

ปัจจัยที่สัมพันธ์ต่อการเสียชีวิตของผู้ป่วยโควิด-19

กรณีศึกษาโรงพยาบาลระยอง

FACTORS ASSOCIATED WITH DEATHS OF COVID-19 PATIENTS:
CASE STUDY OF RAYONG HOSPITAL



ภูริทัต เมฆวิไลพันธ์

PHOORITAT MEKWILAIPHAN

การค้นคว้าอิสระนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาสถิติและการวิเคราะห์ธุรกิจ
ภาควิชาสถิติ คณะวิทยาศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
พ.ศ. 2565

KMITL-2022-SC-M-050-137

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

FACTORS ASSOCIATED WITH DEATHS OF COVID-19 PATIENTS:
CASE STUDY OF RAYONG HOSPITAL



An INDEPENDENT STUDY SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENT FOR THE DEGREE OF MASTER OF SCIENCE IN STATISTICS AND
BUSINESS ANALYTICS

DEPARTMENT OF STATISTICS SCHOOL OF SCIENCE
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

2022

KMITL-2022-SC-M-050-137

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



COPYRIGHT 2022

SCHOOL OF SCIENCE

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อการค้นคว้าอิสระ	ปัจจัยที่สัมพันธ์ต่อการเสียชีวิตของผู้ป่วยโควิด-19 กรณีศึกษา โรงพยาบาลระยอง
ชื่อนักศึกษา	ภูริทัต เมฆวิไลพันธ์
รหัสประจำตัว	63605133
ปริญญา	วิทยาศาสตรมหาบัณฑิต (สถิติและการวิเคราะห์ธุรกิจ)
ภาควิชา	สถิติ
พ.ศ.	2565
อาจารย์ที่ปรึกษาการค้นคว้าอิสระ	ผู้ช่วยศาสตราจารย์ ดร.พรพิมล ชัยวุฒิศักดิ์

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาปัจจัยที่สัมพันธ์ต่อการเสียชีวิตของผู้ป่วยโควิด-19 กรณีศึกษาโรงพยาบาลระยอง และเปรียบเทียบประสิทธิภาพของวิธีจำแนกกลุ่ม 3 เทคนิคได้แก่ การวิเคราะห์การถดถอยลอจิสติก ต้นไม้ตัดสินใจ และโครงข่ายประสาทเทียมแบบหลายลำดับชั้น เพื่อทำนายการเสียชีวิตของผู้ป่วยโควิด-19 ผลการศึกษาพบว่า โครงข่ายประสาทเทียมแบบหลายลำดับชั้น โดยกำหนดให้จำนวนชั้นซ่อนเท่ากับ 3 พร้อมทั้งฟังก์ชันเรกติไฟด์ลีนีเยียมูนิต เป็นฟังก์ชันกระตุ้น และฟังก์ชันซิกมอยด์ เป็นฟังก์ชันกระตุ้นในชั้นข้อมูลออก เป็นวิธีที่ให้ประสิทธิภาพในการทำนายดีที่สุดด้วยค่าความแม่นยำเป็นร้อยละ 95 ค่าความเที่ยงคิดเป็นร้อยละ 95 ค่าเรียกค้นคิดเป็นร้อยละ 100 และค่าความถ่วงดุลคิดเป็นร้อยละ 95 โดยปัจจัยสำคัญที่เกี่ยวข้องต่อการเสียชีวิตของผู้ป่วยโควิด-19 ได้แก่ การใช้เครื่องช่วยหายใจแบบใส่ท่อช่วยหายใจ การป่วยเป็นโรคไต อายุ จำนวนวันนับตั้งแต่วันตรวจพบเชื้อจนถึงวันจำหน่ายการรักษา และการป่วยเป็นโรคหัวใจ

คำสำคัญ : ปัจจัยที่สัมพันธ์ต่อการเสียชีวิต โรคโควิด-19 การวิเคราะห์การถดถอยลอจิสติก ต้นไม้ตัดสินใจ โครงข่ายประสาทเทียมแบบหลายลำดับชั้น

Independent Study Title	Factors Associated with Deaths of COVID-19 Patients: Case Study of Rayong Hospital
Student Name	Phooritat Mekwilaiphan
Student ID	63605133
Degree	Master of Science (Statistics and Business Analytics)
Department	Statistics
Year	2022
Independent Study Advisor	Asst.Prof.Dr.Pornpimol Chaiwuttisak

Abstract

The purposes of this study were to investigate the factors related to deaths in COVID-19 patients in a case study of Rayong Hospital and compare the performance of three classification techniques: Logistic Regression, Decision Tree and Multilayer Perceptron Neural Network, to predict the mortality in COVID-19 cases. The results showed that the Multilayer Perceptron Neural Network model containing 3 hidden layers with Rectified Linear Unit function and an output layer with a sigmoid activation function provided outstanding potentiality, with 95% accuracy, 95% precision, 100% recall, and 95% F1-score. The key factors involved in the death of COVID-19 patients such as the use of Invasive mechanical ventilation, being diagnosed with kidney disease, age, the interval range in days from the date of confirmation case to the date of discharge and being diagnosed with heart disease.

Keywords: Factors associated with death, COVID-19, Logistic Regression, Decision Tree, Multilayer Neural Networks

กิตติกรรมประกาศ

การค้นคว้าอิสระฉบับนี้สำเร็จลุล่วงได้ เนื่องด้วยความกรุณาจากอาจารย์ที่ปรึกษาการค้นคว้าอิสระ ผู้ช่วยศาสตราจารย์ ดร.พรพิมล ชัยวุฒิศักดิ์ ที่ให้คำปรึกษา ให้คำแนะนำ ให้ความช่วยเหลือ และตรวจสอบแก้ไขข้อบกพร่องต่าง ๆ ตลอดจนติดตามผลการดำเนินงานในการทำการค้นคว้าอิสระนี้ ให้สำเร็จลุล่วงสมบูรณ์ ข้าพเจ้าขอกราบขอบพระคุณเป็นอย่างสูง มา ณ โอกาสนี้

ขอกราบขอบพระคุณ รองศาสตราจารย์ สายชล สินสมบุรณ์ทอง และผู้ช่วยศาสตราจารย์ ดร. สมศรี บัณฑิตวิไล อาจารย์กรรมการทั้งสองท่าน ที่กรุณาสละเวลาอันมีค่ายิ่ง มาเป็นกรรมการสอบ การค้นคว้าอิสระของข้าพเจ้าในครั้งนี้ ตลอดจนให้คำแนะนำ และข้อเสนอแนะที่เป็นประโยชน์ เพื่อนำไปปรับปรุงแก้ไขข้อบกพร่อง ทำให้การค้นคว้าอิสระฉบับนี้สมบูรณ์ยิ่งขึ้น

ขอขอบพระคุณโรงพยาบาลระยอง ที่ให้ความอนุเคราะห์ข้อมูลสำหรับการทำการค้นคว้าอิสระ ในครั้งนี้

สุดท้ายนี้ ขอกราบขอบพระคุณครอบครัวของข้าพเจ้า ที่คอยเป็นกำลังใจให้การสนับสนุน และ คำปรึกษา อีกทั้งช่วยเหลือข้าพเจ้าในทุกด้านมาโดยตลอด จนการค้นคว้าอิสระนี้สำเร็จลุล่วงไปได้ ด้วยดี

นายภูริทัต เมฆวิไลพันธ์

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ก
บทคัดย่อภาษาอังกฤษ.....	ข
กิตติกรรมประกาศ.....	ค
สารบัญ	ง
สารบัญตาราง	ช
สารบัญรูป.....	ฉ
บทที่ 1 บทนำ.....	1
1.1 ความเป็นมาและความสำคัญของปัญหา.....	1
1.2 วัตถุประสงค์ของงานวิจัย.....	3
1.3 ขอบเขตของงานวิจัย.....	3
1.4 ประโยชน์ที่คาดว่าจะได้รับ.....	3
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง.....	4
2.1 แนวคิดและทฤษฎีเกี่ยวกับโรคโควิด-19.....	4
2.1.1 ความหมายของไวรัสโคโรนา.....	4
2.1.2 รูปร่างและลักษณะของไวรัสโคโรนา.....	4
2.1.3 ประเภทของไวรัสโคโรนา.....	4
2.1.4 สายพันธุ์ของเชื้อ SARS-CoV-2.....	5
2.2 แนวคิดและทฤษฎีเกี่ยวกับความรุนแรงของโรคโควิด-19	6
2.2.1 ความรุนแรงของไวรัสโคโรนา	6
2.2.2 ลักษณะอาการของโรคโควิด-19	7
2.2.3 ความรุนแรงของโรคโควิด-19	7
2.2.4 อัตราการแพร่เชื้อและความรุนแรง.....	8
2.2.5 กลุ่มโรคที่เสี่ยงมีอาการรุนแรงในผู้ป่วยโควิด-19	8
2.3 แนวทางเวชปฏิบัติ การวินิจฉัย ดูแลรักษา และป้องกันการติดเชื้อในโรงพยาบาล..	9
2.4 การวิเคราะห์การถดถอยลอจิสติก.....	10
2.4.1 การเลือกตัวแปรอิสระด้วยวิธี Backward Stepwise.....	11
2.4.2 การตรวจสอบความเหมาะสมของตัวแบบการวิเคราะห์การถดถอยลอจิสติก	12
2.5 ต้นไม้ตัดสินใจ	14
2.6 โครงข่ายประสาทเทียมแบบหลายลำดับชั้น.....	15
2.6.1 ฟังก์ชันกระตุ้น.....	16

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

	หน้า
2.7 วิธีสุ่มเพิ่มจำนวนข้อมูล	17
2.8 ค่ามาตรฐานแฮต	17
2.9 การเปรียบเทียบประสิทธิภาพของตัวแบบ	17
2.10 งานวิจัยที่เกี่ยวข้อง	19
บทที่ 3 วิธีการดำเนินงานวิจัย	22
3.1 การเก็บรวบรวมข้อมูล.....	22
3.2 ตัวแปรที่ใช้ในการวิจัย.....	22
3.3 การเตรียมข้อมูล.....	25
3.3.1 การทำความสะอาดข้อมูล.....	25
3.3.2 การแปลงข้อมูล	25
3.3.3 การแปลงข้อมูลให้อยู่ในค่ามาตรฐานแฮต	27
3.3.4 การสุ่มเพิ่มจำนวนข้อมูล.....	28
3.3.5 การแบ่งข้อมูลเป็นชุดข้อมูลฝึกหัด และชุดข้อมูลทดสอบ.....	28
3.4 การวิเคราะห์การจำแนกกลุ่ม	28
3.4.1 วิธีการวิเคราะห์การถดถอยลอจิสติก	28
3.4.2 วิธีต้นไม้ตัดสินใจ.....	28
3.4.3 วิธีโครงข่ายประสาทเทียมแบบหลายลำดับชั้น	28
3.5 ค่าความสำคัญของตัวแปร.....	29
3.6 การเปรียบเทียบประสิทธิภาพของตัวแบบ	29
3.7 เครื่องมือที่ใช้ในการวิจัย.....	29
บทที่ 4 ผลการวิจัยและการอภิปรายผล.....	30
4.1 ลักษณะทั่วไปของกลุ่มตัวอย่าง	30
4.2 การตรวจสอบความเหมาะสมของตัวแบบด้วยวิธีการวิเคราะห์การถดถอยลอจิสติก	34
4.3 การวิเคราะห์ความสำคัญของตัวแปร	37
4.4 การทำนายการเสียชีวิตของผู้ป่วยผู้ป่วยโควิด-19	39
4.4.1 วิธีการวิเคราะห์การถดถอยลอจิสติก	39
4.4.2 วิธีต้นไม้ตัดสินใจ.....	41
4.4.3 วิธีโครงข่ายประสาทเทียมแบบหลายลำดับชั้น	42
4.5 ผลการเปรียบเทียบประสิทธิภาพในการทำนาย	44
4.6 อภิปรายผล	44

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

	หน้า
บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ	46
5.1 สรุปผลการวิจัย	46
5.2 ข้อเสนอแนะ.....	47
เอกสารอ้างอิง.....	48
ภาคผนวก	52
ประวัติผู้เขียน	60



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

ตารางที่	หน้า
2.1 เมตริกซ์ความสับสน	18
3.1 ตัวแปรอิสระปัจจัยส่วนบุคคล	23
3.2 ตัวแปรอิสระปัจจัยด้านสุขภาพ	23
3.3 ตัวแปรอิสระปัจจัยด้านแรกรับ	23
3.4 ตัวแปรอิสระปัจจัยด้านการรักษา	24
3.5 ตัวแปรตาม	24
3.6 การแปลงข้อมูลตัวแปรอิสระที่ใช้ในการวิจัย	25
3.7 การแปลงข้อมูลตัวแปรตามที่ใช้ในการวิจัย	27
4.1 ตัวแปรอิสระปัจจัยส่วนบุคคล	30
4.2 ตัวแปรอิสระปัจจัยด้านสุขภาพ	31
4.3 ตัวแปรอิสระปัจจัยด้านแรกรับ	32
4.4 ตัวแปรอิสระปัจจัยด้านการรักษา	32
4.5 ตัวแปรตาม	33
4.6 ผลการสุ่มเพิ่มจำนวนข้อมูลด้วยเทคนิค SMOTE	33
4.7 ผลการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรอิสระที่เป็นข้อมูลเชิงปริมาณ	35
4.8 ผลการวิเคราะห์ตัวแบบการถดถอยลอจิสติกที่มีการคัดเลือกตัวแปรแบบ Backward	36
4.9 ค่าความเป็นไปได้ และค่าสถิติทดสอบระดับความสัมพันธ์	36
4.10 ผลการวิเคราะห์ความสำคัญของตัวแปรอิสระ	37
4.11 เมตริกซ์ความสับสนจากการทำนายการเสียชีวิตของผู้ป่วยผู้ป่วยโควิด-19 ด้วยวิธีการวิเคราะห์การถดถอยลอจิสติก	39
4.12 ค่าประสิทธิภาพของตัวแบบจำแนกรายละเอียดในแต่ละคำตอบจากการทำนายการเสียชีวิตของผู้ป่วยผู้ป่วยโควิด-19 ด้วยวิธีการวิเคราะห์การถดถอยลอจิสติก	39
4.13 เมตริกซ์ความสับสนจากการทำนายการเสียชีวิตของผู้ป่วยผู้ป่วยโควิด-19 ด้วยวิธีต้นไม้ตัดสินใจ	41
4.14 ค่าประสิทธิภาพของตัวแบบจำแนกรายละเอียดในแต่ละคำตอบจากการทำนายการเสียชีวิตของผู้ป่วยผู้ป่วยโควิด-19 ด้วยวิธีต้นไม้ตัดสินใจ	41
4.15 เมตริกซ์ความสับสนจากการทำนายการเสียชีวิตของผู้ป่วยผู้ป่วยโควิด-19 ด้วยวิธีโครงข่ายประสาทเทียมแบบหลายลำดับชั้น	42

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง (ต่อ)

ตารางที่	หน้า
4.16 ค่าประสิทธิภาพของตัวแบบจำแนกรายละเอียดในแต่ละคำตอบจากการทำนายการเสียชีวิตของผู้ป่วยผู้ป่วยโควิด-19 ด้วยวิธีโครงข่ายประสาทเทียมแบบหลายลำดับชั้น.....	43
4.17 ผลการเปรียบเทียบประสิทธิภาพในการทำนายของวิธีการจำแนกกลุ่มทั้ง 3 วิธี.....	44



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป

รูปที่	หน้า
4.1 AUC ของวิธีการวิเคราะห์การถดถอยลอจิสติก.....	40
4.2 AUC ของวิธีต้นไม้ตัดสินใจ.....	42
4.3 AUC ของวิธีโครงข่ายประสาทเทียมแบบหลายลำดับชั้น.....	43



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

การระบาดของโรคติดเชื้อไวรัสโคโรนา 2019 (COVID-19) ได้เริ่มต้นขึ้นในเดือนธันวาคม 2562 โดยพบครั้งแรกในนครอู่ฮั่น เมืองหลวงของมณฑลหูเป่ย์ สาธารณรัฐประชาชนจีน ซึ่งสาเหตุที่เป็นไปได้มากที่สุดในการติดต่อสู่คนคือการสัมผัสกับเนื้อสัตว์ประเภทต่าง ๆ ที่วางขายในตลาด หลังจากพบการระบาดของเชื้อไวรัสสายพันธุ์ใหม่ในเมืองอู่ฮั่นประเทศจีน องค์การอนามัยโลก (The World Health Organization: WHO) ได้ออกมาระบุว่า ไวรัสชนิดดังกล่าว คือ SARS-CoV-2 เรียกว่า Corona Virus Disease 2019 (COVID-19) ตามการประกาศชื่ออย่างเป็นทางการที่ใช้เรียกโรคทางเดินหายใจที่เกิดจากไวรัสโคโรนาสายพันธุ์ใหม่ 2019 และพบการแพร่เชื้อจากคนสู่คนผ่านละอองฝอยขนาดเล็ก (Aerosol) องค์การอนามัยโลกได้ประกาศให้การระบาดนี้เป็นภาวะฉุกเฉินทางสาธารณสุขระหว่างประเทศ (Public Health Emergency of International Concern)

วันที่ 11 มีนาคม 2563 องค์การอนามัยโลกประกาศโรคติดเชื้อไวรัสโคโรนา 2019 เป็นการระบาดใหญ่ (Coronavirus Pandemic) มีจำนวนผู้ป่วยเพิ่มขึ้นนอกเหนือจากประเทศจีนเป็นจำนวนมากอย่างรวดเร็ว พบผู้ป่วยรายใหม่ทั่วโลกเพิ่มขึ้นวันละประมาณ 3,000 - 4,000 ราย มีผู้เสียชีวิตเพิ่มขึ้นวันละประมาณ 200 - 300 ราย (กรมควบคุมโรค, 2564ก)

ประเทศไทยพบผู้ติดเชื้อรายแรกที่เข้ารับการรักษาเมื่อวันที่ 13 มกราคม 2563 เป็นชาวจีนที่ติดเชื้อจากประเทศจีน และได้เดินทางมาประเทศไทย ส่วนผู้ป่วยที่ติดเชื้อในประเทศไทยรายแรกมีการรายงานเมื่อ 31 มกราคม 2563 (คณะแพทยศาสตร์โรงพยาบาลรามาธิบดี มหาวิทยาลัยมหิดล, 2563) วันที่ 1 มีนาคม 2563 กระทรวงสาธารณสุข รายงานพบผู้เสียชีวิตรายแรกของประเทศ เป็นชายไทย อายุ 35 ปี ซึ่งเสียชีวิตด้วยกลุ่มอาการทำหน้าที่ผิดปกติของหลายอวัยวะ (กรมควบคุมโรค, 2563)

การระบาดในระลอกแรกนั้นอยู่ในช่วงเดือนมีนาคม - เมษายน 2563 และสามารถควบคุมโรคได้ภายใน 2 เดือน และเกิดการระบาดระลอกใหม่ในประเทศไทย เป็นการระบาดของเชื้อสายพันธุ์ใหม่คือสายพันธุ์ G614 ซึ่งเป็นสายพันธุ์ที่พบในสหภาพพม่า เมื่อเดือนธันวาคม 2563 ตรวจพบคนไทยติดเชื้อรายแรก และแพร่เชื้อจากตลาดกลางกุ้ง จังหวัดสมุทรสาคร (กรมควบคุมโรค, 2564ข)

สำหรับจังหวัดระยอง เมื่อวันที่ 19 มีนาคม 2563 พบผู้ป่วยติดเชื้อไวรัสโคโรนา รายแรกมีประวัติเดินทางไปทำงานที่กรุงเทพมหานคร (คณะกรรมการโรคติดต่อจังหวัดระยอง, 2563) วันที่ 28 ธันวาคม 2563 พบผู้ป่วยโควิด-19 เสียชีวิตรายแรก เป็นชายอายุ 45 ปี มีโรคประจำตัวคือ โรคหัวใจขาดเลือด โรคเบาหวานไม่ได้ทานยา (สาริต, 2563)

การระบาดของไวรัสโคโรนากลายพันธุ์ โดยมีรหัสที่เรียกว่า โคโรนสายพันธุ์ B.1.1.529 หรือ โอไมครอน (Omicron) ถูกรายงานว่าพบครั้งแรก ในตอนใต้ของทวีปแอฟริกาใต้ ที่ประเทศ บอตสวานา (โรงพยาบาลลิชเวซ อินเตอร์เนชั่นแนล หนองแขม, 2564) จากงานวิจัยพบว่า เชื้อไวรัสโคโรนา-19 สายพันธุ์โอไมครอน มีการกลายพันธุ์ถึง 50 ตำแหน่ง และมีการกลายพันธุ์ บนโปรตีนหนาม (Spike Protein) จำนวน 32 ตำแหน่ง ซึ่งเป็นเหตุให้ไวรัสสามารถแพร่กระจาย และติดต่อกันได้เร็วขึ้น และยังทำให้ตัวไวรัสมีความสามารถที่จะหลบหลีกภูมิคุ้มกันได้อีกด้วย (โรงพยาบาลพริ้นซ์ สุวรรณภูมิ, 2564) โดยส่วนใหญ่ของผู้ติดเชื้อสายพันธุ์โอไมครอนจะไม่แสดง อาการ หรือแสดงอาการเพียงเล็กน้อย เมื่อเปรียบเทียบกับอาการของผู้ติดเชื้อสายพันธุ์เดลต้าพบว่า ความรุนแรงของเชื้อสายพันธุ์โอไมครอนจะมีความรุนแรงน้อยกว่า แต่เนื่องจากเชื้อไวรัสโคโรนา 2019 สายพันธุ์โอไมครอน สามารถแพร่กระจายได้อย่างรวดเร็ว ก็อาจก่อให้เกิดความล้มเหลวทางด้าน สาธารณสุขได้เช่นกัน (กรมควบคุมโรค, 2564ค) เมื่อวันที่ 6 มกราคม 2565 กระทรวงสาธารณสุขได้ ประกาศยกระดับเตือนภัยโควิด-19 จากเดิมระดับ 3 เป็นระดับ 4 หลังจากที่จำนวนของผู้ติดเชื้อ เพิ่มขึ้นอย่างมีนัยสำคัญ และมีการแพร่ระบาดของเชื้อโควิด-19 สายพันธุ์โอไมครอนในหลาย ๆ จังหวัด จากการสถานการณ์การระบาดของโรคติดเชื้อไวรัสโคโรนา 2019 ในประเทศ คาดการณ์ว่าจะ พบจำนวนผู้ติดเชื้อเพิ่มสูงขึ้นเรื่อย ๆ แต่จำนวนผู้เสียชีวิตอาจเพิ่มสูงขึ้นเล็กน้อยหรือคงที่ และจะมีการแพร่ระบาดของเชื้อโควิด-19 สายพันธุ์โอไมครอนอย่างต่อเนื่อง และอาจครอบคลุมทุกจังหวัดของ ประเทศ และอาจประสบกับปัญหาเดียวกันกับการระบาดของเชื้อในระลอกเดือนเมษายนที่ผ่านมา (กรมการแพทย์, 2565)

จากการทบทวนวรรณกรรมที่เกี่ยวข้องกับการสร้างตัวแบบในการทำนายโอกาสการเสียชีวิตของผู้ป่วยโควิด-19 พบว่า การศึกษาของ Das. et al. (2020), Sánchez-Montañés. et al. (2021), Ferreira. et al. (2021), Muhammad. et al. (2020), Li et al. (2020), Agbelusi and Olayemi (2020), Khan. et al. (2021) และ Pourhomayoun and Shakibi (2021) ทำการสร้างตัวแบบ ในการทำนายโอกาสการเสียชีวิตของผู้ป่วยโควิด-19 เปรียบเทียบโดยใช้วิธีการเรียนรู้ของเครื่อง และ วิธีโครงข่ายประสาทเทียมแบบหลายลำดับชั้น ในการศึกษาของ Das. et al. (2020) และ Sánchez-Montañés. et al. (2021) ผลการศึกษาพบว่าวิธีการวิเคราะห์การถดถอยลอจิสติก มีประสิทธิภาพดี ที่สุด เปรียบเทียบประสิทธิภาพด้วยค่าความแม่นยำในการศึกษาของ Das. et al. (2020) มีค่าความแม่นยำ สูงที่สุดเท่ากับ 0.96 และเปรียบเทียบประสิทธิภาพด้วยค่าพื้นที่ใต้โค้งในการศึกษาของ Sánchez-Montañés. et al. (2021) มีค่าพื้นที่ใต้โค้งมากที่สุดเท่ากับ 0.89 ในการศึกษาของ Ferreira. et al. (2021) และ Muhammad. et al. (2020) ผลการศึกษาพบว่าวิธีต้นไม้ตัดสินใจ มีประสิทธิภาพดีที่สุด เปรียบเทียบประสิทธิภาพด้วยค่าความแม่นยำ การศึกษาของ Ferreira. et al. (2021) และ Muhammad. et al. (2020) มีค่าความแม่นยำเท่ากับ 0.90 และ 0.95 ตามลำดับ และในการศึกษาของ Li et al. (2020), Agbelusi and Olayemi (2020), Khan. et al. (2021) และ Pourhomayoun and Shakibi (2021) ผลการศึกษาพบว่าวิธีโครงข่ายประสาทเทียมแบบหลายลำดับชั้นมีประสิทธิภาพ เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ในทางอื่นไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ดีที่สุด โดยเปรียบเทียบประสิทธิภาพด้วยค่าความแม่นยำ การศึกษาของ Li et al. (2020), Agbelusi and Olayemi (2020), Khan. et al. (2021) และ Pourhomayoun and Shakibi (2021) มีค่าความแม่นยำเท่ากับ 0.85, 0.85, 0.97 และ 0.90 ตามลำดับ จากการทบทวนวรรณกรรมที่เกี่ยวข้อง ผู้วิจัยจึงดำเนินการสร้างตัวแบบในการทำนายโอกาสการเสียชีวิตของผู้ป่วยโควิด-19 ด้วยวิธีการวิเคราะห์การถดถอยลอจิสติก วิธีต้นไม้ตัดสินใจ และวิธีโครงข่ายประสาทเทียมแบบหลายลำดับชั้น

ด้วยสถานการณ์ปัจจุบันจำนวนผู้ป่วยที่เพิ่มสูงขึ้น ประกอบกับแนวทางการรักษาโรคโควิด-19 มีการปรับเปลี่ยนอยู่ตลอดเวลา และจำนวนเตียงที่มีอยู่อย่างจำกัด กระบวนการคัดกรองผู้ป่วยเข้ารับการรักษาในโรงพยาบาลจึงเป็นสิ่งสำคัญ ความรู้ หรือข้อมูลต่าง ๆ ที่ใช้ในการอ้างอิงส่วนใหญ่ จะเป็นการศึกษากลุ่มตัวอย่างในต่างประเทศซึ่งอาจไม่เพียงพอ ดังนั้นผู้วิจัยเห็นความสำคัญในการศึกษาปัจจัยที่มีความสัมพันธ์ต่อการเสียชีวิตของผู้ป่วยโควิด-19 ในโรงพยาบาลระยอง โดยศึกษาปัจจัยส่วนบุคคล ปัจจัยด้านสุขภาพ ปัจจัยอาการแรกเริ่ม และปัจจัยด้านการรักษา เพื่อนำผลการทำนายนำมาใช้เป็นแนวทางในการเฝ้าระวังผู้ป่วยที่มีความเสี่ยงสูงในการเสียชีวิตและทำการรักษาอย่างมีประสิทธิภาพมากขึ้น

1.2 วัตถุประสงค์ของงานวิจัย

- 1) เพื่อศึกษาปัจจัยที่สัมพันธ์กับการเสียชีวิตของผู้ป่วยโควิด-19
- 2) เพื่อวิเคราะห์ตัวแบบที่เหมาะสมในการทำนายโอกาสเสียชีวิตของผู้ป่วยโควิด-19

1.3 ขอบเขตของงานวิจัย

งานวิจัยนี้เป็นการศึกษาแบบย้อนหลัง (Retrospective) ในผู้ป่วยทุกคนที่ได้รับการวินิจฉัยร่วมว่าเป็นโรคโควิด-19 หรือที่มีการให้รหัสโรค U071 (COVID-19, Virus Identified (Lab Confirmed Case)) เป็นการวินิจฉัยร่วม (Comorbidity) และได้รับการเก็บสารคัดหลั่งด้วยวิธีสวอปโพรงจมูก (Nasopharyngeal Swabs) และสวอปคอ (Throat Swabs) เพื่อตรวจยืนยันว่ามีการติดเชื้อ SARS-CoV-2 ด้วยวิธีการตรวจแบบ Real-Time Reverse Transcription Polymerase Chain Reaction (RT-PCR) ทั้งผู้ป่วยที่เป็นโรคโควิดครั้งแรก และผู้ป่วยที่กลับมาเป็นซ้ำ ที่เข้ารับการรักษาในโรงพยาบาลระยอง และจำหน่ายการรักษาออกจากโรงพยาบาลระหว่างวันที่ 1 เมษายน 2564 ถึงวันที่ 31 มกราคม 2565

1.4 ประโยชน์ที่คาดว่าจะได้รับ

- 1) ทราบถึงปัจจัยที่สัมพันธ์กับการเสียชีวิตของผู้ป่วยโควิด-19
- 2) สามารถนำตัวแบบในการทำนายไปใช้ในการคำนวณหาโอกาสที่ผู้ป่วยโควิด-19 จะเสียชีวิต

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 แนวคิดและทฤษฎีเกี่ยวกับโรคโควิด-19

2.1.1 ความหมายของไวรัสโคโรนา

ไวรัสโคโรนาคือไวรัสสายเดี่ยว (Single-Stranded) ที่มีสารพันธุกรรมชนิด RNA ไวรัสโคโรนาจัดอยู่ในสายพันธุ์เดียวกันกับนิโดไวรัสเลส (Nidovirales) และไวรัสโคโรนาไวรัสเด (Coronaviridae) เป็นสายพันธุ์รอง มาจากไวรัสสอโตโคโรนาไวรัสเด (Orthocoronaviridae) โดยจากการศึกษาพันธุกรรมของ ไวรัสสามารถแบ่งประเภทได้ดังนี้ อัลฟา (α), เบต้า (β), แกมมา (γ) และ เดลต้า (δ) ไวรัสโคโรนารวมอยู่ในกลุ่มไวรัสโคโรนาไวรัสเด ชื่อของไวรัสที่ถูกตั้งตามลักษณะของไวรัสที่เมื่อดูจากกล้องอิเล็กตรอนไมโครสโคปจะเห็นเป็นรูปร่างคล้ายมงกุฎ

2.1.2 รูปร่างและลักษณะของไวรัสโคโรนา

ไวรัสโคโรนามีเปลือกที่ห่อหุ้มสารพันธุกรรมชนิด RNA เป็นไวรัสที่มีส่วนประกอบครบสมบูรณ์ (Virion) มีลักษณะกลมหรือเป็นรูปวงรี ส่วนใหญ่มีภาวะหลายรูปแบบ มีขนาดเส้นผ่านศูนย์กลางตั้งแต่ 50 ถึง 200 นาโนเมตร โดยไวรัสโคโรนามีเส้นผ่านศูนย์กลาง 60 ถึง 140 นาโนเมตร จะมีปุ่มอยู่บนพื้นผิวภายนอก และก่อตัวในลักษณะเป็นท่อนโปรตีนหนามเป็นโปรตีนแอนติเจนหลักของไวรัสที่ใช้ในการเพิ่มปริมาณ มีโปรตีนนิวคลีโอแคปซิด (Nucleocapsid Protein) ห่อหุ้มสารพันธุกรรมที่เกิดจากไวรัส และสามารถใช้เป็นแอนติเจนวินิจฉัยสาเหตุโรค

2.1.3 ประเภทของไวรัสโคโรนา

ไวรัสโคโรนาส่วนมากจะติดต่อในสัตว์ ปัจจุบันไวรัสโคโรนาที่ตรวจพบในมนุษย์ 2 จีเนส ได้แก่ ไวรัสอัลฟาโคโรนา (229E, NL63), ไวรัสเบต้าโคโรนา (OC43, HKU1), โรคทางเดินหายใจตะวันออกกลาง (MERS-CoV) และโรคทางเดินหายใจเฉียบพลันรุนแรงหรือโรคใช้หัดมรณะ (SARS-CoV) เมื่อไม่นานมานี้ ไวรัสโคโรนาสายพันธุ์ใหม่ถูกพบครั้งแรกในการระบาดที่เมืองอู่ฮั่น มณฑลหูเป่ย์ สาธารณรัฐประชาชนจีน ในช่วงปลายปี 2019 คนไข้มีภาวะปอดอักเสบโดยไม่ทราบสาเหตุ องค์การอนามัยโลก เรียกโรคนี้ว่า โควิด-19 (COVID-19) ในขณะที่คณะกรรมการระหว่างประเทศว่าด้วยอนุกรมวิธานวิทยาของไวรัส (International Committee on Taxonomy of Viruses: ICTV) ตั้งชื่อไวรัสนี้ว่า SARS-CoV-2 ซึ่งต่อมาไวรัสชนิดนี้ได้รับการยืนยันว่าสามารถติดต่อจากคนสู่คนได้ ไวรัสโคโรนาสายพันธุ์ใหม่ที่พบเจอนี้มีความใกล้เคียงทางพันธุกรรมกับไวรัสที่พบในค้างคาว จากการวิเคราะห์ลำดับพันธุกรรมที่คล้ายคลึงกันของไวรัสสายพันธุ์ใหม่นี้แสดงให้เห็นว่ามันมีความคล้ายคลึง

กับไวรัสโรคซาร์สอย่างมาก ไวรัสโคโรนาสายพันธุ์ใหม่ถูกจัดอยู่ในประเภทเบต้าโคโรนาไวรัส (โรงพยาบาลพญาไท, 2564)

2.1.4 สายพันธุ์ของเชื้อ SARS-CoV-2

เนื่องจากไวรัสที่จะพัฒนาไปสู่เชื้อ SARS-CoV-2 มีวิวัฒนาการต่อเนื่องและยังคงมีการเปลี่ยนแปลงของข้อมูลที่ทำให้เข้าใจผลกระทบของสายพันธุ์หลายพันธุเหล่านี้ การปรับและการจัดประเภทสายพันธุ์ของเชื้อ SARS-CoV-2 การกระจายทางภูมิศาสตร์ของสายพันธุ์ที่น่ากังวล และการสรุปคุณสมบัติทางพันธุกรรมของไวรัส (ความสามารถในการแพร่กระจาย ความรุนแรงของโรค ความเสี่ยงในการติดเชื้อซ้ำ และผลกระทบต่อการศึกษาวิจัยและการทำงานของวัคซีน) ซึ่งสายพันธุ์ที่ไม่มีคุณสมบัติตามเกณฑ์ทุกข้อของสายพันธุ์ที่น่ากังวล สายพันธุ์ที่น่าสนใจ และสายพันธุ์ที่ต้องเฝ้าติดตาม อาจได้รับการกำหนดใหม่ และสายพันธุ์ที่มีความเสี่ยงลดลงเมื่อเทียบกับสายพันธุ์อื่นที่แพร่กระจายอยู่ อาจได้รับการจัดประเภทใหม่ โดยการหารือกับคณะที่ปรึกษาทางวิชาการด้านวิวัฒนาการของไวรัส ซึ่งก่อนหน้านี้เรียกว่า “คณะทำงานทางวิชาการด้านวิวัฒนาการของไวรัส”

2.1.4.1 สายพันธุ์ที่น่ากังวล (Variants of Concern: VOC)

สายพันธุ์ของเชื้อ SARS-CoV-2 ที่ตรงตามนิยามของสายพันธุ์ที่น่าสนใจ และผ่านการประเมินเปรียบเทียบ แสดงให้เห็นถึงความสัมพันธ์กับข้อใดข้อหนึ่ง หรือมากกว่านั้น ในระดับที่มีนัยสำคัญทางสาธารณสุข ดังนี้

- 1) มีความสามารถในการแพร่ระบาดเพิ่มขึ้น หรือมีการเปลี่ยนแปลงด้านระบาดวิทยาของโรคโควิด-19
- 2) มีความรุนแรงเพิ่มขึ้น หรือการเปลี่ยนแปลงทางคลินิกของโรค
- 3) มาตรการทางสาธารณสุข และทางสังคม การตรวจวินิจฉัยที่เป็นไปได้ การจัดการด้านวัคซีนกระบวนการรักษา มีประสิทธิผลลดลง

2.1.4.2 สายพันธุ์ย่อยของสายพันธุ์ที่น่ากังวลที่ต้องเฝ้าติดตาม (Variants of Concern Lineages Under Monitoring: VOC-LUM)

สายพันธุ์ย่อยที่อยู่ในวงศ์เดียวกันกับสายพันธุ์ที่น่ากังวลซึ่งกำลังแพร่กระจายอยู่ ตามที่ได้มีการวิเคราะห์วิวัฒนาการของสายพันธุ์ และส่งสัญญาณว่ามีข้อได้เปรียบในความสามารถทางการแพร่กระจาย เมื่อเปรียบเทียบกับสายพันธุ์ย่อยอื่น ๆ ในวงศ์เดียวกัน และมีการเปลี่ยนแปลงเพิ่มเติมของกรดอะมิโน ซึ่งเป็นที่ทราบชัดเจนหรือต้องสงสัยว่าก่อให้เกิดการเปลี่ยนแปลงที่สังเกตได้ทางระบาดวิทยา และมีข้อได้เปรียบทางความแข็งแรงเมื่อเทียบกับสายพันธุ์หลายพันธุอื่น ๆ ซึ่งแพร่กระจายอยู่

2.1.4.3 สายพันธุ์ที่น่าสนใจ (Variants of Interest: VOI) ที่เคยแพร่กระจายอยู่ก่อนหน้านี

มีการเปลี่ยนแปลงทางพันธุกรรมที่คาดการณ์ไว้ หรือเป็นที่ทราบกันดีว่าจะมีผลต่อคุณสมบัติของไวรัส เช่น ความสามารถในการแพร่กระจาย ความรุนแรงของโรค การหลบเลี่ยงภูมิ การวินิจฉัย ยารักษา และได้รับการระบุว่าก่อให้เกิดการแพร่ระบาดในชุมชน หรือกลุ่มก้อนของผู้ป่วยโรคโควิด-19 จำนวนมาก ในหลายประเทศโดยมีความชุกเพิ่มขึ้นควบคู่ไปกับการเพิ่มจำนวนของผู้ป่วยเมื่อเวลาผ่านไปหรือเกิดผลกระทบทางระบาดวิทยา ซึ่งแนะนำว่าเป็นความเสี่ยงที่อุบัติขึ้นต่อระบบสาธารณสุขโลก

2.1.4.4 สายพันธุ์ที่ต้องเฝ้าติดตาม (Variant Under Monitoring: VUM)

2.1.4.5 สายพันธุ์ที่เคยเฝ้าติดตาม (Formerly Monitored Variants: FMV)

สายพันธุ์ที่เคยได้รับการกำหนดให้เป็นสายพันธุ์น่ากังวล สายพันธุ์ที่น่าสนใจ สายพันธุ์ที่ต้องเฝ้าติดตาม และได้รับการจัดประเภทใหม่ตามเกณฑ์อย่างน้อยข้อใดข้อหนึ่ง ดังต่อไปนี้

- 1) สายพันธุ์ไม่ได้แพร่กระจายในระดับที่มีนัยสำคัญต่อสาธารณสุขระดับโลกอีกต่อไป
- 2) สายพันธุ์แพร่กระจายอยู่เป็นเวลานาน แต่ไม่มีผลกระทบต่อสถานการณ์ทางระบาดวิทยาโดยรวม
- 3) หลักฐานทางวิทยาศาสตร์บ่งชี้ว่า สายพันธุ์ไม่ได้มีความเกี่ยวข้องกับคุณสมบัติที่น่ากังวล (องค์การอนามัยโลก, 2565)

2.2 แนวคิดและทฤษฎีเกี่ยวกับความรุนแรงของโรคโควิด-19

2.2.1 ความรุนแรงของไวรัสโคโรนา

ไวรัสโคโรนาโดยปกติติดเชื้อได้มากในผู้ใหญ่ และเด็กโต ทำให้เป็นหวัดธรรมดา ในบางสายพันธุ์อาจทำให้มีอาการท้องร่วงในผู้ใหญ่ ไวรัสเหล่านี้ ติดต่อกันได้โดยผ่านละอองฝอย การแพร่ระบาดของไวรัสโคโรนาอยู่ในช่วงฤดูหนาว และฤดูใบไม้ผลิ ระยะเวลาฟักตัว มักจะอยู่ที่ 3-7 วัน หลังจากที่ ไข้หวัดกลายเป็น 2019-nCoV ระยะเวลาการฟักตัวอยู่ที่ 2-14 วัน อย่างไรก็ตาม มีบางกรณีพบว่าระยะเวลาฟักตัวนั้นมากถึง 24 วัน ระดับความอันตรายของไวรัสดังกล่าวนี้ มีความรุนแรงของโรคทั้งในระดับการติดเชื้อ และเสียชีวิต ไวรัสโคโรนาสายพันธุ์ใหม่มีอัตราการติดเชื้อในระดับสูงมาก และเป็นอันตรายต่อชีวิต

2.2.2 ลักษณะอาการของโรคโควิด-19

อาการทั่วไปของโรคโควิด-19 ที่พบมากที่สุดคือ ไข้ ไอ ลื่นไม่รับรส จมูกไม่ได้กลิ่น และอ่อนเพลีย อาการที่พบน้อยกว่าแต่อาจมีผลต่อผู้ป่วยบางรายคือ ปวดเมื่อย ปวดหัว คัดจมูก น้ำมูกไหล เจ็บคอ ท้องเสีย ตาแดง หรือผื่นตามผิวหนัง หรือสีผิวเปลี่ยนตามนิ้วมือนิ้วเท้า อาการเหล่านี้มักจะไม่รุนแรงนัก และค่อย ๆ เริ่มทีละน้อย บางรายติดเชื้อแต่มีอาการไม่รุนแรง ผู้ป่วยส่วนมากร้อยละ 80 หายป่วยได้โดยไม่ต้องเข้ารับการรักษาในโรงพยาบาล ประมาณ 1 ใน 5 ของผู้ติดเชื้อโควิด-19 มีอาการหนัก และหายใจลำบาก ผู้สูงอายุและมีโรคประจำตัวเช่น ความดันโลหิตสูง โรคหัวใจ โรคเบาหวาน หรือมะเร็งมีแนวโน้มที่จะมีอาการป่วยรุนแรงกว่า

ภาวะแทรกซ้อนที่อาจเกิดขึ้นได้ คือ ความล้มเหลวของระบบทางเดินหายใจ โรคทางเดินหายใจเฉียบพลัน (Acute Respiratory Distress Syndrome: ARDs) ภาวะติดเชื้อในกระแสเลือด ภาวะช็อก ภาวะลิ่มเลือดอุดตัน และความล้มเหลวของหลายอวัยวะ รวมถึงการบาดเจ็บของหัวใจ ตับ หรือไต มีรายงานว่าเด็ก และวัยรุ่นที่ติดเชื้อโควิด-19 อาจมีอาการที่หายากแต่ร้ายแรงกว่า เรียกว่า กลุ่มอาการอักเสบจากระบบหลายระบบในเด็ก ซึ่งนำไปสู่ความล้มเหลวของอวัยวะหลายส่วน และเกิดภาวะช็อก

2.2.3 ความรุนแรงของโรคโควิด-19

องค์การอนามัยโลกแบ่งความรุนแรงของโรคโควิด-19 ออกเป็น 5 ประเภท ดังนี้

1) ความรุนแรงอยู่ในระดับเล็กน้อย (Mild) อาการและลักษณะทางคลินิก ไข้ (อุณหภูมิมากกว่า 37.5 องศาเซลเซียส) มีอาการของระบบทางเดินหายใจเล็กน้อย หรืออาการคล้ายไข้หวัด โดยไม่มีภาวะพร่องออกซิเจนในเลือด (Hypoxemia) SpO₂ มากกว่าหรือเท่ากับร้อยละ 93 เจ็บคอ ไอแห้ง ๆ การรับรสที่ลดลง และอาจมีท้องเสียร่วมด้วยได้ ไม่พบความผิดปกติกรณีส่งตรวจภาพรังสีทรวงอกหรือเอกซเรย์คอมพิวเตอร์ทรวงอก

2) ความรุนแรงอยู่ในระดับปานกลาง (Moderate) เริ่มมีการอักเสบเพิ่มขึ้น อาการและลักษณะทางคลินิก ไข้ หายใจสั้น ไอ มีภาวะพร่องออกซิเจน SpO₂ น้อยกว่าร้อยละ 93 ระดับ IL-6 ต่ำกว่า 95 pg/mL การดำเนินโรคของกลุ่มนี้อาจแย่งลงได้ ควรเฝ้าระวังอาการและติดตามค่า SpO₂ กรณีส่งตรวจภาพรังสีทรวงอกหรือเอกซเรย์คอมพิวเตอร์ทรวงอก พบลักษณะของภาวะหลอดลมใหญ่ และปอดอักเสบ (Broncho Pneumonia)

3) ความรุนแรงอยู่ในระดับรุนแรง (Severe) อาการและลักษณะทางคลินิก ไข้ มีภาวะพร่องออกซิเจนรุนแรงมากขึ้น SpO₂ น้อยกว่าร้อยละ 88 มีค่า P/F ratio (PaO₂/FiO₂) ต่ำกว่า 200 มม.ปรอท อัตราการหายใจมากกว่า 25 ครั้งต่อนาที ระดับ IL-6 สูงกว่า 94 pg/mL การดำเนินของโรคทรุดลงมากจากลักษณะที่ 2 กรณีส่งตรวจภาพรังสีทรวงอก หรือเอกซเรย์คอมพิวเตอร์ทรวงอกพบลักษณะของ Organizing Pneumonia

4) ความรุนแรงอยู่ในระดับรุนแรง (Severe หรือ Pre Critical Illness) เป็นระยะของการเกิดภาวะหายใจลำบากเฉียบพลัน อาการและลักษณะทางคลินิก ใช้ และมีภาวะพร่อง ออกซิเจนรุนแรง มี Low Elastance หรือ High Compliance กรณีส่งตรวจภาพรังสีทรวงอกหรือเอกซเรย์คอมพิวเตอร์ทรวงอกพบลักษณะของ Ground Glass Opacities และปริมาณของ Non-Aerated Lung Area มีน้อย ทำให้ไม่ตอบสนองต่อการรักษาด้วยระดับ Positive End Expiratory Pressure (PEEP) ที่สูง และการใช้ใน Lung Recruitment Maneuver ซึ่งแตกต่างจากภาวะหายใจลำบากเฉียบพลันทั่วไป นอกจากนี้พบว่าระยะเวลาที่เริ่มปรากฏอาการคือ 8-12 วัน ต่างจาก ภาวะหายใจลำบากเฉียบพลันทั่วไปซึ่งจะเกิดภายใน 7 วันนับจากมีสาเหตุกระตุ้นหรืออาการทางระบบหายใจที่แยลง

5) ความรุนแรงอยู่ในระดับวิกฤต (Critical) เป็นกลุ่มที่มีการดำเนินโรครุนแรงมากขึ้น อาการและลักษณะทาง คลินิก ใช้ และมีภาวะพร่องออกซิเจนรุนแรง มี High Elastance หรือ Low Compliance พบน้ำในปอด และปอดแฟบเพิ่มขึ้น ทำให้ตอบสนองดีต่อ PEEP ที่สูง หรือใช้ Recruitment Maneuver คล้ายคลึงกับภาวะหายใจ ลำบากเฉียบพลันทั่วไป นอกจากนี้อาจทำให้เกิดภาวะหายใจล้มเหลว (Respiratory Failure) กลุ่มอาการตอบสนองต่อการอักเสบทั่วร่างกาย (Systemic Inflammatory Response Syndrome) ช็อก (Shock) หรือหัวใจล้มเหลว (Cardiac Failure) (ณยฎา, 2564)

2.2.4 อัตราการแพร่เชื้อและความรุนแรง

1) Basic Reproductive Rate (R0) คือ ค่าเฉลี่ยที่ผู้ป่วย 1 คนจะแพร่เชื้อให้ผู้อื่นในกลุ่ม ประชากรที่มีความไวต่อการรับเชื้อ ค่า R0 ของโรคโควิด-19 อยู่ที่ประมาณ 2 และ 4

2) Clinical Onset Interval คือ ช่วงเวลาแสดงอาการในผู้ป่วยรายต่อ ๆ กันในห่วงโซ่ของการแพร่ระบาด ในงานวิจัยส่วนใหญ่เวลาเฉลี่ยอยู่ที่ 4 และ 5 วัน

3) Case Fatality Ratio (CFR) หรืออัตราการเสียชีวิต คือ สัดส่วนของผู้ป่วยที่ถึงแก่ชีวิต อัตราคร่าว ๆ ของทั่วโลกอยู่ที่ร้อยละ 7 (ธ.ค. 2562 - พ.ค. 2563) (องค์การอนามัยโลก, 2564)

2.2.5 กลุ่มโรคที่เสี่ยงมีอาการรุนแรงในผู้ป่วยโควิด-19

1) โรคเบาหวาน ชนิดที่ 1 และชนิดที่ 2 รวมถึงผู้ป่วยโรคเบาหวานขณะตั้งครรภ์ มีความเสี่ยงในการเกิดอาการรุนแรงหากติดโรคโควิด-19 เนื่องจากผู้ป่วยโรคเบาหวานที่ควบคุมระดับ น้ำตาลได้ไม่ดี และมีระดับน้ำตาลในเลือดสูงกว่าเป้าหมาย มักมีโรคแทรกซ้อนอื่น ๆ ร่วมด้วย นอกจากนี้การติดเชื้อยังทำให้ระดับน้ำตาลในเลือดผันผวน และควบคุมได้ยาก ทำให้ภูมิคุ้มกันต่ำลง และติดเชื้อได้ง่ายขึ้น รวมถึงเชื้อโรคเจริญเติบโตได้ดีในภาวะน้ำตาลในเลือดสูง

2) โรคไตเรื้อรัง ผู้ป่วยโรคไตที่อยู่ในขั้นที่ 3 ถึงขั้นที่ 5 ผู้ป่วยที่รักษาด้วยการฟอกเลือด และผู้ป่วยที่เปลี่ยนถ่ายไตอยู่ในกลุ่มเสี่ยงในการมีอาการรุนแรง ผู้ป่วยที่รักษาด้วยการฟอกเลือดอาจมี ภูมิคุ้มกันต่ำกว่าคนปกติทั่วไป นอกจากนี้ผู้ป่วยที่ได้รับการเปลี่ยนถ่ายไตต้องได้รับยากดภูมิคุ้มกันทำให้ภูมิคุ้มกันลดต่ำลงทำให้ติดเชื้อได้ง่ายขึ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3) โรคปอดและทางเดินหายใจ ผู้ป่วยโรคหืดหอบระดับปานกลางถึงรุนแรง และผู้ป่วยโรคปอดเรื้อรัง เช่น โรคหลอดลมอุดกั้นเรื้อรัง (COPD) โรคปอดอักเสบเรื้อรัง และโรคซิสติกไฟโบรซิส เนื่องจากโควิด-19 มีผลกระทบต่อระบบทางเดินหายใจ กระตุ้นอาการหอบหืด และอาจทำให้เกิดโรคปอดบวม และโรคร้ายแรงอื่น ๆ

4) โรคอ้วน ผู้ที่มีค่าดัชนีมวลกายตั้งแต่ 40 ขึ้นไปหรือตั้งแต่ 30 ขึ้นไปสำหรับชาวเอเชีย ผู้ที่เป็นโรคอ้วนชนิดรุนแรงมีความเสี่ยงในการเกิดกลุ่มอาการหายใจลำบากเฉียบพลัน

5) โรคตับ ผู้ป่วยโรคตับเรื้อรัง เช่น โรคตับแข็ง ผู้ป่วยปลูกถ่ายตับที่ใช้ยากดภูมิคุ้มกัน ผู้ป่วยภาวะที่มีตับอักเสบจากภูมิไวเกิน และผู้ป่วยมะเร็งตับที่อยู่ระหว่างการรักษาด้วยเคมีบำบัด

6) ผู้มีภาวะภูมิคุ้มกันบกพร่อง กลุ่มผู้ป่วยโรคมะเร็ง กลุ่มผู้ป่วยโรคเอดส์ มีประวัติสูบบุหรี่ล้วนมีความเสี่ยงในการเกิดอาการรุนแรงเนื่องจากภูมิคุ้มกันที่อ่อนแอทำให้ร่างกายไม่สามารถต่อสู้กับเชื้อโรคได้อย่างมีประสิทธิภาพ

7) ผู้ป่วยที่ใช้ยากดภูมิคุ้มกัน ยากดภูมิคุ้มกันทำให้ภูมิคุ้มกันร่างกายอ่อนแอลง ทำให้มีความเสี่ยงต่อการติดเชื้อง่ายขึ้น และอาจมีความรุนแรงของการติดเชื้อสูงขึ้น

8) ผู้ป่วยโรคหัวใจ หากติดโควิด-19 อาจมีอาการหัวใจวายหรือภาวะหัวใจล้มเหลวได้ อาการโรคหัวใจที่รุนแรงยิ่งขึ้นนี้เกิดจากอาการป่วยของการติดเชื้อไวรัสและการที่หัวใจต้องทำงานหนักขึ้น เช่น หัวใจเต้นเร็วขึ้น ภาวะกล้ามเนื้อหัวใจอักเสบเฉียบพลัน (Myocarditis) เป็นต้น (โรงพยาบาลบำรุงราษฎร์, 2563)

2.3 แนวทางเวชปฏิบัติ การวินิจฉัย ดูแลรักษา และป้องกันการติดเชื้อในโรงพยาบาล

เมื่อวันที่ 22 เมษายน 2565 คณะทำงานด้านการรักษาพยาบาลและการป้องกันการติดเชื้อในโรงพยาบาล ได้ประกาศแนวทางเวชปฏิบัติ การวินิจฉัย ดูแลรักษา และป้องกันการติดเชื้อในโรงพยาบาล กรณีผู้ป่วยติดเชื้อไวรัสโคโรนา 2019 สำหรับแพทย์และบุคลากรสาธารณสุข ในประเด็นของการให้ยาด้านไวรัสในผู้ป่วยกลุ่มอาการต่าง ๆ ซึ่งได้แก่ กลุ่มผู้ป่วยที่ไม่มีอาการหรือสบายดี ให้ดูแลรักษาผู้ป่วยโดยไม่ต้องให้ยาด้านไวรัส กลุ่มผู้ป่วยที่มีอาการไม่รุนแรง ไม่มีปอดอักเสบ ไม่มีปัจจัยเสี่ยงต่อการเป็นโรครุนแรง/โรคร่วมสำคัญ และภาพถ่ายรังสีปอดปกติ อาจพิจารณาให้ยาฟาวิพิราเวียร์ (Favipiravir) กลุ่มผู้ป่วยที่มีอาการไม่รุนแรง แต่มีปัจจัยเสี่ยงต่อการเป็นโรครุนแรงหรือมีโรคร่วมสำคัญ หรือผู้ป่วยที่มีปอดอักเสบเล็กน้อย ยังไม่ต้องให้ออกซิเจน ให้พิจารณาให้ยาด้านไวรัสเพียง 1 ชนิด คือ ยาแพ็กซ์โลวิด (Paxlovid) ยาเรมเดซิเวียร์ (Remdesivir) หรือยาโมลนูพิราเวียร์ (Molnupiravir) อย่างใดอย่างหนึ่ง และกลุ่มผู้ป่วยปอดอักเสบที่มีภาวะพร่องออกซิเจน (Resting O₂ Saturation ≤ 94 %) หรือมีภาวะลดลงของออกซิเจน SpO₂ ≥ 3% ของค่าที่วัดได้ขณะออกแรง (Exercise-Induced Hypoxemia) หรือภาพรังสีทรวงอกมีความผิดปกติในเนื้อปอดที่บ่งสภาวะการอักเสบหรือติดเชื้อ แนะนำให้ยาเรมเดซิเวียร์ (Remdesivir) ร่วมกับให้ยาคอร์ติโคสเตียรอยด์ (Corticosteroid) (กรมการแพทย์, 2565)

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ไว้เพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.4 การวิเคราะห์การถดถอยลอจิสติก (Logistic Regression Analysis)

เป็นการศึกษาเพื่อทำนายว่าจะเกิดเหตุการณ์หนึ่งขึ้นหรือไม่ หรือมีโอกาสเกิดขึ้นมากน้อยเพียงใด โดยมีการกำหนดค่าตัวแปรตัวหนึ่งหรือหลายตัวที่คาดว่าจะส่งผลต่อการเกิดเหตุการณ์นั้น ๆ งานวิจัยนี้ใช้การวิเคราะห์การถดถอยลอจิสติกในการศึกษาเนื่องจากตัวแปรตามเป็นตัวแปรเชิงกลุ่ม โดยตัวแปรตามมีลักษณะเป็นตัวแปรแบ่งสอง (Dichotomous Variable) สำหรับการวิเคราะห์การถดถอยลอจิสติกทวิ (Binary Logistic Regression) เพื่อวิเคราะห์หาปัจจัยเสี่ยงที่สัมพันธ์ต่อการเสียชีวิตของผู้ป่วยโควิด-19 ซึ่งเป็นการหาความสัมพันธ์ระหว่างตัวแปรตาม 1 ตัวและตัวแปรอิสระมากกว่า 1 ตัว โดยตัวแปรตามสามารถมีค่าได้เพียง 2 ค่า คือ 1 และ 0 ซึ่งจะแสดงถึงการเสียชีวิตหรือรอดชีวิต ส่วนตัวแปรอิสระที่ใช้ในการวิเคราะห์สามารถเป็นได้ทั้งตัวแปรเชิงปริมาณและตัวแปรเชิงคุณภาพ อย่างไรก็ตาม การวิเคราะห์ด้วยวิธีนี้มีเงื่อนไขดังต่อไปนี้

- 1) ค่าความคลาดเคลื่อนของการสำรวจต้องเป็นอิสระจากกัน (Independence of Errors)
- 2) ตัวแปรตามมีความสัมพันธ์เชิงเส้นตรงกับฟังก์ชันลอจิสติกของตัวแปรอิสระแบบต่อเนื่อง
- 3) ตัวแปรอิสระไม่มีความสัมพันธ์กันหรือไม่เกิดปัญหาความสัมพันธ์เชิงเส้นพหุ (Multicollinearity)
- 4) ไม่มีค่าผิดปกติ (Outlier) (Stoltzfus, 2011)

การวิเคราะห์การวิเคราะห์การถดถอยลอจิสติกทวิ จะต้องใช้ขนาดตัวอย่าง $n \geq 30p$ โดยที่ p คือ จำนวนตัวแปรอิสระ (กัลยา, 2546)

สำหรับการวิเคราะห์การถดถอยลอจิสติกทวิ ที่ค่าตัวแปรตาม หรือ y สามารถเป็นไปได้เพียง 2 ค่าคือ ไม่เกิดเหตุการณ์ ($y=0$) หรือเกิดเหตุการณ์ ($y=1$) มีความสัมพันธ์กับตัวแปรอิสระไม่อยู่ในรูปเชิงเส้น โดยมีตัวแบบในการทำนายข้อมูลประชากรตามสมการที่ 3.1 ดังนี้

$$P_y = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}} \quad (3.1)$$

และตัวแบบในการทำนายข้อมูลตัวอย่างตามสมการที่ 3.2 ดังนี้

$$P_y = \frac{e^{b_0 + b_1 x_1 + \dots + b_p x_p}}{1 + e^{b_0 + b_1 x_1 + \dots + b_p x_p}} \quad (3.2)$$

โดยที่ P_y แทนความน่าจะเป็นของการเกิดเหตุการณ์ที่สนใจ

x_i แทนตัวแปรอิสระ i

β_i แทนค่าสัมประสิทธิ์ของตัวแปร i ของประชากร

b_i แทนค่าสัมประสิทธิ์ของตัวแปร i ของตัวอย่าง

e แทนค่าคงที่ ($e = 2.71828$)

เมื่อพิจารณาความสัมพันธ์ระหว่างตัวแปรอิสระ และตัวแปรตามของการวิเคราะห์การถดถอย ลอจิสติกที่ไม่เป็นรูปเชิงเส้น จึงต้องมีการปรับให้ความสัมพันธ์อยู่ในรูปเชิงเส้นในรูปของ Odds ซึ่งหมายถึง อัตราส่วนระหว่างโอกาสที่จะเกิดเหตุการณ์ที่สนใจ ($y=1$) และโอกาสที่จะไม่เกิดเหตุการณ์ที่สนใจ ($y=0$) ดังแสดงในสมการที่ 3.3

$$Odds = \frac{P_y}{Q_y} \quad (3.3)$$

โดยที่ P_y แทนความน่าจะเป็นของการเกิดเหตุการณ์ที่สนใจ

Q_y แทนความน่าจะเป็นของการไม่เกิดเหตุการณ์ที่สนใจ

และ Odds Ratio จะหมายถึงอัตราส่วนของค่า Odds เปรียบเทียบกันระหว่าง 2 ค่าของตัวแปรอิสระ ซึ่งจะบ่งบอกถึงโอกาสในการเกิดเหตุการณ์ที่สนใจว่าเป็นกี่เท่าเมื่อเทียบระหว่างค่าของตัวแปรอิสระ 2 ค่า ตัวอย่างเช่น Odds Ratio ของโอกาสที่ผู้ป่วยจะเสียชีวิตเป็นดังนี้

$$Odds = \frac{P_{Died}}{P_{Survival}} \quad (3.4)$$

1) ค่า Odds Ratio มีค่าเท่ากับ 1 หมายความว่าโอกาสที่ผู้ป่วยจะเสียชีวิตและรอดชีวิตมีโอกาสเกิดขึ้นเท่า ๆ กัน

2) ค่า Odds Ratio มีค่ามากกว่า 1 หมายความว่าโอกาสที่ผู้ป่วยจะเสียชีวิตมากกว่าโอกาสที่ผู้ป่วยจะรอดชีวิต

3) ค่า Odds Ratio มีค่าน้อยกว่า 1 หมายความว่าโอกาสที่ผู้ป่วยจะรอดชีวิตมากกว่าโอกาสที่ผู้ป่วยจะเสียชีวิต

จากสมการที่ 3.4 หากค่าฟังก์ชันตอบสนองลอจิสติก (Logit Response Function: Logit) พบว่า ลอจิสติกของข้อมูลประชากรอยู่ในรูปเชิงเส้น ดังสมการที่ 3.5

$$\log(Odds) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p \quad (3.5)$$

และลอจิสติกของข้อมูลตัวอย่างอยู่ในรูปเชิงเส้น ดังสมการที่ 3.6

$$\log(Odds) = b_0 + b_1 x_1 + \dots + b_p x_p \quad (3.6)$$

2.4.1 การเลือกตัวแปรอิสระด้วยวิธี Backward Stepwise

วิธี Backward Stepwise เป็นวิธีที่นำตัวแปรอิสระทั้ง p ตัว ($x_1, x_2, x_3, \dots, x_p$) เข้าสมการพร้อมกันก่อนจากนั้นพิจารณาตัวแปรอิสระที่อธิบายความผันแปรของตัวแปรตามได้น้อยที่สุดออกจากสมการก่อน ทำเช่นนี้ไปเรื่อย ๆ จนกระทั่งเหลือตัวแปรอิสระที่สามารถอธิบายความผันแปรของตัวแปรตามได้อย่างมีนัยสำคัญทางสถิติ

2.4.2 การตรวจสอบความเหมาะสมของตัวแบบการวิเคราะห์การถดถอยลอจิสติก

2.4.2.1 พิจารณาค่าความเป็นไปได้ (Likelihood Value)

การวิเคราะห์การถดถอยแบบปกติจะประเมินคุณภาพของตัวแบบหรือสมการทำนายจากค่าผลบวกกำลังสองของส่วนเหลือ (Residual Sum of Squares) แต่สำหรับการวิเคราะห์การถดถอยลอจิสติก จะประเมินจากค่า Log-Likelihood (LL) ซึ่งเทียบเคียงได้กับค่าส่วนเหลือ (Residuals) นั่นก็คือ ยิ่งค่า LL น้อย ยิ่งแสดงว่าตัวแบบการถดถอยลอจิสติกที่สร้างขึ้นมีคุณภาพหรือกลมกลืนกับข้อมูลดี โดยจะศึกษาจากค่า 2LL (-2 Log-likelihood) ซึ่งเป็นค่ามาจาก Log-likelihood ที่คูณด้วย -2 เพื่อต้องการให้ค่าที่ได้มีการแจกแจง χ^2 สำหรับการทดสอบนัยสำคัญทางสถิติ การพิจารณาค่า 2LL ถ้ามีค่าต่ำ หมายความว่าตัวแบบลอจิสติกมีความเหมาะสม ในการทดสอบนัยสำคัญความเหมาะสมของตัวแบบลอจิสติก ใช้สถิติ χ^2 ที่ $df = p$ (จำนวนตัวแปรอิสระ) ทดสอบสมมติฐาน (อรทัย, 2560) ดังนี้

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$$

$$H_1 : \beta_i \neq 0 \text{ อย่างน้อย 1 ค่า ; } i = 1, 2, \dots, p$$

โดยที่
$$\text{Log-Likelihood} = \sum_{i=1}^n [y_i \ln(\hat{y}_i) + (1 - y_i) \ln(1 - \hat{y}_i)] \quad (3.7)$$

และ
$$\chi^2 = 2[L(1) - L(0)] \quad (3.8)$$

โดยที่ $L(0)$ แทนค่า Null deviance หรือค่า Log-Likelihood สำหรับ Constant-Only Model

$L(1)$ แทนค่า Residual deviance หรือค่า Log-Likelihood สำหรับ Full Model

การทดสอบถ้าค่าสถิติ χ^2 มีนัยสำคัญหรือยอมรับ H_1 แสดงว่า กลุ่มของตัวแปรอิสระสามารถร่วมกันทำนายโอกาสของการเกิดเหตุการณ์ที่สนใจ ($y=1$) ได้ด้วยความเชื่อมั่น $(1 - \alpha) \times 100\%$

2.4.2.2 การทดสอบของสัมประสิทธิ์ถดถอยลอจิสติก

การทดสอบของสัมประสิทธิ์ถดถอยลอจิสติกเป็นการทดสอบว่าตัวแปรอิสระมีผลต่อการเปลี่ยนแปลงของตัวแปรตามหรือไม่ ทดสอบด้วยสถิติทดสอบของวัลด์ (Wald Statistic) เป็นการทดสอบสมมติฐาน ดังนี้

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$$

$$H_1 : \beta_i \neq 0 \text{ อย่างน้อย 1 ค่า ; } i = 1, 2, \dots, p$$

สถิติทดสอบของวัลด์คือ

$$w = \frac{\beta_i}{\text{estimate SE}(\beta_i)} \quad (3.9)$$

ในการทดสอบถ้าผลการทดสอบไม่มีนัยสำคัญ หรือยอมรับ H_0 แสดงว่าตัวแปรอิสระที่ i ไม่มีผลต่อการเปลี่ยนแปลง Odds Ratio ดังนั้นตัวแปรอิสระที่ i จึงไม่มีผลต่อความน่าจะเป็นของการเกิดเหตุการณ์นั้น และหากมีนัยสำคัญ หรือยอมรับ H_1 โดยสถิติทดสอบของวัลด์จะมีการแจกแจงแบบแซด (z)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.4.2.3 สถิติทดสอบระดับความสัมพันธ์ McFadden's R square

สถิติทดสอบ McFadden's R Square เป็นการพิจารณาหรือตรวจสอบความสอดคล้องของตัวแบบ หรือเปอร์เซ็นต์ที่สามารถอธิบายความแปรปรวน หรือความผันแปรในการวิเคราะห์ถดถอยลอจิสติก สถิตินี้จะคล้ายกับค่า R^2 ในการวิเคราะห์ถดถอยพหุคูณปกติ คำนวณได้จากสมการดังนี้

$$R^2 = 1 - \frac{\ln L(1)}{\ln L(0)} \quad (3.10)$$

2.4.2.4 สัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน (Pearson's Correlation)

สัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันเป็นการศึกษาความสัมพันธ์เชิงเส้น (Linear Relationship) ระหว่างตัวแปร 2 ตัวหรือข้อมูล 2 ชุด โดยมีข้อตกลงเบื้องต้นคือตัวแปรทั้งสองต้องเป็นตัวแปรต่อเนื่อง อยู่ในมาตราอันตรภาคชั้น (Interval Scale) ขึ้นไป และมีการแจกแจงปกติสองตัวแปร (Bivariate Normal Distribution) โดยค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันจะมีค่าอยู่ระหว่าง -1 ถึง 1 ถ้าเข้าใกล้ ± 1 หมายความว่าตัวแปรทั้งสองมีความสัมพันธ์กันอยู่ในระดับมาก ถ้าสัมประสิทธิ์สหสัมพันธ์มีค่าเข้าใกล้ 0 หมายความว่าตัวแปรทั้งสองมีความสัมพันธ์กันในระดับน้อยหรือไม่สัมพันธ์กันเลย เครื่องหมายบวก (+) จะแสดงทิศทางของความสัมพันธ์ว่ามีทิศทางเป็นแบบใด โดยถ้าสัมประสิทธิ์สหสัมพันธ์มีเครื่องหมายบวก (+) หมายความว่าตัวแปรทั้งสองมีความสัมพันธ์ไปในทิศทางเดียวกัน หากมีเครื่องหมายลบ (-) หมายความว่าตัวแปรทั้งสองมีความสัมพันธ์ไม่ไปในทิศทางเดียวกันหรือตรงกันข้าม สัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันสามารถคำนวณได้จากสูตร (ชินันท์, 2563) ดังนี้

$$r_{x_i, x_j} = \frac{n(\sum x_i x_j) - (\sum x_i)(\sum x_j)}{\sqrt{[n(\sum x_i^2) - (\sum x_i)^2][n(\sum x_j^2) - (\sum x_j)^2]}} \quad (3.11)$$

โดยที่ r_{x_i, x_j} แทนสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน
 $\sum x_i$ แทนผลรวมของข้อมูลที่วัดจากตัวแปร x_i
 $\sum x_j$ แทนผลรวมของข้อมูลที่วัดจากตัวแปร x_j
 $\sum x_i x_j$ แทนผลรวมของผลคูณของตัวแปร x_i และ x_j
 $\sum x_i^2$ แทนผลรวมกำลังสองของข้อมูลที่วัดจากตัวแปร x_i
 $\sum x_j^2$ แทนผลรวมกำลังสองของข้อมูลที่วัดจากตัวแปร x_j
 n แทนขนาดตัวอย่าง

2.5 ต้นไม้ตัดสินใจ (Decision Tree)

ในช่วงปลายของยุค 1970 J. Ross Quinlan ได้คิดค้นขั้นตอนวิธี (Algorithm) สำหรับสร้างต้นไม้ตัดสินใจที่มีชื่อว่า ID3 (Iterative Dichotomiser) ต่อมาเขาได้พัฒนาต่อยอด ID3 ไปเป็น C4.5 ซึ่งได้กลายมาเป็นขั้นตอนวิธีพื้นฐานที่ใช้สำหรับเปรียบเทียบประสิทธิภาพการทำงานของขั้นตอนวิธีต่าง ๆ ทางด้านการเรียนรู้แบบมีผู้สอน (Supervised Learning) (โกเมศ, ม.ป.ป.)

การสร้างต้นไม้ตัดสินใจ C4.5 ใช้ค่ามาตรฐานอัตราส่วนเกน (Gain Ratio) เพื่อเลือกลักษณะเฉพาะที่จะใช้เป็นรากหรือโหนด ถ้าให้ชุดของข้อมูล M ประกอบด้วยค่าที่เป็นไปได้คือ $\{m_1, m_2, \dots, m_n\}$ และให้ความน่าจะเป็นที่จะเกิดค่า m_i มีค่าเท่ากับ $P(m_i)$ จะได้ว่า ค่าเกนสารสนเทศ (Information Gain) หรือที่เรียกอีกอย่างหนึ่งว่า ค่าเอนโทรปี (Entropy) ของ M เขียนแทนด้วย $I(M)$ คำนวณได้ดังสมการที่ 3.12

$$I(M) = -\sum_{i=1}^n P(m_i) \log_2 P(m_i) \quad (3.12)$$

ถ้าให้ข้อมูลสอน คือ T และลักษณะเฉพาะที่เป็นโหนด คือ x และมีค่าทั้งหมดที่เป็นไปได้ n ค่า โหนดปัจจุบันจะแบ่งตัวอย่าง T ออกตามกิ่งเป็น $\{t_1, t_2, \dots, t_n\}$ ตามค่าที่เป็นไปได้ของ x ดังนั้นจึงสามารถคำนวณค่าเกนสารสนเทศ หลังจากแบ่งตามลักษณะเฉพาะ x ได้ดังสมการที่ 3.13

$$I_x(T) = \sum_{i=1}^n \frac{|t_i|}{|T|} I(t_i) \quad (3.13)$$

ค่ามาตรฐานเกน (Gain) ของลักษณะเฉพาะ x ได้ดังสมการ

$$Gain(x) = I(T) - I_x(T) \quad (3.14)$$

ถ้าให้ T คือ ชุดของตัวอย่าง เมื่อแบ่งตัวอย่างนี้ตามลักษณะเฉพาะ x จะได้ชุดของตัวอย่างย่อยในแต่ละกิ่ง คือ $\{t_1, t_2, \dots, t_n\}$ จำนวน n ชุด ตามค่าที่เป็นไปได้ในลักษณะเฉพาะ x จะสามารถคำนวณค่าสารสนเทศของการแบ่งแยก (Split Information) ของลักษณะเฉพาะแต่ละตัวได้ดังสมการที่ 3.15

$$Split\ Information(T) = -\sum_{i=1}^n \frac{|t_i|}{|T|} \log_2 \frac{|t_i|}{|T|} \quad (3.15)$$

ค่ามาตรฐานอัตราส่วนเกน (Gain Ratio) ของลักษณะเฉพาะ x ได้ดังสมการ

$$Gain\ Ratio(x) = \frac{Gain(x)}{Split\ Information(T)} \quad (3.16)$$

ท้ายสุดจึงเลือกค่ามาตรฐานอัตราส่วนเกนสูงสุดเป็นลักษณะเฉพาะเริ่มต้น และเลือกลักษณะเฉพาะถัดไปตามค่ามาตรฐานอัตราส่วนเกนน้อยลงตามลำดับ (พรพล และคณะ, 2551)

2.6 โครงข่ายประสาทเทียมแบบหลายลำดับชั้น (Multilayer Perceptron Artificial Neural Network)

โครงข่ายประสาทเทียมแบบหลายลำดับชั้นเป็นภาพแบบหนึ่งของโครงข่ายประสาทเทียมที่มีโครงสร้างเป็นแบบหลาย ๆ ชั้น ใช้สำหรับงานที่มีความซับซ้อนโดยมีกระบวนการฝึกฝนเป็นแบบมีผู้สอน และใช้ขั้นตอนการส่งค่าย้อนกลับ (Back Propagation) สำหรับการฝึกฝนกระบวนการส่งค่าย้อนกลับ ประกอบด้วย 2 ส่วนย่อยคือ การส่งผ่านไปข้างหน้า (Forward Pass) การส่งผ่านย้อนกลับ (Backward Pass) สำหรับการส่งผ่านไปข้างหน้า ข้อมูลจะผ่านเข้าโครงข่ายประสาทเทียมที่ชั้นข้อมูลเข้า และจะส่งผ่านจากอีกชั้นหนึ่งไปสู่อีกชั้นหนึ่งจนกระทั่งถึงชั้นข้อมูลออก ส่วนการส่งผ่านย้อนกลับ คำนวณน้ำหนักการเชื่อมต่อจะถูกปรับเปลี่ยนให้สอดคล้องกับกฎการแก้ข้อผิดพลาด (Error-correction) คือผลต่างของผลตอบที่แท้จริง (Actual Response) กับผลตอบเป้าหมาย (Target Response) เกิดเป็นสัญญาณผิดพลาด (Error Signal) ซึ่งสัญญาณผิดพลาดนี้จะถูกส่งย้อนกลับเข้าสู่โครงข่ายประสาทเทียมในทิศทางตรงกันข้ามกับการเชื่อมต่อ และค่าน้ำหนักของการเชื่อมต่อจะถูกปรับจนกระทั่งผลตอบที่แท้จริงเข้าใกล้ผลตอบเป้าหมาย สัญญาณที่มีโครงข่ายประสาทเทียมแบบหลายลำดับชั้น มี 2 ประเภทคือ

- 1) สัญญาณฟังก์ชัน (Function Signal) เป็นสัญญาณเข้าที่มาจากโหนดในชั้นก่อนหน้า และจะส่งผ่านไปข้างหน้าจากโหนดหนึ่งไปสู่อีกโหนดหนึ่ง
- 2) สัญญาณผิดพลาด เป็นสัญญาณย้อนกลับที่เกิดขึ้นที่โหนดในชั้นข้อมูลออกของโครงข่ายประสาทเทียม และถูกส่งผ่านย้อนกลับจากชั้นหนึ่งไปสู่อีกชั้นหนึ่ง (นิเวศ, 2553)

โดยโครงข่ายประสาทเทียมแบบหลายลำดับชั้น มีการเชื่อมโยงกัน 3 ชั้น ประกอบด้วยชั้นข้อมูลเข้า (Input Layer) ถัดมาเป็นชั้นซ่อน (Hidden Layer) และชั้นสุดท้ายคือชั้นข้อมูลออก (Output Layer) โดยส่วนประกอบที่ถูกบรรจุอยู่ในเซลล์ประสาทแต่ละตัวประกอบด้วย 2 ฟังก์ชันย่อย คือ ฟังก์ชันผลรวม (Summation Function) และฟังก์ชันกระตุ้น (Activation Function)

ฟังก์ชันผลรวมทำหน้าที่ในการคำนวณผลรวมของข้อมูลที่ได้จากชั้นข้อมูลเข้าซึ่งคำนวณได้ดังสมการที่ 3.17

$$g = \sum_{i=1}^z x_i w_i + \beta \quad (3.17)$$

โดยที่ g แทนค่าข้อมูลออกจากฟังก์ชันผลรวม

x_i แทนค่าข้อมูลเข้าตัวที่ i

w_i แทนค่าถ่วงน้ำหนักของข้อมูลเข้าตัวที่ i

z แทนจำนวนโหนดในชั้นซ่อน

β แทนค่าความเอนเอียง (Bias)

ฟังก์ชันกระตุ้นทำหน้าที่ปรับเปลี่ยนค่าของข้อมูลที่ได้จากฟังก์ชันผลรวมให้อยู่ในช่วงที่ต้องการ ฟังก์ชันกระตุ้นที่นิยม ได้แก่ ฟังก์ชันเชิงเส้น (Linear Function) ฟังก์ชันซิกมอยด์ (Sigmoid Function) และฟังก์ชันไฮเพอร์โบลิกแทนเจนต์ (Hyperbolic Tangent Function) (สายชล, 2563)

โดยในแต่ละชั้นไม่จำเป็นต้องเป็นฟังก์ชันกระตุ้นเดียวกันก็ได้ ชั้นซ่อนนั้นมีหน้าที่สำคัญคือ จะพยายามแปลงข้อมูลที่เข้ามาในชั้น (Layer) ให้สามารถแยกแยะความแตกต่างโดยใช้เส้นตรงเส้นเดียว (Linearly Separable) และก่อนที่ข้อมูลจะถูกส่งไปถึงชั้นข้อมูลออก ในบางครั้งอาจจำเป็นต้องใช้ชั้นซ่อนมากกว่า 1 ชั้น ในการแปลงข้อมูลให้อยู่ในภาพเส้นตรงเส้นเดียว ในการคำนวณหาข้อมูลออก (Output) ในปัญหาการจำแนกทำได้โดยการใส่ข้อมูลเข้า (Input) เข้าไปในโครงข่ายประสาทเทียมที่เราได้ทำการหาไว้แล้ว จากนั้นให้ทำการเปรียบเทียบค่าของข้อมูลออกในชั้นข้อมูลออก และให้ทำการเลือกค่าของข้อมูลออก ที่มีค่าเซลล์ประสาท (Neuron) ที่สูงกว่า และทำการรับค่าของทำนายที่ตรงกับเซลล์ประสาทที่เลือก และให้นำมาเปรียบเทียบกับค่าที่ยอมรับได้ หากอยู่ในช่วงที่รับได้ (ค่าความคลาดเคลื่อน (Error) น้อยกว่าค่าความคลาดเคลื่อนที่กำหนด) ก็ให้ทำการรับข้อมูลชุดถัดไป แต่หากมากกว่าค่าที่ยอมรับได้ ให้ทำการปรับค่าน้ำหนัก และค่าความเอนเอียง (Bias) ตามขั้นตอนที่ได้กล่าวไว้ข้างต้น เมื่อทำการปรับน้ำหนักเรียบร้อยแล้ว ให้ทำการรับข้อมูลชุดถัดไป และทำตามขั้นตอนซ้ำอีกรอบจนกระทั่งถึงข้อมูลชุดสุดท้าย และเมื่อทำข้อมูลชุดสุดท้ายเสร็จ จะนับเป็น 1 รอบของการคำนวณ (1 Epoch) จากนั้นจะทำการหาค่าผิดพลาดรวมเฉลี่ย จากค่าความผิดพลาดที่ได้เก็บเอาไว้ เพื่อใช้ในการตรวจสอบว่าค่าความผิดพลาดโดยเฉลี่ยในการจำแนกนั้น มีค่าน้อยกว่าค่าผิดพลาดที่ยอมรับได้หรือไม่ ถ้าใช่แสดงว่าโครงข่ายประสาทเทียมที่สร้างขึ้นนั้นสามารถให้ผลลัพธ์ที่ถูกต้องของทุก ๆ ข้อมูลแล้ว จึงทำการจบการเรียนรู้ได้ แต่ถ้าไม่ใช่ ให้กลับไปทำตามขั้นตอนแรกโดยเริ่มรับข้อมูลชุดที่ 1 ใหม่ (นิเวศ, 2553)

2.6.1 ฟังก์ชันกระตุ้น (Activation Function)

1) ฟังก์ชันซิกมอยด์ (Sigmoid Function)

$$f(x) = \frac{1}{1 + e^{-x}} \quad (3.18)$$

2) ฟังก์ชันเรคตีไฟด์ลินีเยยูนิต (Rectified Linear Unit Function: ReLU)

$$f(x) = \max(0, x) \quad (3.19)$$

2.7 วิธีสุ่มเพิ่มจำนวนข้อมูล (Synthetic Minority Oversampling Technique: SMOTE)

SMOTE เป็นเทคนิคที่ใช้ในการแก้ปัญหาการจำแนกข้อมูลที่ไม่สมดุล เนื่องจากข้อมูลในแต่ละคลาสมีจำนวนแตกต่างกันมาก ทำให้ผลลัพธ์ของการจำแนกอยู่ในข้อมูลกลุ่มมาก ดังนั้นวิธี SMOTE เป็นการเพิ่มจำนวนข้อมูลกลุ่มน้อยให้มีจำนวนเพิ่มขึ้น โดยการเพิ่มข้อมูลในกลุ่มน้อยนั้นทำให้การกระจายของกลุ่มข้อมูลมีความสมดุลมากขึ้น โดยทำการสุ่มค่าข้อมูลที่อยู่ในกลุ่มข้อมูลน้อยขึ้นมา 1 ค่า หลังจากนั้นพิจารณาค่าเพื่อนบ้านใกล้สุด K ตัว (K-Nearest Neighbor) แล้วคำนวณค่าระยะทาง (Euclidean Distance) ระหว่างค่าที่สุ่มกับค่าข้อมูลใกล้เคียงแต่ละค่า เพื่อหาค่าระยะทางที่น้อยที่สุดระหว่างค่าที่สุ่มกับค่าข้อมูลใกล้เคียง จากนั้นจึงสร้างข้อมูลเทียมระหว่างค่าข้อมูลที่สุ่มกับค่าข้อมูลใกล้เคียงตัวที่ให้ค่าระยะทางที่น้อยที่สุด (พุทธิพร, 2562)

2.8 ค่ามาตรฐานแซด (Z Standardization)

ค่ามาตรฐานแซด เป็นการแปลงข้อมูลเพื่อให้สามารถเปรียบเทียบกันได้ระหว่างข้อมูล หรือระหว่างชุดข้อมูลที่ค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐานไม่เท่ากันโดยมีสมการ ดังนี้

$$z_i = \frac{x_i - \mu}{\sigma} \quad (3.20)$$

โดยที่ z_i แทนค่ามาตรฐานตัวที่ i

x_i แทนค่าข้อมูลตัวแปรอิสระตัวที่ i

μ แทนค่าเฉลี่ยของตัวแปร x

σ แทนค่าส่วนเบี่ยงเบนมาตรฐานของตัวแปร x

จะได้ว่า z_i จะมีการแจกแจงปกติด้วยค่าเฉลี่ยเท่ากับ 0 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1

2.9 การเปรียบเทียบประสิทธิภาพของตัวแบบ

นำผลการวิเคราะห์ของแต่ละวิธีทั้ง 3 วิธี มาเปรียบเทียบประสิทธิภาพโดยพิจารณาจากเมทริกซ์ความสับสน (Confusion matrix) เป็นรูปแบบตารางที่เฉพาะเจาะจงที่นำผลลัพธ์จากการทำนายมาใส่ในรูปตารางเมทริกซ์ซึ่งจะช่วยให้ง่ายต่อการอ่านค่าทำนายของขั้นตอนวิธี ดังตารางที่ 2.1

ตารางที่ 2.1 เมทริกซ์ความสับสน

		ค่าทำนาย	
		คำตอบเป็นบวก	คำตอบเป็นลบ
ค่าจริง	คำตอบเป็นบวก	TP	FN
	คำตอบเป็นลบ	FP	TN

ค่าบวกจริง (True Positive: TP) คือ จำนวนข้อมูลที่ทำนายถูกว่าเป็นบวก ซึ่งค่าที่แท้จริงเป็นบวก
 ค่าลบจริง (True Negative: TN) คือ จำนวนข้อมูลที่ทำนายถูกว่าเป็นลบ ซึ่งค่าที่แท้จริงเป็นลบ
 ค่าบวกเท็จ (False Positive: FP) คือ จำนวนข้อมูลที่ทำนายผิดว่าเป็นบวก ซึ่งค่าที่แท้จริงเป็นลบ
 ค่าลบเท็จ (False Negative: FN) คือ จำนวนข้อมูลที่ทำนายผิดว่าเป็นลบ ซึ่งค่าที่แท้จริงเป็นบวก
 (สายชล, 2563)

ค่าความแม่นยำ (Accuracy) คือ การแสดงการวัดที่ได้มีความถูกต้องในรูปอัตราส่วนของจำนวนข้อมูลที่ทำนายถูกทั้งบวกและลบ จากจำนวนข้อมูลทั้งหมด ดังสมการ 3.21

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.21)$$

ค่าความเที่ยง (Precision) คือ อัตราส่วนของจำนวนข้อมูลที่ทำนายถูกว่าเป็นบวกจากจำนวนข้อมูลที่ทำนายถูกว่าเป็นบวกทั้งหมด ดังสมการที่ 3.22

$$Precision = \frac{TP}{TP + FP} \quad (3.22)$$

ค่าเรียกคืน (Recall) คือ อัตราส่วนของจำนวนข้อมูลที่ทำนายถูกว่าเป็นลบ จากจำนวนข้อมูลแท้จริงที่เป็นลบทั้งหมด ดังสมการที่ 3.23

$$Recall = \frac{TP}{TP + FN} \quad (3.23)$$

ค่าความถ่วงดุล (F1-score) คือ ค่าที่แสดงความสัมพันธ์ระหว่างค่าความเที่ยง และค่าเรียกคืน ดังสมการที่ 3.24

$$F1 - score = \frac{2(Precision)(Recall)}{(Precision)(Recall)} \quad (3.24)$$

ค่าอัตราความถูกต้องเชิงบวก (True Positive Rate) คืออัตราส่วนของจำนวนข้อมูลที่ทำนายถูกว่าเป็นบวก จากจำนวนข้อมูลแท้จริงที่เป็นบวกทั้งหมด ดังสมการที่ 3.25

$$True Positive Rate = \frac{TP}{TP + FN} \quad (3.25)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ค่าอัตราความผิดพลาดเชิงบวก (False Positive Rate) คืออัตราส่วนของจำนวนข้อมูลที่ทำนายผิดว่าเป็นบวก ซึ่งค่าที่แท้จริงเป็นลบ จากจำนวนข้อมูลแท้จริงที่เป็นลบทั้งหมด ดังสมการที่ 3.26

$$\text{False Positive Rate} = \frac{FP}{FP + TN} \quad (3.26)$$

ค่าพื้นที่ใต้โค้ง (Area Under the Curve: AUC) คือ ค่าพื้นที่ใต้เส้นโค้ง ROC (Receiver Operating Characteristic Curve) โดยค่าในแนวแกน x จะเป็นค่าอัตราความผิดพลาดเชิงบวก ส่วนเส้นในแกน y จะเป็นค่าอัตราความถูกต้องเชิงบวก ที่จุดตัดต่าง ๆ เพื่อแบ่งผลลัพธ์ของการทำนาย ออกเป็นกลุ่มเหตุการณ์ที่สนใจ และเหตุการณ์ที่ไม่ได้สนใจ พื้นที่ใต้โค้งเป็นดัชนีที่ใช้ในการบ่งชี้ความถูกต้องหรือความน่าเชื่อถือของตัวแบบ ตัวแบบใดที่มีพื้นที่ใต้โค้งมากที่สุดจะถือว่าเป็นตัวแบบที่ดีที่สุด

2.10 งานวิจัยที่เกี่ยวข้อง

Das. et al. (2020) ทำนายการเสียชีวิตของผู้ป่วยโควิด-19 ในประเทศเกาหลีใต้ จากชุดข้อมูลสาธารณะที่เผยแพร่โดยศูนย์ควบคุมและป้องกันโรค ประเทศเกาหลีใต้ มีจำนวน 3,524 ราย ตั้งแต่วันที่ 20 มกราคม ถึง 30 พฤษภาคม พ.ศ. 2563 โดยใช้ตัวแบบวิธีการเรียนรู้ของเครื่อง (Machine Learning) ซึ่งได้แก่วิธีการวิเคราะห์การถดถอยลอจิสติก (Logistic Regression) วิธีการซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine) วิธีเพื่อนบ้านใกล้สุด K ตัว (K-Nearest Neighbor) วิธีป่าสุ่ม (Random Forest) และวิธีเอ็กซ์ทรีมกราดิเอนท์บูตติ้ง (Extreme Gradient Boosting) พบว่าวิธีการวิเคราะห์การถดถอยลอจิสติก มีประสิทธิภาพเหนือกว่าวิธีอื่นโดยมีค่าความแม่นยำ 0.97 และค่าพื้นที่ใต้โค้ง 0.83

Sánchez-Montañés. et al. (2021) วิเคราะห์ข้อมูลตัวอย่างผู้ป่วยที่เข้ารับการรักษาที่โรงพยาบาลด้วยโรคโควิด-19 ในเมืองมาดริด ประเทศสเปน จากชุดข้อมูลในระบบ HER ซึ่งประกอบไปด้วยข้อมูลผู้ป่วยจำนวน 2,307 ราย ตัวแปรอิสระจำนวน 29 ตัวแปร โดยแบ่งข้อมูลออกเป็น 2 ส่วน โดยชุดข้อมูลเรียนรู้ (Training Set) 70% ใช้ในการสร้างตัวแบบ และชุดข้อมูลทดสอบ (Test Set) 30% ใช้ในการทดสอบประสิทธิภาพตัวแบบ ด้วยวิธีการวิเคราะห์การอยู่รอด (Survival Analysis) วิธีการวิเคราะห์การถดถอยลอจิสติก ต้นไม้ตัดสินใจ (Decision Tree) วิธีป่าสุ่ม และวิธีนาอิวเบส (Naïve Bayes) ผลลัพธ์แสดงให้เห็นว่า วิธีการวิเคราะห์การถดถอยลอจิสติก มีประสิทธิภาพมากที่สุด โดยมีค่าพื้นที่ใต้โค้ง 0.89 โดยตัวแปรอิสระที่ส่งผลกระทบต่ออัตราการเสียชีวิตอย่างโดดเด่นได้แก่ อายุ ปริมาณออกซิเจนในเลือด และการที่ผู้ป่วยเข้ารับการรักษาตัวในแผนกฉุกเฉิน

Ferreira. et al. (2021) วิเคราะห์ข้อมูลผู้ป่วยโควิด-19 จาก Directorate-General of Health of Portugal โดยใช้วิธีการวิเคราะห์การถดถอยลอจิสติก วิธีต้นไม้ตัดสินใจ และวิธีการเรียนรู้เชิงลึก (Deep Learning) เพื่อสร้างตัวแบบในการทำนายประเภทการจำหน่ายผู้ป่วยว่า เสียชีวิต หรือรอดชีวิต โดยใช้โรคร่วม อายุ และอาการของผู้ป่วยพร้อมกับใช้เทคนิคการสุ่มเพิ่มจำนวนข้อมูล พบว่าตัวแบบทำนายที่สร้างด้วยวิธีต้นไม้ตัดสินใจ มีค่าความแม่นยำมากที่สุด 0.90 และค่าเรียกคืนเท่ากับ 0.95

Li et al. (2020) เป็นการศึกษาแบบย้อนหลัง มีวัตถุประสงค์เพื่อสร้างตัวแบบในการทำนายการรับเข้าไอซียู และการเสียชีวิตในโรงพยาบาลของผู้ป่วยโควิด-19 โดยใช้วิธีการเรียนรู้เชิงลึกและ Risk-Score System ชุดข้อมูลประกอบด้วยข้อมูลทางคลินิกจากผู้ป่วย 1,108 ราย โดย 837 ราย เข้ารับการรักษาทั่วไป และผู้ป่วย 271 ราย เข้ารับการรักษาในไอซียู จากจำนวนผู้ป่วยทั่วไป 837 ราย (ออกจากโรงพยาบาลแล้ว 772 ราย และเสียชีวิต 65 ราย) มีผู้ป่วยรับการรักษาในห้องไอซียู 271 ราย (ยังอยู่ในไอซียู 86 ราย ออกจากโรงพยาบาลแล้ว 108 ราย และเสียชีวิต 77 ราย) ผู้ป่วยทั้งหมดเข้ารับการรักษาที่โรงพยาบาล Stony Brook University ในนิวยอร์ก การคัดเลือกตัวแปรอิสระโดยขั้นตอนวิธีป่าสุ่ม จากนั้นตัวทำนายสูงสุดถูกนำไปสร้างตัวแบบที่ประกอบด้วยชั้นซ่อน 5 ชั้นที่เชื่อมต่อกันอย่างเต็มที่ สำหรับชั้นซ่อนจะใช้ฟังก์ชันเรกติไฟด์ลินเนียร์เป็นฟังก์ชันกระตุ้น ในขณะที่ฟังก์ชันซิกมอยด์นั้น ใช้สำหรับชั้นข้อมูลออก โดยประสิทธิภาพตัวแบบในการทำนายการรับเข้าไอซียู มีความแม่นยำเท่ากับ 0.721 ค่าพื้นที่ใต้โค้งเท่ากับ 0.728 และตัวแบบในการทำนายโอกาสการเสียชีวิต มีความแม่นยำเท่ากับ 0.853 ค่าพื้นที่ใต้โค้งเท่ากับ 0.844 โดยตัวแปรสำคัญสำหรับการเสียชีวิตคือ Procalcitonin, Lactate dehydrogenase, C-reactive protein, cardiac troponin, อายุ และออกซิเจนในกระแสเลือด อย่างไรก็ตาม โรคร่วมไม่ได้เป็นตัวทำนายที่สำคัญของการเข้ารับการรักษาในห้องไอซียู และการเสียชีวิต

Agbelusi and Olayemi (2020) ได้พัฒนาตัวแบบในการทำนายอัตราการเสียชีวิตของผู้ป่วยติดเชื้อไวรัสโคโรนาในประเทศไนจีเรีย โดยใช้เทคนิคการทำเหมืองข้อมูล 4 วิธีได้แก่ วิธีโครงข่ายประสาทเทียมแบบหลายลำดับชั้น วิธีต้นไม้ตัดสินใจ วิธีต้นไม้ตัดสินใจ และกฎการตัดสินใจ (Decision Rule) ผลการศึกษาแสดงให้เห็นว่าวิธีโครงข่ายประสาทเทียมแบบหลายลำดับชั้น มีประสิทธิภาพในการทำนายสูงที่สุด โดยมีค่าความแม่นยำอยู่ที่ 85%

Khan. et al. (2021) ได้นำตัวแบบวิธีการเรียนรู้เชิงลึกที่มีการกำหนดชั้นซ่อนทั้งหมด 6 ชั้น โดยมีฟังก์ชันเรกติไฟด์ลินเนียร์เป็นฟังก์ชันกระตุ้น และฟังก์ชันซิกมอยด์ เป็นฟังก์ชันกระตุ้นในชั้นท้ายสุด รวมไปถึงตัวแบบวิธีการเรียนรู้ของเครื่อง ได้แก่ ต้นไม้ตัดสินใจ วิธีการวิเคราะห์การถดถอยลอจิสติก วิธีป่าสุ่ม วิธีเอ็กซ์ทริมกาเดียนบูตตัง และวิธีเพื่อนบ้านใกล้สุด K ตัว เพื่อใช้ทำนายโอกาสเสียชีวิตในผู้ป่วยโควิด-19 โดยใช้ข้อมูลผู้ป่วยที่รับการรักษายืนยันว่าติดเชื้อไวรัสโคโรนาจากทั้งหมด 146 ประเทศพบว่า ตัวแบบวิธีการเรียนรู้เชิงลึกมีประสิทธิภาพในการทำนายสูงที่สุด ด้วยค่าความแม่นยำ 0.97

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Moulaei. et al. (2022) ได้ศึกษาผู้ป่วยโควิด-19 จำนวน 1,500 ราย (รอดชีวิต 1,386 ราย เสียชีวิต 114 ราย) ของโรงพยาบาล Ayatollah Taleghani เมือง Abadan ประเทศอิหร่าน โดยใช้ขั้นตอนวิธีการเรียนรู้ของเครื่อง ในการสร้างตัวแบบทำนายโอกาสในการเสียชีวิตของผู้ป่วยโควิด-19 พบว่า จากตัวแปรทำนายทั้งหมด 38 ตัวแปร ภาวะหายใจลำบาก การเข้ารับการรักษาในห้องไอซียู และการได้รับการบำบัดด้วยออกซิเจน เป็นตัวแปรทำนาย 3 อันดับแรกที่ส่งผลต่อโอกาสการเสียชีวิตของผู้ป่วยโควิด-19 มากที่สุด และวิธีป่าสุ่ม มีประสิทธิภาพสูงที่สุดด้วย ด้วยค่าความแม่นยำเท่ากับ 95.03%

Patsadu. et al. (2020) ได้นำเสนอตัวแบบในการตรวจหาผู้ป่วยมะเร็งระดับ และประเมินระดับความผิดปกติของตับ โดยใช้เทคนิคในการจำแนกกลุ่มซึ่งประกอบด้วยวิธีโครงข่ายประสาทเทียมแบบหลายลำดับชั้น วิธีการวิเคราะห์การถดถอยลอจิสติก และวิธีป่าสุ่ม สร้างตัวแบบด้วยชุดข้อมูลผู้ป่วยโรคตับในประเทศอินเดีย ผลการทดลองแสดงให้เห็นว่า วิธีป่าสุ่ม มีความแม่นยำสูงที่สุด 76.83% ในการตรวจหาผู้ป่วยมะเร็งระดับ และวิธีการวิเคราะห์การถดถอยลอจิสติก สามารถประเมินระดับความผิดปกติของตับด้วย DeRitis Ratio มีความแม่นยำที่สุด 91.46%

Pourhomayoun and Shakibi (2021) ได้พัฒนาตัวแบบในการทำนายโอกาสเสียชีวิตของผู้ป่วยโควิด-19 โดยใช้วิธีการเรียนรู้ของเครื่อง ด้วยชุดข้อมูลผู้ป่วยโควิด-19 ที่ได้รับการยืนยันจากห้องปฏิบัติการมากกว่า 2,670,000 ราย จาก 146 ประเทศทั่วโลก ในการสร้างตัวแบบในการทำนายโอกาสเสียชีวิตของผู้ป่วยโควิด-19 ด้วยวิธีการซัพพอร์ตเวกเตอร์แมชชีน วิธีโครงข่ายประสาทเทียม (Artificial Neural Networks) วิธีป่าสุ่ม วิธีต้นไม้ตัดสินใจ วิธีการวิเคราะห์การถดถอยลอจิสติก และวิธีเพื่อนบ้านใกล้ที่สุด พบว่าวิธีโครงข่ายประสาทเทียม มีประสิทธิภาพสูงที่สุด โดยมีค่าความแม่นยำเท่ากับ 89.98%

Muhammad. et al. (2020) ได้ทำการสร้างตัวแบบในการทำนายโอกาสการติดเชื้อโคโรนาไวรัส จากชุดข้อมูลการระบาดวิทยาของผู้ป่วยโควิด-19 ทั้งที่ตรวจพบเชื้อและไม่พบเชื้อ ในประเทศเม็กซิโก โดยจำแนกเป็นผู้ป่วยที่ตรวจพบเชื้อ จำนวนทั้งสิ้น 263,007 ราย ตัวแปรอิสระ 41 ตัวแปร สร้างตัวแบบในการทำนายด้วยวิธีการวิเคราะห์การถดถอยลอจิสติก วิธีต้นไม้ตัดสินใจ วิธีการซัพพอร์ตเวกเตอร์แมชชีน วิธีนาอีฟเบส และโครงข่ายประสาทเทียม โดยใช้ชุดข้อมูลเรียนรู้ในการสร้างตัวแบบ 80% และตรวจสอบประสิทธิภาพการทำนายของแบบจำลองกับชุดข้อมูลทดสอบ 20% โดยพิจารณาจากค่าความแม่นยำแสดงให้เห็นว่า วิธีต้นไม้ตัดสินใจ มีค่าความแม่นยำสูงที่สุด 94.99%

บทที่ 3

วิธีการดำเนินงานวิจัย

จากการศึกษาและทบทวนวรรณกรรมที่เกี่ยวข้องทั้งในประเทศ และต่างประเทศพบว่า งานวิจัยที่ศึกษาเกี่ยวกับการสร้างตัวแบบในการทำนายโอกาสการเสียชีวิตของผู้ป่วยโควิด-19 โดยส่วนมากสร้างตัวแบบในการทำนายด้วยวิธีการวิเคราะห์การถดถอยลอจิสติก วิธีต้นไม้ตัดสินใจ และวิธีโครงข่ายประสาทเทียมแบบหลายลำดับชั้น และเมื่อเปรียบเทียบประสิทธิภาพของตัวแบบโดยพิจารณาจากค่าความแม่นยำ พบว่า ทั้ง 3 วิธีมีความแม่นยำสูงที่สุด เมื่อเทียบกับวิธีการจำแนกกลุ่มวิธีอื่น ผู้วิจัยจึงเลือกวิธีจำแนกกลุ่มทั้ง 3 วิธีดังกล่าว ในการนำไปสร้างและเปรียบเทียบประสิทธิภาพของตัวแบบสำหรับการวิเคราะห์หาปัจจัยที่สัมพันธ์ต่อการเสียชีวิตของผู้ป่วยโควิด-19 กรณีศึกษาโรงพยาบาลระยอง เพื่อเปรียบเทียบประสิทธิภาพการทำนายของตัวแบบที่สร้างขึ้นในแต่ละวิธี ตลอดจนศึกษาปัจจัยที่สัมพันธ์ต่อการเสียชีวิตของผู้ป่วยโควิด-19

3.1 การเก็บรวบรวมข้อมูล

ผู้วิจัยได้ดำเนินการเก็บรวบรวมข้อมูลผู้ป่วยโรคโควิด-19 ที่เข้ารับการรักษาในโรงพยาบาลระยอง และจำหน่ายออกจากโรงพยาบาลตั้งแต่วันที่ 1 เมษายน 2564 ถึงวันที่ 31 มกราคม 2565 โดยจะมีการเก็บรวบรวมข้อมูลลักษณะประชากรทั่วไป ได้แก่ เพศ อายุ สัญชาติ และดัชนีมวลกาย จากฐานข้อมูลผู้ป่วยในระบบสารสนเทศโรงพยาบาล (Hospital Information Systems: HIS) และเก็บรวบรวมข้อมูลอาการแรกรับของผู้ป่วยแต่ละราย ได้แก่ ผลภาพรังสีทรวงอก ความดันโลหิต อัตราการเต้นของหัวใจ ปริมาณออกซิเจนในเลือด อุณหภูมิร่างกาย อัตราการหายใจ ข้อมูลด้านการรักษาพยาบาลทั้งในเรื่องของการใช้ยา จำนวนวันนับตั้งแต่วันที่ตรวจพบเชื้อจนถึงวันจำหน่ายการรักษา การใช้เครื่องช่วยหายใจ รวมไปถึงโรคประจำตัว และโรคแทรกซ้อนที่แพทย์ผู้ให้การรักษาสรุปไว้ในเวชระเบียนผู้ป่วย รวมไปถึงประเภทการจำหน่ายของผู้ป่วยทุกราย

3.2 ตัวแปรที่ใช้ในการวิจัย

งานวิจัยนี้ จำแนกตัวแปรอิสระออกเป็นปัจจัยด้านต่าง ๆ ทั้งหมด 4 ด้าน ได้แก่ ปัจจัยส่วนบุคคล (ตารางที่ 3.1) ปัจจัยด้านสุขภาพ (ตารางที่ 3.2) ปัจจัยอาการแรกรับ (ตารางที่ 3.3) และปัจจัยด้านการรักษา (ตารางที่ 3.4) โดยที่ตัวแปรตามคือ ประเภทการจำหน่ายของผู้ป่วยโดยจำแนกเป็นรอดชีวิต และเสียชีวิต (ตารางที่ 3.5)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.1 ตัวแปรอิสระปัจจัยส่วนบุคคล

ตัวแปรอิสระ	ความหมาย	สเกลข้อมูล
X_1	เพศ	นามบัญญัติ
X_2	อายุ (หน่วย : ปี)	อัตราส่วน
X_3	สัญชาติ	นามบัญญัติ
X_4	ดัชนีมวลกาย (หน่วย : เมตร/กิโลกรัม ²)	อัตราส่วน

ตารางที่ 3.2 ปัจจัยด้านสุขภาพ

ตัวแปรอิสระ	ความหมาย	สเกลข้อมูล
X_5	โรคหลอดเลือดสมอง	นามบัญญัติ
X_6	โรคปอดอุดกั้นเรื้อรัง	นามบัญญัติ
X_7	โรคเบาหวาน	นามบัญญัติ
X_8	โรคหัวใจ	นามบัญญัติ
X_9	โรคเอดส์	นามบัญญัติ
X_{10}	โรคไต	นามบัญญัติ
X_{11}	โรคตับ	นามบัญญัติ
X_{12}	โรคปอด	นามบัญญัติ
X_{13}	โรคมะเร็ง	นามบัญญัติ
X_{14}	วัคซีนแอสตราเซนเก้า (หน่วย : เข็ม)	อัตราส่วน
X_{15}	วัคซีนซิโนแวค (หน่วย : เข็ม)	อัตราส่วน
X_{16}	วัคซีนไฟเซอร์ (หน่วย : เข็ม)	อัตราส่วน
X_{17}	วัคซีนซิโนฟาร์ม (หน่วย : เข็ม)	อัตราส่วน

ตารางที่ 3.3 ตัวแปรอิสระปัจจัยอาการแรกเริ่ม

ตัวแปรอิสระ	ความหมาย	สเกลข้อมูล
X_{18}	ภาพรังสีทรวงอกรูปแบบ Patchy Opacity	นามบัญญัติ
X_{19}	ภาพรังสีทรวงอกรูปแบบ Ground Glass Opacity	นามบัญญัติ
X_{20}	ภาพรังสีทรวงอกรูปแบบ Reticulonodular Opacity	นามบัญญัติ
X_{21}	การกระจายตัวของรอยโรคในปอดเด่นที่ Peripheral	นามบัญญัติ
X_{22}	การกระจายตัวของรอยโรคเด่นที่ Perihilar	นามบัญญัติ
X_{23}	การกระจายตัวของรอยโรคเด่นที่ปอดข้างซ้าย	นามบัญญัติ
X_{24}	การกระจายตัวของรอยโรคเด่นที่ปอดข้างขวา	นามบัญญัติ
X_{25}	การกระจายตัวของรอยโรคเด่นที่ปอดทั้ง 2 ข้าง	นามบัญญัติ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.3 ตัวแปรอิสระปัจจัยอาการแรกเริ่ม (ต่อ)

ตัวแปรอิสระ	ความหมาย	สเกลข้อมูล
X_{26}	การกระจายตัวของรอยโรคเด่นที่ปอดส่วนล่าง	นามบัญญัติ
X_{27}	การกระจายตัวของรอยโรคเด่นที่ปอดส่วนบน	นามบัญญัติ
X_{28}	การกระจายตัวของรอยโรคเด่นที่ปอดส่วนกลาง	นามบัญญัติ
X_{29}	ความดันโลหิตสูงสุดขณะหัวใจห้องล่างบีบตัว (หน่วย : มิลลิเมตรปรอท)	อัตราส่วน
X_{30}	ความดันโลหิตที่ต่ำสุดขณะหัวใจห้องล่างคลายตัว (หน่วย : มิลลิเมตรปรอท)	อัตราส่วน
X_{31}	อัตราการหายใจ (หน่วย : ครั้ง/นาที)	อัตราส่วน
X_{32}	อัตราการเต้นของหัวใจ (หน่วย : ครั้ง/นาที)	อัตราส่วน
X_{33}	ปริมาณออกซิเจนในเลือด (หน่วย : เปอร์เซ็นต์)	อัตราส่วน
X_{34}	อุณหภูมิร่างกาย (หน่วย : องศาเซลเซียส)	อันตรภาค

ตารางที่ 3.4 ตัวแปรอิสระปัจจัยด้านการรักษา

ตัวแปรอิสระ	ความหมาย	สเกลข้อมูล
X_{35}	เครื่องช่วยหายใจแบบใส่ท่อช่วยหายใจ	นามบัญญัติ
X_{36}	เครื่องช่วยหายใจแบบไม่ใส่ท่อช่วยหายใจ	นามบัญญัติ
X_{37}	การให้ออกซิเจนผ่านหน้ากาก หรือสายออกซิเจน	นามบัญญัติ
X_{38}	ยาฟิวโรซิมิด	นามบัญญัติ
X_{39}	ยาคาซิวิแมบ และยาอิมตีวิแมบ	นามบัญญัติ
X_{40}	ยาเรมเดซิเวียร์	นามบัญญัติ
X_{41}	ยาฟ้าทลายโจร	นามบัญญัติ
X_{42}	ยาแพ็กซ์โลวิด	นามบัญญัติ
X_{43}	ยาเดกซาเมทาโซน	นามบัญญัติ
X_{44}	ยาเมทิลเพรดนิโซโลน	นามบัญญัติ
X_{45}	ยาไฮโดรคอร์ติโซน	นามบัญญัติ
X_{46}	จำนวนวันนับตั้งแต่วันตรวจพบเชื้อจนถึงวันจำหน่าย การรักษา (หน่วย : วัน)	อัตราส่วน

ตารางที่ 3.5 ตัวแปรตาม

ตัวแปรตาม	ความหมาย	สเกลข้อมูล
y	ประเภทการจำหน่าย	นามบัญญัติ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.3 การเตรียมข้อมูล

3.3.1 การทำความสะอาดข้อมูล (Data Cleaning)

เป็นขั้นตอนการจัดการกับข้อมูลที่สูญหาย โดยใช้คำสั่ง `dropna` โมดูล ของ `pandas` โปรแกรมภาษาไพธอน (Python Programming Language) พบว่าข้อมูลไม่พบค่าสูญหาย

3.3.2 การแปลงข้อมูล (Data Transform)

เนื่องจากตัวแปรอิสระบางตัว และตัวแปรตามเป็นตัวแปรเชิงคุณภาพ มีความจำเป็นต้องทำการแปลงข้อมูล ก่อนที่จะนำไปสร้างตัวแบบในการทำนาย ผู้วิจัยจึงใช้วิธีการแทนค่ากลุ่มโดยใช้คำสั่ง `LabelEncoder` ใน `preprocessing` โมดูล ของ `scikit-learn` ในโปรแกรมภาษาไพธอน ดังแสดงตาราง 3.6 และตาราง 3.7

ตารางที่ 3.6 การแปลงข้อมูลตัวแปรอิสระที่ใช้ในการวิจัย

ตัวแปรอิสระ	ความหมาย	แปลงข้อมูล
X_1	เพศ	0 = หญิง 1 = ชาย
X_3	สัญชาติ	0 = ต่างชาติ 1 = ไทย
X_5	โรคหลอดเลือดสมอง	0 = ไม่เป็น 1 = เป็น
X_6	โรคปอดอุดกั้นเรื้อรัง	0 = ไม่เป็น 1 = เป็น
X_7	โรคเบาหวาน	0 = ไม่เป็น 1 = เป็น
X_8	โรคหัวใจ	0 = ไม่เป็น 1 = เป็น
X_9	โรคเอดส์	0 = ไม่เป็น 1 = เป็น
X_{10}	โรคไต	0 = ไม่เป็น 1 = เป็น
X_{11}	โรคตับ	0 = ไม่เป็น 1 = เป็น
X_{12}	โรคปอด	0 = ไม่เป็น 1 = เป็น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับกรใช้ภายในเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้ทำไปใช้ประโยชน์อื่นใด
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.6 การแปลงข้อมูลตัวแปรอิสระที่ใช้ในการวิจัย (ต่อ)

ตัวแปรอิสระ	ความหมาย	แปลงข้อมูล
X_{13}	โรคมะเร็ง	0 = ไม่เป็น 1 = เป็น
X_{18}	ภาพรังสีทรวงอกรูปแบบ Patchy Opacity	0 = ไม่พบ 1 = พบ
X_{19}	ภาพรังสีทรวงอกรูปแบบ Ground Glass Opacity	0 = ไม่พบ 1 = พบ
X_{20}	ภาพรังสีทรวงอกรูปแบบ Reticulonodular Opacity	0 = ไม่พบ 1 = พบ
X_{21}	การกระจายตัวของรอยโรคในปอดเด่นที่ Peripheral	0 = ไม่พบ 1 = พบ
X_{22}	การกระจายตัวของรอยโรคเด่นที่ Perihilar	0 = ไม่พบ 1 = พบ
X_{23}	การกระจายตัวของรอยโรคเด่นที่ปอดข้างซ้าย	0 = ไม่พบ 1 = พบ
X_{24}	การกระจายตัวของรอยโรคเด่นที่ปอดข้างขวา	0 = ไม่พบ 1 = พบ
X_{25}	การกระจายตัวของรอยโรคเด่นที่ปอดทั้ง 2 ข้าง	0 = ไม่พบ 1 = พบ
X_{26}	การกระจายตัวของรอยโรคเด่นที่ปอดส่วนล่าง	0 = ไม่พบ 1 = พบ
X_{27}	การกระจายตัวของรอยโรคเด่นที่ปอดส่วนบน	0 = ไม่พบ 1 = พบ
X_{28}	การกระจายตัวของรอยโรคเด่นที่ปอดส่วนกลาง	0 = ไม่พบ 1 = พบ
X_{35}	เครื่องช่วยหายใจแบบใส่ท่อช่วยหายใจ	0 = ไม่ใช่ 1 = ใช่
X_{36}	เครื่องช่วยหายใจแบบไม่ใส่ท่อช่วยหายใจ	0 = ไม่ใช่ 1 = ใช่
X_{37}	การให้ออกซิเจนผ่านหน้ากาก หรือสายออกซิเจน	0 = ไม่ใช่ 1 = ใช่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.6 การแปลงข้อมูลตัวแปรอิสระที่ใช้ในการวิจัย (ต่อ)

ตัวแปรอิสระ	ความหมาย	แปลงข้อมูล
X_{38}	ยาฟาวิพิราเวียร์	0 = ไม่ใช่ 1 = ใช่
X_{39}	ยาคาซิรีวิแมบ และยาอิมติวิแมบ	0 = ไม่ใช่ 1 = ใช่
X_{40}	ยาเรมเดซิเวียร์	0 = ไม่ใช่ 1 = ใช่
X_{41}	ยาฟ้าทลายโจร	0 = ไม่ใช่ 1 = ใช่
X_{42}	ยาแพ็กซ์โลวิด	0 = ไม่ใช่ 1 = ใช่
X_{43}	ยาเดกซาเมทาโซน	0 = ไม่ใช่ 1 = ใช่
X_{44}	ยาเมทิลเพรดนิโซโลน	0 = ไม่ใช่ 1 = ใช่
X_{45}	ยาไฮโดรคอร์ติโซน	0 = ไม่ใช่ 1 = ใช่

ตารางที่ 3.7 การแปลงข้อมูลตัวแปรตามที่ใช้ในการวิจัย

ตัวแปรอิสระ	ความหมาย	แปลงข้อมูล
y	ประเภทการจำหน่าย	0 = รอดชีวิต 1 = เสียชีวิต

3.3.3 การแปลงข้อมูลให้อยู่ในค่ามาตรฐานแซด

การแปลงข้อมูลให้อยู่ในค่ามาตรฐานแซด เพื่อปรับให้ค่าของข้อมูลในตัวแปรอิสระแต่ละตัว ให้มีค่าเฉลี่ยเท่ากับ 1 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0 โดยใช้คำสั่ง StandardScaler ใน preprocessing โมดูล ของ scikit-learn ในโปรแกรมภาษาไพธอน

3.3.4 การสุ่มเพิ่มจำนวนข้อมูล

เพื่อแก้ไขปัญหาข้อมูลไม่สมดุล (Imbalanced Data) ดังนั้นผู้วิจัยจึงนำวิธีสุ่มเพิ่มจำนวนข้อมูลด้วยเทคนิค SMOTE มาใช้ปรับสมดุลข้อมูล เป็นการเพิ่มจำนวนข้อมูลกลุ่มน้อย ซึ่งก็คือจำนวนข้อมูลของผู้ป่วยโรคโควิด-19 มีประเภทการจำหน่ายเป็นเสียชีวิต ให้มีอัตราส่วนเท่ากับข้อมูลกลุ่มมาก ซึ่งก็คือจำนวนข้อมูลของผู้ป่วยโรคโควิด-19 มีสถานการณจำหน่ายเป็นรอดชีวิต โดยใช้คำสั่ง SMOTE ใน over_sampling โมดูล ของ Imbalanced-learn ในโปรแกรมภาษาไพธอน

3.3.5 การแบ่งข้อมูลเป็นชุดข้อมูลฝึกหัด และชุดข้อมูลทดสอบ

ในงานวิจัยนี้จะทำการแบ่งข้อมูลจากการสุ่มเพิ่มจำนวนข้อมูลทั้งหมดออกเป็น ชุดข้อมูลเรียนรู้ร้อยละ 70 และชุดข้อมูลทดสอบร้อยละ 30 โดยใช้คำสั่ง train_test_split ใน selection โมดูล ของ scikit-learn ในโปรแกรมภาษาไพธอน

3.4 การวิเคราะห์การจำแนกกลุ่ม

ในงานวิจัยนี้ ได้มีการนำวิธีการจำแนกกลุ่มของชุดข้อมูลมาเปรียบเทียบกัน ทั้งหมด 3 วิธี ประกอบไปด้วย

3.4.1 วิธีการวิเคราะห์การถดถอยลอจิสติก

วิธีการวิเคราะห์การถดถอยลอจิสติก โดยทำการตรวจสอบว่าตัวแปรอิสระไม่มีความสัมพันธ์กันหรือไม่เกิดปัญหาความสัมพันธ์เชิงเส้นพหุ โดยพิจารณาจากค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน และคัดเลือกตัวแปรอิสระด้วยวิธี Backward Stepwise โดยเป็นการคัดเลือกตัวแปรอิสระที่สามารถอธิบายความผันแปรของตัวแปรตามได้อย่างมีนัยสำคัญทางสถิติที่ระดับนัยสำคัญ 0.05 และตรวจสอบความเหมาะสมของตัวแบบ โดยพิจารณาจากค่าความเป็นไปได้ทดสอบสัมประสิทธิ์การถดถอย (Regression Coefficient) ด้วยสถิติทดสอบของวีลด์ และพิจารณาค่าสถิติทดสอบระดับความสัมพันธ์ McFadden's R square

3.4.2 วิธีต้นไม้ตัดสินใจ

กำหนดให้ค่าเอนโทรปี สูงสุดเป็นลักษณะเฉพาะเริ่มต้น และเลือกลักษณะ เฉพาะถัดไปตามค่าเอนโทรปี น้อยลงตามลำดับ

3.4.3 วิธีโครงข่ายประสาทเทียมแบบหลายลำดับชั้น

โดยกำหนดค่าอัตราการเรียนรู้ (Learning Rate) เป็น 0.1 ค่าโมเมนตัม (Momentum) เป็น 0.9 จำนวนรอบการสอน (Training Time) 1,000 รอบ ชั้นซ่อนจำนวน 3 ชั้น โดยชั้นแรกมีจำนวนโหนด 500 โหนด ชั้นถัดมามีจำนวน 1000 โหนด และชั้นสุดท้ายมีจำนวน 500 โหนด ใช้ฟังก์ชันเรกติไฟต์ลินีเยียนิต เป็นฟังก์ชันกระตุ้นในทุกชั้นซ่อน และฟังก์ชันซิกมอยด์เป็นฟังก์ชันกระตุ้นในชั้นข้อมูลออก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.5 ค่าความสำคัญของตัวแปร (Feature Importance)

ค่าความสำคัญของตัวแปรเป็นการศึกษาว่าตัวแปรอิสระตัวใดที่ส่งผลต่อตัวแบบในการทำนายมาก หรือน้อยกว่ากัน โดยในวิธีการวิเคราะห์การถดถอยลอจิสติก จะพิจารณาจากค่าสัมประสิทธิ์การถดถอย วิธีต้นไม้ตัดสินใจ พิจารณาจากค่าเอนโทรปี และวิธีโครงข่ายประสาทเทียม คำนวณค่าความสำคัญจากวิธี Permutation Importance โดยใช้คำสั่ง `permutation_importance` ใน `inspection` โมดูลของ `scikit-learn` ในโปรแกรมภาษาไพธอน โดยหลักการทำงานคือ พิจารณาตัวแปรอิสระทีละตัว โดยทำการสลับตำแหน่งของข้อมูลในตัวแปรอิสระที่พิจารณา โดยที่ตำแหน่งของตัวแปรอิสระอื่น ๆ ยังคงเดิม แล้วนำไปสร้างตัวแบบในการทำนาย ถ้าประสิทธิภาพของตัวแบบลดลงจากตัวแบบก่อนหน้าที่ไม่มีการสลับตำแหน่งตัวแปรอิสระแสดงว่า ตัวแปรอิสระนั้นมีความสำคัญมาก โดยประสิทธิภาพของตัวแบบที่ลดลงมากกว่า แปลว่าตัวแปรอิสระนั้นมีความสำคัญมากกว่า

3.6 การเปรียบเทียบประสิทธิภาพของตัวแบบ

การเปรียบเทียบประสิทธิภาพของตัวแบบที่เหมาะสม โดยนำแบบจำลองที่สร้างขึ้นด้วยวิธีการวิเคราะห์การถดถอยลอจิสติก วิธีต้นไม้การตัดสินใจ และวิธีโครงข่ายประสาทเทียมแบบหลายลำดับชั้น นำมาเปรียบเทียบประสิทธิภาพ โดยพิจารณาจากค่าความแม่นยำ ค่าความเที่ยง ค่าเรียกคืน ค่าความถ่วงดุล และค่าพื้นที่ใต้โค้ง โดยถ้าค่ายิ่งมากแสดงว่า ตัวแบบมีประสิทธิภาพในการทำนายสูง

3.7 เครื่องมือที่ใช้ในการวิจัย

ในขั้นตอนวิธีการสร้างตัวแบบ ผู้วิจัยดำเนินการโดยใช้โปรแกรมภาษาไพธอน รันบนระบบปฏิบัติการ Windows 10 Pro ในเครื่องคอมพิวเตอร์โน้ตบุ๊ก Dell inspiron 5558 Intel® Core™ i3-4005U RAM 4.00 GB 64-bit โดยติดตั้งไลบรารี (Library) ที่จำเป็นในการสร้างตัวแบบดังนี้

- 1) Pandas เป็นไลบรารีสำหรับการจัดการข้อมูล (Data Manipulation)
- 2) NumPy เป็นไลบรารีสำหรับคณิตศาสตร์และสถิติ (Mathematics and Statistics)
- 3) Scikit-Learn เป็นไลบรารีสำหรับสร้างตัวแบบการเรียนรู้ของเครื่อง
- 4) Matplotlib, Seaborn และ ggplot เป็นไลบรารีสำหรับการแสดงผลข้อมูล (Data Visualization)

ผลการวิจัยและการอภิปรายผล

งานวิจัยนี้ใช้วิธีการสร้างตัวแบบในการทำนายการเสียชีวิตของผู้ป่วยโรคโควิด-19 ซึ่งได้แก่ วิธีการวิเคราะห์การถดถอยลอจิสติก วิธีต้นไม้ตัดสินใจ และวิธีโครงข่ายประสาทเทียมแบบหลายลำดับชั้น เพื่อเปรียบเทียบประสิทธิภาพในการทำนายของตัวแบบพร้อมทั้งวิเคราะห์ปัจจัยที่ปัจจัยที่สัมพันธ์ต่อการเสียชีวิตของผู้ป่วยโควิด-19 โดยรายละเอียดมีดังต่อไปนี้

4.1 ลักษณะทั่วไปของกลุ่มตัวอย่าง

จากการตรวจสอบข้อมูลผู้ป่วยที่ได้รับการวินิจฉัยร่วมว่าเป็นโรคโควิด-19 ที่เข้ารับการรักษาในโรงพยาบาลระยองจำนวนทั้งสิ้น 1,006 ราย จำแนกเป็นชายร้อยละ 45.13 และหญิงร้อยละ 54.87 อายุเฉลี่ย 60 ปี ผู้ป่วยร้อยละ 89.36 มีสัญชาติไทย ที่เหลือเป็นผู้ป่วยชาวต่างชาติ โดยที่ผู้ป่วยทั้งหมดมีดัชนีมวลกายเฉลี่ยอยู่ที่ 26.25 กิโลกรัม/เมตร² (ตารางที่ 4.1)

ตารางที่ 4.1 ตัวแปรอิสระปัจจัยส่วนบุคคล

		ตัวแปรอิสระ	จำนวน (%)
X ₁	เพศ		
	หญิง		552 (54.87%)
	ชาย		454 (45.13%)
X ₂	อายุ (mean ± SD)		55.99 ± 17.24
X ₃	สัญชาติ		
	ไทย		899 (89.36%)
	ต่างชาติ		107 (10.64%)
X ₄	ดัชนีมวลกาย (mean ± SD)		26.25 ± 5.90

จากตารางที่ 4.2 โรคเบาหวานเป็นโรคที่พบมากที่สุดร้อยละ 27.04 รองลงมาคือโรคไตร้อยละ 13.62 โดยโรคปอดเป็นโรคที่พบน้อยที่สุดอยู่ที่ร้อยละ 0.89 ผู้ป่วยส่วนมากได้รับวัคซีนแอสตราเซนเนกา โดยจำแนกเป็นได้รับวัคซีน 1 เข็มร้อยละ 5.27 ได้รับวัคซีน 2 เข็มร้อยละ 1.69

ตารางที่ 4.2 ตัวแปรอิสระปัจจัยด้านสุขภาพ

	ตัวแปรอิสระ	จำนวน (%)
X ₅	โรคหลอดเลือดสมอง	43 (4.27%)
X ₆	โรคปอดอุดกั้นเรื้อรัง	14 (1.39%)
X ₇	โรคเบาหวาน	272 (27.04%)
X ₈	โรคหัวใจ	84 (8.35%)
X ₉	โรคเอดส์	14 (1.39%)
X ₁₀	โรคไต	137 (13.62%)
X ₁₁	โรคตับ	31 (3.08%)
X ₁₂	โรคปอด	9 (0.89%)
X ₁₃	โรคมะเร็ง	19 (1.89%)
X ₁₄	วัคซีนแอสตราเซนเนกา (mean ± SD)	0.09 ± 0.34
X ₁₅	วัคซีนซิโนแวค (mean ± SD)	0.06 ± 0.26
X ₁₆	วัคซีนไฟเซอร์ (mean ± SD)	0.01 ± 0.08
X ₁₇	วัคซีนซิโนฟาร์ม (mean ± SD)	0.01 ± 0.12

ลักษณะภาพรังสีทรวงอกที่พบส่วนใหญ่เป็นแบบ Ground glass opacity ร้อยละ 50.30 รอยโรคมักกระจายตัวอยู่ที่ปอดส่วน peripheral 13.32 โดยพบความผิดปกติของปอดทั้ง 2 ข้างมากที่สุดร้อยละ 54.37 และปอดส่วนล่าง 38.97 ผู้ป่วยมีความดันโลหิตเฉลี่ย 130/79 มิลลิเมตรปรอท อัตราการหายใจเฉลี่ย 22 ครั้ง/นาที อัตราการเต้นของหัวใจเฉลี่ย 92 ครั้ง/นาที ปริมาณออกซิเจนในเลือดเฉลี่ย 95% และอุณหภูมิร่างกายเฉลี่ยอยู่ที่ 36.72 องศาเซลเซียส (ตารางที่ 4.3)

ตารางที่ 4.3 ตัวแปรอิสระปัจจัยด้านแรกรับ

	ตัวแปรอิสระ	จำนวน (%)
X ₁₈	ภาพรังสีทรวงอกแบบ Patchy Opacity	506 (50.30%)
X ₁₉	ภาพรังสีทรวงอกแบบ Ground Glass Opacity	333 (33.10%)
X ₂₀	ภาพรังสีทรวงอกแบบ Reticulonodular Opacity	56 (5.57%)
X ₂₁	การกระจายตัวของรอยโรคในปอดเด่นที่ Peripheral	134 (13.32%)
X ₂₂	การกระจายตัวของรอยโรคเด่นที่ Perihilar	6 (0.60%)
X ₂₃	การกระจายตัวของรอยโรคเด่นที่ปอดข้างซ้าย	125 (12.43%)
X ₂₄	การกระจายตัวของรอยโรคเด่นที่ปอดข้างขวา	111 (11.03%)
X ₂₅	การกระจายตัวของรอยโรคเด่นที่ปอดทั้ง 2 ข้าง	547 (54.37%)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.3 ตัวแปรอิสระปัจจัยด้านแรกรับ (ต่อ)

	ตัวแปรอิสระ	จำนวน (%)
X ₂₆	การกระจายตัวของรอยโรคเด่นที่ปอดส่วนล่าง	392 (38.97%)
X ₂₇	การกระจายตัวของรอยโรคเด่นที่ปอดส่วนบน	73 (7.26%)
X ₂₈	การกระจายตัวของรอยโรคเด่นที่ปอดส่วนกลาง	287 (28.53%)
X ₂₉	ความดันโลหิตสูงสุดขณะหัวใจห้องล่างบีบตัว (mean ± SD)	130.56 ± 25.07
X ₃₀	ความดันโลหิตที่ต่ำสุดขณะหัวใจห้องล่างคลายตัว (mean ± SD)	79.82 ± 15.54
X ₃₁	อัตราการหายใจ (mean ± SD)	22.21 ± 3.91
X ₃₂	อัตราการเต้นของหัวใจ (mean ± SD)	92.38 ± 18.33
X ₃₃	ปริมาณออกซิเจนในเลือด (mean ± SD)	95.02 ± 7.09
X ₃₄	อุณหภูมิร่างกาย (mean ± SD)	36.72 ± 0.67

ตารางที่ 4.4 ตัวแปรอิสระปัจจัยด้านการรักษา

	ตัวแปรอิสระ	จำนวน (%)
X ₃₅	เครื่องช่วยหายใจแบบใส่ท่อช่วยหายใจ	119 (11.83%)
X ₃₆	เครื่องช่วยหายใจแบบไม่ใส่ท่อช่วยหายใจ	135 (13.42%)
X ₃₇	การให้ออกซิเจนผ่านหน้ากาก หรือสายออกซิเจน	198 (19.68%)
X ₃₈	ยาฟิวโรซิมาเวียร์	877 (87.18%)
X ₃₉	ยาคาซิริเวแมบ และยาอิมตีเวแมบ	19 (1.89%)
X ₄₀	ยาเรมเดซิเวียร์	71 (7.06%)
X ₄₁	ยาฟ้าทลายโจร	43 (4.27%)
X ₄₂	ยาแพ็กซ์โลวิด	1 (0.1%)
X ₄₃	ยาเดกซาเมทาโซน	730 (72.56%)
X ₄₄	ยาเมทิลเพรดนิโซโลน	414 (41.15%)
X ₄₅	ยาไฮโดรคอร์ติโซน	25 (2.49%)
X ₄₆	จำนวนวันนับตั้งแต่วันตรวจพบเชื้อจนถึงวันจำหน่ายการรักษา (mean ± SD)	15.52 ± 6.56

ผู้ป่วยส่วนใหญ่ให้ออกซิเจนผ่านหน้ากากหรือสายออกซิเจนร้อยละ 19.68 รองลงมาใช้เครื่องช่วยหายใจแบบไม่ใส่ท่อช่วยหายใจร้อยละ 13.42 และใช้เครื่องช่วยหายใจแบบใส่ท่อช่วยหายใจร้อยละ 11.83 ยาที่พบมากที่สุดที่ใช้ในการรักษาผู้ป่วยคือยาฟิวโรซิมาเวียร์ร้อยละ 87.18 รองลงมาคือยาเดกซาเมทาโซนร้อยละ 72.56 จำนวนวันนับตั้งแต่วันตรวจพบเชื้อจนถึงวันจำหน่ายการรักษาเฉลี่ย 16 วัน (ตารางที่ 4.4)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.5 ตัวแปรตาม

ตัวแปรตาม	จำนวน (%)
ประเภทการจำหน่าย	1,006 (100.00%)
รอดชีวิต	814 (80.91%)
เสียชีวิต	192 (19.09%)

จากตารางที่ 4.5 พบว่าผู้ป่วยที่ได้รับการวินิจฉัยร่วมว่าเป็นโรคโควิด-19 ที่เข้ารับการรักษาในโรงพยาบาลระยองจำนวนทั้งสิ้น 1,006 ราย มีเพียง 192 รายเท่านั้น หรือคิดเป็นร้อยละ 19.09 ที่มีประเภทการจำหน่ายเป็นเสียชีวิต ซึ่งเกิดปัญหาข้อมูลไม่สมดุล ดังนั้นผู้วิจัยจึงนำวิธีสุ่มเพิ่มจำนวนข้อมูลด้วยเทคนิค SMOTE มาใช้ปรับสมดุลข้อมูล เป็นการเพิ่มจำนวนข้อมูลกลุ่มน้อย ซึ่งก็คือจำนวนข้อมูลของผู้ป่วยโรคโควิด-19 มีประเภทการจำหน่ายเป็นเสียชีวิต ก่อนนำเข้าสู่กระบวนการสร้างตัวแบบในการทำนายต่อไป โดยผลการสุ่มเพิ่มจำนวนข้อมูลด้วยเทคนิค SMOTE แสดงในตารางที่ 4.6

ตารางที่ 4.6 ผลการสุ่มเพิ่มจำนวนข้อมูลด้วยเทคนิค SMOTE

ตัวแปรตาม	จำนวน (%)
ประเภทการจำหน่าย	1,152 (100.00%)
รอดชีวิต	576 (50.00%)
เสียชีวิต	576 (50.00%)

จากตารางที่ 4.6 พบว่า หลังจากสุ่มเพิ่มจำนวนข้อมูลด้วยเทคนิค SMOTE จำนวนของข้อมูลกลุ่มน้อย (กลุ่มเสียชีวิต) มีจำนวนเพิ่มขึ้น และมีอัตราส่วนของข้อมูลเท่ากับข้อมูลในกลุ่มมาก (กลุ่มรอดชีวิต) เป็น 576 รายทั้งสองกลุ่ม รวมทั้งสิ้น 1,152 ราย

สำหรับการสร้างตัวแบบทำนายด้วยวิธีการวิเคราะห์การถดถอยลอจิสติก และวิธีโครงข่ายประสาทเทียมแบบหลายลำดับชั้น ผู้วิจัยได้ดำเนินการปรับค่าของตัวแปรอิสระของทุกตัวแปรให้อยู่ในค่ามาตรฐาน นั่นคือ ค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ 1 ยกเว้นวิธีต้นไม้ตัดสินใจที่ไม่ได้ปรับค่าของตัวแปรอิสระให้อยู่ในค่ามาตรฐานเนื่องจาก การปรับค่าของตัวแปรอิสระ จะส่งผลต่อขั้นตอนการเลือกลักษณะเฉพาะที่จะใช้เป็นรากหรือโหนด ในวิธีต้นไม้ตัดสินใจ

สำหรับการวิเคราะห์ข้อมูล จะแบ่งข้อมูลออกเป็น 2 ส่วนดังนี้

ส่วนที่ 1 : ชุดข้อมูลชุดเรียนรู้ เพื่อนำไปสร้างตัวแบบ มีข้อมูล 70% ของชุดข้อมูลจากการสุ่มเพิ่มจำนวนข้อมูลด้วยเทคนิค SMOTE ซึ่งจะได้ข้อมูลในส่วนที่ 1 จำนวน 806 ราย

ส่วนที่ 2 : ชุดข้อมูลทดสอบ เพื่อนำไปทดสอบความถูกต้องของตัวแบบ มีข้อมูล 30% ของชุดข้อมูลจากการสุ่มเพิ่มจำนวนข้อมูลด้วยเทคนิค SMOTE ซึ่งจะได้ข้อมูลในส่วนที่ 2 จำนวน 346 ราย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2 การตรวจสอบความเหมาะสมของตัวแบบด้วยวิธีการวิเคราะห์การถดถอยลอจิสติก

จากเงื่อนไขของวิธีการวิเคราะห์การถดถอยลอจิสติกจะต้องกำจัดตัวแปรอิสระที่มีความสัมพันธ์กันเองก่อน จากค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน (ตารางที่ 4.7) พบว่าไม่มีตัวแปรอิสระที่มีความสัมพันธ์กันเอง ($r > 0.8$) (Stevens, 1996) แสดงว่าข้อมูลชุดนี้สามารถวิเคราะห์การถดถอยลอจิสติกได้เนื่องจากตัวแปรอิสระไม่มีความสัมพันธ์กันมากหรือไม่มีปัญหาปัญหาความสัมพันธ์เชิงเส้นพหุเกิดขึ้น

จากผลการวิเคราะห์ตัวแบบการถดถอยลอจิสติกพบว่าตัวแปรอิสระ ที่มีค่าสัมประสิทธิ์เป็นบวก ซึ่งหมายความว่า การที่ผู้ป่วยที่มีอายุ (X_2) สูงขึ้นเป็นปัจจัยสำคัญที่สุดที่สัมพันธ์ต่อการเสียชีวิตของผู้ป่วยโควิด-19 กล่าวคือเมื่อผู้ป่วยมีอายุสูงขึ้น ส่งผลให้โอกาสที่ผู้ป่วยจะเสียชีวิตเพิ่มมากขึ้นเช่นกัน ปัจจัยที่มีความสำคัญรองลงมาคือ ผู้ป่วยที่มีสัญชาติไทย (X_3) มีโอกาสเสียชีวิตสูงกว่าผู้ป่วยที่เป็นชาวต่างชาติ รวมไปถึงการใช้เครื่องช่วยหายใจแบบใส่ท่อช่วยหายใจ (X_{35}) ผู้ป่วยที่มีโรคไต (X_{10}) เป็นโรคประจำตัว ผู้ป่วยที่ได้รับยาเมทิลเพรดนิโซโลน (X_{44}) ผู้ป่วยที่มีโรคหัวใจ (X_8) เป็นโรคประจำตัว ผู้ป่วยที่ใช้เครื่องช่วยหายใจแบบไม่ใส่ท่อช่วยหายใจ (X_{36}) ผู้ป่วยที่มีโรคมะเร็ง (X_{13}) เป็นโรคประจำตัว ผู้ป่วยที่มีความผิดปกติของภาพรังสีทรวงอกรูปแบบ Patchy Opacity (X_{18}) อัตราการหายใจ (X_{31}) และผู้ป่วยที่ได้รับยาไฮโดรคอร์ติโซน (X_{45}) ส่งผลให้โอกาสที่ผู้ป่วยจะเสียชีวิตเพิ่มมากขึ้นตามลำดับ ในขณะที่ตัวแปรอิสระที่มีค่าสัมประสิทธิ์เป็นลบ ซึ่งได้แก่จำนวนวันระหว่างวันพบเชื้อและวันจะหน่วยการรักษา (X_{46}) พบว่าจำนวนวันที่เพิ่มมากขึ้นนั้นส่งผลให้โอกาสการเสียชีวิตลดลง ในทำนองเดียวกัน ผู้ป่วยที่รับฉีดวัคซีนแอสตราเซนเนก้า (X_{14}) ก่อนติดเชื้อพบว่าช่วยลดโอกาสการเสียชีวิตได้ เช่นเดียวกับผู้ป่วยที่ได้รับยาเดกซาเมทาโซน (X_{43}) ในการรักษาพบว่ามีโอกาสเสียชีวิตลดลง รวมไปถึงกลุ่มผู้ป่วยมีลักษณะภาพรังสีทรวงอกที่พบความผิดปกติของปอดข้างขวา (X_{24}) หรือผู้ป่วยที่มีโรคเบาหวาน (X_7) เป็นโรคประจำตัว พบว่าโอกาสเสียชีวิตลดลงในกลุ่มผู้ป่วยดังกล่าว (ตารางที่ 4.8)

ตารางที่ 4.7 ผลการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรอิสระที่เป็นข้อมูลเชิงปริมาณ

ตัวแปรอิสระ	X_2	X_4	X_{14}	X_{15}	X_{16}	X_{17}	X_{29}	X_{30}	X_{31}	X_{32}	X_{33}	X_{34}	X_{46}
X_2	1.000												
X_4	-0.182	1.000											
X_{14}	0.093	0.019	1.000										
X_{15}	-0.065	0.003	0.127	1.000									
X_{16}	-0.030	0.031	-0.013	0.244	1.000								
X_{17}	-0.080	-0.035	-0.012	-0.012	-0.003	1.000							
X_{29}	0.197	0.075	0.089	0.048	-0.018	-0.008	1.000						
X_{30}	-0.118	0.107	0.022	0.058	0.021	0.009	0.598	1.000					
X_{31}	0.100	0.252	-0.096	-0.109	-0.028	-0.032	0.000	-0.013	1.000				
X_{32}	-0.100	0.053	-0.061	-0.014	0.017	-0.001	-0.025	0.189	0.299	1.000			
X_{33}	-0.115	-0.175	0.095	0.073	0.029	0.031	0.042	0.051	-0.326	-0.100	1.000		
X_{34}	-0.113	0.000	-0.008	0.043	0.013	-0.020	0.090	0.057	-0.014	0.134	0.067	1.000	
X_{46}	-0.088	0.150	0.058	0.095	0.007	-0.025	0.197	0.230	-0.098	0.000	0.073	0.067	1.000

ตารางที่ 4.8 ผลการวิเคราะห์ตัวแบบการถดถอยลอจิสติกที่มีการคัดเลือกตัวแปรแบบ Backward

ตัวแปรอิสระ	coef	Std err	z	p> z
Intercept	-2.3734	0.282	-8.429	0.000
X ₂ อายุ	1.5329	0.190	8.083	0.000
X ₃ สัญชาติ	1.2438	0.459	2.709	0.007
X ₇ โรคเบาหวาน	-0.4733	0.161	-2.945	0.003
X ₈ โรคหัวใจ	0.7133	0.142	5.012	0.000
X ₁₀ โรคไต	1.0027	0.139	7.232	0.000
X ₁₃ โรคมะเร็ง	0.3997	0.109	3.659	0.000
X ₁₄ วัคซีนแอสตราเซนเนก้า	-0.8361	0.241	-3.473	0.001
X ₁₈ ภาพรังสีทรวงอกรูปแบบ Patchy Opacity	0.3409	0.154	2.212	0.027
X ₂₄ การกระจายตัวของรอยโรคเด่นที่ปอดข้างขวา	-0.4972	0.167	-2.976	0.003
X ₃₁ อัตราการหายใจ	0.3385	0.154	2.201	0.028
X ₃₅ เครื่องช่วยหายใจแบบใส่ท่อช่วยหายใจ	1.2081	0.146	8.281	0.000
X ₃₆ เครื่องช่วยหายใจแบบไม่ใส่ท่อช่วยหายใจ	0.4201	0.147	2.855	0.004
X ₄₃ ยาเดกซาเมทาโซน	-0.6407	0.203	-3.154	0.002
X ₄₄ ยาเมทิลเพรดนิโซโลน	0.9584	0.201	4.769	0.000
X ₄₅ ยาไฮโดรคอร์ติโซน	0.3075	0.113	2.718	0.007
X ₄₆ จำนวนวันนับตั้งแต่วันตรวจพบเชื้อจนถึงวันจำหน่ายการรักษา	-0.9561	0.151	-6.346	0.000

ตารางที่ 4.9 ค่าความเป็นไปได้ และค่าสถิติทดสอบระดับความสัมพันธ์

Log likelihood		Pseudo R Square
Full Model	Constant-Only Model	
-165.30	-558.04	0.7038

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตารางที่ 4.9 ผลการวิเคราะห์ตัวแบบการถดถอยลอจิสติกที่มีการคัดเลือกตัวแปรแบบ Backward พบว่ามีตัวแปรอิสระทั้งหมด 16 ตัวแปร มีนัยสำคัญทางสถิติที่ระดับนัยสำคัญ 0.05 ค่าความเป็นไปได้ เมื่อพิจารณาความเหมาะสมค่า -2Log-Likelihood มีค่าเท่ากับ 330.60 ค่า Null deviance หรือค่า Log-Likelihood สำหรับ Constant-Only Model (L0) เท่ากับ -558.04 และค่า Residual deviance หรือค่า Log Likelihood สำหรับ Full Model (L1) เท่ากับ -165.3 ค่า McFadden's R Square เท่ากับ 0.7038 นั้นหมายความว่า ตัวแบบสามารถทำนายความผันแปรของโอกาสที่ผู้ป่วยโควิด-19 ที่จะเสียชีวิต และรอดชีวิตได้ร้อยละ 70.38 สามารถเขียนสมการ Logit ได้ดังนี้

$$\begin{aligned} \log(\text{Odds}) = & -2.3734 + 1.5329X_2 + 1.2438X_3 - 0.4733X_7 + 0.7133X_8 + 1.0027X_{10} \\ & + 0.3997X_{13} - 0.8361X_{14} + 0.3409X_{18} - 0.4972X_{24} + 0.3385X_{31} \\ & + 1.2081X_{35} + 0.4201X_{36} - 0.6407X_{43} + 0.9584X_{44} + 0.3075X_{45} - 0.9561X_{46} \end{aligned}$$

4.3 การวิเคราะห์ความสำคัญของตัวแปร (Feature Importance)

จากการวิเคราะห์ความสำคัญของตัวแปร ด้วยวิธีการวิเคราะห์การถดถอยลอจิสติก (LR) วิธีต้นไม้ตัดสินใจ (DT) และวิธีโครงข่ายประสาทเทียมแบบหลายลำดับชั้น (NN) พบว่าวิธีการวิเคราะห์การถดถอยลอจิสติกมีตัวแปรอิสระที่มีค่าความสำคัญสูงสุด 5 อันดับแรกที่ส่งผลต่อโอกาสการเสียชีวิตของผู้ป่วยโควิด-19 ได้แก่ อายุ สัญชาติ การใช้เครื่องช่วยหายใจแบบใส่ท่อช่วยหายใจ โรคไต และการใช้ยาเมทิลเพรดนิโซโลน ตามลำดับ ในส่วนของวิธีต้นไม้ตัดสินใจพบว่า ตัวแปรอิสระ ได้แก่ จำนวนวันนับตั้งแต่วันที่ตรวจพบเชื้อจนถึงวันจำหน่ายการรักษา อายุ การใช้เครื่องช่วยหายใจแบบใส่ท่อช่วยหายใจ อัตราการช่วยหายใจ และปริมาณออกซิเจนในเลือด เป็นตัวแปรอิสระที่มีความสำคัญในการทำนายโอกาสการเสียชีวิตของผู้ป่วย และวิธีโครงข่ายประสาทเทียมแบบหลายลำดับชั้น พบว่าตัวแปรอิสระที่มีค่าความสำคัญสูงสุด 5 อันดับแรกมีความคล้ายคลึงกันกับทั้งสองวิธีที่ผ่านมา โดยที่การใช้เครื่องช่วยหายใจแบบใส่ท่อช่วยหายใจ มีความสำคัญมากที่สุด รองลงมาคือ โรคไต อายุ จำนวนวันนับตั้งแต่วันที่ตรวจพบเชื้อจนถึงวันจำหน่ายการรักษา และการใช้ยาไฮโดรคอร์ติโซน ตามลำดับ ดังแสดงในตารางที่ 4.10

ตารางที่ 4.10 ผลการวิเคราะห์ค่าความสำคัญของตัวแปรอิสระ (Feature Importance)

ตัวแปรอิสระ	LR	DT	NN
X_{35} เครื่องช่วยหายใจแบบใส่ท่อช่วยหายใจ	1.2081	0.1432	0.0873
X_{10} โรคไต	1.0027	0.0081	0.0551
X_2 อายุ	1.5329	0.1823	0.0496
X_{46} จำนวนวันนับตั้งแต่วันที่ตรวจพบเชื้อจนถึงวันจำหน่าย การรักษา	-0.9561	0.2533	0.0300
X_{45} ยาไฮโดรคอร์ติโซน	0.3075	0.0000	0.0208

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.10 ผลการวิเคราะห์ค่าความสำคัญของตัวแปรอิสระ (Feature Importance) (ต่อ)

ตัวแปรอิสระ	LR	DT	NN
X ₈ โรคหัวใจ	0.7133	0.0189	0.0203
X ₃₉ ยาคาซิรีวิแมบ และยาอิมติวิแมบ	0.0000	0.0111	0.0117
X ₇ โรคเบาหวาน	-0.4733	0.0043	0.0109
X ₁₃ โรคมะเร็ง	0.3997	0.0000	0.0094
X ₄₃ ยาเด็กซาเมทาโซน	-0.6407	0.0271	0.0089
X ₆ โรคปอดอุดกั้นเรื้อรัง	0.0000	0.0000	0.0084
X ₁₄ วัคซีนแอสตราเซนเนก้า	-0.8361	0.0097	0.0082
X ₃ สัญชาติ	1.2438	0.0000	0.0082
X ₃₃ ปริมาณออกซิเจนในเลือด	0.0000	0.0610	0.0077
X ₃₆ เครื่องช่วยหายใจแบบไม่ใส่ท่อช่วยหายใจ	0.4201	0.0099	0.0072
X ₁₅ วัคซีนซิโนแวค	0.0000	0.0000	0.0055
X ₄₀ ยาเรมเดซิเวียร์	0.0000	0.0000	0.0050
X ₁₁ โรคตับ	0.0000	0.0000	0.0047
X ₄₄ ยาเมทิลเฟรดนิโซโลน	0.9584	0.0063	0.0047
X ₂₉ ความดันโลหิตสูงสุดขณะหัวใจห้องล่างบีบตัว	0.0000	0.0368	0.0045
X ₅ โรคหลอดเลือดสมอง	0.0000	0.0000	0.0037
X ₁₆ วัคซีนไฟเซอร์	0.0000	0.0000	0.0030
X ₃₂ อัตราการเต้นของหัวใจ	0.0000	0.0115	0.0030
X ₃₁ อัตราการหายใจ	0.3385	0.0681	0.0027
X ₂₃ การกระจายตัวของรอยโรคเด่นที่ปอดข้างซ้าย	0.0000	0.0086	0.0022
X ₃₇ การให้ออกซิเจนผ่านหน้ากาก หรือสายออกซิเจน	0.0000	0.0000	0.0017
X ₉ โรคเอดส์	0.0000	0.0000	0.0017
X ₂₄ การกระจายตัวของรอยโรคเด่นที่ปอดข้างขวา	-0.4972	0.0317	0.0015
X ₁₈ ภาพรังสีทรวงอกรูปแบบ Patchy Opacity	0.0000	0.0040	0.0012
X ₂₇ การกระจายตัวของรอยโรคเด่นที่ปอดส่วนบน	0.0000	0.0000	0.0012
X ₃₄ อุณหภูมิร่างกาย	0.0000	0.0414	0.0012
X ₂₀ ภาพรังสีทรวงอกรูปแบบ Reticulonodular Opacity	0.0000	0.0000	0.0010
X ₂₁ การกระจายตัวของรอยโรคในปอดเด่นที่ Peripheral	0.0000	0.0031	0.0010
X ₁ เพศ	0.0000	0.0126	0.0007
X ₁₂ โรคปอด	0.0000	0.0000	0.0007
X ₁₇ วัคซีนซิโนฟาร์ม	0.0000	0.0000	0.0007

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์และห้ามการใช้นิติกรรมที่ออกโดยหน่วยงานนี้ ไม่อนุญาตให้นำไปเผยแพร่โดยไม่ขออนุญาต

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.10 ผลการวิเคราะห์ค่าความสำคัญของตัวแปรอิสระ (Feature Importance) (ต่อ)

ตัวแปรอิสระ	LR	DT	NN
X ₁₉ ภาพรังสีทรวงอกรูปแบบ Ground Glass Opacity	0.3409	0.0000	0.0007
X ₃₀ ความดันโลหิตที่ต่ำสุดขณะหัวใจห้องล่างคลายตัว	0.0000	0.0209	0.0007
X ₄₂ ยาแพ็กซ์โลวิด	0.0000	0.0000	0.0007
X ₂₂ การกระจายตัวของรอยโรคเด่นที่ Perihilar	0.0000	0.0000	0.0005
X ₄ ดัชนีมวลกาย	0.0000	0.0197	0.0002
X ₄₁ ยาฟ้าทลายโจร	0.0000	0.0000	0.0002
X ₂₅ การกระจายตัวของรอยโรคเด่นที่ปอดทั้ง 2 ข้าง	0.0000	0.0000	0.0000
X ₂₆ การกระจายตัวของรอยโรคเด่นที่ปอดส่วนล่าง	0.0000	0.0065	0.0000
X ₂₈ การกระจายตัวของรอยโรคเด่นที่ปอดส่วนกลาง	0.0000	0.0000	0.0000
X ₃₈ ยาฟาวิพิราเวียร์	0.0000	0.0000	0.0000

หมายเหตุ เรียงลำดับค่าความสำคัญจากและวิธีโครงข่ายประสาทเทียมแบบหลายลำดับชั้น (NN) จากมากไปน้อย

4.4 การทำนายการเสียชีวิตของผู้ป่วยผู้ป่วยโควิด-19

การทำนายการเสียชีวิตของผู้ป่วยผู้ป่วยโควิด-19 จากชุดข้อมูลทดสอบ จำนวน 346 ราย โดยวิธีการวิเคราะห์การถดถอยลอจิสติก วิธีต้นไม้ตัดสินใจ และวิธีโครงข่ายประสาทเทียมแบบหลายลำดับชั้น ได้ผลดังนี้

4.4.1 วิธีการวิเคราะห์การถดถอยลอจิสติก

ตารางที่ 4.11 เมตริกซ์ความสับสนจากการทำนายการเสียชีวิตของผู้ป่วยผู้ป่วยโควิด-19 ด้วยวิธีการวิเคราะห์การถดถอยลอจิสติก

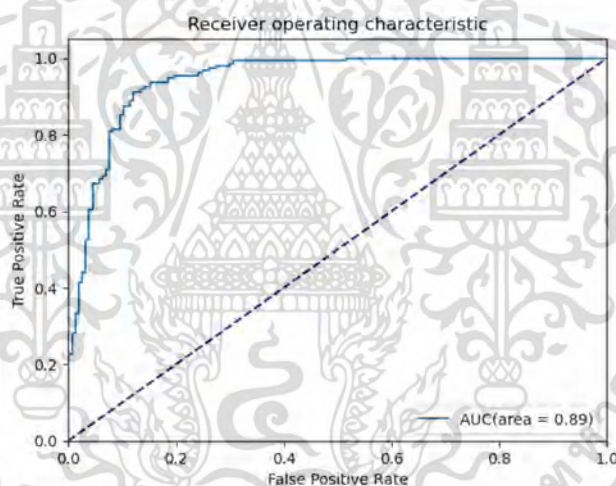
		ค่าทำนาย	
		รอดชีวิต	เสียชีวิต
ค่าจริง	รอดชีวิต	136	21
	เสียชีวิต	16	173

ตารางที่ 4.10 มีข้อมูล 346 ราย ตัวแบบสามารถทำนายข้อมูลได้ถูกต้อง 309 ราย โดยมีจำนวนข้อมูลที่จำแนกถูกว่ารอดชีวิต 136 ราย และข้อมูลที่จำแนกถูกว่าเสียชีวิต 173 ราย และตัวแบบทำนายข้อมูลไม่ถูกต้อง 37 ราย โดยมีจำนวนข้อมูลที่จำแนกผิดว่ารอดชีวิต ซึ่งผลที่แท้จริงเสียชีวิต 16 ราย และข้อมูลที่จำแนกผิดว่าเสียชีวิต ซึ่งผลที่แท้จริงแล้วรอดชีวิต 21 ราย

ตารางที่ 4.12 ค่าประสิทธิภาพของตัวแบบจำแนกรายละเอียดในแต่ละคำตอบจากการทำนายเสียชีวิตของผู้ป่วยผู้ป่วยโควิด-19 ด้วยวิธีการวิเคราะห์การถดถอยลอจิสติก

Class	Precision	Recall	F1- score	Accuracy
0	0.89	0.87	0.88	0.89
1	0.89	0.92	0.90	
macro avg	0.89	0.89	0.89	
weighted avg	0.89	0.89	0.89	

ตารางที่ 4.12 ประสิทธิภาพการทำนายของตัวแบบมีค่าความแม่นยำเท่ากับ 0.89 สำหรับค่าทำนายรอดชีวิตมีความเที่ยงเท่ากับ 0.89 ค่าเรียกคืนเท่ากับ 0.87 และค่าความถ่วงดุลเท่ากับ 0.88 ส่วนค่าทำนายเสียชีวิตมีความเที่ยงเท่ากับ 0.89 ค่าเรียกคืนเท่ากับ 0.92 และค่าความถ่วงดุลเท่ากับ 0.90



รูปที่ 4.1 AUC ของวิธีการวิเคราะห์การถดถอยลอจิสติก

จากรูปที่ 4.1 พบว่ากราฟ ROC curve มีลักษณะโค้งเข้าใกล้มุมบนซ้าย และมีค่าพื้นที่ใต้โค้งเท่ากับ 0.89 ซึ่งมีค่าเข้าใกล้ 1 แสดงว่าตัวแบบวิธีการวิเคราะห์การถดถอยลอจิสติก มีความเหมาะสม

4.4.2 วิธีต้นไม้ตัดสินใจ

ตารางที่ 4.13 เมตริกซ์ความสับสนจากการทำนายการเสียชีวิตของผู้ป่วยผู้ป่วยโควิด-19 ด้วยวิธีต้นไม้ตัดสินใจ

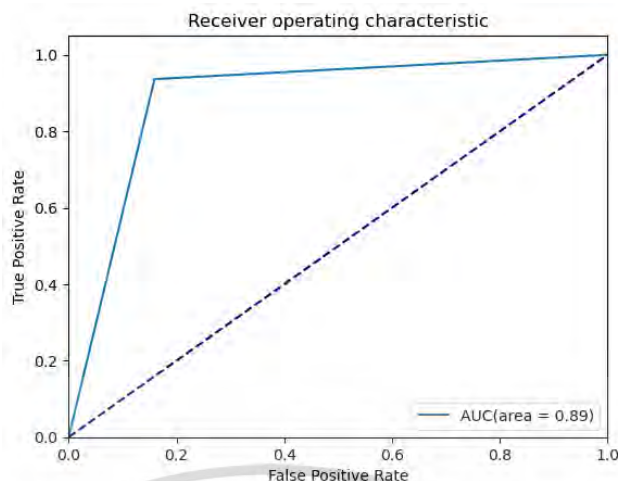
		ค่าทำนาย	
		รอดชีวิต	เสียชีวิต
ค่าจริง	รอดชีวิต	132	25
	เสียชีวิต	12	177

ตารางที่ 4.13 มีข้อมูล 346 ราย ตัวแบบสามารถทำนายข้อมูลได้ถูกต้อง 309 ราย โดยมีจำนวนข้อมูลที่จำแนกถูกกว่ารอดชีวิต 132 ราย และข้อมูลที่จำแนกถูกกว่าเสียชีวิต 177 ราย และตัวแบบทำนายข้อมูลไม่ถูกต้อง 37 ราย โดยมีจำนวนข้อมูลที่จำแนกผิดว่ารอดชีวิต ซึ่งผลที่แท้จริงเสียชีวิต 12 ราย และข้อมูลที่จำแนกผิดว่าเสียชีวิต ซึ่งผลที่แท้จริงแล้วรอดชีวิต 25 ราย

ตารางที่ 4.14 ค่าประสิทธิภาพของตัวแบบจำแนกรายละเอียดในแต่ละค่าตอบจากการทำนายการเสียชีวิตของผู้ป่วยผู้ป่วยโควิด-19 ด้วยวิธีต้นไม้ตัดสินใจ

Class	Precision	Recall	F1- score	Accuracy
0	0.92	0.84	0.88	0.89
1	0.88	0.94	0.91	
macro avg	0.90	0.88	0.89	
weighted avg	0.89	0.89	0.89	

ตารางที่ 4.14 ประสิทธิภาพการทำนายของตัวแบบมีค่าความแม่นยำเท่ากับ 0.89 สำหรับค่าทำนายรอดชีวิตมีค่าความเที่ยงเท่ากับ 0.92 ค่าเรียกคืนเท่ากับ 0.84 และค่าความถ่วงดุลเท่ากับ 0.88 ส่วนค่าทำนายเสียชีวิตมีค่าความเที่ยงเท่ากับ 0.88 ค่าเรียกคืนเท่ากับ 0.94 และค่าความถ่วงดุลเท่ากับ 0.91



รูปที่ 4.2 AUC ของวิธีต้นไม้ตัดสินใจ

จากรูปที่ 4.2 พบว่ากราฟ ROC curve มีลักษณะโค้งเข้าใกล้มุมบนซ้าย และมีค่าพื้นที่ใต้โค้งเท่ากับ 0.89 ซึ่งมีค่าเข้าใกล้ 1 แสดงว่าตัวแบบวิธีต้นไม้ตัดสินใจ มีความเหมาะสม

4.4.3 วิธีโครงข่ายประสาทเทียมแบบหลายลำดับชั้น

ตารางที่ 4.15 เมตริกซ์ความสับสนจากการทำนายการเสียชีวิตของผู้ป่วยผู้ป่วยโควิด-19 ด้วยวิธีโครงข่ายประสาทเทียมแบบหลายลำดับชั้น

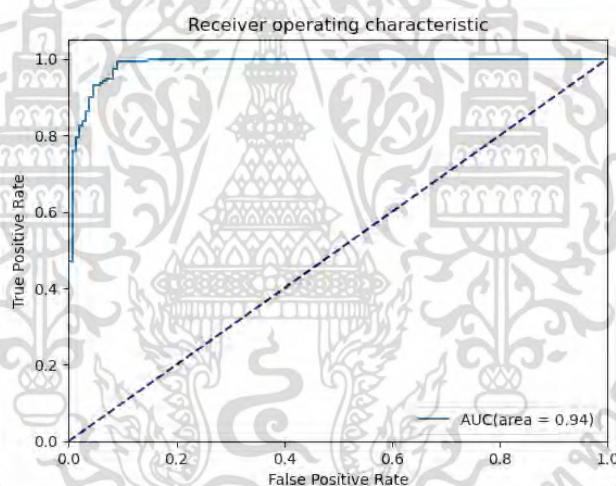
		ค่าทำนาย	
		รอดชีวิต	เสียชีวิต
ค่าจริง	รอดชีวิต	139	18
	เสียชีวิต	0	189

ตารางที่ 4.15 มีข้อมูล 346 ราย ตัวแบบสามารถทำนายข้อมูลได้ถูกต้อง 328 ราย โดยมีจำนวนข้อมูลที่จำแนกถูกว่ารอดชีวิต 139 ราย และข้อมูลที่จำแนกถูกว่าเสียชีวิต 189 ราย และตัวแบบทำนายข้อมูลไม่ถูกต้อง 18 ราย โดยทั้งหมดเป็นการจำแนกผิดว่าเสียชีวิต ซึ่งผลที่แท้จริงแล้วรอดชีวิต 18 ราย

ตารางที่ 4.16 ค่าประสิทธิภาพของตัวแบบจำแนกรายละเอียดในแต่ละคำตอบจากการทำนายการเสียชีวิตของผู้ป่วยผู้ป่วยโควิด-19 ด้วยวิธีโครงข่ายประสาทเทียมแบบหลายลำดับชั้น

Class	Precision	Recall	F1- score	Accuracy
0	1.00	0.89	0.94	0.95
1	0.91	1.00	0.95	
macro avg	0.96	0.94	0.95	
weighted avg	0.95	0.95	0.95	

ตารางที่ 4.16 ประสิทธิภาพการทำนายของตัวแบบมีความแม่นยำเท่ากับ 0.95 สำหรับค่าทำนายรอดชีวิตมีความเที่ยงเท่ากับ 1.00 ค่าเรียกคืนเท่ากับ 0.89 และค่าความถ่วงดุลเท่ากับ 0.94 ส่วนค่าทำนายเสียชีวิตมีความเที่ยงเท่ากับ 0.91 ค่าเรียกคืนเท่ากับ 1.00 และค่าความถ่วงดุลเท่ากับ 0.95



รูปที่ 4.3 AUC ของวิธีโครงข่ายประสาทเทียมแบบหลายลำดับชั้น

จากรูปที่ 4.3 พบว่ากราฟ ROC curve มีลักษณะโค้งเข้าใกล้มุมบนซ้าย และมีค่าพื้นที่ใต้โค้งเท่ากับ 0.94 ซึ่งมีค่าเข้าใกล้ 1 แสดงว่าตัวแบบวิธีโครงข่ายประสาทเทียมแบบหลายลำดับชั้นมีความเหมาะสม

4.5 ผลการเปรียบเทียบประสิทธิภาพในการทำนาย

การเปรียบเทียบประสิทธิภาพในการทำนายการเสียชีวิตของผู้ป่วยผู้ป่วยโควิด-19 ระหว่างวิธีการวิเคราะห์การถดถอยลอจิสติก วิธีต้นไม้ตัดสินใจ และวิธีโครงข่ายประสาทเทียมแบบหลายลำดับชั้น โดยพิจารณาจากค่าความแม่นยำ ค่าความเที่ยง ค่าเรียกคืน และค่าความถ่วงดุล

ตารางที่ 4.17 ผลการเปรียบเทียบประสิทธิภาพในการทำนายของวิธีการจำแนกกลุ่มทั้ง 3 วิธี

วิธีจำแนกกลุ่ม	Accuracy	Precision	Recall	F1-score
วิธีการวิเคราะห์การถดถอยลอจิสติก	0.89	0.89	0.92	0.90
วิธีต้นไม้ตัดสินใจ	0.89	0.88	0.94	0.91
วิธีโครงข่ายประสาทเทียมแบบหลายลำดับชั้น	0.95	0.91	1.00	0.95

ตารางที่ 4.17 พบว่าวิธีโครงข่ายประสาทเทียมแบบหลายลำดับชั้น มีค่าความแม่นยำมากที่สุดเท่ากับ 0.95 มีค่าความเที่ยงเท่ากับ 0.91 ค่าเรียกคืนเท่ากับ 1.00 และค่าความถ่วงดุลเท่ากับ 0.95 ดังนั้นวิธีโครงข่ายประสาทเทียมแบบหลายลำดับชั้น มีประสิทธิภาพในการทำนายดีที่สุด

4.6 อภิปรายผล

ผลการศึกษาปัจจัยที่สัมพันธ์ต่อการเสียชีวิตของผู้ป่วยโควิด-19 กรณีศึกษาโรงพยาบาลระยอง โดยพิจารณาจากค่าความแม่นยำ พบว่าวิธีโครงข่ายประสาทเทียมแบบหลายลำดับชั้นมีประสิทธิภาพในการทำนายผลดีที่สุดซึ่งสอดคล้องกับงานวิจัยของ Agbelusi and Olayemi (2020) ได้พัฒนาตัวแบบในการทำนายอัตราการเสียชีวิตของผู้ป่วยติดเชื้อไวรัสโคโรนาในประเทศไนจีเรีย โดยใช้เทคนิคการทำเหมืองข้อมูล 4 วิธีได้แก่ วิธีโครงข่ายประสาทเทียมแบบหลายลำดับชั้น วิธีนาอิวส์ วิธีต้นไม้ตัดสินใจ และกฎการตัดสินใจ ผลการศึกษาแสดงให้เห็นว่าวิธีโครงข่ายประสาทเทียมแบบหลายลำดับชั้น มีประสิทธิภาพในการทำนายสูงที่สุด โดยมีค่าความแม่นยำอยู่ที่ร้อยละ 85 และยังสอดคล้องกับงานวิจัยของ Khan. et al. (2021) ได้นำตัวแบบการเรียนรู้เชิงลึกที่มีการกำหนดชั้นซ่อนทั้งหมด 6 ชั้น โดยมีฟังก์ชันเรกติไฟต์ลินเนี่ยนิต เป็นฟังก์ชันกระตุ้น และฟังก์ชันซิกมอยด์ เป็นกระตุ้นในชั้นท้ายสุด รวมไปถึงตัวแบบการเรียนรู้ของเครื่อง ได้แก่ วิธีต้นไม้ตัดสินใจ วิธีการวิเคราะห์การถดถอยลอจิสติก วิธีป่าสุ่ม วิธีเอ็กซ์ทรีมกาเดียนบูตตึง และวิธีเพื่อนบ้านใกล้สุด K ตัว เพื่อใช้ทำนายโอกาสเสียชีวิตในผู้ป่วยโควิด-19 โดยใช้ข้อมูลผู้ป่วยที่รับการยืนยันว่าติดเชื้อไวรัสโคโรนาจากทั้งหมด 146 ประเทศพบว่าตัวแบบการเรียนรู้เชิงลึกมีประสิทธิภาพในการทำนายสูงที่สุด ด้วยค่าความแม่นยำ 0.97

ส่วนปัจจัยที่สัมพันธ์ต่อการเสียชีวิตของผู้ป่วยโควิด-19 จากการทำนายด้วยตัวแบบวิธีโครงข่ายประสาทเทียมแบบหลายลำดับชั้น พบว่าการใช้เครื่องช่วยหายใจแบบใส่ท่อช่วยหายใจ เป็นปัจจัยที่สำคัญที่สุดที่ส่งผลต่อการเสียชีวิตของผู้ป่วยโควิด-19 ซึ่งสอดคล้องกับการศึกษาของ Moulaei. et al. (2022) ได้ศึกษาผู้ป่วยโควิด-19 จำนวน 1,500 ราย ของโรงพยาบาล Ayatollah Taleghani เมือง Abadan ประเทศอิหร่าน โดยใช้ขั้นตอนวิธีการเรียนรู้ของเครื่อง ในการสร้างตัวแบบทำนายพบว่าจากตัวแปรทำนายทั้งหมด 38 ตัวแปร ภาวะหายใจลำบาก การเข้ารับการรักษาในห้องไอซียู และการได้รับการบำบัดด้วยออกซิเจน เป็นตัวแปรทำนาย 3 อันดับแรกที่ส่งผลต่อโอกาสการเสียชีวิตของผู้ป่วยโควิด-19 มากที่สุด



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สรุปผลการวิจัยและข้อเสนอแนะ

5.1 สรุปผลการวิจัย

การศึกษาปัจจัยที่สัมพันธ์ต่อการเสียชีวิตของผู้ป่วยโควิด-19 กรณีศึกษาโรงพยาบาลระยอง มีวัตถุประสงค์เพื่อศึกษาปัจจัยที่สัมพันธ์กับการเสียชีวิตของผู้ป่วยโควิด-19 จากข้อมูลผู้ป่วยโควิด-19 จำนวน 1,006 ราย ที่เข้ารับการรักษาในโรงพยาบาลระยอง โดยมีตัวแปรอิสระทั้งหมด 46 ตัวแปร นำไปสร้างตัวแบบในการทำนายโอกาสการเสียชีวิตของผู้ป่วยโควิด-19 ด้วยวิธีการวิเคราะห์การถดถอยลอจิสติก วิธีต้นไม้ตัดสินใจ และวิธีโครงข่ายประสาทเทียมแบบหลายลำดับชั้น ผลสรุปดังนี้

ประสิทธิภาพของตัวแบบวิธีการวิเคราะห์การถดถอยลอจิสติกในการทำนายชุดข้อมูลทดสอบพบว่า มีค่าความแม่นยำเท่ากับ 0.89 ค่าความเที่ยงเท่ากับ 0.89 ค่าเรียกคืนเท่ากับ 0.92 และค่าความถ่วงดุลเท่ากับ 0.90 โดยตัวแปรที่ส่งผลต่อโอกาสการเสียชีวิตของผู้ป่วยโควิด-19 จากการทำนายด้วยตัวแบบวิธีการวิเคราะห์การถดถอยลอจิสติก ได้แก่ อายุ สัญชาติ การใช้เครื่องช่วยหายใจแบบใส่ท่อช่วยหายใจ โรคไต และการใช้ยาเมทิลเพรดนิโซโลน ตามลำดับ

ประสิทธิภาพของตัวแบบวิธีต้นไม้ตัดสินใจในการทำนายชุดข้อมูลทดสอบพบว่า มีค่าความแม่นยำเท่ากับ 0.89 ค่าความเที่ยงเท่ากับ 0.88 ค่าเรียกคืนเท่ากับ 0.94 และค่าความถ่วงดุลเท่ากับ 0.91 โดยในส่วนของตัวแปรอิสระ ได้แก่ จำนวนวันนับตั้งแต่วันที่ตรวจพบเชื้อจนถึงวันจำหน่ายการรักษา อายุ การใช้เครื่องช่วยหายใจแบบใส่ท่อช่วยหายใจ อัตราการหายใจ และปริมาณออกซิเจนในเลือดเป็นตัวแปรอิสระที่มีความสำคัญในการทำนายโอกาสการเสียชีวิตของผู้ป่วยโควิด-19 ด้วยตัวแบบวิธีต้นไม้ตัดสินใจ

ประสิทธิภาพของตัวแบบวิธีโครงข่ายประสาทเทียมแบบหลายลำดับชั้น ในการทำนายชุดข้อมูลทดสอบพบว่า มีค่าความแม่นยำเท่ากับ 0.95 ค่าความเที่ยงเท่ากับ 0.91 ค่าเรียกคืนเท่ากับ 1.00 และค่าความถ่วงดุลเท่ากับ 0.95 โดยตัวแปรอิสระที่มีความสำคัญสูงสุด 5 อันดับแรก ได้แก่ การใช้เครื่องช่วยหายใจแบบใส่ท่อช่วยหายใจ มีค่าความสำคัญมากที่สุด รองลงมาคือ โรคไต อายุ จำนวนวันนับตั้งแต่วันที่ตรวจพบเชื้อจนถึงวันจำหน่ายการรักษา และการใช้ยาไฮโดรคอร์ติโซนในการรักษาตามลำดับ

จากประสิทธิภาพของตัวแบบทั้ง 3 วิธี ในการทำนายชุดข้อมูลทดสอบพบว่า วิธีโครงข่ายประสาทเทียมแบบหลายลำดับชั้น มีประสิทธิภาพสูงสุด รองลงมาคือวิธีต้นไม้ตัดสินใจ และวิธีการวิเคราะห์การถดถอยลอจิสติก ตามลำดับ โดยตัวแปรอิสระที่มีความสำคัญสูงสุด 5 อันดับแรกจากตัวแบบทั้ง 3 วิธีพบว่า อายุ และการใช้เครื่องช่วยหายใจแบบใส่ท่อช่วยหายใจ เป็นตัวแปรอิสระที่มีความสำคัญสูงสุดในการทำนายโอกาสการเสียชีวิตของผู้ป่วยโควิด-19 ด้วยตัวแบบทั้ง 3 วิธี โรคไต

เป็นตัวแปรอิสระที่มีความสำคัญในการทำนายด้วยตัวแบบวิธีการวิเคราะห์การถดถอยลอจิสติก และเอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาค้นคว้าเท่านั้น ไม่อนุญาตให้นำไปเผยแพร่หรือนำไปใช้ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

วิธีโครงข่ายประสาทเทียมแบบหลายลำดับชั้น จำนวนวันนับตั้งแต่วันตรวจพบเชื้อจนถึงวันจำหน่าย การรักษาเป็นตัวแปรอิสระที่มีความสำคัญในการทำนายด้วยตัวแบบวิธีต้นไม้ตัดสินใจ และวิธีโครงข่ายประสาทเทียมแบบหลายลำดับชั้น สัญชาติ และการใช้ยาเมทิลเพรดนิโซโลน เป็นตัวแปรอิสระที่มีความสำคัญในการทำนายด้วยตัวแบบวิธีการวิเคราะห์การถดถอยลอจิสติก อัตราการหายใจ และปริมาณออกซิเจนในเลือด เป็นตัวแปรอิสระที่มีความสำคัญในการทำนายด้วยตัวแบบวิธีต้นไม้ตัดสินใจ และการใช้ยาไฮโดรคอร์ติโซนในการรักษาเป็นตัวแปรอิสระที่มีความสำคัญในการทำนายด้วยตัวแบบวิธีโครงข่ายประสาทเทียมแบบหลายลำดับชั้น

5.2 ข้อเสนอแนะ

- 1) ตัวแปรอิสระที่นำมาใช้ในงานวิจัยนี้เป็นเพียงบางส่วนของปัจจัยที่สัมพันธ์ต่อการเสียชีวิตของผู้ป่วยโควิด-19 เท่านั้น เพื่อให้ตัวแบบมีประสิทธิภาพเพิ่มมากขึ้น ควรเพิ่มตัวแปรอิสระอื่น ๆ ที่เกี่ยวข้อง เช่น ผลการตรวจทางห้องปฏิบัติการ อาการแสดงของผู้ป่วย เป็นต้น
- 2) ใช้วิธีการเรียนรู้ของเครื่อง อื่น ๆ ในการสร้างตัวแบบสำหรับปัญหาจำแนกกลุ่ม เช่น วิธีป่าสุ่ม วิธีเอ็กซ์ทรีมกาเดียนบูทตั้ง วิธีการซัพพอร์ตเวกเตอร์แมชชีน วิธีนาอิวเบส และวิธีเพื่อนบ้านใกล้สุด K ตัว เป็นต้น

เอกสารอ้างอิง

- กรมควบคุมโรค. 2564ก. สถานการณ์โรคติดเชื้อไวรัสโคโรนา 2019 (COVID-19) มาตรการสาธารณสุข และปัญหาอุปสรรคการป้องกันควบคุมโรคในผู้เดินทาง วันที่ 18 สิงหาคม 2564. [Online]. เข้าถึงได้จาก : <https://ddc.moph.go.th/uploads/files/2017420210820025238.pdf>
- คณะแพทยศาสตร์โรงพยาบาลรามาธิบดี มหาวิทยาลัยมหิดล. 2563. ความรู้พื้นฐาน COVID-19 ตอนที่ 1. [Online]. เข้าถึงได้จาก : <https://www.rama.mahidol.ac.th/ceb/News/2020/Covid-19>
- กรมควบคุมโรค. 2563. รายงานสถานการณ์โรคติดเชื้อไวรัสโคโรนา 2019 ฉบับที่ 58 วันที่ 1 มีนาคม 2563. [Online]. เข้าถึงได้จาก : <https://ddc.moph.go.th/viralpneumonia/file/situation/situation-no58-010363.pdf>
- กรมควบคุมโรค. 2564ข. สถานการณ์โรคติดเชื้อไวรัสโคโรนา 2019 (COVID-19) มาตรการสาธารณสุข และปัญหาอุปสรรคการป้องกันควบคุมโรคในผู้เดินทาง วันที่ 14 มกราคม 2564. [Online]. เข้าถึงได้จาก : [https://ddc.moph.go.th/uploads/ckeditor2//files/1\)%20วิเคราะห์สถานการณ์%20covid%20\(15%20มค64cp\).pdf](https://ddc.moph.go.th/uploads/ckeditor2//files/1)%20วิเคราะห์สถานการณ์%20covid%20(15%20มค64cp).pdf)
- คณะกรรมการโรคติดต่อจังหวัดระยอง. 2563. แฉลงการณ์จังหวัดระยอง ฉบับที่ 1 เรื่อง สถานการณ์ โรคติดเชื้อไวรัสโคโรนา 2019 (COVID-19) ในพื้นที่จังหวัดระยอง วันที่ 19 มีนาคม 2563.
- สาธิต ปิตุเตชะ. 2563. ผู้ป่วยโควิด-19 เพศชาย อายุ 45 ปี เสียชีวิต 1 ราย ที่ระยอง. [Online]. เข้าถึงได้จาก : https://www.facebook.com/permalink.php?story_fbid=3648069868612657&id=119568378129508
- โรงพยาบาลวิชัยเวช อินเตอร์เนชั่นแนล หนองแขม. 2564. โควิดสายพันธุ์โอไมครอนคืออะไร น่ากลัวแค่ไหน. [Online]. เข้าถึงได้จาก : <https://vichaivej-nongkhaem.com/healthinfo/โควิดโอไมครอน/>
- โรงพยาบาลพริ้นซ์ สุวรรณภูมิ. 2564. เรื่องควรรู้ เกี่ยวกับเชื้อไวรัสโควิด-19 สายพันธุ์โอไมครอน (Omicron). [Online]. เข้าถึงได้จาก : <https://www.princsuvarnabhumi.com/content-omicron/>
- กรมควบคุมโรค. 2564ค. รายงานสถานการณ์โรคติดเชื้อไวรัสโคโรนา 2019 ฉบับที่ 724 วันที่ 27 ธันวาคม 2564. [Online]. เข้าถึงได้จาก : <https://ddc.moph.go.th/viralpneumonia/file/situation/situation-no724-271264.pdf>

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- กรมการแพทย์. 2565. แนวทางเวชปฏิบัติ การวินิจฉัย ดูแลรักษา และป้องกันการติดเชื้อใน
โรงพยาบาล กรณีโรคติดเชื้อไวรัสโคโรนา 2019 (COVID-19) สำหรับแพทย์และบุคลากร
สาธารณสุข ฉบับปรับปรุง วันที่ 22 เมษายน 2565. [Online]. เข้าถึงได้จาก : [https://
covid19.dms.go.th/backend/Content/Content_File/Covid_Health/Attach/2565042
2162203PM_CPG_COVID-19_n_v.22_20220422.pdf](https://covid19.dms.go.th/backend/Content/Content_File/Covid_Health/Attach/25650422162203PM_CPG_COVID-19_n_v.22_20220422.pdf)
- Das, A. K., Mishra, S., and Gopalan, S. S. 2020. “Predicting CoVID-19 community mortality
risk using machine learning and development of an online prognostic tool.” *PeerJ.*
8 : e10083.
- Sánchez-Montañés, M., Rodríguez-Belenguer, P., Serrano-López, A. J., Soria-Olivas, E., and
Alakhdar-Mohmara, Y. 2020. “Machine learning for mortality analysis in patients
with COVID-19.” *International journal of environmental research and public
health.* 17(22) : 8386.
- Ferreira, A. T., Fernandes, C., Vieira, J., and Portela, F. 2021. “Pervasive intelligent models
to predict the outcome of COVID-19 patients.” *Future Internet.* 13(4) : 102.
- Muhammad, L. J., Algehyne, E. A., Usman, S. S., Ahmad, A., Chakraborty, C., and
Mohammed, I. A. 2021. “Supervised machine learning models for prediction of
COVID-19 infection using epidemiology dataset.” *SN computer science.* 2(1) : 1-13.
- Li, X., Ge, P., Zhu, J., Li, H., Graham, J., Singer, A., Richman, P.S. and Duong, T.Q. 2020.
“Deep learning prediction of likelihood of ICU admission and mortality in COVID-
19 patients using clinical variables.” *PeerJ.* 8 : e10337.
- Agbelusi, O., and Olayemi, O. C. 2020. “Prediction of mortality rate of COVID-19
patients using machine learning techniques in nigeria.” *International journal of
computer science and software engineering.* 9(5) : 30-34.
- Khan, I. U., Aslam, N., Aljabri, M., Aljameel, S. S., Kamaleldin, M. M. A., Alshamrani, F.
M., and Chrouf, S. M. B. 2021. “Computational intelligence-based model for
mortality rate prediction in COVID-19 patients.” *International journal of
environmental research and public health.* 18(12) : 6429.
- Pourhomayoun, M., and Shakibi, M. 2021. “Predicting mortality risk in patients with
COVID-19 using machine learning to help medical decision-making.” *Smart
Health.* 20 : 100178.
- โรงพยาบาลพญาไท. 2564. COVID-19 HANDBOOK ทำความรู้จักไวรัส CORONA. [Online].
เข้าถึงได้จาก : <https://www.phyathai.com/covid19healthyalert/>

องค์การอนามัยโลก. 2565. การติดตามสายพันธุ์ของเชื้อ SARS-CoV-2. [Online]. เข้าถึงได้จาก : <https://www.who.int/thailand/news/feature-stories/detail/tracking-SARS-CoV-2-variants>

ณญา ธนกิจธรรมกุล. 2564. “การพยาบาลผู้ป่วยโรคติดเชื้อโคโรนาไวรัส 19 ในระยะที่มีภาวะหายใจลำบากเฉียบพลัน.” *วารสารสภาการพยาบาล*. 36(3) : 16-30

องค์การอนามัยโลก. 2564. **Coronavirus disease (COVID-19)**. [Online]. เข้าถึงได้จาก : <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/coronavirus-disease-covid-19>

โรงพยาบาลบำรุงราษฎร์. 2563. 8 กลุ่มผู้ป่วยที่หากติด Covid-19 เสี่ยงมีอาการรุนแรง. [Online]. เข้าถึงได้จาก : <https://www.bumrungrad.com/th/health-blog/april2020/covid-19-there-are-eight-risk-groups>

Stoltzfus, J. C. 2011. “Logistic regression: a brief primer.” *Academic Emergency Medicine*. 18(10) : 1099-1104.

กัลยา วานิชย์บัญชา. 2546. การวิเคราะห์สถิติขั้นสูงด้วย SPSS for Windows. พิมพ์ครั้งที่ 3. ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย. กรุงเทพฯ.

อรทัย เจริญสิทธิ์. 2560. “การวิเคราะห์การถดถอยลอจิสติกแบบไบนารีสำหรับการวิจัยทางสังคมศาสตร์.” *วารสารมนุษยศาสตร์และสังคมศาสตร์มหาวิทยาลัยเอเชียอาคเนย์*. 1(2) : 1-9.

ชนินันท์ พลกษ์ประมุข, นัฐิกา เจริญตะคุ และสิวะโชติ ศรีสุทธิยากร. 2563. “ประสิทธิภาพของสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน สเปียร์แมน และเคนดอลล์ เมื่อข้อมูลแจกแจงแบบไม่ปกติ.” *An Online Journal of Education*. 15(2) : 16 หน้า

โกเมศ อัมพวัน. ม.ป.ป. **การทำเหมืองข้อมูล**. [Online]. เข้าถึงได้จาก : <https://staff.informatics.buu.ac.th/~komate/886464/%5B6%5D-Classification.pdf>

พรพล ธรรมรงค์รัตน์, ลัดดา ปรีชาวีรกุล และวิภาดา เวทย์ประสิทธิ์. 2551. “การจำแนกประเภทเว็บเพจโดยใช้ค่าความถี่เอกสารและซอฟต์แวร์เวกเตอร์แมชชีน.” หน้า 498-504. ใน **การประชุมวิชาการวิทยาการคอมพิวเตอร์และวิศวกรรมคอมพิวเตอร์แห่งชาติ ครั้งที่ 12 (NCSEC 2008)**. ชลบุรี.

สายชล สันสมบุรณ์ทอง. 2563. “การเปรียบเทียบประสิทธิภาพในการทำนายผลค่านอกเกณฑ์ด้วยการจำแนก 6 วิธี.” *Thai Journal of Science and Technology*, 9(3) : 255-268.

นิเวศ จิระวิชุดชัย. 2553. “การค้นหาค่าเทคนิคเหมืองข้อมูลเพื่อสร้างตัวแบบการวิเคราะห์โรคอัตโนมัติ.” กรุงเทพฯ : มหาวิทยาลัยราชภัฏสวนสุนันทา สถาบันวิจัยและพัฒนา.

พุทธิพร ธนธรรมเมธี. และเยาวเรศ ศิริสถิตย์กุล. 2562. “เทคนิคการจำแนกข้อมูลที่พัฒนาสำหรับชุดข้อมูลที่ไม่สมดุลของภาวะข้อเข่าเสื่อมในผู้สูงอายุ.” *วารสารวิทยาศาสตร์และเทคโนโลยี*. 27(6) : 1164-1178.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- Moulaei, K., Shanbehzadeh, M., Mohammadi-Taghiabad, Z., and Kazemi-Arpanahi, H. 2022. “Comparing machine learning algorithms for predicting COVID-19 mortality.” *BMC medical informatics and decision making*. 22(1) : 1-12.
- Patsadu, O., Tangchitwilakun, P. and Lowsuwankul, S. 2020. “Liver Cancer Patient Classification on a Multiple-Stage using Hybrid Classification Methods.” *Walailak Journal of Science and Technology (WJST)*. 18(10) : 9169-14.
- Stevens, J. 1996. “Applied multivariate statistics for the social science.” Routledge.



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ชุดคำสั่งโปรแกรมภาษาไพธอน

```
!pip install imblearn
```

```
!pip install mlxtend
```

```
!pip install eli5
```

```
#===== Import Library =====
```

```
import pandas as pd
```

```
import numpy as np
```

```
from sklearn import preprocessing
```

```
import matplotlib.pyplot as plt
```

```
#===== Import Data =====
```

```
data = pd.read_excel('Data IS.xlsx')
```

```
data = data.dropna()
```

```
random_seed = 0
```

```
#===== Data exploration =====
```

```
data.shape
```

```
data.columns
```

```
## Numerical Features
```

```
numerical_features = ['Age', 'BMI', 'AstraZeneca', 'Sinovac', 'Pfizer', 'Sinopharm',  
                      'Bp_systolic', 'Bp_diastolic', 'Respiration_rate', 'Pulse_rate',  
                      'Oxygen_saturation', 'Temperature',  
                      'Day_between_detected_and_discharge']
```

```
## Categorical Features
```

```
categorical_features = data.drop(columns = numerical_features)
```

```
categorical_features = categorical_features.drop(columns = 'Died')
```

```
categorical_features.columns
```

```
### Create dummy variables
```

```
#### - Nationality
```

```
from sklearn.preprocessing import LabelEncoder
```

```
np.unique(data['Nationality'])
```

```
encoder = LabelEncoder()
```

```
labels = encoder.fit_transform(data['Nationality'])
```

```
mappings = {index: label for index, label in enumerate(encoder.classes_)}
```

```
data['Nationality'] = labels
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- Gender

```

data['Gender'].unique()
encoder = LabelEncoder()
labels = encoder.fit_transform(data['Gender'])
mappings = {index: label for index, label in enumerate(encoder.classes_)}
data['Gender'] = labels

```

Define X, y

```

X = data.drop(columns = ['Died'])
y = data[['Died']]

```

===== Normalization =====

```

from sklearn.preprocessing import StandardScaler
feature_norm = StandardScaler().fit_transform(X)
X = pd.DataFrame(feature_norm, columns = X.columns)

```

===== Imbalance class =====**## Check imbalance class**

```

count_survival = len(y[y['Died'] == 0])
print("Number of survival is", count_survival)
count_died = len(y[y['Died'] == 1])
print("Number of died is", count_died)
pct_of_survival = count_survival/(count_survival+count_died)
print("Percentage of survival data is", pct_of_survival*100)
pct_of_died = count_died/(count_survival+count_died)
print("Percentage of died data is", pct_of_died*100)

```

Synthetic Minority Oversampling Technique (SMOTE)

```

from sklearn.model_selection import train_test_split
from imblearn.over_sampling import SMOTE
os = SMOTE(random_state = random_seed)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.3, random_state =
                                                    random_seed)
smote_data_X,smote_data_y = os.fit_resample(X_train, y_train)
smote_data_X = pd.DataFrame(data = smote_data_X, columns = X_train.columns )
smote_data_y = pd.DataFrame(data = smote_data_y, columns = ['Died'])

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

print("Length of oversampled data is", len(smote_data_X))
print("No. of survival in SMOTE data", len(smote_data_y[smote_data_y['Died'] == 0]))
print("No. of died in SMOTE data", len(smote_data_y[smote_data_y['Died'] == 1]))
print("Percentage of survival data in oversampled data is", len(smote_data_y
    [smote_data_y['Died'] == 0])/len(smote_data_X)*100)
print("Percentage of died data in oversampled data is", len(smote_data_y
    [smote_data_y['Died'] == 1])/len(smote_data_X)*100)
X = smote_data_X
y = smote_data_y

# ===== Multicollinearity =====

## Correlation
import seaborn as sb
X_numerical_features = X[numerical_features]
data_corr = X_numerical_features.corr()
pd.set_option('display.max_columns', None)
pd.set_option('display.max_rows', None)
data_corr
sb.heatmap(data_corr)
def corration(data, threshold):
    correlated_cols = set()
    corr_matrix = data.corr()
    for i in range(len(corr_matrix.columns)):
        for j in range(i):
            if abs(corr_matrix.iloc[i, j]) > threshold:
                colname = corr_matrix.columns[i]
                correlated_cols.add(colname)
    return correlated_cols
features_corr = corration(X, 0.8)
X.drop(labels = features_corr, axis=1, inplace = True )
features_corr

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

# ===== Feature selection =====

## Backward Elimination
X = X[['Day_between_detected_and_discharge', 'Age', 'Nationality', 'Respiration_rate',
      'AstraZeneca', 'Invasive_mechanical_ventilation',
      'Non_invasive_mechanical_ventilation', 'Dexamethasone', 'Methylprednisolone',
      'Hydrocortisone', 'Patchy_opacity', 'Right', 'Diabetes', 'Heart_diseases',
      'Kidney_diseases', 'Cancer']]

# ===== Split training set and test set =====

from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.3,
                                                  random_state = random_seed)

# ===== Model =====

## Binary Logistic Regression
import statsmodels.formula.api as smf
from sklearn.linear_model import LogisticRegression
df_LoR = pd.DataFrame(X_train)
df_LoR = df_LoR.join(y_train)
clf = smf.logit("Died ~ Day_between_detected_and_discharge + Age + Nationality +
              Respiration_rate + AstraZeneca + Invasive_mechanical_ventilation +
              Non_invasive_mechanical_ventilation + Dexamethasone + Methylprednisolone
              + Hydrocortisone + Patchy_opacity + Right + Diabetes + Heart_diseases +
              Kidney_diseases + Cancer", data = df_LoR).fit()

print(clf.summary())
clf = LogisticRegression(penalty = 'none', random_state = random_seed)
clf.fit(X_train, y_train)

## Decision Tree
from sklearn.tree import DecisionTreeClassifier
from sklearn import tree
clf = DecisionTreeClassifier(criterion = 'entropy', splitter = 'best', random_state =
                          random_seed)
clf.fit(X_train, y_train)
tree.plot_tree(clf)

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Neural network

```

from sklearn.neural_network import MLPClassifier
import matplotlib.pyplot as plt
clf = MLPClassifier(solver = "sgd",
                    activation = "relu",
                    hidden_layer_sizes = (500, 1000, 500),
                    learning_rate_init = 0.1,
                    max_iter = 1000,
                    batch_size = X_train.shape[0],
                    verbose = 1)
clf.fit(X_train, y_train)
plt.plot(clf.loss_curve_)
# ===== Performance measures =====
from sklearn.metrics import classification_report
from sklearn import metrics
import matplotlib.pyplot as plt
clf.score(X_test, y_test)
y_pred = clf.predict(X_test)
print(classification_report(y_test, y_pred))
confusion_matrix = metrics.confusion_matrix(y_test, y_pred)
confusion_matrix
cm_display = metrics.ConfusionMatrixDisplay(confusion_matrix = confusion_matrix,
display_labels = [False, True])
cm_display.plot()
plt.show()
## ROC Cuve
from sklearn.metrics import roc_auc_score
from sklearn.metrics import roc_curve
logit_roc_auc = roc_auc_score(y_test, clf.predict(X_test))
fpr, tpr, thresholds = roc_curve(y_test, clf.predict_proba(X_test)[:, 1])
plt.figure()
plt.plot(fpr, tpr, label='AUC(area = %0.2f)' % logit_roc_auc)
plt.plot([0, 1], [0, 1], color="navy", linestyle = "--")

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

plt.xlim([0.0, 1.0])
plt.ylim([0.0, 1.05])
plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')
plt.title('Receiver operating characteristic')
plt.legend(loc="lower right")
plt.show()

# ===== Feature Importance =====

##Binary Logistic Regression
w = clf.coef_[0]

##Decision Tree
w = clf.feature_importances_

##Neural network
from sklearn.inspection import permutation_importance
perm = permutation_importance(clf, X_train, y_train, random_state = random_seed)
w = perm.importances_mean

##Important Features
feature_importance = pd.DataFrame(X_train.columns.tolist(), columns = ["feature"])
feature_importance["importance"] = w
feature_importance = feature_importance.sort_values(by = ["importance"], ascending
= False)

feature_importance = feature_importance[:len(X_train.columns)].sort_values(by =
["importance"], ascending=False)

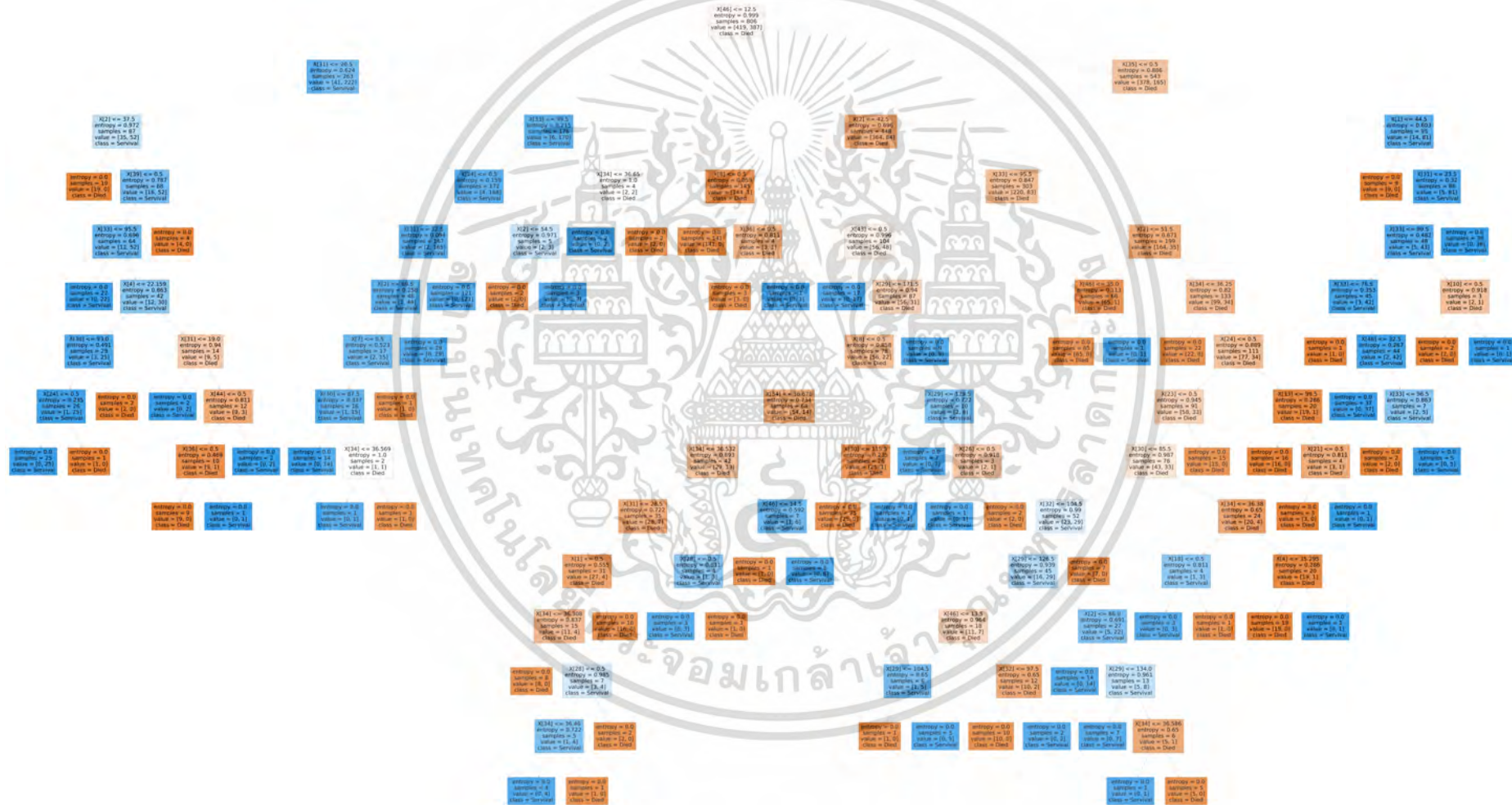
##Visualiztion
ax = feature_importance.sort_values(by = ["importance"], ascending =
True).plot.barh(x = 'feature', y = 'importance')
plt.title('Important Features')
plt.show()

##Important features
feature_importance

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

แผนภาพต้นไม้ตัดสินใจ



ประวัติผู้เขียน

ชื่อ นายภูริทัต เมฆวิไลพันธ์
 วัน เดือน ปีเกิด 26 พฤษภาคม พ.ศ.2539
 ที่อยู่ปัจจุบัน 9/4 หมู่ 1 ตำบลหนองตะพาน อำเภอบ้านค่าย จังหวัดระยอง
 ประวัติการศึกษา (2560) วิทยาศาสตรบัณฑิต สาขาคณิตศาสตร์ประยุกต์ เกรดเฉลี่ย 3.28
 มหาวิทยาลัยเกษตรศาสตร์



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
แบบฟอร์มการตรวจสอบการคัดลอกผลงานทางวิชาการ

ข้าพเจ้า (นาย/นาง/นางสาว) ...**ภูริทัต เมฆวิไลพันธ์**.....รหัสนักศึกษา.....**63605133**.....
ระดับ (ปริญญาตรี/ปริญญาโท/ปริญญาเอก) สาขาวิชา...**สถิติและการวิเคราะห์ธุรกิจ**... ภาควิชา...**สถิติ**.....
คณะ...**วิทยาศาสตร์**.....

ได้เสนอ

วิทยานิพนธ์	โครงการพิเศษ	สหกิจศึกษา
<input checked="" type="checkbox"/> การค้นคว้าอิสระ	ปัญหาพิเศษ	เทียบเท่า ระบุ.....
ปริญญาานิพนธ์	การศึกษาอิสระ	

หัวข้อเรื่อง

(ไทย).....**ปัจจัยที่สัมพันธ์ต่อการเสียชีวิตของผู้ป่วยโควิด-19 กรณีศึกษา โรงพยาบาลระยอง**.....
(อังกฤษ).....**Factors Associated with Deaths of COVID-19 Patients: Case Study of Rayong Hospital**.....

ได้ตรวจเช็คผลงานวิชาการข้างต้นแล้ว ในภาคเรียนที่...**1/2565**...วันที่...**4**...เดือน...**มกราคม**...ปี...**2566**.....

โดยใช้โปรแกรม

อักษราวิสุทธิ์ **TURNITIN**

ทั้งนี้ ตรวจสอบพบความเหมือนของเนื้อหา.....**10.78**.....% โดยอาจารย์ที่ปรึกษายอมรับได้ว่าไม่ได้
คัดลอกข้อความที่มีสาระสำคัญจากผลงานของผู้อื่น

ลายมือชื่อนักศึกษา..........

(**นายภูริทัต เมฆวิไลพันธ์**)

วันที่.....**4 มกราคม 2566**.....

ได้รับความเห็นชอบจากอาจารย์ที่ปรึกษา

ลายมือชื่ออาจารย์ที่ปรึกษา.....**พรพิมล ชัยวุฒิศักดิ์**.....

(**ผศ.ดร.พรพิมล ชัยวุฒิศักดิ์**)

วันที่.....**4 มกราคม 2566**.....

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้