



การเปรียบเทียบประสิทธิภาพของตัวแบบการจำแนกข้อมูลด้านสุขภาพ



วิทยานิพนธ์เสนอบัณฑิตวิทยาลัย มหาวิทยาลัยนเรศวร
เพื่อเป็นส่วนหนึ่งของการศึกษา หลักสูตรวิทยาศาสตรมหาบัณฑิต
สาขาวิชาสถิติ
ปีการศึกษา 2566
ลิขสิทธิ์เป็นของมหาวิทยาลัยนเรศวร

การเปรียบเทียบประสิทธิภาพของตัวแบบการจำแนกข้อมูลด้านสุขภาพ



วิทยานิพนธ์เสนอบัณฑิตวิทยาลัย มหาวิทยาลัยนเรศวร
เพื่อเป็นส่วนหนึ่งของการศึกษา หลักสูตรวิทยาศาสตรมหาบัณฑิต
สาขาวิชาสถิติ
ปีการศึกษา 2566
ลิขสิทธิ์เป็นของมหาวิทยาลัยนเรศวร

วิทยานิพนธ์ เรื่อง "การเปรียบเทียบประสิทธิภาพของตัวแบบการจำแนกข้อมูลด้านสุขภาพ"
ของ จีรนนท์ ป้องท้าว
ได้รับการพิจารณาให้นับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาสถิติ

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการสอบวิทยานิพนธ์
(รองศาสตราจารย์ ดร.อชมา อระวีพร)

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

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

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

อนุมัติ

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

ชื่อเรื่อง	การเปรียบเทียบประสิทธิภาพของตัวแบบการจำแนกข้อมูลด้าน สุขภาพ
ผู้วิจัย	จิรนนท์ ป้องท้าว
ประธานที่ปรึกษา	รองศาสตราจารย์ ดร.อนามัย นาอุดม
กรรมการที่ปรึกษา	ผู้ช่วยศาสตราจารย์ ดร.จรัสศรี รุ่งรัตนอุบล
ประเภทสารนิพนธ์	วิทยานิพนธ์ วท.ม. สถิติ, มหาวิทยาลัยนเรศวร, 2566
คำสำคัญ	เทคนิคการถดถอยลอจิสติก, เทคนิคต้นไม้ตัดสินใจ, เทคนิคป่าสุ่ม, เทคนิคซัพพอร์ตเวกเตอร์แมชชีน, เทคนิคการคัดเลือกตัวแปรอิสระ, ข้อมูลด้านสุขภาพ

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาการทำงานและเปรียบเทียบประสิทธิภาพการจำแนกของเทคนิคการจำแนกทั้ง 4 เทคนิค ประกอบด้วย เทคนิคการถดถอยลอจิสติก เทคนิคต้นไม้ตัดสินใจ เทคนิคป่าสุ่ม และเทคนิคซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ข้อมูลด้านสุขภาพที่มีจำนวนตัวแปรอิสระเชิงปริมาณและตัวแปรอิสระเชิงคุณภาพแตกต่างกันทั้งหมด 3 ชุด ได้แก่ ชุดข้อมูลโรคหัวใจและหลอดเลือดซึ่งเป็นชุดข้อมูลที่มีจำนวนตัวแปรอิสระเชิงปริมาณใกล้เคียงตัวแปรอิสระเชิงคุณภาพ ชุดข้อมูลโรคเบาหวานซึ่งเป็นชุดข้อมูลที่มีจำนวนตัวแปรอิสระเชิงปริมาณน้อยกว่าตัวแปรอิสระเชิงคุณภาพ และชุดข้อมูลการเลิกสูบบุหรี่ซึ่งเป็นชุดข้อมูลที่มีจำนวนตัวแปรอิสระเชิงปริมาณมากกว่าตัวแปรอิสระเชิงคุณภาพ และคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติก และค่าสัมประสิทธิ์สหสัมพันธ์ โดยทดสอบประสิทธิภาพของตัวแบบการจำแนกด้วยหลักการ 10 – Fold Cross-Validation จากนั้นวัดประสิทธิภาพของตัวแบบการจำแนกด้วยค่าความไว ค่าความจำเพาะ ค่าความแม่นยำ ค่าความแม่นยำที่สมดุลแบบปรับน้ำหนักถ่วง และค่าประสิทธิภาพโดยรวมแบบปรับน้ำหนักถ่วง จากการศึกษาพบว่า ชุดข้อมูลโรคหัวใจและหลอดเลือด และชุดข้อมูลโรคเบาหวาน เทคนิคการจำแนกที่มีประสิทธิภาพมากที่สุดคือ เทคนิคซัพพอร์ตเวกเตอร์แมชชีนโดยใช้ตัวแปรอิสระทั้งหมด ซึ่งมีค่าความแม่นยำที่สมดุลแบบปรับถ่วงน้ำหนักเท่ากับ 73.76% และ 75.39% ตามลำดับ และชุดข้อมูลการเลิกสูบบุหรี่ เทคนิคการจำแนกที่มีประสิทธิภาพมากที่สุด คือ เทคนิคซัพพอร์ตเวกเตอร์แมชชีนที่มีการคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกโดยใช้ข้อมูลทั้งหมด ซึ่งมีค่าความแม่นยำที่สมดุลแบบปรับน้ำหนักถ่วงเท่ากับ 75.41%

อีกทั้งยังพบว่า การคัดเลือกตัวแปรอิสระทำให้ตัวแบบการจำแนกด้วยเทคนิคการถดถอย
ลอจิสติก และเทคนิคต้นไม้ตัดสินใจมีประสิทธิภาพในการจำแนกดีขึ้น ในขณะที่เทคนิคป่าสุ่มและ
เทคนิคซัพพอร์ตเวกเตอร์แมชชีนมีประสิทธิภาพในการจำแนกลดลง



Title	EFFICIENCY COMPARISON OF CLASSIFICATION MODEL FOR HEALTH DATA
Author	Jiranan Pongthao
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, 2023
Keywords	Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, Variable Selection, Health Data

ABSTRACT

The purpose of this research is to study and compare the performance of four classification techniques, including logistic regression, decision tree, random forest, and support vector machine, by using three health datasets with different numbers of quantitative and qualitative independent variables, namely the cardiovascular disease dataset, with a similar number of quantitative and qualitative independent variables; the diabetes dataset, which consists of smaller number of quantitative than qualitative independent variables; and the smoking dataset, which includes larger number of quantitative than qualitative independent variables. The independent variables were selected using logistic regression and the correlation coefficient. The performance of the models was evaluated using the 10-fold cross-validation technique, by considering various criteria, which are sensitivity, specificity, accuracy, weighted balance accuracy, and $f1_{\text{weight}}$. The results show that in the case of the cardiovascular disease dataset and the diabetes dataset, the most effective classification technique is support vector machines using all independent variables, which provides a weighted balanced accuracy value of 73.76% and 75.39%, respectively. For the smoking dataset, the most effective classification technique is the support vector machine with independent variables selection using the logistic regression techniques based on the total dataset, which has a weighted balanced accuracy value of 75.41%.

It is also found that, the selection of independent variables enhances the logistic regression and the decision tree classifier in terms of better classification performance. Whereas the performances of the random forest and support vector machine classifiers decrease when some independent variables are selected.



ประกาศคุณูปการ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงได้ด้วยดีเนื่องจากความกรุณาและช่วยเหลือจากหลายๆ ท่าน เป็นอย่างดี ผู้วิจัยขอขอบพระคุณทุกท่านเป็นอย่างสูง

ผู้วิจัยขอขอบพระคุณ รองศาสตราจารย์ ดร.อนามัย นาอุดม ประธานที่ปรึกษาวิทยานิพนธ์ ผู้ช่วยศาสตราจารย์ ดร. จรัสศรี รุ่งรัตนอุบล กรรมการที่ปรึกษาวิทยานิพนธ์ เป็นอย่างสูงที่ได้ให้เวลาในการให้คำแนะนำ ให้ความรู้ ให้คำปรึกษา ช่วยเหลือตรวจทานและแก้ไขข้อบกพร่องตลอดระยะเวลาในการทำวิทยานิพนธ์ฉบับนี้ และขอขอบพระคุณ รองศาสตราจารย์ ดร. อัจฉมา อระวีพร ประธานกรรมการ สอบวิทยานิพนธ์ และกรรมการผู้ทรงคุณวุฒิภายใน ที่ได้กรุณาให้คำแนะนำแก้ไขในส่วนที่บกพร่องของงานวิทยานิพนธ์ด้วยความเอาใจใส่ จนทำให้วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงได้อย่างสมบูรณ์

ขอขอบพระคุณอาจารย์ภาควิชาคณิตศาสตร์ รวมทั้งเจ้าหน้าที่ทุกท่าน ที่กรุณาให้คำแนะนำ ให้คำปรึกษา และให้ความช่วยเหลือ ขอขอบพระคุณภาควิชาวิทยาการคอมพิวเตอร์และเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์ มหาวิทยาลัยนเรศวร ที่ให้ความช่วยเหลือในการจัดเตรียมสิ่งอำนวยความสะดวกสำหรับการดำเนินการทำวิจัย

ขอขอบพระคุณ พ่อ แม่ ญาติพี่น้อง ที่คอยเป็นกำลังใจ เป็นแรงผลักดัน และให้การสนับสนุนในทุกๆ ด้านอย่างดีที่สุดเสมอมา และขอขอบคุณเพื่อนร่วมรุ่นปริญญาโท และรุ่นพี่ปริญญาเอก และทุกท่านที่ไม่ได้เอ่ยนาม ที่ให้กำลังใจ ให้คำปรึกษา ให้คำแนะนำ และความช่วยเหลือที่ดีตลอดมา

ผู้วิจัยหวังเป็นอย่างยิ่งว่า งานวิจัยนี้จะเป็นประโยชน์ต่อประชาชน บุคลากรทางการแพทย์ และบุคคลที่มีความสนใจที่จะนำวิทยานิพนธ์ฉบับนี้ไปต่อยอดเพื่อพัฒนาองค์ความรู้ใหม่ๆ ต่อไป

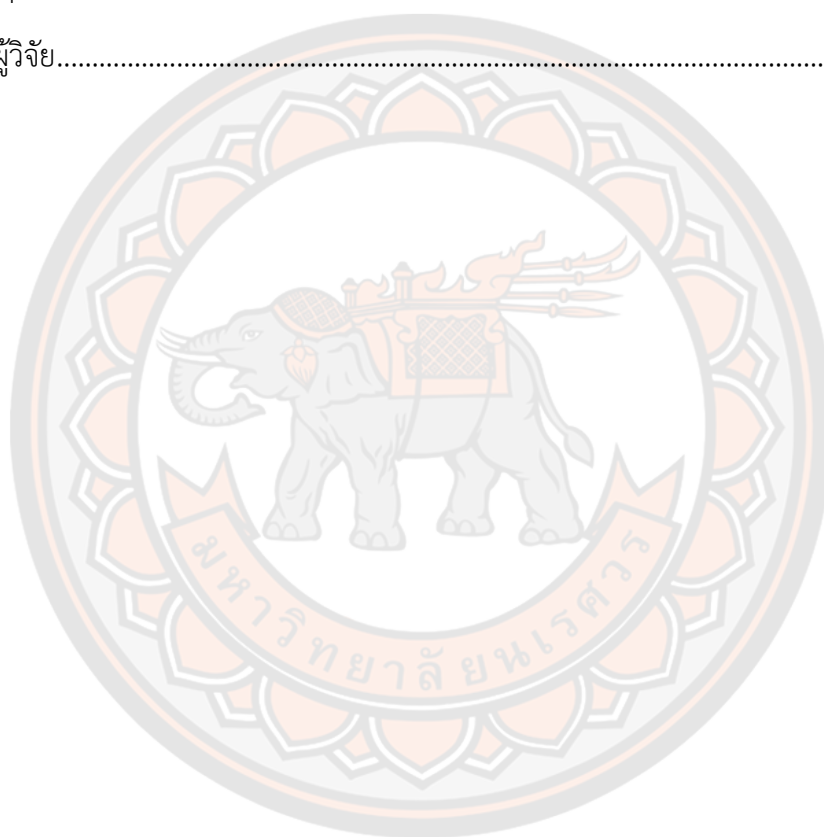
จิรนนท์ ป้องท้าว

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ค
บทคัดย่อภาษาอังกฤษ.....	จ
ประกาศคุณูปการ.....	ช
สารบัญ.....	ซ
สารบัญตาราง.....	ฎ
สารบัญภาพ.....	ฅ
บทที่ 1 บทนำ.....	1
1.1. ความเป็นมาและความสำคัญของปัญหา.....	1
1.2. จุดมุ่งหมายของการวิจัย.....	3
1.3. ขอบเขตการวิจัย.....	3
1.4. ประโยชน์ที่คาดว่าจะได้รับ.....	4
1.5. นิยามศัพท์เฉพาะ.....	4
บทที่ 2 เอกสารและงานวิจัยที่เกี่ยวข้อง.....	6
2.1. กระบวนการทำงานของเหมืองข้อมูล (Data Mining Process).....	6
2.2. เทคนิคการถดถอยลอจิสติก (Logistic Regression).....	9
2.2.1. ประเภทของการถดถอยลอจิสติก.....	9
2.2.2. วัตถุประสงค์ของการถดถอยลอจิสติก.....	9
2.2.3. ข้อตกลงเบื้องต้นของตัวแบบการถดถอยลอจิสติกแบบทวิ.....	9
2.2.4. ตัวแบบการถดถอยลอจิสติกแบบทวิ.....	10
2.3. เทคนิคต้นไม้ตัดสินใจ (Decision Tree : DT).....	12
2.4. เทคนิคป่าสุ่ม (Random Forest : RF).....	19

2.5. เทคนิคซ์พอร์ตเวกเตอร์แมชชีน (Support Vector Machine : SVM)	20
2.6. หลักการพัฒนาตัวแบบการจำแนกและเครื่องมือวัดประสิทธิภาพ	26
2.7. สัมประสิทธิ์สหสัมพันธ์	29
2.8. งานวิจัยที่เกี่ยวข้อง	30
บทที่ 3 วิธีการดำเนินการวิจัย	43
3.1. ข้อมูลที่ใช้ในการศึกษา	43
3.2. เครื่องมือที่ใช้ในงานวิจัย	44
3.3. ขั้นตอนการเตรียมข้อมูลสำหรับการวิเคราะห์	45
3.3.1. ข้อมูลชุดที่ 1 โรคหัวใจและหลอดเลือด	46
3.3.2. ข้อมูลชุดที่ 2 โรคเบาหวาน	48
3.3.3. ข้อมูลชุดที่ 3 การเลิกสูบบุหรี่	50
3.4. ขั้นตอนการพัฒนาตัวแบบ	53
3.4.1. พัฒนาตัวแบบการจำแนกโดยการใช้ตัวแปรอิสระทั้งหมด	53
3.4.2. พัฒนาตัวแบบการจำแนกโดยการคัดเลือกตัวแปรอิสระ	54
3.5. การประเมินประสิทธิภาพ	66
บทที่ 4 ผลการวิจัย	67
4.1. ผลการเปรียบเทียบประสิทธิภาพของตัวแบบการจำแนก โดยใช้ตัวแปรอิสระทั้งหมด	67
4.2. ผลการเปรียบเทียบประสิทธิภาพของตัวแบบการจำแนก โดยคัดเลือกตัวแปรอิสระ	70
4.2.1. ผลการเปรียบเทียบประสิทธิภาพของตัวแบบการจำแนก โดยคัดเลือกตัวแปรอิสระ ด้วยเทคนิคการถดถอยลอจิสติก (Logistic Regression)	70
4.2.2. ผลการเปรียบเทียบประสิทธิภาพของตัวแบบการจำแนกโดยคัดเลือกตัวแปรอิสระด้วย ค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient)	83
4.3. ผลการเปรียบเทียบประสิทธิภาพของตัวแบบการจำแนก โดยคัดเลือกตัวแปรอิสระเทียบ กับการใช้ตัวแปรอิสระทั้งหมด	85
4.4. ผลการเปรียบเทียบประสิทธิภาพของตัวแบบการจำแนกที่ดีที่สุด	89

บทที่ 5 สรุปลผลการวิจัย	90
5.1. อภิปรายผลวิจัย	91
5.2. ข้อเสนอแนะ	92
ภาคผนวก ก ผลการวิเคราะห์ข้อมูล.....	93
ภาคผนวก ข โปรแกรมอาร์.....	178
บรรณานุกรม.....	183
ประวัติผู้วิจัย.....	187



สารบัญตาราง

หน้า

ตาราง 1 แสดงตัวอย่างข้อมูลการผ่าคลอด	14
ตาราง 2 เมทริกซ์ความสับสน (Confusion Matrix)	28
ตาราง 3 สรุปงานวิจัยที่เกี่ยวข้องที่ประยุกต์ใช้วิธีการทางสถิติและเทคนิคเหมืองข้อมูลในการจำแนกข้อมูลด้านสุขภาพ.....	36
ตาราง 4 คุณลักษณะและจำนวนข้อมูลของแต่ละชุดข้อมูลด้านสุขภาพจากฐานข้อมูล Kaggle.....	44
ตาราง 5 คุณลักษณะและจำนวนข้อมูลที่ใช้ในการศึกษาหลังการตัดข้อมูลบางส่วนทิ้ง	44
ตาราง 6 รายละเอียดตัวแปรข้อมูลชุดที่ 1.....	46
ตาราง 7 รายละเอียดค่าสัมประสิทธิ์สหสัมพันธ์ของข้อมูลชุดที่ 1	46
ตาราง 8 ตัวอย่างข้อมูลหลังการแปลงค่าให้อยู่ในรูปปกติของข้อมูลชุดที่ 1	47
ตาราง 9 รายละเอียดตัวแปรข้อมูลชุดที่ 2.....	48
ตาราง 10 รายละเอียดค่าสัมประสิทธิ์สหสัมพันธ์ของข้อมูลชุดที่ 2.....	49
ตาราง 11 ตัวอย่างข้อมูลหลังการแปลงค่าให้อยู่ในรูปปกติของข้อมูลชุดที่ 2	50
ตาราง 12 รายละเอียดตัวแปรข้อมูลชุดที่ 3	50
ตาราง 13 รายละเอียดค่าสัมประสิทธิ์สหสัมพันธ์ของข้อมูลชุดที่ 3.....	51
ตาราง 14 ตัวอย่างข้อมูลหลังการแปลงค่าให้อยู่ในรูปปกติของข้อมูลชุดที่ 3.....	52
ตาราง 15 การแบ่งข้อมูลเป็น 10 ชุดของข้อมูลทั้ง 3 ชุด.....	53
ตาราง 16 การคัดเลือกตัวแปรอิสระในข้อมูลชุดที่ 1 ด้วยวิธีการทั้ง 3 รูปแบบ	57
ตาราง 17 การคัดเลือกตัวแปรอิสระในข้อมูลชุดที่ 2 ด้วยวิธีการทั้ง 3 รูปแบบ	59
ตาราง 18 การคัดเลือกตัวแปรอิสระในข้อมูลชุดที่ 3 ด้วยวิธีการทั้ง 3 รูปแบบ	61
ตาราง 19 ผลการจำแนกข้อมูลภายใต้ชุดข้อมูลที่ใช้ตัวแปรอิสระทั้งหมด ในชุดข้อมูลโรคหัวใจและหลอดเลือด.....	67
ตาราง 20 ผลการจำแนกข้อมูลภายใต้ชุดข้อมูลที่ใช้ตัวแปรอิสระทั้งหมด ในชุดข้อมูลโรคเบาหวาน.....	68

ตาราง 33 ผลการจำแนกข้อมูลภายใต้ชุดข้อมูลที่มีการคัดเลือกตัวแปรอิสระโดยใช้ข้อมูลเรียนรู้ทำซ้ำ 10 ครั้ง ด้วยเทคนิคการถดถอยลอจิสติก โดยวิธีการเลือกแบบไปข้างหน้า สำหรับชุดข้อมูลการเลิกสูบบุหรี่	77
ตาราง 34 ผลการจำแนกข้อมูลภายใต้ชุดข้อมูลที่มีการคัดเลือกตัวแปรอิสระโดยใช้ข้อมูลเรียนรู้ทำซ้ำ 10 ครั้ง ด้วยเทคนิคการถดถอยลอจิสติก โดยวิธีการกำจัดแบบถอยหลัง สำหรับชุดข้อมูลโรคหัวใจและหลอดเลือด.....	78
ตาราง 35 ผลการจำแนกข้อมูลภายใต้ชุดข้อมูลที่มีการคัดเลือกตัวแปรอิสระโดยใช้ข้อมูลเรียนรู้ทำซ้ำ 10 ครั้ง ด้วยเทคนิคการถดถอยลอจิสติก โดยวิธีการกำจัดแบบถอยหลัง สำหรับชุดข้อมูลโรคเบาหวาน	79
ตาราง 36 ผลการจำแนกข้อมูลภายใต้ชุดข้อมูลที่มีการคัดเลือกตัวแปรอิสระโดยใช้ข้อมูลเรียนรู้ทำซ้ำ 10 ครั้ง ด้วยเทคนิคการถดถอยลอจิสติก โดยวิธีการกำจัดแบบถอยหลัง สำหรับชุดข้อมูลการเลิกสูบบุหรี่	80
ตาราง 37 ผลการจำแนกข้อมูลภายใต้ชุดข้อมูลที่มีการคัดเลือกตัวแปรอิสระโดยใช้ข้อมูลเรียนรู้ทำซ้ำ 10 ครั้ง ด้วยเทคนิคการถดถอยลอจิสติก โดยวิธีการถดถอยทีละชั้น สำหรับชุดข้อมูลโรคหัวใจและหลอดเลือด.....	80
ตาราง 38 ผลการจำแนกข้อมูลภายใต้ชุดข้อมูลที่มีการคัดเลือกตัวแปรอิสระโดยใช้ข้อมูลเรียนรู้ทำซ้ำ 10 ครั้ง ด้วยเทคนิคการถดถอยลอจิสติก โดยวิธีการถดถอยทีละชั้น สำหรับชุดข้อมูลโรคเบาหวาน.	81
ตาราง 39 ผลการจำแนกข้อมูลภายใต้ชุดข้อมูลที่มีการคัดเลือกตัวแปรอิสระโดยใช้ข้อมูลเรียนรู้ทำซ้ำ 10 ครั้ง ด้วยเทคนิคการถดถอยลอจิสติก โดยวิธีการถดถอยทีละชั้น สำหรับชุดข้อมูลการเลิกสูบบุหรี่	82
ตาราง 40 ผลการจำแนกข้อมูลภายใต้ชุดข้อมูลที่มีการคัดเลือกตัวแปรอิสระด้วยค่าสัมประสิทธิ์สหสัมพันธ์ สำหรับชุดข้อมูลโรคหัวใจและหลอดเลือด.....	83
ตาราง 41 ผลการจำแนกข้อมูลภายใต้ชุดข้อมูลที่มีการคัดเลือกตัวแปรอิสระด้วยค่าสัมประสิทธิ์สหสัมพันธ์ สำหรับชุดข้อมูลโรคเบาหวาน.....	83
ตาราง 42 ผลการจำแนกข้อมูลภายใต้ชุดข้อมูลที่มีการคัดเลือกตัวแปรอิสระด้วยค่าสัมประสิทธิ์สหสัมพันธ์ สำหรับชุดข้อมูลการเลิกสูบบุหรี่	84
ตาราง 43 ผลการจำแนกด้วยเทคนิคการถดถอยลอจิสติก ในกรณีที่มีการคัดเลือกตัวแปรอิสระเทียบกับการใช้ตัวแปรอิสระทั้งหมด เมื่อวัดด้วยค่าความแม่นยำที่สมดุลแบบปรับน้ำหนักถ่วง.....	85

- ตาราง 44 ผลการจำแนกด้วยเทคนิคต้นไม้ตัดสินใจ ในกรณีที่มีการคัดเลือกตัวแปรอิสระเทียบกับการใช้ตัวแปรอิสระทั้งหมด เมื่อวัดด้วยค่าความแม่นยำที่สมดุลแบบปรับน้ำหนักถ่วง 86
- ตาราง 45 ผลการจำแนกด้วยเทคนิคป่าสุ่ม ในกรณีที่มีการคัดเลือกตัวแปรอิสระเทียบกับการใช้ตัวแปรอิสระทั้งหมด เมื่อวัดด้วยค่าความแม่นยำที่สมดุลแบบปรับน้ำหนักถ่วง 87
- ตาราง 46 ผลการจำแนกด้วยเทคนิคซัพพอร์ตเวกเตอร์แมชชีน ในกรณีที่มีการคัดเลือกตัวแปรอิสระเทียบกับการใช้ตัวแปรอิสระทั้งหมด เมื่อวัดด้วยค่าความแม่นยำที่สมดุลแบบปรับน้ำหนักถ่วง 88
- ตาราง 47 ผลการเปรียบเทียบประสิทธิภาพของตัวแบบการจำแนกที่ดีที่สุด..... 89



สารบัญภาพ

	หน้า
ภาพ 1 แบบจำลองกระบวนการวิเคราะห์ข้อมูลแบบ CRISP – DM	7
ภาพ 2 องค์ประกอบของต้นไม้ตัดสินใจ.....	12
ภาพ 3 เหตุการณ์การเลือกตัวแปรอิสระเป็นโหนดราก	18
ภาพ 4 แผนภาพต้นไม้ตัดสินใจของข้อมูลการผ่าคลอด.....	19
ภาพ 5 หลักการทำงานของเทคนิคป่าสุ่ม (Random Forest)	20
ภาพ 6 SVM แบบเชิงเส้นกรณี Hard Margin	21
ภาพ 7 SVM แบบเชิงเส้นกรณี Soft Margin	23
ภาพ 8 ตัวอย่างค่า C ที่มีผลต่อขนาดของระยะขอบ	24
ภาพ 9 แสดงวิธีการแบบ 10 – Fold Cross Validation	27
ภาพ 10 ผลงานกระบวนการสร้างตัวแบบการจำแนกโดยใช้ตัวแปรอิสระทั้งหมด	64
ภาพ 11 ผลงานกระบวนการสร้างตัวแบบการจำแนกโดยการคัดเลือกตัวแปรอิสระ ด้วยเทคนิคการ ถดถอยลอจิสติก โดยใช้ข้อมูลทั้งหมด	64
ภาพ 12 ผลงานกระบวนการสร้างตัวแบบการจำแนกโดยการคัดเลือกตัวแปรอิสระ ด้วยเทคนิคการ ถดถอยลอจิสติก โดยการสร้างตัวแบบ 10 ครั้ง ตามรูปแบบ 10-Fold Cross-Validation	65
ภาพ 13 ผลงานกระบวนการสร้างตัวแบบการจำแนกโดยการคัดเลือกตัวแปรอิสระ ด้วยค่า สัมประสิทธิ์สหสัมพันธ์สเปียร์แมน.....	65

บทที่ 1

บทนำ

1.1. ความเป็นมาและความสำคัญของปัญหา

ในปัจจุบันเทคโนโลยีได้เข้ามามีบทบาทมากขึ้น ซึ่งการเปลี่ยนแปลงด้านเทคโนโลยีส่งผลโดยตรงต่อองค์กรทั้งภาครัฐและเอกชนอย่างหลีกเลี่ยงไม่ได้ การใช้เทคโนโลยีมาจัดเก็บข้อมูลการทำธุรกรรมออนไลน์ และการใช้สื่อสังคมออนไลน์ รวมถึงอุปกรณ์รับส่งสัญญาณอัตโนมัติต่างๆ ทำให้ข้อมูลถูกสร้างขึ้นอย่างต่อเนื่องตลอดเวลา ทำให้ข้อมูลมีจำนวนมหาศาล และมีความหลากหลาย จนมีคำนิยาม ข้อมูลขนาดใหญ่ (Big Data) ขึ้นมาในปัจจุบัน โดยข้อมูลเหล่านี้ยังไม่ได้นำมาจัดเก็บอย่างมีระบบและหาความสัมพันธ์ของข้อมูล และมีการวิเคราะห์เพื่อนำมาใช้ประโยชน์ในด้านต่างๆ ซึ่งการเกิดขึ้นของชุดข้อมูลมหาศาลเหล่านี้ นำไปสู่การพัฒนาเทคโนโลยีสารสนเทศด้วยปัญญาประดิษฐ์ (Artificial Intelligence : AI) การทำเหมืองข้อมูล (Data Mining) ที่เป็นการทำงานเพื่อสกัดสารสนเทศที่สำคัญและนำไปใช้ให้เกิดประโยชน์เป็นส่วนหนึ่งในการทำงานปัญญาประดิษฐ์ตั้งแต่ปี ค.ศ. 1960 (Liao et al., 2012) ในทางด้านวงการแพทย์นั้นข้อมูลก็มีจำนวนมหาศาล และมีความหลากหลาย ไม่ว่าจะเป็น ประวัติผู้ป่วย บันทึกของแพทย์ ผลการทดสอบทางคลินิก ข้อมูลยา ได้มีการนำมาใช้รวบรวมและวิเคราะห์ข้อมูลสุขภาพในระดับมหภาค เพื่อแจ้งเตือนสถานการณ์โรคระบาด เพื่อจัดแพทย์ผู้เชี่ยวชาญที่เหมาะสม เพื่อคาดการณ์โรคระบาด และวางแผนป้องกันแก้ไขในอนาคต และยังช่วยให้ผู้ป่วยที่ไม่ทราบว่าตนป่วยเป็นโรคอะไรสามารถวิเคราะห์อาการของโรคในเบื้องต้นได้

การทำเหมืองข้อมูล (Data Mining) เป็นเทคนิคการวิเคราะห์ข้อมูล เพื่อค้นหารูปแบบความสัมพันธ์ของข้อมูลที่ซ่อนอยู่ในข้อมูลขนาดใหญ่ (Big Data) แล้วนำผลการวิเคราะห์ที่ได้มาใช้ประโยชน์ เทคนิคเหมืองข้อมูลมักจะถูกจัดให้เป็นส่วนหนึ่งของการเรียนรู้ของเครื่อง (Machine learning) โดยการเรียนรู้ของเครื่องแบ่งออกเป็น 3 ประเภทตามรูปแบบการเรียนรู้

1. การเรียนรู้แบบมีผู้สอน (Supervised learning) เป็นการทำให้เครื่องจักรสามารถหาคำตอบของปัญหาได้ด้วยตัวเอง หลังจากเรียนรู้จากชุดข้อมูลตัวอย่างไปแล้วระยะหนึ่ง ซึ่งการเรียนรู้แยกได้เป็นการถดถอย (Regression) และการจำแนกประเภท (Classification) ตามลักษณะของผลลัพธ์

2. การเรียนรู้แบบไม่มีผู้สอน (Unsupervised learning) เป็นการเรียนรู้ที่ให้เครื่องจักรนั้นสามารถเรียนรู้ด้วยตัวเอง โดยไม่ต้องมีค่าเป้าหมายของแต่ละข้อมูล ส่วนใหญ่ใช้เพื่อแบ่งกลุ่มการเรียนรู้

3. การเรียนรู้แบบเสริมกำลัง (Reinforcement Learning) เป็นการเรียนรู้ที่เกิดจากการปฏิสัมพันธ์ (Interaction) ระหว่างผู้เรียนรู้ (Agent) กับสิ่งแวดล้อม (Environment) โดยกำหนดเงื่อนไขบางอย่างให้กับคอมพิวเตอร์ ทำให้คอมพิวเตอร์ทำตามเงื่อนไขนั้นให้ได้ ผ่านการลองผิดลองถูกภายใต้สถานการณ์นั้น

ตัวแบบเพื่อการจำแนกประเภทข้อมูล (Data Classification) เป็นการเรียนรู้แบบมีผู้สอนและเป็นการทำเหมืองข้อมูลที่ได้รับคามนิยมในการนำไปใช้เพื่อจำแนกข้อมูลที่มีตัวแปรตามเป็นตัวแปรเชิงคุณภาพ การทำเหมืองข้อมูลเป็นวิธีการที่ช่วยในการวิเคราะห์ที่ดี แต่จำเป็นต้องเลือกเทคนิคที่จะนำมาใช้ในการวิเคราะห์ให้เหมาะสมกับประเภทของข้อมูล (Han et al., 2011) ปัจจุบันเหมืองข้อมูลและการสร้างตัวแบบเพื่อการจำแนกถูกนำไปใช้ในงานด้านต่างๆ เช่น ด้านธุรกิจ ด้านอุตสาหกรรม ด้านเศรษฐศาสตร์ ด้านการศึกษา ด้านวิศวกรรมศาสตร์ และด้านการแพทย์ เป็นต้น

งานวิจัยในด้านการแพทย์มีผู้วิจัยหลายท่านได้ทำการศึกษาและพัฒนาตัวแบบการจำแนกบนข้อมูลทางด้านสุขภาพ เช่น ข้อมูลโรคเบาหวาน (รุ่งโรจน์ บุญมา และนิเวศ จิระวิชิตชัย ,2562)(รัชนิวรรณ ไพศาลวรเกียรติ, 2564)(ปพนน์ศรณ์ สิวส์ำแดงเดช, 2565)(Astuti et al., 2018) ข้อมูลโรคมะเร็งเต้านม (Ali et al., 2020)(Aroef et al., 2020)(อุกฤษฏ์ ศรีสุข และจारी ทองคำ, 2563) ข้อมูลโรคหัวใจและหลอดเลือด (Assari et al., 2017)(Nai-arun & Moungrmai, 2020) ข้อมูลโรคไตเรื้อรัง (สุรวัชร ศรีเปารยะ และสายชล สิ้นสมบุรณ์ทอง, 2560) และข้อมูลโรคไฮเปอร์ไทรอยด์ (อุกฤษฏ์ ศรีสุข และจारी ทองคำ, 2563) เป็นต้น จากการพัฒนาตัวแบบบนข้อมูลผู้ป่วยโรคเบาหวานพบว่าเทคนิคซัพพอร์ตเวกเตอร์แมชชีนมีประสิทธิภาพที่ดีที่สุด (รุ่งโรจน์ บุญมา และนิเวศ จิระวิชิตชัย ,2562) ในการเปรียบเทียบตัวแบบการถดถอยลอจิสติกและเทคนิคเหมืองข้อมูลสำหรับการจำแนกการเป็นโรคเบาหวานและคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติก พบว่าเทคนิคโครงข่ายประสาทเทียมที่คัดเลือกตัวแปรอิสระด้วยวิธีการเลือกแบบไปข้างหน้าและวิธีการถดถอยทีละขั้นมีประสิทธิภาพสูงที่สุด (รัชนิวรรณ ไพศาลวรเกียรติ, 2564) การศึกษาและพัฒนาตัวแบบการจำแนกโอกาสการเป็นโรคเบาหวาน พบว่าเทคนิคป่าสุ่มให้ค่าความแม่นยำสูงที่สุด (ปพนน์ศรณ์ สิวส์ำแดงเดช ,2565) จากการศึกษาเปรียบเทียบการจำแนกผู้ป่วยโรคเบาหวานด้วยเทคนิคการถดถอยลอจิสติกและเทคนิคซัพพอร์ตเวกเตอร์แมชชีน พบว่าเทคนิคการถดถอยลอจิสติกโดยคัดเลือกตัวแปรอิสระด้วยวิธีการเลือกแบบไปข้างหน้ามีประสิทธิภาพสูงที่สุด (Astuti et al., 2018) ในการเปรียบเทียบซัพพอร์ตเวกเตอร์แมชชีนและเทคนิคการถดถอยลอจิสติก พบว่าเทคนิคการถดถอยลอจิสติกมีค่าความแม่นยำสูงที่สุด (Ali et al., 2020) การพัฒนาตัวแบบการจำแนกของข้อมูลโรคมะเร็งเต้านมพบว่าเทคนิคซัพพอร์ตเวกเตอร์แมชชีนมีค่าความแม่นยำสูงที่สุด (Aroef et al., 2020) การพัฒนาตัวแบบการจำแนกโอกาสการเป็นโรคมะเร็งเต้านมและการเป็นโรคไฮเปอร์ไทรอยด์ พบว่าเทคนิคต้นไม้ตัดสินใจมีค่าความแม่นยำมากที่สุด (อุกฤษฏ์ ศรีสุข และจारी ทองคำ, 2563) นอกจากนี้การพัฒนาตัว

จำแนกการเกิดโรคหัวใจพบว่าเทคนิคซัพพอร์ตเวกเตอร์แมชชีนมีความแม่นยำสูงสุด (Assari et al., 2017) การพัฒนาตัวแบบการจำแนกความเสี่ยงของการเกิดโรคหัวใจและหลอดเลือด พบว่าเทคนิคการถดถอยลอจิสติกมีค่าความแม่นยำสูงสุด (Nai-arun & Moungrmai, 2020) การพัฒนาตัวแบบการจำแนกของข้อมูลโรคไตเรื้อรังพบว่า เทคนิคต้นไม้ตัดสินใจมีค่าความแม่นยำสูงสุด (สุรวุฒ ศรีเปารยะ และสายชล สนิสมบุญทอง, 2560)

จากการศึกษาที่กล่าวมาทำให้ผู้วิจัยต้องการศึกษาการพัฒนาตัวแบบการจำแนกด้วยเทคนิคแบบต่างๆ บนข้อมูลด้านสุขภาพ โดยในที่นี้ได้เลือกเทคนิคการถดถอยลอจิสติก เทคนิคต้นไม้ตัดสินใจ เทคนิคป่าสุ่ม และเทคนิคซัพพอร์ตเวกเตอร์แมชชีน ซึ่งเป็นเทคนิคทางเหมืองข้อมูลที่มีประสิทธิภาพบนชุดข้อมูลทางด้านสุขภาพ และศึกษาพัฒนาตัวแบบการจำแนกด้วยการคัดเลือกตัวแปรอิสระโดยใช้เทคนิคการถดถอยลอจิสติก และค่าสัมประสิทธิ์สหสัมพันธ์ พัฒนาตัวแบบการจำแนกบนชุดข้อมูลด้านสุขภาพที่มีคุณลักษณะของตัวแปรอิสระแตกต่างกันทั้งหมด 3 ชุดข้อมูล ได้แก่ ชุดข้อมูลที่มีตัวแปรอิสระเชิงปริมาณใกล้เคียงกับเชิงคุณภาพ ชุดข้อมูลที่มีตัวแปรอิสระเชิงปริมาณน้อยกว่าเชิงคุณภาพ และชุดข้อมูลที่มีตัวแปรเชิงปริมาณมากกว่าเชิงคุณภาพ โดยพัฒนาตัวแบบและทดสอบประสิทธิภาพด้วยวิธีตรวจสอบไขว้ (k -Fold Cross-Validation) และเปรียบเทียบประสิทธิภาพของตัวแบบการจำแนกด้วยเกณฑ์ต่อไปนี้ ได้แก่ ค่าความไว (Sensitivity) ค่าความจำเพาะ (Specificity) ค่าความแม่นยำ (Accuracy) ค่าความแม่นยำที่สมดุลแบบปรับน้ำหนักถ่วง (Weighted Balanced Accuracy) และค่าประสิทธิภาพโดยรวมแบบปรับน้ำหนักถ่วง ($F1_{weight}$)

1.2. จุดมุ่งหมายของการวิจัย

1. เพื่อศึกษาการทำงานของเทคนิคการถดถอยลอจิสติก (Logistic Regression) เทคนิคต้นไม้ตัดสินใจ (Decision Tree) เทคนิคป่าสุ่ม (Random Forest) และเทคนิคซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine)

2. เพื่อพัฒนาตัวแบบและเปรียบเทียบประสิทธิภาพการจำแนกข้อมูลด้านสุขภาพที่มีคุณลักษณะของตัวแปรอิสระที่แตกต่างกัน โดยใช้เทคนิคการถดถอยลอจิสติก เทคนิคต้นไม้ตัดสินใจ เทคนิคป่าสุ่ม และเทคนิคซัพพอร์ตเวกเตอร์แมชชีน

3. เพื่อศึกษาอิทธิพลของการคัดเลือกตัวแปรอิสระโดยใช้เทคนิคการถดถอยลอจิสติก และค่าสัมประสิทธิ์สหสัมพันธ์ที่มีต่อประสิทธิภาพการจำแนก

1.3. ขอบเขตการวิจัย

1. ข้อมูลที่ใช้ในการศึกษาครั้งนี้เป็นข้อมูลทางด้านสุขภาพจำนวน 3 ชุด ได้แก่ ชุดที่ 1 ข้อมูลโรคหัวใจและหลอดเลือดจำนวน 70,000 คน มีจำนวนตัวแปรอิสระ 11 ตัวแปร แบ่งออกเป็นตัวแปร

เชิงปริมาณ 5 ตัวแปร และตัวแปรเชิงคุณภาพ 6 ตัวแปร ชุดที่ 2 ข้อมูลโรคเบาหวานจำนวน 70,692 คน มีตัวแปรอิสระจำนวน 21 ตัวแปร เป็นตัวแปรเชิงปริมาณ 3 ตัวแปร และตัวแปรเชิงคุณภาพ 18 ตัวแปร และชุดที่ 3 ข้อมูลการเลิกสูบบุหรี่จำนวน 38,984 คน มีตัวแปรอิสระจำนวน 22 ตัวแปร แบ่งเป็นตัวแปรเชิงปริมาณ 19 ตัวแปร และตัวแปรเชิงคุณภาพ 3 ตัวแปร จากฐานข้อมูล Kaggle

2. พัฒนาตัวแบบการจำแนกด้วยเทคนิคการถดถอยลอจิสติก เทคนิคต้นไม้ตัดสินใจ เทคนิคป่าสุ่ม และเทคนิคซัพพอร์ตเวกเตอร์แมชชีน โดยแบ่งการทดสอบแบบ 10 – Fold Cross Validation และวัดประสิทธิภาพด้วยค่าความไว (Sensitivity) ค่าความจำเพาะ (Specificity) ค่าความแม่นยำ (Accuracy) ค่าความแม่นยำที่สมดุลแบบปรับน้ำหนักถ่วง (Weighted Balanced Accuracy) และค่าประสิทธิภาพโดยรวมแบบปรับน้ำหนักถ่วง ($F1_{weight}$)

3. ชุดข้อมูลที่ศึกษามีจำนวนตัวแปรอิสระเชิงปริมาณ และตัวแปรอิสระเชิงคุณภาพแตกต่างกันประกอบไปด้วย ชุดข้อมูลที่มีจำนวนตัวแปรอิสระเชิงปริมาณใกล้เคียงตัวแปรอิสระเชิงคุณภาพ (ชุดข้อมูลโรคหัวใจและหลอดเลือด) ชุดข้อมูลที่มีจำนวนตัวแปรอิสระเชิงปริมาณน้อยกว่าตัวแปรอิสระเชิงคุณภาพ (ชุดข้อมูลโรคเบาหวาน) และชุดข้อมูลที่มีจำนวนตัวแปรอิสระเชิงปริมาณมากกว่าตัวแปรอิสระเชิงคุณภาพ (ชุดข้อมูลการเลิกสูบบุหรี่)

1.4. ประโยชน์ที่คาดว่าจะได้รับ

1. ทราบกระบวนการทำงานของเทคนิคการถดถอยลอจิสติก เทคนิคต้นไม้ตัดสินใจ เทคนิคป่าสุ่ม และเทคนิคซัพพอร์ตเวกเตอร์แมชชีน
2. ได้ตัวแบบการจำแนกและทราบประสิทธิภาพการจำแนกข้อมูลของแต่ละเทคนิค ในกรณีที่ข้อมูลมีคุณลักษณะของตัวแปรอิสระแตกต่างกัน
3. ได้ตัวแบบการจำแนกและทราบประสิทธิภาพการจำแนกข้อมูลของแต่ละเทคนิค ในกรณีที่มีการคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติก และค่าสัมประสิทธิ์สหสัมพันธ์

1.5. นิยามศัพท์เฉพาะ

TP	แทน	ผลบวกจริง
TN	แทน	ผลลบจริง
FP	แทน	ผลบวกปลอม
FN	แทน	ผลลบปลอม
Pre.	แทน	ค่าความเที่ยง (Precision)
Sen.	แทน	ค่าความไว (Sensitivity)
Spe.	แทน	ค่าความจำเพาะ (Specificity)

Acc.	แทน	ค่าความแม่นยำ (Accuracy)
BAcc.	แทน	ค่าความแม่นยำที่สมดุลแบบปรับน้ำหนักถ่วง (Weighted Balanced Accuracy)
F1 _{weight}	แทน	ค่าประสิทธิภาพโดยรวมแบบปรับน้ำหนักถ่วง
LR	แทน	เทคนิคการถดถอยลอจิสติก
DT	แทน	เทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม C4.5
RF	แทน	เทคนิคป่าสุ่ม
SVM	แทน	เทคนิคซัพพอร์ตเวกเตอร์แมชชีน
r _s	แทน	ค่าสัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมน
A _f	แทน	เซตของตัวแปรอิสระที่ใช้วิธีการเลือกแบบไปข้างหน้า
A _b	แทน	เซตของตัวแปรอิสระที่ใช้วิธีการกำจัดแบบถอยหลัง
A _s	แทน	เซตของตัวแปรอิสระที่ใช้วิธีการถดถอยทีละชั้น
Full	แทน	กรณีใช้ตัวแปรอิสระทั้งหมด
LR _{ข้อมูลทั้งหมด}	แทน	การคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติก โดยใช้ข้อมูลทั้งหมด
LR _{ข้อมูลเรียนรู้}	แทน	การคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติก โดยใช้ข้อมูลเรียนรู้ทำซ้ำ 10 ครั้ง
COR	แทน	การคัดเลือกตัวแปรอิสระด้วยค่าสัมประสิทธิ์สหสัมพันธ์
f	แทน	วิธีการเลือกแบบไปข้างหน้า
s	แทน	วิธีการกำจัดแบบถอยหลัง
b	แทน	วิธีการถดถอยทีละชั้น

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

การวิจัยครั้งนี้ มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของตัวแบบการถดถอยลอจิสติก กับเทคนิคการทำเหมืองข้อมูล เพื่อจำแนกข้อมูลด้านสุขภาพที่มีคุณลักษณะของตัวแปรอิสระแตกต่างกัน โดยผู้วิจัยได้ศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องเพื่อเป็นแนวทางในการศึกษา ดังรายละเอียดต่อไปนี้

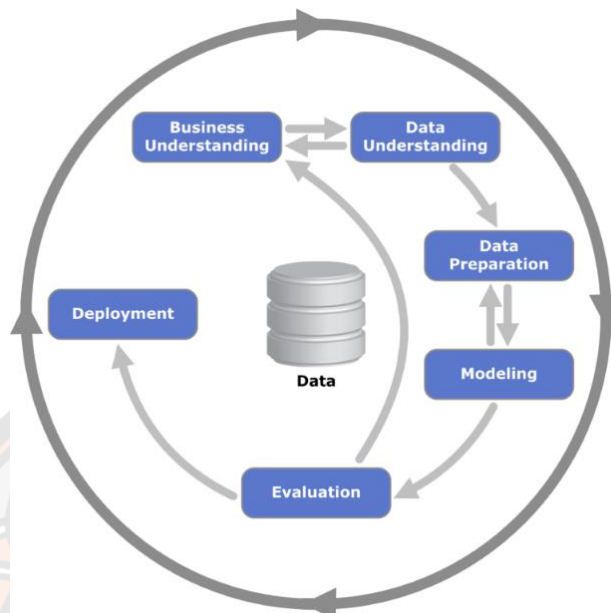
1. กระบวนการทำงานของเหมืองข้อมูล
2. เทคนิคการถดถอยลอจิสติก
3. เทคนิคต้นไม้ตัดสินใจ
4. เทคนิคป่าสุ่ม
5. เทคนิคซัพพอร์ตเวกเตอร์แมชชีน
6. หลักการพัฒนาตัวแบบการจำแนกและเครื่องมือวัดประสิทธิภาพ
7. สัมประสิทธิ์สหสัมพันธ์
8. งานวิจัยที่เกี่ยวข้อง

2.1. กระบวนการทำงานของเหมืองข้อมูล (Data Mining Process)

การทำเหมืองข้อมูล (Data Mining) เป็นกระบวนการค้นหาข้อมูลเชิงลึก เพื่อค้นหารูปแบบความสัมพันธ์ของข้อมูลที่ซ่อนอยู่ในชุดข้อมูลขนาดใหญ่ โดยอาศัยหลักสถิติ การรู้จำ การเรียนรู้ของเครื่อง และหลักคณิตศาสตร์ เพื่อให้ได้องค์ความรู้ใหม่ที่สามารถนำไปประกอบการตัดสินใจในด้านต่างๆ เช่น ด้านการวางแผนบริหารจัดการธุรกิจ ด้านเศรษฐกิจและสังคม ด้านการเงิน หรือด้านการแพทย์ เป็นต้น โดยด้านการแพทย์ การทำเหมืองข้อมูลมีบทบาทที่หลากหลาย ทั้งในการวินิจฉัยโรค การพยากรณ์โรค การรักษาโรค และการศึกษากลไกของโรค การทำเหมืองข้อมูลเปรียบเสมือนวิวัฒนาการหนึ่งในการจัดเก็บและตีความหมายข้อมูล จากเดิมที่มีการจัดเก็บข้อมูลอย่างง่าย มาสู่การจัดเก็บในรูปแบบข้อมูลที่สามารถดึงข้อมูลสารสนเทศมาใช้ได้ จนถึงการทำเหมืองข้อมูลที่สามารถค้นพบความรู้ที่ซ่อนอยู่ในชุดข้อมูลนั้น

ขั้นตอนการทำเหมืองข้อมูล (Process of Data Mining) เป็นกระบวนการในการค้นหาลักษณะแฝงของข้อมูล (Pattern) ที่ซ่อนอยู่ในฐานข้อมูล โดยขั้นตอนการวิเคราะห์ข้อมูลจะใช้กระบวนการวิเคราะห์ข้อมูลแบบ CRISP – DM (Cross – Industry Standard Process for Data

Mining) ซึ่งเป็นกระบวนการมาตรฐานในการวิเคราะห์ข้อมูลด้านเทคนิคเหมืองข้อมูล มีขั้นตอนทั้งหมด 6 ขั้นตอน ดังแสดงในภาพ 1



ภาพ 1 แบบจำลองกระบวนการวิเคราะห์ข้อมูลแบบ CRISP – DM

ที่มา : เอกสิทธิ์ พัชรวงศ์ศักดิ์ (2563, หน้า 125)

กระบวนการวิเคราะห์ข้อมูลแบบ CRISP – DM มี 6 ขั้นตอนดังนี้

1. ทำความเข้าใจปัญหา (Business Understanding) เป็นขั้นตอนแรกในกระบวนการ CRISP – DM ซึ่งเน้นไปที่การทำความเข้าใจปัญหา และแปลงปัญหาที่ได้ให้อยู่ในรูปโจทย์ของการวิเคราะห์ข้อมูลทางการทำเหมืองข้อมูล (Data Mining) ระบุเป้าหมายที่ต้องการได้จากการวิเคราะห์ข้อมูลด้วยการทำเหมืองข้อมูล และวางแผนดำเนินการ

2. ทำความเข้าใจข้อมูล (Data Understanding) ในขั้นตอนนี้เป็นการวิเคราะห์ว่ามีข้อมูลอะไรบ้างที่ช่วยในการตอบโจทย์ในขั้นตอนแรก ซึ่งขั้นตอนนี้เป็นการรวบรวมข้อมูลที่เกี่ยวข้องข้อมูล ถูกต้องและน่าเชื่อถือ เป็นการทำความเข้าใจว่าข้อมูลแต่ละตัวแปรมีเงื่อนไขในการจัดเก็บหรือข้อยกเว้นอะไร และข้อมูลที่ได้มีความเหมาะสม มีรายละเอียดเพียงพอสำหรับการนำไปวิเคราะห์หรือไม่

3. เตรียมข้อมูล (Data Preparation) เป็นขั้นตอนที่ใช้เวลามากที่สุด เนื่องจากตัวแบบที่ได้จากการทำเหมืองข้อมูลจะให้ผลลัพธ์ที่ถูกต้องหรือไม่ ขึ้นอยู่กับคุณภาพของข้อมูลที่ใช้ โดยขั้นตอนนี้จะแบ่งออกเป็น 3 ขั้นตอนได้แก่

3.1. การคัดเลือกข้อมูล (Data Selection) เป็นขั้นตอนของการเลือกใช้เฉพาะข้อมูลที่เกี่ยวข้องกับสิ่งที่ต้องการวิเคราะห์ และเลือกช่วงเวลาที่วิเคราะห์ข้อมูล

3.2. การกลั่นกรองข้อมูล (Data Cleansing) เป็นขั้นตอนการหาข้อมูลที่ผิดปกติ เช่น ข้อมูลมีค่าที่สูญหายไป ข้อมูลมีความซ้ำซ้อน หรือข้อมูลที่มีลักษณะผิดปกติมากกว่าข้อมูลอื่นมากๆ แล้วทำการแก้ไขข้อมูลที่ผิดปกติถ้ามี

3.3. การแปลงข้อมูล (Data Transformation) เป็นการแปลงข้อมูลให้เหมาะสม สำหรับการวิเคราะห์ข้อมูลตามเทคนิคเหมืองข้อมูลที่ใช้ โดยปกติแล้วข้อมูลจะไม่ได้อยู่ในรูปแบบที่นำไปวิเคราะห์ได้ จึงจำเป็นต้องมีการแปลงข้อมูลให้อยู่ในรูปแบบที่สามารถนำมาวิเคราะห์ได้ง่ายขึ้น

4. การสร้างแบบจำลอง (Modeling) เป็นขั้นตอนการวิเคราะห์ข้อมูลด้วยเทคนิคเหมืองข้อมูล ซึ่งแบ่งได้เป็น

4.1. การหาความสัมพันธ์ (Association Rules) เป็นการหาความสัมพันธ์ของข้อมูลที่เกิดขึ้นร่วมกัน ในการหาความสัมพันธ์นี้มีเทคนิคหาความสัมพันธ์หลายเทคนิค แต่เทคนิคที่นิยมใช้อย่างแพร่หลายคือ เทคนิค Apriori

4.2. การแบ่งกลุ่มข้อมูล (Clustering) เป็นการจัดกลุ่มข้อมูลที่มีลักษณะเหมือนกันหรือคล้ายกันเข้าไว้ในกลุ่มเดียวกันอัตโนมัติ ซึ่งเป็นการวิเคราะห์โดยไม่พิจารณาจัดกลุ่มตามประเภทที่มีหรือรู้จัก แต่จะใช้ขั้นตอนวิธีการจัดกลุ่มเพื่อหากลุ่มที่สามารถยอมรับได้เพื่อจัดเข้ากลุ่ม ซึ่งเทคนิคที่นิยมใช้อย่างแพร่หลาย คือ การแบ่งกลุ่มแบบเคมีน (K-Means Clustering)

4.3. การจำแนกประเภทข้อมูล (Classification) เป็นการวิเคราะห์แบบพยากรณ์ (Predictive Analytics) โดยมีการจดจำหรือเรียนรู้จากรูปแบบของข้อมูลในอดีต และนำมาสร้างแบบจำลองเพื่อใช้ในการจำแนกประเภทข้อมูล โดยพยากรณ์ค่าให้กับข้อมูลใหม่ที่พบ ซึ่งตัวแบบจำลองอาจจะอยู่ในรูปของกฎหรือเงื่อนไขต่างๆ หรืออยู่ในรูปของสมการทางคณิตศาสตร์

5. การประเมินผล (Evaluation) เป็นขั้นตอนการวัดผลลัพธ์ ซึ่งวิธีการวัดผลอาจใช้การเปรียบเทียบจากผลลัพธ์ที่ได้ของแต่ละเทคนิคที่ใช้งานเพื่อเลือกตัวแบบจำลองที่ให้ความถูกต้องดีกว่า และนำไปพิจารณาว่าตอบคำถามในขั้นตอนแรกหรือไม่ ถ้ายังไม่ได้สามารถย้อนกลับไปทำซ้ำในขั้นตอนก่อนหน้าได้เรื่อย ๆ แต่ถ้าผลลัพธ์สามารถตอบคำถามขั้นตอนแรกได้ก็ดำเนินการต่อในขั้นตอนถัดไป

6. การนำแบบจำลองไปใช้ (Deployment) เป็นขั้นตอนของการนำผลลัพธ์หรือแบบจำลองที่ได้จากการวิเคราะห์ข้อมูลด้วยเทคนิคเหมืองข้อมูลไปใช้งานจริงในองค์กรหรือบริษัท การใช้งานขึ้นอยู่กับว่าจะนำข้อมูลเชิงลึกที่ได้ไปใช้งานในรูปแบบใด

2.2. เทคนิคการถดถอยลอจิสติก (Logistic Regression)

การถดถอยลอจิสติกเป็นวิธีการทางสถิติที่มีวัตถุประสงค์เพื่อประมาณค่าหรือพยากรณ์ความน่าจะเป็นของการเกิดเหตุการณ์ที่สนใจหรือการไม่เกิดเหตุการณ์ที่สนใจภายใต้อิทธิพลของตัวปัจจัยหรือตัวแปรอิสระ โดยที่ตัวแปรตามเป็นตัวแปรเชิงคุณภาพ (กัลยา วานิชย์บัญชา, 2552)

2.2.1. ประเภทของการถดถอยลอจิสติก

1. แบ่งตามจำนวนตัวแปรตาม แบ่งออกเป็น 2 ประเภทย่อยดังนี้
 - การถดถอยลอจิสติกแบบทวิ (Binary Logistic Regression) คือตัวแปรตามมีได้ 2 ค่า
 - การถดถอยลอจิสติกแบบพหุกลุ่ม (Multinomial Logistic Regression) คือตัวแปรตามมีมากกว่า 2 ค่าขึ้นไป
2. แบ่งตามจำนวนตัวแปรอิสระ โดยที่ตัวแปรอิสระเป็นได้ทั้งตัวแปรเชิงปริมาณและตัวแปรเชิงคุณภาพ แบ่งออกเป็น 2 ประเภทดังนี้
 - การถดถอยลอจิสติกอย่างง่าย (Simple Logistic Regression) คือตัวแปรอิสระมีเพียง 1 ตัวแปร
 - การถดถอยลอจิสติกเชิงพหุ (Multiple Logistic Regression) คือตัวแปรอิสระตั้งแต่ 2 ตัวขึ้นไป

ในงานวิจัยนี้จะศึกษาเฉพาะการถดถอยลอจิสติกแบบทวิ (Binary Logistic Regression) และตัวแปรอิสระตั้งแต่ 2 ตัวขึ้นไป (Multiple Logistic Regression)

2.2.2. วัตถุประสงค์ของการถดถอยลอจิสติก

1. เพื่อศึกษาความสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระที่สามารถอธิบายโอกาสการเกิดเหตุการณ์ที่สนใจหรือไม่เกิดเหตุการณ์ที่สนใจ พร้อมทั้งศึกษาระดับความสัมพันธ์ระหว่างตัวแปรอิสระแต่ละตัวกับตัวแปรตาม
2. เพื่อพยากรณ์โอกาสที่จะเกิดเหตุการณ์ที่สนใจ จากตัวแบบการถดถอยลอจิสติกแบบทวิ ที่ % ความถูกต้องในการพยากรณ์สูงที่สุด

2.2.3. ข้อตกลงเบื้องต้นของตัวแบบการถดถอยลอจิสติกแบบทวิ

1. ตัวแปรตามเป็นตัวแปรเชิงคุณภาพจะกำหนด 2 ค่า คือ 0 และ 1 เท่านั้น
2. ค่าคาดหวังของค่าคลาดเคลื่อนเป็นศูนย์ นั่นคือ $E(e) = 0$
3. e_i และ e_j เป็นอิสระกัน
4. e_i และ X_i เป็นอิสระกัน
5. ตัวแปรอิสระไม่มีความสัมพันธ์กันหรือไม่เกิดปัญหาความสัมพันธ์เชิงเส้นพหุ (Multicollinearity)

2.2.4. ตัวแบบการถดถอยลอจิสติกแบบทวิ

การถดถอยลอจิสติกแบบทวิ คือตัวแปรตามจะเป็นตัวแปรทวิ (Binary Variable) ที่มีเพียง 2 ค่า คือ 0 และ 1 เมื่อให้ Y_i แทนตัวแปรตาม และ $P(Y_i)$ คือความน่าจะเป็นที่จะเกิดเหตุการณ์ จะได้ว่า

○ ความน่าจะเป็นที่จะไม่เกิดเหตุการณ์ที่สนใจ Y_i เป็น 0 เท่ากับ $1 - \pi_i$

หรือ $P(Y_i = 0) = 1 - \pi_i$

○ ความน่าจะเป็นที่จะเกิดเหตุการณ์ที่สนใจ Y_i เป็น 1 เท่ากับ π_i หรือ

$P(Y_i = 1) = \pi_i$ ซึ่งค่าคาดหวังของ Y หรือ $E(Y_i) = \pi_i$ (ทรงศิริ แต่สมบัติ, 2548, หน้า 238)

ตัวแบบการถดถอยลอจิสติกเชิงพหุมีรูปแบบดังนี้ (สายชล สินสมบูรณ์ทอง, 2553, หน้า 445)

$$E(Y_i) = \pi_i = \frac{\exp(\beta_0 + \beta_1 X_{i1} + \dots + \beta_k X_{ik})}{1 + \exp(\beta_0 + \beta_1 X_{i1} + \dots + \beta_k X_{ik})} ; i = 1, 2, \dots, n$$

$$= \frac{\exp(X_i' \beta)}{1 + \exp(X_i' \beta)} \quad (2.1)$$

เมื่อ k แทน จำนวนตัวแปรอิสระ

n แทน จำนวนข้อมูล

X_{ij} แทน ตัวแปรอิสระ เมื่อ $j = 0, \dots, k$ และ $X_{i0} = 1$

β_j แทน สัมประสิทธิ์ถดถอย เมื่อ $j = 0, \dots, k$

ในรูปแบบของอัตราส่วนออกดส์ (Odds Ratio) นั่นคืออัตราส่วนระหว่างความน่าจะเป็นที่จะเกิดเหตุการณ์ที่สนใจต่อความน่าจะเป็นของการไม่เกิดเหตุการณ์ที่สนใจ จากสมการ (2.1) สามารถเขียนอัตราส่วนออกดส์ได้ดังนี้

$$Odds = \frac{P(Y_i = 1)}{P(Y_i = 0)} = \frac{\pi_i}{1 - \pi_i} \quad (2.2)$$

เนื่องจากความสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระไม่อยู่ในรูปแบบเชิงเส้น ในการวิเคราะห์การถดถอยลอจิสติกจึงต้องมีการปรับให้ความสัมพันธ์อยู่ในรูปแบบเชิงเส้น นั่นคือจะเขียนให้อยู่ในรูป log ของ Odds เรียกว่า ฟังก์ชันตอบสนองลอจิท (Logit Response function) สามารถเขียนได้ดังนี้

$$\log_e(odds) = \log_e \left[\frac{\pi_i}{1 - \pi_i} \right] = X_i' \beta \quad (2.3)$$

เป้าหมายของการวิเคราะห์การถดถอยลอจิสติก คือ การประมาณค่าสัมประสิทธิ์การถดถอยลอจิสติก (β) เนื่องจากค่าสังเกต Y_i มีการแจกแจงแบบเบอร์นูลลี สามารถเขียนฟังก์ชันความน่าจะเป็นดังแสดงในสมการ (2.4)

$$f_i(y_i) = (\pi_i)^{y_i}(1 - \pi_i)^{1-y_i} ; i = 1, 2, \dots, n \quad (2.4)$$

เนื่องจากค่าสังเกต Y_i แต่ละค่าเป็นอิสระกัน ดังนั้นฟังก์ชันความน่าจะเป็น (Likelihood Function) คือ

$$\begin{aligned} L(y_1, \dots, y_n; \beta) &= \prod_{i=1}^n f_i(y_i) \\ &= \prod_{i=1}^n (\pi_i)^{y_i}(1 - \pi_i)^{1-y_i} \end{aligned} \quad (2.5)$$

หาค่าประมาณความน่าจะเป็นสูงสุดเพื่อประมาณพารามิเตอร์ โดยการหาล็อกการิทึมของฟังก์ชันความน่าจะเป็น ดังนี้

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

ดังนั้น $\log_e L(y_1, \dots, y_n; \beta) = \sum_{i=1}^n y_i (X_i' \beta) - \sum_{i=1}^n \log_e (1 + \exp(X_i' \beta)) \quad (2.6)$

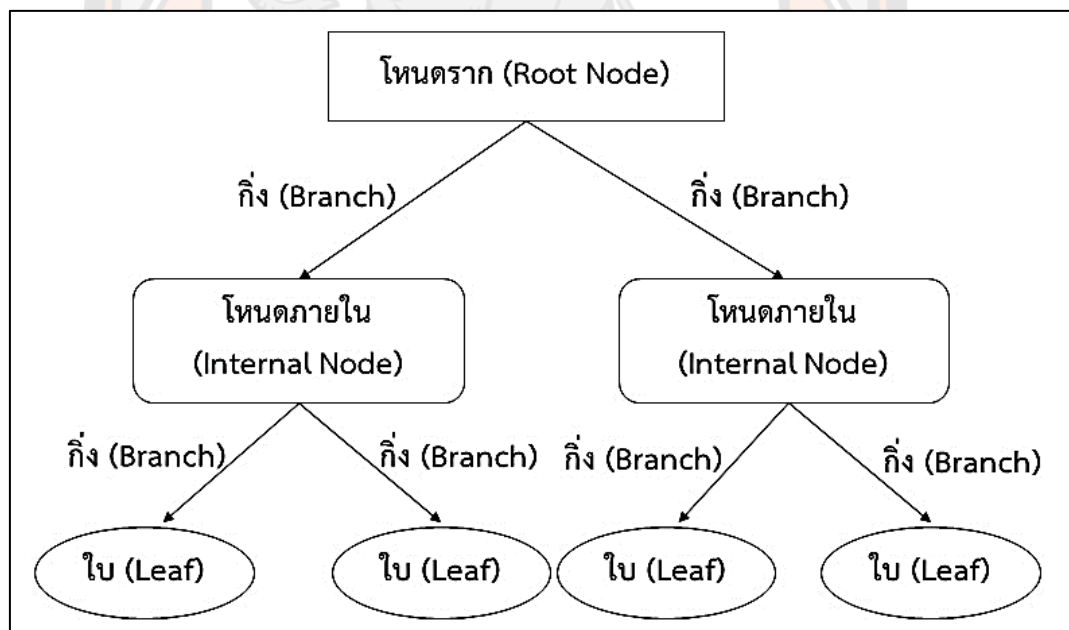
จากที่กล่าวมา การประมาณค่าสัมประสิทธิ์การถดถอยลอจิสติก (β) จะใช้หลักการของค่าความน่าจะเป็นสูงสุด (Maximum Likelihood) หรือประมาณค่า β ที่ทำให้ $\log_e L(y_1, \dots, y_n; \beta)$ ในสมการที่ (2.6) มีค่ามากที่สุด โดยใช้วิธีการหาอนุพันธ์บางส่วนเทียบกับ $\beta_0, \beta_1, \dots, \beta_k$ แล้วกำหนดให้เท่ากับ 0 เนื่องจากสมการไม่ได้อยู่ในรูปเชิงเส้น จึงต้องใช้เทคนิคการทำซ้ำ (Iteration techniques) โดย (Haberman, 1978) ได้ศึกษาโดยใช้วิธีการของนิวตัน - ราฟสัน (Newton -

Raphson method) ซึ่งพบว่าเป็นวิธีที่มีประสิทธิภาพ ดังนั้นส่วนใหญ่มักใช้โปรแกรมสำเร็จรูปทางสถิติในการประมาณค่าสัมประสิทธิ์ (กัลยา วานิชย์บัญชา, 2552)

2.3. เทคนิคต้นไม้ตัดสินใจ (Decision Tree : DT)

เทคนิคต้นไม้ตัดสินใจ (Decision Tree) (เอกรัฐ บุญเชิยง, 2561) เป็นวิธีการเรียนรู้ของเครื่อง ซึ่งเป็นการเรียนรู้แบบมีผู้สอน (Supervised Learning) โดยการจำแนก (Classification) ข้อมูลออกเป็นคลาส (Class) ต่างๆ โดยใช้คุณลักษณะ (Attribute) ของข้อมูลในการจำแนก ภายในต้นไม้ตัดสินใจประกอบด้วย 3 องค์ประกอบหลัก ได้แก่

1. โหนด (Node) แบ่งออกเป็น โหนดราก (Root Node) กับโหนดภายใน (Internal Node) และมีคุณลักษณะ (Attribute) เป็นตัวแปรที่ใช้ในการทดสอบ
 2. กิ่ง (Branch) เป็นกิ่งของต้นไม้ที่แสดงถึงค่าที่เป็นไปได้ของคุณลักษณะที่ถูกทดสอบ
 3. ใบ (Leaf) แสดงถึงกลุ่มของข้อมูล (Class) หรือผลลัพธ์ในการจำแนกประเภทข้อมูล
- ดังตัวอย่างที่แสดงในภาพ 2



ภาพ 2 องค์ประกอบของต้นไม้ตัดสินใจ

ที่มา : Akash, N. H. (2020).

ซึ่งหลักการพื้นฐานของการสร้างต้นไม้ตัดสินใจจะเป็นการสร้างจากบนลงล่าง (Top - Down) และการสร้างต้นไม้ตัดสินใจมีหลายวิธีโดยแต่ละวิธีถูกพัฒนามาจากวิธีพื้นฐาน คือ อัลกอริทึม ID3

อัลกอริทึม ID3 (Iterative Dichotomiser 3) เป็นต้นไม้ตัดสินใจที่ถูกเสนอขึ้นโดย John Ross Quinlan (Quinlan, 1986) โดยการใช้ค่าเกนสารสนเทศ (Information gain) ดังสมการ (2.11) มาช่วยในการเลือกคุณลักษณะที่เหมาะสมในแต่ละโหนด ซึ่งทำได้โดยการคำนวณค่าเอนโทรปี (Entropy) ของคุณลักษณะแต่ละตัว ดังสมการ (2.7)

คำนวณค่าเอนโทรปี (Entropy) ของเงื่อนไขในแต่ละคุณลักษณะแสดงดังสมการ (2.7)

$$Entropy(t) = - \sum_{i=1}^c p_i(t) \log_2 p_i(t) \quad (2.7)$$

โดย $p_i(t)$ คือ อัตราส่วนของเงื่อนไขในแต่ละคุณลักษณะ (t) ที่สนใจที่ตัวแปรตามมีค่า i
 c คือ จำนวนค่าจำแนกหรือจำนวนคลาส

จากนั้นนำค่าเอนโทรปีในแต่ละเงื่อนไขมาแทนในสมการ (2.8) เพื่อหาค่าเกนสารสนเทศ (Information gain)

$$Gain_{split} = Entropy(p) - \sum_{i=1}^k \frac{n_i}{n} Entropy(i) \quad (2.8)$$

โดย $Gain_{split}$ คือ ค่าเกนสารสนเทศ
 $Entropy(p)$ คือ ค่าเอนโทรปีที่พิจารณาจากกลุ่มคำตอบของตัวแปรตาม
 $Entropy(i)$ คือ ค่าเอนโทรปีของข้อมูลทุกตัวแปรอิสระ
 n_i คือ จำนวนข้อมูลในส่วนที่ i
 k คือ จำนวนค่าคุณสมบัติของคุณสมบัติ i

อัลกอริทึม C4.5 หรือ J48 เป็นวิธีที่ถูกปรับปรุงต่อเนื่องมาจากอัลกอริทึม ID3 โดยเพิ่ม Generalization เพื่อเลี่ยงปัญหา Overfitting และรองรับข้อมูลในรูปแบบที่หลากหลายมากขึ้น เช่น ข้อมูลที่มีค่าต่อเนื่องและไม่ต่อเนื่อง เป็นต้น (Quinlan, 1993) หลักการทำงานของอัลกอริทึม C4.5 ได้เพิ่มการใช้ค่าอัตราส่วนเกน (Gain Ratio) ในการตัดสินใจเลือกคุณลักษณะที่ใช้เป็นโหนด เนื่องจากค่าเกนมาตรฐาน (Gain Criterion) จะมีค่าเอนโทรปี (Bias) มากกับข้อมูลที่ประกอบด้วยคุณลักษณะที่

มีค่าเป็นไปได้จำนวนมากๆ เช่น ตัวอย่างที่ไม่ซ้ำกัน หรือตัวอย่างที่มีกลุ่มของข้อมูลกลุ่มเดียว ซึ่งสามารถแก้ไขได้ด้วยค่าอัตราส่วนเกิน โดยใช้ค่าสารสนเทศของการจำแนก (Split Information) ของคุณลักษณะแต่ละตัว ดังสมการ (2.9) เมื่อได้ค่าสารสนเทศของการจำแนกแล้ว สามารถคำนวณหาค่าอัตราส่วนเกิน (Gain Ratio) เพื่อลดความเอนเอียง ดังสมการ (2.10)

$$Split\ Info = - \sum_{i=1}^k \frac{n_i}{n} \log_2 \frac{n_i}{n} \quad (2.9)$$

$$Gain\ Ratio = \frac{Gain_{split}}{Split\ Info} \quad (2.10)$$

ตาราง 1 แสดงตัวอย่างข้อมูลการผ่าคลอด

Age	Delivery No	Blood of Pressure	Heart Problem	Caesarian
18	Timely	Normal	apt	No
20	Timely	High	inept	Yes
21	Premature	Low	inept	Yes
26	Latecomer	Normal	apt	Yes
27	Premature	Normal	apt	No
28	Timely	High	apt	Yes
30	Latecomer	High	apt	No
33	Latecomer	Normal	inept	No
35	Premature	Low	apt	No
37	Timely	Normal	inept	Yes
38	Latecomer	High	inept	Yes

ที่มา : UCI Machine Learning Repository

จากตารางที่ 1 ข้อมูลการตัดสินใจผ่าคลอด เป็นข้อมูลที่น่ามาใช้ศึกษาวิธีการสร้างต้นไม้ตัดสินใจโดยข้อมูลที่น่ามาศึกษาเป็นข้อมูลจากฐานข้อมูล UCI ซึ่งนำข้อมูลมาเพียงบางส่วนเท่านั้น มีข้อมูลจำนวน 11 ชุด มีคุณลักษณะทั้งหมด 4 คุณลักษณะ ได้แก่ อายุ (Age) เวลาคลอด (Delivery No) ความดันโลหิต (Blood of Pressure) และสภาวะหัวใจ (Heart Problem) สามารถแสดงขั้นตอนวิธีการสร้างต้นไม้ตัดสินใจ C4.5 ได้ดังนี้

ขั้นตอนที่ 1 คำนวณค่าเอนโทรปี โดยพิจารณาจากกลุ่มคำตอบของตัวแปรตาม

$$Entropy(Caesarian) = -\frac{6}{11}\log_2\left(\frac{6}{11}\right) - \frac{5}{11}\log_2\left(\frac{5}{11}\right) \approx 0.9940$$

ขั้นตอนที่ 2 คำนวณค่าเอนโทรปีของข้อมูลทุกคุณลักษณะหรือทุกตัวแปรอิสระ

- พิจารณาตัวแปรอิสระตัวที่ 1 คือ อายุ (Age) เนื่องจากข้อมูลเป็นค่าต่อเนื่อง ขั้นแรกต้องทำการเรียงข้อมูลเพื่อต้องการทราบจุดเปลี่ยนระหว่างค่า Yes เป็น No หรือ No เป็น Yes จากตารางที่ 1 พบจุดเปลี่ยนทั้งหมด 5 จุด จากนั้นใช้ค่ากลางเป็นค่าขีดขั้น ได้แก่ $(18 + 20)/2 = 19$, $(26 + 27)/2 = 26.5$, $(27 + 28)/2 = 27.5$, $(28 + 30)/2 = 36$ และ $(35 + 37)/2 = 36$

- ค่ากลางเท่ากับ 19

$$Entropy(\leq 19) = -\frac{1}{1}\log_2\left(\frac{1}{1}\right) = 0$$

$$Entropy(> 19) = -\frac{6}{10}\log_2\left(\frac{6}{10}\right) - \frac{4}{10}\log_2\left(\frac{4}{10}\right) \approx 0.9710$$

$$\therefore \sum_{i=1}^k \frac{n_i}{n} Entropy(i) = \frac{1}{11}(0) + \frac{10}{11}(0.9710) \approx 0.8827$$

- ค่ากลางเท่ากับ 26.5

$$Entropy(\leq 26.5) = -\frac{3}{4}\log_2\left(\frac{3}{4}\right) - \frac{1}{4}\log_2\left(\frac{1}{4}\right) \approx 0.8113$$

$$Entropy(> 26.5) = -\frac{3}{7}\log_2\left(\frac{3}{7}\right) - \frac{4}{7}\log_2\left(\frac{4}{7}\right) \approx 0.9852$$

$$\therefore \sum_{i=1}^k \frac{n_i}{n} Entropy(i) = \frac{4}{11}(0.8113) + \frac{7}{11}(0.9852) \approx 0.9220$$

- ค่ากลางเท่ากับ 27.5

$$Entropy(\leq 27.5) = -\frac{3}{5}\log_2\left(\frac{3}{5}\right) - \frac{2}{5}\log_2\left(\frac{2}{5}\right) \approx 0.9710$$

$$Entropy(> 27.5) = -\frac{3}{6}\log_2\left(\frac{3}{6}\right) - \frac{3}{6}\log_2\left(\frac{3}{6}\right) = 1$$

$$\therefore \sum_{i=1}^k \frac{n_i}{n} Entropy(i) = \frac{5}{11}(0.9710) + \frac{6}{11}(1) \approx 0.9868$$

- ค่ากลางเท่ากับ 29

$$Entropy(\leq 29) = -\frac{4}{6}\log_2\left(\frac{4}{6}\right) - \frac{2}{6}\log_2\left(\frac{2}{6}\right) \approx 0.9183$$

$$Entropy(> 29) = -\frac{2}{5}\log_2\left(\frac{2}{5}\right) - \frac{3}{5}\log_2\left(\frac{3}{5}\right) \approx 0.9710$$

$$\therefore \sum_{i=1}^k \frac{n_i}{n} Entropy(i) = \frac{6}{11}(0.9183) + \frac{5}{11}(0.9710) \approx 0.9423$$

- ค่ากลางเท่ากับ 36

$$Entropy(\leq 36) = -\frac{4}{9}\log_2\left(\frac{4}{9}\right) - \frac{5}{9}\log_2\left(\frac{5}{9}\right) \approx 0.9911$$

$$Entropy(> 36) = -\frac{2}{2}\log_2\left(\frac{2}{2}\right) = 0$$

$$\therefore \sum_{i=1}^k \frac{n_i}{n} Entropy(i) = \frac{9}{11}(0.9911) + \frac{2}{11}(0) \approx 0.8109$$

○ พิจารณาตัวแปรอิสระตัวที่ 2 คือ เวลาคลอด (Delivery No)

$$Entropy(Timely) = -\frac{3}{4}\log_2\left(\frac{3}{4}\right) - \frac{1}{4}\log_2\left(\frac{1}{4}\right) \approx 0.8113$$

$$Entropy(Premature) = -\frac{1}{3}\log_2\left(\frac{1}{3}\right) - \frac{2}{3}\log_2\left(\frac{2}{3}\right) \approx 0.9183$$

$$Entropy(Latecomer) = -\frac{2}{4}\log_2\left(\frac{2}{4}\right) - \frac{2}{4}\log_2\left(\frac{2}{4}\right) = 1$$

$$\therefore \sum_{i=1}^k \frac{n_i}{n} Entropy(i) = \frac{4}{11}(0.8113) + \frac{3}{11}(0.9183) + \frac{4}{11}(1) \approx 0.9091$$

○ พิจารณาตัวแปรอิสระตัวที่ 3 คือ ความดันโลหิต (Blood of Pressure)

$$Entropy(Low) = -\frac{1}{2}\log_2\left(\frac{1}{2}\right) - \frac{1}{2}\log_2\left(\frac{1}{2}\right) = 1$$

$$Entropy(Normal) = -\frac{2}{5}\log_2\left(\frac{2}{5}\right) - \frac{3}{5}\log_2\left(\frac{3}{5}\right) \approx 0.9710$$

$$Entropy(High) = -\frac{3}{4}\log_2\left(\frac{3}{4}\right) - \frac{1}{4}\log_2\left(\frac{1}{4}\right) \approx 0.8113$$

$$\therefore \sum_{i=1}^k \frac{n_i}{n} Entropy(i) = \frac{2}{11}(1) + \frac{5}{11}(0.9710) + \frac{4}{11}(0.8113) \approx 0.9182$$

○ พิจารณาตัวแปรอิสระตัวที่ 4 คือ สภาวะหัวใจ (Heart Problem)

$$Entropy(apt) = -\frac{2}{6}\log_2\left(\frac{2}{6}\right) - \frac{4}{6}\log_2\left(\frac{4}{6}\right) \approx 0.9183$$

$$Entropy(inept) = -\frac{4}{5}\log_2\left(\frac{4}{5}\right) - \frac{1}{5}\log_2\left(\frac{1}{5}\right) \approx 0.7219$$

$$\therefore \sum_{i=1}^k \frac{n_i}{n} Entropy(i) = \frac{6}{11}(0.9183) + \frac{5}{11}(0.7219) \approx 0.8290$$

ขั้นตอนที่ 3 คำนวณค่าเกนสารสนเทศ (Information gain) ของทุกตัวแปรอิสระ

$$Gain_{split}(Caesarian, Age: 19) = 0.9940 - 0.8827 = 0.1113$$

$$Gain_{split}(Caesarian, Age: 26.5) = 0.9940 - 0.9220 = 0.0720$$

$$Gain_{split}(Caesarian, Age: 27.5) = 0.9940 - 0.9868 = 0.0072$$

$$Gain_{split}(Caesarian, Age: 29) = 0.9940 - 0.9423 = 0.0517$$

$$Gain_{split}(Caesarian, Age: 36) = 0.9940 - 0.8109 = 0.1831$$

$$Gain_{split}(Caesarian, DeliveryNo) = 0.9940 - 0.9091 = 0.0849$$

$$Gain_{split}(Caesarian, BloodOfPressure) = 0.9940 - 0.9182 = 0.0758$$

$$Gain_{split}(Caesarian, HeartProblem) = 0.9940 - 0.8290 = 0.1650$$

จากขั้นตอนที่ 3 จะเห็นว่าค่าเกณฑ์สารสนเทศที่มีค่ามากที่สุดคือ อายุ (Age) ที่มีค่ากลางเท่ากับ 36 ดังนั้นจึงเลือก อายุ เป็นโหนดราก ในส่วนของอัลกอริทึม ID3 จะหยุดการทำงานในขั้นตอนที่ 3 และจะทำซ้ำในขั้นตอนที่ 1 - 3 เพื่อที่จะเลือกโหนดภายใน ส่วนอัลกอริทึม C4.5 จะเริ่มสร้างตั้งแต่ขั้นตอนที่ 1 - 3 และพิจารณาขั้นตอนต่อไป

ขั้นตอนที่ 4 คำนวณหาค่าเกณฑ์สารสนเทศของการจำแนกของทุกตัวแปรอิสระ

$$Split\ Info(Age: 19) = -\frac{1}{11}\log_2\left(\frac{1}{11}\right) - \frac{10}{11}\log_2\left(\frac{10}{11}\right) \approx 0.4395$$

$$Split\ Info(Age: 26.5) = -\frac{4}{11}\log_2\left(\frac{4}{11}\right) - \frac{7}{11}\log_2\left(\frac{7}{11}\right) \approx 0.9457$$

$$Split\ Info(Age: 27.5) = -\frac{5}{11}\log_2\left(\frac{5}{11}\right) - \frac{6}{11}\log_2\left(\frac{6}{11}\right) \approx 0.9940$$

$$Split\ Info(Age: 29) = -\frac{6}{11}\log_2\left(\frac{6}{11}\right) - \frac{5}{11}\log_2\left(\frac{5}{11}\right) \approx 0.9940$$

$$Split\ Info(Age: 36) = -\frac{9}{11}\log_2\left(\frac{9}{11}\right) - \frac{2}{11}\log_2\left(\frac{2}{11}\right) \approx 0.6840$$

$$Split\ Info(DeliveryNo) = -\frac{4}{11}\log_2\left(\frac{4}{11}\right) - \frac{3}{11}\log_2\left(\frac{3}{11}\right) - \frac{4}{11}\log_2\left(\frac{4}{11}\right) \approx 1.5726$$

$$Split\ Info(BloodOfPressure) = -\frac{2}{11}\log_2\left(\frac{2}{11}\right) - \frac{5}{11}\log_2\left(\frac{5}{11}\right) - \frac{4}{11}\log_2\left(\frac{4}{11}\right) \approx 1.4949$$

$$Split\ Info(HeartProblem) = -\frac{5}{11}\log_2\left(\frac{5}{11}\right) - \frac{6}{11}\log_2\left(\frac{6}{11}\right) \approx 0.9940$$

ขั้นตอนที่ 5 คำนวณหาค่าอัตราส่วนเกณฑ์สารสนเทศของทุกตัวแปรอิสระ

$$GainRatio(Age: 19) = \frac{0.1113}{0.4395} = 0.2532$$

$$GainRatio(Age: 26.5) = \frac{0.0720}{0.9457} = 0.0761$$

$$GainRatio(Age: 27.5) = \frac{0.0072}{0.9940} = 0.0072$$

$$GainRatio(Age: 29) = \frac{0.0517}{0.9940} = 0.0520$$

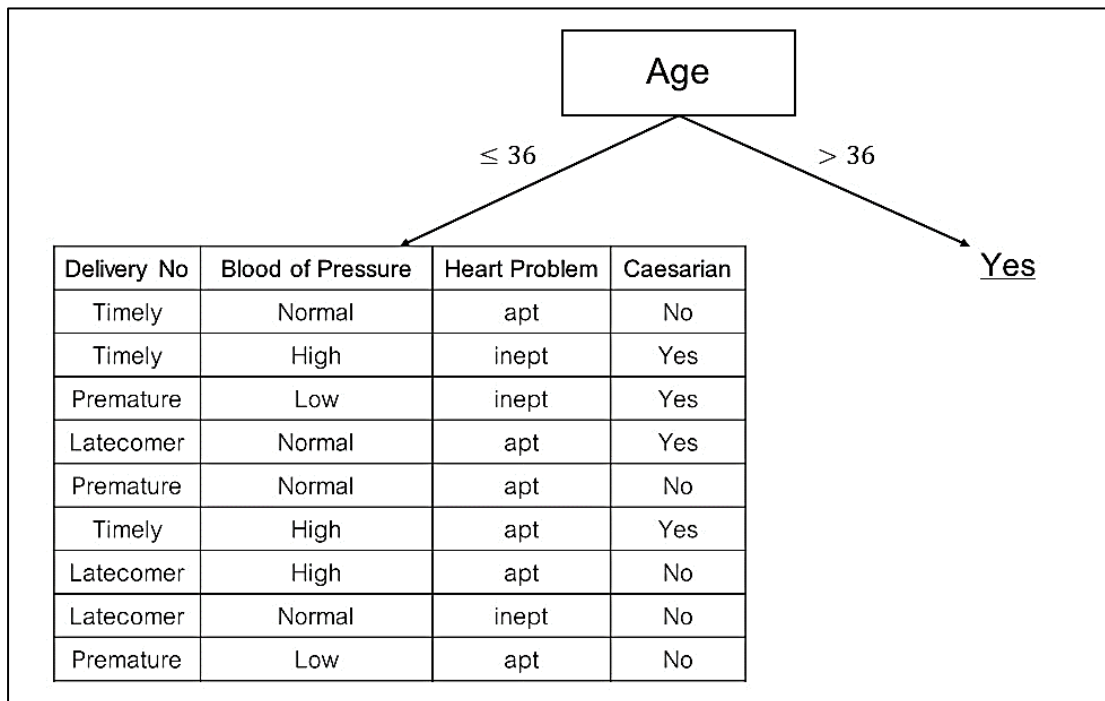
$$GainRatio(Age: 36) = \frac{0.1831}{0.6840} = 0.2677$$

$$GainRatio(DeliveryNo) = \frac{0.0849}{1.5726} = 0.0540$$

$$\text{GainRatio}(\text{BloodOfPressure}) = \frac{0.0758}{1.4949} = 0.0507$$

$$\text{GainRatio}(\text{HeartProblem}) = \frac{0.1650}{0.9940} = 0.1660$$

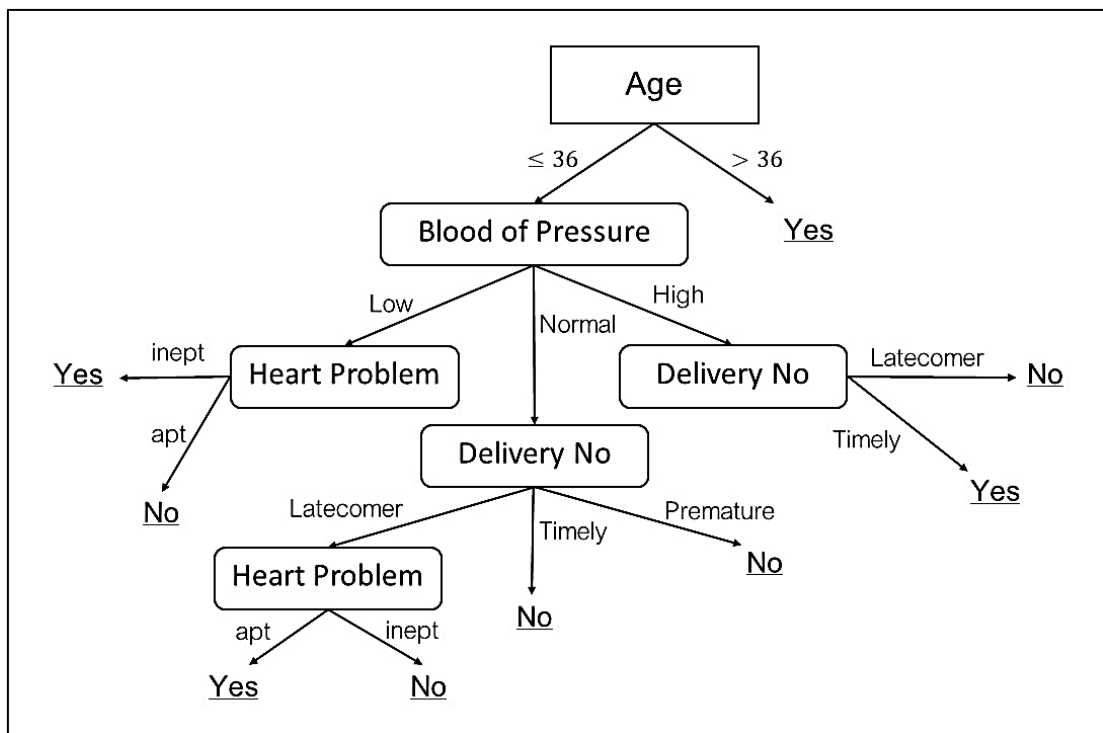
จากขั้นตอนที่ 5 ค่าอัตราส่วนเกณฑ์ที่มีค่ามากที่สุดคือ อายุ (Age) ดังนั้นจึงเลือกอายุที่มีค่ากลางเท่ากับ 36 เป็นโหนดรากดังภาพ 3



ภาพ 3 เหตุการณ์การเลือกตัวแปรอิสระเป็นโหนดราก

จากภาพ 3 จะเห็นว่าโหนดภายใน อายุมากกว่า 36 ปี มีใบหนึ่งใบคือ yes

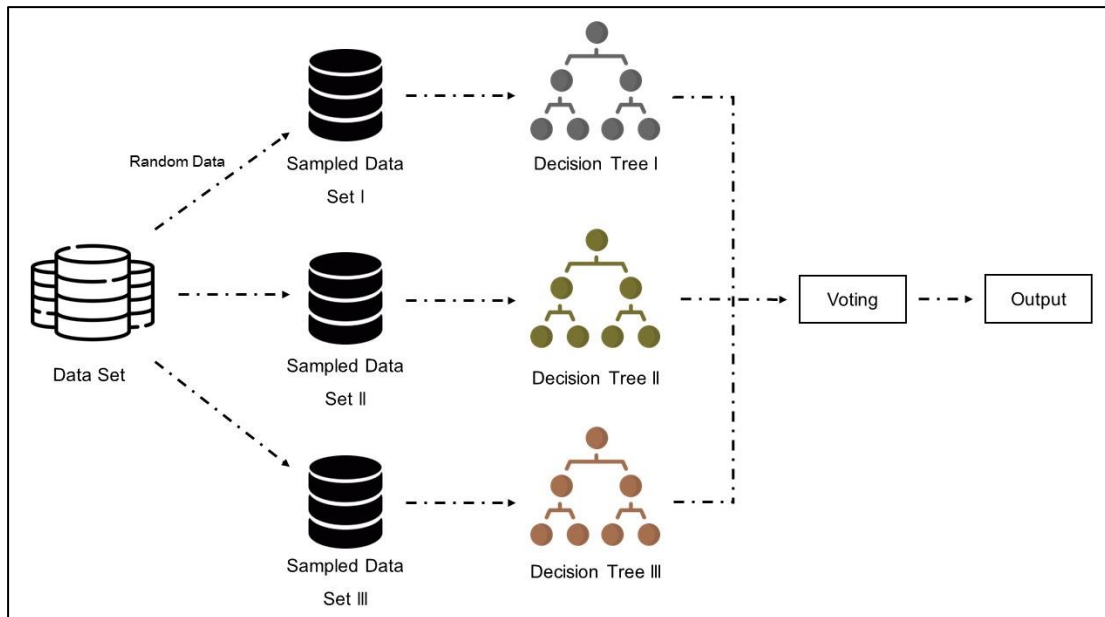
ขั้นตอนที่ 6 ทำซ้ำในขั้นตอนที่ 1 – 5 เพื่อเลือกโหนดภายใน และจะทำให้ได้ต้นไม้ตัดสินใจภาพ 4



ภาพ 4 แผนภาพต้นไม้ตัดสินใจของข้อมูลการผ่าคลอด

2.4. เทคนิคป่าสุ่ม (Random Forest : RF)

เทคนิคป่าสุ่ม (Random Forest) เป็นเทคนิคหนึ่งของการเรียนรู้ของเครื่อง ถูกนำเสนอครั้งแรกโดย Tin Kam เมื่อปี ค.ศ. 1995 และต่อมาถูกพัฒนาโดย Leo Breiman สามารถใช้ได้กับปัญหาการจำแนกประเภท (Classification Problem) และปัญหาการถดถอย (Regression Problem) เป็นเทคนิคที่พัฒนามาจากเทคนิคต้นไม้ตัดสินใจ (Decision Tree) โดยการทำงานของเทคนิคป่าสุ่มคือ จะสร้างต้นไม้ตัดสินใจหลายๆ ต้น ซึ่งแต่ละต้นจะมีคุณลักษณะ (Attribute) และข้อมูล (Data) ที่แตกต่างกัน เพื่อให้ต้นไม้ที่ได้มีความหลากหลายและเป็นอิสระกัน



ภาพ 5 หลักการทำงานของเทคนิคป่าสุ่ม (Random Forest)

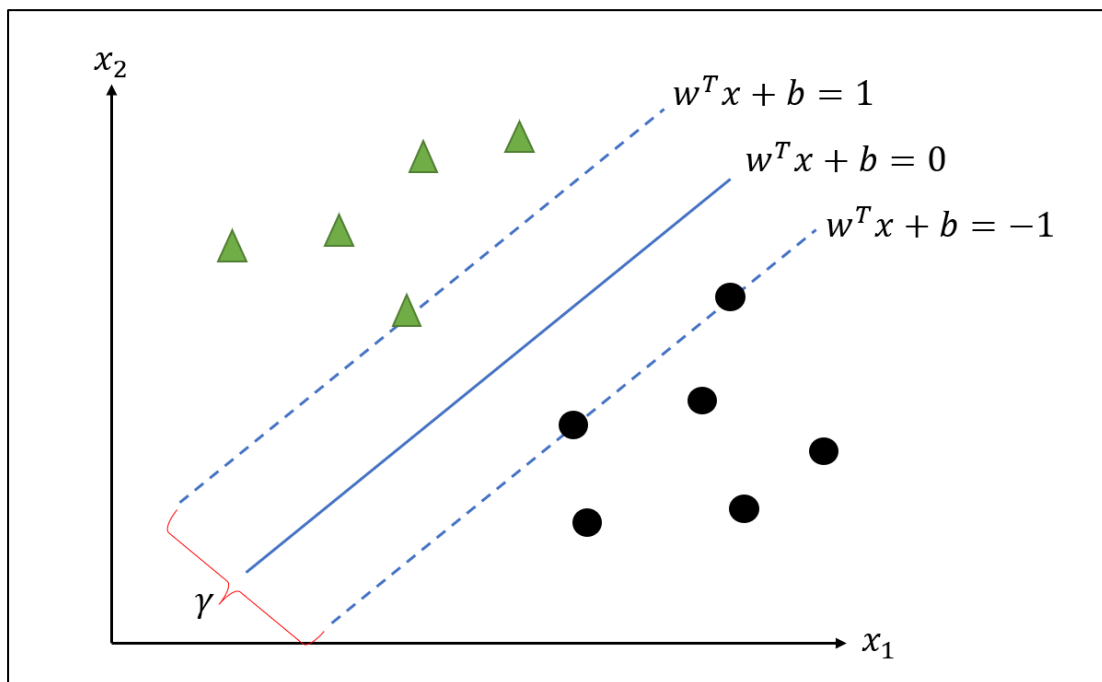
ที่มา : วิชญ์พงศ์ ดรุณธรรม, (2018)

จากภาพ 5 หลักการทำงานของเทคนิคป่าสุ่มเริ่มจากการสุ่มข้อมูลและสุ่มคุณลักษณะเพื่อสร้างชุดข้อมูลใหม่จากข้อมูลทั้งหมดให้ได้ออกมาตามจำนวนที่ต้องการ นำชุดข้อมูลใหม่มาสร้างต้นไม้ตัดสินใจ (Breiman, 2001) และทำการหาคำตอบจากตัวแบบของต้นไม้แต่ละต้น จากนั้นหาคำคำตอบสุดท้าย กรณีที่เป็นปัญหาการจำแนกประเภทจะใช้วิธีผลโหวตมากที่สุด (Majority vote) นั่นคือหาคำตอบใดมีผลโหวตมากที่สุดจะเป็นคำตอบของชุดข้อมูลนั้น ในส่วนปัญหาการถดถอยจะใช้วิธีการหาค่าเฉลี่ย (Mean) โดยนำคำตอบของแต่ละตัวแบบมาหาค่าเฉลี่ย

2.5. เทคนิคซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine : SVM)

ซัพพอร์ตเวกเตอร์แมชชีน (SVM) พัฒนาโดย Cortes และ Vapnik เมื่อปีค.ศ. 1995 ได้รับการพัฒนาขึ้นเพื่อแก้ปัญหาการจำแนกประเภท (Classification Problem) และต่อมาพบว่าช่วยแก้ปัญหาการถดถอย (Regression Problem) ได้ ซัพพอร์ตเวกเตอร์แมชชีนเป็นตัวแยกประเภทที่กำหนดระนาบเกน (Hyperplane) เพื่อจำแนกประเภทข้อมูล โดยมีวัตถุประสงค์ที่พยายามลดความผิดพลาดในการจำแนก (Minimize error) และเพิ่มระยะขอบให้มากที่สุด (Maximized Margin) อีกทั้งยังรองรับข้อมูลที่มีลักษณะเป็นเชิงเส้นและไม่เชิงเส้น เนื่องจากข้อมูลส่วนใหญ่ไม่เป็นเชิงเส้น จะใช้กลวิธีเคอร์เนล (Kernel Trick) มาลดความซับซ้อนของข้อมูลโดยการส่ง (Mapping) ไปยังปริภูมิพิเศษที่มีความเป็นเชิงเส้น ฟังก์ชันนี้สามารถช่วยหลีกเลี่ยงการเกิดปัญหา Overfitting (Duan, 2018)

ในซัพพอร์ตเวกเตอร์แมชชีน (SVM) ระยะขอบที่มากที่สุดระหว่างคลาสหนึ่งไปอีกคลาสหนึ่ง ถูกใช้เป็นบรรทัดฐานในการตัดสินใจเลือกระนาบเกินที่เหมาะสมที่สุด (Optimal Hyperplane)(Vapnik, 1995) และข้อมูลบริเวณขอบของแต่ละคลาส เรียกว่า เวกเตอร์ซัพพอร์ต (Support Vector) ดังแสดงในภาพ 6



ภาพ 6 SVM แบบเชิงเส้นกรณี Hard Margin

ที่มา : Gholami, R., & Fakhari, N. (2017)

ความสัมพันธ์ของระยะขอบกับระนาบเกินแสดงได้ดังนี้ (ปริญา สงวนสัตย์, 2562, หน้า 151)

$$\gamma = \frac{2}{\|w\|} = \frac{2}{\sqrt{w^T w}} \propto \frac{2}{w^T w} \quad (2.11)$$

ข้อมูลที่ใช้ในกระบวนการเรียนรู้เพื่อหาระนาบเกินที่เหมาะสมที่สุด เมื่อเราพิจารณาข้อมูลที่ประกอบด้วยข้อมูล 2 กลุ่ม แสดงดังสมการ (2.12)

$$D = \{(x_i, y_i); i = 1, 2, \dots, n\} \quad (2.12)$$

เมื่อ $x_i = (x_{i1}, x_{i2}, \dots, x_{im}) \in R^m$

$y_i \in \{-1, +1\}$ โดย 1 คือ ข้อมูลกลุ่ม 1 และ -1 คือข้อมูลกลุ่ม 2

n คือ จำนวนข้อมูล

m คือ จำนวนตัวแปรอิสระ
สามารถเขียนฟังก์ชันเชิงเส้นของระนาบเกิน ได้ดังนี้

$$f(x) = w^T x + b \quad (2.13)$$

เมื่อ b คือ ค่าเอนเอียง และ w คือ ค่าเวกเตอร์น้ำหนัก ซึ่งสมการ (2.13) ใช้สำหรับกรณีที่สามารถจำแนกประเภทข้อมูลเชิงเส้น โดยระนาบเกินอยู่ภายใต้เงื่อนไขดังนี้

$$\begin{aligned} w^T x_i + b > 0 & \text{ ถ้า } y_i = +1 \\ w^T x_i + b < 0 & \text{ ถ้า } y_i = -1 \end{aligned} \quad (2.14)$$

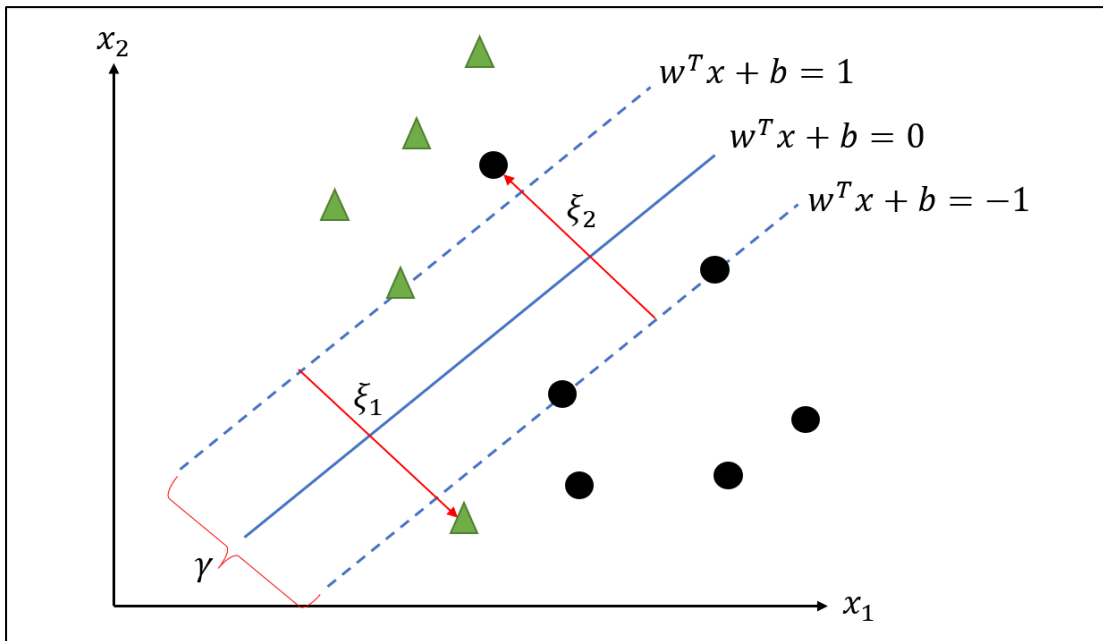
สามารถเขียนในรูปแบบอสมการทางคณิตศาสตร์ได้ดังนี้

$$y_i(w^T x_i + b) \geq 1, \forall i \quad (2.15)$$

จากความสัมพันธ์ของระยะขอบกับระนาบเกินในสมการ (2.11) และสมการ (2.15) ดังนั้นปัญหาการหาค่าที่เหมาะสมที่สุดเป็นดังนี้

$$\begin{aligned} \text{Minimize: } & \frac{1}{2} w^T w \\ \text{Subject to: } & y_i(w^T x_i + b) \geq 1, \forall i \end{aligned}$$

ซึ่งเป็นการแก้ปัญหาขอบแข็ง (Hard Margin) แต่ความเป็นจริงข้อมูลอาจไม่เป็นดังภาพ 6 เสมอไป กล่าวคือ อาจจะมีข้อมูลที่จัดอยู่ในกลุ่มที่ 1 ไปอยู่ในกลุ่มที่ 2 ดังภาพ 7 ดังนั้นจึงมีการแก้ปัญหาเพื่อหาค่าขอบที่เรียกว่า การแก้ปัญหาขอบอ่อน (Soft Margin) โดยเพิ่มตัวแปร Slack (ξ_i) เข้าไปในปัญหาเพื่อยอมรับความผิดพลาด



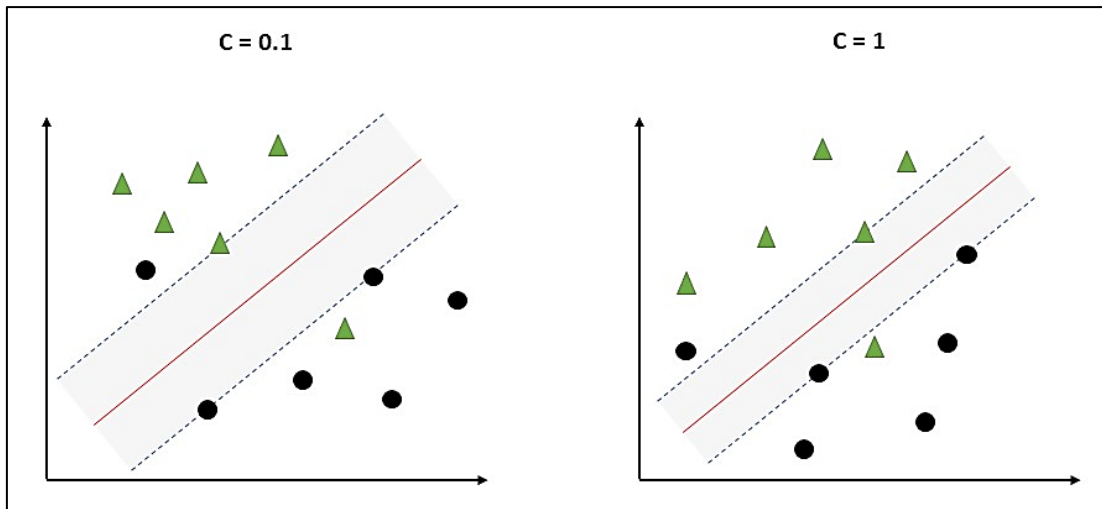
ภาพ 7 SVM แบบเชิงเส้นกรณี Soft Margin

ที่มา : Gholami, R., & Fakhari, N. (2017)

ดังนั้น ปัญหาการหาค่าเหมาะสมที่สุดเป็นดังนี้

$$\begin{aligned} \text{Minimize: } & \frac{1}{2} w^T w + C \sum_{i=1}^n \xi_i \\ \text{Subject to: } & y_i(w^T x_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \forall i \end{aligned}$$

โดยที่ C คือ ค่าคงที่ ซึ่งเป็นตัวกำหนดปริมาณความผิดพลาดในการพยากรณ์ หากมีค่ามาก หมายความว่ายอมให้เกิดความผิดพลาดน้อย ระยะขอบแคบ แต่หากมีค่าน้อยนั้นหมายความว่ายอมให้เกิดความผิดพลาดมาก ระยะขอบกว้าง แสดงดังภาพ 8 ดังนั้นการเลือก C จะส่งผลต่อประสิทธิภาพการทำงานของเทคนิคซัพพอร์ตเวกเตอร์แมชชีน



ภาพ 8 ตัวอย่างค่า C ที่มีผลต่อขนาดของระยะขอบ

ที่มา : <https://medium.com/@pradyasin/support-vector-machines-svm-943f9a732a69>

ซึ่งการแก้ปัญหาค่าที่เหมาะสมสามารถทำได้โดยการพิจารณาปัญหาควบคู่ (Dual Problem) โดยการใช้ตัวคูณลากรางจ์ (Lagrange Multipliers) ด้วยเงื่อนไข Karush-Kuhn-Tucker (KKT) ดังนั้นฟังก์ชันวัตถุประสงค์ของปัญหาการหาค่าที่เหมาะสมที่สุดแสดงดังนี้ (ปริญา สงวนสัตย์, 2562, หน้า 161)

$$L(w, b, \alpha, \xi, r) = \frac{1}{2} w^T w + C \sum_{i=1}^n \xi_i - \sum_{i=1}^n \alpha_i [y_i (w^T x_i + b) - 1 + \xi_i] - \sum_{i=1}^n r_i \xi_i \quad (2.16)$$

โดย α_i เรียกว่า ตัวคูณลากรางจ์ด้วยเงื่อนไข KKT ซึ่ง $\alpha_i \geq 0, \forall i$ และ $r_i \geq 0, \forall i$ หาอนุพันธ์ย่อยเทียบตัวแปรต่าง ๆ เพื่อหาค่าที่เหมาะสมที่สุด ได้ดังนี้

$$\frac{\partial L(w, b, \alpha, \xi, r)}{\partial w} = w - \sum_{i=1}^n y_i \alpha_i x_i = 0 \quad (2.17)$$

$$\frac{\partial L(w, b, \alpha, \xi, r)}{\partial \xi_i} = C - \alpha_i - r_i = 0 \quad (2.18)$$

$$\frac{\partial L(w, b, \alpha, \xi, r)}{\partial b} = \sum_{i=1}^n y_i \alpha_i = 0 \quad (2.19)$$

หรือสมการ (2.17) คือ
$$w = \sum_{i=1}^n y_i \alpha_i x_i \quad (2.20)$$

และสมการ (2.18) คือ
$$C = \alpha_i - r_i \quad (2.21)$$

นำค่าที่ได้จากสมการ (2.19), (2.20) และ (2.21) แทนในสมการ (2.16) จะได้ฟังก์ชันวัตถุประสงค์ดังนี้

$$\begin{aligned} L(w, b, \alpha, \xi, r) &= \frac{1}{2} w^T w + C \sum_{i=1}^n \xi_i - \sum_{i=1}^n \alpha_i [y_i (w^T x_i + b) - 1 + \xi_i] - \sum_{i=1}^n r_i \xi_i \\ &= \frac{1}{2} w^T w + (\alpha_i + r_i) \sum_{i=1}^n \xi_i - \sum_{i=1}^n \alpha_i [y_i (w^T x_i + b) - 1 + \xi_i] \\ &\quad - \sum_{i=1}^n r_i \xi_i \\ &= \frac{1}{2} w^T w - \sum_{i=1}^n \alpha_i [y_i (w^T x_i + b) - 1] \\ &= \frac{1}{2} \sum_{i,j=1}^n y_i y_j \alpha_i \alpha_j x_i^T x_j - \sum_{i,j=1}^n y_i y_j \alpha_i \alpha_j x_i^T x_j + \sum_{i=1}^n \alpha_i \\ &= \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n y_i y_j \alpha_i \alpha_j x_i^T x_j \end{aligned} \quad (2.22)$$

จะได้ปัญหาค่าที่เหมาะสมแสดงได้ดังนี้

$$\begin{aligned} \text{Minimize:} & \quad \frac{1}{2} \sum_{i,j=1}^n y_i y_j \alpha_i \alpha_j x_i^T x_j - \sum_{i=1}^n \alpha_i \\ \text{Subject to:} & \quad \sum_{i=1}^n y_i \alpha_i = 0, 0 \leq \alpha_i \leq C, \quad \forall i \end{aligned}$$

จากที่กล่าวมาเป็นข้อมูลแบบเชิงเส้น สามารถใช้ระนาบเกินเชิงเส้นเพื่อจำแนกประเภทของข้อมูลได้ แต่ในกรณีที่ข้อมูลแบบไม่เชิงเส้น ต้องทำการแปลงปริภูมิไม่เชิงเส้น (Nonlinear Space) ให้เป็นปริภูมิแบบเชิงเส้น โดยการหาฟังก์ชันการส่ง (Mapping Function; $\Phi(x)$) ที่สามารถแปลงปริภูมิแบบไม่เชิงเส้นเป็นปริภูมิแบบเชิงเส้นได้ ด้วยการใช้ทริคเคอร์เนล (Kernel Trick)

ฟังก์ชันเคอร์เนล (Kernel Function; $K(x_i, x_j)$) เป็นฟังก์ชันการส่งที่อยู่ในลักษณะรูปเมทริกซ์ Gram เกิดจากผลคูณภายใน (Inner Product) เป็นการแปลงข้อมูลจากพื้นที่ข้อมูลนำเข้า (Input Space) ให้เป็นพื้นที่คุณลักษณะ (Feature Space) ซึ่งฟังก์ชันเคอร์เนลเป็นฟังก์ชันที่แก้ไข

ภายใต้เงื่อนไขของเมอร์เซอร์ (Mercer's condition)(Courant and Hilbert, 1953) มีค่าเท่ากับผลคูณภายในของ x_i กับ x_j ในพื้นที่คุณลักษณะ $\Phi(x_i)$ และ $\Phi(x_j)$

$$K(x_i, x_j) = \Phi(x_i)^T \Phi(x_j) \quad (2.23)$$

ซึ่งฟังก์ชันเคอร์เนลมีหลายวิธี เช่น โพลีโนเมียล (Polynomial) เรเดียลเบสซิสฟังก์ชัน (Radial Basis Function : RBF) และ ซิกมอยด์ (Sigmoid) เป็นต้น โดยในงานวิจัยนี้เลือกใช้เรเดียลเบสซิสฟังก์ชัน (Radial Basis Function : RBF) ในการศึกษาแสดงดังสมการ (2.24)

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \quad (2.24)$$

กำหนดให้ค่า γ เป็นค่าพารามิเตอร์ที่ผู้ใช้สามารถกำหนดค่าให้ก่อนเริ่มทำงาน จากการแปลงปริภูมิแบบไม่เชิงเส้นเป็นปริภูมิแบบเชิงเส้นโดยอาศัย กลวิธีเคอร์เนล ทำให้ได้ปัญหาการหาค่าที่เหมาะสมดังนี้

$$\begin{aligned} \text{Minimize: } & \frac{1}{2} \sum_{i,j=1}^n y_i y_j \alpha_i \alpha_j K(x_i, x_j) - \sum_{i=1}^n \alpha_i \\ \text{Subject to: } & \sum_{i=1}^n y_i \alpha_i = 0, 0 \leq \alpha_i \leq C, \quad \forall i \end{aligned}$$

เมื่อได้ค่าที่เหมาะสม α_i ดังนั้น ฟังก์ชันของระนาบเกินแสดงดังสมการ (2.25)

$$f(x) = \text{sgn} \left(\sum_{i=1}^n y_i \alpha_i K(x_i, x) + b \right) \quad (2.25)$$

โดย $b = y_i - \sum_{j=1}^n \alpha_j y_j K(x_i, x_j)$ และ x_i คือ เวกเตอร์ซัพพอร์ต

2.6. หลักการพัฒนาตัวแบบการจำแนกและเครื่องมือวัดประสิทธิภาพ

ขั้นตอนพัฒนาตัวแบบการจำแนกแบ่งได้เป็น 2 ขั้นตอน ในขั้นแรกนำชุดข้อมูลมาทำการแบ่งออกเป็น 2 ชุด ประกอบด้วย ชุดข้อมูลเรียนรู้ (Training Set) และชุดข้อมูลทดสอบ (Test Set) ขั้นที่สองนำชุดข้อมูลเรียนรู้ไปสร้างตัวแบบการจำแนก และใช้ชุดข้อมูลทดสอบมาเป็นตัววัดประสิทธิภาพของตัวแบบการจำแนก ดังนั้นประสิทธิภาพของตัวแบบการจำแนกจะขึ้นอยู่กับความถูกต้องในการจำแนกประเภทข้อมูลในชุดข้อมูลทดสอบ

การแบ่งข้อมูลนั้นมีด้วยกันหลายวิธี ในที่นี้ผู้วิจัยใช้การแบ่งข้อมูลด้วยวิธีตรวจสอบไขว้ (k - Fold Cross-Validation) เพื่อป้องกันและลดโอกาสการเกิด Overfitting ของตัวแบบการจำแนก หลักการทำงาน คือ จะทำการแบ่งข้อมูลออกเป็น k กลุ่มที่มีจำนวนข้อมูลเท่ากัน และทำการคำนวณ k รอบ โดยแต่ละรอบข้อมูลหนึ่งกลุ่มจะถูกนำมาเป็นตัวทดสอบประสิทธิภาพของตัวแบบการจำแนก เรียกว่าชุดข้อมูลทดสอบ (Test Set) และข้อมูลอีก k-1 กลุ่มจะถูกนำไปสร้างตัวแบบการจำแนก เรียกว่าชุดข้อมูลเรียนรู้ (Training Set) ซึ่งข้อมูลแต่ละส่วนจะถูกวนให้เป็นชุดข้อมูลเรียนรู้และชุดข้อมูลทดสอบจนครบจำนวนที่แบ่งไว้คือ k รอบการทดสอบ ดังภาพ 9 ซึ่งเป็นการทำงานของ 10 - Fold Cross-Validation นั่นคือข้อมูลทั้งหมด 100% จะถูกแบ่งเป็นกลุ่มกลุ่มละ 10% จากนั้นจะทำการทดสอบโดยรอบแรกให้ชุดที่ 1 เป็นชุดข้อมูลทดสอบ และชุดที่ 2-10 เป็นชุดข้อมูลเรียนรู้ ทำวนไปจนครบทั้งหมด 10 รอบการทดสอบ แล้วนำผลลัพธ์ที่ได้ในแต่ละรอบมาหาค่าเฉลี่ย เพื่อเป็นผลลัพธ์ของประสิทธิภาพของตัวแบบการจำแนก

รอบที่ 1	----->	1	2	3	4	5	6	7	8	9	10
รอบที่ 2	----->	1	2	3	4	5	6	7	8	9	10
รอบที่ 3	----->	1	2	3	4	5	6	7	8	9	10
รอบที่ 4	----->	1	2	3	4	5	6	7	8	9	10
รอบที่ 5	----->	1	2	3	4	5	6	7	8	9	10
รอบที่ 6	----->	1	2	3	4	5	6	7	8	9	10
รอบที่ 7	----->	1	2	3	4	5	6	7	8	9	10
รอบที่ 8	----->	1	2	3	4	5	6	7	8	9	10
รอบที่ 9	----->	1	2	3	4	5	6	7	8	9	10
รอบที่ 10	----->	1	2	3	4	5	6	7	8	9	10

	คือ ชุดข้อมูลเรียนรู้		คือ ชุดข้อมูลทดสอบ
--	-----------------------	--	--------------------

ภาพ 9 แสดงวิธีการแบบ 10 - Fold Cross Validation

เครื่องมือวัดประสิทธิภาพของตัวแบบการจำแนกจะใช้ค่าที่ได้จากตารางเมทริกซ์ความสับสน (Confusion Matrix) ซึ่งเป็นตารางสรุปการจำแนกประเภทข้อมูลของตัวแบบการจำแนก โดยสามารถวัดประสิทธิภาพของตัวแบบด้วยการหาค่าความไว (Sensitivity) ค่าความจำเพาะ (Specificity) ค่าความแม่นยำ (Accuracy) ค่าความแม่นยำที่สมดุลแบบปรับน้ำหนักถ่วง (Weight Balanced Accuracy) และค่าประสิทธิภาพโดยรวม ($F1_{weight}$) รายละเอียดการคำนวณแสดงได้ดังต่อไปนี้

ตาราง 2 เมทริกซ์ความสับสน (Confusion Matrix)

ค่าจริง (Actual Class)	ค่าคำตอบจากตัวจำแนกหรือพยากรณ์ (Predicted Class)	
	Yes	No
Yes	TP	FN
No	FP	TN

โดยที่

ผลบวกจริง (TP) คือ ค่าพยากรณ์ตรงกับค่าจริง ในกรณีค่าพยากรณ์เป็นจริง และค่าจริงเป็น

จริง

ผลลบจริง (TN) คือ ค่าพยากรณ์ตรงกับค่าจริง ในกรณีค่าพยากรณ์เป็นเท็จ และค่าจริงเป็น

เท็จ

ผลบวกปลอม (FP) คือ ค่าพยากรณ์ไม่ตรงกับค่าจริง ในกรณีค่าพยากรณ์เป็นจริง แต่ค่าจริง

เป็นเท็จ

ผลลบปลอม (FN) คือ ค่าพยากรณ์ไม่ตรงกับค่าจริง ในกรณีค่าพยากรณ์เป็นเท็จ แต่ค่าจริง

เป็นจริง

ค่าความเที่ยง (Precision) คือ ความแม่นยำในการจำแนกของตัวแบบการจำแนกโดยพิจารณาในแต่ละคลาสคำตอบ ดังสมการ (2.26) และ (2.27)

$$\text{Precision (Yes)} = \frac{TP}{TP+FP} \quad (2.26)$$

$$\text{Precision (No)} = \frac{TN}{TN+FN} \quad (2.27)$$

ค่าเรียกคืน (Recall กลุ่ม Yes) หรือค่าความไว (Sensitivity) คือ สัดส่วนของผลบวกที่เป็นจริงสำหรับภาวะนั้นๆ ดังสมการ (2.28)

$$\text{Recall (Yes) หรือ Sensitivity} = \frac{TP}{TP+FN} \quad (2.28)$$

ค่าเรียกคืน (Recall กลุ่ม No) หรือค่าความจำเพาะ (Specificity) คือ สัดส่วนของผลลบที่เป็นจริงสำหรับภาวะนั้นๆ ดังสมการ (2.29)

$$\text{Recall (No) หรือ Specificity} = \frac{TN}{TN+FP} \quad (2.29)$$

ค่าความแม่นยำ (Accuracy) คือ จำนวนข้อมูลที่ตัวแบบจำแนกได้ถูกต้องต่อข้อมูลทั้งหมด ดังสมการ (2.30)

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN} \quad (2.30)$$

ค่าความแม่นยำที่สมดุลแบบปรับน้ำหนักถ่วง (Weighted Balanced Accuracy) ดังสมการ (2.31)

$$B\text{Acc} = \left(\frac{n_{Yes}}{N} \times \text{Recall}(Yes) \right) + \left(\frac{n_{No}}{N} \times \text{Recall}(No) \right) \quad (2.31)$$

ค่าประสิทธิภาพโดยรวมแบบปรับน้ำหนักถ่วง ($F1_{weight}$) ดังสมการ (2.32)

$$F1_{(weight)} = \frac{n_{Yes}}{N} (F1_{(Yes)}) + \frac{n_{No}}{N} (F1_{(No)}) \quad (2.32)$$

$$\text{โดย } F1_{(Yes)} = \frac{2 \times \text{Recall}(Yes) \times \text{Precision}(Yes)}{\text{Recall}(Yes) + \text{Precision}(Yes)}$$

$$F1_{(No)} = \frac{2 \times \text{Recall}(No) \times \text{Precision}(No)}{\text{Recall}(No) + \text{Precision}(No)}$$

เมื่อ n_{Yes} คือ จำนวนข้อมูลที่มีค่าจริงเป็นจริง (Yes)
 n_{No} คือ จำนวนข้อมูลที่มีค่าจริงเป็นเท็จ (No)
 N คือ จำนวนข้อมูลทั้งหมด

2.7. สัมประสิทธิ์สหสัมพันธ์

การวิเคราะห์สหสัมพันธ์ (Correlation Analysis) เป็นการศึกษาความสัมพันธ์ระหว่างข้อมูลหรือตัวแปรตั้งแต่ 2 ตัวแปรขึ้นไป ว่ามีความสัมพันธ์กันในระดับใด และมีความสัมพันธ์ในทิศทางใด โดยตัวสถิติที่ใช้วัดคือ ค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation coefficient) ซึ่งมีค่าอยู่ในช่วง (-1, 1) โดยที่ +1 หมายถึงข้อมูลหรือตัวแปรมีความสัมพันธ์กันในทิศทางเดียวกันหรือเชิงบวกอย่างสมบูรณ์ ในขณะที่ 0 หมายถึง ข้อมูลหรือตัวแปรไม่มีความสัมพันธ์กัน และ -1 หมายถึงข้อมูลหรือตัวแปรมีความสัมพันธ์กันในทิศทางตรงกันข้ามหรือเชิงลบอย่างสมบูรณ์ (Hair et al., 2010) วิธีการคำนวณค่าสัมประสิทธิ์สหสัมพันธ์มีหลายวิธีขึ้นอยู่กับประเภทของข้อมูลหรือตัวแปร เช่น สัมประสิทธิ์

สหสัมพันธ์สเปียร์แมน สัมประสิทธิ์สหสัมพันธ์พี และสัมประสิทธิ์สหสัมพันธ์เพียร์สัน เป็นต้น โดยในงานวิจัยได้ใช้สัมประสิทธิ์สหสัมพันธ์สเปียร์แมน (Spearman rank correlation coefficient) สำหรับพิจารณาความสัมพันธ์ของข้อมูลหรือตัวแปร คำนวณได้จากสูตรดังต่อไปนี้

$$r_s = \frac{cov(R(X), R(Y))}{\sigma_{R(X)}\sigma_{R(Y)}} \quad (2.33)$$

โดยที่ r_s คือ ค่าสัมประสิทธิ์สหสัมพันธ์สเปียร์แมน

$R(X)$ คือ แรงค์ (Rank) ของตัวแปรอิสระ (X)

$R(Y)$ คือ แรงค์ (Rank) ของตัวแปรตาม (Y)

$\sigma_{R(X)}$ คือ ส่วนเบี่ยงเบนมาตรฐานของแรงค์ตัวแปรอิสระ

$\sigma_{R(Y)}$ คือ ส่วนเบี่ยงเบนมาตรฐานของแรงค์ตัวแปรตาม

การทดสอบความมีนัยสำคัญทางสถิติของสัมประสิทธิ์สหสัมพันธ์สเปียร์แมนทำได้ดังนี้

1. สมมติฐานการทดสอบ H_0 : ตัวแปรทั้งสองไม่มีความสัมพันธ์กัน
 H_1 : ตัวแปรทั้งสองมีความสัมพันธ์กัน

2. ค่าสถิติทดสอบกรณีตัวอย่างขนาดใหญ่เป็นดังนี้

$$Z = r_s\sqrt{n-1} \quad (2.34)$$

เมื่อ n คือ จำนวนคู่ข้อมูล

3. สรุปผลโดยการเปรียบเทียบค่าสถิติทดสอบกับค่าวิกฤต เมื่อกำหนดระดับนัยสำคัญของการทดสอบเท่ากับ α โดยที่

ยอมรับ H_0 เมื่อค่าสถิติทดสอบที่คำนวณได้มีค่าน้อยกว่าค่าวิกฤต ($Z_{\alpha/2}$)

ปฏิเสธ H_0 เมื่อค่าสถิติทดสอบที่คำนวณได้มีค่ามากกว่าหรือเท่ากับค่าวิกฤต ($Z_{\alpha/2}$)

2.8. งานวิจัยที่เกี่ยวข้อง

สุรวุฒ ศรีเปารยะ และ สายชล สิ้นสมบูรณ์ทอง (2560) ได้สร้างแบบจำลองเพื่อเปรียบเทียบประสิทธิภาพวิธีการจำแนกกลุ่ม ข้อมูลที่ใช้เป็นข้อมูลเกี่ยวกับโรคไตเรื้อรังมีจำนวนข้อมูลทั้งหมด 400 แถวจากฐานข้อมูล UCI แบ่งเป็นชุดข้อมูลเรียนรู้ 70% และชุดข้อมูลทดสอบ 30% และใช้วิธีการจำแนกกลุ่ม 7 เทคนิค ได้แก่ เทคนิคเคเนียร์เนสเนเบอร์ (K-nearest Neighbors : K-NN) เทคนิคต้นไม้ตัดสินใจ (Decision Tree : DT) เทคนิคโครงข่ายประสาทเทียม (Artificial Neural Network : ANN) เทคนิคซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine : SVM) เทคนิคการถดถอยลอ-

จิสติก (Logistic Regression : LR) เทคนิคฐานกฎ (Rule -Based) และเทคนิคนาอิวเบย์ (Naïve Bayes : NB) ผลการศึกษาพบว่า วิธีการจำแนกกลุ่มที่มีประสิทธิภาพการจำแนกที่ดีที่สุดคือ เทคนิคต้นไม้ตัดสินใจ โดยมีค่าความแม่นยำ (Accuracy) 100% และค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error) 0.0059

รุ่งโรจน์ บุญมา และ นิเวศ จิระวิจิตชัย (2562) ได้สร้างแบบจำลองการจำแนกประเภทผู้ป่วยโรคเบาหวานโดยใช้เทคนิคเหมือนข้อมูลและการเลือกคุณลักษณะจากความสัมพันธ์ของข้อมูล สำหรับข้อมูลที่ใช้ในการศึกษาคือ กลุ่มผู้ป่วยโรคเบาหวานจำนวน 768 แถวจากฐานข้อมูล UCI และทำการเปรียบเทียบประสิทธิภาพของแบบจำลองของเหมือนข้อมูล 4 ประเภท เทคนิคที่ใช้คือ เทคนิคต้นไม้ตัดสินใจ เทคนิคซัพพอร์ตเวกเตอร์แมชชีน เทคนิคเนอิวเบย์ และเทคนิคเคเนียร์เรนเนอร์ จากผลการศึกษาพบว่าเทคนิคซัพพอร์ตเวกเตอร์แมชชีน มีประสิทธิภาพในการทำนายสูงสุด โดยมีค่าความแม่นยำ (Accuracy) 76.95%

ณัฐวดี หงส์บุญมี และ ประภาสสิริ ตรีพานิชกุล (2562) เปรียบเทียบประสิทธิภาพของการจำแนกข้อมูลเพื่อวิเคราะห์ปัจจัยความเสี่ยงที่ส่งผลต่อการเกิดโรคไฮเปอร์ไทรอยด์ ด้วยเทคนิคเหมือนข้อมูล 3 เทคนิคคือ เทคนิคโครงข่ายประสาทเทียม เทคนิคนาอิวเบย์ และเทคนิคต้นไม้ตัดสินใจ ซึ่งข้อมูลที่น่าสนใจในการทดลองเป็นข้อมูลผู้ป่วยโรคไฮเปอร์ไทรอยด์จากโรงพยาบาลพิษณุโลกจำนวน 323 แถว มีจำนวน 12 ปัจจัย โดยแบ่งชุดข้อมูลทดสอบแบบ k – fold Cross Validation ผลการเปรียบเทียบประสิทธิภาพการจำแนกข้อมูลพบว่าเทคนิคโครงข่ายประสาทเทียม มีค่าความแม่นยำ 82.97% ซึ่งมากกว่าเทคนิคต้นไม้ตัดสินใจ และเทคนิคนาอิวเบย์ที่มีค่าความแม่นยำ 79.87% และ 68.11%ตามลำดับ และพบว่าปัจจัยที่มีความสำคัญคือ เพศ อารมณ์แปรปรวนและเหนื่อยง่าย

อุกฤษฏ์ ศรีสุข และ จาริ ทองคำ (2563) ศึกษาประสิทธิภาพของเทคนิคเหมือนข้อมูลที่หลากหลาย ข้อมูลที่ใช้ในการศึกษาวิจัยเป็นข้อมูลผู้ป่วยโรคมะเร็งเต้านม ผู้ป่วยโรคเบาหวาน และผู้ป่วยโรคไฮเปอร์ไทรอยด์ จากฐานข้อมูล UCI เทคนิคเหมือนข้อมูลที่ใช้ศึกษามี 5 เทคนิค ได้แก่ เทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม C4.5 เทคนิคนาอิวเบย์ เทคนิคโครงข่ายประสาทเทียม เทคนิคป่าสุ่ม และเทคนิคการเรียนรู้เชิงลึก (Deep Learning : DL) ทดสอบประเมินประสิทธิภาพของแบบจำลองด้วยวิธี 10 – Fold Cross Validation จากการทดลองพบว่าเทคนิคต้นไม้ตัดสินใจ C4.5 เป็นเทคนิคเหมาะสมในชุดข้อมูลผู้ป่วยโรคไฮเปอร์ไทรอยด์ โดยมีค่าความแม่นยำ 99.86% ค่าความไว 99.85% และค่าความจำเพาะ 100% ชุดข้อมูลผู้ป่วยโรคเบาหวานเทคนิคที่เหมาะสมคือ เทคนิคการเรียนรู้เชิงลึก มีค่าความแม่นยำ 77.47% ค่าความไว 87.40% และค่าความจำเพาะ 58.96% และชุดข้อมูลผู้ป่วยโรคมะเร็งเต้านมเทคนิคที่เหมาะสมคือ เทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม C4.5 มีค่าความแม่นยำ 75.52% ค่าความไว 96.02% และค่าความจำเพาะ 27.05% สรุปได้ว่าเทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม C4.5 มีความเหมาะสมในการนำมาสร้างแบบจำลองเพื่อพยากรณ์โรคไฮเปอร์

ไทรอยด์ โรคเบาหวาน และโรคมะเร็งเต้านม เนื่องจากผลการพยากรณ์ข้อมูลมีความแม่นยำค่อนข้างสูง

นงเยาว์ ในอรุณ (2564) ได้สร้างแบบจำลองการทำนายความเสี่ยงโรคหัวใจและหลอดเลือด โดยใช้อัลกอริทึมเหมือนข้อมูล ได้แก่ เทคนิคโครงข่ายประสาทเทียม เทคนิคป่าสุ่ม (Random Forest : RF) เทคนิคเคเนียร์สเนเบอร์ เทคนิคนาอ็ฟเบย์ และเทคนิคต้นไม้ตัดสินใจ พร้อมการเลือกคุณสมบัติของแอตทริบิวต์ และเปรียบเทียบประสิทธิภาพของแบบจำลองด้วยวิธี 10 – Fold Cross Validation ข้อมูลที่ใช้เป็นข้อมูลผู้ป่วยโรคหัวใจและหลอดเลือดมาจากสำนักงานสาธารณสุขจังหวัดสระบุรีจำนวน 31,929 แถว ผลการวิจัยพบว่า แบบจำลองที่มีประสิทธิภาพการทำนายสูงสุดคือ แบบจำลองจากเทคนิคโครงข่ายประสาทเทียมพร้อมการเลือกคุณสมบัติ มีความแม่นยำ 99.29% ควรนำไปพัฒนาเว็บแอปพลิเคชันเพื่อการวินิจฉัยโรคเบื้องต้นเกี่ยวกับความเสี่ยงโรคหัวใจและหลอดเลือด

รัชนิวรรณ ไพศาลวรกิจ (2564) ได้วิเคราะห์ปัจจัยที่มีผลต่อโรคเบาหวานด้วยเทคนิคการถดถอยลอจิสติก และเปรียบเทียบตัวแบบการจำแนกด้วยเทคนิคการถดถอยลอจิสติก เทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม J48 และ LMT และเทคนิคโครงข่ายประสาทเทียม ข้อมูลที่ใช้เป็นข้อมูลผู้ป่วยที่เข้ารับบริการในโรงพยาบาลมหาวิทยาลัยนเรศวรจำนวน 5,081 ชุด โดยแบ่งการทดสอบเป็นชุดข้อมูลเรียนรู้และชุดข้อมูลทดสอบ คือ 70:30 และ 80:20 ผลการศึกษาพบว่า เทคนิคโครงข่ายประสาทเทียมมีประสิทธิภาพสูงสุด มีความแม่นยำเท่ากับ 81.78% (70:30) และ 81.42% (80:20) รองลงมาคือเทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม J48 และ LMT และเทคนิคการถดถอยลอจิสติกตามลำดับ

ปพนันศรณ์ สิวสำแดงเดช (2565) ได้ประยุกต์ใช้เทคนิคการทำเหมืองข้อมูลเพื่อพยากรณ์ผู้ป่วยโรคเบาหวาน และเปรียบเทียบประสิทธิภาพแบบจำลองที่เหมาะสมที่สุดสำหรับการพยากรณ์ผู้ป่วยโรคเบาหวานด้วยเทคนิคต้นไม้ตัดสินใจ เทคนิคนาอ็ฟเบย์ เทคนิคโหวตรวม (Vote Ensemble) และเทคนิคป่าสุ่ม ข้อมูลที่ใช้เป็นข้อมูลผู้ป่วยโรคเบาหวานจากระบบฐานข้อมูลโรงพยาบาลศูนย์อุดรธานี โดยแบ่งการทดสอบแบบ 10 – Fold Cross Validation ผลของการวิจัยพบว่า เทคนิคป่าสุ่ม ให้ค่าความแม่นยำในการทำนายผลเป็นโรคเบาหวานมากที่สุดคือ 88.03% มีค่าความเที่ยง 88.22% ค่าเรียกคืน (Recall) 90.36% และค่าประสิทธิภาพโดยรวม (F-Measure) 89.28%

วรสิทธิ์พล แสงทองรัตนโชติ (2565) ศึกษาเปรียบเทียบเทคนิคการเรียนรู้ของเครื่องเพื่อสร้างตัวแบบการจำแนก ด้วยการปรับปรุงข้อมูลสมมูลกับชุดข้อมูลที่มีจำนวนของตัวแปรอิสระเชิงคุณภาพและเชิงปริมาณแตกต่างกันทั้งหมด 3 ชุดข้อมูล โดยใช้เทคนิคการวิเคราะห์จำแนกกลุ่มเชิงเส้นโดยวิธีของฟิชเชอร์ (Fisher's linear Discriminant Analysis : LDA) เทคนิคนาอ็ฟเบย์ เทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม C4.5 เทคนิคป่าสุ่มและเทคนิคโครงข่ายประสาทเทียม และทดสอบประสิทธิภาพด้วย 5-Fold Cross-Validation พบว่า เทคนิคป่าสุ่มทำงานได้ดีภายใต้ชุดข้อมูลที่มี

จำนวนตัวแปรอิสระเชิงปริมาณเท่ากับเชิงคุณภาพ เทคนิคการจำแนกการวิเคราะห์จำแนกกลุ่มเชิงเส้น โดยวิธีของพีชเซอร์ทำงานได้ดีภายใต้ชุดข้อมูลที่มีตัวแปรอิสระเชิงปริมาณทุกตัว และเทคนิคโครงข่ายประสาทเทียมทำงานได้ดีภายใต้ชุดข้อมูลที่มีจำนวนตัวแปรอิสระเชิงปริมาณมากกว่าเชิงคุณภาพ และพบว่า การปรับปรุงข้อมูลให้สมดุลด้วยการสุ่มตัวอย่างแบบง่ายให้ประสิทธิภาพดีกว่าการแบ่งกลุ่มข้อมูลแบบเคมีน

Assari et al. (2017) ศึกษาพยากรณ์การเกิดโรคหัวใจโดยการนำไปประยุกต์ใช้กับเทคนิคเหมืองข้อมูล 4 เทคนิคคือ เทคนิคต้นไม้ตัดสินใจ เทคนิคนาอูฟเบย์ เทคนิคเคเนียร์เซนเบอร์ และเทคนิคซัพพอร์ตเวกเตอร์แมชชีน โดยได้อธิบายว่าโรคหัวใจเป็นสาเหตุสำคัญของการเสียชีวิตทั่วโลก แต่ก็ยังเป็นโรคที่สามารถรักษาและป้องกันได้ และเชื่อว่าการวินิจฉัยโรคหัวใจในระยะแรก และทันเวลา จะช่วยป้องกันได้ ข้อมูลที่ใช้ในการศึกษามีจำนวน 303 แถว ผลการศึกษาพบว่าเทคนิคซัพพอร์ตเวกเตอร์แมชชีนมีค่าความความแม่นยำ 84.33% ซึ่งมีมากกว่าเทคนิคนาอูฟเบย์ที่มีค่าความแม่นยำ 83.66% เทคนิคเคเนียร์เซนเบอร์ที่มีค่าความแม่นยำ 81% และเทคนิคต้นไม้ตัดสินใจที่มีค่าความแม่นยำ 79%

Olivera et al. (2017) เปรียบเทียบประสิทธิภาพของเทคนิคเหมืองข้อมูลต่างๆ โดยใช้ข้อมูลโรคเบาหวานจาก ELSA-Brasil เทคนิคที่ใช้คือ เทคนิคการถดถอยลอจิสติก เทคนิคโครงข่ายประสาทเทียม เทคนิคนาอูฟเบย์ เทคนิคเคเนียร์เซนเบอร์ และเทคนิคป่าสุ่ม และคัดเลือกตัวแปรด้วยวิธีการเลือกแบบไปข้างหน้า (Forward Selection) และ Wrapper Strategy จากนั้นทดสอบประสิทธิภาพโดยใช้ 10-Fold Cross-Validation พบว่า เทคนิคเคเนียร์เซนเบอร์ และเทคนิคการถดถอยลอจิสติก เป็นเทคนิคที่ดีที่สุดมีพื้นที่เฉลี่ยภายใต้เส้นโค้งเท่ากับ 75.24% และ 74.98% ตามลำดับ

Astuti et al. (2018) ได้ศึกษาเปรียบเทียบวิธีแบ่งข้อมูล ได้แก่ Random Repeated Holdout, Stratified Repeated Holdout, Random Cross-Validation, และ Stratified Cross-Validation ในชุดข้อมูลโรคเบาหวาน โดยใช้เทคนิคการถดถอยลอจิสติก และเทคนิคซัพพอร์ตเวกเตอร์แมชชีน ช่วยในการวิเคราะห์ประเมินผล และคัดเลือกตัวแปรด้วยวิธีการเลือกแบบไปข้างหน้า (Forward Selection) และวิธีกำจัดแบบถดถอยหลัง (Backward Elimination) พบว่าการคัดเลือกตัวแปรช่วยให้เทคนิคการถดถอยลอจิสติกมีประสิทธิภาพเพิ่มขึ้น แต่ตรงกันข้ามกับเทคนิคซัพพอร์ตเวกเตอร์แมชชีน และไม่พบการวิเคราะห์ที่ชัดเจนว่าวิธีการแบ่งข้อมูลแบบไหนดีกว่ากัน

Ali et al. (2020) ศึกษาเปรียบเทียบการจำแนกประเภทมะเร็งเต้านม โดยใช้เทคนิคซัพพอร์ตเวกเตอร์แมชชีน และเทคนิคการถดถอยลอจิสติก โดยแบ่งเป็นชุดข้อมูลเรียนรู้ (Training Data) 80% และชุดข้อมูลทดสอบ (Test Data) 20% ผลการวิจัยพบว่าเทคนิคการถดถอยลอจิสติกมีค่าความแม่นยำ 98.61% ซึ่งมากกว่าเทคนิคซัพพอร์ตเวกเตอร์แมชชีน

Aroef et al. (2020) ได้ศึกษาเปรียบเทียบเทคนิคป่าสุ่ม และเทคนิคซัพพอร์ตเวกเตอร์-

แมชชีน สำหรับการจำแนกประเภทข้อมูลโรคมะเร็งเต้านม โดยได้อธิบายว่ามะเร็งมีมากกว่า 100 ชนิดทั่วโลก หนึ่งในนั้นมะเร็งที่พบบ่อยคือ มะเร็งเต้านม โดยข้อมูลจากกระทรวงสาธารณสุขระบุว่า ในปี ค.ศ. 2018 ในอินโดนีเซียพบผู้ป่วยโรคมะเร็งเต้านม 42 รายจาก 100,000 ราย เสียชีวิตประมาณ 17 ราย และการศึกษานี้ได้นำ Boruta มาใช้ในการเลือกลักษณะที่สำคัญที่สุดในการสร้างแบบจำลอง ผลการวิจัยพบว่าเทคนิคซัพพอร์ตเวกเตอร์แมชชีน มีค่าความแม่นยำ 95.45% และเทคนิคป่าสุ่มมีค่าความแม่นยำ 90.9%

Nai-arun and Moungrmai (2020) นำเสนอแบบจำลองเพื่อพยากรณ์ความเสี่ยงของโรคหัวใจและหลอดเลือดโดยใช้เทคนิคการทำเหมืองข้อมูล และทำการแบ่งข้อมูลแบบ 10 – Fold Cross Validation พบว่าเทคนิคการถดถอยลอจิสติกมีค่าความแม่นยำ 99.94% ซึ่งมากกว่าเทคนิคโครงข่ายประสาทเทียมแบบแพร่ย้อนหลัง (Back-propagation Neural Network) เทคนิคป่าสุ่ม เทคนิคต้นไม้ตัดสินใจ เทคนิคเคเนียร์สเนเบอร์ และเทคนิคนาอิวเบย์

Dissanayake and Johar (2021) ได้เปรียบเทียบการทำนายโรคหัวใจโดยใช้เทคนิคการเลือกคุณลักษณะ 10 เทคนิค ได้แก่ ANOVA, Chi-square, mutual information, ReliefF, forward feature selection, backward feature selection, exhaustive feature selection, recursive feature elimination, Lasso regression, และ Ridge regression และเทคนิคการจำแนก 6 เทคนิค ได้แก่ เทคนิคต้นไม้ตัดสินใจ เทคนิคป่าสุ่ม เทคนิคซัพพอร์ตเวกเตอร์แมชชีน เทคนิคเคเนียร์สเนเบอร์ เทคนิคการถดถอยลอจิสติก และเทคนิคเกาส์เซียนนาอิวเบย์ (Gaussian naïve Bayes : GNB) พบว่าเทคนิคต้นไม้ตัดสินใจที่เลือกคุณลักษณะด้วยวิธีการกำจัดแบบถดถอยหลัง มีประสิทธิภาพมากที่สุด และให้ค่าความแม่นยำเท่ากับ 88.52%

Dritsas and Trigka (2022) ได้ศึกษาการทำงานและเปรียบเทียบประสิทธิภาพของเทคนิคการจำแนกในชุดข้อมูลโรคเบาหวาน ประกอบด้วย เทคนิคนาอิวเบย์ เทคนิคเครือข่ายแบบเบย์ (Bayesian Network) เทคนิคซัพพอร์ตเวกเตอร์แมชชีน เทคนิคการถดถอยลอจิสติก เทคนิคโครงข่ายประสาทเทียม เทคนิคเคเนียร์สเนเบอร์ เทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม J48 เทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม LMT เทคนิค REPT เทคนิค RT เทคนิค RotF เทคนิคป่าสุ่ม เทคนิคเอดาบูทเอ็มวัน (AdaBoost M1) เทคนิค SGD และเทคนิค Stacking ทดสอบประสิทธิภาพของตัวแบบการจำแนกโดยใช้ 10-Fold Cross-Validation และแบ่งชุดข้อมูลเรียนรู้ 80% และชุดข้อมูลทดสอบ 20% พบว่าเทคนิคป่าสุ่ม และเทคนิคเคเนียร์สเนเบอร์ เป็นตัวแบบที่มีประสิทธิภาพดีที่สุด

Fridausanti et al. (2022) ได้เปรียบเทียบเทคนิคการถดถอยลอจิสติกกับเทคนิคซัพพอร์ตเวกเตอร์แมชชีน ในชุดข้อมูลคลื่นไฟฟ้าหัวใจ โดยคัดเลือกคุณลักษณะด้วยวิธีการกำจัดแบบถดถอยหลัง และแบ่งข้อมูลเรียนรู้ ข้อมูลทดสอบด้วยวิธีการตรวจสอบไขว้ (Cross Validation) และวิธี holdout

พบว่า เทคนิคซ์พอร์ตเวกเตอร์แมชชีนที่แบ่งข้อมูลด้วย unstratified holdout มีประสิทธิภาพมากที่สุด มีค่าความแม่นยำเท่ากับ 91.54%



ตาราง 3 สรุปงานวิจัยที่เกี่ยวข้องที่ประยุกต์ใช้วิธีการทางสถิติและเทคนิคเหมืองข้อมูลในการจำแนกข้อมูลด้านสุขภาพ

งานวิจัย	ชุดข้อมูล	การแบ่งข้อมูล	เทคนิคที่ใช้	คัดเลือกตัวแปร		เทคนิคที่เหมาะสม
				อัตรา	เชิงปริมาณ	
สุวัชร ศรีเปรายะ และ สายชล สินสมบูรณ์ทอง (2560)	โรคไตเรื้อรัง	ข้อมูล 400 ชุด แบ่งเป็นชุด เรียนรู้ 70% และชุดทดสอบ 30%	K-NN, DT, SVM (polynomial Kernel), LR, NB, Rule- Based, ANN	ไม่มีการคัดเลือก ตัวแปรอิสระ	13	11 DT (ค่าความแม่นยำ 100%, ค่า ความคลาดเคลื่อน กำลังสองเฉลี่ย 0.0059)
รุ่งโรจน์ บุญมา และนิเวศ จิระ- วิจิตชัย (2562)	โรคเบาหวาน	ข้อมูล 768 ชุด	DT, SVM, K-NN, NB	Correlation Based Feature Selection (CFS)	-	8 SVM (ค่าความแม่นยำ 76.95%)
ณัฐวดี หงส์บุญมี และ ประภาสิริ ตรีพานิชกุล (2562)	โรคไฮเปอร์ ไทรอยด์	ข้อมูล 323 ชุด แบ่งการทดสอบ แบบ 2-10 Fold Cross- Validation	ANN, NB, DT	ไม่มีการคัดเลือก ตัวแปรอิสระ	12	- ANN (ค่าความแม่นยำ 82.97%)

งานวิจัย	ชุดข้อมูล	การแบ่งข้อมูล	เทคนิคที่ใช้	คัดเลือกตัวแปร		ลักษณะตัวแปร		เทคนิคที่เหมาะสม
				อิสระ	เชิงคุณภาพ	เชิงปริมาณ	เชิงปริมาณ	
อุกฤษณ์ ศรีสุข และจารี ทองคำ (2563)	โรคมะเร็งเต้านม	ข้อมูลโรคมะเร็งเต้านม 286 ชุด	DT, NB, ANN, RF, DL	Chi-Square	โรคมะเร็งเต้านม 9, โรคมะเร็งเต้านม 0, โรคมะเร็งเต้านม	โรคมะเร็งเต้านม 9, โรคมะเร็งเต้านม 0, โรคมะเร็งเต้านม	DT (โรคมะเร็งเต้านม)	DT (โรคมะเร็งเต้านม)
	โรคมะเร็งเต้านม	ข้อมูลโรคมะเร็งเต้านม 768 ชุด และ โรคมะเร็งเต้านม 3,772 ชุด แบ่ง การทดสอบแบบ 10-Fold Cross Validation			โรคมะเร็งเต้านม 23, โรคมะเร็งเต้านม 0	โรคมะเร็งเต้านม 23, โรคมะเร็งเต้านม 0	โรคมะเร็งเต้านม 6, โรคมะเร็งเต้านม 8	โรคมะเร็งเต้านม 6, โรคมะเร็งเต้านม 8
นงเยาว์ ใสใจและ ใสใจ (2564)	โรคหัวใจและหลอดเลือด	ชุดแบ่งการทดสอบแบบ	NB, ANN, RF, K-NN, DT	ไม่มีการคัดเลือกตัวแปรอิสระ	3	5	ANN (ค่าความแม่นยำ 99.29%)	ANN (ค่าความแม่นยำ 99.29%)

งานวิจัย	ชุดข้อมูล	การแบ่งข้อมูล	เทคนิคที่ใช้	คัดเลือกตัวแปร		ลักษณะตัวแปร		เทคนิคที่เหมาะสม
				อิสระ	เชิงคุณภาพ	เชิงปริมาณ	เชิงปริมาณ	
รัชนิวรรณ ไพศาลวร- เกียรติ (2564)	โรคเบาหวาน	10-Fold Cross Validation	LR, DT, ANN	Full, Forward Selection, Backward elimination, Stepwise Regression	1	12	ANN (ค่าความแม่นยำ ที่สมมูล 73.97%)	
		ข้อมูล 5,081 ชุดแบ่งการ ทดสอบเป็นชุด เรียนรู้ 70% กับ 80% และชุด ทดสอบ 30% กับ 20%						
ปพนรัตน์ ลีวถ้ำแดงเดช (2565)	โรคเบาหวาน	ข้อมูล 70,421 ชุด และข้อมูล 80,201 ชุด แบ่ง การทดสอบแบบ 10-Fold Cross Validation	DT, NB, K-NN, RF, Vote Ensemble	ไม่มีการคัดเลือก ตัวแปรอิสระ	5	8	RF (ค่าความแม่นยำ 88.03%)	

งานวิจัย	ชุดข้อมูล	การแบ่งข้อมูล	เทคนิคที่ใช้	คัดเลือกตัวแปร		ลักษณะตัวแปร		เทคนิคที่เหมาะสม
				อิสระ	เชิงคุณภาพ	เชิงปริมาณ	เชิงปริมาณ	
วริทธิพล แสงทองรัตน์- โชติ (2565)	สถาบันการเงิน, สายพันธุ์ข้าว, นักวิทยาศาสตร์ ข้อมูล	สถาบันการเงิน 41,188 ชุด สายพันธุ์ข้าว 3,810 ชุด นักวิทยาศาสตร์ 8,955 ชุด แบ่งการ ทดสอบแบบ 5- Fold Cross Validation	LDA, NB, DT, RF, ANN	ไม่มีการคัดเลือก ตัวแปรอิสระ	สถาบันการเงิน 10 สายพันธุ์ข้าว 0 นักวิทยาศาสตร์ 9	สถาบันการเงิน 10 สายพันธุ์ข้าว 7 นักวิทยาศาสตร์ 1	RF (สถาบันการเงิน) LDA (สายพันธุ์ข้าว) ANN (นักวิทยาศาสตร์ ข้อมูล)	
Assari et al. (2017)	โรคหัวใจ	ข้อมูล 303 ชุด แบ่งการทดสอบ แบบ 10-Fold Cross Validation	DT, NB, K-NN, SVM	ไม่มีการคัดเลือก ตัวแปรอิสระ	8	5	SVM (ค่าความแม่นยำ 84.33%)	

งานวิจัย	ชุดข้อมูล	การแบ่งข้อมูล	เทคนิคที่ใช้	คัดเลือกตัวแปร		ลักษณะตัวแปร		เทคนิคที่เหมาะสม
				อิสระ	เชิงคุณภาพ	เชิงปริมาณ	เชิงปริมาณ	
Olivera et al. (2017)	โรคเบาหวาน	10-Fold Cross Validation	LR, ANN, NB, K-NN, RF	Forward Selection, Wrapper Strategy	24	4	ANN, LR (พื้นที่เฉลี่ยภายใต้เส้นโค้ง 75.24%, 74.98%)	
Astuti et al. (2018)	โรคเบาหวาน	ข้อมูล 768 ชุด แบ่งการทดสอบ แบบ 5-Fold Cross Validation และ holdout	LR, SVM (linear, polynomial, RBF)	Full, Forward Selection, Backward Selection	-	8	LR (ค่าความแม่นยำ 80.5%)	
Aroef et al. (2020)	โรคมะเร็งเต้านม	ข้อมูล 116 ชุด	RF, SVM (RBF)	Boruta	-	9	SVM (ค่าความแม่นยำ 95.45%)	
Ali et al. (2020)	โรคมะเร็งเต้านม	ข้อมูล 286 ชุด แบ่งเป็นชุด	SVM, LR	Full, Boruta, LASSO	-	4	LR (ค่าความแม่นยำ 98.61%)	

งานวิจัย	ชุดข้อมูล	การแบ่งข้อมูล	เทคนิคที่ใช้	คัดเลือกตัวแปร		ลักษณะตัวแปร		เทคนิคที่เหมาะสม
				อิสระ	เชิงคุณภาพ	เชิงปริมาณ	เชิงปริมาณ	
		เรียนรู้ 80% และชุดทดสอบ 20%						SVM (ค่าความแม่นยำ 69.44%)
Nai-arun and Moungmai (2020)	โรคหัวใจและหลอดเลือด	ข้อมูล 44,674 ชุด แบ่งการทดสอบแบบ 10-Fold Cross Validation	LR, Back-propagation, Neural Network, RF, DT, K-NN, NB	เลือกด้วยโปรแกรม WEKA	3	5		LR (ค่าความแม่นยำ 99.94%)
Dissanayake and Johar (2021)	โรคหัวใจ	ข้อมูล 303 ชุด	RF, SVM (linear), DT, KNIN, LR, GNB	Full, ANOVA, Chi-Square, MI, Relief, FFS, BFS, EFS, RFE, Lasso, Ridge	8	5		DT (ค่าความแม่นยำ 88.52%)

งานวิจัย	ชุดข้อมูล	การแบ่งข้อมูล	เทคนิคที่ใช้	คัดเลือกตัวแปรอิสระ	ลักษณะตัวแปร		เทคนิคที่เหมาะสม
					เชิงคุณภาพ	เชิงปริมาณ	
Dritsas and Trigka (2022)	ความเสี่ยงโรคเบาหวาน	ข้อมูล 520 ชุด แบ่งเป็นชุดเรียนรู้ 80% ชุดทดสอบ 20% และ 10-Fold Cross Validation	Bayesian Network, NB, SVM, LR, ANIN, K-NN, J48, LMT, RF, RT, REPT, RotF, AdaBoostM1, SGD, Stacking	ไม่มีการคัดเลือกตัวแปรอิสระ	15	1	RF, K-NN (10-Fold Cross Validation), KNN (80:20)
Firdausanti et al. (2022)	คลื่นไฟฟ้าหัวใจ	ข้อมูล 132 ชุด แบ่งการทดสอบแบบ 5-Fold Cross Validation และ holdout (80:20)	LR, SVM (linear)	Backward elimination	-	7	SVM (ค่าความแม่นยำ 91.54%)

บทที่ 3

วิธีการดำเนินการวิจัย

ในการศึกษาวิจัยในครั้งนี้เพื่อเปรียบเทียบประสิทธิภาพตัวแบบการจำแนกข้อมูลด้านสุขภาพ ด้วยวิธีการทางสถิติและวิธีการทางเทคนิคเหมืองข้อมูล โดยมีขั้นตอนการดำเนินงานวิจัยดังต่อไปนี้

3.1. ข้อมูลที่ใช้ในการศึกษา

ในงานวิจัยนี้ผู้วิจัยคัดเลือกชุดข้อมูลด้านสุขภาพทั้งหมด 3 ชุดข้อมูล ได้แก่ ชุดข้อมูลโรคหัวใจ และหลอดเลือด ชุดข้อมูลโรคเบาหวาน และชุดข้อมูลการเลิกสูบบุหรี่ จากฐานข้อมูล Kaggle โดยแต่ละชุดข้อมูลมีลักษณะโดยสรุปดังตารางที่ 4

1. ข้อมูลชุดที่ 1 คือ ข้อมูลโรคหัวใจและหลอดเลือด มีจำนวนข้อมูลทั้งหมด 70,000 คน ประกอบด้วยตัวแปรทั้งหมด 12 ตัวแปร เป็นตัวแปรอิสระ 11 ตัวแปร แบ่งเป็นตัวแปรอิสระเชิงปริมาณ 5 ตัวแปรและตัวแปรอิสระเชิงคุณภาพ 6 ตัวแปร และตัวแปรตามเชิงกลุ่ม 1 ตัวแปร ประกอบด้วย 2 กลุ่ม คือ กลุ่มผู้ป่วยที่เป็นโรคหัวใจและหลอดเลือดจำนวน 34,979 คน (แทนค่ากลุ่มด้วย Yes) และกลุ่มผู้ป่วยที่ไม่เป็นโรคหัวใจและหลอดเลือดจำนวน 35,021 คน (แทนค่ากลุ่มด้วย No)

2. ข้อมูลชุดที่ 2 คือ ข้อมูลโรคเบาหวาน มีจำนวนข้อมูลทั้งหมด 70,692 คน ประกอบด้วยตัวแปรทั้งหมด 22 ตัวแปร เป็นตัวแปรอิสระ 21 ตัวแปร แบ่งเป็นตัวแปรอิสระเชิงปริมาณ 3 ตัวแปรและตัวแปรอิสระเชิงคุณภาพ 18 ตัวแปร และตัวแปรตามเชิงกลุ่ม 1 ตัวแปร ประกอบด้วย 2 กลุ่ม คือ กลุ่มผู้ป่วยที่เป็นโรคเบาหวานจำนวน 35,346 คน (แทนค่ากลุ่มด้วย Yes) และกลุ่มผู้ป่วยที่ไม่เป็นโรคเบาหวานจำนวน 35,346 คน (แทนค่ากลุ่มด้วย No)

3. ข้อมูลชุดที่ 3 คือ ข้อมูลการเลิกสูบบุหรี่ มีจำนวนข้อมูลทั้งหมด 38,984 คน ประกอบด้วยตัวแปรทั้งหมด 23 ตัวแปร เป็นตัวแปรอิสระ 22 ตัวแปร แบ่งเป็นตัวแปรอิสระเชิงปริมาณ 18 ตัวแปรและตัวแปรอิสระเชิงคุณภาพ 4 ตัวแปร และตัวแปรตามเชิงกลุ่ม 1 ตัวแปร ประกอบด้วย 2 กลุ่ม คือ กลุ่มผู้ที่เลิกสูบบุหรี่จำนวน 14,318 คน (แทนค่ากลุ่มด้วย Yes) และกลุ่มผู้ที่ไม่เลิกสูบบุหรี่จำนวน 24,666 คน (แทนค่ากลุ่มด้วย No)

ตาราง 4 คุณลักษณะและจำนวนข้อมูลของแต่ละชุดข้อมูลด้านสุขภาพจากฐานข้อมูล Kaggle

ข้อมูลชุดที่	จำนวนข้อมูล	จำนวนตัวแปรอิสระ		ค่าตัวแปรตาม		สัดส่วนกลุ่มข้อมูลตัวแปรตาม
		เชิงคุณภาพ	เชิงปริมาณ	Yes	No	
1. โรคหัวใจและหลอดเลือด	70,000	6	5	34,979	35,021	1:1.001
2. โรคเบาหวาน	70,692	18	3	35,346	35,346	1:1
3. การเลิกสูบบุหรี่	38,984	4	18	14,318	24,666	1:1.723

จากการศึกษาพบว่าภายในชุดข้อมูลด้านสุขภาพทั้ง 3 ชุดมีข้อมูลผิดปกติจากการบันทึกข้อมูลสูญหาย (Missing Data) และข้อมูลที่มีความซ้ำซ้อนกัน ดังนั้นผู้วิจัยจึงได้ตัดข้อมูลเหล่านั้นออกทำให้ชุดข้อมูลทั้ง 3 ชุด มีจำนวนข้อมูลแสดงดังตารางที่ 5

ตาราง 5 คุณลักษณะและจำนวนข้อมูลที่ใช้ในการศึกษาหลังการตัดข้อมูลบางส่วนทิ้ง

ข้อมูลชุดที่	จำนวนข้อมูล	จำนวนตัวแปรอิสระ		ค่าตัวแปรตาม		สัดส่วนกลุ่มข้อมูลตัวแปรตาม
		เชิงคุณภาพ	เชิงปริมาณ	Yes	No	
1. โรคหัวใจและหลอดเลือด	68,433	6	5	33,830	34,603	1:1.023
2. โรคเบาหวาน	68,134	18	3	34,394	33,740	1.019:1
3. การเลิกสูบบุหรี่	27,285	4	18	9,173	18,112	1:1.975

3.2. เครื่องมือที่ใช้ในงานวิจัย

การพัฒนาตัวแบบการจำแนกด้วยเทคนิคต่างๆ ในชุดข้อมูลทั้ง 3 ชุด เพื่อเปรียบเทียบประสิทธิภาพของตัวแบบการจำแนกผู้วิจัยได้ใช้เครื่องมือในงานวิจัย ดังต่อไปนี้

1. ด้านคอมพิวเตอร์

CPU: 2.20 GHz, RAM: 8.00 GB

2. ด้านโปรแกรมคอมพิวเตอร์ หรือ ซอฟต์แวร์

2.1. โปรแกรม Microsoft Excel เพื่อใช้ในการจัดการข้อมูลเบื้องต้น

2.2. โปรแกรม SPSS เพื่อใช้ในการหาความสัมพันธ์ของข้อมูล

2.3. โปรแกรม RStudio Version 4.1.0 เพื่อใช้ในการพัฒนาตัวแบบการจำแนก

2.3.1. พัฒนาตัวแบบการจำแนกด้วยเทคนิคการถดถอยลอจิสติกโดยแพ็คเกจที่ใช้คือ `blorr` และใช้ฟังก์ชัน `glm()`

2.3.2. พัฒนาตัวแบบการจำแนกด้วยเทคนิคต้นไม้ตัดสินใจ โดยแพ็คเกจที่ใช้คือ `RWeka` และใช้ฟังก์ชัน `J48()`

2.3.3. พัฒนาตัวแบบการจำแนกด้วยเทคนิคป่าสุ่ม โดยแพ็คเกจที่ใช้คือ `randomForest` และใช้ฟังก์ชัน `randomForest()`

2.3.4. พัฒนาตัวแบบการจำแนกด้วยเทคนิคเครื่องเวกเตอร์ค้ำยันโดยใช้แพ็คเกจ `e1071` และใช้ฟังก์ชัน `svm()`

3.3. ขั้นตอนการเตรียมข้อมูลสำหรับการวิเคราะห์

ในส่วนนี้จะเป็นการรายงานการวิเคราะห์และจัดเตรียมข้อมูลโดยแบ่งการรายงานเป็น 3 ประเด็น

ประเด็นที่ 1 ศึกษารายละเอียดของแต่ละตัวแปรในชุดข้อมูล

ประเด็นที่ 2 รายงานการวิเคราะห์ความสัมพันธ์ของตัวแปรอิสระกับตัวแปรตามด้วยสัมประสิทธิ์สหสัมพันธ์สเปียร์แมน (Spearman rank correlation coefficient)

ประเด็นที่ 3 แสดงการแปลงข้อมูลเชิงปริมาณก่อนนำไปสร้างตัวแบบการจำแนก โดยทำการแปลงค่าตัวแปรให้อยู่ในรูปปกติ (Normalization) ซึ่งมีค่าอยู่ในช่วง $[0,1]$ คำนวณได้จากสูตรดังต่อไปนี้

$$X^* = \frac{X - \min(X)}{\max(X) - \min(X)} \quad (3.2)$$

โดยที่ X^* คือ ค่าที่ได้หลังการแปลงค่าให้อยู่ในรูปปกติ

X คือ ค่าข้อมูลที่ต้องการแปลง

$\min(X)$ คือ ค่าข้อมูลที่น้อยที่สุดในตัวแปร

$\max(X)$ คือ ค่าข้อมูลที่มากที่สุดในตัวแปร

3.3.1. ข้อมูลชุดที่ 1 โรคหัวใจและหลอดเลือด

ในข้อมูลชุดที่ 1 แบ่งการรายงานการวิเคราะห์และจัดเตรียมข้อมูลออกเป็น 3 ประเด็น โดยมีรายละเอียดดังนี้

ประเด็นที่ 1 รายละเอียดของตัวแปรในข้อมูลชุดที่ 1

ตาราง 6 รายละเอียดตัวแปรข้อมูลชุดที่ 1

ตัวแปร	ความหมาย	ลักษณะตัวแปร
age	อายุ (Day)	เชิงปริมาณ
gender	เพศ	เชิงคุณภาพ
height	ส่วนสูง (cm.)	เชิงปริมาณ
weight	น้ำหนัก (kg.)	เชิงปริมาณ
ap_hi	ค่าความดันขณะหัวใจบีบตัว (mmHg.)	เชิงปริมาณ
ap_lo	ค่าความดันขณะหัวใจคลายตัว (mmHg.)	เชิงปริมาณ
cholesterol	ระดับคอเลสเตอรอลแบ่งเป็น 3 ระดับ	เชิงคุณภาพ
gluc	ระดับน้ำตาลกลูโคสแบ่งเป็น 3 ระดับ	เชิงคุณภาพ
smoke	การสูบบุหรี่	เชิงคุณภาพ
alco	การดื่มแอลกอฮอล์	เชิงคุณภาพ
active	การออกกำลังกาย	เชิงคุณภาพ
class	การเป็นโรคหัวใจและหลอดเลือด	เชิงคุณภาพ

โดยรายละเอียดข้อมูลของแต่ละตัวแปรโดยใช้สถิติพรรณนา แสดงได้ดังตารางภาคผนวก 1 (ในภาคผนวก ก)

ประเด็นที่ 2 รายงานการวิเคราะห์ความสัมพันธ์ของตัวแปรอิสระกับตัวแปรตามด้วยสัมประสิทธิ์สหสัมพันธ์สเปียร์แมน

ตาราง 7 รายละเอียดค่าสัมประสิทธิ์สหสัมพันธ์ของข้อมูลชุดที่ 1

ตัวแปร	r_s	p-value
age	0.236*	<0.001
gender	0.007	0.074
height	-0.013*	0.001
weight	0.180*	<0.001

ตัวแปร	r_s	p-value
ap_hi	0.451*	<0.001
ap_lo	0.356*	<0.001
cholesterol	0.215*	<0.001
gluc	0.091*	<0.001
smoke	-0.017*	<0.001
alco	-0.009*	0.021
active	-0.038*	<0.001

*มีนัยสำคัญ ที่ระดับนัยสำคัญ 0.05

จากตารางที่ 7 พบว่า ตัวแปรอิสระที่มีความสัมพันธ์กับตัวแปรตามที่ระดับนัยสำคัญ 0.05 ได้แก่ อายุ (age), ส่วนสูง (height), น้ำหนัก (weight), ค่าความดันขณะหัวใจบีบตัว (ap_hi), ค่าความดันขณะหัวใจคลายตัว (ap_lo), ระดับคอเลสเตอรอล (cholesterol), ระดับน้ำตาลกลูโคส (gluc), การสูบบุหรี่ (smoke), การดื่มแอลกอฮอล์ (alco), การออกกำลังกาย (active)

เมื่อพิจารณาจากค่าสัมประสิทธิ์สหสัมพันธ์ พบว่า ค่าความดันขณะหัวใจบีบตัว (ap_hi) มีความสัมพันธ์กับการเป็นโรคหัวใจและหลอดเลือดมากที่สุด รองลงมาคือ ค่าความดันขณะหัวใจคลายตัว (ap_lo) อายุ (age) และระดับคอเลสเตอรอล (cholesterol) ตามลำดับ

ประเด็นที่ 3 แปลงข้อมูลให้อยู่ในรูปปกติ

ตาราง 8 ตัวอย่างข้อมูลหลังการแปลงค่าให้อยู่ในรูปปกติของข้อมูลชุดที่ 1

age	gender	height	weight	ap_hi	ap_lo	smoke	alco	active	...	class
0.5881	men	0.4179	0.1882	0.3077	0.4286	No	No	Yes	...	0
0.5169	women	0.2388	0.1529	0.2308	0.1429	No	No	No	...	0
0.8607	women	0.1642	0.2176	0.3846	0.4286	No	No	No	...	0
0.8761	women	0.2537	0.3706	0.4615	0.4286	No	No	Yes	...	0
0.9084	women	0.4328	0.2941	0.3846	0.4286	No	No	Yes	...	0
0.6208	men	0.4925	0.1765	0.3846	0.4286	No	No	Yes	...	0
0.7981	women	0.2537	0.3765	0.3846	0.4286	No	No	No	...	1
0.4114	women	0.3134	0.2353	0.2308	0.2857	No	No	No	...	1
0.4639	women	0.0746	0.1588	0.5385	0.4286	No	No	Yes	...	1
0.8603	men	0.4328	0.2000	0.6923	0.7143	No	No	No	...	1
0.8500	women	0.2239	0.2059	0.6154	0.5714	No	No	Yes	...	1
0.9598	women	0.1791	0.3471	0.6923	0.2857	No	No	No	...	1

3.3.2. ข้อมูลชุดที่ 2 โรคเบาหวาน

ในข้อมูลชุดที่ 2 แบ่งการรายงานการวิเคราะห์และจัดเตรียมข้อมูลออกเป็น 3 ประเด็น โดยมีรายละเอียดดังนี้

ประเด็นที่ 1 รายละเอียดของตัวแปรในข้อมูลชุดที่ 2

ตาราง 9 รายละเอียดตัวแปรข้อมูลชุดที่ 2

ตัวแปร	ความหมาย	ลักษณะตัวแปร
HighBP	โรคความดันโลหิตสูง	เชิงคุณภาพ
HighChol	คอเลสเตอรอลสูง	เชิงคุณภาพ
CholCheck	รับการตรวจคอเลสเตอรอลทุกๆ 5 ปี	เชิงคุณภาพ
BMI	ค่าดัชนีมวลกาย (kg/m ²)	เชิงปริมาณ
Smoker	สูบบุหรี่อย่างน้อย 100 มวน	เชิงคุณภาพ
Stroke	โรคหลอดเลือดสมอง	เชิงคุณภาพ
HeartDiseaseorAttack	โรคหลอดเลือดหัวใจ (CHD) หรือกล้ามเนื้อหัวใจตาย (MI)	เชิงคุณภาพ
PhysActivity	การออกกำลังกาย	เชิงคุณภาพ
Fruits	ทานผลไม้อย่างน้อย 1 ครั้งต่อสัปดาห์	เชิงคุณภาพ
Veggies	ทานผักอย่างน้อย 1 ครั้งต่อสัปดาห์	เชิงคุณภาพ
HvyAlcoholConsump	ภาวะติดแอลกอฮอล์	เชิงคุณภาพ
AnyHealthcare	การคุ้มครองด้านการดูแลสุขภาพ	เชิงคุณภาพ
NoDocbcCost	ไม่สามารถพบแพทย์ได้เนื่องจากปัญหาค่าใช้จ่าย	เชิงคุณภาพ
GenHlth	ระดับคะแนนสุขภาพแบ่งเป็น 5 ระดับ	เชิงคุณภาพ
MentHlth	จำนวนวันที่สุขภาพจิตแย่ใน 1 เดือน (Day)	เชิงปริมาณ
PhysHlth	จำนวนวันที่เจ็บป่วยและบาดเจ็บใน 1 เดือน (Day)	เชิงปริมาณ
DiffWalk	ปัญหาในการเดินหรือขึ้นบันได	เชิงคุณภาพ
Sex	เพศ	เชิงคุณภาพ
Age	อายุแบ่งเป็น 13 ช่วง	เชิงคุณภาพ
Education	ระดับการศึกษาแบ่งเป็น 6 ระดับ	เชิงคุณภาพ
Income	รายได้แบ่งเป็น 8 ช่วง	เชิงคุณภาพ
Class	การเป็นโรคเบาหวาน	เชิงคุณภาพ

โดยรายละเอียดข้อมูลของแต่ละตัวแปรโดยใช้สถิติพรรณนา แสดงได้ดังตารางภาคผนวก 2 (ในภาคผนวก ก)

ประเด็นที่ 2 รายงานการวิเคราะห์ความสัมพันธ์ของตัวแปรอิสระกับตัวแปรตามด้วยสัมประสิทธิ์สหสัมพันธ์สเปียร์แมน

ตาราง 10 รายละเอียดค่าสัมประสิทธิ์สหสัมพันธ์ของข้อมูลชุดที่ 2

ตัวแปร	r_s	p-value
HighBP	0.371*	<0.001
HighChol	0.282*	<0.001
CholCheck	0.118*	<0.001
BMI	0.316*	<0.001
Smoker	0.077*	<0.001
Stroke	0.124*	<0.001
HeartDiseaseorAttack	0.208*	<0.001
PhysActivity	-0.146*	<0.001
Fruits	-0.044*	<0.001
Veggies	-0.072*	<0.001
HvyAlcoholConsump	-0.098*	<0.001
AnyHealthcare	0.028*	<0.001
NoDocbcCost	0.035*	<0.001
GenHlth	0.400*	<0.001
MentHlth	0.077*	<0.001
PhysHlth	0.202*	<0.001
DiffWalk	0.262*	<0.001
Sex	0.045*	<0.001
Age	0.260*	<0.001
Education	-0.156*	<0.001
Income	-0.217*	<0.001

*มีนัยสำคัญ ที่ระดับนัยสำคัญ 0.05

จากตารางที่ 10 พบว่า ตัวแปรอิสระทุกตัวมีความสัมพันธ์กับตัวแปรตามที่มีระดับนัยสำคัญ 0.05

เมื่อพิจารณาจากค่าสัมประสิทธิ์สหสัมพันธ์ พบว่า ระดับคะแนนสุขภาพ (GenHlth) มีความสัมพันธ์กับการเป็นโรคเบาหวานมากที่สุด รองลงมาคือ โรคความดันโลหิตสูง (HighBP) ค่าดัชนีมวลกาย (BMI) และคอเลสเตอรอลสูง (HighChol) ตามลำดับ

ประเด็นที่ 3 แปลงข้อมูลให้อยู่ในรูปปกติ

ตาราง 11 ตัวอย่างข้อมูลหลังการแปลงค่าให้อยู่ในรูปปกติของข้อมูลชุดที่ 2

HighBP	HighChol	CholCheck	BMI	Smoker	Stroke	HeartDiseaseorAttack	PhysActivity	Fruits	Veggies	...	Class
Yes	No	Yes	0.3684	No	No	No	Yes	No	Yes	...	0
Yes	Yes	Yes	0.3684	Yes	Yes	No	No	Yes	No	...	0
No	No	Yes	0.3684	No	No	No	Yes	Yes	Yes	...	0
Yes	Yes	Yes	0.4211	Yes	No	No	Yes	Yes	Yes	...	0
No	No	Yes	0.4474	Yes	No	No	Yes	Yes	Yes	...	0
No	No	Yes	0.1579	No	No	No	Yes	Yes	Yes	...	0
No	Yes	Yes	0.7368	No	No	No	Yes	No	Yes	...	1
Yes	Yes	Yes	0.5526	Yes	Yes	Yes	No	No	No	...	1
No	Yes	Yes	0.5263	No	No	No	Yes	No	Yes	...	1
No	Yes	Yes	0.3684	No	No	No	Yes	No	No	...	1
Yes	Yes	Yes	0.9474	No	No	No	No	Yes	Yes	...	1
Yes	Yes	Yes	0.2895	Yes	No	No	Yes	No	Yes	...	1

3.3.3. ข้อมูลชุดที่ 3 การเลิกสูบบุหรี่

ในข้อมูลชุดที่ 3 แบ่งการรายงานการวิเคราะห์และจัดเตรียมข้อมูลออกเป็น 3 ประเด็น โดยมีรายละเอียดดังนี้

ประเด็นที่ 1 รายละเอียดของตัวแปรในข้อมูลชุดที่ 3

ตาราง 12 รายละเอียดตัวแปรข้อมูลชุดที่ 3

ตัวแปร	ความหมาย	ลักษณะตัวแปร
age	อายุ (Year)	เชิงปริมาณ
height	ส่วนสูง (cm.)	เชิงปริมาณ
weight	น้ำหนัก (kg.)	เชิงปริมาณ
waist	รอบเอว (cm.)	เชิงปริมาณ
eyesight_left	ความยาวสายตาด้านซ้าย (Diopter)	เชิงปริมาณ
eyesight_right	ความยาวสายตาด้านขวา (Diopter)	เชิงปริมาณ
hearing_left	การได้ยินของหูด้านซ้าย	เชิงคุณภาพ
hearing_right	การได้ยินของหูด้านขวา	เชิงคุณภาพ
systolic	ค่าความดันขณะหัวใจบีบตัว (mmHg.)	เชิงปริมาณ
relaxation	ค่าความดันของหัวใจคลายตัว (mmHg.)	เชิงปริมาณ

ตัวแปร	ความหมาย	ลักษณะตัวแปร
fasting_blood_sugar	ค่าน้ำตาลในเลือด (FBS) (mg/dL)	เชิงปริมาณ
Cholesterol	ค่าคอเลสเตอรอล (mg/dL)	เชิงปริมาณ
triglyceride	ค่าไตรกลีเซอไรด์ (mg/dL)	เชิงปริมาณ
HDL	ค่าคอเลสเตอรอลชนิดที่ดี (mg/dL)	เชิงปริมาณ
LDL	ค่าคอเลสเตอรอลชนิดที่ไม่ดี (mg/dL)	เชิงปริมาณ
hemoglobin	ค่าฮีโมโกลบิน (g/dL)	เชิงปริมาณ
Urine_protein	โปรตีนในปัสสาวะ	เชิงคุณภาพ
serum_creatinine	ซีรัมครีเอตินิน (Scr) (mg/dL)	เชิงปริมาณ
AST	ค่าเอนไซม์ AST ในเลือด (U/L)	เชิงปริมาณ
ALT	ค่าเอนไซม์ ALT ในเลือด (U/L)	เชิงปริมาณ
Gtp	ค่าจีโปรตีน (kDa)	เชิงปริมาณ
dental_caries	โรคฟันผุ	เชิงคุณภาพ
class	การเลิกสูบบุหรี่	เชิงคุณภาพ

โดยรายละเอียดข้อมูลของแต่ละตัวแปรโดยใช้สถิติพรรณนา แสดงได้ดังตารางภาคผนวก 3 (ในภาคผนวก ก)

ประเด็นที่ 2 รายงานการวิเคราะห์ความสัมพันธ์ของตัวแปรอิสระกับตัวแปรตามด้วยสัมประสิทธิ์สหสัมพันธ์สเปียร์แมน

ตาราง 13 รายละเอียดค่าสัมประสิทธิ์สหสัมพันธ์ของข้อมูลชุดที่ 3

ตัวแปร	r_s	p-value
age	-0.178*	<0.001
height	0.403*	<0.001
weight	0.311*	<0.001
waist	0.212*	<0.001
eyesight_left	0.093*	<0.001
eyesight_right	0.105*	<0.001
hearing_left	-0.024*	<0.001
hearing_right	-0.018*	.003

ตัวแปร	r_s	p-value
systolic	0.057*	<0.001
relaxation	0.086*	<0.001
fasting_blood_sugar	0.071*	<0.001
Cholesterol	-0.047*	<0.001
triglyceride	0.222*	<0.001
HDL	-0.199*	<0.001
LDL	-0.051*	<0.001
hemoglobin	0.412*	<0.001
Urine_protein	-0.004	.472
serum_creatinine	0.273*	<0.001
AST	0.052*	<0.001
ALT	0.173*	<0.001
Gtp	0.350*	<0.001
dental_caries	0.113*	<0.001

*มีนัยสำคัญ ที่ระดับนัยสำคัญ 0.05

จากตารางที่ 13 พบว่า ตัวแปรอิสระทุกตัวมีความสัมพันธ์กับตัวแปรตามที่ระดับนัยสำคัญ 0.05 ยกเว้นโปรตีนในปัสสาวะ (Urine_protein)

เมื่อพิจารณาจากค่าสัมประสิทธิ์สหสัมพันธ์ พบว่า ค่าฮีโมโกลบิน (hemoglobin) มีความสัมพันธ์กับการเลิกสูบบุหรี่มากที่สุด รองลงมาคือ ส่วนสูง (height) ค่าจีโปรตีน (Gtp) และ น้ำหนัก (weight) ตามลำดับ

ประเด็นที่ 3 แปลงข้อมูลให้อยู่ในรูปปกติ

ตาราง 14 ตัวอย่างข้อมูลหลังการแปลงค่าให้อยู่ในรูปปกติของข้อมูลชุดที่ 3

age	height	weight	waist	eyesight_left	eyesight_right	hearing_left	hearing_right	systolic	relaxation	...	Class
0.6154	0.5000	0.1905	0.3462	0.0918	0.0816	1	1	0.3395	0.3302	...	0
0.3077	0.7500	0.5714	0.5641	0.0816	0.0918	0	0	0.3642	0.4528	...	0
0.3077	0.8333	0.4286	0.4359	0.1429	0.1429	0	0	0.2407	0.1887	...	0
0.6154	0.5000	0.1905	0.3462	0.0918	0.0816	1	1	0.3395	0.3302	...	0
0.5385	0.3333	0.3333	0.3769	0.0918	0.0612	1	0	0.2778	0.2642	...	0
0.1538	0.6667	0.4762	0.4679	0.1122	0.1429	0	0	0.4136	0.3774	...	0
0.3077	0.3333	0.2857	0.4615	0.1122	0.1122	1	0	0.3025	0.2830	...	1
0.3077	0.5000	0.1905	0.3077	0.1122	0.0612	0	0	0.2778	0.4057	...	1
0.6154	0.6667	0.3810	0.4231	0.0918	0.0714	0	0	0.2901	0.3396	...	1
0.6154	0.5833	0.3810	0.5897	0.0612	0.0612	1	1	0.3642	0.3585	...	1

age	height	weight	waist	eyesight_left	eyesight_right	hearing_left	hearing_right	systolic	relaxation	...	Class
0.3077	0.7500	0.4762	0.4615	0.0612	0.0408	0	0	0.2840	0.2925	...	1
0.3077	0.5833	0.2857	0.3333	0.0918	0.0918	0	0	0.3951	0.3396	...	1

3.4. ขั้นตอนการพัฒนาตัวแบบ

3.4.1. พัฒนาตัวแบบการจำแนกโดยใช้ตัวแปรอิสระทั้งหมด

ในส่วนนี้เป็นกระบวนการพัฒนาตัวแบบโดยใช้ตัวแปรอิสระทั้งหมดด้วย 4 เทคนิค ได้แก่ เทคนิคการถดถอยลอจิสติก (Logistic Regression : LR) เทคนิคต้นไม้ตัดสินใจ (Decision Tree : DT) เทคนิคป่าสุ่ม (Random Forest : RF) และเทคนิคเครื่องเวกเตอร์ค้ำยัน (Support Vector Machine : SVM) โดยใช้การพัฒนาตัวแบบ 10-Fold Cross-Validation มีขั้นตอนดังต่อไปนี้ แสดงได้ดังผังงานในภาพที่ 10

ขั้นตอนที่ 1 แบ่งชุดข้อมูลออกเป็น 10 ชุด ตามหลักการ 10-Fold Cross-Validation โดยการสุ่มตัวอย่างแบบแบ่งเป็นชั้นภูมิ (Stratified Random Sampling) ดังตารางที่ 15

ตาราง 15 การแบ่งข้อมูลเป็น 10 ชุดของข้อมูลทั้ง 3 ชุด

ชุด	ข้อมูลชุดที่ 1 โรคหัวใจและหลอดเลือด		ข้อมูลชุดที่ 2 โรคเบาหวาน		ข้อมูลชุดที่ 3 การเลิกสูบบุหรี่	
	Yes	No	Yes	No	Yes	No
ชุด 1	3,383	3,461	3,440	3,374	918	1,812
ชุด 2	3,383	3,461	3,440	3,374	918	1,812
ชุด 3	3,383	3,461	3,440	3,374	918	1,811
ชุด 4	3,383	3,460	3,440	3,374	917	1,811
ชุด 5	3,383	3,460	3,439	3,374	917	1,811
ชุด 6	3,383	3,460	3,439	3,374	917	1,811
ชุด 7	3,383	3,460	3,439	3,374	917	1,811
ชุด 8	3,383	3,460	3,439	3,374	917	1,811
ชุด 9	3,383	3,460	3,439	3,374	917	1,811
ชุด 10	3,383	3,460	3,439	3,374	917	1,811
รวม	33,830	34,603	34,394	33,740	9,173	18,112

ขั้นตอนที่ 2 พัฒนาตัวแบบการจำแนกตามหลักการ 10-Fold Cross-Validation โดยนำชุดข้อมูลเรียนรู้ (9 ชุดข้อมูล) มาพัฒนาตัวแบบและทดสอบประสิทธิภาพด้วยชุดข้อมูลทดสอบ (1 ชุดที่เหลือ) ด้วย 4 เทคนิคดังนี้

1. เทคนิคการถดถอยลอจิสติก (Logistic Regression : LR)
2. เทคนิคต้นไม้ตัดสินใจ (Decision Tree : DT)
3. เทคนิคป่าสุ่ม (Random Forest : RF) ซึ่งมีการกำหนดพารามิเตอร์ดังนี้
 - 3.1. mtry ของข้อมูลชุดที่ 1 เท่ากับ 3 และ 4 และข้อมูลชุดที่ 2, 3 เท่ากับ 4 และ 5 เป็นการสุ่มตัวแปรอิสระเพื่อนำมาสร้างต้นไม้ตัดสินใจมีค่าเท่ากับ \sqrt{n} (n แทนจำนวนตัวแปรอิสระทั้งหมด)(Rebala et al., 2019)
 - 3.2. ntree เท่ากับ 250 และ 500
 - 3.3. importance กำหนดเป็น TRUE
4. เทคนิคเครื่องเวกเตอร์ค้ำยัน (Support Vector Machine : SVM) ซึ่งมีการกำหนดพารามิเตอร์ดังนี้
 - 4.1. kernel กำหนดเป็น radial
 - 4.2. gamma เท่ากับ 0.001, 0.01, 0.1 และ 0.2
 - 4.3. cost เท่ากับ 1 และ 10
 โดยจะพัฒนาตัวแบบทั้งหมด 10 รอบ และประเมินประสิทธิภาพบนชุดข้อมูลทดสอบ (ค่าเฉลี่ย)

ขั้นตอนที่ 3 รายงานประสิทธิภาพด้วยค่าความไว (Sensitivity) ค่าความจำเพาะ (Specificity) ค่าความแม่นยำ (Accuracy) ค่าความแม่นยำที่สมดุลแบบปรับน้ำหนักถ่วง (Weighted Balanced Accuracy) และค่าประสิทธิภาพโดยรวมแบบปรับน้ำหนักถ่วง ($F1_{weight}$)

3.4.2. พัฒนาตัวแบบการจำแนกโดยการคัดเลือกตัวแปรอิสระ

ในส่วนนี้เป็นการพัฒนาตัวแบบโดยการคัดเลือกตัวแปรอิสระ ซึ่งแบ่งวิธีการคัดเลือกออกเป็น 2 วิธีดังนี้

วิธีที่ 1 คัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติก (Logistic Regression)

วิธีที่ 2 คัดเลือกตัวแปรอิสระด้วยค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient)

วิธีที่ 1 คัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติก (Logistic Regression)

การคัดเลือกตัวแปรอิสระในเทคนิคการถดถอยลอจิสติกมีวิธีเลือก 3 วิธีประกอบด้วย

1. วิธีการเลือกแบบไปข้างหน้า (Forward Selection) เป็นวิธีการเลือกตัวแปรอิสระด้วยการเพิ่มตัวแปรอิสระเข้าไปในตัวแบบครั้งละ 1 ตัว โดยเลือกตัวแปรที่มีสัมประสิทธิ์สหสัมพันธ์กับตัว

แปรตามสูงสุดเข้าเป็นตัวแรก และทำการทดสอบนัยสำคัญของตัวแปรอิสระ เมื่อตัวแปรอิสระถูกคัดเลือกเข้ามาอยู่ในตัวแบบแล้วจะไม่มีผลการพิจารณาตัดออกภายหลัง ขั้นตอนถัดไปทำการคัดเลือกตัวแปรอิสระที่มีความสัมพันธ์รองลงมา และทดสอบอิทธิพลของตัวแปรอิสระที่เพิ่มเข้ามา เมื่อกำหนดให้ตัวแปรที่ถูกเลือกก่อนหน้านี้ที่อยู่ในตัวแบบ และจะทำไปเรื่อยๆ จนกระทั่งไม่มีตัวแปรอิสระใดที่สามารถอธิบายความผันแปรของตัวแปรตามได้อีก

2. วิธีกำจัดแบบถอยหลัง (Backward Elimination) เป็นวิธีการเลือกตัวแปรอิสระที่เริ่มต้นด้วยตัวแบบที่ประกอบด้วยตัวแปรอิสระทั้งหมด แล้วจึงตัดตัวแปรอิสระที่อธิบายความสัมพันธ์กับตัวแปรตามได้น้อยออกไปครั้งละ 1 ตัว เมื่อตัวแปรอิสระตัวใดถูกตัดออกแล้ว จะไม่พิจารณาเพิ่มเข้ามาอยู่ในตัวแบบอีก และจะทำไปเรื่อยๆ จนกระทั่งตัวแปรอิสระที่เหลืออยู่ในตัวแบบสามารถอธิบายความสัมพันธ์กับตัวแปรตามอย่างมีนัยสำคัญเชิงสถิติ ก็จะหยุดการทำงาน

3. วิธีการถดถอยทีละขั้น (Stepwise Regression) เป็นวิธีการคัดเลือกตัวแปรอิสระที่ประยุกต์มาจากวิธีการเลือกแบบไปข้างหน้า โดยมีการเลือกตัวแปรอิสระเข้าไปในตัวแบบครั้งละ 1 ตัวเช่นกัน แต่เพิ่มการทดสอบอิทธิพลของตัวแปรอิสระที่อยู่ในตัวแบบก่อนหน้านี้แล้วทุกตัว เมื่อกำหนดให้ตัวแปรอิสระอื่นยังคงอยู่ในตัวแบบ ดังนั้นตัวแปรอิสระที่เพิ่มเข้ามาในตัวแบบขั้นแรกๆ สามารถตัดออกได้ในภายหลัง การพิจารณาเลือกตัวแปรอิสระจะทำไปเรื่อยๆ จนกระทั่งไม่สามารถหาตัวแปรอิสระอื่นเข้ามาในตัวแบบได้อีก และขณะเดียวกันไม่สามารถตัดตัวแปรอิสระที่อยู่ในตัวแบบออกได้

และในงานวิจัยนี้ผู้วิจัยได้ออกแบบการคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกออกเป็น 2 แบบดังต่อไปนี้

แบบที่ 1 การคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติก โดยใช้ข้อมูลทั้งหมด มีขั้นตอนการพัฒนาตัวแบบดังต่อไปนี้ แสดงได้ดังผังงานในภาพที่ 11

ขั้นตอนที่ 1 นำข้อมูลทั้งหมดมาสร้างตัวแบบด้วยเทคนิคการถดถอยลอจิสติก ด้วยวิธีการคัดเลือกตัวแปรอิสระทั้ง 3 วิธี ที่ระดับนัยสำคัญ 0.05 จะได้เซตของตัวแปรอิสระที่ปรากฏในแต่ละตัวแบบด้วยการคัดเลือกทั้ง 3 วิธี ได้แก่ A_F แทน เซตของตัวแปรอิสระที่ใช้วิธีการเลือกแบบไปข้างหน้า, A_B แทน เซตของตัวแปรอิสระที่ใช้วิธีกำจัดแบบถอยหลัง และ A_S แทน เซตของตัวแปรอิสระที่ใช้วิธีการถดถอยทีละขั้น

ขั้นตอนที่ 2 พัฒนาตัวแบบโดยใช้ตัวแปรอิสระที่ได้จากการคัดเลือกในขั้นตอนที่ 1 จากนั้นทำตามขั้นตอนที่ 1-3 ในหัวข้อ 3.4.1

หมายเหตุ เทคนิคป่าสุ่ม กำหนด mtry ของข้อมูลชุดที่ 1, 3 เท่ากับ 3 และ 4 และข้อมูลชุดที่ 2 เท่ากับ 4 และ 5

แบบที่ 2 การคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติก โดยใช้ข้อมูลเรียนรู้ทำซ้ำ 10 ครั้ง มีขั้นตอนการพัฒนาตัวแบบดังต่อไปนี้ แสดงได้ดังผังงานในภาพที่ 12

ขั้นตอนที่ 1 แบ่งชุดข้อมูลออกเป็น 10 ชุด ตามหลักการ 10-Fold Cross-Validation โดยการสุ่มตัวอย่างแบบแบ่งเป็นชั้นภูมิ (Stratified Random Sampling) ดังตารางที่ 15

ขั้นตอนที่ 2 นำชุดข้อมูลเรียนรู้ในแต่ละชุดจากขั้นตอนที่ 1 มาสร้างตัวแบบด้วยเทคนิคการถดถอย-ลอจิสติก และทำการคัดเลือกตัวแปรอิสระด้วยวิธีเลือกตัวแปรทั้ง 3 วิธี ที่ระดับนัยสำคัญ 0.05 จะได้เซตของตัวแปรอิสระจากตัวแบบทั้ง 10 ตัวแบบ และคัดเลือกตัวแปรอิสระโดยพิจารณาตัวแปรอิสระที่ปรากฏในตัวแบบทั้ง 10 ตัวแบบ ดังนั้นจะได้ A_F , A_B และ A_S ตามนิยามต่อไปนี้

$$A_F = \bigcap_{i=1}^{10} A_{Fi}$$

เมื่อ A_{Fi} แทน เซตของตัวแปรอิสระที่ใช้วิธีการเลือกแบบไปข้างหน้าของชุดที่ i

$$A_B = \bigcap_{i=1}^{10} A_{Bi}$$

เมื่อ A_{Bi} แทน เซตของตัวแปรอิสระที่ใช้วิธีการกำจัดแบบถดถอยหลังของชุดที่ i

$$A_S = \bigcap_{i=1}^{10} A_{Si}$$

เมื่อ A_{Si} แทน เซตของตัวแปรอิสระที่ใช้วิธีการถดถอยทีละขั้นของชุดที่ i

ขั้นตอนที่ 3 พัฒนาตัวแบบโดยใช้ตัวแปรอิสระที่ได้จากการคัดเลือกในขั้นตอนที่ 2 จากนั้นทำตามขั้นตอนที่ 1-3 ในหัวข้อ 3.4.1

หมายเหตุ เทคนิคป่าสุ่ม mtry ของข้อมูลชุดที่ 1, 2 และ 3 เท่ากับ 3 และ 4

วิธีที่ 2 คัดเลือกตัวแปรอิสระด้วยค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient) แสดงได้ดังผังงานในภาพที่ 13

ขั้นตอนที่ 1 นำข้อมูลทั้งหมดมาหาค่าสัมประสิทธิ์สหสัมพันธ์ของตัวแปรอิสระกับตัวแปรตาม จะได้ค่าดังตารางที่ 7, 10 และ 13 ทำการคัดเลือกตัวแปรอิสระ โดยพิจารณาจำนวนตัวแปรอิสระที่ใช้จากวิธีที่ 1 ทั้ง 2 แบบ โดยเลือกใช้จำนวนตัวแปรอิสระตามแบบที่ได้จำนวนตัวแปรอิสระมากที่สุด ยกตัวอย่างข้อมูลชุดที่ 2 คัดเลือกด้วยเทคนิคการถดถอยลอจิสติกแบบที่ 1 มีจำนวนตัวแปรอิสระ 16 ตัว และแบบที่ 2 มีจำนวนตัวแปรอิสระ 14 ตัว ดังนั้นการคัดเลือกตัวแปรอิสระด้วยค่าสัมประสิทธิ์สหสัมพันธ์จะคัดเลือกมา 16 ตัว

ขั้นตอนที่ 2 พัฒนาตัวแบบโดยใช้ตัวแปรอิสระที่ได้จากการคัดเลือกในขั้นตอนที่ 1 จากนั้นทำตามขั้นตอนที่ 1-3 ในหัวข้อ 3.4.1

หมายเหตุ เทคนิคป่าสุ่ม mtry ของข้อมูลชุดที่ 1, 3 เท่ากับ 3 และ 4 และข้อมูลชุดที่ 2 เท่ากับ 4 และ 5

ตาราง 16 การคัดเลือกตัวแปรอิสระในข้อมูลชุดที่ 1 ด้วยวิธีการทั้ง 3 รูปแบบ

คัดเลือกตัวแปรอิสระ	วิธี	ตัวแปรอิสระ
เทคนิคการถดถอยลอจิสติก โดยใช้ข้อมูลทั้งหมด	การเลือกแบบไปข้างหน้า (Forward Selection)	อายุ, ส่วนสูง, น้ำหนัก, ค่าความดันขณะหัวใจบีบตัว, ค่าความดันขณะหัวใจคลายตัว, ระดับคอเลสเตอรอล, ระดับน้ำตาลกลูโคส, การสูบบุหรี่, การดื่มแอลกอฮอล์, การออกกำลังกาย
	กำจัดแบบถดถอยหลัง (Backward Elimination)	อายุ, ส่วนสูง, น้ำหนัก, ค่าความดันขณะหัวใจบีบตัว, ค่าความดันขณะหัวใจคลายตัว, ระดับคอเลสเตอรอล, ระดับน้ำตาลกลูโคส, การสูบบุหรี่, การดื่มแอลกอฮอล์, การออกกำลังกาย
	การถดถอยทีละขั้น (Stepwise Regression)	อายุ, ส่วนสูง, น้ำหนัก, ค่าความดันขณะหัวใจบีบตัว, ค่าความดันขณะหัวใจคลายตัว, ระดับคอเลสเตอรอล, ระดับน้ำตาล

คัดเลือกตัวแปรอิสระ	วิธี	ตัวแปรอิสระ
		กลูโคส, การสูบบุหรี่, การดื่มแอลกอฮอล์, การออกกำลังกาย
เทคนิคการถดถอยลอ- จิสติก โดยใช้ข้อมูล เรียนรู้ทำซ้ำ 10 ครั้ง	การเลือกแบบไปข้างหน้า (Forward Selection)	อายุ, ส่วนสูง, น้ำหนัก, ค่าความดันขณะ หัวใจบีบตัว, ค่าความดันขณะหัวใจคลาย ตัว, ระดับคอเลสเตอรอล, ระดับน้ำตาล กลูโคส, การสูบบุหรี่, การดื่มแอลกอฮอล์, การออกกำลังกาย
	กำจัดแบบถดถอยหลัง (Backward Elimination)	อายุ, ส่วนสูง, น้ำหนัก, ค่าความดันขณะ หัวใจบีบตัว, ค่าความดันขณะหัวใจคลาย ตัว, ระดับคอเลสเตอรอล, ระดับน้ำตาล กลูโคส, การสูบบุหรี่, การดื่มแอลกอฮอล์, การออกกำลังกาย
	การถดถอยทีละขั้น (Stepwise Regression)	อายุ, ส่วนสูง, น้ำหนัก, ค่าความดันขณะ หัวใจบีบตัว, ค่าความดันขณะหัวใจคลาย ตัว, ระดับคอเลสเตอรอล, ระดับน้ำตาล กลูโคส, การสูบบุหรี่, การดื่มแอลกอฮอล์, การออกกำลังกาย
ค่าสัมประสิทธิ์ สหสัมพันธ์	สเปียร์แมน (Spearman)	อายุ, ส่วนสูง, น้ำหนัก, ค่าความดันขณะ หัวใจบีบตัว, ค่าความดันขณะหัวใจคลาย ตัว, ระดับคอเลสเตอรอล, ระดับน้ำตาล กลูโคส, การสูบบุหรี่, การดื่มแอลกอฮอล์, การออกกำลังกาย

จากตารางที่ 16 การคัดเลือกตัวแปรอิสระด้วยวิธีการทั้ง 3 รูปแบบสำหรับชุดข้อมูลที่ 1 พบว่าตัวแปรอิสระที่ถูกคัดเลือกเข้าตัวแบบมีจำนวนเท่ากันคือ 10 ตัว และได้ตัวแปรเหมือนกันทั้ง 3 รูปแบบ โดยมีตัวแปรที่ถูกคัดออกจากตัวแบบเพียงตัวแปรเดียวคือ เพศ

ตาราง 17 การคัดเลือกตัวแปรอิสระในข้อมูลชุดที่ 2 ด้วยวิธีการทั้ง 3 รูปแบบ

คัดเลือกตัวแปรอิสระ	วิธี	ตัวแปรอิสระ
เทคนิคการถดถอยล- จิสติก โดยใช้ข้อมูล ทั้งหมด	การเลือกแบบไปข้างหน้า (Forward Selection)	โรคความดันโลหิตสูง, คอเลสเทอรอลสูง, รับการตรวจคอเลสเทอรอลทุกๆ 5 ปี, ค่า ดัชนีมวลกาย, โรคหลอดเลือดสมอง, โรค หลอดเลือดหัวใจ (CHD) หรือกล้ามเนื้อ หัวใจตาย (MI), ทานผักอย่างน้อย 1 ครั้ง ต่อสัปดาห์, ภาวะติดแอลกอฮอล์, ระดับ คะแนนสุขภาพ, จำนวนวันที่สุขภาพจิตแย่ ใน 1 เดือน, จำนวนวันที่เจ็บป่วยหรือ บาดเจ็บใน 1 เดือน, ปัญหาในการเดินหรือ ขึ้นบันได, เพศ, อายุ, ระดับการศึกษา, รายได้
	กำจัดแบบถดถอยหลัง (Backward Elimination)	โรคความดันโลหิตสูง, คอเลสเทอรอลสูง, รับการตรวจคอเลสเทอรอลทุกๆ 5 ปี, ค่า ดัชนีมวลกาย, โรคหลอดเลือดสมอง, โรค หลอดเลือดหัวใจ (CHD) หรือกล้ามเนื้อ หัวใจตาย (MI), ทานผักอย่างน้อย 1 ครั้ง ต่อสัปดาห์, ภาวะติดแอลกอฮอล์, ระดับ คะแนนสุขภาพ, จำนวนวันที่สุขภาพจิตแย่ ใน 1 เดือน, จำนวนวันที่เจ็บป่วยหรือ บาดเจ็บใน 1 เดือน, ปัญหาในการเดินหรือ ขึ้นบันได, เพศ, อายุ, ระดับการศึกษา, รายได้
	การถดถอยทีละขั้น (Stepwise Regression)	โรคความดันโลหิตสูง, คอเลสเทอรอลสูง, รับการตรวจคอเลสเทอรอลทุกๆ 5 ปี, ค่า ดัชนีมวลกาย, โรคหลอดเลือดสมอง, โรค หลอดเลือดหัวใจ (CHD) หรือกล้ามเนื้อ หัวใจตาย (MI), ทานผักอย่างน้อย 1 ครั้ง ต่อสัปดาห์, ภาวะติดแอลกอฮอล์, ระดับ

คัดเลือกตัวแปรอิสระ	วิธี	ตัวแปรอิสระ
		คะแนนสุขภาพ, จำนวนวันที่สุขภาพจิตแย่ใน 1 เดือน, จำนวนวันที่เจ็บป่วยหรือบาดเจ็บใน 1 เดือน, ปัญหาในการเดินหรือขึ้นบันได, เพศ, อายุ, ระดับการศึกษา, รายได้
เทคนิคการถดถอยลง- จิสติก โดยใช้ข้อมูล เรียนรู้ทำซ้ำ 10 ครั้ง	การเลือกแบบไปข้างหน้า (Forward Selection)	โรคความดันโลหิตสูง, คอเลสเทอรอลสูง, รับการตรวจคอเลสเทอรอลทุกๆ 5 ปี, ค่าดัชนีมวลกาย, โรคหลอดเลือดสมอง, โรคหลอดเลือดหัวใจ (CHD) หรือกล้ามเนื้อหัวใจตาย (MI), ภาวะติดแอลกอฮอล์, ระดับคะแนนสุขภาพ, จำนวนวันที่สุขภาพจิตแย่ใน 1 เดือน, จำนวนวันที่เจ็บป่วยหรือบาดเจ็บใน 1 เดือน, ปัญหาในการเดินหรือขึ้นบันได, เพศ, อายุ, รายได้
	กำจัดแบบถดถอยหลัง (Backward Elimination)	โรคความดันโลหิตสูง, คอเลสเทอรอลสูง, รับการตรวจคอเลสเทอรอลทุกๆ 5 ปี, ค่าดัชนีมวลกาย, โรคหลอดเลือดสมอง, โรคหลอดเลือดหัวใจ (CHD) หรือกล้ามเนื้อหัวใจตาย (MI), ภาวะติดแอลกอฮอล์, ระดับคะแนนสุขภาพ, จำนวนวันที่สุขภาพจิตแย่ใน 1 เดือน, จำนวนวันที่เจ็บป่วยหรือบาดเจ็บใน 1 เดือน, ปัญหาในการเดินหรือขึ้นบันได, เพศ, อายุ, รายได้
	การถดถอยทีละขั้น (Stepwise Regression)	โรคความดันโลหิตสูง, คอเลสเทอรอลสูง, รับการตรวจคอเลสเทอรอลทุกๆ 5 ปี, ค่าดัชนีมวลกาย, โรคหลอดเลือดสมอง, โรคหลอดเลือดหัวใจ (CHD) หรือกล้ามเนื้อหัวใจตาย (MI), ภาวะติดแอลกอฮอล์, ระดับคะแนนสุขภาพ, จำนวนวันที่

คัดเลือกตัวแปรอิสระ	วิธี	ตัวแปรอิสระ
		สุขภาพจิตแย่ใน 1 เดือน, จำนวนวันที่เจ็บป่วยหรือขาดงานใน 1 เดือน, ปัญหาในการเดินหรือขึ้นบันได, เพศ, อายุ, รายได้
ค่าสัมประสิทธิ์สหสัมพันธ์	สเปียร์แมน (Spearman)	โรคความดันโลหิตสูง, คอเลสเตอรอลสูง, รับการตรวจคอเลสเตอรอลทุกๆ 5 ปี, ค่าดัชนีมวลกาย, สูบบุหรี่อย่างน้อย 100 มวน, โรคหลอดเลือดสมอง, โรคหลอดเลือดหัวใจ (CHD) หรือกล้ามเนื้อหัวใจตาย (MI), การออกกำลังกาย, ภาวะติดแอลกอฮอล์, ระดับคะแนนสุขภาพ, จำนวนวันที่สุขภาพจิตแย่ใน 1 เดือน, จำนวนวันที่เจ็บป่วยหรือขาดงานใน 1 เดือน, ปัญหาในการเดินหรือขึ้นบันได, อายุ, ระดับการศึกษา, รายได้

จากตารางที่ 17 การคัดเลือกตัวแปรอิสระด้วยวิธีการทั้ง 3 รูปแบบสำหรับชุดข้อมูลที่ 2 พบว่า ตัวแปรอิสระที่ถูกคัดออกจากตัวแบบเหมือนกัน ได้แก่ ทานผลไม้อย่างน้อย 1 ครั้งต่อสัปดาห์ (Fruits), การคุ้มครองด้านการดูแลสุขภาพ (AnyHealthcare) และไม่สามารถพบแพทย์ได้เนื่องจากปัญหาค่าใช้จ่าย (NoDocbcCost) ซึ่งการคัดเลือกทั้ง 3 รูปแบบได้ตัวแปรอิสระที่แตกต่างกัน โดยรูปแบบที่ 1 มีจำนวนตัวแปรอิสระที่ถูกคัดเข้าตัวแบบ 14 ตัว และรูปแบบที่ 2, 3 มีจำนวนตัวแปรอิสระที่ถูกคัดเข้าตัวแบบ 16 ตัว

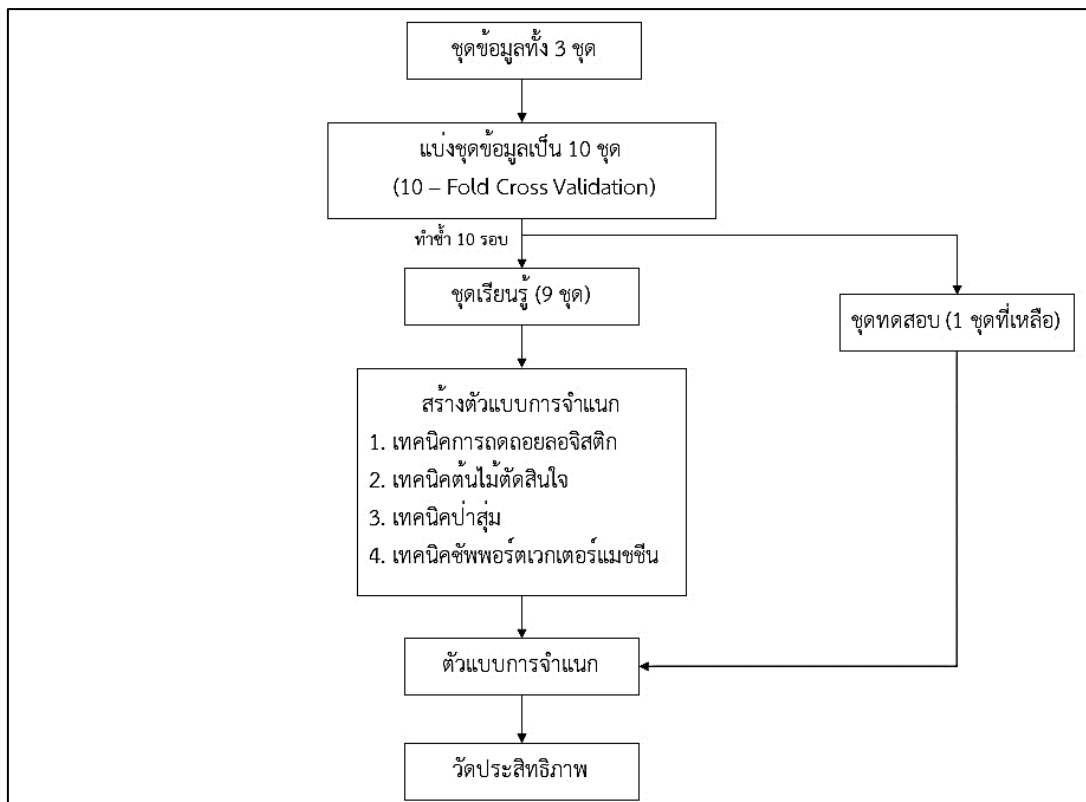
ตาราง 18 การคัดเลือกตัวแปรอิสระในข้อมูลชุดที่ 3 ด้วยวิธีการทั้ง 3 รูปแบบ

คัดเลือกตัวแปรอิสระ	วิธี	ตัวแปรอิสระ
เทคนิคการถดถอยลอจิสติก โดยใช้ข้อมูลทั้งหมด	การเลือกแบบไปข้างหน้า (Forward Selection)	ส่วนสูง, น้ำหนัก, รอบเอว, ค่าความดันขณะหัวใจบีบตัว, ค่าน้ำตาลในเลือด, ค่าคอเลสเตอรอล, ค่าไตรกลีเซอไรด์, ค่าคอเลสเตอรอลชนิดที่ดี, ค่าฮีโมโกลบิน,

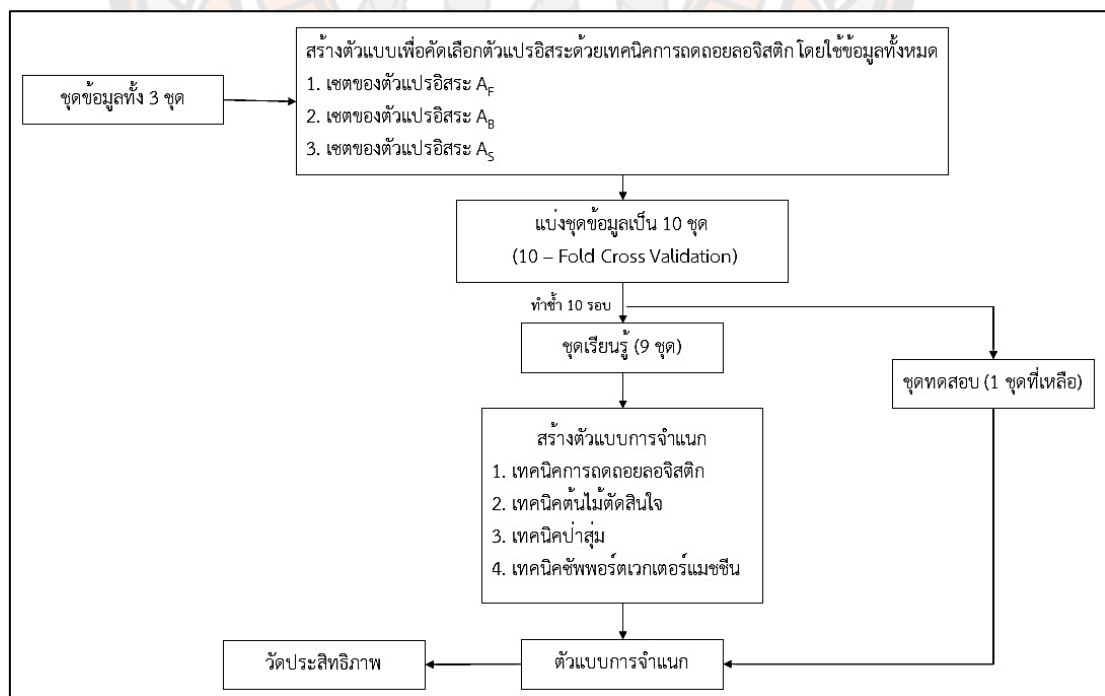
คัดเลือกตัวแปรอิสระ	วิธี	ตัวแปรอิสระ
		ซีรัมครีเอตินิน, ค่าเอนไซม์ AST ในเลือด, ค่าเอนไซม์ ALT ในเลือด, ค่าจีโปรตีน, โรคพิษณุ
กำจัดแบบถอยหลัง (Backward Elimination)		ส่วนสูง, น้ำหนัก, รอบเอว, ค่าความดัน ขณะหัวใจบีบตัว, ค่าน้ำตาลในเลือด, ค่า ไตรกลีเซอไรด์, ค่าคอเลสเตอรอลชนิดที่ ดี, ค่าคอเลสเตอรอลชนิดที่ไม่ดี, ค่า ฮีโมโกลบิน, ซีรัมครีเอตินิน, ค่าเอนไซม์ AST ในเลือด, ค่าเอนไซม์ ALT ในเลือด, ค่าจีโปรตีน, โรคพิษณุ
	การถดถอยทีละขั้น (Stepwise Regression)	ส่วนสูง, น้ำหนัก, รอบเอว, ค่าความดัน ขณะหัวใจบีบตัว, ค่าน้ำตาลในเลือด, ค่า คอเลสเตอรอล, ค่าไตรกลีเซอไรด์, ค่า คอเลสเตอรอลชนิดที่ดี, ค่าฮีโมโกลบิน, ซีรัมครีเอตินิน, ค่าเอนไซม์ AST ในเลือด, ค่าเอนไซม์ ALT ในเลือด, ค่าจีโปรตีน, โรค พิษณุ
เทคนิคการถดถอยโล- จิสติก โดยใช้ข้อมูล เรียนรู้ทำซ้ำ 10 ครั้ง	การเลือกแบบไปข้างหน้า (Forward Selection)	ส่วนสูง, น้ำหนัก, รอบเอว, ค่าความดัน ขณะหัวใจบีบตัว, ค่าน้ำตาลในเลือด, ค่า คอเลสเตอรอล, ค่าไตรกลีเซอไรด์, ค่า คอเลสเตอรอลชนิดที่ดี, ค่าฮีโมโกลบิน, ค่า เอนไซม์ ALT ในเลือด, ค่าจีโปรตีน, โรคพิษ ณุ
	กำจัดแบบถอยหลัง (Backward Elimination)	ส่วนสูง, น้ำหนัก, รอบเอว, ค่าความดัน ขณะหัวใจบีบตัว, ค่าน้ำตาลในเลือด, ค่า ไตรกลีเซอไรด์, ค่าคอเลสเตอรอลชนิดที่ ดี, ค่าฮีโมโกลบิน, ค่าเอนไซม์ ALT ในเลือด , ค่าจีโปรตีน, โรคพิษณุ

คัดเลือกตัวแปรอิสระ	วิธี	ตัวแปรอิสระ
	การถดถอยทีละขั้น (Stepwise Regression)	ส่วนสูง, น้ำหนัก, รอบเอว, ค่าความดัน ขณะหัวใจบีบตัว, ค่าน้ำตาลในเลือด, ค่า คอเลสเตอรอล, ค่าไตรกลีเซอไรด์, ค่า คอเลสเตอรอลชนิดที่ดี, ค่าฮีโมโกลบิน, ค่า เอนไซม์ ALT ในเลือด, ค่าจีโปรตีน, โรคฟัน ผุ
ค่าสัมประสิทธิ์ สหสัมพันธ์	สเปียร์แมน (Spearman)	อายุ, ส่วนสูง, น้ำหนัก, รอบเอว, ความยาว สายตาด้านซ้าย, ความยาวสายตาด้านขวา, ค่าความดันของหัวใจคลายตัว, ค่าไตรกลีเซ อไรด์, ค่าคอเลสเตอรอลชนิดที่ดี, ค่า ฮีโมโกลบิน, ซีรัมครีเอตินิน, ค่าเอนไซม์ ALT ในเลือด, ค่าจีโปรตีน, โรคฟันผุ

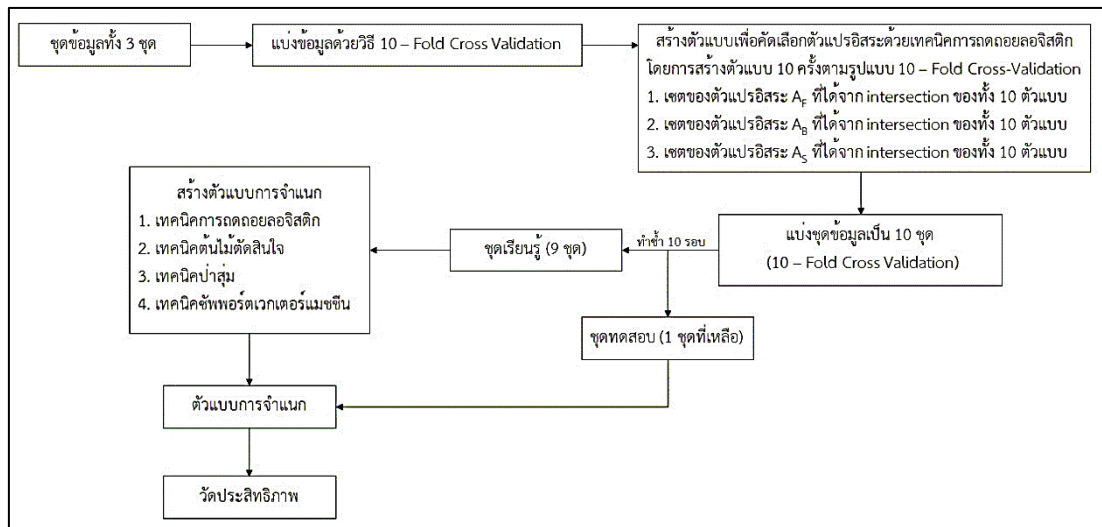
จากตารางที่ 18 การคัดเลือกตัวแปรอิสระทั้ง 3 รูปแบบสำหรับชุดข้อมูลที่ 3 พบว่า ตัวแปรอิสระที่ไม่ถูกคัดเลือกเข้าตัวแบบเหมือนกัน ได้แก่ การได้ยินของหูด้านซ้าย (hearing_left), การได้ยินของหูด้านขวา (hearing_right) และโปรตีนในปัสสาวะ (Urine_protein) ซึ่งการคัดเลือกทั้ง 3 รูปแบบได้ตัวแปรอิสระที่แตกต่างกัน โดยรูปแบบที่ 1 คัดเลือกด้วยวิธีการเลือกแบบไปข้างหน้า และวิธีการถดถอยทีละขั้นมีจำนวนตัวแปรอิสระที่ถูกคัดเลือกเข้าตัวแบบ 12 ตัว ส่วนวิธีการจัดแบบถดถอยหลังมีจำนวนตัวแปรอิสระที่ถูกคัดเลือกเข้าตัวแบบ 11 ตัว และรูปแบบที่ 2, 3 มีจำนวนตัวแปรอิสระที่ถูกคัดเลือกเข้าตัวแบบ 14 ตัว



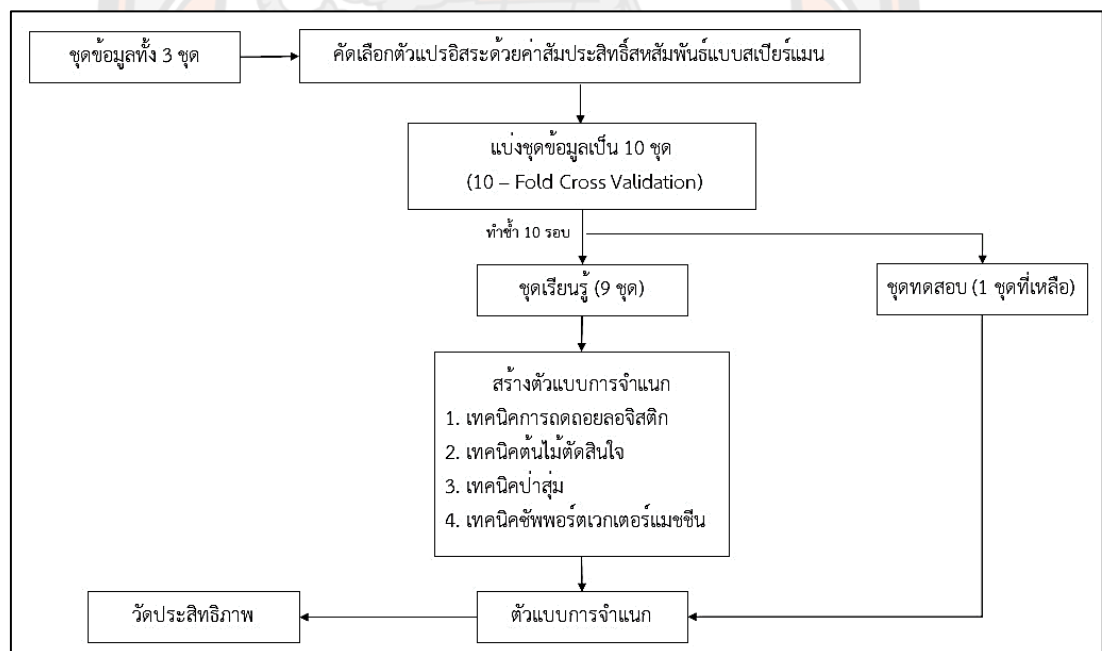
ภาพ 10 ผังงานกระบวนการสร้างตัวแบบการจำแนกโดยใช้ตัวแปรอิสระทั้งหมด



ภาพ 11 ผังงานกระบวนการสร้างตัวแบบการจำแนกโดยการคัดเลือกตัวแปรอิสระด้วยเทคนิคการลดถอยลอจิสติก โดยใช้ข้อมูลทั้งหมด



ภาพ 12 ผังงานกระบวนการสร้างตัวแบบการจำแนกโดยการคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติก โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation

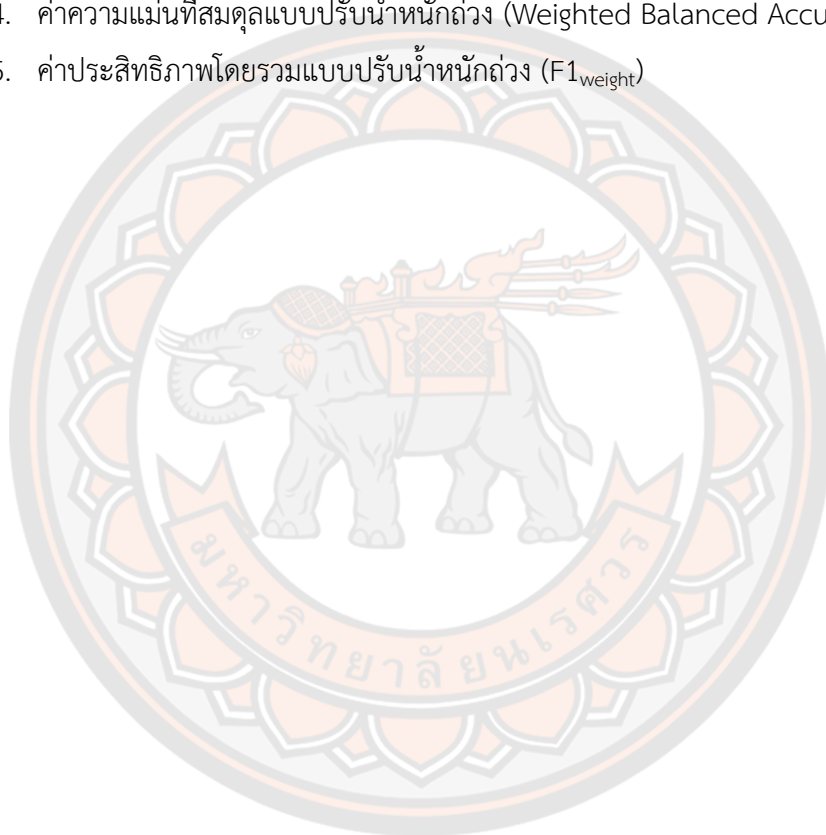


ภาพ 13 ผังงานกระบวนการสร้างตัวแบบการจำแนกโดยการคัดเลือกตัวแปรอิสระด้วยค่าสัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมน

3.5. การประเมินประสิทธิภาพ

ประเมินประสิทธิภาพโดยใช้ค่าที่ได้จากตารางเมทริกซ์ความสับสน (Confusion Matrix) ซึ่งเป็นตารางสรุปการจำแนกประเภทข้อมูลของตัวแบบการจำแนก โดยสามารถวัดประสิทธิภาพของตัวแบบการจำแนกจากเกณฑ์การวัดดังนี้

1. ค่าความไว (Sensitivity)
2. ค่าความจำเพาะ (Specificity)
3. ค่าความแม่นยำ (Accuracy)
4. ค่าความแม่นยำที่สมดุลแบบปรับน้ำหนักถ่วง (Weighted Balanced Accuracy)
5. ค่าประสิทธิภาพโดยรวมแบบปรับน้ำหนักถ่วง ($F1_{\text{weight}}$)



บทที่ 4 ผลการวิจัย

ในบทนี้จะนำเสนอผลการวิจัยที่ได้จากการพัฒนาตัวแบบการจำแนกของชุดข้อมูลโรคหัวใจและหลอดเลือด ชุดข้อมูลโรคเบาหวาน และชุดข้อมูลการเลิกสูบบุหรี่ โดยใช้เทคนิคการจำแนกทั้งหมด 4 เทคนิค ได้แก่ เทคนิคการถดถอยลอจิสติก (Logistic Regression : LR) เทคนิคต้นไม้ตัดสินใจ (Decision Tree : DT) เทคนิคป่าสุ่ม (Random Forest : RF) และเทคนิคซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine : SVM) และคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติก และค่าสัมประสิทธิ์สหสัมพันธ์ ในการสร้างตัวแบบการจำแนก และทดสอบประสิทธิภาพของตัวแบบการจำแนกด้วยเทคนิค 10 – Fold Cross Validation ซึ่งมีเกณฑ์ที่ใช้วัดประสิทธิภาพ คือ ค่าความแม่นยำ (Accuracy : Acc.) ค่าความไว (Sensitivity : Sen.) ค่าความจำเพาะ (Specificity : Spe.) ค่าความแม่นยำที่สมดุลแบบปรับน้ำหนักถ่วง (Weighted Balanced Accuracy : BAcc.) และค่าประสิทธิภาพโดยรวมแบบปรับน้ำหนักถ่วง ($F1_{weight}$) ซึ่งได้ผลการวิเคราะห์ข้อมูลดังนี้

4.1. ผลการเปรียบเทียบประสิทธิภาพของตัวแบบการจำแนก โดยใช้ตัวแปรอิสระทั้งหมด

ในหัวข้อนี้ผู้วิจัยจะนำเสนอผลการเปรียบเทียบประสิทธิภาพของตัวแบบการจำแนกข้อมูลโดยใช้ตัวแปรอิสระทั้งหมด สำหรับชุดข้อมูลโรคหัวใจและหลอดเลือด ชุดข้อมูลโรคเบาหวาน และชุดข้อมูลการเลิกสูบบุหรี่ ดังตารางที่ 19 – 21 ตามลำดับ

ตาราง 19 ผลการจำแนกข้อมูลภายใต้ชุดข้อมูลที่ใช้ตัวแปรอิสระทั้งหมด ในชุดข้อมูลโรคหัวใจและหลอดเลือด

เทคนิค	ชุดเรียนรู้					ชุดทดสอบ					
	การ จำแนก	Acc. (%)	Sen. (%)	Spe. (%)	BAcc. (%)	F1 _{weight} (%)	Acc. (%)	Sen. (%)	Spe. (%)	BAcc. (%)	F1 _{weight} (%)
LR		72.73	66.69	78.64	72.73	72.58	72.71	66.66	78.63	72.71	72.56
DT		75.23	71.04	79.33	75.23	75.19	73.02	68.89	<u>77.06</u>	73.02	72.97
RF (3, 500)*		<u>88.22</u>	<u>91.53</u>	<u>85.47</u>	<u>88.47</u>	<u>88.19</u>	73.16	74.55	71.98	73.25	73.12
SVM (0.1, 1)*		73.70	76.61	71.49	74.02	73.59	<u>73.45</u>	<u>76.32</u>	71.25	<u>73.76</u>	<u>73.34</u>

* RF (จำนวนตัวแปรอิสระ, จำนวนต้นไม้) และ SVM (ค่าแกมมา, ค่าคงที่)

จากตารางที่ 19 เมื่อพิจารณาข้อมูลภายใต้ชุดข้อมูลที่ใช้ตัวแปรอิสระทั้งหมด ในชุดข้อมูลโรคหัวใจและหลอดเลือด พบว่าในชุดข้อมูลเรียนรู้เทคนิคป่าสุ่มให้ค่าประสิทธิภาพทุกเกณฑ์สูงสุด (ค่าความแม่นยำ ค่าความไว ค่าความจำเพาะ ค่าความแม่นยำที่สมดุลแบบปรับน้ำหนักถ่วง และค่าประสิทธิภาพโดยรวมแบบปรับน้ำหนักถ่วง) แต่ในชุดทดสอบพบว่าเทคนิคซัพพอร์ตเวกเตอร์แมชชีนให้ค่าสูงสุดในทุกค่า ยกเว้นค่าความจำเพาะที่เทคนิคต้นไม้ตัดสินใจมีค่าสูงสุด และเมื่อเปรียบเทียบประสิทธิภาพการจำแนกข้อมูลด้วยค่าความแม่นยำที่สมดุลแบบปรับน้ำหนักถ่วง พบว่า เทคนิคซัพพอร์ตเวกเตอร์แมชชีนมีประสิทธิภาพมากที่สุด มีค่าเท่ากับ 73.76%

ตาราง 20 ผลการจำแนกข้อมูลภายใต้ชุดข้อมูลที่ใช้ตัวแปรอิสระทั้งหมด ในชุดข้อมูลโรคเบาหวาน

เทคนิคการ จำแนก	ชุดเรียนรู้					ชุดทดสอบ				
	Acc. (%)	Sen. (%)	Spe. (%)	BAcc. (%)	F1 _{weight} (%)	Acc. (%)	Sen. (%)	Spe. (%)	BAcc. (%)	F1 _{weight} (%)
LR	74.48	76.90	72.02	74.49	74.45	74.43	76.80	72.01	74.43	74.40
DT	83.72	86.28	81.12	83.72	83.71	71.75	74.30	69.16	71.76	71.73
RF (4, 500)*	<u>94.34</u>	<u>94.38</u>	<u>94.30</u>	<u>94.34</u>	<u>94.34</u>	74.01	<u>78.93</u>	72.00	74.01	73.94
SVM (0.01,1)*	74.85	71.58	79.50	75.64	74.66	<u>74.74</u>	71.50	<u>79.37</u>	<u>75.40</u>	<u>74.55</u>

* RF (จำนวนตัวแปรอิสระ, จำนวนต้นไม้) และ SVM (ค่าแกมมา, ค่าคงที่)

จากตารางที่ 20 เมื่อพิจารณาข้อมูลภายใต้ชุดข้อมูลที่ใช้ตัวแปรอิสระทั้งหมด ในชุดข้อมูลโรคเบาหวาน พบว่าในชุดข้อมูลเรียนรู้เทคนิคป่าสุ่มให้ค่าประสิทธิภาพทุกเกณฑ์สูงสุด (ค่าความแม่นยำ ค่าความไว ค่าความจำเพาะ ค่าความแม่นยำที่สมดุลแบบปรับน้ำหนักถ่วง และค่าประสิทธิภาพโดยรวมแบบปรับน้ำหนักถ่วง) แต่ในชุดทดสอบพบว่าเทคนิคซัพพอร์ตเวกเตอร์แมชชีนให้ค่าสูงสุดในทุกค่า ยกเว้นค่าความไวที่เทคนิคป่าสุ่มมีค่าสูงสุด และเมื่อเปรียบเทียบประสิทธิภาพการจำแนกข้อมูลด้วยค่าความแม่นยำที่สมดุลแบบปรับน้ำหนักถ่วง พบว่า เทคนิคซัพพอร์ตเวกเตอร์แมชชีนมีประสิทธิภาพมากที่สุด มีค่าเท่ากับ 75.40%

ตาราง 21 ผลการจำแนกข้อมูลภายใต้ชุดข้อมูลที่ใช้ตัวแปรอิสระทั้งหมด ในชุดข้อมูลการเลิกสูบบุหรี่

เทคนิคการ จำแนก	ชุดเรียนรู้					ชุดทดสอบ				
	Acc. (%)	Sen. (%)	Spe. (%)	BAcc. (%)	F1 _{weight} (%)	Acc. (%)	Sen. (%)	Spe. (%)	BAcc. (%)	F1 _{weight} (%)
LR	73.78	52.94	84.33	73.78	65.46	73.64	52.72	84.24	73.64	65.28
DT	90.32	82.98	94.04	90.32	90.26	70.80	53.82	79.40	70.80	70.59
RF (4, 500)*	<u>100</u>	<u>100</u>	<u>100</u>	<u>100</u>	<u>100</u>	75.32	64.26	<u>80.39</u>	74.97	75.10
SVM (0.01,10)*	77.06	67.48	81.28	76.64	76.77	<u>75.78</u>	<u>65.43</u>	80.34	<u>75.33</u>	<u>75.47</u>

* RF (จำนวนตัวแปรอิสระ, จำนวนต้นไม้) และ SVM (ค่าแกมมา, ค่าคงที่)

จากตารางที่ 21 เมื่อพิจารณาข้อมูลภายใต้ชุดข้อมูลที่ใช้ตัวแปรอิสระทั้งหมด ในชุดข้อมูลการเลิกสูบบุหรี่ พบว่าในชุดข้อมูลเรียนรู้เทคนิคป่าสุ่มให้ค่าประสิทธิภาพทุกเกณฑ์สูงสุด (ค่าความแม่นยำ ค่าความไว ค่าความจำเพาะ ค่าความแม่นยำที่สมดุลแบบปรับน้ำหนักถ่วง และค่าประสิทธิภาพโดยรวมแบบปรับน้ำหนักถ่วง) แต่ในชุดทดสอบพบว่าเทคนิคซัพพอร์ตเวกเตอร์แมชชีนให้ค่าสูงสุดในทุกค่า ยกเว้นค่าความจำเพาะที่เทคนิคป่าสุ่มมีค่าสูงสุด และเมื่อเปรียบเทียบประสิทธิภาพการจำแนกข้อมูลด้วยค่าความแม่นยำที่สมดุลแบบปรับน้ำหนักถ่วง พบว่า เทคนิคซัพพอร์ตเวกเตอร์แมชชีนมีประสิทธิภาพมากที่สุด มีค่าเท่ากับ 75.33% เนื่องจากชุดข้อมูลการเลิกสูบบุหรี่เป็นชุดข้อมูลที่มีสัดส่วนคำตอบ Yes และ No แตกต่างกัน พบว่าค่าความไว และค่าความจำเพาะแตกต่างกัน จึงเป็นข้อสังเกต เมื่อชุดข้อมูลมีจำนวนคลาสคำตอบที่แตกต่างกันมาก (ชุดข้อมูลสมดุล)

จากตารางที่ 19 – 21 สามารถสรุปได้ว่า เทคนิคป่าสุ่มให้ค่าประสิทธิภาพสูงสุดในชุดเรียนรู้ แต่มีประสิทธิผลลดลงในชุดทดสอบ เนื่องจากเกิด Overfit กับข้อมูลในชุดเรียนรู้สูง จึงทำให้ตัวแบบที่นำมาใช้กับชุดทดสอบให้ค่าประสิทธิภาพที่ลดลง และเทคนิคซัพพอร์ตเวกเตอร์แมชชีนมีประสิทธิภาพมากที่สุดในชุดข้อมูลทดสอบทั้ง 3 ชุด

4.2. ผลการเปรียบเทียบประสิทธิภาพของตัวแบบการจำแนก โดยการคัดเลือกตัวแปรอิสระ

ในหัวข้อนี้ผู้วิจัยจะนำเสนอผลการเปรียบเทียบประสิทธิภาพของตัวแบบการจำแนกโดยการคัดเลือกตัวแปรอิสระ สำหรับชุดข้อมูลโรคหัวใจและหลอดเลือด ชุดข้อมูลโรคเบาหวาน และชุดข้อมูลการเลิกสูบบุหรี่ ผู้วิจัยได้กำหนดการคัดเลือกตัวแปรอิสระเป็นดังนี้

4.2.1. ผลการเปรียบเทียบประสิทธิภาพของตัวแบบการจำแนก โดยการคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติก (Logistic Regression)

ในการคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติก ผู้วิจัยได้ออกแบบการคัดเลือกเป็น 2 แบบ ดังต่อไปนี้

แบบที่ 1 การคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติก โดยใช้ข้อมูลทั้งหมด

ผลการพัฒนาตัวแบบการจำแนกด้วยการคัดเลือกตัวแปรอิสระที่ใช้ข้อมูลทั้งหมดด้วยเทคนิคการถดถอยลอจิสติก โดยวิธีการเลือกแบบไปข้างหน้า

ตาราง 22 ผลการจำแนกข้อมูลที่มีการคัดเลือกตัวแปรอิสระที่ใช้ข้อมูลทั้งหมด ด้วยเทคนิคการถดถอยลอจิสติก โดยวิธีการเลือกแบบไปข้างหน้าสำหรับชุดข้อมูลโรคหัวใจและหลอดเลือด

เทคนิคการจำแนก	ชุดเรียนรู้					ชุดทดสอบ				
	Acc. (%)	Sen. (%)	Spe. (%)	BAcc. (%)	F1 _{weight} (%)	Acc. (%)	Sen. (%)	Spe. (%)	BAcc. (%)	F1 _{weight} (%)
LR	72.73	66.67	78.66	72.73	72.63	72.71	66.65	78.63	72.71	72.60
DT	74.75	70.27	79.13	74.75	74.70	73.01	68.55	77.37	73.01	72.95
RF (3, 500)*	<u>89.06</u>	<u>84.76</u>	<u>93.86</u>	<u>89.36</u>	<u>89.04</u>	73.06	<u>69.22</u>	76.81	73.06	73.02
SVM (0.2, 1)*	73.94	67.50	80.25	73.94	73.83	<u>73.41</u>	66.94	<u>79.73</u>	<u>73.41</u>	<u>73.29</u>

* RF (จำนวนตัวแปรอิสระ, จำนวนต้นไม้) และ SVM (ค่าแกมมา, ค่าคงที่)

จากตารางที่ 22 แสดงประสิทธิภาพของตัวแบบจำแนกที่ใช้ข้อมูลทั้งหมด เพื่อคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติก โดยวิธีการเลือกแบบไปข้างหน้า ในชุดข้อมูลโรคหัวใจและหลอดเลือด พบว่าจากข้อมูลที่มีตัวแปรอิสระทั้งหมด 11 ตัวถูกคัดเลือกเพียง 10 ตัว และเมื่อวัดประสิทธิภาพด้วยค่าความแม่นยำที่สมดุลแบบปรับน้ำหนักถ่วง บนชุดทดสอบพบว่าเทคนิคซัพพอร์ตเวกเตอร์แมชชีนมีประสิทธิภาพมากที่สุด มีค่าเท่ากับ 73.41% รองลงมาคือ เทคนิคป่าสุ่ม มีค่าเท่ากับ 73.06% เทคนิคต้นไม้ตัดสินใจ มีค่าเท่ากับ 73.01% และเทคนิคการถดถอยลอจิสติก มีค่าเท่ากับ 72.71% ตามลำดับ

ตาราง 23 ผลการจำแนกข้อมูลที่มีการคัดเลือกตัวแปรอิสระที่ใช้ข้อมูลทั้งหมด ด้วยเทคนิคการถดถอยลอจิสติก โดยวิธีการเลือกแบบไปข้างหน้าสำหรับชุดข้อมูลโรคเบาหวาน

เทคนิคการ จำแนก	ชุดเรียนรู้					ชุดทดสอบ				
	Acc. (%)	Sen. (%)	Spe. (%)	BAcc. (%)	F1 _{weight} (%)	Acc. (%)	Sen. (%)	Spe. (%)	BAcc. (%)	F1 _{weight} (%)
LR	74.48	76.88	72.02	74.48	74.46	74.43	76.81	<u>71.99</u>	74.43	74.41
DT	81.23	84.59	77.80	81.23	81.20	72.44	75.92	68.90	72.44	72.40
RF (4, 500)*	<u>92.51</u>	<u>92.94</u>	<u>92.06</u>	<u>92.50</u>	<u>92.50</u>	73.77	78.48	68.97	73.77	73.70
SVM (0.01,10)*	74.92	82.86	66.84	74.98	74.75	<u>74.76</u>	<u>82.67</u>	66.69	<u>74.79</u>	<u>74.58</u>

* RF (จำนวนตัวแปรอิสระ, จำนวนต้นไม้) และ SVM (ค่าแกมมา, ค่าคงที่)

จากตารางที่ 23 แสดงประสิทธิภาพของตัวแบบจำแนกที่ใช้ข้อมูลทั้งหมด เพื่อคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติก โดยวิธีการเลือกแบบไปข้างหน้า ในชุดข้อมูลโรคเบาหวานพบว่าจากข้อมูลที่มีตัวแปรอิสระทั้งหมด 21 ตัวถูกคัดเหลือเพียง 16 ตัว และเมื่อวัดประสิทธิภาพด้วยค่าความแม่นยำที่สมดุลแบบปรับน้ำหนักถ่วง บนชุดทดสอบพบว่าเทคนิคซัพพอร์ตเวกเตอร์แมชชีนมีประสิทธิภาพมากที่สุด มีค่าเท่ากับ 74.79% รองลงมาคือ เทคนิคการถดถอยลอจิสติก มีค่าเท่ากับ 74.43% เทคนิคป่าสุ่ม มีค่าเท่ากับ 73.77% และเทคนิคต้นไม้ตัดสินใจ มีค่าเท่ากับ 72.44% ตามลำดับ

ตาราง 24 ผลการจำแนกข้อมูลที่มีการคัดเลือกตัวแปรอิสระที่ใช้ข้อมูลทั้งหมด ด้วยเทคนิคการถดถอยลอจิสติก โดยวิธีการเลือกแบบไปข้างหน้าสำหรับชุดข้อมูลการเลิกสูบบุหรี่

เทคนิค การ จำแนก	ชุดเรียนรู้					ชุดทดสอบ				
	Acc. (%)	Sen. (%)	Spe. (%)	BAcc. (%)	F1 _{weight} (%)	Acc. (%)	Sen. (%)	Spe. (%)	BAcc. (%)	F1 _{weight} (%)
LR	73.74	52.89	84.30	73.74	73.11	73.66	52.68	84.29	73.66	73.01
DT	85.10	75.69	89.87	85.10	85.02	71.58	55.12	79.91	71.57	71.38
RF (3, 250)*	<u>100</u>	<u>100</u>	<u>100</u>	<u>100</u>	<u>100</u>	74.69	<u>59.59</u>	82.34	74.69	74.51
SVM (0.1, 1)*	79.41	63.86	86.09	78.61	78.36	<u>75.39</u>	59.07	<u>83.65</u>	<u>75.39</u>	<u>75.09</u>

* RF (จำนวนตัวแปรอิสระ, จำนวนต้นไม้) และ SVM (ค่าแกมมา, ค่าคงที่)

จากตารางที่ 24 แสดงประสิทธิภาพของตัวแบบจำแนกที่ใช้ข้อมูลทั้งหมด เพื่อคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติก โดยวิธีการเลือกแบบไปข้างหน้า ในชุดข้อมูลการเลิกสูบบุหรี่ พบว่าจากข้อมูลที่มีตัวแปรอิสระทั้งหมด 22 ตัวถูกคัดเหลือเพียง 14 ตัว และเมื่อวัดประสิทธิภาพด้วยค่าความแม่นยำที่สมดุลแบบปรับน้ำหนักถ่วง บนชุดทดสอบพบว่าเทคนิคซัพพอร์ตเวกเตอร์แมชชีนมีประสิทธิภาพมากที่สุด มีค่าเท่ากับ 75.39% รองลงมาคือ เทคนิคป่าสุ่ม มีค่าเท่ากับ 74.69% เทคนิคการถดถอยลอจิสติก มีค่าเท่ากับ 73.66% และเทคนิคต้นไม้ตัดสินใจ มีค่าเท่ากับ 71.57% ตามลำดับ

ผลการพัฒนาตัวแบบการจำแนกด้วยการคัดเลือกตัวแปรอิสระที่ใช้ข้อมูลทั้งหมดด้วยเทคนิคการถดถอยลอจิสติก โดยวิธีกำจัดแบบถอยหลัง

ตาราง 25 ผลการจำแนกข้อมูลที่มีการคัดเลือกตัวแปรอิสระที่ใช้ข้อมูลทั้งหมด ด้วยเทคนิคการถดถอยลอจิสติก โดยวิธีกำจัดแบบถอยหลัง สำหรับชุดข้อมูลโรคหัวใจและหลอดเลือด

เทคนิคการจำแนก	ชุดเรียนรู้					ชุดทดสอบ				
	Acc. (%)	Sen. (%)	Spe. (%)	BAcc. (%)	F1 _{weight} (%)	Acc. (%)	Sen. (%)	Spe. (%)	BAcc. (%)	F1 _{weight} (%)
LR	72.73	66.67	78.66	72.73	72.63	72.71	66.65	78.63	72.71	72.60
DT	74.75	70.27	79.13	74.75	74.70	73.01	68.55	77.37	73.01	72.95
RF (3, 500)*	<u>89.06</u>	<u>84.76</u>	<u>93.86</u>	<u>89.36</u>	<u>89.04</u>	73.06	<u>69.22</u>	76.81	73.06	73.02
SVM (0.2, 1)*	73.94	67.50	80.25	73.94	73.83	<u>73.41</u>	66.94	<u>79.73</u>	<u>73.41</u>	<u>73.29</u>

* RF (จำนวนตัวแปรอิสระ, จำนวนต้นไม้) และ SVM (ค่าแกมมา, ค่าคงที่)

จากตารางที่ 25 แสดงประสิทธิภาพของตัวแบบจำแนกที่ใช้ข้อมูลทั้งหมด เพื่อคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติก โดยวิธีกำจัดแบบถอยหลัง ในชุดข้อมูลโรคหัวใจและหลอดเลือด พบว่าจากข้อมูลที่มีตัวแปรอิสระทั้งหมด 11 ตัวถูกคัดเหลือเพียง 10 ตัว และเมื่อวัดประสิทธิภาพด้วยค่าความแม่นยำที่สมดุลแบบปรับน้ำหนักถ่วง บนชุดทดสอบพบว่าเทคนิคซัพพอร์ตเวกเตอร์แมชชีนมีประสิทธิภาพมากที่สุด มีค่าเท่ากับ 73.41% รองลงมาคือ เทคนิคป่าสุ่ม มีค่าเท่ากับ 73.06% เทคนิคต้นไม้ตัดสินใจ มีค่าเท่ากับ 73.01% และเทคนิคการถดถอยลอจิสติก มีค่าเท่ากับ 72.71% ตามลำดับ

ตาราง 26 ผลการจำแนกข้อมูลที่มีการคัดเลือกตัวแปรอิสระที่ใช้ข้อมูลทั้งหมด ด้วยเทคนิคการถดถอยลอจิสติก โดยวิธีกำจัดแบบถอยหลัง สำหรับชุดข้อมูลโรคเบาหวาน

เทคนิคการ จำแนก	ชุดเรียนรู้					ชุดทดสอบ				
	Acc. (%)	Sen. (%)	Spe. (%)	BAcc. (%)	F1 _{weight} (%)	Acc. (%)	Sen. (%)	Spe. (%)	BAcc. (%)	F1 _{weight} (%)
LR	74.48	76.88	72.02	74.48	74.46	74.43	76.81	<u>71.99</u>	74.43	74.41
DT	81.23	84.59	77.80	81.23	81.20	72.44	75.92	68.90	72.44	72.40
RF (4, 500)*	<u>92.51</u>	<u>92.94</u>	<u>92.06</u>	<u>92.50</u>	<u>92.50</u>	73.77	78.48	68.97	73.77	73.70
SVM (0.01,10)*	74.92	82.86	66.84	74.98	74.75	<u>74.76</u>	<u>82.67</u>	66.69	<u>74.79</u>	<u>74.58</u>

* RF (จำนวนตัวแปรอิสระ, จำนวนต้นไม้) และ SVM (ค่าแกมมา, ค่าคงที่)

จากตารางที่ 26 แสดงประสิทธิภาพของตัวแบบจำแนกที่ใช้ข้อมูลทั้งหมด เพื่อคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติก โดยวิธีกำจัดแบบถอยหลัง ในชุดข้อมูลโรคเบาหวาน พบว่าจากข้อมูลที่มีตัวแปรอิสระทั้งหมด 21 ตัวถูกคัดเหลือเพียง 16 ตัว และเมื่อวัดประสิทธิภาพด้วยค่าความแม่นยำที่สมดุลแบบปรับน้ำหนักถ่วง บนชุดทดสอบพบว่าเทคนิคซัพพอร์ตเวกเตอร์แมชชีนมีประสิทธิภาพมากที่สุด มีค่าเท่ากับ 74.79% รองลงมาคือ เทคนิคการถดถอยลอจิสติก มีค่าเท่ากับ 74.43% เทคนิคป่าสุ่ม มีค่าเท่ากับ 73.77% และเทคนิคต้นไม้ตัดสินใจ มีค่าเท่ากับ 72.44% ตามลำดับ

ตาราง 27 ผลการจำแนกข้อมูลที่มีการคัดเลือกตัวแปรอิสระที่ใช้ข้อมูลทั้งหมด ด้วยเทคนิคการถดถอยลอจิสติก โดยวิธีกำจัดแบบถอยหลัง สำหรับชุดข้อมูลการเลิกสูบบุหรี่

เทคนิคการ จำแนก	ชุดเรียนรู้					ชุดทดสอบ				
	Acc. (%)	Sen. (%)	Spe. (%)	BAcc. (%)	F1 _{weight} (%)	Acc. (%)	Sen. (%)	Spe. (%)	BAcc. (%)	F1 _{weight} (%)
LR	73.75	52.92	84.30	73.75	73.12	73.66	52.70	84.28	73.66	73.02
DT	85.21	76.12	89.81	85.21	85.14	71.50	55.41	79.64	71.50	71.33
RF (3, 250)*	<u>100</u>	<u>100</u>	<u>100</u>	<u>100</u>	<u>100</u>	74.72	<u>59.38</u>	82.49	74.72	74.51
SVM (0.01,10)*	75.99	59.01	84.59	75.99	75.63	<u>75.41</u>	58.13	<u>84.17</u>	<u>75.41</u>	<u>75.04</u>

* RF (จำนวนตัวแปรอิสระ, จำนวนต้นไม้) และ SVM (ค่าแกมมา, ค่าคงที่)

จากตารางที่ 27 แสดงประสิทธิภาพของตัวแบบจำแนกที่ใช้ข้อมูลทั้งหมด เพื่อคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติก โดยวิธีกำจัดแบบถอยหลัง ในชุดข้อมูลการเลิกสูบบุหรี่ พบว่าจากข้อมูลที่มีตัวแปรอิสระทั้งหมด 22 ตัวถูกคัดเหลือเพียง 14 ตัว และเมื่อวัดประสิทธิภาพด้วยค่าความแม่นยำที่สมดุลแบบปรับน้ำหนักถ่วง บนชุดทดสอบพบว่าเทคนิคซัพพอร์ตเวกเตอร์แมชชีนมีประสิทธิภาพมากที่สุด มีค่าเท่ากับ 75.41% รองลงมาคือ เทคนิคป่าสุ่ม มีค่าเท่ากับ 74.72% เทคนิคการถดถอยลอจิสติก มีค่าเท่ากับ 73.66% และเทคนิคต้นไม้ตัดสินใจ มีค่าเท่ากับ 71.50% ตามลำดับ

ผลการพัฒนาตัวแบบการจำแนกด้วยการคัดเลือกคุณลักษณะที่ใช้ข้อมูลทั้งหมดด้วยเทคนิคการถดถอยลอจิสติก โดยวิธีการถดถอยทีละขั้น

ตาราง 28 ผลการจำแนกข้อมูลที่มีการคัดเลือกตัวแปรอิสระที่ใช้ข้อมูลทั้งหมด ด้วยเทคนิคการถดถอยลอจิสติก โดยวิธีการถดถอยทีละขั้น สำหรับชุดข้อมูลโรคหัวใจและหลอดเลือด

เทคนิคการจำแนก	ชุดเรียนรู้					ชุดทดสอบ				
	Acc. (%)	Sen. (%)	Spe. (%)	BAcc. (%)	F1 _{weight} (%)	Acc. (%)	Sen. (%)	Spe. (%)	BAcc. (%)	F1 _{weight} (%)
LR	72.73	66.67	78.66	72.73	72.63	72.71	66.65	78.63	72.71	72.60
DT	74.75	70.27	79.13	74.75	74.70	73.01	68.55	77.37	73.01	72.95
RF (3, 500)*	<u>89.06</u>	<u>84.76</u>	<u>93.86</u>	<u>89.36</u>	<u>89.04</u>	73.06	<u>69.22</u>	76.81	73.06	73.02
SVM (0.2, 1)*	73.94	67.50	80.25	73.94	73.83	<u>73.41</u>	66.94	<u>79.73</u>	<u>73.41</u>	<u>73.29</u>

* RF (จำนวนตัวแปรอิสระ, จำนวนต้นไม้) และ SVM (ค่าแกมมา, ค่าคงที่)

จากตารางที่ 28 แสดงประสิทธิภาพของตัวแบบจำแนกที่ใช้ข้อมูลทั้งหมด เพื่อคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติก โดยวิธีการถดถอยทีละขั้น ในชุดข้อมูลโรคหัวใจและหลอดเลือด พบว่าจากข้อมูลที่มีตัวแปรอิสระทั้งหมด 11 ตัวถูกคัดเหลือเพียง 10 ตัว และเมื่อวัดประสิทธิภาพด้วยค่าความแม่นยำที่สมดุลแบบปรับน้ำหนักถ่วง บนชุดทดสอบพบว่าเทคนิคซัพพอร์ตเวกเตอร์แมชชีนมีประสิทธิภาพมากที่สุด มีค่าเท่ากับ 73.41% รองลงมาคือ เทคนิคป่าสุ่ม มีค่าเท่ากับ 73.06% เทคนิคต้นไม้ตัดสินใจ มีค่าเท่ากับ 73.01% และเทคนิคการถดถอยลอจิสติก มีค่าเท่ากับ 72.71% ตามลำดับ

ตาราง 29 ผลการจำแนกข้อมูลที่มีการคัดเลือกตัวแปรอิสระที่ใช้ข้อมูลทั้งหมด ด้วยเทคนิคการถดถอยลอจิสติก โดยวิธีการถดถอยทีละชั้น สำหรับชุดข้อมูลโรคเบาหวาน

เทคนิคการ จำแนก	ชุดเรียนรู้					ชุดทดสอบ				
	Acc. (%)	Sen. (%)	Spe. (%)	BAcc. (%)	F1 _{weight} (%)	Acc. (%)	Sen. (%)	Spe. (%)	BAcc. (%)	F1 _{weight} (%)
LR	74.48	76.88	72.02	74.48	74.46	74.43	76.81	<u>71.99</u>	74.43	74.41
DT	81.23	84.59	77.80	81.23	81.20	72.44	75.92	68.90	72.44	72.40
RF (4, 500)*	<u>92.51</u>	<u>92.94</u>	<u>92.06</u>	<u>92.50</u>	<u>92.50</u>	73.77	78.48	68.97	73.77	73.70
SVM (0.01,10)*	74.92	82.86	66.84	74.98	74.75	<u>74.76</u>	<u>82.67</u>	66.69	<u>74.79</u>	<u>74.58</u>

* RF (จำนวนตัวแปรอิสระ, จำนวนต้นไม้) และ SVM (ค่าแกมมา, ค่าคงที่)

จากตารางที่ 29 แสดงประสิทธิภาพของตัวแบบจำแนกที่ใช้ข้อมูลทั้งหมด เพื่อคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติก โดยวิธีการถดถอยทีละชั้น ในชุดข้อมูลโรคเบาหวาน พบว่าจากข้อมูลที่มีตัวแปรอิสระทั้งหมด 21 ตัวถูกคัดเหลือเพียง 16 ตัว และเมื่อวัดประสิทธิภาพด้วยค่าความแม่นยำที่สมดุลแบบปรับน้ำหนักถ่วง บนชุดทดสอบพบว่าเทคนิคซัพพอร์ตเวกเตอร์แมชชีนมีประสิทธิภาพมากที่สุด มีค่าเท่ากับ 74.79% รองลงมาคือ เทคนิคการถดถอยลอจิสติก มีค่าเท่ากับ 74.43% เทคนิคป่าสุ่ม มีค่าเท่ากับ 73.77% และเทคนิคต้นไม้ตัดสินใจ มีค่าเท่ากับ 72.44% ตามลำดับ

ตาราง 30 ผลการจำแนกข้อมูลที่มีการคัดเลือกตัวแปรอิสระที่ใช้ข้อมูลทั้งหมด ด้วยเทคนิคการถดถอยลอจิสติกโดยวิธีการถดถอยทีละชั้น สำหรับชุดข้อมูลการเลิกสูบบุหรี่

เทคนิค การ จำแนก	ชุดเรียนรู้					ชุดทดสอบ				
	Acc. (%)	Sen. (%)	Spe. (%)	BAcc. (%)	F1 _{weight} (%)	Acc. (%)	Sen. (%)	Spe. (%)	BAcc. (%)	F1 _{weight} (%)
LR	73.74	52.89	84.30	73.74	73.11	73.66	52.68	84.29	73.66	73.01
DT	85.10	75.69	89.87	85.10	85.02	71.58	55.12	79.91	71.57	71.38
RF (3, 250)*	<u>100</u>	<u>100</u>	<u>100</u>	<u>100</u>	<u>100</u>	74.69	<u>59.59</u>	82.34	74.69	74.51
SVM (0.1,1)*	79.41	63.86	86.09	78.61	78.36	<u>75.39</u>	59.07	<u>83.65</u>	<u>75.39</u>	<u>75.09</u>

* RF (จำนวนตัวแปรอิสระ, จำนวนต้นไม้) และ SVM (ค่าแกมมา, ค่าคงที่)

จากตารางที่ 30 แสดงประสิทธิภาพของตัวแบบจำแนกที่ใช้ข้อมูลทั้งหมด เพื่อคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติก โดยวิธีการถดถอยทีละชั้น ในชุดข้อมูลการเลิกสูบบุหรี่ พบว่าจากข้อมูลที่มีตัวแปรอิสระทั้งหมด 22 ตัวถูกคัดเหลือเพียง 14 ตัว และเมื่อวัดประสิทธิภาพด้วยค่าความแม่นยำที่สมดุลแบบปรับน้ำหนักถ่วง บนชุดทดสอบพบว่าเทคนิคซัพพอร์ตเวกเตอร์แมชชีนมีประสิทธิภาพมากที่สุด มีค่าเท่ากับ 75.39% รองลงมาคือ เทคนิคป่าสุ่ม มีค่าเท่ากับ 74.69% เทคนิคการถดถอยลอจิสติก มีค่าเท่ากับ 73.66% และเทคนิคต้นไม้ตัดสินใจ มีค่าเท่ากับ 71.57% ตามลำดับ

แบบที่ 2 การคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติก โดยใช้ข้อมูลเรียนรู้ทำซ้ำ 10 ครั้ง

ผลการพัฒนาตัวแบบการจำแนกด้วยการคัดเลือกตัวแปรอิสระโดยใช้ข้อมูลเรียนรู้ทำซ้ำ

10 ครั้ง ด้วยเทคนิคการถดถอยลอจิสติก โดยวิธีการเลือกแบบไปข้างหน้า

ตาราง 31 ผลการจำแนกข้อมูลภายใต้ชุดข้อมูลที่มีการคัดเลือกตัวแปรอิสระโดยใช้ข้อมูลเรียนรู้ทำซ้ำ 10 ครั้ง ด้วยเทคนิคการถดถอยลอจิสติก โดยวิธีการเลือกแบบไปข้างหน้า สำหรับชุดข้อมูลโรคหัวใจและหลอดเลือด

เทคนิค	ชุดเรียนรู้					ชุดทดสอบ					
	การ จำแนก	Acc. (%)	Sen. (%)	Spe. (%)	BAcc. (%)	F1 _{weight} (%)	Acc. (%)	Sen. (%)	Spe. (%)	BAcc. (%)	F1 _{weight} (%)
LR		72.73	66.67	78.66	72.73	72.63	72.71	66.65	78.63	72.71	72.60
DT		74.75	70.27	79.13	74.75	74.70	73.01	68.55	77.37	73.01	72.95
RF (3, 500)*		<u>89.06</u>	<u>84.76</u>	<u>93.86</u>	<u>89.36</u>	<u>89.04</u>	73.06	<u>69.22</u>	76.81	73.06	73.02
SVM (0.2, 1)*		73.94	67.50	80.25	73.94	73.83	<u>73.41</u>	66.94	<u>79.73</u>	<u>73.41</u>	<u>73.29</u>

* RF (จำนวนตัวแปรอิสระ, จำนวนต้นไม้) และ SVM (ค่าแกมมา, ค่าคงที่)

จากตารางที่ 31 แสดงประสิทธิภาพของตัวแบบจำแนกที่ใช้ข้อมูลเรียนรู้ทำซ้ำ 10 ครั้ง เพื่อคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติก โดยวิธีการเลือกแบบไปข้างหน้า ในชุดข้อมูลโรคหัวใจและหลอดเลือด พบว่าจากข้อมูลที่มีตัวแปรอิสระทั้งหมด 11 ตัวถูกคัดเหลือเพียง 10 ตัว และเมื่อวัดประสิทธิภาพด้วยค่าความแม่นยำที่สมดุลแบบปรับน้ำหนักถ่วง บนชุดทดสอบพบว่าเทคนิคซัพพอร์ตเวกเตอร์แมชชีนมีประสิทธิภาพมากที่สุด มีค่าเท่ากับ 73.41% รองลงมาคือ เทคนิคป่าสุ่ม มีค่าเท่ากับ 73.06% เทคนิคต้นไม้ตัดสินใจ มีค่าเท่ากับ 73.01% และเทคนิคการถดถอยลอจิสติก มีค่าเท่ากับ 72.71% ตามลำดับ

ตาราง 32 ผลการจำแนกข้อมูลภายใต้ชุดข้อมูลที่มีการคัดเลือกตัวแปรอิสระโดยใช้ข้อมูลเรียนรู้ทำซ้ำ 10 ครั้ง ด้วยเทคนิคการถดถอยลอจิสติก โดยวิธีการเลือกแบบไปข้างหน้า สำหรับชุดข้อมูลโรคเบาหวาน

เทคนิคการ จำแนก	ชุดเรียนรู้					ชุดทดสอบ				
	Acc. (%)	Sen. (%)	Spe. (%)	BAcc. (%)	F1 _{weight} (%)	Acc. (%)	Sen. (%)	Spe. (%)	BAcc. (%)	F1 _{weight} (%)
LR	74.47	76.89	71.99	74.47	74.45	74.42	76.81	<u>71.99</u>	74.42	74.41
DT	79.14	83.39	74.82	79.14	79.10	73.13	77.52	68.67	73.13	73.07
RF (3, 500)*	<u>83.05</u>	<u>86.80</u>	<u>79.22</u>	<u>83.04</u>	<u>83.02</u>	74.42	79.41	69.34	74.42	74.35
SVM (0.01,10)*	74.92	82.86	66.84	74.92	74.75	<u>74.76</u>	<u>82.67</u>	66.69	<u>74.76</u>	<u>74.58</u>

* RF (จำนวนตัวแปรอิสระ, จำนวนต้นไม้) และ SVM (ค่าแกมมา, ค่าคงที่)

จากตารางที่ 32 แสดงประสิทธิภาพของตัวแบบจำแนกที่ใช้ข้อมูลเรียนรู้ทำซ้ำ 10 ครั้ง เพื่อคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติก โดยวิธีการเลือกแบบไปข้างหน้า ในชุดข้อมูลโรคเบาหวาน พบว่าจากข้อมูลที่มีตัวแปรอิสระทั้งหมด 21 ตัวถูกคัดเหลือเพียง 14 ตัว และเมื่อวัดประสิทธิภาพด้วยค่าความแม่นยำที่สมดุลแบบปรับน้ำหนักถ่วง บนชุดทดสอบพบว่าเทคนิคซัพพอร์ตเวกเตอร์แมชชีนมีประสิทธิภาพมากที่สุด มีค่าเท่ากับ 74.76% รองลงมาคือ เทคนิคการถดถอยลอจิสติก มีค่าเท่ากับ 74.42% เทคนิคป่าสุ่ม มีค่าเท่ากับ 74.42% และเทคนิคต้นไม้ตัดสินใจ มีค่าเท่ากับ 73.13% ตามลำดับ

ตาราง 33 ผลการจำแนกข้อมูลภายใต้ชุดข้อมูลที่มีการคัดเลือกตัวแปรอิสระโดยใช้ข้อมูลเรียนรู้ทำซ้ำ 10 ครั้ง ด้วยเทคนิคการถดถอยลอจิสติก โดยวิธีการเลือกแบบไปข้างหน้า สำหรับชุดข้อมูลการเลิกสูบบุหรี่

เทคนิคการ จำแนก	ชุดเรียนรู้					ชุดทดสอบ				
	Acc. (%)	Sen. (%)	Spe. (%)	BAcc. (%)	F1 _{weight} (%)	Acc. (%)	Sen. (%)	Spe. (%)	BAcc. (%)	F1 _{weight} (%)
LR	73.72	52.70	84.36	73.72	73.07	73.76	52.70	84.42	73.76	73.10
DT	81.84	71.06	87.31	81.84	81.75	72.61	57.08	80.47	72.61	72.45
RF (3, 500)*	<u>100</u>	<u>100</u>	<u>100</u>	<u>100</u>	<u>100</u>	74.44	<u>59.22</u>	82.15	74.44	74.26

เทคนิคการ จำแนก	ชุดเรียนรู้					ชุดทดสอบ				
	Acc. (%)	Sen. (%)	Spe. (%)	BAcc. (%)	F1 _{weight} (%)	Acc. (%)	Sen. (%)	Spe. (%)	BAcc. (%)	F1 _{weight} (%)
SVM (0.01,10)*	75.37	56.83	84.77	75.38	74.91	<u>75.03</u>	56.33	<u>84.51</u>	<u>75.03</u>	<u>74.56</u>

* RF (จำนวนตัวแปรอิสระ, จำนวนต้นไม้) และ SVM (ค่าแกมมา, ค่าคงที่)

จากตารางที่ 33 แสดงประสิทธิภาพของตัวแบบจำแนกที่ใช้ข้อมูลเรียนรู้ทำซ้ำ 10 ครั้ง เพื่อคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติก โดยวิธีการเลือกแบบไปข้างหน้า ในชุดข้อมูลการเลิกสูบบุหรี่ พบว่าจากข้อมูลที่มีตัวแปรอิสระทั้งหมด 22 ตัวถูกคัดเลือกเพียง 12 ตัว และเมื่อวัดประสิทธิภาพด้วยค่าความแม่นยำที่สมดุลแบบปรับน้ำหนักถ่วง บนชุดทดสอบพบว่าเทคนิคซัพพอร์ตเวกเตอร์แมชชีนมีประสิทธิภาพมากที่สุด มีค่าเท่ากับ 75.03% รองลงมาคือ เทคนิคป่าสุ่ม มีค่าเท่ากับ 74.44% เทคนิคการถดถอยลอจิสติก มีค่าเท่ากับ 73.76% และเทคนิคต้นไม้ตัดสินใจ มีค่าเท่ากับ 72.61% ตามลำดับ

ผลการพัฒนาตัวแบบการจำแนกด้วยการคัดเลือกตัวแปรอิสระโดยใช้ข้อมูลเรียนรู้ทำซ้ำ 10 ครั้ง ด้วยเทคนิคการถดถอยลอจิสติก โดยวิธีกำจัดแบบถอยหลัง

ตาราง 34 ผลการจำแนกข้อมูลภายใต้ชุดข้อมูลที่มีการคัดเลือกตัวแปรอิสระโดยใช้ข้อมูลเรียนรู้ทำซ้ำ 10 ครั้ง ด้วยเทคนิคการถดถอยลอจิสติก โดยวิธีกำจัดแบบถอยหลัง สำหรับชุดข้อมูลโรคหัวใจและหลอดเลือด

เทคนิค การ จำแนก	ชุดเรียนรู้					ชุดทดสอบ				
	Acc. (%)	Sen. (%)	Spe. (%)	BAcc. (%)	F1 _{weight} (%)	Acc. (%)	Sen. (%)	Spe. (%)	BAcc. (%)	F1 _{weight} (%)
LR	72.73	66.67	78.66	72.73	72.63	72.71	66.65	78.63	72.71	72.60
DT	74.75	70.27	79.13	74.75	74.70	73.01	68.55	77.37	73.01	72.95
RF (3, 500)*	<u>89.06</u>	<u>84.76</u>	<u>93.86</u>	<u>89.36</u>	<u>89.04</u>	73.06	<u>69.22</u>	76.81	73.06	73.02
SVM (0.2, 1)*	73.94	67.50	80.25	73.94	73.83	<u>73.41</u>	66.94	<u>79.73</u>	<u>73.41</u>	<u>73.29</u>

* RF (จำนวนตัวแปรอิสระ, จำนวนต้นไม้) และ SVM (ค่าแกมมา, ค่าคงที่)

จากตารางที่ 34 แสดงประสิทธิภาพของตัวแบบจำแนกที่ใช้ข้อมูลเรียนรู้ทำซ้ำ 10 ครั้ง เพื่อคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติก โดยวิธีกำจัดแบบถอยหลัง ในชุดข้อมูลโรคหัวใจ

และหลุดเลือด พบว่าจากข้อมูลที่มีตัวแปรอิสระทั้งหมด 11 ตัวถูกคัดเหลือเพียง 10 ตัว และเมื่อวัดประสิทธิภาพด้วยค่าความแม่นยำที่สมดุลแบบปรับน้ำหนักถ่วง บนชุดทดสอบพบว่าเทคนิคซัพพอร์ตเวกเตอร์แมชชีนมีประสิทธิภาพมากที่สุด มีค่าเท่ากับ 73.41% รองลงมาคือ เทคนิคป่าสุ่ม มีค่าเท่ากับ 73.06% เทคนิคต้นไม้ตัดสินใจ มีค่าเท่ากับ 73.01% และเทคนิคการถดถอยลอจิสติก มีค่าเท่ากับ 72.71% ตามลำดับ

ตาราง 35 ผลการจำแนกข้อมูลภายใต้ชุดข้อมูลที่มีการคัดเลือกตัวแปรอิสระโดยใช้ข้อมูลเรียนรู้ทำซ้ำ 10 ครั้ง ด้วยเทคนิคการถดถอยลอจิสติก โดยวิธีกำจัดแบบถอยหลัง สำหรับชุดข้อมูลโรคเบาหวาน

เทคนิคการ จำแนก	ชุดเรียนรู้					ชุดทดสอบ				
	Acc. (%)	Sen. (%)	Spe. (%)	BAcc. (%)	F1 _{weight} (%)	Acc. (%)	Sen. (%)	Spe. (%)	BAcc. (%)	F1 _{weight} (%)
LR	74.47	76.89	71.99	74.47	74.45	74.42	76.81	<u>71.99</u>	74.42	74.41
DT	79.14	83.39	74.82	79.14	79.10	73.13	77.52	68.67	73.13	73.07
RF (3, 500)*	<u>83.05</u>	<u>86.80</u>	<u>79.22</u>	<u>83.04</u>	<u>83.02</u>	74.42	79.41	69.34	74.42	74.35
SVM (0.01,10)*	74.92	82.86	66.84	74.92	74.75	<u>74.76</u>	<u>82.67</u>	66.69	<u>74.76</u>	<u>74.58</u>

* RF (จำนวนตัวแปรอิสระ, จำนวนต้นไม้) และ SVM (ค่าแกมมา, ค่าคงที่)

จากตารางที่ 35 แสดงประสิทธิภาพของตัวแบบจำแนกที่ใช้ข้อมูลเรียนรู้ทำซ้ำ 10 ครั้ง เพื่อคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติก โดยวิธีกำจัดแบบถอยหลัง ในชุดข้อมูลโรคเบาหวาน พบว่าจากข้อมูลที่มีตัวแปรอิสระทั้งหมด 21 ตัวถูกคัดเหลือเพียง 14 ตัว และเมื่อวัดประสิทธิภาพด้วยค่าความแม่นยำที่สมดุลแบบปรับน้ำหนักถ่วง บนชุดทดสอบพบว่าเทคนิคซัพพอร์ตเวกเตอร์แมชชีนมีประสิทธิภาพมากที่สุด มีค่าเท่ากับ 74.76% รองลงมาคือ เทคนิคการถดถอยลอจิสติก มีค่าเท่ากับ 74.42% เทคนิคป่าสุ่ม มีค่าเท่ากับ 74.42% และเทคนิคต้นไม้ตัดสินใจ มีค่าเท่ากับ 73.13% ตามลำดับ

ตาราง 36 ผลการจำแนกข้อมูลภายใต้ชุดข้อมูลที่มีการคัดเลือกตัวแปรอิสระโดยใช้ข้อมูลเรียนรู้ทำซ้ำ 10 ครั้ง ด้วยเทคนิคการถดถอยลอจิสติก โดยวิธีกำจัดแบบถอยหลัง สำหรับชุดข้อมูลการเลิกสูบบุหรี่

เทคนิคการ จำแนก	ชุดเรียนรู้					ชุดทดสอบ				
	Acc. (%)	Sen. (%)	Spe. (%)	BAcc. (%)	F1 _{weight} (%)	Acc. (%)	Sen. (%)	Spe. (%)	BAcc. (%)	F1 _{weight} (%)
LR	73.54	52.09	84.40	73.54	72.85	73.76	52.70	84.42	73.40	72.70
DT	80.94	69.23	86.87	80.94	80.82	72.70	57.04	80.63	72.70	72.53
RF (4, 500)*	<u>100</u>	<u>100</u>	<u>100</u>	<u>100</u>	<u>100</u>	74.26	<u>59.49</u>	81.74	74.26	74.11
SVM (0.01,10)*	75.27	56.28	84.89	75.27	74.77	<u>75.02</u>	55.99	<u>84.65</u>	<u>75.02</u>	<u>74.51</u>

* RF (จำนวนตัวแปรอิสระ, จำนวนต้นไม้) และ SVM (ค่าแกมมา, ค่าคงที่)

จากตารางที่ 36 แสดงประสิทธิภาพของตัวแบบจำแนกที่ใช้ข้อมูลเรียนรู้ทำซ้ำ 10 ครั้ง เพื่อคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติก โดยวิธีกำจัดแบบถอยหลัง ในชุดข้อมูลการเลิกสูบบุหรี่ พบว่าจากข้อมูลที่มีตัวแปรอิสระทั้งหมด 22 ตัวถูกคัดเลือกเหลือเพียง 11 ตัว และเมื่อวัดประสิทธิภาพด้วยค่าความแม่นยำที่สมดุลแบบปรับน้ำหนักถ่วง บนชุดทดสอบพบว่าเทคนิคซัพพอร์ตเวกเตอร์แมชชีนมีประสิทธิภาพมากที่สุด มีค่าเท่ากับ 75.02% รองลงมาคือ เทคนิคป่าสุ่ม มีค่าเท่ากับ 74.26% เทคนิคการถดถอยลอจิสติก มีค่าเท่ากับ 73.40% และเทคนิคต้นไม้ตัดสินใจ มีค่าเท่ากับ 72.70% ตามลำดับ

ผลการพัฒนาตัวแบบการจำแนกด้วยการคัดเลือกตัวแปรอิสระโดยใช้ข้อมูลเรียนรู้ทำซ้ำ 10 ครั้ง ด้วยเทคนิคการถดถอยลอจิสติก โดยวิธีการถดถอยทีละขั้น

ตาราง 37 ผลการจำแนกข้อมูลภายใต้ชุดข้อมูลที่มีการคัดเลือกตัวแปรอิสระโดยใช้ข้อมูลเรียนรู้ทำซ้ำ 10 ครั้ง ด้วยเทคนิคการถดถอยลอจิสติก โดยวิธีการถดถอยทีละขั้น สำหรับชุดข้อมูลโรคหัวใจและหลอดเลือด

เทคนิค การ จำแนก	ชุดเรียนรู้					ชุดทดสอบ				
	Acc. (%)	Sen. (%)	Spe. (%)	BAcc. (%)	F1 _{weight} (%)	Acc. (%)	Sen. (%)	Spe. (%)	BAcc. (%)	F1 _{weight} (%)
LR	72.73	66.67	78.66	72.73	72.63	72.71	66.65	78.63	72.71	72.60
DT	74.75	70.27	79.13	74.75	74.70	73.01	68.55	77.37	73.01	72.95
RF	<u>89.06</u>	<u>84.76</u>	<u>93.86</u>	<u>89.36</u>	<u>89.04</u>	73.06	<u>69.22</u>	76.81	73.06	73.02

เทคนิค	ชุดเรียนรู้					ชุดทดสอบ					
	การ จำแนก	Acc. (%)	Sen. (%)	Spe. (%)	BAcc. (%)	F1 _{weight} (%)	Acc. (%)	Sen. (%)	Spe. (%)	BAcc. (%)	F1 _{weight} (%)
(3, 500)*											
SVM (0.2, 1)*		73.94	67.50	80.25	73.94	73.83	<u>73.41</u>	66.94	<u>79.73</u>	<u>73.41</u>	<u>73.29</u>

* RF (จำนวนตัวแปรอิสระ, จำนวนต้นไม้) และ SVM (ค่าแกมมา, ค่าคงที่)

จากตารางที่ 37 แสดงประสิทธิภาพของตัวแบบจำแนกที่ใช้ข้อมูลเรียนรู้ทำซ้ำ 10 ครั้ง เพื่อคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติก โดยวิธีการถดถอยทีละชั้น ในชุดข้อมูลโรคหัวใจ และหลอดเลือด พบว่าจากข้อมูลที่มีตัวแปรอิสระทั้งหมด 11 ตัวถูกคัดเลือกเหลือเพียง 10 ตัว และเมื่อวัดประสิทธิภาพด้วยค่าความแม่นยำที่สมดุลแบบปรับน้ำหนักถ่วง บนชุดทดสอบพบว่าเทคนิคซัพพอร์ตเวกเตอร์แมชชีนมีประสิทธิภาพมากที่สุด มีค่าเท่ากับ 73.41% รองลงมาคือ เทคนิคป่าสุ่ม มีค่าเท่ากับ 73.06% เทคนิคต้นไม้ตัดสินใจ มีค่าเท่ากับ 73.01% และเทคนิคการถดถอยลอจิสติก มีค่าเท่ากับ 72.71% ตามลำดับ

ตาราง 38 ผลการจำแนกข้อมูลภายใต้ชุดข้อมูลที่มีการคัดเลือกตัวแปรอิสระโดยใช้ข้อมูลเรียนรู้ทำซ้ำ 10 ครั้ง ด้วยเทคนิคการถดถอยลอจิสติก โดยวิธีการถดถอยทีละชั้น สำหรับชุดข้อมูลโรคเบาหวาน

เทคนิคการ จำแนก	ชุดเรียนรู้					ชุดทดสอบ				
	Acc. (%)	Sen. (%)	Spe. (%)	BAcc. (%)	F1 _{weight} (%)	Acc. (%)	Sen. (%)	Spe. (%)	BAcc. (%)	F1 _{weight} (%)
LR	74.47	76.89	71.99	74.47	74.45	74.42	76.81	<u>71.99</u>	74.42	74.41
DT	79.14	83.39	74.82	79.14	79.10	73.13	77.52	68.67	73.13	73.07
RF (3, 500)*	<u>83.05</u>	<u>86.80</u>	<u>79.22</u>	<u>83.04</u>	<u>83.02</u>	74.42	79.41	69.34	74.42	74.35
SVM (0.01,10)*	74.92	82.86	66.84	74.92	74.75	<u>74.76</u>	<u>82.67</u>	66.69	<u>74.76</u>	<u>74.58</u>

* RF (จำนวนตัวแปรอิสระ, จำนวนต้นไม้) และ SVM (ค่าแกมมา, ค่าคงที่)

จากตารางที่ 38 แสดงประสิทธิภาพของตัวแบบจำแนกที่ใช้ข้อมูลเรียนรู้ทำซ้ำ 10 ครั้ง เพื่อคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติก โดยวิธีการถดถอยทีละชั้น ในชุดข้อมูลโรคเบาหวาน พบว่าจากข้อมูลที่มีตัวแปรอิสระทั้งหมด 21 ตัวถูกคัดเลือกเหลือเพียง 14 ตัว และเมื่อวัด

ประสิทธิภาพด้วยค่าความแม่นยำที่สมดุลแบบปรับน้ำหนักถ่วง บนชุดทดสอบพบว่าเทคนิคซัพพอร์ตเวกเตอร์แมชชีนมีประสิทธิภาพมากที่สุด มีค่าเท่ากับ 74.76% รองลงมาคือ เทคนิคการถดถอยลอจิสติก มีค่าเท่ากับ 74.42% เทคนิคป่าสุ่ม มีค่าเท่ากับ 74.42% และเทคนิคต้นไม้ตัดสินใจ มีค่าเท่ากับ 73.13% ตามลำดับ

ตาราง 39 ผลการจำแนกข้อมูลภายใต้ชุดข้อมูลที่มีการคัดเลือกตัวแปรอิสระโดยใช้ข้อมูลเรียนรู้ทำซ้ำ 10 ครั้ง ด้วยเทคนิคการถดถอยลอจิสติก โดยวิธีการถดถอยทีละชั้น สำหรับชุดข้อมูลการเลิกสูบบุหรี่

เทคนิคการ จำแนก	ชุดเรียนรู้					ชุดทดสอบ				
	Acc. (%)	Sen. (%)	Spe. (%)	BAcc. (%)	F1 _{weight} (%)	Acc. (%)	Sen. (%)	Spe. (%)	BAcc. (%)	F1 _{weight} (%)
LR	73.72	52.70	84.36	73.72	73.07	73.76	52.70	84.42	73.76	73.10
DT	81.84	71.06	87.31	81.84	81.75	72.61	57.08	80.47	72.61	72.45
RF (3, 500)*	<u>100</u>	<u>100</u>	<u>100</u>	<u>100</u>	<u>100</u>	74.44	<u>59.22</u>	82.15	74.44	74.26
SVM (0.01, 10)*	75.37	56.83	84.77	75.38	74.91	<u>75.03</u>	56.33	<u>84.51</u>	<u>75.03</u>	<u>74.56</u>

* RF (จำนวนตัวแปรอิสระ, จำนวนต้นไม้) และ SVM (ค่าแกมมา, ค่าคงที่)

จากตารางที่ 39 แสดงประสิทธิภาพของตัวแบบจำแนกที่ใช้ข้อมูลเรียนรู้ทำซ้ำ 10 ครั้ง เพื่อคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติก โดยวิธีการถดถอยทีละชั้น ในชุดข้อมูลการเลิกสูบบุหรี่ พบว่าจากข้อมูลที่มีตัวแปรอิสระทั้งหมด 22 ตัวถูกคัดเหลือเพียง 12 ตัว และเมื่อวัดประสิทธิภาพด้วยค่าความแม่นยำที่สมดุลแบบปรับน้ำหนักถ่วง บนชุดทดสอบพบว่าเทคนิคซัพพอร์ตเวกเตอร์แมชชีนมีประสิทธิภาพมากที่สุด มีค่าเท่ากับ 75.03% รองลงมาคือ เทคนิคป่าสุ่ม มีค่าเท่ากับ 74.44% เทคนิคการถดถอยลอจิสติก มีค่าเท่ากับ 73.76% และเทคนิคต้นไม้ตัดสินใจ มีค่าเท่ากับ 72.61% ตามลำดับ

4.2.2. ผลการเปรียบเทียบประสิทธิภาพของตัวแบบการจำแนกโดยคัดเลือกตัวแปรอิสระ ด้วยค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient)

ตาราง 40 ผลการจำแนกข้อมูลภายใต้ชุดข้อมูลที่มีการคัดเลือกตัวแปรอิสระด้วยค่าสัมประสิทธิ์สหสัมพันธ์ สำหรับชุดข้อมูลโรคหัวใจและหลอดเลือด

เทคนิคการจำแนก	ชุดเรียนรู้					ชุดทดสอบ				
	Acc. (%)	Sen. (%)	Spe. (%)	BAcc. (%)	F1 _{weight} (%)	Acc. (%)	Sen. (%)	Spe. (%)	BAcc. (%)	F1 _{weight} (%)
LR	72.73	66.67	78.66	72.73	72.63	72.71	66.65	78.63	72.71	72.60
DT	74.75	70.27	79.13	74.75	74.70	73.01	68.55	77.37	73.01	72.95
RF (3, 500)*	<u>89.06</u>	<u>84.76</u>	<u>93.86</u>	<u>89.36</u>	<u>89.04</u>	73.06	<u>69.22</u>	76.81	73.06	73.02
SVM (0.2, 1)*	73.94	67.50	80.25	73.94	73.83	<u>73.41</u>	66.94	<u>79.73</u>	<u>73.41</u>	<u>73.29</u>

* RF (จำนวนตัวแปรอิสระ, จำนวนต้นไม้) และ SVM (ค่าแกมมา, ค่าคงที่)

จากตารางที่ 40 แสดงประสิทธิภาพของตัวแบบจำแนกที่ใช้ค่าสัมประสิทธิ์สหสัมพันธ์คัดเลือกตัวแปรอิสระ ในชุดข้อมูลโรคหัวใจและหลอดเลือด พบว่าจากข้อมูลที่มีตัวแปรอิสระทั้งหมด 11 ตัวถูกคัดเลือกเหลือเพียง 10 ตัว และเมื่อวัดประสิทธิภาพด้วยค่าความแม่นยำที่สมดุลแบบปรับน้ำหนักถ่วง บนชุดทดสอบพบว่าเทคนิคซัพพอร์ตเวกเตอร์แมชชีนมีประสิทธิภาพมากที่สุด มีค่าเท่ากับ 73.41% รองลงมาคือ เทคนิคป่าสุ่ม มีค่าเท่ากับ 73.06% เทคนิคต้นไม้ตัดสินใจ มีค่าเท่ากับ 73.01% และเทคนิคการถดถอยลอจิสติก มีค่าเท่ากับ 72.71% ตามลำดับ

ตาราง 41 ผลการจำแนกข้อมูลภายใต้ชุดข้อมูลที่มีการคัดเลือกตัวแปรอิสระด้วยค่าสัมประสิทธิ์สหสัมพันธ์ สำหรับชุดข้อมูลโรคเบาหวาน

เทคนิคการจำแนก	ชุดเรียนรู้					ชุดทดสอบ				
	Acc. (%)	Sen. (%)	Spe. (%)	BAcc. (%)	F1 _{weight} (%)	Acc. (%)	Sen. (%)	Spe. (%)	BAcc. (%)	F1 _{weight} (%)
LR	74.32	76.66	71.93	74.32	74.30	74.27	76.60	<u>71.90</u>	74.27	74.25
DT	81.51	84.90	78.05	81.48	81.49	72.30	75.61	68.93	72.30	72.27
RF (4, 250)*	<u>92.67</u>	<u>93.05</u>	<u>92.29</u>	<u>92.67</u>	<u>92.67</u>	73.74	78.34	69.04	73.74	73.68
SVM	74.80	82.35	67.10	74.80	74.64	<u>74.61</u>	<u>82.15</u>	66.91	<u>74.61</u>	<u>74.45</u>

เทคนิคการ จำแนก	ชุดเรียนรู้					ชุดทดสอบ				
	Acc. (%)	Sen. (%)	Spe. (%)	BAcc. (%)	F1 _{weight} (%)	Acc. (%)	Sen. (%)	Spe. (%)	BAcc. (%)	F1 _{weight} (%)
(0.01,10)*										

* RF (จำนวนตัวแปรอิสระ, จำนวนต้นไม้) และ SVM (ค่าแกมมา, ค่าคงที่)

จากตารางที่ 41 แสดงประสิทธิภาพของตัวแบบจำแนกที่ใช้ค่าสัมประสิทธิ์สหสัมพันธ์คัดเลือกตัวแปรอิสระ ในชุดข้อมูลโรคเบาหวาน พบว่าจากข้อมูลที่มีตัวแปรอิสระทั้งหมด 21 ตัวถูกคัดเหลือเพียง 16 ตัว และเมื่อวัดประสิทธิภาพด้วยค่าความแม่นยำที่สมดุลแบบปรับน้ำหนักถ่วง บนชุดทดสอบพบว่าเทคนิคซัพพอร์ตเวกเตอร์แมชชีนมีประสิทธิภาพมากที่สุด มีค่าเท่ากับ 74.61% รองลงมาคือ เทคนิคการถดถอยลอจิสติก มีค่าเท่ากับ 74.27% เทคนิคป่าสุ่ม มีค่าเท่ากับ 73.74% และเทคนิคต้นไม้ตัดสินใจ มีค่าเท่ากับ 72.30% ตามลำดับ

ตาราง 42 ผลการจำแนกข้อมูลภายใต้ชุดข้อมูลที่มีการคัดเลือกตัวแปรอิสระด้วยค่าสัมประสิทธิ์สหสัมพันธ์ สำหรับชุดข้อมูลการเลิกสูบบุหรี่

เทคนิค การ จำแนก	ชุดเรียนรู้					ชุดทดสอบ				
	Acc. (%)	Sen. (%)	Spe. (%)	BAcc. (%)	F1 _{weight} (%)	Acc. (%)	Sen. (%)	Spe. (%)	BAcc. (%)	F1 _{weight} (%)
LR	73.40	52.01	84.24	73.40	72.72	73.40	51.91	<u>84.28</u>	73.40	72.71
DT	87.53	80.21	91.23	87.53	87.49	71.57	56.41	79.25	71.57	71.48
RF (4, 500)*	<u>100</u>	<u>100</u>	<u>100</u>	<u>100</u>	<u>100</u>	<u>75.34</u>	<u>60.10</u>	83.06	<u>75.34</u>	<u>75.12</u>
SVM (0.1, 1)*	78.70	63.99	86.14	78.70	78.44	<u>75.34</u>	58.80	83.71	<u>75.34</u>	75.03

* RF (จำนวนตัวแปรอิสระ, จำนวนต้นไม้) และ SVM (ค่าแกมมา, ค่าคงที่)

จากตารางที่ 42 แสดงประสิทธิภาพของตัวแบบจำแนกที่ใช้ค่าสัมประสิทธิ์สหสัมพันธ์คัดเลือกตัวแปรอิสระ ในชุดข้อมูลการเลิกสูบบุหรี่ พบว่าจากข้อมูลที่มีตัวแปรอิสระทั้งหมด 22 ตัวถูกคัดเหลือเพียง 14 ตัว และเมื่อวัดประสิทธิภาพด้วยค่าความแม่นยำที่สมดุลแบบปรับน้ำหนักถ่วง บนชุดทดสอบพบว่าเทคนิคซัพพอร์ตเวกเตอร์แมชชีนมีประสิทธิภาพมากที่สุด มีค่าเท่ากับ 75.34% รองลงมาคือ เทคนิคป่าสุ่ม มีค่าเท่ากับ 75.34% เทคนิคการถดถอยลอจิสติก มีค่าเท่ากับ 73.40% และเทคนิคต้นไม้ตัดสินใจ มีค่าเท่ากับ 71.57% ตามลำดับ

4.3. ผลการเปรียบเทียบประสิทธิภาพของตัวแบบการจำแนก โดยการคัดเลือกตัวแปรอิสระเทียบกับการใช้ตัวแปรอิสระทั้งหมด

ผู้วิจัยได้ศึกษาเปรียบเทียบประสิทธิภาพของตัวแบบการจำแนก โดยการคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติก และค่าสัมประสิทธิ์สหสัมพันธ์เทียบกับการใช้ตัวแปรอิสระทั้งหมด สำหรับชุดข้อมูลโรคหัวใจและหลอดเลือด ชุดข้อมูลโรคเบาหวาน และชุดข้อมูลการเลิกสูบบุหรี่ รายละเอียดดังตาราง 43 - 46

ตาราง 43 ผลการจำแนกด้วยเทคนิคการถดถอยลอจิสติก ในกรณีที่มีการคัดเลือกตัวแปรอิสระเทียบกับการใช้ตัวแปรอิสระทั้งหมด เมื่อวัดด้วยค่าความแม่นยำที่สมดุลแบบปรับน้ำหนักถ่วง

ชุดข้อมูล	การคัดเลือกตัวแปรอิสระ	จำนวนตัวแปรอิสระ		BAcc. (%)	%change
		ทั้งหมด	คัดเลือก		
โรคหัวใจและหลอดเลือด	Full	11	-	72.71	-
	LRข้อมูลทั้งหมด	11	10	72.71 _{f,b,s}	0.0000
	LRข้อมูลเรียนรู้	11	10	72.71 _{f,b,s}	0.0000
	COR	11	10	72.71	0.0000
โรคเบาหวาน	Full	21	-	74.43	-
	LRข้อมูลทั้งหมด	21	16	74.43 _{f,b,s}	0.0000
	LRข้อมูลเรียนรู้	21	14	74.42 _{f,b,s}	-0.0134
	COR	21	16	74.27	-0.2150
การเลิกสูบบุหรี่	Full	22	-	73.64	-
	LRข้อมูลทั้งหมด	22	14	73.66 _{f,b,s}	0.0272
	LRข้อมูลเรียนรู้	22	12	73.76 _{f,s}	0.1630
	COR	22	14	73.40	-0.3259

จากตารางที่ 43 พบว่า

1. วิธีการคัดเลือกตัวแปรอิสระกับชุดข้อมูลโรคหัวใจและหลอดเลือดให้ผลการคัดเลือกเหมือนกันทุกวิธี จึงทำให้ค่าความแม่นยำที่สมดุลแบบปรับน้ำหนักถ่วงมีค่าเท่ากัน และค่าความแม่นยำที่สมดุลแบบปรับน้ำหนักถ่วงเท่ากันกับกรณีที่ใช้ตัวแปรอิสระทั้งหมด

2. ชุดข้อมูลโรคเบาหวานการคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกที่ใช้ข้อมูลทั้งหมด ให้ค่าความแม่นยำที่สมดุลแบบปรับน้ำหนักถ่วงสูงที่สุด และมีค่าเท่ากับกับกรณีที่ใช้ตัวแปรอิสระทั้งหมด

3. ชุดข้อมูลการเลิกสูบบุหรี่การคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติก โดยใช้ข้อมูลเรียนรู้ทำซ้ำ 10 ครั้ง ให้ค่าความแม่นยำที่สมดุลแบบปรับน้ำหนักถ่วงสูงที่สุด และมีค่าสูงกว่ากรณีที่ใช้ตัวแปรอิสระทั้งหมด คิดเป็นเปอร์เซ็นต์เปลี่ยนแปลงเท่ากับ 0.1630%

ตาราง 44 ผลการจำแนกด้วยเทคนิคต้นไม้ตัดสินใจ ในกรณีที่มีการคัดเลือกตัวแปรอิสระเทียบกับการใช้ตัวแปรอิสระทั้งหมด เมื่อวัดด้วยค่าความแม่นยำที่สมดุลแบบปรับน้ำหนักถ่วง

ชุดข้อมูล	การคัดเลือกตัวแปรอิสระ	จำนวนตัวแปรอิสระ		BAcc. (%)	%change
		ทั้งหมด	คัดเลือก		
โรคหัวใจและหลอดเลือด	Full	11	-	73.02	-
	LRข้อมูลทั้งหมด	11	10	73.01 _{f,b,s}	-0.0137
	LRข้อมูลเรียนรู้	11	10	73.01 _{f,b,s}	-0.0137
	COR	11	10	73.01	-0.0137
โรคเบาหวาน	Full	21	-	71.76	-
	LRข้อมูลทั้งหมด	21	16	72.44 _{f,b,s}	0.9476
	LRข้อมูลเรียนรู้	21	14	73.13 _{f,b,s}	1.9091
	COR	21	16	72.30	0.7525
การเลิกสูบบุหรี่	Full	22	-	70.80	-
	LRข้อมูลทั้งหมด	22	14	71.57 _{f,s}	1.0876
	LRข้อมูลเรียนรู้	22	12	72.70 _b	2.6836
	COR	22	14	71.57	1.0876

จากตารางที่ 44 พบว่า

1. ชุดข้อมูลโรคหัวใจและหลอดเลือดกรณีที่มีการคัดเลือกตัวแปรอิสระให้ค่าความแม่นยำที่สมดุลแบบปรับน้ำหนักถ่วงต่ำกว่ากรณีที่ใช้ตัวแปรอิสระทั้งหมด คิดเป็น%เปลี่ยนแปลงเท่ากับ 0.0137%

2. ชุดข้อมูลโรคเบาหวานการคัดเลือกตัวแปรอิสระให้ค่าความแม่นยำที่สมดุลแบบปรับน้ำหนักกว้างสูงกว่าการใช้ตัวแปรอิสระทั้งหมด และการคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติก โดยใช้ข้อมูลเรียนรู้ทำซ้ำ 10 ครั้ง ให้ค่าความแม่นยำที่สมดุลแบบปรับน้ำหนักกว้างสูงสุด คิดเป็น%เปลี่ยนแปลงเท่ากับ 1.9091%

3. ชุดข้อมูลการเลิกสูบบุหรี่การคัดเลือกตัวแปรอิสระให้ค่าความแม่นยำที่สมดุลแบบปรับน้ำหนักกว้างสูงกว่าการใช้ตัวแปรอิสระทั้งหมด และการคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอย-ลอจิสติก โดยใช้ข้อมูลเรียนรู้ทำซ้ำ 10 ครั้ง ให้ค่าความแม่นยำที่สมดุลแบบปรับน้ำหนักกว้างสูงสุด คิดเป็นเปอร์เซ็นต์เปลี่ยนแปลงเท่ากับ 2.6836%

ตาราง 45 ผลการจำแนกด้วยเทคนิคป่าสุ่ม ในกรณีที่มีการคัดเลือกตัวแปรอิสระเทียบกับการใช้ตัวแปรอิสระทั้งหมด เมื่อวัดด้วยค่าความแม่นยำที่สมดุลแบบปรับน้ำหนักกว้าง

ชุดข้อมูล	การคัดเลือกตัวแปรอิสระ	จำนวนตัวแปรอิสระ		เทคนิคที่ดีที่สุด	BAcc. (%)	%change
		ทั้งหมด	คัดเลือก			
โรคหัวใจและหลอดเลือด	Full	11	-	RF (3, 500)*	73.25	-
	LRข้อมูลทั้งหมด	11	10	RF (3, 500)*	73.06 _{f,b,s}	-0.2594
	LRข้อมูลเรียนรู้	11	10	RF (3, 500)*	73.06 _{f,b,s}	-0.2594
	COR	11	10	RF (3, 500)*	73.06	-0.2594
โรคเบาหวาน	Full	21	-	RF (4, 500)*	74.01	-
	LRข้อมูลทั้งหมด	21	16	RF (4, 500)*	73.77 _{f,b,s}	-0.3243
	LRข้อมูลเรียนรู้	21	14	RF (3, 500)*	74.42 _{f,b,s}	0.5540
	COR	21	16	RF (4, 250)*	73.74	-0.3648
การเลิกสูบบุหรี่	Full	22	-	RF (4, 500)*	74.97	-
	LRข้อมูลทั้งหมด	22	14	RF (3, 250)*	74.72 _b	-0.3335
	LRข้อมูลเรียนรู้	22	12	RF (3, 500)*	74.44 _{f,s}	-0.7069
	COR	22	14	RF (4, 500)*	75.34	0.4935

* RF (จำนวนตัวแปรอิสระ, จำนวนต้นไม้)

จากตารางที่ 45 พบว่า

1. ชุดข้อมูลโรคหัวใจและหลอดเลือดกรณีที่มีการคัดเลือกตัวแปรอิสระให้ค่าความแม่นยำที่สมดุลแบบปรับน้ำหนักกว้างต่ำกว่ากรณีที่ใช้ตัวแปรอิสระทั้งหมด คิดเป็น%เปลี่ยนแปลงเท่ากับ 0.2594%

2. ชุดข้อมูลโรคเบาหวานการคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติก โดยใช้ข้อมูลเรียนรู้ทำซ้ำ 10 ครั้ง ให้ค่าความแม่นยำที่สมดุลแบบปรับน้ำหนักถ่วงสูงที่สุด และมีค่าสูงกว่ากรณีที่ใช้ตัวแปรอิสระทั้งหมด คิดเป็น%เปลี่ยนแปลงเท่ากับ 0.5540%

3. ชุดข้อมูลการเลิกสูบบุหรี่การคัดเลือกตัวแปรอิสระด้วยค่าสัมประสิทธิ์สหสัมพันธ์ให้ค่าความแม่นยำที่สมดุลแบบปรับน้ำหนักถ่วงสูงที่สุด และมีค่าสูงกว่ากรณีที่ใช้ตัวแปรอิสระทั้งหมด คิดเป็นเปอร์เซ็นต์เปลี่ยนแปลงเท่ากับ 0.4935%

ตาราง 46 ผลการจำแนกด้วยเทคนิคซัพพอร์ตเวกเตอร์แมชชีน ในกรณีที่มีการคัดเลือกตัวแปรอิสระ เทียบกับการใช้ตัวแปรอิสระทั้งหมด เมื่อวัดด้วยค่าความแม่นยำที่สมดุลแบบปรับน้ำหนักถ่วง

ชุดข้อมูล	การคัดเลือกตัวแปรอิสระ	จำนวนตัวแปรอิสระ		เทคนิคที่ดีที่สุด	BAcc. (%)	%change
		ทั้งหมด	คัดเลือก			
โรคหัวใจและหลอดเลือด	Full	11	-	SVM (0.1, 1)*	73.76	-
	LRข้อมูลทั้งหมด	11	10	SVM (0.2, 1)*	73.41 _{f,b,s}	-0.4745
	LRข้อมูลเรียนรู้	11	10	SVM (0.2, 1)*	73.41 _{f,b,s}	-0.4745
	COR	11	10	SVM (0.2, 1)*	73.41	-0.4745
โรคเบาหวาน	Full	21	-	SVM (0.01, 1)*	75.40	-
	LRข้อมูลทั้งหมด	21	16	SVM (0.01, 10)*	74.79 _{f,b,s}	-0.8090
	LRข้อมูลเรียนรู้	21	14	SVM (0.01, 10)*	74.76 _{f,b,s}	-0.8488
	COR	21	16	SVM (0.01, 10)*	74.61	-1.0477
การเลิกสูบบุหรี่	Full	22	-	SVM (0.01, 10)*	75.33	-
	LRข้อมูลทั้งหมด	22	14	SVM (0.01, 10)*	75.41 _b	0.1062
	LRข้อมูลเรียนรู้	22	12	SVM (0.01, 10)*	75.03 _{f,s}	-0.3982
	COR	22	14	SVM (0.1, 1)*	75.34	0.0133

* SVM (ค่าแกมมา, ค่าคงที่)

จากตารางที่ 46 พบว่า

1. ชุดข้อมูลโรคหัวใจและหลอดเลือดกรณีที่มีการคัดเลือกตัวแปรอิสระให้ค่าความแม่นยำที่สมดุลแบบปรับน้ำหนักถ่วงต่ำกว่ากรณีที่ใช้ตัวแปรอิสระทั้งหมด คิดเป็น%เปลี่ยนแปลงเท่ากับ 0.4745%

2. ชุดข้อมูลโรคเบาหวานกรณีที่มีการคัดเลือกตัวแปรอิสระให้ค่าความแม่นยำที่สมดุลแบบปรับน้ำหนักถ่วงต่ำกว่ากรณีที่ใช้ตัวแปรอิสระทั้งหมด

3. ชุดข้อมูลการเลิกสูบบุหรี่การคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกที่ใช้ข้อมูลทั้งหมด ให้ค่าความแม่นยำที่สมดุลแบบปรับน้ำหนักถ่วงสูงสุด และมีค่าสูงกว่ากรณีที่ใช้ตัวแปรอิสระทั้งหมด คิดเป็นเปอร์เซ็นต์เปลี่ยนแปลงเท่ากับ 0.1062%

4.4. ผลการเปรียบเทียบประสิทธิภาพของตัวแบบการจำแนกที่ดีที่สุด

ตาราง 47 ผลการเปรียบเทียบประสิทธิภาพของตัวแบบการจำแนกที่ดีที่สุด

ชุดข้อมูล	เทคนิคการจำแนกที่ดีที่สุด	การคัดเลือกตัวแปรอิสระ	BAcc.
1. โรคหัวใจและหลอดเลือด	SVM (0.1, 1)*	Full	73.76%
2. โรคเบาหวาน	SVM (0.01, 1)*	Full	75.39%
3. การเลิกสูบบุหรี่	SVM (0.01, 10)*	LR ^{ข้อมูลทั้งหมด}	75.41%

* SVM (ค่าแกมมา, ค่าคงที่)

จากตาราง 47 พบว่าตัวแบบการจำแนกที่ได้จากเทคนิคซัพพอร์ตเวกเตอร์แมชชีน กรณีที่ใช้ตัวแปรอิสระทั้งหมดมีประสิทธิภาพมากที่สุดสำหรับชุดข้อมูลโรคหัวใจและหลอดเลือด และชุดข้อมูลโรคเบาหวาน โดยมีค่าความแม่นยำที่สมดุลแบบปรับน้ำหนักถ่วงเท่ากับ 73.76% และ 75.39% ตามลำดับ และตัวแบบการจำแนกที่มีประสิทธิภาพมากที่สุดสำหรับชุดข้อมูลการเลิกสูบบุหรี่ คือ ตัวแบบการจำแนกจากเทคนิคซัพพอร์ตเวกเตอร์แมชชีน กรณีที่มีการคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติก โดยใช้ข้อมูลทั้งหมดมีค่าความแม่นยำที่สมดุลแบบปรับน้ำหนักถ่วงเท่ากับ 75.41%

บทที่ 5

สรุปผลการวิจัย

งานวิจัยเรื่องการเปรียบเทียบประสิทธิภาพของตัวแบบการจำแนกข้อมูลด้านสุขภาพ มีวัตถุประสงค์เพื่อศึกษาการทำงานและเปรียบเทียบประสิทธิภาพการจำแนกข้อมูลด้านสุขภาพที่มีจำนวนตัวแปรอิสระเชิงปริมาณและตัวแปรอิสระเชิงคุณภาพแตกต่างกันทั้งหมด 3 ชุด ได้แก่ ชุดข้อมูลโรคหัวใจและหลอดเลือด (ชุดข้อมูลที่มีจำนวนตัวแปรอิสระเชิงปริมาณใกล้เคียงตัวแปรอิสระเชิงคุณภาพ) ชุดข้อมูลโรคเบาหวาน (ชุดข้อมูลที่มีจำนวนตัวแปรอิสระเชิงปริมาณน้อยกว่าตัวแปรอิสระเชิงคุณภาพ) และชุดข้อมูลการเลิกสูบบุหรี่ (ชุดข้อมูลที่มีจำนวนตัวแปรอิสระเชิงปริมาณมากกว่าตัวแปรอิสระเชิงคุณภาพ) โดยใช้เทคนิคการจำแนก 4 เทคนิค ได้แก่ เทคนิคการถดถอยลอจิสติก เทคนิคต้นไม้ตัดสินใจ เทคนิคป่าสุ่ม และเทคนิคซัพพอร์ตเวกเตอร์แมชชีน และศึกษาวิเคราะห์เกี่ยวกับการคัดเลือกตัวแปรอิสระโดยใช้เทคนิคการถดถอยลอจิสติก ที่ระดับนัยสำคัญ 0.05 และค่าสัมประสิทธิ์สหสัมพันธ์ ในขั้นตอนการพัฒนาตัวแบบการจำแนกได้ทำการทดสอบประสิทธิภาพของตัวแบบการจำแนกโดยใช้ 10 – Fold Cross-Validation จากนั้นทำการเปรียบเทียบประสิทธิภาพของตัวแบบการจำแนกด้วยค่าเกณฑ์การวัดประสิทธิภาพ คือ ค่าความแม่นยำที่สมดุลแบบปรับน้ำหนักถ่วง

ผลที่ได้จากการศึกษาสามารถสรุปได้ดังนี้

1. จากการศึกษาพัฒนาตัวแบบการจำแนกโดยใช้ตัวแปรอิสระทั้งหมด พบว่า เทคนิคซัพพอร์ตเวกเตอร์แมชชีนมีประสิทธิภาพมากที่สุดสำหรับชุดข้อมูลด้านสุขภาพทั้ง 3 ชุดที่มีคุณลักษณะของตัวแปรอิสระที่แตกต่างกัน เมื่อวัดจากค่าความแม่นยำที่สมดุลแบบปรับน้ำหนักถ่วง ดังแสดงผลลัพธ์ในตารางที่ 19 – 21

2. จากการศึกษายังไม่สามารถระบุได้ชัดเจนว่าเทคนิคการจำแนกใดทำงานได้ดีภายใต้ชุดข้อมูลที่มีคุณลักษณะของตัวแปรอิสระในลักษณะใด เนื่องจากชุดข้อมูลทั้ง 3 ชุด เทคนิคซัพพอร์ตเวกเตอร์แมชชีนมีประสิทธิภาพมากที่สุด

3. จากการพัฒนาตัวแบบการจำแนก เมื่อวิเคราะห์ผลลัพธ์ในตารางที่ 43 – 46 พบว่า

- 3.1. กรณีที่มีการคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติก โดยใช้ข้อมูลทั้งหมด สำหรับชุดข้อมูลทั้ง 3 ชุด เทคนิคการจำแนกที่มีประสิทธิภาพมากที่สุด คือ เทคนิคซัพพอร์ตเวกเตอร์แมชชีน โดยชุดข้อมูลโรคหัวใจและหลอดเลือด และชุดข้อมูลโรคเบาหวาน ได้จากวิธีการ

เลือกแบบไปข้างหน้า วิธีการกำจัดแบบถอยหลัง และวิธีการถดถอยทีละชั้น สำหรับชุดข้อมูลการเลิกสูบบุหรี่ ได้จากวิธีการกำจัดแบบถอยหลัง

3.2. กรณีที่มีการคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติก ที่ใช้ข้อมูลเรียนรู้ทำซ้ำ 10 ครั้ง สำหรับชุดข้อมูลทั้ง 3 ชุด เทคนิคการจำแนกที่มีประสิทธิภาพมากที่สุด คือ เทคนิคซัพพอร์ตเวกเตอร์แมชชีน โดยชุดข้อมูลโรคหัวใจและหลอดเลือด และชุดข้อมูลโรคเบาหวาน ได้จากวิธีการเลือกแบบไปข้างหน้า วิธีการกำจัดแบบถอยหลัง และวิธีการถดถอยทีละชั้น สำหรับชุดข้อมูลการเลิกสูบบุหรี่ ได้จากวิธีการเลือกแบบไปข้างหน้า และวิธีการถดถอยทีละชั้น

3.3. กรณีที่มีการคัดเลือกตัวแปรอิสระด้วยค่าสัมประสิทธิ์สหสัมพันธ์ สำหรับชุดข้อมูลโรคหัวใจและหลอดเลือด และชุดข้อมูลโรคเบาหวาน เทคนิคการจำแนกที่มีประสิทธิภาพมากที่สุด คือ เทคนิคซัพพอร์ตเวกเตอร์แมชชีน สำหรับชุดข้อมูลการเลิกสูบบุหรี่ เทคนิคการจำแนกที่มีประสิทธิภาพมากที่สุด คือ เทคนิคซัพพอร์ตเวกเตอร์แมชชีน และเทคนิคป่าสุ่ม

3.4. การพัฒนาตัวแบบด้วยเทคนิคการถดถอยลอจิสติก และเทคนิคต้นไม้ตัดสินใจ พบว่า เมื่อมีการคัดเลือกตัวแปรอิสระทำให้มีประสิทธิภาพในการจำแนกดีขึ้น และจากการเปรียบเทียบพบว่าวิธีที่เหมาะสม คือ การคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติก ที่ใช้ข้อมูลเรียนรู้ทำซ้ำ 10 ครั้ง

3.5. การพัฒนาตัวแบบด้วยเทคนิคป่าสุ่ม และเทคนิคซัพพอร์ตเวกเตอร์แมชชีน พบว่า เมื่อมีการคัดเลือกตัวแปรอิสระจากผลลัพธ์ส่วนมากทำให้ประสิทธิภาพในการจำแนกลดลง เนื่องจากหลักการทำงานของทั้งสองเทคนิคจะจำแนกได้ดีในกรณีที่มีจำนวนตัวแปรอิสระ และจำนวนข้อมูลที่มากพอ

5.1. อภิปรายผลวิจัย

จากงานวิจัยเรื่องการเปรียบเทียบประสิทธิภาพของตัวแบบการจำแนกข้อมูลด้านสุขภาพ เพื่อหาเทคนิคการจำแนกที่เหมาะสมกับข้อมูลที่ใช้ในงานวิจัย พบว่า เทคนิคการถดถอยลอจิสติก และเทคนิคต้นไม้ตัดสินใจ มีประสิทธิภาพในการจำแนกดีขึ้นในกรณีที่มีการคัดเลือกตัวแปรอิสระ ซึ่งสอดคล้องกับงานวิจัยของรัชนิวรรณ ไพศาลวรเกียรติ (2564) และ Dissanayake and Johar (2021)

5.2. ข้อเสนอแนะ

จากการวิจัยในครั้งนี้ เพื่อเป็นแนวทางในการพัฒนางานวิจัยให้มีประสิทธิภาพมากขึ้น ผู้วิจัยมีข้อเสนอแนะดังนี้

1. คัดเลือกชุดข้อมูลที่นำมาวิเคราะห์ให้มีความหลากหลายทั้งคุณลักษณะของตัวแปรอิสระและจำนวนข้อมูล เพื่อสามารถวัดประสิทธิภาพในการจำแนกข้อมูลของแต่ละเทคนิคได้ชัดเจนมากขึ้น
2. คัดเลือกตัวแปรอิสระด้วยวิธีอื่นๆ เช่น Chi-Square, LASSO, Boruta เป็นต้น
3. ประยุกต์ใช้เทคนิคการจำแนกอื่นๆ เช่น เทคนิคโครงข่ายประสาทเทียม เทคนิคนาอ็ฟเบย์ และเทคนิคเคเนียร์เซนเบอร์ เป็นต้น
4. การแบ่งข้อมูลเพื่อนำมาทดสอบประสิทธิภาพของตัวแบบการจำแนกวิธีอื่นๆ เช่น Split Test, 5-Fold Cross-Validation เป็นต้น
5. เลือกใช้ฟังก์ชันเคอร์เนลในเทคนิคซัพพอร์ตเวกเตอร์แมชชีนด้วยฟังก์ชันอื่นๆ เช่น โพลีโนเมียล (Polynomial) ซิกมอยด์ (Sigmoid) เป็นต้น
6. ประยุกต์ใช้วิธีการปรับชุดข้อมูลสมดุลให้สมดุล เช่น วิธีการสุ่มเกิน วิธีการสุ่มเกินโดยเทคนิค SMOTE และวิธีการสุ่มลด เป็นต้น



ภาคผนวก ก
ผลการวิเคราะห์ข้อมูล

ตารางภาคผนวกที่ 1 รายละเอียดชุดข้อมูลโรคหัวใจและหลอดเลือด โดยแยกเป็นกรณีผู้ที่เป็นโรคหัวใจและหลอดเลือดกับผู้ที่ไม่เป็นโรคหัวใจและหลอดเลือด

ตัวแปร	เป็นโรคหัวใจและหลอดเลือด				ไม่เป็นโรคหัวใจและหลอดเลือด			
	จำนวน	Min	Max	Avg±SD	จำนวน	Min	Max	Avg±SD
อายุ (Day)		14,275	23,713	20,061.60± 2,316.33		10,798	23,678	18,879.07± 2,472.26
เพศ	ช 11,908 ญ 21,922				ช 11,955 ญ 22,648			
ส่วนสูง (cm.)		140	198	164.36±7.90		140	207	164.56±7.70
น้ำหนัก (kg.)		30	200	76.69±14.80		30	200	71.58±13.24
ค่าความดันขณะหัวใจบีบตัว (mmHg.)		70	200	133.73±17.00		70	200	119.54±12.47
ค่าความดันขณะหัวใจคลายตัว (mmHg.)		50	120	84.51±9.36		50	120	78.15±8.14
ระดับคอเลสเตอรอลแบ่งเป็น 3 ระดับ	ปกติ 22,338 เริ่มสูง 5,516 สูง 5,976				ปกติ 28,994 เริ่มสูง 3,743 สูง 1,866			
ระดับน้ำตาลกลูโคสแบ่งเป็น 3 ระดับ	ปกติ 27,660 เริ่มสูง 2,959 สูง 3,211				ปกติ 30,529 เริ่มสูง 2,083 สูง 1,991			
การสูบบุหรี่	สูบ 2,817 ไม่สูบ 31,013				สูบ 3,207 ไม่สูบ 31,396			
การดื่มแอลกอฮอล์	ดื่ม 1,741 ไม่ดื่ม 32,089				ดื่ม 1,918 ไม่ดื่ม 32,685			
การออกกำลังกาย	ออก 26,665 ไม่ออก 7,165				ออก 28,313 ไม่ออก 6,290			

ตารางภาคผนวกที่ 2 รายละเอียดชุดข้อมูลโรคเบาหวาน โดยแยกเป็นกรณีผู้ที่เป็นโรคเบาหวานกับผู้ที่ไม่เป็นโรคเบาหวาน

ตัวแปร	เป็นโรคเบาหวาน				ไม่เป็นโรคเบาหวาน			
	จำนวน	Min	Max	Avg±SD	จำนวน	Min	Max	Avg±SD
ค่าดัชนีมวลกาย (kg/m ²)		13	50	31.42±6.26		12	50	27.65±5.48
จำนวนวันที่สุขภาพจิตแย่ใน 1 เดือน (Day)		0	30	4.41±8.90		0	30	3.15±7.31
จำนวนวันที่เจ็บป่วยและขาดเจ็บใน 1 เดือน (Day)		0	30	7.87±11.25		0	30	3.79±8.20
โรคความดันโลหิตสูง	เป็น = 25,810 ไม่เป็น = 8,584				เป็น = 12,931 ไม่เป็น = 20,809			
คอเลสเตอรอลสูง	ใช่ = 11,346 ไม่ = 23,048				ใช่ = 13,114 ไม่ = 20,626			
รับการตรวจคอเลสเตอรอลทุกๆ 5 ปี	ใช่ = 34,157 ไม่ = 237				ใช่ = 32,256 ไม่ = 1,484			
สูบบุหรี่อย่างน้อย 100 มวน	ใช่ = 17,907 ไม่ = 16,487				ใช่ = 14,971 ไม่ = 18,769			
โรคหลอดเลือดสมอง	เป็น = 3,212 ไม่เป็น = 31,182				เป็น = 1,117 ไม่เป็น = 32,623			
โรคหลอดเลือดหัวใจ (CHD) หรือกล้ามเนื้อหัวใจตาย (MI)	เป็น = 7,726 ไม่เป็น = 26,668				เป็น = 2,549 ไม่เป็น = 31,191			
การออกกำลังกาย	ออก = 21,795 ไม่ออก = 12,599				ออก = 25,902 ไม่ออก = 7,838			
ทานผลไม้อย่างน้อย 1 ครั้งต่อสัปดาห์	ใช่ = 20,154 ไม่ = 14,240				ใช่ = 21,206 ไม่ = 12,534			
ทานผักอย่างน้อย 1 ครั้งต่อสัปดาห์	ใช่ = 25,983 ไม่ = 8,411				ใช่ = 27,478 ไม่ = 6,262			
ภาวะติดแอลกอฮอล์	เป็น = 828 ไม่เป็น = 33,566				เป็น = 2,165 ไม่เป็น = 31,575			
การคุ้มครองด้านการดูแลสุขภาพ	มี = 33,017 ไม่มี = 1,377				มี = 32,000 ไม่มี = 1,740			
ไม่สามารถพบแพทย์ได้เนื่องจากปัญหาค่าใช้จ่าย	ใช่ = 3,620 ไม่ = 30,774				ใช่ = 2,867 ไม่ = 30,873			
ระดับคะแนนสุขภาพแบ่งเป็น 5 ระดับ	ยอดเยี่ยม = 1,126 ดีมาก = 6,237 ดี = 13,143 พอใช้ = 9,499 แย่มาก = 4,389				ยอดเยี่ยม = 6,456 ดีมาก = 12,773 ดี = 9,841 พอใช้ = 3,458 แย่มาก = 1,212			
ปัญหาในการเดินหรือขึ้นบันได	มี = 12,602 ไม่มี = 21,792				มี = 4,659 ไม่มี = 29,081			
เพศ	ช = 16,527 ญ = 17,866				ช = 14,686 ญ = 19,054			
อายุแบ่งเป็น 13 ช่วง	1 = 76				1 = 891			

ตัวแปร	เป็นโรคเบาหวาน			ไม่เป็นโรคเบาหวาน				
	จำนวน	Min	Max	Avg±SD	จำนวน	Min	Max	Avg±SD
	2 = 127				2 = 1,227			
	3 = 294				3 = 1,657			
	4 = 594				4 = 2,044			
	5 = 997				5 = 2,320			
	6 = 1,670				6 = 2,708			
	7 = 2,962				7 = 3,525			
	8 = 4,122				8 = 4,124			
	9 = 5,556				9 = 4,166			
	10 = 6,383				10 = 4,078			
	11 = 5,054				11 = 2,835			
	12 = 3,370				12 = 1,971			
	13 = 3,189				13 = 2,194			
ระดับการศึกษาแบ่งเป็น 6 ระดับ	1 = 46				1 = 27			
	2 = 1,154				2 = 460			
	3 = 2,239				3 = 1,143			
	4 = 10,795				4 = 8,306			
	5 = 10,084				5 = 9,462			
	6 = 10,076				6 = 14,342			
รายได้แบ่งเป็น 8 ช่วง	1 = 2,280				1 = 1,218			
	2 = 2,992				2 = 1,390			
	3 = 3,477				3 = 1,967			
	4 = 3,949				4 = 2,577			
	5 = 4,408				5 = 3,458			
	6 = 5,178				6 = 4,901			
	7 = 5,144				7 = 6,017			
	8 = 6,966				8 = 12,212			

ตารางภาคผนวกที่ 3 รายละเอียดชุดข้อมูลการเลิกสูบบุหรี่ โดยแยกเป็นกรณีที่เลิกสูบบุหรี่กับไม่เลิกสูบบุหรี่

ตัวแปร	จำนวน	เลิกสูบบุหรี่			ไม่เลิกสูบบุหรี่			
		Min	Max	Avg±SD	จำนวน	Min	Max	Avg±SD
อายุ (Year)		20	85	40.98±11.52		20	85	45.51±12.09
ส่วนสูง (cm.)		140	190	169.53±6.78		140	190	161.80±9.27
น้ำหนัก (kg.)		35	115	69.86±11.22		30	115	62.19±11.79
รอบเอว (cm.)		54	112.1	83.70±8.24		54	112	79.75±8.83
ความยาวสายตาด้านซ้าย (Diopter)		0.1	1.8	1.03±0.31		0.1	1.6	0.97±0.32
ความยาวสายตาด้านขวา (Diopter)		0.1	1.6	1.03±0.31		0.1	1.6	0.96±0.32
การได้ยินของหูด้านซ้าย	1 = 8,996				1 = 17,618			
การได้ยินของหูด้านขวา	2 = 177				2 = 494			
การได้ยินของหูด้านซ้าย	1 = 8,982				1 = 17,626			
การได้ยินของหูด้านขวา	2 = 191				2 = 486			

ตัวแปร	จำนวน	เล็กสูบบุหรี่			จำนวน	ไม่เล็กสูบบุหรี่		
		Min	Max	Avg±SD		Min	Max	Avg±SD
ค่าความดันขณะหัวใจ บีบตัว (mmHg.)		81	166	121.36±12.19		72	166	119.87±13.22
ค่าความดันของหัวใจ คลายตัว (mmHg.)		46	106	76.32±8.81		42	106	74.67±9.24
ค่าน้ำตาลในเลือด (FBS) (mg/dL)		57	134	96.67±11.43		56	134	95.05±10.81
ค่าคอเลสเตอรอล (mg/dL)		77	316	193.43±33.98		93	315	197.06±34.55
ค่าไตรกลีเซอไรด์ (mg/dL)		11	330	135.25±65.22		8	330	106.94±55.21
ค่าคอเลสเตอรอลชนิด ที่ดี (mg/dL)		30	120	53.99±12.97		22	104	59.53±13.82
ค่าคอเลสเตอรอลชนิด ที่ไม่ดี (mg/dL)		12	230	112.67±32.05		7	229	116.26±31.71
ค่าฮีโมโกลบิน (g/dL)		9.6	20	15.38±1.08		9.4	19.1	14.13±1.47
ค่าโปรตีนในปัสสาวะ	1 = 8,749 2 = 276 3 = 114 4 = 25 5 = 9 6 = 0				1 = 17,240 2 = 551 3 = 251 4 = 53 5 = 15 6 = 2			
ซีรั่มครีเอตินิน (Scr) (mg/dL)		0.5	1.4	0.95±0.16		0.5	1.4	0.85±0.19
ค่าเอนไซม์ AST ใน เลือด (U/L)		9	50	23.74±6.53		7	50	23.07±6.39
ค่าเอนไซม์ ALT ใน เลือด (U/L)		3	64	24.70±11.33		1	64	21.06±10.17
ค่าจีโปรตีน (kDa)		7	98	36.98±19.87		3	98	24.88±15.38
โรคฟันผุ	เป็น = 2,511 ไม่เป็น = 6,662				เป็น = 3,198 ไม่เป็น = 14,914			

ชุดข้อมูลโรคหัวใจและหลอดเลือด

ตารางภาคผนวกที่ 4 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคการถดถอยลอจิสติก โดยใช้ตัวแปรอิสระทั้งหมด ภายใต้ชุดข้อมูลโรคหัวใจและหลอดเลือด (ชุดข้อมูลเรียนรู้)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	20340	6642	10107	24500	0.7281	0.6680	0.7867	0.7280	0.7270	0.7538	0.7079
2	20324	6659	10123	24483	0.7275	0.6675	0.7862	0.7275	0.7265	0.7532	0.7075
3	20265	6670	10182	24472	0.7264	0.6656	0.7858	0.7264	0.7253	0.7524	0.7062
4	20286	6640	10161	24503	0.7272	0.6663	0.7868	0.7272	0.7261	0.7534	0.7069
5	20341	6663	10106	24480	0.7277	0.6681	0.7861	0.7278	0.7267	0.7533	0.7078
6	20292	6675	10155	24468	0.7267	0.6665	0.7857	0.7268	0.7257	0.7525	0.7067
7	20305	6653	10142	24490	0.7273	0.6669	0.7864	0.7273	0.7263	0.7532	0.7071
8	20285	6662	10162	24481	0.7268	0.6662	0.7861	0.7268	0.7258	0.7528	0.7067
9	20286	6628	10161	24515	0.7274	0.6663	0.7872	0.7274	0.7263	0.7537	0.7070
10	20311	6621	10136	24522	0.7279	0.6671	0.7874	0.7279	0.7269	0.7542	0.7075
Average					0.7273	0.6669	0.7864	0.7273	0.7263	0.7532	0.7071

ตารางภาคผนวกที่ 5 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคการถดถอยลอจิสติก โดยใช้ตัวแปรอิสระทั้งหมด ภายใต้ชุดข้อมูลโรคหัวใจและหลอดเลือด (ชุดข้อมูลทดสอบ)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	2215	742	1168	2719	0.7209	0.6547	0.7856	0.7209	0.7196	0.7491	0.6995
2	2255	753	1128	2708	0.7252	0.6666	0.7824	0.7252	0.7242	0.7497	0.7059
3	2266	709	1117	2752	0.7332	0.6698	0.7951	0.7332	0.7320	0.7617	0.7113
4	2253	719	1130	2741	0.7298	0.6660	0.7922	0.7298	0.7286	0.7581	0.7081
5	2238	753	1145	2707	0.7226	0.6615	0.7824	0.7226	0.7215	0.7482	0.7028
6	2288	744	1095	2716	0.7313	0.6763	0.7850	0.7313	0.7304	0.7546	0.7127
7	2251	728	1135	2732	0.7282	0.6654	0.7896	0.7282	0.7269	0.7556	0.7065
8	2276	729	1107	2731	0.7317	0.6728	0.7893	0.7317	0.7307	0.7574	0.7116
9	2271	763	1112	2697	0.7260	0.6713	0.7795	0.7260	0.7251	0.7485	0.7081
10	2238	756	1145	2704	0.7222	0.6615	0.7815	0.7222	0.7211	0.7475	0.7025
Average					0.7271	0.6666	0.7863	0.7271	0.7260	0.7530	0.7069

ตารางภาคผนวกที่ 6 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคการถดถอยลอจิสติก
กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการเลือกแบบไปข้างหน้า โดยใช้ข้อมูล
ทั้งหมด ภายใต้ชุดข้อมูลโรคหัวใจและหลอดเลือด (ชุดข้อมูลเรียนรู้)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	20333	6639	24503	10114	0.7280	0.7868	0.6678	0.7280	0.7269	0.7539	0.7078
2	20316	6652	24490	10131	0.7275	0.7864	0.6673	0.7275	0.7265	0.7533	0.7074
3	20264	6667	24475	10183	0.7264	0.7859	0.6655	0.7264	0.7253	0.7524	0.7062
4	20286	6630	24513	10161	0.7274	0.7871	0.6663	0.7274	0.7263	0.7537	0.7070
5	20334	6667	24476	10113	0.7276	0.7859	0.6678	0.7275	0.7265	0.7531	0.7076
6	20298	6676	24467	10149	0.7268	0.7856	0.6667	0.7268	0.7258	0.7525	0.7068
7	20306	6641	24502	10141	0.7275	0.7868	0.6669	0.7275	0.7265	0.7536	0.7073
8	20275	6654	24489	10172	0.7268	0.7863	0.6659	0.7268	0.7257	0.7529	0.7065
9	20271	6619	24524	10176	0.7273	0.7875	0.6658	0.7273	0.7262	0.7538	0.7067
10	20306	6620	24523	10141	0.7279	0.7874	0.6669	0.7278	0.7268	0.7541	0.7074
Average					0.7273	0.7866	0.6667	0.7273	0.7262	0.7533	0.7071

ตารางภาคผนวกที่ 7 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคการถดถอยลอจิสติก
กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการเลือกแบบไปข้างหน้า โดยใช้ข้อมูล
ทั้งหมด ภายใต้ชุดข้อมูลโรคหัวใจและหลอดเลือด (ชุดข้อมูลทดสอบ)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	2215	743	2718	1168	0.7208	0.7853	0.6547	0.7207	0.7195	0.7488	0.6994
2	2251	750	2711	1132	0.7250	0.7833	0.6654	0.7250	0.7240	0.7501	0.7054
3	2262	708	2753	1121	0.7328	0.7954	0.6686	0.7327	0.7316	0.7616	0.7106
4	2254	719	2741	1129	0.7299	0.7922	0.6663	0.7300	0.7288	0.7582	0.7083
5	2236	752	2708	1147	0.7225	0.7827	0.6610	0.7225	0.7214	0.7483	0.7025
6	2290	743	2717	1093	0.7317	0.7853	0.6769	0.7317	0.7308	0.7550	0.7131
7	2251	733	2727	1132	0.7275	0.7882	0.6654	0.7275	0.7264	0.7544	0.7067
8	2278	726	2734	1105	0.7324	0.7902	0.6734	0.7325	0.7315	0.7583	0.7122
9	2274	766	2694	1109	0.7260	0.7786	0.6722	0.7260	0.7252	0.7480	0.7084
10	2238	755	2705	1145	0.7223	0.7818	0.6615	0.7223	0.7212	0.7477	0.7026
Average					0.7271	0.7863	0.6665	0.7271	0.7260	0.7530	0.7069

ตารางภาคผนวกที่ 8 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคการถดถอยลอจิสติก
กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีกำจัดแบบถอยหลัง โดยใช้ข้อมูลทั้งหมด
ภายใต้ชุดข้อมูลโรคหัวใจและหลอดเลือด (ชุดข้อมูลเรียนรู้)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	20333	6639	24503	10114	0.7280	0.7868	0.6678	0.7280	0.7269	0.7539	0.7078
2	20316	6652	24490	10131	0.7275	0.7864	0.6673	0.7275	0.7265	0.7533	0.7074
3	20264	6667	24475	10183	0.7264	0.7859	0.6655	0.7264	0.7253	0.7524	0.7062
4	20286	6630	24513	10161	0.7274	0.7871	0.6663	0.7274	0.7263	0.7537	0.7070
5	20334	6667	24476	10113	0.7276	0.7859	0.6678	0.7275	0.7265	0.7531	0.7076
6	20298	6676	24467	10149	0.7268	0.7856	0.6667	0.7268	0.7258	0.7525	0.7068
7	20306	6641	24502	10141	0.7275	0.7868	0.6669	0.7275	0.7265	0.7536	0.7073
8	20275	6654	24489	10172	0.7268	0.7863	0.6659	0.7268	0.7257	0.7529	0.7065
9	20271	6619	24524	10176	0.7273	0.7875	0.6658	0.7273	0.7262	0.7538	0.7067
10	20306	6620	24523	10141	0.7279	0.7874	0.6669	0.7278	0.7268	0.7541	0.7074
Average					0.7273	0.7866	0.6667	0.7273	0.7262	0.7533	0.7071

ตารางภาคผนวกที่ 9 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคการถดถอยลอจิสติก
กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีกำจัดแบบถอยหลัง โดยใช้ข้อมูลทั้งหมด
ภายใต้ชุดข้อมูลโรคหัวใจและหลอดเลือด (ชุดข้อมูลทดสอบ)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	2215	743	2718	1168	0.7208	0.7853	0.6547	0.7207	0.7195	0.7488	0.6994
2	2251	750	2711	1132	0.7250	0.7833	0.6654	0.7250	0.7240	0.7501	0.7054
3	2262	708	2753	1121	0.7328	0.7954	0.6686	0.7327	0.7316	0.7616	0.7106
4	2254	719	2741	1129	0.7299	0.7922	0.6663	0.7300	0.7288	0.7582	0.7083
5	2236	752	2708	1147	0.7225	0.7827	0.6610	0.7225	0.7214	0.7483	0.7025
6	2290	743	2717	1093	0.7317	0.7853	0.6769	0.7317	0.7308	0.7550	0.7131
7	2251	733	2727	1132	0.7275	0.7882	0.6654	0.7275	0.7264	0.7544	0.7067
8	2278	726	2734	1105	0.7324	0.7902	0.6734	0.7325	0.7315	0.7583	0.7122
9	2274	766	2694	1109	0.7260	0.7786	0.6722	0.7260	0.7252	0.7480	0.7084
10	2238	755	2705	1145	0.7223	0.7818	0.6615	0.7223	0.7212	0.7477	0.7026
Average					0.7271	0.7863	0.6665	0.7271	0.7260	0.7530	0.7069

ตารางภาคผนวกที่ 10 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคการถดถอยลอจิสติก
กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการถดถอยทีละชั้น โดยใช้ข้อมูลทั้งหมด
ภายใต้ชุดข้อมูลโรคหัวใจและหลอดเลือด (ชุดข้อมูลเรียนรู้)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	20333	6639	24503	10114	0.7280	0.7868	0.6678	0.7280	0.7269	0.7539	0.7078
2	20316	6652	24490	10131	0.7275	0.7864	0.6673	0.7275	0.7265	0.7533	0.7074
3	20264	6667	24475	10183	0.7264	0.7859	0.6655	0.7264	0.7253	0.7524	0.7062
4	20286	6630	24513	10161	0.7274	0.7871	0.6663	0.7274	0.7263	0.7537	0.7070
5	20334	6667	24476	10113	0.7276	0.7859	0.6678	0.7275	0.7265	0.7531	0.7076
6	20298	6676	24467	10149	0.7268	0.7856	0.6667	0.7268	0.7258	0.7525	0.7068
7	20306	6641	24502	10141	0.7275	0.7868	0.6669	0.7275	0.7265	0.7536	0.7073
8	20275	6654	24489	10172	0.7268	0.7863	0.6659	0.7268	0.7257	0.7529	0.7065
9	20271	6619	24524	10176	0.7273	0.7875	0.6658	0.7273	0.7262	0.7538	0.7067
10	20306	6620	24523	10141	0.7279	0.7874	0.6669	0.7278	0.7268	0.7541	0.7074
Average					0.7273	0.7866	0.6667	0.7273	0.7262	0.7533	0.7071

ตารางภาคผนวกที่ 11 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคการถดถอยลอจิสติก
กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการถดถอยทีละชั้น โดยใช้ข้อมูลทั้งหมด
ภายใต้ชุดข้อมูลโรคหัวใจและหลอดเลือด (ชุดข้อมูลทดสอบ)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	2215	743	2718	1168	0.7208	0.7853	0.6547	0.7207	0.7195	0.7488	0.6994
2	2251	750	2711	1132	0.7250	0.7833	0.6654	0.7250	0.7240	0.7501	0.7054
3	2262	708	2753	1121	0.7328	0.7954	0.6686	0.7327	0.7316	0.7616	0.7106
4	2254	719	2741	1129	0.7299	0.7922	0.6663	0.7300	0.7288	0.7582	0.7083
5	2236	752	2708	1147	0.7225	0.7827	0.6610	0.7225	0.7214	0.7483	0.7025
6	2290	743	2717	1093	0.7317	0.7853	0.6769	0.7317	0.7308	0.7550	0.7131
7	2251	733	2727	1132	0.7275	0.7882	0.6654	0.7275	0.7264	0.7544	0.7067
8	2278	726	2734	1105	0.7324	0.7902	0.6734	0.7325	0.7315	0.7583	0.7122
9	2274	766	2694	1109	0.7260	0.7786	0.6722	0.7260	0.7252	0.7480	0.7084
10	2238	755	2705	1145	0.7223	0.7818	0.6615	0.7223	0.7212	0.7477	0.7026
Average					0.7271	0.7863	0.6665	0.7271	0.7260	0.7530	0.7069

ตารางภาคผนวกที่ 12 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคการถดถอยลอจิสติก กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการเลือกแบบไปข้างหน้า โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้อุณหภูมิข้อมูลโรคหัวใจและหลอดเลือด (ชุดข้อมูลเรียนรู้)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	20333	6639	24503	10114	0.7280	0.7868	0.6678	0.7280	0.7269	0.7539	0.7078
2	20316	6652	24490	10131	0.7275	0.7864	0.6673	0.7275	0.7265	0.7533	0.7074
3	20264	6667	24475	10183	0.7264	0.7859	0.6655	0.7264	0.7253	0.7524	0.7062
4	20286	6630	24513	10161	0.7274	0.7871	0.6663	0.7274	0.7263	0.7537	0.7070
5	20334	6667	24476	10113	0.7276	0.7859	0.6678	0.7275	0.7265	0.7531	0.7076
6	20298	6676	24467	10149	0.7268	0.7856	0.6667	0.7268	0.7258	0.7525	0.7068
7	20306	6641	24502	10141	0.7275	0.7868	0.6669	0.7275	0.7265	0.7536	0.7073
8	20275	6654	24489	10172	0.7268	0.7863	0.6659	0.7268	0.7257	0.7529	0.7065
9	20271	6619	24524	10176	0.7273	0.7875	0.6658	0.7273	0.7262	0.7538	0.7067
10	20306	6620	24523	10141	0.7279	0.7874	0.6669	0.7278	0.7268	0.7541	0.7074
Average					0.7273	0.7866	0.6667	0.7273	0.7262	0.7533	0.7071

ตารางภาคผนวกที่ 13 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคการถดถอยลอจิสติก กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการเลือกแบบไปข้างหน้า โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้อุณหภูมิข้อมูลโรคหัวใจและหลอดเลือด (ชุดข้อมูลทดสอบ)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	2215	743	2718	1168	0.7208	0.7853	0.6547	0.7207	0.7195	0.7488	0.6994
2	2251	750	2711	1132	0.7250	0.7833	0.6654	0.7250	0.7240	0.7501	0.7054
3	2262	708	2753	1121	0.7328	0.7954	0.6686	0.7327	0.7316	0.7616	0.7106
4	2254	719	2741	1129	0.7299	0.7922	0.6663	0.7300	0.7288	0.7582	0.7083
5	2236	752	2708	1147	0.7225	0.7827	0.6610	0.7225	0.7214	0.7483	0.7025
6	2290	743	2717	1093	0.7317	0.7853	0.6769	0.7317	0.7308	0.7550	0.7131
7	2251	733	2727	1132	0.7275	0.7882	0.6654	0.7275	0.7264	0.7544	0.7067
8	2278	726	2734	1105	0.7324	0.7902	0.6734	0.7325	0.7315	0.7583	0.7122
9	2274	766	2694	1109	0.7260	0.7786	0.6722	0.7260	0.7252	0.7480	0.7084
10	2238	755	2705	1145	0.7223	0.7818	0.6615	0.7223	0.7212	0.7477	0.7026
Average					0.7271	0.7863	0.6665	0.7271	0.7260	0.7530	0.7069

ตารางภาคผนวกที่ 14 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคการถดถอยลอจิสติก กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีกำจัดแบบถอยหลัง โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้ชุดข้อมูลโรคหัวใจและหลอดเลือด (ชุดข้อมูลเรียนรู้)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	20333	6639	24503	10114	0.7280	0.7868	0.6678	0.7280	0.7269	0.7539	0.7078
2	20316	6652	24490	10131	0.7275	0.7864	0.6673	0.7275	0.7265	0.7533	0.7074
3	20264	6667	24475	10183	0.7264	0.7859	0.6655	0.7264	0.7253	0.7524	0.7062
4	20286	6630	24513	10161	0.7274	0.7871	0.6663	0.7274	0.7263	0.7537	0.7070
5	20334	6667	24476	10113	0.7276	0.7859	0.6678	0.7275	0.7265	0.7531	0.7076
6	20298	6676	24467	10149	0.7268	0.7856	0.6667	0.7268	0.7258	0.7525	0.7068
7	20306	6641	24502	10141	0.7275	0.7868	0.6669	0.7275	0.7265	0.7536	0.7073
8	20275	6654	24489	10172	0.7268	0.7863	0.6659	0.7268	0.7257	0.7529	0.7065
9	20271	6619	24524	10176	0.7273	0.7875	0.6658	0.7273	0.7262	0.7538	0.7067
10	20306	6620	24523	10141	0.7279	0.7874	0.6669	0.7278	0.7268	0.7541	0.7074
Average					0.7273	0.7866	0.6667	0.7273	0.7262	0.7533	0.7071

ตารางภาคผนวกที่ 15 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคการถดถอยลอจิสติก กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีกำจัดแบบถอยหลัง โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้ชุดข้อมูลโรคหัวใจและหลอดเลือด (ชุดข้อมูลทดสอบ)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	2215	743	2718	1168	0.7208	0.7853	0.6547	0.7207	0.7195	0.7488	0.6994
2	2251	750	2711	1132	0.7250	0.7833	0.6654	0.7250	0.7240	0.7501	0.7054
3	2262	708	2753	1121	0.7328	0.7954	0.6686	0.7327	0.7316	0.7616	0.7106
4	2254	719	2741	1129	0.7299	0.7922	0.6663	0.7300	0.7288	0.7582	0.7083
5	2236	752	2708	1147	0.7225	0.7827	0.6610	0.7225	0.7214	0.7483	0.7025
6	2290	743	2717	1093	0.7317	0.7853	0.6769	0.7317	0.7308	0.7550	0.7131
7	2251	733	2727	1132	0.7275	0.7882	0.6654	0.7275	0.7264	0.7544	0.7067
8	2278	726	2734	1105	0.7324	0.7902	0.6734	0.7325	0.7315	0.7583	0.7122
9	2274	766	2694	1109	0.7260	0.7786	0.6722	0.7260	0.7252	0.7480	0.7084
10	2238	755	2705	1145	0.7223	0.7818	0.6615	0.7223	0.7212	0.7477	0.7026
Average					0.7271	0.7863	0.6665	0.7271	0.7260	0.7530	0.7069

ตารางภาคผนวกที่ 16 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคการถดถอยลอจิสติก กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการถดถอยทีละชั้น โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้ชุดข้อมูลโรคหัวใจและหลอดเลือด (ชุดข้อมูล เรียนรู้)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	20333	6639	24503	10114	0.7280	0.7868	0.6678	0.7280	0.7269	0.7539	0.7078
2	20316	6652	24490	10131	0.7275	0.7864	0.6673	0.7275	0.7265	0.7533	0.7074
3	20264	6667	24475	10183	0.7264	0.7859	0.6655	0.7264	0.7253	0.7524	0.7062
4	20286	6630	24513	10161	0.7274	0.7871	0.6663	0.7274	0.7263	0.7537	0.7070
5	20334	6667	24476	10113	0.7276	0.7859	0.6678	0.7275	0.7265	0.7531	0.7076
6	20298	6676	24467	10149	0.7268	0.7856	0.6667	0.7268	0.7258	0.7525	0.7068
7	20306	6641	24502	10141	0.7275	0.7868	0.6669	0.7275	0.7265	0.7536	0.7073
8	20275	6654	24489	10172	0.7268	0.7863	0.6659	0.7268	0.7257	0.7529	0.7065
9	20271	6619	24524	10176	0.7273	0.7875	0.6658	0.7273	0.7262	0.7538	0.7067
10	20306	6620	24523	10141	0.7279	0.7874	0.6669	0.7278	0.7268	0.7541	0.7074
Average					0.7273	0.7866	0.6667	0.7273	0.7262	0.7533	0.7071

ตารางภาคผนวกที่ 17 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคการถดถอยลอจิสติก กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการถดถอยทีละชั้น โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้ชุดข้อมูลโรคหัวใจและหลอดเลือด (ชุดข้อมูล ทดสอบ)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	2215	743	2718	1168	0.7208	0.7853	0.6547	0.7207	0.7195	0.7488	0.6994
2	2251	750	2711	1132	0.7250	0.7833	0.6654	0.7250	0.7240	0.7501	0.7054
3	2262	708	2753	1121	0.7328	0.7954	0.6686	0.7327	0.7316	0.7616	0.7106
4	2254	719	2741	1129	0.7299	0.7922	0.6663	0.7300	0.7288	0.7582	0.7083
5	2236	752	2708	1147	0.7225	0.7827	0.6610	0.7225	0.7214	0.7483	0.7025
6	2290	743	2717	1093	0.7317	0.7853	0.6769	0.7317	0.7308	0.7550	0.7131
7	2251	733	2727	1132	0.7275	0.7882	0.6654	0.7275	0.7264	0.7544	0.7067
8	2278	726	2734	1105	0.7324	0.7902	0.6734	0.7325	0.7315	0.7583	0.7122
9	2274	766	2694	1109	0.7260	0.7786	0.6722	0.7260	0.7252	0.7480	0.7084
10	2238	755	2705	1145	0.7223	0.7818	0.6615	0.7223	0.7212	0.7477	0.7026
Average					0.7271	0.7863	0.6665	0.7271	0.7260	0.7530	0.7069

ตารางภาคผนวกที่ 18 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคการถดถอยลอจิสติก
กรณีคัดเลือกตัวแปรอิสระด้วยค่าสัมประสิทธิ์สหสัมพันธ์ ภายใต้ชุดข้อมูลโรคหัวใจและหลอดเลือด
(ชุดข้อมูลเรียนรู้)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	20333	6639	24503	10114	0.7280	0.7868	0.6678	0.7280	0.7269	0.7539	0.7078
2	20316	6652	24490	10131	0.7275	0.7864	0.6673	0.7275	0.7265	0.7533	0.7074
3	20264	6667	24475	10183	0.7264	0.7859	0.6655	0.7264	0.7253	0.7524	0.7062
4	20286	6630	24513	10161	0.7274	0.7871	0.6663	0.7274	0.7263	0.7537	0.7070
5	20334	6667	24476	10113	0.7276	0.7859	0.6678	0.7275	0.7265	0.7531	0.7076
6	20298	6676	24467	10149	0.7268	0.7856	0.6667	0.7268	0.7258	0.7525	0.7068
7	20306	6641	24502	10141	0.7275	0.7868	0.6669	0.7275	0.7265	0.7536	0.7073
8	20275	6654	24489	10172	0.7268	0.7863	0.6659	0.7268	0.7257	0.7529	0.7065
9	20271	6619	24524	10176	0.7273	0.7875	0.6658	0.7273	0.7262	0.7538	0.7067
10	20306	6620	24523	10141	0.7279	0.7874	0.6669	0.7278	0.7268	0.7541	0.7074
Average					0.7273	0.7866	0.6667	0.7273	0.7262	0.7533	0.7071

ตารางภาคผนวกที่ 19 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคการถดถอยลอจิสติก
กรณีคัดเลือกตัวแปรอิสระด้วยค่าสัมประสิทธิ์สหสัมพันธ์ ภายใต้ชุดข้อมูลโรคหัวใจและหลอดเลือด
(ชุดข้อมูลทดสอบ)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	2215	743	2718	1168	0.7208	0.7853	0.6547	0.7207	0.7195	0.7488	0.6994
2	2251	750	2711	1132	0.7250	0.7833	0.6654	0.7250	0.7240	0.7501	0.7054
3	2262	708	2753	1121	0.7328	0.7954	0.6686	0.7327	0.7316	0.7616	0.7106
4	2254	719	2741	1129	0.7299	0.7922	0.6663	0.7300	0.7288	0.7582	0.7083
5	2236	752	2708	1147	0.7225	0.7827	0.6610	0.7225	0.7214	0.7483	0.7025
6	2290	743	2717	1093	0.7317	0.7853	0.6769	0.7317	0.7308	0.7550	0.7131
7	2251	733	2727	1132	0.7275	0.7882	0.6654	0.7275	0.7264	0.7544	0.7067
8	2278	726	2734	1105	0.7324	0.7902	0.6734	0.7325	0.7315	0.7583	0.7122
9	2274	766	2694	1109	0.7260	0.7786	0.6722	0.7260	0.7252	0.7480	0.7084
10	2238	755	2705	1145	0.7223	0.7818	0.6615	0.7223	0.7212	0.7477	0.7026
Average					0.7271	0.7863	0.6665	0.7271	0.7260	0.7530	0.7069

ตารางภาคผนวกที่ 20 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคต้นไม้ตัดสินใจ โดยใช้ตัวแปรอิสระทั้งหมด ภายใต้ชุดข้อมูลโรคหัวใจและหลอดเลือด (ชุดข้อมูลเรียนรู้)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	21976	6603	8471	24539	0.7552	0.7218	0.7880	0.7553	0.7550	0.7690	0.7434
2	21906	6604	8541	24538	0.7541	0.7195	0.7879	0.7541	0.7538	0.7684	0.7418
3	21682	6559	8765	24583	0.7512	0.7121	0.7894	0.7512	0.7508	0.7677	0.7372
4	21548	6420	8899	24723	0.7513	0.7077	0.7939	0.7513	0.7508	0.7705	0.7353
5	21471	6367	8976	24776	0.7509	0.7052	0.7956	0.7509	0.7503	0.7713	0.7341
6	21584	6417	8863	24726	0.7519	0.7089	0.7940	0.7519	0.7514	0.7708	0.7361
7	21632	6511	8815	24632	0.7512	0.7105	0.7909	0.7512	0.7507	0.7686	0.7364
8	21516	6415	8931	24728	0.7508	0.7067	0.7940	0.7508	0.7503	0.7703	0.7347
9	21458	6125	8989	25018	0.7546	0.7048	0.8033	0.7546	0.7539	0.7779	0.7357
10	21527	6352	8920	24791	0.7520	0.7070	0.7960	0.7520	0.7515	0.7722	0.7354
Average					0.7523	0.7104	0.7933	0.7523	0.7518	0.7707	0.7370

ตารางภาคผนวกที่ 21 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคต้นไม้ตัดสินใจ โดยใช้ตัวแปรอิสระทั้งหมด ภายใต้ชุดข้อมูลโรคหัวใจและหลอดเลือด (ชุดข้อมูลทดสอบ)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	2351	825	1032	2636	0.7287	0.6949	0.7616	0.7286	0.7283	0.7402	0.7186
2	2346	852	1037	2609	0.7240	0.6935	0.7538	0.7240	0.7237	0.7336	0.7156
3	2373	772	1010	2689	0.7396	0.7014	0.7769	0.7396	0.7392	0.7545	0.7270
4	2328	771	1055	2689	0.7332	0.6881	0.7772	0.7332	0.7326	0.7512	0.7182
5	2289	795	1094	2665	0.7240	0.6766	0.7702	0.7239	0.7233	0.7422	0.7090
6	2371	786	1012	2674	0.7372	0.7009	0.7728	0.7373	0.7369	0.7510	0.7254
7	2321	791	1062	2669	0.7292	0.6861	0.7714	0.7292	0.7287	0.7458	0.7154
8	2329	777	1054	2683	0.7324	0.6884	0.7754	0.7324	0.7318	0.7498	0.7180
9	2302	803	1081	2657	0.7247	0.6805	0.7679	0.7247	0.7241	0.7414	0.7108
10	2294	764	1089	2696	0.7292	0.6781	0.7792	0.7292	0.7285	0.7502	0.7123
Average					0.7302	0.6889	0.7706	0.7302	0.7297	0.7460	0.7170

ตารางภาคผนวกที่ 22 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคต้นไม้ตัดสินใจ กรณี
คัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการเลือกแบบไปข้างหน้า โดยใช้ข้อมูล
ทั้งหมด ภายใต้ชุดข้อมูลโรคหัวใจและหลอดเลือด (ชุดข้อมูลเรียนรู้)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	21573	6595	8874	24547	0.7488	0.7085	0.7882	0.7488	0.7484	0.7659	0.7345
2	22190	7200	8257	23942	0.7490	0.7288	0.7688	0.7490	0.7489	0.7550	0.7436
3	21120	6288	9327	24854	0.7465	0.6937	0.7981	0.7465	0.7457	0.7706	0.7271
4	21029	5996	9418	25147	0.7497	0.6907	0.8075	0.7498	0.7488	0.7781	0.7275
5	20611	5709	9836	25434	0.7476	0.6769	0.8167	0.7476	0.7463	0.7831	0.7211
6	21493	6529	8954	24614	0.7486	0.7059	0.7904	0.7486	0.7481	0.7670	0.7333
7	21422	6713	9025	24430	0.7445	0.7036	0.7844	0.7445	0.7440	0.7614	0.7302
8	21688	6879	8759	24264	0.7461	0.7123	0.7791	0.7461	0.7458	0.7592	0.7348
9	21822	6925	8625	24218	0.7475	0.7167	0.7776	0.7475	0.7472	0.7591	0.7374
10	21001	6152	9446	24991	0.7467	0.6898	0.8025	0.7468	0.7459	0.7734	0.7257
Average					0.7475	0.7027	0.7913	0.7475	0.7469	0.7673	0.7315

ตารางภาคผนวกที่ 23 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคต้นไม้ตัดสินใจ กรณี
คัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการเลือกแบบไปข้างหน้า โดยใช้ข้อมูล
ทั้งหมด ภายใต้ชุดข้อมูลโรคหัวใจและหลอดเลือด (ชุดข้อมูลทดสอบ)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	2323	819	1060	2642	0.7255	0.6867	0.7634	0.7255	0.7250	0.7393	0.7137
2	2389	914	994	2547	0.7212	0.7062	0.7359	0.7212	0.7211	0.7233	0.7193
3	2318	736	1065	2725	0.7368	0.6852	0.7873	0.7368	0.7361	0.7590	0.7190
4	2285	711	1098	2749	0.7356	0.6754	0.7945	0.7356	0.7346	0.7627	0.7146
5	2202	702	1181	2758	0.7248	0.6509	0.7971	0.7248	0.7233	0.7583	0.7002
6	2356	780	1027	2680	0.7359	0.6964	0.7746	0.7359	0.7355	0.7513	0.7230
7	2306	786	1077	2674	0.7278	0.6816	0.7728	0.7277	0.7271	0.7458	0.7129
8	2380	803	1003	2657	0.7361	0.7035	0.7679	0.7361	0.7358	0.7477	0.7260
9	2381	850	1002	2610	0.7294	0.7038	0.7543	0.7293	0.7291	0.7369	0.7226
10	2250	729	1133	2731	0.7279	0.6651	0.7893	0.7279	0.7268	0.7553	0.7068
Average					0.7301	0.6855	0.7737	0.7301	0.7294	0.7480	0.7158

ตารางภาคผนวกที่ 24 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคต้นไม้ตัดสินใจ กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีกำจัดแบบถอยหลัง โดยใช้ข้อมูลทั้งหมด ภายใต้ชุดข้อมูลโรคหัวใจและหลอดเลือด (ชุดข้อมูลเรียนรู้)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	21573	6595	8874	24547	0.7488	0.7085	0.7882	0.7488	0.7484	0.7659	0.7345
2	22190	7200	8257	23942	0.7490	0.7288	0.7688	0.7490	0.7489	0.7550	0.7436
3	21120	6288	9327	24854	0.7465	0.6937	0.7981	0.7465	0.7457	0.7706	0.7271
4	21029	5996	9418	25147	0.7497	0.6907	0.8075	0.7498	0.7488	0.7781	0.7275
5	20611	5709	9836	25434	0.7476	0.6769	0.8167	0.7476	0.7463	0.7831	0.7211
6	21493	6529	8954	24614	0.7486	0.7059	0.7904	0.7486	0.7481	0.7670	0.7333
7	21422	6713	9025	24430	0.7445	0.7036	0.7844	0.7445	0.7440	0.7614	0.7302
8	21688	6879	8759	24264	0.7461	0.7123	0.7791	0.7461	0.7458	0.7592	0.7348
9	21822	6925	8625	24218	0.7475	0.7167	0.7776	0.7475	0.7472	0.7591	0.7374
10	21001	6152	9446	24991	0.7467	0.6898	0.8025	0.7468	0.7459	0.7734	0.7257
Average					0.7475	0.7027	0.7913	0.7475	0.7469	0.7673	0.7315

ตารางภาคผนวกที่ 25 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคต้นไม้ตัดสินใจ กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีกำจัดแบบถอยหลัง โดยใช้ข้อมูลทั้งหมด ภายใต้ชุดข้อมูลโรคหัวใจและหลอดเลือด (ชุดข้อมูลทดสอบ)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	2323	819	1060	2642	0.7255	0.6867	0.7634	0.7255	0.7250	0.7393	0.7137
2	2389	914	994	2547	0.7212	0.7062	0.7359	0.7212	0.7211	0.7233	0.7193
3	2318	736	1065	2725	0.7368	0.6852	0.7873	0.7368	0.7361	0.7590	0.7190
4	2285	711	1098	2749	0.7356	0.6754	0.7945	0.7356	0.7346	0.7627	0.7146
5	2202	702	1181	2758	0.7248	0.6509	0.7971	0.7248	0.7233	0.7583	0.7002
6	2356	780	1027	2680	0.7359	0.6964	0.7746	0.7359	0.7355	0.7513	0.7230
7	2306	786	1077	2674	0.7278	0.6816	0.7728	0.7277	0.7271	0.7458	0.7129
8	2380	803	1003	2657	0.7361	0.7035	0.7679	0.7361	0.7358	0.7477	0.7260
9	2381	850	1002	2610	0.7294	0.7038	0.7543	0.7293	0.7291	0.7369	0.7226
10	2250	729	1133	2731	0.7279	0.6651	0.7893	0.7279	0.7268	0.7553	0.7068
Average					0.7301	0.6855	0.7737	0.7301	0.7294	0.7480	0.7158

ตารางภาคผนวกที่ 26 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคต้นไม้ตัดสินใจ กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการถดถอยทีละชั้น โดยใช้ข้อมูลทั้งหมดภายใต้ชุดข้อมูลโรคหัวใจและหลอดเลือด (ชุดข้อมูลเรียนรู้)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	21573	6595	8874	24547	0.7488	0.7085	0.7882	0.7488	0.7484	0.7659	0.7345
2	22190	7200	8257	23942	0.7490	0.7288	0.7688	0.7490	0.7489	0.7550	0.7436
3	21120	6288	9327	24854	0.7465	0.6937	0.7981	0.7465	0.7457	0.7706	0.7271
4	21029	5996	9418	25147	0.7497	0.6907	0.8075	0.7498	0.7488	0.7781	0.7275
5	20611	5709	9836	25434	0.7476	0.6769	0.8167	0.7476	0.7463	0.7831	0.7211
6	21493	6529	8954	24614	0.7486	0.7059	0.7904	0.7486	0.7481	0.7670	0.7333
7	21422	6713	9025	24430	0.7445	0.7036	0.7844	0.7445	0.7440	0.7614	0.7302
8	21688	6879	8759	24264	0.7461	0.7123	0.7791	0.7461	0.7458	0.7592	0.7348
9	21822	6925	8625	24218	0.7475	0.7167	0.7776	0.7475	0.7472	0.7591	0.7374
10	21001	6152	9446	24991	0.7467	0.6898	0.8025	0.7468	0.7459	0.7734	0.7257
Average					0.7475	0.7027	0.7913	0.7475	0.7469	0.7673	0.7315

ตารางภาคผนวกที่ 27 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคต้นไม้ตัดสินใจ กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการถดถอยทีละชั้น โดยใช้ข้อมูลทั้งหมดภายใต้ชุดข้อมูลโรคหัวใจและหลอดเลือด (ชุดข้อมูลทดสอบ)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	2323	819	1060	2642	0.7255	0.6867	0.7634	0.7255	0.7250	0.7393	0.7137
2	2389	914	994	2547	0.7212	0.7062	0.7359	0.7212	0.7211	0.7233	0.7193
3	2318	736	1065	2725	0.7368	0.6852	0.7873	0.7368	0.7361	0.7590	0.7190
4	2285	711	1098	2749	0.7356	0.6754	0.7945	0.7356	0.7346	0.7627	0.7146
5	2202	702	1181	2758	0.7248	0.6509	0.7971	0.7248	0.7233	0.7583	0.7002
6	2356	780	1027	2680	0.7359	0.6964	0.7746	0.7359	0.7355	0.7513	0.7230
7	2306	786	1077	2674	0.7278	0.6816	0.7728	0.7277	0.7271	0.7458	0.7129
8	2380	803	1003	2657	0.7361	0.7035	0.7679	0.7361	0.7358	0.7477	0.7260
9	2381	850	1002	2610	0.7294	0.7038	0.7543	0.7293	0.7291	0.7369	0.7226
10	2250	729	1133	2731	0.7279	0.6651	0.7893	0.7279	0.7268	0.7553	0.7068
Average					0.7301	0.6855	0.7737	0.7301	0.7294	0.7480	0.7158

ตารางภาคผนวกที่ 28 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคต้นไม้ตัดสินใจ กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการเลือกแบบไปข้างหน้า โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้อัตราข้อมูลโรคหัวใจและหลอดเลือด (ชุดข้อมูลเรียนรู้)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	21573	6595	8874	24547	0.7488	0.7085	0.7882	0.7488	0.7484	0.7659	0.7345
2	22190	7200	8257	23942	0.7490	0.7288	0.7688	0.7490	0.7489	0.7550	0.7436
3	21120	6288	9327	24854	0.7465	0.6937	0.7981	0.7465	0.7457	0.7706	0.7271
4	21029	5996	9418	25147	0.7497	0.6907	0.8075	0.7498	0.7488	0.7781	0.7275
5	20611	5709	9836	25434	0.7476	0.6769	0.8167	0.7476	0.7463	0.7831	0.7211
6	21493	6529	8954	24614	0.7486	0.7059	0.7904	0.7486	0.7481	0.7670	0.7333
7	21422	6713	9025	24430	0.7445	0.7036	0.7844	0.7445	0.7440	0.7614	0.7302
8	21688	6879	8759	24264	0.7461	0.7123	0.7791	0.7461	0.7458	0.7592	0.7348
9	21822	6925	8625	24218	0.7475	0.7167	0.7776	0.7475	0.7472	0.7591	0.7374
10	21001	6152	9446	24991	0.7467	0.6898	0.8025	0.7468	0.7459	0.7734	0.7257
Average					0.7475	0.7027	0.7913	0.7475	0.7469	0.7673	0.7315

ตารางภาคผนวกที่ 29 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคต้นไม้ตัดสินใจ กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการเลือกแบบไปข้างหน้า โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้อัตราข้อมูลโรคหัวใจและหลอดเลือด (ชุดข้อมูลทดสอบ)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	2323	819	1060	2642	0.7255	0.6867	0.7634	0.7255	0.7250	0.7393	0.7137
2	2389	914	994	2547	0.7212	0.7062	0.7359	0.7212	0.7211	0.7233	0.7193
3	2318	736	1065	2725	0.7368	0.6852	0.7873	0.7368	0.7361	0.7590	0.7190
4	2285	711	1098	2749	0.7356	0.6754	0.7945	0.7356	0.7346	0.7627	0.7146
5	2202	702	1181	2758	0.7248	0.6509	0.7971	0.7248	0.7233	0.7583	0.7002
6	2356	780	1027	2680	0.7359	0.6964	0.7746	0.7359	0.7355	0.7513	0.7230
7	2306	786	1077	2674	0.7278	0.6816	0.7728	0.7277	0.7271	0.7458	0.7129
8	2380	803	1003	2657	0.7361	0.7035	0.7679	0.7361	0.7358	0.7477	0.7260
9	2381	850	1002	2610	0.7294	0.7038	0.7543	0.7293	0.7291	0.7369	0.7226
10	2250	729	1133	2731	0.7279	0.6651	0.7893	0.7279	0.7268	0.7553	0.7068
Average					0.7301	0.6855	0.7737	0.7301	0.7294	0.7480	0.7158

ตารางภาคผนวกที่ 30 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคต้นไม้ตัดสินใจ กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีกำจัดแบบถอยหลัง โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้ชุดข้อมูลโรคหัวใจและหลอดเลือด (ชุดข้อมูลเรียนรู้)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	21573	6595	8874	24547	0.7488	0.7085	0.7882	0.7488	0.7484	0.7659	0.7345
2	22190	7200	8257	23942	0.7490	0.7288	0.7688	0.7490	0.7489	0.7550	0.7436
3	21120	6288	9327	24854	0.7465	0.6937	0.7981	0.7465	0.7457	0.7706	0.7271
4	21029	5996	9418	25147	0.7497	0.6907	0.8075	0.7498	0.7488	0.7781	0.7275
5	20611	5709	9836	25434	0.7476	0.6769	0.8167	0.7476	0.7463	0.7831	0.7211
6	21493	6529	8954	24614	0.7486	0.7059	0.7904	0.7486	0.7481	0.7670	0.7333
7	21422	6713	9025	24430	0.7445	0.7036	0.7844	0.7445	0.7440	0.7614	0.7302
8	21688	6879	8759	24264	0.7461	0.7123	0.7791	0.7461	0.7458	0.7592	0.7348
9	21822	6925	8625	24218	0.7475	0.7167	0.7776	0.7475	0.7472	0.7591	0.7374
10	21001	6152	9446	24991	0.7467	0.6898	0.8025	0.7468	0.7459	0.7734	0.7257
Average					0.7475	0.7027	0.7913	0.7475	0.7469	0.7673	0.7315

ตารางภาคผนวกที่ 31 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคต้นไม้ตัดสินใจ กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีกำจัดแบบถอยหลัง โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้ชุดข้อมูลโรคหัวใจและหลอดเลือด (ชุดข้อมูลทดสอบ)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	2323	819	1060	2642	0.7255	0.6867	0.7634	0.7255	0.7250	0.7393	0.7137
2	2389	914	994	2547	0.7212	0.7062	0.7359	0.7212	0.7211	0.7233	0.7193
3	2318	736	1065	2725	0.7368	0.6852	0.7873	0.7368	0.7361	0.7590	0.7190
4	2285	711	1098	2749	0.7356	0.6754	0.7945	0.7356	0.7346	0.7627	0.7146
5	2202	702	1181	2758	0.7248	0.6509	0.7971	0.7248	0.7233	0.7583	0.7002
6	2356	780	1027	2680	0.7359	0.6964	0.7746	0.7359	0.7355	0.7513	0.7230
7	2306	786	1077	2674	0.7278	0.6816	0.7728	0.7277	0.7271	0.7458	0.7129
8	2380	803	1003	2657	0.7361	0.7035	0.7679	0.7361	0.7358	0.7477	0.7260
9	2381	850	1002	2610	0.7294	0.7038	0.7543	0.7293	0.7291	0.7369	0.7226
10	2250	729	1133	2731	0.7279	0.6651	0.7893	0.7279	0.7268	0.7553	0.7068
Average					0.7301	0.6855	0.7737	0.7301	0.7294	0.7480	0.7158

ตารางภาคผนวกที่ 32 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคต้นไม้ตัดสินใจ กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการถดถอยทีละชั้น โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้ชุดข้อมูลโรคหัวใจและหลอดเลือด (ชุดข้อมูลเรียนรู้)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	21573	6595	8874	24547	0.7488	0.7085	0.7882	0.7488	0.7484	0.7659	0.7345
2	22190	7200	8257	23942	0.7490	0.7288	0.7688	0.7490	0.7489	0.7550	0.7436
3	21120	6288	9327	24854	0.7465	0.6937	0.7981	0.7465	0.7457	0.7706	0.7271
4	21029	5996	9418	25147	0.7497	0.6907	0.8075	0.7498	0.7488	0.7781	0.7275
5	20611	5709	9836	25434	0.7476	0.6769	0.8167	0.7476	0.7463	0.7831	0.7211
6	21493	6529	8954	24614	0.7486	0.7059	0.7904	0.7486	0.7481	0.7670	0.7333
7	21422	6713	9025	24430	0.7445	0.7036	0.7844	0.7445	0.7440	0.7614	0.7302
8	21688	6879	8759	24264	0.7461	0.7123	0.7791	0.7461	0.7458	0.7592	0.7348
9	21822	6925	8625	24218	0.7475	0.7167	0.7776	0.7475	0.7472	0.7591	0.7374
10	21001	6152	9446	24991	0.7467	0.6898	0.8025	0.7468	0.7459	0.7734	0.7257
Average					0.7475	0.7027	0.7913	0.7475	0.7469	0.7673	0.7315

ตารางภาคผนวกที่ 33 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคต้นไม้ตัดสินใจ กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการถดถอยทีละชั้น โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้ชุดข้อมูลโรคหัวใจและหลอดเลือด (ชุดข้อมูลทดสอบ)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	2323	819	1060	2642	0.7255	0.6867	0.7634	0.7255	0.7250	0.7393	0.7137
2	2389	914	994	2547	0.7212	0.7062	0.7359	0.7212	0.7211	0.7233	0.7193
3	2318	736	1065	2725	0.7368	0.6852	0.7873	0.7368	0.7361	0.7590	0.7190
4	2285	711	1098	2749	0.7356	0.6754	0.7945	0.7356	0.7346	0.7627	0.7146
5	2202	702	1181	2758	0.7248	0.6509	0.7971	0.7248	0.7233	0.7583	0.7002
6	2356	780	1027	2680	0.7359	0.6964	0.7746	0.7359	0.7355	0.7513	0.7230
7	2306	786	1077	2674	0.7278	0.6816	0.7728	0.7277	0.7271	0.7458	0.7129
8	2380	803	1003	2657	0.7361	0.7035	0.7679	0.7361	0.7358	0.7477	0.7260
9	2381	850	1002	2610	0.7294	0.7038	0.7543	0.7293	0.7291	0.7369	0.7226
10	2250	729	1133	2731	0.7279	0.6651	0.7893	0.7279	0.7268	0.7553	0.7068
Average					0.7301	0.6855	0.7737	0.7301	0.7294	0.7480	0.7158

ตารางภาคผนวกที่ 34 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคต้นไม้ตัดสินใจ กรณีคัดเลือกตัวแปรอิสระด้วยค่าสัมประสิทธิ์สหสัมพันธ์ ภายใต้ชุดข้อมูลโรคหัวใจและหลอดเลือด (ชุดข้อมูลเรียนรู้)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	21573	6595	8874	24547	0.7488	0.7085	0.7882	0.7488	0.7484	0.7659	0.7345
2	22190	7200	8257	23942	0.7490	0.7288	0.7688	0.7490	0.7489	0.7550	0.7436
3	21120	6288	9327	24854	0.7465	0.6937	0.7981	0.7465	0.7457	0.7706	0.7271
4	21029	5996	9418	25147	0.7497	0.6907	0.8075	0.7498	0.7488	0.7781	0.7275
5	20611	5709	9836	25434	0.7476	0.6769	0.8167	0.7476	0.7463	0.7831	0.7211
6	21493	6529	8954	24614	0.7486	0.7059	0.7904	0.7486	0.7481	0.7670	0.7333
7	21422	6713	9025	24430	0.7445	0.7036	0.7844	0.7445	0.7440	0.7614	0.7302
8	21688	6879	8759	24264	0.7461	0.7123	0.7791	0.7461	0.7458	0.7592	0.7348
9	21822	6925	8625	24218	0.7475	0.7167	0.7776	0.7475	0.7472	0.7591	0.7374
10	21001	6152	9446	24991	0.7467	0.6898	0.8025	0.7468	0.7459	0.7734	0.7257
Average					0.7475	0.7027	0.7913	0.7475	0.7469	0.7673	0.7315

ตารางภาคผนวกที่ 35 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคต้นไม้ตัดสินใจ กรณีคัดเลือกตัวแปรอิสระด้วยค่าสัมประสิทธิ์สหสัมพันธ์ ภายใต้ชุดข้อมูลโรคหัวใจและหลอดเลือด (ชุดข้อมูลทดสอบ)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	2323	819	1060	2642	0.7255	0.6867	0.7634	0.7255	0.7250	0.7393	0.7137
2	2389	914	994	2547	0.7212	0.7062	0.7359	0.7212	0.7211	0.7233	0.7193
3	2318	736	1065	2725	0.7368	0.6852	0.7873	0.7368	0.7361	0.7590	0.7190
4	2285	711	1098	2749	0.7356	0.6754	0.7945	0.7356	0.7346	0.7627	0.7146
5	2202	702	1181	2758	0.7248	0.6509	0.7971	0.7248	0.7233	0.7583	0.7002
6	2356	780	1027	2680	0.7359	0.6964	0.7746	0.7359	0.7355	0.7513	0.7230
7	2306	786	1077	2674	0.7278	0.6816	0.7728	0.7277	0.7271	0.7458	0.7129
8	2380	803	1003	2657	0.7361	0.7035	0.7679	0.7361	0.7358	0.7477	0.7260
9	2381	850	1002	2610	0.7294	0.7038	0.7543	0.7293	0.7291	0.7369	0.7226
10	2250	729	1133	2731	0.7279	0.6651	0.7893	0.7279	0.7268	0.7553	0.7068
Average					0.7301	0.6855	0.7737	0.7301	0.7294	0.7480	0.7158

ตารางภาคผนวกที่ 36 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคป่าสุ่ม โดยใช้ตัวแปรอิสระทั้งหมด ภายใต้ชุดข้อมูลโรคหัวใจและหลอดเลือด (ชุดข้อมูลเรียนรู้)

(mtry, ntree)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(3, 250)	0.8818	0.9148	0.8544	0.8843	0.8815	0.8390	0.9236
(3, 500)	0.8822	0.9153	0.8547	0.8847	0.8819	0.8394	0.9240
(4, 250)	0.9623	0.9864	0.9410	0.9634	0.9623	0.9367	0.9873
(4, 500)	0.9634	0.9877	0.9420	0.9646	0.9634	0.9377	0.9886

ตารางภาคผนวกที่ 37 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคป่าสุ่ม โดยใช้ตัวแปรอิสระทั้งหมด ภายใต้ชุดข้อมูลโรคหัวใจและหลอดเลือด (ชุดข้อมูลทดสอบ)

(mtry, ntree)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(3, 250)	0.7314	0.7453	0.7196	0.7323	0.7310	0.6939	0.7681
(3, 500)	0.7316	0.7455	0.7198	0.7325	0.7311	0.6940	0.7683
(4, 250)	0.7249	0.7341	0.7171	0.7255	0.7247	0.6954	0.7537
(4, 500)	0.7255	0.7347	0.7169	0.7257	0.7249	0.6952	0.7545

ตารางภาคผนวกที่ 38 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคป่าสุ่ม กรณีคัดเลือกตัวแปรอิสระ ด้วยเทคนิคการถดถอยลอจิสติกวิธีการเลือกแบบไปข้างหน้า โดยใช้ข้อมูลทั้งหมด ภายใต้ชุดข้อมูลโรคหัวใจและหลอดเลือด (ชุดข้อมูลเรียนรู้)

(mtry, ntree)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(3, 250)	0.8897	0.8470	0.9314	0.8897	0.8894	0.9235	0.8616
(3, 500)	0.8906	0.8476	0.9386	0.8936	0.8917	0.9248	0.8623
(4, 250)	0.9695	0.9478	0.9907	0.9695	0.9695	0.9901	0.9510
(4, 500)	0.9702	0.9484	0.9916	0.9703	0.9702	0.9910	0.9516

ตารางภาคผนวกที่ 39 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคป่าสุ่ม กรณีคัดเลือกตัวแปรอิสระ ด้วยเทคนิคการถดถอยลอจิสติกวิธีการเลือกแบบไปข้างหน้า โดยใช้ข้อมูลทั้งหมด ภายใต้ชุดข้อมูลโรคหัวใจและหลอดเลือด (ชุดข้อมูลทดสอบ)

(mtry, ntree)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(3, 250)	0.7300	0.6911	0.7681	0.7301	0.7296	0.7445	0.7178
(3, 500)	0.7306	0.6922	0.7681	0.7306	0.7301	0.7448	0.7185
(4, 250)	0.7229	0.6949	0.7503	0.7229	0.7227	0.7312	0.7156
(4, 500)	0.7232	0.6943	0.7514	0.7232	0.7229	0.7320	0.7154

ตารางภาคผนวกที่ 40 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคป่าสุ่ม กรณีคัดเลือกตัวแปรอิสระ ด้วยเทคนิคการถดถอยลอจิสติกวิธีกำจัดแบบถอยหลัง โดยใช้ข้อมูลทั้งหมด ภายใต้ชุดข้อมูลโรคหัวใจ และหลอดเลือด (ชุดข้อมูลเรียนรู้)

(mtry, ntree)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(3, 250)	0.8897	0.8470	0.9314	0.8897	0.8894	0.9235	0.8616
(3, 500)	0.8906	0.8476	0.9386	0.8936	0.8917	0.9248	0.8623
(4, 250)	0.9695	0.9478	0.9907	0.9695	0.9695	0.9901	0.9510
(4, 500)	0.9702	0.9484	0.9916	0.9703	0.9702	0.9910	0.9516

ตารางภาคผนวกที่ 41 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคป่าสุ่ม กรณีคัดเลือกตัวแปรอิสระ ด้วยเทคนิคการถดถอยลอจิสติกวิธีกำจัดแบบถอยหลัง โดยใช้ข้อมูลทั้งหมด ภายใต้ชุดข้อมูลโรคหัวใจ และหลอดเลือด (ชุดข้อมูลทดสอบ)

(mtry, ntree)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(3, 250)	0.7300	0.6911	0.7681	0.7301	0.7296	0.7445	0.7178
(3, 500)	0.7306	0.6922	0.7681	0.7306	0.7301	0.7448	0.7185
(4, 250)	0.7229	0.6949	0.7503	0.7229	0.7227	0.7312	0.7156
(4, 500)	0.7232	0.6943	0.7514	0.7232	0.7229	0.7320	0.7154

ตารางภาคผนวกที่ 42 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคป่าสุ่ม กรณีคัดเลือกตัวแปรอิสระ ด้วยเทคนิคการถดถอยลอจิสติกวิธีการถดถอยทีละชั้น โดยใช้ข้อมูลทั้งหมด ภายใต้ชุดข้อมูลโรคหัวใจ และหลอดเลือด (ชุดข้อมูลเรียนรู้)

(mtry, ntree)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(3, 250)	0.8897	0.8470	0.9314	0.8897	0.8894	0.9235	0.8616
(3, 500)	0.8906	0.8476	0.9386	0.8936	0.8917	0.9248	0.8623
(4, 250)	0.9695	0.9478	0.9907	0.9695	0.9695	0.9901	0.9510
(4, 500)	0.9702	0.9484	0.9916	0.9703	0.9702	0.9910	0.9516

ตารางภาคผนวกที่ 43 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคป่าสุ่ม กรณีคัดเลือกตัวแปรอิสระ ด้วยเทคนิคการถดถอยลอจิสติกวิธีการถดถอยทีละชั้น โดยใช้ข้อมูลทั้งหมด ภายใต้ชุดข้อมูลโรคหัวใจ และหลอดเลือด (ชุดข้อมูลทดสอบ)

(mtry, ntree)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(3, 250)	0.7300	0.6911	0.7681	0.7301	0.7296	0.7445	0.7178
(3, 500)	0.7306	0.6922	0.7681	0.7306	0.7301	0.7448	0.7185
(4, 250)	0.7229	0.6949	0.7503	0.7229	0.7227	0.7312	0.7156

(mtry, ntree)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(4, 500)	0.7232	0.6943	0.7514	0.7232	0.7229	0.7320	0.7154

ตารางภาคผนวกที่ 44 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคป่าสุ่ม กรณีคัดเลือกตัวแปรอิสระ ด้วยเทคนิคการถดถอยลอจิสติกวิธีการเลือกแบบไปข้างหน้า โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้อุณหภูมิและหอดเลืด (ชุดข้อมูลเรียนรู้)

(mtry, ntree)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(3, 250)	0.8897	0.8470	0.9314	0.8897	0.8894	0.9235	0.8616
(3, 500)	0.8906	0.8476	0.9386	0.8936	0.8917	0.9248	0.8623
(4, 250)	0.9695	0.9478	0.9907	0.9695	0.9695	0.9901	0.9510
(4, 500)	0.9702	0.9484	0.9916	0.9703	0.9702	0.9910	0.9516

ตารางภาคผนวกที่ 45 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคป่าสุ่ม กรณีคัดเลือกตัวแปรอิสระ ด้วยเทคนิคการถดถอยลอจิสติกวิธีการเลือกแบบไปข้างหน้า โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้อุณหภูมิและหอดเลืด (ชุดข้อมูลทดสอบ)

(mtry, ntree)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(3, 250)	0.7300	0.6911	0.7681	0.7301	0.7296	0.7445	0.7178
(3, 500)	0.7306	0.6922	0.7681	0.7306	0.7301	0.7448	0.7185
(4, 250)	0.7229	0.6949	0.7503	0.7229	0.7227	0.7312	0.7156
(4, 500)	0.7232	0.6943	0.7514	0.7232	0.7229	0.7320	0.7154

ตารางภาคผนวกที่ 46 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคป่าสุ่ม กรณีคัดเลือกตัวแปรอิสระ ด้วยเทคนิคการถดถอยลอจิสติกวิธีการกำจัดแบบถดถอยหลัง โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้อุณหภูมิและหอดเลืด (ชุดข้อมูลเรียนรู้)

(mtry, ntree)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(3, 250)	0.8897	0.8470	0.9314	0.8897	0.8894	0.9235	0.8616
(3, 500)	0.8906	0.8476	0.9386	0.8936	0.8917	0.9248	0.8623
(4, 250)	0.9695	0.9478	0.9907	0.9695	0.9695	0.9901	0.9510
(4, 500)	0.9702	0.9484	0.9916	0.9703	0.9702	0.9910	0.9516

ตารางภาคผนวกที่ 47 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคป่าสุ่ม กรณีคัดเลือกตัวแปรอิสระ ด้วยเทคนิคการถดถอยลอจิสติกวิธีกำจัดแบบถอยหลัง โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้อุณหภูมิของข้อมูลโรควิวและหลอดเลือด (ชุดข้อมูลทดสอบ)

(mtry, ntree)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(3, 250)	0.7300	0.6911	0.7681	0.7301	0.7296	0.7445	0.7178
(3, 500)	0.7306	0.6922	0.7681	0.7306	0.7301	0.7448	0.7185
(4, 250)	0.7229	0.6949	0.7503	0.7229	0.7227	0.7312	0.7156
(4, 500)	0.7232	0.6943	0.7514	0.7232	0.7229	0.7320	0.7154

ตารางภาคผนวกที่ 48 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคป่าสุ่ม กรณีคัดเลือกตัวแปรอิสระ ด้วยเทคนิคการถดถอยลอจิสติกวิธีการถดถอยทีละขั้น โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้อุณหภูมิของข้อมูลโรควิวและหลอดเลือด (ชุดข้อมูลเรียนรู้)

(mtry, ntree)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(3, 250)	0.8897	0.8470	0.9314	0.8897	0.8894	0.9235	0.8616
(3, 500)	0.8906	0.8476	0.9386	0.8936	0.8917	0.9248	0.8623
(4, 250)	0.9695	0.9478	0.9907	0.9695	0.9695	0.9901	0.9510
(4, 500)	0.9702	0.9484	0.9916	0.9703	0.9702	0.9910	0.9516

ตารางภาคผนวกที่ 49 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคป่าสุ่ม กรณีคัดเลือกตัวแปรอิสระ ด้วยเทคนิคการถดถอยลอจิสติกวิธีการถดถอยทีละขั้น โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้อุณหภูมิของข้อมูลโรควิวและหลอดเลือด (ชุดข้อมูลทดสอบ)

(mtry, ntree)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(3, 250)	0.7300	0.6911	0.7681	0.7301	0.7296	0.7445	0.7178
(3, 500)	0.7306	0.6922	0.7681	0.7306	0.7301	0.7448	0.7185
(4, 250)	0.7229	0.6949	0.7503	0.7229	0.7227	0.7312	0.7156
(4, 500)	0.7232	0.6943	0.7514	0.7232	0.7229	0.7320	0.7154

ตารางภาคผนวกที่ 50 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคป่าสุ่ม กรณีคัดเลือกตัวแปรอิสระ ด้วยค่าสัมประสิทธิ์สหสัมพันธ์ ภายใต้อุณหภูมิของข้อมูลโรควิวและหลอดเลือด (ชุดข้อมูลเรียนรู้)

(mtry, ntree)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(3, 250)	0.8897	0.8470	0.9314	0.8897	0.8894	0.9235	0.8616
(3, 500)	0.8906	0.8476	0.9386	0.8936	0.8917	0.9248	0.8623
(4, 250)	0.9695	0.9478	0.9907	0.9695	0.9695	0.9901	0.9510
(4, 500)	0.9702	0.9484	0.9916	0.9703	0.9702	0.9910	0.9516

ตารางภาคผนวกที่ 51 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคป่าสุ่ม กรณีคัดเลือกตัวแปรอิสระ ด้วยค่าสัมประสิทธิ์สหสัมพันธ์ ภายใต้ชุดข้อมูลโรคหัวใจและหลอดเลือด (ชุดข้อมูลทดสอบ)

(mtry, ntree)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(3, 250)	0.7300	0.6911	0.7681	0.7301	0.7296	0.7445	0.7178
(3, 500)	0.7306	0.6922	0.7681	0.7306	0.7301	0.7448	0.7185
(4, 250)	0.7229	0.6949	0.7503	0.7229	0.7227	0.7312	0.7156
(4, 500)	0.7232	0.6943	0.7514	0.7232	0.7229	0.7320	0.7154

ตารางภาคผนวกที่ 52 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ตัวแปรอิสระทั้งหมด ภายใต้ชุดข้อมูลโรคหัวใจและหลอดเลือด (ชุดข้อมูลเรียนรู้)

(gamma, C)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(0.001, 1)	0.7273	0.7734	0.6959	0.7342	0.7248	0.6341	0.8184
(0.01, 1)	0.7314	0.7602	0.7093	0.7345	0.7302	0.6669	0.7944
(0.1, 1)	0.7370	0.7661	0.7149	0.7402	0.7360	0.6738	0.7989
(0.2, 1)	0.7408	0.7698	0.7184	0.7438	0.7397	0.6786	0.8016
(0.001, 10)	0.7296	0.7706	0.7007	0.7352	0.7276	0.6450	0.8123
(0.01, 10)	0.7327	0.7651	0.7085	0.7365	0.7314	0.6630	0.8010
(0.1, 10)	0.7419	0.7722	0.7187	0.7452	0.7408	0.6780	0.8045
(0.2, 10)	0.7561	0.7868	0.7323	0.7592	0.7551	0.6948	0.8159

ตารางภาคผนวกที่ 53 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ตัวแปรอิสระทั้งหมด ภายใต้ชุดข้อมูลโรคหัวใจและหลอดเลือด (ชุดข้อมูลทดสอบ)

(gamma, C)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(0.001, 1)	0.7272	0.7734	0.6957	0.7341	0.7247	0.6339	0.8184
(0.01, 1)	0.7309	0.7600	0.7089	0.7342	0.7298	0.6665	0.7939
(0.1, 1)	0.7345	0.7632	0.7125	0.7376	0.7334	0.6714	0.7963
(0.2, 1)	0.7341	0.7623	0.7125	0.7371	0.7330	0.6717	0.7952
(0.001, 10)	0.7295	0.7704	0.7006	0.7351	0.7275	0.6451	0.8120
(0.01, 10)	0.7319	0.7641	0.7078	0.7357	0.7306	0.6621	0.8002
(0.1, 10)	0.7334	0.7625	0.7111	0.7365	0.7322	0.6691	0.7962
(0.2, 10)	0.7294	0.7556	0.7090	0.7320	0.7283	0.6691	0.7883

ตารางภาคผนวกที่ 54 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคซ์พอร์ตเวกเตอร์แมชชีน กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการเลือกแบบไปข้างหน้า โดยใช้ข้อมูลทั้งหมด ภายใต้ชุดข้อมูลโรคหัวใจและหลอดเลือด (ชุดข้อมูลเรียนรู้)

(gamma, C)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(0.001, 1)	0.7273	0.6341	0.8184	0.7273	0.7248	0.7734	0.6958
(0.01, 1)	0.7314	0.6670	0.7944	0.7314	0.7302	0.7602	0.7093
(0.1, 1)	0.7363	0.6711	0.8000	0.7263	0.7351	0.7664	0.7133
(0.2, 1)	0.7394	0.6750	0.8025	0.7394	0.7383	0.7696	0.7163
(0.001, 10)	0.7293	0.6449	0.8121	0.7294	0.7274	0.7704	0.7005
(0.01, 10)	0.7326	0.6621	0.8015	0.7326	0.7312	0.7653	0.7081
(0.1, 10)	0.7401	0.6741	0.8047	0.7401	0.7389	0.7714	0.7164
(0.2, 10)	0.7519	0.6904	0.8120	0.7519	0.7509	0.7822	0.7285

ตารางภาคผนวกที่ 55 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคซ์พอร์ตเวกเตอร์แมชชีน กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการเลือกแบบไปข้างหน้า โดยใช้ข้อมูลทั้งหมด ภายใต้ชุดข้อมูลโรคหัวใจและหลอดเลือด (ชุดข้อมูลทดสอบ)

(gamma, C)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(0.001, 1)	0.7271	0.6337	0.8184	0.7271	0.7246	0.7733	0.6956
(0.01, 1)	0.7310	0.6665	0.7941	0.7310	0.7298	0.7599	0.7089
(0.1, 1)	0.7330	0.6690	0.7975	0.7340	0.7328	0.7636	0.7113
(0.2, 1)	0.7341	0.6694	0.7973	0.7341	0.7329	0.7635	0.7116
(0.001, 10)	0.7296	0.6448	0.8121	0.7294	0.7275	0.7705	0.7007
(0.01, 10)	0.7320	0.6617	0.8008	0.7320	0.7306	0.7646	0.7077
(0.1, 10)	0.7337	0.6677	0.7982	0.7337	0.7324	0.7639	0.7107
(0.2, 10)	0.7301	0.6690	0.7899	0.7301	0.7290	0.7569	0.7094

ตารางภาคผนวกที่ 56 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคซ์พอร์ตเวกเตอร์แมชชีน กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการกำจัดแบบถอยหลัง โดยใช้ข้อมูลทั้งหมด ภายใต้ชุดข้อมูลโรคหัวใจและหลอดเลือด (ชุดข้อมูลเรียนรู้)

(gamma, C)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(0.001, 1)	0.7273	0.6341	0.8184	0.7273	0.7248	0.7734	0.6958
(0.01, 1)	0.7314	0.6670	0.7944	0.7314	0.7302	0.7602	0.7093
(0.1, 1)	0.7363	0.6711	0.8000	0.7263	0.7351	0.7664	0.7133
(0.2, 1)	0.7394	0.6750	0.8025	0.7394	0.7383	0.7696	0.7163
(0.001, 10)	0.7293	0.6449	0.8121	0.7294	0.7274	0.7704	0.7005
(0.01, 10)	0.7326	0.6621	0.8015	0.7326	0.7312	0.7653	0.7081

(gamma, C)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(0.1, 10)	0.7401	0.6741	0.8047	0.7401	0.7389	0.7714	0.7164
(0.2, 10)	0.7519	0.6904	0.8120	0.7519	0.7509	0.7822	0.7285

ตารางภาคผนวกที่ 57 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคซ์พอร์ตเวกเตอร์แมชชีน กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีกำจัดแบบถอยหลัง โดยใช้ข้อมูลทั้งหมด ภายใต้ชุดข้อมูลโรคหัวใจและหลอดเลือด (ชุดข้อมูลทดสอบ)

(gamma, C)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(0.001, 1)	0.7271	0.6337	0.8184	0.7271	0.7246	0.7733	0.6956
(0.01, 1)	0.7310	0.6665	0.7941	0.7310	0.7298	0.7599	0.7089
(0.1, 1)	0.7330	0.6690	0.7975	0.7340	0.7328	0.7636	0.7113
(0.2, 1)	0.7341	0.6694	0.7973	0.7341	0.7329	0.7635	0.7116
(0.001, 10)	0.7296	0.6448	0.8121	0.7294	0.7275	0.7705	0.7007
(0.01, 10)	0.7320	0.6617	0.8008	0.7320	0.7306	0.7646	0.7077
(0.1, 10)	0.7337	0.6677	0.7982	0.7337	0.7324	0.7639	0.7107
(0.2, 10)	0.7301	0.6690	0.7899	0.7301	0.7290	0.7569	0.7094

ตารางภาคผนวกที่ 58 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคซ์พอร์ตเวกเตอร์แมชชีน กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีกำจัดแบบถอยทีละชั้น โดยใช้ข้อมูลทั้งหมด ภายใต้ชุดข้อมูลโรคหัวใจและหลอดเลือด (ชุดข้อมูลเรียนรู้)

(gamma, C)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(0.001, 1)	0.7273	0.6341	0.8184	0.7273	0.7248	0.7734	0.6958
(0.01, 1)	0.7314	0.6670	0.7944	0.7314	0.7302	0.7602	0.7093
(0.1, 1)	0.7363	0.6711	0.8000	0.7263	0.7351	0.7664	0.7133
(0.2, 1)	0.7394	0.6750	0.8025	0.7394	0.7383	0.7696	0.7163
(0.001, 10)	0.7293	0.6449	0.8121	0.7294	0.7274	0.7704	0.7005
(0.01, 10)	0.7326	0.6621	0.8015	0.7326	0.7312	0.7653	0.7081
(0.1, 10)	0.7401	0.6741	0.8047	0.7401	0.7389	0.7714	0.7164
(0.2, 10)	0.7519	0.6904	0.8120	0.7519	0.7509	0.7822	0.7285

ตารางภาคผนวกที่ 59 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคซ์พอร์ตเวกเตอร์แมชชีน กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีกำจัดแบบถอยทีละชั้น โดยใช้ข้อมูลทั้งหมด ภายใต้ชุดข้อมูลโรคหัวใจและหลอดเลือด (ชุดข้อมูลทดสอบ)

(gamma, C)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(0.001, 1)	0.7271	0.6337	0.8184	0.7271	0.7246	0.7733	0.6956

(gamma, C)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(0.01, 1)	0.7310	0.6665	0.7941	0.7310	0.7298	0.7599	0.7089
(0.1, 1)	0.7330	0.6690	0.7975	0.7340	0.7328	0.7636	0.7113
(0.2, 1)	0.7341	0.6694	0.7973	0.7341	0.7329	0.7635	0.7116
(0.001, 10)	0.7296	0.6448	0.8121	0.7294	0.7275	0.7705	0.7007
(0.01, 10)	0.7320	0.6617	0.8008	0.7320	0.7306	0.7646	0.7077
(0.1, 10)	0.7337	0.6677	0.7982	0.7337	0.7324	0.7639	0.7107
(0.2, 10)	0.7301	0.6690	0.7899	0.7301	0.7290	0.7569	0.7094

ตารางภาคผนวกที่ 60 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคซัพพอร์ตเวกเตอร์แมชชีน กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการเลือกแบบไปข้างหน้า โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้อัตราข้อมูลโรคหัวใจและหลอดเลือด (ชุดข้อมูลเรียนรู้)

(gamma, C)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(0.001, 1)	0.7273	0.6341	0.8184	0.7273	0.7248	0.7734	0.6958
(0.01, 1)	0.7314	0.6670	0.7944	0.7314	0.7302	0.7602	0.7093
(0.1, 1)	0.7363	0.6711	0.8000	0.7263	0.7351	0.7664	0.7133
(0.2, 1)	0.7394	0.6750	0.8025	0.7394	0.7383	0.7696	0.7163
(0.001, 10)	0.7293	0.6449	0.8121	0.7294	0.7274	0.7704	0.7005
(0.01, 10)	0.7326	0.6621	0.8015	0.7326	0.7312	0.7653	0.7081
(0.1, 10)	0.7401	0.6741	0.8047	0.7401	0.7389	0.7714	0.7164
(0.2, 10)	0.7519	0.6904	0.8120	0.7519	0.7509	0.7822	0.7285

ตารางภาคผนวกที่ 61 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคซัพพอร์ตเวกเตอร์แมชชีน กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการเลือกแบบไปข้างหน้า โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้อัตราข้อมูลโรคหัวใจและหลอดเลือด (ชุดข้อมูลทดสอบ)

(gamma, C)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(0.001, 1)	0.7271	0.6337	0.8184	0.7271	0.7246	0.7733	0.6956
(0.01, 1)	0.7310	0.6665	0.7941	0.7310	0.7298	0.7599	0.7089
(0.1, 1)	0.7330	0.6690	0.7975	0.7340	0.7328	0.7636	0.7113
(0.2, 1)	0.7341	0.6694	0.7973	0.7341	0.7329	0.7635	0.7116
(0.001, 10)	0.7296	0.6448	0.8121	0.7294	0.7275	0.7705	0.7007
(0.01, 10)	0.7320	0.6617	0.8008	0.7320	0.7306	0.7646	0.7077
(0.1, 10)	0.7337	0.6677	0.7982	0.7337	0.7324	0.7639	0.7107
(0.2, 10)	0.7301	0.6690	0.7899	0.7301	0.7290	0.7569	0.7094

ตารางภาคผนวกที่ 62 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคซ์พอร์ดเวกเตอร์แมชชีน กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีกำจัดแบบถอยหลัง โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้ชุดข้อมูลโรคหัวใจและหลอดเลือด (ชุดข้อมูลเรียนรู้)

(gamma, C)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(0.001, 1)	0.7273	0.6341	0.8184	0.7273	0.7248	0.7734	0.6958
(0.01, 1)	0.7314	0.6670	0.7944	0.7314	0.7302	0.7602	0.7093
(0.1, 1)	0.7363	0.6711	0.8000	0.7263	0.7351	0.7664	0.7133
(0.2, 1)	0.7394	0.6750	0.8025	0.7394	0.7383	0.7696	0.7163
(0.001, 10)	0.7293	0.6449	0.8121	0.7294	0.7274	0.7704	0.7005
(0.01, 10)	0.7326	0.6621	0.8015	0.7326	0.7312	0.7653	0.7081
(0.1, 10)	0.7401	0.6741	0.8047	0.7401	0.7389	0.7714	0.7164
(0.2, 10)	0.7519	0.6904	0.8120	0.7519	0.7509	0.7822	0.7285

ตารางภาคผนวกที่ 63 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคซ์พอร์ดเวกเตอร์แมชชีน กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีกำจัดแบบถอยหลัง โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้ชุดข้อมูลโรคหัวใจและหลอดเลือด (ชุดข้อมูลทดสอบ)

(gamma, C)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(0.001, 1)	0.7271	0.6337	0.8184	0.7271	0.7246	0.7733	0.6956
(0.01, 1)	0.7310	0.6665	0.7941	0.7310	0.7298	0.7599	0.7089
(0.1, 1)	0.7330	0.6690	0.7975	0.7340	0.7328	0.7636	0.7113
(0.2, 1)	0.7341	0.6694	0.7973	0.7341	0.7329	0.7635	0.7116
(0.001, 10)	0.7296	0.6448	0.8121	0.7294	0.7275	0.7705	0.7007
(0.01, 10)	0.7320	0.6617	0.8008	0.7320	0.7306	0.7646	0.7077
(0.1, 10)	0.7337	0.6677	0.7982	0.7337	0.7324	0.7639	0.7107
(0.2, 10)	0.7301	0.6690	0.7899	0.7301	0.7290	0.7569	0.7094

ตารางภาคผนวกที่ 64 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคซ์พอร์ดเวกเตอร์แมชชีน กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการถดถอยทีละชั้น โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้ชุดข้อมูลโรคหัวใจและหลอดเลือด (ชุดข้อมูลเรียนรู้)

(gamma, C)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(0.001, 1)	0.7273	0.6341	0.8184	0.7273	0.7248	0.7734	0.6958

(gamma, C)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(0.01, 1)	0.7314	0.6670	0.7944	0.7314	0.7302	0.7602	0.7093
(0.1, 1)	0.7363	0.6711	0.8000	0.7263	0.7351	0.7664	0.7133
(0.2, 1)	0.7394	0.6750	0.8025	0.7394	0.7383	0.7696	0.7163
(0.001, 10)	0.7293	0.6449	0.8121	0.7294	0.7274	0.7704	0.7005
(0.01, 10)	0.7326	0.6621	0.8015	0.7326	0.7312	0.7653	0.7081
(0.1, 10)	0.7401	0.6741	0.8047	0.7401	0.7389	0.7714	0.7164
(0.2, 10)	0.7519	0.6904	0.8120	0.7519	0.7509	0.7822	0.7285

ตารางภาคผนวกที่ 65 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคซ์พอร์ตเวกเตอร์แมชชีน กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการถดถอยทีละชั้น โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้ชุดข้อมูลโรคหัวใจและหลอดเลือด (ชุดข้อมูลทดสอบ)

(gamma, C)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(0.001, 1)	0.7271	0.6337	0.8184	0.7271	0.7246	0.7733	0.6956
(0.01, 1)	0.7310	0.6665	0.7941	0.7310	0.7298	0.7599	0.7089
(0.1, 1)	0.7330	0.6690	0.7975	0.7340	0.7328	0.7636	0.7113
(0.2, 1)	0.7341	0.6694	0.7973	0.7341	0.7329	0.7635	0.7116
(0.001, 10)	0.7296	0.6448	0.8121	0.7294	0.7275	0.7705	0.7007
(0.01, 10)	0.7320	0.6617	0.8008	0.7320	0.7306	0.7646	0.7077
(0.1, 10)	0.7337	0.6677	0.7982	0.7337	0.7324	0.7639	0.7107
(0.2, 10)	0.7301	0.6690	0.7899	0.7301	0.7290	0.7569	0.7094

ตารางภาคผนวกที่ 66 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคซ์พอร์ตเวกเตอร์แมชชีน กรณีคัดเลือกตัวแปรอิสระด้วยค่าสัมประสิทธิ์สหสัมพันธ์ ภายใต้ชุดข้อมูลโรคหัวใจและหลอดเลือด (ชุดข้อมูลเรียนรู้)

(gamma, C)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(0.001, 1)	0.7273	0.6341	0.8184	0.7273	0.7248	0.7734	0.6958
(0.01, 1)	0.7314	0.6670	0.7944	0.7314	0.7302	0.7602	0.7093
(0.1, 1)	0.7363	0.6711	0.8000	0.7263	0.7351	0.7664	0.7133
(0.2, 1)	0.7394	0.6750	0.8025	0.7394	0.7383	0.7696	0.7163
(0.001, 10)	0.7293	0.6449	0.8121	0.7294	0.7274	0.7704	0.7005
(0.01, 10)	0.7326	0.6621	0.8015	0.7326	0.7312	0.7653	0.7081
(0.1, 10)	0.7401	0.6741	0.8047	0.7401	0.7389	0.7714	0.7164
(0.2, 10)	0.7519	0.6904	0.8120	0.7519	0.7509	0.7822	0.7285

ตารางภาคผนวกที่ 67 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคซ์พอร์ตเวกเตอร์แมชชีน กรณีคัดเลือกตัวแปรอิสระด้วยค่าสัมประสิทธิ์สหสัมพันธ์ ภายใต้ชุดข้อมูลโรคหัวใจและหลอดเลือด (ชุดข้อมูลทดสอบ)

(gamma, C)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(0.001, 1)	0.7271	0.6337	0.8184	0.7271	0.7246	0.7733	0.6956
(0.01, 1)	0.7310	0.6665	0.7941	0.7310	0.7298	0.7599	0.7089
(0.1, 1)	0.7330	0.6690	0.7975	0.7340	0.7328	0.7636	0.7113
(0.2, 1)	0.7341	0.6694	0.7973	0.7341	0.7329	0.7635	0.7116
(0.001, 10)	0.7296	0.6448	0.8121	0.7294	0.7275	0.7705	0.7007
(0.01, 10)	0.7320	0.6617	0.8008	0.7320	0.7306	0.7646	0.7077
(0.1, 10)	0.7337	0.6677	0.7982	0.7337	0.7324	0.7639	0.7107
(0.2, 10)	0.7301	0.6690	0.7899	0.7301	0.7290	0.7569	0.7094

ชุดข้อมูลโรคเบาหวาน

ตารางภาคผนวกที่ 68 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคการถดถอยลอจิสติก โดยใช้ตัวแปรอิสระทั้งหมด ภายใต้ชุดข้อมูลโรคเบาหวาน (ชุดข้อมูลเรียนรู้)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	23804	8484	21882	7150	0.7450	0.7690	0.7206	0.7450	0.7449	0.7372	0.7537
2	23828	8462	21904	7126	0.7458	0.7698	0.7213	0.7458	0.7456	0.7379	0.7545
3	23764	8500	21866	7190	0.7441	0.7677	0.7201	0.7441	0.7440	0.7365	0.7525
4	23842	8473	21893	7112	0.7458	0.7702	0.7210	0.7458	0.7457	0.7378	0.7548
5	23834	8491	21875	7121	0.7454	0.7700	0.7204	0.7454	0.7452	0.7373	0.7544
6	23773	8482	21884	7182	0.7446	0.7680	0.7207	0.7446	0.7444	0.7370	0.7529
7	23817	8555	21811	7138	0.7441	0.7694	0.7183	0.7441	0.7439	0.7357	0.7534
8	23783	8496	21870	7172	0.7445	0.7683	0.7202	0.7445	0.7443	0.7368	0.7530
9	23792	8504	21862	7163	0.7445	0.7686	0.7199	0.7445	0.7443	0.7367	0.7532
10	23803	8509	21857	7152	0.7446	0.7690	0.7198	0.7446	0.7444	0.7367	0.7535
Average					0.7448	0.7690	0.7202	0.7448	0.7447	0.7370	0.7536

ตารางภาคผนวกที่ 69 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคการถดถอยลอจิสติก โดยใช้ตัวแปรอิสระทั้งหมด ภายใต้ชุดข้อมูลโรคเบาหวาน (ชุดข้อมูลทดสอบ)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	2585	918	2456	855	0.7398	0.7515	0.7279	0.7398	0.7398	0.7379	0.7418

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
2	2571	933	2441	869	0.7355	0.7474	0.7235	0.7356	0.7355	0.7337	0.7375
3	2679	952	2422	761	0.7486	0.7788	0.7178	0.7486	0.7483	0.7378	0.7609
4	2631	972	2402	809	0.7386	0.7648	0.7119	0.7386	0.7384	0.7302	0.7481
5	2632	945	2429	807	0.7428	0.7653	0.7199	0.7428	0.7427	0.7358	0.7506
6	2617	917	2457	822	0.7448	0.7610	0.7282	0.7448	0.7447	0.7405	0.7493
7	2641	908	2466	798	0.7496	0.7680	0.7309	0.7496	0.7495	0.7442	0.7555
8	2722	992	2382	717	0.7492	0.7915	0.7060	0.7492	0.7487	0.7329	0.7686
9	2634	934	2440	805	0.7448	0.7659	0.7232	0.7448	0.7446	0.7382	0.7519
10	2702	972	2402	737	0.7492	0.7857	0.7119	0.7492	0.7488	0.7354	0.7652
Average					0.7443	0.7680	0.7201	0.7443	0.7441	0.7367	0.7529

ตารางภาคผนวกที่ 70 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคการถดถอยลอจิสติก
กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการเลือกแบบไปข้างหน้า โดยใช้ข้อมูล
ทั้งหมด ภายใต้ชุดข้อมูลโรคเบาหวาน (ชุดข้อมูลเรียนรู้)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	23807	8493	7147	21873	0.7449	0.7691	0.7203	0.7449	0.7448	0.7371	0.7537
2	23815	8461	7139	21905	0.7456	0.7694	0.7214	0.7456	0.7454	0.7379	0.7542
3	23756	8496	7198	21870	0.7441	0.7675	0.7202	0.7441	0.7439	0.7366	0.7524
4	23827	8484	7127	21882	0.7454	0.7698	0.7206	0.7454	0.7452	0.7374	0.7543
5	23837	8500	7118	21866	0.7453	0.7701	0.7201	0.7453	0.7451	0.7371	0.7544
6	23772	8480	7183	21886	0.7446	0.7680	0.7207	0.7446	0.7444	0.7371	0.7529
7	23805	8549	7150	21817	0.7440	0.7690	0.7185	0.7440	0.7438	0.7358	0.7532
8	23777	8503	7178	21863	0.7443	0.7681	0.7200	0.7443	0.7441	0.7366	0.7528
9	23797	8494	7158	21872	0.7448	0.7688	0.7203	0.7448	0.7446	0.7370	0.7534
10	23783	8493	7172	21873	0.7445	0.7683	0.7203	0.7445	0.7444	0.7369	0.7531
Average					0.7448	0.7688	0.7202	0.7448	0.7446	0.7369	0.7534

ตารางภาคผนวกที่ 71 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคการถดถอยลอจิสติก
กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการเลือกแบบไปข้างหน้า โดยใช้ข้อมูล
ทั้งหมด ภายใต้ชุดข้อมูลโรคเบาหวาน (ชุดข้อมูลทดสอบ)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	2582	924	858	2450	0.7385	0.7506	0.7261	0.7406	0.7406	0.7406	0.7406
2	2571	936	869	2438	0.7351	0.7474	0.7226	0.7372	0.7372	0.7372	0.7372

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
3	2684	945	756	2429	0.7504	0.7802	0.7199	0.7626	0.7626	0.7626	0.7626
4	2630	972	810	2402	0.7385	0.7645	0.7119	0.7478	0.7478	0.7478	0.7478
5	2626	947	813	2427	0.7417	0.7636	0.7193	0.7491	0.7491	0.7491	0.7491
6	2617	916	822	2458	0.7449	0.7610	0.7285	0.7494	0.7494	0.7494	0.7494
7	2644	905	795	2469	0.7505	0.7688	0.7318	0.7564	0.7564	0.7564	0.7564
8	2726	992	713	2382	0.7497	0.7927	0.7060	0.7696	0.7696	0.7696	0.7696
9	2634	934	805	2440	0.7448	0.7659	0.7232	0.7519	0.7519	0.7519	0.7519
10	2704	978	735	2396	0.7486	0.7863	0.7101	0.7653	0.7653	0.7653	0.7653
Average					0.7443	0.7681	0.7199	0.7443	0.7441	0.7366	0.7530

ตารางภาคผนวกที่ 72 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคการถดถอยลอจิสติกกรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีกำจัดแบบถอยหลัง โดยใช้ข้อมูลทั้งหมดภายใต้ชุดข้อมูลโรคเบาหวาน (ชุดข้อมูลเรียนรู้)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	23807	8493	7147	21873	0.7449	0.7691	0.7203	0.7449	0.7448	0.7371	0.7537
2	23815	8461	7139	21905	0.7456	0.7694	0.7214	0.7456	0.7454	0.7379	0.7542
3	23756	8496	7198	21870	0.7441	0.7675	0.7202	0.7441	0.7439	0.7366	0.7524
4	23827	8484	7127	21882	0.7454	0.7698	0.7206	0.7454	0.7452	0.7374	0.7543
5	23837	8500	7118	21866	0.7453	0.7701	0.7201	0.7453	0.7451	0.7371	0.7544
6	23772	8480	7183	21886	0.7446	0.7680	0.7207	0.7446	0.7444	0.7371	0.7529
7	23805	8549	7150	21817	0.7440	0.7690	0.7185	0.7440	0.7438	0.7358	0.7532
8	23777	8503	7178	21863	0.7443	0.7681	0.7200	0.7443	0.7441	0.7366	0.7528
9	23797	8494	7158	21872	0.7448	0.7688	0.7203	0.7448	0.7446	0.7370	0.7534
10	23783	8493	7172	21873	0.7445	0.7683	0.7203	0.7445	0.7444	0.7369	0.7531
Average					0.7448	0.7688	0.7202	0.7448	0.7446	0.7369	0.7534

ตารางภาคผนวกที่ 73 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคการถดถอยลอจิสติกกรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีกำจัดแบบถอยหลัง โดยใช้ข้อมูลทั้งหมดภายใต้ชุดข้อมูลโรคเบาหวาน (ชุดข้อมูลทดสอบ)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	2582	924	858	2450	0.7385	0.7506	0.7261	0.7406	0.7406	0.7406	0.7406
2	2571	936	869	2438	0.7351	0.7474	0.7226	0.7372	0.7372	0.7372	0.7372
3	2684	945	756	2429	0.7504	0.7802	0.7199	0.7626	0.7626	0.7626	0.7626

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
4	2630	972	810	2402	0.7385	0.7645	0.7119	0.7478	0.7478	0.7478	0.7478
5	2626	947	813	2427	0.7417	0.7636	0.7193	0.7491	0.7491	0.7491	0.7491
6	2617	916	822	2458	0.7449	0.7610	0.7285	0.7494	0.7494	0.7494	0.7494
7	2644	905	795	2469	0.7505	0.7688	0.7318	0.7564	0.7564	0.7564	0.7564
8	2726	992	713	2382	0.7497	0.7927	0.7060	0.7696	0.7696	0.7696	0.7696
9	2634	934	805	2440	0.7448	0.7659	0.7232	0.7519	0.7519	0.7519	0.7519
10	2704	978	735	2396	0.7486	0.7863	0.7101	0.7653	0.7653	0.7653	0.7653
Average					0.7443	0.7681	0.7199	0.7443	0.7441	0.7366	0.7530

ตารางภาคผนวกที่ 74 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคการถดถอยลอจิสติกกรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการถดถอยทีละชั้น โดยใช้ข้อมูลทั้งหมดภายใต้ชุดข้อมูลโรคเบาหวาน (ชุดข้อมูลเรียนรู้)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	23807	8493	7147	21873	0.7449	0.7691	0.7203	0.7449	0.7448	0.7371	0.7537
2	23815	8461	7139	21905	0.7456	0.7694	0.7214	0.7456	0.7454	0.7379	0.7542
3	23756	8496	7198	21870	0.7441	0.7675	0.7202	0.7441	0.7439	0.7366	0.7524
4	23827	8484	7127	21882	0.7454	0.7698	0.7206	0.7454	0.7452	0.7374	0.7543
5	23837	8500	7118	21866	0.7453	0.7701	0.7201	0.7453	0.7451	0.7371	0.7544
6	23772	8480	7183	21886	0.7446	0.7680	0.7207	0.7446	0.7444	0.7371	0.7529
7	23805	8549	7150	21817	0.7440	0.7690	0.7185	0.7440	0.7438	0.7358	0.7532
8	23777	8503	7178	21863	0.7443	0.7681	0.7200	0.7443	0.7441	0.7366	0.7528
9	23797	8494	7158	21872	0.7448	0.7688	0.7203	0.7448	0.7446	0.7370	0.7534
10	23783	8493	7172	21873	0.7445	0.7683	0.7203	0.7445	0.7444	0.7369	0.7531
Average					0.7448	0.7688	0.7202	0.7448	0.7446	0.7369	0.7534

ตารางภาคผนวกที่ 75 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคการถดถอยลอจิสติกกรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการถดถอยทีละชั้น โดยใช้ข้อมูลทั้งหมดภายใต้ชุดข้อมูลโรคเบาหวาน (ชุดข้อมูลทดสอบ)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	2582	924	858	2450	0.7385	0.7506	0.7261	0.7406	0.7406	0.7406	0.7406
2	2571	936	869	2438	0.7351	0.7474	0.7226	0.7372	0.7372	0.7372	0.7372
3	2684	945	756	2429	0.7504	0.7802	0.7199	0.7626	0.7626	0.7626	0.7626
4	2630	972	810	2402	0.7385	0.7645	0.7119	0.7478	0.7478	0.7478	0.7478

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
5	2626	947	813	2427	0.7417	0.7636	0.7193	0.7491	0.7491	0.7491	0.7491
6	2617	916	822	2458	0.7449	0.7610	0.7285	0.7494	0.7494	0.7494	0.7494
7	2644	905	795	2469	0.7505	0.7688	0.7318	0.7564	0.7564	0.7564	0.7564
8	2726	992	713	2382	0.7497	0.7927	0.7060	0.7696	0.7696	0.7696	0.7696
9	2634	934	805	2440	0.7448	0.7659	0.7232	0.7519	0.7519	0.7519	0.7519
10	2704	978	735	2396	0.7486	0.7863	0.7101	0.7653	0.7653	0.7653	0.7653
Average					0.7443	0.7681	0.7199	0.7443	0.7441	0.7366	0.7530

ตารางภาคผนวกที่ 76 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคการถดถอยลอจิสติก กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการเลือกแบบไปข้างหน้า โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้อัตราข้อมูลโรคเบาหวาน (ชุดข้อมูลเรียนรู้)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	23818	8486	7136	21880	0.7452	0.7695	0.7205	0.7452	0.7451	0.7373	0.7541
2	23818	8464	7136	21902	0.7456	0.7695	0.7213	0.7456	0.7454	0.7378	0.7543
3	23775	8521	7179	21845	0.7440	0.7681	0.7194	0.7440	0.7438	0.7362	0.7527
4	23823	8502	7131	21864	0.7451	0.7696	0.7200	0.7450	0.7449	0.7370	0.7541
5	23822	8505	7133	21861	0.7450	0.7696	0.7199	0.7450	0.7448	0.7369	0.7540
6	23790	8500	7165	21866	0.7445	0.7685	0.7201	0.7445	0.7444	0.7368	0.7532
7	23811	8537	7144	21829	0.7443	0.7692	0.7189	0.7443	0.7441	0.7361	0.7534
8	23778	8518	7177	21848	0.7441	0.7681	0.7195	0.7440	0.7439	0.7363	0.7527
9	23803	8499	7152	21867	0.7448	0.7690	0.7201	0.7448	0.7446	0.7369	0.7535
10	23781	8518	7174	21848	0.7441	0.7682	0.7195	0.7441	0.7439	0.7363	0.7528
Average					0.7447	0.7689	0.7199	0.7447	0.7445	0.7367	0.7535

ตารางภาคผนวกที่ 77 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคการถดถอยลอจิสติก กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการเลือกแบบไปข้างหน้า โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้อัตราข้อมูลโรคเบาหวาน (ชุดข้อมูลทดสอบ)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	2593	925	847	2449	0.7399	0.7538	0.7258	0.7399	0.7399	0.7371	0.7430
2	2571	935	869	2439	0.7353	0.7474	0.7229	0.7353	0.7352	0.7333	0.7373

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
3	2680	939	760	2435	0.7507	0.7791	0.7217	0.7507	0.7504	0.7405	0.7621
4	2631	968	809	2406	0.7392	0.7648	0.7131	0.7392	0.7390	0.7310	0.7484
5	2630	945	809	2429	0.7426	0.7648	0.7199	0.7426	0.7424	0.7357	0.7502
6	2611	916	828	2458	0.7440	0.7592	0.7285	0.7440	0.7439	0.7403	0.7480
7	2642	915	797	2459	0.7487	0.7682	0.7288	0.7487	0.7486	0.7428	0.7552
8	2724	991	715	2383	0.7496	0.7921	0.7063	0.7496	0.7491	0.7332	0.7692
9	2633	939	806	2435	0.7439	0.7656	0.7217	0.7439	0.7437	0.7371	0.7513
10	2704	979	735	2395	0.7484	0.7863	0.7098	0.7484	0.7480	0.7342	0.7652
Average					0.7442	0.7681	0.7199	0.7442	0.7440	0.7365	0.7530

ตารางภาคผนวกที่ 78 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคการถดถอยลอจิสติก กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีกำจัดแบบลอยหลัง โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้อุณหภูมิข้อมูลโรคเบาหวาน (ชุดข้อมูลเรียนรู้)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	23818	8486	7136	21880	0.7452	0.7695	0.7205	0.7452	0.7451	0.7373	0.7541
2	23818	8464	7136	21902	0.7456	0.7695	0.7213	0.7456	0.7454	0.7378	0.7543
3	23775	8521	7179	21845	0.7440	0.7681	0.7194	0.7440	0.7438	0.7362	0.7527
4	23823	8502	7131	21864	0.7451	0.7696	0.7200	0.7450	0.7449	0.7370	0.7541
5	23822	8505	7133	21861	0.7450	0.7696	0.7199	0.7450	0.7448	0.7369	0.7540
6	23790	8500	7165	21866	0.7445	0.7685	0.7201	0.7445	0.7444	0.7368	0.7532
7	23811	8537	7144	21829	0.7443	0.7692	0.7189	0.7443	0.7441	0.7361	0.7534
8	23778	8518	7177	21848	0.7441	0.7681	0.7195	0.7440	0.7439	0.7363	0.7527
9	23803	8499	7152	21867	0.7448	0.7690	0.7201	0.7448	0.7446	0.7369	0.7535
10	23781	8518	7174	21848	0.7441	0.7682	0.7195	0.7441	0.7439	0.7363	0.7528
Average					0.7447	0.7689	0.7199	0.7447	0.7445	0.7367	0.7535

ตารางภาคผนวกที่ 79 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคการถดถอยลอจิสติก
กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีกำจัดแบบถอยหลัง โดยการสร้างตัว
แบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้อัตราข้อมูลโรคเบาหวาน (ชุดข้อมูล
ทดสอบ)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	2593	925	847	2449	0.7399	0.7538	0.7258	0.7399	0.7399	0.7371	0.7430
2	2571	935	869	2439	0.7353	0.7474	0.7229	0.7353	0.7352	0.7333	0.7373
3	2680	939	760	2435	0.7507	0.7791	0.7217	0.7507	0.7504	0.7405	0.7621
4	2631	968	809	2406	0.7392	0.7648	0.7131	0.7392	0.7390	0.7310	0.7484
5	2630	945	809	2429	0.7426	0.7648	0.7199	0.7426	0.7424	0.7357	0.7502
6	2611	916	828	2458	0.7440	0.7592	0.7285	0.7440	0.7439	0.7403	0.7480
7	2642	915	797	2459	0.7487	0.7682	0.7288	0.7487	0.7486	0.7428	0.7552
8	2724	991	715	2383	0.7496	0.7921	0.7063	0.7496	0.7491	0.7332	0.7692
9	2633	939	806	2435	0.7439	0.7656	0.7217	0.7439	0.7437	0.7371	0.7513
10	2704	979	735	2395	0.7484	0.7863	0.7098	0.7484	0.7480	0.7342	0.7652
Average					0.7442	0.7681	0.7199	0.7442	0.7440	0.7365	0.7530

ตารางภาคผนวกที่ 80 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคการถดถอยลอจิสติก
กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการถดถอยทีละขั้น โดยการสร้างตัวแบบ
10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้อัตราข้อมูลโรคเบาหวาน (ชุดข้อมูลเรียนรู้)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	23818	8486	7136	21880	0.7452	0.7695	0.7205	0.7452	0.7451	0.7373	0.7541
2	23818	8464	7136	21902	0.7456	0.7695	0.7213	0.7456	0.7454	0.7378	0.7543
3	23775	8521	7179	21845	0.7440	0.7681	0.7194	0.7440	0.7438	0.7362	0.7527
4	23823	8502	7131	21864	0.7451	0.7696	0.7200	0.7450	0.7449	0.7370	0.7541
5	23822	8505	7133	21861	0.7450	0.7696	0.7199	0.7450	0.7448	0.7369	0.7540
6	23790	8500	7165	21866	0.7445	0.7685	0.7201	0.7445	0.7444	0.7368	0.7532
7	23811	8537	7144	21829	0.7443	0.7692	0.7189	0.7443	0.7441	0.7361	0.7534
8	23778	8518	7177	21848	0.7441	0.7681	0.7195	0.7440	0.7439	0.7363	0.7527
9	23803	8499	7152	21867	0.7448	0.7690	0.7201	0.7448	0.7446	0.7369	0.7535
10	23781	8518	7174	21848	0.7441	0.7682	0.7195	0.7441	0.7439	0.7363	0.7528
Average					0.7447	0.7689	0.7199	0.7447	0.7445	0.7367	0.7535

ตารางภาคผนวกที่ 81 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคการถดถอยลอจิสติก
กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการถดถอยที่ละชั้น โดยการสร้างตัวแบบ
10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้อชุดข้อมูลโรคเบาหวาน (ชุดข้อมูลทดสอบ)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	2593	925	847	2449	0.7399	0.7538	0.7258	0.7399	0.7399	0.7371	0.7430
2	2571	935	869	2439	0.7353	0.7474	0.7229	0.7353	0.7352	0.7333	0.7373
3	2680	939	760	2435	0.7507	0.7791	0.7217	0.7507	0.7504	0.7405	0.7621
4	2631	968	809	2406	0.7392	0.7648	0.7131	0.7392	0.7390	0.7310	0.7484
5	2630	945	809	2429	0.7426	0.7648	0.7199	0.7426	0.7424	0.7357	0.7502
6	2611	916	828	2458	0.7440	0.7592	0.7285	0.7440	0.7439	0.7403	0.7480
7	2642	915	797	2459	0.7487	0.7682	0.7288	0.7487	0.7486	0.7428	0.7552
8	2724	991	715	2383	0.7496	0.7921	0.7063	0.7496	0.7491	0.7332	0.7692
9	2633	939	806	2435	0.7439	0.7656	0.7217	0.7439	0.7437	0.7371	0.7513
10	2704	979	735	2395	0.7484	0.7863	0.7098	0.7484	0.7480	0.7342	0.7652
Average					0.7442	0.7681	0.7199	0.7442	0.7440	0.7365	0.7530

ตารางภาคผนวกที่ 82 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคการถดถอยลอจิสติก
กรณีคัดเลือกตัวแปรอิสระด้วยค่าสัมประสิทธิ์สหสัมพันธ์ ภายใต้อชุดข้อมูลโรคเบาหวาน (ชุดข้อมูล
เรียนรู้)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	23742	8507	7212	21859	0.7437	0.7670	0.7199	0.7437	0.7435	0.7362	0.7519
2	23747	8472	7207	21894	0.7443	0.7672	0.7210	0.7443	0.7442	0.7370	0.7523
3	23701	8527	7253	21839	0.7427	0.7657	0.7192	0.7427	0.7425	0.7354	0.7507
4	23748	8540	7206	21826	0.7432	0.7672	0.7188	0.7432	0.7430	0.7355	0.7518
5	23747	8531	7208	21835	0.7433	0.7671	0.7191	0.7433	0.7432	0.7357	0.7518
6	23708	8481	7247	21885	0.7435	0.7659	0.7207	0.7435	0.7434	0.7365	0.7512
7	23715	8563	7240	21803	0.7423	0.7661	0.7180	0.7423	0.7421	0.7347	0.7507
8	23724	8527	7231	21839	0.7430	0.7664	0.7192	0.7430	0.7429	0.7356	0.7513
9	23753	8514	7202	21852	0.7437	0.7673	0.7196	0.7437	0.7435	0.7361	0.7521
10	23716	8565	7239	21801	0.7423	0.7661	0.7179	0.7422	0.7421	0.7347	0.7507
Average					0.7432	0.7666	0.7193	0.7432	0.7430	0.7358	0.7515

ตารางภาคผนวกที่ 83 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคการถดถอยลอจิสติก
กรณีคัดเลือกตัวแปรอิสระด้วยค่าสัมประสิทธิ์สหสัมพันธ์ ภายใต้ชุดข้อมูลโรคเบาหวาน (ชุดข้อมูล
ทดสอบ)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	2577	914	863	2460	0.7392	0.7491	0.7291	0.7403	0.7403	0.7403	0.7403
2	2562	943	878	2431	0.7328	0.7448	0.7205	0.7347	0.7347	0.7347	0.7347
3	2677	949	763	2425	0.7488	0.7782	0.7187	0.7607	0.7607	0.7607	0.7607
4	2630	963	810	2411	0.7398	0.7645	0.7146	0.7485	0.7485	0.7485	0.7485
5	2629	947	810	2427	0.7421	0.7645	0.7193	0.7498	0.7498	0.7498	0.7498
6	2615	935	824	2439	0.7418	0.7604	0.7229	0.7475	0.7475	0.7475	0.7475
7	2648	917	791	2457	0.7493	0.7700	0.7282	0.7565	0.7565	0.7565	0.7565
8	2702	1013	737	2361	0.7431	0.7857	0.6998	0.7621	0.7621	0.7621	0.7621
9	2619	937	820	2437	0.7421	0.7616	0.7223	0.7482	0.7482	0.7482	0.7482
10	2685	963	754	2411	0.7480	0.7808	0.7146	0.7618	0.7618	0.7618	0.7618
Average					0.7427	0.7660	0.7190	0.7427	0.7425	0.7354	0.7510

ตารางภาคผนวกที่ 84 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคต้นไม้ตัดสินใจ โดยใช้
ตัวแปรอิสระทั้งหมด ภายใต้ชุดข้อมูลโรคเบาหวาน (ชุดข้อมูลเรียนรู้)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	26863	5588	4091	24778	0.8422	0.8678	0.8160	0.8421	0.8420	0.8278	0.8583
2	26827	5811	4127	24555	0.8379	0.8667	0.8086	0.8379	0.8378	0.8220	0.8561
3	26632	5726	4322	24640	0.8361	0.8604	0.8114	0.8361	0.8360	0.8230	0.8508
4	26744	5786	4210	24580	0.8370	0.8640	0.8095	0.8370	0.8369	0.8221	0.8538
5	26684	5983	4271	24383	0.8328	0.8620	0.8030	0.8328	0.8326	0.8168	0.8509
6	26780	5681	4175	24685	0.8393	0.8651	0.8129	0.8393	0.8391	0.8250	0.8553
7	26624	5576	4331	24790	0.8384	0.8601	0.8164	0.8385	0.8384	0.8268	0.8513
8	26619	5639	4336	24727	0.8373	0.8599	0.8143	0.8373	0.8372	0.8252	0.8508
9	26670	5585	4285	24781	0.8390	0.8616	0.8161	0.8391	0.8390	0.8268	0.8526
10	26640	5974	4315	24392	0.8322	0.8606	0.8033	0.8322	0.8321	0.8168	0.8497
Average					0.8372	0.8628	0.8112	0.8372	0.8371	0.8232	0.8530

ตารางภาคผนวกที่ 85 เมตริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคต้นไม้ตัดสินใจ โดยใช้ตัวแปรอิสระทั้งหมด ภายใต้ชุดข้อมูลโรคเบาหวาน (ชุดข้อมูลทดสอบ)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	2498	1030	942	2344	0.7106	0.7262	0.6947	0.7106	0.7105	0.7080	0.7133
2	2508	1031	932	2343	0.7119	0.7291	0.6944	0.7119	0.7118	0.7087	0.7154
3	2569	1003	871	2371	0.7250	0.7468	0.7027	0.7250	0.7248	0.7192	0.7313
4	2562	1073	878	2301	0.7137	0.7448	0.6820	0.7137	0.7134	0.7048	0.7238
5	2583	1028	856	2346	0.7235	0.7511	0.6953	0.7235	0.7232	0.7153	0.7327
6	2522	1021	917	2353	0.7155	0.7334	0.6974	0.7156	0.7155	0.7118	0.7196
7	2587	1005	852	2369	0.7274	0.7523	0.7021	0.7274	0.7272	0.7202	0.7355
8	2595	1092	844	2282	0.7158	0.7546	0.6763	0.7158	0.7154	0.7038	0.7300
9	2524	1066	915	2308	0.7092	0.7339	0.6841	0.7092	0.7090	0.7031	0.7161
10	2606	1056	833	2318	0.7227	0.7578	0.6870	0.7227	0.7224	0.7116	0.7356
Average					0.7175	0.7430	0.6916	0.7175	0.7173	0.7107	0.7253

ตารางภาคผนวกที่ 86 เมตริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคต้นไม้ตัดสินใจ กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการเลือกแบบไปข้างหน้า โดยใช้ข้อมูลทั้งหมด ภายใต้ชุดข้อมูลโรคเบาหวาน (ชุดข้อมูลเรียนรู้)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	26363	6767	4591	23599	0.8148	0.8517	0.7772	0.8148	0.8145	0.7957	0.8371
2	26195	6894	4759	23472	0.8100	0.8463	0.7730	0.8100	0.8097	0.7917	0.8314
3	25948	6273	5006	24093	0.8161	0.8383	0.7934	0.8161	0.8159	0.8053	0.8280
4	26401	6894	4553	23472	0.8133	0.8529	0.7730	0.8133	0.8130	0.7929	0.8375
5	26112	6780	4843	23586	0.8105	0.8435	0.7767	0.8104	0.8102	0.7939	0.8296
6	26196	6760	4759	23606	0.8122	0.8463	0.7774	0.8122	0.8119	0.7949	0.8322
7	25976	6347	4979	24019	0.8153	0.8392	0.7910	0.8153	0.8152	0.8036	0.8283
8	26210	7068	4745	23298	0.8074	0.8467	0.7672	0.8073	0.8070	0.7876	0.8308
9	26351	6778	4604	23588	0.8144	0.8513	0.7768	0.8144	0.8141	0.7954	0.8367
10	26073	6857	4882	23509	0.8086	0.8423	0.7742	0.8086	0.8083	0.7918	0.8280
Average					0.8123	0.8459	0.7780	0.8122	0.8120	0.7953	0.8320

ตารางภาคผนวกที่ 87 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคต้นไม้ตัดสินใจ กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการเลือกแบบไปข้างหน้า โดยใช้ข้อมูลทั้งหมด ภายใต้ชุดข้อมูลโรคเบาหวาน (ชุดข้อมูลทดสอบ)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	2593	1042	847	2332	0.7228	0.7538	0.6912	0.7228	0.7225	0.7133	0.7336
2	2526	1046	914	2328	0.7124	0.7343	0.6900	0.7124	0.7122	0.7072	0.7181
3	2599	983	841	2391	0.7323	0.7555	0.7087	0.7323	0.7322	0.7256	0.7398
4	2621	1105	819	2269	0.7176	0.7619	0.6725	0.7176	0.7170	0.7034	0.7348
5	2647	1047	792	2327	0.7301	0.7697	0.6897	0.7301	0.7296	0.7166	0.7461
6	2582	1037	857	2337	0.7220	0.7508	0.6926	0.7220	0.7217	0.7135	0.7317
7	2618	1009	821	2365	0.7314	0.7613	0.7009	0.7314	0.7311	0.7218	0.7423
8	2708	1095	731	2279	0.7320	0.7874	0.6755	0.7320	0.7311	0.7121	0.7571
9	2610	1061	829	2313	0.7226	0.7589	0.6855	0.7226	0.7222	0.7110	0.7362
10	2607	1070	832	2304	0.7208	0.7581	0.6829	0.7209	0.7204	0.7090	0.7347
Average					0.7244	0.7592	0.6890	0.7244	0.7240	0.7133	0.7374

ตารางภาคผนวกที่ 88 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคต้นไม้ตัดสินใจ กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการกำจัดแบบถอยหลัง โดยใช้ข้อมูลทั้งหมด ภายใต้ชุดข้อมูลโรคเบาหวาน (ชุดข้อมูลเรียนรู้)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	26363	6767	4591	23599	0.8148	0.8517	0.7772	0.8148	0.8145	0.7957	0.8371
2	26195	6894	4759	23472	0.8100	0.8463	0.7730	0.8100	0.8097	0.7917	0.8314
3	25948	6273	5006	24093	0.8161	0.8383	0.7934	0.8161	0.8159	0.8053	0.8280
4	26401	6894	4553	23472	0.8133	0.8529	0.7730	0.8133	0.8130	0.7929	0.8375
5	26112	6780	4843	23586	0.8105	0.8435	0.7767	0.8104	0.8102	0.7939	0.8296
6	26196	6760	4759	23606	0.8122	0.8463	0.7774	0.8122	0.8119	0.7949	0.8322
7	25976	6347	4979	24019	0.8153	0.8392	0.7910	0.8153	0.8152	0.8036	0.8283
8	26210	7068	4745	23298	0.8074	0.8467	0.7672	0.8073	0.8070	0.7876	0.8308
9	26351	6778	4604	23588	0.8144	0.8513	0.7768	0.8144	0.8141	0.7954	0.8367
10	26073	6857	4882	23509	0.8086	0.8423	0.7742	0.8086	0.8083	0.7918	0.8280
Average					0.8123	0.8459	0.7780	0.8122	0.8120	0.7953	0.8320

ตารางภาคผนวกที่ 89 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคต้นไม้ตัดสินใจ กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีกำจัดแบบถอยหลัง โดยใช้ข้อมูลทั้งหมด ภายใต้ชุดข้อมูลโรคเบาหวาน (ชุดข้อมูลทดสอบ)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	2593	1042	847	2332	0.7228	0.7538	0.6912	0.7228	0.7225	0.7133	0.7336
2	2526	1046	914	2328	0.7124	0.7343	0.6900	0.7124	0.7122	0.7072	0.7181
3	2599	983	841	2391	0.7323	0.7555	0.7087	0.7323	0.7322	0.7256	0.7398
4	2621	1105	819	2269	0.7176	0.7619	0.6725	0.7176	0.7170	0.7034	0.7348
5	2647	1047	792	2327	0.7301	0.7697	0.6897	0.7301	0.7296	0.7166	0.7461
6	2582	1037	857	2337	0.7220	0.7508	0.6926	0.7220	0.7217	0.7135	0.7317
7	2618	1009	821	2365	0.7314	0.7613	0.7009	0.7314	0.7311	0.7218	0.7423
8	2708	1095	731	2279	0.7320	0.7874	0.6755	0.7320	0.7311	0.7121	0.7571
9	2610	1061	829	2313	0.7226	0.7589	0.6855	0.7226	0.7222	0.7110	0.7362
10	2607	1070	832	2304	0.7208	0.7581	0.6829	0.7209	0.7204	0.7090	0.7347
Average					0.7244	0.7592	0.6890	0.7244	0.7240	0.7133	0.7374

ตารางภาคผนวกที่ 90 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคต้นไม้ตัดสินใจ กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการถดถอยทีละชั้น โดยใช้ข้อมูลทั้งหมด ภายใต้ชุดข้อมูลโรคเบาหวาน (ชุดข้อมูลเรียนรู้)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	26363	6767	4591	23599	0.8148	0.8517	0.7772	0.8148	0.8145	0.7957	0.8371
2	26195	6894	4759	23472	0.8100	0.8463	0.7730	0.8100	0.8097	0.7917	0.8314
3	25948	6273	5006	24093	0.8161	0.8383	0.7934	0.8161	0.8159	0.8053	0.8280
4	26401	6894	4553	23472	0.8133	0.8529	0.7730	0.8133	0.8130	0.7929	0.8375
5	26112	6780	4843	23586	0.8105	0.8435	0.7767	0.8104	0.8102	0.7939	0.8296
6	26196	6760	4759	23606	0.8122	0.8463	0.7774	0.8122	0.8119	0.7949	0.8322
7	25976	6347	4979	24019	0.8153	0.8392	0.7910	0.8153	0.8152	0.8036	0.8283
8	26210	7068	4745	23298	0.8074	0.8467	0.7672	0.8073	0.8070	0.7876	0.8308
9	26351	6778	4604	23588	0.8144	0.8513	0.7768	0.8144	0.8141	0.7954	0.8367
10	26073	6857	4882	23509	0.8086	0.8423	0.7742	0.8086	0.8083	0.7918	0.8280
Average					0.8123	0.8459	0.7780	0.8122	0.8120	0.7953	0.8320

ตารางภาคผนวกที่ 91 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคต้นไม้ตัดสินใจ กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการถดถอยทีละชั้น โดยใช้ข้อมูลทั้งหมด ภายใต้ชุดข้อมูลโรคเบาหวาน (ชุดข้อมูลทดสอบ)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	2593	1042	847	2332	0.7228	0.7538	0.6912	0.7228	0.7225	0.7133	0.7336
2	2526	1046	914	2328	0.7124	0.7343	0.6900	0.7124	0.7122	0.7072	0.7181
3	2599	983	841	2391	0.7323	0.7555	0.7087	0.7323	0.7322	0.7256	0.7398
4	2621	1105	819	2269	0.7176	0.7619	0.6725	0.7176	0.7170	0.7034	0.7348
5	2647	1047	792	2327	0.7301	0.7697	0.6897	0.7301	0.7296	0.7166	0.7461
6	2582	1037	857	2337	0.7220	0.7508	0.6926	0.7220	0.7217	0.7135	0.7317
7	2618	1009	821	2365	0.7314	0.7613	0.7009	0.7314	0.7311	0.7218	0.7423
8	2708	1095	731	2279	0.7320	0.7874	0.6755	0.7320	0.7311	0.7121	0.7571
9	2610	1061	829	2313	0.7226	0.7589	0.6855	0.7226	0.7222	0.7110	0.7362
10	2607	1070	832	2304	0.7208	0.7581	0.6829	0.7209	0.7204	0.7090	0.7347
Average					0.7244	0.7592	0.6890	0.7244	0.7240	0.7133	0.7374

ตารางภาคผนวกที่ 92 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคต้นไม้ตัดสินใจ กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการเลือกแบบไปข้างหน้า โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้ชุดข้อมูลโรคเบาหวาน (ชุดข้อมูลเรียนรู้)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	25902	7451	5052	22915	0.7961	0.8368	0.7546	0.7961	0.7957	0.7766	0.8194
2	25750	7197	5204	23169	0.7978	0.8319	0.7630	0.7978	0.7975	0.7816	0.8166
3	25554	7290	5400	23076	0.7931	0.8255	0.7599	0.7930	0.7928	0.7780	0.8104
4	26107	7878	4847	22488	0.7925	0.8434	0.7406	0.7925	0.7919	0.7682	0.8227
5	25475	7430	5480	22936	0.7895	0.8230	0.7553	0.7895	0.7892	0.7742	0.8072
6	26013	7798	4942	22568	0.7922	0.8403	0.7432	0.7922	0.7917	0.7694	0.8204
7	25789	7846	5166	22520	0.7878	0.8331	0.7416	0.7878	0.7873	0.7667	0.8134
8	25963	8053	4992	22313	0.7873	0.8387	0.7348	0.7872	0.7866	0.7633	0.8172
9	25954	7887	5001	22479	0.7898	0.8384	0.7403	0.7898	0.7893	0.7669	0.8180
10	25620	7644	5335	22722	0.7883	0.8277	0.7483	0.7884	0.7880	0.7702	0.8099
Average					0.7914	0.8339	0.7482	0.7914	0.7910	0.7715	0.8155

ตารางภาคผนวกที่ 93 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคต้นไม้ตัดสินใจ กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการเลือกแบบไปข้างหน้า โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้อัตราข้อมูลโรคเบาหวาน (ชุดข้อมูลทดสอบ)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	2614	1017	826	2357	0.7295	0.7599	0.6986	0.7405	0.7405	0.7405	0.7405
2	2553	1025	887	2349	0.7194	0.7422	0.6962	0.7259	0.7259	0.7259	0.7259
3	2656	1001	784	2373	0.7380	0.7721	0.7033	0.7517	0.7517	0.7517	0.7517
4	2703	1114	737	2260	0.7284	0.7858	0.6698	0.7541	0.7541	0.7541	0.7541
5	2667	1028	772	2346	0.7358	0.7755	0.6953	0.7524	0.7524	0.7524	0.7524
6	2645	1060	794	2314	0.7279	0.7691	0.6858	0.7445	0.7445	0.7445	0.7445
7	2700	1040	739	2334	0.7389	0.7851	0.6918	0.7595	0.7595	0.7595	0.7595
8	2769	1158	670	2216	0.7317	0.8052	0.6568	0.7678	0.7678	0.7678	0.7678
9	2680	1075	759	2299	0.7308	0.7793	0.6814	0.7518	0.7518	0.7518	0.7518
10	2673	1054	766	2320	0.7329	0.7773	0.6876	0.7518	0.7518	0.7518	0.7518
Average					0.7313	0.7752	0.6867	0.7313	0.7307	0.7162	0.7500

ตารางภาคผนวกที่ 94 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคต้นไม้ตัดสินใจ กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการจัดแบบถอยหลัง โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้อัตราข้อมูลโรคเบาหวาน (ชุดข้อมูลเรียนรู้)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	25902	7451	5052	22915	0.7961	0.8368	0.7546	0.7961	0.7957	0.7766	0.8194
2	25750	7197	5204	23169	0.7978	0.8319	0.7630	0.7978	0.7975	0.7816	0.8166
3	25554	7290	5400	23076	0.7931	0.8255	0.7599	0.7930	0.7928	0.7780	0.8104
4	26107	7878	4847	22488	0.7925	0.8434	0.7406	0.7925	0.7919	0.7682	0.8227
5	25475	7430	5480	22936	0.7895	0.8230	0.7553	0.7895	0.7892	0.7742	0.8072
6	26013	7798	4942	22568	0.7922	0.8403	0.7432	0.7922	0.7917	0.7694	0.8204
7	25789	7846	5166	22520	0.7878	0.8331	0.7416	0.7878	0.7873	0.7667	0.8134
8	25963	8053	4992	22313	0.7873	0.8387	0.7348	0.7872	0.7866	0.7633	0.8172
9	25954	7887	5001	22479	0.7898	0.8384	0.7403	0.7898	0.7893	0.7669	0.8180
10	25620	7644	5335	22722	0.7883	0.8277	0.7483	0.7884	0.7880	0.7702	0.8099
Average					0.7914	0.8339	0.7482	0.7914	0.7910	0.7715	0.8155

ตารางภาคผนวกที่ 95 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคต้นไม้ตัดสินใจ กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีกำจัดแบบถอยหลัง โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้อข้อมูลโรคเบาหวาน (ชุดข้อมูลทดสอบ)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	2614	1017	826	2357	0.7295	0.7599	0.6986	0.7405	0.7405	0.7405	0.7405
2	2553	1025	887	2349	0.7194	0.7422	0.6962	0.7259	0.7259	0.7259	0.7259
3	2656	1001	784	2373	0.7380	0.7721	0.7033	0.7517	0.7517	0.7517	0.7517
4	2703	1114	737	2260	0.7284	0.7858	0.6698	0.7541	0.7541	0.7541	0.7541
5	2667	1028	772	2346	0.7358	0.7755	0.6953	0.7524	0.7524	0.7524	0.7524
6	2645	1060	794	2314	0.7279	0.7691	0.6858	0.7445	0.7445	0.7445	0.7445
7	2700	1040	739	2334	0.7389	0.7851	0.6918	0.7595	0.7595	0.7595	0.7595
8	2769	1158	670	2216	0.7317	0.8052	0.6568	0.7678	0.7678	0.7678	0.7678
9	2680	1075	759	2299	0.7308	0.7793	0.6814	0.7518	0.7518	0.7518	0.7518
10	2673	1054	766	2320	0.7329	0.7773	0.6876	0.7518	0.7518	0.7518	0.7518
Average					0.7313	0.7752	0.6867	0.7313	0.7307	0.7162	0.7500

ตารางภาคผนวกที่ 96 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคต้นไม้ตัดสินใจ กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการถดถอยทีละชั้น โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้อข้อมูลโรคเบาหวาน (ชุดข้อมูลเรียนรู้)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	25902	7451	5052	22915	0.7961	0.8368	0.7546	0.7961	0.7957	0.7766	0.8194
2	25750	7197	5204	23169	0.7978	0.8319	0.7630	0.7978	0.7975	0.7816	0.8166
3	25554	7290	5400	23076	0.7931	0.8255	0.7599	0.7930	0.7928	0.7780	0.8104
4	26107	7878	4847	22488	0.7925	0.8434	0.7406	0.7925	0.7919	0.7682	0.8227
5	25475	7430	5480	22936	0.7895	0.8230	0.7553	0.7895	0.7892	0.7742	0.8072
6	26013	7798	4942	22568	0.7922	0.8403	0.7432	0.7922	0.7917	0.7694	0.8204
7	25789	7846	5166	22520	0.7878	0.8331	0.7416	0.7878	0.7873	0.7667	0.8134
8	25963	8053	4992	22313	0.7873	0.8387	0.7348	0.7872	0.7866	0.7633	0.8172
9	25954	7887	5001	22479	0.7898	0.8384	0.7403	0.7898	0.7893	0.7669	0.8180
10	25620	7644	5335	22722	0.7883	0.8277	0.7483	0.7884	0.7880	0.7702	0.8099
Average					0.7914	0.8339	0.7482	0.7914	0.7910	0.7715	0.8155

ตารางภาคผนวกที่ 97 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคต้นไม้ตัดสินใจ กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการถดถอยทีละชั้น โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้อัตราข้อมูลโรคเบาหวาน (ชุดข้อมูลทดสอบ)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	2614	1017	826	2357	0.7295	0.7599	0.6986	0.7405	0.7405	0.7405	0.7405
2	2553	1025	887	2349	0.7194	0.7422	0.6962	0.7259	0.7259	0.7259	0.7259
3	2656	1001	784	2373	0.7380	0.7721	0.7033	0.7517	0.7517	0.7517	0.7517
4	2703	1114	737	2260	0.7284	0.7858	0.6698	0.7541	0.7541	0.7541	0.7541
5	2667	1028	772	2346	0.7358	0.7755	0.6953	0.7524	0.7524	0.7524	0.7524
6	2645	1060	794	2314	0.7279	0.7691	0.6858	0.7445	0.7445	0.7445	0.7445
7	2700	1040	739	2334	0.7389	0.7851	0.6918	0.7595	0.7595	0.7595	0.7595
8	2769	1158	670	2216	0.7317	0.8052	0.6568	0.7678	0.7678	0.7678	0.7678
9	2680	1075	759	2299	0.7308	0.7793	0.6814	0.7518	0.7518	0.7518	0.7518
10	2673	1054	766	2320	0.7329	0.7773	0.6876	0.7518	0.7518	0.7518	0.7518
Average					0.7313	0.7752	0.6867	0.7313	0.7307	0.7162	0.7500

ตารางภาคผนวกที่ 98 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคต้นไม้ตัดสินใจ กรณีคัดเลือกตัวแปรอิสระด้วยค่าสัมประสิทธิ์สหสัมพันธ์ ภายใต้อัตราข้อมูลโรคเบาหวาน (ชุดข้อมูลเรียนรู้)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	26340	6611	4614	23755	0.8169	0.8509	0.7823	0.8169	0.8167	0.7994	0.8374
2	26463	7049	4491	23317	0.8118	0.8549	0.7679	0.8118	0.8114	0.7897	0.8385
3	26179	6497	4775	23869	0.8162	0.8457	0.7860	0.8161	0.8160	0.8012	0.8333
4	26241	6498	4713	23868	0.8172	0.8477	0.7860	0.8171	0.8170	0.8015	0.8351
5	26296	6882	4659	23484	0.8118	0.8495	0.7734	0.8118	0.8115	0.7926	0.8345
6	26151	6904	4804	23462	0.8091	0.8448	0.7726	0.8090	0.8088	0.7911	0.8300
7	26353	6739	4602	23627	0.8151	0.8513	0.7781	0.8151	0.8148	0.7964	0.8370
8	26497	6954	4458	23412	0.8139	0.8560	0.7710	0.8139	0.8135	0.7921	0.8400
9	26273	6187	4682	24179	0.8228	0.8487	0.7963	0.8228	0.8226	0.8094	0.8378
10	26021	6324	4934	24042	0.8164	0.8406	0.7917	0.8164	0.8163	0.8045	0.8297
Average					0.8151	0.8490	0.7805	0.8151	0.8148	0.7978	0.8353

ตารางภาคผนวกที่ 99 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคต้นไม้ตัดสินใจ กรณีคัดเลือกตัวแปรอิสระด้วยค่าสัมประสิทธิ์สหสัมพันธ์ ภายใต้ชุดข้อมูลโรคเบาหวาน (ชุดข้อมูลทดสอบ)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	2571	1025	869	2349	0.7220	0.7474	0.6962	0.7220	0.7218	0.7150	0.7300
2	2581	1064	859	2310	0.7178	0.7503	0.6846	0.7178	0.7174	0.7081	0.7289
3	2595	1019	845	2355	0.7264	0.7544	0.6980	0.7265	0.7262	0.7180	0.7359
4	2581	1050	859	2324	0.7198	0.7503	0.6888	0.7198	0.7195	0.7108	0.7301
5	2653	1047	786	2327	0.7310	0.7714	0.6897	0.7309	0.7305	0.7170	0.7475
6	2583	1030	856	2344	0.7232	0.7511	0.6947	0.7232	0.7229	0.7149	0.7325
7	2609	1034	830	2340	0.7264	0.7587	0.6935	0.7264	0.7261	0.7162	0.7382
8	2682	1130	757	2244	0.7230	0.7799	0.6651	0.7231	0.7221	0.7036	0.7478
9	2567	1060	872	2314	0.7164	0.7464	0.6858	0.7164	0.7161	0.7077	0.7263
10	2582	1024	857	2350	0.7239	0.7508	0.6965	0.7239	0.7237	0.7160	0.7328
Average					0.7230	0.7561	0.6893	0.7230	0.7226	0.7127	0.7350

ตารางภาคผนวกที่ 100 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคป่าสุ่ม โดยใช้ตัวแปรอิสระทั้งหมด ภายใต้ชุดข้อมูลโรคเบาหวาน (ชุดข้อมูลเรียนรู้)

(mtry, ntree)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(4, 250)	0.9431	0.9438	0.9423	0.9431	0.9431	0.9437	0.9429
(4, 500)	0.9434	0.9438	0.9430	0.9434	0.9434	0.9443	0.9428
(5, 250)	0.9726	0.9675	0.9777	0.9725	0.9725	0.9780	0.9673
(5, 500)	0.9733	0.9681	0.9785	0.9733	0.9733	0.9788	0.9679

ตารางภาคผนวกที่ 101 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคป่าสุ่ม โดยใช้ตัวแปรอิสระทั้งหมด ภายใต้ชุดข้อมูลโรคเบาหวาน (ชุดข้อมูลทดสอบ)

(mtry, ntree)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(4, 250)	0.7401	0.7885	0.6908	0.7401	0.7393	0.7224	0.7627
(4, 500)	0.7401	0.7893	0.6900	0.7401	0.7393	0.7221	0.7632
(5, 250)	0.7355	0.7812	0.6889	0.7355	0.7348	0.7193	0.7560
(5, 500)	0.7360	0.7822	0.6889	0.7360	0.7353	0.7196	0.7568

ตารางภาคผนวกที่ 102 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคป่าสุ่ม กรณีคัดเลือกตัวแปรอิสระ ด้วยเทคนิคการถดถอยลอจิสติกวิธีการเลือกแบบไปข้างหน้า โดยใช้ข้อมูลทั้งหมด ภายใต้ชุดข้อมูล โรคเบาหวาน (ชุดข้อมูลเรียนรู้)

(mtry, ntree)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(4, 250)	0.9245	0.9291	0.9198	0.9245	0.9245	0.9222	0.9273
(4, 500)	0.9251	0.9294	0.9206	0.9250	0.9250	0.9229	0.9277
(5, 250)	0.9577	0.9547	0.9607	0.9577	0.9576	0.9613	0.9542
(5, 500)	0.9583	0.9551	0.9616	0.9583	0.9583	0.9622	0.9547

ตารางภาคผนวกที่ 103 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคป่าสุ่ม กรณีคัดเลือกตัวแปรอิสระ ด้วยเทคนิคการถดถอยลอจิสติกวิธีการเลือกแบบไปข้างหน้า โดยใช้ข้อมูลทั้งหมด ภายใต้ชุดข้อมูล โรคเบาหวาน (ชุดข้อมูลทดสอบ)

(mtry, ntree)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(4, 250)	0.7373	0.7847	0.6890	0.7373	0.7366	0.7591	0.7203
(4, 500)	0.7377	0.7848	0.6897	0.7377	0.7369	0.7207	0.7593
(5, 250)	0.7315	0.7748	0.6873	0.7315	0.7308	0.7166	0.7502
(5, 500)	0.7316	0.7745	0.6879	0.7316	0.7310	0.7169	0.7501

ตารางภาคผนวกที่ 104 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคป่าสุ่ม กรณีคัดเลือกตัวแปรอิสระ ด้วยเทคนิคการถดถอยลอจิสติกวิธีการกำจัดแบบถอยหลัง โดยใช้ข้อมูลทั้งหมด ภายใต้ชุดข้อมูล โรคเบาหวาน (ชุดข้อมูลเรียนรู้)

(mtry, ntree)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(4, 250)	0.9245	0.9291	0.9198	0.9245	0.9245	0.9222	0.9273
(4, 500)	0.9251	0.9294	0.9206	0.9250	0.9250	0.9229	0.9277
(5, 250)	0.9577	0.9547	0.9607	0.9577	0.9576	0.9613	0.9542
(5, 500)	0.9583	0.9551	0.9616	0.9583	0.9583	0.9622	0.9547

ตารางภาคผนวกที่ 105 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคป่าสุ่ม กรณีคัดเลือกตัวแปรอิสระ ด้วยเทคนิคการถดถอยลอจิสติกวิธีการกำจัดแบบถอยหลัง โดยใช้ข้อมูลทั้งหมด ภายใต้ชุดข้อมูล โรคเบาหวาน (ชุดข้อมูลทดสอบ)

(mtry, ntree)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(4, 250)	0.7373	0.7847	0.6890	0.7373	0.7366	0.7591	0.7203
(4, 500)	0.7377	0.7848	0.6897	0.7377	0.7369	0.7207	0.7593
(5, 250)	0.7315	0.7748	0.6873	0.7315	0.7308	0.7166	0.7502

(mtry, ntree)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(5, 500)	0.7316	0.7745	0.6879	0.7316	0.7310	0.7169	0.7501

ตารางภาคผนวกที่ 106 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคป่าสุ่ม กรณีคัดเลือกตัวแปรอิสระ ด้วยเทคนิคการถดถอยลอจิสติกวิธีการถดถอยทีละชั้น โดยใช้ข้อมูลทั้งหมด ภายใต้อุดข้อมูลโรคเบาหวาน (ชุดข้อมูลเรียนรู้)

(mtry, ntree)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(4, 250)	0.9245	0.9291	0.9198	0.9245	0.9245	0.9222	0.9273
(4, 500)	0.9251	0.9294	0.9206	0.9250	0.9250	0.9229	0.9277
(5, 250)	0.9577	0.9547	0.9607	0.9577	0.9576	0.9613	0.9542
(5, 500)	0.9583	0.9551	0.9616	0.9583	0.9583	0.9622	0.9547

ตารางภาคผนวกที่ 107 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคป่าสุ่ม กรณีคัดเลือกตัวแปรอิสระ ด้วยเทคนิคการถดถอยลอจิสติกวิธีการถดถอยทีละชั้น โดยใช้ข้อมูลทั้งหมด ภายใต้อุดข้อมูลโรคเบาหวาน (ชุดข้อมูลทดสอบ)

(mtry, ntree)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(4, 250)	0.7373	0.7847	0.6890	0.7373	0.7366	0.7591	0.7203
(4, 500)	0.7377	0.7848	0.6897	0.7377	0.7369	0.7207	0.7593
(5, 250)	0.7315	0.7748	0.6873	0.7315	0.7308	0.7166	0.7502
(5, 500)	0.7316	0.7745	0.6879	0.7316	0.7310	0.7169	0.7501

ตารางภาคผนวกที่ 108 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคป่าสุ่ม กรณีคัดเลือกตัวแปรอิสระ ด้วยเทคนิคการถดถอยลอจิสติกวิธีการเลือกแบบไปข้างหน้า โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้อุดข้อมูลโรคเบาหวาน (ชุดข้อมูลเรียนรู้)

(mtry, ntree)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(3, 250)	0.8302	0.8684	0.7913	0.8302	0.8299	0.8093	0.8551
(3, 500)	0.8305	0.8680	0.7922	0.8304	0.8301	0.8099	0.8549
(4, 250)	0.8914	0.9078	0.8747	0.8914	0.8913	0.8810	0.9031
(4, 500)	0.8917	0.9080	0.8751	0.8917	0.8917	0.8814	0.9033

ตารางภาคผนวกที่ 109 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคป่าสุ่ม กรณีคัดเลือกตัวแปรอิสระ ด้วยเทคนิคการถดถอยลอจิสติกวิธีการเลือกแบบไปข้างหน้า โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้อัตราข้อมูลโรคเบาหวาน (ชุดข้อมูลทดสอบ)

(mtry, ntree)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(3, 250)	0.7439	0.7946	0.6923	0.7439	0.7431	0.7248	0.7682
(3, 500)	0.7442	0.7941	0.6934	0.7442	0.7435	0.7254	0.7681
(4, 250)	0.7382	0.7844	0.6911	0.7382	0.7375	0.7215	0.7592
(4, 500)	0.7386	0.7855	0.6907	0.7386	0.7378	0.7215	0.7601

ตารางภาคผนวกที่ 110 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคป่าสุ่ม กรณีคัดเลือกตัวแปรอิสระ ด้วยเทคนิคการถดถอยลอจิสติกวิธีการกำจัดแบบถดถอยหลัง โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้อัตราข้อมูลโรคเบาหวาน (ชุดข้อมูลเรียนรู้)

(mtry, ntree)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(3, 250)	0.8302	0.8684	0.7913	0.8302	0.8299	0.8093	0.8551
(3, 500)	0.8305	0.8680	0.7922	0.8304	0.8301	0.8099	0.8549
(4, 250)	0.8914	0.9078	0.8747	0.8914	0.8913	0.8810	0.9031
(4, 500)	0.8917	0.9080	0.8751	0.8917	0.8917	0.8814	0.9033

ตารางภาคผนวกที่ 111 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคป่าสุ่ม กรณีคัดเลือกตัวแปรอิสระ ด้วยเทคนิคการถดถอยลอจิสติกวิธีการกำจัดแบบถดถอยหลัง โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้อัตราข้อมูลโรคเบาหวาน (ชุดข้อมูลทดสอบ)

(mtry, ntree)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(3, 250)	0.7439	0.7946	0.6923	0.7439	0.7431	0.7248	0.7682
(3, 500)	0.7442	0.7941	0.6934	0.7442	0.7435	0.7254	0.7681
(4, 250)	0.7382	0.7844	0.6911	0.7382	0.7375	0.7215	0.7592
(4, 500)	0.7386	0.7855	0.6907	0.7386	0.7378	0.7215	0.7601

ตารางภาคผนวกที่ 112 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคป่าสุ่ม กรณีคัดเลือกตัวแปรอิสระ ด้วยเทคนิคการถดถอยลอจิสติกวิธีการถดถอยทีละขั้น โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้อัตราข้อมูลโรคเบาหวาน (ชุดข้อมูลเรียนรู้)

(mtry, ntree)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(3, 250)	0.8302	0.8684	0.7913	0.8302	0.8299	0.8093	0.8551
(3, 500)	0.8305	0.8680	0.7922	0.8304	0.8301	0.8099	0.8549
(4, 250)	0.8914	0.9078	0.8747	0.8914	0.8913	0.8810	0.9031

(mtry, ntree)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(4, 500)	0.8917	0.9080	0.8751	0.8917	0.8917	0.8814	0.9033

ตารางภาคผนวกที่ 113 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคป่าสุ่ม กรณีคัดเลือกตัวแปรอิสระ ด้วยเทคนิคการถดถอยลอจิสติกวิธีการถดถอยที่ละชั้น โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้อุดข้อมูลโรคเบาหวาน (ชุดข้อมูลทดสอบ)

(mtry, ntree)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(3, 250)	0.7439	0.7946	0.6923	0.7439	0.7431	0.7248	0.7682
(3, 500)	0.7442	0.7941	0.6934	0.7442	0.7435	0.7254	0.7681
(4, 250)	0.7382	0.7844	0.6911	0.7382	0.7375	0.7215	0.7592
(4, 500)	0.7386	0.7855	0.6907	0.7386	0.7378	0.7215	0.7601

ตารางภาคผนวกที่ 114 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคป่าสุ่ม กรณีคัดเลือกตัวแปรอิสระ ด้วยค่าสัมประสิทธิ์สหสัมพันธ์ ภายใต้อุดข้อมูลโรคเบาหวาน (ชุดข้อมูลเรียนรู้)

(mtry, ntree)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(4, 250)	0.9267	0.9305	0.9229	0.9267	0.9267	0.9250	0.9288
(4, 500)	0.9273	0.9306	0.9239	0.9273	0.9273	0.9260	0.9290
(5, 250)	0.9591	0.9550	0.9604	0.9577	0.9584	0.9639	0.9547
(5, 500)	0.9595	0.9550	0.9640	0.9595	0.9595	0.9645	0.9548

ตารางภาคผนวกที่ 115 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคป่าสุ่ม กรณีคัดเลือกตัวแปรอิสระ ด้วยค่าสัมประสิทธิ์สหสัมพันธ์ ภายใต้อุดข้อมูลโรคเบาหวาน (ชุดข้อมูลทดสอบ)

(mtry, ntree)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(4, 250)	0.7374	0.7834	0.6904	0.7374	0.7367	0.7209	0.7582
(4, 500)	0.7373	0.7844	0.6893	0.7373	0.7366	0.7204	0.7588
(5, 250)	0.7309	0.7740	0.6869	0.7309	0.7302	0.7162	0.7494
(5, 500)	0.7317	0.7752	0.6875	0.7317	0.7311	0.7168	0.7505

ตารางภาคผนวกที่ 116 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ตัวแปรอิสระทั้งหมด ภายใต้อุดข้อมูลโรคเบาหวาน (ชุดข้อมูลเรียนรู้)

(gamma, C)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(0.001, 1)	0.7447	0.7258	0.7682	0.7468	0.7441	0.7946	0.6940
(0.01, 1)	0.7485	0.7158	0.7950	0.7550	0.7466	0.8322	0.6631
(0.1, 1)	0.7579	0.7332	0.7899	0.7613	0.7570	0.8183	0.6964

(gamma, C)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(0.2, 1)	0.7759	0.7534	0.8038	0.7784	0.7752	0.8267	0.7241
(0.001, 10)	0.7476	0.7229	0.7798	0.7511	0.7465	0.8108	0.6832
(0.01, 10)	0.7513	0.7214	0.7921	0.7564	0.7497	0.8263	0.6747
(0.1, 10)	0.7853	0.7647	0.8103	0.7873	0.7848	0.8301	0.7396
(0.2, 10)	0.8471	0.8370	0.8582	0.8475	0.8471	0.8657	0.8282

ตารางภาคผนวกที่ 117 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคซ์พอร์ตเวกเตอร์แมชชีน โดยใช้ตัวแปรอิสระทั้งหมด ภายใต้ชุดข้อมูลโรคเบาหวาน (ชุดข้อมูลทดสอบ)

(gamma, C)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(0.001, 1)	0.7441	0.7254	0.7675	0.7464	0.7434	0.7936	0.6936
(0.01, 1)	0.7474	0.7150	0.7937	0.7540	0.7455	0.8309	0.6623
(0.1, 1)	0.7470	0.7234	0.7777	0.7503	0.7460	0.8077	0.6851
(0.2, 1)	0.7453	0.7252	0.7708	0.7478	0.7445	0.7980	0.6916
(0.001, 10)	0.7467	0.7225	0.7785	0.7503	0.7456	0.8091	0.6832
(0.01, 10)	0.7487	0.7194	0.7891	0.7539	0.7471	0.8235	0.6724
(0.1, 10)	0.7408	0.7236	0.7620	0.7426	0.7401	0.7873	0.6935
(0.2, 10)	0.7238	0.7170	0.7315	0.7242	0.7236	0.7482	0.6989

ตารางภาคผนวกที่ 118 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคซ์พอร์ตเวกเตอร์แมชชีน กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการเลือกแบบไปข้างหน้า โดยใช้ข้อมูลทั้งหมด ภายใต้ชุดข้อมูลโรคเบาหวาน (ชุดข้อมูลเรียนรู้)

(gamma, C)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(0.001, 1)	0.7445	0.7944	0.6938	0.7446	0.7439	0.7256	0.7680
(0.01, 1)	0.7480	0.8327	0.6616	0.7480	0.7460	0.7150	0.7951
(0.1, 1)	0.7532	0.8149	0.6903	0.7532	0.7522	0.7284	0.7853
(0.2, 1)	0.7639	0.8193	0.9444	0.7327	0.7329	0.7406	0.7934
(0.001, 10)	0.7473	0.8105	0.6830	0.7474	0.7462	0.7227	0.7795
(0.01, 10)	0.7498	0.8268	0.6712	0.7498	0.7482	0.7194	0.7918
(0.1, 10)	0.7664	0.8206	0.7111	0.7664	0.7656	0.7433	0.7954
(0.2, 10)	0.8013	0.8392	0.7627	0.8013	0.8010	0.7829	0.8231

ตารางภาคผนวกที่ 119 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคซ์พอร์ตเวกเตอร์แมชชีน กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการเลือกแบบไปข้างหน้า โดยใช้ข้อมูลทั้งหมด ภายใต้ชุดข้อมูลโรคเบาหวาน (ชุดข้อมูลทดสอบ)

(gamma, C)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(0.001, 1)	0.7441	0.7937	0.6937	0.7441	0.7434	0.7254	0.7675

(gamma, C)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(0.01, 1)	0.7473	0.8315	0.6624	0.7477	0.7456	0.7146	0.7941
(0.1, 1)	0.7462	0.8078	0.6834	0.7462	0.7451	0.7223	0.7774
(0.2, 1)	0.7455	0.8015	0.6884	0.7455	0.7446	0.7239	0.7731
(0.001, 10)	0.7468	0.8093	0.6832	0.7468	0.7457	0.7225	0.7787
(0.01, 10)	0.7479	0.8244	0.6700	0.7479	0.7463	0.7181	0.7894
(0.1, 10)	0.7441	0.7993	0.6879	0.7441	0.7432	0.7231	0.7710
(0.2, 10)	0.7355	0.7765	0.6937	0.7355	0.7350	0.7211	0.7530

ตารางภาคผนวกที่ 120 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคซัพพอร์ตเวกเตอร์แมชชีน กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีกำจัดแบบถอยหลัง โดยใช้ข้อมูลทั้งหมด ภายใต้ชุดข้อมูลโรคเบาหวาน (ชุดข้อมูลเรียนรู้)

(gamma, C)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(0.001, 1)	0.7445	0.7944	0.6938	0.7446	0.7439	0.7256	0.7680
(0.01, 1)	0.7480	0.8327	0.6616	0.7480	0.7460	0.7150	0.7951
(0.1, 1)	0.7532	0.8149	0.6903	0.7532	0.7522	0.7284	0.7853
(0.2, 1)	0.7639	0.8193	0.9444	0.7327	0.7329	0.7406	0.7934
(0.001, 10)	0.7473	0.8105	0.6830	0.7474	0.7462	0.7227	0.7795
(0.01, 10)	0.7498	0.8268	0.6712	0.7498	0.7482	0.7194	0.7918
(0.1, 10)	0.7664	0.8206	0.7111	0.7664	0.7656	0.7433	0.7954
(0.2, 10)	0.8013	0.8392	0.7627	0.8013	0.8010	0.7829	0.8231

ตารางภาคผนวกที่ 121 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคซัพพอร์ตเวกเตอร์แมชชีน กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีกำจัดแบบถอยหลัง โดยใช้ข้อมูลทั้งหมด ภายใต้ชุดข้อมูลโรคเบาหวาน (ชุดข้อมูลทดสอบ)

(gamma, C)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(0.001, 1)	0.7441	0.7937	0.6937	0.7441	0.7434	0.7254	0.7675
(0.01, 1)	0.7473	0.8315	0.6624	0.7477	0.7456	0.7146	0.7941
(0.1, 1)	0.7462	0.8078	0.6834	0.7462	0.7451	0.7223	0.7774
(0.2, 1)	0.7455	0.8015	0.6884	0.7455	0.7446	0.7239	0.7731
(0.001, 10)	0.7468	0.8093	0.6832	0.7468	0.7457	0.7225	0.7787
(0.01, 10)	0.7479	0.8244	0.6700	0.7479	0.7463	0.7181	0.7894
(0.1, 10)	0.7441	0.7993	0.6879	0.7441	0.7432	0.7231	0.7710
(0.2, 10)	0.7355	0.7765	0.6937	0.7355	0.7350	0.7211	0.7530

ตารางภาคผนวกที่ 122 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคซ์พอร์ตเวกเตอร์แมชชีน กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการถดถอยทีละชั้น โดยใช้ข้อมูลทั้งหมด ภายใต้ชุดข้อมูลโรคเบาหวาน (ชุดข้อมูลเรียนรู้)

(gamma, C)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(0.001, 1)	0.7445	0.7944	0.6938	0.7446	0.7439	0.7256	0.7680
(0.01, 1)	0.7480	0.8327	0.6616	0.7480	0.7460	0.7150	0.7951
(0.1, 1)	0.7532	0.8149	0.6903	0.7532	0.7522	0.7284	0.7853
(0.2, 1)	0.7639	0.8193	0.9444	0.7327	0.7329	0.7406	0.7934
(0.001, 10)	0.7473	0.8105	0.6830	0.7474	0.7462	0.7227	0.7795
(0.01, 10)	0.7498	0.8268	0.6712	0.7498	0.7482	0.7194	0.7918
(0.1, 10)	0.7664	0.8206	0.7111	0.7664	0.7656	0.7433	0.7954
(0.2, 10)	0.8013	0.8392	0.7627	0.8013	0.8010	0.7829	0.8231

ตารางภาคผนวกที่ 123 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคซ์พอร์ตเวกเตอร์แมชชีน กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการถดถอยทีละชั้น โดยใช้ข้อมูลทั้งหมด ภายใต้ชุดข้อมูลโรคเบาหวาน (ชุดข้อมูลทดสอบ)

(gamma, C)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(0.001, 1)	0.7441	0.7937	0.6937	0.7441	0.7434	0.7254	0.7675
(0.01, 1)	0.7473	0.8315	0.6624	0.7477	0.7456	0.7146	0.7941
(0.1, 1)	0.7462	0.8078	0.6834	0.7462	0.7451	0.7223	0.7774
(0.2, 1)	0.7455	0.8015	0.6884	0.7455	0.7446	0.7239	0.7731
(0.001, 10)	0.7468	0.8093	0.6832	0.7468	0.7457	0.7225	0.7787
(0.01, 10)	0.7479	0.8244	0.6700	0.7479	0.7463	0.7181	0.7894
(0.1, 10)	0.7441	0.7993	0.6879	0.7441	0.7432	0.7231	0.7710
(0.2, 10)	0.7355	0.7765	0.6937	0.7355	0.7350	0.7211	0.7530

ตารางภาคผนวกที่ 124 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคซ์พอร์ตเวกเตอร์แมชชีน กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการเลือกแบบไปข้างหน้า โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้ชุดข้อมูลโรคเบาหวาน (ชุดข้อมูลเรียนรู้)

(gamma, C)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(0.001, 1)	0.7447	0.7950	0.6933	0.7446	0.7439	0.7255	0.7684
(0.01, 1)	0.7473	0.8324	0.6606	0.7473	0.7453	0.7143	0.7942
(0.1, 1)	0.7504	0.8124	0.6872	0.7504	0.7493	0.7258	0.7823
(0.2, 1)	0.7567	0.8135	0.6989	0.7567	0.7559	0.7336	0.7861

(gamma, C)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(0.001, 10)	0.7470	0.8094	0.6833	0.7469	0.7477	0.7295	0.7786
(0.01, 10)	0.7492	0.8286	0.6634	0.7492	0.7475	0.7181	0.7927
(0.1, 10)	0.7577	0.8148	0.6995	0.7577	0.7568	0.7343	0.7875
(0.2, 10)	0.7761	0.8247	0.7266	0.7761	0.7755	0.7546	0.8026

ตารางภาคผนวกที่ 125 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคซ์พอร์ตเวกเตอร์แมชชีน กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการเลือกแบบไปข้างหน้า โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้อุณหภูมิข้อมูลโรคเบาหวาน (ชุดข้อมูลทดสอบ)

(gamma, C)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(0.001, 1)	0.7445	0.7946	0.6934	0.7445	0.7438	0.7255	0.7683
(0.01, 1)	0.7469	0.8316	0.6606	0.7469	0.7449	0.7141	0.7940
(0.1, 1)	0.7456	0.8068	0.6833	0.7457	0.7446	0.7220	0.7764
(0.2, 1)	0.7454	0.8020	0.6877	0.7454	0.7445	0.7236	0.7733
(0.001, 10)	0.7466	0.8086	0.6833	0.7466	0.7455	0.7225	0.7781
(0.01, 10)	0.7476	0.8267	0.6669	0.7476	0.7458	0.7168	0.7908
(0.1, 10)	0.7452	0.8023	0.6871	0.7452	0.7443	0.7233	0.7734
(0.2, 10)	0.7412	0.7905	0.6909	0.7412	0.7404	0.7228	0.7641

ตารางภาคผนวกที่ 126 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคซ์พอร์ตเวกเตอร์แมชชีน กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการกำจัดแบบถอยหลัง โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้อุณหภูมิโรคเบาหวาน (ชุดข้อมูลเรียนรู้)

(gamma, C)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(0.001, 1)	0.7447	0.7950	0.6933	0.7446	0.7439	0.7255	0.7684
(0.01, 1)	0.7473	0.8324	0.6606	0.7473	0.7453	0.7143	0.7942
(0.1, 1)	0.7504	0.8124	0.6872	0.7504	0.7493	0.7258	0.7823
(0.2, 1)	0.7567	0.8135	0.6989	0.7567	0.7559	0.7336	0.7861
(0.001, 10)	0.7470	0.8094	0.6833	0.7469	0.7477	0.7295	0.7786
(0.01, 10)	0.7492	0.8286	0.6634	0.7492	0.7475	0.7181	0.7927
(0.1, 10)	0.7577	0.8148	0.6995	0.7577	0.7568	0.7343	0.7875
(0.2, 10)	0.7761	0.8247	0.7266	0.7761	0.7755	0.7546	0.8026

ตารางภาคผนวกที่ 127 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคซ์พอร์ตเวกเตอร์แมชชีน กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีกำจัดแบบถอยหลัง โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้อุณหภูมิของข้อมูลโรคเบาหวาน (ชุดข้อมูลทดสอบ)

(gamma, C)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(0.001, 1)	0.7445	0.7946	0.6934	0.7445	0.7438	0.7255	0.7683
(0.01, 1)	0.7469	0.8316	0.6606	0.7469	0.7449	0.7141	0.7940
(0.1, 1)	0.7456	0.8068	0.6833	0.7457	0.7446	0.7220	0.7764
(0.2, 1)	0.7454	0.8020	0.6877	0.7454	0.7445	0.7236	0.7733
(0.001, 10)	0.7466	0.8086	0.6833	0.7466	0.7455	0.7225	0.7781
(0.01, 10)	0.7476	0.8267	0.6669	0.7476	0.7458	0.7168	0.7908
(0.1, 10)	0.7452	0.8023	0.6871	0.7452	0.7443	0.7233	0.7734
(0.2, 10)	0.7412	0.7905	0.6909	0.7412	0.7404	0.7228	0.7641

ตารางภาคผนวกที่ 128 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคซ์พอร์ตเวกเตอร์แมชชีน กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการถดถอยทีละชั้น โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้อุณหภูมิของข้อมูลโรคเบาหวาน (ชุดข้อมูลเรียนรู้)

(gamma, C)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(0.001, 1)	0.7447	0.7950	0.6933	0.7446	0.7439	0.7255	0.7684
(0.01, 1)	0.7473	0.8324	0.6606	0.7473	0.7453	0.7143	0.7942
(0.1, 1)	0.7504	0.8124	0.6872	0.7504	0.7493	0.7258	0.7823
(0.2, 1)	0.7567	0.8135	0.6989	0.7567	0.7559	0.7336	0.7861
(0.001, 10)	0.7470	0.8094	0.6833	0.7469	0.7477	0.7295	0.7786
(0.01, 10)	0.7492	0.8286	0.6634	0.7492	0.7475	0.7181	0.7927
(0.1, 10)	0.7577	0.8148	0.6995	0.7577	0.7568	0.7343	0.7875
(0.2, 10)	0.7761	0.8247	0.7266	0.7761	0.7755	0.7546	0.8026

ตารางภาคผนวกที่ 129 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคซ์พอร์ตเวกเตอร์แมชชีน กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการถดถอยทีละชั้น โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้อุณหภูมิของข้อมูลโรคเบาหวาน (ชุดข้อมูลทดสอบ)

(gamma, C)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(0.001, 1)	0.7445	0.7946	0.6934	0.7445	0.7438	0.7255	0.7683
(0.01, 1)	0.7469	0.8316	0.6606	0.7469	0.7449	0.7141	0.7940
(0.1, 1)	0.7456	0.8068	0.6833	0.7457	0.7446	0.7220	0.7764
(0.2, 1)	0.7454	0.8020	0.6877	0.7454	0.7445	0.7236	0.7733
(0.001, 10)	0.7466	0.8086	0.6833	0.7466	0.7455	0.7225	0.7781
(0.01, 10)	0.7476	0.8267	0.6669	0.7476	0.7458	0.7168	0.7908

(gamma, C)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(0.1, 10)	0.7452	0.8023	0.6871	0.7452	0.7443	0.7233	0.7734
(0.2, 10)	0.7412	0.7905	0.6909	0.7412	0.7404	0.7228	0.7641

ตารางภาคผนวกที่ 130 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคซัพพอร์ตเวกเตอร์แมชชีน กรณีคัดเลือกตัวแปรอิสระด้วยค่าสัมประสิทธิ์สหสัมพันธ์ ภายใต้ชุดข้อมูลโรคเบาหวาน (ชุดข้อมูลเรียนรู้)

(gamma, C)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(0.001, 1)	0.7436	0.7928	0.6934	0.7436	0.7429	0.7250	0.7665
(0.01, 1)	0.7465	0.8298	0.6617	0.7465	0.7446	0.7143	0.7922
(0.1, 1)	0.7517	0.8127	0.6895	0.7517	0.7507	0.7274	0.7831
(0.2, 1)	0.7605	0.8157	0.7082	0.7625	0.7617	0.7402	0.7903
(0.001, 10)	0.7461	0.8090	0.6820	0.7461	0.7450	0.7217	0.7779
(0.01, 10)	0.7480	0.8235	0.6710	0.7480	0.7464	0.7184	0.7886
(0.1, 10)	0.7648	0.8164	0.7122	0.7648	0.7641	0.7431	0.7919
(0.2, 10)	0.8013	0.8352	0.7668	0.8014	0.8011	0.7850	0.8203

ตารางภาคผนวกที่ 131 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคซัพพอร์ตเวกเตอร์แมชชีน กรณีคัดเลือกตัวแปรอิสระด้วยค่าสัมประสิทธิ์สหสัมพันธ์ ภายใต้ชุดข้อมูลโรคเบาหวาน (ชุดข้อมูลทดสอบ)

(gamma, C)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(0.001, 1)	0.7435	0.7924	0.6936	0.7435	0.7428	0.7250	0.7665
(0.01, 1)	0.7459	0.8290	0.6612	0.7459	0.7440	0.7138	0.7916
(0.1, 1)	0.7444	0.8053	0.6824	0.7444	0.7434	0.7211	0.7748
(0.2, 1)	0.7441	0.7983	0.6889	0.7441	0.7433	0.7235	0.7704
(0.001, 10)	0.7455	0.8082	0.6814	0.7455	0.7443	0.7212	0.7773
(0.01, 10)	0.7461	0.8215	0.6691	0.7461	0.7445	0.7169	0.7865
(0.1, 10)	0.7431	0.7953	0.6898	0.7431	0.7423	0.7233	0.7679
(0.2, 10)	0.7332	0.7707	0.6949	0.7332	0.7327	0.7204	0.7485

ชุดข้อมูลการเลิกสูบบุหรี่

ตารางภาคผนวกที่ 132 เมตริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคการถดถอยลอจิสติก โดยใช้ตัวแปรอิสระทั้งหมด ภายใต้ชุดข้อมูลการเลิกสูบบุหรี่ (ชุดข้อมูลเรียนรู้)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	4365	2555	13745	3890	0.7375	0.5288	0.8433	0.7376	0.7108	0.6308	0.7794
2	4376	2565	13735	3879	0.7376	0.5301	0.8426	0.7375	0.7112	0.6305	0.7798
3	4362	2566	13735	3893	0.7370	0.5284	0.8426	0.7370	0.7104	0.6296	0.7792

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
4	4397	2561	13740	3859	0.7386	0.5326	0.8429	0.7386	0.7126	0.6319	0.7807
5	4346	2536	13765	3910	0.7375	0.5264	0.8444	0.7375	0.7100	0.6315	0.7788
6	4353	2563	13738	3903	0.7367	0.5273	0.8428	0.7367	0.7099	0.6294	0.7788
7	4388	2547	13754	3868	0.7388	0.5315	0.8438	0.7388	0.7123	0.6327	0.7805
8	4395	2550	13751	3861	0.7389	0.5323	0.8436	0.7389	0.7127	0.6328	0.7808
9	4348	2546	13755	3908	0.7372	0.5266	0.8438	0.7372	0.7099	0.6307	0.7787
10	4376	2555	13746	3880	0.7380	0.5300	0.8433	0.7380	0.7114	0.6314	0.7799
Average					0.7378	0.5294	0.8433	0.7378	0.7111	0.6311	0.7797

ตารางภาคผนวกที่ 133 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคการถดถอยลอจิสติก โดยใช้ตัวแปรอิสระทั้งหมด ภายใต้ชุดข้อมูลการเลิกสูบบุหรี่ (ชุดข้อมูลทดสอบ)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	482	279	1533	436	0.7381	0.5251	0.8460	0.7381	0.7313	0.6334	0.7786
2	487	286	1526	431	0.7374	0.5305	0.8422	0.7374	0.7312	0.6300	0.7798
3	478	255	1556	440	0.7453	0.5207	0.8592	0.7454	0.7373	0.6521	0.7796
4	482	308	1503	435	0.7276	0.5256	0.8299	0.7276	0.7221	0.6101	0.7755
5	484	281	1530	433	0.7383	0.5278	0.8448	0.7382	0.7317	0.6327	0.7794
6	497	271	1540	420	0.7467	0.5420	0.8504	0.7467	0.7405	0.6471	0.7857
7	491	304	1507	426	0.7324	0.5354	0.8321	0.7324	0.7272	0.6176	0.7796
8	482	327	1484	435	0.7207	0.5256	0.8194	0.7206	0.7160	0.5958	0.7733
9	479	263	1548	438	0.7430	0.5224	0.8548	0.7430	0.7354	0.6456	0.7795
10	474	281	1530	443	0.7346	0.5169	0.8448	0.7346	0.7274	0.6278	0.7755
Average					0.7364	0.5272	0.8424	0.7364	0.7300	0.6292	0.7786

ตารางภาคผนวกที่ 134 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคการถดถอยลอจิสติก กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการเลือกแบบไปข้างหน้า โดยใช้ข้อมูลทั้งหมด ภายใต้ชุดข้อมูลการเลิกสูบบุหรี่ (ชุดข้อมูลเรียนรู้)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	4352	2569	3903	13731	0.7364	0.5272	0.8424	0.7364	0.7300	0.6288	0.7787
2	4375	2563	3880	13737	0.7376	0.5300	0.8428	0.7376	0.7313	0.6306	0.7798
3	4357	2569	3898	13732	0.7366	0.5278	0.8424	0.7366	0.7303	0.6291	0.7789
4	4399	2561	3857	13740	0.7386	0.5328	0.8429	0.7386	0.7325	0.6320	0.7808
5	4350	2544	3906	13757	0.7373	0.5269	0.8439	0.7373	0.7308	0.6310	0.7789

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
6	4340	2569	3916	13732	0.7359	0.5257	0.8424	0.7359	0.7294	0.6282	0.7781
7	4384	2546	3872	13755	0.7386	0.5310	0.8438	0.7386	0.7323	0.6326	0.7803
8	4395	2551	3861	13750	0.7389	0.5323	0.8435	0.7389	0.7327	0.6327	0.7808
9	4345	2557	3911	13744	0.7366	0.5263	0.8431	0.7366	0.7301	0.6295	0.7785
10	4370	2556	3886	13745	0.7377	0.5293	0.8432	0.7377	0.7313	0.6310	0.7796
Average					0.7374	0.5289	0.8430	0.7374	0.7311	0.6305	0.7794

ตารางภาคผนวกที่ 135 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคการถดถอยลอจิสติก
กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการเลือกแบบไปข้างหน้า โดยใช้ข้อมูล
ทั้งหมด ภายใต้ชุดข้อมูลการเลิกสูบบุหรี่ (ชุดข้อมูลทดสอบ)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	482	277	436	1535	0.7388	0.5251	0.8471	0.7388	0.7320	0.6350	0.7788
2	483	292	435	1520	0.7337	0.5261	0.8389	0.7337	0.7275	0.6232	0.7775
3	480	254	438	1557	0.7464	0.5229	0.8597	0.7465	0.7385	0.6540	0.7805
4	484	309	433	1502	0.7280	0.5278	0.8294	0.7280	0.7226	0.6103	0.7762
5	482	276	435	1535	0.7394	0.5256	0.8476	0.7393	0.7325	0.6359	0.7792
6	492	266	425	1545	0.7467	0.5365	0.8531	0.7467	0.7400	0.6491	0.7843
7	487	302	430	1509	0.7317	0.5311	0.8332	0.7316	0.7262	0.6172	0.7782
8	480	321	437	1490	0.7221	0.5234	0.8227	0.7221	0.7170	0.5993	0.7732
9	481	261	436	1550	0.7445	0.5245	0.8559	0.7445	0.7369	0.6482	0.7805
10	481	288	436	1523	0.7346	0.5245	0.8410	0.7346	0.7282	0.6255	0.7774
Average					0.7366	0.5268	0.8429	0.7366	0.7301	0.6298	0.7786

ตารางภาคผนวกที่ 136 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคการถดถอยลอจิสติก
กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการกำจัดแบบถอยหลัง โดยใช้ข้อมูลทั้งหมด
ภายใต้ชุดข้อมูลการเลิกสูบบุหรี่ (ชุดข้อมูลเรียนรู้)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	4360	2569	3895	13731	0.7368	0.5282	0.8424	0.7368	0.7304	0.6292	0.7790
2	4376	2563	3879	13737	0.7377	0.5301	0.8428	0.7377	0.7314	0.6306	0.7798
3	4360	2574	3895	13727	0.7366	0.5282	0.8421	0.7366	0.7302	0.6288	0.7790
4	4400	2561	3856	13740	0.7387	0.5329	0.8429	0.7387	0.7326	0.6321	0.7809
5	4349	2547	3907	13754	0.7372	0.5268	0.8438	0.7372	0.7307	0.6307	0.7788
6	4346	2567	3910	13734	0.7362	0.5264	0.8425	0.7362	0.7298	0.6287	0.7784

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
7	4379	2546	3877	13755	0.7384	0.5304	0.8438	0.7384	0.7321	0.6323	0.7801
8	4395	2545	3861	13756	0.7391	0.5323	0.8439	0.7391	0.7329	0.6333	0.7808
9	4349	2557	3907	13744	0.7368	0.5268	0.8431	0.7368	0.7303	0.6297	0.7787
10	4371	2560	3885	13741	0.7375	0.5294	0.8430	0.7376	0.7312	0.6306	0.7796
Average					0.7375	0.5292	0.8430	0.7375	0.7312	0.6306	0.7795

ตารางภาคผนวกที่ 137 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคการถดถอยลอจิสติกกรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีกำจัดแบบถอยหลัง โดยใช้ข้อมูลทั้งหมดภายใต้ชุดข้อมูลการเลิกสูบบุหรี่ (ชุดข้อมูลทดสอบ)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	482	280	436	1532	0.7377	0.5251	0.8455	0.7378	0.7310	0.6325	0.7785
2	484	291	434	1521	0.7344	0.5272	0.8394	0.7344	0.7283	0.6245	0.7780
3	480	253	438	1558	0.7468	0.5229	0.8603	0.7469	0.7388	0.6548	0.7806
4	486	309	431	1502	0.7287	0.5300	0.8294	0.7287	0.7235	0.6113	0.7770
5	481	279	436	1532	0.7379	0.5245	0.8459	0.7378	0.7310	0.6329	0.7785
6	492	267	425	1544	0.7463	0.5365	0.8526	0.7463	0.7397	0.6482	0.7842
7	488	301	429	1510	0.7324	0.5322	0.8338	0.7324	0.7269	0.6185	0.7788
8	480	320	437	1491	0.7225	0.5234	0.8233	0.7225	0.7174	0.6000	0.7733
9	482	262	435	1549	0.7445	0.5256	0.8553	0.7445	0.7370	0.6478	0.7807
10	479	286	438	1525	0.7346	0.5224	0.8421	0.7346	0.7280	0.6261	0.7769
Average					0.7366	0.5270	0.8428	0.7366	0.7302	0.6297	0.7786

ตารางภาคผนวกที่ 138 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคการถดถอยลอจิสติกกรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการถดถอยทีละชั้น โดยใช้ข้อมูลทั้งหมดภายใต้ชุดข้อมูลการเลิกสูบบุหรี่ (ชุดข้อมูลเรียนรู้)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	4352	2569	3903	13731	0.7364	0.5272	0.8424	0.7364	0.7300	0.6288	0.7787
2	4375	2563	3880	13737	0.7376	0.5300	0.8428	0.7376	0.7313	0.6306	0.7798
3	4357	2569	3898	13732	0.7366	0.5278	0.8424	0.7366	0.7303	0.6291	0.7789
4	4399	2561	3857	13740	0.7386	0.5328	0.8429	0.7386	0.7325	0.6320	0.7808
5	4350	2544	3906	13757	0.7373	0.5269	0.8439	0.7373	0.7308	0.6310	0.7789
6	4340	2569	3916	13732	0.7359	0.5257	0.8424	0.7359	0.7294	0.6282	0.7781
7	4384	2546	3872	13755	0.7386	0.5310	0.8438	0.7386	0.7323	0.6326	0.7803

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
8	4395	2551	3861	13750	0.7389	0.5323	0.8435	0.7389	0.7327	0.6327	0.7808
9	4345	2557	3911	13744	0.7366	0.5263	0.8431	0.7366	0.7301	0.6295	0.7785
10	4370	2556	3886	13745	0.7377	0.5293	0.8432	0.7377	0.7313	0.6310	0.7796
Average					0.7374	0.5289	0.8430	0.7374	0.7311	0.6305	0.7794

ตารางภาคผนวกที่ 139 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคการถดถอยลอจิสติก กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการถดถอยทีละชั้น โดยใช้ข้อมูลทั้งหมด ภายใต้ชุดข้อมูลการเลิกสูบบุหรี่ (ชุดข้อมูลทดสอบ)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	482	277	436	1535	0.7388	0.5251	0.8471	0.7388	0.7320	0.6350	0.7788
2	483	292	435	1520	0.7337	0.5261	0.8389	0.7337	0.7275	0.6232	0.7775
3	480	254	438	1557	0.7464	0.5229	0.8597	0.7465	0.7385	0.6540	0.7805
4	484	309	433	1502	0.7280	0.5278	0.8294	0.7280	0.7226	0.6103	0.7762
5	482	276	435	1535	0.7394	0.5256	0.8476	0.7393	0.7325	0.6359	0.7792
6	492	266	425	1545	0.7467	0.5365	0.8531	0.7467	0.7400	0.6491	0.7843
7	487	302	430	1509	0.7317	0.5311	0.8332	0.7316	0.7262	0.6172	0.7782
8	480	321	437	1490	0.7221	0.5234	0.8227	0.7221	0.7170	0.5993	0.7732
9	481	261	436	1550	0.7445	0.5245	0.8559	0.7445	0.7369	0.6482	0.7805
10	481	288	436	1523	0.7346	0.5245	0.8410	0.7346	0.7282	0.6255	0.7774
Average					0.7366	0.5268	0.8429	0.7366	0.7301	0.6298	0.7786

ตารางภาคผนวกที่ 140 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคการถดถอยลอจิสติก กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการเลือกแบบไปข้างหน้า โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้ชุดข้อมูลการเลิกสูบบุหรี่ (ชุดข้อมูลเรียนรู้)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	4353	2566	3902	13734	0.7366	0.5273	0.8426	0.7366	0.7302	0.6291	0.7787
2	4356	2559	3899	13741	0.7370	0.5277	0.8430	0.7370	0.7306	0.6299	0.7790
3	4346	2560	3909	13741	0.7366	0.5265	0.8430	0.7366	0.7301	0.6293	0.7785
4	4370	2549	3886	13752	0.7380	0.5293	0.8436	0.7379	0.7316	0.6316	0.7797
5	4325	2537	3931	13764	0.7366	0.5239	0.8444	0.7367	0.7299	0.6303	0.7778
6	4325	2560	3931	13741	0.7357	0.5239	0.8430	0.7357	0.7291	0.6282	0.7776

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
7	4360	2534	3896	13767	0.7382	0.5281	0.8445	0.7381	0.7316	0.6324	0.7794
8	4368	2527	3888	13774	0.7388	0.5291	0.8450	0.7388	0.7323	0.6335	0.7799
9	4337	2544	3919	13757	0.7368	0.5253	0.8439	0.7368	0.7302	0.6303	0.7783
10	4364	2553	3892	13748	0.7375	0.5286	0.8434	0.7376	0.7312	0.6309	0.7794
Average					0.7372	0.5270	0.8436	0.7372	0.7307	0.6306	0.7788

ตารางภาคผนวกที่ 141 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคการถดถอยลอจิสติก กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการเลือกแบบไปข้างหน้า โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้อัตราข้อมูลการเลิกสูบบุหรี่ (ชุดข้อมูลทดสอบ)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	482	270	436	1542	0.7414	0.5251	0.8510	0.7414	0.7342	0.6410	0.7796
2	480	286	438	1526	0.7348	0.5229	0.8422	0.7349	0.7282	0.6266	0.7770
3	473	256	445	1555	0.7431	0.5153	0.8586	0.7432	0.7348	0.6488	0.7775
4	482	304	435	1507	0.7291	0.5256	0.8321	0.7291	0.7234	0.6132	0.7760
5	490	274	427	1537	0.7430	0.5344	0.8487	0.7430	0.7365	0.6414	0.7826
6	496	268	421	1543	0.7474	0.5409	0.8520	0.7474	0.7410	0.6492	0.7856
7	487	304	430	1507	0.7309	0.5311	0.8321	0.7309	0.7255	0.6157	0.7780
8	484	319	433	1492	0.7243	0.5278	0.8239	0.7244	0.7194	0.6027	0.7751
9	479	260	438	1551	0.7441	0.5224	0.8564	0.7441	0.7364	0.6482	0.7798
10	481	280	436	1531	0.7375	0.5245	0.8454	0.7375	0.7307	0.6321	0.7783
Average					0.7376	0.5270	0.8442	0.7376	0.7310	0.6319	0.7789

ตารางภาคผนวกที่ 142 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคการถดถอยลอจิสติก กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการกำจัดแบบถอยหลัง โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้อัตราข้อมูลการเลิกสูบบุหรี่ (ชุดข้อมูลเรียนรู้)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	4295	2547	3960	13753	0.7350	0.5203	0.8437	0.7350	0.7281	0.6277	0.7764
2	4316	2553	3939	13747	0.7356	0.5228	0.8434	0.7356	0.7289	0.6283	0.7773
3	4273	2552	3982	13749	0.7339	0.5176	0.8434	0.7339	0.7269	0.6261	0.7754
4	4333	2544	3923	13757	0.7367	0.5248	0.8439	0.7366	0.7300	0.6301	0.7781

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
5	4296	2543	3960	13758	0.7352	0.5203	0.8440	0.7352	0.7283	0.6282	0.7765
6	4278	2556	3978	13745	0.7339	0.5182	0.8432	0.7339	0.7270	0.6260	0.7755
7	4299	2535	3957	13766	0.7356	0.5207	0.8445	0.7356	0.7287	0.6291	0.7767
8	4322	2530	3934	13771	0.7368	0.5235	0.8448	0.7368	0.7300	0.6308	0.7778
9	4286	2520	3970	13781	0.7357	0.5191	0.8454	0.7357	0.7286	0.6297	0.7764
10	4305	2544	3951	13757	0.7355	0.5214	0.8439	0.7355	0.7286	0.6286	0.7769
Average					0.7354	0.5209	0.8440	0.7354	0.7285	0.6284	0.7767

ตารางภาคผนวกที่ 143 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคการถดถอยลอจิสติก กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีกำจัดแบบถอยหลัง โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้อัตราข้อมูลการเลิกสูบบุหรี่ (ชุดข้อมูลทดสอบ)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	475	275	443	1537	0.7370	0.5174	0.8482	0.7763	0.7763	0.7763	0.7763
2	472	292	446	1520	0.7297	0.5142	0.8389	0.7731	0.7731	0.7731	0.7731
3	477	248	441	1563	0.7475	0.5196	0.8631	0.7799	0.7799	0.7799	0.7799
4	477	313	440	1498	0.7240	0.5202	0.8272	0.7730	0.7730	0.7730	0.7730
5	464	262	453	1549	0.7379	0.5060	0.8553	0.7737	0.7737	0.7737	0.7737
6	486	273	431	1538	0.7419	0.5300	0.8493	0.7811	0.7811	0.7811	0.7811
7	488	311	429	1500	0.7287	0.5322	0.8283	0.7776	0.7776	0.7776	0.7776
8	475	306	442	1505	0.7258	0.5180	0.8310	0.7730	0.7730	0.7730	0.7730
9	466	275	451	1536	0.7339	0.5082	0.8482	0.7730	0.7730	0.7730	0.7730
10	474	285	443	1526	0.7331	0.5169	0.8426	0.7750	0.7750	0.7750	0.7750
Average					0.7340	0.5183	0.8432	0.7340	0.7270	0.6265	0.7756

ตารางภาคผนวกที่ 144 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคการถดถอยลอจิสติก กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการถดถอยทีละชั้น โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้อัตราข้อมูลการเลิกสูบบุหรี่ (ชุดข้อมูลเรียนรู้)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	4353	2566	3902	13734	0.7366	0.5273	0.8426	0.7366	0.7302	0.6291	0.7787
2	4356	2559	3899	13741	0.7370	0.5277	0.8430	0.7370	0.7306	0.6299	0.7790
3	4346	2560	3909	13741	0.7366	0.5265	0.8430	0.7366	0.7301	0.6293	0.7785

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
4	4370	2549	3886	13752	0.7380	0.5293	0.8436	0.7379	0.7316	0.6316	0.7797
5	4325	2537	3931	13764	0.7366	0.5239	0.8444	0.7367	0.7299	0.6303	0.7778
6	4325	2560	3931	13741	0.7357	0.5239	0.8430	0.7357	0.7291	0.6282	0.7776
7	4360	2534	3896	13767	0.7382	0.5281	0.8445	0.7381	0.7316	0.6324	0.7794
8	4368	2527	3888	13774	0.7388	0.5291	0.8450	0.7388	0.7323	0.6335	0.7799
9	4337	2544	3919	13757	0.7368	0.5253	0.8439	0.7368	0.7302	0.6303	0.7783
10	4364	2553	3892	13748	0.7375	0.5286	0.8434	0.7376	0.7312	0.6309	0.7794
Average					0.7372	0.5270	0.8436	0.7372	0.7307	0.6306	0.7788

ตารางภาคผนวกที่ 145 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคการถดถอยลอจิสติกกรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการถดถอยที่ละชั้น โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้ชุดข้อมูลการเลิกสูบบุหรี่ (ชุดข้อมูลทดสอบ)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	482	270	436	1542	0.7414	0.5251	0.8510	0.7414	0.7342	0.6410	0.7796
2	480	286	438	1526	0.7348	0.5229	0.8422	0.7349	0.7282	0.6266	0.7770
3	473	256	445	1555	0.7431	0.5153	0.8586	0.7432	0.7348	0.6488	0.7775
4	482	304	435	1507	0.7291	0.5256	0.8321	0.7291	0.7234	0.6132	0.7760
5	490	274	427	1537	0.7430	0.5344	0.8487	0.7430	0.7365	0.6414	0.7826
6	496	268	421	1543	0.7474	0.5409	0.8520	0.7474	0.7410	0.6492	0.7856
7	487	304	430	1507	0.7309	0.5311	0.8321	0.7309	0.7255	0.6157	0.7780
8	484	319	433	1492	0.7243	0.5278	0.8239	0.7244	0.7194	0.6027	0.7751
9	479	260	438	1551	0.7441	0.5224	0.8564	0.7441	0.7364	0.6482	0.7798
10	481	280	436	1531	0.7375	0.5245	0.8454	0.7375	0.7307	0.6321	0.7783
Average					0.7376	0.5270	0.8442	0.7376	0.7310	0.6319	0.7789

ตารางภาคผนวกที่ 146 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคการถดถอยลอจิสติกกรณีคัดเลือกตัวแปรอิสระด้วยค่าสัมประสิทธิ์สหสัมพันธ์ ภายใต้ชุดข้อมูลการเลิกสูบบุหรี่ (ชุดข้อมูลเรียนรู้)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	4304	2578	3951	13722	0.7341	0.5214	0.8418	0.7341	0.7274	0.6254	0.7764
2	4306	2574	3949	13726	0.7344	0.5216	0.8421	0.7344	0.7277	0.6259	0.7766
3	4279	2581	3976	13720	0.7330	0.5184	0.8417	0.7330	0.7261	0.6238	0.7753
4	4325	2574	3931	13727	0.7351	0.5239	0.8421	0.7351	0.7285	0.6269	0.7774

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
5	4281	2569	3975	13732	0.7335	0.5185	0.8424	0.7335	0.7266	0.6250	0.7755
6	4274	2577	3982	13724	0.7329	0.5177	0.8419	0.7329	0.7260	0.6239	0.7751
7	4298	2561	3958	13740	0.7345	0.5206	0.8429	0.7345	0.7277	0.6266	0.7764
8	4311	2553	3945	13748	0.7354	0.5222	0.8434	0.7354	0.7286	0.6281	0.7770
9	4262	2553	3994	13748	0.7334	0.5162	0.8434	0.7334	0.7263	0.6254	0.7749
10	4295	2569	3961	13732	0.7341	0.5202	0.8424	0.7341	0.7273	0.6257	0.7761
Average					0.7340	0.5201	0.8424	0.7340	0.7272	0.6257	0.7761

ตารางภาคผนวกที่ 147 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคการถดถอยลอจิสติกกรณีคัดเลือกตัวแปรอิสระด้วยค่าสัมประสิทธิ์สหสัมพันธ์ ภายใต้ชุดข้อมูลการเลิกสูบบุหรี่ (ชุดข้อมูลทดสอบ)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	470	274	448	1538	0.7355	0.5120	0.8488	0.7356	0.7278	0.6317	0.7744
2	479	300	439	1512	0.7293	0.5218	0.8344	0.7293	0.7232	0.6149	0.7750
3	475	258	443	1553	0.7431	0.5174	0.8575	0.7432	0.7350	0.6480	0.7781
4	475	312	442	1499	0.7236	0.5180	0.8277	0.7236	0.7178	0.6036	0.7723
5	461	261	456	1550	0.7372	0.5027	0.8559	0.7372	0.7282	0.6385	0.7727
6	487	272	430	1539	0.7427	0.5311	0.8498	0.7427	0.7359	0.6416	0.7816
7	481	305	436	1506	0.7284	0.5245	0.8316	0.7284	0.7227	0.6120	0.7755
8	475	313	442	1498	0.7232	0.5180	0.8272	0.7232	0.7175	0.6028	0.7722
9	477	275	440	1536	0.7379	0.5202	0.8482	0.7379	0.7307	0.6343	0.7773
10	482	278	435	1533	0.7386	0.5256	0.8465	0.7386	0.7318	0.6342	0.7790
Average					0.7340	0.5191	0.8428	0.7340	0.7271	0.6262	0.7758

ตารางภาคผนวกที่ 148 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคต้นไม้ตัดสินใจ โดยใช้ตัวแปรอิสระทั้งหมด ภายใต้ชุดข้อมูลการเลิกสูบบุหรี่ (ชุดข้อมูลเรียนรู้)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	6793	866	1462	15434	0.9052	0.8229	0.9469	0.9052	0.9043	0.8869	0.9135
2	6721	1209	1534	15091	0.8883	0.8142	0.9258	0.8883	0.8877	0.8475	0.9077
3	6940	913	1315	15388	0.9093	0.8407	0.9440	0.9093	0.9087	0.8837	0.9213
4	6804	807	1452	15494	0.9080	0.8241	0.9505	0.9080	0.9070	0.8940	0.9143
5	6885	798	1371	15503	0.9117	0.8339	0.9510	0.9116	0.9108	0.8961	0.9188

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
6	7064	1206	1192	15095	0.9023	0.8556	0.9260	0.9023	0.9024	0.8542	0.9268
7	6998	929	1258	15372	0.9109	0.8476	0.9430	0.9109	0.9105	0.8828	0.9244
8	6695	1089	1561	15212	0.8921	0.8109	0.9332	0.8921	0.8913	0.8601	0.9069
9	6521	747	1735	15554	0.8989	0.7898	0.9542	0.8989	0.8972	0.8972	0.8996
10	7089	1151	1167	15150	0.9056	0.8586	0.9294	0.9056	0.9056	0.8603	0.9285
Average					0.9032	0.8298	0.9404	0.9032	0.9025	0.8763	0.9162

ตารางภาคผนวกที่ 149 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคต้นไม้ตัดสินใจ โดยใช้ตัวแปรอิสระทั้งหมด ภายใต้ชุดข้อมูลการเลิกสูบบุหรี่ (ชุดข้อมูลทดสอบ)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	466	374	452	1438	0.6974	0.5076	0.7936	0.6974	0.6939	0.5548	0.7608
2	500	401	418	1411	0.7000	0.5447	0.7787	0.7000	0.6993	0.5549	0.7715
3	481	352	437	1459	0.7109	0.5240	0.8056	0.7109	0.7072	0.5774	0.7695
4	498	376	419	1435	0.7086	0.5431	0.7924	0.7086	0.7068	0.5698	0.7740
5	473	350	444	1461	0.7089	0.5158	0.8067	0.7089	0.7047	0.5747	0.7669
6	519	382	398	1429	0.7141	0.5660	0.7891	0.7141	0.7135	0.5760	0.7822
7	514	372	403	1439	0.7159	0.5605	0.7946	0.7159	0.7147	0.5801	0.7812
8	498	368	419	1443	0.7115	0.5431	0.7968	0.7115	0.7094	0.5751	0.7750
9	468	337	449	1474	0.7119	0.5104	0.8139	0.7119	0.7068	0.5814	0.7665
10	520	420	397	1391	0.7005	0.5671	0.7681	0.7005	0.7014	0.5532	0.7780
Average					0.7080	0.5382	0.7940	0.7080	0.7058	0.5697	0.7726

ตารางภาคผนวกที่ 150 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคต้นไม้ตัดสินใจ กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลงจิสติวิธีการเลือกแบบไปข้างหน้า โดยใช้ข้อมูลทั้งหมด ภายใต้ชุดข้อมูลการเลิกสูบบุหรี่ (ชุดข้อมูลเรียนรู้)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	6171	1677	2084	14623	0.8468	0.7475	0.8971	0.8468	0.8458	0.7863	0.8753
2	6254	1917	2001	14383	0.8404	0.7576	0.8824	0.8404	0.8402	0.7654	0.8779
3	6122	1797	2133	14504	0.8400	0.7416	0.8898	0.8400	0.8391	0.7731	0.8718
4	6546	1523	1710	14778	0.8683	0.7929	0.9066	0.8684	0.8680	0.8113	0.8963
5	6310	1446	1946	14855	0.8619	0.7643	0.9113	0.8619	0.8608	0.8136	0.8842
6	6146	1713	2110	14588	0.8443	0.7444	0.8949	0.8443	0.8433	0.7820	0.8736
7	6049	1518	2207	14783	0.8483	0.7327	0.9069	0.8483	0.8466	0.7994	0.8701

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
8	6394	1529	1862	14772	0.8619	0.7745	0.9062	0.8619	0.8612	0.8070	0.8881
9	6051	1380	2205	14921	0.8540	0.7329	0.9153	0.8540	0.8520	0.8143	0.8712
10	6445	2021	1811	14280	0.8440	0.7806	0.8760	0.8439	0.8444	0.7613	0.8875
Average					0.8510	0.7569	0.8987	0.8510	0.8501	0.7914	0.8796

ตารางภาคผนวกที่ 151 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคต้นไม้ตัดสินใจ กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการเลือกแบบไปข้างหน้า โดยใช้ข้อมูลทั้งหมด ภายใต้ชุดข้อมูลการเลิกสูบบุหรี่ (ชุดข้อมูลทดสอบ)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	500	341	418	1471	0.7220	0.5447	0.8118	0.7220	0.7188	0.5945	0.7787
2	525	403	393	1409	0.7084	0.5719	0.7776	0.7084	0.7088	0.5657	0.7819
3	512	335	406	1476	0.7285	0.5577	0.8150	0.7285	0.7257	0.6045	0.7843
4	523	373	394	1438	0.7188	0.5703	0.7940	0.7188	0.7180	0.5837	0.7849
5	481	373	436	1438	0.7034	0.5245	0.7940	0.7034	0.7007	0.5632	0.7673
6	494	339	423	1472	0.7207	0.5387	0.8128	0.7206	0.7171	0.5930	0.7768
7	509	346	408	1465	0.7236	0.5551	0.8089	0.7236	0.7211	0.5953	0.7822
8	494	391	423	1420	0.7016	0.5387	0.7841	0.7016	0.7003	0.5582	0.7705
9	468	327	449	1484	0.7155	0.5104	0.8194	0.7155	0.7100	0.5887	0.7677
10	550	410	367	1401	0.7152	0.5998	0.7736	0.7152	0.7167	0.5729	0.7924
Average					0.7158	0.5512	0.7991	0.7158	0.7137	0.5820	0.7787

ตารางภาคผนวกที่ 152 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคต้นไม้ตัดสินใจ กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการกำจัดแบบถอยหลัง โดยใช้ข้อมูลทั้งหมด ภายใต้ชุดข้อมูลการเลิกสูบบุหรี่ (ชุดข้อมูลเรียนรู้)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	6295	1781	1960	14519	0.8476	0.7626	0.8907	0.8476	0.8472	0.7795	0.8811
2	6402	1882	1853	14418	0.8479	0.7755	0.8845	0.8479	0.8479	0.7728	0.8861
3	6363	1738	1892	14563	0.8522	0.7708	0.8934	0.8522	0.8518	0.7855	0.8850
4	6472	2048	1784	14253	0.8440	0.7839	0.8744	0.8440	0.8446	0.7596	0.8888
5	6281	1452	1975	14849	0.8604	0.7608	0.9109	0.8604	0.8593	0.8122	0.8826
6	6183	1590	2073	14711	0.8508	0.7489	0.9025	0.8509	0.8497	0.7954	0.8765
7	6309	1420	1947	14881	0.8629	0.7642	0.9129	0.8629	0.8617	0.8163	0.8843
8	6267	1383	1989	14918	0.8627	0.7591	0.9152	0.8627	0.8613	0.8192	0.8824

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
9	5812	1374	2444	14927	0.8445	0.7040	0.9157	0.8445	0.8416	0.8088	0.8593
10	6459	1938	1797	14363	0.8479	0.7823	0.8811	0.8479	0.8482	0.7692	0.8888
Average					0.8521	0.7612	0.8981	0.8521	0.8513	0.7919	0.8815

ตารางภาคผนวกที่ 153 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคต้นไม้ตัดสินใจ กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการจัดแบบถอยหลัง โดยใช้ข้อมูลทั้งหมด ภายใต้ชุดข้อมูลการเลิกสูบบุหรี่ (ชุดข้อมูลทดสอบ)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	527	365	391	1447	0.7231	0.5741	0.7986	0.7231	0.7221	0.5908	0.7873
2	527	419	391	1393	0.7033	0.5741	0.7688	0.7033	0.7044	0.5571	0.7808
3	522	358	396	1453	0.7237	0.5686	0.8023	0.7237	0.7223	0.5932	0.7858
4	533	403	384	1408	0.7115	0.5812	0.7775	0.7115	0.7122	0.5694	0.7857
5	481	369	436	1442	0.7049	0.5245	0.7962	0.7049	0.7020	0.5659	0.7678
6	500	331	417	1480	0.7258	0.5453	0.8172	0.7258	0.7222	0.6017	0.7802
7	514	369	403	1442	0.7170	0.5605	0.7962	0.7170	0.7156	0.5821	0.7816
8	491	365	426	1446	0.7100	0.5354	0.7985	0.7100	0.7075	0.5736	0.7724
9	450	308	467	1503	0.7159	0.4907	0.8299	0.7159	0.7084	0.5937	0.7629
10	538	400	379	1411	0.7144	0.5867	0.7791	0.7144	0.7152	0.5736	0.7883
Average					0.7150	0.5541	0.7964	0.7150	0.7132	0.5801	0.7793

ตารางภาคผนวกที่ 154 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคต้นไม้ตัดสินใจ กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการถดถอยทีละชั้น โดยใช้ข้อมูลทั้งหมด ภายใต้ชุดข้อมูลการเลิกสูบบุหรี่ (ชุดข้อมูลเรียนรู้)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	6171	1677	2084	14623	0.8468	0.7475	0.8971	0.8468	0.8458	0.7863	0.8753
2	6254	1917	2001	14383	0.8404	0.7576	0.8824	0.8404	0.8402	0.7654	0.8779
3	6122	1797	2133	14504	0.8400	0.7416	0.8898	0.8400	0.8391	0.7731	0.8718
4	6546	1523	1710	14778	0.8683	0.7929	0.9066	0.8684	0.8680	0.8113	0.8963
5	6310	1446	1946	14855	0.8619	0.7643	0.9113	0.8619	0.8608	0.8136	0.8842
6	6146	1713	2110	14588	0.8443	0.7444	0.8949	0.8443	0.8433	0.7820	0.8736
7	6049	1518	2207	14783	0.8483	0.7327	0.9069	0.8483	0.8466	0.7994	0.8701
8	6394	1529	1862	14772	0.8619	0.7745	0.9062	0.8619	0.8612	0.8070	0.8881
9	6051	1380	2205	14921	0.8540	0.7329	0.9153	0.8540	0.8520	0.8143	0.8712

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
10	6445	2021	1811	14280	0.8440	0.7806	0.8760	0.8439	0.8444	0.7613	0.8875
Average					0.8510	0.7569	0.8987	0.8510	0.8501	0.7914	0.8796

ตารางภาคผนวกที่ 155 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคต้นไม้ตัดสินใจ กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการถดถอยทีละชั้น โดยใช้ข้อมูลทั้งหมด ภายใต้ชุดข้อมูลการเลิกสูบบุหรี่ (ชุดข้อมูลทดสอบ)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	500	341	418	1471	0.7220	0.5447	0.8118	0.7220	0.7188	0.5945	0.7787
2	525	403	393	1409	0.7084	0.5719	0.7776	0.7084	0.7088	0.5657	0.7819
3	512	335	406	1476	0.7285	0.5577	0.8150	0.7285	0.7257	0.6045	0.7843
4	523	373	394	1438	0.7188	0.5703	0.7940	0.7188	0.7180	0.5837	0.7849
5	481	373	436	1438	0.7034	0.5245	0.7940	0.7034	0.7007	0.5632	0.7673
6	494	339	423	1472	0.7207	0.5387	0.8128	0.7206	0.7171	0.5930	0.7768
7	509	346	408	1465	0.7236	0.5551	0.8089	0.7236	0.7211	0.5953	0.7822
8	494	391	423	1420	0.7016	0.5387	0.7841	0.7016	0.7003	0.5582	0.7705
9	468	327	449	1484	0.7155	0.5104	0.8194	0.7155	0.7100	0.5887	0.7677
10	550	410	367	1401	0.7152	0.5998	0.7736	0.7152	0.7167	0.5729	0.7924
Average					0.7158	0.5512	0.7991	0.7158	0.7137	0.5820	0.7787

ตารางภาคผนวกที่ 156 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคต้นไม้ตัดสินใจ กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการเลือกแบบไปข้างหน้า โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้ชุดข้อมูลการเลิกสูบบุหรี่ (ชุดข้อมูลเรียนรู้)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	5888	2123	2367	14177	0.8171	0.7133	0.8698	0.8172	0.8165	0.7350	0.8569
2	5963	2219	2292	14081	0.8163	0.7224	0.8639	0.8163	0.8161	0.7288	0.8600
3	5773	2346	2482	13955	0.8034	0.6993	0.8561	0.8034	0.8030	0.7110	0.8490
4	6263	2347	1993	13954	0.8233	0.7586	0.8560	0.8233	0.8241	0.7274	0.8750
5	5621	1978	2635	14323	0.8122	0.6808	0.8787	0.8122	0.8101	0.7397	0.8446
6	5859	2073	2397	14228	0.8180	0.7097	0.8728	0.8180	0.8170	0.7387	0.8558
7	5736	1784	2520	14517	0.8247	0.6948	0.8906	0.8248	0.8226	0.7628	0.8521
8	6152	1952	2104	14349	0.8348	0.7452	0.8803	0.8349	0.8345	0.7591	0.8721

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
9	5749	2025	2507	14276	0.8154	0.6963	0.8758	0.8155	0.8140	0.7395	0.8506
10	5659	1847	2597	14454	0.8190	0.6854	0.8867	0.8190	0.8168	0.7539	0.8477
Average					0.8184	0.7106	0.8731	0.8184	0.8175	0.7396	0.8564

ตารางภาคผนวกที่ 157 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคต้นไม้ตัดสินใจ กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการเลือกแบบไปข้างหน้า โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้อุณหภูมิการเลิกสุบหรี (ชุดข้อมูลทดสอบ)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	508	353	410	1459	0.7205	0.5534	0.8052	0.7205	0.7182	0.5900	0.7806
2	526	378	392	1434	0.7179	0.5730	0.7914	0.7180	0.7174	0.5819	0.7853
3	544	346	374	1465	0.7362	0.5926	0.8089	0.7362	0.7352	0.6112	0.7966
4	566	375	351	1436	0.7339	0.6172	0.7929	0.7338	0.7347	0.6015	0.8036
5	508	336	409	1475	0.7269	0.5540	0.8145	0.7269	0.7239	0.6019	0.7829
6	538	334	379	1477	0.7386	0.5867	0.8156	0.7386	0.7370	0.6170	0.7958
7	514	342	403	1469	0.7269	0.5605	0.8112	0.7269	0.7245	0.6005	0.7847
8	526	372	391	1439	0.7203	0.5736	0.7946	0.7203	0.7196	0.5857	0.7863
9	498	355	419	1456	0.7163	0.5431	0.8040	0.7163	0.7136	0.5838	0.7765
10	508	346	409	1465	0.7232	0.5540	0.8089	0.7232	0.7207	0.5948	0.7818
Average					0.7261	0.5708	0.8047	0.7261	0.7245	0.5968	0.7874

ตารางภาคผนวกที่ 158 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคต้นไม้ตัดสินใจ กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการกำจัดแบบถอยหลัง โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้อุณหภูมิการเลิกสุบหรี (ชุดข้อมูลเรียนรู้)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	5708	2119	2547	14181	0.8100	0.6915	0.8700	0.8100	0.8087	0.7293	0.8477
2	5768	2249	2487	14051	0.8071	0.6987	0.8620	0.8071	0.8064	0.7195	0.8496
3	5790	2367	2465	13934	0.8032	0.7014	0.8548	0.8032	0.8029	0.7098	0.8497
4	6101	2500	2155	13801	0.8104	0.7390	0.8466	0.8104	0.8114	0.7093	0.8649
5	5413	1970	2843	14331	0.8040	0.6556	0.8791	0.8040	0.8011	0.7332	0.8345
6	5578	2132	2678	14169	0.8041	0.6756	0.8692	0.8041	0.8024	0.7235	0.8410
7	5631	1993	2625	14308	0.8119	0.6820	0.8777	0.8119	0.8100	0.7386	0.8450

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
8	6061	2155	2195	14146	0.8229	0.7341	0.8678	0.8229	0.8227	0.7377	0.8657
9	5590	2102	2666	14199	0.8058	0.6771	0.8711	0.8059	0.8041	0.7267	0.8419
10	5517	1813	2739	14488	0.8146	0.6682	0.8888	0.8146	0.8117	0.7527	0.8410
Average					0.8094	0.6923	0.8687	0.8094	0.8081	0.7280	0.8481

ตารางภาคผนวกที่ 159 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคต้นไม้ตัดสินใจ กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีกำจัดแบบถอยหลัง โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้อัตราข้อมูลการเลิกสูบบุหรี่ (ชุดข้อมูลทดสอบ)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	515	341	403	1471	0.7275	0.5610	0.8118	0.7275	0.7250	0.6016	0.7850
2	530	372	388	1440	0.7216	0.5773	0.7947	0.7216	0.7210	0.5876	0.7877
3	540	345	378	1466	0.7351	0.5882	0.8095	0.7351	0.7339	0.6102	0.7950
4	580	400	337	1411	0.7298	0.6325	0.7791	0.7298	0.7319	0.5918	0.8072
5	491	324	426	1487	0.7251	0.5354	0.8211	0.7250	0.7207	0.6025	0.7773
6	533	325	384	1486	0.7401	0.5812	0.8205	0.7400	0.7378	0.6212	0.7947
7	526	346	391	1465	0.7298	0.5736	0.8089	0.7298	0.7281	0.6032	0.7893
8	531	376	386	1435	0.7207	0.5791	0.7924	0.7207	0.7203	0.5854	0.7880
9	486	349	431	1462	0.7141	0.5300	0.8073	0.7141	0.7105	0.5820	0.7723
10	500	330	417	1481	0.7262	0.5453	0.8178	0.7262	0.7226	0.6024	0.7803
Average					0.7270	0.5704	0.8063	0.7270	0.7252	0.5988	0.7877

ตารางภาคผนวกที่ 160 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคต้นไม้ตัดสินใจ กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการถดถอยทีละชั้น โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้อัตราข้อมูลการเลิกสูบบุหรี่ (ชุดข้อมูลเรียนรู้)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	5888	2123	2367	14177	0.8171	0.7133	0.8698	0.8172	0.8165	0.7350	0.8569
2	5963	2219	2292	14081	0.8163	0.7224	0.8639	0.8163	0.8161	0.7288	0.8600
3	5773	2346	2482	13955	0.8034	0.6993	0.8561	0.8034	0.8030	0.7110	0.8490
4	6263	2347	1993	13954	0.8233	0.7586	0.8560	0.8233	0.8241	0.7274	0.8750
5	5621	1978	2635	14323	0.8122	0.6808	0.8787	0.8122	0.8101	0.7397	0.8446
6	5859	2073	2397	14228	0.8180	0.7097	0.8728	0.8180	0.8170	0.7387	0.8558
7	5736	1784	2520	14517	0.8247	0.6948	0.8906	0.8248	0.8226	0.7628	0.8521

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
8	6152	1952	2104	14349	0.8348	0.7452	0.8803	0.8349	0.8345	0.7591	0.8721
9	5749	2025	2507	14276	0.8154	0.6963	0.8758	0.8155	0.8140	0.7395	0.8506
10	5659	1847	2597	14454	0.8190	0.6854	0.8867	0.8190	0.8168	0.7539	0.8477
Average					0.8184	0.7106	0.8731	0.8184	0.8175	0.7396	0.8564

ตารางภาคผนวกที่ 161 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคต้นไม้ตัดสินใจ กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการถดถอยทีละชั้น โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้อัตราข้อมูลการเลิกสูบบุหรี่ (ชุดข้อมูลทดสอบ)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	508	353	410	1459	0.7205	0.5534	0.8052	0.7205	0.7182	0.5900	0.7806
2	526	378	392	1434	0.7179	0.5730	0.7914	0.7180	0.7174	0.5819	0.7853
3	544	346	374	1465	0.7362	0.5926	0.8089	0.7362	0.7352	0.6112	0.7966
4	566	375	351	1436	0.7339	0.6172	0.7929	0.7338	0.7347	0.6015	0.8036
5	508	336	409	1475	0.7269	0.5540	0.8145	0.7269	0.7239	0.6019	0.7829
6	538	334	379	1477	0.7386	0.5867	0.8156	0.7386	0.7370	0.6170	0.7958
7	514	342	403	1469	0.7269	0.5605	0.8112	0.7269	0.7245	0.6005	0.7847
8	526	372	391	1439	0.7203	0.5736	0.7946	0.7203	0.7196	0.5857	0.7863
9	498	355	419	1456	0.7163	0.5431	0.8040	0.7163	0.7136	0.5838	0.7765
10	508	346	409	1465	0.7232	0.5540	0.8089	0.7232	0.7207	0.5948	0.7818
Average					0.7261	0.5708	0.8047	0.7261	0.7245	0.5968	0.7874

ตารางภาคผนวกที่ 162 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคต้นไม้ตัดสินใจ กรณีคัดเลือกตัวแปรอิสระด้วยค่าสัมประสิทธิ์สหสัมพันธ์ ภายใต้อัตราข้อมูลการเลิกสูบบุหรี่ (ชุดข้อมูลเรียนรู้)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	6730	1525	1525	14775	0.8758	0.8153	0.9064	0.8758	0.8758	0.8153	0.9064
2	6520	1587	1735	14713	0.8647	0.7898	0.9026	0.8647	0.8644	0.8042	0.8945
3	6759	1625	1496	14676	0.8729	0.8188	0.9003	0.8729	0.8731	0.8062	0.9075
4	6436	1319	1820	14982	0.8722	0.7796	0.9191	0.8722	0.8712	0.8299	0.8917
5	6556	1485	1700	14816	0.8703	0.7941	0.9089	0.8703	0.8699	0.8153	0.8971
6	6553	1728	1703	14573	0.8603	0.7937	0.8940	0.8603	0.8603	0.7913	0.8954
7	6597	1034	1659	15267	0.8903	0.7991	0.9366	0.8904	0.8892	0.8645	0.9020
8	6680	1262	1576	15039	0.8844	0.8091	0.9226	0.8844	0.8839	0.8411	0.9051
9	6553	1265	1703	15036	0.8791	0.7937	0.9224	0.8791	0.8783	0.8382	0.8983

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
10	6836	1464	1420	14837	0.8826	0.8280	0.9102	0.8826	0.8826	0.8236	0.9127
Average					0.8753	0.8021	0.9123	0.8753	0.8749	0.8230	0.9011

ตารางภาคผนวกที่ 163 เมทริกซ์ความสับสนของตัวแบบการจำแนกจากเทคนิคต้นไม้ตัดสินใจ กรณีคัดเลือกตัวแปรอิสระด้วยค่าสัมประสิทธิ์สหสัมพันธ์ ภายใต้ชุดข้อมูลการเลิกสูบบุหรี่ (ชุดข้อมูลทดสอบ)

Fold	TP	FP	TN	FN	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
1	506	382	412	1430	0.7092	0.5512	0.7892	0.7092	0.7080	0.5698	0.7763
2	547	398	371	1414	0.7183	0.5959	0.7804	0.7184	0.7193	0.5788	0.7922
3	535	364	383	1447	0.7263	0.5828	0.7990	0.7263	0.7256	0.5951	0.7907
4	476	372	441	1439	0.7020	0.5191	0.7946	0.7020	0.6989	0.5613	0.7654
5	526	360	391	1451	0.7247	0.5736	0.8012	0.7247	0.7235	0.5937	0.7877
6	549	384	368	1427	0.7243	0.5987	0.7880	0.7244	0.7249	0.5884	0.7950
7	493	372	424	1439	0.7082	0.5376	0.7946	0.7082	0.7060	0.5699	0.7724
8	496	367	421	1444	0.7111	0.5409	0.7973	0.7111	0.7089	0.5747	0.7743
9	497	361	420	1450	0.7137	0.5420	0.8007	0.7137	0.7112	0.5793	0.7754
10	549	398	368	1413	0.7192	0.5987	0.7802	0.7192	0.7203	0.5797	0.7934
Average					0.7157	0.5641	0.7925	0.7157	0.7147	0.5791	0.7823

ตารางภาคผนวกที่ 164 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคป่าสุ่ม โดยใช้ตัวแปรอิสระทั้งหมด ภายใต้ชุดข้อมูลการเลิกสูบบุหรี่ (ชุดข้อมูลเรียนรู้)

(mtry, ntree)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(4, 250)	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
(4, 500)	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
(5, 250)	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
(5, 500)	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

ตารางภาคผนวกที่ 165 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคป่าสุ่ม โดยใช้ตัวแปรอิสระทั้งหมด ภายใต้ชุดข้อมูลการเลิกสูบบุหรี่ (ชุดข้อมูลทดสอบ)

(mtry, ntree)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(4, 250)	0.7524	0.6409	0.8036	0.7489	0.7502	0.5995	0.8298
(4, 500)	0.7532	0.6426	0.8039	0.7497	0.7509	0.5997	0.8309
(5, 250)	0.7520	0.6399	0.8037	0.7486	0.7499	0.6002	0.8288

(mtry, ntree)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(4, 500)	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

ตารางภาคผนวกที่ 173 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคป่าสุ่ม กรณีคัดเลือกตัวแปรอิสระ ด้วยเทคนิคการถดถอยลอจิสติกวิธีการเลือกแบบไปข้างหน้า โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้อุณหภูมิการสุ่ม (ชุดข้อมูลทดสอบ)

(mtry, ntree)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(3, 250)	0.7427	0.5906	0.8196	0.7426	0.7408	0.6239	0.7981
(3, 500)	0.7444	0.5922	0.8215	0.7443	0.7426	0.6271	0.7991
(4, 250)	0.7430	0.5972	0.8169	0.7430	0.7416	0.6231	0.8002
(4, 500)	0.7433	0.5981	0.8169	0.7433	0.7419	0.6234	0.8005

ตารางภาคผนวกที่ 174 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคป่าสุ่ม กรณีคัดเลือกตัวแปรอิสระ ด้วยเทคนิคการถดถอยลอจิสติกวิธีการกำจัดแบบถอยหลัง โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้อุณหภูมิการสุ่ม (ชุดข้อมูลเรียนรู้)

(mtry, ntree)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(3, 250)	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
(3, 500)	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
(4, 250)	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
(4, 500)	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

ตารางภาคผนวกที่ 175 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคป่าสุ่ม กรณีคัดเลือกตัวแปรอิสระ ด้วยเทคนิคการถดถอยลอจิสติกวิธีการกำจัดแบบถอยหลัง โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้อุณหภูมิการสุ่ม (ชุดข้อมูลทดสอบ)

(mtry, ntree)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(3, 250)	0.7423	0.5955	0.8167	0.7423	0.7409	0.6220	0.7995
(3, 500)	0.7421	0.5957	0.8162	0.7421	0.7407	0.6215	0.7995
(4, 250)	0.7417	0.5945	0.8162	0.7416	0.7382	0.6097	0.7990
(4, 500)	0.7426	0.5949	0.8174	0.7426	0.7411	0.6227	0.7994

ตารางภาคผนวกที่ 176 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคป่าสุ่ม กรณีคัดเลือกตัวแปรอิสระ ด้วยเทคนิคการถดถอยลอจิสติกวิธีการถดถอยที่ละชั้น โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้ชุดข้อมูลการเลิกสูบบุหรี่ (ชุดข้อมูลเรียนรู้)

(mtry, ntree)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(3, 250)	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
(3, 500)	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
(4, 250)	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
(4, 500)	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

ตารางภาคผนวกที่ 177 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคป่าสุ่ม กรณีคัดเลือกตัวแปรอิสระ ด้วยเทคนิคการถดถอยลอจิสติกวิธีการถดถอยที่ละชั้น โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้ชุดข้อมูลการเลิกสูบบุหรี่ (ชุดข้อมูลทดสอบ)

(mtry, ntree)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(3, 250)	0.7427	0.5906	0.8196	0.7426	0.7408	0.6239	0.7981
(3, 500)	0.7444	0.5922	0.8215	0.7443	0.7426	0.6271	0.7991
(4, 250)	0.7430	0.5972	0.8169	0.7430	0.7416	0.6231	0.8002
(4, 500)	0.7433	0.5981	0.8169	0.7433	0.7419	0.6234	0.8005

ตารางภาคผนวกที่ 178 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคป่าสุ่ม กรณีคัดเลือกตัวแปรอิสระ ด้วยค่าสัมประสิทธิ์สหสัมพันธ์ ภายใต้ชุดข้อมูลการเลิกสูบบุหรี่ (ชุดข้อมูลเรียนรู้)

(mtry, ntree)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(3, 250)	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
(3, 500)	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
(4, 250)	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
(4, 500)	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

ตารางภาคผนวกที่ 179 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคป่าสุ่ม กรณีคัดเลือกตัวแปรอิสระ ด้วยค่าสัมประสิทธิ์สหสัมพันธ์ ภายใต้ชุดข้อมูลการเลิกสูบบุหรี่ (ชุดข้อมูลทดสอบ)

(mtry, ntree)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(3, 250)	0.7518	0.5985	0.8294	0.7518	0.7496	0.6400	0.8031
(3, 500)	0.7534	0.6009	0.8306	0.7534	0.7512	0.6426	0.8043
(4, 250)	0.7495	0.6018	0.8244	0.7496	0.7478	0.6346	0.8035
(4, 500)	0.7504	0.6020	0.8255	0.7503	0.7486	0.6361	0.8037

ตารางภาคผนวกที่ 180 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคซ์พอร์ตเวกเตอร์แมชชีน โดยใช้ตัวแปรอิสระทั้งหมด ภายใต้ชุดข้อมูลการเลิกสูบบุหรี่ (ชุดข้อมูลเรียนรู้)

(gamma, C)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(0.001, 1)	0.7412	0.6436	0.7772	0.7323	0.7331	0.5158	0.8554
(0.01, 1)	0.7595	0.6623	0.8002	0.7538	0.7553	0.5809	0.8500
(0.1, 1)	0.8326	0.7710	0.8605	0.8304	0.8310	0.7143	0.8926
(0.2, 1)	0.9080	0.8843	0.9191	0.9074	0.9074	0.8358	0.9446
(0.001, 10)	0.7491	0.6516	0.7873	0.7416	0.7429	0.5452	0.8524
(0.01, 10)	0.7706	0.6748	0.8128	0.7664	0.7677	0.6133	0.8503
(0.1, 10)	0.9407	0.9275	0.9471	0.9405	0.9404	0.8936	0.9646
(0.2, 10)	0.9968	0.9974	0.9966	0.9969	0.9968	0.9932	0.9987

ตารางภาคผนวกที่ 181 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคซ์พอร์ตเวกเตอร์แมชชีน โดยใช้ตัวแปรอิสระทั้งหมด ภายใต้ชุดข้อมูลการเลิกสูบบุหรี่ (ชุดข้อมูลทดสอบ)

(gamma, C)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(0.001, 1)	0.7400	0.6420	0.7763	0.7311	0.7319	0.5136	0.8547
(0.01, 1)	0.7538	0.6534	0.7958	0.7479	0.7493	0.5710	0.8464
(0.1, 1)	0.7516	0.6439	0.7992	0.7470	0.7485	0.5853	0.8359
(0.2, 1)	0.7426	0.6361	0.7863	0.7358	0.7374	0.5488	0.8408
(0.001, 10)	0.7465	0.6474	0.7856	0.7391	0.7403	0.5417	0.8503
(0.01, 10)	0.7578	0.6543	0.8034	0.7533	0.7547	0.5936	0.8410
(0.1, 10)	0.7245	0.5963	0.7837	0.7207	0.7222	0.5597	0.8080
(0.2, 10)	0.7133	0.5798	0.7734	0.7083	0.7103	0.5347	0.8038

ตารางภาคผนวกที่ 182 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคซ์พอร์ตเวกเตอร์แมชชีน กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการเลือกแบบไปข้างหน้า โดยใช้ข้อมูลทั้งหมด ภายใต้ชุดข้อมูลการเลิกสูบบุหรี่ (ชุดข้อมูลเรียนรู้)

(gamma, C)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(0.001, 1)	0.7411	0.5182	0.8541	0.7411	0.7333	0.6426	0.7778
(0.01, 1)	0.7556	0.5746	0.8473	0.7556	0.7513	0.6558	0.7973
(0.1, 1)	0.7941	0.6386	0.8609	0.7861	0.7836	0.6992	0.8247
(0.2, 1)	0.8330	0.7099	0.8953	0.8330	0.8311	0.7745	0.8590
(0.001, 10)	0.7466	0.5400	0.8511	0.7465	0.7402	0.6476	0.7852
(0.01, 10)	0.7596	0.5898	0.8455	0.7596	0.7560	0.6592	0.8028
(0.1, 10)	0.8490	0.7318	0.9084	0.8490	0.8496	0.8191	0.8699
(0.2, 10)	0.9396	0.8857	0.9669	0.9396	0.9392	0.9313	0.9435

ตารางภาคผนวกที่ 183 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคซ์พอร์ตเวกเตอร์แมชชีน กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการเลือกแบบไปข้างหน้า โดยใช้ข้อมูลทั้งหมด ภายใต้ชุดข้อมูลการเลิกสูบบุหรี่ (ชุดข้อมูลทดสอบ)

(gamma, C)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(0.001, 1)	0.7402	0.5165	0.8535	0.7402	0.7323	0.6414	0.7771
(0.01, 1)	0.7531	0.5701	0.8458	0.7531	0.7486	0.6522	0.7953
(0.1, 1)	0.7539	0.5907	0.8365	0.7539	0.7509	0.6468	0.8014
(0.2, 1)	0.7478	0.5786	0.8335	0.7478	0.7445	0.6378	0.7962
(0.001, 10)	0.7453	0.5392	0.8496	0.7453	0.7390	0.6453	0.7845
(0.01, 10)	0.7538	0.5813	0.8411	0.7538	0.7501	0.6499	0.7987
(0.1, 10)	0.7361	0.5592	0.8251	0.7358	0.7314	0.6188	0.7875
(0.2, 10)	0.27143	0.5397	0.8027	0.7143	0.7116	0.5809	0.7750

ตารางภาคผนวกที่ 184 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคซ์พอร์ตเวกเตอร์แมชชีน กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการกำจัดแบบถอยหลัง โดยใช้ข้อมูลทั้งหมด ภายใต้ชุดข้อมูลการเลิกสูบบุหรี่ (ชุดข้อมูลเรียนรู้)

(gamma, C)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(0.001, 1)	0.7411	0.5186	0.8537	0.7411	0.7333	0.6423	0.7779
(0.01, 1)	0.7558	0.5761	0.8468	0.7558	0.7515	0.6556	0.7977
(0.1, 1)	0.7862	0.6391	0.8607	0.7862	0.7837	0.6992	0.8248
(0.2, 1)	0.8328	0.7086	0.8957	0.8328	0.8309	0.7749	0.8585
(0.001, 10)	0.7469	0.5427	0.8508	0.7472	0.7408	0.6478	0.7857
(0.01, 10)	0.7599	0.5901	0.8459	0.7599	0.7563	0.6598	0.8029
(0.1, 10)	0.8501	0.7330	0.9094	0.8501	0.8483	0.8038	0.8705
(0.2, 10)	0.9405	0.8870	0.9676	0.9405	0.9401	0.9328	0.9442

ตารางภาคผนวกที่ 185 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคซ์พอร์ตเวกเตอร์แมชชีน กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการกำจัดแบบถอยหลัง โดยใช้ข้อมูลทั้งหมด ภายใต้ชุดข้อมูลการเลิกสูบบุหรี่ (ชุดข้อมูลทดสอบ)

(gamma, C)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(0.001, 1)	0.7401	0.5171	0.8531	0.7401	0.7323	0.6410	0.7772
(0.01, 1)	0.7530	0.5711	0.8451	0.7530	0.7486	0.6516	0.7956
(0.1, 1)	0.7533	0.5903	0.8359	0.7533	0.7504	0.6458	0.8012
(0.2, 1)	0.7485	0.5798	0.8339	0.7485	0.7453	0.6389	0.7967
(0.001, 10)	0.7452	0.5404	0.8489	0.7452	0.7390	0.6447	0.7848
(0.01, 10)	0.7541	0.5813	0.8417	0.7541	0.7504	0.6507	0.7988

(gamma, C)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(0.1, 10)	0.7376	0.5619	0.8266	0.7376	0.7341	0.6214	0.7884
(0.2, 10)	0.7143	0.5406	0.8024	0.7144	0.7117	0.5807	0.7753

ตารางภาคผนวกที่ 186 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคซัพพอร์ตเวกเตอร์แมชชีน กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการถดถอยทีละชั้น โดยใช้ข้อมูลทั้งหมด ภายใต้ชุดข้อมูลการเลิกสูบบุหรี่ (ชุดข้อมูลเรียนรู้)

(gamma, C)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(0.001, 1)	0.7411	0.5182	0.8541	0.7411	0.7333	0.6426	0.7778
(0.01, 1)	0.7556	0.5746	0.8473	0.7556	0.7513	0.6558	0.7973
(0.1, 1)	0.7941	0.6386	0.8609	0.7861	0.7836	0.6992	0.8247
(0.2, 1)	0.8330	0.7099	0.8953	0.8330	0.8311	0.7745	0.8590
(0.001, 10)	0.7466	0.5400	0.8511	0.7465	0.7402	0.6476	0.7852
(0.01, 10)	0.7596	0.5898	0.8455	0.7596	0.7560	0.6592	0.8028
(0.1, 10)	0.8490	0.7318	0.9084	0.8490	0.8496	0.8191	0.8699
(0.2, 10)	0.9396	0.8857	0.9669	0.9396	0.9392	0.9313	0.9435

ตารางภาคผนวกที่ 187 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคซัพพอร์ตเวกเตอร์แมชชีน กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการถดถอยทีละชั้น โดยใช้ข้อมูลทั้งหมด ภายใต้ชุดข้อมูลการเลิกสูบบุหรี่ (ชุดข้อมูลทดสอบ)

(gamma, C)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(0.001, 1)	0.7402	0.5165	0.8535	0.7402	0.7323	0.6414	0.7771
(0.01, 1)	0.7531	0.5701	0.8458	0.7531	0.7486	0.6522	0.7953
(0.1, 1)	0.7539	0.5907	0.8365	0.7539	0.7509	0.6468	0.8014
(0.2, 1)	0.7478	0.5786	0.8335	0.7478	0.7445	0.6378	0.7962
(0.001, 10)	0.7453	0.5392	0.8496	0.7453	0.7390	0.6453	0.7845
(0.01, 10)	0.7538	0.5813	0.8411	0.7538	0.7501	0.6499	0.7987
(0.1, 10)	0.7361	0.5592	0.8251	0.7358	0.7314	0.6188	0.7875
(0.2, 10)	0.27143	0.5397	0.8027	0.7143	0.7116	0.5809	0.7750

ตารางภาคผนวกที่ 188 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคซ์ฟอर्टเวกเตอร์แมชชีน กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการเลือกแบบไปข้างหน้า โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้อัตราข้อมูลการเลิกสูบบุหรี่ (ชุดข้อมูลเรียนรู้)

(gamma, C)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(0.001, 1)	0.7403	0.5139	0.8550	0.7403	0.7322	0.6423	0.7764
(0.01, 1)	0.7511	0.5553	0.8503	0.7511	0.7456	0.6527	0.7906
(0.1, 1)	0.7707	0.6090	0.8526	0.7707	0.7675	0.6766	0.8115
(0.2, 1)	0.8020	0.6591	0.8743	0.8020	0.7994	0.7265	0.8351
(0.001, 10)	0.7450	0.5348	0.8515	0.7450	0.7383	0.6459	0.7833
(0.01, 10)	0.7537	0.5683	0.8477	0.7538	0.7491	0.6539	0.7950
(0.1, 10)	0.8102	0.6680	0.8823	0.8102	0.8076	0.7418	0.8399
(0.2, 10)	0.8884	0.7927	0.9368	0.8884	0.8871	0.8640	0.8992

ตารางภาคผนวกที่ 189 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคซ์ฟอर्टเวกเตอร์แมชชีน กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการเลือกแบบไปข้างหน้า โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้อัตราข้อมูลการเลิกสูบบุหรี่ (ชุดข้อมูลทดสอบ)

(gamma, C)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(0.001, 1)	0.7395	0.5126	0.8544	0.7395	0.7313	0.6411	0.7758
(0.01, 1)	0.7494	0.5531	0.8488	0.7494	0.7439	0.6497	0.7895
(0.1, 1)	0.7492	0.5750	0.8374	0.7492	0.7455	0.6420	0.7955
(0.2, 1)	0.7455	0.5744	0.8321	0.7455	0.7421	0.6343	0.7943
(0.001, 10)	0.7440	0.5336	0.8505	0.7440	0.7386	0.6543	0.7827
(0.01, 10)	0.7503	0.5633	0.8451	0.7503	0.7456	0.6484	0.7926
(0.1, 10)	0.7415	0.5624	0.8323	0.7416	0.7377	0.6296	0.7896
(0.2, 10)	0.7203	0.5364	0.8134	0.7203	0.7166	0.5930	0.7760

ตารางภาคผนวกที่ 190 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคซ์ฟอर्टเวกเตอร์แมชชีน กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการกำจัดแบบถอยหลัง โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้อัตราข้อมูลการเลิกสูบบุหรี่ (ชุดข้อมูลเรียนรู้)

(gamma, C)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(0.001, 1)	0.7397	0.5120	0.8556	0.7397	0.7314	0.6418	0.7755
(0.01, 1)	0.7500	0.5501	0.8513	0.7500	0.7442	0.6519	0.7888
(0.1, 1)	0.7636	0.5950	0.8489	0.7636	0.7600	0.6661	0.8054

(gamma, C)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(0.2, 1)	0.7855	0.6289	0.8649	0.7855	0.7824	0.7022	0.8215
(0.001, 10)	0.7443	0.5307	0.8525	0.7443	0.7373	0.6456	0.7820
(0.01, 10)	0.7527	0.5628	0.8489	0.7527	0.7477	0.6535	0.7931
(0.1, 10)	0.7921	0.6322	0.8731	0.7921	0.7886	0.7161	0.8242
(0.2, 10)	0.8546	0.7308	0.9173	0.8546	0.8524	0.8173	0.8706

ตารางภาคผนวกที่ 191 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคซัพพอร์ตเวกเตอร์แมชชีน กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีกำจัดแบบถอยหลัง โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้อัตราข้อมูลการเลิกสูบบุหรี่ (ชุดข้อมูลทดสอบ)

(gamma, C)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(0.001, 1)	0.7391	0.5098	0.8553	0.7391	0.7307	0.6414	0.7750
(0.01, 1)	0.7481	0.5478	0.8495	0.7481	0.7422	0.6487	0.7877
(0.1, 1)	0.7483	0.5717	0.8378	0.7483	0.7445	0.6412	0.7943
(0.2, 1)	0.7440	0.5660	0.8341	0.7440	0.7401	0.6336	0.7915
(0.001, 10)	0.7433	0.5293	0.8518	0.7433	0.7363	0.6443	0.7813
(0.01, 10)	0.7502	0.5599	0.8465	0.7502	0.7451	0.6491	0.7916
(0.1, 10)	0.7402	0.5546	0.8343	0.7402	0.7358	0.6291	0.7871
(0.2, 10)	0.7254	0.5342	0.8223	0.7254	0.7209	0.6036	0.7771

ตารางภาคผนวกที่ 192 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคซัพพอร์ตเวกเตอร์แมชชีน กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการถดถอยทีละขั้น โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้อัตราข้อมูลการเลิกสูบบุหรี่ (ชุดข้อมูลเรียนรู้)

(gamma, C)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(0.001, 1)	0.7403	0.5139	0.8550	0.7403	0.7322	0.6423	0.7764
(0.01, 1)	0.7511	0.5553	0.8503	0.7511	0.7456	0.6527	0.7906
(0.1, 1)	0.7707	0.6090	0.8526	0.7707	0.7675	0.6766	0.8115
(0.2, 1)	0.8020	0.6591	0.8743	0.8020	0.7994	0.7265	0.8351
(0.001, 10)	0.7450	0.5348	0.8515	0.7450	0.7383	0.6459	0.7833
(0.01, 10)	0.7537	0.5683	0.8477	0.7538	0.7491	0.6539	0.7950
(0.1, 10)	0.8102	0.6680	0.8823	0.8102	0.8076	0.7418	0.8399
(0.2, 10)	0.8884	0.7927	0.9368	0.8884	0.8871	0.8640	0.8992

ตารางภาคผนวกที่ 193 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคซ์ฟอว์ตเวกเตอร์แมชชีน กรณีคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยลอจิสติกวิธีการถดถอยทีละชั้น โดยการสร้างตัวแบบ 10 ครั้งตามรูปแบบ 10-Fold Cross-Validation ภายใต้อุณหภูมิการเลิกสุบบุหรี (ชุดข้อมูลทดสอบ)

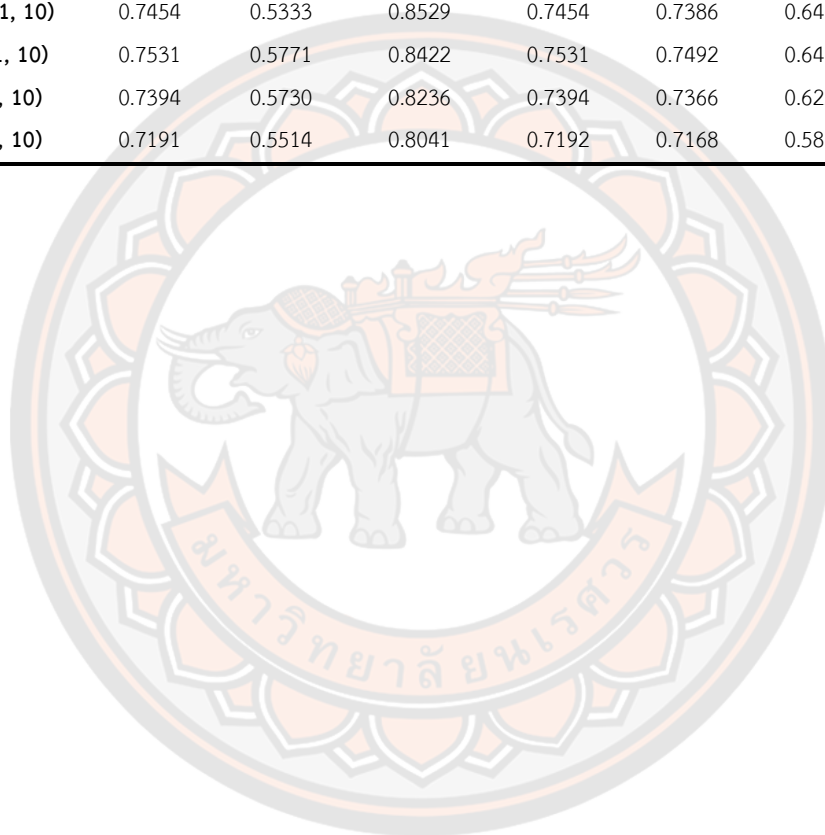
(gamma, C)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(0.001, 1)	0.7395	0.5126	0.8544	0.7395	0.7313	0.6411	0.7758
(0.01, 1)	0.7494	0.5531	0.8488	0.7494	0.7439	0.6497	0.7895
(0.1, 1)	0.7492	0.5750	0.8374	0.7492	0.7455	0.6420	0.7955
(0.2, 1)	0.7455	0.5744	0.8321	0.7455	0.7421	0.6343	0.7943
(0.001, 10)	0.7440	0.5336	0.8505	0.7440	0.7386	0.6543	0.7827
(0.01, 10)	0.7503	0.5633	0.8451	0.7503	0.7456	0.6484	0.7926
(0.1, 10)	0.7415	0.5624	0.8323	0.7416	0.7377	0.6296	0.7896
(0.2, 10)	0.7203	0.5364	0.8134	0.7203	0.7166	0.5930	0.7760

ตารางภาคผนวกที่ 194 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคซ์ฟอว์ตเวกเตอร์แมชชีน กรณีคัดเลือกตัวแปรอิสระด้วยค่าสัมประสิทธิ์สหสัมพันธ์ ภายใต้อุณหภูมิการเลิกสุบบุหรี (ชุดข้อมูลเรียนรู้)

(gamma, C)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(0.001, 1)	0.7405	0.5098	0.8574	0.7405	0.7319	0.6441	0.7754
(0.01, 1)	0.7539	0.5678	0.8481	0.7539	0.7492	0.6544	0.7949
(0.1, 1)	0.7870	0.6399	0.8614	0.7870	0.7844	0.7005	0.8253
(0.2, 1)	0.8330	0.7139	0.8933	0.8330	0.8313	0.7722	0.8604
(0.001, 10)	0.7470	0.5357	0.8540	0.7470	0.7402	0.6502	0.7840
(0.01, 10)	0.7592	0.5880	0.8459	0.7592	0.7555	0.6589	0.8021
(0.1, 10)	0.8502	0.7405	0.9058	0.8502	0.8487	0.7992	0.8733
(0.2, 10)	0.9429	0.8960	0.9666	0.9429	0.9426	0.9314	0.9483

ตารางภาคผนวกที่ 195 ผลลัพธ์ของตัวแบบการจำแนกจากเทคนิคซ์ฟอर्टเวกเตอร์แมชชีน กรณี
คัดเลือกตัวแปรอิสระด้วยค่าสัมประสิทธิ์สหสัมพันธ์ ภายใต้ชุดข้อมูลการเลิกสูบบุหรี่ (ชุดข้อมูล
ทดสอบ)

(gamma, C)	Acc.	Sen.	Spe.	BAcc.	F1	Pre. (Yes)	Pre. (No)
(0.001, 1)	0.7404	0.5086	0.8578	0.7404	0.7316	0.6447	0.7751
(0.01, 1)	0.7516	0.5642	0.8465	0.7516	0.7468	0.6510	0.7932
(0.1, 1)	0.7534	0.5880	0.8371	0.7534	0.7503	0.6467	0.8005
(0.2, 1)	0.7499	0.5852	0.8333	0.7499	0.7469	0.6403	0.7987
(0.001, 10)	0.7454	0.5333	0.8529	0.7454	0.7386	0.6478	0.7830
(0.01, 10)	0.7531	0.5771	0.8422	0.7531	0.7492	0.6499	0.7973
(0.1, 10)	0.7394	0.5730	0.8236	0.7394	0.7366	0.6221	0.7921
(0.2, 10)	0.7191	0.5514	0.8041	0.7192	0.7168	0.5878	0.7797





ภาคผนวก ข
โปรแกรมอาร์

มหาวิทยาลัยบูรณะ

การสร้างตัวแบบการจำแนกด้วยเทคนิคการถดถอยลอจิสติก

```

library(readxl)
library(caret)
library(blorr)
trainfile <- c("NTrain_Cardio1.xlsx","NTrain_Cardio2.xlsx","NTrain_Cardio3.xlsx","NTrain_Cardio4.xlsx",
              "NTrain_Cardio5.xlsx","NTrain_Cardio6.xlsx","NTrain_Cardio7.xlsx","NTrain_Cardio8.xlsx",
              "NTrain_Cardio9.xlsx", "NTrain_Cardio10.xlsx") #ข้อมูลโรคหัวใจและหลอดเลือด
testfile <- c("NTest_Cardio1.xlsx","NTest_Cardio2.xlsx","NTest_Cardio3.xlsx","NTest_Cardio4.xlsx",
             "NTest_Cardio5.xlsx","NTest_Cardio6.xlsx","NTest_Cardio7.xlsx","NTest_Cardio8.xlsx",
             "NTest_Cardio9.xlsx","NTest_Cardio10.xlsx") #ข้อมูลโรคหัวใจและหลอดเลือด

for (i in 1:10) {
  print(i)
  train <- read_excel(trainfile[i])
  test <- read_excel(testfile[i])
  set.seed(1)
  train$class <- factor(train$class, levels = c(1,0))
  train$gender <- factor(train$gender, levels = c("women","men"))
  train$smoke <- factor(train$smoke, levels = c("No","Yes"))
  train$alco <- factor(train$alco, levels = c("No","Yes"))
  train$active <- factor(train$active, levels = c("No","Yes"))
  test$class <- factor(test$class, levels = c(1,0))
  test$gender <- factor(test$gender, levels = c("women","men"))
  test$smoke <- factor(test$smoke, levels = c("No","Yes"))
  test$alco <- factor(test$alco, levels = c("No","Yes"))
  test$active <- factor(test$active, levels = c("No","Yes"))
  #----- ตัวแบบเทคนิคการถดถอยลอจิสติก -----
  model_LR <- glm(class ~ .,data = train, family = binomial("logit"))
  #----- ผลการวิเคราะห์ -----
  trainResult<- blr_confusion_matrix(model_LR,data = train) #confusion Matrix & Accuracy train
  testResult <- blr_confusion_matrix(model_LR,data = test) #confusion Matrix & Accuracy test
  print(trainResult)
  print(testResult)
}

```

การสร้างตัวแบบการจำแนกด้วยเทคนิคต้นไม้ตัดสินใจ

```

library(RWeka)
library(readxl)
library(caret)

trainfile <- c("NTrain_Cardio1.xlsx","NTrain_Cardio2.xlsx","NTrain_Cardio3.xlsx","NTrain_Cardio4.xlsx",
              "NTrain_Cardio5.xlsx","NTrain_Cardio6.xlsx","NTrain_Cardio7.xlsx","NTrain_Cardio8.xlsx",
              "NTrain_Cardio9.xlsx", "NTrain_Cardio10.xlsx") #ข้อมูลโรคหัวใจและหลอดเลือด
testfile <- c("NTest_Cardio1.xlsx","NTest_Cardio2.xlsx","NTest_Cardio3.xlsx","NTest_Cardio4.xlsx",
              "NTest_Cardio5.xlsx","NTest_Cardio6.xlsx","NTest_Cardio7.xlsx","NTest_Cardio8.xlsx",
              "NTest_Cardio9.xlsx","NTest_Cardio10.xlsx") #ข้อมูลโรคหัวใจและหลอดเลือด

for (i in 1:10) {
  print(i)
  train <- read_excel(trainfile[i])
  test <- read_excel(testfile[i])
  set.seed(1)
  train$class <- factor(train$class, levels = c(1,0))
  train$gender <- factor(train$gender, levels = c("women","men"))
  train$smoke <- factor(train$smoke, levels = c("No","Yes"))
  train$alco <- factor(train$alco, levels = c("No","Yes"))
  train$active <- factor(train$active, levels = c("No","Yes"))
  test$class <- factor(test$class, levels = c(1,0))
  test$gender <- factor(test$gender, levels = c("women","men"))
  test$smoke <- factor(test$smoke, levels = c("No","Yes"))
  test$alco <- factor(test$alco, levels = c("No","Yes"))
  test$active <- factor(test$active, levels = c("No","Yes"))
  #----- ตัวแบบเทคนิคต้นไม้ตัดสินใจ -----
  model_DT <- J48(class ~., data = train)
  #----- ผลการวิเคราะห์ -----
  ypred <- predict(model_DT, newdata = train)
  Train_Result <- confusionMatrix(table(train$class,ypred))
  print(Train_Result)
  ypred <- predict(model_DT, newdata = test)
  Test_Result <- confusionMatrix(table(test$class,ypred))
  print(Test_Result)
}

```

การสร้างตัวแบบการจำแนกด้วยเทคนิคป่าสุ่ม

```

library(randomForest)
library(readxl)
library(caret)
trainfile <- c("NTrain_Cardio1.xlsx","NTrain_Cardio2.xlsx","NTrain_Cardio3.xlsx","NTrain_Cardio4.xlsx",
              "NTrain_Cardio5.xlsx","NTrain_Cardio6.xlsx","NTrain_Cardio7.xlsx","NTrain_Cardio8.xlsx",
              "NTrain_Cardio9.xlsx", "NTrain_Cardio10.xlsx") #ข้อมูลโรคหัวใจและหลอดเลือด
testfile <- c("NTest_Cardio1.xlsx","NTest_Cardio2.xlsx","NTest_Cardio3.xlsx","NTest_Cardio4.xlsx",
             "NTest_Cardio5.xlsx","NTest_Cardio6.xlsx","NTest_Cardio7.xlsx","NTest_Cardio8.xlsx",
             "NTest_Cardio9.xlsx","NTest_Cardio10.xlsx") #ข้อมูลโรคหัวใจและหลอดเลือด

for (i in 1:10) {
  print(i)
  train <- read_excel(trainfile[i])
  test <- read_excel(testfile[i])
  set.seed(1)
  train$class <- factor(train$class, levels = c(1,0))
  train$gender <- factor(train$gender, levels = c("women","men"))
  train$smoke <- factor(train$smoke, levels = c("No","Yes"))
  train$alco <- factor(train$alco, levels = c("No","Yes"))
  train$active <- factor(train$active, levels = c("No","Yes"))
  test$class <- factor(test$class, levels = c(1,0))
  test$gender <- factor(test$gender, levels = c("women","men"))
  test$smoke <- factor(test$smoke, levels = c("No","Yes"))
  test$alco <- factor(test$alco, levels = c("No","Yes"))
  test$active <- factor(test$active, levels = c("No","Yes"))
  #----- ตัวแบบเทคนิคป่าสุ่ม -----
  model_RF <- randomForest(class ~ ., data = train, mtry = 4, ntree = 250, importance = TRUE)
  #----- ผลการวิเคราะห์ -----
  ypred <- predict(model_RF, newdata = train)
  Train_Result <- confusionMatrix(table(train$class,ypred))
  print(Train_Result)
  ypred <- predict(model_RF, newdata = test)
  Test_Result <- confusionMatrix(table(test$class,ypred))
  print(Test_Result)
}

```


การสร้างตัวแบบการจำแนกด้วยเทคนิคซ์พอร์ตเวกเตอร์แมชชีน

```

library(e1071)
library(readxl)
library(caret)
trainfile <- c("NTrain_Cardio1.xlsx","NTrain_Cardio2.xlsx","NTrain_Cardio3.xlsx","NTrain_Cardio4.xlsx",
              "NTrain_Cardio5.xlsx","NTrain_Cardio6.xlsx","NTrain_Cardio7.xlsx","NTrain_Cardio8.xlsx",
              "NTrain_Cardio9.xlsx", "NTrain_Cardio10.xlsx") #ข้อมูลโรคหัวใจและหลอดเลือด
testfile <- c("NTest_Cardio1.xlsx","NTest_Cardio2.xlsx","NTest_Cardio3.xlsx","NTest_Cardio4.xlsx",
             "NTest_Cardio5.xlsx","NTest_Cardio6.xlsx","NTest_Cardio7.xlsx","NTest_Cardio8.xlsx",
             "NTest_Cardio9.xlsx","NTest_Cardio10.xlsx") #ข้อมูลโรคหัวใจและหลอดเลือด

for (i in 1:10) {
  print(i)
  train <- read_excel(trainfile[i])
  test <- read_excel(testfile[i])
  set.seed(1)
  train$class <- factor(train$class, levels = c(1,0))
  train$gender <- factor(train$gender, levels = c("women","men"))
  train$smoke <- factor(train$smoke, levels = c("No","Yes"))
  train$alco <- factor(train$alco, levels = c("No","Yes"))
  train$active <- factor(train$active, levels = c("No","Yes"))
  test$class <- factor(test$class, levels = c(1,0))
  test$gender <- factor(test$gender, levels = c("women","men"))
  test$smoke <- factor(test$smoke, levels = c("No","Yes"))
  test$alco <- factor(test$alco, levels = c("No","Yes"))
  test$active <- factor(test$active, levels = c("No","Yes"))
  #----- ตัวแบบเทคนิคซ์พอร์ตเวกเตอร์แมชชีน -----
  model_svm <- svm(class ~ ., data = train, kernel = "radial", gamma = 0.001, cost = 1)
  #----- ผลการวิเคราะห์ -----
  ypred <- predict(model_svm, newdata = train)
  Train_Result <- confusionMatrix(table(train$class,ypred))
  print(Train_Result)
  ypred <- predict(model_svm, newdata = test)
  Test_Result <- confusionMatrix(table(test$class,ypred))
  print(Test_Result)
}

```

บรรณานุกรม



บรรณานุกรม

- Akash, N. H. (2020). What is the Decision tree and how it work?. Retrieved January 23, 2023, from <https://medium.com/@bantzviper21st/what-is-decision-tree-and-how-it-works-2637178915fc>
- Ali, N. M., Aziz, N. A. & Besar, R. (2020). Comparison of microarray breast cancer classification using support vector machine and logistic regression with LASSO and Boruta feature selection. *Indonesian Journal of Electrical Engineering and Computer Science*, 20(2), 712-719.
- Aroef, C., Rivan, Y. & Rustam, Z. (2020). Comparing random forest and support vector machines for breast cancer classification. *TELKOMNIKA Telecommunication Computing Electronics and Control*, 18(2), 815-821.
- Assari, R., Azimi, P. & Taghva, M. R. (2017). Heart Disease Diagnosis Using Data Mining Techniques. *International Journal of Economics & Management Sciences*, 6(3), 72-79.
- Astuti, B. S. F., Fridausanti, N. A., & Purnami, S. W. (2018). Model Evaluation for Logistic Regression and Support Vector Machines in Diabetes Problem. *INFERENSI*, 1(2), 77-82. <http://dx.doi.org/10.12962/j27213862.v1i2.6728>
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5-32.
- Duan, M. (2018). Short-Time Prediction of Traffic Flow Based on PSO Optimized SVM. *International Conference on Intelligent Transportation, big data & smart city (ICITBS)*, Xiamen, 41-45.
- Dissanayake, K., & Johar, M. G. (2021). Comparative Study on Heart Disease Prediction Using Feature Selection Techniques on Classification Algorithms. *Applied Computational Intelligence and Soft Computing*. <https://doi.org/10.1155/2021/5581806>
- Dritsas, E., & Trigka, M. (2022). Data-Driven Machine-Learning Methods for Diabetes Risk Prediction. *Sensors* 2022, 22, 5304. <https://doi.org/10.3390/s22145304>
- Firdausanti, N. A., Ningrum, R. A., & Qomariyah, S. (2022). Comparisons of Logistic Regression and Support Vector Machine in Classification of Echocardiogram

- Dataset. *INFERENSI*, 5(2), 85-90. <http://dx.doi.org/10.12962/j27213862.v5i2.14121>
- Gholami, R., & Fakhari, N. (2017). Support Vector Machine: Principle, Parameters, and Applications, 515-535.
- Haberman, S. J. (1978). *Analysis of Qualitative Data*. New York.
- Hair, J.F., William C., Babin, B.J., & Anderson, R.E. (2010). *Multivariate Data Analysis*. 7th ed. NJ: Pearson
- Han, J., Pei, J., & Kamber, M. (2011). *Data mining Concepts and Techniques*. 3rd ed. Elsevier.
- Laudon, J.P. & Laudon, K.C. (2001). *Essentials of Management Information Systems: Organization and Technology in the Enterprise*. 4th ed. New Jersey: Prentice Hall.
- Liao, S. H., Chu, P. H., & Hsiao, P. Y. (2012). Data mining techniques and applications – A decade review from 2000 to 2011. *Expert Systems with Applications*, 39(12), 11303-11311.
- Nai-arun, N. & Moungrmai, R. (2020). Diagnostic Prediction Models for Cardiovascular Disease Risk using Data Mining Techniques. *ECTI Transactions on Computer and Information Technology (ECIT-CIT)*, 14(2), 113-121.
- Olivera, A. R., Roesler, V., Lochpe, C., Schmidt, M. I., Vigo, A., Barreto, S. M., & Duncan, B. B. (2017). Comparison of machine-learning algorithms to build a predictive model for detecting undiagnosed diabetes – ELSA-Brasil: accuracy study. *Sao Paulo Medical Journal*, 135(03), 234-246.
<https://doi.org/10.1590/1516-3180.2016.0309010217>
- Quinlan, J. R. (1986). Induction of Decision Trees. *Machine Learning*, 1(1), 81-106.
- Quinlan, J. R. (1993). *C4.5: programs for machine learning*: Morgan Kaufmann Publishers Inc.
- Rebala, G., Ravi, A., & Churiwala, S. (2019). *An Introduction to Machine Learning*. Cham: Springer.
- Vapnik, V. N. (1995). *The nature of statistical learning. theory*: Springer – Verlag New York, Inc.

- กัลยา วานิชย์บัญชา. (2552). การวิเคราะห์ข้อมูลหลายตัวแปร (พิมพ์ครั้งที่ 4). กรุงเทพฯ: ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย.
- ณัฐวดี หงส์บุญมี และประภาสศิริ ตรีพานิชกุล. (2562). การเปรียบเทียบประสิทธิภาพการจำแนกข้อมูลเพื่อวิเคราะห์ปัจจัยความเสี่ยงที่ส่งผลต่อการเกิดโรคไฮโปไทรอยด์ด้วยเทคนิคเหมืองข้อมูล. *วารสารวิทยาศาสตร์และเทคโนโลยี*, 9(1), 41-51.
- ทรงศิริ แต่สมบัติ. (2548). *การวิเคราะห์การถดถอย* (พิมพ์ครั้งที่ 3). กรุงเทพฯ: สำนักพิมพ์มหาวิทยาลัยเกษตรศาสตร์.
- นงเยาว์ ในอรุณ. (2564). การเปรียบเทียบประสิทธิภาพของแบบจำลองการทำนายความเสี่ยงโรคหัวใจและหลอดเลือดโดยใช้อัลกอริทึมเหมืองข้อมูล. *วารสารวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยมหาสารคาม*, 40(2), 137-147.
- นิภา นิรุตติกุล. (2558). *การพยากรณ์การขาย* (พิมพ์ครั้งที่ 7). กรุงเทพฯ: สำนักพิมพ์มหาวิทยาลัยเกษตรศาสตร์.
- ปพนน์ศรณ สิวส์แดงเดช. (2565). *การจำแนกผู้ป่วยเบาหวานโดยใช้เทคนิคการโหวตรวมกรณีศึกษา : โรงพยาบาลศูนย์อุดรธานี*. วิทยานิพนธ์ปริญญาโทมหาบัณฑิต, มหาวิทยาลัยมหาสารคาม, มหาสารคาม.
- ปริญญา สงวนสัตย์. (2562). *Artificial intelligence with machine learning : AI สร้างได้ด้วยแมชชีนเลิร์นนิ่ง* (พิมพ์ครั้งที่ 1). นนทบุรี: บริษัท ไอดีซี พรีเมียร์ จำกัด.
- รุ่งโรจน์ บุญมา และนิเวศ จิระวิจิตชัย. (2562). การจำแนกผู้ป่วยโรคเบาหวานโดยใช้เทคนิคเหมืองข้อมูลและการเลือกคุณลักษณะจากความสัมพันธ์ของข้อมูล. *วารสารวิชาการชาชนันท์เทคโนโลยี*, 5(2), 11-18.
- รัชนิวรรณ ไพศาลรเกียรติ. (2564). *การเปรียบเทียบตัวแบบการถดถอยลอจิสติกและเทคนิคเหมืองข้อมูลสำหรับพยากรณ์การเป็นโรคเบาหวาน* (วิทยานิพนธ์ปริญญาโทมหาบัณฑิต, มหาวิทยาลัยนเรศวร). NU Intellectual Repository.
<https://nuir.lib.nu.ac.th/dspace/handle/123456789/3943>
- วิชญ์พงศ์ ธรรมธรรม. (2561). เจาะลึก Random Forest. สืบค้น 15 พฤศจิกายน 2565, จาก <https://medium.com/@witchapongdaroontham>
- วริทธิ์พล แสงทองรัตนโชติ. (2565). *การเปรียบเทียบเทคนิคการเรียนรู้ของเครื่องเพื่อสร้างตัวแบบการจำแนกด้วยการปรับปรุงชุดข้อมูลสมดุล* (วิทยานิพนธ์ปริญญาโทมหาบัณฑิต, มหาวิทยาลัยนเรศวร). NU Intellectual Repository.
<https://nuir.lib.nu.ac.th/dspace/handle/123456789/5332>

- ศิริชัย พงษ์วิชัย. (2549). *การวิเคราะห์ข้อมูลทางสถิติด้วยคอมพิวเตอร์* (พิมพ์ครั้งที่ 16). กรุงเทพฯ: จุฬาลงกรณ์มหาวิทยาลัย.
- สุรวัชร์ ศรีเปารยะ และสายชล สิ้นสมบุญ. (2560). การเปรียบเทียบประสิทธิภาพวิธีการจำแนกกลุ่ม การเป็นโรคไตเรื้อรัง : กรณีศึกษาโรงพยาบาลแห่งหนึ่งในประเทศอินเดีย. *วารสารวิทยาศาสตร์และเทคโนโลยี*, 25(5), 839–853.
- สายชล สิ้นสมบุญทอง. (2553). *การวิเคราะห์เชิงสถิติ* (พิมพ์ครั้งที่ 1). กรุงเทพฯ: จามจุรีโปรดักท์.
- อุกฤษฏ์ ศรีสุข และจारी ทองคำ. (2563). การเปรียบเทียบประสิทธิภาพของเทคนิคเหมืองข้อมูล สำหรับพยากรณ์การเกิดโรค. *วารสารวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยมหาสารคาม*, 40(2), 157-163.
- เอกรัฐ บุญเชียง. (2561). การแบ่งกลุ่มและการแบ่งประเภทข้อมูล. เชียงใหม่: สยามพิมพ์นานาชาติ.
- เอกสิทธิ์ พัทธวงศ์ศักดิ์. (2563). *A Little Book of Big Data and Machine Learning* (พิมพ์ครั้งที่ 2). นนทบุรี: บริษัท ไอดีซี พรีเมียร์ จำกัด.

