

Received: 29 เม.ย. 2567

Revised: 2 ก.ค. 2567

Accepted: 5 ก.ค. 2567

การพัฒนาตัวแบบสำหรับคัดกรองผู้ป่วยภาวะสมองเสื่อมในระยะเริ่มแรก
ด้วยเทคนิคการเรียนรู้ของเครื่อง

Development Of A Model For Screening Patients With Dementia In The Early
Stages with Machine Learning Techniques

เลิศศักดิ์ โพธิ์ทอง¹, ชุติมิน จำปี¹, เพ็ญพิชชา จันทร์โท¹ และ ประภาภรณ์ ชุบสุวรรณ^{1*}

¹สาขาคอมพิวเตอร์ธุรกิจ คณะการบัญชีและการจัดการ มหาวิทยาลัยมหาสารคาม

Lersak Phothong¹, Chutimon Jampee¹, Phenphitcha Chantho¹ and Prapaporn
Chubsuwan^{1*}

¹Department of Business Computer, Mahasarakham Business School,
Mahasarakham University

* ผู้นิพนธ์ประสานงาน: ประภาภรณ์ ชุบสุวรรณ อีเมล: prapaporn.c@mbs.msu.ac.th

Abstract

Accurate early detection of dementia is crucial for effective management. This study aims to develop and compare the performance of three machine learning models: Random Forest (RF), Logistic Regression (LR), and Support Vector Machine (SVM) for screening brain conditions. Using a dataset of 1,842 samples, including dementia data and 18 health factors, the data was analyzed, and models were developed according to the standard CRISP-DM data mining processes. Performance was assessed using accuracy, F-measure, sensitivity, and specificity. The results showed that LR had the highest accuracy (94.03%), followed by RF (93.60%) and SVM (92.56%). LR also demonstrated superior F-measure (96.76%), sensitivity (99.51%), and specificity (47.63%). This study indicates the potential of LR in dementia screening, outperforming RF and SVM in terms of accuracy, efficiency, and balanced classification. These findings support the development of machine learning-based tools to aid clinical decision-making in dementia diagnosis and effective treatment planning.

Keywords: *Dementia; Machine Learning; Logistic Regression; Classification
Performance; Dementia Screening*

บทคัดย่อ

การตรวจพบภาวะสมองเสื่อมตั้งแต่ระยะเริ่มต้นอย่างแม่นยำเป็นสิ่งสำคัญต่อการจัดการที่มีประสิทธิภาพ การศึกษานี้มุ่งพัฒนาและเปรียบเทียบประสิทธิภาพของตัวแบบการเรียนรู้ของเครื่อง 3 เทคนิค ได้แก่ เทคนิคต้นไม้ป่าสุ่ม เทคนิคการวิเคราะห์การถดถอยโลจิสติก และเทคนิคซัพพอร์ตเวกเตอร์แมชชีน สำหรับการคัดกรองภาวะสมองเสื่อม โดยใช้ชุดข้อมูล 1,842 ตัวอย่าง ประกอบด้วยข้อมูลภาวะสมองเสื่อมและข้อมูลปัจจัยด้านสุขภาพ 18 รายการ นำมาวิเคราะห์ข้อมูลและพัฒนาแบบจำลองตามกระบวนการมาตรฐานในการทำเหมืองข้อมูล (CRISP-DM) ประเมินประสิทธิภาพโดยใช้ความแม่นยำ ประสิทธิภาพโดยรวม ความไว และความจำเพาะ ผลการศึกษาพบว่าเทคนิคการวิเคราะห์การถดถอยโลจิสติกมีความแม่นยำสูงสุด (94.03%) ตามด้วยเทคนิคต้นไม้ป่าสุ่ม (93.60%) และเทคนิคซัพพอร์ตเวกเตอร์แมชชีน (92.56%) อีกทั้งเทคนิคการวิเคราะห์การถดถอยโลจิสติกยังแสดงให้เห็นถึงประสิทธิภาพโดยรวมที่สูงกว่า (96.76%) ความไวที่ (99.51%) และความจำเพาะ (47.63%) การศึกษานี้บ่งชี้ถึงศักยภาพของเทคนิคการวิเคราะห์การถดถอยโลจิสติกในการคัดกรองภาวะสมองเสื่อม ซึ่งมีประสิทธิภาพเหนือกว่าเทคนิคต้นไม้ป่าสุ่มและเทคนิคซัพพอร์ตเวกเตอร์แมชชีน ในแง่ของความแม่นยำ ประสิทธิภาพ และการจำแนกประเภทที่สมดุล การค้นพบนี้สนับสนุนการพัฒนาเครื่องมือที่ใช้การเรียนรู้ของเครื่องเพื่อช่วยในการตัดสินใจทางคลินิกในการวินิจฉัยภาวะสมองเสื่อมและการวางแผนการรักษาอย่างมีประสิทธิภาพ

คำสำคัญ: ภาวะสมองเสื่อม; การเรียนรู้ของเครื่อง; การวิเคราะห์การถดถอยโลจิสติก; ประสิทธิภาพการจำแนก; การคัดกรองโรคสมองเสื่อม

บทนำ

ภาวะสมองเสื่อมเป็นปัญหาสุขภาพสำคัญที่มีผลกระทบเพิ่มขึ้นในประชากรผู้สูงอายุทั่วโลก การตรวจจับและการคัดกรองภาวะสมองเสื่อมอย่างแม่นยำในระยะแรกเป็นสิ่งสำคัญสำหรับการจัดการและการดูแลที่มีประสิทธิภาพ ภาวะสมองเสื่อมเป็นภาวะการลดลงของการทำงานทางปัญญาและสังคม ซึ่งส่งผลกระทบต่อหน่วยความจำ ภาษา การแก้ปัญหา และพฤติกรรม (Denning, 2023) นำไปสู่การพึ่งพาผู้อื่นสำหรับกิจกรรมประจำวัน (Werner, 2023) ความชุกของภาวะสมองเสื่อมเพิ่มขึ้นทั่วโลก โดยประมาณการคาดการณ์ว่าบุคคลที่ได้รับผลกระทบจะเพิ่มขึ้นอย่างมีนัยสำคัญภายในปี 2050 (Kumari, 2023)

ปัจจัยที่หลากหลาย เช่น โรคทางพันธุกรรม โรคหลอดเลือดสมอง และการติดเชื้อในสมอง ก็สามารถนำไปสู่ภาวะสมองเสื่อมได้ ภาวะสมองเสื่อมไม่เพียงแต่แสดงออกถึงการลดลงของความรู้เท่านั้น แต่ยังรวมถึงอาการทางพฤติกรรมและจิตใจด้วย แม้ว่าภาวะสมองเสื่อมจะไม่ติดต่อ แต่ความ

บกพร่องทางพันธุกรรมสามารถมีบทบาทในการพัฒนาของโรคในบางกรณี การค้นหาวิธีที่มีประสิทธิภาพและเชื่อถือได้ในการวินิจฉัยโรคในระยะเริ่มต้นเป็นสิ่งสำคัญสำหรับการจัดการและการดูแลผู้ที่มีภาวะสมองเสื่อมอย่างมีประสิทธิภาพ

ในประเทศไทย ภาวะสมองเสื่อมในหมู่ผู้สูงอายุกำลังเพิ่มขึ้นอย่างมีนัยสำคัญ ซึ่งส่งผลกระทบต่อผู้ป่วยและผู้ดูแล การศึกษาชี้ให้เห็นว่า อาการทางประสาทจิตเวช (Neuropsychiatric Symptoms, NPS) พบได้บ่อยในหมู่ผู้ป่วยภาวะสมองเสื่อม อาการเหล่านี้ ได้แก่ ความตื่นเต้น ความวิตกกังวล ภาพหลอน และภาวะซึมเศร้า ซึ่งส่งผลกระทบต่อคุณภาพชีวิตของผู้ป่วยและผู้ดูแล (Senanarong et al, 2023; Hinton et al, 2022) ความรุนแรงของ NPS เกี่ยวข้องกับความทุกข์ ภาระ และภาวะซึมเศร้าที่เพิ่มขึ้นในหมู่ผู้ดูแล ซึ่งเน้นถึงความจำเป็นในการประเมินและกลยุทธ์การจัดการที่มีประสิทธิภาพ (Wang et al, 2022; Hinton et al, 2022) นอกจากนี้ ความบกพร่องทางปัญญายังเป็นที่แพร่หลายในผู้สูงอายุที่เป็นโรคเบาหวานในประเทศไทย โดยประมาณ 54.5% ของผู้ป่วยเบาหวานประสบปัญหาทางปัญญา ซึ่งเน้นถึงความสำคัญของการจัดการสุขภาพทางปัญญา (Panyawattanakit et al, 2022)

เทคโนโลยีการเรียนรู้ของเครื่อง เช่น ป่าแบบสุ่ม การถดถอยโลจิสติก และเทคนิคการสนับสนุนเครื่องเวกเตอร์ ได้ปรับปรุงการวินิจฉัยและการรักษาทางการแพทย์อย่างมีนัยสำคัญ (Rogerson & Hall, 2023; Langenberger et al, 2023; An et al, 2023; Austin et al, 2022) ความก้าวหน้าเหล่านี้ได้ปฏิวัติการดูแลสุขภาพโดยการทำงานอัตโนมัติเช่นการจำแนกประเภทและการตรวจวินิจฉัยทางการแพทย์นำไปสู่ประสิทธิภาพและความแม่นยำในการวินิจฉัยที่ดีขึ้น (Domínguez-Rodríguez et al, 2022) ตัวอย่างเช่น โมเดล RF แสดงให้เห็นประสิทธิภาพที่เหนือกว่าในการทำนายผู้ป่วยที่มีต้นทุนสูง โดยมีประสิทธิภาพเหนือกว่า LR และเครือข่ายประสาทเทียม ในทำนองเดียวกัน LR และแบบจำลอง RF ถูกนำมาใช้อย่างมีประสิทธิภาพในการวิจัยบริการด้านสุขภาพเพื่อวิเคราะห์ผลลัพธ์สำหรับผู้ป่วยที่เป็นโรคหลอดเลือดแดงส่วนปลายและเบาหวาน แสดงให้เห็นถึงศักยภาพของการเรียนรู้ของเครื่องในการระบุปัจจัยที่ปรับเปลี่ยนได้เพื่อผลลัพธ์ของผู้ป่วยที่ดีขึ้น

ในบทความนี้นำเสนอการเปรียบเทียบประสิทธิภาพของเทคนิคการเรียนรู้ของเครื่องสามแบบ ได้แก่ ป่าแบบสุ่ม การถดถอยโลจิสติก และเทคนิคการสนับสนุนเครื่องเวกเตอร์ ในการคัดกรองภาวะสมองเสื่อมจากข้อมูลทางสุขภาพ เพื่อพัฒนาวิธีการที่มีประสิทธิภาพในการตรวจวินิจฉัยและจัดการภาวะสมองเสื่อมในระยะเริ่มต้น การค้นพบนี้สามารถนำไปสู่การพัฒนาเครื่องมือที่ใช้การเรียนรู้ของเครื่องเพื่อช่วยในการตัดสินใจทางคลินิกในการวินิจฉัยภาวะสมองเสื่อมและการวางแผนการรักษาอย่างมีประสิทธิภาพ

วัตถุประสงค์การวิจัย

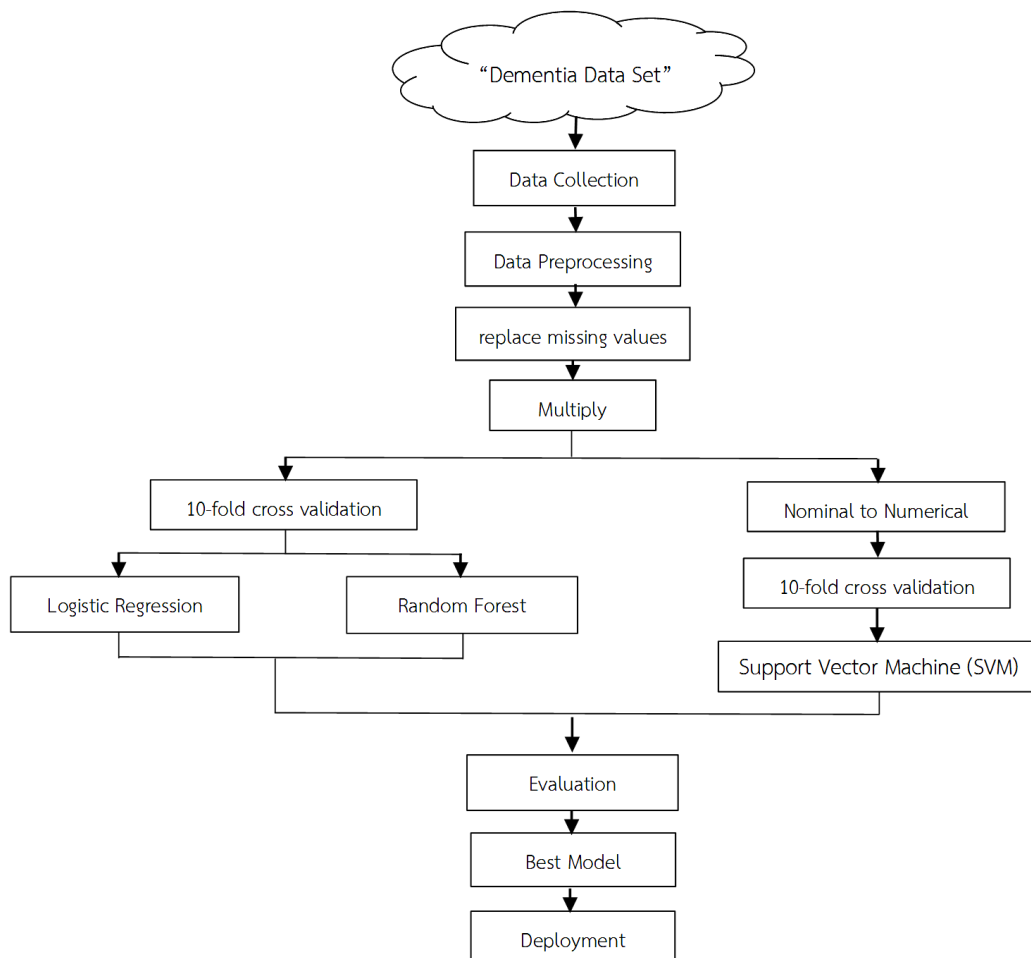
- 1) เพื่อสร้างตัวแบบสำหรับการคัดกรองผู้ป่วยโรคสมองเสื่อม
- 2) เพื่อเปรียบเทียบประสิทธิภาพของตัวแบบสำหรับการคัดกรองผู้ป่วยโรคสมองเสื่อม

วิธีดำเนินการวิจัย

ผู้วิจัยได้ดำเนินการรวบรวมข้อมูลจากเว็บไซต์ www.kaggle.com และนำข้อมูลมาผ่านกระบวนการเตรียมข้อมูลเพื่อให้ได้ชุดข้อมูลที่มีคุณภาพและเหมาะสมสำหรับการวิเคราะห์ต่อไป กระบวนการเตรียมข้อมูลรวมถึงการแทนค่าข้อมูลที่ขาดหายไปและการแปลงข้อมูลจากรูปแบบนามมาเป็นตัวเลข จากนั้นข้อมูลถูกขยายเพื่อเพิ่มจำนวนให้เพียงพอต่อการวิเคราะห์ โดยใช้เทคนิคการทำเหมืองข้อมูล (Data Mining) ในการพัฒนาตัวแบบการพยากรณ์โรคสมองเสื่อม

การวิจัยเริ่มต้นจากการรวบรวมข้อมูล (Data Collection) และดำเนินการเตรียมข้อมูล (Data Preprocessing) ซึ่งรวมถึงการแทนค่าข้อมูลที่ขาดหายไป (replace missing values) และการขยายจำนวนข้อมูล (Multiply) ข้อมูลที่ได้ถูกนำมาผ่านกระบวนการแปลงจากรูปแบบนามมาเป็นตัวเลข (Nominal to Numerical) หลังจากนั้นแบ่งข้อมูลเพื่อทำการทดสอบประสิทธิภาพของแบบจำลองด้วยวิธีการ Cross-Validation โดยใช้เทคนิคการจำแนกประเภทข้อมูล 3 เทคนิค ประกอบด้วย Random Forest (RF), Logistic Regression (LR), และ Support Vector Machine (SVM)

การประเมินผล (Evaluation) ของแบบจำลองทำโดยใช้เกณฑ์การวัดประสิทธิภาพ 4 ค่า ได้แก่ ค่าความแม่นยำ (Accuracy) ค่าประสิทธิภาพโดยรวม (F-measure) ค่าความไว (Sensitivity) และค่าความจำเพาะ (Specificity) เพื่อเปรียบเทียบประสิทธิภาพของตัวแบบต่างๆ โดยมุ่งหวังว่าจะได้ตัวแบบที่ดีที่สุด สำหรับการคัดกรองผู้ป่วยที่มีภาวะสมองเสื่อมและนำไปใช้จริง ดังแสดงในภาพที่ 1



ภาพที่ 1 ขั้นตอนการดำเนินงาน

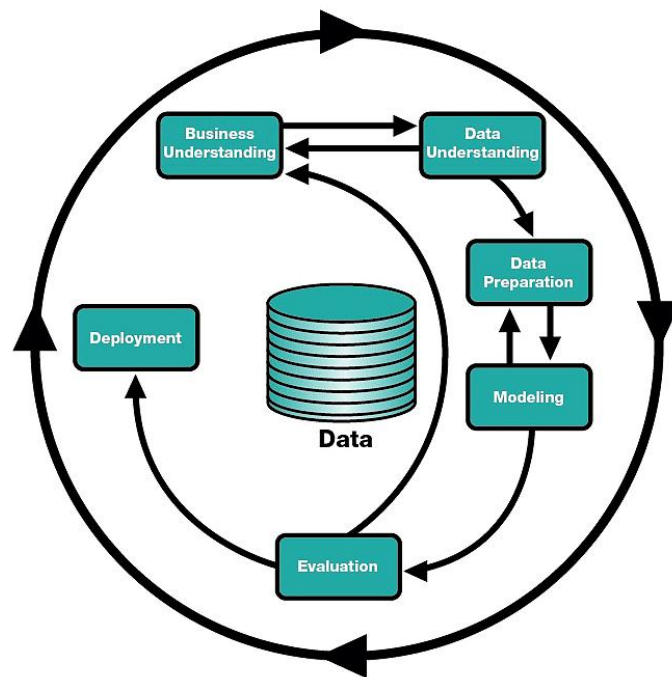
กระบวนการมาตรฐานในการทำเหมืองข้อมูล

การสร้างตัวแบบและการทดสอบประสิทธิภาพของตัวแบบสำหรับการคัดกรองผู้ป่วยภาวะสมองเสื่อม ผู้วิจัยได้ดำเนินการตามกระบวนการมาตรฐานของการทำเหมืองข้อมูล (Cross-Industry Standard Process for Data Mining: CRISP-DM) ประกอบด้วย 6 ขั้นตอน (อนุพงศ์ สุขประเสริฐ, 2024) ดังภาพที่ 2

1. การทำความเข้าใจกับปัญหา (Business Understanding)

ในการศึกษาเบื้องต้นผู้วิจัยได้พิจารณาถึงปัญหาเกี่ยวกับภาวะสมองเสื่อมซึ่งเป็นภาวะที่เกิดจากการทำงานลดลงของเซลล์สมองอย่างรวดเร็ว ทำให้เซลล์ฝ่อเหี่ยว ผลที่ตามมาคือการลดลงของความจำ ความคิด อารมณ์ และสติปัญญา อาการดังกล่าวส่งผลกระทบต่อคุณภาพชีวิตประจำวันและบุคลิกภาพของผู้ป่วย ปัจจุบันสาเหตุของภาวะสมองเสื่อมยังไม่ถูกพิสูจน์อย่างชัดเจน ซึ่งส่งผลให้การ

วินิจัยเกิดความท้าทาย ด้วยเหตุนี้ผู้วิจัยจึงมุ่งเน้นที่จะใช้เทคนิคการทำเหมืองข้อมูล (Data Mining) เพื่อพัฒนาแบบจำลองการพยากรณ์ที่สามารถทำนายความเสี่ยงของการเกิดภาวะสมองเสื่อม โดยหวังว่าผลลัพธ์จะสามารถช่วยเสริมสร้างความเข้าใจและประโยชน์ในการตัดสินใจของแพทย์ และปรับปรุงกระบวนการตรวจสอบสำหรับผู้ที่ได้รับการวินิจฉัยภาวะสมองเสื่อมในอนาคต



ภาพที่ 2 แสดงกระบวนการมาตรฐานการทำเหมืองข้อมูล (CRISP-DM)

หมายเหตุ. จาก <https://dslytics.wordpress.com/2017/11/15/crisp-dm/>

2. การทำความเข้าใจเกี่ยวกับข้อมูล (Data Understanding)

ผู้วิจัยได้ทำการรวบรวมข้อมูลจากเว็บไซต์ www.kaggle.com ซึ่งเป็นชุดข้อมูลที่สร้างโดย Mehrparvar, F., (2024) วิศวกรชีวการแพทย์, และสามารถเข้าถึงได้อย่างเปิดเผย ชุดข้อมูลนี้มีชื่อว่า “Optimal combined 3 studies” ถูกรวบรวมในปี พ.ศ. 2567 และมีข้อมูลทั้งหมด 1,842 แถว ในชุดข้อมูลนี้ประกอบด้วย 20 แอททริบิวต์ ดังแสดงในตารางที่ 1 โดยการวิเคราะห์ชุดข้อมูลนี้จะช่วยให้ผู้วิจัยสามารถสร้างตัวแบบและเปรียบเทียบประสิทธิภาพของตัวแบบสำหรับการคัดกรองภาวะสมองเสื่อมได้

3. การเตรียมข้อมูล (Data Preparation)

การเตรียมข้อมูล (Data Preparation) เป็นขั้นตอนสำคัญที่มีผลต่อคุณภาพและความถูกต้องของการวิเคราะห์ข้อมูลเพื่อพยากรณ์ผลลัพธ์ที่เกี่ยวข้องกับสุขภาพสมอง เริ่มต้นด้วยการกำหนดหน้าที่ของแอตทริบิวต์ต่าง ๆ ในชุดข้อมูล เพื่อให้เป็นไปตามความต้องการและวัตถุประสงค์ของการวิเคราะห์ ซึ่งได้แก่การกำหนดให้แอตทริบิวต์ที่ชื่อ "ID" ทำหน้าที่เป็นตัวแทนประจำแถวข้อมูล (Identifier) ซึ่งจะช่วยในการระบุและการติดตามข้อมูลของแต่ละแถว และแอตทริบิวต์ที่ชื่อ "CMB count" ใช้เป็นตัวเป้าหมาย (Label) สำหรับการพยากรณ์ผลลัพธ์ โดยที่แอตทริบิวต์อื่น ๆ ทำหน้าที่เป็นตัวแปรต้นที่จะนำมาใช้ในการวิเคราะห์ข้อมูล

ตารางที่ 1 ข้อมูลสำหรับการวิเคราะห์ข้อมูล

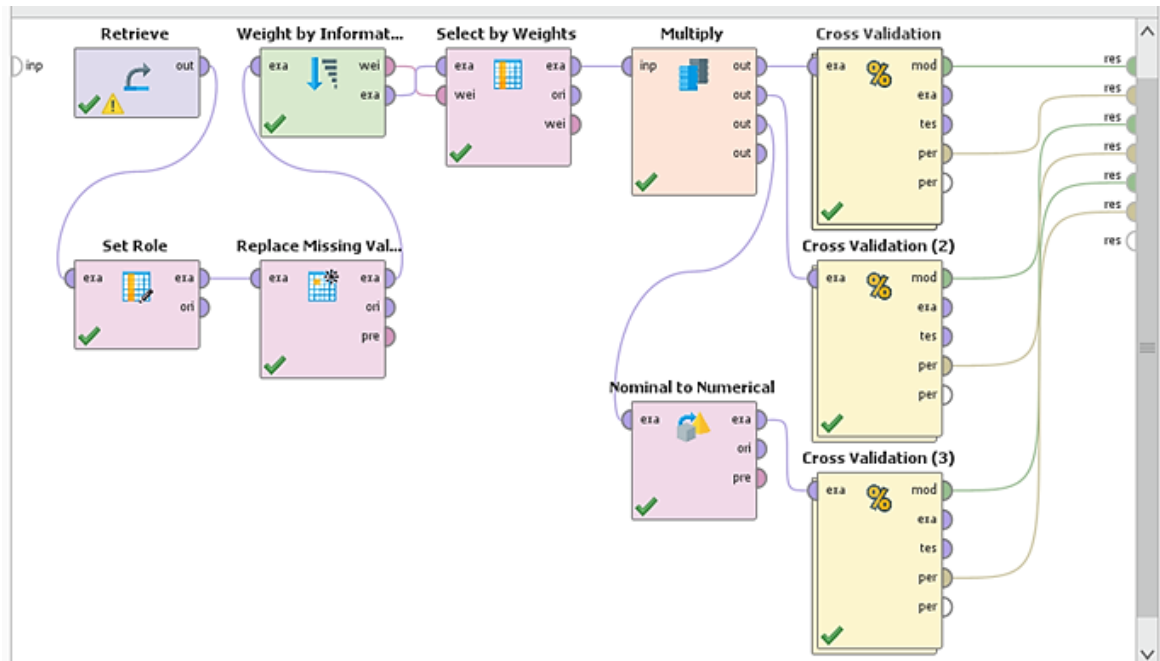
No	Name	Data Type	Description
1	ID	Integer	ลำดับที่
2	Age	Integer	อายุ
3	Gender	Binominal	เพศของผู้ป่วย
4	Dementia all	Binominal	การคาดการณ์ที่จะเป็นภาวะสมองเสื่อม
5	Educationyears	Integer	ระยะเวลาการศึกษาใช้เวลากี่ปี
6	EF: Executive function	Real	การประเมินการทำงานของความคิดและประสิทธิภาพของสมอง
7	PS: Processing speed	Real	ความเร็วในการประมวลผล ประเมินทักษะการรับรู้
8	Global	Real	คะแนนความจำ
9	Diabetes	Binominal	โรคเบาหวาน
10	Smoking	Polynominal	การสูบบุหรี่
11	hypertension	Binominal	ความดันโลหิตสูง
12	hypercholesterolemia	Binominal	ไขมันในเลือดสูง
13	lacunes num	Binominal	ภาวะสมองขาดเลือด
14	Fazekas cat	Binominal	หมวดหมู่ที่แสดงระดับความรุนแรงของโรคจากสารสีขาว
15	study	Polynominal	ข้อมูลการศึกษา
16	SVD Simple Score	Integer	คะแนนการวินิจฉัยโรคหลอดเลือดสมองตีบโดยง่าย
17	SVD Amended Score	Integer	คะแนนแก้ไขการวินิจฉัยโรคหลอดเลือดสมองตีบ
18	Fazekas	Integer	ระดับความรุนแรงของโรคจากสารสีขาว
19	lac count	Polynominal	หมวดหมู่ที่แสดงภาวะสมองขาดเลือด
20	CMB count	Binominal (Label)	ผลวินิจฉัยภาวะสมองเสื่อม

หลังจากนั้น เราได้ดำเนินการทำความสะอาดข้อมูล (Data Cleaning) ซึ่งเป็นขั้นตอนสำคัญในการเตรียมข้อมูล ซึ่งพบว่ามีค่าที่ไม่ทราบ (N/A) เกิดขึ้นในแอตทริบิวต์หลายตัว เช่น การประเมินการทำงานของความคิดและประสิทธิภาพของสมอง (EF: Executive Function) ความเร็วในการประมวลผลทักษะการรับรู้ (PS: Processing Speed) คะแนนความจำรวม (Global) คะแนนการวินิจฉัยโรคหลอดเลือดสมองตึบโดยง่าย (SVD Simple Score) และคะแนนแก้ไขการวินิจฉัยโรคหลอดเลือดสมองตึบ (SVD Amended Score) ในการจัดการค่าไม่ทราบ (N/A) เหล่านี้ ทางผู้วิจัยได้ทำการแทนค่าด้วยค่าฐานนิยม ซึ่งเป็นวิธีที่เหมาะสมและนิยมใช้ในกรณีที่ไม่มีข้อมูลที่จะช่วยในการแทนค่าอื่น ๆ ที่หายไป

ขั้นตอนสุดท้ายคือการตรวจสอบและปรับชนิดข้อมูลให้ตรงกับรูปแบบที่จัดเก็บในไฟล์ .xlsx หลังจากการเตรียมข้อมูลเสร็จสมบูรณ์ ข้อมูลจะพร้อมสำหรับการนำไปใช้ในการวิเคราะห์ต่อไป ดังที่แสดงในตารางที่ 1

4. การสร้างแบบจำลอง (Modeling)

ในขั้นตอนการสร้างตัวแบบ การใช้งานเทคนิคการทำเหมืองข้อมูลในการจำแนกประเภทข้อมูลได้ถูกดำเนินการผ่านโปรแกรม RapidMiner Studio เวอร์ชัน 10 การวิเคราะห์ในโครงการนี้ได้นำเทคนิคการจำแนกประเภทข้อมูล 3 ประเภท ได้แก่ Random Forest (RF), Logistic Regression (LR) และ Support Vector Machine (SVM) เพื่อสร้างและทดสอบตัวแบบการพยากรณ์สำหรับการคัดกรองผู้ป่วยที่มีความเสี่ยงต่อการเกิดภาวะสมองเสื่อม ผลลัพธ์ของการวิเคราะห์นี้จะได้รับการนำเสนอในรูปแบบกราฟิก ดังแสดงในภาพที่ 3



ภาพที่ 3 การสร้างและการเปรียบเทียบประสิทธิภาพของแบบจำลอง โดยโปรแกรม

RapidMiner Studio Version 10

4.1 เทคนิคต้นไม้ป่าสุ่ม (Random Forest: RF) อัลกอริทึมต้นไม้ป่าสุ่ม (RF) เป็นส่วนขยายของ Decision Trees ช่วยเพิ่มประสิทธิภาพและความแม่นยำโดยการใช้ต้นไม้จำนวนมากขึ้น RF ได้รับความนิยมในการเรียนรู้ของเครื่องสำหรับการวิเคราะห์โรค (Chrobak et al, 2023; Teja & Veeramani, 2022) พบว่าวิธี Random Forest Fuzzy Entropy (RFFE) เป็นวิธีที่มีประสิทธิภาพในการทำนายโรคเบาหวาน โดยมีความแม่นยำสูงถึง 98 เปอร์เซ็นต์ (Ruby et al, 2023) และความแม่นยำที่ดีขึ้นเมื่อเทียบกับต้นไม้ตัดสินใจการค้นพบเหล่านี้เน้นย้ำถึงความสำคัญของ RF ในการวิเคราะห์โรค แสดงให้เห็นถึงศักยภาพในการปฏิบัติกระบวนการวินิจฉัยและปรับปรุงผลลัพธ์ทางการแพทย์

4.2 เทคนิคการวิเคราะห์การถดถอยโลจิสติก (Logistic Regression: LR) เป็นวิธีการทางสถิติที่ใช้ในการทำนายความน่าจะเป็นของผลลัพธ์แบบไบนารี เช่น ใช่หรือไม่ โดยอิงจากตัวแปรอิสระอย่างน้อยหนึ่งตัวแปรที่มีอิทธิพลต่อเหตุการณ์ที่น่าสนใจมีคุณค่าเป็นพิเศษในการวิเคราะห์โรค เช่น โรคหัวใจและหลอดเลือด (Starbuck, 2023) แบบจำลอง LR ได้รับการออกแบบมาเพื่อประเมินโอกาสที่เหตุการณ์เกิดขึ้นโดยการสร้างสมการที่พิจารณาผลกระทบของตัวแปรอิสระต่อผลลัพธ์ (Sayed, 2022) วิธีนี้มีความสำคัญในการวินิจฉัยโรคในช่วงต้น ดังที่เห็นในการประยุกต์ใช้เทคนิค LR ในการจำแนกโรคหัวใจด้วยความแม่นยำ 87.10% (Ambrish et al, 2022) การถดถอยทางโลจิสติกส์นั้นแตกต่างจากการถดถอยเชิงเส้น เนื่องจากการได้รับการปรับแต่งสำหรับผลลัพธ์แบบไบนารี ทำให้เป็น

เครื่องมือที่มีประสิทธิภาพในด้านต่างๆ รวมถึงการดูแลสุขภาพ ซึ่งการคาดการณ์ผลลัพธ์เช่นการหัตถ์ของโรคหรืออัตราการรับกลับเป็นสิ่งจำเป็น

4.3 เทคนิคการซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine: SVM) ใช้กันอย่างแพร่หลายในการวิเคราะห์โรค ดำเนินการโดยการสร้างไฮเปอร์เพลนที่เหมาะสมเพื่อแยกข้อมูลออกเป็นหมวดหมู่ที่แตกต่างกันตามตัวแปรอินพุต (Lumbanraja et al, 2022; Hurtado et al, 2022) ประสิทธิภาพของ SVM ในการจำแนกโรคต่างๆ เช่นโรคเบาหวานโรคเรื้อรังโรคเบาหวานและภาวะเรื้อรังเช่นความดันโลหิตสูงและโรคอ้วนได้รับการแสดงให้เห็นจากการศึกษาต่างๆ (Chelladurai & Pandian, 2021; Prayogo et al, 2022) ความสามารถของ SVM ในการทำนายความรุนแรงของโรคอย่างแม่นยำและอำนวยความสะดวกในการวินิจฉัยในช่วงต้นมีความสำคัญในการปรับปรุงผลลัพธ์ทางการแพทย์และลดอัตราการเสียชีวิต นอกจากนี้ การใช้ประโยชน์ของ SVM ในแบบจำลองการจำแนกประเภทโรค รวมกับเทคนิคเช่นตรรกะแบบฟuzzy ได้แสดงให้เห็นผลลัพธ์ที่น่าประทับใจในการทำนายโรคเรื้อรังที่ตามปัจจัยเสี่ยงของแต่ละบุคคล โดยรวมแล้ว บทบาทของ SVM ในการวิเคราะห์โรคอยู่ที่ความสามารถในการแยกความแตกต่างระหว่างกลุ่มโรคได้อย่างมีประสิทธิภาพ ช่วยในการวินิจฉัยและการแทรกแซงการรักษาอย่างทันท่วงที

5. การประเมินผล (Evaluation)

ในการประเมินผลของตัวแบบที่พัฒนาขึ้น ผู้วิจัยได้ทำการแบ่งชุดข้อมูลออกเป็นสองส่วนหลัก: ชุดสำหรับการฝึกสอน (Training Set) และชุดสำหรับการทดสอบ (Testing Set) เพื่อประเมินประสิทธิภาพของตัวแบบที่ถูกสร้างขึ้น การประเมินนี้ใช้วิธีการ Cross Validation แบบ 10-fold ซึ่งเป็นการแบ่งข้อมูลออกเป็น 10 ส่วนเท่าๆ กันและใช้ส่วนหนึ่งเป็นชุดทดสอบในขณะที่ส่วนที่เหลือใช้เป็นชุดฝึกสอน เพื่อทดสอบประสิทธิภาพของตัวแบบซ้ำๆ 10 ครั้ง ในแต่ละครั้งจะสลับส่วนของชุดทดสอบไปเรื่อยๆ ผลการทดสอบประสิทธิภาพถูกวัดโดยใช้ 4 เกณฑ์คือ ค่าความแม่นยำ (Accuracy) ซึ่งเป็นตัววัดที่ชี้แจงถึงความสามารถของตัวแบบในการทำนายข้อมูลที่ถูกต้อง ค่าประสิทธิภาพโดยรวม (F-measure) ซึ่งเป็นค่าเฉลี่ยแบบถ่วงน้ำหนักของความแม่นยำและความระลึก ค่าความไว (Sensitivity) ที่วัดจำนวนของผู้ป่วยที่มีโรคและถูกตรวจพบโดยตัวแบบว่ามีโรค และค่าความจำเพาะ (Specificity) ที่วัดจำนวนของบุคคลปกติที่ไม่มีโรคและถูกตรวจพบโดยตัวแบบว่าไม่มีโรค ผลลัพธ์เหล่านี้ให้ข้อมูลที่สำคัญในการประเมินว่าตัวแบบสามารถใช้งานในสถานการณ์จริงเพื่อคัดกรองภาวะสมองเสื่อมได้อย่างไร

6. การนำไปใช้ (Deployment)

เมื่อการวิเคราะห์ข้อมูลในทุกขั้นตอนของกระบวนการทำเหมืองข้อมูลเสร็จสิ้น ผลลัพธ์ที่ได้จะเป็นการเลือกเทคนิคที่เหมาะสมที่สุดสำหรับการสร้างตัวแบบการจำแนกระดับความเสี่ยงของภาวะสมองเสื่อม การใช้งานตัวแบบนี้มีประโยชน์ในการจำแนกความเสี่ยงของภาวะสมองเสื่อมอย่างมีประสิทธิภาพ ซึ่งจะช่วยให้แพทย์สามารถลดระยะเวลาในการประเมินและวินิจฉัยได้ นอกจากนี้ในอนาคตตัวแบบดังกล่าวยังมีศักยภาพที่จะถูกประยุกต์ใช้เพื่อสร้างระบบสนับสนุนการตัดสินใจซึ่งสามารถช่วยเสริมการทำงานของแพทย์ในการจำแนกโรค และการจัดการกับผู้ที่เข้ารับการตรวจสอบภาวะสมองเสื่อม ผลลัพธ์จากการวิเคราะห์นี้ส่งเสริมให้เกิดความก้าวหน้าในการทำนายและจัดการกับภาวะสมองเสื่อมอย่างมีประสิทธิภาพ

ผลการวิจัย

จากการวิเคราะห์การสร้างตัวแบบ และการเปรียบเทียบประสิทธิภาพของตัวแบบสำหรับการคัดกรองผู้ป่วยภาวะสมองเสื่อม โดยใช้เทคนิคการจำแนกประเภทข้อมูล (Classification) จำนวน 3 เทคนิค ได้แก่ เทคนิค RF LR และ SVM จากนั้นทำการเปรียบเทียบประสิทธิภาพของการจำแนกประเภทข้อมูล ด้วยค่าความแม่นยำ ค่าประสิทธิภาพโดยรวม ค่าความไว และค่าความจำเพาะ เพื่อทำการศึกษาผลลัพธ์ของประสิทธิภาพตัวแบบแต่ละเทคนิคว่าตัวแบบใดเหมาะสำหรับการนำไปใช้คัดกรองผู้ป่วยภาวะสมองเสื่อมดีที่สุด ซึ่งผลการวิเคราะห์ข้อมูลปรากฏ ดังตารางที่ 2

ตารางที่ 2 การเปรียบเทียบประสิทธิภาพของการจำแนกประเภทข้อมูล

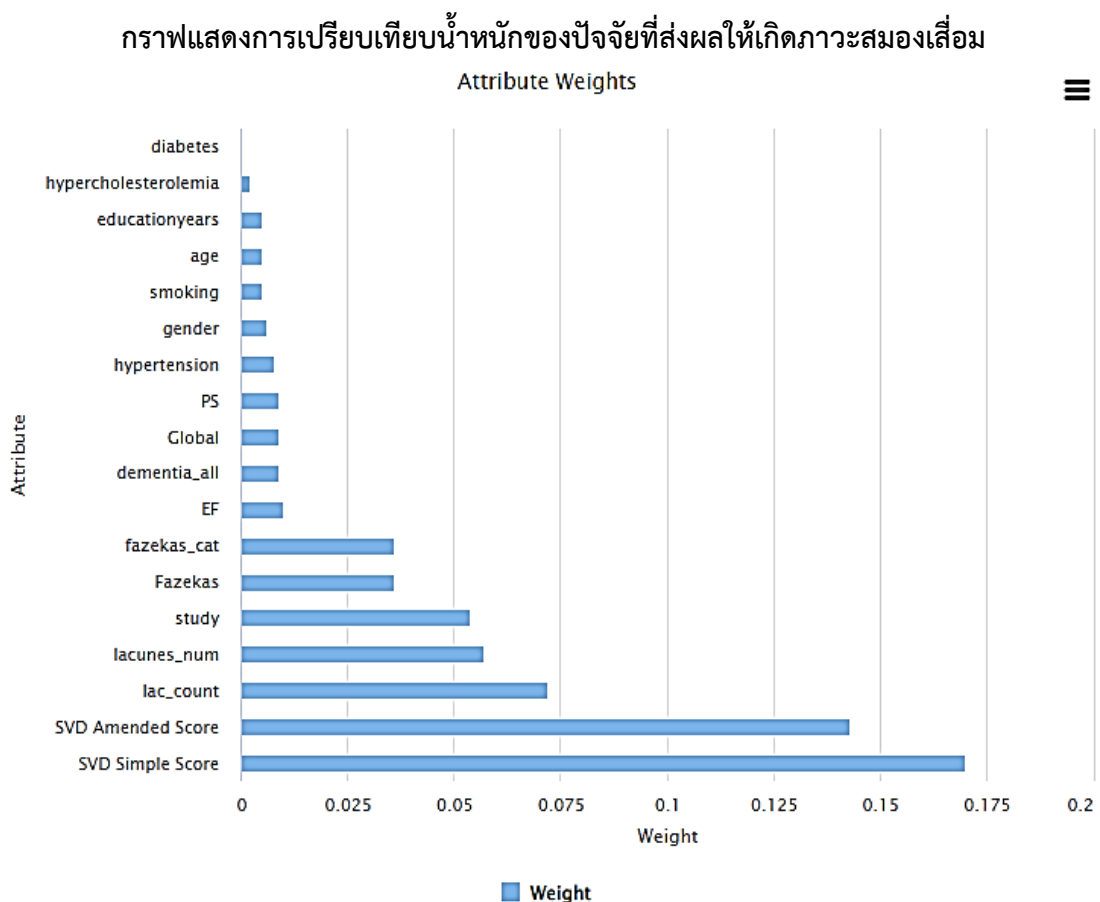
Classification Technique	Classification Performance			
	Accuracy	F-measure	Sensitivity	Specificity
Logistic Regression*	94.03	96.76	99.51	47.63
Random Forest	93.60	96.51	98.91	48.68
Support Vector Machine	92.56	95.92	97.69	49.21

* คือเทคนิคที่มีความเหมาะสมสำหรับการคัดกรองผู้ป่วยภาวะสมองเสื่อม ได้แก่ เทคนิค LR

จากตารางที่ 2 จะเห็นได้ว่าเทคนิคที่ให้ค่าความแม่นยำมากที่สุด คือ เทคนิค LR โดยให้ค่าความแม่นยำ เท่ากับ 94.03% ค่าประสิทธิภาพโดยรวม เท่ากับ 96.76% ค่าความไว เท่ากับ 99.51% ค่าความจำเพาะ เท่ากับ 47.63% การประยุกต์ใช้ตัวแบบที่มีความแม่นยำสูงนี้ในปฏิบัติการทางการแพทย์ไม่เพียงแต่ช่วยในการปรับแต่งยุทธศาสตร์การรักษาให้เหมาะสมกับแต่ละบุคคล แต่ยังช่วยใน

การคำนวณโอกาสในการพัฒนาภาวะสมองเสื่อม ซึ่งสามารถนำไปสู่การตัดสินใจที่ดีขึ้นเกี่ยวกับการรักษา การเปลี่ยนแปลงวิถีชีวิต และการให้คำปรึกษาทางการแพทย์

การวิจัยนี้เน้นการใช้เทคนิคการเรียนรู้ของเครื่อง ซึ่งให้ความแม่นยำสูงสุดในการทำนายภาวะสมองเสื่อม โดยการนำปัจจัยที่มีน้ำหนักสำคัญเหล่านี้มาเป็นปัจจัยหลัก (features) ตัวแบบที่พัฒนาขึ้นช่วยวินิจฉัยได้ละเอียดและครอบคลุมมากขึ้น จากภาพที่ 4 ปัจจัยคะแนนการวินิจฉัยโรคหลอดเลือดสมองตีบโดยง่าย (SVD Simple Score) มีน้ำหนัก (weight) สูงที่สุด บ่งชี้ว่ามีผลต่อการเกิดภาวะสมองเสื่อมมากที่สุด รองลงมาเป็นปัจจัยคะแนนแก้ไขการวินิจฉัยโรคหลอดเลือดสมองตีบ (SVD Amended Score), ปัจจัยหมวดหมู่ที่แสดงภาวะสมองขาดเลือด (lac count) ตัวแบบที่สร้างขึ้นสามารถจำแนกและประเมินความเสี่ยงของโรคได้อย่างละเอียดและครอบคลุมมากขึ้นการผสานข้อมูลจากหลากหลายแหล่งนี้ไม่เพียงแต่เพิ่มความแม่นยำในการคาดการณ์เท่านั้น แต่ยังช่วยให้การวินิจฉัยและการรักษามีประสิทธิภาพและรวดเร็วขึ้น ผลลัพธ์จากการวิจัยนี้มีศักยภาพที่จะนำไปสู่การพัฒนาวิธีการวินิจฉัยภาวะสุขภาพอื่นที่มีความซับซ้อน ทำให้การดูแลรักษาผู้ป่วยมีประสิทธิภาพมากขึ้น



ภาพที่ 4 การเปรียบเทียบน้ำหนักของปัจจัยที่ส่งผลให้เกิดภาวะสมองเสื่อม

อภิปรายผล

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อสร้างและเปรียบเทียบประสิทธิภาพของแบบจำลองสำหรับการคัดกรองผู้ป่วยภาวะสมองเสื่อม โดยเน้นการใช้เทคนิค Logistic Regression (LR) ซึ่งแสดงให้เห็นประสิทธิภาพที่เหนือกว่าในการทำนายภาวะสมองเสื่อม ด้วยค่าความแม่นยำที่ 94.03% ซึ่งสูงกว่าเทคนิค Random Forest (RF) และ Support Vector Machine (SVM) ที่ให้ค่าความแม่นยำ 93.60% และ 92.56% ตามลำดับ นอกจากนี้ ผลการวิจัยของเรายังแสดงให้เห็นว่าแบบจำลอง LR มีประสิทธิภาพสูงกว่าผลการวิจัยของ James et al. (2021) ที่ประเมินประสิทธิภาพของอัลกอริทึมการเรียนรู้ของเครื่องในการทำนายการพัฒนาไปสู่ภาวะสมองเสื่อมในผู้ป่วยคลินิกความจำ โดยในงานวิจัยของ James อัลกอริทึม Gradient Boosting Trees ให้ค่าความแม่นยำโดยรวม 92% ความไว 0.45 ความจำเพาะ 0.97 และพื้นที่ใต้เส้นโค้ง (AUC) 0.92

ผลลัพธ์ที่ดีกว่าในการศึกษาครั้งนี้สามารถอธิบายได้จากหลายปัจจัยสำคัญ ได้แก่ การใช้ชุดข้อมูลที่มีความหลากหลายและครอบคลุมมากขึ้น ซึ่งรวมถึงข้อมูลทางสุขภาพที่เกี่ยวข้องกับภาวะสมองเสื่อมและปัจจัยเสี่ยงอื่นๆ ทำให้แบบจำลองสามารถเรียนรู้และจับคู่ลักษณะพิเศษของข้อมูลกับประเภทของภาวะสมองเสื่อมได้ดีขึ้น การปรับแต่งอัลกอริทึมการวิเคราะห์อย่างละเอียด ทั้งในส่วนของการพารามิเตอร์และการเลือกตัวแปรที่เหมาะสม ทำให้แบบจำลองมีความแม่นยำสูงขึ้นในการทำนายผลลัพธ์ นอกจากนี้ การใช้เทคนิคการประมวลผลล่วงหน้า (preprocessing) และการเลือกตัวแปรที่เหมาะสม (feature selection) อย่างมีประสิทธิภาพช่วยเพิ่มความสามารถของแบบจำลองในการทำนายผลลัพธ์ที่แม่นยำ สะท้อนถึงการประยุกต์ใช้งานที่มีประสิทธิผลในสถานการณ์จริงได้ดีกว่าการศึกษาก่อนหน้า การใช้เทคนิคการวิเคราะห์ข้อมูลที่ทันสมัยและมีประสิทธิภาพมากขึ้น ซึ่งรวมถึงการใช้เทคนิคการเพิ่มประสิทธิภาพของอัลกอริทึม (algorithm optimization) ทำให้สามารถปรับปรุงประสิทธิภาพของแบบจำลองได้ดีกว่าผลงานของ James et al. (2021)

หากการวิจัยนี้ได้รับการพัฒนาต่อยอดและนำมาประยุกต์ใช้กับข้อมูลภาวะสมองเสื่อมของประเทศไทย จะสามารถขยายผลการทำนายที่มีความเที่ยงตรงและเป็นส่วนตัว ซึ่งจะไม่เพียงแต่เพิ่มความแม่นยำในการคัดกรองและวินิจฉัยภาวะสมองเสื่อมเท่านั้น แต่ยังช่วยให้การดูแลผู้ป่วยแต่ละรายมีความเหมาะสมมากขึ้นด้วยการคำนึงถึงลักษณะพิเศษและความต้องการของผู้ป่วย ผลจากการวิจัยนี้จะช่วยพัฒนามาตรฐานใหม่ในการวินิจฉัยและการรักษาภาวะสมองเสื่อมในประเทศไทย อีกทั้งยังช่วยให้ผู้ป่วยและครอบครัวสามารถวางแผนและเตรียมตัวสำหรับอนาคตได้ดีขึ้น

สำหรับการพัฒนาต่อไป การวิจัยควรรวมการประเมินผลกระทบของตัวแปรสังคม-เศรษฐกิจ และพฤติกรรมในการวินิจฉัยภาวะสมองเสื่อมเพิ่มเติม ซึ่งจะนำไปสู่การพัฒนาแนวทางการรักษาที่ครอบคลุมมากขึ้น นำไปสู่การรักษาที่ทันทั่วถึงและป้องกันผลกระทบระยะยาวจากโรคได้ดียิ่งขึ้น รวมถึงการนำแบบจำลองไปใช้ในสถานการณ์จริง โดยทดสอบความสามารถของแบบจำลองในการตัดสินใจทางคลินิกเพื่อประเมินศักยภาพการประยุกต์ใช้งานจริง พร้อมทั้งพัฒนาแบบจำลองการคัดกรองภาวะสมองเสื่อมแบบเฉพาะบุคคล ซึ่งสามารถปรับแต่งได้ตามความต้องการของแต่ละบุคคล เช่น อายุ ประวัติครอบครัว และข้อมูลสุขภาพพื้นฐานอื่นๆ เพื่อส่งเสริมการตรวจพบและการรักษาที่เหมาะสมและมีประสิทธิภาพยิ่งขึ้น ซึ่งเป็นความท้าทายในการพัฒนางานวิจัยต่อไป

กิตติกรรมประกาศ

คณะผู้วิจัยขอขอบคุณ www.kaggle.com สำหรับชุดข้อมูลการคัดกรองผู้ป่วยโรคสมองเสื่อมที่ใช้เป็นข้อมูลในการทดลองสำหรับงานวิจัยครั้งนี้ นอกจากนี้ คณะผู้วิจัยขอแสดงความขอบคุณอย่างสูงต่อคณะกรรมการบัญชีและการจัดการ มหาวิทยาลัยมหาสารคาม ที่ให้การสนับสนุนการวิจัยนี้จนสำเร็จลุล่วงอย่างเป็นผลสำเร็จ

เอกสารอ้างอิง

- อนุพงศ์ สุขประเสริฐ. (2024). คู่มือการทำเหมืองข้อมูล โปรแกรม RAPIDMINER STUDIO. พิมพ์ครั้งที่ 5. มหาสารคาม : สาขาคอมพิวเตอร์ธุรกิจ คณะการบัญชีและการจัดการ มหาวิทยาลัยมหาสารคาม.
- Ambrish, G., Ganesh, B., Ganesh, A., Srinivas, C., & Mensinkal, K. (2022). Logistic regression technique for prediction of cardiovascular disease. **Global Transitions Proceedings**, 3(1), 127-130.
<https://doi.org/10.1016/j.gltp.2022.04.008>
- An, Q., Rahman, S., Zhou, J., & Kang, J. J. (2023). A comprehensive review on machine learning in healthcare industry: classification, restrictions, opportunities and challenges. **Sensors**, 23(9), 4178. doi: 10.3390/s23094178
- Austin, A. M., Ramkumar, N., Gladders, B., Barnes, J. A., Eid, M. A., Moore, K. O., ... & Goodney, P. P. (2022). Using a cohort study of diabetes and peripheral artery disease to compare logistic regression and machine learning via random

forest modeling. **BMC Medical Research Methodology**, 22(1), 300. doi: 10.1186/s12874-022-01774-8

Chelladurai, U., & Pandian, S. (2021). Machine Learning based Early Prediction of Disease with Risk Factors Data of the Patient Using Support Vector Machines. **In Machine Learning, Deep Learning and Computational Intelligence for Wireless Communication: Proceedings of MDCWC 2020** (pp. 519-534). Springer Singapore. https://www.doi.org/10.1007/978-981-16-0289-4_38

Chrobak, D., Kołodziejczak, M., Kozłowska, P., Krzemińska, A., & Miller, T. (2023). Leveraging random forest techniques for enhanced microbiological analysis: a machine learning approach to investigating microbial communities and their interactions. **Scientific Collection «InterConf+»**, (32 (151)), 386-398.

Dening, K. H. (2023). Dementia: recognition and cognitive testing in community and primary care settings. **British Journal of Community Nursing**, 28(7), 332-336. <https://doi.org/10.12968/bjcn.2023.28.7.332>

Domínguez-Rodríguez, S., Serna-Pascual, M., Oletto, A., Barnabas, S., Zuidewind, P., Dobbels, E., ... & EPIICAL Consortium. (2022). Machine learning outperformed logistic regression classification even with limit sample size: A model to predict pediatric HIV mortality and clinical progression to AIDS. **Plos one**, 17(10), e0276116. doi: 10.1371/journal.pone.0276116

Hinton, L., Wang, K., Levkoff, S. E., Chuengsatiansup, K., Sihapark, S., Gallagher-Thompson, D., & Chen, H. (2022). Dementia neuropsychiatric prevalence, severity, and correlates in community-dwelling Thai older adults. **Alzheimer's & Dementia**, 18, <https://doi.org/10.1002/alz.066876>

Hurtado, R., Matute, J., & Boni, J. (2022). An analysis model for Machine Learning using Support Vector Machine for the prediction of Diabetic Retinopathy.

Artificial Intelligence and Social Computing, 28(28).

<http://doi.org/10.54941/ahfe1001450>

James, C., Ranson, J. M., Everson, R., & Llewellyn, D. J. (2021). Performance of machine learning algorithms for predicting progression to dementia in memory clinic patients. **JAMA network open**, 4(12), e2136553-e2136553. doi:10.1001/jamanetworkopen.2021.36553

Kumari, S., Bagri, K., & Deshmukh, R. (2023). Dementia: A journey from cause to cure. **In Nanomedicine-Based Approaches for the Treatment of Dementia** (pp. 37-56). Academic Press. <https://doi.org/10.1016/B978-0-12-824331-2.00011-X>

Langenberger, B., Schulte, T., & Groene, O. (2023). The application of machine learning to predict high-cost patients: A performance-comparison of different models using healthcare claims data. **PloS one**, 18(1), e0279540. doi: 10.1371/journal.pone.0279540

Lumbanraja, F. R., Lufiana, F., Heningtyas, Y., & Muludi, K. (2022). Implementasi Support Vector Machine (SVM) untuk Klasifikasi Penderita Diabetes Mellitus. **Jurnal Komputasi**, 10(1), 75-83. <http://repository.lppm.unila.ac.id/id/eprint/46370>

Mehrparvar, F., (2024) Dementia prediction. **Kaggle**. <https://www.kaggle.com/datasets/fatemehmehrparvar/dementia>

Panyawattanakit, C., Wongpradit, W., Kanhasing, R., & Kulalert, P. (2022). Cognitive impairment and associated factors among older adults with diabetes in a suburban primary health center in Thailand. **Dementia and Geriatric Cognitive Disorders**, 51(2), 175-181. <https://doi.org/10.1159/000524132>

Prayogo, R., Anggraeni, D., & Hadi, A. F. (2022). Classification of Cardiovascular Disease Gene Data Using Discriminant Analysis and Support Vector Machine (SVM). **BERKALA SAINSTEK**, 10(3), 124-132. <https://doi.org/10.19184/bst.v10i3.22259>

- Rogerson, C., & Hall, M. (2023). Methodological progress note: Machine learning methods in healthcare research. **Journal of Hospital Medicine**, 18(5), 431-434. doi: 10.1002/jhm.13078
- Ruby, A. U., Chandran, J. G. C., Jain, T. S., Chaithanya, B. N., & Patil, R. (2023). RFFE–Random Forest Fuzzy Entropy for the classification of Diabetes Mellitus. **AIMS Public Health**, 10(2), 422.
<https://doi.org/10.3934%2Fpublichealth.2023030>
- Sayed, A. H. (2022). **Logistic Regression. In Inference and Learning from Data: Learning** (pp. 2457–2498). chapter, Cambridge: Cambridge University Press.
- Senanarong, V., Rattanabannakit, C., Hunnangkul, S., Wongkom, N., Likitjaroen, Y., Witoonpanich, P., & Phanthumchinda, K. (2023). Five year dementia registry in Thailand: Regional distribution, etiologies, and outcome of dementia. **Alzheimer's & Dementia**, 19. <https://doi.org/10.1002/alz.061560>
- Starbuck, C. M. (2023) Logistic Regression. In: The Fundamentals of People Analytics. **Springer, Cham**. https://doi.org/10.1007/978-3-031-28674-2_12
- Teja, P. P. S., & Veeramani, T. (2022). Supervised study of Novel Random Forest Algorithm for prediction of heart disease in Comparison with The Decision Tree Algorithm. **Cardiometry**, (25), 1483-1490.
DOI:10.18137/cardiometry.2022.25.14831490
- Wang, K., Hinton, L., Levkoff, S., Sihapark, S., Chuengsatiansup, K., & Chen, H. (2022). PSYCHOMETRIC PROPERTIES OF THE NEUROPSYCHIATRIC INVENTORY QUESTIONNAIRE FOR THAI OLDER ADULTS WITH DEMENTIA. **Innovation in Aging**, 6(Supplement_1), 383-383.
<https://doi.org/10.1093/geroni/igac059.1511>

Werner, P. (2023). Like beauty and contact lenses, the meaning of dementia behavioral changes is in the eyes of the beholder. **International Psychogeriatrics**, 35(2), 59–61. doi:10.1017/S104161022200120X