ็นต์ และ 73.5604 เปอร์เซ็นต์

ผลการสร้างตัวแบบพยากรณ์จากเทคนิคต้นไม้ตัดสินใจ LMT วิธีการเลือกแบบไปข้างหน้า

สำหรับผลลัพธ์ของตัวแบบพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม LMT วิธีการเลือกแบบไปข้างหน้าแสดงดังตาราง 50 และ 51 ตามลำดับ

ตาราง 50 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม LMT วิธีการเลือกแบบไปข้างหน้า ของชุดข้อมูลเรียนรู้ 70 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	503	306	62.1755
ไม่เป็นโรคเบาหวาน	299	2447	89.1114
ค่าความถูกต้องของตัวแบบพยากรณ์			82.9817

จากตาราง 50 พบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 82.9817 เปอร์เซ็นต์ และ 75.6435 เปอร์เซ็นต์

ตาราง 51 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม LMT วิธีการเลือกแบบไปข้างหน้า ของชุดข้อมูลทดสอบ 30 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	198	140	58.5799
ไม่เป็นโรคเบาหวาน	147	1041	87.6263
ค่าความถูกต้องของตัวแบบพยากรณ์			81.1927

จากตาราง 51 พบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 81.1927 เปอร์เซ็นต์ และ 73.1031 เปอร์เซ็นต์

ผลการสร้างตัวแบบพยากรณ์จากเทคนิคต้นไม้ตัดสินใจ LMT วิธีการกำจัดแบบถอยหลัง

สำหรับผลลัพธ์ของตัวแบบพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม LMT วิธีการกำจัดแบบถอยหลังแสดงดังตาราง 52 และ 53 ตามลำดับ

ตาราง 52 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม LMT วิธีการกำจัดแบบถอยหลัง ของชุดข้อมูลเรียนรู้ 70 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	482	284	62.9243
ไม่เป็นโรคเบาหวาน	320	2469	88.5264
ค่าความถูกต้องของตัวแบบพยากรณ์			83.0098

จากตาราง 52 พบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 83.0098 เปอร์เซ็นต์ และ 75.7253 เปอร์เซ็นต์

ตาราง 53 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม LMT วิธีการกำจัดแบบถอยหลัง ของชุดข้อมูลทดสอบ 30 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	186	121	60.5863
ไม่เป็นโรคเบาหวาน	159	1060	86.9565
ค่าความถูกต้องของตัวแบบพยากรณ์			81.6514

จากตาราง 53 พบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 81.6514 เปอร์เซ็นต์ และ 73.7714 เปอร์เซ็นต์

ผลการสร้างตัวแบบพยากรณ์จากเทคนิคต้นไม้ตัดสินใจ LMT วิธีการถดถอยทีละชั้น

สำหรับผลลัพธ์ของตัวแบบพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วย อัลกอริทึม LMT วิธีการถดถอยทีละชั้นแสดงดังตาราง 54 และ 55 ตามลำดับ

ตาราง 54 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วย อัลกอริทึม LMT วิธีการถดถอยทีละชั้น ของชุดข้อมูลเรียนรู้ 70 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	503	306	62.1755
ไม่เป็นโรคเบาหวาน	299	2447	89.1114
ค่าความถูกต้องของตัวแบบพยากรณ์			82.9817

จากตาราง 54 พบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 82.9817 เปอร์เซ็นต์ และ 75.6435 เปอร์เซ็นต์

ตาราง 55 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วย อัลกอริทึม LMT วิธีการถดถอยทีละชั้น ของชุดข้อมูลทดสอบ 30 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	198	140	58.5799
ไม่เป็นโรคเบาหวาน	147	1041	87.6263
ค่าความถูกต้องของตัวแบบพยากรณ์			81.1927

จากตาราง 55 พบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 81.1927 เปอร์เซ็นต์ และ 73.1031 เปอร์เซ็นต์

ผลการเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์จากเทคนิคต้นไม้ตัดสินใจด้วย
อัลกอริทึม LMT ของชุดข้อมูลแบบที่ 1

ตาราง 56 แสดงผลการเปรียบเทียบประสิทธิภาพการพยากรณ์ ด้วยเทคนิคต้นไม้ตัดสินใจด้วย
อัลกอริทึม LMT ของชุดข้อมูลแบบที่ 1

เทคนิคต้นไม้ตัดสินใจ อัลกอริทึม LMT	ชุดเรียนรู้		ชุดทดสอบ	
	ACC	BCA	ACC	BCA
กรณีตัวแปรอิสระทุกตัว	84.2757	77.5876	81.3893	73.3590
วิธีการเลือกแบบพื้นฐาน	82.8973	75.5512	81.5203	73.5604
วิธีการเลือกแบบไปข้างหน้า	82.9817	75.6435	81.1927	73.1031
วิธีการกำจัดแบบถอยหลัง	83.0098	75.7253	81.6514	73.7714
วิธีการถดถอยทีละขั้น	82.9817	75.6435	81.1927	73.1031

จากตาราง 56 พบว่าเทคนิคต้นไม้ตัดสินใจอัลกอริทึม LMT กรณีการคัดเลือกตัวแปรอิสระ
ด้วยวิธีการกำจัดแบบถอยหลังในเทคนิคการถดถอยลอจิสติกทวิภาค ประกอบไปด้วยตัวแปรอิสระ คือ
ค่าความดันขณะหัวใจบีบตัว ค่าความดันขณะหัวใจคลายตัว ค่าอัตราการเต้นของหัวใจ น้ำหนัก ความ
สูง ระดับน้ำตาลในเลือด ระดับไขมันในเลือดที่มีความหนาแน่นสูง และระดับไขมันในเลือดที่มีความ
หนาแน่นต่ำ มีค่าความถูกต้องในชุดข้อมูลทดสอบดีที่สุด 81.6514 เปอร์เซ็นต์ ค่าความถูกต้องในการ
จำแนกแบบสมดุล 73.7714 เปอร์เซ็นต์

4.2.4 เทคนิคโครงข่ายประสาทเทียม

กำหนดค่าพารามิเตอร์ของเทคนิคโครงข่ายประสาทเทียมดังนี้

1. ฟังก์ชันกระตุ้น คือฟังก์ชันซิกมอยด์
2. จำนวนชั้นซ่อน 1 ชั้น
3. จำนวนโหนดในชั้นซ่อน กำหนดเป็น 5, 6, 8, 9 และ 10
4. อัตราการเรียนรู้ คือ 0.1, 0.2 และ 0.3
5. ค่าโมเมนตัม คือ 0.1, 0.2 และ 0.3
6. จำนวนรอบ 100 รอบ

จากการกำหนดพารามิเตอร์ดังกล่าวข้างต้น ทำให้ต้องทำซ้ำ 45 ครั้ง ในการสร้างตัวแบบ
พยากรณ์ของแต่ละกรณีการเลือกตัวแปรอิสระทั้ง 5 รูปแบบ กรณีตัวแปรอิสระทุกตัว วิธีการเลือก
แบบพื้นฐาน วิธีการเลือกแบบไปข้างหน้า วิธีการกำจัดแบบถอยหลัง และวิธีการถดถอยทีละขั้น

ผลการสร้างตัวแบบพยากรณ์จากเทคนิคโครงข่ายประสาทเทียม กรณีตัวแปรอิสระทุกตัว

สำหรับผลลัพธ์ของตัวแบบพยากรณ์การเป็นโรคเบาหวานจากเทคนิคโครงข่ายประสาทเทียม กรณีตัวแปรอิสระทุกตัว โดยโครงข่ายประสาทเทียมที่มีประสิทธิภาพดีที่สุด คือตัวแบบพยากรณ์ที่มีพารามิเตอร์ดังนี้ จำนวนโหนดในชั้นซ่อน คือ 8 โหนด อัตราการเรียนรู้ คือ 0.3 โมเมนตัม คือ 0.2 แสดงดังตาราง 57 และ 58 ตามลำดับ

ตาราง 57 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคโครงข่ายประสาทเทียม กรณีตัวแปรอิสระทุกตัว ของชุดข้อมูลเรียนรู้ 70 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	283	146	65.9674
ไม่เป็นโรคเบาหวาน	519	2607	83.3973
ค่าความถูกต้องของตัวแบบพยากรณ์			81.2940

จากตาราง 57 พบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 81.2940 เปอร์เซ็นต์ และ 74.6823 เปอร์เซ็นต์

ตาราง 58 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคโครงข่ายประสาทเทียม กรณีตัวแปรอิสระทุกตัว ของชุดข้อมูลทดสอบ 30 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	159	102	60.9195
ไม่เป็นโรคเบาหวาน	186	1079	85.2964
ค่าความถูกต้องของตัวแบบพยากรณ์			81.1271

จากตาราง 58 มีพบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 81.1271 เปอร์เซ็นต์ และ 73.1080 เปอร์เซ็นต์

ผลการสร้างตัวแบบพยากรณ์จากเทคนิคโครงข่ายประสาทเทียมวิธีการเลือกแบบพื้นฐาน

สำหรับผลลัพธ์ของตัวแบบพยากรณ์การเป็นโรคเบาหวานจากเทคนิคโครงข่ายประสาทเทียมวิธีการเลือกแบบพื้นฐานโดยโครงข่ายประสาทเทียมที่มีประสิทธิภาพดีที่สุด คือตัวแบบพยากรณ์ที่มีพารามิเตอร์ดังนี้ จำนวนโหนดในชั้นซ่อน คือ 6 โหนด อัตราการเรียนรู้ คือ 0.3 โมเมนตัม คือ 0.3 แสดงดังตาราง 59 และ 60 ตามลำดับ

ตาราง 59 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคโครงข่ายประสาทเทียมวิธีการเลือกแบบพื้นฐาน ของชุดข้อมูลเรียนรู้ 70 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	348	189	64.8045
ไม่เป็นโรคเบาหวาน	454	2564	84.9569
ค่าความถูกต้องของตัวแบบพยากรณ์			81.9128

จากตาราง 59 พบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 81.9128 เปอร์เซ็นต์ และ 74.8807 เปอร์เซ็นต์

ตาราง 60 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคโครงข่ายประสาทเทียมวิธีการเลือกแบบพื้นฐาน ของชุดข้อมูลทดสอบ 30 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	185	122	60.2606
ไม่เป็นโรคเบาหวาน	160	1059	86.8745
ค่าความถูกต้องของตัวแบบพยากรณ์			81.5203

จากตาราง 60 มีพบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 81.5203 เปอร์เซ็นต์ และ 73.5675 เปอร์เซ็นต์

ผลการสร้างตัวแบบพยากรณ์จากเทคนิคโครงข่ายประสาทเทียมวิธีการเลือกแบบไปข้างหน้า

สำหรับผลลัพธ์ของตัวแบบพยากรณ์การเป็นโรคเบาหวาน จากเทคนิคโครงข่ายประสาทเทียมวิธีการเลือกแบบไปข้างหน้า โดยโครงข่ายประสาทเทียมที่มีประสิทธิภาพดีที่สุดในตัวแบบพยากรณ์ที่มีพารามิเตอร์ดังนี้ จำนวนโหนดในชั้นซ่อน คือ 6 โหนด อัตราการเรียนรู้ คือ 0.2 โมเมนตัม คือ 0.3 แสดงดังตาราง 61 และ 62 ตามลำดับ

ตาราง 61 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคโครงข่ายประสาทเทียมวิธีการเลือกแบบไปข้างหน้า ของชุดข้อมูลเรียนรู้ 70 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	336	182	64.8649
ไม่เป็นโรคเบาหวาน	466	2571	84.6559
ค่าความถูกต้องของตัวแบบพยากรณ์			81.7721

จากตาราง 61 พบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 81.7721 เปอร์เซ็นต์ และ 74.7604 เปอร์เซ็นต์

ตาราง 62 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคโครงข่ายประสาทเทียมวิธีการเลือกแบบไปข้างหน้า ของชุดข้อมูลทดสอบ 30 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	188	121	60.8414
ไม่เป็นโรคเบาหวาน	157	1060	87.0994
ค่าความถูกต้องของตัวแบบพยากรณ์			81.7824

จากตาราง 62 พบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 81.7824เปอร์เซ็นต์ และ 73.9704 เปอร์เซ็นต์

ผลการสร้างตัวแบบพยากรณ์จากเทคนิคโครงข่ายประสาทเทียมวิธีการกำจัดแบบถอยหลัง

สำหรับผลลัพธ์ของตัวแบบพยากรณ์การเป็นโรคเบาหวานจากเทคนิคโครงข่ายประสาทเทียมวิธีการกำจัดแบบถอยหลัง โดยโครงข่ายประสาทเทียมที่มีประสิทธิภาพดีที่สุด คือตัวแบบพยากรณ์ที่มีพารามิเตอร์ดังนี้ จำนวนโหนดในชั้นซ่อน คือ 6 โหนด อัตราการเรียนรู้ คือ 0.3 โมเมนตัม คือ 0.3 แสดงดังตาราง 63 และ 64 ตามลำดับ

ตาราง 63 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคโครงข่ายประสาทเทียมวิธีการกำจัดแบบถอยหลัง ของชุดข้อมูลเรียนรู้ 70 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	267	140	65.6020
ไม่เป็นโรคเบาหวาน	535	2613	83.0051
ค่าความถูกต้องของตัวแบบพยากรณ์			81.0127

จากตาราง 63 พบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 81.0127 เปอร์เซ็นต์ และ 74.3035 เปอร์เซ็นต์

ตาราง 64 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคโครงข่ายประสาทเทียมวิธีการกำจัดแบบถอยหลัง ของชุดข้อมูลทดสอบ 30 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	157	98	61.5686
ไม่เป็นโรคเบาหวาน	188	1083	85.2085
ค่าความถูกต้องของตัวแบบพยากรณ์			81.2582

จากตาราง 64 พบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 81.2582 เปอร์เซ็นต์ และ 73.3886 เปอร์เซ็นต์

ผลการสร้างตัวแบบพยากรณ์จากเทคนิคโครงข่ายประสาทเทียมวิธีการถดถอยทีละชั้น

สำหรับผลลัพธ์ของตัวแบบพยากรณ์การเป็นโรคเบาหวานจากเทคนิคโครงข่ายประสาทเทียมวิธีการถดถอยทีละชั้น โดยโครงข่ายประสาทเทียมที่มีประสิทธิภาพดีที่สุดในชุดข้อมูลที่มีพารามิเตอร์ดังนี้ จำนวนโหนดในชั้นซ่อน คือ 6 โหนด อัตราการเรียนรู้ คือ 0.2 โมเมนตัม คือ 0.3 แสดงดังตาราง 65 และ 66 ตามลำดับ

ตาราง 65 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคโครงข่ายประสาทเทียม วิธีการถดถอยทีละชั้น ของชุดข้อมูลเรียนรู้ 70 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	336	182	64.8649
ไม่เป็นโรคเบาหวาน	466	2571	84.6559
ค่าความถูกต้องของตัวแบบพยากรณ์			81.7721

จากตาราง 65 พบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 81.7721 เปอร์เซ็นต์ และ 74.7604 เปอร์เซ็นต์

ตาราง 66 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคโครงข่ายประสาทเทียม วิธีการถดถอยทีละชั้น ของชุดข้อมูลทดสอบ 30 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	188	121	60.8414
ไม่เป็นโรคเบาหวาน	157	1060	87.0994
ค่าความถูกต้องของตัวแบบพยากรณ์			81.7824

จากตาราง 66 พบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 81.7824เปอร์เซ็นต์ และ 73.9704 เปอร์เซ็นต์

ผลการเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์จากเทคนิคโครงข่ายประสาทเทียม ของชุดข้อมูลแบบที่ 1

ตาราง 67 แสดงผลการเปรียบเทียบประสิทธิภาพการพยากรณ์ ด้วยเทคนิคโครงข่ายประสาทเทียม ของชุดข้อมูลแบบที่ 1

เทคนิคโครงข่ายประสาทเทียม (จำนวนชั้นซ่อน, อัตราการเรียนรู้, โมเมนตัม)	ชุดเรียนรู้		ชุดทดสอบ	
	ACC	BCA	ACC	BCA
กรณีตัวแปรอิสระทุกตัว (8, 0.3, 0.2)	81.2940	74.6823	81.1271	73.1080
วิธีการเลือกแบบพื้นฐาน (6, 0.3, 0.3)	81.9128	74.8807	81.5203	73.5675
วิธีการเลือกแบบไปข้างหน้า (6, 0.2, 0.3)	81.7721	74.7604	81.7824	73.9704
วิธีการกำจัดแบบถอยหลัง (6, 0.3, 0.3)	81.0127	74.3035	81.2582	73.3886
วิธีการถดถอยทีละชั้น (6, 0.2, 0.3)	81.7721	74.7604	81.7824	73.9704

จากตาราง 67 พบว่าเทคนิคโครงข่ายประสาทเทียม กรณีการคัดเลือกตัวแปรอิสระด้วยวิธีการเลือกแบบไปข้างหน้า และวิธีการถดถอยทีละชั้นในเทคนิคการถดถอยลอจิสติกทวิภาค ประกอบไปด้วยตัวแปรอิสระ คือ ระดับน้ำตาลในเลือด ดัชนีมวลกาย ระดับคอเลสเตอรอลในเลือด ค่าความดันขณะหัวใจบีบตัว ค่าความดันขณะหัวใจคลายตัว และค่าอัตราการเต้นของหัวใจ โดยมีจำนวนชั้นซ่อน อัตราการเรียนรู้ และโมเมนตัม คือ 6, 0.2 และ 0.3 ตามลำดับ มีค่าความถูกต้องในชุดข้อมูลทดสอบดีที่สุด 81.7824 เปอร์เซ็นต์ ค่าความถูกต้องในการจำแนกแบบสมดุล 73.9704 เปอร์เซ็นต์ และในชุดเรียนรู้ 81.7721 เปอร์เซ็นต์ ค่าความถูกต้องในการจำแนกแบบสมดุล 74.7604 เปอร์เซ็นต์

4.3 ผลการเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์ในชุดข้อมูลแบบที่ 1 (70:30)

ตาราง 68 แสดงผลการเปรียบเทียบประสิทธิภาพตัวแบบพยากรณ์ด้วยเทคนิคต่าง ๆ ของชุดข้อมูลแบบที่ 1

	ชุดข้อมูลทดสอบ	
	ACC	BCA
การถดถอยลอจิสติกทวิภาค (วิธีการเลือกแบบพื้นฐาน)	80.6029	62.5387
ต้นไม้ตัดสินใจด้วยอัลกอริทึม J48 (วิธีการกำจัดแบบถอยหลัง)	81.6514	73.8231
ต้นไม้ตัดสินใจด้วยอัลกอริทึม LMT (วิธีการกำจัดแบบถอยหลัง)	81.6514	73.7714
โครงข่ายประสาทเทียม (วิธีการเลือกแบบไปข้างหน้า, วิธีการถดถอยทีละชั้น)	81.7824	73.9704
จำนวนโหนดในชั้นซ่อน, อัตราการเรียนรู้, โมเมนตัม (6, 0.2, 0.3)		

จากตาราง 68 พบว่าเทคนิคโครงข่ายประสาทเทียม มีประสิทธิภาพดีที่สุดในชุดข้อมูลแบบที่ 1 ชุดข้อมูลเรียนรู้ 70 เปอร์เซ็นต์ และชุดข้อมูลทดสอบ 30 เปอร์เซ็นต์

4.4 ผลการสร้างและพัฒนาตัวแบบพยากรณ์ชุดข้อมูลแบบที่ 2 (80:20)

4.4.1 เทคนิคการถดถอยลอจิสติกทวิภาค

ผู้วิจัยเริ่มสร้างตัวแบบพยากรณ์ในชุดข้อมูลแบบที่ 2 ซึ่งมีข้อมูลเรียนรู้ 80 เปอร์เซ็นต์ ชุดข้อมูลทดสอบ 20 เปอร์เซ็นต์ และใช้วิธีการคัดเลือกตัวแปรอิสระซึ่งประกอบไปด้วย วิธีการเลือกแบบพื้นฐาน วิธีการเลือกแบบไปข้างหน้า วิธีการกำจัดแบบถอยหลัง และวิธีการถดถอยทีละชั้น กำหนดระดับนัยสำคัญ 0.05

ผลการสร้างตัวแบบพยากรณ์วิธีการเลือกแบบพื้นฐาน

ตาราง 69 แสดงผลลัพธ์จากการสร้างตัวแบบพยากรณ์ด้วยเทคนิคการถดถอยลอจิสติกทวิภาค วิธีการเลือกแบบพื้นฐาน ของชุดข้อมูลแบบที่ 2

ตัวแปรอิสระ	β	S.E.	Wald	OR	p
ระดับน้ำตาลในเลือด (X_9)	0.0304	0.0015	410.7378	1.0308	<0.001*
น้ำหนัก(X_6)	0.0243	0.0034	51.0804	1.0246	<0.001*
ความสูง (X_7)	-0.0250	0.0057	19.2367	0.9753	<0.001*
ค่าความดันขณะหัวใจคลายตัว(X_4)	-0.0308	0.0041	56.4331	0.9697	<0.001*
ค่าความดันขณะหัวใจบีบตัว (X_3)	0.0136	0.0025	29.5936	1.0137	<0.001*
อัตราการเต้นของหัวใจ (X_5)	0.0141	0.0033	18.2562	1.0142	<0.001*

ตาราง 69 (ต่อ)

ตัวแปรอิสระ	β	S.E.	Wald	OR	p
ค่าคงที่ (Constant)	-3.2396	0.9343	12.0236	0.0392	<0.001*

$-2 \text{ Log likelihood} = 3435.551$ (df = 7)
 AIC = 3449.551
 Cox & Snell $R^2 = 0.1995$, Nagelkerke $R^2 = 0.3040$

* มีนัยสำคัญเชิงสถิติ 0.05

จากตาราง 69 สามารถวิเคราะห์ที่ได้ดังนี้ ตัวแปรอิสระที่ส่งผลให้เกิดโรคเบาหวาน ได้แก่ ระดับน้ำตาลในเลือด น้ำหนัก ความสูง ค่าความดันขณะหัวใจคลายตัว ค่าความดันขณะหัวใจบีบตัว และอัตราการเต้นของหัวใจอย่างมีนัยสำคัญเชิงสถิติ 0.05

จากค่าอัตราส่วนออดส์ ซึ่งจะเห็นว่าเมื่อระดับน้ำตาลในเลือด น้ำหนัก ค่าความดันขณะหัวใจบีบตัว และอัตราการเต้นของหัวใจ ถ้าค่าเหล่านี้เพิ่มขึ้น 1 หน่วยจะทำให้มีโอกาสในการเป็นโรคเบาหวานเพิ่มขึ้นประมาณ 1 เท่า ในทางตรงกันข้ามเมื่อความสูง และค่าความดันขณะหัวใจคลายตัว ถ้าค่าเหล่านี้เพิ่มขึ้น 1 หน่วย ทำให้มีโอกาสในการเป็นโรคเบาหวานลดลงประมาณ 0.02 เท่า โดยสามารถนำมาเขียนสมการตัวแบบพยากรณ์ได้ดังสมการ (38)

$$P(Y = 1) = \pi_i = \frac{e^{w_5}}{1 + e^{w_5}} \quad (38)$$

เมื่อ w_5 คือ $-3.2396 + 0.0136X_3 - 0.0308X_4 + 0.0141X_5 + 0.0243X_6 - 0.0250X_7 + 0.0304X_9$

และเมื่อพิจารณาค่า $-2 \text{ Log likelihood}$ มีค่าเท่ากับ 3435.551 และค่า AIC เท่ากับ 3449.551 เมื่อพิจารณาความสามารถในการพยากรณ์ตัวแบบนี้ พบว่าสามารถอธิบายความผันแปรของการเป็นโรคเบาหวานได้ร้อยละ 19.95 จากนั้นผู้วิจัยทำการตรวจสอบความสัมพันธ์ของตัวแปรอิสระได้ผลลัพธ์ดังตาราง 70

ตาราง 70 แสดงการตรวจสอบความสัมพันธ์ระหว่างตัวแปรอิสระวิธีการเลือกแบบพื้นฐาน

ตัวแปรอิสระ	Tolerance	VIF
ระดับน้ำตาลในเลือด (X_9)	0.9276	1.0781
น้ำหนัก (X_6)	0.6970	1.4347
ความสูง (X_7)	0.7202	1.3885
ค่าความดันขณะหัวใจคลายตัว (X_4)	0.7136	1.4014
ค่าความดันขณะหัวใจบีบตัว (X_3)	0.7413	1.3490
อัตราการเต้นของหัวใจ (X_5)	0.9434	1.0600

จากตาราง 70 พบว่าค่า Variance Inflation Factor (VIF) และค่าความคลาดเคลื่อน ยินยอม (Tolerance) ของตัวแปรอิสระไม่มีความสัมพันธ์กัน

ผลการสร้างตัวแบบพยากรณ์จากเทคนิคการถดถอยลอจิสติกทวิภาควิธีการเลือกแบบ พื้นฐาน

สำหรับผลลัพธ์ของตัวแบบพยากรณ์การเป็นโรคเบาหวานจากเทคนิคการถดถอยลอจิสติก ทวิภาควิธีการเลือกแบบพื้นฐาน แสดงดังตาราง 71 และ 72 ตามลำดับ

ตาราง 71 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคการถดถอยลอจิสติก ทวิภาควิธีการเลือกแบบพื้นฐาน ของชุดข้อมูลเรียนรู้ 80 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	294	623	32.0611
ไม่เป็นโรคเบาหวาน	151	2996	95.2018
ค่าความถูกต้องของตัวแบบพยากรณ์			80.9547

จากตาราง 71 พบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 80.9547 เปอร์เซ็นต์ และ 63.6314 เปอร์เซ็นต์

ตาราง 72 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคการถดถอยลอจิสติก ทวิภาควิธีการเลือกแบบพื้นฐาน ของชุดข้อมูลทดสอบ 20 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	63	167	27.3913
ไม่เป็นโรคเบาหวาน	39	755	95.0882
ค่าความถูกต้องของตัวแบบพยากรณ์			80.4326

จากตาราง 72 พบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 80.4326 เปอร์เซ็นต์ และ 61.6626 เปอร์เซ็นต์

ผลการสร้างตัวแบบพยากรณ์วิธีการเลือกแบบไปข้างหน้า

ตาราง 73 แสดงผลการคัดเลือกตัวแปรอิสระวิธีการเลือกแบบไปข้างหน้า ของชุดข้อมูลแบบที่ 2

ตัวแปรอิสระ	β	S.E.	Wald	OR	p
ระดับน้ำตาลในเลือด (X_9)	0.0299	0.0015	422.6819	1.0303	<0.001*
ดัชนีมวลกาย (X_8)	0.0530	0.0085	39.1274	1.0545	<0.001*
ระดับคอเลสเตอรอลในเลือด (X_{10})	-0.0068	0.0012	30.2992	0.9933	<0.001*
ค่าความดันขณะหัวใจคลายตัว (X_4)	-0.0283	0.0040	51.0333	0.9721	<0.001*
ค่าความดันขณะหัวใจบีบตัว (X_3)	0.0134	0.0025	28.1849	1.0134	<0.001*
อัตราการเต้นของหัวใจ (X_5)	0.0152	0.0033	20.8470	1.0153	<0.001*
ค่าคงที่ (Constant)	-5.9794	0.4899	148.9520	0.0025	<0.001*

$-2 \text{ Log likelihood} = 3416.357$ ($df = 7$)
 AIC = 3430.357
 Cox & Snell $R^2 = 0.2033$, Nagelkerke $R^2 = 0.3098$

* มีนัยสำคัญเชิงสถิติที่ระดับ 0.05

จากตาราง 73 สามารถวิเคราะห์ได้ดังนี้ ตัวแปรอิสระที่ส่งผลให้เกิดโรคเบาหวาน ได้แก่ ระดับน้ำตาลในเลือด ดัชนีมวลกาย ระดับคอเลสเตอรอลในเลือด ค่าความดันขณะหัวใจคลายตัว ค่าความดันขณะหัวใจบีบตัว และค่าอัตราการเต้นของหัวใจ อย่างมีนัยสำคัญเชิงสถิติ 0.05

จากค่าอัตราส่วนออดส์ ซึ่งจะเห็นว่าเมื่อระดับน้ำตาลในเลือด ดัชนีมวลกาย ค่าความดันขณะหัวใจบีบตัว และอัตราการเต้นของหัวใจ ถ้าค่าเหล่านี้เพิ่มขึ้น 1 หน่วยจะทำให้มีโอกาสในการเป็นโรคเบาหวานเพิ่มขึ้นประมาณ 1 เท่า ในทางตรงกันข้ามเมื่อระดับคอเลสเตอรอลในเลือด และค่าความดันขณะหัวใจคลายตัว ถ้าค่าเหล่านี้เพิ่มขึ้น 1 หน่วย ทำให้มีโอกาสในการเป็นโรคเบาหวานลดลงประมาณ 0.03 เท่า

โดยสามารถนำมาเขียนสมการตัวแบบพยากรณ์ได้ดังสมการ (39)

$$P(Y = 1) = \pi_i = \frac{e^{w_6}}{1 + e^{w_6}} \quad (39)$$

เมื่อ w_6 คือ $-5.9794 + 0.0134X_3 - 0.0283X_4 + 0.0152X_5 + 0.0530X_8 + 0.0299X_9 - 0.0068X_{10}$

และเมื่อพิจารณาค่า $-2 \text{ Log likelihood}$ มีค่าเท่ากับ 3416.357 และค่า AIC เท่ากับ 3430.357 เมื่อพิจารณาความสามารถในการพยากรณ์ตัวแบบนี้ พบว่าสามารถอธิบายความผันแปรของการเป็นโรคเบาหวานได้ร้อยละ 20.33 จากนั้นผู้วิจัยทำการตรวจสอบความสัมพันธ์ของตัวแปรอิสระได้ผลลัพธ์ดังตาราง 74

ตาราง 74 แสดงการตรวจสอบความสัมพันธ์ระหว่างตัวแปรอิสระวิธีการเลือกแบบไปข้างหน้า
ของชุดข้อมูลแบบที่ 2

ตัวแปรอิสระ	Tolerance	VIF
ระดับน้ำตาลในเลือด (X_9)	0.9268	1.0790
ดัชนีมวลกาย (X_8)	0.9370	1.0672
ระดับคอเลสเตอรอลในเลือด (X_{10})	0.9813	1.0191
ค่าความดันขณะหัวใจคลายตัว (X_4)	0.7438	1.3444
ค่าความดันขณะหัวใจบีบตัว (X_3)	0.7549	1.3247
อัตราการเต้นของหัวใจ (X_5)	0.9466	1.0564

จากตาราง 74 พบว่าค่า Variance Inflation Factor (VIF) และค่าความคลาดเคลื่อน
ยินยอม (Tolerance) ของตัวแปรอิสระทุกตัวไม่มีความสัมพันธ์กัน

ผลการสร้างตัวแบบพยากรณ์จากเทคนิคการถดถอยลอจิสติกทวิภาคีวิธีการเลือกแบบ ไปข้างหน้า

สำหรับผลลัพธ์ของตัวแบบพยากรณ์การเป็นโรคเบาหวานจากเทคนิคการถดถอยลอจิสติก
ทวิภาคีวิธีการเลือกแบบไปข้างหน้า แสดงดังตาราง 75 และ 76 ตามลำดับ

ตาราง 75 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคการถดถอยลอจิสติก
ทวิภาคีวิธีการเลือกแบบไปข้างหน้า ของชุดข้อมูลเรียนรู้ 80 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	291	626	31.7339
ไม่เป็นโรคเบาหวาน	157	2990	95.0111
ค่าความถูกต้องของตัวแบบพยากรณ์			80.7333

จากตาราง 75 มีพบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ
80.7333 เปอร์เซ็นต์ และ 63.3725 เปอร์เซ็นต์

ตาราง 76 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคการถดถอยลอจิสติก
ทวิภาควิธีการเลือกแบบไปข้างหน้า ของชุดข้อมูลทดสอบ 20 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	62	168	26.9565
ไม่เป็นโรคเบาหวาน	32	755	95.9339
ค่าความถูกต้องของตัวแบบพยากรณ์			80.3343

จากตาราง 76 พบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 80.3343 เปอร์เซ็นต์ และ 61.4452 เปอร์เซ็นต์

ผลการสร้างตัวแบบพยากรณ์ด้วยวิธีการกำจัดแบบถอยหลัง

ตาราง 77 แสดงผลการคัดเลือกตัวแปรอิสระวิธีการกำจัดแบบถอยหลัง ของชุดข้อมูลแบบที่ 2

ตัวแปรอิสระ	β	S.E.	Wald	OR	p
ค่าความดันขณะหัวใจบีบตัว (X_3)	0.0129	0.0025	25.4651	1.0129	<0.001*
ค่าความดันขณะหัวใจคลายตัว (X_4)	-0.0287	0.0041	49.2263	0.9717	<0.001*
อัตราการเต้นของหัวใจ (X_5)	0.0148	0.0033	19.4263	1.0149	<0.001*
น้ำหนัก (X_6)	0.0235	0.0035	46.2676	1.0237	<0.001*
ความสูง (X_7)	-0.0256	0.0057	19.8927	0.9747	<0.001*
ระดับน้ำตาลในเลือด (X_9)	0.0293	0.0015	401.6730	1.0297	<0.001*
ระดับไขมันในเลือดที่มีความหนาแน่นสูง (X_{12})	-0.0092	0.0034	7.1520	0.9908	0.0075*
ระดับไขมันในเลือดที่มีความหนาแน่นต่ำ (X_{13})	-0.0069	0.0014	24.8990	0.9931	<0.001*
ค่าคงที่ (Constant)	-1.9231	0.9774	3.8712	0.1462	0.0491*

$-2 \text{ Log likelihood} = 3400.367$ ($df = 9$)

AIC = 3418.367

Cox & Snell $R^2 = 0.2064$, Nagelkerke $R^2 = 0.3145$

* มีนัยสำคัญเชิงสถิติที่ระดับ 0.05

จากตาราง 77 สามารถวิเคราะห์ได้ดังนี้ ตัวแปรอิสระที่ส่งผลให้เกิดโรคเบาหวาน ได้แก่ ค่าความดันขณะหัวใจบีบตัว ค่าความดันขณะหัวใจคลายตัว ค่าอัตราการเต้นของหัวใจ น้ำหนัก ความ

สูง ระดับน้ำตาลในเลือด ระดับไขมันในเลือดที่มีความหนาแน่นสูง และระดับไขมันในเลือดที่มีความหนาแน่นต่ำ อย่างมีนัยสำคัญเชิงสถิติ 0.05

จากค่าอัตราส่วนออดส์ ซึ่งจะเห็นว่าเมื่อค่าความดันขณะหัวใจบีบตัว ค่าอัตราการเต้นของหัวใจ น้ำหนัก และระดับน้ำตาลในเลือด ถ้าค่าเหล่านี้เพิ่มขึ้น 1 หน่วยจะทำให้มีโอกาสในการเป็นโรคเบาหวานเพิ่มขึ้นประมาณ 1 เท่า ในทางตรงกันข้ามเมื่อค่าความดันขณะหัวใจคลายตัว ความสูง ระดับไขมันในเลือดที่มีความหนาแน่นสูง และระดับไขมันในเลือดที่มีความหนาแน่นต่ำ ถ้าค่าเหล่านี้เพิ่มขึ้น 1 หน่วย ทำให้มีโอกาสในการเป็นโรคเบาหวานลดลงประมาณ 0.03 เท่า

โดยสามารถนำมาเขียนสมการตัวแบบพยากรณ์ได้ดังสมการ (40)

$$P(Y = 1) = \pi_i = \frac{e^{w_7}}{1 + e^{w_7}} \quad (40)$$

เมื่อ w_7 คือ $-1.9231 + 0.0129X_3 - 0.0287X_4 + 0.0148X_5 + 0.0235X_6 - 0.0256X_7 + 0.0293X_9 - 0.0092X_{12} - 0.0069X_{13}$

และเมื่อพิจารณาค่า $-2 \text{ Log likelihood}$ มีค่าเท่ากับ 3400.367 และค่า AIC เท่ากับ 3418.367 เมื่อพิจารณาความสามารถในการพยากรณ์ตัวแบบนี้ พบว่าสามารถอธิบายความผันแปรของการเป็นโรคเบาหวานได้ร้อยละ 20.64 จากนั้นผู้วิจัยทำการตรวจสอบความสัมพันธ์ของตัวแปรอิสระได้ผลลัพธ์ดังตาราง 78

ตาราง 78 แสดงการตรวจสอบความสัมพันธ์ระหว่างตัวแปรอิสระด้วยวิธีการกำจัดแบบถอยหลัง

ตัวแปรอิสระ	Tolerance	VIF
ค่าความดันขณะหัวใจบีบตัว (X_3)	0.7360	1.3587
ค่าความดันขณะหัวใจคลายตัว (X_4)	0.7057	1.4171
อัตราการเต้นของหัวใจ (X_5)	0.9424	1.0611
น้ำหนัก (X_6)	0.6696	1.4934
ความสูง (X_7)	0.7197	1.3895
ระดับน้ำตาลในเลือด (X_9)	0.9097	1.0992
ระดับไขมันในเลือดที่มีความหนาแน่นสูง (X_{12})	0.9125	1.0959
ระดับไขมันในเลือดที่มีความหนาแน่นต่ำ (X_{13})	0.9741	1.0266

ตาราง 78 พบว่าค่า Variance Inflation Factor (VIF) และค่าความคลาดเคลื่อนยินยอม (Tolerance) ของตัวแปรอิสระทุกตัวไม่มีความสัมพันธ์กัน

**ผลการสร้างตัวแบบพยากรณ์จากเทคนิคการถดถอยลอจิสติกทวิภาควิธีการกำจัดแบบ
ถอยหลัง**

สำหรับผลลัพธ์ของตัวแบบพยากรณ์การเป็นโรคเบาหวาน จากเทคนิคการถดถอยลอจิสติก
ทวิภาคด้วยวิธีการกำจัดแบบถอยหลัง แสดงดังตาราง 79 และ 80 ตามลำดับ

**ตาราง 79 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคการถดถอยลอจิสติก
ทวิภาควิธีการกำจัดแบบถอยหลัง ของชุดข้อมูลเรียนรู้ 80 เปอร์เซ็นต์**

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	296	621	67.7208
ไม่เป็นโรคเบาหวาน	158	2989	94.9793
ค่าความถูกต้องของตัวแบบพยากรณ์			80.8317

จากตาราง 79 พบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุลคือ
80.8317 เปอร์เซ็นต์ และ 63.6293 เปอร์เซ็นต์

**ตาราง 80 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคการถดถอยลอจิสติก
ทวิภาควิธีการกำจัดแบบถอยหลัง ของชุดข้อมูลทดสอบ 20 เปอร์เซ็นต์**

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	62	168	26.9565
ไม่เป็นโรคเบาหวาน	36	751	95.4257
ค่าความถูกต้องของตัวแบบพยากรณ์			79.9410

จากตาราง 80 พบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ
79.9410 เปอร์เซ็นต์ และ 61.1911 เปอร์เซ็นต์

ผลการสร้างตัวแบบพยากรณ์วิธีการถดถอยทีละขั้น

ตาราง 81 แสดงผลการคัดเลือกตัวแปรอิสระวิธีการถดถอยทีละขั้น ของชุดข้อมูลแบบที่ 2

ตัวแปรอิสระ	β	S.E.	Wald	OR	p
ระดับน้ำตาลในเลือด (X_9)	0.0299	0.0015	422.6819	1.0303	<0.001*
ดัชนีมวลกาย (X_8)	0.0530	0.0085	39.1274	1.0545	<0.001*
ระดับคอเลสเตอรอลในเลือด (X_{10})	-0.0068	0.0012	30.2992	0.9933	<0.001*
ค่าความดันขณะหัวใจคลายตัว (X_4)	-0.0283	0.0040	51.0333	0.9721	<0.001*
ค่าความดันขณะหัวใจบีบตัว (X_3)	0.0134	0.0025	28.1849	1.0134	<0.001*
อัตราการเต้นของหัวใจ (X_5)	0.0152	0.0033	20.8470	1.0153	<0.001*
ค่าคงที่ (Constant)	-5.9794	0.4899	148.9520	0.0025	<0.001*

$-2 \text{ Log likelihood} = 3416.357$ ($df = 7$)
 AIC = 3430.357
 Cox & Snell $R^2 = 0.2033$, Nagelkerke $R^2 = 0.3098$

* มีนัยสำคัญเชิงสถิติ 0.05

จากตาราง 81 สามารถวิเคราะห์ได้ดังนี้ ตัวแปรอิสระที่ส่งผลให้เกิดโรคเบาหวาน ได้แก่ ระดับน้ำตาลในเลือด ดัชนีมวลกาย ระดับคอเลสเตอรอลในเลือด ค่าความดันขณะหัวใจคลายตัว ค่าความดันขณะหัวใจบีบตัว และค่าอัตราการเต้นของหัวใจ อย่างมีนัยสำคัญเชิงสถิติ 0.05

จากค่าอัตราส่วนออดส์ ซึ่งจะเห็นว่าเมื่อระดับน้ำตาลในเลือด ดัชนีมวลกาย ค่าความดันขณะหัวใจบีบตัว และอัตราการเต้นของหัวใจถ้าค่าเหล่านี้เพิ่มขึ้น 1 หน่วยจะทำให้มีโอกาสในการเป็นโรคเบาหวานเพิ่มขึ้นประมาณ 1 เท่า ในทางตรงกันข้ามเมื่อระดับคอเลสเตอรอลในเลือด และค่าความดันขณะหัวใจคลายตัว ถ้าค่าเหล่านี้เพิ่มขึ้น 1 หน่วย ทำให้มีโอกาสในการเป็นโรคเบาหวานลดลงประมาณ 0.03 เท่า

โดยสามารถนำมาเขียนสมการตัวแบบพยากรณ์ได้ดังสมการ (41)

$$P(Y = 1) = \pi_i = \frac{e^{w_8}}{1 + e^{w_8}} \quad (41)$$

เมื่อ w_8 คือ $-5.9794 + 0.0134X_3 - 0.0283X_4 + 0.0152X_5 + 0.0530X_8 + 0.0299X_9 - 0.0068X_{10}$

และเมื่อพิจารณาค่า $-2 \text{ Log likelihood}$ มีค่าเท่ากับ 3416.357 และค่า AIC เท่ากับ 3430.357

เมื่อพิจารณาความสามารถในการพยากรณ์ตัวแบบนี้ พบว่าสามารถอธิบายความผันแปรของการเป็นโรคเบาหวานได้ร้อยละ 20.33 จากนั้นผู้วิจัยทำการตรวจสอบความสัมพันธ์ของตัวแปรอิสระได้ผลลัพธ์ดังตาราง 82

ตาราง 82 แสดงการตรวจสอบความสัมพันธ์ระหว่างตัวแปรอิสระด้วยวิธีการถดถอยทีละขั้น

ตัวแปรอิสระ	Tolerance	VIF
ระดับน้ำตาลในเลือด (X_9)	0.9237	1.0825
ดัชนีมวลกาย (X_8)	0.9381	1.0659
ระดับคอเลสเตอรอลในเลือด (X_{10})	0.9794	1.0210
ค่าความดันขณะหัวใจคลายตัว (X_4)	0.7329	1.3644
ค่าความดันขณะหัวใจบีบตัว (X_3)	0.7452	1.3419
อัตราการเต้นของหัวใจ (X_5)	0.9448	1.0585

จากตาราง 82 พบว่าค่า Variance Inflation Factor (VIF) และค่าความคลาดเคลื่อน ยินยอม (Tolerance) ของตัวแปรอิสระทุกตัวไม่มีความสัมพันธ์กัน

ผลการสร้างตัวแบบพยากรณ์จากเทคนิคการถดถอยลอจิสติกทวิภาคีวิธีการถดถอยทีละขั้น

สำหรับผลลัพธ์ของตัวแบบพยากรณ์การเป็นโรคเบาหวานจากเทคนิคการถดถอยลอจิสติก ทวิภาคีวิธีการถดถอยทีละขั้น แสดงดังตาราง 83 และ 84 ตามลำดับ

ตาราง 83 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคการถดถอยลอจิสติก ทวิภาคีวิธีการถดถอยทีละขั้น ของชุดข้อมูลเรียนรู้ 80 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	291	626	31.7339
ไม่เป็นโรคเบาหวาน	157	2990	95.0111
ค่าความถูกต้องของตัวแบบพยากรณ์			80.7333

จากตาราง 83 มีพบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 80.7333 เปอร์เซ็นต์ และ 63.3725 เปอร์เซ็นต์

ตาราง 84 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคการถดถอยลอจิสติก
ทวิภาควิธีถดถอยทีละชั้น ของชุดข้อมูลทดสอบ 20 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	62	168	26.9565
ไม่เป็นโรคเบาหวาน	32	755	95.9339
ค่าความถูกต้องของตัวแบบพยากรณ์			80.3343

จากตาราง 84 พบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 80.3343 เปอร์เซ็นต์ และ 61.4452 เปอร์เซ็นต์

หลังจากนั้นผู้วิจัยทำการตรวจสอบความเหมาะสมของตัวแบบการถดถอยลอจิสติกทวิภาค โดยใช้การทดสอบอัตราส่วนภาวะน่าจะเป็น และการทดสอบไคกำลังสอง และค่าเกณฑ์สารสนเทศอะกะอิเกะ ได้ผลลัพธ์ดังตาราง 85

ตาราง 85 แสดงผลการเปรียบเทียบการตรวจสอบความเหมาะสมของตัวแบบพยากรณ์ ด้วย
เทคนิคการถดถอยลอจิสติกทวิภาค ของชุดข้อมูลแบบที่ 2

เทคนิคการถดถอยลอจิสติกทวิภาค	-2LL	AIC
วิธีการเลือกแบบพื้นฐาน	3435.551	3449.551
วิธีการเลือกแบบไปข้างหน้า	3416.357	3430.357
วิธีการกำจัดแบบถอยหลัง	3400.367	3418.367
วิธีการถดถอยทีละชั้น	3416.357	3430.357

จากตาราง 85 เทคนิคการถดถอยลอจิสติกทวิภาควิธีการเลือกแบบไปข้างหน้า และวิธีการถดถอยทีละชั้น พบว่ามีค่าลอการิทึมของภาวะน่าจะเป็น ($-2 \text{ Log likelihood}$) เหมาะสม และค่าเกณฑ์สารสนเทศอะกะอิเกะ (AIC) เหมาะสมในวิธีการกำจัดแบบถอยหลัง ซึ่งเราจะพิจารณาระดับความน่าเชื่อถือของตัวแบบพยากรณ์ร่วมด้วยเพื่อให้ตัวแบบที่ได้มีความเหมาะสมมากที่สุด แสดงดังตาราง 86

ตาราง 86 แสดงผลระดับความน่าเชื่อถือของตัวแบบพยากรณ์จากเทคนิคการถดถอยลอจิสติก
ทวิภาคของชุดข้อมูลแบบที่ 2

เทคนิคการถดถอยลอจิสติกทวิภาค	Cox & Snell R^2	Nagelkerke R^2
วิธีการเลือกแบบพื้นฐาน	19.9522	30.4022
วิธีการเลือกแบบไปข้างหน้า	20.3293	30.9769
วิธีการกำจัดแบบถอยหลัง	20.6422	31.4537
วิธีการถดถอยทีละขั้น	20.3293	30.9769

จากตาราง 86 เทคนิคการถดถอยลอจิสติกทวิภาควิธีการกำจัดแบบถอยหลัง พบว่าสถิติทดสอบระดับความน่าเชื่อถือของตัวแบบพยากรณ์ Cox & Snell R^2 ตัวแบบพยากรณ์สามารถอธิบายการเป็นโรคเบาหวานได้ร้อยละ 20.6422 และ Nagelkerke R^2 พบว่าตัวแบบพยากรณ์สามารถอธิบายการเป็นโรคเบาหวานได้ร้อยละ 31.4537 สามารถอธิบายการเป็นโรคเบาหวานได้ดีที่สุด

ผลการเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์จากเทคนิคการถดถอยลอจิสติกทวิภาคของชุดข้อมูลแบบที่ 2

ตาราง 87 แสดงผลการเปรียบเทียบประสิทธิภาพจากเทคนิคการถดถอยลอจิสติกทวิภาค ของชุดข้อมูลแบบที่ 2

เทคนิคการถดถอยลอจิสติกทวิภาค	ชุดเรียนรู้		ชุดทดสอบ	
	ACC	BCA	ACC	BCA
วิธีการเลือกแบบพื้นฐาน	80.9547	63.6314	80.4326	61.6626
วิธีการเลือกแบบไปข้างหน้า	80.7333	63.3725	80.3343	61.4452
วิธีการกำจัดแบบถอยหลัง	80.8317	63.6293	79.9410	61.1911
วิธีการถดถอยทีละขั้น	80.7333	63.3725	80.3343	61.4452

จากตาราง 87 พบว่าเทคนิคการถดถอยลอจิสติกทวิภาควิธีการเลือกแบบพื้นฐาน มีความถูกต้องในชุดข้อมูลทดสอบ 80.4326 เปอร์เซ็นต์ ค่าความถูกต้องในการจำแนกแบบสมดุล 61.6626 เปอร์เซ็นต์

หลังจากผู้วิจัยเริ่มสร้างตัวแบบพยากรณ์ในชุดข้อมูลแบบที่ 2 ซึ่งมีข้อมูลเรียนรู้ 80 เปอร์เซ็นต์ ชุดข้อมูลทดสอบ 20 เปอร์เซ็นต์ ตัวด้วยเทคนิคการถดถอยลอจิสติกทวิภาคแล้ว ต่อมาผู้วิจัยพิจารณาสร้างตัวแบบพยากรณ์ด้วยเทคนิคต้นไม้ตัดสินใจอัลกอริทึม J48 และอัลกอริทึม LMT ตามลำดับ

4.4.2 เทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม J48

ผู้วิจัยเลือกใช้ค่าพารามิเตอร์ดังต่อไปนี้ กำหนดช่วงความเชื่อมั่นที่ใช้ในการแตกกิ่งต้นไม้เท่ากับ 95% และจำนวนการแตกกิ่งเท่ากับ 2

สำหรับผลลัพธ์ของตัวแบบพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม J48 โดยมีวิธีการคัดเลือกตัวแปรอิสระดังนี้

- 1) กรณีตัวแปรอิสระทุกตัว ได้แก่ ตัวแปรอิสระ $X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9, X_{10}, X_{11}, X_{12}$ และ X_{13}
- 2) วิธีการเลือกแบบพื้นฐาน ได้แก่ ตัวแปรอิสระ X_3, X_4, X_5, X_6, X_7 และ X_9
- 3) วิธีการเลือกแบบไปข้างหน้า X_3, X_4, X_5, X_8, X_9 และ X_{10}
- 4) วิธีการกำจัดแบบถอยหลัง $X_3, X_4, X_5, X_6, X_7, X_9, X_{12}$ และ X_{13}
- 5) วิธีการถดถอยทีละขั้น X_3, X_4, X_5, X_8, X_9 และ X_{10}

ผลการสร้างตัวแบบพยากรณ์จากเทคนิคต้นไม้ตัดสินใจ J48 กรณีตัวแปรอิสระทุกตัว

สำหรับผลลัพธ์ของตัวแบบพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม J48 กรณีตัวแปรอิสระทุกตัวแสดงดังตาราง 88 และ 89 ตามลำดับ

ตาราง 88 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม J48 กรณีตัวแปรอิสระทุกตัว ของชุดข้อมูลเรียนรู้ 80 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	608	272	69.0909
ไม่เป็นโรคเบาหวาน	309	2875	90.2952
	ค่าความถูกต้องของตัวแบบพยากรณ์		85.7037

จากตาราง 88 มีพบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 85.7037 เปอร์เซ็นต์ และ 79.6931 เปอร์เซ็นต์

ตาราง 89 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วย
อัลกอริทึม J48 กรณีตัวแปรอิสระทุกตัว ของชุดข้อมูลทดสอบ 20 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	123	92	57.2093
ไม่เป็นโรคเบาหวาน	107	695	86.6584
ค่าความถูกต้องของตัวแบบพยากรณ์			80.4327

จากตาราง 89 พบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 80.4327 เปอร์เซ็นต์ และ 71.9338 เปอร์เซ็นต์

ผลการสร้างตัวแบบพยากรณ์จากเทคนิคต้นไม้ตัดสินใจ J48 วิธีการเลือกแบบพื้นฐาน สำหรับผลลัพธ์ของตัวแบบพยากรณ์การเป็นโรคเบาหวาน จากเทคนิคต้นไม้ตัดสินใจด้วย อัลกอริทึม J48 วิธีการเลือกแบบพื้นฐาน แสดงดังตาราง 90 และ 91 ตามลำดับ

ตาราง 90 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วย
อัลกอริทึม J48 วิธีการเลือกแบบพื้นฐาน ของชุดข้อมูลเรียนรู้ 80 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	497	262	65.4809
ไม่เป็นโรคเบาหวาน	420	2885	87.2920
ค่าความถูกต้องของตัวแบบพยากรณ์			83.2185

จากตาราง 90 มีพบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 83.2185 เปอร์เซ็นต์ และ 76.3864 เปอร์เซ็นต์

ตาราง 91 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วย
อัลกอริทึม J48 วิธีการเลือกแบบพื้นฐาน ของชุดข้อมูลทดสอบ 20 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	108	72	60.0000
ไม่เป็นโรคเบาหวาน	122	715	85.4241
ค่าความถูกต้องของตัวแบบพยากรณ์			80.9243

จากตาราง 91 พบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 80.9243 เปอร์เซ็นต์ และ 72.7121 เปอร์เซ็นต์

ผลการสร้างตัวแบบพยากรณ์จากเทคนิคต้นไม้ตัดสินใจ J48 วิธีการเลือกแบบไปข้างหน้า

สำหรับผลลัพธ์ของตัวแบบพยากรณ์การเป็นโรคเบาหวาน จากเทคนิคต้นไม้ตัดสินใจด้วย อัลกอริทึม J48 วิธีการเลือกแบบไปข้างหน้า แสดงดังตาราง 92 และ 93 ตามลำดับ

ตาราง 92 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วย อัลกอริทึม J48 วิธีการเลือกแบบไปข้างหน้า ของชุดข้อมูลเรียนรู้ 80 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	574	317	64.4220
ไม่เป็นโรคเบาหวาน	343	2830	89.1900
ค่าความถูกต้องของตัวแบบพยากรณ์			83.7598

จากตาราง 92 มีพบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 83.7598 เปอร์เซ็นต์ และ 76.8060 เปอร์เซ็นต์

ตาราง 93 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วย อัลกอริทึม J48 วิธีการเลือกแบบไปข้างหน้า ของชุดข้อมูลทดสอบ 20 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	124	92	57.4074
ไม่เป็นโรคเบาหวาน	106	695	86.7665
ค่าความถูกต้องของตัวแบบพยากรณ์			80.5310

จากตาราง 93 พบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 80.5310 เปอร์เซ็นต์ และ 72.0870 เปอร์เซ็นต์

ผลการสร้างตัวแบบพยากรณ์จากเทคนิคต้นไม้ตัดสินใจ J48 วิธีการกำจัดแบบถอยหลัง
 สำหรับผลลัพธ์ของตัวแบบพยากรณ์การเป็นโรคเบาหวาน จากเทคนิคต้นไม้ตัดสินใจด้วย
 อัลกอริทึม J48 วิธีการกำจัดแบบถอยหลัง แสดงดังตาราง 94 และ 95 ตามลำดับ

**ตาราง 94 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วย
 อัลกอริทึม J48 วิธีการกำจัดแบบถอยหลัง ของชุดข้อมูลทดสอบ 80 เปอร์เซ็นต์**

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	602	265	69.4348
ไม่เป็นโรคเบาหวาน	315	2882	90.1470
ค่าความถูกต้องของตัวแบบพยากรณ์			85.7283

จากตาราง 94 มีพบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ
 85.7283 เปอร์เซ็นต์ และ 79.7909 เปอร์เซ็นต์

**ตาราง 95 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วย
 อัลกอริทึม J48 วิธีการกำจัดแบบถอยหลัง ของชุดข้อมูลทดสอบ 20 เปอร์เซ็นต์**

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	117	96	54.9296
ไม่เป็นโรคเบาหวาน	113	691	85.9453
ค่าความถูกต้องของตัวแบบพยากรณ์			79.4494

จากตาราง 95 พบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ
 79.4494 เปอร์เซ็นต์ และ 70.4374 เปอร์เซ็นต์

ผลการสร้างตัวแบบพยากรณ์จากเทคนิคต้นไม้ตัดสินใจ J48 วิธีการถดถอยทีละชั้น

สำหรับผลลัพธ์ของตัวแบบพยากรณ์การเป็นโรคเบาหวาน จากเทคนิคต้นไม้ตัดสินใจด้วย อัลกอริทึม J48 วิธีการถดถอยทีละชั้น แสดงดังตาราง 96 และ 97 ตามลำดับ

ตาราง 96 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วย อัลกอริทึม J48 วิธีการถดถอยทีละชั้น ของชุดข้อมูลทดสอบ 80 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	574	317	64.4220
ไม่เป็นโรคเบาหวาน	343	2830	89.1900
ค่าความถูกต้องของตัวแบบพยากรณ์			83.7598

จากตาราง 96 พบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 83.7598 เปอร์เซ็นต์ และ 76.8060 เปอร์เซ็นต์

ตาราง 97 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วย อัลกอริทึม J48 วิธีการถดถอยทีละชั้น ของชุดข้อมูลทดสอบ 20 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	124	92	57.4074
ไม่เป็นโรคเบาหวาน	106	695	86.7665
ค่าความถูกต้องของตัวแบบพยากรณ์			80.5310

จากตาราง 97 พบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 80.5310 เปอร์เซ็นต์ และ 72.0870 เปอร์เซ็นต์

ผลการเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์จากเทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม J48 ของชุดข้อมูลแบบที่ 2

ตาราง 98 แสดงผลการเปรียบเทียบประสิทธิภาพด้วยเทคนิคต้นไม้ตัดสินใจอัลกอริทึม J48 ของชุดข้อมูลแบบที่ 2

เทคนิคต้นไม้ตัดสินใจ อัลกอริทึม J48	ชุดเรียนรู้		ชุดทดสอบ	
	ACC	BCA	ACC	BCA
กรณีตัวแปรอิสระทุกตัว	85.7037	79.6931	80.4327	71.9338
วิธีการเลือกแบบพื้นฐาน	83.2185	76.3864	80.9243	72.7121
วิธีการเลือกแบบไปข้างหน้า	83.7598	76.8060	80.5310	72.0870
วิธีการกำจัดแบบถอยหลัง	85.7283	79.7909	79.4494	70.4374
วิธีการถดถอยทีละชั้น	83.7598	76.8060	80.5310	72.0870

จากตาราง 98 พบว่าเทคนิคต้นไม้ตัดสินใจอัลกอริทึม J48 กรณีการคัดเลือกตัวแปรอิสระด้วยวิธีการเลือกแบบพื้นฐานในเทคนิคการถดถอยลอจิสติกทวิภาค ประกอบไปด้วยตัวแปรอิสระ ค่าความดันขณะหัวใจบีบตัว ค่าความดันขณะหัวใจคลายตัว ค่าอัตราการเต้นของหัวใจ น้ำหนัก ความสูง และระดับน้ำตาลในเลือด มีค่าความถูกต้องในชุดข้อมูลทดสอบดีที่สุดที่สุด 80.9243 เปอร์เซ็นต์ ค่าความถูกต้องในการจำแนกแบบสมดุล 72.7121 เปอร์เซ็นต์

4.4.3 เทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม LMT

ผู้วิจัยกำหนดจำนวนการทำซ้ำสำหรับอัลกอริทึมลอจิสติกบูส (LogitBoost) เท่ากับ -1 ซึ่งหมายถึง การทำซ้ำแบบการตรวจสอบข้าม (Cross-Validation) จำนวนตัวแปรอิสระขั้นต่ำในโหนดที่ถูกพิจารณา กำหนดค่าเท่ากับ 15

สำหรับผลลัพธ์ของตัวแบบพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม LMT โดยมีวิธีการคัดเลือกตัวแปรอิสระทั้ง 5 รูปแบบ กรณีตัวแปรอิสระทุกตัว วิธีการเลือกแบบพื้นฐาน วิธีการเลือกแบบไปข้างหน้า วิธีการกำจัดแบบถอยหลัง และวิธีการถดถอยทีละชั้น

ผลการสร้างตัวแบบพยากรณ์จากเทคนิคต้นไม้ตัดสินใจ LMT กรณีตัวแปรอิสระทุกตัว สำหรับผลลัพธ์ของตัวแบบพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วย อัลกอริทึม LMT กรณีตัวแปรอิสระทุกตัวแสดงดังตาราง 99 แล 100 ตามลำดับ

ตาราง 99 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วย อัลกอริทึม LMT กรณีตัวแปรอิสระทุกตัว ของชุดข้อมูลเรียนรู้ 80 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	538	278	65.9314
ไม่เป็นโรคเบาหวาน	379	2869	88.3312
ค่าความถูกต้องของตัวแบบพยากรณ์			83.8337

จากตาราง 99 พบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 83.8337 เปอร์เซ็นต์ และ 77.1313 เปอร์เซ็นต์

ตาราง 100 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วย อัลกอริทึม LMT กรณีตัวแปรอิสระทุกตัว ของชุดข้อมูลทดสอบ 20 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	119	78	60.4061
ไม่เป็นโรคเบาหวาน	111	709	86.4634
ค่าความถูกต้องของตัวแบบพยากรณ์			81.4159

จากตาราง 100 พบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 81.4159 เปอร์เซ็นต์ และ 73.4348 เปอร์เซ็นต์

ผลการสร้างตัวแบบพยากรณ์จากเทคนิคต้นไม้ตัดสินใจ LMT วิธีการเลือกแบบพื้นฐาน สำหรับผลลัพธ์ของตัวแบบพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วย อัลกอริทึม LMT วิธีการเลือกแบบพื้นฐานแสดงดังตาราง 101 และ 102 ตามลำดับ

ตาราง 101 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วย อัลกอริทึม LMT วิธีการเลือกแบบพื้นฐาน ของชุดข้อมูลเรียนรู้ 80 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	547	333	62.1591
ไม่เป็นโรคเบาหวาน	370	2814	88.3794
ค่าความถูกต้องของตัวแบบพยากรณ์			82.7018

จากตาราง 101 พบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 82.7018 เปอร์เซ็นต์ และ 75.2692 เปอร์เซ็นต์

ตาราง 102 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วย อัลกอริทึม LMT วิธีการเลือกแบบพื้นฐาน ของชุดข้อมูลทดสอบ 20 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	123	83	59.7087
ไม่เป็นโรคเบาหวาน	107	704	86.8064
ค่าความถูกต้องของตัวแบบพยากรณ์			81.3176

จากตาราง 102 มีพบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 81.3176 เปอร์เซ็นต์ และ 73.2576 เปอร์เซ็นต์

ผลการสร้างตัวแบบพยากรณ์จากเทคนิคต้นไม้ตัดสินใจ LMT วิธีการเลือกแบบไปข้างหน้า

สำหรับผลลัพธ์ของตัวแบบพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม LMT วิธีการเลือกแบบไปข้างหน้า แสดงดังตาราง 103 แล 104 ตามลำดับ

ตาราง 103 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม LMT วิธีการเลือกแบบไปข้างหน้า ของชุดข้อมูลเรียนรู้ 80 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	563	336	62.6251
ไม่เป็นโรคเบาหวาน	354	2811	88.8152
ค่าความถูกต้องของตัวแบบพยากรณ์			83.0217

จากตาราง 103 พบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 83.0217 เปอร์เซ็นต์ และ 75.7202 เปอร์เซ็นต์

ตาราง 104 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม LMT วิธีการเลือกแบบไปข้างหน้า ของชุดข้อมูลทดสอบ 20 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	126	90	58.3333
ไม่เป็นโรคเบาหวาน	104	697	87.0162
ค่าความถูกต้องของตัวแบบพยากรณ์			80.9243

จากตาราง 104 พบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 80.9243 เปอร์เซ็นต์ และ 72.6748 เปอร์เซ็นต์

ผลการสร้างตัวแบบพยากรณ์จากเทคนิคต้นไม้ตัดสินใจ LMT วิธีการกำจัดแบบถอยหลัง

สำหรับผลลัพธ์ของตัวแบบพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม LMT วิธีการกำจัดแบบถอยหลัง แสดงดังตาราง 105 และ 106 ตามลำดับ

ตาราง 105 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม LMT วิธีการกำจัดแบบถอยหลัง ของชุดข้อมูลเรียนรู้ 80 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	542	289	65.2226
ไม่เป็นโรคเบาหวาน	375	2858	88.4009
ค่าความถูกต้องของตัวแบบพยากรณ์			83.6614

จากตาราง 105 พบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 83.6614 เปอร์เซ็นต์ และ 76.8117 เปอร์เซ็นต์

ตาราง 106 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม LMT วิธีการกำจัดแบบถอยหลัง ของชุดข้อมูลทดสอบ 20 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	113	79	58.8542
ไม่เป็นโรคเบาหวาน	117	708	85.8182
ค่าความถูกต้องของตัวแบบพยากรณ์			80.7276

จากตาราง 106 พบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 80.7276 เปอร์เซ็นต์ และ 72.3362 เปอร์เซ็นต์

ผลการสร้างตัวแบบพยากรณ์จากเทคนิคต้นไม้ตัดสินใจ LMT วิธีการถดถอยทีละชั้น

สำหรับผลลัพธ์ของตัวแบบพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วย อัลกอริทึม LMT วิธีการถดถอยทีละชั้น แสดงดังตาราง 107 แล 108 ตามลำดับ

ตาราง 107 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วย อัลกอริทึม LMT วิธีการถดถอยทีละชั้น ของชุดข้อมูลเรียนรู้ 80 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	563	336	62.6251
ไม่เป็นโรคเบาหวาน	354	2811	88.8152
ค่าความถูกต้องของตัวแบบพยากรณ์			83.0217

จากตาราง 107 พบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 83.0217 เปอร์เซ็นต์ และ 75.7202 เปอร์เซ็นต์

ตาราง 108 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคต้นไม้ตัดสินใจด้วย อัลกอริทึม LMT วิธีการถดถอยทีละชั้น ของชุดข้อมูลทดสอบ 20 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	126	90	58.3333
ไม่เป็นโรคเบาหวาน	104	697	87.0162
ค่าความถูกต้องของตัวแบบพยากรณ์			80.9243

จากตาราง 108 พบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 80.9243 เปอร์เซ็นต์ และ 72.6748 เปอร์เซ็นต์

ผลการเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์จากเทคนิคต้นไม้ตัดสินใจด้วย
อัลกอริทึม LMT ของชุดข้อมูลแบบที่ 2

ตาราง 109 แสดงผลการเปรียบเทียบประสิทธิภาพจากเทคนิคต้นไม้ตัดสินใจอัลกอริทึม LMT
ของชุดข้อมูลแบบที่ 2

เทคนิคต้นไม้ตัดสินใจ อัลกอริทึม LMT	ชุดเรียนรู้		ชุดทดสอบ	
	ACC	BCA	ACC	BCA
กรณีตัวแปรอิสระทุกตัว	83.8337	77.1313	81.4159	73.4348
วิธีการเลือกแบบพื้นฐาน	82.7018	75.2692	81.3176	73.2576
วิธีการเลือกแบบไปข้างหน้า	83.0217	75.7202	80.9243	72.6748
วิธีการกำจัดแบบถอยหลัง	83.6614	76.8117	80.7276	72.3362
วิธีการถดถอยทีละขั้น	83.0217	75.7202	80.9243	72.6748

จากตาราง 109 พบว่าเทคนิคต้นไม้ตัดสินใจอัลกอริทึม LMT กรณีตัวแปรอิสระทุกตัว มีความถูกต้องในชุดข้อมูลทดสอบดีที่สุด 81.4159 เปอร์เซ็นต์ ค่าความถูกต้องในการจำแนกแบบสมดุล 73.4348 เปอร์เซ็นต์

4.4.4 เทคนิคโครงข่ายประสาทเทียม

กำหนดค่าพารามิเตอร์ของเทคนิคโครงข่ายประสาทเทียมดังนี้

- 1) ฟังก์ชันกระตุ้น คือฟังก์ชันซิกมอยด์
- 2) จำนวนชั้นซ่อน 1 ชั้น
- 3) จำนวนโหนดในชั้นซ่อน กำหนดเป็น 5, 6, 8, 9 และ 10
- 4) อัตราการเรียนรู้ คือ 0.1, 0.2 และ 0.3
- 5) ค่าโมเมนตัม คือ 0.1, 0.2 และ 0.3
- 6) จำนวนรอบ 100 รอบ

จากการกำหนดพารามิเตอร์ดังกล่าวข้างต้น ทำให้ต้องทำซ้ำ 45 ครั้ง ในการสร้างตัวแบบพยากรณ์ของแต่ละกรณีการเลือกตัวแปรอิสระทั้ง 5 รูปแบบ กรณีตัวแปรอิสระทุกตัว วิธีการเลือกแบบพื้นฐาน วิธีการเลือกแบบไปข้างหน้า วิธีการกำจัดแบบถอยหลัง และวิธีการถดถอยทีละขั้น

ผลการสร้างตัวแบบพยากรณ์จากเทคนิคโครงข่ายประสาทเทียม กรณีตัวแปรอิสระทุกตัว

สำหรับผลลัพธ์ของตัวแบบพยากรณ์การเป็นโรคเบาหวานจากเทคนิคโครงข่ายประสาทเทียมกรณีตัวแปรอิสระทุกตัว โดยโครงข่ายประสาทเทียมที่มีประสิทธิภาพดีที่สุด คือตัวแบบพยากรณ์ที่มีพารามิเตอร์ดังนี้ จำนวนโหนดในชั้นซ่อน คือ 5 โหนด อัตราการเรียนรู้ คือ 0.2 โมเมนตัม คือ 0.2 แสดงดังตาราง 110 และ 111 ตามลำดับ

ตาราง 110 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคโครงข่ายประสาทเทียม กรณีตัวแปรอิสระทุกตัว ของชุดข้อมูลเรียนรู้ 80 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	261	133	66.2437
ไม่เป็นโรคเบาหวาน	656	3014	82.1253
ค่าความถูกต้องของตัวแบบพยากรณ์			80.5856

จากตาราง 110 พบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 80.5856 เปอร์เซ็นต์ และ 74.1845 เปอร์เซ็นต์

ตาราง 111 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคโครงข่ายประสาทเทียม กรณีตัวแปรอิสระทุกตัว ของชุดข้อมูลทดสอบ 20 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	167	133	55.6667
ไม่เป็นโรคเบาหวาน	63	654	91.2134
ค่าความถูกต้องของตัวแบบพยากรณ์			80.7276

จากตาราง 111 พบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 80.7276 เปอร์เซ็นต์ และ 73.4400 เปอร์เซ็นต์

ผลการสร้างตัวแบบพยากรณ์จากเทคนิคโครงข่ายประสาทเทียมวิธีการเลือกแบบพื้นฐาน

สำหรับผลลัพธ์ของตัวแบบพยากรณ์การเป็นโรคเบาหวานจากเทคนิคโครงข่ายประสาทเทียมวิธีการเลือกแบบพื้นฐาน โดยโครงข่ายประสาทเทียมที่มีประสิทธิภาพดีที่สุด คือตัวแบบพยากรณ์ที่มีพารามิเตอร์ดังนี้ จำนวนโหนดในชั้นซ่อน คือ 9 โหนด อัตราการเรียนรู้ คือ 0.3 โมเมนตัม คือ 0.1 แสดงดังตาราง 112 และ 113 ตามลำดับ

ตาราง 112 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคโครงข่ายประสาทเทียมวิธีการเลือกแบบพื้นฐาน ของชุดข้อมูลเรียนรู้ 80 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	273	143	65.6250
ไม่เป็นโรคเบาหวาน	644	3004	82.3465
ค่าความถูกต้องของตัวแบบพยากรณ์			80.6348

จากตาราง 112 พบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 80.6348 เปอร์เซ็นต์ และ 73.9857 เปอร์เซ็นต์

ตาราง 113 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคโครงข่ายประสาทเทียมวิธีการเลือกแบบพื้นฐาน ของชุดข้อมูลทดสอบ 20 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	153	112	57.7358
ไม่เป็นโรคเบาหวาน	77	675	89.7606
ค่าความถูกต้องของตัวแบบพยากรณ์			81.4159

จากตาราง 113 พบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 81.4159 เปอร์เซ็นต์ และ 73.7482 เปอร์เซ็นต์

ผลการสร้างตัวแบบพยากรณ์จากเทคนิคโครงข่ายประสาทเทียมวิธีการเลือกแบบไปข้างหน้า

สำหรับผลลัพธ์ของตัวแบบพยากรณ์การเป็นโรคเบาหวานจากเทคนิคโครงข่ายประสาทเทียมวิธีการเลือกไปข้างหน้า โดยโครงข่ายประสาทเทียมที่มีประสิทธิภาพดีที่สุดในที่นี้คือตัวแบบพยากรณ์ที่มีพารามิเตอร์ดังนี้ จำนวนโหนดในชั้นซ่อน คือ 5 โหนด อัตราการเรียนรู้ คือ 0.3 โมเมนตัม คือ 0.3 แสดงดังตาราง 114 และ 115 ตามลำดับ

ตาราง 114 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคโครงข่ายประสาทเทียมวิธีการเลือกแบบไปข้างหน้า ของชุดข้อมูลเรียนรู้ 80 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	264	142	65.0246
ไม่เป็นโรคเบาหวาน	653	3005	82.1487
ค่าความถูกต้องของตัวแบบพยากรณ์			80.4380

จากตาราง 114 พบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 80.4380 เปอร์เซ็นต์ และ 73.5867 เปอร์เซ็นต์

ตาราง 115 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคโครงข่ายประสาทเทียมวิธีการเลือกแบบไปข้างหน้า ของชุดข้อมูลทดสอบ 20 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	166	140	54.2484
ไม่เป็นโรคเบาหวาน	64	647	90.9986
ค่าความถูกต้องของตัวแบบพยากรณ์			79.9410

จากตาราง 115 พบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 79.9410 เปอร์เซ็นต์ และ 72.6235 เปอร์เซ็นต์

ผลการสร้างตัวแบบพยากรณ์จากเทคนิคโครงข่ายประสาทเทียมวิธีการกำจัดแบบถอยหลัง

สำหรับผลลัพธ์ของตัวแบบพยากรณ์การเป็นโรคเบาหวานจากเทคนิคโครงข่ายประสาทเทียมวิธีการกำจัดแบบถอยหลัง โดยโครงข่ายประสาทเทียมที่มีประสิทธิภาพดีที่สุด คือตัวแบบพยากรณ์ที่มีพารามิเตอร์ดังนี้ จำนวนโหนดในชั้นซ่อน คือ 10 โหนด อัตราการเรียนรู้ คือ 0.2 โมเมนตัม คือ 0.3 แสดงดังตาราง 116 และ 117 ตามลำดับ

ตาราง 116 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคโครงข่ายประสาทเทียมวิธีการกำจัดแบบถอยหลัง ของชุดข้อมูลทดสอบ 20 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	286	145	66.3573
ไม่เป็นโรคเบาหวาน	631	3002	82.6314
ค่าความถูกต้องของตัวแบบพยากรณ์			66

จากตาราง 116 พบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 80.9055 เปอร์เซ็นต์ และ 74.4944 เปอร์เซ็นต์

ตาราง 117 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคโครงข่ายประสาทเทียมวิธีการกำจัดแบบถอยหลัง ของชุดข้อมูลทดสอบ 20 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	162	128	55.8621
ไม่เป็นโรคเบาหวาน	68	659	90.6465
ค่าความถูกต้องของตัวแบบพยากรณ์			80.7276

จากตาราง 117 พบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 80.7276 เปอร์เซ็นต์ และ 73.2543 เปอร์เซ็นต์

ผลการสร้างตัวแบบพยากรณ์จากเทคนิคโครงข่ายประสาทเทียมวิธีการถดถอยทีละชั้น

สำหรับผลลัพธ์ของตัวแบบพยากรณ์การเป็นโรคเบาหวานจากเทคนิคโครงข่ายประสาทเทียมวิธีการถดถอยทีละชั้น โดยโครงข่ายประสาทเทียมที่มีประสิทธิภาพดีที่สุด คือตัวแบบพยากรณ์ที่มีพารามิเตอร์ดังนี้ จำนวนโหนดในชั้นซ่อน คือ 5 โหนด อัตราการเรียนรู้ คือ 0.3 โมเมนตัม คือ 0.3 แสดงดังตาราง 118 และ 119 ตามลำดับ

ตาราง 118 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคโครงข่ายประสาทเทียมวิธีการถดถอยทีละชั้น ของชุดข้อมูลเรียนรู้ 80 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	264	142	65.0246
ไม่เป็นโรคเบาหวาน	653	3005	82.1487
ค่าความถูกต้องของตัวแบบพยากรณ์			80.4380

จากตาราง 118 พบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 80.4380 เปอร์เซ็นต์ และ 73.5867 เปอร์เซ็นต์

ตาราง 119 แสดงผลลัพธ์การพยากรณ์การเป็นโรคเบาหวานจากเทคนิคโครงข่ายประสาทเทียมวิธีการถดถอยทีละชั้น ของชุดข้อมูลทดสอบ 20 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การพยากรณ์ (Predicted Class)		ค่าความถูกต้อง (%)
	เป็นโรคเบาหวาน	ไม่เป็นโรคเบาหวาน	
เป็นโรคเบาหวาน	166	140	54.2484
ไม่เป็นโรคเบาหวาน	64	647	90.9986
ค่าความถูกต้องของตัวแบบพยากรณ์			79.9410

จากตาราง 119 พบว่าค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล คือ 79.9410 เปอร์เซ็นต์ และ 72.6235 เปอร์เซ็นต์

ผลการเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์จากเทคนิคโครงข่ายประสาทเทียม ของชุดข้อมูลแบบที่ 2

ตาราง 120 แสดงผลการเปรียบเทียบประสิทธิภาพการพยากรณ์ด้วยเทคนิคโครงข่ายประสาทเทียม ของชุดข้อมูลแบบที่ 2

เทคนิคโครงข่ายประสาทเทียม (จำนวนชั้นซ่อน, อัตราการเรียนรู้, โมเมนตัม)	ชุดเรียนรู้		ชุดทดสอบ	
	ACC	BCA	ACC	BCA
กรณีตัวแปรอิสระทุกตัว (5, 0.2, 0.2)	80.5856	74.1845	80.7276	73.4400
วิธีการเลือกแบบพื้นฐาน (9, 0.3, 0.1)	80.6348	73.9857	81.4159	73.7482
วิธีการเลือกแบบไปข้างหน้า (5, 0.3, 0.3)	80.4380	73.5867	79.9410	72.6235
วิธีการกำจัดแบบถอยหลัง (10, 0.2, 0.3)	80.9055	74.4944	80.7276	73.2543
วิธีการถดถอยทีละชั้น (5, 0.3, 0.3)	80.4380	73.5867	79.9410	72.6235

จากตาราง 120 พบว่าเทคนิคโครงข่ายประสาทเทียม กรณีตัวแปรอิสระทุกตัว โดยมีจำนวนชั้นซ่อน อัตราการเรียนรู้ และโมเมนตัม คือ 9, 0.3 และ 0.1 ตามลำดับ มีค่าความถูกต้องในชุดข้อมูลทดสอบดีที่สุด 81.4159 เปอร์เซ็นต์ ค่าความถูกต้องในการจำแนกแบบสมดุล 73.7482 เปอร์เซ็นต์

จากที่กล่าวมาข้างต้นผลลัพธ์ของประสิทธิภาพของเทคนิคทั้ง 3 เทคนิค ในชุดข้อมูลแบบที่ 1 และชุดข้อมูลแบบที่ 2 สามารถสรุปได้ดังตาราง 121 และ 122 ตามลำดับ

4.5 ผลการเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์ในชุดข้อมูลแบบที่ 2 (80:20)

ตาราง 121 แสดงผลการเปรียบเทียบประสิทธิภาพตัวแบบพยากรณ์ด้วยเทคนิคต่าง ๆ ของชุดข้อมูลแบบที่ 2

	ชุดข้อมูลทดสอบ	
	ACC	BCA
การถดถอยลอจิสติกทวิภาค (วิธีการเลือกแบบพื้นฐาน)	80.4326	61.6626
ต้นไม้ตัดสินใจด้วยอัลกอริทึม J48 (วิธีการเลือกแบบพื้นฐาน)	80.9243	72.7121
ต้นไม้ตัดสินใจด้วยอัลกอริทึม LMT (กรณีตัวแปรอิสระทุกตัว)	81.4159	73.4348
โครงข่ายประสาทเทียม (วิธีการเลือกแบบพื้นฐาน)	81.4159	73.7482
จำนวนโหนดในชั้นซ่อน, อัตราการเรียนรู้, โมเมนตัม (9, 0.3, 0.1)		

จากตาราง 121 พบว่าเทคนิคโครงข่ายประสาทเทียม มีประสิทธิภาพดีที่สุดในชุดข้อมูลแบบที่ 2 ชุดข้อมูลเรียนรู้ 80 เปอร์เซ็นต์ และชุดข้อมูลทดสอบ 20 เปอร์เซ็นต์

4.6 ผลการเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์ทั้งสองชุดข้อมูล

ตาราง 122 แสดงผลการเปรียบเทียบประสิทธิภาพตัวแบบพยากรณ์ที่ดีที่สุด

ชุดข้อมูล		ชุดข้อมูลทดสอบ	
		ACC	BCA
1	โครงข่ายประสาทเทียม (วิธีการเลือกแบบไปข้างหน้า, วิธีการถดถอยทีละชั้น)	81.7824	73.9704
2	โครงข่ายประสาทเทียม (วิธีการเลือกแบบพื้นฐาน)	81.4159	73.7482

จากตาราง 122 พบว่าตัวแบบพยากรณ์ที่มีประสิทธิภาพดีที่สุดในชุดข้อมูลแบบที่ 1 คือ ตัวแบบพยากรณ์ด้วยเทคนิคโครงข่ายประสาทเทียม มีค่าความถูกต้องในชุดข้อมูลทดสอบดีที่สุด 81.7824 เปอร์เซ็นต์ ค่าความถูกต้องในการจำแนกแบบสมดุล 73.9704 เปอร์เซ็นต์

บทที่ 5

สรุปผลการวิจัย

การวิจัยเรื่องการเปรียบเทียบตัวแบบการถดถอยลอจิสติกและเทคนิคเหมืองข้อมูลสำหรับพยากรณ์การเป็นโรคเบาหวาน มีวัตถุประสงค์เพื่อวิเคราะห์ปัจจัยที่ส่งผลต่อการเป็นโรคเบาหวานโดยใช้การถดถอยลอจิสติกทวิภาค ที่ระดับนัยสำคัญ 0.05 และเปรียบเทียบประสิทธิภาพของตัวแบบการถดถอยลอจิสติกทวิภาค ต้นไม้ตัดสินใจ และ โครงข่ายประสาทเทียมในการพยากรณ์การเป็นโรคเบาหวาน โดยใช้ข้อมูลผู้ป่วยทั่วไปที่เข้ามารับการรักษาในโรงพยาบาลมหาวิทยาลัยนครสวรรค์ ตั้งแต่วันที่ 1 มกราคม 2563 – 31 มิถุนายน 2563 โดยคัดเลือกตัวแปรอิสระที่เกี่ยวข้องจากการทบทวนวรรณกรรม ซึ่งได้ตัวแปรอิสระที่ใช้ในงานวิจัยนี้มีทั้งหมด 13 ตัวแปร ได้แก่ เพศ อายุ ค่าความดันขณะหัวใจบีบตัว ค่าความดันขณะหัวใจคลายตัว อัตราการเต้นของหัวใจ น้ำหนัก ความสูง ดัชนีมวลกาย ระดับน้ำตาลในเลือด ระดับคอเลสเตอรอลในเลือด ระดับไตรกลีเซอไรด์ในเลือด ระดับไขมันในเลือดที่มีความหนาแน่นสูง และระดับไขมันในเลือดที่มีความหนาแน่นต่ำ และตัวแปรตาม คือ การเป็นโรคเบาหวาน จากนั้นนำข้อมูลตามตัวแปรที่กล่าวมาข้างต้นมาใช้ในการสร้างตัวแบบพยากรณ์การเป็นโรคเบาหวานด้วยเทคนิค 4 เทคนิค ได้แก่ เทคนิคการถดถอยลอจิสติกทวิภาค เทคนิคต้นไม้ตัดสินใจ ด้วยอัลกอริทึม J48 และอัลกอริทึม LMT และเทคนิคโครงข่ายประสาทเทียม จากนั้นทำการเปรียบเทียบประสิทธิภาพด้วยค่าความถูกต้อง และค่าความถูกต้องในการจำแนกแบบสมดุล

โดยผลที่ได้จากการศึกษาสามารถสรุปได้ดังนี้

1. จากการเตรียมข้อมูลเพื่อสร้างตัวแบบพยากรณ์ ผู้วิจัยได้ทำการคัดกรองและคัดเลือกข้อมูล พบว่าข้อมูลที่น่ามาใช้มีจำนวน 5,081 ชุด ตัวแปรอิสระที่ใช้มีทั้งหมด 13 ตัวแปร และตัวแปรตาม คือการเป็นโรคเบาหวาน จากการวิเคราะห์พบว่าปัจจัยที่ส่งผลต่อการเป็นโรคเบาหวานด้วยเทคนิคการถดถอยลอจิสติกทวิภาคที่ระดับนัยสำคัญ 0.05 มีดังนี้

- 1) ค่าความดันขณะหัวใจบีบตัว
- 2) ค่าความดันขณะหัวใจคลายตัว
- 3) อัตราการเต้นของหัวใจ
- 4) น้ำหนัก
- 5) ความสูง
- 6) ระดับน้ำตาลในเลือด

2. ตัวแปรอิสระที่ได้จากวิธีการเลือกแบบต่าง ๆ ตามแนวทางของเทคนิคการถดถอยลอจิสติกทวิภาค ในชุดข้อมูลแบบที่ 1 และแบบที่ 2 มีความเหมือนกันดังแสดงเป็นข้อสรุปได้ดังตาราง 123

ตาราง 123 แสดงผลการคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกทวิภาค

วิธี	ตัวแปรอิสระ
วิธีการเลือกแบบพื้นฐาน	ค่าความดันขณะหัวใจบีบตัว, ค่าความดันขณะหัวใจคลายตัว, อัตราการเต้นของหัวใจ, น้ำหนัก, ความสูง, ระดับน้ำตาลในเลือด
วิธีการเลือกแบบไปข้างหน้า	ค่าความดันขณะหัวใจบีบตัว, ค่าความดันขณะหัวใจคลายตัว, อัตราการเต้นของหัวใจ, ดัชนีมวลกาย, ระดับน้ำตาลในเลือด, ระดับคอเลสเตอรอลในเลือด
วิธีการกำจัดแบบถอยหลัง	ค่าความดันขณะหัวใจบีบตัว, ค่าความดันขณะหัวใจคลายตัว, อัตราการเต้นของหัวใจ, ดัชนีมวลกาย, ระดับน้ำตาลในเลือด, ระดับคอเลสเตอรอลในเลือด
วิธีการถดถอยทีละขั้น	ค่าความดันขณะหัวใจบีบตัว, ค่าความดันขณะหัวใจคลายตัว, อัตราการเต้นของหัวใจ, น้ำหนัก, ความสูง, ระดับน้ำตาลในเลือด, ระดับไขมันในเลือดที่มีความหนาแน่นสูง, ระดับไขมันในเลือดที่มีความหนาแน่นต่ำ

จากตาราง 123 พบว่าตัวแปรอิสระ 4 ตัวแปร ได้แก่ ค่าความดันขณะหัวใจบีบตัว ค่าความดันขณะหัวใจคลายตัว อัตราการเต้นของหัวใจ และระดับน้ำตาลในเลือด ได้รับการคัดเลือกทั้ง 4 วิธี

3. การเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์ที่ดีที่สุดของชุดข้อมูลแบบที่ 1 และชุดข้อมูลแบบที่ 2 สามารถสรุปได้ดังตาราง 124

ตาราง 124 แสดงผลการเปรียบเทียบประสิทธิภาพตัวแบบพยากรณ์ที่ดีที่สุด

ชุดข้อมูล		ชุดข้อมูลทดสอบ	
		ACC	BCA
1	โครงข่ายประสาทเทียม	81.7824	73.9704
2	โครงข่ายประสาทเทียม	81.4159	73.7482

จากตาราง 124 พบว่าตัวแบบพยากรณ์ที่มีประสิทธิภาพดีที่สุดในชุดข้อมูลแบบที่ 1 คือ ตัวแบบพยากรณ์ด้วยเทคนิคโครงข่ายประสาทเทียม มีค่าความถูกต้องในชุดข้อมูลทดสอบดีที่สุดที่ 81.7824 เปอร์เซ็นต์ ค่าความถูกต้องในการจำแนกแบบสมดุล 73.9704 เปอร์เซ็นต์

4. การคัดเลือกตัวแปรอิสระด้วยวิธีการเลือกตามเทคนิคการถดถอยลอจิสติกทวิภาค ด้วยวิธีต่าง ๆ เพื่อใช้สร้างตัวแบบพยากรณ์ที่มีประสิทธิภาพ สามารถสรุปได้ดังตาราง 125

ตาราง 125 แสดงเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์ที่ดีที่สุดในแต่ละชุดข้อมูล

ชุดข้อมูลแบบที่ 1 (70 : 30)	ชุดข้อมูลแบบที่ 2 (80 : 20)
เทคนิคการถดถอยลอจิสติกทวิภาค	เทคนิคการถดถอยลอจิสติกทวิภาค
<ul style="list-style-type: none"> • วิธีการเลือกแบบพื้นฐาน 	<ul style="list-style-type: none"> • วิธีการเลือกแบบพื้นฐาน
เทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม J48	เทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม J48
<ul style="list-style-type: none"> • วิธีการกำจัดแบบถอยหลัง 	<ul style="list-style-type: none"> • วิธีการเลือกแบบพื้นฐาน
เทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม LMT	เทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม LMT
<ul style="list-style-type: none"> • วิธีการกำจัดแบบถอยหลัง 	<ul style="list-style-type: none"> • กรณีสองตัวแปรอิสระทุกตัว
เทคนิคโครงข่ายประสาทเทียม	เทคนิคโครงข่ายประสาทเทียม
<ul style="list-style-type: none"> • วิธีการเลือกแบบไปข้างหน้า • วิธีการถดถอยทีละชั้น • (6, 0.2, 0.3) 	<ul style="list-style-type: none"> • วิธีการเลือกแบบพื้นฐาน • (9, 0.3, 0.1)

จากตาราง 125 พบว่า

- ตัวแบบพยากรณ์จากเทคนิคการถดถอยลอจิสติกทวิภาคที่มีประสิทธิภาพที่ดีที่สุด ได้จากวิธีการเลือกแบบพื้นฐาน
- ตัวแบบพยากรณ์จากเทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม J48 ที่มีประสิทธิภาพที่ดีที่สุด ได้จากวิธีการกำจัดแบบถอยหลัง และวิธีการเลือกแบบพื้นฐาน
- ตัวแบบพยากรณ์จากเทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม LMT ที่มีประสิทธิภาพที่ดีที่สุด ได้จากวิธีการกำจัดแบบถอยหลัง และกรณีสองตัวแปรอิสระทุกตัว
- ตัวแบบพยากรณ์จากเทคนิคโครงข่ายประสาทเทียม ที่มีประสิทธิภาพที่ดีที่สุด ได้จากวิธีการเลือกแบบไปข้างหน้า วิธีการถดถอยทีละชั้น และวิธีการเลือกแบบพื้นฐาน

ซึ่งในที่นี้พบว่า ตัวแบบพยากรณ์ที่มีประสิทธิภาพเป็นเทคนิคเหมือนข้อมูลที่ได้จากวิธีการคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกทวิภาคทั้งสิ้น กล่าวคือไม่ได้เกิดจากการใช้ตัวแปรอิสระทั้งหมด 13 ตัวแปร

5. การนำรูปแบบการคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกทวิภาค มาประยุกต์ใช้ในการสร้างและพัฒนาตัวแบบพยากรณ์ในเทคนิคต้นไม้ตัดสินใจ และเทคนิคโครงข่ายประสาทเทียม (เทคนิคเหมือนข้อมูล) สามารถเพิ่มประสิทธิภาพได้ และการใช้ตัวแปรอิสระทุกตัวในการพัฒนาตัวแบบพยากรณ์อาจไม่ได้ตัวแบบที่มีประสิทธิภาพเสมอไป

5.1 อภิปรายผล

จากการวิจัยเรื่องการเปรียบเทียบตัวแบบการถดถอยลอจิสติกและเทคนิคเหมืองข้อมูลสำหรับพยากรณ์การเป็นโรคเบาหวาน เพื่อหาเทคนิคที่ดีที่สุดที่เหมาะสมกับข้อมูลที่ใช้ในการวิจัย พบว่า

1. เทคนิคโครงข่ายประสาทเทียมเหมาะสมในการสร้างตัวแบบพยากรณ์ในชุดข้อมูลแบบที่ 1 ประกอบไปด้วยชุดข้อมูลเรียนรู้ 70 เปอร์เซ็นต์ และชุดข้อมูลทดสอบ 30 เปอร์เซ็นต์ ซึ่งสอดคล้องกับงานวิจัยของสายชล สินสมบูรณ์ทอง (2561)

2. เทคนิคที่มีความเหมาะสมรองลงมา คือเทคนิคต้นไม้ตัดสินใจ อาจเป็นเพราะเทคนิคต้นไม้ตัดสินใจพัฒนามาจากแนวคิดข้อมูลจำแนกประเภท (Categorical Data) แต่ข้อมูลที่ใช้ในการวิจัยเป็นข้อมูลเชิงตัวเลข (Numerical Data) แต่เทคนิคต้นไม้ตัดสินใจเป็นเทคนิคที่เหมาะสมสำหรับข้อมูลจำแนกประเภท (Categorical Data) ซึ่งสอดคล้องกับงานวิจัยของ Manimaran and Muthuraman (2017)

3. การประยุกต์ใช้วิธีการคัดเลือกตัวแปรอิสระในเทคนิคการถดถอยลอจิสติกทวิภาค เพื่อคัดเลือกตัวแปรอิสระที่ส่งผลต่อการเป็นโรคเบาหวาน สามารถพัฒนาตัวแบบพยากรณ์ให้มีประสิทธิภาพมากยิ่งขึ้นได้ ซึ่งสอดคล้องกับงานวิจัย Paul and Latha (2017)

5.2 ข้อเสนอแนะ

จากการวิจัยในครั้งนี้ เพื่อให้เป็นแนวทางในการพัฒนางานวิจัยให้มีประสิทธิภาพมากขึ้น ผู้วิจัยมีข้อเสนอแนะดังนี้

1. การประยุกต์ใช้เทคนิคอื่น ๆ เช่น วิธีบูตสแตรป์ (Bootstrap) เป็นต้นเพื่อแก้ปัญหากรณีชุดข้อมูลไม่สมดุล

2. วิธีการสุ่มตัวอย่างจากชุดข้อมูล และการแบ่งชุดข้อมูลเรียนรู้ และชุดข้อมูลทดสอบ สามารถเลือกใช้เทคนิคอื่น ๆ เช่น 10 Fold Cross – Validation เป็นต้น

3. การประยุกต์ใช้เทคนิคเหมืองข้อมูลเพื่อการพยากรณ์แบบอื่น ๆ เช่น เทคนิคเพื่อนบ้านใกล้สุด k ตัว เทคนิคนาอ็พเบย์ เทคนิคซัพพอร์ตเวกเตอร์แมชชีน และเทคนิคการเรียนรู้เชิงลึก (Deep Learning) เป็นต้น

4. เนื่องจากข้อมูลที่ใช้ในงานวิจัยนี้เป็นข้อมูลผู้ป่วยทั่วไปที่เป็นโรคเบาหวาน ทำให้ตัวแบบพยากรณ์ที่ได้อาจไม่มีประสิทธิภาพมากพอ ดังนั้นการหาข้อมูลผู้ป่วยโรคเบาหวานที่มีความเจาะจงและลักษณะเฉพาะต่อการเป็นโรคเบาหวานอาจจะส่งผลให้ตัวแบบพยากรณ์มีประสิทธิภาพมากยิ่งขึ้น



ภาคผนวก



ภาคผนวก ก

โปรแกรมอาร์

มหาวิทยาลัยพระนคร

โปรแกรมอาร์ ด้วยวิธีการถดถอยลอจิสติกทวิภาค

```
rm(list=ls())  
# install new package  
#install.packages("mlbench")  
library(tidyverse)  
#install.packages("car")  
library(car)  
library(broom)  
library(caret) # confusion matrix  
library(mlbench)  
# Read the Data  
library(MASS)  
library(dplyr)  
library(readr)  
library(readxl)  
library(lmtest)  
library(blorr)  
  
train_d <- read_excel("70new.xlsx", col_types = c("text",  
                                                  "numeric", "numeric", "numeric", "numeric",  
                                                  "numeric", "numeric", "numeric", "numeric",  
                                                  "numeric", "numeric", "numeric", "numeric",  
                                                  "numeric"))  
  
test_d <- read_excel("30new.xlsx", col_types = c("text",  
                                                  "numeric", "numeric", "numeric", "numeric",  
                                                  "numeric", "numeric", "numeric", "numeric",  
                                                  "numeric", "numeric", "numeric", "numeric",  
                                                  "numeric"))  
  
nrow(train_d); nrow(test_d)
```



```

#-----Null-----#
Null<- glm(Outcome ~1, data = train_d , family = binomial("logit"))
summary(Null)
#-----Full Model-----#
fit1 <- glm(Outcome ~ ., data = train_d , family = binomial("logit"))
summary(fit1)
(cbind(OR=exp(coef(fit1)),exp(confint.default(fit1))))
vif(fit1)
blr_vif_tol(fit1)
blr_rsqu_cox_snell(fit1)
blr_rsqu_nagelkerke(fit1)
blr_test_hosmer_lemeshow(fit1, data = train_d)
library(blorr)
#-----เพิ่มกรณีตัวแปรอิสระ sign* จากตัวแปรอิสระทุกตัว-----#
S<-glm(Outcome ~ FBS+bw+ht+dia+sys+pr, data = train_d , family = binomial("logit"))
summary(S)
blr_vif_tol(S)
(cbind(OR=exp(coef(S)),exp(confint.default(S))))
blr_rsqu_cox_snell(S)
blr_rsqu_nagelkerke(S)
#-----forward-----#
blr_step_p_forward(fit1) #เข้า 11 ตัวแปรยกเว้น LDL, Triglyceride
blr_step_p_forward(fit1, penter = 0.05, details = FALSE) #เข้า 6 ตัวแปร
forward2.2<-glm(Outcome ~ FBS+bmi+Cholesterol+dia+sys+pr, data = train_d ,
family = binomial("logit"))
summary(forward2.2)
blr_vif_tol(forward2.2)
(cbind(OR=exp(coef(forward2.2)),exp(confint.default(forward2.2))))
blr_rsqu_cox_snell(forward2.2)
blr_rsqu_nagelkerke(forward2.2)
#-----backward-----#

```

```

blr_step_p_backward(fit1,prem=0.05,details = FALSE) #เข้า 8 ตัวแปร
backward2.2<-glm(Outcome ~ sys+dia+pr+bw+ht+FBS+HDL+LDL, data = train_d ,
family = binomial("logit"))
summary(backward2.2)
blr_vif_tol(backward2.2)
(cbind(OR=exp(coef(backward2.2)),exp(confint.default(backward2.2))))
blr_rsqa_cox_snell(backward2.2)
blr_rsqa_nagelkerke(backward2.2)

#-----Stepwise-----#
blr_step_p_both(fit1, pent = 0.05, prem = 0.05, details = FALSE)
stepwise2.2<-glm(Outcome ~ FBS+bmi+Cholesterol+dia+sys+pr, data = train_d ,
family = binomial("logit"))
summary(stepwise2.2)
blr_vif_tol(stepwise2.2)
(cbind(OR=exp(coef(stepwise2.2)),exp(confint.default(stepwise2.2))))
#-----ทดสอบความเหมาะสมของตัวแบบ-----#
#-----Likelihood Ratio Test & Chi-squared test-----#
#ถ้ายอมรับหรือ p value >0.05 แสดงว่า ตัวแบบนั้นมีความเหมาะสม
#ถ้ายอมรับหรือ p value <0.05 แสดงว่า ตัวแบบนั้นไม่เหมาะสม
anova(forward2.2,fit1, test = "Chisq") #forward2.2ดีกว่า
anova(backward2.2, fit1, test = "Chisq") #fit1 ดีกว่า
AIC(forward2.2,backward2.2)
#-----การทดสอบความมีนัยสำคัญของสัมประสิทธิ์ถดถอย ----#
#ถ้ายอมรับหรือ p value >0.05 แสดงว่า ตัวแบบนั้นไม่ส่งผลต่อ Y#
#ถ้ายอมรับหรือ p value <0.05 แสดงว่า ตัวแบบนั้นส่งผลต่อ Y#
waldtest(forward2.2,fit1) #forward2.2 เป็นตัวแบบที่ส่งผลต่อ Y
waldtest(backward2.2,fit1) #backward2.2 เป็นตัวแบบที่ไม่ส่งผลต่อ Y

#-----สถิติทดสอบระดับความสัมพันธ์ -----#
blr_rsqa_cox_snell(fit1) #0.2141985

```

```

blr_rsqa_cox_snell(forward2.2) #0.2101525
blr_rsqa_cox_snell(backward2.2) #0.2131245

#-----Accuracy-----#
#-----Full-----!
blr_confusion_matrix(fit1,data = train_d) #Accuracy : 0.8104
blr_confusion_matrix(fit1,data = test_d) #Accuracy : 0.8014
#-----forward-----#
blr_confusion_matrix(forward2.2,data = train_d) #Accuracy : 0.8087
blr_confusion_matrix(forward2.2,data = test_d) #Accuracy : 0.8014
#-----backward กรณี p-value <=0.05-----#
blr_confusion_matrix(backward2.2,data = train_d) #Accuracy : 0.8087
blr_confusion_matrix(backward2.2,data = test_d) #Accuracy : 0.7995
#-----stepwise-----#
blr_confusion_matrix(stepwise2.2,data = train_d) #Accuracy : 0.8087
blr_confusion_matrix(stepwise2.2,data = test_d) #Accuracy : 0.8014
#-----เพิ่มกรณีตัวแปรอิสระ sig* จากตัวแปรอิสระทุกตัว-----#
blr_confusion_matrix(S,data = train_d)
blr_confusion_matrix(S,data = test_d)
#-----เพิ่มกรณี correlation -----#
fit1.c <- glm(Outcome
~FBS+Cholesterol+bmi+dia+sys+pr+Triglyceride+LDL+HDL+bw+ht+age, data =
train_d , family = binomial("logit"))
summary(fit1.c)
blr_vif_tol(fit1.c)
(cbind(OR=exp(coef(fit1.c)),exp(confint.default(fit1.c))))
#-----correlation-----#
blr_confusion_matrix(fit1.c,data = train_d) #Accuracy : 0.8107
blr_confusion_matrix(fit1.c,data = test_d) #Accuracy : 0.8014

```

โปรแกรมอาร์ ด้วยวิธีต้นไม้ตัดสินใจ อัลกอริทึม J48

```
rm(list=ls())
# install new package
#install.packages("mlbench")
library(tidyverse)
#install.packages("car")
library(car)
library(broom)
library(caret) # confusion matrix
library(mlbench)
# Read the Data
library(MASS)
library(dplyr)
library(readr)
library(readxl)

#-----กรณี J48-----#ต้องใช้ไฟล์ arff เปิด
library(RWeka)
library(rJava)
#-----Model J48-----#
train_Dt<- read.arff(file= "Train_Yes_No.arff")
test_Dt<- read.arff(file= "Test_Yes_No.arff")
## Identify a decision tree.
DT_fit1<- J48(Outcome~., data = train_Dt)
summary(DT_fit1)
PredTest <- predict(DT_fit1, newdata = test_Dt )
table(test_Dt$Outcome, PredTest) #Accuracy=0.8028
mean(PredTest == test_Dt$Outcome)
```

```

#-----เพิ่มกรณีตัวแปรอิสระ sign* จากตัวแปรอิสระทุกตัว-----#
DT_S<- J48(Outcome~FBS+bw+ht+dia+sys+pr, data = train_Dt)

summary(DT_S)

PredTestS <- predict(DT_S, newdata = test_Dt )

table(test_Dt$Outcome, PredTestS)

mean(PredTestS == test_Dt$Outcome)

#-----forward=FBS+bmi+Cholesterol+dia+sys+pr~~~~~
DT_fit1.1<- J48(Outcome~FBS+bmi+Cholesterol+dia+sys+pr, data = train_Dt)
summary(DT_fit1.1)
PredTest1.1 <- predict(DT_fit1.1, newdata = test_Dt )
table(test_Dt$Outcome, PredTest1.1)
mean(PredTest1.1 == test_Dt$Outcome)

#-----backward=FBS+bw+ht+dia+sys+pr+HDL+LDL~~~~~
DT_fit1.2<- J48(Outcome~FBS+bw+ht+dia+sys+pr+HDL+LDL, data = train_Dt)
summary(DT_fit1.2)
PredTest1.2 <- predict(DT_fit1.2, newdata = test_Dt )
table(test_Dt$Outcome, PredTest1.2)
mean(PredTest1.2 == test_Dt$Outcome)

```

โปรแกรมอาร์ ด้วยวิธีต้นไม้ตัดสินใจ อัลกอริทึม LMT

```
rm(list=ls())
# install new package
#install.packages("mlbench")
library(tidyverse)
#install.packages("car")
library(car)
library(broom)
library(caret) # confusion matrix
library(mlbench)
# Read the Data
library(MASS)
library(dplyr)
library(readr)
library(readxl)

#-----กรณี LMT-----#ต้องใช้ไฟล์ arff เปิด
library(RWeka)
library(rJava)
#-----Model LMT-----#
train_Dt<- read.arff(file= "Train_Yes_No.arff")
test_Dt<- read.arff(file= "Test_Yes_No.arff")
#-----LMT-----#
LM_fit1<-LMT(Outcome~., data = train_Dt)
summary(LM_fit1)
PredTestLM1<- predict(LM_fit1, newdata = test_Dt )
table(test_Dt$Outcome, PredTestLM1)
mean(PredTestLM1== test_Dt$Outcome)
```

```

#-----เพิ่มกรณีตัวแปรอิสระ sign* จากตัวแปรอิสระทุกตัว-----#
LM_S<- LMT(Outcome~FBS+bw+ht+dia+sys+pr, data = train_Dt)
summary(LM_S)
PredTestLMS<- predict(LM_S, newdata = test_Dt )
table(test_Dt$Outcome, PredTestLMS)
mean(PredTestLMS== test_Dt$Outcome)
#-----forward=FBS+bmi+Cholesterol+dia+sys+pr----#
LM_fit1.1<-LMT(Outcome~FBS+bmi+Cholesterol+dia+sys+pr, data = train_Dt)
summary(LM_fit1.1)
PredTestLM1.1<- predict(LM_fit1.1, newdata = test_Dt )
table(test_Dt$Outcome, PredTestLM1.1)
mean(PredTestLM1.1== test_Dt$Outcome)
#-----backward=FBS+bw+ht+dia+sys+pr+HDL+LDL~~~~~
LM_fit1.2<-LMT(Outcome~FBS+bw+ht+dia+sys+pr+HDL+LDL, data = train_Dt)
summary(LM_fit1.2)
PredTestLM1.2<- predict(LM_fit1.2, newdata = test_Dt )
table(test_Dt$Outcome, PredTestLM1.2)
mean(PredTestLM1.2== test_Dt$Outcome)
#-----Stepwise=FBS+bmi+Cholesterol+dia+sys+pr----#
LM_fit1.1<-LMT(Outcome~FBS+bmi+Cholesterol+dia+sys+pr, data = train_Dt)
summary(LM_fit1.1)
PredTestLM1.1<- predict(LM_fit1.1, newdata = test_Dt )
table(test_Dt$Outcome, PredTestLM1.1)
mean(PredTestLM1.1== test_Dt$Outcome)

```

โปรแกรมอาร์ ด้วยวิธีโครงข่ายประสาทเทียม

```
rm(list=ls())
# install new package
#install.packages("mlbench")
library(tidyverse)
#install.packages("car")
library(car)
library(broom)
library(caret) # confusion matrix
library(mlbench)
# Read the Data
library(MASS)
library(dplyr)
library(readr)
library(readxl)

Train_yes<-read.table("Train_Yes_1147.txt",header=T)
Test_yes<-read.table("Test_Yes_1147.txt",header=T)
Train_no<-read.table("Train_No_3934.txt",header=T)
Test_no<-read.table("Test_No_3934.txt",header=T)

## Split data จากการแยกข้อมูล แบ่งข้อมูลออกเป็น 2 ส่วน แต่ละส่วน 70%= train , 30% =
Test
set.seed(0)
train_d <- rbind(Train_yes,Train_no)
test_d<- rbind(Test_yes,Test_no)
nrow(train_d); nrow(test_d)
scaleddata<-scale(train_d)
str(train_d)
```



```

#-----Train-----#
normalize_train <- function(train_d) {
  return ((train_d - min(train_d)) / (max(train_d) - min(train_d)))
}
Trainset <- as.data.frame(lapply(train_d, normalize_train))

#-----Test-----#
normalize_test <- function(test_d) {
  return ((test_d - min(test_d)) / (max(test_d) - min(test_d)))
}
Testset <- as.data.frame(lapply(test_d, normalize_test))

#-----ANN2-----#-----Full model-----#
# Train a neural network on the iris dataset
library(ANN2)
X_train <- Trainset[,1:13]
y_train <- Trainset$Outcome
X_test <- Testset[,1:13]
y_test <- Testset$Outcome
NN <- neuralnetwork(X_train, y_train, hidden.layers = 1,standardize = FALSE,
  loss.type = "log",
  huber.delta = 1,
  activ.functions = "sigmoid",sgd.momentum = 0.1,
  learn.rates = 0.1, val.prop = 0.5, n.epochs = 100,random.seed = 0)

#-----Train set-----#
# Make predictions
y_pred1 <- predict(NN, newdata = X_train)
correct1 <- (y_train == y_pred1$predictions)
table(y_train,y_pred1$predictions)
mean(y_train == y_pred1$predictions)

```

```

#-----Test set-----#
# Make predictions
y_pred2 <- predict(NN, newdata = X_test)
correct <- (y_test == y_pred2$predictions)
table(y_test,y_pred2$predictions)
mean(y_test == y_pred2$predictions)
#-----ANN2-----#----- Significant จาก Full model-----#
# Train a neural network on the iris dataset
#view(Trainset)
library(dplyr)
#view(Trainset)
#เฉพาะตัวแปรอิสระ dia,sys,pr,bw,ht,FBS
D1 <- Trainset[,3:7]
D2 <- Trainset[,9]
X_train_D<-cbind(D1,D2)
#เฉพาะตัวแปรตาม y-----
y_train_D <- Trainset$Outcome
#เฉพาะตัวแปรอิสระ dia,sys,pr,bw,ht,FBSL
DD1 <- Testset[,3:7]
DD2 <- Testset[,9]
X_test_D<-cbind(DD1,DD2)
#เฉพาะตัวแปรตาม y-----
y_test_D <- Testset$Outcome
#-----สร้างตัวแบบ-----
NND <- neuralnetwork(X_train_D, y_train_D, hidden.layers = 6,standardize = FALSE,
                    loss.type = "log",
                    huber.delta = 1,
                    activ.functions = "sigmoid",learn.rates = 0.3,sgd.momentum = 0.3,
                    val.prop = 0.5, n.epochs = 100,random.seed = 0)

```

```

#-----Train set-----#
# Make predictions
y_pred1D <- predict(NND, newdata = X_train_D)
correct1D <- (y_train_D == y_pred1D$predictions)
table(y_train_D,y_pred1D$predictions)
mean(y_train_D == y_pred1D$predictions)
#-----Test set-----#
# Make predictions
y_pred2D <- predict(NND, newdata = X_test_D)
correctD <- (y_test_D == y_pred2D$predictions)
table(y_test_D,y_pred2D$predictions)
mean(y_test_D == y_pred2D$predictions)
#-----#
#-----ANN2-----#----- Forward-----#
# Train a neural network on the iris dataset
library(dplyr)
#view(Trainset)
#เฉพาะตัวแปรอิสระ dia,sys,pr,FBS,bmi,cholesterol
X1 <- Trainset[,3:5]
X2 <- Trainset[,8:10]
X_train_F<-cbind(X1,X2)
#เฉพาะตัวแปรตาม y-----
y_train_F <- Trainset$Outcome
#เฉพาะตัวแปรอิสระ dia,sys,pr,FBS,bmi,cholesterol
XX1 <- Testset[,3:5]
XX2 <- Testset[,8:10]
X_test_F<-cbind(XX1,XX2)
#เฉพาะตัวแปรตาม y-----
y_test_F <- Testset$Outcome

```

```

#-----สร้างตัวแบบ-----
NNF <- neuralnetwork(X_train_F, y_train_F, hidden.layers = 6,standardize = FALSE,
  loss.type = "log",
  huber.delta = 1,
  activ.functions = "sigmoid",learn.rates = 0.2,sgd.momentum = 0.3,
  val.prop = 0.5, n.epochs = 100,random.seed = 0)

#-----Train set-----#
# Make predictions
y_pred1F <- predict(NNF, newdata = X_train_F)
correct1F <- (y_train_F == y_pred1F$predictions)
table(y_train_F,y_pred1F$predictions)
mean(y_train_F == y_pred1F$predictions)
#-----Test set-----#
# Make predictions
y_pred2F <- predict(NNF, newdata = X_test_F)
correctF <- (y_test_F == y_pred2F$predictions)
table(y_test_F,y_pred2F$predictions)
mean(y_test_F == y_pred2F$predictions)
#-----#

#-----ANN2-----#----- Backward -----#
# Train a neural network on the iris dataset
library(dplyr)
#view(Trainset)
#เฉพาะตัวแปรอิสระ dia,sys,pr,bw,ht,FBS,HDL,LDL
Z1 <- Trainset[,3:7]
Z2 <- Trainset[,9]
Z3 <- Trainset[,12:13]
X_train_B<-cbind(Z1,Z2,Z3)
#เฉพาะตัวแปรตาม y-----
y_train_B <- Trainset$Outcome

```

```

#เฉพาะตัวแปรอิสระ dia,sys,pr,bw,ht,FBS,HDL,LDL
ZZ1 <- Testset[,3:7]
ZZ2 <- Testset[,9]
ZZ3 <- Testset[,12:13]
X_test_B<-cbind(ZZ1,ZZ2,ZZ3)
#เฉพาะตัวแปรตาม y-----
y_test_B <- Testset$Outcome
#-----สร้างตัวแบบ-----
NNB <- neuralnetwork(X_train_B, y_train_B, hidden.layers = 10,standardize = FALSE,
  loss.type = "log",
  huber.delta = 1,
  activ.functions = "sigmoid",learn.rates = 0.3,sgd.momentum = 0.2,
  val.prop = 0.5, n.epochs = 100,random.seed = 0)

# Plot the loss during training

#-----Train set-----#
# Make predictions
y_pred1B <- predict(NNB, newdata = X_train_B)
correct1B <- (y_train_B == y_pred1B$predictions)
table(y_train_B,y_pred1B$predictions)
mean(y_train_B == y_pred1B$predictions)
#-----Test set-----#
# Make predictions
y_pred2B <- predict(NNB, newdata = X_test_B)
correctB <- (y_test_B == y_pred2B$predictions)
table(y_test_B,y_pred2B$predictions)
mean(y_test_B == y_pred2B$predictions)
#-----#

```



ภาคผนวก ข

ผลลัพธ์จากเทคนิคโครงข่ายประสาทเทียม

มหาวิทยาลัยสุรินทร์

แสดงผลประสิทธิภาพของชุดข้อมูลแบบที่ 1 ข้อมูลเรียนรู้ 70 เปอร์เซนต์ และชุดข้อมูลทดสอบ 30 เปอร์เซนต์ จากเทคนิคโครงข่ายประสาทเทียม กรณีตัวแปรอิสระทุกตัว

Node	learning/momentum 0.1/0.1		learning/momentum 0.1/0.2		learning/momentum 0.1/0.3		learning/momentum 0.2/0.1		learning/momentum 0.2/0.2		learning/momentum 0.2/0.3		learning/momentum 0.3/0.1		learning/momentum 0.3/0.2		learning/momentum 0.3/0.3		
	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	
5	80.0281	79.8820	80.0000	79.6199	80.3657	79.9476	81.5190	80.5374	81.7159	80.3408	81.5753	80.6029	81.7159	80.4718	81.7440	80.6684	81.8847	80.4718	81.8847
6	80.1969	79.6199	80.0563	79.7510	80.4219	79.6199	81.6034	80.0131	81.3783	79.4233	81.6034	79.3578	81.6878	79.2923	81.5471	79.2923	81.4909	79.4233	81.4909
8	80.1688	80.2097	80.2813	80.2097	80.4219	80.1442	80.9564	80.8650	80.8720	80.7339	80.7314	80.9305	81.4909	80.9305	81.2940	81.1271	80.9564	81.0616	80.9564
9	79.3530	79.2267	79.2686	79.8820	79.5781	80.2097	79.8875	80.2097	80.1125	80.5374	80.0563	80.3408	80.0563	80.2752	80.1688	80.6684	80.1688	80.6684	80.1688
10	79.8875	79.6855	80.2250	79.8820	80.4782	80.0131	81.3783	80.2097	81.2377	80.1441	81.2658	80.3408	81.5471	80.4718	81.5471	80.6684	81.4627	80.4718	81.4627

แสดงผลประสิทธิภาพของชุดข้อมูลแบบที่ 1 ข้อมูลเรียนรู้ 70 เปอร์เซนต์ และชุดข้อมูลทดสอบ 30 เปอร์เซนต์ จากเทคนิคโครงข่ายประสาทเทียมวิธีการเลือกแบบพื้นฐาน

Node	learning/momentum 0.1/0.1		learning/momentum 0.1/0.2		learning/momentum 0.1/0.3		learning/momentum 0.2/0.1		learning/momentum 0.2/0.2		learning/momentum 0.2/0.3		learning/momentum 0.3/0.1		learning/momentum 0.3/0.2		learning/momentum 0.3/0.3		
	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	
5	79.8031	79.4233	80.1688	79.9476	80.2250	79.9476	80.7595	80.7995	80.7595	80.7339	80.8158	80.7339	80.9283	80.9305	80.9283	81.1271	80.9845	81.1271	80.9845
6	80.1125	80.4063	80.0563	80.4063	80.4782	80.6029	81.3221	80.8650	81.5471	80.9961	81.9691	81.1927	81.4346	81.1271	81.9410	81.3893	81.9128	81.5203	81.9128
8	79.9719	79.8820	80.1969	80.6684	80.6470	80.2097	81.4909	80.9305	81.6315	81.1271	81.6596	81.2582	81.8847	81.2582	81.9691	81.1927	81.9409	81.1927	81.9409
9	80.3938	80.3408	80.3657	80.4063	80.1688	80.9305	80.6189	80.8650	80.6470	80.9961	80.6751	80.9305	80.8158	80.9961	80.8720	80.9961	80.7876	81.1271	80.7876
10	80.0563	87.2752	80.3376	80.7339	80.2532	80.8650	80.3094	87.8650	80.3938	80.8650	80.6751	80.7995	80.3657	80.9961	80.5626	80.7339	80.6189	80.8650	80.6189

แสดงผลประสิทธิภาพของชุดข้อมูลแบบที่ 1 ข้อมูลเรียนรู้ 70 เปอร์เซนต์ และชุดข้อมูลทดสอบ 30 เปอร์เซนต์ จากเทคนิคโครงข่ายประสาทเทียมวิธีการเลือกแบบไปข้างหน้า

Node	0.1/0.1		0.1/0.2		0.1/0.3		0.2/0.1		0.2/0.2		0.2/0.3		0.3/0.1		0.3/0.2		0.3/0.3		
	learning/momentum	Test	learning/momentum	Test	learning/momentum	Test	learning/momentum	Test	learning/momentum	Test	learning/momentum	Test	learning/momentum	Test	learning/momentum	Test	learning/momentum	Test	
5	79.8031	79.7510	80.1407	80.0131	80.1125	80.4718	80.9564	80.8650	81.0971	80.9305	81.3221	81.0616	81.2377	80.7339	81.3221	80.9305	81.3221	80.9305	81.1271
6	80.0281	79.8165	80.2813	80.2097	80.2250	80.6684	81.1252	81.3237	81.5471	81.5203	81.7721	81.7824	81.5753	81.7169	81.8284	81.7169	81.8284	81.7169	81.5859
8	80.0563	79.8820	80.5345	80.8650	80.7314	80.6684	81.4065	80.7339	81.4627	80.9305	81.4909	80.7339	81.6596	81.0616	81.6596	80.9961	81.7440	81.2582	81.2582
9	79.7187	79.7510	80.0000	80.1442	80.2250	80.8650	80.5907	81.2582	80.5345	81.4548	80.6470	81.4548	80.9283	81.3237	80.8439	81.3237	80.7876	81.2582	81.2582
10	79.7468	79.6855	80.0563	79.8820	80.0563	80.2752	80.1407	80.6684	80.0281	80.7995	80.1125	81.0616	80.1407	80.7995	80.1407	80.7339	80.0844	81.2582	81.2582

แสดงผลประสิทธิภาพของชุดข้อมูลแบบที่ 1 ข้อมูลเรียนรู้ 70 เปอร์เซนต์ และชุดข้อมูลทดสอบ 30 เปอร์เซนต์ จากเทคนิคโครงข่ายประสาทเทียมวิธีการจัดแบบถอยหลัง

Node	0.1/0.1		0.1/0.2		0.1/0.3		0.2/0.1		0.2/0.2		0.2/0.3		0.3/0.1		0.3/0.2		0.3/0.3		
	learning/momentum	Test	learning/momentum	Test	learning/momentum	Test	learning/momentum	Test	learning/momentum	Test	learning/momentum	Test	learning/momentum	Test	learning/momentum	Test	learning/momentum	Test	
5	80.0563	79.9476	79.8594	80.0786	80.0563	80.0131	80.8439	80.7995	80.8439	80.6684	81.2096	80.9305	81.1533	80.9305	81.2377	81.0616	81.3783	81.0616	81.0616
6	79.8875	80.3408	80.3376	80.0786	80.3376	80.2752	80.7876	80.9305	80.8439	80.8650	80.8158	80.9961	80.7876	81.1271	80.8158	81.2582	81.0127	81.2582	81.2582
8	80.1125	80.0786	80.3376	80.0786	80.3094	80.2752	80.2813	80.6029	80.2813	80.7339	80.3938	80.7339	80.3094	80.7339	80.3094	80.7995	80.3938	80.9961	80.9961
9	80.0281	79.6199	80.2532	79.8820	80.5345	79.8820	81.3502	80.0786	81.6596	79.9476	81.6315	80.0131	81.6596	80.0786	81.8847	80.3408	81.8003	80.0786	80.0786
10	80.0563	80.0786	80.1688	80.2097	80.3376	80.2752	80.6470	80.4718	80.9001	80.5374	81.0408	80.7339	81.2377	80.7339	81.2377	80.7995	81.2940	80.8650	80.8650

แสดงผลประสิทธิภาพของชุดข้อมูลแบบที่ 1 ข้อมูลเรียนรู้ 70 เปอร์เซ็นต์ และชุดข้อมูลทดสอบ 30 เปอร์เซ็นต์ จากเทคนิคโครงข่ายประสาทเทียมวิธีการถดถอยที่ละชั้น

Node	0.1/0.1		0.1/0.2		0.1/0.3		0.2/0.1		0.2/0.2		0.2/0.3		0.3/0.1		0.3/0.2		0.3/0.3	
	learning/momentum	Test	learning/momentum	Test	learning/momentum	Test	learning/momentum	Test	learning/momentum	Test	learning/momentum	Test	learning/momentum	Test	learning/momentum	Test	learning/momentum	Test
5	79.8031	79.7510	80.1407	80.0131	80.1125	80.4718	80.9564	80.8650	81.0971	80.9305	81.3221	81.0616	81.2377	80.7339	81.3221	80.9305	81.3783	81.1271
6	80.0281	79.8165	80.2813	80.2097	80.2250	80.6684	81.1252	81.3237	81.5471	81.5203	81.7721	81.7824	81.5753	81.7169	81.8284	81.7169	82.0253	81.5859
8	80.0563	79.8820	80.5345	80.8650	80.7314	80.6684	81.4065	80.7339	81.4627	80.9305	81.4909	80.7339	81.6596	81.0616	81.6596	80.9961	81.7440	81.2582
9	79.7187	79.7510	80.0000	80.1442	80.2250	80.8650	80.5907	81.2582	80.5345	81.4548	80.6470	81.4548	80.9283	81.3237	80.8439	81.3237	80.7876	81.2582
10	79.7468	79.6855	80.0563	79.8820	80.0563	80.2752	80.1407	80.6684	80.0281	80.7995	80.1125	81.0616	80.1407	80.7995	80.1407	80.7339	80.0844	81.2582



แสดงผลประสิทธิภาพของชุดข้อมูลแบบที่ 2 ข้อมูลเรียนรู้ 80 เปอร์เซ็นต์ และชุดข้อมูลทดสอบ 20 เปอร์เซ็นต์ จากเทคนิคโครงข่ายประสาทเทียม กรณีตัวแปรอิสระทุกตัว

Node	learning/momentum 0.1/0.1		learning/momentum 0.1/0.2		learning/momentum 0.1/0.3		learning/momentum 0.2/0.1		learning/momentum 0.2/0.2		learning/momentum 0.2/0.3		learning/momentum 0.3/0.1		learning/momentum 0.3/0.2		learning/momentum 0.3/0.3		
	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	
5	79.7490	78.2694	80.0197	79.0561	80.2165	79.5477	80.6102	80.3343	80.5856	80.7276	80.6841	80.6293	80.7333	80.5310	80.7579	80.5310	80.7579	80.5310	80.7276
6	79.7982	78.8594	80.0689	78.7611	80.3642	79.0561	80.9793	79.0561	81.1270	79.1544	81.1270	78.5644	81.1762	79.2527	81.2008	78.2694	81.4715	77.7778	77.7778
8	80.2658	77.7778	80.6841	77.8761	81.1270	77.9744	81.5945	76.2045	81.5699	76.2045	81.5699	76.3029	81.6683	75.3196	81.8652	75.9095	81.9144	76.4995	76.4995
9	79.9951	79.3510	80.3642	79.4494	80.3396	79.6460	80.5856	80.6293	80.6595	80.2360	80.6348	80.4327	80.5364	80.7276	80.8071	80.6293	80.9793	80.1377	80.1377
10	79.8967	78.7611	80.0935	78.6627	80.2411	79.1544	80.5364	80.2360	80.5364	80.5310	80.5610	80.5310	80.4626	80.2693	80.4626	80.7276	80.4380	80.7276	80.7276

แสดงผลประสิทธิภาพของชุดข้อมูลแบบที่ 2 ข้อมูลเรียนรู้ 80 เปอร์เซ็นต์ และชุดข้อมูลทดสอบ 20 เปอร์เซ็นต์ จากเทคนิคโครงข่ายประสาทเทียมวิธีการเลือกแบบพื้นฐาน

Node	learning/momentum 0.1/0.1		learning/momentum 0.1/0.2		learning/momentum 0.1/0.3		learning/momentum 0.2/0.1		learning/momentum 0.2/0.2		learning/momentum 0.2/0.3		learning/momentum 0.3/0.1		learning/momentum 0.3/0.2		learning/momentum 0.3/0.3		
	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	
5	79.7982	79.8427	80.2411	79.5477	80.2411	80.3343	80.3642	80.7276	80.5118	80.7276	80.5364	80.7276	80.5118	80.9243	80.5610	80.8260	80.4626	81.1210	81.1210
6	80.069	80.0393	80.4134	79.2527	80.7579	78.2694	81.0532	79.0561	81.2254	78.6627	81.4469	78.5644	81.0778	78.8594	81.3238	78.5644	81.5699	77.9744	77.9744
8	80.2411	80.4327	80.2411	79.8427	80.6595	79.5477	80.8810	79.7444	81.0532	78.6672	81.1270	78.5644	80.9793	80.0393	81.1516	78.9577	81.3484	78.5644	78.5644
9	80.1673	80.7276	80.0935	80.8260	80.2658	81.0226	80.5856	81.2193	80.7087	81.1209	80.7579	81.0226	80.6348	81.4159	80.6595	81.3176	80.7825	80.6293	80.6293
10	80.3396	78.8594	80.7579	78.2694	80.7825	78.1711	81.1762	78.0728	81.2254	78.0728	81.1516	78.5644	81.6437	78.2694	81.6437	78.5644	81.5945	78.5644	78.5644

แสดงผลประสิทธิภาพของชุดข้อมูลแบบที่ 2 ข้อมูลเรียนรู้ 80 เปอร์เซนต์ และชุดข้อมูลทดสอบ 20 เปอร์เซนต์ จากเทคนิคโครงข่ายประสาทเทียมวิธีการเลือกแบบไปข้างหน้า

Node	0.1/0.1		0.1/0.2		0.1/0.3		0.2/0.1		0.2/0.2		0.2/0.3		0.3/0.1		0.3/0.2		0.3/0.3		
	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	
5	79.7490	79.3510	79.9951	79.6460	79.9951	79.1544	80.2904	79.6460	80.3396	79.6460	80.3150	79.6460	80.4134	79.8427	80.3888	79.9410	80.4380	79.9410	80.4380
6	79.9459	78.8594	80.0935	77.9744	80.4380	76.0079	81.0039	75.7129	81.3238	74.8279	81.5699	74.1396	81.4222	75.4179	81.4715	74.8279	81.5945	74.0413	74.0413
8	80.2904	78.8594	80.4134	78.6627	80.6841	77.5811	81.1270	77.1878	81.2254	76.3029	81.4961	75.8112	81.1516	77.4828	81.4961	76.7945	81.5207	76.0079	76.0079
9	79.8228	79.8427	79.9459	79.4994	79.9705	79.7444	80.3396	79.1544	80.4134	79.4494	80.5364	79.2527	80.4872	79.5477	80.5610	79.5477	80.6841	79.0561	79.0561
10	80.4380	78.1711	80.5610	77.9744	80.9301	76.6962	81.3484	75.9095	81.4469	76.1062	81.3976	76.4012	81.4222	75.8112	81.5207	75.9095	81.5945	75.9095	75.9095

แสดงผลประสิทธิภาพของชุดข้อมูลแบบที่ 2 ข้อมูลเรียนรู้ 80 เปอร์เซนต์ และชุดข้อมูลทดสอบ 20 เปอร์เซนต์ จากเทคนิคโครงข่ายประสาทเทียมวิธีการกำจัดแบบเคออส

Node	0.1/0.1		0.1/0.2		0.1/0.3		0.2/0.1		0.2/0.2		0.2/0.3		0.3/0.1		0.3/0.2		0.3/0.3		
	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	
5	80.1427	80.1377	80.0443	80.4327	80.1919	80.0393	80.9793	80.0393	80.9055	79.9410	80.9793	80.0393	81.4469	79.6460	81.4715	79.6460	81.3976	79.7444	79.7444
6	80.0197	79.8427	79.9213	79.9410	80.2904	79.9410	81.1270	79.5477	81.2254	79.6460	81.2764	79.8427	81.5699	79.3510	81.5945	79.6460	81.7175	79.6460	79.6460
8	80.3396	79.1544	80.5856	79.0561	80.6595	79.5477	81.1024	79.3510	81.1762	79.4494	81.1024	79.1544	81.2746	79.4494	81.2500	79.3510	81.2254	79.2527	79.2527
9	79.9705	79.8427	80.1673	80.2360	80.2165	79.6460	81.0039	79.5477	80.2992	79.3510	81.1516	79.8427	81.2500	79.8427	81.2746	79.6460	81.2992	79.4494	79.4494
10	79.9213	80.1377	80.0443	80.1377	80.1919	80.7276	80.4872	80.3343	80.5610	80.6293	80.9055	80.7276	80.6841	80.4327	80.9547	80.5310	81.0285	80.1377	80.1377

แสดงผลประสิทธิภาพของชุดคอมแบบที่ 2 ข้อมูลเรียนรู้ 80 เปอร์เซ็นต์ และชุดข้อมูลทดสอบ 20 เปอร์เซ็นต์ จากเทคนิคโครงข่ายประสาทเทียมวิธีการดอยทีละชั้น

Node	0.1/0.1		0.1/0.2		0.1/0.3		0.2/0.1		0.2/0.2		0.2/0.3		0.3/0.1		0.3/0.2		0.3/0.3	
	learning/momentum	Test	learning/momentum	Test	learning/momentum	Test	learning/momentum	Test	learning/momentum	Test	learning/momentum	Test	learning/momentum	Test	learning/momentum	Test	learning/momentum	Test
5	79.7490	79.3510	79.9951	79.6460	79.9951	79.1544	80.2904	79.6460	80.3396	79.6460	80.3150	79.6460	80.4134	79.8427	80.3888	79.9410	80.4380	79.9410
6	79.9459	78.8594	80.0935	77.9744	80.4380	76.0079	81.0039	75.7129	81.3238	74.8279	81.5699	74.1396	81.4222	75.4179	81.4715	74.8279	81.5945	74.0413
8	80.2904	78.8594	80.4134	78.6627	80.6841	77.5811	81.1270	77.1878	81.2254	76.3029	81.4961	75.8112	81.1516	77.4828	81.4961	76.7945	81.5207	76.0079
9	79.8228	79.8427	79.9459	79.4994	79.9705	79.7444	80.3396	79.1544	80.4134	79.4494	80.5364	79.2527	80.4872	79.5477	80.5610	79.5477	80.6841	79.0561
10	80.4380	78.1711	80.5610	77.9744	80.9301	76.6962	81.3484	75.9095	81.4469	76.1062	81.3976	76.4012	81.4222	75.8112	81.5207	75.9095	81.5945	75.9095



บรรณานุกรม

- Haberman, S. J. (1978). *Analysis of Qualitative Data*. New York.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and techniques* (3 ed.). CA: Morgan Kaufmann.
- IDF Diabetes Atlas 9th Edition 2019. (2019). (9, Ed.) www.diabetesatlas.org
- Kassambara, A. (2018). *Machine Learning Essentials: Practical Guide in R*. sthda. <https://books.google.co.th/books?id=745ODwAAOBAJ>
- Kumar, P., & Umatejaswi, V. (2017). Diagnosing Diabetes using Data Mining Techniques. *International Journal of Scientific and Research Publications*, 7(6), 785-793.
- Larose, D. T., & Larose, C. D. (2014). *Discovering knowledge in data : an introduction to data mining* (2nd ed. ed.) [Non-fiction]. Wiley-Interscience. <http://search.ebscohost.com/login.aspx?direct=true&AuthType=sso&db=cat07210a&AN=nur.367322&site=eds-live&custid=ns004377>
- Manimaran, R., & Muthuraman, V. (2017). Prediction of Diabetes Disease Using Classification Data Mining Techniques. *International Journal of Engineering and Technology*, 9, 3610-3614. <https://doi.org/10.21817/ijet/2017/v9i5/170905319>
- Maulana, M., & Defriani, M. (2020). Logistic Model Tree and Decision Tree J48 Algorithms for Predicting the Length of Study Period. *PIKSEL : Penelitian Ilmu Komputer Sistem Embedded and Logic*, 8, 39-48. <https://doi.org/10.33558/piksel.v8i1.2018>
- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2020). *Introduction to linear regression analysis* (Fifth edition. ed.). Wiley,.
- Myers, R. H. (1986). *Classical and modern regression with applications*. Duxbury Press.
- Myers, R. H. (1990). *Classical and modern regression with applications* (2nd ed.). PWS-KENT. Publisher description <http://www.loc.gov/catdir/enhancements/fy1515/90146188-d.html>
- Table of contents only <http://www.loc.gov/catdir/enhancements/fy1515/90146188-t.html>
- Paul, S., . and Latha,C.B.,. (2017). Prediction of Diabetes using a Classification model Dr.Sujni Paul. *Al Dar Research Journal For Sustainability*, 2(2), 1-12.

- Perveen, S., Shahbaz, M., Guergachi, A., & Keshavjee, K. (2016). Performance Analysis of Data Mining Classification Techniques to Predict Diabetes. *Procedia Computer Science*, 82, 115-121. <https://doi.org/https://doi.org/10.1016/j.procs.2016.04.016>
- Sisodia, D., & Sisodia, D. S. (2018). Prediction of Diabetes using Classification Algorithms. *Procedia Computer Science*, 132, 1578-1585. <https://doi.org/https://doi.org/10.1016/j.procs.2018.05.122>
- Sunge, A. S., Warnar, H. L. H. S., Heryadi, Y., Abdurachman, E., Soewito, B., & Gaol, F. L. (2019). Prediction Diabetes Mellitus Using Decision Tree Models. *International Congress on Applied Information Technology* 4(1), 189-198. <https://doi.org/10.1109/AIT49014.2019.9144971>
- Yang, H., & Fong, S. (2011, 24-26 Oct. 2011). Optimized very fast decision tree with balanced classification accuracy and compact tree size. The 3rd International Conference on Data Mining and Intelligent Information Technology Applications, กัลยา วาณิชย์บัญชา. (2555). การวิเคราะห์ข้อมูลหลายตัวแปร (พิมพ์ครั้งที่ 3 ed.). ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย.
- ณัฐกรกุล ไชยสงคราม. (2561). สรุปรายงานการเฝ้าระวังโรคประจำปี 2561. สำนักระบาดวิทยา กรมควบคุมโรค กระทรวงสาธารณสุข. https://apps.doe.moph.go.th/boeeng/download/AW_Annual_Mix%206212_14_r1.pdf
- พจนา แวสวัสดิ์. (2547). เกณฑ์การคัดเลือกตัวแบบสำหรับความถดถอยเชิงเส้นตรง. *วารสารศรีปทุมปริทัศน์*, 4(1), 5-20.
- ภักสุภางค์ มาปรีดา. (2560). ตัวแบบการถดถอยลอจิสติกในการพยากรณ์ความน่าจะเป็นของการชำระหนี้ได้ของครัวเรือน กรณีศึกษาจังหวัดปทุมธานี มหาวิทยาลัยธรรมศาสตร์. ปทุมธานี.
- ยุทธ ไกยวรรณ. (2555). หลักการและการใช้การวิเคราะห์การถดถอยลอจิสติกสำหรับการวิจัย. *วารสารวิจัยมหาวิทยาลัยเทคโนโลยีราชมงคลศรีวิชัย*, 4(1), 1-12.
- ศิริชัย พงษ์วิชัย. (2549). การวิเคราะห์ข้อมูลทางสถิติด้วยคอมพิวเตอร์ (พิมพ์ครั้งที่ 16 ed.). จุฬาลงกรณ์มหาวิทยาลัย.
- สมาคมโรคเบาหวานแห่งประเทศไทยในพระราชูปถัมภ์สมเด็จพระเทพรัตนราชสุดาฯ สยามบรมราชกุมารี. (2560). แนวทางเวชปฏิบัติสำหรับโรคเบาหวาน 2560 (พิมพ์ครั้งที่ 3 ed.). รมเย็นมีเดีย. <https://www.dmthai.org/attachments/article/443/guideline-diabetes-care-2017.pdf>

- สายชล สิ้นสมบูรณ์ทอง. (2560). การทำเหมืองข้อมูลเล่ม 1 การค้นหาความรู้จากข้อมูล (พิมพ์ครั้งที่ 2 ed.). จามจุรีโปรดักส์.
- สายชล สิ้นสมบูรณ์ทอง. (2561). การเปรียบเทียบประสิทธิภาพในการทำนายผลการเป็นโรคเบาหวาน. วารสารวิทยาศาสตร์และเทคโนโลยี, 26(2), 195-207.
- สำราญ มีแจ่ม. (2546). สถิติขั้นสูงสำหรับการวิจัย (พิมพ์ครั้งที่ 2 ed.). นิชินแอตเวอร์ไทซิงกรุ๊ป.



