การประยุกต์ใช้เทคนิคแบบรวมเพื่อเพิ่มประสิทธิภาพของแบบจำลองตามกฎ

วิทยานิพนธ์

ของ

ปพิชญา กลางนอก

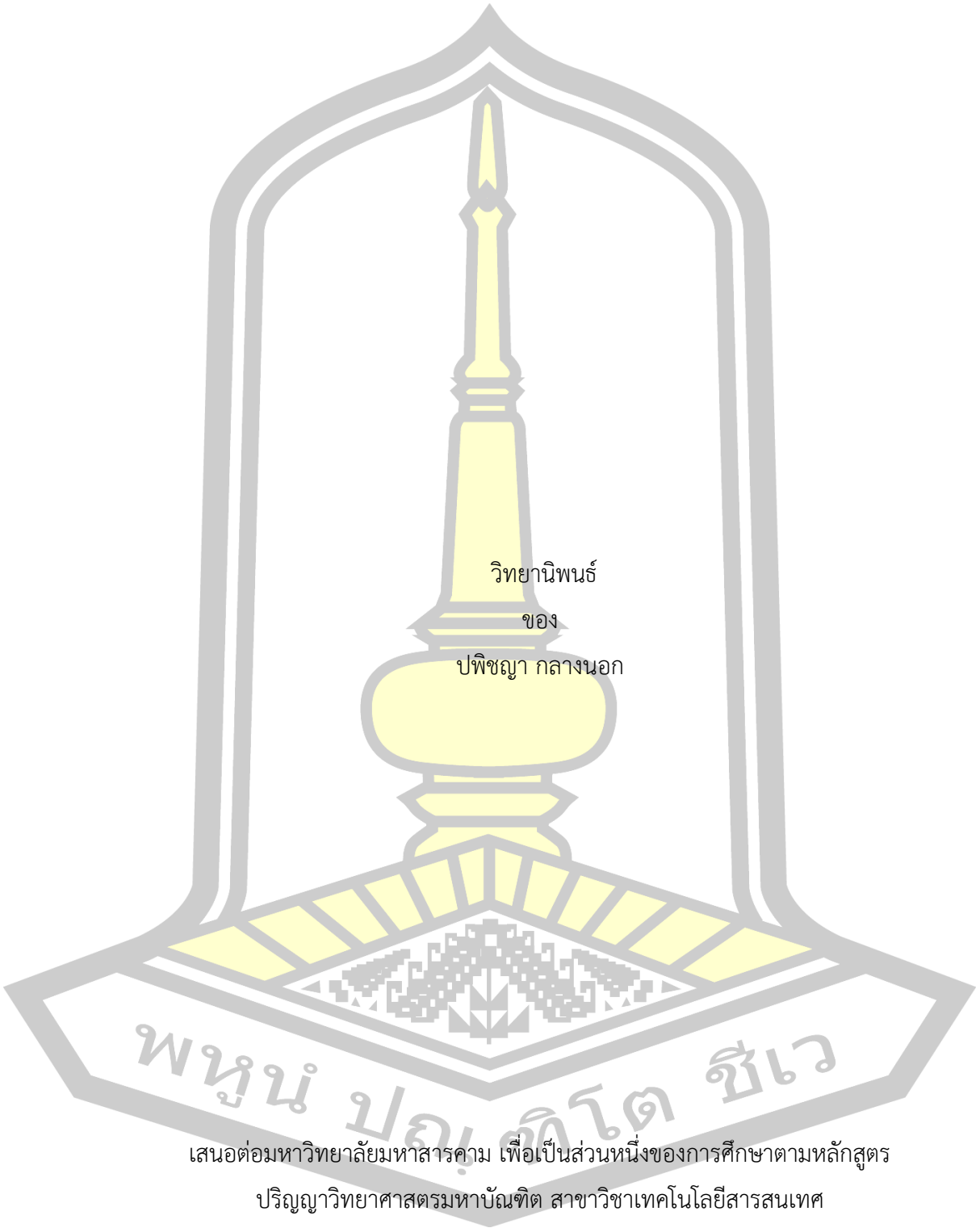
เสนอต่อมหาวิทยาลัยมหาสารคาม เพื่อเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ

กันยายน 2561

สงวนลิขสิทธิ์เป็นของมหาวิทยาลัยมหาสารคาม

การประยุกต์ใช้เทคนิคแบบรวมเพื่อเพิ่มประสิทธิภาพของแบบจำลองตามกฎ



วิทยานิพนธ์

ของ

ปพิชญา กลางนอก

เสนอต่อมหาวิทยาลัยมหาสารคาม เพื่อเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

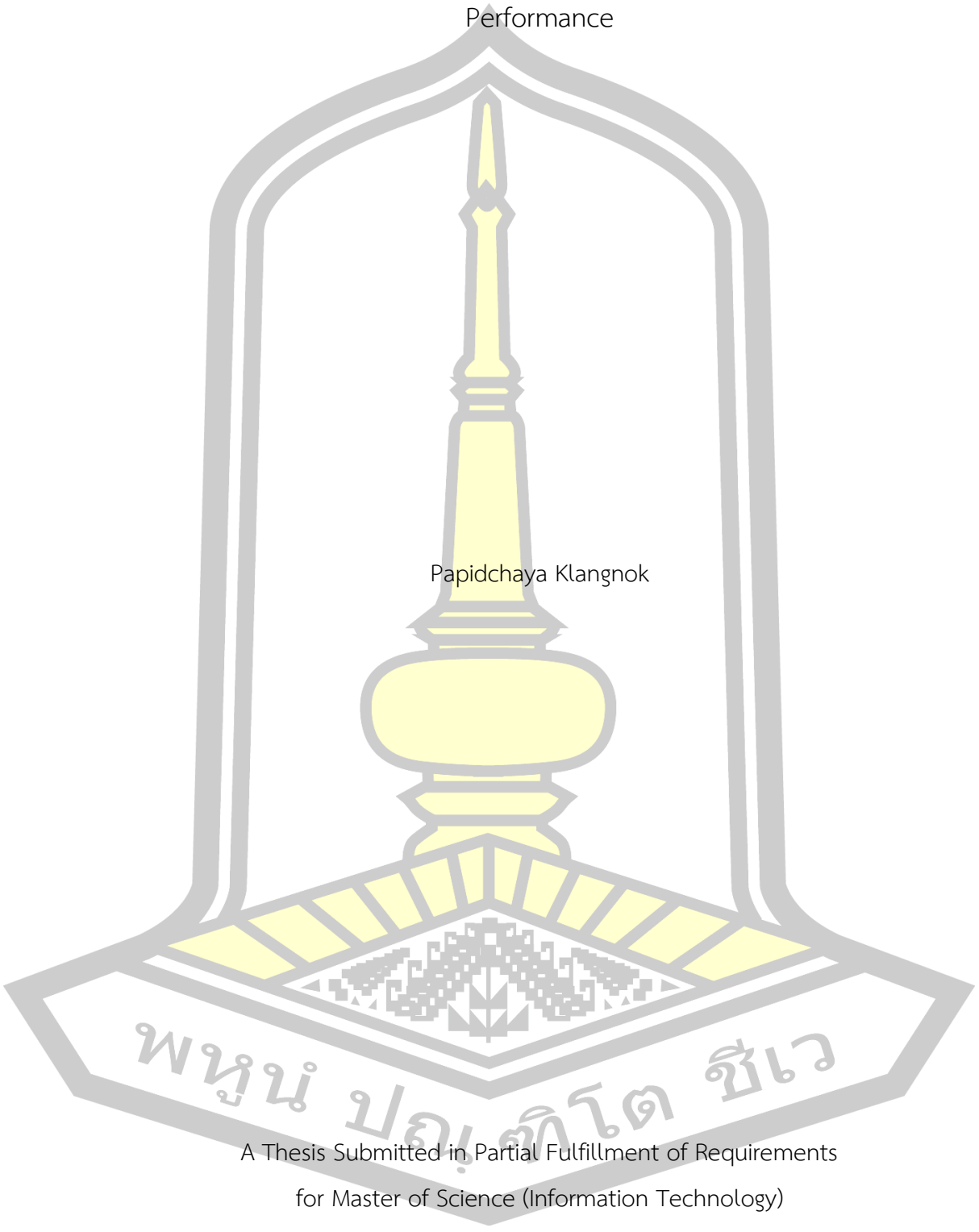
ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ

กันยายน 2561

สงวนลิขสิทธิ์เป็นของมหาวิทยาลัยมหาสารคาม

Applying the Ensemble Technique for Improving Rule-based Models Performance

Papidchaya Klangnok



A Thesis Submitted in Partial Fulfillment of Requirements

for Master of Science (Information Technology)

September 2018

Copyright of Mahasarakham University

คณะกรรมการสอบวิทยานิพนธ์ ได้พิจารณาวิทยานิพนธ์ของนางสาวปพิชญา กลางนอก แล้วเห็นสมควรรับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญา วิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ ของมหาวิทยาลัยมหาสารคาม

คณะกรรมการสอบวิทยานิพนธ์

.....ประธานกรรมการ
(ผศ. ดร. แกมกาญจน์ สมประเสริฐศรี)

.....อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(ผศ. ดร. จารี ทองคำ)

.....กรรมการ
(อ. ดร. สาธิต แสงประดิษฐ์)

.....กรรมการผู้ทรงคุณวุฒิภายนอก
(รศ. ดร. สิทธิชัย บุษหมั่น)

มหาวิทยาลัยอนุมัติให้รับวิทยานิพนธ์ฉบับนี้ เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญา วิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ ของมหาวิทยาลัยมหาสารคาม

.....
(ผศ. ดร. สุจิน บุตรดีสุวรรณ)
คณบดีคณะวิทยาการสารสนเทศ

.....
(ผศ. ดร. กริสน์ ชัยมูล)
คณบดีบัณฑิตวิทยาลัย

วัน.....เดือน.....ปี.....

ชื่อเรื่อง การประยุกต์ใช้เทคนิคแบบรวมเพื่อเพิ่มประสิทธิภาพของแบบจำลองตามกฎ
ผู้วิจัย ปพิชญา กลางนอก
อาจารย์ที่ปรึกษา ผู้ช่วยศาสตราจารย์ ดร. จารี ทองคำ
ปริญญา วิทยาศาสตรมหาบัณฑิต **สาขาวิชา** เทคโนโลยีสารสนเทศ
มหาวิทยาลัย มหาวิทยาลัยมหาสารคาม **ปีการศึกษา** 2561

บทคัดย่อ

การสร้างแบบจำลองเพื่อการพยากรณ์ตามกฎให้มีประสิทธิภาพสูง เป็นงานวิจัยที่ท้าท้าย งานวิจัยนี้จึงมีวัตถุประสงค์ เพื่อสร้างแบบจำลองที่มีประสิทธิภาพในการพยากรณ์โดยใช้แบบจำลองพื้นฐาน คือ FURIA, MODLEM และ RIPPER และเทคนิคแบบรวมที่เป็นที่นิยม คือ Bagging และWeighted Instances Handler Wrapper (WI) โดยใช้ข้อมูลผู้ป่วยโรคมะเร็งเต้านม จำนวน 699 คน และข้อมูลผู้ป่วยโรคเบาหวาน จำนวน 768 คน การวัดประสิทธิภาพแบบจำลองได้ทดลองแบ่งชุดข้อมูลออกเป็นชุดข้อมูลเรียนรู้และชุดข้อมูลทดสอบด้วยการใช้หลักการ10-fold cross validation และ ทำการทดลอง 10 รอบ เพื่อลดความลำเอียงของการทดลอง ในการวัดประสิทธิภาพการพยากรณ์ของแบบจำลองที่สร้างจากแต่ละเทคนิคด้วยค่า Sensitivity, Specificity และ Accuracy จากการศึกษาพบว่า เทคนิค Bagging สามารถเพิ่มค่า Accuracy ในการพยากรณ์การเกิดโรคมะเร็งเต้านมได้สูงถึง 4.91%

คำสำคัญ : แบบจำลองตามกฎ, เหมืองข้อมูล, เทคนิคแบบรวม

พหุน ปณ ฑิโต ชีเว

TITLE Applying the Ensemble Technique for Improving Rule-based Models Performance

AUTHOR Papidchaya Klangnok

ADVISORS Assistant Professor Jaree Thongkam , Ph.D.

DEGREE Master of Science **MAJOR** Information Technology

UNIVERSITY Mahasarakham University **YEAR** 2018

ABSTRACT

Modeling for high-performance forecasting is a challenging research. This research aims to enhance the performance of basic models including FURIA, MODLEM and RIPPER with popular integration techniques, including Bagging and Weighted Instances Handler Wrapper (WI). Data were collected from 699 breast cancer patients and 768 diabetic patients. In order to evaluate the prediction model, 10-fold cross validation was applied to divide dataset into training and testing sets. 10 experiments were conducted to reduce the bias of the experiment. Sensitivity, specificity and accuracy were used to measure the predictive performance of the model generated by each technique. Based on the study, it was found that bagging can increase the accuracy of breast cancer prediction by 4.91%.

Keyword : Rule-based Models, Data Mining, Ensemble Technique

พหุน ปณ ฑิโต ชีเว

กิตติกรรมประกาศ

งานวิจัยฉบับนี้สำเร็จลงได้ด้วยดี เนื่องจากได้รับความกรุณาอย่างสูงจาก ผศ.ดร.จารี ทองคำ อาจารย์ที่ปรึกษางานวิจัย ที่กรุณาให้คำแนะนำตลอดจนปรับปรุงแก้ไขข้อบกพร่องต่างๆ ด้วยความเอาใจใส่อย่างดียิ่ง ผู้วิจัยตระหนักถึงความตั้งใจจริงและความทุ่มเทของอาจารย์ และขอกราบขอบพระคุณเป็นอย่างสูงไว้ ณ ที่นี้

ผู้วิจัยขอกราบขอบพระคุณ ผศ.ดร.แกมกาญจน์ สมประเสริฐศรี รศ.ดร.สิทธิชัย บุษหมั่น และอาจารย์ ดร.สาธิต แสงประดิษฐ์ ที่กรุณาเป็นกรรมการสอบวิทยานิพนธ์และให้คำปรึกษาตลอดจนคำแนะนำต่างๆ รวมถึงทุกท่านที่มีส่วนช่วยให้งานวิจัยนี้สำเร็จลงได้ด้วยดี ผู้วิจัยขอขอบคุณเป็นอย่างสูงไว้ ณ โอกาส นี้

อนึ่ง ผู้วิจัยหวังว่า งานวิจัยฉบับนี้จะเป็นประโยชน์อยู่ไม่น้อย จึงขอมอบส่วนดีทั้งหมดนี้ ให้แก่เหล่าคณาจารย์ ที่ได้ประสิทธิประสาทวิชาจนทำให้ผลงานวิจัยเป็นประโยชน์ต่อผู้ที่เกี่ยวข้อง

และขอมอบความกตัญญูกตเวทิตาคุณ แด่บิดา มารดา และผู้มีพระคุณทุกท่าน สำหรับข้อบกพร่องต่างๆ ที่ อาจจะเกิดขึ้นนั้น ผู้วิจัยขอน้อมรับผิดเพียงผู้เดียว และยินดีที่จะรับฟังคำแนะนำจากทุกท่านที่ได้เข้ามาศึกษา เพื่อประโยชน์ในการพัฒนางานวิจัยนี้ต่อไป

ปพิชญา กลางนอก

พหูน ปณฺฑิโต ชีเว

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ง

บทคัดย่อภาษาอังกฤษ ..... จ

กิตติกรรมประกาศ ..... ฉ

สารบัญ ..... ช

สารบัญตาราง ..... ญ

สารบัญภาพ ..... ญ

บทที่ 1 บทนำ 1

1.1 ความเป็นมาและความสำคัญของปัญหา 1

1.2 ความมุ่งหมายของการวิจัย 2

1.3 ความสำคัญของการศึกษา 2

1.4 ขอบเขตการศึกษา 3

1.5 นิยามคำศัพท์เฉพาะ 3

บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง 4

2.1 ทฤษฎีที่เกี่ยวข้อง 4

2.1.1 โรคมะเร็งเต้านม 4

2.1.2 โรคเบาหวาน 4

2.1.3 เหมืองข้อมูล 4

2.1.4 เทคนิคการพยากรณ์ตามกฎในเหมืองข้อมูล 5

2.1.5 การวัดประสิทธิภาพแบบจำลอง 16

2.2 งานวิจัยที่เกี่ยวข้อง 18

บทที่ 3 วิธีการดำเนินการวิจัย 20

3.1 การเตรียมข้อมูล 20
3.2 กระบวนการก่อนการสร้างแบบจำลอง 21
3.3 การสร้างแบบจำลอง 21
3.3.1 เทคนิค FURIA 22
3.3.2 เทคนิค MODLEM 22
3.3.3 เทคนิค RIPPER 23
3.3.4 เทคนิค Bagging 23
3.3.5 เทคนิค Weighted Instances Handler Wrapper (WI) 24
3.4 การวัดประสิทธิภาพของแบบจำลอง 24
บทที่ 4 ผลการศึกษา 25
4.1 ผลการเปรียบเทียบประสิทธิภาพของแบบจำลองตามกฎ 25
4.2 แบบจำลองเพื่อการพยากรณ์ Bagging และ FURIA 29
บทที่ 5 สรุป วิจารณ์ และข้อเสนอแนะ 31
5.1 สรุป 31
5.2 วิจารณ์ 31
5.3 ข้อเสนอแนะ 32
บรรณานุกรม 33
ภาคผนวก 38
ประวัติผู้เขียน 42

พหุน ปณ ฑิโต ชีเว

สารบัญตาราง

ตารางที่ 2.1 ตารางการตัดสินใจที่ผสมหัวข้อ 7
ตารางที่ 2.2 ลักษณะของตารางการตัดสินใจ 8
ตารางที่ 3.1 ตัวแปรโรคมะเร็งเต้านม 20
ตารางที่ 3.2 ตัวแปรโรคเบาหวาน 21
ตารางที่ 4.1 การเปรียบเทียบ ค่า Sensitivity ของแบบจำลองในการพยากรณ์การเกิดโรคมะเร็ง... 25
ตารางที่ 4.2 การเปรียบเทียบ ค่า Specificity ของแบบจำลองในการพยากรณ์การเกิดโรคมะเร็ง .. 27
ตารางที่ 4.3 การเปรียบเทียบ ค่า Accuracy ของแบบจำลองในการพยากรณ์การเกิดโรคมะเร็ง 28
ตารางที่ 4.4 แบบจำลองตามกฎของผู้ป่วยโรคมะเร็งเต้านมและผู้ป่วยโรคเบาหวาน 29

พหูน ปณฺฑิโต ชีเว

สารบัญภาพ

ภาพประกอบที่ 2.1 แผนภาพการตัดสินใจ 9
ภาพประกอบที่ 2.2 การกำหนดคอลัมน์ 9
ภาพประกอบที่ 2.3 การใส่เงื่อนไข และการกระทำ 10
ภาพประกอบที่ 2.4 การแสดงเงื่อนไข 10
ภาพประกอบที่ 2.5 การลดเงื่อนไขที่ซ้ำซ้อน 11
ภาพประกอบที่ 2.6 เงื่อนไขการแสดงกฎ RIPPER 12
ภาพประกอบที่ 2.7 เงื่อนไขการแสดงกฎ FURIA 13
ภาพประกอบที่ 2.8 เงื่อนไขการแสดงกฎ MODLEM 14
ภาพประกอบที่ 2.9 การทำงานของเทคนิค Bagging [24] 15
ภาพประกอบที่ 2.10 การแบ่งชุดเรียนรู้และชุดทดสอบ 16
ภาพประกอบที่ 2.11 คอนฟิวชันเมตริกซ์ 17

พหูน ปณ ฑิโต ชีเว

บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

การสร้างแบบจำลองกฎการพยากรณ์ที่มีประสิทธิภาพในการพยากรณ์ ซึ่งเป็นงานวิจัยที่ท้าทายในเหมืองข้อมูล เช่น ปวีณา ชัยวนารมย์ [1] ได้ค้นหาแบบจำลองที่มีประสิทธิภาพในการพยากรณ์ความเครียดด้วยเทคนิคการ Bayesian Network, Naïve Bayesian, C4.5, Decision Table, Partial Rules (PART) และ Multilayer Perceptron (MLP) จากการทดสอบแบบจำลองทั้งหมด จะเห็นว่า การสร้างแบบจำลองจากอัลกอริทึม Bayesian Network อัลกอริทึม Naïve Bayesian อัลกอริทึม Decision Tree:4.5 อัลกอริทึม Decision Table และ อัลกอริทึม Partial Rules พบว่าอัลกอริทึม Bayesian Network ดีที่สุด ส่วนเสถียร วิชาเรือง [2] ได้ใช้ 3 เทคนิคในการทำเหมืองข้อมูลในกระบวนการอนุมัติสินเชื่อ คือ C4.5, Decision Rules และ Decision Table ผลการทดลองพบว่า Decision Rules ให้ผลลัพธ์ที่มีความถูกต้องมากที่สุดเพียง 72.70% ส่วนเทคนิค FURIA เป็นเทคนิคที่ใช้หลักการสุ่มกลุ่มตัวอย่างในการสร้างกฎสัมพันธ์ที่มีความแม่นยำสูง และเทคนิค MODLEM [3] เป็น เทคนิคที่ใช้ในการสร้างกฎความสัมพันธ์ โดยกฎที่สร้างขึ้นเกิดจากการประมาณค่าระดับชั้นเป็นกฎจากการประมาณบนในแง่ของทฤษฎีคร่าว ๆ ข้อมูลที่ใช้สามารถใช้ข้อมูลชนิดตัวเลขจากแอตทริบิวต์ที่มีอยู่แล้ว ซึ่งทำให้ชุดกฎที่สร้างจากเทคนิคนี้ จะมีความง่ายและแม่นยำมากขึ้น

Rule-Based [4] เป็นการสร้างกฎเพื่อนำมาใช้ในการพยากรณ์ ซึ่งกฎนั้นได้มาจากการประมวลผลของการทำเหมืองข้อมูลเพื่อเป็นค้นหากฎความสัมพันธ์ ซึ่ง Rule – Based นั้นมีทั้งข้อดีและข้อเสีย โดยมีข้อดี คือ ง่ายต่อการแปลผล และมีข้อเสียคือ มีประสิทธิภาพในการพยากรณ์ที่ไม่สูงนัก ดังนั้นจึงต้องนำเอาข้อดีของสองเทคนิคมารวมกัน คือ ระหว่างเทคนิคการทำเหมืองข้อมูลเพื่อค้นหาความสัมพันธ์ และเทคนิคการจำแนกประเภทข้อมูล ซึ่งกฎที่ได้ผลลัพธ์จำนวนมากทำให้ยากในการนำไปใช้จำแนกข้อมูล เพื่อเป็นการลดจำนวนกฎความสัมพันธ์และสร้างกฎความสัมพันธ์ให้สามารถจำแนกประเภทข้อมูลได้อย่างมีประสิทธิภาพนั้นจะต้องนำเอาข้อดีของเทคนิคการจำแนกประเภทข้อมูลมาช่วยในการลดกฎ

การทำเหมืองข้อมูล (Data Mining) เป็นกระบวนการค้นหาความรู้ที่มีอยู่ในฐานข้อมูลขนาดใหญ่ โดยนำเทคโนโลยีเข้ามาช่วยในการเปลี่ยนแปลงข้อมูลเหล่านั้น เพื่อนำข้อมูลที่ได้ไปประยุกต์ใช้ในงานด้านต่างๆได้ ซึ่งองค์กรต่างๆ ไม่ค่อยให้ความสำคัญและมักจะมองไม่เห็นถึงประโยชน์ของข้อมูลที่เก็บไว้ จึงมีนักวิจัยหลายท่านได้นำกระบวนของการทำเหมืองข้อมูลมาใช้ในการสร้างแบบจำลองเพื่อใช้ใน

การพยากรณ์ เช่น กิตติศักดิ์ สุมามาลย์ [5] ได้ยกตัวอย่างการวิเคราะห์ข้อมูลจากแบบสำรวจสภาวะสุขภาพประชาชนในการใช้ข้อมูลของประชาชนในเขตเทศบาลเมืองจังหวัดบุรีรัมย์ ปี พ.ศ. 2555 ซึ่งในการศึกษาการคัดกรองสุขภาพเบื้องต้นของประชาชน ผลการทดลองพบว่า อัลกอริทึม Partial Rules ให้ความถูกต้องมากกว่าอัลกอริทึมต้นไม้ตัดสินใจ C4.5 โดยมีค่าความถูกต้อง (Precision) เท่ากับ 88.60% ค่าระลึก (Recall) เท่ากับ 89.20% และค่าความเหวี่ยง (F-measure) เท่ากับ 88.80%

แต่อย่างไรก็ตามเทคนิคกฎการพยากรณ์ยังมีจุดด้อยเรื่องประสิทธิภาพในการพยากรณ์ต่ำ ซึ่งในปัจจุบันได้มีนักวิจัยได้นำเสนอการรวมเทคนิค เช่น Basterrech และ Mesa [6] ได้ใช้เทคนิคถุงจำแนกในการลดความผิดพลาดในการพยากรณ์โดยใช้เทคนิคนี้ร่วมกับเทคนิคโครงข่ายประสาทเทียม ปรากฏว่าได้ผลเป็นอย่างดี และ Karasuyama, Harada, Sugiyama และ Takeuchi [7] นำเอา support vector machine (SVM) เพื่อเพิ่มประสิทธิภาพในการวิเคราะห์ข้อมูลและการสร้างแบบจำลอง โดยนำเทคนิค weighted instance เพื่อพยากรณ์การจำแนกแบบจำลอง พบว่าให้ประสิทธิภาพค่าความถูกต้องสูงสุด

ดังนั้นในงานวิจัยนี้ผู้วิจัยมีแนวคิดในการนำเอาเทคนิคแบบรวม รวมถึง Bagging และ Weighted Instances Handler Wrapper (WI) มาใช้ในการเพิ่มประสิทธิภาพของแบบจำลองตามกฎ การพยากรณ์การเกิดมะเร็งเต้านม และการเกิดโรคเบาหวาน ซึ่งเป็นโรคที่นิยมเป็นกันมากทั่วโลก รวมถึง FURIA, MODLEM และ RIPPER ใช้ 10 รอบของ 10-fold cross validation ในการแบ่งกลุ่มข้อมูลเป็นชุดข้อมูลเรียนรู้ และชุดข้อมูลทดสอบ เพื่อลดความลำเอียงของผลการทดลอง คณะผู้วิจัยจึงได้ทำการทดลอง 10 ครั้งของ 10-fold cross validation และวัดประสิทธิภาพการพยากรณ์ของแบบจำลองตามกฎด้วยค่า Sensitivity, Specificity และ Accuracy

1.2 ความมุ่งหมายของการวิจัย

เพื่อศึกษาเทคนิคแบบรวมเพื่อเพิ่มประสิทธิภาพของแบบจำลองตามกฎ

1.3 ความสำคัญของการศึกษา

ได้แบบจำลองตามกฎจากเทคนิคแบบรวมที่มีประสิทธิภาพมากขึ้น

1.4 ขอบเขตการศึกษา

การศึกษาครั้งนี้มุ่งเน้นการเพิ่มประสิทธิภาพของแบบจำลองกฎการพยากรณ์ข้อมูลผู้ป่วยโรคมะเร็งเต้านมและข้อมูลผู้ป่วยโรคเบาหวาน ด้วยการเทคนิคแบบรวม โดยใช้ข้อมูลจากฐานข้อมูล UCI ที่มีมาตรฐาน

1. ข้อมูลผู้ป่วยโรคมะเร็งเต้านม เป็นข้อมูลผู้ป่วยจากศูนย์การแพทย์มหาวิทยาลัย สถาบันมะเร็งวิทยา ลูบลิยานา ยูโกสลาเวีย

2. ข้อมูลผู้ป่วยโรคเบาหวาน เป็นผู้ป่วยจากการดูแลรักษาทางคลินิก 10 ปี (1999-2008) ที่โรงพยาบาลในสหรัฐฯ 130 แห่งและเครือข่ายการจัดส่งแบบรวม ประกอบด้วยคุณลักษณะมากกว่า 50 รายการที่แสดงถึงผลลัพธ์ของผู้ป่วยและโรงพยาบาล

1.5 นิยามคำศัพท์เฉพาะ

1. เทคนิคแบบรวม หมายถึง เทคนิค Bagging และ Weighted Instances Handler Wrapper (WI) ที่ใช้ในการรวมกันกับเทคนิคพื้นฐานเพื่อเพิ่มประสิทธิภาพของแบบจำลองแบบ

2. การพยากรณ์ หมายถึง การคาดการณ์การเกิดโรคมะเร็งเต้านมและโรคเบาหวาน

พหุน ปณฺฑิโต ชีเว

บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

การศึกษาครั้งนี้ผู้ศึกษาได้ศึกษาค้นคว้าทฤษฎี ตำรา และงานวิจัยที่เกี่ยวข้องกับการการประยุกต์ใช้เทคนิคถุงจำแนกเพื่อพัฒนาประสิทธิภาพของแบบจำลองตามกฎ ผู้ศึกษาได้เสนอทฤษฎีและงานวิจัยที่เกี่ยวข้อง ดังต่อไปนี้

2.1 ทฤษฎีที่เกี่ยวข้อง

2.1.1 โรคมะเร็งเต้านม

โรคมะเร็งเต้านม คือ มะเร็งที่เกิดจากความผิดปกติของเซลที่ซึ่งอยู่ต่อมน้ำนมหรือท่อน้ำนม ซึ่งเซลเหล่านี้มีการการแบ่งตัวผิดปกติทำให้ไม่สามารถควบคุมได้ แล้วจะแพร่กระจายไปตามทางเดินทางน้ำเหลืองทำให้ไปสู่อวัยวะที่อยู่ใกล้เคียง และเมื่อเซลมะเร็งมีจำนวนมากขึ้นเรื่อยๆ ก็จะทำให้แย่งสารอาหารแล้วไปปล่อยสารบางอย่างซึ่งเป็นอันตรายนำพาไปสู่การทำลายอวัยวะต่างๆ ในที่สุดก็ทำให้ผู้ป่วยเสียชีวิต [8] โดยส่วนใหญ่การวิเคราะห์การเกิดโรคแพทย์จะทำการวินิจฉัยจากตัวแปร ต่างๆ เช่นความหนาของเซลล์ ความสม่ำเสมอของขนาดเซลล์ ความสม่ำเสมอของรูปร่างเซลล์ การยึดเกาะด้านนอก ขนาดเซลล์เยื่อบุเดี่ยว นิวเคลียสเปลือย ไบร์ทโครมาติน ภาวะปกติในการถ่ายทอดทางกรรมพันธุ์ กระบวนการแบ่งเซลล์ และอื่นๆ เป็นต้น

2.1.2 โรคเบาหวาน

โรคเบาหวาน คือ โรคที่เกิดจากความผิดปกติจากเมตาบอลิซึม โดยมีอาการระดับน้ำตาลในเลือดสูง จึงเป็นผลทำให้เกิดความผิดปกติของการหลั่งอินซูลิน (Insulin) และการออกฤทธิ์ของอินซูลิน ในภาวะที่ระดับน้ำตาลในเลือดสูงอย่างต่อเนื่อง จึงทำให้เกิดการเสื่อมของอวัยวะในร่างกายยาว ทำให้เกิดการเสียหน้าที่และอวัยวะสำคัญจำนวนหลายอวัยวะมีผลต่อการทำงานล้มเหลว คือ ตา ไต หัวใจ ระบบประสาทและหลอดเลือด [9] โดยส่วนใหญ่จะทำการวิเคราะห์การเกิดโรคจากตัวแปรต่างๆ เช่น ความดันโลหิต ความหนาของผิวพับ ดัชนีมวลกาย กรรมพันธุ์ เป็นต้น

2.1.3 เหมืองข้อมูล

การทำเหมืองข้อมูล (Data Mining) [10] เป็นวิธีการที่ใช้ทำกับข้อมูลที่มีจำนวนมากเพื่อใช้ในการค้นหารูปแบบและความสัมพันธ์ที่มองไม่เห็นอยู่ในชุดข้อมูลนั้นๆ และเป็นกระบวนการที่ใช้ในการจัดเก็บและสื่อความหมายของข้อมูลเดิมที่จัดเก็บง่ายๆ เพื่อนำไปสู่การจัดเก็บข้อมูลที่อยู่ในแบบฐานข้อมูล ซึ่งสามารถดึงข้อมูลสารสนเทศเพื่อใช้ในการทำเหมืองข้อมูลที่ค้นหาความรู้ที่ซ่อนอยู่ในข้อมูล พบว่าในการปัจจุบันการทำเหมืองข้อมูลสามารถนำไปนำไปประยุกต์ใช้ได้กับงานเป็นจำนวน

มาก ทั้งในทางธุรกิจที่ใช้สำหรับช่วยในการตัดใจของผู้บริหาร ในด้านวิทยาศาสตร์ และการแพทย์ ด้านเศรษฐกิจ และสังคม

การทำเหมืองข้อมูล มีขั้นตอนหลัก 4 ขั้นตอน [11] คือ

ขั้นตอนการทำเหมืองข้อมูล

1. การเตรียมข้อมูล คือ การศึกษาและรวบรวมข้อมูล ข้อมูลผู้ป่วยโรคมะเร็งเต้านม และข้อมูลผู้ป่วยโรคเบาหวาน

2. การทำกระบวนการก่อนการสร้างแบบจำลอง ประกอบด้วย 1. การแปลงข้อมูล คือ ข้อมูลที่ได้จากการเก็บรวบรวมข้อมูลผู้ป่วยโรคมะเร็งเต้านม และข้อมูลโรคเบาหวาน 2. การวิเคราะห์ปัจจัย คือ เป็นการคัดเลือกแอตทริบิวต์ที่สามารถนำไปเป็นตัวแทนของกลุ่มแอตทริบิวต์สำหรับใช้ลดจำนวนแอตทริบิวต์ในการพยากรณ์

3. การสร้างแบบจำลอง เป็นการสร้างแบบจำลองเพื่อวิเคราะห์ข้อมูลที่ได้จากการแปลงข้อมูลพร้อมทดสอบผลลัพธ์แบบจำลองเพื่อให้ได้ค่าคำตอบที่ดีที่สุด

4. การวัดประสิทธิภาพแบบจำลอง คือ การแบ่งกลุ่มข้อมูลออกเป็นกลุ่มข้อมูล

2.1.4 เทคนิคการพยากรณ์ตามกฎในเหมืองข้อมูล

เทคนิคการพยากรณ์ตามกฎในเหมืองข้อมูล (Rule Based Technique) เป็นการพยากรณ์ข้อมูลด้วยกฎนั้นเป็นโมเดลที่แสดงผลด้วยเซตของกฎที่มีลักษณะแบบ ‘IF-THEN’ ซึ่งกฎหนึ่งๆ จะถูกแสดงอยู่ในรูปฟอร์ม ดังนี้

IF condition THEN conclusion

ในงานวิจัยฉบับนี้ ผู้วิจัยได้ทำการศึกษาค้นคว้างานวิจัยที่เกี่ยวข้องจึงได้เทคนิคที่มีประสิทธิภาพในการสร้างแบบจำลองตามกฎดังนี้

Juneja และ Amar [12] ได้นำเอา Decision table มาแก้ปัญหางบประมาณระหว่างองค์กรของบริษัท ซึ่งข้อมูลที่นำมาใช้ในการเคราะห์ข้อมูลงบการเงินตั้งแต่ปี 1991-2015 ผลปรากฏว่า Decision table ใช้งานง่ายและลดต้นทุน

เสถียร วิชาเรือง [2] การประยุกต์ใช้เทคนิคการทำเหมืองข้อมูลในกระบวนการอนุมัติสินเชื่อ สำหรับการนำความรู้ที่ได้จากการทำเหมืองข้อมูลมาเป็นช่วยในการสนับสนุนการตัดสินใจการอนุมัติสินเชื่อ สำหรับการศึกษาครั้งนี้ได้ทดลองจำแนกกลุ่มข้อมูลด้วยเทคนิคจำนวน 3 เทคนิค ในการให้ผลลัพธ์เป็นกฎ คือ 1. Decision Tree: C4.5 2. Decision Tree: Partial Rules 3. vDecision Table ซึ่งสำหรับผลการทดลองจะพบว่าการจำแนกกลุ่มแบบ Decision Tree: Partial Rules นั้นจะให้ผลลัพธ์ที่มีความถูกต้องมากที่สุดร้อยละ 72.7 และได้กฎทั้งสิ้นจำนวน 346 กฎ

Zhang และ Zhang [13] ได้ศึกษาพบว่ากฎที่ได้เป็นผลลัพธ์มีจำนวนมากและยากต่อการนำไปใช้ในการจำแนกข้อมูล จึงต้องนำข้อดีของเทคนิคการจำแนกประเภทข้อมูลเข้ามาช่วยลดจำนวนกฎความสัมพันธ์และสร้างกฎความสัมพันธ์ ให้สามารถใช้ในการจำแนกประเภทข้อมูลได้ดังตัวอย่างข้อแตกต่างระหว่างกฎความสัมพันธ์และกฎจำแนกประเภทข้อมูล

Yu, Ni, Dan และ Xu [14] ได้นำเอา Decision Rule มาใช้กับข้อมูลที่มีคลาสที่ไม่สมดุล ข้อมูลที่ใช้ คือ cancer gene expression datasets ผลปรากฏว่าการทดลองได้น้อยกฎแต่เป็นกฎที่เข้าใจง่าย

Gil-Herrera และคณะ [15] ได้ทำการศึกษาเทคนิค MODLEM ในการแนะนำบ้านพักคนชรา พบว่าประสิทธิภาพของ MODLEM เมื่อวัดด้วยค่า Sensitivity เท่ากับ 64.47% และค่า Specificity เท่ากับ 68.62% ส่วน Thakur Meenakshi และ Priya [16] ได้วัดประสิทธิภาพ ชุดข้อมูลการฝึกอบรมของ URL โดยใช้เทคนิค RIPPER พบว่ามีค่าความถูกต้อง 82%

Lotte, Lecuyer และ Arnaldi [17] ได้นำเทคนิค FURIA ในการจำแนกคุณลักษณะเฉพาะของสมองคอมพิวเตอร์ โดยใช้เทคนิค FURIA พบว่ามีความแม่นยำสูงสุดที่ 86.00%

นิภาพร ชนะมาร และพรรณี สิทธิเดช [18] ได้ทำการวิเคราะห์ปัจจัยการเรียนรู้ด้วยการคัดเลือกคุณสมบัติและการพยากรณ์ผลสัมฤทธิ์ทางการเรียนของนิสิต จำนวน 180 ระเบียน โดยใช้เทคนิค BPNN และเทคนิค SVMs ผลการวิเคราะห์ปัจจัยการเรียนรู้พบว่า มีความผิดพลาดอยู่ในระดับต่ำ ดังนั้นจึงได้ทดลองใช้เทคนิคการรวมกลุ่มด้วยวิธี Bagging ร่วมกับ BPNN และ SVMs พบว่า Bagging ร่วมกับ BPNN มีค่าความผิดพลาดอยู่ในระดับต่ำสุด (RMSE = 0.1051) เหมาะสำหรับใช้ในการพยากรณ์ได้อย่างมีประสิทธิภาพ

Karasuyama, Harada, Sugiyama และ Takeuchi [7] นำเอา support vector machine (SVM) ในการวิเคราะห์ในการวิเคราะห์ข้อมูลและการสร้างแบบจำลอง โดยนำเทคนิค weighted instance ในการเพิ่มประสิทธิภาพแบบจำลอง พบว่า ให้ประสิทธิภาพสูงสุด

2.1.4.1 เทคนิค Decision Table

เทคนิค Decision Table [19] เป็นการจัดกลุ่มในตารางการตัดสินใจโดยการนำผลที่ได้จากการเรียนรู้จากข้อมูลที่นำเข้ามาที่มีลักษณะของข้อมูลที่ใกล้เคียงกันให้มาอยู่ในกลุ่มเดียวกัน โดยแอตทริบิวต์ตัวสุดท้ายของตารางจะใช้กำหนดกลุ่มที่ต้องการแบ่งดังตัวอย่าง ดังตารางที่ 2.1

ตารางที่ 2.1 ตารางการตัดสินใจที่ผสมหัวข้อ

การสั่งสินค้า	**1**	**2**	**3**	**4**	**5**	**6**
ราคาสินค้า	-	ต่ำกว่า 100 บาท	100–1,000 บาท	มากกว่า 1,000 บาท	ต่ำกว่า 100 บาท	มากกว่า 1,000 บาท
จำนวนที่สั่ง	ต่ำกว่า 5 ชิ้น	50–100 ชิ้น	50–100 ชิ้น	50–100 ชิ้น	เกินกว่า 100 ชิ้น	เกินกว่า 100 ชิ้น
สั่งทันที		x				
สั่งตามกำหนดเวลา	x		x			
รอผู้บริหารอนุมัติ				x	x	
ประกวดราคา						x

จากตารางที่ 2.1 แสดงถึงการตัดสินใจที่ผสมหัวข้อ (Mixed-Entry Form) เป็นตารางการตัดสินใจที่ผสมระหว่างการจำกัดหัวข้อ และไม่จำกัดหัวข้อในการตรวจสอบเงื่อนไขและการกระทำกิจกรรม ซึ่งมีนักวิจัยหลายท่านได้นำเทคนิคนี้ไปใช้ เช่น Juneja และ Amar [12] ได้นำเอา Decision table มาแก้ปัญหางบประมาณระหว่างองค์กรของบริษัท ซึ่งข้อมูลที่นำมาใช้ในการเคราะห์ข้อมูลงบการเงินตั้งแต่ปี 1991-2015 ผลปรากฏว่า Decision table ใช้งานง่ายและลดต้นทุน ส่วนสายชล สินสมบูรณ์ทอง [20] เทคนิค Decision Table เป็นการกำหนดวิธีสำหรับการเรียนรู้ข้อมูลให้เป็นรูแบบชุดฝึกอบรม หลังจากนั้นก็เริ่มเข้าสู่การวิเคราะห์ผลด้วยการทำนายที่ได้จากค่าความถูกต้องในการทำนายคำตอบที่แท้จริงของจำนวนแถว และอัตราส่วนของความน่าจะเป็นที่ค่าว่าจะเป็น หรือดูว่าข้อมูลนั้นมีความสอดคล้องกันหรือไม่ และ Confusion Matrix จะเป็นการวัดประสิทธิภาพของโมเดลประเภทข้อมูลเหล่านั้น โดยจะอาจจะอาศัย Confusion Matrix สำหรับการเก็บข้อมูลจำนวนแถวที่จำแนกที่ได้จากกลุ่มข้อมูลจริง และจากกลุ่มข้อมูลการทำนาย แต่ส่วน Goyal และ Modi [21] ได้นำการทำเหมืองข้อมูลที่เป็นกระบวนการสำหรับการวิเคราะห์ข้อมูลที่มีมุมมองที่แตกต่างและเป็นข้อมูลที่ใช้เป็นประโยชน์ สำหรับการจำแนกบทความนี้จะประกอบไปด้วยวิธีการจัดหมวดหมู่ตารางการตัดสินใจที่เพิ่มขึ้นในการจัดประเภทชุดข้อมูลในการ ช้อปปิ้งออนไลน์ซึ่งสำหรับการจัดหมวดหมู่ชุดข้อมูล Decision Table เป็นวิธีที่ได้ผลลัพธ์ที่ดีที่สุดสำหรับการจัดหมวดหมู่ชุดข้อมูล

2.1.4.2 เทคนิค Decision Table Naive Bayes (DTNB)

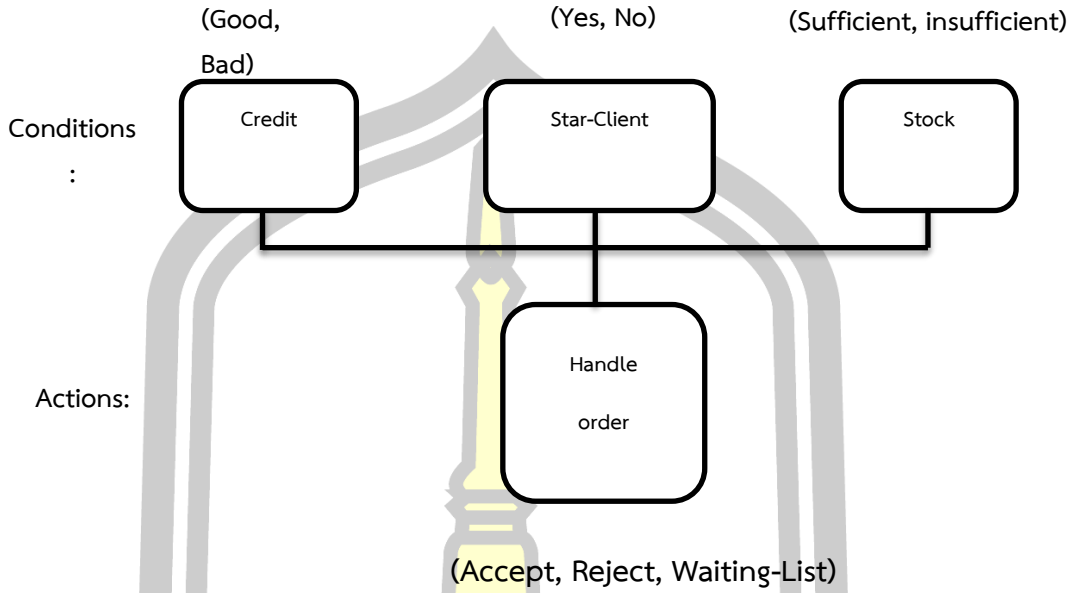
เทคนิค DTNB [22] เป็นการจัดกลุ่มโดยให้ผลลัพธ์ที่ได้จะอยู่ในรูปของกฎที่นำเอาข้อมูลแต่ละแอทริบิวต์มาเป็นเงื่อนไขในการตัดสินใจลักษณะของข้อมูลที่ใกล้เคียงกันให้มาอยู่ในกลุ่มเดียวกันโดยค่าของข้อมูลของแอทริบิวต์ที่มีค่าต่อเนื่องจะใช้เครื่องหมาย “<”, “<=”, “=>”, “>” ในการพิจารณา ส่วนข้อมูลที่เป็นค่าไม่ต่อเนื่องจะใช้เครื่องหมาย “=” และใช้เงื่อนไข “AND” ในการเชื่อมโยงระหว่างแอทริบิวต์ จะเชื่อมโยงไปจนถึงแอทริบิวต์สุดท้ายที่จะแบ่งกลุ่มได้ ดังตารางที่ 2.2

ตารางที่ 2.2 ลักษณะของตารางการตัดสินใจ

เงื่อนไข	กฎสำหรับการตัดสินใจ/การกระทำ
การระบุเงื่อนไขสำหรับการพิจารณาการทำงาน	กฎที่เป็นไปได้ภายใต้เงื่อนไขที่ระบุ
การกระทำที่เป็นไปได้	การระบุการเลือกการกระทำภายใต้กฎเกณฑ์

ตัวอย่างเช่น บริษัท ABC ต้องการที่จะหาวิธีจัดการสั่งซื้อของลูกค้าโดยการสั่งซื้อทั้งหมดของลูกค้าที่ไม่ใช่ลูกค้าประเภทลูกค้าชั้นดี และลูกค้าที่ผิดสัญญาจะถูกปฏิเสธ order (reject) ถ้าสินค้าคงคลังมีเพียงพอจะยอมรับการสั่งซื้อนอกเหนือจากนั้นจะให้รอ (waiting-list) โดยเราเขียนแผนภาพการตัดสินใจ (Decision Diagram) ได้ดังภาพประกอบที่ 1.1

พหุน ปณ ฑิโต ชีเว



ภาพประกอบที่ 2.1 แผนภาพการตัดสินใจ

จำนวนของ rules จะเท่ากับ จำนวนของ values (cond1) x ... x จำนวนของ values (condN) ซึ่งนั่นก็คือ 2 x 2 x 2 = 8 rules จากนั้นกำหนดตารางให้สอดคล้องกับกฎ ดังตัวอย่าง ถ้ามีกฎ 8 กฎ จะต้องมีการสร้างตารางทั้งหมด 8 คอลัมน์ โดยมี Conditions และ Actions ดังภาพประกอบที่ 2.1 และใส่เงื่อนไข และ Actions ต่างๆลงไป ดังภาพประกอบที่ 2.2

	Rule 1	Rule 2	Rule 3	Rule 4	Rule 5	Rule 6	Rule 7	Rule 8
conditions: Credit								
conditions: Star-Clients								
conditions: Stock								
Actions: Order Handling								

ภาพประกอบที่ 2.2 การกำหนดคอลัมน์

		Rule1	Rule2	Rule 3	Rule4	Rule5	Rule6	Rule7	Rule8
conditions	Credit	Good	Good	Good	Good	Bad	Bad	Bad	Bad
	Star-Clients	Yes	Yes	No	No	Yes	Yes	No	No
	Stock	Suf	Insuf	Suf	Insuf	Suf	Insuf	Suf	Insuf
Actions	Order Handling	Accept	Wait	Accept	Wait	Accept	Wait	Reject	Reject

ภาพประกอบที่ 2.3 การใส่เงื่อนไข และการกระทำ

จากตารางที่ได้มานั้นจะเห็นได้ว่ามีขนาดใหญ่เกินไป และมีอีกหลายกรณีที่เงื่อนไขบางเงื่อนไขไม่มีความหมาย ซ้ำซ้อน และให้ผลลัพธ์ออกมาที่เหมือนกัน ดังตัวอย่างเช่น ถ้ามี credit ดี และมีเงื่อนไขของ stock ก็จะทำให้เพียงพอต่อการที่จะเกิดการกระทำ (Actions) แล้ว client นั้นจะไม่ส่งผลต่อการตัดสินใจ ดังภาพประกอบที่ 2.4

		Rule1	Rule2	Rule3	Rule4	Rule5	Rule6	Rule7	Rule8
conditions	Credit	Good	Good	Good	Good	Bad	Bad	Bad	Bad
	Star-Clients	Yes	Yes	No	No	Yes	Yes	No	No
	Stock	Suf	Insuf	Suf	Insuf	Suf	Insuf	Suf	Insuf
Actions	Order Handling	Accept	Wait	Accept	Wait	Accept	Wait	Reject	Reject

ภาพประกอบที่ 2.4 การแสดงเงื่อนไข

โดยเราจะทำตารางให้กะทัดรัด ด้วยการลดเงื่อนไขที่ซ้ำซ้อนกันลง คือเงื่อนไขที่ไม่ได้ใช้เราจะเขียนแทนด้วยเครื่องหมาย " - " ในแถวตั้งที่ไม่จำเป็นนั้นเราจะยุบ column ให้เหลือเพียง 1 column โดยถ้าเมื่อยุบแล้วจะได้ในรูปแบบ ดังภาพประกอบที่ 2.5

conditions	Credit	Good		Bad		Bad
	Star-Clients	-		Yes		No
	Stock	Suf	Insuf	Suf	Insuf	-
Actions	Order Handling	Accept	Wait	Accept	Wait	Reject

ภาพประกอบที่ 2.5 การลดเงื่อนไขที่ซ้ำซ้อน

จาก Decision Table นี้ สรุปได้ดังนี้

R1: If credit = good And Stock = suf Then Handling = accept

R2: If credit = good And Stock = insuf Then Handling = wait

R3: If credit = bad And Clients = yes And Stock = suf Then Handling = accept

R4: If credit = bad And Clients = yes And Stock = insuf Then Handling = wait

R5: If credit = bad And Clients = no Then Handling = reject

เป็นการสร้างและใช้ตารางการตัดสินใจในลักษณะผสมแบบ Naïve Bayes ซึ่งแต่ละจุดในการค้นหาอัลกอริทึมจะประเมินคุณค่าของการแบ่งแอตทริบิวต์เป็นสองส่วนที่ไม่เหมือนกัน คือ สำหรับตารางการตัดสินใจและสำหรับ Naïve Bayes ที่ใช้การค้นหาการเลือกแบบไปข้างหน้าโดยที่ในแต่ละขั้นตอน แอตทริบิวต์ที่เลือกจะถูกจำลองโดย Naïve Bayes และส่วนที่เหลือโดยตารางตัดสินใจและแอตทริบิวต์ทั้งหมดจะถูกจำลอง โดยตารางการตัดสินใจในขั้นต้น ในแต่ละขั้นตอนอัลกอริทึมจะพิจารณาลดแอตทริบิวต์ทั้งหมดจากโมเดล ซึ่งมีนักวิจัยหลายท่านได้นำเทคนิคนี้ไปใช้ เช่น Yu, Ni, Dan และ Xu [14] ได้นำเอา Decision Rule มาใช้กับข้อมูลที่มีคลาส ที่ไม่สมดุล ข้อมูลที่ใช้ คือ cancer gene expression datasets ผลปรากฏว่าการทดลองได้น้อยกฎแต่เป็นกฎที่เข้าใจง่าย

2.1.4.3 เทคนิค RIPPER

เทคนิค RIPPER ถูกสร้างโดย Cohen [23] ในปี 1995 ได้พัฒนามาจากอัลกอริทึม IREP ซึ่งใช้หลักการ (Growing and Pruning) คือ การแยกกฎการจำแนกประเภทข้อมูลให้เป็นไปตามกลุ่มหมวดหมู่ เป้าหมาย สำหรับใช้หลักการตัดกิ่งในการเลือกกฎที่สามารถทำให้เกิดผลในการทำงานมากกว่าอัตราความผิดพลาด (Error Rate) ซึ่งได้กำหนดสำหรับการใช้ในการทำนาย

ข้อมูล ริปเปอร์ ประกอบด้วย 2 เฟส คือ เฟสที่ 1 ระบุกฎเริ่มต้น เฟสที่ 2 ระบุค่า porst – process rule optimization โดยการเรียนรู้จากข้อมูลที่ได้กำหนดคลาสไว้เรียบร้อยแล้ว (Traning data) ซึ่งแบ่งเป็น Growing set และ pruning set เช่นนั้นอัลกอริธึมจะสร้างกฎความสัมพันธ์ใน greedy fashion แต่ส่วนริปเปอร์นั้น จะหาค่าที่ดีที่สุด growing set ใน rule space อธิบายได้จาก BNF จากที่ได้ Growing set แล้วจะทำการ pruning ข้อมูลทันที หลังจาดนั้นก็จะได้กลุ่มตัวอย่างที่เหมือนกันออกมาที่รวมถึงกฎของ training set หลังจากนั้นก็จะลบทิ้ง ซึ่ง Training data ที่ยังคงเหลือจะถูกแบ่งใหม่อีกครั้ง หลังจากที่ทำการเรียนรู้ตามกฎแล้ว สำหรับเป็นการช่วยแก้ปัญหาที่เกิดมาจากการแบ่งแยกกลุ่มที่ผิดพลาด ทั้งนี้จะกระทำจนกว่าผลออกมาแล้วเป็นที่พอใจมากที่สุด ตัวอย่างเช่น เงื่อนไขที่ 1 อาศัยอยู่ในน้ำ เท่ากับ ใช่ แสดงสัตว์นั้นเป็นปลา มีขา เท่ากับ ไม่ แสดงว่าสัตว์นั้นเป็นสัตว์เลื้อยคลาน เงื่อนไขที่ 2ให้กำเนิด เท่ากับ ไม่สามารถบินได้ เท่ากับ ไม่ หรือ อาศัยอยู่ในน้ำ เท่ากับ ไม่ แสดงว่าสัตว์นั้นเป็นสัตว์เลื้อยคลาน และเงื่อนไขที่ 3 สามารถบินได้ เท่ากับ ใช่ หรือ ให้กำเนิด เท่ากับ ไม่ แสดงว่าสัตว์นั้นเป็นนก แสดงว่าสัตว์นั้นเป็นสัตว์เลี้ยงลูกด้วยนม ดังภาพประกอบที่ 2.6

(Live in Water=Yes) ⟶ Fishes (Have Legs=No) ⟶ Reptiles
(Give Birth=No, Can Fly=No, Live In Water=No) ⟶ Reptiles
(Can Fly=Yes,Give Birth=No) ⟶ Birds () ⟶ Mammals

ภาพประกอบที่ 2.6 เงื่อนไขการแสดงกฎ RIPPER

ซึ่งมีนักวิจัยหลายท่านได้นำเทคนิคนี้ไปใช้ เช่น เสถียร วิชาเรือง [2] ใช้เทคนิค C4.5, Decision Rules และ Decision Table มาทำการพยากรณ์การให้สินเชื่อแก่ลูกค้า พบว่า Decision Rules ให้ผลลัพธ์ที่มีความถูกต้องมากที่สุดเพียง 72.70% ส่วน Lotte, Lecuyer และ Arnaldi [17] ได้นำเทคนิค FURIA ในการจำแนกคุณลักษณะเฉพาะของสมองคอมพิวเตอร์ โดยใช้เทคนิค FURIA พบว่ามีความแม่นยำสูงสุดที่ 86.00%

2.1.4.4 เทคนิค FURIA

เทคนิค FURIA [24] เป็นเทคนิคมีอัลกอริธึมในการเหนี่ยวนำกฎที่ไม่เรียงลำดับตามฟัซซี่ ซึ่งฟัซซี่ คือ ความคลุมเครือหรือความไม่แน่นอนในการจำแนกประเภทข้อมูลด้วยกฎความสัมพันธ์การจำแนกประเภทข้อมูลด้วยกฎความสัมพันธ์เป็นวิธีการสร้างกฎเพื่อนำมาใช้ในการ

พยากรณ์ข้อมูล ซึ่งกฎทีนำมาสร้าง คือ กฎความสัมพันธ์ที่ได้มาจากการประมวลผลของการทำเหมืองข้อมูลเพื่อค้นหากฎความสัมพันธ์ (Association Rule Mining) หรืออธิบายการจำแนกข้อมูล ด้วยกฎความสัมพันธ์ นั่นคือ การนำข้อดีทั้งสองเทคนิคในทางด้านเหมืองข้อมูลมา รวมกันระหว่างเทคนิคการทำเหมืองข้อมูลเพื่อค้นหาความสัมพันธ์และเทคนิคการจำแนกประเภทข้อมูล ได้โดยการนำเทคนิคการทำเหมืองข้อมูลเพื่อค้นหากฎความสัมพันธ์ เช่น Zhang และ Zhang [13] ได้ศึกษาพบว่ากฎที่ได้เป็นผลลัพธ์มีจำนวนมากและยากต่อการนำไปใช้ในการจำแนกข้อมูล จึงต้องนำข้อดีของเทคนิคการจำแนกประเภทข้อมูลเข้ามาช่วยลดจำนวนกฎความสัมพันธ์และสร้างกฎความสัมพันธ์ ให้สามารถใช้ในการจำแนกประเภทข้อมูลได้ดังตัวอย่างข้อแตกต่างระหว่างกฎความสัมพันธ์และกฎจำแนกประเภทข้อมูล ดังภาพประกอบที่ 2.7

กฎที 1 IF B then D and E
กฎที 2 IF B then Class=Yes

ภาพประกอบที่ 2.7 เงื่อนไขการแสดงกฎ

จากภาพประกอบที่ 2.7 กฎที่ 1 คือกฎความสัมพันธ์ที่ว่า ถ้า B เกิดขึ้นแล้ว D และ E ก็จะเกิดขึ้นด้วย แต่ถ้าเป็นกฎที่ 2 ซึ่งเป็นกฎการจำแนกประเภทข้อมูลว่า ถ้า B เกิดขึ้นแล้วนั้นจะพยากรณ์ได้ว่า จะเป็นจริง เท่ากับ Yes ซึ่งจะเห็นข้อแตกต่างของทั้งสองกฎนี้นั้นก็คือ กฎที่ 1 สามารถมีแอททริบิวต์ (Attribute) ข้างหลัง then (Consequent) ได้มากกว่า 1 แอททริบิวต์ คือ D และ E แต่กฎที่ 2 สามารถมีแอททริบิวต์ข้างหลัง then ได้เพียงหนึ่งแอททริบิวต์ ซึ่งแอททริบิวต์ดังกล่าวจะเป็นแอททริบิวต์ คลาสเป้าหมาย (Class Target) ซึ่งมีนักวิจัยหลายท่านได้นำเทคนิคนี้ไปใช้ เช่น Lotte, Lecuyer และ Arnaldi [17] ได้นำเทคนิค FURIA ในการจำแนกคุณลักษณะเฉพาะของสมองคอมพิวเตอร์ โดยใช้เทคนิค FURIA พบว่ามีความแม่นยำสูงสุดที่ 86.00%

2.1.4.5 เทคนิค MODELM

เทคนิค MODELM [25] เป็นเทคนิคที่ใช้เพื่อสร้างกฎที่กำหนดไว้สำหรับการจำแนกประเภทที่มีประสิทธิภาพและเพิ่มความถูกต้องในการจำแนกประเภทโดยการรวมตัวจำแนกประเภทต่างๆ อัลกอริทึมการเหนี่ยวนำกฎ คือ เทคนิคMODLEM ในกรอบของตัวจำแนกประเภทรวมกันคือเครื่องบรรจุถุง n 2-classifier และ combiner aggregation ผลของการทดลองที่ดำเนินการได้แสดงให้เห็นว่าอัลกอริธึม MODLEM สามารถนำมาใช้อย่างมีประสิทธิภาพภายในกรอบ

ของตัวจำแนกรวมที่พิจารณาที่จะครอบคลุมทุกตัวอย่างที่เป็นบวกของแนวคิดการตัดสินใจที่กำหนดต่อไปนั้นแสดงว่าเป็น B และไม่ครอบคลุมตัวอย่างเชิงลบใด ๆ (เช่น U \ B) หลักขั้นตอนสำหรับการเหนี่ยวนำกฎเริ่มต้นจากการสร้างกฎข้อแรกโดยการเลือกลำดับขั้นตอนที่ดีที่สุดตามเกณฑ์ที่เลือก เมื่อกฎถูกเก็บไว้การเรียนรู้ในเชิงบวกทั้งหมดตัวอย่างที่ตรงกับกฎนี้จะถูกตัดออกจากการพิจารณากระบวนการซ้ำในขณะที่บางตัวอย่างที่ดีของแนวคิดการตัดสินใจยังคงอยู่เปิด จากนั้นขั้นตอนจะถูกทำซ้ำตามลำดับสำหรับแต่ละชุดตัวอย่างจากแนวคิดการตัดสินใจที่ประสบความสำเร็จ ในอัลกอริทึม MODLEM อัลกอริทึมตัวเลขจะถูกจัดการระหว่างกฎ induction ในขณะที่เงื่อนไขเบื้องต้นของกฎจะถูกสร้างขึ้น เงื่อนไขเหล่านี้คือแสดงเป็นอย่างใดอย่างหนึ่ง ดังภาพประกอบที่ 2.8

กฎข้อที่ 1 (a <va) หรือ (a ≥ va)
กฎข้อที่ 2 (a = [v1, v2)) (a <v2) และ (a ≥ v1)

ภาพประกอบที่ 2.8 เงื่อนไขการแสดงกฎ MODLEM

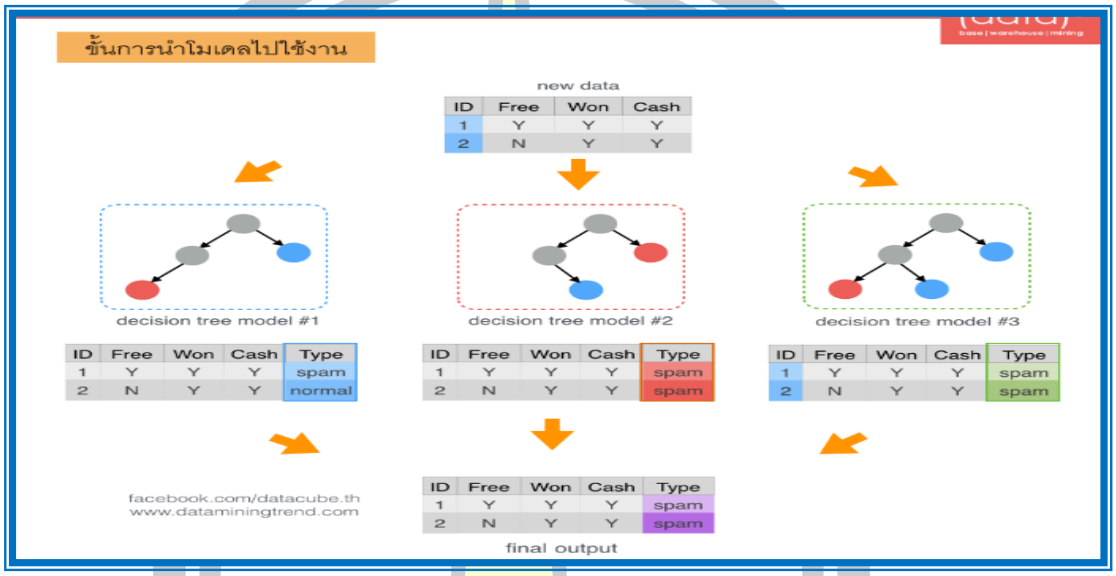
จากภาพประกอบที่ 2.8 กฎข้อที่ 1 คือ ในขณะที่เงื่อนไขเบื้องต้นของกฎจะถูกสร้างขึ้น หรือ ค่าของแอตทริบิวต์คือ va กฎข้อที่ 2 คือ ถ้าเลือกแอตทริบิวต์เดียวกันสองครั้งในขณะสร้างกฎหนึ่งข้อจะได้รับเงื่อนไขที่เกิดจุดตัดสองเงื่อนไข เช่น v1 <v2 สำหรับคุณลักษณะที่ระบุเงื่อนไขเหล่านี้คือ a = va ซึ่งมีนักวิจัยหลายท่านได้นำเทคนิคนี้ไปใช้ เช่น Gil-Herrera และคณะ [15] ได้ทำการศึกษาเทคนิค MODLEM ในการแนะนำบ้านพักคนชรา พบว่าประสิทธิภาพของ MODLEM เมื่อวัดด้วยค่า Sensitivity เท่ากับ 64.47% และค่า Specificity เท่ากับ 68.62% ส่วน Thakur, Meenakshi และPriya [16] ได้วัดประสิทธิภาพ ชุดข้อมูลการฝึกอบรมของ URL โดยใช้เทคนิค RIPPER พบว่ามีค่าความถูกต้อง 82%

2.1.4.6 เทคนิค Ensemble

เทคนิค Ensemble [26] คือเทคนิคแบบรวม เพื่อเพิ่มประสิทธิภาพของแบบจำลองแบบที่สร้างจากเทคนิคเดียวในการจำแนกข้อมูลโดยเทคนิคแบบเดียว (เทคนิคพื้นฐาน) สามารถทำงานร่วมกัน เป็นวิธีการอาศัยตัวจำแนกที่มีมากกว่าหนึ่งตัว โดยแต่ละตัวที่จำแนกนั้นจะมีกระบวนการทำงานของตัวเอง และทุกตัวที่จำแนกจะต้องกระทำกับข้อมูลเดียวกัน ถ้าได้ผลการจำแนกของแต่ละตัวแล้ว จะต้องนำผลเหล่านี้มาผ่านวิธีการรวบรวม Combination Integration หรือ Vote และตัดสินสุดท้าย เพื่อให้ผลที่จะได้ในการจำแนกมีเพียงผลเดียวเท่านั้น เช่น

2.1.4.6.1 Bagging [27] เป็นเทคนิคที่ใช้วิธีการสุ่มข้อมูลการฝึกอบรมโดยวิธี Bootstrap และแบ่งข้อมูลออกเป็นหลายชุด แล้วจึงเอามาสร้างแบบจำลองด้วยเทคนิคพื้นฐาน

แล้วจึงทำการทดสอบหาชุดข้อมูลที่สร้างแบบจำลองแล้วมีค่าการพยากรณ์ต่ำที่สุด แบบจำลองที่สร้างจากเทคนิคพื้นฐานอาจมีข้อผิดพลาดมาก แต่เมื่อมีการรวมตัวกันกับ Bagging แล้วจะสามารถทำให้เกิดข้อผิดพลาดในการพยากรณ์น้อยลง ดังภาพประกอบที่ 2.9



ภาพประกอบที่ 2.9 การทำงานของเทคนิค Bagging [24]

มีนักวิจัยหลายท่านได้ศึกษาเกี่ยวกับ Bagging เช่น นิภาพร ชนะมาร และพรรณี สิทธิเดช [18] ได้ทำการวิเคราะห์ปัจจัยการเรียนรู้ด้วยการคัดเลือกคุณสมบัติและการพยากรณ์ผลสัมฤทธิ์ทางการเรียนของนิสิต จำนวน 180 ระเบียน โดยใช้เทคนิค BPNN และเทคนิค SVMs ผลการวิเคราะห์ปัจจัยการเรียนรู้พบว่า มีความผิดพลาดอยู่ในระดับต่ำ ดังนั้นจึงได้ทดลองใช้เทคนิคการรวมกลุ่มด้วยวิธี Bagging ร่วมกับ BPNN และ SVMs พบว่า Bagging ร่วมกับ BPNN มีค่าความผิดพลาดอยู่ในระดับต่ำสุด (RMSE = 0.1051) เหมาะสำหรับใช้ในการพยากรณ์ได้อย่างมีประสิทธิภาพ

2.1.4.6.2 เทคนิค Weighted Instances Handler Wrapper (WI) เป็นเทคนิคที่ใช้การสุ่มกลุ่มตัวอย่างด้วยน้ำหนัก Golinko และ Zhu [28] นำเอาอัลกอริธึมการเรียนรู้การฝังตัวคุณลักษณะโดยรวม GFEL ซึ่งเรียนรู้คุณลักษณะโดยการจำแนกประเภทและคุณลักษณะในการจัดกลุ่มและจัดหมวดหมู่ โดยใช้เทคนิค weighted instance เพื่อเพิ่มประสิทธิภาพในการพยากรณ์ โดยใช้ข้อมูลชุดทดสอบจำนวน 28 ชุด พบว่า ในการจัดกลุ่มและจัดหมวดหมู่ได้อย่างมีประสิทธิภาพสูงสุด Karasuyama, Harada, Sugiyama และ Takeuchi [7] นำเอา support vector machine

(SVM) ในการวิเคราะห์ในการวิเคราะห์ข้อมูลและการสร้างแบบจำลอง โดยนำเทคนิค weighted instance ในการเพิ่มประสิทธิภาพแบบจำลอง พบว่า ให้ประสิทธิภาพสูงสุด

2.1.5 การวัดประสิทธิภาพแบบจำลอง

การวัดประสิทธิภาพแบบจำลอง ในงานวิจัยนี้ ผู้วิจัยได้ทำการทดลองวัดประสิทธิภาพของแบบจำลอง โดยการแบ่งข้อมูลออกเป็น ชุดข้อมูลสอน และชุดข้อมูลทดสอบ โดยใช้หลักการ 10-Fold cross - validation โดยจะทำการแบ่งข้อมูลออกเป็น 10 ชุดเท่าๆกัน จากนั้นจะทำการทดสอบทั้งหมด 10 รอบ โดยในแต่ละรอบจะใช้ข้อมูล 1 ชุดเป็นชุดทดสอบและอีก 9 ชุดที่เหลือเป็นชุดเรียนรู้ ในรอบต่อไปก็ใช้ชุดข้อมูลถัดไปเป็นชุดทดสอบจนครบทั้ง 10 ชุดข้อมูล โดยค่าที่ใช้ในการแสดงวัดประสิทธิภาพของแบบจำลอง คือได้ Sensitivity, Specificity และ Accuracy ดังภาพประกอบที่ 2.10

ภาพประกอบที่ 2.10 การแบ่งชุดเรียนรู้และชุดทดสอบ

จากภาพประกอบที่ 2.10 การแบ่งชุดเรียนรู้และชุดทดสอบออกเป็นรอบโดยในแต่ละรอบจะใช้ข้อมูล 1 ชุดเป็นชุดทดสอบและอีก 9 ชุดที่เหลือเป็นชุดเรียนรู้ในรอบที่ 2 นำชุดที่ 2 มาทำการทดสอบ ชุดที่ 1 และชุดที่เหลือ นำไปเป็นชุดเรียนรู้ และทำอย่างนี้ไปจนครบรอบ ทั้งหมด 10 รอบ โดยค่าที่ใช้ในการแสดงวัดประสิทธิภาพของแบบจำลอง คือได้ Precision, Recall, F-measure, Sensitivity, Specificity และ Accuracy ดังสมการที่ 1, 2, 3, 4 และ 5 ซึ่งได้จากภาพประกอบที่ 2.11

			Predicted Class	
			เกิดโรค	ไม่เกิดโรค
Actual Outcomes		เกิดโรค	TP	FN
		ไม่เกิดโรค	FP	TN

ภาพประกอบที่ 2.11 คอนฟิวชันเมตริกซ์

โดย True Positive (TP) คือ จำนวนข้อมูลที่พยากรณ์ถูกต้องในคลาสที่เกิดโรค

True Negative (TN) คือ จำนวนข้อมูลที่พยากรณ์ถูกต้องในคลาสที่ไม่เกิดโรค

False Positive (FP) คือ จำนวนข้อมูลที่พยากรณ์ไม่ถูกต้องในคลาสที่เกิดโรค

False Negative (FN) คือ จำนวนข้อมูลที่พยากรณ์ไม่ถูกต้องในคลาสที่ไม่เกิดโรค

ค่าความไว (Sensitivity)

ค่าความไว คือ ค่าของความถูกต้องในการพยากรณ์ของคลาสที่เกิดโรคต่อจำนวนทั้งหมดในกลุ่มของคลาสที่พยากรณ์ว่าเกิดโรค ค่าความไวที่เกิดโรคจริงให้ผลบวกเป็นเปอร์เซ็นต์ ดังสมการที่ 2.1

$$\text{Sensitivity} = \frac{TP}{TP + FN} \qquad (2.1)$$

ค่าความจำเพาะ (Specificity)

ค่าความจำเพาะ คือ ค่าความถูกต้องในการพยากรณ์ของคลาสที่เกิดโรคต่อจำนวนทั้งหมดที่เกิดโรคจริง ค่าความจำเพาะที่ไม่เกิดโรคให้ผลลบเป็นเปอร์เซ็นต์ ดังสมการที่ 2.2

$$\text{Specificity} = \frac{TN}{FP + TN} \qquad (2.2)$$

พหุน ปณ ฑิโต ชีเว

ค่าความแม่นยำ (Accuracy)

ค่าความแม่นยำ คือ ค่าที่เกิดจากการเปรียบเทียบระหว่างค่า ความไว และค่าความจำเพาะ ดังสมการที่ 2.3

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \quad (2.3)$$

จากสมการข้างต้นจะเห็นว่าค่า Recall และ Sensitivity เป็นค่าเดียวกัน ซึ่งค่าที่นำมาใช้แสดงผลของการพยากรณ์แบบ binary class คือ Sensitivity, Specificity และ Accuracy

2.2 งานวิจัยที่เกี่ยวข้อง

Ashraf, Le และ Huang [29] ได้ศึกษาทำการวินิจฉัยโรคมะเร็งเต้านมโดยอัตโนมัติ ซึ่งเป้าหมายหลัก คือ การเตรียมข้อมูลสำหรับกระบวนการทำเหมืองข้อมูล รวมถึงวิธีการเตรียมชุดข้อมูลในอนาคต โดยใช้อัลกอริทึมถ่วงน้ำหนักใกล้ที่สุดในการจำแนกระยะทางวัดค่าความถูกต้องในการจัดหมวดหมู่ของ 0.005 ในชุดข้อมูลที่สร้างขึ้น ซึ่งค่าที่หายไปจำแนกค่าความถูกต้องสูงสุด คือ 0.9698 เมื่อ K= 1

Mumtaz, Sheriff และ Duraiswamy [30] ทำการวิจัยเกี่ยวกับการวินิจฉัยและการรักษาโรคมะเร็ง ซึ่งการวินิจฉัยโรคมะเร็งเต้านมเป็นปัญหาสำหรับแพทย์มากและเป็นเรื่องที่สำคัญที่สุดของโลก โดยงานวิจัยนี้เป็นการประเมินโครงข่ายประสาทเทียม 3 แบบ ได้แก่ 1. Adaptive Resonance Theory Based Network (ART), 2. เครือข่ายประสาทเทียม Neuro Network Based Networks (SOM) และเครือข่าย Neuro network propagation (BPA) โดยใช้ฐานข้อมูลมะเร็งเต้านมของ Wisconsin

Sirisomboonrat และSinapiromsaran [31] ศึกษาทำการวิจัยเกี่ยวกับการวินิจฉัยโรคมะเร็งเต้านมซึ่งสามารถช่วยในการตรวจหาโรคมะเร็งเต้านมในระยะแรก โดยใช้ต้นไม้ตัดสินใจในการจำแนกประเภท ซึ่งได้นำเอาเทคนิคใหม่ๆ มาใช้ โดยใช้ข้อมูลจาก UCI และชุดข้อมูลที่ใช้จะเป็นลักษณะของผู้ป่วยโรคมะเร็งเต้านม ซึ่งผลแสดงให้เห็นว่าอัลกอริธึมต้นไม้ตัดสินใจดีกว่าอัลกอริธึม C4.5

Ghosh, Mondal และ Ghosh [32] ได้ศึกษาวิจัยเกี่ยวกับการสำรวจข้อมูลทางการแพทย์พบว่าโรคมะเร็งเต้านมนั้นเกิดขึ้นทั่วโลกในอัตรา 22.9% ของโรคมะเร็งทั้งหมดในผู้หญิงและอัตรา 13.7% เกิดการเสียชีวิตจากโรคมะเร็งเต้านม ดังนั้นงานวิจัยนี้จึงได้นำเทคนิคในการจำแนกประเภทต่างๆ โดยการใช้เทคนิคการวิเคราะห์องค์ประกอบหลัก (PCA) เพื่อลดขนาดของชุดข้อมูล โดยการวินิจฉัยและวิเคราะห์โรคมะเร็งเต้านมด้วยความช่วยเหลือของผู้จำแนกประเภท ได้แก่ MLP โดยใช้ Backpropagation NN (MLP BPN) และ Support Vector Machine (SVM) และได้วัด

ประสิทธิภาพด้วยค่าความถูกต้อง ค่าความแม่นยำ ค่าRecall ค่าF-Measure และ Kappa Statistic Etc

ณัฐวุฒิ ศรีวิบูล [33] ได้ทำการวิจัยศึกษาการวินิจฉัยการเกิดโรคมะเร็งด้วยการสอบถามประวัติและการทดสอบเลือดหรือของเหลวในร่างกาย โดยการนำเทคนิคการทำเหมืองข้อมูลมาประยุกต์ใช้กับการวิเคราะห์การเกิดโรคมะเร็ง โดยใช้เทคนิคในการเปรียบเทียบอัลกอริธึมการทำเหมืองข้อมูลด้วยเทคนิค 3 เทคนิค คือ C4.5, K-Nearest Neighbor และ Naïve Bayes ผลการเปรียบเทียบพบว่า เทคนิค C4.5 มีประสิทธิภาพสูงสุด 98.63%

Isaac และSureshkumar [34] เป็นการศึกษาวิจัยเกี่ยวกับการป้องกันโรคมะเร็งเต้านมและเพื่อเพิ่มอายุการใช้งานของมนุษย์หลังการเกิดโรคมะเร็งเต้านม โดยการนำเทคนิค K-Means และDecision Tree เพื่อพยากรณ์การป้องกันโรคมะเร็งเต้านมและเพื่อเพิ่มอายุการใช้งานของมนุษย์หลังการเกิดโรคมะเร็งเต้านม ซึ่งได้วัดค่าประสิทธิภาพของแบบจำลองในการพยากรณ์ด้วยค่า specificity, accuracy, และ F-measure พบว่า เทคนิค K-Means มีประสิทธิภาพสูงสุด มีค่า accuracy สูงถึง 98.91%

Islam, Iqbal, Haque และ Hasan [35] ได้ทำการศึกษาวิจัยการการวินิจฉัยข้อมูลการเกิดโรคมะเร็งเต้านม โดยใช้ข้อมูลจาก UCI ซึ่งได้นำเทคนิคในการพยากรณ์การเกิดโรคมะเร็งเต้านม คือ Support Vector Machine และ K-Nearest Neighbors รวมถึงการใช้หลักการ10-fold cross validation ในการแบ่งกลุ่มข้อมูลเป็นชุดข้อมูลอบรม และชุดข้อมูลทดสอบ พบว่ามีค่าความแม่นยำ 98.57% และ91.14% และค่าความจำเพาะ 95.65% และ92.31% ตามลำดับ

Adegoke, Chen, Banissi และ Barikzai [36] เป็นการศึกษาวิจัยเกี่ยวกับการพยากรณ์การรอดชีวิตจากโรคมะเร็งด้วยการทดสอบโดยใช้ AdaBoost ด้วยอัลกอริทึม Random Forest, Radial Basis Function Network และอัลกอริทึม Neural Network เป็นอัลกอริทึมเป็นพื้นฐาน ทำให้การพยากรณ์ยังไม่มีประสิทธิภาพมากนัก จึงได้นำ AdaBoost สามแบบคือ AdaBoostM1, Real AdaBoost และ MultiAdaBoostAB เพื่อเพิ่มประสิทธิภาพในการพยากรณ์การรอดชีวิตจากโรคมะเร็ง พบว่า ROC AdaM1 + RF และ MBAB + ANN มีความแม่นยำในการพยากรณ์สูงสุดเท่ากับ 0.97

NirmalaDevi, Alias Balamurugan และ Swathi [37] ได้ศึกษาวิจัยเกี่ยวกับการพัฒนารูปแบบสำหรับการจำแนกฐานข้อมูลโรคเบาหวานการเกิดโรคเบาหวานเพื่อพยากรณ์การเกิดโรคเบาหวานในอนาคต โดยใช้เทคนิค K-means และ K-Nearest Neighbour (KNN) พบว่า เทคนิค K-Nearest Neighbour มีประสิทธิภาพสูงสุดในการจำแนกการเกิดโรคเบาหวาน ซึ่งมีค่าความถูกต้อง 97.4%

บทที่ 3
วิธีการดำเนินการวิจัย

วิธีการดำเนินการวิจัยในงานวิจัยสำหรับงานวิจัยนี้ได้นำเอาขั้นตอนการทำเหมืองข้อมูลมาใช้เพื่อการประยุกต์ใช้เทคนิคถุงจำแนกเพื่อพัฒนาประสิทธิภาพของแบบจำลองตามกฎ ซึ่งมีขั้นตอนการดำเนินการดังต่อไปนี้ การเตรียมข้อมูล การทำกระบวนการก่อนการสร้างแบบจำลอง การสร้างแบบจำลอง และการวัดประสิทธิภาพแบบจำลอง

3.1 การเตรียมข้อมูล

การศึกษาและรวบรวมข้อมูลของการศึกษาค้นคว้าอิสระครั้งนี้ ได้ทำการเก็บรวบรวมข้อมูลของผู้ป่วยจากเว็บไซต์ UCI (https://archive.ics.uci.edu/ml/datasets.html) ประกอบด้วยข้อมูลผู้ป่วยโรคมะเร็งเต้านม จำนวน 699 คน และข้อมูลผู้ป่วยโรคเบาหวาน จำนวน 768 คน สามารถแสดงตัวแปรได้ดังตารางที่ 4 และตารางที่ 5 ตามลำดับ

ตารางที่ 3.1 ตัวแปรโรคมะเร็งเต้านม

ชื่อตัวแปร	คำอธิบายตัวแปร	ชนิดของตัวแปร
Clump Thickness	ความหนาของเซลล์	Nominal
Uniformity of Cell Size	ความสม่ำเสมอของขนาดเซลล์	Nominal
Uniformity of Cell Shape	ความสม่ำเสมอของรูปร่างเซลล์	Nominal
Marginal Adhesion	การยึดเกาะด้านนอก	Nominal
Single Epithelial Cell Size	ขนาดเซลล์เยื่อบุเดี่ยว	Nominal
Bare Nuclei	นิวเคลียสเปลือย	Nominal
Bland Chromatin	ไบร์ทโครมาติน	Nominal
Normal Nucleoli	ภาวะปกติในการถ่ายทอดทางกรรมพันธุ์	Nominal
Mitoses	กระบวนการแบ่งเซลล์	Nominal
Class	คลาส	Nominal

ตารางที่ 3.2 ตัวแปรโรคเบาหวาน

ชื่อตัวแปร	คำอธิบายตัวแปร	ชนิดของตัวแปร
Number of times pregnant	จำนวนครั้งที่ตั้งครรภ์	Number
Plasma glucose concentration a 2 hours in an oral glucose tolerance test	ความเข้มข้นของกลูโคสในพลาสมาเป็นเวลา 2 ชั่วโมงในการทดสอบความทนทานต่อกลูโคสในช่องปาก	Number
Diastolic blood pressure (mm Hg)	ความดันโลหิต	Number
Triceps skin fold thickness (mm)	ความหนาของผิวพับ	Number
2-Hour serum insulin (mu U/ml)	อินซูลินซีรั่ม 2 ชม.	Number
Body mass index	ดัชนีมวลกาย	Number
Diabetes pedigree function	กรรมพันธุ์โรคเบาหวาน	Number
Age	อายุ	Number
Class	คลาส	Nominal

3.2 กระบวนการก่อนการสร้างแบบจำลอง

ในกระบวนการก่อนการสร้างแบบจำลองในที่นี้ ผู้วิจัยได้ทำการตรวจสอบข้อมูล นำข้อมูลที่ไม่สมบูรณ์ออกจากชุดข้อมูล และทำการวิเคราะห์ตัวแปรต้น (ปัจจัย) เพื่อแสดงถึงความสัมพันธ์ระหว่างตัวแปรต้นและตัวแปรตามต่อไป

3.3 การสร้างแบบจำลอง

การสร้างแบบจำลองการประยุกต์ใช้เทคนิคถุงจำแนกเพื่อพัฒนาประสิทธิภาพของแบบจำลองตามกฎ เพื่อช่วยให้ทราบปัจจัยผลที่เกิดขึ้นจากการเกิดโรคมะเร็งเต้านม และข้อมูลผู้ป่วยโรคเบาหวาน ด้วยตัวแปรทั้ง 10 และ 9 ตัวแปรตามลำดับ ที่ได้ทำการเก็บรวบรวมข้อมูลของผู้ป่วยจากเว็บไซต์ UCI (https://archive.ics.uci.edu/ml/datasets.html) นั้น ด้วยการทำเหมืองข้อมูล

ด้วยโปรแกรม Weka 3.8.2 และเทคนิคที่นำมาใช้ในการสร้างแบบจำลองทั้งหมดมีจำนวน 5 เทคนิค ดังต่อไปนี้

3.3.1 เทคนิค FURIA เป็นเทคนิคที่ใช้หลักการสุ่มกลุ่มตัวอย่างในการสร้างกฎสัมพันธ์ที่มีความแม่นยำสูง ซึ่งในงานวิจัยนี้ได้ใช้พารามิเตอร์ดังนี้

1. บรรทัดฐาน (TNorm) = บรรทัดฐานมาตรฐาน
2. จำนวนของข้อมูลที่ใช้ในการสร้างแบบจำลอง (batchSize) = 100
3. ตรวจสอบอัตราความผิดพลาดหรือไม่ (checkErrorRate) = True
4. debug หรือไม่ (debug) = False
5. ตรวจสอบความสามารถในการจำแนกประเภท (doNotCheckCapabilities) = False
6. จำนวนการแบ่งชุดข้อมูลที่ใช้ในการทดสอบ (folds) = 3
7. น้ำหนักรวมต่ำสุด (minNo) = 2.0
8. จำนวนตำแหน่งทศนิยมที่จะใช้สำหรับผลลัพธ์ของตัวเลขในโมเดล (numDecimalPlaces) = 2
9. จำนวนการเพิ่มประสิทธิภาพ (optimizations) = 2
10. ข้อมูลการสุ่มตัวอย่าง (seed) = 1
11. การใช้กฎ (uncovAction) = การใช้กฎมาตรฐาน

3.3.2 เทคนิค MODLEM เป็นเทคนิคเทคนิคที่ใช้ในการสร้างกฎความสัมพันธ์ โดยกฎที่สร้างขึ้นเกิดจากการประมาณค่าระดับชั้น ซึ่งทำให้ชุดกฎที่สร้างจากเทคนิคนี้ จะมีความง่ายและแม่นยำมากขึ้น ซึ่งในงานวิจัยนี้ได้ใช้พารามิเตอร์ดังนี้

1. จำนวนของข้อมูลที่ใช้ในการสร้างแบบจำลอง (batchSize) = 100
2. วิธีการจำแนกประเภท (classificationStrategy) = m-Estimate
3. เงื่อนไขการวัด (conditionsMeasure) = Laplace estimator
4. debug หรือไม่ (debug) = False
5. ตรวจสอบความสามารถในการจำแนกประเภท (doNotCheckCapabilities) = False
6. เลือกประเภทในการจับคู่ (matching Type) = Full matching
7. จำนวนผลลัพธ์ในโมเดล (numDecimalPlaces) = 2
8. ชนิดของกฎที่สร้างขึ้น (rulesType) = lower approximation - certain rules

3.3.3 เทคนิค RIPPER เป็นเทคนิคเทคนิคที่ใช้การแยกกฎการจำแนกประเภทข้อมูลให้เป็นไปตามกลุ่มหมวดหมู่เป้าหมาย ซึ่งในงานวิจัยนี้ได้ใช้พารามิเตอร์ดังนี้

1. จำนวนของข้อมูลที่ใช้ในการสร้างแบบจำลอง (batchSize) = 100
2. ตรวจสอบอัตราความผิดพลาดหรือไม่ (checkErrorRate) = True
3. debug หรือไม่ (debug) = False
4. ตรวจสอบความสามารถในการจำแนกประเภท (doNotCheckCapabilities) = False
5. จำนวนการแบ่งชุดข้อมูลที่ใช้ในการทดสอบ (folds) = 3
6. น้ำหนักรวมต่ำสุด (minNo) = 2.0
7. จำนวนผลลัพธ์ในโมเดล (numDecimalPlaces) = 2
8. จำนวนการเพิ่มประสิทธิภาพ (optimizations) = 2
9. ข้อมูลการสุ่มตัวอย่าง (seed) = 1
10. ใช้การการทดสอบ หรือไม่ (usePruning) = True

3.3.4 เทคนิค Bagging เป็นเทคนิคที่ใช้วิธีการสุ่มข้อมูลการฝึกอบรมโดยวิธี Boostap และแบ่งข้อมูลออกเป็นหลายชุด แล้วจึงเอามาสร้างแบบจำลองด้วยเทคนิคพื้นฐาน แล้วจึงทำการทดสอบหาชุดข้อมูลที่สร้างแบบจำลองแล้วมีค่าการพยากรณ์ต่ำที่สุด ซึ่งในงานวิจัยนี้ได้ใช้พารามิเตอร์ดังนี้

1. ขนาดของถุงที่ใช้การชุดทดสอบ (bagSizePercent) = 100
2. จำนวนของข้อมูลที่ใช้ในการสร้างแบบจำลอง (batchSize) = 100
3. การคำนวณมีข้อผิดพลาดหรือไม่ (calcOutOfBag) = False
4. คลิกซ้ายเพื่อแก้ไขหรือคลิกขวาเพื่อเลือกเทคนิคที่ต้องการ (classifier) = choose
5. debug หรือไม่ (debug) = False
6. ตรวจสอบความสามารถในการจำแนกประเภท (doNotCheckCapabilities) = False
7. จำนวนผลลัพธ์ในโมเดล (numDecimalPlaces) = 2
8. จำนวนเทคนิคที่ใช้ในการทดสอบข้อมูล (numExecutionSlots) = 1
9. จำนวนการรอบในการทดสอบ (numIterrations) = 10
10. สถิติที่ซับซ้อนหรือไม่ (outputOutOfBagComplexityStatistics) = False
11. ใช้แบบสุ่มหรือไม่ (printClassifiers) = False
12. การพยากรณ์คาดเคลื่อนหรือไม่ (representCopiesUsingWeights) = False
13. ข้อมูลการสุ่มตัวอย่าง (seed) = 1

3.3.5 เทคนิค Weighted Instances Handler Wrapper (WI) เป็นเทคนิคเทคนิคที่ใช้การเพิ่มประสิทธิภาพของแบบจำลอง ซึ่งในงานวิจัยนี้ได้ทำพารามิเตอร์ดังนี้

1. จำนวนของข้อมูลที่ใช้ในการสร้างแบบจำลอง (batchSize) = 100
2. คลิกซ้ายเพื่อแก้ไขหรือคลิกขวาเพื่อเลือกเทคนิคที่ต้องการ (classifier) = choose
3. debug หรือไม่ (debug) = False
4. ตรวจสอบความสามารถในการจำแนกประเภท (doNotCheckCapabilities) = False
5. การจำแนกสามารถจัดการกับน้ำหนักตัวอย่างได้หรือไม่ (forceResampleWithWeights) = False
6. จำนวนผลลัพธ์ในโมเดล (numDecimalPlaces) = 2
7. ข้อมูลการสุ่มตัวอย่าง (seed) = 1

3.4 การวัดประสิทธิภาพของแบบจำลอง

การวัดประสิทธิภาพของแบบจำลองนั้น ผู้วิจัยได้ทำการทดลองแบ่งชุดข้อมูลออกเป็น 2 ชุด คือ 1. ชุดข้อมูลเรียนรู้ 2. ชุดข้อมูลทดสอบ ด้วยการใช้หลักการ 10-fold cross validation และวัดประสิทธิภาพของการพยากรณ์ของ Sensitivity, Specificity และ Accuracy

พหุน ปณ ฑิโต ชีเว

บทที่ 4
ผลการศึกษา

ผลการดำเนินการวิจัยในงานวิจัยนี้ผู้วิจัยได้ใช้เทคนิค FURIA, MODLEM และRIPPER ในการสร้างแบบจำลองเพื่อพยากรณ์ ในการวัดประสิทธิภาพแบบจำลองได้ทดลองแบ่งชุดข้อมูลออกเป็นชุดข้อมูลเรียนรู้ และชุดข้อมูลทดสอบ ด้วย 10 รอบของ 10-fold cross validation และวัดประสิทธิภาพการพยากรณ์ของแบบจำลองที่สร้างจากแต่ละเทคนิคโดยใช้ค่า Sensitivity, Specificity และ Accuracy และแสดงตัวอย่างการแปลผลการพยากรณ์

4.1 ผลการเปรียบเทียบประสิทธิภาพของแบบจำลองตามกฎ

การเปรียบเทียบประสิทธิภาพของแบบจำลองตามกฎ ในงานวิจัยนี้ได้นำเอาค่า Sensitivity, Specificity และ Accuracy มาทำการเปรียบเทียบได้ผลการทดลองดังตารางที่ 4.1 ตารางที่ 4.2 และ ตารางที่ 4.3 ตามลำดับ

ตารางที่ 4.1 การเปรียบเทียบ ค่า Sensitivity ของแบบจำลองในการพยากรณ์การเกิดโรคมะเร็งเต้านม และโรคเบาหวาน

Techniques	Breast cancer (%)		Diabetes (%)	
	Sensitivity	Improvement	Sensitivity	Improvement
DT	84.41±1.28	-	96.64±1.30	-
DTNB	83.01±0.88	-	96.76±0.27	-
FURIA	87.01±1.18	-	97.75±0.49	-
MODLEM	89.52±1.01	-	97.86±0.13	-
RIPPER	84.68±0.75	-	95.78±1.11	-
Bag+DT	84.70±0.41	0.28	96.98±0.20	0.34
Bag+DTNB	85.87±0.41	2.86	97.53±0.10	0.77
Bag+FURIA	93.37±0.58	6.36	97.86±0.27	0.11
Bag+MODLEM	87.20±0.69	-2.32	97.55±0.50	-0.31
Bag+RIPPER	85.20±0.53	0.52	96.53±0.43	0.75
WI+DT	82.89±0.89	-1.52	96.82±0.27	0.19

ตารางที่ 4.1 การเปรียบเทียบ ค่า Sensitivity ของแบบจำลองในการพยากรณ์การเกิดโรคมะเร็งเต้านม และโรคเบาหวาน (ต่อ)

Techniques	Breast cancer (%)		Diabetes (%)	
	Sensitivity	Improvement	Sensitivity	Improvement
WI+DTNB	86.91±1.20	3.90	97.63±0.51	0.87
WI+FURIA	89.51±1.02	2.50	97.80±0.14	0.05
WI+MODLEM	84.95±0.82	-4.57	96.02±1.02	-1.84
WI+RIPPER	83.39±2.51	-1.29	94.76±0.84	-1.02

จากตารางที่ 4.1 แสดงค่า Sensitivity ของแบบจำลองในการพยากรณ์การเกิดโรคมะเร็งเต้านมและโรคเบาหวาน พบว่าเทคนิคพื้นฐาน MODLEM มีค่ามากที่สุดถึง 89.52% จากชุดข้อมูลโรคมะเร็งเต้านม และมากที่สุดถึง 97.86% จากชุดข้อมูลโรคเบาหวาน แต่เทคนิคพื้นฐาน DTNB มีค่าต่ำสุด 83.01% จากชุดข้อมูลโรคมะเร็งเต้านม และเทคนิคพื้นฐาน RIPPER 95.78% จากชุดข้อมูลโรคเบาหวาน เมื่อวิเคราะห์ประสิทธิภาพของแบบจำลองจากเทคนิคพื้นฐาน รวมกับเทคนิคแบบรวมในการเกิดโรคมะเร็งเต้านมและโรคเบาหวาน ค่า Sensitivity ของแบบจำลองในการพยากรณ์ พบว่าเทคนิคแบบรวม Bagging สามารถเพิ่มค่า Sensitivity ของแบบจำลองที่สร้าง ด้วยเทคนิค FURIA สูงขึ้นมากที่สุดถึง 6.36% จากชุดข้อมูลโรคมะเร็งเต้านม และเทคนิคDTNB สูงขึ้น 0.87% จากชุดข้อมูลโรคเบาหวาน แต่เทคนิคแบบรวม WI ทำให้ ค่าSensitivity ของแบบจำลอง MODLEM ลดลง -4.57% จากชุดข้อมูลโรคมะเร็งเต้านม และลดลง -1.84% จากชุดข้อมูลโรคเบาหวาน

ตารางที่ 4.2 การเปรียบเทียบ ค่า Specificity ของแบบจำลองในการพยากรณ์การเกิดโรคมะเร็งเต้านม และโรคเบาหวาน

Techniques	Breast cancer (%)		Diabetes (%)	
	Specificity	Improvement	Specificity	Improvement
DT	79.73±1.53	-	62.35±0.91	-
DTNB	94.37±1.51	-	67.98±1.02	-
FURIA	87.83±0.64	-	63.44±0.88	-
MODLEM	87.82±1.07	-	59.89±0.84	-
RIPPER	86.75±1.19	-	65.49±1.19	-
Bag+DT	88.31±1.52	8.58	65.77±0.85	3.42
Bag+DTNB	96.70±0.52	2.33	67.83±0.88	-0.15
Bag+FURIA	87.21±2.07	-0.62	66.36±0.71	2.92
Bag+MODLEM	90.37±0.98	2.55	56.36±0.78	-3.53
Bag+RIPPER	86.57±1.29	-0.18	65.36±0.70	-0.13
WI+DT	81.16±1.49	1.43	65.29±1.07	2.94
WI+DTNB	93.71±1.62	-0.66	67.01±1.05	-0.98
WI+FURIA	85.82±1.01	-2.01	63.76±0.85	0.32
WI+MODLEM	90.36±0.83	2.54	59.11±1.05	-0.78
WI+RIPPER	86.29±1.73	-0.46	66.17±1.00	0.68

จากตารางที่ 4.2 แสดงค่า Specificity ของแบบจำลองในการพยากรณ์การเกิด โรคมะเร็งเต้านมและโรคเบาหวาน พบว่าเทคนิคพื้นฐาน DTNB มีค่ามากที่สุดถึง 94.37% จากชุดข้อมูลโรคมะเร็งเต้านม มีค่ามากที่สุดถึง 67.98% จากชุดข้อมูลโรคเบาหวาน แต่เทคนิคพื้นฐาน DT มีค่าต่ำสุด 79.73% จากชุดข้อมูลโรคมะเร็งเต้านม และเทคนิคพื้นฐาน MODLEM มีค่าต่ำสุด 59.89% จากชุดข้อมูลโรคเบาหวาน เมื่อวิเคราะห์ประสิทธิภาพของแบบจำลองจากเทคนิคพื้นฐาน รวมกับเทคนิคแบบรวมในการเกิดโรคมะเร็งเต้านมและโรคเบาหวาน ค่า Specificity ของแบบจำลองในการพยากรณ์ พบว่า เทคนิคแบบรวม Bagging รวมกับเทคนิค DT สูงขึ้น 8.58% จากชุดข้อมูลโรคมะเร็งเต้านม และสูงขึ้น 3.42% จากชุดข้อมูลโรคเบาหวาน แต่เมื่อรวมเทคนิคแบบรวม WI ทำให้ Specificity ของแบบจำลอง FURIA ลดลง -2.01% จากชุดข้อมูลโรคมะเร็งเต้านม และ Bagging รวมกับเทคนิค MODLEM ลดลง -3.53% จากชุดข้อมูลโรคเบาหวาน

ตารางที่ 4.3 การเปรียบเทียบ ค่า Accuracy ของแบบจำลองในการพยากรณ์การเกิดโรคมะเร็งเต้านม และโรคเบาหวาน

Techniques	Breast cancer (%)		Diabetes (%)	
	Accuracy	Improvement	Accuracy	Improvement
DT	83.17± 1.06	-	82.01± 0.59	-
DTNB	85.81± 0.82	-	84.03± 0.47	-
FURIA	87.21±0.84	-	82.75±0.55	-
MODLEM	89.14±0.74	-	80.87±0.64	-
RIPPER	85.22±0.77	-	82.84±0.72	-
Bag+DT	85.65±0.48	2.48	83.16±0.44	1.15
Bag+DTNB	88.49±0.48	2.68	84.68±0.54	0.65
Bag+FURIA	92.12±0.58	4.91	84.17±0.32	1.42
Bag+MODLEM	87.98±0.63	-1.16	79.37±0.47	-1.5
Bag+RIPPER	85.56±0.41	0.34	83.01±0.39	0.17
WI+DT	82.41±0.49	-0.76	82.73±0.49	0.73
WI+DTNB	88.49±1.04	2.68	84.25±0.67	0.21
WI+FURIA	88.69±0.92	1.48	82.84±0.46	0.09
WI+MODLEM	86.30±0.69	-2.84	79.84±0.85	-1.03
WI+RIPPER	84.14±1.67	-1.08	82.49±0.68	-0.35

จากตารางที่ 4.3 แสดงค่า Accuracy ของแบบจำลองในการพยากรณ์การเกิดโรคมะเร็งเต้านม และ โรคเบาหวาน พบว่าเทคนิคพื้นฐาน MODLEM มีค่ามากที่สุดถึง 89.14% จากชุดข้อมูลโรคมะเร็งเต้านม และเทคนิคพื้นฐาน DTNB มีค่ามากที่สุดถึง 84.03% จากชุดข้อมูลโรคเบาหวาน แต่เทคนิคพื้นฐาน DT มีค่าต่ำสุด 83.17% จากชุดข้อมูลโรคมะเร็งเต้านม และเทคนิคพื้นฐาน MODLEM มีค่าต่ำสุด 80.87% จากชุดข้อมูลโรคเบาหวานเมื่อวิเคราะห์ประสิทธิภาพของแบบจำลองจากเทคนิคพื้นฐาน รวมกับเทคนิคแบบรวมในการเกิดโรคมะเร็งเต้านมและโรคเบาหวาน ค่า Accuracy ของแบบจำลองในการพยากรณ์ พบว่า เทคนิคแบบรวม Bagging สามารถเพิ่มค่า Accuracy ของแบบจำลองที่สร้างด้วยเทคนิค FURIA สูงขึ้นมากที่สุดถึง 4.91% จากชุดข้อมูลโรคมะเร็งเต้านม และสูงขึ้น 1.42% จากชุดข้อมูลโรคเบาหวาน แต่เมื่อรวมเทคนิคแบบรวม

โรคเบาหวาน แต่เทคนิคแบบรวม WI ทำให้ Accuracy ของแบบจำลอง MODLEM ลดลง -2.84% จากชุดข้อมูลโรคมะเร็งเต้านม และลดลง -1.03% จากชุดข้อมูลโรคเบาหวาน

4.2 แบบจำลองเพื่อการพยากรณ์ Bagging และ FURIA

จากการทดลองพบว่าแบบจำลองที่สร้างจากเทคนิค Bagging ร่วมกับเทคนิค FURIA ให้ค่าความถูกต้องมากที่สุดคณะผู้วิจัยจึงได้นำแบบจำลองตามกฎของผู้ป่วยโรคมะเร็งเต้านมและผู้ป่วยโรคเบาหวานจำนวน 5 กฎมาแสดงให้เห็นถึงความง่ายในการพยากรณ์ ตารางที่ 4.4

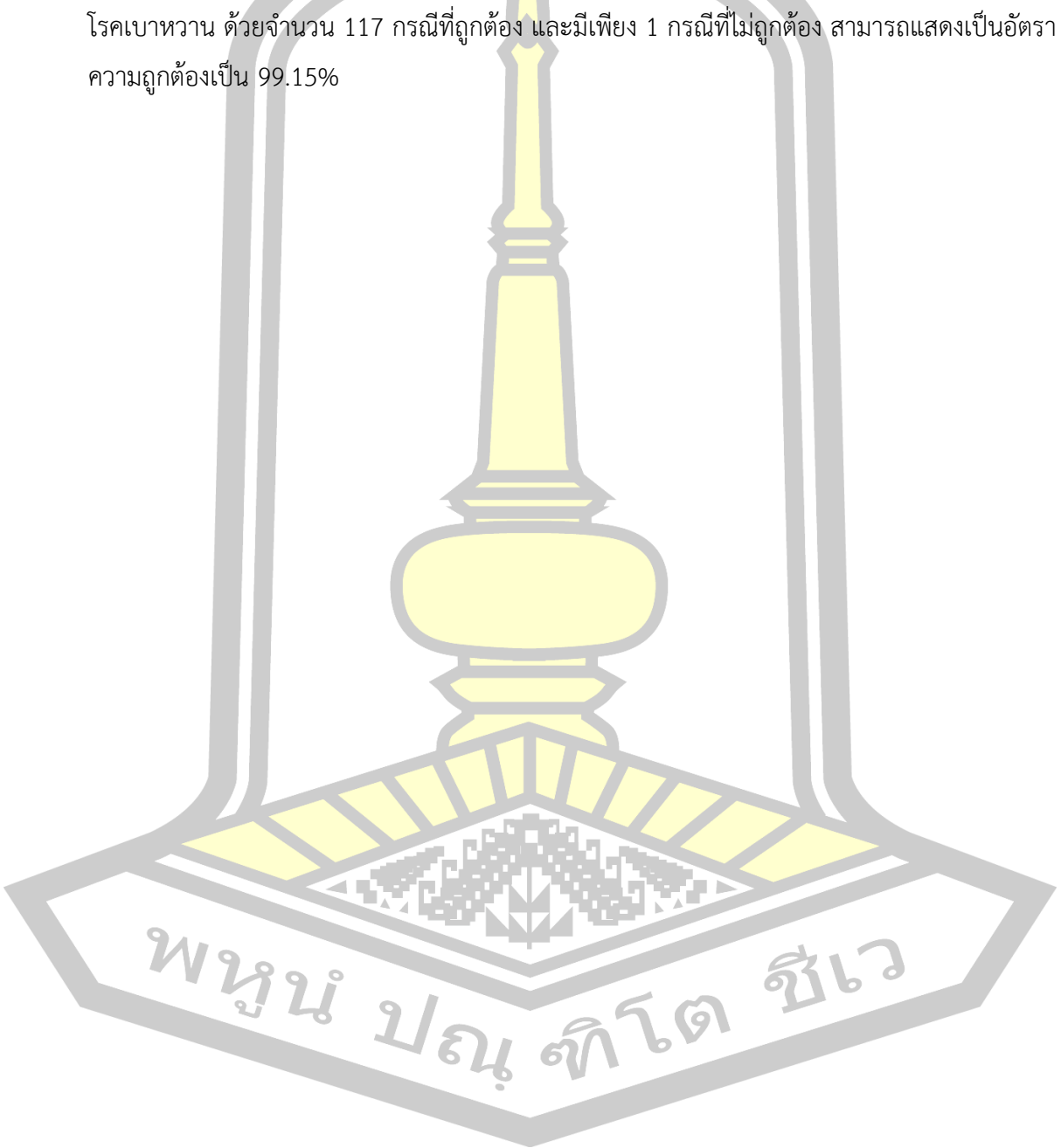
ตารางที่ 4.4 แบบจำลองตามกฎของผู้ป่วยโรคมะเร็งเต้านมและผู้ป่วยโรคเบาหวาน

Order No.	Breast Cancer Model	Diabetic Model
กฎที่ 1	Nuclei = 1 AND cellSize = 1 : tested_positive (336.0)	plas <= 127 AND mass <= 26.4 AND preg <= 7: tested_negative (117.0/1.0)
กฎที่ 2	cellShape=10 : tested_negative (58.0)	Plas>154ANDmass>29.8 : tested_positive (100.0/14.0)
กฎที่ 3	cellShape=10 : tested_negative (58.0)	plas <= 99 AND age <= 25 AND age <= 22: tested_negative (33.0)
กฎที่ 4	Nuclei=10 : tested_negative (94.0/2.0)	age <= 28 AND skin > 0 AND skin <= 34 AND age > 22 AND preg<=ANDplas<=127: tested_negative (61.0/7.0)
กฎที่ 5	cellSize = 2 : tested_positive (30.0/6.0)	plas <= 99 AND insu <= 88 AND insu <= 18 AND skin <= 21 : tested_negative (26.0/1.0)

จากตารางที่ 4.4 แบบจำลองตามกฎของผู้ป่วยโรคมะเร็งเต้านมและผู้ป่วยโรคเบาหวาน จากตัวอย่างแบบจำลองตามกฎของผู้ป่วยโรคมะเร็งเต้านม กฎข้อที่ 1 ถ้าการถ่ายทอดทางกรรมพันธุ์ในผู้ป่วยโรคมะเร็งเต้านมมีเป็นปกติ และความสม่ำเสมอของขนาดเซลล์ในผู้ป่วยโรคมะเร็งเต้านม

มีค่าเท่ากับ 1 พยากรณ์ว่าผู้ป่วยไม่เกิดโรคมะเร็งเต้านมด้วยความเป็นไปได้ด้วยจำนวน 336 กรณีที่ถูกต้อง สามารถแสดงเป็นอัตราความถูกต้องเป็น 100%

ส่วนแบบกฎข้อที่1ของผู้ป่วยโรคเบาหวาน สามารถแปลได้ดังนี้คือ ถ้าระดับน้ำตาลในเลือดในผู้ป่วยมีน้อยกว่าหรือเท่ากับ 127 และดัชนีมวลกายมีค่าน้อยกว่าหรือเท่ากับ 26.4 และ จำนวนครั้งที่ตั้งครรภ์ในผู้ป่วยโรคเบาหวานมีค่าน้อยกว่าหรือเท่ากับ 7 ครั้ง พยากรณ์ว่าผู้ป่วยไม่เป็นโรคเบาหวาน ด้วยจำนวน 117 กรณีที่ถูกต้อง และมีเพียง 1 กรณีที่ไม่ถูกต้อง สามารถแสดงเป็นอัตราความถูกต้องเป็น 99.15%



บทที่ 5
สรุป วิจารณ์ และข้อเสนอแนะ

การศึกษาในครั้งนี้ มีวัตถุประสงค์เพื่อศึกษาเทคนิคแบบรวมเพื่อเพิ่มประสิทธิภาพของแบบจำลองตามกฎ สามารถสรุปผลการศึกษา อภิปรายผล รวมทั้งข้อเสนอแนะแนวทางในการวิจัยได้ดังต่อไปนี้

5.1 สรุป

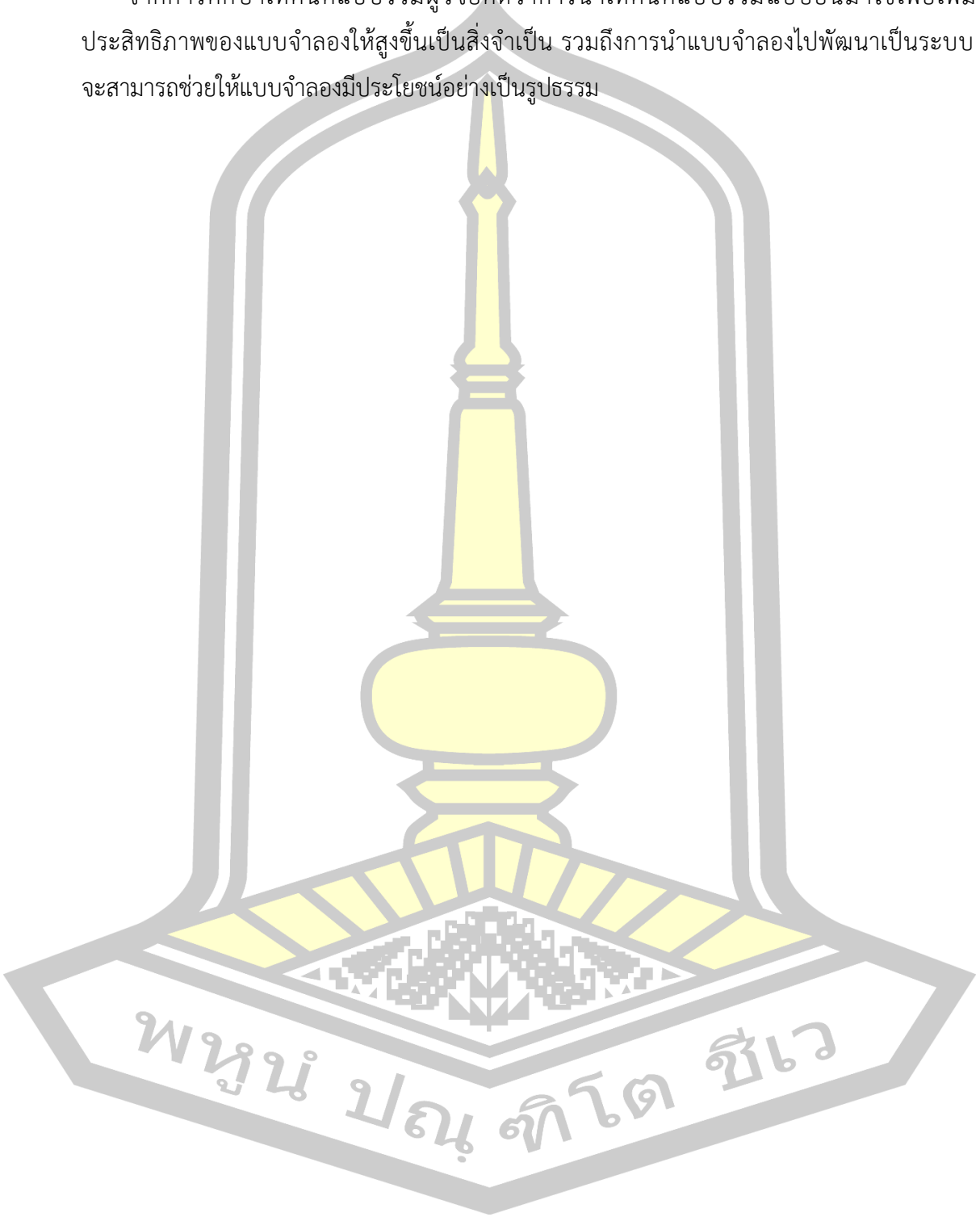
จากการทดลองการสร้างกฎความสัมพันธ์พยากรณ์การเกิดโรคมะเร็งเต้านมและโรคเบาหวาน สรุปได้ว่าการเพิ่มประสิทธิภาพของแบบจำลองตามกฎพื้นฐาน 5 เทคนิค รวมถึง DT, DTNB, FURIA, MODLEM และ RIPPER โดยใช้เทคนิค Bagging และ Weighted Instances Handler Wrapper พบว่าค่า Sensitivity เมื่อ Bagging รวมกับ FURIA มีค่ามากที่สุดถึง 93.37% และ 97.86% ส่วนค่า Specificity เมื่อ Bagging รวมกับ DTNB มีค่ามากที่สุดถึง 96.70% จากชุดข้อมูลผู้ป่วยโรคมะเร็งเต้านม และ Bagging 96.72% จากชุดข้อมูลผู้ป่วยโรคเบาหวาน ค่า Accuracy เมื่อ Bagging รวมกับ FURIA มีค่ามากที่สุดถึง 92.12% จากชุดข้อมูลโรคมะเร็งเต้านม และ Bagging รวมกับ DTNB มีค่ามากที่สุดถึง 84.68% จากชุดข้อมูลโรคเบาหวาน เมื่อ Bagging รวมกับ FURIA สามารถทำให้ค่า Accuracy ในข้อมูลผู้ป่วยโรคเบาหวาน สูงขึ้น 92.12% และเมื่อ Bagging รวมกับ DTNB สูงขึ้น 84.68% ยิ่งไปกว่านั้นเมื่อ Bagging รวมกับ DT สามารถทำให้ค่า Specificity มีประสิทธิภาพเพิ่มขึ้นถึง 8.58% และ Bagging รวมกับ FURIA สามารถทำให้ค่า Sensitivity มีประสิทธิภาพเพิ่มขึ้นถึง 6.36%

5.2 วิจารณ์

จากการสร้างกฎความสัมพันธ์พยากรณ์การเกิดโรคมะเร็งเต้านมและโรคเบาหวานข้างต้นแสดงให้เห็นว่าการการพยากรณ์การเกิดโรคด้วยเทคนิคแบบรวมเพื่อเพิ่มประสิทธิภาพของแบบจำลองตามกฎมีประสิทธิภาพเพิ่มขึ้น ซึ่งสอดคล้องกับผลงานวิจัยของ Belhadj, Benmoussat และ Krachai [9] ได้ทำการเพิ่มประสิทธิภาพในการตรวจสอบข้อมูลของ BMI โดยใช้เทคนิค Bagging ทำให้เทคนิคพื้นฐานสูงขึ้น Golinko และ Zhu [28] นำเอาอัลกอริธึมการเรียนรู้การฝังตัวคุณลักษณะโดยรวม GFEL ซึ่งเรียนรู้คุณลักษณะโดยการจำแนกประเภทและคุณลักษณะในการจัดกลุ่มและจัดหมวดหมู่โดยใช้เทคนิค weighted instance เพื่อเพิ่มประสิทธิภาพในการพยากรณ์ โดยใช้ข้อมูลชุดทดสอบจำนวน 28 ชุด พบว่า ในการจัดกลุ่มและจัดหมวดหมู่ได้อย่างมีประสิทธิภาพสูงสุด

5.3 ข้อเสนอแนะ

จากการศึกษาเทคนิคแบบรวมผู้วิจัยคิดว่าการนำเทคนิคแบบรวมแบบอื่นมาใช้เพื่อเพิ่มประสิทธิภาพของแบบจำลองให้สูงขึ้นเป็นสิ่งจำเป็น รวมถึงการนำแบบจำลองไปพัฒนาเป็นระบบจะสามารถช่วยให้แบบจำลองมีประโยชน์อย่างเป็นรูปธรรม



บรรณานุกรม

บรรณานุกรม

[1] ปวีณา ชัยวนารมย์, "การพัฒนาแบบจำลองเพื่อพยากรณ์การเกิดความเครียดในหลายระดับด้วยเทคนิคการทำเหมืองข้อมูล," ในการประชุมวิชาการระดับชาติมหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ ครั้งที่ 1, 2016, หน้า. 406-416.

[2] เสถียร วิชาเรือง, การประยุกต์ใช้เทคนิคการทำเหมืองข้อมูลในกระบวนการอนุมัติสินเชื่อ. มหาวิทยาลัยธุรกิจบัณฑิตย์, กรุงเทพมหานคร, 2553.

[3] Stefanowski, J., "The rough set based rule induction technique for classification problems," in Proceedings of the 6th European Conference on Intelligent Techniques and Soft Computing EUFIT, 1998.

[4] Jian, Y., Papazoglou, M. P., Orriens, B., and Heuvel, W. J. v., "A rule based approach to the service composition life-cycle," in Proceedings of the 4th International Conference on Web Information Systems Engineering, 2003, pp. 295-298.

[5] กิตติศักดิ์ สุมามาลย์, "การคัดกรองสุขภาพเบื้องต้นโดยใช้เทคนิคเหมืองข้อมูล," วิทยาศาสตรมหาบัณฑิต, มหาวิทยาลัยธุรกิจบัณฑิตย์, กรุงเทพมหานคร, 2555.

[6] Basterrech, S., and Mesa, A., "Bagging technique using temporal expansion functions," in Proceedings of the Fifth Internal Conference on Innovation in Bio-Inspired Computer and Application, 2014, pp. 395-404.

[7] Karasuyama, M., Harada, N., Sugiyama, M., and Takeuchi, I., "Multi-parametric solution-path algorithm for instance-weighted support vector machines," in Proceedings of the International Workshop on Machine Learning for Signal Processing, 2011, pp. 1-6.

[8] Denniston, R. W., and "The breast cancer digest : a guide to medical care, emotional support,education programs and resource," National Medical Association, vol. 72(2), p. 150, 1980.

[9] American Diabetes Association. Classification and Diagnosis of Diabetes. สืบค้นเมื่อ: 1/10/2560 สืบค้นจาก: http://care.diabetesjournals.org/content/38/Supplement_1/S8.short.

[10] Han, J., and Kamber, M., Data Mining Concepts and Techniques. Elsevier, Waltham, 2006.

[11] Rahman, H., Data Mining Applications for Empowering Knowledge Societies. Information Science Reference, New York, 2009.

[12] Juneja, J., and Amar, A. D., "A capital allocation decision rule for knowledge-intensive organizations," in Proceedings of the International Conference on Technology & Engineering Management Conference, 2017, pp. 376-383.

[13] Zhang, C., and Zhang, S., "Association rule mining," in Models and Algorithms, ed: Springer, VerlagBerlin, 2002, pp. 1-23.

[14] Yu, H., Ni, J., Dan, Y., and Xu, S., "Mining and integrating reliable decision rules for imbalanced cancer gene expression data sets," Tsinghua Science and Technology, vol. 17, pp. 666-673, 2012.

[15] Gil-Herrera, E., Aden-Buie, G., Yalcin, A., Tsalatsanis, A., Barnes, L. E., and Djulbegovic, B., "Rough set theory based prognostic classification models for hospice referral," BMC Medical Informaitics & Decision Making, vol. 15, pp. 1-18, 2015.

[16] Thakur, S., Meenakshi, E., and Priya, A., "Detection of malicious URLs in big data using RIPPER algorithm," in Proceedings of the 2nd International Conference on Recent Trends in Electronics, Information & Communication Technology, 2017, pp. 1296-1301.

[17] Lotte, F., Lecuyer, A., and Arnaldi, B., "FuRIA: An Inverse Solution Based Feature Extraction Algorithm Using Fuzzy Set Theory for Brain– Computer Interfaces," IEEE Transactions on Signal Processing, vol. 57, pp. 3253-3263, 2009.

[18] นิภาพร ชนะมาร และพรรณี สิทธิเดช, "การวิเคราะห์ปัจจัยการเรียนรู้ด้วยการคัดเลือกคุณสมบัติและการพยากรณ์," หน้า. 31-46, 2557.

[19] Kohavi, R., "The Power of Decision Tables,"in Proceedings of 8th International Conference on European Conference on Machine Lerning, pp. 1-17 1995.

[20] สายชล สินสมบูรณ์ทอง, การทำเหมืองข้อมูล. กรุงเทพฯ: จามจุรีโปรดักส์, 2558.

[21] Goyal, s., and Modi, n., "Data mining using enhanced dedision table classifier for online shopping," in Proceedings of the 7th International Conference on Cloud Computing, Data Science & Engineering – Confluence, 2017, pp. 313-318,

[22] Frank, E., and H, I., "Witten:Generating Accurate Rule Sets Without Global Optimization," in Proceedings of the 15th International Conference on Machine Learning, 1998, pp. 1-15.

[23] Cohen, W. W., "Fast effective rule induction," in Proceedings of the 12th International Conference on Machine Learning, 1995, pp. 1-19.

[24] Hühn, J. C., and Hüllermeier, E., "FURIA: an algorithm for unordered fuzzy rule induction," Data Mining and Knowledge Discovery. vol. 19, pp. 293-319 2009.

[25] J.W, G.-B., "LERS - a system for learning from examples based on rough sets," in Intelligent Decision Support, Kluwer Academic Publishers. vol. 11, 1992, pp. 3-18.

[26] Shixin, Y., "Feature selection and classifier ensembles:A study on Hyperspectral Remote Sensing data," in Ph.D. Thesis, University of Antwerp, 2003.

[27] Breiman, L., "Bagging predictors," Machine Learning, vol. 24, pp. 123-140, 1996.

[28] Golinko, E., and Zhu, X., "GFEL: Generalized feature embedding learning using weighted instance matching," in Proceedings of the International Conference on Information Reuse and Integration, 2017, pp. 235-244.

[29] Ashraf, M., Le, K., and Huang, X., "Iterative weighted k-NN for constructing missing feature values in Wisconsin breast cancer dataset," in Proceedings of the 3rd International Conference on Data Mining and Intelligent Information Technology Applications, 2011, pp. 23-27.

[30] Mumtaz, K., Sheriff, S. A., and Duraiswamy, K., "Evaluation of three neural network models using Wisconsin breast cancer database," in Proceedings of the International Conference on Control, Automation, Communication and Energy Conservation, 2009, pp. 1-7.

[31] Sirisomboonrat, C., and Sinapiromsaran, K., "Breast cancer diagnosis using multi-attributed lens recursive partitioning algorithm," in Proceedings of the 10th International Conference on ICT and Knowledge Engineering, 2012, pp. 40-45.

[32] Ghosh, S., Mondal, S., and Ghosh, B., "A comparative study of breast cancer detection based on SVM and MLP BPN classifier," in Proceedings of the 1st International Conference on Automation, Control, Energy and Systems, 2014, pp. 1-4.

[33] ณัฐวุฒิ ศรีวิบูลย์, "การเปรียบเทียบประสิทธิภาพอัลกอริธึมเหมืองข้อมูลเพื่อวิเคราะห์ปัจจัยที่ส่งผลต่อการเกิดโรคมะเร็งเต้านม," วารสาร SNRU สถาบันวิจัยและพัฒนา, vol. 8, pp. 344-352, 2016.

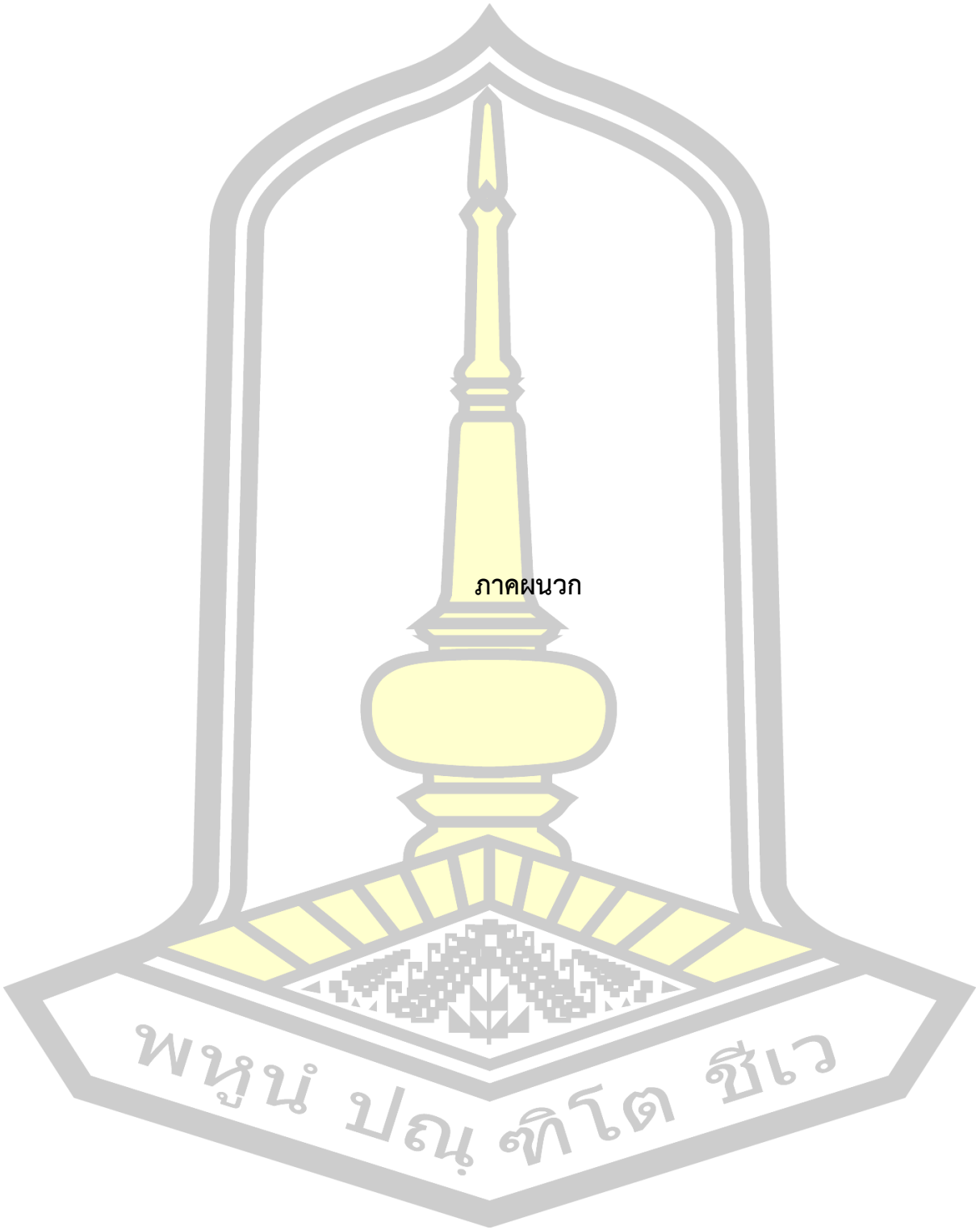
[34] Isaac, L. D., and Sureshkumar, C., "Diagnosis prognosis and prevention of breast cancer based on present scenario of human life," in Proceedings of the International Conference on Communication information and Computing Technology, 2018, pp. 1-7.

[35] Islam, M. M., Iqbal, H., Haque, M. R., and Hasan, M. K., "Prediction of breast cancer using support vector machine and K-Nearest neighbors," in Proceedings of the Region 10 Humanitarian Technology Conference, 2017, pp. 226-229.

[36] Adegoke, V. F., Chen, D., Banissi, E., and Barikzai, S., "Prediction of breast cancer survivability using ensemble algorithms," in Proceedings of the International Conference on Smart Systems and Technologies, 2017, pp. 223-231.

[37] NirmalaDevi, M., Appavu, S., and Swathi, U., "An amalgam KNN to predict diabetes mellitus," in Proceedings of the International Conference onEmerging Trends in Computing, Communication and Nanotechnology, 2013, pp. 691-695.

พหุน ปณ ฑิโต ชีเว



ภาคผนวก

ผลการทดลองการเปรียบเทียบ ค่า Sensitivity ของแบบจำลองในการพยากรณ์การเกิดโรคมะเร็งเต้านม

Rounds	FURIA	MODLEM	RIPPER	DT	DTNB	Bag+FURIA	Bag+MODLEM	Bag+RIPPER	Bag+DT	Bag+DTNB	WI+FURIA	WI+MODLEM	WI+RIPPER	WI+DT	WI+DTNB
1	87.35	89.12	85.37	81.77	85.50	94.34	86.68	84.08	84.86	85.80	89.14	85.77	84.34	83.37	87.22
2	86.45	88.93	85.71	84.29	85.66	93.33	88.55	85.29	84.28	85.49	88.93	85.94	80.68	83.46	86.35
3	86.06	90.06	84.79	84.31	82.22	94.13	87.65	85.85	84.12	87.78	90.00	85.29	79.47	81.84	86.00
4	86.68	90.79	85.01	85.51	84.81	92.93	86.85	84.95	85.29	85.21	90.83	83.14	87.93	84.66	86.52
5	87.65	91.16	85.03	84.57	82.25	93.32	87.35	85.40	85.01	85.66	91.16	85.24	85.89	81.94	87.60
6	86.17	89.51	84.38	85.49	82.79	92.72	87.70	85.66	84.28	85.83	89.40	84.47	84.54	82.85	86.12
7	88.34	88.57	85.86	82.57	85.37	92.92	86.60	85.29	85.12	85.18	88.52	84.11	84.69	83.37	88.32
8	86.71	90.44	84.57	86.14	85.56	93.52	87.68	84.74	85.12	86.35	90.48	84.63	81.70	83.46	86.57
9	85.19	88.75	84.82	84.28	81.94	93.93	86.97	85.88	84.54	86.17	88.82	85.27	83.80	81.87	85.01
10	89.48	87.85	85.28	85.21	82.03	92.54	86.00	84.84	84.34	85.21	87.85	85.60	80.88	82.13	89.38
Average	87.01	89.52	84.68	84.41	85.01	93.37	87.20	85.20	84.70	85.87	89.51	84.95	83.39	82.89	86.91
SD.	1.18	1.01	0.75	1.28	0.88	0.58	0.69	0.53	0.41	0.74	1.02	0.82	2.51	0.89	1.20

ผลการทดลองการเปรียบเทียบ ค่า Sensitivity ของแบบจำลองในการพยากรณ์การเกิดโรคเบาหวาน

Rounds	FURIA	MODLEM	RIPPER	DT	DTNB	Bag+FURIA	Bag+MODLEM	Bag+RIPPER	Bag+DT	Bag+DTNB	WI+FURIA	WI+MODLEM	WI+RIPPER	WI+DT	WI+DTNB
1	97.33	98.03	94.27	100.00	96.74	97.50	98.54	96.22	97.12	97.47	97.98	94.50	94.48	96.66	97.19
2	97.30	97.81	95.97	97.00	96.97	97.49	96.90	95.90	97.09	97.45	97.74	96.12	94.49	97.00	97.20
3	97.79	97.80	95.22	97.15	96.50	98.15	97.29	97.30	97.31	97.70	97.79	95.48	94.57	96.46	97.64
4	98.25	97.78	96.83	95.72	96.67	98.15	96.88	96.62	97.29	97.67	97.82	96.92	95.74	96.81	98.07
5	98.41	98.06	97.45	96.64	97.19	97.74	97.71	96.70	96.87	97.48	97.96	97.63	95.68	97.29	98.34
6	97.70	97.80	95.94	95.45	96.66	97.51	98.13	95.81	96.86	97.68	97.70	96.11	92.79	96.77	97.59
7	96.85	98.05	97.25	96.79	96.22	98.00	97.13	96.48	96.90	97.44	97.95	97.41	95.72	96.35	96.67
8	98.41	97.85	94.28	96.61	97.02	97.70	97.49	96.66	96.88	97.48	97.74	94.80	94.28	97.05	98.37
9	97.51	97.61	95.89	95.10	96.93	97.95	97.69	96.70	96.86	97.45	97.51	96.15	94.81	97.05	97.39
10	97.99	97.86	94.65	95.92	96.67	98.17	97.71	96.90	96.64	97.47	97.77	95.02	95.08	96.80	97.84
Average	97.75	97.86	95.78	96.64	96.76	97.83	97.55	96.53	96.98	97.53	97.80	96.02	94.76	96.82	97.63
SD.	0.49	0.13	1.11	1.30	0.27	0.27	0.50	0.43	0.20	0.10	0.14	1.02	0.84	0.27	0.51

ผลการทดลองการเปรียบเทียบ ค่า Specificity ของแบบจำลองในการพยากรณ์การเกิดโรคมะเร็งเต้านม

Rounds	FURIA	MODLEM	RIPPER	DT	DTNB	Bag+FURIA	Bag+MODLEM	Bag+RIPPER	Bag+DT	Bag+DTNB	WI+FURIA	WI+MODLEM	WI+RIPPER	WI+DT	WI+DTNB
1	86.54	85.94	88.14	77.60	95.60	88.60	90.42	87.85	88.83	96.30	83.97	89.66	88.14	78.76	95.07
2	88.27	89.55	88.69	78.21	94.48	84.48	90.00	86.36	85.71	96.34	86.33	91.41	89.56	79.79	94.08
3	87.43	89.05	86.63	81.97	93.60	86.82	89.44	84.21	88.33	97.44	85.31	91.53	86.98	82.99	92.99
4	88.34	87.23	85.03	78.80	94.64	87.39	90.80	86.08	89.06	96.97	86.62	89.83	82.72	80.30	94.12
5	88.48	87.59	87.78	82.26	93.57	90.27	91.57	88.51	87.03	97.33	86.99	89.77	86.42	82.90	92.99
6	88.24	86.96	87.43	78.19	96.27	85.48	91.36	85.25	88.70	95.65	86.96	89.47	86.42	79.08	95.74
7	87.34	88.46	86.67	80.21	97.08	88.07	88.17	88.33	86.03	96.84	85.19	91.23	85.45	81.37	96.50
8	88.55	86.81	85.79	79.77	91.95	88.98	91.25	85.88	88.65	96.25	86.81	89.20	85.14	82.05	91.30
9	87.88	88.64	86.36	80.98	92.66	83.33	90.80	86.70	90.16	96.93	85.40	91.02	86.52	82.41	91.77
10	87.18	88.00	85.03	79.33	93.85	88.70	89.94	86.52	90.56	96.91	84.62	90.45	85.55	81.95	92.52
Averege	87.83	87.82	86.75	79.73	94.37	87.21	90.37	86.57	88.31	96.70	85.82	90.36	86.29	81.16	93.71
SD.	0.64	1.07	1.19	1.53	1.51	2.07	0.98	1.29	1.52	0.52	1.01	0.83	1.73	1.49	1.62

ผลการทดลองการเปรียบเทียบ ค่า Specificity ของแบบจำลองในการพยากรณ์การเกิดโรคเบาหวาน

Rounds	FURIA	MODLEM	RIPPER	DT	DTNB	Bag+FURIA	Bag+MODLEI	Bag+RIPPER	Bag+DT	Bag+DTNB	WI+FURIA	WI+MODLEM	WI+RIPPER	WI+DT	WI+DTNB
1	62.11	58.64	65.64	62.05	66.28	66.47	55.59	66.09	65.46	68.73	62.75	57.53	66.57	64.41	65.27
2	63.66	60.22	64.37	61.09	67.24	66.57	56.19	66.57	64.82	66.86	64.18	59.23	65.61	64.64	65.77
3	62.75	59.42	65.64	63.92	69.09	65.98	55.56	65.61	65.36	66.47	63.28	58.31	66.47	66.67	68.22
4	62.50	58.31	62.61	61.35	69.70	66.87	56.08	64.08	64.11	68.09	62.61	57.18	63.69	64.61	68.75
5	64.52	60.61	65.74	61.08	67.57	66.37	56.76	64.94	66.01	68.48	64.52	59.83	66.36	63.38	67.07
6	64.14	59.84	65.48	62.70	68.86	66.87	55.06	65.01	67.15	68.21	64.66	59.30	65.68	65.90	67.90
7	64.43	60.22	67.18	62.66	68.53	65.98	57.61	64.57	66.86	66.97	64.43	59.78	67.68	66.38	67.38
8	63.85	60.70	66.36	62.39	67.46	67.47	56.61	65.32	66.10	69.40	64.37	59.72	66.97	64.76	66.77
9	62.32	59.89	65.36	63.00	66.86	66.37	57.48	65.71	65.83	67.26	62.43	59.45	66.47	65.33	66.07
10	64.12	61.05	66.46	63.26	68.24	64.64	56.66	65.68	66.01	67.87	64.43	60.72	66.15	66.86	66.87
Averege	63.44	59.89	65.49	62.35	67.98	66.36	56.36	65.36	65.77	67.83	63.76	59.11	66.17	65.29	67.01
SD.	0.88	0.84	1.19	0.91	1.02	0.71	0.78	0.70	0.85	0.88	0.85	1.05	1.00	1.07	1.05

ผลการทดลองการเปรียบเทียบ ค่า Accuracy ของแบบจำลองในการพยากรณ์การเกิดโรคมะเร็งเต้านม

Rounds	FURIA	MODLEM	RIPPER	DT	DTNB	Bag+FURIA	Bag+MODLEM	Bag+RIPPER	Bag+DT	Bag+DTNB	WI+FURIA	WI+MODLEM	WI+PART	WI+DT	WI+DTNB
1	87.16	88.46	86.10	80.68	86.35	93.19	87.61	85.06	85.94	88.37	88.05	86.76	94.62	82.10	88.98
2	86.90	89.07	86.47	82.70	86.26	91.57	88.92	85.59	84.66	88.16	88.36	87.28	96.03	82.45	88.15
3	86.40	89.84	85.29	83.68	85.04	92.53	88.09	85.41	85.22	90.11	88.92	86.90	96.32	82.15	87.67
4	87.09	89.98	83.55	83.70	87.26	91.81	87.82	85.26	86.33	88.10	89.87	84.84	95.94	83.43	88.31
5	87.86	90.32	85.76	83.94	85.04	92.72	88.40	86.20	85.55	88.34	90.18	86.40	97.45	82.20	88.91
6	86.66	88.94	85.15	83.50	85.99	91.20	88.60	85.55	85.42	88.22	88.85	85.71	96.15	81.82	88.24
7	88.10	88.55	84.59	81.93	86.77	92.00	87.00	86.09	85.36	87.95	87.80	85.88	97.22	82.80	90.17
8	87.16	89.60	84.89	84.50	85.67	92.60	88.55	85.03	86.07	88.75	89.63	85.80	95.51	83.07	87.73
9	85.84	88.73	85.22	83.41	84.64	91.83	87.92	86.11	86.02	88.82	88.08	86.69	95.88	82.02	86.62
10	88.92	87.88	85.22	83.67	85.04	91.78	86.94	85.28	85.96	88.04	87.18	86.75	96.11	82.08	90.11
Average	87.21	89.14	85.22	83.17	85.81	92.12	87.98	85.56	85.65	88.49	88.69	86.30	96.12	82.41	88.49
SD.	0.84	0.74	0.77	1.06	0.82	0.58	0.63	0.41	0.48	0.61	0.92	0.69	0.76	0.49	1.04

ผลการทดลองการเปรียบเทียบ ค่า Accuracy ของแบบจำลองในการพยากรณ์การเกิดโรคเบาหวาน

Rounds	FURIA	MODLEM	RIPPER	DT	DTNB	Bag+FURIA	Bag+MODLEM	Bag+RIPPER	Bag+DT	Bag+DTNB	WI+FURIA	WI+MODLEM	WI+RIPPER	WI+DT	WI+DTNB
1	81.88	80.07	82.31	82.90	83.18	84.11	79.70	83.14	83.09	85.22	82.32	78.22	82.61	81.89	83.18
2	82.61	80.92	82.46	81.52	83.66	84.22	78.78	83.40	82.67	84.01	82.93	79.93	82.00	82.29	83.44
3	82.54	80.39	82.61	82.90	84.58	84.05	78.66	83.44	83.08	84.00	82.65	79.09	82.47	83.16	84.97
4	82.69	79.73	82.01	81.17	84.80	84.51	78.90	82.32	82.30	84.85	82.51	79.10	81.72	82.39	85.29
5	83.61	81.60	83.86	81.41	84.21	84.12	79.72	82.92	83.25	84.99	83.40	81.24	83.12	82.21	84.67
6	82.88	80.63	82.80	81.69	84.33	84.37	78.98	82.56	83.86	85.03	82.99	79.74	81.28	83.01	84.57
7	82.72	81.28	84.47	82.45	83.88	84.18	80.12	82.61	83.83	84.21	83.25	80.89	83.81	83.02	83.85
8	83.27	81.44	82.39	82.08	84.18	84.62	79.46	83.02	83.31	85.68	83.06	79.79	82.53	82.78	84.70
9	81.89	80.82	82.73	81.57	83.53	84.17	79.84	83.21	83.06	84.16	81.91	80.05	82.44	83.02	83.58
10	83.38	81.81	82.75	82.36	83.97	83.35	79.49	83.52	83.13	84.64	83.33	80.38	82.90	83.55	84.23
Average	82.75	80.87	82.84	82.01	84.03	84.17	79.37	83.01	83.16	84.68	82.84	79.84	82.49	82.73	84.25
SD.	0.55	0.64	0.72	0.59	0.47	0.32	0.47	0.39	0.44	0.54	0.46	0.85	0.68	0.49	0.67

ประวัติผู้เขียน

ชื่อ	นางสาวปพิชญา กลางนอก
วันเกิด	วันที่ 5 กันยายน พ.ศ. 2529
สถานที่เกิด	อำเภอเมือง จังหวัดอุดรธานี
สถานที่อยู่ปัจจุบัน	บ้านเลขที่ 155 หมู่ 6 ตำบลเชียงพิณ อำเภอเมือง จังหวัดอุดรธานี รหัสไปรษณีย์ 41000
ตำแหน่งหน้าที่การงาน	นักจัดการงานทั่วไป
สถานที่ทำงานปัจจุบัน	คณะวิทยาศาสร์การกีฬาและสุขภาพ สถาบันการพลศึกษา วิทยาเขตอุดรธานี เลขที่ 380 ถนนอุดร-หนองบัวลำภู ตำบลหมากแข้ง อำเภอเมือง จังหวัดอุดรธานี รหัสไปรษณีย์ 41000
ประวัติการศึกษา	พ.ศ. 2545 มัธยมศึกษาตอนต้น โรงเรียนอุดรพิชัยรักษ์พิทยา ตำบลหมากแข้ง อำเภอเมือง จังหวัดอุดรธานี รหัสไปรษณีย์ 41000 พ.ศ. 2548 ประกาศนียบัตรวิชาชีพ (ปวช.) สาขาการบัญชี วิทยาลัยอาชีวศึกษาอุดรธานี ตำบลหมากแข้ง อำเภอเมือง จังหวัดอุดรธานี รหัสไปรษณีย์ 41000 พ.ศ. 2552 ปริญญาบริหารธุรกิจบัณฑิต (บธ.บ.) สาขาการจัดการทั่วไป มหาวิทยาลัยราชภัฏอุดรธานี พ.ศ. 2561 ปริญญาวิทยาศาสตรมหาบัณฑิต (วท.ม.) สาขาวิชาเทคโนโลยีสารสนเทศ มหาวิทยาลัยมหาสารคาม

พหูน ปณ ฑิโต ชีเว