

การเปรียบเทียบประสิทธิภาพในการทำนายผลระดับความรุนแรง
ของอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019

EFFICIENCY COMPARISON OF SEVERE PREDICTION OF
CORONAVIRUS DISEASE 2019 PATIENTS



ปัญหาพิเศษนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาวิทยาศาสตรบัณฑิต (สถิติประยุกต์)
ภาควิชาสถิติ คณะวิทยาศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
ปีการศึกษา 2566

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

EFFICIENCY COMPARISON OF SEVERE PREDICTION OF
CORONAVIRUS DISEASE 2019 PATIENTS



A SPECIAL PROBLEM SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENT FOR
THE DEGREE OF BACHELOR OF SCIENCE (APPLIED STATISTICS)
DEPARTMENT OF STATISTICS, SCHOOL OF SCIENCE
KING MONGUT'S INSTITUTE OF TECHNOLOGY LADKRABANG




เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้ **ACADEMIC YEAR 2023** ภาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อปัญหาพิเศษ การเปรียบเทียบประสิทธิภาพในการทำนายผลระดับความรุนแรงของ
 อาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019
 Efficiency Comparison of Severe Prediction of Coronavirus
 Disease 2019 Patients

ชื่อนักศึกษา นางสาวฐานิตา เงามาม รหัสนักศึกษา 63050618
 นางสาวณิพัทร เพียงลิ้ม รหัสนักศึกษา 63050621
 นายรชต พรรณโรจน์ รหัสนักศึกษา 63050655

ปริญญา วิทยาศาสตรบัณฑิต (สถิติประยุกต์)
ภาควิชา สถิติ
คณะ วิทยาศาสตร์
ปีการศึกษา 2566
อาจารย์ที่ปรึกษา รศ.สายชล สีนสมบูรณ์ทอง

คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง (สจล.) อนุมัติให้
 ปัญหาพิเศษนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรบัณฑิต (สถิติประยุกต์)
 ประจำปีการศึกษา 2566

คณะกรรมการตรวจสอบ	ลายมือชื่อ
ผศ.ดร.พรรณทิพา วาณิชยจิรัฐติกาล ประธานกรรมการ	
ดร.ธัญรดา ชัยขจรวัฒน์ กรรมการ	
รศ.สายชล สีนสมบูรณ์ทอง กรรมการและอาจารย์ที่ปรึกษา	

ลิขสิทธิของคณะวิทยาศาสตร์
 สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อปัญหาพิเศษ	การเปรียบเทียบประสิทธิภาพในการทำนายผลระดับความรุนแรงของอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019	
ชื่อนักศึกษา	นางสาวฐานิตา เกางาม	รหัสนักศึกษา 63050618
	นางสาวณิพัทร เพียงลิ้ม	รหัสนักศึกษา 63050621
	นายรชต พรรณโรจน์	รหัสนักศึกษา 63050655
ปริญญา	วิทยาศาสตร์บัณฑิต (สถิติประยุกต์)	
ภาควิชา	สถิติ	
คณะ	วิทยาศาสตร์	
ปีการศึกษา	2566	
อาจารย์ที่ปรึกษา	รศ.สายชล สีนสมบุรณ์ทอง	

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อทำนายปัจจัยที่มีผลต่อระดับความรุนแรงของอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 และเปรียบเทียบประสิทธิภาพในการทำนายผลระดับความรุนแรงของอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 ด้วยการวิเคราะห์การถดถอยลอจิสติกทวิภาคและโครงข่ายประสาทเทียม โดยใช้ข้อมูลของผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 ที่เข้ารับการรักษาที่กรมอนามัย กระทรวงสาธารณสุขแห่งหนึ่งในกรุงเทพมหานคร ระหว่างเดือนมีนาคม-ธันวาคม พ.ศ. 2565 คัดเลือกกลุ่มตัวอย่างแบบเจาะจงคือผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 ที่มีผลตรวจ ATK เป็นบวก ซึ่งเก็บข้อมูลจากผู้ป่วยโดยตรง ดำเนินการโดยเจ้าหน้าที่กรมอนามัยทำการสอบถามอาการของผู้ป่วยตามแบบสอบถามผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 ได้ข้อมูลจำนวน 652 ชุด โดยแบ่งข้อมูลเป็นชุดข้อมูลเรียนรู้ 80 เปอร์เซ็นต์ และชุดข้อมูลทดสอบ 20 เปอร์เซ็นต์ ทำการเปรียบเทียบประสิทธิภาพการทำนายบนชุดข้อมูลทดสอบด้วย ค่าความแม่นยำ ค่าความไว ค่าวัดประสิทธิภาพโดยใช้พื้นที่ใต้กราฟ และค่าคลาดเคลื่อนกำลังสองเฉลี่ย จากการศึกษาพบว่าปัจจัยที่ส่งผลกระทบต่อระดับความรุนแรงของอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 คือ อายุ ค่าดัชนีมวลกาย ค่าความดันโลหิตขณะหัวใจบีบตัว ค่าความดันโลหิตขณะหัวใจคลายตัว การสูบบุหรี่ การดื่มเครื่องดื่มที่มีแอลกอฮอล์ การเป็นโรคความดันโลหิตสูง การเป็นโรคเบาหวาน การเป็นโรคหลอดเลือดหัวใจ และการเป็นโรคปอด นอกจากนี้ยังพบว่าโครงข่ายประสาทเทียมมีประสิทธิภาพในการทำนายดีที่สุดในตัวแบบการทำนายที่มีจำนวนโหนดในชั้นซ่อน 8 โหนด อัตราการเรียนรู้ 0.05 และโมเมนตัม 0.5 โดยให้ค่าความแม่นยำร้อยละ 94.62 ค่าความไวร้อยละ 83.33 ค่าวัดประสิทธิภาพโดยใช้พื้นที่ใต้กราฟคือ 0.817 และค่าคลาดเคลื่อนกำลังสองเฉลี่ยคือ 0.286 ส่วนการวิเคราะห์การถดถอยลอจิสติกทวิภาคให้ค่าความแม่นยำร้อยละ 88.46 ค่าความไวร้อยละ 76.67 ค่าวัดประสิทธิภาพโดยใช้พื้นที่ใต้กราฟคือ 0.707 และค่าคลาดเคลื่อนกำลังสองเฉลี่ยคือ 0.312

เอกสารนี้ **คำสำคัญ** : การวิเคราะห์การถดถอยลอจิสติกทวิภาค โครงข่ายประสาทเทียม ใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Title	Efficiency Comparison of Severe Prediction of Coronavirus Disease 2019 Patients
Students	Miss. Thanita Ngaongam Student ID 63050618 Miss. Nipaporn Phianglim Student ID 63050621 Mr. Rachata Pannaroj Student ID 63050655
Degree	Bachelor of Science (Applied Statistics)
Department	Statistics
School	Science
University	King Mongkut's Institute of Technology Ladkrabang (KMITL)
Academic Year	2023
Advisor	Assoc. Prof. Saichon Sinsomboonthong

Abstract

This research aims to predict factors influencing the severity level of symptoms in patients with COVID-19 and to compare the prediction performance of severity levels of symptoms in COVID-19 patients using binary logistic regression analysis and artificial neural networks. Data were collected from COVID-19 patients treated at a specific department of health in Bangkok between March and December, 2022. A targeted sample group was selected, comprising COVID-19 patients with positive ATK test results. Data were collected directly from patients by health department staff through interviews based on the COVID-19 patient investigation form, yielding a total of 652 data sets. The dataset were divided into a learning dataset (80%) and a testing dataset (20%). Prediction performance was compared on the testing dataset using accuracy, sensitivity, specificity, area under the curve (AUC), and mean squared error. The study found that factors influencing the severity level of COVID-19 symptoms included age, body mass index, systolic blood pressure, diastolic blood pressure, smoking, alcohol consumption, hypertension, diabetes, coronary artery disease, and lung disease. The neural network demonstrated the best prediction performance in a model with 8 hidden layer nodes, a learning rate of 0.05, and momentum of 0.5, achieving an accuracy of 94.62%, sensitivity of 83.33%, specificity of 0.817, and mean squared error of 0.286 The binary logistic regression analysis yielded an accuracy of 88.46%, sensitivity of 76.67%, specificity of 0.707, and mean squared error of 0.312.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
Keywords: Binary Logistic Regression Analysis, Artificial Neural Networks

กิตติกรรมประกาศ

ปัญหาพิเศษนี้สำเร็จสมบูรณ์ด้วยดีโดยได้รับความกรุณาจาก รศ.สายชล สิ้นสมบูรณ์ทอง อาจารย์ที่ปรึกษาปัญหาพิเศษ ผู้ให้คำแนะนำ ข้อเสนอแนะ คำปรึกษา เอื้อเพื่อเอกสารต่าง ๆ ที่ใช้เป็นแนวทางในการวิเคราะห์ข้อมูลและตรวจทานแก้ไขความถูกต้องตลอดจนติดตามผลงานทุก ๆ ขั้นตอนของการดำเนินงานในการทำปัญหาพิเศษนี้จนกระทั่งเสร็จสมบูรณ์ จึงขอขอบพระคุณด้วยความเคารพอย่างสูงไว้ ณ โอกาสนี้ด้วย

ขอขอบพระคุณ คณะกรรมการปัญหาพิเศษที่ให้คำปรึกษา และคำแนะนำที่ทำให้ปัญหาพิเศษนี้มีความสมบูรณ์ยิ่งขึ้น

ขอขอบพระคุณคณะอาจารย์ ภาควิชาสถิติ ทุกท่านที่ได้ประสิทธิ์ประสาทวิชาความรู้ และช่วยเหลือให้คำแนะนำในเรื่องต่าง ๆ มาโดยตลอด

สุดท้ายนี้ ขอขอบพระคุณบิดามารดา ที่คอยให้กำลังใจและสนับสนุนผู้จัดทำปัญหาพิเศษมาโดยตลอด และขอขอบคุณเพื่อน ๆ พี่น้อง ทุกคนที่ให้คำปรึกษา ช่วยเหลือการทำงานมาโดยตลอดจนปัญหาพิเศษนี้สำเร็จลุล่วงมาได้ด้วยดี

ฐานิตา เกางาม
ณิพัทร เพียงลิ้ม
รชต พรธโรจน์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

	หน้า
บทคัดย่อภาษาไทย	ก
บทคัดย่อภาษาอังกฤษ	ข
กิตติกรรมประกาศ	ค
สารบัญ	ง
สารบัญตาราง	ฉ
สารบัญรูป	ช
คำย่อ	ณ
บทที่ 1 บทนำ	1
1.1 ความเป็นมาและความสำคัญของปัญหา	1
1.2 วัตถุประสงค์ของงานวิจัย	3
1.3 ขอบเขตของงานวิจัย	3
1.4 ประโยชน์ที่คาดว่าจะได้รับ	4
1.5 นิยามศัพท์เฉพาะ	4
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง	5
2.1 การเตรียมข้อมูล	5
2.1.1 การทำความเข้าใจข้อมูล	5
2.1.2 การทำความสะอาดข้อมูล	5
2.1.3 การแปลงข้อมูล	5
2.2 การวิเคราะห์การถดถอยลอจิสติกทวิภาค	7
2.3 โครงข่ายประสาทเทียม	11
2.4 การเปรียบเทียบประสิทธิภาพการทำนาย	14
2.4.1 ค่าความแม่นยำ	14
2.4.2 ค่าเรียกคืน	15
2.4.3 ค่าวัดประสิทธิภาพโดยใช้พื้นที่ใต้กราฟ	15
2.4.4 ค่าคลาดเคลื่อนกำลังสองเฉลี่ย	15
2.5 งานวิจัยที่เกี่ยวข้อง	15
บทที่ 3 วิธีการดำเนินงานวิจัย	17
3.1 การวางแผนการวิจัย	17
3.1.1 กลุ่มตัวอย่าง	17
3.1.2 เครื่องมือที่ใช้ในการวิจัย	17
3.1.1 การรวบรวมข้อมูล	17

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

	หน้า
3.2 การวิธีดำเนินการวิจัย	18
3.2.1 การเตรียมข้อมูล	18
3.3 การสร้างตัวแบบทำนาย	22
3.3.1 การถดถอยลอจิสติกทวิภาค	22
3.3.2 โครงข่ายประสาทเทียม	23
3.4 การเปรียบเทียบประสิทธิภาพการทำนาย	24
3.5 แผนการดำเนินงานวิจัย	25
บทที่ 4 ผลการวิจัย	26
4.1 ผลที่ได้จากการเตรียมข้อมูลเพื่อสร้างตัวแบบทำนาย	26
4.2 ผลการสร้างและพัฒนาตัวแบบทำนาย	26
4.2.1 การวิเคราะห์การถดถอยลอจิสติกทวิภาค	26
4.2.2 โครงข่ายประสาทเทียม	30
4.3 ผลการเปรียบเทียบประสิทธิภาพของตัวแบบทำนาย	35
บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ	36
5.1 สรุปผลการวิจัย	36
5.2 ข้อเสนอแนะ	37
เอกสารอ้างอิง	38
ภาคผนวก	41
ภาคผนวก ก	42
ภาคผนวก ข	44

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

ตารางที่	หน้า
2.1 เมทริกซ์ความสับสน	14
3.1 ตัวอย่างข้อมูลทั้งหมดของผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 ที่เข้ารับการรักษาใน กรมอนามัย กระทรวงสาธารณสุขแห่งหนึ่งในกรุงเทพมหานคร ตั้งแต่เดือนมีนาคม- ธันวาคม พ.ศ. 2565	18
3.2 รายละเอียดข้อมูลผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019	19
3.3 สถิติเชิงพรรณนาของข้อมูลผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019	20
3.4 การแปลงข้อมูลให้อยู่ในรูปปรกติมาตรฐาน	21
3.5 รายละเอียดตัวแปรหลังจากแปลงข้อมูลให้อยู่ในรูปปรกติมาตรฐาน	22
4.1 ผลลัพธ์จากการตรวจสอบความสัมพันธ์ระหว่างตัวแปรอิสระด้วยวิธีทีละชั้น แบบถอยหลัง	27
4.2 ผลการคัดเลือกตัวแปรอิสระด้วยวิธีทีละชั้นแบบถอยหลัง	27
4.3 ผลลัพธ์การทำนายระดับความรุนแรงของอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 จากการถดถอยลอจิสติกทวิภาคโดยใช้วิธีทีละชั้นแบบถอยหลังชุดการเรียนรู้ 80 เปอร์เซนต์	29
4.4 ผลลัพธ์การทำนายระดับความรุนแรงของอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 จากการถดถอยลอจิสติกทวิภาคโดยใช้วิธีทีละชั้นแบบถอยหลังชุดการทดสอบ 80 เปอร์เซนต์	29
4.5 ค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวแบบการทำนายที่มีจำนวนโหนดในชั้นซ่อน 1 โหนด	31
4.6 ค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวแบบการทำนายที่มีจำนวนโหนดในชั้นซ่อน 2 โหนด	31
4.7 ค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวแบบการทำนายที่มีจำนวนโหนดในชั้นซ่อน 3 โหนด	31
4.8 ค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวแบบการทำนายที่มีจำนวนโหนดในชั้นซ่อน 4 โหนด	31
4.9 ค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวแบบการทำนายที่มีจำนวนโหนดในชั้นซ่อน 5 โหนด	31
4.10 ค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวแบบการทำนายที่มีจำนวนโหนดในชั้นซ่อน 6 โหนด	32

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง (ต่อ)

ตารางที่	หน้า
4.11 ค่าตลาดเคลื่อนกำลังสองเฉลี่ยของตัวแบบการทำนายที่มีจำนวนโหนดในชั้นซ่อน 7 โหนด	32
4.12 ค่าตลาดเคลื่อนกำลังสองเฉลี่ยของตัวแบบการทำนายที่มีจำนวนโหนดในชั้นซ่อน 8 โหนด	32
4.13 ค่าตลาดเคลื่อนกำลังสองเฉลี่ยของตัวแบบการทำนายที่มีจำนวนโหนดในชั้นซ่อน 9 โหนด	32
4.14 ค่าตลาดเคลื่อนกำลังสองเฉลี่ยของตัวแบบการทำนายที่มีจำนวนโหนดในชั้นซ่อน 10 โหนด	32
4.15 ผลลัพธ์การทำนายระดับความรุนแรงของอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 จากโครงข่ายประสาทเทียมโดยใช้วิธีแพร่ย้อนกลับของชุดข้อมูลเรียนรู้ 80 เปอร์เซ็นต์ และอัตราการเรียนรู้ 0.05	33
4.16 ผลลัพธ์การทำนายระดับความรุนแรงของอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 จากโครงข่ายประสาทเทียมโดยใช้วิธีแพร่ย้อนกลับของชุดข้อมูลเรียนรู้ 80 เปอร์เซ็นต์ และอัตราการเรียนรู้ 0.05	34
4.17 ผลลัพธ์การทำนายระดับความรุนแรงของอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 จากโครงข่ายประสาทเทียมโดยใช้วิธีแพร่ย้อนกลับของชุดข้อมูลทดสอบ 20 เปอร์เซ็นต์ และอัตราการเรียนรู้ 0.05	35

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป

รูปที่		หน้า
2.1	เซลล์ประสาทกับแบบจำลองการทำงานของเซลล์ประสาทเทียม	11
2.2	แบบจำลองการทำงานของโครงข่ายประสาทเทียม	11
3.1	แผนผังการดำเนินงานวิจัย	25



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

คำย่อ

คำย่อ/สัญลักษณ์	คำอธิบาย
Sex	เพศ
Age	อายุ
Weight	น้ำหนัก
Height	ส่วนสูง
BMI	ค่าดัชนีมวลกาย
SBP	ค่าความดันโลหิตขณะหัวใจบีบตัว
DBP	ค่าความดันโลหิตขณะหัวใจคลายตัว
SpO2	ระดับออกซิเจนในเลือด
HR	อัตราการเต้นของหัวใจ
T	อุณหภูมิร่างกาย
NOV	จำนวนวัคซีนที่ได้รับ
SMK	การสูบบุหรี่
DAB	การดื่มเครื่องดื่มที่มีแอลกอฮอล์
HT	การเป็นโรคความดันโลหิตสูง
DM	การเป็นโรคเบาหวาน
CAD	การเป็นโรคหลอดเลือดหัวใจ
PN	การเป็นโรคปอด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

โรคติดเชื้อไวรัสโคโรนา 2019 หรือโรคโควิด-19 (Coronavirus disease 2019; COVID-19) เกิดจากไวรัสโคโรนา กลุ่มอาการทางเดินหายใจเฉียบพลันรุนแรง 2 (Severe Acute Respiratory Syndrome-Coronavirus-2; SARS-CoV-2) ซึ่งเริ่มปรากฏในเดือนธันวาคม 2562 ที่นครอู่ฮั่น เมืองเอกของมณฑลหูเป่ย์ ประเทศจีน โรคโควิด-19 สามารถติดต่อโดยตรงจากการสัมผัสสารคัดหลั่งของผู้ป่วย และสามารถติดต่อผ่านทางละอองที่ถูกขับออกมาเมื่อผู้ป่วยไอหรือจาม อาการของผู้ป่วยโรคโควิด-19 เป็นอาการที่เกี่ยวข้องกับระบบทางเดินหายใจ เช่น ไอ เจ็บคอ หายใจลำบาก มีไข้สูง และปวดกล้ามเนื้อ ในบางรายอาจมีอาการรุนแรง ทำให้ระบบทางเดินหายใจล้มเหลวและเสียชีวิตได้ โดยในวันที่ 11 มีนาคม 2563 องค์การอนามัยโลกได้ประกาศให้โรคโควิด-19 เป็นโรคระบาดใหญ่ทั่วโลก (Pandemic) ซึ่งในขณะนั้นมีผู้ป่วยมากกว่า 118,000 ราย ใน 114 ประเทศ และมีผู้เสียชีวิตแล้ว 4,291 ราย (World Health Organization, 2022)

ปัจจุบัน SARS-CoV-2 มีการแพร่ระบาดอยู่หลากหลายสายพันธุ์ เช่น สายพันธุ์อัลฟา สายพันธุ์เบต้า สายพันธุ์เดลต้า และสายพันธุ์โอไมครอนหรือโอไมครอน การกลายพันธุ์เหล่านี้มีความเกี่ยวข้องกับลักษณะของสมรรถภาพที่เพิ่มขึ้น แพร่เชื้อได้รวดเร็วขึ้น และแต่ละสายพันธุ์มีระดับความรุนแรงของอาการแตกต่างกัน ยิ่งกว่านั้นหลังจากการเกิดขึ้นของสายพันธุ์โอไมครอน วิวัฒนาการของ SARS-CoV-2 มีความสามารถในการหลบเลี่ยงภูมิคุ้มกัน ทำให้ประสิทธิภาพของวัคซีนต่อ SARS-CoV-2 ลดลง ผู้ที่เคยติดเชื้อแล้วสามารถติดเชื้อซ้ำได้อีก การฉีดวัคซีนไม่สามารถป้องกันการติดเชื้อได้ แต่การฉีดวัคซีนทำให้ร่างกายมีภูมิคุ้มกันมากขึ้น เมื่อมีการติดเชื้อหรือติดเชื้อซ้ำระดับความรุนแรงของโรคจะลดน้อยลง

Rawat (2019) ทำนายการเป็นโรคหัวใจด้วยการประยุกต์ใช้กระบวนการเรียนรู้ของเครื่อง (Machine Learning) 7 แบบ เพื่อเปรียบเทียบประสิทธิภาพในการทำนายปัจจัยที่ส่งผลต่อการเป็นโรคหัวใจ ได้แก่ Logistic Regression, Random Forest Classifier, SVM, Naive Bayes Classifier, Decision Tree Classifier, LightGBM และ XGBoost ในการประเมินประสิทธิภาพการทำนาย ผลการทดสอบพบว่า การวิเคราะห์การถดถอยลอจิสติกทวิภาคมีความแม่นยำ 86.4% สำหรับข้อมูลชุดเรียนรู้และ 80.4% สำหรับข้อมูลชุดทดสอบ สรุปว่าการวิเคราะห์การถดถอยลอจิสติกทวิภาคเป็นแบบจำลองที่มีประสิทธิภาพสูงที่สุดในกรณีนี้ การวิเคราะห์การถดถอยลอจิสติกจึงนิยมใช้ในการวิเคราะห์หาปัจจัยเสี่ยงของโรคต่างๆ

Abdulaal et al. (2020) นำเสนอปัญญาประดิษฐ์ (AI) ในการสร้างระบบการให้คะแนนความเสี่ยงต่อการเสียชีวิต ณ จุดรับเข้า โดยใช้โครงข่ายประสาทเทียม (Artificial Neural Networks: ANN) มีวัตถุประสงค์เพื่อสร้างระบบคำนวณความเสี่ยงต่อการเสียชีวิตของผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 ช่วยในการตัดสินใจด้านการจัดการทางคลินิกได้อย่างรวดเร็ว สามารถคาดการณ์การเสียชีวิตเฉพาะผู้ป่วยด้วยความแม่นยำ (Accuracy) 86.25% ความไว (Sensitivity) 87.50% ความจำเพาะ (Specificity) 85.94% ค่าทำนายเชิงบวก (Positive Predictive Value) 60.87% และค่าทำนายเชิงลบ (Negative Predictive Value) 96.49%

เอกสารนี้เป็นเอกสารสงวนไว้สำหรับการใช้งานเพื่อการศึกษานานาชาติเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Zhou et al. (2020) การศึกษาพบว่าผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 ที่มีอายุมากขึ้น มีโอกาสเสี่ยงต่อการเสียชีวิตมากขึ้น เป็นการพบความสัมพันธ์ที่มีความสัมพันธ์สถิติอย่างมีนัยสำคัญ ($p = 0.0043$) และมักพบว่าผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 ที่เสียชีวิตมักมีโรคร่วมที่พบมากที่สุดคือโรคความดันโลหิตสูง, โรคเบาหวาน และโรคหลอดเลือดหัวใจ นี่เป็นความสอดคล้องกับ Mohammadi et al. (2021) ที่พบว่าผู้ป่วยโรคติดเชื้อ SARS-CoV-2 มักมีโรคร่วมที่พบบ่อย เช่น โรคเบาหวาน, โรคความดันโลหิตสูง, โรคปอดอุดกั้นเรื้อรัง, และโรคหลอดเลือดหัวใจ โครงข่ายประสาทเทียมหลายชั้น Perceptron มีประสิทธิภาพดีกว่าการวิเคราะห์การถดถอยลอจิสติกทวิภาค ในการทำนายความเสี่ยงในการเสียชีวิตของผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 สรุปได้ว่า โครงข่ายประสาทเทียมหลายชั้น Perceptron เป็นวิธีการทำนายที่มีประสิทธิภาพและเชื่อถือได้ ในการทำนายความเสี่ยงในการเสียชีวิตของผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019

Rashidi et al. (2021) ได้สำรวจการประยุกต์ปัญญาประดิษฐ์ ในการจัดการกับโรคติดเชื้อไวรัสโคโรนา 2019 พบว่า AI สามารถช่วยในการวินิจฉัยโรคติดเชื้อไวรัสโคโรนา 2019 ทำนายความเสี่ยงต่อการเสียชีวิต คาดการณ์ผลของการรักษา และทำนายการแพร่ระบาด สรุปได้ว่า AI มีบทบาทสำคัญในการเพิ่มประสิทธิภาพในการรักษาทางการแพทย์ โดยการช่วยให้การวินิจฉัยและการรักษารวดเร็วและมีความแม่นยำมากขึ้น

Ding et al. (2021) พัฒนาตัวแบบเพื่อทำนายการเสียชีวิตในโรงพยาบาลในระยะเริ่มแรกของผู้ป่วยโรคตับอ่อนอักเสบเฉียบพลัน (AP) และเปรียบเทียบประสิทธิภาพในการทำนายการเสียชีวิตของผู้ป่วยโรคตับอ่อนอักเสบเฉียบพลัน โดยผลการวิจัยพบว่า โครงข่ายประสาทเทียมแบบแพร่ย้อนกลับ (Back Propagation Neural Network) สามารถระบุปัจจัยที่มีผลต่อการเสียชีวิตของผู้ป่วยโรคตับอ่อนอักเสบเฉียบพลันได้ทั้งหมด 12 ปัจจัยด้วยความแม่นยำ 76.9% ความไว 66.6% ความจำเพาะ 66.1% ค่าทำนายเชิงบวก 56.3% ค่าทำนายเชิงลบ 91.6% และมีค่าวัดประสิทธิภาพโดยใช้พื้นที่ใต้กราฟ (Area Under a Curve: AUC) เท่ากับ 0.769 และการวิเคราะห์การถดถอยลอจิสติกทวิภาค (Binary Logistic Regression) สามารถระบุปัจจัยที่มีผลต่อการเสียชีวิตของผู้ป่วยโรคตับอ่อนอักเสบเฉียบพลันได้ 3 ปัจจัยด้วยความแม่นยำ 66% ความไว 34.6% ความจำเพาะ 92.3% ค่าทำนายเชิงบวก 36% ค่าทำนายเชิงลบ 91.9% และมีค่าวัดประสิทธิภาพโดยใช้พื้นที่ใต้กราฟเท่ากับ 0.607 สรุปได้ว่าการใช้โครงข่ายประสาทเทียมแบบแพร่ย้อนกลับทำนายการเสียชีวิตในโรงพยาบาลในผู้ป่วยโรคตับอ่อนอักเสบเฉียบพลันเป็นครั้งแรก สามารถคัดกรองผู้ป่วยโรคตับอ่อนอักเสบเฉียบพลันที่มีความเสี่ยงสูงต่อการเสียชีวิตได้ดีกว่าการวิเคราะห์การถดถอยลอจิสติกทวิภาค

จากการทบทวนวรรณกรรมและงานวิจัยที่เกี่ยวข้องกับระดับความรุนแรงของอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 ผู้วิจัยจึงเลือกประยุกต์ใช้การวิเคราะห์การถดถอยลอจิสติกทวิภาคและโครงข่ายประสาทเทียมสำหรับการทำนายปัจจัยเสี่ยงของโรคติดเชื้อไวรัสโคโรนา 2019 เนื่องจากการวิเคราะห์การถดถอยลอจิสติกทวิภาค มีประสิทธิภาพสูงสุดในกระบวนการเรียนรู้ของเครื่อง ทั้ง 7 แบบ (Rawat, 2019) และโครงข่ายประสาทเทียมมีความแม่นยำสูง (Abdulaal et al., 2020, Ding et al., 2021 and Zhou et al., 2020) การวิจัยในครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพในการทำนายปัจจัยที่มีผลต่อระดับความรุนแรงของผู้ป่วยโรคติดเชื้อไวรัส โคโรนา 2019 ด้วยวิธีการวิเคราะห์การถดถอยลอจิสติกทวิภาคกับวิธีโครงข่ายประสาทเทียม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.2 วัตถุประสงค์

1.2.1 เพื่อเปรียบเทียบประสิทธิภาพในการทำนายผลระดับความรุนแรงของอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 ด้วยวิธีการวิเคราะห์การถดถอยลอจิสติกทวิภาคกับวิธีโครงข่ายประสาทเทียม

1.2.2 เพื่อทำนายปัจจัยที่มีผลต่อระดับความรุนแรงของอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019

1.3 ขอบเขตการวิจัย

กำหนดขอบเขตการศึกษาดังนี้

1.3.1 ขอบเขตด้านข้อมูล

การวิจัยครั้งนี้ใช้ข้อมูลย้อนหลังที่เป็นข้อมูลทุติยภูมิ (Secondary Data) ได้มาจากการรวบรวมข้อมูลของผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 ที่เข้ารับการรักษาในกรมอนามัย กระทรวงสาธารณสุขแห่งหนึ่งในกรุงเทพมหานคร ตั้งแต่เดือนมีนาคม-ธันวาคม พ.ศ. 2565 จำนวน 652 ชุด ตัวแปรที่ใช้ในการวิเคราะห์ระดับความรุนแรงของอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 มีดังนี้

1. ตัวแปรอิสระ จำนวน 17 ตัว ประกอบด้วย

1) ปัจจัยส่วนบุคคล ได้แก่

- เพศ
- อายุ (ปี)
- น้ำหนัก (กิโลกรัม)
- ส่วนสูง (เซนติเมตร)
- ค่าดัชนีมวลกาย (กิโลกรัม/เมตร²)
- ค่าความดันโลหิตขณะหัวใจบีบตัว (มิลลิเมตรปรอท)
- ค่าความดันโลหิตขณะหัวใจคลายตัว (มิลลิเมตรปรอท)
- ระดับออกซิเจนในเลือด (เปอร์เซ็นต์)
- อัตราการเต้นของหัวใจ (ครั้ง/นาที)
- อุณหภูมิร่างกาย (องศาเซลเซียส)
- จำนวนวัคซีนที่ได้รับ (เข็ม)

2) ปัจจัยด้านสุขภาพ ได้แก่

- การสูบบุหรี่ (ไม่สูบ, สูบ)
- การดื่มเครื่องดื่มที่มีแอลกอฮอล์ (ไม่ดื่ม, ดื่ม)
- การเป็นโรคความดันโลหิตสูง (ไม่เป็น, เป็น)
- การเป็นโรคเบาหวาน (ไม่เป็น, เป็น)
- การเป็นโรคหลอดเลือดหัวใจ (ไม่เป็น, เป็น)
- การเป็นโรคปอด (ไม่เป็น, เป็น)

2. ตัวแปรตาม จำนวน 1 ตัว คือ

ระดับความรุนแรงของอาการผู้ป่วย (ไม่รุนแรง, รุนแรง)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.3.2 ขอบเขตด้านเนื้อหา

1) สร้างตัวแบบในการทำนายระดับความรุนแรงของอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 โดยใช้วิธีการวิเคราะห์การถดถอยลอจิสติกทวิภาคและวิธีโครงข่ายประสาทเทียมโดยแบ่งข้อมูลออกเป็นชุดเรียนรู้ (Training Data Set) 80% และชุดทดสอบ (Testing Data Set) 20%

2) การเปรียบเทียบประสิทธิภาพในการทำนายผลระดับความรุนแรงของอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 โดยใช้เครื่องมือที่ใช้วัดประสิทธิภาพดังนี้ คือ ค่าความแม่นยำ (Accuracy) ค่าความไว (Sensitivity) ค่าวัดประสิทธิภาพโดยใช้พื้นที่ใต้กราฟ (Area Under a Curve: AUC) และความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error: MSE)

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1) ทำให้ทราบประสิทธิภาพในการทำนายปัจจัยที่ส่งผลต่อระดับความรุนแรงของอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019

2) ทำให้ทราบปัจจัยที่ส่งผลต่อระดับความรุนแรงของอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019

1.5 นิยามศัพท์เฉพาะ

1. ระดับความรุนแรง (Severity Level: SL) คือความรุนแรงของอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 โดยพิจารณาตามการวินิจฉัยของแพทย์ที่รักษาแบ่งเป็น 2 ระดับ ดังนี้

กลุ่มที่ 1 ไม่รุนแรง (Not Severe) ได้แก่ กลุ่มผู้ป่วยที่ไม่มีอาการ (Asymptomatic Case) หรือ กลุ่มผู้ป่วยที่มีอาการเล็กน้อย ไม่มีภาวะเสี่ยง และไม่มีโรคร่วมสำคัญ

กลุ่มที่ 2 รุนแรง (Severe) ได้แก่ กลุ่มผู้ป่วยที่เสียชีวิต ระดับออกซิเจนในเลือดน้อยกว่า 95% และผู้ป่วยโรคปอดบวม ทั้งที่มีอาการรุนแรงและไม่รุนแรง

2. ผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 หมายถึง ผู้ป่วยที่มีอาการตรงตามเกณฑ์การสอบสวนโรคและได้รับการตรวจด้วยวิธีมาตรฐาน RT-PCR ว่าติดเชื้อไวรัสโคโรนา (SARS-CoV-2)

3. การวิเคราะห์การถดถอยลอจิสติกทวิภาค (Binary Logistic Regression Analysis) คือ การศึกษาความสัมพันธ์ของตัวแปรตาม ซึ่งเป็นตัวแปรเชิงคุณภาพที่มีค่าเพียง 2 ค่า กำหนดให้ตัวแปร Y และ X มีความสัมพันธ์ภายใต้ตัวแบบการถดถอยลอจิสติกทวิภาค โดย Y มีการแจกแจงแบร์นูลลี (Bernoulli Distribution) และมีตัวแปรอิสระจำนวน p ตัว ซึ่งความน่าจะเป็นที่จะเกิดการเกิดเหตุการณ์ที่สนใจ คือ π

4. โครงข่ายประสาทเทียม (Artificial Neural Networks) คือการกำหนดรูปแบบการให้เหตุผล จากรูปแบบการคิดของสมองมนุษย์ที่ประกอบด้วยเซลล์ประสาทที่เชื่อมต่อกันอย่างหนาแน่น แต่ละหน่วยจะเปรียบเสมือนข้อมูลแต่ละตัวที่ต้องนำมาหาความสัมพันธ์ระหว่างกัน เรียกว่า เซลล์ประสาท (Neuron) สมองมนุษย์ประกอบด้วยเซลล์ประสาทประมาณ 1 หมื่นล้านเซลล์มีการเชื่อมต่อระหว่างกิ่งก้านของเซลล์ประสาทหนึ่งไปอีกเซลล์ประสาทหนึ่ง 60 ล้านล้านตำแหน่ง ซึ่งการเชื่อมต่อเซลล์ประสาทภายในสมองจะมีการประมวลผลเร็วกว่าคอมพิวเตอร์ที่มีอยู่ในปัจจุบัน การทำงานของโครงข่ายประสาทเทียมมี 3 ชั้น คือ ชั้นข้อมูลเข้า (Input Layer) ชั้นซ่อน (Hidden Layers) และชั้น

เอกสารนี้ ข้อมูลออก (Output Layer) การใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

สำหรับปัญหาพิเศษนี้ผู้วิจัยมีความสนใจที่จะเปรียบเทียบประสิทธิภาพในการทำนายผลระดับความรุนแรงของอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 ด้วยการวิเคราะห์การถดถอยลอจิสติกทวิภาคและโครงข่ายประสาทเทียม

2.1 การเตรียมข้อมูล (Data Preprocessing)

การเตรียมข้อมูลเป็นขั้นตอนการจัดกระทำกับข้อมูลที่ได้รับมาแล้ว แต่ยังมีรูปแบบไม่ตรงกับความต้องการที่จะนำมาใช้ หรือยังไม่ครบถ้วนสมบูรณ์ จึงจำเป็นต้องมีวิธีการจัดกระทำกับข้อมูลนั้น ๆ เสียก่อน โดยการเตรียมข้อมูลสามารถดำเนินการได้หลายวิธี ซึ่งแต่ละวิธีที่เลือกนำมาใช้ขึ้นอยู่กับลักษณะดั้งเดิมของข้อมูลที่มีความแตกต่างกัน (ไกรศักดิ์, 2564)

2.1.1 การทำความเข้าใจข้อมูล (Data Understanding) เป็นขั้นตอนที่สำคัญในกระบวนการวิเคราะห์ข้อมูล ซึ่งมีเป้าหมายเพื่อเข้าใจลึกซึ้งเกี่ยวกับข้อมูลที่มีอยู่ โดยต้องทำความเข้าใจความหมาย มาตราวัด หน่วยวัด และรายละเอียดของตัวแปรทุกตัวที่จะใช้ในงานวิจัย

2.1.2 การทำความสะอาดข้อมูล (Data Cleaning) เป็นกระบวนการตรวจสอบและแก้ไขข้อมูลที่ไม่ถูกต้อง หรือผิดปกติ แตกต่างไปจากข้อมูลอื่น ในไฟล์ข้อมูลเดียวกัน ซึ่งเป็นหลักการพื้นฐานของระบบฐานข้อมูล นั่นคือ ความไม่สมบูรณ์ (Anomaly) ความไม่สัมพันธ์กับข้อมูลอื่น ๆ เป็นต้น จึงต้องมีการแทนที่ การปรับปรุง การแทนที่ หรือบางครั้งอาจจำเป็นต้องลบข้อมูลที่ไม่ถูกต้องเหล่านี้ออกไป เพื่อให้ข้อมูลมีคุณภาพพร้อมที่จะนำไปดำเนินการขั้นต่อไปได้ ส่วนสาเหตุที่ทำให้ “ข้อมูลไม่สมบูรณ์” เกิดได้จากความผิดพลาดจากหลายปัจจัย เช่น การกำหนดคุณลักษณะของข้อมูลไม่ถูกต้อง มีการพิมพ์ผิดระหว่างกรบันทึกข้อมูล การส่งข้อมูลไม่สมบูรณ์ เป็นต้น ทั้งนี้ลักษณะของข้อมูลที่ไม่สมบูรณ์ที่สามารถพบเห็นได้ คือ มีการจัดเก็บข้อมูลที่ซ้ำซ้อน (Duplicate Data) ข้อมูลไม่ถูกต้อง (Incorrectly Data) ข้อมูลเก่าล้าสมัย (Expired Data) เกิดการสูญหายของข้อมูลในบางส่วน (Missing Value) มีค่าข้อมูลที่มีความผิดปกติหรือแตกต่างไปจากข้อมูลในกลุ่ม (Outliers)

2.1.3 การแปลงข้อมูล (Data Transformation) เป็นการแปลงค่าของข้อมูลควรจะทำตัวแปรเชิงปริมาณให้เป็นปรกติ (Normalize) เพื่อทำให้สเกลของอิทธิพลของตัวแปรแต่ละตัวที่มีต่อผลลัพธ์เป็นมาตรฐานและเพื่อทำให้ตัวแปรตัวหนึ่งไม่มีอิทธิพลเหนือตัวแปรตัวอื่น ๆ การทำให้เป็นปรกติ (Normalization) ที่ใช้กันแพร่หลายมีอยู่หลายวิธี (สายชล, 2560) ดังนี้

1) การทำให้เป็นปรกติน้อยที่สุด-มากที่สุด (Min-Max Normalization) พิจารณาค่าของข้อมูลมากกว่าค่าน้อยที่สุด $\min(X)$ อยู่เท่าไร และหารความแตกต่างนี้ด้วยพิสัย โดยมีสูตรการคำนวณดังนี้

$$X^* = \frac{X - \min(X)}{\text{range}(X)}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งาน $X^* = \frac{X - \min(X)}{\max(X) - \min(X)}$ (2.1) เป็นการคำนวณค่า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยที่

X^*	คือ ค่าที่ได้จากการแปลงค่าอยู่ในรูปการทำให้เป็นปรกติ
X	คือ ค่าปัจจุบันที่นำมาแปลงค่าอยู่ในรูปการทำให้เป็นปรกติ
$\min(X)$	คือ ค่าข้อมูลที่มีค่าน้อยที่สุดในชุดข้อมูล
$\max(X)$	คือ ค่าข้อมูลที่มีค่ามากที่สุดในชุดข้อมูล

2) คะแนนมาตรฐานแซด (Z-Score Standardization) คะแนนมาตรฐานแซดใช้กันแพร่หลายอย่างมากในการวิเคราะห์เชิงสถิติ โดยการหาความแตกต่างระหว่างค่าของข้อมูลและค่าเฉลี่ยของข้อมูล แล้วหารความแตกต่างนี้ด้วยส่วนเบี่ยงเบนมาตรฐานของค่าของข้อมูล โดยมีสูตรการคำนวณดังนี้

$$X^* = \frac{X - \text{mean}(X)}{\text{SD}(X)} \quad (2.2)$$

3) การทำให้เป็นปรกติโดยใช้เลขทศนิยม (Normalization by Decimal Scaling) เป็นการแปลงค่าของข้อมูลเดิมให้เป็นเลขทศนิยม ตำแหน่งทศนิยมกำหนดโดยค่าสัมบูรณ์ที่มีค่ามากที่สุด โดยมีสูตรการคำนวณดังนี้

$$X^* = \frac{X}{10^j} \quad (2.3)$$

โดยที่

j คือ จำนวนหลักของค่าที่มากที่สุด

4) การทำให้เป็นปรกติโดยใช้ค่ามัธยฐาน (Median Normalization) วิธีมัธยฐานทำให้แต่ละตัวอย่างเป็นปรกติด้วยค่ามัธยฐานของข้อมูลเข้าทั้งหมดในตัวอย่าง วิธีนี้ใช้เมื่อต้องการคำนวณอัตราส่วนระหว่างตัวอย่างผสม 2 ตัวอย่าง ค่ามัธยฐานจะไม่ได้รับอิทธิพลโดยค่าของส่วนเบี่ยงเบนสุดขีด (Extreme Deviation) (Jayalakshmi and Santhakumaran, 2011) โดยมีสูตรการคำนวณดังนี้

$$X^* = \frac{X}{\text{median}} \quad (2.4)$$

5) การทำให้เป็นปรกติแบบซิกมอยด์ (Sigmoid Normalization) ฟังก์ชันการทำให้เป็นปรกติแบบซิกมอยด์ใช้กับมาตรวัดของตัวอย่างที่อยู่ในช่วง 0 ถึง 1 หรือ -1 ถึง 1 มีฟังก์ชันซิกมอยด์ที่เป็นประโยชน์หลายชนิด เช่น ฟังก์ชันซิกมอยด์แทนเจนต์เป็นวิธีที่ดีที่ทำให้กระบวนการเป็นปรกติรวดเร็วยิ่งขึ้น (Jayalakshmi and Santhakumaran, 2011) โดยมีสูตรการคำนวณดังนี้

$$X^* = \frac{e^X - e^{-X}}{e^X + e^{-X}} \quad (2.5)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

6) การทำให้เป็นปกติแบบสดมภ์เชิงสถิติ (Statistical Column Normalization) วิธีการทำให้เป็นปกติแบบสดมภ์เชิงสถิติทำให้แต่ละตัวอย่างเป็นปกติด้วยสดมภ์ที่เป็นปกติ โดยคำนวณการทำให้เป็นปกติของแต่ละสดมภ์ด้วยทำให้สดมภ์เป็นปกติที่มีความยาวเท่ากับ 1 ทำการคำนวณแต่ละตัวอย่างโดยหารด้วยสดมภ์ที่เป็นปกติและคูณด้วยค่าความเอนเอียง 0.1 (Jayalakshmi and Santhakumaran, 2011) โดยมีสูตรการคำนวณดังนี้

$$X^* = \frac{X - n(c_a)}{n(c_a)} \times 0.1 \quad (2.6)$$

2.2 การวิเคราะห์การถดถอยลอจิสติกทวิภาค (Binary Logistic Regression Analysis)

การทำนายโอกาสการติดเชื้อไวรัสโคโรนา 2019 ในระดับรุนแรงหรือในระดับไม่รุนแรงของปัจจัยที่ส่งผลต่อระดับความรุนแรงของอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 และกำหนดระดับนัยสำคัญทางสถิติที่ระดับ 0.05 (ยุทธ, 2555)

ข้อตกลงเบื้องต้น (Assumption) ของการวิเคราะห์การถดถอยลอจิสติกทวิภาค

1) ตัวแปรอิสระหรือตัวแปรทำนาย (X) เป็นตัวแปรที่ข้อมูลอยู่ในระดับช่วง (Interval Scale) เป็นอย่างต่ำ กรณีที่เป็นข้อมูลเชิงกลุ่มให้แปลงเป็นตัวแปรหุ่น (Dummy Variable) มีค่าเป็น 0 กับ 1 ส่วนตัวแปรตามหรือตัวแปรเกณฑ์ (Y) กรณีที่เป็นการวิเคราะห์ลอจิสติกทวิภาค จะกำหนด 2 ค่า คือ 0, 1

2) ตัวแปรอิสระไม่มีความสัมพันธ์กัน จะใช้เกณฑ์ความสัมพันธ์เหมือนกับการวิเคราะห์การถดถอยพหุคูณ โดยถ้าใช้เกณฑ์ของ Burns and Grove (1993) จะใช้ค่า r ไม่เกิน 0.65 ถ้าใช้เกณฑ์ของ Stevens (1996) ใช้ค่า r ไม่เกิน 0.80 (กัลยา, 2549)

3) ค่าความคลาดเคลื่อน (e) มีค่าเฉลี่ยเท่ากับศูนย์ เงื่อนไขข้อนี้เป็นจริงเนื่องจากเมื่อใช้วิธีกำลังสองน้อยที่สุด (Least Square Method) ในการประมาณค่าสัมประสิทธิ์การถดถอยจะทำให้ $e = 0$ เสมอ (เพชรน้อย, 2549)

4) การวิเคราะห์การถดถอยลอจิสติกทวิภาคจะต้องใช้ขนาดตัวอย่าง n มากกว่าการวิเคราะห์การถดถอยพหุคูณ โดยจะใช้ขนาดตัวอย่างเท่ากับ $n \geq 30p$ โดยที่ p คือ จำนวนตัวแปรทำนาย (กัลยา, 2549)

การวิเคราะห์การถดถอยลอจิสติกทวิภาค เนื่องจากตัวแปรตาม (Y) เป็นตัวแปรเชิงคุณภาพมีค่าได้ 2 ค่า คือ ไม่เกิดเหตุการณ์ ($Y = 0$) หรือเกิดเหตุการณ์ ($Y = 1$) จึงทำให้ความสัมพันธ์ระหว่างตัวแปรตาม (Y) กับตัวแปรอิสระ (X) ไม่อยู่ในรูปเชิงเส้น แต่จะอยู่ในรูป

$$E\{Y\} = p = \frac{e^w}{1 + e^w} \quad (2.7)$$

$$\text{เมื่อ } W = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$$

เอกสารนี้ตั้งนัยการทำนายค่าของตัวแปรตาม Y แทนค่าด้วยตัวแปรอิสระ X เป็นการทำนายค่าความน่าจะเป็นว่าของ Y ที่จะเกิดขึ้นเมื่อรู้ค่าของตัวแปรอิสระ X จะได้ดังนี้ถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$P(Y=1) = P(\text{เกิดเหตุการณ์}) = \frac{1}{1+e^{-w}} \quad (2.8)$$

$$\text{เมื่อ } W = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$$

$$P(Y=0) = P(\text{ไม่เกิดเหตุการณ์}) = 1 - P \quad (2.9)$$

$$= \frac{1}{1+e^w}$$

$$\text{เมื่อ } W = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$$

การพิจารณาโอกาสที่จะเกิดเหตุการณ์ที่สนใจอธิบายได้ด้วยค่าอัตราส่วนออดส์ (Odd Ratio: OR) โดยกำหนดให้อัตราส่วนออดส์เป็นอัตราส่วนระหว่างโอกาสที่จะเกิดเหตุการณ์กับโอกาสที่จะไม่เกิดเหตุการณ์ ดังสมการ

$$\text{Odd Ratio(OR)} = \frac{P(\text{เกิดเหตุการณ์})}{P(\text{ไม่เกิดเหตุการณ์})} = \frac{P}{1-P} = e^{\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k} \quad (2.10)$$

ถ้าค่า $OR > 1$ แสดงว่า โอกาสที่จะเกิดเหตุการณ์มากกว่าโอกาสที่จะไม่เกิดเหตุการณ์

ถ้าค่า $OR = 1$ แสดงว่า โอกาสที่จะเกิดเหตุการณ์และโอกาสที่จะไม่เกิดเหตุการณ์เท่ากัน

ถ้าค่า $OR < 1$ แสดงว่า โอกาสที่จะเกิดเหตุการณ์น้อยกว่าโอกาสที่จะไม่เกิดเหตุการณ์

ดังนั้น ค่าอัตราส่วนออดส์ (Odd Ratio: OR) จึงเป็นค่าที่แสดงให้เห็นว่ามีความน่าจะเป็นที่จะเกิดเหตุการณ์เป็นกี่เท่าของความน่าจะเป็นที่จะไม่เกิดเหตุการณ์

$$\begin{aligned} \text{ถ้า } \ln\left(\frac{P}{1-P}\right) &= \ln\left(e^{\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k}\right) \\ &= \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k \end{aligned} \quad (2.11)$$

หรือกรณีใช้ข้อมูลตัวอย่าง จะได้

$$\text{Logit}(P) = b_0 + b_1 X_1 + \dots + b_k X_k \quad (2.12)$$

เมื่อ P คือ โอกาสของการเกิดเหตุการณ์ที่สนใจ

b_i คือ ค่าประมาณสัมประสิทธิ์การถดถอยตัวที่ i

P คือ ค่าประมาณสัมประสิทธิ์การถดถอยคงที่

X_i คือ ตัวแปรอิสระตัวที่ i

เรียกสมการที่ได้ว่า การแปลงลอจิท (Logit transformation) ซึ่งเป็นสมการลอจิสติกที่แสดงความสัมพันธ์ในรูปสมการถดถอยเชิงเส้นพหุคูณ แต่แสดงในรูปของ $\ln(\text{Odds})$ ซึ่งสมการที่ได้นี้จะง่ายต่อการใช้วิธีการประมาณค่าแบบกำลังสองน้อยที่สุดในการวิเคราะห์ข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ขั้นตอนการวิเคราะห์การถดถอยลอจิสติกทวิภาค (ยุทธ, 2555)

ขั้นตอนที่ 1 กำหนดตัวแปรตามการถดถอย ถ้าตัวแปรตามมีค่าเพียง 2 ค่า จะกำหนดเป็นแบบการถดถอยลอจิสติกทวิภาค (Binary Logistic Regression) ถ้ามีตัวแปรตามมากกว่า 2 ค่า จะกำหนดเป็นแบบการถดถอยลอจิสติกอเนกนาม (Multinomial Logistic Regression)

ขั้นตอนที่ 2 คัดเลือกตัวแปรอิสระเข้าในตัวแบบ โดยใช้วิธีทีละขั้นแบบถอยหลัง: อัตราส่วนภาวะน่าเป็น (Backward Stepwise: Likelihood Ratio) เป็นวิธีที่นำตัวแปรทำนายทั้ง p ตัว ($X_1, X_2, X_3, \dots, X_p$) เข้าสมการพร้อมกัน จากนั้นพิจารณาตัวแปรทำนายที่อธิบายความผันแปรของตัวแปรตามได้น้อยที่สุดออกจากสมการ ทำเช่นนี้ไปเรื่อย ๆ จนได้ตัวแปรทำนายที่สามารถอธิบายความผันแปรของตัวแปรตามได้อย่างมีนัยสำคัญทางสถิติ (ยุทธ, 2555)

โดยในงานวิจัยนี้พิจารณาการคัดเลือกตัวแปรอิสระจากเกณฑ์สารสนเทศอะกะอิเกะ (Akaike's Information Criterion: AIC) เกณฑ์นี้จะเลือกตัวแบบที่ประกอบไปด้วยตัวแปรอิสระที่ให้ค่า AIC ต่ำที่สุด

$$AIC_k = -2 \ln \hat{L}_k + 2p_k \quad (2.13)$$

โดยที่

\ln คือ ลอการิทึมธรรมชาติ

\hat{L}_k คือ ฟังก์ชันภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Function) ของตัวแบบที่ k

p_k คือ จำนวนตัวแปรอิสระของตัวแบบที่ k

กล่าวคือตัวแบบที่ได้มีแนวโน้มให้ค่าทำนายที่ใกล้เคียงกับค่าจริง โดยขั้นตอนการคัดเลือกตัวแปรอิสระจะเริ่มด้วยการสร้างตัวแบบเต็ม หลังจากนั้นจะนำตัวแปรอิสระออกจากสมการถดถอยครั้งละ 1 ตัว (อริญชัย, 2562)

ขั้นตอนที่ 3 ตรวจสอบข้อตกลงเบื้องต้น (Assumption) ของการวิเคราะห์การถดถอยลอจิสติกทวิภาค (พิชญา, 2564)

ข้อตกลงเบื้องต้นข้อที่ 1 ตัวแปรตามเป็นการวิเคราะห์ลอจิสติกทวิภาคที่มี 2 ค่า คือ 0, 1

ข้อตกลงเบื้องต้นข้อที่ 2 ค่าความคลาดเคลื่อน (e) มีค่าเฉลี่ยเท่ากับศูนย์ เงื่อนไขข้อนี้เป็นจริงเนื่องจากเมื่อใช้วิธีกำลังสองน้อยที่สุดในการประมาณค่าสัมประสิทธิ์การถดถอยจะทำให้ $e = 0$ เสมอ

ข้อตกลงเบื้องต้นข้อที่ 3 ตัวแปรอิสระต้องเป็นอิสระต่อกันหรือไม่มีความสัมพันธ์กัน ไม่เกิดปัญหาความสัมพันธ์เชิงเส้นพหุ (Multicollinearity) โดยพิจารณาจากค่า Variance Inflation Factors (VIF) และค่าคลาดเคลื่อนยินยอม (Tolerance)

ค่า VIF เป็นค่าที่ใช้วัดตัวแปรอิสระแต่ละตัวซึ่งอธิบายด้วยตัวแปรอื่นว่าอยู่ในระดับใด ถ้าค่า VIF มาก แสดงว่าตัวแปรอิสระมีความสัมพันธ์กับตัวแปรอื่นมาก ซึ่งถ้าค่า VIF มากกว่า 10 ขึ้นไป จะถือว่า อยู่ในระดับที่ทำให้เกิดปัญหาความสัมพันธ์เชิงเส้นพหุได้ ซึ่งค่า VIF สามารถคำนวณได้จากสมการ

$$VIF = \frac{1}{1 - R_i^2} \quad (2.14)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ค่า Tolerance เป็นค่าที่ใช้วัดตัวแปรอิสระแต่ละตัว ซึ่งถูกอธิบายโดยตัวแปรอิสระอื่นใน ระดับใด ซึ่งค่า Tolerance มีค่าอยู่ระหว่าง 0-1 ถ้าค่า Tolerance ของตัวแปรใดมีค่าเข้าใกล้ 0 แสดง ว่าตัวแปรอิสระนั้นมีความสัมพันธ์กับตัวแปรอื่นมาก ดังนั้นค่า Tolerance ของตัวแปรใดมีค่า เข้าใกล้ 1 จะถือว่าดี

$$Tolerance = 1 - R_i^2 = \frac{1}{VIF} \quad (2.15)$$

โดยที่

R_i^2 แทน ค่าสัมประสิทธิ์การกำหนดการถดถอยของตัวแปรอิสระ

ข้อตกลงเบื้องต้นข้อที่ 4 การวิเคราะห์การถดถอยลอจิสติกทวิภาคจะต้องใช้ขนาดตัวอย่าง เท่ากับ $n \geq 30$

ขั้นตอนที่ 4 ประมาณค่าสัมประสิทธิ์การถดถอย β_i ด้วยวิธีภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Method) คำนวณหาค่าอัตราส่วนออดส์ (Odd Ratio) และสร้างสมการ ลอจิต (Logit)

ขั้นตอนที่ 5 ตรวจสอบความเหมาะสมของตัวแบบ วิธีการตรวจสอบความเหมาะสมของ ตัวแบบที่นิยมใช้มีหลายวิธีซึ่งจะเลือกใช้วิธีไหนก็ได้ตามความเหมาะสมของข้อมูล ในงานวิจัยนี้เลือก ทดสอบความเหมาะสมด้วยวิธีฮอสเมอร์-เลเมสโชว์ (Hosmer-Lemeshow)

สมมติฐานในการทดสอบคือ

H_0 : ตัวแบบเหมาะสม

H_1 : ตัวแบบไม่เหมาะสม

ถ้า p-value $> \alpha$ จะไม่สามารถปฏิเสธ H_0 นั่นคือตัวแบบมีความเหมาะสม

ขั้นตอนที่ 6 วัดระดับความสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระ โดยตรวจสอบ ความสอดคล้องของตัวแบบการทำนาย ตัวสถิติทดสอบที่ใช้ ได้แก่ ค็อก-สเนลล์ ($Cox \& snell R_{CS}^2$) และนาเจลเคอร์คิ ($Nagelkerki R_N^2$) โดยมีสูตรดังนี้

ค็อก-สเนลล์ ($Cox \& snell R_{CS}^2$)

$$R_{CS}^2 = 1 - \left[\frac{L(0)}{L(B)} \right]^{\frac{2}{n}} \quad (2.16)$$

โดยที่

$L(0)$ คือ ฟังก์ชันภาวะน่าจะเป็นของฟังก์ชันที่มีค่าคงที่เท่านั้น

$L(B)$ คือ ฟังก์ชันภาวะน่าจะเป็นของฟังก์ชันที่มีตัวแปรอิสระ

พบว่าค่า R_{CS}^2 จะน้อยกว่า 1 เสมอ

นาเจลเคอร์คิ ($Nagelkerki R_N^2$)

$$R_N^2 = \frac{R_{CS}^2}{R_{CS:\max}^2} \quad (2.17)$$

เมื่อ $R_{CS:\max}^2 = 1 - [L(0)]^{\frac{2}{n}}$ และ $0 \leq R_N^2 \leq 1$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

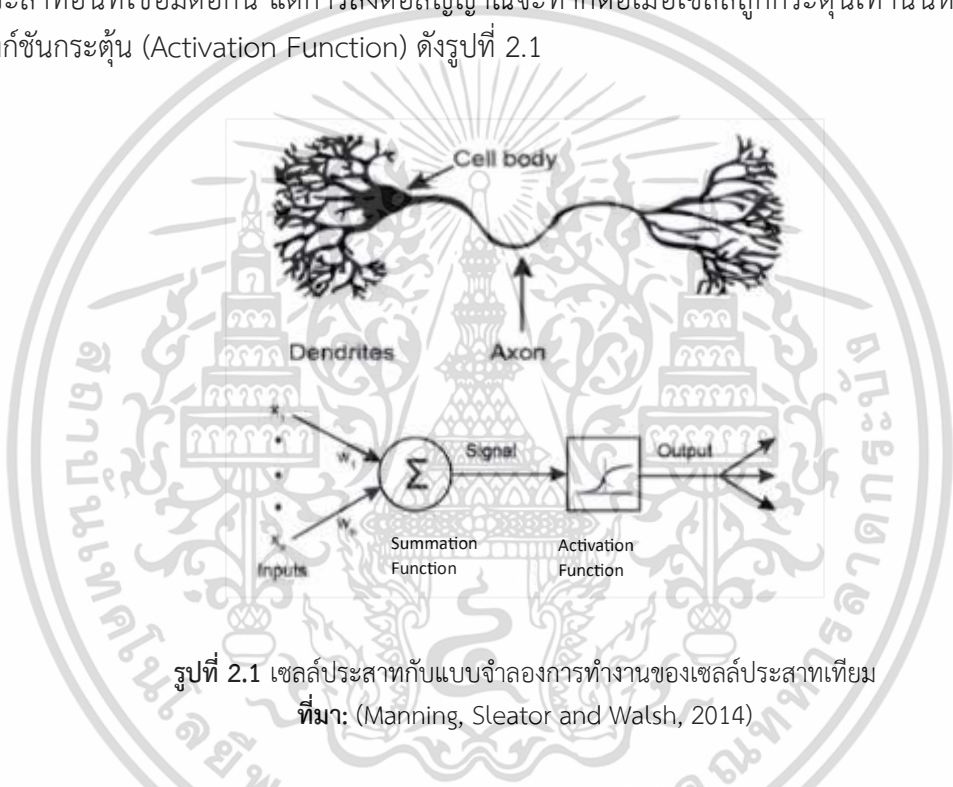
โดยที่

R_N^2 มีค่ามากกว่า R_{CS}^2 เสมอ

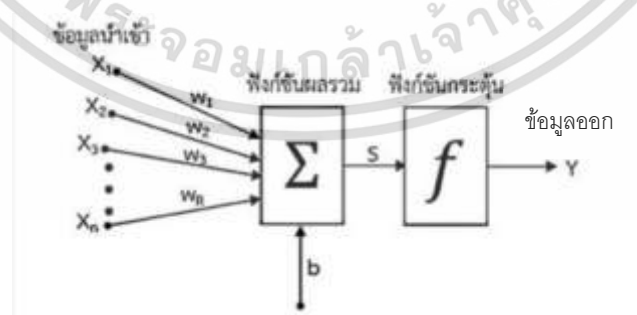
ขั้นตอนที่ 7 นำตัวแบบไปใช้ทำนายตัวแปรตาม

2.3 โครงข่ายประสาทเทียม (Artificial Neural Networks) (คັນสนีย์, 2556)

โครงข่ายประสาทเทียมถูกสร้างขึ้นมาเพื่อเลียนแบบการทำงานของที่ซับซ้อนของเซลล์สมองมนุษย์ประกอบด้วยเซลล์ประสาท (Neuron) เป็นจำนวนมาก ส่วนประกอบของเซลล์ที่สำคัญ ได้แก่ โยประสาทนำเข้า (Dendrites) ทำหน้าที่รับสัญญาณจากเซลล์ประสาทอื่นๆ ที่เชื่อมกันเข้าสู่ตัวเซลล์ (Cell Body) ทำการรวมสัญญาณ ส่วนโยประสาทนำออก (Axon) ทำการส่งต่อข้อมูลไปยังเซลล์ประสาทอื่นที่เชื่อมต่อกัน แต่การส่งต่อสัญญาณจะทำก็ต่อเมื่อเซลล์ถูกกระตุ้นเท่านั้นหรือเรียกว่า ฟังก์ชันกระตุ้น (Activation Function) ดังรูปที่ 2.1



รูปที่ 2.1 เซลล์ประสาทกับแบบจำลองการทำงานของเซลล์ประสาทเทียม
ที่มา: (Manning, Sleanor and Walsh, 2014)



รูปที่ 2.2 แบบจำลองการทำงานของโครงข่ายประสาทเทียม
ที่มา: (อกนิษฐ์ และคณะ, 2562)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปที่ 2.2 จะเห็นว่าโครงข่ายประสาทเทียมแบ่งการทำงานออกเป็น 3 ชั้น คือ ชั้นข้อมูลเข้า (Input Layer) เป็นชั้นที่นำข้อมูลสู่โครงข่ายประสาทเทียม และส่งข้อมูลไปยังชั้นซ่อน (Hidden layer) ทำหน้าที่เพิ่มประสิทธิภาพในการจัดกลุ่มข้อมูล เพื่อให้ได้ผลลัพธ์ที่ดีที่สุด และส่งออกไปยังชั้นข้อมูลออก (Output Layer) ทำหน้าที่ส่งออกข้อมูลโดยผ่านการประมวลผลจากการเรียนรู้ของโครงข่ายประสาทเทียม (รัชนีวรรณ, 2564) การหาผลลัพธ์ของโครงข่ายประสาทเทียม มี 5 องค์ประกอบ ดังนี้

1. นำข้อมูลเข้า (Input Data) คือ X_1, X_2, \dots, X_n สู่ชั้นข้อมูลเข้า (Input Layer)
2. ค่าน้ำหนักถ่วง (Weights) คือ $W_{1j}, W_{2j}, \dots, W_{nj}$ มาจากการเรียนรู้ของโครงข่ายประสาทเทียม และปรับค่าให้เหมาะสม ซึ่งวิธีที่นิยมใช้ในการหาค่าน้ำหนักถ่วงคือการทำงานแบบการส่งค่าย้อนกลับ (Back Propagation)

หลักการหาค่าน้ำหนักถ่วงด้วยการทำงานแบบการส่งค่าย้อนกลับจะมีการคำนวณเพื่อหาค่าน้ำหนักถ่วงใหม่ทุก ๆ ชุดข้อมูล ทำเช่นนี้จนครบทุกชุดข้อมูล นับเป็นจำนวนรอบในการเรียนรู้ 1 รอบ (Epoch) ซึ่งในแต่ละรอบในการเรียนรู้จะให้ผลลัพธ์ที่แตกต่างกัน จะทำให้ความคลาดเคลื่อนระหว่างค่าจริงกับค่าทำนายลดลงด้วยวิธีเคลื่อนลงตามความชัน (Gradient Descent Method)

$$\nabla SSE(W) = \left[\frac{\partial SSE}{\partial W_1}, \frac{\partial SSE}{\partial W_2}, \dots, \frac{\partial SSE}{\partial W_n} \right] \quad (2.18)$$

การหาค่าน้ำหนักถ่วงด้วยการทำงานแบบการส่งค่าย้อนกลับมีพารามิเตอร์ที่เกี่ยวข้อง ได้แก่ อัตราการเรียนรู้ (Learning Rate: η) และค่าโมเมนตัม (Momentum: α) ที่มีค่าอยู่ในช่วง $[0,1]$ โดยการปรับค่าน้ำหนักถ่วงสามารถคำนวณได้

$$W_{ij,new} = W_{ij,current} + \Delta W_{ij,current} + \alpha \Delta W_{ij,previous}; \Delta W_{ij,current} = \eta \delta_j X_{ij} \quad (2.19)$$

เมื่อ $W_{ij,new}$ คือ น้ำหนักถ่วงใหม่ของของตัวแปรอิสระที่ i ไปยัง โหนด j
 $W_{ij,current}$ คือ น้ำหนักถ่วงเดิมของของตัวแปรอิสระที่ i ไปยัง โหนด j
 η คือ อัตราการเรียนรู้
 X_{ij} คือ ค่าตัวแปรอิสระที่ i ไปยัง โหนด j
 α คือ ค่าโมเมนตัม
 $\Delta W_{ij,previous}$ คือ $\Delta W_{ij,current}$ ของจุดทดลองก่อนหน้า
 δ_j คือ ค่าความคลาดเคลื่อนของโหนด j

$\delta_j = \hat{Y}(1 - \hat{Y})(Y - \hat{Y})$ เมื่อโหนด j อยู่ในชั้นข้อมูลออก (Output Layer)

$\delta_j = output_j(1 - output_j) \sum_{downstream} W_{jk} \delta_k$ เมื่อโหนด j อยู่ในชั้นซ่อนและโหนด k อยู่ชั้นถัดไป

3. ฟังก์ชันผลรวม (Combination Function) มาจากผลรวมของข้อมูลในชั้นข้อมูลเข้าและค่าน้ำหนักถ่วง โดยที่ \sum จะทำหน้าที่รวมผลคูณของตัวแปรอิสระและค่าน้ำหนักถ่วงแต่ละตัว รวมทั้งความเอนเอียงหรือน้ำหนักถ่วง (Bias; b) ของแต่ละโหนดไว้ในฟังก์ชันผลรวม

ไม่ว่าการันเตฯ ฟังสน อักทิงหำมีมเห็ดดแบ่สงเนยหำ และตองยอึงองเงงใจ ของเอ็กส่ำวทุคครั้งที่มีกรนำไปใช้

4. ฟังก์ชันกระตุ้น (Activation Function) สำหรับการหาผลลัพธ์ของโครงข่ายประสาทเทียม จะใช้ฟังก์ชันซิกมอยด์ (Sigmoid Function) เป็นฟังก์ชันกระตุ้น คือ ฟังก์ชัน f

โดยทั่วไปฟังก์ชันกระตุ้นที่ใช้จะขึ้นอยู่กับลักษณะของผลลัพธ์ ในงานวิจัยนี้ผลลัพธ์มีค่าระหว่าง 0 และ 1 จึงใช้ฟังก์ชันซิกมอยด์เพื่อทำนายความน่าจะเป็น (Probability) ของข้อมูลออก (Output Data) ดังสมการ

$$f(x) = \frac{1}{1 + e^{-x}} \quad (2.20)$$

5. ข้อมูลออก (Output Data) หลังเสร็จสิ้นกระบวนการเรียนรู้ของโครงข่ายประสาทเทียม จะทำการปรับค่าต่างๆ ให้เหมาะสมกับชุดข้อมูลแล้วจะได้ข้อมูลออกที่ดีที่สุด ซึ่งกระบวนการดังกล่าวสามารถเขียนอยู่ในรูปสมการ ดังนี้

$$Y = f\left(\sum_{i=1}^d X_i W_i + b\right) \quad (2.21)$$

สำหรับการทำนายข้อมูลออก (Output) จะขึ้นอยู่กับค่าเทรชโฮลด์ (Threshold) หรือค่าเริ่มต้นที่เหมาะสม โดยทั่วไปถ้าค่าการจำแนกมี 2 ค่า ผลรวมของข้อมูลเข้าและค่าน้ำหนักถ่วง (Weight) มากกว่าหรือเท่ากับ 0.5 สามารถส่งข้อมูลไปชั้นถัดไปได้ แต่ถ้าผลรวมของข้อมูลเข้าและค่าน้ำหนักถ่วงน้อยกว่า 0.5 จะไม่มีข้อมูลออก

การหาผลลัพธ์ของโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับ

การออกแบบโครงสร้างของโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับประกอบด้วยจำนวนตัวแปรนำเข้า จำนวนชั้นซ่อน จำนวนโหนด และจำนวนตัวแปรออกหรือผลลัพธ์ ซึ่งสามารถกำหนดฟังก์ชันกระตุ้น อัตราการเรียนรู้ (Learning Rate) ค่าโมเมนตัม (Momentum) ตามความเหมาะสม และกำหนดจำนวนรอบในการเรียนรู้ (Epoch) โดยมีการกำหนดค่าพารามิเตอร์ ดังนี้

- 1) จำนวนชั้นข้อมูลเข้า (Input Layer) มี 1 ชั้น
- 2) จำนวนชั้นข้อมูลเข้า (Input Node) มีจำนวนเท่ากับปัจจัยหรือตัวแปรอิสระ (X_i) ที่ส่งผลต่อตัวแปรตาม (Y)
- 3) จำนวนชั้นซ่อน (Hidden Layer) มี 1 ชั้น เนื่องจากเป็นโครงข่ายแบบแพร่ย้อนกลับ
- 4) จำนวนโหนดในชั้นซ่อน (Hidden Node) ขึ้นอยู่กับความซับซ้อนของปัญหาไม่สามารถกำหนดเป็นเกณฑ์ตายตัวได้ จึงต้องเริ่มต้นด้วยหน่วยที่น้อยที่สุดที่สามารถเป็นไปได้ อย่างไรก็ตามปกติแล้วมักจะเลือกจำนวนโหนดในชั้นซ่อนให้มีจำนวนใกล้เคียงกับจำนวนโหนดในชั้นข้อมูลเข้า แต่ไม่จำเป็นต้องเท่ากันทุกครั้ง (รสรินและอรรษาพรณ, 2563)

5) อัตราการเรียนรู้ (Learning Rate) โดยปกติแล้วอัตราการเรียนรู้จะควบคุมขั้นตอนของการเรียนรู้ ถ้าอัตราการเรียนรู้มีค่าน้อยเกินไป การปรับค่าน้ำหนักจะเป็นไปทีละน้อย แต่ถ้าอัตราการเรียนรู้มีค่ามากเกินไป การปรับค่าน้ำหนักจะมีค่ามากเกินไปในแต่ละครั้ง ซึ่งอาจจะทำให้ค่าน้ำหนักแกว่งกวัด (Oscillate) ดังนั้นการเลือกอัตราการเรียนรู้ที่ดีที่สุดสามารถทำได้โดยใช้การลองผิดลองถูก (Trial and Error) โดยทั่วไปจะกำหนดอัตราการเรียนรู้ให้อยู่ระหว่าง 0.05-0.25 (คันสนีย์, 2556)

6) โมเมนตัม (Momentum) เป็นค่าที่อยู่ช่วงในการถ่วงน้ำหนักที่เกิดการแกว่งกวัดจากอัตราการเรียนรู้ และการเลือกค่าโมเมนตัมที่ดีที่สุดก็ทำได้เช่นเดียวกับการเลือกอัตราการเรียนรู้ที่ดีที่สุด โดยทั่วไปจะกำหนดให้อยู่ระหว่าง 0.6-0.9 (Larose and Larose, 2014)

7) จำนวนรอบในการเรียนรู้ (Epoch) หากกำหนดจำนวนรอบในการเรียนรู้ให้กับระบบโครงข่ายประสาทเทียมน้อยไป ผลการทำนายค่าก็จะให้ผลลัพธ์ที่ผิดพลาดมาก เนื่องจากกระบวนการเรียนรู้ยังไม่เสร็จสิ้น แต่ถ้าหากกำหนดจำนวนรอบในการเรียนรู้ให้กับระบบโครงข่ายประสาทเทียมมากเกินไปจะทำให้ระบบโครงข่ายประสาทเทียมเกิดการต่อต้านกับข้อมูลที่เรียนรู้ ซึ่งส่งผลให้การทำนายค่าได้ผลลัพธ์ที่ผิดพลาด (รสรินและอัจฉราพรรณ, 2563)

8) จำนวนชั้นข้อมูลออก (Output Layer) มี 1 ชั้น

9) จำนวนโหนดในชั้นข้อมูลออก (Output Node) ขึ้นอยู่กับการกำหนดค่าของตัวแปรตาม ถ้าตัวแปรตามมี 2 ค่า คือ 0, 1 จำนวนโหนดในชั้นข้อมูลออกจะมี 2 โหนด

2.4 การเปรียบเทียบประสิทธิภาพการทำนาย

จากการเปรียบเทียบตัวแบบโครงข่ายประสาทเทียมด้วยค่าคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) ได้ตัวแบบโครงข่ายประสาทเทียมที่ดีที่สุด จากนั้นนำตัวแบบโครงข่ายประสาทเทียมที่ดีที่สุดมาเปรียบเทียบกับตัวแบบการวิเคราะห์การถดถอยลوجิสติกทวิภาค โดยใช้เกณฑ์การวัดประสิทธิภาพดังนี้

ตารางที่ 2.1 เมทริกซ์ความสับสน (Confusion Matrix)

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์ที่ได้จากการทำนาย (Predicted Class)	
	ระดับอาการรุนแรง	ระดับอาการไม่รุนแรง
ระดับอาการรุนแรง	TP	FN
ระดับอาการไม่รุนแรง	FP	TN

โดยที่

True Positive (TP) คือ ระดับความรุนแรงของผู้ป่วยจากการทำนายมีอาการรุนแรง และผลลัพธ์จริงมีอาการรุนแรง

True Negative (TN) คือ ระดับความรุนแรงของผู้ป่วยจากการทำนายมีอาการไม่รุนแรง และผลลัพธ์จริงมีอาการไม่รุนแรง

False Positive (FP) คือ ระดับความรุนแรงของผู้ป่วยจากการทำนายมีอาการรุนแรง และผลลัพธ์จริงมีอาการไม่รุนแรง

False Negative (FN) คือ ระดับความรุนแรงของผู้ป่วยจากการทำนายมีอาการไม่รุนแรง และผลลัพธ์จริงมีอาการรุนแรง

2.4.1 ค่าความแม่นยำ (Accuracy) คือ อัตราส่วนของจำนวนการทำนายถูกต้องว่าเป็นจริงหรือเป็นเท็จต่อข้อมูลทั้งหมดที่ใช้ในการทำนาย ความแม่นยำวัดประสิทธิภาพของการทำนายในรูปแบบของอัตราส่วน (เอกพันธ์, 2563) มีสูตรการคำนวณ ดังนี้

$$accuracy = \frac{TP + TN}{TP + FN + TN + FP} \quad (2.22)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.4.2 ค่าเรียกคืน (Recall) ในทางสถิติเรียกว่า ค่าความไว (Sensitivity) คือ อัตราส่วนการทำนายที่เป็นจริงต่อจำนวนที่ทำนายเป็นจริงทั้งหมด (เอกพันธ์, 2563) มีสูตรการคำนวณดังนี้

$$recall = \frac{TP}{TP + FN} \quad (2.23)$$

2.4.3 ค่าวัดประสิทธิภาพโดยใช้พื้นที่ใต้กราฟ (Area Under a Curve: AUC) คือ พื้นที่ใต้กราฟโค้ง (Receiver Operating Characteristic: ROC) เป็นการวัดประสิทธิภาพของตัวแบบการทำนายว่าสามารถทำนายได้ถูกต้องและแม่นยำมากเท่าใด ถ้าตัวแบบการทำนายใดมีค่าพื้นที่ใต้กราฟ ROC Curve (AUC) สูง หมายความว่าตัวแบบการทำนายนั้นมีประสิทธิภาพสูง (เอกชัย, 2557)

ค่า AUC > 0.9 แสดงว่าตัวแบบที่ใช้ในการทำนายมีประสิทธิภาพสูงมาก

ค่า AUC อยู่ระหว่าง 0.75-0.9 แสดงว่าตัวแบบที่ใช้ในการทำนายมีประสิทธิภาพสูง

ค่า AUC อยู่ระหว่าง 0.5-0.74 แสดงว่าตัวแบบที่ใช้ในการทำนายมีประสิทธิภาพค่อนข้างต่ำ

ค่า AUC < 0.5 แสดงว่าตัวแบบที่ใช้ในการทำนายไม่มีประสิทธิภาพ (Ray et al., 2010)

2.4.4 ค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error: MSE) เป็นวิธีการวัดประสิทธิภาพที่พิจารณาความแตกต่างระหว่างค่าจริงกับค่าทำนายแล้วยกกำลังสอง เพื่อประสิทธิภาพที่ดีที่สุด จะต้องมียค่า MSE น้อยที่สุด (อกนิษฐ์และคณะ, 2562) มีสูตรการคำนวณ MSE เป็นดังนี้

$$MSE = \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n} \quad (2.24)$$

เมื่อ Y_i คือ ค่าที่เกิดขึ้นจริง

\hat{Y}_i คือ ค่าที่เกิดจากการทำนาย

n คือ จำนวนตัวอย่างทั้งหมด

2.5 งานวิจัยที่เกี่ยวข้อง

Rawat (2019) การทำนายการเป็นโรคหัวใจ ด้วยการเรียนรู้ของเครื่อง (Machine Learning) 7 แบบ เพื่อเปรียบเทียบประสิทธิภาพในการทำนายปัจจัยที่ส่งผลต่อการเป็นโรคหัวใจ ได้แก่ Logistic Regression, Random Forest Classifier, SVM, Naive Bayes Classifier, Decision Tree Classifier, LightGBM และ XGBoost บนชุดข้อมูล 303 ชุด โดยแบ่งข้อมูลออกเป็นข้อมูลชุดเรียนรู้ 80% และข้อมูลชุดทดสอบ 20% ซึ่งจะใช้เมทริกซ์ความสับสน (Confusion Matrix) ในการประเมินประสิทธิภาพการทำนาย ผลการทดสอบพบว่าการวิเคราะห์การถดถอยลอจิสติกมีความแม่นยำ 86.4% สำหรับข้อมูลชุดเรียนรู้ และ 80.4% สำหรับข้อมูลชุดทดสอบ เนื่องจากการวิเคราะห์การถดถอยลอจิสติกมีความแตกต่างต่ำสุดระหว่างความแม่นยำของชุดเรียนรู้และชุดทดสอบ สรุปว่าการวิเคราะห์การถดถอยลอจิสติกเป็นแบบจำลองที่มีประสิทธิภาพสูงที่สุดในกรณีนี้ การวิเคราะห์การถดถอยลอจิสติกทวิภาคจึงนิยมใช้ในการวิเคราะห์หาปัจจัยเสี่ยงของโรคต่างๆ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Abdulaal et al. (2020) นำเสนอ AI ในการสร้างระบบการให้คะแนนความเสี่ยงต่อการเสียชีวิต โดยใช้โครงข่ายประสาทเทียม (Artificial Neural Networks) มีวัตถุประสงค์เพื่อสร้างระบบคำนวณความเสี่ยงต่อการเสียชีวิตของผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 ช่วยในการตัดสินใจด้านการจัดการทางคลินิกได้อย่างรวดเร็ว ทำให้อัตราความเสี่ยงต่อการเสียชีวิตของผู้ป่วยลดลงและสามารถคาดการณ์การเสียชีวิตเฉพาะผู้ป่วยด้วยความแม่นยำ (Accuracy) 86.25% ความไว (Sensitivity) 87.50% และความจำเพาะ (Specificity) 85.94%

Zhou et al. (2020) ศึกษาปัจจัยเสี่ยงต่อการเสียชีวิตของในผู้ใหญ่ที่ติดเชื้อไวรัสโคโรนา 2019 โดยใช้การวิเคราะห์การถดถอยลอจิสติกทวิภาค เพื่อทำนายความเสี่ยงในการเสียชีวิตของผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 เปรียบเทียบระหว่างผู้ป่วยที่รอดชีวิตกับไม่รอดชีวิต ผลการวิจัยสรุปว่าโอกาสของการเสียชีวิตของผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 ในโรงพยาบาลมีความสัมพันธ์กับอายุของผู้ป่วยคือผู้ป่วยที่มีอายุมากขึ้น โอกาสในการเสียชีวิตก็เพิ่มขึ้นด้วย (Odds Ratio = 1.10, $p = 0.0043$) และผู้ป่วยส่วนใหญ่เสียชีวิตโดยมีโรคร่วมที่พบมากที่สุดคือโรคความดันโลหิตสูง โรคเบาหวาน และโรคหลอดเลือดหัวใจ

Rashidi et al. (2021) ได้สำรวจการประยุกต์ปัญญาประดิษฐ์ (AI) ในการจัดการกับโรคติดเชื้อไวรัสโคโรนา 2019 พบว่า AI สามารถช่วยในการวินิจฉัยโรคติดเชื้อไวรัสโคโรนา 2019 ทำนายความเสี่ยงต่อการเสียชีวิต คาดการณ์ผลของการรักษา และทำนายการแพร่ระบาด สรุปได้ว่า AI มีบทบาทสำคัญในการเพิ่มประสิทธิภาพในการรักษาทางการแพทย์ โดยการช่วยให้การวินิจฉัยและการรักษารวดเร็วและมีความแม่นยำมากขึ้น

Mohammadi et al. (2021) ศึกษาเปรียบเทียบลักษณะของผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 ใน 6 จังหวัดของอิหร่าน โดยใช้แบบจำลองโครงข่ายประสาทเทียมหลายชั้น Perceptron และการวิเคราะห์การถดถอยลอจิสติกทวิภาค เพื่อระบุลักษณะของผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 ผลการวิจัยพบว่าลักษณะผู้ป่วยที่ติดเชื้อ SARS-CoV-2 มีโรคร่วมที่พบบ่อย ได้แก่ โรคความดันโลหิตสูง โรคเบาหวาน โรคปอดเรื้อรัง และโรคหลอดเลือดหัวใจ โดยที่การวิเคราะห์การถดถอยลอจิสติกทวิภาค มีค่า Cox&Snell R Square เท่ากับ 0.683 บ่งชี้ว่าแบบจำลองมีความเหมาะสมกับข้อมูล มีค่าวัดประสิทธิภาพโดยใช้พื้นที่ใต้กราฟเท่ากับ 0.992 ความแม่นยำ 98.7% และความไว 99.1% ส่วนโครงข่ายประสาทเทียมหลายชั้น Perceptron มีค่าวัดประสิทธิภาพโดยใช้พื้นที่ใต้กราฟ 0.999 ความแม่นยำ 99.4% และความไว 100% สรุปได้ว่าโครงข่ายประสาทเทียมหลายชั้น Perceptron มีประสิทธิภาพดีกว่าการวิเคราะห์การถดถอยลอจิสติกทวิภาค

Ding et al. (2021) พัฒนาตัวแบบเพื่อทำนายการเสียชีวิตในโรงพยาบาลในระยะเริ่มแรกของผู้ป่วยโรคตับอ่อนอักเสบเฉียบพลัน (AP) และเปรียบเทียบประสิทธิภาพในการทำนายการเสียชีวิตของผู้ป่วยโรคตับอ่อนอักเสบเฉียบพลันสำหรับข้อมูล 337 ชุด โดยแบ่งข้อมูลออกเป็นชุดเรียนรู้ 80% และชุดทดสอบ 20% โดยผลการวิจัยพบว่า โครงข่ายประสาทเทียมแบบแพร่ย้อนกลับมีความแม่นยำ 76.9% ความไว 66.6% ความจำเพาะ 66.1% และมีค่าวัดประสิทธิภาพโดยใช้พื้นที่ใต้กราฟ 0.769 และเทคนิคการวิเคราะห์การถดถอยลอจิสติกทวิภาคมีความแม่นยำ 66% ความไว 34.6% ความจำเพาะ 92.3% และมีค่าวัดประสิทธิภาพโดยใช้พื้นที่ใต้กราฟ เท่ากับ 0.607 สรุปได้ว่าการใช้โครงข่ายประสาทเทียมแบบแพร่ย้อนกลับทำนายการเสียชีวิตในโรงพยาบาลในผู้ป่วยโรคตับอ่อนอักเสบเฉียบพลันเป็นครั้งแรก สามารถคัดกรองผู้ป่วยโรคตับอ่อนอักเสบเฉียบพลันที่มีความเสี่ยงสูงต่อการเสียชีวิตได้ดีกว่าการวิเคราะห์การถดถอยลอจิสติกทวิภาค

ไม่ว่าการนี้... เอกสารนี้เป็นลิขสิทธิ์ของ... ไม่ว่าการนี้... เอกสารนี้เป็นลิขสิทธิ์ของ... ไม่ว่าการนี้... เอกสารนี้เป็นลิขสิทธิ์ของ...

บทที่ 3

วิธีการดำเนินงานวิจัย

การวิจัยในครั้งนี้ทำการเปรียบเทียบประสิทธิภาพในการทำนายผลระดับความรุนแรงของอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 ด้วยการวิเคราะห์การถดถอยลอจิสติกทวิภาคและโครงข่ายประสาทเทียม ซึ่งเป็นการศึกษาวิจัยเชิงสำรวจ (Survey Research) รูปแบบการศึกษาย้อนหลัง (Retrospective Study) เพื่อวิเคราะห์และเปรียบเทียบประสิทธิภาพในการทำนายปัจจัยที่มีผลต่อระดับความรุนแรงของผู้ป่วย โดยมีขั้นตอนดังต่อไปนี้

- 3.1 การวางแผนการวิจัย
- 3.2 วิธีดำเนินการวิจัย
- 3.3 การสร้างตัวแบบทำนาย
 - 3.3.1 การถดถอยลอจิสติกทวิภาค
 - 3.3.2 โครงข่ายประสาทเทียม
- 3.4 การเปรียบเทียบประสิทธิภาพตัวแบบทำนาย
- 3.5 แผนการดำเนินงานวิจัย

3.1 การวางแผนการวิจัย

3.1.1 กลุ่มตัวอย่าง คือ ผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 ที่เข้ารับการรักษาที่กรมอนามัย กระทรวงสาธารณสุขแห่งหนึ่งในกรุงเทพมหานคร ระหว่างเดือนมีนาคม-ธันวาคม พ.ศ. 2565 คัดเลือกกลุ่มตัวอย่างแบบเจาะจง (Purposive Sampling) ได้แก่ ผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 ที่มีผลตรวจ Antigen test kit (ATK) เป็นบวก (Covid Positive) จึงได้กลุ่มตัวอย่างที่เข้าเกณฑ์การคัดเลือกจำนวนทั้งหมด 652 คน ซึ่งแบ่งเป็นผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 ในระดับรุนแรง 152 คน และไม่รุนแรง 500 คน

3.1.2 เครื่องที่ใช้ในการวิจัย คือ แบบสอบถามผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 (Novel Corona-2) โดยรายละเอียดของแบบสอบถามข้อมูลผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 มี 3 ส่วน ดังนี้

ส่วนที่ 1 ข้อมูลทั่วไป ประกอบด้วย เพศ อายุ อาชีพ น้ำหนัก ส่วนสูง โรคประจำตัว การสูบบุหรี่ และการดื่มเครื่องดื่มที่มีแอลกอฮอล์

ส่วนที่ 2 ข้อมูลทางคลินิก ประกอบด้วย ผลการตรวจหาเชื้อไวรัสโคโรนา 2019 ลักษณะอาการ ประวัติการเจ็บป่วย ประวัติการรักษา และปัจจัยเสี่ยงต่างๆ

ส่วนที่ 3 ประวัติการได้รับวัคซีนป้องกันโรคติดเชื้อไวรัสโคโรนา 2019

3.1.3 การรวบรวมข้อมูล ทำการเก็บข้อมูลจากผู้ป่วยโดยตรง ดำเนินการโดยเจ้าหน้าที่กรมอนามัยทำการสอบถามอาการของผู้ป่วยตามแบบสอบถามผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 (Novel Corona-2) เก็บข้อมูลตั้งแต่เดือนมีนาคม-ธันวาคม พ.ศ. 2565 ซึ่งได้ข้อมูลจำนวน 652 ชุด ผู้วิจัยจึงทำการป้อนข้อมูลที่ได้อ้างอิงในโปรแกรม Microsoft Office Excel เพื่อเป็นฐานข้อมูลและใช้โปรแกรม RStudio ในการสร้างตัวแบบสำหรับการทำนายด้วยการวิเคราะห์การถดถอยลอจิสติกทวิภาคและโครงข่ายประสาทเทียม แสดงดังตารางที่ 3.1

เอกสารนี้เป็นเอกสารสงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.1 ตัวอย่างข้อมูลทั้งหมดของผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 ที่เข้ารับการรักษาในกรมอนามัย กระทรวงสาธารณสุขแห่งหนึ่งในกรุงเทพมหานคร ตั้งแต่เดือนมีนาคม-ธันวาคม พ.ศ. 2565

No.	Sex	Age	Weight	Height	BMI	SBP	DBP	SpO2	HR	T	NOV	SMK	DAB	HT	DM	CAD	PN	SL
1	1	20	49	155	20.4	112	65	97	113	37.5	3	0	0	0	0	0	0	1
2	0	14	45	168	15.94	109	75	97	95	35.1	2	0	0	0	0	0	0	0
3	1	29	56	169	19.61	96	72	97	87	36.4	3	0	0	0	0	0	0	0
4	0	41	65	170	22.49	104	79	96	103	36.5	2	0	0	0	0	0	0	1
5	1	32	45	155	18.73	93	67	95	84	36.4	4	0	0	0	0	0	0	0
6	0	38	73	175	23.84	106	76	98	106	37	2	1	1	0	0	0	0	0
7	1	32	53	160	20.7	111	84	98/102	96	36.1		0	0	0	0	0	0	0
8	0	44	80	183	23.89	106	76	96	102	37	3	1	1	0	0	0	0	0
9	1	43	47	160	18.36	130	84	96	89	37.8	3	0	0	0	0	0	0	1
10	0	43	60	170	20.76	93	67	91	80	36.5	3	0	0	0	0	0	0	1
11	1	31	55	160	21.48	98	55	98	92	35.1	4	0	0	0	0	0	0	0
12	1	34	49	151	21.49	106	76	96	101	37.7	2	1	1	0	0	0	0	0
13	1	38	70	168	24.8	103	69	96	115	36	4	0	0	0	0	0	0	0
14	0	23	55	157	22.31	101	57	96	134	36	2	0	0	0	0	0	0	1
15	1	18	57	157	23.12				98	36.3	2							0
16	1	22	39	158	15.62	130	68	97	95	37.2	4	0	0	0	0	0	0	0
17	1	59	89	165	32.69	115	82	96	96	37.6	4	0	0	0	1	0	0	0
18	0	28	81	179	25.28	124	95	98	61	37.2	2	0	0	0	0	0	0	0
19	1	40	86	165	31.59	115	82	97	94	36	3	0	0	0	1	0	0	1
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
652	1	20	53	64	19.71	142	80	97	90	38.2	4	0	1	0	0	0	0	0

จากตารางที่ 3.1 พบว่าข้อมูลชุดที่ 7 ระดับออกซิเจนในเลือดมีการบันทึกข้อมูลซ้ำ คือ 98/102 ทำให้เกิดความซ้ำซ้อน และการสูญหามีข้อมูลสูญหาย ส่วนข้อมูลชุดที่ 15 ค่าความดันโลหิตขณะหัวใจบีบตัว ค่าความดันโลหิตขณะหัวใจคลายตัว ระดับออกซิเจนในเลือด การสูบบุหรี่ การดื่มเครื่องดื่มที่มีแอลกอฮอล์มีข้อมูลสูญหาย

3.2 วิธีดำเนินการวิจัย

3.2.1 การเตรียมข้อมูล (Data Preprocessing) คือ การตรวจสอบคุณภาพของข้อมูล ทำข้อมูลให้อยู่ในรูปแบบที่เหมาะสมก่อนนำไปใช้งานในขั้นตอนอื่น เพื่อช่วยให้การวิเคราะห์ข้อมูลมีประสิทธิภาพมากยิ่งขึ้น (ไกรศักดิ์, 2564)

1) การทำความเข้าใจข้อมูล (Data Understanding) จากการคัดกรองข้อมูลก่อนหน้านี ทำให้ข้อมูลประกอบด้วยตัวแปรอิสระ 17 ตัวแปร ได้แก่ เพศ อายุ น้ำหนัก ส่วนสูง ดัชนีมวลกาย ค่าความดันโลหิตขณะหัวใจบีบตัว ค่าความดันโลหิตขณะหัวใจคลายตัว ระดับออกซิเจนในเลือด อัตราการเต้นของหัวใจ อุณหภูมิร่างกาย จำนวนวัคซีนที่ได้รับ การสูบบุหรี่ การดื่มเครื่องดื่มที่มีแอลกอฮอล์ การเป็นโรคความดันโลหิตสูง การเป็นโรคเบาหวาน การเป็นโรคหลอดเลือดหัวใจ การเป็นโรคปอด และตัวแปรตามคือระดับความรุนแรงของอาการผู้ป่วย โดยตัวแปรอิสระแต่ละตัวมีความหมาย มาตราวัด หน่วยวัด และรายละเอียดตัวแปร ดังตารางที่ 3.2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.2 รายละเอียดข้อมูลของผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019

ตัวแปร	ความหมาย	มาตรวัด	หน่วย	รายละเอียดตัวแปร
X_1	เพศ	นามบัญญัติ	ชาย, หญิง	0 = ชาย 1 = หญิง
X_2	อายุ	เชิงตัวเลข	ปี	ค่าตัวเลขอายุ
X_3	น้ำหนัก	เชิงตัวเลข	กิโลกรัม	ค่าตัวเลขน้ำหนัก
X_4	ส่วนสูง	เชิงตัวเลข	เซนติเมตร	ค่าตัวเลขส่วนสูง
X_5	ค่าดัชนีมวลกาย	เชิงตัวเลข	กิโลกรัม/ตารางเมตร	ค่าตัวเลขดัชนีมวลกาย
X_6	ค่าความดันโลหิตขณะหัวใจบีบตัว	เชิงตัวเลข	มิลลิเมตรปรอท	ค่าตัวเลขค่าตัวเลข
X_7	ค่าความดันโลหิตขณะหัวใจคลายตัว	เชิงตัวเลข	มิลลิเมตรปรอท	ค่าตัวเลขค่าตัวเลข
X_8	ระดับออกซิเจนในเลือด	เชิงตัวเลข	เปอร์เซ็นต์	ค่าตัวเลขระดับออกซิเจนในเลือด
X_9	อัตราการเต้นของหัวใจ	เชิงตัวเลข	ครั้ง/นาที	ค่าตัวเลขอัตราการเต้นของหัวใจ
X_{10}	อุณหภูมิร่างกาย	เชิงตัวเลข	องศาเซลเซียส	ค่าตัวเลขอุณหภูมิร่างกาย
X_{11}	จำนวนวัคซีนที่ได้รับ	เชิงตัวเลข	เข็ม	ค่าตัวเลขจำนวนวัคซีนที่ได้รับ
X_{12}	การสูบบุหรี่	นามบัญญัติ	ไม่สูบ, สูบ	0 = ไม่สูบ 1 = สูบ
X_{13}	การดื่มเครื่องดื่มที่มีแอลกอฮอล์	นามบัญญัติ	ไม่ดื่ม, ดื่ม	0 = ไม่ดื่ม 1 = ดื่ม
กลุ่มโรคประจำตัว				
X_{14}	การเป็นโรคความดันโลหิตสูง	นามบัญญัติ	ไม่เป็น, เป็น	0 = ไม่เป็น 1 = เป็น
X_{15}	การเป็นโรคเบาหวาน	นามบัญญัติ	ไม่เป็น, เป็น	0 = ไม่เป็น 1 = เป็น
X_{16}	การเป็นโรคหลอดเลือดหัวใจ	นามบัญญัติ	ไม่เป็น, เป็น	0 = ไม่เป็น 1 = เป็น
X_{17}	การเป็นโรคปอดบวม	นามบัญญัติ	ไม่เป็น, เป็น	0 = ไม่เป็น 1 = เป็น
Y	ระดับความรุนแรงของอาการผู้ป่วย	นามบัญญัติ	ไม่รุนแรง, รุนแรง	0 = ไม่รุนแรง 1 = รุนแรง

2) การทำความสะอาดข้อมูล (Data Cleaning) ในการเตรียมข้อมูลพบว่ามีข้อมูลบางส่วนที่มีความซ้ำซ้อนและสูญหายคือข้อมูลชุดที่ 7 มีการบันทึกข้อมูลซ้ำ และข้อมูลชุดที่ 15 มีข้อมูลสูญหาย ผู้วิจัยจึงตัดข้อมูลเหล่านี้ออก 2 ชุด จากข้อมูลทั้งหมด 652 ชุด จึงเหลือข้อมูลที่ใช้ในการวิเคราะห์ทั้งสิ้น 650 ชุด ซึ่งแบ่งเป็นผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 ในระดับรุนแรง 152 คน และไม่รุนแรง 498 คน เอกสารนี้คือเอกสารที่ส่งมอบไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.3 สถิติเชิงพรรณนาของข้อมูลผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019

ตัวแปร	จำนวน	ค่าต่ำสุด	ค่าสูงสุด	ค่าเฉลี่ย
เพศ	ชาย 300 หญิง 350	-	-	-
อายุ	-	2	97	39.32
น้ำหนัก	-	11	130	61.27
ส่วนสูง	-	60	190	161.08
ค่าดัชนีมวลกาย	-	10.81	47.22	23.41
ค่าความดันโลหิตขณะหัวใจบีบตัว	-	74	156	112.86
ค่าความดันโลหิตขณะหัวใจคลายตัว	-	30	103	73.62
ระดับออกซิเจนในเลือด	-	91	99	96.89
อัตราการเต้นของหัวใจ	-	60	136	96.82
อุณหภูมิร่างกาย	-	35.1	40.3	37.04
จำนวนวัคซีนที่ได้รับ	-	0	4	2.58
การสูบบุหรี่	ไม่สูบ 547 สูบ 103	-	-	-
การดื่มเครื่องดื่มที่มีแอลกอฮอล์	ไม่ดื่ม 518 ดื่ม 132	-	-	-
การเป็นโรคความดันโลหิต	ไม่เป็น 587 เป็น 63	-	-	-
การเป็นโรคเบาหวาน	ไม่เป็น 588 เป็น 62	-	-	-
การเป็นโรคหลอดเลือดหัวใจ	ไม่เป็น 602 เป็น 48	-	-	-
การเป็นโรคปอดบวม	ไม่เป็น 601 เป็น 49	-	-	-

จากตารางที่ 3.3 จะพบว่าข้อมูลของผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 ทั้งหมด 650 คน เป็นเพศชาย 300 คน และเพศหญิง 350 คน จากผู้ป่วยทั้งหมดมีอายุเฉลี่ย 39.32 ปี มีน้ำหนักเฉลี่ย 61.27 กิโลกรัม มีส่วนสูงเฉลี่ย 161.08 เซนติเมตร มีค่าดัชนีมวลกายเฉลี่ย 23.41 มีค่าความดันโลหิตขณะหัวใจบีบตัวเฉลี่ย 112.86 มีค่าความดันโลหิตขณะหัวใจคลายตัว 73.62 ระดับออกซิเจนในเลือดเฉลี่ย 96.89 อัตราการเต้นของหัวใจเฉลี่ย 96.82 อุณหภูมิร่างกายเฉลี่ย 37.04 องศา จำนวนวัคซีนที่ได้รับเฉลี่ย 2.58 เข็ม และพบว่าไม่สูบบุหรี่ 103 คน ไม่สูบบุหรี่ 547 คน มีผู้ที่ดื่มเครื่องดื่มที่มีแอลกอฮอล์ 132 คน ไม่ดื่มเครื่องดื่มที่มีแอลกอฮอล์ 518 คน มีผู้ที่เป็นโรคความดันโลหิต 63 คน ไม่เป็นโรคความดันโลหิต 587 คน มีผู้ที่เป็นโรคเบาหวาน 62 คน ไม่เป็นโรคเบาหวาน 588 คน มีผู้ที่เป็นโรคหลอดเลือดหัวใจ 48 คน ไม่เป็นโรคหลอดเลือดหัวใจ 602 คน และมีผู้ที่เป็นโรคปอดบวม 49 คน ไม่เป็นโรคปอดบวม 601 คน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3) การแปลงข้อมูล (Data Transformation) เป็นการแปลงค่าของข้อมูลให้อยู่ในรูปมาตรฐาน (Normalization) โดยการแปลงค่าของตัวแปรอิสระให้อยู่ในช่วงที่กำหนด [0, 1] เพื่อให้ตัวแปรอิสระแต่ละตัวมีความเท่าเทียมกัน โดยมีสูตรการคำนวณดังนี้

$$X^* = \frac{X - \min(X)}{\max(X) - \min(X)} \tag{2.25}$$

โดยที่

- X^* คือ ค่าที่ได้จากการแปลงค่าอยู่ในรูปการทำให้เป็นปรกติ
- X คือ ค่าปัจจุบันที่นำมาแปลงค่าอยู่ในรูปการทำให้เป็นปรกติ
- $\min(X)$ คือ ค่าข้อมูลที่มีค่าน้อยที่สุดในชุดข้อมูล
- $\max(X)$ คือ ค่าข้อมูลที่มีค่ามากที่สุดในการชุดข้อมูล

ตารางที่ 3.4 การแปลงข้อมูลให้อยู่ในรูปปรกติมาตรฐาน

No	x_1^*	x_2^*	x_3^*	x_4^*	x_5^*	x_6^*	x_7^*	x_8^*	x_9^*	x_{10}^*	x_{11}^*	x_{12}^*	x_{13}^*	x_{14}^*	x_{15}^*	x_{16}^*	x_{17}^*	Y
1	1	0.187	0.336	0.738	0.278	0.524	1.000	0.750	0.413	0.288	0.50	0	1	0	0	0	0	1
2	1	0.250	0.487	0.730	0.491	0.292	0.342	0.750	0.613	0.192	0.50	0	0	0	0	0	0	0
3	1	0.052	0.100	0.469	0.134	0.146	0.301	1.000	0.493	0.326	0.75	0	0	0	0	0	0	0
4	1	0.812	0.411	0.638	0.508	0.865	0.630	0.750	0.293	0.192	0.50	0	0	1	0	0	0	1
5	1	0.302	0.537	0.869	0.391	0.329	0.369	0.750	0.493	0.346	0.75	0	0	0	0	0	0	0
6	0	0.385	0.512	0.907	0.327	0.231	0.506	0.625	0.453	0.423	1.00	0	0	0	0	0	0	0
7	1	0.437	0.428	0.830	0.306	0.609	0.904	0.625	0.586	0.500	0.75	0	0	0	0	0	0	0
8	1	0.072	0.117	0.538	0.109	0.670	0.753	0.875	0.573	0.000	0.50	0	0	0	0	0	0	0
9	0	0.416	0.327	0.807	0.207	0.268	0.575	0.875	0.453	0.173	0.50	1	1	0	0	0	0	1
10	0	0.427	0.411	0.884	0.241	0.829	0.684	0.875	0.386	0.288	0.50	1	1	1	0	0	0	1
11	1	0.614	0.310	0.730	0.251	0.609	0.904	0.500	0.386	0.403	0.75	0	0	0	0	0	0	0
12	0	0.239	0.294	0.823	0.156	0.768	0.808	1.000	0.826	0.192	0.00	0	0	0	0	0	0	0
13	0	0.208	0.663	0.884	0.510	0.402	0.657	0.625	0.973	0.192	0.50	0	0	0	0	0	0	0
14	1	0.666	0.361	0.800	0.254	0.804	0.630	0.625	0.186	0.403	0.00	0	0	1	0	0	0	1
15	0	0.895	0.344	0.769	0.260	0.122	0.301	0.750	0.560	0.442	0.00	0	0	0	0	0	0	0
16	0	0.708	0.437	0.823	0.323	0.609	0.657	0.625	0.560	0.230	1.00	0	0	0	0	0	0	0
17	1	0.177	0.344	0.784	0.247	0.609	0.904	0.875	0.426	0.557	0.25	0	0	0	0	0	0	0
18	1	0.000	0.000	0.153	0.175	0.146	0.301	0.750	0.293	0.403	0.00	0	0	0	0	0	0	0
19	0	0.697	0.411	0.700	0.425	0.329	0.369	0.875	0.480	0.288	0.50	0	1	0	0	0	0	1
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
650	1	0.187	0.352	0.030	0.244	0.829	0.684	0.750	0.386	0.596	0	0	1	0	0	0	0	0

จากตารางที่ 3.4 จะพบว่าหลังจากการแปลงข้อมูลให้อยู่ในรูปปรกติมาตรฐานด้วยวิธีการทำให้เป็นปรกติน้อยที่สุด-มากที่สุด ตัวแปรอิสระทุกตัวมีค่าอยู่ในช่วง [0,1]

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.5 รายละเอียดตัวแปรหลังจากแปลงข้อมูลให้อยู่ในรูปปรกติมาตรฐาน

ตัวแปร	ความหมาย	ตัวแปร	ความหมาย
X_1^*	เพศ	X_{10}^*	อุณหภูมิร่างกาย
X_2^*	อายุ	X_{11}^*	จำนวนวัคซีนที่ได้รับ
X_3^*	น้ำหนัก	X_{12}^*	การสูบบุหรี่
X_4^*	ส่วนสูง	X_{13}^*	การดื่มเครื่องดื่มที่มีแอลกอฮอล์
X_5^*	ค่าดัชนีมวลกาย	X_{14}^*	การเป็นโรคความดันโลหิตสูง
X_6^*	ค่าความดันโลหิตขณะหัวใจบีบตัว	X_{15}^*	การเป็นโรคเบาหวาน
X_7^*	ค่าความดันโลหิตขณะหัวใจคลายตัว	X_{16}^*	การเป็นโรคหลอดเลือดหัวใจ
X_8^*	ระดับออกซิเจนในเลือด	X_{17}^*	การเป็นโรคปอดบวม
X_9^*	อัตราการเต้นของหัวใจ	Y	ระดับความรุนแรงของอาการผู้ป่วย

3.3 การสร้างตัวแบบทำนาย

การสร้างตัวแบบที่ใช้ในการทำนายเพื่อเปรียบเทียบประสิทธิภาพการทำนายผลระดับความรุนแรงของอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 จากข้อมูลทั้งหมด 650 ชุด แบ่งข้อมูลออกเป็น 2 ชุด คือ

- 1) ข้อมูลชุดเรียนรู้ (Training Data Set) 80% จากข้อมูลทั้งหมด เพื่อนำไปสร้างตัวแบบทำนาย ซึ่งจะได้ข้อมูลชุดเรียนรู้ จำนวน 520 ชุด
- 2) ข้อมูลชุดทดสอบ (Testing Data Set) 20% จากข้อมูลทั้งหมด เพื่อเปรียบเทียบประสิทธิภาพของตัวแบบทำนาย ซึ่งจะได้ข้อมูลชุดทดสอบ จำนวน 130 ชุด

Ding et al. (2021) พบว่าโครงข่ายประสาทเทียมมีประสิทธิภาพสูงและ Rawat (2019) พบว่าการวิเคราะห์การถดถอยลอจิสติกทวิภาคีมีประสิทธิภาพสูงเมื่อทำการแบ่งข้อมูลเป็น 80:20 ดังนั้นในงานวิจัยนี้จึงใช้ข้อมูลชุดเรียนรู้ 80% และข้อมูลชุดทดสอบ 20%

3.3.1 การถดถอยลอจิสติกทวิภาค

ขั้นตอนที่ 1 กำหนดตัวแบบสมการการถดถอย ตัวแปรตามมีค่าเพียง 2 ค่าจะกำหนดเป็นแบบการถดถอยลอจิสติกทวิภาค (Binary Logistic Regression)

ขั้นตอนที่ 2 ตรวจสอบข้อตกลงเบื้องต้นของการวิเคราะห์การถดถอยลอจิสติกทวิภาค

ข้อตกลงเบื้องต้นข้อที่ 1 ตัวแปรตามเป็นการวิเคราะห์ลอจิสติกมี 2 ค่า คือ 0, 1 ในงานวิจัยนี้กำหนดค่า 0 คือ อาการของผู้ป่วยอยู่ในระดับไม่รุนแรง และ 1 คือ อาการของผู้ป่วยอยู่ในระดับรุนแรง

ข้อตกลงเบื้องต้นข้อที่ 2 ตัวแปรอิสระต้องเป็นอิสระต่อกันหรือไม่มีความสัมพันธ์กัน ไม่เกิดปัญหาความสัมพันธ์เชิงเส้นพหุ (Multicollinearity) โดยพิจารณาจากค่า Variance Inflation Factors (VIF) และค่าคลาดเคลื่อนยินยอม (Tolerance)

ค่า VIF เป็นค่าที่ใช้วัดตัวแปรอิสระแต่ละตัวซึ่งอธิบายด้วยตัวแปรอื่นว่าอยู่ในระดับใด ถ้าค่า VIF มาก แสดงว่าตัวแปรอิสระมีความสัมพันธ์กับตัวแปรอื่นมาก ซึ่งถ้าค่า VIF มากกว่า 10 ขึ้นไปจะถือว่าอยู่ในระดับที่ทำให้เกิดปัญหาความสัมพันธ์เชิงเส้นพหุได้

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้คัดลอกเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ค่า Tolerance เป็นค่าที่ใช้วัดตัวแปรอิสระแต่ละตัว ซึ่งถูกอธิบายโดยตัวแปรอิสระอื่นในระดับใด ซึ่งค่า Tolerance มีค่าอยู่ระหว่าง 0-1 ถ้าค่า Tolerance ของตัวแปรใดมีค่าเข้าใกล้ 0 แสดงว่าตัวแปรอิสระนั้นมีความสัมพันธ์กับตัวแปรอื่นมาก ดังนั้นค่า Tolerance ของตัวแปรใดมีค่าเข้าใกล้ 1 จะถือว่าดี

ข้อตกลงเบื้องต้นข้อที่ 3 ค่าความคลาดเคลื่อน (e) มีค่าเฉลี่ยเท่ากับศูนย์ เงื่อนไขข้อนี้เป็นจริงเนื่องจากเมื่อใช้วิธีกำลังสองน้อยที่สุดในการประมาณค่าสัมประสิทธิ์การถดถอยจะทำให้ $e = 0$ เสมอ

ข้อตกลงเบื้องต้นข้อที่ 4 การวิเคราะห์การถดถอยลอจิสติกทวิภาคจะต้องใช้ขนาดตัวอย่างเท่ากับ $n \geq 30$ โดยงานวิจัยนี้มีขนาดตัวอย่างเท่ากับ $n = 650$

ขั้นตอนที่ 3 คัดเลือกตัวแปรอิสระเข้าในตัวแบบ โดยใช้วิธีทีละขั้นแบบถอยหลัง: อัตราส่วนภาวะน่าจะเป็น (Backward Stepwise: Likelihood Ratio) นำตัวแปรทำนายทั้ง 17 ตัว ($X_1^*, X_2^*, X_3^*, \dots, X_{17}^*$) เข้าสมการพร้อมกัน จากนั้นพิจารณาตัวแปรทำนายที่อธิบายความผันแปรของตัวแปรตามได้น้อยที่สุดออกจากสมการ ทำเช่นนี้ไปเรื่อย ๆ จนได้ตัวแปรอิสระที่ให้ค่า AIC ต่ำที่สุด

ขั้นตอนที่ 4 ประมาณค่าสัมประสิทธิ์ การถดถอย β_i ด้วยวิธีภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Method) คำนวณหาอัตราส่วนออก (Odd Ratio) และสร้างสมการลอจิสติก (Logit)

ขั้นตอนที่ 5 ตรวจสอบความเหมาะสมของตัวแบบด้วยวิธีฮอสเมอร์-เลเมสโชว์ (Hosmer and Lemeshow) โดยที่ค่า $p\text{-value} > \alpha$ จะไม่สามารถปฏิเสธ H_0 นั่นคือตัวแบบมีความเหมาะสม

ขั้นตอนที่ 6 วัดระดับความสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระ โดยตรวจสอบความสอดคล้องหรือความผันแปรของตัวแบบทำนาย ตัวสถิติทดสอบที่ใช้คือ ค็อก-สเนลล์ ($Cox \& snell R_{CS}^2$) และนาเจลเคอร์คิ ($Nagelkerki R_N^2$) โดยที่ค่า R_{CS}^2 และ R_N^2 จะอยู่ระหว่าง 0 ถึง 1 โดยที่มีค่าความใกล้เคียง 1 แสดงถึงการเป็นตัวแบบที่เหมาะสมมาก และค่ามีความใกล้เคียง 0 แสดงถึงความไม่เหมาะสมของตัวแบบ

ขั้นตอนที่ 7 นำตัวแบบไปใช้ทำนายตัวแปรตาม

3.3.2 โครงข่ายประสาทเทียม

สำหรับงานวิจัยนี้ใช้โครงข่ายประสาทเทียมแบบแพร่ย้อนกลับ (Back Propagation Neural Networks) การออกแบบโครงสร้างของโครงข่ายประกอบด้วยจำนวนชั้นข้อมูลเข้า จำนวนโหนดในชั้นข้อมูลเข้า จำนวนชั้นซ่อน จำนวนโหนดในชั้นซ่อน จำนวนชั้นข้อมูลออก และจำนวนโหนดในชั้นข้อมูลออก ซึ่งสามารถกำหนดฟังก์ชันกระตุ้น อัตราการเรียนรู้ ค่าโมเมนตัม ตามความเหมาะสม และกำหนดจำนวนรอบการเรียนรู้ โดยงานวิจัยนี้มีการกำหนดค่าพารามิเตอร์ ดังนี้

1) จำนวนชั้นข้อมูลเข้า (Input Layer) มี 1 ชั้น

2) จำนวนโหนดในชั้นข้อมูลเข้า (Input Node) มีจำนวนเท่ากับปัจจัย (X_i) ที่ส่งผลกระทบต่อความรุนแรง (Y) ของอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 ซึ่งได้มาจากการคัดเลือกตัวแปรอิสระตามผลลัพธ์ที่ได้จากการสร้างตัวแบบทำนายการวิเคราะห์การถดถอยลอจิสติกทวิภาค ด้วยวิธีทีละขั้นแบบถอยหลัง

3) จำนวนชั้นซ่อน (Hidden Layer) มี 1 ชั้น

4) จำนวนโหนดในชั้นซ่อน (Hidden Node) คือ 1, 2, 3, 4, 5, 6, 7, 8, 9 และ 10

5) จำนวนชั้นข้อมูลออก (Output Layer) มี 1 ชั้น

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ไว้เพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น หากมีข้อสงสัยหรือต้องการข้อมูลเพิ่มเติม กรุณาติดต่อผู้จัดทำเอกสารทุกครั้งที่มีกรณำไปใช้

- 6) จำนวนโหนดในชั้นข้อมูลออก (Output Node) กำหนดเป็น 2 โหนด ได้แก่ ระดับความรุนแรงของอาการป่วย 2 ค่า คือ ไม่รุนแรงและรุนแรง
- 7) ฟังก์ชันกระตุ้น (Activation Function) ใช้ฟังก์ชันซิกมอยด์ (Sigmoid Function)
- 8) อัตราการเรียนรู้ คือ 0.05, 0.1, 0.2 และ 0.3
- 9) ค่าโมเมนตัม คือ 0.5, 0.7 และ 0.9
- 10) จำนวนรอบการเรียนรู้ 100 รอบ

3.4 การเปรียบเทียบประสิทธิภาพตัวแบบทำนาย

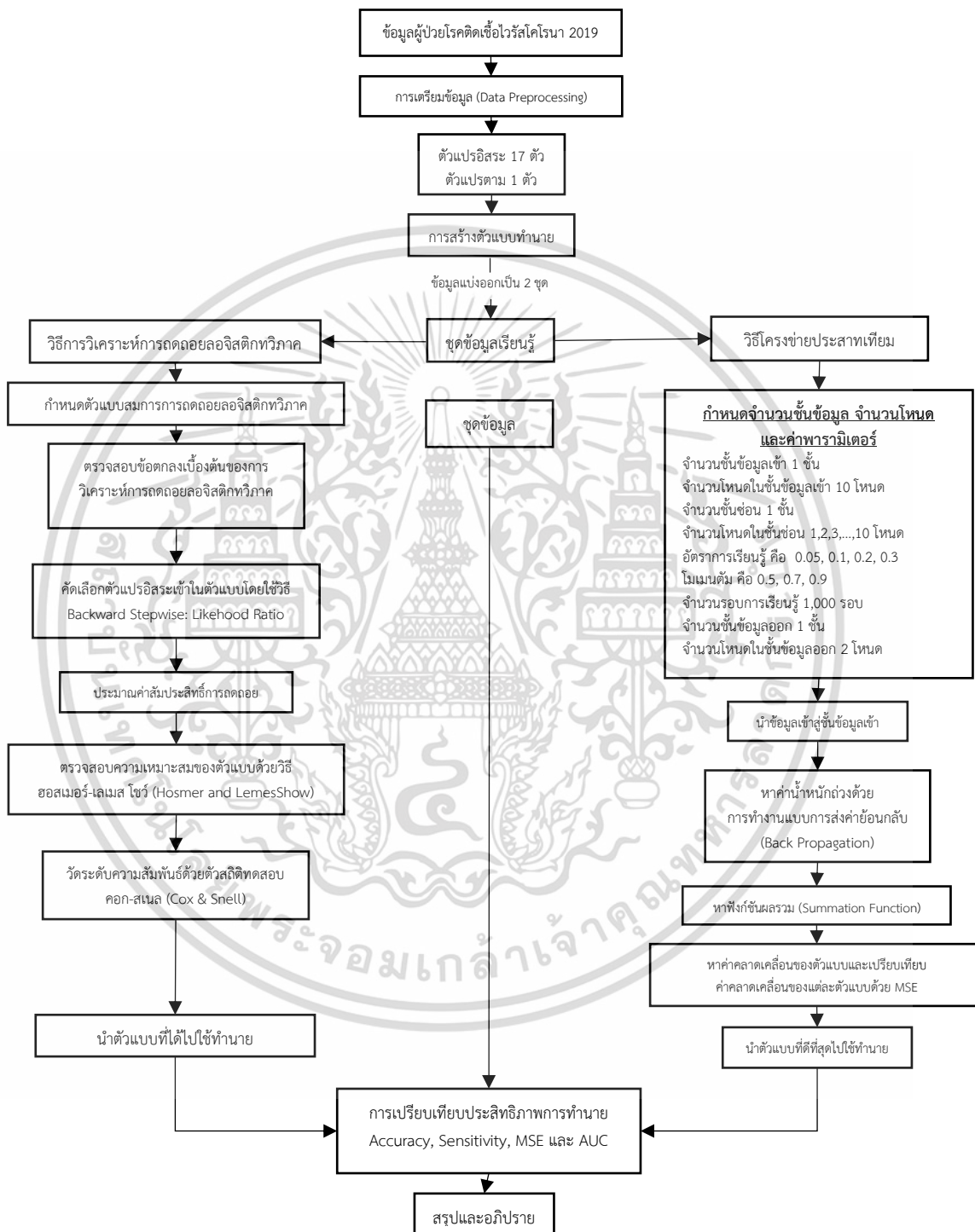
การประเมินประสิทธิภาพตัวแบบการทำนายของการถดถอยลอจิสติกทวิภาคและโครงข่ายประสาทเทียม ผู้วิจัยเลือกใช้เกณฑ์การประเมินประสิทธิภาพดังนี้

- 1) ค่าความแม่นยำ (Accuracy)
- 2) ค่าเรียกคืน (Recall) ในทางสถิติเรียกว่า ค่าความไว (Sensitivity)
- 3) ค่าวัดประสิทธิภาพโดยใช้พื้นที่ใต้กราฟ (Area Under a Curve: AUC)
- 4) ค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error: MSE)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.5 แผนการดำเนินงานวิจัย

กระบวนการดำเนินงานการเปรียบเทียบประสิทธิภาพในการทำนายผลระดับความรุนแรงของอาการผู้โรคติดเชื้อไวรัสโคโรนา 2019 สามารถสรุปได้เป็นแผนผังดังนี้



รูปที่ 3.1 แผนผังการดำเนินงานวิจัย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไมอนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4

ผลการวิจัยและอภิปรายผล

ในบทนี้จะนำเสนอผลการวิจัยการสร้างและพัฒนาตัวแบบทำนายระดับความรุนแรงของอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 โดยมีขั้นตอนดังนี้

1. ผลที่ได้จากการเตรียมข้อมูลเพื่อสร้างตัวแบบทำนาย
2. ผลการสร้างและพัฒนาตัวแบบทำนาย
3. ผลการเปรียบเทียบประสิทธิภาพของตัวแบบทำนาย

4.1 ผลที่ได้จากการเตรียมข้อมูลเพื่อสร้างตัวแบบทำนาย

จากข้อมูลผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 ที่เข้ารับการรักษาในกรมอนามัย กระทรวงสาธารณสุขแห่งหนึ่งในกรุงเทพมหานคร กระทรวงสาธารณสุข ตั้งแต่เดือนมีนาคม-ธันวาคม พ.ศ. 2565 จำนวน 652 ชุด ซึ่งในการเตรียมข้อมูล ผู้วิจัยได้ทำความสะอาดข้อมูล (Data Cleaning) พบว่ามีข้อมูลบางส่วนที่มีความซ้ำซ้อนและสูญหาย ทำให้ผู้วิจัยตัดข้อมูลเหล่านี้ออกไป 2 ชุด และเหลือข้อมูลที่ใช้ในการวิเคราะห์ทั้งสิ้น 650 ชุด ประกอบด้วยชุดข้อมูลผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 ในระดับรุนแรง 152 คน และไม่รุนแรง 498 คน ซึ่งมีตัวแปรอิสระที่ใช้ในการวิเคราะห์ระดับความรุนแรงของอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 มีจำนวน 17 ตัวแปร ได้แก่ เพศ อายุ น้ำหนัก ส่วนสูง ค่าดัชนีมวลกาย ค่าความดันโลหิตขณะหัวใจบีบตัว ค่าความดันโลหิตขณะหัวใจคลายตัว ระดับออกซิเจนในเลือด อัตราการเต้นของหัวใจ อุณหภูมิร่างกาย จำนวนวัคซีนที่ได้รับการสูบบุหรี่ การดื่มเครื่องดื่มที่มีแอลกอฮอล์ การเป็นโรคความดันโลหิตสูง การเป็นโรคเบาหวาน การเป็นโรคหลอดเลือดหัวใจ การเป็นโรคปอด และตัวแปรตามคือระดับความรุนแรงของอาการป่วยผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019

ผู้วิจัยเลือกใช้วิธีการทางสถิติ ได้แก่ การวิเคราะห์การถดถอยลอจิสติกทวิภาคและโครงข่ายประสาทเทียม ซึ่งได้มีการปรับแต่งโครงข่าย ค่าอัตราการเรียนรู้ และค่าโมเมนตัม เพื่อให้ได้ตัวแบบทำนายที่มีประสิทธิภาพที่ดีที่สุด

ในงานวิจัยนี้ผู้วิจัยต้องการเปรียบเทียบประสิทธิภาพของการวิเคราะห์การถดถอยลอจิสติกทวิภาคและโครงข่ายประสาทเทียม โดยแบ่งชุดข้อมูลเรียนรู้ 80 เปอร์เซ็นต์ และข้อมูลชุดทดสอบ 20 เปอร์เซ็นต์

4.2 ผลการสร้างและพัฒนาตัวแบบทำนาย

4.2.1 การวิเคราะห์การถดถอยลอจิสติกทวิภาค

ผู้วิจัยเริ่มสร้างตัวแบบทำนายในชุดข้อมูลเรียนรู้ 80 เปอร์เซ็นต์ ชุดข้อมูลทดสอบ 20 เปอร์เซ็นต์ โดยใช้การคัดเลือกตัวแปรอิสระด้วยวิธีที่ละชั้นแบบถอยหลัง: อัตราส่วนภาวะน่าจะเป็น (Backward Stepwise: Likelihood Ratio) ระดับนัยสำคัญเท่ากับ 0.05 ทำให้เลือกตัวแปรอิสระที่ส่งผลต่อตัวแปร Y ได้ดังตารางที่ 4.1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.1 ผลลัพธ์การตรวจสอบความสัมพันธ์ระหว่างตัวแปรอิสระด้วย วิธีทีละขั้นแบบถอยหลัง

ตัวแปรอิสระ	Tolerance	VIF
อายุ (X_2^*)	0.798	1.253
ค่าดัชนีมวลกาย (X_5^*)	0.935	1.069
ค่าความดันโลหิตขณะหัวใจบีบตัว (X_6^*)	0.485	2.062
ค่าความดันโลหิตขณะหัวใจคลายตัว (X_7^*)	0.576	1.736
การสูบบุหรี่ (X_{12}^*)	0.805	1.243
การดื่มเครื่องดื่มที่มีแอลกอฮอล์ (X_{13}^*)	0.860	1.163
การเป็นโรคความดันโลหิตสูง (X_{14}^*)	0.742	1.348
การเป็นโรคเบาหวาน (X_{15}^*)	0.761	1.313
การเป็นโรคหลอดเลือดหัวใจ (X_{16}^*)	0.808	1.237
การเป็นโรคปอด (X_{17}^*)	0.895	1.117

จากตารางที่ 4.1 จะพบว่าค่า VIF ของแปรอิสระทุกตัวมีค่าไม่เกิน 10 แสดงว่าตัวแปรอิสระไม่มีความสัมพันธ์กัน ส่วนค่า Tolerance ของตัวแปรอิสระทุกตัวก็มีค่าไม่ใกล้เคียง 0 แสดงว่าตัวแปรอิสระไม่มีความสัมพันธ์กัน

ตารางที่ 4.2 ผลการคัดเลือกตัวแปรอิสระด้วยวิธีทีละขั้นแบบถอยหลัง

ตัวแปรอิสระ	B	S.E.	Wald	OR	p-value
อายุ (X_2^*)	0.716	6.989	0.011	1.006	<0.001
ค่าดัชนีมวลกาย (X_5^*)	-0.662	11.104	0.004	0.981	<0.001
ค่าความดันโลหิตขณะหัวใจบีบตัว (X_6^*)	10.846	9.598	1.280	1.055	<0.001
ค่าความดันโลหิตขณะหัวใจคลายตัว (X_7^*)	-10.386	12.322	0.711	0.993	<0.001
การสูบบุหรี่ (X_{12}^*)	23.802	4356.736	0.000	1.006	<0.001
การดื่มเครื่องดื่มที่มีแอลกอฮอล์ (X_{13}^*)	-1.355	2.083	0.423	0.965	<0.001
การเป็นโรคความดันโลหิตสูง (X_{14}^*)	46.007	12911.631	0.000	1.091	<0.001
การเป็นโรคเบาหวาน (X_{15}^*)	47.944	12663.227	0.000	1.042	<0.001
การเป็นโรคหลอดเลือดหัวใจ (X_{16}^*)	44.584	14361.562	0.000	1.026	<0.001
การเป็นโรคปอด (X_{17}^*)	49.581	14667.604	0.000	1.003	<0.001
ค่าคงที่ (Constant)	-24.620	4356.724	0.000	0.009	<0.001

เอกสารนี้เป็นเอกสารลิขสิทธิ์ของมหาวิทยาลัยราชภัฏวไลยอลงกรณ์ ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 ไม่ว่าในรูปแบบใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การตรวจสอบความเหมาะสมของตัวแบบด้วยวิธีฮอสเมอร์-เลเมสโชว์ (Hosmer-Lemeshow)

สมมติฐานในการทดสอบคือ

H_0 : ตัวแบบเหมาะสม

H_1 : ตัวแบบไม่เหมาะสม

จากตารางที่ 4.2 จะพบว่า $p\text{-value} = 0.612 > \alpha$ ไม่สามารถปฏิเสธ H_0 นั่นคือตัวแบบมีความเหมาะสม สามารถวิเคราะห์ที่ได้ดังนี้ ตัวแปรอิสระที่ส่งผลกระทบต่อระดับความรุนแรงของอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 ได้แก่ อายุ (X_2^*) ค่าดัชนีมวลกาย (X_5^*) ค่าความดันโลหิตขณะหัวใจบีบตัว (X_6^*) ค่าความดันโลหิตขณะหัวใจคลายตัว (X_7^*) การสูบบุหรี่ (X_{12}^*) การดื่มเครื่องดื่มที่มีแอลกอฮอล์ (X_{13}^*) การเป็นโรคความดันโลหิตสูง (X_{14}^*) การเป็นโรคเบาหวาน (X_{15}^*) การเป็นโรคหลอดเลือดหัวใจ (X_{16}^*) และการเป็นโรคปอด (X_{17}^*) อย่างมีนัยสำคัญเชิงสถิติ 0.05 มีสมการดังนี้

$$\hat{Y} = -24.6200 + 0.7157X_2^* - 0.6616X_5^* + 10.8462X_6^* - 10.3860X_7^* + 23.8018X_{12}^* - 1.3553X_{13}^* + 46.0069X_{14}^* + 47.944X_{15}^* + 47.5841X_{16}^* + 49.5809X_{17}^*$$

โดยที่ X^* คือ ค่าของตัวแปรที่แปลงให้อยู่ให้รูปปรกติมาตรฐาน

จากค่าอัตราส่วนออดส์ (Odd Ratio) ซึ่งจะเห็นได้ว่าเมื่ออายุ (X_2^*) ค่าความดันโลหิตขณะหัวใจบีบตัว (X_6^*) การสูบบุหรี่ (X_{12}^*) การเป็นโรคความดันโลหิตสูง (X_{14}^*) การเป็นโรคเบาหวาน (X_{15}^*) การเป็นโรคหลอดเลือดหัวใจ (X_{16}^*) และการเป็นโรคปอด (X_{17}^*) เพิ่มขึ้น 1 หน่วย ทำให้โอกาสที่อาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 อยู่ในระดับรุนแรงเพิ่มขึ้นประมาณ 1 เท่าของโอกาสที่อาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 อยู่ในระดับไม่รุนแรง

ส่วนตัวแปรอิสระที่ไม่ส่งผลกระทบต่อระดับความรุนแรงของอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 คือเพศ (X_1^*) น้ำหนัก (X_3^*) ส่วนสูง (X_4^*) ระดับออกซิเจนในเลือด (X_8^*) อัตราการเต้นของหัวใจ (X_9^*) อุณหภูมิร่างกาย (X_{10}^*) และจำนวนวัคซีนที่ได้รับ (X_{11}^*) เมื่อพิจารณาความสามารถในการทำนายของตัวแบบทำนายนี้พบว่าตัวแปรอิสระทั้ง 10 ตัวสามารถอธิบายความแปรผันของระดับความรุนแรงของอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 ได้ร้อยละ 65.7 ส่วนที่เหลืออธิบายได้โดยตัวแปรอิสระอื่น ๆ อีกร้อยละ 34.3

ผลการสร้างตัวแบบทำนายจากการถดถอยลอจิสติกทวิภาคด้วย วิธีที่ละชั้นแบบถอยหลัง สำหรับผลลัพธ์ของตัวแบบทำนายระดับความรุนแรงของอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 จากการถดถอยลอจิสติกทวิภาคโดยใช้วิธีที่ละชั้นแบบถอยหลัง แสดงดังตาราง 4.3

ตารางที่ 4.3 ผลลัพธ์การทำนายระดับความรุนแรงของอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 จากการถดถอยลอจิสติกทวิภาคโดยใช้วิธีที่ละชั้นแบบถอยหลังของชุดข้อมูลเรียนรู้ 80 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การทำนาย (Predicted Class)		ค่าความไว (%)
	อาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 อยู่ในระดับรุนแรง	อาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 อยู่ในระดับไม่รุนแรง	
อาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 อยู่ในระดับรุนแรง	109	13	89.34
อาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 อยู่ในระดับไม่รุนแรง	24	374	93.97
ค่าความแม่นยำของตัวแบบทำนาย			92.89

จากตารางที่ 4.3 มีข้อมูล 520 คน ตัวแบบสามารถทำนายข้อมูลได้ถูกต้อง 483 คน โดยมีจำนวนข้อมูลที่จำแนกถูกกว่าอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 อยู่ในระดับรุนแรง 109 คน มีค่าความไวร้อยละ 89.34 และจำนวนข้อมูลที่จำแนกถูกกว่าอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 อยู่ในระดับไม่รุนแรง 374 คน มีค่าความไวร้อยละ 93.97 ตัวแบบทำนายข้อมูลไม่ถูกต้อง 37 คน โดยมีจำนวนข้อมูลที่จำแนกผิดว่าอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 อยู่ในระดับรุนแรง ซึ่งผลที่แท้จริงแล้วอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 อยู่ในระดับไม่รุนแรง 24 คน และจำนวนข้อมูลที่จำแนกผิดว่าอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 อยู่ในระดับไม่รุนแรง ซึ่งผลที่แท้จริงแล้วอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 อยู่ในระดับรุนแรง 13 คน มีค่าความแม่นยำร้อยละ 92.89

ตารางที่ 4.4 ผลลัพธ์การทำนายระดับความรุนแรงของอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 จากการถดถอยลอจิสติกทวิภาคโดยใช้วิธีที่ละชั้นแบบถอยหลังของชุดข้อมูลทดสอบ 20 เปอร์เซ็นต์

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การทำนาย (Predicted Class)		ค่าความไว (%)
	อาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 อยู่ในระดับรุนแรง	อาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 อยู่ในระดับไม่รุนแรง	
อาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 อยู่ในระดับรุนแรง	23	7	76.67
อาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 อยู่ในระดับไม่รุนแรง	8	92	92.00
ค่าความแม่นยำของตัวแบบทำนาย			88.46

เอกสารนี้เป็นทรัพย์สินทางปัญญาของสถาบันฯ ใช้เพื่อการเรียนการสอนเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่าในรูปแบบใดๆ ทั้งสิ้น อีกทั้งห้ามเผยแพร่ข้อมูลนี้ไปยังบุคคลอื่นโดยไม่ได้รับอนุญาตจากสถาบันฯ

จากตารางที่ 4.4 มีข้อมูล 130 คน ตัวแบบสามารถทำนายข้อมูลได้ถูกต้อง 115 คน โดยมีจำนวนข้อมูลที่จำแนกถูกกว่าอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 อยู่ในระดับรุนแรง 23 คน มีค่าความไวร้อยละ 76.67 และจำนวนข้อมูลที่จำแนกถูกกว่าอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 อยู่ในระดับไม่รุนแรง 92 คน มีค่าความไวร้อยละ 92.00 ตัวแบบทำนายข้อมูลไม่ถูกต้อง 15 คน โดยมีจำนวนข้อมูลที่จำแนกผิดว่าอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 อยู่ในระดับรุนแรง ซึ่งผลที่แท้จริงแล้วอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 อยู่ในระดับไม่รุนแรง 8 คน และจำนวนข้อมูลที่จำแนกผิดว่าอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 อยู่ในระดับไม่รุนแรง ซึ่งผลที่แท้จริงแล้วอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 อยู่ในระดับรุนแรง 7 คน มีค่าความแม่นยำร้อยละ 88.46

4.2.2 โครงข่ายประสาทเทียม

กำหนดค่าพารามิเตอร์ของโครงข่ายประสาทเทียมดังนี้

- 1) จำนวนชั้นข้อมูลเข้า 1 ชั้น
- 2) จำนวนโหนดในชั้นข้อมูลเข้า มี 10 โหนด ซึ่งได้มาจากการคัดเลือกตัวแปรอิสระตามผลลัพธ์จากการสร้างตัวแบบทำนายของการวิเคราะห์การถดถอยลอจิสติกทวิภาคจำนวน 10 ตัวแปรได้แก่

X_2^*	แทน	อายุ
X_5^*	แทน	ค่าดัชนีมวลกาย
X_6^*	แทน	ค่าความดันโลหิตขณะหัวใจบีบตัว
X_7^*	แทน	ค่าความดันโลหิตขณะหัวใจคลายตัว
X_{12}^*	แทน	การสูบบุหรี่
X_{13}^*	แทน	การดื่มเครื่องดื่มที่มีแอลกอฮอล์
X_{14}^*	แทน	การเป็นโรคความดันโลหิตสูง
X_{15}^*	แทน	การเป็นโรคเบาหวาน
X_{16}^*	แทน	การเป็นโรคหลอดเลือดหัวใจ
X_{17}^*	แทน	การเป็นโรคปอด

- 3) จำนวนชั้นซ่อน 1 ชั้น
- 4) จำนวนโหนดในชั้นซ่อน คือ 1, 2, 3, 4, 5, 6, 7, 8, 9 และ 10
- 5) จำนวนชั้นข้อมูลออก 1 ชั้น
- 6) จำนวนโหนดในชั้นข้อมูลออก 2 โหนดคืออาการของผู้ป่วยอยู่ในระดับรุนแรงและไม่รุนแรง
- 7) ฟังก์ชันกระตุ้นใช้ฟังก์ชันซิกมอยด์ (Sigmoid Function)
- 8) อัตราการเรียนรู้ คือ 0.05, 0.1, 0.2 และ 0.3
- 9) ค่าโมเมนตัม คือ 0.5, 0.7 และ 0.9
- 10) จำนวนรอบการเรียนรู้ 100 รอบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ผลการสร้างตัวแบบทำนายจากโครงข่ายประสาทเทียม

ตารางที่ 4.5 ค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวแบบทำนายที่มีจำนวนโหนดในชั้นซ่อน 1 โหนด

โมเมนตัม	อัตราการเรียนรู้			
	0.05	0.1	0.2	0.3
0.5	0.289	0.316	0.331	0.335
0.7	0.306	0.328	0.335	0.338
0.9	0.333	0.337	0.339	0.340

ตารางที่ 4.6 ค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวแบบทำนายที่มีจำนวนโหนดในชั้นซ่อน 2 โหนด

โมเมนตัม	อัตราการเรียนรู้			
	0.05	0.1	0.2	0.3
0.5	0.296	0.322	0.332	0.336
0.7	0.317	0.330	0.337	0.338
0.9	0.336	0.338	0.339	0.339

ตารางที่ 4.7 ค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวแบบทำนายที่มีจำนวนโหนดในชั้นซ่อน 3 โหนด

โมเมนตัม	อัตราการเรียนรู้			
	0.05	0.1	0.2	0.3
0.5	0.292	0.323	0.330	0.333
0.7	0.317	0.329	0.334	0.336
0.9	0.332	0.335	0.340	0.341

ตารางที่ 4.8 ค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวแบบทำนายที่มีจำนวนโหนดในชั้นซ่อน 4 โหนด

โมเมนตัม	อัตราการเรียนรู้			
	0.05	0.1	0.2	0.3
0.5	0.298	0.327	0.337	0.339
0.7	0.322	0.335	0.339	0.341
0.9	0.338	0.341	0.341	0.342

ตารางที่ 4.9 ค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวแบบทำนายที่มีจำนวนโหนดในชั้นซ่อน 5 โหนด

โมเมนตัม	อัตราการเรียนรู้			
	0.05	0.1	0.2	0.3
0.5	0.291	0.324	0.331	0.333
0.7	0.319	0.330	0.334	0.335
0.9	0.332	0.335	0.339	0.339

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับครูใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้คัดลอกเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.10 ค่าตลาดเคลื่อนกำลังสองเฉลี่ยของตัวแบบทำนายที่มีจำนวนโหนดในชั้นซ่อน 6 โหนด

โมเมนตัม	อัตราการเรียนรู้			
	0.05	0.1	0.2	0.3
0.5	0.306	0.320	0.329	0.332
0.7	0.319	0.328	0.333	0.335
0.9	0.331	0.335	0.338	0.339

ตารางที่ 4.11 ค่าตลาดเคลื่อนกำลังสองเฉลี่ยของตัวแบบทำนายที่มีจำนวนโหนดในชั้นซ่อน 7 โหนด

โมเมนตัม	อัตราการเรียนรู้			
	0.05	0.1	0.2	0.3
0.5	0.296	0.320	0.330	0.333
0.7	0.317	0.328	0.333	0.335
0.9	0.331	0.335	0.338	0.339

ตารางที่ 4.12 ค่าตลาดเคลื่อนกำลังสองเฉลี่ยของตัวแบบทำนายที่มีจำนวนโหนดในชั้นซ่อน 8 โหนด

โมเมนตัม	อัตราการเรียนรู้			
	0.05	0.1	0.2	0.3
0.5	0.286	0.321	0.329	0.331
0.7	0.315	0.328	0.332	0.333
0.9	0.331	0.333	0.337	0.340

ตารางที่ 4.13 ค่าตลาดเคลื่อนกำลังสองเฉลี่ยของตัวแบบทำนายที่มีจำนวนโหนดในชั้นซ่อน 9 โหนด

โมเมนตัม	อัตราการเรียนรู้			
	0.05	0.1	0.2	0.3
0.5	0.302	0.321	0.329	0.331
0.7	0.316	0.327	0.332	0.336
0.9	0.331	0.334	0.335	0.335

ตารางที่ 4.14 ค่าตลาดเคลื่อนกำลังสองเฉลี่ยของตัวแบบทำนายที่มีจำนวนโหนดในชั้นซ่อน 10 โหนด

โมเมนตัม	อัตราการเรียนรู้			
	0.05	0.1	0.2	0.3
0.5	0.302	0.324	0.332	0.334
0.7	0.319	0.331	0.334	0.335
0.9	0.333	0.335	0.339	0.340

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตารางที่ 4.5 – 4.14 แสดงค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวแบบการทำนายระดับความรุนแรงของอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 จากโครงข่ายประสาทเทียมโดยใช้วิธีการแพร่ย้อนกลับ พบว่าตัวแบบการทำนายที่มีค่าคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุดคือ ตัวแบบการทำนายที่มีพารามิเตอร์ดังนี้ จำนวนโหนดในชั้นซ่อนคือ 8 โหนด อัตราการเรียนรู้คือ 0.05 โมเมนตัมคือ 0.5 แสดงดังตารางที่ 4.12

ตารางที่ 4.15 ผลลัพธ์การทำนายระดับความรุนแรงของอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 จากโครงข่ายประสาทเทียมโดยใช้วิธีการแพร่ย้อนกลับของชุดข้อมูลเรียนรู้ 80 เปอร์เซ็นต์ และอัตราการเรียนรู้ 0.05

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การทำนาย (Predicted Class)		ค่าความไว (%)
	อาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 อยู่ในระดับรุนแรง	อาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 อยู่ในระดับไม่รุนแรง	
อาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 อยู่ในระดับรุนแรง	109	13	89.34
อาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 อยู่ในระดับไม่รุนแรง	12	386	96.99
ค่าความแม่นยำของตัวแบบทำนาย			95.19

จากตารางที่ 4.15 มีข้อมูล 520 คน ตัวแบบสามารถทำนายข้อมูลได้ถูกต้อง 495 คน โดยมีจำนวนข้อมูลที่จำแนกถูกกว่าอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 อยู่ในระดับรุนแรง 109 คน มีค่าความไวร้อยละ 89.34 และจำนวนข้อมูลที่จำแนกถูกกว่าอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 อยู่ในระดับไม่รุนแรง 386 คน มีค่าความไวร้อยละ 96.99 ตัวแบบทำนายข้อมูลไม่ถูกต้อง 25 คน โดยมีจำนวนข้อมูลที่จำแนกผิดว่าอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 อยู่ในระดับรุนแรง ซึ่งผลที่แท้จริงแล้วอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 อยู่ในระดับไม่รุนแรง 12 คน และจำนวนข้อมูลที่จำแนกผิดว่าอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 อยู่ในระดับไม่รุนแรง ซึ่งผลที่แท้จริงแล้วอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 อยู่ในระดับรุนแรง 13 คน มีค่าความแม่นยำร้อยละ 95.19

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.16 ผลลัพธ์การทำนายระดับความรุนแรงของอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 จากวิธีเทคนิคโครงข่ายประสาทเทียมวิธีการการจัดแบบถอยหลังของชุดข้อมูลทดสอบ 20 เปอร์เซนต์ และอัตราการเรียนรู้ 0.05

ผลลัพธ์จริง (Actual Class)	ผลลัพธ์การทำนาย (Predicted Class)		ค่าความไว (%)
	อาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 อยู่ในระดับรุนแรง	อาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 อยู่ในระดับไม่รุนแรง	
อาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 อยู่ในระดับรุนแรง	25	5	83.33
อาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 อยู่ในระดับไม่รุนแรง	2	98	98
ค่าความแม่นยำของตัวแบบทำนาย			94.62

จากตารางที่ 4.16 มีข้อมูล 130 คน ตัวแบบสามารถทำนายข้อมูลได้ถูกต้อง 123 คน โดยมีจำนวนข้อมูลที่จำแนกถูกกว่าอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 อยู่ในระดับรุนแรง 25 คน มีค่าความไวร้อยละ 83.33 และจำนวนข้อมูลที่จำแนกถูกกว่าอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 อยู่ในระดับไม่รุนแรง 98 คน มีค่าความไวร้อยละ 98 ตัวแบบทำนายข้อมูลไม่ถูกต้อง 7 คน โดยมีจำนวนข้อมูลที่จำแนกผิดว่าอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 อยู่ในระดับรุนแรง ซึ่งผลที่แท้จริงแล้วอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 อยู่ในระดับไม่รุนแรง 2 คน และจำนวนข้อมูลที่จำแนกผิดว่าอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 อยู่ในระดับไม่รุนแรง ซึ่งผลที่แท้จริงแล้วอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 อยู่ในระดับรุนแรง 5 คน มีค่าความแม่นยำร้อยละ 94.62

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.3 ผลการเปรียบเทียบประสิทธิภาพของตัวแบบทำนาย

การเปรียบเทียบประสิทธิภาพตัวแบบการทำนายระดับความรุนแรงของอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 ด้วยการถดถอยลอจิสติกทวิภาคและโครงข่ายประสาทเทียม

ตารางที่ 4.17 ค่าความแม่นยำ ค่าความไว ค่าคลาดเคลื่อนเฉลี่ยกำลังสอง และค่าวัดประสิทธิภาพโดยใช้พื้นที่ใต้กราฟของตัวแบบทำนายระดับความรุนแรงของอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 จากการถดถอยลอจิสติกทวิภาคและโครงข่ายประสาทเทียม

	ชุดข้อมูลทดสอบ			
	ความแม่นยำ (%)	ความไว (%)	MSE	AUC
การถดถอยลอจิสติกทวิภาค	88.46	76.67	0.312	0.707
โครงข่ายประสาทเทียม	94.62	83.33	0.286	0.817

จากตารางที่ 4.17 พบว่าตัวแบบทำนายระดับความรุนแรงของอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 ที่มีประสิทธิภาพดีที่สุดในชุดข้อมูลทดสอบคือ ตัวแบบทำนายด้วยโครงข่ายประสาทเทียม มีค่าความแม่นยำร้อยละ 94.62 ค่าความไวร้อยละ 83.33 ค่าคลาดเคลื่อนกำลังสองเฉลี่ยคือ 0.286 และค่าวัดประสิทธิภาพโดยใช้พื้นที่ใต้กราฟคือ 0.817 ซึ่งมีค่าอยู่ระหว่าง 0.75-0.9 แสดงว่าตัวแบบที่ใช้ในการทำนายมีประสิทธิภาพสูง

4.4 การอภิปรายผล

จากการวิจัยเรื่องการเปรียบเทียบประสิทธิภาพในการทำนายผลระดับความรุนแรงของอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 เพื่อหาวิธีที่ดีที่สุดที่เหมาะสมกับข้อมูลที่ใช้ในการวิจัย พบว่าโครงข่ายประสาทเทียมเหมาะสมในการสร้างตัวแบบทำนายผลระดับความรุนแรงของอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 ด้วยชุดข้อมูลทดสอบ 20 เปรอร์เซ็นต์ โดยมีค่าความแม่นยำร้อยละ 94.62 ค่าความไวร้อยละ 83.33 ค่าคลาดเคลื่อนกำลังสองเฉลี่ยคือ 0.286 และค่าวัดประสิทธิภาพโดยใช้พื้นที่ใต้กราฟคือ 0.817 ซึ่งสอดคล้องกับงานวิจัยของ Ding et al. (2021) พบว่า โครงข่ายประสาทเทียมสามารถระบุปัจจัยที่มีผลต่อการเสียชีวิตของผู้ป่วยโรคตับอ่อนอักเสบเฉียบพลันได้ทั้งหมด 12 ปัจจัยด้วยความแม่นยำร้อยละ 76.9 ความไวร้อยละ 66.6 และมีค่าวัดประสิทธิภาพโดยใช้พื้นที่ใต้กราฟคือ 0.769

ปัจจัยที่มีผลต่อระดับความรุนแรงของอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 ได้แก่ อายุ ค่าความดันโลหิตขณะหัวใจบีบตัว ค่าความดันโลหิตขณะหัวใจคลายตัว ค่าดัชนีมวลกาย การสูบบุหรี่ การดื่มเครื่องดื่มที่มีแอลกอฮอล์ การเป็นโรคความดันโลหิตสูง การเป็นโรคเบาหวาน การเป็นโรคหลอดเลือดหัวใจ และการเป็นโรคปอด ซึ่งมาจากการวิเคราะห์การถดถอยลอจิสติกทวิภาค โดยมีค่า Cox & Snell R Square เท่ากับ 0.657 ซึ่งสอดคล้องกับ Mohammadi et al. (2021) พบว่าลักษณะของผู้ป่วยที่ติดเชื้อ SARS-CoV-2 มีโรคร่วมที่พบบ่อย ได้แก่ โรคเบาหวาน โรคความดันโลหิตสูง โรคปอดอุดกั้นเรื้อรัง และโรคหลอดเลือดหัวใจ โดยที่การวิเคราะห์การถดถอยลอจิสติกทวิภาคมี

ค่า Cox & Snell R Square เท่ากับ 0.683 บ่งชี้ว่าแบบจำลองการถดถอยลอจิสติกทวิภาคมีความเหมาะสมกับข้อมูลและเชื่อถือได้

ไม่ว่าการันตีว่าสิ่งนี้ถูกต้องทั้งหมดมีแต่เพียงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5

สรุปผลการวิจัยและข้อเสนอแนะ

งานวิจัยนี้มีวัตถุประสงค์เพื่อการศึกษาเปรียบเทียบประสิทธิภาพในการทำนายผลระดับความรุนแรงของอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 ด้วยการวิเคราะห์การถดถอยลอจิสติกทวิภาคและโครงข่ายประสาทเทียม ซึ่งเป็นการศึกษาวิจัยเชิงสำรวจ รูปแบบการศึกษาย้อนหลัง เพื่อวิเคราะห์และเปรียบเทียบประสิทธิภาพในการทำนายปัจจัยที่มีผลต่อระดับความรุนแรงของผู้ป่วย โดยกลุ่มตัวอย่างในการศึกษาคือข้อมูลผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 ที่เข้ารับรักษาในกรมอนามัย กระทรวงสาธารณสุขแห่งหนึ่งในกรุงเทพมหานคร ตั้งแต่เดือนมีนาคม-ธันวาคม พ.ศ. 2565 จำนวน 652 ชุด ซึ่งในการเตรียมข้อมูลผู้วิจัยได้ทำความสะอาดข้อมูล พบว่ามีข้อมูลบางส่วนที่มีความซ้ำซ้อนและสูญหาย ทำให้ผู้วิจัยตัดข้อมูลเหล่านี้ออกไป 2 ชุดและเหลือข้อมูลที่ใช้ในการวิเคราะห์ทั้งสิ้น 650 ชุด แบ่งเป็นผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 ในระดับรุนแรง 152 คน และไม่รุนแรง 498 คน ซึ่งมีจำนวน 17 ตัวแปร ได้แก่ เพศ อายุ น้ำหนัก ส่วนสูง ค่าดัชนีมวลกาย ค่าความดันโลหิตขณะหัวใจบีบตัว ค่าความดันโลหิตขณะหัวใจคลายตัว ระดับออกซิเจนในเลือด อัตราการเต้นของหัวใจ อุณหภูมิร่างกาย จำนวนวัคซีนที่ได้รับ การสูบบุหรี่ การดื่มเครื่องดื่มที่มีแอลกอฮอล์ การเป็นโรคความดันโลหิตสูง การเป็นโรคเบาหวาน การเป็นโรคหลอดเลือดหัวใจ การเป็นโรคปอด และตัวแปรตามคือระดับความรุนแรงของอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 ในการวิเคราะห์ข้อมูลจะแบ่งข้อมูลออกเป็นชุดเรียนรู้ 80% และชุดทดสอบ 20% โดยการถดถอยลอจิสติกทวิภาคใช้การคัดเลือกตัวแปรอิสระด้วยวิธีที่ละเอียดแบบถดถอยหลัง: อัตราส่วนภาวะน่าจะเป็น จากนั้นตรวจสอบความเหมาะสมของตัวแบบด้วยวิธีฮอสเมอร์-เลเมสโซว์ ส่วนวิธีโครงข่ายประสาทเทียมจะใช้ตัวแปรที่ได้มาจากการคัดเลือกตัวแปรอิสระตามผลลัพธ์จากการสร้างตัวแบบทำนายของการวิเคราะห์การถดถอยลอจิสติกทวิภาค กำหนดค่าพารามิเตอร์ต่าง ๆ เพื่อหาตัวแบบการทำนายที่มีค่าคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุด จากนั้นทำการเปรียบเทียบประสิทธิภาพการทำนายโดยเครื่องมือที่ใช้เปรียบเทียบประสิทธิภาพคือค่าความแม่นยำ ค่าความไว ค่าวัดประสิทธิภาพโดยใช้พื้นที่ใต้กราฟ และความคลาดเคลื่อนกำลังสองเฉลี่ย

5.1 สรุปผลการวิจัย

จากผลการศึกษาเปรียบเทียบประสิทธิภาพในการทำนายผลระดับความรุนแรงของอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 ด้วยการวิเคราะห์การถดถอยลอจิสติกทวิภาคและโครงข่ายประสาทเทียมพบว่าตัวแบบทำนายระดับความรุนแรงของอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 ที่มีประสิทธิภาพที่ดีที่สุดคือข้อมูลชุดทดสอบ คือ ตัวแบบทำนายด้วยโครงข่ายประสาทเทียม มีค่าความแม่นยำร้อยละ 94.62 ค่าความไวร้อยละ 83.33 ค่าคลาดเคลื่อนกำลังสองเฉลี่ยคือ 0.286 และค่าวัดสัมประสิทธิภาพโดยใช้พื้นที่ใต้กราฟคือ 0.817 มีค่าอยู่ระหว่าง 0.75-0.9 แสดงว่าตัวแบบที่ใช้ในการทำนายมีประสิทธิภาพสูง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สถิติเชิงพรรณนาของข้อมูลผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 พบว่าส่วนใหญ่เป็นเพศหญิง 350 คน จากผู้ป่วยทั้งหมดมีอายุเฉลี่ย 39.32 ปี มีน้ำหนักเฉลี่ย 61.27 กิโลกรัม มีส่วนสูงเฉลี่ย 161.08 เซนติเมตร มีค่าดัชนีมวลกายเฉลี่ย 23.41 กิโลกรัม/ตารางเมตร มีค่าความดันโลหิตขณะหัวใจบีบตัวเฉลี่ย 112.86 มิลลิเมตรปรอท มีค่าความดันโลหิตขณะหัวใจคลายตัว 73.62 มิลลิเมตรปรอท ระดับออกซิเจนในเลือดเฉลี่ยร้อยละ 96.89 อัตราการเต้นของหัวใจเฉลี่ย 96.82 ครั้ง/นาที อุณหภูมิร่างกายเฉลี่ย 37.04 องศา จำนวนวัคซีนที่ได้รับเฉลี่ย 2.58 เข็ม และพบว่าส่วนใหญ่ไม่สูบบุหรี่ 547 คน ไม่ดื่มเครื่องดื่มที่มีแอลกอฮอล์ 518 คน ไม่เป็นโรคความดันโลหิต 587 คน ไม่เป็นโรคเบาหวาน 588 คน ไม่เป็นโรคหลอดเลือดหัวใจ 602 คน และไม่เป็นโรคปอดบวม 601 คน

การตรวจสอบความสัมพันธ์ระหว่างตัวแปรอิสระด้วยวิธีที่ละขั้นตอนถอยหลังของการวิเคราะห์การถดถอยลอจิสติกทวิภาคพบว่าตัวแปรอิสระแต่ละตัวไม่มีความสัมพันธ์กัน เนื่องจากค่า VIF ของตัวแปรอิสระทุกตัวมีค่าไม่เกิน 10 และค่า Tolerance ของตัวแปรอิสระทุกตัวก็มีค่าไม่ใกล้เคียง 0 ซึ่งได้ตัวแปรอิสระที่ส่งผลต่อระดับความรุนแรงของอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 มีทั้งหมด 10 ตัว ได้แก่ อายุ (X_2^*) ค่าดัชนีมวลกาย (X_5^*) ค่าความดันโลหิตขณะหัวใจบีบตัว (X_6^*) ค่าความดันโลหิตขณะหัวใจคลายตัว (X_7^*) การสูบบุหรี่ (X_{12}^*) การดื่มเครื่องดื่มที่มีแอลกอฮอล์ (X_{13}^*) การเป็นโรคความดันโลหิตสูง (X_{14}^*) การเป็นโรคเบาหวาน (X_{15}^*) การเป็นโรคหลอดเลือดหัวใจ (X_{16}^*) และการเป็นโรคปอด (X_{17}^*) อย่างมีนัยสำคัญทางสถิติ 0.05

เมื่ออายุ (X_2^*) ค่าความดันโลหิตขณะหัวใจบีบตัว (X_6^*) การสูบบุหรี่ (X_{12}^*) การเป็นโรคความดันโลหิตสูง (X_{14}^*) การเป็นโรคเบาหวาน (X_{15}^*) การเป็นโรคหลอดเลือดหัวใจ (X_{16}^*) และการเป็นโรคปอด (X_{17}^*) เพิ่มขึ้น 1 หน่วย ทำให้โอกาสที่อาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 อยู่ในระดับรุนแรงเพิ่มขึ้นประมาณ 1 เท่าของโอกาสที่อาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 อยู่ในระดับไม่รุนแรง

5.2 ข้อเสนอแนะ

5.2.1 การศึกษาครั้งต่อไปอาจประยุกต์ใช้กระบวนการเรียนรู้ของเครื่อง (Machine Learning) เพื่อเปรียบเทียบประสิทธิภาพในทำนายแบบอื่น ๆ เช่น Random Forest Classifier, SVM, Naive Bayes Classifier, Decision Tree Classifier, LightGBM และ XGBoost

5.2.2 การศึกษาครั้งต่อไปอาจแบ่งชุดข้อมูลเรียนรู้ และชุดข้อมูลทดสอบที่แตกต่างกัน เช่น ชุดข้อมูลเรียนรู้ 70 เปอร์เซ็นต์ และชุดข้อมูลทดสอบ 30 เปอร์เซ็นต์ เพื่อหาสัดส่วนของชุดข้อมูลที่เหมาะสมที่สุดในการสร้างตัวแบบการทำนาย

5.2.3 เนื่องจากข้อมูลที่ใช้ในการวิจัยนี้เป็นข้อมูลย้อนหลังของผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 ทำให้ตัวแบบที่ได้ อาจไม่มีประสิทธิภาพมากพอ ดังนั้นการหาข้อมูลของผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 ที่มีความเจาะจงและมีลักษณะเฉพาะต่อการเป็นโรคติดเชื้อไวรัสโคโรนา 2019 อาจส่งผลให้ตัวแบบทำนายมีประสิทธิภาพมากยิ่งขึ้น

5.2.4 ตัวแปรอิสระที่ส่งผลต่อระดับความรุนแรงของอาการผู้ป่วยอาจยังเก็บมาไม่ครบถ้วน เช่น ผู้ป่วยที่มีภาวะภูมิคุ้มกันบกพร่อง ผู้ป่วยที่ช้ำยากภูมิคุ้มกัน เป็นต้น

5.2.5 เนื่องจากการวิจัยนี้มีข้อมูลสูญหายและมีข้อมูลซ้ำซ้อนการศึกษาครั้งต่อไปจึงควรคัดเลือกข้อมูลที่มีคุณภาพมากขึ้น

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ของบัณฑิตวิทยาลัย มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี หากมีการนำเอกสารนี้ไปใช้โดยไม่ได้รับอนุญาตจากบัณฑิตวิทยาลัย มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี ถือว่าผิดกฎหมาย

เอกสารอ้างอิง

- กัลยา วานิชย์บัญชา. 2549. สถิติสำหรับงานวิจัย. พิมพ์ครั้งที่ 2. สำนักงานพิมพ์จุฬาลงกรณ์มหาวิทยาลัย.
- ไกรศักดิ์ เกษร. 2564. ศิลปะแห่งวิทยาการข้อมูลสำหรับผู้เริ่มต้นโดยใช้ Python และ Azure ML. วิทยาการคอมพิวเตอร์. พิษณุโลก: มหาวิทยาลัยนเรศวร.
- พิชญา ศักดิ์ศรีพาณิชย์. 2564. ปัจจัยที่มีผลต่อผลการรักษาภาวะเนื้อเยื่อปลูกถ่ายเอกพันธ์ของไต ทำหน้าที่ผิดปกติเรื้อรังโดยใช้แผนการรักษาด้วยยาเอเวโรลิมีสเป็นหลัก. เชียงใหม่: มหาวิทยาลัยเชียงใหม่.
- เพชรน้อย สิ่งช่างชัย. 2549. หลักการและการใช้สถิติการวิเคราะห์ตัวแปรหลายตัวสำหรับการวิจัยทางการแพทย์. พิมพ์ครั้งที่ 3. ชานเมืองการพิมพ์. สงขลา.
- ยุทธ ไถยวรรณ. 2555. หลักการและการใช้การวิเคราะห์การถดถอยโลจิสติกส์สำหรับการวิจัย. 4(1).
- รสริน โคตรเสนา และอัจฉราพรรณ ประทุมมณี. 2563. การพยากรณ์ความต้องการสินค้าโดยวิธีโครงข่ายประสาทเทียมของโรงงานผลิตขนมแห่งหนึ่ง. โครงการงาน (วิศวกรรมการจัดการและโลจิสติกส์). กรุงเทพฯ: มหาวิทยาลัยธุรกิจบัณฑิต.
- รัชนีวรรณ ไพศาลวรเกียรติ. 2564. การเปรียบเทียบตัวแบบการถดถอยลอจิสติกและเทคนิคเหมืองข้อมูลสำหรับการทำนายการเป็นโรคเบาหวาน. วิทยานิพนธ์ (สาขาวิชาสถิติ). พิษณุโลก: มหาวิทยาลัยนเรศวร.
- ศันสนีย์ เอื้อพันธ์วิริยะกุล. 2556. บทนำความเชิงคำนวณสำหรับวิศวกรรมคอมพิวเตอร์. เชียงใหม่: มหาวิทยาลัยเชียงใหม่.
- สายชล สีนสมบูรณ์ทอง. 2560. การทำเหมืองข้อมูล เล่ม 1: การค้นหาความรู้จากข้อมูล. พิมพ์ครั้งที่ 2 ฉบับปรับปรุง. กรุงเทพฯ: จามจุรีโปรดักส์ จำกัด.
- อกนิษฐ์ ทองจิตร, พูลพงศ์ สุขสว่าง และจตุภัทร เมฆพ่ายพ. 2562). การพัฒนาวิธีจำแนกประเภทข้อมูลโดยใช้โครงข่ายประสาทเทียมแบบปรับเหมาะผสมผสานการหาค่าเหมาะสมที่สุดแบบกลุ่มอนุภาคสำหรับการจำแนกประเภทกลุ่มเสี่ยงในการเป็นโรคเบาหวาน, 17(2), 83.
- อริญชัย บุญมีคำ, วินัย โพธิ์สุวรรณ และธิดาพร ศุภภากร. 2562. ตัวแบบการถดถอยลอจิสติกส์สำหรับการวิเคราะห์การขาดอายุของกรมธรรม์ประกันชีวิต. กรุงเทพฯ: มหาวิทยาลัยเกษตรศาสตร์.
- เอกพันธ์ บุญเสริม. 2563. การประยุกต์ใช้การเรียนรู้ของเครื่องในการทำนายความรุนแรงของผู้บาดเจ็บจากอุบัติเหตุทางถนนในช่วงเทศกาลปีใหม่จากข้อมูลเปิดภาครัฐของประเทศไทย. สารนิพนธ์ (วิทยาศาสตร์มหาบัณฑิต). กรุงเทพฯ: มหาวิทยาลัยศรีนครินทรวิโรฒ.
- Abdulaal, A., Patel, A., Charani, E., Denny, S., Mughal, N. and Moore, L. 2020. Prognostic Modeling of COVID-19 Using Artificial Intelligence in the United Kingdom: Model Development and Validation. 22(8).
- เอกสารนี้ Burns, N. and Grove, S. K. 1993. The Practice of Nursing Research: Conduct, Critique and Utilization. W.B. Saunders Company, Philadelphia. ไม่ว่ากรณีใดๆ ทั้งสิ้น ยกเว้นที่ผู้พิมพ์เห็นชอบและต้องขออนุญาตจากเจ้าของลิขสิทธิ์ที่มีการนำไปใช้

เอกสารอ้างอิง (ต่อ)

- Ding, N., Guo, C., Li, C., Zhou, Y. and Chai, X. 2021. **An Artificial Neural Networks Model for Early Predicting In-Hospital Mortality in Acute Pancreatitis in MIMIC-III.** 8.
- El-Rashidi N., Abdelrazik, S., Abuhmed, T., Amer, E., Ali, F., Hu, J. W. and Sappagh, S. 2021. **Comprehensive Survey of Using Machine Learning in the COVID19 Pandemic.** 11(7), 1155.
- Jayalakshmi, T. and Santhakumaran, A. 2011. **Statistical Normalization and Back Propagation for Classification.** International Journal of Computer Theory and Engineering, 1793-8201.
- Larose, D. T. and Larose, C. D. 2014. **Discovering Knowledge in Data: An Introduction to Data Mining.** 2nd ed. Wiley-Interscience.
- Mohammadi, F., Pourzamani, H., Karimi, H., Mohammadi M., Mohammadi, M., Ardalan, N., Khoshravesh, R., Pooresmaeil, H., Shahab, I S., Sabahi, M., Sadat Miryonesi F, Najafi, M., Yavari, Z., Mohammad, I F., Teir, I H. and Jannati M. 2021. **Artificial Neural Network and Logistic Regression Modelling to Characterize COVID-19 Infected Patients in Local Areas of Iran.** 44(3), 304-316.
- Rawat, S. 2019. **Heart Disease Prediction Cleveland Heart Disease (UCI Repository).** Towards Data Science.
- Ray, P., Le Manach, Y., Riou, B. and Houle, T. T. 2010. **Statistical Evaluation of a Biomarker.** *Anesthesiology.* 112(4), 1023-40.
- Stevens, J. 1996. **Applied Multivariate Statistics for Social Science.** Lawrence Erlbaum Associate, Inc., Mahwah, New Jersey.
- World Health Organization. 2020. **Coronavirus Disease (COVID-19).** Retrieved 19 Sep 2023, From <https://www.who.int/emergencies/diseases/novelcoronavirus2019/questionand-answer/hub/q-a-detail/coronavirus-disease-covid-19>.
- Zhou, F., Yu, T., Du, R., Fan, G., Liu, Y., Liu, Z., Xiang, J., Wang, Y., Song, B., Gu, X., Guan, L., Wei, Y., Li, H., Wu, X., Xu, J., Tu, S., Zhang, Y., Chen, H. and Cao, B. 2020. **Clinical Course and Risk Factors for Mortality of Adult in Patients with COVID-19 in Wuhan.** China: A Retrospective Cohort Study, 395(10229), 1054.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ก

.....

คำสั่งโปรแกรมอาร์ การถดถอยลอจิสติก
สำหรับการทำนายผลระดับความรุนแรงของอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019

.....

```
#install new package
install.packages("mlbench")
library(tidyverse)
install.packages("car")
library(carData)
library(broom)
install.packages("caret")
library(caret) # confusion matrix
library(mlbench)
#Read the Data
install.packages("dplyr")
library(MASS)
library(dplyr)
library(readr)
library(readxl)
install.packages("lmttest")
library(lmttest)
install.packages("blorr")
library(blorr)
train_d <- read_excel("C:/Theses/Datatraining.xlsx",
                      col_types = c("text", "numeric", "numeric",
                                     "numeric", "numeric", "numeric",
                                     "numeric", "numeric", "numeric",
                                     "numeric", "numeric", "numeric",
                                     "numeric", "numeric", "numeric"))
test_d <- read_excel("C:/Theses/Datatest.xlsx",
                     col_types = c("text", "numeric", "numeric",
                                    "numeric", "numeric", "numeric",
                                    "numeric", "numeric", "numeric",
                                    "numeric", "numeric", "numeric",
                                    "numeric", "numeric", "numeric"))
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงชื่อของเอกสารทุกครั้งที่มีการนำไปใช้

```

"numeric", "numeric", "numeric",
"numeric", "numeric", "numeric"))

#View(train_d)
nrow(train_d); nrow(test_d)
#-----Null-----#
Null<- glm(Y ~1, data = train_d , family = binomial("logit"))
summary(Null)
#-----Full Model-----#
fit1 <- glm(Y ~ ., data = train_d , family = binomial("logit"))
summary(fit1)
(cbind(OR=exp(coef(fit1)),exp(confint.default(fit1))))
vif(fit1)
blr_vif_tol(fit1)
blr_rsqa_cox_snell(fit1)
blr_rsqa_nagelkerke(fit1)
blr_test_hosmer_lemeshow(fit1, data = train_d)
library(blorr)
#-----เพิ่มกรณีตัวแปรอิสระ sign* จากตัวแปรอิสระทุกตัว-----#
S<-glm(Y ~
X1+X2+X3+X4+X5+X6+X7+X8+X9+X10+X11+X12+X13+X14+X15+X16+X17, data =
train_d , family = binomial("logit"))
summary(S)
blr_vif_tol(S)
(cbind(OR=exp(coef(S)),exp(confint.default(S))))
blr_rsqa_cox_snell(S)
blr_rsqa_nagelkerke(S)
#-----backward-----#
blr_step_p_backward(fit1,prem=0.05,details = FALSE) #เข้า12ตัวแปร
backward2.2<-glm(Y ~ X2+X6+X7+X5+X12+X13+X14+X15+X16+X17, data = train_d ,
family = binomial("logit"))
summary(backward2.2)
blr_vif_tol(backward2.2)
(cbind(OR=exp(coef(backward2.2)),exp(confint.default(backward2.2))))
blr_rsqa_cox_snell(backward2.2)
blr_rsqa_nagelkerke(backward2.2)
#-----ทดสอบความเหมาะสมของตัวแบบ-----#

```

#-----Likelihood Ratio Test & Chi-squared test-----#
 เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์หรือที่สงวนสิทธิ์ในบางประการ ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 #ถ้ายอมรับหรือ p value >0.05 แสดงว่า ตัวแบบนี้มีความเหมาะสม
 ไม่ว่าจะกรณีใดๆ ทั้งสิ้น ยกเว้นหากมีเหตุที่แบบสงวนเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

#ถ้ายอมรับหรือ p value <0.05 แสดงว่า ตัวแบบนั้นไม่เหมาะสม
anova(backward2.2, fit1, test = "Chisq")
AIC(backward2.2)
#-----การทดสอบความมีนัยสำคัญของสัมประสิทธิ์ถดถอย ----#
#ถ้ายอมรับหรือ p value >0.05 แสดงว่า ตัวแบบนั้นไม่ส่งผลต่อ Y#
#ถ้ายอมรับหรือ p value <0.05 แสดงว่า ตัวแบบนั้นส่งผลต่อ Y#
waldtest(backward2.2,fit1) #backward2.2 เป็นตัวแบบที่ไม่ส่งผลต่อ Y
#-----สถิติทดสอบระดับความสัมพันธ์ -----#
blr_rsqu_cox_snell(backward2.2)
#-----Accuracy-----#
#-----Full-----!
blr_confusion_matrix(fit1,data = train_d) #Accuracy :
blr_confusion_matrix(fit1,data = test_d) #Accuracy : 0.9439
#-----backward กรณี p-value <=0.05-----#
blr_confusion_matrix(backward2.2,data = train_d) #Accuracy : 0.8087
blr_confusion_matrix(backward2.2,data = test_d) #Accuracy : 0.7995

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ข

คำสั่งโปรแกรมอาร์ โครงข่ายประสาทเทียม

สำหรับการทำนายผลระดับความรุนแรงของอาการผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019

```

rm(list=ls())
# install new package
install.packages("mlbench")
library(tidyverse)
install.packages("car")
library(car)
library(broom)
install.packages("caret")
library(caret) # confusion matrix
library(mlbench)
# Read the Data
install.packages("dplyr")
library(MASS)
library(dplyr)
library(readr)
library(readxl)
Train_yes<- read_excel("C:/Theses/Train_Yes.xlsx")
Test_yes<- read_excel("C:/Theses/Test_Yes.xlsx")
Train_no<- read_excel("C:/Theses/Train_No.xlsx")
Test_no<- read_excel("C:/Theses/Test_No.xlsx")
## Split data จากการแยกข้อมูล แบ่งข้อมูลออกเป็น 2 ส่วน แต่ละส่วน 80% = train, 20%
=Test
set.seed(0)
train_d <- rbind(Train_yes,Train_no)
test_d<- rbind(Test_yes,Test_no)
nrow(train_d); nrow(test_d)
scaleddata<-scale(train_d)
str(train_d)
#-----Train-----#
normalize_train <- function(train_d) {
  return ((train_d - min(train_d)) / (max(train_d) - min(train_d)))
}

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 ไม่ว่ากรณีใดๆ ทั้งสิ้น หากมีข้อสงสัยหรือต้องการข้อมูลเพิ่มเติม กรุณาติดต่อฝ่ายวิชาการของมหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี โทร. 02-214-9434 หรือ อีเมล: info@kmutd.ac.th

```

}
Trainset <- as.data.frame(lapply(train_d, normalize_train))
#-----Test-----#
normalize_test <- function(test_d) {
  return ((test_d - min(test_d)) / (max(test_d) - min(test_d)))
}
Testset <- as.data.frame(lapply(test_d, normalize_test))
#-----ANN2-----#-----Full model-----#
# Train a neural network on the iris dataset
library(ANN2)
X_train <- Trainset[,1:17]
y_train <- Trainset$Y
X_test <- Testset[,1:17]
y_test <- Testset$Y
NN <- neuralnetwork(X_train, y_train, hidden.layers = 1,standardize = FALSE,
  loss.type = "log",
  huber.delta = 1,
  activ.functions = "sigmoid",sgd.momentum = 0.1,
  learn.rates = 0.1, val.prop = 0.5, n.epochs = 100,random.seed = 0)
#-----Train set-----#
# Make predictions
y_pred1 <- predict(NN, newdata = X_train)
correct1 <- (y_train == y_pred1$predictions)
table(y_train,y_pred1$predictions)
mean(y_train == y_pred1$predictions)
133
#-----Test set-----#
# Make predictions
y_pred2 <- predict(NN, newdata = X_test)
correct <- (y_test == y_pred2$predictions)
table(y_test,y_pred2$predictions)
mean(y_test == y_pred2$predictions)
#-----#
#Learning Rate 0.1 Momentum 0.05
#-----ANN2-----#----- Backward -----#
# Train a neural network on the iris dataset

```

เอกสารนี้เป็นทรัพย์สินทางปัญญาของมหาวิทยาลัยเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
 ไม่ควรเผยแพร่หรือใช้เพื่อวัตถุประสงค์อื่นโดยไม่ได้รับอนุญาตจากทางมหาวิทยาลัย

```

#เฉพาะตัวแปรอิสระ X2,X5,X6,X7,X12,X13,X14,X15,X16,X17
Z1 <- Trainset[,2]
Z2 <- Trainset[,5:7]
Z3 <- Trainset[,12:17]
X_train_B<-cbind(Z1,Z2,Z3)
#เฉพาะตัวแปรตาม y-----
y_train_B <- Trainset$Y
#เฉพาะตัวแปรอิสระ X2,X5,X6,X7,X12,X13,X14,X15,X16,X17
ZZ1 <- Testset[,2]
ZZ2 <- Testset[,5:7]
ZZ3 <- Testset[,12:17]
X_test_B<-cbind(ZZ1,ZZ2,ZZ3)
#เฉพาะตัวแปรตาม y-----
y_test_B <- Testset$Y
#-----สร้างตัวแบบ-----
NNB <- neuralnetwork(X_train_B, y_train_B, hidden.layers = 10,standardize = FALSE,
  loss.type = "log",
  huber.delta = 1,
  activ.functions = "sigmoid",
  learn.rates = 0.05
  ,sgd.momentum = 0.5
  ,val.prop = 0.5, n.epochs = 100,random.seed = 0)
# Plot the loss during training
#-----Train set-----#
# Make predictions
y_pred1B <- predict(NNB, newdata = X_train_B)
correct1B <- (y_train_B == y_pred1B$predictions)
table(y_train_B,y_pred1B$predictions)
mean(y_train_B == y_pred1B$predictions)
#-----Test set-----#
# Make predictions
y_pred2B <- predict(NNB, newdata = X_test_B)
correctB <- (y_test_B == y_pred2B$predictions)
table(y_test_B,y_pred2B$predictions)
mean(y_test_B == y_pred2B$predictions)
predictions <- predict(NNB, X_test_B)

```

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ไว้เพื่อใช้ในการเรียนการสอนเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 ไม่ว่ากรณีใดๆ ทั้งสิ้น ยกเว้น ห้า มิมีเหตุเปลี่ยนแปลงเนื้อหา และต้องยกย่องเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

# คำนวณค่า Mean Squared Error (MSE)
mse <- mean((predictions_numeric - y_test_B)^2)
print(paste("Mean Squared Error (MSE):", mse))
install.packages("pROC")
library(pROC)
roc_curve <- roc(X_test$Y, predict(predictions, newdata = X_test_B, type =
"response"))
auc_value <- auc(roc_curve)
print(auc_value)
# แสดงกราฟของ "ROC curve" พร้อมกับคำนวณค่า AUC
plot(roc_curve, main = "ROC Curve", col = "blue")
# คำนวณค่า AUC
auc_value <- auc(roc_curve)
print(auc_value)
#-----#

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



งานทะเบียนคณะวิทยาศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
คำรับรองเล่มปัญหาพิเศษ

วันที่ 18 เดือน เมษายน พ.ศ 2567

ข้าพเจ้า นางสาวฐานิตา เงามาม รหัสประจำตัว 63050618
นางสาวณิพัทร เพียงลิ้ม รหัสประจำตัว 63050621
นายรชต พรรณโรจน์ รหัสประจำตัว 63050655

นักศึกษาหลักสูตรวิทยาศาสตรบัณฑิต สาขาวิชา สถิติประยุกต์ ภาควิชา สถิติ

ขอรับรองว่าปัญหาพิเศษ เรื่อง

การเปรียบเทียบประสิทธิภาพในการทำนายผลระดับความรุนแรงของอาการผู้ป่วย

โรคติดเชื้อไวรัสโคโรนา 2019

EFFICIENCY COMPARISON OF SEVERE PREDICTION OF
CORONAVIRUS DISEASE 2019 PATIENTS

ปีการศึกษา 2566

เป็นผลงานวิจัยที่ได้คัดลอกหรือละเมิดลิขสิทธิ์ของผู้อื่นและได้ผ่านการตรวจสอบความซ้ำซ้อน

เรียบร้อยแล้ว และได้แนบเอกสารการตรวจสอบการลอกเลียนงานวรรณกรรมที่ตรวจสอบจากเล่ม

ปัญหาพิเศษฉบับสมบูรณ์แล้ว

โปรแกรมอักษราวิสุทธิ 2.46 %

ลงชื่อ ฐานิตา เงามาม

(นางสาวฐานิตา เงามาม)

นักศึกษา

ลงชื่อ ณิพัทร เพียงลิ้ม

(นางสาวณิพัทร เพียงลิ้ม)

นักศึกษา

ลงชื่อ รชต

(นายรชต พรรณโรจน์)

นักศึกษา

ข้าพเจ้า รองศาสตราจารย์สายชล สินสมบูรณ์ทอง อาจารย์ที่ปรึกษาปัญหาพิเศษ ได้ตรวจสอบปัญหาพิเศษของนักศึกษาข้างต้น แล้ว ขอรับรองว่าเป็นผลงานวิจัยของนักศึกษาจริงและมีเนื้อหาสมบูรณ์ จึงลงชื่อไว้เป็นหลักฐาน

ลงชื่อ สายชล สินสมบูรณ์ทอง

อาจารย์ที่ปรึกษา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้