แบบจำลองการประเมินความเสี่ยงที่เกี่ยวข้องกับการติดเชื้อเอชไอวีด้วยการเรียนรู้ด้วยเครื่อง

MACHINE LEARNING TECHNIQUES FOR ASSESSING RISK BEHAVIOUR ASSOCIATED WITH CONTRACTING HIV

แพรพรรณ พุ่มโพธิ์สุวรรณ

บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ

2566

แบบจำลองการประเมินความเสี่ยงที่เกี่ยวข้องกับการติดเชื้อเอชไอวีด้วยการเรียนรู้ด้วยเครื่อง

แพรพรรณ พุ่มโพธิ์สุวรรณ

สารนิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
วิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการข้อมูล
คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ
ปีการศึกษา 2566
ลิขสิทธิ์ของมหาวิทยาลัยศรีนครินทรวิโรฒ

MACHINE LEARNING TECHNIQUES FOR ASSESSING RISK BEHAVIOUR ASSOCIATED WITH CONTRACTING HIV

PEARPARN PUMPHOSUWAN

A Master's Project Submitted in Partial Fulfillment of the Requirements

for the Degree of MASTER OF SCIENCE

(Data Science)

Faculty of Science, Srinakharinwirot University

2023

Copyright of Srinakharinwirot University

สารนิพนธ์

เรื่อง

แบบจำลองการประเมินความเสี่ยงที่เกี่ยวข้องกับการติดเชื้อเอชไอวีด้วยการเรียนรู้ด้วยเครื่อง

ของ

แพรพรรณ พุ่มโพธิ์สุวรรณ

ได้รับอนุมัติจากบัณฑิตวิทยาลัยให้นับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการข้อมูล

ของมหาวิทยาลัยศรีนครินทรวิโรฒ

(รองศาสตราจารย์ นายแพทย์ฉัตรชัย เอกปัญญาสกุล)

คณบดีบัณฑิตวิทยาลัย

คณะกรรมการสอบปากเปล่าสารนิพนธ์

..... ที่ปรึกษาหลัก

(ผู้ช่วยศาสตราจารย์ ดร.นภา แซ่เบ๊)

..... ประธาน

(อาจารย์ ดร.นิดา ชาติวัฒนศิริ)

..... กรรมการ

(ผู้ช่วยศาสตราจารย์ ดร.ศิริสรรพ เหล่าหะเกียรติ)

ชื่อเรื่อง	แบบจำลองการประเมินความเสี่ยงที่เกี่ยวข้องกับการติดเชื้อเอชไอวีด้วยการเรียนรู้ด้วยเครื่อง
ผู้วิจัย	แพรพรรณ พุ่มโพธิ์สุวรรณ
ปริญญา	วิทยาศาสตรมหาบัณฑิต
ปีการศึกษา	2566
อาจารย์ที่ปรึกษา	ผู้ช่วยศาสตราจารย์ ดร. นภา แซ่เบ๊

การศึกษาวิจัยนี้มีวัตถุประสงค์เพื่อสร้างแบบจำลองการทำนายระดับความเสี่ยงที่เกี่ยวข้องกับการติดเชื้อเอชไอวีโดยอาศัยกลุ่มตัวแปร 4 กลุ่ม คือ ข้อมูลทางด้านประชากร (Demographics), รูปแบบการดำเนินชีวิต (Lifestyles), พฤติกรรมทางเพศ (Sexual Behaviors) และ อาการโรคติดต่ออื่นๆ (Symptoms) โดยใช้เทคนิคการเรียนรู้ของเครื่อง 7 แบบ คือ Decision tree, Random Forest, XGBoost Classifier, Support Vector Machine (SVM), XGBoost Regressor, Support Vector Regressor (SVR) และ Regression Neural Network โดยใช้ชุดข้อมูลจากสถาบันเพื่อการวิจัยและนวัตกรรมด้านเอชไอวี ในช่วงวันที่ 1 มกราคม 2564 จนถึง วันที่ 30 ธันวาคม 2565 จำนวนข้อมูลทั้งหมด 3,621 รายการ ซึ่งเป็นข้อมูลของผู้ที่มารับบริการเกี่ยวกับด้านการตรวจเอชไอวี โดยพิจารณาเฉพาะรายการผลการตรวจเอชไอวีที่เป็นลบเท่านั้น ในแต่ละรายการ ผู้ประเมินของสถาบันจะประเมินความเสี่ยงต่อการติดเชื้อเอชไอวีเป็น 4 ระดับ คือ ไม่มีความเสี่ยง (No risk: N) ความเสี่ยงต่ำ (Low risk: L) ความเสี่ยงปานกลาง (Moderate risk: M) และความเสี่ยงสูง (High risk: H) ในงานวิจัยนี้จะมีการแบ่งกลุ่มการทดลองตามรูปแบบผลการทำนายระดับความเสี่ยงออกเป็น 4 กลุ่ม ดังนี้ Multi-Class (N : L : M : H), Binary Class A (N, L, M : H), Binary Class B (N, L : M, H), Binary Class C (N : L, M, H) ผลการศึกษาพบว่า การทำนายระดับความเสี่ยงออกเป็น 2 กลุ่ม แบบ C (N : L, M, H) สามารถให้ประสิทธิภาพการทำนายได้ดีที่สุด ซึ่งมีค่าพื้นที่ใต้เส้นโค้งรับสมรรถนะ (AUC) 0.88 โดยอาศัยแบบจำลอง XGBoost Regressor การศึกษาวิจัยนี้แสดงให้เห็นถึงศักยภาพของแบบจำลองการเรียนรู้ของเครื่องในการช่วยประเมินระดับความเสี่ยงต่อการติดเชื้อเอชไอวี และมีแนวโน้มที่จะสามารถนำมาประยุกต์ใช้ในการคัดกรองความเสี่ยงได้อย่างมีประสิทธิภาพ

คำสำคัญ : การประเมินความเสี่ยงเอชไอวี, การเรียนรู้ด้วยเครื่อง

Title	MACHINE LEARNING TECHNIQUES FOR ASSESSING RISK BEHAVIOUR ASSOCIATED WITH CONTRACTING HIV
Author	PEARPARN PUMPHOSUWAN
Degree	MASTER OF SCIENCE
Academic Year	2023
Thesis Advisor	Assistant Professor Dr. Napa Sae-bae

The aim of this study 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: Decision tree, Random Forest, XGBoost Classifier, Support Vector Machine (SVM), XGBoost Regressor, Support Vector Regressor (SVR), and Regression Neural Network. The dataset was obtained from the research and innovation institution in the field of HIV, spanning from January 1, 2021, to December 30, 2021, comprising a total of 3,621 entries. The data pertains to individuals receiving HIV-related services, where only entries with negative HIV test results were considered. Manual risk assessment for HIV infection by the institution's evaluator was categorized into four levels: No risk (N), Low risk (L), Moderate risk (M), and High risk (H). The experiments were divided into four tasks: 1) Four-Class (N : L : M : H), Binary Class A (N, L, M : H), Binary Class B (N, L : M, H), and Binary Class C (N : L, M, H). The results indicated that the best performance was observed in Binary Class C (No vs. Low/Moderate/High) by the XGBoost Regressor model, achieving an Area Under the Curve (AUC) value of 0.88. This study demonstrates the potential of machine learning models to assist in assessing the risk levels associated with HIV infection, and it shows promise for effective application in risk screening.

Keyword : HIV Risk Assessment, Machine Learning

กิตติกรรมประกาศ

การจัดทำวิจัยได้รับการสนับสนุนข้อมูลจากสถาบันเพื่อการวิจัยและนวัตกรรมด้านเอชไอวี และได้รับการสนับสนุนจากบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ ในการนำเสนอผลงานวิจัย ผู้วิจัยจึงขอขอบคุณมา ณ ที่นี้

แพรพรรณ พุ่มโพธิ์สุวรรณ

สารบัญ

หน้า

สารบัญตาราง

หน้า

สารบัญรูปภาพ

หน้า

บทที่ 1
บทนำ

1.1 ที่มาและความสำคัญของงานวิจัย

เอชไอวีและเอดส์เป็นหนึ่งในสาเหตุหลักของการเสียชีวิตทั่วโลก และยังเป็นโรคที่คุกคามมนุษยชาติที่ร้ายแรงที่สุดโรคหนึ่ง ในปัจจุบันนี้องค์การอนามัยโลกได้เปิดเผยข้อมูลว่าในปี 2015 มีประชากรติดเชื้อเอชไอวีหรือเอดส์จำนวน 36.7 ล้านคนทั่วโลก (กีระสุนทรพงษ์, 2016) ปัจจุบันโลกกำลังเริ่มต้นใน Fast-Track เพื่อยุติการแพร่ระบาดของโรคเอดส์ภายในปี 2030 ซึ่งเป้าหมาย คือ 95-95-95 สำหรับ การทดสอบเชื้อเอชไอวี การรักษา และการลดระดับจำนวนเชื้อไวรัสสำเร็จ (Frescura et al., 2022) ในประเทศไทย กรมควบคุมโลกได้เปิดเผยข้อมูลเกี่ยวกับ สถานการณ์เอชไอวีปี 2022 ว่าจำนวนผู้ติดเชื้อเอชไอวีรายใหม่ 9,230 คน/ปี แบ่งเป็น เด็ก (<15 ปี) 54 คน เยาวชน(15-24 ปี) 4,379 คน ผู้ใหญ่ (>24 ปี) 4,797 คน และการติดเชื้อรายใหม่ที่ติดเชื้อจากการฉีดสารเสพติดที่ไม่ปลอดภัย 4% และ 96% เกิดจากการมีเพศสัมพันธ์ที่ไม่ป้องกัน ซึ่งแบ่งเป็นเพศสัมพันธ์ระหว่างชายกับชาย 68% คู่ผลเลือดต่าง (คู่อยู่กิน/คู่ประจำ) 19% คู่นอนชั่วคราวและนอกสมรส 8% เพศสัมพันธ์จากการซื้อขายบริการ 1% ซึ่งผู้ติดเชื้อจากเอดส์ในปี 2022 ทั้งหมด 10,972คน/ปี แบ่งเป็น เด็ก (<15 ปี) 81 คน เยาวชน (15-24 ปี) 237 คน ผู้ใหญ่ (>24 ปี) 10,654 คน และในด้านการดูแลรักษาตามเป้าหมาย 95-95-95 ในประเทศไทย 90%ของผู้ติดเชื้อเอชไอวีที่รู้สถานะว่าตนเองติดเชื้อฯ 90% ของผู้ติดเชื้อฯที่รู้สถานะ กำลังรับยาต้านไวรัส 97%ของผู้ที่รับยาต้านไวรัสสำเร็จ จะสังเกตได้ว่ายังมีผู้ติดเชื้อเอชไอวีที่ไม่ทราบสถานะว่าตนติดเชื้อเอชไอวี (กรมควบคุมโรค, 2022)

1.2 วัตถุประสงค์ของงานวิจัย

1. เพื่อสร้างแบบจำลองการทำนายระดับความเสี่ยงที่เกี่ยวข้องกับการติดเชื้อเอชไอวีโดยอาศัยกลุ่มตัวแปร 4 กลุ่ม คือ ข้อมูลทางด้านประชากร (Demographics), รูปแบบการดำเนินชีวิต (Lifestyles), พฤติกรรมทางเพศ (Sexual Behaviors) และ อาการอื่นๆ (Symptoms) โดยอาศัยแบบจำลอง 7 แบบ คือ Decision Tree, Random Forest, XGBoost Classifier, Support Vector Machine (SVM), XGBoost Regressor, Support Vector Regressor (SVR) และ Regression Neural Network

2. เพื่อทดสอบและเปรียบเทียบประสิทธิภาพของแบบจำลองรูปแบบต่างๆ

3. เพื่อศึกษาเปรียบเทียบความสำคัญของกลุ่มตัวแปรแต่ละกลุ่ม ต่อความแม่นยำในการทำนายระดับความเสี่ยงการติดเชื้อเอชไอวี

1.3 เป้าหมายและขอบเขตการวิจัย

งานวิจัยนี้รวบรวมข้อมูลจาก สถาบันเพื่อการวิจัยและนวัตกรรมด้านเอชไอวี ในช่วงวันที่ 1 มกราคม 2564 จนถึง วันที่ 30 ธันวาคม 2565 ซึ่งเป็นข้อมูลของผู้ที่มารับบริการเกี่ยวกับด้านการตรวจเอชไอวี การรับยา เพร็พ (PrEP) หรือ เป็ป (PEP) เพื่อป้องกันการติดเชื้อเอชไอวี และ การรักษาโรคติดต่อทางเพศสัมพันธุ์ ซึ่งจะต้องมีผลการตรวจเอชไอวี เป็น ลบ(Negative) เท่านั้น เพื่อประเมินความเสี่ยงการติดเชื้อเอชไอวี โดยแบ่งออกเป็น 4 ระดับ คือ ไม่มีความเสี่ยง (No Risk (N)), ความเสี่ยงต่ำ (Low Risk (L)), ความเสี่ยงปานกลาง (Medium Risk (M)) และ ความเสี่ยงสูง (High Risk (H))

ตัวแปรที่ศึกษา

1. ตัวแปรอิสระ แบ่งเป็นดังนี้

1.1 ข้อมูลทางด้านประชากร (Demographics): อายุ, เพศ, อัตลักษณ์ทางเพศ, การศึกษา, อาชีพ

1.2 รูปแบบการดำเนินชีวิต (Lifestyles): การใช้สารเสพติด, เหตุผลที่มาตรวจเอชไอวี, การรับยาเพร็พ (PrEP), การรับยาเป็ป (PEP)

1.3 พฤติกรรมทางเพศ (Sexual Behaviors): เคยมีเพศสัมพันธ์แบบสอดใส่หรือไม่, ในช่วงสามเดือนที่ผ่านมาใช้ถุงยางอนามัยระหว่างการมีเพศสัมพันธ์หรือไม่, ในช่วงสามเดือนที่ผ่านมาใช้ถุงยางบ่อยแค่ไหน, ถุงยางแตก / ถุงยางรั่วหรือไม่, ถุงยางหลุดในระหว่างการมีเพศสัมพันธ์หรือไม่, คู่นอนถอดถุงยางอนามัยขณะมีเพศสัมพันธ์หรือไม่, มีการสอดใส่ แต่ไม่เคยเกิดเหตุการณ์ใน 3 ข้อ (ถุงยางแตก, ถุงยางหลุด, คู่นอนถอดถุงยาง) ที่กล่าวมาหรือไม่, ในช่วงสามเดือนที่ผ่านมามีเพศสัมพันธ์เพื่อแลกกับเงินหรือสิ่งของตอบแทนหรือไม่

1.4 อาการอื่นๆ (Symptoms): การรักษาเอชไอวี, มีอาการของโรคหนองในหนองในเทียม หรือซิฟิลิส หรือไม่, ผลวินิจฉัยโรค syphilis, ผลวินิจฉัยโรคหนองใน หนองในเทียม

1.5 เจ้าหน้าที่ (Staff): หมอ, พยาบาล และ ผู้ให้คำปรึกษา

2. ตัวแปรตาม ได้แก่ ระดับความเสี่ยงที่จะติดเชื้อเอชไอวี มี 4 ระดับ ได้แก่ ไม่มีความเสี่ยง มีความเสี่ยงน้อย มีความเสี่ยงปานกลาง และ มีความเสี่ยงมาก

1.4 ขั้นตอนในการทำงานวิจัย

1. ศึกษาปัจจัยที่ทำให้เกิดความเสี่ยงการติดเชื้อเอชไอวี
2. ศึกษาการสร้างแบบจำลองการเรียนรู้ด้วยเครื่อง
3. รวบรวมข้อมูลจากสถาบันเพื่อการวิจัยและนวัตกรรมด้านเอชไอวี
4. นำข้อมูลมาเตรียมให้อยู่ในรูปแบบที่พร้อมใช้งาน

5. นำข้อมูลไปฝึกสอนในแบบจำลองและปรับปรุงไฮเปอร์พารามิเตอร์

6. บันทึกผลและเปรียบเทียบแบบจำลองต่างๆ

7. เลือกแบบจำลองที่ได้ผลที่ดีที่สุดเพื่อนำมาประเมินความเสี่ยงการติดเชื้อเอชไอวี

8. วิเคราะห์ผลการทดลอง

9. ประมาณค่าความสำคัญของตัวแปร

10. สรุปผลการทดลอง

1.5 ประโยชน์ของการทำงานวิจัย

1. นำแบบจำลองไปประเมินผู้ที่สงสัยว่ามีความเสี่ยงที่จะติดเชื้อเอชไอวี เพื่อเป็นการประเมินเบื้องต้นสำหรับผู้ที่คิดว่ามีความเสี่ยงที่จะติดเชื้อเอชไอวี และพิจารณาการรับยา PrEP หรือ PEP สำหรับผู้ที่มีความเสี่ยงสูง เพื่อเป็นการป้องกันการติดเชื้อเอชไอวี

2. สามารถพิจารณาตัวแปรสำคัญที่จะทำให้ติดเชื้อเอชไอวี และหลีกเลี่ยงปัจจัยนั้นๆ

บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 การติดเชื้อเอชไอวี และการป้องกัน

2.1.1 โรคเอดส์ และ การติดเชื้อไวรัสเอชไอวี

โรคเอดส์ (Immune Deficiency Syndrome: AIDS) หรือภาวะภูมิคุ้มกันบกพร่อง คือ ระยะสุดท้ายของการติดเชื้อไวรัสเอชไอวี (Human Immunodeficiency Virus: HIV)ที่ทำให้เซลล์เม็ดเลือดขาว CD4 หรือ T-cells ในระบบภูมิคุ้มกันร่างกายถูกทำลาย ส่งผลให้ภูมิคุ้มกันร่างกายลดต่ำ ทำให้ร่างกายอ่อนแอ และนำไปสู่การเกิดโรคแทรกซ้อนหรือโรคติดเชื้อฉวยโอกาส (มอร์ด, 2023)

ระยะการติดเชื้อ HIV มี 3 ระยะ ได้แก่

ระยะที่ 1 ระยะแรกเริ่มของการติดเชื้อเอชไอวี (Primary Infections: Acute HIV) เป็นระยะแรกของการติดเชื้อเอชไอวี ในช่วง 2-4 สัปดาห์แรกหลังการได้รับเชื้อ ในระยะนี้ เป็นระยะที่ผู้ติดเชื้อสามารถแพร่เชื้อสู่ผู้อื่นได้ จึงจำเป็นต้องป้องกันการแพร่เชื้ออย่างเคร่งครัด อย่างไรก็ตาม ผู้ติดเชื้อบางรายอาจไม่แสดงอาการใด ๆ เลย จึงทำให้สูญเสียโอกาสในการพบแพทย์เพื่อรับการตรวจวินิจฉัยและรับยาต้านไวรัส ตั้งแต่ระยะแรกเริ่ม จนนำไปสู่การแพร่เชื้อสู่ผู้อื่นในเวลาต่อมา

ระยะที่ 2 ระยะติดเชื้อโดยไม่มีอาการ (Clinical Latent Infection: Chronic HIV) เป็นการติดเชื้อระยะแฝงที่เชื้อไวรัสเอชไอวี อยู่ในร่างกายโดยไม่แสดงอาการใด ๆ (Asymptomatic HIV infection) ผู้ที่ติดเชื้อส่วนใหญ่จะมีสภาพร่างกายเป็นปกติเหมือนบุคคลทั่วไป ในระยะนี้เชื้อเอชไอวี จะค่อย ๆ ทำลายเซลล์เม็ดเลือดขาว (CD4) ทำให้ภูมิคุ้มกันร่างกายลดต่ำ และเกิดการเจ็บป่วยง่ายขึ้น โดยทั่วไป การดำเนินโรคในระยะนี้จะใช้เวลาประมาณ 5-10 ปี ผู้ติดเชื้อที่มีการดำเนินโรคเร็ว (Rapid Progressor) อาจใช้เวลาในระยะนี้เพียง 2-5 ปี แต่ในผู้ติดเชื้อที่ร่างกายสามารถควบคุมเชื้อได้ดีเป็นพิเศษ (Elite Controller) อาจยืดระยะเวลาการดำเนินโรคในระยะนี้ได้ 10-15 ปี

ระยะที่ 3 ระยะเอดส์เต็มขั้นหรือระยะโรคเอดส์ (Progression to AIDS) เป็นระยะที่การติดเชื้อเอชไอวี ได้พัฒนากลายเป็นโรคเอดส์โดยสมบูรณ์ ผู้ติดเชื้อในระยะนี้จะมีระดับ CD4 ในร่างกายน้อยกว่า 200 เซลล์ต่อลูกบาศก์มิลลิเมตร ทำให้ร่างกายอ่อนแอลงอย่างมากจนนำไปสู่การการเกิดโรคแทรกซ้อน และโรคติดเชื้อฉวยโอกาส (Opportunistic Infection: OIs)

2.1.2 ยาป้องกันการติดเชื้อเอชไอวี

การป้องกันการติดเชื้อก่อนการสัมผัส หรือ ยาเพร็พ (Pre-ExPosure Prophylaxis: PrEP) คือ ยาต้านไวรัสเอชไอวี ที่กินก่อนการสัมผัสเชื้อ เพื่อลดโอกาสในการติดเชื้อ มีประสิทธิภาพในการป้องกันสูงสุดถึงร้อยละ 99 หากกินอย่างสม่ำเสมอ ได้รับการรับรองจากองค์การอาหารและยาของสหรัฐอเมริกา หรือ FDA ตั้งแต่ปี พ.ศ. 2555 และในเดือนกันยายน ปี พ.ศ. 2558 ยาเพร็พ ได้ถูกบรรจุเป็นส่วนหนึ่งในชุดบริการเพื่อป้องกันการติดเชื้อเอชไอวีโดยองค์การอนามัยโลก (WHO) (ยงค์เจริญชัย, 2020)

การป้องกันการติดเชื้อหลังการสัมผัส หรือ ยาเป็ป (Post-ExPosure Prophylaxis: PEP) แบ่งเป็น 2 ชนิด คือ

1. การป้องกันการติดเชื้อเอชไอวีในบุคลากรทางการแพทย์หลังการสัมผัสจากการทำงาน หรือ HIV Occupational PEP (HIV OPEP) สำหรับบุคลากรทางการแพทย์ซึ่งสัมผัสเลือดและสารคัดหลั่งต่างๆจากการทำงานผ่านทางผิวหนังเช่น ถูกเข็มตำ ผ่านทางเยื่อบุเช่น กระเซ็นเข้าตา ปากหรือผ่าน ผิวหนังที่ไม่ปกติเช่น มีบาดแผลรอยแตก มีผื่น เป็นต้น

2. การป้องกันการติดเชื้อเอชไอวีหลังการสัมผัสที่ไม่ใช่จากการทำงาน หรือ HIV Non-Occupational PEP (nPEP) สำหรับการสัมผัสเลือดและสารคัดหลั่งที่เกิดจากการมีเพศสัมพันธ์การใช้เข็มฉีดยาร่วมกัน การถูกเข็มตำนอกสถานพยาบาล การถูกล่วงละเมิดทางเพศ และการได้รับบาดเจ็บ ซึ่งทำให้ผู้สัมผัสมีความเสี่ยงต่อการติดเชื้อเอชไอวี (Khongsra, 2019)

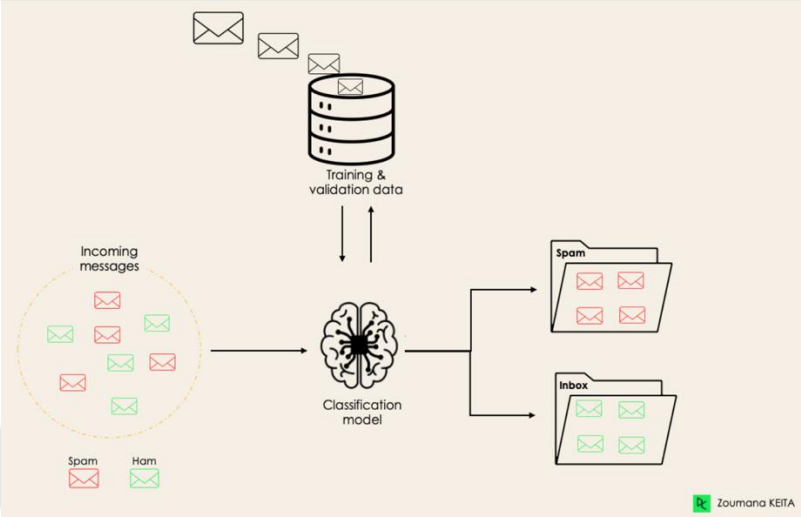
2.2 ทฤษฎีการวิเคราะห์ข้อมูลด้วยหลักการเรียนรู้ด้วยเครื่องแบบมีผู้สอน (Supervised Learning)

การเรียนรู้ด้วยเครื่องแบบมีผู้สอน เป็นหนึ่งในวิธีการในการเรียนรู้ของเครื่อง ที่ใช้ในการสร้างแบบจำลองที่สามารถทำนายผลลัพธ์จากข้อมูลที่มีคำตอบ (Label) ที่ระบุไว้ล่วงหน้า ซึ่งการสร้างแบบจำลองในการเรียนรู้ด้วยเครื่องแบบมีผู้สอน จะเกิดขึ้นด้วยการให้แบบจำลองเรียนรู้ จากข้อมูลที่มีป้ายกำกับ โดยการปรับพารามิเตอร์ของแบบจำลองเพื่อให้มีความสอดคล้องกับข้อมูลที่ให้มา โดยประเมินความสอดคล้องนี้ผ่านกระบวนการทดสอบกับข้อมูลที่ไม่เคยเห็นมาก่อน (Test Data) เพื่อให้เกิดการทำนาย (Prediction) ที่มีประสิทธิภาพสูงสุด

การเลือกใช้การเรียนรู้ด้วยเครื่องแบบมีผู้สอน แบบไหนขึ้นอยู่กับลักษณะของปัญหาและลักษณะของข้อมูลที่มีอยู่ เช่น

1. Classification เมื่อต้องการจำแนกเป็นประเภทหรือคลาส ซึ่งจะสามารถอธิบายได้โดยการใช้ตัวอย่างง่าย ๆ เช่น การจำแนกประเภทของอีเมลเป็นสแปมหรือไม่สแปม โดยแบ่งข้อมูลเป็น ข้อมูลที่ใช้สำหรับเรียนรู้ของแบบจำลอง (Train Set Data) และ ข้อมูลที่ใช้สำหรับทดสอบ

แบบจำลอง (Test Set Data) เมื่อนำข้อมูลเข้า Classification Model จึงสามารถแยกแยะได้ว่าอีเมลเป็นสแปมหรือไม่สแปม (Keita, 2022) ดังภาพประกอบ 1



ภาพประกอบ 1 อธิบายการทำงานของแบบจำลองการจำแนกประเภท

2. Regression เมื่อต้องการทำนายค่าต่อเนื่อง ตัวอย่างเช่นการทำนายราคาของบ้านจากข้อมูลต่าง ๆ เช่น พื้นที่ของบ้าน, จำนวนห้องนอน, ตำแหน่งที่ตั้ง เป็นต้น ซึ่งราคาบ้านนั้นเป็นค่าที่มีความต่อเนื่อง (Continuous Value)

2.2.1 แบบจำลองการจำแนกประเภท (Classification Model)

1. Decision Tree เป็นวิธีการสร้างแบบจำลองที่ใช้ในการทำนายหรือจำแนกประเภทของข้อมูลตามเงื่อนไขที่ถูกกำหนดไว้ในรูปของโครงสร้างต้นไม้ตัดสินใจ โดยที่แต่ละโหนด (Nodes) ในต้นไม้จะแทนด้วยเงื่อนไข (Conditions) เพื่อทำนายผลลัพธ์ของข้อมูล ซึ่งอาจเป็นการจำแนกประเภทหรือการทำนายค่าต่อเนื่อง ต้นไม้จะแบ่งข้อมูลออกเป็นกลุ่มย่อยๆ โดยมีเงื่อนไขที่ต่างกันอย่างต่อเนื่อง จนกระทั่งมีลูกของต้นไม้ที่เรียกว่าใบ (Leaf) ซึ่งใบจะส่งผลลัพธ์ที่เป็นคำตอบหรือการทำนายที่ถูกต้องสำหรับข้อมูลนั้น

2. Random Forest เป็นการรวมกันของหลาย Decision Trees เพื่อทำนายผลลัพธ์หรือจำแนกประเภทของข้อมูล โดยที่แต่ละ Decision Tree จะถูกสร้างขึ้นโดยการสุ่มสร้างต้นไม้แยกต่างหาก และการตัดสินใจในแต่ละต้นไม้จะถูกใช้เพื่อทำนายผลลัพธ์

3. XGBoost (Extreme Gradient Boosting) เป็นอัลกอริทึมที่ใช้วิธีการเรียนรู้ที่เรียกว่า Gradient Boosting ซึ่ง เริ่มต้นด้วยการเรียนรู้ของแต่ละต้นไม้เรียกว่า Weak Learner โดยแบบจำลองจะพยายามทำนายผลลัพธ์ที่ยังไม่เป็นที่รู้จักได้ โดยการใช้ Gradient Descent เพื่อ

ปรับแบบจำลองให้มีความแม่นยำมากขึ้นในทุก ๆ รอบ หลังจากการสร้างต้นไม้แต่ละต้น แบบจำลองจะพยายามปรับแก้ความผิดพลาดของแบบจำลองก่อนหน้าด้วยการเพิ่ม Weak Learner ใหม่เข้าไป โดยแบบจำลองจะให้ความสำคัญมากขึ้นกับข้อมูลที่ทำนายผลไม่ถูกต้อง ซึ่งช่วยให้แบบจำลองมีความแม่นยำมากขึ้น

4. Support Vector Machine (SVM) เป็นวิธีการในการจำแนกประเภทของข้อมูลโดยใช้การสร้างเส้นแบ่ง (Hyperplane) เพื่อแบ่งข้อมูลให้อยู่ในกลุ่มต่าง ๆ โดยที่มีระยะห่างระหว่างข้อมูลที่ใกล้สุดกับเส้นแบ่งนั้นมากที่สุด (Margin) ซึ่ง Support Vector จะเป็นตัวอย่างข้อมูลที่อยู่บนขอบเขตของการแยกกลุ่ม โดยที่มีความหลากหลายหรือสำคัญที่สุดในการหาเส้นแบ่ง ซึ่ง SVM จะพยายามหาเส้นแบ่งที่ดีที่สุดเพื่อแบ่งข้อมูลออกเป็นกลุ่มต่าง ๆ เพื่อให้เส้นแบ่งแยกกลุ่มอย่างชัดเจนที่สุด และมีระยะห่างระหว่างข้อมูลกับเส้นแบ่งมากที่สุด และจะพยายามสร้างระยะห่าง ที่กว้างที่สุดระหว่างเส้นแบ่งกับข้อมูลที่อยู่บนขอบของกลุ่ม เพื่อป้องกันการ Overfitting และเพิ่มความแม่นยำของแบบจำลอง

2.2.2 แบบจำลองการถดถอย (Regression Model)

1. XGBoost Regressor เป็นอัลกอริทึมสำหรับปัญหาการทำนาย โดยเฉพาะในกรณีที่ต้องการทำนายค่าตัวเลข (Numeric Values) เช่น การทำนายราคาหุ้น, การทำนายยอดขาย, หรือการทำนายค่าอื่น ๆ ที่เป็นตัวเลข

2. Support Vector Regressor (SVR) เป็นเส้นแบ่งที่ใช้ใน Regression โดยหาก SVM ใช้ในการจำแนกประเภทของข้อมูล SVR จะใช้ในการทำนายค่าต่อเนื่องของข้อมูล คล้ายกับการปรับแบบพารามิเตอร์ในทิศทางของข้อมูลที่ใกล้เคียงกันในพื้นที่ของข้อมูลหรือเส้นแบ่ง แต่จะใช้ในกรณีที่ตัวแปรตามไม่ได้มีค่าแบบหมายเลขและต้องการทำนายค่าต่อเนื่อง เช่น การทำนายราคาของอสังหาริมทรัพย์จากตัวแปรต่าง ๆ เช่น พื้นที่, จำนวนห้องนอน ฯลฯ

3. Regression Neural Network เป็นแบบจำลอง Neural Network ที่ใช้สำหรับการทำนายค่าต่อเนื่อง (Continuous Value) ซึ่งเหมาะสำหรับการแก้ปัญหา Regression ซึ่งมักจะใช้ในการทำนายค่าต่าง ๆ โดยโครงสร้างจะประกอบด้วยชั้นข้อมูล (Layers) ที่ต่อกัน โดยปกติจะมีชั้นดังนี้

- Input Layer: รับข้อมูลตัวแปรอิสระ เข้าสู่แบบจำลอง โดยจำนวนโหนดจะมีจำนวนเท่ากับจำนวนตัวแปรอิสระ

- Hidden Layers: ประกอบด้วยหลายๆ ชั้นที่มีโหนดต่อกัน และแต่ละโหนดจะเป็นเหมือนเซลล์ประมวลผลที่นำข้อมูลตัวแปรอิสระ มาทำการประมวลผลด้วยฟังก์ชัน Activation

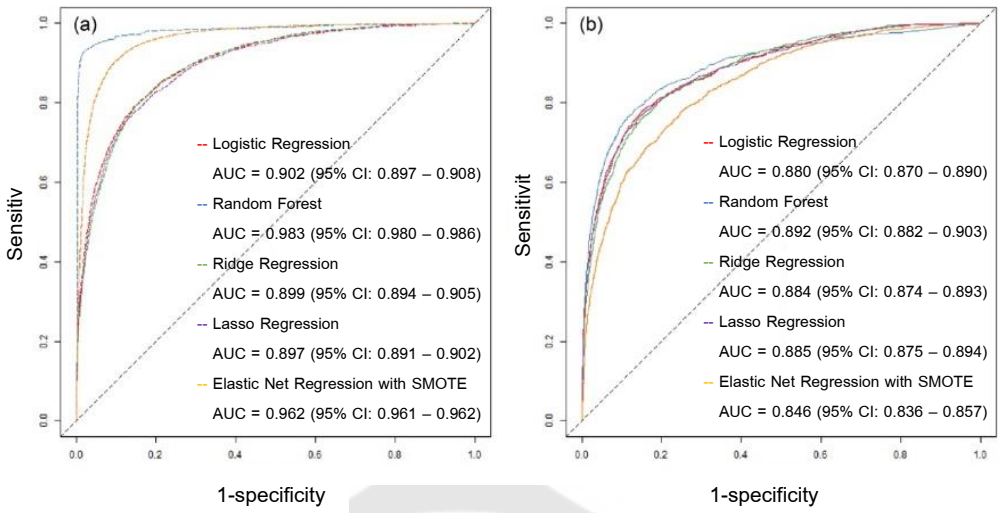
เพื่อสร้างความสัมพันธ์ระหว่างตัวแปรอิสระ และ ตัวแปรตาม โดยที่ลำดับ และจำนวนโหนดในแต่ละ Hidden Layer สามารถปรับได้ตามความซับซ้อนของปัญหา

- Output Layer: เป็นชั้นสุดท้ายที่จะสร้างค่าผลลัพธ์ที่เป็นค่าต่อเนื่อง ซึ่งจำนวนโหนดในชั้นนี้จะขึ้นอยู่กับประเภทของ Regression ที่ต้องการ ตัวอย่างเช่น หากต้องการทำนายค่าต่อเนื่องเพียงค่าเดียว ก็จะมีเพียงหนึ่งโหนด

2.3 งานวิจัยที่เกี่ยวข้องกับการติดเชื้อเอชไอวี

2.3.1 การจัดกลุ่มตัวแปร

จากงานวิจัย Algorithmic Prediction of HIV Status Using Nation-Wide Electronic Registry Data (Ahlström et al., 2019) ใช้ข้อมูลข้อมูลทะเบียนอิเล็กทรอนิกส์ทั่วประเทศเดนมาร์กนำมาทำนายสถานะเอชไอวีโดยใช้อัลกอริธึมการเรียนรู้ของเครื่องโดยใช้ตัวแปรน้อยที่สุดในการประเมิน โดยแบ่งกลุ่มแบบจำลอง Model I: สำหรับแบบจำลองที่รวมอายุ เพศ และโรคติดต่อทางเพศสัมพันธ์, Model II: ข้อมูลเกี่ยวกับประวัติทางการแพทย์, Mode III: รวมทุกตัวแปร ซึ่งใช้แบบจำลอง Random Forest, Logistic Regression, Ridge Regression , Lasso Regression และ Elastic Net Regression ปรับสมดุลชุดข้อมูลโดยเทคนิค Synthetic Minority Oversampling Technique (SMOTE) และใช้วิธี Area Under the Receiver Operating Characteristic Curve (AUROC) ในการวัดประสิทธิภาพแบบจำลอง ค้นพบว่าจำนวนตัวแปรที่มากขึ้นส่งผลให้ประสิทธิภาพในการทำนายเพิ่มขึ้น ในชุดการตรวจสอบรวมทุกตัวแปร (Mode III) มี AUC ที่ 0.899 (95% CI: 0·894–0·905) และ AUC ที่ 0.800 (95% CI: 0·793–0·807) สำหรับแบบจำลองที่รวมอายุ เพศ และโรคติดต่อทางเพศสัมพันธ์ (Model I) และ AUC ที่ 0·780 (95% CI: 0·769–0·792) สำหรับชุดการตรวจสอบ ที่รวมข้อมูลเพิ่มเติมเกี่ยวกับประวัติทางการแพทย์ (Model II) จะเห็นไว้ว่าในชุดการตรวจสอบรวมทุกตัวแปร (Mode III) ดีกว่าแบบจำลองอื่น ๆ ในทุกรายการวัดผล จึงนำ Mode III ไปใช้ทำนายในแบบจำลองต่างๆ ซึ่งแบบจำลองที่ดีที่สุดคือ Random Forest ที่ AUC 0.892



(a) ROC Curve สำหรับชุดข้อมูลการฝึก (b) ROC Curve ROC สำหรับชุดข้อมูลการตรวจสอบ

ภาพประกอบ 2 ROC Curve เปรียบเทียบแบบจำลองในแต่ละชุดข้อมูล

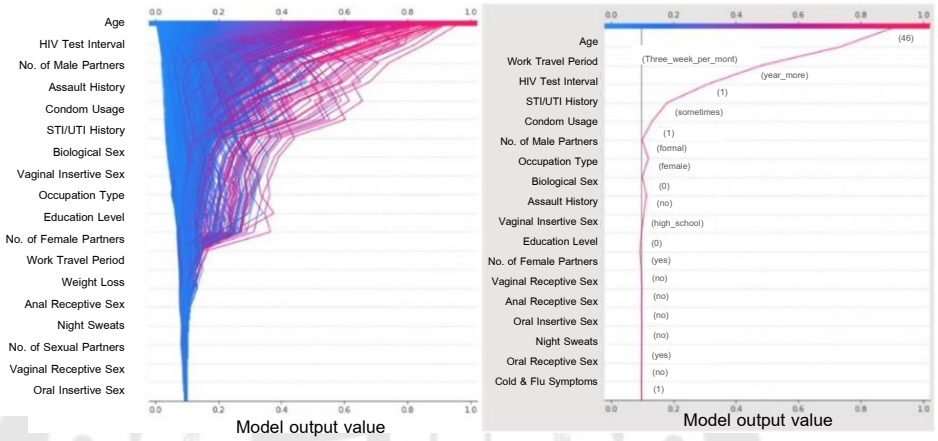
2.3.2 การเลือกแบบจำลองสำหรับข้อมูลที่มีความไม่สมดุล

จากงานวิจัย Prediction of HIV Infections Among Individuals with Sexual Risk Behaviours in Rwanda using Machine Learning Algorithms (Muhimpundu & Dr. Pierre Claver) เป็นการสร้างแบบจำลองในการทำนายการติดเชื้อเอชไอวี ในกลุ่มที่มีพฤติกรรมเสี่ยงทางเพศ งานวิจัยนี้ใช้ข้อมูลจาก RPHIA ปี 2018-2019 โดยมี ผู้ตอบแบบสำรวจ 30,709 คน ซึ่งเป็นผู้ติดเชื้อเอชไอวี จำนวน 934 (0.03%) และผู้ไม่ติดเชื้อเอชไอวี จำนวน 29,775 (99.97%) โดยใช้ 3 แบบจำลองคือ Logistic Regression, Gradient Boost และ Random Tree Forest พบว่า Random Tree Forest เป็นแบบจำลองที่ดีที่สุด ด้วย Accuracy 71.15%, Precision 61.2%, Recall 84.5% ซึ่งค่าที่ทำนายได้ 261 True Negatives, 163 False Positives, 47 False Negatives and 257 True Positives จากการวิเคราะห์พบว่า Random Tree Forest ช่วยลด False Negatives, เพิ่ม True Positives, Recall, และ F1-Score และพื้นที่ใต้เส้นโค้ง (AUC) คือ 0.75

2.3.3 ความสำคัญของตัวแปร (Feature Importances)

จากงานวิจัย Utility of a Machine-Guided Tool for Assessing Risk Behaviour Associated with Contracting HIV in Three Sites in South Africa (Majam et al., 2023) เป็นงานวิจัยที่ใช้ Machine Learning ประเมินผลของการติดเชื้อเอชไอวี ซึ่งข้อมูลมาจาก 3 ไซต์ใน

ประเทศแอฟริกาใต้ แบบจำลองที่ได้ผลลัพธ์ที่ดีที่สุดสำหรับงานวิจัยนี้คือ Gradient Boosted Tree Model และมีผล Sensitivity 84%, Specificity 71% งานวิจัยนี้สามารถทำนายผลของผู้ที่ไม่ติดเชื้อเอชไอวี(Negative) ได้ถึง 95% จากนั้นงานวิจัยนี้ได้ใช้วิธีการ SHAP (Shapley Additive Explanations) โดยทำหน้าที่แสดงความสำคัญของแต่ละตัวแปรที่มีผลต่อการทำนาย จากภาพประกอบ 3 จะสังเกตได้ว่า 5 ตัวแปรที่สำคัญต่อการสร้างแบบจำลองในงานวิจัยนี้คือ อายุ, การทดสอบ HIV ครั้งล่าสุด, จำนวนคู่เพศชาย, ประวัติการโดนทำร้าย และ ประวัติการใช้ถุงยาง



(a) กราฟ SHAP แสดงค่าความสำคัญของตัวแปรในทุกแบบจำลอง (b) กราฟ SHAP แสดงค่าความสำคัญของตัวแปรในแบบจำลอง Gradient Boosted Tree Model

ภาพประกอบ 3 เปรียบเทียบค่าความสำคัญของตัวแปรโดยวิธี SHAP

บทที่ 3
วิธีดำเนินการวิจัย

3.1 **การรวบรวมข้อมูล** (Data Acquisition)

เนื่องจากสถาบันเพื่อการวิจัยและนวัตกรรมด้านเอชไอวีมีการเก็บข้อมูลของผู้เข้ารับบริการที่เกี่ยวข้องกับด้านการติดเชื้อเอชไอวีต่าง ๆ จึงขอความอนุเคราะห์ด้านข้อมูลจากสถาบันดังกล่าว ชุดข้อมูลแยกตามแบบฟอร์มในการเก็บข้อมูล 3 แบบฟอร์ม คือ แบบสอบถามข้อมูลประชากร แบบสอบถามของผู้ให้คำปรึกษา และ แบบสอบถามเรื่องพฤติกรรมเสี่ยงต่อการติดเชื้อเอชไอวี ชุดข้อมูลมีการแยกนัดหมาย โดยที่เมื่อผู้เข้ารับบริการมาครั้งแรก จะกำหนดให้เป็น Baseline และเมื่อผู้เข้ารับบริการมาติดตามนัดครั้งถัดไป กำหนดให้เป็น Follow up ซึ่งจะเลือกเฉพาะข้อมูลผู้เข้ารับบริการที่มาครั้งแรก ที่มีผลตรวจเอชไอวีเป็นลบ และมาเข้ารับบริการ ด้านการตรวจเอชไอวีหรือ การรับยาป้องกันการติดเชื้อเอชไอวี(PrEP, PEP)หรือ การรักษาโรคติดต่อทางเพศสัมพันธุ์ และเลือกตัวแปรที่อาจมีส่วนเกี่ยวข้องที่จะทำให้เกิดความเสี่ยงการติดเชื้อเอชไอวี โดยการประเมินความเสี่ยงนี้ดำเนินการโดยเจ้าหน้าที่ที่มีความเชี่ยวชาญ และมีการแบ่งกลุ่มตัวแปรดังนี้ ข้อมูลทางด้านประชากร (Demographics), รูปแบบการดำเนินชีวิต (Lifestyles), พฤติกรรมทางเพศ (Sexual Behaviors) และ อาการอื่นๆ (Symptoms)

3.2 **การเตรียมข้อมูล** (Data Preprocessing)

3.2.1 จัดการข้อมูล (Data Management)

เนื่องจากชุดข้อมูลแยกตามแบบฟอร์มในการเก็บข้อมูล จึงทำการรวมข้อมูลในอยู่ในชีตเดียวโดยมี UID ซึ่งเป็นรหัสของผู้รับบริการ เป็นหลัก และเลือกข้อมูลเฉพาะนัดหมาย Baseline ของผู้เข้ารับบริการ จัดการกับข้อมูลที่ขาดหายไป (Missing Value) โดยการแปลงให้เป็นค่า ไม่ขอตอบ (Prefer Not to Answer (PNA)) เนื่องจาก สำหรับบางคำถาม การไม่สะดวกใจตอบไม่ได้หมายความว่าข้อมูลนั้นหายไป แต่เป็นการแสดงถึงความปรารถนาของบุคคลที่ไม่ต้องการระบุข้อมูลนั้นๆ อย่างชัดเจน ดังนั้นจึงเป็นการแสดงความไม่ต้องการให้ระบุข้อมูลนั้นในแบบฟอร์ม โดยไม่ใช่การบ่งชี้ว่าข้อมูลนั้นหายไป

3.2.2 การคัดเลือกคุณลักษณะ (Feature Selection)

ในการศึกษานี้จะทำการทดสอบและคัดเลือกคุณลักษณะด้วยวิธีการพ่วง (Wrapper approach) (Eakasit Pacharawongsakda, 2015) ซึ่งเป็นการคัดเลือกคุณลักษณะด้วยการสร้างแบบจำลองการจำแนกประเภท (Classification Model) จากกลุ่มคุณลักษณะที่กำหนดไว้ 4 กลุ่ม คือ ข้อมูลทางด้านประชากร , รูปแบบการดำเนินชีวิต , พฤติกรรมทางเพศ และ อาการอื่น ๆ ดังนี้

1. ข้อมูลทางด้านประชากร

ตาราง 1 คุณลักษณะกลุ่มข้อมูลทางด้านประชากร

ลำดับ	ตัวแปร	รายละเอียด
1	อายุ (Age)	เป็นค่าที่มีความต่อเนื่องที่อยู่ระหว่าง 15-73 ปี
2	เพศ (Sex)	1. ชาย (Male) 2. หญิง (Female) 3. เพศกำกวม (Intersex)
3	อัตลักษณ์ทางเพศ (Gender)	1. ชาย (Male) 2. หญิง (Female) 3. ชายผู้มีเพศสัมพันธ์กับชาย (MSM) 4. เกย์ (Gay) 5. หญิงข้ามเพศ (Transgender Women) 6. ชายข้ามเพศ (Transgender Men) 7. เลสเบียน (Lesbian) 8. ไบเซคชวล (Bisexual) 9. ไม่ใช่ชายหญิง (Non-Binary) 10. ไม่แน่ใจ(Not Sure) 11. ไม่ขอตอบ (PNA)
4	การศึกษา (Education)	1. น้อยกว่าประถมศึกษา (Lower than Primary School) 2. ประถมศึกษา (Primary School) 3. มัธยมศึกษาตอนต้น (Junior High School, Secondary School) 4. มัธยมศึกษาตอนปลาย/ปวช. (Senior High School, Secondary School) 5. ปวส./อนุปริญญา (Diploma/High Vocational Certificate) 6. ปริญญาตรี (Bachelor Degrees) 7. ปริญญาโทหรือสูงกว่า (Master Degree or Higher) 8. ไม่มีประวัติศึกษา (None) 9. ไม่ขอตอบ (PNA)

ตาราง 1 (ต่อ)

ลำดับ	ตัวแปร	รายละเอียด
5	อาชีพ (Occupation)	1. พนักงานบริษัทเอกชน (Company Employee) 2. ข้าราชการ/รัฐวิสาหกิจ (Government Employee) 3. เกษตรกร/ปศุสัตว์ (Farmer) 4. พนักงานโรงงานอุตสาหกรรม (Industrial Workers) 5. ธุรกิจส่วนตัว (Business Owner) 6. พนักงานบริการทางเพศ (Sex Workers) 7. ฟรีแลนส์ (Freelance) 8. ขายบริการ (Sex Worker) 9. ไม่ทำงาน (No Work) 10. ว่างงาน (Unemployed) 11. อื่นๆ (Others)

2) รูปแบบการดำเนินชีวิต

ตาราง 2 คุณลักษณะกลุ่มรูปแบบการดำเนินชีวิต

ลำดับ	ตัวแปร	รายละเอียด
1	การใช้สารเสพติด (Drug use)	1. ไม่ใช้สารเสพติดในช่วง 3 เดือนที่ผ่านมา (Drug-Free for the Past 3 Months) 2. ใช้สารเสพติดในช่วง 3 เดือนที่ผ่านมา (Using Drugs in the Last 3 Months) 3. ไม่เกี่ยวข้อง (Irrelevant)

ตาราง 2 (ต่อ)

ลำดับ	ตัวแปร	รายละเอียด
2	เหตุผลที่มาตรวจเอชไอวี (Reason HIV Test)	1. มาตรวจเอง เพราะ ทราบว่าคู่นอนมีผลเลือดเป็นบวก (Self Index Pos) 2. คู่ที่มีผลเลือดเป็นบวก แนะนำให้มาตรวจ (Partner Index Pos) 3. คู่ที่มีผลเลือดเป็นลบ แนะนำให้มาตรวจ (Partner) 4. เพื่อนกลุ่มเสี่ยงแนะนำให้มาตรวจเอชไอวี (Friends at Risk Together, Recommend Coming for an HIV Test) 5. ไปตรวจ HIV กับเพื่อน (Getting an HIV Test with a Friend) 6. มีคู่ที่รับประทาน PrEP (I Have a Partner Who Takes PrEP) 7. ฉันต้องการรับยา PrEP/PEP (I Would Like to Receive PrEP/PEP) 8. ฉันมีผื่นและอาการทางร่างกายต่างๆที่ทำให้เกิดความกังวล (I Have a Rash and Various Physical Symptoms That are Causing Concern) 9. ฉันมีคู่นอนที่เป็นโรคติดต่อ (I Have a Sexual Partner With a Contagious Disease) 10. ตรวจสุขภาพประจำปี (Annual Health Check-Up) 11. เชื่อว่ามีความเสี่ยงต่อการมีเพศสัมพันธ์ (Believe There is a Risk of Sexual exPosure) 12. มาตรวจตามนัด (Coming For a Scheduled Appointment) 13. เจ้าหน้าที่แนะนำให้ไปตรวจเอชไอวี (The Staff Recommends Getting an HIV Test) 14. ยืนยันการตรวจเอชไอวี (Confirm HIV Test)

ตาราง 2 (ต่อ)

ลำดับ	ตัวแปร	รายละเอียด
2	เหตุผลที่มาตรวจเอชไอวี (Reason HIV Test)	15. ลองตรวจเอชไอวี (Try getting an HIV Test) 16. เพื่อใช้สำหรับการรับรองในการจ้างงานหรือการศึกษาต่อ (To use for Certification in Employment or Further Studies) 17. อื่นๆ (Other) 18. ไม่ขอตอบ (PNA)
3	การรับยาเพร็พ (PrEP)	1. รับ (Y) 2. ไม่รับ (N)
4	การรับยาเป็ป (PEP)	1. รับ (Y) 2. ไม่รับ (N)

3) พฤติกรรมทางเพศ

ตาราง 3 คุณลักษณะกลุ่มพฤติกรรมทางเพศ

ลำดับ	ตัวแปร	รายละเอียด
1	เคยมีเพศสัมพันธ์แบบสอดใส่หรือไม่	1. ใช่ รวมถึงในช่วง 3 เดือนที่ผ่านมา (Yes, Including in the Past 3 Months) 2. ใช่ แต่ไม่ใช่ในช่วง 3 เดือนที่ผ่านมา (Yes, But not in the Past 3 Months) 3. ไม่เคยมีเพศสัมพันธ์ (Never) 4. ไม่ขอตอบ (PNA)
2	ในช่วงสามเดือนที่ผ่านมาใช้ถุงยางอนามัยระหว่างการมีเพศสัมพันธ์หรือไม่	1. ไม่เคยใช้ถุงยางอนามัยเลย (N) 2. ใช้ถุงยางอนามัย (Y) 3. ไม่เคยมีเพศสัมพันธ์ ใน 3 เดือน (NC)

ตาราง 3 (ต่อ)

ลำดับ	ตัวแปร	รายละเอียด
3	ในช่วงสามเดือนที่ผ่านมาใช้ถุงยางบ่อยแค่ไหน	1. บางครั้ง (Sometimes) 2. ส่วนใหญ่ (Often) 3. ทุกครั้ง (Always) 4. ไม่ขอตอบ (PNA)
4	ถุงยางแตก / ถุงยางรั่ว	1. ใช่ (Y) 2. ไม่ใช่ (N)
5	ถุงยางหลุดในระหว่างการมีเพศสัมพันธ์หรือไม่	1. ใช่ (Y) 2. ไม่ใช่ (N)
6	คู่นอนถอดถุงยางอนามัยขณะมีเพศสัมพันธ์หรือไม่	1. ใช่ (Y) 2. ไม่ใช่ (N)
7	มีการสอดใส่ แต่ไม่เคยเกิดเหตุการณ์ใน 3 ข้อ (ถุงยางแตก, ถุงยางหลุด, คู่นอนถอดถุงยาง) ที่กล่าวมาหรือไม่	1. ใช่ (Y) 2. ไม่ใช่ (N)
8	ในช่วงสามเดือนที่ผ่านมามีเพศสัมพันธ์เพื่อแลกกับเงินหรือสิ่งของตอบแทนหรือไม่	1. ไม่มีเพศสัมพันธ์เพื่อแลกกับเงินหรือสิ่งของตอบแทน (No) 2. มีและใช้ถุงยาง (Yes and used Condom) 3. มีและไม่ได้ใช้ถุงยาง (Yes and did not use Condom) 4. ไม่ขอตอบ (PNA)

4) อาการอื่น ๆ

ตาราง 4 คุณลักษณะกลุ่มอาการอื่นๆ

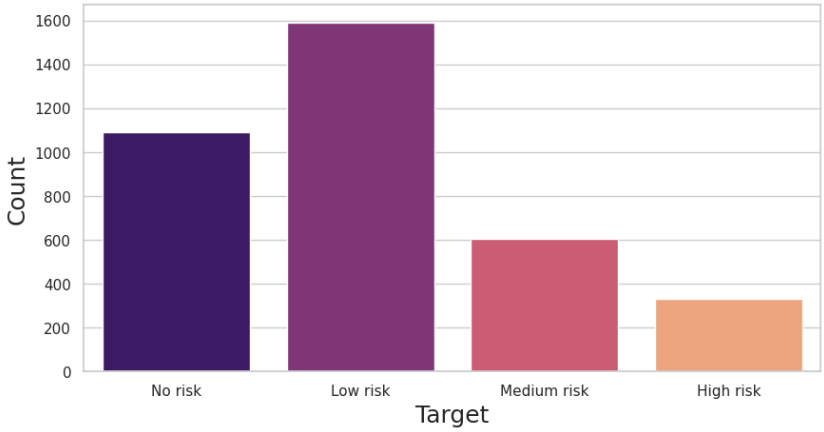
ลำดับ	ตัวแปร	รายละเอียด
1	การรักษาเอชไอวี	1. ใช่ (Y) 2. ไม่ใช่ (N) 3. ไม่เกี่ยวข้อง (Irrelevant)
2	มีอาการของโรคหนองใน หนองในเทียม หรือซิฟิลิส	1. ใช่ (Y) 2. ไม่ใช่ (N) 3. ไม่เกี่ยวข้อง (Irrelevant)
3	ผลวินิจฉัยโรค syphilis	1. ไม่เป็นโรคซิฟิลิส (Negative Result for Syphilis) 2. ซิฟิลิสระยะแรก (Early Syphilis) 3. ซิฟิลิสตอนปลาย (Late Syphilis) 4. กำลังรักษาโรคซิฟิลิส (Treated Syphilis) 5. ระหว่างการรักษาพร้อมติดตามผล (During Treatment with Follow Up) 6. ความล้มเหลวในการรักษา (Treatment Failure) 7. ซิฟิลิสติดเชื้อซ้ำ (Re-Infection Syphilis) 8. อื่นๆ (Other) 9. ไม่เกี่ยวข้อง (Irrelevant)
4	ผลวินิจฉัยโรคหนองใน หนองในเทียม	1. เป็นโรค (R) 2. ไม่เป็นโรค (NR) 3. ไม่ตรวจ (Irrelevant)

3.2.3 การสำรวจข้อมูลเบื้องต้น (Exploratory Data Analysis (EDA))

การสำรวจข้อมูลเบื้องต้นเกี่ยวกับการกระจายตัวของข้อมูลที่อยู่ในชุดข้อมูลดังกล่าวและความสัมพันธ์ระหว่างข้อมูล ดังนี้

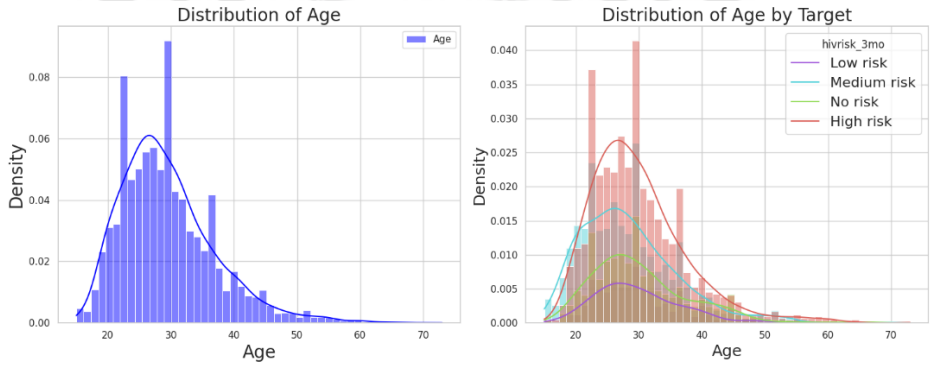
3.2.3.1 ข้อมูลระดับความเสี่ยงซึ่งกำหนดเป็นผลลัพธ์ของแบบจำลองที่จะสร้างขึ้นมีการกระจายตัวดังภาพประกอบ 4 ซึ่งสามารถสังเกตได้ว่าข้อมูลของ ความเสี่ยงต่ำ (Low Risk) มีจำนวนมากที่สุดจำนวน 1591 แถว รองลงมาคือ ไม่มีความเสี่ยง (No Risk) จำนวน 1091

แถว, ความเสี่ยงปานกลาง (Medium Risk) จำนวน 606 แถว, ความเสี่ยงมาก (High Risk) จำนวน 333 แถว ตามลำดับ จะสังเกตได้ว่าข้อมูลตัวแปรตามเป็นข้อมูลที่มีความไม่สมดุล (Imbalanced Data)

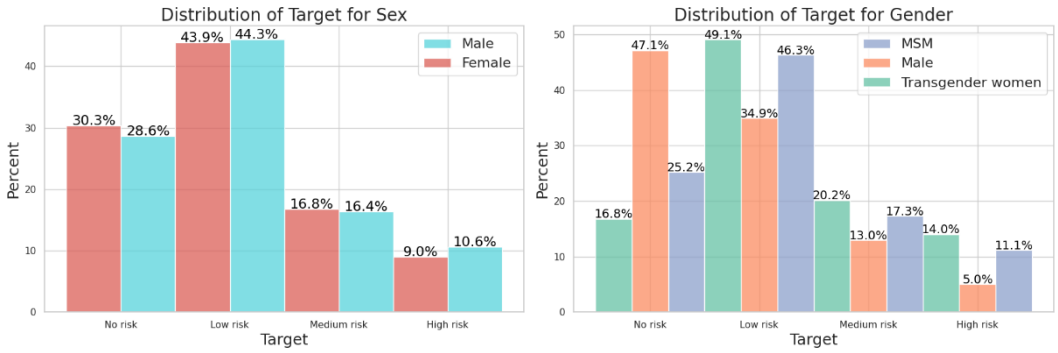


ภาพประกอบ 4 กราฟแท่งแสดงกระจายตัวของระดับความเสี่ยงเอชไอวี

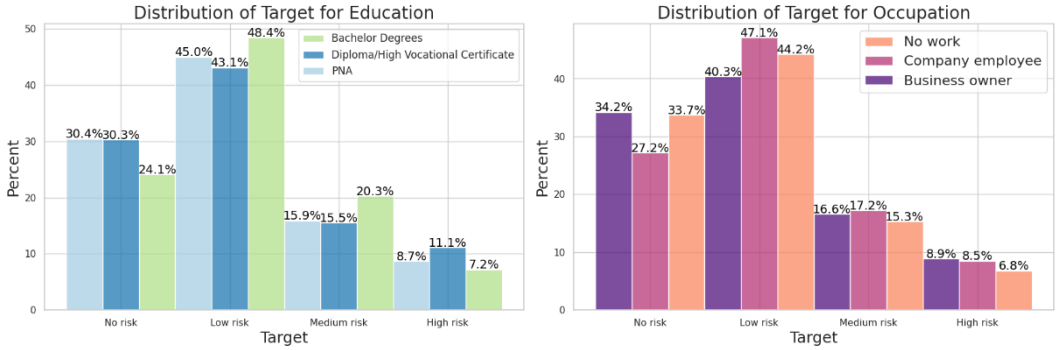
3.2.3.2 การกระจายตัวของข้อมูลสำหรับกลุ่มข้อมูลทางด้านประชากร (Demographics) และการกระจายตัวของข้อมูลดังกล่าวสำหรับแต่ละระดับความเสี่ยง แสดงดังภาพประกอบ 5(d) ซึ่งจะสังเกตได้ว่า ในตัวแปรอัตลักษณ์ทางเพศ หญิงข้ามเพศ (Transgender Women) แสดงผลกระจายตัวของข้อมูลที่อาจทำให้มีผลต่อแบบจำลองในคลาสไม่มีความเสี่ยง



5(a) แสดงการกระจายตัวของอายุ 5(b) แสดงการกระจายตัวของอายุกับความเสี่ยง



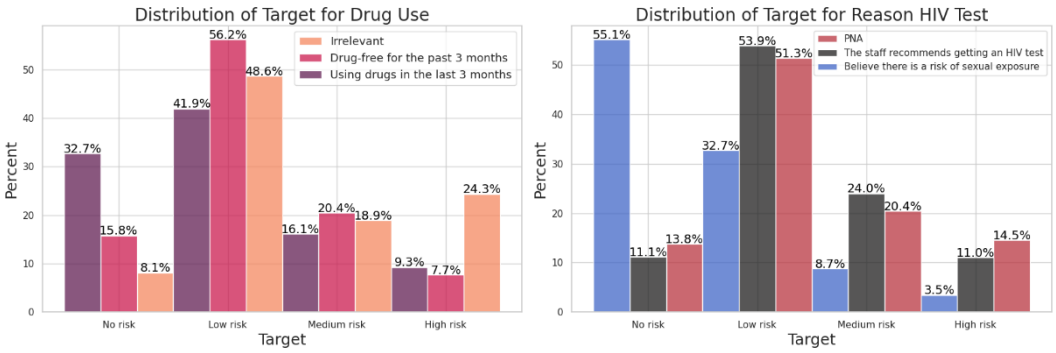
5(c) แสดงการกระจายตัวของเพศกับความเสี่ยง 5(d) แสดงการกระจายตัวของอัตลักษณ์ทางเพศกับความเสี่ยง



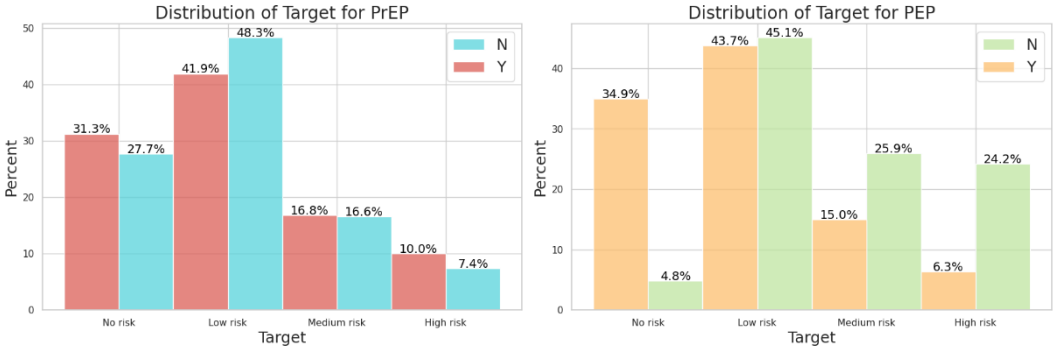
5(e) แสดงการกระจายตัวของการศึกษากับความเสี่ยง 5(f) แสดงการกระจายตัวของอาชีพกับความเสี่ยง

ภาพประกอบ 5 กราฟแท่งแสดงกระจายตัวของกลุ่มข้อมูลทางด้านประชากรกับระดับความเสี่ยงเอชไอวี

3.2.3.3 การกระจายตัวของข้อมูลสำหรับกลุ่มรูปแบบการดำเนินชีวิต (Lifestyles) และการกระจายตัวของข้อมูลดังกล่าวสำหรับแต่ละระดับความเสี่ยง แสดงดังภาพประกอบ 6(b) ซึ่งจะสังเกตได้ว่า เจ้าหน้าที่แนะนำให้ไปตรวจเอชไอวี (The Staff Recommends Getting an HIV Test) แสดงผลกระจายตัวของข้อมูลที่อาจทำให้มีผลต่อแบบจำลองในคลาสไม่มีความเสี่ยง และภาพประกอบ 6(d) ไม่รับ (N) จะมีผลต่อคลาสไม่มีความเสี่ยง ในขณะที่ รับ (Y) จะมีผลต่อกลุ่มมีความเสี่ยงสูง อาจเพราะยา PEP เป็นยาที่มีไว้สำหรับผู้ที่คาดว่ามีการสัมผัสเชื้อเอชไอวี มาภายในระยะเวลาไม่เกิน 72 ชั่วโมง ซึ่งเป็นการป้องกันแบบฉุกเฉิน



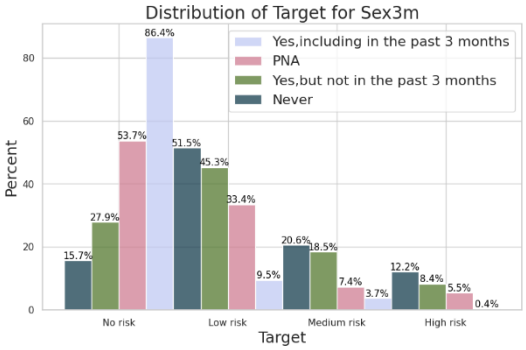
6(a) แสดงการกระจายตัวของการใช้สารเสพติดกับความเสี่ยง 6(b) แสดงการกระจายตัวของเหตุผลที่มาตรวจเอชไอวีกับความเสี่ยง

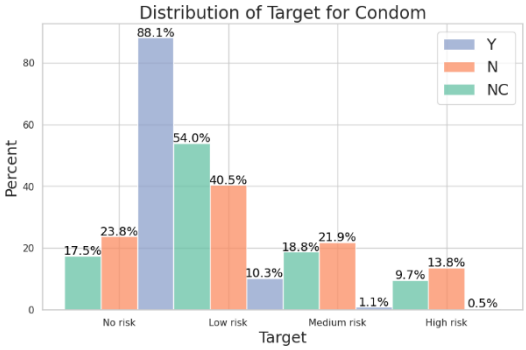


6(c) แสดงการกระจายตัวของการรับยาเพร็พกับความเสี่ยง 6(d) แสดงการกระจายตัวของการรับยาเป็ปกับความเสี่ยง

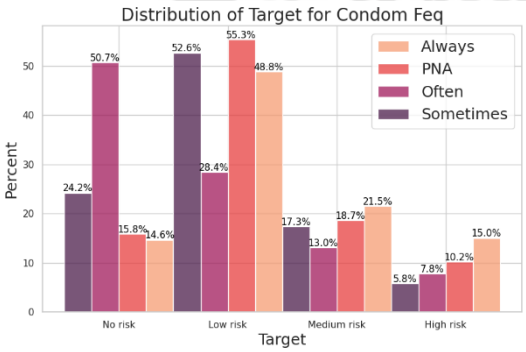
ภาพประกอบ 6 กราฟแท่งแสดงกระจายตัวของกลุ่มรูปแบบการดำเนินชีวิตกับระดับความเสี่ยงเอชไอวี

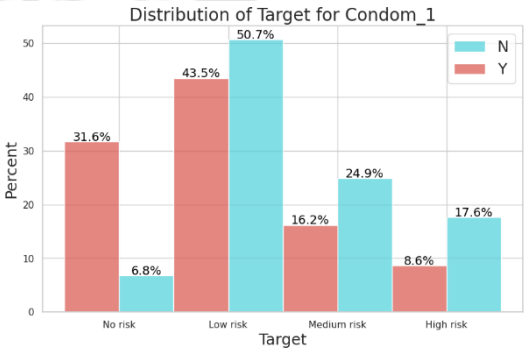
3.2.3.4 การกระจายตัวของข้อมูลสำหรับกลุ่มพฤติกรรมทางเพศ (Sexual Behaviors) และการกระจายตัวของข้อมูลดังกล่าวสำหรับแต่ละระดับความเสี่ยง แสดงดังภาพประกอบ 7(a) ซึ่งจะสังเกตได้ว่า ไม่เคย (Never) จะมีผลต่อกลุ่มไม่มีความเสี่ยง เนื่องจากการการที่มีเพศสัมพันธ์จึงจะทำให้เกิดความเสี่ยงเอชไอวีได้ และภาพประกอบ 7(h) มีและไม่ได้ใช้ถุงยาง (Yes and did not use Condom) จะมีผลต่อความเสี่ยงสูง



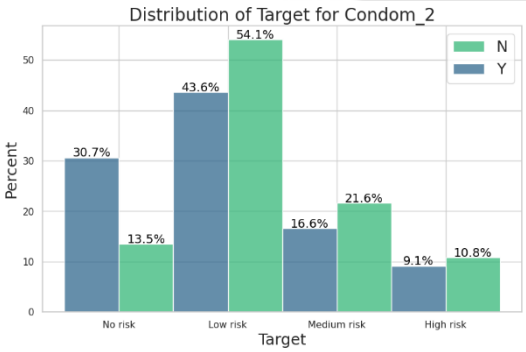


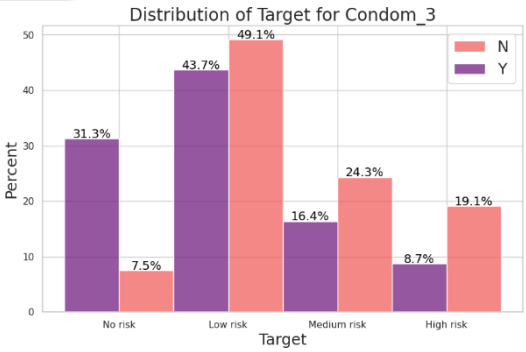
7(a) แสดงการกระจายตัวของเคยมีเพศสัมพันธ์แบบสอดใส่หรือไม่กับความเสี่ยง 7(b) แสดงการกระจายตัวของในช่วงสามเดือนที่ผ่านมาใช้ถุงยางอนามัยระหว่างการมีเพศสัมพันธ์หรือไม่กับความเสี่ยง



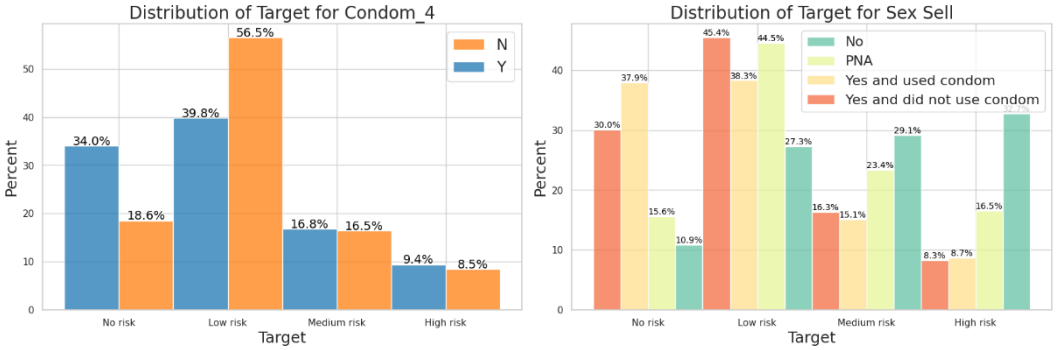


7(c) แสดงการกระจายตัวของในช่วงสามเดือนที่ผ่านมาใช้ถุงยางบ่อยแค่ไหนกับความเสี่ยง 7(d) แสดงการกระจายตัวของถุงยางแตก/ถุงยางรั่วกับความเสี่ยง





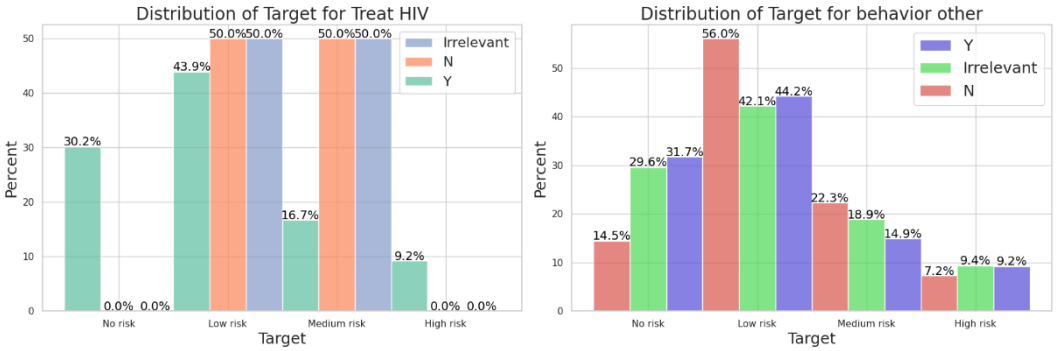
7(e) แสดงการกระจายตัวของถุงยางหลุดในระหว่างการมีเพศสัมพันธ์หรือไม่กับความเสี่ยง 7(f) แสดงการกระจายตัวของคู่นอนถอดถุงยางอนามัยขณะมีเพศสัมพันธ์หรือไม่กับความเสี่ยง



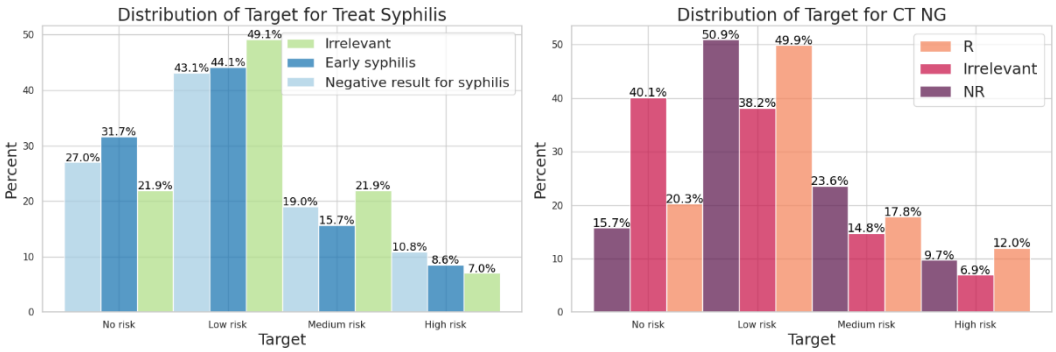
7(g) แสดงการกระจายตัวของมีการสอดใส่ แต่ไม่เคยเกิดเหตุการณ์ใน 3 ข้อ (ถุงยางแตก, ถุงยางหลุด, คู่นอนถอดถุงยาง) ที่กล่าวมาหรือไม่กับความเสี่ยง 7(h) แสดงการกระจายตัวของในช่วงสามเดือนที่ผ่านมามีเพศสัมพันธ์เพื่อแลกกับเงินหรือสิ่งของตอบแทนหรือไม่กับความเสี่ยง

ภาพประกอบ 7 กราฟแท่งแสดงกระจายตัวของกลุ่มพฤติกรรมทางเพศกับระดับความเสี่ยงเอชไอวี

3.2.3.5 การกระจายตัวของข้อมูลสำหรับกลุ่มอาการอื่นๆ (Symptoms) และการกระจายตัวของข้อมูลดังกล่าวสำหรับแต่ละระดับความเสี่ยง แสดงดังภาพประกอบ 8(a) ซึ่งจะสังเกตได้ว่า ไม่เกี่ยวข้อง (Irrelevant) มีผลทั้ง ไม่มีความเสี่ยง และความเสี่ยงสูง อาจเพราะ ผู้วิจัยเลือกข้อมูลเฉพาะครั้งแรกที่ผู้รับบริการเข้ามาตรวจเอชไอวี จึงไม่มีข้อมูลของผู้ที่รับการรักษาเอชไอวีในกลุ่มคนที่มีความเสี่ยงสูง



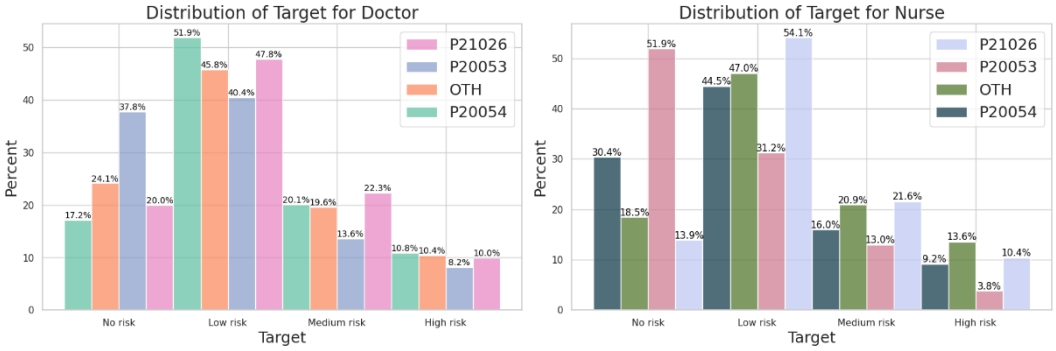
8(a) แสดงการกระจายตัวของการรักษาเอชไอวีกับความเสี่ยง 8(b) แสดงการกระจายตัวของมีอาการของโรคหนองใน หนองในเทียม หรือซิฟิลิสกับความเสี่ยง



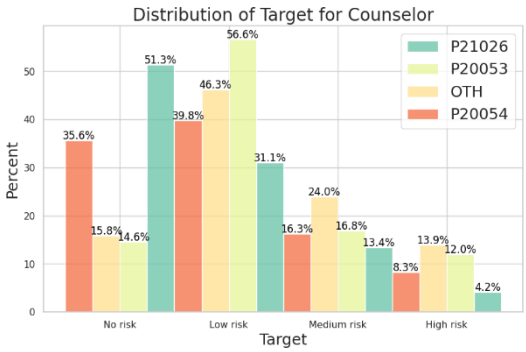
8(c) แสดงการกระจายตัวของผลวินิจฉัยโรค syphilis กับความเสี่ยง 8(d) แสดงการกระจายตัวของผลวินิจฉัยโรคหนองใน หนองในเทียมกับความเสี่ยง

ภาพประกอบ 8 กราฟแท่งแสดงกระจายตัวของกลุ่มอาการอื่นๆกับระดับความเสี่ยงเอชไอวี

3.2.3.6 การกระจายตัวของข้อมูลสำหรับกลุ่มเจ้าหน้าที่ (Staff) และการกระจายตัวของข้อมูลดังกล่าวสำหรับแต่ละระดับความเสี่ยง แสดงดังภาพประกอบ 9(c) ซึ่งจะสังเกตได้ว่าอื่นๆ (OTH) จะมีผลต่อทุกๆคลาส เนื่องจากในคลินิกที่ผู้วิจัยขอข้อมูลมามีพนักงานที่ยังไม่มีรหัสของผู้ให้คำปรึกษาค่อนข้างเยอะ และผู้ที่ให้คำปรึกษาได้เยอะที่สุดคือ P20057



9(a) แสดงการกระจายตัวของหมอกับความเสี่ยง 9(b) แสดงการกระจายตัวของมีพยาบาลกับความเสี่ยง



9(c) แสดงการกระจายตัวของผู้ให้คำปรึกษากับความเสี่ยง

ภาพประกอบ 9 กราฟแท่งแสดงกระจายตัวของกลุ่มเจ้าหน้าที่กับระดับความเสี่ยงเอชไอวี

3.2.4 การแบ่งข้อมูล (Train-Test Split)

ในขั้นตอนนี้ข้อมูลทั้งหมดจะถูกแบ่งออกเป็น ชุดข้อมูลสำหรับฝึกสอน (Training) 80% และ ชุดข้อมูลสำหรับทดสอบ (Test) 20% ด้วยการพยายามให้การแบ่งแยกข้อมูลเป็นกลุ่มย่อยสอดคล้องกับการกระจายของข้อมูลต้นฉบับ (Stratify)

3.2.5 การสุ่มตัวอย่างใหม่แบบเพิ่มจำนวน (Oversampling)

ในชุดข้อมูลสำหรับฝึก จากการสำรวจข้อมูล ข้อมูลตัวแปรตามเป็นข้อมูลที่มีความไม่สมดุล (Imbalanced Data) จึงทำใช้ เครื่องมือสุ่มเพิ่มตัวอย่าง (Random Over Sampler) เพื่อจัดการกับปัญหาข้อมูลที่ไม่สมดุลในการเรียนรู้ของเครื่องจักร โดยมีวัตถุประสงค์เพื่อเพิ่มจำนวนตัวอย่างในคลาสที่มีจำนวนน้อย (Minority Class) เพื่อให้มีความสมดุลกับคลาสที่มีจำนวนมาก (Majority Class)

3.2.6 การเข้ารหัสแบบหนึ่งต่อหนึ่ง (One-Hot encoding)

ทำการเข้ารหัสแบบหนึ่งต่อหนึ่งก่อนนำข้อมูลเข้า Model โดยมีข้อมูล 3621 รายการ และมี 90 คุณลักษณะ

3.3 การสร้างแบบจำลอง (Modeling)

การสร้างแบบจำลองโดยการใส่ทีละกลุ่ม และทำการวัดประสิทธิภาพการทำงานของแบบจำลองที่สร้างขึ้น โดยอาศัยแบบจำลองทั้งสองประเภท ดังนี้

3.3.1 แบบจำลองเพื่อจัดกลุ่มระดับความเสี่ยง (Classification Model)

แบบจำลองเพื่อจัดกลุ่มระดับความเสี่ยงมีจำนวน 4 แบบจำลองได้แก่ Decision Tree, Random Forest, XGBoost Classifier และ Support Vector Machine (SVM)

3.3.2 แบบจำลองเพื่อทำนายค่าระดับความเสี่ยง (Regression Mode)

แบบจำลองเพื่อทำนายค่าระดับความเสี่ยงมีจำนวน 3 แบบจำลอง ได้แก่ XGBoost Regressor, Support Vector Regressor (SVR) และ Regression Neural Network

โดยในการฝึกแบบจำลองจะมีการค้นหาไฮเปอร์พารามิเตอร์ที่ดีที่สุดของแต่ละแบบจำลอง (Hyperparameter) ด้วยเทคนิคการค้นหาแบบกริด (Grid Search)

3.4 การประเมินแบบจำลอง (Evaluation)

3.4.1 การวัดประสิทธิภาพแบบจำลองการจำแนกประเภท

ตัวชี้วัดประสิทธิภาพของแบบจำลองการจำแนกประเภทในการทดลองนี้มี 4 ตัวชี้วัดดังนี้

3.4.1.1 Confusion Matrix

Confusion Matrix ในรูปแบบ Actual-Predict (Actual vs. Predicted) คือตารางที่ใช้ในการแสดงผลการทำนายของแบบจำลอง Classification (Visitsora-at, 2019) โดยจะแบ่งผลการทำนายออกเป็น 4 ส่วนหลัก ๆ ได้แก่

- TP คือ ผลบวกจริง (True Positive) หมายถึงจำนวนข้อมูลที่เป็น Positive และแบบจำลองสามารถทำนายออกมาได้ถูกต้องว่าเป็น Positive

- TN คือ ผลลบจริง (True Negative) หมายถึงจำนวนข้อมูลที่เป็น Negative และแบบจำลองสามารถทำนายออกมาได้ถูกต้องว่าเป็น Negative

- FP คือ ผลบวกลวง (False Positive) หมายถึงจำนวนข้อมูลที่เป็น Negative แต่แบบจำลองทำนายออกมาว่าเป็น Positive

- FN คือ ผลลบลวง (False Positive) หมายถึงจำนวนข้อมูลที่เป็น Positive แต่แบบจำลองทำนายออกมาว่าเป็น Negative

ซึ่งสามารถแสดงในรูปของตาราง 5 ดังนี้

ตาราง 5 การแบ่งผลการทำนาย

Actual Class	Predicted Class	
	Positive	Negative
Positive	TP	FN
Negative	FP	TN

3.4.1.2 Precision (ความแม่นยำ)

Precision คือ อัตราส่วนของตัวอย่างที่ทำนายว่าเป็น Positive และทำนายถูกต้องเมื่อเทียบกับตัวอย่างทั้งหมดที่ทำนายว่าเป็น Positive สูตรคำนวณ คือ

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

Precision จะเพิ่มขึ้นเมื่อแบบจำลองมีการทำนาย Positive ที่แม่นยำมากขึ้น และลดลงเมื่อแบบจำลองมีการทำนาย Positive ที่ไม่แม่นยำ (Visitsora-at, 2019)

3.4.1.3 Recall (ความครอบคลุม) หรือ Sensitivity

Recall คือ อัตราส่วนของตัวอย่างที่มีค่าจริงเป็น Positive และถูกทำนายถูกต้องเมื่อเทียบกับตัวอย่างทั้งหมดที่มีค่าจริงเป็น Positive สูตรคำนวณ คือ

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

Recall จะเพิ่มขึ้นเมื่อแบบจำลองมีการจำแนกคลาส Positive ที่แม่นยำมากขึ้น และลดลงเมื่อแบบจำลองมีการจำแนกคลาส Positive ที่ไม่แม่นยำ

3.4.1.4 F1-Score

F1-Score คือ จะนับความสมดุลระหว่าง Precision และ Recall ของแบบจำลองด้วยกัน ซึ่งเป็นค่าที่สมบูรณ์แบบสำหรับการประเมินความสามารถในการค้นหาตัวอย่างของคลาสที่เป็น Positive อย่างถูกต้อง และความสามารถในการหลีกเลี่ยงคลาสที่เป็น Negative อย่างมีประสิทธิภาพ สูตรคำนวณ คือ

$$F1 - Score = \frac{2 * precision * recall}{(precision + recall)}$$

3.4.1.5 Receiver Operating Characteristic Curve (ROC Curve)

ROC Curve เป็นกราฟที่ใช้ในการประเมินประสิทธิภาพแบบจำลองโดยเฉพาะอย่างยิ่งในกรณีที่มีการประมวลผลข้อมูลที่มีความไม่สมดุล (Imbalanced Data) หรือมีค่าความสำคัญที่แตกต่างกันระหว่าง Positive และ Negative โดยแสดงความสัมพันธ์ระหว่าง Precision กับ Recall โดยมีลักษณะดังนี้

Precision เป็นอัตราส่วนของตัวอย่างที่ถูกต้องที่แบบจำลองจำแนกได้ (True Positive) ต่อจำนวนทั้งหมดของตัวอย่างที่แบบจำลองจำแนกได้ว่าเป็นบวก (Predicted Positive)

Recall เป็นอัตราส่วนของตัวอย่างที่ถูกต้องที่แบบจำลองจำแนกได้ (True Positive) ต่อจำนวนทั้งหมดของตัวอย่างที่เป็นจริงในกลุ่มที่เป็นบวก (Actual Positive)

ROC Curve มีความสำคัญในกรณีที่ข้อมูลมีความไม่สมดุล (Imbalanced data) หรือจำนวนตัวอย่างในแต่ละกลุ่มมีความแตกต่างกันมาก โดยเฉพาะเมื่อจำนวนตัวอย่างในกลุ่มที่เป็นบวกน้อยกว่ามาก แบบจำลองที่มีประสิทธิภาพสูงจะแสดงค่า Precision และ Recall สูงสำหรับกลุ่มที่เป็นบวก

3.4.2 การวัดประสิทธิภาพแบบจำลองการถดถอย

ตัวชี้วัดประสิทธิภาพของแบบจำลองการถดถอยในการทดลองนี้มี 2 ตัวชี้วัดดังนี้

3.4.2.1 MAE หรือ Mean Absolute Error

ตัววัดประสิทธิภาพของแบบจำลองทางสถิติที่ใช้วัดความคลาดเคลื่อนโดยการหาค่าเฉลี่ยของความต่างระหว่างค่าที่ทำนายได้จากแบบจำลองกับค่าจริงที่ได้ ซึ่งจะทนต่อ Data ที่มี Outlier ได้มากกว่า (Promrit, 2020) สูตรคำนวณคือ

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

โดยกำหนดให้

n คือจำนวนตัวอย่างในชุดข้อมูล

y_i คือค่าจริงของตัวอย่างที่ i

$\hat{y}_i$ คือค่าทำนายของตัวอย่างที่ i

ซึ่งจะให้มุมมองเฉลี่ยของความคลาดเคลื่อนในทุกรายการข้อมูล ถ้า MAE น้อย แสดงว่าแบบจำลองมีประสิทธิภาพที่ดีในการทำนาย.

3.4.2.2 Root Mean Squared Error (RMSE)

เป็นการปรับปรุงของ Mean Squared Error (MSE) โดยที่ในการคำนวณ RMSE จะทำการรากที่สองของค่า MSE เพื่อให้ได้ค่าที่มีหน่วยเดียว กับข้อมูลต้นฉบับ (เช่น หน่วยของข้อมูลที่ถูกวัด) สูตรที่ใช้ในการคำนวณ RMSE คือ

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

RMSE มักถูกใช้ในการวัดความคลาดเคลื่อนของแบบจำลองในการทำนายค่าต่อค่าในกรณีของ Regression Models ค่า RMSE ที่น้อยกว่าจะแสดงถึงความแม่นยำของแบบจำลองที่ดีกว่า

บทที่ 4
ผลการดำเนินงานวิจัย

ผู้วิจัยได้ดำเนินการวิจัยโดยแบบจำลองการจำแนกประเภท (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 โดยแบ่งกลุ่มสำหรับทดลองแบบจำลองต่างๆ ตามตัวแปรตามออกเป็น 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)

4.1 ความสำคัญของแต่ละกลุ่มคุณลักษณะ (Feature Grouping)

ในการทดลองนี้ จะทำการศึกษาด้วยการแบ่งตัวแปรอิสระ ออกเป็นกลุ่มๆ เพื่อดูว่ากลุ่มไหนที่จะทำให้แบบจำลองนั้นมีประสิทธิภาพ โดยแบ่งข้อมูลออกเป็นกลุ่มตามลักษณะของกลุ่มคุณลักษณะ(Feature) คือ ทุกคุณลักษณะ (All Features), ข้อมูลทางด้านประชากร (Demographics), รูปแบบการดำเนินชีวิต (Lifestyles), พฤติกรรมทางเพศ (Sexual Behaviors) และ อาการอื่นๆ (Symptoms) ซึ่งมีการวัดผลด้วยค่า Precision, Recall, F1-Score , และ Accuracy สำหรับแต่ละกลุ่มของคลาสที่เราสนใจ (0 (N), 1 (L), 2 (M), และ 3 (H)) ซึ่งผู้วิจัยจะสนใจคลาส 3 (H) เป็นอันดับแรกเนื่องจากเป็นกลุ่มที่มีความเสี่ยงสูง โดยเราสามารถสรุปผลของแต่ละแบบจำลองในแต่ละกลุ่มคุณลักษณะได้จากตาราง 6 จะสังเกตได้ว่า ในกลุ่ม ทุกคุณลักษณะ, ข้อมูลทางด้านประชากร, รูปแบบการดำเนินชีวิต, พฤติกรรมทางเพศ และ อาการอื่นๆ มีค่า Precision สำหรับคลาส 3 (H) คือ (0.25, 0.11, 0.16, 0.13, 0.12) ตามลำดับ และค่า Recall (0.37, 0.51, 0.49, 0.35, 0.46) ตามลำดับ ซึ่งแสดงให้เห็นว่าแบบจำลองที่ใช้ข้อมูลทุกคุณลักษณะมีความสามารถในการจำแนกคลาส 3 (H) ได้ดีกว่าแบบจำลองในกลุ่ม ข้อมูลทางด้านประชากร, รูปแบบการดำเนินชีวิต, พฤติกรรมทางเพศ และ อาการอื่นๆ ที่มีค่า Precision ต่ำกว่า ดังนั้นการทดลองต่อไปนี้จึงเลือกใช้กลุ่มคุณลักษณะ ทุกคุณลักษณะ ในการทดสอบแต่ละแบบจำลอง

ตาราง 6 ค่า Precision, Recall, F1-Score , และ Accuracy สำหรับแต่ละกลุ่มในแต่ละแบบจำลองการจำแนกประเภท

Model	Feature Grouping	Class	Precision	Recall	F1-Score	Accuracy
Decision Tree	All Features	0 (N)	0.67	0.64	0.65	0.40
		1 (L)	0.54	0.38	0.44	
		2 (M)	0.21	0.24	0.23	
		3 (H)	0.21	0.48	0.29	
	Demographics	0 (N)	0.48	0.52	0.50	0.30
		1 (L)	0.44	0.15	0.22	
		2 (M)	0.25	0.19	0.21	
		3 (H)	0.11	0.51	0.18	
	Lifestyles	0 (N)	0.52	0.71	0.60	0.42
		1 (L)	0.54	0.33	0.41	
		2 (M)	0.28	0.10	0.15	
		3 (H)	0.16	0.49	0.24	
	Sexual behaviors	0 (N)	0.81	0.52	0.63	0.43
		1 (L)	0.51	0.43	0.46	
		2 (M)	0.24	0.32	0.28	
		3 (H)	0.14	0.35	0.20	
	Symptoms	0 (N)	0.44	0.52	0.48	0.27
		1 (L)	0.00	0.00	0.00	
		2 (M)	0.23	0.40	0.29	
		3 (H)	0.12	0.46	0.19	

ตาราง 6 (ต่อ)

Model	Feature Grouping	Class	Precision	Recall	F1-Score	Accuracy
Random Forest	All Features	0 (N)	0.70	0.63	0.66	0.46
		1 (L)	0.56	0.50	0.53	
		2 (M)	0.18	0.22	0.20	
		3 (H)	0.25	0.37	0.30	
	Demographics	0 (N)	0.51	0.55	0.53	0.35
		1 (L)	0.51	0.23	0.32	
		2 (M)	0.32	0.27	0.29	
		3 (H)	0.11	0.44	0.17	
	Lifestyles	0 (N)	0.51	0.72	0.60	0.43
		1 (L)	0.54	0.31	0.39	
		2 (M)	0.28	0.27	0.28	
		3 (H)	0.21	0.39	0.27	
	Sexual behaviors	0 (N)	0.73	0.54	0.62	0.42
		1 (L)	0.51	0.40	0.45	
		2 (M)	0.24	0.33	0.28	
		3 (H)	0.13	0.30	0.18	
	Symptoms	0 (N)	0.43	0.52	0.47	0.26
		1 (L)	0.56	0.03	0.06	
		2 (M)	0.24	0.16	0.19	
		3 (H)	0.11	0.67	0.19	
XGBoost Classifier	All Features	0 (N)	0.72	0.61	0.66	0.48
		1 (L)	0.58	0.51	0.54	
		2 (M)	0.18	0.21	0.19	
		3 (H)	0.27	0.45	0.34	

ตาราง 6 (ต่อ)

Model	Feature Grouping	Class	Precision	Recall	F1-Score	Accuracy
XGBoost Classifier	Demographics	0 (N)	0.51	0.57	0.54	0.34
		1 (L)	0.50	0.23	0.31	
		2 (M)	0.22	0.19	0.21	
		3 (H)	0.10	0.37	0.15	
	Lifestyles	0 (N)	0.52	0.71	0.60	0.43
		1 (L)	0.52	0.32	0.40	
		2 (M)	0.24	0.21	0.22	
		3 (H)	0.21	0.40	0.28	
	Sexual behaviors	0 (N)	0.85	0.52	0.65	0.45
		1 (L)	0.52	0.47	0.49	
		2 (M)	0.26	0.31	0.28	
		3 (H)	0.16	0.40	0.22	
	Symptoms	0 (N)	0.42	0.52	0.47	0.25
		1 (L)	0.48	0.04	0.07	
		2 (M)	0.23	0.12	0.16	
		3 (H)	0.09	0.58	0.16	
Support Vector Machine	All Features	0 (N)	0.69	0.59	0.64	0.45
		1 (L)	0.57	0.43	0.49	
		2 (M)	0.15	0.20	0.17	
		3 (H)	0.23	0.46	0.30	
	Demographics	0 (N)	0.47	0.64	0.54	0.32
		1 (L)	0.52	0.14	0.22	
		2 (M)	0.28	0.12	0.17	
		3 (H)	0.10	0.49	0.17	

ตาราง 6 (ต่อ)

Model	Feature Grouping	Class	Precision	Recall	F1-Score	Accuracy
	Lifestyles	0 (N)	0.48	0.73	0.58	0.45
		1 (L)	0.51	0.45	0.48	
		2 (M)	0.00	0.00	0.00	
		3 (H)	0.23	0.46	0.30	
	Sexual behaviors	0 (N)	0.90	0.50	0.64	0.44
		1 (L)	0.50	0.56	0.53	
		2 (M)	0.18	0.04	0.07	
		3 (H)	0.11	0.44	0.18	
	Symptoms	0 (N)	0.43	0.55	0.48	0.38
		1 (L)	0.46	0.37	0.41	
		2 (M)	0.22	0.24	0.23	
		3 (H)	0.13	0.09	0.10	

4.2 Multi-Class หรือ 4 Class

โดยมี Class Target ดังนี้ 0: No Risk

1: Low Risk

2: Medium risk

3: High Risk

ในการทดลองแรกผู้วิจัยมีจุดประสงค์เพื่อดูประสิทธิภาพของการทำนายใน Multi-Class โดยผู้วิจัยใช้เทคนิคการค้นหาแบบกริดเพื่อหา Hyperparameters ที่ดีที่สุดโดยในแบบจำลองการจำแนกประเภท ซึ่งมีรายละเอียดดังตาราง 7

ตาราง 7 Hyperparameters ในแบบจำลองการจำแนกประเภทของกลุ่ม Multi-Class

Model Type	Hyperparameters
Decision Tree	Criterion: Entropy
	Max Depth: 30
	Max Features: log2
	Min Samples Leaf: 4
	Min Samples Split: 10
Random Forest	Max Depth: 10
	Max Features: 0.5
	Min Samples Leaf: 2
	Min Samples Split: 10
	n_estimators: 50
XGBoost Classifier	Learning Rate: 0.1
	Max Depth: 5
	n_estimators: 50
	Subsample: 0.9
Support Vector Machine (SVM)	C: 0.1
	Gamma: Scale

ตาราง 8 แสดงผลการทดลองโดยแสดง Confusion matrix, Precision, Recall และ F1-Score ของในกลุ่ม Multi-Class ในแบบจำลองการจำแนกประเภท จากผลการทดลองจะสังเกตได้ว่าใน Class 2 (M) และ Class 3 (H) จะมี Precision, Recall และ F1-Score ที่ค่อนข้างต่ำ แสดงให้เห็นว่าแบบจำลองการจำแนกประเภทยังทำนายผลใน Class 2 (M) และ Class 3 (H) ได้ไม่แม่นยำในทั้ง 4 แบบจำลอง

ตาราง 8 ค่า Confusion matrix, Precision, Recall และ F1-Score ในแบบจำลองการจำแนกประเภทของกลุ่ม Multi-Class

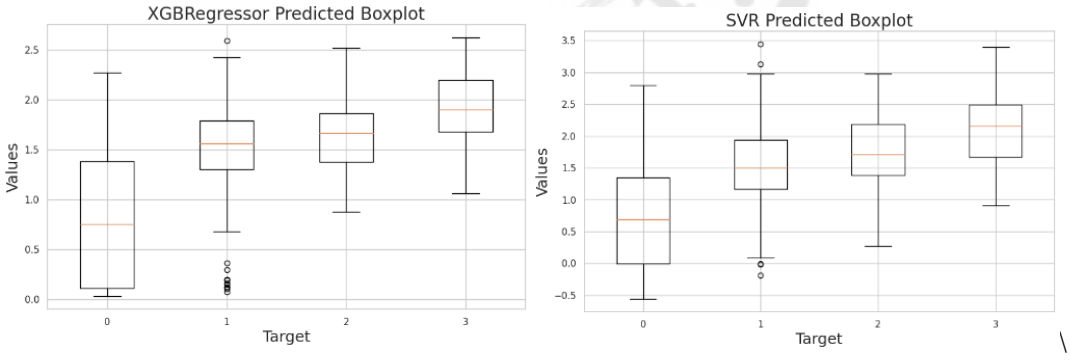
Model	Class	Confusion matrix				Precision	Recall	F1-Score
		0 (N)	1 (L)	2 (M)	3 (H)			
Decision Tree	0 (N)	139	41	27	11	0.62	0.64	0.63
	1 (L)	58	134	75	52	0.55	0.42	0.48
	2 (M)	21	52	24	24	0.16	0.20	0.18
	3 (H)	8	17	20	22	0.20	0.33	0.25
Random Forest	0 (N)	126	64	23	5	0.75	0.58	0.65
	1 (L)	31	167	67	54	0.57	0.52	0.54
	2 (M)	8	49	33	31	0.23	0.27	0.25
	3 (H)	3	15	19	30	0.25	0.45	0.32
XGBoost Classifier	0 (N)	133	52	18	15	0.70	0.61	0.65
	1 (L)	42	152	67	58	0.59	0.48	0.53
	2 (M)	15	48	23	35	0.18	0.19	0.18
	3 (H)	0	7	23	37	0.26	0.55	0.35
Support Vector Machine	0 (N)	131	48	28	11	0.67	0.60	0.63
	1 (L)	53	127	78	61	0.56	0.40	0.47
	2 (M)	11	43	31	36	0.20	0.26	0.23
	3 (H)	0	9	16	42	0.28	0.63	0.39

เนื่องจากผลการทดลองในแบบจำลองการจำแนกประเภท มีผลลัพธ์การทำนายยังไม่แม่นยำ และจากผลการทดลองจะสังเกตได้ว่าแบบจำลองการจำแนกประเภทได้ทำนายข้ามระดับความเสี่ยงไปค่อนค้างมาก อาจเป็นเพราะแบบจำลองการจำแนกประเภทไม่ได้คำนึงถึงข้อมูลที่เป็นระดับ (Ordinal) จึงทำการทดลองต่อด้วยแบบจำลองการถดถอย จากตาราง 9 แสดง Hyperparameters ที่ดีที่สุด และผลการทดลองโดยแสดง MAE และ RMSE ของกลุ่ม Multi-Class ในแบบจำลองการถดถอย จากผลการทดลองจะสังเกตได้ว่า ค่า MAE และ RMSE มีค่าที่ค่อนข้าง

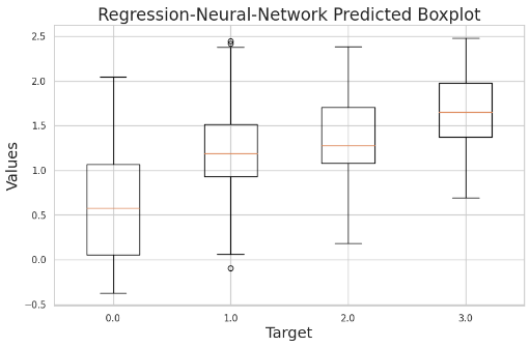
สูงซึ่งหมายถึงแบบจำลองทำนายได้ไม่แม่นยำ จากนั้นผู้วิจัยได้ Boxplot เพื่อแสดงค่าการทำนายในแต่ละคลาสดังภาพประกอบ 10

ตาราง 9 Hyperparameters และค่า MAE และ RMSE ในแบบจำลองการการถดถอยของกลุ่ม Multi-Class

Model Type	Hyperparameters	MAE	RMSE
XGBoost Regressor	Learning Rate: 0.015 Max Depth: 4 n_estimators: 300	0.67	0.84
Support Vector Regressor (SVR)	C: 0.1 Epsilon: 0.2 Gamma: scale	0.67	0.84
Regression Neural Network	Epochs=100 Batch Size=50 Verbose=1	0.61	0.79



10(a) แสดงค่าการทำนายของแบบจำลอง XGBoost Regressor 10(b) แสดงค่าการทำนายของแบบจำลอง Support Vector Regressor



10(c) แสดงค่าการทำนายของแบบจำลอง Regression Neural Network

ภาพประกอบ 10 Boxplot แสดงค่าการทำนายในแต่ละระดับความเสี่ยงเอชไอวีของกลุ่ม Multi-Class

จากนั้นผู้วิจัยได้แปลงค่าทำทายของแบบจำลองการถดถอยให้เป็นคลาส 0, 1, 2, 3 ด้วยวิธีการปัดเศษทศนิยม (Round) ให้เป็นจำนวนเต็ม และนำมาเปรียบเทียบกับผลการทดลองด้วยแบบจำลองการจำแนกประเภทดังตาราง 10

ตาราง 10 ค่า Confusion matrix, Precision, Recall และ F1-Score ในแบบจำลองการจำแนกประเภทเปรียบเทียบกับแบบจำลองการถดถอย ของกลุ่ม Multi-Class

Model	Class	Confusion matrix				Precision	Recall	F1-Score
		0 (N)	1 (L)	2 (M)	3 (H)			
Decision Tree	0 (N)	139	41	27	11	0.62	0.64	0.63
	1 (L)	58	134	75	52	0.55	0.42	0.48
	2 (M)	21	52	24	24	0.16	0.20	0.18
	3 (H)	8	17	20	22	0.20	0.33	0.25
Random Forest	0 (N)	126	64	23	5	0.75	0.58	0.65
	1 (L)	31	167	67	54	0.57	0.52	0.54
	2 (M)	8	49	33	31	0.23	0.27	0.25
	3 (H)	3	15	19	30	0.25	0.45	0.32

ตาราง 10 (ต่อ)

Model	Class	Confusion matrix				Precision	Recall	F1-Score
		0 (N)	1 (L)	2 (M)	3 (H)			
XGBoost Classifier	0 (N)	133	52	18	15	0.70	0.61	0.65
	1 (L)	42	152	67	58	0.59	0.48	0.53
	2 (M)	15	48	23	35	0.18	0.19	0.18
	3 (H)	0	7	23	37	0.26	0.55	0.35
Support Vector Machine	0 (N)	131	48	28	11	0.67	0.60	0.63
	1 (L)	53	127	78	61	0.56	0.40	0.47
	2 (M)	11	43	31	36	0.20	0.26	0.23
	3 (H)	0	9	16	42	0.28	0.63	0.39
XGBoost Regressor	0 (N)	101	72	45	0	0.91	0.46	0.61
	1 (L)	10	139	168	2	0.53	0.44	0.48
	2 (M)	0	43	76	2	0.22	0.63	0.33
	3 (H)	0	7	57	3	0.43	0.04	0.08
Support Vector Regressor	0 (N)	97	74	43	1	0.89	0.44	0.59
	1 (L)	11	141	151	15	0.54	0.44	0.48
	2 (M)	1	42	61	17	0.21	0.50	0.29
	3 (H)	0	6	41	20	0.38	0.30	0.33
Regression Neural Network	0 (N)	107	83	25	0	0.87	0.49	0.63
	1 (L)	16	175	126	2	0.53	0.55	0.54
	2 (M)	0	59	62	0	0.23	0.51	0.32
	3 (H)	0	12	52	3	0.60	0.04	0.08

จะสังเกตได้ว่าผลการทดลองในแต่ละแบบจำลองทำนายผลได้ไม่แม่นยำโดยเฉพาะในคลาสที่เกี่ยวข้องกับผู้มีความเสี่ยงสูงซึ่งเป็นกลุ่มเป้าหมายการคัดกรองในงานวิจัยนี้ ผู้วิจัยจึงรวมคลาสจาก Multi-Class ให้เป็น Binary Class เพื่อทดลองว่าแบบจำลองจะสามารถทำนายได้แม่นยำมากขึ้นหรือไม่

4.3 Binary Class หรือ 2 Class

แบ่งกลุ่มการทดลองจากคลาสเป็น Binary Class หรือ 2 Class โดยแบ่ง Class Target ได้ 3 แบบ ดังนี้

แบบ A (N, L, M : H) 0: No Risk, Low Risk, Medium Risk

1: High Risk

แบบ B (N, L : M, H) 0: No Risk, Low Risk

1: Medium risk, High Risk

แบบ C (N : L, M, H) 0: No Risk

1: Low Risk, Medium risk, High Risk

4.3.1 Binary Class A (N, L, M : H)

จะเป็นการเทรนแบบจำลองโดยที่มีการรวมคลาสให้เหลือเป็น 2 Class โดย Binary Class A แบ่ง Class ได้ดังนี้

0: No Risk, Low Risk, Medium Risk

1: High Risk

Hyperparameters ที่ดีที่สุดในแบบจำลองการจำแนกประเภทดังตาราง 11

ตาราง 11 Hyperparameters ในแบบจำลองการจำแนกประเภทของกลุ่ม Binary Class A (N, L, M : H)

Model Type	Hyperparameters
Decision Tree	Criterion: Entropy
	Max Depth: 30
	Max Features: log2
	Min Samples Leaf: 4
	Min Samples Split: 5
Random Forest	Max Depth: 10
	Max Features: log2
	Min Samples Leaf: 2
	Min Samples Split: 5
	n_estimators: 10

ตาราง 11 (ต่อ)

Model Type	Hyperparameters
XGBoost Classifier	Learning Rate: 0.1 Max Depth: 4 n_estimators: 50 Subsample: 1.0
Support Vector Machine (SVM)	C: 0.1 Gamma: scale

ตาราง 12 แสดงผลการทดลองโดยแสดง Confusion matrix, Precision, Recall และ F1-Score ของในกลุ่ม Binary Class A (N, L, M : H) ในแบบจำลองการจำแนกประเภท จากผลการทดลองจะสังเกตได้ว่าแบบจำลองทำนายได้แม่นยำดีใน Class 0 ซึ่งมี Precision, Recall และ F1-Score ที่ค่อนข้างสูงมีค่าระหว่าง (0.93-0.97), (0.72-0.83), (0.82-0.88) ตามลำดับ และทำนายได้ไม่แม่นยำใน Class 1 ซึ่งมี Precision, Recall และ F1-Score ที่ค่อนข้างต่ำซึ่งอยู่ระหว่าง (0.16-0.23), (0.43-0.75), (0.23-0.35) ตามลำดับ

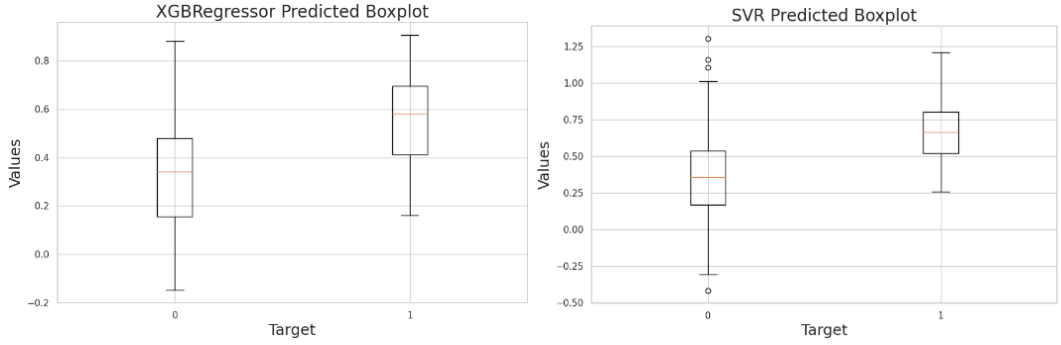
ตาราง 12 ค่า Confusion matrix, Precision, Recall และ F1-Score ในแบบจำลองการจำแนกประเภทของกลุ่ม Binary Class A (N, L, M : H)

Model	Class	Confusion matrix		Precision	Recall	F1-Score
		0 (N,L,M)	1 (H)			
Decision Tree	0 (N,L,M)	501	157	0.93	0.76	0.84
	1 (H)	38	29	0.16	0.43	0.23
Random Forest	0 (N,L,M)	544	114	0.94	0.83	0.88
	1 (H)	33	34	0.23	0.51	0.32
XGBoost Classifier	0 (N,L,M)	500	158	0.96	0.76	0.85
	1 (H)	20	47	0.23	0.70	0.35
Support Vector Machine	0 (N,L,M)	471	187	0.97	0.72	0.82
	1 (H)	17	50	0.21	0.75	0.33

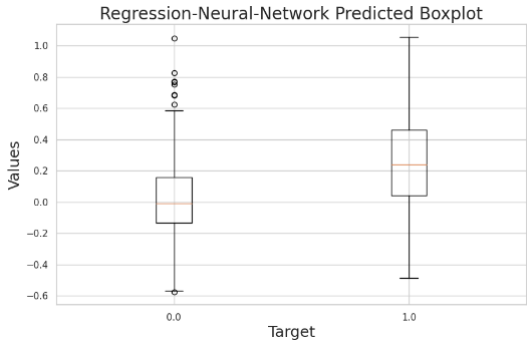
เนื่องจากผลการทดลองในแบบจำลองการจำแนกประเภท มีผลลัพธ์การทำนายยังไม่แม่นยำใน Class 1 (H) จากนั้นจึงทำการทดลองต่อด้วยแบบจำลองการถดถอย จากตาราง 13 แสดง Hyperparameters ที่ดีที่สุด และผลการทดลองโดยแสดง MAE และ RMSE ของกลุ่ม Binary Class A (N, L, M : H) ในแบบจำลองการถดถอย จากผลการทดลองจะสังเกตได้ว่า ค่า MAE และ RMSE มีค่าที่ค่อนข้างต่ำ ซึ่งหมายถึงแบบจำลองทำนายได้แม่นยำมากขึ้นมีค่าอยู่ระหว่าง (0.26-0.37), (0.33-0.44) ตามลำดับ ผู้วิจัยคาดว่าจากค่า MAE และ RMSE ที่มีค่าต่ำลงอาจเป็นเพราะแบบจำลองสามารถทำนายคลาส0ได้ดีมาก ผู้วิจัยจึงได้ Boxplot เพื่อแสดงค่าการทำนายในแต่ละคลาส ดังภาพประกอบ 11

ตาราง 13 Hyperparameters และค่า MAE และ RMSE ในแบบจำลองการการถดถอยของกลุ่ม Binary Class A (N, L, M : H)

Model Type	Hyperparameters	MAE	RMSE
XGBoost Regressor	Learning Rate: 0.015 Max Depth: 4 n_estimators: 300	0.34	0.40
Support Vector Regressor (SVR)	C: 0.1 Epsilon: 0.2 Gamma: scale	0.37	0.44
Regression Neural Network	Epochs=100 Batch Size=50 Verbose=1	0.26	0.33



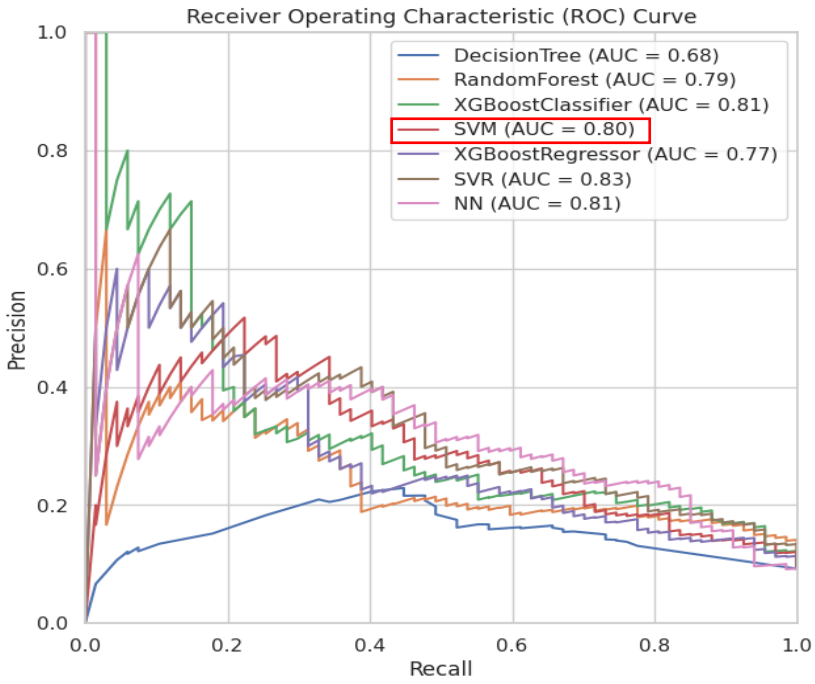
11(a) แสดงค่าการทำนายของแบบจำลอง XGBoost Regressor 11(b) แสดงค่าการทำนายของแบบจำลอง Support Vector Regressor



11(c) แสดงค่าการทำนายของแบบจำลอง Regression Neural Network

ภาพประกอบ 11 Boxplot แสดงค่าการทำนายในแต่ละระดับความเสี่ยงเอชไอวีของกลุ่ม Binary Class A (N, L, M : H)

ภาพประกอบ 12 ผู้วิจัยแสดงกราฟ ROC Curve (Receiver Operating Characteristic Curve) ในกลุ่ม Binary Class A (N, L, M : H) เพื่อแสดงแบบจำลองที่มีประสิทธิภาพในการจำแนกและคัดแยก Positive Class และ Negative Class ได้อย่างเหมาะสม จะสังเกตได้ว่า Support Vector Regressor (SVR) นั้น มีค่า AUC สูงที่สุดที่ 0.83 ซึ่งแบบแบบจำลองที่ดีที่สุดในกลุ่ม Binary Class A



ภาพประกอบ 12 ROC Curve เปรียบเทียบแบบจำลองของกลุ่ม Binary Class A (N, L, M : H)

4.3.2 Binary Class B (N, L : M, H)

เนื่องจากการทดลอง Binary Class A (N, L, M : H) มีผลการทดลองแบบจำลองการจำแนกประเภทมีค่า Precision ในคลาส 0 และ1 ที่ห่างกันมาก อาจเพราะในคลาส0 นั้นมีจำนวนตัวอย่างค่อนข้างเยอะ ดังนั้นจึงจัดกลุ่ม Binary Class ใหม่ โดย Binary Class B (N, L : M, H) แบ่ง Class ได้ดังนี้

0: No Risk, Low Risk

1: Medium risk, High Risk

ซึ่ง Hyperparameters ที่ดีที่สุดโดยในแบบจำลองการจำแนกประเภท มีรายละเอียดดังตาราง 14

ตาราง 14 Hyperparameters ในแบบจำลองการจำแนกประเภทของกลุ่ม Binary Class B (N, L : M, H)

Model Type	Hyperparameters
Decision Tree	Criterion: Qini
	Max Depth: 20
	Max Features: log2
	Min Samples Leaf: 4
	Min Samples Split: 10
Random Forest	Max Depth: 10
	Max Features: sqrt
	Min Samples Leaf: 2
	Min Samples Split: 2
	n_estimators: 50
XGBoost Classifier	Learning Rate: 0.1
	Max Depth: 4
	n_estimators: 50
	Subsample: 0.8
Support Vector Machine (SVM)	C: 0.1
	Gamma: scale

ตาราง 15 แสดงผลการทดลองโดยแสดง Confusion matrix, Precision, Recall และ F1-Score ของในกลุ่ม Binary Class B (N, L : M, H) ในแบบจำลองการจำแนกประเภท จากผลการทดลองจะสังเกตได้ว่าแบบจำลองทำนายได้แม่นยำดีใน Class 0 ซึ่งมี Precision, Recall และ F1-Score ที่ค่อนข้างสูงมีค่าระหว่าง (0.78-0.85), (0.57-0.65), (0.68-0.73) ตามลำดับ และทำนายได้ดีขึ้นใน Class 1 ซึ่งมี Precision, Recall และ F1-Score อยู่ระหว่าง (0.32-0.39), (0.49-0.71), (0.39-0.48) ตามลำดับ

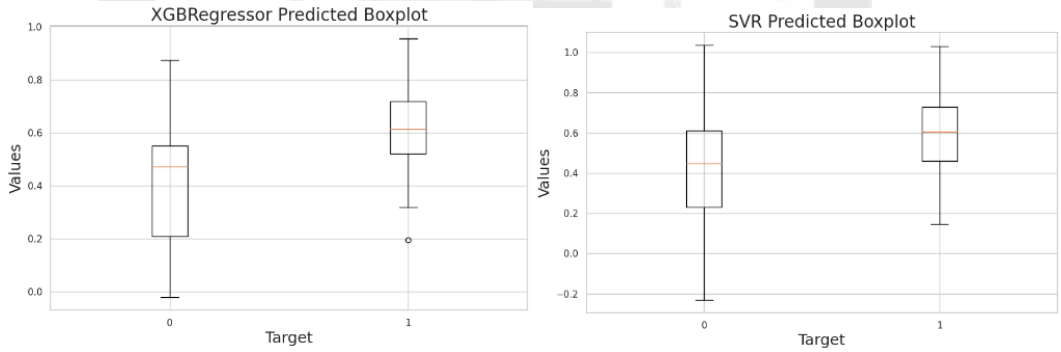
ตาราง 15 ค่า Confusion matrix, Precision, Recall และ F1-Score ในแบบจำลองการจำแนกประเภทของกลุ่ม Binary Class B (N, L : M, H)

Model	Class	Confusion matrix		Precision	Recall	F1-Score
		0 (N,L)	1 (M,H)			
Decision Tree	0 (N,L)	337	200	0.78	0.63	0.70
	1 (M,H)	95	93	0.32	0.49	0.39
Random Forest	0 (N,L)	348	189	0.83	0.65	0.73
	1 (M,H)	69	119	0.39	0.63	0.48
XGBoost Classifier	0 (N,L)	341	196	0.83	0.64	0.72
	1 (M,H)	70	118	0.38	0.63	0.47
Support Vector Machine	0 (N,L)	308	229	0.85	0.57	0.68
	1 (M,H)	55	133	0.37	0.71	0.48

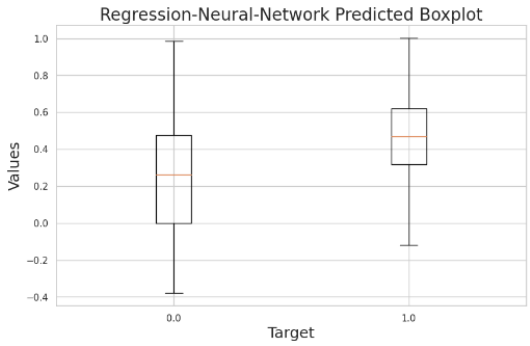
ตาราง 16 แสดง Hyperparameters ที่ดีที่สุด และผลการทดลองโดยแสดง MAE และ RMSE ของกลุ่ม Binary Class B (N, L : M, H) ในแบบจำลองการถดถอย จากผลการทดลองจะสังเกตได้ว่า ค่า MAE และ RMSE มีค่าอยู่ระหว่าง (0.36-0.41), (0.43-0.47) ตามลำดับ ซึ่งมีค่าที่สูงขึ้นจากกลุ่ม Binary Class A (N, L, M : H) ผู้วิจัยคาดว่าจากค่า MAE และ RMSE ที่มีค่าสูงขึ้นอาจเป็นเพราะแบบจำลองสามารถทำนายคลาส1ได้ดีขึ้นจากจำนวนตัวอย่างที่มากขึ้น แต่ก็ทำนายคลาส0 ได้น้อยลงเช่นกัน ผู้วิจัยจึงได้ Boxplot เพื่อแสดงค่าการทำนายในแต่ละคลาสดังภาพประกอบ 13

ตาราง 16 Hyperparameters และค่า MAE และ RMSE ในแบบจำลองการการถดถอยของกลุ่ม Binary Class B (N, L : M, H)

Model Type	Hyperparameters	MAE	RMSE
XGBoost Regressor	Learning Rate: 0.015 Max Depth: 4 n_estimators: 300	0.38	0.44
Support Vector Regressor (SVR)	C: 0.1 Epsilon: 0.3 Gamma: scale	0.41	0.47
Regression Neural Network	Epochs=100 Batch Size=50 Verbose=1	0.36	0.43



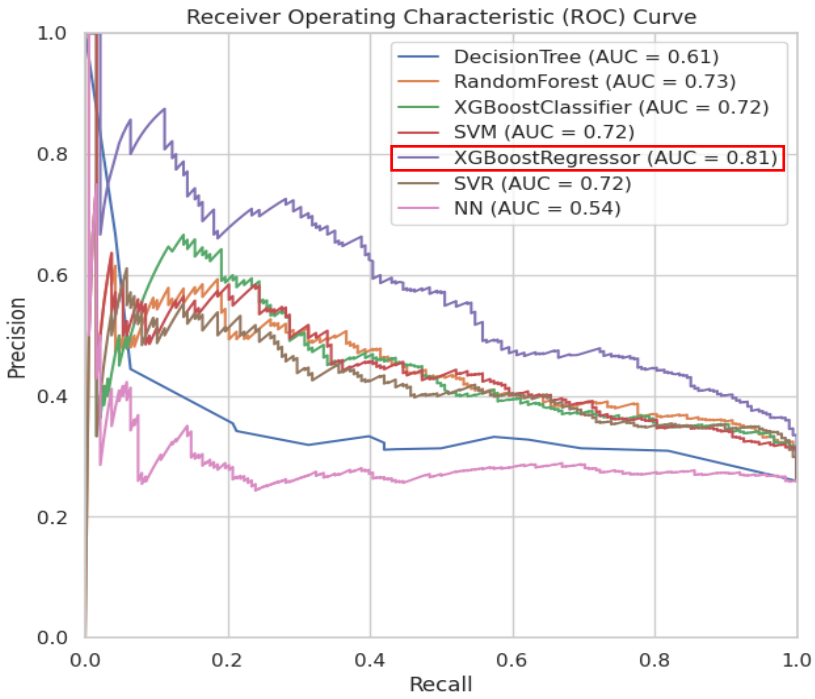
13(a) แสดงค่าการทำนายของแบบจำลอง XGBoost Regressor 13(b) แสดงค่าการทำนายของแบบจำลอง Support Vector Regressor



13(c) แสดงค่าการทำนายของแบบจำลอง Regression Neural Network

ภาพประกอบ 13 Boxplot แสดงค่าการทำนายในแต่ละระดับความเสี่ยงเอชไอวีของกลุ่ม Binary Class B (N, L : M, H)

ภาพประกอบ 14 ผู้วิจัยแสดงกราฟ ROC Curve (Receiver Operating Characteristic Curve) ในกลุ่ม Binary Class B (N, L : M, H) เพื่อแสดงแบบจำลองที่มีประสิทธิภาพในการจำแนกและคัดแยก Positive Class และ Negative Class ได้อย่างเหมาะสม จะสังเกตได้ว่า XGBoost Regressor นั้น มีค่า AUC สูงที่สุดที่ 0.81 ซึ่งแบบแบบจำลองที่ดีที่สุดในกลุ่ม Binary Class B (N, L : M, H)



ภาพประกอบ 14 ROC Curve เปรียบเทียบแบบจำลองของกลุ่ม Binary Class B (N, L : M, H)

4.3.3 Binary Class C (N : L, M, H)

จะเป็นการเทรนแบบจำลองโดยที่มีการรวมคลาส ให้เหลือเป็น 2 class โดย Binary Class C (N : L, M, H) แบ่ง Class ได้ดังนี้

0: No Risk

1: Low Risk, Medium risk, High Risk

ซึ่ง Hyperparameters ที่ดีที่สุดโดยในแบบจำลองการจำแนกประเภท มีรายละเอียดดังตาราง 17

ตาราง 17 Hyperparameters ในแบบจำลองการจำแนกประเภทของกลุ่ม Binary Class C (N : L, M, H)

Model Type	Hyperparameters
Decision Tree	Criterion: Entropy Max Depth: 30 Max Features: log2 Min Samples Leaf: 4 Min Samples Split: 2
Random Forest	Max Depth: 20 Max Features: log2 Min Samples Leaf: 1 Min Samples Split: 10 n_estimators: 100
XGBoost Classifier	Learning Rate: 0.1 Max Depth: 5 n_estimators: 100 Subsample: 1.0
Support Vector Machine (SVM)	C: 0.1 Gamma: scale

ตาราง 18 แสดงผลการทดลองโดยแสดง Confusion matrix, Precision, Recall และ F1-Score ของในกลุ่ม Binary Class C (N : L, M, H) ในแบบจำลองการจำแนกประเภท จากผลการทดลองจะสังเกตได้ว่าแบบจำลองทำนายได้ค่อนข้างแม่นยำดีในทั้งสองคลาส ซึ่งมี Precision, Recall และ F1-Score ใน Class 0 มีค่าระหว่าง (0.53-0.73), (0.61-0.71), (0.59-0.67) ตามลำดับ และใน Class 1 มีค่าอยู่ระหว่าง (0.83-0.87), (0.76-0.91), (0.79-0.87) ตามลำดับ

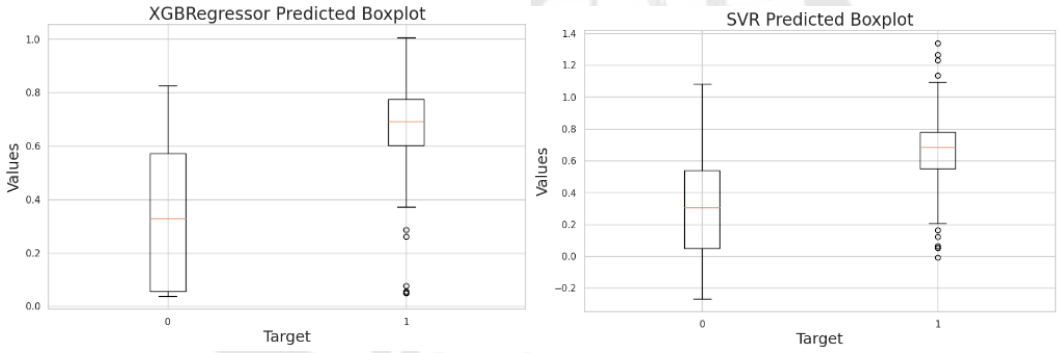
ตาราง 18 ค่า Confusion matrix, Precision, Recall และ F1-Score ในแบบจำลองการจำแนกประเภทของกลุ่ม Binary Class C (N : L, M, H)

Model	Class	Confusion matrix		Precision	Recall	F1-Score
		0 (N)	1 (L,M,H)			
Decision Tree	0 (N)	141	77	0.53	0.65	0.59
	1 (L,M,H)	123	384	0.83	0.76	0.79
Random Forest	0 (N)	133	85	0.73	0.61	0.67
	1 (L,M,H)	48	459	0.84	0.91	0.87
XGBoost Classifier	0 (N)	149	69	0.64	0.68	0.66
	1 (L,M,H)	85	422	0.86	0.83	0.85
Support Vector	0 (N)	155	63	0.64	0.71	0.67
Machine	1 (L,M,H)	87	420	0.87	0.83	0.85

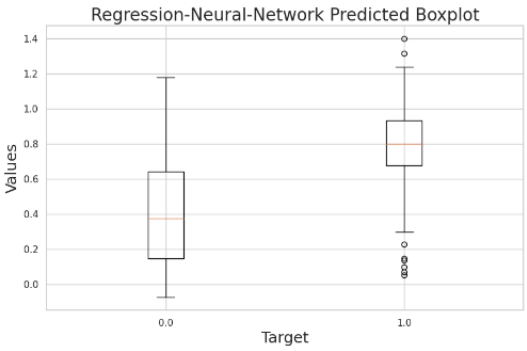
ตาราง 19 แสดง Hyperparameters ที่ดีที่สุด และผลการทดลองโดยแสดง MAE และ RMSE ของกลุ่ม Binary Class C (N : L, M, H) ในแบบจำลองการถดถอย จากผลการทดลองจะสังเกตได้ว่า ค่า MAE และ RMSE มีค่าที่ค่อนข้างต่ำ ซึ่งหมายถึงแบบจำลองทำนายได้แม่นยำดี ซึ่งมีค่าอยู่ระหว่าง (0.28-0.34), (0.36-0.40) ตามลำดับ จากนั้นผู้วิจัยจึงได้ Boxplot เพื่อแสดงค่าการทำนายในแต่ละ Class ดังภาพประกอบ 15

ตาราง 19 Hyperparameters และค่า MAE และ RMSE ในแบบจำลองการการถดถอยของกลุ่ม Binary Class C (N : L, M, H)

Model Type	Hyperparameters	MAE	RMSE
XGBoost Regressor	Learning Rate: 0.01 Max Depth: 4 n_estimators: 300	0.31	0.36
Support Vector Regressor (SVR)	C: 0.1 Epsilon: 0.3 Gamma: scale	0.34	0.40
Regression Neural Network	Epochs=100 Batch Size=50 Verbose=1	0.28	0.37



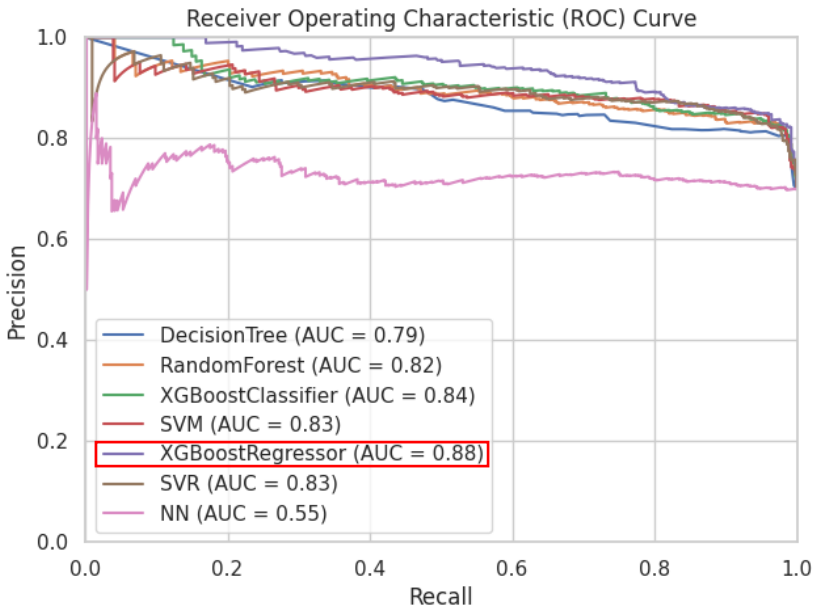
(a) แสดงค่าการทำนายของแบบจำลอง XGBoost Regressor (b) แสดงค่าการทำนายของแบบจำลอง Support Vector Regressor



(c) แสดงค่าการทำนายของแบบจำลอง Regression Neural Network

ภาพประกอบ 15 Boxplot แสดงค่าการทำนายในแต่ละระดับความเสี่ยงเอชไอวีของกลุ่ม Binary Class C (N : L, M, H)

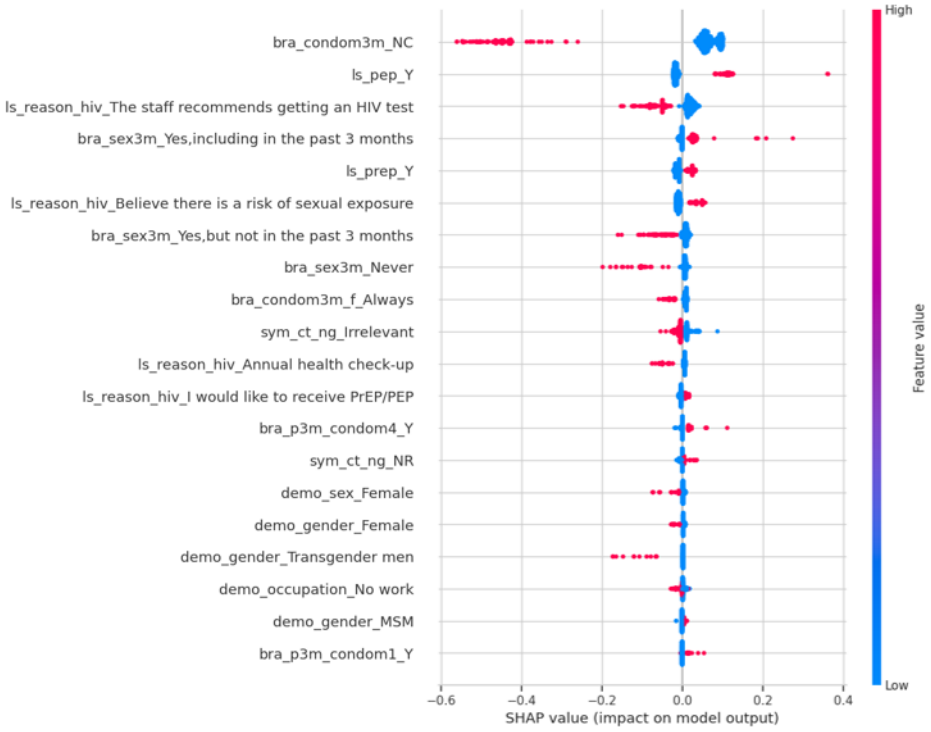
ภาพประกอบ 16 ผู้วิจัยแสดงกราฟ ROC Curve (Receiver Operating Characteristic Curve) ในกลุ่ม Binary Class C (N : L, M, H) เพื่อแสดงแบบจำลองที่มีประสิทธิภาพในการจำแนกและคัดแยก Positive Class และ Negative Class ได้อย่างเหมาะสม จะสังเกตได้ว่า XGBoost Regressor นั้น มีค่า AUC สูงที่สุดที่ 0.88 ซึ่งแบบแบบจำลองที่ดีที่สุดในกลุ่ม Binary Class C



ภาพประกอบ 16 ROC Curve เปรียบเทียบแบบจำลองของกลุ่ม Binary Class C (N : L, M, H)

4.4 ประเมินความสำคัญของตัวแปร (Feature Importances)

จากผลการทดลองทุกกลุ่มคลาสแบบจำลองที่ดีที่สุดในการประเมินความเสี่ยงคือแบบจำลองการถดถอย XGBoost Regressor ที่มีค่า AUC สูงสุดที่ 0.88 และอยู่ในกลุ่มที่เป็นกลุ่มเป้าหมายการคัดกรองในงานวิจัยนี้ดังนั้นผู้วิจัยจึงประเมินความสำคัญของตัวแปรเพื่อดูว่าตัวแปรไหนที่ให้ความสำคัญแบบใดกับแบบจำลองด้วยวิธี SHAP Value (Awan, 2023) ดังภาพประกอบ 17



ภาพประกอบ 17 ประเมินความสำคัญของตัวแปร ด้วยวิธี SHAP Value

จากภาพตัวแปรที่มีความสำคัญ 5 อันดับแรกคือ ไม่เคยมีเพศสัมพันธ์ ใน 3 เดือน, การรับยาเป็ป, เจ้าหน้าที่แนะนำให้ไปตรวจเอชไอวี, การมีเพศสัมพันธ์รวมถึงในช่วง 3 เดือนที่ผ่านมา, การรับยาเพร็พ (PrEP) ตามลำดับ

บทที่ 5
สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

5.1 สรุปผลการวิจัย

งานวิจัยนี้มุ่งเน้นประเมินความเสี่ยงการติดเชื้อเอชไอวี โดยแบ่งออกเป็น 4 ระดับ คือ ไม่มีความเสี่ยง (No Risk (N)), ความเสี่ยงต่ำ (Low Risk (L)), ความเสี่ยงปานกลาง (Medium Risk (M)) และความเสี่ยงสูง (High Risk (H)) โดยมีการสร้างแบบจำลองเพื่อทำนายความเสี่ยงการติดเชื้อเอชไอวีเพื่อทำนายความเสี่ยงทั้ง 4 ระดับ และการสร้างแบบจำลองเพื่อทำนายความเสี่ยงการติดเชื้อเอชไอวีเพื่อทำนายความเสี่ยง 2 ระดับ โดยทำการจัดกลุ่มระดับความเสี่ยงใหม่เพื่อให้เหมาะสมในการนำไปใช้งานที่แตกต่างกัน ซึ่งสามารถสรุปออกเป็นตาราง 20 ของแบบจำลองการจำแนกประเภท และตาราง 21 ของแบบจำลองการถดถอย ดังนี้

ตาราง 20 สรุปผลลัพธ์การทำนายแบบจำลองการจำแนกประเภทในแต่ละกลุ่มคลาส

Group Class	Model Type	Class	Precision	Recall	F1-Score
Multi-Target	Decision Tree	0 (N)	0.62	0.64	0.63
		1 (L)	0.55	0.42	0.48
		2 (M)	0.16	0.20	0.18
		3 (H)	0.20	0.33	0.25
	Random Forest	0 (N)	0.75	0.58	0.65
		1 (L)	0.57	0.52	0.54
		2 (M)	0.23	0.27	0.25
		3 (H)	0.25	0.45	0.32
	XGBoost Classifier	0 (N)	0.70	0.61	0.65
		1 (L)	0.59	0.48	0.53
		2 (M)	0.18	0.19	0.18
		3 (H)	0.26	0.55	0.35

ตาราง 20 (ต่อ)

Group Class	Model Type	Class	Precision	Recall	F1-Score
	Support Vector Machine	0 (N)	0.67	0.60	0.63
		1 (L)	0.56	0.40	0.47
		2 (M)	0.20	0.26	0.23
		3 (H)	0.28	0.63	0.39
Binary Class A (N, L, M : H)	Decision Tree	0 (N,L,M)	0.93	0.76	0.84
		1 (H)	0.16	0.43	0.23
	Random Forest	0 (N,L,M)	0.94	0.83	0.88
		1 (H)	0.23	0.51	0.32
	XGBoost Classifier	0 (N,L,M)	0.96	0.76	0.85
		1 (H)	0.23	0.70	0.35
	Support Vector Machine	0 (N,L,M)	0.97	0.72	0.82
		1 (H)	0.21	0.75	0.33
Binary Class B (N, L : M, H)	Decision Tree	0 (N,L)	0.78	0.63	0.70
		1 (M,H)	0.32	0.49	0.39
	Random Forest	0 (N,L)	0.83	0.65	0.73
		1 (M,H)	0.39	0.63	0.48
	XGBoost Classifier	0 (N,L)	0.83	0.64	0.72
		1 (M,H)	0.38	0.63	0.47
	Support Vector Machine	0 (N,L)	0.85	0.57	0.68
		1 (M,H)	0.37	0.71	0.48

ตาราง 20 (ต่อ)

Group Class	Model Type	Class	Precision	Recall	F1-Score
Binary Class C (N : L, M, H)	Decision Tree	0 (N)	0.53	0.65	0.59
		1 (L,M,H)	0.83	0.76	0.79
	Random Forest	0 (N)	0.73	0.61	0.67
		1 (L,M,H)	0.84	0.91	0.87
	XGBoost Classifier	0 (N)	0.64	0.68	0.66
		1 (L,M,H)	0.86	0.83	0.85
	Support Vector Machine	0 (N)	0.64	0.71	0.67
		1 (L,M,H)	0.87	0.83	0.85

ตาราง 21 ตารางสรุปผลลัพธ์การทำนายแบบจำลองการถดถอยในแต่ละกลุ่มคลาส

Group Class	Model	MAE	RMSE
Multi-Target	XGBoost Regressor	0.67	0.84
	Support Vector Regressor	0.67	0.84
	Regression Neural Network	0.61	0.79
Binary Class A (N, L, M : H)	XGBoost Regressor	0.34	0.40
	Support Vector Regressor	0.37	0.44
	Regression Neural Network	0.26	0.33
Binary Class B (N, L : M, H)	XGBoost Regressor	0.38	0.44
	Support Vector Regressor	0.41	0.47
	Regression Neural Network	0.36	0.43
Binary Class C (N : L, M, H)	XGBoost Regressor	0.31	0.36
	Support Vector Regressor	0.34	0.40
	Regression Neural Network	0.28	0.37

จากนั้นผู้วิจัยได้เปรียบเทียบแต่ละแบบจำลองของแต่ละกลุ่มคลาสด้วยการแสดงกราฟ ROC Curve ดังตาราง 22

ตาราง 22 ค่า AUC ในแต่ละแบบจำลองในแต่ละกลุ่มคลาส

Group Class	Model type	AUC
Binary Class A (N, L, M : H)	Decision Tree	0.68
	Random Forest	0.79
	XGBoost Classifier	0.81
	Support Vector Machine	0.80
	XGBoost Regressor	0.77
	Support Vector Regressor	0.83
	Regression Neural Network	0.81
Binary Class B (N, L : M, H)	Decision Tree	0.61
	Random Forest	0.73
	XGBoost Classifier	0.72
	Support Vector Machine	0.72
	XGBoost Regressor	0.81
	Support Vector Regressor	0.72
	Regression Neural Network	0.54
Binary Class C (N : L, M, H)	Decision Tree	0.79
	Random Forest	0.82
	XGBoost Classifier	0.84
	Support Vector Machine	0.83
	XGBoost Regressor	0.88
	Support Vector Regressor	0.83
	Regression Neural Network	0.55

ต่อมาผู้วิจัยได้ประเมินความสำคัญของตัวแปรด้วยวิธี SHAP Value ซึ่งประเมินกับแบบจำลองการถดถอย XGBoost Regressor ซึ่งเป็นแบบจำลองที่ดีที่สุดใน Binary Class C (N : L, M, H) จึงสรุปผล 20 อันดับแรกซึ่งประกอบด้วยตัวแปรที่มีความสำคัญในคลาส 0 และตัวแปรที่มีความสำคัญในคลาส 1 ดังตาราง 23

ตาราง 23 สรุปผลการประเมินความสำคัญของตัวแปร

ตัวแปรที่มีความสำคัญในคลาส 0	**กลุ่มคุณลักษณะ**	**ตัวแปรที่มีความสำคัญในคลาส 1**	**กลุ่มคุณลักษณะ**
1. ไม่เคยมีเพศสัมพันธ์ใน 3 เดือน	พฤติกรรมทางเพศ	1. การรับยาเป็ป	รูปแบบการดำเนินชีวิต
2. เจ้าหน้าที่แนะนำให้ไปตรวจเอชไอวี	รูปแบบการดำเนินชีวิต	2. เคยมีเพศสัมพันธ์แบบสอดใส่ในช่วง 3 เดือนที่ผ่านมา	พฤติกรรมทางเพศ
3. เคยมีเพศสัมพันธ์แต่ไม่ใช่ในช่วง 3 เดือนที่ผ่านมา	พฤติกรรมทางเพศ	3. การรับยาเพร็พ	รูปแบบการดำเนินชีวิต
4. ไม่เคยมีเพศสัมพันธ์แบบสอดใส่	พฤติกรรมทางเพศ	4. เชื่อว่ามีความเสี่ยงต่อการมีเพศสัมพันธ์จึงตรวจเอชไอวี	รูปแบบการดำเนินชีวิต
5. ในช่วงสามเดือนที่ผ่านมาใช้ถุงยางทุกครั้ง	พฤติกรรมทางเพศ	5. ฉันต้องการรับยา PrEP/PEP จึงตรวจเอชไอวี	รูปแบบการดำเนินชีวิต
6. ไม่ตรวจโรคหนองในหนองในเทียม	อาการอื่นๆ	6. มีการสอดใส่เมื่อมีเพศสัมพันธ์ แต่ไม่เคยเกิดเหตุการณ์ใน 3 ข้อ (ถุงยางแตก, ถุงยางหลุด, คู่นอนถอดถุงยาง) ที่กล่าวมา	พฤติกรรมทางเพศ

ตาราง 23 (ต่อ)

ตัวแปรที่มีความสำคัญในคลาส 0	กลุ่มคุณลักษณะ	ตัวแปรที่มีความสำคัญในคลาส 1	กลุ่มคุณลักษณะ
7. ตรวจสุขภาพประจำปี	รูปแบบการดำเนินชีวิต	7. ไม่เป็นโรคหนองในหรือ หนองในเทียม	อาการอื่นๆ
8. เพศ หญิง	ข้อมูลทางด้านประชากร	8. อัตลักษณ์ ชายผู้มีเพศสัมพันธ์กับชาย	ข้อมูลทางด้านประชากร
9. อัตลักษณ์ หญิง	ข้อมูลทางด้านประชากร	9. ถุงยางแตก / ถุงยางรั่ว	พฤติกรรมทางเพศ
10. อัตลักษณ์ ชายข้ามเพศ	ข้อมูลทางด้านประชากร		
11. ไม่มีอาชีพ	ข้อมูลทางด้านประชากร		

5.2 อภิปรายผล

จากการสรุปผลการทดลองพบว่าแบบจำลองการถดถอยมักจะให้ผลลัพธ์ที่ดีในการจำแนกคลาสในทุกกลุ่มซึ่งผู้วิจัยคาดว่าเพราะแบบจำลองการถดถอยมักจะยืดหยุ่นกว่าในการจัดการกับข้อมูลที่มี Noise เนื่องจากโดยทั่วไปแล้วแบบจำลองการถดถอยมักมีการใช้งานตัวควบคุมการเรียนรู้ (Regularizes) เพื่อลดผลกระทบจากข้อมูลที่มี Noise โดยจากสรุปผลการทดลอง Binary Class A (N, L, M : H) มีแบบจำลองที่ดีที่สุดคือ Support Vector Regressor ที่ AUC 0.83, Binary Class B (N, L : M, H) มีแบบจำลองที่ดีที่สุดคือ XGBoost Regressor ที่ AUC 0.81 และ Binary Class C (N : L, M, H) มีแบบจำลองที่ดีที่สุดคือ XGBoost Regressor ที่ AUC 0.88 ในขณะที่ Regression Neural Network มีประสิทธิภาพต่ำสุดในทุกกลุ่มด้วย AUC ต่ำที่สุด ซึ่งแบบจำลองที่มีค่า AUC สูงสุดที่ 0.88 คือแบบจำลอง XGBoost Regressor ในกลุ่ม Binary Class C (N : L, M, H) ซึ่งอยู่ในกลุ่มที่ในงานวิจัยนี้ให้ความสนใจเป็นอันดับแรก เนื่องจากเป็นการเน้นกลุ่มที่มีความเสี่ยงเพื่อไม่ให้พลาดการทำนายผู้ป่วยที่มีความเสี่ยงที่จะติดเชื้อเอชไอวี

การเลือกกลุ่มคลาสเพื่อนำไปใช้จึงจะขึ้นอยู่กับวัตถุประสงค์ที่ต้องการนำไปใช้ ในงานวิจัยนี้ผู้วิจัยต้องการประเมินเบื้องต้นสำหรับผู้ที่คิดว่ามีความเสี่ยงที่จะติดเชื้อเอชไอวี และพิจารณาการรับยา PrEP หรือ PEP สำหรับผู้ที่มีความเสี่ยง เพื่อเป็นการป้องกันการติดเชื้อเอชไอวี จึงต้องการให้แบบจำลองทำนายในกลุ่มของผู้มีความเสี่ยงได้แม่นยำ มากกว่ากลุ่มที่ไม่มีความ

เสี่ยง แต่สำหรับกลุ่มที่เน้นผู้ที่ไม่มีความเสี่ยง หรือความเสี่ยงต่ำ อาจจะถูกนำไปใช้เพื่อแก้ปัญหาอื่นๆได้เช่นกัน เช่น สำหรับผู้ที่จะเข้าโครงการวิจัยด้านสุขภาพ ซึ่งมีเงื่อนไขการเข้าโครงการเป็นผู้รับบริการที่ไม่มีความเสี่ยง เป็นต้น

คุณลักษณะที่มีผลต่อแบบจำลองเพื่อประเมินความเสี่ยงเอชไอวีใน 3 อันดับแรกคือ การรับยาเป็ป, เคยมีเพศสัมพันธ์แบบสอดใส่ในช่วง 3 เดือนที่ผ่านมา และการรับยาเพร็พ ซึ่งเป็นตัวแปรที่ทำให้เกิดความเสี่ยง และตัวแปรที่ไม่ทำให้เกิดความเสี่ยงใน 3 อันดับแรกคือ ไม่เคยมีเพศสัมพันธ์ ใน 3 เดือน, เจ้าหน้าที่แนะนำให้ไปตรวจเอชไอวี และ เคยมีเพศสัมพันธ์แต่ไม่ใช่ในช่วง 3 เดือนที่ผ่านมา จากที่กล่าวมานั้นการรับยาเป็ปและเพร็พนั้นเป็นการป้องกันการติดเชื้อเอชไอวีจึงมีผลมากที่ทำให้เกิดความเสี่ยงเนื่องจากผู้ที่รับบริการเมื่อมีความเสี่ยงจึงรับยาเพื่อป้องกันไว้ก่อน แต่การเคยมีเพศสัมพันธ์แบบสอดใส่ในช่วง 3 เดือนที่ผ่านมานั้นเป็นความเสี่ยงที่ผู้เข้ารับบริการสามารถหลีกเลี่ยงได้ และยังมีถุงยางแตก / ถุงยางรั่ว ซึ่งเป็นหนึ่งในตัวแปรที่ทำให้เกิดความเสี่ยงที่ผู้รับบริการควรหลีกเลี่ยงเช่นกัน

5.3 **ข้อเสนอแนะ**

งานวิจัยนี้เป็นการศึกษาการประเมินความเสี่ยงการติดเชื้อเอชไอวีเพื่อให้ผู้ที่มีความเสี่ยงตระหนักรู้และไม่เกิดความประมาทในการแพร่เชื้อต่อผู้อื่น โดยศึกษากับแบบจำลองการจำแนกประเภท และแบบจำลองการถดถอย ซึ่งข้อเสนอแนะเพื่อเพิ่มประสิทธิภาพในการประเมินความเสี่ยงในอนาคต สามารถพัฒนาและศึกษาเทคนิคการเรียนรู้ของเครื่อง และวิธีการจัดกลุ่มข้อมูลแบบอื่น ๆ ที่มีการนำมาใช้ในการประเมินความเสี่ยงการติดเชื้อเอชไอวี ตัวอย่างเช่น การใช้ข้อมูลผู้ให้คำปรึกษา ศึกษาวิธีการนำข้อมูลจากผู้ให้คำปรึกษาหรือผู้ที่มีประสบการณ์เกี่ยวกับการป้องกันเอชไอวี เข้ามาในการประเมินความเสี่ยง ซึ่งอาจจะช่วยให้มีการสร้างแบบจำลองที่สามารถรับฟังข้อมูลและแนะนำวิธีการป้องกันที่เหมาะสมตามบุคคลและสถานการณ์

บรรณานุกรม

Ahlström, M. G., Ronit, A., Omland, L. H., Vedel, S., & Obel, N. (2019). Algorithmic prediction of HIV status using nation-wide electronic registry data. *EClinicalMedicine*, 17, 100203. https://doi.org/10.1016/j.eclinm.2019.10.016

Awan, A. A. (2023). *An Introduction to SHAP Values and Machine Learning Interpretability*. https://www.datacamp.com/tutorial/introduction-to-shap-values-machine-learning-interpretability

Eakasit Pacharawongsakda, P. D. (2015). การคัดเลือก Feature selection. https://th.linkedin.com/pulse/%E0%B8%81%E0%B8%B2%E0%B8%A3%E0%B8%84%E0%B8%94%E0%B9%80%E0%B8%A5%E0%B8%AD%E0%B8%81-feature-selection-%E0%B8%94%E0%B8%A7%E0%B8%A2%E0%B8%A7%E0%B8%98-information-gain-pacharawongsakda

files/22/การคดเลอก-feature-selection-ดวยวธ-information-gain-pacharawongsakda.html

Frescura, L., Godfrey-Faussett, P., Feizzadeh A, A., El-Sadr, W., Syarif, O., Ghys, P. D., on, & behalf of the testing treatment target Working, G. (2022). Achieving the 95 95 95 targets for all: A pathway to ending AIDS. *PloS One*, 17(8), e0272405. https://doi.org/10.1371/journal.pone.0272405

Keita, Z. (2022). *Classification in Machine Learning: An Introduction*. https://www.datacamp.com/blog/classification-machine-learning

Khongsra, L. (2019). เพร็พ (PREP) ,เป็ป (PEP). ศูนย์วิจัยโรคเอดส์ สภากาชาดไทย (คลีนิคนิรนาม). https://th.trcarc.org/เพร็พ-prep-เป็ป-pep-คืออะไร/

files/12/เพร็พ-prep-เป็ป-pep-คืออะไร.html

Majam, M., Segal, B., Fieggen, J., Smith, E., Hermans, L., Singh, L., Phatsoane, M., Arora, L., & Lalla-Edward, S. T. (2023). Utility of a machine-guided tool for assessing risk behaviour associated with contracting HIV in three sites in South Africa. *Informatics in Medicine Unlocked*, 37, 101192. https://doi.org/10.1016/j.imu.2023.101192

Muhimpundu, L., & Dr. Pierre Claver, R. Prediction of HIV infections among individuals with sexual risk behaviours in Rwanda using machine learning algorithms. http://dr.ur.ac.rw/bitstream/handle/123456789/1988/Muhimpundu%20%20Lorra

ine.pdf?sequence=1&isAllowed=y

Promrit, N. (2020). การเลือกใช้ Loss Function ในการพัฒนา Deep Learning Model (ตอนที่ 1). *PJJOP*. https://blog.pjjop.org/loss-functions-for-training-deep-learning-model-part1/

files/42/loss-functions-for-training-deep-learning-model-part1.html

Visitsora-at, P. (2019). Metrics พื้นฐานสำหรับวัดประสิทธิภาพของโมเดล Machine Learning. *Medium*. https://medium.com/@615162020027/metrics-%E0%B8%9E%E0%B8%B7%E0%B9%89%E0%B8%99%E0%B8%90%E0%B8%B2%E0%B8%99%E0%B8%AA%E0%B8%B3%E0%B8%AB%E0%B8%A3%E0%B8%B1%E0%B8%9A%E0%B8%A7%E0%B8%B1%E0%B8%94%E0%B8%9B%E0%B8%A3%E0%B8%B0%E0%B8%AA%E0%B8%B4%E0%B8%97%E0%B8%98%E0%B8%B4%E0%B8%A0%E0%B8%B2%E0%B8%9E%E0%B8%82%E0%B8%AD%E0%B8%87%E0%B9%82%E0%B8%A1%E0%B9%80%E0%B8%94%E0%B8%A5-machine-learning-c00fcc32fa30

files/40/metrics-พื้นฐานสำหรับวัดประสิทธิภาพขอ.html

กรมควบคุมโรค. (2022). สถานการณ์เอชไอวีประเทศไทย ปี 2565. In.

กีระสุนทรพงษ์, น. อ. (2016). ความรู้เกี่ยวกับเอชไอวีและเอดส์. https://www.bumrungrad.com/th/health-blog/november-2016/hiv-aids-infection-treatment

files/2/hiv-aids-infection-treatment.html

มอร์ด, พ. ร. ร. (2023). โรคเอดส์ (HIV/AIDS). *MedPark Hospital*. http://www.medparkhospital.com/disease-and-treatment/hiv-aids

files/10/hiv-aids.html

ยงค์เจริญชัย, ช. (2020). เอดส์: 6 ปีผ่านไป ไทยใช้ "ยาเพร็ป" ป้องกันเอชไอวีได้ดีแค่ไหน. *BBC News* ไทย. https://www.bbc.com/thai/thailand-55141812

files/18/thailand-55141812.html

ประวัติผู้เขียน