



สถาบันวิจัยระบบสาธารณสุข

รายงานฉบับสมบูรณ์
การพัฒนาแบบจำลองทำนายความเสี่ยงและภาวะแทรกซ้อนจากโรคเบาหวานและ
ความดันโลหิตสูงในประเทศไทยเพื่อประเมินผลกระทบทางเศรษฐกิจต่อระบบ
หลักประกันสุขภาพและระบบสุขภาพ

Development of Predictive Risk Models for Diabetes and
Hypertension in Thailand to Evaluate Economic Implications on Thai
Health Systems and Health Financing

หัวหน้าโครงการ

ดร.นพ.ปิยะ หาญวรวงศ์ชัย Dr. Piya Hanvoravongchai

ผู้ร่วมงานวิจัย

ศ.ดร.เสกสรร เกียรติสุไพบูลย์

Prof. Dr. Seksan Kiatsupaibul

ผศ.ดร.วิฐุรา พึ่งพาพงศ์

Asst.Prof. Dr. Vitara Pungpapong

รศ.ดร.ชญ.อรลักษณ์ พัฒนาประทีป

Assoc.Prof. Dr. Oraluck Pattanaprteep

รศ.ดร.ชญ.มนตร์ธมภ์ ถาวรเจริญทรัพย์

Assoc.Prof. Dr Montarat Thavorncharoensap

นพ.วศิน เลหาวินิจ

Dr. Wasin Laohavinij

พญ.จิตาภา หาญวรวงศ์ชัย

Dr. Jidapa Hanvoravongchai

มูลนิธิสาธารณสุขแห่งชาติ (มสช.)

โครงการนี้ได้รับทุนอุดหนุนจากสถาบันวิจัยระบบสาธารณสุข (สวรส.) ความคิดเห็นและ
ข้อเสนอแนะที่ปรากฏในเอกสารนี้เป็นของผู้วิจัย มิใช่ความเห็นของ
สถาบันวิจัยระบบสาธารณสุข

ตุลาคม 2566

รายงานฉบับสมบูรณ์การพัฒนาแบบจำลองทำนายความเสี่ยงและภาวะแทรกซ้อนจากโรคเบาหวานและความ
ดันโลหิตสูงในประเทศไทยเพื่อประเมินผลกระทบทางเศรษฐกิจต่อระบบหลักประกันสุขภาพ
และระบบสุขภาพ

Development of Predictive Risk Models for Diabetes and Hypertension in Thailand to
Evaluate Economic Implications on Thai Health Systems and Health Financing

หัวหน้าโครงการ

ดร.นพ.ปิยะ หาญวรวงศ์ชัย Dr. Piya Hanvoravongchai

ผู้ร่วมงานวิจัย

ศ.ดร.เสกสรร เกียรติสุไพบูลย์

Prof. Dr. Seksan Kiatsupaibul

ผศ.ดร.วิฐุรา พึ่งพาพงศ์

Asst.Prof. Dr. Vitara Pungpapong

รศ.ดร.ชญ.อรลักษณ์ พัฒนาประทีป

Assoc.Prof. Dr. Oraluck Pattanaprteep

รศ.ดร.ชญ.มนตร์ตม์ ถาวรเจริญทรัพย์

Assoc.Prof. Dr. Montarat Thavorncharoensap

นพ.วศิน เลาหวินิจ

Dr. Wasin Laohavinij

พญ.จิตาภา หาญวรวงศ์ชัย

Dr. Jidapa Hanvoravongchai

บทคัดย่อ

โครงการนี้มีวัตถุประสงค์เพื่อ (1) สร้างแบบจำลองทำนายความเสี่ยง 2 แบบจำลองหลัก คือ 1.1. แบบจำลองทำนายความเสี่ยงการเกิดโรคเบาหวานและความดันโลหิตสูงในประชากรทั่วไป และ 1.2. แบบจำลองทำนายความเสี่ยงการเกิดภาวะแทรกซ้อนและการเสียชีวิตอันเนื่องมาจากโรคเบาหวานและความดันโลหิตสูง และ (2) ประเมินการภาระรายจ่ายทางสุขภาพและความสูญเสียด้านผลิตภาพการทำงานอันเนื่องมาจากโรคเบาหวานและความดันโลหิตสูงในปัจจุบัน โดยเป็นการศึกษาเชิงปริมาณ และใช้ข้อมูลการใช้บริการสาธารณสุขผู้ป่วยนอกและผู้ป่วยในจากฐานข้อมูล 43 แห่ง กระทรวงสาธารณสุข กลุ่มตัวอย่างของการศึกษานี้มีจำนวนรวมทั้งสิ้น 1.2 ล้านคน ถูกติดตามเป็นระยะเวลา 6 ปี 2 เดือน ระหว่างวันที่ 1 มกราคม พ.ศ. 2560 ถึง 28 กุมภาพันธ์ พ.ศ.2566 สถิติหลักที่ใช้การสร้างแบบจำลอง คือ การวิเคราะห์การรอดชีพตัวแปรเวลาแบบไม่ต่อเนื่องและใช้วิธีการเรียนรู้ของเครื่องในการจำแนกประเภทข้อมูล 2 วิธี ได้แก่ 1. การถดถอยโลจิสติกแบบ Elastic Net และ 2. วิธี XGBoost (Extreme Gradient Boosting) เพื่อค้นหาแบบจำลองที่มีความแม่นยำสูงสุดประเมินจากพื้นที่ใต้กราฟ หรือ ค่า AUC (Area Under ROC Curve)

ผลการศึกษา พบว่า วิธี XGBoost มีประสิทธิภาพในการทำนายสูงกว่าการถดถอยโลจิสติกแบบ Elastic Net ในทุกประเภทของแบบจำลอง โดย ประสิทธิภาพการทำนายของแบบจำลองทำนายความเสี่ยงการเกิดโรคเบาหวานและการเกิดโรคความดันโลหิตสูงด้วยวิธี XGBoost ในกลุ่มประชากรทั่วไปอยู่ในเกณฑ์ดี มีค่า AUC อยู่ระหว่าง 0.86-0.88 และ 0.84-0.87 ตามลำดับ อย่างไรก็ตาม ประสิทธิภาพการทำนายของแบบจำลองทำนายความเสี่ยงการเกิดภาวะแทรกซ้อนในกลุ่มผู้ป่วยโรคเบาหวานและโรคความดันโลหิตสูงให้ผลลัพธ์แตกต่างกันขึ้นอยู่กับประเภทของภาวะแทรกซ้อนโดยมีช่วงระหว่าง 0.63-0.84 ซึ่งการทำนายเกิดภาวะไตเสื่อมเรื้อรังมีค่า AUC ดีที่สุดอยู่ที่ 0.82-0.84 และ 0.80-0.82 ในผู้ป่วยโรคเบาหวานและผู้ป่วยโรคความดันโลหิตสูง ตามลำดับ สำหรับผลการศึกษาที่เกี่ยวข้องกับการใช้บริการสุขภาพของผู้ป่วยโรคเบาหวานและความดันโลหิตสูง พบว่า ค่าใช้จ่ายทางสุขภาพทั้งหมด ประกอบด้วย ค่ารักษาพยาบาล ค่าเดินทาง ค่าอาหาร ค่าเสียโอกาสของผู้ป่วยเนื่องจากต้องมารับการรักษาที่โรงพยาบาล และ ค่าเสียโอกาสของผู้ดูแลของผู้ที่ไม่เป็นทั้งโรคเบาหวานและความดันโลหิตสูง ผู้ที่เป็นโรคเบาหวานอย่างเดียว ผู้ที่เป็นโรคความดันโลหิตสูงอย่างเดียว และ ผู้ที่เป็นทั้ง 2 โรคอยู่ที่ 6,114 25,322 18,124 และ 37,575 บาทต่อคนต่อปี ตามลำดับ โดยปัจจัยที่ส่งผลให้ค่าใช้จ่ายทางสุขภาพของผู้ป่วยโรคเบาหวานอย่างเดียวและโรคความดันโลหิตสูงอย่างเดียวแตกต่างกัน คือ ค่าเสียโอกาสของผู้ดูแล

กระทรวงสาธารณสุข สามารถนำแบบจำลองทำนายความเสี่ยงการเกิดโรคเบาหวานและการเกิดโรคความดันโลหิตสูงในประชากรทั่วไปที่พัฒนาขึ้นไปประยุกต์ใช้ในการคัดกรองโรคเบาหวานและความดันโลหิตสูงในประชาชนอายุ 35 ปีขึ้นไปของประเทศเพื่อเพิ่มประสิทธิภาพการประเมินความเสี่ยงแทนที่การใช้เพียงผลการตรวจระดับน้ำตาลปลายนิ้วและการวัดความดันโลหิตเพียงอย่างเดียว และควรสร้างเกณฑ์การจัดกลุ่มผู้ป่วยตามค่าความเสี่ยงและข้อมูลที่จะใช้ในการสื่อสารความเสี่ยงทางสุขภาพ เช่น การแบ่งกลุ่มผู้ป่วยตามเกณฑ์ เพื่อให้หน่วยงานในพื้นที่สามารถสื่อสารกับผู้ที่มีการคัดกรองโรคได้อย่างมีประสิทธิภาพก่อนนำแบบจำลองไปใช้จริงในพื้นที่

Abstract

The primary objectives of this study are as follows: (1) Develop two primary risk prediction models: 1.1. for determining the likelihood of developing diabetes and hypertension in the normal population; and 1.2. for determining the risk of developing diabetes and hypertension-related complications and mortality. In addition, (2) estimates the health expenses and productivity loss due to diabetes and hypertension.

This is a quantitative study utilizing Ministry of Public Health outpatient and inpatient data. The study consisted of 1.2 million samples and was followed over a period of 6 years, from January 1st, 2017, to February 28th, 2023. The primary statistic employed in model development is discrete-time survival analysis. For data classification, two machine learning techniques, namely Lasso Logistic Regression and XGBoost (Extreme Gradient Boosting), were utilized. Additionally, model accuracy was assessed using the area under the Receiver Operating Characteristic (ROC) Curve.

The study found that the XGBoost method had higher prediction efficiency than Lasso Logistic Regression in all models. The diabetes and hypertension risk models for the general population that used the XGBoost method were accurate, with the area under the curve (AUC) values ranging from 0.86 to 0.88 for diabetes and 0.84 to 0.87 for hypertension. However, the predictive performance for the risk of complications in diabetic and hypertensive patients varied depending on the type of complications, with the AUC ranging from 0.63 to 0.84. Predicting chronic kidney disease yielded the best AUC, ranging from 0.82 to 0.84 and 0.80 to 0.82 in diabetic and hypertensive patients, respectively.

Additionally, the present study examined the comprehensive health expenditure of individuals diagnosed with diabetes and hypertension. This expenditure comprised direct medical costs, direct non-medical costs (including food and travel), and opportunity costs incurred by both patients and their caregivers. The annual total expenditures per patient vary as follows: 6,114 baht, 25,322 baht, 18,124 baht, and 37,575 baht for patients without both diabetes and hypertension, patients with only diabetes, patients with only hypertension, and patients with both conditions. Caregiver opportunity cost is the determining factor for health expenditure differences between patients with diabetes and hypertension alone.

It is recommended that the Ministry of Public Health incorporate the diabetes and hypertension risk models into nationwide diabetes and hypertension screening programs for individuals aged 35 years and older. These models offer a more comprehensive assessment of the risks involved in diabetes and hypertension than point-of-care glucose tests and blood pressure measurements alone. In addition, prior to implementing the risk model, criteria should be established to categorize patients according to their risk, and health

recommendations should be tailored to each risk group. This will enable local agencies to efficiently disseminate information to the public.

บทสรุปของผู้บริหาร

โครงการพัฒนาแบบจำลองทำนายความเสี่ยงและภาวะแทรกซ้อนจากโรคเบาหวานและความดันโลหิตสูงในประเทศไทยเพื่อประเมินผลกระทบทางเศรษฐกิจต่อระบบหลักประกันสุขภาพและระบบสุขภาพ มีวัตถุประสงค์เพื่อ (1) สร้างแบบจำลองทำนายความเสี่ยง 2 แบบจำลองหลักตามลักษณะประชากรที่สำคัญของประเทศ แบ่งออกเป็น 1.1. แบบจำลองทำนายความเสี่ยงการเกิดโรคเบาหวานและความดันโลหิตสูงในประชากรทั่วไป และ 1.2. แบบจำลองทำนายความเสี่ยงการเกิดภาวะแทรกซ้อนและการเสียชีวิตอันเนื่องมาจากโรคเบาหวานและความดันโลหิตสูง โดยกำหนดภาวะแทรกซ้อนเป้าหมาย 4 กลุ่มหลัก ได้แก่ 1. โรคหลอดเลือดหัวใจ 2. โรคเส้นเลือดสมองแตกและตีบ 3. โรคไตวายเรื้อรังที่ต้องได้รับการบำบัดทดแทนไต และ 4. โรคแทรกซ้อนทางตา และ (2) ประเมินการกระจายจ่ายทางสุขภาพและความสูญเสียด้านผลิตภาพการทำงานอันเนื่องมาจากโรคเบาหวานและความดันโลหิตสูงในปัจจุบัน

การศึกษานี้เป็นการศึกษาเชิงปริมาณ (Quantitative study) โดยใช้ข้อมูลการใช้บริการสาธารณสุขผู้ป่วยนอกและผู้ป่วยในจากฐานข้อมูล 43 แฟ้ม กระทรวงสาธารณสุข กลุ่มตัวอย่างของการศึกษานี้มีจำนวนรวมทั้งสิ้น 1.2 ล้านคน ถูกติดตามเป็นระยะเวลา 6 ปี 2 เดือน ระหว่างวันที่ 1 มกราคม พ.ศ. 2560 ถึง 28 กุมภาพันธ์ พ.ศ. 2566 สำหรับการสร้างแบบจำลองทำนายความเสี่ยง ผู้วิจัยแบ่งกลุ่มตัวอย่างออกเป็น 2 กลุ่ม คือ ข้อมูลชุดฝึกฝน (Training Set) และ ข้อมูลชุดทดสอบ (Test Set) ที่อัตราส่วน 70 ต่อ 30 โดยผู้ป่วย 1 คนจะอยู่ได้เพียงชุดข้อมูลเพียงชุดเดียว สำหรับสถิติที่เลือกใช้ในการวิเคราะห์ คือ การวิเคราะห์การรอดชีพตัวแปรเวลาแบบไม่ต่อเนื่อง (Discrete-time Survival Analysis) โดยใช้วิธีการเรียนรู้ของเครื่องในการจำแนกประเภทข้อมูล 2 วิธี ได้แก่ 1. การถดถอยโลจิสติกแบบ Elastic Net (Lasso Logistic Regression) และ 2. วิธี XGBoost (Extreme Gradient Boosting) เพื่อค้นหาแบบจำลองที่มีความแม่นยำสูงสุดประเมินจากพื้นที่ใต้กราฟ หรือ ค่า AUC (Area Under ROC Curve)

ผลการศึกษาคำสำคัญที่เกี่ยวข้องกับแบบจำลองทำนายความเสี่ยง พบว่า

- 1) จากการตรวจสอบข้อมูลในฐานข้อมูล 43 แฟ้ม พบปัญหา 2 กรณี ได้แก่ (1) ปัญหาค่านอกเกณฑ์ (Outlier) หรือค่าที่มีความผิดปกติหรือแตกต่างไปจากค่าสังเกตส่วนใหญ่ และ (2) ปัญหาข้อมูลสูญหาย (Missing Data) ดังนั้น จึงต้องมีการจัดการทำความสะอาดข้อมูลดังกล่าวก่อนนำไปวิเคราะห์ ซึ่งรายละเอียดการจัดการและทำความสะอาดข้อมูลแต่ละประเภทมีความแตกต่างกันขึ้นอยู่กับชนิดของข้อมูลและวรรณกรรมอื่น ๆ ในอดีต
- 2) วิธี XGBoost (Extreme Gradient Boosting) มีประสิทธิภาพในการทำนายได้ดีกว่าการถดถอยโลจิสติกแบบ Elastic Net (Lasso Logistic Regression) ในแบบจำลองทำนายความเสี่ยงทุกประเภท
- 3) แบบจำลองทำนายความเสี่ยงการเกิดโรคเบาหวานและโรคความดันโลหิตสูงในประชากรทั่วไป ซึ่งมีปัจจัยตัวแปรทำนายมีความซับซ้อนน้อยกว่าและมีอุบัติการณ์ของโรคสูงกว่าเมื่อเทียบกับแบบจำลองทำนายการเกิดภาวะแทรกซ้อนในผู้ป่วยโรคเบาหวานและความดันโลหิตสูง ส่งผลให้ประสิทธิภาพการทำนายของแบบจำลองการเกิดโรคเบาหวานและความดันโลหิตสูงในกลุ่มประชากรทั่วไปมีประสิทธิภาพสูงกว่าแบบจำลองทำนายความเสี่ยงการเกิดภาวะแทรกซ้อน
- 4) ประสิทธิภาพการทำนายของแบบจำลองทำนายความเสี่ยงการเกิดโรคเบาหวานด้วยวิธี XGBoost มีค่า AUC อยู่ระหว่าง 0.86-0.88 ซึ่งใกล้เคียงกับแบบจำลองทำนายความเสี่ยงการเกิดโรคเบาหวานมาตรฐานของ

ต่างประเทศที่มีค่า AUC อยู่ที่ 0.76-0.88

- 5) ประสิทธิภาพการทำนายของแบบจำลองทำนายความเสี่ยงการเกิดโรคความดันโลหิตสูงด้วยวิธี XGBoost ให้ผลไปในทิศทางเดียวกันโดยมีค่า AUC อยู่ระหว่าง 0.84-0.87 ซึ่งใกล้เคียงกับแบบจำลองทำนายความเสี่ยงการเกิดโรคความดันโลหิตสูงมาตรฐานของต่างประเทศที่มีประสิทธิภาพการทำนายเฉลี่ย (pooled AUC) อยู่ที่ 0.75 (0.68-0.80)
- 6) ประสิทธิภาพการทำนายของแบบจำลองทำนายความเสี่ยงการเกิดภาวะแทรกซ้อนในผู้ป่วยโรคเบาหวานด้วยวิธี XGBoost ให้ผลลัพธ์แตกต่างกันขึ้นอยู่กับประเภทของภาวะแทรกซ้อนโดยมีค่า AUC อยู่ระหว่าง 0.63-0.84
- 7) ประสิทธิภาพการทำนายของแบบจำลองทำนายความเสี่ยงการเกิดภาวะแทรกซ้อนในผู้ป่วยโรคความดันโลหิตสูงด้วยวิธี XGBoost ให้ผลลัพธ์ในทิศทางเดียวกันโดยมีประสิทธิภาพการทำนายขึ้นอยู่กับประเภทของภาวะแทรกซ้อนและมีค่า AUC อยู่ระหว่าง 0.63-0.84

ผลการศึกษาคำคัญที่เกี่ยวข้องกับการใช้บริการสุขภาพของผู้ป่วยโรคเบาหวานและความดันโลหิตสูงพบว่า

- 8) เมื่อแบ่งผู้ป่วยออกเป็น 4 กลุ่ม คือ 1. ผู้ที่ไม่เป็นทั้งโรคเบาหวานและความดันโลหิตสูง 2. ผู้ที่เป็นโรคเบาหวานอย่างเดียว 3. ผู้ที่เป็นโรคความดันโลหิตสูงอย่างเดียว และ 4. ผู้ที่เป็นทั้ง 2 โรค พบว่า จำนวนการให้บริการและค่าใช้จ่ายทางสุขภาพเพิ่มขึ้นตามจำนวนโรคของผู้ป่วยโดยผู้ที่เป็นโรคเบาหวานอย่างเดียวมีการให้บริการและค่าใช้จ่ายทางสุขภาพใกล้เคียงกับผู้ที่เป็นโรคความดันโลหิตสูงอย่างเดียว และผู้ที่เป็นทั้ง 2 โรคจะมีการใช้ทรัพยากรสูงสุด ซึ่งลักษณะดังกล่าวพบทั้งในการใช้บริการผู้ป่วยนอกและผู้ป่วยใน
- 9) เมื่อแบ่งผู้ป่วยเพิ่มเติมแยกตามจำนวนภาวะแทรกซ้อนพบผลสำรวจในทิศทางเดียวกัน คือ จำนวนการให้บริการและค่าใช้จ่ายทางสุขภาพเพิ่มขึ้นตามจำนวนภาวะแทรกซ้อนของผู้ป่วยทั้งในการใช้บริการผู้ป่วยนอกและผู้ป่วยใน
- 10) เมื่อประเมินค่าใช้จ่ายทางสุขภาพทั้งหมด ประกอบด้วย ค่ารักษาพยาบาล ค่าเดินทาง ค่าอาหาร ค่าเสียโอกาสของผู้ป่วยเนื่องจากต้องมารับการรักษาที่โรงพยาบาล และ ค่าเสียโอกาสของผู้ดูแล พบว่า ค่าใช้จ่ายทางสุขภาพต่อของผู้ที่ไม่เป็นทั้งโรคเบาหวานและความดันโลหิตสูง ผู้ที่เป็นโรคเบาหวานอย่างเดียว ผู้ที่เป็นโรคความดันโลหิตสูงอย่างเดียว และ ผู้ที่เป็นทั้ง 2 โรคอยู่ที่ 6,114 25,322 18,124 และ 37,575 บาทต่อคนต่อปีตามลำดับ
- 11) สำหรับสาเหตุที่ส่งผลให้ค่าใช้จ่ายทางสุขภาพของผู้ป่วยโรคเบาหวานอย่างเดียวและผู้ป่วยโรคความดันโลหิตสูงอย่างเดียวแตกต่างกัน คือ ค่าเสียโอกาสของผู้ดูแลซึ่งได้จากการทบทวนวรรณกรรม โดย ค่าเสียโอกาสของผู้ดูแลของผู้ป่วยโรคเบาหวานและโรคความดันโลหิตสูงอยู่ที่ 8,790 และ 1,326 บาทต่อคนต่อปี ตามลำดับ ซึ่งหากไม่นำค่าใช้จ่ายหมวดนี้มาวิเคราะห์ พบว่า ค่าใช้จ่ายทางสุขภาพของผู้ป่วยโรคเบาหวานอย่างเดียวและผู้ป่วยโรคความดันโลหิตสูงอย่างเดียวอยู่ที่ 16,531 และ 16,798 บาทต่อคนต่อปี

ข้อเสนอแนะเชิงนโยบาย

- 1) กระทรวงสาธารณสุข สามารถนำแบบจำลองทำนายความเสี่ยงการเกิดโรคเบาหวานและการเกิดโรคความดันโลหิตสูงในประชากรทั่วไปไปประยุกต์ใช้ในการคัดกรองโรคเบาหวานและความดันโลหิตสูงในประชาชนอายุ 35 ปีขึ้นไปของประเทศ เพื่อเพิ่มประสิทธิภาพการประเมินความเสี่ยงแทนที่การใช้เพียงผลการตรวจระดับน้ำตาลปลายนิ้วและการวัดความดันโลหิตเพียงอย่างเดียว
- 2) กระทรวงสาธารณสุขและหน่วยงานที่เกี่ยวข้อง ควรร่วมกันสร้างเกณฑ์การจัดกลุ่มผู้ป่วยตามค่าความเสี่ยงและข้อมูลที่จะใช้ในการสื่อสารความเสี่ยงทางสุขภาพ เช่น การแบ่งกลุ่มผู้ป่วยตามเฉดสี เพื่อให้หน่วยงานในพื้นที่สามารถสื่อสารกับผู้ที่มีการคัดกรองโรคได้อย่างมีประสิทธิภาพก่อนนำแบบจำลองไปใช้จริงในพื้นที่
- 3) กระทรวงสาธารณสุข ควรสนับสนุนหรือเป็นผู้นำในการพัฒนาแบบจำลองทำนายความเสี่ยงอย่างต่อเนื่อง โดยการติดตามข้อมูลผลลัพธ์ทางสุขภาพระยะยาวจะช่วยให้แบบจำลองทำนายความเสี่ยงมีความแม่นยำมากขึ้น โดยเฉพาะแบบจำลองทำนายความเสี่ยงการเกิดภาวะแทรกซ้อน
- 4) กระทรวงสาธารณสุขและสำนักงานหลักประกันสุขภาพแห่งชาติสามารถประยุกต์ใช้แบบจำลองทำนายความเสี่ยงในการประเมินผลกระทบทางสุขภาพอันเนื่องมาจากการลงทุนทางด้านโครงสร้างเสริมป้องกันโรค โดยสามารถเปรียบเทียบงบประมาณการลงทุนกับค่าใช้จ่ายทางสุขภาพที่เปลี่ยนแปลงไปอันเนื่องมาจากโอกาสเกิดโรคลดลงของกลุ่มเป้าหมาย

สารบัญ

| | หน้าที่ |
|---|---------|
| บทคัดย่อ | ก-ค |
| บทสรุปของผู้บริหาร | ง-ฉ |
| บทที่ 1 บทนำ | 1-5 |
| บทที่ 2 การทบทวนวรรณกรรม | 6-14 |
| บทที่ 3 ระเบียบวิธีวิจัย | |
| ระยะที่ 1 การทบทวนวรรณกรรมเพื่อคัดเลือกตัวแปรสำคัญสำหรับการสร้างแบบจำลองทำนายความเสี่ยง | 15-16 |
| ระยะที่ 2 การวิเคราะห์ทางระบาดวิทยาและการสร้างแบบจำลองทำนายความเสี่ยง | 16-27 |
| ระยะที่ 3 การประเมินผลกระทบทางเศรษฐกิจอันเนื่องมาจากโรคเบาหวานและความดันโลหิตสูง | 27-28 |
| บทที่ 4 ผลการศึกษา | |
| แบบจำลองที่ 1 ทำนายความเสี่ยงการเกิดโรคเบาหวานในประชากรทั่วไปจาก Cohort ของผู้ที่ยังไม่เป็นโรคเบาหวาน (Cohort 1.1) | 29-59 |
| แบบจำลองที่ 2 ทำนายความเสี่ยงการเกิดโรคโรคความดันโลหิตสูงในประชากรทั่วไปจาก Cohort ของผู้ที่ยังไม่เป็นโรคความดันโลหิตสูง (Cohort 2.1) | 59-84 |
| แบบจำลองที่ 3 ทำนายความเสี่ยงการเกิดภาวะแทรกซ้อนและการเสียชีวิตจาก Cohort ของผู้ที่เป็นโรคเบาหวาน (Cohort 1.2) | 84-153 |
| แบบจำลองที่ 4 ทำนายความเสี่ยงการเกิดภาวะแทรกซ้อนและการเสียชีวิตจาก Cohort ของผู้ที่เป็นโรคความดันโลหิตสูง (Cohort 2.2) | 153-206 |
| ผลกระทบทางเศรษฐกิจของโรคเบาหวานและความดันโลหิตสูง | 207-221 |
| บทที่ 5 อภิปรายและสรุปผลการศึกษา | 222-228 |
| บรรณานุกรม | |
| ภาคผนวก | ผ1-6 |

สารบัญตาราง

| | | หน้าที่ |
|-------------|---|---------|
| ตารางที่ 1 | เปรียบเทียบแบบจำลองทำนายความเสี่ยงของประเทศไทยเทียบกับแบบจำลองทำนายความเสี่ยงที่ทางผู้วิจัยพัฒนาขึ้น | 10-12 |
| ตารางที่ 2 | แสดงรายการเพิ่มข้อมูลของฐานข้อมูล 43 แฟ้มที่นำมาใช้ในการสร้างแบบจำลองทำนายความเสี่ยง | 15-16 |
| ตารางที่ 3 | เมทริกซ์สับสน | 25 |
| ตารางที่ 4 | ตัวแปรที่นำมาพิจารณาในการสร้างตัวแบบทำนายความเสี่ยงของการเกิดโรคเบาหวาน | 32-35 |
| ตารางที่ 5 | ค่าเฉลี่ยเลขคณิต มัธยฐาน ส่วนเบี่ยงเบนมาตรฐาน ค่าต่ำสุด ค่าสูงสุดของอายุผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคเบาหวานเมื่อเริ่มต้น Cohort เท่านั้น | 38 |
| ตารางที่ 6 | ค่าเฉลี่ยเลขคณิต มัธยฐาน ส่วนเบี่ยงเบนมาตรฐาน ค่าต่ำสุด ค่าสูงสุดของส่วนสูงผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคเบาหวานเมื่อเริ่มต้น Cohort เท่านั้น | 40 |
| ตารางที่ 7 | ค่าเฉลี่ยเลขคณิต มัธยฐาน ส่วนเบี่ยงเบนมาตรฐาน ค่าต่ำสุด ค่าสูงสุดของดัชนีมวลกาย (BMI) ผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคเบาหวานเมื่อเริ่มต้น Cohort เท่านั้น | 40 |
| ตารางที่ 8 | ค่าเฉลี่ยเลขคณิต มัธยฐาน ส่วนเบี่ยงเบนมาตรฐาน ค่าต่ำสุด ค่าสูงสุด ขนาดเส้นรอบเอวของผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคเบาหวานเมื่อเริ่มต้น Cohort เท่านั้น | 42 |
| ตารางที่ 9 | ค่าเฉลี่ยเลขคณิต มัธยฐาน ส่วนเบี่ยงเบนมาตรฐาน ค่าต่ำสุด ค่าสูงสุดของความดันช่วงหัวใจบีบตัว สำหรับผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคเบาหวานเมื่อเริ่มต้น Cohort เท่านั้น | 43 |
| ตารางที่ 10 | ค่าเฉลี่ยเลขคณิต มัธยฐาน ส่วนเบี่ยงเบนมาตรฐาน ค่าต่ำสุด ค่าสูงสุดของความดันช่วงหัวใจคลายตัว สำหรับผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคเบาหวานเมื่อเริ่มต้น Cohort เท่านั้น | 43 |
| ตารางที่ 11 | ค่าเฉลี่ยเลขคณิต มัธยฐาน ส่วนเบี่ยงเบนมาตรฐาน ค่าต่ำสุด ค่าสูงสุดของระดับน้ำตาลในเลือด ผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคเบาหวานเมื่อเริ่มต้น Cohort เท่านั้น | 46 |
| ตารางที่ 12 | ค่าเฉลี่ยเลขคณิต มัธยฐาน ส่วนเบี่ยงเบนมาตรฐาน ค่าต่ำสุด ค่าสูงสุดของระดับไขมันดี สำหรับผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคเบาหวานเมื่อเริ่มต้น Cohort เท่านั้น | 49 |

| | หน้าที่ | |
|-------------|---|-------|
| ตารางที่ 13 | ค่าสัมประสิทธิ์สหสัมพันธ์ในการวิเคราะห์ความสัมพันธ์ระหว่าง ตัวแปรต่างๆ ที่ใช้ในการทำนายอุบัติการณ์การเป็นเบาหวาน สำหรับผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคเบาหวานเมื่อเริ่มต้น Cohort เท่านั้น (เฉพาะคู่ที่มีสัมประสิทธิ์สหสัมพันธ์มากในทิศทางเดียวกัน) | 49 |
| ตารางที่ 14 | ค่า AUC เพื่อวัดประสิทธิภาพการทำนายความเสี่ยงการเกิดโรคเบาหวานในประชากรทั่วไป จาก Cohort ของผู้ที่ยังไม่เป็นโรคเบาหวาน (Cohort 1.1) (วิธีที่ดีที่สุดจะแสดงเป็นตัวหนา) | 51 |
| ตารางที่ 15 | สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ในการทำนายความเสี่ยงการเกิดโรคเบาหวานในประเทศไทย | 55-59 |
| ตารางที่ 16 | ตัวแปรที่นำมาพิจารณาในการสร้างตัวแบบทำนายความเสี่ยงของการเกิดโรคความดันโลหิตสูง | 60-63 |
| ตารางที่ 17 | ค่าเฉลี่ยเลขคณิต มัธยฐาน ส่วนเบี่ยงเบนมาตรฐาน ค่าต่ำสุด ค่าสูงสุดของอายุผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคความดันโลหิตสูงเมื่อเริ่มต้น Cohort เท่านั้น | 64 |
| ตารางที่ 18 | ค่าเฉลี่ยเลขคณิต มัธยฐาน ส่วนเบี่ยงเบนมาตรฐาน ค่าต่ำสุด ค่าสูงสุดของส่วนสูงผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคความดันโลหิตสูงเมื่อเริ่มต้น Cohort เท่านั้น | 66 |
| ตารางที่ 19 | ค่าเฉลี่ยเลขคณิต มัธยฐาน ส่วนเบี่ยงเบนมาตรฐาน ค่าต่ำสุด ค่าสูงสุดของดัชนีมวลกาย (BMI) ผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคเบาหวานเมื่อเริ่มต้น Cohort เท่านั้น | 67 |
| ตารางที่ 20 | ค่าเฉลี่ยเลขคณิต มัธยฐาน ส่วนเบี่ยงเบนมาตรฐาน ค่าต่ำสุด ค่าสูงสุด ขนาดเส้นรอบเอวของผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคความดันโลหิตสูงเมื่อเริ่มต้น Cohort เท่านั้น | 69 |
| ตารางที่ 21 | ค่าเฉลี่ยเลขคณิต มัธยฐาน ส่วนเบี่ยงเบนมาตรฐาน ค่าต่ำสุด ค่าสูงสุดของความดันช่วงหัวใจบีบตัว สำหรับผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคความดันโลหิตสูงเมื่อเริ่มต้น Cohort เท่านั้น | 70 |
| ตารางที่ 22 | ค่าเฉลี่ยเลขคณิต มัธยฐาน ส่วนเบี่ยงเบนมาตรฐาน ค่าต่ำสุด ค่าสูงสุดของความดันช่วงหัวใจคลายตัว สำหรับผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคความดันโลหิตสูงเมื่อเริ่มต้น Cohort เท่านั้น | 70 |
| ตารางที่ 23 | ค่าเฉลี่ยเลขคณิต มัธยฐาน ส่วนเบี่ยงเบนมาตรฐาน ค่าต่ำสุด ค่าสูงสุดของค่า Pulse Pressure สำหรับผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคความดันโลหิตสูงเมื่อเริ่มต้น Cohort เท่านั้น | 71 |

| | หน้าที่ | |
|-------------|---|---------|
| ตารางที่ 24 | ค่าเฉลี่ยเลขคณิต มัธยฐาน ส่วนเบี่ยงเบนมาตรฐาน ค่าต่ำสุด ค่าสูงสุด ค่าระดับไขมันไตรกลีเซอไรด์ ของผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคความดันโลหิตสูงเมื่อเริ่มต้น Cohort เท่านั้น | 74 |
| ตารางที่ 25 | ค่าเฉลี่ยเลขคณิต มัธยฐาน ส่วนเบี่ยงเบนมาตรฐาน ค่าต่ำสุด ค่าสูงสุดของระดับไขมันดี สำหรับผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคความดันโลหิตสูงเมื่อเริ่มต้น Cohort เท่านั้น | 75 |
| ตารางที่ 26 | ค่า AUC เพื่อวัดประสิทธิภาพการทำนายความเสี่ยงการเกิดโรคความดันโลหิตสูงในประชากรทั่วไป จาก Cohort ของผู้ที่ยังไม่เป็นโรคความดันโลหิตสูง (Cohort 2.1) (วิธีที่ดีที่สุดจะแสดงเป็นตัวหนา) | 77 |
| ตารางที่ 27 | สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ในการทำนายความเสี่ยงการเกิดโรคความดันโลหิตสูงในประชากรไทย | 81-84 |
| ตารางที่ 28 | รหัส ICD10 สำหรับกลุ่มโรคแทรกซ้อนอันเนื่องมาจากโรคเบาหวาน | 85 |
| ตารางที่ 29 | ตัวแปรที่นำมาพิจารณาในการสร้างตัวแบบทำนายความเสี่ยงภาวะแทรกซ้อนและการเสียชีวิตในผู้ป่วยที่เป็นโรคเบาหวานและความดันโลหิตสูง | 90-101 |
| ตารางที่ 30 | ค่า AUC เพื่อวัดประสิทธิภาพการทำนายความเสี่ยงภาวะเส้นเลือดหัวใจตีบหรือกล้ามเนื้อหัวใจขาดเลือดในผู้ป่วยโรคเบาหวาน (Cohort 1.2) (วิธีที่ดีที่สุดจะแสดงเป็นตัวหนา) | 103 |
| ตารางที่ 31 | สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ในการทำนายความเสี่ยงภาวะเส้นเลือดหัวใจตีบหรือกล้ามเนื้อหัวใจขาดเลือดในผู้ป่วยโรคเบาหวาน | 107-113 |
| ตารางที่ 32 | ค่า AUC เพื่อวัดประสิทธิภาพการทำนายความเสี่ยงภาวะเส้นเลือดสมองแตกหรือตีบในผู้ป่วยโรคเบาหวาน (Cohort 1.2) (วิธีที่ดีที่สุดจะแสดงเป็นตัวหนา) | 114 |
| ตารางที่ 33 | สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ในการทำนายความเสี่ยงภาวะเส้นเลือดสมองแตกหรือตีบในผู้ป่วยโรคเบาหวาน | 118-123 |
| ตารางที่ 34 | ค่า AUC เพื่อวัดประสิทธิภาพการทำนายความเสี่ยงภาวะแทรกซ้อนทางตาในผู้ป่วยโรคเบาหวาน (Cohort 1.2) (วิธีที่ดีที่สุดจะแสดงเป็นตัวหนา) | 124 |
| ตารางที่ 35 | สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ในการทำนายความเสี่ยงแทรกซ้อนทางตาในผู้ป่วยโรคเบาหวาน | 129-132 |
| ตารางที่ 36 | ค่า AUC เพื่อวัดประสิทธิภาพการทำนายความเสี่ยงภาวะไตเสื่อมเรื้อรังในผู้ป่วยโรคเบาหวาน (Cohort 1.2) (วิธีที่ดีที่สุดจะแสดงเป็นตัวหนา) | 133 |
| ตารางที่ 37 | สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ในการทำนายความเสี่ยงไตเสื่อมเรื้อรังในผู้ป่วยโรคเบาหวาน | 137-142 |

| | หน้าที่ | |
|-------------|---|---------|
| ตารางที่ 38 | ค่า AUC เพื่อวัดประสิทธิภาพการทำนายความเสี่ยงการเสียชีวิตในผู้ป่วยโรคเบาหวาน (Cohort 1.2) (วิธีที่ดีที่สุดจะแสดงเป็นตัวหนา) | 143 |
| ตารางที่ 39 | สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ในการทำนายความเสี่ยงการเสียชีวิตในผู้ป่วยโรคเบาหวาน | 147-153 |
| ตารางที่ 40 | ค่า AUC เพื่อวัดประสิทธิภาพการทำนายความเสี่ยงภาวะเส้นเลือดหัวใจตีบหรือกล้ามเนื้อหัวใจขาดเลือดในผู้ป่วยโรคความดันโลหิตสูง (Cohort 2.2) (วิธีที่ดีที่สุดจะแสดงเป็นตัวหนา) | 156 |
| ตารางที่ 41 | สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ในการทำนายความเสี่ยงภาวะเส้นเลือดหัวใจตีบหรือกล้ามเนื้อหัวใจขาดเลือดในผู้ป่วยโรคความดันโลหิตสูง | 161-167 |
| ตารางที่ 42 | ค่า AUC เพื่อวัดประสิทธิภาพการทำนายความเสี่ยงภาวะเส้นเลือดสมองแตกหรือตีบในผู้ป่วยโรคความดันโลหิตสูง (Cohort 2.2) (วิธีที่ดีที่สุดจะแสดงเป็นตัวหนา) | 168 |
| ตารางที่ 43 | สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ในการทำนายความเสี่ยงภาวะเส้นเลือดสมองแตกหรือตีบในผู้ป่วยโรคความดันโลหิตสูง | 173-179 |
| ตารางที่ 44 | ค่า AUC เพื่อวัดประสิทธิภาพการทำนายความเสี่ยงภาวะแทรกซ้อนทางตาในผู้ป่วยโรคความดันโลหิตสูง (Cohort 2.2) (วิธีที่ดีที่สุดจะแสดงเป็นตัวหนา) | 180 |
| ตารางที่ 45 | สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ในการทำนายความเสี่ยงภาวะแทรกซ้อนทางตาในผู้ป่วยโรคความดันโลหิตสูง | 185-187 |
| ตารางที่ 46 | ค่า AUC เพื่อวัดประสิทธิภาพการทำนายความเสี่ยงภาวะไตเสื่อมเรื้อรังในผู้ป่วยโรคความดันโลหิตสูง(Cohort 2.2) (วิธีที่ดีที่สุดจะแสดงเป็นตัวหนา) | 188 |
| ตารางที่ 47 | สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ในการทำนายความเสี่ยงภาวะไตเสื่อมเรื้อรังหรือตีบในผู้ป่วยโรคความดันโลหิตสูง | 192-196 |
| ตารางที่ 48 | ค่า AUC เพื่อวัดประสิทธิภาพการทำนายความเสี่ยงการเสียชีวิตในผู้ป่วยโรคความดันโลหิตสูง (Cohort 2.2) (วิธีที่ดีที่สุดจะแสดงเป็นตัวหนา) | 197 |
| ตารางที่ 49 | สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ในการทำนายความเสี่ยงภาวะเส้นเลือดสมองแตกหรือตีบในผู้ป่วยโรคความดันโลหิตสูง | 201-206 |
| ตารางที่ 50 | จำนวนผู้ใช้บริการและการให้บริการผู้ป่วยนอกและผู้ป่วยในระหว่างปี พ.ศ.2560 ถึง 2565 แยกรายประเภทผู้ป่วย | 207 |

| | | หน้าที่ |
|--------------------|---|---------|
| ตารางที่ 51 | จำนวนการรับบริการผู้ป่วยนอกและค่าบริการผู้ป่วยนอกเฉลี่ยต่อคนต่อปีแยกรายประเภทกลุ่มผู้ป่วย | 208 |
| ตารางที่ 52 | จำนวนการรับบริการผู้ป่วยนอกและค่าบริการผู้ป่วยนอกเฉลี่ยต่อคนต่อปีแยกรายเพศ | 208-209 |
| ตารางที่ 53 | สัมประสิทธิ์ การถดถอยจากตัวแบบการถดถอยลิเนียร์ในการอธิบายค่ารักษาพยาบาลผู้ป่วยนอกต่อครั้ง (บาท) | 210 |
| ตารางที่ 54 | จำนวนการรับบริการผู้ป่วยนอกและค่าบริการผู้ป่วยนอกเฉลี่ยต่อปีแยกรายจำนวนภาวะแทรกซ้อน | 210-211 |
| ตารางที่ 55 | จำนวนการรับบริการผู้ป่วยในและค่าบริการผู้ป่วยในเฉลี่ยต่อคนต่อปีแยกรายประเภทกลุ่มผู้ป่วย | 211 |
| ตารางที่ 56 | จำนวนการรับบริการผู้ป่วยในและค่าบริการผู้ป่วยในเฉลี่ยต่อคนต่อปีแยกรายเพศ | 212 |
| ตารางที่ 57 | สัมประสิทธิ์ การถดถอยจากตัวแบบการถดถอยลิเนียร์ในการอธิบายค่ารักษาพยาบาลผู้ป่วยในต่อครั้ง (บาท) | 214 |
| ตารางที่ 58 | จำนวนการรับบริการผู้ป่วยในและค่าบริการผู้ป่วยในเฉลี่ยต่อปีแยกรายจำนวนภาวะแทรกซ้อน | 215 |
| ตารางที่ 59 | ต้นทุนทางตรงที่ไม่เกี่ยวข้องกับการแพทย์ของการมารับบริการผู้ป่วยนอกและผู้ป่วยใน 1 ครั้ง | 216 |
| ตารางที่ 60 | ต้นทุนค่าเสียโอกาสของผู้ป่วยของการมารับบริการผู้ป่วยนอกและผู้ป่วยใน 1 ครั้ง | 217 |
| ตารางที่ 61 | ต้นทุนค่าเสียโอกาสของผู้ดูแลของโรคเบาหวานและความดันโลหิตสูงต่อปี | 218 |
| ตารางที่ 62 | ค่าใช้จ่ายทางสุขภาพเฉลี่ยต่อคนต่อปีแยกรายประเภทผู้ป่วยและรายเพศ | 220 |
| ตารางที่ 63 | ค่าใช้จ่ายทางสุขภาพเฉลี่ยต่อคนต่อปีแยกรายจำนวนภาวะแทรกซ้อน | 221 |
| ตารางที่ 64 | สรุปประสิทธิภาพการทำนายของแบบจำลองทำนายความเสี่ยงของการศึกษา (วิธีที่ดีที่สุดจะแสดงเป็นตัวหนา) | 223 |
| ตารางที่ 65 | ข้อดีและข้อด้อยของวิธี Elastic Net และ XGBoost | 224 |
| ตารางผนวก ที่ 1 | จำนวนการรับบริการผู้ป่วยนอกและค่าบริการผู้ป่วยนอกเฉลี่ยต่อคนต่อปีแยกรายประเภทสถานะของโรค | ผ1-3 |
| ตารางผนวก ที่ 2 | จำนวนการรับบริการผู้ป่วยในและค่าบริการผู้ป่วยในเฉลี่ยต่อคนต่อปีแยกรายประเภทสถานะของโรค | ผ4-6 |

สารบัญญภาพ

| | หน้าที่ | |
|-----------|---|-------|
| รูปที่ 1 | กรอบแนวคิดการวิจัย | 4 |
| รูปที่ 2 | ขั้นตอนการวิเคราะห์ข้อมูลเพื่อสร้างแบบจำลองทำนายความเสี่ยง | 17 |
| รูปที่ 3 | ตัวอย่างข้อมูลเวลาจนกระทั่งเกิดเหตุการณ์ที่สนใจ | 19 |
| รูปที่ 4 | ตัวอย่างการแปลงข้อมูลเวลาจนกระทั่งเกิดเหตุการณ์ที่สนใจให้อยู่ในรูปของเวลาที่ไม่ต่อเนื่อง | 20-21 |
| รูปที่ 5 | ตัวอย่างต้นไม้ตัดสินใจ | 23 |
| รูปที่ 6 | กราฟแสดงความสำคัญของตัวแปรจากตัวแบบต้นไม้ตัดสินใจ | 23 |
| รูปที่ 7 | กระบวนการทำงานของวิธี XGBoost (η คืออัตราการเรียนรู้ (Learning Rate)) | 24 |
| รูปที่ 8 | ตัวอย่างกราฟแสดงค่า SHAP | 25 |
| รูปที่ 9 | ตัวอย่างกราฟ ROC | 27 |
| รูปที่ 10 | บุคคลที่นำเข้าในการสร้าง Cohort ของผู้ที่ยังไม่เป็นโรคเบาหวาน (Cohort 1.1) | 30 |
| รูปที่ 11 | การเติมสถานะอุบัติการณ์การเป็นเบาหวานสำหรับข้อมูลสูญหาย | 31 |
| รูปที่ 12 | แผนภาพกล่อง แสดงถึงการกระจายข้อมูลอายุของผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคเบาหวานเมื่อเริ่มต้น Cohort เท่านั้น | 39 |
| รูปที่ 13 | แผนภูมิวงกลม แสดงถึงข้อมูลร้อยละ ผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคเบาหวานเมื่อเริ่มต้น Cohort เท่านั้น จำแนกตามเพศ | 39 |
| รูปที่ 14 | แผนภูมิวงกลมแสดงถึงข้อมูลร้อยละ ผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคเบาหวานเมื่อเริ่มต้น Cohort เท่านั้น จำแนกตามพฤติกรรมการสูบบุหรี่ | 40 |
| รูปที่ 15 | แผนภาพกล่องแสดงถึงการกระจายข้อมูลค่าดัชนีมวลกาย (BMI) ของผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคเบาหวานเมื่อเริ่มต้น Cohort เท่านั้น | 41 |
| รูปที่ 16 | แผนภูมิวงกลมแสดงถึงข้อมูลร้อยละ ผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคเบาหวานเมื่อเริ่มต้น Cohort เท่านั้น จำแนกตามดัชนีมวลกายแบบที่ 1 | 41 |
| รูปที่ 17 | แผนภูมิวงกลม แสดงถึงข้อมูลร้อยละ ผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคเบาหวานเมื่อเริ่มต้น Cohort เท่านั้น จำแนกตามดัชนีมวลกายแบบที่ 2 | 42 |
| รูปที่ 18 | แผนภูมิวงกลม แสดงถึงข้อมูลร้อยละ ผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคเบาหวานเมื่อเริ่มต้น Cohort เท่านั้น จำแนกตามขนาดเส้นรอบเอว | 43 |
| รูปที่ 19 | แผนภูมิวงกลมแสดงถึงข้อมูลร้อยละ ผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคเบาหวานเมื่อเริ่มต้น Cohort เท่านั้น จำแนกตามการมีประวัติเบาหวานในญาติสายตรง | 44 |

| | หน้าที่ | |
|-----------|---|----|
| รูปที่ 20 | แผนภูมิวงกลมแสดงถึงข้อมูลร้อยละ ผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคเบาหวานเมื่อเริ่มต้น Cohort เท่านั้น จำแนกตามการมีประวัติความดันโลหิตสูงในญาติสายตรง | 45 |
| รูปที่ 21 | แผนภูมิวงกลมแสดงถึงข้อมูลร้อยละ ผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคเบาหวานเมื่อเริ่มต้น Cohort เท่านั้น จำแนกตามการเกิดภาวะความดันโลหิตสูง | 45 |
| รูปที่ 22 | แผนภาพกล่องแสดงถึงการกระจายข้อมูลค่าระดับน้ำตาลในเลือด ของผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคเบาหวานเมื่อเริ่มต้น Cohort เท่านั้น | 46 |
| รูปที่ 23 | แผนภูมิวงกลมแสดงถึงข้อมูลร้อยละ ผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคเบาหวานเมื่อเริ่มต้น Cohort เท่านั้น จำแนกตามพฤติกรรมการอดอาหารก่อนเจาะเลือดเพื่อวัดระดับน้ำตาลในเลือด | 47 |
| รูปที่ 24 | แผนภูมิวงกลมแสดงถึงข้อมูลร้อยละ ผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคเบาหวานเมื่อเริ่มต้น Cohort เท่านั้น จำแนกตามระดับไขมันไตรกลีเซอไรด์แบบที่ 1 | 48 |
| รูปที่ 25 | แผนภูมิวงกลมแสดงถึงข้อมูลร้อยละ ผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคเบาหวานเมื่อเริ่มต้น Cohort เท่านั้น จำแนกตามระดับไขมันไตรกลีเซอไรด์แบบที่ 2 | 48 |
| รูปที่ 26 | ค่าสัมประสิทธิ์สหสัมพันธ์ในการวิเคราะห์ความสัมพันธ์ระหว่าง ตัวแปรต่างๆ ที่ใช้ในการทำนายอุบัติการณ์การเป็นเบาหวาน | 50 |
| รูปที่ 27 | ความสำคัญของตัวแปรทำนายจากวิธี XGBoost ในการทำนายความเสี่ยงการเกิดโรคเบาหวานในประเทศไทย | 53 |
| รูปที่ 28 | ค่า SHAP จากตัวแบบ XGBoost ในการทำนายความเสี่ยงการเกิดโรคเบาหวานในประเทศไทย | 54 |
| รูปที่ 29 | แผนภาพกล่อง แสดงถึงการกระจายข้อมูลอายุของผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคความดันโลหิตสูงเมื่อเริ่มต้น Cohort เท่านั้น | 64 |
| รูปที่ 30 | แผนภูมิวงกลม แสดงถึงข้อมูลร้อยละ ผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคความดันโลหิตสูงเมื่อเริ่มต้น Cohort เท่านั้น จำแนกตามช่วงอายุ | 65 |
| รูปที่ 31 | แผนภูมิวงกลม แสดงถึงข้อมูลร้อยละ ผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคความดันโลหิตสูงเมื่อเริ่มต้น Cohort เท่านั้น จำแนกตามเพศ | 65 |
| รูปที่ 32 | แผนภูมิวงกลมแสดงถึงข้อมูลร้อยละ ผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคความดันโลหิตสูงเมื่อเริ่มต้น Cohort เท่านั้น จำแนกตามพฤติกรรมการสูบบุหรี่ | 66 |
| รูปที่ 33 | แผนภาพกล่องแสดงถึงการกระจายข้อมูลค่าดัชนีมวลกาย (BMI) ของผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคความดันโลหิตสูงเมื่อเริ่มต้น Cohort เท่านั้น | 67 |

| | หน้าที่ |
|---|---------|
| รูปที่ 34 แผนภูมิวงกลมแสดงถึงข้อมูลร้อยละ ผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคความดันโลหิตสูงเมื่อเริ่มต้น Cohort เท่านั้น จำแนกตามดัชนีมวลกายแบบที่ 1 | 68 |
| รูปที่ 35 แผนภูมิวงกลมแสดงถึงข้อมูลร้อยละ ผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคความดันโลหิตสูงเมื่อเริ่มต้น Cohort เท่านั้น จำแนกตามดัชนีมวลกายแบบที่ 2 | 68 |
| รูปที่ 36 แผนภาพกล่องแสดงถึงการกระจายข้อมูลค่าเส้นรอบเอว ของผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคความดันโลหิตสูงเมื่อเริ่มต้น Cohort เท่านั้น | 69 |
| รูปที่ 37 แผนภาพกล่องแสดงถึงการกระจายข้อมูลค่าความดันช่วงหัวใจบีบตัว ของผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคความดันโลหิตสูงเมื่อเริ่มต้น Cohort เท่านั้น | 70 |
| รูปที่ 38 แผนภาพกล่องแสดงถึงการกระจายข้อมูลค่าความดันช่วงหัวใจคลายตัว ของผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคความดันโลหิตสูงเมื่อเริ่มต้น Cohort เท่านั้น | 71 |
| รูปที่ 39 แผนภาพกล่องแสดงถึงการกระจายข้อมูลค่า Pulse Pressure ของผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคความดันโลหิตสูงเมื่อเริ่มต้น Cohort เท่านั้น | 72 |
| รูปที่ 40 แผนภูมิวงกลมแสดงถึงข้อมูลร้อยละ ผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคความดันโลหิตสูงเมื่อเริ่มต้น Cohort เท่านั้น จำแนกตามการมีประวัติเบาหวานในญาติสายตรง | 72 |
| รูปที่ 41 แผนภูมิวงกลมแสดงถึงข้อมูลร้อยละ ผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคความดันโลหิตสูงเมื่อเริ่มต้น Cohort เท่านั้น จำแนกตามการมีประวัติความดันโลหิตสูงในญาติสายตรง | 73 |
| รูปที่ 42 แผนภูมิวงกลมแสดงถึงข้อมูลร้อยละ ผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคความดันโลหิตสูงเมื่อเริ่มต้น Cohort เท่านั้น จำแนกตามการเป็นโรคเบาหวาน | 74 |
| รูปที่ 43 แผนภูมิวงกลมแสดงถึงข้อมูลร้อยละ ผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคความดันโลหิตสูงเมื่อเริ่มต้น Cohort เท่านั้น จำแนกตามการทราบค่าไขมันไตรกลีเซอไรด์ | 75 |
| รูปที่ 44 แผนภูมิวงกลมแสดงถึงข้อมูลร้อยละ ผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคความดันโลหิตสูงเมื่อเริ่มต้น Cohort เท่านั้น จำแนกตามการทราบค่าระดับไขมันดี | 76 |
| รูปที่ 45 แผนภูมิวงกลม แสดงถึงข้อมูลร้อยละ ผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคความดันโลหิตสูงเมื่อเริ่มต้น Cohort เท่านั้น จำแนกตามระดับไขมันดี | 76 |
| รูปที่ 46 ความสำคัญของตัวแปรทำนายจากวิธี XGBoost ในการทำนายความเสี่ยงการเกิดโรคความดันโลหิตสูงในประชากรไทย | 78 |

| | หน้าที่ | |
|-----------|--|-----|
| รูปที่ 47 | ค่า SHAP จากตัวแบบ XGBoost ในการทำนายความเสี่ยงการเกิดโรคความดันโลหิตสูงในประชากรไทย | 80 |
| รูปที่ 48 | เมทริกซ์เปลี่ยนสถานะรายปีของผู้ป่วยที่เป็นโรคเบาหวานในมาตราส่วนเชิงลอการิทึม | 87 |
| รูปที่ 49 | เมทริกซ์เปลี่ยนสถานะรายปีของผู้ป่วยที่เป็นโรคเบาหวานและความดันโลหิตสูงในมาตราส่วนเชิงลอการิทึม | 88 |
| รูปที่ 50 | ความสำคัญของตัวแปรทำนายจากวิธี XGBoost ในการทำนายความเสี่ยงภาวะเส้นเลือดหัวใจตีบหรือกล้ามเนื้อหัวใจขาดเลือดในผู้ป่วยโรคเบาหวาน | 104 |
| รูปที่ 51 | ค่า SHAP จากตัวแบบ XGBoost ในการทำนายความเสี่ยงภาวะเส้นเลือดหัวใจตีบหรือกล้ามเนื้อหัวใจขาดเลือดในผู้ป่วยโรคเบาหวาน | 106 |
| รูปที่ 52 | ความสำคัญของตัวแปรทำนายจากวิธี XGBoost ในการทำนายความเสี่ยงภาวะเส้นเลือดสมองแตกหรือตีบในผู้ป่วยโรคเบาหวาน | 115 |
| รูปที่ 53 | ค่า SHAP จากตัวแบบ XGBoost ในการทำนายความเสี่ยงภาวะเส้นเลือดสมองแตกหรือตีบในผู้ป่วยโรคเบาหวาน | 117 |
| รูปที่ 54 | ความสำคัญของตัวแปรทำนายจากวิธี XGBoost ในการทำนายภาวะแทรกซ้อนทางตาในผู้ป่วยโรคเบาหวาน | 126 |
| รูปที่ 55 | ค่า SHAP จากตัวแบบ XGBoost ในการทำนายความเสี่ยงภาวะแทรกซ้อนทางตาในผู้ป่วยโรคเบาหวาน | 128 |
| รูปที่ 56 | ความสำคัญของตัวแปรทำนายจากวิธี XGBoost ในการทำนายภาวะไตเสื่อมเรื้อรังในผู้ป่วยโรคเบาหวาน | 134 |
| รูปที่ 57 | ค่า SHAP จากตัวแบบ XGBoost ในการทำนายความเสี่ยงภาวะไตเสื่อมเรื้อรังในผู้ป่วยโรคเบาหวาน | 136 |
| รูปที่ 58 | ความสำคัญของตัวแปรทำนายจากวิธี XGBoost ในการทำนายการเสียชีวิตในผู้ป่วยโรคเบาหวาน | 144 |
| รูปที่ 59 | ค่า SHAP จากตัวแบบ XGBoost ในการทำนายความเสี่ยงการเสียชีวิตในผู้ป่วยโรคเบาหวาน | 146 |
| รูปที่ 60 | เมทริกซ์เปลี่ยนสถานะรายปีของผู้ป่วยที่เป็นโรคความดันโลหิตสูงในมาตราส่วนเชิงลอการิทึม | 155 |
| รูปที่ 61 | ความสำคัญของตัวแปรทำนายจากวิธี XGBoost ในการทำนายความเสี่ยงภาวะเส้นเลือดหัวใจตีบหรือกล้ามเนื้อหัวใจขาดเลือดในผู้ป่วยโรคความดันโลหิตสูง | 158 |
| รูปที่ 62 | ค่า SHAP จากตัวแบบ XGBoost ในการทำนายความเสี่ยงภาวะเส้นเลือดหัวใจตีบหรือกล้ามเนื้อหัวใจขาดเลือดในผู้ป่วยโรคความดันโลหิตสูง | 160 |
| รูปที่ 63 | ความสำคัญของตัวแปรทำนายจากวิธี XGBoost ในการทำนายความเสี่ยงเส้นเลือดสมองแตกหรือตีบในผู้ป่วยโรคความดันโลหิตสูง | 170 |
| รูปที่ 64 | ค่า SHAP จากตัวแบบ XGBoost ในการทำนายความเสี่ยงภาวะเส้นเลือดสมองแตกหรือตีบในผู้ป่วยโรคความดันโลหิตสูง | 172 |

| | หน้าที่ |
|-----------|--|
| รูปที่ 65 | ความสำคัญของตัวแปรทำนายจากวิธี XGBoost ในการทำนายภาวะแทรกซ้อนทางตาในผู้ป่วยโรคความดันโลหิตสูง 182 |
| รูปที่ 66 | ค่า SHAP จากตัวแบบ XGBoost ในการทำนายความเสี่ยงภาวะแทรกซ้อนทางตาในผู้ป่วยโรคความดันโลหิตสูง 184 |
| รูปที่ 67 | ความสำคัญของตัวแปรทำนายจากวิธี XGBoost ในการทำนายภาวะไตเสื่อมเรื้อรังในผู้ป่วยโรคความดันโลหิตสูง 190 |
| รูปที่ 68 | ค่า SHAP จากตัวแบบ XGBoost ในการทำนายความเสี่ยงภาวะไตเสื่อมเรื้อรังในผู้ป่วยโรคความดันโลหิตสูง 191 |
| รูปที่ 69 | ความสำคัญของตัวแปรทำนายจากวิธี XGBoost ในการทำนายการเสียชีวิตในผู้ป่วยโรคความดันโลหิตสูง 199 |
| รูปที่ 70 | ค่า SHAP จากตัวแบบ XGBoost ในการทำนายความเสี่ยงการเสียชีวิตในผู้ป่วยโรคความดันโลหิตสูง 200 |
| รูปที่ 71 | ค่ารักษาพยาบาลผู้ป่วยนอกเฉลี่ยต่อคนต่อปี (บาท) แยกรายประเภทผู้ป่วยและช่วงอายุ 209 |
| รูปที่ 72 | ค่ารักษาพยาบาลผู้ป่วยในเฉลี่ยต่อคนต่อปี (บาท) แยกรายประเภทผู้ป่วยและช่วงอายุ 213 |
| รูปที่ 73 | จำนวนวันนอนเฉลี่ยต่อการรักษาตัวในโรงพยาบาล 1 ครั้งแยกรายประเภทผู้ป่วยและช่วงอายุ 213 |
| รูปที่ 74 | ต้นทุนใช้จ่ายสุขภาพเฉลี่ยต่อคนต่อปี (บาท) แยกรายองค์ประกอบของต้นทุน 218 |
| รูปที่ 75 | ต้นทุนค่าใช้จ่ายสุขภาพเฉลี่ยต่อคนต่อปี (บาท) แยกรายประเภทผู้ป่วย 219 |

บทที่ 1 บทนำ

1.1 หลักการและเหตุผล

ในปัจจุบัน โรคเบาหวานและโรคความดันโลหิตสูงเป็นปัญหาใหญ่ทางด้านระบบสาธารณสุขของประเทศ ไทยและต่างประเทศ โดยข้อมูลจากองค์การอนามัยโลก (WHO) พบว่าระหว่างปี พ.ศ. 2523 ถึง 2557 มีผู้ป่วย เบาหวานทั่วโลกเพิ่มขึ้นจาก 180 เป็น 422 ล้านคน และพบแนวโน้มของภาระโรคความดันโลหิตสูงในทิศทาง เดียวกันโดยทั่วโลกมีผู้ป่วยโรคความดันโลหิตสูงเพิ่มขึ้นจาก 594 ล้าน เป็น 1.13 พันล้านคนระหว่างปี พ.ศ. 2518 และ 2558 ตามลำดับ นอกจากนี้ การศึกษาต่าง ๆ ยังพบว่าความชุกของโรคเบาหวานและ ความดันโลหิตสูงในกลุ่มประเทศที่มีรายได้ต่ำและปานกลางซึ่งรวมประเทศไทยอยู่ด้วยนั้นมีแนวโน้มสูงขึ้น มากกว่าประเทศที่มีรายได้สูง ซึ่งส่งผลกระทบต่อภาวะแทรกซ้อนและค่าใช้จ่ายทางสุขภาพที่ตามมา โดยมีการ ประมาณค่าไว้ว่าค่าใช้จ่ายทางตรงและทางอ้อมในการรักษาพยาบาลและการขาดรายได้จากการขาดงาน หรือพิการอันเนื่องมาจากโรคเบาหวานและความดันโลหิตสูงทั่วโลกสูงถึง 1.7 และ 3.6 ล้านล้านเหรียญ ดอลลาร์สหรัฐ ตามลำดับ

สำหรับสถานการณ์ของประเทศไทย จากสำรวจสุขภาพประชาชนไทยโดยการตรวจร่างกายครั้งที่ 4 และ 5 ในปี พ.ศ. 2552 และ 2557 ตามลำดับ พบว่า ประเทศไทยมีผู้ป่วยเบาหวานที่อายุ 15 ปีขึ้นไปเพิ่มขึ้นจาก ร้อยละ 6.9 เป็น 8.9 และมีสัดส่วนของผู้ป่วยที่ไม่ทราบว่าตนเองเป็นเบาหวานเพิ่มขึ้นจากร้อยละ 31.2 เป็น 43.1 ตามลำดับ ซึ่งแสดงให้เห็นถึงการเพิ่มขึ้นของภาระโรคเบาหวานในประเทศไทย รวมทั้งจากการสำรวจ เดียวกันนี้ยังพบว่า แนวโน้มของภาระโรคความดันโลหิตสูงมีแนวโน้มคล้ายคลึงกัน โดย พบว่ามีผู้ป่วย โรคความดันโลหิตสูงที่อายุ 15 ปีขึ้นไปเพิ่มขึ้นจากร้อยละ 21.4 เป็น 24.7 แต่มีสัดส่วนของผู้ป่วยที่ไม่ได้รับการ วินิจฉัยลดลงจากร้อยละ 50.3 เหลือ 44.7 ตามลำดับ ในขณะที่ประเทศไทยกำลังเข้าสู่ช่วงเปลี่ยน ผ่านของโครงสร้างประชากรเข้าสู่สังคมผู้สูงอายุ โดยข้อมูลจากองค์การสหประชาชาติ คาดว่า ประเทศไทยจะมี สัดส่วนของผู้ที่มีอายุ 65 ปี ขึ้นไป เพิ่มขึ้นจากร้อยละ 12.9 เป็น 25.8 ระหว่างปี พ.ศ. 2563 ถึง 2583 ตามลำดับ จากปัจจัยดังกล่าว ส่งผลให้โรคไม่ติดต่อเรื้อรัง ซึ่ง ณ ปัจจุบันถือว่าเป็นภาระโรคที่สำคัญของ ประเทศมีแนวโน้มทวีความรุนแรงมากขึ้น โดยเฉพาะโรคเบาหวานและความดันโลหิตสูงซึ่งพบเพิ่มสูงขึ้นตาม ช่วงอายุ และเป็นจุดเริ่มต้นซึ่งของการเกิดโรคไม่ติดต่อเรื้อรังประเภทต่าง ๆ เช่น โรคหลอดเลือดหัวใจตีบ โรคเส้นเลือดสมองตีบหรือแตก และ โรคไตวายเรื้อรัง เป็นต้น ซึ่งการเพิ่มขึ้นของภาระโรคกลุ่มนี้ล้วนส่งผล กระทบต่อทั้งผู้ป่วย ครอบครัว สังคม ระบบสาธารณสุขและเศรษฐกิจของประเทศเป็นอย่างมาก

จากเหตุผลดังกล่าว คณะกรรมการปฏิรูปประเทศด้านสาธารณสุข ได้เล็งเห็นถึงความสำคัญใน การแก้ไขปัญหาโรคไม่ติดต่อเรื้อรัง โดยเฉพาะโรคเบาหวานและความดันโลหิตสูงซึ่งเป็นการแก้ไขปัญหาที่ สอดคล้องกับแผนยุทธศาสตร์แห่งชาติ ในส่วนของการพัฒนาและเสริมสร้างศักยภาพทรัพยากรมนุษย์ ให้ ประชาชนของประเทศมีสุขภาพที่ดีในทุกช่วงวัย มีความรอบรู้ด้านสุขภาพ และสามารถป้องกันหรือหลีกเลี่ยง ปัจจัยเสี่ยงทางสุขภาพได้ ผ่านทางการพัฒนาการจัดระบบบริการ มาตรการทางภาษี และการเพิ่ม ประสิทธิภาพและประสิทธิผลของการสร้างเสริมป้องกันโรค ซึ่งการมีแบบจำลองทำนายความเสี่ยงระดับบุคคล และผลกระทบระดับมหภาค สามารถเป็นหนึ่งในวิธีการที่ช่วยเพิ่มประสิทธิภาพและประสิทธิผลของการสร้าง

เสริมป้องกันโรค ผ่านทางการสร้างความรอบรู้เกี่ยวกับความเสี่ยงทางสุขภาพของทั้งประชาชนทั่วไปและผู้ป่วยโรคเบาหวานและความดันโลหิตสูงได้อย่างชัดเจนมากขึ้น เมื่อเทียบกับข้อมูลระดับน้ำตาลในเลือดหรือค่าความดันโลหิตตามมาตรฐานเดิม นอกจากนี้ ผู้ที่ใช้บริการแบบจำลองทำนายความเสี่ยง สามารถทราบได้ถึงแนวโน้มผลลัพธ์ทางสุขภาพที่จะเปลี่ยนแปลงไป หากตนเองได้ปรับเปลี่ยนพฤติกรรม เช่น ลดน้ำหนัก หรือ เลิกบุหรี่ ส่งผลให้เกิดแรงจูงใจในการปรับเปลี่ยนพฤติกรรมเพื่อชะลอการเกิดโรคเบาหวานและความดันโลหิตสูงในประชากรทั่วไป และชะลอการเกิดภาวะแทรกซ้อน หรือเสียชีวิตก่อนวัยอันควรในผู้ป่วย และข้อมูลที่ได้ยังสามารถนำไปคาดการณ์ผลกระทบทางสุขภาพและทางเศรษฐกิจในระดับมหภาค รวมถึงในประชากรกลุ่มเปราะบางได้อีกด้วย

ในช่วงทศวรรษที่ผ่านมา บทบาทของปัญญาประดิษฐ์และฐานข้อมูลขนาดใหญ่ได้มีการพัฒนาอย่างก้าวกระโดดทั่วโลก โดย ข้อมูลที่มีการบันทึกไว้ในฐานข้อมูลได้ถูกนำไปใช้ในการวิเคราะห์และประกอบการตัดสินใจเชิงนโยบายและการวางแผนการดำเนินงานในหลากหลายมิติทั้งในองค์กรของรัฐและเอกชน รวมทั้งในส่วนของงานทางด้านสาธารณสุขเองได้มีการนำประโยชน์ของเทคโนโลยีทั้งสองนี้ไปใช้ในการพัฒนางานหลากหลายมิติ เช่น การวินิจฉัยโรค การพัฒนา ยา การวางแผนทรัพยากรสาธารณสุข และการพัฒนาเทคโนโลยีเพื่อช่วยในการประกอบการตัดสินใจ เป็นต้น โดยเมื่อมุ่งเป้าไปยังงานวิจัยทางการพัฒนาแบบจำลองทำนายความเสี่ยงในช่วง 5 ปีที่ผ่านมา พบว่าเทคโนโลยีดังกล่าวได้เริ่มมีบทบาทมากขึ้นและสามารถพัฒนาแบบจำลองทำนายความเสี่ยงให้มีความแม่นยำสูงและสามารถใช้งานได้จริง เพื่อให้แบบจำลองทำนายความเสี่ยงมีความสามารถในการทำนายสูงและมีความแม่นยำเหมาะสมกับกลุ่มประชากรทั้งในส่วนของประชาชนทั่วไปที่ยังไม่เกิดโรคและกลุ่มผู้ป่วยที่เกิดโรคแล้วซึ่งมีจุดมุ่งเน้นเพื่อป้องกันการเกิดภาวะแทรกซ้อนและการเสียชีวิตก่อนวัยอันควร การมีข้อมูลประชากรที่มากเพียงพอและมีความหลากหลายทั้งในมิติของลักษณะประชากร พฤติกรรมสุขภาพและสถานะสุขภาพจึงมีความจำเป็นต่อการพัฒนาแบบจำลองทำนายความเสี่ยงทั้งสิ้น ดังนั้น ฐานข้อมูลสำนักงานหลักประกันสุขภาพแห่งชาติซึ่งเป็นฐานข้อมูลสาธารณสุขขนาดใหญ่ของประเทศไทยที่มีการเก็บบันทึกข้อมูลดังกล่าวจึงสามารถใช้เป็นข้อมูลตั้งต้นในการพัฒนาแบบจำลองทำนายความเสี่ยงได้ ถึงแม้ว่าฐานข้อมูลดังกล่าวจะมีข้อจำกัดในแง่ของความสมบูรณ์ของข้อมูลไปบ้าง หากเทียบกับงานวิจัยทั่วไปที่มีการติดตามและเก็บข้อมูลผู้ป่วยระยะยาว เช่น การศึกษาระยะยาวถึงอิทธิพลของปัจจัยเสี่ยงต่อการเกิดโรคหัวใจหลอดเลือดและเมแทบอลิซึมในพนักงานการไฟฟ้าฝ่ายผลิตแห่งประเทศไทย หรืออีกชื่อหนึ่ง คือ EGAT Study การที่มีฐานข้อมูลที่มีขนาดใหญ่เพียงพอจากสำนักงานหลักประกันสุขภาพแห่งชาติ ซึ่งทำให้ผู้วิจัยสามารถเลือกกลุ่มประชากรที่มีข้อมูลสมบูรณ์มาการผนวกกับการใช้เทคนิค Machine Learning ในการพัฒนาแบบจำลองสามารถช่วยเติมเต็มข้อจำกัดเรื่องความสมบูรณ์ของข้อมูลได้ รวมทั้งเทคนิค Machine Learning สามารถนำมาใช้ในการหาความสัมพันธ์และสร้างสมการระหว่างข้อมูลตัวแปรที่มีความซับซ้อนซึ่งแบบจำลองทำนายความเสี่ยงเดิมไม่สามารถทำได้ให้มีความสามารถในการทำนายและความแม่นยำที่สูงขึ้นอีกด้วย

ดังนั้น ชุดการศึกษานี้จึงออกแบบมาเพื่อสร้างแบบจำลองทำนายความเสี่ยง ผ่านทางการคำนวณข้อมูลทางระบาดวิทยาของโรคเบาหวานและโรคความดันโลหิตสูงในประเทศไทยระหว่างปี พ.ศ. 2558 ถึง 2563 โดยใช้ข้อมูลจากฐานข้อมูลการใช้บริการสาธารณสุขระดับประเทศของสำนักงานหลักประกันสุขภาพแห่งชาติ (สปสช.) และทำการวิเคราะห์หาปัจจัยที่ส่งผลต่ออุบัติการณ์ของโรค ภาวะแทรกซ้อนและการเสียชีวิตอันเนื่องมาจากโรคเบาหวานและโรคความดันโลหิตสูงตามลำดับ และทำการพัฒนาแบบจำลองเพื่อการคาดการณ์

การเกิดโรคในประชากรทั่วไปและแบบจำลองทำนายความเสี่ยงการเกิดภาวะแทรกซ้อนประเภทต่าง ๆ และการเสียชีวิตในกลุ่มผู้ป่วยเพิ่มเติมด้วยวิธีการทางปัญญาประดิษฐ์แบบ Machine Learning ตรรกะและแบบจำลองที่ถูกสร้างขึ้นจะถูกนำไปใช้พัฒนาเครื่องมือประเมินผลกระทบทางสุขภาพรายบุคคลที่สามารถบรรจุในแอปพลิเคชันบนคอมพิวเตอร์หรือสมาร์ทโฟน และยังนำไปคาดการณ์ผลกระทบทางเศรษฐกิจอันเนื่องมาจากโรคเบาหวานและความดันโลหิตสูงที่จะเกิดขึ้นในอนาคตเพื่อประกอบการตัดสินใจเชิงนโยบายในการวางแผน ควบคุมและป้องกันรักษาโรคเบาหวานและความดันโลหิตสูงในประเทศไทยต่อไป

1.2 วัตถุประสงค์

1. วิเคราะห์ความเสี่ยงทางระบาดวิทยา ประกอบด้วย
 - a. ปัจจัยและขนาดของความเสี่ยงที่เกี่ยวข้องกับการเกิดโรคเบาหวานและความดันโลหิตสูงในกลุ่มประชากรทั่วไป
 - b. ปัจจัยและขนาดของความเสี่ยงที่เกี่ยวข้องกับการเกิดภาวะแทรกซ้อน 4 ประเภท ได้แก่ โรคหลอดเลือดหัวใจ โรคเส้นเลือดสมองตีบหรือแตก โรคไตเสื่อมเรื้อรังระยะสุดท้าย และภาวะแทรกซ้อนทางตาในผู้ป่วยโรคเบาหวานและความดันโลหิตสูง
 - c. ปัจจัยและขนาดของความเสี่ยงที่เกี่ยวข้องกับการเสียชีวิตในผู้ป่วยโรคเบาหวานและความดันโลหิตสูง
2. สร้างแบบจำลองทำนายความเสี่ยง 4 แบบจำลอง ประกอบด้วย
 - a. แบบจำลองทำนายความเสี่ยงการเกิดโรคเบาหวานในกลุ่มประชากรทั่วไป
 - b. แบบจำลองทำนายความเสี่ยงการเกิดภาวะแทรกซ้อนและการเสียชีวิตอันเนื่องมาจากโรคเบาหวาน
 - c. แบบจำลองทำนายความเสี่ยงการเกิดโรคความดันโลหิตสูง
 - d. แบบจำลองทำนายความเสี่ยงการเกิดภาวะแทรกซ้อนและการเสียชีวิตอันเนื่องมาจากโรคความดันโลหิตสูง
3. คำนวณค่าใช้จ่ายทางสุขภาพและความสูญเสียด้านผลิตภาพการทำงาน
 - a. ค่าใช้จ่ายของผู้ที่ยังไม่เกิดโรค
 - b. ค่าใช้จ่ายทางสุขภาพของผู้ที่เป็นโรคเบาหวาน/ความดันโลหิตสูงแต่ยังไม่มีภาวะแทรกซ้อน
 - c. ค่าใช้จ่ายทางสุขภาพของผู้ที่เป็นโรคเบาหวาน/ความดันโลหิตสูงและมีภาวะแทรกซ้อน

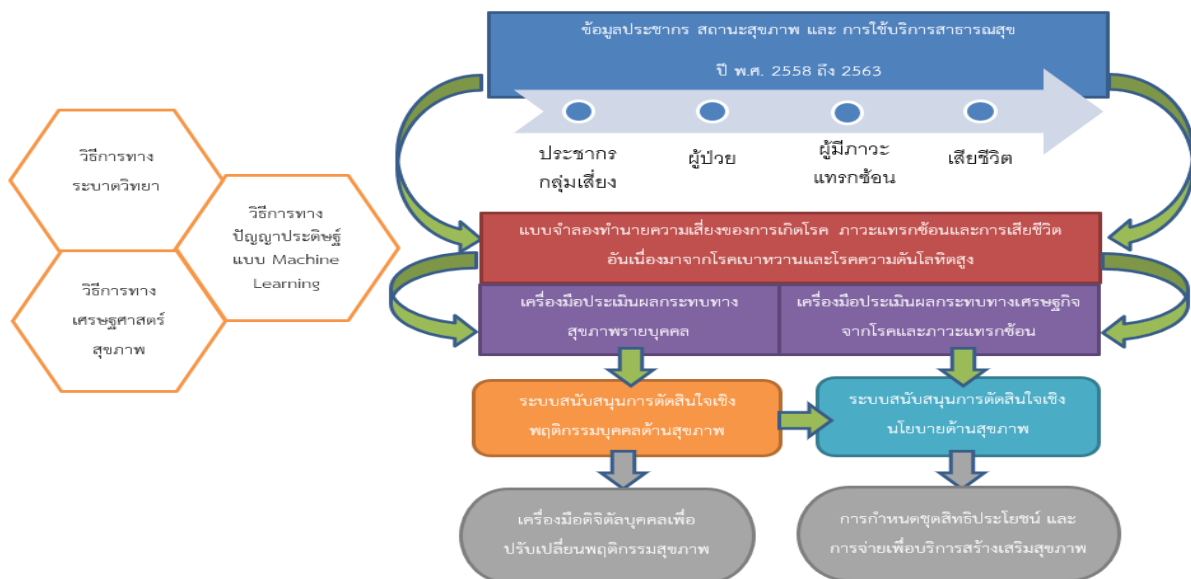
1.3 กรอบแนวคิดการวิจัย (Conceptual framework)

การศึกษานี้มีจุดมุ่งเน้นเพื่อพัฒนาแบบจำลองทำนายความเสี่ยงผ่านทางการใช้ข้อมูลการใช้บริการสาธารณสุขขนาดใหญ่ของประเทศของสำนักงานหลักประกันสุขภาพแห่งชาติ โดยนำข้อมูลจากฐานข้อมูลดังกล่าวมาคำนวณผลกระทบทางระบาดวิทยาของโรคเบาหวานและความดันโลหิตสูง ทั้งในแง่มุมมองการเพิ่มขึ้นของภาระโรค (Incidence) ของโรคเบาหวานและความดันโลหิตสูงเองในกลุ่มประชากรทั่วไปและ

การเพิ่มขึ้นของภาวะแทรกซ้อนและการเสียชีวิตก่อนวัยอันควรอันเนื่องมาจากโรคทั้งสองในกลุ่มผู้ป่วยผ่านทางการสร้าง Cohort ติดตามกลุ่มเป้าหมายระยะยาวระยะเวลาทั้งสิ้น 6 ปี 2 เดือนระหว่างเดือนมกราคม พ.ศ. 2560 ถึง เดือนกุมภาพันธ์ 2566 ซึ่งเป็นช่วงเวลาที่ฐานข้อมูลมีข้อมูลสำคัญทั้งในส่วนของปัจจัยเสี่ยงและข้อมูลผลลัพธ์ของการศึกษาที่สมบูรณ์

นอกจากนี้ ข้อมูลทางระบาดวิทยาที่คำนวณได้และข้อมูลความสัมพันธ์ระหว่างปัจจัยเสี่ยงและผลลัพธ์ของการศึกษาจะถูกนำไปใช้ในการพัฒนาแบบจำลองทำนายความเสี่ยง 2 ประเภทหลัก คือ 1. แบบจำลองทำนายความเสี่ยงการเกิดโรคเบาหวานและความดันโลหิตสูงสำหรับประชากรทั่วไปที่ยังไม่เกิดโรค และ 2. แบบจำลองทำนายความเสี่ยงการเกิดภาวะแทรกซ้อนและการเสียชีวิตอันเนื่องมาจากโรคเบาหวานและความดันโลหิตสูงสำหรับผู้ป่วยโรคดังกล่าว โดยนำจุดเด่นของฐานข้อมูลขนาดใหญ่ที่มีความหลากหลายทางประชากร สถานะสุขภาพ และ พฤติกรรมสุขภาพมาผสานกับเทคนิคทาง Machine Learning เพื่อให้แบบจำลองทำนายความเสี่ยงที่ถูกพัฒนาขึ้นมีความแม่นยำสูง มีความเป็นตัวแทนทางประชากรของประเทศสามารถนำไปใช้ในการให้ข้อมูลประกอบการตัดสินใจทั้งในมุมมองของประชาชนและผู้ให้บริการทางสาธารณสุขได้ในบริบทของสถานการณ์จริงภายในประเทศ

มากกว่านั้น ข้อมูลความเสี่ยงที่แบบจำลองได้พัฒนาขึ้นจะถูกนำไปใช้ในการประเมินผลกระทบทางเศรษฐกิจ โดยใช้ข้อมูลต้นทุนการให้บริการสุขภาพในฝั่งของ Provider ที่มีการบันทึกไว้ในฐานข้อมูลสุขภาพ 43 แห่งมาผสานกับต้นทุนทางฝั่งผู้มารับบริการที่มีการบันทึกไว้ในวรรณกรรมของประเทศไทยเพื่อให้สามารถคำนวณ ผลกระทบทางเศรษฐกิจที่เกิดขึ้น ณ ปัจจุบัน และผลกระทบที่เกิดขึ้นในอนาคตผ่านการวิเคราะห์ Scenario ต่าง ๆ เพื่อให้ทราบถึงประโยชน์และความคุ้มค่าของนโยบายที่จะนำไปใช้ และเป็นส่วนหนึ่งของการสร้าง Evidence Based Policies ต่อไป



รูปที่ 1 กรอบแนวคิดการวิจัย

1.4 ขอบเขตการดำเนินงาน

1. การศึกษานี้จะใช้เพียงข้อมูลจากฐานข้อมูลสุขภาพ 43 แฟ้มเท่านั้น ซึ่งครอบคลุมเฉพาะประชาชนที่รับบริการการรักษาโรคและการสร้างเสริมป้องกันโรคกับโรงพยาบาลที่มีการนำส่งข้อมูลให้กับกระทรวงสาธารณสุข โดยโรงพยาบาลส่วนใหญ่เป็นโรงพยาบาลของรัฐและสังกัดสำนักงานปลัดกระทรวงสาธารณสุข
2. การประเมินผลกระทบทางเศรษฐกิจอันเนื่องมาจากโรคเบาหวานและความดันโลหิตสูง นอกเหนือจากข้อมูลที่ได้รับจากแบบจำลองทำนายความเสี่ยงแล้ว หากทางผู้วิจัยจำเป็นต้องใช้ข้อมูลส่วนเสริมอื่น ๆ จะเลือกใช้ข้อมูลที่เป็นตัวแทนของประเทศไทยก่อน หากไม่มีข้อมูลจึงจะเลือกจากวรรณกรรมของต่างประเทศที่มีลักษณะกลุ่มตัวอย่างใกล้เคียงกับประชาชนไทย เพื่อให้การประเมินผลกระทบทางเศรษฐกิจมีความใกล้เคียงความเป็นจริงมากที่สุด

บทที่ 2

การทบทวนวรรณกรรม

โรคเบาหวานและความดันโลหิตสูงเป็นโรคไม่ติดต่อเรื้อรังที่สร้างปัญหาให้กับระบบสาธารณสุขทั่วโลก จากการศึกษาต่าง ๆ พบว่าความชุกของโรคเบาหวานทั่วโลกมีแนวโน้มสูงขึ้น¹⁻⁶ โดยในปี พ.ศ. 2538 ความชุกของโรคเบาหวานทั่วโลกอยู่ที่ร้อยละ 4 หรือคิดเป็นผู้ป่วยประมาณ 135 ล้านคน⁵ และเพิ่มขึ้นเป็นร้อยละ 8.8 หรือคิดเป็นผู้ป่วยประมาณ 425 ล้านคนและหากอุบัติการณ์ของโรคยังคงสูงขึ้นด้วยอัตราเดิม จำนวนผู้ป่วยโรคเบาหวานมีโอกาสูงถึง 629 ล้านคนในปี พ.ศ.2588² ซึ่งการเพิ่มขึ้นของโรคเบาหวานนั้นพบมากในกลุ่มประเทศที่มีรายได้ต่ำและปานกลางซึ่งประเทศไทยได้รวมอยู่ด้วย โดยกลุ่มประเทศเหล่านี้มีอัตราส่วนของผู้ป่วยเบาหวานสูงถึงร้อยละ 79 ของจำนวนผู้ป่วยทั่วโลก¹⁻³ สำหรับสถานการณ์ของโรคความดันโลหิตสูงนั้นมีแนวโน้มเช่นเดียวกับโรคเบาหวาน โดยจากการศึกษาแนวโน้มโรคความดันโลหิตสูงของโลก พบว่า ผู้ป่วยโรคความดันโลหิตสูงมีปริมาณเพิ่มขึ้นจาก 594 ล้านเป็น 1.13 พันล้านคนระหว่างปี พ.ศ. 2518 และ 2558⁷ นอกจากนี้ เมื่อทำการวิเคราะห์ความชุกของโรคความดันโลหิตสูงแยกตามสถานะเศรษฐกิจของประเทศ พบว่า กลุ่มประเทศที่มีรายได้ต่ำและปานกลางมีความชุกของโรคความดันโลหิตสูงมากกว่ากลุ่มประเทศที่มีรายได้สูง เช่นเดียวกับโรคเบาหวาน^{4,7,8}

จากปัจจัยดังกล่าวส่งผลให้เกิดภาวะแทรกซ้อนและการเสียชีวิตต่าง ๆ อันเนื่องมาจากโรคเบาหวาน และความดันโลหิตสูงเป็นจำนวนมาก เช่น โรคเส้นเลือดหัวใจตีบ โรคเส้นเลือดสมองแตกหรือตีบ โรคไตวายเรื้อรัง ภาวะพิการจากการที่ต้องถูกตัดขา ภาวะพิการทางสายตา เส้นประสาทถูกทำลายอันเนื่องมาจากโรคเบาหวาน โรคทางพันธุกรรม ภาวะเกิดโรคีพของทารกในหญิงตั้งครรภ์ที่เป็นเบาหวาน และ ครรภ์เป็นพิษ ในหญิงตั้งครรภ์ที่เป็นโรคความดันโลหิตสูง เป็นต้น^{1,2,8-17} และนำไปสู่ความสูญเสียทางด้านทรัพยากรของระบบสาธารณสุขและผลกระทบต่อเศรษฐกิจผ่านทางรายจ่ายสุขภาพที่เพิ่มสูงขึ้นและการขาดรายได้จากการขาดงานหรือพิการ โดยพบว่าค่าใช้จ่ายทางสุขภาพอันเนื่องมาจากโรคเบาหวานและความดันโลหิตสูงทั่วโลก สูงถึง 1.7 และ 3.6 ล้านล้านเหรียญดอลลาร์สหรัฐ ตามลำดับ^{18,19} นอกจากนี้ โรคเบาหวานและความดันโลหิตสูงยังส่งผลให้ผู้ป่วยเกิดภาวะสิ้นเนื้อประดาตัวจากการเจ็บป่วยโดยเฉพาะในกลุ่มผู้ป่วยที่มีโรคไม่ติดต่อเรื้อรังหลายโรคร่วมและผู้ป่วยในกลุ่มประเทศที่มีรายได้ต่ำ^{20,21}

สำหรับสถานการณ์ของประเทศไทย จากสำรวจสุขภาพประชาชนไทยโดยการตรวจร่างกาย ครั้งที่ 4 และ 5 ในปี พ.ศ. 2552 และ 2557 ตามลำดับ พบว่า ประเทศไทยมีผู้ป่วยเบาหวานที่อายุ 15 ปีขึ้นไปเพิ่มขึ้นจากร้อยละ 6.9 เป็น 8.9 และมีสัดส่วนของผู้ป่วยที่ไม่ทราบว่าตนเองเป็นเบาหวานเพิ่มขึ้นจากร้อยละ 31.2 เป็น 43.1 ตามลำดับ ซึ่งแสดงให้เห็นถึงการเพิ่มขึ้นของภาระโรคเบาหวานในประเทศไทย รวมทั้งจากการสำรวจเดียวกันนี้ยังพบว่า แนวโน้มของภาระโรคความดันโลหิตสูงมีแนวโน้มคล้ายคลึงกัน โดย พบว่ามีผู้ป่วยโรคความดันโลหิตสูงที่อายุ 15 ปีขึ้นไปเพิ่มขึ้นจากร้อยละ 21.4 เป็น 24.7 แต่มีสัดส่วนของผู้ป่วยที่ไม่ได้รับการวินิจฉัยลดลงจากร้อยละ 50.3 เหลือ 44.7 ตามลำดับ^{22,23} ในขณะเดียวกัน ประเทศไทยกำลังเข้าสู่ช่วงเปลี่ยนผ่านของโครงสร้างประชากรเข้าสู่สังคมผู้สูงอายุ โดยข้อมูลจากองค์การสหประชาชาติ คาดว่า ประเทศไทยจะมีสัดส่วนของผู้ที่มีอายุ 65 ปี ขึ้นไป เพิ่มจากร้อยละ 12.9 เป็น 25.8 ระหว่างปี พ.ศ. 2563 ถึง 2583 ตามลำดับ²⁴ จากปัจจัยดังกล่าว ส่งผลให้โรคไม่ติดต่อเรื้อรัง ซึ่ง ณ ปัจจุบันถือว่าเป็นภาระโรคที่สำคัญ

ของประเทศมีแนวโน้มทวีความรุนแรงมากขึ้น โดยเฉพาะโรคเบาหวานและความดันโลหิตสูงซึ่งพบเพิ่มสูงขึ้นตามช่วงอายุ และเป็นจุดเริ่มต้นซึ่งของการเกิดโรคไม่ติดต่อเรื้อรังประเภทต่าง ๆ เช่น โรคหลอดเลือดหัวใจตีบ โรคเส้นเลือดสมองตีบหรือแตก และ โรคไตวายเรื้อรัง เป็นต้น^{2,8,11,13,14} ซึ่งการเพิ่มขึ้นของภาระโรคกลุ่มนี้ล้วนส่งผลกระทบต่อทั้งผู้ป่วย ครอบครัว สังคม ระบบสาธารณสุขและเศรษฐกิจของประเทศเป็นอย่างมาก

สำหรับแนวคิดเรื่องการใช้แบบจำลองทำนายความเสี่ยงเข้ามาช่วยทำนายการเกิดโรคเบาหวานและความดันโลหิตสูงในประชาชนทั่วไปนั้น พบว่า มีงานศึกษาวิจัยรองรับและประสบความสำเร็จในการสร้างแบบจำลองทำนายความเสี่ยงมาก่อนหน้าแล้ว โดยจากการทบทวนวรรณกรรมอย่างเป็นระบบในฐานข้อมูลระดับนานาชาติถึงจำนวนของแบบจำลองทำนายความเสี่ยงการเกิดโรคเบาหวานและความดันโลหิตสูงและปัจจัยสำคัญที่ใช้ในการสร้างแบบจำลอง ณ พ.ศ. 2558 พบว่า มีการศึกษาจำนวน 39 และ 11 ศึกษาซึ่งรายงานแบบจำลองทำนายความเสี่ยงการเกิดโรคเบาหวานและความดันโลหิตสูงตามลำดับ^{25,26} โดยปัจจัยเสี่ยงสำคัญที่นำมาใช้ในการสร้างแบบจำลองของทั้ง 2 โรคมักเป็นปัจจัยเสี่ยงที่พบร่วมกันและสัมพันธ์กับข้อมูลทางระบาดวิทยา เช่น เพศ อายุ เชื้อชาติ ค่า BMI เส้นรอบเอว ประวัติการสูบบุหรี่ ประวัติโรคเบาหวานและความดันโลหิตสูงในครอบครัว และค่าระดับความดันโลหิตหรือระดับน้ำตาลของผู้ที่ถูกประเมินซึ่งข้อมูลเหล่านี้ล้วนแต่เป็นข้อมูลที่มีการบันทึกไว้ในฐานข้อมูลสุขภาพ 43 แห่งทั้งสิ้น

ในแง่มุมมองของการพัฒนาแบบจำลองทำนายความเสี่ยงการเกิดภาวะแทรกซ้อนและการเสียชีวิตอันเนื่องมาจากโรคเบาหวานและความดันโลหิตสูง มีการพัฒนาที่ก้าวหน้ามากกว่าการทำนายความเสี่ยงการเกิดโรคมก เนื่องจากภาวะแทรกซ้อนของโรคเบาหวานและความดันโลหิตสูงมีความหลากหลายมากและความรุนแรงของภาวะแทรกซ้อนเหล่านี้เป็นอันตรายถึงชีวิตดังที่ได้กล่าวไว้ข้างต้น โดยแบบจำลองทำนายความเสี่ยงการเกิดภาวะแทรกซ้อนที่มักถูกพูดถึง คือ UK Prospective Diabetes Study (UKPDS) ที่ถูกพัฒนาด้วยเทคนิค Markov Model โดยมีวัตถุประสงค์เพื่อประเมินความเสี่ยงของการเกิดภาวะแทรกซ้อนที่ลดได้อันเนื่องมาจากโรคเบาหวาน หากผู้ป่วยสามารถควบคุมระดับน้ำตาลในเลือดเฉลี่ย (HbA_{1c}) ได้ดีขึ้น²⁷ โดยพบว่าหากสามารถลดระดับ HbA_{1c} ได้ 1% จะสามารถลดโอกาสในการเกิด microvascular complication เช่น ภาวะเบาหวานขึ้นจอประสาทตา ภาวะไตเสื่อม และภาวะเส้นประสาทเสื่อมได้ถึง 35% ลดโอกาสเกิดโรคหลอดเลือดหัวใจตีบและการเสียชีวิตจากภาวะแทรกซ้อนของโรคเบาหวานได้ 18 และ 25% ตามลำดับ²⁸ จากแบบจำลองทำนายความเสี่ยงดังกล่าว ส่งผลให้เกิดการพัฒนาแบบจำลองทำนายความเสี่ยงการเกิดภาวะแทรกซ้อนอันเนื่องมาจากโรคเบาหวานและความดันโลหิตสูงตามมาเป็นจำนวนมากในกลุ่มผู้ป่วยหลากหลายเชื้อชาติ²⁹⁻³² โดยเฉพาะในกลุ่มโรคเบาหวานที่มีการจัดตั้งเครือข่าย Mt Hood Diabetes Challenge Network สำหรับการพัฒนาแบบจำลองทำนายความเสี่ยงโรคเบาหวานเพื่อใช้ในการทำนายความเสี่ยงและวิเคราะห์ผลกระทบทางเศรษฐกิจ³³ ซึ่งข้อมูลปัจจัยเสี่ยงที่ใช้ในการสร้างแบบจำลองมีความคล้ายคลึงกับแบบจำลองทำนายความเสี่ยงการเกิดโรคเพียงแต่จะมีการเพิ่มผลทางห้องปฏิบัติการเฉพาะ เช่น ค่า creatinine หรือ eGFR ซึ่งแสดงถึงประสิทธิภาพของไต ค่าไขมันในเลือด ระยะเวลาที่ป่วยด้วยโรคเบาหวานหรือความดันโลหิตสูง และสถานะของผู้ป่วยว่ามีภาวะแทรกซ้อนประเภทใดบ้างและระยะเวลาในการทำนายของแบบจำลองที่ได้ยกตัวอย่างในการทบทวนวรรณกรรมอยู่ที่ระหว่าง 5 – 20 ปี ตามลำดับ นอกจากนี้ แบบจำลองทำนายความเสี่ยงการเกิดภาวะเส้นเลือดหัวใจตีบและเส้นเลือดสมองแตกหรือตีบที่มี

การพัฒนาขึ้นและได้รับรองในประเทศออสเตรเลียและนิวซีแลนด์มีความสามารถในการทำนายภาวะดังกล่าว
อยู่ที่ระยะเวลา 5 ปี^{34,35}

อย่างไรก็ตาม ข้อจำกัดของฐานข้อมูลของประเทศไทยในการสร้างแบบจำลองทำนายความเสี่ยงนั้นอยู่ที่
ผลการตรวจทางห้องปฏิบัติการระดับน้ำตาลเฉลี่ยในเลือด (HbA_{1c}) ที่แบบจำลองทำนายความเสี่ยงทั่วโลกใช้
เป็นปัจจัยหลักในการประเมินความรุนแรงของโรคเบาหวาน เพียงแต่การตรวจทางห้องปฏิบัติการชนิดนี้ยังเป็น
สัดส่วนน้อย เมื่อเทียบกับการตรวจระดับน้ำตาลในเลือดหรือปลายนิ้วหลังอดอาหาร ซึ่งเป็นการตรวจทาง
ห้องปฏิบัติการที่ใช้มากกว่าในการให้บริการคัดกรองโรคและตรวจติดตามผู้ป่วยโรคเบาหวาน ประเด็นดังกล่าว
จึงเป็นความท้าทายของงานวิจัยขึ้นนี้ในการพัฒนาแบบจำลองทำนายความเสี่ยงให้เข้ากับบริบทของการบริการ
สุขภาพของประเทศ นอกจากนี้ เนื่องจากระบบฐานข้อมูลทั่วโลกมีการพัฒนาอย่างก้าวกระโดด ทำให้การเก็บ
รวบรวมข้อมูลและการวิเคราะห์ข้อมูลขนาดใหญ่มีความเป็นไปได้ง่ายขึ้น การพัฒนาแบบจำลองทำนาย
ความเสี่ยงที่ต้องใช้ข้อมูลปริมาณมากเพื่อพัฒนาให้การทำนายมีความแม่นยำและสามารถนำผลการทำนายไป
ใช้ประโยชน์ได้ในหลากหลายมิติจึงมีการนำเทคนิค Machine Learning เข้ามาใช้ในการพัฒนาแบบจำลอง
ดังที่เห็นได้จากปีที่ตีพิมพ์ผลงานวิจัย โดยปัจจัยเสี่ยงสำคัญที่นำมาใช้ยังคงเป็นปัจจัยเสี่ยงเดิมกับงานวิจัย
การพัฒนาแบบจำลองก่อนหน้า แต่ใช้เทคนิค Machine Learning เข้ามาช่วยหาความสัมพันธ์และสร้าง
สมการระหว่างข้อมูลตัวแปรที่การสร้างแบบจำลองทำนายความเสี่ยงเดิมไม่สามารถทำได้³⁶⁻³⁸

เมื่อเปรียบเทียบแบบจำลองทำนายความเสี่ยงที่จะพัฒนาขึ้นกับแบบจำลองทำนายความเสี่ยง CV Risk
Score ที่ได้มีการพัฒนาจากการศึกษา EGAT Study^{39,40} ความแตกต่างสำคัญที่คาดว่าจะได้รับจากการศึกษา
นี้คือ ความแม่นยำของแบบจำลอง ซึ่งเกิดจากการที่แบบจำลองทำนายความเสี่ยงใหม่จะถูกพัฒนาจากข้อมูล
สุขภาพของประชากรทั้งประเทศซึ่งมีความหลากหลายทางลักษณะประชากรและมีจำนวนตัวอย่างข้อมูลสูง
กว่าแบบจำลองเดิมอยู่มาก (จำนวนกลุ่มตัวอย่างที่ใช้มีจำนวนในหลักล้านราย) ซึ่งทางผู้วิจัยคาดว่า ปัจจัยนี้
เป็นปัจจัยสำคัญที่ทำให้แบบจำลองทำนายความเสี่ยงใหม่จะมีความสามารถในการทำนายที่สูงกว่าและ
สามารถนำไปขยายผลใช้กับประชากรจริงได้ดีกว่าแบบจำลองเดิม ถึงแม้ว่าระยะเวลาติดตามของการศึกษานี้
ในระยะแรกจะสั้นกว่าโดยอยู่ที่ 6 ปี โดยแบบจำลองทำนายความเสี่ยงที่จะถูกพัฒนาขึ้นจะมีการพัฒนาให้ตรงกับ
รูปแบบการใช้งานของ User ผ่านทางการใช้ข้อมูลประกอบการวิเคราะห์ที่แตกต่างกัน เช่น

1. **การตรวจคัดกรองเชิงรุกของกระทรวงสาธารณสุข** แบบจำลองทำนายความเสี่ยงการเกิดโรคเบาหวานจะใช้
เพียงข้อมูลที่สามารถรับจากการซักประวัติและการตรวจร่างกายโดยไม่จำเป็นต้องใช้ผลการตรวจทาง
ห้องปฏิบัติการ เพื่อให้ทาง User สามารถคัดกรองความเสี่ยงเบื้องต้นได้ และหากมีข้อมูลผลการตรวจทาง
ห้องปฏิบัติการ เช่น ระดับน้ำตาลในเลือด จะสามารถช่วยให้ความแม่นยำในการทำนายการเกิดโรคสูงขึ้น
2. **การตรวจสุขภาพประจำปี** ซึ่งจะประกอบด้วยผลการตรวจทางห้องปฏิบัติการและการตรวจร่างกายอยู่แล้ว ใน
กรณีนี้ แบบจำลองทำนายความเสี่ยงสามารถใช้ข้อมูลจากทั้ง 2 องค์ประกอบในการประมวลผลความเสี่ยงที่จะ
เกิดขึ้นได้
3. **การตรวจติดตามการรักษาโรคเบาหวานและความดันโลหิตสูง** ซึ่งโดยทั่วไป ผู้ป่วยกลุ่มนี้จะต้องได้รับ
การรักษาต่อเนื่อง ทำให้มีข้อมูลการติดตามระยะยาว เช่น ผลทางห้องปฏิบัติการที่เปลี่ยนแปลงไปในแต่ละ
ช่วงเวลา ประเภทของยาที่ใช้ในการควบคุมโรค ภาวะแทรกซ้อนที่เกิดขึ้นระหว่างการรักษา เป็นต้น ข้อมูล

เหล่านี้ จะถูกนำมาประกอบการคำนวณความเสี่ยงและทำให้มีความสามารถในการทำนายได้มากกว่า
แบบจำลองเดิม ซึ่งเป็นแบบจำลองในลักษณะของข้อ 2

ตารางที่ 1 เปรียบเทียบแบบจำลองทำนายความเสี่ยงของประเทศไทยเทียบกับแบบจำลองทำนายความเสี่ยงที่ทางผู้วิจัยพัฒนาขึ้น

| ประเด็นสำคัญของแบบจำลองทำนายความเสี่ยง | ประเทศไทย (EGAT Study) ^{39,40} | แบบจำลองทำนายความเสี่ยงที่ผู้วิจัยพัฒนาขึ้น |
|---|---|--|
| วัตถุประสงค์ | ทำนายความเสี่ยงในการเกิดโรคในช่วงเวลา 10 ปีข้างหน้า โดย ณ ปัจจุบัน มีแบบจำลองทำนายความเสี่ยง 3 กลุ่มโรคหลัก คือ <ol style="list-style-type: none"> แบบจำลองทำนายความเสี่ยงการเกิดโรคหัวใจและหลอดเลือดสมอง (Thai CV risk Score)⁴¹ แบบจำลองทำนายความเสี่ยงการเกิดโรคไตวายเรื้อรัง⁴² แบบจำลองทำนายความเสี่ยงการเกิดโรคเบาหวาน (RAMA-EGAT DIABETES SCORE)⁴³ | ทำนายความเสี่ยงในการเกิดโรคในช่วงเวลา 5 ปีข้างหน้า โดยแบ่งเป็น <ol style="list-style-type: none"> การทำนายความเสี่ยงการเกิดโรคเบาหวานและความดันโลหิตสูงในประชาชนทั่วไป (แบบจำลองที่ 1) การทำนายความเสี่ยงการเกิดภาวะแทรกซ้อนในกลุ่มผู้ป่วยที่เป็นโรคเบาหวานและความดันโลหิตสูง (แบบจำลองที่ 2.1) การทำนายความเสี่ยงการเสียชีวิตในกลุ่มผู้ป่วยที่เป็นโรคเบาหวานและความดันโลหิตสูง (แบบจำลองที่ 2.2) |
| กลุ่มตัวอย่างที่ใช้ในการสร้างแบบจำลองทำนายความเสี่ยง | ข้อมูลสุขภาพของพนักงานการไฟฟ้าผลิตแห่งประเทศไทย จำนวนกลุ่มตัวอย่างประมาณ 9,100 คน | ข้อมูลสุขภาพของประชาชนไทยนอกเขตพื้นที่ กรุงเทพมหานคร และมีสิทธิการรักษาหลักประกันสุขภาพถ้วนหน้า จำนวนกลุ่มตัวอย่างประมาณ 36 ล้านคนโดยการศึกษานี้ สุ่มเลือกกลุ่มตัวอย่างมาทั้งสิ้น 2 ล้านคน |
| ประเภทของข้อมูลที่ใช้ในการสร้างแบบจำลองทำนายความเสี่ยง | ลักษณะทางประชากร: เพศ อายุ ระดับการศึกษา, อาชีพ ประวัติโรคในครอบครัว: โรคเบาหวาน โรคหลอดเลือดหัวใจ พฤติกรรมเสี่ยง: การสูบบุหรี่ และการดื่มเครื่องดื่มแอลกอฮอล์ สถานะสุขภาพ ณ วันที่สำรวจว่ามีโรคประจำตัวใดบ้าง ผลการตรวจร่างกาย: ความดันโลหิต น้ำหนัก ส่วนสูง เส้นรอบเอว ชีพจร | ลักษณะทางประชากร: เพศ และ อายุ ประวัติโรคในครอบครัว: โรคเบาหวาน และ โรคความดันโลหิตสูง พฤติกรรมเสี่ยง: การสูบบุหรี่ และการดื่มเครื่องดื่มแอลกอฮอล์ สถานะสุขภาพ จากฐานข้อมูลการใช้บริการสาธารณสุข ผลการตรวจร่างกาย: ความดันโลหิต น้ำหนัก ส่วนสูง เส้นรอบเอว |

ตารางที่ 1 เปรียบเทียบแบบจำลองทำนายความเสี่ยงของประเทศไทยเทียบกับแบบจำลองทำนายความเสี่ยงที่ทางผู้วิจัยพัฒนาขึ้น (ต่อ)

| ประเด็นสำคัญของแบบจำลองทำนายความเสี่ยง | ประเทศไทย (EGAT Study) ^{39,40} | แบบจำลองทำนายความเสี่ยงที่ผู้วิจัยพัฒนาขึ้น |
|---|---|--|
| <p>ประเภทของข้อมูลที่ใช้ในการสร้างแบบจำลองทำนายความเสี่ยง (ต่อ)</p> | <p>ผลทางห้องปฏิบัติการ: ความสมบูรณ์ของเม็ดเลือด ผลการตรวจปัสสาวะ ระดับน้ำตาลและไขมันในเลือด การทำงานของไตและตับ เป็นต้น (คล้ายคลึงกับการตรวจสุขภาพประจำปี)</p> <p>ข้อมูลการเกิดโรคเป้าหมาย ได้รับจากการวิเคราะห์รหัสโรค (ICD) ในข้อมูลผู้ป่วยนอกและผู้ป่วย ของ ฐานข้อมูลการเบิกจ่ายค่ารักษาพยาบาลที่ขึ้นตรงกับสิทธิ์การรักษาของผู้ป่วย และ/หรือ การโทรศัพท์สัมภาษณ์</p> <p>ข้อมูลการเสียชีวิต ได้รับจากฐานข้อมูลมรณบัตร กระทรวงมหาดไทยที่มีการเชื่อมฐานข้อมูลกับ สปสช.</p> | <p>ผลทางห้องปฏิบัติการ กลุ่มประชาชนทั่วไป (แบบจำลองที่ 1) ระดับน้ำตาลในเลือดจากการคัดกรองประจำปีและผลตรวจทางห้องปฏิบัติการอื่น ๆ ที่ประชาชนรายดังกล่าวเคยได้รับการตรวจที่โรงพยาบาล</p> <p>กลุ่มผู้ป่วยโรคเบาหวานและความดันโลหิตสูง (แบบจำลองที่ 2) ระดับน้ำตาลในเลือด การทำงานของไต ซึ่งเป็นการตรวจทางห้องปฏิบัติการที่สำคัญที่พบประกอบการมารับบริการแบบผู้ป่วยนอก รวมทั้งผลตรวจทางห้องปฏิบัติการอื่น ๆ ที่ผู้ป่วยได้รับการตรวจ จะถูกนำมาวิเคราะห์เพื่อพัฒนาแบบจำลองร่วมด้วย</p> <p>ประเภทของยาที่ใช้ในการรักษาโรค (เพิ่มเติมจาก EGAT Study) ข้อมูลการเกิดโรคเป้าหมาย ได้รับจากการวิเคราะห์รหัสโรค (ICD) ในข้อมูลผู้ป่วยนอกและผู้ป่วยในของฐานข้อมูล 43 แพ้ม</p> <p>ข้อมูลการเสียชีวิต ได้รับจากการบันทึกวันที่เสียชีวิตในฐานข้อมูล 43 แพ้ม</p> |

ตารางที่ 1 เปรียบเทียบแบบจำลองทำนายความเสี่ยงของประเทศไทยเทียบกับแบบจำลองทำนายความเสี่ยงที่ทางผู้วิจัยพัฒนาขึ้น (ต่อ)

| ประเด็นสำคัญของแบบจำลองทำนายความเสี่ยง | ประเทศไทย (EGAT Study) ^{39,40} | แบบจำลองทำนายความเสี่ยงที่ผู้วิจัยพัฒนาขึ้น |
|--|---|--|
| รูปแบบการนำไปใช้ประโยชน์ | <p>สามารถใช้ทำนายความเสี่ยงได้ทั้งกรณีที่มีและไม่มีผลทางห้องปฏิบัติการ โดยหากมีผลทางห้องปฏิบัติการร่วมด้วยจะมีความแม่นยำในการทำนายความเสี่ยงเพิ่มขึ้น^{43,44}</p> <p>Thai CV Risk Score ถูกกำหนดให้ใช้เป็นเครื่องมือในการประเมินความเสี่ยงที่สำคัญและถูกบรรจุเป็นหนึ่งในตัวชี้วัดของกระทรวงสาธารณสุข⁴⁴</p> | <p>แบบจำลองทำนายความเสี่ยงที่วางแผนไว้จะพัฒนาขึ้นในลักษณะเดียวกัน โดยจะพัฒนาแบบจำลองออกเป็น 3 ส่วนหลักเพื่อตอบโจทย์กลุ่มเป้าหมายที่แตกต่างกัน โดย</p> <ol style="list-style-type: none"> 1. แบบจำลองที่ไม่จำเป็นต้องใช้ผลการตรวจทางห้องปฏิบัติการ 2. แบบจำลองที่ใช้ผลการตรวจทางห้องปฏิบัติการร่วมด้วย เช่น ผลการตรวจสุขภาพประจำปี 3. แบบจำลองที่ใช้ข้อมูลการรักษาโรคต่อเนืองระยะยาว เช่น ผลทางห้องปฏิบัติการที่เปลี่ยนแปลงไปในแต่ละช่วงเวลา ประเภทของยาที่ใช้ในการรักษา เป็นต้น <p>คาดว่าจะถูกนำไปใช้ประกอบการประเมินความเสี่ยงทั้งในส่วนของหน่วยงานรัฐและเอกชน โดยแบบจำลองทำนายความเสี่ยงแต่ละรูปแบบที่ถูกพัฒนาขึ้นจะใช้ในการตอบโจทย์ในแต่ละประเด็นตามที่ได้อธิบายไว้ในเบื้องต้น</p> |

จากสถานการณ์ภาวะโรคเบาหวานและความดันโลหิตสูงของประเทศไทยดังที่ได้กล่าวไว้ในข้างต้น คณะกรรมการปฏิรูปประเทศด้านสาธารณสุข ได้เล็งเห็นถึงความสำคัญในการแก้ไขปัญหาโรคไม่ติดต่อเรื้อรัง โดยเฉพาะโรคเบาหวานและความดันโลหิตสูงซึ่งเป็นการแก้ไขปัญหาที่สอดคล้องกับแผนยุทธศาสตร์ชาติ ในด้านการพัฒนาและเสริมสร้างศักยภาพทรัพยากรมนุษย์ ให้ประชาชนของประเทศไทยมีสุขภาวะที่ดีในทุกช่วงวัย มีความรอบรู้ด้านสุขภาวะ และสามารถป้องกันหรือหลีกเลี่ยงปัจจัยเสี่ยงทางสุขภาพได้ โดยดำเนินการจัดทำหัวข้อการปฏิรูปหลัก “กิจกรรมปฏิรูปที่ ๒ การปฏิรูปเพื่อเพิ่มประสิทธิภาพและประสิทธิผลของการสร้างเสริมสุขภาพ ความรอบรู้ด้านสุขภาพ การป้องกันและดูแลรักษาโรคไม่ติดต่อสำหรับประชาชนและผู้ป่วย” โดยมุ่งเน้นในส่วนของกลุ่มวัยทำงานที่มีความเสี่ยงต่อการเกิดโรคและกลุ่มผู้ป่วยให้สามารถเข้าถึงบริการการรักษาและสร้างเสริมป้องกันโรค ผ่านทางกลไก 5 วิธีหลัก ดังต่อไปนี้

1. การออกนโยบายให้สถานที่ทำงาน มีการสร้างเสริมสุขภาพและการคัดกรองโรคเป็นส่วนหนึ่งของสวัสดิการและการบริหารทรัพยากรมนุษย์ทั้งในองค์กรของภาครัฐและเอกชน ผ่านทางความร่วมมือกับกระทรวงแรงงานและสำนักงาน กพ.
2. การพัฒนาชุดสิทธิประโยชน์ให้กับผู้ป่วยและผู้ที่มีความเสี่ยงต่อโรคเบาหวานและความดันโลหิตสูงผ่านการสร้างเสริมสุขภาพ การป้องกันโรคและการเพิ่มความรอบรู้ทางสุขภาพ โดยดำเนินการจัดสรรงบประมาณด้านการสร้างเสริมสุขภาพและการป้องกันโรคเพิ่มเติมเพื่อให้ได้ผลลัพธ์ตรงตามเป้าหมาย
3. การจัดระบบฐานข้อมูลที่สามารถเชื่อมโยงข้อมูลสุขภาพของโรคเบาหวานและความดันโลหิตสูงระหว่างหน่วยงานที่เกี่ยวข้อง รวมทั้งการมีระบบส่งข้อมูลกลับไปให้ผู้ป่วยและองค์การบริหารส่วนท้องถิ่นเพื่อให้ผู้ป่วยทราบถึงสถานะทางสุขภาพของตนเองและหน่วยงานภาคท้องถิ่นสามารถนำข้อมูลที่ได้รับไปใช้ในการปรับการจัดบริการเพื่อตอบสนองต่อภาวะโรคในแต่ละพื้นที่ได้อย่างตรงจุด
4. การนำเข้าเทคโนโลยีสุขภาพโดยเฉพาะเทคโนโลยีในกลุ่มอุปกรณ์ดิจิทัลส่วนบุคคลเพื่อช่วยเพิ่มประสิทธิภาพในการดูแลผู้ป่วยโรคเบาหวานและความดันโลหิตสูงและการเพิ่มการยอมรับของเทคโนโลยีทั้งในส่วนของบุคลากรระดับปฏิบัติการและผู้ป่วย
5. ดำเนินนโยบายมาตรการทางภาษีสำหรับอาหารและสารเคมีที่เป็นสาเหตุของโรคไม่ติดต่อ โดยมุ่งเน้นไปที่ น้ำตาล โซเดียม และสารเคมีอันตรายของภาคการเกษตร และนางบประมาณที่จัดเก็บได้ไปใช้ในการดำเนินงานด้านการสร้างเสริมสุขภาพและการเพิ่มการเข้าถึงเทคโนโลยีสุขภาพของประชาชน

การมีแบบจำลองทำนายความเสี่ยงระดับบุคคลและผลกระทบระดับมหภาคจึงสามารถเป็นหนึ่งในวิธีการที่ช่วยเพิ่มประสิทธิภาพและประสิทธิผลของการสร้างเสริมป้องกันโรค ผ่านทางการสร้างความรอบรู้เกี่ยวกับความเสี่ยงทางสุขภาพของทั้งประชาชนทั่วไปและผู้ป่วยโรคเบาหวานและความดันโลหิตสูงได้อย่างชัดเจนมากขึ้น เมื่อเทียบกับข้อมูลระดับน้ำตาลในเลือดหรือค่าความดันโลหิตตามมาตรฐานเดิม นอกจากนี้ ผู้ที่ใช้บริการแบบจำลองทำนายความเสี่ยง สามารถทราบได้ถึงแนวโน้มผลลัพธ์ทางสุขภาพที่จะเปลี่ยนแปลงไป หากตนเองได้ปรับเปลี่ยนพฤติกรรม เช่น ลดน้ำหนัก หรือ เลิกบุหรี่ ส่งผลให้เกิดแรงจูงใจในการปรับเปลี่ยนพฤติกรรมเพื่อชะลอการเกิดโรคเบาหวานและความดันโลหิตสูงในประชากรทั่วไป และชะลอการเกิดภาวะแทรกซ้อน หรือเสียชีวิตก่อนวัยอันควรในผู้ป่วย และข้อมูลที่ได้ยังสามารถนำไปคาดการณ์ผลกระทบทางสุขภาพและทางเศรษฐกิจในระดับมหภาครวมถึงในประชากรกลุ่มเปราะบางได้อีกด้วย

ดังนั้น ชุดการศึกษานี้จึงออกแบบมาเพื่อสร้างแบบจำลองทำนายความเสี่ยง ผ่านทางการคำนวณข้อมูลทางระบาดวิทยาของโรคเบาหวานและโรคความดันโลหิตสูงในประเทศไทยระหว่างปี พ.ศ. 2560 ถึง 2566 โดยใช้ข้อมูลจากฐานข้อมูลการใช้บริการสาธารณสุขระดับประเทศของฐานข้อมูล 43 แห่ง กระทรวงสาธารณสุข ผ่านทางการพัฒนาแบบจำลองเพื่อการคาดการณ์การเกิดโรคในประชากรทั่วไปและแบบจำลองทำนายความเสี่ยงการเกิดภาวะแทรกซ้อนประเภทต่าง ๆ และการเสียชีวิตในกลุ่มผู้ป่วยเพิ่มเติมด้วยวิธีการทางปัญญาประดิษฐ์แบบ Machine Learning ตรรกะและแบบจำลองที่ถูกสร้างขึ้นจะถูกนำไปใช้พัฒนาเครื่องมือประเมินผลกระทบทางสุขภาพรายบุคคลที่สามารถบรรจุในแอปพลิเคชันบนคอมพิวเตอร์หรือสมาร์ทโฟน และยังสามารถนำไปคาดการณ์ผลกระทบทางเศรษฐกิจอันเนื่องมาจากโรคเบาหวานและความดันโลหิตสูงที่จะเกิดขึ้นในอนาคตเพื่อประกอบการตัดสินใจเชิงนโยบายในการวางแผน ควบคุมและป้องกันรักษาโรคเบาหวานและความดันโลหิตสูงในประเทศไทยต่อไป

บทที่ 3 ระเบียบวิธีวิจัย

การศึกษานี้จะมีขั้นตอนทั้งหมด 3 ระยะที่เชื่อมโยงต่อกันดังต่อไปนี้

ระยะที่ 1 การทบทวนวรรณกรรมเพื่อคัดเลือกตัวแปรสำคัญสำหรับการสร้างแบบจำลองทำนายความเสี่ยง

ผู้วิจัยได้ทบทวนวรรณกรรมเกี่ยวข้องกับตัวแปรที่เคยมีการใช้ในการสร้างแบบจำลองทำนายความเสี่ยงโดยทำการสืบค้นข้อมูลจากฐานข้อมูลระดับนานาชาติโดยการสืบค้นจะมุ่งเป้าไปที่

1. รายการตัวแปรที่เกี่ยวข้องกับการเกิดโรคเบาหวานและความดันโลหิตสูงในกลุ่มประชากรทั่วไป
2. รายการตัวแปรที่เกี่ยวข้องกับการเกิดภาวะแทรกซ้อน ได้แก่ โรคหลอดเลือดหัวใจตีบ โรคเส้นเลือดสมองแตก หรือตีบ โรคไตวายเรื้อรัง และ ภาวะแทรกซ้อนทางตา
3. รายการตัวแปรที่เกี่ยวข้องกับการเสียชีวิตในกลุ่มผู้ป่วยโรคเบาหวานและความดันโลหิตสูง

เมื่อได้รับรายการปัจจัยที่เกี่ยวข้องจากการทบทวนวรรณกรรม ผู้วิจัยได้ประเมินความเป็นไปได้ของการนำปัจจัยไปใช้โดยวิเคราะห์ว่ามีการบันทึกข้อมูลตัวแปรดังกล่าวในฐานข้อมูล 43 แฟ้ม กระทรวงสาธารณสุขหรือไม่ โดยมีรายละเอียดดังตารางที่ 2⁴⁵ สำหรับตัวอย่างของรายการตัวแปรที่พบในการทบทวนวรรณกรรมแต่ถูกคัดออกไม่นำเข้าสู่การสร้างแบบจำลอง เป็นดังตัวอย่างที่แสดง

1. ตัวแปรเรื่องเชื้อชาติ (Caucasian Hispanic African American Asia) ที่พบจากการทบทวนวรรณกรรมถูกคัดออกจากการศึกษาเนื่องจาก บริบทไม่เกี่ยวข้องกับประเทศไทย
2. ตัวแปรเรื่องปริมาณกาแฟที่ดื่มต่อวัน ที่พบจากการทบทวนวรรณกรรมถูกคัดออกจากการศึกษาเนื่องจาก ไม่มีการเก็บข้อมูลตัวแปรดังกล่าวในฐานข้อมูล 43 แฟ้ม

ตารางที่ 2 แสดงรายการแฟ้มข้อมูลของฐานข้อมูล 43 แฟ้มที่นำมาใช้ในการสร้างแบบจำลองทำนายความเสี่ยง

| ประเภทของข้อมูล | แฟ้มโครงสร้างมาตรฐาน |
|--|---|
| ข้อมูลปัจจัยส่วนบุคคล เช่น เพศ และ อายุ | PERSON |
| ข้อมูลการเสียชีวิต | DEATH |
| ข้อมูลประวัติครอบครัวของโรคเบาหวานและความดันโลหิตสูง | NCDSSCREEN |
| ข้อมูลพฤติกรรมสุขภาพ เช่น การสูบบุหรี่ | NCDSSCREEN |
| ข้อมูลลักษณะทางกายภาพ เช่น น้ำหนัก ส่วนสูง | NCDSSCREEN |
| เส้นรอบเอว ค่าความดันโลหิต เป็นต้น | CHRONICFU |
| ข้อมูลสถานะสุขภาพและโรคประจำตัว | CHRONIC DIAGNOSIS_OPD DIAGNOSIS_IPD |

ตารางที่ 2 แสดงรายการเพิ่มข้อมูลของฐานข้อมูล 43 เพิ่มที่นำมาใช้ในการสร้างแบบจำลองทำนายความเสี่ยง
(ต่อ)

| ประเภทของข้อมูล | เพิ่มโครงสร้างมาตรฐาน |
|---|--|
| ข้อมูลผลการตรวจทางห้องปฏิบัติการ | LABFU NCDScreen สำหรับการคัดกรองระดับน้ำตาลในเลือดในประชากร |
| ข้อมูลยา | DRUG_OPD, IPD |
| ข้อมูลค่ารักษาพยาบาลที่เกี่ยวข้อง (ใช้ในการประเมินผลกระทบทางเศรษฐกิจ) | SERVICE ADMISSION |

ระยะที่ 2 การวิเคราะห์ทางระบาดวิทยาและการสร้างแบบจำลองทำนายความเสี่ยง

การศึกษานี้ใช้ข้อมูลจากฐานข้อมูล 43 เพิ่มของกระทรวงสาธารณสุข ตั้งแต่ปี พ.ศ. 2560 จนถึงเดือนกุมภาพันธ์ พ.ศ. 2566 มีคนไข้ประมาณ 29 ล้านคนที่มีการบันทึกไว้ในฐานข้อมูล การศึกษานี้มีขั้นตอนการจัดการและวิเคราะห์ข้อมูลดังต่อไปนี้

การจัดการข้อมูล

- ประเมินคุณภาพ ความครบถ้วนของข้อมูลแต่ละตัวแปรที่นำมาใช้ในการสร้างแบบจำลองทำนายความเสี่ยง สำหรับรายละเอียดวิธีการจัดการข้อมูลมีการระบุในรายละเอียดในเนื้อหาบทถัดไป
- กำหนดนิยามปฏิบัติการของการวินิจฉัยโรคเบาหวาน ความดันโลหิตสูง และภาวะแทรกซ้อนแต่ละประเภทสำหรับรายละเอียดวิธีการจัดการข้อมูลมีการระบุในรายละเอียดในเนื้อหาบทถัดไป
- จัดทำระบบฐานข้อมูลให้เป็นลักษณะของ Cohort Study แยกตามประเภทของคนไข้ คือ 1. ผู้ที่ยังไม่เกิดโรค และ 2. ผู้ป่วยที่เป็นโรคแล้วแต่ยังไม่มีภาวะแทรกซ้อน โดยมีการ subset ลงในแต่ละ Cohort ตามประเภทของโรค ดังนี้
 - Cohort 1.1: Cohort ของผู้ที่ยังไม่เป็นโรคเบาหวาน ณ จุดเริ่มต้นของ Cohort
 - Cohort 1.2: Cohort ของผู้ป่วยโรคเบาหวาน ณ จุดเริ่มต้นของ Cohort
 - Cohort 2.1: Cohort ของผู้ที่ยังไม่เป็นโรคความดันโลหิตสูง ณ จุดเริ่มต้นของ Cohort
 - Cohort 2.2: Cohort ของผู้ป่วยโรคความดันโลหิตสูง ณ จุดเริ่มต้นของ Cohort

การวิเคราะห์ข้อมูลเพื่อสร้างแบบจำลองทำนายความเสี่ยง

หลังจากการเตรียมข้อมูลและทำความสะอาดข้อมูล ในการวิเคราะห์ข้อมูลเพื่อสร้างตัวแบบจำลองทำนายความเสี่ยงโรคเบาหวาน โรคความดันโลหิตสูง และโรคแทรกซ้อนหรือเสียชีวิต จะประกอบไปด้วยขั้นตอนดังรูปที่ 2



รูปที่ 2 ขั้นตอนการวิเคราะห์ข้อมูลเพื่อสร้างแบบจำลองทำนายความเสี่ยง

จากรูปที่ 2 แสดงขั้นตอนในการสร้างตัวแบบจำลองทำนายความเสี่ยงโรคเบาหวาน โรคความดันโลหิตสูง และโรคแทรกซ้อนหรือเสียชีวิต ซึ่งมีรายละเอียดดังต่อไปนี้

1. นำข้อมูลมาแบ่งเป็น 2 ชุด โดยแบ่งข้อมูลออกเป็นข้อมูลชุดฝึกฝน (Training Set) ประมาณ 70% ของตัวอย่างข้อมูลผู้ป่วยทั้งหมด และข้อมูลชุดทดสอบ (Test Set) อีกประมาณ 30% ของตัวอย่างข้อมูลผู้ป่วยทั้งหมด โดยที่ผู้ป่วย 1 คน จะอยู่ได้เพียงชุดข้อมูลเพียงชุดเดียว คือ ชุดข้อมูลฝึกฝน หรือ ชุดข้อมูลทดสอบ
2. นำข้อมูลจากชุดฝึกฝนไปสร้างตัวแบบทำนาย สำหรับตัวแบบทำนายนี้จะใช้ตัวแบบประเภทความเสี่ยงเวลาแบบไม่ต่อเนื่อง (Discrete-Time Hazard Model) ซึ่งจะต้องเตรียมข้อมูลในการทำนายสถานะของโรคเบาหวาน โรคความดันโลหิตสูง และโรคแทรกซ้อนหรือเสียชีวิตในปีถัดไป จากตัวแปรทำนาย (Predictors) ต่างๆ ที่วัดได้ในปี

ก่อนหน้า หลังจากเตรียมข้อมูลพร้อมแล้ว จะสร้างตัวแบบด้วยเทคนิคการเรียนรู้ของเครื่องในการจำแนกประเภทข้อมูล (Classification) ในการสร้างตัวแบบนี้จะต้องคำนึงถึงการปรับพารามิเตอร์ต่างๆ และการคัดเลือกตัวแปรทำนายเข้าตัวแบบเพื่อให้ผลการทำนายมีประสิทธิภาพที่ดีขึ้น

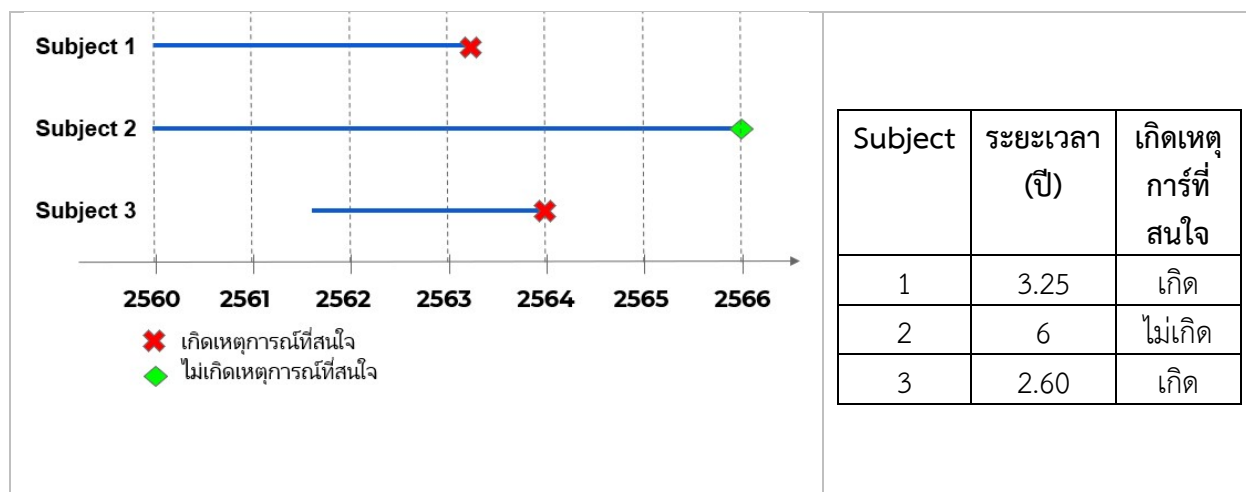
3. นำตัวแบบทำนายที่สร้างขึ้นจากข้อมูลชุดฝึกฝนไปพยากรณ์ค่าความเสี่ยงโรคเบาหวาน โรคความดันโลหิตสูง และโรคแทรกซ้อนหรือเสียชีวิตของข้อมูลในชุดทดสอบ
4. นำค่าพยากรณ์ที่ได้ในข้อ 3. และสถานะของโรคเบาหวาน โรคความดันโลหิตสูง และโรคแทรกซ้อนหรือเสียชีวิตที่เกิดขึ้นจริงในข้อมูลฝึกฝนไปคำนวณตัววัดประสิทธิภาพการทำนายต่าง ๆ
5. หากพบว่าตัวแบบทำนายมีประสิทธิภาพในการทำนายเป็นที่พึงพอใจแล้ว จะนำตัวแบบทำนายไปใช้งานต่อไป แต่หากพบว่าประสิทธิภาพในการทำนายยังไม่ดีพอ ให้กลับไปยังขั้นตอนที่ 2 เพื่อปรับพารามิเตอร์ต่าง ๆ และการคัดเลือกตัวแปรทำนายเข้าตัวแบบเพื่อให้ผลการทำนายมีประสิทธิภาพที่ดีขึ้น

สำหรับเนื้อหาการสร้างแบบจำลองทำนายความเสี่ยง ผู้วิจัยจะขอขยายความรายละเอียดในส่วนต่าง ๆ ดังต่อไปนี้
หัวข้อ 1 ตัวแบบความเสี่ยงเวลาแบบไม่ต่อเนื่อง
หัวข้อ 2 เทคนิคการเรียนรู้ของเครื่องในการจำแนกประเภทข้อมูล
หัวข้อ 3 ตัววัดประสิทธิภาพในการทำนาย

หัวข้อที่ 1 ตัวแบบความเสี่ยงเวลาแบบไม่ต่อเนื่อง

การวิเคราะห์การรอดชีพ (Survival Analysis) มีวัตถุประสงค์เพื่อพยากรณ์เวลาจนกระทั่งเกิดเหตุการณ์ที่สนใจ (Time-to-event Outcome) เช่น เวลาจนกระทั่งผู้เข้ารับการตรวจคัดกรองเบาหวานจะมีภาวะเป็นโรคเบาหวาน หรือเวลาที่ผู้ป่วยโรคเบาหวานจะเกิดภาวะแทรกซ้อนทางตา เป็นต้น ตัวแบบการรอดชีพมักจะสนใจพยากรณ์ความเสี่ยงหรือความน่าจะเป็นของการเกิดเหตุการณ์ที่สนใจของคนไข้แต่ละคน โดยอาศัยปัจจัยที่มีอิทธิพลต่อความน่าจะเป็นของการเกิดเหตุการณ์เป็นตัวแปรทำนายในตัวแบบ

ในการสร้างตัวแบบการรอดชีพนั้นจะต้องใช้ข้อมูลเวลาจนกระทั่งเกิดเหตุการณ์ที่สนใจของคนไข้ ซึ่งอาจพบว่าเกิดเหตุการณ์กับคนไข้บางคนในช่วงเวลาที่ติดตามคนไข้ (Follow-up Time) แต่ก็อาจจะมีคนไข้บางคนที่ยังไม่พบว่ามีเหตุการณ์ขึ้นในช่วงเวลาที่ติดตาม อย่างไรก็ตาม เราไม่สามารถสรุปได้ว่าในคนไข้ที่ยังไม่พบเหตุการณ์ที่เกิดขึ้นในช่วงเวลาที่ติดตามนั้นจะเกิดเหตุการณ์ที่สนใจหรือไม่เกิดเหตุการณ์ที่สนใจหลังจากช่วงเวลาติดตามคนไข้ ข้อมูลดังกล่าวจึงถือเป็นข้อมูลที่ไม่สมบูรณ์ (Incomplete Data) ซึ่งมักเรียกว่าข้อมูลตรวจตัด (Censored Data) ดังแสดงในรูปที่ 3 ซึ่งแสดงตัวอย่างข้อมูลเวลาจนกระทั่งเกิดเหตุการณ์ที่สนใจ จากรูปพบว่าข้อมูลของคนไข้คนที่ 2 เป็นข้อมูลตรวจตัด



รูปที่ 3 ตัวอย่างข้อมูลเวลาจนกระทั่งเกิดเหตุการณ์ที่สนใจ

วิธีการสร้างตัวแบบการรอดชีพมีหลากหลายวิธี ไม่ว่าจะเป็นวิธีการทางสถิติ หรือวิธีการเรียนรู้ของเครื่อง วิธีการทางสถิติมักจะใช้เวลาจนกระทั่งเกิดเหตุการณ์ที่สนใจเป็นตัวแปรเชิงปริมาณแบบต่อเนื่อง (Continuous Random Variable) เช่น การถดถอยการรอดชีพ (Survival Regression) ซึ่งถือเป็นวิธีการวิเคราะห์แบบอิงพารามิเตอร์ (Parametric Method) ที่มีการให้ข้อสมมติเกี่ยวกับการแจกแจงของเวลาจนกระทั่งเกิดเหตุการณ์ที่สนใจ การเลือกการแจกแจงที่เหมาะสมมีความสำคัญซึ่งอาจทำให้ผลพยากรณ์ของตัวแบบไม่มีประสิทธิภาพ วิธีการทางสถิติอีกวิธีหนึ่งที่ได้รับการนิยมนำมาใช้วิเคราะห์การรอดชีพ คือ ตัวแบบพีบีดีเชิงสัดส่วนของค็อกซ์ (Cox Proportional Hazards Model)⁴⁶ ซึ่งวิธีการวิเคราะห์แบบกึ่งอิงพารามิเตอร์ (Semi-parametric Method) ซึ่งไม่ต้องตั้งข้อสมมติเกี่ยวกับการแจกแจงของตัวแปรเวลา อย่างไรก็ตาม ตัวแบบพีบีดีเชิงสัดส่วนของค็อกซ์มีข้อสมมติให้ฟังก์ชันพีบีดี (Hazard Function) เมื่อตัวแปรทำนายเปลี่ยนไป มีความสัมพันธ์เชิงสัดส่วนและไม่ขึ้นอยู่กับเวลา ซึ่งในความเป็นจริงข้อสมมตินี้มักถูกละเมิดและอาจทำให้ผลพยากรณ์ของตัวแบบไม่มีประสิทธิภาพก็เป็นได้⁴⁷

สำหรับการใช้วิธีการเรียนรู้ของเครื่องในการวิเคราะห์การรอดชีพนั้นจะมีวัตถุประสงค์หลักเพื่อให้ได้ประสิทธิภาพการทำนายสูงสุดโดยไม่มีข้อกำหนดข้อสมมติต่างๆ อีกทั้งยังสามารถในการสร้างตัวแบบที่รองรับข้อมูลขนาดใหญ่ที่มีรูปแบบความสัมพันธ์ระหว่างตัวแปรทำนายและผลลัพธ์ที่มีความซับซ้อนได้ การใช้วิธีการเรียนรู้ของเครื่องในการวิเคราะห์การรอดชีพสามารถทำได้ 2 วิธี คือ 1) ใช้เวลาจนกระทั่งเกิดเหตุการณ์ที่สนใจเป็นตัวแปรเชิงปริมาณที่ต่อเนื่อง และ 2) แปลงตัวแปรเวลาจนกระทั่งเกิดเหตุการณ์ที่สนใจเป็นตัวแปรที่ไม่ต่อเนื่อง ซึ่งมักจะทำการแปลงเป็นตัวแปรแบบทวิ (Binary Variable) ดังนั้น การวิเคราะห์การรอดชีพตัวแปรเวลาแบบไม่ต่อเนื่อง (Discrete-time Survival Analysis) จึงมีข้อดี กล่าวคือ สามารถใช้อัลกอริทึมสำหรับการจำแนกประเภทใดๆ ในการวิเคราะห์ได้ ซึ่งปัจจุบันมีเทคนิคการจำแนกประเภทจำนวนมากที่ให้ผลการพยากรณ์ที่แม่นยำ

ในการแปลงเวลาจนกระทั่งเกิดเหตุการณ์ที่สนใจซึ่งเป็นตัวแปรแบบต่อเนื่องให้เป็นตัวแปรแบบไม่ต่อเนื่องสามารถทำได้โดยการแบ่งช่วงเวลาที่ต่อเนื่องออกเป็นช่วง ๆ เช่น แบ่งช่วงเวลาตั้งแต่ปี พ.ศ. 2560 – 2566 เป็นช่วงเวลารายปี 7 ช่วงเวลา ดังนี้

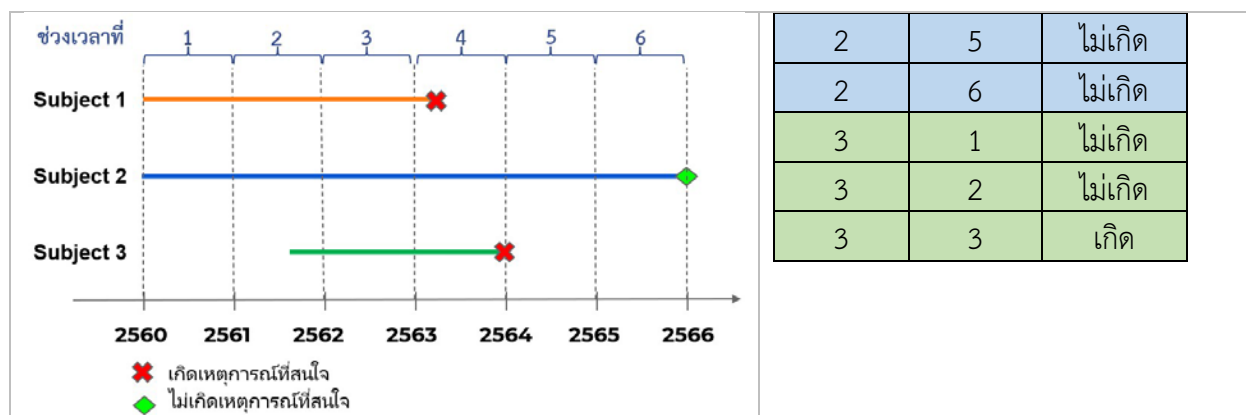
- 1) 1 มกราคม 2560 – 31 ธันวาคม 2560
- 2) 1 มกราคม 2561 – 31 ธันวาคม 2561
- 3) 1 มกราคม 2562 – 31 ธันวาคม 2562
- 4) 1 มกราคม 2563 – 31 ธันวาคม 2563
- 5) 1 มกราคม 2564 – 31 ธันวาคม 2564
- 6) 1 มกราคม 2565 – 31 ธันวาคม 2565
- 7) 1 มกราคม 2566 – 28 กุมภาพันธ์ 2566 (หมายเหตุ ข้อมูลในปี 2566 มีถึงเพียงเดือนกุมภาพันธ์)

หลังจากนั้นจึงนำข้อมูลของคนไข้แต่ละคน (Person Data) มาแปลงเป็นข้อมูลในลักษณะข้อมูลคนไข้แต่ละคนในแต่ละช่วงเวลา (Person-period Data) และแปลงตัวแปรเวลาเป็นข้อมูลทวิวาเกิดเหตุการณ์ที่สนใจในช่วงเวลาดังกล่าวหรือไม่ หากมีคนไข้ที่ไม่ได้เก็บข้อมูลเริ่มต้น ณ ตำแหน่งเวลาเดียวกัน จะนับจำนวนช่วงเริ่มตั้งแต่เวลา ณ ตำแหน่งที่เริ่มเก็บข้อมูล ดังแสดงในคนไข้คนที่ 3 จากรูปที่ 4

หากพิจารณาข้อมูลดั้งเดิม จะพบว่าข้อมูลมีเพียง 1 แถวสำหรับข้อมูลหนึ่งคน แต่หลังจากการแปลงข้อมูลแล้ว ข้อมูลของคนๆ หนึ่งจะถูกแปลงเป็นข้อมูลในหลายแถว ซึ่งหากพิจารณาแล้วข้อมูลหลายแถวที่มาจากบุคคลเดียวกันน่าจะไม่เป็นอิสระกัน อย่างไรก็ตาม งานวิจัยหลายงานแสดงให้เห็นว่าข้อมูลแต่ละแถวในระดับ Person-period Data เป็นอิสระต่อกัน ซึ่งทำให้ง่ายต่อการนำข้อมูลที่แปลงแล้วไปใช้เทคนิคการเรียนรู้ของเครื่องต่อไป^{48,49}

จากการทำการแปลงข้อมูลเวลาจนกระทั่งเกิดเหตุการณ์ที่สนใจให้เป็นตัวแปรที่ไม่ต่อเนื่องดังแสดงตัวอย่างในรูปที่ 4 และนำไปสร้างตัวแบบทำนายโอกาสหรือความเสี่ยงของการเกิดเหตุการณ์ที่สนใจในช่วงเวลาถัดไปหรือในตัวอย่างนี้ซึ่งแบ่งช่วงเวลาออกเป็นรายปีคือจะทำนายความเสี่ยงของการเกิดเหตุการณ์ที่สนใจในปีถัดไปนั่นเอง ตัวแบบที่ได้นี้จะเรียกว่าตัวแบบความเสี่ยงเวลาแบบไม่ต่อเนื่อง (Discrete-time Hazard Model)

| Subject | ระยะเวลา (ปี) | เกิดเหตุการณ์ที่สนใจ | Subject | ช่วงเวลา | เกิดเหตุการณ์ที่สนใจ |
|---------|---------------|----------------------|---------|----------|----------------------|
| 1 | 3.25 | เกิด | 1 | 1 | ไม่เกิด |
| 2 | 6 | ไม่เกิด | 1 | 2 | ไม่เกิด |
| 3 | 2.60 | เกิด | 1 | 3 | ไม่เกิด |
| | | | 1 | 4 | เกิด |
| | | | 2 | 1 | ไม่เกิด |
| | | | 2 | 2 | ไม่เกิด |
| | | | 2 | 3 | ไม่เกิด |
| | | | 2 | 4 | ไม่เกิด |



รูปที่ 4 ตัวอย่างการแปลงข้อมูลเวลาจนกระทั่งเกิดเหตุการณ์ที่สนใจให้อยู่ในรูปของเวลาที่ต่อเนื่อง

หัวข้อที่ 2 เทคนิคการเรียนรู้ของเครื่องในการจำแนกประเภทข้อมูล

เทคนิคการเรียนรู้ของเครื่องในการจำแนกประเภท ในงานวิจัยนี้ จะขอกกล่าวถึง 2 เทคนิคที่ใช้ในงานวิจัยชิ้นนี้ คือ

1. การถดถอยโลจิสติกแบบ Elastic Net (Lasso Logistic Regression)
2. วิธี XGBoost (Extreme Gradient Boosting)

ซึ่งรายละเอียดของวิธีทั้ง 2 จะได้กล่าวถึงในหัวข้อถัดไป ดังนี้

1. การถดถอยโลจิสติกแบบ Elastic Net

จากหัวข้อที่ 1 จะเห็นได้ว่าข้อมูลระยะเวลาจนกระทั่งเกิดเหตุการณ์ที่สนใจซึ่งเป็นข้อมูลรายบุคคล ได้ถูกแปลงเป็นข้อมูลของแต่ละบุคคลในแต่ละช่วงเวลา ซึ่งจะมีตัวแปรทวิ (Binary Variable) ที่บ่งบอกถึงสถานะของการเกิดเหตุการณ์ที่สนใจสำหรับบุคคลดังกล่าวในช่วงเวลาดังกล่าว จึงสามารถนำข้อมูลไปสร้างตัวแบบการถดถอยโลจิสติกทวิ (Binary Logistic Regression) ได้

ให้ ข้อมูลที่เก็บคือ (Y_i, X_i) เมื่อ $i = 1, \dots, n$

โดยที่ Y_i เป็นตัวแปรตามซึ่งบ่งบอกสถานะของการเกิดเหตุการณ์ของข้อมูลแถวที่ i

$$Y_i = \begin{cases} 1 & ; \text{เกิดเหตุการณ์} \\ 0 & ; \text{ไม่เกิดเหตุการณ์} \end{cases}$$

$X_i = (X_{i1}, \dots, X_{ip})$ เป็นเวกเตอร์ของตัวแปรทำนายขนาด p ของข้อมูลแถวที่ i

และ n เป็นจำนวนแถวในตารางข้อมูล

ให้ $\pi_i = P(Y_i = 1)$ หรือโอกาสที่จะเกิดเหตุการณ์ วิธีถดถอยโลจิสติกทวิสมมติให้ตัวแบบเขียนได้ดังสมการที่ (1)

$$\log\left(\frac{\pi_i}{1 - \pi_i}\right) = \beta_0 + \beta_1 X_{i1} + \dots + \beta_p X_{ip} \quad (1)$$

และสามารถประมาณค่าสัมประสิทธิ์การถดถอย $\beta = (\beta_0, \beta_1, \dots, \beta_p)$ ด้วยวิธีการภาวะความน่าจะเป็นสูงสุด (Maximum Likelihood Method) หรือคือการหาค่า β ที่ทำให้ฟังก์ชันในสมการที่ (2) สูงสุด

$$\hat{\beta} = \operatorname{argmax}_{\beta} \log L(\beta) \quad (2)$$

เมื่อ $\log L(\beta)$ คือฟังก์ชันล็อกภาวะน่าจะเป็น (Loglikelihood Function) ในการหาค่า β สมการที่ (2) สามารถเขียนได้ดังสมการที่ (3)

$$\hat{\beta} = \operatorname{argmin}_{\beta} -\log L(\beta) \quad (3)$$

เมื่อตัวแปรทำนาย X มีจำนวนมาก หรือ p มีขนาดใหญ่ จะมีความเป็นไปได้ว่าตัวแปรทำนายอาจมีความสัมพันธ์กันเองสูง ทำให้การหาค่าประมาณของ β ทำได้ยาก นอกจากนี้ วิธีการถดถอยโลจิสติกแบบทั่วไป จะไม่สามารถเลือกตัวแปรเข้าหรือออกตัวแบบอย่างอัตโนมัติ ซึ่งจะทำให้การแปลผลตัวแบบทำได้ยาก

สำหรับวิธีการถดถอยโลจิสติกแบบ Elastic Net เป็นวิธีการที่ต่อยอดมาจากวิธีการถดถอยโลจิสติกแบบทั่วไป เพื่อแก้ไขปัญหาในการประมาณค่าสัมประสิทธิ์การถดถอยในกรณีที่ตัวแปรทำนายมีจำนวนมากและมีความสัมพันธ์กันเองสูง นอกจากนี้ วิธีการถดถอยโลจิสติกแบบ Elastic Net ยังสามารถคัดเลือกตัวแปรเข้าหรือออกจากตัวแบบได้อย่างอัตโนมัติ โดยค่าสัมประสิทธิ์การถดถอยแบบ Elastic Net จะหาได้จากสมการที่ (4)

$$\hat{\beta}(\alpha) = \operatorname{argmin}_{\beta} \{-\log L(\beta) + \alpha \|\beta\|_1 + (1 - \alpha) \|\beta\|^2\} \quad (4)$$

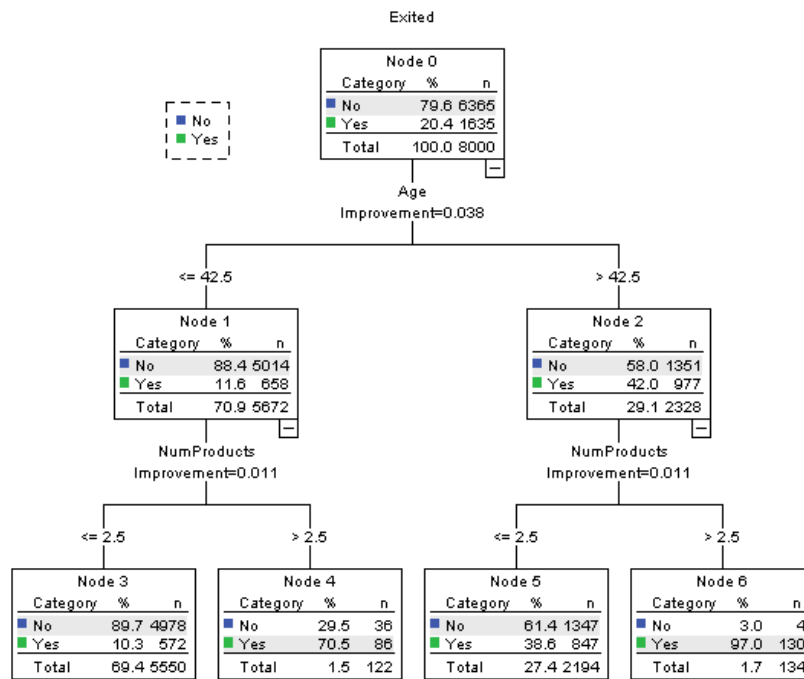
จากสมการที่ (4) จะปรับปรุงสมการจากสมการที่ (3) โดยเพิ่มเทอมฟังก์ชันการลงโทษ (Penalty Term หรือ Regularization Term) 2 เทอม ได้แก่ 1) $\alpha \|\beta\|_1 = \sum_{i=1}^p |\beta_i|$ และ 2) $(1 - \alpha) \|\beta\|^2 = \sum_{i=1}^p \beta_i^2$ โดยที่ α คือพารามิเตอร์ที่ใช้ปรับจูน หรือคือค่าถ่วงน้ำหนักของฟังก์ชันการลงโทษ โดยที่ $0 \leq \alpha \leq 1$

ค่าสัมประสิทธิ์การถดถอยที่ได้จากวิธีการถดถอยโลจิสติกแบบ Elastic Net จะให้ค่าสัมประสิทธิ์บางค่าเท่ากับ 0 หรือเทียบเท่ากับการที่ตัวแปรทำนายดังกล่าวไม่ได้มีผลต่อค่าพยากรณ์ความน่าจะเป็นของเหตุการณ์ที่จะเกิดขึ้น ดังนั้นวิธีการถดถอยโลจิสติกแบบ Elastic Net จึงสามารถทำการคัดเลือกตัวแปรโดยอัตโนมัติ

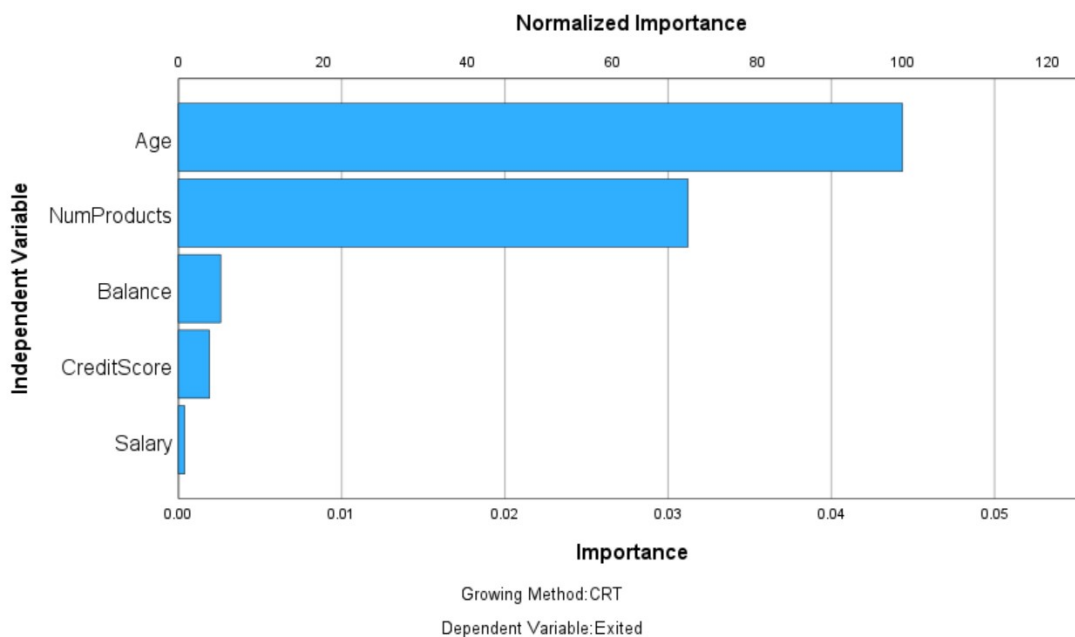
2. วิธี XGBoost

ต้นไม้ตัดสินใจเป็นเทคนิคการเรียนรู้ของเครื่องประเภทหนึ่ง ซึ่งสามารถทำนายค่าที่เป็นตัวเลขหรือจำแนกประเภทก็ได้ โดยผลลัพธ์ตัวแบบสามารถแสดงออกมาเป็นแผนผังต้นไม้ตัดสินใจได้ ในการสร้างต้นไม้ตัดสินใจ โดยทั่วไปจะอาศัยวิธีแบ่งแยกและเอาชนะ (Divide and Conquer Strategy) ผ่านกระบวนการ ค้นหาอย่าง

ละโมบ (Greedy Search) ในการค้นหาตัวแปรทำนายรวมถึงค่าในการแบ่งแยกที่เหมาะสมที่สุดซึ่งทำให้ข้อมูลหลังจากแบ่งแล้วมีความคล้ายคลึงกันมากที่สุด ตัวอย่างตัวแบบต้นไม้ตัดสินใจแสดงดังรูปที่ 5 และ ต้นไม้ตัดสินใจมีความสามารถในการคัดเลือกตัวแปรเข้าตัวแบบอัตโนมัติ ตัวแบบสามารถทำความเข้าใจได้ง่ายและสามารถแสดงความสำคัญของตัวแปรแต่ละตัวได้ดังรูปที่ 6



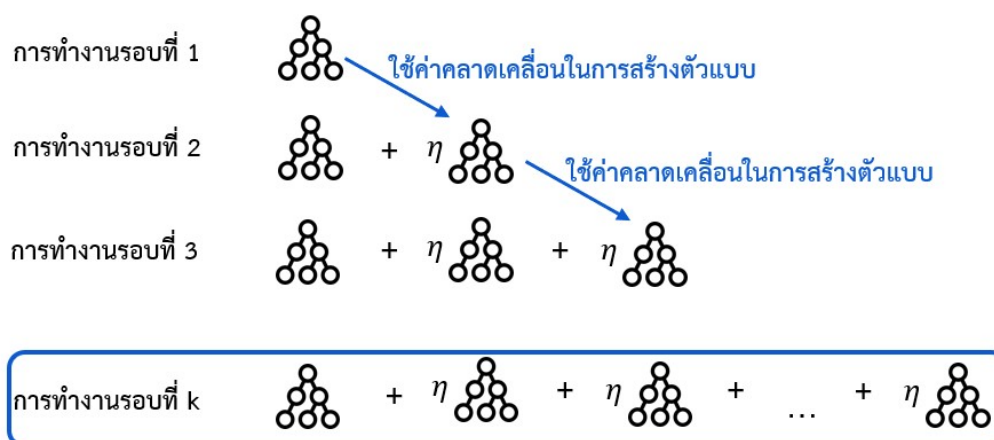
รูปที่ 5 ตัวอย่างต้นไม้ตัดสินใจ



รูปที่ 6 กราฟแสดงความสำคัญของตัวแปรจากตัวแบบต้นไม้ตัดสินใจ

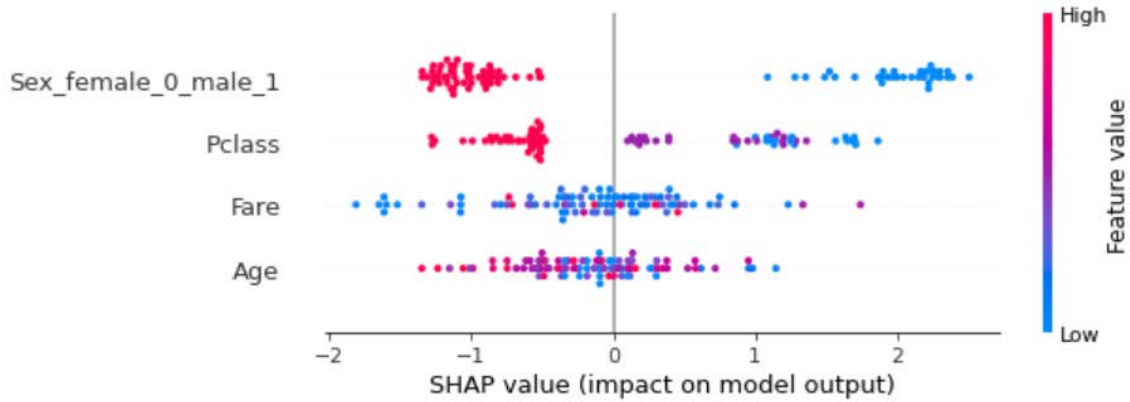
วิธีบูสตีง (Boosting) เป็นวิธีการสร้างตัวแบบจากการเรียนรู้จากผู้เรียนรู้ที่มีความแม่นยำต่ำ (Weak Learner) หลายๆ ตัวหรือในที่นี้คือต้นไม้ตัดสินใจหลายๆ ต้นที่มีเพียงไม่กี่ชั้นเพื่อป้องกันปัญหาพอดีเกินไป (Overfitting Problem) ทั้งนี้ เพื่อเพิ่มประสิทธิภาพการทำนายให้กับตัวแบบ วิธี XGBoost⁵⁰ ใช้กระบวนการบูสตีงในการสร้างต้นไม้ตัดสินใจหลายต้นต่อเนื่องกัน โดยต้นไม้ตัดสินใจที่สร้างขึ้นในแต่ละต้นจะเรียนรู้จากค่าความผิดพลาดที่ต้นไม้ต้นก่อนหน้าที่ยังไม่สามารถทำนายข้อมูลให้ถูกต้องได้ ผลลัพธ์ที่ได้จากวิธีการ XGBoost นี้จึงประกอบไปด้วยต้นไม้หลายๆ ต้นต่อเนื่องกัน ซึ่งหากจำนวนต้นไม้ต่อเนื่องกันมีความลึกมากพอ ผลลัพธ์ที่ได้จากตัวแบบในการทำนายก็จะมีแม่นยำมากขึ้น

กระบวนการการทำงานของ XGBoost แสดงดังรูปที่ 7 ซึ่งจะเห็นว่าในการทำงานแต่ละรอบจะใช้ค่าคลาดเคลื่อนในรอบก่อนหน้าในการสร้างตัวแบบต้นไม้ตัดสินใจ และหากดำเนินการวนซ้ำไปหลายๆ รอบ เช่น k รอบ ตัวแบบ XGBoost ที่จะนำไปพยากรณ์ก็คือตัวแบบที่วงไว้ในรูปที่ 7



รูปที่ 7 กระบวนการทำงานของวิธี XGBoost (η คืออัตราการเรียนรู้ (Learning Rate))

ในการแปลผลตัวแบบที่ได้จากวิธี XGBoost สามารถสร้างกราฟแสดงความสำคัญของตัวแปรแต่ละตัวได้ เช่นเดียวกับต้นไม้ตัดสินใจดังรูปที่ 6 นอกจากนี้ ในการแปลผลตัวแบบการเรียนรู้ของเครื่องต่าง ๆ ยังสามารถทำได้ผ่านค่า SHAP (Shapley Additive Explanations) ซึ่งเป็นตัววัดผลกระทบจากตัวแปรแต่ละตัวที่มีต่อค่าพยากรณ์ โดยเฉลี่ย ดังนั้นค่า SHAP จึงสามารถระบุความสำคัญของตัวแปรแต่ละตัว รวมถึงสามารถระบุทิศทางของความสัมพันธ์ได้ว่าตัวแปรแต่ละตัวมีความสัมพันธ์ไปในทิศทางเดียวกันหรือตรงกันข้ามกับค่าพยากรณ์ ดังแสดงดังรูปที่ 8



รูปที่ 8 ตัวอย่างกราฟแสดงค่า SHAP

หัวข้อที่ 3 ตัววัดประสิทธิภาพในการทำนาย

ตัวแบบการเรียนรู้ของเครื่องในการจำแนกประเภทที่ได้กล่าวถึงไปทั้ง 2 วิธีไม่ว่าจะเป็นวิธีการถดถอยโลจิสติกแบบ Elastic Net และวิธี XGBoost สามารถทำนายค่าความน่าจะเป็นหรือโอกาสที่เหตุการณ์ที่สนใจเกิดขึ้นได้ หรือก็คือสามารถพยากรณ์ $\pi_i = P(Y_i = 1)$ โดยจะนำตัวแบบที่สร้างจากข้อมูลชุดฝึกฝนไปทำนายค่าความน่าจะเป็นในชุดข้อมูลทดสอบเพื่อวัดประสิทธิภาพในการทำนาย

จากค่าความน่าจะเป็นที่ทำนายได้ สามารถเลือกใช้เกณฑ์หรือ Cutoff Value ที่เหมาะสมในการพยากรณ์สถานะการเกิดของเหตุการณ์ได้ เช่น หากเลือกใช้เกณฑ์ 0.5 นั้นหมายความว่า

- หากโอกาสที่พยากรณ์ได้ $\hat{\pi}_i \geq 0.5$ จะพยากรณ์สถานะของเหตุการณ์ว่าเหตุการณ์ที่สนใจเกิดขึ้น ($\hat{Y}_i = 1$)
- หากโอกาสที่พยากรณ์ได้ $\hat{\pi}_i < 0.5$ จะพยากรณ์สถานะของเหตุการณ์ว่าเหตุการณ์ที่สนใจไม่เกิดขึ้น ($\hat{Y}_i = 0$)

จากสถานะของเหตุการณ์ที่ทำนายได้ สามารถสร้างเมทริกซ์สับสน (Confusion Matrix) ได้ดังตารางที่ 3

ตารางที่ 3 เมทริกซ์สับสน

| | | ค่าพยากรณ์ \hat{Y} | |
|----------------|---|----------------------|----|
| | | 1 | 0 |
| ค่าจริง Y | 1 | TP | FN |
| | 0 | FP | TN |

โดยที่ TP เรียกว่า จำนวนผลบวกจริง (True Positive)

TN เรียกว่า จำนวนผลลบจริง (True Negative)

FP เรียกว่า จำนวนผลบวกหลง (False Positive)

FN เรียกว่า จำนวนผลลบหลง (False Negative)

จากเมตริกซ์สับสนที่ได้ จะสามารถวัดประสิทธิภาพผ่านตัววัดต่างๆ ดังนี้

- ความเที่ยง (Precision)

$$\text{Precision} = \frac{TP}{TP + FP}$$

- ความไว (Sensitivity หรือ Recall)

$$\text{Recall} = \frac{TP}{TP + FN}$$

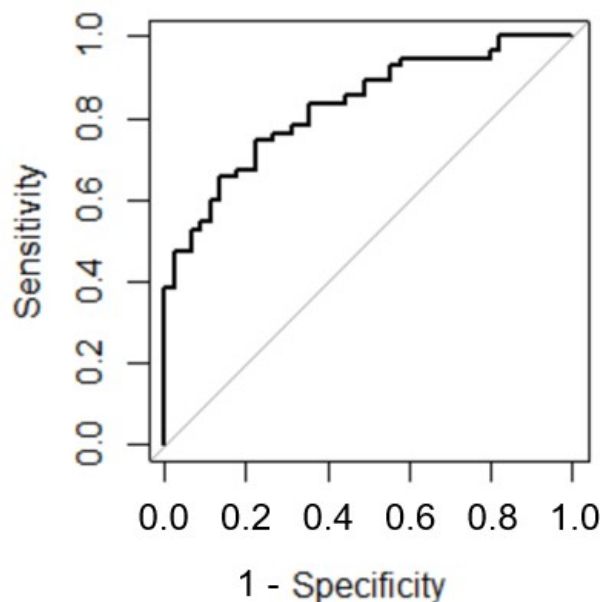
- ความจำเพาะ (Specificity)

$$\text{Specificity} = \frac{TN}{TN + FP}$$

- F1-score

$$\text{F1 - score} = \frac{2(\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}}$$

อย่างไรก็ตาม การวัดประสิทธิภาพการทำนายด้วยตัววัดข้างต้นทั้งหมดนั้น ยังมีข้อจำกัด กล่าวคือ ผลการทำนายสถานะการเกิดเหตุการณ์นั้นขึ้นอยู่กับทางเลือกใช้เกณฑ์หรือ Cutoff Value ที่เหมาะสม ดังนั้น หากจะวัดประสิทธิภาพการทำนายในภาพรวมโดยพิจารณาเกณฑ์ทั้งหมดที่เป็นไปได้นั้นสามารถทำได้โดยใช้กราฟ ROC (Receiver Operating Characteristic Curve) ซึ่งเป็นการพล็อตระหว่างค่าความไว (หรือที่เรียกว่าอัตราผลบวกจริง (True Positive Rate)) และค่า (1 - ความจำเพาะ) (หรือที่เรียกว่าอัตราผลบวกหลง (False Positive Rate)) ที่ Cutoff Value ต่างๆ ดังแสดงตัวอย่างในรูปที่ 9 และใช้ตัววัดประสิทธิภาพการทำนายออกมาเป็นพื้นที่ใต้กราฟ ROC หรือมักเรียกว่าค่า AUC (Area Under ROC Curve) ซึ่งมีค่าอยู่ระหว่าง 0 ถึง 1 ยิ่งค่า AUC มากนั้นหมายความว่าตัวแบบมีความสามารถในการจำแนกตัวแปรตามทั้งสองกลุ่มได้ดีขึ้น หรือกล่าวอีกนัยหนึ่งคือมีความสามารถในการทำนายที่ดีขึ้นนั่นเอง



รูปที่ 9 ตัวอย่างกราฟ ROC

นอกจากค่า AUC แบบดั้งเดิมซึ่งวัดประสิทธิภาพการทำนายในชุดข้อมูลทดลอง เนื่องจากเป็นการศึกษานี้เก็บข้อมูลติดตามของแต่ละบุคคลในแต่ละปี จึงต้องการที่จะวัดประสิทธิภาพการทำนายรายบุคคลผ่านตัววัดผล AUC ที่คำนวณด้วยความน่าจะเป็นสะสมของบุคคลตลอด Cohort (Cumulative Probability) กับสถานะอุบัติการณ์ปีสุดท้ายใน Cohort ของบุคคลนั้น เรียกตัววัดผลนี้ว่า Cumulative AUC โดยหากกำหนดให้ T_i คือจำนวนปีที่บุคคลที่ i ที่อยู่ใน Cohort และ $P(Y_{it} = 1)$ คือโอกาสที่จะเกิดเหตุการณ์ในปีที่ t ความน่าจะเป็นสะสมของบุคคล i คำนวณได้ดังนี้

$$\text{Cumulative Probability} = 1 - \left(\prod_{t=1}^{T_i} \{1 - P(Y_{it} = 1)\} \right)^{1/T_i}$$

ระยะที่ 3 การประเมินผลกระทบทางเศรษฐกิจอันเนื่องมาจากโรคเบาหวานและความดันโลหิตสูง

เพื่อให้ทราบถึงผลกระทบทางเศรษฐกิจอันเนื่องมาจากโรคเบาหวานและความดันโลหิตสูง การศึกษานี้ประเมินค่าใช้จ่ายทางสุขภาพที่เกิดขึ้นในกลุ่มประชากร 4 กลุ่มหลัก ประกอบด้วย 1. ผู้ที่ยังไม่เป็นโรคเบาหวานและความดันโลหิตสูง 2. ผู้ป่วยโรคเบาหวานเพียงอย่างเดียว 3. ผู้ป่วยโรคความดันโลหิตสูงเพียงอย่างเดียว และ 4. ผู้ป่วยโรคเบาหวานและความดันโลหิตสูง และมีการประเมินความแตกต่างของค่าใช้จ่ายทางสุขภาพเพิ่มเติมในปัจจุบันสำคัญที่คาดว่าจะส่งผลต่อค่าใช้จ่าย ได้แก่ 1. เพศ 2. ช่วงอายุ และ 3. ประเภทและจำนวนภาวะแทรกซ้อน

ข้อมูลที่ใช้ในการคำนวณได้จากการติดตามกลุ่มตัวอย่างเป็นระยะเวลา 6 ปี ระหว่างเดือนมกราคม พ.ศ. 2560 ถึง ธันวาคม พ.ศ. 2565 และข้อมูลค่าใช้จ่ายที่เกิดขึ้นในแต่ละปีถูกนำมาปรับมูลค่าปีฐานตามค่า Consumer Price Index ของเดือนกันยายน ปี พ.ศ. 2566 โดยใช้ข้อมูลหมวดค่าตรวจรักษาและค่ายา (MEDICAL CARE) และรวมทุกรายการ (ALL COMMODITIES) ในการปรับมูลค่าปีฐานของค่ารักษาพยาบาล และ ค่าใช้จ่ายอื่นๆ ที่เกี่ยวข้องตามลำดับ⁵¹ สำหรับองค์ประกอบของค่าใช้จ่ายทางสุขภาพถูกแบ่งออกเป็น 2 ประเภทหลักดังนี้

1. ต้นทุนทางตรง (Direct Cost) ประกอบด้วย

a. ต้นทุนทางตรงทางการแพทย์ (Direct Medical Cost)

คำนวณจากค่ารักษาพยาบาลผู้ป่วยนอกและผู้ป่วยในทั้งหมดที่มีการบันทึกไว้ในฐานข้อมูล 43 แฟ้ม⁴⁵ ประกอบด้วย ค่าบริการทางการแพทย์ ค่าตรวจทางห้องปฏิบัติการ ค่ายา และ ค่าใช้จ่ายในการทำหัตถการ เป็นต้น

b. ต้นทุนทางตรงที่ไม่เกี่ยวข้องกับการแพทย์ (Direct Non-Medical Cost)

ประกอบด้วย ค่าอาหาร และ ค่าเดินทางมารับบริการผู้ป่วยนอกและผู้ป่วยใน อ้างอิงจากรายการต้นทุนมาตรฐานเพื่อการประเมินเทคโนโลยีด้านสุขภาพ โครงการประเมินเทคโนโลยีและนโยบายด้านสุขภาพ กระทรวงสาธารณสุข⁵²

2. ต้นทุนทางอ้อม (Indirect cost) ประกอบด้วย

a. ค่าเสียโอกาสของผู้ป่วย (Patient's Productivity loss) เนื่องจากต้องลางานมารับการรักษาโรค โดย

i. ระยะเวลาที่ใช้ในการรับบริการผู้ป่วยนอก (รวมระยะเวลาที่ใช้ในการเดินทาง) อ้างอิงจากรายการต้นทุนมาตรฐานเพื่อการประเมินเทคโนโลยีด้านสุขภาพ

ii. ระยะเวลาที่ใช้ในการรับบริการผู้ป่วยใน คำนวณจากจำนวนวันนอนเฉลี่ยในการรับการรักษาตัวในโรงพยาบาลซึ่งถูกบันทึกไว้ในฐานข้อมูล 43 แฟ้ม และ ระยะเวลาที่ใช้ในการเดินทางมาโรงพยาบาล อ้างอิงจากรายการต้นทุนมาตรฐานเพื่อการประเมินเทคโนโลยีด้านสุขภาพ

b. ค่าเสียโอกาสของผู้ดูแล (Informal Care Cost) โดยอ้างอิงจากจำนวนชั่วโมงที่ใช้ในการดูแลผู้ป่วยโรคเบาหวานและ/หรือความดันโลหิตสูง ซึ่งได้จากการทบทวนวรรณกรรมทั้งในไทยและต่างประเทศ

53,54

สำหรับการคำนวณค่าเสียโอกาส ทางผู้วิจัยใช้ค่ารายได้ประชาชาติต่อหัว (GNI per capita) ณ ปี พ.ศ. 2565 จากฐานข้อมูล World Bank⁵⁵ เป็นตัวแทนของค่าเสียโอกาส ดังสมการนี้

$$\text{ค่าเสียโอกาสต่อชั่วโมง} = \frac{\text{รายได้ประชาชาติต่อหัวต่อปี}}{365 \text{ วัน} \times 8 \text{ ชั่วโมง}}$$

นอกจากนี้ เพื่อให้สะท้อนภาพรวมของค่าใช้จ่ายทางสุขภาพ ผู้วิจัยเลือกใช้ค่าเฉลี่ยระหว่างข้อมูลของโรงพยาบาลชุมชนและโรงพยาบาลทั่วไปของจากรายการต้นทุนมาตรฐานเพื่อการประเมินเทคโนโลยีด้านสุขภาพ ยกตัวอย่างเช่น ข้อมูลเวลาที่ใช้ทั้งหมดในการรับบริการที่หน่วยผู้ป่วยนอกของโรงพยาบาลชุมชน และ โรงพยาบาลทั่วไปที่ระบุในรายการต้นทุนมาตรฐาน คือ 175 และ 361 นาที ตามลำดับ ส่งผลให้ ข้อมูลเวลาที่ใช้ทั้งหมดในการรับบริการที่หน่วยผู้ป่วยนอกของการศึกษานี้ คือ $\frac{175+361}{2} = 268$ นาที

บทที่ 4 ผลการศึกษา

แบบจำลองที่ 1 ทำนายความเสี่ยงการเกิดโรคเบาหวานในประชากรทั่วไป จาก Cohort ของผู้ที่ยังไม่เป็นโรคเบาหวาน (Cohort 1.1)

ในการสร้าง Cohort ของผู้ที่ยังไม่เป็นโรคเบาหวานเมื่อเริ่มต้น Cohort จำเป็นต้องระบุให้แน่ชัดถึงเกณฑ์การพิจารณาสถานะอุบัติการณ์การเป็นเบาหวานของคนใช้แต่ละคน จึงจะสร้างฐานข้อมูลของ Cohort 1.1 ได้ นอกจากนี้ ในการสร้างแบบจำลองทำนายความเสี่ยงการเกิดโรคเบาหวานในประชากรที่เข้ามาคัดกรองโรคเบาหวานนั้น จะต้องนำตัวแปรอื่นๆ ที่คาดการณ์ว่าจะมีความสัมพันธ์กับความเสี่ยงที่จะเป็นโรคเบาหวานมาพิจารณาเพื่อสร้างแบบทำนายให้มีความแม่นยำมากขึ้น อนึ่ง ในการทำงานกับข้อมูลจริง พบว่าข้อมูลจริงมีความซับซ้อน และไม่ครบถ้วน เช่น คนไข้บางคนอาจจะไม่ได้เข้ามาคัดกรองในทุกๆ ปี นอกจากนี้ ยังมีข้อมูลที่ขาดหายไปหรือมีค่าผิดปกติเกินเกณฑ์ เป็นต้น ซึ่งในบทนี้จะได้กล่าวถึงเกณฑ์ที่ใช้ในการกำหนดสถานะอุบัติการณ์การเป็นเบาหวาน ตลอดจนการเตรียมข้อมูลตัวแปรต่างๆ ที่ใช้ในตัวแบบ จนถึงผลลัพธ์ที่ได้จากตัวแบบเพื่อนำไปอภิปรายผลต่อไป

การเตรียมตัวแปรสถานะอุบัติการณ์การเป็นเบาหวาน

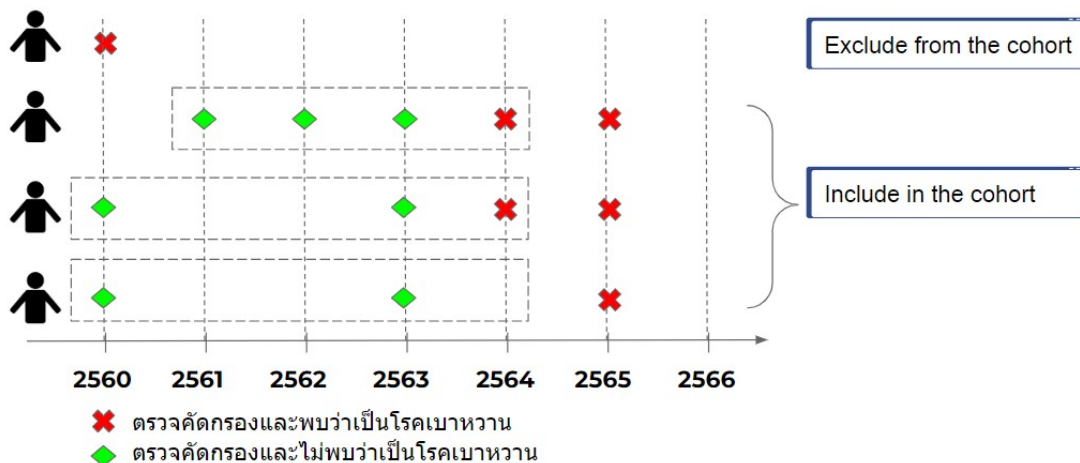
1. เกณฑ์ที่ใช้กำหนดสถานะอุบัติการณ์การเป็นเบาหวาน

ในการกำหนดสถานะอุบัติการณ์การเป็นเบาหวาน ผู้ป่วยจะต้องมีประวัติการวินิจฉัยการเป็นโรคเบาหวาน (ICD10: E11) จากการรักษาผู้ป่วยนอก (OP) หรือการรักษาผู้ป่วยใน (IP) หรือข้อมูลโรคเรื้อรัง (แฟ้ม CHRONIC) นอกจากนี้ ผู้ป่วยที่ได้รับการวินิจฉัยตามข้างต้นจะต้องมีลักษณะตรงตามเงื่อนไขดังต่อไปนี้อย่างน้อย 1 อย่างในปีเดียวกับที่มีการวินิจฉัยโรค

- (1) มีระดับน้ำตาลในเลือดหลังอดอาหาร ≥ 126 mg/dl
- (2) มีระดับน้ำตาลสะสม $\geq 6.5\%$
- (3) ได้รับการรักษาเบาหวานจากการรักษาผู้ป่วยนอก (OP) หรือผู้ป่วยใน (IP)

2. การจัดการและทำความสะอาดตัวแปรตาม

หลังจากกำหนดสถานะอุบัติการณ์การเป็นเบาหวานแล้ว ในการสร้าง Cohort 1.1 จะดึงข้อมูลเฉพาะบุคคลที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคเบาหวานเมื่อเริ่มต้น Cohort เท่านั้น หากบุคคลใดเข้ารับการคัดกรองและตรวจพบว่าเป็นโรคเบาหวาน จะไม่นำข้อมูลของบุคคลดังกล่าวเข้ามาสร้าง Cohort 1.1 ดังแสดงในรูปที่ 10

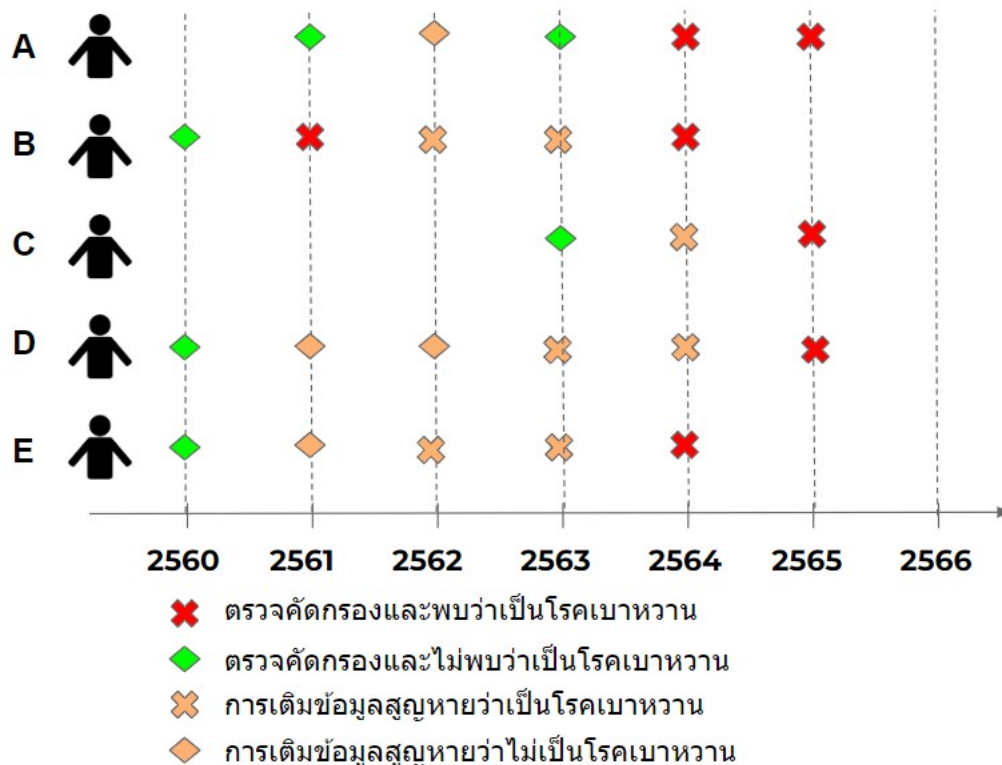


รูปที่ 10 บุคคลที่นำเข้าในการสร้าง Cohort ของผู้ที่ยังไม่เป็นโรคเบาหวาน (Cohort 1.1)

เนื่องจากการวิเคราะห์ข้อมูลจะทำนายสถานะอุบัติการณ์การเป็นเบาหวานแบบรายปี ในการเตรียมข้อมูลเพื่อนำไปสร้างตัวแบบทำนายกรณีที่มีบุคคลคนหนึ่งมีการตรวจคัดกรองมากกว่า 1 ครั้ง ต่อปี หากพบว่าบุคคลคนหนึ่งตรวจพบว่าไม่เป็นโรคเบาหวาน และเป็นโรคเบาหวาน (ทั้งชนิดไม่มีโรคแทรกซ้อน หรือชนิดที่มีโรคแทรกซ้อน) จะกำหนดสถานะอุบัติการณ์การเป็นเบาหวานของปีนั้นว่าเป็นโรคเบาหวาน

นอกจากนี้ บุคคลที่เข้าตรวจคัดกรองอาจจะไม่ได้เข้ามาตรวจคัดกรองทุกปี ข้อมูลจึงมีการขาดหาย ระหว่างปี ซึ่งจะต้องประมาณค่าและเติมข้อมูลที่สูญหายไป (Data Imputation) สำหรับขั้นตอนในการเติมข้อมูล สถานะอุบัติการณ์การเป็นเบาหวานที่สูญหายไปเป็นดังต่อไปนี้

- (1) กรณีที่ปีที่ข้อมูลสูญหายอยู่ระหว่างปีที่พบว่าไม่เป็นโรคเบาหวานทั้ง 2 ปี จะเติมข้อมูลสถานะอุบัติการณ์ที่สูญหายไปเป็นสถานะไม่เป็นโรคเบาหวาน (รูปที่ 11 - A)
- (2) กรณีที่ปีที่ข้อมูลสูญหายอยู่ระหว่างปีที่พบว่าเป็นโรคเบาหวานทั้ง 2 ปี จะเติมข้อมูลสถานะอุบัติการณ์ที่สูญหายไปเป็นสถานะเป็นโรคเบาหวาน (รูปที่ 11 - B)
- (3) กรณีที่ข้อมูลสูญหาย 1 ปี ซึ่งปีที่ข้อมูลสูญหายอยู่ระหว่างปีที่พบว่าไม่เป็นโรคเบาหวานและเป็นโรคเบาหวาน จะเติมข้อมูลสถานะอุบัติการณ์ที่สูญหายไปเป็นสถานะเป็นโรคเบาหวาน (รูปที่ 11 - C)
- (4) กรณีที่ข้อมูลสูญหายมากกว่า 1 ปี ซึ่งปีที่ข้อมูลสูญหายอยู่ระหว่างปีที่พบว่าไม่เป็นโรคเบาหวานและเป็นโรคเบาหวาน จะเติมข้อมูลสถานะอุบัติการณ์ ณ ปีที่ข้อมูลสูญหายกึ่งหนึ่งเป็นสถานะไม่เป็นโรคเบาหวาน และอีกกึ่งหนึ่งเป็นสถานะโรคเบาหวาน ในกรณีที่จำนวนปีที่ข้อมูลสูญหายเป็นเลขคี่ จะเติมข้อมูลสถานะอุบัติการณ์เริ่มจากปีกึ่งกลางที่สูญหายเป็นสถานะโรคเบาหวาน (รูปที่ 11 - D และ E)



รูปที่ 11 การเติมสถานะอุบัติการณ์การเป็นเบาหวานสำหรับข้อมูลสูญหาย

การเตรียมตัวแปรต่างๆ เพื่อใช้ทำนายอุบัติการณ์การเป็นเบาหวาน

1. ตัวแปรต่างๆ ที่ใช้ในการทำนายอุบัติการณ์การเป็นเบาหวาน

จากการทบทวนวรรณกรรม ทางทีมผู้วิจัยได้เตรียมข้อมูลตัวแปรต่างๆ ที่คาดการณ์ว่ามีความสัมพันธ์กับความเสี่ยงของการเกิดโรคเบาหวาน ซึ่งจะนำไปใช้เป็นตัวแปรอิสระในการสร้างตัวแบบเพื่อทำนายอุบัติการณ์การเป็นเบาหวานในปีถัดไป ทั้งนี้ รายละเอียดของตัวแปรต่างๆ ที่คาดการณ์ว่ามีความเสี่ยงต่อการเป็นโรคเบาหวานและมีอยู่ในฐานข้อมูล 43 แฟ้มของกระทรวงสาธารณสุข จะแสดงดังตารางที่ 4

ตารางที่ 4 ตัวแปรที่นำมาพิจารณาในการสร้างตัวแบบทำนายความเสี่ยงของการเกิดโรคเบาหวาน

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ในตัวแบบ | Reference |
|-----------------------------|---|--|-----------|
| Demographic | | | |
| Age | อายุ (ปี) | age | 56-63 |
| Age category | ช่วงอายุ - 35 - 39 ปี - 40 - 44 ปี - 45 - 49 ปี - >=50 ปี | age_cat_1 age_cat_2 age_cat_3 | 43 |
| Gender | เพศ - ชาย - หญิง | t.sex_2 | 43 |
| Health Behavior | | | |
| Smoking | พฤติกรรมการสูบบุหรี่ - สูบ - ไม่สูบ | SMOKE_1 | 58,62-64 |
| Physical measurement | | | |
| Height | ส่วนสูง (เซนติเมตร) | t.height | 65 |
| BMI | ดัชนีมวลกาย | BMI | 59 |
| BMI_CAT1 | ดัชนีมวลกาย - < 23 - 23 - 27.5 - >= 27.5 | bmi_cat_1_2 bmi_cat_1_3 | 43 |
| BMI_CAT2 | ดัชนีมวลกาย - < 18.5 - 18.5 - 22.9 - 23 - 24.9 - 25 - 29.9 - >= 30 | bmi_cat_2_2 bmi_cat_2_3 bmi_cat_2_4 bmi_cat_2_5 | 66 |
| WAIST_CM | เส้นรอบเอว (เซนติเมตร) | WAIST_CM | 56-58,61 |

ตารางที่ 4 ตัวแปรที่นำมาพิจารณาในการสร้างตัวแบบทำนายความเสี่ยงของการเกิดโรคเบาหวาน (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ในตัวแบบ | Reference |
|---------------------------|---|--------------------------|---|
| WAIST_CAT | เส้นรอบเอว - ปกติ - ผิดปกติ เกณฑ์ผิดปกติในเพศชาย คือ ≥ 90 เซนติเมตร และในหญิงคือ ≥ 80 เซนติเมตร | WAIST_CM_cat_2 | 43 |
| SBP | Systolic Blood Pressure (mm Hg) | SBP | 56,57,59,60 |
| DBP | Diastolic Blood Pressure (mm Hg) | DBP | 59 |
| Family History | | | |
| Family Hx of DM | ประวัติเบาหวานในญาติ สายตรง - มี - ไม่มี/ไม่ทราบ | DMFAMILY_1 | 43,56,57,59,60,62,63 |
| Family Hx of HT | ประวัติความดันโลหิตสูง ในญาติสายตรง - มี - ไม่มี/ไม่ทราบ | HTFAMILY_1 | No references: Data was recorded in NCDScreen Database, try test for associations |
| Underlying Disease | | | |
| Hypertension | ภาวะความดันโลหิตสูง - ปกติ - ผิดปกติ เกณฑ์ผิดปกติเมื่อ SBP ≥ 140 หรือ DBP ≥ 90 | HT_1 | 43,56,58,61,63 |

ตารางที่ 4 ตัวแปรที่นำมาพิจารณาในการสร้างตัวแบบทำนายความเสี่ยงของการเกิดโรคเบาหวาน (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ในตัวแบบ | Reference |
|--------------------|--|----------------------------------|---|
| Hx of high glucose | มีประวัติระดับน้ำตาลสูง - มี - ไม่มี/ไม่ทราบ | hx_of_high_glucose | 56,58,61 |
| Lab | | | |
| Bs_level | ระดับน้ำตาลในเลือด (mg/dL) | t.bslevel | 56,57,59,60 |
| Is_na_bslevel* | ทราบค่าระดับน้ำตาลในเลือดหรือไม่ - ทราบ - ไม่ทราบ | is_na_bslevel_1 | For adjusting data |
| Fasting Glucose | ประเภทการอดอาหารก่อนเจาะเลือด) - อดอาหารก่อน - ไม่ได้อดอาหาร/ไม่ทราบ | fasting_glucose | For adjusting data, NCDScreen recorded both fasting and non-fasting glucose |
| FG_cat | ระดับน้ำตาล (อดอาหารก่อนเจาะเลือด) - < 100 mg/dL - 100 - 125 mg/dL - > 125 mg/dL - ไม่ทราบ | FG_cat_1 FG_cat_2 FG_cat_3 | 43 |
| TG | ระดับไขมันไตรกลีเซอไรด์ (mg/dL) | TG_labresult | 57,59 |
| TG_CAT1 | ระดับไขมันไตรกลีเซอไรด์ - < 200 mg/dL - >=200 mg/dL - ไม่ทราบ | TG_cat_1_1 TG_cat_1_2 | 43 |

ตารางที่ 4 ตัวแปรที่นำมาพิจารณาในการสร้างตัวแบบทำนายความเสี่ยงของการเกิดโรคเบาหวาน (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ในตัวแบบ | Reference |
|------------|---|----------------------------|---------------------------|
| TG_CAT2 | ระดับไขมันไตรกลีเซอไรด์ - < 150 mg/dL - >= 150 mg/dL - ไม่ทราบ | TG_cat_2_1 TG_cat_2_2 | Triglyceride normal range |
| Is_na_TG* | ทราบค่าระดับไขมันไตรกลีเซอไรด์หรือไม่ - ทราบ - ไม่ทราบ | is_na_TG_1 | For adjusting data |
| HDL | ระดับไขมันดี (mg/dL) | HDL_labresult | 56,57,59,60 |
| HDL_CAT1 | ระดับไขมันดี - ผิดปกติ - ปกติ - ไม่ทราบ เกณฑ์ผิดปกติในเพศชายคือ < 40 mg/dL และในเพศหญิงคือ < 50 mg/dL | HDL_cat_1_1 HDL_cat_1_2 | |
| Is_na_HDL* | ทราบค่าระดับไขมันดีหรือไม่ - ทราบ - ไม่ทราบ | is_na_HDL_1 | For adjusting data |

หมายเหตุ * หมายถึงตัวแปรเพิ่มเติมที่ได้จากการเตรียมข้อมูล

จากตารางที่ 4 จะเห็นว่าตัวแปรเชิงปริมาณจำนวนหนึ่งมีการนำไปจัดกลุ่ม เช่น BMI HDL และ TG เป็นต้น นอกจากนั้น ยังพบว่าตัวแปรเชิงปริมาณที่นำไปจัดกลุ่มนั้นอาจมีวิธีการมากกว่า 1 วิธีในการจัดกลุ่มก็เป็นได้ ทั้งนี้ ในการสร้างตัวแบบทำนายจะพิจารณาตัวแปรทำนายทั้งหมด ไม่ว่าจะเป็นตัวแปรเชิงปริมาณหรือตัวแปรเชิงคุณภาพซึ่งได้จากการจัดกลุ่มจากตัวแปรตัวเดียวกัน โดยจะหาเซตย่อยของตัวแปรทำนายทั้งหมดที่ทำให้ได้ตัวแบบที่ดีมีความทำนายแม่นยำที่สุด

2. การจัดการและการทำความสะอาดตัวทำนาย

จากการตรวจสอบข้อมูลตัวแปรทำนาย พบปัญหา 2 กรณี ได้แก่ (1) ปัญหาค่านอกเกณฑ์ (Outlier) หรือค่าที่มีความผิดปกติหรือแตกต่างไปจากค่าสังเกตส่วนใหญ่ และ (2) ปัญหาข้อมูลสูญหาย (Missing Data) ดังนั้น จึงต้องมีการจัดการทำความสะอาดข้อมูลดังกล่าว ดังนี้

(1) การจัดการปัญหาค่านอกเกณฑ์ (Outlier)

ในการตรวจสอบค่านอกเกณฑ์สำหรับตัวแปรเชิงปริมาณ จะพิจารณาเฉพาะตัวแปรอายุ และตัวแปรเชิงปริมาณในหมวดของ Physical measurement ซึ่งมีโอกาสในการวัดที่คลาดเคลื่อนได้ โดยใช้เกณฑ์ในการตรวจจับค่านอกเกณฑ์ดังต่อไปนี้

- ตัวแปรอายุ (Age)

ใช้เกณฑ์ตรวจจับค่านอกเกณฑ์เช่นเดียวกับการใช้ Box Plot ในการตรวจสอบค่านอกเกณฑ์ กล่าวคือ หากข้อมูลไม่ตกอยู่ในช่วง

$$(Q_1 - 1.5 \times IQR, Q_3 + 1.5 \times IQR)$$

เมื่อ Q_1 คือ ควอร์ไทล์ที่ 1

Q_3 คือ ควอร์ไทล์ที่ 3

IQR คือ ส่วนต่างระหว่างควอร์ไทล์ที่ 3 และ 1 ($Q_3 - Q_1$)

ในกรณีที่พบว่า Age มีค่าต่ำกว่า $Q_1 - 1.5 \times IQR$ ปี จะแทนด้วยค่า $Q_1 - 1.5 \times IQR$ เซนติเมตร และหากพบว่า Age มีค่าสูงกว่า $Q_3 + 1.5 \times IQR$ เซนติเมตร จะแทนด้วยค่า $Q_3 + 1.5 \times IQR$ ปี

- ตัวแปร Systolic Blood Pressure (SBP)

ใช้เกณฑ์ตรวจจับค่านอกเกณฑ์ในกรณีที่ค่าข้อมูลไม่ตกอยู่ในช่วง 75 – 250 mm Hg ซึ่งจากการใช้เกณฑ์ตรวจจับดังกล่าวพบว่าข้อมูล 0.1855% อยู่นอกเกณฑ์ปกติ ในกรณีที่พบว่า SBP มีค่าต่ำกว่า 75 mm Hg จะแทนด้วยค่า 75 mm Hg และหากพบว่า SBP มีค่าสูงกว่า 250 mm Hg จะแทนด้วยค่า 250 mm Hg

- ตัวแปร Diastolic Blood Pressure (DBP)

ใช้เกณฑ์ตรวจจับค่านอกเกณฑ์ในกรณีที่ค่าข้อมูลไม่ตกอยู่ในช่วง 30 – 150 mm Hg ซึ่งจากการใช้เกณฑ์ตรวจจับดังกล่าวพบว่าข้อมูล 0.1819% อยู่นอกเกณฑ์ปกติ ในกรณีที่พบว่า SBP มีค่าต่ำกว่า 30 mm Hg จะแทนด้วยค่า 30 mm Hg และหากพบว่า SBP มีค่าสูงกว่า 150 mm Hg จะแทนด้วยค่า 150 mm Hg

- ตัวแปรเส้นรอบเอว (WAIST_CM)

ใช้เกณฑ์ตรวจจับค่านอกเกณฑ์เช่นเดียวกับการใช้ Box Plot ในการตรวจสอบค่านอกเกณฑ์ กล่าวคือ หากข้อมูลไม่ตกอยู่ในช่วง

$$(Q_1 - 1.5 \times IQR, Q_3 + 1.5 \times IQR)$$

เมื่อ Q_1 คือ ควอร์ไทล์ที่ 1

Q_3 คือ ควอร์ไทล์ที่ 3

IQR คือ ส่วนต่างระหว่างควอร์ไทล์ที่ 3 และ 1 ($Q_3 - Q_1$)

ในกรณีที่พบว่า WAIST_CM มีค่าต่ำกว่า $Q_1 - 1.5 \times IQR$ เซนติเมตร จะแทนด้วยค่า $Q_1 - 1.5 \times IQR$ เซนติเมตร และหากพบว่า WAIST_CM มีค่าสูงกว่า $Q_3 + 1.5 \times IQR$ เซนติเมตร จะแทนด้วยค่า $Q_3 + 1.5 \times IQR$ เซนติเมตร

จากการใช้เกณฑ์ดังกล่าวในการตรวจจับค่าผิดปกติของตัวแปร WAIST_CM พบว่าข้อมูลคิดเป็น 2.9989% อยู่นอกเกณฑ์ปกติ

- ตัวแปรดัชนีมวลกาย (BMI)

ใช้เกณฑ์ตรวจจับค่าผิดปกติและแทนค่านอกเกณฑ์ด้วยวิธีเดียวกับที่ใช้กับตัวแปร WAIST_CM จากการวิเคราะห์ดังกล่าวในการตรวจจับค่าผิดปกติของตัวแปร BMI พบว่าข้อมูลคิดเป็น 2.8843% อยู่นอกเกณฑ์ปกติ

(2) การจัดการปัญหาข้อมูลสูญหาย (Missing Data)

การจัดการปัญหาข้อมูลสูญหายสำหรับตัวแปรทำนายจะขอเสนอแบ่งตามหมวดของตัวแปรทำนาย ดังนี้

- ตัวแปรทำนายในหมวด Physical Measurement

ตัวแปรทำนายในหมวด Physical Measurement ทั้งหมดที่เป็นตัวแปรเชิงปริมาณ ได้แก่ ส่วนสูง (Height) ดัชนีมวลกาย (BMI) เส้นรอบเอว (WAIST_CM), Systolic Blood Pressure (SBP) และ Diastolic Blood Pressure (DBP) หากมีค่าสูญหาย จะแทนค่าสูญหายนั้นด้วยค่ามัธยฐานของแต่ละตัวแปรของแต่ละบุคคล ในกรณีที่มีการแทนค่าสูญหายของตัวแปร BMI และ WAIST_CM จะนำค่าดังกล่าวไปจัดกลุ่มต่อเพื่อแทนค่าสูญหายสำหรับตัวแปร BMI_CAT1 BMI_CAT2 และ WAIST_CAT ต่อไป

- ตัวแปรทำนายในหมวด Family History

ตัวแปรทำนายในหมวด Family History มี 2 ตัวแปร ซึ่งเป็นตัวแปรเชิงคุณภาพทั้งคู่ ได้แก่ ประวัติเบาหวานในญาติสายตรง (Family Hx of DM) และ ประวัติความดันโลหิตสูงในญาติสายตรง (Family Hx of HT) ซึ่งหากมีข้อมูลสูญหาย จะแทนข้อมูลสูญหายเป็นกลุ่ม “ไม่มี”

- ตัวแปรทำนายในหมวด Underlying Disease

ตัวแปรทำนายในหมวด Underlying Disease มี 2 ตัวแปร ซึ่งเป็นตัวแปรเชิงคุณภาพทั้งคู่ ได้แก่ ภาวะความดันโลหิตสูง (Hypertension) และ ประวัติระดับน้ำตาลสูง (Hx of high glucose) ซึ่งหากมีข้อมูลสูญหาย จะแทนข้อมูลสูญหายเป็นกลุ่ม “ปกติ” และ “ไม่มีประวัติระดับน้ำตาลสูง” ตามลำดับ

- ตัวแปรทำนายในหมวด Lab

โดยทั่วไปเนื่องจากตัวแปรทำนายในหมวด Lab มีการวัดค่าที่แน่ชัดและเชื่อถือได้ และน่าจะเป็นปัจจัยหลักในการทำนายความเสี่ยงการเป็นโรคเบาหวานในอนาคต ตัวแปรทำนายในหมวด Lab ได้แก่ ระดับน้ำตาลในเลือด (Bs_level) ระดับไขมันไตรกลีเซอไรด์ (TG) และระดับไขมันดี (HDL) กรณีที่มีข้อมูลในตัวแปรดังกล่าวสูญหาย จะไม่มีการเติมข้อมูล แต่จะสร้าง Flag ว่าข้อมูลใดบ้างเป็นข้อมูลสูญหาย และจะพิจารณาใช้ Flag ดังกล่าวเป็นตัวแปรทำนายเพิ่มเติมในการสร้างตัวแบบด้วย

การวิเคราะห์ข้อมูลเชิงสำรวจ

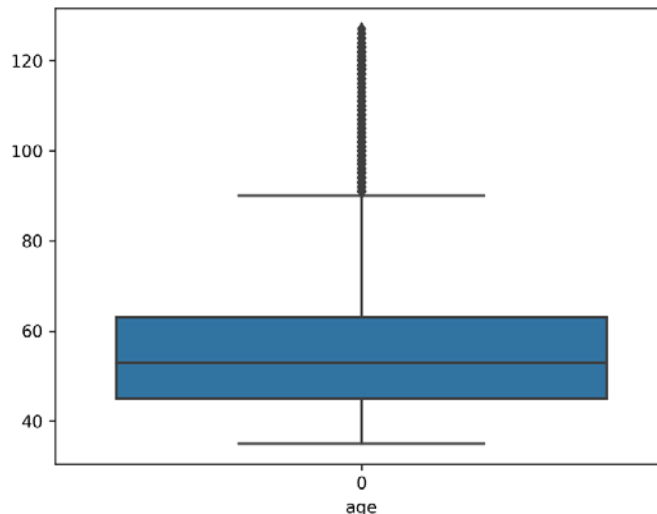
ในเบื้องต้นของการวิเคราะห์ ผู้วิจัยได้ทำการวิเคราะห์ข้อมูลเชิงสำรวจเพื่อดูการกระจายของข้อมูลตัวแปรต่างๆ ที่เกี่ยวข้อง โดยพบว่าข้อมูลคิดเป็น

- อายุ (Age)

อายุของผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคเบาหวานเมื่อเริ่มต้น Cohort เท่านั้น มีอายุเฉลี่ยอยู่ที่ 53.77 ปี และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 11.81 ปี ดังแสดงในตารางที่ 5 และรูปที่ 12

ตารางที่ 5 ค่าเฉลี่ยเลขคณิต มัธยฐาน ส่วนเบี่ยงเบนมาตรฐาน ค่าต่ำสุด ค่าสูงสุดของอายุผู้ป่วยที่เข้ารับ
การคัดกรองและพบว่าไม่เป็นโรคเบาหวานเมื่อเริ่มต้น Cohort เท่านั้น

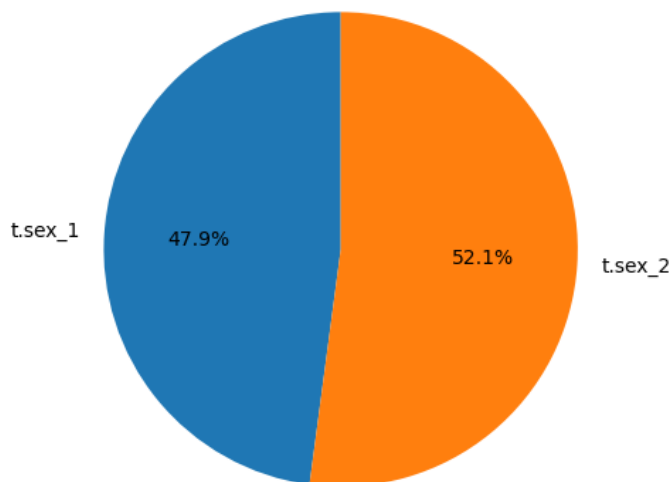
| ตัวแปร | ค่าเฉลี่ย | มัธยฐาน | S.D. | Min | Max |
|-----------|-----------|---------|-------|-----|-----|
| อายุ (ปี) | 53.77 | 52 | 11.81 | 35 | 87 |



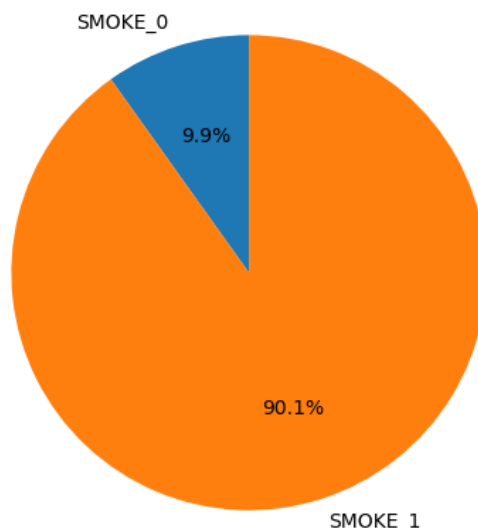
รูปที่ 12 แผนภาพกล่อง แสดงถึงการกระจายข้อมูลอายุของผู้ป่วยที่เข้ารับการรักษาและพบว่าไม่เป็นโรคเบาหวานเมื่อเริ่มต้น Cohort เท่านั้น

- เพศ (Gender)
เพศของผู้ป่วยที่เข้ารับการรักษาและพบว่าไม่เป็นโรคเบาหวานเมื่อเริ่มต้น Cohort เท่านั้น ส่วนใหญ่เป็นเพศหญิง (t.sex_2) ร้อยละ 52.1 และเป็นเพศชาย (t.sex_1) ร้อยละ 47.9 ดังแสดงในรูปที่ 13

- Smoking (พฤติกรรมการสูบบุหรี่)
พฤติกรรมการสูบบุหรี่ของผู้ป่วยที่เข้ารับการรักษาและพบว่าไม่เป็นโรคเบาหวานเมื่อเริ่มต้น Cohort เท่านั้น ส่วนใหญ่เป็นผู้ที่ไม่สูบบุหรี่ (SMOKE_1) ร้อยละ 90.1 และเป็นผู้ที่สูบบุหรี่ (SMOKE_0) ร้อยละ 9.9 ดังแสดงในรูปที่ 14



รูปที่ 13 แผนภูมิวงกลม แสดงถึงข้อมูลร้อยละ ผู้ป่วยที่เข้ารับการรักษาและพบว่าไม่เป็นโรคเบาหวานเมื่อเริ่มต้น Cohort เท่านั้น จำแนกตามเพศ



รูปที่ 14 แผนภูมิวงกลมแสดงถึงข้อมูลร้อยละ ผู้ป่วยที่เข้ารับการรักษาและพบว่าไม่เป็นโรคเบาหวานเมื่อเริ่มต้น Cohort เท่านั้น จำแนกตามพฤติกรรมการสูบบุหรี่

- Height (ส่วนสูง)
ส่วนสูงของผู้ป่วยที่เข้ารับการรักษาและพบว่าไม่เป็นโรคเบาหวานเมื่อเริ่มต้น Cohort เท่านั้น มีส่วนสูงเฉลี่ยอยู่ที่ 160.33 เซนติเมตร ดังแสดงในตารางที่ 6

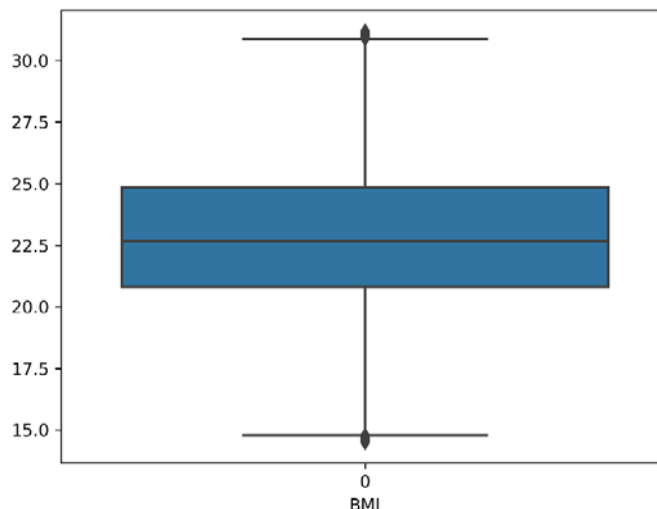
ตารางที่ 6 ค่าเฉลี่ยเลขคณิต มัธยฐาน ส่วนเบี่ยงเบนมาตรฐาน ค่าต่ำสุด ค่าสูงสุดของส่วนสูงผู้ป่วยที่เข้ารับการรักษาและพบว่าไม่เป็นโรคเบาหวานเมื่อเริ่มต้น Cohort เท่านั้น

| ตัวแปร | ค่าเฉลี่ย | มัธยฐาน | S.D. | Min | Max |
|------------------------|-----------|---------|------|-----|-----|
| ส่วนสูง (เซนติเมตร) | 160.33 | 160 | 7.44 | 140 | 180 |

- BMI (ดัชนีมวลกาย)
ค่าดัชนีมวลกายของผู้ป่วยที่เข้ารับการรักษาและพบว่าไม่เป็นโรคเบาหวานเมื่อเริ่มต้น Cohort เท่านั้น มีค่าเฉลี่ยอยู่ที่ 23.05 ดังแสดงในตารางที่ 7 และรูปที่ 15

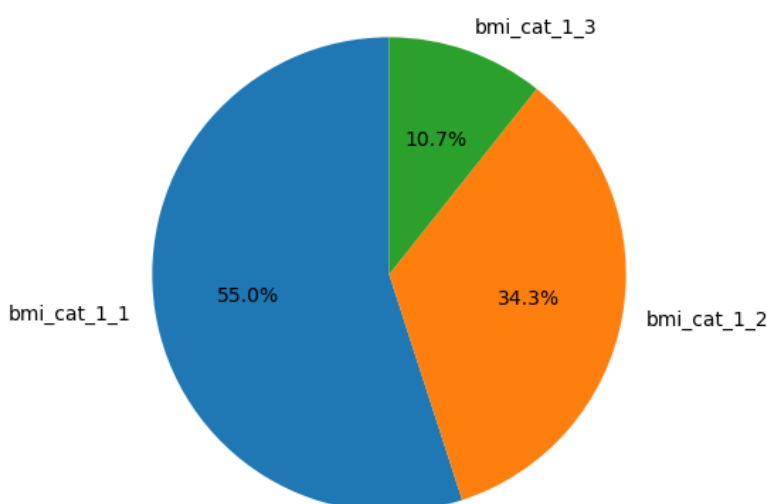
ตารางที่ 7 ค่าเฉลี่ยเลขคณิต มัธยฐาน ส่วนเบี่ยงเบนมาตรฐาน ค่าต่ำสุด ค่าสูงสุดของดัชนีมวลกาย (BMI) ผู้ป่วยที่เข้ารับการรักษาและพบว่าไม่เป็นโรคเบาหวานเมื่อเริ่มต้น Cohort เท่านั้น

| ตัวแปร | ค่าเฉลี่ย | มัธยฐาน | S.D. | Min | Max |
|----------------------|-----------|---------|------|------|------|
| ดัชนีมวลกาย (BMI) | 23.05 | 22.66 | 3.30 | 14.5 | 31.2 |

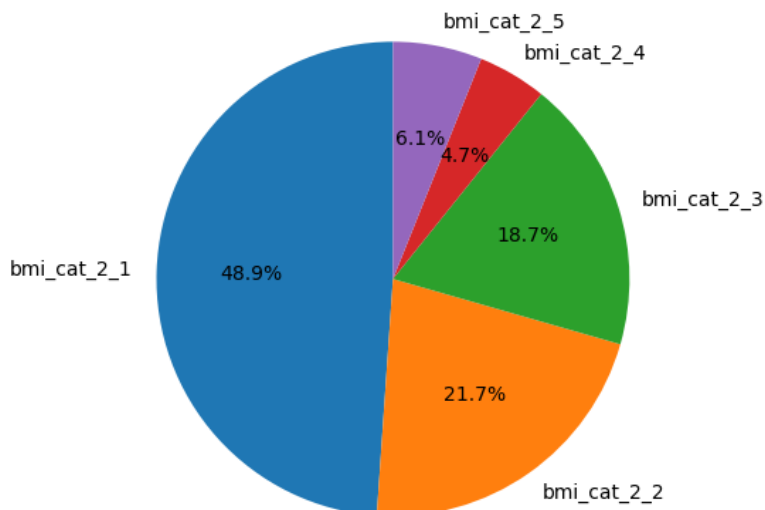


รูปที่ 15 แผนภาพกล่องแสดงถึงการกระจายข้อมูลค่าดัชนีมวลกาย (BMI) ของผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคเบาหวานเมื่อเริ่มต้น Cohort เท่านั้น

เมื่อจำแนกดัชนีมวลกาย (BMI) ตามกลุ่มแบบที่ 1 พบว่า จำนวนผู้ป่วยสูงสุด อยู่ในกลุ่มที่มีค่า BMI น้อยกว่า 23 (bmi_cat_1_1) ร้อยละ 55.0 รองลงมา กลุ่มที่มีค่า BMI อยู่ในช่วง 23 – 27.5 (bmi_cat_1_2) ร้อยละ 34.5 และกลุ่มที่มีค่า BMI มากกว่าหรือเท่ากับ 27.5 (bmi_cat_1_3) ร้อยละ 10.7 ดังแสดงในรูปที่ 16 และเมื่อจำแนกดัชนีมวลกาย (BMI) ตามกลุ่มแบบที่ 2 พบว่า จำนวนผู้ป่วยสูงสุด อยู่ใน กลุ่มที่มีค่า BMI น้อยกว่า 18.5 (bmi_cat_2_1) ร้อยละ 48.9 รองลงมา กลุ่มที่มีค่า BMI อยู่ในช่วง 18.5 – 22.9 (bmi_cat_2_2) ร้อยละ 21.7 กลุ่มที่มีค่า BMI อยู่ในช่วง 23 – 24.9 (bmi_cat_2_3) ร้อยละ 18.7 กลุ่มที่มีค่า BMI มากกว่าหรือเท่ากับ 30 (bmi_cat_2_5) ร้อยละ 6.1 และกลุ่มที่มีค่า BMI อยู่ในช่วง 25 – 29.9 (bmi_cat_2_4) ร้อยละ 4.7 ดังแสดงในรูปที่ 17



รูปที่ 16 แผนภูมิวงกลมแสดงถึงข้อมูลร้อยละ ผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคเบาหวานเมื่อเริ่มต้น Cohort เท่านั้น จำแนกตามดัชนีมวลกายแบบที่ 1



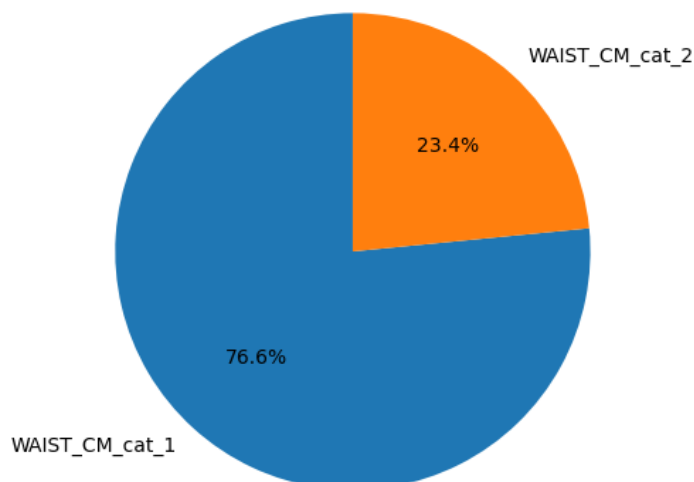
รูปที่ 17 แผนภูมิวงกลม แสดงถึงข้อมูลร้อยละ ผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคเบาหวานเมื่อเริ่มต้น Cohort เท่านั้น จำแนกตามดัชนีมวลกายแบบที่ 2

- WAIST (เส้นรอบเอว)

เส้นรอบเอวของผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคเบาหวานเมื่อเริ่มต้น Cohort เท่านั้น มีขนาดเฉลี่ยอยู่ที่ 80.43 เซนติเมตร ดังแสดงในตารางที่ 8 และเมื่อจำแนกกลุ่มตามความผิดปกติของเส้นรอบเอว โดยที่เกณฑ์ผิดปกติในเพศชายคือมากกว่าหรือเท่ากับ 90 เซนติเมตร และในหญิงคือมากกว่าหรือเท่ากับ 80 เซนติเมตร พบว่าส่วนใหญ่เป็นผู้ที่ไม่ผิดปกติ (WAIST_CM_cat_1) ร้อยละ 76.6 และเป็นผู้ที่ผิดปกติ (WAIST_CM_cat_2) ร้อยละ 23.4 ดังแสดงในรูปที่ 18

ตารางที่ 8 ค่าเฉลี่ยเลขคณิต มัธยฐาน ส่วนเบี่ยงเบนมาตรฐาน ค่าต่ำสุด ค่าสูงสุด ขนาดเส้นรอบเอวของผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคเบาหวานเมื่อเริ่มต้น Cohort เท่านั้น

| ตัวแปร | ค่าเฉลี่ย | มัธยฐาน | S.D. | Min | Max |
|---------------------------|-----------|---------|------|-----|-----|
| เส้นรอบเอว (เซนติเมตร) | 80.43 | 80 | 7.91 | 60 | 100 |



รูปที่ 18 แผนภูมิวงกลม แสดงถึงข้อมูลร้อยละ ผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคเบาหวานเมื่อเริ่มต้น Cohort เท่านั้น จำแนกตามขนาดเส้นรอบเอว

- SBP (ความดันช่วงหัวใจบีบตัว)

ค่าความดันช่วงหัวใจบีบตัวของผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคเบาหวานเมื่อเริ่มต้น Cohort เท่านั้น มีค่าเฉลี่ยอยู่ที่ 111.38 มิลลิเมตรปรอท ดังแสดงในตารางที่ 9

ตารางที่ 9 ค่าเฉลี่ยเลขคณิต มัธยฐาน ส่วนเบี่ยงเบนมาตรฐาน ค่าต่ำสุด ค่าสูงสุดของความดันช่วงหัวใจบีบตัว สำหรับผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคเบาหวานเมื่อเริ่มต้น Cohort เท่านั้น

| ตัวแปร | ค่าเฉลี่ย | มัธยฐาน | S.D. | Min | Max |
|-------------|-----------|---------|-------|-----|-----|
| SBP (mm Hg) | 111.38 | 113 | 16.66 | 75 | 250 |

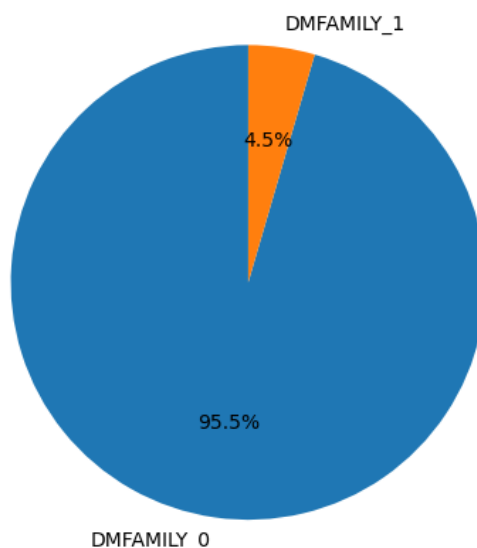
- DBP (ความดันช่วงหัวใจคลายตัว)

ค่าความดันช่วงหัวใจคลายตัวของผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคเบาหวานเมื่อเริ่มต้น Cohort เท่านั้น มีค่าเฉลี่ยอยู่ที่ 68.68 มิลลิเมตรปรอท ดังแสดงในตารางที่ 10

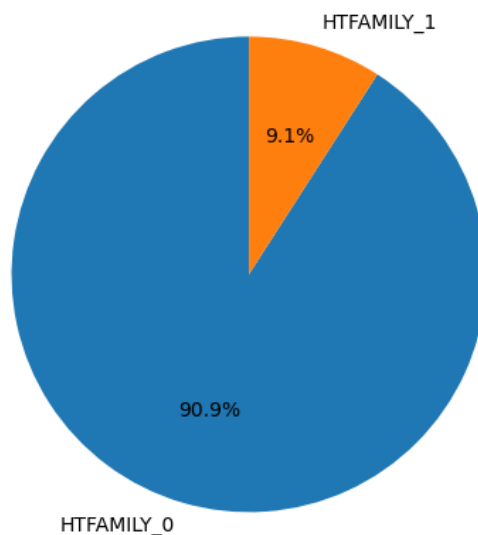
ตารางที่ 10 ค่าเฉลี่ยเลขคณิต มัธยฐาน ส่วนเบี่ยงเบนมาตรฐาน ค่าต่ำสุด ค่าสูงสุดของความดันช่วงหัวใจคลายตัว สำหรับผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคเบาหวานเมื่อเริ่มต้น Cohort เท่านั้น

| ตัวแปร | ค่าเฉลี่ย | มัธยฐาน | S.D. | Min | Max |
|-------------|-----------|---------|-------|-----|-----|
| DBP (mm Hg) | 68.68 | 71 | 13.46 | 30 | 150 |

- Family Hx of DM (ประวัติเบาหวานในญาติสายตรง)
ประวัติเบาหวานในญาติสายตรงของผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคเบาหวานเมื่อเริ่มต้น Cohort เท่านั้น ส่วนใหญ่เป็นผู้ที่ไม่มีประวัติเบาหวานในญาติสายตรง (DMFAMILY_0) ร้อยละ 95.5 และเป็นผู้ที่มีประวัติเบาหวานในญาติสายตรง (DMFAMILY_1) ร้อยละ 4.5 ดังแสดงในรูปที่ 19
- Family Hx of HT (ประวัติความดันโลหิตสูงในญาติสายตรง)
ประวัติความดันโลหิตสูงในญาติสายตรงของผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคเบาหวานเมื่อเริ่มต้น Cohort เท่านั้น ส่วนใหญ่เป็นผู้ที่ไม่มีประวัติความดันโลหิตสูงในญาติสายตรง (HTFAMILY_0) ร้อยละ 90.9 และเป็นผู้ที่มีประวัติความดันโลหิตสูงในญาติสายตรง (HTFAMILY_1) ร้อยละ 9.1 ดังแสดงในรูปที่ 20



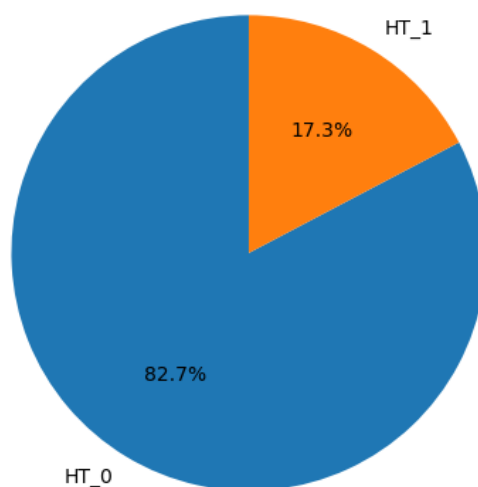
รูปที่ 19 แผนภูมิวงกลมแสดงถึงข้อมูลร้อยละ ผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคเบาหวานเมื่อเริ่มต้น Cohort เท่านั้น จำแนกตามการมีประวัติเบาหวานในญาติสายตรง



รูปที่ 20 แผนภูมิวงกลมแสดงถึงข้อมูลร้อยละ ผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคเบาหวานเมื่อเริ่มต้น Cohort เท่านั้น จำแนกตามการมีประวัติความดันโลหิตสูงในญาติสายตรง

- Hypertension (ภาวะความดันโลหิตสูง)

การเกิดภาวะความดันโลหิตสูงของผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคเบาหวานเมื่อเริ่มต้น Cohort เท่านั้น เมื่อจำแนกกลุ่มตามความผิดปกติของการเกิดภาวะความดันโลหิตสูง โดยถือว่าเกณฑ์ผิดปกตินั้นจะเกิดขึ้นเมื่อค่า SBP มากกว่าหรือเท่ากับ 140 มิลลิเมตรปรอท หรือค่า DBP มากกว่าหรือเท่ากับ 90 มิลลิเมตรปรอท พบว่าส่วนใหญ่เป็นผู้ที่ไม่ผิดปกติ (HT_0) ร้อยละ 82.7 และเป็นผู้ที่ผิดปกติ (HT_1) ร้อยละ 17.3 ดังแสดงในรูปที่ 21



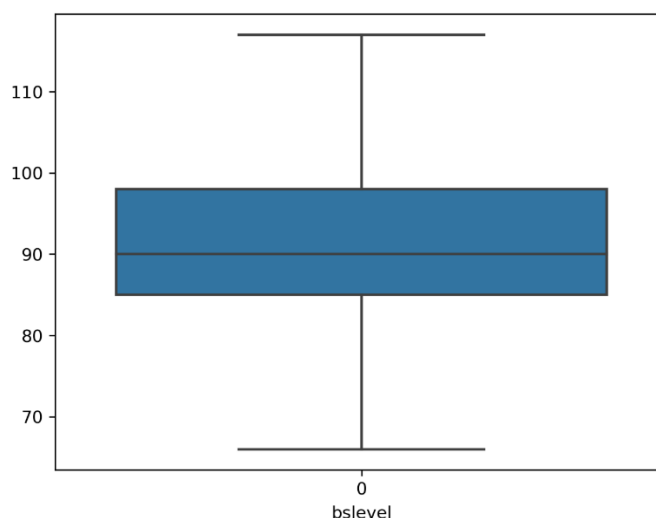
รูปที่ 21 แผนภูมิวงกลมแสดงถึงข้อมูลร้อยละ ผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคเบาหวานเมื่อเริ่มต้น Cohort เท่านั้น จำแนกตามการเกิดภาวะความดันโลหิตสูง

- BSLlevel (ระดับน้ำตาลในเลือด)

ระดับน้ำตาลในเลือดของผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคเบาหวานเมื่อเริ่มต้น Cohort เท่านั้น มีค่าเฉลี่ยอยู่ที่ 91.38 มิลลิกรัมต่อเดซิลิตร ทั้งนี้ ส่วนใหญ่เป็นผู้ป่วยที่ทราบค่าระดับน้ำตาลในเลือดย้อยละ 99.8 และเป็นผู้ป่วยที่ไม่ทราบค่าระดับน้ำตาลในเลือดย้อยละ 0.2 ดังแสดงในตารางที่ 11 และรูปที่ 22

ตารางที่ 11 ค่าเฉลี่ยเลขคณิต มัธยฐาน ส่วนเบี่ยงเบนมาตรฐาน ค่าต่ำสุด ค่าสูงสุดของระดับน้ำตาลในเลือด ผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคเบาหวานเมื่อเริ่มต้น Cohort เท่านั้น

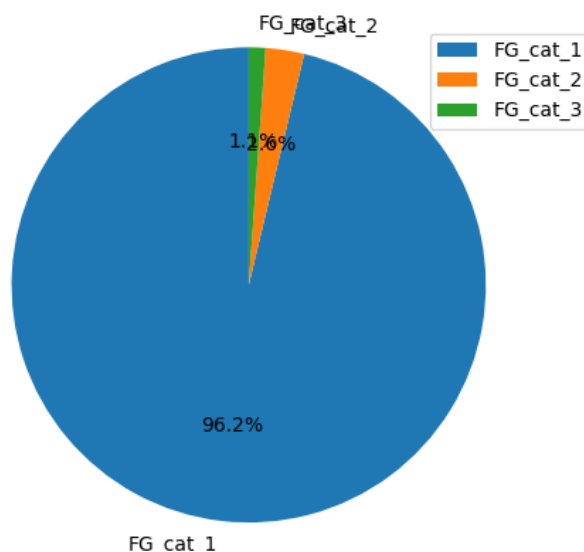
| ตัวแปร | ค่าเฉลี่ย | มัธยฐาน | S.D. | Min | Max |
|----------------------------|-----------|---------|------|-----|-----|
| ระดับน้ำตาลในเลือด (mg/dL) | 91.38 | 90 | 12.5 | 0 | 180 |



รูปที่ 22 แผนภาพกล่องแสดงถึงการกระจายข้อมูลค่าระดับน้ำตาลในเลือด ของผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคเบาหวานเมื่อเริ่มต้น Cohort เท่านั้น

- Fasting Glucose (ประเภทการอดอาหารก่อนเจาะเลือด)

พฤติกรรมการอดอาหารก่อนเจาะเลือดเพื่อวัดระดับน้ำตาลในเลือดของผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคเบาหวานเมื่อเริ่มต้น Cohort เท่านั้น ส่วนใหญ่เป็นผู้ที่อดอาหารก่อนเจาะเลือดเพื่อวัดระดับน้ำตาลในเลือด (FG_cat1) ร้อยละ 96.2 และเป็นผู้ที่ไม่ได้อดอาหารก่อนเจาะเลือดเพื่อวัดระดับน้ำตาลในเลือด หรือไม่ทราบข้อมูลพฤติกรรม (FG_cat2 & FG_cat3) ร้อยละ 3.8 ดังแสดงในรูปที่ 23

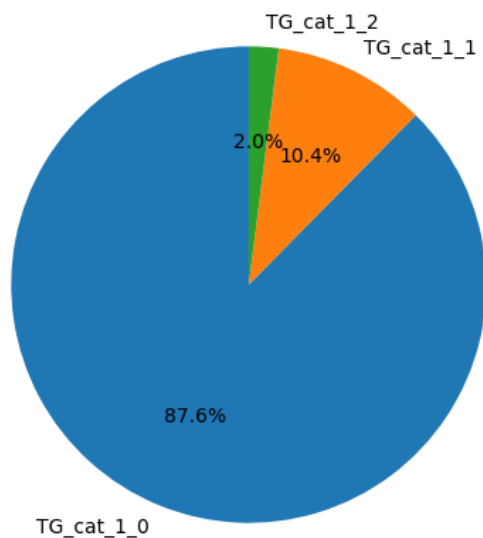


รูปที่ 23 แผนภูมิวงกลมแสดงถึงข้อมูลร้อยละ ผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคเบาหวานเมื่อเริ่มต้น Cohort เท่านั้น จำแนกตามพฤติกรรมการอดอาหารก่อนเจาะเลือดเพื่อวัดระดับน้ำตาลในเลือด

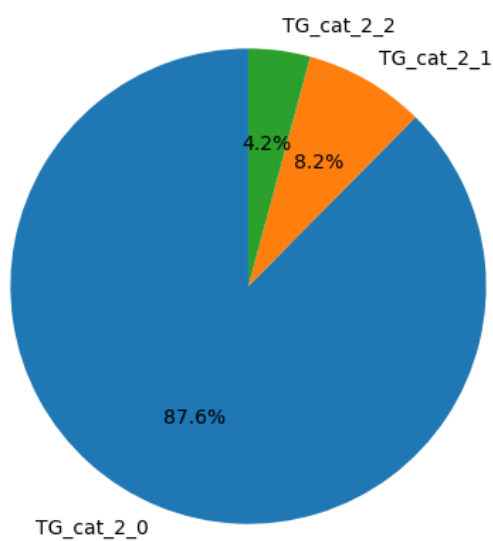
- TG (ระดับไขมันไตรกลีเซอไรด์)

ผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคเบาหวานเมื่อเริ่มต้น Cohort เท่านั้น ส่วนใหญ่ทราบระดับไขมันไตรกลีเซอไรด์ ร้อยละ 87.6 ไม่ทราบระดับไขมันไตรกลีเซอไรด์ ร้อยละ 12.4 และเมื่อจำแนกระดับไขมันไตรกลีเซอไรด์ของผู้ป่วย ตามกลุ่มแบบที่ 1 พบว่า จำนวนผู้ป่วยสูงสุด อยู่ใน กลุ่มที่ไม่ทราบระดับไขมันไตรกลีเซอไรด์ ร้อยละ 87.6 มีค่าระดับไขมันไตรกลีเซอไรด์น้อยกว่า 200 มิลลิกรัมต่อเดซิลิตร ร้อยละ 10.4 และมีค่าระดับไขมันไตรกลีเซอไรด์มากกว่าหรือเท่ากับ 200 มิลลิกรัมต่อเดซิลิตร ร้อยละ 2.0 ดังแสดงในรูปที่ 24

นอกจากนี้ หากจำแนกระดับไขมันไตรกลีเซอไรด์ของผู้ป่วย ตามกลุ่มแบบที่ 2 พบว่า จำนวนผู้ป่วยสูงสุด อยู่ใน กลุ่มที่ไม่ทราบระดับไขมันไตรกลีเซอไรด์ ร้อยละ 87.6 มีค่าระดับไขมันไตรกลีเซอไรด์น้อยกว่า 150 มิลลิกรัมต่อเดซิลิตร ร้อยละ 8.2 และมีค่าระดับไขมันไตรกลีเซอไรด์มากกว่าหรือเท่ากับ 150 มิลลิกรัมต่อเดซิลิตร ร้อยละ 4.2 ดังแสดงในรูปที่ 25



รูปที่ 24 แผนภูมิวงกลมแสดงถึงข้อมูลร้อยละ ผู้ป่วยที่เข้ารับการรักษาและพบว่าไม่เป็นโรคเบาหวาน
เมื่อเริ่มต้น Cohort เท่านั้น จำแนกตามระดับไขมันไตรกลีเซอไรด์แบบที่ 1



รูปที่ 25 แผนภูมิวงกลมแสดงถึงข้อมูลร้อยละ ผู้ป่วยที่เข้ารับการรักษาและพบว่าไม่เป็นโรคเบาหวาน
เมื่อเริ่มต้น Cohort เท่านั้น จำแนกตามระดับไขมันไตรกลีเซอไรด์แบบที่ 2

- HDL (ระดับไขมันดี)
ค่าระดับไขมันดีของผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคเบาหวานเมื่อเริ่มต้น Cohort เท่านั้น มีระดับไขมันดีเฉลี่ยอยู่ที่ 6.41 มิลลิกรัมต่อเดซิลิตร ดังแสดงในตารางที่ 12

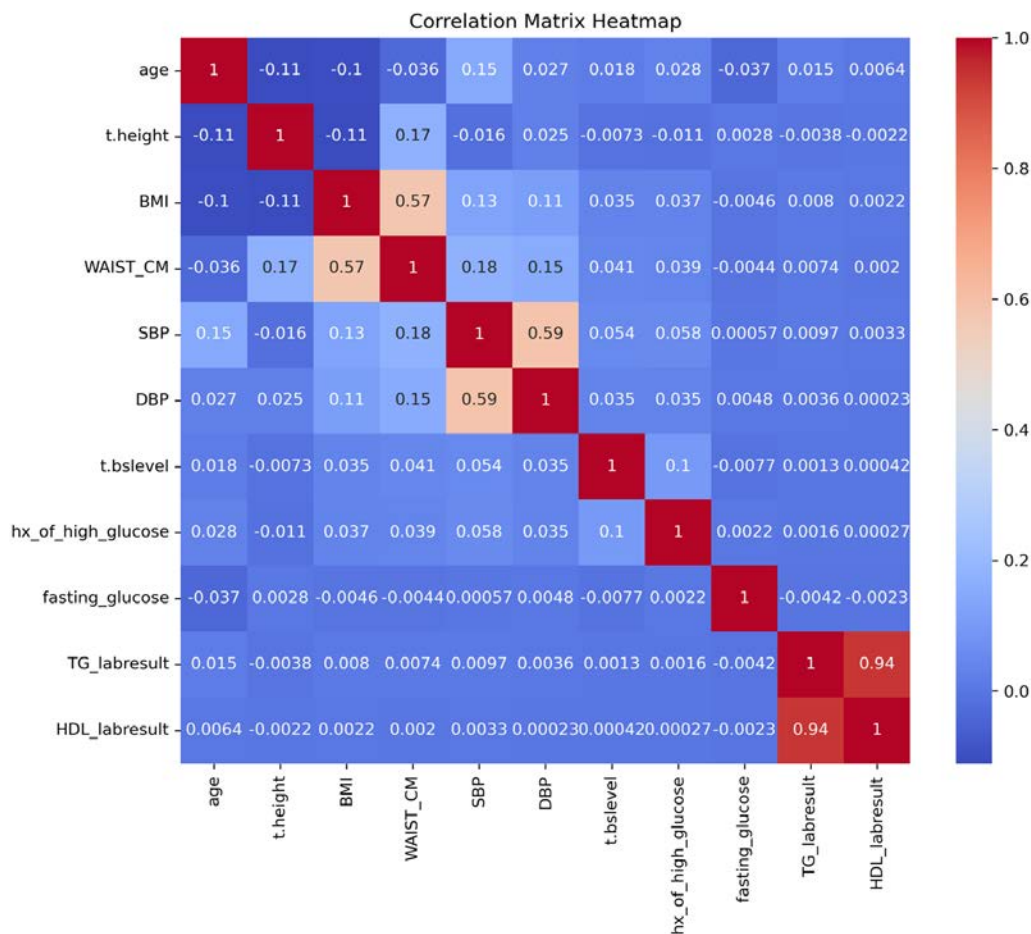
ตารางที่ 12 ค่าเฉลี่ยเลขคณิต มัธยฐาน ส่วนเบี่ยงเบนมาตรฐาน ค่าต่ำสุด ค่าสูงสุดของระดับไขมันดี สำหรับผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคเบาหวานเมื่อเริ่มต้น Cohort เท่านั้น

| ตัวแปร | ค่าเฉลี่ย | มัธยฐาน | S.D. | Min | Max |
|-------------|-----------|---------|-------|-----|-----|
| HDL (mg/dL) | 6.41 | 0 | 17.15 | 0 | 87 |

- ความสัมพันธ์ระหว่างตัวแปรต่าง ๆ
นอกจากการวิเคราะห์ข้อมูลเชิงสำรวจโดยพิจารณาทีละหนึ่งตัวแปร (Univariate Analysis) ทางผู้วิจัยได้ทำการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรต่าง ๆ โดยแสดงผ่านแผนภาพ Heatmap ดังรูปที่ 26 เพื่อดูความสัมพันธ์เชิงเส้นว่ามีความสัมพันธ์มาก-น้อยระหว่างตัวแปรโดยใช้สถิติทดสอบสัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient) ซึ่งพบว่า 3 คู่ตัวแปรที่มีความสัมพันธ์กันมากในทิศทางเดียวกันดังตารางที่ 13

ตารางที่ 13 ค่าสัมประสิทธิ์สหสัมพันธ์ในการวิเคราะห์ความสัมพันธ์ระหว่าง ตัวแปรต่างๆ ที่ใช้ในการทำนาย อุบัติการณ์การเป็นเบาหวาน สำหรับผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคเบาหวานเมื่อเริ่มต้น Cohort เท่านั้น (เฉพาะคู่ที่มีสัมประสิทธิ์สหสัมพันธ์มากในทิศทางเดียวกัน)

| ตัวแปรที่ 1 | ตัวแปรที่ 2 | Correlation Coefficient |
|--|--|-------------------------|
| TG_labresult (การตรวจพบ ระดับไขมันไตรกลีเซอไรด์) | HDL_labresult (การตรวจพบ ระดับไขมัน HDL) | 0.94 |
| SBP (ความดันช่วงหัวใจบีบตัว) | DBP (ความดันช่วงหัวใจคลายตัว) | 0.59 |
| WAIST_CM (เส้นรอบเอว) | BMI (ดัชนีมวลกาย) | 0.57 |



รูปที่ 26 ค่าสัมประสิทธิ์สหสัมพันธ์ในการวิเคราะห์ความสัมพันธ์ระหว่าง ตัวแปรต่างๆ ที่ใช้ในการทำนาย
อุบัติการณ์การเป็นเบาหวาน

ผลลัพธ์จากตัวแบบทำนายความเสี่ยงการเกิดโรคเบาหวานในประชากรทั่วไป

จากกลุ่มตัวอย่างของผู้เข้ารับการคัดกรองเบาหวานจำนวน 1,083,255 คน ซึ่งประกอบได้ด้วยข้อมูลจำนวน 4,211,132 แถวหรือจำนวนครั้งที่เข้ารับการคัดกรองทั้งหมด จะถูกสุ่มแบ่งออกเป็นชุดข้อมูลฝึกฝนและชุดข้อมูลทดสอบคิดเป็นร้อยละ 70 และ 30 โดยประมาณ ตามลำดับ ซึ่งจากการแบ่งข้อมูลจะได้ข้อมูลดังต่อไปนี้

- ชุดข้อมูลฝึกฝน ประกอบด้วยข้อมูลจากผู้เข้ารับการคัดกรองเบาหวานจำนวน 758,277 คน ซึ่งมีข้อมูลจำนวน 2,948,334 แถว
- ชุดข้อมูลทดสอบ ประกอบด้วยข้อมูลจากผู้เข้ารับการคัดกรองเบาหวานจำนวน 324,978 คน ซึ่งมีข้อมูลจำนวน 1,262,798 แถว

จากนั้นจึงสร้างตัวแบบจากข้อมูลชุดฝึกฝน ซึ่งจะพิจารณา 2 ตัวแบบ ได้แก่

1. การถดถอยโลจิสติกแบบ Elastic Net โดยจะทำขั้นตอน 2 ขั้นตอนในการคัดเลือกตัวแปร โดยขั้นที่ 1 หากตัวแปรทำนายบางตัวมีความเกี่ยวข้องกัน เช่น ตัวแปร TG TG_CAT1 และ TG_CAT2 ต่างก็เป็นตัวแปรระดับไขมันไตรกลีเซอไรด์ทั้งสามตัวแปร เป็นต้น จะพิจารณาหาเซตของตัวแปรทำนายที่ให้ผลทำนายดีที่สุด และในขั้นที่ 2 จะนำเซตของตัวแปรทำนายเหล่านั้นเข้ากระบวนการสร้างตัวแบบของ EN ในการคัดเลือกตัวแปรอัตโนมัติอีกหนึ่ง วิธีการนี้ขอเรียกว่า “EN”
2. วิธี XGBoost ที่พิจารณาตัวแปรทำนายทุกตัวแปรและให้วิธีการ XGBoost คัดเลือกตัวแปรให้อัตโนมัติ โดยวิธีการนี้ขอเรียกว่า “XGBoost”

หลังจากได้ตัวแบบทั้ง 2 แล้ว จึงนำมาทำนายข้อมูลในชุดทดสอบเพื่อวัดประสิทธิภาพการทำนายโดยพิจารณาจาก

1. AUC
2. Cumulative AUC

ตารางที่ 14 แสดงตัววัดประสิทธิภาพทั้ง 2 ตัววัด ซึ่งจะเห็นได้ชัดว่า วิธี XGBoost มีประสิทธิภาพในการทำนายความเสี่ยงการเกิดโรคเบาหวานในประชากรทั่วไป จาก Cohort ของผู้ที่ยังไม่เป็นโรคเบาหวานได้ดีกว่าวิธีการถดถอยโลจิสติกแบบ Elastic Net อย่างไรก็ตาม ประสิทธิภาพการทำนายของวิธีการถดถอยโลจิสติกแบบ Elastic Net ยังอยู่ในเกณฑ์ดี คือ มีค่า AUC ทั้งสองตัวเกิน 0.8

ตารางที่ 14 ค่า AUC เพื่อวัดประสิทธิภาพการทำนายความเสี่ยงการเกิดโรคเบาหวานในประชากรทั่วไป จาก Cohort ของผู้ที่ยังไม่เป็นโรคเบาหวาน (Cohort 1.1) (วิธีที่ดีที่สุดจะแสดงเป็นตัวหนา)

| ตัวแบบ | AUC | Cumulative AUC |
|----------------|---------------|----------------|
| EN | 0.8452 | 0.8565 |
| XGBoost | 0.8573 | 0.8760 |

การแปลผลตัวแบบทำนายความเสี่ยงการเกิดโรคเบาหวานในประชากรทั่วไป

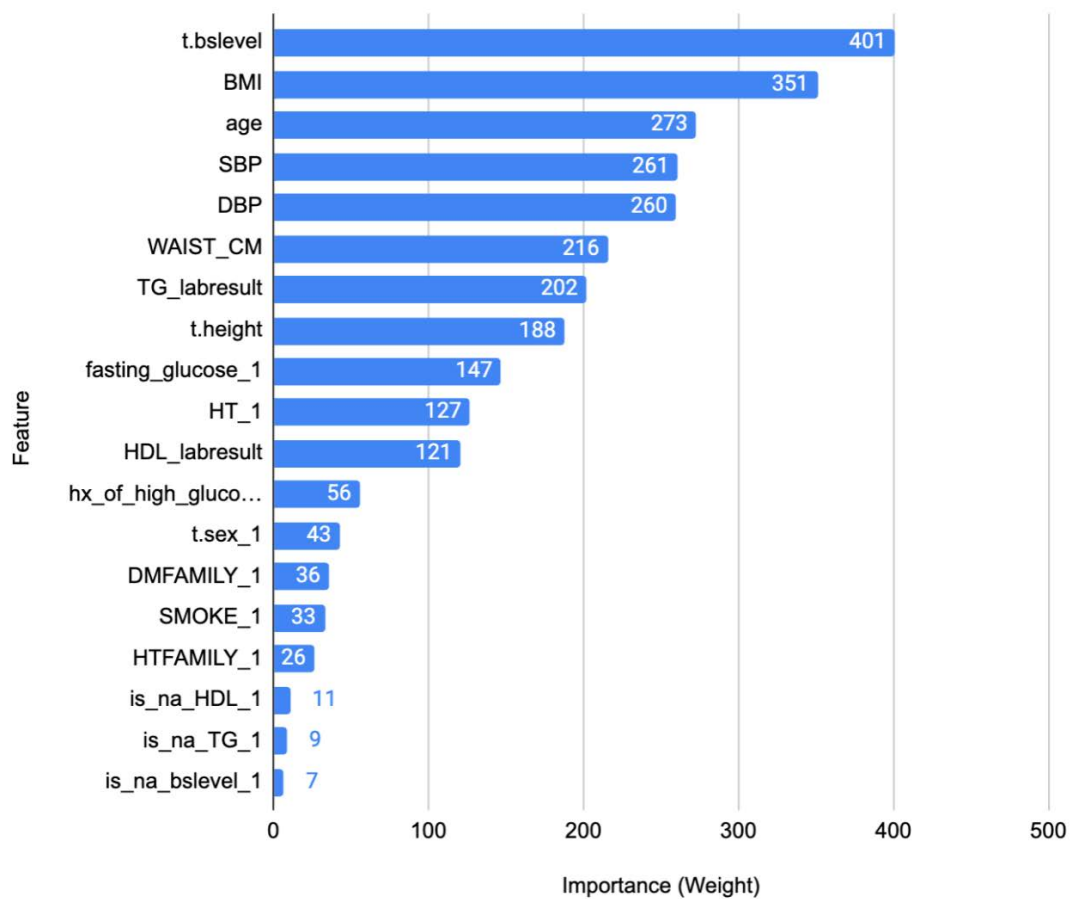
จากตัวแบบ XGBoost ที่ให้ประสิทธิภาพการทำนายดีที่สุดสามารถพิจารณาได้ว่าปัจจัยใดที่มีความสำคัญในการทำนายความเสี่ยงการเกิดโรคเบาหวานในประชากรทั่วไปดังแสดงในรูปที่ 27 ซึ่งจะพบว่าตัวแปรที่มีความสำคัญในการทำนายความเสี่ยง 10 อันดับแรก ได้แก่

1. ระดับน้ำตาลในเลือดซึ่งวัดโดยที่ไม่ได้ทำการอดอาหาร (t.bslevel)
2. ดัชนีมวลกาย (BMI)
3. อายุ (Age)
4. Systolic Blood Pressure (SBP)

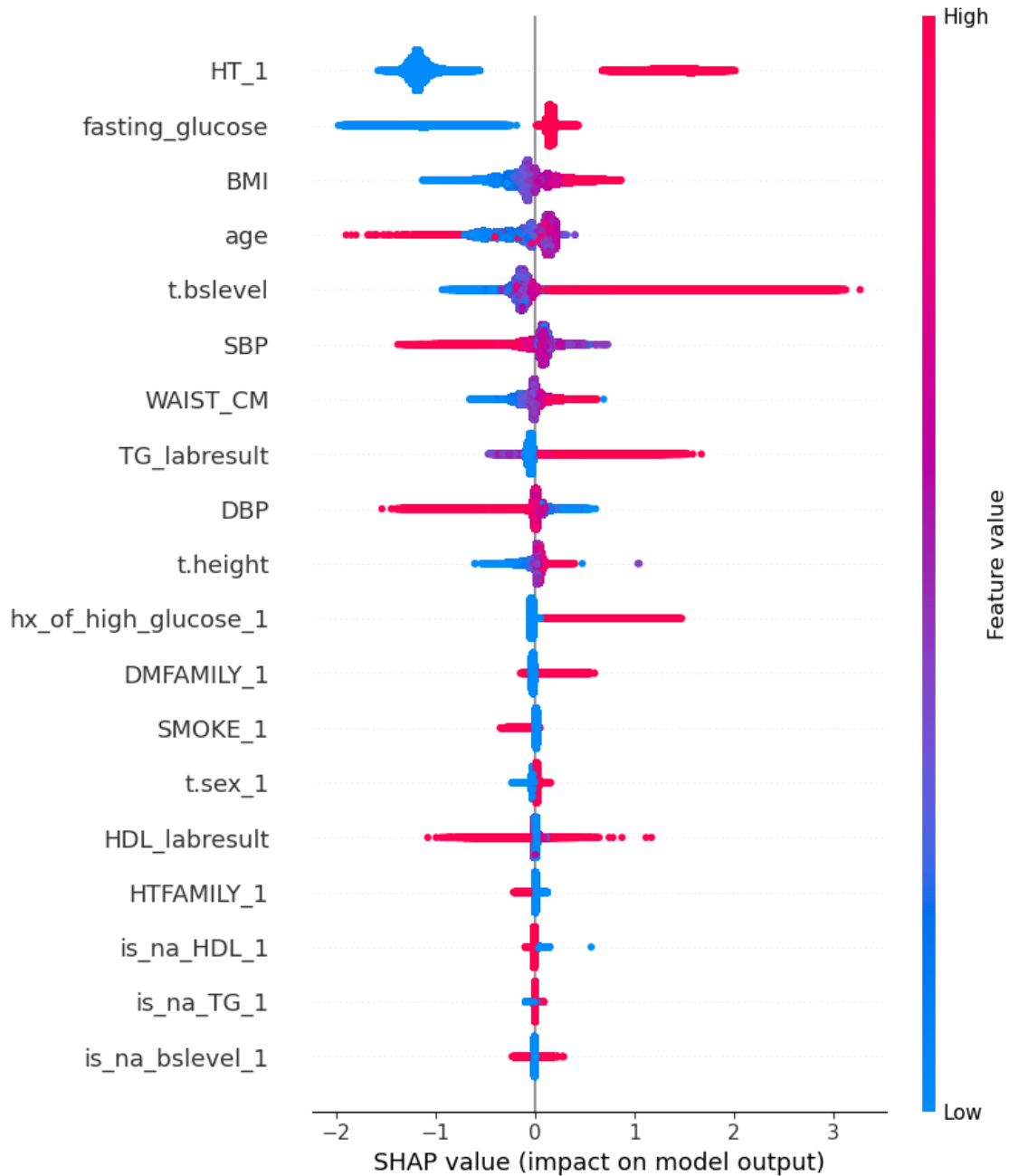
5. Diastolic Blood Pressure (DBP)
6. เส้นรอบเอว (WAIST_CM)
7. ระดับไขมันไตรกลีเซอไรด์ (TG_labresult)
8. ส่วนสูง (t.height)
9. ปรากฏการอดอาหารก่อนเจาะเลือดเพื่อวัดระดับน้ำตาล (fasting_glucose)
10. ภาวะความดันโลหิตสูง (HT_1)

นอกจากนี้ หากพิจารณาคุณค่า SHAP ดังรูปที่ 28 จะสามารถแปลผลในส่วนทิศทางความสัมพันธ์ของตัวแปรแต่ละตัว ซึ่งจากรูปที่ 28 สามารถสรุปความสัมพันธ์ที่เด่นชัดได้ ดังนี้

- หากผู้คัดกรองมีระดับความดันโลหิตผิดปกติ (HT_1) จะมีแนวโน้มที่จะความเสี่ยงการเป็นโรคเบาหวานที่เพิ่มขึ้น
- ระดับน้ำตาลในเลือดซึ่งวัดโดยที่ไม่ได้ทำการอดอาหาร (t.bslevel) หากมีค่าสูงจะมีแนวโน้มที่จะความเสี่ยงการเป็นโรคเบาหวานที่เพิ่มขึ้น
- ดัชนีมวลกาย (BMI) หากมีค่าสูงจะมีแนวโน้มที่จะความเสี่ยงการเป็นโรคเบาหวานที่เพิ่มขึ้น
- Systolic Blood Pressure (SBP) หากมีค่าสูงจะมีแนวโน้มที่จะความเสี่ยงการเป็นโรคเบาหวานที่ลดลง
- Diastolic Blood Pressure (DBP) หากมีค่าสูงจะมีแนวโน้มที่จะความเสี่ยงการเป็นโรคเบาหวานที่ลดลง
- ส่วนสูง (t.height) หากมีค่าสูงจะมีแนวโน้มที่จะความเสี่ยงการเป็นโรคเบาหวานที่เพิ่มขึ้น
- เส้นรอบเอว (WAIST_CM) หากมีค่าสูงจะมีแนวโน้มที่จะความเสี่ยงการเป็นโรคเบาหวานที่เพิ่มขึ้น



รูปที่ 27 ความสำคัญของตัวแปรทำนายจากวิธี XGBoost ในการทำนายความเสี่ยงการเกิดโรคเบาหวานใน
ประชากรไทย



รูปที่ 28 ค่า SHAP จากตัวแบบ XGBoost ในการทำนายความเสี่ยงการเกิดโรคเบาหวานในประเทศไทย

แม้ว่าวิธี XGBoost สามารถระบุความสำคัญของตัวแปรทำนายที่ใช้ในตัวแบบได้ แต่ไม่สามารถแปลผลตัวแบบในรูปของอัตราเสี่ยงได้ ในขณะที่วิธีการถดถอยโลจิสติกแบบ Elastic Net ให้ค่าสัมประสิทธิ์การถดถอยเช่นเดียวกับตัวแบบการถดถอยโลจิสติกแบบปกติ ดังนั้นจึงสามารถแปลผลตัวแบบในรูปของอัตราเสี่ยงได้ อย่างไรก็ตาม วิธี Elastic Net จะไม่ให้ค่า p-value ในการทดสอบสมมติฐานว่าปัจจัยใดบ้างที่มีความสัมพันธ์กับความเสี่ยงการเกิดโรคเบาหวาน

อย่างไรก็ดี การที่วิธี Elastic Net มีคุณสมบัติในการประมาณค่าสัมประสิทธิ์การถดถอยของตัวแปรทำนายบางตัวเป็นศูนย์ ซึ่งถือเป็นการเลือกตัวแปรทำนายเข้าตัวแบบไปพร้อมๆ กับการประมาณค่าสัมประสิทธิ์การถดถอย หรืออีกนัยหนึ่งคือตัวแปรทำนายที่มีค่าสัมประสิทธิ์การถดถอยไม่เท่ากับศูนย์นั้น ย่อมหมายความว่าตัวแปรดังกล่าวมีความสำคัญในการทำนายความเสี่ยงการเกิดโรคเบาหวานนั่นเอง

ตารางที่ 15 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ในการทำนายความเสี่ยงการเกิดโรคเบาหวานในประเทศไทย

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ในตัวแบบ | สัมประสิทธิ์การถดถอย | Odds Ratio |
|-----------------------------|----------------------|--------------------------|----------------------|------------|
| Demographic | | | | |
| Age | อายุ (ปี) | age | | |
| Age category | ช่วงอายุ | | | |
| | - 35 – 39 ปี | age_cat_1 | 0.1877 | 1.2065 |
| | - 40 – 44 ปี | age_cat_2 | 0.3731 | 1.4522 |
| | - 45 – 49 ปี | age_cat_3 | 0.5474 | 1.7288 |
| | - >=50 ปี | | | |
| Gender | เพศ | | | |
| | - ชาย | | | |
| | - หญิง | t.sex_2 | 0.1292 | 1.1379 |
| Health Behavior | | | | |
| Smoking | พฤติกรรมการสูบบุหรี่ | | | |
| | - สูบ | | | |
| | - ไม่สูบ | SMOKE_1 | -0.2486 | 0.7799 |
| Physical measurement | | | | |
| Height | ส่วนสูง (เซนติเมตร) | t.height | 0.0166 | 1.0167 |
| BMI | ดัชนีมวลกาย | BMI | 0.0722 | 1.0749 |
| BMI_CAT1 | ดัชนีมวลกาย | | | |
| | - < 23 | | | |
| | - 23 - 27.5 | bmi_cat_1_2 | | |
| | - >= 27.5 | bmi_cat_1_3 | | |

ตารางที่ 15 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net
ในการทำนายความเสี่ยงการเกิดโรคเบาหวานในประชากรไทย (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ในตัวแบบ | สัมประสิทธิ์การถดถอย | Odds Ratio |
|-----------------------|--|--|----------------------|------------|
| BMI_CAT2 | ดัชนีมวลกาย - < 18.5 - 18.5 – 22.9 - 23 – 24.9 - 25 – 29.9 - >= 30 | bmi_cat_2_2 bmi_cat_2_3 bmi_cat_2_4 bmi_cat_2_5 | | |
| WAIST_CM | เส้นรอบเอว (เซนติเมตร) | WAIST_CM | -0.0019 | 0.9981 |
| WAIST_CAT | เส้นรอบเอว - ปกติ - ผิดปกติ เกณฑ์ผิดปกติในเพศชายคือ >= 90 เซนติเมตร และในหญิงคือ >=80 เซนติเมตร | WAIST_CM_cat_2 | | |
| SBP | Systolic Blood Pressure (mm Hg) | SBP | -0.0187 | 0.9815 |
| DBP | Diastolic Blood Pressure (mm Hg) | DBP | 0.0000 | 1.0000 |
| Family History | | | | |
| Family Hx of DM | ประวัติเบาหวานในญาติสายตรง - มี - ไม่มี | DMFAMILY_1 | 0.1251 | 1.1332 |
| Family Hx of HT | ประวัติความดันโลหิตสูงในญาติสายตรง - มี - ไม่มี | HTFAMILY_1 | -0.1557 | 0.8558 |

ตารางที่ 15 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net
ในการทำนายความเสี่ยงการเกิดโรคเบาหวานในประชากรไทย (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ใน ตัวแบบ | สัมประสิทธิ์ การถดถอย | Odds Ratio |
|---------------------------|---|------------------------------|--------------------------|------------|
| Underlying Disease | | | | |
| Hypertension | ภาวะความดันโลหิตสูง - ปกติ - ผิดปกติ เกณฑ์ผิดปกติเมื่อ SBP >= 140 หรือ DBP >= 90 | HT_1 | 2.5340 | 12.6033 |
| Hx of high glucose | มีประวัติระดับน้ำตาลสูง - มี - ไม่มี | hx_of_high_glucose | 0.2101 | 1.2338 |
| Lab | | | | |
| Bs_level | ระดับน้ำตาลในเลือด (mg/dL) | t.bslevel | 0.0220 | 1.0222 |
| Is_na_bslevel* | ทราบค่าระดับน้ำตาลในเลือดหรือไม่ - ทราบ - ไม่ทราบ | is_na_bslevel_1 | 2.2253 | 9.2560 |
| Fasting Glucose | ประเภทการอดอาหารก่อนเจาะเลือด - อดอาหารก่อน - ไม่ได้อดอาหาร/ ไม่ทราบ | fasting_glucose | 1.2838 | 3.6104 |

ตารางที่ 15 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net
ในการทำนายความเสี่ยงการเกิดโรคเบาหวานในประชากรไทย (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ในตัวแบบ | สัมประสิทธิ์การถดถอย | Odds Ratio |
|-----------|--|----------------------------------|----------------------|------------------|
| FG_cat | ระดับน้ำตาล (อดอาหารก่อนเจาะเลือด) - < 100 mg/dL - 100 - 125 mg/dL - > 125 mg/dL - ไม่ทราบ | FG_cat_1 FG_cat_2 FG_cat_3 | | |
| TG | ระดับไขมันไตรกลีเซอไรด์ (mg/dL) | TG_labresult | | |
| TG_CAT1 | ระดับไขมันไตรกลีเซอไรด์ - < 200 mg/dL - >=200 mg/dL - ไม่ทราบ | TG_cat_1_1 TG_cat_1_2 | | |
| TG_CAT2 | ระดับไขมันไตรกลีเซอไรด์ - < 15- mg/dL - >= 150 mg/dL - ไม่ทราบ | TG_cat_2_1 TG_cat_2_2 | 0.0000 0.5239 | 1.0000 1.6887 |
| Is_na_TG* | ทราบค่าระดับไขมันไตรกลีเซอไรด์หรือไม่ - ทราบ - ไม่ทราบ | is_na_TG_1 | -0.0061 | 0.9939 |
| HDL | ระดับไขมันดี (mg/dL) | HDL_labresult | | |

ตารางที่ 15 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net
ในการทำนายความเสี่ยงการเกิดโรคเบาหวานในประชากรไทย (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ใน ตัวแบบ | สัมประสิทธิ์ การถดถอย | Odds Ratio |
|------------|---|------------------------------|--------------------------|------------|
| HDL_CAT1 | ระดับไขมันดี - ผิดปกติ - ปกติ - ไม่ทราบ เกณฑ์ผิดปกติในเพศ ชายคือ < 40 mg/dL และในเพศหญิงคือ < 50 mg/dL | HDL_cat_1_1 | 0.0000 | 1.0000 |
| | | HDL_cat_1_2 | 0.1405 | 1.1508 |
| Is_na_HDL* | ทราบค่าระดับไขมันดี หรือไม่ - ทราบ - ไม่ทราบ | is_na_HDL_1 | -0.0138 | 0.9862 |

หมายเหตุ * หมายถึงตัวแปรเพิ่มเติมที่ได้จากการเตรียมข้อมูล

แบบจำลองที่ 2 ทำนายความเสี่ยงการเกิดโรคโรคความดันโลหิตสูงในประชากรทั่วไป จาก Cohort ของผู้ที่ยัง
ไม่เป็นโรคความดันโลหิตสูง (Cohort 2.1)

การเตรียมตัวแปรสถานะอุบัติการณ์การเป็นโรคความดันโลหิตสูง

1. เกณฑ์ที่ใช้กำหนดสถานะอุบัติการณ์การเป็นโรคความดันโลหิตสูง

ในการกำหนดสถานะอุบัติการณ์การเป็นโรคความดันโลหิตสูง ผู้ป่วยจะต้องมีลักษณะตรงตามเงื่อนไข
ดังต่อไปนี้อย่างน้อย 1 อย่าง ดังนี้

- (1) มีประวัติการวินิจฉัยการเป็นโรคความดันโลหิตสูง (ICD10: 110)
- (2) มีระดับความดันโลหิตเกินเกณฑ์ ณ วันที่มารับบริการคัดกรองโรค ดังนี้
 - (2.1) มีผลตรวจระดับความดันโลหิต Systolic Blood Pressure (SBP) \geq 180 mm Hg
 - (2.2) มีผลตรวจระดับความดันโลหิต Diastolic Blood Pressure (DBP) \geq 110 mm Hg

ด้วยเงื่อนไขข้างต้น พบว่าผู้ป่วยที่เป็นโรคความดันโลหิตสูงประมาณ 16.5% ของผู้ที่เข้ารับการคัดกรองทั้งหมด

2. การจัดการและทำความสะอาดตัวแปรตาม

หลังจากกำหนดสถานะอุบัติการณ์การเป็นโรคความดันโลหิตสูงแล้ว ในการสร้าง Cohort 2.1 จะดึงข้อมูลเฉพาะบุคคลที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคความดันโลหิตสูงเมื่อเริ่มต้น Cohort เท่านั้น สำหรับกรณีผู้ป่วยไม่ได้เข้ารับการคัดกรองทุกๆ ปี ซึ่งทำให้เกิดข้อมูลสูญหายบางปี จะเติมข้อมูลสถานะอุบัติการณ์การเป็นโรคความดันโลหิตสูงที่สูญหายไปด้วยกระบวนการเช่นเดียวกับวิธีการเติมข้อมูลสถานะอุบัติการณ์การเป็นโรคเบาหวาน

การเตรียมตัวแปรต่างๆ เพื่อใช้ทำนายอุบัติการณ์การเป็นโรคความดันโลหิตสูง

1. ตัวแปรต่างๆ ที่ใช้ในการทำนายอุบัติการณ์การเป็นโรคความดันโลหิตสูง

จากการทบทวนวรรณกรรม ทางทีมผู้วิจัยได้เตรียมข้อมูลตัวแปรต่างๆ ที่คาดการณ์ว่ามีความสัมพันธ์กับความเสี่ยงของการเกิดโรคความดันโลหิตสูง ซึ่งจะนำไปใช้เป็นตัวแปรอิสระในการสร้างตัวแบบเพื่อทำนายอุบัติการณ์การเป็นโรคความดันโลหิตสูงในปีถัดไป ทั้งนี้ รายละเอียดของตัวแปรต่างๆ ที่คาดการณ์ว่ามีความเสี่ยงต่อการเป็นโรคความดันโลหิตสูงและมีอยู่ในฐานข้อมูล 43 แฟ้มของกระทรวงสาธารณสุข จะแสดงดังตารางที่ 16

ตารางที่ 16 ตัวแปรที่นำมาพิจารณาในการสร้างตัวแบบทำนายความเสี่ยงของการเกิดโรคความดันโลหิตสูง

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ในตัวแบบ | Reference |
|------------------------|---|--|-------------------------------------|
| Demographic | | | |
| Age | อายุ (ปี) | age | 36,67–86,86 |
| Age category | ช่วงอายุ - 35 – 39 ปี - 40 – 44 ปี - 45 – 49 ปี - >=50 ปี | age_cat age_cat_1 age_cat_2 age_cat_3 | Use DM model categorical data |
| Gender | เพศ - ชาย - หญิง | t.sex_1 t.sex_2 | 36,67,69,71,72,74,75,77,79,81,82,85 |
| Health Behavior | | | |
| Smoking | พฤติกรรมการสูบบุหรี่ - สูบ - ไม่สูบ | SMOKE_1 | 36,67,69,71–73,75,76,78–82,85,86 |

ตารางที่ 16 ตัวแปรที่นำมาพิจารณาในการสร้างตัวแบบทำนายความเสี่ยงของการเกิดโรคความดันโลหิตสูง (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ในตัวแบบ | Reference |
|----------------------|---|--|-------------------------------|
| Physical measurement | | | |
| Height | ส่วนสูง (เซนติเมตร) | t.height | |
| BMI | ดัชนีมวลกาย | BMI | 67-72,74-77,79-83,85-87 |
| BMI_CAT1 | ดัชนีมวลกาย - < 23 - 23 - 27.5 - >= 27.5 | bmi_cat_1_2 bmi_cat_1_3 | Use DM model categorical data |
| BMI_CAT2 | ดัชนีมวลกาย - < 18.5 - 18.5 - 22.9 - 23 - 24.9 - 25 - 29.9 - >= 30 | bmi_cat_2_2 bmi_cat_2_3 bmi_cat_2_4 bmi_cat_2_5 | Use DM model categorical data |
| WAIST_CM | เส้นรอบเอว (เซนติเมตร) | WAIST_CM | 67,73,77,84,87 |
| WAIST_CAT | เส้นรอบเอว - ปกติ - ผิดปกติ เกณฑ์ผิดปกติในเพศชาย คือ >= 90 เซนติเมตร และในหญิงคือ >=80 เซนติเมตร | WAIST_CM_cat_2 | 36,78 |
| SBP | Systolic Blood Pressure (mm Hg) | SBP | 68-77,79,80,82-85,87 |
| DBP | Diastolic Blood Pressure (mm Hg) | DBP | 69-77,79,80,82-85,87 |
| PP | Pulse Pressure คำนวณจาก SBP - DBP | PP | Try test variables |

ตารางที่ 16 ตัวแปรที่นำมาพิจารณาในการสร้างตัวแบบทำนายความเสี่ยงของการเกิดโรคความดันโลหิตสูง (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ในตัวแบบ | Reference |
|---------------------------|---|----------------------------------|----------------------------------|
| PP_CAT | ตัวแปร PP ที่นำมาจัด กลุ่ม < 40 40 – 60 > 60 | PP_CAT PP_CAT2 PP_CAT3 | Try test variables |
| Family History | | | |
| Family Hx of DM | ประวัติเบาหวานในญาติ สายตรง - มี - ไม่มี | DMFAMILY_1 | Try test variables |
| Family Hx of HT | ประวัติความดันโลหิตสูง ในญาติสายตรง - มี - ไม่มี | HTFAMILY_1 | 36,68,69,71,75,76,82- 86 |
| Underlying Disease | | | |
| DM | ภาวะการเป็น โรคเบาหวาน - ปกติ - ผิดปกติ เบาหวาน ไม่มีโรคแทรกซ้อน - ผิดปกติ เบาหวาน และมีโรคแทรกซ้อน | DM_1 DM_2 | 36,72,85 |
| Lab | | | |
| TG | ระดับไขมันไตรกลีเซอไรด์ (mg/dL) | TG_labresult | 86,87 |
| TG_CAT1 | ระดับไขมันไตรกลีเซอไรด์ - < 200 mg/dL - >=200 mg/dL - ไม่ทราบ | TG_cat_1_1 TG_cat_1_2 | Use DM model categorical data |

ตารางที่ 16 ตัวแปรที่นำมาพิจารณาในการสร้างตัวแบบทำนายความเสี่ยงของการเกิดโรคความดันโลหิตสูง (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ในตัวแบบ | Reference |
|------------|--|----------------------------|----------------------------------|
| TG_CAT2 | ระดับไขมันไตรกลีเซอไรด์ - < 150 mg/dL - >= 150 mg/dL - ไม่ทราบ | TG_cat_2_1 TG_cat_2_2 | Use DM model categorical data |
| Is_na_TG* | ทราบค่าระดับไขมันไตร กลีเซอไรด์หรือไม่ - ทราบ - ไม่ทราบ | is_na_TG_1 | For adjusting data |
| HDL | ระดับไขมันดี (mg/dL) | HDL_labresult | 77 |
| HDL_CAT1 | ระดับไขมันดี - ผิดปกติ - ปกติ - ไม่ทราบ เกณฑ์ผิดปกติในเพศชาย คือ < 40 mg/dL และใน เพศหญิงคือ < 50 mg/dL | HDL_cat_1_1 HDL_cat_1_2 | Use DM model categorical data |
| Is_na_HDL* | ทราบค่าระดับไขมันดี หรือไม่ - ทราบ - ไม่ทราบ | is_na_HDL_1 | For adjusting data |

หมายเหตุ * หมายถึงตัวแปรเพิ่มเติมที่ได้จากการเตรียมข้อมูล

2. การจัดการและการทำความสะอาดตัวทำนาย

จากตารางที่ 16 จะเห็นได้ชัดว่าตัวแปรอิสระในการทำนายความเสี่ยงการเกิดโรคความดันโลหิตสูงมีความคล้ายคลึงกับตัวแปรอิสระที่ใช้ในการทำนายความเสี่ยงการเกิดโรคเบาหวานในประชากรไทยเป็นอย่างมาก จึงใช้หลักการการจัดการและการทำความสะอาดตัวทำนายเช่นเดียวกันกับที่ได้อธิบายไว้ก่อนหน้านี้

การวิเคราะห์ข้อมูลเชิงสำรวจ

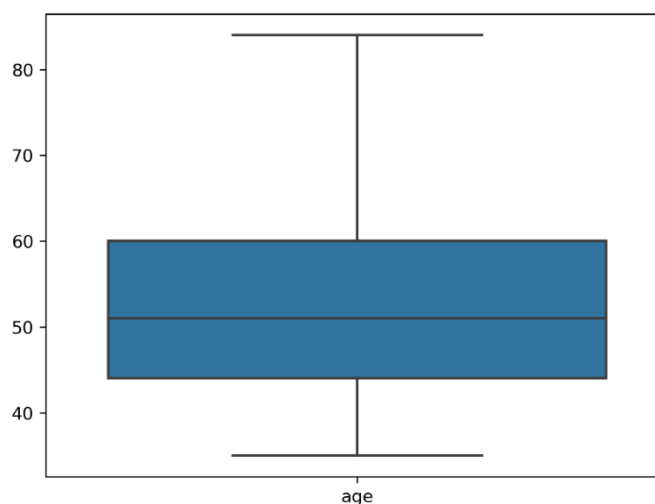
ในเบื้องต้นของการวิเคราะห์ ผู้วิจัยได้ทำการวิเคราะห์ข้อมูลเชิงสำรวจเพื่อดูการกระจายของข้อมูลตัวแปรต่าง ๆ ที่เกี่ยวข้อง โดยพบว่า

- อายุ (Age)

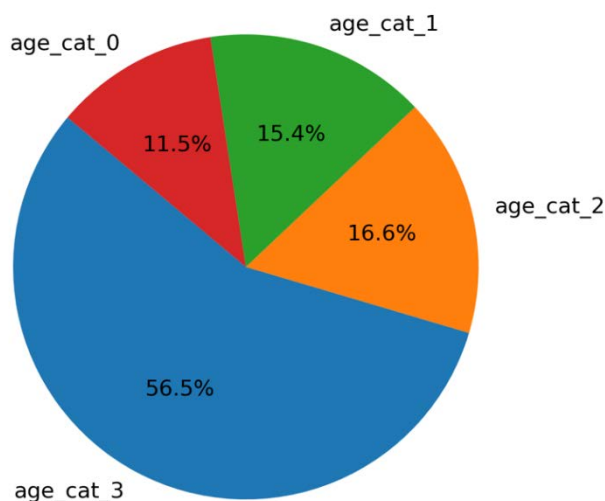
อายุของผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคโรคความดันโลหิตสูง เมื่อเริ่มต้น Cohort เท่านั้น มีอายุเฉลี่ยอยู่ที่ 52.76 ปี และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 11.22 ปี ดังแสดงในตารางที่ 17 และรูปที่ 29 และเมื่อจำแนกช่วงอายุ พบว่าส่วนใหญ่อยู่ในช่วงอายุ 45 - 49 ปี (age_cat_3) ร้อยละ 52.1 ช่วงอายุ 40 - 44 ปี (age_cat_2) ร้อยละ 16.6 ช่วงอายุ 35 - 39 ปี (age_cat_1) ร้อยละ 15.4 และช่วงอายุมากกว่าหรือเท่ากับ 50 ปี ขึ้นไป (age_cat_0) ร้อยละ 11.5 ดังแสดงในรูปที่ 30

ตารางที่ 17 ค่าเฉลี่ยเลขคณิต มัธยฐาน ส่วนเบี่ยงเบนมาตรฐาน ค่าต่ำสุด ค่าสูงสุดของอายุผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคความดันโลหิตสูงเมื่อเริ่มต้น Cohort เท่านั้น

| ตัวแปร | ค่าเฉลี่ย | มัธยฐาน | S.D. | Min | Max |
|-----------|-----------|---------|-------|-----|-----|
| อายุ (ปี) | 52.76 | 51 | 11.22 | 35 | 84 |



รูปที่ 29 แผนภาพกล่อง แสดงถึงการกระจายข้อมูลอายุของผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคความดันโลหิตสูงเมื่อเริ่มต้น Cohort เท่านั้น



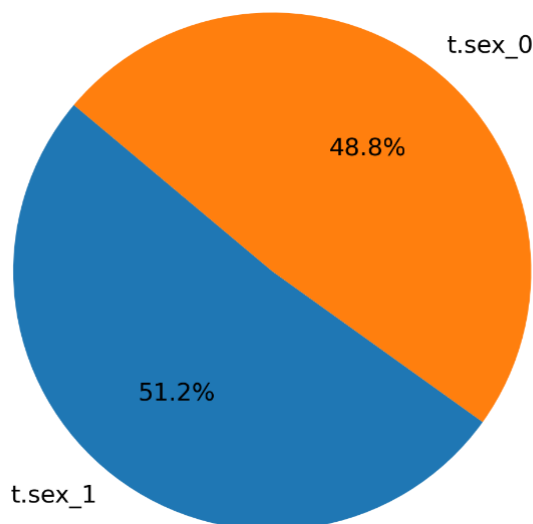
รูปที่ 30 แผนภูมิวงกลม แสดงถึงข้อมูลร้อยละ ผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคความดันโลหิตสูงเมื่อเริ่มต้น Cohort เท่านั้น จำแนกตามช่วงอายุ

- เพศ (Gender)

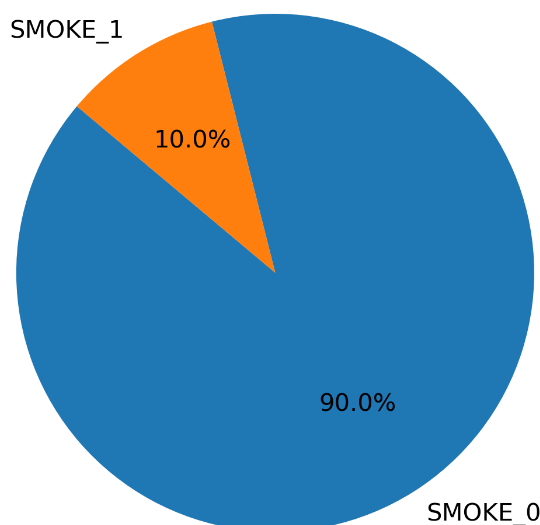
เพศของผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคความดันโลหิตสูงเมื่อเริ่มต้น Cohort เท่านั้น พบว่าเป็นเพศชาย (t.sex_1) ร้อยละ 51.2 และเป็นเพศหญิง (t.sex_2) ร้อยละ 48.8 ดังแสดงในรูปที่ 31

- Smoking (พฤติกรรมการสูบบุหรี่)

พฤติกรรมการสูบบุหรี่ของผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคความดันโลหิตสูงเมื่อเริ่มต้น Cohort เท่านั้น ส่วนใหญ่เป็นผู้ที่ไม่สูบบุหรี่ (SMOKE_0) ร้อยละ 90 และเป็นผู้ที่สูบบุหรี่ (SMOKE_1) ร้อยละ 10 ดังแสดงในรูปที่ 32



รูปที่ 31 แผนภูมิวงกลม แสดงถึงข้อมูลร้อยละ ผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคความดันโลหิตสูงเมื่อเริ่มต้น Cohort เท่านั้น จำแนกตามเพศ



รูปที่ 32 แผนภูมิวงกลมแสดงถึงข้อมูลร้อยละ ผู้ป่วยที่เข้ารับการศัลยกรรมและพบว่าไม่เป็นโรคความดันโลหิตสูงเมื่อเริ่มต้น Cohort เท่านั้น จำแนกตามพฤติกรรมการสูบบุหรี่

- Height (ส่วนสูง)

ส่วนสูงของผู้ป่วยที่เข้ารับการศัลยกรรมและพบว่าไม่เป็นโรคความดันโลหิตสูงเมื่อเริ่มต้น Cohort เท่านั้น มีส่วนสูงเฉลี่ยอยู่ที่ 160.33 เซนติเมตร ดังแสดงในตารางที่ 18

ตารางที่ 18 ค่าเฉลี่ยเลขคณิต มัธยฐาน ส่วนเบี่ยงเบนมาตรฐาน ค่าต่ำสุด ค่าสูงสุดของส่วนสูงผู้ป่วยที่เข้ารับการศัลยกรรมและพบว่าไม่เป็นโรคความดันโลหิตสูงเมื่อเริ่มต้น Cohort เท่านั้น

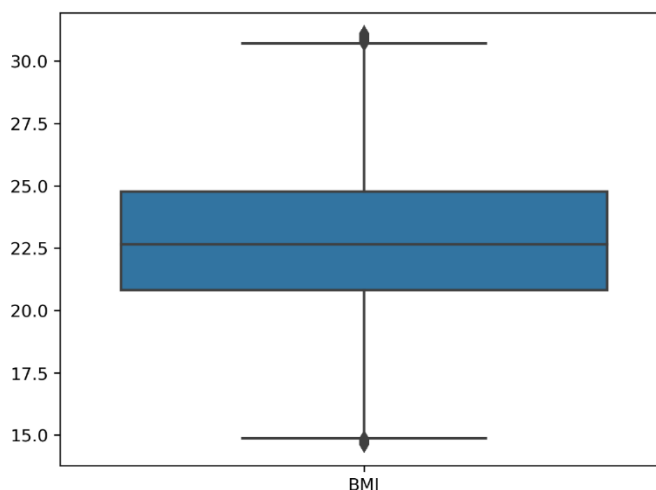
| ตัวแปร | ค่าเฉลี่ย | มัธยฐาน | S.D. | Min | Max |
|------------------------|-----------|---------|------|-----|-----|
| ส่วนสูง (เซนติเมตร) | 160.33 | 160 | 7.44 | 140 | 180 |

- BMI (ดัชนีมวลกาย)

ค่าดัชนีมวลกายของผู้ป่วยที่เข้ารับการศัลยกรรมและพบว่าไม่เป็นโรคความดันโลหิตสูงเมื่อเริ่มต้น Cohort เท่านั้น มีค่าเฉลี่ยอยู่ที่ 23.02 ดังแสดงในตารางที่ 19 และรูปที่ 33

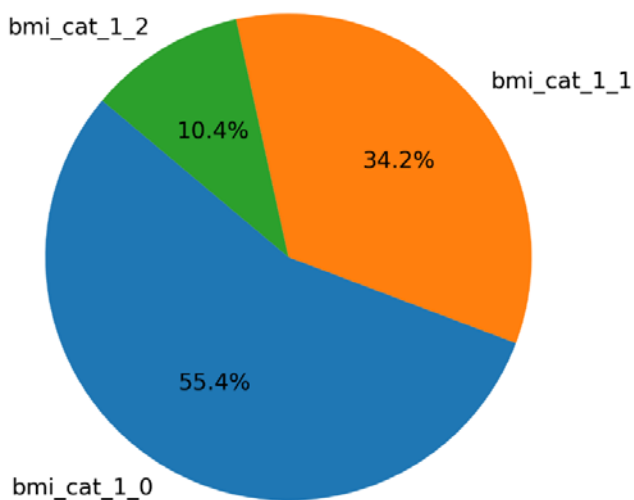
ตารางที่ 19 ค่าเฉลี่ยเลขคณิต มัธยฐาน ส่วนเบี่ยงเบนมาตรฐาน ค่าต่ำสุด ค่าสูงสุดของดัชนีมวลกาย (BMI) ผู้ป่วยที่
เข้ารับการคัดกรองและพบว่าไม่เป็นโรคเบาหวานเมื่อเริ่มต้น Cohort เท่านั้น

| ตัวแปร | ค่าเฉลี่ย | มัธยฐาน | S.D. | Min | Max |
|----------------------|-----------|---------|------|------|------|
| ดัชนีมวลกาย (BMI) | 23.02 | 22.66 | 3.26 | 14.6 | 31.1 |

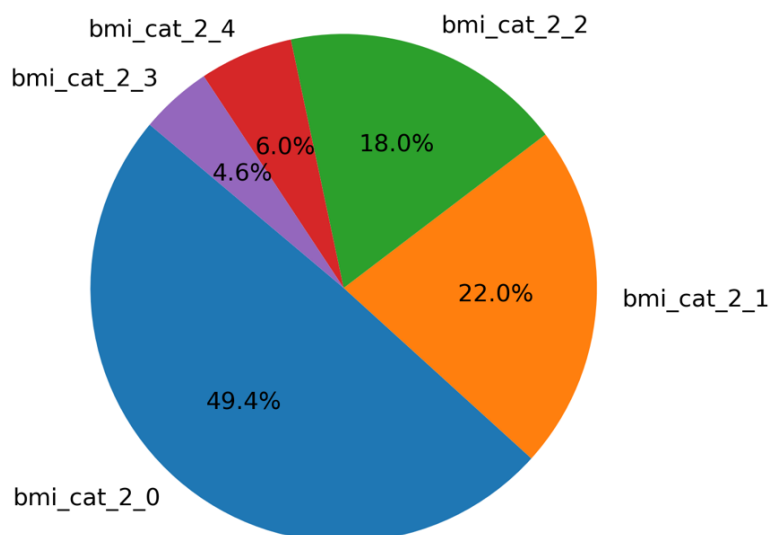


รูปที่ 33 แผนภาพกล่องแสดงถึงการกระจายข้อมูลค่าดัชนีมวลกาย (BMI) ของผู้ป่วยที่เข้ารับการคัดกรองและ
พบว่าไม่เป็นโรคความดันโลหิตสูงเมื่อเริ่มต้น Cohort เท่านั้น

เมื่อจำแนกดัชนีมวลกาย (BMI) ตามกลุ่มแบบที่ 1 พบว่า จำนวนผู้ป่วยสูงสุด อยู่ใน กลุ่มที่มีค่า BMI น้อยกว่า 23 (bmi_cat_1_0) ร้อยละ 55.4 รองลงมา กลุ่มที่มีค่า BMI อยู่ในช่วง 23 – 27.5 (bmi_cat_1_1) ร้อยละ 34.2 และกลุ่มที่มีค่า BMI มากกว่าหรือเท่ากับ 27.5 (bmi_cat_1_2) ร้อยละ 10.4 ดังแสดงในรูปที่ 34 และเมื่อจำแนกดัชนีมวลกาย (BMI) ตามกลุ่มแบบที่ 2 พบว่า จำนวนผู้ป่วยสูงสุด อยู่ใน กลุ่มที่มีค่า BMI น้อยกว่า 18.5 (bmi_cat_2_0) ร้อยละ 49.4 รองลงมา กลุ่มที่มีค่า BMI อยู่ในช่วง 18.5 – 22.9 (bmi_cat_2_1) ร้อยละ 22.0 กลุ่มที่มีค่า BMI อยู่ในช่วง 23 – 24.9 (bmi_cat_2_2) ร้อยละ 18.0 กลุ่มที่มีค่า BMI มากกว่าหรือเท่ากับ 30 (bmi_cat_2_4) ร้อยละ 6.0 และกลุ่มที่มีค่า BMI อยู่ในช่วง 25 – 29.9 (bmi_cat_2_3) ร้อยละ 4.6 ดังแสดงในรูปที่ 35



รูปที่ 34 แผนภูมิวงกลมแสดงถึงข้อมูลร้อยละ ผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคความดันโลหิตสูงเมื่อเริ่มต้น Cohort เท่านั้น จำแนกตามดัชนีมวลกายแบบที่ 1



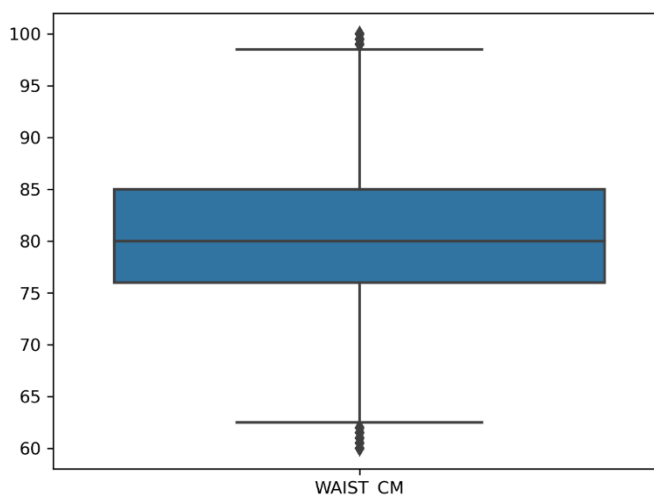
รูปที่ 35 แผนภูมิวงกลม แสดงถึงข้อมูลร้อยละ ผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคความดันโลหิตสูงเมื่อเริ่มต้น Cohort เท่านั้น จำแนกตามดัชนีมวลกายแบบที่ 2

- WAIST (เส้นรอบเอว)

เส้นรอบเอวของผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคความดันโลหิตสูงเมื่อเริ่มต้น Cohort เท่านั้น มีขนาดเฉลี่ยอยู่ที่ 80.36 เซนติเมตร ดังแสดงในตารางที่ 20

ตารางที่ 20 ค่าเฉลี่ยเลขคณิต มัธยฐาน ส่วนเบี่ยงเบนมาตรฐาน ค่าต่ำสุด ค่าสูงสุด ขนาดเส้นรอบเอวของผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคความดันโลหิตสูงเมื่อเริ่มต้น Cohort เท่านั้น

| ตัวแปร | ค่าเฉลี่ย | มัธยฐาน | S.D. | Min | Max |
|---------------------------|-----------|---------|------|-----|-----|
| เส้นรอบเอว (เซนติเมตร) | 80.36 | 80 | 7.91 | 60 | 100 |



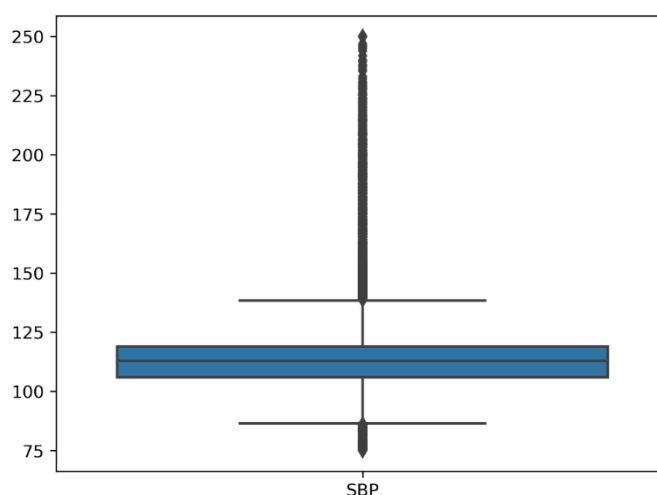
รูปที่ 36 แผนภาพกล่องแสดงถึงการกระจายข้อมูลค่าเส้นรอบเอว ของผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคความดันโลหิตสูงเมื่อเริ่มต้น Cohort เท่านั้น

- SBP (ความดันช่วงหัวใจบีบตัว)

ค่าความดันช่วงหัวใจบีบตัวของผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคความดันโลหิตสูงเมื่อเริ่มต้น Cohort เท่านั้น มีค่าเฉลี่ยอยู่ที่ 110.86 มิลลิเมตรปรอท ดังแสดงในตารางที่ 21

ตารางที่ 21 ค่าเฉลี่ยเลขคณิต มัธยฐาน ส่วนเบี่ยงเบนมาตรฐาน ค่าต่ำสุด ค่าสูงสุดของความดันช่วงหัวใจบีบตัว สำหรับผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคความดันโลหิตสูงเมื่อเริ่มต้น Cohort เท่านั้น

| ตัวแปร | ค่าเฉลี่ย | มัธยฐาน | S.D. | Min | Max |
|-------------|-----------|---------|-------|-----|-----|
| SBP (mm Hg) | 110.86 | 113 | 16.61 | 75 | 250 |



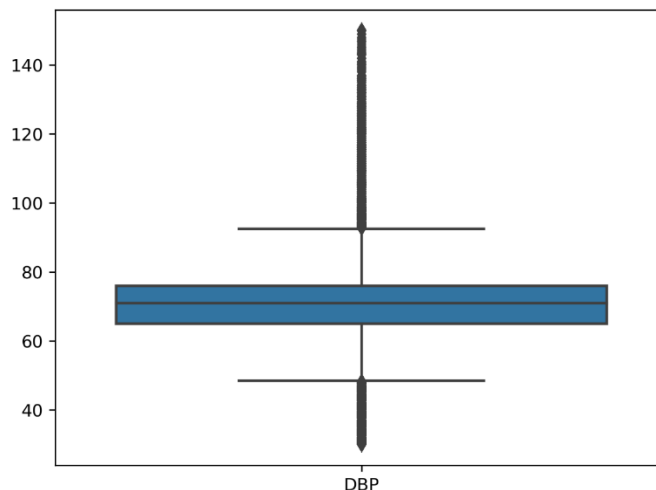
รูปที่ 37 แผนภาพกล่องแสดงถึงการกระจายข้อมูลค่าความดันช่วงหัวใจบีบตัว ของผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคความดันโลหิตสูงเมื่อเริ่มต้น Cohort เท่านั้น

- DBP (ความดันช่วงหัวใจคลายตัว)

ค่าความดันช่วงหัวใจคลายตัวของผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคความดันโลหิตสูงเมื่อเริ่มต้น Cohort เท่านั้น มีค่าเฉลี่ยอยู่ที่ 68.43 มิลลิเมตรปรอท ดังแสดงในตารางที่ 22

ตารางที่ 22 ค่าเฉลี่ยเลขคณิต มัธยฐาน ส่วนเบี่ยงเบนมาตรฐาน ค่าต่ำสุด ค่าสูงสุดของความดันช่วงหัวใจคลายตัว สำหรับผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคความดันโลหิตสูงเมื่อเริ่มต้น Cohort เท่านั้น

| ตัวแปร | ค่าเฉลี่ย | มัธยฐาน | S.D. | Min | Max |
|-------------|-----------|---------|-------|-----|-----|
| DBP (mm Hg) | 68.43 | 71 | 13.62 | 30 | 150 |



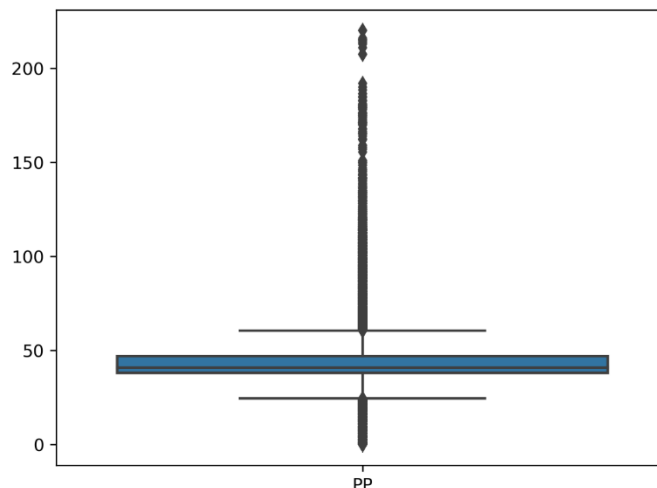
รูปที่ 38 แผนภาพกล่องแสดงถึงการกระจายข้อมูลค่าความดันช่วงหัวใจคลายตัว ของผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคความดันโลหิตสูงเมื่อเริ่มต้น Cohort เท่านั้น

- PP (Pulse Pressure คำนวณจาก SBP - DBP)

ค่า Pulse Pressure ซึ่งถูกคำนวณจากผลต่างระหว่างค่าความดันช่วงหัวใจบีบตัว กับค่าความดันช่วงหัวใจคลายตัว ของผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคความดันโลหิตสูงเมื่อเริ่มต้น Cohort เท่านั้น มีค่าเฉลี่ยอยู่ที่ 42.44 มิลลิเมตรปรอท ดังแสดงในตารางที่ 23

ตารางที่ 23 ค่าเฉลี่ยเลขคณิต มัธยฐาน ส่วนเบี่ยงเบนมาตรฐาน ค่าต่ำสุด ค่าสูงสุดของค่า Pulse Pressure สำหรับผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคความดันโลหิตสูงเมื่อเริ่มต้น Cohort เท่านั้น

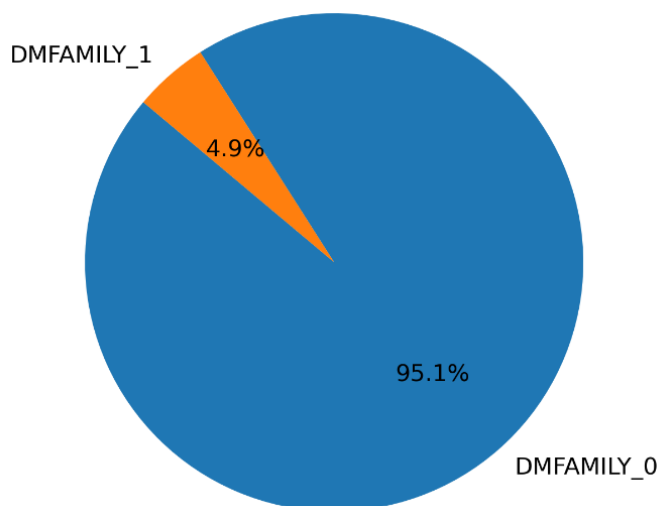
| ตัวแปร | ค่าเฉลี่ย | มัธยฐาน | S.D. | Min | Max |
|------------|-----------|---------|------|-----|-----|
| PP (mm Hg) | 42.44 | 41 | 8.78 | 0 | 220 |



รูปที่ 39 แผนภาพกล่องแสดงถึงการกระจายข้อมูลค่า Pulse Pressure ของผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคความดันโลหิตสูงเมื่อเริ่มต้น Cohort เท่านั้น

- Family Hx of DM (ประวัติเบาหวานในญาติสายตรง)

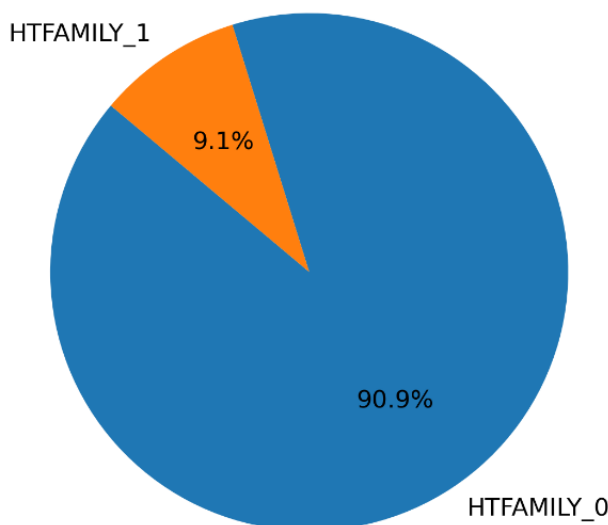
ประวัติเบาหวานในญาติสายตรงของผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคความดันโลหิตสูงเมื่อเริ่มต้น Cohort เท่านั้น ส่วนใหญ่เป็นผู้ที่ไม่มีประวัติเบาหวานในญาติสายตรง (DMFAMILY_0) ร้อยละ 95.1 และเป็นผู้ที่มีประวัติเบาหวานในญาติสายตรง (DMFAMILY_1) ร้อยละ 4.9 ดังแสดงในรูปที่ 40



รูปที่ 40 แผนภูมิวงกลมแสดงถึงข้อมูลร้อยละ ผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคความดันโลหิตสูงเมื่อเริ่มต้น Cohort เท่านั้น จำแนกตามการมีประวัติเบาหวานในญาติสายตรง

- Family Hx of HT (ประวัติความดันโลหิตสูงในญาติสายตรง)

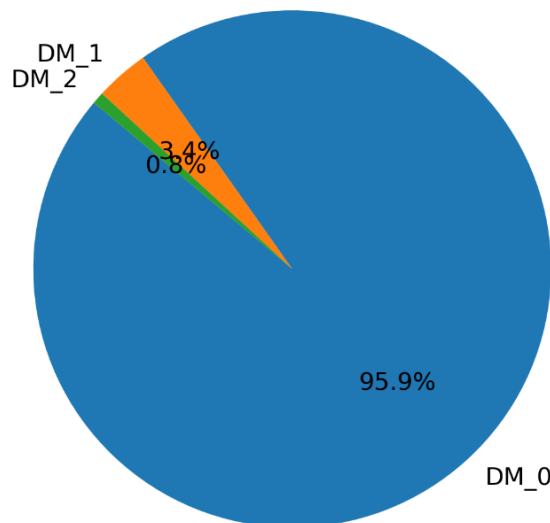
ประวัติความดันโลหิตสูงในญาติสายตรงของผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคความดันโลหิตสูง เมื่อเริ่มต้น Cohort เท่านั้น ส่วนใหญ่เป็นผู้ที่ไม่มีประวัติความดันโลหิตสูงในญาติสายตรง (HTFAMILY_0) ร้อยละ 90.9 และเป็นผู้ที่มีประวัติความดันโลหิตสูงในญาติสายตรง (HTFAMILY_1) ร้อยละ 9.1 ดังแสดงในรูปที่ 41



รูปที่ 41 แผนภูมิวงกลมแสดงถึงข้อมูลร้อยละ ผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคความดันโลหิตสูง เมื่อเริ่มต้น Cohort เท่านั้น จำแนกตามการมีประวัติความดันโลหิตสูงในญาติสายตรง

- DM (ภาวะการเป็นโรคเบาหวาน)

การเกิดภาวะการเป็นโรคเบาหวาน ของผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคความดันโลหิตสูง เมื่อเริ่มต้น Cohort เท่านั้น เมื่อจำแนกกลุ่มตามความผิดปกติของการเป็นโรคเบาหวาน พบว่าส่วนใหญ่เป็นผู้ที่ไม่ผิดปกติ (DM_0) ร้อยละ 95.9 เป็นผู้ที่ผิดปกติ เบาหวาน ไม่มีโรคแทรกซ้อน (DM_1) ร้อยละ 3.4 และเป็นผู้ที่ผิดปกติ เบาหวานและมีโรคแทรกซ้อน (DM_2) ร้อยละ 0.8 ดังแสดงในรูปที่ 42



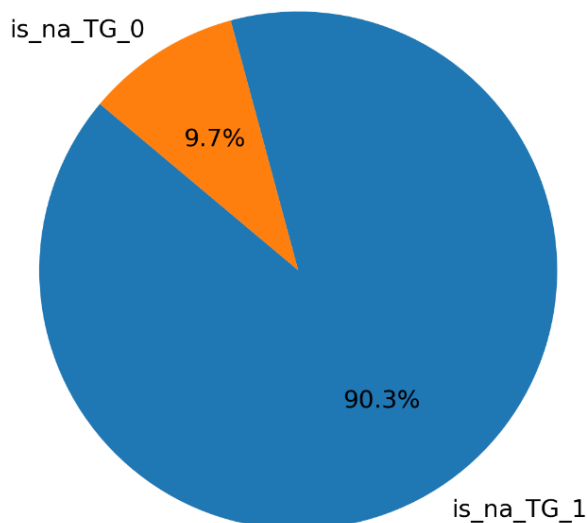
รูปที่ 42 แผนภูมิวงกลมแสดงถึงข้อมูลร้อยละ ผู้ป่วยที่เข้ารับการรักษาและพบว่าไม่เป็นโรคความดันโลหิตสูงเมื่อเริ่มต้น Cohort เท่านั้น จำแนกตามการเป็นโรคเบาหวาน

- TG (ระดับไขมันไตรกลีเซอไรด์)

ค่าระดับไขมันไตรกลีเซอไรด์ ของผู้ป่วยที่เข้ารับการรักษาและพบว่าไม่เป็นโรคความดันโลหิตสูงเมื่อเริ่มต้น Cohort เท่านั้น มีค่าเฉลี่ยอยู่ที่ 13.90 มิลลิกรัม/เดซิลิตร ดังแสดงในตารางที่ 24 ส่วนใหญ่ไม่ทราบระดับไขมันไตรกลีเซอไรด์ ร้อยละ 90.3 และทราบระดับไขมันไตรกลีเซอไรด์ ร้อยละ 9.7 ดังแสดงในรูปที่ 43

ตารางที่ 24 ค่าเฉลี่ยเลขคณิต มัธยฐาน ส่วนเบี่ยงเบนมาตรฐาน ค่าต่ำสุด ค่าสูงสุด ค่าระดับไขมันไตรกลีเซอไรด์ของผู้ป่วยที่เข้ารับการรักษาและพบว่าไม่เป็นโรคความดันโลหิตสูงเมื่อเริ่มต้น Cohort เท่านั้น

| ตัวแปร | ค่าเฉลี่ย | มัธยฐาน | S.D. | Min | Max |
|---|-----------|---------|-------|-----|-------|
| ระดับไขมันไตรกลีเซอไรด์ (มิลลิกรัม/เดซิลิตร) | 13.90 | 0 | 47.77 | 0 | 312.4 |



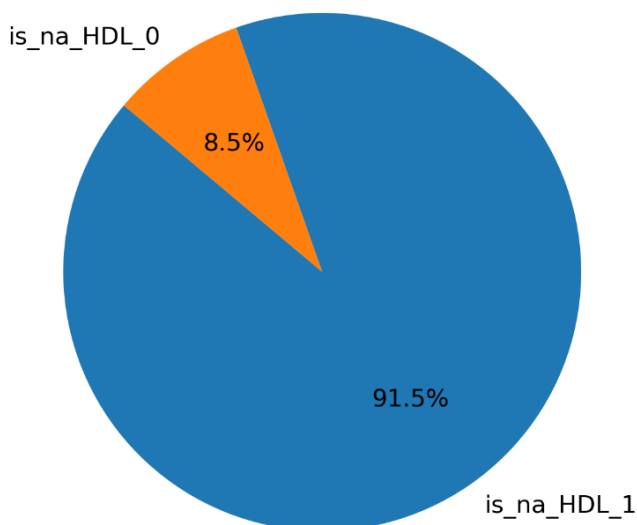
รูปที่ 43 แผนภูมิวงกลมแสดงถึงข้อมูลร้อยละ ผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคความดันโลหิตสูงเมื่อเริ่มต้น Cohort เท่านั้น จำแนกตามการทราบค่าไขมันไตรกลีเซอไรด์

- HDL (ระดับไขมันดี)

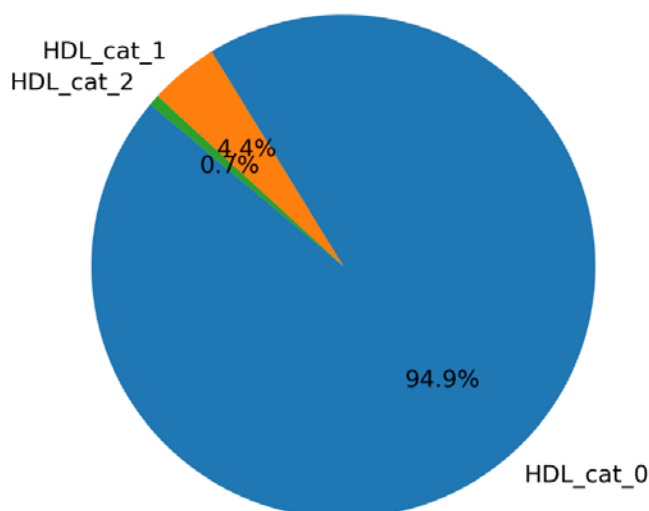
ค่าระดับไขมันดีของผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคความดันโลหิตสูงเมื่อเริ่มต้น Cohort เท่านั้น มีระดับไขมันดีเฉลี่ยอยู่ที่ 4.39 มิลลิกรัมต่อเดซิลิตร ดังแสดงในตารางที่ 25 และเมื่อจำแนกผู้ป่วยตามการทราบระดับไขมันดี ส่วนใหญ่ไม่ทราบระดับไขมันดี ร้อยละ 91.5 และทราบระดับไขมันไตรกลีเซอไรด์ ร้อยละ 8.5 ดังแสดงในรูปที่ 44 นอกจากนี้ เมื่อจำแนกกลุ่มระดับไขมันดี ที่มีเกณฑ์ผิดปกติในเพศชาย คือ < 40 mg/dL และในเพศหญิงคือ < 50 mg/dL พบว่าส่วนใหญ่เป็นผู้ที่ผิดปกติ (HDL_cat_0) ร้อยละ 94.9 เป็นผู้ปกติ (HDL_cat_1) ร้อยละ 4.4 และเป็นผู้ที่ไม่ทราบ(HDL_cat_2) ร้อยละ 0.7 ดังแสดงในรูปที่ 45

ตารางที่ 25 ค่าเฉลี่ยเลขคณิต มัธยฐาน ส่วนเบี่ยงเบนมาตรฐาน ค่าต่ำสุด ค่าสูงสุดของระดับไขมันดี สำหรับผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคความดันโลหิตสูงเมื่อเริ่มต้น Cohort เท่านั้น

| ตัวแปร | ค่าเฉลี่ย | มัธยฐาน | S.D. | Min | Max |
|-------------|-----------|---------|-------|-----|-----|
| HDL (mg/dL) | 4.39 | 0 | 14.97 | 0 | 87 |



รูปที่ 44 แผนภูมิวงกลมแสดงถึงข้อมูลร้อยละ ผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคความดันโลหิตสูงเมื่อเริ่มต้น Cohort เท่านั้น จำแนกตามการทราบค่าระดับไขมันดี



รูปที่ 45 แผนภูมิวงกลม แสดงถึงข้อมูลร้อยละ ผู้ป่วยที่เข้ารับการคัดกรองและพบว่าไม่เป็นโรคความดันโลหิตสูงเมื่อเริ่มต้น Cohort เท่านั้น จำแนกตามระดับไขมันดี

ผลลัพธ์จากตัวแบบทำนายความเสี่ยงการเกิดโรคความดันโลหิตสูงในประชากรทั่วไป

จากกลุ่มตัวอย่างของผู้เข้ารับการคัดกรองโรคความดันโลหิตสูงจำนวน 1,016,194 คน ซึ่งประกอบได้ด้วยข้อมูลจำนวน 4,190,538 แถวหรือจำนวนครั้งที่เข้ารับการคัดกรองทั้งหมด จะถูกสุ่มแบ่งออกเป็นชุดข้อมูลฝึกฝนและชุดข้อมูลทดสอบคิดเป็นร้อยละ 70 และ 30 โดยประมาณ ตามลำดับ ซึ่งจากการแบ่งข้อมูลจะได้ข้อมูลดังต่อไปนี้

- ชุดข้อมูลฝึกฝน ประกอบด้วยข้อมูลจากผู้เข้ารับการคัดกรองโรคความดันโลหิตสูงจำนวน 711,335 คน ซึ่งมีข้อมูลจำนวน 2,931,963 แถว

- ชุดข้อมูลทดสอบ ประกอบด้วยข้อมูลจากผู้เข้ารับการคัดกรองโรคความดันโลหิตสูงจำนวน 304,859 คน ซึ่งมีข้อมูลจำนวน 1,258,575 แถว

จากนั้นจึงสร้างตัวแบบจากข้อมูลชุดฝึกฝน ซึ่งจะพิจารณา 2 ตัวแบบ ได้แก่

1. การถดถอยโลจิสติกแบบ Elastic Net โดยจะทำขั้นตอน 2 ขั้นตอนในการคัดเลือกตัวแปร โดยขั้นที่ 1 หากตัวแปรทำนายบางตัวมีความเกี่ยวข้องกัน เช่น ตัวแปร TG TG_CAT1 และ TG_CAT2 ต่างก็คือตัวแปรระดับไขมันไตรกลีเซอไรด์ทั้งสามตัวแปร เป็นต้น จะพิจารณาหาเซตของตัวแปรทำนายที่ให้ผลทำนายดีที่สุด และในขั้นที่ 2 จะนำเซตของตัวแปรทำนายเหล่านั้นเข้ากระบวนการสร้างตัวแบบของ EN ในการคัดเลือกตัวแปรอัตโนมัติอีกหนึ่ง วิธีการนี้ขอเรียกว่า “EN”
2. วิธี XGBoost ที่พิจารณาตัวแปรทำนายทุกตัวแปรและให้วิธีการ XGBoost คัดเลือกตัวแปรให้อัตโนมัติ โดยวิธีการนี้ขอเรียกว่า “XGBoost”

หลังจากได้ตัวแบบทั้ง 2 แล้ว จึงนำมาทำนายข้อมูลในชุดทดสอบเพื่อวัดประสิทธิภาพการทำนายโดยพิจารณาจาก

1. AUC
2. Cumulative AUC

ตารางที่ 26 แสดงตัววัดประสิทธิภาพทั้ง 2 ตัววัด ซึ่งจะเห็นได้ชัดว่า วิธี XGBoost มีประสิทธิภาพในการทำนายความเสี่ยงการเกิดโรคความดันโลหิตสูงในประชากรทั่วไป จาก Cohort ของผู้ที่ยังไม่เป็นโรคความดันโลหิตสูงได้ดีกว่าวิธีการถดถอยโลจิสติกแบบ Elastic Net อย่างไรก็ตาม ประสิทธิภาพการทำนายของวิธีการถดถอยโลจิสติกแบบ Elastic Net ยังอยู่ในเกณฑ์ดี คือ มีค่า AUC ทั้งสองตัวเกิน 0.8

ตารางที่ 26 ค่า AUC เพื่อวัดประสิทธิภาพการทำนายความเสี่ยงการเกิดโรคความดันโลหิตสูงในประชากรทั่วไป จาก Cohort ของผู้ที่ยังไม่เป็นโรคความดันโลหิตสูง (Cohort 2.1) (วิธีที่ดีที่สุดจะแสดงเป็นตัวหนา)

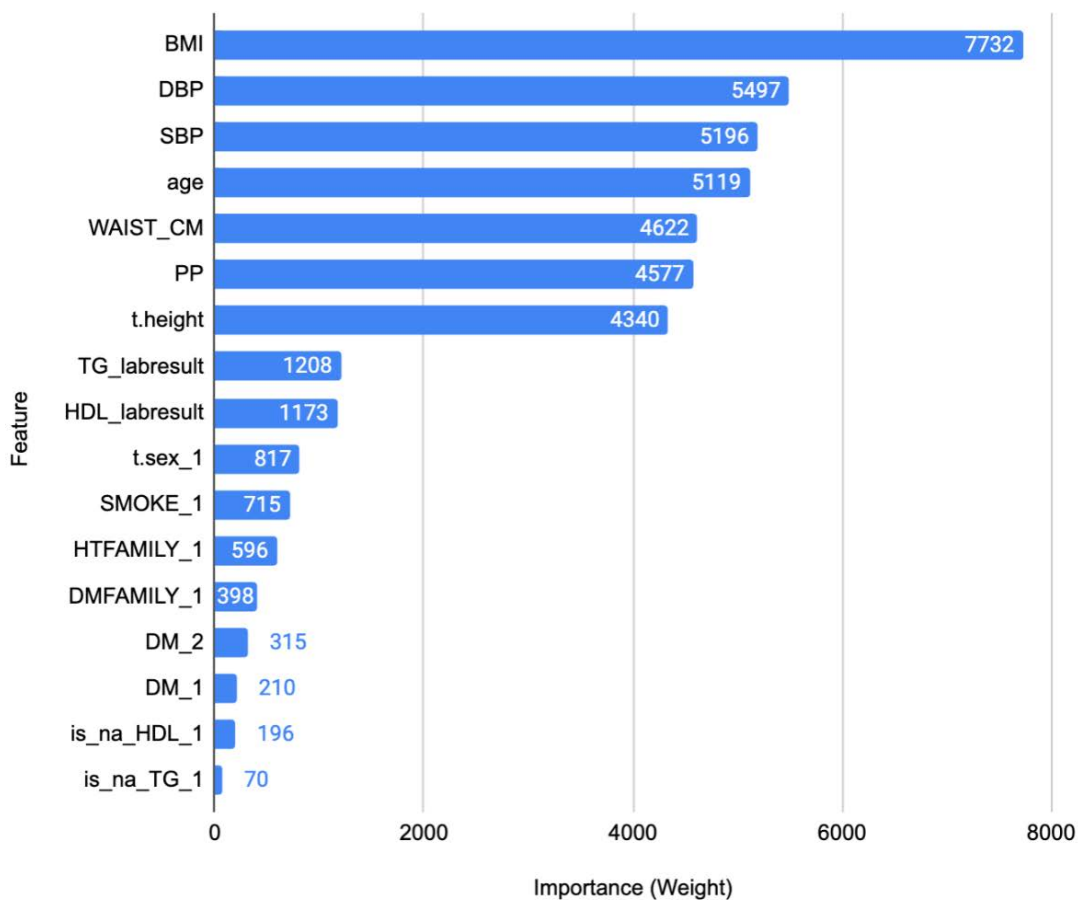
| ตัวแบบ | AUC | Cumulative AUC |
|---------|---------------|----------------|
| EN | 0.8286 | 0.8657 |
| XGBoost | 0.8372 | 0.8721 |

การแปลผลตัวแบบทำนายความเสี่ยงการเกิดโรคความดันโลหิตสูงในประชากรทั่วไป

จากตัวแบบ XGBoost ที่ให้ประสิทธิภาพการทำนายดีที่สุดสามารถพิจารณาได้ว่าปัจจัยใดที่มีความสำคัญในการทำนายความเสี่ยงการเกิดโรคเบาหวานในประชากรทั่วไปดังแสดงในรูปที่ 46 ซึ่งจะพบว่าตัวแปรที่มีความสำคัญในการทำนายความเสี่ยง 10 อันดับแรก ได้แก่

1. ดัชนีมวลกาย (BMI)
2. Diastolic Blood Pressure (DBP)

3. Systolic Blood Pressure (SBP)
4. อายุ (Age)
5. เส้นรอบเอว (WAIST_CM)
6. Pulse Pressure (PP)
7. ส่วนสูง (t.height)
8. ระดับไขมันไตรกลีเซอไรด์ (TG_labresult)
9. ระดับไขมันดี (HDL_labresult)
10. เพศ (t.sex)

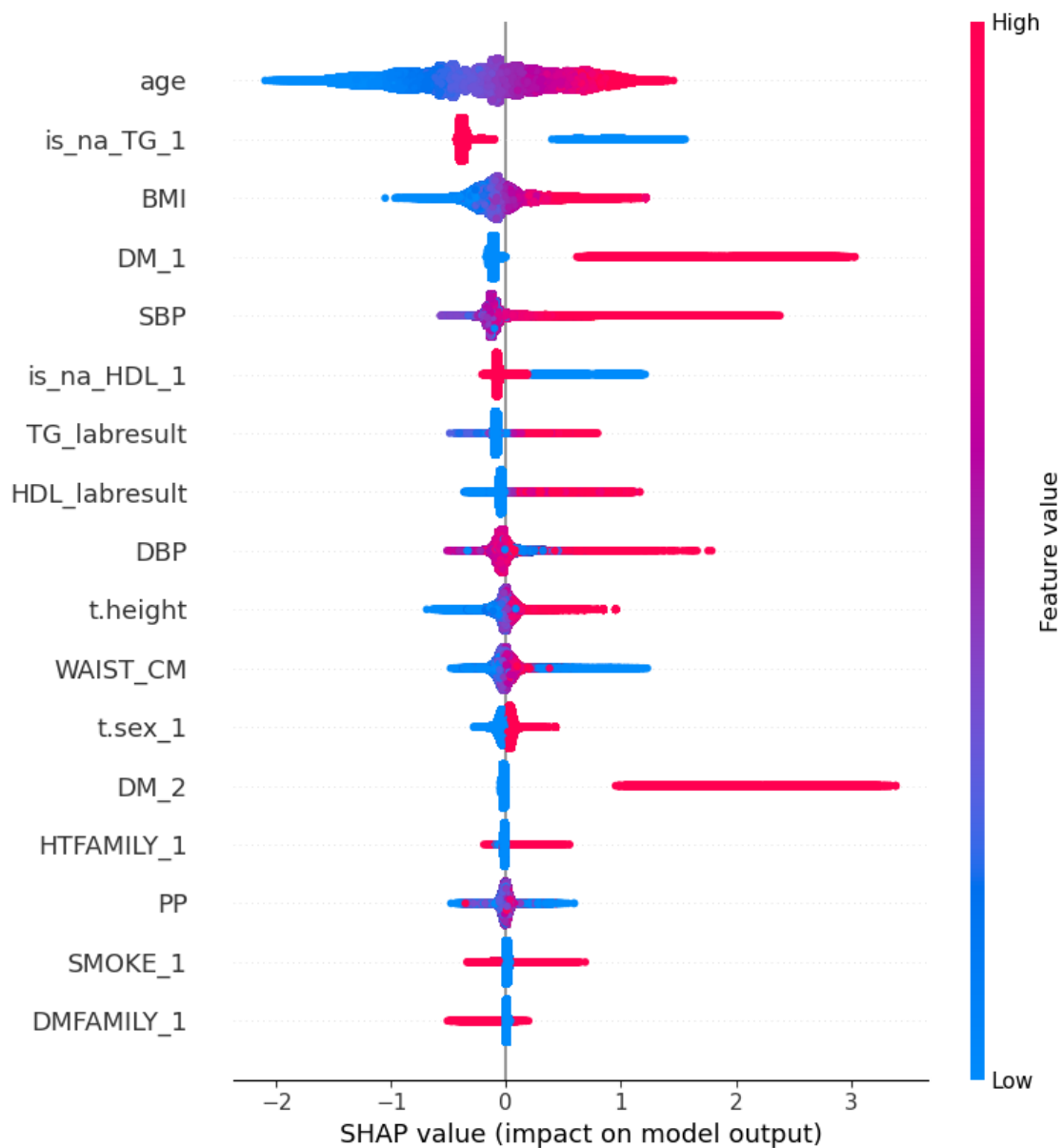


รูปที่ 46 ความสำคัญของตัวแปรทำนายจากวิธี XGBoost ในการทำนายความเสี่ยงการเกิดโรคความดันโลหิตสูงในประชากรไทย

นอกจากนี้ หากพิจารณาคุณค่า SHAP ดังรูปที่ 47 จะสามารถแปลผลในส่วนทิศทางความสัมพันธ์ของตัวแปรแต่ละตัว ซึ่งจากรูปที่ 47 สามารถสรุปความสัมพันธ์ที่เด่นชัดได้ ดังนี้

- ผู้คัดกรองที่มีอายุ (age) มากจะมีแนวโน้มที่จะความเสี่ยงการเป็นโรคความดันโลหิตสูงที่เพิ่มขึ้น
- ผู้ป่วยที่มีภาวะโรคเบาหวานรูปแบบที่ 1 (Type-I Diabetes) (DM_1.0) และ 2 (Type-II Diabetes) (DM_2.0) มีแนวโน้มที่จะความเสี่ยงการเป็นโรคความดันโลหิตสูงที่เพิ่มขึ้น
- ผู้คัดกรองที่มีระดับไขมันไตรกลีเซอไรด์ (TG_labresult) สูงจะมีแนวโน้มที่จะความเสี่ยงการเป็นโรคความดันโลหิตสูงที่เพิ่มขึ้น
- ผู้คัดกรองที่มีระดับไขมันดี (HDL_labresult) สูงจะมีแนวโน้มที่จะความเสี่ยงการเป็นโรคความดันโลหิตสูงที่เพิ่มขึ้น
- ผู้คัดกรองที่มีดัชนีมวลกาย (BMI) สูงจะมีแนวโน้มที่จะความเสี่ยงการเป็นโรคความดันโลหิตสูงที่เพิ่มขึ้น

แม้ว่าวิธี XGBoost สามารถระบุความสำคัญของตัวแปรทำนายที่ใช้ในตัวแบบได้ แต่ไม่สามารถแปลผลตัวแบบในรูปของอัตราเสี่ยงได้ ในขณะที่วิธีการถดถอยโลจิสติกแบบ Elastic Net ให้ค่าสัมประสิทธิ์การถดถอยเช่นเดียวกับตัวแบบการถดถอยโลจิสติกแบบปกติ ดังนั้นจึงสามารถแปลผลตัวแบบในรูปของอัตราเสี่ยงได้ ตารางที่ 27



รูปที่ 47 ค่า SHAP จากตัวแบบ XGBoost ในการทำนายความเสี่ยงการเกิดโรคความดันโลหิตสูงในประชากรไทย

ตารางที่ 27 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net
ในการทำนายความเสี่ยงการเกิดโรคความดันโลหิตสูงในประเทศไทย

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ในตัว แบบ | สัมประสิทธิ์การ ถดถอย | Odds Ratio |
|-----------------------------|---|--|--------------------------|------------|
| Demographic | | | | |
| Age | อายุ (ปี) | age | 0.0486 | 1.0498 |
| Age category | ช่วงอายุ - 35 – 39 ปี - 40 – 44 ปี - 45 – 49 ปี - >=50 ปี | age_cat age_cat_1 age_cat_2 age_cat_3 | | |
| Gender | เพศ - ชาย - หญิง | t.sex_2 | 0.1291 | 1.1378 |
| Health Behavior | | | | |
| Smoking | พฤติกรรมการสูบบุหรี่ - สูบ - ไม่สูบ | SMOKE_1 | 0.0274 | 1.0278 |
| Physical measurement | | | | |
| Height | ส่วนสูง (เซนติเมตร) | t.height | 0.0082 | 1.0082 |
| BMI | ดัชนีมวลกาย | BMI | 0.0750 | 1.0779 |
| BMI_CAT1 | ดัชนีมวลกาย - < 23 - 23 - 27.5 - >= 27.5 | bmi_cat_1_2 bmi_cat_1_3 | | |
| BMI_CAT2 | ดัชนีมวลกาย - < 18.5 - 18.5 – 22.9 - 23 – 24.9 - 25 – 29.9 - >= 30 | bmi_cat_2_2 bmi_cat_2_3 bmi_cat_2_4 bmi_cat_2_5 | | |
| WAIST_CM | เส้นรอบเอว (เซนติเมตร) | WAIST_CM | 0.0059 | 1.0059 |

ตารางที่ 27 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net
ในการทำนายความเสี่ยงการเกิดโรคความดันโลหิตสูงในประชากรไทย (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ในตัวแบบ | สัมประสิทธิ์การถดถอย | Odds Ratio |
|-----------------------|---|------------------------------|----------------------|------------------|
| WAIST_CAT | เส้นรอบเอว - ปกติ - ผิดปกติ เกณฑ์ผิดปกติในเพศชาย คือ ≥ 90 เซนติเมตร และในหญิงคือ ≥ 80 เซนติเมตร | WAIST_CM_cat_2 | | |
| SBP | Systolic Blood Pressure (mm Hg) | SBP | 0.0162 | 1.0163 |
| DBP | Diastolic Blood Pressure (mm Hg) | DBP | -0.0104 | 0.9896 |
| PP | Pulse Pressure คำนวณจาก SBP - DBP | PP | | |
| PP_CAT | ตัวแปร PP ที่นำมาจัดกลุ่ม < 40 40 – 60 > 60 | PP_CAT PP_CAT1 PP_CAT2 | 0.0851 0.2852 | 1.0888 1.3300 |
| Family History | | | | |
| Family Hx of DM | ประวัติเบาหวานในญาติสายตรง - มี - ไม่มี | DMFAMILY_1 | -0.0575 | 0.9441 |
| Family Hx of HT | ประวัติความดันโลหิตสูงในญาติสายตรง - มี - ไม่มี | HTFAMILY_1 | 0.1936 | 1.2136 |

ตารางที่ 27 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net
ในการทำนายความเสี่ยงการเกิดโรคความดันโลหิตสูงในประชากรไทย (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ในตัวแบบ | สัมประสิทธิ์การถดถอย | Odds Ratio |
|--------------------|--|--------------------------|----------------------|------------------|
| Underlying Disease | | | | |
| DM | ภาวะการเป็นโรคเบาหวาน - ปกติ - ผิดปกติ เบาหวานไม่มีโรคแทรกซ้อน - ผิดปกติ เบาหวานมีโรคแทรกซ้อน | DM_1 DM_2 | 1.4002 1.8455 | 4.0560 6.3313 |
| Lab | | | | |
| TG | ระดับไขมันไตรกลีเซอไรด์ (mg/dL) | TG_labresult | | |
| TG_CAT1 | ระดับไขมันไตรกลีเซอไรด์ - < 200 mg/dL - >=200 mg/dL - ไม่ทราบ | TG_cat_1_1 TG_cat_1_2 | | |
| TG_CAT2 | ระดับไขมันไตรกลีเซอไรด์ - < 150 mg/dL - >= 150 mg/dL - ไม่ทราบ | TG_cat_2_1 TG_cat_2_2 | 0.0000 0.0528 | 1.0000 1.0543 |
| Is_na_TG* | ทราบค่าระดับไขมันไตรกลีเซอไรด์หรือไม่ - ทราบ - ไม่ทราบ | is_na_TG_1 | -1.5992 | 0.2021 |
| HDL | ระดับไขมันดี (mg/dL) | HDL_labresult | | |

ตารางที่ 27 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net
ในการทำนายความเสี่ยงการเกิดโรคความดันโลหิตสูงในประชากรไทย (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ในตัวแบบ | สัมประสิทธิ์การถดถอย | Odds Ratio |
|------------|--|--------------------------|----------------------|------------|
| HDL_CAT1 | ระดับไขมันดี - ผิดปกติ - ปกติ - ไม่ทราบ เกณฑ์ผิดปกติในเพศชาย คือ < 40 mg/dL และใน เพศหญิงคือ < 50 mg/dL | HDL_cat_1_1 | -0.2208 | 0.8019 |
| | | HDL_cat_1_2 | -0.4472 | 0.6394 |
| Is_na_HDL* | ทราบค่าระดับไขมันดี หรือไม่ - ทราบ - ไม่ทราบ | is_na_HDL_1 | -0.9524 | 0.3858 |

แบบจำลองที่ 3 ทำนายความเสี่ยงการเกิดภาวะแทรกซ้อนและการเสียชีวิตจาก Cohort ของผู้ที่เป็นโรคเบาหวาน (Cohort 1.2)

กลุ่มโรคภาวะแทรกซ้อนและการเตรียมตัวแปรสถานะอุบัติการณ์โรคแทรกซ้อน

อุบัติการณ์ของภาวะแทรกซ้อนที่สำคัญ อันเนื่องมาจากโรคเบาหวานและความดันโลหิตสูง ประกอบด้วย 5 กลุ่มโรคกลุ่มโรค ได้แก่

1. ภาวะเส้นเลือดหัวใจตีบหรือกล้ามเนื้อหัวใจขาดเลือด (Cardiovascular Disease)
2. ภาวะเส้นเลือดสมองแตกหรือตีบ (Cerebrovascular Disease)
3. ภาวะแทรกซ้อนทางตาอันเนื่องมาจากโรคเบาหวานและความดันโลหิตสูง
4. ภาวะไตเสื่อมเรื้อรังระยะที่ 5
5. ภาวะแทรกซ้อนทางเส้นประสาทและเส้นเลือดส่วนปลายอันจำเป็นต้องเกิดการตัดรยางค์ (Amputation)

นอกจาก 5 กลุ่มโรคข้างต้นแล้ว ผู้ป่วยที่เป็นโรคเบาหวานอยู่ก่อนมักมีโอกาเป็นภาวะความดันโลหิตสูงร่วมด้วย และอาจมีโอกาเสียชีวิตได้ ดังนั้นในการสร้างแบบจำลองทำนายความเสี่ยงจึงพิจารณาการเกิดโรคแทรกซ้อน 5 กลุ่ม รวมไปถึงโรคความดันโลหิตสูงและการเสียชีวิตอีกด้วย สำหรับการเตรียมตัวแปรสถานะอุบัติการณ์โรคแทรกซ้อน 5 กลุ่มโรคนั้นจะพิจารณาจากรหัส ICD10 ที่ผู้ป่วยได้รับการวินิจฉัย ไม่ว่าจะรหัส ICD10 จะปรากฏในแฟ้มข้อมูลการวินิจฉัยโดยการตรวจผู้ป่วยนอก (แฟ้ม DIAGNOSIS_OPD) หรือ แฟ้มข้อมูล

การวินิจฉัยโดยการตรวจผู้ป่วยใน (เพิ่ม DIAGNOSIS_IPD) หรือเพิ่มข้อมูลการป่วยเรื้อรัง (เพิ่ม CHRONIC) โดยวันที่เริ่มเป็นภาวะแทรกซ้อน จะพิจารณาจากวันแรกที่ถูกระบุวินิจฉัยโรคและบันทึกระหว่าง 3 เพิ่มข้างต้น สำหรับรหัส ICD10 ในแต่ละกลุ่มโรคทั้ง 5 กลุ่มโรคแสดงดังตารางที่ 28

กลุ่มโรคทางไต นอกเหนือจากการกำหนดสถานะอุบัติการณ์โดยดูจากรหัส ICD10 ที่ผู้ป่วยได้รับการวินิจฉัยแล้ว ผู้ป่วยต้องมีค่า Estimated Glomerular Filtration Rate (eGFR) ต่ำกว่า 15 mL/min/1.73m² จึงถือว่าผู้ป่วยดังกล่าวมีโรคแทรกซ้อนทางไต สำหรับค่า eGFR ในผู้ป่วยเพศชาย คำนวณโดยใช้สูตรด้านล่าง

$$eGFR = 375.5 \times CR(Enz)^{-0.848} \times Age^{-0.364}$$

และสำหรับผู้หญิง ค่า eGFR สามารถคำนวณได้ ดังนี้

$$eGFR = 375.5 \times CR(Enz)^{-0.848} \times Age^{-0.364} \times 0.712$$

โดยที่ CR(Enz) คือค่า Enzymatic Creatinine และ Age คืออายุของผู้ป่วย

ตารางที่ 28 รหัส ICD10 สำหรับกลุ่มโรคแทรกซ้อนอันเนื่องมาจากโรคเบาหวาน

| ลำดับที่ | โรคแทรกซ้อน | รหัส ICD10 |
|----------|-----------------------------|---|
| 1 | Cardiovascular | 1.1 + 1.2. |
| 1.1. | Acute Myocardial Infarction | I20,I21,I22,I25 |
| 1.2. | Congestive Heart Failure | I099, I110, I130, I132, I255, I420, I425-I429, I43, I50, P290 |
| 2 | Cerebrovascular Disease | G45, G46, H340, I60-I69 |
| 3 | Eye | E113, H33, H350, H3535, H356, H358, H359, H431, H54 |
| 4 | Amputation | Z89 |
| 5 | Kidney | N18, N250, Z49, Z992, E112, I12 |

จากการเตรียมข้อมูลตัวแปรสถานะอุบัติการณ์โรคแทรกซ้อน 5 กลุ่ม ไม่พบผู้ป่วยที่มีภาวะแทรกซ้อนทางเส้นประสาทและเส้นเลือดส่วนปลายอันจำเป็นต้องเกิดการตัดรยางค์ (Amputation) ดังนั้น ในการสร้างตัวแบบทำนายความเสี่ยงของโรคแทรกซ้อนอันเนื่องมาจากโรคเบาหวานนั้น จะพิจารณาโรคแทรกซ้อนเพียง 4 กลุ่ม ได้แก่ 1) ภาวะเส้นเลือดหัวใจตีบหรือกล้ามเนื้อหัวใจขาดเลือด (Cardiovascular Disease) 2) ภาวะเส้นเลือดสมองแตกหรือตีบ (Cerebrovascular Disease) 3) ภาวะแทรกซ้อนทางตาอันเนื่องมาจากโรคเบาหวานและความดันโลหิตสูง และ 4) ภาวะไตเสื่อมเรื้อรัง ซึ่งหากรวมกับการเสียชีวิต จะได้ตัวแบบทำนายความเสี่ยงทั้งหมด 5 ตัวแบบ

การวิเคราะห์ข้อมูลเชิงสำรวจ

ในการวิเคราะห์ข้อมูลเชิงสำรวจ รูปที่ 48 แสดงเมทริกซ์เปลี่ยนสถานะ (Transition Matrix) ซึ่งแสดงค่าลอการิทึมของสัดส่วนการเปลี่ยนสถานะของอุบัติการณ์โรคในผู้ป่วยจากปีหนึ่งไปยังปีต่อไป เนื่องจากจากการสำรวจเบื้องต้นพบว่าผู้ป่วยที่เป็นโรคเบาหวานจำนวนหนึ่งจะมีภาวะของโรคความดันโลหิตสูงร่วมด้วย ดังนั้น

ในการวิเคราะห์ นอกจากจะพิจารณาสถานะของโรคเบาหวานเพียงอย่างเดียวแล้ว (DB) ยังพิจารณาสถานะของโรคเบาหวานร่วมกับภาวะโรคความดันโลหิตสูงอีกด้วย (DB-HT) สำหรับกลุ่มโรคแทรกซ้อนอื่น ๆ 4 กลุ่ม ได้แก่

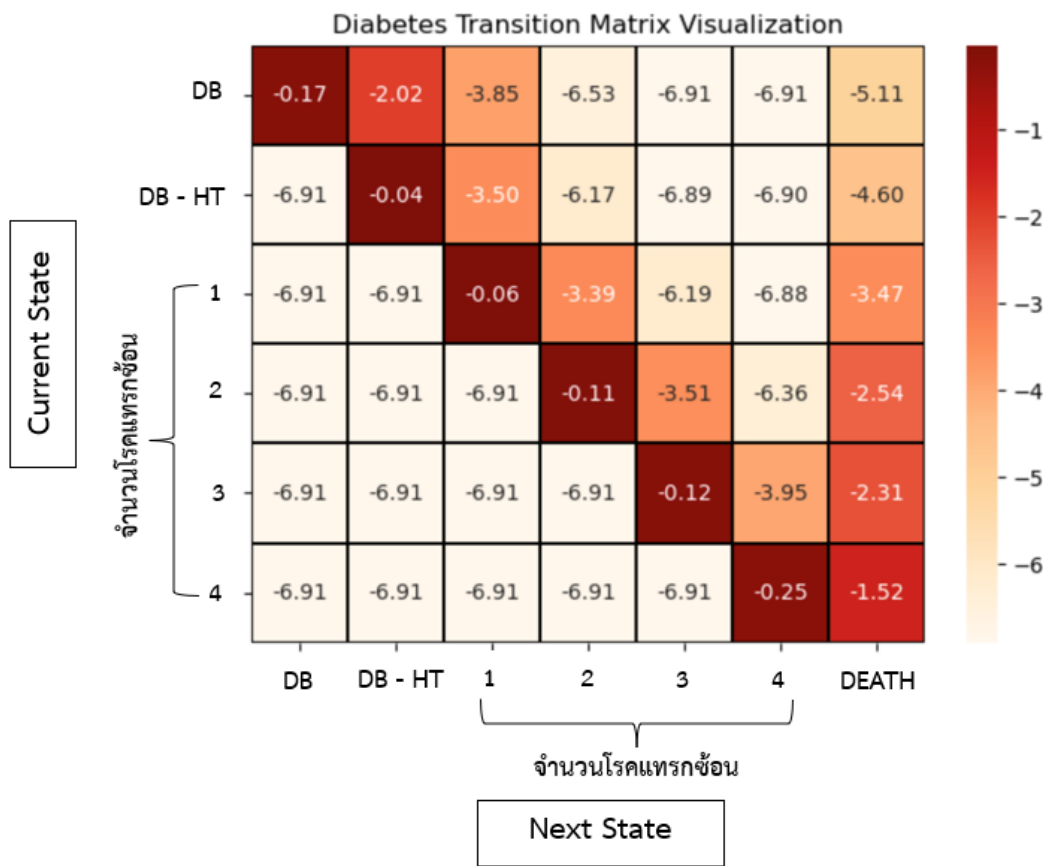
1) ภาวะเส้นเลือดหัวใจตีบหรือกล้ามเนื้อหัวใจขาดเลือด (Cardiovascular Disease) 2) ภาวะเส้นเลือดสมองแตกหรือตีบ (Cerebrovascular Disease) 3) ภาวะแทรกซ้อนทางตาอันเนื่องมาจากโรคเบาหวานและความดันโลหิตสูง และ 4) ภาวะไตเสื่อมเรื้อรัง จะพิจารณาโดยนับจำนวนโรคแทรกซ้อนที่เกิดขึ้นแทน

จากรูปที่ 48 โดยภาพรวม พบว่าส่วนใหญ่ผู้ป่วยจะไม่เปลี่ยนสถานะหรืออยู่สถานะเดิมในปีถัดไป อย่างไรก็ตาม ยังพบว่าผู้ป่วยที่เป็นโรคเบาหวานเพียงอย่างเดียวมีสัดส่วนถึง $e^{-2.02} \times 100\% = 13.27\%$ ที่ในปีถัดไปจะเป็นโรคความดันโลหิตสูงร่วมด้วย สำหรับโอกาสเกิดโรคแทรกซ้อนในปีถัดไป โดยภาพรวม พบว่าผู้ป่วยเบาหวานที่ในปีถัดไปมีโรคแทรกซ้อนเพิ่มขึ้นส่วนใหญ่จะเพิ่มปีละ 1 โรค นอกจากนี้ ยังเห็นได้ชัดเจนว่าผู้ป่วยเบาหวานยังมีจำนวนโรคแทรกซ้อนมากขึ้น โอกาสในการเสียชีวิตก็จะเพิ่มขึ้นตามไปด้วย

หากจำแนกประเภททุก ๆ ความเป็นไปได้ของกลุ่มโรคแทรกซ้อนที่เป็นไปได้ทั้งหมดของผู้ป่วยทั้ง 3 กลุ่ม ได้แก่

- 1) ผู้ป่วยที่เป็นโรคเบาหวานเพียงอย่างเดียว
- 2) ผู้ป่วยที่เป็นโรคความดันโลหิตสูงเป็นอย่างเดียว
- 3) ผู้ป่วยที่เป็นโรคเบาหวานร่วมกับความดันโลหิตสูง

สัดส่วน (ในมาตราส่วนเชิงลอการิทึม) ของทุก ๆ สถานะของโรคจากปีหนึ่งไปยังอีกปีหนึ่ง สามารถแสดงผ่านเมทริกซ์เปลี่ยนสถานะดังรูปที่ 49

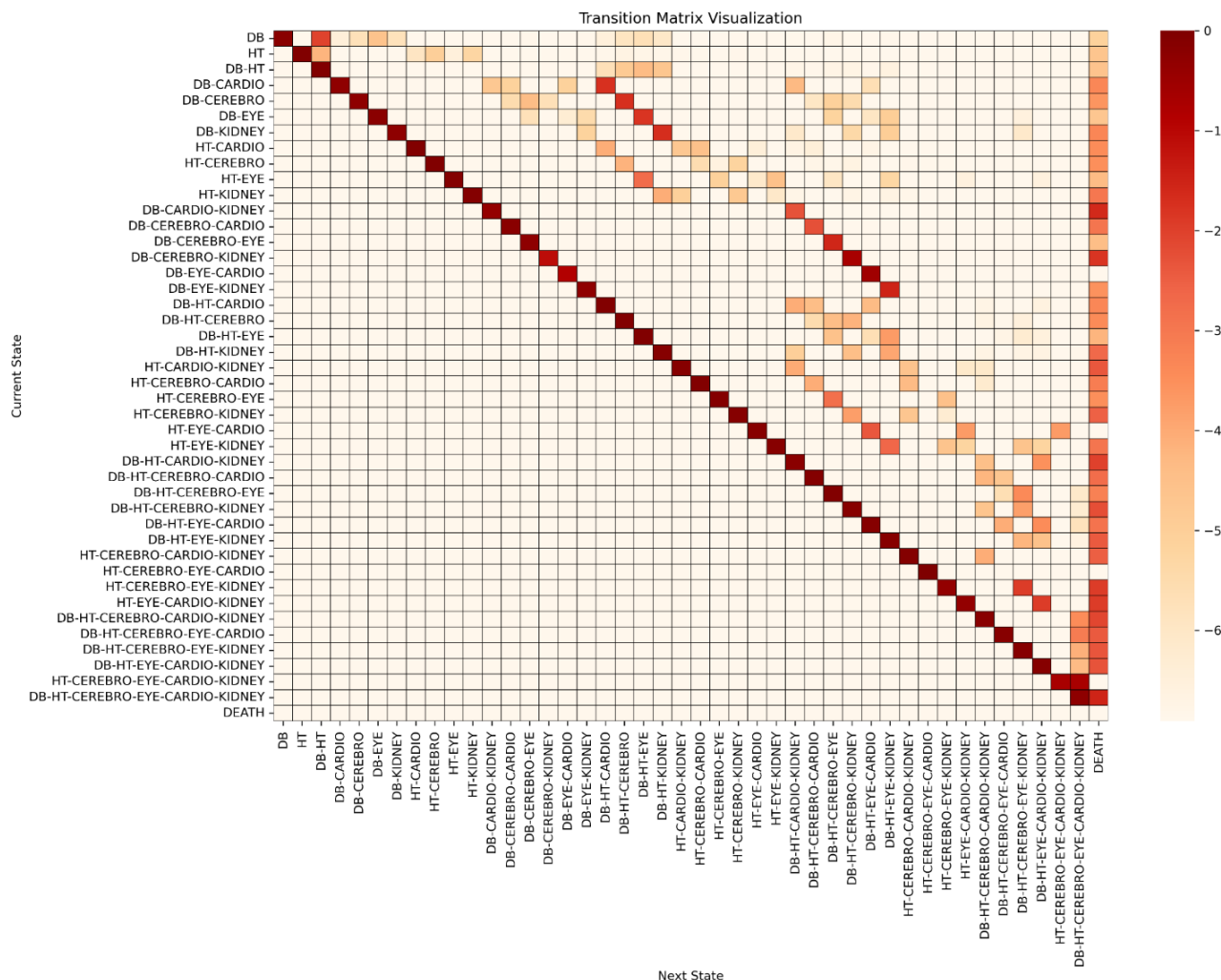


รูปที่ 48 เมทริกซ์เปลี่ยนสถานะรายปีของผู้ป่วยที่เป็นโรคเบาหวานในมาตราส่วนเชิงลอการิทึม

จากรูปที่ 49 พบว่าผู้ป่วยส่วนใหญ่ไม่เปลี่ยนแปลงสถานะหรืออยู่ในสถานะเดิมในปีถัดไป นอกจากนี้แล้วยังพบว่าผู้ป่วยที่ไม่ว่าจะเป็นโรคเบาหวานเพียงอย่างเดียว หรือเป็นโรคเบาหวานร่วมกับโรคแทรกซ้อนอื่น ๆ จะมีโอกาสมีภาวะความดันโลหิตสูงร่วมด้วยในปีถัดไปมากกว่าโรคอื่น ๆ และในทางกลับกัน ในผู้ป่วยที่เป็นโรคความดันโลหิตสูงเพียงอย่างเดียว หรือเป็นโรคความดันโลหิตสูงร่วมกับโรคแทรกซ้อนอื่น ๆ จะมีโอกาสเป็นโรคเบาหวานร่วมด้วยในปีถัดไปมากกว่าโรคอื่น ๆ แม้ว่าโอกาสการเป็นโรคเบาหวานหลังจากเป็นโรคความดันโลหิตสูง จะน้อยกว่าโอกาสการเป็นโรคความดันโลหิตสูงหลังจากเป็นโรคเบาหวานก็ตาม

สำหรับโอกาสในการเสียชีวิต พบว่า โอกาสการเสียชีวิตของผู้ป่วยทั้ง 3 กลุ่ม ได้แก่ 1) ผู้ป่วยที่เป็นโรคเบาหวานเพียงอย่างเดียว 2) ผู้ป่วยที่เป็นโรคความดันโลหิตสูงเป็นอย่างเดียว และ 3) ผู้ป่วยที่เป็นโรคเบาหวานร่วมกับความดันโลหิตสูงนั้นไม่ได้สูงนัก นอกจากนี้แล้ว ข้อมูลชุดนี้ยังแสดงให้เห็นว่าผู้ป่วยเบาหวานและความดันโลหิตสูงที่มีจำนวนโรคแทรกซ้อน 3 - 4 กลุ่มโรคจะมีความเสี่ยงในการเสียชีวิตค่อนข้างสูง อย่างไรก็ตาม ยังพบประเด็นที่น่าสนใจในผู้ป่วยเบาหวานที่มีโรคแทรกซ้อน 2 กลุ่มโรค ซึ่งพบว่าผู้ป่วยทั้ง 2 กลุ่มนี้มีความเสี่ยงสูงในการเสียชีวิตในปีถัดไปได้แก่

- 1) ภาวะเส้นเลือดหัวใจตีบหรือกล้ามเนื้อหัวใจขาดเลือดร่วมกับโรคไตเสื่อมเรื้อรัง (DB-CARDIO-KIDNEY)
- 2) ภาวะเส้นเลือดสมองแตกหรือตีบร่วมกับโรคไตเสื่อมเรื้อรัง (DB-CEREBRO-KIDNEY)



รูปที่ 49 เมทริกซ์เปลี่ยนสถานะรายปีของผู้ป่วยที่เป็นโรคเบาหวานและความดันโลหิตสูงในมาตราส่วนเชิงลอการิทึม

ตัวแปรทำนายที่ใช้ในตัวแบบทำนายความเสี่ยงภาวะแทรกซ้อนและการเสียชีวิตในผู้ป่วยที่เป็นโรคเบาหวาน (Cohort 1.2)

ในการสร้างตัวแบบทำนายความเสี่ยงภาวะแทรกซ้อน ทีมผู้วิจัยจะสร้างตัวแบบทำนายทั้งหมด 5 ตัวแบบ จำแนกตามภาวะแทรกซ้อนและการเสียชีวิต ดังนี้

1. ตัวแบบทำนายความเสี่ยงภาวะเส้นเลือดหัวใจตีบหรือกล้ามเนื้อหัวใจขาดเลือดในผู้ป่วยโรคเบาหวาน
2. ตัวแบบทำนายความเสี่ยงภาวะเส้นเลือดสมองแตกหรือตีบในผู้ป่วยโรคเบาหวาน
3. ตัวแบบทำนายความเสี่ยงภาวะแทรกซ้อนทางตาในผู้ป่วยโรคเบาหวาน
4. ตัวแบบทำนายความเสี่ยงภาวะไตเสื่อมเรื้อรังในผู้ป่วยโรคเบาหวาน

5. ตัวแบบทำนายความเสี่ยงการเสียชีวิตในผู้ป่วยโรคเบาหวาน
สำหรับตัวแปรที่จะใช้ในการทำนาย จากการทบทวนวรรณกรรมสามารถสรุปตัวแปรอิสระในแต่ละตัวแบบได้
ดังตารางที่ 29

ตารางที่ 29 ตัวแปรที่นำมาพิจารณาในการสร้างตัวแบบทำนายความเสี่ยงภาวะแทรกซ้อนและการเสียชีวิตในผู้ป่วยที่เป็นโรคเบาหวาน
และความดันโลหิตสูง

ข้อมูลรายการตัวแปรสรุปมาจากงานวิจัย Meta-Analysis ขนาดใหญ่⁸⁸ ประกอบด้วย งานวิจัยจำนวน 78 ฉบับและแบบจำลองทำนายความเสี่ยง 260 แบบจำลอง ครอบคลุมภาวะแทรกซ้อน ได้แก่ หลอดเลือดหัวใจและสมอง (CVD) 42 งานวิจัย ไตวายเรื้อรัง (CKD) 14 งานวิจัย ตา (EYE) 10 งานวิจัย การเสียชีวิต (DEATH) 16 งานวิจัย และ ภาวะแทรกซ้อนอื่น ๆ อีก 16 งานวิจัย โดยตารางจะสรุปจำนวนงานวิจัยที่มีการใช้ตัวแปรตามที่แสดงในตารางแยกประเภทภาวะแทรกซ้อน

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ในตัวแบบ | ภาวะแทรกซ้อน | | | | | Reference |
|-----------------------------|---|--------------------------|--------------|------|----|----|-----------|--|
| | | | หัวใจ | สมอง | ตา | ไต | เสียชีวิต | |
| Demographic | | | | | | | | |
| Age | อายุ (ปี) | age | ✓ | ✓ | ✓ | ✓ | ✓ | CVD: 24 papers EYE: 4 papers CKD: 9 papers DEATH: 14 papers |
| Gender | เพศ - ชาย - หญิง | t.sex | ✓ | ✓ | ✓ | ✓ | ✓ | CVD: 28 papers EYE: 2 papers CKD: 5 papers DEATH: 10 papers |
| Health Behavior | | | | | | | | |
| Smoking | พฤติกรรมการสูบบุหรี่ - สูบ - ไม่สูบ | t.smoke | ✓ | ✓ | ✓ | ✓ | ✓ | CVD: 23 papers EYE: 1 paper CKD: 8 papers DEATH: 10 papers |
| Physical measurement | | | | | | | | |
| Height | ส่วนสูง (เซนติเมตร) | t.height | ✓ | ✓ | | | | CVD: 2 papers |

ตารางที่ 29 ตัวแปรที่นำมาพิจารณาในการสร้างตัวแบบทำนายความเสี่ยงภาวะแทรกซ้อนและการเสียชีวิตในผู้ป่วยที่เป็นโรคเบาหวาน
และความดันโลหิตสูง (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ในตัวแบบ | ภาวะแทรกซ้อน | | | | | Reference |
|---------------------------|--|--------------------------|--------------|------|----|----|-----------|--|
| | | | หัวใจ | สมอง | ตา | ไต | เสียชีวิต | |
| BMI | ดัชนีมวลกาย | BMI | ✓ | ✓ | ✓ | ✓ | ✓ | CVD: 15 papers EYE: 3 papers CKD: 8 papers DEATH: 11 papers |
| WAIST_CM | เส้นรอบเอว (เซนติเมตร) | WAIST_CM | | | | ✓ | | CKD: 1 paper |
| SBP | Systolic Blood Pressure (mm Hg) | SBP | ✓ | ✓ | ✓ | ✓ | ✓ | CVD: 23 papers EYE: 5 papers CKD: 7 papers DEATH: 7 papers |
| DBP | Diastolic Blood Pressure (mm Hg) | DBP | ✓ | ✓ | | ✓ | ✓ | CVD: 6 papers CKD: 1 paper DEATH: 3 papers |
| PP | Pulse Pressure คำนวณจาก SBP – DBP | PP | ✓ | ✓ | | | | CVD: 1 paper |
| Underlying Disease | | | | | | | | |
| HT | สถานะโรคความดันโลหิตสูง - เป็น - ไม่เป็น | HT | ✓ | ✓ | ✓ | ✓ | ✓ | CVD: 4 papers EYE: 2 papers CKD: 3 papers DEATH: 2 papers |

ตารางที่ 29 ตัวแปรที่นำมาพิจารณาในการสร้างตัวแบบทำนายความเสี่ยงภาวะแทรกซ้อนและการเสียชีวิตในผู้ป่วยที่เป็นโรคเบาหวาน
และความดันโลหิตสูง (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ในตัวแบบ | ภาวะแทรกซ้อน | | | | | Reference |
|--|---|--------------------------|--------------|------|----|----|-----------|---|
| | | | หัวใจ | สมอง | ตา | ไต | เสียชีวิต | |
| HT Duration | ระยะเวลาที่เป็นความดันโลหิตสูง (ปี) | ht_duration | ✓ | ✓ | ✓ | ✓ | ✓ | Try test variables |
| HT Age of Diagnosis | อายุที่ได้รับการวินิจฉัยว่าเป็นโรคความดันโลหิตสูง (ปี) | ht_age_of_diag | ✓ | ✓ | | | ✓ | Try test variables |
| DLP | สถานะภาวะไขมันผิดปกติ - เป็น - ไม่เป็น | flag_dlp | ✓ | ✓ | | ✓ | ✓ | CVD: 2 papers CKD: 2 papers DEATH: 1 paper |
| DM Duration | ระยะเวลาที่เป็นโรคเบาหวาน (ปี) | db_duration | ✓ | ✓ | ✓ | ✓ | ✓ | CVD: 19 papers EYE: 2 papers CKD: 5 papers DEATH: 5 papers |
| DM Age of Diagnosis | อายุที่ได้รับการวินิจฉัยว่าเป็นโรคเบาหวาน (ปี) | db_age_of_diag | ✓ | ✓ | | | | CVD: 6 papers |
| Heart Disease (Cardiovascular Disease) | สถานะการเป็นภาวะเส้นเลือดหัวใจตีบหรือกล้ามเนื้อหัวใจขาดเลือด - เป็น - ไม่เป็น | CARDIO | | ✓ | ✓ | ✓ | ✓ | CVD: 11 papers EYE: 1 paper CKD: 3 papers DEATH: 1 paper |

ตารางที่ 29 ตัวแปรที่นำมาพิจารณาในการสร้างตัวแบบทำนายความเสี่ยงภาวะแทรกซ้อนและการเสียชีวิตในผู้ป่วยที่เป็นโรคเบาหวาน
และความดันโลหิตสูง (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ในตัวแบบ | ภาวะแทรกซ้อน | | | | | Reference |
|---------------|---|--------------------------|--------------|------|----|----|-----------|--|
| | | | หัวใจ | สมอง | ตา | ไต | เสียชีวิต | |
| Heart Disease | สถานะการเป็นโรคหัวใจ ซึ่งรวมถึงการเป็นโรคหัวใจเด่นผิดปกติ - เป็น - ไม่เป็น | flag_HEART | | ✓ | ✓ | ✓ | ✓ | CVD: 18 papers EYE: 2 papers CKD: 8 papers DEATH: 15 papers |
| CKD | สถานะการเป็นโรคไตเรื้อรัง โดยดูจากรหัสโรคที่วินิจฉัย - เป็น - ไม่เป็น | flag_CKD | ✓ | ✓ | | | ✓ | CVD: 4 papers DEATH: papers |
| KIDNEY | สถานะการเป็นโรคไตเรื้อรัง ในระยะสุดท้าย (พิจารณาจาก CKD และค่า eGFR) - เป็น - ไม่เป็น | KIDNEY | ✓ | ✓ | | | ✓ | CVD: 4 papers DEATH: 2 papers |
| Stroke | สถานะการเป็นโรคหลอดเลือดสมอง - เป็น - ไม่เป็น | CEREBRO | ✓ | | | ✓ | ✓ | CVD: 4 papers CKD: 2 papers DEATH: 4 papers |

ตารางที่ 29 ตัวแปรที่นำมาพิจารณาในการสร้างตัวแบบทำนายความเสี่ยงภาวะแทรกซ้อนและการเสียชีวิตในผู้ป่วยที่เป็นโรคเบาหวาน
และความดันโลหิตสูง (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ในตัวแบบ | ภาวะแทรกซ้อน | | | | | Reference |
|-------------|--|--------------------------|--------------|------|----|----|-----------|---|
| | | | หัวใจ | สมอง | ตา | ไต | เสียชีวิต | |
| flag_stroke | สถานะการเป็นโรคหลอดเลือดสมอง พิจารณาจากรหัสโรคที่วินิจฉัย - เป็น - ไม่เป็น | flag_stroke | ✓ | | | ✓ | ✓ | CVD: 4 papers CKD: 2 papers DEATH: 4 papers |
| Eye | สถานะการเป็นโรคตาอันเนื่องมาจากเบาหวาน - เป็น - ไม่เป็น | EYE | ✓ | ✓ | | ✓ | ✓ | CVD: 2 papers CKD: 4 papers DEATH: 1 paper |
| flag_eye | สถานะการเป็นโรคตา พิจารณาจากรหัสโรคที่วินิจฉัย - เป็น - ไม่เป็น | flag_eye | ✓ | ✓ | | ✓ | ✓ | CVD: 2 papers CKD: 4 papers DEATH: 1 paper |
| COPD | สถานะการเป็นโรคถุงลมโป่งพอง - เป็น - ไม่เป็น | flag_copd | ✓ | ✓ | | ✓ | ✓ | CVD: 1 paper CKD: 1 paper DEATH: 2 papers |

ตารางที่ 29 ตัวแปรที่นำมาพิจารณาในการสร้างตัวแบบทำนายความเสี่ยงภาวะแทรกซ้อนและการเสียชีวิตในผู้ป่วยที่เป็นโรคเบาหวาน
และความดันโลหิตสูง (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ในตัวแบบ | ภาวะแทรกซ้อน | | | | | Reference |
|----------------------|--|--------------------------|--------------|------|----|----|-----------|--|
| | | | หัวใจ | สมอง | ตา | ไต | เสียชีวิต | |
| Lab | | | | | | | | |
| TC | ระดับคอเลสเตอรอล (mg/dL) | tcl_labresult | ✓ | ✓ | ✓ | ✓ | ✓ | CVD: 5 papers EYE: 2 papers CKD: 2 papers DEATH: 2 papers |
| is_na_tcl_labresult* | ทราบค่า TC หรือไม่ - ทราบ - ไม่ทราบ | is_na_tcl_labresult | ✓ | ✓ | ✓ | ✓ | ✓ | |
| HDL | ระดับไขมันดี (mg/dL) | hdl_labresult | ✓ | ✓ | ✓ | ✓ | ✓ | CVD: 7 papers EYE: 2 papers CKD: 4 papers DEATH: 4 papers |
| is_na_hdl_labresult* | ทราบค่า HDL หรือไม่ - ทราบ - ไม่ทราบ | is_na_hdl_labresult | ✓ | ✓ | ✓ | ✓ | ✓ | |
| LDL | ระดับไขมันไม่ดี (mg/dL) | ldl_labresult | ✓ | ✓ | ✓ | ✓ | ✓ | CVD: 7 papers EYE: 2 papers CKD: 3 papers DEATH: 4 papers |

ตารางที่ 29 ตัวแปรที่นำมาพิจารณาในการสร้างตัวแบบทำนายความเสี่ยงภาวะแทรกซ้อนและการเสียชีวิตในผู้ป่วยที่เป็นโรคเบาหวาน
และความดันโลหิตสูง (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ในตัวแบบ | ภาวะแทรกซ้อน | | | | | Reference |
|----------------------|--|--------------------------|--------------|------|----|----|-----------|--|
| | | | หัวใจ | สมอง | ตา | ไต | เสียชีวิต | |
| is_na_ldl_labresult* | ทราบค่า HDL หรือไม่ - ทราบ - ไม่ทราบ | is_na_ldl_labresult | ✓ | ✓ | ✓ | ✓ | ✓ | |
| TG | ระดับไขมันไตรกลีเซอไรด์ (mg/dL) | tg_labresult | ✓ | ✓ | | ✓ | ✓ | CVD: 4 papers CKD: 1 paper DEATH: 3 papers |
| is_na_tg_labresult* | ทราบค่า TG หรือไม่ - ทราบ - ไม่ทราบ | is_na_tg_labresult | ✓ | ✓ | | ✓ | ✓ | |
| non-HDL | TC – HDL | non_hdl | ✓ | ✓ | | | ✓ | CVD: 4 papers DEATH: 1 paper |
| is_na_non_hdl* | ทราบค่า non-HDL หรือไม่ - ทราบ - ไม่ทราบ | is_na_non_hdl | ✓ | ✓ | | | ✓ | |
| total Cho:HDL | TC/HDL | total_cho_to_hdl | ✓ | ✓ | | ✓ | ✓ | CVD: 10 papers CKD: 2 papers DEATH: 2 papers |

ตารางที่ 29 ตัวแปรที่นำมาพิจารณาในการสร้างตัวแบบทำนายความเสี่ยงภาวะแทรกซ้อนและการเสียชีวิตในผู้ป่วยที่เป็นโรคเบาหวาน
และความดันโลหิตสูง (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ในตัวแบบ | ภาวะแทรกซ้อน | | | | | Reference |
|-------------------------|---|--------------------------|--------------|------|----|----|-----------|--|
| | | | หัวใจ | สมอง | ตา | ไต | เสียชีวิต | |
| is_na_total_cho_to_hdl* | ทราบค่า total Cho:HDL หรือไม่ - ทราบ - ไม่ทราบ | is_na_total_cho_to_hdl | ✓ | ✓ | | ✓ | ✓ | |
| non HDL:HDL | (TC – HDL)/HDL | no_hdl_to_hdl | ✓ | ✓ | | | | CVD: 1 paper |
| is_na_no_hdl_to_hdl* | ทราบค่า non HDL:HDL หรือไม่ - ทราบ - ไม่ทราบ | is_na_no_hdl_to_hdl | ✓ | ✓ | | | | |
| HBA1c | ค่าน้ำตาลเฉลี่ยสะสมใน เลือด (mmol/mol) | HBA1c_labresult | ✓ | ✓ | ✓ | ✓ | ✓ | CVD: 23 papers EYE: 5 papers CKD: 10 papers DEATH: 9 papers |
| is_na_HBA1C_labresult* | ทราบค่า HBA1c หรือไม่ - ทราบ - ไม่ทราบ | is_na_HBA1C_labresult | ✓ | ✓ | ✓ | ✓ | ✓ | |
| FBS | Fasting Blood Sugar (mg/dL) | FBS_labresult | ✓ | ✓ | | ✓ | ✓ | CVD: 3 papers CKD: 1 paper DEATH: 2 papers |

ตารางที่ 29 ตัวแปรที่นำมาพิจารณาในการสร้างตัวแบบทำนายความเสี่ยงภาวะแทรกซ้อนและการเสียชีวิตในผู้ป่วยที่เป็นโรคเบาหวาน
และความดันโลหิตสูง (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ในตัวแบบ | ภาวะแทรกซ้อน | | | | | Reference |
|-----------------------|---|-----------------------------|--------------|------|----|----|-----------|---|
| | | | หัวใจ | สมอง | ตา | ไต | เสียชีวิต | |
| is_na_FBS_labresult* | ทราบค่า FBS หรือไม่ - ทราบ - ไม่ทราบ | is_na_FBS_labresult | ✓ | ✓ | | ✓ | ✓ | |
| Macro-albuminuria | สถานะความเสี่ยงการเป็นโรคไตโดยวัดจาก Macro-albuminuria - เป็น - ไม่เป็น | macro_albuminuria_labresult | ✓ | ✓ | | ✓ | ✓ | CVD: 3 papers CKD: 1 paper DEATH: 1 paper |
| Micro-albuminuria | สถานะความเสี่ยงการเป็นโรคไตโดยวัดจาก Micro-albuminuria - เป็น - ไม่เป็น | micro_albuminuria_labresult | ✓ | ✓ | | | ✓ | CVD: 4 papers DEATH: 1 paper |
| eGFR | Estimated Glomerular Filtration Rate (mL/min/1.73m ²) | egfr_labresult | ✓ | ✓ | ✓ | ✓ | ✓ | CVD: 6 papers EYE: 1 paper CKD: 4 papers DEATH: 5 papers |
| is_na_egfr_labresult* | ทราบค่า eGFR หรือไม่ - ทราบ - ไม่ทราบ | is_na_egfr_labresult | ✓ | ✓ | ✓ | ✓ | ✓ | |

ตารางที่ 29 ตัวแปรที่นำมาพิจารณาในการสร้างตัวแบบทำนายความเสี่ยงภาวะแทรกซ้อนและการเสียชีวิตในผู้ป่วยที่เป็นโรคเบาหวาน
และความดันโลหิตสูง (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ในตัวแบบ | ภาวะแทรกซ้อน | | | | | Reference |
|-----------------------------|---|----------------------------|--------------|------|----|----|-----------|--|
| | | | หัวใจ | สมอง | ตา | ไต | เสียชีวิต | |
| Creatinine | ระดับ Creatinine (mg/dL) | creatinine_labresult | ✓ | ✓ | ✓ | ✓ | ✓ | CVD: 3 papers EYE: 1 paper CKD: 4 papers DEATH: 3 papers |
| is_na_creatinine_labresult* | ทราบค่า Creatinine หรือไม่ - ทราบ - ไม่ทราบ | is_na_creatinine_labresult | ✓ | ✓ | ✓ | ✓ | ✓ | |
| UPCR | ระดับโปรตีนในปัสสาวะ (mg/g) | upcr_labresult | ✓ | ✓ | ✓ | ✓ | ✓ | CVD: 8 papers EYE: 3 papers CKD: 6 papers DEATH: 5 papers |
| is_na_upcr_labresult* | ทราบค่า UPCR หรือไม่ - ทราบ - ไม่ทราบ | is_na_upcr_labresult | ✓ | ✓ | ✓ | ✓ | ✓ | |
| Hemoglobin | ระดับฮีโมโกลบินในเลือด (g/dL) | hemoglobin_labresult | ✓ | ✓ | | ✓ | ✓ | CVD: 2 papers CKD: 1 paper DEATH: 3 papers |

ตารางที่ 29 ตัวแปรที่นำมาพิจารณาในการสร้างตัวแบบทำนายความเสี่ยงภาวะแทรกซ้อนและการเสียชีวิตในผู้ป่วยที่เป็นโรคเบาหวาน
และความดันโลหิตสูง (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ในตัวแบบ | ภาวะแทรกซ้อน | | | | | Reference |
|-----------------------------|---|----------------------------|--------------|------|----|----|-----------|--|
| | | | หัวใจ | สมอง | ตา | ไต | เสียชีวิต | |
| is_na_hemoglobin_labresult* | ทราบค่า Hemoglobin หรือไม่ - ทราบ - ไม่ทราบ | is_na_hemoglobin_labresult | ✓ | ✓ | | ✓ | ✓ | |
| Albumin | ระดับ Albumin (g/dL) | albumin_labresult | ✓ | ✓ | | | | CVD: 2 papers |
| is_na_albumin_labresult* | ทราบค่า Albumin หรือไม่ - ทราบ - ไม่ทราบ | is_na_albumin_labresult | ✓ | ✓ | | | | |
| Medication | | | | | | | | |
| Lipid-lowering med | การใช้อยาลดไขมัน - ใช้ - ไม่ใช้ | flag_lipid_lowering | ✓ | ✓ | ✓ | ✓ | ✓ | CVD: 5 papers EYE: 1 paper CKD: 2 papers DEATH: 4 papers |
| Non-insulin med | การใชยาเบาหวานที่ไม่ต้องพึ่งอินซูลิน - ใช้ - ไม่ใช้ | flag_non_insulin_med | ✓ | ✓ | ✓ | ✓ | ✓ | CVD: 3 papers EYE: 2 papers CKD: 3 papers DEATH: 3 papers |

ตารางที่ 29 ตัวแปรที่นำมาพิจารณาในการสร้างตัวแบบทำนายความเสี่ยงภาวะแทรกซ้อนและการเสียชีวิตในผู้ป่วยที่เป็นโรคเบาหวาน
และความดันโลหิตสูง (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ในตัวแบบ | ภาวะแทรกซ้อน | | | | | Reference |
|-------------|---|--------------------------|--------------|------|----|----|-----------|---|
| | | | หัวใจ | สมอง | ตา | ไต | เสียชีวิต | |
| Insulin med | การใช้ยาเบาหวานที่ต้องพึ่งอินซูลิน - ใช้ - ไม่ใช้ | flag_insulin_med | ✓ | ✓ | | ✓ | ✓ | CVD: 6 papers CKD: 2 papers DEATH: 7 papers |
| Aspirin | การใช้ยาแอสไพริน - ใช้ - ไม่ใช้ | flag_aspirin_med | | ✓ | | ✓ | ✓ | CVD: 2 papers CKD: 1 paper DEATH: 1 paper |

หมายเหตุ* หมายถึงตัวแปรเพิ่มเติมที่ได้จากการเตรียมข้อมูล

ผลลัพธ์จากตัวแบบทำนายความเสี่ยงการเกิดภาวะแทรกซ้อนและการเสียชีวิตในผู้ป่วยที่เป็นโรคเบาหวาน (Cohort 1.2)

ตัวแบบทำนายความเสี่ยงภาวะเส้นเลือดหัวใจตีบหรือกล้ามเนื้อหัวใจขาดเลือดในผู้ป่วยโรคเบาหวาน

จากกลุ่มตัวอย่างของผู้ป่วยโรคเบาหวานจำนวน 90,242 คน ซึ่งประกอบไปด้วยข้อมูลจำนวน 541,910 แถว จะถูกสุ่มแบ่งออกเป็นชุดข้อมูลฝึกฝนและชุดข้อมูลทดสอบคิดเป็นร้อยละ 70 และ 30 โดยประมาณ ตามลำดับ โดยที่ชุดข้อมูลฝึกฝนและชุดข้อมูลทดสอบจะมีสัดส่วนของผู้ป่วยโรคเบาหวานที่พบภาวะเส้นเลือดหัวใจตีบหรือกล้ามเนื้อหัวใจขาดเลือดพอ ๆ กัน ซึ่งในที่นี้พบว่ามีความเท่ากับ 0.002917 หรือคือมีผู้ป่วยโรคเบาหวานคิดเป็นร้อยละ 0.2917 ที่มีภาวะเส้นเลือดหัวใจตีบหรือกล้ามเนื้อหัวใจขาดเลือดนั่นเอง จากการแบ่งข้อมูลจะได้ข้อมูลดังต่อไปนี้

- ชุดข้อมูลฝึกฝน ประกอบด้วยข้อมูลจากผู้เป็นโรคเบาหวานจำนวน 63,169 คน ซึ่งมีข้อมูลจำนวน 379,152 แถว
- ชุดข้อมูลทดสอบ ประกอบด้วยข้อมูลจากผู้เป็นโรคเบาหวานจำนวน 27,073 คน ซึ่งมีข้อมูลจำนวน 162,758 แถว

จากนั้นจึงสร้างตัวแบบจากข้อมูลชุดฝึกฝน ซึ่งจะพิจารณา 2 ตัวแบบ ได้แก่

1. การถดถอยโลจิสติกแบบ Elastic Net วิธีการนี้ขอเรียกว่า “EN”
2. วิธี XGBoost โดยวิธีการนี้ขอเรียกว่า “XGBoost”

หลังจากได้ตัวแบบทั้ง 2 แล้ว จึงนำมาทำนายข้อมูลในชุดทดสอบเพื่อวัดประสิทธิภาพการทำนายโดยพิจารณาจาก

1. AUC
2. Cumulative AUC

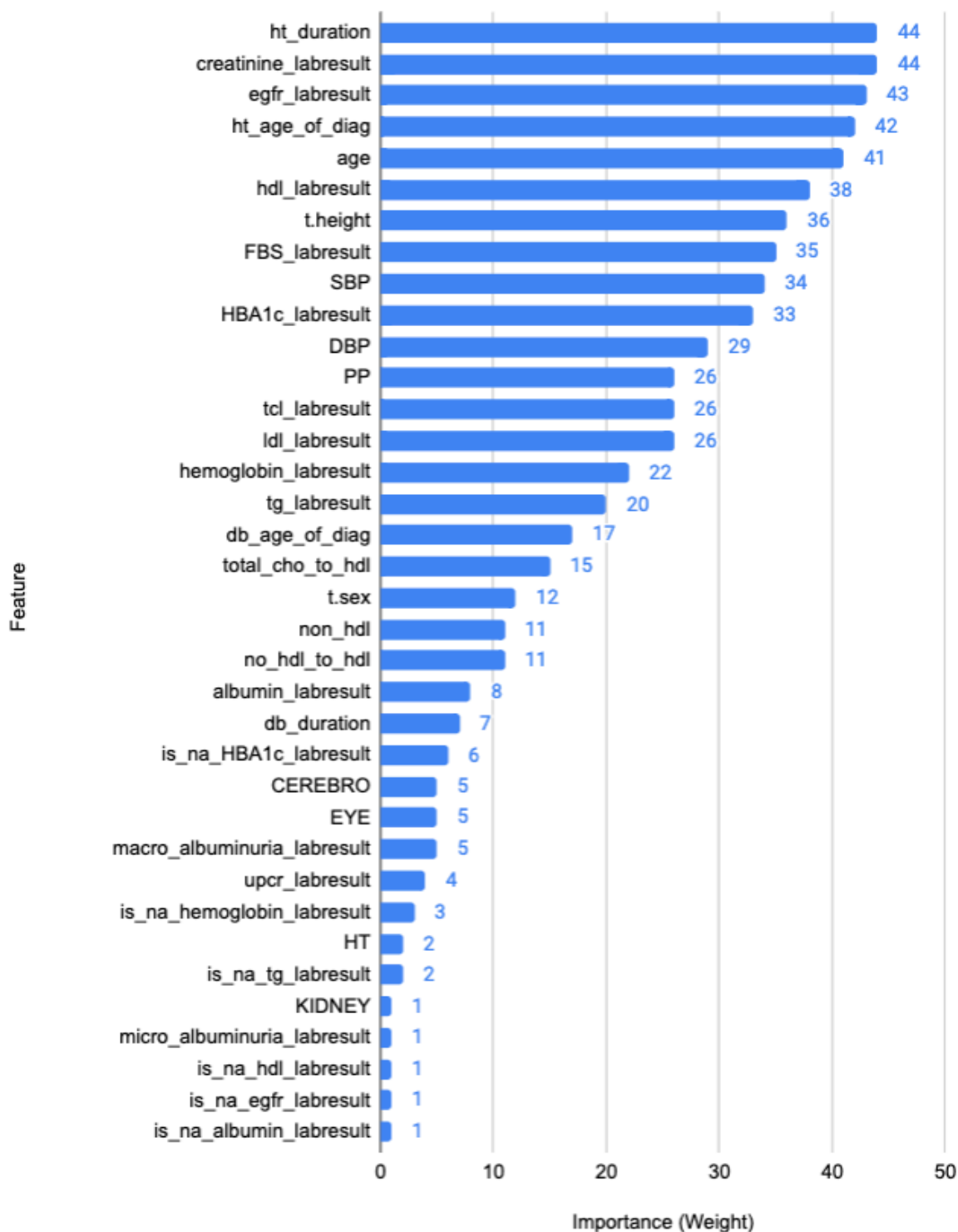
ตารางที่ 30 แสดงค่า AUC จาก 2 วิธี ซึ่งพบว่าวิธี XGBoost มีประสิทธิภาพในการทำนายความเสี่ยงการภาวะเส้นเลือดหัวใจตีบหรือกล้ามเนื้อหัวใจขาดเลือดในผู้ป่วยโรคเบาหวานได้ดีกว่าวิธีการถดถอยโลจิสติกแบบ Elastic Net แม้ว่าค่า AUC ที่ได้จากวิธี XGBoost จะดีกว่าวิธีการถดถอยโลจิสติกแบบ Elastic Net ไม่มาก แต่หากพิจารณาค่า Cumulative AUC ซึ่งวัดประสิทธิภาพการทำนายรายบุคคลผ่านตัววัดผล AUC ที่คำนวณด้วยความน่าจะเป็นสะสมของบุคคลตลอด Cohort (Cumulative Probability) กับสถานะอุบัติการณ์ปีสุดท้าย จะพบว่าวิธี XGBoost มีประสิทธิภาพเหนือกว่าวิธี Elastic Net ค่อนข้างมาก อย่างไรก็ตาม พบว่าวิธีทั้งสองอาจจะมีค่า AUC ที่ยังอยู่ในเกณฑ์ไม่สูง หรือมีค่าต่ำกว่า 0.8 ซึ่งในอนาคตควรจะต้องพิจารณาตัวแบบอื่นๆ รวมทั้งพิจารณาตัวแปรหรือใช้ขนาดตัวอย่างเพิ่มเติม เพื่อที่จะปรับปรุงตัวแบบในการทำนายให้มีประสิทธิภาพมากขึ้น

ตารางที่ 30 ค่า AUC เพื่อวัดประสิทธิภาพการทำนายความเสี่ยงภาวะเส้นเลือดหัวใจตีบหรือกล้ามเนื้อหัวใจขาด
เลือดในผู้ป่วยโรคเบาหวาน (Cohort 1.2) (วิธีที่ดีที่สุดจะแสดงเป็นตัวหนา)

| ตัวแบบ | AUC | Cumulative AUC |
|---------|---------------|-------------------|
| EN | 0.6524 | 0.5996 |
| XGBoost | 0.6713 | 0.6820 |

สำหรับการแปลผลตัวแบบโดยอ้างอิงจากตัวแบบ XGBoost ที่ให้ประสิทธิภาพการทำนายที่ดีที่สุดสามารถ
พิจารณาได้ว่าปัจจัยใดที่มีความสำคัญในการทำนายความเสี่ยงการเกิดโรคเบาหวานในประชากรทั่วไปดังแสดงใน
รูปที่ 50 ซึ่งจะพบว่าตัวแปรที่มีความสำคัญในการทำนายความเสี่ยง 10 อันดับแรก ได้แก่

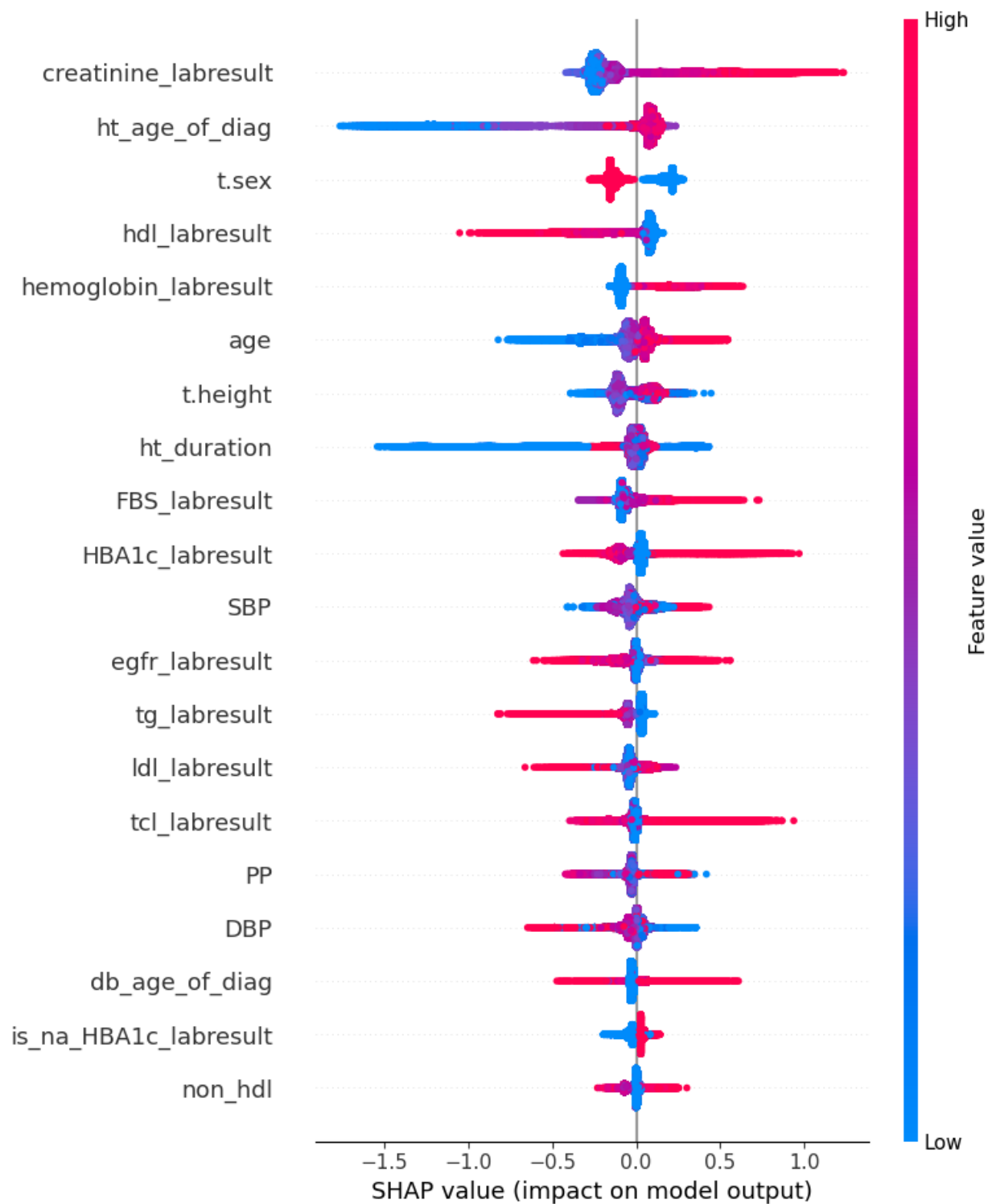
1. ระยะเวลาที่เป็นความดันโลหิตสูง (ปี) (ht_duration)
2. ระดับ Creatinine (mg/dL) (creatinine_labresult)
3. Estimated Glomerular Filtration Rate (mL/min/1.73m²) (egfr_labresult)
4. อายุที่ได้รับการวินิจฉัยว่าเป็นโรคความดันโลหิตสูง (ปี) (ht_age_of_diag)
5. อายุ (age)
6. ระดับไขมันดี (mg/dL) (hdl_labresult)
7. ความสูง (t.height)
8. Fasting Blood Sugar (mg/dL) (FBS_labresult)
9. Systolic Blood Pressure (mm Hg) (SBP)
10. ค่าน้ำตาลเฉลี่ยสะสมในเลือด (mmol/mol) (HBA1c_labresult)



รูปที่ 50 ความสำคัญของตัวแปรทำนายจากวิธี XGBoost ในการทำนายความเสี่ยงภาวะเส้นเลือดหัวใจตีบหรือกล้ามเนื้อหัวใจขาดเลือดในผู้ป่วยโรคเบาหวาน

นอกจากนี้ หากพิจารณาคุณค่า SHAP ของตัวแปร 20 ตัวแปรที่มีความสำคัญมากที่สุดในการทำนายความเสี่ยง ภาวะเส้นเลือดหัวใจตีบหรือกล้ามเนื้อหัวใจขาดเลือดในผู้ป่วยเบาหวาน ดังแสดงในรูปที่ 51 จะสามารถแปลผลใน ส่วนทิศทางการสัมพันธ์ของตัวแปรแต่ละตัว ซึ่งสามารถสรุปความสัมพันธ์ที่เด่นชัดได้ ดังนี้

- ผู้ป่วยเบาหวานที่มีระดับ Creatinine (mg/dL) (creatinine_labresult) สูง จะมีแนวโน้มที่จะมีความเสี่ยงในการมีภาวะเส้นเลือดหัวใจตีบหรือกล้ามเนื้อหัวใจขาดเลือดสูงขึ้น
- ผู้ป่วยเบาหวานที่ได้รับการวินิจฉัยว่าเป็นโรคความดันโลหิตสูงตอนช่วงอายุยังน้อย (ht_age_of_diag) จะมีแนวโน้มที่จะมีความเสี่ยงในการมีภาวะเส้นเลือดหัวใจตีบหรือกล้ามเนื้อหัวใจขาดเลือดต่ำลง
- ผู้ป่วยเบาหวานที่มีระดับไขมันดี (mg/dL) (hdl_labresult) สูง จะมีแนวโน้มที่จะมีความเสี่ยงในการมีภาวะเส้นเลือดหัวใจตีบหรือกล้ามเนื้อหัวใจขาดเลือดลดลง
- ผู้ป่วยเบาหวานที่มีระดับฮีโมโกลบินในเลือด (g/dL) (hemoglobin_labresult) สูง จะมีแนวโน้มที่จะมีความเสี่ยงในการมีภาวะเส้นเลือดหัวใจตีบหรือกล้ามเนื้อหัวใจขาดเลือดสูงขึ้น



รูปที่ 51 ค่า SHAP จากตัวแบบ XGBoost ในการทำนายความเสี่ยงภาวะเส้นเลือดหัวใจตีบหรือกล้ามเนื้อหัวใจขาดเลือดในผู้ป่วยโรคเบาหวาน

แม้ว่าวิธี XGBoost สามารถระบุความสำคัญของตัวแปรทำนายที่ใช้ในตัวแบบได้ แต่ไม่สามารถแปลผลตัวแบบในรูปของอัตราเสี่ยงได้ ในขณะที่วิธีการถดถอยโลจิสติกแบบ Elastic Net ให้ค่าสัมประสิทธิ์การถดถอยเช่นเดียวกับตัวแบบการถดถอยโลจิสติกแบบปกติ ดังนั้นจึงสามารถแปลผลตัวแบบในรูปของอัตราเสี่ยงได้ ตารางที่ 31 แสดงผลลัพธ์ที่ได้ของสัมประสิทธิ์การถดถอยของตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ในการทำนายความเสี่ยงภาวะเส้นเลือดหัวใจตีบหรือกล้ามเนื้อหัวใจขาดเลือดในผู้ป่วยโรคเบาหวาน

ตารางที่ 31 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ในการทำนายความเสี่ยงภาวะเส้นเลือดหัวใจตีบหรือกล้ามเนื้อหัวใจขาดเลือดในผู้ป่วยโรคเบาหวาน

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ในตัวแบบ | สัมประสิทธิ์การถดถอย | Odds Ratio |
|-----------------------------|---|--------------------------|----------------------|------------|
| Demographic | | | | |
| Age | อายุ (ปี) | age | -0.0060 | 0.9941 |
| Gender | เพศ - ชาย - หญิง | t.sex | -0.5004 | 0.6063 |
| Health Behavior | | | | |
| Smoking | พฤติกรรมการสูบบุหรี่ - สูบ - ไม่สูบ | t.smoke | -0.0172 | 0.9830 |
| Physical measurement | | | | |
| Height | ส่วนสูง (เซนติเมตร) | t.height | 0.0050 | 1.0050 |
| BMI | ดัชนีมวลกาย | BMI | 0.0144 | 1.0145 |
| SBP | Systolic Blood Pressure (mm Hg) | SBP | 0 | 1 |
| DBP | Diastolic Blood Pressure (mm Hg) | DBP | -0.0011 | 0.9989 |
| PP | Pulse Pressure คำนวณจาก SBP – DBP | PP | 0.0027 | 1.0027 |

ตารางที่ 31 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ใน
การทำนายความเสี่ยงภาวะเส้นเลือดหัวใจตีบหรือกล้ามเนื้อหัวใจขาดเลือดในผู้ป่วยโรคเบาหวาน (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ใน ตัวแบบ | สัมประสิทธิ์ การถดถอย | Odds Ratio |
|---------------------------|---|------------------------------|--------------------------|------------|
| Underlying Disease | | | | |
| HT | สถานะโรคความดันโลหิต สูง - เป็น - ไม่เป็น | HT | 0.0574 | 1.0591 |
| HT Duration | ระยะเวลาที่เป็นความดัน โลหิตสูง (ปี) | ht_duration | 0.0166 | 1.0167 |
| HT Age of Diag | อายุที่ได้รับการวินิจฉัยว่า เป็นโรคความดันโลหิตสูง (ปี) | ht_age_of_diag | 0.0180 | 1.0182 |
| DLP | สถานะภาวะไขมันผิดปกติ - เป็น - ไม่เป็น | flag_dlp | 0 | 1 |
| DM Duration | ระยะเวลาที่เป็น โรคเบาหวาน (ปี) | db_duration | -0.0167 | 0.9834 |
| DM Age of Diag | อายุที่ได้รับการวินิจฉัยว่า เป็นโรคเบาหวาน (ปี) | db_age_of_diag | 0.0050 | 1.0050 |
| CKD | สถานะการเป็นโรคไต เรื้อรังโดยดูจากรหัสโรคที่ วินิจฉัย - เป็น - ไม่เป็น | flag_CKD | 0 | 1 |
| KIDNEY | สถานะการเป็นโรคไต เรื้อรังในระยะสุดท้าย (พิจารณาจาก CKD และ ค่า eGFR) - เป็น - ไม่เป็น | KIDNEY | 0.2281 | 1.2562 |

ตารางที่ 31 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ใน
การทำนายความเสี่ยงภาวะเส้นเลือดหัวใจตีบหรือกล้ามเนื้อหัวใจขาดเลือดในผู้ป่วยโรคเบาหวาน (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ใน ตัวแบบ | สัมประสิทธิ์ การถดถอย | Odds Ratio |
|----------------------|--|------------------------------|--------------------------|------------|
| Stroke | สถานะการเป็นโรคหลอดเลือดสมอง - เป็น - ไม่เป็น | CEREBRO | 0.2392 | 1.2703 |
| flag_stroke | สถานะการเป็นโรคหลอดเลือดสมอง พิจารณาจากรหัสโรคที่วินิจฉัย - เป็น - ไม่เป็น | flag_stroke | 0 | 1 |
| Eye | สถานะการเป็นโรคตาอันเนื่องมาจากเบาหวาน - เป็น - ไม่เป็น | EYE | 0.2074 | 1.2305 |
| flag_eye | สถานะการเป็นโรคตา พิจารณาจากรหัสโรคที่วินิจฉัย - เป็น - ไม่เป็น | flag_eye | 0 | 1 |
| COPD | สถานะการเป็นโรคถุงลมโป่งพอง - เป็น - ไม่เป็น | flag_copd | 0 | 1 |
| Lab | | | | |
| TC | ระดับคอเลสเตอรอล (mg/dL) | tcl_labresult | 0.0001 | 1.0001 |
| is_na_tcl_labresult* | ทราบค่า TC หรือไม่ - ทราบ - ไม่ทราบ | is_na_tcl_labresult | | |
| HDL | ระดับไขมันดี (mg/dL) | hdl_labresult | -0.0043 | 0.9957 |

ตารางที่ 31 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ใน
การทำนายความเสี่ยงภาวะเส้นเลือดหัวใจตีบหรือกล้ามเนื้อหัวใจขาดเลือดในผู้ป่วยโรคเบาหวาน (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ใน ตัวแบบ | สัมประสิทธิ์ การถดถอย | Odds Ratio |
|-------------------------|---|------------------------------|--------------------------|------------|
| is_na_hdl_labresult* | ทราบค่า HDL หรือไม่ - ทราบ - ไม่ทราบ | is_na_hdl_labresult | 0 | 1 |
| LDL | ระดับไขมันไม่ดี (mg/dL) | ldl_labresult | 0.0027 | 1.0027 |
| is_na_ldl_labresult* | ทราบค่า HDL หรือไม่ - ทราบ - ไม่ทราบ | is_na_ldl_labresult | 0.3785 | 1.4601 |
| TG | ระดับไขมันไตรกลีเซอไรด์ (mg/dL) | tg_labresult | 0.0002 | 1.0002 |
| is_na_tg_labresult* | ทราบค่า TG หรือไม่ - ทราบ - ไม่ทราบ | is_na_tg_labresult | 0.1944 | 1.2146 |
| non-HDL | TC – HDL | non_hdl | 0 | 1 |
| is_na_non_hdl* | ทราบค่า non-HDL หรือไม่ - ทราบ - ไม่ทราบ | is_na_non_hdl | 0 | 1 |
| total Cho:HDL | TC/HDL | total_cho_to_hdl | 0 | 1 |
| is_na_total_cho_to_hdl* | ทราบค่า total Cho:HDL หรือไม่ - ทราบ - ไม่ทราบ | is_na_total_cho_to_hdl | 0 | 1 |
| non HDL:HDL | (TC – HDL)/HDL | no_hdl_to_hdl | 0 | 1 |

ตารางที่ 31 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ใน
การทำนายความเสี่ยงภาวะเส้นเลือดหัวใจตีบหรือกล้ามเนื้อหัวใจขาดเลือดในผู้ป่วยโรคเบาหวาน (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ในตัวแบบ | สัมประสิทธิ์การถดถอย | Odds Ratio |
|------------------------|---|-----------------------------|----------------------|------------|
| is_na_no_hdl_to_hdl* | ทราบค่า non HDL:HDL หรือไม่ - ทราบ - ไม่ทราบ | is_na_no_hdl_to_hdl | 0 | 1 |
| HBA1c | ค่าน้ำตาลเฉลี่ยสะสมในเลือด (mmol/mol) | HBA1c_labresult | 0.0486 | 1.0498 |
| is_na_HBA1C_labresult* | ทราบค่า HBA1c หรือไม่ - ทราบ - ไม่ทราบ | is_na_HBA1C_labresult | 0.6103 | 1.8410 |
| FBS | Fasting Blood Sugar (mg/dL) | FBS_labresult | 0.0014 | 1.0014 |
| is_na_FBS_labresult* | ทราบค่า FBS หรือไม่ - ทราบ - ไม่ทราบ | is_na_FBS_labresult | 0 | 1 |
| Macro-albuminuria | สถานะความเสี่ยงการเป็นโรคไตโดยวัดจาก Macro-albuminuria - เป็น - ไม่เป็น | macro_albuminuria_labresult | 0.2504 | 1.2846 |
| Micro-albuminuria | สถานะความเสี่ยงการเป็นโรคไตโดยวัดจาก Micro-albuminuria - เป็น - ไม่เป็น | micro_albuminuria_labresult | 0.1703 | 1.1857 |
| eGFR | Estimated Glomerular Filtration Rate (mL/min/1.73m ²) | egfr_labresult | -0.0119 | 0.9882 |

ตารางที่ 31 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ใน
การทำนายความเสี่ยงภาวะเส้นเลือดหัวใจตีบหรือกล้ามเนื้อหัวใจขาดเลือดในผู้ป่วยโรคเบาหวาน (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ใน ตัวแบบ | สัมประสิทธิ์ การถดถอย | Odds Ratio |
|-----------------------------|--|------------------------------|--------------------------|------------|
| is_na_egfr_labresult* | ทราบค่า eGFR หรือไม่ - ทราบ - ไม่ทราบ | is_na_egfr_labresult | -0.6097 | 0.5435 |
| Creatinine | ระดับ Creatinine (mg/dL) | creatinine_labresult | -0.0022 | 0.9978 |
| is_na_creatinine_labresult* | ทราบค่า Creatinine หรือไม่ - ทราบ - ไม่ทราบ | is_na_creatinine_labresult | -0.6097 | 0.5435 |
| UPCR | ระดับโปรตีนในปัสสาวะ (mg/g) | upcr_labresult | 0 | 1 |
| is_na_upcr_labresult* | ทราบค่า UPCR หรือไม่ - ทราบ - ไม่ทราบ | is_na_upcr_labresult | 0 | 1 |
| Hemoglobin | ระดับฮีโมโกลบินในเลือด (g/dL) | hemoglobin_labresult | 0 | 1 |
| is_na_hemoglobin_labresult* | ทราบค่า Hemoglobin หรือไม่ - ทราบ - ไม่ทราบ | is_na_hemoglobin_labresult | -0.4132 | 0.6615 |
| Albumin | ระดับ Albumin (g/dL) | albumin_labresult | 0 | 1 |

ตารางที่ 31 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ใน
การทำนายความเสี่ยงภาวะเส้นเลือดหัวใจตีบหรือกล้ามเนื้อหัวใจขาดเลือดในผู้ป่วยโรคเบาหวาน (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ในตัวแบบ | สัมประสิทธิ์การถดถอย | Odds Ratio |
|---------------------------|--|--------------------------|----------------------|------------|
| is_na_albumin_labresult * | ทราบค่า Albumin หรือไม่ - ทราบ - ไม่ทราบ | is_na_albumin_labresult | -0.0209 | 0.9793 |
| Medication | | | | |
| Non-insulin med | การใช้ยาเบาหวานที่ไม่ต้องพึ่งอินซูลิน - ใช้ - ไม่ใช้ | flag_non_insulin_med | 0 | 1 |
| Insulin med | การใช้ยาเบาหวานที่ต้องพึ่งอินซูลิน - ใช้ - ไม่ใช้ | flag_insulin_med | 0 | 1 |

ตัวแบบทำนายความเสี่ยงภาวะเส้นเลือดสมองแตกหรือตีบในผู้ป่วยโรคเบาหวาน

จากกลุ่มตัวอย่างของผู้ป่วยโรคเบาหวานจำนวน 89,166 คน ซึ่งประกอบไปด้วยข้อมูลจำนวน 530,482 แถว จะถูกสุ่มแบ่งออกเป็นชุดข้อมูลฝึกฝนและชุดข้อมูลทดสอบคิดเป็นร้อยละ 70 และ 30 โดยประมาณตามลำดับ โดยที่ชุดข้อมูลฝึกฝนและชุดข้อมูลทดสอบจะมีสัดส่วนของผู้ป่วยโรคเบาหวานที่พบภาวะเส้นเลือดสมองแตกหรือตีบพอ ๆ กัน ซึ่งในที่นี้พบว่ามีความเท่ากับ 0.008336 หรือคือมีผู้ป่วยโรคเบาหวานคิดเป็นร้อยละ 0.8336 ที่มีภาวะเส้นเลือดสมองแตกหรือตีบนั่นเอง จากการแบ่งข้อมูลจะได้ข้อมูลดังต่อไปนี้

- ชุดข้อมูลฝึกฝน ประกอบด้วยข้อมูลจากผู้เป็นโรคเบาหวานจำนวน 62,417 คน ซึ่งมีข้อมูลจำนวน 371,043 แถว
- ชุดข้อมูลทดสอบ ประกอบด้วยข้อมูลจากผู้เป็นโรคเบาหวานจำนวน 26,749 คน ซึ่งมีข้อมูลจำนวน 159,439 แถว

จากนั้นจึงสร้างตัวแบบจากข้อมูลชุดฝึกฝน ซึ่งจะพิจารณา 2 ตัวแบบ ได้แก่

1. การถดถอยโลจิสติกแบบ Elastic Net วิธีการนี้ขอเรียกว่า “EN”
2. วิธี XGBoost โดยวิธีการนี้ขอเรียกว่า “XGBoost”

หลังจากได้ตัวแบบทั้ง 2 แล้ว จึงนำมาทำนายข้อมูลในชุดทดสอบเพื่อวัดประสิทธิภาพการทำนายโดยพิจารณาจาก

1. AUC
2. Cumulative AUC

ตารางที่ 32 แสดงค่า AUC จาก 2 วิธี ซึ่งพบว่าวิธี XGBoost มีประสิทธิภาพในการทำนายความเสี่ยงการภาวะเส้นเลือดสมองแตกหรือตีบในผู้ป่วยโรคเบาหวานได้ดีกว่าวิธีการถดถอยโลจิสติกแบบ Elastic Net แม้ว่าค่า AUC ที่ได้จากวิธี XGBoost จะดีกว่าวิธีการถดถอยโลจิสติกแบบ Elastic Net ไม่มาก แต่หากพิจารณาค่า Cumulative AUC ซึ่งวัดประสิทธิภาพการทำนายรายบุคคลผ่านตัววัดผล AUC ที่คำนวณด้วยความน่าจะเป็นสะสมของบุคคลตลอด Cohort (Cumulative Probability) กับสถานะอุบัติการณ์ปีสุดท้าย จะพบว่าวิธี XGBoost มีประสิทธิภาพเหนือกว่าวิธี Elastic Net ค่อนข้างมาก อย่างไรก็ตาม พบว่าวิธีทั้งสองอาจจะมีค่า AUC ที่ยังอยู่ในเกณฑ์ไม่สูง หรือมีค่าต่ำกว่า 0.8 ซึ่งในอนาคตควรจะต้องพิจารณาตัวแบบอื่นๆ รวมทั้งพิจารณาตัวแปรหรือใช้ขนาดตัวอย่างเพิ่มเติม เพื่อที่จะปรับปรุงตัวแบบในการทำนายให้มีประสิทธิภาพมากขึ้น

ตารางที่ 32 ค่า AUC เพื่อวัดประสิทธิภาพการทำนายความเสี่ยงภาวะเส้นเลือดสมองแตกหรือตีบในผู้ป่วยโรคเบาหวาน (Cohort 1.2) (วิธีที่ดีที่สุดจะแสดงเป็นตัวหนา)

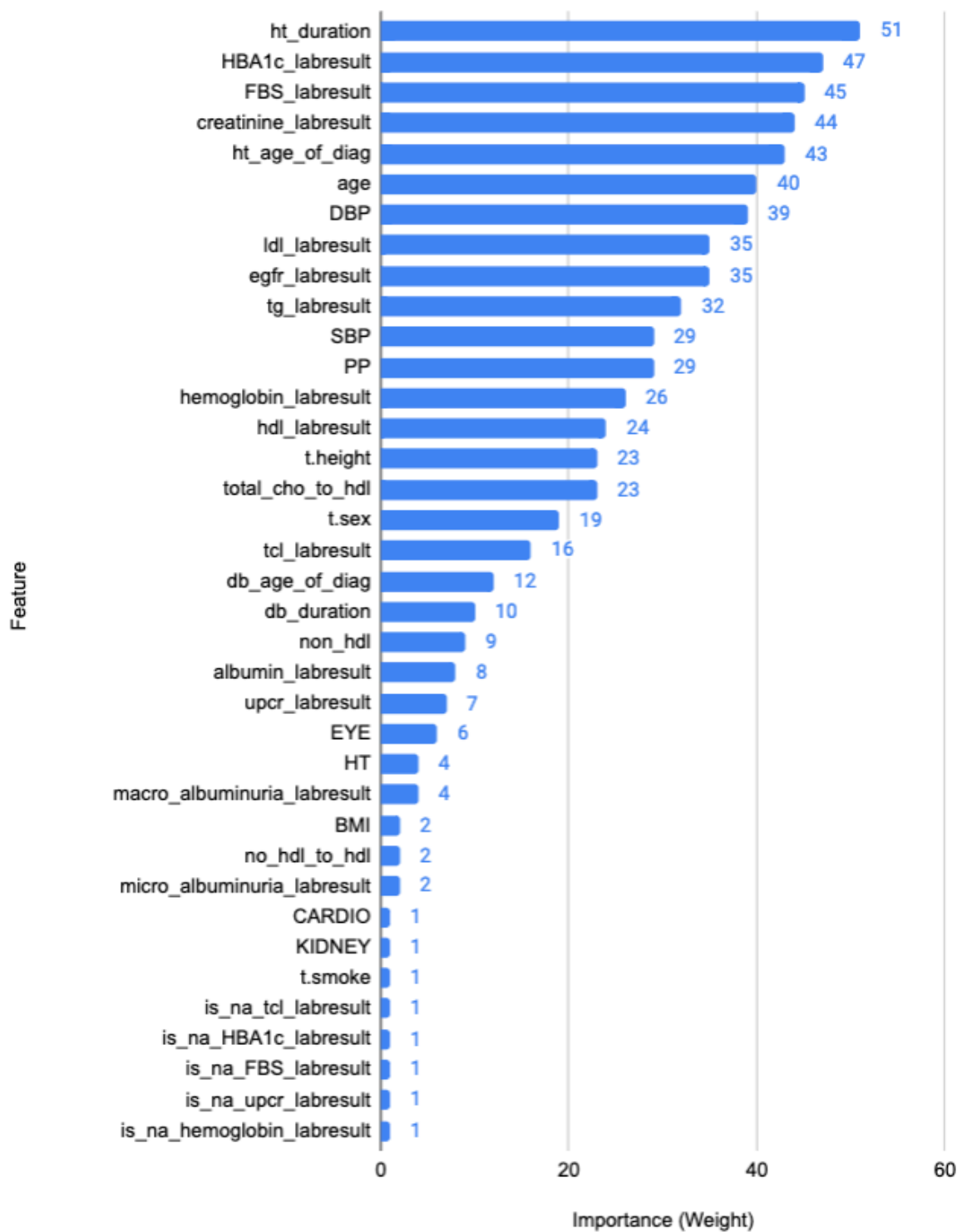
| ตัวแบบ | AUC | Cumulative AUC |
|----------------|---------------|----------------|
| EN | 0.6501 | 0.6248 |
| XGBoost | 0.6666 | 0.6378 |

สำหรับการแปลผลตัวแบบโดยอ้างอิงจากตัวแบบ XGBoost ที่ให้ประสิทธิภาพการทำนายดีที่สุดที่สามารถพิจารณาได้ว่าปัจจัยใดที่มีความสำคัญในการทำนายความเสี่ยงการเกิดโรคเบาหวานในประชากรทั่วไป ดังแสดงในรูปที่ 52 ซึ่งจะพบว่าตัวแปรที่มีความสำคัญในการทำนายความเสี่ยง 10 อันดับแรก ได้แก่

1. ระยะเวลาที่เป็นความดันโลหิตสูง (ปี) (ht_duration)
2. ค่าน้ำตาลเฉลี่ยสะสมในเลือด (mmol/mol) (HBA1c_labresult)
3. Fasting Blood Sugar (mg/dL) (FBS_labresult)
4. ระดับ Creatinine (mg/dL) (creatinine_labresult)
5. อายุที่ได้รับการวินิจฉัยว่าเป็นโรคความดันโลหิตสูง (ปี) (ht_age_of_diag)
6. อายุ (age)
7. Diastolic Blood Pressure (mm Hg) (DBP)
8. ระดับไขมันไม่ดี (mg/dL) (ldl_labresult)

9. Estimated Glomerular Filtration Rate (mL/min/1.73m²) (egfr_labresult)

10. ระดับไขมันไตรกลีเซอไรด์ (mg/dL) (tg_labresult)

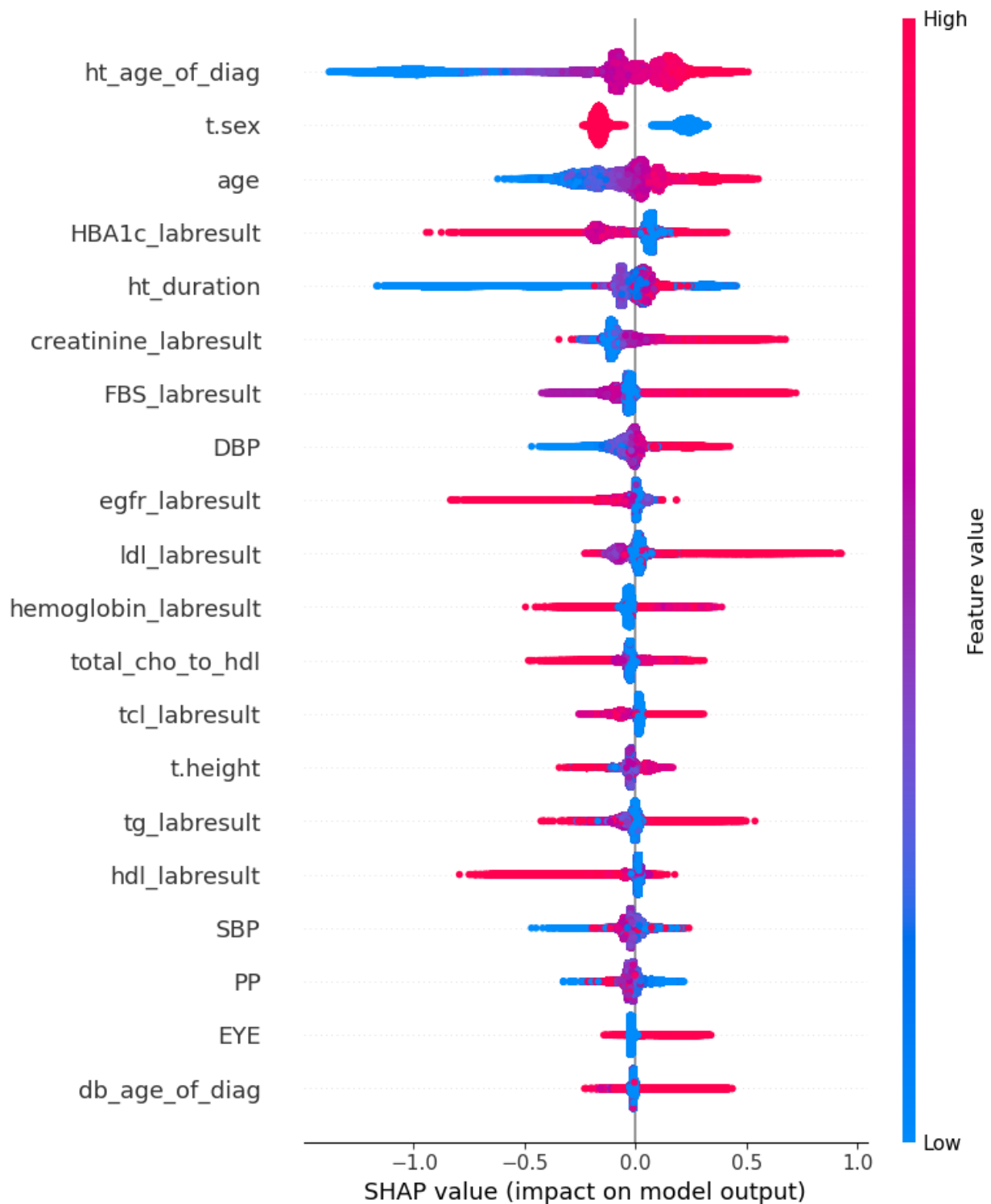


รูปที่ 52 ความสำคัญของตัวแปรทำนายจากวิธี XGBoost ในการทำนายความเสี่ยงภาวะเส้นเลือดสมองแตกหรือตีบในผู้ป่วยโรคเบาหวาน

นอกจากนี้ หากพิจารณาตัว SHAP ของตัวแปร 20 ตัวแปรที่มีความสำคัญมากที่สุดในการทำนายความเสี่ยง ภาวะเส้นเลือดสมองแตกหรือตีบในผู้ป่วยเบาหวาน ดังแสดงในรูปที่ 53 จะสามารถแปลผลในส่วนทิศทาง ความสัมพันธ์ของตัวแปรแต่ละตัว ซึ่งสามารถสรุปความสัมพันธ์ที่เด่นชัดได้ ดังนี้

- ผู้ป่วยเบาหวานที่ได้รับการวินิจฉัยว่าเป็นโรคความดันโลหิตสูงตอนช่วงอายุยังน้อย (ht_age_of_diag) จะมีแนวโน้มที่จะมีความเสี่ยงในการมีภาวะเส้นเลือดสมองแตกหรือตีบต่ำลง
- ผู้ป่วยเบาหวานที่มีระยะเวลาที่เป็นความดันโลหิตสูง (ปี) (ht_duration) น้อย จะมีแนวโน้มที่จะมีความเสี่ยงในการมีภาวะเส้นเลือดสมองแตกหรือตีบลดลง
- ผู้ป่วยเบาหวานที่มีระดับไขมันไม่ดี (mg/dL) (ldl_labresult) สูง จะมีแนวโน้มที่จะมีความเสี่ยงในการมีภาวะเส้นเลือดสมองแตกหรือตีบสูงขึ้น
- ผู้ป่วยเบาหวานที่มีระดับ Creatinine (mg/dL) (creatinine_labresult) สูง จะมีแนวโน้มที่จะมีความเสี่ยงในการมีภาวะเส้นเลือดสมองแตกหรือตีบสูงขึ้น

แม้ว่าวิธี XGBoost สามารถระบุความสำคัญของตัวแปรทำนายที่ใช้ในตัวแบบได้ แต่ไม่สามารถแปลผลตัวแบบในรูปของอัตราเสี่ยงได้ ในขณะที่วิธีการถดถอยโลจิสติกแบบ Elastic Net ให้ค่าสัมประสิทธิ์ การถดถอยเช่นเดียวกับตัวแบบการถดถอยโลจิสติกแบบปกติ ดังนั้นจึงสามารถแปลผลตัวแบบในรูปของอัตราเสี่ยง ได้ ตารางที่ 33 แสดงผลลัพธ์ที่ได้ของสัมประสิทธิ์การถดถอยของตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ในการทำนายความเสี่ยงภาวะเส้นเลือดสมองแตกหรือตีบในผู้ป่วยโรคเบาหวาน



รูปที่ 53 ค่า SHAP จากตัวแบบ XGBoost ในการทำนายความเสี่ยงภาวะเส้นเลือดสมองแตกหรือตีบในผู้ป่วยโรคเบาหวาน

ตารางที่ 33 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net
ในการทำนายความเสี่ยงภาวะเส้นเลือดสมองแตกหรือตีบในผู้ป่วยโรคเบาหวาน

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ในตัวแบบ | สัมประสิทธิ์การถดถอย | Odds Ratio |
|-----------------------------|--|--------------------------|----------------------|------------|
| Demographic | | | | |
| Age | อายุ (ปี) | age | 0.0084 | 1.0084 |
| Gender | เพศ - ชาย - หญิง | t.sex | -0.5113 | 0.5997 |
| Health Behavior | | | | |
| Smoking | พฤติกรรมการสูบบุหรี่ - สูบ - ไม่สูบ | t.smoke | 0.0546 | 1.0562 |
| Physical measurement | | | | |
| Height | ส่วนสูง (เซนติเมตร) | t.height | 0.0006 | 1.0006 |
| BMI | ดัชนีมวลกาย | BMI | 0.0127 | 1.1028 |
| SBP | Systolic Blood Pressure (mm Hg) | SBP | 0 | 1 |
| DBP | Diastolic Blood Pressure (mm Hg) | DBP | 0.0081 | 1.0081 |
| PP | Pulse Pressure คำนวณจาก SBP - DBP | PP | -0.0014 | 0.9986 |
| Underlying Disease | | | | |
| HT | สถานะโรคความดันโลหิตสูง - เป็น - ไม่เป็น | HT | -0.1016 | 0.9034 |
| HT Duration | ระยะเวลาที่เป็นความดันโลหิตสูง (ปี) | ht_duration | 0.2227 | 1.0229 |
| HT Age of Diag | อายุที่ได้รับการวินิจฉัยว่าเป็นโรคความดันโลหิตสูง (ปี) | ht_age_of_diag | 0.0185 | 1.0187 |

ตารางที่ 33 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net
ในการทำนายความเสี่ยงภาวะเส้นเลือดสมองแตกหรือตีบในผู้ป่วยโรคเบาหวาน (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ในตัวแบบ | สัมประสิทธิ์การถดถอย | Odds Ratio |
|--|--|--------------------------|----------------------|------------|
| DLP | สถานะภาวะไขมันผิดปกติ - เป็น - ไม่เป็น | flag_dlp | 0 | 1 |
| DM Duration | ระยะเวลาที่เป็นโรคเบาหวาน (ปี) | db_duration | -0.0017 | 0.9983 |
| DM Age of Diag | อายุที่ได้รับการวินิจฉัยว่าเป็นโรคเบาหวาน (ปี) | db_age_of_diag | 0.0018 | 1.0018 |
| CKD | สถานะการเป็นโรคไตเรื้อรังโดยดูจากรหัสโรคที่วินิจฉัย - เป็น - ไม่เป็น | flag_CKD | 0 | 1 |
| KIDNEY | สถานะการเป็นโรคไตเรื้อรังในระยะสุดท้าย (พิจารณาจาก CKD และค่า eGFR) - เป็น - ไม่เป็น | KIDNEY | -0.0115 | 0.9886 |
| Heart Disease (Cardiovascular Disease) | สถานะการเป็นภาวะเส้นเลือดหัวใจตีบหรือกล้ามเนื้อหัวใจขาดเลือด - เป็น - ไม่เป็น | CARDIO | 0.1889 | 1.2080 |
| Heart Disease | สถานะการเป็นโรคหัวใจซึ่งรวมถึงการเป็นโรคหัวใจเต้นผิดปกติ - เป็น - ไม่เป็น | flag_HEART | 0 | 1 |

ตารางที่ 33 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net
ในการทำนายความเสี่ยงภาวะเส้นเลือดสมองแตกหรือตีบในผู้ป่วยโรคเบาหวาน (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ในตัวแบบ | สัมประสิทธิ์การถดถอย | Odds Ratio |
|--------------------------|--|--------------------------|----------------------|------------|
| Eye | สถานะการเป็นโรคตาอัน เนื่องมาจากเบาหวาน - เป็น - ไม่เป็น | EYE | 0.3524 | 1.4225 |
| flag_eye | สถานะการเป็นโรคตา พิจารณาจากรหัสโรคที่วินิจฉัย - เป็น - ไม่เป็น | flag_eye | 0 | 1 |
| COPD | สถานะการเป็นโรคถุงลมโป่ง พอง - เป็น - ไม่เป็น | flag_copd | 0 | 1 |
| Lab | | | | |
| TC | ระดับคลอเรสเตอรอล (mg/dL) | tcl_labresult | 0.0001 | 1.0001 |
| is_na_tcl_labresult* | ทราบค่า TC หรือไม่ - ทราบ - ไม่ทราบ | is_na_tcl_labresult | 0.0308 | 1.0313 |
| HDL | ระดับไขมันดี (mg/dL) | hdl_labresult | -0.0058 | 0.9942 |
| is_na_hdl_labresult * | ทราบค่า HDL หรือไม่ - ทราบ - ไม่ทราบ | is_na_hdl_labresult | -0.5994 | 0.5491 |
| LDL | ระดับไขมันไม่ดี (mg/dL) | ldl_labresult | 0.0044 | 1.0044 |
| is_na_ldl_labresult* | ทราบค่า HDL หรือไม่ - ทราบ - ไม่ทราบ | is_na_ldl_labresult | 0.7318 | 2.0787 |
| TG | ระดับไขมันไตรกลีเซอไรด์ (mg/dL) | tg_labresult | 0.0006 | 1.0006 |

ตารางที่ 33 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net
ในการทำนายความเสี่ยงภาวะเส้นเลือดสมองแตกหรือตีบในผู้ป่วยโรคเบาหวาน (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ในตัวแบบ | สัมประสิทธิ์การถดถอย | Odds Ratio |
|-------------------------|--|--------------------------|----------------------|------------|
| is_na_tg_labresult* | ทราบค่า TG หรือไม่ - ทราบ - ไม่ทราบ | is_na_tg_labresult | 0.1372 | 1.1470 |
| non-HDL | TC – HDL | non_hdl | 0.0012 | 1.0012 |
| is_na_non_hdl* | ทราบค่า non-HDL หรือไม่ - ทราบ - ไม่ทราบ | is_na_non_hdl | 0 | 1 |
| total Cho:HDL | TC/HDL | total_cho_to_hdl | -0.0281 | 0.9723 |
| is_na_total_cho_to_hdl* | ทราบค่า total Cho:HDL หรือไม่ - ทราบ - ไม่ทราบ | is_na_total_cho_to_hdl | 0.1568 | 1.1698 |
| non HDL:HDL | (TC – HDL)/HDL | no_hdl_to_hdl | 0 | 1 |
| is_na_no_hdl_to_hdl* | ทราบค่า non HDL:HDL หรือไม่ - ทราบ - ไม่ทราบ | is_na_no_hdl_to_hdl | 0.1568 | 1.1698 |
| HBA1c | ค่าน้ำตาลเฉลี่ยสะสมในเลือด (mmol/mol) | HBA1c_labresult | 0.0235 | 1.0238 |
| is_na_HBA1C_labresult* | ทราบค่า HBA1c หรือไม่ - ทราบ - ไม่ทราบ | is_na_HBA1C_labresult | 0.3934 | 1.4820 |
| FBS | Fasting Blood Sugar (mg/dL) | FBS_labresult | 0.0026 | 1.0026 |
| is_na_FBS_labresult* | ทราบค่า FBS หรือไม่ - ทราบ - ไม่ทราบ | is_na_FBS_labresult | 0.3539 | 1.4247 |

ตารางที่ 33 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net
ในการทำนายความเสี่ยงภาวะเส้นเลือดสมองแตกหรือตีบในผู้ป่วยโรคเบาหวาน (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ในตัวแบบ | สัมประสิทธิ์การถดถอย | Odds Ratio |
|-----------------------------|---|-----------------------------|----------------------|------------|
| Macro-albuminuria | สถานะความเสี่ยงการเป็นโรคไตโดยวัดจาก Macro-albuminuria - เป็น - ไม่เป็น | macro_albuminuria_labresult | 0.2249 | 1.2522 |
| Micro-albuminuria | สถานะความเสี่ยงการเป็นโรคไตโดยวัดจาก Micro-albuminuria - เป็น - ไม่เป็น | micro_albuminuria_labresult | 0.0574 | 1.0590 |
| eGFR | Estimated Glomerular Filtration Rate (mL/min/1.73m ²) | egfr_labresult | -0.0061 | 0.9939 |
| is_na_egfr_labresult* | ทราบค่า eGFR หรือไม่ - ทราบ - ไม่ทราบ | is_na_egfr_labresult | -0.3522 | 0.7031 |
| Creatinine | ระดับ Creatinine (mg/dL) | creatinine_labresult | -0.0052 | 0.9949 |
| is_na_creatinine_labresult* | ทราบค่า Creatinine หรือไม่ - ทราบ - ไม่ทราบ | is_na_creatinine_labresult | -0.3239 | 0.7233 |
| UPCR | ระดับโปรตีนในปัสสาวะ (mg/g) | upcr_labresult | -0.0009 | 0.9991 |
| is_na_upcr_labresult* | ทราบค่า UPCR หรือไม่ - ทราบ - ไม่ทราบ | is_na_upcr_labresult | -0.0094 | 0.9907 |

ตารางที่ 33 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net
ในการทำนายความเสี่ยงภาวะเส้นเลือดสมองแตกหรือตีบในผู้ป่วยโรคเบาหวาน (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ในตัวแบบ | สัมประสิทธิ์การถดถอย | Odds Ratio |
|-----------------------------|--|----------------------------|----------------------|------------|
| Hemoglobin | ระดับฮีโมโกลบินในเลือด (g/dL) | hemoglobin_labresult | -0.0225 | 0.9778 |
| is_na_hemoglobin_labresult* | ทราบค่า Hemoglobin หรือไม่ - ทราบ - ไม่ทราบ | is_na_hemoglobin_labresult | -0.4609 | 0.6307 |
| Albumin | ระดับ Albumin (g/dL) | albumin_labresult | -0.1335 | 0.8750 |
| is_na_albumin_labresult* | ทราบค่า Albumin หรือไม่ - ทราบ - ไม่ทราบ | is_na_albumin_labresult | -0.5522 | 0.5757 |
| Medication | | | | |
| Lipid-lowering med | การใช้ยาลดไขมัน - ใช้ - ไม่ใช้ | flag_lipid_lowering | 0 | 1 |
| Non-insulin med | การใช้ยาเบาหวานที่ไม่ต้องพึ่งอินซูลิน - ใช้ - ไม่ใช้ | flag_non_insulin_med | 0 | 1 |
| Insulin med | การใช้ยาเบาหวานที่ต้องพึ่งอินซูลิน - ใช้ - ไม่ใช้ | flag_insulin_med | 0 | 1 |
| Aspirin | การใช้ยาแอสไพริน - ใช้ - ไม่ใช้ | flag_aspirin_med | 0 | 1 |

ตัวแบบทำนายความเสี่ยงภาวะแทรกซ้อนทางตาในผู้ป่วยโรคเบาหวาน

จากกลุ่มตัวอย่างของผู้ป่วยโรคเบาหวานจำนวน 88,958 คน ซึ่งประกอบไปด้วยข้อมูลจำนวน 514,077 แถว จะถูกสุ่มแบ่งออกเป็นชุดข้อมูลฝึกฝนและชุดข้อมูลทดสอบคิดเป็นร้อยละ 70 และ 30 โดยประมาณ ตามลำดับ โดยที่ชุดข้อมูลฝึกฝนและชุดข้อมูลทดสอบจะมีสัดส่วนของผู้ป่วยโรคเบาหวานที่พบภาวะแทรกซ้อนทางตาพอ ๆ กัน ซึ่งในที่นี้พบว่ามีค่าเท่ากับ 0.013611 หรือคือมีผู้ป่วยโรคเบาหวานคิดเป็นร้อยละ 1.3611 ที่มีภาวะแทรกซ้อนทางตานั่นเอง จากการแบ่งข้อมูลจะได้ข้อมูลดังต่อไปนี้

- ชุดข้อมูลฝึกฝน ประกอบด้วยข้อมูลจากผู้เป็นโรคเบาหวานจำนวน 62,271 คน ซึ่งมีข้อมูลจำนวน 359,863 แถว
- ชุดข้อมูลทดสอบ ประกอบด้วยข้อมูลจากผู้เป็นโรคเบาหวานจำนวน 26,687 คน ซึ่งมีข้อมูลจำนวน 154,214 แถว

จากนั้นจึงสร้างตัวแบบจากข้อมูลชุดฝึกฝน ซึ่งจะพิจารณา 2 ตัวแบบ ได้แก่

1. การถดถอยโลจิสติกแบบ Elastic Net วิธีการนี้ขอเรียกว่า “EN”
2. วิธี XGBoost โดยวิธีการนี้ขอเรียกว่า “XGBoost”

หลังจากได้ตัวแบบทั้ง 2 แล้ว จึงนำมาทำนายข้อมูลในชุดทดสอบเพื่อวัดประสิทธิภาพการทำนายโดยพิจารณาจาก

1. AUC
2. Cumulative AUC

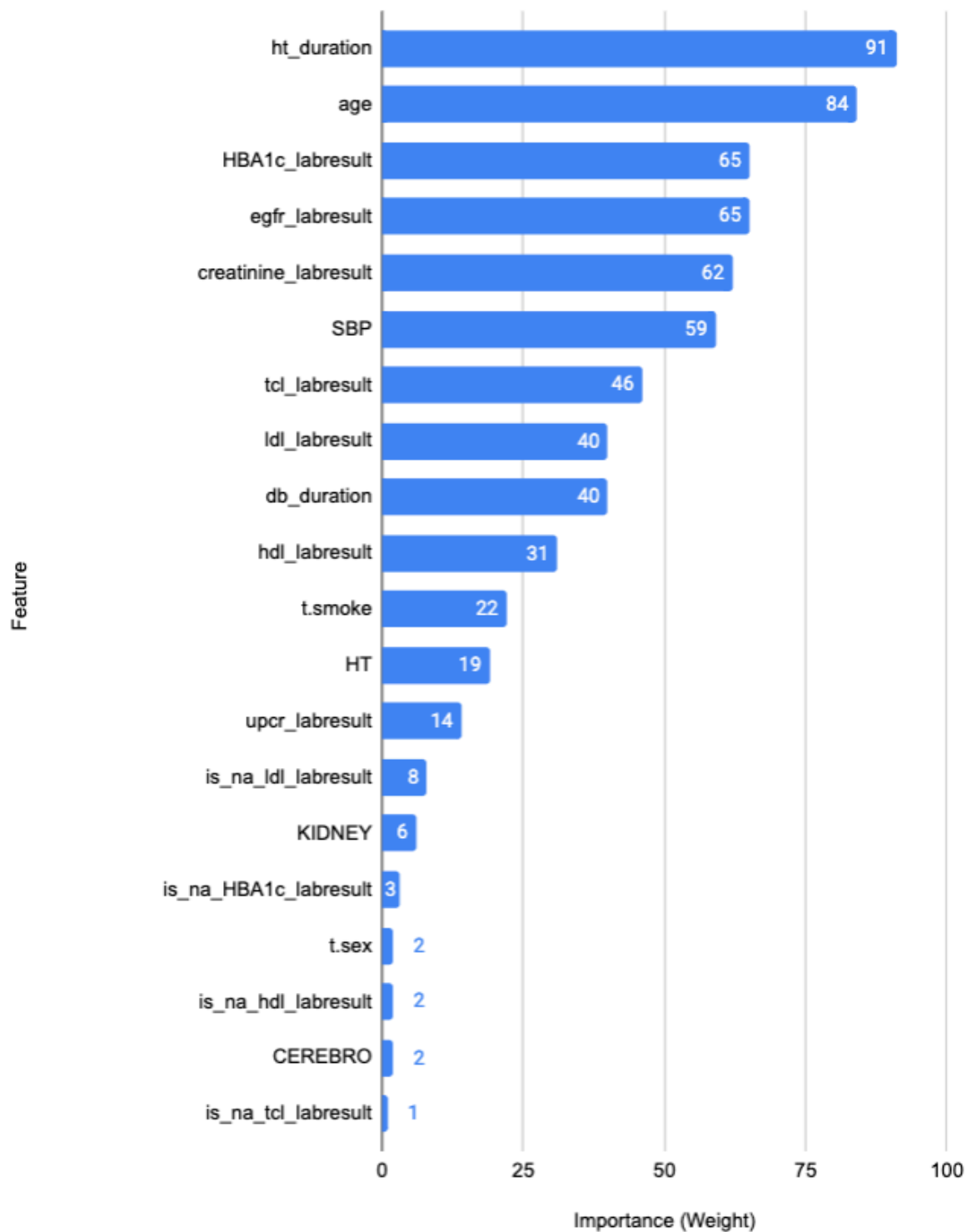
ตารางที่ 34 แสดงค่า AUC จาก 2 วิธี ซึ่งพบว่าวิธี XGBoost มีประสิทธิภาพในการทำนายความเสี่ยงการภาวะแทรกซ้อนทางตาในผู้ป่วยโรคเบาหวานได้ดีกว่าวิธีการถดถอยโลจิสติกแบบ Elastic Net แม้ว่าค่า AUC ที่ได้จากวิธี XGBoost จะดีกว่าวิธีการถดถอยโลจิสติกแบบ Elastic Net ไม่มาก แต่หากพิจารณาค่า Cumulative AUC ซึ่งวัดประสิทธิภาพการทำนายรายบุคคลผ่านตัววัดผล AUC ที่คำนวณด้วยความน่าจะเป็นสะสมของบุคคลตลอด Cohort (Cumulative Probability) กับสถานะอุบัติการณ์ปีสุดท้าย จะพบว่าวิธี XGBoost มีประสิทธิภาพเหนือกว่าวิธี Elastic Net ค่อนข้างมาก อย่างไรก็ตาม พบว่าวิธีทั้งสองอาจจะมีค่า AUC ที่ยังอยู่ในเกณฑ์ไม่สูง หรือมีค่าต่ำกว่า 0.8 ซึ่งในอนาคตควรจะต้องพิจารณาตัวแบบอื่นๆ รวมทั้งพิจารณาตัวแปรหรือใช้ขนาดตัวอย่างเพิ่มเติม เพื่อที่จะปรับปรุงตัวแบบในการทำนายให้มีประสิทธิภาพมากขึ้น

ตารางที่ 34 ค่า AUC เพื่อวัดประสิทธิภาพการทำนายความเสี่ยงภาวะแทรกซ้อนทางตาในผู้ป่วยโรคเบาหวาน (Cohort 1.2) (วิธีที่ดีที่สุดจะแสดงเป็นตัวหนา)

| ตัวแบบ | AUC | Cumulative AUC |
|----------------|---------------|----------------|
| EN | 0.6005 | 0.6471 |
| XGBoost | 0.6322 | 0.677 |

สำหรับการแปลผลตัวแบบโดยอ้างอิงจากตัวแบบ XGBoost ที่ให้ประสิทธิภาพการทำนายดีที่สุดที่สามารถพิจารณาได้ว่าปัจจัยใดที่มีความสำคัญในการทำนายความเสี่ยงการเกิดโรคเบาหวานในประชากรทั่วไป ดังแสดงในรูปที่ 54 ซึ่งจะพบว่าตัวแปรที่มีความสำคัญในการทำนายความเสี่ยง 10 อันดับแรก ได้แก่

1. ระยะเวลาที่เป็นความดันโลหิตสูง (ปี) (ht_duration)
2. อายุ (age)
3. ค่าน้ำตาลเฉลี่ยสะสมในเลือด (mmol/mol) (HBA1c_labresult)
4. Estimated Glomerular Filtration Rate (mL/min/1.73m²) (egfr_labresult)
5. ระดับ Creatinine (mg/dL) (creatinine_labresult)
6. Systolic Blood Pressure (mm Hg) (SBP)
7. ระดับคลอเลสเตอรอล (mg/dL) (tcl_labresult)
8. ระดับไขมันไม่ดี (mg/dL) (ldl_labresult)
9. ระยะเวลาที่เป็นโรคเบาหวาน (ปี) (db_duration)
10. ระดับไขมันดี (mg/dL) (hdl_labresult)

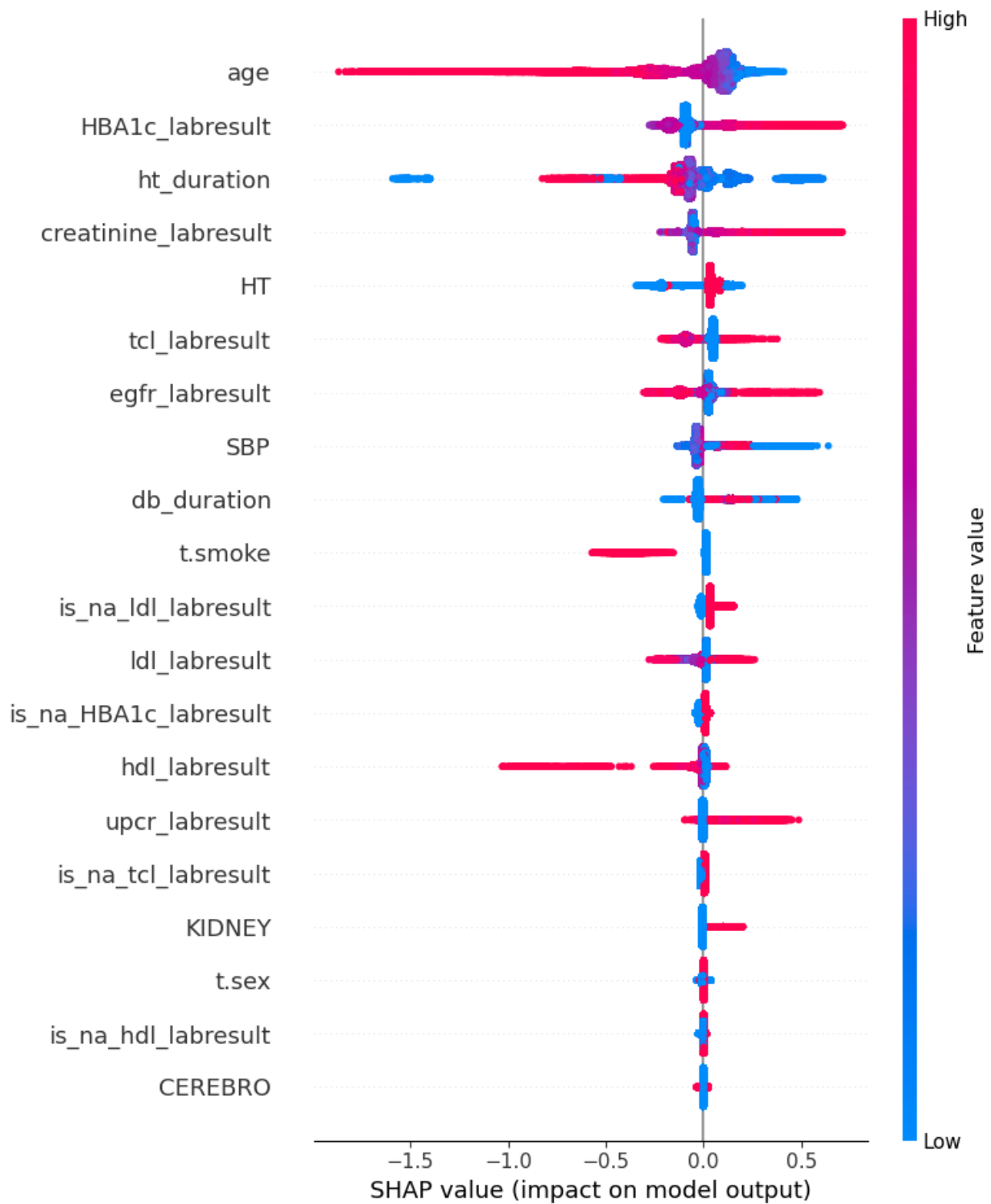


รูปที่ 54 ความสำคัญของตัวแปรทำนายจากวิธี XGBoost ในการทำนายภาวะแทรกซ้อนทางตาในผู้ป่วยโรคเบาหวาน

นอกจากนี้ หากพิจารณาคุณค่า SHAP ของตัวแปร 20 ตัวแปรที่มีความสำคัญมากที่สุดในการทำนายความเสี่ยงภาวะแทรกซ้อนทางตาในผู้ป่วยเบาหวาน ดังแสดงในรูปที่ 55 จะสามารถแปลผลในส่วนทิศทางความสัมพันธ์ของตัวแปรแต่ละตัว ซึ่งสามารถสรุปความสัมพันธ์ที่เด่นชัดได้ ดังนี้

- ผู้ป่วยเบาหวานที่ได้รับการวินิจฉัยว่ามีอายุน้อย (age) จะมีแนวโน้มที่จะมีความเสี่ยงในการมีภาวะแทรกซ้อนทางตาสูงขึ้น
- ผู้ป่วยเบาหวานที่มีระยะเวลาที่เป็นความดันโลหิตสูง (ปี) (ht_duration) น้อย จะมีแนวโน้มที่จะมีความเสี่ยงในการมีภาวะแทรกซ้อนทางตาสูงขึ้น
- ผู้ป่วยเบาหวานที่มีค่าน้ำตาลเฉลี่ยสะสมในเลือด (mmol/mol) (HBA1c_labresult) สูง จะมีแนวโน้มที่จะมีความเสี่ยงในการมีภาวะแทรกซ้อนทางตาสูงขึ้น
- ผู้ป่วยเบาหวานที่มีระดับ Creatinine (mg/dL) (creatinine_labresult) สูง จะมีแนวโน้มที่จะมีความเสี่ยงในการมีภาวะแทรกซ้อนทางตาสูงขึ้น

แม้ว่าวิธี XGBoost สามารถระบุความสำคัญของตัวแปรทำนายที่ใช้ในตัวแบบได้ แต่ไม่สามารถแปลผลตัวแบบในรูปของอัตราเสี่ยงได้ ในขณะที่วิธีการถดถอยโลจิสติกแบบ Elastic Net ให้ค่าสัมประสิทธิ์การถดถอยเช่นเดียวกับตัวแบบการถดถอยโลจิสติกแบบปกติ ดังนั้นจึงสามารถแปลผลตัวแบบในรูปของอัตราเสี่ยงได้ตารางที่ 31 แสดงผลลัพธ์ที่ได้ของสัมประสิทธิ์การถดถอยของตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ในการทำนายความเสี่ยงภาวะแทรกซ้อนทางตาในผู้ป่วยโรคเบาหวาน



รูปที่ 55 ค่า SHAP จากตัวแบบ XGBoost ในการทำนายความเสี่ยงภาวะแทรกซ้อนทางตาในผู้ป่วยโรคเบาหวาน

ตารางที่ 35 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ใน
การทำนายความเสี่ยงแทรกซ้อนทางตาในผู้ป่วยโรคเบาหวาน

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ใน ตัวแบบ | สัมประสิทธิ์ การถดถอย | Odds Ratio |
|-----------------------------|--|------------------------------|--------------------------|------------|
| Demographic | | | | |
| Age | อายุ (ปี) | age | -0.0206 | 0.9796 |
| Gender | เพศ - ชาย - หญิง | t.sex | -0.0133 | 0.9868 |
| Health Behavior | | | | |
| Smoking | พฤติกรรมการสูบบุหรี่ - สูบ - ไม่สูบ | t.smoke | -0.3891 | 0.6777 |
| Physical measurement | | | | |
| BMI | ดัชนีมวลกาย | BMI | 0.0009 | 1.0009 |
| SBP | Systolic Blood Pressure (mm Hg) | SBP | 0.0026 | 1.0026 |
| Underlying Disease | | | | |
| HT | สถานะโรคความดันโลหิตสูง - เป็น - ไม่เป็น | HT | 0.3145 | 1.3696 |
| HT Duration | ระยะเวลาที่เป็นความดันโลหิตสูง (ปี) | ht_duration | -0.0297 | 0.9708 |
| DM Duration | ระยะเวลาที่เป็นโรคเบาหวาน (ปี) | db_duration | 0.0200 | 1.0202 |
| KIDNEY | สถานะการเป็นโรคไตเรื้อรังในระยะสุดท้าย (พิจารณาจาก CKD และค่า eGFR) - เป็น - ไม่เป็น | KIDNEY | 0.3119 | 1.3660 |

ตารางที่ 35 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ใน
การทำนายความเสี่ยงแทรกซ้อนทางตาในผู้ป่วยโรคเบาหวาน (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ใน ตัวแบบ | สัมประสิทธิ์ การถดถอย | Odds Ratio |
|---|---|------------------------------|--------------------------|------------|
| Heart Disease (Cardiovascular Disease) | สถานะการเป็นภาวะ เส้นเลือดหัวใจตีบ หรือกล้ามเนื้อหัวใจ ขาดเลือด - เป็น - ไม่เป็น | CARDIO | 0 | 1 |
| Heart Disease | สถานะการเป็น โรคหัวใจ ซึ่งรวมถึง การเป็นโรคหัวใจเต้น ผิดปกติ - เป็น - ไม่เป็น | flag_HEART | 0 | 1 |
| Stroke | สถานะการเป็นโรค หลอดเลือดสมอง - เป็น - ไม่เป็น | CEREBRO | 0.0580 | 1.0597 |
| Lab | | | | |
| TC | ระดับคลอ เรสเตอรอล (mg/dL) | tcl_labresult | 0.0009 | 1.0009 |
| is_na_tcl_labresult* | ทราบค่า TC หรือไม่ - ทราบ - ไม่ทราบ | is_na_tcl_labresult | 0.3722 | 1.4509 |
| HDL | ระดับไขมันดี (mg/dL) | hdl_labresult | -0.0036 | 0.9964 |
| is_na_hdl_labresult* | ทราบค่า HDL หรือไม่ - ทราบ - ไม่ทราบ | is_na_hdl_labresult | -0.2819 | 0.7544 |
| LDL | ระดับไขมันไม่ดี (mg/dL) | ldl_labresult | 0.0020 | 1.0020 |

ตารางที่ 35 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ใน
การทำนายความเสี่ยงแทรกซ้อนทางตาในผู้ป่วยโรคเบาหวาน (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ใน ตัวแบบ | สัมประสิทธิ์ การถดถอย | Odds Ratio |
|---------------------------------|---|--------------------------------|--------------------------|------------|
| is_na_ldl_labresult* | ทราบค่า HDL หรือไม่ - ทราบ - ไม่ทราบ | is_na_ldl_labresult | 0.4907 | 1.6335 |
| HBA1c | ค่าน้ำตาลเฉลี่ยสะสม ในเลือด (mmol/mol) | HBA1c_labresult | 0.0694 | 1.0719 |
| is_na_HBA1C_labresult* | ทราบค่า HBA1c หรือไม่ - ทราบ - ไม่ทราบ | is_na_HBA1C_labre sult | 0.4271 | 1.5328 |
| eGFR | Estimated Glomerular Filtration Rate (mL/min/1.73m2) | egfr_labresult | -0.0067 | 0.9933 |
| is_na_egfr_labresult* | ทราบค่า eGFR หรือไม่ - ทราบ - ไม่ทราบ | is_na_egfr_labresul t | -0.3771 | 0.6858 |
| Creatinine | ระดับ Creatinine (mg/dL) | creatinine_labresult | -0.0131 | 0.9870 |
| is_na_creatinine_labresult * | ทราบค่า Creatinine หรือไม่ - ทราบ - ไม่ทราบ | is_na_creatinine_la bresult | -0.3554 | 0.7009 |
| UPCR | ระดับโปรตีนใน ปัสสาวะ (mg/g) | upcr_labresult | 0.0036 | 1.0036 |

ตารางที่ 35 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ใน
การทำนายความเสี่ยงแทรกซ้อนทางตาในผู้ป่วยโรคเบาหวาน (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ใน ตัวแบบ | สัมประสิทธิ์ การถดถอย | Odds Ratio |
|-----------------------|--|------------------------------|--------------------------|------------|
| is_na_upcr_labresult* | ทราบค่า URCR หรือไม่ - ทราบ - ไม่ทราบ | is_na_upcr_ labresult | 0 | 1 |
| Medication | | | | |
| Lipid-lowering med | การใช้ยาลดไขมัน - ใช้ - ไม่ใช่ | flag_lipid_lowering | 0 | 1 |
| Non-insulin med | การใช้ยาเบาหวานที่ ไม่ต้องพึ่งอินซูลิน - ใช้ - ไม่ใช่ | flag_non_insulin_m ed | 0 | 1 |

ตัวแบบทำนายความเสี่ยงภาวะไตเสื่อมเรื้อรังในผู้ป่วยโรคเบาหวาน

จากกลุ่มตัวอย่างของผู้ป่วยโรคเบาหวานจำนวน 89,846 คน ซึ่งประกอบไปด้วยข้อมูลจำนวน 530,881 แถว จะถูกสุ่มแบ่งออกเป็นชุดข้อมูลฝึกฝนและชุดข้อมูลทดสอบคิดเป็นร้อยละ 70 และ 30 โดยประมาณ ตามลำดับ โดยที่ชุดข้อมูลฝึกฝนและชุดข้อมูลทดสอบจะมีสัดส่วนของผู้ป่วยโรคเบาหวานที่พบภาวะไตเสื่อมเรื้อรังพอ ๆ กัน ซึ่งในที่นี้พบว่ามีความเท่ากับ 0.010424 หรือคือมีผู้ป่วยโรคเบาหวานคิดเป็นร้อยละ 1.0424 ที่มีภาวะไตเสื่อมเรื้อรังนั่นเอง จากการแบ่งข้อมูลจะได้ข้อมูลดังต่อไปนี้

- ชุดข้อมูลฝึกฝน ประกอบด้วยข้อมูลจากผู้เป็นโรคเบาหวานจำนวน 62,892 คน ซึ่งมีข้อมูลจำนวน 371,529 แถว
- ชุดข้อมูลทดสอบ ประกอบด้วยข้อมูลจากผู้เป็นโรคเบาหวานจำนวน 26,954 คน ซึ่งมีข้อมูลจำนวน 159,352 แถว

จากนั้นจึงสร้างตัวแบบจากข้อมูลชุดฝึกฝน ซึ่งจะพิจารณา 2 ตัวแบบ ได้แก่

1. การถดถอยโลจิสติกแบบ Elastic Net วิธีการนี้ขอเรียกว่า “EN”
2. วิธี XGBoost โดยวิธีการนี้ขอเรียกว่า “XGBoost”

หลังจากได้ตัวแบบทั้ง 2 แล้ว จึงนำมาทำนายข้อมูลในชุดทดสอบเพื่อวัดประสิทธิภาพการทำนายโดยพิจารณาจาก

1. AUC
2. Cumulative AUC

ตารางที่ 36 แสดงค่า AUC จาก 2 วิธี ซึ่งพบว่าวิธี XGBoost มีประสิทธิภาพในการทำนายความเสี่ยงการภาวะไตเสื่อมเรื้อรังในผู้ป่วยโรคเบาหวานได้ดีกว่าวิธีการถดถอยโลจิสติกแบบ Elastic Net แม้ว่าค่า AUC ที่ได้จากวิธี XGBoost จะดีกว่าวิธีการถดถอยโลจิสติกแบบ Elastic Net ไม่มาก แต่หากพิจารณาค่า Cumulative AUC ซึ่งวัดประสิทธิภาพการทำนายรายบุคคลผ่านตัววัดผล AUC ที่คำนวณด้วยความน่าจะเป็นสะสมของบุคคลตลอด Cohort (Cumulative Probability) กับสถานะอุบัติการณ์ปีสุดท้าย จะพบว่าวิธี XGBoost มีประสิทธิภาพเหนือกว่าวิธี Elastic Net และมีค่า AUC สูงกว่า 0.8 ซึ่งอยู่ในเกณฑ์ดี

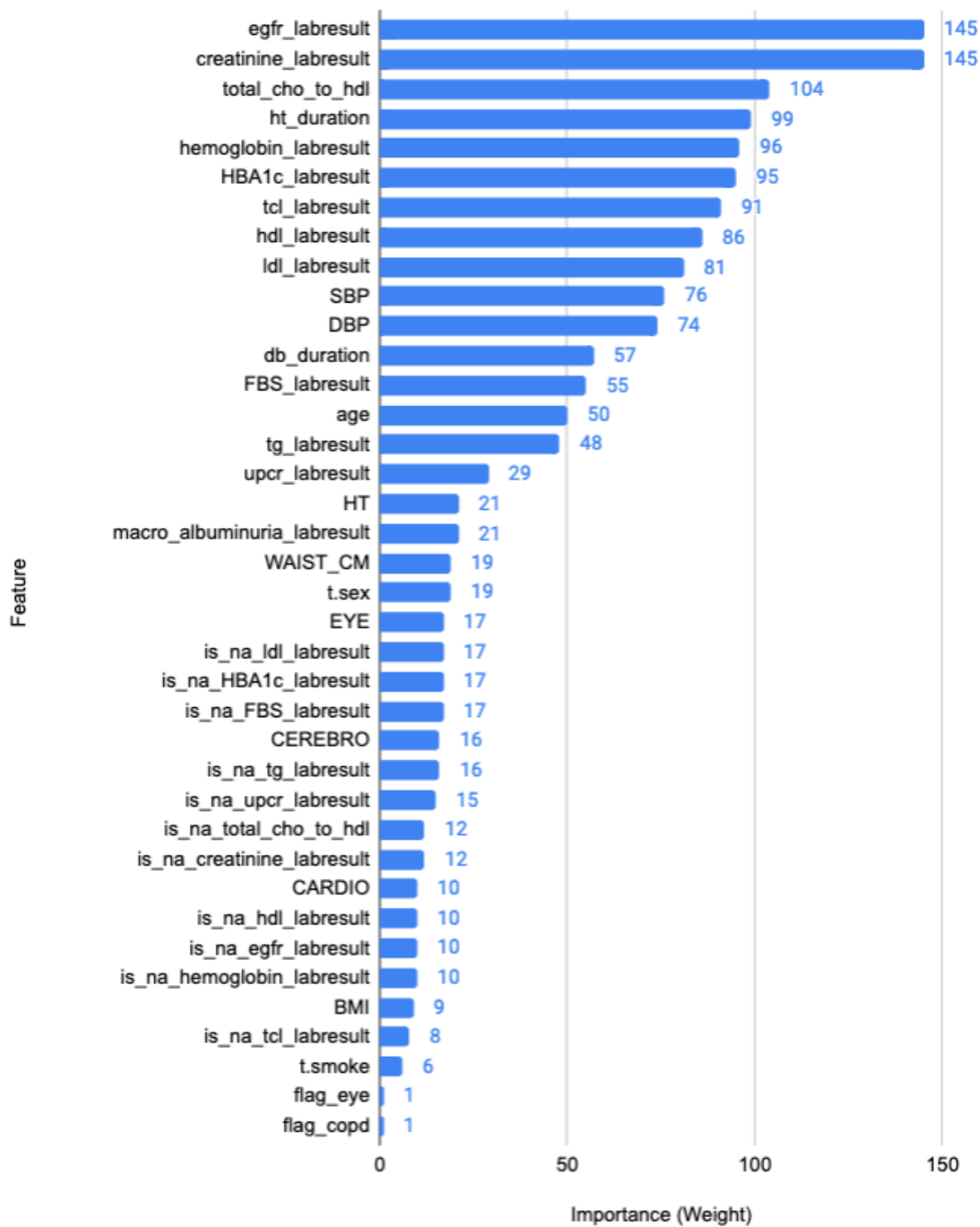
ตารางที่ 36 ค่า AUC เพื่อวัดประสิทธิภาพการทำนายความเสี่ยงภาวะไตเสื่อมเรื้อรังในผู้ป่วยโรคเบาหวาน (Cohort 1.2) (วิธีที่ดีที่สุดจะแสดงเป็นตัวหนา)

| ตัวแบบ | AUC | Cumulative AUC |
|----------------|---------------|----------------|
| EN | 0.7859 | 0.8038 |
| XGBoost | 0.8147 | 0.8373 |

สำหรับการแปลผลตัวแบบโดยอ้างอิงจากตัวแบบ XGBoost ที่ให้ประสิทธิภาพการทำนายดีที่สุดที่สุดสามารถพิจารณาได้ว่าปัจจัยใดที่มีความสำคัญในการทำนายความเสี่ยงการเกิดโรคเบาหวานในประชากรทั่วไป ดังแสดงในรูปที่ 56 รูปที่ 56 ซึ่งจะพบว่าตัวแปรที่มีความสำคัญในการทำนายความเสี่ยง 10 อันดับแรก ได้แก่

1. Estimated Glomerular Filtration Rate (mL/min/1.73m²) (egfr_labresult)
2. ระดับ Creatinine (mg/dL) (creatinine_labresult)
3. TC/HDL (total_cho_to_hdl)
4. ระยะเวลาที่เป็นความดันโลหิตสูง (ปี) (ht_duration)
5. ระดับฮีโมโกลบินในเลือด (g/dL) (hemoglobin_labresult)
6. ค่าน้ำตาลเฉลี่ยสะสมในเลือด (mmol/mol) (HBA1c_labresult)
7. ระดับคลอเรสเตอรอล (mg/dL) (tcl_labresult)
8. ระดับไขมันดี (mg/dL) (hdl_labresult)
9. ระดับไขมันไม่ดี (mg/dL) (ldl_labresult)

10. Systolic Blood Pressure (mm Hg) (SBP)

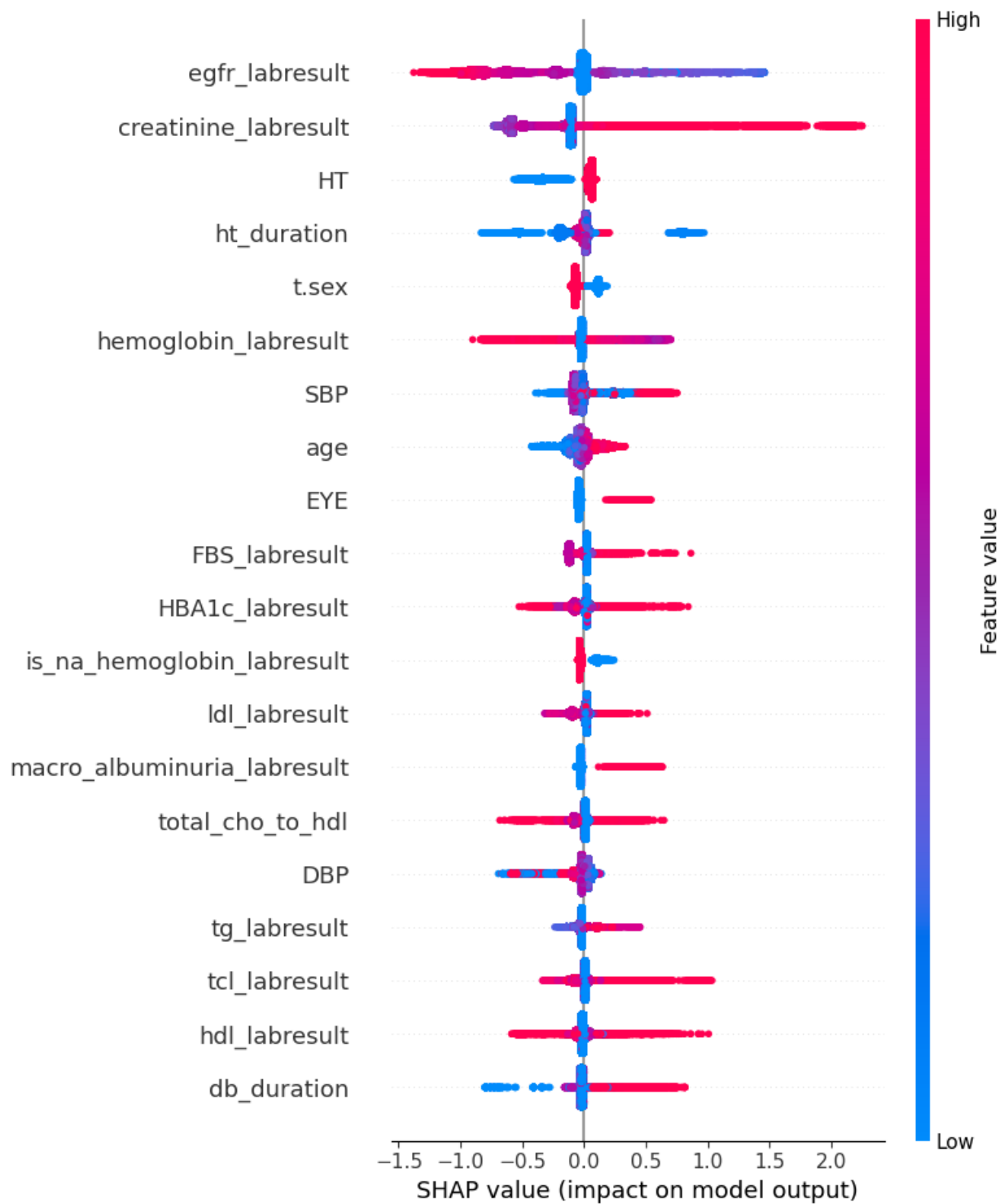


รูปที่ 56 ความสำคัญของตัวแปรทำนายจากวิธี XGBoost ในการทำนายภาวะไตเสื่อมเรื้อรังในผู้ป่วยโรคเบาหวาน

นอกจากนี้ หากพิจารณาคุณค่า SHAP ของตัวแปร 20 ตัวแปรที่มีความสำคัญมากที่สุดในการทำนายความเสี่ยงภาวะไตเสื่อมเรื้อรังในผู้ป่วยเบาหวาน ดังแสดงในรูปที่ 57 จะสามารถแปลผลในส่วนทิศทางความสัมพันธ์ของตัวแปรแต่ละตัว ซึ่งสามารถสรุปความสัมพันธ์ที่เด่นชัดได้ ดังนี้

- ผู้ป่วยเบาหวานที่มี Estimated Glomerular Filtration Rate (mL/min/1.73m²) (egfr_labresult) ต่ำ จะมีความเสี่ยงในการมีภาวะไตเสื่อมเรื้อรังสูงขึ้น
- ผู้ป่วยเบาหวานที่มีระดับ Creatinine (mg/dL) (creatinine_labresult) สูง จะมีแนวโน้มที่จะมีความเสี่ยงในการมีภาวะไตเสื่อมเรื้อรังสูงขึ้น
- ผู้ป่วยเบาหวานที่มีระยะเวลาที่เป็นความดันโลหิตสูง (ปี) (ht_duration) น้อย จะมีแนวโน้มที่จะมีความเสี่ยงในการมีภาวะไตเสื่อมเรื้อรังน้อยลง
- ผู้ป่วยเบาหวานที่มีค่า Systolic Blood Pressure (mm Hg) (SBP) สูง จะมีแนวโน้มที่จะมีความเสี่ยงในการมีภาวะไตเสื่อมเรื้อรังสูงขึ้น

แม้ว่าวิธี XGBoost สามารถระบุความสำคัญของตัวแปรทำนายที่ใช้ในตัวแบบได้ แต่ไม่สามารถแปลผลตัวแบบในรูปของอัตราเสี่ยงได้ ในขณะที่วิธีการถดถอยโลจิสติกแบบ Elastic Net ให้ค่าสัมประสิทธิ์การถดถอยเช่นเดียวกับตัวแบบการถดถอยโลจิสติกแบบปกติ ดังนั้นจึงสามารถแปลผลตัวแบบในรูปของอัตราเสี่ยงได้ ตารางที่ 37 แสดงผลลัพธ์ที่ได้ของสัมประสิทธิ์การถดถอยของตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ในการทำนายความเสี่ยงภาวะไตเสื่อมเรื้อรังในผู้ป่วยโรคเบาหวาน



รูปที่ 57 ค่า SHAP จากตัวแบบ XGBoost ในการทำนายความเสี่ยงภาวะไตเสื่อมเรื้อรังในผู้ป่วยโรคเบาหวาน

ตารางที่ 37 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ใน
การทำนายความเสี่ยงไตเสื่อมเรื้อรังในผู้ป่วยโรคเบาหวาน

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ในตัวแบบ | สัมประสิทธิ์การถดถอย | Odds Ratio |
|-----------------------------|--|--------------------------|----------------------|------------|
| Demographic | | | | |
| Age | อายุ (ปี) | age | -0.0094 | 0.9907 |
| Gender | เพศ - ชาย - หญิง | t.sex | -0.5886 | 0.5551 |
| Health Behavior | | | | |
| Smoking | พฤติกรรมการสูบบุหรี่ - สูบ - ไม่สูบ | t.smoke | -0.0673 | 0.9350 |
| Physical measurement | | | | |
| BMI | ดัชนีมวลกาย | BMI | 0.0026 | 1.0026 |
| WAIST_CM | เส้นรอบเอว (เซนติเมตร) | WAIST_CM | -0.0209 | 0.9793 |
| SBP | Systolic Blood Pressure (mm Hg) | SBP | 0.0026 | 1.0026 |
| DBP | Diastolic Blood Pressure (mm Hg) | DBP | -0.0060 | 0.9940 |
| Underlying Disease | | | | |
| HT | สถานะโรคความดันโลหิตสูง - เป็น - ไม่เป็น | HT | 0.5482 | 1.7301 |
| HT Duration | ระยะเวลาที่เป็นความดันโลหิตสูง (ปี) | ht_duration | -0.0109 | 0.9892 |
| DLP | สถานะภาวะไขมันผิดปกติ - เป็น - ไม่เป็น | flag_dlp | -0.3601 | 0.6976 |

ตารางที่ 37 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ใน
การทำนายความเสี่ยงไตเสื่อมเรื้อรังในผู้ป่วยโรคเบาหวาน (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ใน ตัวแบบ | สัมประสิทธิ์ การถดถอย | Odds Ratio |
|---|---|------------------------------|--------------------------|------------|
| DM Duration | ระยะเวลาที่เป็น โรคเบาหวาน (ปี) | db_duration | 0.0171 | 1.0172 |
| Heart Disease (Cardiovascular Disease) | สถานะการเป็นภาวะเส้น เลือดหัวใจตีบหรือ กล้ามเนื้อหัวใจขาดเลือด - เป็น - ไม่เป็น | CARDIO | 0.2582 | 1.2946 |
| Heart Disease | สถานะการเป็นโรคหัวใจ ซึ่งรวมถึงการเป็น โรคหัวใจเต้นผิดปกติ - เป็น - ไม่เป็น | flag_HEART | -0.0834 | 0.9200 |
| Stroke | สถานะการเป็นโรคหลอดเลือด สมอง - เป็น - ไม่เป็น | CEREBRO | 0.2627 | 1.3005 |
| flag_stroke | สถานะการเป็นโรคหลอดเลือด สมอง พิจารณาจาก รหัสโรคที่วินิจฉัย - เป็น - ไม่เป็น | flag_stroke | -0.1983 | 0.8201 |
| Eye | สถานะการเป็นโรคตาอัน เนื่องมาจากเบาหวาน - เป็น - ไม่เป็น | EYE | 0.5966 | 1.8159 |
| flag_eye | สถานะการเป็นโรคตา พิจารณาจากรหัสโรคที่ วินิจฉัย - เป็น - ไม่เป็น | flag_eye | -0.0834 | 0.9200 |

ตารางที่ 37 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ใน
การทำนายความเสี่ยงไตเสื่อมเรื้อรังในผู้ป่วยโรคเบาหวาน (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ในตัว แบบ | สัมประสิทธิ์ การถดถอย | Odds Ratio |
|----------------------|--|------------------------------|--------------------------|------------|
| COPD | สถานะการเป็นโรคถุง ลมโป่งพอง - เป็น - ไม่เป็น | flag_copd | -0.0834 | 0.9200 |
| Lab | | | | |
| TC | ระดับคลอ เรสเตอรอล (mg/dL) | tcl_labresult | 0.0005 | 1.0005 |
| is_na_tcl_labresult* | ทราบค่า TC หรือไม่ - ทราบ - ไม่ทราบ | is_na_tcl_labresult | 0.0603 | 1.0622 |
| HDL | ระดับไขมันดี (mg/dL) | hdl_labresult | 0.0002 | 1.0002 |
| is_na_hdl_labresult* | ทราบค่า HDL หรือไม่ - ทราบ - ไม่ทราบ | is_na_hdl_labresult | -0.2934 | 0.7457 |
| LDL | ระดับไขมันไม่ดี (mg/dL) | ldl_labresult | 0.0029 | 1.0029 |
| is_na_ldl_labresult* | ทราบค่า HDL หรือไม่ - ทราบ - ไม่ทราบ | is_na_ldl_labresult | 0.5256 | 1.6915 |
| TG | ระดับไขมันไตรกลี เซอไรด์ (mg/dL) | tg_labresult | 0.0007 | 1.0007 |
| is_na_tg_labresult* | ทราบค่า TG หรือไม่ - ทราบ - ไม่ทราบ | is_na_tg_labresult | 0.1419 | 1.1525 |
| total Cho:HDL | TC/HDL | total_cho_to_hdl | 0.0503 | 1.0515 |

ตารางที่ 37 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ใน
การทำนายความเสี่ยงไตเสื่อมเรื้อรังในผู้ป่วยโรคเบาหวาน (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ในตัว แบบ | สัมประสิทธิ์ การถดถอย | Odds Ratio |
|-------------------------|--|---------------------------------|--------------------------|------------|
| is_na_total_cho_to_hdl* | ทราบค่า total Cho:HDL หรือไม่ - ทราบ - ไม่ทราบ | is_na_total_cho_to _hdl | 0.5495 | 1.7323 |
| HBA1c | ค่าน้ำตาลเฉลี่ยสะสม ในเลือด (mmol/mol) | HBA1c_labresult | 0.0138 | 1.0139 |
| is_na_HBA1C_labresult* | ทราบค่า HBA1c หรือไม่ - ทราบ - ไม่ทราบ | is_na_HBA1C_labre sult | 0.2281 | 1.2563 |
| FBS | Fasting Blood Sugar (mg/dL) | FBS_labresult | 0.0016 | 1.0016 |
| is_na_FBS_labresult* | ทราบค่า FBS หรือไม่ - ทราบ - ไม่ทราบ | is_na_FBS_labresult | 0.2842 | 1.3288 |
| Macro-albuminuria | สถานะความเสี่ยงการ เป็นโรคไตโดยวัดจาก Macro-albuminuria - เป็น - ไม่เป็น | macro_albuminuria _labresult | 0.5740 | 1.7754 |
| eGFR | Estimated Glomerular Filtration Rate (mL/min/1.73m ²) | egfr_labresult | -0.0809 | 0.9223 |
| is_na_egfr_labresult* | ทราบค่า eGFR หรือไม่ - ทราบ - ไม่ทราบ | is_na_egfr_labresul t | -2.8781 | 0.0562 |

ตารางที่ 37 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ใน
การทำนายความเสี่ยงไตเสื่อมเร็วขึ้นในผู้ป่วยโรคเบาหวาน (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ในตัวแบบ | สัมประสิทธิ์การถดถอย | Odds Ratio |
|-----------------------------|---|----------------------------|----------------------|------------|
| Creatinine | ระดับ Creatinine (mg/dL) | creatinine_labresult | -0.1420 | 0.8676 |
| is_na_creatinine_labresult* | ทราบค่า Creatinine หรือไม่ - ทราบ - ไม่ทราบ | is_na_creatinine_labresult | -2.7730 | 0.0625 |
| UPCR | ระดับโปรตีนในปัสสาวะ (mg/g) | upcr_labresult | 0.0023 | 1.0024 |
| is_na_upcr_labresult* | ทราบค่า URCR หรือไม่ - ทราบ - ไม่ทราบ | is_na_upcr_labresult | 0.1963 | 1.2169 |
| Hemoglobin | ระดับฮีโมโกลบินในเลือด (g/dL) | hemoglobin_labresult | -0.0364 | 0.9643 |
| is_na_hemoglobin_labresult* | ทราบค่า Hemoglobin หรือไม่ - ทราบ - ไม่ทราบ | is_na_hemoglobin_labresult | -0.6389 | 0.5278 |
| | | | | |
| Medication | | | | |
| Lipid-lowering med | การใช้อยาลดไขมัน - ใช้ - ไม่ใช่ | flag_lipid_lowering | -0.4707 | 0.6246 |

ตารางที่ 37 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ใน
การทำนายความเสี่ยงไตเสื่อมเรื้อรังในผู้ป่วยโรคเบาหวาน (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ในตัว แบบ | สัมประสิทธิ์ การถดถอย | Odds Ratio |
|-----------------|--|------------------------------|--------------------------|------------|
| Non-insulin med | การใช้จ่ายเบาหวานที่ ไม่ต้องพึ่งอินซูลิน - ใช้ - ไม่ใช่ | flag_non_insulin_med | -0.4707 | 0.6246 |
| Insulin med | การใช้จ่ายเบาหวานที่ ต้องพึ่งอินซูลิน - ใช้ - ไม่ใช่ | flag_insulin_med | -0.0047 | 0.9953 |
| Aspirin | การใช้จ่ายแอสไพริน - ใช้ - ไม่ใช่ | flag_aspirin_med | -0.2919 | 0.7468 |

ตัวแบบทำนายความเสี่ยงการเสียชีวิตในผู้ป่วยโรคเบาหวาน

จากกลุ่มตัวอย่างของผู้ป่วยโรคเบาหวานจำนวน 90,941 คน ซึ่งประกอบไปด้วยข้อมูลจำนวน 548,903 แถว จะถูกสุ่มแบ่งออกเป็นชุดข้อมูลฝึกฝนและชุดข้อมูลทดสอบคิดเป็นร้อยละ 70 และ 30 โดยประมาณ ตามลำดับ โดยที่ชุดข้อมูลฝึกฝนและชุดข้อมูลทดสอบจะมีสัดส่วนของผู้ป่วยโรคเบาหวานที่เสียชีวิตพอ ๆ กัน ซึ่งในที่นี้พบว่ามีค่าเท่ากับ 0.012312 หรือคือมีผู้ป่วยโรคเบาหวานคิดเป็นร้อยละ 1.2312 ที่เสียชีวิตนั่นเอง จากการแบ่งข้อมูลจะได้ข้อมูลดังต่อไปนี้

- ชุดข้อมูลฝึกฝน ประกอบด้วยข้อมูลจากผู้เป็นโรคเบาหวานจำนวน 63,658 คน ซึ่งมีข้อมูลจำนวน 383,861 แถว
- ชุดข้อมูลทดสอบ ประกอบด้วยข้อมูลจากผู้เป็นโรคเบาหวานจำนวน 27,283 คน ซึ่งมีข้อมูลจำนวน 165,042 แถว

จากนั้นจึงสร้างตัวแบบจากข้อมูลชุดฝึกฝน ซึ่งจะพิจารณา 2 ตัวแบบ ได้แก่

1. การถดถอยโลจิสติกแบบ Elastic Net วิธีการนี้ขอเรียกว่า “EN”
2. วิธี XGBoost โดยวิธีการนี้ขอเรียกว่า “XGBoost”

หลังจากได้ตัวแบบทั้ง 2 แล้ว จึงนำมาทำนายข้อมูลในชุดทดสอบเพื่อวัดประสิทธิภาพการทำนายโดยพิจารณาจาก

1. AUC
2. Cumulative AUC

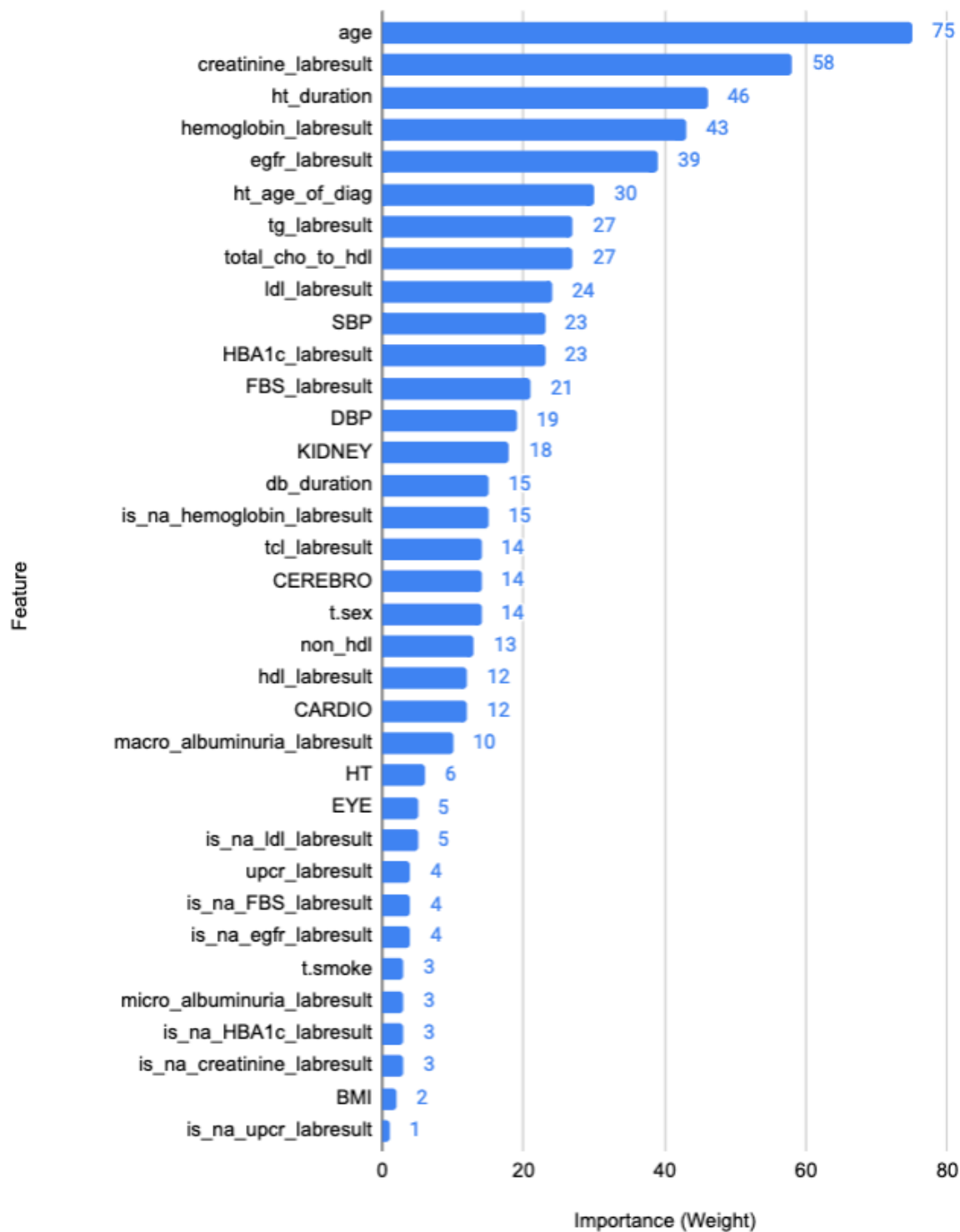
ตารางที่ 38 แสดงค่า AUC จาก 2 วิธี ซึ่งพบว่าวิธี XGBoost มีประสิทธิภาพในการทำนายความเสี่ยงการเสียชีวิตในผู้ป่วยโรคเบาหวานได้ดีกว่าวิธีการถดถอยโลจิสติกแบบ Elastic Net แม้ว่าค่า AUC ที่ได้จากวิธี XGBoost จะดีกว่าวิธีการถดถอยโลจิสติกแบบ Elastic Net ไม่มาก แต่หากพิจารณาค่า Cumulative AUC ซึ่งวัดประสิทธิภาพการทำนายรายบุคคลผ่านตัววัดผล AUC ที่คำนวณด้วยความน่าจะเป็นสะสมของบุคคลตลอด Cohort (Cumulative Probability) กับสถานะอุบัติการณ์ปีสุดท้าย จะพบว่าวิธี XGBoost มีประสิทธิภาพเหนือกว่าวิธี Elastic Net และมีค่า AUC สูงกว่า 0.8 ซึ่งอยู่ในเกณฑ์ดี

ตารางที่ 38 ค่า AUC เพื่อวัดประสิทธิภาพการทำนายความเสี่ยงการเสียชีวิตในผู้ป่วยโรคเบาหวาน (Cohort 1.2) (วิธีที่ดีที่สุดจะแสดงเป็นตัวหนา)

| ตัวแบบ | AUC | Cumulative AUC |
|----------------|---------------|----------------|
| EN | 0.7886 | 0.7157 |
| XGBoost | 0.8145 | 0.701 |

สำหรับการแปลผลตัวแบบโดยอ้างอิงจากตัวแบบ XGBoost ที่ให้ประสิทธิภาพการทำนายที่ดีที่สุดสามารถพิจารณาได้ว่าปัจจัยใดที่มีความสำคัญในการทำนายความเสี่ยงการเกิดโรคเบาหวานในประชากรทั่วไปดังแสดงในรูปที่ 58 ซึ่งจะพบว่าตัวแปรที่มีความสำคัญในการทำนายความเสี่ยง 10 อันดับแรก ได้แก่

1. อายุ (ปี) (age)
2. ระดับ Creatinine (mg/dL) (creatinine_labresult)
3. ระยะเวลาที่เป็นความดันโลหิตสูง (ปี) (ht_duration)
4. ระดับฮีโมโกลบินในเลือด (g/dL) (hemoglobin_labresult)
5. Estimated Glomerular Filtration Rate (mL/min/1.73m²) (egfr_labresult)
6. อายุที่ได้รับการวินิจฉัยว่าเป็นโรคความดันโลหิตสูง (ปี) (ht_age_of_diag)
7. ระดับไขมันไตรกลีเซอไรด์ (mg/dL) (tg_labresult)
8. TC/HDL (total_cho_to_hdl)
9. ระดับไขมันไม่ดี (mg/dL) (ldl_labresult)
10. Systolic Blood Pressure (mm Hg) (SBP)

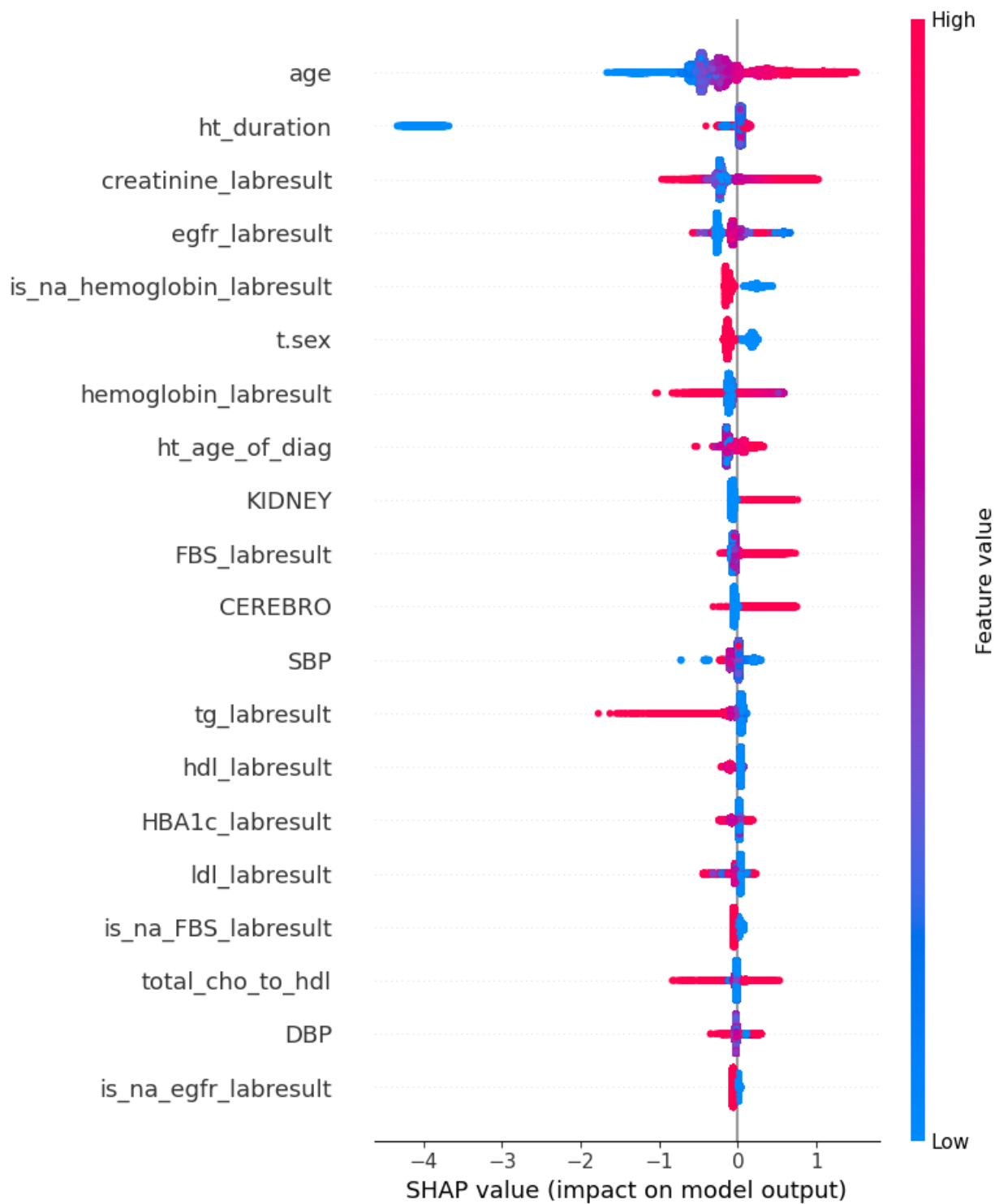


รูปที่ 58 ความสำคัญของตัวแปรทำนายจากวิธี XGBoost ในการทำนายการเสียชีวิตในผู้ป่วยโรคเบาหวาน

นอกจากนี้ หากพิจารณาคุณค่า SHAP ของตัวแปร 20 ตัวแปรที่มีความสำคัญมากที่สุดในการทำนายการเสียชีวิตในผู้ป่วยเบาหวาน ดังแสดงในรูปที่ 59 จะสามารถแปลผลในส่วนทิศทางความสัมพันธ์ของตัวแปรแต่ละตัว ซึ่งสามารถสรุปความสัมพันธ์ที่เด่นชัดได้ ดังนี้

- ผู้ป่วยเบาหวานที่มีอายุ (ปี) (age) สูง จะมีความเสี่ยงในการเสียชีวิตสูงขึ้น
- ผู้ป่วยเบาหวานที่มีระยะเวลาที่เป็นความดันโลหิตสูง (ปี) (ht_duration) น้อย จะมีแนวโน้มที่จะมีความเสี่ยงในการมีการเสียชีวิตน้อยลง
- ผู้ป่วยเบาหวานที่มีค่าระดับไขมันไตรกลีเซอไรด์ (mg/dL) สูง จะมีแนวโน้มที่จะมีความเสี่ยงในการมีการเสียชีวิตน้อยลง
- ผู้ป่วยที่มีเป็นโรคไตเรื้อรังในระยะสุดท้าย (พิจารณาจาก CKD และค่า eGFR) จะมีแนวโน้มที่จะมีความเสี่ยงในการมีการเสียชีวิตสูง

แม้ว่าวิธี XGBoost สามารถระบุความสำคัญของตัวแปรทำนายที่ใช้ในตัวแบบได้ แต่ไม่สามารถแปลผลตัวแบบในรูปของอัตราเสี่ยงได้ ในขณะที่วิธีการถดถอยโลจิสติกแบบ Elastic Net ให้ค่าสัมประสิทธิ์การถดถอยเช่นเดียวกับตัวแบบการถดถอยโลจิสติกแบบปกติ ดังนั้นจึงสามารถแปลผลตัวแบบในรูปของอัตราเสี่ยงได้ ตารางที่ 39 แสดงผลลัพธ์ที่ได้ของสัมประสิทธิ์การถดถอยของตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ในการทำนายความเสี่ยงการเสียชีวิตในผู้ป่วยโรคเบาหวาน



รูปที่ 59 ค่า SHAP จากตัวแบบ XGBoost ในการทำนายความเสี่ยงการเสียชีวิตในผู้ป่วยโรคเบาหวาน

ตารางที่ 39 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ใน
การทำนายความเสี่ยงการเสียชีวิตในผู้ป่วยโรคเบาหวาน

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ใน ตัวแบบ | สัมประสิทธิ์ การถดถอย | Odds Ratio |
|-----------------------------|---|------------------------------|--------------------------|------------|
| Demographic | | | | |
| Age | อายุ (ปี) | age | 0.0590 | 1.0608 |
| Gender | เพศ - ชาย - หญิง | t.sex | -0.4299 | 0.6506 |
| Health Behavior | | | | |
| Smoking | พฤติกรรมการสูบบุหรี่ - สูบ - ไม่สูบ | t.smoke | 0.1504 | 1.1623 |
| Physical measurement | | | | |
| BMI | ดัชนีมวลกาย | BMI | 0.0107 | 1.0107 |
| SBP | Systolic Blood Pressure (mm Hg) | SBP | -0.0042 | 0.9958 |
| DBP | Diastolic Blood Pressure (mm Hg) | DBP | 0.0038 | 1.0038 |
| Underlying Disease | | | | |
| HT | สถานะโรคความดันโลหิต สูง - เป็น - ไม่เป็น | HT | 0.2789 | 1.3216 |
| HT Duration | ระยะเวลาที่เป็นความดัน โลหิตสูง (ปี) | ht_duration | -0.0037 | 0.9963 |
| HT Age of Diag | อายุที่ได้รับการวินิจฉัยว่า เป็นโรคความดันโลหิตสูง (ปี) | ht_age_of_diag | -0.0006 | 0.994 |
| DLP | สถานะภาวะไขมัน ผิดปกติ - เป็น - ไม่เป็น | flag_dlp | 0 | 1 |

ตารางที่ 39 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ใน
การทำนายความเสี่ยงการเสียชีวิตในผู้ป่วยโรคเบาหวาน (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ใน ตัวแบบ | สัมประสิทธิ์ การถดถอย | Odds Ratio |
|---|---|------------------------------|--------------------------|------------|
| DM Duration | ระยะเวลาที่เป็น โรคเบาหวาน (ปี) | db_duration | -0.0031 | 0.9970 |
| Heart Disease (Cardiovascular Disease) | สถานะการเป็นภาวะเส้น เลือดหัวใจตีบหรือ กล้ามเนื้อหัวใจขาดเลือด - เป็น - ไม่เป็น | CARDIO | 0.6524 | 1.9202 |
| Heart Disease | สถานะการเป็นโรคหัวใจ ซึ่งรวมถึงการเป็น โรคหัวใจเต้นผิดปกติ - เป็น - ไม่เป็น | flag_HEART | 0 | 1 |
| CKD | สถานะการเป็นโรคไต เรื้อรังโดยดูจากรหัสโรคที่ วินิจฉัย - เป็น - ไม่เป็น | flag_CKD | 0 | 1 |
| KIDNEY | สถานะการเป็นโรคไต เรื้อรังในระยะสุดท้าย (พิจารณาจาก CKD และ ค่า eGFR) - เป็น - ไม่เป็น | KIDNEY | 0.8596 | 2.3621 |
| Stroke | สถานะการเป็นโรคหลอดเลือด สมอง - เป็น - ไม่เป็น | CEREBRO | 0.6717 | 1.9576 |

ตารางที่ 39 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ใน
การทำนายความเสี่ยงการเสียชีวิตในผู้ป่วยโรคเบาหวาน (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ใน ตัวแบบ | สัมประสิทธิ์ การถดถอย | Odds Ratio |
|----------------------|---|------------------------------|--------------------------|------------|
| flag_stroke | สถานะการเป็นโรค หลอดเลือดสมอง พิจารณาจากรหัส โรคที่วินิจฉัย - เป็น - ไม่เป็น | flag_stroke | 0 | 1 |
| Eye | สถานะการเป็นโรค ตาอันเนื่องมาจาก เบาหวาน - เป็น - ไม่เป็น | EYE | 0.2117 | 1.2358 |
| flag_eye | สถานะการเป็นโรค ตา พิจารณาจาก รหัสโรคที่วินิจฉัย - เป็น - ไม่เป็น | flag_eye | 0 | 1 |
| COPD | สถานะการเป็นโรค ถุงลมโป่งพอง - เป็น - ไม่เป็น | flag_copd | 0 | 1 |
| Lab | | | | |
| TC | ระดับคลอ เรสเตอรอล (mg/dL) | tcl_labresult | 0.0004 | 1.0004 |
| is_na_tcl_labresult* | ทราบค่า TC หรือไม่ - ทราบ - ไม่ทราบ | is_na_tcl_labresu lt | 0 | 1 |
| HDL | ระดับไขมันดี (mg/dL) | hdl_labresult | -0.0006 | 0.9994 |

ตารางที่ 39 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ใน
การทำนายความเสี่ยงการเสียชีวิตในผู้ป่วยโรคเบาหวาน (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ใน ตัวแบบ | สัมประสิทธิ์ การถดถอย | Odds Ratio |
|-------------------------|---|------------------------------|--------------------------|------------|
| is_na_hdl_labresult* | ทราบค่า HDL หรือไม่ - ทราบ - ไม่ทราบ | is_na_hdl_labres ult | 0 | 1 |
| LDL | ระดับไขมันไม่ดี (mg/dL) | ldl_labresult | 0.0011 | 1.0011 |
| is_na_ldl_labresult* | ทราบค่า HDL หรือไม่ - ทราบ - ไม่ทราบ | is_na_ldl_labres ult | 0.4627 | 1.5884 |
| TG | ระดับไขมันไตรกลี เซอไรด์ (mg/dL) | tg_labresult | -0.0010 | 0.9990 |
| is_na_tg_labresult* | ทราบค่า TG หรือไม่ - ทราบ - ไม่ทราบ | is_na_tg_labresul t | -0.0020 | 0.9980 |
| non-HDL | TC – HDL | non_hdl | -0.0022 | 0.9978 |
| is_na_non_hdl* | ทราบค่า non-HDL หรือไม่ - ทราบ - ไม่ทราบ | is_na_non_hdl | 0.2410 | 1.2726 |
| total Cho:HDL | TC/HDL | total_cho_to_hd l | 0.1257 | 1.1339 |
| is_na_total_cho_to_hdl* | ทราบค่า total Cho:HDL หรือไม่ - ทราบ - ไม่ทราบ | is_na_total_cho_ to_hdl | 0 | 1 |

ตารางที่ 39 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ใน
การทำนายความเสี่ยงการเสียชีวิตในผู้ป่วยโรคเบาหวาน (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ในตัวแบบ | สัมประสิทธิ์การถดถอย | Odds Ratio |
|------------------------|---|-----------------------------|----------------------|------------|
| HBA1c | ค่าน้ำตาลเฉลี่ย สะสมในเลือด (mmol/mol) | HBA1c_labresult | 0.0100 | 1.0100 |
| is_na_HBA1C_labresult* | ทราบค่า HBA1c หรือไม่ - ทราบ - ไม่ทราบ | is_na_HBA1C_labresult | 0.2769 | 1.3191 |
| FBS | Fasting Blood Sugar (mg/dL) | FBS_labresult | 0.0031 | 1.0031 |
| is_na_FBS_labresult* | ทราบค่า FBS หรือไม่ - ทราบ - ไม่ทราบ | is_na_FBS_labresult | 0.2305 | 1.2592 |
| Macro-albuminuria | สถานะความเสี่ยง การเป็นโรคไตโดย วัดจาก Macro- albuminuria - เป็น - ไม่เป็น | macro_albuminuria_labresult | 0.2939 | 1.3416 |
| Micro-albuminuria | สถานะความเสี่ยง การเป็นโรคไตโดย วัดจาก Micro- albuminuria - เป็น - ไม่เป็น | micro_albuminuria_labresult | 0.0019 | 1.0019 |
| eGFR | Estimated Glomerular Filtration Rate (mL/min/1.73m ²) | egfr_labresult | -0.0087 | 0.9913 |

ตารางที่ 39 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ใน
การทำนายความเสี่ยงการเสียชีวิตในผู้ป่วยโรคเบาหวาน (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ใน ตัวแบบ | สัมประสิทธิ์ การถดถอย | Odds Ratio |
|-----------------------------|---|--------------------------------|--------------------------|------------|
| is_na_egfr_labresult* | ทราบค่า eGFR หรือไม่ - ทราบ - ไม่ทราบ | is_na_egfr_labres ult | -0.7356 | 0.4792 |
| Creatinine | ระดับ Creatinine (mg/dL) | creatinine_labres ult | 0.0189 | 1.0191 |
| is_na_creatinine_labresult* | ทราบค่า Creatinine หรือไม่ - ทราบ - ไม่ทราบ | is_na_creatinine_ labresult | -0.7326 | 0.4806 |
| UPCR | ระดับโปรตีนใน ปัสสาวะ (mg/g) | upcr_labresult | 0.0023 | 1.0023 |
| is_na_upcr_labresult* | ทราบค่า URCR หรือไม่ - ทราบ - ไม่ทราบ | is_na_upcr_labre sult | 0.1852 | 1.2035 |
| Hemoglobin | ระดับฮีโมโกลบินใน เลือด (g/dL) | hemoglobin_labr esult | -0.0619 | 0.9400 |
| is_na_hemoglobin_labresult* | ทราบค่า Hemoglobin หรือไม่ - ทราบ - ไม่ทราบ | is_na_hemoglobi n_labresult | -1.3572 | 0.2574 |

ตารางที่ 39 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ใน
การทำนายความเสี่ยงการเสียชีวิตในผู้ป่วยโรคเบาหวาน (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ในตัวแบบ | สัมประสิทธิ์การถดถอย | Odds Ratio |
|--------------------|--|--------------------------|----------------------|------------|
| Medication | | | | |
| Lipid-lowering med | การใช้ยาลดไขมัน - ใช้ - ไม่ใช่ | flag_lipid_lowering | 0 | 1 |
| Non-insulin med | การใช้ยาเบาหวานที่ไม่ต้องพึ่งอินซูลิน - ใช้ - ไม่ใช่ | flag_non_insulin_med | 0 | 1 |
| Insulin med | การใช้ยาเบาหวานที่ต้องพึ่งอินซูลิน - ใช้ - ไม่ใช่ | flag_insulin_med | 0 | 1 |
| Aspirin | การใช้ยาแอสไพริน - ใช้ - ไม่ใช่ | flag_aspirin_med | 0 | 1 |

แบบจำลองที่ 4 ทำนายความเสี่ยงการเกิดภาวะแทรกซ้อนและการเสียชีวิตจาก Cohort ของผู้ที่เป็นโรคความดันโลหิตสูง (Cohort 2.2)

กลุ่มโรคภาวะแทรกซ้อนและการเตรียมตัวแปรสถานะอุบัติการณ์โรคแทรกซ้อน

กลุ่มโรคภาวะแทรกซ้อนที่สำคัญในผู้ป่วยที่เป็นโรคความดันโลหิตสูง จะเป็นกลุ่มโรคภาวะแทรกซ้อนเช่นเดียวกันกับผู้ป่วยที่เป็นโรคเบาหวาน ซึ่งได้แก่

1. ภาวะเส้นเลือดหัวใจตีบหรือกล้ามเนื้อหัวใจขาดเลือด (Cardiovascular Disease)
2. ภาวะเส้นเลือดสมองแตกหรือตีบ (Cerebrovascular Disease)
3. ภาวะแทรกซ้อนทางตาอันเนื่องมาจากโรคเบาหวานและความดันโลหิตสูง
4. ภาวะไตเสื่อมเรื้อรัง

5. ภาวะแทรกซ้อนทางเส้นประสาทและเส้นเลือดส่วนปลายอันจำเป็นต้องเกิดการตัดรยางค์ (Amputation)

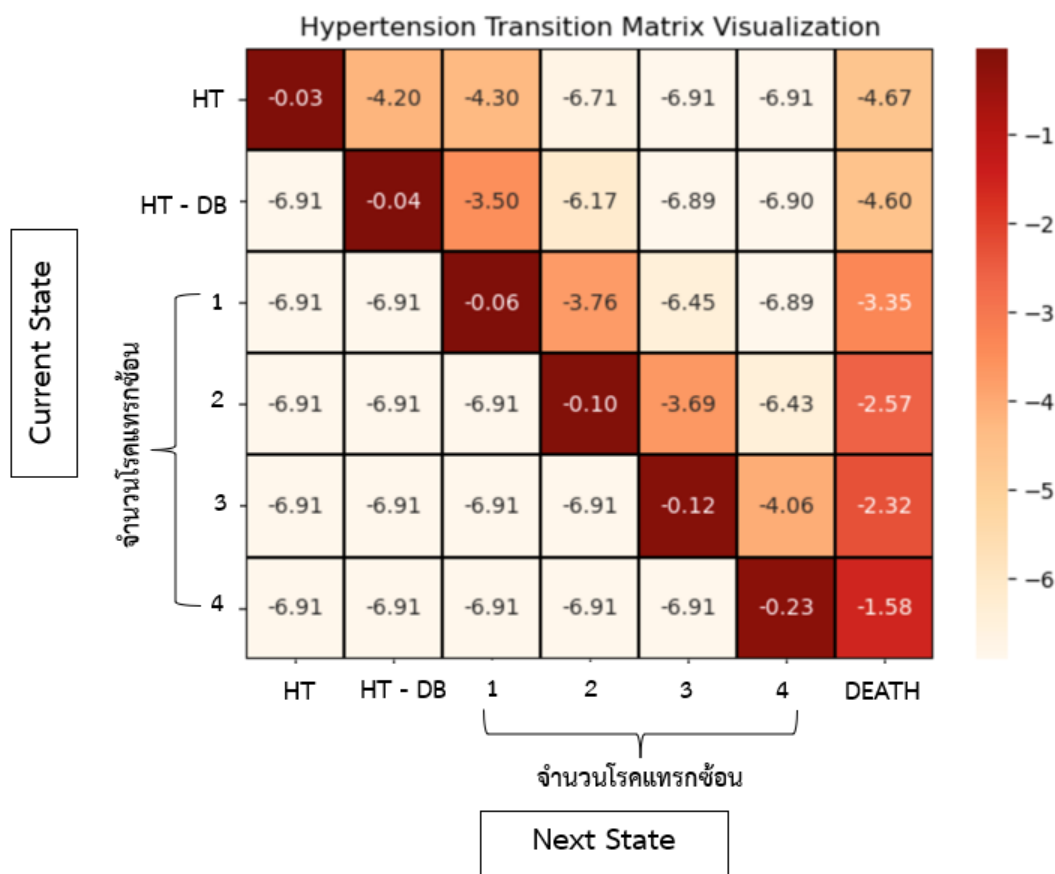
ดังนั้น ในการเตรียมตัวแปรสถานะอุบัติการณ์โรคแทรกซ้อน 5 กลุ่มโรคจึงพิจารณาจากรหัส ICD10 เช่นเดียวกับที่ได้อธิบายไว้ในเนื้อหาด้านบน จากการเตรียมสถานะอุบัติการณ์ภาวะโรคแทรกซ้อน ไม่พบผู้ป่วยที่มีภาวะแทรกซ้อนทางเส้นประสาทและเส้นเลือดส่วนปลายอันจำเป็นต้องเกิดการตัดรยางค์ (Amputation) ดังนั้นในการสร้างตัวแบบจำลองทำนายความเสี่ยงภาวะแทรกซ้อนและการเสียชีวิต จึงพิจารณา 5 ตัวแบบ ดังนี้

1. ตัวแบบทำนายความเสี่ยงภาวะเส้นเลือดหัวใจตีบหรือกล้ามเนื้อหัวใจขาดเลือดในผู้ป่วยโรคความดันโลหิตสูง
2. ตัวแบบทำนายความเสี่ยงภาวะเส้นเลือดสมองแตกหรือตีบในผู้ป่วยโรคความดันโลหิตสูง
3. ตัวแบบทำนายความเสี่ยงภาวะแทรกซ้อนทางตาในผู้ป่วยโรคความดันโลหิตสูง
4. ตัวแบบทำนายความเสี่ยงภาวะไตเสื่อมเรื้อรังในผู้ป่วยโรคความดันโลหิตสูง
5. ตัวแบบทำนายความเสี่ยงการเสียชีวิตในผู้ป่วยโรคความดันโลหิตสูง

การวิเคราะห์ข้อมูลเชิงสำรวจ

ในการวิเคราะห์ข้อมูลเชิงสำรวจ รูปที่ 60 แสดงเมทริกซ์เปลี่ยนสถานะ (Transition Matrix) ซึ่งแสดงค่าลอการิทึมของสัดส่วนการเปลี่ยนสถานะของอุบัติการณ์โรคในผู้ป่วยจากปีหนึ่งไปยังปีต่อไป เนื่องจากจากการสำรวจเบื้องต้นพบว่าผู้ป่วยที่เป็นโรคความดันโลหิตสูงจำนวนหนึ่งจะมีภาวะของโรคเบาหวานร่วมด้วย ดังนั้นในการวิเคราะห์ นอกจากจะพิจารณาสถานะของโรคความดันโลหิตสูงเพียงอย่างเดียวแล้ว (HT) ยังพิจารณาสถานะของโรคความดันโลหิตสูงร่วมกับภาวะโรคเบาหวานอีกด้วย (HT-DB) สำหรับกลุ่มโรคแทรกซ้อนอื่นๆ 4 กลุ่ม ได้แก่ 1) ภาวะเส้นเลือดหัวใจตีบหรือกล้ามเนื้อหัวใจขาดเลือด (Cardiovascular Disease) 2) ภาวะเส้นเลือดสมองแตกหรือตีบ (Cerebrovascular Disease) 3) ภาวะแทรกซ้อนทางตาอันเนื่องมาจากโรคเบาหวานและความดันโลหิตสูง และ 4) ภาวะไตเสื่อมเรื้อรัง จะพิจารณาโดยนับจำนวนโรคแทรกซ้อนที่เกิดขึ้นแทน

จากรูปที่ 60 โดยภาพรวม พบว่าส่วนใหญ่ผู้ป่วยจะไม่เปลี่ยนสถานะหรืออยู่สถานะเดิมในปีถัดไป อย่างไรก็ตาม ยังพบว่าผู้ป่วยที่เป็นโรคความดันโลหิตสูงเพียงอย่างเดียวมีสัดส่วนถึง $e^{-4.20} \times 100\% = 1.50\%$ ที่ในปีถัดไปจะเป็นโรคเบาหวานร่วมด้วย สำหรับโอกาสเกิดโรคแทรกซ้อนในปีถัดไป โดยภาพรวม พบว่าผู้ป่วยความดันโลหิตสูงที่ในปีถัดไปมีโรคแทรกซ้อนเพิ่มขึ้นส่วนใหญ่มักจะเพิ่มปีละ 1 โรค นอกจากนี้ ยังเห็นได้ชัดเจนว่าผู้ป่วยความดันโลหิตสูงยังมีจำนวนโรคแทรกซ้อนมากขึ้น โอกาสในการเสียชีวิตก็จะเพิ่มขึ้นตามไปด้วย



รูปที่ 60 เมทริกซ์เปลี่ยนสถานะรายปีของผู้ป่วยที่เป็นโรคความดันโลหิตสูงในมาตราส่วนเชิงลอการิทึม

ตัวแปรทำนายที่ใช้ในแบบทำนายความเสี่ยงภาวะแทรกซ้อนและการเสียชีวิต

ในการสร้างแบบทำนายความเสี่ยงภาวะแทรกซ้อน ทีมผู้วิจัยจะสร้างแบบทำนายทั้งหมด 5 ตัวแบบ เช่นเดียวกับแบบจำลองของผู้ป่วยโรคเบาหวานโดยใช้รายการตัวแปรจากการทบทวนวรรณกรรมที่ได้สรุปไว้ในตารางที่ 29 ที่ได้แสดงไว้แล้วก่อนหน้านี้

ผลลัพธ์จากแบบทำนายความเสี่ยงการเกิดภาวะแทรกซ้อนและการเสียชีวิตในผู้ป่วยที่เป็นโรคความดันโลหิตสูง (Cohort 2.2)

แบบทำนายความเสี่ยงภาวะเส้นเลือดหัวใจตีบหรือกล้ามเนื้อหัวใจขาดเลือดในผู้ป่วยโรคความดันโลหิตสูง

จากกลุ่มตัวอย่างของผู้ป่วยโรคความดันโลหิตสูงจำนวน 235,365 คน ซึ่งประกอบไปด้วยข้อมูลจำนวน 1,498,669 แถว จะถูกสุ่มแบ่งออกเป็นชุดข้อมูลฝึกฝนและชุดข้อมูลทดสอบคิดเป็นร้อยละ 70 และ 30 โดยประมาณ ตามลำดับ โดยที่ชุดข้อมูลฝึกฝนและชุดข้อมูลทดสอบจะมีสัดส่วนของผู้ป่วยโรคความดันโลหิตสูงที่พบภาวะเส้นเลือดหัวใจตีบหรือกล้ามเนื้อหัวใจขาดเลือดพอ ๆ กัน ซึ่งในที่นี้พบว่ามีความเท่ากับ 0.002561 หรือคือมีผู้ป่วยโรคความดันโลหิตสูงคิดเป็นร้อยละ 0.2561 ที่มีภาวะเส้นเลือดหัวใจตีบหรือกล้ามเนื้อหัวใจขาดเลือดนั่นเอง จากการแบ่งข้อมูลจะได้ข้อมูลดังต่อไปนี้

- ชุดข้อมูลฝึกฝน ประกอบด้วยข้อมูลจากผู้เป็นโรคความดันโลหิตสูงจำนวน 164,756 คน ซึ่งมีข้อมูลจำนวน 1,048,807 แถว
- ชุดข้อมูลทดสอบ ประกอบด้วยข้อมูลจากผู้เป็นโรคความดันโลหิตสูงจำนวน 70,609 คน ซึ่งมีข้อมูลจำนวน 449,862 แถว

จากนั้นจึงสร้างตัวแบบจากข้อมูลชุดฝึกฝน ซึ่งจะพิจารณา 2 ตัวแบบ ได้แก่

1. การถดถอยโลจิสติกแบบ Elastic Net วิธีการนี้ขอเรียกว่า “EN”
2. วิธี XGBoost โดยวิธีการนี้ขอเรียกว่า “XGBoost”

หลังจากได้ตัวแบบทั้ง 2 แล้ว จึงนำมาทำนายข้อมูลในชุดทดสอบเพื่อวัดประสิทธิภาพการทำนาย โดยพิจารณาจาก

1. AUC
2. Cumulative AUC

ตารางที่ 40 แสดงค่า AUC จาก 2 วิธี ซึ่งพบว่าวิธี XGBoost มีประสิทธิภาพในการทำนายความเสี่ยงการภาวะเส้นเลือดหัวใจตีบหรือกล้ามเนื้อหัวใจขาดเลือดในผู้ป่วยโรคความดันโลหิตสูงได้ดีกว่าวิธีการถดถอยโลจิสติกแบบ Elastic Net แม้ว่าค่า AUC ที่ได้จากวิธี XGBoost จะดีกว่าวิธีการถดถอยโลจิสติกแบบ Elastic Net ไม่มาก แต่หากพิจารณาค่า Cumulative AUC ซึ่งวัดประสิทธิภาพการทำนายรายบุคคลผ่านตัววัดผล AUC ที่คำนวณด้วยความน่าจะเป็นสะสมของบุคคลตลอด Cohort (Cumulative Probability) กับสถานะอุบัติการณ์ปีสุดท้าย จะพบว่าวิธี XGBoost มีประสิทธิภาพเหนือกว่าวิธี Elastic Net ค่อนข้างมาก อย่างไรก็ตาม พบว่าวิธีทั้งสองอาจจะยังมีค่า AUC ที่ยังอยู่ในเกณฑ์ไม่สูง หรือมีค่าต่ำกว่า 0.8 ซึ่งในอนาคตควรจะต้องพิจารณาตัวแบบอื่นๆ รวมทั้งพิจารณาตัวแปรหรือใช้ขนาดตัวอย่างเพิ่มเติม เพื่อที่จะปรับปรุงตัวแบบในการทำนายให้มีประสิทธิภาพมากขึ้น

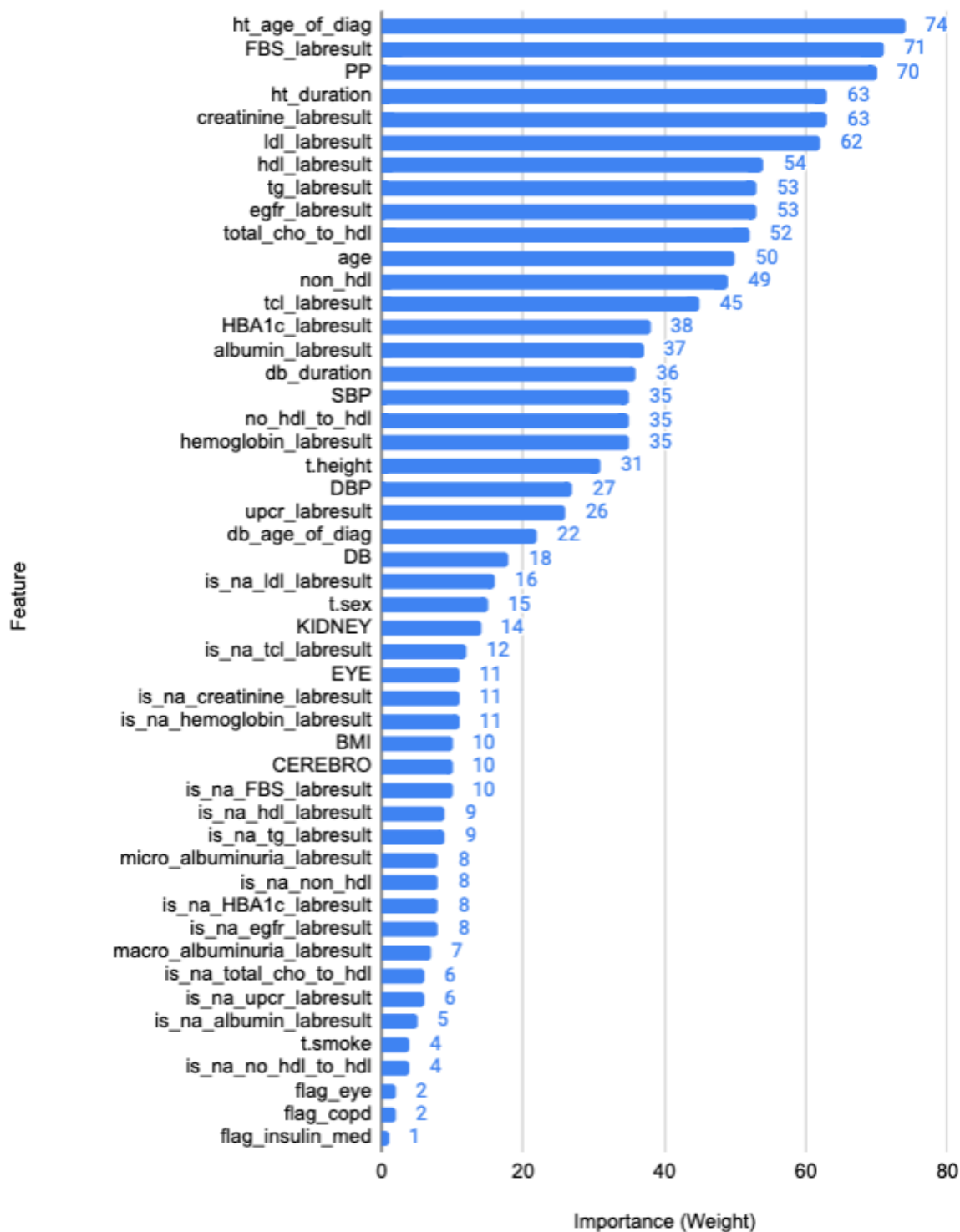
ตารางที่ 40 ค่า AUC เพื่อวัดประสิทธิภาพการทำนายความเสี่ยงภาวะเส้นเลือดหัวใจตีบหรือกล้ามเนื้อหัวใจขาดเลือดในผู้ป่วยโรคความดันโลหิตสูง (Cohort 2.2) (วิธีที่ดีที่สุดจะแสดงเป็นตัวหนา)

| ตัวแบบ | AUC | Cumulative AUC |
|----------------|--------------|----------------|
| EN | 0.6557 | 0.6217 |
| XGBoost | 0.663 | 0.6402 |

สำหรับการแปลผลตัวแบบโดยอ้างอิงจากตัวแบบ XGBoost ที่ให้ประสิทธิภาพการทำนายที่ดีที่สุดสามารถพิจารณาได้ว่าปัจจัยใดที่มีความสำคัญในการทำนายความเสี่ยงการเกิดโรคความดันโลหิตสูงในประชากรทั่วไปดังแสดงในรูปที่ 61 ซึ่งจะพบว่าตัวแปรที่มีความสำคัญในการทำนายความเสี่ยง 10 อันดับแรก ได้แก่

1. อายุที่ได้รับการวินิจฉัยว่าเป็นโรคความดันโลหิตสูง (ปี) (ht_age_of_diag)
2. Fasting Blood Sugar (mg/dL) (FBS_labresult)
3. Pulse Pressure คำนวณจาก SBP – DBP (PP)

4. อายุที่ได้รับการวินิจฉัยว่าเป็นโรคความดันโลหิตสูง (ปี) (ht_age_of_diag)
5. ระยะเวลาที่เป็นความดันโลหิตสูง (ปี) (ht_duration)
6. ระดับ Creatinine (mg/dL) (creatinine_labresult)
7. ระดับไขมันไม่ดี (mg/dL) (ldl_labresult)
8. ระดับไขมันดี (mg/dL) (hdl_labresult)
9. ระดับไขมันไตรกลีเซอไรด์ (mg/dL) (tg_labresult)
10. Estimated Glomerular Filtration Rate (mL/min/1.73m²) (egfr_labresult)

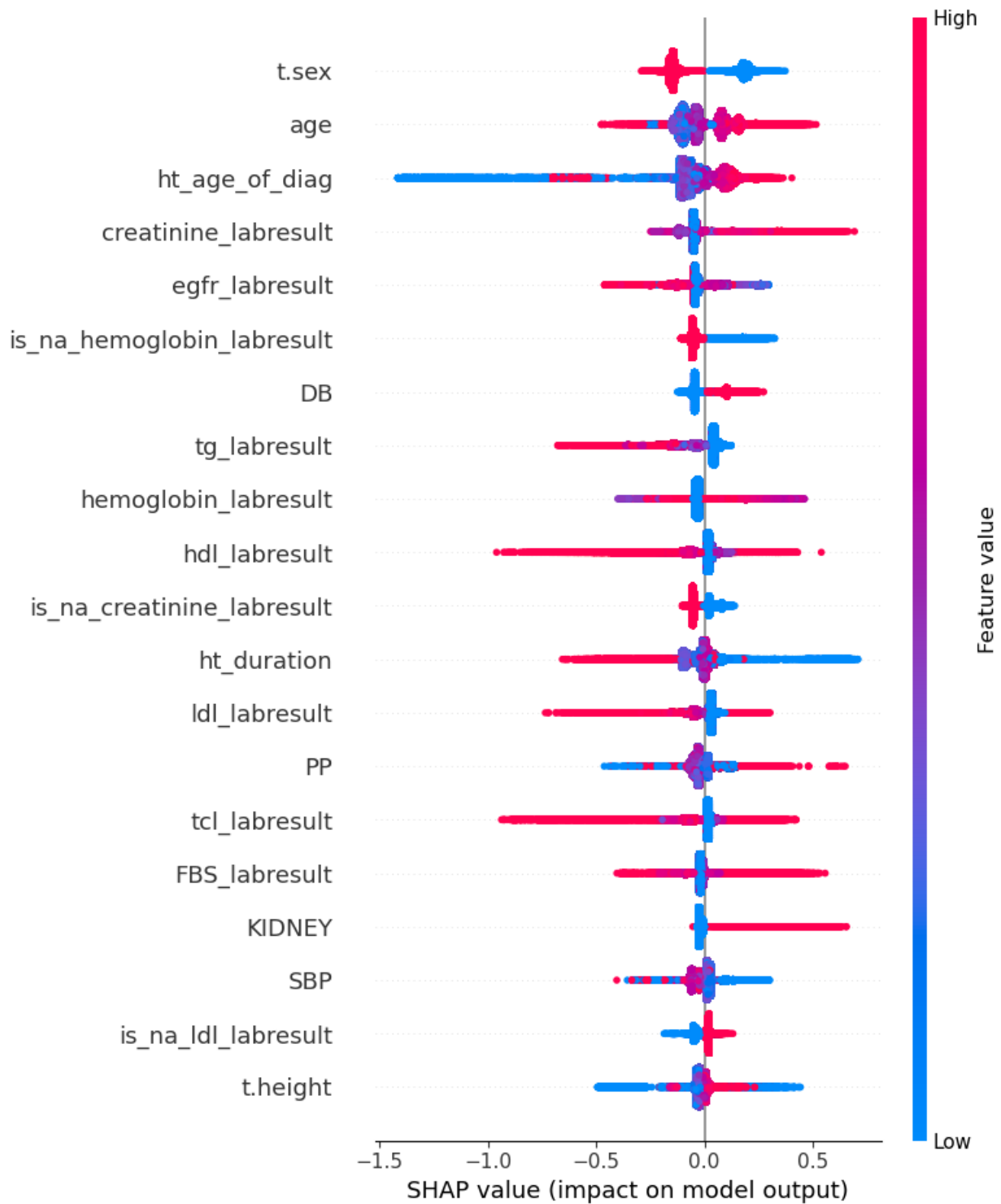


รูปที่ 61 ความสำคัญของตัวแปรทำนายจากวิธี XGBoost ในการทำนายความเสี่ยงภาวะเส้นเลือดหัวใจตีบหรือกล้ามเนื้อหัวใจขาดเลือดในผู้ป่วยโรคความดันโลหิตสูง

นอกจากนี้ หากพิจารณาคุณค่า SHAP ของตัวแปร 20 ตัวแปรที่มีความสำคัญมากที่สุดในการทำนายความเสี่ยงภาวะเส้นเลือดหัวใจตีบหรือกล้ามเนื้อหัวใจขาดเลือดในผู้ป่วยโรคความดันโลหิตสูง ดังแสดงในรูปที่ 62 จะสามารถแปลผลในส่วนทิศทางการความสัมพันธ์ของตัวแปรแต่ละตัว ซึ่งสามารถสรุปความสัมพันธ์ที่เด่นชัดได้ ดังนี้

- ผู้ป่วยความดันโลหิตสูงที่ได้รับการวินิจฉัยว่าเป็นโรคความดันโลหิตสูงตอนช่วงอายุน้อย (ht_age_of_diag) จะมีแนวโน้มที่จะมีความเสี่ยงในการมีภาวะเส้นเลือดหัวใจตีบหรือกล้ามเนื้อหัวใจขาดเลือดสูงต่ำ
- ผู้ป่วยความดันโลหิตสูงที่มีระดับ Creatinine (mg/dL) (creatinine_labresult) สูง จะมีแนวโน้มที่จะมีความเสี่ยงในการมีภาวะเส้นเลือดหัวใจตีบหรือกล้ามเนื้อหัวใจขาดเลือดสูงขึ้น
- ผู้ป่วยความดันโลหิตสูงที่มีระดับ Estimated Glomerular Filtration Rate (mL/min/1.73m²) (egfr_labresult) สูง จะมีแนวโน้มที่จะมีความเสี่ยงในการมีภาวะเส้นเลือดหัวใจตีบหรือกล้ามเนื้อหัวใจขาดเลือดต่ำลง
- ผู้ป่วยความดันโลหิตสูงที่มีระดับไขมันดี (mg/dL) (hdl_labresult) สูง จะมีแนวโน้มที่จะมีความเสี่ยงในการมีภาวะเส้นเลือดหัวใจตีบหรือกล้ามเนื้อหัวใจขาดเลือดลดลง

แม้ว่าวิธี XGBoost สามารถระบุความสำคัญของตัวแปรทำนายที่ใช้ในตัวแบบได้ แต่ไม่สามารถแปลผลตัวแบบในรูปของอัตราเสี่ยงได้ ในขณะที่วิธีการถดถอยโลจิสติกแบบ Elastic Net ให้ค่าสัมประสิทธิ์การถดถอย เช่นเดียวกับตัวแบบการถดถอยโลจิสติกแบบปกติ ดังนั้นจึงสามารถแปลผลตัวแบบในรูปของอัตราเสี่ยงได้ตารางที่ 41 แสดงผลลัพธ์ที่ได้ของสัมประสิทธิ์การถดถอยของตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ในการทำนายความเสี่ยงภาวะเส้นเลือดหัวใจตีบหรือกล้ามเนื้อหัวใจขาดเลือดในผู้ป่วยโรคความดันโลหิตสูง



รูปที่ 62 ค่า SHAP จากตัวแบบ XGBoost ในการทำนายความเสี่ยงภาวะเส้นเลือดหัวใจตีบหรือกล้ามเนื้อหัวใจขาดเลือดในผู้ป่วยโรคความดันโลหิตสูง

ตารางที่ 41 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ใน
การทำนายความเสี่ยงภาวะเส้นเลือดหัวใจตีบหรือกล้ามเนื้อหัวใจขาดเลือดในผู้ป่วยโรคความดันโลหิตสูง

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ใน ตัวแบบ | สัมประสิทธิ์ การถดถอย | Odds Ratio |
|-----------------------------|--|------------------------------|--------------------------|------------|
| Demographic | | | | |
| Age | อายุ (ปี) | age | -0.0050 | 0.9950 |
| Gender | เพศ - ชาย - หญิง | t.sex | -0.4458 | 0.6403 |
| Health Behavior | | | | |
| Smoking | พฤติกรรมการสูบบุหรี่ - สูบ - ไม่สูบ | t.smoke | 0 | 1 |
| Physical measurement | | | | |
| Height | ส่วนสูง (เซนติเมตร) | t.height | 0.0015 | 1.0015 |
| BMI | ดัชนีมวลกาย | BMI | 0.0157 | 1.0158 |
| SBP | Systolic Blood Pressure (mm Hg) | SBP | -0.0004 | 0.9996 |
| DBP | Diastolic Blood Pressure (mm Hg) | DBP | -0.0022 | 0.9978 |
| PP | Pulse Pressure คำนวณจาก SBP - DBP | PP | 0 | 1 |
| Underlying Disease | | | | |
| HT Duration | ระยะเวลาที่เป็นความดันโลหิตสูง (ปี) | ht_duration | 0.0061 | 1.0062 |
| HT Age of Diag | อายุที่ได้รับการวินิจฉัยว่าเป็นโรคความดันโลหิตสูง (ปี) | ht_age_of_diag | 0.0159 | 1.016 |
| DLP | สถานะภาวะไขมันผิดปกติ - เป็น - ไม่เป็น | flag_dlp | 0 | 1 |

ตารางที่ 41 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ใน
การทำนายความเสี่ยงภาวะเส้นเลือดหัวใจตีบหรือกล้ามเนื้อหัวใจขาดเลือดในผู้ป่วยโรคความดันโลหิตสูง (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ใน ตัวแบบ | สัมประสิทธิ์ การถดถอย | Odds Ratio |
|----------------|---|------------------------------|--------------------------|------------|
| DM | สถานะภาวะเบาหวาน - เป็น - ไม่เป็น | DB | 0.1556 | 1.1683 |
| DM Duration | ระยะเวลาที่เป็น โรคเบาหวาน (ปี) | db_duration | -0.0266 | 0.9738 |
| DM Age of Diag | อายุที่ได้รับการวินิจฉัยว่า เป็นโรคเบาหวาน (ปี) | db_age_of_diag | 0.0061 | 1.0061 |
| CKD | สถานะการเป็นโรคไต เรื้อรังโดยดูจากรหัสโรคที่ วินิจฉัย - เป็น - ไม่เป็น | flag_CKD | 0 | 1 |
| KIDNEY | สถานะการเป็นโรคไต เรื้อรังในระยะสุดท้าย (พิจารณาจาก CKD และ ค่า eGFR) - เป็น - ไม่เป็น | KIDNEY | 0.3865 | 1.4719 |
| Stroke | สถานะการเป็นโรคหลอดเลือด สมอง - เป็น - ไม่เป็น | CEREBRO | 0.1532 | 1.1655 |
| flag_stroke | สถานะการเป็นโรคหลอดเลือด สมอง พิจารณาจาก รหัสโรคที่วินิจฉัย - เป็น - ไม่เป็น | flag_stroke | 0 | 1 |

ตารางที่ 41 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ใน
การทำนายความเสี่ยงภาวะเส้นเลือดหัวใจตีบหรือกล้ามเนื้อหัวใจขาดเลือดในผู้ป่วยโรคความดันโลหิตสูง (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ใน ตัวแบบ | สัมประสิทธิ์ การถดถอย | Odds Ratio |
|----------------------|---|------------------------------|--------------------------|------------|
| Eye | สถานะการเป็นโรคตาอัน เนื่องมาจากเบาหวาน - เป็น - ไม่เป็น | EYE | 0.1647 | 1.179 |
| flag_eye | สถานะการเป็นโรคตา พิจารณาจาการหีสโรคที่ วินิจฉัย - เป็น - ไม่เป็น | flag_eye | 0 | 1 |
| COPD | สถานะการเป็นโรคถุงลม โป่งพอง - เป็น - ไม่เป็น | flag_copd | 0 | 1 |
| Lab | | | | |
| TC | ระดับคอเลสเตอรอล (mg/dL) | tcl_labresult | -0.0008 | 0.9992 |
| is_na_tcl_labresult* | ทราบค่า TC หรือไม่ - ทราบ - ไม่ทราบ | is_na_tcl_labresu lt | -0.0295 | 0.9709 |
| HDL | ระดับไขมันดี (mg/dL) | hdl_labresult | -0.0058 | 0.9943 |
| is_na_hdl_labresult* | ทราบค่า HDL หรือไม่ - ทราบ - ไม่ทราบ | is_na_hdl_labres ult | -0.276 | 0.7588 |
| LDL | ระดับไขมันไม่ดี (mg/dL) | ldl_labresult | 0.0009 | 1.0009 |
| is_na_ldl_labresult* | ทราบค่า HDL หรือไม่ - ทราบ - ไม่ทราบ | is_na_ldl_labres ult | 0.4923 | 1.636 |

ตารางที่ 41 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ใน
การทำนายความเสี่ยงภาวะเส้นเลือดหัวใจตีบหรือกล้ามเนื้อหัวใจขาดเลือดในผู้ป่วยโรคความดันโลหิตสูง (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ใน ตัวแบบ | สัมประสิทธิ์ การถดถอย | Odds Ratio |
|-------------------------|---|------------------------------|--------------------------|------------|
| TG | ระดับไขมันไตรกลีเซอไรด์ (mg/dL) | tg_labresult | -0.0003 | 0.9997 |
| is_na_tg_labresult* | ทราบค่า TG หรือไม่ - ทราบ - ไม่ทราบ | is_na_tg_labresult | 0.0914 | 1.0957 |
| non-HDL | TC – HDL | non_hdl | -0.0003 | 0.9997 |
| is_na_non_hdl* | ทราบค่า non-HDL หรือไม่ - ทราบ - ไม่ทราบ | is_na_non_hdl | 0 | 1 |
| total Cho:HDL | TC/HDL | total_cho_to_hdl | 0 | 1 |
| is_na_total_cho_to_hdl* | ทราบค่า total Cho:HDL หรือไม่ - ทราบ - ไม่ทราบ | is_na_total_cho_to_hdl | 0 | 1 |
| non HDL:HDL | (TC – HDL)/HDL | no_hdl_to_hdl | 0.0007 | 1.0007 |
| is_na_no_hdl_to_hdl* | ทราบค่า non HDL:HDL หรือไม่ - ทราบ - ไม่ทราบ | is_na_no_hdl_to_hdl | 0 | 1 |
| HBA1c | ค่าน้ำตาลเฉลี่ย สะสมในเลือด (mmol/mol) | HBA1c_labresult | 0.0343 | 1.0349 |

ตารางที่ 41 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ใน
การทำนายความเสี่ยงภาวะเส้นเลือดหัวใจตีบหรือกล้ามเนื้อหัวใจขาดเลือดในผู้ป่วยโรคความดันโลหิตสูง (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ใน ตัวแบบ | สัมประสิทธิ์ การถดถอย | Odds Ratio |
|------------------------|---|---------------------------------|--------------------------|------------|
| is_na_HBA1C_labresult* | ทราบค่า HBA1c หรือไม่ - ทราบ - ไม่ทราบ | is_na_HBA1C_lab result | 0.1807 | 1.1981 |
| FBS | Fasting Blood Sugar (mg/dL) | FBS_labresult | 0.0016 | 1.0016 |
| is_na_FBS_labresult* | ทราบค่า FBS หรือไม่ - ทราบ - ไม่ทราบ | is_na_FBS_labres ult | 0.1059 | 1.1118 |
| Macro-albuminuria | สถานะความเสี่ยง การเป็นโรคไตโดย วัดจาก Macro- albuminuria - เป็น - ไม่เป็น | macro_albuminu ria_labresult | 0.1819 | 1.1995 |
| Micro-albuminuria | สถานะความเสี่ยง การเป็นโรคไตโดย วัดจาก Micro- albuminuria - เป็น - ไม่เป็น | micro_albuminur ia_labresult | 0.067 | 1.0693 |
| eGFR | Estimated Glomerular Filtration Rate (mL/min/1.73m2) | egfr_labresult | -0.0126 | 0.9875 |

ตารางที่ 41 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ใน
การทำนายความเสี่ยงภาวะเส้นเลือดหัวใจตีบหรือกล้ามเนื้อหัวใจขาดเลือดในผู้ป่วยโรคความดันโลหิตสูง (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ใน ตัวแบบ | สัมประสิทธิ์ การถดถอย | Odds Ratio |
|-----------------------------|---|--------------------------------|--------------------------|------------|
| is_na_egfr_labresult* | ทราบค่า eGFR หรือไม่ - ทราบ - ไม่ทราบ | is_na_egfr_labres ult | -0.6978 | 0.4977 |
| Creatinine | ระดับ Creatinine (mg/dL) | creatinine_labres ult | -0.0094 | 0.9907 |
| is_na_creatinine_labresult* | ทราบค่า Creatinine หรือไม่ - ทราบ - ไม่ทราบ | is_na_creatinine_ labresult | -0.6925 | 0.5003 |
| UPCR | ระดับโปรตีนใน ปัสสาวะ (mg/g) | upcr_labresult | 0 | 1 |
| is_na_upcr_labresult* | ทราบค่า URCR หรือไม่ - ทราบ - ไม่ทราบ | is_na_upcr_labre sult | 0.1482 | 1.1597 |
| Hemoglobin | ระดับฮีโมโกลบินใน เลือด (g/dL) | hemoglobin_labr esult | -0.0011 | 0.9989 |
| is_na_hemoglobin_labresult* | ทราบค่า Hemoglobin หรือไม่ - ทราบ - ไม่ทราบ | is_na_hemoglobi n_labresult | -0.4065 | 0.666 |
| Albumin | ระดับ Albumin (g/dL) | albumin_labresu lt | -0.1719 | 0.842 |

ตารางที่ 41 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ใน
การทำนายความเสี่ยงภาวะเส้นเลือดหัวใจตีบหรือกล้ามเนื้อหัวใจขาดเลือดในผู้ป่วยโรคความดันโลหิตสูง (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ใน ตัวแบบ | สัมประสิทธิ์ การถดถอย | Odds Ratio |
|--------------------------|---|------------------------------|--------------------------|------------|
| is_na_albumin_labresult* | ทราบค่า Albumin หรือไม่ - ทราบ - ไม่ทราบ | is_na_albumin_l abresult | -0.8675 | 0.42 |
| Medication | | | | |
| Non-insulin med | การใช้อาหารเบาหวานที่ ไม่ต้องพึ่งอินซูลิน - ใช้ - ไม่ใช่ | flag_non_insulin _med | 0 | 1 |
| Insulin med | การใช้อาหารเบาหวานที่ ต้องพึ่งอินซูลิน - ใช้ - ไม่ใช่ | flag_insulin_med | 0 | 1 |

ตัวแบบทำนายความเสี่ยงภาวะเส้นเลือดสมองแตกหรือตีบในผู้ป่วยโรคความดันโลหิตสูง

จากกลุ่มตัวอย่างของผู้ป่วยโรคความดันโลหิตสูงจำนวน 232,571 คน ซึ่งประกอบไปด้วยข้อมูลจำนวน 1,464,009 แถว จะถูกสุ่มแบ่งออกเป็นชุดข้อมูลฝึกฝนและชุดข้อมูลทดสอบคิดเป็นร้อยละ 70 และ 30 โดยประมาณ ตามลำดับ โดยที่ชุดข้อมูลฝึกฝนและชุดข้อมูลทดสอบจะมีสัดส่วนของผู้ป่วยโรคความดันโลหิตสูงที่พบภาวะเส้นเลือดสมองแตกหรือตีบพอ ๆ กัน ซึ่งในที่นี้พบว่ามีความเท่ากันกับ 0.007903 หรือคือมีผู้ป่วยโรคความดันโลหิตสูงคิดเป็นร้อยละ 0.7903 ที่มีภาวะเส้นเลือดสมองแตกหรือตีบนั่นเอง จากการแบ่งข้อมูลจะได้ข้อมูลดังต่อไปนี้

- ชุดข้อมูลฝึกฝน ประกอบด้วยข้อมูลจากผู้เป็นโรคความดันโลหิตสูงจำนวน 162,800 คน ซึ่งมีข้อมูลจำนวน 1,024,885 แถว
- ชุดข้อมูลทดสอบ ประกอบด้วยข้อมูลจากผู้เป็นโรคความดันโลหิตสูงจำนวน 69,771 คน ซึ่งมีข้อมูลจำนวน 439,124 แถว

จากนั้นจึงสร้างตัวแบบจากข้อมูลชุดฝึกฝน ซึ่งจะพิจารณา 2 ตัวแบบ ได้แก่

1. การถดถอยโลจิสติกแบบ Elastic Net วิธีการนี้ขอเรียกว่า “EN”

2. วิธี XGBoost โดยวิธีการนี้ขอเรียกว่า “XGBoost”

หลังจากได้ตัวแบบทั้ง 2 แล้ว จึงนำมาทำนายข้อมูลในชุดทดสอบเพื่อวัดประสิทธิภาพการทำนายโดยพิจารณาจาก

1. AUC
2. Cumulative AUC

ตารางที่ 42 แสดงค่า AUC จาก 2 วิธี ซึ่งพบว่าวิธี XGBoost มีประสิทธิภาพในการทำนายความเสี่ยงการภาวะเส้นเลือดสมองแตกหรือตีบในผู้ป่วยโรคความดันโลหิตสูงได้ดีกว่าวิธีการถดถอยโลจิสติกแบบ Elastic Net แม้ว่าค่า AUC ที่ได้จากวิธี XGBoost จะดีกว่าวิธีการถดถอยโลจิสติกแบบ Elastic Net ไม่มาก แต่หากพิจารณาค่า Cumulative AUC ซึ่งวัดประสิทธิภาพการทำนายรายบุคคลผ่านตัววัดผล AUC ที่คำนวณด้วยความน่าจะเป็นสะสมของบุคคลตลอด Cohort (Cumulative Probability) กับสถานะอุบัติการณ์ปีสุดท้าย จะพบว่าวิธี XGBoost มีประสิทธิภาพเหนือกว่าวิธี Elastic Net ค่อนข้างมาก อย่างไรก็ตาม พบว่าวิธีทั้งสองอาจจะมีค่า AUC ที่ยังอยู่ในเกณฑ์ไม่สูง หรือมีค่าต่ำกว่า 0.8 ซึ่งในอนาคตควรจะต้องพิจารณาตัวแบบอื่นๆ รวมทั้งพิจารณาตัวแปรหรือใช้ขนาดตัวอย่างเพิ่มเติม เพื่อที่จะปรับปรุงตัวแบบในการทำนายให้มีประสิทธิภาพมากขึ้น

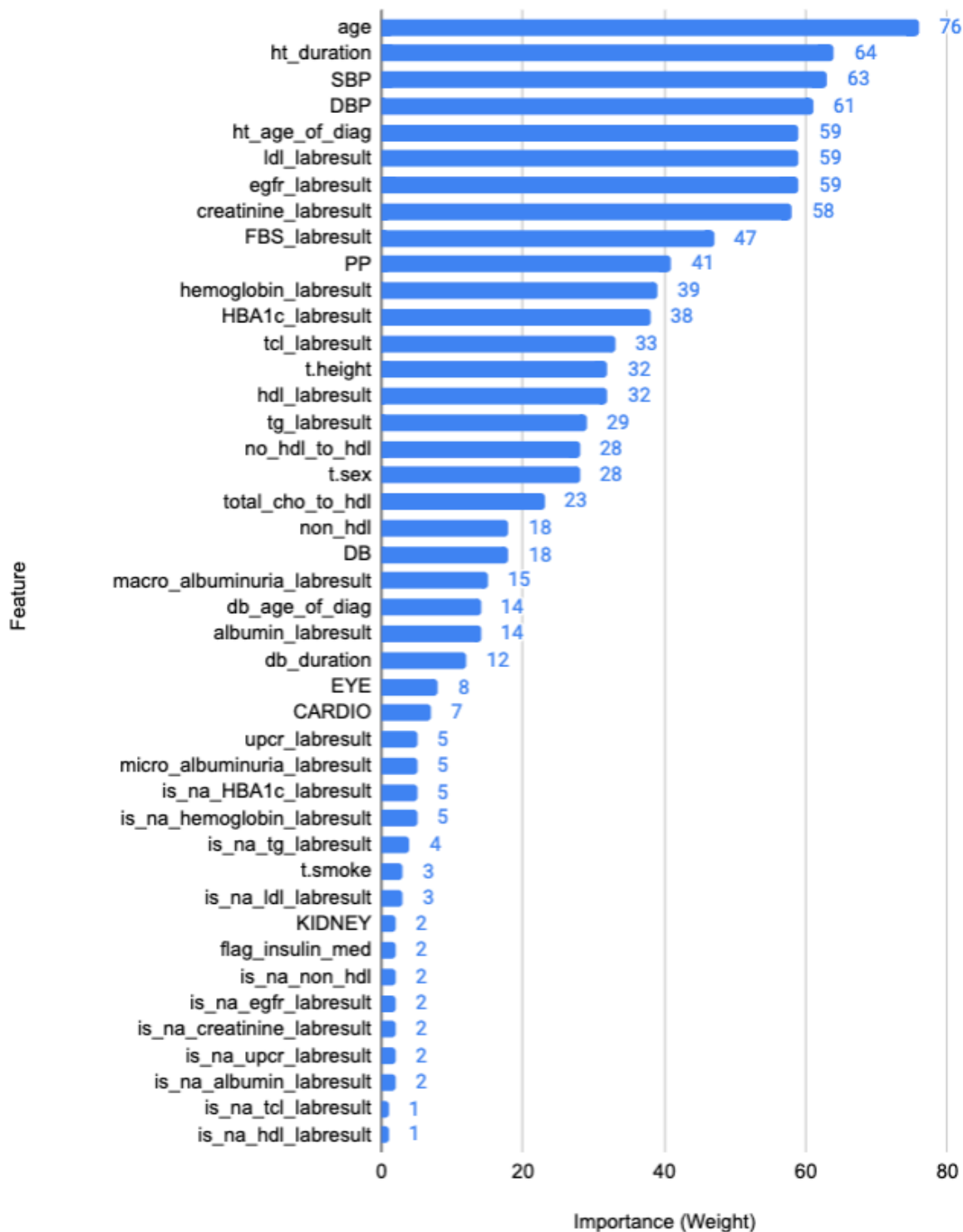
ตารางที่ 42 ค่า AUC เพื่อวัดประสิทธิภาพการทำนายความเสี่ยงภาวะเส้นเลือดสมองแตกหรือตีบในผู้ป่วยโรคความดันโลหิตสูง (Cohort 2.2) (วิธีที่ดีที่สุดจะแสดงเป็นตัวหนา)

| ตัวแบบ | AUC | Cumulative AUC |
|----------------|---------------|----------------|
| EN | 0.6418 | 0.6181 |
| XGBoost | 0.6467 | 0.6253 |

สำหรับการแปลผลตัวแบบโดยอ้างอิงจากตัวแบบ XGBoost ที่ให้ประสิทธิภาพการทำนายที่ดีที่สุดสามารถพิจารณาได้ว่าปัจจัยใดที่มีความสำคัญในการทำนายความเสี่ยงการเกิดโรคความดันโลหิตสูงในประชากรทั่วไปดังแสดงในรูปที่ 63 ซึ่งจะพบว่าตัวแปรที่มีความสำคัญในการทำนายความเสี่ยง 10 อันดับแรก ได้แก่

1. อายุ (age)
2. ระยะเวลาที่เป็นความดันโลหิตสูง (ปี) (ht_duration)
3. Systolic Blood Pressure (mm Hg) (SBP)
4. Diastolic Blood Pressure (mm Hg) (DBP)
5. อายุที่ได้รับการวินิจฉัยว่าเป็นโรคความดันโลหิตสูง (ปี) (ht_age_of_diag)
6. ระดับไขมันไม่ดี (mg/dL) (ldl_labresult)
7. Estimated Glomerular Filtration Rate (mL/min/1.73m²) (egfr_labresult)

8. ระดับ Creatinine (mg/dL) (creatinine_labresult)
9. Fasting Blood Sugar (mg/dL) (FBS_labresult)
10. Pulse Pressure คำนวณจาก SBP – DBP (PP)

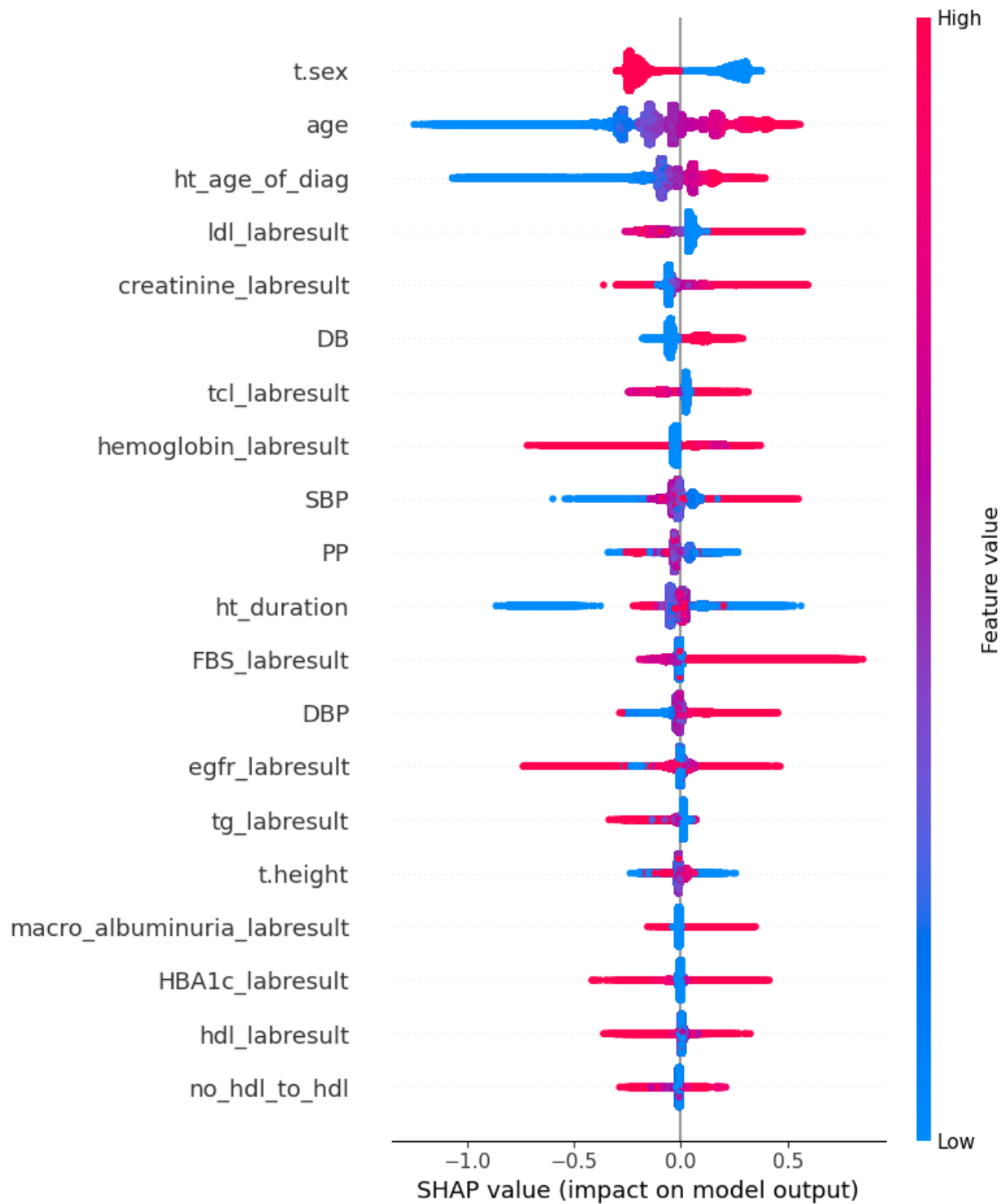


รูปที่ 63 ความสำคัญของตัวแปรทำนายจากวิธี XGBoost ในการทำนายความเสี่ยงเส้นเลือดสมองแตกหรือตีบในผู้ป่วยโรคความดันโลหิตสูง

นอกจากนี้ หากพิจารณาคุณค่า SHAP ของตัวแปร 20 ตัวแปรที่มีความสำคัญมากที่สุดในการทำนาย ความเสี่ยงภาวะเส้นเลือดสมองแตกหรือตีบในผู้ป่วยความดันโลหิตสูง ดังแสดงในรูปที่ 64 จะสามารถแปลผลใน ส่วนทิศทางการสัมพันธ์ของตัวแปรแต่ละตัว ซึ่งสามารถสรุปความสัมพันธ์ที่เด่นชัดได้ ดังนี้

- ผู้ป่วยความดันโลหิตสูงที่มีอายุน้อย (age) จะมีแนวโน้มที่จะมีความเสี่ยงในการมีภาวะเส้นเลือดสมองแตกหรือตีบน้อย
- ผู้ป่วยความดันโลหิตสูงที่ได้รับการวินิจฉัยว่าเป็นโรคความดันโลหิตสูงตอนช่วงอายุยังน้อย (ht_age_of_diag) จะมีแนวโน้มที่จะมีความเสี่ยงในการมีภาวะเส้นเลือดสมองแตกหรือตีบต่ำลง
- ผู้ป่วยความดันโลหิตสูงที่มีระดับ Creatinine (mg/dL) (creatinine_labresult) สูง จะมีแนวโน้มที่จะมีความเสี่ยงในการมีภาวะเส้นเลือดสมองแตกหรือตีบสูงขึ้น
- ผู้ป่วยความดันโลหิตสูงที่มีค่า Fasting Blood Sugar (mg/dL) (FBS_labresult) สูง จะมีแนวโน้มที่จะมีความเสี่ยงในการมีภาวะเส้นเลือดสมองแตกหรือตีบสูงขึ้น

แม้ว่าวิธี XGBoost สามารถระบุความสำคัญของตัวแปรทำนายที่ใช้ในตัวแบบได้ แต่ไม่สามารถแปลผลตัวแบบในรูปของอัตราเสี่ยงได้ ในขณะที่วิธีการถดถอยโลจิสติกแบบ Elastic Net ให้ค่าสัมประสิทธิ์การถดถอย เช่นเดียวกับตัวแบบการถดถอยโลจิสติกแบบปกติ ดังนั้นจึงสามารถแปลผลตัวแบบในรูปของอัตราเสี่ยงได้ตารางที่ 43 แสดงผลลัพธ์ที่ได้ของสัมประสิทธิ์การถดถอยของตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ในการทำนาย ความเสี่ยงภาวะเส้นเลือดสมองแตกหรือตีบในผู้ป่วยโรคความดันโลหิตสูง



รูปที่ 64 ค่า SHAP จากตัวแบบ XGBoost ในการทำนายความเสี่ยงภาวะเส้นเลือดสมองแตกหรือตีบในผู้ป่วยโรคความดันโลหิตสูง

ตารางที่ 43 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ใน
การทำนายความเสี่ยงภาวะเส้นเลือดสมองแตกหรือตีบในผู้ป่วยโรคความดันโลหิตสูง

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ใน ตัวแบบ | สัมประสิทธิ์ การถดถอย | Odds Ratio |
|-----------------------------|--|------------------------------|--------------------------|------------|
| Demographic | | | | |
| Age | อายุ (ปี) | age | 0.0124 | 1.0125 |
| Gender | เพศ - ชาย - หญิง | t.sex | -0.5436 | 0.5806 |
| Health Behavior | | | | |
| Smoking | พฤติกรรมการสูบบุหรี่ - สูบ - ไม่สูบ | t.smoke | 0.1046 | 1.1103 |
| Physical measurement | | | | |
| Height | ส่วนสูง (เซนติเมตร) | t.height | -0.0001 | 0.9999 |
| BMI | ดัชนีมวลกาย | BMI | 0.0058 | 1.0058 |
| SBP | Systolic Blood Pressure (mm Hg) | SBP | 0 | 1 |
| DBP | Diastolic Blood Pressure (mm Hg) | DBP | 0.0053 | 1.0054 |
| PP | Pulse Pressure คำนวณจาก SBP - DBP | PP | -0.0012 | 0.9988 |
| Underlying Disease | | | | |
| HT Duration | ระยะเวลาที่เป็นความดันโลหิตสูง (ปี) | ht_duration | 0.006 | 1.006 |
| HT Age of Diag | อายุที่ได้รับการวินิจฉัยว่าเป็นโรคความดันโลหิตสูง (ปี) | ht_age_of_diag | 0.0144 | 1.0145 |
| DLP | สถานะภาวะไขมันผิดปกติ - เป็น - ไม่เป็น | flag_dlp | 0 | 1 |

ตารางที่ 43 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ใน
การทำนายความเสี่ยงภาวะเส้นเลือดสมองแตกหรือตีบในผู้ป่วยโรคความดันโลหิตสูง (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ใน ตัวแบบ | สัมประสิทธิ์ การถดถอย | Odds Ratio |
|----------------|---|------------------------------|--------------------------|------------|
| DM | สถานะภาวะเบาหวาน - เป็น - ไม่เป็น | DB | 0.2095 | 1.2331 |
| DM Duration | ระยะเวลาที่เป็น โรคเบาหวาน (ปี) | db_duration | 0.0089 | 1.009 |
| DM Age of Diag | อายุที่ได้รับการวินิจฉัยว่า เป็นโรคเบาหวาน (ปี) | db_age_of_diag | 0 | 1 |
| Heart Disease | สถานะการเป็นโรคหัวใจ ซึ่งรวมถึงการเป็นโรคหัวใจ เต้นผิดปกติ - เป็น - ไม่เป็น | flag_HEART | 0 | 1 |
| CKD | สถานะการเป็นโรคไต เรื้อรังโดยดูจากรหัสโรคที่ วินิจฉัย - เป็น - ไม่เป็น | flag_CKD | 0 | 1 |
| KIDNEY | สถานะการเป็นโรคไต เรื้อรังในระยะสุดท้าย (พิจารณาจาก CKD และ ค่า eGFR) - เป็น - ไม่เป็น | KIDNEY | 0.0647 | 1.0668 |
| Eye | สถานะการเป็นโรคตาอัน เนื่องมาจากเบาหวาน - เป็น - ไม่เป็น | EYE | 0.303 | 1.3539 |

ตารางที่ 43 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ใน
การทำนายความเสี่ยงภาวะเส้นเลือดสมองแตกหรือตีบในผู้ป่วยโรคความดันโลหิตสูง (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ใน ตัวแบบ | สัมประสิทธิ์ การถดถอย | Odds Ratio |
|----------------------|--|------------------------------|--------------------------|------------|
| flag_eye | สถานะการเป็นโรคตา พิจารณาจากรหัสโรคที่ วินิจฉัย - เป็น - ไม่เป็น | flag_eye | 0 | 1 |
| COPD | สถานะการเป็นโรคถุงลม โป่งพอง - เป็น - ไม่เป็น | flag_copd | 0 | 1 |
| Lab | | | | |
| TC | ระดับคลอเรสเตอรอล (mg/dL) | tcl_labresult | 0.0003 | 1.0003 |
| is_na_tcl_labresult* | ทราบค่า TC หรือไม่ - ทราบ - ไม่ทราบ | is_na_tcl_labresu lt | 0.1147 | 1.1215 |
| HDL | ระดับไขมันดี (mg/dL) | hdl_labresult | -0.0037 | 0.9963 |
| is_na_hdl_labresult* | ทราบค่า HDL หรือไม่ - ทราบ - ไม่ทราบ | is_na_hdl_labres ult | -0.5035 | 0.6044 |
| LDL | ระดับไขมันไม่ดี (mg/dL) | ldl_labresult | 0.0013 | 1.0013 |
| is_na_ldl_labresult* | ทราบค่า HDL หรือไม่ - ทราบ - ไม่ทราบ | is_na_ldl_labres ult | 0.4798 | 1.6157 |
| TG | ระดับไขมันไตรกลีเซอไรด์ (mg/dL) | tg_labresult | 0.0003 | 1.0003 |

ตารางที่ 43 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ใน
การทำนายความเสี่ยงภาวะเส้นเลือดสมองแตกหรือตีบในผู้ป่วยโรคความดันโลหิตสูง (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ในตัว แบบ | สัมประสิทธิ์ การถดถอย | Odds Ratio |
|-------------------------|---|------------------------------|--------------------------|------------|
| is_na_tg_labresult* | ทราบค่า TG หรือไม่ - ทราบ - ไม่ทราบ | is_na_tg_labresult | 0.1696 | 1.1848 |
| non-HDL | TC – HDL | non_hdl | 0 | 1 |
| is_na_non_hdl* | ทราบค่า non-HDL หรือไม่ - ทราบ - ไม่ทราบ | is_na_non_hdl | 0.1794 | 1.1965 |
| total Cho:HDL | TC/HDL | total_cho_to_hdl | -0.0019 | 0.9981 |
| is_na_total_cho_to_hdl* | ทราบค่า total Cho:HDL หรือไม่ - ทราบ - ไม่ทราบ | is_na_total_cho_to_ hdl | 0 | 1 |
| non HDL:HDL | (TC – HDL)/HDL | no_hdl_to_hdl | 0 | 1 |
| is_na_no_hdl_to_hdl* | ทราบค่า non HDL:HDL หรือไม่ - ทราบ - ไม่ทราบ | is_na_no_hdl_to_hd l | 0 | 1 |
| HBA1c | ค่าน้ำตาลเฉลี่ยสะสม ในเลือด (mmol/mol) | HBA1c_labresult | 0.0162 | 1.0163 |
| is_na_HBA1C_labresult* | ทราบค่า HBA1c หรือไม่ - ทราบ - ไม่ทราบ | is_na_HBA1C_labres ult | 0.2178 | 1.2433 |
| FBS | Fasting Blood Sugar (mg/dL) | FBS_labresult | 0.0026 | 1.0026 |

ตารางที่ 43 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ใน
การทำนายความเสี่ยงภาวะเส้นเลือดสมองแตกหรือตีบในผู้ป่วยโรคความดันโลหิตสูง (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ใน ตัวแบบ | สัมประสิทธิ์ การถดถอย | Odds Ratio |
|-----------------------|---|---------------------------------|--------------------------|------------|
| is_na_FBS_labresult* | ทราบค่า FBS หรือไม่ - ทราบ - ไม่ทราบ | is_na_FBS_labres ult | 0.3354 | 1.3985 |
| Macro-albuminuria | สถานะความเสี่ยง การเป็นโรคไตโดย วัดจาก Macro- albuminuria - เป็น - ไม่เป็น | macro_albuminu ria_labresult | 0.3147 | 1.3699 |
| Micro-albuminuria | สถานะความเสี่ยง การเป็นโรคไตโดย วัดจาก Micro- albuminuria - เป็น - ไม่เป็น | micro_albuminur ia_labresult | 0.1148 | 1.1216 |
| eGFR | Estimated Glomerular Filtration Rate (mL/min/1.73m ²) | egfr_labresult | -0.0046 | 0.9954 |
| is_na_egfr_labresult* | ทราบค่า eGFR หรือไม่ - ทราบ - ไม่ทราบ | is_na_egfr_labres ult | -0.3109 | 0.7328 |
| Creatinine | ระดับ Creatinine (mg/dL) | creatinine_labres ult | -0.0145 | 0.9856 |

ตารางที่ 43 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ใน
การทำนายความเสี่ยงภาวะเส้นเลือดสมองแตกหรือตีบในผู้ป่วยโรคความดันโลหิตสูง (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ในตัวแบบ | สัมประสิทธิ์การถดถอย | Odds Ratio |
|-----------------------------|---|----------------------------|----------------------|------------|
| is_na_creatinine_labresult* | ทราบค่า Creatinine หรือไม่ - ทราบ - ไม่ทราบ | is_na_creatinine_labresult | -0.3036 | 0.7382 |
| UPCR | ระดับโปรตีนในปัสสาวะ (mg/g) | upcr_labresult | -0.0005 | 0.9995 |
| is_na_upcr_labresult* | ทราบค่า URCR หรือไม่ - ทราบ - ไม่ทราบ | is_na_upcr_labresult | 0.0122 | 1.0123 |
| Hemoglobin | ระดับฮีโมโกลบินในเลือด (g/dL) | hemoglobin_labresult | -0.0223 | 0.9779 |
| is_na_hemoglobin_labresult* | ทราบค่า Hemoglobin หรือไม่ - ทราบ - ไม่ทราบ | is_na_hemoglobin_labresult | -0.463 | 0.6294 |
| Albumin | ระดับ Albumin (g/dL) | albumin_labresult | -0.1432 | 0.8666 |
| is_na_albumin_labresult* | ทราบค่า Albumin หรือไม่ - ทราบ - ไม่ทราบ | is_na_albumin_labresult | -0.6317 | 0.5317 |

ตารางที่ 43 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ใน
การทำนายความเสี่ยงภาวะเส้นเลือดสมองแตกหรือตีบในผู้ป่วยโรคความดันโลหิตสูง (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ในตัว แบบ | สัมประสิทธิ์ การถดถอย | Odds Ratio |
|--------------------|--|------------------------------|--------------------------|------------|
| Medication | | | | |
| Lipid-lowering med | การใช้ยาลดไขมัน - ใช้ | flag_lipid_lowering | 0 | 1 |
| Non-insulin med | - ไม่ใช่ การใช้ยาเบาหวานที่ไม่ ต้องพึ่งอินซูลิน - ใช้ | flag_non_insulin_med | 0 | 1 |
| Insulin med | - ไม่ใช่ การใช้ยาเบาหวานที่ ต้องพึ่งอินซูลิน - ใช้ | flag_insulin_med | -0.5589 | 0.5719 |
| Aspirin | - ไม่ใช่ การใช้ยาแอสไพริน - ใช้ | flag_aspirin_med | 0 | 1 |
| | - ไม่ใช่ | | | |

ตัวแบบทำนายความเสี่ยงภาวะแทรกซ้อนทางตาในผู้ป่วยโรคความดันโลหิตสูง

จากกลุ่มตัวอย่างของผู้ป่วยโรคความดันโลหิตสูงจำนวน 234,443 คน ซึ่งประกอบไปด้วยข้อมูลจำนวน 1,484,774 แถว จะถูกสุ่มแบ่งออกเป็นชุดข้อมูลฝึกฝนและชุดข้อมูลทดสอบคิดเป็นร้อยละ 70 และ 30 โดยประมาณ ตามลำดับ โดยที่ชุดข้อมูลฝึกฝนและชุดข้อมูลทดสอบจะมีสัดส่วนของผู้ป่วยโรคความดันโลหิตสูงที่พบภาวะแทรกซ้อนทางตาพอ ๆ กัน ซึ่งในที่นี้พบว่ามีความเท่ากับ 0.004232 หรือคือมีผู้ป่วยโรคความดันโลหิตสูงคิดเป็นร้อยละ 0.4232 ที่มีภาวะแทรกซ้อนทางตานั่นเอง จากการแบ่งข้อมูลจะได้ข้อมูลดังต่อไปนี้

- ชุดข้อมูลฝึกฝน ประกอบด้วยข้อมูลจากผู้เป็นโรคความดันโลหิตสูงจำนวน 164,110 คน ซึ่งมีข้อมูลจำนวน 1,039,993 แถว
- ชุดข้อมูลทดสอบ ประกอบด้วยข้อมูลจากผู้เป็นโรคความดันโลหิตสูงจำนวน 70,333 คน ซึ่งมีข้อมูลจำนวน 444,781 แถว

จากนั้นจึงสร้างตัวแบบจากข้อมูลชุดฝึกฝน ซึ่งจะพิจารณา 2 ตัวแบบ ได้แก่

1. การถดถอยโลจิสติกแบบ Elastic Net วิธีการนี้ขอเรียกว่า “EN”

2. วิธี XGBoost โดยวิธีการนี้ขอเรียกว่า “XGBoost”

หลังจากได้ตัวแบบทั้ง 2 แล้ว จึงนำมาทำนายข้อมูลในชุดทดสอบเพื่อวัดประสิทธิภาพการทำนายโดยพิจารณาจาก

1. AUC
2. Cumulative AUC

ตารางที่ 44 แสดงค่า AUC จาก 2 วิธี ซึ่งพบว่าวิธี XGBoost มีประสิทธิภาพในการทำนายความเสี่ยงการภาวะแทรกซ้อนทางตาในผู้ป่วยโรคความดันโลหิตสูงได้ดีกว่าวิธีการถดถอยโลจิสติกแบบ Elastic Net แม้ว่าค่า AUC ที่ได้จากวิธี XGBoost จะดีกว่าวิธีการถดถอยโลจิสติกแบบ Elastic Net ไม่มาก แต่หากพิจารณาค่า Cumulative AUC ซึ่งวัดประสิทธิภาพการทำนายรายบุคคลผ่านตัววัดผล AUC ที่คำนวณด้วยความน่าจะเป็นสะสมของบุคคลตลอด Cohort (Cumulative Probability) กับสถานะอุบัติการณ์ปีสุดท้าย จะพบว่าวิธี XGBoost มีประสิทธิภาพเหนือกว่าวิธี Elastic Net ค่อนข้างมาก อย่างไรก็ตาม พบว่าวิธีทั้งสองมีค่า AUC สูงกว่า 0.8 ซึ่งอยู่ในเกณฑ์ที่ดีและรับได้

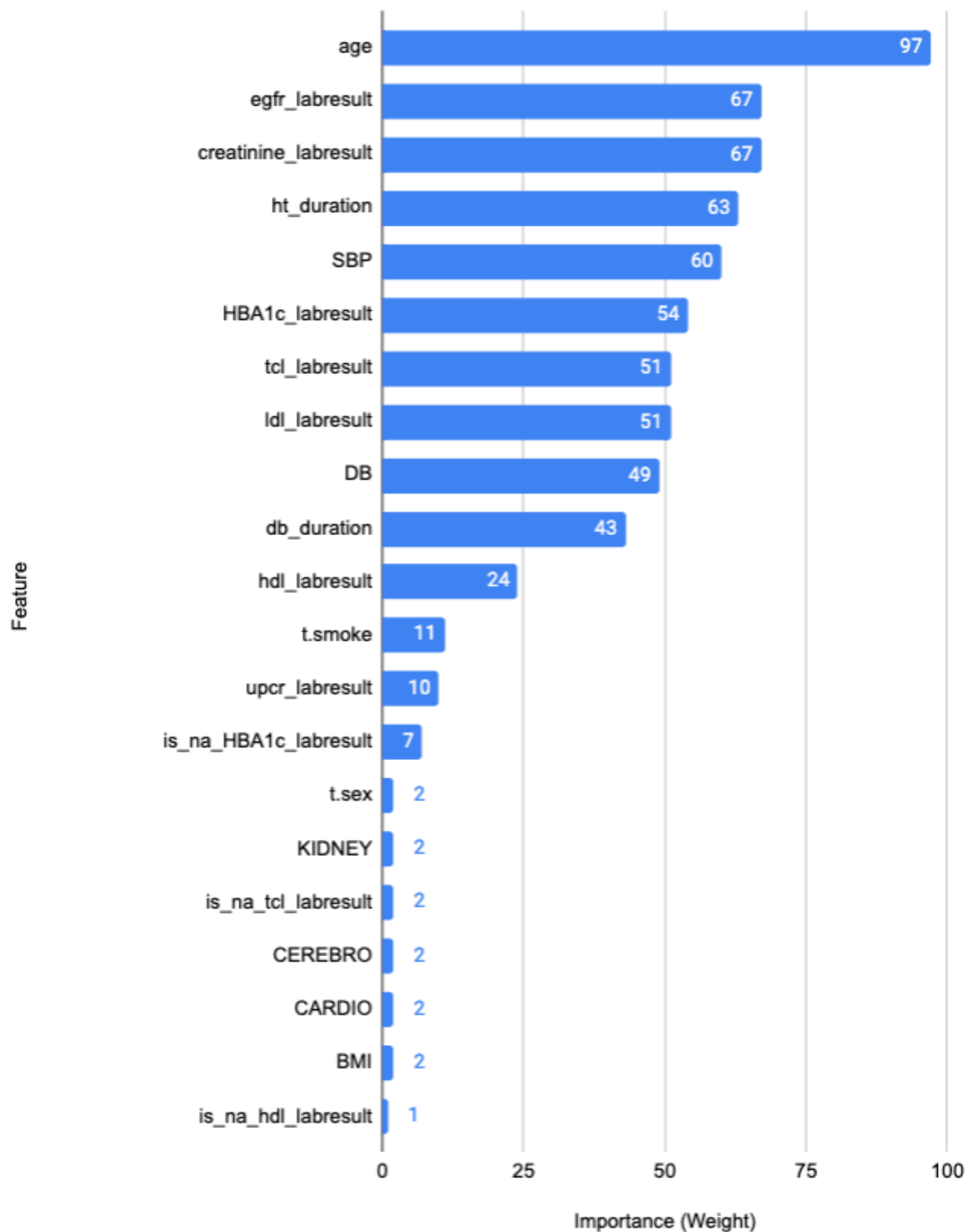
ตารางที่ 44 ค่า AUC เพื่อวัดประสิทธิภาพการทำนายความเสี่ยงภาวะแทรกซ้อนทางตาในผู้ป่วยโรคความดันโลหิตสูง (Cohort 2.2) (วิธีที่ดีที่สุดจะแสดงเป็นตัวหนา)

| ตัวแบบ | AUC | Cumulative AUC |
|---------|---------------|----------------|
| EN | 0.8147 | 0.8367 |
| XGBoost | 0.8415 | 0.8413 |

สำหรับการแปลผลตัวแบบโดยอ้างอิงจากตัวแบบ XGBoost ที่ให้ประสิทธิภาพการทำนายที่ดีที่สุดสามารถพิจารณาได้ว่าปัจจัยใดที่มีความสำคัญในการทำนายความเสี่ยงการเกิดโรคความดันโลหิตสูงในประชากรทั่วไปดังแสดงในรูปที่ 65 ซึ่งจะพบว่าตัวแปรที่มีความสำคัญในการทำนายความเสี่ยง 10 อันดับแรก ได้แก่

1. อายุ (age)
2. Estimated Glomerular Filtration Rate (mL/min/1.73m²) (egfr_labresult)
3. ระดับ Creatinine (mg/dL) (creatinine_labresult)
4. ระยะเวลาที่เป็นความดันโลหิตสูง (ปี) (ht_duration)
5. Systolic Blood Pressure (mm Hg) (SBP)
6. ค่าน้ำตาลเฉลี่ยสะสมในเลือด (mmol/mol) (HBA1c_labresult)
7. ระดับคอเลสเตอรอล (mg/dL) (tcl_labresult)

8. ระดับไขมันไม่ดี (mg/dL) (ldl_labresult)
9. สถานะโรคเบาหวาน (DB)
10. ระยะเวลาที่เป็นโรคเบาหวาน (ปี) (db_duration)

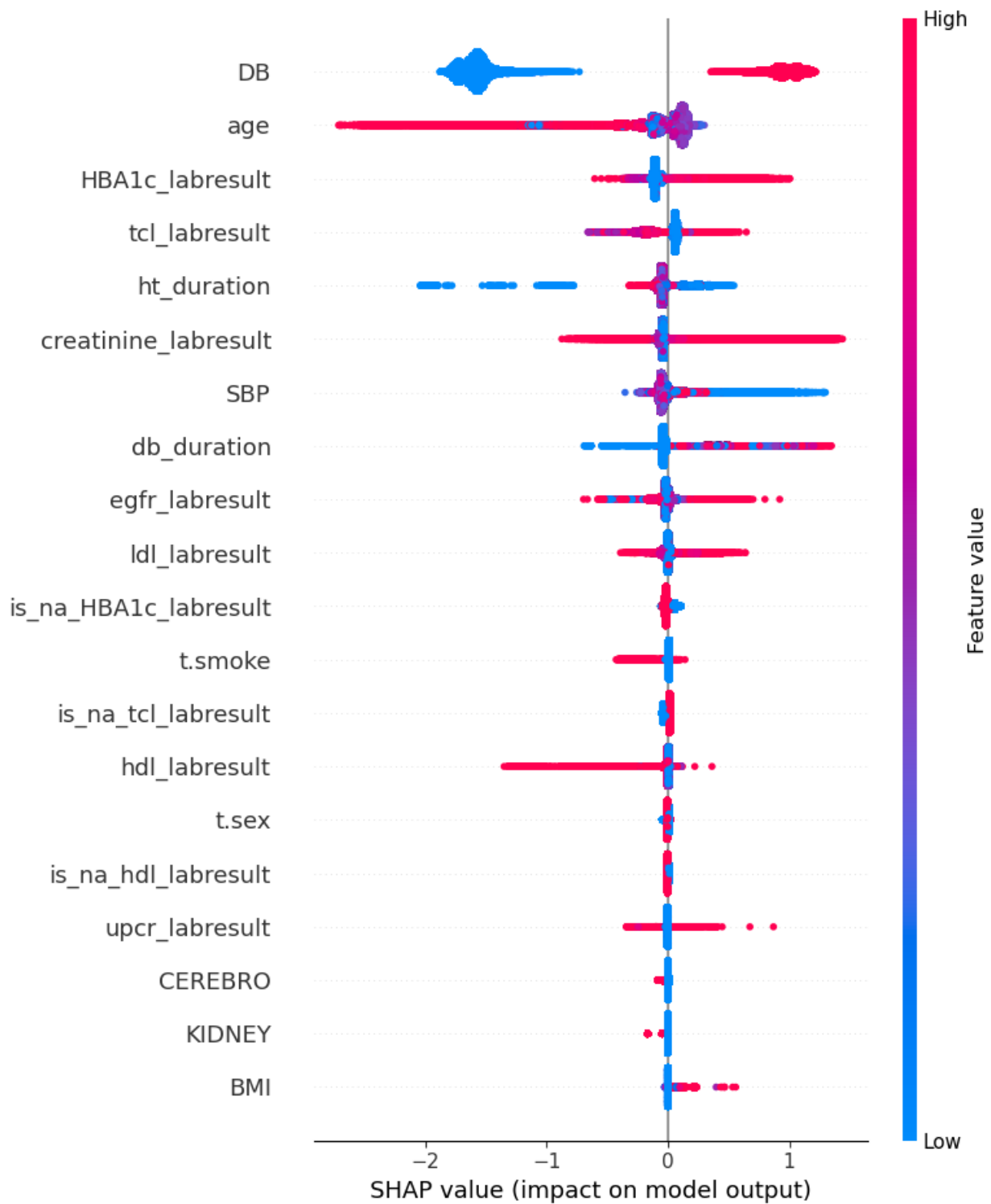


รูปที่ 65 ความสำคัญของตัวแปรทำนายจากวิธี XGBoost ในการทำนายภาวะแทรกซ้อนทางตาในผู้ป่วยโรคความดันโลหิตสูง

นอกจากนี้ หากพิจารณาดูค่า SHAP ของตัวแปร 20 ตัวแปรที่มีความสำคัญมากที่สุดในการทำนายความเสี่ยงภาวะแทรกซ้อนทางตาในผู้ป่วยความดันโลหิตสูง ดังแสดงในรูปที่ 66 จะสามารถแปลผลในส่วนทิศทางความสัมพันธ์ของตัวแปรแต่ละตัว ซึ่งสามารถสรุปความสัมพันธ์ที่เด่นชัดได้ ดังนี้

- ผู้ป่วยความดันโลหิตสูง ที่ได้รับการวินิจฉัยว่าเป็นโรคเบาหวานด้วย (DB) จะมีแนวโน้มที่จะมีความเสี่ยงในการมีภาวะแทรกซ้อนทางตาสูงขึ้น
- ผู้ป่วยความดันโลหิตสูง ที่ได้รับการวินิจฉัยว่ามีอายุมาก (age) จะมีแนวโน้มที่จะมีความเสี่ยงในการมีภาวะแทรกซ้อนทางตาน้อยลง
- ผู้ป่วยความดันโลหิตสูง ที่มีค่า Systolic Blood Pressure (mm Hg) (SBP) ต่ำ จะมีแนวโน้มที่จะมีความเสี่ยงในการมีภาวะแทรกซ้อนทางตาสูงขึ้น
- ผู้ป่วยความดันโลหิตสูง ที่มีค่าระดับไขมันดี (mg/dL) (hdl_labresult) สูง จะมีแนวโน้มที่จะมีความเสี่ยงในการมีภาวะแทรกซ้อนทางตาน้อยลง

แม้ว่าวิธี XGBoost สามารถระบุความสำคัญของตัวแปรทำนายที่ใช้ในตัวแบบได้ แต่ไม่สามารถแปลผลตัวแบบในรูปของอัตราเสี่ยงได้ ในขณะที่วิธีการถดถอยโลจิสติกแบบ Elastic Net ให้ค่าสัมประสิทธิ์การถดถอยเช่นเดียวกับตัวแบบการถดถอยโลจิสติกแบบปกติ ดังนั้นจึงสามารถแปลผลตัวแบบในรูปของอัตราเสี่ยงได้ ตารางที่ 45 แสดงผลลัพธ์ที่ได้ของสัมประสิทธิ์การถดถอยของตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ในการทำนายความเสี่ยงภาวะแทรกซ้อนทางตาในผู้ป่วยโรคความดันโลหิตสูง



รูปที่ 66 ค่า SHAP จากตัวแบบ XGBoost ในการทำนายความเสี่ยงภาวะแทรกซ้อนทางตาในผู้ป่วยโรคความดันโลหิตสูง

ตารางที่ 45 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ใน
การทำนายความเสี่ยงภาวะแทรกซ้อนทางตาในผู้ป่วยโรคความดันโลหิตสูง

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ใน ตัวแบบ | สัมประสิทธิ์ การถดถอย | Odds Ratio |
|-----------------------------|---|------------------------------|--------------------------|------------|
| Demographic | | | | |
| Age | อายุ (ปี) | age | -0.0192 | 0.981 |
| Gender | เพศ - ชาย - หญิง | t.sex | -0.0412 | 0.9597 |
| Health Behavior | | | | |
| Smoking | พฤติกรรมการสูบบุหรี่ - สูบ - ไม่สูบ | t.smoke | -0.3836 | 0.6814 |
| Physical measurement | | | | |
| BMI | ดัชนีมวลกาย | BMI | 0.0102 | 1.0102 |
| SBP | Systolic Blood Pressure (mm Hg) | SBP | 0.0016 | 1.0016 |
| Underlying Disease | | | | |
| HT Duration | ระยะเวลาที่เป็นความดัน โลหิตสูง (ปี) | ht_duration | -0.0224 | 0.9779 |
| DM | สถานะภาวะเบาหวาน - เป็น - ไม่เป็น | DB | 2.8763 | 17.7482 |
| DM Duration | ระยะเวลาที่เป็น โรคเบาหวาน (ปี) | db_duration | 0.0235 | 1.0238 |
| Heart Disease | สถานะการเป็นโรคหัวใจ ซึ่งรวมถึงการเป็น โรคหัวใจเต้นผิดปกติ - เป็น - ไม่เป็น | flag_HEART | 0 | 1 |
| Lab | | | | |
| TC | ระดับคอเลสเตอรอล (mg/dL) | tcl_labresult | 0.0008 | 1.0008 |

ตารางที่ 45 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ใน
การทำนายความเสี่ยงภาวะแทรกซ้อนทางตาในผู้ป่วยโรคความดันโลหิตสูง (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ใน ตัวแบบ | สัมประสิทธิ์ การถดถอย | Odds Ratio |
|------------------------|---|------------------------------|--------------------------|------------|
| is_na_tcl_labresult* | ทราบค่า TC หรือไม่ - ทราบ - ไม่ทราบ | is_na_tcl_labresult | 0.4308 | 1.5385 |
| HDL | ระดับไขมันดี (mg/dL) | hdl_labresult | -0.0027 | 0.9973 |
| is_na_hdl_labresult* | ทราบค่า HDL หรือไม่ - ทราบ - ไม่ทราบ | is_na_hdl_labresult | -0.3258 | 0.7219 |
| LDL | ระดับไขมันไม่ดี (mg/dL) | ldl_labresult | 0.0025 | 1.0025 |
| is_na_ldl_labresult* | ทราบค่า HDL หรือไม่ - ทราบ - ไม่ทราบ | is_na_ldl_labresult | 0.5764 | 1.7796 |
| HBA1c | ค่าน้ำตาลเฉลี่ยสะสมในเลือด (mmol/mol) | HBA1c_labresult | 0.073 | 1.0757 |
| is_na_HBA1C_labresult* | ทราบค่า HBA1c หรือไม่ - ทราบ - ไม่ทราบ | is_na_HBA1C_labresult | 0.47 | 1.6 |
| eGFR | Estimated Glomerular Filtration Rate (mL/min/1.73m ²) | egfr_labresult | -0.0077 | 0.9923 |
| is_na_egfr_labresult* | ทราบค่า eGFR หรือไม่ - ทราบ - ไม่ทราบ | is_na_egfr_labresult | -0.465 | 0.6282 |
| Creatinine | ระดับ Creatinine (mg/dL) | creatinine_labresult | -0.0212 | 0.9791 |

ตารางที่ 45 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ใน
การทำนายความเสี่ยงภาวะแทรกซ้อนทางตาในผู้ป่วยโรคความดันโลหิตสูง (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ในตัวแบบ | สัมประสิทธิ์การถดถอย | Odds Ratio |
|-----------------------------|--|----------------------------|----------------------|------------|
| is_na_creatinine_labresult* | ทราบค่า Creatinine หรือไม่ - ทราบ - ไม่ทราบ | is_na_creatinine_labresult | -0.4144 | 0.6608 |
| UPCR | ระดับโปรตีนในปัสสาวะ (mg/g) | upcr_labresult | 0.003 | 1.003 |
| is_na_upcr_labresult* | ทราบค่า URCR หรือไม่ - ทราบ - ไม่ทราบ | is_na_upcr_labresult | 0.0892 | 1.0933 |
| Medication | | | | |
| Lipid-lowering med | การใช้ยาลดไขมัน - ใช้ - ไม่ใช่ | flag_lipid_lowering | 0 | 1 |
| Non-insulin med | การใช้ยาเบาหวานที่ไม่ต้องพึ่งอินซูลิน - ใช้ - ไม่ใช่ | flag_non_insulin_med | 0 | 1 |

ตัวแบบทำนายความเสี่ยงภาวะไตเสื่อมเรื้อรังในผู้ป่วยโรคความดันโลหิตสูง

จากกลุ่มตัวอย่างของผู้ป่วยโรคความดันโลหิตสูงจำนวน 234,471 คน ซึ่งประกอบไปด้วยข้อมูลจำนวน 1,476,892 แถว จะถูกสุ่มแบ่งออกเป็นชุดข้อมูลฝึกฝนและชุดข้อมูลทดสอบคิดเป็นร้อยละ 70 และ 30 โดยประมาณ ตามลำดับ โดยที่ชุดข้อมูลฝึกฝนและชุดข้อมูลทดสอบจะมีสัดส่วนของผู้ป่วยโรคความดันโลหิตสูงที่พบภาวะไตเสื่อมเรื้อรังพอ ๆ กัน ซึ่งในที่นี้พบว่ามีความเท่ากับ 0.007087 หรือคือมีผู้ป่วยโรคความดันโลหิตสูงคิดเป็นร้อยละ 0.7087 ที่มีภาวะไตเสื่อมเรื้อรังนั่นเอง จากการแบ่งข้อมูลจะได้ข้อมูลดังต่อไปนี้

- ชุดข้อมูลฝึกฝน ประกอบด้วยข้อมูลจากผู้เป็นโรคความดันโลหิตสูงจำนวน 164,129 คน ซึ่งมีข้อมูลจำนวน 1,033,254 แถว

- ชุดข้อมูลทดสอบ ประกอบด้วยข้อมูลจากผู้เป็นโรคความดันโลหิตสูงจำนวน 70,342 คน ซึ่งมีข้อมูลจำนวน 443,638 แถว

จากนั้นจึงสร้างตัวแบบจากข้อมูลชุดฝึกฝน ซึ่งจะพิจารณา 2 ตัวแบบ ได้แก่

1. การถดถอยโลจิสติกแบบ Elastic Net วิธีการนี้ขอเรียกว่า “EN”
2. วิธี XGBoost โดยวิธีการนี้ขอเรียกว่า “XGBoost”

หลังจากได้ตัวแบบทั้ง 2 แล้ว จึงนำมาทำนายข้อมูลในชุดทดสอบเพื่อวัดประสิทธิภาพการทำนายโดยพิจารณาจาก

1. AUC
2. Cumulative AUC

ตารางที่ 46 แสดงค่า AUC จาก 2 วิธี ซึ่งพบว่าวิธี XGBoost มีประสิทธิภาพในการทำนายความเสี่ยงการภาวะไตเสื่อมเรื้อรังในผู้ป่วยโรคความดันโลหิตสูงได้ดีกว่าวิธีการถดถอยโลจิสติกแบบ Elastic Net แม้ว่าค่า AUC ที่ได้จากวิธี XGBoost จะดีกว่าวิธีการถดถอยโลจิสติกแบบ Elastic Net ไม่มาก แต่หากพิจารณาค่า Cumulative AUC ซึ่งวัดประสิทธิภาพการทำนายรายบุคคลผ่านตัววัดผล AUC ที่คำนวณด้วยความน่าจะเป็นสะสมของบุคคลตลอด Cohort (Cumulative Probability) กับสถานะปฏิบัติการณปีสุดท้าย จะพบว่าวิธี XGBoost มีประสิทธิภาพเหนือกว่าวิธี Elastic Net ค่อนข้างมาก อย่างไรก็ตาม พบว่าวิธีทั้งสองอาจจะมีค่า AUC ที่ยังอยู่ในเกณฑ์ไม่สูง หรือมีค่าต่ำกว่า 0.8 ซึ่งในอนาคตควรจะต้องพิจารณาตัวแบบอื่นๆ รวมทั้งพิจารณาตัวแปรหรือใช้ขนาดตัวอย่างเพิ่มเติม เพื่อที่จะปรับปรุงตัวแบบในการทำนายให้มีประสิทธิภาพมากขึ้น

ตารางที่ 46 ค่า AUC เพื่อวัดประสิทธิภาพการทำนายความเสี่ยงภาวะไตเสื่อมเรื้อรังในผู้ป่วยโรคความดันโลหิตสูง (Cohort 2.2) (วิธีที่ดีที่สุดจะแสดงเป็นตัวหนา)

| ตัวแบบ | AUC | Cumulative AUC |
|---------|---------------|----------------|
| EN | 0.7803 | 0.7981 |
| XGBoost | 0.7973 | 0.8166 |

สำหรับการแปลผลตัวแบบโดยอ้างอิงจากตัวแบบ XGBoost ที่ให้ประสิทธิภาพการทำนายดีที่สุดที่สุดสามารถพิจารณาได้ว่าปัจจัยใดที่มีความสำคัญในการทำนายความเสี่ยงการเกิดโรคความดันโลหิตสูงในประชากรทั่วไปดังแสดงในรูปที่ 67 ซึ่งจะพบว่าตัวแปรที่มีความสำคัญในการทำนายความเสี่ยง 10 อันดับแรก ได้แก่

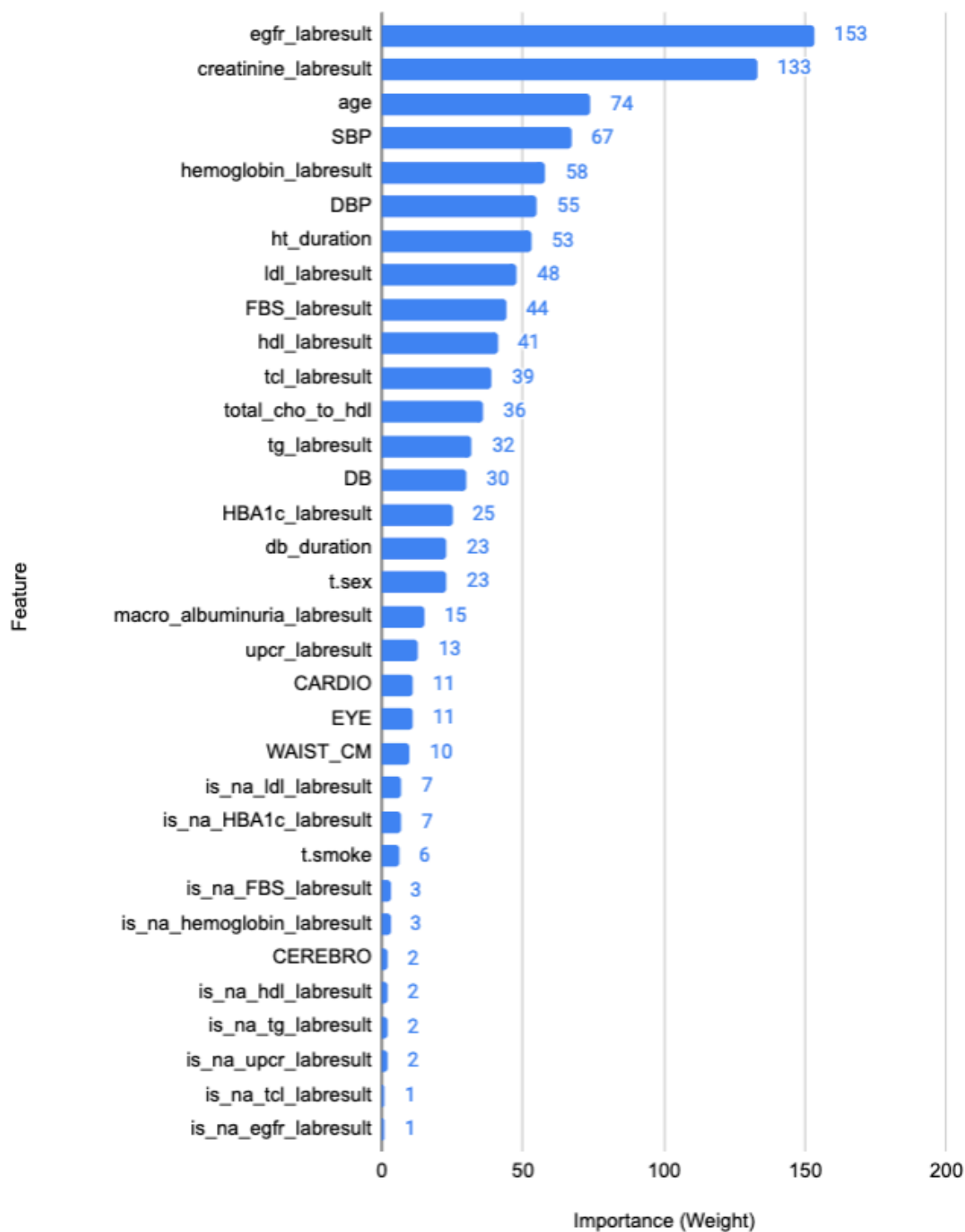
1. Estimated Glomerular Filtration Rate (mL/min/1.73m²) (egfr_labresult)
2. ระดับ Creatinine (mg/dL) (creatinine_labresult)
3. อายุ (age)

4. Systolic Blood Pressure (mm Hg) (SBP)
5. ระดับฮีโมโกลบินในเลือด (g/dL) (hemoglobin_labresult)
6. Diastolic Blood Pressure (mm Hg) (DBP)
7. ระยะเวลาที่เป็นความดันโลหิตสูง (ปี) (ht_duration)
8. ระดับไขมันไม่ดี (mg/dL) (ldl_labresult)
9. Fasting Blood Sugar (mg/dL) (FBS_labresult)
10. ระดับไขมันดี (mg/dL) (hdl_labresult)

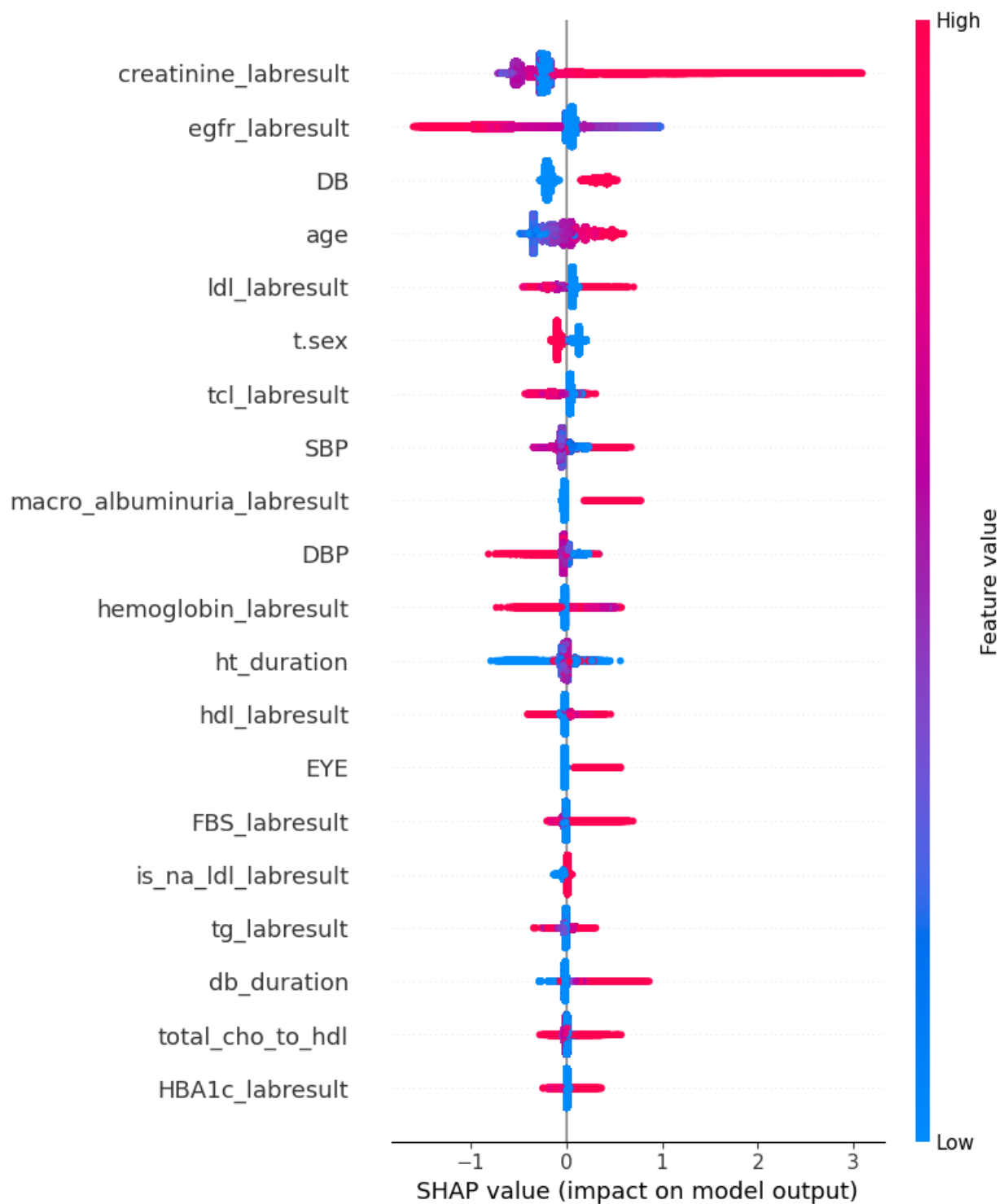
นอกจากนี้ หากพิจารณาคุณค่า SHAP ของตัวแปร 20 ตัวแปรที่มีความสำคัญมากที่สุดในการทำนายความเสี่ยงภาวะไตเสื่อมเรื้อรังในผู้ป่วยโรคความดันโลหิตสูง ดังแสดงในรูปที่ 68 จะสามารถแปลผลในส่วนทิศทางความสัมพันธ์ของตัวแปรแต่ละตัว ซึ่งสามารถสรุปความสัมพันธ์ที่เด่นชัดได้ ดังนี้

- ผู้ป่วยโรคความดันโลหิตสูง ที่มีระดับ Creatinine (mg/dL) (creatinine_labresult) สูง จะมีแนวโน้มที่จะมีความเสี่ยงในการมีภาวะไตเสื่อมเรื้อรังสูงขึ้น
- ผู้ป่วยโรคความดันโลหิตสูง ที่มี Estimated Glomerular Filtration Rate (mL/min/1.73m²) (egfr_labresult) สูง จะมีความเสี่ยงในการมีภาวะไตเสื่อมเรื้อรังต่ำลง
- ผู้ป่วยโรคความดันโลหิตสูง ที่มีอายุ (age) มาก จะมีแนวโน้มที่จะมีความเสี่ยงในการมีภาวะไตเสื่อมเรื้อรังสูงขึ้น
- ผู้ป่วยโรคความดันโลหิตสูง ที่มีค่า Systolic Blood Pressure (mm Hg) (SBP) สูง จะมีแนวโน้มที่จะมีความเสี่ยงในการมีภาวะไตเสื่อมเรื้อรังสูงขึ้น

แม้ว่าวิธี XGBoost สามารถระบุความสำคัญของตัวแปรทำนายที่ใช้ในตัวแบบได้ แต่ไม่สามารถแปลผลตัวแบบในรูปของอัตราเสี่ยงได้ ในขณะที่วิธีการถดถอยโลจิสติกแบบ Elastic Net ให้ค่าสัมประสิทธิ์การถดถอย เช่นเดียวกับตัวแบบการถดถอยโลจิสติกแบบปกติ ดังนั้นจึงสามารถแปลผลตัวแบบในรูปของอัตราเสี่ยงได้ ตารางที่ 47 แสดงผลลัพธ์ที่ได้ของสัมประสิทธิ์การถดถอยของตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ในการทำนายความเสี่ยงภาวะไตเสื่อมเรื้อรังในผู้ป่วยโรคความดันโลหิตสูง



รูปที่ 67 ความสำคัญของตัวแปรทำนายจากวิธี XGBoost ในการทำนายภาวะไตเสื่อมเรื้อรังในผู้ป่วยโรคความดันโลหิตสูง



รูปที่ 68 ค่า SHAP จากตัวแบบ XGBoost ในการทำนายความเสี่ยงภาวะไตเสื่อมเรื้อรังในผู้ป่วยโรคความดันโลหิตสูง

ตารางที่ 47 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ใน
การทำนายความเสี่ยงภาวะไตเสื่อมเรื้อรังหรือตีบในผู้ป่วยโรคความดันโลหิตสูง

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ใน ตัวแบบ | สัมประสิทธิ์ การถดถอย | Odds Ratio |
|-----------------------------|--|------------------------------|--------------------------|------------|
| Demographic | | | | |
| Age | อายุ (ปี) | age | 0.0027 | 1.0027 |
| Gender | เพศ - ชาย - หญิง | t.sex | -0.5491 | 0.5775 |
| Health Behavior | | | | |
| Smoking | พฤติกรรมการสูบบุหรี่ - สูบ - ไม่สูบ | t.smoke | -0.0776 | 0.9254 |
| Physical measurement | | | | |
| BMI | ดัชนีมวลกาย | BMI | 0.0046 | 1.0046 |
| SBP | Systolic Blood Pressure (mm Hg) | SBP | 0.0009 | 1.0009 |
| DBP | Diastolic Blood Pressure (mm Hg) | DBP | -0.0044 | 0.9956 |
| Underlying Disease | | | | |
| HT Duration | ระยะเวลาที่เป็นความดัน โลหิตสูง (ปี) | ht_duration | -0.0012 | 0.9988 |
| DLP | สถานะภาวะไขมัน ผิดปกติ - เป็น - ไม่เป็น | flag_dlp | 0 | 1 |
| DM | สถานะภาวะเบาหวาน - เป็น - ไม่เป็น | DB | 0.5821 | 1.7898 |
| DM Duration | ระยะเวลาที่เป็น โรคเบาหวาน (ปี) | db_duration | 0.0134 | 1.0135 |

ตารางที่ 47 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ใน
การทำนายความเสี่ยงภาวะไตเสื่อมเรื้อรังหรือตบในผู้ป่วยโรคความดันโลหิตสูง (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ในตัว แบบ | สัมประสิทธิ์ การถดถอย | Odds Ratio |
|----------------------|--|------------------------------|--------------------------|------------|
| Heart Disease | สถานะการเป็นโรคหัวใจ ซึ่ง รวมถึงการเป็นโรคหัวใจเต้น ผิดปกติ - เป็น - ไม่เป็น | flag_HEART | 0 | 1 |
| Stroke | สถานะการเป็นโรคหลอดเลือด สมอง - เป็น - ไม่เป็น | CEREBRO | 0.0741 | 1.0769 |
| Eye | สถานะการเป็นโรคตาอัน เนื่องมาจากเบาหวาน - เป็น - ไม่เป็น | EYE | 0.6041 | 1.8296 |
| flag_eye | สถานะการเป็นโรคตา พิจารณาจากรหัสโรคที่ วินิจฉัย - เป็น - ไม่เป็น | flag_eye | 0 | 1 |
| COPD | สถานะการเป็นโรคถุงลมโป่ง พอง - เป็น - ไม่เป็น | flag_copd | 0 | 1 |
| Lab | | | | |
| TC | ระดับคลอเรสเตอรอล (mg/dL) | tcl_labresult | 0.0001 | 1.0001 |
| is_na_tcl_labresult* | ทราบค่า TC หรือไม่ - ทราบ - ไม่ทราบ | is_na_tcl_labresult | 0 | 1 |

ตารางที่ 47 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ใน
การทำนายความเสี่ยงภาวะไตเสื่อมเรื้อรังหรือตบในผู้ป่วยโรคความดันโลหิตสูง (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ในตัว แบบ | สัมประสิทธิ์ การถดถอย | Odds Ratio |
|-------------------------|---|------------------------------|--------------------------|------------|
| HDL | ระดับไขมันดี (mg/dL) | hdl_labresult | 0.0001 | 1.0001 |
| is_na_hdl_labresult* | ทราบค่า HDL หรือไม่ - ทราบ - ไม่ทราบ | is_na_hdl_labresul t | -0.3153 | 0.7296 |
| LDL | ระดับไขมันไม่ดี (mg/dL) | ldl_labresult | 0.0017 | 1.0017 |
| is_na_ldl_labresult* | ทราบค่า HDL หรือไม่ - ทราบ - ไม่ทราบ | is_na_ldl_labresult | 0.5205 | 1.6828 |
| TG | ระดับไขมันไตรกลีเซอ ไรด์ (mg/dL) | tg_labresult | 0.0007 | 1.0007 |
| is_na_tg_labresult* | ทราบค่า TG หรือไม่ - ทราบ - ไม่ทราบ | is_na_tg_labresult | 0.0745 | 1.0774 |
| total Cho:HDL | TC/HDL | total_cho_to_hdl | 0.0533 | 1.0548 |
| is_na_total_cho_to_hdl* | ทราบค่า total Cho:HDL หรือไม่ - ทราบ - ไม่ทราบ | is_na_total_cho_t o_hdl | 0.6004 | 1.8229 |
| HBA1c | ค่าน้ำตาลเฉลี่ยสะสม ในเลือด (mmol/mol) | HBA1c_labresult | -0.0036 | 0.9964 |
| is_na_HBA1C_labresult* | ทราบค่า HBA1c หรือไม่ - ทราบ - ไม่ทราบ | is_na_HBA1C_labr esult | 0.1493 | 1.1611 |
| FBS | Fasting Blood Sugar (mg/dL) | FBS_labresult | 0.0014 | 1.0014 |

ตารางที่ 47 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ใน
การทำนายความเสี่ยงภาวะไตเสื่อมเรื้อรังหรือตีบในผู้ป่วยโรคความดันโลหิตสูง (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ในตัว แบบ | สัมประสิทธิ์ การถดถอย | Odds Ratio |
|---------------------------------|--|---------------------------------|--------------------------|------------|
| is_na_FBS_labresult* | ทราบค่า FBS หรือไม่ - ทราบ - ไม่ทราบ | is_na_FBS_labresul t | 0.1238 | 1.1318 |
| Macro-albuminuria | สถานะความเสี่ยงการ เป็นโรคไตโดยวัดจาก Macro-albuminuria - เป็น - ไม่เป็น | macro_albuminuri a_labresult | 0.5715 | 1.771 |
| eGFR | Estimated Glomerular Filtration Rate (mL/min/1.73m ²) | egfr_labresult | -0.083 | 0.9203 |
| is_na_egfr_labresult* | ทราบค่า eGFR หรือไม่ - ทราบ - ไม่ทราบ | is_na_egfr_labresul t | -2.8812 | 0.0561 |
| Creatinine | ระดับ Creatinine (mg/dL) | creatinine_labresul t | -0.1357 | 0.8731 |
| is_na_creatinine_labresu lt* | ทราบค่า Creatinine หรือไม่ - ทราบ - ไม่ทราบ | is_na_creatinine_la bresult | -2.8208 | 0.0596 |
| UPCR | ระดับโปรตีนใน ปัสสาวะ (mg/g) | upcr_labresult | 0.0018 | 1.0018 |
| is_na_upcr_labresult* | ทราบค่า UPCR หรือไม่ - ทราบ - ไม่ทราบ | is_na_upcr_labres ult | 0.0001 | 1.0001 |

ตารางที่ 47 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ใน
การทำนายความเสี่ยงภาวะไตเสื่อมเรื้อรังหรือตบในผู้ป่วยโรคความดันโลหิตสูง (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ในตัว แบบ | สัมประสิทธิ์ การถดถอย | Odds Ratio |
|---------------------------------|--|--------------------------------|--------------------------|------------|
| Hemoglobin | ระดับฮีโมโกลบินใน เลือด (g/dL) | hemoglobin_labre sult | -0.0203 | 0.98 |
| is_na_hemoglobin_labre sult* | ทราบค่า Hemoglobin หรือไม่ - ทราบ - ไม่ทราบ | is_na_hemoglobin _labresult | -0.461 | 0.6306 |
| Medication | | | | |
| Lipid-lowering med | การใช้ยาลดไขมัน - ใช้ - ไม่ใช่ | flag_lipid_lowering | 0 | 1 |
| Non-insulin med | การใช้ยาเบาหวานที่ไม่ ต้องพึ่งอินซูลิน - ใช้ - ไม่ใช่ | flag_non_insulin_ med | 0 | 1 |
| Insulin med | การใช้ยาเบาหวานที่ ต้องพึ่งอินซูลิน - ใช้ - ไม่ใช่ | flag_insulin_med | 0 | 1 |
| Aspirin | การใช้ยาแอสไพริน - ใช้ - ไม่ใช่ | flag_aspirin_med | 0 | 1 |

ตัวแบบทำนายความเสี่ยงการเสียชีวิตในผู้ป่วยโรคความดันโลหิตสูง

จากกลุ่มตัวอย่างของผู้ป่วยโรคความดันโลหิตสูงจำนวน 236,950 คน ซึ่งประกอบไปด้วยข้อมูลจำนวน 1,517,640 แถว จะถูกสุ่มแบ่งออกเป็นชุดข้อมูลฝึกฝนและชุดข้อมูลทดสอบคิดเป็นร้อยละ 70 และ 30 โดยประมาณ ตามลำดับ โดยที่ชุดข้อมูลฝึกฝนและชุดข้อมูลทดสอบจะมีสัดส่วนของผู้ป่วยโรคความดันโลหิตสูงที่

เสียชีวิตพอ ๆ กัน ซึ่งในที่นี้พบว่ามีค่าเท่ากับ 0.011993 หรือคือมีผู้ป่วยโรคความดันโลหิตสูงคิดเป็นร้อยละ 1.1993 ที่เสียชีวิตนั่นเอง จากการแบ่งข้อมูลจะได้ข้อมูลดังต่อไปนี้

- ชุดข้อมูลฝึกฝน ประกอบด้วยข้อมูลจากผู้เป็นโรคความดันโลหิตสูงจำนวน 165,865 คน ซึ่งมีข้อมูลจำนวน 1,062,246 แถว
- ชุดข้อมูลทดสอบ ประกอบด้วยข้อมูลจากผู้เป็นโรคความดันโลหิตสูงจำนวน 71,085 คน ซึ่งมีข้อมูลจำนวน 455,394 แถว

จากนั้นจึงสร้างตัวแบบจากข้อมูลชุดฝึกฝน ซึ่งจะพิจารณา 2 ตัวแบบ ได้แก่

1. การถดถอยโลจิสติกแบบ Elastic Net วิธีการนี้ขอเรียกว่า “EN”
2. วิธี XGBoost โดยวิธีการนี้ขอเรียกว่า “XGBoost”

หลังจากได้ตัวแบบทั้ง 2 แล้ว จึงนำมาทำนายข้อมูลในชุดทดสอบเพื่อวัดประสิทธิภาพการทำนายโดยพิจารณาจาก

1. AUC
2. Cumulative AUC

ตารางที่ 48 แสดงค่า AUC จาก 2 วิธี ซึ่งพบว่าวิธี XGBoost มีประสิทธิภาพในการทำนายความเสี่ยงการเสียชีวิตในผู้ป่วยโรคความดันโลหิตสูงได้ดีกว่าวิธีการถดถอยโลจิสติกแบบ Elastic Net แม้ว่าค่า AUC ที่ได้จากวิธี XGBoost จะดีกว่าวิธีการถดถอยโลจิสติกแบบ Elastic Net ไม่มาก แต่หากพิจารณาค่า Cumulative AUC ซึ่งวัดประสิทธิภาพการทำนายรายบุคคลผ่านตัววัดผล AUC ที่คำนวณด้วยความน่าจะเป็นสะสมของบุคคลตลอด Cohort (Cumulative Probability) กับสถานะอุบัติการณ์ปีสุดท้าย จะพบว่าวิธี XGBoost มีประสิทธิภาพเหนือกว่าวิธี Elastic Net ค่อนข้างมาก อย่างไรก็ตาม พบว่าวิธีทั้งสองอาจจะมีค่า AUC ที่ยังอยู่ในเกณฑ์ไม่สูงหรือมีค่าต่ำกว่า 0.8 ซึ่งในอนาคตควรจะต้องพิจารณาตัวแบบอื่นๆ รวมทั้งพิจารณาตัวแปรหรือใช้ขนาดตัวอย่างเพิ่มเติม เพื่อที่จะปรับปรุงตัวแบบในการทำนายให้มีประสิทธิภาพมากขึ้น

ตารางที่ 48 ค่า AUC เพื่อวัดประสิทธิภาพการทำนายความเสี่ยงการเสียชีวิตในผู้ป่วยโรคความดันโลหิตสูง (Cohort 2.2) (วิธีที่ดีที่สุดจะแสดงเป็นตัวหนา)

| ตัวแบบ | AUC | Cumulative AUC |
|----------------|---------------|----------------|
| EN | 0.7975 | 0.7534 |
| XGBoost | 0.8197 | 0.7673 |

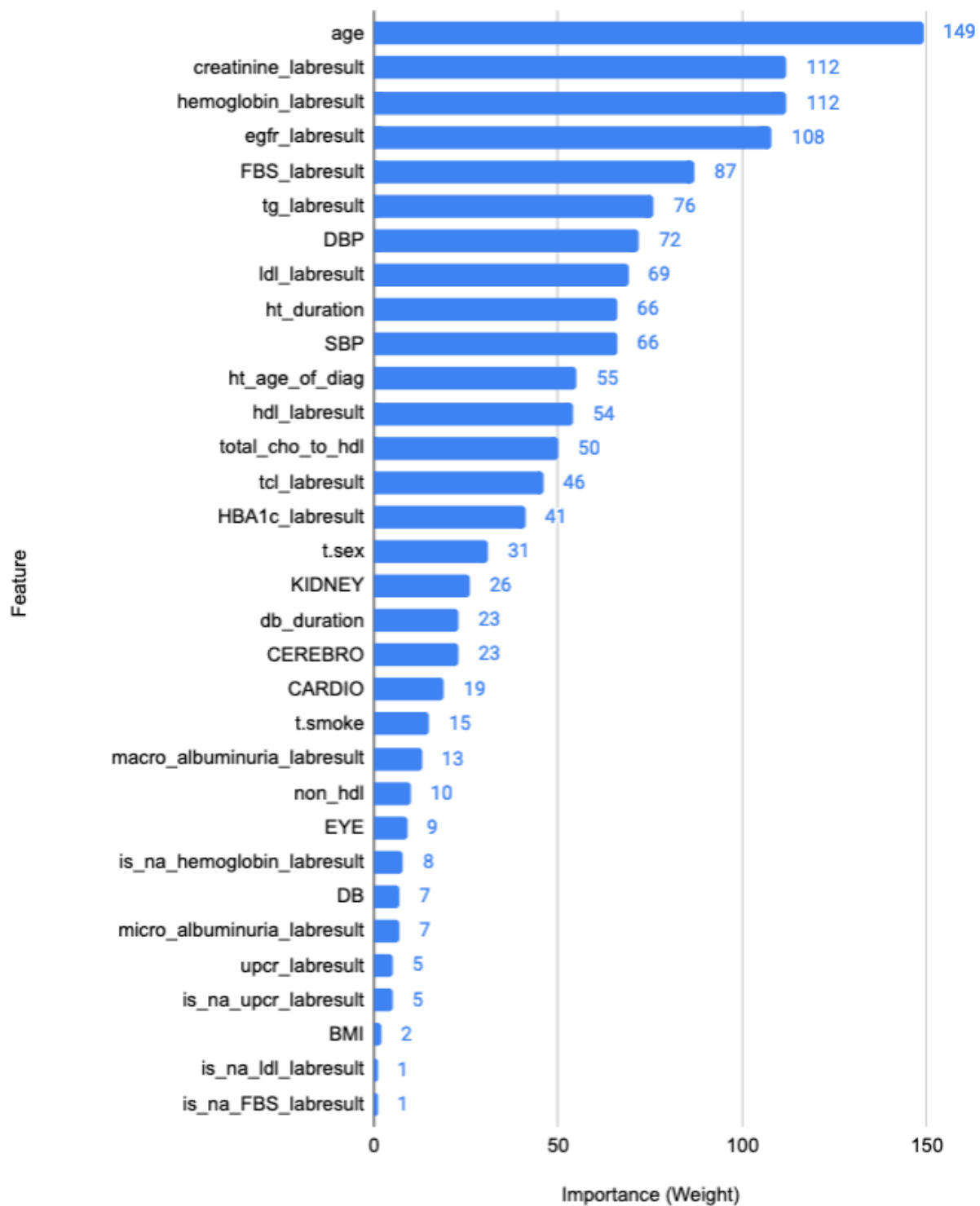
สำหรับการแปลผลตัวแบบโดยอ้างอิงจากตัวแบบ XGBoost ที่ให้ประสิทธิภาพการทำนายที่ดีที่สุดสามารถพิจารณาได้ว่าปัจจัยใดที่มีความสำคัญในการทำนายความเสี่ยงการเกิดโรคความดันโลหิตสูงในประชากรทั่วไปดังแสดงในรูปที่ 69 ซึ่งจะพบว่าตัวแปรที่มีความสำคัญในการทำนายความเสี่ยง 10 อันดับแรก ได้แก่

1. อายุ (ปี) (age)
2. ระดับ Creatinine (mg/dL) (creatinine_labresult)
3. ระดับฮีโมโกลบินในเลือด (g/dL) (hemoglobin_labresult)
4. Estimated Glomerular Filtration Rate (mL/min/1.73m²) (egfr_labresult)
5. Fasting Blood Sugar (mg/dL) (FBS_labresult)
6. ระดับไขมันไตรกลีเซอไรด์ (mg/dL) (tg_labresult)
7. Diastolic Blood Pressure (mm Hg) (DBP)
8. ระดับไขมันไม่ดี (mg/dL) (ldl_labresult)
9. ระยะเวลาที่เป็นความดันโลหิตสูง (ปี) (ht_duration)
10. Systolic Blood Pressure (mm Hg) (SBP)

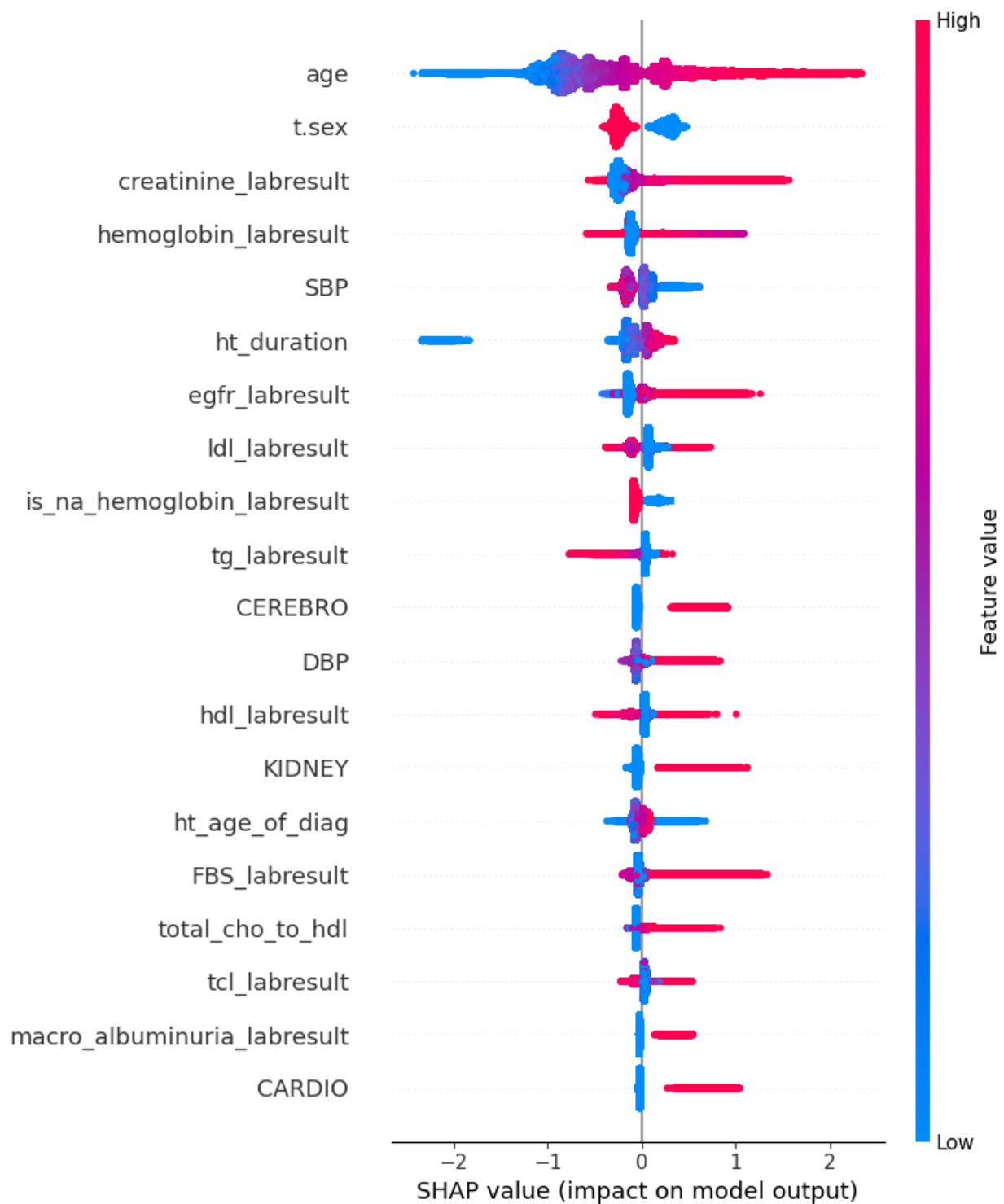
นอกจากนี้ หากพิจารณาคุณค่า SHAP ของตัวแปร 20 ตัวแปรที่มีความสำคัญมากที่สุดในการทำนายการเสียชีวิตในผู้ป่วยความดันโลหิตสูง ดังแสดงในรูปที่ 70 จะสามารถแปลผลในส่วนทิศทางการความสัมพันธ์ของตัวแปรแต่ละตัว ซึ่งสามารถสรุปความสัมพันธ์ที่เด่นชัดได้ ดังนี้

- ผู้ป่วยความดันโลหิตสูงที่มีอายุ (ปี) (age) สูง จะมีความเสี่ยงในการเสียชีวิตสูงขึ้น
- ผู้ป่วยความดันโลหิตสูงที่มีค่าระดับ Creatinine (mg/dL) (creatinine_labresult) สูง จะมีแนวโน้มที่จะมีความเสี่ยงในการมีการเสียชีวิตสูง
- ผู้ป่วยความดันโลหิตสูงที่มีค่าระดับฮีโมโกลบินในเลือด (g/dL) (hemoglobin_labresult) สูง จะมีแนวโน้มที่จะมีความเสี่ยงในการมีการเสียชีวิตสูง
- ผู้ป่วยความดันโลหิตสูงที่มีค่า Estimated Glomerular Filtration Rate (mL/min/1.73m²) (egfr_labresult) สูง จะมีแนวโน้มที่จะมีความเสี่ยงในการมีการเสียชีวิตสูง

แม้ว่าวิธี XGBoost สามารถระบุความสำคัญของตัวแปรทำนายที่ใช้ในตัวแบบได้ แต่ไม่สามารถแปลผลตัวแบบในรูปของอัตราเสี่ยงได้ ในขณะที่วิธีการถดถอยโลจิสติกแบบ Elastic Net ให้ค่าสัมประสิทธิ์การถดถอยเช่นเดียวกับตัวแบบการถดถอยโลจิสติกแบบปกติ ดังนั้นจึงสามารถแปลผลตัวแบบในรูปของอัตราเสี่ยงได้ ตารางที่ 49 แสดงผลลัพธ์ที่ได้ของสัมประสิทธิ์การถดถอยของตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ในการทำนายความเสี่ยงการเสียชีวิตในผู้ป่วยโรคความดันโลหิตสูง



รูปที่ 69 ความสำคัญของตัวแปรทำนายจากวิธี XGBoost ในการทำนายการเสียชีวิตในผู้ป่วยโรคความดันโลหิตสูง



รูปที่ 70 ค่า SHAP จากตัวแบบ XGBoost ในการทำนายความเสี่ยงการเสียชีวิตในผู้ป่วยโรคความดันโลหิตสูง

ตารางที่ 49 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ใน
การทำนายความเสี่ยงภาวะเส้นเลือดสมองแตกหรือตีบในผู้ป่วยโรคความดันโลหิตสูง

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ใน ตัวแบบ | สัมประสิทธิ์ การถดถอย | Odds Ratio |
|-----------------------------|---|------------------------------|--------------------------|------------|
| Demographic | | | | |
| Age | อายุ (ปี) | age | 0.0723 | 1.075 |
| Gender | เพศ - ชาย - หญิง | t.sex | -0.4614 | 0.6304 |
| Health Behavior | | | | |
| Smoking | พฤติกรรมการสูบบุหรี่ - สูบ - ไม่สูบ | t.smoke | 0.2677 | 1.3069 |
| Physical measurement | | | | |
| BMI | ดัชนีมวลกาย | BMI | 0.0016 | 1.0016 |
| SBP | Systolic Blood Pressure (mm Hg) | SBP | -0.0068 | 0.9933 |
| DBP | Diastolic Blood Pressure (mm Hg) | DBP | 0.0069 | 1.0069 |
| Underlying Disease | | | | |
| HT Duration | ระยะเวลาที่เป็นความดัน โลหิตสูง (ปี) | ht_duration | -0.0008 | 0.9992 |
| HT Age of Diag | อายุที่ได้รับการวินิจฉัยว่า เป็นโรคความดันโลหิตสูง (ปี) | ht_age_of_diag | -0.0071 | 0.9929 |
| DLP | สถานะภาวะไขมัน ผิดปกติ - เป็น - ไม่เป็น | flag_dlp | 0 | 1 |
| DM | สถานะภาวะเบาหวาน - เป็น - ไม่เป็น | DB | 0.0722 | 1.0749 |
| DM Duration | ระยะเวลาที่เป็น โรคเบาหวาน (ปี) | db_duration | -0.0006 | 0.9994 |

ตารางที่ 49 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ใน
การทำนายความเสี่ยงภาวะเส้นเลือดสมองแตกหรือตีบในผู้ป่วยโรคความดันโลหิตสูง (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ใน ตัวแบบ | สัมประสิทธิ์ การถดถอย | Odds Ratio |
|---------------|---|------------------------------|--------------------------|------------|
| Heart Disease | สถานะการเป็นโรคหัวใจ ซึ่ง รวมถึงการเป็นโรคหัวใจเต้น ผิดปกติ | | | |
| | - เป็น | flag_HEART | 0 | 1 |
| | - ไม่เป็น | | | |
| CKD | สถานะการเป็นโรคไตเรื้อรัง โดยดูจากระดับน้ำตาลที่วินิจฉัย - เป็น - ไม่เป็น | flag_CKD | 0 | 1 |
| KIDNEY | สถานะการเป็นโรคไตเรื้อรังใน ระยะสุดท้าย (พิจารณาจาก CKD และค่า eGFR) - เป็น - ไม่เป็น | KIDNEY | 0.8686 | 2.3837 |
| Eye | สถานะการเป็นโรคตาอัน เนื่องมาจากเบาหวาน - เป็น - ไม่เป็น | EYE | 0.2674 | 1.3066 |
| flag_eye | สถานะการเป็นโรคตา พิจารณาจากระดับน้ำตาลที่วินิจฉัย - เป็น - ไม่เป็น | flag_eye | 0 | 1 |
| COPD | สถานะการเป็นโรคถุงลมโป่ง พอง - เป็น - ไม่เป็น | flag_copd | 0 | 1 |
| Lab | | | | |
| TC | ระดับคอเลสเตอรอล (mg/dL) | tcl_labresult | 0 | 1 |

ตารางที่ 49 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ใน
การทำนายความเสี่ยงภาวะเส้นเลือดสมองแตกหรือตีบในผู้ป่วยโรคความดันโลหิตสูง (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ใน ตัวแบบ | สัมประสิทธิ์ การถดถอย | Odds Ratio |
|--------------------------|--|------------------------------|--------------------------|------------|
| is_na_tcl_labresult* | ทราบค่า TC หรือไม่ - ทราบ - ไม่ทราบ | is_na_tcl_labresu lt | -0.0504 | 0.9509 |
| HDL | ระดับไขมันดี (mg/dL) | hdl_labresult | -0.0027 | 0.9973 |
| is_na_hdl_labresult * | ทราบค่า HDL หรือไม่ - ทราบ - ไม่ทราบ | is_na_hdl_labres ult | -0.1675 | 0.8458 |
| LDL | ระดับไขมันไม่ดี (mg/dL) | ldl_labresult | 0.0011 | 1.0011 |
| is_na_ldl_labresult* | ทราบค่า HDL หรือไม่ - ทราบ - ไม่ทราบ | is_na_ldl_labresu lt | 0.6309 | 1.8792 |
| TG | ระดับไขมันไตรกลีเซอไรด์ (mg/dL) | tg_labresult | -0.0012 | 0.9988 |
| is_na_tg_labresult* | ทราบค่า TG หรือไม่ - ทราบ - ไม่ทราบ | is_na_tg_labresul t | -0.0542 | 0.9473 |
| non-HDL | TC – HDL | non_hdl | -0.0029 | 0.9972 |
| is_na_non_hdl* | ทราบค่า non-HDL หรือไม่ - ทราบ - ไม่ทราบ | is_na_non_hdl | 0.1183 | 1.1256 |
| total Cho:HDL | TC/HDL | total_cho_to_hdl | 0.136 | 1.1457 |

ตารางที่ 49 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ใน
การทำนายความเสี่ยงภาวะเส้นเลือดสมองแตกหรือตีบในผู้ป่วยโรคความดันโลหิตสูง (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ใน ตัวแบบ | สัมประสิทธิ์ การถดถอย | Odds Ratio |
|-------------------------|---|------------------------------|--------------------------|------------|
| is_na_total_cho_to_hdl* | ทราบค่า total Cho:HDL หรือไม่ - ทราบ - ไม่ทราบ | is_na_total_cho_to_hdl | 0.1591 | 1.1725 |
| HBA1c | ค่าน้ำตาลเฉลี่ยสะสมในเลือด (mmol/mol) | HBA1c_labresult | 0.0201 | 1.0203 |
| is_na_HBA1C_labresult* | ทราบค่า HBA1c หรือไม่ - ทราบ - ไม่ทราบ | is_na_HBA1C_labresult | 0.1737 | 1.1897 |
| FBS | Fasting Blood Sugar (mg/dL) | FBS_labresult | 0.0036 | 1.0036 |
| is_na_FBS_labresult* | ทราบค่า FBS หรือไม่ - ทราบ - ไม่ทราบ | is_na_FBS_labresult | 0.4273 | 1.5331 |
| Macro-albuminuria | สถานะความเสี่ยงการเป็นโรค ไตโดยวัดจาก Macro- albuminuria - เป็น - ไม่เป็น | macro_albuminuria_labresult | 0.3679 | 1.4447 |
| Micro-albuminuria | สถานะความเสี่ยงการเป็นโรค ไตโดยวัดจาก Micro- albuminuria - เป็น - ไม่เป็น | micro_albuminuria_labresult | 0.0359 | 1.0366 |

ตารางที่ 49 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ใน
การทำนายความเสี่ยงภาวะเส้นเลือดสมองแตกหรือตีบในผู้ป่วยโรคความดันโลหิตสูง (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ใน ตัวแบบ | สัมประสิทธิ์ การถดถอย | Odds Ratio |
|---------------------------------|---|--------------------------------|--------------------------|------------|
| eGFR | Estimated Glomerular Filtration Rate (mL/min/1.73m ²) | egfr_labresult | -0.0048 | 0.9952 |
| is_na_egfr_labresult * | ทราบค่า eGFR หรือไม่ - ทราบ - ไม่ทราบ | is_na_egfr_labres ult | -0.5358 | 0.5852 |
| Creatinine | ระดับ Creatinine (mg/dL) | creatinine_labres ult | 0.0275 | 1.0279 |
| is_na_creatinine_lab result* | ทราบค่า Creatinine หรือไม่ - ทราบ - ไม่ทราบ | is_na_creatinine_ labresult | -0.5357 | 0.5853 |
| UPCR | ระดับโปรตีนในปัสสาวะ (mg/g) | upcr_labresult | 0.0017 | 1.0017 |
| is_na_upcr_labresul t* | ทราบค่า UPCR หรือไม่ - ทราบ - ไม่ทราบ | is_na_upcr_labre sult | 0.1678 | 1.1827 |
| Hemoglobin | ระดับฮีโมโกลบินในเลือด (g/dL) | hemoglobin_labr esult | -0.0944 | 0.9099 |
| is_na_hemoglobin_l abresult* | ทราบค่า Hemoglobin หรือไม่ - ทราบ - ไม่ทราบ | is_na_hemoglobi n_labresult | -1.7387 | 0.1757 |

ตารางที่ 49 สัมประสิทธิ์การถดถอยและ Odds Ratio จากตัวแบบการถดถอยโลจิสติกแบบ Elastic Net ใน
การทำนายความเสี่ยงภาวะเส้นเลือดสมองแตกหรือตีบในผู้ป่วยโรคความดันโลหิตสูง (ต่อ)

| ตัวแปร | รายละเอียด | ชื่อตัวแปรที่ใช้ในตัวแบบ | สัมประสิทธิ์การถดถอย | Odds Ratio |
|--------------------------|--|--------------------------|----------------------|------------|
| Albumin | ระดับ Albumin (g/dL) | albumin_labresult | -0.1432 | 0.8666 |
| is_na_albumin_labresult* | ทราบค่า Albumin หรือไม่ - ทราบ - ไม่ทราบ | is_na_albumin_labresult | -0.6317 | 0.5317 |
| Medication | | | | |
| Lipid-lowering med | การใช้ยาลดไขมัน - ใช้ - ไม่ใช่ | flag_lipid_lowering | 0 | 1 |
| Non-insulin med | การใช้ยาเบาหวานที่ไม่ต้องพึ่งอินซูลิน - ใช้ - ไม่ใช่ | flag_non_insulin_med | 0 | 1 |
| Insulin med | การใช้ยาเบาหวานที่ต้องพึ่งอินซูลิน - ใช้ - ไม่ใช่ | flag_insulin_med | -0.5589 | 0.5719 |
| Aspirin | การใช้ยาแอสไพริน - ใช้ - ไม่ใช่ | flag_aspirin_med | 0 | 1 |

ผลกระทบทางเศรษฐกิจของโรคเบาหวานและความดันโลหิตสูง

จากข้อมูลค่าใช้จ่ายสุขภาพที่มีการบันทึกไว้ในฐานข้อมูล 43 แฟ้มระยะเวลา 6 ปี ระหว่างเดือนมกราคม พ.ศ. 2560 ถึง ธันวาคม พ.ศ. 2565 พบว่า มีกลุ่มตัวอย่างที่เข้ารับการรักษาแบบผู้ป่วยนอกและผู้ป่วยในรวมทั้งสิ้น 1,189,228 และ 309,934 คนและมีจำนวนการรับบริการผู้ป่วยนอกและผู้ป่วยในรวมทั้งสิ้น 37,790,574 และ 691,377 ครั้ง แบ่งออกเป็นตามประเภทของผู้ป่วยดังตารางที่ 50

ตารางที่ 50 จำนวนผู้ใช้บริการและการให้บริการผู้ป่วยนอกและผู้ป่วยในระหว่างปี พ.ศ.2560 ถึง 2565
แยกรายประเภทผู้ป่วย

| ประเภท | จำนวน ผู้ใช้บริการผู้ป่วย นอก (คน) | จำนวนการรับ บริการผู้ป่วย นอก (ครั้ง) | จำนวน ผู้ใช้บริการผู้ป่วย ใน (คน) | จำนวนการรับ บริการผู้ป่วยใน (ครั้ง) |
|---|--|---|---|---|
| ผู้ที่ไม่เป็นโรคเบาหวาน และความดันโลหิตสูง | 1,001,258 | 19,336,803 | 172,244 | 309,303 |
| ผู้ป่วยโรคเบาหวานอย่าง เดียว | 48,225 | 1,482,571 | 13,374 | 26,236 |
| ผู้ป่วยโรคความดันโลหิต สูงอย่างเดียว | 273,265 | 11,278,739 | 94,196 | 218,543 |
| ผู้ป่วยที่เป็นทั้ง 2 โรค | 99,493 | 5,692,461 | 45,704 | 137,295 |
| รวมทั้งสิ้น | 1,189,228 | 37,790,574 | 309,934 | 691,377 |

หมายเหตุ: เมื่อนำจำนวนผู้ป่วยของทั้ง 4 กลุ่มที่แสดงในตารางมารวมกันจะได้ค่ามากกว่าจำนวนผู้ป่วยทั้งหมด เนื่องจากกระหว่างการติดตาม ผู้ป่วยสามารถเปลี่ยนสถานะได้ เช่น จากไม่เป็นโรคเบาหวานเป็นเกิดโรค ส่งผลให้ผู้ป่วย 1 รายสามารถอยู่ได้มากกว่า 1 ประเภท

หัวข้อที่ 1 ต้นทุนทางตรง (Direct Cost)

หัวข้อที่ 1.1. ต้นทุนทางตรงทางการแพทย์ (Direct Medical Cost)

การศึกษานี้ใช้ค่ารักษาพยาบาลที่โรงพยาบาลเรียกเก็บเป็นตัวแทนของต้นทุนทางตรงทางการแพทย์ จากการวิเคราะห์ พบว่า กลุ่มตัวอย่างมีการใช้บริการผู้ป่วยนอกและผู้ป่วยในรวมถึงมีค่ารักษาพยาบาลที่เกี่ยวข้องแตกต่างกันขึ้นอยู่กับสถานะของโรคและจำนวนภาวะแทรกซ้อนดังตารางผนวกที่ 1 และ 2

หัวข้อที่ 1.1.1. ค่ารักษาพยาบาลผู้ป่วยนอก

จากการวิเคราะห์ พบว่า ผู้ที่ไม่เป็นโรคเบาหวานและความดันโลหิตสูงจะมีค่ารักษาพยาบาลผู้ป่วยนอกเฉลี่ย (SD) อยู่ที่ 447 (3,753) บาทต่อครั้ง และ มีการใช้บริการเฉลี่ยอยู่ที่ 4.5 ครั้งต่อปี หรือคิดเป็นค่ารักษาพยาบาลผู้ป่วยนอกเฉลี่ย 2,030 บาทต่อคนต่อปี สำหรับผู้ป่วยที่เป็นโรคเบาหวาน และ/หรือ ความดันโลหิตสูงจะมีค่ารักษาพยาบาลผู้ป่วยนอกเฉลี่ยต่อครั้งสูงกว่า และ ต้องมารับบริการที่โรงพยาบาลมากกว่า

กลุ่มที่ไม่เป็นโรค โดยค่าบริการผู้ป่วยนอกเฉลี่ยต่อคนต่อปีของผู้ที่เป็นโรคเบาหวานอย่างเดียว เป็นโรคความดันโลหิตสูงอย่างเดียว และเป็นทั้ง 2 โรคเฉลี่ยอยู่ที่ 6,563 6,696 และ 12,021 บาท ตามลำดับดังตารางที่ 51

ตารางที่ 51 จำนวนการรับบริการผู้ป่วยนอกและค่าบริการผู้ป่วยนอกเฉลี่ยต่อคนต่อปีแยกประเภทกลุ่มผู้ป่วย

| ประเภท | ค่าบริการผู้ป่วยนอกเฉลี่ย (SD) ต่อครั้ง (บาท) | จำนวนการมารับบริการผู้ป่วยนอกเฉลี่ยต่อปี (ครั้ง) | ค่าบริการผู้ป่วยนอกเฉลี่ยต่อคนต่อปี (บาท) |
|---|---|--|---|
| ผู้ที่ไม่เป็นโรคเบาหวาน และ ความดันโลหิตสูง | 447 (3,753) | 4.5 | 2,030 |
| ผู้ป่วยโรคเบาหวานอย่างเดียว | 636 (4,368) | 10.3 | 6,563 |
| ผู้ป่วยโรคความดันโลหิตสูงอย่างเดียว | 690 (5,292) | 9.7 | 6,696 |
| ผู้ป่วยที่เป็นทั้ง 2 โรค | 893 (5,411) | 13.5 | 12,021 |

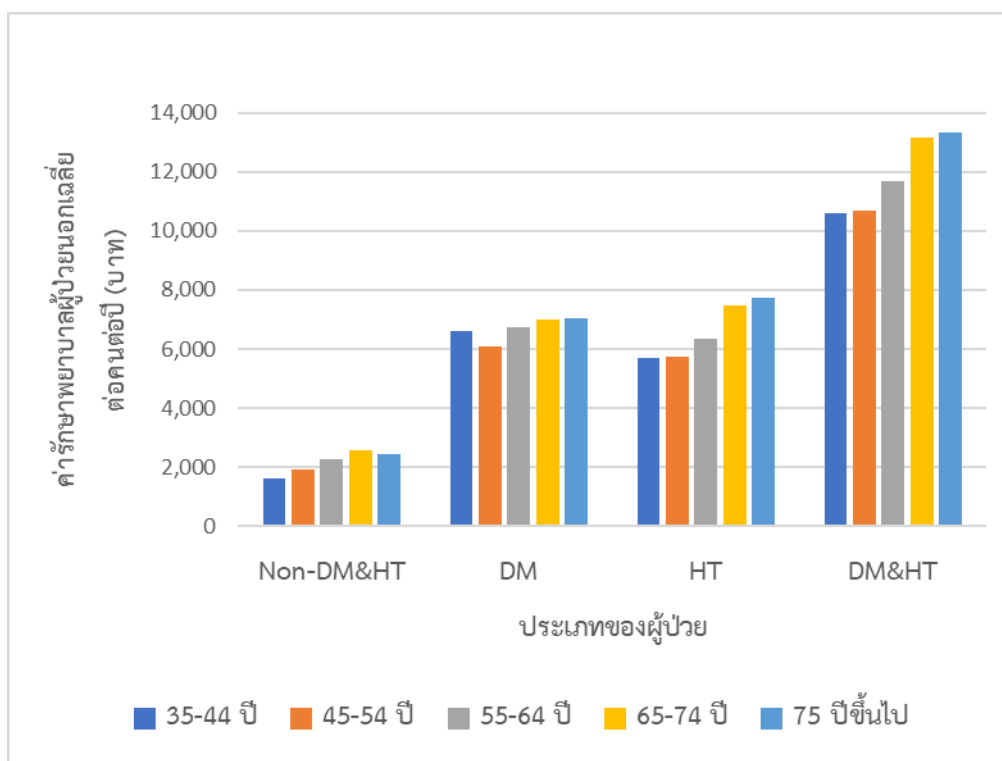
เมื่อวิเคราะห์แยกชายเพศและช่วงอายุของผู้ป่วย พบว่า เพศชายมีค่ารักษาพยาบาลผู้ป่วยนอกเฉลี่ยต่อคนต่อปีสูงกว่าเพศหญิงในทุกประเภทของผู้ป่วยยกเว้นในกลุ่มผู้ที่ไม่เป็นโรคเบาหวานและความดันโลหิตสูงซึ่งเพศชายและหญิงมีค่ารักษาพยาบาลผู้ป่วยนอกเฉลี่ยต่อคนต่อปีใกล้เคียงกันดังตารางที่ 52 นอกจากนี้ ค่ารักษาพยาบาลผู้ป่วยนอกเฉลี่ยต่อคนต่อปีมีแนวโน้มเพิ่มสูงขึ้นตามช่วงอายุโดยเห็นได้ชัดเจนที่สุดในกลุ่มผู้ป่วยที่ไม่เป็นโรคเบาหวานและความดันโลหิตสูง โดยค่ารักษาพยาบาลผู้ป่วยนอกของกลุ่มอายุ 75 ปีขึ้นไปมีค่าสูงกว่าค่ารักษาพยาบาลผู้ป่วยนอกของกลุ่มอายุ 35-44 ปีอยู่ถึง 51% เมื่อประเมินในลักษณะเดียวกันสำหรับผู้ป่วยโรคเบาหวาน ผู้ป่วยโรคความดันโลหิตสูง และผู้ป่วยที่เป็นทั้ง 2 โรค พบว่าความแตกต่างของค่ารักษาพยาบาลผู้ป่วยนอกระหว่างกลุ่มอายุสูงและต่ำสุดอยู่ที่ 7 37 และ 26% ตามลำดับ ดังรูปที่ 71

ตารางที่ 52 จำนวนการรับบริการผู้ป่วยนอกและค่าบริการผู้ป่วยนอกเฉลี่ยต่อคนต่อปีแยกชายเพศ

| ประเภท | ค่าบริการผู้ป่วยนอกเฉลี่ย (SD) ต่อครั้ง (บาท) | จำนวนการมารับบริการผู้ป่วยนอกเฉลี่ยต่อปี (ครั้ง) | ค่าบริการผู้ป่วยนอกเฉลี่ยต่อคนต่อปี (บาท) |
|--|---|--|---|
| เพศชาย | | | |
| ผู้ที่ไม่เป็นโรคเบาหวาน และความดันโลหิตสูง | 499 (4,515) | 4.0 | 2,000 |
| ผู้ป่วยโรคเบาหวานอย่างเดียว | 745 (5,494) | 9.7 | 7,262 |
| ผู้ป่วยโรคความดันโลหิตสูงอย่างเดียว | 835 (6,579) | 9.4 | 7,821 |
| ผู้ป่วยที่เป็นทั้ง 2 โรค | 1,040 (6,593) | 13.4 | 13,963 |

ตารางที่ 52 จำนวนการรับบริการผู้ป่วยนอกและค่าบริการผู้ป่วยนอกเฉลี่ยต่อคนต่อปีแยกรายเพศ (ต่อ)

| ประเภท | ค่าบริการผู้ป่วยนอกเฉลี่ย (SD) ต่อครั้ง (บาท) | จำนวนการมารับบริการผู้ป่วยนอกเฉลี่ยต่อปี (ครั้ง) | ค่าบริการผู้ป่วยนอกเฉลี่ยต่อคนต่อปี (บาท) |
|--|---|--|---|
| เพศหญิง | | | |
| ผู้ที่ไม่เป็นโรคเบาหวาน และความดันโลหิตสูง | 411 (3,111) | 5.0 | 2,056 |
| ผู้ป่วยโรคเบาหวานอย่างเดียว | 571 (3,533) | 10.7 | 6,108 |
| ผู้ป่วยโรคความดันโลหิตสูงอย่างเดียว | 595 (4,222) | 9.9 | 5,903 |
| ผู้ป่วยที่เป็นทั้ง 2 โรค | 808 (4,596) | 13.5 | 10,901 |



รูปที่ 71 ค่ารักษาพยาบาลผู้ป่วยนอกเฉลี่ยต่อคนต่อปี (บาท) แยกรายประเภทผู้ป่วยและช่วงอายุ

เมื่อพิจารณาถึงปัจจัยเรื่องภาวะแทรกซ้อน (ตารางผนวกที่ 1) พบว่า การเพิ่มขึ้นของภาวะแทรกซ้อนแต่ละประเภทส่งผลให้ค่ารักษาพยาบาลผู้ป่วยนอกเฉลี่ยต่อคนต่อปีที่สูงขึ้น โดยภาวะแทรกซ้อนแต่ละประเภทให้ผลกระทบที่แตกต่างกันเรียงลำดับจากมากไปน้อย ได้แก่ โรคหลอดเลือดหัวใจ โรคเส้นเลือดสมองแตกหรือตีบ โรคไตเสื่อมเรื้อรัง และ ภาวะแทรกซ้อนทางตา ดังตารางที่ 53

อย่างไรก็ตาม ความเป็นไปได้ของการจัดกลุ่มปัจจัยแทรกซ้อนมีได้หลายรูปแบบ (2 โรคหลัก และ 4 ภาวะแทรกซ้อน) ส่งผลให้กลุ่มย่อยบางประเภทมีจำนวนผู้ป่วยที่เข้าเกณฑ์และจำนวนครั้งที่มาใช้บริการน้อย

หรือไม่มีเลย จากเหตุการณ์ดังกล่าวอาจทำให้ข้อมูลค่าใช้จ่ายบางประเภทมีความผันผวนและอาจไม่ตรงกับความเป็นจริง เพื่อให้ง่ายต่อการประเมิน ทางผู้วิจัยจึงจัดกลุ่มค่าใช้จ่ายของผู้ป่วยโรคเบาหวานและ/หรือ ความดันโลหิตสูงตามจำนวนภาวะแทรกซ้อนดังตารางที่ 55 จากข้อมูล พบว่า เมื่อผู้ป่วยที่มีจำนวนภาวะแทรกซ้อนเพิ่มขึ้น จำนวนการรับบริการผู้ป่วยนอกและค่ารักษาพยาบาลผู้ป่วยนอกจะสูงขึ้นตามลำดับและผู้ที่เป็นทั้งโรคเบาหวานและความดันโลหิตสูงจะมีการใช้บริการผู้ป่วยนอกและค่าใช้จ่ายสูงกว่าผู้ที่ป่วยเพียงโรคเดียว

ตารางที่ 53 สัมประสิทธิ์การถดถอยจากตัวแบบการถดถอยลิเนียร์ในการอธิบาย
ค่ารักษาพยาบาลผู้ป่วยนอกต่อครั้ง (บาท)

| กลุ่มโรค | สัมประสิทธิ์ | 95% Confidence Interval |
|----------------------------|--------------|-------------------------|
| โรคเบาหวาน | 226 | 218 to 234 |
| โรคความดันโลหิตสูง | 224 | 219 to 230 |
| โรคหลอดเลือดหัวใจตีบ | 860 | 833 to 886 |
| โรคเส้นเลือดสมองแตกหรือตีบ | 419 | 403 to 435 |
| โรคไตเสื่อมเรื้อรัง | 353 | 338 to 366 |
| ภาวะแทรกซ้อนทางตา | 247 | 227 to 268 |
| Constant | 382 | 379 to 385 |

ตารางที่ 54 จำนวนการรับบริการผู้ป่วยนอกและค่าบริการผู้ป่วยนอกเฉลี่ยต่อปีแยกรายจำนวนภาวะแทรกซ้อน

| ประเภทของผู้ป่วย | จำนวนการรับ บริการผู้ป่วยนอก (ครั้ง) | ค่าบริการผู้ป่วย นอกเฉลี่ย (SD) ต่อครั้ง (บาท) | จำนวนการรับ บริการผู้ป่วยนอก เฉลี่ยต่อปี (ครั้ง) | ค่าบริการผู้ป่วย นอกเฉลี่ยต่อคน ต่อปี (บาท) |
|---|--|--|--|---|
| เป็นโรคเบาหวานอย่างเดียว | | | | |
| ไม่มีภาวะแทรกซ้อน | 1,254,519 | 609 (4,235) | 10.2 | 6,198 |
| มีภาวะแทรกซ้อน 1 ประเภท | 211,951 | 772 (4,526) | 11.2 | 8,615 |
| มีภาวะแทรกซ้อน 2 ประเภท | 15,563 | 937 (9,545) | 11.5 | 10,729 |
| มีภาวะแทรกซ้อน 3 ประเภท | 537 | 1,129 (6,822) | 12.5 | 14,096 |
| มีภาวะแทรกซ้อน 4 ประเภท* | - | - | - | - |
| เป็นโรคความดันโลหิตสูงอย่างเดียว | | | | |
| ไม่มีภาวะแทรกซ้อน | 9,736,475 | 592 4,829 | 9.2 | 5,450 |
| มีภาวะแทรกซ้อน 1 ประเภท | 1,390,580 | 1,292 7,430 | 14.4 | 18,615 |
| มีภาวะแทรกซ้อน 2 ประเภท | 106,808 | 1,799 9,558 | 21.1 | 37,910 |
| มีภาวะแทรกซ้อน 3 ประเภท | 5,226 | 1,802 9,086 | 29.5 | 53,211 |
| มีภาวะแทรกซ้อน 4 ประเภท | 50 | 1,604 3,139 | 16.7 | 26,737 |

ตารางที่ 54 จำนวนการรับบริการผู้ป่วยนอกและค่าบริการผู้ป่วยนอกเฉลี่ยต่อปี
แยกรายจำนวนภาวะแทรกซ้อน (ต่อ)

| ประเภทของผู้ป่วย | จำนวนการรับ บริการผู้ป่วยนอก (ครั้ง) | ค่าบริการผู้ป่วย นอกเฉลี่ย (SD) ต่อครั้ง (บาท) | จำนวนการรับ บริการผู้ป่วยนอก เฉลี่ยต่อปี (ครั้ง) | ค่าบริการผู้ป่วย นอกเฉลี่ยต่อคน ต่อปี (บาท) |
|---|--|--|--|---|
| เป็นทั้งโรคเบาหวานและความดันโลหิตสูง ไม่มีภาวะแทรกซ้อน | 3,719,454 | 742 (4,470) | 12.2 | 9,021 |
| มีภาวะแทรกซ้อน 1 ประเภท | 1,546,390 | 1,066 (6,271) | 15.7 | 16,711 |
| มีภาวะแทรกซ้อน 2 ประเภท | 376,988 | 1,524 (8,353) | 22.8 | 34,805 |
| มีภาวะแทรกซ้อน 3 ประเภท | 47,365 | 1,959 (9,754) | 31.8 | 62,325 |
| มีภาวะแทรกซ้อน 4 ประเภท | 2,264 | 2,547 (13,456) | 32.3 | 82,390 |

* ไม่มีข้อมูลการใช้บริการผู้ป่วยนอกหรือมีจำนวนการใช้บริการผู้ป่วยนอกน้อยกว่า 5 ครั้ง ส่งผลให้ข้อมูลที่คำนวณได้อาจไม่สามารถเป็นตัวแทนของกลุ่มประชากรได้

หัวข้อที่ 1.1.2. ค่ารักษาพยาบาลผู้ป่วยใน

สำหรับการรักษาพยาบาลแบบผู้ป่วยใน พบว่า ผู้ที่ไม่เป็นโรคเบาหวานและความดันโลหิตสูงจะมีค่ารักษาพยาบาลผู้ป่วยในเฉลี่ย (SD) อยู่ที่ 18,357 (48,527) บาทต่อครั้ง และ มีการใช้บริการเฉลี่ยอยู่ที่ 0.07 ครั้งต่อปีหรือคิดเป็นค่ารักษาพยาบาลผู้ป่วยในเฉลี่ย 1,333 บาทต่อคนต่อปี สำหรับผู้ป่วยที่เป็นโรคเบาหวาน และ/หรือความดันโลหิตสูงจะมีค่ารักษาพยาบาลผู้ป่วยในเฉลี่ยต่อครั้งสูงกว่า และ ต้องมารับบริการที่โรงพยาบาลมากกว่ากลุ่มที่ไม่เป็นโรคเช่นเดียวกับการรับบริการผู้ป่วยนอกโดยค่าบริการผู้ป่วยในเฉลี่ยต่อคนต่อปีของผู้ที่เป็นโรคเบาหวานอย่างเดียว เป็นโรคความดันโลหิตสูงอย่างเดียว และเป็นทั้ง 2 โรคเฉลี่ยอยู่ที่ 3,605 4,091 และ 6,866 บาท ตามลำดับดังตารางที่ 55

ตารางที่ 55 จำนวนการรับบริการผู้ป่วยในและค่าบริการผู้ป่วยในเฉลี่ยต่อคนต่อปีแยกรายประเภทกลุ่มผู้ป่วย

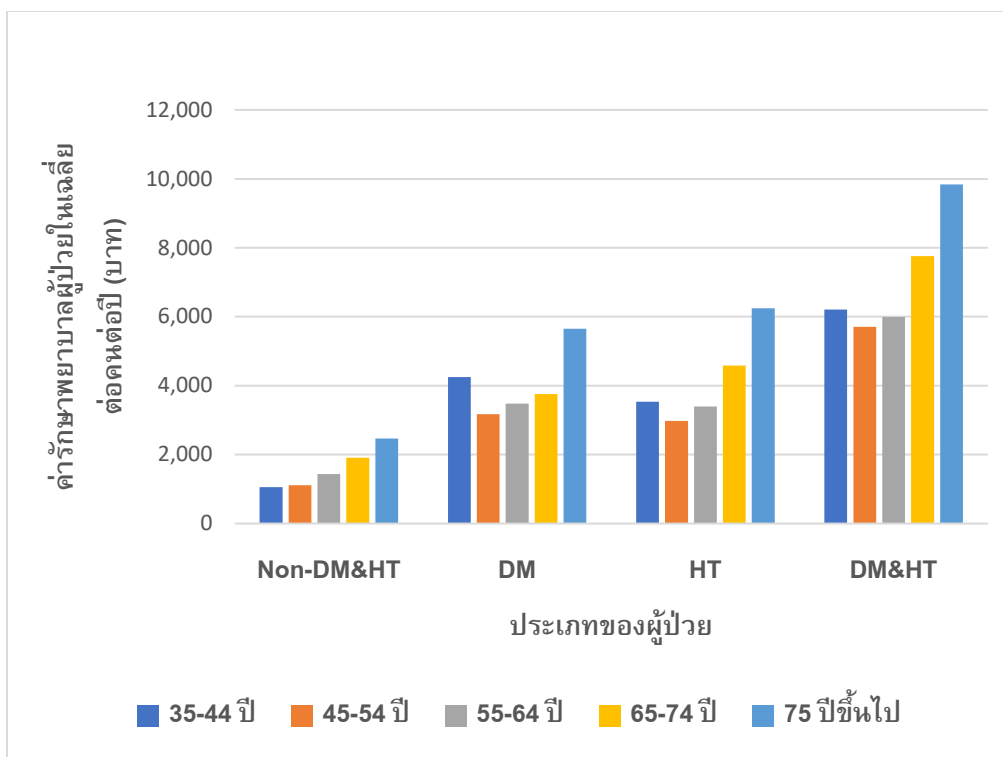
| ประเภท | จำนวนวันนอน เฉลี่ย (SD) ต่อ ครั้ง (วัน) | ค่าบริการผู้ป่วย ในเฉลี่ย (SD) ต่อครั้ง (บาท) | จำนวนการรับ บริการผู้ป่วยใน เฉลี่ยต่อปี (ครั้ง) | ค่าบริการผู้ป่วย ในเฉลี่ยต่อคนต่อ ปี (บาท) |
|--|---|---|---|--|
| ผู้ที่ไม่เป็นโรคเบาหวาน และความดันโลหิตสูง | 4.1 (5.4) | 18,357 (48,527) | 0.07 | 1,333 |
| ผู้ป่วยโรคเบาหวานอย่างเดียว | 4.6 (6.0) | 19,745 (40,684) | 0.18 | 3,605 |
| ผู้ป่วยโรคความดันโลหิตสูงอย่างเดียว | 4.4 (6.2) | 21,693 (52,192) | 0.19 | 4,091 |
| ผู้ป่วยที่เป็นทั้ง 2 โรค | 4.4 (5.8) | 21,144 (47,808) | 0.32 | 6,866 |

เมื่อวิเคราะห์แยกรายเพศและช่วงอายุของผู้ป่วย พบว่า เพศชายมีค่ารักษาพยาบาลผู้ป่วยในเฉลี่ยต่อคนต่อปีสูงกว่าเพศหญิงในทุกประเภทของผู้ป่วยดังตารางที่ 56 นอกจากนี้ ค่ารักษาพยาบาลผู้ป่วยในเฉลี่ยต่อคนต่อปี

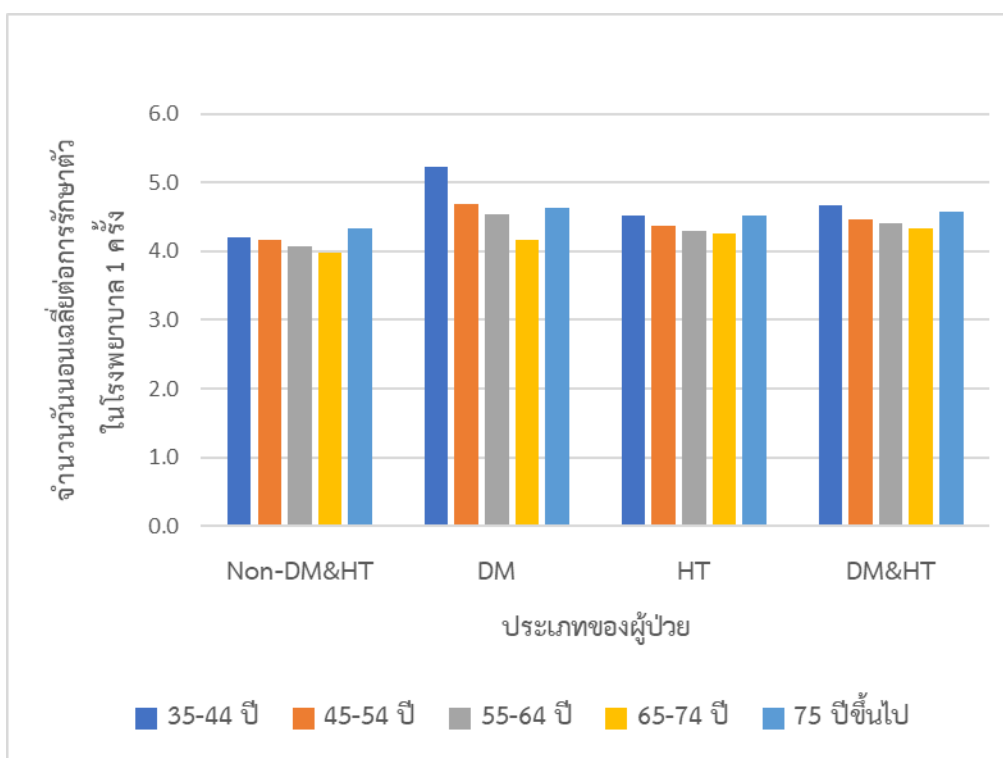
มีแนวโน้มเพิ่มสูงขึ้นตามช่วงอายุและมีลักษณะเช่นเดียวกับค่ารักษาพยาบาลผู้ป่วยนอก โดยความแตกต่างของค่ารักษาพยาบาลผู้ป่วยในระหว่างกลุ่มอายุสูงและต่ำสุดของผู้ที่ไม่เป็นโรค ผู้ป่วยโรคเบาหวาน ผู้ป่วยโรคความดันโลหิตสูง และ ผู้ป่วยที่เป็นทั้ง 2 โรคมีค่าเท่ากับ 134 33 76 และ 58% ตามลำดับดังรูปที่ 72 มากไปกว่านั้นจำนวนวันนอนเฉลี่ยต่อการรับการรักษาตัวในโรงพยาบาล 1 ครั้งในแต่ละช่วงอายุมีความใกล้เคียงกันและพบลักษณะเดียวกันในทุกประเภทของผู้ป่วยดังรูปที่ 73

ตารางที่ 56 จำนวนการรับบริการผู้ป่วยในและค่าบริการผู้ป่วยในเฉลี่ยต่อคนต่อปีแยกชายเพศ

| ประเภท | จำนวนวันนอนเฉลี่ย (SD) ต่อครั้ง (วัน) | ค่าบริการผู้ป่วยในเฉลี่ย (SD) ต่อครั้ง (บาท) | จำนวนการรับบริการผู้ป่วยในเฉลี่ยต่อปี (ครั้ง) | ค่าบริการผู้ป่วยในเฉลี่ยต่อคนต่อปี (บาท) |
|--|---------------------------------------|--|---|--|
| เพศชาย | | | | |
| ผู้ที่ไม่เป็นโรคเบาหวาน และความดันโลหิตสูง | 4.4 (6.0) | 19,567 (54,745) | 0.08 | 1,593 |
| ผู้ป่วยโรคเบาหวานอย่างเดียว | 5.0 (6.7) | 21,839 (46,414) | 0.24 | 5,224 |
| ผู้ป่วยโรคความดันโลหิตสูงอย่างเดียว | 4.5 (6.5) | 22,993 (51,136) | 0.23 | 5,313 |
| ผู้ป่วยที่เป็นทั้ง 2 โรค | 4.6 (6.2) | 23,142 (50,182) | 0.38 | 8,760 |
| เพศหญิง | | | | |
| ผู้ที่ไม่เป็นโรคเบาหวาน และความดันโลหิตสูง | 3.9 (4.8) | 17,033 (40,610) | 0.06 | 1,107 |
| ผู้ป่วยโรคเบาหวานอย่างเดียว | 4.2 (5.2) | 17,513 (33,369) | 0.15 | 2,553 |
| ผู้ป่วยโรคความดันโลหิตสูงอย่างเดียว | 4.2 (5.9) | 20,359 (53,221) | 0.16 | 3,230 |
| ผู้ป่วยที่เป็นทั้ง 2 โรค | 4.3 (5.5) | 19,660 (45,908) | 0.29 | 5,774 |



รูปที่ 72 ค่ารักษาพยาบาลผู้ป่วยในเฉลี่ยต่อคนต่อปี (บาท) แยกรายประเภทผู้ป่วยและช่วงอายุ



รูปที่ 73 จำนวนวันนอนเฉลี่ยต่อการรักษาตัวในโรงพยาบาล 1 ครั้งแยกรายประเภทผู้ป่วยและช่วงอายุ

เมื่อพิจารณาถึงปัจจัยเรื่องภาวะแทรกซ้อน (ตารางผนวกที่ 2) พบว่า มีแนวโน้มใกล้เคียงกับค่ารักษาพยาบาลผู้ป่วยนอก โดยการเพิ่มขึ้นของภาวะแทรกซ้อนแต่ละประเภทส่งผลให้ค่ารักษาพยาบาลผู้ป่วยในเฉลี่ยต่อคนต่อปีที่สูงขึ้น อย่างไรก็ตาม จากการวิเคราะห์ถดถอยลิเนียร์ของค่ารักษาพยาบาลผู้ป่วยในรายครั้ง พบว่า ผู้ป่วยที่มีภาวะไตวายเรื้อรังจะมีค่ารักษาพยาบาลผู้ป่วยในต่อครั้งน้อยกว่าผู้ป่วยที่ไม่มีภาวะแทรกซ้อนดังกล่าวดังตารางที่ 57

นอกจากนี้ เมื่อพิจารณาค่ารักษาพยาบาลผู้ป่วยในต่อคนต่อปีตามจำนวนภาวะแทรกซ้อน พบว่า เมื่อผู้ป่วยมีจำนวนภาวะแทรกซ้อนเพิ่มขึ้น จำนวนการรับบริการผู้ป่วยใน จำนวนวันนอนและค่ารักษาพยาบาลผู้ป่วยในมีแนวโน้มเพิ่มสูงขึ้นตามลำดับ โดยผู้ที่เป็นทั้งโรคเบาหวานและความดันโลหิตสูงจะมีการใช้บริการผู้ป่วยในจำนวนวันนอน และค่าใช้จ่ายสูงกว่าผู้ที่ป่วยเพียงโรคเดียวเช่นเดียวกับค่ารักษาพยาบาลผู้ป่วยนอกดังตารางที่ 58

ตารางที่ 57 สัมประสิทธิ์การถดถอยจากตัวแบบการถดถอยลิเนียร์ในการอธิบายค่ารักษาพยาบาลผู้ป่วยในต่อครั้ง (บาท)

| กลุ่มโรค | สัมประสิทธิ์ | 95% Confidence Interval |
|----------------------------|--------------|-------------------------|
| โรคเบาหวาน | 94 | -219 to 407 |
| โรคความดันโลหิตสูง | 2,753 | 2,489 to 3,016 |
| โรคหลอดเลือดหัวใจตีบ | 6,792 | 6,102 to 7,482 |
| โรคเส้นเลือดสมองแตกหรือตีบ | 1,777 | 1,302 to 2,251 |
| โรคไตเสื่อมเรื้อรัง | -2,026 | -2,439 to -1,612 |
| ภาวะแทรกซ้อนทางตา | 702 | -4.4 to 1,408 |
| Constant | 18,452 | 18,284 to 18,621 |

ตารางที่ 58 จำนวนการรับบริการผู้ป่วยในและค่าบริการผู้ป่วยในเฉลี่ยต่อปีแยกรายจำนวนภาวะแทรกซ้อน

| ประเภทของผู้ป่วย | จำนวนการรับบริการผู้ป่วยใน (ครั้ง) | จำนวนวันนอนเฉลี่ย (SD) ต่อครั้ง (วัน) | ค่าบริการผู้ป่วยในเฉลี่ย (SD) ต่อครั้ง (บาท) | จำนวนการรับบริการผู้ป่วยในเฉลี่ยต่อปี (ครั้ง) | ค่าบริการผู้ป่วยในเฉลี่ยต่อคนต่อปี (บาท) |
|---|------------------------------------|---------------------------------------|--|---|--|
| เป็นโรคเบาหวานอย่างเดียว | | | | | |
| ไม่มีภาวะแทรกซ้อน | 20,786 | 4.7 (6.0) | 19,795 (40,941) | 0.17 | 3,338 |
| มีภาวะแทรกซ้อน 1 ประเภท | 4,907 | 4.4 (6.4) | 19,573 (40,059) | 0.26 | 5,054 |
| มีภาวะแทรกซ้อน 2 ประเภท | 522 | 4.3 (5.3) | 19,378 (36,350) | 0.38 | 7,443 |
| มีภาวะแทรกซ้อน 3 ประเภท | 21 | 3.9 (3.1) | 20,429 (32,840) | 0.49 | 9,977 |
| มีภาวะแทรกซ้อน 4 ประเภท* | - | - | - | - | - |
| เป็นโรคความดันโลหิตสูงอย่างเดียว | | | | | |
| ไม่มีภาวะแทรกซ้อน | 159,887 | 4.3 (6.0) | 21,264 (52,981) | 0.15 | 3,216 |
| มีภาวะแทรกซ้อน 1 ประเภท | 53,068 | 4.5 (6.7) | 22,695 (49,985) | 0.55 | 12,481 |
| มีภาวะแทรกซ้อน 2 ประเภท | 5,274 | 4.8 (6.3) | 24,336 (49,530) | 1.04 | 25,325 |
| มีภาวะแทรกซ้อน 3 ประเภท | 311 | 5.1 (6.8) | 26,588 (52,643) | 1.76 | 46,717 |
| มีภาวะแทรกซ้อน 4 ประเภท* | - | - | - | - | - |
| เป็นทั้งโรคเบาหวานและความดันโลหิตสูง | | | | | |
| ไม่มีภาวะแทรกซ้อน | 65,991 | 4.4 (5.7) | 21,150 (49,232) | 0.22 | 4,560 |
| มีภาวะแทรกซ้อน 1 ประเภท | 50,743 | 4.4 (5.9) | 20,710 (45,506) | 0.51 | 10,653 |
| มีภาวะแทรกซ้อน 2 ประเภท | 17,580 | 4.5 (5.9) | 21,753 (44,319) | 1.07 | 23,173 |
| มีภาวะแทรกซ้อน 3 ประเภท | 2,747 | 4.8 (6.9) | 24,201 (69,698) | 1.84 | 44,648 |
| มีภาวะแทรกซ้อน 4 ประเภท | 234 | 5.5 (7.7) | 32,168 (59,886) | 3.34 | 107,534 |

* ไม่มีข้อมูลการใช้บริการผู้ป่วยในหรือมีจำนวนการใช้บริการผู้ป่วยในน้อยกว่า 5 ครั้ง ส่งผลให้ข้อมูลที่คำนวณได้อาจไม่สามารถเป็นตัวแทนของกลุ่มประชากรได้

หัวข้อที่ 1.2. ต้นทุนทางตรงที่ไม่เกี่ยวข้องกับการแพทย์ (Direct Non-Medical Cost)

ต้นทุนทางตรงที่ไม่เกี่ยวข้องกับการแพทย์ของการศึกษานี้ ประกอบด้วย ค่าอาหาร และ ค่าเดินทางมารับบริการอ้างอิงจากรายการต้นทุนมาตรฐานเพื่อการประเมินเทคโนโลยีด้านสุขภาพ โดยสามารถสรุปค่าใช้จ่ายของการรับบริการผู้ป่วยนอกและผู้ป่วยในต่อ 1 ครั้งดังตารางที่ 59

ตารางที่ 59 ต้นทุนทางตรงที่ไม่เกี่ยวข้องกับการแพทย์ของการมารับบริการผู้ป่วยนอกและผู้ป่วยใน 1 ครั้ง

| รายการ | ค่าใช้จ่าย ณ ปี พ.ศ. 2552 (บาท) | ค่าใช้จ่าย ณ ปี พ.ศ. 2566 (บาท) |
|--|------------------------------------|------------------------------------|
| การมารับบริการผู้ป่วยนอก 1 ครั้ง | | |
| ค่าอาหารของผู้ป่วย 1 มื้อ | 39.4 | 50.1 |
| ค่าเดินทางไปกลับบ้าน 1 ครั้ง | 107.4 | 136.6 |
| รวมทั้งสิ้น | 146.8 | 186.7 |
| การมารับบริการผู้ป่วยใน 1 ครั้ง | | |
| ค่าอาหารของผู้ป่วย 2 มื้อ (วันที่มาและกลับจาก รพ.) | 78.7 | 100.1 |
| ค่าเดินทางไปกลับบ้าน 2 ครั้ง | 214.9 | 273.3 |
| รวมทั้งสิ้น | 293.6 | 373.4 |

หัวข้อที่ 2 ต้นทุนทางอ้อม (Indirect Cost)

การคำนวณค่าเสียโอกาสของการศึกษาอ้างอิงจากข้อมูลรายได้ประชาชาติต่อหัว (GNI per capita) ณ ปี พ.ศ. 2565 จากฐานข้อมูล World Bank โดยมีมูลค่าอยู่ที่ 235,219 บาท หรือ คิดเป็น 240,146 บาท ณ ปีมูลค่าฐาน พ.ศ. 2566 โดยเมื่อคำนวณเป็นค่าเสียโอกาสต่อชั่วโมง และต่อวัน ดังสมการที่แสดง

$$\text{ค่าเสียโอกาสต่อชั่วโมง} = \frac{\text{รายได้ประชาชาติต่อหัวต่อปี}}{365 \text{ วัน} \times 8 \text{ ชั่วโมง}} = \frac{240,146}{365 \text{ วัน} \times 8 \text{ ชั่วโมง}} = 82.2 \text{ บาทต่อชั่วโมง}$$

$$\text{ค่าเสียโอกาสต่อวัน} = 82.2 \text{ บาทต่อชั่วโมง} \times \text{ระยะเวลาการทำงาน 8 ชั่วโมงต่อวัน} = 658 \text{ บาทต่อวัน}$$

หัวข้อที่ 2.1. ค่าเสียโอกาสของผู้ป่วย (Patient's Productivity loss)

ค่าเสียโอกาสของผู้ป่วยของการมารับบริการผู้ป่วยนอกและผู้ป่วยใน 1 ครั้ง สามารถสรุปได้ดังตารางที่ 60 โดย ค่าเสียโอกาสของผู้ป่วยในการมารับบริการผู้ป่วยนอก คือ 367.3 บาท และ การมารับบริการผู้ป่วยใน คือ 137.1 บาท (ระยะเวลาที่ใช้ในการเดินทาง) + จำนวนวันนอนโรงพยาบาลโดยคิดอัตราที่ 658 บาทต่อวัน

ตารางที่ 60 ต้นทุนค่าเสียโอกาสของผู้ป่วยของการมารับบริการผู้ป่วยนอกและผู้ป่วยใน 1 ครั้ง

| รายการ | เวลา (ชั่วโมง) | ค่าเสียโอกาส (บาท) |
|--|----------------|--------------------|
| การมารับบริการผู้ป่วยนอก 1 ครั้ง เวลาที่ใช้ทั้งหมดในการมารับบริการที่หน่วยผู้ป่วยนอก (รวมระยะเวลาเดินทางจากบ้านไปกลับโรงพยาบาล) | 4.5 | 367.3 |
| การมารับบริการผู้ป่วยใน 1 ครั้ง เวลาที่ใช้ในการเดินทางกลับบ้านและโรงพยาบาล จำนวนวันนอนโรงพยาบาล (LOS) ขึ้นอยู่กับประเภทของผู้ป่วย | 1.7 - - | 137.1 658 x LOS |

หัวข้อที่ 2.2. ค่าเสียโอกาสของผู้ดูแล (Informal Care Cost)

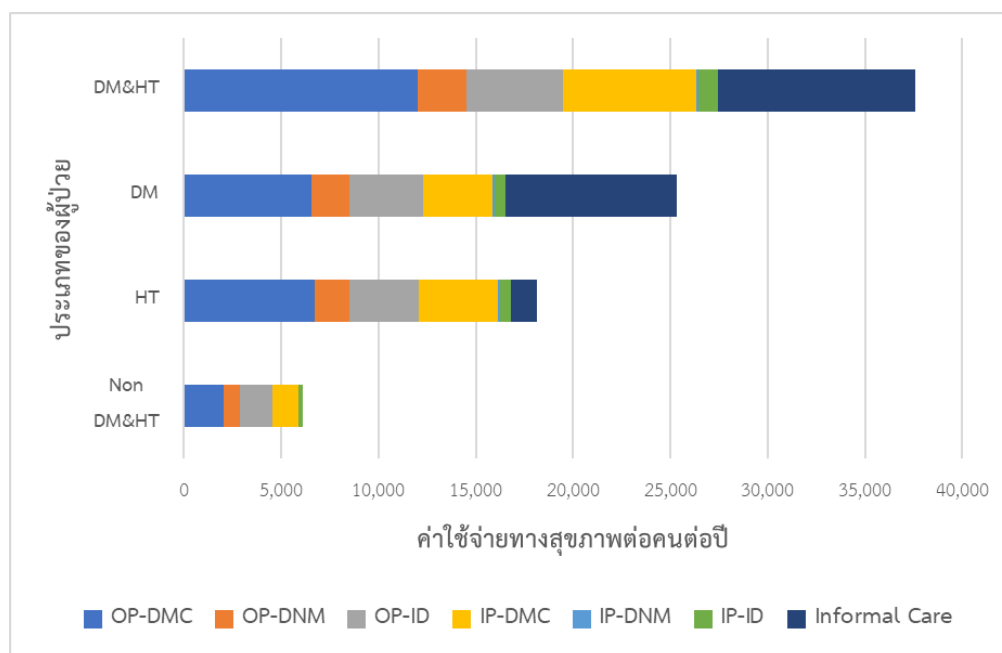
จากการทบทวนวรรณกรรม พบว่า จำนวนชั่วโมงที่ผู้ดูแลต้องใช้เพิ่มเติมในการดูแลผู้ป่วยโรคความดันโลหิตสูง โรคเบาหวานที่ใช้เฉพาะยาทาน และ โรคเบาหวานที่ใช้ยาอินซูลินร่วมด้วยอยู่ที่ 0.31 1.9 และ 4 ชั่วโมงต่อสัปดาห์ตามลำดับ โดยความชุกของผู้ป่วยโรคเบาหวานที่ใช้ยาอินซูลินอยู่ที่ 7.4% ของผู้ป่วยเบาหวานทั้งหมด⁸⁹ ส่งผลให้ค่าเสียโอกาสของผู้ดูแลเป็นดังตารางที่ 61

หัวข้อที่ 3 ต้นทุนค่าใช้จ่ายสุขภาพทั้งหมด

จากข้อมูลต้นทุนทั้งหมดดังที่ได้กล่าวไป เมื่อนำองค์ประกอบของต้นทุนทั้งหมดมารวมกันส่งผลให้ต้นทุนค่าใช้จ่ายทางสุขภาพในแต่ละกลุ่มเพิ่มขึ้นอย่างมีนัยสำคัญ โดย ต้นทุนทางตรงทางการแพทย์คิดเป็นสัดส่วนประมาณ 40-60% ของต้นทุนทั้งหมด สำหรับข้อแตกต่างที่สำคัญระหว่างกลุ่มผู้ที่ไม่เป็นโรคเมื่อเทียบกับผู้ที่เป็นโรคเบาหวานและ/หรือความดันโลหิตสูง คือ ต้นทุนค่าเสียโอกาสของผู้ดูแลโดยเฉพาะโรคเบาหวานที่มีสัดส่วนสูงถึง 35% ของต้นทุนทั้งหมดดังรูปที่ 74

ตารางที่ 61 ต้นทุนค่าเสียโอกาสของผู้ดูแลของโรคเบาหวานและความดันโลหิตสูงต่อปี

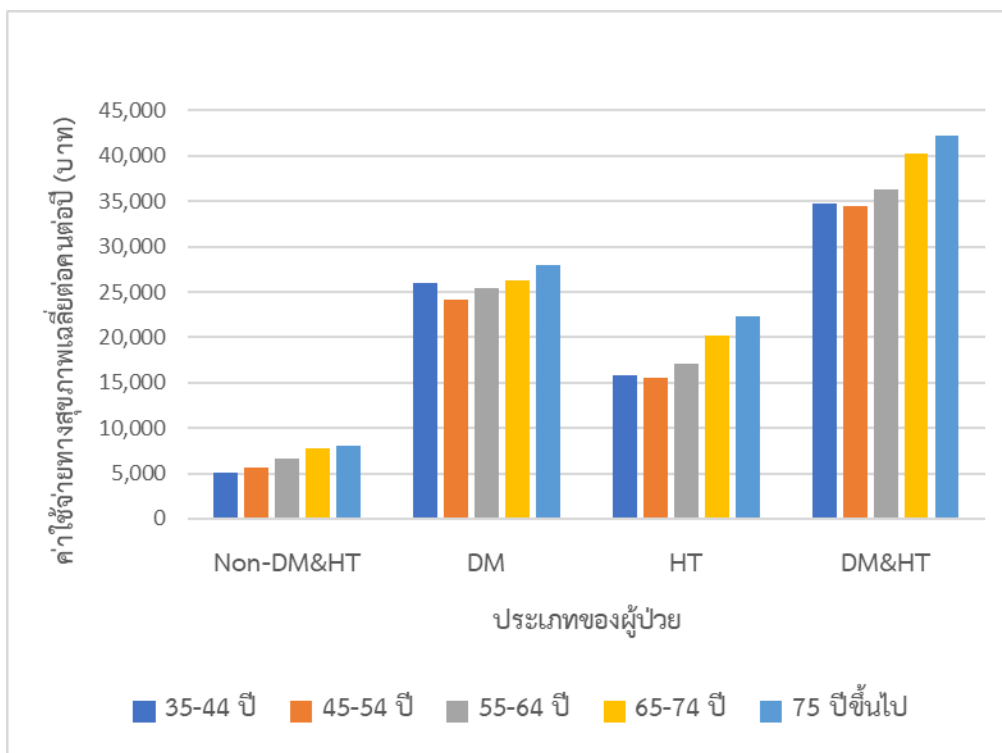
| รายการ | เวลาที่ใช้เพิ่มเติม ต่อสัปดาห์ (ชั่วโมง) | เวลาที่ใช้เพิ่มเติม ต่อปี (ชั่วโมง) | ค่าเสียโอกาส ต่อปี (บาท) |
|--|--|--|-----------------------------|
| โรคความดันโลหิตสูง | 0.31 | 16.1 | 1,326 |
| โรคเบาหวานที่ใช้เฉพาะยาทาน | 1.9 | 98.8 | 8,125 |
| โรคเบาหวานที่ใช้ยาอินซูลิน | 4.0 | 208.0 | 17,106 |
| โรคเบาหวาน: Weight Average ตาม สัดส่วนของผู้ป่วยที่ใช้ยาอินซูลิน (7.4%) | 2.1 | 106.9 | 8,790 |



OP: ผู้ป่วยนอก, IP: ผู้ป่วยใน, DMC: Direct Medical Cost, DNM: Direct Non-medical Cost, ID: Indirect Cost

รูปที่ 74 ต้นทุนใช้จ่ายสุขภาพเฉลี่ยต่อคนต่อปี (บาท) แยกรายองค์ประกอบของต้นทุน

เมื่อทำการวิเคราะห์แยกรายเพศ พบว่า ผู้ป่วยเพศชายยังคงมีค่าใช้จ่ายทางสุขภาพต่อคนต่อปีสูงกว่าเพศหญิงดังตารางที่ 62 ค่าใช้จ่ายทางสุขภาพต่อคนต่อปีมีแนวโน้มเพิ่มสูงขึ้นตามช่วงอายุและมีลักษณะเช่นเดียวกับค่ารักษาพยาบาลผู้ป่วยนอกและผู้ป่วยใน โดยความแตกต่างของค่าใช้จ่ายทางสุขภาพต่อคนต่อปีระหว่างกลุ่มอายุสูงและต่ำสุดของผู้ที่ไม่เป็นโรค ผู้ป่วยโรคเบาหวาน ผู้ป่วยโรคความดันโลหิตสูง และ ผู้ป่วยที่เป็นทั้ง 2 โรคมีค่าเท่ากับ 59 8 41 และ 22% ตามลำดับดังรูปที่ 75 และ จำนวนภาวะแทรกซ้อนที่เพิ่มขึ้นส่งผลให้ค่าใช้จ่ายทางสุขภาพในภาพรวมสูงขึ้นเช่นเดียวกับการวิเคราะห์ของต้นทุนทางตรงทางการแพทย์ดังตารางที่ 63



รูปที่ 75 ต้นทุนค่าใช้จ่ายสุขภาพเฉลี่ยต่อคนต่อปี (บาท) แยกรายประเภทผู้ป่วย

ตารางที่ 62 ค่าใช้จ่ายทางสุขภาพเฉลี่ยต่อคนต่อปีแยกประเภทผู้ป่วยและรายเพศ

| ประเภท | ผู้ป่วยนอก | | | ผู้ป่วยใน | | | Informal Care Cost | รวมทั้งสิ้น (ต่อคนต่อปี) |
|---|---------------------|-------------------------|---------------|---------------------|-------------------------|---------------|--------------------|--------------------------|
| | Direct Medical Cost | Direct Non-medical Cost | Indirect Cost | Direct Medical Cost | Direct Non-medical Cost | Indirect Cost | | |
| ทั้งหมด | | | | | | | | |
| ผู้ที่ไม่เป็นโรคเบาหวานและความดันโลหิตสูง | 2,030 | 848 | 1,668 | 1,333 | 27 | 208 | - | 6,114 |
| ผู้ป่วยโรคเบาหวานอย่างเดียว | 6,563 | 1,927 | 3,790 | 3,605 | 68 | 578 | 8,790 | 25,322 |
| ผู้ป่วยโรคความดันโลหิตสูงอย่างเดียว | 6,696 | 1,811 | 3,562 | 4,091 | 70 | 568 | 1,326 | 18,124 |
| ผู้ป่วยที่เป็นทั้ง 2 โรค | 12,021 | 2,514 | 4,946 | 6,866 | 121 | 992 | 10,116 | 37,575 |
| เพศชาย | | | | | | | | |
| ผู้ที่ไม่เป็นโรคเบาหวานและความดันโลหิตสูง | 2,000 | 749 | 1,473 | 1,593 | 30 | 246 | - | 6,090 |
| ผู้ป่วยโรคเบาหวานอย่างเดียว | 7,262 | 1,820 | 3,580 | 5,224 | 89 | 815 | 8,790 | 27,581 |
| ผู้ป่วยโรคความดันโลหิตสูงอย่างเดียว | 7,821 | 1,749 | 3,442 | 5,313 | 86 | 723 | 1,326 | 20,461 |
| ผู้ป่วยที่เป็นทั้ง 2 โรค | 13,963 | 2,507 | 4,933 | 8,760 | 141 | 1,205 | 10,116 | 41,626 |
| เพศหญิง | | | | | | | | |
| ผู้ที่ไม่เป็นโรคเบาหวานและความดันโลหิตสูง | 2,056 | 935 | 1,839 | 1,107 | 24 | 174 | - | 6,135 |
| ผู้ป่วยโรคเบาหวานอย่างเดียว | 6,108 | 1,996 | 3,927 | 2,553 | 54 | 425 | 8,790 | 23,854 |
| ผู้ป่วยโรคความดันโลหิตสูงอย่างเดียว | 5,903 | 1,854 | 3,647 | 3,230 | 59 | 458 | 1,326 | 16,477 |
| ผู้ป่วยที่เป็นทั้ง 2 โรค | 10,901 | 2,517 | 4,953 | 5,774 | 110 | 869 | 10,116 | 35,239 |

ตารางที่ 63 ค่าใช้จ่ายทางสุขภาพเฉลี่ยต่อคนต่อปีแยกรายจำนวนภาวะแทรกซ้อน

| ประเภท | ผู้ป่วยนอก | | | ผู้ป่วยใน | | | Informal Care Cost | รวมทั้งสิ้น (ต่อคนต่อปี) |
|---|---------------------|-------------------------|---------------|---------------------|-------------------------|---------------|--------------------|--------------------------|
| | Direct Medical Cost | Direct Non-medical Cost | Indirect Cost | Direct Medical Cost | Direct Non-medical Cost | Indirect Cost | | |
| เป็นโรคเบาหวานอย่างเดียว | | | | | | | | |
| ไม่มีภาวะแทรกซ้อน | 6,198 | 1,900 | 3,738 | 3,338 | 63 | 540 | 8,790 | 24,567 |
| มีภาวะแทรกซ้อน 1 ประเภท | 8,615 | 2,083 | 4,097 | 5,054 | 96 | 787 | 8,790 | 29,522 |
| มีภาวะแทรกซ้อน 2 ประเภท | 10,729 | 2,138 | 4,207 | 7,443 | 143 | 1,143 | 8,790 | 34,594 |
| มีภาวะแทรกซ้อน 3 ประเภท | 14,096 | 2,332 | 4,588 | 9,977 | 182 | 1,316 | 8,790 | 41,280 |
| มีภาวะแทรกซ้อน 4 ประเภท* | - | - | - | - | - | - | - | - |
| เป็นโรคความดันโลหิตสูงอย่างเดียว | | | | | | | | |
| ไม่มีภาวะแทรกซ้อน | 5,450 | 1,720 | 3,383 | 3,216 | 56 | 448 | 1,326 | 15,599 |
| มีภาวะแทรกซ้อน 1 ประเภท | 18,615 | 2,691 | 5,294 | 12,481 | 205 | 1,721 | 1,326 | 42,333 |
| มีภาวะแทรกซ้อน 2 ประเภท | 37,910 | 3,935 | 7,742 | 25,325 | 389 | 3,399 | 1,326 | 80,025 |
| มีภาวะแทรกซ้อน 3 ประเภท | 53,211 | 5,513 | 10,846 | 46,717 | 656 | 6,104 | 1,326 | 124,373 |
| มีภาวะแทรกซ้อน 4 ประเภท* | - | - | - | - | - | - | - | - |
| เป็นทั้งโรคเบาหวานและความดันโลหิตสูง | | | | | | | | |
| ไม่มีภาวะแทรกซ้อน | 9,021 | 2,269 | 4,463 | 4,560 | 81 | 658 | 10,116 | 31,166 |
| มีภาวะแทรกซ้อน 1 ประเภท | 16,711 | 2,927 | 5,759 | 10,653 | 192 | 1,556 | 10,116 | 47,914 |
| มีภาวะแทรกซ้อน 2 ประเภท | 34,805 | 4,265 | 8,392 | 23,173 | 398 | 3,312 | 10,116 | 84,459 |
| มีภาวะแทรกซ้อน 3 ประเภท | 62,325 | 5,939 | 11,685 | 44,648 | 689 | 6,086 | 10,116 | 141,488 |
| มีภาวะแทรกซ้อน 4 ประเภท | 82,390 | 6,039 | 11,881 | 107,534 | 1,248 | 12,522 | 10,116 | 231,730 |

* ไม่มีข้อมูลการใช้บริการผู้ป่วยในหรือมีจำนวนการใช้บริการผู้ป่วยในน้อยกว่า 5 ครั้ง ส่งผลให้ข้อมูลที่คำนวณได้อาจไม่สามารถเป็นตัวแทนของกลุ่มประชากรได้

บทที่ 5 อภิปรายและสรุปผลการศึกษา

อภิปรายผลการศึกษา

การศึกษาเรื่องการพัฒนาแบบจำลองทำนายความเสี่ยงและภาวะแทรกซ้อนจากโรคเบาหวานและความดันโลหิตสูงในประเทศไทยเพื่อประเมินผลกระทบทางเศรษฐกิจต่อระบบหลักประกันสุขภาพและระบบสุขภาพ พบว่า แบบจำลองทำนายความเสี่ยงที่สร้างขึ้นหลายแบบจำลองดังตารางที่ 64 มีประสิทธิภาพอยู่ในเกณฑ์ดี โดยเฉพาะแบบจำลองทำนายความเสี่ยงการเกิดโรคเบาหวานและการเกิดโรคความดันโลหิตสูงซึ่งถือว่ามีประสิทธิภาพการทำนายสูงมากเมื่อเทียบกับแบบจำลองทำนายความเสี่ยงมาตรฐานของต่างประเทศที่มีการใช้งานทั่วโลก

ประสิทธิภาพการทำนายการเกิดโรคเบาหวานที่พัฒนาขึ้นด้วยวิธี XGBoost มีค่า AUC อยู่ระหว่าง 0.86-0.88 ซึ่งใกล้เคียงกับแบบจำลองทำนายความเสี่ยงการเกิดโรคเบาหวานของ Diabetes Risk Score ที่มีค่า AUC อยู่ที่ 0.85-0.87⁶¹ และ The Finnish Diabetes Risk Score (FINDRISC) ที่มีค่า AUC อยู่ที่ 0.76-0.79⁹⁰ รวมทั้งแบบจำลองที่พัฒนาขึ้นยังมีประสิทธิภาพสูงกว่าแบบจำลองทำนายความเสี่ยงมาตรฐานของประเทศไทยที่มีค่า AUC อยู่ที่ 0.74-0.78⁴³ นอกจากนี้ ประสิทธิภาพการทำนายการเกิดโรคความดันโลหิตสูงที่พัฒนาขึ้นด้วยวิธี XGBoost นั้นมีค่า AUC อยู่ที่ 0.84-0.87 ซึ่งเทียบเท่ากับแบบจำลองทำนายความเสี่ยงการเกิดโรคความดันโลหิตสูง The Framingham Heart Study ซึ่งมีประสิทธิภาพการทำนายเฉลี่ย (pooled AUC) อยู่ที่ 0.75 (0.68-0.80)⁹¹ นอกเหนือจากประสิทธิภาพการทำนายที่สูงแล้ว แบบจำลองทำนายความเสี่ยงที่สร้างขึ้นนั้นใช้ข้อมูลสุขภาพของฐานข้อมูล 43 แฟ้มซึ่งเป็นตัวแทนส่วนใหญ่ของประชากรไทยที่รับการรักษาในโรงพยาบาลรัฐบาลทำให้เชื่อได้ว่า เมื่อนำแบบจำลองทำนายความเสี่ยงดังกล่าวไปใช้จริงสำหรับการคัดกรองโรคเบาหวานและความดันโลหิตสูงประจำปีในกลุ่มประชากรทั่วไป แบบจำลองที่สร้างขึ้นจะมีประสิทธิภาพดีเพียงพอในการทำนายและให้ข้อมูลความเสี่ยงทางสุขภาพให้กับประชาชนได้

อย่างไรก็ตาม เมื่อพิจารณาผลการสร้างแบบจำลองทำนายความเสี่ยงของการศึกษาจำนวน 24 แบบจำลองด้วยค่า Annual AUC และ Cumulative AUC มีข้อค้นพบดังนี้

ตารางที่ 64 สรุปประสิทธิภาพการทำนายของแบบจำลองทำนายความเสี่ยงของการศึกษา
(วิธีที่ดีที่สุดจะแสดงเป็นตัวหนา)

| ประเภทแบบจำลอง | วิธี Elastic Net | | วิธี XGBoost | |
|--|------------------|----------------|---------------|----------------|
| | Annual AUC | Cumulative AUC | Annual AUC | Cumulative AUC |
| กลุ่มประชากรทั่วไป | | | | |
| การเกิดโรคเบาหวาน | 0.8452 | 0.8565 | 0.8573 | 0.8760 |
| การเกิดโรคความดันโลหิตสูง | 0.8286 | 0.8657 | 0.8372 | 0.8721 |
| กลุ่มผู้ป่วยโรคเบาหวาน | | | | |
| การเกิดภาวะเส้นเลือดหัวใจตีบ หรือ กล้ามเนื้อหัวใจขาดเลือด | 0.6524 | 0.5996 | 0.6713 | 0.6820 |
| การเกิดภาวะเส้นเลือดสมองแตก หรือตีบ | 0.6501 | 0.6248 | 0.6666 | 0.6378 |
| การเกิดภาวะแทรกซ้อนทางตา | 0.6005 | 0.6471 | 0.6322 | 0.677 |
| การเกิดภาวะไตเสื่อมเรื้อรัง | 0.7859 | 0.8038 | 0.8147 | 0.8373 |
| การเสียชีวิต | 0.7886 | 0.7157 | 0.8145 | 0.701 |
| กลุ่มผู้ป่วยโรคความดันโลหิตสูง | | | | |
| การเกิดภาวะเส้นเลือดหัวใจตีบ หรือ กล้ามเนื้อหัวใจขาดเลือด | 0.6557 | 0.6217 | 0.663 | 0.6402 |
| การเกิดภาวะเส้นเลือดสมองแตก หรือตีบ | 0.6418 | 0.6181 | 0.6467 | 0.6253 |
| การเกิดภาวะแทรกซ้อนทางตา | 0.8147 | 0.8367 | 0.8415 | 0.8413 |
| การเกิดภาวะไตเสื่อมเรื้อรัง | 0.7803 | 0.7981 | 0.7973 | 0.8166 |
| การเสียชีวิต | 0.7975 | 0.7534 | 0.8197 | 0.7673 |

ประเด็นที่ 1 ความแตกต่างระหว่าง linear model กับ XGBoost

โมเดลในงานวิจัยนี้อาศัยระเบียบวิธี 2 วิธี คือ Elastic Net และ XGBoost ซึ่ง Elastic Net เป็นตัวแบบเส้นตรงสำหรับการแยกประเภทซึ่งมีการ regularization ด้วยขนาดของพารามิเตอร์กำลังหนึ่งและกำลังสอง การ regularization เป็นการลดความแปรปรวนของค่าทำนายของตัวแบบเส้นตรง ส่วน XGBoost เป็นวิธีในกลุ่มของต้นไม้ตัดสินใจซึ่งจัดเป็นตัวแบบที่ไม่เป็นเส้นตรงโดยกระบวนการทำนายของ XGBoost จะทำนายค่าเป้าหมายด้วยต้นไม้ตัดสินใจเป็นลำดับชั้น ในแต่ละชั้นจะทำการทำนายความผิดพลาดจากชั้นที่ผ่านมาเพื่อให้ค่าทำนายเข้าใกล้เป้าหมายมากขึ้นเป็นลำดับ โมเดล Elastic Net และ XGBoost มีส่วนที่ร่วมกันคือทั้งสองโมเดลเป็นโมเดลเชิงบวก (Additive model) ซึ่งอาจกล่าวได้ว่า Elastic Net เป็นโมเดลเชิงบวก (additive model) ที่มีองค์ประกอบเป็นฟังก์ชันเส้นตรง ส่วน XGBoost เป็นโมเดลเชิงบวกที่มีองค์ประกอบเป็นฟังก์ชันที่ไม่เป็นเส้นตรง และสามารถสรุปข้อดีข้อด้อยของแต่ละโมเดลเป็นดังตารางที่ 65

ตารางที่ 65 ข้อดีและข้อด้อยของวิธี Elastic Net และ XGBoost

| | Elastic Net | XGBoost |
|---------|--|---|
| ข้อดี | <ul style="list-style-type: none"> - ติดตั้งใช้งานง่าย - ดีความง่าย - ป้องกันการ overfit ได้ง่าย | <ul style="list-style-type: none"> - ความถูกต้องสูง - ใช้ทรัพยากรในการคำนวณน้อยกว่า Elastic Net เมื่อใช้งานกับข้อมูลขนาดใหญ่ - สามารถรองรับข้อมูลได้หลากหลาย |
| ข้อด้อย | <ul style="list-style-type: none"> - ใช้ทรัพยากรในการคำนวณสูงเมื่อใช้งานกับข้อมูลขนาดใหญ่ - ความถูกต้องไม่สูงนัก | <ul style="list-style-type: none"> - ติดตั้งใช้งานยากกว่า Elastic Net - ดีความยาก - มีโอกาสเกิด overfitting ได้ง่าย - เมื่อใช้กับข้อมูลขนาดเล็กจะใช้ทรัพยากรการคำนวณมากกว่า Elastic Net |

ประเด็นที่ 2 การขยายเวลาการทำนายการเกิดโรค เนื่องจากแบบจำลองที่สร้างขึ้นใช้ข้อมูลติดตามกลุ่มตัวอย่างเป็นระยะเวลา 6 ปี 2 เดือน ส่งผลให้ประสิทธิภาพการทำนายที่ดี หรือ Cumulative AUC จะอยู่ในช่วงระยะเวลา 5 ปี การขยายระยะเวลาการทำนายเพิ่มขึ้นสามารถทำได้โดยตรงไปตรงมา เช่น จากระยะเวลา 5 เป็น 10 ปีสามารถทำได้โดยการสร้างโมเดลจากข้อมูลที่เพิ่มขึ้นโดยมีการติดตามผู้ป่วยให้ครอบคลุมช่วงเวลาที่สนใจ

อย่างไรก็ตาม ด้วยโมเดลปัจจุบันที่ทำนายความเสี่ยงแบบปีต่อปี (Annual AUC) แบบจำลองที่สร้างขึ้นยังสามารถขยายการทำนายความเสี่ยงการเกิดโรคให้ไกลกว่า 5 ปีได้โดยตรงโดยไม่ต้องเพิ่มข้อมูลโดยจะทำนายความเสี่ยงการเกิดโรคแบบปีต่อปีไปจนถึงช่วงเวลาที่สนใจแต่ประสิทธิภาพการทำนายความเสี่ยงที่ระยะไกลขึ้นอาจจะขึ้นอยู่กับหลายปัจจัย เช่น การแจกแจงของประชากรในช่วงที่ไกลกว่า 5 ปี (เช่น 10 ปี) ว่ามีความแตกต่างเล็กน้อยเพียงใด หากการแจกแจงของประชากรในสองช่วงเวลาใกล้เคียงกัน โมเดลปัจจุบันอาจคงประสิทธิภาพที่มีอยู่ไว้ได้

นอกจากนี้ ปัจจัยเรื่องอายุเป็นอีกหนึ่งปัจจัยสำคัญที่ส่งผลต่อการขยายระยะเวลาการทำนาย เนื่องจาก ณ ปัจจุบัน ตัวแปรอายุหนึ่งในตัวแปรอิสระที่สำคัญ หากช่วงระยะเวลาที่ขยายการทำนายเพิ่มเติมยาวเกินไป ขนาดความสัมพันธ์ของการเกิดโรคกับอายุอาจเปลี่ยนแปลงได้ เช่น ผู้หญิงที่หมดประจำเดือนจะมีความเสี่ยงต่อการเกิดโรคกลุ่มเมทาบอลิกสูงขึ้น²² และส่งผลให้ความเสี่ยงในการเกิดโรคไม่ติดต่อเรื้อรังเพิ่มขึ้นด้วย การเปลี่ยนแปลงดังกล่าวอาจส่งผลให้ระดับความเสี่ยงที่คำนวณได้ต่ำกว่าความเป็นจริง เนื่องจากแบบจำลองไม่ได้คำนึงถึงความเสี่ยงของการเกิดโรคที่เปลี่ยนแปลงไป

ประเด็นที่ 3 ประสิทธิภาพในการทำนายของแบบจำลองทำนายความเสี่ยงการเกิดโรคเบาหวานและความดันโลหิตสูงในประชากรทั่วไปมีประสิทธิภาพสูงกว่าแบบจำลองทำนายความเสี่ยงการเกิดภาวะแทรกซ้อนในกลุ่มผู้ป่วยโรคเบาหวานและโรคความดันโลหิตสูงบางประเภท ปรากฏการณ์ดังกล่าวสามารถอธิบายได้จากหลายสาเหตุได้แก่

- 1) การที่ภาวะแทรกซ้อนบางประเภทมีอัตราการเกิดที่ต่ำมากทำให้ไม่สามารถตรวจจับได้ด้วยโมเดล ซึ่งสามารถแก้ไขปัญหาดังกล่าวได้จากการอาศัยชุดข้อมูลที่ใหญ่ขึ้นในการสร้างโมเดลเพื่อเพิ่มประสิทธิภาพของ

การทำนาย เมื่อวิเคราะห์กับลักษณะฐานข้อมูล 43 แฟ้มที่ใช้ในการศึกษานี้ พบว่า ยังมีปริมาณข้อมูลเพียงพอที่สามารถขยายชุดข้อมูลและพัฒนาให้แบบจำลองมีศักยภาพเพิ่มเติมได้

2) ปัจจัยที่ใช้ในการทำนายการเกิดภาวะแทรกซ้อนอาจมีความซับซ้อนมากกว่าปัจจัยที่ใช้ทำนายความเสี่ยงโรคเบาหวานและโรคความดันโลหิตสูง โดยสามารถอธิบายได้จาก 3 ประเด็นหลัก คือ

a) พยาธิสรีรวิทยาของภาวะแทรกซ้อนมีปัจจัยที่เกี่ยวข้องมากกว่าโรคเบาหวานและความดันโลหิตสูง ถึงแม้ว่ารายการตัวแปรที่นำมาใช้ในการสร้างแบบจำลองทำนายการเกิดภาวะแทรกซ้อนจะมีความครอบคลุมมากกว่ารายการที่ใช้ในการทำนายการเกิดโรคเบาหวานและความดันโลหิตสูงอยู่แล้ว ปัจจัยที่ถูกเลือกเข้ามาในแบบจำลองอาจยังไม่ครอบคลุมและไม่ได้รวมปัจจัยที่มีความจำเพาะ (idiosyncratic risk) เช่น ปัจจัยทางด้านพันธุกรรม หรือ การตรวจทางห้องปฏิบัติการและรังสีวิทยาที่ซับซ้อนและมีความจำเพาะในการทำนายโรคแทรกซ้อนได้ดี ยกตัวอย่างเช่น ข้อมูลการ Ultrasound เส้นเลือดแดง Carotid ที่คอ หรือ การเอกเรย์คอมพิวเตอร์เพื่อประเมินปริมาณ Calcium ในเส้นเลือดหัวใจ ซึ่งจำเพาะต่อการประเมินความเสี่ยงการเกิดโรคเส้นเลือดสมองตีบและเส้นเลือดหัวใจตีบ ตามลำดับ อย่างไรก็ตาม วัตถุประสงค์ของการพัฒนาแบบจำลองทำนายความเสี่ยง คือ การใช้ข้อมูลที่สามารถพบได้ตามขั้นตอนการรักษามาตรฐาน เพื่อค้นหาผู้ป่วยที่มีความเสี่ยงสูงให้เข้าสู่การตรวจรักษาเพิ่มเติมที่จำเพาะ การคัดเลือกตัวแปรในการพัฒนาแบบจำลองคงต้องชั่งน้ำหนักระหว่างค่าใช้จ่ายในการได้มาซึ่งข้อมูลและความแม่นยำส่วนเพิ่มที่ได้จากแบบจำลอง

b) ปัจจัยกวนอันเนื่องมาจากการได้รับการรักษาด้วยยา หากเปรียบเทียบกลุ่มประชากรที่ใช้ในการสร้างแบบจำลองจะพบได้ว่า กลุ่มประชากรที่ใช้ในการสร้างแบบจำลองการเกิดโรคเบาหวานและความดันโลหิตสูงนั้นเป็นกลุ่มประชากรทั่วไปซึ่งยังไม่ได้ได้รับการรักษาด้วยยาในการควบคุมระดับความดันโลหิตและระดับน้ำตาล ส่งผลให้ เมื่อรายการตัวแปร เช่น ค่าความดันโลหิตและค่าระดับน้ำตาลในเลือดมาใช้ในการทำนาย จะพบได้ว่า แบบจำลองมีความแม่นยำเพียงพอประกอบกับเมื่อรวมปัจจัยทางด้านพฤติกรรมครอบครัวและสถานะสุขภาพอื่นๆ ที่เกี่ยวข้องการโรคไม่ติดต่อเรื้อรังเพิ่มเติม เช่น ประวัติการสูบบุหรี่ ประวัติโรคความดันโลหิตสูงหรือเบาหวานในครอบครัว และดัชนีมวลกาย เป็นต้น ส่งผลให้แบบจำลองที่ได้รับมีประสิทธิภาพในการทำนายที่ดี ในทางกลับกัน กลุ่มประชากรที่ใช้ในการสร้างแบบจำลองการเกิดภาวะแทรกซ้อนจะมีปัจจัยเรื่องการรักษาด้วยยาถือว่าเป็นหนึ่งในปัจจัยกวนที่สำคัญในการทำนาย ถึงแม้ว่าแบบจำลองที่พัฒนาขึ้นจะมีการนำรายการยาที่สำคัญเข้ามาในแบบจำลองแล้ว และค่าผลทางห้องปฏิบัติการอาจพอเป็นตัวแทนบอกสถานะสุขภาพของผู้ป่วยได้ เช่น หากผู้ป่วยควบคุมระดับน้ำตาลได้ดีอาจอนุมานได้ว่าผู้ป่วยรายนี้น่าจะมีความเสี่ยงในการเกิดภาวะแทรกซ้อนต่ำกว่าผู้ป่วยที่ควบคุมระดับน้ำตาลไม่ได้ อย่างไรก็ตาม การศึกษานี้ ยังไม่ได้ประเมินถึงปริมาณของยาที่ผู้ป่วยใช้ในการรักษาและพฤติกรรมการใช้ยาซึ่งคาดว่าจะเป็นอย่างอื่นในปัจจัยสำคัญที่ส่งผลต่อการเกิดภาวะแทรกซ้อนได้นอกเหนือจากปัจจัยที่ใช้ในการสร้างแบบจำลอง

c) เนื่องจากโรคเบาหวานและความดันโลหิตสูงเป็นโรคเรื้อรัง การติดตามผลการรักษาในลักษณะของ Time Series และประเมิน Variation ของผลการตรวจทางห้องปฏิบัติการ เช่น ระดับความแปรปรวนของระดับ

น้ำตาลในเลือดหรือความดันโลหิตของผู้ป่วยอาจเป็นข้อมูลที่ช่วยให้แบบจำลองทำนายความเสี่ยงมีความแม่นยำเพิ่มขึ้นซึ่งสมมติฐานดังกล่าวอ้างอิงมาจากการศึกษาติดตามผู้ป่วยโรคความดันโลหิตสูงที่พบว่า เมื่อควบคุมปัจจัยกวนเรื่องระดับความดันโลหิตแล้ว ผู้ป่วยที่มีระดับความแปรปรวนของค่าความดันโลหิตสูงระหว่างการตรวจรักษาผู้ป่วยนอกแต่ละครั้งจะมีความเสี่ยงต่อการเกิดโรคหลอดเลือดหัวใจตีบ เส้นเลือดสมองแตกหรือตีบและโรคหัวใจวายได้มากกว่าผู้ป่วยที่มีระดับความแปรปรวนของค่าความดันโลหิตต่ำ⁹³ นอกจากนี้ ปัจจัยเรื่องความแปรปรวนของค่าความดันโลหิตยังเกี่ยวข้องกับการเกิดโรคเส้นเลือดสมองตีบซ้ำและผลการรักษาที่ไม่ดีในผู้ป่วยโรคเส้นเลือดสมองตีบอีกด้วย^{94,95} การพัฒนาแบบจำลองทำนายความเสี่ยงการเกิดภาวะแทรกซ้อนในอนาคตจึงควรนำปัจจัยเรื่องความแปรปรวนของผลการตรวจทางห้องปฏิบัติการเข้าไปวิเคราะห์เพิ่มเติมด้วย เนื่องจากเป็นข้อมูลที่มีการเก็บรวบรวมอยู่แล้วและมีความเป็นไปได้มากกว่าการเลือกการตรวจที่ราคาสูงดังที่กล่าวไว้ในเบื้องต้น

นอกเหนือจากจุดเด่นของการศึกษานี้ที่ได้กล่าวไป 2 ข้อหลัก คือ 1. การวิเคราะห์ด้วยฐานข้อมูลขนาดใหญ่ และ 2. ฐานข้อมูลที่เลือกใช้สามารถเป็นตัวแทนประชากรส่วนใหญ่ของประเทศได้ดี จุดด้อยที่สำคัญของการศึกษานี้มีอยู่ 2 ประเด็น คือ

1. เนื่องจากการศึกษานี้เป็นการใช้ข้อมูลทุติยภูมิที่มีการเก็บรวบรวมไว้อยู่แล้ว ข้อมูลพฤติกรรมทางสุขภาพที่สำคัญหลายรายการที่มีหลักฐานเชิงประจักษ์ในวรรณกรรมทั้งในและต่างประเทศ เช่น การออกกำลังกาย และการรับประทานผักและผลไม้ ซึ่งเป็นหนึ่งในปัจจัยที่เกี่ยวข้องกับการเกิดโรคไม่ติดต่อเรื้อรังไม่สามารถนำมาใช้ในการพัฒนาแบบจำลองได้
2. คุณภาพและความสมบูรณ์ของข้อมูล การศึกษานี้ประสบปัญหาเรื่องความครบถ้วนข้อมูลซึ่งอาจเกิดจากที่ผู้ป่วยไม่ไปรับบริการที่สถานพยาบาลอย่างต่อเนื่อง หรือย้ายไปรับการรักษาที่สถานพยาบาลอื่นที่ไม่ได้ส่งข้อมูลให้กับฐานข้อมูล 43 แพ้ ผู้วิจัยจึงต้องตั้งสมมติฐานในการเติมข้อมูลดังที่แสดงไว้ในระเบียบวิธีวิจัย ซึ่งข้อมูลที่เติมเข้าไบนั้นอาจจะไม่ตรงกับความเป็นจริงส่งผลให้เกิดปัญหาเรื่อง Misclassification of Exposure และ Outcome ได้

นอกเหนือจากการนำแบบจำลองทำนายความเสี่ยงการเกิดโรคเบาหวานและความดันโลหิตสูงไปใช้ในกลุ่มประชากรทั่วไป แบบจำลองทำนายความเสี่ยงยังสามารถนำมาประยุกต์ใช้ได้หลายแง่มุม เช่น การวิเคราะห์ความสัมพันธ์ระหว่างปัจจัยที่ใช้ในการทำนายและโอกาสเกิดโรคเบาหวานและความดันโลหิตสูงด้วยวิธีการเรียนรู้ของเครื่องสามารถช่วยระบุกลุ่มเสี่ยงที่สำคัญซึ่งผู้กำหนดนโยบายและบุคลากรทางการแพทย์ควรมุ่งเน้นในการคัดกรองโรคโดยประเมินจากค่า SHAP กล่าวคือ ผู้ป่วยโรคความดันโลหิตสูงและผู้ที่มีน้ำหนักตัวเกินเกณฑ์สะท้อนจากค่าดัชนีมวลกายหรือขนาดเส้นรอบเอวจะเป็นกลุ่มเสี่ยงหลักต่อการเกิดโรคเบาหวานดังรูปที่ 28 และ ผู้ป่วยโรคเบาหวาน ผู้ที่มีน้ำหนักตัวเกินเกณฑ์สะท้อนจากค่าดัชนีมวลกายหรือขนาดเส้นรอบเอว และ ผู้ที่มีค่าไขมันไตรกลีเซอไรด์สูงจะเป็นกลุ่มเสี่ยงหลักต่อการเกิดโรคความดันโลหิตสูงดังรูปที่ 47 ซึ่งการประเมินปัจจัยเสี่ยงด้วยค่า SHAP นั้นสอดคล้องกับการศึกษาระยะยาวในต่างประเทศที่พบว่า เมื่อควบคุมปัจจัยกวนที่เกี่ยวข้องกับการเกิดโรคเบาหวาน ประกอบด้วย เพศ ดัชนีมวลกาย เศรษฐฐานะทางสังคม พฤติกรรมสุขภาพ และผลการตรวจทางห้องปฏิบัติการต่าง ๆ เช่น ระดับน้ำตาลในเลือด และ ค่าไขมันในเลือด เป็นต้น ผู้ที่มีโรคความดันโลหิตสูงเดิมและผู้

ที่อยู่ในระยะก่อนความดันโลหิตสูง (Prehypertension) จะมีโอกาสเสี่ยงในการเกิดโรคเบาหวานได้มากกว่าผู้ที่มีความดันโลหิตปกติอยู่ที่ 1.51 และ 1.27 เท่าตามลำดับ⁹⁶ มากไปกว่านั้น การศึกษาระยะยาวถึงปัจจัยเสี่ยงการเกิดโรคความดันโลหิตสูงพบในลักษณะเดียวกัน กล่าวคือ เมื่อควบคุมปัจจัยกวนต่าง ๆ แล้ว ผู้ที่เป็นโรคเบาหวานมีโอกาสเสี่ยงในการเกิดโรคความดันโลหิตสูงคิดเป็น 3.14 เท่าเมื่อเทียบกับผู้ที่ไม่เป็นโรคเบาหวาน⁹⁷ ข้อค้นพบดังกล่าวสามารถช่วยให้ผู้กำหนดนโยบายและบุคลากรทางการแพทย์เฝ้าระวัง ติดตามและให้ความสำคัญกับการคัดกรองโรคเบาหวานและความดันโลหิตสูงเพิ่มเติมในผู้ป่วยโรคความดันโลหิตสูงอย่างเดียว และ ผู้ป่วยโรคเบาหวานอย่างเดียว ตามลำดับ

มากไปกว่านั้น แบบจำลองดังกล่าวอาจมีประสิทธิภาพในการทำนายการเกิดโรคเบาหวานและโรคความดันโลหิตสูงในกลุ่มประชากรอื่น ๆ และสามารถต่อยอดเป็นเครื่องมือช่วยในการตัดสินใจในการตรวจสุขภาพประจำปี เช่น ผู้รับบริการในโรงพยาบาลเอกชน โรงพยาบาลขนาดใหญ่ในเขตเมือง หรือ โรงเรียนแพทย์ เป็นต้น แต่ควรมีการประเมิน External Validation เพิ่มเติมเนื่องจากฐานข้อมูลที่นำมาใช้อาจมีประชาชนกลุ่มดังกล่าว น้อยหรือไม่มีเลย นอกจากนี้ ข้อจำกัดหลักที่อาจทำให้ไม่สามารถ generalize แบบจำลองไปใช้ในกลุ่มประชากรที่รับบริการในโรงพยาบาลเอกชน โรงพยาบาลขนาดใหญ่ หรือ โรงเรียนแพทย์ได้ คือ ความแตกต่างของปัจจัยทางด้านเศรษฐกิจทางสังคมซึ่งเป็นตัวแทนสะท้อนถึงพฤติกรรมสุขภาพและกิจวัตรประจำวันที่มีความสัมพันธ์โดยตรงกับการเกิดโรคไม่ติดต่อเรื้อรัง โดยการศึกษาในต่างประเทศ พบว่า กลุ่มเศรษฐกิจทางสังคมต่ำจะมีโอกาสดื่มเครื่องดื่มแอลกอฮอล์ สูบบุหรี่ และ รับประทานผักและผลไม้ไม่เพียงพอ และ กลุ่มเศรษฐกิจทางสังคมสูงจะมีพฤติกรรมเนือยนิ่ง (Sedentary Lifestyle) และรับประทานอาหารแปรรูปที่มีเกลือและไขมันสูง⁹⁸ ซึ่งปัจจัยทั้งหมดเหล่านี้ล้วนเกี่ยวข้องกับโรคไม่ติดต่อเรื้อรังโดยเฉพาะโรคเบาหวานและความดันโลหิตสูง เพื่อลดข้อจำกัดดังกล่าว การปรับปรุงแบบจำลองทำนายความเสี่ยงผ่านการเพิ่มกลุ่มประชากรที่มีความหลากหลายและพัฒนาประสิทธิภาพการทำนายอย่างต่อเนื่องโดยให้แบบจำลองเรียนรู้จากผลการทำนายในอดีตสามารถช่วยให้แบบจำลองมีความแม่นยำมากขึ้นและเหมาะสมต่อกลุ่มเป้าหมาย

นอกเหนือจากการวิเคราะห์ความเสี่ยงรายบุคคล แบบจำลองทำนายความเสี่ยงยังสามารถนำไปประยุกต์ใช้ในการวางแผนการจัดสรรทรัพยากรและการเตรียมความพร้อมทางด้านงบประมาณเพื่อรองรับภาระโรคไม่ติดต่อเรื้อรังที่สูงขึ้น โดยแบบจำลองทำนายความเสี่ยงที่พัฒนาขึ้นทั้งแบบจำลองทำนายความเสี่ยงการเกิดโรคเบาหวานและความดันโลหิตสูงสำหรับประชากรทั่วไป และ แบบจำลองทำนายความเสี่ยงการเกิดภาวะแทรกซ้อนในผู้ที่เป็นโรคแล้วสามารถประเมินฉันทันทีของภาระโรคไม่ติดต่อเรื้อรังในอนาคตได้รวมทั้งสามารถทราบถึงภาระงบประมาณที่จะเพิ่มขึ้นเพื่อรองรับต่อการรักษาโรค นอกจากนี้ ผู้กำหนดนโยบายสามารถประยุกต์ใช้แบบจำลองทำนายความเสี่ยงในการประเมินความคุ้มค่าการลงทุนทางด้านสร้างเสริมสุขภาพและการป้องกันโรคผ่านทางเปรียบเทียบงบประมาณที่ต้องใช้ในการดำเนินโครงการกับค่ารักษาพยาบาลที่ลดลงอันเนื่องมาจากความเสี่ยงทางสุขภาพของประชากรที่เปลี่ยนแปลงไปดังที่มีการศึกษาในต่างประเทศ⁹⁹

สรุปผลการศึกษา

แบบจำลองทำนายความเสี่ยงที่ถูกพัฒนาขึ้นโดยเฉพาะแบบจำลองทำนายการเกิดโรคเบาหวานและความดันโลหิตสูงมีประสิทธิภาพดีเพียงพอที่จะนำไปใช้ในสถานการณ์จริงโดยเฉพาะการคัดกรองโรคเบาหวานและความดันโลหิตสูงในกลุ่มประชากรทั่วไป เนื่องจาก กลุ่มตัวอย่างที่ใช้ในการสร้างแบบจำลองนั้นมีความใกล้เคียงกับผู้รับบริการ นอกจากนี้ แบบจำลองทำนายการเกิดโรคเบาหวานและความดันโลหิตสูงน่าจะมีประสิทธิภาพเพียงพอในการประเมินความเสี่ยงของกลุ่มวัยแรงงานที่เข้ารับการตรวจสุขภาพประจำปีตามโรงพยาบาลต่าง ๆ ในเขตเมืองและโรงพยาบาลเอกชน อย่างไรก็ตาม ควรมีการทำ External Validation เพิ่มเติมเพื่อประเมินศักยภาพในการทำนายก่อนนำไปใช้จริง

สำหรับแบบจำลองทำนายความเสี่ยงการเกิดภาวะแทรกซ้อนบางรายการนั้นอาจมีความจำเป็นต้องพัฒนาประสิทธิภาพของการทำนายเพิ่มเติมดังที่ได้ยกตัวอย่างไว้ในเบื้องต้น เช่น การขยายปริมาณกลุ่มตัวอย่างและการเพิ่มรายการตัวแปรที่มีความซับซ้อนเพิ่มขึ้นก่อนนำไปใช้ในสถานการณ์จริง เพื่อให้แบบจำลองทำนายความเสี่ยงการเกิดภาวะแทรกซ้อนสามารถทำงานและให้ข้อมูลที่มีคุณภาพกับบุคลากรทางการแพทย์และผู้ป่วยเพื่อใช้ในการประกอบการตัดสินใจการรักษาโรคเบาหวานและโรคความดันโลหิตสูงอย่างมีคุณภาพ

บรรณานุกรม

1. Global report on diabetes [Internet]. [cited 2023 Nov 3]. Available from: <https://www.who.int/publications-detail-redirect/9789241565257>
2. International Diabetes Federation. IDF Diabetes Atlas [Internet]. 8th ed. Available from: https://diabetesatlas.org/upload/resources/previous/files/8/IDF_DA_8e-EN-final.pdf
3. Guariguata L, Whiting DR, Hambleton I, Beagley J, Linnenkamp U, Shaw JE. Global estimates of diabetes prevalence for 2013 and projections for 2035. *Diabetes Res Clin Pract*. 2014 Feb;103(2):137–49.
4. Mills KT, Bundy JD, Kelly TN, Reed JE, Kearney PM, Reynolds K, et al. Global Disparities of Hypertension Prevalence and Control: A Systematic Analysis of Population-based Studies from 90 Countries. *Circulation*. 2016 Aug 9;134(6):441–50.
5. King H, Aubert RE, Herman WH. Global burden of diabetes, 1995-2025: prevalence, numerical estimates, and projections. *Diabetes Care*. 1998 Sep;21(9):1414–31.
6. Wild S, Roglic G, Green A, Sicree R, King H. Global prevalence of diabetes: estimates for the year 2000 and projections for 2030. *Diabetes Care*. 2004 May;27(5):1047–53.
7. Noncommunicable diseases country profiles 2018 [Internet]. [cited 2023 Nov 3]. Available from: <https://www.who.int/publications-detail-redirect/9789241514620>
8. A global brief on hypertension : silent killer, global public health crisis: World Health Day 2013 [Internet]. [cited 2023 Nov 3]. Available from: <https://www.who.int/publications-detail-redirect/a-global-brief-on-hypertension-silent-killer-global-public-health-crisis-world-health-day-2013>
9. HYPERTENSION MECHANISMS CAUSING STROKE - Johansson - 1999 - Clinical and Experimental Pharmacology and Physiology - Wiley Online Library [Internet]. [cited 2023 Nov 3]. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1440-1681.1999.03081.x>
10. Beckman JA, Paneni F, Cosentino F, Creager MA. Diabetes and vascular disease: pathophysiology, clinical consequences, and medical therapy: part II. *Eur Heart J*. 2013 Aug;34(31):2444–52.

11. Emerging Risk Factors Collaboration, Sarwar N, Gao P, Seshasai SRK, Gobin R, Kaptoge S, et al. Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: a collaborative meta-analysis of 102 prospective studies. *Lancet Lond Engl*. 2010 Jun 26;375(9733):2215–22.
12. Flaxman SR, Bourne RRA, Resnikoff S, Ackland P, Braithwaite T, Cicinelli MV, et al. Global causes of blindness and distance vision impairment 1990–2020: a systematic review and meta-analysis. *Lancet Glob Health*. 2017 Dec 1;5(12):e1221–34.
13. Jha V, Garcia-Garcia G, Iseki K, Li Z, Naicker S, Plattner B, et al. Chronic kidney disease: global dimension and perspectives. *Lancet Lond Engl*. 2013 Jul 20;382(9888):260–72.
14. Shen Y, Cai R, Sun J, Dong X, Huang R, Tian S, et al. Diabetes mellitus as a risk factor for incident chronic kidney disease and end-stage renal disease in women compared with men: a systematic review and meta-analysis. *Endocrine*. 2017 Jan;55(1):66–76.
15. Zhou D, Xi B, Zhao M, Wang L, Veeranki SP. Uncontrolled hypertension increases risk of all-cause and cardiovascular disease mortality in US adults: the NHANES III Linked Mortality Study. *Sci Rep*. 2018 Jun 20;8(1):9418.
16. Hypertensive retinopathy: incidence, risk factors, and comorbidities - PubMed [Internet]. [cited 2023 Nov 3]. Available from: <https://pubmed.ncbi.nlm.nih.gov/22468968/>
17. Chronic hypertension and pregnancy outcomes: systematic review and meta-analysis | The BMJ [Internet]. [cited 2023 Nov 3]. Available from: <https://www.bmj.com/content/348/bmj.g2301>
18. Bloom DE, Cafiero ET, Jané-Llopis E, Abrahams-Gessel S, Bloom LR, Fathima S, et al. The Global Economic Burden of Noncommunicable Diseases. Geneva: World Economic Forum; 2011.
19. Gaziano TA, Bitton A, Anand S, Weinstein MC, International Society of Hypertension. The global cost of nonoptimal blood pressure. *J Hypertens*. 2009 Jul;27(7):1472–7.
20. Smith-Spangler CM, Bhattacharya J, Goldhaber-Fiebert JD. Diabetes, Its Treatment, and Catastrophic Medical Spending in 35 Developing Countries. *Diabetes Care*. 2012 Feb;35(2):319–26.

21. Si Y, Zhou Z, Su M, Ma M, Xu Y, Heitner J. Catastrophic healthcare expenditure and its inequality for households with hypertension: evidence from the rural areas of Shaanxi Province in China. *Int J Equity Health*. 2017 Dec;16(1):1–12.
22. เอกพลากร ว, ปรปักษ์ขาม ๒, ฐานีพานิชสกุล ส, พรรคเจริญ ห, เสถียรนพแก้ว ว, ไทยกล้า ก. การสำรวจสุขภาพประชาชนไทยโดยการตรวจร่างกายครั้งที่ 4 พ.ศ. 2551-2552 [Internet]. สำนักงานสำรวจสุขภาพประชาชนไทย; 2553 [cited 2023 Nov 3]. Available from: <https://kb.hsri.or.th/dspace/handle/11228/2976>
23. สถาบันวิจัยระบบสาธารณสุข. การสำรวจสภาวะสุขภาพอนามัยของประชาชนไทยโดยการตรวจร่างกายครั้งที่ 5 พ.ศ. 2557. สำนักนโยบายและยุทธศาสตร์ สำนักงานปลัดกระทรวงสาธารณสุข; 2559.
24. Population Division. *World Population Prospects, The 2017 Revision Key Findings and Advance Tables*. Department of Economic and Social Affairs, United Nations; 2017.
25. Collins GS, Mallett S, Omar O, Yu LM. Developing risk prediction models for type 2 diabetes: a systematic review of methodology and reporting. *BMC Med*. 2011 Sep 8;9(1):103.
26. Echouffo-Tcheugui JB, Batty GD, Kivimäki M, Kengne AP. Risk Models to Predict Hypertension: A Systematic Review. *PLoS ONE*. 2013 Jul 5;8(7):e67370.
27. King P, Peacock I, Donnelly R. The UK prospective diabetes study (UKPDS): clinical and therapeutic implications for type 2 diabetes. *Br J Clin Pharmacol*. 1999 Nov;48(5):643–8.
28. Implications of the United kingdom prospective diabetes study - PubMed [Internet]. [cited 2023 Nov 3]. Available from: <https://pubmed.ncbi.nlm.nih.gov/12502617/>
29. Quan J, Pang D, Li TK, Choi CH, Siu SC, Tang SY, et al. Risk Prediction Scores for Mortality, Cerebrovascular, and Heart Disease Among Chinese People With Type 2 Diabetes. *J Clin Endocrinol Metab*. 2019 Dec 1;104(12):5823–30.
30. van der Heijden A a. WA, Feenstra TL, Hoogenveen RT, Niessen LW, de Bruijne MC, Dekker JM, et al. Policy evaluation in diabetes prevention and treatment using a population-based macro simulation model: the MICADO model. *Diabet Med J Br Diabet Assoc*. 2015 Dec;32(12):1580–7.

31. Willis M, Johansen P, Nilsson A, Asseburg C. Validation of the Economic and Health Outcomes Model of Type 2 Diabetes Mellitus (ECHO-T2DM). *PharmacoEconomics*. 2017 Mar;35(3):375–96.
32. Lee W, Lee J, Lee H, Jun CH, Park IS, Kang SH. Prediction of Hypertension Complications Risk Using Classification Techniques. *Ind Eng Manag Syst*. 2014;13(4):449–53.
33. Mt hood diabetes challenge network [Internet]. [cited 2023 Nov 3]. Available from: <https://www.mthooddiabeteschallenge.com/>
34. Australian cardiovascular disease risk calculator | AusCVDRisk [Internet]. [cited 2023 Nov 3]. Available from: <https://www.cvdcheck.org.au/calculator>
35. CVD Risk Assessment [Internet]. [cited 2023 Nov 3]. Available from: <https://www.nzssd.org.nz/cvd/>
36. AlKaabi LA, Ahmed LS, Al Attiyah MF, Abdel-Rahman ME. Predicting hypertension using machine learning: Findings from Qatar Biobank Study. *PLoS ONE*. 2020 Oct 16;15(10):e0240370.
37. McCammon JM, Bandhakavi S, Salek D, Liu Z, Ni X, Benner N, et al. Use of machine learning to predict hypertension-related complication outcomes of varying severity [Internet]. *medRxiv*; 2020 [cited 2023 Nov 3]. p. 2020.10.30.20169615. Available from: <https://www.medrxiv.org/content/10.1101/2020.10.30.20169615v1>
38. Mujumdar A, Vaidehi V. Diabetes Prediction using Machine Learning Algorithms. *Procedia Comput Sci*. 2019 Jan 1;165:292–9.
39. Vathesatogkit P, Woodward M, Tanomsup S, Ratanachaiwong W, Vanavanan S, Yamwong S, et al. Cohort profile: the electricity generating authority of Thailand study. *Int J Epidemiol*. 2012 Apr;41(2):359–65.
40. EGAT STUDY | คณะแพทยศาสตร์โรงพยาบาลรามาธิบดี มหาวิทยาลัยมหิดล [Internet]. [cited 2023 Nov 3]. Available from: https://www.rama.mahidol.ac.th/cvmc/th/egatstudy_history
41. Thai CV risk score [Internet]. [cited 2023 Nov 3]. Available from: https://www.rama.mahidol.ac.th/cardio_vascular_risk/thai_cv_risk_score/

42. Saranburut K, Vathesatogkit P, Thongmung N, Chittamma A, Vanavanan S, Tangstheanphan T, et al. Risk scores to predict decreased glomerular filtration rate at 10 years in an Asian general population. BMC Nephrol [Internet]. 2017;18(1). Available from: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85024372638&doi=10.1186%2fs12882-017-0653-z&partnerID=40&md5=a383b6f11e9c4d23b41ef1c5da99183a>
43. Aekplakorn W, Bunnag P, Woodward M, Sritara P, Cheepudomwit S, Yamwong S, et al. A risk score for predicting incident diabetes in the Thai population. Diabetes Care. 2006;29(8):1872–7.
44. ร้อยละของผู้ป่วยเบาหวาน ความดันโลหิตสูงที่ขึ้นทะเบียนได้รับการประเมินโอกาสเสี่ยงต่อโรคหัวใจและหลอดเลือด (CVD Risk) [Internet]. [cited 2023 Nov 3]. Available from: <http://healthkpi.moph.go.th/kpi/kpi-list/view/?id=33>
45. คู่มือโครงสร้างมาตรฐานข้อมูล 43 แฟ้ม [Internet]. [cited 2023 Nov 3]. Available from: <https://backenddc.anamai.moph.go.th/index.php?r=handbook%2Findex&groupId=1>
46. Cox DR, Oakes D. Analysis of Survival Data.
47. Therneau TM, Grambsch PM. Modeling Survival Data: Extending the Cox Model [Internet]. New York, NY: Springer; 2000 [cited 2023 Nov 3]. (Dietz K, Gail M, Krickeberg K, Samet J, Tsiatis A, editors. Statistics for Biology and Health). Available from: <http://link.springer.com/10.1007/978-1-4757-3294-8>
48. Shumway T. Forecasting Bankruptcy More Accurately: A Simple Hazard Model. J Bus. 2001;74(1):101–24.
49. Suresh K, Severn C, Ghosh D. Survival prediction models: an introduction to discrete-time modeling. BMC Med Res Methodol. 2022 Jul 26;22(1):207.
50. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining [Internet]. New York, NY, USA: Association for Computing Machinery; 2016 [cited 2023 Nov 2]. p. 785–94. (KDD '16). Available from: <https://dl.acm.org/doi/10.1145/2939672.2939785>
51. Bureau of Trade and Economic Indices(CPI) [Internet]. [cited 2023 Nov 3]. Available from: http://www.price.moc.go.th/price/cpi/index_new_all.asp

52. รายการต้นทุนมาตรฐานเพื่อการประเมินเทคโนโลยีด้านสุขภาพ- ค้นหาข้อมูล [Internet]. [cited 2023 Nov 3]. Available from: <https://costingmenu.hitap.net/>
53. Dunbar SB, Khavjou OA, Bakas T, Hunt G, Kirch RA, Leib AR, et al. Projected Costs of Informal Caregiving for Cardiovascular Disease: 2015 to 2035: A Policy Statement From the American Heart Association. *Circulation*. 2018 May 8;137(19):e558–77.
54. Langa KM, Vijan S, Hayward RA, Chernew ME, Blaum CS, Kabeto MU, et al. Informal caregiving for diabetes and diabetic complications among elderly americans. *J Gerontol B Psychol Sci Soc Sci*. 2002 May;57(3):S177-186.
55. World Bank Open Data [Internet]. World Bank Open Data. [cited 2023 Nov 3]. Available from: <https://data.worldbank.org>
56. Risk models and scores for type 2 diabetes: systematic review | The BMJ [Internet]. [cited 2023 Nov 3]. Available from: <https://www.bmj.com/content/343/bmj.d7163>
57. Schmidt MI, Duncan BB, Bang H, Pankow JS, Ballantyne CM, Golden SH, et al. Identifying individuals at high risk for diabetes: The Atherosclerosis Risk in Communities study. *Diabetes Care*. 2005 Aug;28(8):2013–8.
58. AUSDRISK: an Australian Type 2 Diabetes Risk Assessment Tool based on demographic, lifestyle and simple anthropometric measures - PubMed [Internet]. [cited 2023 Nov 3]. Available from: <https://pubmed.ncbi.nlm.nih.gov/20170456/>
59. Prediction of incident diabetes mellitus in middle-aged adults: the Framingham Offspring Study - PubMed [Internet]. [cited 2023 Nov 3]. Available from: <https://pubmed.ncbi.nlm.nih.gov/17533210/>
60. Stern MP, Williams K, Haffner SM. Identification of persons at high risk for type 2 diabetes mellitus: do we need the oral glucose tolerance test? *Ann Intern Med*. 2002 Apr 16;136(8):575–81.
61. The diabetes risk score: a practical tool to predict type 2 diabetes risk - PubMed [Internet]. [cited 2023 Nov 3]. Available from: <https://pubmed.ncbi.nlm.nih.gov/12610029/>
62. Predicting risk of type 2 diabetes in England and Wales: prospective derivation and validation of QDScore | The BMJ [Internet]. [cited 2023 Nov 3]. Available from: <https://www.bmj.com/content/338/bmj.b880>

63. Rahman M, Simmons RK, Harding AH, Wareham NJ, Griffin SJ. A simple risk score identifies individuals at high risk of developing Type 2 diabetes: a prospective cohort study. *Fam Pract.* 2008 Jun;25(3):191–6.
64. Noble A, Amerasinghe P, Manthritilake H, Arasalingam S. Review of literature on chronic kidney disease of unknown etiology (CKDu) in Sri Lanka [Internet]. International Water Management Institute; 2014. (IWMI Working Papers; vol. 158). Available from: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85049444805&doi=10.5337%2f2014.206&partnerID=40&md5=29da203e8fd9a53246aad04e294b99d5>
65. Flor LS, Wilson S, Bhatt P, Bryant M, Burnett A, Camarda JN, et al. Community-based interventions for detection and management of diabetes and hypertension in underserved communities: a mixed-methods evaluation in Brazil, India, South Africa and the USA. *BMJ Glob Health.* 2020 Jun 1;5(6):e001959.
66. Lim JU, Lee JH, Kim JS, Hwang YI, Kim TH, Lim SY, et al. Comparison of World Health Organization and Asia-Pacific body mass index classifications in COPD patients. *Int J Chron Obstruct Pulmon Dis.* 2017 Aug 21;12:2465–75.
67. Zhao H, Zhang X, Xu Y, Gao L, Ma Z, Sun Y, et al. Predicting the Risk of Hypertension Based on Several Easy-to-Collect Risk Factors: A Machine Learning Method. *Front Public Health.* 2021 Sep 24;9:619429.
68. Pearson TA, LaCroix AZ, Mead LA, Liang KY. The prediction of midlife coronary heart disease and hypertension in young adults: the Johns Hopkins multiple risk equations. *Am J Prev Med.* 1990;6(2 Suppl):23–8.
69. Parikh NI, Pencina MJ, Wang TJ, Benjamin EJ, Lanier KJ, Levy D, et al. A risk score for predicting near-term incidence of hypertension: the Framingham Heart Study. *Ann Intern Med.* 2008 Jan 15;148(2):102–10.
70. Prediction of incident hypertension risk in women with currently normal blood pressure - PubMed [Internet]. [cited 2023 Nov 3]. Available from: <https://pubmed.ncbi.nlm.nih.gov/19375556/>

71. Kivimäki M, Batty GD, Singh-Manoux A, Ferrie JE, Tabak AG, Jokela M, et al. Validating the Framingham Hypertension Risk Score: results from the Whitehall II study. *Hypertens Dallas Tex* 1979. 2009 Sep;54(3):496–501.
72. Kshirsagar AV, Chiu YL, Bombback AS, August PA, Viera AJ, Colindres RE, et al. A hypertension risk score for middle-aged and older adults. *J Clin Hypertens Greenwich Conn*. 2010 Oct;12(10):800–8.
73. Bozorgmanesh M, Hadaegh F, Mehrabi Y, Azizi F. A point-score system superior to blood pressure measures alone for predicting incident hypertension: Tehran Lipid and Glucose Study. *J Hypertens*. 2011 Aug;29(8):1486–93.
74. Chien KL, Hsu HC, Su TC, Chang WT, Sung FC, Chen MF, et al. Prediction models for the risk of new-onset hypertension in ethnic Chinese in Taiwan. *J Hum Hypertens*. 2011 May;25(5):294–303.
75. Predicting the Risk of Incident Hypertension in a Korean Middle-Aged Population: Korean Genome and Epidemiology Study - Lim - 2013 - The Journal of Clinical Hypertension - Wiley Online Library [Internet]. [cited 2023 Nov 3]. Available from: <https://onlinelibrary.wiley.com/doi/10.1111/jch.12080>
76. Development of a risk prediction model for incident hypertension in a working-age Japanese male population - PubMed [Internet]. [cited 2023 Nov 3]. Available from: <https://pubmed.ncbi.nlm.nih.gov/25391458/>
77. Asgari S, Khalili D, Mehrabi Y, Kazempour-Ardebili S, Azizi F, Hadaegh F. Incidence and risk factors of isolated systolic and diastolic hypertension: a 10 year follow-up of the Tehran Lipids and Glucose Study. *Blood Press*. 2016 Jun;25(3):177–83.
78. Sathish T, Kannan S, Sarma PS, Razum O, Thrift AG, Thankappan KR. A Risk Score to Predict Hypertension in Primary Care Settings in Rural India. *Asia Pac J Public Health*. 2016 Jan;28(1 Suppl):26S-31S.
79. Kanegae H, Oikawa T, Suzuki K, Okawara Y, Kario K. Developing and validating a new precise risk-prediction model for new-onset hypertension: The Jichi Genki hypertension prediction model (JG model). *J Clin Hypertens Greenwich Conn*. 2018 May;20(5):880–90.

80. Chen Y, Wang C, Liu Y, Yuan Z, Zhang W, Li X, et al. Incident hypertension and its prediction model in a prospective northern urban Han Chinese cohort study. *J Hum Hypertens*. 2016 Dec;30(12):794–800.
81. Wang Y, Ma Z, Xu C, Wang Z, Yang X. Prediction of transfer among multiple states of blood pressure based on Markov model: an 18-year cohort study. *J Hypertens*. 2018 Jul;36(7):1506–13.
82. Syllos DH, Calsavara VF, Bensenor IM, Lotufo PA. Validating the Framingham Hypertension Risk Score: A 4-year follow-up from the Brazilian Longitudinal Study of the Adult Health (ELSA-Brasil). *J Clin Hypertens Greenwich Conn*. 2020 May;22(5):850–6.
83. Wang B, Liu Y, Sun X, Yin Z, Li H, Ren Y, et al. Prediction model and assessment of probability of incident hypertension: the Rural Chinese Cohort Study. *J Hum Hypertens*. 2021 Jan;35(1):74–84.
84. Xu F, Zhu J, Sun N, Wang L, Xie C, Tang Q, et al. Development and validation of prediction models for hypertension risks in rural Chinese populations. *J Glob Health*. 2019 Dec;9(2):020601.
85. Kadomatsu Y, Tsukamoto M, Sasakabe T, Kawai S, Naito M, Kubo Y, et al. A risk score predicting new incidence of hypertension in Japan. *J Hum Hypertens*. 2019 Oct;33(10):748–55.
86. Ture M, Kurt I, Turhan Kurum A, Ozdamar K. Comparing classification techniques for predicting essential hypertension. *Expert Syst Appl*. 2005 Oct 1;29(3):583–8.
87. Kanegae H, Suzuki K, Fukatani K, Ito T, Harada N, Kario K. Highly precise risk prediction model for new-onset hypertension using artificial intelligence techniques. *J Clin Hypertens*. 2020;22(3):445–50.
88. Ndjaboue R, Ngueta G, Rochefort-Brihay C, Delorme S, Guay D, Ivers N, et al. Prediction models of diabetes complications: a scoping review. *J Epidemiol Community Health*. 2022 Oct 1;76(10):896–904.
89. Basu S, Yudkin JS, Kehlenbrink S, Davies J. Estimation of Global Insulin Utilisation for Type 2 Diabetes Mellitus, 2018 to 2030.

90. Lim HM, Chia YC, Koay ZL. Performance of the Finnish Diabetes Risk Score (FINDRISC) and Modified Asian FINDRISC (ModAsian FINDRISC) for screening of undiagnosed type 2 diabetes mellitus and dysglycaemia in primary care. *Prim Care Diabetes*. 2020 Oct 1;14(5):494–500.
91. Chowdhury MZI, Naeem I, Quan H, Leung AA, Sikdar KC, O’Beirne M, et al. Prediction of hypertension using traditional regression and machine learning models: A systematic review and meta-analysis. *PLOS ONE*. 2022 Apr 7;17(4):e0266334.
92. Mumusoglu S, Yildiz BO. Metabolic Syndrome During Menopause. *Curr Vasc Pharmacol*. 2019;17(6):595–603.
93. Muntner P, Whittle J, Lynch AI, Colantonio LD, Simpson LM, Einhorn PT, et al. Visit-to-visit variability of blood pressure and coronary heart disease, stroke, heart failure and mortality: A cohort study. *Ann Intern Med*. 2015 Sep 1;163(5):329–38.
94. de Havenon A, Fino NF, Johnson B, Wong KH, Majersik JJ, Tirschwell D, et al. Blood Pressure Variability and Cardiovascular Outcomes in Patients With Prior Stroke. *Stroke*. 2019 Nov;50(11):3170–6.
95. Frontiers | Day-by-Day Blood Pressure Variability Is Associated With Neurological Functional Outcome After Acute Ischemic Stroke [Internet]. [cited 2023 Nov 3]. Available from: <https://www.frontiersin.org/articles/10.3389/fneur.2020.566825/full>
96. Kim MJ, Lim NK, Choi SJ, Park HY. Hypertension is an independent risk factor for type 2 diabetes: the Korean genome and epidemiology study. *Hypertens Res*. 2015 Nov;38(11):783–9.
97. Tsimihodimos V, Gonzalez-Villalpando C, Meigs JB, Ferrannini E. Hypertension and Diabetes Mellitus. *Hypertension*. 2018 Mar;71(3):422–8.
98. Allen L, Williams J, Townsend N, Mikkelsen B, Roberts N, Foster C, et al. Socioeconomic status and non-communicable disease behavioural risk factors in low-income and lower-middle-income countries: a systematic review. *Lancet Glob Health*. 2017 Mar 1;5(3):e277–89.
99. Fu YH, Xu TL, Rao ZZ, Liu JM, Li RT, Liu M, et al. [Forecasting the burden of disease from diabetes under the scenarios of specific risk factors control in China in 2030]. *Zhonghua Liu Xing Bing Xue Za Zhi Zhonghua Liuxingbingxue Zazhi*. 2023 Apr 10;44(4):581–6.

ภาคผนวก

ตารางผนวกที่ 1 จำนวนการรับบริการผู้ป่วยนอกและค่าบริการผู้ป่วยนอกเฉลี่ยต่อคนต่อปีแยกรายประเภทสถานะของโรค

| ประเภทของผู้ป่วย | จำนวนการรับบริการ ผู้ป่วยนอก (ครั้ง) | ค่าบริการผู้ป่วยนอก เฉลี่ย (SD) ต่อครั้ง (บาท) | จำนวนการรับบริการ ผู้ป่วยนอกเฉลี่ยต่อปี (ครั้ง) | ค่าบริการผู้ป่วยนอก เฉลี่ยต่อคนต่อปี (บาท) |
|-------------------|---|--|---|--|
| Non-DM & HT | 19,336,803 | 447 (3,753) | 4.5 | 2,030 |
| DM | 1,254,519 | 609 (4,235) | 10.2 | 6,198 |
| HT | 9,736,475 | 592 (4,829) | 9.2 | 5,450 |
| DM-HT | 3,719,454 | 742 (4,470) | 12.2 | 9,021 |
| DM-CARDIO | 6,945 | 1,588 (11,858) | 9.4 | 14,960 |
| DM-CEREBRO | 28,485 | 926 (6,283) | 11.1 | 10,304 |
| DM-EYE | 66,403 | 834 (4,125) | 11.5 | 9,626 |
| DM-KIDNEY | 110,118 | 644 (3,164) | 11.1 | 7,126 |
| HT-CARDIO | 157,612 | 1,384 (11,268) | 11.4 | 15,752 |
| HT-CEREBRO | 556,642 | 992 (7,316) | 11.3 | 11,170 |
| HT-EYE | 36,333 | 936 (5,293) | 10.6 | 9,951 |
| HT-KIDNEY | 639,993 | 1,550 (6,354) | 21.5 | 33,307 |
| DM-HT-CARDIO | 69,061 | 1,729 (12,038) | 13.6 | 23,602 |
| DM-HT-CEREBRO | 208,916 | 1,059 (7,618) | 13.6 | 14,367 |
| DM-HT-EYE | 321,646 | 939 (4,406) | 14.4 | 13,536 |
| DM-HT-KIDNEY | 946,767 | 1,062 (5,850) | 16.9 | 18,001 |
| DM-CARDIO-KIDNEY | 969 | 1,469 (13,023) | 11.4 | 16,745 |
| DM-CEREBRO-CARDIO | 589 | 1,572 (11,135) | 13.1 | 20,582 |
| DM-CEREBRO-EYE | 1,595 | 1,267 (7,596) | 10.2 | 12,959 |
| DM-CEREBRO-KIDNEY | 1,835 | 1,433 (23,444) | 9.1 | 13,085 |
| DM-EYE-CARDIO | 184 | 758 (1,166) | 7.7 | 5,810 |

| ประเภทของผู้ป่วย | จำนวนการรับบริการ ผู้ป่วยนอก (ครั้ง) | ค่าบริการผู้ป่วยนอก เฉลี่ย (SD) ต่อครั้ง (บาท) | จำนวนการรับบริการ ผู้ป่วยนอกเฉลี่ยต่อปี (ครั้ง) | ค่าบริการผู้ป่วยนอก เฉลี่ยต่อคนต่อปี (บาท) |
|-----------------------------|---|--|---|--|
| DM-EYE-KIDNEY | 10,391 | 716 (2,755) | 12.3 | 8,773 |
| HT-CARDIO-KIDNEY | 37,272 | 1,702 (9,141) | 32.5 | 55,249 |
| HT-CEREBRO-CARDIO | 18,624 | 1,485 (10,431) | 13.5 | 20,078 |
| HT-CEREBRO-EYE | 2,013 | 890 (4,046) | 12.4 | 11,060 |
| HT-CEREBRO-KIDNEY | 33,997 | 1,712 (10,538) | 18.6 | 31,893 |
| HT-EYE-CARDIO | 965 | 1,625 (9,526) | 15.3 | 24,896 |
| HT-EYE-KIDNEY | 13,937 | 2,833 (7,104) | 28.3 | 80,102 |
| DM-HT-CARDIO-KIDNEY | 43,397 | 2,150 (13,468) | 27.0 | 58,073 |
| DM-HT-CEREBRO-CARDIO | 6,331 | 2,709 (21,155) | 15.3 | 41,528 |
| DM-HT-CEREBRO-EYE | 24,064 | 1,271 (6,499) | 16.3 | 20,682 |
| DM-HT-CEREBRO-KIDNEY | 75,908 | 1,320 (5,895) | 20.4 | 26,959 |
| DM-HT-EYE-CARDIO | 7,963 | 1,662 (11,155) | 17.9 | 29,811 |
| DM-HT-EYE-KIDNEY | 219,325 | 1,458 (7,056) | 24.8 | 36,174 |
| DM-CEREBRO-EYE-KIDNEY | 338 | 798 (1,450) | 12.1 | 9,630 |
| DM-EYE-CARDIO-KIDNEY | 94 | 3,059 (15,936) | 10.4 | 31,950 |
| DM-CEREBRO-CARDIO-KIDNEY | 100 | 470 (1,369) | 20.0 | 9,403 |
| DM-CEREBRO-CARDIO-EYE | 5 | 386 (512) | 5.0 | 1,928 |
| HT-CEREBRO-CARDIO-KIDNEY | 3,832 | 1,715 (10,156) | 33.9 | 58,151 |
| HT-CEREBRO-EYE-CARDIO | 60 | 1,043 (3,267) | 12.0 | 12,520 |
| HT-CEREBRO-EYE-KIDNEY | 672 | 2,005 (5,255) | 15.3 | 30,623 |
| HT-EYE-CARDIO-KIDNEY | 662 | 2,171 (5,048) | 44.1 | 95,821 |
| DM-HT-CEREBRO-CARDIO-KIDNEY | 5,828 | 2,474 (14,323) | 27.9 | 68,989 |

| ประเภทของผู้ป่วย | จำนวนการรับบริการ ผู้ป่วยนอก (ครั้ง) | ค่าบริการผู้ป่วยนอก เฉลี่ย (SD) ต่อครั้ง (บาท) | จำนวนการรับบริการ ผู้ป่วยนอกเฉลี่ยต่อปี (ครั้ง) | ค่าบริการผู้ป่วยนอก เฉลี่ยต่อคนต่อปี (บาท) |
|---------------------------------|---|--|---|--|
| DM-HT-CEREBRO-EYE-CARDIO | 786 | 1,005 (3,833) | 19.2 | 19,259 |
| DM-HT-CEREBRO-EYE-KIDNEY | 25,306 | 1,661 (8,379) | 29.5 | 48,939 |
| DM-HT-EYE-CARDIO-KIDNEY | 15,445 | 2,302 (9,910) | 40.6 | 93,564 |
| DM-CEREBRO-EYE-CARDIO-KIDNEY | 1 | 1,632 (.) | 1.0 | 1,632 |
| HT-CEREBRO-EYE-CARDIO-KIDNEY | 50 | 1,604 (3,139) | 16.7 | 26,737 |
| DM-HT-CEREBRO-EYE-CARDIO-KIDNEY | 2,264 | 2,547 (13,456) | 32.3 | 82,390 |

Non-DM & HT: ผู้ที่ไม่เป็นโรคเบาหวานและความดันโลหิตสูง, DM: โรคเบาหวาน, HT: โรคความดันโลหิตสูง, CARDIO: โรคหลอดเลือดหัวใจตีบ, CEREBRO: โรคหลอดเลือดหัวใจแตกหรือตีบ, EYE: ภาวะแทรกซ้อนทางตา, KIDNEY: โรคไตวายเรื้อรังระยะที่ 5

ตารางผนวกที่ 2 จำนวนการรับบริการผู้ป่วยในและค่าบริการผู้ป่วยในเฉลี่ยต่อคนต่อปีแยกแยะประเภทสถานะของโรค

| ประเภทของผู้ป่วย | จำนวนการรับบริการผู้ป่วยใน (ครั้ง) | จำนวนวันนอนเฉลี่ย (SD) ต่อครั้ง (วัน) | ค่าบริการผู้ป่วยในเฉลี่ย (SD) ต่อครั้ง (บาท) | จำนวนการรับบริการผู้ป่วยในเฉลี่ยต่อปี (ครั้ง) | ค่าบริการผู้ป่วยในเฉลี่ยต่อคนต่อปี (บาท) |
|-------------------|------------------------------------|---------------------------------------|--|---|--|
| Non-DM&HT | 309,303 | 4.1 (5.4) | 18,357 (48,527) | 0.07 | 1,333 |
| DM | 20,786 | 4.7 (6.0) | 19,795 (40,941) | 0.17 | 3,338 |
| HT | 159,887 | 4.3 (6.0) | 21,264 (52,981) | 0.15 | 3,216 |
| DM-HT | 65,991 | 4.4 (5.7) | 21,150 (49,232) | 0.22 | 4,560 |
| DM-CARDIO | 381 | 4.6 (5.7) | 34,076 (62,504) | 0.52 | 17,616 |
| DM-CEREBRO | 1,340 | 4.3 (5.9) | 22,360 (46,035) | 0.52 | 11,704 |
| DM-EYE | 1,013 | 4.3 (6.8) | 18,405 (31,423) | 0.18 | 3,241 |
| DM-KIDNEY | 2,173 | 4.5 (6.6) | 15,856 (33,444) | 0.22 | 3,462 |
| HT-CARDIO | 8,192 | 4.2 (5.6) | 27,792 (60,717) | 0.59 | 16,441 |
| HT-CEREBRO | 23,901 | 4.8 (7.5) | 23,499 (51,156) | 0.48 | 11,357 |
| HT-EYE | 704 | 3.7 (5.2) | 24,964 (44,710) | 0.21 | 5,145 |
| HT-KIDNEY | 20,271 | 4.4 (6.0) | 19,607 (43,308) | 0.68 | 13,349 |
| DM-HT-CARDIO | 3,727 | 4.3 (5.4) | 29,176 (56,082) | 0.74 | 21,490 |
| DM-HT-CEREBRO | 10,150 | 4.6 (7.0) | 22,520 (48,940) | 0.66 | 14,839 |
| DM-HT-EYE | 6,682 | 4.2 (5.5) | 20,796 (37,645) | 0.30 | 6,228 |
| DM-HT-KIDNEY | 30,184 | 4.4(5.6) | 19,036 (44,287) | 0.54 | 10,285 |
| DM-CARDIO-KIDNEY | 46 | 4.4(3.7) | 30,590 (60,459) | 0.54 | 16,555 |
| DM-CEREBRO-CARDIO | 57 | 5.3(7.8) | 30,861 (59,006) | 1.27 | 39,090 |
| DM-CEREBRO-EYE | 75 | 3.6(4.5) | 19,452 (36,449) | 0.48 | 9,352 |
| DM-CEREBRO-KIDNEY | 133 | 3.6(4.2) | 12,904 (18,855) | 0.66 | 8,538 |

| ประเภทของผู้ป่วย | จำนวนการรับบริการ ผู้ป่วยใน (ครั้ง) | จำนวนวันนอน เฉลี่ย (SD) ต่อ ครั้ง (วัน) | ค่าบริการผู้ป่วยใน เฉลี่ย (SD) ต่อครั้ง (บาท) | จำนวนการรับ บริการผู้ป่วยใน เฉลี่ยต่อปี (ครั้ง) | ค่าบริการ ผู้ป่วยในเฉลี่ย ต่อคนต่อปี (บาท) |
|--------------------------|--|---|---|---|---|
| DM-EYE-CARDIO | 8 | 3.2(1.8) | 16,906 (13,895) | 0.33 | 5,635 |
| DM-EYE-KIDNEY | 203 | 4.8(5.6) | 17,925 (28,406) | 0.24 | 4,291 |
| HT-CARDIO-KIDNEY | 1,774 | 4.7(6.4) | 23,119 (44,710) | 1.55 | 35,725 |
| HT-CEREBRO-CARDIO | 1,172 | 4.9(7.0) | 27,900 (54,150) | 0.85 | 23,747 |
| HT-CEREBRO-EYE | 72 | 5.1(9.2) | 22,757 (43,732) | 0.44 | 10,114 |
| HT-CEREBRO-KIDNEY | 1,825 | 4.9(6.0) | 24,262 (54,029) | 1.00 | 24,262 |
| HT-EYE-CARDIO | 26 | 4.7(8.0) | 35,379 (55,857) | 0.41 | 14,601 |
| HT-EYE-KIDNEY | 405 | 3.8(4.7) | 19,255 (30,991) | 0.82 | 15,818 |
| DM-HT-CARDIO-KIDNEY | 2,614 | 4.5(5.5) | 25,202 (51,013) | 1.63 | 40,995 |
| DM-HT-CEREBRO-CARDIO | 562 | 5.0(7.7) | 28,397 (57,377) | 1.36 | 38,642 |
| DM-HT-CEREBRO-EYE | 1,186 | 4.7(7.4) | 21,287 (36,938) | 0.80 | 17,070 |
| DM-HT-CEREBRO-KIDNEY | 4,489 | 4.5(6.2) | 19,752 (43,640) | 1.21 | 23,848 |
| DM-HT-EYE-CARDIO | 388 | 5.4 (6.1) | 33,577 (62,074) | 0.87 | 29,342 |
| DM-HT-EYE-KIDNEY | 8,341 | 4.4 (5.4) | 20,818 (41,104) | 0.94 | 19,638 |
| DM-CEREBRO-EYE-KIDNEY | 8 | 2.7 (1.9) | 7,846 (6,935) | 0.29 | 2,242 |
| DM-EYE-CARDIO-KIDNEY | 9 | 4.0 (2.7) | 30,673 (46,236) | 1.00 | 30,673 |
| DM-CEREBRO-CARDIO-KIDNEY | 4 | 6.0 (5.3) | 22,543 (25,177) | 0.80 | 18,034 |
| HT-CEREBRO-CARDIO-KIDNEY | 260 | 5.3 (7.0) | 27,187 (53,947) | 2.30 | 62,554 |
| HT-CEREBRO-EYE-CARDIO | 3 | 1.8 (0.9) | 27,817 (16,608) | 0.60 | 16,690 |
| HT-CEREBRO-EYE-KIDNEY | 35 | 3.8 (4.5) | 20,426 (49,324) | 0.80 | 16,248 |
| HT-EYE-CARDIO-KIDNEY | 13 | 5.4 (8.7) | 30,921 (41,220) | 0.87 | 26,798 |

| ประเภทของผู้ป่วย | จำนวนการรับบริการ ผู้ป่วยใน (ครั้ง) | จำนวนวันนอน เฉลี่ย (SD) ต่อ ครั้ง (วัน) | ค่าบริการผู้ป่วยใน เฉลี่ย (SD) ต่อครั้ง (บาท) | จำนวนการรับ บริการผู้ป่วยใน เฉลี่ยต่อปี (ครั้ง) | ค่าบริการ ผู้ป่วยในเฉลี่ย ต่อคนต่อปี (บาท) |
|---------------------------------|--|---|---|---|---|
| DM-HT-CEREBRO-CARDIO-KIDNEY | 444 | 5.2 (6.6) | 25,741 (48,985) | 2.12 | 54,683 |
| DM-HT-CEREBRO-EYE-CARDIO | 57 | 3.7 (3.8) | 21,092 (38,347) | 1.39 | 29,323 |
| DM-HT-CEREBRO-EYE-KIDNEY | 1,393 | 4.3 (5.7) | 20,251 (37,299) | 1.62 | 32,840 |
| DM-HT-EYE-CARDIO-KIDNEY | 853 | 5.4 (8.7) | 30,058 (109,442) | 2.24 | 67,472 |
| HT-CEREBRO-EYE-CARDIO-KIDNEY | 3 | 5.2 (3.1) | 32,966 (1,913) | 1.00 | 32,966 |
| DM-HT-CEREBRO-EYE-CARDIO-KIDNEY | 234 | 5.5 (7.7) | 32,168 (59,886) | 3.34 | 107,534 |

Non-DM & HT: ผู้ที่ไม่เป็นโรคเบาหวานและความดันโลหิตสูง, DM: โรคเบาหวาน, HT: โรคความดันโลหิตสูง, CARDIO: โรคหลอดเลือดหัวใจตีบ, CEREBRO: โรคหลอดเลือดหัวใจแตกหรือตีบ, EYE: ภาวะแทรกซ้อนทางตา, KIDNEY: โรคไตวายเรื้อรังระยะที่ 5