การประยุกต์ใช้เทคนิคการเรียนรู้ของคอมพิวเตอร์เพื่อพัฒนาตัวแบบทางเหมืองข้อมูลสำหรับพยากรณ์ระดับฮีโมโกลบินของผู้บริจาคโลหิต

นายสาธิต เทศสมบูรณ์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิทยาศาสตร์ระดับโมเลกุลทางจุลชีววิทยาทางการแพทย์และวิทยาภูมิคุ้มกัน ภาควิชาเวชศาสตร์การธนาคารเลือดและจุลชีววิทยาคลินิก
คณะสหเวชศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2562

Application of machine learning techniques to develop data mining models for predicting hemoglobin levels of blood donors

Mr. Sathit Tadsomboon

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science in Molecular Science of Medical Microbiology and Immunology
Department of Transfusion Medicine and Clinical Microbiology
FACULTY OF ALLIED HEALTH SCIENCES
Chulalongkorn University
Academic Year 2019

หัวข้อวิทยานิพนธ์	การประยุกต์ใช้เทคนิคการเรียนรู้ของคอมพิวเตอร์เพื่อพัฒนาตัวแบบทางเหมืองข้อมูลสำหรับพยากรณ์ระดับฮีโมโกลบินของผู้บริจาคโลหิต
โดย	นายสาธิต เทศสมบูรณ์
สาขาวิชา	วิทยาศาสตร์ระดับโมเลกุลทางจุลชีววิทยาทางการแพทย์และวิทยาภูมิคุ้มกัน
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	ผู้ช่วยศาสตราจารย์ ดร.เทวฤทธิ์ สะระชนะ

คณะสหเวชศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

..... คณบดีคณะสหเวชศาสตร์
(รองศาสตราจารย์ ดร.ปาลนี อัมรานนท์)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ
(ดร.สุนทรี กุลกีรติยุต)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(ผู้ช่วยศาสตราจารย์ ดร.เทวฤทธิ์ สะระชนะ)

..... กรรมการ
(อาจารย์ทัศนีย์ สกุลดำรงค์พานิช)

..... กรรมการภายนอกมหาวิทยาลัย
(ดร.ลิขิต ปรียานนท์)

สาธิต เทศสมบูรณ์ : การประยุกต์ใช้เทคนิคการเรียนรู้ของคอมพิวเตอร์เพื่อพัฒนาตัวแบบทางเหมืองข้อมูลสำหรับพยากรณ์ระดับฮีโมโกลบินของผู้บริจาคโลหิต. (Application of machine learning techniques to develop data mining models for predicting hemoglobin levels of blood donors) อ.ที่ปรึกษาหลัก : ผศ. ดร.เทวฤทธิ์ สะระชนะ

ปริมาณโลหิตและส่วนประกอบโลหิตที่เพียงพอในธนาคารเลือดนั้นมีความสำคัญในการรักษาผู้ป่วย ดังนั้นธนาคารเลือดทุกแห่งต้องส่งเสริมให้ผู้บริจาคโลหิตมีสุขภาพดีและมีคุณสมบัติเหมาะสมในการบริจาคโลหิตอย่างสม่ำเสมอจากข้อมูลของศูนย์บริการโลหิตแห่งชาติ สภากาชาดไทยพบผู้ถูกปฏิเสธการบริจาคโลหิตประมาณร้อยละ 15-20 จากจำนวนผู้ประสงค์จะบริจาคโลหิตทั้งหมด ส่วนใหญ่มีสาเหตุจากค่าฮีโมโกลบิน (hemoglobin; Hb) ไม่ผ่านเกณฑ์ นำไปสู่ความผิดหวังความไม่พอใจเพราะผู้บริจาคโลหิตต้องเสียเวลา ค่าใช้จ่ายในการเดินทางแต่ไม่ได้บริจาคโลหิตและส่งผลกระทบต่อการจัดหาโลหิตโดยตรง หากสามารถพยากรณ์ผลตรวจ Hb ได้ล่วงหน้าจะช่วยลดผลกระทบปัญหา เป็นประโยชน์ทั้งต่อผู้บริจาคโลหิต ธนาคารเลือดและผู้ป่วย การศึกษานี้มีวัตถุประสงค์เพื่อพัฒนาและเปรียบเทียบประสิทธิภาพเทคนิคการเรียนรู้ของคอมพิวเตอร์ในการพยากรณ์จำแนกกลุ่มผลตรวจ Hb ของผู้บริจาคโลหิต โดยเก็บข้อมูล 44 ตัวแปรของผู้บริจาคโลหิตจำนวน 2,180 รายจากภาคบริการโลหิตแห่งชาติ 12 แห่งและสถานีกาชาดหัวหินเฉลิมพระเกียรติตั้งแต่ 1 ต.ค. 2561 ถึง 31 พ.ค. 2562 นำมากลั่นกรองข้อมูล คัดเลือกตัวแปร พัฒนาตัวแบบทางเหมืองข้อมูลได้แก่ ต้นไม้ตัดสินใจ ซัพพอร์ตเวกเตอร์แมชชีน การจำแนกแบบเบส์อย่างง่ายและโครงข่ายประสาทเทียม จากนั้นเปรียบเทียบประสิทธิภาพการจำแนกกลุ่มของตัวแบบพยากรณ์พบว่าต้นไม้ตัดสินใจเป็นตัวแบบการจำแนกกลุ่มที่เหมาะสมที่สุดโดยให้ค่าความถูกต้อง ค่าความไว ค่าความจำเพาะ ค่าการพยากรณ์ผลบวก ค่าการพยากรณ์ผลลบสูงที่สุดและค่า AUC เท่ากับร้อยละ 92.20, 82.98, 94.74, 81.25, 95.29 และ 0.943 ตามลำดับ ซึ่งต้นไม้ตัดสินใจที่ได้จากศึกษานี้อาจนำไปพัฒนาต่อเป็นระบบประเมินออนไลน์ก่อนเดินทางมาบริจาคโลหิตได้

สาขาวิชา วิทยาศาสตร์ระดับโมเลกุลทางจุลชีววิทยาทางการแพทย์และวิทยาภูมิคุ้มกัน ลายมือชื่อนิสิต .....

ปีการศึกษา 2562 ลายมือชื่อ อ.ที่ปรึกษาหลัก .....

5976756637 : MAJOR MOLECULAR SCIENCE OF MEDICAL MICROBIOLOGY AND IMMUNOLOGY

KEYWORD: predictive model, machine learning, hemoglobin, blood bank, data mining, health informatics

Sathit Tadsomboon : Application of machine learning techniques to develop data mining models for predicting hemoglobin levels of blood donors. Advisor: Asst. Prof. Tewarit Sarachana, Ph.D.

Adequate blood and blood components stored in blood banks are critical for patients. It is therefore necessary for blood banks to encourage healthy blood donors to donate their blood regularly. Thailand’s National Blood Centre has to reject as many as 15-20% of potential donors, mostly due to low hemoglobin (Hb) levels. In many cases, such rejections lead to unsatisfactory feelings, causing inadequate of blood. Therefore, it would be helpful for all, if the Hb levels of blood donors can be predicted prior to traveling to the donation sites. This study aims to develop and compare the classification efficiency of machine learning techniques in predicting Hb levels of blood donors. We obtained the information of blood donors (n = 2,180 cases) who visited 13 Regional Blood Centers from 1st October 2018 to 31st May 2018 related to as many as 44 aspects. Following data cleaning and predictive models analyses of blood donor data were performed to predict the Hb levels which indicated acceptance/rejection of blood donation. We found that decision tree show respective accuracy, sensitivity, specificity, positive predictive value, negative predictive value and AUC were 92.20%, 82.98%, 94.74%, 81.25%, 95.29%, and 0.943, respectively. This study provides the information about the efficiency of decision tree analyses as predictive tools, which warrant further research in the future and could lead to further development of an online assessment application.

Field of Study: Molecular Science of Medical Microbiology and Immunology Student's Signature .....

Academic Year: 2019 Advisor's Signature .....

กิตติกรรมประกาศ

ขอกราบขอบพระคุณ ผศ.ดร. เทวฤทธิ์ สะระชนะ อาจารย์ที่ปรึกษาวิทยานิพนธ์ เป็นอย่างสูง ยิ่งที่ได้กรุณาให้คำแนะนำในการดำเนินการวิจัยตลอดจนเสียสละเวลาในการให้คำปรึกษา ตรวจแก้ไขวิทยานิพนธ์ ทั้งยังให้ความช่วยเหลือช่วยวางแผนการศึกษา ติดตามความก้าวหน้าจนทำให้การจัดทาวิทยานิพนธ์เสร็จสมบูรณ์ด้วยดี

ขอกราบขอบพระคุณ ดร.สุนทรี กุลกีรติยุต ดร.ลิขิต ปรียานนท์และอาจารย์ทัศนีย์ สกุลดำรงค์พานิช คณะกรรมการสอบวิทยานิพนธ์เป็นอย่างยิ่ง ที่กรุณาให้คำแนะนำต่าง ๆ ในการปรับปรุงการดำเนินงานการแก้ไขปัญหา เพื่อให้การทำงานเป็นไปอย่างราบรื่น

ขอขอบพระคุณ คณาจารย์ทุกท่านในภาควิชาเวชศาสตร์การธนาคารเลือดและจุลชีววิทยาคลินิก และคณาจารย์จากภาควิชาอื่น ๆ ของคณะสหเวชศาสตร์ จุฬาลงกรณ์มหาวิทยาลัยที่ให้ความรู้อย่างดียิ่งตลอดหลักสูตร

ขอขอบพระคุณ หัวหน้าภาคบริการโลหิตแห่งชาติทุกท่านและสถานีกาชาดหัวหินเฉลิมพระเกียรติ ที่ได้ให้ความอนุเคราะห์เก็บข้อมูลผู้บริจาคโลหิต เพื่อนำมาใช้ในวิทยานิพนธ์นี้

ขอขอบคุณ ศูนย์บริการโลหิตแห่งชาติ สภากาชาดไทย ที่ได้ให้ทุนในการศึกษาแก่ผู้วิจัย

สุดท้ายนี้ขอขอบพระคุณ บิดา มารดา ที่ได้อบรมสั่งสอนเลี้ยงดูผู้วิจัยมาด้วยความรัก ความเมตตาและขอขอบคุณ เพื่อน ๆ พี่ ๆ น้อง ๆ ทุกท่านที่ช่วยเหลือ ให้กำลังใจในการศึกษาด้วยดีเสมอมา

สาธิต เทศสมบูรณ์

สารบัญ

หน้า

สารบัญตาราง

หน้า

สารบัญรูปภาพ

หน้า

บทที่ 1
บทนำ

1.1 ที่มาและความสำคัญของปัญหา

เป็นที่ทราบดีว่าโลหิตและส่วนประกอบโลหิตเป็นสิ่งสำคัญในการรักษาผู้ป่วย โลหิตที่มีความปลอดภัยในระดับที่เป็นมาตรฐานนั้น ต้องเป็นโลหิตที่ได้รับจากผู้บริจาคโลหิตที่มีความเสี่ยงต่ำ(1) ผ่านการตรวจวิเคราะห์ทางห้องปฏิบัติการ เตรียมเป็นส่วนประกอบโลหิตที่มีคุณภาพ ผ่านการตรวจเข้ากันได้ของเลือดระหว่างผู้ป่วยและผู้บริจาค จนถึงการนำไปให้ผู้ป่วยอย่างถูกต้อง โดยตลอดกระบวนการต้องมีการควบคุมห่วงโซ่ความเย็นตลอดกระบวนการ เพื่อรักษาสภาพของโลหิตให้มีคุณภาพ เป็นที่ยอมรับทั่วโลกว่าโลหิตที่มีความปลอดภัยสูงสุดคือโลหิตที่ได้รับจากผู้บริจาคโลหิตประจำ ไม่หวังสิ่งตอบแทน ดังนั้นนอกจากการพยายามเพิ่มจำนวนผู้บริจาคโลหิตรายใหม่ หน่วยงานที่ทำหน้าที่ในการจัดหาโลหิตจำเป็นต้องทำการรณรงค์ประชาสัมพันธ์รักษาผู้บริจาคโลหิตประจำให้สามารถบริจาคโลหิตได้ต่อเนื่อง จะทำให้ได้โลหิตที่มีคุณภาพ มีปริมาณเพียงพออย่างสม่ำเสมอ

ภาคบริการโลหิตแห่งชาติที่ 8 จังหวัดนครสวรรค์ ทำหน้าที่จัดหาโลหิตให้เพียงพอต่อความต้องการใช้ของผู้ป่วยของโรงพยาบาลต่าง ๆ ในเขตจังหวัดนครสวรรค์ อุทัยธานี พิจิตร ชัยนาท ตาก และกำแพงเพชร จากผลการดำเนินงานในปี 2560 พบว่ามีอัตราการปฏิเสธการรับบริจาคโลหิตสูงถึงร้อยละ 23.9 ของจำนวนผู้ประสงค์จะบริจาคโลหิต โดยร้อยละ 50 ของผู้ถูกปฏิเสธ มีสาเหตุมาจากค่าฮีโมโกลบินต่ำกว่าเกณฑ์(2) ซึ่งสามารถพบได้บ่อยในผู้บริจาคโลหิตที่บริจาคโลหิตเป็นประจำ จากปัญหาดังกล่าวนอกจากจะทำให้การจัดหาโลหิตไม่เป็นไปตามเป้าหมายแล้ว ยังส่งผลให้จำนวนผู้บริจาคโลหิตรายเก่ามีจำนวนน้อยลงในทุกปี(3, 4) ดังนั้นหากทราบปัจจัยที่เกี่ยวข้องและสัมพันธ์กับค่าฮีโมโกลบินต่ำกว่าเกณฑ์ในผู้บริจาคโลหิตรายเก่าและพยากรณ์ผลตรวจได้ล่วงหน้า เจ้าหน้าที่สามารถให้คำแนะนำแก่ผู้บริจาคโลหิตให้ตระหนักถึงการดูแลสุขภาพก่อนมาบริจาคโลหิตครั้งต่อไป ทำให้สามารถบริจาคโลหิตได้ยาวนานขึ้น อีกทั้งยังสามารถนำข้อมูลที่ได้มาหาแนวทางในการรณรงค์การบริจาคโลหิตเพื่อป้องกันปัญหาการถูกปฏิเสธรับบริจาคโลหิต ช่วยให้ผู้บริจาคโลหิตไม่เสียเวลาค่าใช้จ่ายในการเดินทางมาบริจาคโลหิตแต่ไม่สามารถบริจาคได้ ลดความผิดหวังในการทำความดี สำหรับหน่วยงานที่ทำหน้าที่จัดหาโลหิตเป็นการช่วยให้มีจำนวนโลหิตเพียงพอและช่วยผู้ป่วยได้รับโลหิตที่มีความปลอดภัยด้วย

ปัจจัยที่มีผลต่อค่าการตรวจฮีโมโกลบินในผู้บริจาคโลหิตนั้นอาจแบ่งออกเป็น 2 ปัจจัยหลัก คือปัจจัยภายใน เช่น เพศ อายุ น้ำหนัก พฤติกรรมการบริโภค เจตคติภายใน การคล้อยตามกลุ่มอ้างอิง การควบคุมพฤติกรรม(5-7) และปัจจัยภายนอกเช่น ความถี่ในการบริจาค ลักษณะทางสังคม รายได้ การศึกษา เป็นต้น ดังนั้นในการศึกษาจำเป็นต้องคำนึงถึงปัจจัยต่าง ๆ ที่อาจเกี่ยวข้องทั้งหมด

จากปัญหาดังกล่าวผู้วิจัยจึงมีแนวคิดประยุกต์ใช้การเรียนรู้ของคอมพิวเตอร์ (machine learning; ML) ในการพัฒนาตัวแบบพยากรณ์ค่าฮีโมโกลบินในผู้บริจาคโลหิตรายเก่า โดยอาศัยข้อมูลผู้บริจาคโลหิตที่ภาคบริการโลหิตแห่งชาติ 12 แห่งและสถานีกาชาดหัวหินเฉลิมพระเกียรติ

1.2 วัตถุประสงค์ของงานวิจัย

เพื่อพัฒนาตัวแบบพยากรณ์ด้วยโปรแกรม RapidMiner และเปรียบเทียบประสิทธิภาพในการพยากรณ์ผลการตรวจฮีโมโกลบินในผู้บริจาคโลหิต โดยใช้การวิเคราะห์ด้วยเทคนิคการเรียนรู้ของคอมพิวเตอร์ (machine learning) แบบต่าง ๆ ได้แก่ ต้นไม้ตัดสินใจ (decision tree) ซัพพอร์ตเวกเตอร์แมชชีน (Support vector machine) การจำแนกแบบเบย์อย่างง่าย (Naïve Bayesain classifier) และโครงข่ายประสาทเทียม (Artificial Neural Networks)

1.3 ขอบเขตงานวิจัย

1. งานวิจัยนี้เป็นการจำแนกกลุ่มค่าฮีโมโกลบินในผู้บริจาคโลหิตเท่านั้น
2. งานวิจัยนี้เป็นการจำแนกกลุ่มค่าฮีโมโกลบินเป็น 2 กลุ่มคือ ผ่านเกณฑ์และไม่ผ่านเกณฑ์เท่านั้น
3. ข้อมูลที่นำมาใช้ในการวิจัยนี้ นำมาจากข้อมูลผู้บริจาคโลหิตของภาคบริการโลหิตแห่งชาติและงานรับบริจาคโลหิต สถานีกาชาดหัวหินเฉลิมพระเกียรติ
4. งานวิจัยนี้เป็นการนำข้อมูลของผู้บริจาคโลหิตมาใช้พยากรณ์การจำแนกกลุ่มค่าฮีโมโกลบินในปัจจุบันเท่านั้น

1.4 ขั้นตอนและวิธีการดำเนินงานวิจัย

1. ศึกษางานวิจัยที่เกี่ยวข้องกับการเรียนรู้ของคอมพิวเตอร์
2. ศึกษาความรู้และทฤษฎีที่เกี่ยวข้องกับงานวิจัย
3. ศึกษาการสร้างตัวแบบพยากรณ์ด้วยโปรแกรม RapidMiner
4. ออกแบบกระบวนการสร้างตัวแบบพยากรณ์
5. เก็บข้อมูลผู้บริจาคโลหิต

6. การทำความสะอาดข้อมูลและการแปลงค่า
7. สร้างตัวแบบพยากรณ์การจำแนกกลุ่ม
8. การฝึกหัดตัวแบบพยากรณ์ในการจำแนกกลุ่ม
9. ประเมินประสิทธิภาพของตัวแบบพยากรณ์
10. วิเคราะห์ผลการประเมินประสิทธิภาพ
11. สรุปผลและจัดทำเล่มวิทยานิพนธ์

1.5 ประโยชน์ที่คาดว่าจะได้รับ

หากผลการศึกษานี้เป็นไปตามสมมุติฐานนั้นคือผลการตรวจฮีโมโกลบินของผู้บริจาคโลหิตสามารถพยากรณ์ผลได้ล่วงหน้า จะก่อให้เกิดประโยชน์ในการดูแลผู้บริจาคโลหิต โดยเจ้าหน้าที่ทางการแพทย์สามารถให้คำแนะนำพิเศษเพิ่มเติมในกลุ่มที่มีผลการพยากรณ์อาจมีค่าฮีโมโกลบินไม่ผ่านเกณฑ์ในการบริจาคครั้งต่อไป ทำให้ผู้บริจาคโลหิตมีความตระหนักในการดูแลสุขภาพ ลดปัญหาการขาดแคลนโลหิตได้อีกทางหนึ่ง นอกจากนี้ยังเป็นข้อมูลพื้นฐานในการพัฒนาระบบประเมินตนเองล่วงหน้าออนไลน์ก่อนเดินทางมาบริจาคโลหิตของผู้ประสงค์จะบริจาคโลหิต หากพบว่าผลพยากรณ์มีโอกาสไม่ผ่านเกณฑ์จะช่วยละภาระค่าใช้จ่ายในการเดินทาง การเสียเวลาเดินทางโดยไม่จำเป็น ลดความผิดหวังของผู้บริจาคโลหิตที่ตั้งใจทำความดี นอกจากนี้ยังลดการใช้งบประมาณของหน่วยงานที่ทำหน้าที่ในการจัดหาโลหิตหากมีผู้ที่ถูกปฏิเสธให้บริจาคโลหิตน้อยลง ถือเป็นการช่วยลดการใช้งบประมาณของประเทศในการจัดหาโลหิต

1.6 โครงสร้างของเนื้อหาวิทยานิพนธ์

โครงสร้างของเนื้อหาวิทยานิพนธ์ประกอบไปด้วย 5 บท มีรายละเอียดดังต่อไปนี้

บทที่ 1 กล่าวถึงที่มาและความสำคัญของปัญหา วัตถุประสงค์ ขอบเขตของงานวิจัย ขั้นตอนและวิธีการดำเนินงานวิจัย ประโยชน์ที่คาดว่าจะได้รับ

บทที่ 2 กล่าวถึงทฤษฎีและงานวิจัยที่เกี่ยวข้อง

บทที่ 3 กล่าวถึงแนวคิดและวิธีการดำเนินงาน

บทที่ 4 กล่าวถึงการทดสอบและประเมินผลงานวิจัย

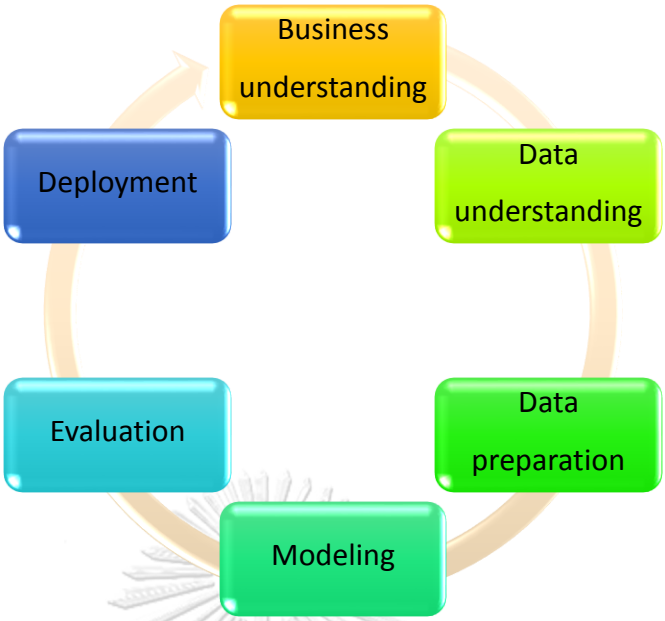
บทที่ 5 กล่าวถึงบทสรุป

บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 ทฤษฎีที่เกี่ยวข้อง

2.1.1 การเรียนรู้ของเครื่องจักร (Machine learning)

การเรียนรู้ของเครื่องจักร (Machine learning; ML) หมายถึงการนำข้อมูลต่าง ๆ มาฝึกให้เครื่องจักรหรือคอมพิวเตอร์ได้เรียนรู้หาคำตอบ การเรียนรู้หมายถึงการเรียนรู้วิธีหาคำตอบจากข้อมูลฝึกหัด จากนั้นจึงนำวิธีการหาคำตอบที่ได้ไปใช้พยากรณ์คาดการณ์สิ่งที่จะเกิดขึ้นจากข้อมูลชุดใหม่ ต่างจากการทำเหมืองข้อมูลเพราะการทำเหมืองข้อมูลเป็นการวิเคราะห์สืบค้นหาความรู้ที่เป็นประโยชน์และมีความสำคัญที่แฝงอยู่ในฐานข้อมูลขนาดใหญ่ที่ยังไม่ทราบคำตอบหรือความสัมพันธ์ของปัจจัยต่าง ๆ ที่แฝงอยู่(8) ในการฝึกให้คอมพิวเตอร์เรียนรู้จะใช้ข้อมูลในอดีตจำนวนมากนำมาวิเคราะห์ด้วยเทคนิคของเหมืองข้อมูล เพื่อสร้างวิธีการเรียนรู้ที่เป็นต้นแบบใช้พยากรณ์หรือทำนายสิ่งที่อาจจะเกิดในอนาคต โดยกระบวนการเรียนรู้ประกอบด้วย 6 ขั้นตอน คือ 1) การทำความเข้าใจการวิจัยและธุรกิจ เป็นขั้นตอนที่ช่วยให้สามารถกำหนดวัตถุประสงค์ได้ชัดเจนเข้าใจเป้าหมายในการดำเนินงานเก็บข้อมูลได้ถูกต้อง 2) การทำความเข้าใจข้อมูล ทำให้การเก็บข้อมูลมีความถูกต้องครบถ้วนเกิดความเข้าใจข้อมูลในเบื้องต้นและทราบถึงคุณภาพของข้อมูล 3) การจัดเตรียมข้อมูล เป็นการจัดเตรียมข้อมูลให้พร้อมสำหรับขั้นตอนต่อไปที่จำเป็นต้องแปลงข้อมูลหรือปรับแต่งข้อมูลที่ไม่ครบถ้วน 4) การสร้างตัวแบบ ข้อมูลแต่ละประเภทจำเป็นต้องมีการเลือกใช้ตัวแบบที่เหมาะสมซึ่งมีเทคนิคที่แตกต่างกัน 5) การประเมินตัวแบบ เป็นการทดสอบว่าตัวแบบนั้นเป็นไปตามวัตถุประสงค์หรือไม่ และ 6) การใช้งาน คือการนำตัวแบบไปใช้งานจริง (9)



รูปที่ 1 กรอบแนวคิดขั้นตอนการทำเหมืองข้อมูล (Data mining framework)(9, 10)

การเรียนรู้ของคอมพิวเตอร์อาจแบ่งหลัก ๆ ได้เป็น 3 ประเภท(11-13) คือ

1. การสร้างต้นแบบพยากรณ์หรือทำนาย (Predictive modeling)

อาจเรียกได้อีกอย่างคือการเรียนรู้แบบมีผู้สอน (Supervised learning) เป็นการนำข้อมูลหรือตัวแปรในอดีตที่ทราบคำตอบแล้วมาใช้ฝึกหัด (training data) โดยตัวแปรจะมีคุณสมบัติในการแบ่งแยกผลลัพธ์ ถ้าตัวแปรมีค่าไม่ต่อเนื่องจะเรียกกระบวนการแบ่งแยกว่าการจำแนกกลุ่ม (classification) แต่ถ้าข้อมูลหรือตัวแปรมีค่าต่อเนื่องจะเรียกการแบ่งว่าการถดถอย (regression) หรือการพยากรณ์ (prediction) เช่น การจำแนกด้วยแผนภูมิต้นไม้ (decision tree) โครงข่ายประสาทเทียม (neural network) เป็นต้น

2. การสร้างต้นแบบในการพรรณนาอธิบายหรือบรรยาย (Descriptive modeling)

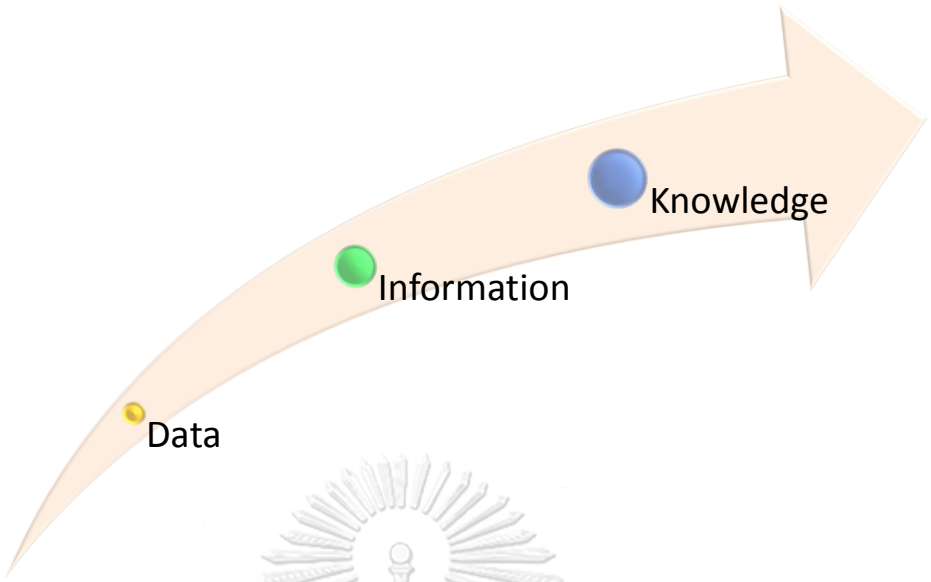
คือการเรียนรู้แบบไม่มีผู้สอน (Unsupervised learning) เป็นการนำข้อมูลที่ยังไม่ทราบคำตอบหรือความสัมพันธ์ระหว่างปัจจัยต่าง ๆ นำมาศึกษาหาความสัมพันธ์ (association) หรือการจัดกลุ่ม (clustering) โดยไม่ได้มีจุดมุ่งหมายเพื่อการทำนายหรือพยากรณ์คำตอบในอนาคตแต่เป็นการพยายามหาความสัมพันธ์ของข้อมูล เช่น การจัดกลุ่มโครงข่ายโคโฮเนน หรือการหากฎความสัมพันธ์ เป็นต้น

3. การสร้างการเรียนรู้แบบเสริมกำลัง (Reinforcement Learning; RL)

คือการสร้างการเรียนรู้ของคอมพิวเตอร์จากการลองผิดลองถูกภายใต้สถานการณ์ต่าง ๆ ส่วนใหญ่ใช้ในการพัฒนาเกมส์หรืองานด้าน Robot โดยอาศัยพื้นฐานกระบวนการตัดสินใจของมาร์คอฟ (Markov decision process; MDP) โดยระบบการเรียนรู้ที่เราสร้างจะเรียกว่า Agent หรือ Robot ซึ่ง MDP ประกอบด้วย A, S, R และ P โดย A=Action คือทางเลือกที่ Agent สามารถเลือกทำได้ S=State คือสถานการณ์หรือรับรู้ปัจจุบันของ Agent R=Reward คือผลของ Action โดยมีเป้าหมายทำให้ได้มากที่สุดและ P= Probabilities คือความน่าจะเป็นของย้าย State จาก S_0 ไป S_1 ที่เกิดจากการกระทำ A_0 ของ Agent ซึ่งตัวแบบการเรียนรู้ชนิดนี้จะอาศัยการทำซ้ำ ๆ ใน Action ต่าง ๆ จะกว่าจะได้ R สูงสุดนั้นก็คือเป้าหมายหรือการชนะของเกมส์นั้นเองตัวอย่างการพัฒนา RL เช่น AlphaGO เป็นต้น

2.1.2 ข้อมูล (Data)

ข้อมูล (data) คือ ข้อเท็จจริงของสิ่งต่าง ๆ ในโลกที่อยู่รอบตัวเราไม่ว่าจะเป็นคน สัตว์ สิ่งของหรือเหตุการณ์ ซึ่งข้อมูลจะอยู่ในรูปแบบต่าง ๆ ที่ได้จากการสังเกตบันทึกทั้งข้อมูลที่มีโครงสร้างและไม่มีโครงสร้างเมื่อข้อมูลถูกตรวจวัดกำหนดค่าหรือแปลความหมายของข้อมูลเพื่อนำไปใช้ประโยชน์จะเรียกว่าข้อมูลสารสนเทศ (information) เมื่อนำข้อมูลสารสนเทศถูกนำมาช่วยในการคิดหรือตัดสินใจและกำหนดเป็นหลักเกณฑ์หรือเงื่อนไขในการกระทำใด ๆ ข้อมูลสารสนเทศนั้นจะเรียกว่าความรู้ (knowledge) สามารถสรุปการเปลี่ยนรูปของข้อมูลได้ดังรูปที่ 2 โดยทั่วไปข้อมูลจะมีคุณลักษณะที่หลากหลายแต่สามารถแบ่งได้เป็น 2 ประเภทหลัก ๆ คือ ข้อมูลที่มีคุณสมบัติเป็นเชิงตัวเลข (numerical) และเป็นหมวดหมู่ (categorical) ทั้ง 2 ประเภทยังสามารถแยกย่อยได้อีก เช่น ข้อมูลเชิงตัวเลขที่มีค่าต่อเนื่อง (continuous) ข้อมูลเชิงตัวเลขที่มีค่าไม่ต่อเนื่อง (discrete) ข้อมูลหมวดหมู่ที่แบ่งได้เป็น 2 กลุ่ม (nominal) และที่แบ่งได้มากกว่า 2 กลุ่ม (ordinal) เป็นต้น (14)

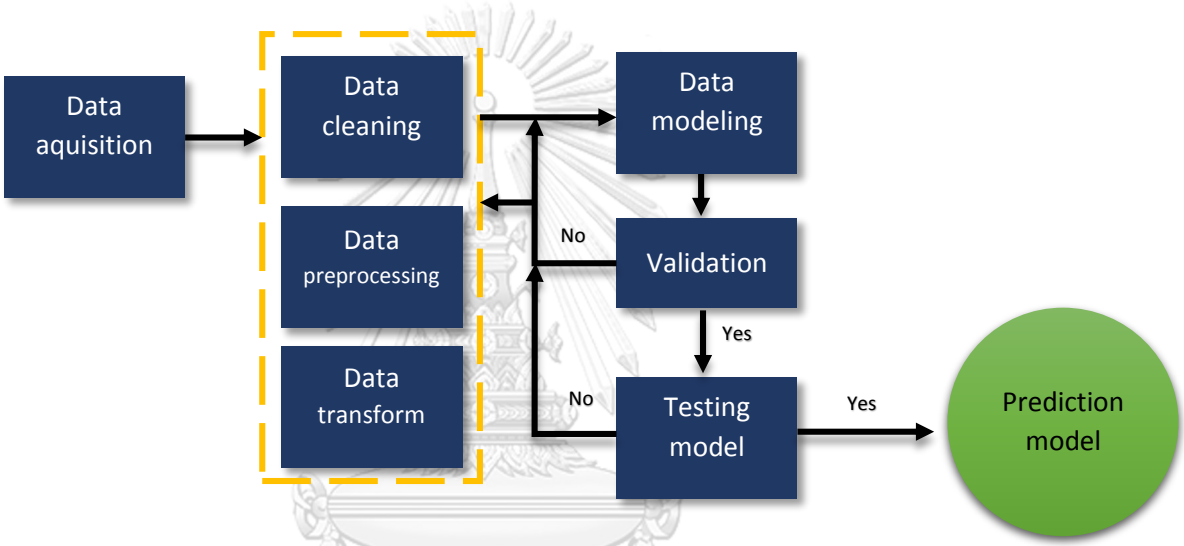


รูปที่ 2 การเปลี่ยนแปลงรูปแบบของข้อมูล(14)

ทุกวันมีข้อมูลเพิ่มขึ้นอย่างมากมายและรวดเร็วจนมนุษย์ไม่สามารถสรุปความรู้จากข้อมูลเหล่านั้นได้ด้วยสมองของมนุษย์เอง ต้องอาศัยคอมพิวเตอร์ในการวิเคราะห์หาคำตอบที่ซ้อนอยู่ในข้อมูลประเภทต่าง ๆ (14) ถ้าข้อมูลนั้นมีความซับซ้อนมากการจะทำให้คอมพิวเตอร์วิเคราะห์หาคำตอบได้ถูกต้อง ข้อมูลที่นำมาใช้สำหรับการเรียนรู้ของคอมพิวเตอร์ (machine learning) ต้องถูกทำให้เหมาะสมต่อการวิเคราะห์ เรียกว่ากระบวนการจัดเตรียมข้อมูล (data preparation) แบ่งออกเป็น 3 ขั้นตอนประกอบด้วย 1) cleaning data เป็นขั้นตอนตรวจสอบข้อมูลเบื้องต้นเพื่อกำจัดข้อมูลที่ไม่ถูกต้อง มีค่าเกินช่วงยอมรับหรือข้อมูลที่มีการลงบันทึกผิดพลาด 2) preprocessing data เป็นการปรับแต่งค่าให้เหมาะสมเช่นการเติมค่าที่ขาดหาดไปโดยไม่มีผลกระทบต่อความถูกต้องของข้อมูลและ 3) transformations data เป็นขั้นตอนการแปลงค่าของข้อมูลให้อยู่ในรูปแบบที่คอมพิวเตอร์สามารถนำไปคำนวณหรือวิเคราะห์ได้ก่อนนำไปสร้างตัวแบบพยากรณ์

ในการสร้างตัวแบบพยากรณ์จำเป็นต้องมีการฝึกสอนให้คอมพิวเตอร์ได้เรียนรู้และทดสอบเงื่อนไขในค้นหาคำตอบหรือความสัมพันธ์ของข้อมูลกับผลลัพธ์ เรียกข้อมูลที่ใช้ฝึกสอนว่าชุดข้อมูล (dataset) ซึ่งในการทำการเรียนรู้ของคอมพิวเตอร์ ชุดข้อมูลจะถูกแบ่งออกเป็น 3 ชุดได้แก่ 1) ชุดข้อมูลฝึกหัด (training dataset) ใช้ในการทดลองฝึกหัดให้คอมพิวเตอร์ได้เรียนรู้เงื่อนไขในการ

ค้นหาคำตอบ ซึ่งจะแบ่งประมาณร้อยละ 70-75 ของข้อมูล 2) ชุดข้อมูลประเมินผล (validation dataset) เป็นข้อมูลใช้ประเมินความถูกต้องของเงื่อนไขที่ได้จากการเรียนรู้ซึ่งจะมีประมาณร้อยละ 10-15 ของข้อมูล ในการเรียนรู้จะทำซ้ำขั้นตอน 1 และ 2 จนกว่าจะได้เงื่อนไขการเรียนรู้ที่ให้คำตอบที่ถูกต้องที่สุด 3) ชุดข้อมูลทดสอบ (testing dataset) เป็นชุดข้อมูลที่ถูกแบ่งออกมาเพื่อทดสอบเงื่อนไขที่ได้จากการฝึกหัดเพื่อประเมินความถูกต้องของคำตอบ โดยชุดข้อมูลทดสอบจะมีประมาณร้อยละ 15-20 ของข้อมูล(7)



รูปที่ 3 แผนผังขั้นตอนการสร้างตัวแบบพยากรณ์(15)

2.1.3 การคัดเลือกคุณสมบัติ (Feature Selection)

ในการค้นหาความสัมพันธ์ของตัวแปรพยากรณ์หลายตัวกับ 1 ตัวแปรตาม วิธีที่ได้รับความนิยมคือการสร้างสมการถดถอยเชิงเส้นพหุ (multiple linear regression) เป็นการหาความสัมพันธ์เชิงเส้นของตัวแปรพยากรณ์หลายตัวกับตัวแปรตามโดยมีวัตถุประสงค์พยายามพยากรณ์ค่าของตัวแปรตาม ในการสร้างตัวแบบความสัมพันธ์มักพบว่าบางตัวแปรไม่มีส่วนช่วยในการทำนายหรือเมื่อนำเข้ามาคำนวณแล้วทำให้ค่าการพยากรณ์แย่ลง(16-18)หรือตัวแปรพยากรณ์มีความสัมพันธ์กันเองเกิดปัญหาสหสัมพันธ์เชิงเส้นพหุ (multicollinearity) ทำให้ค่าพยากรณ์มีความคลาดเคลื่อนสูง แต่ในทางกลับกันถ้าตัวแบบพยากรณ์ขาดตัวแปรพยากรณ์ที่สำคัญไปก็จะทำ

ให้เกิดความคลาดเคลื่อนสูงเช่นกัน จึงจำเป็นต้องทำการคัดเลือกตัวแปรพยากรณ์ที่เหมาะสมเข้าสู่ตัวแบบ(19) วิธีการคัดเลือกตัวแปรที่นิยมแบ่งได้เป็น 4 วิธี คือ

2.1.3.1 การเลือกตัวแปรโดยวิธีเพิ่มตัวแปร (Forward Selection)

เป็นวิธีที่ต้องการได้โมเดลประหยัดคือจะเลือกเฉพาะตัวแปรพยากรณ์ที่ดีที่สุดซึ่งพิจารณาจากค่า R สูงสุดเข้าสู่สมการตัวแรก จากนั้นจะคัดเลือกตัวแปรตัวต่อไปที่ให้ค่าพยากรณ์สูงขึ้นเข้าสมการ จะทำซ้ำไปจนครบทุกตัวแปรกระทั่งไม่เหลือตัวแปรพยากรณ์หรือเหลือเฉพาะตัวแปรที่นำเข้ามาแล้วทำให้ค่าพยากรณ์แย่ลงกว่าเกณฑ์ยอมรับ วิธีการเพิ่มตัวแปรก็จะสิ้นสุด(17, 18, 20) จุดบกพร่องของวิธีนี้คือไม่ได้ตรวจสอบผลกระทบที่เกิดจากตัวแปรพยากรณ์ตัวใหม่ที่เข้าไปในสมการแล้วทำให้ตัวแปรเก่าที่นำเข้ามาก่อนมีค่าพยากรณ์แย่ลงหรือไม่

2.1.3.2 การเลือกตัวแปรโดยวิธีลดตัวแปร (Backward Elimination)

เป็นวิธีตรงข้ามกับ forward แต่เป็นวิธีคัดเลือกตัวแปรที่ดีและได้โมเดลประหยัดเช่นเดียวกัน โดยเริ่มต้นจะนำตัวแปรพยากรณ์ทุกตัวเข้ามาในสมการและพิจารณาตัวแปรพยากรณ์ที่มีค่าสัมประสิทธิ์สหสัมพันธ์บางส่วน (Partial Correlation) กับตัวแปรตาม โดยควบคุมอิทธิพลของตัวแปรพยากรณ์อื่น ๆ ซึ่งมีค่าต่ำที่สุดออกจากสมการ (18, 20) แล้วจึงดำเนินการทดสอบว่าค่า R2 ลดลงอย่างมีนัยสำคัญทางสถิติหรือไม่ ถ้าพบว่าลดลงอย่างไม่มีนัยสำคัญทางสถิติแสดงว่าตัวแปรพยากรณ์ ดังกล่าวไม่ได้มีส่วนทำให้การพยากรณ์ตัวแปรตามดีขึ้นสามารถขจัดออกจากสมการได้

2.1.3.3 การเลือกตัวแปรโดยวิธีเพิ่มตัวแปรอิสระแบบขั้นตอน (Stepwise Regression)

เป็นวิธีที่มีความเหมาะสมในการพิจารณาคัดเลือกตัวแปรพยากรณ์ที่ดีที่สุดและได้โมเดลที่ประหยัดที่สุด ซึ่งลำดับขั้นตอนจะคล้ายกับวิธี Forward เพียงแต่การวิเคราะห์ด้วยวิธี Stepwise จะทำการ ทดสอบตัวแปรพยากรณ์ที่เข้าสมการไปแล้วทุกครั้งที่มีการนำตัวแปรใหม่เข้าในสมการหมายความว่าตัวแปรพยากรณ์บางตัวที่เข้าไปในสมการแล้วก็สามารถถูกขจัดออกจากสมการได้ หากพบว่าตัวแปรพยากรณ์ตัวนั้นไม่ได้ส่งผลให้ค่า R2 เพิ่มขึ้นอย่างมีนัยสำคัญทางสถิติ(17, 18, 21, 22)

2.1.3.4 การเลือกตัวแปรโดยวิธีนำตัวแปรเข้าทั้งหมด (Enter Regression)

วิธีที่ไม่ได้มีการคัดเลือกตัวแปร โดยนำตัวแปรพยากรณ์ทุกตัวเข้าสมการเป็นการวิเคราะห์เพียงขั้นตอนเดียว เริ่มต้นการวิเคราะห์โดยใช้ตัวแปรพยากรณ์ที่ศึกษานำเข้าสมการพยากรณ์พร้อมกันทุกตัว ถึงแม้ว่าตัวแปรพยากรณ์จะไม่ทำให้ค่าการพยากรณ์ดีขึ้นก็ตาม วิธีนี้มักจะใช้ในกรณีที่ต้องการทราบว่าตัวแปรแต่ละตัวที่ทำการศึกษาจะสามารถพยากรณ์ตัวแปรตามได้หรือไม่มากน้อยเพียงใด ข้อด้อยคือ เป็นการวิเคราะห์ที่ไม่ได้คัดเลือกตัวแปรเข้าสู่สมการถดถอยที่เหมาะสมทำให้เป็นวิธีที่ไม่ประหยัด

2.1.4 อัลกอริทึมการเรียนรู้ของคอมพิวเตอร์ (Machine learning algorithms)

ในการเลือกใช้เทคนิคการทำเหมืองข้อมูล (data mining technique) ที่นำมาใช้ในการเรียนรู้ของคอมพิวเตอร์ (machine learning) นอกจากจะแบ่งได้เป็นแบบ supervise learning กับ unsupervised learning แล้วยังสามารถแบ่งได้ตามประเภทของข้อมูลและคำตอบที่ต้องการ โดยในการพยากรณ์ข้อมูลแบบ supervise learning ที่ต้องเรียนรู้จากข้อมูลฝึกหัด (training dataset) ก่อนจากนั้นจึงประเมินประสิทธิภาพด้วยชุดข้อมูลทดสอบ (testing dataset) แบ่งได้หลัก 2 ประเภทคือ

2.1.4.1 การพยากรณ์ด้วยสมการถดถอย (Regression)

เป็นอัลกอริทึมที่ใช้สำหรับทำนายหรือหาความสัมพันธ์ของข้อมูล โดยอาศัยข้อมูลเก่าที่เป็นตัวแปรพยากรณ์หนึ่งตัวหรือหลายตัวต่อตัวแปรตามหรือตัวแปรเป้าหมายประเภทตัวเลขหนึ่งตัว ด้วยการหาความสัมพันธ์จากสมการถดถอยเชิงเส้นพหุ (multiple linear regression) วิธีนี้ประสิทธิภาพในการพยากรณ์ขึ้นอยู่กับการคัดเลือกตัวแปรพยากรณ์ที่เหมาะสมเข้าสมการ โดยพิจารณาจากค่า F ซึ่งได้จากอัตราส่วนระหว่างค่ากำลังเฉลี่ยของการถดถอยต่อค่ากำลังสองเฉลี่ยของความคลาดเคลื่อน

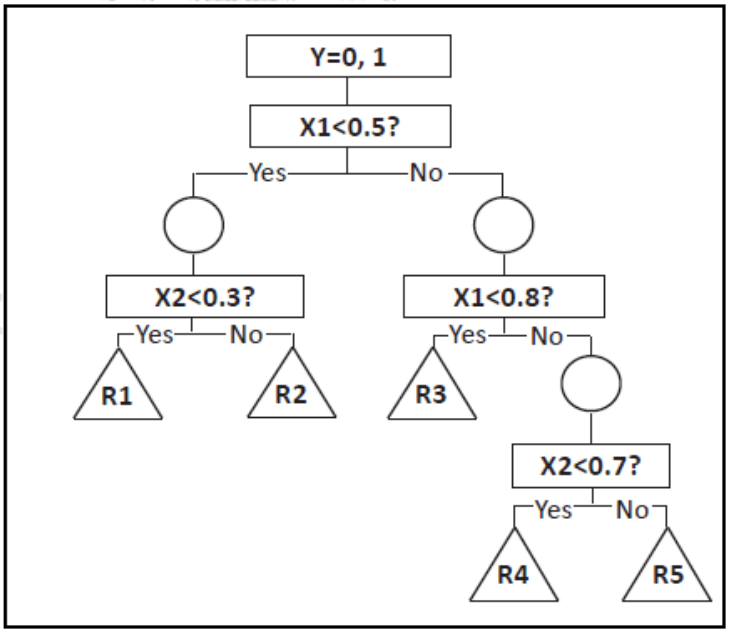
2.1.4.2 การพยากรณ์ด้วยการจำแนกกลุ่ม (Classification)

เป็นวิธีที่คล้ายกับการพยากรณ์ด้วยสมการถดถอย (regression) แต่ต่างกันที่ตัวแปรตามหรือตัวแปรเป้าหมายของการพยากรณ์ด้วยการจำแนกกลุ่มไม่ใช่ค่าตัวเลขที่มีค่าต่อเนื่องเช่น ค่าบวกลบ ตัวแปรที่มีค่า ลำดับชั้น เป็นต้น การจำแนกกลุ่มนั้นมีหลาย

วิธีการเลือกใช้ขึ้นอยู่กับปัญหาของงานและคำตอบที่ต้องการจึงจะเลือกอัลกอริทึมที่สามารถแก้ปัญหานั้นได้

2.1.4.2.1 ต้นไม้ตัดสินใจ (Decision Tree)

เรียกอีกอย่างว่าต้นไม้จำแนก (classification tree) เป็นวิธีการจำแนกกลุ่มที่ง่ายและเป็นที่นิยมด้วยการรวบรวมโหนดตัดสินใจ (decision node) แล้วเชื่อมต่อกันออกไปด้วยกิ่งก้านคล้ายต้นไม้ โดยโหนดบนสุดเรียกว่าโหนดราก (root node) ส่วนโหนดที่ต่อเป็นกิ่งก้านเรียกว่าโหนดใบ (leaf node) แต่ละโหนดจะทำหน้าที่ทดสอบคุณลักษณะ (attribute) และตัดสินใจเลือกกลุ่มในแต่ละกิ่งก้านจะนำไปสู่โหนดตัดสินใจอีกโหนด ในการเรียนรู้ของต้นไม้ตัดสินใจต้องอาศัยตัวแปรที่มีความสัมพันธ์ที่เป็นระบบและมีระเบียงที่ครบถ้วนจึงจะมีประสิทธิภาพในการจำแนก โดยคำตอบ (class) ของต้นไม้ตัดสินใจจะต้องไม่เป็นค่าต่อเนื่องหมายความว่าตัวแปรตามจะต้องถูกกำหนดค่าไว้ชัดเจน(11, 14, 23, 24)



รูปที่ 4 แผนภูมิต้นไม้ตัดสินใจอย่างง่ายที่ตัวแปรเป้าหมายเป็นประเภท nominal(18)

อัลกอริทึมต้นไม้ตัดสินใจเป็นเทคนิคที่แพร่หลายและสามารถเข้าใจผลลัพธ์ได้ง่าย แต่เทคนิคต้นไม้ตัดสินใจจะจำกัดจำนวนตัวแปรตามเพียงหนึ่งตัวหากข้อมูลมีตัวแปร

ตามมากกว่าหนึ่งตัวจะต้องสร้างตัวแบบพยากรณ์เพิ่มขึ้นตามจำนวนของตัวแปรตาม เทคนิคต้นไม้ตัดสินใจมีหลายเทคนิคให้เลือกใช้เช่น อัลกอริทึม ต้นไม้ตัดสินใจการจำและและการถดถอย (classification and regression tree algorithms; CART), C4.5 (C4.5 algorithms), ID3 (ID3 algorithms) และ C5.0 (C5.0 algorithms) เป็นต้น(25) โดยแต่ละเทคนิคอาศัยหลักการคล้ายกันโดยเลือกตัวแปรพยากรณ์ที่มีค่าเกน (gain) สูงที่สุดเป็นโหนดราก (root node) ซึ่งค่าเกนพิจารณาจากความน่าจะเป็นของคุณสมบัติตัวแปรตามในที่นี้ให้เป็นคลาส P และ N ต่อคุณสมบัติของตัวแปรพยากรณ์สามารถคำนวณได้ดังสมการที่ 1-3(26, 27)

$$I(p,n) = \frac{p}{p+n} log_2 \frac{p}{p+n} - \frac{n}{p+n} log_2 \frac{n}{p+n} \quad (1)$$

I(p,n) คือค่าความสับสนในการในการแยกคลาส P กับ N โดยพิจารณาจากคุณลักษณะต่าง ๆ

$$E(A) = \sum_{i=1}^{v} \frac{pi+ni}{p+n} I(pi,ni) \quad (2)$$

E(A) คือค่าคาดคะเนของข้อมูล (entropy) ที่แยกโดยการใช้คุณลักษณะ A ซึ่งกำหนด A คือ คุณลักษณะที่แบ่งข้อมูลคลาส P ได้จำนวน pi และแบ่งข้อมูลคลาส N ได้จำนวน ni

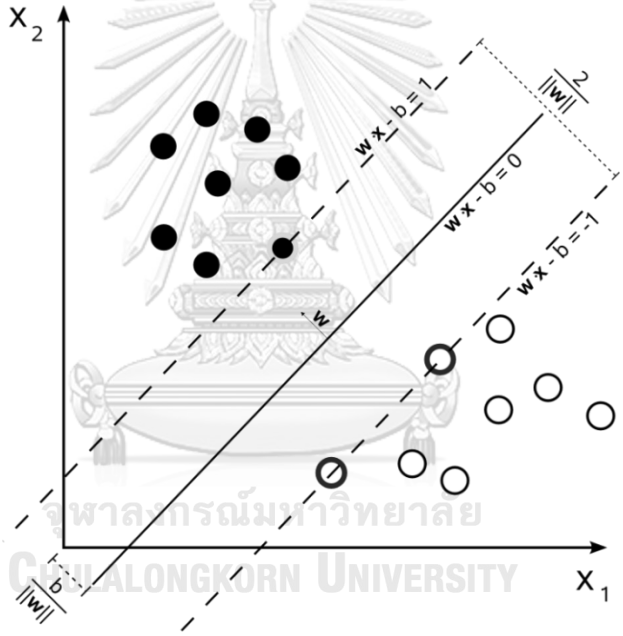
$$Gain(A) = I(p,n) - E(A) \quad (3)$$

Gain(A) คือค่าของคะแนนที่มีหน่วยเป็นบิต ในการพิจารณาจะเลือกค่าเกนสูงสุดเป็น root node และพิจารณาคุณลักษณะตัวต่อไปที่ให้ค่าเกนจาก root node สูงสุดจะถูกเลือกเป็นตัวต่อไป

2.1.4.2.2 ซัพพอร์ตเวกเตอร์แมชชีน (Support vector machine; SVM)

ซัพพอร์ตเวกเตอร์แมชชีนเป็นอัลกอริทึมที่นำมาใช้กันอย่างกว้างขวางในการจำแนกกลุ่ม(15) หลักการของ SVM คือการเพิ่มมิติให้ข้อมูลตัวแปรพยากรณ์ที่ใช้ฝึกหัด โดยกลุ่มของคุณลักษณะของตัวแปรพยากรณ์ที่สามารถอธิบายตัวแปรตามได้จะเรียกว่าเวคเตอร์ (vector) จะแสดงเวคเตอร์ในแบบสเปซ 2 มิติอยู่ในระนาบ xy หรือ 3 มิติในระนาบ xyz โดยทำการสร้างไฮเปอร์เพลน (hyperplane) ด้วยฟังก์ชั่นเครเนล (kernel function) ไฮเปอร์เพลนคือเส้นตรงระนาบที่แยกกลุ่มตัว

แปรตามออกจากกันได้ดีที่สุดนั้นหมายถึงเส้นแบ่งมีระยะห่างจากตัวแปรมากที่สุด ซึ่ง SVM เป็นเทคนิคที่ช่วยในการเรียนรู้ ซึ่งมีลักษณะการทำงานคล้ายกับโครงข่ายประสาทเทียม (artificial neural network) แต่มีความแตกต่างกันคือเทคนิคโครงข่ายประสาทเทียมจะใช้หลักการลดความเสี่ยงเชิงการทดสอบให้มีค่าต่ำที่สุด (empirical risk minimization) ส่วน SVM การแก้สมการหาค่าน้ำหนักใช้ในการแก้สมาการ quadratic ที่มีข้อบังคับเชิงเส้น (linear constrained) ลดความเสี่ยงเชิงโครงสร้างให้ต่ำที่สุด (structural risk minimization) โดย SVM ประยุกต์การใช้งานได้ 2 รูปแบบ คือการวิเคราะห์การถดถอย (regression) หรือการประมาณค่าของฟังก์ชันและการจำแนกประเภท (classification)



รูปที่ 5 ภาพเส้นไฮเปอร์เพลนในการแยกกลุ่มของเวกเตอร์(28)

2.1.4.2.3 การจำแนกแบบเบย์อย่างง่าย (Naïve Bayesain classifier)

การจำแนกแบบเบย์อย่างง่ายเป็นการจำแนกที่อาศัยทฤษฎีความน่าจะเป็นของเบย์ (bayes's theorem) และเป็นสมมุติฐานที่ให้แต่ละเหตุการณ์เป็นอิสระต่อกัน โดยสามารถคำนวณได้จากสมการที่ 4

$$P(h|D) = \frac{[P(D|h) * P(h)]}{P(D)} \quad (4)$$

P(h|D) คือ ความน่าจะเป็นในการเกิดเหตุการณ์ h เมื่อเกิดเหตุการณ์ D

P(D|h) คือ ความน่าจะเป็นในการเกิดเหตุการณ์ D เมื่อเกิดเหตุการณ์ h

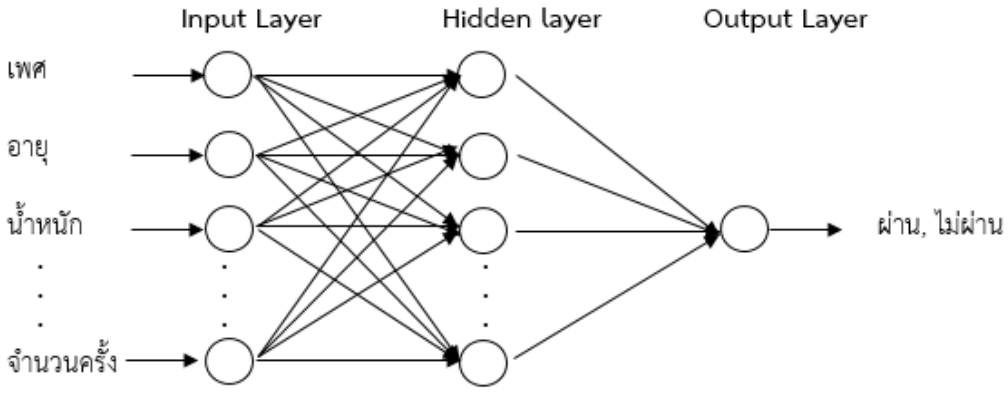
P(h) คือ ความน่าจะเป็นที่จะเกิดเหตุการณ์ h

P(D) คือ ความน่าจะเป็นที่จะเกิดเหตุการณ์ D

การทำงานของอัลกอริทึมการจำแนกแบบเบย์อย่างง่าย (Naïve Bayesain classifier) คือการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรพยากรณ์แต่ละตัวกับตัวแปรตาม นำมาสร้างเงื่อนไขค่าความน่าจะเป็นของแต่ละความสัมพันธ์ แต่การจำแนกแบบเบย์อย่างง่ายจะมีความแม่นยำเมื่อตัวแปรพยากรณ์แต่ละตัวเป็นอิสระต่อกัน ซึ่งในการปฏิบัติตัวแปรพยากรณ์มักพบความสัมพันธ์กันเอง นอกจากนี้การจำแนกแบบเบย์อย่างง่าย (Naïve Bayesain classifier) ยังไม่รองรับข้อมูลตัวแปรที่เป็นค่าต่อเนื่อง ดังนั้นหากข้อมูลตัวแปรเป็นค่าต่อเนื่องจะต้องแปลงเป็นช่วงค่าที่เหมาะสมเพื่อให้การจำแนกมีความแม่นยำ

2.1.4.2.4 โครงข่ายประสาทเทียม (Artificial neural networks: ANN)

เทคนิคโครงข่ายประสาทเทียม (artificial neural network) จุดเด่นคือสามารถที่จะเรียนรู้จดจำสิ่งต่าง ๆ ได้ มีรูปแบบการทำงานเรียนแบบสมองของมนุษย์(24)



รูปที่ 6 แบบโครงข่ายประสาทเทียม (artificial neural network)(29)

เทคนิคโครงข่ายประสาทเทียมชนิด neural network ประกอบด้วย ชั้นรับข้อมูล (input Layer) ชั้นซ่อนเร้น (hidden Layer) และชั้นแสดงผล (output layer) โดยแบ่งข้อมูลเป็น 2 กลุ่มประกอบด้วย training data ประมาณร้อยละ 80

ของข้อมูลและ testing data ร้อยละ 20 ของข้อมูล(14, 30) กำหนดจำนวนโหนดในชั้นรับข้อมูล (input node) ให้มีจำนวนเท่ากับจำนวนตัวแปรกำหนดจำนวนโหนดในชั้นซ่อนเร้น (hidden node) และกำหนดจำนวนโหนดในชั้นแสดงผล (output node) เท่ากับ 2 โหนด กำหนดจำนวนรอบในการเรียนรู้ (Epoch) เช่น 1,000 รอบ ค่าผิดพลาดที่ยอมรับได้เช่น 0.00001 โดยโครงข่ายประสาทเทียมจะให้ค่าน้ำหนักความสัมพันธ์ระหว่างชั้นนำเข้า (input node) ในแต่ละโหนดกับชั้นซ่อนเร้นในแต่ละรอบของการเรียนรู้เพื่อทำนายการจำแนกที่ดีที่สุด

2.1.5 การรับบริจาคโลหิต

การรับบริจาคโลหิตสามารถแบ่งออกเป็น 5 ขั้นตอนหลักได้ดังนี้

1. ขั้นตอนการสมัครบริจาคโลหิต

   เป็นขั้นตอนที่ผู้ประสงค์บริจาคโลหิตต้องกรอกใบสมัครบริจาคโลหิตทั้งข้อมูลทั่วไปและตอบคำถามในใบสมัครเพื่อประเมินสุขภาพ ความพร้อมของตนเองในการบริจาคโลหิต ซึ่งเป็นขั้นตอนที่สำคัญผู้ประสงค์บริจาคโลหิตต้องกรอกข้อมูลและตอบแบบสอบถามด้วยตนเองเท่านั้น

2. ขั้นตอนการตรวจคัดกรองสุขภาพ

   เป็นขั้นตอนการตรวจวัดความดันโลหิต ชีพจรและการซักประวัติ คัดกรองสุขภาพด้วยแพทย์หรือพยาบาลที่ผ่านการฝึกอบรมเพื่อประเมินว่าผู้ประสงค์บริจาคโลหิตมีสุขภาพแข็งแรง ปลอดภัยพร้อมบริจาคโลหิต

3. ขั้นตอนการตรวจหมู่โลหิตและค่าฮีโมโกลบิน

   ผู้ประสงค์บริจาคโลหิตรายใหม่ทุกรายจะต้องผ่านการตรวจหมู่โลหิต ABO เบื้องต้นด้วยวิธีสไลด์ แต่สำหรับผู้ประสงค์บริจาคโลหิตรายเก่าจะใช้ประวัติผลตรวจครั้งที่ผ่านมา แต่ผู้ประสงค์บริจาคโลหิตทุกรายต้องได้รับการตรวจระดับค่าฮีโมโกลบิน โดยสามารถตรวจได้ 2 วิธีคือ 1) ตรวจโดยใช้สารละลายคอปเปอร์ซัลเฟต แบ่งเป็น ความเข้มข้นร้อยละ 80 สำหรับผู้บริจาคโลหิตเพศหญิง และ ความเข้มข้นร้อยละ 90 สำหรับผู้บริจาคโลหิตเพศชาย ซึ่งผลที่ได้จะรายงานเพียงผ่านหรือไม่ผ่านเกณฑ์เท่านั้น 2) ตรวจด้วยเครื่องตรวจค่าฮีโมโกลบิน (Hemoglobin meter) ผลตรวจที่ได้จะเป็นค่าฮีโมโกลบินหน่วย

เป็นมิลลิกรัมต่อเดซิลิตร ซึ่งศูนย์บริการโลหิตแห่งชาติ สภากาชาดไทย ได้กำหนดว่าผู้หญิงต้องมีค่าฮีโมโกลบินมากกว่าหรือเท่ากับ 12.5 มิลลิกรัมต่อเดซิลิตรและผู้ชายต้องมีค่าฮีโมโกลบินมากกว่าหรือเท่ากับ 13 มิลลิกรัมต่อเดซิลิตร จึงจะถือว่าผ่านเกณฑ์สามารถบริจาคโลหิตได้

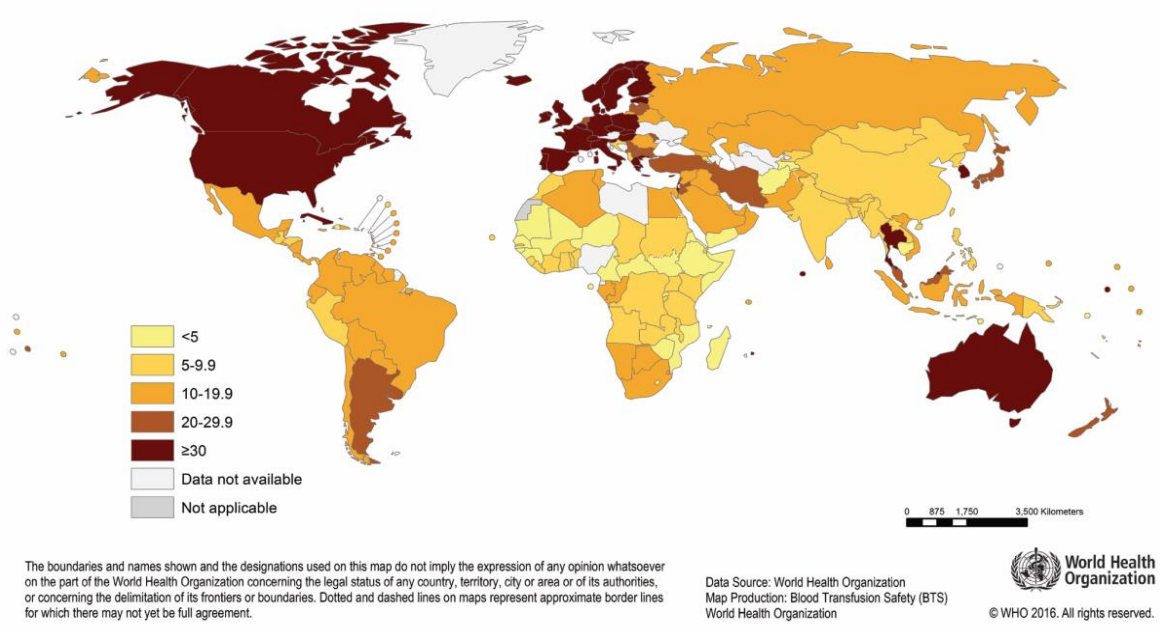
4. ขั้นตอนการเจาะเก็บโลหิต

เป็นการเจาะเก็บโลหิตบริจาค ซึ่งประเทศไทยแบ่งปริมาตรการเจาะเก็บตามน้ำหนักตัวได้ 2 ชนิดคือ โดยผู้ที่น้ำหนักตัวมากกว่าหรือเท่ากับ 45 กิโลกรัมแต่น้อยกว่า 50 กิโลกรัมสามารถบริจาคโลหิตได้ปริมาตร 350 มิลลิลิตรและผู้ที่มีน้ำหนักมากกว่าหรือเท่ากับ 50 กิโลกรัมสามารถบริจาคโลหิตปริมาตร 450 มิลลิลิตร

5. ขั้นตอนการพักหลังบริจาคโลหิต

หลังการบริจาคโลหิตผู้บริจาคโลหิตต้องนั่งพัก ดื่มน้ำหรือน้ำหวานเพื่อเป็นการชดเชยโลหิตที่บริจาคไป ซึ่งเป็นการป้องกันอุบัติเหตุที่อาจเกิดได้หลังบริจาคโลหิต

ในปี 2016 องค์การอนามัยโลกสำรวจข้อมูลการบริจาคโลหิตพบว่าทั่วโลกมีการบริจาคโลหิตประมาณ 112.5 ล้านยูนิต แบ่งเป็นบริจาคแบบโลหิตรวม 100.6 ล้านยูนิตและแบบบริจาคโลหิตเฉพาะส่วน 11.9 ล้านยูนิตและมี 67 ประเทศที่มีการบริจาคน้อยกว่า 10 ยูนิตต่อประชากร 1,000 คน (รูปที่ 2) และพบว่ามีผู้บริจาคโลหิตที่ไม่หวังสิ่งตอบแทนเพิ่มขึ้น 10.7 ล้านยูนิตจากปี 2008 ถึงปี 2013 จาก 159 ประเทศ โดยพบว่าประเทศแถบเอเชียตะวันออกเฉียงใต้มีอัตราเพิ่มขึ้นสูงสุดคือ 75%(1)



รูปที่ 7 จำนวนผู้บริจาคโลหิตต่อจำนวนประชากร 1,000 คน ในปี 2013(1)

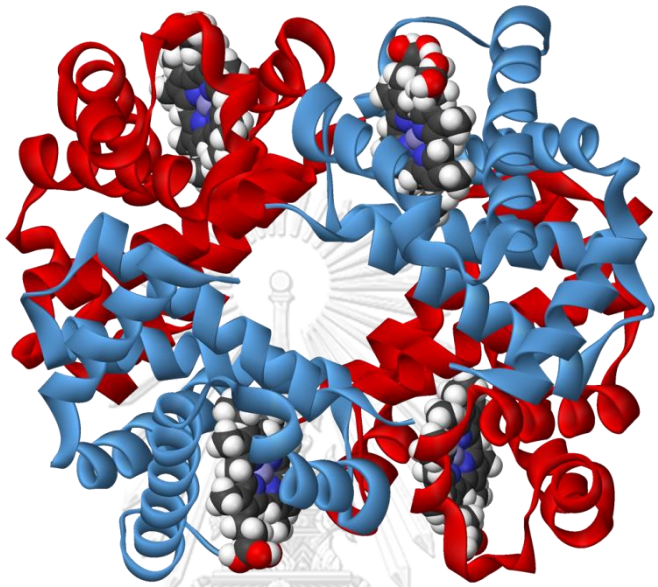
ในประเทศไทย ศูนย์บริการโลหิตแห่งชาติ สภากาชาดไทยเป็นองค์กรหลักในการดำเนินการจัดหาโลหิตให้กับโรงพยาบาลต่าง ๆ ทั่วประเทศ โดยมีภาคบริการโลหิตแห่งชาติ 12 แห่งและ 1 สถานีกาชาดหัวหินเฉลิมพระเกียรติทำหน้าที่เป็นสาขาในการให้บริการกับโรงพยาบาล ในส่วนภูมิภาคจากข้อมูลปีงบประมาณ 2559 พบว่าทั้งประเทศมีการบริจาคโลหิตประมาณ 2,250,262 ยูนิต หรือประมาณร้อยละ 3.413 ของจำนวนประชากร เฉพาะผู้บริจาคโลหิตของศูนย์บริการโลหิตแห่งชาติพบว่ามีจำนวนไม่ผ่านคัดกรองจำนวน 115,474 ราย คิดเป็นร้อยละ 14.87 ประมาณร้อยละ 50 ของจำนวนที่ปฏิเสธทั้งหมดมีสาเหตุมาจากผลตรวจวัดค่าฮีโมโกลบินไม่ผ่านเกณฑ์(2)

2.1.6 ฮีโมโกลบิน

ฮีโมโกลบิน (Hemoglobin, Hb) คือโมเลกุลของโปรตีนที่อยู่บนเม็ดเลือดแดง ทำหน้าที่ในการขนส่งออกซิเจนจากปอดไปยังเนื้อเยื่อต่าง ๆ ทั่วร่างกายและขนส่งคาร์บอนไดออกไซด์จากเนื้อเยื่อกลับเข้าสู่ปอด

ฮีโมโกลบิน (Hemoglobin) ประกอบไปด้วยโกลบิน (Globin) 4 สายประกอบกัน โดยในผู้ใหญ่ปกติพบว่าฮีโมโกลบินประกอบด้วย แอลฟาโกลบิน (α globin) 2 สายและเบต้าโกลบิน (β

globin) 2 สาย ในขณะที่เด็กปกติแรกเกิดจะพบประกอบด้วยสายแอลฟาโกลบิน 2 สายและแกมมาโกลบิน (ϒ globin) 2 สาย ภายหลังที่เด็กเจริญเติบโตขึ้นสายแกมม่าโกลบินจะถูกแทนที่ด้วยเบต้าโกลบิน ทั้งนี้สายโกลบินแต่ละสายจะประกอบด้วยโมเลกุลที่สำคัญคือฮีม (Heme) ซึ่งทำหน้าที่ในการจับกับออกซิเจนและคาร์บอนไดออกไซด์(31, 32)



https://www.kindpng.com/imgv/bTixRh_haemoglobin-3d-ribbons-haemoglobin-structure-3d-transparent-hd/

รูปที่ 8 โครงสร้าง 3 มิติของฮีโมโกลบิน

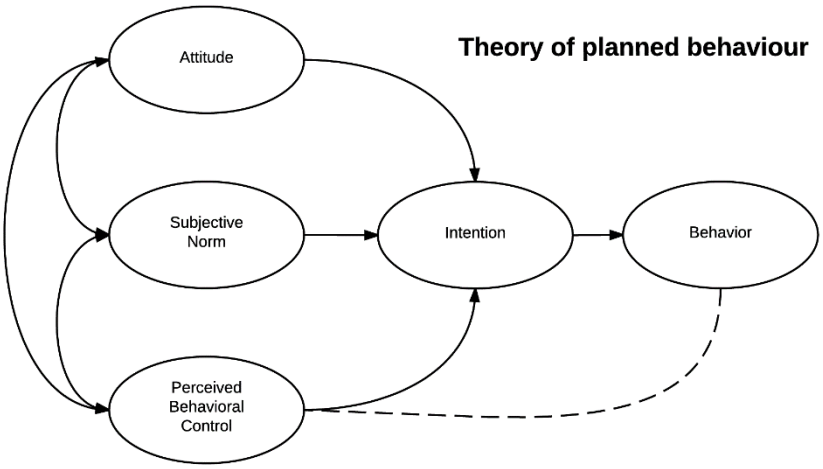
ภาวะโลหิตจางหรือที่เรียกกันอีกอย่างว่าภาวะซีด หมายถึงสภาวะที่ร่างกายมีปริมาณเม็ดโลหิตแดงน้อยกว่าปกติ(33)โดยสามารถแบ่งสาเหตุออกเป็น 3 กลุ่มลักษณะได้แก่

1. การสร้างเม็ดโลหิตแดงน้อยลง อาจเกิดจากโรคทางไขกระดูก โรคไตเรื้อรังหรือการขาดสารอาหารที่จำเป็นต่อการสร้างเม็ดโลหิตแดง
2. การที่เม็ดโลหิตแดงถูกทำลายมากกว่าปกติ โรคกลุ่มนี้เป็นสาเหตุให้เม็ดเลือดแดงของผู้ป่วยแตกง่ายโรคที่ประเทศไทยพบบ่อยได้แก่ โรคธาลัสซีเมีย โรคขาดเอนไซม์ G-6-PD และโรคภูมิคุ้มกันทำลายเม็ดโลหิตแดง
3. การเสียเลือด แบ่งออกเป็น 2 สาเหตุคือการเสียเลือดอย่างเฉียบพลัน เช่น การเสียเลือดจากการผ่าตัด อุบัติเหตุและการเสียเลือดอย่างเรื้อรังแบบค่อยเป็นค่อยไป เช่น ผู้ป่วยที่มีเลือดออกในทางเดินอาหาร เป็นต้น

2.1.7 ทฤษฎีพฤติกรรมตามแผน (Theory of Planned Behavior)

เป็นทฤษฎีที่พัฒนามาจากทฤษฎีการกระทำด้วยเหตุผล (A Theory of Reasoned Action) หรือ TRA ซึ่งถูกนำมาใช้ในการทำนายพฤติกรรมมนุษย์ โดยเชื่อว่ามนุษย์เป็นสิ่งมีชีวิตที่มีเหตุผล ข้อมูลด้านต่าง ๆ ที่ได้รับจะนำมาใช้ในการตัดสินใจอย่างเป็นระบบ ต่อมาในปี 1985 ไอส์เซ็น (5)ได้ปรับทฤษฎีให้สามารถทำนายพฤติกรรมที่มนุษย์ไม่สามารถควบคุมได้ จำเป็นต้องอาศัยปัจจัยอื่น ๆ เข้ามาร่วมในการตัดสินใจเช่น เวลา ความช่วยเหลือจากบุคคลอื่น เป็นต้น เรียกทฤษฎีนี้ว่า ทฤษฎีพฤติกรรมตามแผน ซึ่งต่างจากทฤษฎีการกระทำด้วยเหตุผลตรงที่ปัจจัยในการรับรู้ความสามารถในการควบคุมพฤติกรรม (Perceived behavioral control)

ทฤษฎีพฤติกรรมตามแผน (Theory of Planned Behavior) ของไอส์เซ็นเชื่อว่าความสำเร็จไม่ได้ขึ้นอยู่กับปัจจัยเจตนาหรือความต้องการของบุคคลอย่างเดียวแต่มีปัจจัยที่ไม่ใช่ สิ่งจูงใจด้วย เช่น ปัจจัยด้านทรัพยากร การเงิน เป็นต้น สรุปคือบุคคลที่มีเจตนาจะทำพฤติกรรมและมีปัจจัยอื่น ๆ สนับสนุนก็น่าจะทำให้บุคคลนั้นทำพฤติกรรมได้สำเร็จ ดังนั้นผลการตรวจฮีโมโกลบินในการบริจาคโลหิตอาจมีผลจากปัจจัยที่ไม่ใช่สิ่งจูงใจ เช่น รายได้ จำนวนบุตร ที่อยู่อาศัย เป็นต้น และผลจากปัจจัยเจตนาหรือความเชื่อ เช่น ศาสนา เป็นต้น



รูปที่ 9 แผนภูมิปัจจัยความเชื่อของมนุษย์ต่อการแสดงออกพฤติกรรมตามทฤษฎีของไอส์เซ็น(5)

2.1.8 เครื่องมือที่ใช้ในการทำงานวิจัย (Tools)

ในการศึกษาวิจัยนี้ใช้เครื่องในการเรียนรู้ของคอมพิวเตอร์ที่มีชื่อว่า RapidMiner (RM) ซึ่งเป็นโปรแกรมที่มีเทคนิคต่าง ๆ ให้เลือกใช้อย่างมากมายทั้งการคัดเลือกตัวแปร การจำแนกกลุ่ม

การแบ่งกลุ่มหรือการหาความสัมพันธ์ อาทิเช่น Decision tree, K-NN, Regression, Artificial neural networks เป็นต้น ปัจจุบันโปรแกรม RapidMiner เป็นที่นิยมอย่างแพร่หลายเนื่องจากการใช้งานที่ง่ายและยังมี Library ในการใช้งานร่วมกันโปรแกรมอื่นในการพัฒนาเช่น ภาษาไพธอน ภาษา R เป็นต้น

2.2 งานวิจัยที่เกี่ยวข้อง

2.2.1 การบริจาคโลหิตและปัจจัยที่เกี่ยวข้อง

การศึกษาของสุรชัย จันทวารีย์(34) ในผู้บริจาคโลหิตของโรงพยาบาลตำรวจ พบว่าผู้บริจาคโลหิตไม่ผ่านการคัดเลือกในปี 2551 และปี 2552 สูงถึงร้อยละ 26.2 และ 27.1 ตามลำดับ โดยพบเพศหญิงมากกว่าเพศชายในจำนวนดังกล่าว พบว่ามีสาเหตุจากความเข้มข้นโลหิตต่ำในเพศชายร้อยละ 21.9 และในเพศหญิงร้อยละ 38.4 และพบในผู้บริจาคโลหิตรายเก่ามากกว่าผู้บริจาคโลหิตครั้งแรก การศึกษาของวิชุดา กลิ่นหอม(35) พบว่ามีผู้ถูกปฏิเสธให้บริจาคโลหิตประมาณร้อยละ 15-20 โดยร้อยละ 50 ของผู้ที่ถูกปฏิเสธให้บริจาคโลหิตมีสาเหตุจากค่าฮีโมโกลบิน (hemoglobin; Hb) ต่ำกว่าเกณฑ์ ซึ่งมักพบในผู้บริจาคโลหิตประจำโดยเฉพาะผู้บริจาคโลหิตเพศหญิง จากปัญหาดังกล่าวนอกจากจะทำให้การจัดหาโลหิตไม่เป็นไปตามเป้าหมายแล้วยัง ส่งผลให้จำนวนผู้บริจาคโลหิตรายเก่ามีจำนวนน้อยลงในทุกปีส่งผลกระทบต่อผู้ป่วยโดยตรง

2.2.2 แนวคิดการประยุกต์ใช้การเรียนรู้ของคอมพิวเตอร์

การศึกษาของ Tanner L. (36) เป็นการศึกษาใช้เทคนิคต้นไม้ตัดสินใจในการวินิจฉัยและพยากรณ์ความรุนแรงของโรคไข้เลือดออกภายใน 72 ชั่วโมงจากที่เริ่มป่วย โดยใช้ข้อมูลทางคลินิกและผลตรวจทางห้องปฏิบัติการเป็นข้อมูลในการพัฒนาตัวแบบพยากรณ์ พบว่าหากให้ผลตรวจ Complete blood count (CBC) เป็นตัวแปรในการสร้างตัวแบบได้ค่าความไว ค่าความจำเพาะ ค่าพยากรณ์ผลบวกและค่าพยากรณ์ผลลบเท่ากับร้อยละ 71.2, 90.1, 57.7 และ 94.4 ตามลำดับโดยค่าความคาดเคลื่อนเท่ากับร้อยละ 15.7 แต่ใช้ผลตรวจ CBC ร่วมกับการตรวจ RT-PCR จะได้ค่า ความไว ค่าความจำเพาะ ค่าพยากรณ์ผลบวกและค่าพยากรณ์ผลลบเท่ากับร้อยละ 78.2, 80.2, 51.1 และ 97.7 ตามลำดับโดยค่าความคาดเคลื่อนเท่ากับร้อยละ 20.5

การศึกษาของ S.Asha Rani(37) เป็นการศึกษาการจำแนกกลุ่มผู้บริจาคโลหิตโดยอาศัยเทคนิคการจำแนกแบบเบย์อย่างง่าย ต้นไม้ตัดสินใจประเภท J48 และ Random tree โดยใช้ข้อมูลระยะห่างจากการบริจาคครั้งล่าสุด จำนวนครั้งของการบริจาคโลหิต ปริมาตรโลหิตที่บริจาค

รวมและจำนวนเดือนตั้งแต่การบริจาคโลหิตครั้งแรก พบว่าค่าความถูกต้องของเทคนิคการจำแนกแบบเบย์อย่างง่าย J48 และ Random tree เท่ากับร้อยละ 75, 80.88 และ 93.18 ตามลำดับ

การศึกษาของ Nasserinejad K.(38) เป็นการศึกษาพยากรณ์ค่าฮีโมโกลบินของผู้บริจาคโลหิตโดยอาศัยเทคนิค Multiple linear regression แบ่งเป็น transition (autoregressive) model และ mixed effects model โดยอาศัยข้อมูล เพศ อายุและฤดู เป็นตัวแปรต้นในการพยากรณ์ค่าฮีโมโกลบิน พบว่าในเพศชายค่า AUC ของเทคนิค transition (autoregressive) model และmixed effects model เท่ากับ 0.83 และ 0.81 ตามลำดับ และในเพศหญิงค่า AUC ของเทคนิค transition (autoregressive) model และ mixed effects model เท่ากับ 0.73 และ 0.72 ตามลำดับ

การศึกษาของ C.H. Yu, M. Bhatnagar(39) ได้ศึกษาพยากรณ์ระดับภาวะโลหิตจางหลังการผ่าตัดโดยอาศัยเทคนิค Multilayer perceptron neural network (MLP) โดยอาศัย 17 ตัวแปรต้น เช่น กระบวนการผ่าตัด วันเวลาเริ่มผ่าตัด วันเวลาผ่าตัดเสร็จ ประเภทผู้ป่วยและข้อมูลการให้โลหิต ส่วนประกอบโลหิต เป็นต้น พบว่าเทคนิค MLP ในการจำแนกกลุ่มเป็น 3 กลุ่มคือ ไม่มีภาวะโลหิตจาง มีภาวะโลหิตจางและกลุ่มมีภาวะโลหิตจางรุนแรง ได้ค่าความถูกต้องร้อยละ 67 และค่าเบี่ยงเบนมาตรฐานเท่ากับร้อยละ 3 และในการจำแนกเป็น 2 กลุ่มคือ ไม่มีภาวะโลหิตจางและมีภาวะโลหิตจาง ได้ค่าความถูกต้องร้อยละ 77 และค่าเบี่ยงเบนมาตรฐานเท่ากับร้อยละ 2

การศึกษาของ Pannaporn K.(40) เป็นการศึกษาการทำเหมืองข้อความเพื่อช่วยจำแนกประเภทโรคจากอาการ อาศัยเทคนิคต้นไม้ตัดสินใจ การเรียนรู้แบบเบส์อย่างง่าย ซัพพอร์ตเวกเตอร์แมชชีนและโครงข่ายประสาทเทียม โดยอาศัยข้อมูลเวชระเบียนประกอบด้วย ข้อมูลในส่วนบันทึกของแพทย์และการวินิจฉัยโรค พบว่าต้นไม้ตัดสินใจ การเรียนรู้แบบเบส์อย่างง่าย ซัพพอร์ตเวกเตอร์แมชชีนและโครงข่ายประสาทเทียม ให้ค่าความถูกต้องเท่ากับร้อยละ 97.13, 92.86, 86.45, 89.50 และค่า AUC เท่ากับ 0.75, 0.84, 0.91, 0.93 ตามลำดับ

การศึกษาของ Priyanka J.(41) ศึกษาพัฒนาตัวแบบพยากรณ์การวินิจฉัยและการพยากรณ์โรคของมะเร็งเต้านม โดยใช้โปรแกรม RapidMiner ในการพัฒนาตัวแบบ 3 ชนิดคือ ซัพพอร์ตเวกเตอร์แมชชีน K-Nearest Neighbor (KNN) และการจำแนกแบบเบส์อย่างง่าย โดยใช้ตัวแปรต้น 16 ตัวแปร พบว่าตัวแบบพยากรณ์ซัพพอร์ตเวกเตอร์แมชชีน K-Nearest Neighbour (KNN) และการเรียนรู้แบบเบส์อย่างง่ายให้ค่าความถูกต้องร้อยละ 80, 75 และ 22 ตามลำดับ

การศึกษาของ Jothikumar R.(42) ได้ศึกษาการสร้างตัวแบบพยากรณ์โรคหัวใจซึ่งพัฒนาด้วยโปรแกรม RapidMiner ประกอบด้วย 4 ชนิดคือ Random Tree, Naïve Bayes, Decision Tree และ Random forest โดยใช้ข้อมูล 14 ตัวแปรต้น เช่น เพศ อายุ ระดับไขมัน เป็นต้น พบว่าตัวแบบพยากรณ์ Random Tree การจำแนกแบบเบส์อย่างง่าย ต้นไม้ตัดสินใจและ Random forest ให้ค่าความถูกต้องร้อยบะ 75.14, 79.25, 78.24 และ 74.16 ตามลำดับ

การศึกษาของ Basharat Naqvi(43) เป็นการศึกษาการพยากรณ์การวินิจฉัยโรคเบาหวาน โดยอาศัยโปรแกรม RapidMiner ในการสร้างตัวแบบพยากรณ์ต้นไม้ตัดสินใจ 4 ชนิด ประกอบด้วย Random Forest, Random Tree, ID3 และ Decision Stump โดยใช้ข้อมูล 50 ตัวแปรของผู้ป่วยจาก 130 โรงพยาบาล พบว่าตัวแบบพยากรณ์ Random Forest, Random Tree, ID3 และ Decision Stump ให้ค่าความถูกต้องและความแม่นยำเท่ากับร้อยละ 78.63, 76.58, 84.94, 74.76 และ 78.63, 76.58, 84.64, 74.76 ตามลำดับ

2.2.3 สรุปผลของงานวิจัยที่เกี่ยวข้อง

จากข้อมูลผลการวิจัยในข้างต้นสามารถนำมาสรุปเป็นตารางเปรียบเทียบความแตกต่างได้ตามตารางที่ 1

ตารางที่ 1 สรุปเปรียบเทียบงานวิจัยที่เกี่ยวข้อง

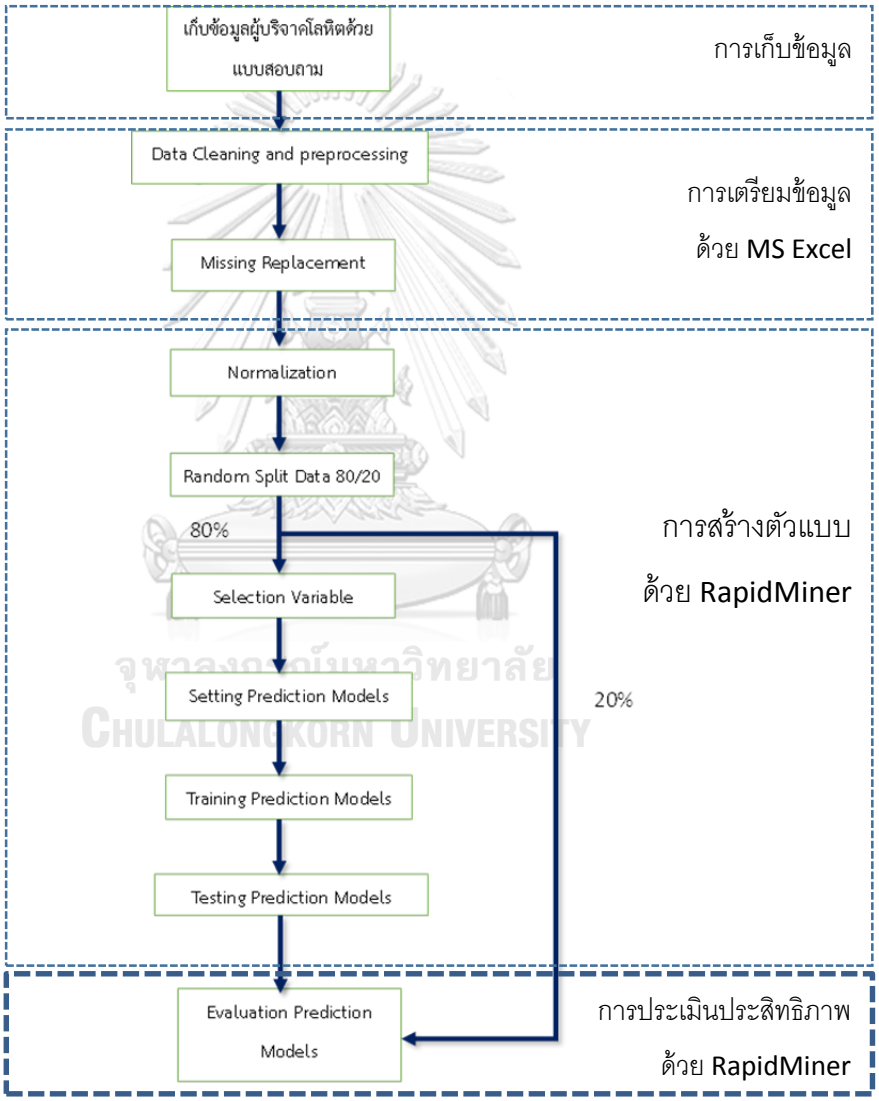
ลำดับ	งานวิจัย	ชุดข้อมูล	ตัวจำแนก	ประเภทตัวชี้วัด	ผลลัพธ์
1	การวินิจฉัยและพยากรณ์ความรุนแรงของโรคไข้เลือดออก	ข้อมูลทางคลินิก ผลตรวจ CBC และ RT-PCR	เทคนิคต้นไม้ตัดสินใจ	ค่าความไว ค่าความจำเพาะ ค่าพยากรณ์ผลบวกและค่าพยากรณ์ผลลบ	ค่าความไว 78.2 ค่าความจำเพาะ 80.2 ค่าพยากรณ์ผลบวก 51.1 ค่าพยากรณ์ผลลบ 97.7
2	การจำแนกกลุ่มผู้บริจาคโลหิต	ทะเบียนข้อมูลของผู้บริจาคโลหิต	จำแนกแบบเบย์อย่างง่าย J48 และ Random tree	ค่าความถูกต้อง Time to taken	ต้นไม้ตัดสินใจชนิด Random tree ให้ค่าความถูกต้องและ Time to taken สูงที่สุดคือ 93.18, 0.01 ตามลำดับ
3	การพยากรณ์ค่าฮีโมโกลบินของผู้บริจาคโลหิต	ข้อมูลผู้บริจาคโลหิต เพศ อายุ และฤดูที่มาบริจาคโลหิต	Transition model และ mixed effects model	AUC	ตัวแบบ Transition model ให้ค่า AUC สูงสุดคือในเพศชาย 0.83 และในเพศหญิง 0.73
4	พยากรณ์ระดับภาวะโลหิตจางหลังการผ่าตัด	ข้อมูลเวชระเบียนผู้ป่วย	Multilayer perceptron neural	ค่าความถูกต้อง ค่าเบี่ยงเบน	ได้ค่าความถูกต้องร้อยละ 77 และค่าเบี่ยงเบน

			network	มาตรฐาน	มาตรฐานเท่ากับร้อยละ 2
5	การทำเหมืองข้อความเพื่อช่วยจำแนกประเภทโรคจากอาการ	ข้อมูลเวชระเบียนผู้ป่วย	ต้นไม้ตัดสินใจ การเรียนรู้แบบเบส์อย่างง่าย ซัพพอร์เวกเตอร์แมชชีนและโครงข่ายประสาทเทียม	TPR FPR ค่าความถูกต้อง ค่าความแม่นยำ AUC	โครงข่ายประสาทเทียมเป็นตัวแบบที่เหมาะสมที่สุดคือให้ค่า TPR และค่า AUC สูงที่สุดเท่ากับ 89.03 และ 0.93
6	พัฒนาตัวแบบพยากรณ์การวินิจฉัยและการพยากรณ์โรคของมะเร็งเต้านม	ข้อมูลเวชระเบียนผู้ป่วย	ซัพพอร์ตเวกเตอร์แมชชีน K-Nearest Neighbor (KNN) และการจำแนกแบบเบส์อย่างง่าย	ค่าความถูกต้อง	ตัวแบบพยากรณ์ซัพพอร์ตเวกเตอร์แมชชีนเป็นตัวแบบที่ดีที่สุดโดยให้ค่าความถูกต้องเท่ากับ 80
7	การสร้างตัวแบบพยากรณ์โรคหัวใจ	ข้อมูลเวชระเบียนผู้ป่วย ข้อมูลจากเว็บไซต์สาธารณะ	Random Tree การจำแนกแบบเบส์อย่างง่าย ต้นไม้ตัดสินใจ และ Random forest	ค่าความถูกต้อง	การจำแนกแบบเบส์อย่างง่ายเป็นตัวแบบที่ดีที่สุด โดยให้ค่าความถูกต้อง 79.25
8	การพยากรณ์การวินิจฉัยโรคเบาหวาน	ข้อมูลเวชระเบียนผู้ป่วย	Random Forest, Random Tree, ID3 และ Decision Stump	ค่าความถูกต้อง ค่าความแม่นยำ	พบว่า ID3 เป็นต้นแบบที่ให้ค่าความถูกต้องและความแม่นยำสูงสุดเท่ากับ 84.94 และ 84.64

จากงานวิจัยที่เกี่ยวข้องจะเห็นได้ว่ามีการนำเทคนิคการทำเหมืองข้อมูลเรียนรู้ของคอมพิวเตอร์มาปรับใช้ในงานทางการแพทย์อย่างหลากหลาย เพื่อช่วยเหลือในการตัดสินใจของแพทย์ ส่วนการพยากรณ์เกี่ยวกับค่าฮีโมโกลบินในผู้บริจาคได้มีการศึกษาแล้วเช่นกันโดยใช้เทคนิคการหาความสัมพันธ์ (Linear Regression) จากตัวแปร เพศ อายุและฤดูกาลที่บริจาคโลหิต แต่ยังไม่เคยมีการวิจัยการพยากรณ์จำแนกกลุ่มค่าฮีโมโกลบิน งานวิจัยนี้จึงมีแนวคิดการนำเทคนิคการเรียนรู้ของคอมพิวเตอร์ที่เป็นที่นิยมในการสร้างตัวแบบพยากรณ์คือ ต้นไม้ตัดสินใจ ซัพพอร์ตเวกเตอร์แมชชีน การจำแนกเบส์อย่างง่ายและโครงข่ายประสาทเทียม มาใช้ในการพยากรณ์ค่าฮีโมโกลบินเพื่อการจำแนกกลุ่มผู้บริจาคโลหิตและเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์ จากนั้นนำประสิทธิภาพของตัวแบบมาเปรียบเทียบกัน เพื่อหาตัวแบบที่เหมาะสมที่สุดของงานวิจัยนี้

บทที่ 3
วิธีการดำเนินงานวิจัย

งานวิจัยนี้มีเป้าหมายในการสร้างตัวแบบพยากรณ์การจำแนกกลุ่มผู้บริจาคโลหิตจากค่าฮีโมโกลบิน โดยใช้ข้อมูลตัวแปรต้นจากแบบสอบถามและข้อมูลผลตรวจค่าฮีโมโกลบินที่ได้จากการเจาะปลายนิ้ว จากนั้นนำตัวแบบต่าง ๆ มาเปรียบเทียบประสิทธิภาพ สามารถแบ่งออกเป็น 4 ขั้นตอนหลักคือ การเก็บข้อมูล การเตรียมข้อมูล การสร้างตัวแบบและการประเมินประสิทธิภาพ



รูปที่ 10 ขั้นตอนการดำเนินการวิจัย

3.1 การเก็บข้อมูล

งานวิจัยนี้ได้รับการรับรองจากคณะกรรมการพิจารณาจริยธรรมการวิจัยในมนุษย์ ศูนย์บริการโลหิตแห่งชาติ สภากาชาดไทย Certificate Number NBC 17/2018 เลขที่โครงการ 17/2561 ทำการเก็บข้อมูลจากผู้ประสงค์บริจาคโลหิตของภาคบริการโลหิตแห่งชาติ 12 แห่งและงานบริการโลหิต สถานีกาชาดหัวหินเฉลิมพระเกียรติ เก็บข้อมูลผู้ประสงค์บริจาคโลหิตทุกรายที่ยินยอมให้ข้อมูล ด้วยแบบสอบถามร่วมกับข้อมูลการบริจาคโลหิตจากระบบฐานข้อมูล Hematos IIG (HIIG) ของศูนย์บริการโลหิตแห่งชาติ สภากาชาดไทย ตั้งแต่ 1 ตุลาคม 2561 ถึง 31 พฤษภาคม 2562

3.1.1 เครื่องมือที่ใช้ในการเก็บข้อมูล

เครื่องมือที่ใช้ในการเก็บข้อมูลในการศึกษานี้คือ แบบสอบถามและข้อมูลในฐานข้อมูลระบบ HIIG ของศูนย์บริการโลหิตแห่งชาติ สภากาชาดไทย ซึ่ง HIIG คือโปรแกรมปฏิบัติการด้านงานบริการโลหิตเก็บข้อมูลตั้งแต่การรับบริจาคโลหิต การเตรียมส่วนประกอบโลหิต การตรวจคัดกรองคุณภาพโลหิตและการจ่ายโลหิตให้ผู้ป่วย จากนั้นบันทึกข้อมูลด้วยโปรแกรม Microsoft Excel 2013 โดยแบบสอบถามประกอบด้วย 3 ส่วน คือ

ส่วนที่ 1 แบบสอบถามข้อมูลทั่วไปของผู้บริจาคโลหิต จำนวน 10 ข้อ

เป็นข้อมูลทั่วไปของผู้เข้าร่วมวิจัยได้แก่ เพศ อายุ น้ำหนัก ส่วนสูง อาชีพ รายได้ เป็นต้น เพื่อใช้เป็นข้อมูลพื้นฐานในการแบ่งกลุ่มเปรียบเทียบ

ส่วนที่ 2 แบบสอบถามเกี่ยวกับการบริจาคโลหิต จำนวน 12 ข้อ

เป็นแบบสอบถามข้อมูลประวัติการบริจาคโลหิตของผู้เข้าร่วมวิจัยได้แก่ หมู่โลหิต จำนวนครั้งที่บริจาค ความถี่ การได้รับข่าวสารการบริจาคโลหิต เป็นต้น

ส่วนที่ 3 แบบสอบถามเกี่ยวกับพฤติกรรมสุขภาพ จำนวน 7 ข้อ

เป็นข้อมูลพฤติกรรมด้านสุขภาพของผู้บริจาคเช่น ความถี่ในการออกกำลังกาย การสูบบุรี่ การดื่มแอลกอฮอล์ การกินยาธาตุเหล็ก เป็นต้น เพื่อศึกษาความสัมพันธ์ปัจจัยการดูแลสุขภาพกับการบริจาคโลหิต

3.1.2 การเก็บข้อมูลตัวแปร

จากแบบสอบถามและข้อมูลในระบบสารสนเทศของศูนย์บริการโลหิตแห่งชาติ สภากาชาดไทย เมื่อนำมาบันทึกข้อมูลใน MS Excel สามารถแจกแจงตัวแปรต้นหรือตัวแปรพยากรณ์ได้จำนวน 44 ตัวแปร แบ่งเป็นตัวแปรต้น 43 ตัวแปร ตัวแปรตาม 1 ตัวแปร ตามตารางที่ 2

ตารางที่ 2 แสดงตัวแปรพยากรณ์ที่ใช้ในการศึกษา

ลำดับ	ตัวแปร	ลำดับ	ตัวแปร
1	เพศ (Gender)	23	หมู่โลหิต Rh (Rh Group)
2	อายุ (Age)	24	สถานที่บริจาคโลหิต (Donation Place)
3	น้ำหนัก (Weight)	25	ประวัติการบริจาคเกล็ดเลือดหรือน้ำเลือด (Single Donation)
4	น้ำหนักเมื่อครั้งที่ผ่านมา (Weight in past 3 months)	26	จำนวนครั้งที่เคยบริจาคเกล็ดเลือดหรือน้ำเลือด (No. Single Donation)
5	น้ำหนักที่เปลี่ยนแปลใน 3 เดือน (Chang Weight in past 3 months)	27	นำหนักที่เปลี่ยนแปลงใน 1 ปี (Chang in Weight past a year)
6	น้ำหนักปีที่ผ่านมา (Weight in past a year)	28	จำนวนครั้งที่บริจาคโลหิตแบบ WB (Whole Blood Donation Times)
7	ส่วนสูง (High)	29	ความถี่การบริจาค WB ในรอบปี (Donation/year)
8	BMI	30	ผลการบริจาค (เต็มถุงหรือไม่เต็มถุง) (Success of Donation)
9	ความดันโลหิต Systolic	31	ขนาดของถุงที่บริจาค (Bag type)
10	ความดันโลหิต Diastolic	32	ประวัติการตรวจฮีโมโกลบินไม่ผ่านเกณฑ์ (EverLowHb)
11	อัตราการเต้นของหัวใจ (Pulse)	33	ประวัติโรคประจำตัว (Disease)
12	ค่าฮีโมโกลบินครั้งที่ผ่านมา (Previous Hb)	34	การรับประทานอาหาร (Food type)
13	ค่าฮีโมโกลบินครั้งนี้ (Current Hb)	35	การสูบบุหรี่ (Smoke)
14	ระยะห่างการบริจาค (Interval donation)	36	การออกกำลังกาย (Exercise)

15	ระดับการศึกษา (Education)	37	การดื่มแอลกอฮอล์ (Alcohol)
16	อาชีพ (Occupation)	38	พฤติกรรมการพักผ่อน (Sleep Type)
17	รายได้ (Income)	39	ชั่วโมงการนอน (Sleep hour)
18	สถานะภาพ (Status)	40	จำนวนบุตร (No. of Child)
19	ศาสนา (Religion)	41	การมีประจำเดือน (Menopause)
20	ที่อยู่ (Address1)	42	การรับประทานธาตุเหล็ก (Fe take)
21	ที่อยู่เขตเมืองหรือชนบท (Address2)	43	สาเหตุการไม่ทานธาตุเหล็ก (Why not Fe take)
22	หมู่โลหิต ABO	44	ช่องทางการได้รับข่าวสารการบริจาคโลหิต (Sources of Donation Information)

3.2 การเตรียมข้อมูล

ข้อมูลดิบ (Raw data) ที่ได้จากแบบสอบถามและข้อมูลจากระบบสารสนเทศ ศูนย์บริการโลหิตแห่งชาติ สภากาชาดไทย เมื่อนำมาบันทึกข้อมูลในโปรแกรม MS Excel อาจพบข้อมูลที่ไม่สมบูรณ์สูญหายหรือข้อมูลมีค่าผิดปกติ (Outlier) ดังนั้นต้องมีการทำความเข้าใจข้อมูล (Data understanding) การจัดเตรียมข้อมูลและตัวแปรบางตัวอาจต้องมีการแปลงข้อมูลเพื่อให้เหมาะต่อการนำไปพัฒนาตัวแบบพยากรณ์ ซึ่งถือเป็นขั้นตอนที่สำคัญอย่างยิ่งในการทำเหมืองข้อมูล

3.2.1 การกลั่นกรองข้อมูล (Data Cleaning)

เป็นการตรวจสอบความคลาดเคลื่อนของข้อมูล เป็นการตรวจสอบข้อมูล แก้ไขข้อมูลที่ผิดพลาด หรือไม่ได้ใช้งาน มีค่าไม่ตรงกับที่ต้องการใช้งานซึ่งอาจเกิดจากความผิดพลาดในการตอบแบบสอบถามของผู้บริจาคโลหิตหรือการบันทึกข้อมูลที่ผิดพลาด

3.2.2 การจัดเตรียมและการแปลงข้อมูล (Data preparation)

ทำการแปลรหัสข้อมูลที่ไม่ใช่ค่าต่อเนื่องให้อยู่ในรูปแบบที่เหมาะสมต่อการนำเข้าตัวแบบพยากรณ์หรือการสร้างตัวแปรใหม่จากข้อมูลตัวแปรเดิมเช่น ตัวแปร BMI สามารถคำนวณได้จากตัวแปรส่วนสูงและน้ำหนัก เป็นต้น

ตารางที่ 3 แสดงการเข้ารหัสข้อมูลให้อยู่ในรูป nominal scale

ตัวแปร	**การเข้ารหัส**	**ตัวแปร**	**การเข้ารหัส**
เพศ	ชาย=0, หญิง=1	หมู่โลหิต	0=A, 1=B, 2=O, 3=AB
สถานะภาพ	0=โสด, 1=สมรส, 2=อย่าร้าง/หม้าย	ประจำเดือน	0=มี 1=ไม่มี
การศึกษา	0=ประถมศึกษาหรือต่ำกว่า 1=มัธยมศึกษาตอนต้น 2=มัธยมศึกษาตอนปลาย/ปวช. 3=อนุปริญญา/ปวส. 4=ปริญญาตรี 5=สูงกว่าปริญญาตรี	การรับประทานอาหาร	0=อาหารปกติ 1=อาหารเจ 2=อาหารมังสะวิรัติ 3=อาหารประเภทไขมัน 4=อาหารdietลดน้ำหนัก
ที่อยู่	Address1 0=อ.เมือง 1=ต่างอำเภอ Address2 0=อยู่ในเขตเทศบาล 1=ไม่อยู่ในเขตเทศบาล	ประวัติโรคประจำตัว	0=ไม่มี 1=เบาหวาน 2=ไขมันในเลือดสูง 3=ความดันโลหิตสูง 4=ภาวะ/โรคโลหิตจาง 5=อื่น ๆ
การศึกษา	0=ประถมหรือต่ำกว่า 1=ม.ต้น 2=ม.ปลาย/ปวช 3=อนุปริญญา/ปวส. 4=ปริญญาตรี 5=สูงกว่าปริญญาตรี	อาชีพ	0=นักเรียน/นักศึกษา 1=เกษตรกร 2=ราชการ/รัฐวิสาหกิจ 3=พนักงานเอกชน 4=ค้าขาย/รับจ้างอิสระ 5=พระภิกษุ/สามเณร/นักบวช 6=อื่น ๆ
สถานที่บริจาค	0=ภาคบริการโลหิตแห่งชาติ 1=โรงพยาบาล 2=หน่วยเคลื่อนที่	ชนิดถุงที่บริจาค	0=เต็มถุงปกติ 450 มล. 1=เต็มถุงปกติ 350 มล. 2=Under Volume
บริจาคโลหิตเฉพาะส่วน	0=เคย 1=ไม่เคย	ประวัติผลตรวจ Hbต่ำ	0=เคย 1=ไม่เคย
การพักผ่อน	0=เป็นเวลา, 1=ไม่เป็นเวลา	การสูบบุหรี่	0=ไม่สูบ, 1=สูบ

ตัวแปร	การเข้ารหัส	ตัวแปร	การเข้ารหัส
ชั่วโมงการพักผ่อน	0=น้อยกว่า 4 ชม. 1=4-6 ชม. 2=มากกว่า 6 ชม.	ทานธาตุเหล็ก	0=ไม่ทาน 1=ทานเป็นประจำ 2=ทานไม่สม่ำเสมอ
การดื่มแอลกอฮอล์	0=ไม่ดื่ม 1=ดื่มน้อยกว่า 2 ครั้ง/สัปดาห์ 2=ดื่ม 3-5 ครั้ง/สัปดาห์ 3=ทุกวัน	รายได้ (ต่อเดือน)	0= น้อยกว่า 10,000 1=10,000-20,000 2=20,001-40,000 3=>40,000
สาเหตุที่ไม่ทานยา	0=ไม่ทาน ก็ยังบริจาคได้ 1=ลืม/ขี้เกียจ 2=ไม่ชอบกลิ่นและสี 3=กลัวอ้วน 4=มีอาการคลื่นไส้/ท้องผูก/ท้องเสีย 5=อื่น ๆ	ช่องทางการรับข่าวสาร	0=วิทยุ 1=โทรทัศน์ 2=เฟซบุค 3=ไลน์ 4= SMS 5= สื่อสิ่งพิมพ์ 6=คนรู้จัก
การออกกำลังกาย	0=ไม่เคยออก 1=น้อยกว่า 2 ครั้ง/สัปดาห์ 2=3-5ครั้ง/สัปดาห์ 3=ทุกวัน		

3.2.3 การจัดการกับข้อมูลสูญหาย (Handling missing data)

งานวิจัยนี้ใช้วิธีการแทนที่ข้อมูลที่สูญหายด้วยค่าเฉลี่ยของข้อมูลสำหรับตัวแปรเชิงปริมาณ และใช้การทดแทนด้วยเลขฐานนิยมสำหรับตัวแปรเชิงกลุ่มหรือเชิงคุณภาพ แต่เฉพาะตัวแปรค่าฮีโมโกลบินครั้งที่ผ่านมาจะทำการชดเชยค่าว่างด้วยค่าเฉลี่ยแยกระหว่างเพศชายและหญิง

3.3 การพัฒนาตัวแบบพยากรณ์ด้วยโปรแกรม RapidMiner

งานวิจัยนี้ใช้โปรแกรม RapidMiner studio ซึ่งเป็น open source software เกี่ยวกับการคำนวณสถิติ การทำเหมืองข้อมูล โดยมีฟังก์ชั่นต่าง ๆ ให้เลือกใช้ทั้งการจำแนก การแบ่งกลุ่ม การหาความสัมพันธ์และการพยากรณ์ข้อมูลนอกจากนี้ยังสามารถใช้งานร่วมกับภาษา R ได้อีกด้วย ในการศึกษานี้ใช้โปรแกรม RapidMiner เวอร์ชั่น 9.2000 ในการสร้างตัวแบบพยากรณ์ประกอบด้วย 4 ต้นแบบคือ

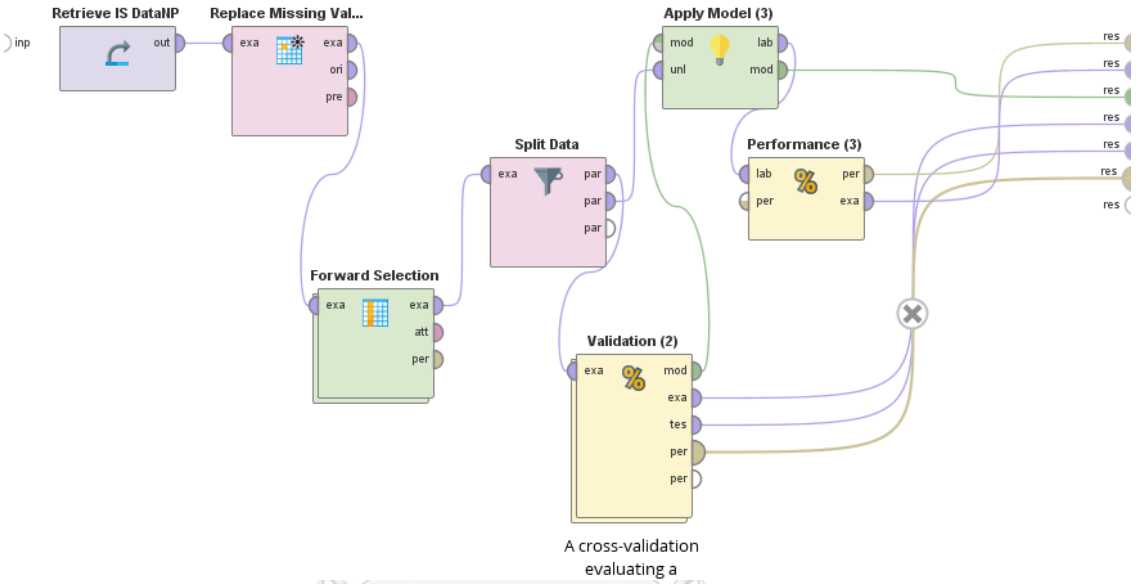
1. ต้นไม้ตัดสินใจ (decision tree)

2. ซัพพอร์ตเวกเตอร์แมชชีน (support vector machine; SVM)

3. การจำแนกแบบเบย์อย่างง่าย (naïve bayesain classifier)

4. โครงข่ายประสาทเทียม (artificial neural networks; ANN)

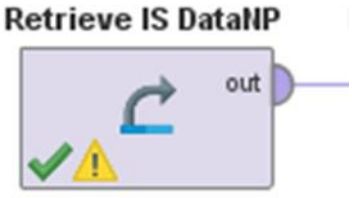
วิธีการพัฒนาตัวแบบพยากรณ์ด้วยโปรแกรม RapidMiner มีขั้นตอน 7 ขั้นตอนดังแสดงในรูปที่ 10



รูปที่ 11 รูปภาพกระบวนการสร้างตัวแบบพยากรณ์ด้วยโปรแกรม RapidMiner

3.3.1 การนำเข้าข้อมูล

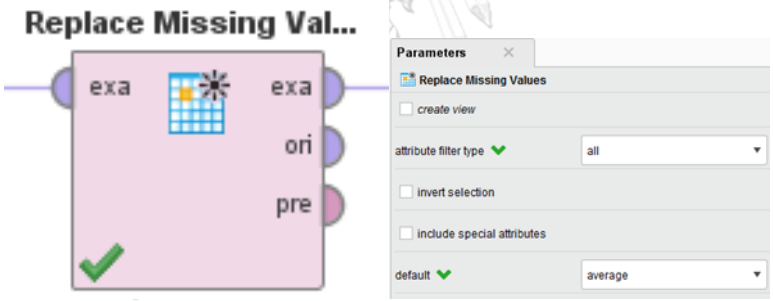
เป็นการนำเข้าข้อมูลจากโปรแกรม Microsoft Excel เข้าสู่โปรแกรม RapidMiner ด้วยฟังก์ชั่น Repository > Add data จากนั้นเลือกไฟล์ข้อมูลที่ต้องการนำมาวางที่ Process view โดยกำหนดให้คอลัมน์ของตัวแปรพยากรณ์มีคุณสมบัติเป็น attribute หรือตัวแปรต้นและกำหนดให้คอลัมน์ของผลลัพธ์มีคุณสมบัติเป็น label หรือตัวแปรตาม โดยกำหนดประเภทข้อมูลตามชนิดของข้อมูลเช่น nominal คือประเภทข้อมูลที่แบ่งได้เป็น 2 กลุ่ม polynominal คือประเภทข้อมูลที่แบ่งได้มากกว่า 2 กลุ่ม และreal คือข้อมูลประเภทค่าต่อเนื่อง เป็นต้น



รูปที่ 12 รูปภาพ Process Data Input

3.3.2 การชดเชยข้อมูลสูญหาย (Handling missing data)

เป็นขั้นตอนกำหนดให้โปรแกรมตรวจสอบข้อมูลที่อาจขาดหายไปถึงแม้จะทำการ data cleaning และ preprocessing ในเบื้องต้นแล้วก็ตามแต่โปรแกรม RapidMiner มีฟังก์ชันช่วยในการทำงานไม่ให้เกิดความผิดพลาดในการเตรียมข้อมูลก่อนการสร้างตัวแบบพยากรณ์ โดยเลือกฟังก์ชั่น Operators > Cleansing > Missing > Replace Missing Values นำมาวางที่ Process view และทำการเชื่อมต่อกับ Data input ในการศึกษานี้กำหนดให้ทำการเติมข้อมูลที่ขาดหายไปด้วยค่าเฉลี่ยของแต่ละชนิดข้อมูล



รูปที่ 13 รูปภาพ Process การเติมค่าที่ขาดหาย

3.3.3 การคัดเลือกตัวแปร

3.3.3.1 การคัดเลือกตัวแปรโดยวิธีนำตัวแปรเข้าทั้งหมด (enter regression)

เป็นการนำตัวแปรทั้งหมดเข้ามาใช้ฝึกสอนและทดสอบตัวแบบพยากรณ์ ซึ่งในการศึกษานี้คือการเชื่อมต่อ Data input เข้าสู่ตัวแบบพยากรณ์โดยไม่ต้องมีการคัดเลือกตัวแปร

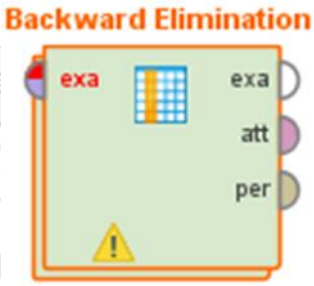
3.3.3.2 การคัดเลือกตัวแปรโดยวิธีเพิ่มตัวแปร (forward selection)

ทำการคัดเลือกตัวแปรด้วยฟังก์ชั่น Operators > Modeling > Predictive > Optimization > Feature Selection > Forward Selection

รูปที่ 14 ฟังก์ชั่นการคัดเลือกตัวแปรด้วย Forward Selection

3.3.3.3 การเลือกตัวแปรโดยวิธีลดตัวแปร (backward elimination)

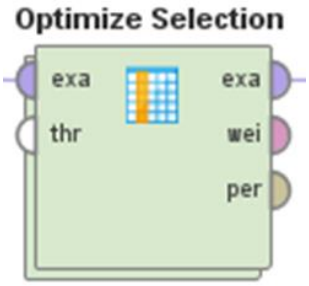
ทำการคัดเลือกตัวแปรด้วยฟังก์ชั่น Operators > Modeling > Predictive > Optimization > Feature Selection > Backward Elimination



รูปที่ 15 ฟังก์ชั่นการคัดเลือกตัวแปรด้วย backward elimination

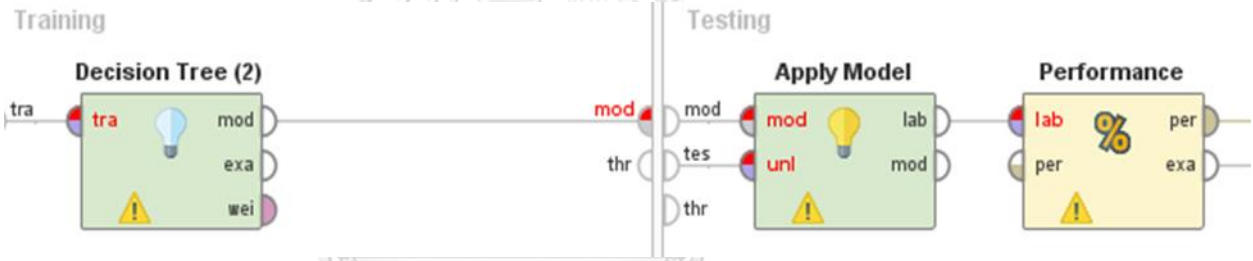
3.3.3.4 การเลือกตัวแปรโดยวิธีเพิ่มตัวแปรและลดตัวแปร (Optimize selection)

เป็นฟังก์ชั่นที่มีหลักการคล้ายกับวิธีเพิ่มตัวแปรอิสระแบบขั้นตอน (Stepwise Regression) โดยทำการคัดเลือกตัวแปรด้วยฟังก์ชั่น Operators > Modeling > Predictive > Optimization > Feature Selection > Optimize selection กำหนดอัตรานำเข้าตัวแปรที่ร้อยละ 78 และอัตราการนำออกตัวแปรที่ร้อยละ 22 ซึ่งเป็นค่าเริ่มต้นที่โปรแกรม RapidMiner กำหนด



รูปที่ 16 ฟังก์ชั่นการคัดเลือกตัวแปรด้วย optimize selection

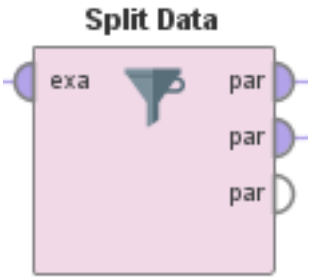
หลังจากเลือก Operation ในการคัดเลือกตัวแปรแต่ละวิธีได้แล้วต้องทำการสร้าง Building block และเลือก Numerical Cross Validation หรือ Nominal Cross Validation ในช่องของ Training box ให้เลือก Operation ตามชนิดของตัวแบบพยากรณ์ที่ต้องการทั้ง 4 ชนิดคือ Decision tree, SVM, Naïve bayesain และ Artificial neural networks



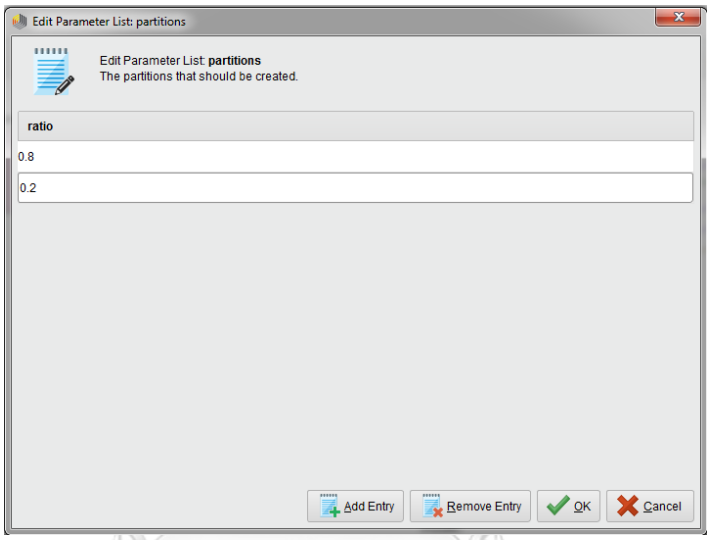
รูปที่ 17 ตัวอย่างฟังก์ชั่นการคัดเลือกตัวแปรด้วยอัลกอริทึมของตัวแบบ Decision tree

3.3.4 การแยกข้อมูล (Split data)

เป็นการสุ่มแยกข้อมูลออกเป็น 2 ส่วนคือ ชุดข้อมูลฝึกหัด (Training dataset) ร้อยละ 80 เพื่อใช้ในการฝึกการเรียนรู้ในการจำแนกของตัวแบบและชุดข้อมูลทดสอบ (Testing dataset) ร้อยละ 20 เพื่อใช้ในการประเมินประสิทธิภาพของตัวแบบพยากรณ์แต่ละชนิดโดยทำการเลือกคำสั่ง Operator > Blending > Examples > Sampling > Split Data



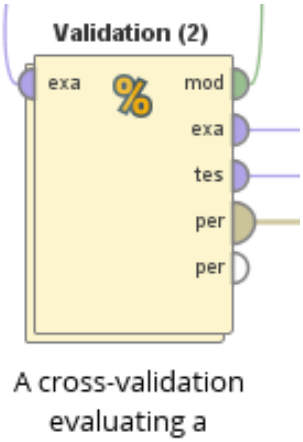
รูปที่ 18 ฟังก์ชั่นการสุ่มแยกข้อมูล



รูปที่ 19 การกำหนดอัตราส่วน Partition ของฟังก์ชั่น Split Data

3.3.5 การฝึกหัดตัวแบบ (Training models)

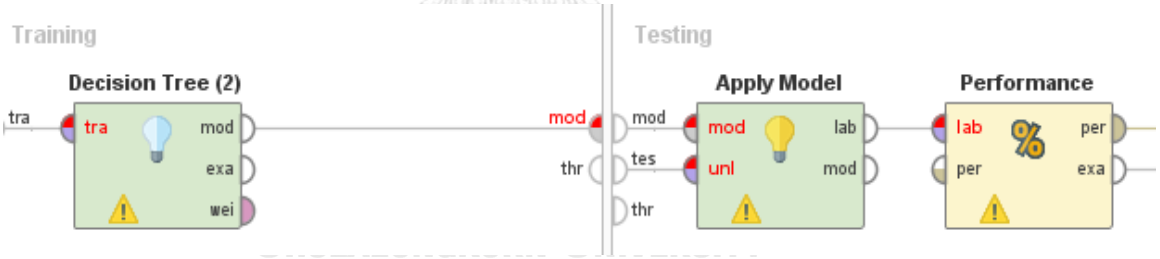
เป็นการฝึกหัดให้ตัวแบบได้เรียนรู้ในการจำแนกกลุ่มระหว่างผู้ที่มีค่าฮีโมโกลบินผ่านเกณฑ์และไม่ผ่านเกณฑ์จากชุดข้อมูลฝึกหัด (Training dataset) โดยทำการสร้าง Building block และเลือก Numerical Cross Validation หรือ Nominal Cross Validation รับข้อมูลจาก Operation Split Data และในช่องของ Training box ให้เลือก Operation ตามชนิดของตัวแบบพยากรณ์ทั้ง 4 ชนิด



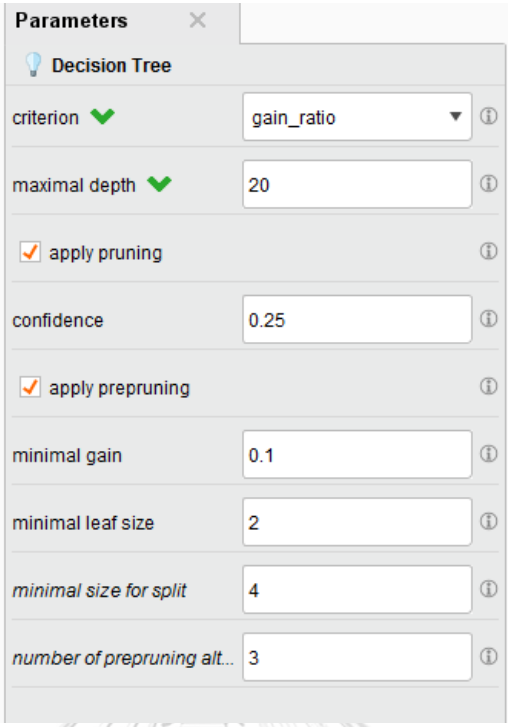
รูปที่ 20 ฟังก์ชั่น Numerical Cross Validation ของการฝึกหัดตัวแบบ

3.3.5.1 ต้นไม้ตัดสินใจ (decision tree)

สร้างต้นไม้ตัดสินใจ (decision tree) ด้วยกันเลือกฟังก์ชั่น Operators > Modeling > Predictive > Tree > Decision Tree จากนั้นนำมาวางในNominal Cross Validation กำหนดเงื่อนไขการ (criterion) แยกลำดับชั้นด้วยค่าอัตราส่วนเกน (gain ratio)



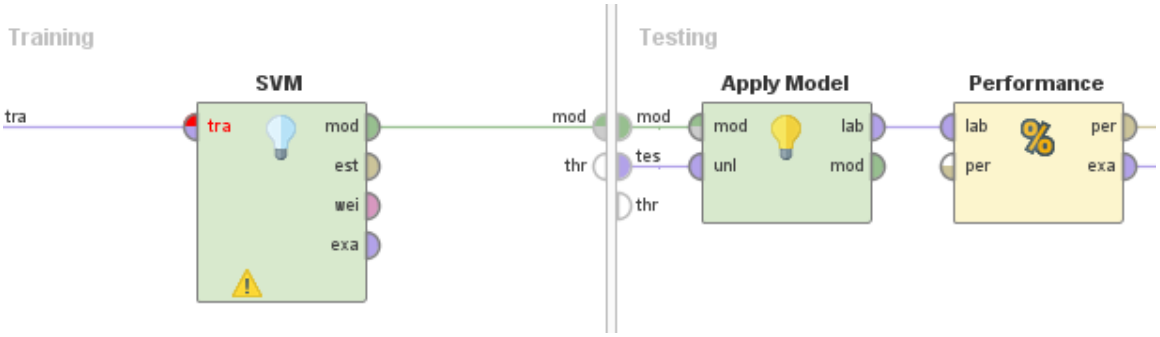
รูปที่ 21 รูปฟังก์ชั่น Decision Tree ใน Nominal Cross Validation



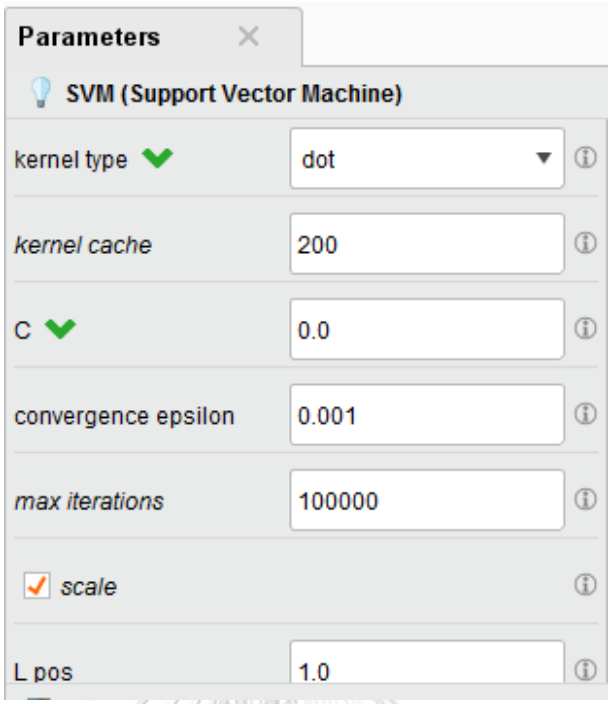
รูปที่ 22 การกำหนดพารามิเตอร์ของ Decision Tree

3.3.5.2 ซัพพอร์ตเวกเตอร์แมชชีน (support vector machine; SVM)

สร้างซัพพอร์ตเวกเตอร์แมชชีน (support vector machine) ด้วยฟังก์ชั่น Operators > Modeling > Predictive > support vector machines > support vector machine จากนั้นนำไปวางใน Numerical Cross Validation จากนั้นกำหนด Kernel type เป็น dot



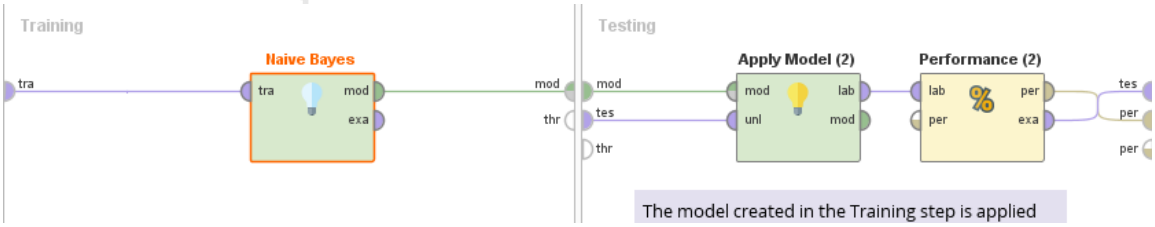
รูปที่ 23 ฟังก์ชั่น SVM ใน Numerical Cross Validation



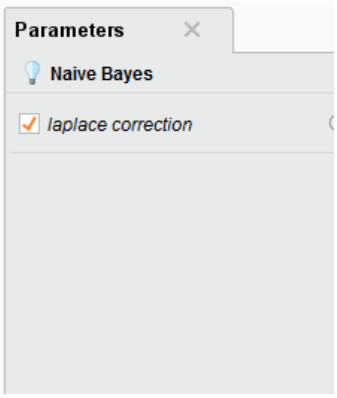
รูปที่ 24 การกำหนดพารามิเตอร์ของ SVM

3.3.5.3 การจำแนกแบบเบย์อย่างง่าย (naïve bayesain classifier)

สร้างการจำแนกแบบเบย์อย่างง่าย (naïve bayesain classifier) ด้วยฟังก์ชั่น Operators > Modeling > Predictive > Bayesain > Naïve Bayes นำไปวางใน Nominal Cross Validation เช่นเดียวกับ Decision Tree



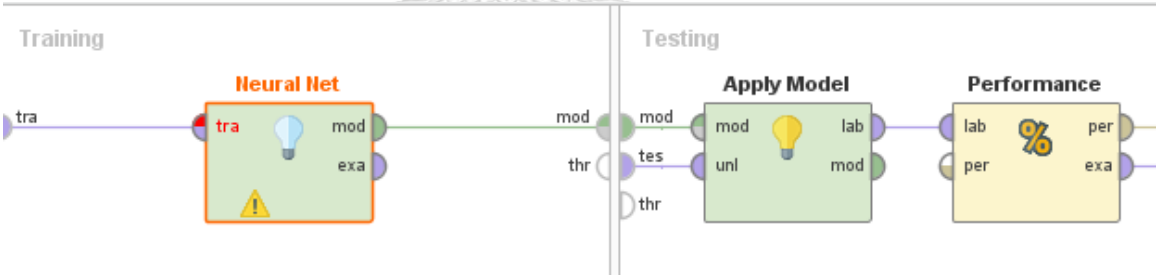
รูปที่ 25 ฟังก์ชั่น Naïve Bayes ใน Nominal Cross Validation



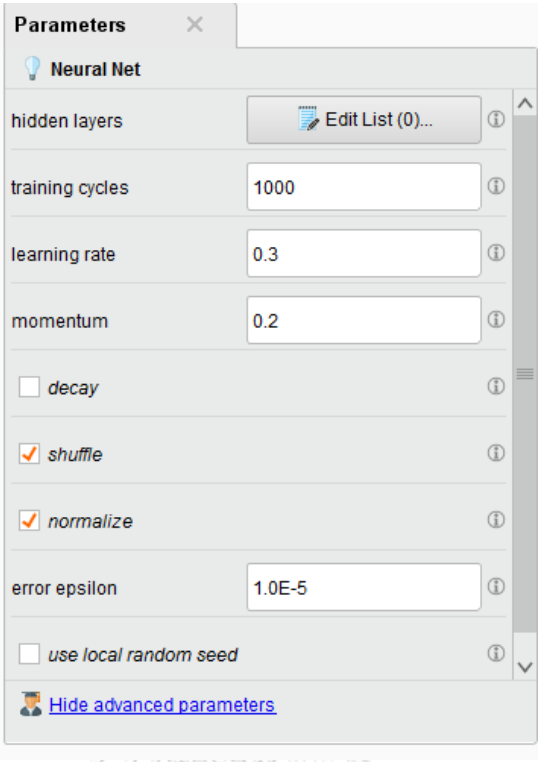
รูปที่ 26 การกำหนดพารามิเตอร์ของ Naïve Bayes

3.3.5.4 โครงข่ายประสาทเทียม (artificial neural networks)

การสร้างโครงข่ายประสาทเทียมโดยเลือกใช้ฟังก์ชั่น Operators > Modeling > Predictive > Neural Nets > Neural Net นำไปวางใน Numerical Cross Validation และกำหนดจำนวนรอบในการสอนเท่ากับ 1,000 รอบ กำหนดการเพิ่มค่าถ่วงน้ำหนักของแต่ละรอบที่ 0.3 กำหนดค่า Momentum เท่ากับ 0.2 และค่ายอมรับความผิดพลาดเท่ากับ 0.00001



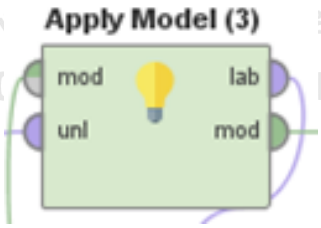
รูปที่ 27 ฟังก์ชั่น Artificial neural networks ใน Numerical Cross Validation



รูปที่ 28 การกำหนดพารามิเตอร์ของโครงข่ายประสาทเทียม

3.3.6 การสร้างตัวแบบที่ได้จากการฝึกหัด (Apply model)

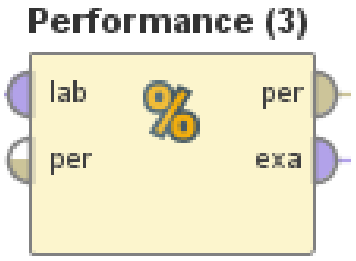
เป็นขั้นตอนการนำตัวแบบที่ได้จากการฝึกหัดมาสร้างเป็นตัวแบบพยากรณ์ โดยเลือก Operator > Scoring > Apply model



รูปที่ 29 ฟังก์ชั่นการสร้างตัวแบบจากตัวแบบที่ได้จากการฝึกหัด

3.4 การประเมินประสิทธิภาพตัวแบบพยากรณ์

ภายหลังจากได้ตัวแบบพยากรณ์แล้วต้องนำชุดข้อมูลทดสอบ (Testing dataset) มาทดสอบตัวแบบพยากรณ์ เพื่อประเมินประสิทธิภาพของตัวแบบแต่ละชนิด โดยเลือก Operation > Validation > Performance > Predictive > Performance



รูปที่ 30 ฟังก์ชั่นประเมินประสิทธิภาพตัวแบบ

การศึกษานี้ทำการวัดผลและเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์ด้วยการวัดผลจากข้อมูล Validation และการวัดผลจากข้อมูลทดสอบด้วยการเปรียบเทียบค่าความถูกต้อง (accuracy) คือค่าร้อยละความถูกต้องในการจำแนกของตัวแบบทั้งในกลุ่มผู้ที่มีค่าฮีโมโกลบินผ่านและไม่ผ่านเกณฑ์ ค่าความไว (sensitivity) คือค่าร้อยละความถูกต้องในจำแนกของตัวแบบเฉพาะในกลุ่มผู้ที่มีค่าฮีโมโกลบินไม่ผ่านเกณฑ์ ค่าความจำเพาะ (specificity) คือค่าร้อยละความถูกต้องในการจำแนกของตัวแบบเฉพาะในกลุ่มที่มีค่าฮีโมโกลบินผ่านเกณฑ์ ค่าการทำนายผลบวก (positive predictive value; PPV) คือค่าร้อยละของความน่าจะเป็นว่าจะมีค่าฮีโมโกลบินไม่ผ่านเกณฑ์เมื่อถูกพยากรณ์จำแนกว่าไม่ผ่านเกณฑ์ ค่าการทำนายผลลบ (negative predictive value: NPV) คือค่าร้อยละของความน่าจะเป็นว่าจะมีค่าฮีโมโกลบินผ่านเกณฑ์เมื่อถูกพยากรณ์จำแนกว่าผ่านเกณฑ์ โดยทั้งหมดสามารถคำนวณได้จากตาราง confusion matrix และค่า AUC ค่าคือพื้นที่ใต้เส้น Receiver Operating Characteristic curve (ROC curve) แสดงถึงความถูกต้องของการพยากรณ์ถ้ามีค่าใกล้ 1 แสดงว่าตัวแบบมีประสิทธิภาพสูง

ตารางที่ 4 Confusion matrix

	True Positive	True Negative
Prediction Positive	a	b
Prediction Negative	c	d

ค่าความถูกต้อง (accuracy) คือค่าที่แสดงว่าตัวแบบพยากรณ์สามารถทำนายได้ถูกต้องเป็นร้อยละเท่าไรคำนวณได้จากสูตรดังนี้

$$Accuracy = \frac{a+d}{a+b+c+d}$$

รูปที่ 31 สูตรการคำนวณค่าความถูกต้อง (Accuracy)

ค่าความระลึก (Class Recall) คือ ค่าความครบถ้วนที่แสดงให้เห็นถึง ความสามารถของระบบในการเลือกข้อมูลหรือคำตอบที่เกี่ยวข้องได้จำนวนมากน้อยเพียงใดแบ่งเป็น 2 ชนิดคือค่าความไวในการตรวจพบค่าบวก (sensitivity) และค่าความจำเพาะในการตรวจพบค่าลบ (specificity) โดยคำนวณจากสมการ

$$Sensitivity = \frac{a}{a+c}$$

รูปที่ 32 สูตรการคำนวณค่าความไวในการตรวจ (Sensitivity)

$$Specificity = \frac{d}{b+d}$$

รูปที่ 33 สูตรการคำนวณค่าความจำเพาะในการตรวจ (Specificity)

ค่าความแม่นยำ (Class Precision) คือ การวัดความสามารถของระบบในการขจัดคำตอบหรือข้อมูลที่ไม่เกี่ยวข้องออกไป ถ้าระบบสามารถขจัดคำตอบหรือข้อมูลที่ไม่เกี่ยวข้องออกไปได้มาก แสดงถึงความแม่นยำของระบบสูง ประกอบด้วยค่าการทำนายผลบวก (positive predictive value; PPV) และค่าการทำนายผลลบ (negative predictive value: NPV) สามารถคำนวณได้จากสมการ

$$PPV = \frac{a}{a+b}$$

รูปที่ 34 สูตรการคำนวณค่าการทำนายผลบวก (positive predictive value; PPV)

$$NPV = \frac{d}{c+d}$$

รูปที่ 35 สูตรการคำนวณค่าการทำนายผลลบ (negative predictive value: NPV)

ค่าพื้นที่ใต้กราฟ ROC (Area Under the Curve; AUC) คือ ค่าแสดงความสามารถในการจำแนกกลุ่มของตัวแบบพยากรณ์หากค่า AUC มีค่าเข้าใกล้ 1 มากนั้นแสดงว่าตัวแบบมีความสามารถในการจำแนกกลุ่มออกจากกันได้อย่างถูกต้อง

บทที่ 4
ผลการทดลอง

การศึกษานี้ทำการเก็บข้อมูลจากแบบสอบถาม โดยมีผู้บริจาคโลหิตตอบแบบสอบถามจำนวนทั้งสิ้น 2,354 ราย จากนั้น ผู้วิจัยได้นำข้อมูลมาศึกษา กลั่นกรองข้อมูลและทำความสะอาดข้อมูล ตรวจสอบความผิดพลาดของข้อมูลพบว่าจากจำนวน 2,354 รายคงเหลือจำนวน 2,180 รายที่มีข้อมูลตัวแปรต่าง ๆ ถูกต้องและจากตัวแปรพยากรณ์จำนวน 43 ตัวแปรพบว่ามี 4 ตัวแปรที่มีผู้ตอบแบบสอบถามน้อยกว่าร้อยละ 70 ของกลุ่มประชากรที่ตอบแบบสอบถามทั้งหมด ได้แก่ ตัวแปรน้ำหนักเมื่อปีที่ผ่านมา (Weight in past a year) น้ำหนักที่เปลี่ยนแปลงในรอบ 1 ปี (Chang Weight in past a Year) ที่อยู่ในเขตเทศบาล (Address2) และขนาดของถุงที่บริจาค (Bag type) คงเหลือตัวแปรพยากรณ์ 39 ตัวแปรและทำการเติมค่าว่างด้วยค่าเฉลี่ยของแต่ละตัวแปรสำหรับตัวแปรเชิงปริมาณและเติมตัวเลขฐานนิยมสำหรับตัวแปรเชิงกลุ่ม แบ่งออกเป็น 2 กลุ่มคือกลุ่มที่มีค่าฮีโมโกลบินผ่านเกณฑ์จำนวน 1,710 รายและไม่ผ่านเกณฑ์ 470 ราย ในการศึกษานี้แจกแจงข้อมูลด้วยโปรแกรม SPSS > Analyze > Descriptive statistics > Descriptives ชดเชยค่าว่างตัวแปรที่เป็นตัวแปรเชิงกลุ่มหรือลำดับชั้นจะชดเชยด้วยเลขฐานนิยม ส่วนตัวแปรที่เป็นค่าตัวเลขหรือค่าต่อเนื่องจะชดเชยด้วยค่าเฉลี่ยของตัวแปรแต่ละตัว ยกเว้นตัวแปรค่าฮีโมโกลบินที่หลังจากทำความเข้าใจข้อมูลและการทดสอบเบื้องต้นของผู้วิจัย พบว่าการใช้ค่าเฉลี่ยร่วมทั้งเพศชายและหญิงนั้นไม่เหมาะสม เพราะเกณฑ์พิจารณาค่า Hb ผ่านเกณฑ์และไม่ผ่านเกณฑ์ระหว่างเพศชายและหญิงนั้นต่างกัน จึงทำการชดเชยค่าเฉลี่ยแยกระหว่างเพศชายและหญิงมีความเหมาะสมกว่า ในเบื้องต้นเมื่อนำข้อมูล 39 ตัวแปรของจำนวนผู้บริจาคโลหิต 2,180 ราย ศึกษาค่าสัมประสิทธิ์ของตัวแปรในการจำแนกกลุ่ม (Classification Function Coefficients) ด้วย SPSS > Analyze > Classify > Discriminant จากนั้นพบ 24 ตัวแปรที่มีความแตกต่างกันระหว่างกลุ่มที่มีผลการตรวจฮีโมโกลบินผ่านและไม่ผ่านเกณฑ์อย่างมีนัยสำคัญทางสถิติที่ p-value น้อยกว่า 0.05

ตารางที่ 5 แจกแจงและการกระจายตัวข้อมูลตัวแปรพยากรณ์

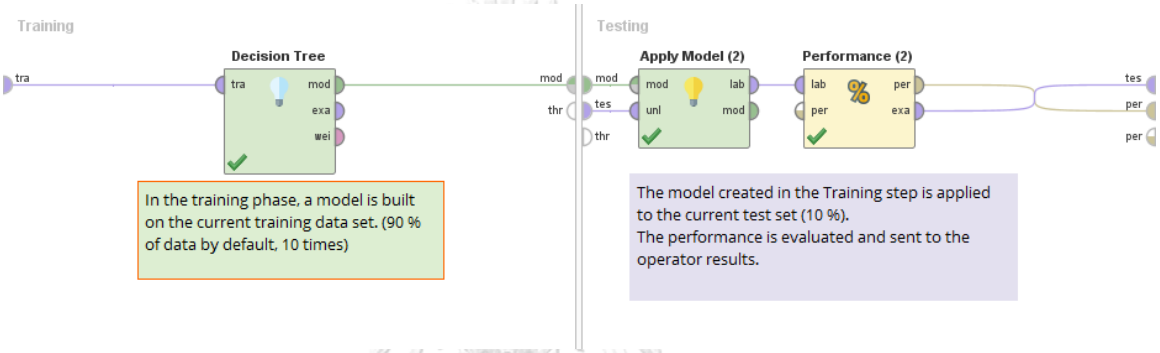
	N	Min	Max	Mean		SD	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic	Std. Error	Statistic	Std. Error
Current Hb	2180	8.60	19.20	13.79	.033	1.53	.28	.052	-.29	.105
Gender	2166	0	1	0.63	.010	0.48	-.53	.053	-1.73	.105
Age	2056	17	72	31.81	.234	10.63	.52	.054	-.50	.108
Menopause	2016	0	1	0.97	.004	0.17	-5.64	.055	29.87	.109
Status	2152	0	2	0.43	.013	0.59	1.01	.053	.02	.105
NumChild	2180	0	5	0.31	.016	0.72	2.50	.052	5.99	.105
Religion	2136	0	3	0.05	.007	0.30	6.58	.053	44.42	.106
High	2127	105	190	164.03	.180	8.29	-.01	.053	1.93	.106
Weight	2155	45.00	130.00	65.46	.306	14.22	1.00	.053	.96	.105
BMI	2124	15.19	74.38	24.29	.099	4.56	1.61	.053	7.89	.106
Weight in past 3 month	1639	44.00	129.00	65.04	.358	14.49	1.02	.060	.97	.121
ChangeWeight in past 3 m	1639	-16.00	33.00	0.39	.072	2.89	1.17	.060	17.62	.121
Education	2148	0	6	3.27	.025	1.16	-.66	.053	-.31	.106
Address1	1981	0	2	0.45	.011	0.50	.25	.055	-1.85	.110
Occupation	2114	0	7	2.58	.043	2.00	.33	.053	-.97	.106
Income	1941	0	3	1.14	.021	0.93	.41	.056	-.71	.111
Donation Place	2032	0	4	0.97	.023	1.03	.33	.054	-1.46	.109
Single Donor	2057	0	1	0.96	.004	0.19	-4.95	.054	22.52	.108
Num Single Donor	2044	0	66	0.21	.048	2.17	19.57	.054	488.26	.108
Donation/year	1920	0	19	1.40	.027	1.18	2.24	.056	27.03	.112
Donation	1959	1	107	10.76	.307	13.59	2.99	.055	11.96	.111
Interval donation	1784	1.0	120.0	7.16	.227	9.60	5.64	.058	44.44	.116
Donation success	1812	0	4	0.02	.004	0.17	15.04	.057	278.97	.115
Previous Hb	2180	8.60	19.20	13.86	.032	1.47	.37	.052	-.14	.105
BP Sys	2061	97	190	124.09	.316	14.34	.48	.054	-.16	.108
BP Dias	2062	40	110	77.14	.223	10.11	.19	.054	-.31	.108
Pulse	1979	34	100	82.60	.247	10.97	-.22	.055	-.45	.110
EverLowHb	2047	0	1	0.33	.010	0.47	.71	.054	-1.49	.108
Disease	2091	0	7	0.21	.021	0.95	4.70	.054	21.66	.107
Public to Known	2016	0	7	4.45	.046	2.08	-.36	.055	-1.31	.109
Sleep Type	1869	0	6	0.35	.012	0.52	1.71	.057	8.88	.113
Sleep hour	1957	0	6	1.60	.012	0.52	-.21	.055	1.37	.111
Exercise	2111	0	10	1.37	.019	0.87	.71	.053	3.99	.106
Smoke	2158	0	2	0.08	.006	0.27	3.25	.053	8.92	.105
Alcohol take	2139	0	4	0.30	.012	0.57	2.04	.053	4.91	.106
Food Type	2151	0	4	0.06	.009	0.42	7.27	.053	54.47	.106
Fe take	2114	0	3	1.14	.020	0.92	-.25	.053	-1.72	.106
Fe Why not	1722	0	6	2.07	.047	1.93	.79	.059	-.79	.118

ตารางที่ 6 ตัวแปรพยากรณ์ที่มีความแตกต่างกันระหว่างผู้ที่มีค่า Hb ผ่านและไม่ผ่านเกณฑ์

	Wilks' Lambda	F	df1	df2	Sig.
Gender	.900	243.101	1	2178	.000
Age	.996	9.272	1	2178	.002
Menopause	.991	19.226	1	2178	.000
Status	.996	9.721	1	2178	.002
No.Child	.997	6.174	1	2178	.013
Religion	1.000	.592	1	2178	.442
High	.962	86.325	1	2178	.000
Weight	.967	73.392	1	2178	.000
BMI	.992	18.473	1	2178	.000
Weight in past 3 Month	.972	63.783	1	2178	.000
Change Weight in past 3 m	1.000	.337	1	2178	.562
Education	1.000	.019	1	2178	.890
Address1	1.000	.007	1	2178	.936
Occupation	.994	14.213	1	2178	.000
Income	.990	22.262	1	2178	.000
ABO	.999	1.706	1	2178	.192
Donation Place	.999	1.384	1	2178	.240
Single Donor	1.000	.384	1	2178	.535
No.Single Donor	1.000	.566	1	2178	.452
Donation/Year	.977	51.496	1	2178	.000
Donation	.984	35.501	1	2178	.000
Interval donation	1.000	.003	1	2178	.954
Donation success	.999	2.534	1	2178	.112
Previous Hb	.666	1093.893	1	2178	.000
BP Sys	.990	22.704	1	2178	.000
BP Dias	.998	5.362	1	2178	.021
Pulse	.995	11.092	1	2178	.001
EverLowHb	.881	294.758	1	2178	.000
Disease	1.000	.938	1	2178	.333
Public to Known	1.000	.078	1	2178	.779
Sleep Type	.997	7.178	1	2178	.007
Sleep hour	.983	37.573	1	2178	.000
Exercise	.982	40.443	1	2178	.000
Smoke	.991	19.929	1	2178	.000
Alcohol take	.988	25.571	1	2178	.000
Food Type	1.000	.830	1	2178	.362
Fe take	1.000	.016	1	2178	.898
Fe Why not	.990	21.655	1	2178	.000

งานวิจัยนี้เป็นการสร้างตัวแบบพยากรณ์ 4 ชนิดด้วยโปรแกรม RapidMiner โดยแต่ละชนิดจะใช้วิธีการคัดเลือกตัวแปรต้น 4 วิธี ด้วยข้อมูลฝึกหัด (Training dataset) ร้อยละ 80 จำนวน 1,744 รายและทำการทดสอบประเมินประสิทธิภาพด้วยข้อมูลทดสอบ (Testing dataset) ร้อย 20 จำนวน 436 ราย จากนั้นเลือกวิธีการคัดเลือกตัวแปรต้นที่ให้ผลการพยากรณ์ที่ดีที่สุดของแต่ละตัวแบบพยากรณ์และนำไปเปรียบเทียบประสิทธิภาพระหว่างตัวแบบพยากรณ์ที่ดีที่สุดของทั้ง 4 ชนิด ซึ่งมีผลการวิจัยดังนี้

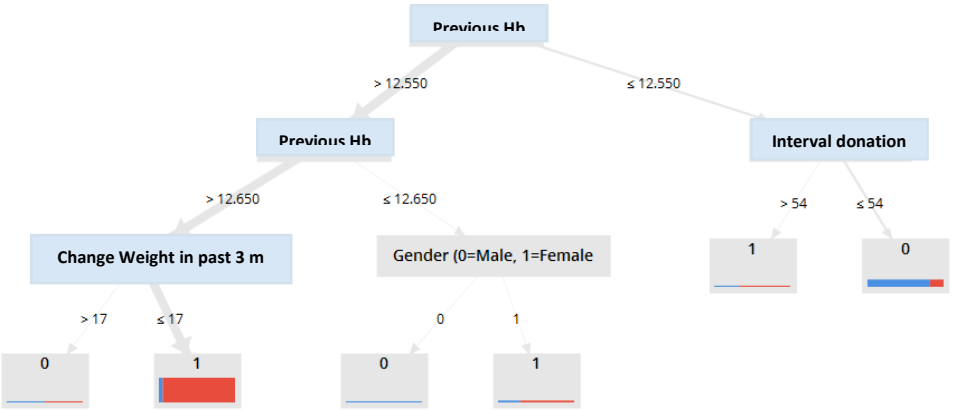
4.1 ผลการทดลองของตัวแบบต้นไม้ตัดสินใจ (decision tree)



รูปที่ 36 การสร้างตัวแบบต้นไม้ตัดสินใจ (decision tree) จากการฝึกหัด

4.1.1 ผลการคัดเลือกตัวแปรวิธีนำตัวแปรเข้าทั้งหมด (enter regression)

เมื่อนำตัวแปรทั้งหมดมาสร้างตัวแบบพยากรณ์ด้วยเทคนิคต้นไม้ตัดสินใจ (Decision tree) พบตัวแปรที่ให้ค่าเกน (Gain) สูงในการสร้างต้นไม้ตัดสินใจ (Decision tree) มี 4 ตัวแปรได้แก่ เพศ (Gender) น้ำหนักที่เปลี่ยนแปลงใน 3 เดือน (Change Weight in past 3 Month) ค่าฮีโมโกลบินครั้งที่ผ่านมา (Previous Hb) และระยะห่างการบริจาคครั้งที่ผ่านมาถึงปัจจุบัน (Interval donation) ประกอบกันเป็นต้นไม้ตัดสินใจที่มี 5 โหนด โดยมีตัวแปรค่าฮีโมโกลบินครั้งที่ผ่านมา (Previous Hb) เป็นโหนดรากและตัวแปรค่าฮีโมโกลบินครั้งที่ผ่านมา (Previous Hb) น้ำหนักที่เปลี่ยนแปลงใน 3 เดือน (Change Weight in past 3 Month) เพศ (Gender) และระยะห่างการบริจาคครั้งที่ผ่านมาถึงปัจจุบัน (Interval donation) เป็นโหนดกิ่ง (รูปที่ 37)



รูปที่ 37 ต้นไม้ตัดสินใจที่ได้จากการนำเข้าตัวแปรทั้งหมด (Enter regression)

ทำการ Validation ตัวแบบพยากรณ์ด้วยชุดข้อมูลฝึกหัด (Training dataset) พบว่าให้ค่าความถูกต้อง (accuracy) เท่ากับร้อยละ 90.65 ค่าความไว (sensitivity) ร้อยละ 76.33 ค่าความจำเพาะ (specificity) ร้อยละ 94.59 ค่าการทำนายผลบวก (positive predictive value) ร้อยละ 79.50 และค่าการทำนายผลลบ (negative predictive value) ร้อยละ 93.56 (ตารางที่ 6) และประเมินประสิทธิภาพตัวแบบพยากรณ์ด้วยข้อมูลทดสอบ (Testing dataset) พบว่าให้ค่าความถูกต้อง (accuracy) เท่ากับร้อยละ 91.06 ค่าความไว (sensitivity) ร้อยละ 79.79 ค่าความจำเพาะ (specificity) ร้อยละ 94.15 ค่าการทำนายผลบวก (positive predictive value) ร้อยละ 78.95 และค่าการทำนายผลลบ (negative predictive value) ร้อยละ 94.43 (ตารางที่ 7) โดยมีค่า AUC เท่ากับ 0.879 (รูปที่ 38)

ตารางที่ 7 Confusion matrix of training dataset using decision tree model with enter regression variables

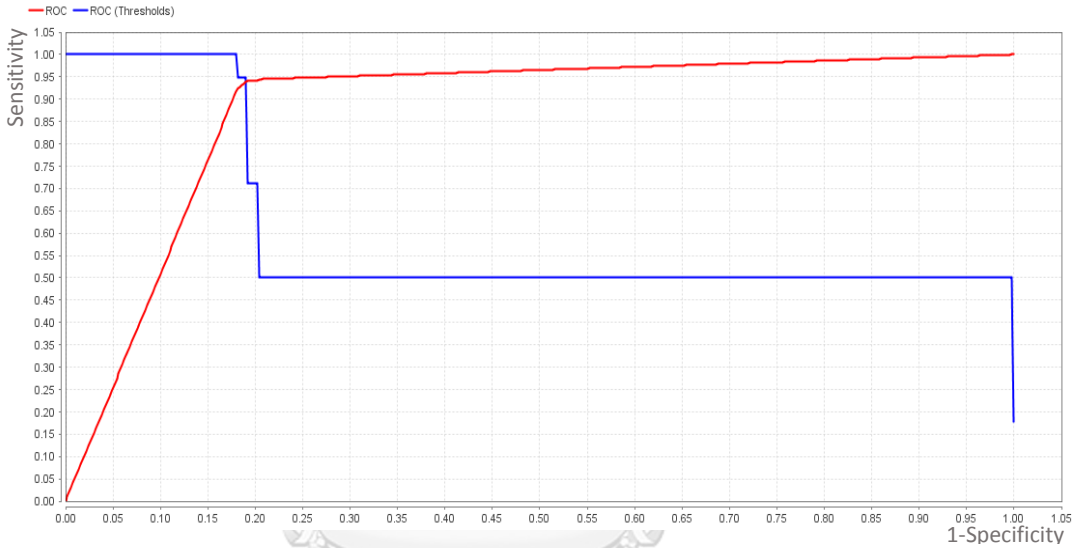
	Hb ไม่ผ่านเกณฑ์	Hb ผ่านเกณฑ์	Class precision**
พยากรณ์ Hb ไม่ผ่านเกณฑ์	287	74	79.50%
พยากรณ์ Hb ผ่านเกณฑ์	89	1294	93.56%
Class recall*	76.33%	94.59%	

*Class recall ประกอบด้วยค่า Sensitivity และ Specificity

**Class precision ประกอบด้วยค่า PPV และค่า NPV

ตารางที่ 8 Confusion matrix of testing dataset using decision tree model with enter regression variables

	Hb ไม่ผ่านเกณฑ์	Hb ผ่านเกณฑ์	Class precision
พยากรณ์ Hb ไม่ผ่านเกณฑ์	75	20	78.95%
พยากรณ์ Hb ผ่านเกณฑ์	19	322	94.43%
Class recall	79.79%	94.15%	

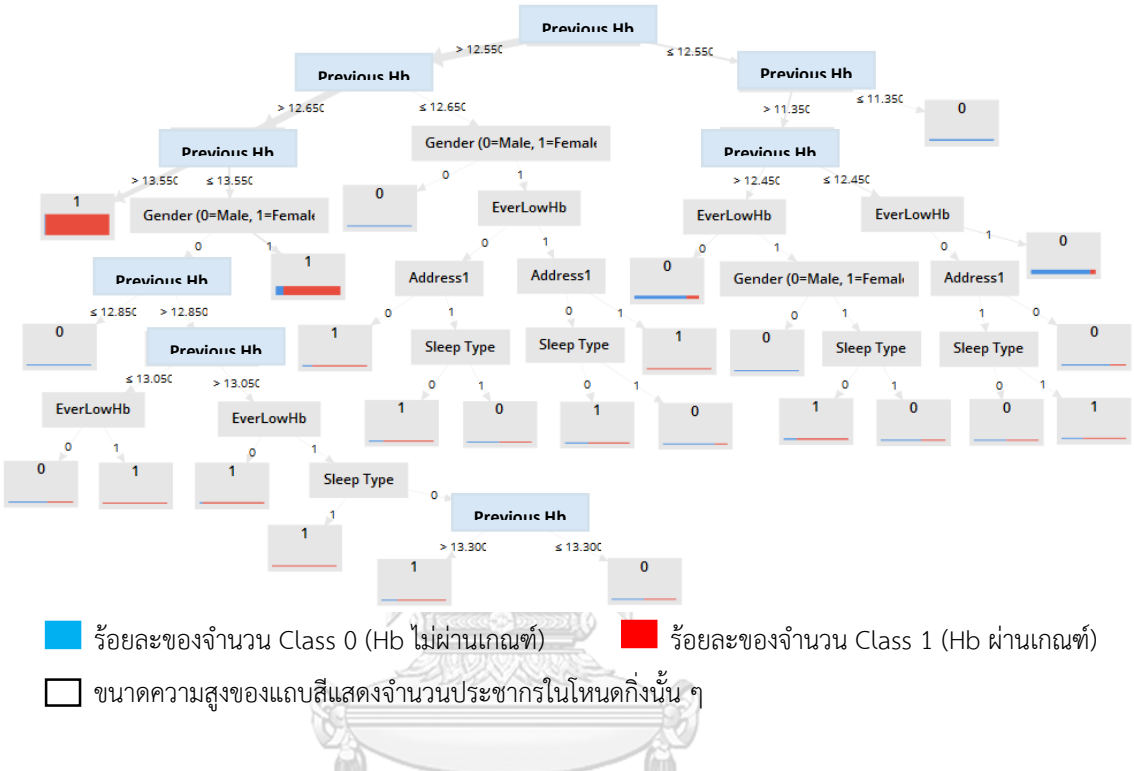


รูปที่ 38 Area under the curve (AUC) of testing dataset using decision tree model with enter regression variables

4.1.2 ผลการคัดเลือกตัวแปรวิธีเพิ่มตัวแปร (forward selection)

ทำการคัดเลือกตัวแปรด้วยวิธีเพิ่มตัวแปร (forward selection) จากนั้นนำตัวแปรที่ได้มาสร้างตัวแบบพยากรณ์ด้วยเทคนิคต้นไม้ตัดสินใจ (Decision tree) พบตัวแปรที่ให้ค่าเกนสูงในการสร้างต้นไม้ตัดสินใจ (Decision tree) มี 5 ตัวแปรคือ เพศ (Gender) ประวัติการตรวจฮีโมโกลบินไม่ผ่านเกณฑ์ (EverLowHb) ค่าฮีโมโกลบินครั้งที่ผ่านมา (Previous Hb) ที่อยู่ (Address1) และพฤติกรรมการพักผ่อน (Sleep type) ประกอบกันเป็นต้นไม้ตัดสินใจจำนวน 24 โหนด โดยมีตัวแปรค่าฮีโมโกลบินครั้งที่ผ่านมา (Previous Hb) เป็นโหนดรากและตัวแปรค่าฮีโมโกลบินครั้งที่ผ่าน

มา (Previous Hb) เป็นโหนดกิ่งจำนวน 7 โหนด ตัวแปรเพศ (Gender) เป็นโหนดกิ่งจำนวน 3 โหนด ตัวแปรประวัติการตรวจฮีโมโกลบินไม่ผ่านเกณฑ์ (EverLowHb) เป็นโหนดกิ่งจำนวน 5 โหนด ตัวแปรที่อยู่ (Address1) เป็นโหนดกิ่งจำนวน 3 โหนดและตัวแปรพฤติกรรมการพักผ่อน (Sleep type) เป็นโหนดกิ่งจำนวน 5 โหนด (รูปที่ 39)



รูปที่ 39 ต้นไม้ตัดสินใจที่ได้จากการคัดเลือกตัวแปรวิธีเพิ่มตัวแปร (forward selection)

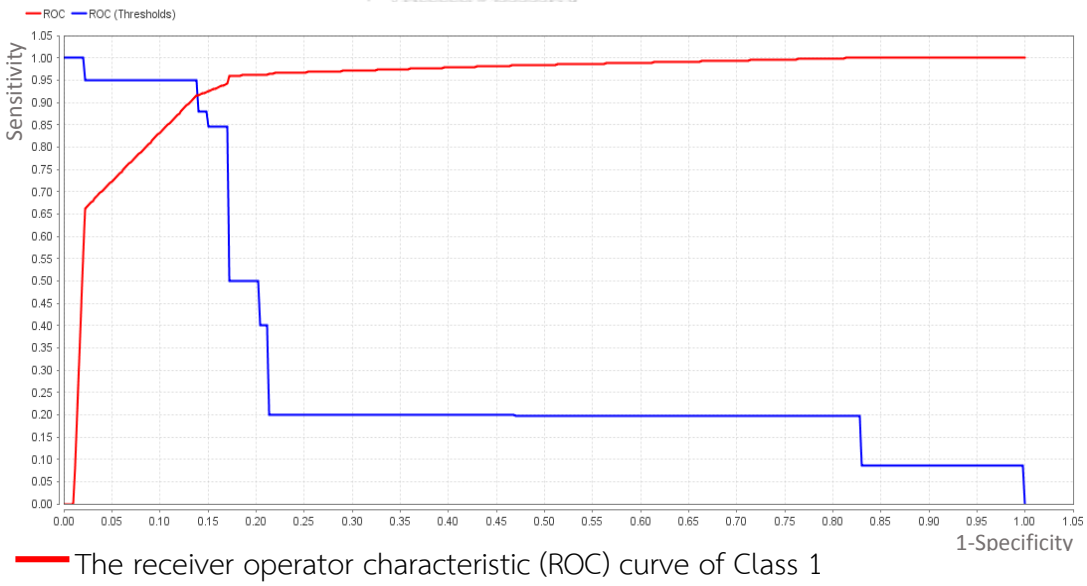
ทำการ Validation ตัวแบบพยากรณ์ด้วยชุดข้อมูลฝึกหัด (Training dataset) พบว่าให้ค่าความถูกต้อง (accuracy) เท่ากับร้อยละ 92.26 ค่าความไว (sensitivity) ร้อยละ78.19 ค่าความจำเพาะ (specificity) ร้อยละ 96.13 ค่าการทำนายผลบวก (positive predictive value) ร้อยละ 84.73 และค่าการทำนายผลลบ (negative predictive value) ร้อยละ 94.13 (ตารางที่ 8) และประเมินประสิทธิภาพตัวแบบพยากรณ์ด้วยข้อมูลทดสอบ (Testing dataset) พบว่าให้ค่าความถูกต้อง (accuracy) เท่ากับร้อยละ 92.20 ค่าความไว (sensitivity) ร้อยละ 82.98 ค่าความจำเพาะ (specificity) ร้อยละ 94.74 ค่าการทำนายผลบวก (positive predictive value) ร้อยละ 81.25 และค่าการทำนายผลลบ (negative predictive value) ร้อยละ 95.29 (ตารางที่ 9) โดยมีค่า AUC เท่ากับ 0.943 (รูปที่ 40)

ตารางที่ 9 Confusion matrix of training dataset using decision tree model with forward selection variables

	Hb ไม่ผ่านเกณฑ์	Hb ผ่านเกณฑ์	Class precision
พยากรณ์ Hb ไม่ผ่านเกณฑ์	294	53	84.73%
พยากรณ์ Hb ผ่านเกณฑ์	82	1315	94.13%
Class recall	78.19%	96.13%	

ตารางที่ 10 Confusion matrix of testing dataset using decision tree model with forward selection variables

	Hb ไม่ผ่านเกณฑ์	Hb ผ่านเกณฑ์	Class precision
พยากรณ์ Hb ไม่ผ่านเกณฑ์	78	18	81.25%
พยากรณ์ Hb ผ่านเกณฑ์	16	324	95.29%
Class recall	82.98%	94.74%	



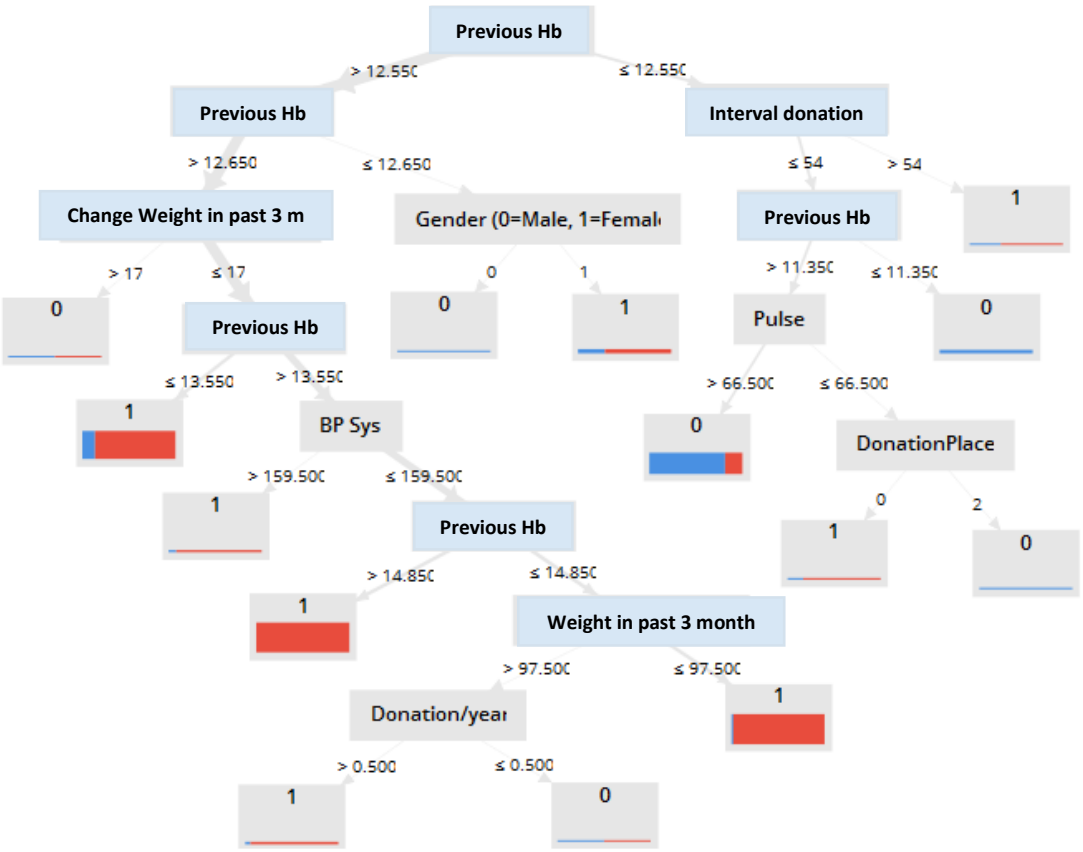
The receiver operator characteristic (ROC) curve of Class 1

ROC (Threshold) show confidence cut off

รูปที่ 40 Area under the curve (AUC) of testing dataset using decision tree model with forward selection variables

4.1.3 ผลการคัดเลือกตัวแปรวิธีลดตัวแปร (backward elimination)

ทำการคัดเลือกตัวแปรด้วยวิธีลดตัวแปร (backward elimination) จากจำนวนตัวแปร 39 ตัวแปรคงเหลือ 38 ตัวแปรเมื่อนำตัวแปรที่ได้มาสร้างตัวแบบพยากรณ์ด้วยเทคนิคต้นไม้ตัดสินใจ (Decision tree) พบตัวแปรที่ให้ค่าเกนสูงในการสร้างต้นไม้ตัดสินใจ (Decision tree) มี 9 ตัวแปรคือ เพศ (Gender) น้ำหนักเมื่อครั้งที่ผ่านมา (Weight in past 3 month) น้ำหนักที่เปลี่ยนแปลใน 3 เดือน (Chang Weight in past 3 month) ค่าฮีโมโกลบินครั้งที่ผ่านมา (Previous Hb) ความดันโลหิต Systolic (BP Sys) อัตราการเต้นหัวใจ (Pulse) ความถี่การบริจาค WB ในรอบปี (Donation/year) สถานที่บริจาคโลหิต (Donation place) และระยะห่างการบริจาค (Interval donation) ประกอบกันเป็นต้นไม้ตัดสินใจที่มี 13 โหนด โดยมีตัวแปรค่าฮีโมโกลบินครั้งที่ผ่านมา (Previous Hb) เป็นโหนดรากและตัวแปรค่าฮีโมโกลบินครั้งที่ผ่านมา (Previous Hb) เป็นโหนดกิ่งจำนวน 4 โหนด ตัวแปรเพศ (Gender) เป็นโหนดกิ่งจำนวน 1 โหนด ตัวแปรน้ำหนักเมื่อครั้งที่ผ่านมา (Weight in past 3 month) เป็นโหนดกิ่งจำนวน 1 โหนด ตัวแปรน้ำหนักที่เปลี่ยนแปลใน 3 เดือน (Chang Weight in past 3 month) เป็นโหนดกิ่งจำนวน 1 โหนด ตัวแปรความดันโลหิต Systolic (BP Sys) เป็นโหนดกิ่งจำนวน 1 โหนด ตัวแปรอัตราการเต้นหัวใจ (Pulse) เป็นโหนดกิ่งจำนวน 1 โหนด ตัวแปรความถี่การบริจาค WB ในรอบปี (Donation/year) เป็นโหนดกิ่งจำนวน 1 โหนด ตัวแปรสถานที่บริจาคโลหิต (Donation place) เป็นโหนดกิ่งจำนวน 1 โหนด และตัวแปรระยะห่างการบริจาค (Interval donation) เป็นโหนดกิ่งจำนวน 1 โหนด (รูปที่ 41)



ร้อยละของจำนวน Class 0 (Hb ไม่ผ่านเกณฑ์) ร้อยละของจำนวน Class 1 (Hb ผ่านเกณฑ์)

ขนาดความสูงของแถบสีแสดงจำนวนประชากรในโหนดกิ่งนั้น ๆ

รูปที่ 41 ต้นไม้ตัดสินใจที่ได้จากการคัดเลือกตัวแปรวิธีลดตัวแปร (backward elimination)

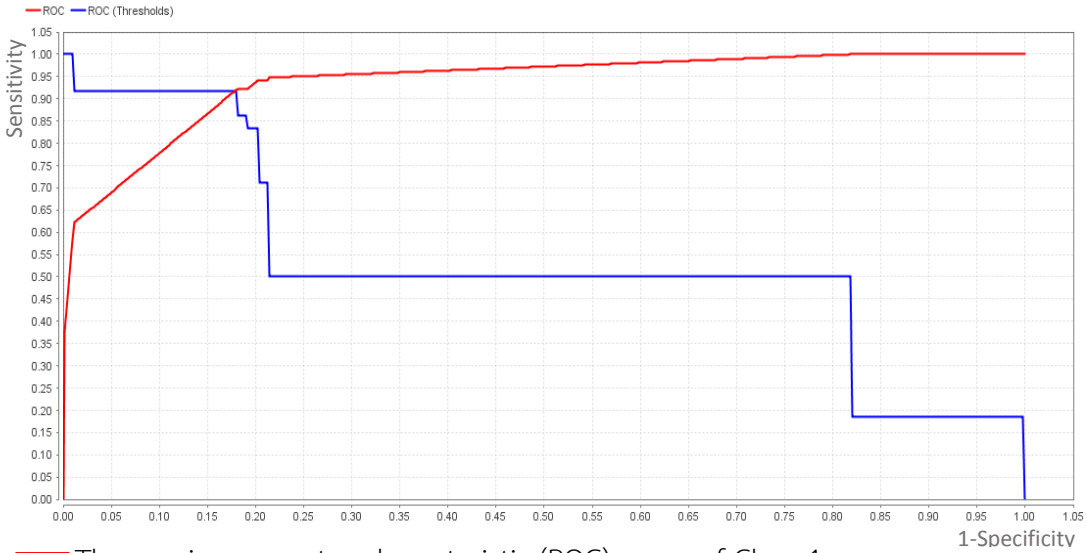
ทำการ Validation ตัวแบบพยากรณ์ด้วยชุดข้อมูลฝึกหัด (Training dataset) พบว่าให้ค่าความถูกต้อง (accuracy) เท่ากับร้อยละ 90.37 ค่าความไว (sensitivity) ร้อยละ75.27 ค่าความจำเพาะ (specificity) ร้อยละ 94.52 ค่าการทำนายผลบวก (positive predictive value) ร้อยละ 79.05 และค่าการทำนายผลลบ (negative predictive value) ร้อยละ 93.29 (ตารางที่ 10) และประเมินประสิทธิภาพตัวแบบพยากรณ์ด้วยข้อมูลทดสอบ (Testing dataset) พบว่าให้ค่าความถูกต้อง (accuracy) เท่ากับร้อยละ 90.83 ค่าความไว (sensitivity) ร้อยละ 78.72 ค่าความจำเพาะ (specificity) ร้อยละ 94.15 ค่าการทำนายผลบวก (positive predictive value) ร้อยละ 78.72 และค่าการทำนายผลลบ (negative predictive value) ร้อยละ 94.15 (ตารางที่ 11) โดยมีค่า AUC เท่ากับ 0.937 (รูปที่ 42)

ตารางที่ 11 Confusion matrix of training dataset using decision tree model with backward elimination variables

	Hb ไม่ผ่านเกณฑ์	Hb ผ่านเกณฑ์	Class precision
พยากรณ์ Hb ไม่ผ่านเกณฑ์	283	75	79.05%
พยากรณ์ Hb ผ่านเกณฑ์	93	1293	93.29%
Class recall	75.27%	94.52%	

ตารางที่ 12 Confusion matrix of testing dataset using decision tree model with backward elimination variables

	Hb ไม่ผ่านเกณฑ์	Hb ผ่านเกณฑ์	Class precision
พยากรณ์ Hb ไม่ผ่านเกณฑ์	74	20	78.72%
พยากรณ์ Hb ผ่านเกณฑ์	20	322	94.15%
Class recall	78.72%	94.15%	



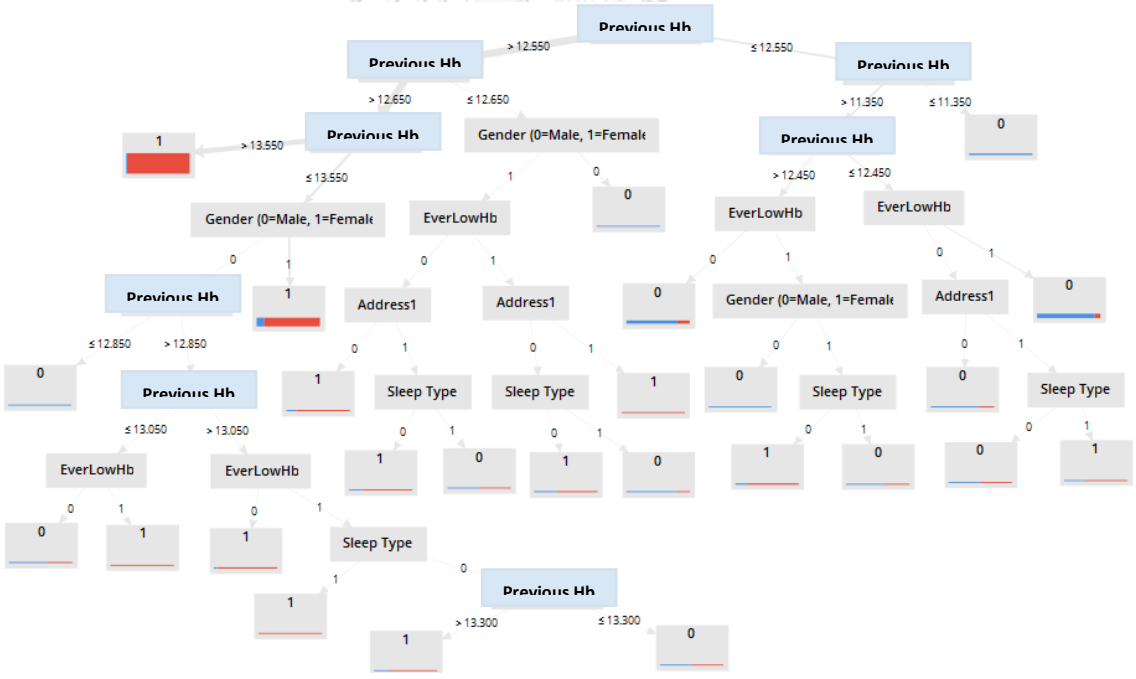
The receiver operator characteristic (ROC) curve of Class 1

ROC (Threshold) show confidence cut off

รูปที่ 42 Area under the curve (AUC) of testing dataset using decision tree model with backward elimination variables

4.1.4 ผลการคัดเลือกตัวแปรวิธีเพิ่มตัวแปรและลดตัวแปร (Optimize selection)

ทำการคัดเลือกตัวแปรด้วยวิธีเพิ่มตัวแปรและลดตัวแปร (Optimize selection) จากจำนวนตัวแปร 39 ตัวแปรคงเหลือ 5 ตัวแปรเมื่อนำตัวแปรที่ได้มาสร้างตัวแบบพยากรณ์ด้วยเทคนิคต้นไม้ตัดสินใจ (Decision tree) พบตัวแปรที่ให้ค่าเกนสูงในการสร้างต้นไม้ตัดสินใจ (Decision tree) ทั้ง 5 ตัวแปรคือ เพศ (Gender) ประวัติการตรวจฮีโมโกลบินไม่ผ่านเกณฑ์ (EverLowHb) ค่าฮีโมโกลบินครั้งที่ผ่านมา (Previous Hb) ที่อยู่ (Address1) และพฤติกรรมการพักผ่อน (Sleep type) ประกอบกันเป็นต้นไม้ตัดสินใจที่มี 24 โหนด โดยมีตัวแปรค่าฮีโมโกลบินครั้งที่ผ่านมา (Previous Hb) เป็นโหนดรากและตัวแปรค่าฮีโมโกลบินครั้งที่ผ่านมา (Previous Hb) เป็นโหนดกิ่งจำนวน 7 โหนด ตัวแปรเพศ (Gender) เป็นโหนดกิ่งจำนวน 3 โหนด ตัวแปรประวัติการตรวจฮีโมโกลบินไม่ผ่านเกณฑ์ (EverLowHb) เป็นโหนดกิ่งจำนวน 5 โหนด ตัวแปรที่อยู่ (Address1) เป็นโหนดกิ่งจำนวน 3 โหนดและตัวแปรพฤติกรรมการพักผ่อน (Sleep type) เป็นโหนดกิ่งจำนวน 5 โหนด (รูปที่ 43)



ร้อยละของจำนวน Class 0 (Hb ไม่ผ่านเกณฑ์) ร้อยละของจำนวน Class 1 (Hb ผ่านเกณฑ์)

ขนาดความสูงของแถบสีแสดงจำนวนประชากรในโหนดกิ่งนั้น ๆ

รูปที่ 43 ต้นไม้ตัดสินใจที่ได้จากการคัดเลือกตัวแปรวิธีเพิ่มตัวแปรและลดตัวแปร (Optimize selection)

ทำการ Validation ตัวแบบพยากรณ์ด้วยชุดข้อมูลฝึกหัด (Training dataset) พบว่าให้ค่าความถูกต้อง (accuracy) เท่ากับร้อยละ 92.26 ค่าความไว (sensitivity) ร้อยละ78.19 ค่าความจำเพาะ (specificity) ร้อยละ 96.13 ค่าการทำนายผลบวก (positive predictive value) ร้อยละ 84.73 และค่าการทำนายผลลบ (negative predictive value) ร้อยละ 94.13 (ตารางที่ 12) และประเมินประสิทธิภาพตัวแบบพยากรณ์ด้วยข้อมูลทดสอบ (Testing dataset) พบว่าให้ค่าความถูกต้อง (accuracy) เท่ากับร้อยละ 92.20 ค่าความไว (sensitivity) ร้อยละ 82.98 ค่าความจำเพาะ (specificity) ร้อยละ 94.74 ค่าการทำนายผลบวก (positive predictive value) ร้อยละ 81.25 และค่าการทำนายผลลบ (negative predictive value) ร้อยละ 95.29 (ตารางที่ 13) โดยมีค่า AUC เท่ากับ 0.943 (รูปที่ 44)

ตารางที่ 13 Confusion matrix of training dataset using decision tree model with optimize selection variables

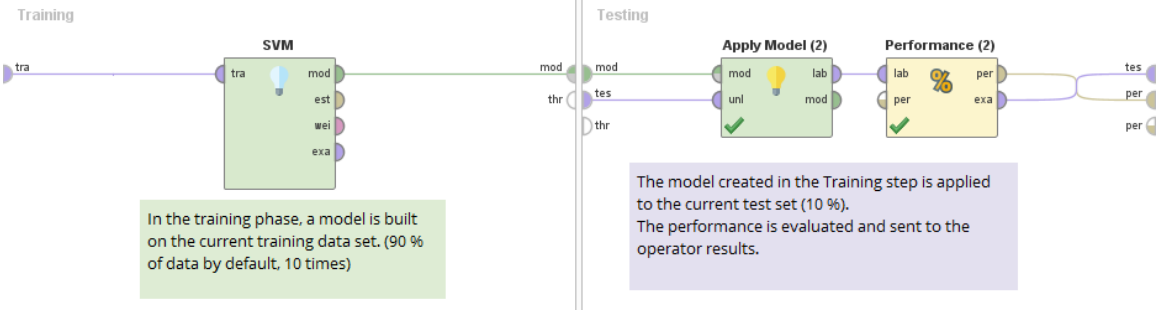
	Hb ไม่ผ่านเกณฑ์	Hb ผ่านเกณฑ์	Class precision
พยากรณ์ Hb ไม่ผ่านเกณฑ์	294	53	84.73%
พยากรณ์ Hb ผ่านเกณฑ์	82	1315	94.13%
Class recall	78.19%	96.13%	

ตารางที่ 14 Confusion matrix of testing dataset using decision tree model with optimize selection variables

	Hb ไม่ผ่านเกณฑ์	Hb ผ่านเกณฑ์	Class precision
พยากรณ์ Hb ไม่ผ่านเกณฑ์	78	18	81.25%
พยากรณ์ Hb ผ่านเกณฑ์	16	324	95.29%
Class recall	82.98%	94.74%	

รูปที่ 44 Area under the curve (AUC) of testing dataset using decision tree model with optimize selection variables

4.2 ผลการทดลองของซัพพอร์ตเวกเตอร์แมชชีน (support vector machine; SVM)



รูปที่ 45 การสร้างตัวแบบซัพพอร์ตเวกเตอร์แมชชีน (support vector machine; SVM) จากการฝึกหัด

ในการสร้างตัวแบบพยากรณ์ด้วย SVM พบว่าตัวแปร Rh มีค่าเป็น Rh positive เหมือนกันเป็นส่วนใหญ่ ในการคัดเลือกตัวแปรจำเป็นต้องนำออกก่อนทำการสร้างและฝึกหัดของตัวแบบ

4.2.1 ผลการคัดเลือกตัวแปรวิธีนำตัวแปรเข้าทั้งหมด (enter regression)

เมื่อนำแปรทั้งหมดมาคำนวณหาค่าน้ำหนักของตัวแปร พบว่าตัวแปรค่าฮีโมโกลบินครั้งที่ผ่านมา (Previous Hb) ให้ค่าน้ำหนักมากที่สุดคือ 1.780 และค่าไบแอสเท่ากับ 1.550

ตารางที่ 15 ค่าน้ำหนักตัวแปรของตัวแบบ SVM โดยการนำตัวแปรเข้าทั้งหมด (enter regression)

Kernel Model	
Attribute	**Weight**
Bias (offset)	1.550
Gender	0.2892105217
Age	-0.1032930991
Menopause	0.0218719072
Status	0.1292183984
No.Child	0.0522317978
Religion	-0.0653763637
High	-0.0899014806
Weight	0.1266697730
BMI	-0.1338682115
Weight in past 3 month	0.0482183837
Change Weight in past 3 Month	0.1217836267
Education	0.0903634963
Address1	0.0402469309
Occupation	0.0960179702
Income	0.0905994650
ABO	0.1392876710
Donation Place	0.1557573683
Single Donor	0.0233343180
No. Single Donor	-0.0267430420
Donation/year	0.0218532060
Donation	0.0022727539
Interval donation	0.0455747974
Donation success	-0.0533309625
Previous Hb	1.7796973135
BP Sys	-0.1160615742
BP Dias	0.0375256198
Pulse	-0.0732845574
EverLowHb	-0.2734668167

Disease	0.0339891174
Public to Known	0.0265744991
Sleep Type	-0.0680850414
Sleep hour	0.2538465280
Exercise	0.1549656638
Smoke	0.0269698003
Alcohol take	-0.0431818660
Food Type	0.0515228545
Fe take	-0.0498952140
Fe Why not	-0.1026760653

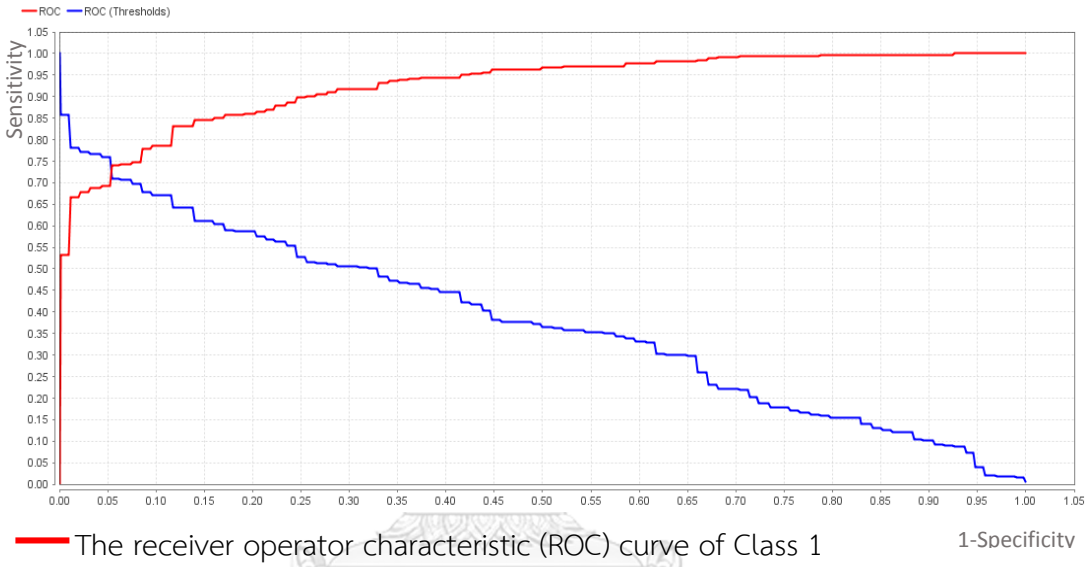
ทำการ Validation ตัวแบบพยากรณ์ด้วยชุดข้อมูลฝึกหัด (Training dataset) พบว่าให้ค่าความถูกต้อง (accuracy) เท่ากับร้อยละ 87.44 ค่าความไว (sensitivity) ร้อยละ 69.15 ค่าความจำเพาะ (specificity) ร้อยละ 92.47 ค่าการทำนายผลบวก (positive predictive value) ร้อยละ 71.63 และค่าการทำนายผลลบ (negative predictive value) ร้อยละ 91.60 (ตารางที่ 15) และประเมินประสิทธิภาพตัวแบบพยากรณ์ด้วยข้อมูลทดสอบ (Testing dataset) พบว่าให้ค่าความถูกต้อง (accuracy) เท่ากับร้อยละ 86.70 ค่าความไว (sensitivity) ร้อยละ 68.09ค่าความจำเพาะ (specificity) ร้อยละ 91.81 ค่าการทำนายผลบวก (positive predictive value) ร้อยละ 69.57 และค่าการทำนายผลลบ (negative predictive value) ร้อยละ 91.28 (ตารางที่ 16) โดยมีค่า AUC เท่ากับ 0.926 (รูปที่ 46)

ตารางที่ 16 Confusion matrix of training dataset using SVM model with enter regression variables

	Hb ไม่ผ่านเกณฑ์	Hb ผ่านเกณฑ์	Class precision
พยากรณ์ Hb ไม่ผ่านเกณฑ์	260	103	71.63%
พยากรณ์ Hb ผ่านเกณฑ์	116	1265	91.60%
Class recall	69.15%	92.47%	

ตารางที่ 17 Confusion matrix of testing dataset using SVM model with enter regression variables

	Hb ไม่ผ่านเกณฑ์	Hb ผ่านเกณฑ์	Class precision
พยากรณ์ Hb ไม่ผ่านเกณฑ์	64	28	69.57%
พยากรณ์ Hb ผ่านเกณฑ์	30	314	91.28%
Class recall	68.09%	91.81%	



รูปที่ 46 Area under the curve (AUC) of testing dataset using SVM model with enter regression variables

4.2.2 ผลการคัดเลือกตัวแปรวิธีเพิ่มตัวแปร (forward selection)

พบว่าคงเหลือตัวแปรตัวเพียง 2 ตัวคือ ค่าฮีโมโกลบินครั้งที่ผ่านมา (Previous Hb) และตัวแปรอาชีพ (Occup) มีค่าน้ำหนักเท่ากับ 3.228 และ 0.003 ตำลำดับ โดยมีค่าไบแอสเท่ากับ 2.934

ตารางที่ 18 ค่าน้ำหนักตัวแปรของตัวแบบ SVM โดยคัดเลือกตัวแปรวิธีเพิ่มตัวแปร (forward selection)

Kernel Model	
Attribute	**Weight**
Bias (offset)	2.934
Previous Hb	3.228328385
Occup	0.003239967

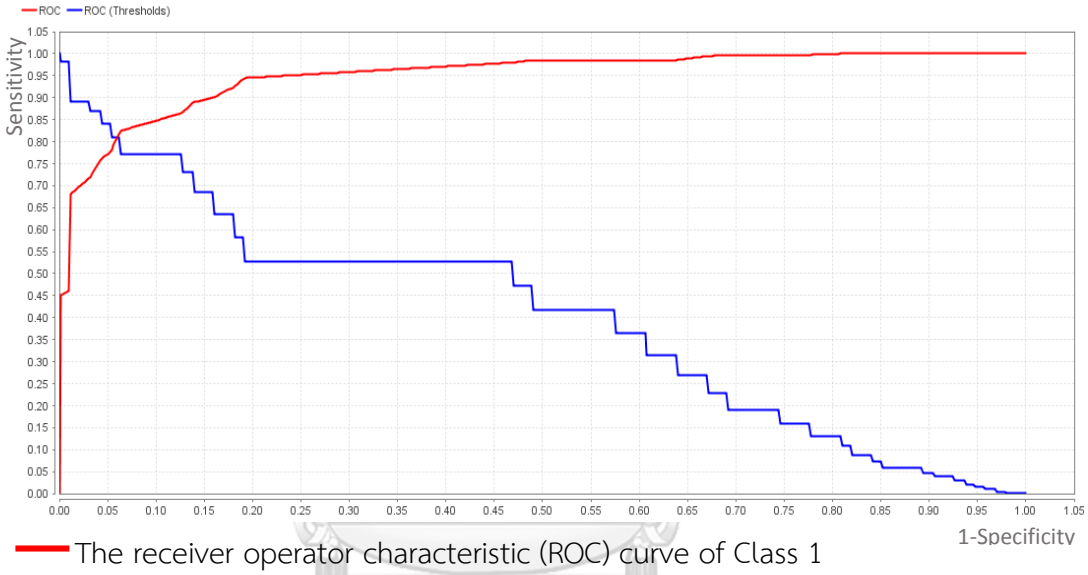
ทำการ Validation ตัวแบบพยากรณ์ด้วยชุดข้อมูลฝึกหัด (Training dataset) พบว่าให้ค่าความถูกต้อง (accuracy) เท่ากับร้อยละ 91.23ค่าความไว (sensitivity) ร้อยละ 76.60 ค่าความจำเพาะ (specificity) ร้อยละ 95.25 ค่าการทำนายผลบวก (positive predictive value) ร้อยละ 81.59 และค่าการทำนายผลลบ (negative predictive value) ร้อยละ 93.67 (ตารางที่ 18) และประเมินประสิทธิภาพตัวแบบพยากรณ์ด้วยข้อมูลทดสอบ (Testing dataset) พบว่าให้ค่าความถูกต้อง (accuracy) เท่ากับร้อยละ 91.51 ค่าความไว (sensitivity) ร้อยละ 80.85 ค่าความจำเพาะ (specificity) ร้อยละ 94.44 ค่าการทำนายผลบวก (positive predictive value) ร้อยละ 80.00 และค่าการทำนายผลลบ (negative predictive value) ร้อยละ 94.72 (ตารางที่ 19) โดยมีค่า AUC เท่ากับ 0.951 (รูปที่ 47)

ตารางที่ 19 Confusion matrix of training dataset using SVM model with forward selection variables

	Hb ไม่ผ่านเกณฑ์	Hb ผ่านเกณฑ์	Class precision
พยากรณ์ Hb ไม่ผ่านเกณฑ์	288	65	81.59%
พยากรณ์ Hb ผ่านเกณฑ์	88	1303	93.67%
Class recall	76.60%	95.25%	

ตารางที่ 20 Confusion matrix of testing dataset using SVM model with forward selection variables

	Hb ไม่ผ่านเกณฑ์	Hb ผ่านเกณฑ์	Class precision
พยากรณ์ Hb ไม่ผ่านเกณฑ์	76	19	80.00%
พยากรณ์ Hb ผ่านเกณฑ์	18	323	94.72%
Class recall	80.85%	94.44%	



รูปที่ 47 Area under the curve (AUC) of testing dataset using SVM model with forward selection variables

4.2.3 ผลการคัดเลือกตัวแปรวิธีลดตัวแปร (backward elimination)

พบว่าคงเหลือตัวแปร 35 ตัวแปร โดยตัวแปรที่มีค่าน้ำหนักสูงสุดคือ ค่าฮีโมโกลบินครั้งที่ผ่านมา (Previous Hb) ซึ่งมีค่าน้ำหนักเท่ากับ 1.956 และมีค่าไบแอสเท่ากับ 1.618

ตารางที่ 21 ค่าน้ำหนักตัวแปรของตัวแบบ SVM โดยคัดเลือกตัวแปรวิธีลดตัวแปร (backward elimination)

Kernel Model	
Attribute	**Weight**
Bias (offset)	1.619
Age	-0.125886921
Menopause	0.025019232
Status	0.119724617
No.Child	0.069689878
Religion	-0.084781126
High	-0.168298771
Weight	0.14426112
BMI	-0.128443502
Weight in past 3 month	0.031150087
Change Weight in past 3 Month	0.111846373
Education	0.101303339
Address1	0.070340359
Occupation	0.154344372
ABO	0.121081609
Donation Place	0.171902529
Single Donor	0.02873739
No. Single Donor	-0.012047244
Donation/year	0.049761939
Donation	0.016344965
Interval donation	0.063104797
Donation success	-0.053999253
Previous Hb	1.955553155
BP Sys	-0.126557436
BP Dias	-0.015807497
Pulse	-0.06637051
Disease	0.036359967

Public to Known	-0.00302628
Sleep Type	-0.081657892
Sleep hour	0.273295905
Exercise	0.131372519
Smoke	0.029953262
Alcohol take	-0.038801974
Food Type	0.027191532
Fe take	-0.072456582
Fe Why not	-0.118307188

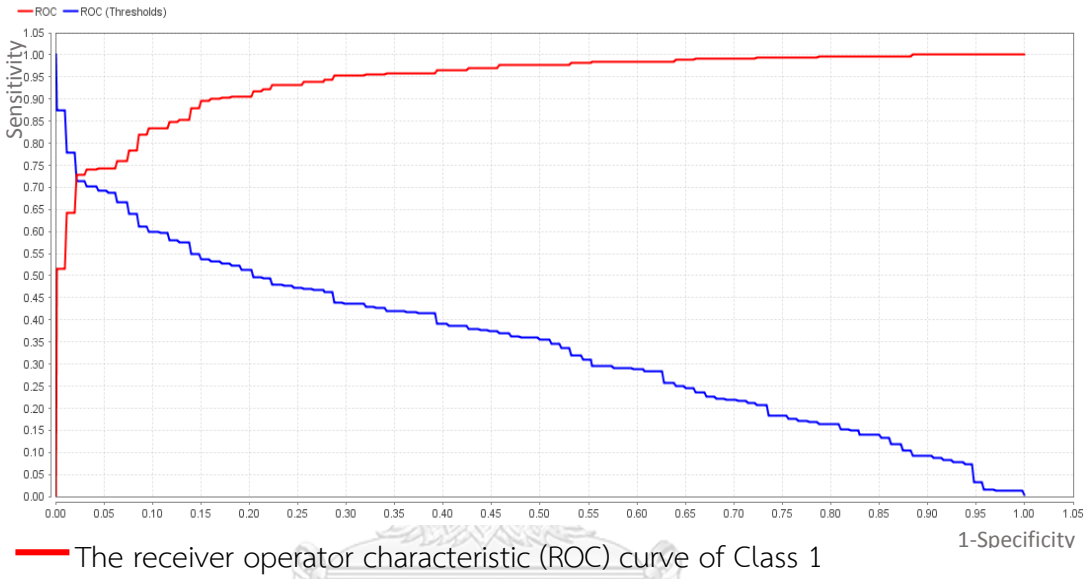
ทำการ Validation ตัวแบบพยากรณ์ด้วยชุดข้อมูลฝึกหัด (Training dataset) พบว่าให้ค่าความถูกต้อง (accuracy) เท่ากับร้อยละ 88.30 ค่าความไว (sensitivity) ร้อยละ 73.40 ค่าความจำเพาะ (specificity) ร้อยละ 92.40 ค่าการทำนายผลบวก (positive predictive value) ร้อยละ 72.63 และค่าการทำนายผลลบ (negative predictive value) ร้อยละ 92.67 (ตารางที่ 21) และประเมินประสิทธิภาพตัวแบบพยากรณ์ด้วยข้อมูลทดสอบ (Testing dataset) พบว่าให้ค่าความถูกต้อง (accuracy) เท่ากับร้อยละ 88.99 ค่าความไว (sensitivity) ร้อยละ 79.79 ค่าความจำเพาะ (specificity) ร้อยละ 91.52 ค่าการทำนายผลบวก (positive predictive value) ร้อยละ 72.12 และค่าการทำนายผลลบ (negative predictive value) ร้อยละ 94.28 (ตารางที่ 22) โดยมีค่า AUC เท่ากับ 0.943 (รูปที่ 48)

ตารางที่ 22 Confusion matrix of training dataset using SVM model with backward elimination variables

	Hb ไม่ผ่านเกณฑ์	Hb ผ่านเกณฑ์	Class precision
พยากรณ์ Hb ไม่ผ่านเกณฑ์	276	104	72.63%
พยากรณ์ Hb ผ่านเกณฑ์	100	1264	92.67%
Class recall	73.40%	92.40%	

ตารางที่ 23 Confusion matrix of testing dataset using SVM model with backward elimination variables

	Hb ไม่ผ่านเกณฑ์	Hb ผ่านเกณฑ์	Class precision
พยากรณ์ Hb ไม่ผ่านเกณฑ์	75	29	72.12%
พยากรณ์ Hb ผ่านเกณฑ์	19	313	94.28%
Class recall	79.79%	91.52%	



The receiver operator characteristic (ROC) curve of Class 1

ROC (Threshold) show confidence cut off

รูปที่ 48 Area under the curve (AUC) of testing dataset using SVM model with backward elimination variables

4.2.4 ผลการคัดเลือกตัวแปรวิธีเพิ่มตัวแปรและลดตัวแปร (optimize selection)

พบว่าคงเหลือตัวแปรตัวเพียง 2 ตัวคือ ค่าฮีโมโกลบินครั้งที่ผ่านมา (Previous Hb) มีค่าน้ำหนักเท่ากับ 3.228 ตัวแปรอาชีพมีค่าน้ำหนักเท่ากับ 0.003 และมีค่าไบแอสเท่ากับ 2.934

ตารางที่ 24 ค่าน้ำหนักตัวแปรของตัวแบบ SVM โดยคัดเลือกตัวแปรด้วยวิธีเพิ่มตัวแปรและลดตัวแปร (optimize selection)

Kernel Model	
Attribute	**Weight**
Bias (offset)	2.934
Previous Hb	3.228328385
Occup	0.003239967

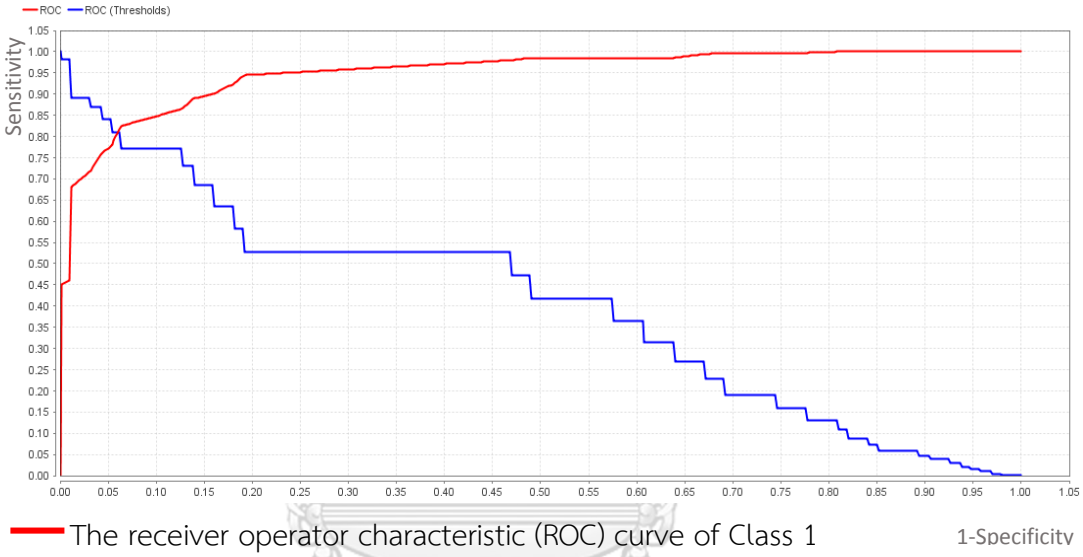
ทำการ Validation ตัวแบบพยากรณ์ด้วยชุดข้อมูลฝึกหัด (Training dataset) พบว่าให้ค่าความถูกต้อง (accuracy) เท่ากับร้อยละ 91.23 ค่าความไว (sensitivity) ร้อยละ 76.60 ค่าความจำเพาะ (specificity) ร้อยละ 95.25 ค่าการทำนายผลบวก (positive predictive value) ร้อยละ 81.59 และค่าการทำนายผลลบ (negative predictive value) ร้อยละ 93.67 (ตารางที่ 24) และประเมินประสิทธิภาพตัวแบบพยากรณ์ด้วยข้อมูลทดสอบ (Testing dataset) พบว่าให้ค่าความถูกต้อง (accuracy) เท่ากับร้อยละ 91.51 ค่าความไว (sensitivity) ร้อยละ 80.85 ค่าความจำเพาะ (specificity) ร้อยละ 94.44 ค่าการทำนายผลบวก (positive predictive value) ร้อยละ 80.00 และค่าการทำนายผลลบ (negative predictive value) ร้อยละ 94.72 (ตารางที่ 25) โดยมีค่า AUC เท่ากับ 0.951 (รูปที่ 49)

ตารางที่ 25 Confusion matrix of training dataset using SVM model with optimize selection variables

	Hb ไม่ผ่านเกณฑ์	Hb ผ่านเกณฑ์	Class precision
พยากรณ์ Hb ไม่ผ่านเกณฑ์	288	65	81.59%
พยากรณ์ Hb ผ่านเกณฑ์	88	1303	93.67%
Class recall	76.60%	95.25%	

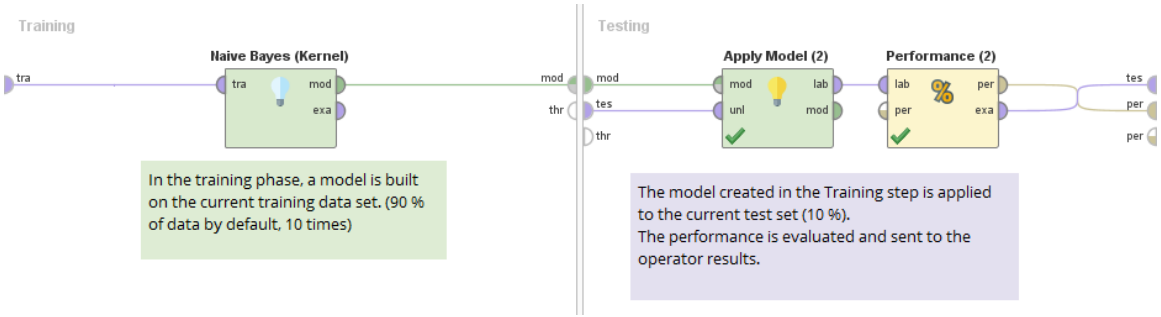
ตารางที่ 26 Confusion matrix of testing dataset using SVM model with optimize selection variables

	Hb ไม่ผ่านเกณฑ์	Hb ผ่านเกณฑ์	Class precision
พยากรณ์ Hb ไม่ผ่านเกณฑ์	76	19	80.00%
พยากรณ์ Hb ผ่านเกณฑ์	18	323	94.72%
Class recall	80.85%	94.44%	



รูปที่ 49 Area under the curve (AUC) of testing dataset using SVM model with optimize selection variables

4.3 ผลการทดลองของการจำแนกแบบเบย์อย่างง่าย (naïve bayesain classifier)



รูปที่ 50 การสร้างตัวแบบการจำแนกแบบเบย์อย่างง่าย (naïve bayesain classifier) จากการฝึกหัด

การสร้างตัวแบบพยากรณ์ด้วยการจำแนกแบบเบย์อย่างง่ายเป็นการคำนวณความน่าจะเป็นในการเกิดเหตุการณ์ โดยตัวแปรที่เป็นชนิดตัวเลขเป็นการคำนวณหาค่าเฉลี่ย (mean) และค่าเบี่ยงเบนมาตรฐาน (standard deviation) ระหว่างกลุ่มที่มีผลตรวจ Hb ผ่านเกณฑ์และไม่ผ่านเกณฑ์ ส่วนตัวแปรเชิงกลุ่มหรือลำดับชั้นจะเป็นการคำนวณความถี่ของประชากรในแต่ละกลุ่มหรือลำดับชั้น

4.3.1 ผลการคัดเลือกตัวแปรวิธีนำตัวแปรเข้าทั้งหมด (enter regression)

ใช้ตัวแปรทั้งหมดในการสร้างตัวแบบ คำนวณหาพารามิเตอร์ของแต่ละตัวแปร (ตารางที่ 26) และแสดงในรูปของการกระจายตัวของประชากรหรือค่าถี่ของแต่ละกลุ่ม (รูปที่ 51-54)

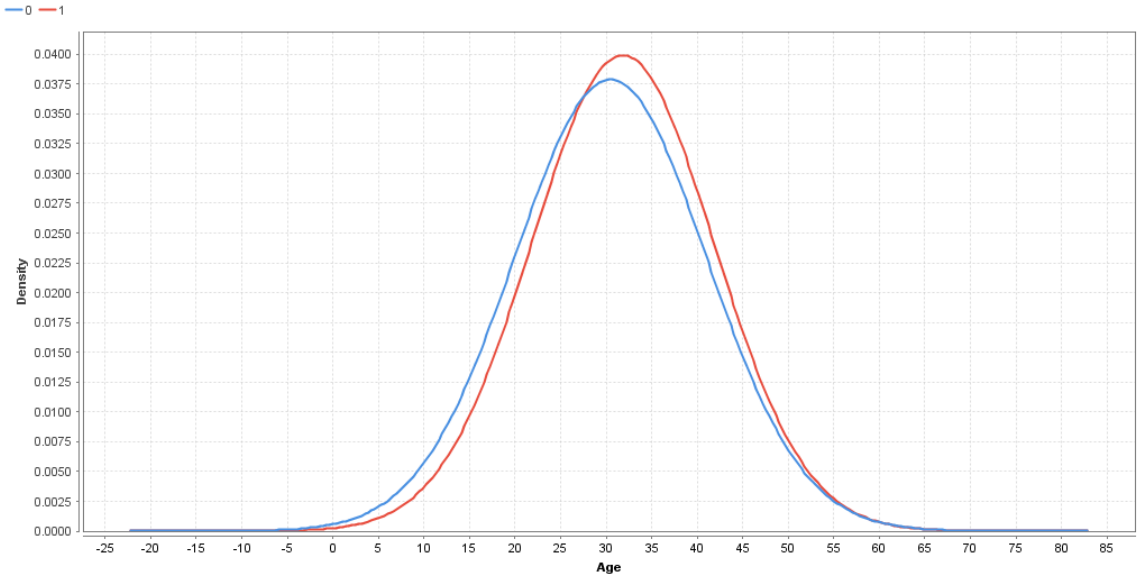
ตารางที่ 27 ค่า Parameter ของตัวแปรต่าง ๆ ที่ได้จากตัวแบบการจำแนกแบบเบย์อย่างง่าย

Attribute	Parameter	0	1
Gender	value=1	0.91489096	0.55043832
	value=0	0.08510752	0.44956126
Age	mean	30.48404255	31.82821637
	standard deviation	10.53551508	9.99636477
Menopause	value=1	0.94946527	0.98391731
	value=0	0.05053321	0.01608227
Status	value=0	0.69680579	0.60964852
	value=1	0.27393602	0.33479518
	value=2	0.02925667	0.05555588
No,Child	mean	0.24734043	0.32529240
	standard deviation	0.65726167	0.74476711
Religion	value=0	0.98137702	0.96856564
	value=2	0.00798019	0.01827523
	value=3	0.00266108	0.00219340
	value=1	0.00798019	0.01096531
High	mean	160.78129047	164.82365878
	standard deviation	6.71826017	8.51909326
Weight	mean	60.46851359	67.01266569
	standard deviation	11.61000509	14.68257182
BMI	mean	23.54449135	24.60204397
	standard deviation	4.11682908	4.69812507
Weight in past 3 month	mean	60.81036150	66.33094867
	standard deviation	9.98327366	13.04239359
Change Weight in past 3 Month	mean	0.28324775	0.40536072
	standard deviation	2.35921110	2.46097635

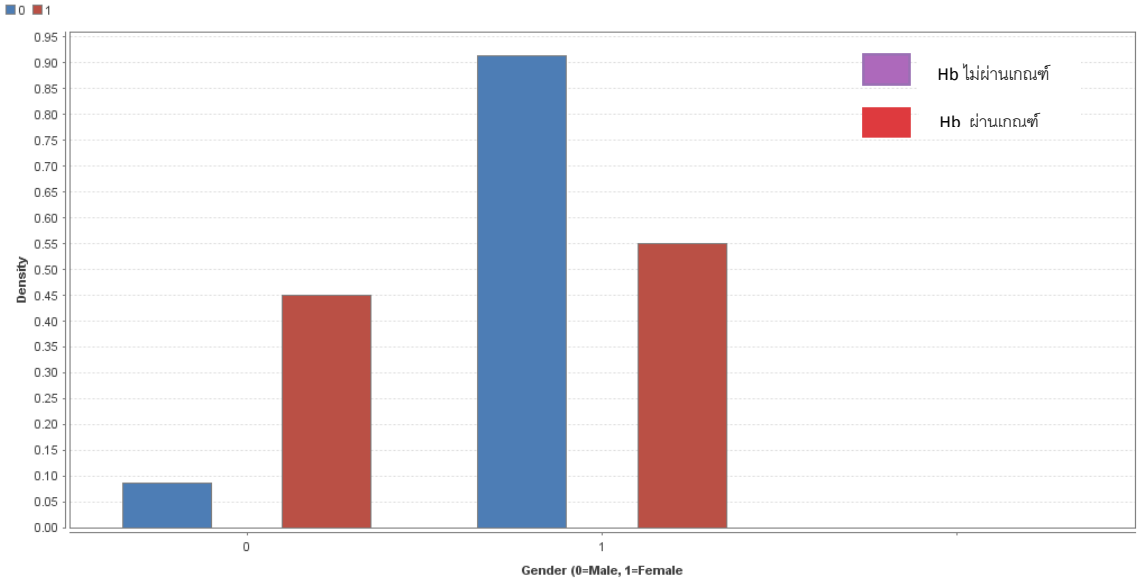
Education	value=2	0.22606260	0.22880082
	value=3	0.13829771	0.12061405
	value=4	0.51063359	0.49122684
	value=1	0.04521374	0.05116984
	value=0	0.01861832	0.01681323
	value=5	0.06117099	0.09064339
	value=6	0.00000152	0.00073141
Address1	value=0	0.60106169	0.59575965
	value=1	0.39893526	0.40204653
	value=2	0.00000152	0.00219340
Occupation	value=0	0.30318885	0.21856685
	value=2	0.33510331	0.32821556
	value=6	0.11702120	0.10964913
	value=5	0.07180905	0.15058465
	value=3	0.15159519	0.17251439
	value=7	0.00000152	0.00219339
	value=4	0.00266106	0.00219339
	value=1	0.01861829	0.01608223
Income	value=1	0.45212574	0.45979478
	value=0	0.31914803	0.22880111
	value=3	0.04255439	0.09502946
	value=2	0.18617032	0.21637423
ABO	value=0	0.22872319	0.20321637
	value=1	0.34840312	0.32163717
	value=2	0.35106268	0.39766040
	value=3	0.07180949	0.07748564
Rh	value=0	0.99999848	0.99999958
Donation Place	value=2	0.37499809	0.34210482
	value=1	0.07712848	0.08040957
	value=0	0.51595425	0.53216282
	value=3	0.03191613	0.04459095
	value=4	0.00000152	0.00073141
Single Donor	value=1	0.97074176	0.96710447
	value=0	0.02925671	0.03289511
No. Single Donor	mean	0.16223404	0.19371345
	standard deviation	1.39819286	2.29908317
Donation/year	mean	1.07669020	1.46617849
	standard deviation	0.97200923	1.05957856
Donation	mean	7.89519606	11.38890210
	standard deviation	8.97205916	13.37233842

Interval donation	mean	7.13795384	7.10569986
	standard deviation	6.53736248	9.13383234
Donation success	value=0	0.98935568	0.99049542
	value=1	0.00532063	0.00950332
	value=3	0.00266108	0.00000042
	value=4	0.00266108	0.00000042
Previous Hb	mean	12.27502660	14.31519737
	standard deviation	0.73085083	1.29064456
BP Sys	mean	121.02672866	124.81625877
	standard deviation	12.41408762	14.27903522
BP Dias	mean	75.83915706	77.46931405
	standard deviation	8.88214588	9.96682406
Pulse	mean	84.16495764	82.06607175
	standard deviation	9.97485809	10.58817905
EverLowHb	value=1	0.62233910	0.22880130
	value=0	0.37765937	0.77119828
Disease	value=0	0.94679704	0.94297932
	value=5	0.02127783	0.02193016
	value=1	0.01063968	0.00365538
	value=6	0.00266106	0.00438637
	value=4	0.00798014	0.00292438
	value=3	0.00266106	0.01754421
	value=2	0.00798014	0.00511736
	value=7	0.00000152	0.00146240
Public to Known	value=6	0.38829407	0.35014530
	value=4	0.04787321	0.15643258
	value=2	0.26329578	0.18859620
	value=1	0.05053275	0.06944460
	value=5	0.06117090	0.03289503
	value=7	0.14893565	0.16081853
	value=0	0.01329921	0.01096529
	value=3	0.02659690	0.03070206
Sleep Type	value=0	0.68084532	0.71564159
	value=1	0.31382796	0.27777738
	value=2	0.00266107	0.00438637
	value=4	0.00000152	0.00146240
	value=3	0.00266107	0.00000042
	value=6	0.00000152	0.00073141

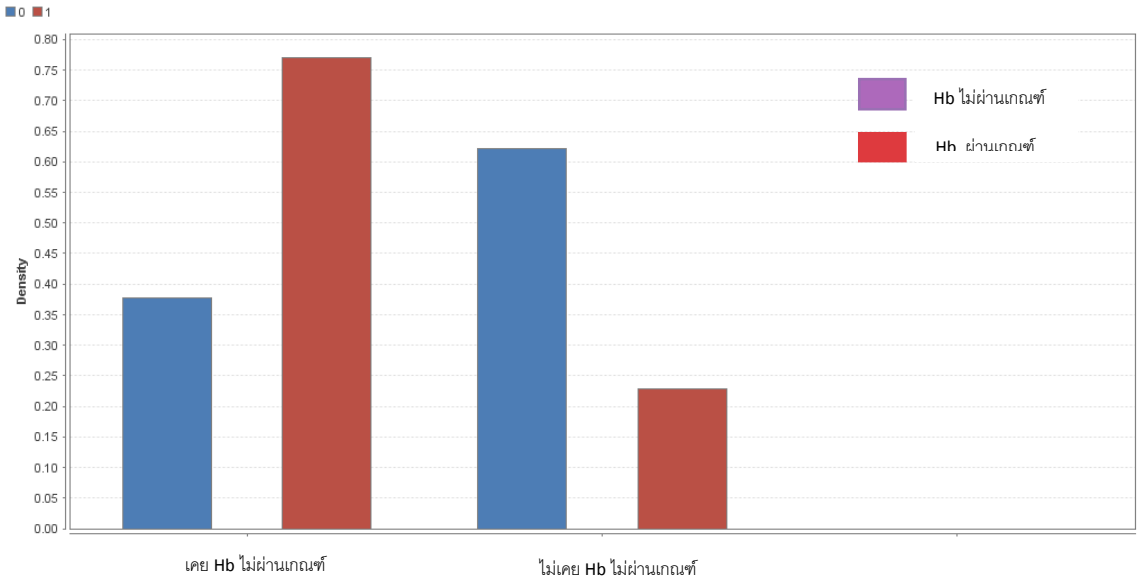
Sleep hour	value=2	0.52925200	0.67982327
	value=1	0.46542280	0.30994116
	value=0	0.00532063	0.00731034
	value=3	0.00000152	0.00219340
	value=6	0.00000152	0.00073141
Exercise	value=0	0.17021247	0.11549716
	value=1	0.56914439	0.45979439
	value=2	0.19148884	0.32090591
	value=3	0.06383063	0.10160831
	value=4	0.00266107	0.00219340
	value=5	0.00266107	0.00000042
Smoke	value=0	0.97074028	0.90496966
	value=1	0.02925667	0.09429851
	value=2	0.00000152	0.00073141
Alcohol take	value=0	0.84041937	0.72368281
	value=1	0.13297903	0.23245597
	value=3	0.00532063	0.00365538
	value=2	0.02127793	0.03801202
	value=4	0.00000152	0.00219340
Food Type	value=0	0.97605642	0.97148920
	value=3	0.01329928	0.00804133
	value=2	0.00000152	0.00657935
	value=1	0.01063973	0.00877233
	value=4	0.00000152	0.00511737
Fe take	value=0	0.38031777	0.34356695
	value=2	0.54786969	0.49342044
	value=1	0.07180949	0.15935681
	value=3	0.00000152	0.00365538
Fe Why not	value=6	0.10106412	0.05628678
	value=1	0.49999543	0.56359502
	value=4	0.10904275	0.10526322
	value=3	0.04521374	0.03435703
	value=5	0.06914962	0.05482480
	value=0	0.09308550	0.16081859



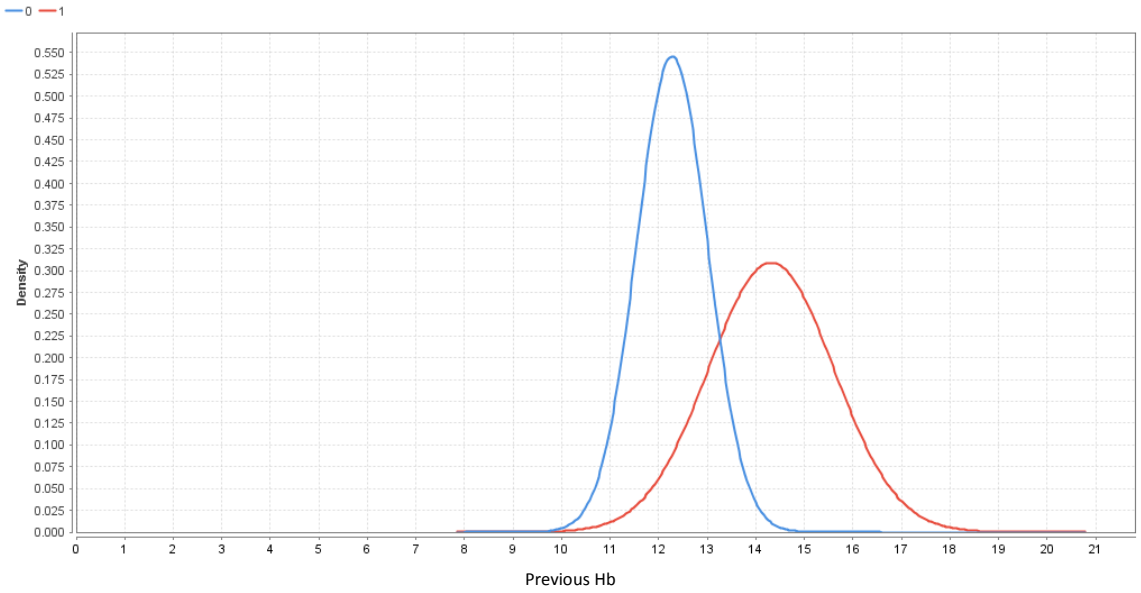
รูปที่ 51 แสดงการแจกแจงตัวแปรอายุ (Age) ระหว่างกลุ่มค่า Hb ผ่านและไม่ผ่านเกณฑ์



รูปที่ 52 แสดงอัตราส่วนของผู้ที่มีค่า Hb ผ่านและไม่ผ่านเกณฑ์ระหว่างเพศชายหญิง (Gender)



รูปที่ 53 แสดงอัตราส่วนของผู้ที่มีค่า Hb ผ่านและไม่ผ่านเกณฑ์ระหว่างผู้ที่เคยมีประวัติค่า Hb ไม่ผ่านเกณฑ์กับผู้ที่ Hb ผ่านเกณฑ์ทุกครั้ง



รูปที่ 54 การแจกแจงค่า Hb ครั้งที่ผ่านมาระหว่างกลุ่มที่มีค่า Hb ครั้งปัจจุบันผ่านเกณฑ์และไม่ผ่านเกณฑ์

ทำการ Validation ตัวแบบพยากรณ์ด้วยชุดข้อมูลฝึกหัด (Training dataset) พบว่าให้ค่าความถูกต้อง (accuracy) เท่ากับร้อยละ 74.77 ค่าความไว (sensitivity) ร้อยละ 79.52 ค่าความจำเพาะ (specificity) ร้อยละ 73.46 ค่าการทำนายผลบวก (positive predictive value)

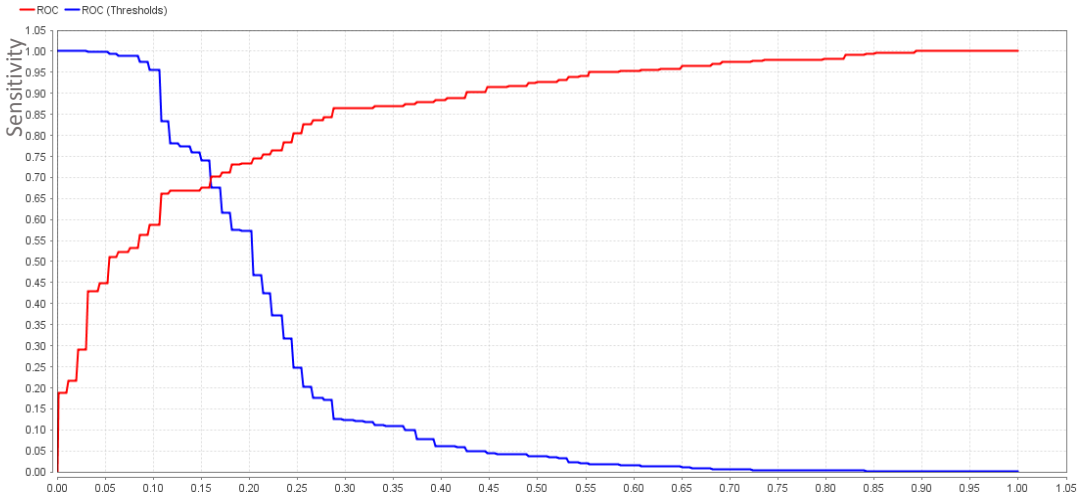
ร้อยละ 45.17 และค่าการทำนายผลลบ (negative predictive value) ร้อยละ 92.88 (ตารางที่ 27) และประเมินประสิทธิภาพตัวแบบพยากรณ์ด้วยข้อมูลทดสอบ (Testing dataset) พบว่าให้ค่าความถูกต้อง (accuracy) เท่ากับร้อยละ 75.46 ค่าความไว (sensitivity) ร้อยละ 79.79 ค่าความจำเพาะ (specificity) ร้อยละ 74.27 ค่าการทำนายผลบวก (positive predictive value) ร้อยละ 46.01 และค่าการทำนายผลลบ (negative predictive value) ร้อยละ 93.04 (ตารางที่ 28) โดยมีค่า AUC เท่ากับ 0.856 (รูปที่ 55)

ตารางที่ 28 Confusion matrix of training dataset using naïve bayesain classifier model with optimize selection variables

	Hb ไม่ผ่านเกณฑ์	Hb ผ่านเกณฑ์	Class precision
พยากรณ์ Hb ไม่ผ่านเกณฑ์	299	363	45.17%
พยากรณ์ Hb ผ่านเกณฑ์	77	1005	92.88%
Class recall	79.52%	73.46%	

ตารางที่ 29 Confusion matrix of testing dataset using naïve bayesain classifier model with optimize selection variables

	Hb ไม่ผ่านเกณฑ์	Hb ผ่านเกณฑ์	Class precision
พยากรณ์ Hb ไม่ผ่านเกณฑ์	75	88	46.01%
พยากรณ์ Hb ผ่านเกณฑ์	19	254	93.04%
Class recall	79.79%	74.27%	



รูปที่ 55 Area under the curve (AUC) of testing dataset using naïve bayesain classifier model with optimize selection variables

4.3.2 ผลการคัดเลือกตัวแปรวิธีเพิ่มตัวแปร (forward selection)

ใช้การคัดเลือกตัวแปรวิธีเพิ่มตัวแปร (forward selection) ในการสร้างตัวแบบ พบว่าคงเหลือตัวแปร 3 ตัวแปรคือ ค่าฮีโมโกลบินครั้งที่ผ่านมา (Previous Hb) การมีประจำเดือน (Mnop) และประวัติการบริจาคเกล็ดเลือดหรือน้ำเลือด (Single Donation)

ตารางที่ 30 ค่า Parameter ของตัวแปรต่าง ๆ ที่ได้จากตัวแบบการจำแนกแบบเบย์โดยใช้วิธีเพิ่มตัวแปร (forward selection)

Attribute	Parameter	0	1
Previous Hb	mean	12.2750266	14.31519737
	standard deviation	0.730850828	1.290644562
Menopause	value=1	0.949465266	0.983917311
	value=0	0.050533209	0.01608227
Single Donor	value=1	0.970741765	0.967104466
	value=0	0.02925671	0.032895115

ทำการ Validation ตัวแบบพยากรณ์ด้วยชุดข้อมูลฝึกหัด (Training dataset) พบว่าให้ค่าความถูกต้อง (accuracy) เท่ากับร้อยละ 90.14 ค่าความไว (sensitivity) ร้อยละ 81.12 ค่า

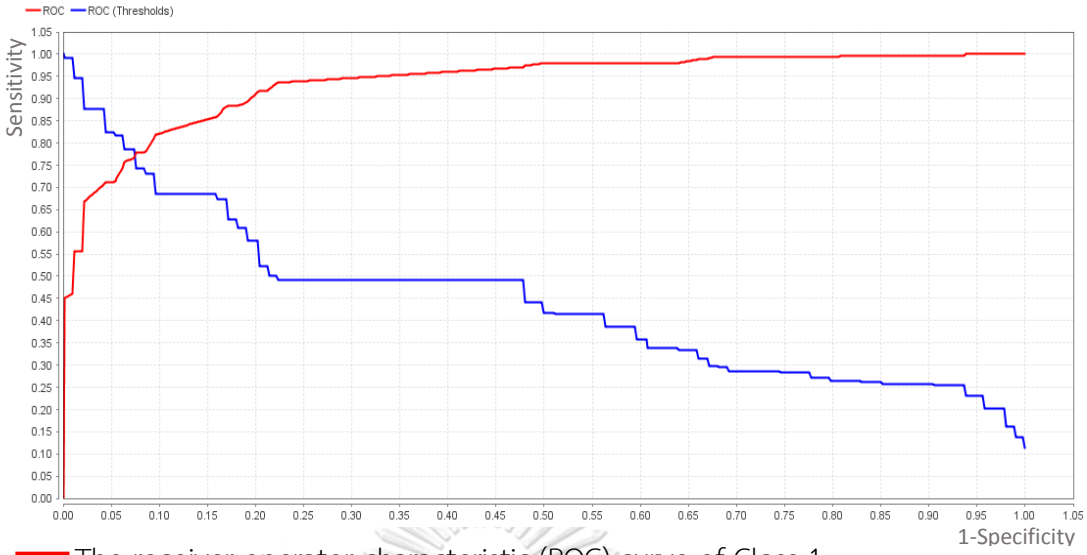
ความจำเพาะ (specificity) ร้อยละ 92.62 ค่าการทำนายผลบวก (positive predictive value) ร้อยละ 75.12 และค่าการทำนายผลลบ (negative predictive value) ร้อยละ 94.69 (ตารางที่ 30) และประเมินประสิทธิภาพตัวแบบพยากรณ์ด้วยข้อมูลทดสอบ (Testing dataset) พบว่าให้ค่าความถูกต้อง (accuracy) เท่ากับร้อยละ 89.22 ค่าความไว (sensitivity) ร้อยละ 78.72 ค่าความจำเพาะ (specificity) ร้อยละ 92.11 ค่าการทำนายผลบวก (positive predictive value) ร้อยละ 73.27 และค่าการทำนายผลลบ (negative predictive value) ร้อยละ 94.03 (ตารางที่ 31) โดยมีค่า AUC เท่ากับ 0.936 (รูปที่ 56)

ตารางที่ 31 Confusion matrix of training dataset using naïve bayesain classifier model with forward selection variables

	Hb ไม่ผ่านเกณฑ์	Hb ผ่านเกณฑ์	Class precision
พยากรณ์ Hb ไม่ผ่านเกณฑ์	305	101	75.12%
พยากรณ์ Hb ผ่านเกณฑ์	71	1267	94.69%
Class recall	81.12%	92.62%	

ตารางที่ 32 Confusion matrix of testing dataset using naïve bayesain classifier model with forward selection variables

	Hb ไม่ผ่านเกณฑ์	Hb ผ่านเกณฑ์	Class precision
พยากรณ์ Hb ไม่ผ่านเกณฑ์	74	27	73.27%
พยากรณ์ Hb ผ่านเกณฑ์	20	315	94.03%
Class recall	78.72%	92.11%	



รูปที่ 56 Area under the curve (AUC) of testing dataset using naïve bayesain classifier model with forward selection variables

4.3.3 ผลการคัดเลือกตัวแปรวิธีลดตัวแปร (backward elimination)

ใช้การคัดเลือกตัวแปรวิธีลดตัวแปร (backward elimination) ในการสร้างตัวแบบ พบตัวแปรคงเหลือทั้งหมด 29 ตัวแปรประกอบด้วย อายุ (Age) การเป็นประจำเดือน (Menop) สภานะ (Status) ศาสนา (Religion) BMI น้ำหนักที่เปลี่ยนแปลใน 3 เดือน (Change Weight 3 Month) การศึกษา (Education) ที่อยู่ (Address1) รายได้ (Income) หมู่โลหิต ABO หมู่โลหิต Rh สถานที่บริจาคโลหิต (Donation Place) ประวัติการบริจาคเกล็ดเลือดหรือน้ำเลือด (Single Donation) ความถี่การบริจาค WB ในรอบปี (Donation/year) ระยะห่างการบริจาค (Interval donation) ผลการบริจาค (เต็มถุงหรือไม่เต็มถุง) (Success Donation) ค่าฮีโมโกลบินครั้งที่ผ่านมา (Previous Hb) ความถี่การบริจาค WB ในรอบปี (Donation/year) ความดันโลหิต Systolic ความดันโลหิต Diastolic อัตราการเต้นของหัวใจ (Pulse) ประวัติการตรวจฮีโมโกลบินไม่ผ่านเกณฑ์ (EverLowHb) ประวัติโรคประจำตัว (Disease) ช่องทางการได้รับข่าวสารการบริจาคโลหิต (Public to know) พฤติกรรมการพักผ่อน (Sleep type) ชั่วโมงการนอน (Sleep hour) การสูบบุรี (Smoke) การรับประทานอาหาร (Food type) และการรับประทานธาตุเหล็ก (Fe take)

ตารางที่ 33 ค่า Parameter ของตัวแปรต่าง ๆ ที่ได้จากตัวแบบการจำแนกแบบเบย์โดยใช้วิธีลดตัวแปร (backward elimination)

Attribute	Parameter	0	1
Age	mean	30.48404255	31.82821637
	standard deviation	10.53551508	9.99636477
Menopause	value=1	0.94946527	0.98391731
	value=0	0.05053321	0.01608227
Status	value=0	0.69680579	0.60964852
	value=1	0.27393602	0.33479518
	value=2	0.02925667	0.05555588
Religion	value=0	0.98137702	0.96856564
	value=2	0.00798019	0.01827523
	value=3	0.00266108	0.00219340
	value=1	0.00798019	0.01096531
BMI	mean	23.54449135	24.60204397
	standard deviation	4.11682908	4.69812507
Change Weight in past 3 Month	mean	0.28324775	0.40536072
	standard deviation	2.35921110	2.46097635
Education	value=2	0.22606260	0.22880082
	value=3	0.13829771	0.12061405
	value=4	0.51063359	0.49122684
	value=1	0.04521374	0.05116984
	value=0	0.01861832	0.01681323
	value=5	0.06117099	0.09064339
	value=6	0.00000152	0.00073141
Address1	value=0	0.60106169	0.59575965
	value=1	0.39893526	0.40204653
	value=2	0.00000152	0.00219340
Income	value=1	0.45212574	0.45979478
	value=0	0.31914803	0.22880111
	value=3	0.04255439	0.09502946
	value=2	0.18617032	0.21637423
ABO	value=0	0.22872319	0.20321637
	value=1	0.34840312	0.32163717
	value=2	0.35106268	0.39766040
	value=3	0.07180949	0.07748564
Rh	value=0	0.99999848	0.99999958
Donation Place	value=2	0.37499809	0.34210482
	value=1	0.07712848	0.08040957

	value=0	0.51595425	0.53216282
	value=3	0.03191613	0.04459095
	value=4	0.00000152	0.00073141
Single Donor	value=1	0.97074176	0.96710447
	value=0	0.02925671	0.03289511
Donation/year	mean	1.07669020	1.46617849
	standard deviation	0.97200923	1.05957856
Interval donation	mean	7.13795384	7.10569986
	standard deviation	6.53736248	9.13383234
Donation success	value=0	0.98935568	0.99049542
	value=1	0.00532063	0.00950332
	value=3	0.00266108	0.00000042
	value=4	0.00266108	0.00000042
Previous Hb	mean	12.27502660	14.31519737
	standard deviation	0.73085083	1.29064456
BP Sys	mean	121.02672866	124.81625877
	standard deviation	12.41408762	14.27903522
BP Dias	mean	75.83915706	77.46931405
	standard deviation	8.88214588	9.96682406
Pulse	mean	84.16495764	82.06607175
	standard deviation	9.97485809	10.58817905
EverLowHb	value=1	0.62233910	0.22880130
	value=0	0.37765937	0.77119828
Disease	value=0	0.94679704	0.94297932
	value=5	0.02127783	0.02193016
	value=1	0.01063968	0.00365538
	value=6	0.00266106	0.00438637
	value=4	0.00798014	0.00292438
	value=3	0.00266106	0.01754421
	value=2	0.00798014	0.00511736
	value=7	0.00000152	0.00146240
Public to Known	value=6	0.38829407	0.35014530
	value=4	0.04787321	0.15643258
Public to Known	value=2	0.26329578	0.18859620
	value=1	0.05053275	0.06944460
	value=5	0.06117090	0.03289503
	value=7	0.14893565	0.16081853
	value=0	0.01329921	0.01096529
	value=3	0.02659690	0.03070206
Sleep Type	value=0	0.68084532	0.71564159

	value=1	0.31382796	0.27777738
	value=2	0.00266107	0.00438637
	value=4	0.00000152	0.00146240
	value=3	0.00266107	0.00000042
	value=6	0.00000152	0.00073141
Sleep hour	value=2	0.52925200	0.67982327
	value=1	0.46542280	0.30994116
	value=0	0.00532063	0.00731034
	value=3	0.00000152	0.00219340
	value=6	0.00000152	0.00073141
Smoke	value=0	0.97074028	0.90496966
	value=1	0.02925667	0.09429851
	value=2	0.00000152	0.00073141
Food Type	value=0	0.97605642	0.97148920
	value=3	0.01329928	0.00804133
	value=2	0.00000152	0.00657935
	value=1	0.01063973	0.00877233
	value=4	0.00000152	0.00511737
Fe take	value=0	0.38031777	0.34356695
	value=2	0.54786969	0.49342044
	value=1	0.07180949	0.15935681
	value=3	0.00000152	0.00365538
Fe Why not	value=6	0.10106412	0.05628678
	value=1	0.49999543	0.56359502
	value=4	0.10904275	0.10526322
	value=3	0.04521374	0.03435703
	value=5	0.06914962	0.05482480
	value=0	0.09308550	0.16081859

ทำการ Validation ตัวแบบพยากรณ์ด้วยชุดข้อมูลฝึกหัด (Training dataset) พบว่าให้ค่าความถูกต้อง (accuracy) เท่ากับร้อยละ 84.58 ค่าความไว (sensitivity) ร้อยละ 73.67 ค่าความจำเพาะ (specificity) ร้อยละ 87.57 ค่าการทำนายผลบวก (positive predictive value) ร้อยละ 61.97 และค่าการทำนายผลลบ (negative predictive value) ร้อยละ 92.37 (ตารางที่ 33) และประเมินประสิทธิภาพตัวแบบพยากรณ์ด้วยข้อมูลทดสอบ (Testing dataset) พบว่าให้ค่าความถูกต้อง (accuracy) เท่ากับร้อยละ 82.34 ค่าความไว (sensitivity) ร้อยละ 68.09 ค่าความจำเพาะ (specificity) ร้อยละ 86.26 ค่าการทำนายผลบวก (positive predictive value)

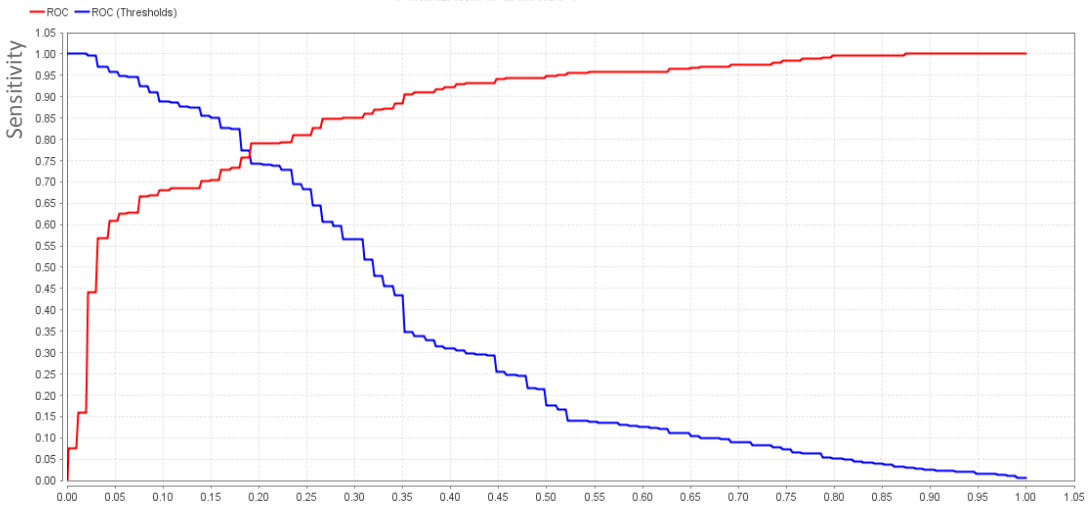
ร้อยละ 57.66 และค่าการทำนายผลลบ (negative predictive value) ร้อยละ 90.77 (ตารางที่ 34) โดยมีค่า AUC เท่ากับ 0.936 (รูปที่ 57)

ตารางที่ 34 Confusion matrix of training dataset using naïve bayesain classifier model with backward elimination variables

	Hb ไม่ผ่านเกณฑ์	Hb ผ่านเกณฑ์	Class precision
พยากรณ์ Hb ไม่ผ่านเกณฑ์	277	170	61.97%
พยากรณ์ Hb ผ่านเกณฑ์	99	1198	92.37%
Class recall	73.67%	87.57%	

ตารางที่ 35 Confusion matrix of testing dataset using naïve bayesain classifier model with backward elimination variables

	Hb ไม่ผ่านเกณฑ์	Hb ผ่านเกณฑ์	Class precision
พยากรณ์ Hb ไม่ผ่านเกณฑ์	64	47	57.66%
พยากรณ์ Hb ผ่านเกณฑ์	30	295	90.77%
Class recall	68.09%	86.26%	



รูปที่ 57 Area under the curve (AUC) of testing dataset using naïve bayesain classifier model with backward elimination variables

4.3.4 ผลการคัดเลือกตัวแปรวิธีเพิ่มตัวแปรและลดตัวแปร (optimize selection)

ใช้การคัดเลือกตัวแปรวิธีเพิ่มตัวแปรและลดตัวแปร (optimize selection)ในการสร้างตัวแบบ พบตัวแปรคงเหลือทั้งหมด 3 ตัวแปร การเป็นประจำเดือน (Menopause) ประวัติการบริจาคเกล็ดเลือดหรือน้ำเลือด (Single Donation) และค่าฮีโมโกลบินครั้งที่ผ่านมา (Previous Hb)

ตารางที่ 36 ค่า Parameter ของตัวแปรต่าง ๆ ที่ได้จากตัวแบบการจำแนกแบบเบย์โดยใช้วิธีเพิ่มตัวแปรและลดตัวแปร (optimize selection)

Attribute	Parameter	0	1
Menopause	value=1	0.94946527	0.98391731
	value=0	0.05053321	0.01608227
Single Donor	value=1	0.97074176	0.96710447
	value=0	0.02925671	0.03289511
Previous Hb	mean	12.27502660	14.31519737
	standard deviation	0.73085083	1.29064456

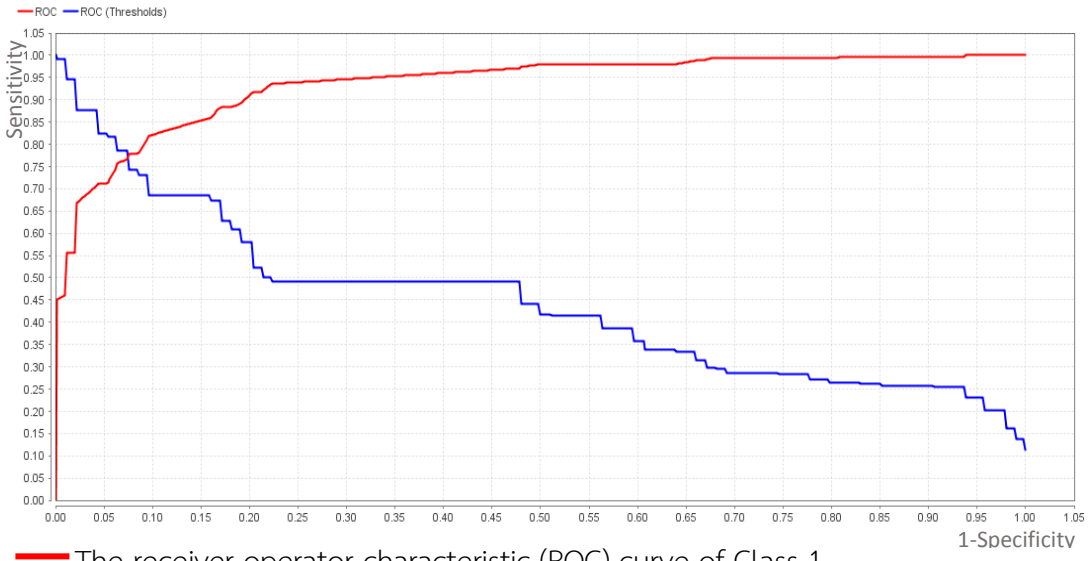
ทำการ Validation ตัวแบบพยากรณ์ด้วยชุดข้อมูลฝึกหัด (Training dataset) พบว่าให้ค่าความถูกต้อง (accuracy) เท่ากับร้อยละ 90.14 ค่าความไว (sensitivity) ร้อยละ 81.12 ค่าความจำเพาะ (specificity) ร้อยละ 92.62 ค่าการทำนายผลบวก (positive predictive value) ร้อยละ 75.12 และค่าการทำนายผลลบ (negative predictive value) ร้อยละ 94.69 (ตารางที่ 36) และประเมินประสิทธิภาพตัวแบบพยากรณ์ด้วยข้อมูลทดสอบ (Testing dataset) พบว่าให้ค่าความถูกต้อง (accuracy) เท่ากับร้อยละ 89.22 ค่าความไว (sensitivity) ร้อยละ 78.72 ค่าความจำเพาะ (specificity) ร้อยละ 92.11 ค่าการทำนายผลบวก (positive predictive value) ร้อยละ 73.27 และค่าการทำนายผลลบ (negative predictive value) ร้อยละ 94.03 (ตารางที่ 37) โดยมีค่า AUC เท่ากับ 0.936 (รูปที่ 58)

ตารางที่ 37 Confusion matrix of training dataset using naïve bayesain classifier model with optimize selection variables

	Hb ไม่ผ่านเกณฑ์	Hb ผ่านเกณฑ์	Class precision
พยากรณ์ Hb ไม่ผ่านเกณฑ์	305	101	75.12%
พยากรณ์ Hb ผ่านเกณฑ์	71	1267	94.69%
Class recall	81.12%	92.62%	

ตารางที่ 38 Confusion matrix of testing dataset using naïve bayesain classifier model with optimize selection variables

	Hb ไม่ผ่านเกณฑ์	Hb ผ่านเกณฑ์	Class precision
พยากรณ์ Hb ไม่ผ่านเกณฑ์	74	27	73.27%
พยากรณ์ Hb ผ่านเกณฑ์	20	315	94.03%
Class recall	78.72%	92.11%	

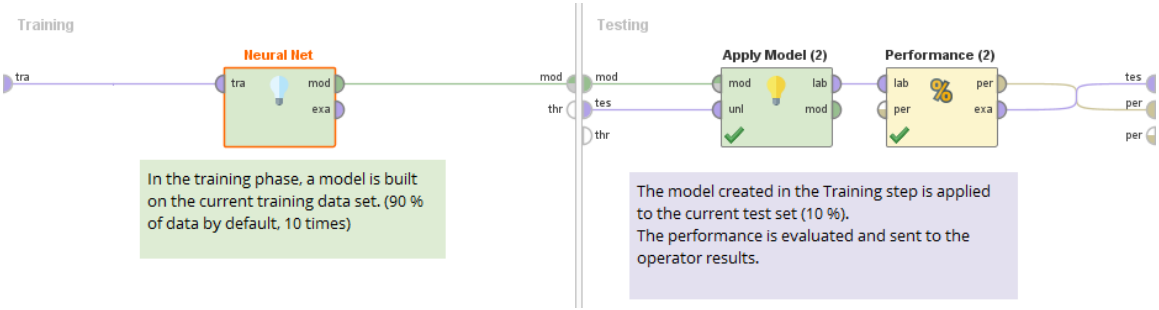


The receiver operator characteristic (ROC) curve of Class 1

ROC (Threshold) show confidence cut off

รูปที่ 58 Area under the curve (AUC) of testing dataset using naïve bayesain classifier model with optimize selection variables

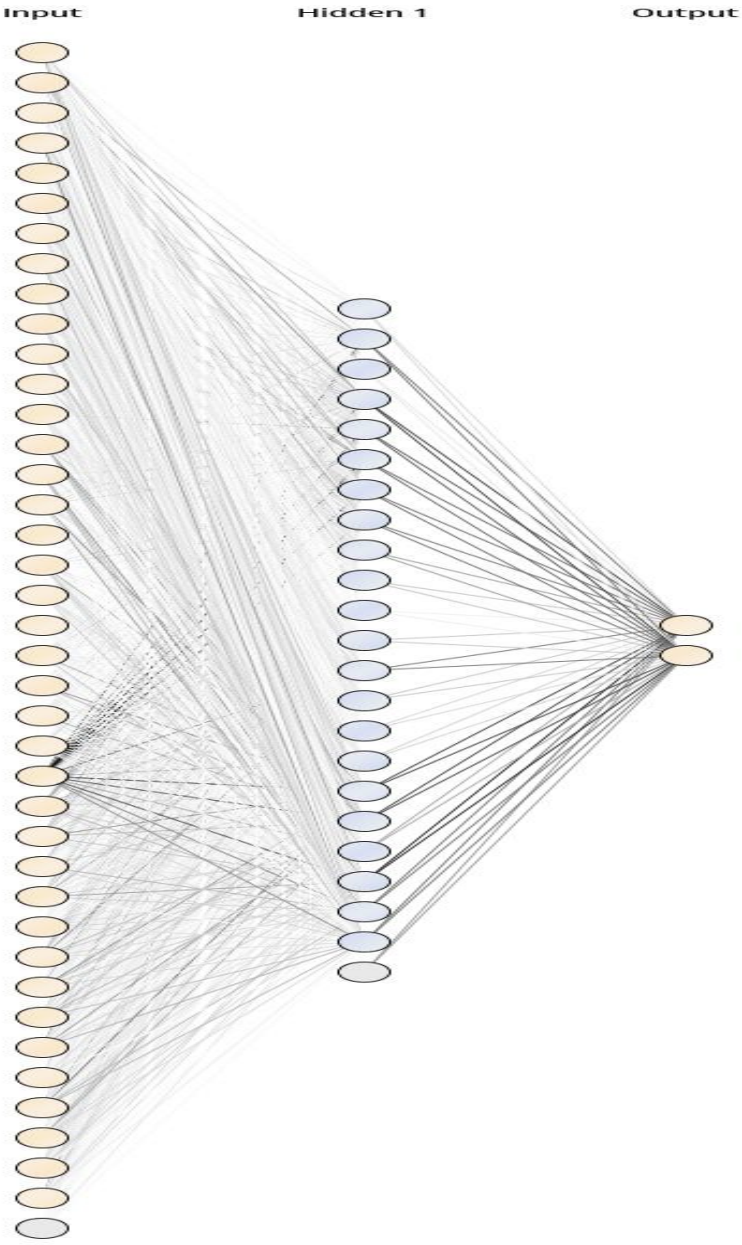
4.4 ผลการทดลองของโครงข่ายประสาทเทียม (artificial neural networks)



รูปที่ 59 การสร้างตัวแบบโครงข่ายประสาทเทียม (artificial neural networks) จากการฝึกหัด

4.4.1 ผลการคัดเลือกตัวแปรวิธีนำตัวแปรเข้าทั้งหมด (enter regression)

ตัวแบบพยากรณ์โครงข่ายประสาทเทียม (artificial neural networks; ANN) โดยวิธีนำตัวแปรเข้าทั้งหมด (enter regression) ได้เป็น ANN ที่มี 3 layer (รูปที่ 60) ประกอบด้วย Input layer จำนวน 40 โหนด โดย 39 โหนดเป็นโหนดนำเข้าและ 1 Threshold (Bias) โหนด และ Hidden layer ประกอบด้วย 22 โหนด ประกอบด้วย 21 โหนดซึ่งแต่ละโหนดมีค่าน้ำหนักของแต่ละตัวแปรพยากรณ์ทั้ง 39 ตัวแปรจากสมการจำแนกแบบ Sigmoid และโหนด Threshold (Bias) 1 โหนด (ตารางที่ 38) ส่วนชั้น Output Layer ประกอบด้วย 2 โหนดคือ กลุ่มที่มีผ่าน Hb ผ่านเกณฑ์และไม่ผ่านเกณฑ์ โดยแต่ละโหนดจะรับค่าน้ำหนักจากแต่ละโหนดของชั้น Hidden 1 Layer (ตารางที่ 39)



รูปที่ 60 โครงข่ายประสาทเทียมที่ได้จากการนำตัวแปรพยากรณ์ทั้งหมดเข้าสร้างตัวแบบ

ตารางที่ 39 ค่าน้ำหนักตัวแปรในชั้น Hidden 1 Layer ของตัวแบบ ANN โดยวิธีนำตัวแปรเข้าทั้งหมด (enter regression)

Input Layer Node	Hidden 1 Layer Node																					
Attribute	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**
Gender	-0.38	-0.03	-0.57	-2.49	1.93	0.17	0.18	-1.20	0.21	-1.20	-0.67	-0.07	-0.73	-0.35	-0.96	0.24	0.20	-2.29	-2.09	-0.95	-0.52	2.92
Age	0.54	-3.54	0.08	-4.65	-0.84	-2.32	-2.79	2.08	0.51	-0.55	0.12	-0.19	1.22	0.30	1.32	-0.72	-2.34	-0.70	-0.72	4.13	0.91	0.00
Menopause	0.10	-0.12	-0.16	-1.14	-1.14	3.01	0.08	-0.08	0.00	-0.19	0.00	-0.21	-1.01	-0.44	-0.09	0.04	0.31	-1.15	-0.30	-1.84	-1.42	0.19
Status	1.34	-0.80	0.19	-0.50	0.71	0.91	-0.56	-3.33	-1.74	1.72	0.17	0.78	-0.28	-0.93	0.35	1.13	-2.38	-1.14	0.76	-0.41	1.82	4.28
No.Child	0.55	2.71	0.38	1.65	-1.36	0.13	0.68	-1.18	-0.60	0.13	0.02	0.18	1.40	0.57	0.81	0.35	1.61	0.07	0.08	-1.94	-0.17	-0.49
Religion	0.23	1.56	0.17	-1.81	-0.63	-0.31	-0.78	0.19	1.12	0.45	0.29	0.00	-1.04	0.75	0.08	-0.40	-1.62	3.02	-0.01	1.46	0.71	-1.68
High	0.40	-1.19	0.05	-0.89	-1.41	-3.41	0.93	-1.60	-0.54	0.68	0.06	0.15	-1.70	-0.52	0.02	0.55	-1.23	3.29	1.16	-0.29	-0.77	1.21
Weight	0.61	0.13	-0.07	1.60	1.43	3.82	0.25	1.10	1.64	0.25	0.02	0.55	-0.74	-1.45	0.66	1.54	-0.17	-1.83	1.09	0.65	-0.61	2.87
BMI	-0.27	-1.16	-0.37	3.09	-0.01	-0.36	-1.56	-0.12	-0.44	-0.94	-0.21	0.79	-0.50	-1.06	-0.22	-0.57	-1.27	0.78	-0.89	1.39	-1.87	-0.99
Weight in past 3 m	0.87	1.18	0.30	2.87	-1.13	0.98	1.61	-1.75	0.74	0.02	0.20	0.57	-0.20	-0.87	1.12	1.50	-0.69	-0.40	0.53	-2.26	-1.38	1.86
Change Weight in past 3 Month	0.04	-0.45	-0.15	1.19	2.62	-0.57	-1.90	-0.35	1.37	0.13	-0.05	-0.12	-0.28	-0.30	0.18	-0.11	1.88	-2.90	0.78	0.61	1.95	-1.22
Education	0.21	-1.85	0.35	-2.65	-0.73	-0.98	2.07	-2.58	1.24	0.58	0.15	-0.77	2.25	0.82	0.00	-0.68	0.05	0.53	1.01	0.78	-0.84	0.25
Address1	-0.28	-0.35	0.14	1.75	5.12	0.62	-1.58	-0.41	1.30	-0.40	-0.06	-0.52	-0.36	0.70	0.53	1.34	1.05	1.26	-1.21	-0.59	0.20	-1.50
Occupation	-1.15	1.29	-0.65	2.91	1.82	0.82	-0.46	1.59	-2.44	-1.99	-0.71	-0.46	2.08	-0.84	-0.24	-0.26	-4.03	-4.30	-2.64	-2.60	0.34	0.09
Income	0.56	-0.22	0.21	-1.09	1.42	2.97	-0.15	-3.02	0.34	-0.27	0.30	-0.81	1.17	-0.21	0.39	-0.12	-0.36	1.02	-0.45	-4.68	-2.80	-2.84
ABO	-0.52	2.94	-0.57	-3.70	-3.26	-4.75	0.03	-0.33	-1.17	-1.14	-0.57	0.75	1.78	-1.31	0.16	-0.03	0.88	-5.12	-0.53	-0.80	-4.05	-0.18
Rh	-0.05	0.02	0.05	-0.04	-0.02	0.05	-0.03	0.02	-0.03	-0.01	-0.01	-0.02	-0.03	0.03	0.03	-0.03	0.03	-0.05	0.02	-0.01	-0.05	-0.04
Donation Place	-1.07	0.54	-0.46	-2.35	1.88	0.44	-3.15	3.36	-1.38	-1.06	-0.09	-0.22	1.78	-1.24	-0.28	-0.19	3.63	-0.77	-0.81	-1.88	-2.29	-1.58
Single Donor	-0.19	0.45	-0.18	0.40	-0.98	1.51	-0.64	1.24	0.18	-0.15	-0.19	-0.10	-0.75	-0.43	-0.47	0.25	-1.23	-1.69	-1.00	-0.58	-2.18	-0.19
No.Single Donor	-0.12	0.05	-0.08	0.46	0.01	-1.26	0.79	-0.27	-0.33	-0.08	0.05	0.03	0.72	-0.08	-0.42	-0.43	-0.02	-0.14	-0.10	0.73	-0.15	-0.34
Donation/year	0.47	-0.55	0.52	-1.83	-0.65	2.54	0.87	1.08	1.38	1.32	0.39	0.16	-1.36	1.32	0.55	0.42	3.73	-0.30	0.14	-2.45	0.56	-0.12
Donation	-0.06	-1.99	-0.03	-0.62	0.14	-1.08	-0.81	-1.55	0.27	0.07	0.09	0.25	-1.23	-0.40	-0.50	-0.71	5.12	-3.62	0.13	-0.49	-1.06	0.49

Input Layer Node	Hidden 1 Layer Node																					
Attribute	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
Interval donation	0.09	-1.68	-0.11	-1.63	0.44	-1.26	-0.52	-1.01	-0.30	0.06	0.05	-0.14	1.88	-0.14	-0.35	0.25	-0.39	-4.03	0.27	-1.47	0.34	1.25
Donation success	-0.25	0.45	-0.09	1.16	-0.34	-2.32	0.60	0.06	-0.62	-0.17	-0.02	0.10	0.74	-0.20	-0.60	-0.47	0.35	0.33	-0.07	1.02	-0.59	-0.60
Previous Hb	1.33	10.20	1.42	-7.16	9.24	12.60	12.36	8.92	4.65	2.17	0.92	-1.05	6.65	2.17	2.18	3.96	10.10	-9.02	4.30	-8.70	5.89	5.45
BP Sys	0.24	2.11	-0.18	4.72	-4.08	3.97	1.79	0.38	-1.97	0.07	0.00	-0.18	-3.00	-1.49	-0.71	-0.99	2.26	-0.20	-1.27	-2.48	-3.42	-1.98
BP Dias	-0.03	5.04	-0.16	2.87	1.09	3.50	-1.55	0.85	-1.77	-0.88	-0.09	-0.32	1.98	-0.99	0.20	-0.56	5.46	3.63	-0.88	2.95	-1.20	-1.54
Pulse	0.39	-0.22	0.52	1.98	-2.76	2.12	-1.39	-0.36	0.84	0.88	0.36	0.20	0.41	0.24	0.09	0.36	2.42	-0.38	-0.80	1.17	3.74	2.06
EverLowHb	0.34	-1.65	0.41	-1.65	-2.86	-2.56	-0.21	-1.34	3.23	0.09	0.32	-1.38	-3.41	-0.32	-1.22	0.00	1.15	-0.51	-2.36	-4.93	-2.40	-0.05
Disease	0.10	2.57	-0.03	0.40	0.74	0.62	-0.02	-0.79	-0.64	-0.26	-0.05	0.24	0.84	-0.10	-0.28	0.81	-0.32	1.42	-1.12	-1.73	2.09	1.68
Public to Known	-0.04	1.55	0.40	1.64	-3.03	3.41	1.86	2.67	-1.42	1.03	0.24	0.65	4.63	1.37	0.24	-0.94	-0.36	3.19	2.38	-0.66	-1.95	-1.66
Sleep Type	-0.30	-1.52	0.01	-0.44	0.36	1.83	0.25	4.46	-0.91	-1.07	-0.06	0.32	0.75	0.68	0.37	-0.05	-4.78	3.02	-0.02	-2.42	-0.69	-2.09
Sleep hour	0.66	3.00	0.89	-2.07	1.56	-1.77	0.52	-0.43	0.79	2.61	0.61	0.15	-0.38	2.16	0.08	0.51	2.59	0.77	3.34	-1.94	0.22	3.23
Exercise	0.07	1.30	0.16	1.78	-5.23	2.94	-1.59	-0.40	3.13	0.99	0.03	-1.04	-2.12	0.42	-0.36	0.83	-0.25	-4.46	-0.15	2.17	1.22	3.87
Smoke	-0.23	1.52	0.11	1.43	-0.08	3.54	-0.85	1.31	0.38	-0.22	-0.04	0.33	-0.79	0.32	-0.51	-0.06	1.64	-0.77	-0.35	-1.15	1.03	0.02
Alcohol take	0.91	-0.05	0.59	-1.68	-0.96	-2.72	-2.04	-0.15	-0.39	0.74	0.36	-0.38	2.74	0.67	1.20	0.17	-5.95	-1.02	0.40	1.71	3.20	3.25
Food Type	0.13	-1.82	0.27	-1.65	0.72	0.68	1.79	-0.40	0.81	0.46	0.31	-0.19	-0.01	0.85	0.55	-0.77	3.05	3.85	0.93	0.83	-1.62	-1.72
Fe take	0.18	-3.48	0.34	-0.13	-1.49	2.17	0.94	-2.92	0.55	1.59	0.47	-1.32	-1.33	0.87	-0.17	0.72	0.45	-0.76	1.45	1.99	-0.95	-0.03
Fe Why not	-0.39	-1.20	0.04	-1.06	-2.41	-4.95	-1.38	2.35	-0.93	-0.56	-0.21	1.60	2.51	1.21	0.68	-0.46	0.35	-0.47	-1.60	-1.45	-0.34	0.85
Bias	0.16	-0.07	0.10	-0.38	-0.03	1.07	-0.77	0.29	0.33	0.09	0.01	-0.01	-0.52	0.04	0.51	0.38	-0.07	-0.09	0.07	-1.03	0.05	0.35

ตารางที่ 40 แสดงค่าน้ำหนักของ Hidden 1 Layer โหนดในชั้น Output Layer

Hidden Layer 1 Node	Output Layer Node	
	Class '0' (Sigmoid)	Class '1' (Sigmoid)
Node 1	-1.698	1.684
Node 2	-6.346	6.347
Node 3	-1.016	1.024
Node 4	9.08	-9.081
Node 5	-7.765	7.767
Node 6	-7.287	7.285
Node 7	-7.114	7.115
Node 8	-6.985	6.984
Node 9	-5.034	5.029
Node 10	-3.129	3.115
Node 11	-0.8	0.857
Node 12	3.008	-3.006
Node 13	-7.186	7.189
Node 14	-2.996	2.995
Node 15	-2.104	2.089
Node 16	-2.251	2.254
Node 17	-9.094	9.094
Node 18	8.854	-8.855
Node 19	-4.116	4.118
Node 20	10.526	-10.525
Node 21	-6.36	6.358
Node 22	-6.423	6.428
Threshold	4.74	-4.744

ทำการ Validation ตัวแบบพยากรณ์ด้วยชุดข้อมูลฝึกหัด (Training dataset) พบว่าให้ค่าความถูกต้อง (accuracy) เท่ากับร้อยละ 88.70 ค่าความไว (sensitivity) ร้อยละ 69.41 ค่าความจำเพาะ (specificity) ร้อยละ 94.01 ค่าการทำนายผลบวก (positive predictive value) ร้อยละ 76.09 และค่าการทำนายผลลบ (negative predictive value) ร้อยละ 91.79 (ตารางที่ 40) และประเมินประสิทธิภาพตัวแบบพยากรณ์ด้วยข้อมูลทดสอบ (Testing dataset) พบว่าให้ค่าความถูกต้อง (accuracy) เท่ากับร้อยละ 87.84 ค่าความไว (sensitivity) ร้อยละ 74.47 ค่า

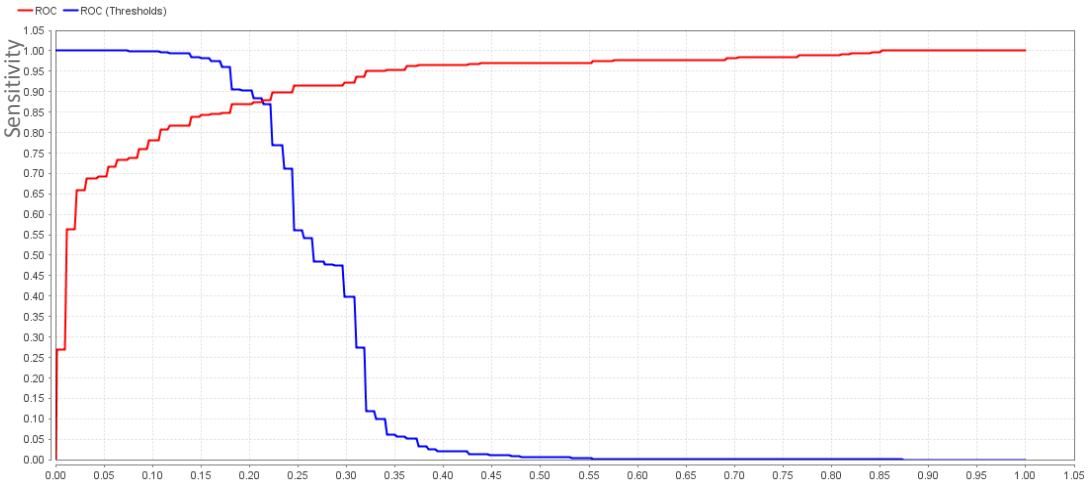
ความจำเพาะ (specificity) ร้อยละ 91.52 ค่าการทำนายผลบวก (positive predictive value) ร้อยละ 70.71 และค่าการทำนายผลลบ (negative predictive value) ร้อยละ 92.88 (ตารางที่ 41) โดยมีค่า AUC เท่ากับ 0.924 (รูปที่ 61)

ตารางที่ 41 Confusion matrix of training dataset using ANN model with enter regression variables

	Hb ไม่ผ่านเกณฑ์	Hb ผ่านเกณฑ์	Class precision
พยากรณ์ Hb ไม่ผ่านเกณฑ์	261	82	76.09%
พยากรณ์ Hb ผ่านเกณฑ์	115	1286	91.79%
Class recall	69.41%	94.01%	

ตารางที่ 42 Confusion matrix of testing dataset using ANN model with enter regression variables

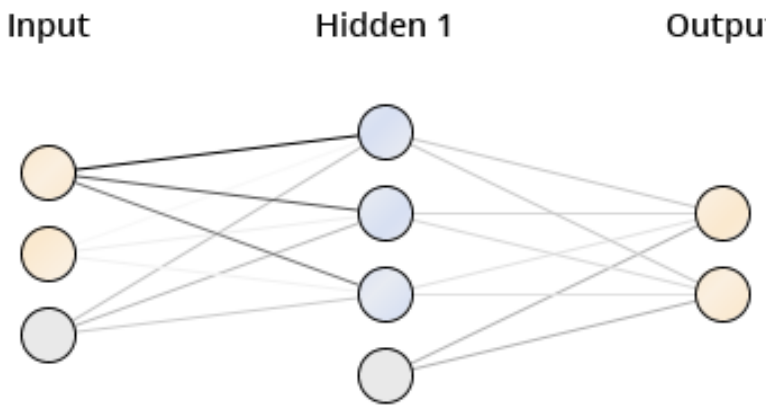
	Hb ไม่ผ่านเกณฑ์	Hb ผ่านเกณฑ์	Class precision
พยากรณ์ Hb ไม่ผ่านเกณฑ์	70	29	70.71%
พยากรณ์ Hb ผ่านเกณฑ์	24	313	92.88%
Class recall	74.47%	91.52%	



รูปที่ 61 Area under the curve (AUC) of testing dataset using ANN model with enter regression variables

4.4.2 ผลการคัดเลือกตัวแปรวิธีเพิ่มตัวแปร (forward selection)

ตัวแบบพยากรณ์โครงข่ายประสาทเทียม (artificial neural networks; ANN) ด้วยวิธีเพิ่มตัวแปร (forward selection) ได้เป็น ANN ที่มี 3 layer (รูปที่ 62) ประกอบด้วย Input layer จำนวน 3 โหนด โดย 2 โหนดเป็นโหนดนำเข้าตัวแปรค่าฮีโมโกลบินครั้งที่ผ่านมา (Previous Hb) และตัวแปรสถานะภาพ (Status) และโหนด 1 Threshold (Bias) ส่วน Hidden layer ประกอบด้วย 4 โหนด โดย 3 โหนดจะเป็นโหนดที่มีค่าน้ำหนักของแต่ละตัวแปรพยากรณ์ทั้ง 2 ตัวแปรจากสมการจำแนกแบบ Sigmoid และโหนด Threshold (Bias) 1 โหนด (ตารางที่ 42) ส่วนชั้น Output Layer ประกอบด้วย 2 โหนดคือ กลุ่มที่มีผ่าน Hb ผ่านเกณฑ์และไม่ผ่านเกณฑ์ โดยแต่ละโหนดจะรับค่าน้ำหนักจากแต่ละโหนดของชั้น Hidden 1 Layer (ตารางที่ 43)



รูปที่ 62 โครงข่ายประสาทเทียมที่ได้จากการคัดเลือกตัวแปรวิธีเพิ่มตัวแปร (forward selection)

ตารางที่ 43 ค่าน้ำหนักตัวแปร Hidden 1 Layer ของตัวแบบ ANN โดยคัดเลือกตัวแปรวิธีเพิ่มตัวแปร (forward selection)

Input Layer Node	Hidden 1 Layer Node		
	1	2	3
Previous Hb	21.308	14.02	11.423
Status	-0.616	1.263	0.99
Bias	4.982	5.656	4.162

ตารางที่ 44 แสดงค่าน้ำหนักของ Hidden 1 Layer โหนดในชั้น Output Layer ของตัวแบบ ANN ที่คัดเลือกตัวแปรด้วยวิธีเพิ่มตัวแปร (forward selection)

Hidden 1 Layer	**Output Node**	
	Class '0' (Sigmoid)	**Class '1' (Sigmoid)**
Node 1	-4.401	4.401
Node 2	-3.155	3.16
Node 3	-2.433	2.428
Threshold	6.467	-6.468

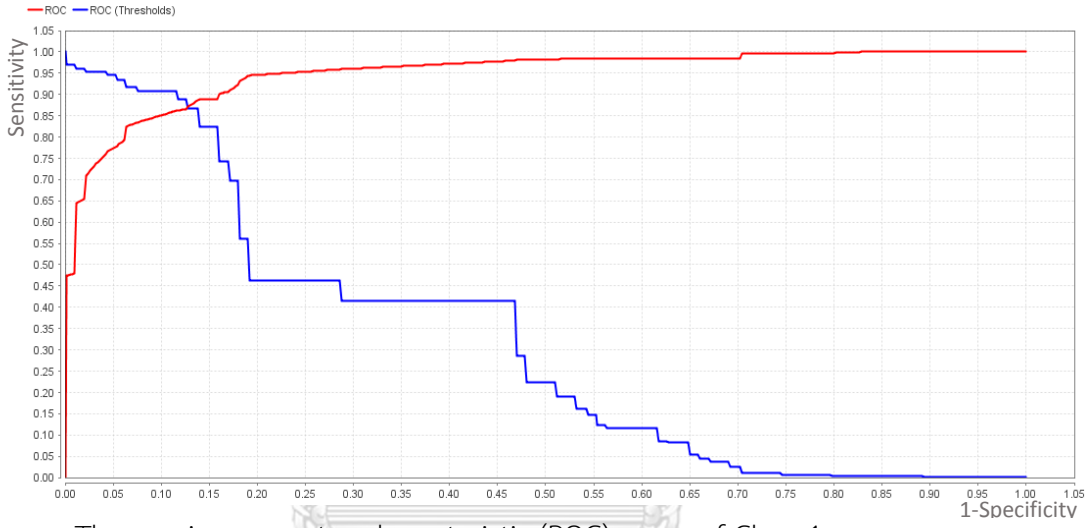
ทำการ Validation ตัวแบบพยากรณ์ด้วยชุดข้อมูลฝึกหัด (Training dataset) พบว่าให้ค่าความถูกต้อง (accuracy) เท่ากับร้อยละ 90.94 ค่าความไว (sensitivity) ร้อยละ 76.06 ค่าความจำเพาะ (specificity) ร้อยละ 95.03 ค่าการทำนายผลบวก (positive predictive value) ร้อยละ 80.79 และค่าการทำนายผลลบ (negative predictive value) ร้อยละ 93.53 (ตารางที่ 44) และประเมินประสิทธิภาพตัวแบบพยากรณ์ด้วยข้อมูลทดสอบ (Testing dataset) พบว่าให้ค่าความถูกต้อง (accuracy) เท่ากับร้อยละ 90.60 ค่าความไว (sensitivity) ร้อยละ 81.91 ค่าความจำเพาะ (specificity) ร้อยละ 92.98 ค่าการทำนายผลบวก (positive predictive value) ร้อยละ 76.24 และค่าการทำนายผลลบ (negative predictive value) ร้อยละ 94.93 (ตารางที่ 45) โดยมีค่า AUC เท่ากับ 0.950 (รูปที่ 63)

ตารางที่ 45 Confusion matrix of training dataset using ANN model with forward selection variables

	Hb ไม่ผ่านเกณฑ์	Hb ผ่านเกณฑ์	Class precision
พยากรณ์ Hb ไม่ผ่านเกณฑ์	286	68	80.79%
พยากรณ์ Hb ผ่านเกณฑ์	90	1300	93.53%
Class recall	76.06%	95.03%	

ตารางที่ 46 Confusion matrix of testing dataset using ANN model with forward selection variables

	Hb ไม่ผ่านเกณฑ์	Hb ผ่านเกณฑ์	Class precision
พยากรณ์ Hb ไม่ผ่านเกณฑ์	77	24	76.24%
พยากรณ์ Hb ผ่านเกณฑ์	17	318	94.93%
Class recall	81.91%	92.98%	

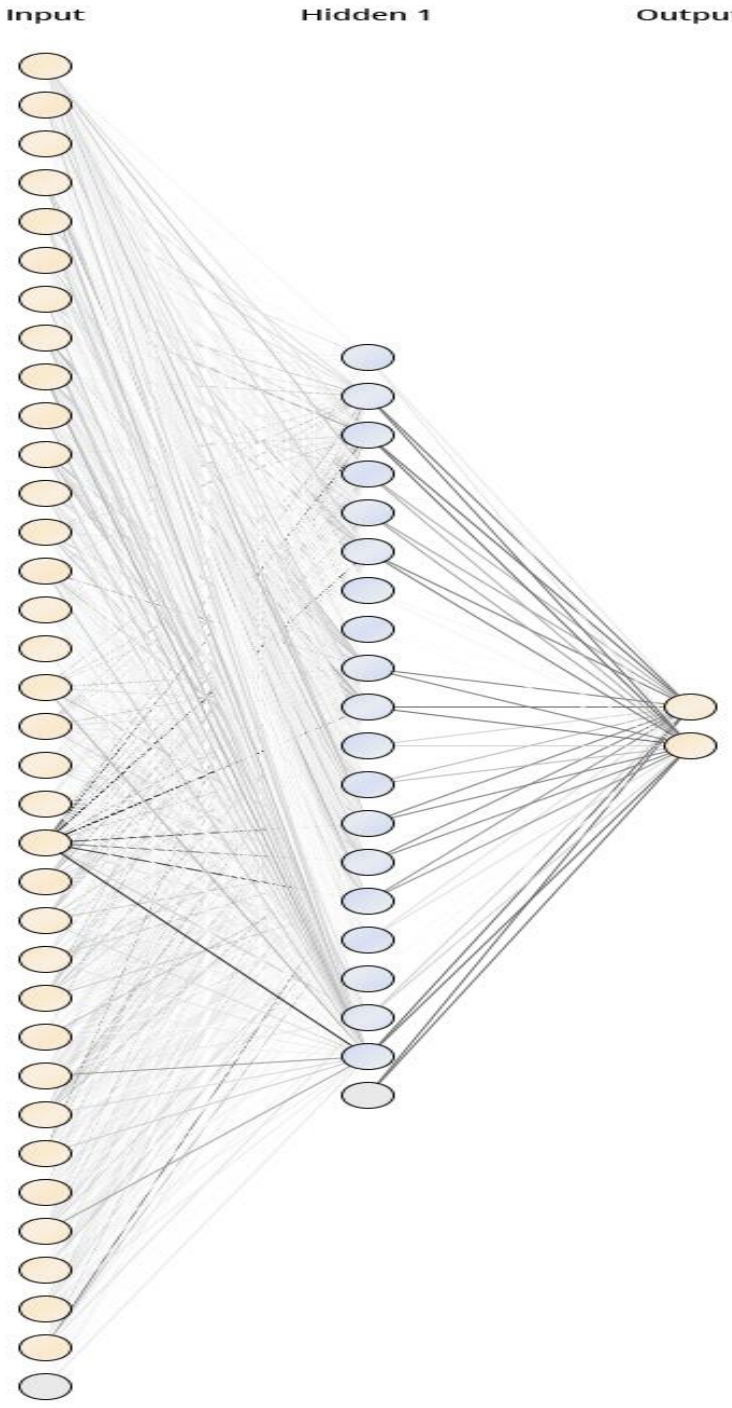


รูปที่ 63 Area under the curve (AUC) of testing dataset using ANN model with forward selection variables

4.4.3 ผลการคัดเลือกตัวแปรวิธีลดตัวแปร (backward elimination)

ตัวแบบพยากรณ์โครงข่ายประสาทเทียม (artificial neural networks; ANN) โดยวิธีลดตัวแปร (backward elimination) ได้เป็น ANN ที่มี 3 layer (รูปที่ 64) ประกอบด้วย Input layer จำนวน 35 โหนด โดย 34 โหนดเป็นโหนดนำเข้าและ 1 Threshold (Bias) โหนด และ Hidden layer ประกอบด้วย 20 โหนด ประกอบด้วย 19 โหนดซึ่งแต่ละโหนดมีค่าน้ำหนักของแต่ละตัวแปรพยากรณ์ทั้ง 39 ตัวแปรจากสมการจำแนกแบบ Sigmoid และโหนด Threshold (Bias) 1 โหนด (ตารางที่ 46) ส่วนชั้น Output Layer ประกอบด้วย 2 โหนดคือ กลุ่มที่มีผ่าน Hb ผ่าน

เกณฑ์และไม่ผ่านเกณฑ์ โดยแต่ละโหนดจะรับค่าน้ำหนักจากแต่ละโหนดของชั้น Hidden 1 Layer (ตารางที่ 47)



รูปที่ 64 โครงข่ายประสาทเทียมที่ได้จากการคัดเลือกตัวแปรวิธีลดตัวแปร (backward elimination)

ตารางที่ 47 ค่าน้ำหนักตัวแปร Hidden 1 Layer ของตัวแบบ ANN โดยคัดเลือกตัวแปรวิธีลดตัวแปร (backward elimination)

Input Layer Node	Hidden 1 Layer Node																		
Attribute	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Gender	-0.71	0.22	-0.20	-3.70	-0.37	2.93	-0.58	-0.57	-1.41	2.69	-1.62	-0.94	3.48	0.94	0.81	-1.12	-0.57	-0.29	-2.40
Age	-0.10	0.32	-1.38	0.19	-1.18	1.03	-0.08	0.03	0.59	1.83	0.28	0.61	-4.18	-0.67	-0.92	0.47	0.03	0.06	2.98
Menopause	0.10	2.09	0.93	0.52	0.32	-0.85	-0.19	-0.28	-0.04	0.80	-0.37	-0.18	-0.43	1.41	0.77	0.17	-0.27	0.25	-0.17
Status	0.35	2.62	-0.28	0.80	1.08	-1.10	-0.15	0.01	-0.03	1.44	0.80	1.73	3.37	3.27	-4.75	-0.09	-0.05	0.79	-4.25
No.Child	1.62	-1.91	1.02	0.79	1.47	-3.61	0.54	0.45	-0.01	-0.41	0.94	-0.88	-0.75	-0.42	1.78	1.69	0.46	2.06	2.91
Religion	-0.32	-0.71	2.81	0.39	-0.23	-1.62	0.03	0.15	3.86	1.12	0.20	-0.12	-2.24	1.15	2.06	0.29	0.21	-0.33	-1.36
High	0.15	-3.51	6.13	-1.28	2.02	4.05	-0.07	-0.12	2.76	0.82	-1.11	0.93	0.45	-0.75	-0.30	-0.36	-0.08	-0.08	3.06
Weight	1.12	2.01	-0.54	1.63	0.29	0.88	0.25	0.29	3.68	-5.95	0.73	0.73	1.29	2.32	-0.39	0.99	0.24	2.58	3.62
Weight in past 3 month	0.94	4.07	-1.77	3.60	0.18	-2.12	0.22	0.31	4.09	-2.72	1.36	1.19	0.38	1.80	-1.68	1.26	0.36	2.84	-5.10
Change Weight in past 3 Month	0.63	1.04	-0.58	0.70	1.36	-0.72	0.14	0.10	0.46	1.61	0.54	-0.87	1.76	1.20	1.36	0.47	0.01	1.19	3.59
Occupation	-0.44	1.35	-3.67	0.86	-0.28	2.90	-0.48	-0.12	2.01	-5.05	-0.27	2.52	1.82	-3.64	3.96	-0.14	-0.13	0.45	-2.79
Income	0.40	3.01	2.57	0.30	2.71	-0.08	0.02	0.13	3.78	-2.04	-0.03	0.87	1.39	1.61	-1.93	0.99	0.13	1.05	-2.51
ABO	0.05	-4.45	-4.05	0.90	-1.40	-3.12	-0.02	0.06	0.56	3.51	0.25	-2.95	-0.75	-3.60	0.62	0.04	0.02	0.89	0.87
Donation Place	-0.12	-3.53	1.18	1.01	0.04	2.09	-0.37	-0.28	-6.98	-0.99	1.27	2.00	3.56	-1.22	1.41	-0.71	-0.26	0.48	3.12
Single Donor	0.42	-1.94	-0.43	1.42	0.80	0.37	0.24	0.10	-0.78	0.79	1.17	0.45	0.63	1.10	-0.13	0.96	0.10	1.33	-0.34
No.Single Donor	0.14	0.09	-0.06	0.32	-0.01	1.04	0.11	0.26	0.22	-0.41	0.31	-0.36	-0.15	0.84	-0.41	0.21	0.29	0.08	1.35
Donation/year	0.73	-0.85	-3.33	0.28	3.69	3.13	0.31	0.33	4.96	3.63	0.37	-1.46	1.48	4.08	1.23	0.77	0.33	0.80	4.79
Donation	-0.45	-1.05	3.13	-2.43	2.07	-0.32	-0.11	0.06	4.53	-2.74	-1.70	-1.76	-0.69	1.22	2.87	-1.19	0.03	-1.59	0.05

Input Layer Node	Hidden 1 Layer Node																		
Attribute	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Interval donation	0.18	0.89	-1.13	0.85	0.30	1.37	0.13	0.27	-1.35	-1.64	0.31	-1.11	1.48	-1.29	2.24	0.15	0.18	0.30	0.67
Donation success	-0.27	-1.73	1.73	-0.44	-0.44	0.17	-0.09	0.12	-0.70	-1.61	-0.18	-0.30	-0.12	0.02	-1.34	-0.26	0.06	-0.42	-0.12
Previous Hb	2.42	6.69	-12.84	3.29	4.74	15.70	1.52	1.18	4.72	17.39	3.75	-0.44	12.12	9.34	13.71	2.12	1.08	3.42	15.09
BP Sys	0.03	-0.37	-3.08	-3.49	-1.76	-4.77	-0.32	-0.21	-2.64	-0.98	-2.34	-1.81	0.74	0.31	-0.72	-0.79	-0.17	-0.49	0.82
BP Dias	-0.27	-0.10	3.08	-2.78	-3.35	-3.64	-0.15	-0.12	-0.55	2.99	-0.17	-1.95	0.44	3.87	0.15	-1.15	-0.17	-1.33	2.44
Pulse	-0.21	-0.26	-2.18	-0.74	0.76	-2.48	-0.23	-0.27	0.49	-1.43	-0.17	0.18	-2.73	3.39	1.67	-0.90	-0.24	-0.73	1.80
EverLowHb	0.21	-4.99	0.65	-3.31	3.54	-1.91	-0.70	-0.46	3.61	1.17	-2.59	0.07	1.20	-1.40	3.70	-0.92	-0.49	-0.48	-1.22
Disease	0.08	4.18	0.79	0.27	-0.50	-0.52	0.10	0.15	0.31	1.05	0.32	0.14	-0.23	4.22	-1.35	0.05	0.23	0.17	1.02
Sleep Type	-0.46	-0.91	-1.81	-1.90	1.35	0.49	0.11	0.11	-1.84	1.68	-0.81	1.22	0.78	-1.95	1.78	-0.21	0.08	-1.61	-7.06
Sleep hour	0.07	1.13	-2.57	-2.73	2.31	-0.63	-0.31	-0.04	4.52	2.33	-2.99	-1.11	-3.41	0.00	1.19	-0.94	-0.10	-0.06	2.70
Exercise	0.50	11.26	-2.04	-1.67	0.61	-3.99	0.31	0.22	-0.19	-1.61	-0.78	0.80	0.32	1.52	-1.27	0.49	0.28	-0.45	3.45
Smoke	-0.34	-2.38	-4.71	-0.46	-0.15	1.57	-0.12	0.09	0.21	2.14	-0.05	-0.44	1.25	-1.17	-0.37	-0.65	0.11	-0.78	0.37
Alcohol take	0.26	1.00	2.13	2.16	-0.21	4.24	0.11	0.21	-2.57	1.12	0.65	-1.71	-2.48	-0.38	2.62	0.98	0.22	1.00	-6.28
Food Type	-0.07	-1.02	1.88	-0.01	-0.24	-0.67	0.04	0.30	-2.07	0.25	0.07	0.30	-1.29	-1.36	-2.95	-0.04	0.21	-0.35	1.00
Fe take	0.04	-2.34	3.38	2.07	1.97	-0.05	-0.24	0.10	-1.46	2.29	1.11	-1.38	1.05	2.13	-0.78	0.72	0.02	0.87	-1.25
Fe Why not	-1.24	-2.98	-1.68	-1.47	0.39	1.94	-0.14	-0.05	0.01	2.97	0.05	-0.77	0.37	-8.24	-3.72	-1.00	-0.10	-1.90	-0.18
Bias	-0.01	0.55	0.08	-0.26	0.20	-1.16	-0.05	-0.22	0.21	0.41	-0.18	0.37	0.05	-1.08	0.39	-0.06	-0.19	0.05	-1.67

ตารางที่ 48 ค่าน้ำหนักของ Hidden 1 Layer โหนดในชั้น Output Layer ของตัวแบบ ANN ที่คัดเลือกตัวแปรด้วยวิธีลดตัวแปร (backward elimination)

Hidden 1 Layer Node	Output Layer Node	
	Class '0' (Sigmoid)	Class '1' (Sigmoid)
Node 1	-1.505	1.493
Node 2	-9.7	9.701
Node 3	10.441	-10.44
Node 4	-5.575	5.592
Node 5	-6.452	6.448
Node 6	-9.451	9.452
Node 7	-0.686	0.655
Node 8	-0.357	0.367
Node 9	-9.368	9.367
Node 10	-10.344	10.346
Node 11	-3.183	3.167
Node 12	4.964	-4.962
Node 13	-9.003	9.001
Node 14	-8.65	8.651
Node 15	-7.586	7.583
Node 16	-2.057	2.059
Node 17	-0.335	0.349
Node 18	-3.104	3.11
Node 19	-10.423	10.427
Threshold	10.176	-10.174

ทำการ Validation ตัวแบบพยากรณ์ด้วยชุดข้อมูลฝึกหัด (Training dataset) พบว่าให้ค่าความถูกต้อง (accuracy) เท่ากับร้อยละ 86.07 ค่าความไว (sensitivity) ร้อยละ 62.23 ค่าความจำเพาะ (specificity) ร้อยละ 92.62 ค่าการทำนายผลบวก (positive predictive value) ร้อยละ 69.85 และค่าการทำนายผลลบ (negative predictive value) ร้อยละ 89.92 (ตารางที่ 48) และประเมินประสิทธิภาพตัวแบบพยากรณ์ด้วยข้อมูลทดสอบ (Testing dataset) พบว่าให้ค่าความถูกต้อง (accuracy) เท่ากับร้อยละ 84.86 ค่าความไว (sensitivity) ร้อยละ 67.02 ค่า

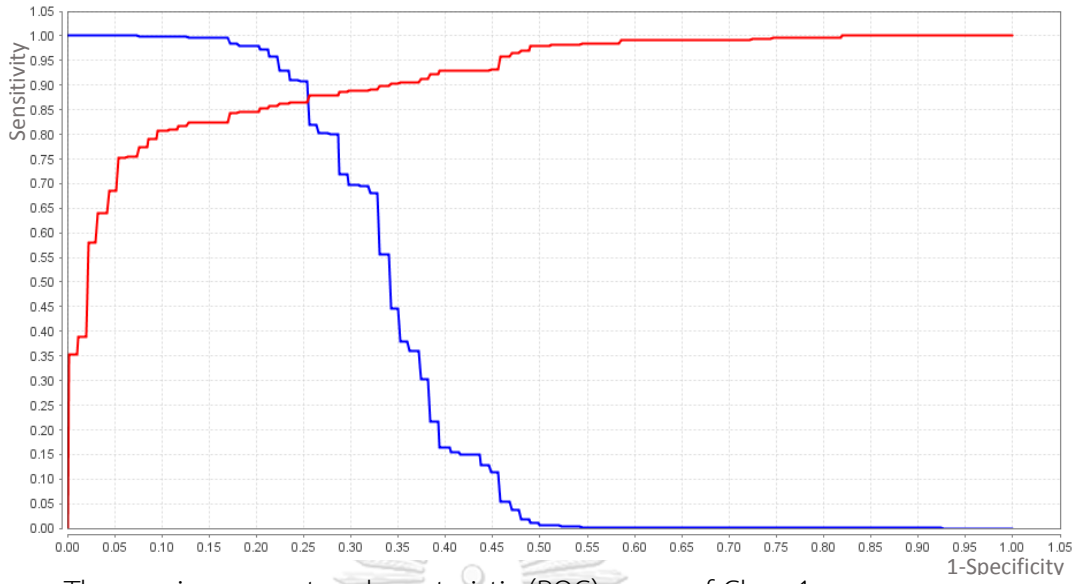
ความจำเพาะ (specificity) ร้อยละ 89.77 ค่าการทำนายผลบวก (positive predictive value) ร้อยละ 64.29 และค่าการทำนายผลลบ (negative predictive value) ร้อยละ 90.83 (ตารางที่ 49) โดยมีค่า AUC เท่ากับ 0.916 (รูปที่ 65)

ตารางที่ 49 Confusion matrix of training dataset using ANN model with backward elimination variables

	Hb ไม่ผ่านเกณฑ์	Hb ผ่านเกณฑ์	Class precision
พยากรณ์ Hb ไม่ผ่านเกณฑ์	234	101	69.85%
พยากรณ์ Hb ผ่านเกณฑ์	142	1267	89.92%
Class recall	62.23%	92.62%	

ตารางที่ 50 Confusion matrix of testing dataset using ANN model with backward elimination variables

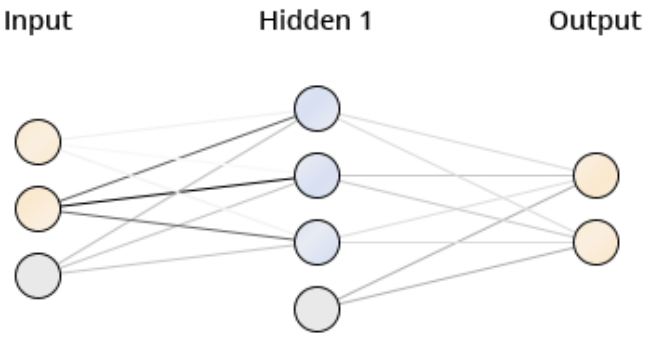
	Hb ไม่ผ่านเกณฑ์	Hb ผ่านเกณฑ์	Class precision
พยากรณ์ Hb ไม่ผ่านเกณฑ์	63	35	64.29%
พยากรณ์ Hb ผ่านเกณฑ์	31	307	90.83%
Class recall	67.02%	89.77%	



รูปที่ 65 Area under the curve (AUC) of testing dataset using ANN model with backward elimination variables

4.4.4 ผลการคัดเลือกตัวแปรวิธีเพิ่มตัวแปรและลดตัวแปร (optimize selection)

ตัวแบบพยากรณ์โครงข่ายประสาทเทียม (artificial neural networks; ANN) โดยวิธีเพิ่มตัวแปรและลดตัวแปร (optimize selection) ได้เป็น ANN ที่มี 3 layer (รูปที่ 66) ประกอบด้วย Input layer จำนวน 3 โหนด โดย 2 โหนดเป็นโหนดนำเข้าตัวแปรค่าฮีโมโกลบินครั้งที่ผ่านมา (Previous Hb) และตัวแปรสภานะภาพ (Status) อีก 1 โหนดคือ Threshold (Bias) ส่วน Hidden layer ประกอบด้วย 4 โหนด โดย 3 โหนดจะเป็นโหนดที่มีค่าน้ำหนักของแต่ละตัวแปรพยากรณ์ทั้ง 2 ตัวแปรจากสมการจำแนกแบบ Sigmoid และโหนด Threshold (Bias) 1 โหนด (ตารางที่ 50) ส่วนชั้น Output Layer ประกอบด้วย 2 โหนดคือ กลุ่มที่มีผ่าน Hb ผ่านเกณฑ์และไม่ผ่านเกณฑ์ โดยแต่ละโหนดจะรับค่าน้ำหนักจากแต่ละโหนดของชั้น Hidden 1 Layer (ตารางที่ 51)



รูปที่ 66 โครงข่ายประสาทเทียมที่ได้จากการคัดเลือกตัวแปรวิธีเพิ่มตัวแปรและลดตัวแปร (optimize selection)

ตารางที่ 51 ค่าน้ำหนักตัวแปร Hidden 1 Layer ของตัวแบบ ANN โดยคัดเลือกตัวแปรวิธีเพิ่มตัวแปรและลดตัวแปร (optimize selection)

Input Layer Node	Hidden 1 Layer Node		
	1	2	3
Previous Hb	1.149	-0.61	1.065
Status	13.264	21.083	12.562
Bias	5.206	4.909	4.816

ตารางที่ 52 แสดงค่าน้ำหนักของ Hidden 1 Layer โหนดในชั้น Output Layer ของตัวแบบ ANN ที่คัดเลือกตัวแปรด้วยวิธีเพิ่มตัวแปรและลดตัวแปร (optimize selection)

Hidden 1 Layer	Output Node	
	Class '0' (Sigmoid)	Class '1' (Sigmoid)
Node 1	-2.95	2.92
Node 2	-4.389	4.388
Node 3	-2.719	2.748
Threshold	6.557	-6.555

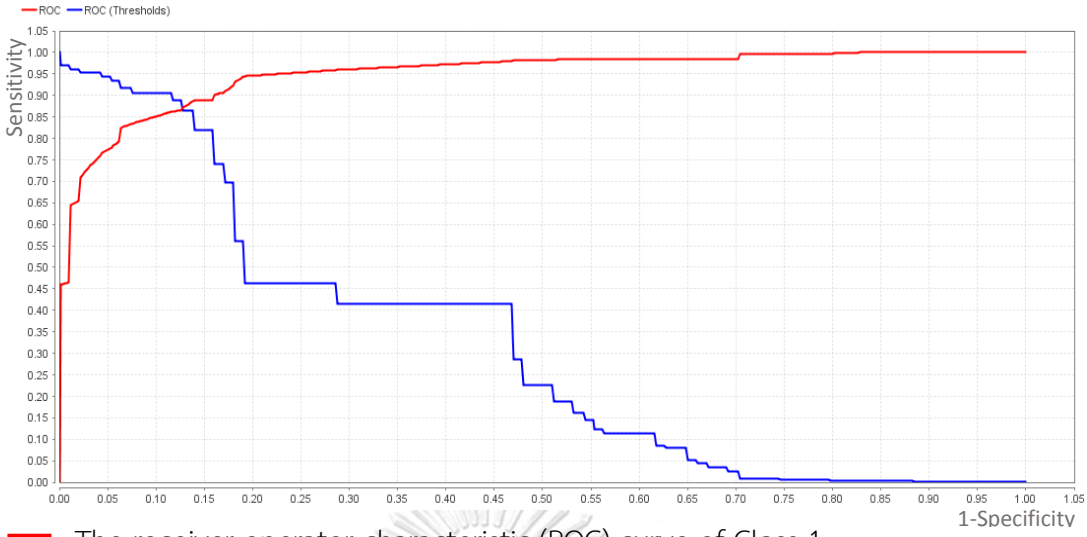
ทำการ Validation ตัวแบบพยากรณ์ด้วยชุดข้อมูลฝึกหัด (Training dataset) พบว่าให้ค่าความถูกต้อง (accuracy) เท่ากับร้อยละ 91.00 ค่าความไว (sensitivity) ร้อยละ 76.06 ค่าความจำเพาะ (specificity) ร้อยละ 95.10 ค่าการทำนายผลบวก (positive predictive value) ร้อยละ 81.02 และค่าการทำนายผลลบ (negative predictive value) ร้อยละ 93.53 (ตารางที่ 52) และประเมินประสิทธิภาพตัวแบบพยากรณ์ด้วยข้อมูลทดสอบ (Testing dataset) พบว่าให้ค่าความถูกต้อง (accuracy) เท่ากับร้อยละ 90.60 ค่าความไว (sensitivity) ร้อยละ 81.91 ค่าความจำเพาะ (specificity) ร้อยละ 92.98 ค่าการทำนายผลบวก (positive predictive value) ร้อยละ 76.24 และค่าการทำนายผลลบ (negative predictive value) ร้อยละ 94.93 (ตารางที่ 53) โดยมีค่า AUC เท่ากับ 0.950 (รูปที่ 67)

ตารางที่ 53 Confusion matrix of training dataset using ANN model with optimize selection variables

	Hb ไม่ผ่านเกณฑ์	Hb ผ่านเกณฑ์	Class precision
พยากรณ์ Hb ไม่ผ่านเกณฑ์	286	67	81.02%
พยากรณ์ Hb ผ่านเกณฑ์	90	1301	93.53%
Class recall	76.06%	95.10%	

ตารางที่ 54 Confusion matrix of testing dataset using ANN model with optimize selection variables

	Hb ไม่ผ่านเกณฑ์	Hb ผ่านเกณฑ์	Class precision
พยากรณ์ Hb ไม่ผ่านเกณฑ์	77	24	76.24%
พยากรณ์ Hb ผ่านเกณฑ์	17	318	94.93%
Class recall	81.91%	92.98%	



รูปที่ 67 Area under the curve (AUC) of testing dataset using ANN model with optimize selection variables

4.5 การประเมินประสิทธิภาพตัวแบบพยากรณ์

ในการพยากรณ์ค่าฮีโมโกลบินในผู้บริจาคหากพยากรณ์ได้ล่วงหน้าจะช่วยให้ผู้บริจาคที่อาจมีผลตรวจไม่ผ่านเกณฑ์ได้ทราบล่วงหน้า เพื่อให้เกิดการตระหนักรู้ในการดูแลสุขภาพหรือรับประทานยาธาตุเหล็กอย่างต่อเนื่องก่อนมาบริจาคครั้งต่อไป ดังนั้นการวิจัยนี้จึงคำนึงถึงค่าความถูกต้อง (accuracy) ค่าความไว (sensitivity) ค่าการทำนายผลบวก (positive predictive value) และค่า AUC มากกว่าค่าความจำเพาะ (specificity) และค่าการทำนายผลลบ (negative predictive value)

จากการวิจัยนี้พบว่าตัวแบบพยากรณ์ต้นไม้ตัดสินใจ (decision tree) ที่ใช้การคัดเลือกตัวแปรวิธีเพิ่มตัวแปร (forward selection) และวิธีเพิ่มตัวแปรและลดตัวแปร (optimize selection) ให้ค่าความถูกต้อง (accuracy) ค่าความไว (sensitivity) ความจำเพาะ (specificity) ค่าการทำนายผลบวก (positive predictive value) และค่าการทำนายผลลบ (negative predictive value) สูงสุดเท่ากันคือร้อยละ 92.20 82.98 94.74 81.25 และ 95.29 ตามลำดับ ส่วนค่า AUC พบว่าตัวแบบพยากรณ์ ANN ที่ใช้การคัดเลือกตัวแปรวิธีเพิ่มตัวแปร (Forward selection) และวิธีเพิ่มตัวแปรและลดตัวแปร (optimize selection) ให้ค่าสูงสุดคือ 0.950 (ตารางที่ 54)

ตารางที่ 55 เปรียบเทียบประสิทธิภาพตัวแบบพยากรณ์

Predictive Models	Selective Variables	Accuracy	Sensitivity	Specificity	PPV	NPV	AUC
Decision tree	Enter Regression	91.06%	79.79%	94.15%	78.95%	94.43%	0.879
	forward selection	**92.20%**	**82.98%**	**94.74%**	**81.25%**	**95.29%**	0.943
	Backward elimination	90.83%	78.72%	94.15%	78.72%	94.15%	0.937
	Optimize selection	**92.20%**	**82.98%**	**94.74%**	**81.25%**	**95.29%**	0.943
SVM	Enter Regression	86.70%	68.09%	91.81%	69.57%	91.28%	0.926
	forward selection	91.51%	80.85%	94.44%	80.00%	94.72%	0.939
	Backward elimination	88.99%	79.79%	91.52%	72.12%	94.28%	0.943
	Optimize selection	91.51%	80.85%	94.44%	80.00%	94.72%	0.945
Naïve Bayesain Classifier	Enter Regression	75.46%	79.79%	74.27%	46.01%	93.04%	0.856
	forward selection	89.22%	78.72%	92.11%	73.27%	94.03%	0.936
	Backward elimination	82.34%	68.09%	86.26%	57.66%	90.77%	0.876
	Optimize selection	89.22%	78.72%	92.11%	73.27%	94.03%	0.936
ANN	Enter Regression	87.84%	74.47%	91.52%	70.71%	92.88%	0.924
	forward selection	90.60%	81.91%	92.98%	76.24%	94.93%	**0.950**
	Backward elimination	84.86%	67.02%	89.77%	64.29%	90.83%	0.916
	Optimize selection	90.60%	81.91%	92.98%	76.24%	94.93%	**0.950**

บทที่ 5
สรุปผลงานวิจัยและอภิปรายผล

5.1 อภิปรายผล

ในการศึกษานี้ได้พัฒนาตัวแบบพยากรณ์ด้วย 4 เทคนิคทางเหมืองข้อมูลได้แก่ ต้นไม้ตัดสินใจ (decision tree) ซัพพอร์ตเวกเตอร์แมชชีน (SVM) การจำแนกแบบเบส์อย่างง่าย (naïve bayesain classifier) และโครงข่ายประสาทเทียม (artificial neural networks) ด้วยโปรแกรม RapidMiner จากนั้นศึกษาการเพิ่มประสิทธิภาพของตัวแบบพยากรณ์แต่ละเทคนิคด้วยวิธีการคัดเลือกตัวแปร 4 วิธีคือ วิธีนำตัวแปรเข้าทั้งหมด (Enter regression) วิธีเพิ่มตัวแปร (Forward selection) วิธีลดตัวแปร (backward elimination) และวิธีเพิ่มตัวแปรและลดตัวแปร (optimize selection) พบว่าตัวแบบพยากรณ์ที่ให้ค่าความถูกต้อง (accuracy) สูงที่สุดคือ ต้นไม้ตัดสินใจ (decision tree) รองลงมาคือ ซัพพอร์ตเวกเตอร์แมชชีน (SVM) โครงข่ายประสาทเทียม (artificial neural networks) และการจำแนกแบบเบส์อย่างง่าย (naïve bayesain classifier) ตามลำดับและพบว่าการคัดเลือกตัวแปรด้วยวิธีเพิ่มตัวแปร (Forward selection) และวิธีเพิ่มตัวแปรและลดตัวแปร (optimize selection) เป็นวิธีที่เหมาะสมกว่าวิธีนำตัวแปรเข้าทั้งหมด (Enter regression) และวิธีลดตัวแปร (backward elimination)

นอกจากนี้ยังพบว่าการคัดเลือกตัวแปรด้วยวิธีเพิ่มตัวแปร (Forward selection) และวิธีเพิ่มตัวแปรและลดตัวแปร (optimize selection) ให้ค่าการวัดประสิทธิภาพของตัวแบบเท่ากันทุกพารามิเตอร์ สาเหตุอาจเกิดจากการคัดเลือกตัวแปรพยากรณ์ด้วยวิธีเพิ่มตัวแปรและลดตัวแปร (optimize selection) ของโปรแกรม RapidMiner นั้นผู้วิจัยกำหนดอัตราการนำเข้าตัวแปรเท่ากับร้อยละ 78 ซึ่งมีค่าสูงกว่าอัตรานำตัวแปรออกที่ร้อยละ 22 มากเกินไปจึงทำให้ได้ผลไม่แตกต่างจากวิธีเพิ่มตัวแปร (Forward selection) นอกจากนี้ยังพบว่าจำนวนตัวแปรที่เหลือจากการคัดเลือกด้วยวิธีเพิ่มตัวแปร (Forward selection) และวิธีเพิ่มตัวแปรและลดตัวแปร (optimize selection) มีจำนวนน้อยกว่าวิธีนำเข้าตัวแปรทั้งหมด (Enter regression) และวิธีลดตัวแปร (backward elimination) แต่ให้ประสิทธิภาพที่สูงกว่าเช่นเดียวกันทั้ง 4 ชนิดของตัวแบบพยากรณ์ ดังนั้นการมีตัวแปรที่ไม่จำเป็นในตัวแบบที่มากเกินไปนอกจากจะเป็นตัวแบบที่ไม่ประหยัดแล้วยังส่งผลให้ประสิทธิภาพของตัวแบบลดลงด้วย โดยพบว่าตัวแปรที่มีความสำคัญที่สุดคือ ค่าฮีโมโกลบินครั้งที่ผ่านมา (Previous Hb) เนื่องจากพบในตัวแบบทุกชนิดและทุกวิธีการคัดเลือกตัวแปร โดยค่าฮีโมโกลบิน

ครั้งที่ผ่านมา (Previous Hb) มีค่าน้ำหนักของตัวแปรสูงที่สุดในทุกตัวแบบ การวิจัยนี้พบว่าตัวแบบพยากรณ์ต้นไม้ตัดสินใจ (Decision tree) ที่ใช้การคัดเลือกตัวแปรด้วยวิธีเพิ่มตัวแปร (Forward selection) และวิธีเพิ่มตัวแปรและลดตัวแปร (optimize selection) เป็นตัวแบบที่มีประสิทธิภาพดีที่สุดให้ค่า Accuracy Sensitivity Specificity PPV NPV และ AUC เท่ากันทุกพารามิเตอร์มีค่าเท่ากับ 92.20 82.98 94.74 81.25 95.29 และ 0.943 ตามลำดับ โดยหลังการคัดเลือกคงเหลือตัวแปรจำนวน 5 ตัวแปรคือ เพศ (Gender) ประวัติการตรวจฮีโมโกลบินไม่ผ่านเกณฑ์ (EverLowHb) ค่าฮีโมโกลบินครั้งที่ผ่านมา (Previous Hb) ที่อยู่ (Address1) และพฤติกรรมการพักผ่อน (Sleep type) จะสังเกตได้ว่าตัวแปรที่อยู่ (Address1) เมื่อทดสอบสถิติเบื้องต้นในตอนแรกพบว่าไม่มีความแตกต่างกันระหว่างกลุ่มที่มีค่า Hb ผ่านเกณฑ์และไม่ผ่านเกณฑ์อย่างมีนัยสำคัญทางสถิติ แต่กลับเป็นตัวแปรที่ผ่านการคัดเลือกเข้าเป็นตัวแบบพยากรณ์ ดังนั้นงานวิจัยนี้จึงแสดงให้เห็นว่าสามารถนำตัวแปรที่สนใจทั้งหมดมาใช้ในการสร้างตัวแบบได้ โดยผ่านการคัดเลือกตัวแปรที่เหมาะสม ไม่จำเป็นต้องนำเข้าเฉพาะตัวแปรที่มีความแตกต่างกันระหว่างกลุ่มอย่างมีนัยสำคัญทางสถิติเท่านั้น นอกจากนี้ยังพบว่าตัวแบบทั้งหมดมีค่าความไว (sensitivity) และค่าการทำนายผลบวก (positive predictive value) ต่ำกว่าค่าความจำเพาะ (specificity) และค่าทำนายผลลบ (negative predictive value) เช่นเดียวกันทั้งตัวแบบ 4 เทคนิค หมายความว่าผู้ที่มีค่าฮีโมโกลบินต่ำกว่าเกณฑ์จะถูกพยากรณ์จำแนกว่าไม่ผ่านเกณฑ์ได้ในอัตราที่ต่ำแต่ผู้ที่มีค่าฮีโมโกลบินผ่านเกณฑ์มีโอกาสจะถูกพยากรณ์ว่าผ่านเกณฑ์ในอัตราที่สูงและถูกต้องกว่า ทั้งนี้อาจเนื่องมาจากระดับค่าฮีโมโกลบินในกลุ่มผู้บริจาคที่มีระดับค่าใกล้เคียงเกณฑ์ผ่านเกณฑ์คือ 12.5 mg/dl ในผู้หญิงและ 13.0 mg/dl ในผู้ชายอาจมีการเปลี่ยนแปลงหรือคลาดเคลื่อนจากปัจจัยอื่น ๆ ในแต่ละรอบของการบริจาคโลหิต ซึ่งไม่ได้เก็บข้อมูลตัวแปรเหล่านั้นในการศึกษานี้เช่น ประวัติสุขภาพ ปริมาณธาตุเหล็กสะสมของแต่ละคน ความถูกต้องความแม่นยำของวิธีการตรวจฮีโมโกลบินในแต่ละครั้ง เป็นต้น ซึ่งในความคาดหวังของพัฒนาตัวแบบพยากรณ์ในอุดมคติควรจะมีค่าความไว (sensitivity) และค่าความจำเพาะ (specificity) สูงที่สุด นั้นแสดงว่าประสิทธิภาพของตัวแบบที่ได้จากการวิจัยนี้มีความเหมาะสมในการนำไปใช้จำแนกผู้ที่มีค่า Hb ผ่านเกณฑ์มากกว่าการนำไปใช้จำแนกผู้ที่มีค่า Hb ไม่ผ่านเกณฑ์ แต่อย่างไรก็ตามค่าความไว (sensitivity) ค่าการทำนายผลบวก (positive predictive value) และค่า AUC ของตัวแบบพยากรณ์ต้นไม้ตัดสินใจ (decision tree) ที่ได้จากการวิจัยนี้ยังถือว่าอยู่ในเกณฑ์ดี จึงอาจสรุปได้ว่าตัวแบบพยากรณ์ต้นไม้ตัดสินใจ (decision tree) ที่ได้จากการวิจัยนี้มีความเหมาะสมในการนำไปใช้พยากรณ์ผลตรวจฮีโมโกลบินในผู้บริจาคโลหิตได้

เมื่อเปรียบเทียบการพยากรณ์ค่าฮีโมโกลบินของการศึกษานี้กับการศึกษาอื่นเกี่ยวข้องกับการพยากรณ์ค่าฮีโมโกลบินในผู้บริจาคโลหิต ได้แก่ Kazem Nasserinejad และคณะ(38) ศึกษาการพยากรณ์ค่า Hb ด้วยเทคนิค Multiple linear regression model, linear mixed effects model และ Transition model ที่พัฒนาด้วยโปรแกรม R ใช้ข้อมูลผู้บริจาคโลหิตรายใหม่และกลับมาบริจาคอีกอย่างน้อย 1 ครั้งในช่วง 1 ม.ค. 2550 ถึง 31 ธ.ค. 2552 จำนวน 15, 625 ราย แยกศึกษาระหว่างเพศชาย-หญิง โดยใช้ตัวแปร 3 ตัวคือ อายุ ฤดูกาลที่มาบริจาคโลหิตและช่วงระยะเวลาห่างจากการมาบริจาคครั้งแรก พบกว่าเทคนิค linear mixed effects model และ Transition model ให้ค่าความถูกต้องการพยากรณ์ที่ดีกว่าเทคนิค Multiple linear regression model เมื่อพิจารณาจากค่า Mean squared prediction error (MSPE) และพบว่าค่า AUC ของเทคนิค linear mixed effects model และ Transition model ในเพศชายอยู่ในช่วงที่ดีคือ 0.83 และ 0.81 ในขนาดที่ค่า AUC ในเพศหญิงอยู่ในช่วงที่ยอมรับได้คือ 0.73 และ 0.72 ตามลำดับ การศึกษาของ Jesse Fokkinga(44) ศึกษาตัวแปรจำนวน 10 ตัวแปรประกอบด้วย อายุ BMI ปริมาณเลือดที่บริจาค ฤดูกาล ค่า Hb ครั้งที่ผ่านมา ค่า zinc protoporphyrin (ZPP) ครั้งที่ผ่านมา เวลาที่มาบริจาค จำนวนครั้งที่บริจาคในรอบ 2 ปี ระยะเวลาห่างจากการบริจาคครั้งที่ผ่านมา การมีประจำเดือนสำหรับเพศหญิง ศึกษาในผู้บริจาคจำนวน 2,215 ราย ระหว่าง ต.ค. 2552 ถึง ก.พ. 2557 จำนวนครั้งการบริจาคโลหิตรวม 14,006 ครั้ง โดยใช้เทคนิค Mixed-effects model, Mixed-effects transition model, decision tree ชนิด Random forest, Gradient tree boosting และ Hierarchical Ornstein–Uhlenbeck พบว่าเทคนิค decision tree ชนิดRandom forest ได้ค่า AUC ดีที่สุดคืออยู่ในช่วงที่ยอมรับได้ในเพศหญิงเท่ากับ 0.717 และ 0.690 ในเพศชาย เนื่องจากการศึกษาทั้งของ Kazem Nasserinejad และ Jesse Fokkinga เป็นการศึกษาทำนายระดับค่า Hb ซึ่งต่างจากการศึกษาฉบับนี้ที่ศึกษาการจำแนกกลุ่มจึงไม่สามารถเปรียบเทียบระดับค่าความถูกต้องของการทำนายได้แต่เมื่อพิจารณาเปรียบเทียบค่า AUC จะพบว่าได้ค่าน้อยกว่าการศึกษาฉบับนี้ถึงแม้ว่าจะใช้จำนวนตัวอย่างที่มากกว่าก็ตามซึ่งเกิดจากการสร้างตัวแบบพยากรณ์เพื่อทำนายระดับค่า Hb นั้นมีความซับซ้อนกว่าการจำแนกกลุ่มที่มีเป้าหมายเพียง 2 กลุ่มคือผ่านและไม่ผ่านเกณฑ์ จึงทำให้มีโอกาสในการทำนายได้ถูกกว่าการทำนายพยากรณ์ระดับค่า Hb ที่อาจมีปัจจัยอื่น ๆ ที่ส่งผลระดับค่าที่ทั้ง Kazem Nasserinejad และ Jesse Fokkinga ไม่ได้นำมาศึกษา

ในการศึกษาครั้งนี้ยังพบว่าข้อมูลผลการตรวจฮีโมโกลบินในครั้งที่ผ่านมา (Previous Hb) เป็นตัวแปรที่มีความสำคัญโดยพิจารณาจากค่าน้ำหนักของตัวแปรที่ได้จากตัวแบบต่าง ๆ ที่มีค่าสูง

ที่สุดและค่า F ทางสถิติที่มีค่า 1093.893 แต่พบว่าผู้บริจาคโลหิตจำนวนมากให้ข้อมูลไม่ครบถ้วน ถูกต้องหรือระบุเพียงว่าผ่านหรือไม่ผ่าน เนื่องมาจากในอดีตการตรวจคัดกรองฮีโมโกลบินใช้วิธีวัดความถ่วงจำเพาะด้วยสารละลายคอปเปอร์ซัลเฟต ทำให้ในการชดเชยค่าที่ขาดหายต้องวิเคราะห์ร่วมกับข้อมูลในระบบ HIIG การให้ข้อมูลตัวแปรประวัติการตรวจฮีโมโกลบินไม่ผ่านเกณฑ์ (EverLowHb) เพื่อพิจารณาว่าจะต้องใช้ค่าค่าเฉลี่ยของตัวแปรหรือค่าผลตรวจเก่าที่มีค่าผลตรวจ Hb ในระบบ HIIG ซึ่งเป็นการบริจาคโลหิตที่ไม่ใช่รอบที่ผ่านมาซึ่งอาจส่งผลให้ตัวแบบพยากรณ์มีความผิดพลาดในการพยากรณ์

5.2 สรุปผลงานวิจัย

ตัวแบบพยากรณ์ต้นไม้ตัดสินใจ (decision tree) ที่ใช้การคัดเลือกตัวแปรด้วยวิธีเพิ่มตัวแปร (Forward selection) และวิธีเพิ่มตัวแปรและลดตัวแปร (optimize selection) เหมาะสมในการนำไปใช้พยากรณ์มากกว่าตัวแบบพยากรณ์ที่ได้จากเทคนิคอื่น ๆ เนื่องจากให้ค่าความถูกต้อง (accuracy) ค่าความไว (sensitivity) ค่าความจำเพาะ (specificity) ค่าพยากรณ์ผลบวก (positive predictive value) และ ค่าพยากรณ์ผลลบ (negative predictive value) สูงสุด ถึงแม้ว่าค่า AUC จะน้อยกว่าตัวแบบพยากรณ์ที่ได้จากเทคนิคตัวแบบโครงข่ายประสาทเทียม (artificial neural networks) แต่ก็จัดอยู่ในเกณฑ์ดีเช่นกัน

เมื่อนำเงื่อนไขการพยากรณ์ที่ได้ตัวแบบต้นไม้ตัดสินใจ (decision tree) ที่ได้จากการศึกษานี้สามารถนำมาสร้างเป็นกฎการตัดสินใจจำแนกกลุ่มได้เป็น 25 กฎแบ่งเป็น กฎการจำแนกกลุ่มผู้บริจาคโลหิตที่มีค่า Hb ผ่านเกณฑ์ 12 กฎ ดังนี้

1. ค่า Hb ครั้งที่ผ่านมา > 13.550 จะถูกพยากรณ์เป็น ผ่านเกณฑ์ โดยโอกาสจะถูกต้องร้อยละ 98.70
2. ค่า Hb ครั้งที่ผ่านมา > 13.050 และค่า Hb ครั้งที่ผ่านมา ≤ 13.550 และเป็นเพศชาย และไม่เคยมีประวัติค่า Hb ไม่ผ่านเกณฑ์ จะถูกพยากรณ์เป็น ผ่านเกณฑ์ โดยโอกาสจะถูกต้องร้อยละ 95.00
3. ค่า Hb ครั้งที่ผ่านมา > 12.650 และค่า Hb ครั้งที่ผ่านมา ≤ 13.550 และเป็นเพศหญิง จะถูกพยากรณ์เป็น ผ่านเกณฑ์ โดยโอกาสจะถูกต้องร้อยละ 88.04
4. ค่า Hb ครั้งที่ผ่านมา > 13.300 และค่า Hb ครั้งที่ผ่านมา ≤ 13.550 และเป็นเพศชาย และเคยมีประวัติค่า Hb ไม่ผ่านเกณฑ์ และพักผ่อนเป็นเวลาแน่นอน จะถูกพยากรณ์เป็น ผ่านเกณฑ์ โดยโอกาสจะถูกต้องร้อยละ 75.00

5. ค่า Hb ครั้งที่ผ่านมา > 13.050 และค่า Hb ครั้งที่ผ่านมา ≤ 13.550 และเป็นเพศชาย และเคยมีประวัติค่า Hb ไม่ผ่านเกณฑ์ 1 และพักผ่อนไม่เป็นเวลาแน่นอน จะถูกพยากรณ์เป็น ผ่านเกณฑ์ โดยโอกาสจะถูกต้องร้อยละ 100

6. ค่า Hb ครั้งที่ผ่านมา > 12.850 และค่า Hb ครั้งที่ผ่านมา ≤ 13.050 และเป็นเพศชาย และเคยมีประวัติค่า Hb ไม่ผ่านเกณฑ์ จะถูกพยากรณ์เป็น ผ่านเกณฑ์ โดยโอกาสจะถูกต้องร้อยละ 100

7. ค่า Hb ครั้งที่ผ่านมา > 12.550 และค่า Hb ครั้งที่ผ่านมา ≤ 12.650 และเป็นเพศหญิง และไม่เคยมีประวัติค่า Hb ไม่ผ่านเกณฑ์ และอาศัยในเขตอำเภอเมือง จะถูกพยากรณ์เป็น ผ่านเกณฑ์ โดยโอกาสจะถูกต้องร้อยละ 84.62

8. ค่า Hb ครั้งที่ผ่านมา > 12.550 และค่า Hb ครั้งที่ผ่านมา ≤ 12.650 และเป็นเพศหญิง และไม่เคยมีประวัติค่า Hb ไม่ผ่านเกณฑ์ และไม่ได้อาศัยในเขตอำเภอเมือง และพักผ่อนเป็นเวลาแน่นอน จะถูกพยากรณ์เป็น ผ่านเกณฑ์ โดยโอกาสจะถูกต้องร้อยละ 77.78

9. ค่า Hb ครั้งที่ผ่านมา > 12.550 และค่า Hb ครั้งที่ผ่านมา ≤ 12.650 และเป็นเพศหญิง และเคยมีประวัติค่า Hb ไม่ผ่านเกณฑ์ 1 และอาศัยในเขตอำเภอเมือง และพักผ่อนเป็นเวลาแน่นอน จะถูกพยากรณ์เป็น ผ่านเกณฑ์ โดยโอกาสจะถูกต้องร้อยละ 63.64

10. ค่า Hb ครั้งที่ผ่านมา > 12.550 ค่า Hb ครั้งที่ผ่านมา ≤ 12.650 และเป็นเพศหญิง และเคยมีประวัติค่า Hb ไม่ผ่านเกณฑ์ และไม่ได้อาศัยในเขตอำเภอเมือง จะถูกพยากรณ์เป็น ผ่านเกณฑ์ โดยโอกาสจะถูกต้องร้อยละ 100

11. ค่า Hb ครั้งที่ผ่านมา > 12.450 และค่า Hb ครั้งที่ผ่านมา ≤ 12.550 และเป็นเพศหญิง และเคยมีประวัติค่า Hb ไม่ผ่านเกณฑ์ และพักผ่อนเป็นเวลาแน่นอน จะถูกพยากรณ์เป็น ผ่านเกณฑ์ โดยโอกาสจะถูกต้องร้อยละ 80.00

12. ค่า Hb ครั้งที่ผ่านมา > 11.350 และค่า Hb ครั้งที่ผ่านมา ≤ 12.450 และไม่เคยมีประวัติค่า Hb ไม่ผ่านเกณฑ์ และไม่ได้อาศัยในเขตอำเภอเมือง และพักผ่อนไม่เป็นเวลาแน่นอน จะถูกพยากรณ์เป็น ผ่านเกณฑ์ โดยโอกาสจะถูกต้องร้อยละ 66.67

และกฎการจำแนกเป็นกลุ่มผู้บริจาคโลหิตที่มีค่า Hb ไม่ผ่านเกณฑ์ 13 กฎ

1. ค่า Hb ครั้งที่ผ่านมา ≤ 11.350 จะถูกพยากรณ์เป็น ไม่ผ่านเกณฑ์ โดยโอกาสจะถูกต้องร้อยละ 100

2. ค่า Hb ครั้งที่ผ่านมา > 11.350 และค่า Hb ครั้งที่ผ่านมา ≤ 12.450 และเคยมีประวัติค่า Hb ไม่ผ่านเกณฑ์ จะถูกพยากรณ์เป็น ไม่ผ่านเกณฑ์ โดยโอกาสจะถูกต้องร้อยละ 91.36

3. ค่า Hb ครั้งที่ผ่านมา > 11.350 และค่า Hb ครั้งที่ผ่านมา ≤ 12.450 และไม่เคยมีประวัติค่า Hb ไม่ผ่านเกณฑ์ และอาศัยในเขตอำเภอเมือง จะถูกพยากรณ์เป็น ไม่ผ่านเกณฑ์ โดยโอกาสจะถูกต้องร้อยละ 75.00

4. ค่า Hb ครั้งที่ผ่านมา > 11.350 และค่า Hb ครั้งที่ผ่านมา ≤ 12.450 และไม่เคยมีประวัติค่า Hb ไม่ผ่านเกณฑ์ และไม่ได้อาศัยในเขตอำเภอเมือง และพักผ่อนเป็นเวลาแน่นอน จะถูกพยากรณ์เป็น ไม่ผ่านเกณฑ์ โดยโอกาสจะถูกต้องร้อยละ 50.00

5. ค่า Hb ครั้งที่ผ่านมา > 12.450 และค่า Hb ครั้งที่ผ่านมา ≤ 12.550 และไม่เคยมีประวัติค่า Hb ไม่ผ่านเกณฑ์ จะถูกพยากรณ์เป็น ไม่ผ่านเกณฑ์ โดยโอกาสจะถูกต้องร้อยละ 80.37

6. ค่า Hb ครั้งที่ผ่านมา > 12.450 และค่า Hb ครั้งที่ผ่านมา ≤ 12.550 และเคยมีประวัติค่า Hb ไม่ผ่านเกณฑ์ และเป็นเพศชาย จะถูกพยากรณ์เป็น ไม่ผ่านเกณฑ์ โดยโอกาสจะถูกต้องร้อยละ 100

7. ค่า Hb ครั้งที่ผ่านมา > 12.450 และค่า Hb ครั้งที่ผ่านมา ≤ 12.550 และเคยมีประวัติค่า Hb ไม่ผ่านเกณฑ์ และเป็นเพศหญิง และพักผ่อนไม่เป็นเวลาแน่นอน จะถูกพยากรณ์เป็น ไม่ผ่านเกณฑ์ โดยโอกาสจะถูกต้องร้อยละ 62.50

8. ค่า Hb ครั้งที่ผ่านมา > 12.650 และค่า Hb ครั้งที่ผ่านมา ≤ 12.85 และเป็นเพศชาย จะถูกพยากรณ์เป็น ไม่ผ่านเกณฑ์ โดยโอกาสจะถูกต้องร้อยละ 100

9. ค่า Hb ครั้งที่ผ่านมา > 13.050 และค่า Hb ครั้งที่ผ่านมา ≤ 13.300 และเป็นเพศชาย และเคยมีประวัติค่า Hb ไม่ผ่านเกณฑ์ และพักผ่อนเป็นเวลาแน่นอน จะถูกพยากรณ์เป็น ไม่ผ่านเกณฑ์ โดยโอกาสจะถูกต้องร้อยละ 50.00

10. ค่า Hb ครั้งที่ผ่านมา > 12.850 และค่า Hb ครั้งที่ผ่านมา ≤ 13.050 และเป็นเพศชาย และไม่เคยมีประวัติค่า Hb ไม่ผ่านเกณฑ์ จะถูกพยากรณ์เป็น ไม่ผ่านเกณฑ์ โดยโอกาสจะถูกต้องร้อยละ 60.00

11. ค่า Hb ครั้งที่ผ่านมา > 12.550 และค่า Hb ครั้งที่ผ่านมา ≤ 12.650 และเป็นเพศชาย จะถูกพยากรณ์เป็น ไม่ผ่านเกณฑ์ โดยโอกาสจะถูกต้องร้อยละ 100

12. ค่า Hb ครั้งที่ผ่านมา > 12.550 และค่า Hb ครั้งที่ผ่านมา ≤ 12.650 และเป็นเพศหญิง และไม่เคยมีประวัติค่า Hb ไม่ผ่านเกณฑ์ และไม่ได้อาศัยในเขตอำเภอเมือง และพักผ่อนไม่เป็นเวลาแน่นอน จะถูกพยากรณ์เป็น ไม่ผ่านเกณฑ์ โดยโอกาสจะถูกต้องร้อยละ 50.00

13. ค่า Hb ครั้งที่ผ่านมา > 12.550 และค่า Hb ครั้งที่ผ่านมา ≤ 12.650 และเป็นเพศหญิง และเคยมีประวัติค่า Hb ไม่ผ่านเกณฑ์ และอาศัยในเขตอำเภอเมือง และพักผ่อนไม่เป็นเวลาแน่นอน จะถูกพยากรณ์เป็น ไม่ผ่านเกณฑ์ โดยโอกาสจะถูกต้องร้อยละ 80.00

5.3 ข้อเสนอแนะ

การวิจัยนี้ต้องอาศัยความถูกต้องของข้อมูลที่นำมาใช้ในการสร้างตัวแบบ ปัจจุบันสภากาชาดไทย ได้ยกเลิกการตรวจค่าฮีโมโกลบินด้วยสารละลายคอปเปอร์ซัลเฟตแล้ว โดยเปลี่ยนมาใช้

เครื่องตรวจฮีโมโกลบินมิเตอร์แทน ดังนั้นหากมีการเก็บข้อมูลเฉพาะผู้บริจาคโลหิตที่มีค่าผลตรวจฮีโมโกลบินในอดีตนำมาใช้ในการสร้างตัวแบบพยากรณ์จะทำให้ได้ผลที่มีความถูกต้องมากกว่าการศึกษานี้ จากการวิจัยนี้ทำให้พบว่าข้อมูลต่าง ๆ หากจะนำมาให้ในการศึกษาเกี่ยวกับเหมืองข้อมูลจำเป็นต้องเริ่มตั้งแต่การออกแบบระบบฐานข้อมูลให้มีความเหมาะสมจัดเก็บข้อมูลและเอื้อต่อการนำข้อมูลออกมาวิเคราะห์ ปัจจุบันหลายหน่วยงานยังไม่ได้ให้ความสำคัญเรื่องการจัดเก็บข้อมูลเพื่อนำมาวิเคราะห์ทางเหมืองข้อมูล ทำให้เป็นอุปสรรคต่อการศึกษาทางด้านนี้

5.4 งานวิจัยในอนาคต

ในอนาคตควรทำการศึกษาเพิ่มเติมโดยใช้กลุ่มประชากรที่มากขึ้นและเก็บข้อมูลเฉพาะผู้ที่มีข้อมูลค่าผลตรวจฮีโมโกลบินเท่านั้น ซึ่งอาจทำให้ค่าความถูกต้องแม่นยำของการพยากรณ์ที่ดียิ่งขึ้น อีกทั้งยังอาจนำผลของการศึกษานี้ไปพัฒนาเป็นโปรแกรมหรือแอพพลิเคชั่นใช้พยากรณ์ค่า Hb ในผู้บริจาคโลหิตจริง ซึ่งเป็นการศึกษาเพื่อพิสูจน์ความถูกต้องของตัวแบบพยากรณ์ต้นไม้ตัดสินใจที่ได้จากการวิจัยนี้

บรรณานุกรม

1. Organization WH. The 2016 global status report on blood safety and availability2017:[166 p.]. Available from: http://www.who.int/iris/handle/10665/254987.

2. National Blood Centre TRCS. Annual Report 2016. 1, editor. National Blood Centre: National Blood Centre; 2017. 200 p.

3. Ratchaneewan M. Hemoglobin and Serum Ferritin Determination in Blood Donor. J Med Tech Assoc Thailand. 2012;40(2):12.

4. Boulahriss M BN. Iron Deficiency in Frequent and First Time Female Blood Donors East African Journal of Public Health. 2008;5(3):3.

5. Icek A. The Theory of Planned Behavior. Organizational Behavior and Human Decision Processes. 1991;50:32.

6. Phuphaibul R NC L-CC. Predicting Alcohol Drinking Intention and Behavior of Thai Adolescents. Pacific Rim Int J Nurs Res. 2011;15(1):11.

7. Gunčar G, Kukar M, Notar M, Brvar M, Černelč P, Notar M, et al. An application of machine learning to haematological diagnosis. Scientific Reports. 2018;8(1):411.

8. Kaur P, Singh M, Josan GS. Classification and Prediction Based Data Mining Algorithms to Predict Slow Learners in Education Sector. Procedia Computer Science. 2015;57:500-8.

9. Chisholm A. Exploring Data with RapidMiner. Packt Publishing Ltd.: Packt Publishing Ltd.; 2013.

10. Kotu V, Deshpande B. CHAPTER 2-Data Mining Process. In: Elliot S, Herbert K, editors. Predictive Analytics and Data Mining. Boston: Morgan Kaufmann; 2015. p. 17-36.

11. R. Sathya AA. Comparison of Supervised and Unsupervised Learning Algorithms for Pattern Classification. International Journal of Advanced Research in Artificial Intelligence. 2013;2(2):5.

12. Vijay Kotu BD. Predictive Analytics and Data Mining. Elsevier Inc.: Elsevier Inc.; 2015.

13. Heidrich-Meisner V, Lauer M, Igel C, Riedmiller M. Reinforcement learning in a Nutshell2007. 277-88 p.

14. Cuesta H. Chapter 1, Getting Started, discusses the principles of data analysis and the data analysis process. In: Cuesta H, editor. Practical Data Analysis. 1 ed. Birmingham: Packt Publishing Ltd.; 2013. p. 7-24.

15. Awad M, Khanna R. Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers. Awad M, Khanna R, editors. Berkeley, CA: Apress Media; 2015 2015//. 248 p.

16. Inaba K, Lustenberger T, Rhee P, Holcomb JB, Blackbourne LH, Shulman I, et al. The Impact of Platelet Transfusion in Massively Transfused Trauma Patients. Journal of the American College of Surgeons. 2010;211(5):573-9.

17. Derksen S, Keselman HJ. Backward, forward and stepwise automated subset selection algorithms: Frequency of obtaining authentic and noise variables. British Journal of Mathematical and Statistical Psychology. 1992;45(2):265-82.

18. Haque MM, Rahman A, Hagare D, Chowdhury KR. A Comparative Assessment of Variable Selection Methods in Urban Water Demand Forecasting. Water. 2018;10(4).

19. Guyon I, Andr, #233, Elisseeff. An introduction to variable and feature selection. J Mach Learn Res. 2003;3:1157-82.

20. Sutter JM, Kalivas JH. Comparison of Forward Selection, Backward Elimination, and Generalized Simulated Annealing for Variable Selection. Microchemical Journal. 1993;47(1):60-6.

21. Zhang Z. Variable selection with stepwise and best subset approaches. Annals of Translational Medicine. 2016;4(7):136.

22. Chong I-G, Jun C-H. Performance of some variable selection methods when multicollinearity is present. Chemometrics and Intelligent Laboratory Systems. 2005;78(1):103-12.

23. Song Y-y, Lu Y. Decision tree methods: applications for classification and prediction. Shanghai Archives of Psychiatry. 2015;27(2):130-5.

24. Kotu V, Deshpande B. Introduction. In: Elliot S, Herbert K, editors. Predictive Analytics and Data Mining. Boston: Morgan Kaufmann; 2015. p. 1-16.

25. Himani Sharma SK. A Survey on Decision Tree Algorithms of Classification in Data Mining. International Journal of Science and Research (IJSR). 2016;5(4):4.

26. Thongkam P, Leesutthipornchai P, editors. Ensemble Features Selection Algorithm by Considering Features Ranking Priority. Recent Advances in Information and Communication Technology 2017; 2018 2018//; Cham: Springer International Publishing.

27. Mongkut K, Pipanmaekaporn L. Miss Puangtip Theanseang Master Project Title : Performance Evaluation of Classification Rule Mining Algorithms Major Field : Computer Sciences 2018.

28. Wei-Zhen L, Wang D. Learning machines: Rationale and application in ground-level ozone prediction. Applied Soft Computing. 2014;24:135–41.

29. M.Mohan R, R. K. S, Dinesh CSB, H. C. S, Anil K. Development of Artificial Neural-Network-BasedModels for

the Simulation of Spring Discharge. Advances in Artificial Intelligence. 2011;2011:11.

30. Walczak S. Artificial neural network medical decision support tool: predicting transfusion requirements of ER patients. IEEE Transactions on Information Technology in Biomedicine. 2005;9(3):468-74.

31. Panyasai S FG, Fucharoen S. Hemoglobin Variants in Northern Thailand: Prevalence, Heterogeneity and Molecular Characteristics. Genet Test Mol Biomarkers. 2016;20(1):7.

32. Srivorakun H SK, Fucharoen G, Sanchaisuriya K, Fucharoen S. A large cohort of hemoglobin variants in Thailand: molecular epidemiological study and diagnostic consideration. PLoS One. 2014;9(9):e108365.

33. Hare GMT, Freedman J, David Mazer C. Review article: Risks of anemia and related management strategies: can perioperative blood management improve patient safety? Canadian Journal of Anesthesia/Journal canadien d'anesthésie. 2013;60(2):168-75.

34. S J. The evaluation of blood donor deferrals at Police General Hospital. Bulletin of Chiang Mai Associated Medical Sciences. 2010;43(3):9.

35. วิชุดา กลิ่นหอม วรรณวิมล มีคงและสมรัก เพชรโฉมฉาย. การศึกษาผู้บริจาคโลหิตที่ไม่ผ่านการคัดกรองสุขภาพก่อนบริจาคโลหิตเพื่อมุ่งเน้นการเพิ่มจำ นวนผู้บริจาคโลหิตของภาคบริการโลหิตแห่งชาติที่ 4 จังหวัดราชบุรี. Thai J Hematol Transf Med. 2015;25(3):1.

36. Tanner L, Schreiber M, Low JGH, Ong A, Tolfvenstam T, Lai YL, et al. Decision Tree Algorithms Predict the Diagnosis and Outcome of Dengue Fever in the Early Phase of Illness. PLOS Neglected Tropical Diseases. 2008;2(3):e196.

37. S.Asha Rani SHG. A comparative study of classification algorithm on blood transfusion. International Journal of Advancements in Research & Technology. June 2014;3(6):4.

38. Nasserinejad K, de Kort W, Baart M, T Komárek A, van Rosmalen J, Lesaffre E. Predicting hemoglobin levels in whole blood donors using transition models and mixed effects models. 2013;13:62.

39. Ching Hao Yu MB, Rachel Hogen, Dilin Mao, Atefeh Farzindar and Kiran Dhanireddy. Anemic Status Prediction using Multilayer Perceptron Neural Network Model. GCAI 2017 3rd Global Conference on Artificial Intelligence. 2017;50:8.

40. Ketpupong P, Piromsopa K. Applying Text Mining for Classifying Disease from Symptoms2018. 467-72 p.

41. Vishwakarma PJaSK. Collaborative Analysis of Cancer Patient Data using Rapid Miner. International Journal of Computer Applications. 2016;145(July 2016):6.

42. Sivabalan RJaRV. Analysis of Classification Algorithms for Heart Disease Prediction and its Accuracies. Middle-East Journal of Scientific Research. 2016;24:7.

43. Basharat Naqvi AA, Muhammad Adnan Hashmi and Muhammad Atif. Prediction Techniques for Diagnosis of Diabetic Disease: A Comparative Study. International Journal of Computer Science and Network Security. 2018;18(8):7.

44. Fokkinga J. Modelling hemoglobin levels of blood donors: Erasmus University Rotterdam; 2019.

ภาคผนวก

ภาคผนวก ก

อักษรย่อที่นำมาใช้ในงานวิจัยนี้

ลำดับ	อักษรย่อ	คำเต็มของอักษรย่อ
1	Hb	Hemoglobin
2	ML	Machine Learning
3	CART	Classification And Regression Tree Algorithms
4	SVM	Support Vector Machine
5	ANN	Artificial Neural Networks
6	RM	RapidMiner
7	K-NN	k-Nearest Neighbour
8	CBC	Complete blood count
9	RT-PCR	Real Time – Polymerase Chain Reaction
10	AUC	Area Under the Curve
11	MLP	Multilayer perceptron neural network
12	TPR	True Positive Rate
13	FPR	False Positive Rate
14	PPV	Positive Predictive Value
15	NPV	Negative Predictive Value
16	ROC curve	Receiver Operating Characteristic curve

ภาคผนวก ข

แบบสอบถาม

แบบสอบถามเพื่อการวิจัย

เรื่อง

การประยุกต์ใช้เทคนิคการเรียนรู้ของคอมพิวเตอร์เพื่อพัฒนาตัวแบบทางเหมืองข้อมูล

สำหรับ

พยากรณ์ระดับฮีโมโกลบินของผู้บริจาคโลหิต

คำชี้แจง

1. แบบสอบถามนี้ แบ่งออกเป็น 3 ส่วน ประกอบด้วย
   ส่วนที่ 1 แบบสอบถามข้อมูลทั่วไปของผู้บริจาคโลหิต จำนวน 10 ข้อ
   ส่วนที่ 2 แบบสอบเกี่ยวกับการบริจาคโลหิต จำนวน 12 ข้อ
   ส่วนที่ 3 แบบสอบถามเกี่ยวกับพฤติกรรมสุขภาพ จำนวน 7 ข้อ

2. กรุณาตอบแบบสอบถามให้ครบทุกข้อตามความคิดเห็น และตามความเป็นจริงเพราะคำตอบของท่านจะเป็นประโยชน์เพื่อใช้ในการเพื่อพัฒนาตัวแบบพยากรณ์ด้วยเทคนิคการเรียนรู้ของคอมพิวเตอร์ (machine learning) แบบต่าง ๆ เพื่อนำมาพัฒนาการจัดหาโลหิตให้เพียงพอต่อความต้องการของผู้ป่วย

3. การตอบแบบสอบถามนี้คำตอบของท่านจะถูกเก็บไว้เป็นความลับ และการนำเสนอผลการวิเคราะห์ข้อมูลเป็นการนำเสนอในภาพรวมของผู้เข้าร่วมวิจัย ซึ่งจะไม่มีผลกระทบต่อผู้ตอบแบบสอบถามเป็นรายบุคคลแต่อย่างใด

1. ข้อมูลทั่วไปของผู้บริจาคโลหิต

1. เพศ ☐ ชาย ☐ หญิง

2. กรณีเพศผู้หญิง ท่านอยู่ในช่วงมีประจำเดือนหรือไม่ ☐ มี ☐ ไม่มี

3. อายุ _____ ปี

4. สภานะภาพ ☐ โสด ☐ สมรส ☐ หย่าร้าง/หม้าย
กรณีตอบ สมรส หย่าร้าง/หม้าย กรุณาระบุจำนวนบุตร ____ คน

5. ศาสนา ☐ พุทธ ☐ คริสต์ ☐ อิสลาม ☐ อื่นๆ.....

6. ส่วนสูง _____ ซม. น้ำหนัก _____ กก. น้ำหนักเมื่อ 3 เดือนที่แล้ว _____ กก.
(น้ำหนักที่เปลี่ยนแปลงในรอบปี: ☐ คงที่ ☐ เพิ่ม___กก. ☐ ลด___กก.)

7.การศึกษา ☐ ประถมศึกษาหรือต่ำกว่า ☐ มัธยมศึกษาต้น ☐ มัธยมศึกษาปลาย/ปวช.
☐ อนุปริญญา/ปวส. ☐ ปริญญาตรี ☐ สูงกว่าปริญญาตรี

8.ที่อยู่ปัจจุบัน -<u>เขตพื้นที่</u> ☐ อ.เมือง ☐ ไม่ใช่ อ.เมือง
<u>โปรดระบุเขตเทศบาล</u>
☐ ในเขตเทศบาล ☐ นอกเขตเทศบาล

9. อาชีพ ☐ นักเรียน/นักศึกษา ☐ เกษตรกร ☐ พนักงานเอกชน
☐ ราชการ/รัฐวิสาหกิจ ☐ พระภิกษุ/สามเณร/นักบวช ☐ ค้าขาย/รับจ้างอิสระ
☐ อื่นๆ.....

10.รายได้ต่อเดือน ☐ น้อยกว่า 10,000 ☐ 10,000-20,000 ☐ 20,001-40,000
☐ >40,000

2. ข้อมูลการบริจาคโลหิตของท่าน

1. หมู่โลหิตของท่าน ☐ A Rh+ ☐ A Rh- ☐ B Rh+ ☐ B Rh-
☐ O Rh+ ☐ O Rh- ☐ AB Rh+ ☐ AB Rh- ☐ ไม่ทราบ

2. สถานที่ที่ท่านบริจาคโลหิตประจำ ☐ ภาคบริการโลหิตแห่งชาติ
☐ โรงพยาบาล ☐ หน่วยเคลื่อนที่

3. ท่านเคยบริจาคโลหิตเฉพาะส่วนหรือไม่ ☐ เคย จำนวน____ครั้ง ☐ ไม่เคย
(เช่น บริจาคเฉพาะเกร็ดเลือดหรือเฉพาะน้ำเลือด)

4. จำนวนครั้งที่ท่านบริจาคโลหิตในรอบ 1 ปี ☐ 1 ครั้ง ☐ 2 ครั้ง ☐ 3 ครั้ง ☐ 4 ครั้ง
(เฉพาะการบริจาคแบบโลหิตรวม)

5. ปัจจุบันท่านบริจาคโลหิตแบบโลหิตรวมทั้งหมด _____ ครั้ง
(ไม่รวมการบริจาคโลหิตเฉพาะส่วนเช่น บริจาคเฉพาะเกร็ดเลือดหรือเฉพาะน้ำเลือด)

6. ระยะห่างของการบริจาคแบบโลหิตรวมครั้งสุดท้ายกับครั้งนี้ _____ เดือน

7. ปริมาตรของโลหิตที่บริจาคครั้งสุดท้ายไม่รวมครั้งนี้
☐ บริจาคได้เต็มถุงปกติ ☐ บริจาคได้ไม่เต็มถุง
โปรดระบุขนาดถุง (ถ้าทราบ)..... ☐ ขนาด 350 มม. ☐ ขนาด 450 มม.

8. ความเข้มข้นเลือด: ครั้งนี้ _____ mg/dl. ครั้งที่ผ่านมา (ถ้าจำได้) _____ mg/dl.

9. ความดันโลหิตของท่านในการบริจาคครั้งนี้ _____/____ มม.ปรอท

10. ท่านเคยตรวจความเข้มข้นไม่ผ่านหรือไม่ ☐ ไม่เคย ☐ เคย

11. ท่านมีโรคประจำตัวหรือไม่ ☐ ไม่มี ☐ เบาหวาน ☐ ไขมันในเลือดสูง
☐ ความดันโลหิตสูง ☐ ภาวะ/โรคโลหิตจาง ☐ มี โปรดระบุ.....

12. ท่านได้รับข่าวสารการบริจาคโลหิตทางใดมากที่สุด

☐ วิทยุ ☐ โทรทัศน์ ☐ เฟสบุ๊ค ☐ ไลน์ ☐ SMS ☐ สิ่งพิมพ์ ☐ คนรู้จัก

3. ข้อมูลพฤติกรรมสุขภาพของผู้บริจาคโลหิต

(คำตอบเป็นพฤติกรรมส่วนใหญ่ที่ท่านมักปฏิบัติเป็นประจำ)

1.พฤติกรรมการนอนหลับพักผ่อน

1.1 เวลา ☐ น้อยกว่า 4 ชม. ☐ 4-6 ชม. ☐ มากกว่า 6 ชม

1.2 รูปแบบ ☐ นอนเป็นเวลาแน่นอน ☐ นอนไม่เป็นเวลาแน่นอน

2.พฤติกรรมการออกกำลังกาย ☐ ไม่เคยออกกำลังกาย ☐ น้อยกว่า 2 ครั้ง/สัปดาห์
☐ 3-5 ครั้ง/สัปดาห์ ☐ ทุกวันสม่ำเสมอ

3.พฤติกรรมการสูบบุหรี่ ☐ ไม่สูบ ☐ สูบ

4.พฤติกรรมการดื่มแอลกอฮอล์ ☐ ไม่ดื่ม ☐ น้อยกว่า 2 ครั้ง/สัปดาห์
☐ ดื่ม 3-5 ครั้ง/สัปดาห์ ☐ ดื่มเป็นประจำทุกวัน

5.พฤติกรรมการทานอาหาร ☐ อาหารปกติ ☐ อาหารเจ ☐ อาหารมังสะวิรัต
☐ อาหารมันๆ ☐ อาหาร Diet ลดน้ำหนัก

6.พฤติกรรมการทานยาธาตุเหล็ก ☐ ไม่ทาน ☐ ทานเป็นประจำทุกวัน
☐ ทานไม่สม่ำเสมอ

7.สาเหตุที่ท่านไม่ทานยาธาตุเหล็ก ☐ ไม่ทาน ก็ยังบริจาคได้ ☐ ลืม/ขี้เกียจ
☐ ไม่ชอบกลิ่นและสี ☐ กลัวอ้วน
☐ มีอาการคลื่นไส้/ท้องผูก/ท้องเสีย ☐ อื่นๆโปรดระบุ.....

ภาคผนวก ค
ข้อจำกัด

1. ในการพัฒนาตัวแบบพยากรณ์ Support Vector Machine ด้วยโปรแกรม Rapid Miner หากตัวแปรพยากรณ์ประเภท Nominal ใดๆ มีค่าส่วนใหญ่เป็นค่าใดค่าหนึ่งมากเกินไป จะทำให้ตัวแบบทำงานช้ากว่าปกติ เช่น จากการศึกษานี้ตัวแปร Rh ข้อมูลเกือบทั้งหมดเป็น Rh positive มีเพียง 19 รายที่เป็น Rh Negative จากทั้งหมด 2180 ราย ในการศึกษานี้จึงต้องนำตัวแปรพยากรณ์ Rh ออกจากการตัวแบบพยากรณ์ Support Vector Machine

2. การศึกษานี้ข้อมูลสำคัญเช่น ค่า Hb ครั้งที่ผ่านมา นั้นส่วนใหญ่ได้จากแบบสอบถามซึ่งผู้บริจาคโลหิตบางรายอาจจำผิดพลาด ซึ่งส่งผลกระทบต่อผลการศึกษาและหลายรายยังเป็นการตรวจด้วยสาร CuSo4 ซึ่งไม่สามารถระบุค่าได้ชัดเจน

ประวัติผู้เขียน

ชื่อ-สกุล	สาธิต เทศสมบูรณ์
วัน เดือน ปี เกิด	19 กันยายน 2524
สถานที่เกิด	จังหวัดชัยนาท
วุฒิการศึกษา	วิทยาศาสตรบัณฑิต (เทตนิคการแพทย์) คณะสหเวชศาสตร์ มหาวิทยาลัยนเรศวร
ที่อยู่ปัจจุบัน	262/3 ม.8 ต.เขากะลา อ.พยุหะคีรี จ.นครสวรรค์ 60130
ผลงานตีพิมพ์	1. นำเสนอในรูปแบบโปสเตอร์เรื่อง “ความชุกของการติดเชื้อซิฟิลิส เอชไอวี ไวรัสตับอักเสบ บี และซี ในผู้บริจาคลิตครั้งแรก ที่ส่งตรวจ ณ ภาคบริการโลหิตแห่งชาติที่ 8 จ.นครสวรรค์ ระหว่างปี พ.ศ. 2550-2554” การประชุมวิชาการงานบริการโลหิตระดับชาติ ครั้งที่ 21 ประจำปี 2556 ศูนย์บริการโลหิตแห่งชาติ สภากาชาดไทย 2 ผลงานวิจัยเรื่อง “ความชุกของการติดเชื้อซิฟิลิส เอชไอวี ไวรัสตับอักเสบ บี และซี ในผู้บริจาคโลหิตครั้งแรก ที่ส่งตรวจ ณ ภาคบริการโลหิตแห่งชาติที่ 8 จ.นครสวรรค์ ระหว่างปี พ.ศ. 2550-2554” วารสารโลหิตวิทยาและเวชศาสตร์บริการโลหิต 2556;23(3):187-193 3 นำเสนอในรูปแบบโปสเตอร์เรื่อง “การประเมินประสิทธิภาพการตรวจหมู่โลหิต ABO, Rh(D) และ Irregular Antibody Screening ด้วยเครื่อง Qwalys3” การประชุมวิชาการงานบริการโลหิตระดับชาติ ครั้งที่ 23 ประจำปี 2558 ศูนย์บริการโลหิตแห่งชาติ สภากาชาดไทย 4 นำเสนอในรูปแบบโปสเตอร์เรื่อง “การบริหารปริมาณโลหิตและส่วนประกอบโลหิตระหว่างภาคบริการโลหิตแห่งชาติด้วย Google spreadsheet” การประชุมวิชาการงานบริการโลหิตระดับชาติ ครั้งที่ 23 ประจำปี 2558 ศูนย์บริการโลหิตแห่งชาติ สภากาชาดไทย 5 ผลงานวิจัยเรื่อง ปริมาณโลหิตที่เพียงพอและปริมาณโลหิตสำรองที่เหมาะสม พุทธชินราชเวชสาร ปีที่ : 29 ฉบับที่ : 1 เลขหน้า : 54-64 ปี พ.ศ. : 2555 6 การพัฒนาการจัดหาโลหิตของจังหวัดนครสวรรค์ ระหว่างปี พ.ศ. 2556-2558 วารสารโลหิตวิทยาและเวชศาสตร์บริการโลหิต J Hematol

Transfus Med Vol. 26 No. 4 October-December 2016

7 นำเสนอในรูปแบบโปสเตอร์เรื่อง “การศึกษาคุณภาพโลหิตและส่วนประกอบโลหิตของภาคบริการโลหิตแห่งชาติที่ 8 จังหวัดนครสวรรค์” การประชุมวิชาการงานบริการโลหิตระดับชาติ ครั้งที่ 27 ประจำปี 2562 ศูนย์บริการโลหิตแห่งชาติ สภากาชาดไทย

8 นำเสนอในรูปแบบโปสเตอร์เรื่อง “การเตรียมเซลล์เม็ดเลือดแดงสำหรับตรวจหมู่โลหิตด้วยเครื่องตรวจวิเคราะห์หมู่โลหิตอัตโนมัติ Qwalys 3” การประชุมวิชาการงานบริการโลหิตระดับชาติ ครั้งที่ 27 ประจำปี 2562 ศูนย์บริการโลหิตแห่งชาติ สภากาชาดไทย

9 นำเสนอในรูปแบบบรรยาย “การประยุกต์ใช้โครงข่ายประสาทเทียมพยากรณ์การจ่ายโลหิตที่เหมาะสมล่วงหน้า 3 วัน” การประชุมวิชาการงานบริการโลหิตระดับชาติ ครั้งที่ 27 ประจำปี 2562 ศูนย์บริการโลหิตแห่งชาติ สภากาชาดไทย

10 นำเสนอในรูปแบบโปสเตอร์เรื่อง “การเปรียบเทียบประสิทธิภาพการจำแนกระดับฮีโมโกลบินของผู้บริจาคโลหิตด้วยเทคนิคการเรียนรู้ของคอมพิวเตอร์” งานประชุมวิชาการระดับชาติมหาวิทยาลัยรังสิต ประจำปี 2562

รางวัลที่ได้รับ บุคลากรดีเด่นระดับองค์ สภากาชาดไทย ประจำปี 2561