

แบบจำลองการประเมินความเสี่ยงที่เกี่ยวข้องกับการติดเชื้อเอชไอวีด้วยการเรียนรู้ด้วยเครื่อง

แพรพรรณ พุ่มโพธิ์สุวรรณ¹, นภา แซ่เปี๊ยะ

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อสร้างแบบจำลองการทำนายระดับความเสี่ยงที่เกี่ยวข้องกับการติดเชื้อเอชไอวีโดยอาศัยกลุ่มตัวแปร 4 กลุ่ม คือ ข้อมูลทางด้านประชากร (Demographics), รูปแบบการดำเนินชีวิต (Lifestyle), พฤติกรรมทางเพศ (Sexual behavior) และ อาการโรคติดต่ออื่นๆ (Symptoms) โดยใช้เทคนิคการเรียนรู้ของเครื่อง 7 แบบ คือ XGBoost Classifier, XGBoost Regressor, Decision tree, Random Forest, Support Vector Machine (SVM), Support Vector Regressor (SVR) และ Regression Neural Network โดยใช้ชุดข้อมูลจากสถาบันเพื่อการวิจัยและนวัตกรรมด้านเอชไอวี ในช่วงวันที่ 1 มกราคม 2564 จนถึง วันที่ 30 ธันวาคม 2565 จำนวนข้อมูลทั้งหมด 3,621 แถว ซึ่งเป็นข้อมูลของผู้ที่มารับบริการเกี่ยวกับด้านการตรวจเอชไอวี การรับยา เพร็พ (PrEP) หรือ เป็ป(PEP) เพื่อป้องกันการติดเชื้อ HIV และ การรักษาโรคติดต่อทางเพศสัมพันธ์ ซึ่งจะต้องมีผลการตรวจเอชไอวี เป็น ลบ(negative) เท่านั้น เพื่อประเมินความเสี่ยงการติดเชื้อเอชไอวี โดยแบ่งออกเป็น 4 ระดับ คือ ไม่มีความเสี่ยง ความเสี่ยงต่ำ ความเสี่ยงปานกลาง และความเสี่ยงสูง ผู้วิจัยสนใจที่จะเปรียบเทียบผลของการจัดกลุ่ม โดยนำข้อมูลแต่ละกลุ่มเข้าแบบจำลองเพื่อทำนายความเสี่ยงการติดเชื้อเอชไอวี ซึ่งการใช้ข้อมูลทั้งหมดโดยไม่ใช้ข้อมูลแค่กลุ่มใดกลุ่มหนึ่งจะให้ผลลัพธ์ในการทำนายได้ดีกว่า จากนั้นแบ่งกลุ่มสำหรับทดลองแบบจำลองต่างๆตามตัวแปรตามออกเป็น 4 กลุ่ม ดังนี้ Multi-Class (N : L : M : H), Binary Class 1 (N, L, M : H), Binary Class 2 (N, L : M, H), Binary Class 3 (N : L, M, H) ผลลัพธ์การเปรียบเทียบผลของแต่ละแบบจำลองกับกลุ่มตัวแปรตาม โดยมีค่า AUC สูงสุดที่ 0.88 ในแบบจำลอง Random Forest ในกลุ่ม Binary Class 2 (N, L : M, H) แต่เนื่องจากงานวิจัยนี้ต้องการให้แบบจำลองทำนายเน้นคลาสความเสี่ยงสูง ซึ่งในกลุ่มที่สามารถเน้นคลาสความเสี่ยงสูงได้ดีคือกลุ่ม Binary Class 3 (N : L, M, H) และแบบจำลองที่มีค่า AUC สูงสุดคือ XGBoost Regressor ที่ 0.79

คำสำคัญ : การประเมินความเสี่ยงเอชไอวี, การเรียนรู้ของเครื่อง

¹ สาขาวิทยาการข้อมูล คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ กรุงเทพฯ 10110

² คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ กรุงเทพฯ 10110

* Corresponding author: Tel.: 081-6262509 E-mail address: pearparn.ppsw@g.swu.ac.th

MACHINE LEARNING TECHNIQUES FOR ASSESSING RISK BEHAVIOUR ASSOCIATED
WITH CONTRACTING HIV

Pearparn Pumphosuwan^{1*}, Napa Saebae²

Abstract

The objective of this research is to develop a predictive model for assessing the risk levels associated with HIV infection, utilizing four groups of variables: Demographics, Lifestyle, Sexual behavior, and Symptoms. Seven machine learning techniques were employed: XGBoost Classifier, XGBoost Regressor, Decision tree, Random Forest, Support Vector Machine (SVM), Support Vector Regressor (SVR), and Regression Neural Network. Data were sourced from an HIV research and innovation institution spanning from January 1, 2021, to December 30, 2022, comprising 3,621 records. These records pertain to individuals seeking services related to HIV testing, receiving medication, PrEP (Pre-Exposure Prophylaxis), or PEP (Post-Exposure Prophylaxis) for preventing HIV infection, and sexually transmitted infection treatment. Only negative HIV test results were considered to evaluate the risk of HIV infection, categorized into four levels: No risk, Low risk, Moderate risk, and High risk. The researchers aimed to compare the grouping effects by inputting each group's data into the model to predict HIV infection risk. Utilizing all data rather than specific group data yielded better prediction results. Subsequently, the groups for model testing were divided into four categories: Multi-Class (N : L : M : H), Binary Class 1 (N, L, M : H), Binary Class 2 (N, L : M, H), and Binary Class 3 (N : L, M, H). The comparison results of each model against variable groups showed the highest AUC value of 0.88 in the Random Forest model within the Binary Class 2 (N, L : M, H). However, as this research aims to prioritize high-risk class predictions, the group that performs best in emphasizing high-risk classes is Binary Class 3 (N : L, M, H), with the highest AUC value achieved by the XGBoost Regressor at 0.79.

Keywords : HIV risk assessment, Machine Learning

¹ Data Science, Faculty of Science, Srinakharinwirot University, Bangkok, 10110, Thailand

² Faculty of Science, Srinakharinwirot University, Bangkok, 10110, Thailand

* Corresponding author: Tel.: 081-6262509 E-mail address: pearparn.ppsw@g.swu.ac.th

บทนำ

เอชไอวีและเอดส์เป็นหนึ่งในสาเหตุหลักของการเสียชีวิตทั่วโลก และยังเป็นโรคที่คุกคามมนุษยชาติที่ร้ายแรงที่สุดโรคหนึ่งในปัจจุบันนี้ องค์การอนามัยโลกได้เปิดเผยข้อมูลว่าในปี 2015 มีประชากรติดเชื้อเอชไอวีหรือเอดส์จำนวน 36.7 ล้านคนทั่วโลก [1] ปัจจุบันโลกกำลังเริ่มต้นใน Fast-Track เพื่อยุติการแพร่ระบาดของโรคเอดส์ภายในปี 2030 ซึ่งเป้าหมายคือ 95-95-95 สำหรับการทดสอบเชื้อเอชไอวี การรักษา และการลดระดับจำนวนเชื้อไวรัสสำเร็จ [2] ในประเทศไทย กรมควบคุมโรคได้เปิดเผยข้อมูลเกี่ยวกับสถานการณ์เอชไอวีปี 2022 ว่าจำนวนผู้ติดเชื้อเอชไอวีรายใหม่ 9,230 คน/ปี แบ่งเป็น เด็ก (<15ปี) 54 คน เยาวชน (15-24ปี) 4,379 คน ผู้ใหญ่ (>24ปี) 4,797 คน และการติดเชื้อรายใหม่ที่ติดเชื้อจากการฉีดสารเสพติดที่ไม่ปลอดภัย 4% และ 96% เกิดจากการมีเพศสัมพันธ์ที่ไม่ป้องกัน ซึ่งแบ่งเป็น เพศสัมพันธ์ระหว่างชายกับชาย 68% คู่ผลเลือดต่าง (คู่อุยกิน/คู่ประจำ) 19% คู่ถอนชั่วคราวและนอกสมรส 8% เพศสัมพันธ์จากการซื้อขายบริการ 1% ซึ่งผู้ติดเชื้อจากเอดส์ในปี 2022 ทั้งหมด 10,972 คน/ปี แบ่งเป็น เด็ก (<15ปี) 81 คน เยาวชน(15-24ปี) 237 คน ผู้ใหญ่ (>24ปี) 10,654 คน และในด้านการดูแลรักษาตามเป้าหมาย 95-95-95 ในประเทศไทย 90%ของผู้ติดเชื้อเอชไอวีที่รู้สถานะว่าตนเองติดเชื้อฯ 90% ของผู้ติดเชื้อฯที่รู้สถานะ กำลังรับยาต้านไวรัส 97%ของผู้ที่รับยาต้านไวรัสสำเร็จ จะเห็นได้ว่ายังมีผู้ติดเชื้อเอชไอวีที่ไม่ทราบสถานะว่าตนติดเชื้อเอชไอวี[3]

วิธีดำเนินการ

ขั้นตอนที่ 1 : ข้อมูลที่ใช้ในการศึกษานี้

ผู้วิจัยนำข้อมูลเกี่ยวกับการประเมินความเสี่ยงการติดเชื้อเอชไอวี โดยใช้ชุดข้อมูลจากสถาบันเพื่อการวิจัยและนวัตกรรมด้านเอชไอวี ในช่วงวันที่ 1 มกราคม 2564 จนถึง วันที่ 30 ธันวาคม 2565 จำนวนข้อมูลทั้งหมด 3,621 แถว ซึ่งเป็นข้อมูลของผู้ที่มาใช้บริการเกี่ยวกับด้านการตรวจเอชไอวี การรับยา เพร็พ (PrEP) หรือ เป็บ(PEP) เพื่อป้องกันการติดเชื้อ HIV และ การรักษาโรคติดต่อทางเพศสัมพันธ์ ซึ่งจะต้องมีผลการตรวจเอชไอวี เป็น ลบ(negative) เท่านั้น โดยประกอบด้วยตัวแปรตามคือ ไม่มีความเสี่ยง (No risk (N)), ความเสี่ยงต่ำ (Low risk (L)), ความเสี่ยงปานกลาง (Medium risk (M)) และ ความเสี่ยงสูง (High risk (H)) และตัวแปรอิสระ 24 ตัวแปร โดยแบ่งกลุ่มของตัวแปรอิสระ เป็น 4 กลุ่มคือ ข้อมูลทางด้านประชากร (Demographics) 5 ตัวแปร, ข้อมูลทางด้านประชากร (Lifestyle) 4 ตัวแปร, พฤติกรรมทางเพศ (Sexual behavior) 8 ตัวแปร และ อาการอื่นๆ (Symptoms) 4 ตัวแปร

ขั้นตอนที่ 2 : การนำเข้าข้อมูล ตรวจสอบข้อมูล และพิจารณาข้อมูล

เนื่องจากชุดข้อมูลแยกตามแบบฟอร์มในการเก็บข้อมูล จึงทำการรวมข้อมูลในอยู่ในชุดเดียวโดยมี UID ซึ่งเป็นรหัสของผู้รับบริการ เป็นหลัก และเลือกข้อมูลเฉพาะนัดหมาย Baseline ของผู้เข้ารับบริการ จัดการกับข้อมูลที่ขาดหายไป (Missing Value) โดยการแปลงให้เป็นค่า ไม่ขอตอบ (prefer not to answer (PNA)) จากนั้นเริ่มกระบวนการตรวจสอบและสำรวจข้อมูล

เบื้องต้น เพื่อหาข้อมูลเชิงลึกจากข้อมูลธุรกรรมโดยใช้ไลบรารี Pandas, Numpy และ Dataprep.eda สืบหาข้อมูลตัวแปรตามกับตัวแปรอิสระ

ขั้นตอนที่ 3 : การเตรียมข้อมูลเพื่อสร้างแบบจำลอง

ผู้วิจัยใช้วิธี One-Hot encoding ก่อนนำข้อมูลเข้าแบบจำลอง โดยหลังจากใช้วิธีดังกล่าวมีข้อมูล 3621 แถว และ 90 คอลัมน์ และจากการสำรวจข้อมูลตัวแปรตามกับตัวแปรอิสระ ข้อมูลมีแนวโน้มเป็นข้อมูลที่ไม่สมดุล (Imbalanced Data) จึงใช้วิธีสุ่มข้อมูลเพิ่ม (Oversampling) เพื่อจัดการกับปัญหาข้อมูลที่ไม่สมดุลในการเรียนรู้ของแบบจำลอง ซึ่งจะทำให้การสุ่มข้อมูลจากคลาสน้อย (minority class) เพื่อสร้างตัวอย่างใหม่ในการเพิ่มจำนวนของคลาสน้อยขึ้นจนสมดุลกับคลาสมาก (majority class)

ขั้นตอนที่ 4 : การสร้างแบบจำลอง

ในการศึกษาครั้งนี้ ผู้วิจัยเปรียบเทียบแบบจำลองการจำแนกประเภท (Classification) 4 แบบ คือ Decision tree, Random Forest, XGBoost Classifier, Support Vector Machine (SVM) และแบบจำลองการถดถอย (Regression) 4 แบบ คือ XGBoost Regressor, Support Vector Regressor (SVR), Regression Neural Network โดยแบ่งกลุ่มสำหรับทดลองแบบจำลองต่างๆตามตัวแปรตามออกเป็น 4 กลุ่ม ดังนี้ Multi-Class (N : L : M : H), Binary Class 1 (N, L, M : H), Binary Class 2 (N, L : M, H), Binary Class 3 (N : L, M, H)

ผลการวิจัยและอภิปรายผลการวิจัย

งานวิจัยนี้มุ่งเน้นประเมินความเสี่ยงการติดเชื้อเอชไอวี โดยการประเมินผลลัพธ์ความแม่นยำของแบบจำลองการจำแนกประเภทที่ต่างกันของแต่ละแบบจำลอง ซึ่งประกอบด้วยค่า Precision, Recall, และ F1-score สำหรับแต่ละคลาส จากตารางที่ 1 ผลการทดลองใน Multi-Class (N : L : M : H) สามารถสังเกตได้ว่าค่า Precision, Recall, และ F1-score สำหรับคลาส 3 (H) จะมีค่าต่ำกว่าคลาสอื่นๆ อาจแสดงถึงความยากลำบากในการจำแนกประเภทของคลาสนี้ ซึ่งจะทำให้พลาดผู้ที่มีความเสี่ยงสูงไป อาจทำให้ผู้ป่วยนั้นเข้าใจผิดได้ว่าไม่มีความเสี่ยงและส่งต่อเชื้อไปให้ผู้อื่น ดังนั้นผู้วิจัยจึงจัดกลุ่มความเสี่ยงใหม่เพื่อเพิ่มจำนวนตัวอย่างให้มากขึ้น

ตารางที่ 1 ผลลัพธ์การทดลองแบบจำลองการจำแนกประเภท ในกลุ่ม Multi-Class (N : L : M : H)

| Group Class | Model Type | Class | Precision | Recall | F1-score |
|--------------------------------|---------------|-------|-----------|--------|----------|
| Multi-Class (N : L : M : H) | Decision tree | 0 (N) | 0.67 | 0.64 | 0.65 |
| | | 1 (L) | 0.54 | 0.38 | 0.44 |
| | | 2 (M) | 0.21 | 0.24 | 0.23 |

| Group Class | Model Type | Class | Precision | Recall | F1-score |
|-------------|------------------------|-------|-----------|--------|----------|
| | Random Forest | 3 (H) | 0.21 | 0.48 | 0.29 |
| | | 0 (N) | 0.70 | 0.63 | 0.66 |
| | | 1 (L) | 0.56 | 0.50 | 0.53 |
| | | 2 (M) | 0.18 | 0.22 | 0.20 |
| | | 3 (H) | 0.25 | 0.37 | 0.30 |
| | XGBoost Classifier | 0 (N) | 0.72 | 0.61 | 0.66 |
| | | 1 (L) | 0.58 | 0.51 | 0.54 |
| | | 2 (M) | 0.18 | 0.21 | 0.19 |
| | | 3 (H) | 0.27 | 0.45 | 0.34 |
| | Support Vector Machine | 0 (N) | 0.69 | 0.59 | 0.64 |
| | | 1 (L) | 0.57 | 0.43 | 0.49 |
| | | 2 (M) | 0.15 | 0.20 | 0.17 |
| | | 3 (H) | 0.23 | 0.46 | 0.30 |

การทดลองในกลุ่ม Binary Class 1 (N, L, M : H) เป็นการประเมินแบบจำลองในรูปแบบ Binary Classification ซึ่งแบ่งข้อมูลออกเป็น 2 คลาส คือ N, L, M, และ H (คลาสที่มีความเสี่ยง) โดยทำการแบ่งคลาสที่มีความเสี่ยงมากเพื่อทดลองว่าแบบจำลองจะสามารถทำนายคลาสที่มีความเสี่ยงสูงได้ดีขึ้นหรือไม่ จากผลการทดลองในตารางที่ 2 สามารถสรุปได้ว่าค่า Precision มีค่าสูงในการจำแนกคลาส 0 (N, L, M) และมีค่าต่ำในการจำแนกคลาสที่มีความเสี่ยง 1 (H) ทั้งนี้เนื่องเพราะการจัดกลุ่มนี้เป็นการแยกคลาสที่มีความเสี่ยงออกมา ซึ่งทำให้มีตัวอย่างที่น้อย

ตารางที่ 2 ผลลัพธ์การทดลองแบบจำลองการจำแนกประเภท ในกลุ่ม Binary Class 1 (N, L, M : H)

| Group Class | Model Type | Class | Precision | Recall | F1-score |
|------------------------------|---------------|-------------|-----------|--------|----------|
| Binary Class 1 (N, L, M : H) | Decision tree | 0 (N, L, M) | 0.94 | 0.76 | 0.84 |
| | | 1 (H) | 0.13 | 0.4 | 0.19 |
| | Random Forest | 0 (N, L, M) | 0.95 | 0.78 | 0.86 |
| | | 1 (H) | 0.18 | 0.54 | 0.27 |

| Group Class | Model Type | Class | Precision | Recall | F1-score |
|-------------|------------------------|-------------|-----------|--------|----------|
| | XGBoost Classifier | 0 (N, L, M) | 0.96 | 0.70 | 0.81 |
| | | 1 (H) | 0.16 | 0.65 | 0.25 |
| | Support Vector Machine | 0 (N, L, M) | 0.97 | 0.71 | 0.82 |
| | | 1 (H) | 0.18 | 0.72 | 0.28 |

จากนั้นผู้วิจัยจึงทำการทดลองในกลุ่ม Binary Class 2 (N, L : M, H) ซึ่งแบ่งข้อมูลออกเป็น 2 คลาส คือ N, L และ M, H เพื่อเพิ่มจำนวนตัวอย่างของกลุ่มที่มีความเสี่ยงขึ้นเพื่อไม่ให้เกิดโอกาสที่ผู้ป่วยจะประมาทจากการที่แบบจำลองทำนายความเสี่ยงสูงได้ต่ำ จากตารางที่ 3 สามารถสังเกตเห็นได้ว่าค่า Precision ยังคงมีค่าสูงในการจำแนกคลาสไม่มีความเสี่ยง และความเสี่ยงต่ำ (N, L) และมีค่าต่ำในการจำแนกคลาสที่มีความเสี่ยงปานกลาง และความเสี่ยงสูง (M, H) ซึ่งอาจสรุปได้ว่าการแบ่งกลุ่มในรูปแบบนี้ก็ยังไม่อาจให้แบบจำลองทำนายความเสี่ยงสูงได้ดีขึ้น

ตารางที่ 3 ผลลัพธ์การทดลองแบบจำลองการจำแนกประเภท ในกลุ่ม Binary Class 2 (N, L : M, H)

| Group Class | Model Type | Class | Precision | Recall | F1-score |
|---------------------------------|------------------------|----------|-----------|--------|----------|
| Binary Class 2 (N, L : M, H) | Decision tree | 0 (N, L) | 0.78 | 0.68 | 0.72 |
| | | 1 (M,H) | 0.34 | 0.46 | 0.39 |
| | Random Forest | 0 (N, L) | 0.82 | 0.65 | 0.73 |
| | | 1 (M,H) | 0.39 | 0.61 | 0.47 |
| | XGBoost Classifier | 0 (N, L) | 0.85 | 0.62 | 0.71 |
| | | 1 (M,H) | 0.40 | 0.69 | 0.50 |
| | Support Vector Machine | 0 (N, L) | 0.84 | 0.63 | 0.72 |
| | | 1 (M,H) | 0.40 | 0.68 | 0.50 |

สำหรับการทดลองในกลุ่ม Binary Class 3 (N : L, M, H) โดยการแบ่งคลาสที่มีความเสี่ยง (L, M, H) ให้เป็นคลาสเดียวกัน เพื่อให้เน้นการจำแนกคลาสที่มีความเสี่ยงได้ ผลการทดลองแสดงค่า Precision, Recall, และ F1-score สูง ซึ่งแบบจำลองที่มีผลการทำนายที่แม่นยำได้ดีคือ Random Forest และ Support Vector Machine

ตารางที่ 4 ผลลัพธ์การทดลองแบบจำลองการจำแนกประเภท ในกลุ่ม Binary Class 3 (N : L, M, H)

| Group Class | Model Type | Class | Precision | Recall | F1-score |
|---------------------------------|---------------------------|--------------|-----------|--------|----------|
| Binary Class 3 (N : L, M, H) | Decision tree | 0 (N) | 0.57 | 0.65 | 0.61 |
| | | 1 (L,M,H) | 0.83 | 0.78 | 0.80 |
| | Random Forest | 0 (N) | 0.68 | 0.62 | 0.65 |
| | | 1 (L,M,H) | 0.84 | 0.87 | 0.85 |
| | XGBoost Classifier | 0 (N) | 0.62 | 0.68 | 0.65 |
| | | 1 (L,M,H) | 0.85 | 0.81 | 0.83 |
| | Support Vector Machine | 0 (N) | 0.64 | 0.69 | 0.67 |
| | | 1 (L,M,H) | 0.86 | 0.83 | 0.84 |

ต่อมาผู้วิจัยได้ทดลองในแบบจำลองการถดถอย (Regression) โดยการประเมินผลลัพธ์ความแม่นยำของแบบจำลองการถดถอยที่แตกต่างกันของแต่ละแบบจำลอง ซึ่งประกอบด้วยค่า MAE, MSE, และ RMSE สำหรับแต่ละคลาส จากตารางที่ 5 จะสังเกตได้ว่า ในกลุ่ม Multi-target จะมีค่า MAE, MSE, และ RMSE ที่ค่อนข้างสูง และกลุ่มที่มีค่าที่ค่อนข้างต่ำที่สุดคือ Binary Class 1 (N, L, M : H)

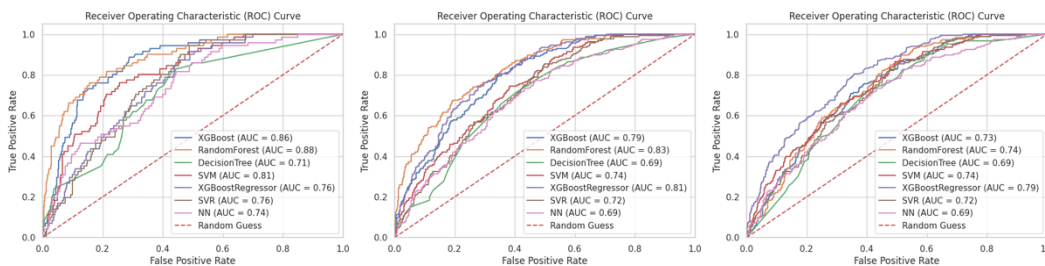
ตารางที่ 5 ผลลัพธ์การทดลองแบบจำลองการถดถอย ในกลุ่มต่างๆ

| Group Class | Model | MAE | MSE | RMSE |
|------------------------------|---------------------------|------|------|------|
| Multi-target | XGBoost Regressor | 0.67 | 0.70 | 0.84 |
| | Support Vector Regressor | 0.67 | 0.70 | 0.84 |
| | Regression Neural Network | 0.67 | 0.71 | 0.84 |
| Binary Class 1 (N, L, M : H) | XGBoost Regressor | 0.35 | 0.17 | 0.41 |
| | Support Vector Regressor | 0.38 | 0.19 | 0.44 |

| Group Class | Model | MAE | MSE | RMSE |
|------------------------------|---------------------------|------|------|------|
| | Regression Neural Network | 0.26 | 0.12 | 0.35 |
| Binary Class 2 (N, L : M, H) | XGBoost Regressor | 0.38 | 0.19 | 0.43 |
| | Support Vector Regressor | 0.42 | 0.21 | 0.46 |
| | Regression Neural Network | 0.36 | 0.19 | 0.43 |
| Binary Class 3 (N : L, M, H) | XGBoost Regressor | 0.40 | 0.19 | 0.44 |
| | Support Vector Regressor | 0.42 | 0.21 | 0.46 |
| | Regression Neural Network | 0.37 | 0.18 | 0.43 |

เพื่อเปรียบเทียบความแม่นยำในการประเมินความเสี่ยงการติดเชื้อเอชไอวีจากแบบจำลองการจำแนกประเภท และแบบจำลองการถดถอยผู้ทดลองจึงการประเมินประสิทธิภาพด้วยวิธี Receiver Operating Characteristic Curve (ROC Curve) จากภาพที่ 1 ในกลุ่ม Binary Class 1 (N, L, M : H) แบบจำลอง Random Forest และ XGBoost Classifier มีค่า AUC สูงสุดที่ 0.88 และ 0.86 ตามลำดับ แสดงถึงความสามารถในการจำแนกคลาสได้อย่างดี ในกลุ่ม Binary Class 2 (N, L : M, H) Random Forest มีค่า AUC สูงสุดที่ 0.83 และในกลุ่ม Binary Class 3 (N : L, M, H) XGBoost Regressor มีค่า AUC สูงสุดที่ 0.79

ภาพที่ 1 แสดงการประเมินประสิทธิภาพแบบจำลองด้วยวิธี Receiver Operating Characteristic Curve (ROC Curve) ในกลุ่ม Binary Class 1 (N, L, M : H), Binary Class 2 (N, L : M, H) และ Binary Class 3 (N : L, M, H) ตามลำดับ



สรุปผลการวิจัย

งานวิจัยนี้เป็นการศึกษาการประเมินความเสี่ยงการติดเชื้อเอชไอวีเพื่อให้ผู้ที่มีความเสี่ยงตระหนักรู้และไม่เกิดความประมาทในการแพร่เชื้อต่อผู้อื่น โดยศึกษากับแบบจำลองการจำแนกประเภท และแบบจำลองการถดถอย ผลการทดลองพบว่า Random Forest และ XGBoost Classifier มักจะให้ผลลัพธ์ที่ดีที่สุดในการจำแนกคลาสในทุกๆ แบบ ในขณะที่แบบจำลองอื่นๆ อาจมีประสิทธิภาพที่แตกต่างกันไปในแต่ละแบบการทดลอง ซึ่งแบบจำลองที่มีค่า AUC สูงสุดที่ 0.88 ในแบบจำลอง Random

Forest ในกลุ่ม Binary Class 2 (N, L : M, H) แต่เนื่องจากงานวิจัยนี้ต้องการให้แบบจำลองทำนายเน้นคลาสความเสี่ยงสูง ซึ่งในกลุ่มที่สามารถเน้นคลาสความเสี่ยงสูงได้ดีคือกลุ่ม Binary Class 3 (N : L, M, H) และแบบจำลองที่มีค่า AUC สูงสุดคือ XGBoost Regressor ที่ 0.79 ซึ่งในงานวิจัยในอนาคต ผู้วิจัยจะศึกษาเทคนิควิธีการเรียนรู้ของเครื่อง ประเภทการจัดกลุ่มของข้อมูลแบบอื่นๆ ที่สามารถนำมาสร้างตัวแบบเพื่อใช้ในการประเมินความเสี่ยงการติดเชื้อเอชไอวีให้มีประสิทธิภาพมากยิ่งขึ้น

กิตติกรรมประกาศ

การจัดทำวิจัยได้รับการสนับสนุนจากบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ ในการนำเสนอผลงานวิจัย ผู้วิจัยจึงขอขอบคุณมา ณ ที่นี้

เอกสารอ้างอิง

- [1] ความรู้เกี่ยวกับเอชไอวีและเอดส์ [Internet]. [cited 2023 Nov 16]. Available from: <https://www.bumrungrad.com/th/health-blog/november-2016/hiv-aids-infection-treatment>
- [2] Frescura L, Godfrey-Faussett P, Feizzadeh A A, El-Sadr W, Syarif O, Ghys PD, et al. Achieving the 95 95 95 targets for all: A pathway to ending AIDS. PloS One. 2022;17(8):e0272405.
- [3] กรมควบคุมโรค. สถานการณ์เอชไอวีประเทศไทย ปี 2565 [Internet]. 2022. Available from:https://hivhub.ddc.moph.go.th/Download/RRTRR/HIV_Factsheet2022_TH_Final_220966.pdf