



Development of Risk Predictive Model for Heart Attack in Elderly Using Data Mining Technique

¹ Mongkol Rodjan, ^{2*} Paranya Palwisut & ² Apinan Junkorn

¹Department of Computer Technology, Nakhon Pathom Rajabhat University, Nakhon Pathom.

²Department of Data Science, Faculty of Science and Technology, Nakhon Pathom
Rajabhat University, Nakhon Pathom.

*Corresponding author: paranya@npru.ac.th

Received
13/05/2024

Reviewed
21/06/2024

Revised
10/12/2024

Accepted
12/12/2024

Abstract

The objectives of this research are 1) to compare risk predictive model performance for heart attack in the elderly using data mining techniques, and 2) to develop a prototype system for predicting the risk of heart attack in the elderly, using the heart attack dataset in the Heart Attack Analysis & Prediction Dataset. This research used the Scikit-Learn Library in Python to create the models. Three classifiers in machine learning, namely Support Vector Machine, Naive Bayes, and Decision Tree were used for comparison. The performance of the models was evaluated using accuracy, precision, recall, and F-measure. The research found that Naive Bayes had the highest performance with an accuracy of 86.05%, precision of 87.44%, recall of 85.82%, and F-measure of 85.86%. Naive Bayes was then used to develop a prototype system for predicting heart attacks in the elderly through web application, to Predict whether the chance of having a heart attack is more or less.

Keywords: Heart Attack; Support Vector Machine; Naive Bayes; Decision Tree



การพัฒนาตัวแบบทำนายความเสี่ยงการเกิดภาวะหัวใจขาดเลือดเฉียบพลัน ในผู้สูงอายุด้วยเทคนิคเหมืองข้อมูล

¹มงคล รอดจันทร์, ^{2*}ภรณ์ยา ปาลวิสุทธิ และ²ภินันท์ จันทกรณ

¹สาขาวิชาเทคโนโลยีคอมพิวเตอร์, มหาวิทยาลัยราชภัฏนครปฐม จังหวัดนครปฐม

²สาขาวิชาวิทยาการข้อมูล, คณะวิทยาศาสตร์และเทคโนโลยี, มหาวิทยาลัยราชภัฏนครปฐม
จังหวัดนครปฐม

*ผู้นิพนธ์หลัก: paranya@npru.ac.th

บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์ คือ 1) เพื่อเปรียบเทียบประสิทธิภาพของตัวแบบทำนายความเสี่ยงการเกิดภาวะหัวใจขาดเลือดเฉียบพลันในผู้สูงอายุด้วยเทคนิคเหมืองข้อมูล และ 2) เพื่อพัฒนาระบบตัวแบบทำนายความเสี่ยงการเกิดภาวะหัวใจขาดเลือดเฉียบพลันในผู้สูงอายุ โดยใช้ชุดข้อมูลหัวใจขาดเลือดเฉียบพลัน ในชุดของ Heart Attack Analysis & Prediction Dataset เครื่องมือในการวิจัยครั้งนี้ใช้ Scikit-Learn Library ของภาษา Python ในการสร้างตัวแบบ เทคนิคเหมืองข้อมูลที่นำมาใช้เปรียบเทียบในงานวิจัย จำนวน 3 เทคนิค ได้แก่ ซัพพอร์ตเวกเตอร์แมชชีน นาอ์ฟเบย์ และต้นไม้ตัดสินใจ และได้ใช้ค่าความถูกต้อง ค่าความแม่นยำ ค่าความระลึก และค่าความถ่วงดุล ในการวัดประสิทธิภาพการพยากรณ์ของตัวแบบ ผลการวิจัยพบว่า นาอ์ฟเบย์ มีประสิทธิภาพสูงที่สุดโดยให้ค่าความถูกต้อง เท่ากับร้อยละ 86.05 ค่าความแม่นยำเท่ากับร้อยละ 87.44 ค่าความระลึกเท่ากับร้อยละ 85.82 และค่าความถ่วงดุลเท่ากับร้อยละ 85.86 และได้นำตัวแบบที่ได้จากนาอ์ฟเบย์ที่มีประสิทธิภาพสูงที่สุดมาพัฒนาระบบตัวแบบทำนายการเกิดภาวะหัวใจขาดเลือดเฉียบพลันในผู้สูงอายุที่พัฒนาในรูปแบบเว็บแอปพลิเคชันให้ทำนายว่าโอกาสการเกิดภาวะหัวใจขาดเลือดเฉียบพลันมีมากหรือน้อย

คำสำคัญ: หัวใจขาดเลือดเฉียบพลัน; ซัพพอร์ตเวกเตอร์แมชชีน; นาอ์ฟเบย์; ต้นไม้ตัดสินใจ



บทนำ

หัวใจขาดเลือดเฉียบพลัน (Heart Attack) (Suesat, 2020) เกิดจากหลอดเลือดหัวใจอุดตันหรือตีบกะทันหัน กล้ามเนื้อหัวใจตายหรือขาดเลือดรุนแรง หัวใจเต้นผิดจังหวะ หรือตกอยู่ในภาวะหัวใจล้มเหลวจนทำให้หัวใจหยุดเต้น ดังนั้นการรู้ทันโรคเป็นเรื่องสำคัญเพื่อช่วยป้องกันและลดโอกาสการเสียชีวิต ซึ่งกลุ่มเสี่ยงที่จะเกิดภาวะหัวใจขาดเลือดเฉียบพลัน ได้แก่ ผู้มีอายุ 50 ปีขึ้นไป มีภาวะหลอดเลือดหัวใจผิดปกติ มีโรคประจำตัว ได้แก่ เบาหวาน ความดันโลหิตสูง ไขมันในเลือดสูง มีประวัติครอบครัวเป็นโรคหลอดเลือดหัวใจ สุกุบหรี มีความเครียดสูง มีน้ำหนักเกินเกณฑ์ ไม่ออกกำลังกาย และผู้สูบบุหรี่เป็นประจำเดือน

ข้อมูลจากองค์การอนามัยโลก (WHO) ในปี 2563 ได้ระบุว่าโรคหัวใจและหลอดเลือดเป็นสาเหตุการเสียชีวิตอันดับ 1 ของโลก (Kaewna, 2023) หัวใจเสียชีวิตด้วยโรคหัวใจและหลอดเลือดสมองมากกว่า 20 ล้านคน และ 80% ของการเสียชีวิตสามารถป้องกันได้ สำหรับประเทศไทยจากรายงานของกระทรวงสาธารณสุข ปี 2565 พบการเสียชีวิตของคนไทยด้วยโรคหัวใจและหลอดเลือดมากถึง 7 หมื่นรายเฉลี่ยชั่วโมงละ 8 คน และคาดว่าจะมีแนวโน้มเพิ่มขึ้นทุกปี และจากข้อมูลกลุ่มเสี่ยงจะเห็นได้ว่าผู้สูงอายุเป็นกลุ่มที่เสี่ยงต่อการเกิดภาวะหัวใจขาดเลือดเฉียบพลัน

การทำเหมืองข้อมูล (Data Mining) (Sinsomboonthong, 2017) กระบวนการที่กระทำกับข้อมูลจำนวนมากเพื่อค้นหารูปแบบและความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น ในปัจจุบันการทำเหมืองข้อมูลได้ถูกนำไปประยุกต์ใช้ในงานหลายประเภท ทั้งในด้านธุรกิจที่ช่วยในการตัดสินใจของผู้บริหาร ในด้านวิทยาศาสตร์และการแพทย์รวมทั้งในด้านเศรษฐกิจและสังคม ได้มีงานวิจัยด้านการแพทย์ที่นำเหมืองข้อมูลมาประยุกต์ใช้เป็นจำนวนมาก เช่น ตัวอย่างงานวิจัยของ Panyatip (2022) ได้ทำวิจัยเรื่องการเพิ่มประสิทธิภาพจำแนกข้อมูลผลกระทบโควิด-19 ต่อผู้ป่วยมะเร็งตับ โดยเปรียบเทียบอัลกอริทึมการจำแนกข้อมูลที่เพิ่มประสิทธิภาพเพื่อหาอัลกอริทึมที่ดีที่สุด ได้แก่ ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine) นาอิวเบย์ (Naive Baye) เพื่อนบ้านใกล้ที่สุด (K-nearest Neighbors) ต้นไม้ตัดสินใจ (Decision Tree) และโครงข่ายประสาทเทียม (Artificial neural networks) Tosasukul (2021) ได้ทำการวิจัยเรื่อง แบบจำลองการพยากรณ์การระบาดของโรคไข้เลือดออกโดยใช้เทคนิคการทำเหมืองข้อมูล 4 วิธี ได้แก่ วิธีต้นไม้ตัดสินใจ วิธีแบบเบย์ วิธีโครงข่ายประสาท และวิธีซัพพอร์ตเวกเตอร์แมชชีน พบว่า วิธีต้นไม้ตัดสินใจเป็นเทคนิคในการสร้างแบบจำลองสำหรับการพยากรณ์การระบาดของโรคไข้เลือดออกที่เหมาะสมที่สุด Mannaisatjatham (2022) ได้ทำการวิจัยเรื่อง การจำลองเปรียบเทียบการสำรองยารักษาโรคมะเร็งโดยคลังยาด้วยระบบการเติมสินค้าโดยผู้ขายผ่านการจำแนกข้อมูลด้วยต้นไม้ตัดสินใจในสถาบันมะเร็งแห่งชาติเพื่อปรับปรุงกระบวนการทำงานและลดค่าใช้จ่ายในการสำรองยาคลังที่ไม่จำเป็น Petwised (2023) ได้วิจัยการพัฒนาตัวแบบทำนายสัญญาณเตือนโรคมะเร็งลำไส้ใหญ่และทวารหนัก ด้วยวิธีการจำแนกกลุ่มมี 5 อัลกอริทึม ได้แก่ นาอิวเบย์ ต้นไม้ตัดสินใจ ปาสุ่ม



(Random Forest) ดีพเลิร์นนิง (Deep Learning) และ การถดถอยโลจิสติก (Logistic Regression Analysis) ผลการวิจัยพบว่า อัลกอริทึมการถดถอยโลจิสติกมีประสิทธิภาพในการทำนายผลดีที่สุด

ดังนั้นบทความวิจัยนี้วัตถุประสงค์หลักคือเพื่อเปรียบเทียบประสิทธิภาพของอัลกอริทึมในการทำนาย 3 อัลกอริทึม ได้แก่ ซัพพอร์ตเวกเตอร์แมชชีน นาอิวเบย์ และต้นไม้ตัดสินใจ ซึ่งเป็นเทคนิคทางเหมืองข้อมูลที่มีประสิทธิภาพบนชุดข้อมูลทางด้านสุขภาพจากการวิจัยที่ผ่านมาในการทำนายความเสี่ยงการเกิดภาวะหัวใจขาดเลือดเฉียบพลันในผู้สูงอายุ เมื่อได้ผลการเปรียบเทียบประสิทธิภาพของอัลกอริทึมทั้ง 3 ในด้านค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าความถ่วงดุล (F-Measure) ที่เหมาะสมที่สุดในการทำนายความเสี่ยงการเกิดภาวะหัวใจขาดเลือดเฉียบพลันในผู้สูงอายุ จากนั้นนำตัวแบบที่ได้จากอัลกอริทึมที่มีประสิทธิภาพที่สุดมาพัฒนาระบบทำนายความเสี่ยงการเกิดภาวะหัวใจขาดเลือดเฉียบพลันในผู้สูงอายุ เพื่อสนับสนุนผู้เชี่ยวชาญทางการแพทย์ในการเลือกวิธีการที่เหมาะสมที่สุดในการป้องกันการเกิดภาวะหัวใจขาดเลือดเฉียบพลันในผู้สูงอายุอย่างมีประสิทธิภาพและแม่นยำยิ่งขึ้น

วัตถุประสงค์การวิจัย

1. เพื่อเปรียบเทียบประสิทธิภาพของตัวแบบทำนายความเสี่ยงการเกิดภาวะหัวใจขาดเลือดเฉียบพลันในผู้สูงอายุใช้เทคนิคซัพพอร์ตเวกเตอร์แมชชีน นาอิวเบย์ และต้นไม้ตัดสินใจ
2. เพื่อพัฒนาระบบตัวแบบทำนายความเสี่ยงการเกิดภาวะหัวใจขาดเลือดเฉียบพลันในผู้สูงอายุ

วิธีการวิจัย

การพัฒนาตัวแบบทำนายความเสี่ยงการเกิดภาวะหัวใจขาดเลือดเฉียบพลันในผู้สูงอายุด้วยเทคนิคเหมืองข้อมูล โดยได้ดำเนินการตามกระบวนการ CRISP-DM (Cross-Industry Standard Process for Data Mining) เป็นกระบวนการมาตรฐานในการวิเคราะห์ข้อมูลด้านเทคนิคเหมืองข้อมูล ดังนี้

1. การทำความเข้าใจปัญหา (Business Understanding)

คณะผู้วิจัยได้ทำการศึกษาเบื้องต้นเพื่อทำความเข้าใจปัญหาที่จะทำ ได้ทำการศึกษาเอกสารความรู้งานวิจัยต่าง ๆ ที่เกี่ยวข้องกับการสร้างตัวแบบทำนายความเสี่ยงการเกิดภาวะหัวใจขาดเลือดเฉียบพลันในผู้สูงอายุทั้งในส่วนของเทคนิคที่ใช้ในการวิเคราะห์ข้อมูล โดยมีรายละเอียด ดังต่อไปนี้

1.1 ศึกษาปัญหา

งานวิจัยนี้วัตถุประสงค์หลักคือเพื่อเปรียบเทียบประสิทธิภาพของตัวแบบในการทำนายความเสี่ยงการเกิดภาวะหัวใจขาดเลือดเฉียบพลันในผู้สูงอายุจำนวน 3 อัลกอริทึม ได้แก่ ซัพพอร์ตเวกเตอร์แมชชีน นาอ์ฟเบย์ และต้นไม้ตัดสินใจ และเพื่อพัฒนาระบบตัวแบบทำนายความเสี่ยงการเกิดภาวะหัวใจขาดเลือดเฉียบพลันในผู้สูงอายุ โดยทำการเปรียบเทียบประสิทธิภาพของอัลกอริทึม 4 ด้าน ได้แก่ ด้านความถูกต้อง ด้านความแม่นยำ ด้านความระลึก และด้านความถ่วงดุล ที่เหมาะสมที่สุดในการทำนายความเสี่ยงการเกิดภาวะหัวใจขาดเลือดเฉียบพลันในผู้สูงอายุจากนั้นจึงนำไปพัฒนาระบบ สรุปกรอบแนวคิดดัง Figure 1

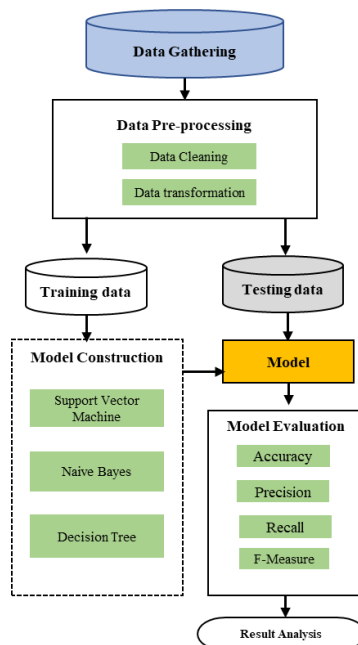


Figure 1 Conceptual framework

1.2 ทฤษฎีที่เกี่ยวข้อง

1) เหมืองข้อมูล

การทำเหมืองข้อมูล (Rungtattanabool, 2023) เป็นหนึ่งในความรู้สมัยใหม่ที่มีความนิยมอย่างมากในปัจจุบัน โดยได้ถูกประยุกต์ใช้ในแทบทุกองค์กร ไม่ว่าจะเป็นภาครัฐและเอกชนเพื่อใช้ในการตัดสินใจ การวางแผนกลยุทธ์ การปรับปรุงการให้บริการและการดำเนินงานต่าง ๆ ในองค์กร เหมืองข้อมูล เป็นการทำงานที่เน้นการค้นหาสารสนเทศหรือองค์ความรู้จากข้อมูลขนาดใหญ่ เพื่อนำสิ่ง

ที่ได้มาใช้ให้เป็นประโยชน์ โดยเหมืองข้อมูลเป็นการผสมผสานศาสตร์ทางสถิติ ปัญญาประดิษฐ์ การรู้จำ และฐานข้อมูลเข้าด้วยกัน

2) เทคนิคเหมืองข้อมูล

เทคนิคเหมืองข้อมูลที่นำมาใช้เปรียบเทียบในงานวิจัย จำนวน 3 อัลกอริทึม มีดังนี้

- ซัพพอร์ตเวกเตอร์แมชชีน (Matachalearnpat, 2022) มีหลักการ คือการหา Hyperplane ที่สามารถแบ่งจุดข้อมูลออกเป็น 2 class ด้วยระยะห่างที่มากที่สุด (Maximum Margin) ในขณะเดียวกันก็สามารถแบ่งจุดข้อมูลได้อย่างถูกต้องมากที่สุดเท่าที่จะเป็นไปได้ จำนวนมิติของ Hyperplane คือจำนวนมิติของข้อมูล ในกรณีที่ข้อมูลมี 2 มิติ (x, y) Hyperplane คือเส้นที่มี Margin จากแต่ละ class มากที่สุด โดยวิธีที่ใช้ในการหา Hyperplane คือ การหาจุดข้อมูลที่อยู่ใกล้กับเส้นแบ่งพรมแดนข้อมูลทั้งสองฝั่งที่จะใช้กำหนดเส้นขอบของแต่ละฝั่ง จุดข้อมูลดังกล่าวจะถูกเรียกว่า Support Vector

- นาอิวเบย์ (Widyawati, 2023) เป็นตัวแบบที่ใช้ในการแบ่งกลุ่มข้อมูลโดยใช้ความน่าจะเป็นตามหลักของ Bayes Theorem โดยตั้งชื่อว่า "Naive" เนื่องจากมันสมมติว่าคุณสมบัติของข้อมูลทั้งหมดนั้นเป็นอิสระต่อกัน (independent) ซึ่งไม่สามารถเกิดขึ้นได้จริงในทางปฏิบัติ เพราะฉะนั้นเรียกว่าเป็น "Naive" หรือง่าย ๆ คือตัดสินใจโดยไม่คำนึงถึงความสัมพันธ์ระหว่างคุณสมบัติของข้อมูล

- ต้นไม้ตัดสินใจ (Njoku, O.C., 2019) หลักการทำงานของ ต้นไม้ตัดสินใจ เริ่มต้นที่ Root node ซึ่งเป็นจุดเริ่มต้นของต้นไม้ การตัดสินใจจะเริ่มต้นที่ Root node โดยตรวจสอบข้อมูลที่ให้มาว่าเป็นไปตามเงื่อนไขที่กำหนดหรือไม่ หากเป็นตามเงื่อนไขนั้น ตัดสินใจจะเลื่อนไปทางลูกศรทางซ้ายที่เชื่อมกับ Child node ด้านซ้ายของ Root node และทำซ้ำกระบวนการนี้จนกว่าจะถึง Leaf node ซึ่งเป็น Node สุดท้ายของต้นไม้และใช้ในการทำนาย เป็นเครื่องมือที่มีประสิทธิภาพในการทำนายและมีความสำคัญในการพัฒนาอัลกอริทึมที่ซับซ้อนมากขึ้น เช่น Random Forest ที่ใช้หลาย ๆ ต้นไม้ตัดสินใจ ร่วมกันเพื่อเพิ่มประสิทธิภาพและความถูกต้องในการทำนาย

2. การทำความเข้าใจข้อมูลที่จัดเก็บ (Data understanding)

เป็นขั้นตอนที่เริ่มต้นจากการเก็บรวบรวมข้อมูล หลังจากนั้นจะเป็นการตรวจสอบข้อมูลที่ได้ทำการรวบรวมมาเพื่อดูความถูกต้องของข้อมูล และพิจารณาว่าจะใช้ข้อมูลทั้งหมดหรือเลือกข้อมูล บางส่วนมาใช้ในการวิเคราะห์ โดยมีขั้นตอนดังนี้

2.1 การเก็บรวบรวมข้อมูล (Data gathering)

งานวิจัยนี้ได้ใช้ชุดข้อมูลหัวใจขาดเลือดเฉียบพลัน ในชุดของ Heart Attack Analysis & Prediction Dataset จากเว็บไซต์ <https://www.kaggle.com/> จำนวนข้อมูลทั้งหมด จำนวน 215 ระเบียบ คุณลักษณะ 14 คุณลักษณะ โดยอยู่ในรูปแบบไฟล์ csv. ตัวอย่างข้อมูลแสดงใน Figure 2

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	age	sex	cp	trtbps	chol	fbs	restecg	thalachh	exng	oldpeak	slp	caa	thall	output	
2	77	1	0	125	304	0	0	162	1	0	2	3	2	0	
3	76	0	2	140	197	0	2	116	0	1.1	1	0	2	1	
4	74	0	1	120	269	0	0	121	1	0.2	2	1	2	1	
5	71	0	1	160	302	0	1	162	0	0.4	2	2	2	1	
6	71	0	2	110	265	1	0	130	0	0	2	1	2	1	
7	71	0	0	112	149	0	1	125	0	1.6	1	0	2	1	
8	70	1	1	156	245	0	0	143	0	0	2	0	2	1	
9	70	1	0	145	174	0	1	125	1	2.6	0	0	3	0	

Figure 2 Heart attack data example

2.2 การอธิบายข้อมูล (Describe data)

ทำการอธิบายข้อมูล เช่น อธิบายถึงคุณลักษณะต่าง ๆ ในตารางอธิบายถึงลักษณะของข้อมูลที่น่ามาใช้ และทำการประเมินว่าข้อมูลที่ได้นั้นตรงกับความต้องการหรือไม่ แสดงรายละเอียดได้ดัง Table 1 และ Figure 3

Table 1 Feature details

	Feature	Description	Data Type	Range
1	Age	Age of the patient	Real	[50, ...,77]
2	Sex	Sex of the patient	Binary	0: Male 1: female
3	Cp	Chest Pain type	nominal	0: typical angina chest pain type 1: atypical angina 2: non-anginal pain 3: asymptomatic
4	Trtbps	resting blood pressure	Real	[94, ..., 200]
5	Chol	cholesterol in mg/dl	Real	[126, ..., 564]



Table 1 continued.

	Feature	Description	Data Type	Range
6	Fbs	fasting blood sugar > 120 mg/dl	Binary	0: false 1: true
7	Restecg	resting electrocardiographi c results	nominal	0: normal 1: having ST-T wave abnormality 2: showing probable or definite left ventricular hypertrophy
8	Thalachh	maximum heart rate achieved	Real	[71, ..., 195]
9	Exng	exercise induced angina	Binary	1: yes 0: no
10	Oldpeak	Previous peak	Real	[0, ...,6.2]
11	Slp	the slope of the peak exercise ST segment	nominal	0: Unsloping 1: Flat sloping 2: Down sloping
12	Caa	number of major vessels	Ordered	0,1,2,3,4
13	Thall	Thal rate	nominal	0-1: normal 2: fixed defect 3: reversable defect
14	output	Target variable	Binary	0: less chance of heart attack 1: more chance of heart attack

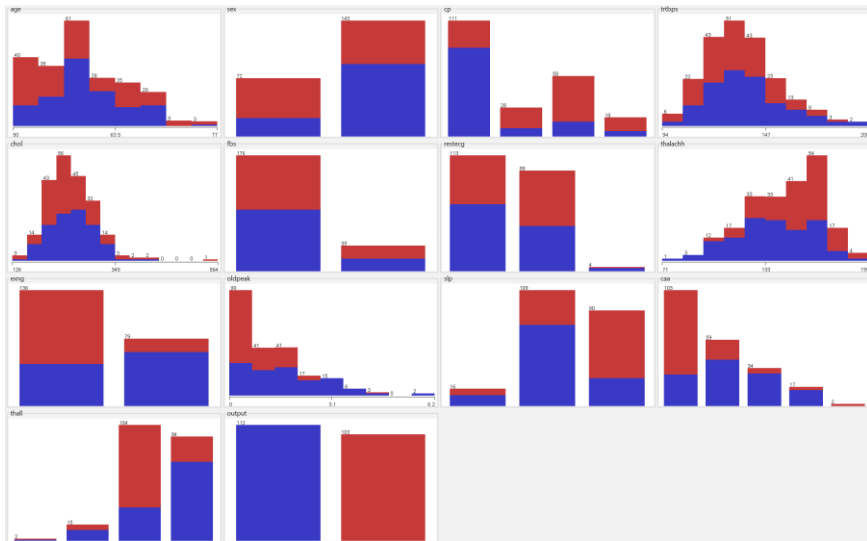


Figure 3 The graph shows the relationship between features and output.

3. การเตรียมข้อมูล (Data Preparation)

ในขั้นตอนการเตรียมข้อมูลสำหรับทำเหมืองข้อมูลที่จะนำไปใช้ในการวิเคราะห์ ผู้วิจัยได้นำข้อมูลที่ได้จากการเก็บรวบรวมข้อมูลจากขั้นตอนก่อนหน้ามาทำความสะอาดข้อมูล (Cleaning data) โดยทำการเช็คค่าว่างของข้อมูลโดยข้อมูลต้องไม่มีค่าว่างและตรวจสอบรายละเอียดข้อมูลจึงเริ่มขั้นต่อไป แสดงตัวอย่างขั้นตอนการทำความสะอาดข้อมูล และทำการแปลงข้อมูล (Data transformation) เป็นขั้นตอนการแปลงข้อมูลให้เหมาะสมสำหรับการใช้งาน ได้ทำการแปลงข้อมูลที่ได้ทำการเก็บรวบรวมมา (raw data) ให้กลายเป็นข้อมูลที่สามารถนำไปวิเคราะห์ในขั้นถัดไปได้

4. การพัฒนาตัวแบบพยากรณ์ (Modeling)

เทคนิคเหมืองข้อมูลที่เลือกนำมาเปรียบเทียบประสิทธิภาพเพื่อนำไปวิเคราะห์ข้อมูลและสร้างตัวแบบทั้งหมด 3 อัลกอริทึม ได้แก่ ซัพพอร์ตเวกเตอร์แมชชีนนาอ์ฟเบย์ และต้นไม้ตัดสินใจ โดยจะหาเทคนิคที่มีประสิทธิภาพที่ดีที่สุดและเหมาะสมกับข้อมูลที่สุดไปพัฒนาระบบตัวแบบทำนายความเสี่ยงการเกิดภาวะหัวใจขาดเลือดเฉียบพลันในผู้สูงอายุ ตัวอย่างการสร้างตัวแบบดัง Figure 4

```
from sklearn.naive_bayes import GaussianNB
model1 = GaussianNB()
model1.fit(x_train, y_train)
y_pred = model1.predict(x_test)
from sklearn.metrics import classification_report, confusion_matrix, ConfusionMatrixDisplay
print('Classification Report:\n', classification_report(y_test, y_pred))

# สร้างและแสดง Confusion Matrix
cm = confusion_matrix(y_test, y_pred)
disp = ConfusionMatrixDisplay(confusion_matrix=cm, display_labels=model.classes_)
disp.plot(cmap='viridis', values_format='d')
plt.title('Confusion Matrix')
plt.show()
```

Figure 4 Example of modeling using Naive Baye

5. การตรวจสอบประสิทธิภาพตัวแบบ (Evaluation)

การตรวจสอบประสิทธิภาพของตัวแบบ ในแต่ละเทคนิควิธีที่นำมาเปรียบเทียบจะต้องทำการเลือกข้อมูลสำหรับเรียนรู้ (Training Set) และข้อมูลสำหรับทดสอบ (Testing Set) งานวิจัยนี้เลือกใช้วิธีการสุ่มเลือกข้อมูลแบบความเที่ยงตรง k กลุ่ม (k-fold Cross Validation) โดยเลือกใช้วิธีการสุ่มเลือกข้อมูลแบบความเที่ยงตรง 5 กลุ่ม (5-fold Cross Validation) แล้วทำการวัดประสิทธิภาพของตัวแบบโดยวัดจากค่าความถูกต้อง ค่าความแม่นยำ ค่าความระลึก และค่าความถ่วงดุล เพื่อหาตัวแบบที่ประสิทธิภาพดีที่สุด การวัดประสิทธิภาพของแบบจำลองใช้ Confusion matrix ในการคำนวณค่าแสดงดัง Figure 5

		Actual Value	
		Positive	Negative
Predicated Value	Positive	TP	FP
	Negative	FN	TN

Figure 5 Confusion matrix



โดยที่

TP หรือ True Positive คือ สิ่งที่โปรแกรมทำนายว่า “จริง” และมีค่าเป็น “จริง”

TN หรือ True Negative คือ สิ่งที่โปรแกรมทำนายว่า “ไม่จริง” และมีค่า “ไม่จริง”

FP หรือ False Positive คือ สิ่งที่โปรแกรมทำนายว่า “จริง” แต่ มีค่าเป็น “ไม่จริง”

FN หรือ False Negative คือ สิ่งที่โปรแกรมทำนายว่า “ไม่จริง” แต่ มีค่าเป็น “จริง”

ตัวชี้วัดที่สามารถคำนวณได้จาก Confusion Matrix

- ค่าความแม่นยำ คือ การวัดความแม่นยำตัวแบบโดยพิจารณาแยกทีละคลาส

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP}) \quad (1)$$

- ค่าความระลึก คือ การวัดความถูกต้องตัวแบบโดยพิจารณาแยกทีละคลาส

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN}) \quad (2)$$

- ค่าความถ่วงดุล คือ การวัดค่าความแม่นยำและค่าความระลึกพร้อมกันของตัวแบบ โดยพิจารณาแยกทีละคลาส

$$\text{F-measure} = (2 \times \text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (3)$$

- ค่าความถูกต้อง คือ การวัดความถูกต้องของตัวแบบโดยพิจารณารวมทุกคลาส

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN}) \quad (4)$$

6. การนำไปใช้งาน (Deployment)

นำแบบตัวแบบที่มีค่าความถูกต้องมากที่สุดมาพัฒนาระบบตัวแบบทำนายความเสี่ยงการเกิดภาวะหัวใจขาดเลือดเฉียบพลันในผู้สูงอายุเพื่อสนับสนุนผู้เชี่ยวชาญทางการแพทย์ในการเลือกวิธีการที่เหมาะสมที่สุดในการป้องกันการเกิดภาวะหัวใจขาดเลือดเฉียบพลันในผู้สูงอายุอย่างมีประสิทธิภาพและแม่นยำยิ่งขึ้น

ผลการวิจัย

1. ผลการเปรียบเทียบประสิทธิภาพ

เทคนิคเหมืองข้อมูลที้นำมาใช้เปรียบเทียบในงานวิจัย จำนวน 3 อัลกอริทึม ได้แก่ ซัพพอร์ตเวกเตอร์แมชชีน นาอ์ฟเบย์ และต้นไม้ตัดสินใจ ได้ทำการตรวจสอบประสิทธิภาพของตัวแบบจากค่าความถูกต้อง ค่าความแม่นยำ ค่าความระลึก และค่าความถ่วงดุล สรุปได้ดัง Table 2 และ Figure 6 ดังนี้



Table 2 Performance Comparison Results

Algorithm	Accuracy	Precision	Recall	F-measure
Support Vector Machine	69.77	69.78	69.70	69.70
Naive Bayes	86.05	87.44	85.82	85.86
Decision Tree	79.07	79.13	79.04	79.02

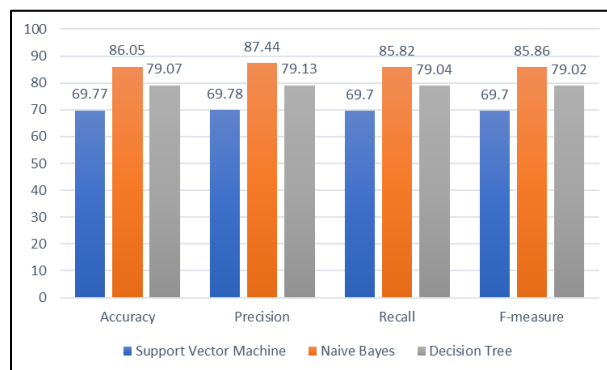


Figure 6 Performance Comparison Results

ซัพพอร์ตเวกเตอร์แมชชีน จากผลการวัดประสิทธิภาพได้ดังนี้ ค่าความถูกต้อง เท่ากับร้อยละ 69.77 ค่าความแม่นยำเท่ากับร้อยละ 69.78 ค่าความระลึกเท่ากับร้อยละ 69.70 และค่าความถ่วงดุลเท่ากับร้อยละ 69.70

นาอ็ฟเบย์ จากผลการวัดประสิทธิภาพ ได้ดังนี้ ค่าความถูกต้อง เท่ากับร้อยละ 86.05 ค่าความแม่นยำเท่ากับร้อยละ 87.44 ค่าความระลึกเท่ากับร้อยละ 85.82 และค่าความถ่วงดุลเท่ากับร้อยละ 85.86

ต้นไม้ตัดสินใจ จากผลการวัดประสิทธิภาพได้ดังนี้ ค่าความถูกต้อง เท่ากับร้อยละ 79.07 ค่าความแม่นยำเท่ากับร้อยละ 79.13 ค่าความระลึกเท่ากับร้อยละ 79.04 และค่าความถ่วงดุลเท่ากับร้อยละ 79.02

เมื่อนำค่าความถูกต้องของแต่ละเทคนิคมาเปรียบเทียบเพื่อหาเทคนิคที่มีประสิทธิภาพสูงสุด พบว่า นาอ็ฟเบย์ มีประสิทธิภาพสูงที่สุดโดยให้ค่าความถูกต้องร้อยละ 86.05 เนื่องจากข้อดีของนาอ็ฟเบย์ที่ไม่ต้องการข้อมูลสำหรับเรียนรู้มากนักเนื่องจากความเป็นอิสระระหว่างฟีเจอร์ต่างๆ สามารถคาดการณ์ได้อย่างรวดเร็วเมื่อโมเดลได้รับการเรียนรู้แล้วและสามารถให้ผลลัพธ์ที่สมเหตุสมผลแม้จะมีข้อมูลการ



เรียนรู้ที่จำกัดก็ตาม และได้ทำการเปรียบเทียบตัวแบบทั้ง 3 แบบด้วยค่า ONE-WAY ANOVA แล้วพบว่า ค่า $F = 47.237$ และ ค่า $p < 0.05$ แสดงว่าตัวแบบทั้ง 3 แบบ แตกต่างกันอย่างมีนัยสำคัญที่ระดับ 0.05

2. ผลการพัฒนาระบบตัวแบบทำนายความเสี่ยงการเกิดภาวะหัวใจขาดเลือดเฉียบพลันในผู้สูงอายุ

จากผลการเปรียบเทียบประสิทธิภาพซึ่งพบว่า ตัวแบบจากอัลกอริทึมไมโครไบเบย์มีประสิทธิภาพสูงที่สุด จึงนำมาพัฒนาระบบ ผู้วิจัยได้พัฒนาระบบตัวแบบทำนายความเสี่ยงการเกิดภาวะหัวใจขาดเลือดเฉียบพลันในผู้สูงอายุ ในรูปแบบเว็บแอปพลิเคชัน (Web Application) ด้วย Python + Streamlit มาใช้ในการทำนายโดยการให้ผู้ใช้สามารถใส่ข้อมูลเพื่อทำนายว่าโอกาสการเกิดภาวะหัวใจขาดเลือดเฉียบพลันมีมากหรือน้อย และบอกข้อมูลทั่วไคราะห์แล้วมาเพื่อให้สามารถทราบและป้องกันหรือลดโอกาส ที่จะเกิดภาวะหัวใจขาดเลือดเฉียบพลันจนถึงขั้นเสียชีวิตตัวอย่างหน้าจอระบบแสดงใน Figure 7

Predictive Model for Heart Attack in Elderly

Age : อายุของผู้ป่วย

30

Sex : เพศของผู้ป่วย (0 : หญิง | 1 : ชาย)

0

Cp : อาการเจ็บหน้าอก
(0 : ไม่พบคลื่นผิดปกติใน ECG | 1 : พบคลื่นผิดปกติใน ECG | 2 : อาการเจ็บหน้าอกในขณะออกกำลังกาย | 3 : ไม่สบาย)

0

Trestps : ความดันโลหิตขณะพัก (mm, mmHg)

120

Chol : cholesterol ใน mg/dl (ค่าปกติ < 200 mg/dl)

194

Fbs : น้ำตาลในเลือดขณะอดอาหาร > 120 mg/dl (1 = 120-130 mg/dl | 2 = 130-140 mg/dl | 3 = 140-160 mg/dl | 4 = 160-180 mg/dl | 5 = 180-200 mg/dl | 6 = 200-220 mg/dl | 7 = 220-240 mg/dl | 8 = 240-260 mg/dl | 9 = 260-280 mg/dl | 10 = 280-300 mg/dl | 11 = 300-320 mg/dl | 12 = 320-340 mg/dl | 13 = 340-360 mg/dl | 14 = 360-380 mg/dl | 15 = 380-400 mg/dl | 16 = 400-420 mg/dl | 17 = 420-440 mg/dl | 18 = 440-460 mg/dl | 19 = 460-480 mg/dl | 20 = 480-500 mg/dl | 21 = 500-520 mg/dl | 22 = 520-540 mg/dl | 23 = 540-560 mg/dl | 24 = 560-580 mg/dl | 25 = 580-600 mg/dl | 26 = 600-620 mg/dl | 27 = 620-640 mg/dl | 28 = 640-660 mg/dl | 29 = 660-680 mg/dl | 30 = 680-700 mg/dl | 31 = 700-720 mg/dl | 32 = 720-740 mg/dl | 33 = 740-760 mg/dl | 34 = 760-780 mg/dl | 35 = 780-800 mg/dl | 36 = 800-820 mg/dl | 37 = 820-840 mg/dl | 38 = 840-860 mg/dl | 39 = 860-880 mg/dl | 40 = 880-900 mg/dl | 41 = 900-920 mg/dl | 42 = 920-940 mg/dl | 43 = 940-960 mg/dl | 44 = 960-980 mg/dl | 45 = 980-1000 mg/dl | 46 = 1000-1020 mg/dl | 47 = 1020-1040 mg/dl | 48 = 1040-1060 mg/dl | 49 = 1060-1080 mg/dl | 50 = 1080-1100 mg/dl | 51 = 1100-1120 mg/dl | 52 = 1120-1140 mg/dl | 53 = 1140-1160 mg/dl | 54 = 1160-1180 mg/dl | 55 = 1180-1200 mg/dl | 56 = 1200-1220 mg/dl | 57 = 1220-1240 mg/dl | 58 = 1240-1260 mg/dl | 59 = 1260-1280 mg/dl | 60 = 1280-1300 mg/dl | 61 = 1300-1320 mg/dl | 62 = 1320-1340 mg/dl | 63 = 1340-1360 mg/dl | 64 = 1360-1380 mg/dl | 65 = 1380-1400 mg/dl | 66 = 1400-1420 mg/dl | 67 = 1420-1440 mg/dl | 68 = 1440-1460 mg/dl | 69 = 1460-1480 mg/dl | 70 = 1480-1500 mg/dl | 71 = 1500-1520 mg/dl | 72 = 1520-1540 mg/dl | 73 = 1540-1560 mg/dl | 74 = 1560-1580 mg/dl | 75 = 1580-1600 mg/dl | 76 = 1600-1620 mg/dl | 77 = 1620-1640 mg/dl | 78 = 1640-1660 mg/dl | 79 = 1660-1680 mg/dl | 80 = 1680-1700 mg/dl | 81 = 1700-1720 mg/dl | 82 = 1720-1740 mg/dl | 83 = 1740-1760 mg/dl | 84 = 1760-1780 mg/dl | 85 = 1780-1800 mg/dl | 86 = 1800-1820 mg/dl | 87 = 1820-1840 mg/dl | 88 = 1840-1860 mg/dl | 89 = 1860-1880 mg/dl | 90 = 1880-1900 mg/dl | 91 = 1900-1920 mg/dl | 92 = 1920-1940 mg/dl | 93 = 1940-1960 mg/dl | 94 = 1960-1980 mg/dl | 95 = 1980-2000 mg/dl | 96 = 2000-2020 mg/dl | 97 = 2020-2040 mg/dl | 98 = 2040-2060 mg/dl | 99 = 2060-2080 mg/dl | 100 = 2080-2100 mg/dl | 101 = 2100-2120 mg/dl | 102 = 2120-2140 mg/dl | 103 = 2140-2160 mg/dl | 104 = 2160-2180 mg/dl | 105 = 2180-2200 mg/dl | 106 = 2200-2220 mg/dl | 107 = 2220-2240 mg/dl | 108 = 2240-2260 mg/dl | 109 = 2260-2280 mg/dl | 110 = 2280-2300 mg/dl | 111 = 2300-2320 mg/dl | 112 = 2320-2340 mg/dl | 113 = 2340-2360 mg/dl | 114 = 2360-2380 mg/dl | 115 = 2380-2400 mg/dl | 116 = 2400-2420 mg/dl | 117 = 2420-2440 mg/dl | 118 = 2440-2460 mg/dl | 119 = 2460-2480 mg/dl | 120 = 2480-2500 mg/dl | 121 = 2500-2520 mg/dl | 122 = 2520-2540 mg/dl | 123 = 2540-2560 mg/dl | 124 = 2560-2580 mg/dl | 125 = 2580-2600 mg/dl | 126 = 2600-2620 mg/dl | 127 = 2620-2640 mg/dl | 128 = 2640-2660 mg/dl | 129 = 2660-2680 mg/dl | 130 = 2680-2700 mg/dl | 131 = 2700-2720 mg/dl | 132 = 2720-2740 mg/dl | 133 = 2740-2760 mg/dl | 134 = 2760-2780 mg/dl | 135 = 2780-2800 mg/dl | 136 = 2800-2820 mg/dl | 137 = 2820-2840 mg/dl | 138 = 2840-2860 mg/dl | 139 = 2860-2880 mg/dl | 140 = 2880-2900 mg/dl | 141 = 2900-2920 mg/dl | 142 = 2920-2940 mg/dl | 143 = 2940-2960 mg/dl | 144 = 2960-2980 mg/dl | 145 = 2980-3000 mg/dl | 146 = 3000-3020 mg/dl | 147 = 3020-3040 mg/dl | 148 = 3040-3060 mg/dl | 149 = 3060-3080 mg/dl | 150 = 3080-3100 mg/dl | 151 = 3100-3120 mg/dl | 152 = 3120-3140 mg/dl | 153 = 3140-3160 mg/dl | 154 = 3160-3180 mg/dl | 155 = 3180-3200 mg/dl | 156 = 3200-3220 mg/dl | 157 = 3220-3240 mg/dl | 158 = 3240-3260 mg/dl | 159 = 3260-3280 mg/dl | 160 = 3280-3300 mg/dl | 161 = 3300-3320 mg/dl | 162 = 3320-3340 mg/dl | 163 = 3340-3360 mg/dl | 164 = 3360-3380 mg/dl | 165 = 3380-3400 mg/dl | 166 = 3400-3420 mg/dl | 167 = 3420-3440 mg/dl | 168 = 3440-3460 mg/dl | 169 = 3460-3480 mg/dl | 170 = 3480-3500 mg/dl | 171 = 3500-3520 mg/dl | 172 = 3520-3540 mg/dl | 173 = 3540-3560 mg/dl | 174 = 3560-3580 mg/dl | 175 = 3580-3600 mg/dl | 176 = 3600-3620 mg/dl | 177 = 3620-3640 mg/dl | 178 = 3640-3660 mg/dl | 179 = 3660-3680 mg/dl | 180 = 3680-3700 mg/dl | 181 = 3700-3720 mg/dl | 182 = 3720-3740 mg/dl | 183 = 3740-3760 mg/dl | 184 = 3760-3780 mg/dl | 185 = 3780-3800 mg/dl | 186 = 3800-3820 mg/dl | 187 = 3820-3840 mg/dl | 188 = 3840-3860 mg/dl | 189 = 3860-3880 mg/dl | 190 = 3880-3900 mg/dl | 191 = 3900-3920 mg/dl | 192 = 3920-3940 mg/dl | 193 = 3940-3960 mg/dl | 194 = 3960-3980 mg/dl | 195 = 3980-4000 mg/dl | 196 = 4000-4020 mg/dl | 197 = 4020-4040 mg/dl | 198 = 4040-4060 mg/dl | 199 = 4060-4080 mg/dl | 200 = 4080-4100 mg/dl | 201 = 4100-4120 mg/dl | 202 = 4120-4140 mg/dl | 203 = 4140-4160 mg/dl | 204 = 4160-4180 mg/dl | 205 = 4180-4200 mg/dl | 206 = 4200-4220 mg/dl | 207 = 4220-4240 mg/dl | 208 = 4240-4260 mg/dl | 209 = 4260-4280 mg/dl | 210 = 4280-4300 mg/dl | 211 = 4300-4320 mg/dl | 212 = 4320-4340 mg/dl | 213 = 4340-4360 mg/dl | 214 = 4360-4380 mg/dl | 215 = 4380-4400 mg/dl | 216 = 4400-4420 mg/dl | 217 = 4420-4440 mg/dl | 218 = 4440-4460 mg/dl | 219 = 4460-4480 mg/dl | 220 = 4480-4500 mg/dl | 221 = 4500-4520 mg/dl | 222 = 4520-4540 mg/dl | 223 = 4540-4560 mg/dl | 224 = 4560-4580 mg/dl | 225 = 4580-4600 mg/dl | 226 = 4600-4620 mg/dl | 227 = 4620-4640 mg/dl | 228 = 4640-4660 mg/dl | 229 = 4660-4680 mg/dl | 230 = 4680-4700 mg/dl | 231 = 4700-4720 mg/dl | 232 = 4720-4740 mg/dl | 233 = 4740-4760 mg/dl | 234 = 4760-4780 mg/dl | 235 = 4780-4800 mg/dl | 236 = 4800-4820 mg/dl | 237 = 4820-4840 mg/dl | 238 = 4840-4860 mg/dl | 239 = 4860-4880 mg/dl | 240 = 4880-4900 mg/dl | 241 = 4900-4920 mg/dl | 242 = 4920-4940 mg/dl | 243 = 4940-4960 mg/dl | 244 = 4960-4980 mg/dl | 245 = 4980-5000 mg/dl | 246 = 5000-5020 mg/dl | 247 = 5020-5040 mg/dl | 248 = 5040-5060 mg/dl | 249 = 5060-5080 mg/dl | 250 = 5080-5100 mg/dl | 251 = 5100-5120 mg/dl | 252 = 5120-5140 mg/dl | 253 = 5140-5160 mg/dl | 254 = 5160-5180 mg/dl | 255 = 5180-5200 mg/dl | 256 = 5200-5220 mg/dl | 257 = 5220-5240 mg/dl | 258 = 5240-5260 mg/dl | 259 = 5260-5280 mg/dl | 260 = 5280-5300 mg/dl | 261 = 5300-5320 mg/dl | 262 = 5320-5340 mg/dl | 263 = 5340-5360 mg/dl | 264 = 5360-5380 mg/dl | 265 = 5380-5400 mg/dl | 266 = 5400-5420 mg/dl | 267 = 5420-5440 mg/dl | 268 = 5440-5460 mg/dl | 269 = 5460-5480 mg/dl | 270 = 5480-5500 mg/dl | 271 = 5500-5520 mg/dl | 272 = 5520-5540 mg/dl | 273 = 5540-5560 mg/dl | 274 = 5560-5580 mg/dl | 275 = 5580-5600 mg/dl | 276 = 5600-5620 mg/dl | 277 = 5620-5640 mg/dl | 278 = 5640-5660 mg/dl | 279 = 5660-5680 mg/dl | 280 = 5680-5700 mg/dl | 281 = 5700-5720 mg/dl | 282 = 5720-5740 mg/dl | 283 = 5740-5760 mg/dl | 284 = 5760-5780 mg/dl | 285 = 5780-5800 mg/dl | 286 = 5800-5820 mg/dl | 287 = 5820-5840 mg/dl | 288 = 5840-5860 mg/dl | 289 = 5860-5880 mg/dl | 290 = 5880-5900 mg/dl | 291 = 5900-5920 mg/dl | 292 = 5920-5940 mg/dl | 293 = 5940-5960 mg/dl | 294 = 5960-5980 mg/dl | 295 = 5980-6000 mg/dl | 296 = 6000-6020 mg/dl | 297 = 6020-6040 mg/dl | 298 = 6040-6060 mg/dl | 299 = 6060-6080 mg/dl | 300 = 6080-6100 mg/dl | 301 = 6100-6120 mg/dl | 302 = 6120-6140 mg/dl | 303 = 6140-6160 mg/dl | 304 = 6160-6180 mg/dl | 305 = 6180-6200 mg/dl | 306 = 6200-6220 mg/dl | 307 = 6220-6240 mg/dl | 308 = 6240-6260 mg/dl | 309 = 6260-6280 mg/dl | 310 = 6280-6300 mg/dl | 311 = 6300-6320 mg/dl | 312 = 6320-6340 mg/dl | 313 = 6340-6360 mg/dl | 314 = 6360-6380 mg/dl | 315 = 6380-6400 mg/dl | 316 = 6400-6420 mg/dl | 317 = 6420-6440 mg/dl | 318 = 6440-6460 mg/dl | 319 = 6460-6480 mg/dl | 320 = 6480-6500 mg/dl | 321 = 6500-6520 mg/dl | 322 = 6520-6540 mg/dl | 323 = 6540-6560 mg/dl | 324 = 6560-6580 mg/dl | 325 = 6580-6600 mg/dl | 326 = 6600-6620 mg/dl | 327 = 6620-6640 mg/dl | 328 = 6640-6660 mg/dl | 329 = 6660-6680 mg/dl | 330 = 6680-6700 mg/dl | 331 = 6700-6720 mg/dl | 332 = 6720-6740 mg/dl | 333 = 6740-6760 mg/dl | 334 = 6760-6780 mg/dl | 335 = 6780-6800 mg/dl | 336 = 6800-6820 mg/dl | 337 = 6820-6840 mg/dl | 338 = 6840-6860 mg/dl | 339 = 6860-6880 mg/dl | 340 = 6880-6900 mg/dl | 341 = 6900-6920 mg/dl | 342 = 6920-6940 mg/dl | 343 = 6940-6960 mg/dl | 344 = 6960-6980 mg/dl | 345 = 6980-7000 mg/dl | 346 = 7000-7020 mg/dl | 347 = 7020-7040 mg/dl | 348 = 7040-7060 mg/dl | 349 = 7060-7080 mg/dl | 350 = 7080-7100 mg/dl | 351 = 7100-7120 mg/dl | 352 = 7120-7140 mg/dl | 353 = 7140-7160 mg/dl | 354 = 7160-7180 mg/dl | 355 = 7180-7200 mg/dl | 356 = 7200-7220 mg/dl | 357 = 7220-7240 mg/dl | 358 = 7240-7260 mg/dl | 359 = 7260-7280 mg/dl | 360 = 7280-7300 mg/dl | 361 = 7300-7320 mg/dl | 362 = 7320-7340 mg/dl | 363 = 7340-7360 mg/dl | 364 = 7360-7380 mg/dl | 365 = 7380-7400 mg/dl | 366 = 7400-7420 mg/dl | 367 = 7420-7440 mg/dl | 368 = 7440-7460 mg/dl | 369 = 7460-7480 mg/dl | 370 = 7480-7500 mg/dl | 371 = 7500-7520 mg/dl | 372 = 7520-7540 mg/dl | 373 = 7540-7560 mg/dl | 374 = 7560-7580 mg/dl | 375 = 7580-7600 mg/dl | 376 = 7600-7620 mg/dl | 377 = 7620-7640 mg/dl | 378 = 7640-7660 mg/dl | 379 = 7660-7680 mg/dl | 380 = 7680-7700 mg/dl | 381 = 7700-7720 mg/dl | 382 = 7720-7740 mg/dl | 383 = 7740-7760 mg/dl | 384 = 7760-7780 mg/dl | 385 = 7780-7800 mg/dl | 386 = 7800-7820 mg/dl | 387 = 7820-7840 mg/dl | 388 = 7840-7860 mg/dl | 389 = 7860-7880 mg/dl | 390 = 7880-7900 mg/dl | 391 = 7900-7920 mg/dl | 392 = 7920-7940 mg/dl | 393 = 7940-7960 mg/dl | 394 = 7960-7980 mg/dl | 395 = 7980-8000 mg/dl | 396 = 8000-8020 mg/dl | 397 = 8020-8040 mg/dl | 398 = 8040-8060 mg/dl | 399 = 8060-8080 mg/dl | 400 = 8080-8100 mg/dl | 401 = 8100-8120 mg/dl | 402 = 8120-8140 mg/dl | 403 = 8140-8160 mg/dl | 404 = 8160-8180 mg/dl | 405 = 8180-8200 mg/dl | 406 = 8200-8220 mg/dl | 407 = 8220-8240 mg/dl | 408 = 8240-8260 mg/dl | 409 = 8260-8280 mg/dl | 410 = 8280-8300 mg/dl | 411 = 8300-8320 mg/dl | 412 = 8320-8340 mg/dl | 413 = 8340-8360 mg/dl | 414 = 8360-8380 mg/dl | 415 = 8380-8400 mg/dl | 416 = 8400-8420 mg/dl | 417 = 8420-8440 mg/dl | 418 = 8440-8460 mg/dl | 419 = 8460-8480 mg/dl | 420 = 8480-8500 mg/dl | 421 = 8500-8520 mg/dl | 422 = 8520-8540 mg/dl | 423 = 8540-8560 mg/dl | 424 = 8560-8580 mg/dl | 425 = 8580-8600 mg/dl | 426 = 8600-8620 mg/dl | 427 = 8620-8640 mg/dl | 428 = 8640-8660 mg/dl | 429 = 8660-8680 mg/dl | 430 = 8680-8700 mg/dl | 431 = 8700-8720 mg/dl | 432 = 8720-8740 mg/dl | 433 = 8740-8760 mg/dl | 434 = 8760-8780 mg/dl | 435 = 8780-8800 mg/dl | 436 = 8800-8820 mg/dl | 437 = 8820-8840 mg/dl | 438 = 8840-8860 mg/dl | 439 = 8860-8880 mg/dl | 440 = 8880-8900 mg/dl | 441 = 8900-8920 mg/dl | 442 = 8920-8940 mg/dl | 443 = 8940-8960 mg/dl | 444 = 8960-8980 mg/dl | 445 = 8980-9000 mg/dl | 446 = 9000-9020 mg/dl | 447 = 9020-9040 mg/dl | 448 = 9040-9060 mg/dl | 449 = 9060-9080 mg/dl | 450 = 9080-9100 mg/dl | 451 = 9100-9120 mg/dl | 452 = 9120-9140 mg/dl | 453 = 9140-9160 mg/dl | 454 = 9160-9180 mg/dl | 455 = 9180-9200 mg/dl | 456 = 9200-9220 mg/dl | 457 = 9220-9240 mg/dl | 458 = 9240-9260 mg/dl | 459 = 9260-9280 mg/dl | 460 = 9280-9300 mg/dl | 461 = 9300-9320 mg/dl | 462 = 9320-9340 mg/dl | 463 = 9340-9360 mg/dl | 464 = 9360-9380 mg/dl | 465 = 9380-9400 mg/dl | 466 = 9400-9420 mg/dl | 467 = 9420-9440 mg/dl | 468 = 9440-9460 mg/dl | 469 = 9460-9480 mg/dl | 470 = 9480-9500 mg/dl | 471 = 9500-9520 mg/dl | 472 = 9520-9540 mg/dl | 473 = 9540-9560 mg/dl | 474 = 9560-9580 mg/dl | 475 = 9580-9600 mg/dl | 476 = 9600-9620 mg/dl | 477 = 9620-9640 mg/dl | 478 = 9640-9660 mg/dl | 479 = 9660-9680 mg/dl | 480 = 9680-9700 mg/dl | 481 = 9700-9720 mg/dl | 482 = 9720-9740 mg/dl | 483 = 9740-9760 mg/dl | 484 = 9760-9780 mg/dl | 485 = 9780-9800 mg/dl | 486 = 9800-9820 mg/dl | 487 = 9820-9840 mg/dl | 488 = 9840-9860 mg/dl | 489 = 9860-9880 mg/dl | 490 = 9880-9900 mg/dl | 491 = 9900-9920 mg/dl | 492 = 9920-9940 mg/dl | 493 = 9940-9960 mg/dl | 494 = 9960-9980 mg/dl | 495 = 9980-10000 mg/dl | 496 = 10000-10020 mg/dl | 497 = 10020-10040 mg/dl | 498 = 10040-10060 mg/dl | 499 = 10060-10080 mg/dl | 500 = 10080-10100 mg/dl | 501 = 10100-10120 mg/dl | 502 = 10120-10140 mg/dl | 503 = 10140-10160 mg/dl | 504 = 10160-10180 mg/dl | 505 = 10180-10200 mg/dl | 506 = 10200-10220 mg/dl | 507 = 10220-10240 mg/dl | 508 = 10240-10260 mg/dl | 509 = 10260-10280 mg/dl | 510 = 10280-10300 mg/dl | 511 = 10300-10320 mg/dl | 512 = 10320-10340 mg/dl | 513 = 10340-10360 mg/dl | 514 = 10360-10380 mg/dl | 515 = 10380-10400 mg/dl | 516 = 10400-10420 mg/dl | 517 = 10420-10440 mg/dl | 518 = 10440-10460 mg/dl | 519 = 10460-10480 mg/dl | 520 = 10480-10500 mg/dl | 521 = 10500-10520 mg/dl | 522 = 10520-10540 mg/dl | 523 = 10540-10560 mg/dl | 524 = 10560-10580 mg/dl | 525 = 10580-10600 mg/dl | 526 = 10600-10620 mg/dl | 527 = 10620-10640 mg/dl | 528 = 10640-10660 mg/dl | 529 = 10660-10680 mg/dl | 530 = 10680-10700 mg/dl | 531 = 10700-10720 mg/dl | 532 = 10720-10740 mg/dl | 533 = 10740-10760 mg/dl | 534 = 10760-10780 mg/dl | 535 = 10780-10800 mg/dl | 536 = 10800-10820 mg/dl | 537 = 10820-10840 mg/dl | 538 = 10840-10860 mg/dl | 539 = 10860-10880 mg/dl | 540 = 10880-10900 mg/dl | 541 = 10900-10920 mg/dl | 542 = 10920-10940 mg/dl | 543 = 10940-10960 mg/dl | 544 = 10960-10980 mg/dl | 545 = 10980-11000 mg/dl | 546 = 11000-11020 mg/dl | 547 = 11020-11040 mg/dl | 548 = 11040-11060 mg/dl | 549 = 11060-11080 mg/dl | 550 = 11080-11100 mg/dl | 551 = 11100-11120 mg/dl | 552 = 11120-11140 mg/dl | 553 = 11140-11160 mg/dl | 554 = 11160-11180 mg/dl | 555 = 11180-11200 mg/dl | 556 = 11200-11220 mg/dl | 557 = 11220-11240 mg/dl | 558 = 11240-11260 mg/dl | 559 = 11260-11280 mg/dl | 560 = 11280-11300 mg/dl | 561 = 11300-11320 mg/dl | 562 = 11320-11340 mg/dl | 563 = 11340-11360 mg/dl | 564 = 11360-11380 mg/dl | 565 = 11380-11400 mg/dl | 566 = 11400-11420 mg/dl | 567 = 11420-11440 mg/dl | 568 = 11440-11460 mg/dl | 569 = 11460-11480 mg/dl | 570 = 11480-11500 mg/dl | 571 = 11500-11520 mg/dl | 572 = 11520-11540 mg/dl | 573 = 11540-11560 mg/dl | 574 = 11560-11580 mg/dl | 575 = 11580-11600 mg/dl | 576 = 11600-11620 mg/dl | 577 = 11620-11640 mg/dl | 578 = 11640-11660 mg/dl | 579 = 11660-11680 mg/dl | 580 = 11680-11700 mg/dl | 581 = 11700-11720 mg/dl | 582 = 11720-11740 mg/dl | 583 = 11740-11760 mg/dl | 584 = 11760-11780 mg/dl | 585 = 11780-11800 mg/dl | 586 = 11800-11820 mg/dl | 587 = 11820-11840 mg/dl | 588 = 11840-11860 mg/dl | 589 = 11860-11880 mg/dl | 590 = 11880-11900 mg/dl | 591 = 11900-11920 mg/dl | 592 = 11920-11940 mg/dl | 593 = 11940-11960 mg/dl | 594 = 11960-11980 mg/dl | 595 = 11980-12000 mg/dl | 596 = 12000-12020 mg/dl | 597 = 12020-12040 mg/dl | 598 = 12040-12060 mg/dl | 599 = 12060-12080 mg/dl | 600 = 12080-12100 mg/dl | 601 = 12100-12120 mg/dl | 602 = 12120-12140 mg/dl | 603 = 12140-12160 mg/dl | 604 = 12160-12180 mg/dl | 605 = 12180-12200 mg/dl | 606 = 12200-12220 mg/dl | 607 = 12220-12240 mg/dl | 608 = 12240-12260 mg/dl | 609 = 12260-12280 mg/dl | 610 = 12280-12300 mg/dl | 611 = 12300-12320 mg/dl | 612 = 12320-12340 mg/dl | 613 = 12340

อภิปรายและสรุปผลการวิจัย

หัวใจขาดเลือดเฉียบพลัน เกิดจากหลอดเลือดหัวใจอุดตันหรือตีบกะทันหัน กล้ามเนื้อหัวใจตายหรือขาดเลือดรุนแรง หัวใจเต้นผิดจังหวะ หรือตกอยู่ในภาวะหัวใจล้มเหลวจนทำให้หัวใจหยุดเต้น ดังนั้นการรู้ทันโรคเป็นเรื่องสำคัญเพื่อช่วยป้องกันและลดโอกาสการเสียชีวิต โดยเฉพาะผู้สูงอายุที่จัดเป็นกลุ่มเสี่ยงในการเกิดโรคนี้ การวิจัยนี้จึงมีวัตถุประสงค์ คือ 1) เพื่อเปรียบเทียบประสิทธิภาพของตัวแบบทำนายความเสี่ยงการเกิดภาวะหัวใจขาดเลือดเฉียบพลันในผู้สูงอายุจำนวน 3 อัลกอริทึม ได้แก่ ซัพพอร์ตเวกเตอร์แมชชีน นาอ์ฟเบย์ และต้นไม้ตัดสินใจ และ 2) เพื่อพัฒนาระบบต้นแบบพยากรณ์การเกิดภาวะหัวใจขาดเลือดเฉียบพลันในผู้สูงอายุ ซึ่งมีขั้นตอนการดำเนินงานคือ 1) การรวบรวมข้อมูลโดยใช้ชุดข้อมูลหัวใจขาดเลือดเฉียบพลัน ในชุดของ Heart Attack Analysis & Prediction Dataset ที่ได้ถูกรวบรวมไว้ในเว็บไซต์ <http://www.kaggle.com/> และทำการอธิบายรายละเอียดของข้อมูลที่นำมาใช้ 2) การเตรียมข้อมูล โดยได้ทำความสะอาดข้อมูลและแปลงข้อมูลสำหรับนำไปสร้างแบบจำลอง 3) การสร้างตัวแบบ โดยเครื่องมือในการวิจัยครั้งนี้ใช้ Scikit-Learn Library ของภาษา Python ในการวิเคราะห์ข้อมูลและสร้างตัวแบบ อัลกอริทึมที่นำมาใช้เปรียบเทียบในงานวิจัย จำนวน 3 อัลกอริทึม ได้แก่ ซัพพอร์ตเวกเตอร์แมชชีน นาอ์ฟเบย์ และต้นไม้ตัดสินใจ 4) การประเมินประสิทธิภาพตัวแบบ ได้ใช้ค่าความถูกต้อง ค่าความแม่นยำ ค่าความระลึก และค่าความถ่วงดุล ในการวัดประสิทธิภาพการทำนายของตัวแบบ และ 5) การใช้งานเป็นการนำแบบจำลองที่ดีที่สุดไปพัฒนาระบบตัวแบบทำนายความเสี่ยงการเกิดภาวะหัวใจขาดเลือดเฉียบพลันในผู้สูงอายุ ผลการวิจัยพบว่า นาอ์ฟเบย์ มีประสิทธิภาพสูงที่สุดโดยให้ค่าความถูกต้องร้อยละ 86.05 รองลงมาคือเทคนิคต้นไม้ตัดสินใจ ให้ค่าความถูกต้องร้อยละ 79.07 สุดท้ายคือซัพพอร์ตเวกเตอร์แมชชีน ให้ค่าความถูกต้องร้อยละ 69.77 และได้นำตัวแบบที่ได้จากนาอ์ฟเบย์ที่มีประสิทธิภาพสูงที่สุดมาพัฒนาระบบตัวแบบทำนายความเสี่ยงการเกิดภาวะหัวใจขาดเลือดเฉียบพลันในผู้สูงอายุที่พัฒนาในรูปแบบเว็บแอปพลิเคชันให้ทำนายว่าโอกาสการเกิดภาวะหัวใจขาดเลือดเฉียบพลันมีมากหรือน้อย ให้ผู้สามารถใช้ทราบและป้องกันหรือลดโอกาสที่จะเกิดภาวะหัวใจขาดเลือดเฉียบพลันจนถึงขั้นเสียชีวิต ในการวิจัยนี้แสดงให้เห็นว่าเทคนิคเหมืองข้อมูลสามารถนำไปใช้ได้อย่างมีประสิทธิภาพในการจำแนกชุดข้อมูลหัวใจขาดเลือดเฉียบพลัน แต่อย่างไรก็ตามตัวแบบการทำนายที่ผู้วิจัยได้นำเสนอนั้นเป็นเพียงการใช้เทคนิคส่วนหนึ่งของการทำเหมืองข้อมูลจากหลากหลายเทคนิคที่มีอยู่ และยังมีข้อจำกัดด้านชุดข้อมูลที่นำมาทำการวิจัยซึ่งเป็นเพียงชุดข้อมูลขนาดเล็ก ถ้าหากได้ชุดข้อมูลที่ใหญ่กว่าจะทำให้ได้รับผลลัพธ์ที่น่าเชื่อถือมากขึ้น



References

- Kaewna, P. (2023). Factors associated with Streptokinase failure in the treatment of STEMI patients in District hospital: a retrospective study. *Academic Journal of Mahasarakham Provincial Public Health Office*, 7(14), 113-122.
- Mannaisatjatham, L. (2022). Simulation of Oncology Drugs Inventory Controls by Drug Inventory Department Using Vendor Managed Inventory through Decision Trees Classification in National Cancer Institute of Thailand. *Journal of The Department of Medical Services*, 46 (4), 106-14.
- Matachalearnpat, R. (2022). *Application of Machine Learning with industrial work*. National Electronics and Computer Technology Center.
- Njoku O.C. (2019). *Decision Trees and Their Application for Classification and Regression Problems*. (Graduate Theses). United State: Missouri State University.
- Panyatip, T. et al. (2022). Optimization Classification of Covid-19 effect on Liver Cancer. *Journal of Technology Management Rajabhat Maha Sarakham University*, 9(2), 66-78.
- Petwised, W. (2023). Development of a Models to Predict Colorectal Cancer Warning Sign. *Journal of Science and Technology*, Rajabhat Maha Sarakham University 6 (1). Maha Sarakham, (75-87).
- Rungrattanaubol, J. (2023). *Data Mining Techniques*. Naresuan University Publishing House, Phitsanulok.
- Suesat, T. et al. (2020). Factors Associated with Treatment Result of Acute Coronary Syndrome Patients. *KKU Journal for Public Health Research*, 13(2), 102-112.
- Sinsomboonthong, S. (2017). *Data mining 1: discovering knowledge in data*. Bangkok: Chamchuree Products Co.,Ltd.
- Tosasukul, J. (2021). A Prediction Modelling for Dengue Fever Outbreaks Using Data Mining Techniques. *PKRU SciTech Journal* 5 (2), 51-60.
- Widyawati, D. et al. (2023). Comparison Analysis of Classification Model Performance in Lung Cancer Prediction Using Decision Tree, Naive Bayes, and Support Vector Machine. *Indonesian Journal of Data and Science*, 4(2), 78-86.

