

การเปรียบเทียบประสิทธิภาพของรูปแบบการแปลงข้อมูลร่วมกับ การจำแนกข้อมูลโรคหัวใจโดยใช้การเรียนรู้ของเครื่อง

ยศภัทร เรืองไพศาล*

สาขาวิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์และเทคโนโลยี
มหาวิทยาลัยเทคโนโลยีราชมงคลตะวันออก ชลบุรี

*Corresponding author email: yotsapat_ru@rmutto.ac.th

ได้รับบทความ: 10 พฤศจิกายน 2565

ได้รับบทความแก้ไข: 3 ธันวาคม 2565

ยอมรับตีพิมพ์: 20 ธันวาคม 2565

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษารูปแบบการแปลงนามบัญญัติเป็นตัวเลขเพื่อใช้ในการจำแนกข้อมูลการทำนายโรคหัวใจ 2 รูปแบบ (1) Unique integers และ (2) Dummy coding ร่วมกับการเรียนรู้ของเครื่อง 5 วิธี ได้แก่ วิธีแบบเบ้อย่างง่าย วิธีโครงข่ายประสาทเทียม วิธีต้นไม้ตัดสินใจ วิธีซัพพอร์ตเวกเตอร์แมชชีน และวิธีเพื่อนบ้านใกล้สุด ซึ่งใช้ชุดข้อมูลโรคหัวใจจากฐานข้อมูลออนไลน์ kaggle ที่มี 12 คุณลักษณะและผู้ถูกสังเกต 918 คน ประเมินประสิทธิภาพของการจำแนกข้อมูลด้วยวิธีการตรวจสอบไขว้กันที่ 10 โดยใช้โปรแกรม rapidMiner ผลการวิจัยพบว่า วิธีการ Unique integers และ Dummy coding ให้ผลลัพธ์ที่แตกต่างกันโดยมีค่าความถูกต้องอยู่ที่ร้อยละ ± 1.00 ซึ่งวิธี Unique integers ร่วมกับวิธีซัพพอร์ตเวกเตอร์แมชชีนให้ผลประสิทธิภาพการจำแนกประเภทข้อมูลที่ดีที่สุดที่ร้อยละ 87.58 รองลงมาเป็นวิธีโครงข่ายประสาทเทียมที่ร้อยละ 85.73 ส่วนของค่าความถูกต้องของวิธีแบบเบ้อย่างง่าย วิธีต้นไม้ตัดสินใจ และวิธีเพื่อนบ้านใกล้สุด มีค่าความถูกต้องที่ร้อยละ 84.42 ร้อยละ 84.09 และร้อยละ 71.45 ตามลำดับ

คำสำคัญ: การจำแนกข้อมูล / โรคหัวใจ / วิธีซัพพอร์ตเวกเตอร์แมชชีน /
วิธีโครงข่ายประสาทเทียม

Comparing the Performance of Data Conversion Model with Heart Disease Data Classification Using Machine Learning

Yotsapat Ruangpaisarn*

Computer Science, Faculty of Science and Technology,
Rajamangala University of Technology Tawan-ok, Chonburi

*Corresponding author email: yotsapat_ru@rmutto.ac.th

Received: 10 November 2022

Revised: 3 December 2022

Accepted: 20 December 2022

Abstract

This research aimed to study 2 types of nominal to numeric for heart disease prediction data classification. (1) Unique integers, and (2) Dummy coding, together with 5 machine learning methods including naive bayes, neural network, decision tree, support vector machine(SVM), and k-nearest neighbor(K-NN). The heart disease dataset was retrieved from the kaggle online database, which contains 12 features and 918 observers. The efficiency of our data classification was evaluated by 10-fold cross-validation using rapidminer. The results showed that the Unique integers and Dummy coding approaches gave different results with an accuracy of ± 1.00 percent. Unique integers with the SVM method provided the best data classification efficiency at 87.58%, followed by the neural network method at 85.73%. The accuracy of the naive bayes, decision tree, and K-NN methods were 84.42%, 84.09%, and 71.45% respectively.

Keywords: Classification / Heart Disease / SVM / Neural Network

บทนำ

สถานการณ์โลกในปัจจุบัน สาเหตุการตายของประชากรโลก ตั้งแต่ปี 2543 ถึง 2562 มีความชัดเจนในสาเหตุการตายด้วยโรคไม่ติดต่อ แตกต่างกับแนวโน้มที่ลดลงของการตายด้วยโรคติดต่อ สาเหตุการตายในลำดับต้นๆ คือ โรคหัวใจขาดเลือด โดยจากข้อมูลมรณบัตรในปี พ.ศ.2562 พบสาเหตุการเสียชีวิต 5 อันดับแรก คือ 1) โรคมะเร็งทุกชนิด 2) โรคหลอดเลือดสมอง 3) โรคปอดอักเสบ 4) โรคหัวใจขาดเลือด 5) อุบัติเหตุจากการคมนาคมขนส่งทางบก โดยในโรคหัวใจและหลอดเลือด (CVDs) ได้คร่าชีวิตผู้คนคิดเป็นร้อยละ 32 ของการเสียชีวิตทั่วโลก สำหรับประเทศไทย ในปี 2562 [1] มีผู้ป่วยจำนวน 350,922 ราย และมีผู้เสียชีวิตจำนวน 20,556 ราย ซึ่งในงานวิจัยของ รวิพรรณ หลักรัตน์ และคณะ[2] ที่เสนอว่าอัตราการเสียชีวิตด้วยโรคหลอดเลือดโคโรนารี จาก พ.ศ. 2556 เท่ากับ 202 คนต่อ 100,000 ประชากร เพิ่มขึ้นเป็น 204 คนต่อ 100,000 ประชากร ในปี พ.ศ. 2559 ดังนั้นผู้ที่เป็โรคหัวใจและหลอดเลือดหรือผู้ที่มีความเสี่ยงต่อโรคหัวใจและหลอดเลือดสูง จำเป็นต้องได้รับการตรวจหาและจัดการตั้งแต่นั้น ๆ

ในงานวิจัยที่เกี่ยวข้องกับการจำแนกข้อมูลโรคหัวใจ คมสัน แสนรักษ์ และคณะ[3] ใช้วิธีโครงข่ายประสาทเทียม วิธีแบบเบย์อย่างง่ายและวิธีเพื่อนบ้านใกล้สุดในการเปรียบเทียบประสิทธิภาพเทคนิคการจำแนกข้อมูลโรคหัวใจด้วยเทคนิคเหมืองข้อมูลสำหรับงานวิจัยการวิเคราะห์ข้อมูลและทำนายความเสี่ยงการเกิดโรคหัวใจด้วยเทคนิคการแบ่งกลุ่มและการหาความสัมพันธ์สำหรับการทำเหมืองข้อมูลของ สิริรัตน์ พิงชมภู และกิริติ อินทวิเศษ[4] ได้ใช้วิธีโครงข่ายประสาทเทียม วิธีต้นไม้ตัดสินใจ และวิธีเพื่อนบ้านใกล้สุด สำหรับวิธีการจำแนกข้อมูลกับโรคอื่นนอกเหนือจากโรคหัวใจ เช่น รุ่งโรจน์ บุญมา และนิเวศ จิระวิชิตชัย[5] ได้ใช้วิธีแบบเบย์อย่างง่าย วิธีเพื่อนบ้านใกล้สุด วิธีต้นไม้ตัดสินใจ และวิธีซัพพอร์ตเวกเตอร์แมชชีน ในงานวิจัยการจำแนกประเภทผู้ป่วยโรคเบาหวานโดยใช้เทคนิคเหมืองข้อมูลและการเลือกคุณลักษณะจากความสัมพันธ์ของข้อมูลของวราวุฒิ นาคบุญนำ [6] ใช้ 5 วิธีการจำแนกข้อมูล คือ วิธีแรนดอมฟลอเรส วิธีโครงข่ายประสาทเทียมแบบหลายชั้น วิธีซัพพอร์ตเวกเตอร์แมชชีน วิธีแบบเบย์อย่างง่าย วิธีโลจิสติกฟูลกลุ่ม ในการพยากรณ์ภาวะซึมเศร้าด้วยวิธีเหมืองข้อมูล

การศึกษาครั้งนี้ผู้วิจัยสนใจที่จะศึกษารูปแบบของการแปลงนามบัญญัติเป็นตัวเลข โดยกำหนดเป็นรูปแบบ Unique integers และแบบ Dummy coding ร่วมกับการเรียนรู้

ของเครื่อง 5 วิธี ได้แก่ วิธีแบบเบรียอย่างง่าย วิธีโครงข่ายประสาทเทียม วิธีต้นไม้ตัดสินใจ วิธีซัพพอร์ตเวกเตอร์แมชชีน และวิธีเพื่อนบ้านใกล้สุด พร้อมทั้งทดสอบแต่ละวิธีด้วยการปรับเปลี่ยนตัวเลือก (Optional) เพื่อหารูปแบบของการแปลงนามบัญญัติเป็นตัวเลขที่เหมาะสมกับวิธีการเรียนรู้ของเครื่อง

วัสดุและวิธีการ

ขั้นตอนการดำเนินงานวิจัย ประกอบด้วย 5 ขั้นตอน ดังนี้

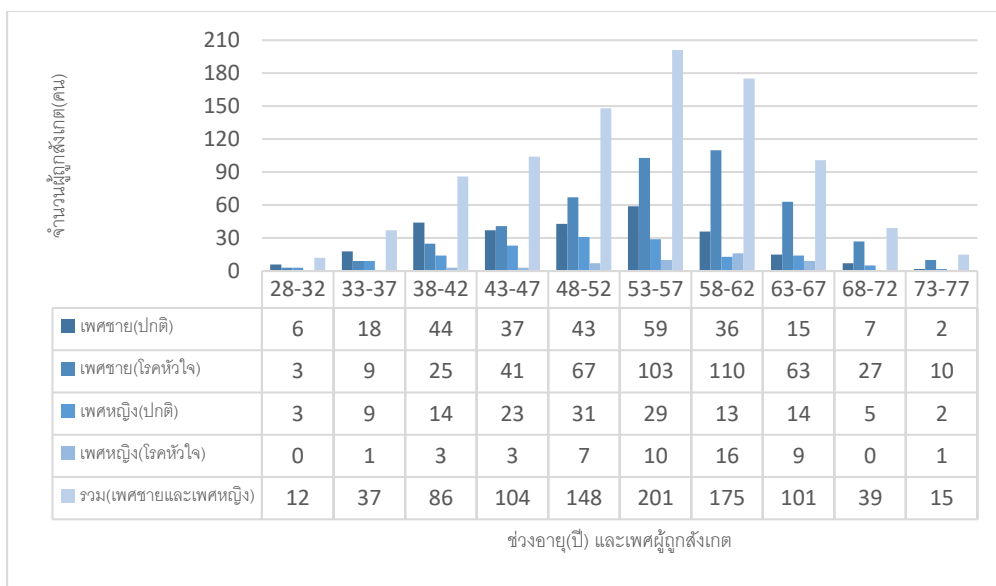
1. การศึกษาข้อมูล

ในงานวิจัยนี้ได้ใช้ชุดข้อมูลทำนายภาวะหัวใจล้มเหลว (Heart Failure Prediction Dataset) จากฐานข้อมูลออนไลน์ Kaggle[7] โดยชุดข้อมูลนี้ได้ถูกรวบรวมจากชุดข้อมูลต่างๆ ที่มีอยู่แล้วโดยอิสระ แต่ไม่เคยรวมกันมาก่อน ซึ่งในชุดข้อมูลนี้มี 12 คุณลักษณะข้อมูล ดังตารางที่ 1

ตารางที่ 1 รายละเอียดคุณลักษณะข้อมูล

ลำดับที่	ชื่อคุณลักษณะ	คำอธิบาย
1	Age	อายุ
2	Sex	เพศ (M คือ ผู้ชาย, F คือ ผู้หญิง)
3	ChestPainType	อาการเจ็บหน้าอก (Typical Angina : TA, Atypical Angina : ATA, Non-Anginal Pain : NAP, Asymptomatic : ASY) [8]
4	RestingBP	อัตราการเต้นของหัวใจขณะพัก
5	Cholesterol	คอเลสเตอรอล
6	FastingBS	ค่าความเข้มข้นน้ำตาลในเลือด
7	RestingECG	ผลการวัดคลื่นไฟฟ้าหัวใจในขณะพัก
8	MaxHR	ค่าชีพจรสูงสุด
9	ExerciseAngina	อาการเจ็บหน้าอกเวลาออกกำลังกาย (Y คือ ใช่, N คือ ไม่ใช่)
10	Oldpeak	ค่าความสูงของ ST ขณะพัก
11	ST_Slope	ค่าความชันของ ST
12	HeartDisease	ค่าคลาส (1 คือ โรคหัวใจ, 0 คือ ปกติ)

ในชุดข้อมูลนี้ มีจำนวนผู้ถูกสังเกต (Observations) จำนวน 918 คน แบ่งเป็นเพศชาย 725 คนและเพศหญิง 193 คน ตั้งแต่ช่วงอายุ 28 ถึง 77 ปี โดยช่วงอายุที่มีจำนวนผู้ถูกสังเกตมากที่สุดอยู่ในช่วงอายุ 53 ถึง 57 ปี ซึ่งมีจำนวน 201 คน และในช่วงอายุ 58 ถึง 62 ปี มีจำนวนเพศชายที่เป็นโรคหัวใจสูงที่สุด คือ 110 คน ดังภาพที่ 1 และตัวอย่างข้อมูลดิบ (Raw data) ดังภาพที่ 2



ภาพที่ 1 กราฟเปรียบเทียบช่วงอายุและเพศต่อจำนวนผู้ถูกสังเกต

Row No.	Age	Sex	ChestPainTy...	RestingBP	Cholesterol	FastingBS	RestingECG	MaxHR	ExerciseAn...	Oldpeak	ST_Slope	HeartDisease
1	40	M	ATA	140	289	0	Normal	172	N	0	Up	0
2	49	F	NAP	160	180	0	Normal	156	N	1	Flat	1
3	37	M	ATA	130	283	0	ST	98	N	0	Up	0
4	48	F	ASY	138	214	0	Normal	108	Y	1.500	Flat	1
5	54	M	NAP	150	195	0	Normal	122	N	0	Up	0
6	39	M	NAP	120	339	0	Normal	170	N	0	Up	0
7	45	F	ATA	130	237	0	Normal	170	N	0	Up	0
8	54	M	ATA	110	208	0	Normal	142	N	0	Up	0
9	37	M	ASY	140	207	0	Normal	130	Y	1.500	Flat	1
10	48	F	ATA	120	284	0	Normal	120	N	0	Up	0
11	37	F	NAP	130	211	0	Normal	142	N	0	Up	0
12	58	M	ATA	136	164	0	ST	99	Y	2	Flat	1

ภาพที่ 2 ตัวอย่างข้อมูลก่อนการเตรียมข้อมูล

2. การเตรียมข้อมูล

งานวิจัยนี้ได้ใช้โปรแกรมสำเร็จรูป RapidMiner [9] เวอร์ชัน 9.10.011 ในการเตรียมข้อมูล ซึ่งได้มีการปรับเปลี่ยนข้อมูลให้อยู่ในรูปแบบที่เหมาะสมกับการใช้งานโมเดลประเภทการจำแนกข้อมูล โดยมีการปรับเปลี่ยนข้อมูล 2 ส่วน ดังนี้

2.1 ข้อมูลคลาส เนื่องจากข้อมูลคลาสนั้นอยู่ในรูปแบบตัวเลข จึงใช้วิธีแปลงตัวเลขเป็นนามบัญญัติ (Numerical to Nominal)

2.2 ข้อมูลประเภทนามบัญญัติ (Nominal) ใช้วิธีแปลงนามบัญญัติเป็นตัวเลข (Nominal to Numerical) ซึ่งในขั้นตอนนี้ กำหนดรูปแบบของข้อมูลสำหรับการทดลองเป็น 2 ชุดข้อมูล คือ

2.2.1 ชุดข้อมูลแบบ Unique integers คือ การแปลงนามบัญญัติเป็นตัวเลขโดยการแทนค่าของคุณลักษณะของข้อมูลเริ่มจาก 0 ถึง n-1 โดยที่ n คือจำนวนของค่าที่เป็นไปได้ในคุณลักษณะข้อมูลนั้น เช่น เพศ(M=0, F=1) อาการเจ็บหน้าอก(ATA=0, NAP=1, ASY=2, TA=3) เป็นต้น ซึ่งข้อมูลหลังผ่านการเตรียมข้อมูลแบบ Unique integers แล้วเป็นดังภาพที่ 3

Row No.	HeartDisease	Sex	ChestPainTy...	RestingECG	ExerciseAn...	ST_Slope	Age	RestingBP	Cholesterol	FastingBS	MaxHR	Oldpeak
1	false	0	0	0	0	0	40	140	289	0	172	0
2	true	1	1	0	0	1	49	160	180	0	156	1
3	false	0	0	1	0	0	37	130	283	0	98	0
4	true	1	2	0	1	1	48	138	214	0	108	1.500
5	false	0	1	0	0	0	54	150	195	0	122	0
6	false	0	1	0	0	0	39	120	339	0	170	0
7	false	1	0	0	0	0	45	130	237	0	170	0
8	false	0	0	0	0	0	54	110	208	0	142	0
9	true	0	2	0	1	1	37	140	207	0	130	1.500
10	false	1	0	0	0	0	48	120	284	0	120	0
11	false	1	1	0	0	0	37	130	211	0	142	0
12	true	0	0	1	1	1	58	136	164	0	99	2

ภาพที่ 3 ตัวอย่างข้อมูลหลังการเตรียมข้อมูลแบบ Unique integers

2.2.2 ชุดข้อมูลแบบ Dummy coding คือ การแปลงนามบัญญัติเป็นตัวเลขโดยนำค่าที่เป็นไปได้ในคุณลักษณะของข้อมูลเปลี่ยนให้เป็นคุณลักษณะของข้อมูล โดยค่าในแต่ละคุณลักษณะของข้อมูลใหม่นั้น จะมีค่าได้เพียง 2 ค่า คือ 0 และ 1 ซึ่งข้อมูลหลังผ่านการเตรียมข้อมูลแบบ Dummy coding แล้วเป็นดังภาพที่ 4

Row No.	HeartDisease	Sex = M	Sex = F	ChestPainTy...	ChestPainTy...	ChestPainTy...	ChestPainTy...	RestingECG ...	RestingECG ...	RestingECG ...	ExerciseAn...	ExerciseAn...
1	false	1	0	1	0	0	0	1	0	0	1	0
2	true	0	1	0	1	0	0	1	0	0	1	0
3	false	1	0	1	0	0	0	0	1	0	1	0
4	true	0	1	0	0	1	0	1	0	0	0	1
5	false	1	0	0	1	0	0	1	0	0	1	0
6	false	1	0	0	1	0	0	1	0	0	1	0
7	false	0	1	1	0	0	0	1	0	0	1	0
8	false	1	0	1	0	0	0	1	0	0	1	0
9	true	1	0	0	0	1	0	1	0	0	0	1
10	false	0	1	1	0	0	0	1	0	0	1	0
11	false	0	1	0	1	0	0	1	0	0	1	0
12	true	1	0	1	0	0	0	0	1	0	0	1

ภาพที่ 4 ตัวอย่างข้อมูลหลังการเตรียมข้อมูลแบบ Dummy coding

3. การออกแบบโมเดล

ในงานวิจัยได้มีการใช้งานโมเดลการจำแนกข้อมูล (Classification) ได้แก่ วิธีแบบเบย์อย่างง่าย วิธีการโครงข่ายประสาทเทียม วิธีต้นไม้ตัดสินใจ วิธีซัพพอร์ตเวกเตอร์แมชชีน และวิธีเพื่อนบ้านใกล้สุด ซึ่งในแต่ละโมเดลได้มีการปรับเปลี่ยนตัวเลือก(Optional) และกำหนดข้อที่ใช้ทดสอบ ดังตารางที่ 2

ตารางที่ 2 คุณลักษณะของโมเดลที่ปรับเปลี่ยนสำหรับการทดสอบ

Model	Optional
Naive Bayes	-
Neural Network	Learning rate = 0.01, momentum = 0.9
Decision tree 1	Criterion = gain_ratio, confidence = 0.25
Decision tree 2	Criterion = information_gain, confidence = 0.25
Decision tree 3	Criterion = gini_index, confidence = 0.25
SVM 1	Kernel type = dot
SVM 2	Kernel type = radial
SVM 3	Kernel type = polynomial
SVM 4	Kernel type = neural
SVM 5	Kernel type = anova
K-NN 1	K = 5
K-NN 2	K = 7
K-NN 3	K = 9
K-NN 4	K = 11

4. การทดสอบประสิทธิภาพของโมเดล

ในการทดสอบประสิทธิภาพของโมเดล ผู้วิจัยได้ใช้การตรวจสอบไขว้กัน (Cross validation) ที่ 10 และมีการวัดประสิทธิภาพของการจำแนกข้อมูล คือ ค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และ ค่าความถ่วงดุล (F-measure)

ผลการศึกษา

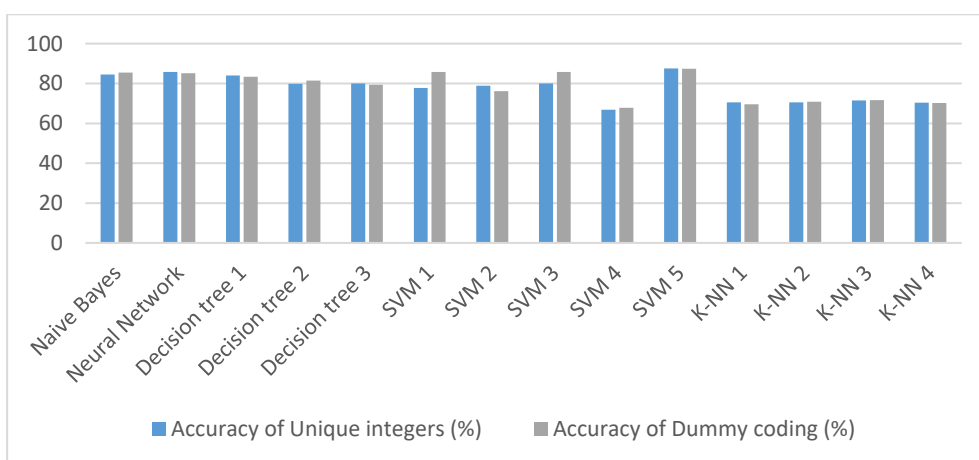
จากการทดลองตามโมเดลที่เลือกไว้ข้างต้น สามารถอธิบายผลการทดลองได้ดังนี้ ผลการเปรียบเทียบประสิทธิภาพในการจำแนกประเภทข้อมูลของแต่ละโมเดล กับชุดข้อมูลผ่านการเตรียมข้อมูลแบบ Unique integers และชุดข้อมูลผ่านการเตรียมข้อมูลแบบ Dummy coding เป็นดังตารางที่ 3 และตารางที่ 4 ตามลำดับ และผลการเปรียบเทียบประสิทธิภาพการจำแนกข้อมูลการทำนายโรคหัวใจแต่ละโมเดลกับชุดข้อมูลผ่านการเตรียมข้อมูลแบบ Unique integers และแบบ Dummy coding ดังภาพที่ 5

ตารางที่ 3 การเปรียบเทียบประสิทธิภาพในการจำแนกประเภทข้อมูลของแต่ละโมเดลกับชุดข้อมูลผ่านการเตรียมข้อมูลแบบ Unique integers

Model	Accuracy (%)	Precision (%)	Recall (%)	F-Measure (%)
Naive Bayes	84.42	86.03	86.03	85.95
Neural Network	85.73	86.91	87.39	87.09
Decision tree 1	84.09	85.91	85.63	85.64
Decision tree 2	79.86	83.48	79.56	81.34
Decision tree 3	79.96	82.64	81.12	81.76
SVM 1	77.78	89.08	68.29	77.11
SVM 2	78.87	75.43	91.93	82.84
SVM 3	80.06	86.70	75.58	80.68
SVM 4	66.78	70.51	69.09	69.66
SVM 5	87.58	87.35	90.93	89.02
K-NN 1	70.47	72.85	74.39	73.41
K-NN 2	70.58	73.31	73.41	73.20
K-NN 3	71.45	73.04	76.37	74.54
K-NN 4	70.37	72.35	74.81	73.34

ตารางที่ 4 การเปรียบเทียบประสิทธิภาพในการจำแนกประเภทข้อมูลของแต่ละโมเดลกับชุดข้อมูลผ่านการเตรียมข้อมูลแบบ Dummy coding

Model	Accuracy (%)	Precision (%)	Recall (%)	F-Measure (%)
Naive Bayes	85.41	87.01	86.84	86.78
Neural Network	85.08	84.97	88.97	86.84
Decision tree 1	83.34	83.39	87.42	85.30
Decision tree 2	81.49	84.60	81.91	82.98
Decision tree 3	79.31	83.43	78.74	80.74
SVM 1	85.73	87.54	86.79	87.01
SVM 2	76.14	71.64	94.49	81.45
SVM 3	85.73	86.18	88.57	87.28
SVM 4	67.76	81.22	57.25	65.61
SVM 5	87.37	87.09	90.93	88.85
K-NN 1	69.61	71.64	74.62	72.99
K-NN 2	70.81	73.06	74.82	73.86
K-NN 3	71.68	73.52	76.20	74.76
K-NN 4	70.15	72.15	75.01	73.49



ภาพที่ 5 ผลการเปรียบเทียบประสิทธิภาพการจำแนกข้อมูลการทำนายโรคหัวใจแต่ละโมเดลกับข้อมูลผ่านการเตรียมข้อมูลแบบ Unique integers และแบบ Dummy coding

วิจารณ์

ในงานวิจัยนี้ ผู้วิจัยได้ศึกษางานวิจัยที่เกี่ยวข้องกับประสิทธิภาพเทคนิคของการจำแนกข้อมูล ซึ่งผลการวิจัยเรียงจากมากไปหาน้อยลำดับค่าความถูกต้องของการจำแนกข้อมูล คือ วิธีซัพพอร์ตเวกเตอร์แมชชีน วิธีโครงข่ายประสาทเทียม วิธีแบบเบย์อย่างง่าย วิธีต้นไม้ตัดสินใจ และวิธีเพื่อนบ้านใกล้สุด ซึ่งสอดคล้องกับ คมสัน แสนรักษ์ และคณะ[3] ได้ทำการวิจัยเรื่องเปรียบเทียบประสิทธิภาพเทคนิคการจำแนกข้อมูลโรคหัวใจด้วยเทคนิคเหมือนข้อมูล ได้ผลวิจัยพบว่า วิธีโครงข่ายประสาทเทียมให้ค่าความถูกต้องร้อยละ 82.17 รองลงมาคือ วิธีแบบเบย์อย่างง่ายให้ค่าความถูกต้องร้อยละ 81.88 และวิธีการเพื่อนบ้านใกล้สุดให้ค่าความถูกต้องร้อยละ 65.66 ในงานวิจัยถัดมาการวิเคราะห์ข้อมูลและทำนายความเสี่ยงการเกิดโรคหัวใจด้วยเทคนิคการแบ่งกลุ่มและการหาความสัมพันธ์สำหรับการทำเหมือนข้อมูลของสิริรัตน์ พิงชมภู และกิริติ อินทวิเศษ[4] ผลวิจัยพบว่า วิธีโครงข่ายประสาทเทียมให้ค่าความถูกต้องร้อยละ 76.50 รองลงมาคือ วิธีต้นไม้ตัดสินใจให้ค่าความถูกต้องร้อยละ 81.88 และวิธีการเพื่อนบ้านใกล้สุด ให้ค่าความถูกต้องร้อยละ 65.66

สรุป

การศึกษารูปแบบการแปลงนามบัญญัติเป็นตัวเลขเพื่อใช้ในการจำแนกข้อมูลการทำนายโรคหัวใจร่วมกับโมเดลการจำแนกข้อมูล 5 วิธี คือ วิธีแบบเบย์อย่างง่าย วิธีโครงข่ายประสาทเทียม วิธีต้นไม้ตัดสินใจ วิธีซัพพอร์ตเวกเตอร์แมชชีน และวิธีเพื่อนบ้านใกล้สุด ร่วมกับการปรับเปลี่ยนตัวเลือกในแต่ละโมเดล พบว่า 1) การแปลงนามบัญญัติเป็นตัวเลขแบบ Unique integers และแบบ Dummy coding ในแต่ละวิธีให้ค่าความถูกต้องของการจำแนกข้อมูลที่ไม่แตกต่างกัน โดยส่วนใหญ่แตกต่างกันไม่เกินร้อยละ 1.00 ยกเว้นในการจำแนกข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีนที่ใช้ Kernel type เป็น dot และ polynomial ที่ใช้ชุดข้อมูลการแปลงนามบัญญัติเป็นตัวเลขแบบ Dummy coding ให้ค่าความถูกต้องสูงกว่าแบบ Unique integers เกินกว่าร้อยละ 5 2) โมเดลการจำแนกข้อมูลเพื่อทำนายโรคหัวใจที่เหมาะสมโดยให้ค่าความถูกต้องในระดับที่เกินกว่าร้อยละ 80.00 เรียงลำดับจากมากไปหาน้อย คือ วิธีซัพพอร์ตเวกเตอร์แมชชีนด้วยการใช้ Kernel type ที่มีค่าเป็น anova ให้ค่าความถูกต้องสูงที่สุด คิดเป็นร้อยละ 87.58 รองลงมาเป็นวิธีโครงข่ายประสาทเทียม ให้ค่าความถูกต้องคิดเป็นร้อยละ 85.73 วิธีแบบเบย์อย่างง่ายให้ค่าความ

ถูกต้องสูงคิดเป็นร้อยละ 84.42 และวิธีต้นไม้ตัดสินใจให้ค่าความถูกต้องสูงคิดเป็นร้อยละ 84.09 สำหรับวิธีเพื่อนบ้านใกล้สุด ให้ค่าความถูกต้องน้อยที่สุดในระดับร้อยละ 70.00

เอกสารอ้างอิง

1. สถาบันการแพทย์ฉุกเฉินแห่งชาติ [อินเทอร์เน็ต]. สถานการณ์และแนวโน้มสุขภาพและการแพทย์ฉุกเฉิน(ระดับโลกและประเทศไทย). [เข้าถึงเมื่อ 7 พ.ย. 2565]. เข้าถึงได้จาก:https://www.niems.go.th/1/UploadAttachFile/2022/EBook/414764_20220208161448.pdf
2. รวีพรรณ หลักรัตน์ พยอม สุขเอนกนันท์ และอุษาวดี สุตะภักดิ์. อัตราการเสียชีวิตด้วยโรคหลอดเลือดโคโรนารีโรงพยาบาลสมเด็จพระยุพราชเดชอุดมจังหวัดอุบลราชธานี ใน พ.ศ. 2556 และ พ.ศ. 2559. เกษตรศาสตร์อีสาน 2562;15(1):93-104.
3. คมสัน แสนรักษ์, ศิริพรรณ กันทะมูล และไกรรุ่ง เสงพระพรหม. การเปรียบเทียบประสิทธิภาพเทคนิคการจำแนกข้อมูลโรคหัวใจด้วยเทคนิคเหมืองข้อมูล. การประชุมวิชาการระดับชาติ มหาวิทยาลัยราชภัฏนครปฐม ครั้งที่ 12 ประจำปี 2563; มหาวิทยาลัยราชภัฏนครปฐม.
4. สิริรัตน์ พิงชมภู และกิริติ อินทวิเศษ. การวิเคราะห์ข้อมูลและทำนายความเสี่ยงการเกิดโรคหัวใจด้วยเทคนิคการแบ่งกลุ่มและการหาความสัมพันธ์สำหรับการทำเหมืองข้อมูล. การประชุมวิชาการระดับชาติ มหาวิทยาลัยทักษิณ ครั้งที่ 30 ประจำปี 2563; สถาบันวิจัยและพัฒนา มหาวิทยาลัยทักษิณ.
5. รุ่งโรจน์ บุญมาและนิเวศ จิระวิชุดชัย. การจำแนกประเภทผู้ป่วยโรคเบาหวานโดยใช้เทคนิคเหมืองข้อมูลและการเลือกคุณลักษณะจากความสัมพันธ์ของข้อมูล. วารสารวิชาการชาชนยเขต มหาวิทยาลัยราชภัฏภูเก็ต 2562;3(2):11-19.
6. วรารุณี นาคบุญนำ. การพยากรณ์ภาวะซึมเศร้าด้วยวิธีเหมืองข้อมูล. [วิทยานิพนธ์ปริญญาปรัชญาดุษฎีบัณฑิต]. มหาสารคาม: มหาวิทยาลัยมหาสารคาม; 2564.
7. fedesoriano. (September 2021). Heart Failure Prediction Dataset.Retrieved [Date Retrieved] from <https://www.kaggle.com/fedesoriano/heart-failure-prediction>.
8. แนวเวชปฏิบัติการดูแลรักษาผู้ป่วยภาวะหัวใจขาดเลือดเรื้อรัง พ.ศ.2564. พิมพ์ครั้งที่ 1. สมาคมแพทย์โรคหัวใจแห่งประเทศไทย ในพระบรมราชูปถัมภ์. 2565.
9. ธรรมนูญก นิลมณี และจรัญ แสนราช. การเปรียบเทียบประสิทธิภาพโมเดลการจำแนกข้อมูลเพื่อทำนายความสอดคล้องอาชีพกับสาขาที่เรียนของนักศึกษาระดับปริญญาตรี

คณะบริหารธุรกิจ มหาวิทยาลัยเทคโนโลยีราชมงคลรัตนโกสินทร์. วารสารก้าวหน้าทันโลก
วิทยาศาสตร์ 2564; 21(2):181-193.