

## ระบบการจำแนกผู้ป่วยโควิด-19 ด้วยเทคนิคป่าสุ่ม



ศิวกร บรรลือทรัพย์

DPU

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรมหาบัณฑิต

สาขาวิชาวิศวกรรมเว็บและการพัฒนาแอปพลิเคชันบนอุปกรณ์พกพา

วิทยาลัยครีเอทีฟดีไซน์ แอนด์ เอ็นเตอร์เทนเมนต์เทคโนโลยี มหาวิทยาลัยธุรกิจบัณฑิตย์

ปีการศึกษา 2564

**A SYSTEM FOR CLASSIFYING COVID-19 PATIENS WITH  
RANDOM FOREST TESCHNIQUE**

**SIWAKORN BANLUESAPY**

**A Thesis Submitted in Partial Fulfilment of the Requirements for the Degree of  
Master of Science Department of Web Engineering and Mobile Application  
Development ,College of Creative Design and Entertainment Technology,**

**Dhurakij Pundit University**

**Academic Year 2021**



## ใบรับรองวิทยานิพนธ์

วิทยาลัยครีเอทีฟดีไซน์ แอนด์ เอ็นเตอร์เทนเมนต์เทคโนโลยี มหาวิทยาลัยธุรกิจบัณฑิตย์  
ปริญญา วิทยาศาสตรมหาบัณฑิต

หัวข้อวิทยานิพนธ์ ระบบการจำแนกผู้ป่วยโควิด-19 ด้วยเทคนิคป่าสุ่ม  
เสนอโดย ศิวกร บรรลือทรัพย์  
สาขาวิชา วิศวกรรมเว็บและการพัฒนาแอปพลิเคชันบนอุปกรณ์พกพา  
อาจารย์ที่ปรึกษา รองศาสตราจารย์ ดร.วราพร จิระพันธุ์ทอง

ได้พิจารณาเห็นชอบโดยคณะกรรมการสอบสารนิพนธ์แล้ว

ลงชื่อ ..... ประธานกรรมการ  
(ผู้ช่วยศาสตราจารย์ ดร.อรวรรณ อิ่มสมบัติ)

ลงชื่อ ..... กรรมการและอาจารย์ที่ปรึกษาวิทยานิพนธ์  
(รองศาสตราจารย์ ดร.วราพร จิระพันธุ์ทอง)

ลงชื่อ ..... กรรมการ  
(ผู้ช่วยศาสตราจารย์ ดร.วิลาวัลย์ อินทร์ชำนาญ)

ลงชื่อ ..... กรรมการ  
(รองศาสตราจารย์ ดร.นุชรี เปรมชัยสวัสดิ์)

วิทยาลัยครีเอทีฟดีไซน์ แอนด์ เอ็นเตอร์เทนเมนต์เทคโนโลยี

..... คณบดี  
(ผู้ช่วยศาสตราจารย์ ดร.วิลาวัลย์ อินทร์ชำนาญ)  
วันที่ ..... เดือน ..... พ.ศ. ๖๕.....

หัวข้อวิทยานิพนธ์	ระบบการจำแนกผู้ป่วยโควิด-19 ด้วยเทคนิคป่าสุ่ม
ชื่อผู้เขียน	ศิวกร บรรลือทรัพย์
อาจารย์ที่ปรึกษา	รองศาสตราจารย์ ดร.วราพร จิระพันธุ์ทอง
หลักสูตร	วิศวกรรมเว็บและการพัฒนาแอปพลิเคชันบนอุปกรณ์พกพา
ปีการศึกษา	2564

### บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อสร้างแบบจำลองสำหรับพัฒนาระบบจำแนกผู้ป่วยโควิด-19 ในรูปแบบเว็บแอปพลิเคชันโดยใช้เทคนิคการเรียนรู้ของเครื่อง โดยจะทำการเปรียบเทียบประสิทธิภาพของอัลกอริทึมในการการจำแนกผู้ป่วยโควิด-19 และได้ทำการศึกษาค้นคว้าข้อมูลผู้ป่วย จำนวน 1,608,923 ราย จากกรมควบคุมโรค โดยทำการรวบรวมข้อมูล 1 มกราคม 2563 ถึง วันที่ 1 ตุลาคม 2564 โดยใช้ อัลกอริทึมการจำแนกข้อมูลทั้งหมด 3 แบบ ได้แก่ Random Forest, Neural network และ Naive Bayes ซึ่งทำการเปรียบเทียบประสิทธิภาพรูปแบบเทคนิคการทำนายระหว่างคุณลักษณะ การทดสอบประสิทธิภาพตัวแบบทำนายด้วยวิธีการ Split Test คือ การแบ่งข้อมูลด้วยการสุ่มออกเป็น 2 ส่วนคือ Training Data (ร้อยละ 80) และ Test Data (ร้อยละ 20) โดยใช้โปรแกรม Orange Data mining จากนั้นทำการทดลองเพื่อหาผลการทดสอบประสิทธิภาพที่มีค่าความถูกต้องที่สูงที่สุด ผลการวิจัยพบว่า Random Forest มีค่าความถูกต้องเท่ากับ 93.33% Neural network มีค่าความถูกต้องเท่ากับ 92.7% Naive Bayes มีค่าความถูกต้องเท่ากับ 92.1% จากผลการเปรียบเทียบในครั้งนี้นี้จึงนำ Random Forest ที่มีค่าสูงสุดไปใช้ในการพัฒนาโปรแกรมระบบในการจำแนกผู้ป่วยโควิด-19 ได้



Thesis Title	A SYSTEM FOR CLASSIFYING COVID-19 PATIENTS WITH RANDOM FOREST TECHNIQUE
Author	Siwakorn Banluesapy
Thesis Advisor	Assoc. Prof. Dr. Waraporn Jirapanthong
Department	Web Engineering and Mobile Application Development
Academic Year	2021

### ABSTRACT

This research aims to create a model for the development of a COVID-19 patient classification system, as a web application using machine learning techniques. It compares the efficiency of the algorithm for classifying COVID-19 patients, and has studied data on 1,608,923 patients from the Department of Disease Control. By collecting data from 1 January 2020 to 1 October 2021. It uses a total of three classification algorithms: Random Forest, Neural network and Naive Bayes. We apply the split test method to evaluate the models' prediction. The data is divided into two parts: training data (80%) and test data (20%). We also apply a software tool, namely Orange Data Mining, for learning process. It is found that the prediction results return with high rate of accuracy. Particularly, the results show that Random Forest, Neural network, and Naive Bayes models have performed the prediction with the accuracy of 93.33%, 92.7%, and 92.1%, consequently. We then apply the Random Forest model for supporting a system of the classification of COVID-19 patients

## กิตกรรมประกาศ

การจัดทำวิทยานิพนธ์ฉบับนี้ สำเร็จลุล่วงได้ด้วยคำแนะนำและความกรุณาของ ท่านอาจารย์ รศ.ดร. วราพร จิระพันธุ์ทอง อาจารย์ที่ปรึกษาวิทยานิพนธ์ ซึ่งท่านได้สละเวลาอันมีค่ามาคอยให้ คำปรึกษาและชี้แนะแนวทางอันเป็นประโยชน์ในการทำวิทยานิพนธ์ อีกทั้งยังให้ความกรุณาเอาใจใส่ และให้กำลังใจอย่างสม่ำเสมอ

ขอกราบขอบพระคุณท่านประธานกรรมการ และท่านคณะกรรมการสอบวิทยานิพนธ์ ตลอดจนคณาจารย์และบุคลากรวิทยาลัยครีเอทีฟดีไซน์แอนด์อินเทอร์เน็ตเทคนิคมেন্টเทคโนโลยี ทุกท่านที่ได้สอนวิชาการความรู้ต่าง ๆ แนะนำการแก้ไขปัญหาต่างๆ และประสิทธิประสาทวิชาความรู้ในด้านต่าง ๆ เป็นอย่างดี ผู้วิจัยขอกราบขอบพระคุณเป็นอย่างสูง ณ โอกาสนี้

ขอขอบพระคุณเพื่อน ๆ พี่ ๆ และน้อง ๆ สาขาวิชาวิศวกรรมเว็บและการพัฒนาแอปพลิเคชัน บนอุปกรณ์พกพา วิทยาลัยครีเอทีฟดีไซน์แอนด์อินเทอร์เน็ตเทคนิคมেন্টเทคโนโลยี ทุกท่านที่ให้ความช่วยเหลือ และเป็นกำลังใจในการทำวิทยานิพนธ์จนสำเร็จลุล่วง

ขอขอบพระคุณห้องปฏิบัติการสาขาวิชาวิศวกรรมเว็บและการพัฒนาแอปพลิเคชันบนอุปกรณ์ พกพา วิทยาลัยครีเอทีฟดีไซน์แอนด์อินเทอร์เน็ตเทคนิคมেন্টเทคโนโลยี มหาวิทยาลัยธุรกิจบัณฑิตย์ซึ่ง ให้อำนาจสถานที่ในการทดลอง และจัดทำวิทยานิพนธ์จนมีความสำเร็จลุล่วงไปด้วยดี

ขอกราบขอบพระคุณคุณพ่อ คุณแม่ และผู้มีพระคุณที่คอยให้การสนับสนุนช่วยเหลือในทุก ๆ ด้านและคอยให้กำลังใจเสมอมา พร้อมได้อบรมปลูกฝังให้ลูกมีสำนึกที่จะทำดีและเป็นคนดีโดยเฉพาะ การให้การศึกษาที่ดีแก่ผู้จัดทำตลอดมา หากความสำเร็จใด ๆ ที่มีในชีวิต ขอมอบให้แก่ผู้มีพระคุณทุก ท่านที่กล่าวมาข้างต้นทั้งหมด

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ฉ
บทคัดย่อภาษาอังกฤษ.....	ง
กิตติกรรมประกาศ.....	จ
สารบัญตาราง.....	ช
สารบัญภาพ.....	ฉ
บทที่.....	
1. บทนำ.....	1
1.1 หลักการและเหตุผล.....	1
1.2 ปัญหาและแรงจูงใจ.....	3
1.3 แนวทางในการแก้ปัญหา.....	3
1.4 วัตถุประสงค์ของการวิจัย.....	4
1.5 ขอบเขตการศึกษา.....	4
1.6 ประโยชน์ที่คาดว่าจะได้รับจากงานวิจัย.....	4
1.7 เครื่องมือและภาษาที่ใช้ในการพัฒนา.....	5
2. แนวคิด ทฤษฎี และงานวิจัยที่เกี่ยวข้อง.....	6
2.1 การประยุกต์ใช้การเรียนรู้ของเครื่องสำหรับโรคโควิด-19.....	6
2.2 Open Data - Open Government Data Center.....	7
2.3 ทฤษฎีและแนวคิดพื้นฐานเกี่ยวกับโรคโควิด-19.....	12
2.4 การเรียนรู้ของเครื่อง (Machine Learning).....	14
2.5 Random Forest.....	17
2.6 Naive Bayes.....	18
2.7 Neural network.....	20
2.8 ตัววัดประสิทธิภาพของแบบจำลองการจำแนกประเภทข้อมูล.....	22
2.9 Data Visualization (การสร้างมโนภาพ) .....	25

## สารบัญ (ต่อ)

บทที่	หน้า
2.10 การพัฒนา Application ตามหลัก Microservice .....	29
2.11 งานวิจัยที่เกี่ยวข้อง .....	31
3. วิธีการดำเนินงานวิจัย.....	37
3.1 Dataset (วิธีการรวบรวมข้อมูล).....	40
3.2 Data Preparation (การเตรียมข้อมูล).....	42
3.3 Modeling (การสร้างแบบจำลอง).....	51
3.4 Evaluation (การวัดประสิทธิภาพแบบจำลอง).....	53
3.5 Deployment (การนำแบบจำลองไปใช้งาน).....	53
4. ผลการทดลองและอภิปรายผล.....	57
4.1 โครงสร้างข้อมูล.....	57
4.2 การใช้โปรแกรม Orange Data Mining เพื่อเปรียบเทียบแบบจำลอง.....	58
4.3 สรุปผลการเปรียบเทียบค่าความถูกต้องและความแม่นยำในแต่ละเทคนิค.....	63
4.4 การนำแบบจำลองไปใช้งาน.....	65
4.5 API สำหรับการทำนาย.....	67
4.6 ส่วนติดต่อผู้ใช้งาน.....	69
5. สรุปและข้อเสนอแนะ.....	74
5.1 สรุปผลการวิจัย.....	74
5.2 สรุปปัญหาและวิธีการแก้ปัญหา.....	75
5.3 ข้อเสนอแนะ.....	76
บรรณานุกรม.....	77
ภาคผนวก.....	85
ภาคผนวก ก. API สำหรับการทำนาย.....	86
ภาคผนวก ข. ส่วนติดต่อผู้ใช้งาน.....	105
ประวัติผู้เขียน.....	112

## สารบัญตาราง

ตารางที่	หน้า
2.1 ชุดข้อมูลบริการรายงาน โควิด-19.....	9
2.2 พจนานุกรมข้อมูลข้อมูลผู้ป่วยโควิด.....	11
2.3 แสดงตาราง Confusion matrix.....	23
2.4 เปรียบเทียบการใช้เทคนิคการเรียนรู้ของเครื่องในงานวิจัยที่ได้ทำการศึกษา.....	34
3.1 คำอธิบาย ชุดข้อมูลผู้ป่วยโควิด-19.....	41
3.2 คำอธิบายข้อมูล patient_type.....	47
3.3 คลาสในการจำแนก.....	50
3.4 แอททริบิวต์ในการวิจัย.....	51
4.1 อัตราส่วนของข้อมูลในแต่ละคลาส.....	59
4.2 Information Gain ของ แอททริบิวต์ในการวิจัย.....	60
4.3 ตัววัดประสิทธิภาพของอัลกอริทึม Naive Bayes.....	61
4.4 ตัววัดประสิทธิภาพของอัลกอริทึม Neural Network.....	62
4.5 ตัววัดประสิทธิภาพของอัลกอริทึม Random Forest.....	63
4.6 พารามิเตอร์ของ Random Forest.....	65

สารบัญภาพ

ภาพที่	หน้า
2.1 เว็บไซต์ข้อมูลบริการรายงานโควิด-19 <a href="https://covid19ddcmophgoth/">https://covid19ddcmophgoth/</a> .....	10
2.2 โปรแกรม Postman ทดสอบ API ในการดึงข้อมูลผู้ป่วยโควิด-19 .....	11
2.3 แสดงอาการผู้ป่วยโควิด-19.....	13
2.4 ตัวอย่างการเรียนรู้แบบมีผู้สอน.....	15
2.5 ตัวอย่างการเรียนรู้แบบไม่มีผู้สอน .....	16
2.6 หลักการทำ Random Forest .....	18
2.7 แบบจำลองการทำงานของ Neural network.....	20
2.8 แผนภูมิแท่ง (Bar Charts) .....	26
2.9 แผนภูมิเส้น (Line Chart) .....	27
2.10 แผนภูมิวงกลม (Pie Charts) .....	28
2.11 แผนภูมิพื้นที่ (Area Charts) .....	28
2.12 สถาปัตยกรรมแบบ Microservice.....	29
3.1 กระบวนการทำงานวิจัย .....	38
3.2 แผนผังกระบวนการทำงานวิจัย .....	39
3.3 ขั้นตอนการพัฒนาส่วนการดาวน์โหลดข้อมูล.....	40
3.4 ตารางข้อมูลโควิด-19 จาก API .....	40
3.5 ตัวอย่างชุดข้อมูลในรูปแบบไฟล์ Microsoft Excel.....	41
3.6 การจัดกลุ่ม gender (เพศ) .....	43
3.7 การจัดกลุ่ม age_range (ช่วงอายุ) .....	43
3.8 การจัดกลุ่ม age_range (ช่วงอายุ) (ต่อ) .....	44
3.9 การจัดกลุ่ม nationality (สัญชาติ) .....	44
3.10 การจัดกลุ่ม risk (ความเสี่ยง) .....	45

สารบัญภาพ (ต่อ)

ภาพที่	หน้า
3.11 การจัดกลุ่ม risk (ความเสี่ยง) (ต่อ) .....	45
3.12 การจัดกลุ่ม risk (ความเสี่ยง) (ต่อ) .....	46
3.13 การจัดกลุ่ม province (จังหวัด) .....	46
3.14 การจัดกลุ่ม patient_type (ประเภทผู้ป่วย) .....	47
3.15 ชุดข้อมูลก่อนทำ Data cleaning .....	50
3.16 แบบจำลองการวิเคราะห์คัดแยกผู้ป่วยโควิด-19 .....	52
3.17 การนำแบบจำลองไปใช้งาน .....	54
3.18 Prediction API ในการคัดแยกผู้ป่วยโควิด-19 .....	55
4.1 โครงสร้างข้อมูล.....	57
4.2 การใช้โปรแกรม Orange Data Mining เพื่อเปรียบเทียบแบบจำลอง.....	58
4.3 การใช้โปรแกรม Orange Data Mining ในการ Split Test .....	59
4.4 การใช้ Confusion Matrix Widget ในการวัดค่า .....	61
4.5 การเปรียบเทียบความแม่นยำในแต่ละอัลกอริทึม .....	64
4.6 การบันทึกแบบจำลอง Random Forest .....	65
4.7 การนำแบบจำลองไปใช้งานกับข้อมูลใหม่ผ่านโปรแกรม Orange Data Mining....	66
4.8 การนำแบบจำลองไปใช้งานกับข้อมูลใหม่ด้วยภาษา Python.....	66
4.9 Fast API (Web Service) .....	67
4.10 การส่งข้อมูลไปทำนายผ่าน Swagger UI .....	68
4.11 รูปแบบการส่งข้อมูลไปทำนาย .....	68
4.12 ผลของการทำนาย .....	69
4.13 แบบฟอร์มวิเคราะห์ข้อมูลผู้ป่วย.....	69
4.14 แบบฟอร์มบันทึกข้อมูลผู้ป่วย.....	70
4.15 ข้อมูลผู้ป่วย .....	70
4.16 ยอดผู้ป่วยในแต่ละเดือน .....	71



## สารบัญภาพ (ต่อ)

ภาพที่	หน้า
4.17 จำนวน Instance ในแต่ละคลาส .....	71
4.18 พื้นที่การติดเชื่อ .....	72
4.19 ยอดรายวัน.....	73
4.20 pgAdmin 4 ระบบฐานข้อมูล PostgreSQL .....	73
ก.1 User.....	87
ก.2 การเรียกดูข้อมูลผู้ใช้งานทั้งหมด .....	87
ก.3 การเพิ่มข้อมูลผู้ใช้งาน .....	88
ก.4 การกำหนดช่วงการดึงข้อมูลผู้ใช้งาน .....	89
ก.5 การแสดงข้อมูลผู้ใช้งานแบบแบ่งหน้า.....	89
ก.6 การกำหนดรหัสผู้ใช้งานเพื่อดูข้อมูลเฉพาะ .....	90
ก.7 ข้อมูลผู้ใช้งาน โดยดึงผ่านรหัสผู้ใช้งาน .....	91
ก.8 แก้ไขข้อมูลผู้ใช้งาน โดยดึงผ่านรหัสผู้ใช้งาน .....	91
ก.9 ข้อตอบกลับหลังจากแก้ไขข้อมูล .....	92
ก.10 การส่งรหัสข้อมูลผู้ใช้งานเพื่อทำการลบ .....	92
ก.11 การส่งข้อมูลเพื่อตรวจสอบตัวตนในการเข้าสู่ระบบ.....	92
ก.12 Patient .....	93
ก.13 การกำหนดช่วงการดึงข้อมูลผู้ป่วย .....	93
ก.14 การแสดงข้อมูลผู้ใช้งานแบบแบ่งหน้า .....	94
ก.15 การกำหนดรหัสผู้ป่วยเพื่อดูข้อมูลเฉพาะ .....	95
ก.16 ข้อมูลผู้ป่วย โดยดึงผ่านรหัสผู้ป่วย.....	95
ก.17 การเพิ่มข้อมูลผู้ป่วย .....	96
ก.18 การส่งรหัสและข้อมูลผู้ป่วยเพื่อแก้ไขข้อมูล.....	97

## สารบัญภาพ (ต่อ)

ก.19 การเพิ่มข้อมูลที่อยู่ผู้ป่วย.....	97
ก.20 การแก้ไขข้อมูลที่อยู่ผู้ป่วย.....	99
ก.21 การจำแนกข้อมูล.....	100
ก.22 ผลลัพธ์ในการจำแนกข้อมูล .....	101
ก.23 MasterData .....	101
ก.24 การแสดงข้อมูลช่วงอายุ.....	101
ก.25 การแสดงข้อมูลประเภทความเสี่ยง .....	102
ก.26 การแสดงข้อมูลประเภทผู้ป่วย.....	102
ก.27 Dashboard.....	103
ก.28 การกำหนดปีแสดงข้อมูลผู้ป่วยรายเดือน .....	103
ก.29 จำนวนยอดผู้ป่วยรายเดือนต่อปี.....	103
ก.30 จำนวนยอดผู้ป่วยโดยแยกตามจังหวัดต่อปี .....	104
ข.1 Cookies Banner .....	106
ข.2 การเข้าสู่ระบบ .....	106
ข.3 แบบฟอร์มวิเคราะห์ข้อมูลผู้ป่วย.....	107
ข.4 แบบฟอร์มบันทึกข้อมูลผู้ป่วย .....	107
ข.5 ข้อมูลผู้ป่วยใหม่.....	108
ข.6 ข้อมูลผู้ป่วยเดิม .....	108
ข.7 กราฟแสดงยอดผู้ป่วยในแต่ละเดือน .....	109
ข.8 จำนวน Instance ในแต่ละคลาส .....	109
ข.9 พื้นที่ติดเชื้อ .....	110
ข.10 ยอดผู้ติดเชื้อ โควิด-19 รายวัน .....	111

## บทที่ 1

### บทนำ



#### 1.1 หลักการและเหตุผล

องค์การอนามัยโลกได้รับแจ้งเมื่อเดือนธันวาคม 2562 ว่าพบกลุ่มผู้ป่วยมีอาการปอดบวมโดยไม่ทราบสาเหตุในเมืองอู่ฮั่น มณฑลหูเป่ย์ สาธารณรัฐประชาชนจีน ต่อมา มีรายงานว่าพบเชื้อไวรัสโคโรนาสายพันธุ์ใหม่หรือ Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV -2) เป็นสาเหตุของโรคและได้แพร่ระบาดไปยังประเทศอื่น ๆ ซึ่งโรคติดเชื้อไวรัสโคโรนา 2019 หรือ COVID - 19 นี้ มีลำดับนิวคลีโอไทด์เหมือนไวรัสซาร์สที่เคยระบาดมาก่อนถึงร้อยละ 80 โดยที่เชื้อ SARS -Co V -2 ก่อโรคปอดบวม สามารถทำให้เกิดระบบหายใจล้มเหลวได้ ผู้ป่วยมีอัตราการเสียชีวิตประมาณร้อยละ 3.4 ผู้สูงอายุที่มีโรคประจำตัวเป็นปัจจัยเสี่ยงเพิ่มความรุนแรง (Manmana et al., 2020) โดยโรคโควิด-19 เป็นโรคติดต่อสายพันธุ์ใหม่ และได้มีการคาดการณ์ว่าเป็นการติดเชื้อจากสัตว์สู่คนโดยมีค้างคาวเป็นพาหะเป็นโรคได้ทั้งคนและสัตว์ (Manmana et al., 2020; Saowapa et al., 2020) เชื้อสามารถติดต่อจากคนสู่คนผ่านทางเดินหายใจจากละอองของเสมหะจากการ ไอ จาม น้ำมูก น้ำลายในระยะประมาณ 1-2 เมตรและการสัมผัสของละอองเชื้อที่ตกอยู่ตามพื้นผิววัสดุ แล้วนำมาสัมผัสปาก จมูก ตา เมื่อได้รับเชื้อโรคผ่านละอองสารคัดหลั่งดังกล่าวแล้วจะทำให้เกิดการติดเชื้อ ซึ่งโดยทั่วไปประมาณว่าคนเราจะสามารถต้านทานไวรัสนี้ได้ประมาณร้อยละ 2 และจะเกิดส่งผลต่อระบบทางเดินหายใจร้ายแรงเมื่อเข้าสู่ร่างกายร้อยละ 5-10 (Rachada, 2020)

สำหรับการระบาดของเชื้อไวรัสโควิด-19 รอบแรกในประเทศไทย พบผู้ป่วยต้องสงสัยรายแรกเมื่อวันที่ 21 มกราคม พ.ศ. 2563 เป็นนักท่องเที่ยวหญิงชาวจีนอายุ 74 ปี ซึ่งเดินทางมาถึง

กรุงเทพมหานครโดยเที่ยวบินจากนครอู่ฮั่น เมืองหลวงของมณฑลหูเป่ย์ สาธารณรัฐประชาชนจีน และ ณ วันที่ 8 เมษายน พ.ศ. 2563 พบอัตราการติดเชื้อในประเทศไทย จำนวน 2,369 คน (Natawan, 2021) หลังจากนั้นพบการระบาดระลอกใหม่ช่วงปลายปี 2563 เริ่มตั้งแต่วันที่ 17 ธันวาคม 2563 พบหญิงไทย อายุ 67 ปี อาชีพค้าขายที่ตลาดกลางกุ้ง ในตำบลมหาชัย อำเภอมหาชัย จังหวัดสมุทรสาคร (โชคกรรชิตไชย, 2563) จากรายงานสำรวจยอดผู้ติดเชื้อตั้งแต่เกิดการระบาดครั้งแรกจนถึงเดือน ตุลาคม 2564 มียอดติดเชื้อรวมที่ยืนยันแล้ว 1,730,364 ราย ยอดเสียชีวิต 17,835 ราย (World Health Organization Thailand, 2564)

กระบวนการตรวจ วินิจฉัยโรคโควิด 19 เป็นการตรวจทางห้องปฏิบัติการ ต้องเสียเวลาในการเตรียมเชื้อและกระบวนการตรวจ เพราะต้องตรวจกับตัวเชื้อจริง ในช่วงแรกของการระบาด ต้องใช้เวลาไม่น้อยกว่า 2 วัน และทำได้เฉพาะในโรงพยาบาลใหญ่หรือโรงพยาบาลศูนย์ ค่าตรวจมีราคาแพง เพราะผลที่ได้ต้องได้ผลชัดเจน และทางกระทรวงสาธารณสุขมีข้อจำกัดในเรื่องจำนวนการตรวจ จึงต้องคัดกรองผู้ต้องสงสัยก่อนเก็บเชื้อส่งตรวจ และใช้วิธีนี้ตั้งแต่เริ่มการระบาด คือ ต้องเก็บเชื้อจากสารคัดหลั่งที่โพรงจมูก ดังนั้นในกระบวนการควบคุมโรค การค้นหา การตรวจหาผู้ป่วย จำเป็นต้องมีกระบวนการคัดกรอง เพื่อคัดกรองเป็นลำดับในการหาผู้สงสัยก่อน (Kasetsart university, 2020) ดังนั้นผู้วิจัยจึงทำการศึกษาและวิเคราะห์ข้อมูลผู้ติดเชื้อ โควิด-19 จากกรมควบคุมโรคตั้งแต่ระลอกที่ 1 – 3 ซึ่งเป็นข้อมูลแบบเปิด โดยทำการรวบรวมข้อมูลตั้งแต่วันที่ 1 มกราคม 2563 ถึง วันที่ 1 ตุลาคม 2564 เป็นจำนวน 1,608,923 ราย (Department of disease control, 2021) ซึ่งเป็นข้อมูลในลักษณะไฟล์คอมพิวเตอร์ (.csv) และ รูปแบบ JSON เพื่อช่วยจำแนกประเภทผู้ป่วยได้เร็วขึ้น

การเรียนรู้ของเครื่อง (Machine Learning) (Puitrakul, 2018) คือ การที่สามารถทำให้คอมพิวเตอร์สามารถเรียนรู้ได้ด้วยตนเอง โดยการเขียนโปรแกรมหรือการสร้างอัลกอริทึมที่สามารถเรียนรู้และพยากรณ์ข้อมูลได้ ทั้งนี้อัลกอริทึมจะต้องอาศัยแบบจำลองที่ได้จากการสร้างชุดของข้อมูลเพื่อพยากรณ์หรือตัดสินใจ

ในการใช้การเรียนรู้ของเครื่องในการวิเคราะห์ข้อมูลที่เกี่ยวข้องกับโรคโควิด-19 นั้นเริ่มมีการถูกใช้อย่างแพร่หลาย เช่น รูปแบบการตรวจจับโควิด-19 แบบอัตโนมัติโดยภาพเอ็กซ์เรย์ทรวงอกดิบ และสร้างแบบจำลอง โดยใช้ Deep Neural networks เพื่อการจำแนกโรคโควิด-19 และ อาการปอดบวม

(Ozturk et al., 2020) การวิเคราะห์โอกาสที่ผู้ป่วยจะมีการอาการรุนแรง/อาการวิกฤตจากโควิด-19 จากข้อมูล 336 รายจากประเทศจีนที่ใช้โดยใช้แบบจำลองแบบจำลอง Support Vector Machine (SVM) (Sun et al., 2020) และการระบุการติดเชื้อโควิด-19 อย่างรวดเร็วเพื่อคัดกรองผู้ต้องสงสัยโดยทันทีโดยใช้แบบจำลองแบบจำลอง Random forest กับการวิเคราะห์จากผลการตรวจเลือด (Wu et al., 2020)

จากงานวิจัยที่ได้ทำการศึกษามาส่วนใหญ่จะทำการจำแนกข้อมูลหรือทำนายโอกาสการเกิดโรคโควิด-19 เท่านั้น ยังไม่มีงานวิจัยที่ใช้เทคนิคการเรียนรู้ของเครื่องมาช่วยในการจำแนกประเภทผู้ป่วย โดยส่วนใหญ่จะเป็นงานในลักษณะ Classification อยู่ในรูปแบบแบบจำลองประเภท Supervised ซึ่งมีตัวแปรที่ใช้ในการวัดเป้าหมาย งานวิจัยนี้จึงสนใจวิเคราะห์ข้อมูลโควิด-19 โดยการจำแนกประเภทผู้ป่วย เพื่อคัดกรองผู้ต้องสงสัยได้เร็วขึ้น โดยจะเป็นการตรวจสอบสำหรับการเฝ้าระวังในกลุ่มต่างๆ โดยจะแบ่งเป็น 4 กลุ่มคือ การเฝ้าระวังในกลุ่มผู้ป่วย หรือมีอาการเข้าได้กับนิยาม (Patient Under Investigation : PUI) การตรวจคัดกรองในประชากรเสี่ยงตามจุดคัดกรองและด่านเข้าออกระหว่างประเทศ (Screening) การเฝ้าระวังในกลุ่มเป้าหมายเฉพาะหรือพื้นที่เฉพาะ (Sentinel Surveillance) การเฝ้าระวังเหตุการณ์ในสถานที่เสี่ยง เก็บตัวอย่างส่งตรวจเมื่อเข้าเกณฑ์ PUI และเป็นกลุ่มก้อน เพื่อเตรียมการรักษาได้ทันเวลา

## 1.2 ปัญหาและแรงจูงใจ

1.2.1 ข้อมูลที่รวบรวมได้ยังเป็นข้อมูลที่มีความซ้ำซ้อนของข้อมูล และการไม่ถูกจัดกลุ่มอย่างถูกต้อง

1.2.2 เนื่องจากมีผู้ป่วยที่ค่อนข้างเยอะ จึงเสียเวลาในการจำแนกผู้ป่วย

## 1.3 แนวทางการแก้ปัญหา

1.3.1 ทำการจัดกลุ่มข้อมูลที่รวบรวมได้ให้ถูกต้อง ลดความซ้ำซ้อนของข้อมูล

1.3.2 หาแบบจำลองในการจำแนกประเภทผู้ป่วย

### 1.3.3 พัฒนาระบบสนับสนุนการประยุกต์ใช้แบบจำลอง

## 1.4 วัตถุประสงค์ของการวิจัย

1.4.1 เพื่อสร้างระบบช่วยในการตัดสินใจในการจำแนกประเภทผู้ป่วยได้เร็วขึ้น

1.4.2 เพื่อศึกษาและหาแบบจำลองการเรียนรู้ของเครื่องในการจำแนกประเภทผู้ป่วยที่มีค่าความแม่นยำที่สุด ระหว่าง 3เทคนิค คือ Random Forest , Neural network และ Naïve Bayes

1.4.3 เพื่อสะท้อนการแพร่ระบาดของโรคโควิด-19 ให้แสดงผลที่อยู่ในรูปแบบที่ดูง่ายการแสดงผลการสร้างมโนภาพ (Data Visualization)

## 1.5 ขอบเขตการศึกษา

ศึกษาและทำการวิเคราะห์ข้อมูลเพื่อจำแนกประเภทผู้ป่วยโควิด-19 จาก 1 มกราคม 2563 ถึงวันที่ 1 ตุลาคม 2564 เป็นจำนวน 1,608,923 ราย จากเว็บไซต์ <https://covid19.ddc.moph.go.th/> หรือ DDC Open data ของกรมควบคุมโรค (Department of disease control, 2021)

## 1.6 ประโยชน์ที่คาดว่าจะได้รับจากงานวิจัย

1.6.1 ได้ระบบตัดสินใจในการจำแนกประเภทผู้ป่วย

1.6.2 ได้แบบจำลองการทำนายสำหรับการจำแนกผู้ป่วยโควิด-19

1.6.3 หน่วยงานที่เกี่ยวข้องกับสาธารณสุขสามารถนำไปใช้ประโยชน์ได้

## 1.7 เครื่องมือและภาษาที่ใช้ในการพัฒนา

### 1.7.1 เครื่องมือ

1. Visual studio code สำหรับการเขียนโค้ดและแก้ไขโค้ดโปรแกรม
2. Navicat สำหรับการออกแบบและคิวรีทดสอบระบบฐานข้อมูล
3. PGAdmin 4 สำหรับการออกแบบและคิวรีทดสอบระบบฐานข้อมูล
4. Orange Data Mining โปรแกรมที่ใช้ทดสอบแบบจำลอง machine learning
5. Pencil สำหรับออกแบบ Mockup Wireframe
6. Adobe XD สำหรับออกแบบ Prototype
7. Microsoft excel
8. บราวเซอร์ เช่น google chrome Opera เป็นต้น

### 1.7.2 ภาษาและเฟรมเวิร์คที่ใช้ในการพัฒนา

1. Python : Fast API Framework
2. PHP : CURL
3. Vue JS : Frond End
4. Node JS
5. SQL
6. Javascript



## บทที่ 2

### แนวคิด ทฤษฎี และงานวิจัยที่เกี่ยวข้อง

#### 2.1 การประยุกต์ใช้การเรียนรู้ของเครื่องสำหรับโรคโควิด-19

ในช่วงเร่งด่วนวาระระดับโลกเมื่อเร็วๆ นี้ มีนักวิทยาศาสตร์ แพทย์ และผู้เชี่ยวชาญด้านการดูแลสุขภาพทั่วโลกยังคงค้นหาเทคโนโลยีใหม่ๆ เพื่อสนับสนุนการแก้ปัญหาการระบาดใหญ่ของโควิด-19 (Lalmuanawma et al., 2020) การเรียนรู้ของเครื่อง (machine learning) ก็เป็นเทคโนโลยีที่เข้ามาช่วยในการแก้ปัญหของโรคโควิด-19 เช่น จีนใช้การวิเคราะห์ข้อมูลขนาดใหญ่ (Big data analytics) เข้ามาควบคุมการแพร่ระบาดให้เกิดประสิทธิผลและเพื่อให้การดำเนินธุรกิจเป็นปกติมากที่สุด (Business-as-usual) ก็เป็นอีกหนึ่งปัจจัยที่ช่วยให้จีนอยู่ในแนวหน้าของการฟื้นตัวทางเศรษฐกิจที่สำคัญแห่งหนึ่งของโลก โดยตัวอย่างการประยุกต์การเรียนรู้ของเครื่องของจีนในช่วงการแพร่ระบาดของไวรัสโควิด-19 ที่น่าสนใจ อาทิ เทคโนโลยี Deep learning สำหรับการประเมินความเป็นไปได้ของการติดเชื้อปอดอักเสบจากการอ่านภาพเอกซเรย์ทรวงอกโดยใช้ระยะเวลาราว 20 วินาที หรือสั้นกว่าระยะเวลาการวินิจฉัยโดยแพทย์ถึง 1 ใน 60 หรือการประยุกต์การเรียนรู้ของเครื่องเข้ากับเทคโนโลยี Internet of things ในภาคการผลิตแบบอัตโนมัติของบริษัทผู้ผลิตคอมพิวเตอร์และอุปกรณ์อิเล็กทรอนิกส์รายใหญ่ของจีนในเมืองอู่ฮั่น โดยไม่ต้องอาศัยพนักงานในช่วงเวลาที่มีมาตรการปิดเมือง (Digest, 2020)

ปัจจุบันมีการพัฒนาอย่างต่อเนื่องในเทคโนโลยีการเรียนรู้ของเครื่องและปัญญาประดิษฐ์ในงานการแพร่ระบาดของโรคโควิด-19 เพื่อวิเคราะห์การใช้ยา การคัดกรอง การคาดการณ์ การติดตามผู้สัมผัส ตลอดจนกระบวนการพัฒนาวัคซีน ของโรคโควิด-19 อย่างมีนัยสำคัญ เพื่อลดการแทรกแซงของมนุษย์ในการปฏิบัติการทางการแพทย์ แต่อย่างไรก็ตามแบบจำลองส่วนใหญ่ไม่ได้ถูกนำไปใช้งาน

มากพอที่จะสามารถแสดงการปฏิบัติงานในโลกแห่งความเป็นจริง แต่ก็ยังสามารถใช้รับมือการแพร่ระบาดของโรคโควิด-19 ได้ (Lalmuanawma et al., 2020)

## 2.2 Open Data - Open Government Data Center

หน่วยงานภาครัฐเปิดเผยข้อมูลข่าวสารอย่างต่อเนื่อง นับตั้งแต่พระราชบัญญัติข้อมูลข่าวสารของราชการ พ.ศ. 2540 มีผลบังคับใช้ เพื่อให้ประชาชนได้รับทราบข้อมูลการดำเนินงานของรัฐ ซึ่งเป็นการสร้างความโปร่งใส การมีส่วนร่วมของประชาชน รวมถึงเป็นประโยชน์ต่อการพัฒนาประเทศด้านการเมือง สังคม และเศรษฐกิจ (สำนักงานพัฒนารัฐบาลดิจิทัล (องค์การมหาชน) (สพร.), 2015)

ปัจจุบัน ประชาชนสามารถเข้าถึงข้อมูลข่าวสารราชการได้ด้วยการเข้าใช้เว็บไซต์ของหน่วยงานภาครัฐ โดยมักพบไฟล์ข้อมูลในรูปแบบ PDF, XLS และ DOC ที่สามารถดาวน์โหลดหรือเปิดอ่านได้ด้วยโปรแกรมที่ประชาชนใช้อยู่แล้วโดยทั่วไป

สำหรับยุคเศรษฐกิจดิจิทัล การผลักดันให้เกิดการเปิดเผยข้อมูลของหน่วยงานภาครัฐในแนวทางเดียวกันมีความสำคัญอย่างยิ่ง เพื่อให้เกิดการจัดทำข้อมูลเปิดภาครัฐ (Open Government Data) รวมถึงชุดข้อมูล (Data set) ที่จะสามารถนำไปใช้สร้างสรรค์นวัตกรรมทางเศรษฐกิจ สังคม และชีวิตความเป็นอยู่ เพื่อนำไปสู่การเป็นประเทศที่มีความเข้มแข็ง และความก้าวหน้าด้วยความสามารถของการเข้าถึงแหล่งข้อมูลความรู้

สำนักงานรัฐบาลอิเล็กทรอนิกส์ (องค์การมหาชน) (สรอ.) หรือ อีจีเอ จึงบริหารจัดการโครงการศูนย์กลางข้อมูลภาครัฐ (Open Government Data Center) ภายใต้ชื่อ Data.go.th โดยมีวัตถุประสงค์เพื่อมุ่งเน้นการพัฒนาระบบศูนย์กลางสำหรับให้บริการข้อมูลเปิดภาครัฐ และเพื่อการเผยแพร่ความรู้ รวมถึงสนับสนุนหน่วยงานภาครัฐในการจัดทำ Open Data

## 2.2.1 แนวทางเบื้องต้นในการจัดทำ Open Data

2.2.1.1 ทำให้ง่ายเข้าไว้ เริ่มต้นด้วยสิ่งที่เล็ก ง่าย และเร็ว ไม่จำเป็นว่าทุกๆ ชุดข้อมูลจะต้องถูกสร้างให้ open ในตอนนี้ เริ่มต้นด้วยการเปิดเพียงแค่ 1 ชุดข้อมูล หรือแม้แต่ 1 ส่วนของชุดข้อมูลขนาดใหญ่ก็เพียงพอแล้วสำหรับการเริ่มต้น

2.2.1.2 มีส่วนร่วมตั้งแต่แรกและบ่อยๆ การมีส่วนร่วมกับผู้ใช้งานจริงที่นำข้อมูลไปใช้ตั้งแต่ในระยะแรกและบ่อยเท่าที่ทำได้ ไม่ว่าจะเป็นภาคประชาชน ภาคธุรกิจ หรือกลุ่มนักพัฒนา จะทำให้การพัฒนาบริการของหน่วยงานนั้น มีความสอดคล้องกับความต้องการของประชาชนได้ง่ายขึ้น เนื่องจากข้อมูลจำนวนมากจะไปไม่ถึงผู้ใช้งานโดยตรง แต่จะต้องผ่าน ‘info-mediaries’ ที่มีการดึงข้อมูลและแปลงหรือเรียบเรียงขึ้นใหม่เพื่อนำมาเสนอ ตัวอย่างเช่น ส่วนใหญ่เราจะไม่ต้องการฐานข้อมูลพิกัด GPS ขนาดใหญ่ แต่เราต้องการแผนที่มากกว่า ดังนั้นจึงต้องทำในส่วนตัวกลางข้อมูลก่อน

2.2.1.3 ขจัดความกลัวและความเข้าใจผิดโดยทั่วไป ซึ่งมีความสำคัญโดยเฉพาะอย่างยิ่งข้อมูลของภาครัฐ เมื่อมีการเปิดเผยข้อมูล อาจจะต้องเจอกับคำถามและความกลัวมากมาย ดังนั้นสิ่งสำคัญคือ (1) ระบุสิ่งที่สำคัญมากที่สุด และ (2) จัดการสิ่งเหล่านั้นในระยะแรกที่สุดเท่าที่เป็นไปได้

## 2.2.3 ระดับการเปิดเผยข้อมูล 5 ระดับ

Data.go.th กำหนดระดับการเปิดเผยข้อมูล เพื่อให้สามารถพิจารณาถึงการนำไปใช้ได้สะดวกมากยิ่งขึ้น โดยจัดระดับการเปิดเผยข้อมูล 5 ระดับ ดังนี้

2.2.3.1 1 ดาว เผยแพร่ข้อมูลในทุกรูปแบบบนเว็บไซต์ ประเภทข้อมูล PDF DOC TXT TIFF JPEG

2.2.3.2 2 ดาว เผยแพร่ข้อมูลในรูปแบบ Structured data ที่เครื่องสามารถอ่านได้ (Machine-readable) เช่น ข้อมูลอยู่ในรูปแบบไฟล์ Excel ประเภทข้อมูล XLS

2.2.3.3 3 ดาว เผยแพร่ข้อมูลในรูปแบบ Non-proprietary format เช่น ข้อมูลในรูปแบบ CSV แทนรูปแบบ Excel ประเภทข้อมูล CSV ODS XML JSON KML SHP KMZ

2.2.3.4 4 ดาว ใช้ URI (Uniform Resource Identifier) ในการระบุตัวตนของทรัพยากร (ข้อมูล) และชี้ไปยังตำแหน่งของทรัพยากรนั้น ประเภทข้อมูล RDF (URIs)

2.2.3.5 5 ดาว ข้อมูลมีการเชื่อมโยงไปสู่แหล่งข้อมูลอื่นๆ ในบริบทที่เกี่ยวข้องกันได้ ประเภทข้อมูล RDF (Linked Data)

### 2.2.3 สถานการณ์ผู้ติดเชื้อ COVID-19 อัปเดตรายวัน รูปแบบ API (Json/CSV Data Format)

สถานการณ์ผู้ติดเชื้อ COVID-19 อัปเดตรายวัน รูปแบบ API (Json/CSV Data Format) เป็นข้อมูลแบบเปิดที่รวบรวมโดยกรมควบคุมโรคจากกระทรวงสาธารณสุข (Department of disease control, 2021) ซึ่งรายละเอียดจะปรากฏดังตารางที่ 2.1

ตารางที่ 2.1 ชุดข้อมูลบริการรายงาน โควิด-19 (Department of disease control, 2021)

คำอธิบายบริการ	Content-Type	API URL
รายงานสถานการณ์ COVID-19 ประจำวัน	application/json	<a href="https://covid19.ddc.moph.go.th/api/Cases/today-cases-all">https://covid19.ddc.moph.go.th/api/Cases/today-cases-all</a>
รายงานสถานการณ์ COVID-19 ประจำวัน แยกตามรายจังหวัด	application/json	<a href="https://covid19.ddc.moph.go.th/api/Cases/today-cases-by-provinces">https://covid19.ddc.moph.go.th/api/Cases/today-cases-by-provinces</a>
ข้อมูลผู้ป่วยประจำวัน (Line Lists)	text/csv	<a href="https://covid19.ddc.moph.go.th/api/Cases/today-cases-line-lists">https://covid19.ddc.moph.go.th/api/Cases/today-cases-line-lists</a>
รายงานสถานการณ์ COVID-19 ระลอก 3 (ตั้งแต่ 01/04/2021 – ปัจจุบัน)	application/json	<a href="https://covid19.ddc.moph.go.th/api/Cases/timeline-cases-all">https://covid19.ddc.moph.go.th/api/Cases/timeline-cases-all</a>
รายงานสถานการณ์ COVID-19 ระลอก 3 (ตั้งแต่ 01/04/2021 – ปัจจุบัน) แยกตามรายจังหวัด	application/json	<a href="https://covid19.ddc.moph.go.th/api/Cases/timeline-cases-by-provinces">https://covid19.ddc.moph.go.th/api/Cases/timeline-cases-by-provinces</a>
ข้อมูลผู้ป่วยระลอก 3 (ตั้งแต่ 01/04/2021 – ปัจจุบัน)	application/json with pagination	<a href="https://covid19.ddc.moph.go.th/api/Cases/round-3-line-lists">https://covid19.ddc.moph.go.th/api/Cases/round-3-line-lists</a>

ตารางที่ 2.1 ชุดข้อมูลบริการรายงาน โควิด-19 (Department of disease control, 2021) (ต่อ)

คำอธิบายบริการ	Content-Type	API URL
ข้อมูลผู้ป่วยระลอก 1 ถึงระลอก 2 (ตั้งแต่ 12/01/2020 – 31/03/2021)	text/csv	<a href="https://covid19.ddc.moph.go.th/api/Cases/round-1to2-line-lists">https://covid19.ddc.moph.go.th/api/Cases/round-1to2-line-lists</a>
รายงานสถานการณ์ COVID-19 ระลอก 1 ถึงระลอก 2 (ตั้งแต่ 12/01/2020 – 31/03/2021)	application/json	<a href="https://covid19.ddc.moph.go.th/api/Cases/round-1to2-all">https://covid19.ddc.moph.go.th/api/Cases/round-1to2-all</a>
รายงานสถานการณ์ COVID-19 ระลอก 1 ถึงระลอก 2 (ตั้งแต่ 12/01/2020 – 31/03/2021) แยกตามรายจังหวัด	application/json	<a href="https://covid19.ddc.moph.go.th/api/Cases/round-1to2-by-provinces">https://covid19.ddc.moph.go.th/api/Cases/round-1to2-by-provinces</a>

จากตารางที่ 2.1 ผู้วิจัยได้ใช้ข้อมูลผู้ป่วยระลอก 1 ถึงระลอก 2 (ตั้งแต่ 12/01/2020 – 31/03/2021) และ ข้อมูลผู้ป่วยระลอก 3 (ตั้งแต่ 01/04/2021 – ปัจจุบัน) นั่นคือตั้งแต่วันที่ 1 มกราคม 2563 ถึง วันที่ 1 ตุลาคม 2564 เป็นจำนวน 1,608,923 ราย ในการทำวิจัยฉบับนี้ โดยหน้าเว็บไซต์จะแสดงดังภาพที่ 2.1

สถานการณ์ผู้ติดเชื้อ COVID-19 อัปเดตรายวัน รูปแบบ API(Json/CSV)		
คำอธิบายบริการ	Content-Type	API U
รายงานสถานการณ์ COVID-19 ประจำวัน	application/json	https
รายงานสถานการณ์ COVID-19 ประจำวัน แยกตามรายจังหวัด	application/json	https
ข้อมูลผู้ป่วยประจำวัน(Line Lists)	text/csv	https
รายงานสถานการณ์ COVID-19 ระลอก 3 (ตั้งแต่ 01/04/2021 – ปัจจุบัน)	application/json	https
รายงานสถานการณ์ COVID-19 ระลอก 3 (ตั้งแต่ 01/04/2021 – ปัจจุบัน) แยกตามรายจังหวัด	application/json	https
ข้อมูลผู้ป่วยระลอก 3 (ตั้งแต่ 01/04/2021 – ปัจจุบัน)	application/json with pagination	https
ข้อมูลผู้ป่วยระลอก 1 ถึงระลอก 2 (ตั้งแต่ 12/01/2020 – 31/03/2021)	text/csv	https
รายงานสถานการณ์ COVID-19 ระลอก 1 ถึงระลอก 2 (ตั้งแต่ 12/01/2020 – 31/03/2021)	application/json	https
รายงานสถานการณ์ COVID-19 ระลอก 1 ถึงระลอก 2 (ตั้งแต่ 12/01/2020 – 31/03/2021) แยกตามรายจังหวัด	application/json	https
คำอธิบายบริการ	Content-Type	API U

ภาพที่ 2.1 เว็บไซต์ข้อมูลบริการรายงาน โควิด-19 <https://covid19.ddc.moph.go.th/>

โดยสามารถดึงข้อมูลผ่าน API ในรูปแบบ JSON ประเภท Method GET ได้โดยมีเส้นทางการร้องขอข้อมูลคือ <https://covid19.ddc.moph.go.th/api/Cases/round-3-line-lists?page=1> จะแสดงผลในการร้องขอข้อมูลดังภาพที่ 2.2

```
{
  "data": [
    {
      "txn_date": "2020-01-12",
      "gender": "หญิง",
      "age_number": "61",
      "age_range": "60-69 ปี",
      "nationality": "Chinese",
      "job": "ซักรถยนต์",
      "risk": "คนต่างชาติเดินทางมาจากต่างประเทศ",
      "patient_type": "1.คัดกรองจากด่าน",
      "province": "กรุงเทพมหานคร",
      "update_date": "2022-06-30 07:26:58"
    }
  ],
  "links": {
    "first": "http://covid19.ddc.moph.go.th/api/Cases/round-3-line-lists?page=1",
    "last": "http://covid19.ddc.moph.go.th/api/Cases/round-3-line-lists?page=907",
    "prev": null,
    "next": "http://covid19.ddc.moph.go.th/api/Cases/round-3-line-lists?page=2"
  },
  "meta": {
```

ภาพที่ 2.2 โปรแกรม Postman ทดสอบ API ในการดึงข้อมูลผู้ป่วยโควิด-19

จากภาพที่ 2.2 สามารถบอกรายละเอียดเป็นพจนานุกรมข้อมูล (Data dictionary) ได้ดังตารางที่ 2.2 ตารางที่ 2.2 พจนานุกรมข้อมูลข้อมูลผู้ป่วยโควิด-19

Attribute	Data Type	Comment
txn_date	date	วันแกลง
gender	varchar	เพศ
age_number	varchar	อายุ
age_range	varchar	ช่วงอายุ
nationality	varchar	สัญชาติ
job	varchar	อาชีพ

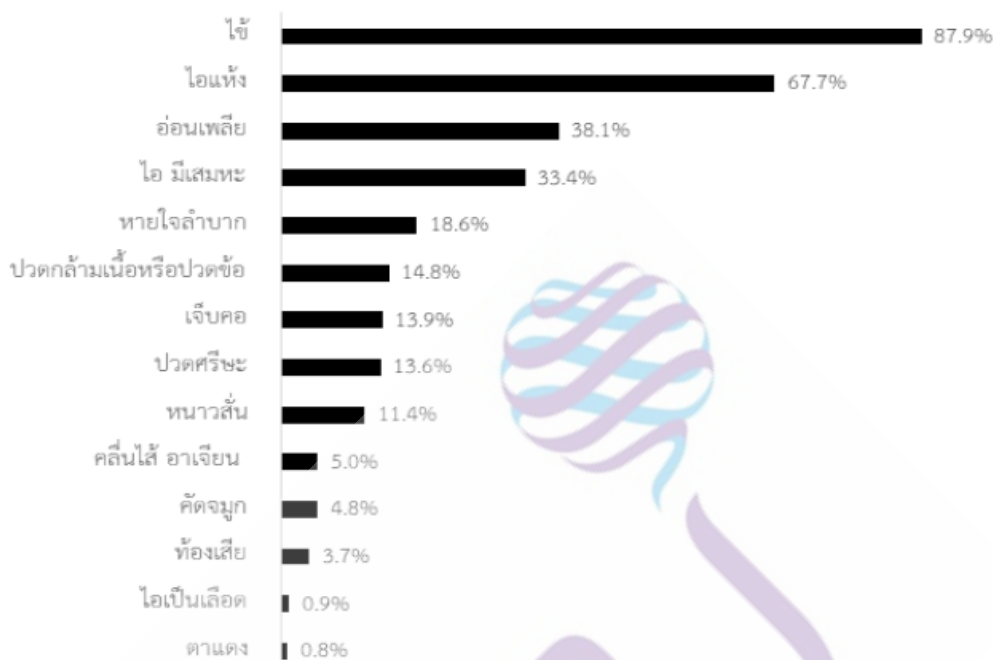
## ตารางที่ 2.2 พจนานุกรมข้อมูลข้อมูลผู้ป่วยโควิด-19 (ต่อ)

Attribute	Data Type	Comment
risk	varchar	ปัจจัยเสี่ยง
patient_type	text	ประเภทผู้ป่วย
province	varchar	จังหวัด
update_date	datetime	วันที่ปรับปรุงข้อมูลล่าสุด

### 2.3 ทฤษฎีและแนวคิดพื้นฐานเกี่ยวกับโรคโควิด-19

โรคโควิด 19 (Coronavirus Disease 2019) เป็นโรคอุบัติใหม่ (หาญสมบุรณ์, 2564) ซึ่งเป็นโรคระบาดร้ายแรง มีการติดต่อโดยละอองเสมหะ จากการไอ และมีการระบาดอย่างรวดเร็ว องค์การอนามัยโรคประกาศเป็นโรคที่ระบาดเป็นวงกว้าง (pandemic disease) ในวันที่ 11 มีนาคม 2563 ในขณะที่ประเทศไทยได้มีประกาศกระทรวงสาธารณสุขลงวันที่ 26 กุมภาพันธ์ 2563 ให้โรคโควิด 19 เป็นโรคติดต่อร้ายแรง พบมีการระบาดทั่วโลก ผู้ติดเชื้อทั่วโลกมากกว่า 170 ล้านคน และมีผู้เสียชีวิตกว่า 3.5 ล้านคน (ข้อมูล ณ วันที่ 1 มิถุนายน 2564 โรคโควิด 19 เกิดจากเชื้อไวรัสโคโรนาสายพันธุ์ใหม่ (Coronavirus) พบครั้งแรกที่นครอู่ฮั่น เมืองหลวงของมณฑล หูเป่ย์ ประเทศจีน ชื่อเต็มของเชื้อไวรัสตัวนี้คือ Severe Acute Respiratory Syndrome Coronavirus 2 (SARS CoV 2) ในวันที่ 11 กุมภาพันธ์ 2563 องค์การอนามัยโลก (WHO) ให้ชื่อโรคใหม่ว่า COVID-19 (Coronavirus Disease 2019) อาการแสดงของโรคนี้ มีตั้งแต่ ไม่มีอาการ อาการเหมือนไข้หวัด ไอ เจ็บคอ ลิ้นไม่รับรส จมูกไม่ได้กลิ่น จนถึงอาการรุนแรงติดเชื้อในปอด ปอดอักเสบ ระบบ ทางเดินหายใจล้มเหลว ไตวายและอาจถึงขั้นเสียชีวิตได้ (เฉลยกิตติ et al., 2563)





ภาพที่ 2.3 แสดงอาการผู้ป่วยโควิด-19 (Manmana et al., 2020)

ลักษณะของเชื้อไวรัสโคโรนาไวรัส ไวรัสหมายถึงสิ่งมีชีวิตที่เป็นจุลินทรีย์ที่มีขนาดเล็ก ไม่จัดเป็นเซลล์ ไม่สามารถมองเห็นด้วยกล้องจุลทรรศน์ธรรมดา มีทั้งชนิดก่อโรคและไม่ก่อโรค จัดเป็นไวรัสชนิดหนึ่งที่แพร่จากคนสู่คนเป็นสาเหตุของการเจ็บป่วยตั้งแต่ไข้หวัดธรรมดาจนถึงภาวะที่ทำให้เกิดการเจ็บป่วยรุนแรง (Rachada, 2020) เช่น โรคระบบทางเดินหายใจเฉียบพลันร้ายแรงหรือโรคซาร์ส (SARS) ในปี 2002-2003 ที่ติดเชื้อมาจากค้างคาวสู่คน โรคไข้หวัด H1N1 ในปี 2009 และโรคทางเดินหายใจตะวันออกกลาง (MERS) ในปี 2012 ที่ติดเชื้อมาจากอูฐสู่คน

เมื่อพิจารณาจากการระบาดของโรคซาร์สและโรคทางเดินหายใจตะวันออกกลาง พบว่า เชื้อไวรัสทั้งสองมีต้นกำเนิดมาจากค้างคาวเช่นเดียวกับโรคโควิด-19 โดยโรคซาร์สและโรค MERS มีสัตว์เลี้ยงลูกด้วยนมเป็นตัวกลางรับไวรัสจากค้างคาว และถ่ายทอดสู่คนอีกต่อหนึ่ง (Intermediate host) (Manmana et al., 2020; Rachada, 2020) ไวรัสโรคซาร์สและโรค MERS มีตัวชะมด (civet) และอูฐหนอกเดียว (dromedary camels) เป็นตัวกลางในการถ่ายทอดไวรัสสู่คนตามลำดับ

องค์การอนามัยโลกหรือ WHO ประกาศว่าการระบาดของโรคโควิด-19 ทำให้เกิดการแพร่กระจายไปทั่วโลก (Pandemic) จัดระดับการแพร่กระจายของโรคติดเชื้อเป็นการระบาดใหญ่ คือ การที่โรคติดเชื้อแพร่ระบาดจากคนสู่คนได้ง่าย โดยเกิดขึ้นพร้อมกันในหลายพื้นที่ทั่วโลก การแพร่ระบาดที่เรียกว่า ระยะ 1,2,3 ไม่ได้เกี่ยวข้องกับระดับความรุนแรงของโรค แต่เป็นการกำหนดระยะเวลาของการระบาดเพื่อวางแผนเสนอนโยบายและวางมาตรการในการควบคุมและป้องกันโรค (ตุ้ยเจริญ et al., 2563) โดยจะเป็นการกำหนดนโยบายของแต่ละประเทศ จึงไม่ได้เป็นบรรทัดฐานเดียวกันทั่วโลก การแพร่ระบาดงานสาธารณสุขแบ่งระยะของการระบาดออกเป็น 3 ระยะ ดังนี้ (กิจผาดิ, 2563)

ระยะที่ 1 พบผู้ติดเชื้อซึ่งเป็นผู้ป่วยจากต่างประเทศและไม่พบการติดเชื้อที่เกิดขึ้นในประเทศ

ระยะที่ 2 พบคนไทยติดเชื้อในประเทศ จากนักท่องเที่ยวหรือคนไทยที่คาดว่าติดเชื้อจากการเดินทางไปยังกลุ่มประเทศเสี่ยง

ระยะที่ 3 มีการติดเชื้อระหว่างคนในประเทศด้วยกันเพิ่มขึ้นอย่างต่อเนื่องระหว่างผู้ที่ไม่เคยมีประวัติเดินทางไปยังต่างประเทศ และทำให้เกิดการแพร่กระจายต่อเป็นวงกว้าง

## 2.4 การเรียนรู้ของเครื่อง (Machine Learning)

การเรียนรู้ของเครื่อง (Machine Learning) เป็นหนึ่งในประเภทของปัญญาประดิษฐ์ ที่จะช่วยให้แอปพลิเคชัน หรือซอฟต์แวร์เกิดการเรียนรู้จากชุดข้อมูลที่จะสามารถทำนายผลได้อย่างแม่นยำ โดยหลักการคือจะมีแบบจำลองทางคณิตศาสตร์และมีอัลกอริทึมในหลากหลายรูปแบบมาเรียนรู้จากชุดข้อมูล (Data set) จนทำให้คอมพิวเตอร์เกิดการเรียนรู้และฉลาดในการตัดสินใจบนข้อมูลที่ไม่เคยได้เห็นมาก่อนได้สำเร็จ (Alpaydin, 2020; Witten et al., 2011; โพธิ์สอาด & โพธิ์สว่าง, 2562)

### 2.4.1 ประเภทการเรียนรู้ของเครื่อง

2.4.1.1 เทคนิคการเรียนรู้แบบมีผู้สอน (Supervised Learning) เทคนิคในประเภทนี้จะเน้นการเรียนรู้จากข้อมูลที่มีอยู่ในอดีตเพื่อนำมาสร้างแบบจำลองสำหรับทำนายหรือคาดการณ์สิ่งที่จะเกิดขึ้นในอนาคต ในแบบจำลองแต่ละแบบจำลองอาจจะเป็นสมการทางคณิตศาสตร์ หรือ กฎต่างๆ โดยเทคนิค

การเรียนรู้แบบมีผู้สอนนี้สามารถแบ่งย่อยได้อีกคือ การจำแนกประเภทข้อมูล (Classification) และการประมาณค่าข้อมูล (Regression) ซึ่งทั้งสองเทคนิคนี้จะมีลักษณะคล้ายกันโดยจะแตกต่างจากคำตอบที่ได้จากการทำนายหรือค่าความจริงที่ปรากฏ (Pacharawongsakda, 2014; มาณวิภา, 2562)

1. **การจำแนกประเภทข้อมูล (Classification)** คือ เทคนิคการจำแนกประเภทข้อมูลที่มีในอดีตมาสอนระบบเพื่อให้เรียนรู้รูปแบบที่เกิดขึ้นในข้อมูลแล้วจึงสร้างเป็นสมการหรือแบบจำลอง (model) ขึ้นมาเพื่อหาคำตอบให้สำหรับข้อมูลใหม่ เช่น การจำแนกออกเป็นอีเมลล์ออกเป็นสแปมหรือแบบปกติ โดยจะต้องมีการนำข้อมูลของอีเมลล์ประเภทสแปมและประเภทปกติมาให้คอมพิวเตอร์ทำการเรียนรู้ และหลังจากนั้นจึงสร้างแบบจำลองโดยการจำแนกประเภทของอีเมลล์ว่าเป็นสแปมหรือแบบปกติ (Pacharawongsakda, 2014)

2. **การประมาณค่าข้อมูลหรือการวิเคราะห์การถดถอย (Regression)** คือ วิธีการวิเคราะห์หาความสัมพันธ์ระหว่างสองตัวแปรขึ้นไป ซึ่งสามารถนำค่าของความสัมพันธ์มาพยากรณ์ค่าของตัวแปรอีกตัวได้ (Fayong, 2019)

โดยจะยกตัวอย่างเทคนิคการเรียนรู้แบบมีผู้สอนคือคอมพิวเตอร์จะเรียนรู้จากชุดข้อมูลของสัตว์ โดยจะใช้ชุดข้อมูลเป็นตัวอย่างสอนคอมพิวเตอร์ให้เข้าใจว่าเป็นแมว โดยการจำแนกแมวเป็นประเภทข้อมูล เช่น แมวจะต้องมี 4 ขา หรือ มี 1 หาง เป็นต้น ด้วยขั้นตอนวิธีนี้ทำให้คอมพิวเตอร์นั้นสามารถแยกแมวออกจากสิ่งอื่นได้ดังภาพที่ 2.4



ภาพที่ 2.4 ตัวอย่างการเรียนรู้แบบมีผู้สอน (Puitrakul, 2018)

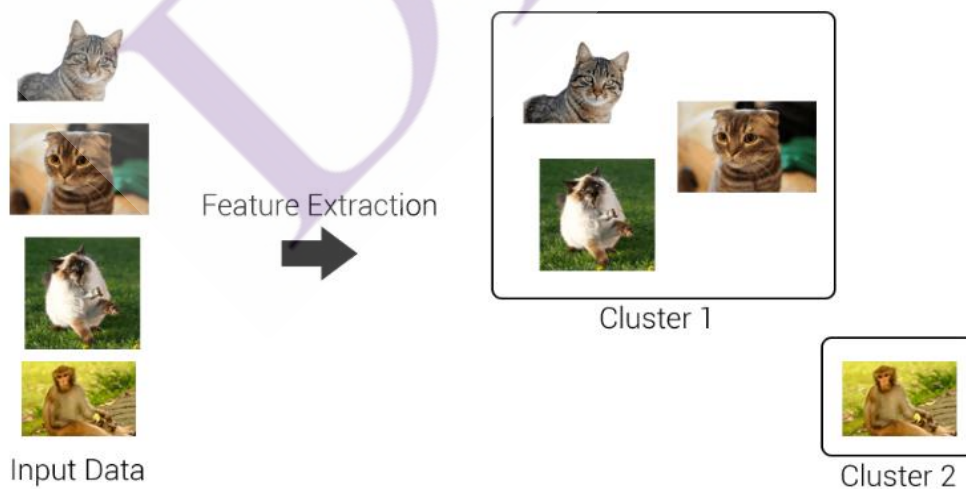
2.4.1.2 เทคนิคการเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) เทคนิคนี้จะเน้นการพิจารณาข้อมูลเป็นหลัก เช่น พิจารณาความสัมพันธ์ของข้อมูลกันในลักษณะแบบใดบ้าง โดยเทคนิคประเภทนี้จะแบ่งย่อยได้อีกคือ เทคนิคการค้นหากฎความสัมพันธ์ (Association rule) (มานวิภา, 2562)

1. เทคนิคการค้นหากฎความสัมพันธ์ Association rule คือ การค้นหากฎความสัมพันธ์จากข้อมูลที่มีขนาดใหญ่จากชุดข้อมูล (Data set) เพื่อนำมาวิเคราะห์หรือทำนายปรากฏการณ์ต่างๆ (สุขสกาฟอง & มีสัง, 2559)

2. การแบ่งกลุ่มข้อมูล (Clustering) คือ เป็นเทคนิคที่ใช้ในการจัดกลุ่มหรือแบ่งกลุ่มข้อมูล เช่น คนสัตว์ สิ่งของ หรือ ต้นไม้ องค์กร บริษัท เป็นต้น โดยกลุ่มที่มีลักษณะเหมือนกันหรือคล้ายกันจะอยู่ในกลุ่มเดียวกัน ดังนั้นการพิจารณาเลือกลักษณะตัวแปรที่จะใช้ในการแบ่งกลุ่มจึงมีความสัมพันธ์ (Srimarong, 2020)

โดยเทคนิควิธีการเรียนรู้แบบไม่มีผู้สอนจะค่อนข้างยากในการที่จะให้คอมพิวเตอร์จัดกลุ่มคำตอบให้ชัดเจน กล่าวคือ คอมพิวเตอร์จะสามารถเรียนรู้จากชุดข้อมูลของสัตว์แต่ไม่รู้ว่าสัตว์ประเภทนี้คือสัตว์อะไร แต่คอมพิวเตอร์สามารถทำได้คือการทำ Feature Extraction และแยกข้อมูลออกมาได้ โดยที่ไม่รู้ว่าประเภทข้อมูลนั้นคืออะไรหรือสัตว์นี้คืออะไร โดยกลุ่มพวกนี้จะเรียกว่าการ Cluster ดังภาพที่

2.5



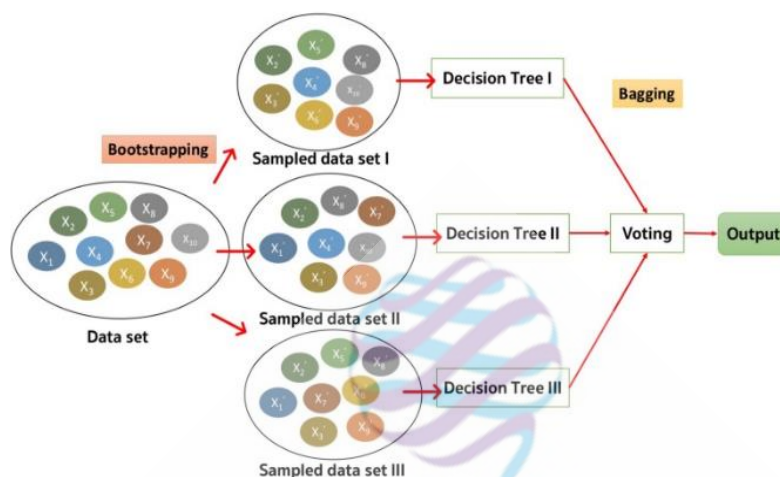
ภาพที่ 2.5 ตัวอย่างการเรียนรู้แบบไม่มีผู้สอน (Puitrakul, 2018)

## 2.5 Random Forest

เป็นการอัลกอริทึมสร้างแบบจำลองที่เป็นหนึ่งในประเภทของอัลกอริทึมต้นไม้ตัดสินใจ (Decision Tree) และเป็นเทคนิคหนึ่งของการเรียนรู้ของเครื่อง (Machine Learning) ที่มีลักษณะแบบไม่ตัดแต่งกิ่ง (Unpruned) หรือ ต้นไม้ถดถอย (Regression Trees) โดยจะถูกสร้างจากการนำข้อมูลฝึกสอนไปสุ่มเลือกตัวอย่างข้อมูลและคุณลักษณะของข้อมูล หลังจากนั้นนำมาสร้างเป็นต้นไม้ตัดสินใจ ซึ่งจะมีตัวอย่างเพียงบางส่วนที่จะไม่ถูกเลือกจะถูกนำมาใช้ในการทดสอบต้นไม้ตัดสินใจ เรียกข้อมูลส่วนนี้ว่า Out-of-Bag (OOB) วิธีการนี้จะเรียกว่าแบ็กกิ้ง Bootstrap Aggregating (Bagging) (ฉันทวานนท์ et al., 2560)

หลักการของ Random Forest คือการสร้าง model จาก Decision Tree หลายๆ model ย่อยๆ (ตั้งแต่ 10 model ถึง มากกว่า 1000 model) โดยแต่ละ model จะได้รับ data set ไม่เหมือนกัน ซึ่งเป็น subset ของ data set ทั้งหมด ตอนทำ prediction จะให้ Decision Tree ทำการ prediction ในแต่ละโหนด และคำนวณผล prediction ด้วยการ vote output ที่ ถูกเลือกโดย Decision Tree มากที่สุด (กรณี classification) หรือ หาค่า mean จาก output ของแต่ละ Decision Tree (กรณี regression) (Patcharacharoenwong et al., 2020)

Decision Tree แต่ละ model ใน Random Forest ถือว่าเป็น weak learner ที่เป็น model ที่ยังไม่มีประสิทธิภาพในการ prediction มากพอ แต่ในการทำงานจะนำ Decision Tree มาทำ prediction ร่วมกัน ก็จะได้ model รวมที่มีความเก่ง และแม่นยำมากกว่า Decision Tree ที่ทำ prediction แบบเดียวๆ



ภาพที่ 2.6 หลักการทำ Random Forest (Daroontham, 2018)

จากภาพที่ 2.6 หลักการทำ Random Forest (Daroontham, 2018)

1. sample ข้อมูล (bootstrapping) จาก data set ทั้งหมด ให้ได้ข้อมูลออกมา  $n$  ชุด ที่ไม่เหมือนกัน ตามจำนวน Decision Tree ใน Random Forest เช่น data set ตั้งต้นมีอยู่ 10 feature ( $X_1, X_2, \dots, X_{10}$ ) แต่ละ Decision Tree จะได้ feature ไปไม่เหมือนกัน และ จะได้ข้อมูลไม่ครบทุก row ด้วยจาก data set ทั้งหมดด้วย ( $X_1 \rightarrow X_1', X_2 \rightarrow X_2', \dots$ )
2. สร้าง model Decision Tree สำหรับแต่ละชุดข้อมูล
3. ทำ aggregation ผลลัพธ์ จากแต่ละ model (bagging) เช่น voting ในกรณี classification หรือ หาค่า mean ในกรณี regression

## 2.6 Naive Bayes

ใช้การคำนวณความน่าจะเป็นซึ่งถูกใช้ในการทำนายผลจัดเป็นเทคนิคในการแก้ปัญหาแบบ Classification ที่สามารถคาดการณ์ผลลัพธ์ได้และสามารถอธิบายได้ โดยจะทำการวิเคราะห์



ความสัมพันธ์ระหว่างตัวแปร เพื่อใช้ในการสร้างเงื่อนไขความน่าจะเป็นสำหรับแต่ละความสัมพันธ์ (Pacharawongsakda, 2014) การเรียนรู้แบบเป็นวิธีจำแนกประเภทข้อมูลที่มีประสิทธิภาพวิธีหนึ่ง ที่การทำงานไม่ซับซ้อนเหมาะกับกรณีของเซตตัวอย่างจำนวนมากและคุณลักษณะของตัวอย่างจะไม่ขึ้นต่อกัน

การจำแนกประเภทข้อมูลด้วยวิธี Naive Bayes เป็นอีกหนึ่งวิธีที่ได้รับความนิยมเนื่องจากการสร้างแบบจำลองง่ายและไม่ซับซ้อน โดยจะอาศัยทฤษฎีความน่าจะเป็น (probability) เป็นหลัก (ปึกเข้ม & จันมณีชัย, 2560)

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad (1)$$

โดยที่

1.  $P(A|B)$  คือ ค่า conditional probability หรือค่าความน่าจะเป็นที่เกิดเหตุการณ์ B ขึ้นก่อนและจะมีเหตุการณ์ A ตามมา
2.  $P(A \cap B)$  คือ ค่า joint probability หรือค่าความน่าจะเป็นที่เหตุการณ์ A และเหตุการณ์ B เกิดขึ้นร่วมกัน
3.  $P(B)$  คือ ค่าความน่าจะเป็นที่เหตุการณ์ B เกิดขึ้น

ในสมการที่เรียกว่า Bayes theorem หรือทฤษฎีของเบย์ในการนำไปใช้งานเพื่อจำแนกประเภทข้อมูล โดยที่ A คือแอตทริบิวต์ (attribute) และ C คือคลาส (class) ดังสมการที่ 2

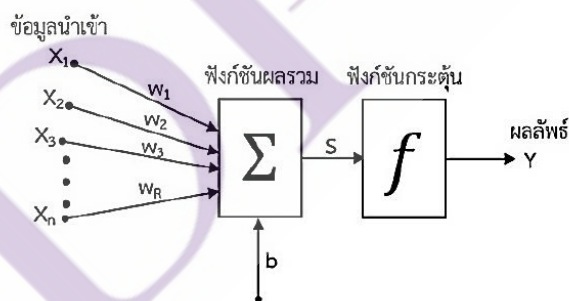
$$P(C|A) = \frac{P(A|C) \times P(C)}{P(A)} \quad (2)$$

1. **Posterior probability** หรือ  $P(C|A)$  คือ ค่าความน่าจะเป็นที่ข้อมูลที่มีแอตทริบิวต์เป็น A จะมีคลาส C
2. **Likelihood** หรือ  $P(A|C)$  คือ ค่าความน่าจะเป็นที่ข้อมูล training data ที่มีคลาส C และมีแอตทริบิวต์ A โดยที่  $A = a_1 \cap a_2 \dots \cap a_M$  โดยที่ M คือจำนวนแอตทริบิวต์ใน training data
3. **Prior probability** หรือ  $P(C)$  คือ ค่าความน่าจะเป็นของคลาส C



## 2.7 Neural network

Neural network หรือโครงข่ายประสาทเทียม โดยลักษณะทั่วไปของโครงข่ายประสาทเทียมคือ โหนดจำลองมาจากรอยประสานประสาท (Synapse) ระหว่างใยประสาทนำเข้า (Dendrite) ของเซลล์ประสาทตัวหนึ่งและแกนประสาทนำออก (Axon) ของเซลล์ประสาทอีกตัวหนึ่ง โดยมีฟังก์ชันการแปลงเป็นตัวกำหนดสัญญาณส่งออก องค์ประกอบในการทำงานของระบบประสาทในสมองมนุษย์กับโครงข่ายประสาทเทียมมีความคล้ายคลึงกัน ในระบบประสาทของมนุษย์การรับกระแสประสาทจะอาศัยเดนไดรต์ (Dendrites) เป็นตัวนำพา เปรียบเทียบได้กับหน่วยข้อมูลป้อนเข้าในระบบโครงข่ายประสาทเทียม ส่วนการนำกระแสประสาทไปยังเซลล์ประสาทอื่นผ่านทางเอกซอน (AXon) ที่เป็นตัวส่งออกกระแสประสาทเปรียบเทียบกับหน่วยผลลัพธ์ในโครงข่ายประสาทเทียม โครงข่ายประสาทเทียมจะผ่านการประมวลผล 2 ขั้นตอน คือ ขั้นตอนการหาผลรวมและขั้นตอนการแปลง (Thongjit et al., 2019) ดังภาพที่ 2.7 โดยโครงข่ายประสาทเทียมจะมีการป้อนข้อมูลเข้าและการกำหนดค่าน้ำหนักซึ่งแบ่งชั้น (Layer) การทำงานออกเป็น 3 ชั้น ดังนี้



ภาพที่ 2.7 แบบจำลองการทำงานของ Neural network (Thongjit et al., 2019)

### คำอธิบายภาพที่ 2.7

1. ชั้นข้อมูลนำเข้า (Input layer) ทำหน้าที่นำเข้าข้อมูลเข้าสู่โครงข่ายประสาทเทียม โดยข้อมูลนี้จะนำไปประมวลผลในแต่ละโหนด (Node) ของชั้นถัดไป

2. ชั้นซ่อน (Hidden layer) ทำหน้าที่รับข้อมูลจากชั้นนำเข้าโดยการกำหนดค่าน้ำหนักของข้อมูลจากชั้นนำเข้าก่อนรับข้อมูลเข้าสู่ชั้นซ่อน ชั้นซ่อนจะทำหน้าที่เพิ่มประสิทธิภาพในการจัดกลุ่มข้อมูลก่อนจะส่งต่อข้อมูลไปยังชั้นต่อไป
3. ชั้นข้อมูลส่งออก (Output layer) ทำหน้าที่ส่งออกข้อมูล โดยผ่านการประมวลผลจากฟังก์ชันผลรวม (Summation function: S) เป็นผลรวมของข้อมูลป้อนเข้าที่ถ่วงด้วยค่าน้ำหนัก (W) กับค่าความเอนเอียง (b) และฟังก์ชันการแปลง (Transfer function) จนได้ผลลัพธ์ ดังสมการที่ (3)

$$S = \sum_{i=1}^n w_i x_i + b \quad (3)$$

ฟังก์ชันกระตุ้น (Activation function) เป็นส่วนที่ทำหน้าที่แปลงผลรวมของข้อมูลป้อนเข้าให้เป็นผลลัพธ์ในชั้นส่งออก ซึ่งสามารถแบ่งได้ 2 ประเภท ดังนี้

1. ฟังก์ชันการแปลงเชิงเส้น เป็นฟังก์ชันที่สามารถเรียนรู้เพียงความสัมพันธ์เชิงเส้นระหว่างชั้นข้อมูลป้อนเข้าและชั้นข้อมูลส่งออก ดังนั้นจะไม่สามารถหาคำตอบได้สำหรับบางกรณีที่ไม่ใช่เส้นตรง ค่าผลลัพธ์ที่ได้จากการคำนวณจะมีค่าเท่ากับค่าที่ได้จากฟังก์ชันการแปลงเสมอเหมาะสำหรับผลลัพธ์ที่ไม่มีกำหนดขอบเขตค่าเป้าหมาย เช่นฟังก์ชันสเต็ป (Step function) ฟังก์ชันเครื่องหมาย (Step function) ฟังก์ชันการแปลงฮาร์ดลิมิต (Hard limit activation function)
2. ฟังก์ชันการแปลงไม่ใช่เชิงเส้น ประกอบด้วยฟังก์ชันการแปลง 2 ชนิด คือ
  - 2.1 ฟังก์ชันซิกมอยด์ (Sigmoid function) ฟังก์ชันการแปลงซิกมอยด์จะบีบช่วงข้อมูลป้อนเข้าที่ไม่จำกัดให้เป็นช่วงของข้อมูลส่งออกที่จำกัด โดยที่ช่วงของข้อมูลส่งออกจะอยู่ในช่วง 0 ถึง 1 ฟังก์ชันการแปลงซิกมอยด์แสดงลักษณะของข้อเท็จจริงที่มีความชันเข้าใกล้ศูนย์เมื่อข้อมูลป้อนเข้ามีจำนวนมาก
  - 2.2 ฟังก์ชันไฮเปอร์โบลิคแทนเจนต์ (Hyperbolic tangent function) มีลักษณะเช่นเดียวกับฟังก์ชันการแปลงซิกมอยด์ แต่ต่างกันเพียงช่วงของข้อมูลส่งออกจะอยู่ในช่วง -1 ถึง +1 โดยในงานวิจัยได้เลือกใช้ฟังก์ชันการแปลงนี้ สามารถคำนวณได้ ดังสมการที่ (4)

$$f(S) = \frac{2}{1 + e^{-2S}} - 1 \quad (4)$$

โครงข่ายประสาทเทียมสามารถแบ่งได้ 3 แบบ ได้แก่ (Gudise & Venayagamoorthy, 2003)

1. โครงข่ายประสาทแบบป้อนไปข้างหน้าชั้นเดียว (Single-layer feed forward neural networks) ประกอบด้วยชั้นสัญญาณประสาทขาเข้าและชั้นสัญญาณประสาทขาออกเท่านั้น
2. โครงข่ายประสาทแบบป้อนไปข้างหน้าหลายชั้น (Multi-layer feed forward neural networks) มีลักษณะเช่นเดียวกับโครงข่ายประสาทแบบป้อนไปข้างหน้าชั้นเดียว แต่จะมีชั้นซ่อนเพิ่มขึ้น โดยอยู่ระหว่างชั้นข้อมูลนำเข้าและชั้นส่งออก
3. โครงข่ายประสาทแบบแพร่ย้อนกลับ (Back propagation neural networks) มีลักษณะเช่นเดียวกับโครงข่ายประสาทแบบป้อนไปข้างหน้า ซึ่งมีชั้นซ่อนอย่างน้อย 1 ชั้น และมีการวนซ้ำแบบป้อนย้อนกลับอย่างน้อยหนึ่งครั้ง โครงข่ายประสาทเทียมมีขั้นตอนการเรียนรู้ที่แตกต่างกัน สำหรับการวิจัยนี้ใช้โครงข่ายประสาทเทียมแบบแพร่ย้อนกลับโดยมีขั้นตอนการดำเนินการ ดังนี้
  - 3.1 กำหนดชั้นข้อมูลนำเข้า ชั้นซ่อนและชั้นข้อมูลส่งออก อัตราการเรียนรู้ จำนวนรอบ ค่าความคลาดเคลื่อนและค่าน้ำหนักให้โครงข่ายประสาท
  - 3.2 คำนวณค่าผลลัพธ์ของโครงข่าย และปรับค่าน้ำหนักในกรณีทีเพอร์เซ็ปตรอน จำแนกข้อมูลผิดพลาด
  - 3.3 ทำซ้ำชุดข้อมูลการเรียนรู้ (Learning data set) จนกระทั่งเพอร์เซ็ปตรอนจำแนกข้อมูลได้ผลลัพธ์ตามค่าเป้าหมายที่กำหนดไว้

## 2.8 ตัววัดประสิทธิภาพของแบบจำลองการจำแนกประเภทข้อมูล (Confusion matrix)

การทำงานของแบบจำลองไปใช้งานจริงนั้นจะเป็นจะต้องทราบประสิทธิภาพของแบบจำลองเสียก่อน โดยทั่วไปแล้วจะมีตัววัดที่นิยมใช้กันในงานวิจัยและการทำงานต่างๆ อยู่ 4 ค่า คือ (chengz, 2019; Duangklang & Kruakaew, 2019)

1. Precision เป็นการวัดความแม่นยำของแบบจำลอง โดยพิจารณาแยกทีละ Class
2. Recall เป็นการวัดความถูกต้องของแบบจำลอง โดยพิจารณาแยกทีละ Class
3. F-measure เป็นการวัดค่า Precision และ Recall พร้อมกันของแบบจำลองโดยพิจารณาแยกทีละ Class
4. Accuray เป็นการวัดความถูกต้องของแบบจำลองโดยพิจารณาทุก Class

**2.8.1 แนวคิดเกี่ยวกับ Confusion matrix** จากบทความ (Pacharawongsakda, 2014) ได้กล่าวไว้ว่า Confusion matrix คือตารางแบบจัตุรัส โดยมีจำนวนแถวเท่ากับจำนวน Class เช่นในตารางที่ 2 มี Class คำตอบอยู่ 2 คำตอบ คือ yes และ no ฉะนั้นตาราง Confusion matrix นี้จะเป็นการสร้างตาราง 2x2 ขึ้นมาดังในตารางที่ 2.3 โดยข้อมูลด้านคอลัมน์คือคลาสที่อยู่ในรูปแบบของเทรนนิ่งดาต้า (Actual) และข้อมูลในแนวแถว คือ Class ที่แบบจำลองทำนายมาได้ (Predicted)

ตารางที่ 2.3 แสดงตาราง Confusion matrix

Predicted / actual	Yes	No
Yes	TP	FP
No	FN	TN

จากตารางที่ 2.3 ค่าที่แสดงในช่องต่างๆ ประกอบไปด้วย

1. True Positive (TP) คือ จำนวนข้อมูลที่ทำนายถูกว่าเป็น Class ซึ่งกำลังสนใจอยู่
2. True Negative (TN) คือ จำนวนข้อมูลที่ทำนายถูกว่าเป็น Class ซึ่งไม่ได้สนใจอยู่
3. False Positive (FP) คือ จำนวนข้อมูลที่ทำนายผิดมาเป็น Class ซึ่งกำลังสนใจอยู่
4. False Negative (FN) คือ จำนวนข้อมูลที่ทำนายผิดมาเป็น Class ซึ่งกำลังสนใจอยู่

ค่าความแม่นยำในการจำแนกประเภทข้อมูลของ Class Yes (Positive) นั้นสามารถคำนวณได้จากสมการที่ 5

$$\text{True Positive Rate (TPR)} = \frac{TP}{TP + FN} \quad (5)$$

ค่าความแม่นยำในการจำแนกประเภทข้อมูลของ Class No (Negative) นั้นสามารถคำนวณได้จากสมการที่ 6

$$\text{True Negative Rate (TNR)} = \frac{TN}{FP + TN} \quad (6)$$

ค่า Precision เป็นการวัดความแม่นยำของแบบจำลอง โดยพิจารณาแยกทีละ Class ดังสมการที่ 7

$$\text{Precision} = \frac{TP}{TP + FP} \quad (7)$$

ค่า Recall เป็นการวัดความถูกต้องของแบบจำลอง โดยพิจารณาแยกทีละ Class ดังสมการที่ 8

$$\text{Recall} = \frac{TP}{TP + FN} \quad (8)$$

ค่า F-measure เป็นการวัดค่า Precision และ Recall พร้อมกันของแบบจำลองโดยพิจารณาแยกทีละ Class ดังสมการที่ 9

$$\text{F - measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (9)$$

ค่า Accuracy เป็นการวัดความถูกต้องของแบบจำลองโดยพิจารณารวมทุก Class คือ จำนวน True Positive (TP) ของทุก Class รวมกัน ดังสมการที่ 10

$$\text{Accuracy} = \frac{TP + TN}{TP + FN + FP + TN} \quad (10)$$

**2.8.2 วิธี Self Consistency Test** หรือบางครั้งเรียกว่า Use Training Set เป็นวิธีการที่ง่ายที่สุด นั่นคือ ข้อมูลที่ใช้ในการสร้างแบบจำลอง (model) และข้อมูลที่ใช้ในการทดสอบแบบจำลองเป็นข้อมูลชุดเดียวกัน กระบวนการนี้เริ่มจาก สร้างแบบจำลองด้วยข้อมูล เทรนนิ่ง คาต้า (training data) หลังจากนั้นนำแบบจำลองที่สร้างได้มาทำนายข้อมูลเทรนนิ่ง คาต้า ชุดเดิม (ม่วงน้อย et al., 2564)

**2.8.3 วิธี Split Test** เป็นการแบ่งข้อมูลด้วยการสุ่มออกเป็น 2 ส่วน เช่น 70% ต่อ 30% หรือ 80% ต่อ 20% โดยข้อมูลส่วนที่หนึ่ง (70% หรือ 80%) ใช้ในการสร้างแบบจำลองและข้อมูลส่วนที่สอง (30% หรือ 20%) ใช้ในการทดสอบประสิทธิภาพของแบบจำลอง (Pacharawongsakda, 2014; ชุตติภาส เจริญ & แสนราช, 2563)

**2.8.4 วิธี Cross-validation Test** วิธีนี้เป็นวิธีที่นิยมในการทำงานวิจัยเพื่อใช้ในการทดสอบประสิทธิภาพของแบบจำลองเนื่องจากผลที่ได้มีความน่าเชื่อถือ การวัด ประสิทธิภาพด้วยวิธี Cross-validation นี้จะทำการแบ่งข้อมูลออกเป็นหลายส่วน (มักจะแสดงด้วยค่า k) เช่น 5-fold cross-validation คือ ทำการแบ่งข้อมูลออกเป็น 5 ส่วน โดยที่แต่ละส่วนมีจำนวนข้อมูลเท่ากัน หรือ 10-fold cross-validation คือ การแบ่งข้อมูลออกเป็น 10 ส่วน โดยที่แต่ละส่วนมีจำนวนข้อมูลเท่ากัน หลังจากนั้นข้อมูลหนึ่งส่วนจะใช้เป็นตัวทดสอบประสิทธิภาพของแบบจำลอง ทำวนไปเช่นนี้จนครบจำนวนที่แบ่งไว้ (ลิ้มตระกูล et al., 2559)

## 2.9 Data Visualization (การสร้างมโนภาพ)

Data Visualization เป็นสิ่งหนึ่ง ที่ถูกนำมาใช้แสดงแทนคำพูด เป็นการใช้อภาพเพื่อแสดงข้อมูลในเชิงปริมาณที่วัดได้ ไม่ว่าจะเป็นตัวเลข แผนภูมิ กราฟ และอื่นๆอีกมากมาย คำว่า Data คือ ข้อมูล ส่วน Visualization คือ การมองเห็นเมื่อนำมารวมกันแล้วหมายถึง ข้อมูลที่มองเห็นได้ด้วยตนเอง (P.Pitchaya, 2020) โดยผู้ใช้สามารถเรียนรู้และจดจำข้อมูลผ่านการมองเห็นได้มากกว่าการใช้ประสาทสัมผัสอื่นๆ หรือจะกล่าวว่า Visualization ก็คือ การสร้างมโนภาพของสิ่งต่างๆที่เราสนใจขึ้นมาในใจ ซึ่งต่อมาได้กลายเป็นการนำภาพมาใช้กับการนำเสนอหรือนำมาเป็นกรอบความคิด ซึ่งได้นำไปใช้ในการสนับสนุนการตัดสินใจ

### 2.9.1 ประโยชน์ของ Data Visualization

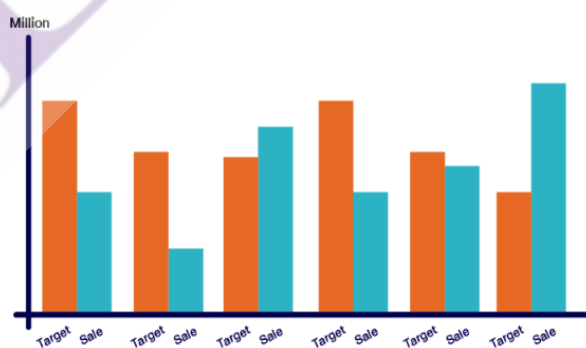
1. ช่วยในการแสดงข้อมูลที่มีปริมาณสูง
2. ช่วยในการรับรู้หรือคาดคะเนสิ่งที่อาจจะเกิดขึ้นในอนาคต

3. Visualization ไม่เพียงแต่แสดงรายละเอียดข้อมูลในตัวเองเท่านั้นยังสามารถแสดงปัญหาที่เกิดขึ้นได้ด้วย
4. ช่วยให้ผู้ใช้สามารถเข้าใจหรือวิเคราะห์ข้อมูลได้สะดวกขึ้น ไม่ว่าจะขอบเขตข้อมูลนั้นจะมีขนาดใหญ่หรือเล็ก
5. ช่วยให้ผู้ใช้สามารถสันนิษฐานข้อมูลได้สะดวกขึ้น

## 2.9.2 การเลือกรูปแบบ Visualization ให้เหมาะสมกับข้อมูล

ในปัจจุบันเป็นยุคที่เทคโนโลยีเข้าถึงทุกคน ทำให้การรับรู้ข่าวสาร ข้อมูลต่างๆ เป็นไปได้ง่าย และรวดเร็วมากขึ้นคนที่นำเสนอข้อมูลจึงต้องนำเสนอข้อมูลให้น่าสนใจ เข้าใจง่าย และรวดเร็ว จึงเกิดการสร้าง Data Visualization ขึ้นมา Data Visualization เป็นการใช้ภาพเพื่อแสดงข้อมูลในเชิงปริมาณที่วัดได้ซึ่งอาจนำเสนอออกมาในรูปแบบ แผนภูมิ กราฟ กราฟิก และอื่นๆ อีกมากมาย (9Expert, 2018; Muhammad, 2017) เพื่อให้เข้าใจได้ง่าย และรวดเร็ว โดยมีแผนภูมิต่างๆ ดังนี้

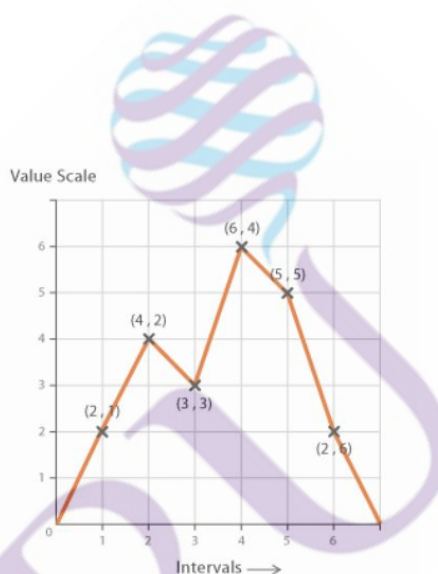
2.9.2.1 แผนภูมิแท่ง (Bar Charts) เป็นแผนภูมิที่ประกอบด้วยแกนนอน แกนตั้ง ที่นิยมแสดงออกมาในรูปแบบสี่เหลี่ยมที่สามารถบอกความสูงได้ เหมาะสำหรับการเปรียบเทียบจำนวนของข้อมูลในแต่ละจุดเช่น รายรับในแต่ละเดือน, ยอดขายที่ขายได้จริง เปรียบเทียบกับเป้าหมายยอดขายที่ตั้งไว้ เป็นต้น



ภาพที่ 2.8 แผนภูมิแท่ง (Bar Charts) (9Expert, 2018)

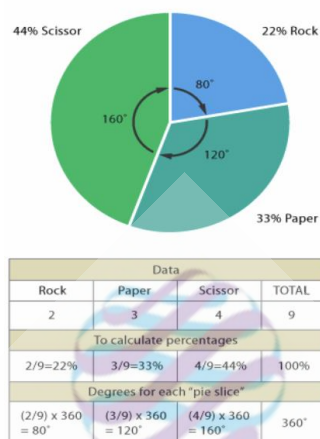


**2.9.2.2 แผนภูมิเส้น (Line Charts)** แผนภูมิเส้น มีลักษณะคล้ายแผนภูมิแท่ง ซึ่งประกอบด้วยแกนตั้งและนอน เพียงแต่เปลี่ยนจากแท่งข้อมูลเป็นจุดบนแผนภูมินั้นเองแผนภูมิประเภทนี้เหมาะกับการนำเสนอข้อมูลตัวเลขที่มีความสัมพันธ์กับข้อมูลที่มีลักษณะเป็นช่วงใช้แสดงให้เห็นการเปลี่ยนแปลงที่เกิดขึ้นตามเวลาเพื่อดูแนวโน้ม รวมถึงสามารถใช้พยากรณ์แนวโน้มในอนาคตได้



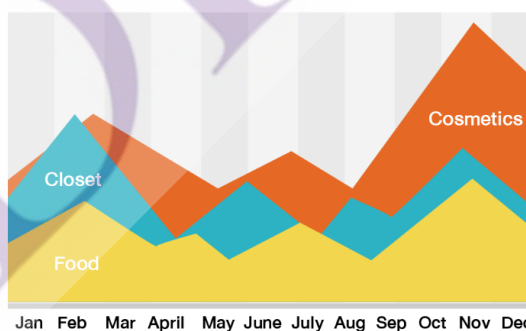
ภาพที่ 2.9 แผนภูมิเส้น (Line Charts) (Muhammad, 2017)]

**2.9.2.3 แผนภูมิมวงกลม (Pie Charts)** แผนภูมิมวงกลมเหมาะกับการนำเสนอข้อมูลที่มีส่วนประกอบย่อยที่รวมกันเป็นส่วนใหญ่ มีการแบ่งส่วนให้ดูง่าย และสวยงามแต่ในทางกลับกันอาจจะดูยากในเรื่องของการประมาณขนาดของแต่ละชิ้น ยิ่งถ้ามีจำนวนชิ้นมาก จะยิ่งแยกยากเพราะต้องใช้หลายสี ในการนำเสนอข้อมูล เช่น ส่วนแบ่งทางการตลาด (Market Share), ข้อมูลแสดงส่วนผสมต่างๆ เป็นต้น



ภาพที่ 2.10 แผนภูมิวงกลม (Pie Charts) (Muhammad, 2017)

**2.9.2.4 แผนภูมิพื้นที่ (Area Charts)** มีหน้าตาคล้ายแผนภูมิเส้น แต่มีการแรเงาพื้นที่ใต้เส้นข้อมูล หรือระหว่าง 2 เส้นเพื่อแสดงให้เห็นปริมาณความแตกต่างระหว่างเส้น เหมาะสำหรับเน้นความสำคัญของการเปลี่ยนแปลงในช่วงเวลาแสดงให้เห็นผลรวมของความแตกต่างระหว่างข้อมูล เช่น ข้อมูลของการซื้อสินค้าในห้างสรรพสินค้าในแต่ละเดือน ตามหมวดหมู่ต่างๆ ไล่ไป เครื่องสำอางค์ เสื้อผ้าแฟชั่น อาหาร ตามลำดับ

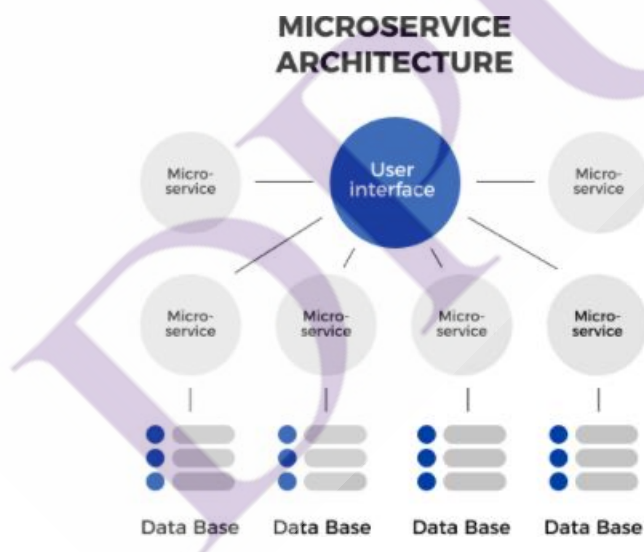


ภาพที่ 2.11 แผนภูมิพื้นที่ (Area Charts) (9Expert, 2018)

จากภาพที่ 2.8 - 2.11 จะเห็นได้ว่ารูปแบบการนำเสนอข้อมูลมีหลากหลาย โดยจะพิจารณาจากข้อมูลที่มีและศึกษาคูณสมบัติต่างๆของรูปแบบการนำเสนอข้อมูลเพื่อให้สามารถเลือกรูปแบบที่เหมาะสมกับการสื่อสารของข้อมูล

## 2.10 การพัฒนา Application ตามหลัก Microservice

Microservices เกิดจากแนวคิดการออกแบบสถาปัตยกรรมให้มีขนาดเล็กและมีความเป็นอิสระต่อกัน โดยแต่ละเซอร์วิสสามารถทำงานร่วมกันได้ ซึ่งการแบ่งเซอร์วิสนั้นต้องมีการกำหนด Bounded context ให้ชัดเจน ซึ่งการออกแบบที่ดีนั้นแต่ละเซอร์วิสจะต้องมีความผูกพันกันน้อยที่สุด (Loose coupling) และสามารถทำงานด้วยตัวของมันเองได้โดยไม่ต้องอาศัยเซอร์วิสอื่น (High cohesion) ดังภาพที่ 2.12 โดยวิธีการแบ่งเซอร์วิสต่างๆ นั้นจะใช้เทคนิคการกำหนด Bounded context โดยการจำกัดขอบเขตของแบบแผนธุรกิจและความต้องการทางธุรกิจให้ออกมาอยู่ในรูปแบบของ Context ดังภาพประกอบที่ 3 ซึ่งจะช่วยให้การออกแบบและพัฒนาเซอร์วิสทำได้ออกมาชัดเจนและง่ายขึ้น เมื่อสามารถกำหนด Bounded context ของแบบแผนธุรกิจทั้งหมดได้แล้ว ในส่วนของการพัฒนานั้นนักพัฒนาสามารถเลือกใช้ภาษาหรือเทคโนโลยีที่เหมาะสมกับความต้องการทางธุรกิจได้ (Karwatka, 2020; ทุมวงษ์, 2562)



ภาพที่ 2.12 สถาปัตยกรรมแบบ Microservice (Karwatka, 2020)

การทำให้เซอร์วิสติดต่อกับเซอร์วิสอื่นๆ นั้นโดยทั่วไปจะทำการติดต่อกันผ่าน REST (Representational state transfer) หรือ Message queue ซึ่งขึ้นอยู่กับการใช้งาน ซึ่งทั้งสองวิธีนี้สามารถพัฒนาโดยใช้ภาษาได้อย่างหลากหลายและมีประสิทธิภาพที่ดี โดยการพัฒนาสถาปัตยกรรมแบบ

Microservices ด้วยวิธีการออกแบบด้วย Bounded context ที่ดีนั้นสถาปัตยกรรมที่ถูกพัฒนาขึ้นมาควรมีคุณสมบัติดังต่อไปนี้

1. ความหลากหลายของเทคโนโลยี (technology heterogeneity) ในแอปพลิเคชันสถาปัตยกรรมที่ประกอบด้วยเซอร์วิสที่มีขนาดเล็กทำงานร่วมกันควรเลือกใช้ซอฟต์แวร์หรือภาษาที่เหมาะสมกับเซอร์วิสนั้น เพื่อให้แอปพลิเคชันสามารถรองรับปริมาณการเข้ามาใช้งานได้อย่างเหมาะสมตาม Context นั้นๆ
2. ความยืดหยุ่น (resilience) แอปพลิเคชันที่ถูกออกแบบมาด้วยแนวคิดแบบ Microservices เมื่อเซอร์วิสหนึ่งไม่สามารถใช้งานได้เซอร์วิสอื่นๆ ควรทำหน้าที่ต่อได้
3. การขยายขนาด (scaling) ในแอปพลิเคชันสถาปัตยกรรมทั่วไปที่ถูกพัฒนาบนแนวคิดแบบ Monolith นั้นการขยายแอปพลิเคชันนั้นจะทำการขยายทั้งแอปพลิเคชันไปด้วยกัน แต่สถาปัตยกรรมที่ถูกพัฒนาด้วย Microservices นั้นสามารถขยายแอปพลิเคชันเฉพาะเซอร์วิสที่มีปริมาณผู้ใช้งานสูงได้
4. ความง่ายต่อการส่งมอบ (ease of deployment) สถาปัตยกรรมแบบ Microservices สามารถพัฒนาและส่งมอบเฉพาะส่วนที่ถูกปรับปรุงหรือพัฒนาเพิ่มเติมได้
5. การจัดการองค์กร (organizational alignment) การใช้แนวคิดแบบ Microservices ในองค์กรใหญ่สามารถแบ่งทีมพัฒนาออกเป็นทีมเล็กๆ ตามความถนัดของทีมนั้นๆ โดยไม่ต้องเรียนรู้ภาษาใหม่ๆ เพื่อนำพัฒนาแอปพลิเคชันที่ใช้เพียงภาษาเดียว
6. การนำกลับมาใช้ใหม่ (composability) การนำเซอร์วิสที่มีฟังก์ชันในการทำงานคล้ายกันนำกลับมาใช้ใหม่เพื่อจุดประสงค์อื่นได้ เช่น การนำเซอร์วิสที่มีฟังก์ชันยืนยันตัวตนจากแบบแผนธุรกิจหนึ่งไปใช้ในอีกแบบแผนธุรกิจหนึ่งโดยไม่จำเป็นต้องพัฒนาแอปพลิเคชันนั้นใหม่
7. ความง่ายต่อการแทนที่ (optimizing for replaceability) การเปลี่ยนแปลงในบางเซอร์วิสนั้นต้องทำได้ง่ายและไม่กระทบต่อเซอร์วิสอื่น

## 2.11 งานวิจัยที่เกี่ยวข้อง

**2.11.1 Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images: Results of 10 convolutional neural networks** (Ardakani et al., 2020) ในงานวิจัยนี้ได้ทำการศึกษาจากผู้ป่วย 108 รายที่เป็นโรคปอดอักเสบจากโควิด-19 จากห้องปฏิบัติการ โดยเป็นชาย 60 คน หญิง 48 คน มีอายุเฉลี่ย  $61.45 \pm 15.04$  ทั้งนี้ชุดข้อมูลได้รวบรวมตั้งแต่ กันยายน 2563 - ธันวาคม 2563 จากโรงพยาบาลในมหาวิทยาลัยที่วิจัย โดยทำการใช้เทคนิค Deep learning ในการจำแนกคุณลักษณะจากภาพเอกซเรย์คอมพิวเตอร์ เพื่อหาความแตกต่างในการจำแนกผู้ป่วยที่เป็นโควิด-19 และไม่เป็นโควิด ซึ่งแบบจำลองที่ใช้จะเป็น ResNet-101 เป็น Deep Convolutional Neural Networks โดยค่า Accuracy อยู่ที่ 99.51% และ Specificity อยู่ที่ 99.0%

**2.11.2 Automated detection of COVID-19 cases using deep neural networks with X-ray images** (Ozturk et al., 2020) งานวิจัยนี้ได้ประยุกต์ใช้การเรียนรู้ของเครื่อง (Machine Learning) ในการทำงานแบบอัตโนมัติ ทั้งนี้ยังได้ใช้ Deep learning ซึ่งเป็นเทคนิคยอดนิยมของปัญญาประดิษฐ์ (AI) โดยชุดข้อมูลเป็นภาพเอกซเรย์ทรวงอก ภาพรังสีวิทยา เป็นชุดข้อมูลแบบเปิดจากเว็บไซต์ <https://github.com/ieee8023/COVID-chestxray-dataset> โดยจะทำการจำแนกระหว่างผู้ป่วยโควิด-19 , โรคปอดอักเสบ และปอดปกติ จากภาพเอกซเรย์ทรวงอก เพื่อลดภาระงานของแพทย์ในการคัดกรองผู้ป่วยและช่วยประหยัดเวลามากขึ้น โดยค่า Accuracy อยู่ที่ 98.08% จากแบบจำลอง neural network

**2.11.3 Combination of four clinical indicators predicts the severe/critical symptom of patients infected COVID-19** (Sun et al., 2020, p. 19) ได้ทำการวิจัยและศึกษาจากตัวอย่างข้อมูลผู้ป่วยทั้งหมด 336 รายที่ติดเชื้อโควิด-19 ในเซี่ยงไฮ้ ประเทศจีน นอกจากนี้ยังรวบรวมการสังเกต / บันทึกทางคลินิกและทางห้องปฏิบัติการ 220 รายการ ตัวบ่งชี้ทางคลินิกสัมพันธ์กับอาการรุนแรง / อาการวิกฤต ได้มีการระบุ และได้มีการพัฒนาแบบจำลองสำหรับการทำนายอาการรุนแรง / วิกฤต โดยใช้ Support Vector Machine (SVM) โดยให้ค่า Accuracy อยู่ที่ 77.5% และ Specificity อยู่ที่ 78.4%

**2.11.4 Covid-19 Classification Using Deep Learning in Chest X-Ray Images** (Karhan & Akal, 2020) ได้ทำการวิจัยข้อมูลผู้ป่วยโรคโควิด-19 จากชุดข้อมูลภาพเอกซเรย์ทรวงอก ภาพรังสีวิทยา

เป็นชุดข้อมูลแบบเปิดจากเว็บไซต์ <https://github.com/ieee8023/COVID-chestxray-dataset> โดยได้ใช้เทคนิค Deep learning ในรูปแบบแบบจำลอง ResNet-18 (Residual Network) ซึ่งเป็น Deep Convolutional Neural Networks ที่ทาง Microsoft เผยแพร่ออกสู่สาธารณชนตั้งแต่ปี 2015 จากผลการทดลองโดยให้ค่า Accuracy 99.5%

#### 2.11.5 COVID-19 Patient Health Prediction Using Boosted Random Forest Algorithm

(Iwendi et al., 2020) งานวิจัยนี้ได้นำเสนอแบบจำลอง Random Forest ที่ปรับแต่งอย่างละเอียด โดยใช้ข้อมูลทางภูมิศาสตร์ การเดินทาง สุขภาพ และข้อมูลประชากรของผู้ป่วยโควิด-19 เพื่อคาดการณ์ความรุนแรงของเคสและผลลัพธ์ที่เป็นไปได้ การฟื้นตัว หรือการเสียชีวิต โดยข้อมูลที่ใช้จะเป็นข้อมูลแบบเปิดชื่อ “Novel Corona Virus 2019 Dataset” จากเว็บไซต์ [www.kaggle.com](http://www.kaggle.com) โดยแบบจำลอง Random Forest ได้ให้ค่า Accuracy อยู่ที่ 94%

**2.11.6 Rapid and accurate identification of COVID-19 infection through machine learning based on clinical available blood test results** (Wu et al., 2020) ได้ทำการวิจัยและศึกษาจากผลการตรวจเลือดเพื่อประเมินผู้ต้องสงสัยในการติดเชื้อ โรคโควิด-19 ในการศึกษานี้ได้รวบรวมตัวอย่างทั้งหมด 253 ตัวอย่าง จากผู้ป่วยต้องสงสัย 169 รายจากในหลายๆ แห่ง รวมทั้งในโรงพยาบาลปอดหลานโจวที่ประเทศจีน โดยทำการตรวจเลือด 49 ราย โดยจะนำค่าจากผลตรวจเลือด 11 ค่ามาวิเคราะห์ผ่านโปรแกรม Covid-19 Assistant Discrimination 2.0 ([http://lishuyan.lzu.edu.cn/COVID2019\\_2/](http://lishuyan.lzu.edu.cn/COVID2019_2/)) เพื่อทำนายโอกาสการติดโรคโควิด-19 จากการวิจัยนี้ได้ใช้แบบจำลอง Random forest โดยให้ค่า Accuracy 95.95% และ Specificity อยู่ที่ 96.95%

#### 2.11.7 COVID-19 detection using federated machine learning (Abdul Salam et al., 2021)

ได้ทำการวิจัยข้อมูลผู้ป่วยโรคโควิด-19 โดยใช้ข้อมูลจากภาพเอ็กซเรย์ทรวงอกซึ่งเป็นข้อมูลแบบเปิดจำนวน 5,144 ภาพ จากเว็บไซต์ [www.kaggle.com](http://www.kaggle.com) เพื่อจำแนกผู้ป่วยและผู้ต้องสงสัยจากโรคโควิด-19 จากงานวิจัยนี้ได้ใช้ Stochastic gradient descent (SDG) ในการวิเคราะห์และจำแนกผู้ป่วยจากภาพเอ็กซเรย์ทรวงอก



### 2.11.8 Machine learning-based prediction of COVID-19 diagnosis based on symptoms

(Zoabi et al., 2021) ได้ทำการวิจัยและสร้างแบบการทำนายผลการวินิจฉัยโรคโควิด-19 ตามอาการโดยใช้การเรียนรู้ของเครื่อง โดยข้อมูลทั้งหมดที่ใช้ในการศึกษานี้ได้มาจากเว็บไซต์กระทรวงสาธารณสุขของอิสราเอลจากเว็บไซต์ <https://github.com/nshomron/covidpred>. ในรูปแบบไฟล์คอมพิวเตอร์ (CSV) จำนวน 4,769 รายตั้งแต่เดือน 22 มีนาคม 2563 – 31 มีนาคม 2563 โดยได้นำเสนอแบบจำลองในการหาผลตรวจที่เป็น positive หรือหาผลบวกจากโรคโควิด-19 และได้ใช้แบบจำลองต้นไม้ตัดสินใจ (Decision-tree) โดยจากผลการทดลองพบค่า Accuracy อยู่ที่ 90%

### 2.11.9 Accurate detection of Covid-19 patients based on Feature Correlated Naive Bayes

(FCNB) classification strategy (Mansour et al., 2021) แนะนำกลยุทธ์การวินิจฉัยโรคโควิด-19 ใหม่โดยจะมี 4 วิธี คือ Feature Selection Phase (FSP) , Feature Clustering Phase (FCP), Master Feature Weighting Phase (MFWP) และ Feature Correlated Naive Bayes Phase (FCNBP) จากผลการทดลองได้พิสูจน์ประสิทธิภาพของกลยุทธ์ FCNBP ซึ่งมีประสิทธิภาพมากที่สุด โดยมีค่า Accuracy อยู่ที่ 99% ซึ่งงานวิจัยนี้ได้แนะนำแบบจำลอง Naive Bayes ในการทำนายข้อมูล

### 2.11.10 Naive Bayes Classifier for Predicting the Novel Coronavirus (Bhatia & Malhotra,

2021) ออกแบบกลไกในการทำนายโรคโควิด-19 โดยใช้การจำแนกคุณลักษณะของอาการที่พบบ่อยที่สุดจากการติดเชื้อ โดยมี 4 คุณลักษณะเช่น Sore Throat Cough , Runny Nose , Fever Difficulty in Breathing โดยเป้าหมายของคุณลักษณะคือ มีผลเป็น positive หรือผลที่เป็นบวกว่าติดโรคโควิด-19 หรือไม่ โดยได้ใช้แบบจำลอง Naive Bayes เป็นตัวขับเคลื่อนการทำงานนี้ ทั้งนี้กระบวนการวิเคราะห์โดยจะแยก 2 ค่า เช่น ติดเชื้อโควิด-19 = yes , ไม่ติดเชื้อโควิด = no มาทำนายอีกกว่าจาก 2 ค่านี้นี้ผู้ป่วยมีความทุกข์ทรมานจากโควิด-19 หรือไม่

### ข้อสรุปงานวิจัยที่เกี่ยวข้องทั้งหมด

จากงานวิจัยที่เกี่ยวข้องผู้วิจัยได้ทำการสรุปและแยกเป็น 2 หมวดหมู่ด้วยกันคือ 1. จำแนกผู้ป่วยโรคโควิด-19 จากภาพเอ็กซเรย์ทรวงอก 2. จำแนกผู้ป่วยโรคโควิด-19 จากชุดข้อมูลหลากหลาย



คุณลักษณะ โดยทุกงานจะมีการใช้อัลกอริทึม (Algorithm) ที่แตกต่างกันออกไป เป็นงานวิจัยเพื่อวัตถุประสงค์ในการจำแนกผู้ป่วย , ทำนายโอกาสการติดเชื้อโควิดหรือทำนายผู้ต้องสงสัยว่าติดเชื้อหรือไม่ ซึ่งจะแตกต่างจากงานวิจัยฉบับนี้ที่ผู้วิจัยได้ให้ความสำคัญกับการจำแนกผู้ป่วยในกลุ่มต่างๆ เพื่อช่วยกรองผู้ต้องสงสัยได้อย่างรวดเร็ว ซึ่งจะช่วยประหยัดเวลาได้เร็วขึ้น (Kasetsart university, 2020) โดยเบื้องต้นผู้วิจัยได้ยึด “แนวทางการดำเนินงานคัดกรอง เฝ้าระวัง และสอบสวนโรคติดเชื้อไวรัสโคโรนา 2019” (กรมควบคุมโรค, 2563) โดยจะเป็นการตรวจสอบสำหรับการเฝ้าระวังในกลุ่มต่างๆ โดยจะแบ่งเป็น 4 กลุ่มดังนี้

1. การเฝ้าระวังในกลุ่มผู้ป่วย หรือมีอาการเข้าได้กับนิยาม (Patient Under Investigation : PUI)
2. การตรวจคัดกรองในประชากรเสี่ยงตามจุดคัดกรองและด่านเข้าออกระหว่างประเทศ (Screening)
3. การเฝ้าระวังในกลุ่มเป้าหมายเฉพาะหรือพื้นที่เฉพาะ (Sentinel Surveillance)
4. การเฝ้าระวังเหตุการณ์ในสถานที่เสี่ยง เก็บตัวอย่างส่งตรวจเมื่อเข้าเกณฑ์ PUI และเป็นกลุ่มก่อน

ตารางที่ 2.4 เปรียบเทียบการใช้เทคนิคการเรียนรู้ของเครื่องในงานวิจัยที่ได้ทำการศึกษา

ผู้ วิ จั ย / เทคนิค	Neural networks	Support vector machine	Random forest	Decision tree	Naive Bayes	Stochastic gradient descent
(Ardakani et al., 2020)	✓					

ตารางที่ 2.4 เปรียบเทียบการใช้เทคนิคการเรียนรู้ของเครื่องในงานวิจัยที่ได้ทำการศึกษา (ต่อ)

ผู้วิจัย / เทคนิค	Neural networks	Support vector machine	Random forest	Decision tree	Naive Bayes	Stochastic gradient descent
(Ozturk et al., 2020)	✓					
(Sun et al., 2020, p. 19)		✓				
(Iwendi et al., 2020)			✓			
(Karhan & Akal, 2020)	✓					
(Wu et al., 2020)			✓			
(Abdul Salam et al., 2021)						✓
(Bhatia & Malhotra, 2021)					✓	
(Mansour et al., 2021)					✓	
(Zoabi et al., 2021)				✓		

จากตารางที่ 2.4 จะเห็นได้ว่างานวิจัยที่ได้ทำการศึกษาทั้งหมดพบว่ามีการใช้เทคนิคแบบจำลอง Neural networks มากที่สุด ถัดมาจะเป็น Random forest และ Naive Bayes ผู้วิจัยจึงได้นำเทคนิคทั้ง 3 แบบจำลองนี้มาเป็นอัลกอริทึมในการทำวิจัยในการทดลองแบบจำลองที่ดีที่สุดจาก 3 แบบจำลองนี้ ว่าแบบจำลองไหนเหมาะสมและให้ค่าความถูกต้องมากที่สุด เมื่อได้ผลลัพธ์ที่ดีที่สุดเพียง 1 แบบจำลองแล้ว จึงจะนำไปพัฒนาระบบการจำแนกผู้ป่วยต่อไป



## บทที่ 3

### วิธีการดำเนินงานวิจัย

แนวคิดหลักการวิจัยนี้คือการนำเทคนิคการเรียนรู้ของเครื่องมาประยุกต์ใช้ในการจำแนกผู้ป่วยโควิด-19 ซึ่งข้อมูลที่ใช้ในการทำวิจัยนี้ได้มาจากกรมควบคุมโรค (Department of disease control, 2021) จากเว็บไซต์ <https://covid19.ddc.moph.go.th/> ในลักษณะรูปแบบ API โดยอัลกอริทึมที่ใช้ในการพัฒนาแบบจำลองจะประกอบไปด้วย 3 เทคนิคคือ Neural network Naive bayes และ Random Forest และหาค่าความถูกต้องหรือค่าความแม่นยำมากที่สุดจาก 3 อัลกอริทึมข้างต้น กรอบงานวิจัยนี้จะแบ่งเป็น 5 ขั้นตอนหลัก ได้แก่

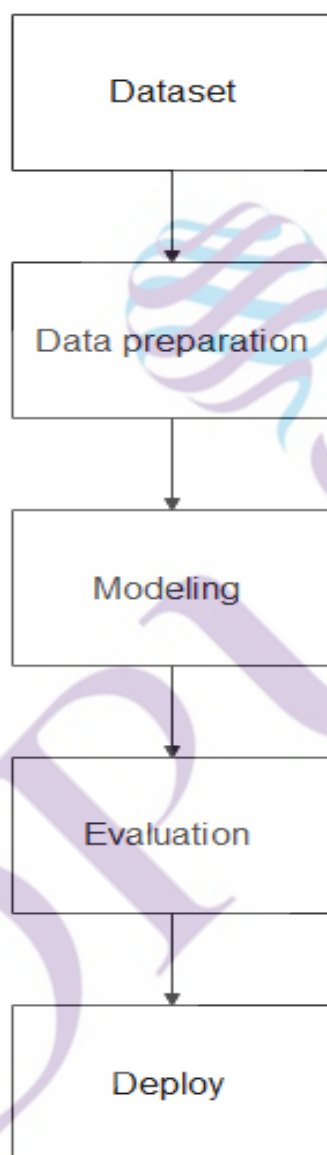
3.1 Dataset : วิธีการรวบรวมข้อมูล

3.2 Data preparation : การเตรียมข้อมูล เช่น การจัดกลุ่มข้อมูลใหม่ และ การทำความสะอาดข้อมูล

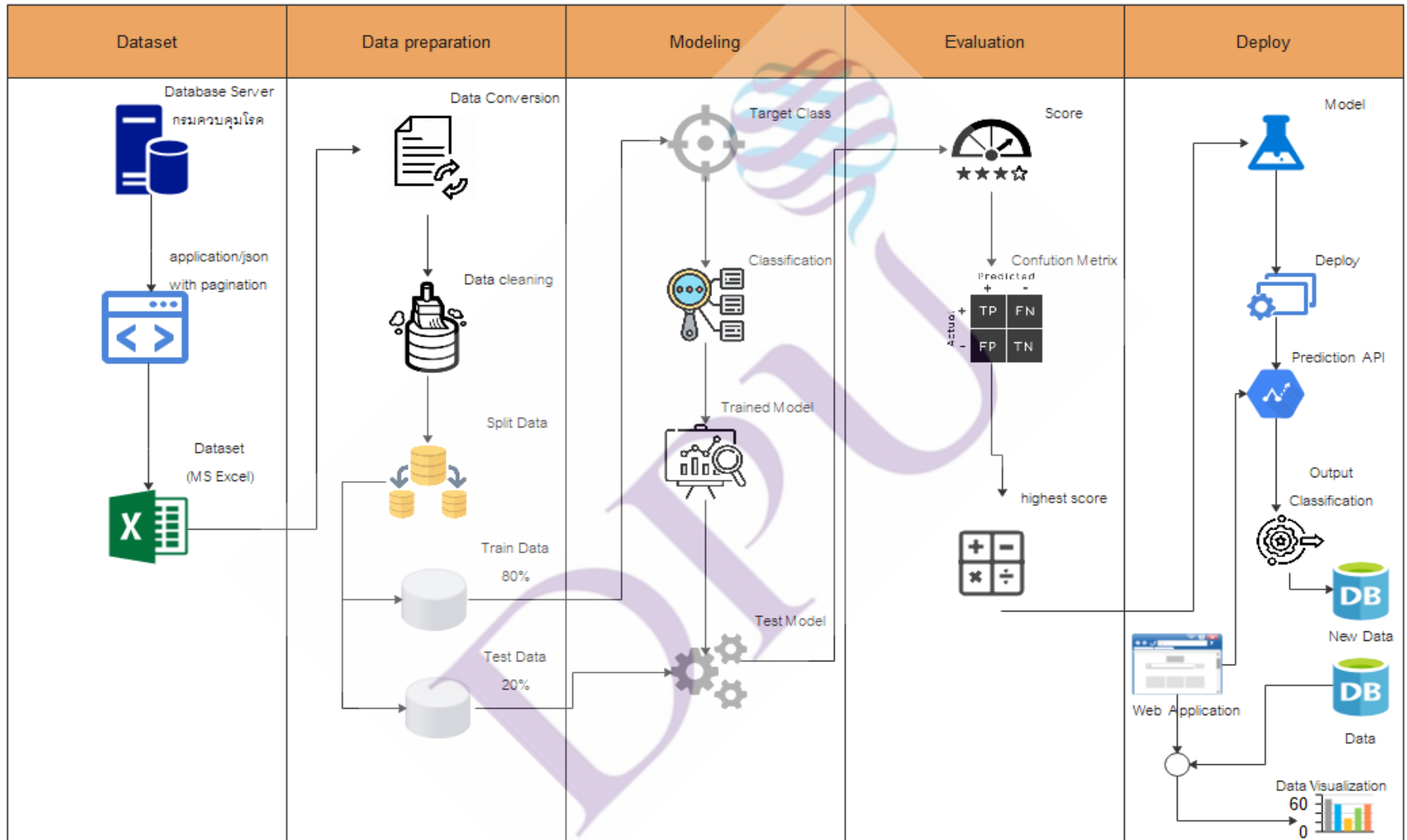
3.3 Modeling : การกำหนดคุณลักษณะที่ใช้ในการวิจัยและการสร้างแบบจำลอง ด้วยอัลกอริทึมไหนให้ค่าความแม่นยำและความถูกต้องมากที่สุด

3.4 Evaluation : การวัดประสิทธิภาพของอัลกอริทึมที่ใช้ในการพัฒนาแบบจำลอง จาก 3 เทคนิค เช่น คือ Neural network ,Naive bayes และ Random Forest

3.5 Deploy : จากหัวข้อที่ 3.4 เมื่อได้อัลกอริทึมที่ดีที่สุดเพียง 1 เทคนิคแล้ว จะนำไปใช้ในการจำแนกผู้ป่วย โดยจะนำไปใช้กับข้อมูลใหม่



ภาพที่ 3.1 กระบวนการทำงานวิจัย



ภาพที่ 3.2 แผนผังกระบวนการทำงานวิจัย

### 3.1 Data Set (วิธีการรวบรวมข้อมูล)

ทำการรวบรวมข้อมูลจากศูนย์กลางข้อมูลภาครัฐ ซึ่งเป็นข้อมูลแบบเปิดจากกรมควบคุมโรค กระทรวงสาธารณสุข โดยจะเป็นข้อมูลผู้ป่วยโควิด-19 ระลอกที่ 1 ถึงระลอกที่ 3 (Department of disease control, 2021) ตั้งแต่ 1 มกราคม 2563 ถึง วันที่ 1 ตุลาคม 2564 เป็นจำนวน 1,608,923 ราย จากนั้นทำการเขียนโปรแกรมเชื่อมต่อข้อมูลผ่าน API ที่ทางกรมควบคุมโรคทำไว้ดังภาพที่ 3.3



ภาพที่ 3.3 ขั้นตอนการพัฒนาส่วนการดาวน์โหลดข้อมูล

เมื่อทำการเชื่อมต่อ API แล้วจากนั้นทำการติดตั้งส่วนเสริมของการพัฒนาเว็บไซต์ที่ชื่อว่า datatables เพื่อแสดงผลข้อมูลในรูปแบบตาราง เพื่อดาวน์โหลดไฟล์ Microsoft Excel โดยจะปรากฏตัวอย่างดังภาพที่ 3.4

txn_date	gender	age_number	age_range	nationality	job	risk	patient_type	province	update_date
2021-09-30	ชาย	20	20-29 ปี	Thai		สัมผัสใกล้ชิดกับผู้ป่วยยืนยันรายก่อนหน้านี้	2. สัมผัสผู้ติดเชื้อ	พังงา	2021-09-30 07:43:00
2021-09-30	ชาย	41	40-49 ปี	Thai		สัมผัสใกล้ชิดกับผู้ป่วยยืนยันรายก่อนหน้านี้	2. สัมผัสผู้ติดเชื้อ	พังงา	2021-09-30 07:43:00
2021-09-30	ชาย	1	< 10 ปี	Thai		สัมผัสใกล้ชิดกับผู้ป่วยยืนยันรายก่อนหน้านี้	2. สัมผัสผู้ติดเชื้อ	พังงา	2021-09-30 07:43:00
2021-09-30	ชาย	45	40-49 ปี	Thai		สัมผัสใกล้ชิดกับผู้ป่วยยืนยันรายก่อนหน้านี้	2. สัมผัสผู้ติดเชื้อ	พังงา	2021-09-30 07:43:00
2021-09-30	หญิง	26	20-29 ปี	Thai		สัมผัสใกล้ชิดกับผู้ป่วยยืนยันรายก่อนหน้านี้	2. สัมผัสผู้ติดเชื้อ	พังงา	2021-09-30 07:43:00
2021-09-30	หญิง	40	40-49 ปี	Thai		สัมผัสใกล้ชิดกับผู้ป่วยยืนยันรายก่อนหน้านี้	2. สัมผัสผู้ติดเชื้อ	พังงา	2021-09-30 07:43:00
2021-09-30	หญิง	34	30-39 ปี	Thai		อื่นๆ	10. อื่นๆ	กรุงเทพมหานคร	2021-09-30 07:43:00
2021-09-30	ชาย	22	20-29 ปี	Thai		อื่นๆ	10. อื่นๆ	กรุงเทพมหานคร	2021-09-30 07:43:00
2021-09-30	หญิง	66	60-69 ปี	Thai		อื่นๆ	10. อื่นๆ	กรุงเทพมหานคร	2021-09-30 07:43:00
2021-09-30	หญิง	46	40-49 ปี	Thai		อื่นๆ	10. อื่นๆ	กรุงเทพมหานคร	2021-09-30 07:43:00

Showing 1 to 10 of 5,000 entries

Previous 1 2 3 4 5 ... 500 Next

ภาพที่ 3.4 ตารางแสดงข้อมูลโควิด-19 จาก API



จากภาพที่ 3.4 ข้อมูลที่เชื่อมต่อกันผ่าน API โดยการแสดงผลข้อมูลจะอยู่ที่ 5,000 รายการ Request API ใน 1 ครั้ง ซึ่งข้อมูลมี 1,608,923 รายการ จึงต้องทำการดาวน์โหลด 313 ครั้ง หรือเท่ากับ ไฟล์ Microsoft Excel 313 ไฟล์ หลังจากนั้นจะนำไฟล์ทั้งหมดมารวมข้อมูลเป็นไฟล์เดียว เพื่อใช้เตรียม ทำ Data Preparation ต่อไป โดยตัวอย่างไฟล์ข้อมูลสามารถแสดงได้ดังภาพที่ 3.5

date	gender	age_number	age_range	nationality	nationality_j	job	risk	patient_type	province	update_date	
1-04-01	ชาย	24	20-29 ปี	Thailand	ไทย		สถานบันเทิง	ผู้ป่วย PUI	กรุงเทพมหานคร	2021-09-01 07:4C	
1-04-01	ชาย	23	20-29 ปี	Thailand	ไทย		สถานบันเทิง	สัมผัสผู้ติดเชื้อ	กรุงเทพมหานคร	2021-09-01 07:4C	
1-04-01	ชาย	29	20-29 ปี	Myanmar	ชาวล่างชาติ	รับจ้างทั่วไป	สัมผัสใกล้ชิดกับผู้ป่วยยืนยันรายก่อนหน้า	สัมผัสผู้ติดเชื้อ	กรุงเทพมหานคร	2021-09-01 07:4C	
1-04-01	ชาย	27	20-29 ปี	Myanmar	ชาวล่างชาติ	พนักงาน Cluster	สมุทรสาคร	ขอตรวจหาเชื้อเอง	สมุทรสาคร	2021-09-01 07:4C	
1-04-01	ชาย	28	20-29 ปี	Myanmar	ชาวล่างชาติ	ว่างงาน	Cluster สมุทรสาคร	ขอตรวจหาเชื้อเอง	สมุทรสาคร	2021-09-01 07:4C	
1-04-01	ชาย	39	30-39 ปี	Thailand	ไทย		ค้าขาย/ธุรกิจ	ไม่สถานที่ชุมชน เช่น ตลาดนัด สถานที่ท่องเที่ยว	ผู้ป่วย PUI	สมุทรปราการ	2021-09-01 07:4C
1-04-01	ชาย	33	30-39 ปี	Thailand	ไทย		ข้าราชการ	สัมผัสใกล้ชิดกับผู้ป่วยยืนยันรายก่อนหน้า	สัมผัสผู้ติดเชื้อ	กรุงเทพมหานคร	2021-09-01 07:4C
1-04-01	ชาย	37	30-39 ปี	Thailand	ไทย		ค้าขาย/ธุรกิจ	สัมผัสใกล้ชิดกับผู้ป่วยยืนยันรายก่อนหน้า	สัมผัสผู้ติดเชื้อ	สมุทรปราการ	2021-09-01 07:4C
1-04-01	ชาย	43	40-49 ปี	Thailand	ไทย		ค้าขาย/ธุรกิจ	ไม่สถานที่ชุมชน เช่น ตลาดนัด สถานที่ท่องเที่ยว	ผู้ป่วย PUI	กรุงเทพมหานคร	2021-09-01 07:4C
1-04-01	ชาย	57	50-59 ปี	Thailand	ไทย		พนักงาน State Quarantine	จากต่างประเทศ อยู่ใน Quarantine	กรุงเทพมหานคร	2021-09-01 07:4C	
1-04-01	ชาย	58	50-59 ปี	Thailand	ไทย		ไม่ระบุ	State Quarantine	ชลบุรี	2021-09-01 07:4C	
1-04-01	หญิง	16	10-19 ปี	Thailand	ไทย		ไม่ระบุ	สัมผัสใกล้ชิดกับผู้ป่วยยืนยันรายก่อนหน้า	สัมผัสผู้ติดเชื้อ	กรุงเทพมหานคร	2021-09-01 07:4C
1-04-01	หญิง	18	10-19 ปี	Myanmar	ชาวล่างชาติ	พนักงาน Cluster	สมุทรสาคร	ขอตรวจหาเชื้อเอง	สมุทรสาคร	2021-09-01 07:4C	
1-04-01	หญิง	28	20-29 ปี	Thailand	ไทย		พนักงานสัมผัสใกล้ชิดกับผู้ป่วยยืนยันรายก่อนหน้า	สัมผัสผู้ติดเชื้อ	กรุงเทพมหานคร	2021-09-01 07:4C	
1-04-01	หญิง	25	20-29 ปี	Myanmar	ชาวล่างชาติ	พนักงาน Cluster	สมุทรสาคร	สำรวจกลุ่มเสี่ยง (survey)	สมุทรสาคร	2021-09-01 07:4C	
1-04-01	หญิง	34	30-39 ปี	Thailand	ไทย		State Quarantine	จากต่างประเทศ อยู่ใน Quarantine	กรุงเทพมหานคร	2021-09-01 07:4C	
1-04-01	หญิง	30	30-39 ปี	Myanmar	ชาวล่างชาติ	พนักงาน Cluster	สมุทรสาคร	สำรวจกลุ่มเสี่ยง (survey)	สมุทรสาคร	2021-09-01 07:4C	
1-04-01	หญิง	30	30-39 ปี	Myanmar	ชาวล่างชาติ	พนักงาน Cluster	สมุทรสาคร	สำรวจกลุ่มเสี่ยง (survey)	สมุทรสาคร	2021-09-01 07:4C	
1-04-01	หญิง	30	30-39 ปี	Myanmar	ชาวล่างชาติ	พนักงาน Cluster	สมุทรสาคร	สำรวจกลุ่มเสี่ยง (survey)	สมุทรสาคร	2021-09-01 07:4C	
1-04-01	หญิง	30	30-39 ปี	Myanmar	ชาวล่างชาติ	พนักงาน Cluster	สมุทรสาคร	สำรวจกลุ่มเสี่ยง (survey)	สมุทรสาคร	2021-09-01 07:4C	
1-04-01	หญิง	30	30-39 ปี	Myanmar	ชาวล่างชาติ	พนักงาน Cluster	สมุทรสาคร	สำรวจกลุ่มเสี่ยง (survey)	สมุทรสาคร	2021-09-01 07:4C	
1-04-01	หญิง	30	30-39 ปี	Myanmar	ชาวล่างชาติ	พนักงาน Cluster	สมุทรสาคร	สำรวจกลุ่มเสี่ยง (survey)	สมุทรสาคร	2021-09-01 07:4C	
1-04-01	หญิง	57	50-59 ปี	Thailand	ไทย		พ่อบ้าน/แม่	สัมผัสใกล้ชิดกับผู้ป่วยยืนยันรายก่อนหน้า	สัมผัสผู้ติดเชื้อ	สมุทรปราการ	2021-09-01 07:4C
1-04-01	หญิง	57	50-59 ปี	Myanmar	ชาวล่างชาติ	พนักงาน Cluster	สมุทรสาคร	สำรวจกลุ่มเสี่ยง (survey)	สมุทรสาคร	2021-09-01 07:4C	
1-04-01	หญิง	62	60-69 ปี	Thailand	ไทย		State Quarantine	จากต่างประเทศ อยู่ใน Quarantine	กรุงเทพมหานคร	2021-09-01 07:4C	

ภาพที่ 3.5 ตัวอย่างชุดข้อมูลในรูปแบบไฟล์ Microsoft Excel

ตารางที่ 3.1 คำอธิบาย ชุดข้อมูลผู้ป่วยโควิด-19

ลำดับ	Attribute	Description
1	txn_date	วันแถลง
2	gender	เพศ
3	age_number	อายุ
4	age_range	ช่วงอายุ
5	nationality	สัญชาติ
6	job	อาชีพ

ตารางที่ 3.1 คำอธิบาย ชุดข้อมูลผู้ป่วยโควิด-19 (ต่อ)

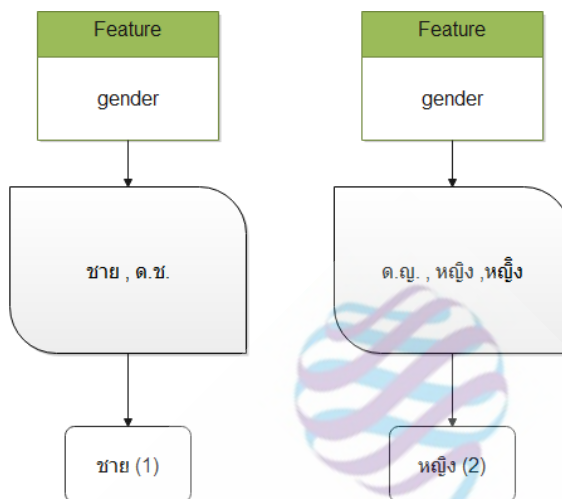
ลำดับ	Attribute	Description
7	risk	ปัจจัยเสี่ยง
8	patient_type	ประเภทผู้ป่วย
9	province	จังหวัด
10	update_date	วันที่ปรับปรุงข้อมูลล่าสุด

### 3.2 Data Preparation (การเตรียมข้อมูล)

#### 3.2.1 Formatting (การจัดรูปแบบข้อมูล)

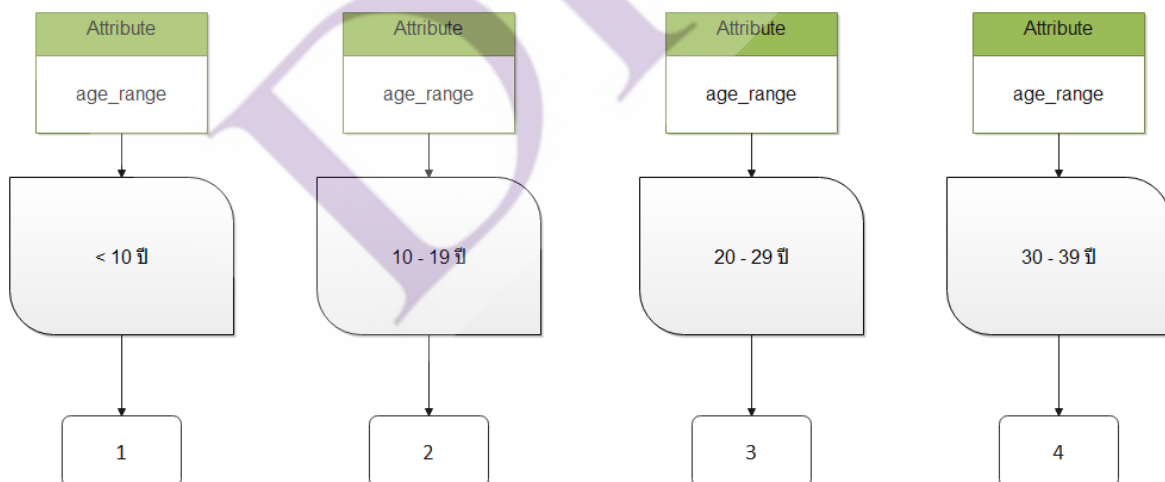
ขั้นตอนนี้เป็นขั้นตอนที่ทำการแปลงข้อมูลที่ได้ทำการเก็บรวบรวมมา ให้กลายเป็นข้อมูลที่สามารถนำไปวิเคราะห์ในขั้นถัดไปได้ ในการแปลงข้อมูลนี้อาจจะต้องมีการทำข้อมูลให้ถูกต้อง เช่น การแปลงข้อมูลให้อยู่ในช่วง (scale) เดียวกัน โดยจะทำการแปลงข้อมูลจากตัวอักษร (text) เป็น ตัวเลข (numeric) เช่น แอททริบิวต์ และ Values ทั้งหมด รวมทั้งจัดกลุ่ม แอททริบิวต์ ของชุดข้อมูล เช่น gender age\_rang nantionality,risk patien\_type และ province

จากแอททริบิวต์ดังกล่าวยังมีคลาสของข้อมูลที่มีความซ้ำซ้อนกัน จึงต้องมีการจัดกลุ่มและรวมคลาส ใหม่ เช่น แอททริบิวต์ ของ gender มี ชาย, หญิง ,ค.ญ ,เด็กหญิง จะเห็นได้ว่าข้อมูลนั้นมีความซ้ำซ้อนหรือทับกันผู้วิจัยจึงจัดกลุ่มใหม่เป็น ชาย => ชาย และ ค.ญ ,เด็กหญิง =>หญิง ดังภาพที่ 3.6

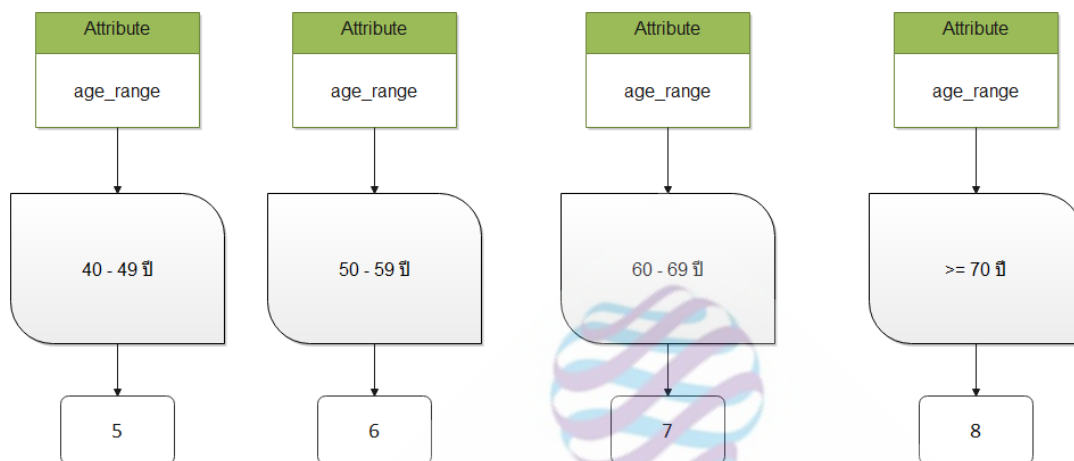


ภาพที่ 3.6 การจัดกลุ่ม gender (เพศ)

จากชุดข้อมูลตามตารางที่ 3.1 มีแอททริบิวต์ `age_number` และ `age_range` ซึ่งเป็นข้อมูลที่มีความสัมพันธ์กัน โดย `age_number` คือ อายุ ส่วน `age_range` คือ กลุ่มอายุ โดยการทำงานจะเลือกเพียง `age_range` มาเป็น แอททริบิวต์ ในการวิจัยเพราะเป็นข้อมูลประเภท categorical และทำการจัดกลุ่มแปลงเป็นตัวเลขดังภาพที่ 3.7 – 3.8

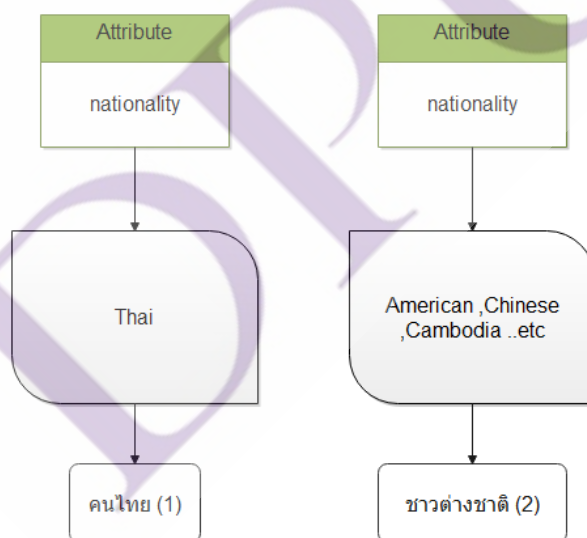


ภาพที่ 3.7 การจัดกลุ่ม age\_range (ช่วงอายุ)



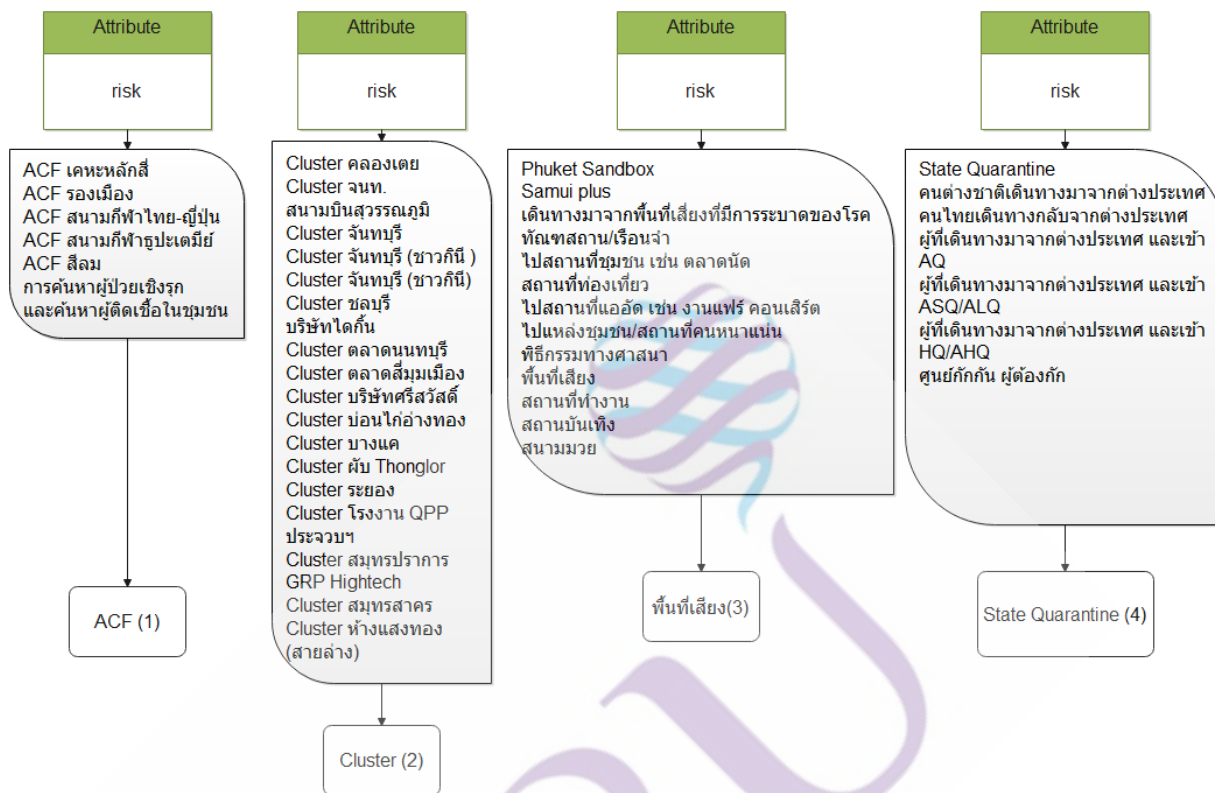
ภาพที่ 3.8 การจัดกลุ่ม age\_range (ช่วงอายุ) (ต่อ)

ถัดมาจะเป็นข้อมูล nationality (สัญชาติ) ซึ่งจากชุดข้อมูลจะมี คลาส จำนวนมากเพราะมีหลายประเทศ จึงจัดกลุ่มเป็น 2 คลาส เท่านั้น โดยจะประกอบไปด้วย คนไทยและคนต่างชาติจากนั้นทำการจัดกลุ่ม แปลงเป็นตัวเลขดังภาพที่ 3.9

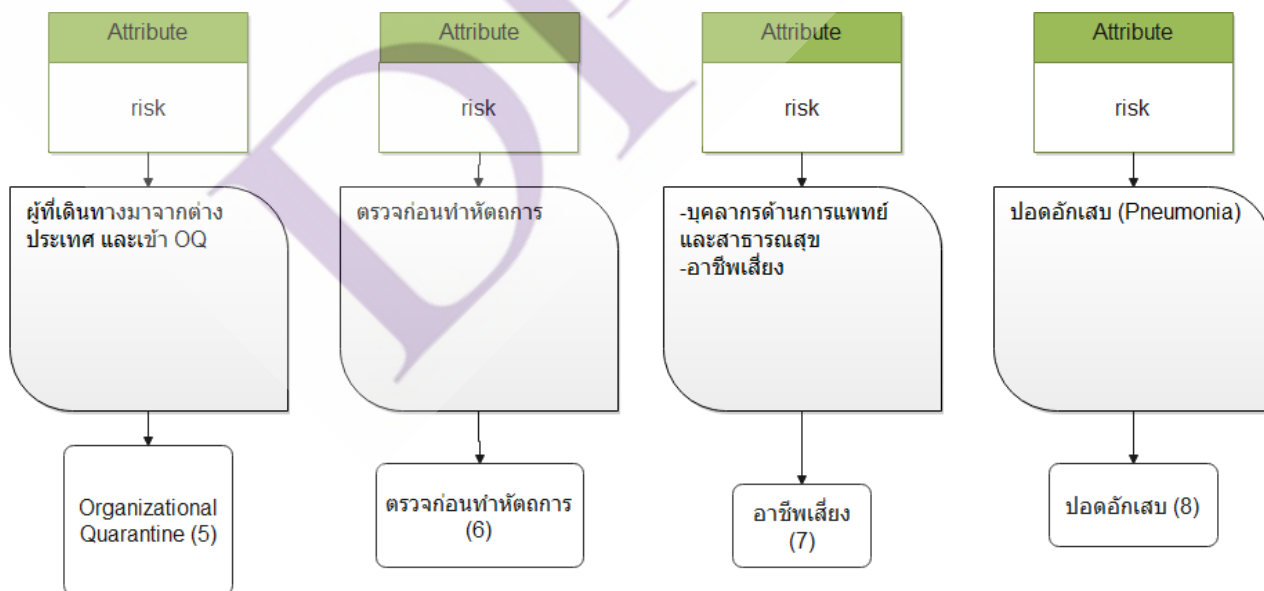


ภาพที่ 3.9 การจัดกลุ่ม nationality (สัญชาติ)

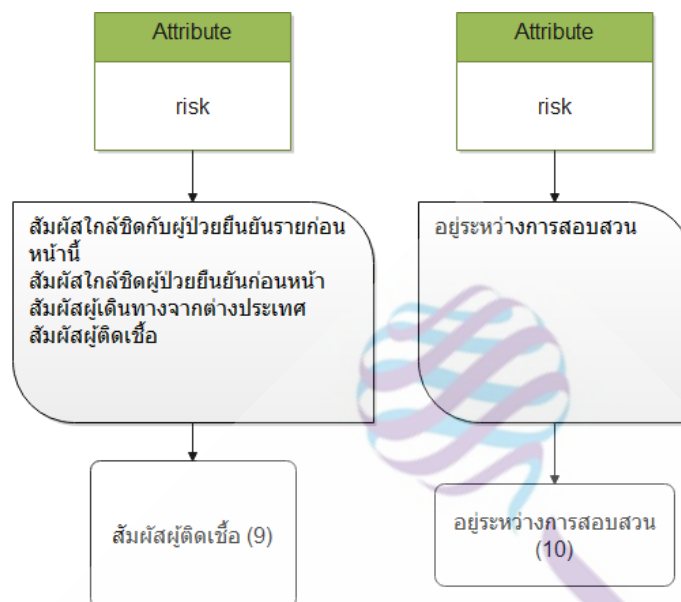
สามารถจัดกลุ่มข้อมูลของ risk (Department of disease control, 2020; กรมควบคุมโรค, 2563; เฉลย กิตติ et al., 2563; สำนักงานสาธารณสุขจังหวัดระยอง, 2564) ได้ดังภาพที่ 3.10 - 3.12



ภาพที่ 3.10 การจัดกลุ่ม risk (ความเสี่ยง)

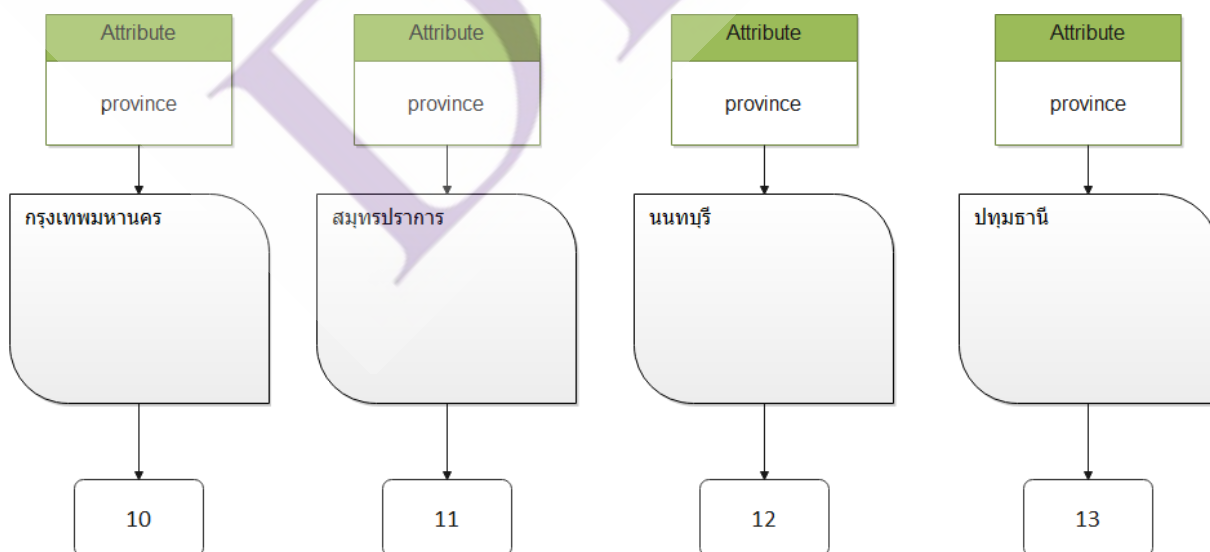


ภาพที่ 3.11 การจัดกลุ่ม risk (ความเสี่ยง) (ต่อ)



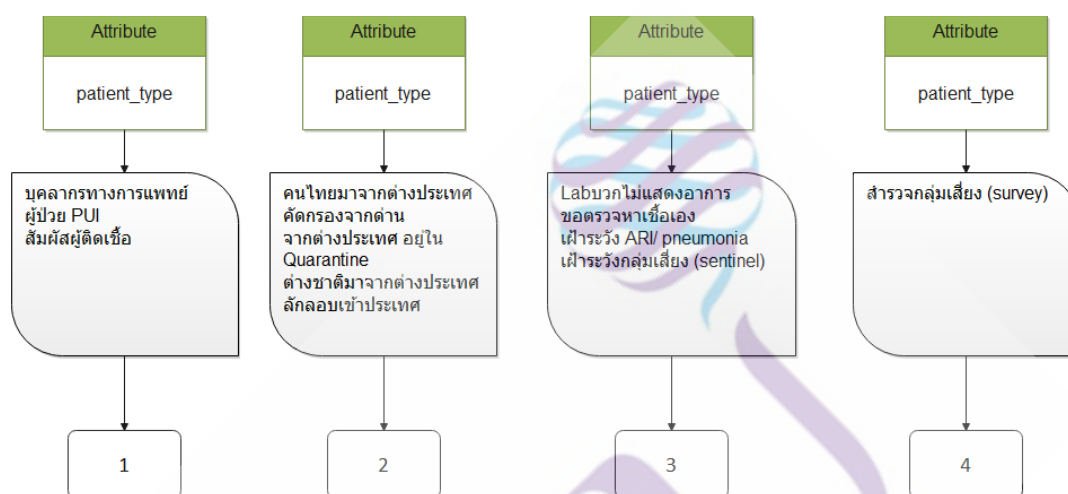
ภาพที่ 3.12 การจัดกลุ่ม risk (ความเสี่ยง) (ต่อ)

ถัดมาจะเป็น province ซึ่ง คลาส ใช้รหัสจังหวัดในประเทศไทย (“ISO 3166-2,” 2021) มาเป็นตัวกำกับกับกลุ่มข้อมูล เช่น จังหวัดกรุงเทพมหานคร มีรหัสจังหวัดคือ 10 และสมุทรสาคร มีรหัสจังหวัดคือ 11 ซึ่ง จะแปลงจากข้อมูล text เป็น ตัวเลข โดยตัวเลขคือรหัสจังหวัดดังภาพที่ 3.13



ภาพที่ 3.13 การจัดกลุ่ม province (จังหวัด)

แอททริบิวต์ ในการวิจัยคือ patient\_type (ประเภทผู้ป่วย) โดยจะมีอยู่ 4 คลาส ด้วยกันจากทั้งหมด 13 คลาส จึงต้องมีการจัดกลุ่มให้ถูกต้อง (กรมควบคุมโรค, 2563; โรงพยาบาลพญาไทศรีราชา, 2020) ดังภาพที่ 3.14



ภาพที่ 3.14 การจัดกลุ่ม patient\_type (ประเภทผู้ป่วย)

จากภาพที่ 3.14 จะเห็นได้ว่าการจัดกลุ่มในแต่ละคลาสให้น้อยลงโดยจะแทนเป็นคลาสที่เป็นตัวเลขได้คำตอบ คือ 1 2 3 4 โดยมีรายละเอียดตามตารางที่ 3.2

ตารางที่ 3.2 คำอธิบายข้อมูล patient\_type

ID	คลาส เดิม	คลาส ใหม่
1	-บุคลากรทางการแพทย์ -ผู้ป่วย PUI -สัมผัสผู้ติดเชื้อ	การเฝ้าระวังในกลุ่มผู้ป่วย หรือผู้ที่มีอาการเข้าได้กับนิยามผู้สงสัยติดเชื้อไวรัสโคโรนา 2019 ที่เข้าเกณฑ์สอบสวนโรค (patient under investigation : PUI)



ตารางที่ 3.2 คำอธิบายข้อมูล patient\_type (ต่อ)

ID	คลาส เดิม	คลาส ใหม่
2	- คนไทยมาจากต่างประเทศ - คัดกรองจากด่าน - จากต่างประเทศ อยู่ใน Quarantine - ต่างชาติมาจากต่างประเทศ - ลักลอบเข้าประเทศ	การตรวจคัดกรองในประชากรเสี่ยงตามจุดคัดกรองและด่านเข้าออกระหว่างประเทศ(screening)
3	- Labบวกไม่แสดงอาการ - ขอตรวจหาเชื้อเอง - เฝ้าระวัง ARI/ pneumonia - เฝ้าระวังกลุ่มเสี่ยง (sentinel)	การเฝ้าระวังในกลุ่มเป้าหมายเฉพาะหรือพื้นที่เฉพาะ (sentinel surveillance)
4	- สํารวจกลุ่มเสี่ยง (survey)	การเฝ้าระวังเหตุการณ์ในสถานที่เสี่ยง เก็บตัวอย่างส่งตรวจเมื่อเข้านิยาม PUI หรือเป็นกลุ่มก้อนของผู้ป่วยทางเดินหายใจ

จากการจัดกลุ่มคลาสใหม่ของ patient\_type ได้ใช้แนวทางการดำเนินงานคัดกรอง เฝ้าระวัง และสอบสวนโรคติดเชื้อไวรัสโคโรนา 2019” (กรมควบคุมโรค, 2563) เพื่อการตรวจสำหรับเฝ้าระวังในกลุ่มต่างๆ โดยมีรายละเอียดดังนี้

1. การเฝ้าระวังในกลุ่มผู้ป่วย หรือมีอาการเข้าได้กับนิยาม (Patient Under Investigation : PUI)

วัตถุประสงค์ เพื่อให้ทราบขนาดของปัญหาและสามารถตรวจจับการระบาดในประชากร โดยทำการเฝ้าระวังในผู้ป่วยที่เข้ารับการรักษาในโรงพยาบาล และกลุ่มก้อนของผู้ป่วยทางเดินหายใจในชุมชน โดยเก็บ ตัวอย่าง nasopharyngeal swab (NPS) ส่งตรวจยืนยันทุกรายที่มีอาการตามนิยามผู้สงสัยติดเชื้อไวรัสโคโรนา 2019 ที่เข้าเกณฑ์สอบสวนโรค (PUI)

## 2. การตรวจคัดกรองในประชากรเสี่ยงตามจุดคัดกรองและด่านเข้าออกระหว่างประเทศ (Screening)

วัตถุประสงค์ เพื่อค้นหาการติดเชื้อของผู้ที่เดินทางมาจากต่างประเทศ เพื่อให้การรักษา และควบคุม ไม่ให้แพร่เชื้อไปยังชุมชน โดยดำเนินการในผู้ที่เดินทางมาจากต่างประเทศเข้ามาในราชอาณาจักรไทย หรือผู้ที่เดินทางมาจากพื้นที่ที่มีการระบาด

## 3. การเฝ้าระวังในกลุ่มเป้าหมายเฉพาะหรือพื้นที่เฉพาะ (Sentinel Surveillance)

วัตถุประสงค์ เพื่อให้ทราบแนวโน้มของโรคติดเชื้อไวรัสโคโรนา 2019 ในกลุ่มเป้าหมายเฉพาะ และสามารถตรวจจับการระบาดได้ทันทั่วทั้งที่ เพื่อตรวจจับการระบาดของโรคในประชากรกลุ่มเสี่ยงได้อย่างทันเวลา และดำเนินการ ในสถานที่ที่คนอยู่รวมกันเป็นจำนวนมาก

## 4. การเฝ้าระวังเหตุการณ์ในสถานที่เสี่ยง เก็บตัวอย่างส่งตรวจเมื่อเข้าเกณฑ์ PUI และเป็นกลุ่มก้อน

วัตถุประสงค์ เพื่อตรวจจับการระบาดของโรคติดเชื้อไวรัสโคโรนา 2019 ในประชากรกลุ่มเสี่ยง ได้ทันเวลา และสามารถควบคุมการระบาดก่อนที่จะมีการแพร่กระจายในวงกว้าง ดำเนินการเฝ้าระวัง ผู้ที่มีอาการทางเดินหายใจในประชากรในสถานที่เสี่ยงที่อาจจะเกิดการระบาดหรือพบผู้เสียชีวิตได้ โดยรายงาน ผ่านระบบเฝ้าระวัง EBS (Event-Based Surveillance) และให้เก็บตัวอย่างส่งตรวจเมื่อมีอาการตามนิยาม PUI การเฝ้าระวังให้ดำเนินการทั่วประเทศ ตามกลุ่มเป้าหมาย

### 3.2.2 Data Cleaning (การทำความสะอาดข้อมูล)

เนื่องจากคะแนนหรือข้อมูลบางส่วนไม่สามารถใช้ในการคำนวณได้ เพราะข้อมูลที่ใส่จะต้องเป็นข้อมูลที่ไม่มีการสูญหายหรือ Missing Value เพราะข้อมูลดังกล่าวจะไม่สามารถจัดกลุ่มได้ จากชุดข้อมูลในงานวิจัยพบว่ายังมีข้อมูลในบางส่วนที่ผิดปกติคือข้อมูล gender nationality risk patient\_type เนื่องจากมีข้อมูลที่ไม่สามารถนำไปวิเคราะห์หรือหาผลได้ ผู้วิจัยจึงได้ทำการลบรายการเหล่านี้ออกดังตัวอย่างในภาพที่ 3.15

หญิง	47	40-49 ปี	Thai	พนักงานบริษัท/ เภสัชกร	สมเด็จพระรัตนโกสินทร์	4. สมเด็จพระรัตนโกสินทร์
ชาย	34	30-39 ปี	Thai	ค้าขาย/ธุรกิจส่วนตัว	สถานบันเทิง	2. สโมสรผู้สูงอายุ
หญิง	39	30-39 ปี	Thai	พนักงานในสถานบันเทิง	ศูนย์ยิมออกกำลังกายต่างประเศ	4. ศูนย์ยิมออกกำลังกายต่างประเศ
หญิง	47	40-49 ปี	Thai	พนักงานบริษัท/โรมาน	ศูนย์ยิมออกกำลังกายต่างประเศ	4. ศูนย์ยิมออกกำลังกายต่างประเศ
หญิง	40	40-49 ปี	Thai		null	4. ศูนย์ยิมออกกำลังกายต่างประเศ
ชาย	41	40-49 ปี	Thai	ข้าราชการ/พนักงานของรัฐ / ผู้รู้สาขาอื่น	ศูนย์ยิมออกกำลังกายต่างประเศ	4. ศูนย์ยิมออกกำลังกายต่างประเศ
ชาย	29	20-29 ปี	Thai	อื่นๆ	อาชีพพิเศษ เช่น ทำงานในสถานเฝ้าระวัง หรือทำงานในกิจการที่มีชาวต่างชาติ เย...	10. อื่นๆ
ชาย	53	50-59 ปี	Thai	รับจ้างทั่วไป / พนักงาน	สนามมวย	10. อื่นๆ
ชาย	39	30-39 ปี	Thai	ไม่ระบุ	สนามมวย	9. ขอตรวจหาเชื้อเอชไอวี
ชาย	21	20-29 ปี	Thai	นักเรียน/นักศึกษา	สนามมวย	10. อื่นๆ
ชาย	41	40-49 ปี	Thai			10. อื่นๆ
ชาย	30	30-39 ปี	Thai	รับจ้างทั่วไป /		10. อื่นๆ
ชาย	59	50-59 ปี	Thai	พนักงานบริษัท		10. อื่นๆ
ชาย	62	60-69 ปี	Thai	ค้าขาย/ธุรกิจ		10. อื่นๆ
ชาย	48	40-49 ปี	Thai	ข้าราชการ/		10. อื่นๆ
ชาย	37	30-39 ปี	Thai	ค้าขาย/ธุรกิจ		10. อื่นๆ
ชาย	39	30-39 ปี	Thai	ค้าขาย/ธุรกิจ		10. อื่นๆ
หญิง	33	30-39 ปี	Thai	พนักงานบริษัท		4. ศูนย์ยิมออกกำลังกายต่างประเศ
หญิง	27	20-29 ปี	Thai	พนักงานบริษัท		10. อื่นๆ
หญิง	30	30-39 ปี	Thai		null	10. อื่นๆ
ชาย	31	30-39 ปี	Thai		ศูนย์ยิมออกกำลังกายต่างประเศ	2. สโมสรผู้สูงอายุ
หญิง	33	30-39 ปี	Thai	รับจ้างทั่วไป / พนักงาน		4. ศูนย์ยิมออกกำลังกายต่างประเศ
หญิง	38	30-39 ปี	Thai	พ่อบ้าน/แม่บ้าน/ดูแลบ้าน	สถานบันเทิง	2. สโมสรผู้สูงอายุ
ชาย	42	40-49 ปี	Thai	ค้าขาย/ธุรกิจส่วนตัว	สนามมวย	10. อื่นๆ
ชาย	58	50-59 ปี	Thai	รับจ้างทั่วไป / พนักงาน	สนามมวย	10. อื่นๆ
ชาย	25	20-29 ปี	Thai		null	10. อื่นๆ
ชาย	46	40-49 ปี	Thai	รับจ้างทั่วไป / พนักงาน	สนามมวย	2. สโมสรผู้สูงอายุ
ชาย	35	30-39 ปี	Thai	อื่นๆ	สนามมวย	10. อื่นๆ
ชาย	73	>= 70 ปี	Thai	ข้าราชการ/พนักงานของรัฐ / ผู้รู้สาขาอื่น	สนามมวย	2. สโมสรผู้สูงอายุ
ชาย	51	50-59 ปี	Thai		null	10. อื่นๆ

ลบข้อมูลที่เป็นค่าว่าง (Missing and empty value)

ภาพที่ 3.15 ชุดข้อมูลก่อนทำ Data cleaning

จากการทำ Data preparation แล้วโดยจะมี คลาส ในการวิจัย 4 คลาส ซึ่งจะเป็นข้อมูลประเภท categorical ดังตารางที่ 3.7

ตารางที่ 3.3 คลาสในการจำแนก

Attribute	Description	Type	Values
patien_type	ประเภทผู้ป่วย	categorical	1= การเฝ้าระวังในกลุ่มผู้ป่วย (PUI) 2=การตรวจคัดกรองในประชากรเสี่ยงตามจุดคัดกรอง 3=การเฝ้าระวังในกลุ่มเป้าหมายเฉพาะหรือพื้นที่เฉพาะ 4=การเฝ้าระวังเหตุการณ์ในสถานที่เสี่ยง

สำหรับ แอททริบิวต์ ในการวิจัย เนื่องจากเป็นงานวิจัยในการแยกและจัดกลุ่มข้อมูล ข้อมูลที่ไม่สามารถจัดกลุ่มได้ เช่น txn\_date age\_number update\_date จะไม่สามารถนำมาวิจัยได้และยังมี job ข้อมูลมีค่าว่างหรือข้อมูลที่ไม่สามารถจัดกลุ่มได้ในจำนวนที่เยอะมาก โดยข้อมูลที่สามารถใช้ได้คิดเป็นอัตราส่วนเพียง 1.69 % เท่านั้น (ข้อมูล 27,294 ราย จากการรวบรวมทั้งหมด 1,608,923 ราย) ข้อมูลจึงไม่เพียงพอต่อการวิจัย จึงคัดแอททริบิวต์นี้ออก ดังตารางที่ 3.4

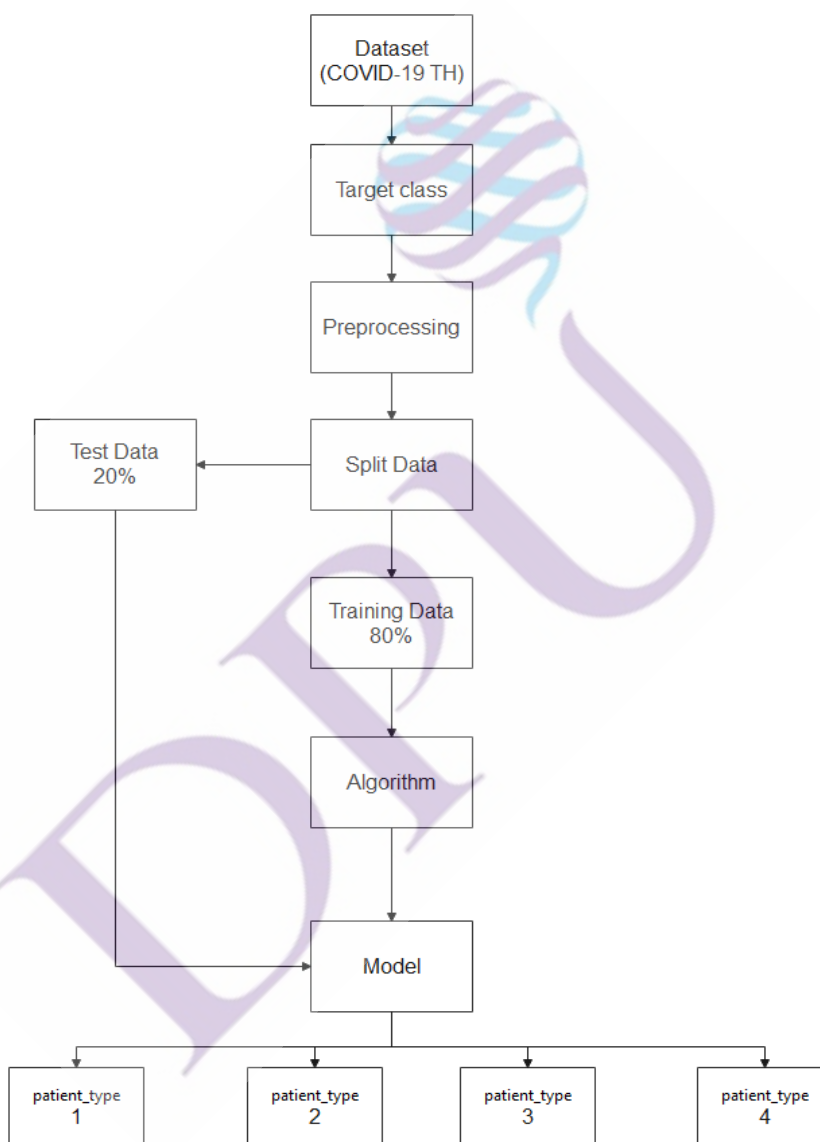
ตารางที่ 3.4 แอททริบิวต์ในการวิจัย

Attribute	Description	Type	Values	Status
gender	เพศ	categorical	1 : ชาย , 2 : หญิง	Active
age_range	ช่วงอายุ	categorical	1 : < 10 ปี 2 : 10 – 19 ปี 3 : 30 – 39 ปี 4 : 40 – 49 ปี 5 : 50 – 59 ปี 6 : 60 – 69 ปี 7 : >= 70 ปี	Active
nationality	สัญชาติ	categorical	1 : ไทย , 2 : ชาวต่างชาติ	Active
province	จังหวัด	categorical	รหัสจังหวัด 77 จังหวัด	Active

### 3.3 Modeling (การสร้างแบบจำลอง)

จากการตั้งค่า แอททริบิวต์ ที่เหมาะสมสำหรับข้อมูลจากกรมควบคุมโรค กระทรวงสาธารณสุข โดยจะเป็นข้อมูลผู้ป่วยโควิด-19 ระลอกที่ 1 ถึงระลอกที่ 3 และทำงานร่วมกับแบบจำลองด้วยวิธีการหาผลลัพธ์ที่ได้ค่าความถูกต้องมากที่สุด โดยการศึกษางานวิจัยที่เกี่ยวข้องตามตารางที่ 2.4 ที่สรุปมาได้ นั้นพบว่าในแต่ละงานวิจัยมี Neural network Random Forest และ Naive Bayes เป็นเทคนิคในการสร้างแบบจำลองที่ให้ค่าความถูกต้องมากที่สุด

งานวิจัยนี้จึงใช้ 3 เทคนิคดังกล่าวมาทำการทดลอง โดยได้ใช้วิธี Split Test ในการแบ่งข้อมูล จำนวน 1,608,923 รายการ สำหรับเป็นชุดการเรียนรู้ 80% และชุดทดสอบ 20% แล้วนำเข้าโปรแกรม Orange Data Mining เพื่อหาแบบจำลองที่ดีที่สุด โดยมีแผนผังการทำงานดังภาพที่ 3.16



ภาพที่ 3.16 แบบจำลองการวิเคราะห์จำแนกผู้ป่วยโควิด-19

จากภาพที่ 3.16 การทำงานจะกำหนด Target คลาส จากการเลือก แอททริบิวต์ ที่จะให้เป็น คลาส ในการวิจัยที่จะหาคำตอบจากการทำนายซึ่งก็คือ patient\_type จากนั้นทำการ Preprocessing และ

ทำการ Split Test โดยการแยกข้อมูลออกเป็น 80% และ 20% โดย คลาส ของข้อมูลจะแบ่งเป็น 4 คลาส คือ

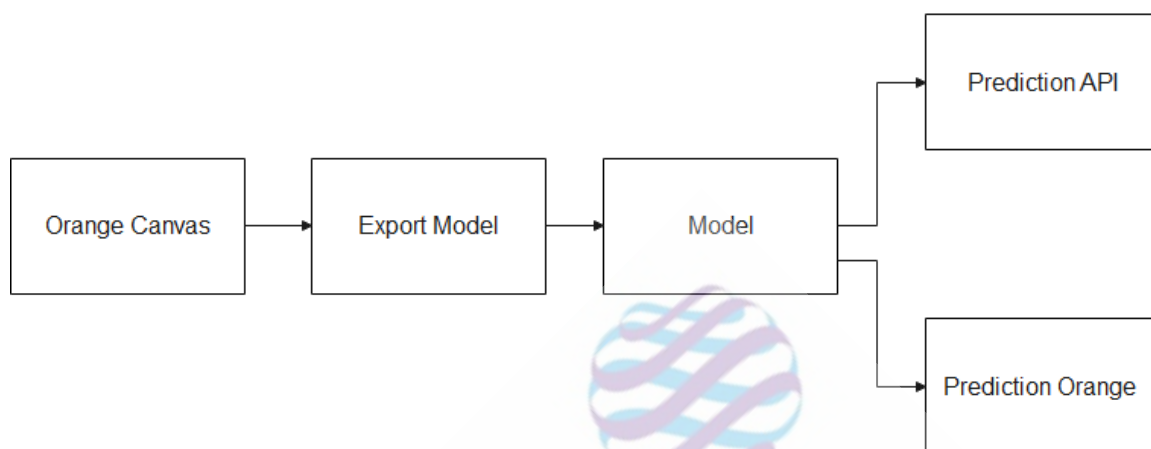
1. การเฝ้าระวังในกลุ่มผู้ป่วย (PUI)
2. การตรวจคัดกรองในประชากรเสี่ยงตามจุดคัดกรอง
3. การเฝ้าระวังในกลุ่มเป้าหมายเฉพาะหรือพื้นที่เฉพาะ
4. การเฝ้าระวังเหตุการณ์ในสถานที่เสี่ยง

### 3.4 Evaluation (การวัดประสิทธิภาพแบบจำลอง)

หลังจากที่แบ่งข้อมูลออกเป็น 2 ส่วนและใช้แบบจำลองที่สร้างได้จากข้อมูลจำนวน 80% มาทำการจำแนกผู้ป่วยโควิด-19 กับข้อมูลผู้ป่วยที่อยู่ใน 20% ที่เหลือ และใช้วัดค่าความถูกต้องโดยใช้ Confusion matrix เพื่อดูว่าเทคนิควิธี Neural network Random Forest และ Naive Bayes เทคนิคที่สร้างแบบจำลองไหนให้ค่าผลทำนายดีที่สุดจึงจะนำแบบจำลองนั้นไปใช้งานต่อไป

### 3.5 Deployment (การนำแบบจำลองไปใช้งาน)

จากแบบจำลองที่สร้างโดย Naive Bayes , Random Forest และ Neural network ที่สร้างขึ้นมาได้จากโปรแกรม Orange Data Mining นั้นสามารถนำไปทำการจำแนกผู้ป่วยโควิด-19 หลังจากนั้นจะต้องพัฒนา Prediction API ของแบบจำลองที่สร้างขึ้นมาเพื่อเป็นเว็บเซอร์วิส (Web service) ในการรองรับการส่งข้อมูลจากเว็บแอปพลิเคชันในการจำแนกผู้ป่วยโควิด-19 ดังภาพที่ 3.17



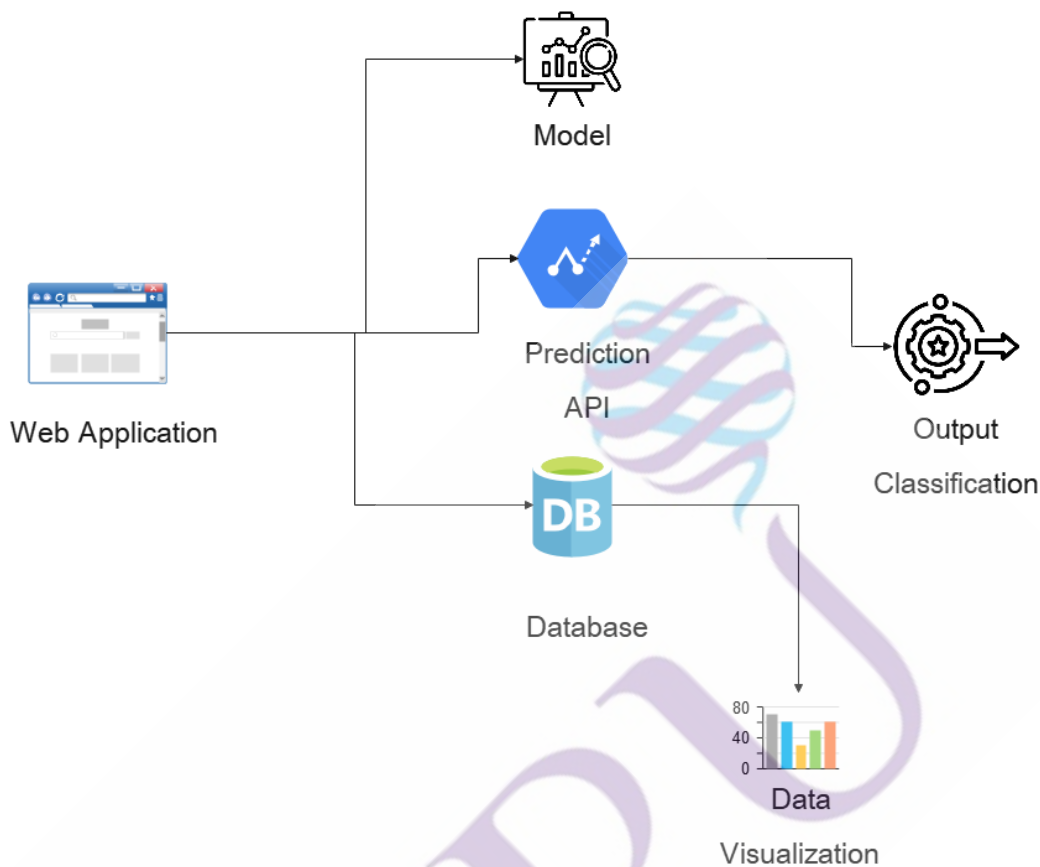
ภาพที่ 3.17 การนำแบบจำลองไปใช้งาน

จากภาพที่ 3.17 เมื่อทำการทดลองผ่าน โปรแกรม Orange Data Mining และได้ผลลัพธ์แล้ว สามารถนำแบบจำลองไปใช้งานได้ 2 วิธี คือ

1. Prediction API นำแบบจำลองไปใช้สำหรับพัฒนาเว็บเซอร์วิสสำหรับการทำนาย
2. Prediction Orange การนำแบบจำลองที่ได้จากการทดลองในส่วนวัดผลไปใช้ในการทำนายได้บน โปรแกรม Orange Data Mining โดยไม่ต้องผ่านกระบวนการ Preprocess

ในงานวิจัยนี้ได้ทำการทดลองเพื่อหาแบบจำลองที่เหมาะสมและการนำแบบจำลองไปใช้กับแอปพลิเคชันในการจำแนกผู้ป่วยโควิด-19 ซึ่งจะต้องให้ความสำคัญในส่วน Prediction API เป็นหลัก ดังรูปที่ 3.18





ภาพที่ 3.18 Prediction API ในการจำแนกผู้ป่วยโควิด-19

จากภาพที่ 3.18 จะเป็นขั้นตอนการทำงานของระบบจำแนกผู้ป่วยโควิด – 19 ซึ่งจะออกแบบในลักษณะ Microservice โดยจะเชื่อมต่อแบบ API โดยมีรายละเอียดดังนี้

**3.5.1 FastAPI Framework (Backend API)** ทำหน้าที่เป็นตัวขับเคลื่อนระบบซึ่งจะทำงานกับแบบจำลองที่ได้จากการทดลองเพื่อจำแนกผลและเชื่อมต่อกับฐานข้อมูล โดย FastAPI Framework จะเป็นภาษา Python (*FastAPI*, n.d.) ซึ่งมีคุณสมบัติดังนี้

1. มีความรวดเร็ว ประสิทธิภาพเทียบเท่ากับ NodeJS และ Go
2. สร้างง่าย เพิ่มความเร็วในการพัฒนา
3. ลดข้อผิดพลาดที่เกิดจากมนุษย์ (นักพัฒนา)
4. เรียนรู้ง่าย

**3.5.2 VueJS Framework (Frontend)** ทำหน้าที่เป็นส่วน User Interface ซึ่งเป็น JavaScript Framework โดยใช้ NodeJS ในการทำงาน (*Vue.js - The Progressive JavaScript Framework | Vue.js, n.d.*) โดยมีคุณสมบัติดังนี้

1. มี Community ที่เติบโตไว้มาก
2. Vue.js สามารถ render หรือแสดงผลได้ไว้มาก เพราะ Vue.js มีขนาดเพียง 30 KB
3. รวมข้อดีของทั้ง React และ Angular มารวมไว้ใน Vue

**3.5.3 Highcharts** ทำหน้าที่แสดงผลการสร้างมโนภาพ (Data Visualization) โดยมี JavaScript Library สำหรับไว้สร้างกราฟบนหน้าเว็บ ซึ่ง Highcharts สามารถสร้างกราฟได้หลายรูปแบบ เช่น กราฟแท่ง กราฟเส้น กราฟวงกลม แผนภูมิ เป็นต้น อีกทั้งยังสามารถแสดงผลได้ทุกเบราว์เซอร์ (Highcharts, n.d.)

**3.5.4 PostgreSQL** ทำหน้าที่ในการเก็บข้อมูลต่างๆ เช่น ข้อมูลการจำแนกผู้ป่วย ข้อมูลผู้ป่วย ข้อมูลผู้ใช้งาน ซึ่ง PostgreSQL เป็นระบบการจัดการฐานข้อมูลเชิงวัตถุ-สัมพันธ์ (object-relational) แบบ ORDBMS โดยสามารถใช้รูปแบบคำสั่งของภาษา SQL ได้เกือบทั้งหมด นอกจากนี้ยังเป็นระบบฐานข้อมูลที่ทันสมัยที่สุดของ Opensource ที่สามารถนำไปใช้งานได้โดยไม่มีค่าใช้จ่ายใด (MDSOFT.CO.TH, n.d.)

จากการศึกษาเทคนิคและวิธีการดำเนินงานแล้วโดยจะใช้ Random Forest, Neural network และ Naive Bayes เพื่อหาแบบจำลองที่เหมาะสมที่สุดเพียง 1 แบบจำลองไปใช้กับการพัฒนา Prediction API จากนั้นทำการพัฒนาเว็บแอปพลิเคชันเพื่อจำแนกผู้ป่วยโควิด-19 โดยหลักการของเว็บแอปพลิเคชันจะใช้ Prediction API ในการขับเคลื่อน โดยผลลัพธ์ของการทดลองและการพัฒนาจะอยู่ในบทถัดไป

## บทที่ 4

### ผลการทดลองและอภิปรายผล

การทดสอบประสิทธิภาพรูปแบบจำลองสำหรับการจำแนกผู้ป่วยโควิด-19 จากชุดข้อมูลผู้ป่วยโควิด-19 จากกรมควบคุมโรค จะทดสอบด้วยค่าความแม่นยำในการจำแนก (Accuracy) ค่าความเที่ยง (Precision) ค่าความไวหรือค่าระลึก (Recall) และระยะเวลาที่ใช้ในการฝึกสอนแบบจำลอง โดยเปรียบเทียบการจำแนกของแบบจำลองที่ถูกฝึกสอนด้วยคุณลักษณะในรูปแบบต่างๆ ตามที่ผู้วิจัยได้ออกแบบไว้

#### 4.1 โครงสร้างข้อมูล

ข้อมูลเพื่อจำแนกประเภทผู้ป่วยโควิด-19 จาก 1 มกราคม 2563 ถึง วันที่ 1 ตุลาคม 2564 เป็นจำนวน 1,608,923 ราย หลังจากผ่านกระบวนการ Data preparation แล้วจะอยู่ที่จำนวน 913,152 ราย มีแอททริบิวต์ 5 อย่าง และคลาส ที่ใช้ในการวิจัย 4 คลาส ดังภาพที่ 4.1

Data Set Name	use_dump_db_all
Data Set Size	Rows: 913152 Columns: 6
Features	Categorical: 5 Numeric: -
Targets	Categorical outcome with 4 values

ภาพที่ 4.1 โครงสร้างข้อมูล

## 4.2 การใช้โปรแกรม Orange Data Mining เพื่อเปรียบเทียบแบบจำลองการเรียนรู้ของเครื่อง

หาผลทดลองโดยใช้ Widget ของโปรแกรม Orange Data Mining ซึ่งจะมี 5 ขั้นตอนหลักๆด้วยกัน เช่น

ขั้นตอนที่ 1 : จัดเตรียมข้อมูล (Load Data)

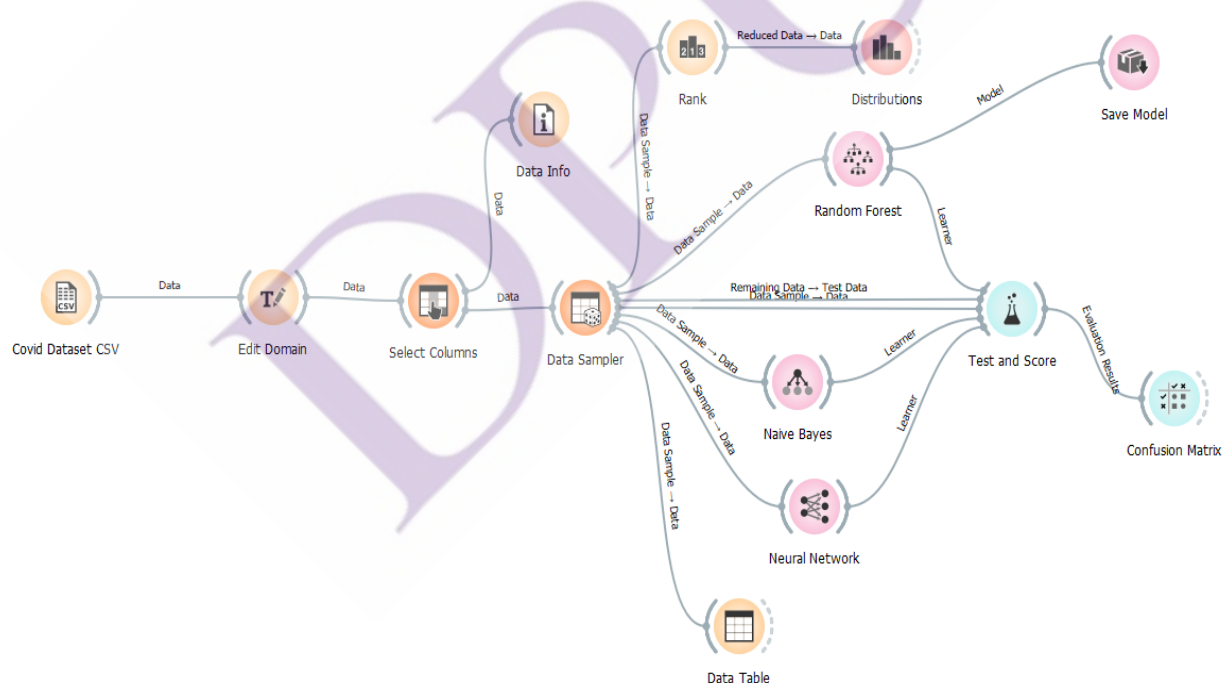
ขั้นตอนที่ 2 : การจัดการกับข้อมูลที่สูญหาย (Data Missing Values)

ขั้นตอนที่ 3 : เลือกและจัดการข้อมูล (Select Data)

ขั้นตอนที่ 4 : การใช้แบบจำลองและการวัดประสิทธิภาพ (Model and Evaluate)

ขั้นตอนที่ 5 : การบันทึกแบบจำลอง (Save Model)

ซึ่งขั้นตอนทั้งหมดจะมีการทำงานดังภาพที่ 4.2



ภาพที่ 4.2 การใช้โปรแกรม Orange Data Mining เพื่อเปรียบเทียบแบบจำลองการเรียนรู้ของเครื่อง

ใช้ Data Sampler Widget เป็นวิธีแบ่งข้อมูลแบบ Split Test จำนวน 80% เป็นข้อมูลสำหรับสร้างแบบจำลอง (Data Sample หรือ Training Data) และข้อมูลจำนวน 20% ใช้เป็นข้อมูลสำหรับทดสอบ (Remaining Data หรือ Test Data) ดังภาพที่ 4.3



ภาพที่ 4.3 การใช้โปรแกรม Orange Data Mining ในการ Split Test

งานวิจัยนี้มีข้อมูลจำนวน 1,608,923 เรคคอร์ด หลังจากผ่านกระบวนการ Data Preparation แล้ว ข้อมูลจะอยู่ที่จำนวน 913,152 เรคคอร์ด ประกอบด้วย คุณลักษณะดังนี้ 1) คำนวณคุณลักษณะของข้อมูลจำนวน 5 แอททริบิวต์ คือ gender age\_rang nationality province risk และ 2) คลาส ที่ใช้ในการวิจัยประกอบไปด้วย 4 คลาส มีจำนวนข้อมูลในแต่ละคลาส ดังนี้ คลาส ที่ 1 มีจำนวน 587,079 เรคคอร์ด คลาส ที่ 2 มีจำนวน 2,755 เรคคอร์ด คลาส ที่ 3 มีจำนวน 52,548 เรคคอร์ด และคลา ส ที่ 4 มีจำนวน 270,770 ดังตารางที่ 4.1

ตารางที่ 4.1 อัตราส่วนของข้อมูลในแต่ละคลาส

คลาส	จำนวน	อัตราส่วน
1	587,049	64.29%
2	2,755	0.30%
3	52,547	5.75%
4	270,770	29.65%

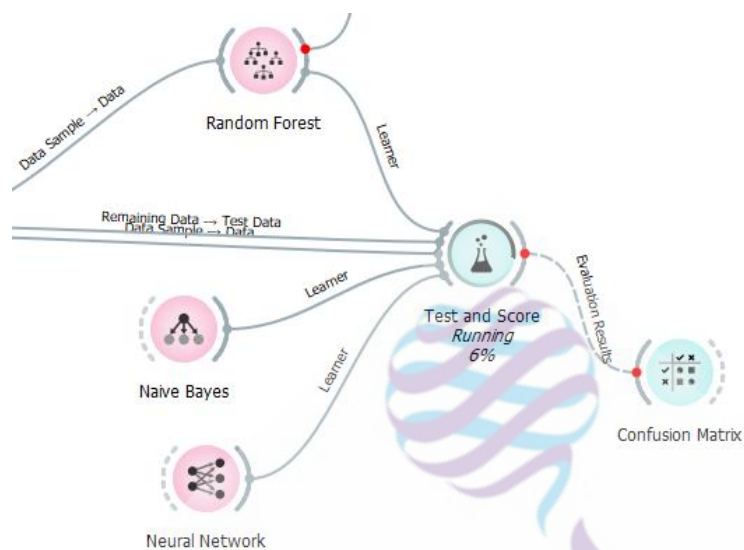
#### 4.2.1 Information Gain ของ แอททริบิวต์ ในการวิจัย

การคัดเลือก แอททริบิวต์ โดยใช้เทคนิคคำนวณค่าน้ำหนักของ แอททริบิวต์ ต่างๆ โดย แอททริบิวต์ ที่มีค่าน้ำหนักมากที่สุดคือ risk ค่าน้ำหนักที่ 0.816 และ แอททริบิวต์ ที่มีค่าน้ำหนักน้อยที่สุดคือ gender ค่าน้ำหนักที่ 0.003 ดังตารางที่ 4.1

ตารางที่ 4.2 Information Gain ของ แอททริบิวต์ในการวิจัย

ตัวแปรที่	Attribute	Information Gain
1	gender	0.003
2	age_rang	0.011
3	nationality	0.019
4	province	0.213
5	risk	0.816

เมื่อได้ค่า Information Gain ในแต่ละ แอททริบิวต์ แล้วนำ แบบจำลอง Widget ของ Neural network Random Forest และ Naive Bayes มาคำนวณค่าความถูกต้องและความแม่นยำโดยใช้มาตรวัด Confusion Matrix Widget ดังภาพที่ 4.4



ภาพที่ 4.4 การใช้ Confusion Matrix Widget ในการวัดค่า

จากการใช้ Confusion Matrix Widget ในการวัดทั้ง 3 เทคนิคแล้วโดยจะเป็นจำนวนผลการทำนาย True Positive และ False Positive ในแต่ละคลาส ที่วิจัย สามารถดูรายละเอียดได้ในตารางที่ 4.3-4.5

#### 4.2.2 ผลการจำแนกข้อมูลด้วย Naive Bayes

ตารางที่ 4.3 ตัววัดประสิทธิภาพของอัลกอริทึม Naive Bayes

Performance Classification (Naive Bayes)					
	true(1)	true(2)	true(3)	true(4)	Precision
pred.(1)	455784	14	7203	6831	97%
pred.(2)	28	2113	3	72	95.35%
pred.(3)	4147	7	24724	13142	58.83%
pred.(4)	21009	36	5060	190349	87.93%
Recall	94.76%	97.14%	66.83%	90.47%	



จากตารางที่ 4.14 พบว่าคลาส 1 True Positive เท่ากับ 455,784 ค่า False Positive เท่ากับ 14,048 ค่า คิดเป็น Precision เท่ากับ 97% และ Recall เท่ากับ 94.76% คลาส 2 True Positive เท่ากับ 2,113 ค่า False Positive เท่ากับ 103 ค่า คิดเป็น Precision เท่ากับ 95.35% และ Recall เท่ากับ 97.14% คลาส 3 True Positive เท่ากับ 24,724 ค่า False Positive เท่ากับ 17,296 ค่า คิดเป็น Precision เท่ากับ 58.83% และ Recall เท่ากับ 66.83% คลาส 4 True Positive เท่ากับ 190,349 ค่า False Positive เท่ากับ 26,105 ค่า คิดเป็น Precision เท่ากับ 87.93% และ Recall เท่ากับ 90.47% สรุปเป็นค่า Accuracy ได้ 92.1%

#### 4.2.3 ผลการจำแนกข้อมูลด้วย Neural Network

ตารางที่ 4.4 ตัววัดประสิทธิภาพของอัลกอริทึม Neural Network

Performance Classification (Neural Network)					
	true(1)	true(2)	true(3)	true(4)	Precision
pred.(1)	455433	19	4544	9836	96.93%
pred.(2)	12	2198	0	6	99.18%
pred.(3)	3408	14	24653	13945	58.66%
pred.(4)	19387	39	1932	195096	90.13%
Recall	95.23%	96.82%	79.19%	89.13%	

จากตารางที่ 6 พบว่าคลาส 1 True Positive เท่ากับ 455,433 ค่า False Positive เท่ากับ 14,399 ค่า คิดเป็น Precision เท่ากับ 96.93% และ Recall เท่ากับ 95.23% คลาส 2 True Positive เท่ากับ 2,198 ค่า False Positive เท่ากับ 18 ค่า คิดเป็น Precision เท่ากับ 99.18% และ Recall เท่ากับ 96.82% คลาส 3 True Positive เท่ากับ 24,653 ค่า False Positive เท่ากับ 17,367 ค่า คิดเป็น Precision เท่ากับ 58.66% และ Recall เท่ากับ 79.19% คลาส 4 True Positive เท่ากับ 195,096 ค่า False Positive เท่ากับ 21,449 ค่า คิดเป็น Precision เท่ากับ 90.13% และ Recall เท่ากับ 89.13% สรุปเป็นค่า Accuracy ได้ 92.7%

#### 4.2.4 ผลการจำแนกข้อมูลด้วย Random Forest

ตารางที่ 4.5 ตัววัดประสิทธิภาพของอัลกอริทึม Random Forest

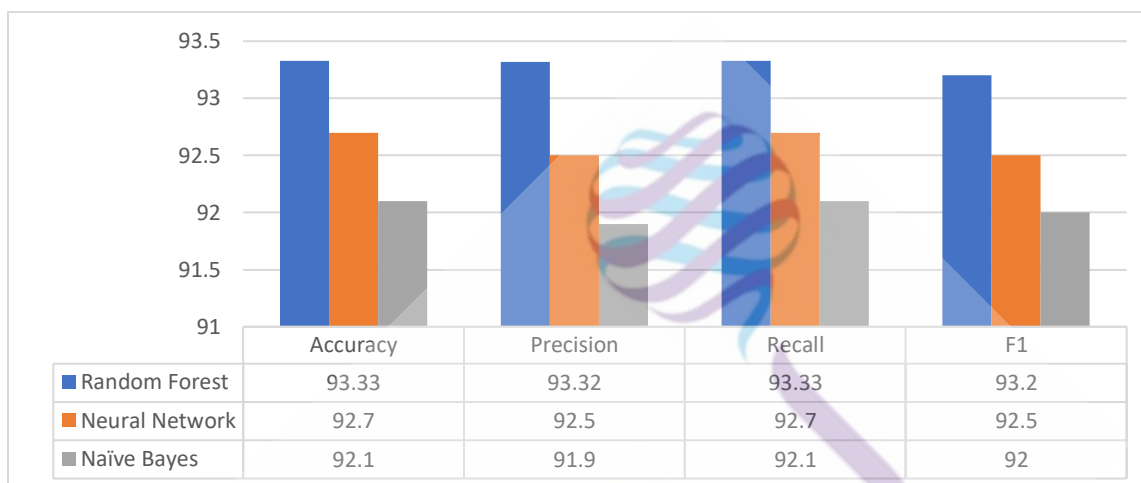
Performance Classification (Random Forest)					
	true(1)	true(2)	true(3)	true(4)	Precision
pred.(1)	459329	19	5254	5230	97.76%
pred.(2)	6	2176	7	27	98.19%
pred.(3)	4091	16	28188	9725	67.08%
pred.(4)	20976	29	3524	191925	88.66%
Recall	94.82%	97.14%	76.23%	92.75%	

จากตารางที่ 4.16 พบว่าคลาส 1 True Positive เท่ากับ 459,329 ค่า False Positive เท่ากับ 10,053 ค่า คิดเป็น Precision เท่ากับ 97.76% และ Recall เท่ากับ 94.82% คลาส 2 True Positive เท่ากับ 2,176 ค่า False Positive เท่ากับ 40 ค่า คิดเป็น Precision เท่ากับ 98.19% และ Recall เท่ากับ 94.14% คลาส 3 True Positive เท่ากับ 28,188 ค่า False Positive เท่ากับ 13,832 ค่า คิดเป็น Precision เท่ากับ 67.08% และ Recall เท่ากับ 76.23% คลาส 4 True Positive เท่ากับ 191,925 ค่า False Positive เท่ากับ 24,529 ค่า คิดเป็น Precision เท่ากับ 88.66 % และ Recall เท่ากับ 92.75% สรุปเป็นค่า Accuracy ได้ 93.33%

#### 4.3 สรุปผลการเปรียบเทียบค่าความถูกต้องและความแม่นยำในแต่ละเทคนิค

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อทดลองสร้างแบบจำลองการทำนาย โดยใช้ข้อมูลผู้ป่วยโควิด-19 จากกรมควบคุมโรค ร่วมกับเทคนิคการเรียนรู้ของเครื่อง ได้แก่ Random Forest , Neural network และ Naive Bayes เพื่อใช้ในการจำแนกผู้ป่วยโควิด-19 ในการเปรียบเทียบประสิทธิภาพความถูกต้องและความแม่นยำในการทำนาย โดย Random Forest มีค่า Accuracy 93.33% ส่วน Neural Network มีค่า Accuracy 92.7% และ Naive Bayes มีค่า Accuracy ที่ 92.1%

ผู้วิจัยจึงใช้เทคนิค Random Forest เนื่องจากให้ค่าความถูกต้องและความแม่นยำมากที่สุดมาสร้างแบบจำลองในการทำนายเพื่อจำแนกผู้ป่วยโควิด-19 ต่อไปดังภาพที่ 4.5



ภาพที่ 4.5 การเปรียบเทียบความแม่นยำในแต่ละอัลกอริทึม

จากข้อมูลตัววัดประสิทธิภาพของอัลกอริทึมตามตารางที่ 4.3 – 4.5 ได้ใช้ข้อมูล Training Data (80 %) จำนวน 730,522 เรคคอร์ด และ Test Data (20%) จำนวน 182,630 เรคคอร์ด โดยจะวัดผลจากชุดข้อมูล Training Data ที่ทำการสอนให้กับคอมพิวเตอร์ และใช้ Test Data มาทดสอบความแม่นยำ ความถูกต้องของแบบจำลองที่เรียนรู้เรียบร้อยแล้ว

#### 4.3.1 สรุปผลการทำงานของ Random Forest

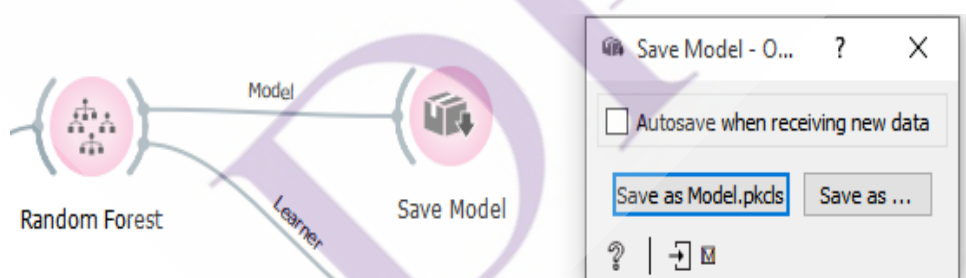
แบบจำลอง Random Forest Classification ให้ผลการทำนายที่มีความแม่นยำมากที่สุดถึง 93.33% แต่ก็ขึ้นอยู่กับข้อมูลความเสี่ยงและจังหวัดในการวัดผลด้วย เนื่องจากในแต่ละจังหวัดยังมีความเสี่ยงที่ใกล้เคียงกัน โดยเฉพาะจังหวัดกรุงเทพมหานครและ ปริมณฑล แต่ในต่างจังหวัดก็จะมี ความเสี่ยงที่แตกต่างกันออกไป ซึ่งจะเห็นได้ว่าการใช้งาน Random Forest Classification สามารถใช้งานได้หลากหลายรูปแบบและได้มีการตั้งค่าพารามิเตอร์ (Parameter) ซึ่งเป็นค่าเริ่มต้นดังตารางที่ 4.6

ตารางที่ 4.6 พารามิเตอร์ของ Random Forest

Parameters	Description	Value
min_samples_split	จำนวนขั้นต่ำที่จำเป็นในการแยกโหนด	5
n_estimators	จำนวนต้นไม้	100
n_jobs	จำนวนงาน	1
criterion	ฟังก์ชันวัดคุณภาพในการแยก	gini

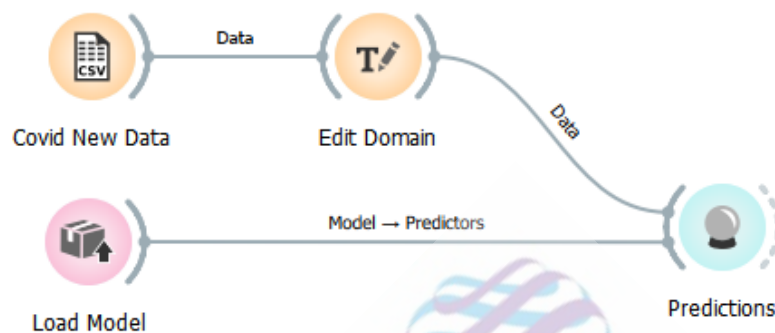
#### 4.4 การนำแบบจำลองไปใช้งาน

แบบจำลองที่สร้างโดย Random Forest นั้นสามารถนำไปทำการจำแนกผู้ป่วยโควิด-19 ได้ และนำไปใช้กับชุดข้อมูลใหม่เพื่อการทำนายการจำแนกผู้ป่วยโควิด-19 ทั้งนี้แบบจำลองสามารถทำการเรียนรู้จากชุดข้อมูลใหม่ได้ทันทีโดยข้อมูลจะต้องอยู่ในโครงสร้างเดิมจากนั้นทำการบันทึกและนำไปใช้ใหม่โดยจะบันทึกแบบจำลองในรูปแบบไฟล์ Model.pkcls ดังภาพที่ 4.6



ภาพที่ 4.6 การบันทึกแบบจำลอง Random Forest

เมื่อได้แบบจำลองแล้วสามารถใช้ Load Model Widget ในการโหลดไฟล์ Model.pkcls ซึ่งเป็นแบบจำลองที่ได้จากการทำการทดลองเพื่อนำมาใช้ทำนายกับชุดข้อมูลใหม่ได้และสามารถใช้งานโปรแกรม Orange Data Mining ได้ทันทีโดยไม่ต้องผ่านการ Preprocess ดังภาพที่ 4.7



ภาพที่ 4.7 การนำแบบจำลองไปใช้งานกับข้อมูลใหม่ผ่านโปรแกรม Orange Data Mining

ในส่วนการพัฒนา Prediction API นั้นสามารถใช้ไฟล์ Model.pkcls ที่ได้จากการทดลองในโปรแกรม Orange Data Mining ไปใช้กับการพัฒนาระบบการจำแนกผู้ป่วยโควิด-19 ด้วยภาษา Python ได้ดังภาพที่ 4.8

```
@router.post("/startpredict")
async def predict(item: Predict):
    # ทำการ Load โมเดลที่เก็บไว้
    model = pickle.load(open("MODEL.pkcls", "rb"))
    gender = int(item.gender)
    age_range = int(item.age_range)
    nationality_type = int(item.nationality_type)
    risk = int(item.risk)
    province = int(item.province)
    # ข้อมูลผู้ป่วยใหม่
    new_data = [[gender, age_range, nationality_type, risk, province]]
    # ทำนาย
    result = model.predict(new_data)
    rsmath = patientType[int(result[0]) - 1]['ptt_name']
    return {'result': str(rsmath), 'type': int(result[0])}
```

ภาพที่ 4.8 การนำแบบจำลองไปใช้งานกับข้อมูลใหม่ด้วยภาษา Python

จากภาพที่ 4.7 – 4.8 จะมีอินพุต (Input) ของข้อมูลที่ใช้สำหรับการจำแนก คือ gender (เพศ) age\_range (ช่วงอายุ) nationality\_type (สัญชาติ) risk (ความเสี่ยง) และ province (จังหวัด)

## 4.5 API สำหรับการทำนาย

จากภาพที่ 4.8 สามารถนำแบบจำลองมาพัฒนา API สำหรับการทำนายได้โดยใช้ Fast API Framework ในการพัฒนา เพื่อเป็นเว็บเซอร์วิสสำหรับการทำงานการจำแนกผู้ป่วยโควิด-19 โดยใช้ภาษา Python ซึ่ง Framework จะแสดงดังภาพที่ 4.9



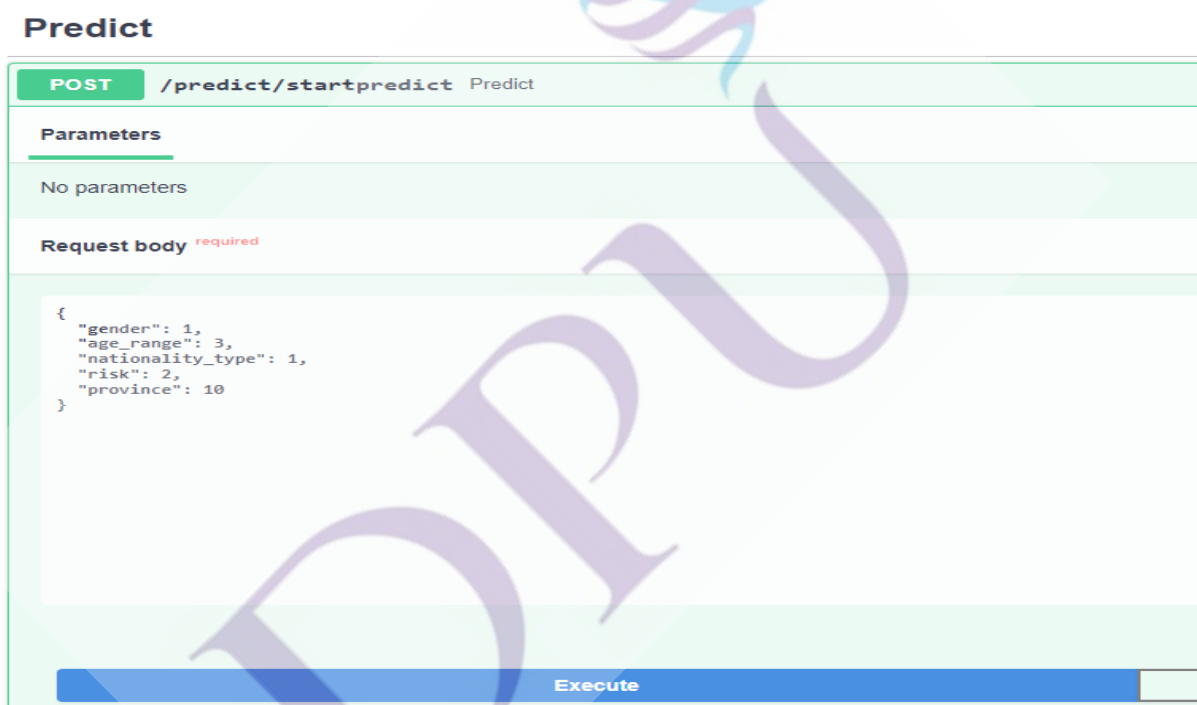
ภาพที่ 4.9 Fast API Framework (Web Service)

จากภาพที่ 4.9 เป็น document สำหรับอธิบายการทำงานของ API ซึ่งมาในรูปแบบ Swagger UI (สามารถดูเพิ่มเติมได้ใน ภาคผนวก ก.) ในแต่ละส่วน ซึ่งจะประกอบไป 5 ส่วนด้วยกัน เช่น

1. User : สำหรับการทำงานของผู้ใช้งาน โดยจะประกอบไปด้วย CRUD และ Login
2. Patient : สำหรับการทำงานข้อมูลผู้ป่วยโดยจะมี CRUD เท่านั้น
3. Predict : สำหรับการส่งข้อมูลแบบ POST เพื่อนำข้อมูลไปทำนายและส่งผลตอบกลับ

4. Master Data : โดยจะประกอบไปด้วยข้อมูลตั้งต้นที่ใช้ประกอบข้อมูลแบบฟอร์มในระบบ เช่น ช่วงอายุ , ประเภทผู้ป่วย และความเสี่ยง
5. Dashboard : สถิติข้อมูลต่างๆ เช่น จำนวนผู้ป่วย และตรวจสอบค่าความถูกต้องต่างๆ

ในส่วนของการทำนายจะส่งข้อมูลในรูปแบบ JSON ประเภท Method POST ซึ่งประกอบไปด้วย gender , age\_range , nationality\_type , risk , province ดังภาพที่ 4.10 - 4.11



ภาพที่ 4.10 การส่งข้อมูลไปทำนายผ่าน Swagger UI

```
{
  "gender": 0,
  "age_range": 0,
  "nationality_type": 0,
  "risk": 0,
  "province": 0
}
```

ภาพที่ 4.11 รูปแบบการส่งข้อมูลไปทำนาย



จากภาพที่ 4.10 - 4.11 จะได้ผลลัพธ์ของการทำนายซึ่งจะเรียนรู้จากแบบจำลองที่สร้างไว้มาหาผล โดยจะส่งค่ากลับมา 2 ค่า คือ result หมายถึง ผลลัพธ์ในรูปแบบ text ของการทำนาย และ type หมายถึง ผลลัพธ์ที่แสดงเป็นตัวเลขของการทำนาย โดยจะแสดงรายละเอียดของ API ดังภาพที่ 4.12

```

Response body
{
  "result": "การเฝ้าระวังเหตุการณ์ในสถานที่เสี่ยง เก็บตัวอย่างส่งตรวจเมื่อเข้านิยาม PUI หรือเป็นกลุ่มก้อนของผู้ป่วยทางเดินหายใจ",
  "type": 4
}

```

ภาพที่ 4.12 ผลของการทำนาย

#### 4.6 ส่วนติดต่อผู้ใช้งาน

ส่วนติดต่อผู้ใช้งานจะถูกพัฒนาด้วย Vue JS และ Node JS ซึ่งจะมีการใช้ API ที่ได้จากข้อที่ 4.5 มาเชื่อมต่อการทำงานของระบบ (สามารถดูเพิ่มเติมได้ใน ภาคผนวก ข.) โดยแบบฟอร์มการทำนายจะแสดงดังภาพที่ 4.13

ภาพที่ 4.13 แบบฟอร์มวิเคราะห์ข้อมูลผู้ป่วย

จากนั้นเมื่อทำนายผลเรียบร้อยแล้ว ระบบจะนำพาไปยังแบบฟอร์มการบันทึกข้อมูลผู้ป่วย โดยจะมีการขอความยินยอมการใช้ข้อมูลกับเจ้าของข้อมูลส่วนบุคคลตาม PDPA ผ่านแบนเนอร์ (banner) ซึ่งเป็นวิธีการที่สะดวกและได้รับความนิยมน เนื่องจากเป็นวิธีที่ไม่ซับซ้อนและถือเป็นการแจ้งให้ผู้ใช้งานเว็บไซต์ในฐานะเจ้าของข้อมูลส่วนบุคคลเห็นได้ชัดเจน ดังภาพที่ 4.14

ภาพที่ 4.14 แบบฟอร์มบันทึกข้อมูลผู้ป่วย

ข้อมูลผู้ป่วยที่ได้จากการบันทึกจะถูกเก็บไว้ที่ฐานข้อมูล และแสดงผลเป็นตารางข้อมูลดังภาพที่ 4.15

วันที่บันทึก/เริ่มป่วย	เพศ	อายุ	ช่วงอายุ	สัญชาติ	ประเภท	จังหวัด	ความเสียหาย	วันที่เริ่มป่วย	จัดการ
2021-10-01	ชาย	16	10 ถึง 19 ปี	ไทย	คนไทย	นครสวรรค์	สัมผัสผู้ติดเชื้อ	2021-09-01	
2021-10-01	ชาย	43	40 ถึง 49 ปี	ไทย	คนไทย	ปราจีนบุรี	พื้นที่เสี่ยง	2021-09-01	
2021-10-01	หญิง	20	20 ถึง 29 ปี	ไทย	คนไทย	นครสวรรค์	สัมผัสผู้ติดเชื้อ	2021-09-01	
2021-10-01	ชาย	42	40 ถึง 49 ปี	ไทย	คนไทย	นครสวรรค์	สัมผัสผู้ติดเชื้อ	2021-09-01	
2021-10-01	ชาย	23	20 ถึง 29 ปี	ไทย	คนไทย	ปราจีนบุรี	พื้นที่เสี่ยง	2021-09-01	
2021-10-01	หญิง	35	30 ถึง 39 ปี	ไทย	คนไทย	ปราจีนบุรี	พื้นที่เสี่ยง	2021-09-01	
2021-10-01	ชาย	12	10 ถึง 19 ปี	ไทย	คนไทย	นครสวรรค์	สัมผัสผู้ติดเชื้อ	2021-09-01	
2021-10-01	หญิง	2	< 10 ปี	ไทย	คนไทย	นครสวรรค์	สัมผัสผู้ติดเชื้อ	2021-09-01	
2021-10-01	ชาย	12	10 ถึง 19 ปี	ไทย	คนไทย	นครสวรรค์	สัมผัสผู้ติดเชื้อ	2021-09-01	
2021-10-01	ชาย	63	60 ถึง 69 ปี	ไทย	คนไทย	นครสวรรค์	สัมผัสผู้ติดเชื้อ	2021-09-01	

ภาพที่ 4.15 ข้อมูลผู้ป่วย

จากการบันทึกข้อมูลผู้ป่วยจะมีกราฟยอดผู้เสียชีวิตรายเดือน เพื่อดูแนวโน้มยอดคิดเชื่อในแต่ละเดือนดังภาพที่ 4.16



ภาพที่ 4.16 ยอดผู้ป่วยในแต่ละเดือน

การสำรวจข้อมูลเพื่อเปรียบเทียบคลาส ในการทำนายตามแบบฟอร์มที่ได้จากการค้นหา เพื่อดูจำนวนของคำตอบในแต่ละคลาส ดังภาพที่ 4.17

#### สำรวจข้อมูล

เพศ: ชาย | ช่วงอายุ: 20 ถึง 29 ปี | สัญชาติ: ไทย

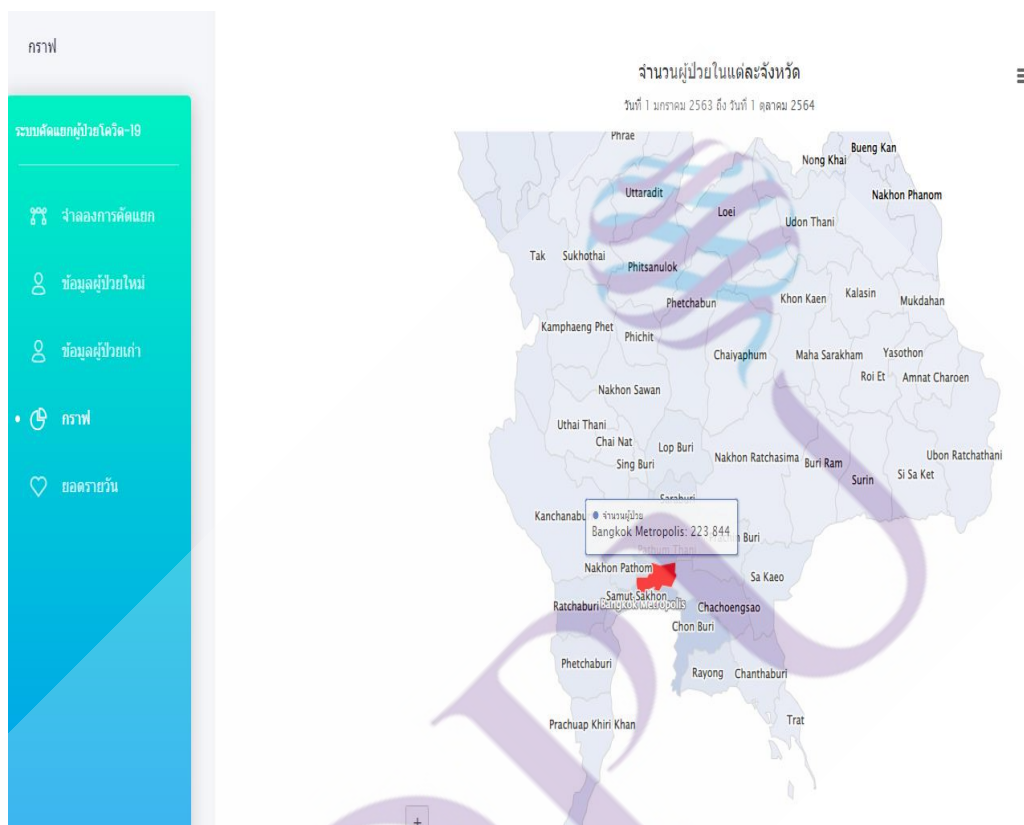
ความเสี่ยง: พื้นที่เสี่ยง | จังหวัด: กรุงเทพมหานคร

ค้นหา

CLASS	TOTAL
การเฝ้าระวังในกลุ่มผู้ป่วย หรือผู้มีโอกาสจะเข้าได้กับนิยามผู้สงสัยติดเชื้อไวรัสโคโรนา 2019 ที่เข้าเกณฑ์สอบสวนโรค (patient under investigation : PUI)	326
การตรวจคัดกรองในประชากรเสี่ยงตามจุดคัดกรองและด่านเข้าออกระหว่างประเทศ(screening)	0
การเฝ้าระวังในกลุ่มเป้าหมายเฉพาะหรือพื้นที่เฉพาะ (sentinel surveillance)	660
การเฝ้าระวังเหตุการณ์ในสถานที่เสี่ยง เก็บตัวอย่างส่งตรวจเมื่อเข้านิยาม PUI หรือเป็นกลุ่มก้อนของผู้ป่วยทางเดินหายใจ	3623

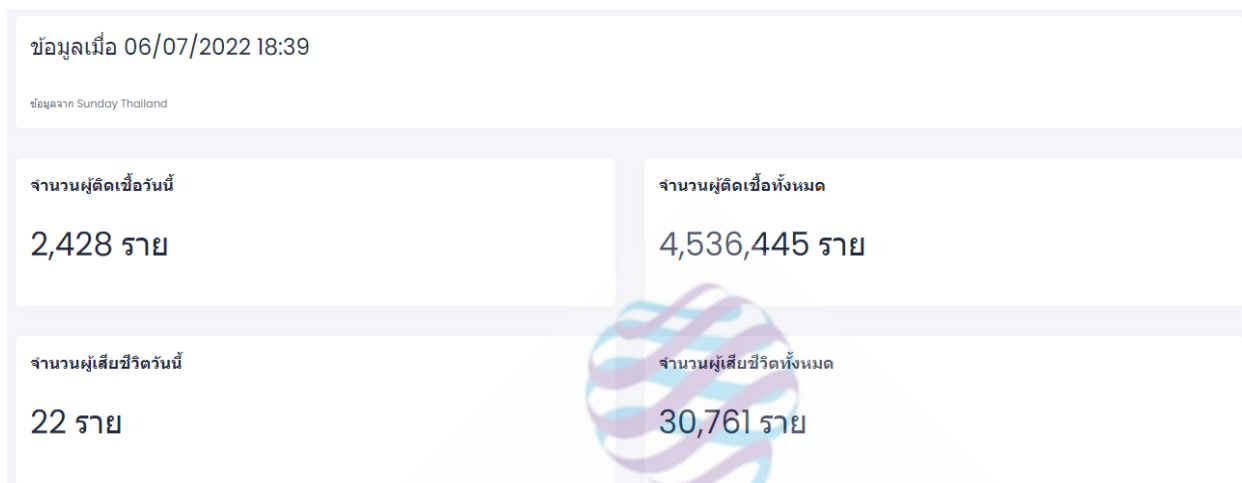
ภาพที่ 4.17 จำนวน Instance ในแต่ละคลาส

การแสดงผลแผนที่เพื่อดูว่าแต่ละจังหวัดมียอดติดเชื้อและมีความหนาแน่นของการติดเชื้อจำนวนเท่าไร โดยจะแสดงในภาพที่ 4.18



ภาพที่ 4.18 พื้นที่การติดเชื้อ

ส่วนเสริมของรายงานยอดผู้ติดเชื้อรายวัน โดยจะแสดงผลข้อมูล เช่น จำนวนผู้ติดเชื้อวันนี้ จำนวนผู้ติดเชื้อทั้งหมด จำนวนผู้เสียชีวิตวันนี้ จำนวนผู้ติดเชื้อทั้งหมด รักษาหายวันนี้ และ รักษาหายแล้วทั้งหมด โดยข้อมูลจะได้อมาจาก <https://static.easysunday.com/covid-19/getTodayCases.json> ซึ่งจะเป็นชุดข้อมูลแบบเปิดมิให้บริการข้อมูลผู้ติดเชื้อ covid-19 ทั่วโลก ซึ่งข้อมูลมีการอัปเดตทุกวันดังภาพที่ 4.19



ภาพที่ 4.19 ยอดรายวัน

ระบบฐานข้อมูลที่ใช้ในการบันทึกและเก็บข้อมูลนั้น ได้ใช้ PostgreSQL ดังภาพที่ 4.20

ptd_id	pd_id	ptd_age	ar_id	pt_id	country_code	ptr_id	province_code	ptd_id	crt_date
1	1	24	3	1	THA	3	10	1	2021-04-01 00:00:00
2	2	23	3	1	THA	3	10	1	2021-04-01 00:00:00
3	3	29	3	2	MMR	9	10	1	2021-04-01 00:00:00
4	4	27	3	2	MMR	2	74	3	2021-04-01 00:00:00
5	5	28	3	2	MMR	2	74	3	2021-04-01 00:00:00
6	6	39	4	1	THA	3	11	1	2021-04-01 00:00:00

ภาพที่ 4.20 pgAdmin 4 ระบบฐานข้อมูล PostgreSQL

จากผลการทดลองแล้วพบว่าได้แบบจำลอง Random Forest มีความเหมาะสมในการทำนายมากที่สุด จึงนำแบบจำลองนี้ไปใช้กับการพัฒนา Prediction API ด้วย FastAPI Framework เป็นภาษา Python และพัฒนาเว็บแอปพลิเคชันด้วย Vue JS เพื่อการจำแนกผู้ป่วยโควิด-19 โดยจะสรุปผลการวิจัยในบทถัดไป

## บทที่ 5

### สรุปและข้อเสนอแนะ

#### 5.1 สรุปผลการวิจัย

จากการทดลองพัฒนาแบบจำลองวิเคราะห์จำแนกผู้ป่วยโควิด-19 ด้วยการเรียนรู้ของเครื่อง ซึ่งการทดลองนี้ได้กำหนดเป้าหมายในการทำนายอยู่ 4 คลาส 1 การเฝ้าระวังในกลุ่มผู้ป่วย (PUI) 2 การตรวจคัดกรองในประชากรเสี่ยงตามจุดคัดกรอง 3 การเฝ้าระวังในกลุ่มเป้าหมายเฉพาะหรือพื้นที่เฉพาะ และ 4 การเฝ้าระวังเหตุการณ์ในสถานที่ ซึ่งการที่จะสร้างแบบจำลองนั้นต้องทำการแปลงข้อมูลเป็นตัวเลขเสียก่อน สำหรับการสร้างแบบจำลองนั้นจะใช้อัลกอริทึม 3 เทคนิคได้แก่ Naive Bayes , Random Forest และ Neural network ซึ่งเป็นอัลกอริทึมที่เหมาะสมกับการ classification ที่ดีที่สุดจากการศึกษาค้นคว้างานวิจัยที่เกี่ยวข้องมาเพื่อนำไปใช้กับการพัฒนาเว็บแอปพลิเคชันเพื่อการจำแนกผู้ป่วยโควิด-19 ได้ผลการสรุปดังต่อไปนี้

1. จากการศึกษางานวิจัยที่เกี่ยวข้อง โดยงานวิจัยที่นำการเรียนรู้ของเครื่องมาใช้จะหาผลจากการทำนายโดยการจำแนกว่าผู้ป่วยมีโอกาสเกิดการติดเชื้อ หรือ มีความรุนแรงของการติดเชื้อโควิด-19 หรือไม่ ซึ่งจะแตกต่างจากงานวิจัยนี้จะจำแนกผู้ป่วยโดยให้ความสำคัญกับการตรวจสำหรับการเฝ้าระวังในกลุ่มต่างๆ โดยยึด “แนวทางการดำเนินงานคัดกรอง เฝ้าระวัง และสอบสวนโรคติดเชื้อไวรัสโคโรนา 2019” ฉบับวันที่ 4 ธันวาคม 2563 จากกรมควบคุมโรคมาใช้ เพื่อให้ทราบขนาดปัญหา ตรวจจับการระบาด และติดตามแนวโน้มการเชื้อในกลุ่มประชากรเสี่ยงและพื้นที่เสี่ยงได้อย่างทันเวลา

2. ข้อมูลจำนวน 1,608,923 เรคคอร์ด หลังจากผ่านกระบวนการ Data Preparation แล้วข้อมูลจะอยู่ที่จำนวน 913,152 เรคคอร์ด

3. คลาส ที่ใช้ในการวิจัยประกอบไปด้วย 4 คลาส มีจำนวนข้อมูลในแต่ละคลาส ดังนี้ คลาส ที่ 1 มีจำนวน 587,079 เรคคอร์ด คลาส ที่ 2 มีจำนวน 2,755 เรคคอร์ด คลาส ที่ 3 มีจำนวน 52,548 เรคคอร์ด และคลาส ที่ 4 มีจำนวน 270,770

4. ข้อมูล Training Data (80 %) จำนวน 730,522 เรคคอร์ด และ Test Data (20%) จำนวน 182,630 เรคคอร์ด โดยจะวัดผลจากชุดข้อมูล Training Data ที่ทำการสอนให้กับคอมพิวเตอร์ และใช้ Test Data มาทดสอบความแม่นยำ ความถูกต้องของแบบจำลองที่เรียนรู้เรียบร้อยแล้ว

5. คำนวณค่าน้ำหนักของ แอททริบิวต์ ต่างๆ โดย แอททริบิวต์ ที่มีค่าน้ำหนักมากที่สุดคือ risk ค่าน้ำหนักที่ 0.816 และ แอททริบิวต์ ที่มีค่าน้ำหนักน้อยที่สุดคือ gender ค่าน้ำหนักที่ 0.003

6. ในการเปรียบเทียบประสิทธิภาพความถูกต้องและความแม่นยำในการทำนาย โดย Random Forest มีค่า Accuracy 93.33% ส่วน Neural Network มีค่า Accuracy 92.7% และ Naive Bayes มีค่า Accuracy ที่ 92.1%

7. ใช้เทคนิค Random Forest เนื่องจากให้ค่าความถูกต้องและความแม่นยำมากที่สุดมาสร้างแบบจำลองในการทำนายเพื่อจำแนกผู้ป่วยโควิด-19

8. นำแบบจำลองที่ได้จากการหาผลในการเปรียบเทียบไปใช้กับการพัฒนา Prediction API และพัฒนาเว็บแอปพลิเคชันด้วย Vue JS เพื่อการจำแนกผู้ป่วยโควิด-19

## 5.2 สรุปปัญหาและวิธีการแก้ปัญหา

### 5.2.1 สรุปปัญหา

1. การรวบรวมข้อมูลเดิมที่ข้อมูลจะเป็นลักษณะ JSON Data จึงไม่สามารถวิเคราะห์ได้
2. ข้อมูลประเทศ เพศ ความเสี่ยง ยังมีการรวบรวมในลักษณะซ้ำซ้อน การใช้คำที่มีความหมายเหมือนกัน แต่ข้อความต่างกันเล็กน้อย จึงทำให้ข้อมูลไม่เป็นระเบียบ



3. ข้อมูลที่ผู้วิจัยรวบรวมจากกรมควบคุมโรคนั้นพบว่าข้อมูลในแต่ละคุณลักษณะยังมี ส่วนที่ขาดและไม่ครบข้างในบางราย คือคุณลักษณะ อาชีพของผู้ป่วย ซึ่งเป็นข้อมูลที่ขาดไป มากคิดเป็นร้อยละ 1.72% จึงไม่สามารถนำไปวิเคราะห์ได้

### 5.2.2 วิธีการแก้ปัญหา

1. เขียนโปรแกรมแสดงข้อมูลและแปลงเป็นไฟล์ Microsoft Excel ก่อนทำ Data preparation
2. ทำการจัดกลุ่มข้อมูลให้อยู่ในรูปแบบ category
3. ทำการลบออกเพราะไม่มีผลกระทบต่อการศึกษา

## 5.3 ข้อเสนอแนะ

### 5.3.1 ข้อเสนอแนะทั่วไป

จากการทดลองพบว่า การที่แบบจำลองจะให้ประสิทธิภาพในการทำนายสูงนั้น ไม่เพียงแต่การเลือกอัลกอริทึมที่เหมาะสมสำหรับการสร้างแบบจำลอง แต่รวมไปถึงการคัดเลือก หรือทำการสร้างคุณลักษณะของข้อมูลที่สามารถนำไปวิเคราะห์ให้เหมาะสมกับประเภทงานที่จะวิเคราะห์ด้วย ซึ่งการที่จะนำไปใช้งานกับชุดข้อมูลประเภทอื่นๆ นั้นจะต้องทำการวิเคราะห์ ข้อมูลและสร้างคุณลักษณะที่เหมาะสมสำหรับการใช้งานด้วย

### 5.3.2 ข้อเสนอแนะสำหรับการวิจัยครั้งต่อไป

1. ผลการวิจัยแสดงให้เห็นว่าอัลกอริทึมการเรียนรู้ของเครื่องนั้นสามารถขยายขอบเขต จากการจำแนกผู้ป่วยโควิด-19 ไปประยุกต์ร่วมใช้กับการจำแนกความเสี่ยงในการเกิดโรค หรือการจำแนกการรักษา
2. ควรเพิ่มคุณลักษณะของข้อมูล เช่น อาการของผู้ป่วย อุณหภูมิร่างกาย เวลารักษา
3. การวิจัยประเภท Regression เพื่อหา Time series ช่วงเวลาของกลุ่มการติดเชื้อ
4. ทดลองกับอัลกอริทึมอื่นๆ และหามาตรวัดแบบจำลองตามประเภทวิจัย



บรรณานุกรม

DPU

## บรรณานุกรม

### ภาษาไทย

กรมควบคุมโรค. (2563). แนวทางการดำเนินงานคัดกรอง เฝ้าระวัง และสอบสวนโรคติดเชื้อไวรัสโคโรนา 2019.

[https://ddc.moph.go.th/viralpneumonia/file/g\\_srrt/g\\_srrt\\_241263.pdf](https://ddc.moph.go.th/viralpneumonia/file/g_srrt/g_srrt_241263.pdf)

กิจผาคติ, นพ. ร. (2563). มุลินธิส่งเสริมการศึกษาใช้หัดใหญ่. มุลินธิส่งเสริมการศึกษาใช้หัดใหญ่.

<http://www.ift2004.org/summarize/iftnews2/11.html>

เฉลยกิตติ, ศ., อาจสันเทียะ, จ., & สุภาพร ดาวดี, ม. (2563). ผลกระทบโรคระบาด COVID-19: การจัดการสอน  
ทางการพยาบาล. *วารสารสุขภาพและการพยาบาล ปีที่ 36, 36(2)*.

ชุตินาสเจริญ, อ., & แสนราช, ดร. จ. (2563). การเปรียบเทียบประสิทธิภาพของอัลกอริทึมและการคัดเลือก  
คุณลักษณะที่เหมาะสมเพื่อการพยากรณ์โอกาสความสำเร็จในการ โอนเงินข้ามประเทศของบุคคลทั่วไป.

6(3). <https://so04.tci-thaijo.org/index.php/gskkuhs/article/download/156370/113480/>

ดุษฎีเจริญ, จ., วงษ์ประกอบ, ณ., หมั่นสระเกษ, ก., & นิมกระ โทก, ธ. (2563). การรับมือกับไวรัสโคโรนา COVID-  
19 ในงานสาธารณสุขมูลฐาน. *วารสารวิทยาศาสตร์สุขภาพ วิทยาลัยพยาบาลบรมราชชนนี สรรพสิทธิ์  
ประสงค์, 3*. <https://he01.tci-thaijo.org/index.php/bcnsp/article/download/241935/167311/>

ทุมวงษ์, น. (2562). การเปรียบเทียบประสิทธิภาพระหว่างสถาปัตยกรรมแบบโมโนลิธิคและไมโครเซอร์วิสโดยใช้ดอก  
เกอร์และคูเบอร์เนตส์ [มหาวิทยาลัยศรีนครินทรวิโรฒ]. <http://ir->

[thesis.swu.ac.th/dspace/bitstream/123456789/611/1/gs611130434.pdf](http://ir-thesis.swu.ac.th/dspace/bitstream/123456789/611/1/gs611130434.pdf)

ชันวานนท์, ป., ยี่งเสรี, ช., พงษ์เพชร, ว., & ผนังะจิตร, ธ. (2560). การประยุกต์ใช้โมเดลการเรียนรู้แบบรวมกลุ่มเพื่อ  
พยากรณ์แนวโน้มของราคาหลักทรัพย์ในตลาดหลักทรัพย์แห่งประเทศไทย. *JOURNAL OF*

*INFORMATION SCIENCE AND TECHNOLOGY, 7(1)*. <https://ph02.tci->

[thaijo.org/index.php/JIST/article/download/135457/101205/](https://ph02.tci-thaijo.org/index.php/JIST/article/download/135457/101205/)

- ปักเข็ม, น., & จันมณีชัย, ช. (2560). การจำแนกประเภทภูมิปัญญาท้องถิ่นของไทยแบบอัตโนมัติโดยวิธีการทางเหมืองข้อมูล. *วารสารมหาวิทยาลัยทักษิณ ฉบับพิเศษ*, 20.
- โพธิ์สอาด, ส., & โปสาวัง, ป. (2562). การจำแนกพฤติกรรมการขับขี่รถโดยสารสาธารณะโดยใช้วิธีการสกัดข้อความและเทคนิคการเรียนรู้ของเครื่อง. *วารสารเทคโนโลยีสารสนเทศ*, 15(1). [https://ph01.tci-thaijo.org/index.php/IT\\_Journal/article/download/199905/139705/](https://ph01.tci-thaijo.org/index.php/IT_Journal/article/download/199905/139705/)
- ม่วงน้อย, เ., พลาคผล, จ., & ปาลวิสุทธิ, ภ. (2564). ตัวประเมินภาวะความเสี่ยงการเป็นโรคซึมเศร้าของนักศึกษาด้วยเทคนิคเหมืองข้อมูล. *วารสารวิชาการประยุกต์ใช้เทคโนโลยีสารสนเทศ*, 7. <https://ph02.tci-thaijo.org/index.php/project-journal/article/download/242196/165791/>
- มานวิภา, ก. (2562). ระบบแนะนำทางการศึกษา และเทคนิคการเรียนรู้ของเครื่องจักร. *วารสารสารสนเทศศาสตร์*, 37(2).
- ลิ่มตระกูล, ป., วิภา, เ., & วิทยา, พ. (2559). การพัฒนาแบบการพยากรณ์ผลผลิตมันสำปะหลังด้วยเทคนิคการทำเหมืองข้อมูล. *สาขาวิทยาศาสตร์และเทคโนโลยี*, 3(3). <https://ph01.tci-thaijo.org/index.php/VESTSU/article/download/63634/52233/>
- สำนักงานพัฒนารัฐบาลดิจิทัล (องค์การมหาชน) (สพร.). (2015, May 29). Open Data: ระดับการเปิดเผยข้อมูล 5 ระดับ. สำนักงานพัฒนารัฐบาลดิจิทัล (องค์การมหาชน) สพร. หรือ DGA. <https://www.dga.or.th/document-sharing/article/35847/>
- สุขสกาฟอง, ป., & มีสัง, พ. (2559). การทำเหมืองข้อมูลกฎความสัมพันธ์ด้วยวิธีค่าถ่วงน้ำหนักสูงสุดจากกราฟบริบูรณ์แบบสมมาตร. *วารสารเทคโนโลยีสารสนเทศ*, 12(1). [https://ph01.tci-thaijo.org/index.php/IT\\_Journal/article/download/72838/58584/](https://ph01.tci-thaijo.org/index.php/IT_Journal/article/download/72838/58584/)
- หาญสมบูรณ์, ป. (2564). ค่ารักษาพยาบาลของโรคโควิด-19. *วารสารกรมการแพทย์ ปีที่ 46*, 46(2). <https://he02.tci-thaijo.org/index.php/JDMS/article/download/253408/172281>

## ภาษาต่างประเทศ

9Expert, T. (2018). การเลือกรูปแบบ *Visualization* ให้เหมาะสมกับข้อมูล. 9Expert Training.

<https://www.9experttraining.com/articles/การเลือกรูปแบบ-visualization-ให้เหมาะสมกับข้อมูล>

Abdul Salam, M., Taha, S., & Ramadan, M. (2021). COVID-19 detection using federated machine learning. *PLOS ONE*, *16*(6), e0252573. <https://doi.org/10.1371/journal.pone.0252573>

Alpaydin, E. (2020). *Introduction to Machine Learning* (New edition). MIT Press Ltd.

<https://www.bookdepository.com/Introduction-to-Machine-Learning-Ethem-Alpaydin-OEzyegin-University/9780262043793>

Ardakani, A. A., Kanafi, A. R., Acharya, U. R., Khadem, N., & Mohammadi, A. (2020). Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images: Results of 10 convolutional neural networks. *Computers in Biology and Medicine*, *121*, 103795. <https://doi.org/10.1016/j.combiomed.2020.103795>

Bhatia, S., & Malhotra, J. (2021). Naïve Bayes Classifier for Predicting the Novel Coronavirus. *2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)*, 880–883. <https://doi.org/10.1109/ICICV50876.2021.9388410>

chengz. (2019, October 3). วัดประสิทธิภาพ Model จาก Confusion Matrix. *Medium*.

<https://medium.com/@cheng3374/วัดประสิทธิภาพ-model-จาก-confusion-matrix-69d391bcd48>

Daroontham, W. (2018). เจาะลึก *Random Forest* “รู้จัก *Decision Tree*, *Random Forest*, และ *XGBoost*.” *Medium*. <https://medium.com/@witchapongdaroontham/เจาะลึก-random-forest-part-2-of-รู้จัก-decision-tree-random-forest-และ-xgboost-79b9f41a1c1c>

Department of disease control. (2021). *DDC OPENDATA Covid19 Thailand*. DDC OPENDATA Covid19 Thailand. <https://covid19.ddc.moph.go.th/>

- Digest, E. (2020, June). *Artificial intelligence (AI)—A catalyst for China’s post-COVID economy*. Kasikorn Research Center. <https://www.kasikornresearch.com/th/analysis/k-social-media/Pages/AI-China-FB300620.aspx>
- Duangklang, P., & Kruakaew, R. (2019). Models for automatic aircraft type prediction. *NKRAFA Journal of Science and Technology*.
- Fayong, S. (2019, December 7). *Machine Learning 4 ประเภท*. Code on the Hill. <http://codeonthehill.com/machine-learning-types/>
- FastAPI*. (n.d.). Retrieved July 5, 2022, from <https://fastapi.tiangolo.com/>
- Highcharts. (n.d.). *สร้างกราฟ ด้วย Highcharts JQuery Javascript Graph*. Retrieved July 5, 2022, from <https://www.mindphp.com/developer/36-jquery/2397-jquery-ทำกราฟ-ด้วย-highcharts.html>
- Gudise, V. G., & Venayagamoorthy, G. K. (2003). Comparison of particle swarm optimization and backpropagation as training algorithms for neural networks. *Proceedings of the 2003 IEEE Swarm Intelligence Symposium. SIS'03 (Cat. No.03EX706)*, 110–117. <https://doi.org/10.1109/SIS.2003.1202255>
- Iwendi, C., Bashir, A. K., Peshkar, A., Sujatha, R., Chatterjee, J. M., Pasupuleti, S., Mishra, R., Pillai, S., & Jo, O. (2020). COVID-19 Patient Health Prediction Using Boosted Random Forest Algorithm. *Frontiers in Public Health*, 8, 357. <https://doi.org/10.3389/fpubh.2020.00357>
- Karhan, Z., & Akal, F. (2020). Covid-19 Classification Using Deep Learning in Chest X-Ray Images. *2020 Medical Technologies Congress (TIPTEKNO)*, 1–4. <https://doi.org/10.1109/TIPTEKNO50054.2020.9299315>
- Karwatka, P. (2020). *Monolithic architecture vs microservices: Which is better?* | Divante. <https://www.divante.com/blog/monolithic-architecture-vs-microservices>

- Kasetsart university. (2020, July). *Covid-19 and epidemiology* [E-Learning]. COVID-19 Course.  
<https://learningcovid.ku.ac.th/course/?c=3&l=1>
- Lalmuanawma, S., Hussain, J., & Chhakchhuak, L. (2020). Applications of machine learning and artificial intelligence for Covid-19 (SARS-CoV-2) pandemic: A review. *Chaos, Solitons & Fractals*, 139, 110059. <https://doi.org/10.1016/j.chaos.2020.110059>
- Manmana, S., Iamsirithaworn, S., & Uttayamakul, S. (2020). Coronavirus Disease-19 (Covid-19). *Journal of Bamrasnaradura Infectious Diseases Institute*, 14(2), 124–133.
- Mansour, N. A., Saleh, A. I., Badawy, M., & Ali, H. A. (2021). Accurate detection of Covid-19 patients based on Feature Correlated Naïve Bayes (FCNB) classification strategy. *Journal of Ambient Intelligence and Humanized Computing*, 1–33. <https://doi.org/10.1007/s12652-020-02883-2>
- Muhammad, B. (2017). *Open Government Data Mining and Visualization: A Case Study of New Year Festival Traffic Accidents in Thailand*. Asian Institute of Technology School of Engineering and Technology Thailand.
- Ozturk, T., Talo, M., Yildirim, E. A., Baloglu, U. B., Yildirim, O., & Rajendra Acharya, U. (2020). Automated detection of COVID-19 cases using deep neural networks with X-ray images. *Computers in Biology and Medicine*, 121, 103792.  
<https://doi.org/10.1016/j.combiomed.2020.103792>
- Pacharawongsakda, E. (2014). *An introduction to Data Mining Techniques* (Vol. 2).
- Patcharacharoenwong, C., Kankawee, H., & Warangkhan, K. (2020). Arrival Time Prediction Model to a Pier for Public Transportation Boats. *Journal of Science Ladkrabang*, 29(2).  
[https://li01.tci-thaijo.org/index.php/science\\_kmitl/article/download/241105/169801/](https://li01.tci-thaijo.org/index.php/science_kmitl/article/download/241105/169801/)



- P.Pitchaya. (2020, November 26). Data and Data Visualization ความสำคัญและอิทธิพลต่อชีวิตเรา. *KBTG Life*. <https://medium.com/kbtg-life/data-and-data-visualization-ความสำคัญและอิทธิพลต่อชีวิตเรา-9e154de3f5ab>
- Puitrakul, A. (2018, January). *What is machine learning* [Blog]. What is machine learning. <https://arnondora.in.th/what-is-machine-learning/>
- Rachada, T.-N. (2020). COVID-19: An Invisible War Against Coronavirus. *THAI FOOD AND DRUG JOURNAL*, 27(2).
- Srimarong, S. (2020, August 28). 4 ประเภทของการแบ่งกลุ่มข้อมูล (Clustering). Big Data Thailand. <https://bigdata.go.th/big-data-101/4-types-of-clustering/>
- Sun, L., Song, F., Shi, N., Liu, F., Li, S., Li, P., Zhang, W., Jiang, X., Zhang, Y., Sun, L., Chen, X., & Shi, Y. (2020). Combination of four clinical indicators predicts the severe/critical symptom of patients infected COVID-19. *Journal of Clinical Virology*, 128, 104431. <https://doi.org/10.1016/j.jcv.2020.104431>
- Thongjit, A., Suksawang, P., & Jatupat, M. (2019). Development of Data Classification using a Hybrid Method of Adaptive Artificial Neural Networks and Particle Swarm Optimization for Identifying Patients at Risk of Diabetes. *Research Methodology & Cognitive Science*, 17(2). <https://so05.tci-thaijo.org/index.php/RMCS/article/download/243179/165072/>
- Vue.js—The Progressive JavaScript Framework* | *Vue.js*. (n.d.). Retrieved July 5, 2022, from <https://vuejs.org/>
- Witten, Ian H., Frank, E., & Hakk, M. A. (2011). *Data Mining: Practical Machine Learning Tools and Techniques*. Elsevier. <https://doi.org/10.1016/C2009-0-19715-5>
- Wu, J., Zhang, P., Zhang, L., Meng, W., Li, J., Tong, C., Li, Y., Cai, J., Yang, Z., Zhu, J., Zhao, M., Huang, H., Xie, X., & Li, S. (2020). *Rapid and accurate identification of COVID-19*

*infection through machine learning based on clinical available blood test results* (p. 2020.04.02.20051136). <https://doi.org/10.1101/2020.04.02.20051136>

Zoabi, Y., Deri-Rozov, S., & Shomron, N. (2021). Machine learning-based prediction of COVID-19 diagnosis based on symptoms. *Npj Digital Medicine*, 4(1), 3. <https://doi.org/10.1038/s41746-020-00372-6>





ภาคผนวก

DPU



**ภาคผนวก ก.**

**API สำหรับการทำนาย**

**DRU**

## API สำหรับการทำนาย

1. **User** คือส่วนจัดการฐานข้อมูลผู้ใช้งาน (User) โดยจะสามารถเพิ่ม ลบ แก้ไข และ ตรวจสอบการเข้าสู่ระบบของผู้ใช้งานได้

User	
GET	/user/users Get Users
POST	/user/users Create User
GET	/user/users/p Get Users With Pagination
GET	/user/user/{user_id} Get User
PUT	/user/user/{user_id} Update User
DELETE	/user/user/{user_id} Delete User
POST	/user/login Login

ภาพที่ ก.1 User

สามารถดูรายชื่อผู้ใช้งานได้โดยใช้ Method GET (/user/users) ดังภาพที่ ก.2

```

Curl
curl -X 'GET' \
  'https://masterthesisproject.herokuapp.com/user/users' \
  -H 'accept: application/json'

Request URL
https://masterthesisproject.herokuapp.com/user/users

Server response
Code    Details
200

Response body
{
  "user_id": "a9b4895f-e00b-44f6-b7d3-ch9f67bc4fd7",
  "username": "siwakorn",
  "first_name": "Siwakorn",
  "last_name": "Bomlueapp",
  "password_hash": "siwakorn",
  "crt_date": "2022-06-29T12:52:29.784000+00:00",
  "upd_date": "2022-06-29T10:17:42.343413+00:00",
  "active": 1,
  "cancelled": 1
}

Response headers
connection: keep-alive
content-length: 267
content-type: application/json
date: Wed, 06 Jul 2022 04:32:05 GMT
server: ovicorn
via: 1.1 vegur

```

ภาพที่ ก.2 การเรียกดูข้อมูลผู้ใช้งานทั้งหมด

จากภาพที่ ก.2 จะสามารถแสดงรายชื่อผู้ใช้งานได้จะต้องมีการเพิ่มข้อมูลผู้ใช้งานก่อนโดยจะใช้ Method POST (/user/users) ซึ่งจะสามารถอธิบายได้ดังรูปที่ ก.3

POST /user/users Create User

Parameters Cancel Reset

No parameters

Request body required application/json

```
{
  "username": "string",
  "first_name": "string",
  "last_name": "string",
  "password_hash": "string",
  "crt_date": "2022-07-06T04:41:37.520Z",
  "udp_date": "2022-07-06 11:28:02.443598+07:00",
  "active": 0,
  "cancelled": 1
}
```

Execute

ภาพที่ ก.3 การเพิ่มข้อมูลผู้ใช้งาน

จากภาพที่ ก.3 การเพิ่มผู้ใช้งานจะประกอบไปด้วยข้อมูล ดังนี้

1. user\_name หมายถึง ชื่อผู้ใช้งานสำหรับการเข้าสู่ระบบ
2. first\_name หมายถึง ชื่อที่แสดงตัวตนการเข้าใช้งานในระบบ
3. last\_name หมายถึง นามสกุลที่แสดงตัวตนการเข้าใช้งานในระบบ
4. password\_hash หมายถึง รหัสการเข้าใช้งาน
5. crt\_date หมายถึง วันที่สร้างผู้ใช้งาน
6. udp\_date หมายถึง วันที่ในการปรับปรุงข้อมูล
7. active หมายถึง ผู้ใช้งานมีการยืนยันความถูกต้อง
8. cancelled หมายถึง ยกเลิกผู้ใช้งานนี้

เมื่อสามารถเพิ่มข้อมูลได้แล้ว ยังมีการเรียกดูข้อมูลผู้ใช้งานอีก 2 วิธีคือการเรียกดูข้อมูลโดยมีการแบ่งหน้า (Pagination) และ เรียกดูข้อมูลโดยอ้างอิงรหัสผู้ใช้งาน เพื่อรองรับการนำไปใช้งานกับเว็บแอปพลิเคชันได้อย่างรวดเร็วโดยไม่ต้องโหลดข้อมูลในปริมาณหลายเรคคอร์ด โดยจะใช้ Method GET (/user/user/p?page=1&per\_page=50) สามารถดูรายละเอียดได้ดังภาพที่ ก.4 - ก.7

GET /user/users/p Get Users With Pagination

Parameters Cancel

Name	Description
page integer (query)	1
per_page integer (query)	50

Execute Clear

Responses

ภาพที่ ก.4 การกำหนดช่วงการดึงข้อมูลผู้ใช้งาน

จากภาพที่ ก.4 จะระบุ 2 ค่าด้วยกันคือ page หมายถึง หมายเลขหน้า และ per\_page หมายถึง จำนวนการแสดงผลต่อหน้าในการดึงข้อมูล โดยทั้ง 2 ค่าดังกล่าวจะต้องเป็นตัวเลขเท่านั้น ซึ่งเมื่อกด Execute แล้วจะแสดงผลข้อมูลดังภาพที่ ก.5

Code Details

200

Response body

```
{
  "status": "Success",
  "message": "Show Data",
  "page": 1,
  "per_page": 50,
  "total_data": 1,
  "total_page": 1,
  "res_data": [
    {
      "user_id": "a9b4895f-e00b-44f6-b7d3-cb9f67bc4fd7",
      "username": "siwakorn",
      "first_name": "Siwakorn",
      "last_name": "Banluesapy",
      "password_hash": "siwakorn",
      "crt_date": "2022-06-29T12:52:29.784000+00:00",
      "upd_date": "2022-06-29T10:17:42.343413+00:00",
      "active": 1,
      "cancelled": 1
    }
  ]
}
```

ภาพที่ ก.5 การแสดงข้อมูลผู้ใช้งานแบบแบ่งหน้า

จากภาพที่ ก.5 จะแสดงการดึงข้อมูลตามค่าที่ระบุในภาพที่ ก.4 โดยจะประกอบไปด้วยดังนี้

1. status หมายถึง สถานะการดึงข้อมูล
2. message หมายถึง ข้อความตอบกลับ



3. page หมายถึง หมายเลขหน้า
4. per\_page หมายถึง จำนวนข้อมูลที่แสดงต่อหน้า
5. total\_data หมายถึง จำนวนข้อมูลทั้งหมด
6. total\_page หมายถึง จำนวนหน้าทั้งหมด
7. res\_data หมายถึง ข้อมูลผู้ใช้งาน

นอกจากนี้การดึงข้อมูลผู้ใช้งานยังสามารถดึงข้อมูลเฉพาะ โดยจะอ้างอิงรหัสผู้ใช้งานเพื่อดึงเฉพาะผู้ใช้งานนั้นซึ่งจะใช้ Method GET (/user/user/{user\_id}) สามารถแสดงผลข้อมูลดังภาพที่ ก.6

The screenshot shows a REST client interface with a 'Parameters' section. It contains a table with two columns: 'Name' and 'Description'. The first row has 'user\_id' in the Name column, which is marked as 'required' in red. The Description column contains the value '3369302d-4a9d-44ad-94d1-6689adaff527'. Below the table are two buttons: 'Execute' and 'Clear'. At the bottom, there is a 'Responses' section which is currently empty.

Name	Description
user_id * required string (path)	3369302d-4a9d-44ad-94d1-6689adaff527

ภาพที่ ก.6 การกำหนดรหัสผู้ใช้งานเพื่อดูข้อมูลเฉพาะ

จากภาพที่ ก.6 จะกำหนดรหัส (user\_id) ของผู้ใช้งานเพื่อดูข้อมูลเฉพาะของผู้ใช้งานนั้น ซึ่งเมื่อกด Execute แล้วจะแสดงผลข้อมูลดังภาพที่ ก.7

Server response

Code	Details
200	<p>Response body</p> <pre>{   "user_id": "3369302d-4a9d-44ad-94d1-6689adaff527",   "username": "faii",   "first_name": "เจนจิรา",   "last_name": "เป็นมงคล",   "password_hash": "12345",   "crt_date": "2022-07-06T04:41:37.520000+00:00",   "upd_date": "2022-07-06T04:28:02.443598+00:00",   "active": 1,   "cancelled": 1 }</pre> <p>Response headers</p> <pre>connection: keep-alive content-length: 285 content-type: application/json date: Wed, 06 Jul 2022 05:07:44 GMT server: uvicorn via: 1.1 vegur</pre>

ภาพที่ ก.7 ข้อมูลผู้ใช้งาน โดยดึงผ่านรหัสผู้ใช้งาน

เมื่อสามารถเพิ่มและเรียกดูข้อมูล แล้วโดยข้อมูลจะต้องสามารถแก้ไขได้ จะใช้ Method PUT (/user/user/{user\_id}) โดจะส่งรหัสผู้ใช้งาน (user\_id) เพื่อแก้ไขข้อมูลนั้นซึ่งจะแสดงผลดังภาพที่ ก.8

PUT /user/user/{user\_id} Update User

Parameters

Name	Description
<b>user_id</b> * required string (path)	3369302d-4a9d-44ad-94d1-6689adaff527

Request body required

```
{
  "username": "faii",
  "first_name": "เจนจิรา",
  "last_name": "เป็นมงคล",
  "password_hash": "12345",
  "crt_date": "2022-07-06T04:41:37.5202",
  "upd_date": "2022-07-06 11:28:02.443598+07:00",
  "active": 1,
  "cancelled": 1
}
```

Execute

ภาพที่ ก.8 แก้ไขข้อมูลผู้ใช้งาน โดยดึงผ่านรหัสผู้ใช้งาน

จากภาพที่ ก.8 จะระบุรหัสผู้ใช้งานเพื่อจะได้แก้ไขข้อมูลได้ถูกต้องตามรหัสที่ระบุมา หลังจากส่งข้อมูลและกด Execute แล้วจะแสดงผลข้อมูลดังภาพที่ ก.9

The screenshot shows the response body and headers for a successful update operation. The response body is a JSON object containing user information, and the headers include access-control and content-type information.

```

Code    Details
200
Response body
{
  "user_id": "3369302d-4a9d-44ad-94d1-6689adaff527",
  "username": "fail",
  "first_name": "เจนจิรา",
  "last_name": "เปี่ยมพนา",
  "password_hash": "12345",
  "crt_date": "2022-07-06T04:41:37.520000+00:00",
  "upd_date": "2022-07-06T04:28:02.443598+00:00",
  "active": 1,
  "cancelled": 1
}
Response headers
access-control-allow-credentials: true
access-control-allow-origin: *
connection: keep-alive
content-length: 285
content-type: application/json
date: Wed, 06 Jul 2022 06:45:01 GMT
server: uvicorn
via: 1.1 vegur
  
```

ภาพที่ ก.9 ข้อตอบกลับหลังจากแก้ไขข้อมูล

ในส่วนสุดท้ายหลังจากที่ได้เพิ่ม ลบ แสดง แล้วจะเป็นส่วนในการลบข้อมูลผู้ใช้งาน ซึ่งจะใช้ Method DELETE (/user/user/{user\_id}) ตามรหัสผู้ใช้งานซึ่งจะแสดงผลดังภาพที่ ก.10

The screenshot shows the configuration for a DELETE request to delete a user. The endpoint is /user/user/{user\_id} and the user\_id parameter is set to 3369302d-4a9d-44ad-94d1-6689adaff527.

DELETE /user/user/{user\_id} Delete User

Parameters

Name	Description
user_id * required	
string	
(path)	

Execute

ภาพที่ ก.10 การส่งรหัสข้อมูลผู้ใช้งานเพื่อทำการลบ

เมื่อจัดการส่วนข้อมูลผู้ใช้งานครบแล้วในส่วนถัดไปจะเป็นการเข้าสู่ระบบโดยจะประกอบไปด้วยข้อมูล 2 แอททริบิวต์ คือ user\_name และ password โดยจะใช้ Method POST ดังภาพที่ ก.11

The screenshot shows the configuration for a POST request to log in. The endpoint is /user/login and the request body is set to application/json.

POST /user/login Login

Parameters

No parameters

Request body required

application/json

Example Value | Schema

```

{
  "user_name": "string",
  "user_password": "string"
}
  
```

ภาพที่ ก.11 การส่งข้อมูลเพื่อตรวจสอบตัวตนในการเข้าสู่ระบบ

## 2. Patient คือส่วนจัดการฐานข้อมูลผู้ป่วย

Patient	
GET	/patient/patient_data/p Get Patient Data With Pagination
GET	/patient/patient_data/{ptd_id} Get Patient Data Id
GET	/patient/patient_data_new/last Get Patient Data Last
POST	/patient/patient_data_new Create Patient Data
GET	/patient/patient_data_new/p Get Patient Data New With Pagination
GET	/patient/patient_data_new/{ptd_id} Get Patient Data Id
PUT	/patient/patient_data_new/{ptd_id} Update Patient Data
DELETE	/patient/patient_data_new/{ptd_id} Delete User
POST	/patient/patient_address Create Patient Address
GET	/patient/patient_address/{ptd_id} Get Patient Address Id
PUT	/patient/patient_address/{ptd_id} Update Patient Data

ภาพที่ ก.12 Patient

Patient	
GET	/patient/patient_data/p Get Patient Data With Pagination
Parameters	
Name	Description
page Integer (query)	1
per_page Integer (query)	50
Execute	
Clear	

ภาพที่ ก.13 การกำหนดช่วงการดึงข้อมูลผู้ป่วย

จากภาพที่ ก.13 จะระบุ 2 ค่าด้วยกันคือ page หมายถึง หมายเลขหน้า และ per\_page หมายถึง จำนวนการแสดงผลต่อหน้าในการดึงข้อมูล ซึ่งทั้ง 2 ค่าดังกล่าวจะต้องเป็นตัวเลขเท่านั้น โดยจะใช้ Method Get (/patient/patient\_data/p?page=1&per\_page=50) หลังจากเมื่อกด Execute แล้วจะแสดงผลข้อมูลดังภาพที่ ก.14

Code	Details
200	<p>Response body</p> <pre> {   "status": "Success",   "message": "Show Data",   "page": 1,   "per_page": 50,   "total_data": 913152,   "total_page": 18264,   "res_data": [     {       "ptd_id": 912907,       "gender": "ชาย",       "ptd_age": 16,       "arType": "10 ถึง 19 ปี",       "nationality": "ไทย",       "nationalityType": "คนไทย",       "province": "นครสวรรค์",       "patient_risk": "สัมผัสผู้ติดเชื้อ",       "crt_date": "2021-10-01 00:00:00+00:00",       "upd_date": "2021-09-01 00:40:49+00:00"     },     {       "ptd_id": 912925,       "gender": "หญิง",       "ptd_age": 1,       "arType": "&lt; 10 ปี",       "nationality": "ไทย",       "nationalityType": "คนไทย",       "province": "นครสวรรค์",       "patient_risk": "สัมผัสผู้ติดเชื้อ",       "crt_date": "2021-10-01 00:00:00+00:00",       "upd_date": "2021-09-01 00:40:49+00:00"     }   ] } </pre>

ภาพที่ ก.14 การแสดงข้อมูลผู้ใช้งานแบบแบ่งหน้า

จากภาพที่ ก.14 จะแสดงการดึงข้อมูลตามค่าที่ระบุในภาพที่ ก.4 โดยจะประกอบไปด้วยดังนี้

1. status หมายถึง สถานะการดึงข้อมูล
2. message หมายถึง ข้อความตอบกลับ
3. page หมายถึง หมายเลขหน้า
4. per\_page หมายถึง จำนวนข้อมูลที่แสดงต่อหน้า
5. total\_data หมายถึง จำนวนข้อมูลทั้งหมด
6. total\_page หมายถึง จำนวนหน้าทั้งหมด
7. res\_data หมายถึง ข้อมูลผู้ใช้งาน

นอกจากนี้การดึงข้อมูลยังสามารถดึงข้อมูลเฉพาะโดยอ้างอิงรหัสผู้ป่วยได้ ซึ่งจะใช้ Method GET (/patient/patient\_data/{ptd\_id}) ดังภาพที่ ก.15

GET /patient/patient\_data/{ptd\_id} Get Patient Data Id

Parameters Cancel

Name	Description
ptd_id * required integer (path)	<input type="text" value="2"/>

Execute Clear

### ภาพที่ ก.15 การกำหนดรหัสผู้ป่วยเพื่อดูข้อมูลเฉพาะ

จากภาพที่ ก.15 เมื่อระบุรหัสผู้ป่วยและทำการกดปุ่ม Execute แล้ว จะแสดงผลข้อมูลดังภาพที่ ก.16

Server response

Code Details

200

Response body

```
{
  "ptd_id": 2,
  "gd_id": 1,
  "ptd_age": 23,
  "ar_id": 3,
  "nt_id": 1,
  "country_code": "THA",
  "ptr_id": 3,
  "province_code": 10,
  "ptt_id": 1,
  "crt_date": "2021-04-01T00:00:00+00:00",
  "upd_date": "2021-09-01T00:40:49+00:00"
}
```

Response headers

### ภาพที่ ก.16 ข้อมูลผู้ป่วยโดยดึงผ่านรหัสผู้ป่วย

จากภาพที่ ก.16 ข้อมูลผู้ป่วยสามารถอธิบายรายละเอียดได้ดังนี้

1. gd\_id หมายถึง รหัสเพศ
2. ptd\_age หมายถึง อายุ
3. ar\_id หมายถึง รหัสช่วงอายุ
4. nt\_id หมายถึง รหัสประเภทสัญชาติ
5. country\_code หมายถึง รหัสประเทศ
6. ptr\_id หมายถึง รหัสความเสี่ยง
7. province หมายถึง รหัสจังหวัด
8. ptt\_id หมายถึง รหัสประเภทผู้ป่วย
9. crt\_date หมายถึง วันที่สร้างผู้ป่วย
10. upd\_date หมายถึง วันที่ในการปรับปรุงข้อมูล

จากรายละเอียดข้อมูลข้างต้นสามารถนำมาใช้กับส่วนสร้างข้อมูลผู้ป่วยได้ โดยจะใช้ Method POST(/patient/patient\_data\_new) ดังภาพที่ ก.17



ภาพที่ ก.17 การเพิ่มข้อมูลผู้ป่วย

เมื่อสามารถเพิ่มข้อมูลได้แล้ว สามารถทำการแก้ไขข้อมูลได้โดยจะใช้ Method PUT (/patient/patient\_data\_new/{ptd\_id}) โดยจะส่งรหัสผู้ป่วย (ptd\_id) เพื่อแก้ไขข้อมูลรหัสนั้น ดังภาพที่ ก.18



**PUT** /patient/patient\_data\_new/{ptd\_id} Update Patient Data

**Parameters**

Name	Description
ptd_id * required integer (path)	1

**Request body** required

```
{
  "gender": "ชาย",
  "ptd_age": 45,
  "arType": "40 ถึง 49 ปี",
  "nationality": "ไทย",
  "nationalityType": "คนไทย",
  "province": "กรุงเทพมหานคร",
  "patient_risk": "Cluster",
  "crt_date": "2022-06-27 00:00:00+00:00",
  "upd_date": "2022-06-29 17:21:09.092143+00:00"
}
```

Execute

### ภาพที่ ก.18 การส่งรหัสและข้อมูลผู้ป่วยเพื่อแก้ไขข้อมูล

เมื่อสามารถเพิ่ม แสดงข้อมูล และแก้ไขข้อมูลได้แล้ว ส่วนถัดมาจะเป็นส่วนเพิ่มที่อยู่ของผู้ป่วย

โดยจะใช้ Method POST (/patient/patient\_address) ดังภาพที่ ก.19

**POST** /patient/patient\_address Create Patient Address

**Parameters**

No parameters

**Request body** required

```
{
  "pta_idcard": "10101010",
  "pta_firstname": "มานะ",
  "pta_lastname": "ใจดี",
  "pta_address_number": "178",
  "pta_email": "sinakorn167@mail.com",
  "pta_phone": "09439080477",
  "pta_img": "https://i.imgur.com/IXH6YEV.jpg",
  "province_code": 330101,
  "amphur_code": 0,
  "district_code": 0,
  "zipcode": 33000,
  "ptd_id": 10,
  "crt_date": "2022-06-22T12:44:04.590000+00:00"
}
```

Execute

### ภาพที่ ก.19 การเพิ่มข้อมูลที่อยู่ผู้ป่วย

จากภาพที่ ก.18 สามารถอธิบายรายละเอียดข้อมูลได้ดังนี้

1. pta\_idcard หมายถึง รหัสบัตรประชาชน
2. pta\_firstname หมายถึง ชื่อผู้ป่วย
3. pta\_lastname หมายถึง นามสกุลผู้ป่วย
4. pta\_address\_number หมายถึง ที่อยู่บ้านเลขที่
5. pta\_email หมายถึง อีเมลผู้ป่วย
6. pta\_phone หมายถึง เบอร์โทรผู้ป่วย
7. pta\_img หมายถึง รูปของผู้ป่วย
8. province\_code หมายถึง รหัสจังหวัด
9. amphur\_code หมายถึง รหัส อำเภอ
10. district\_code หมายถึง รหัสตำบล
11. zipcode หมายถึง รหัสไปรษณีย์
12. ptd\_id หมายถึง รหัสอ้างอิงข้อมูลผู้ป่วย
13. crt\_date หมายถึง วันที่สร้างข้อมูล



**PUT** /patient/patient\_address/{ptd\_id} Update Patient Data

**Parameters**

Name	Description
<b>ptd_id</b> * required integer (path)	<input type="text" value="10"/>

**Request body** required

```
{
  "pta_idcard": "10101010",
  "pta_firstname": "มานะ",
  "pta_lastname": "ใจดี",
  "pta_address_number": "178",
  "pta_email": "siwakorn167@email.com",
  "pta_phone": "09439080477",
  "pta_img": "https://i.imgur.com/IXh6YEV.jpg",
  "province_code": 330101,
  "amphur_code": 0,
  "district_code": 0,
  "zipcode": 33000,
  "ptd_id": 10,
  "crt_date": "2022-06-22T12:44:04.590000+00:00"
}
```

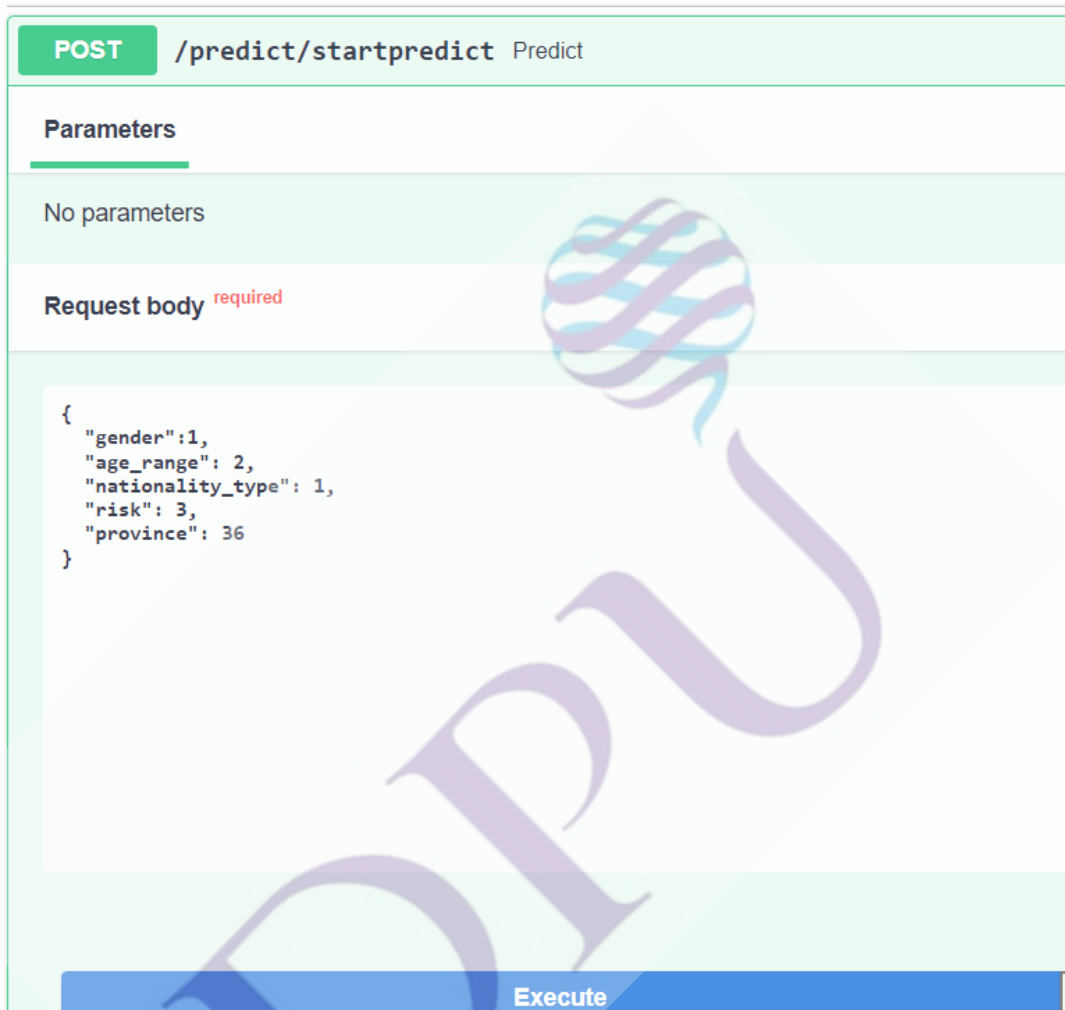
**Execute**

ภาพที่ ก.20 การแก้ไขข้อมูลที่อยู่ผู้ป่วย

จากภาพที่ ก.20 เมื่อเพิ่มข้อมูลได้แล้วสามารถแก้ไขข้อมูลได้โดยจะใช้ Method PUT (/patient/patient\_address)

**3. Predict** คือหัวใจสำคัญของการทำงานคือ การจำแนกข้อมูลโดยทำงานร่วมกับแบบจำลองในการทำนาย โดยจะใช้ Method POST (/predict/startpredict) ดังภาพที่ ก.21

## Predict



**POST** /predict/startpredict Predict

**Parameters**

No parameters

**Request body** required

```
{
  "gender": 1,
  "age_range": 2,
  "nationality_type": 1,
  "risk": 3,
  "province": 36
}
```

Execute

ภาพที่ ก.21 การจำแนกข้อมูล

จากภาพที่ ก.21 สามารถอธิบายรายละเอียดของข้อมูลได้ดังนี้

1. gender หมายถึง รหัสเพศ
2. age\_range หมายถึง ช่วงอายุ
3. nationality\_type หมายถึง ประเภทสัญชาติ
4. risk หมายถึง รหัสความเสี่ยง
5. province รหัสจังหวัด

โดยผลลัพธ์ของการจำแนกจะแสดงผลดังรูปที่ ก.22

Response body

```
{
  "result": "การเฝ้าระวังในกลุ่มเป้าหมายเฉพาะหรือพื้นที่เฉพาะ (sentinel surveillance)",
  "type": 3
}
```

ภาพที่ ก.22 ผลลัพธ์ในการจำแนกข้อมูล

4. **MasterData** เป็นส่วนแสดงผลข้อมูลตั้งต้น เพื่อนำไปใช้กับแบบฟอร์มต่างๆ ดังภาพที่ ก.23

MasterData	
GET	/master_data/agerange Read Agerange
GET	/master_data/patient_risk Read Patient Risk
GET	/master_data/patient_type Read Patient Type

ภาพที่ ก.23 MasterData

ข้อมูลช่วงอายุโดยใช้ Method GET (/master\_data/agerange) จะแสดงดังภาพที่ ก.24

Response body

```
[
  {
    "ar_id": 1,
    "ar_name": "< 10 ปี"
  },
  {
    "ar_id": 2,
    "ar_name": "10 ถึง 19 ปี"
  },
  {
    "ar_id": 3,
    "ar_name": "20 ถึง 29 ปี"
  },
  {
    "ar_id": 4,
    "ar_name": "30 ถึง 39 ปี"
  },
  {
    "ar_id": 5,
    "ar_name": "40 ถึง 49 ปี"
  },
  {

```

ภาพที่ ก.24 การแสดงข้อมูลช่วงอายุ

ข้อมูลประเภทความเสี่ยงโดยใช้ Method GET (/master\_data/patient\_risk) จะแสดงดังภาพที่ ก.25

```

Response body
[
  {
    "ptr_id": 1,
    "ptr_name": "ACF"
  },
  {
    "ptr_id": 2,
    "ptr_name": "Cluster"
  },
  {
    "ptr_id": 3,
    "ptr_name": "พื้นที่เสี่ยง"
  },
  {
    "ptr_id": 4,
    "ptr_name": "State Quarantine"
  },
  {
    "ptr_id": 5,
    "ptr_name": "Organizational Quarantine"
  },
  {
    "ptr_id": 6,
    "ptr_name": "ตรวจก่อนทำหัตถการ"
  },
  {
    "ptr_id": 7,
  }
]

```

ภาพที่ ก.25 การแสดงข้อมูลประเภทความเสี่ยง

ข้อมูลประเภทผู้ป่วยโดยใช้ Method GET (/master\_data/patient\_type) จะแสดงดังภาพที่ ก.26

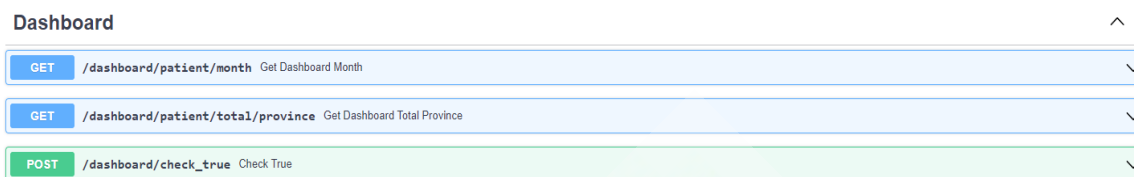
```

Response body
[
  {
    "ptt_id": 1,
    "ptt_name": "การเฝ้าระวังในกลุ่มผู้ป่วย หรือผู้มีอาการเข้าได้กับนิยามผู้สงสัยติดเชื้อไวรัสโคโรนา 2019 ที่เข้าเกณฑ์สอบสวนโรค (patient under investigation : PUI)"
  },
  {
    "ptr_id": 2,
    "ptr_name": "การตรวจคัดกรองในประชากรเสี่ยงตามจุดคัดกรองและด่านเข้าออกระหว่างประเทศ(screening)"
  },
  {
    "ptr_id": 3,
    "ptr_name": "การเฝ้าระวังในกลุ่มเป้าหมายเฉพาะหรือพื้นที่เฉพาะ (sentinel surveillance)"
  },
  {
    "ptr_id": 4,
    "ptr_name": "การเฝ้าระวังเหตุการณ์ในสถานที่เสี่ยง เก็บตัวอย่างส่งตรวจเมื่อเข้านิยาม PUI หรือเป็นกลุ่มก้อนของผู้ป่วยทางเดินหายใจ"
  }
]

```

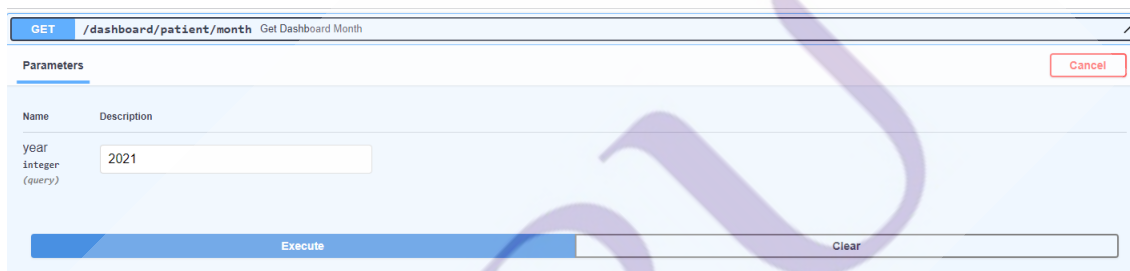
ภาพที่ ก.26 การแสดงข้อมูลประเภทผู้ป่วย

## 5. Dashboard เป็นส่วนการทำงานในการส่งข้อมูลทางสถิติโดยจะแสดงดังภาพที่ ก.27



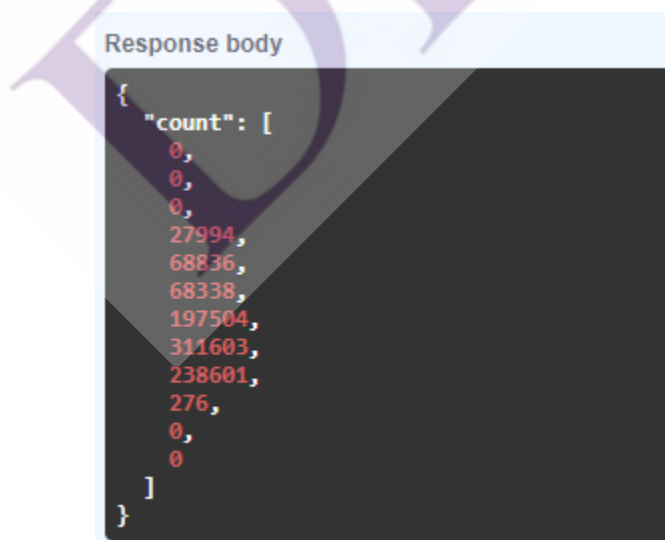
ภาพที่ ก.27 Dashboard

ข้อมูลยอดผู้ป่วยรายเดือน โดยใช้ Method GET (/dashboard/patient/month?year=2021) จะแสดงดังภาพที่ ก.28



ภาพที่ ก.28 การกำหนดปีแสดงข้อมูลผู้ป่วยรายเดือน


จากภาพที่ ก.28 หลังจากกำหนดปีแล้ว จะแสดงข้อมูลดังภาพที่ ก.29



ภาพที่ ก.29 จำนวนยอดผู้ป่วยรายเดือนต่อปี



ข้อมูลยอดผู้ป่วยโดยแยกตามจังหวัด โดยใช้ Method GET (/dashboard/patient/total/province) จะแสดงดังภาพที่ ก.30



```
Details
Response body
{
  "data": [
    [
      "th-bm",
      223844
    ],
    [
      "th-sp",
      73391
    ],
    [
      "th-no",
      47631
    ],
    [
      "th-pt",
      31957
    ],
    [
      "th-pa",
      24514
    ],
    [
      "th-at",
      9562
    ],
    [
      "th-lb",
```

ภาพที่ ก.30 จำนวนยอดผู้ป่วยโดยแยกตามจังหวัดต่อปี

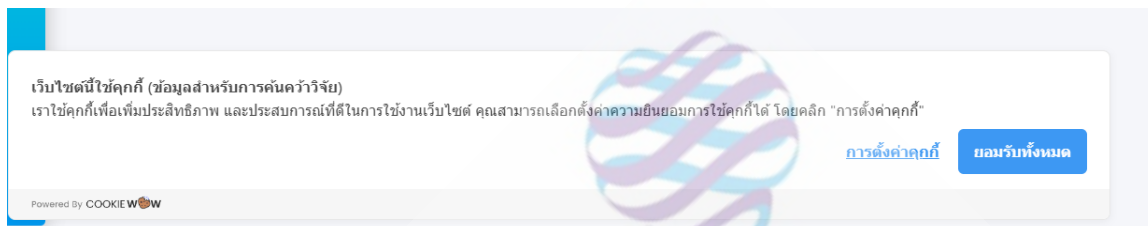


ภาคผนวก ข.

ส่วนติดต่อผู้ใช้งาน

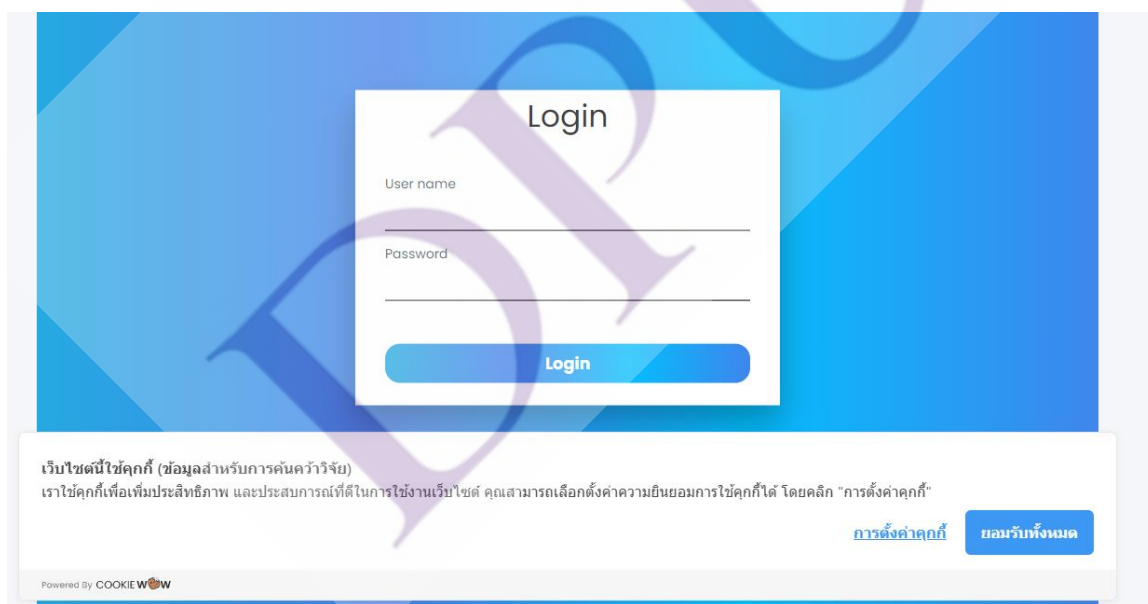
DPU

**1. Cookies Banner** การขอความยินยอมการใช้คุกกี้กับเจ้าของข้อมูลส่วนบุคคลตาม PDPA ผ่านแบนเนอร์ (banner) เป็นวิธีการที่สะดวกและได้รับความนิยม เนื่องจากเป็นวิธีที่ไม่ซับซ้อนและถือเป็นการแจ้งให้ผู้ใช้งานเว็บไซต์ในฐานะเจ้าของข้อมูลส่วนบุคคลเห็นได้ชัดเจน ดังภาพที่ ข.1



ภาพที่ ข. 1 Cookies Banner

## 2. หน้าจอเข้าสู่ระบบ



ภาพที่ ข. 2 การเข้าสู่ระบบ

### 3. แบบฟอร์มวิเคราะห์ข้อมูลผู้ป่วย

ภาพที่ ข. 3 แบบฟอร์มวิเคราะห์ข้อมูลผู้ป่วย

จากภาพที่ ข.3 สามารถอธิบายตามหมายเลขได้ดังนี้

1. แบบฟอร์มวิเคราะห์ผู้ป่วยเพื่อการจำแนกข้อมูล เช่น วันที่เริ่มป่วย เพศ อายุ สัญชาติ ความเสี่ยง และจังหวัด
2. ผลการจำแนก
3. ปุ่มบันทึกผู้ป่วย เมื่อกดปุ่มแล้วระบบจะพาไปยังขั้นตอนกรอกที่อยู่และรายละเอียดผู้ป่วยเพิ่มเติม ดังภาพที่ ข.4

ภาพที่ ข. 4 แบบฟอร์มบันทึกข้อมูลผู้ป่วย

## 4. ข้อมูลผู้ป่วย

ข้อมูลผู้ป่วยใหม่

ระบบคัดแยกผู้ป่วยโควิด-19

หน้าปัจจุบัน: 1 | ข้อมูลทั้งหมด: 5

วันที่บันทึก/เริ่มป่วย	เพศ	อายุ	ช่วงอายุ	สัญชาติ	ประเภท	จังหวัด	ความเสี่ยง	วันที่รับแจ้ง	จัดการ
2022-07-06	ชาย	30	30 ถึง 39 ปี	ไทย	คนไทย	กรุงเทพมหานคร	ปลอดภัยสูง	2022-07-06	
2022-07-05	ชาย	70	>= 70 ปี	ไทย	คนไทย	กรุงเทพมหานคร	พื้นที่เสี่ยง	2022-07-06	
2022-07-05	ชาย	70	>= 70 ปี	ไทย	คนไทย	กรุงเทพมหานคร	พื้นที่เสี่ยง	2022-07-06	
2022-07-01	ชาย	35	30 ถึง 39 ปี	ไทย	คนไทย	สมุทรปราการ	พื้นที่เสี่ยง	2022-07-01	
2022-06-27	ชาย	45	40 ถึง 49 ปี	ไทย	คนไทย	กรุงเทพมหานคร	Cluster	2022-06-29	

หน้าปัจจุบัน: 1 | ข้อมูลทั้งหมด: 5

เว็บไซต์นี้ใช้คุกกี้ (ข้อมูลสำหรับการค้นคว้าวิจัย)  
เราใช้คุกกี้เพื่อเพิ่มประสิทธิภาพ และประสบการณ์ที่ดีในการใช้งานเว็บไซต์ คุณสามารถเลือกตั้งค่าความยินยอมการใช้คุกกี้ได้ โดยคลิก "การตั้งค่าคุกกี้"

[การตั้งค่าคุกกี้](#) [ยอมรับทั้งหมด](#)

Powered By COOKE W@W

ภาพที่ ข. 5 ข้อมูลผู้ป่วยใหม่

ระบบคัดแยกผู้ป่วยโควิด-19

หน้าปัจจุบัน: 1 | ข้อมูลทั้งหมด: 913152

วันที่บันทึก/เริ่มป่วย	เพศ	อายุ	ช่วงอายุ	สัญชาติ	ประเภท	จังหวัด	ความเสี่ยง	วันที่รับแจ้ง	จัดการ
2021-10-01	ชาย	16	10 ถึง 19 ปี	ไทย	คนไทย	นครสวรรค์	สัมผัสผู้ติดเชื้อ	2021-09-01	
2021-10-01	ชาย	43	40 ถึง 49 ปี	ไทย	คนไทย	ปราจีนบุรี	พื้นที่เสี่ยง	2021-09-01	
2021-10-01	หญิง	20	20 ถึง 29 ปี	ไทย	คนไทย	นครสวรรค์	สัมผัสผู้ติดเชื้อ	2021-09-01	
2021-10-01	ชาย	42	40 ถึง 49 ปี	ไทย	คนไทย	นครสวรรค์	สัมผัสผู้ติดเชื้อ	2021-09-01	
2021-10-01	ชาย	23	20 ถึง 29 ปี	ไทย	คนไทย	ปราจีนบุรี	พื้นที่เสี่ยง	2021-09-01	
2021-10-01	หญิง	35	30 ถึง 39 ปี	ไทย	คนไทย	ปราจีนบุรี	พื้นที่เสี่ยง	2021-09-01	
2021-10-01	ชาย	12	10 ถึง 19 ปี	ไทย	คนไทย	นครสวรรค์	สัมผัสผู้ติดเชื้อ	2021-09-01	
2021-10-01	หญิง	2	< 10 ปี	ไทย	คนไทย	นครสวรรค์	สัมผัสผู้ติดเชื้อ	2021-09-01	
2021-10-01	ชาย	12	10 ถึง 19 ปี	ไทย	คนไทย	นครสวรรค์	สัมผัสผู้ติดเชื้อ	2021-09-01	
2021-10-01	ชาย	63	60 ถึง 69 ปี	ไทย	คนไทย	นครสวรรค์	สัมผัสผู้ติดเชื้อ	2021-09-01	

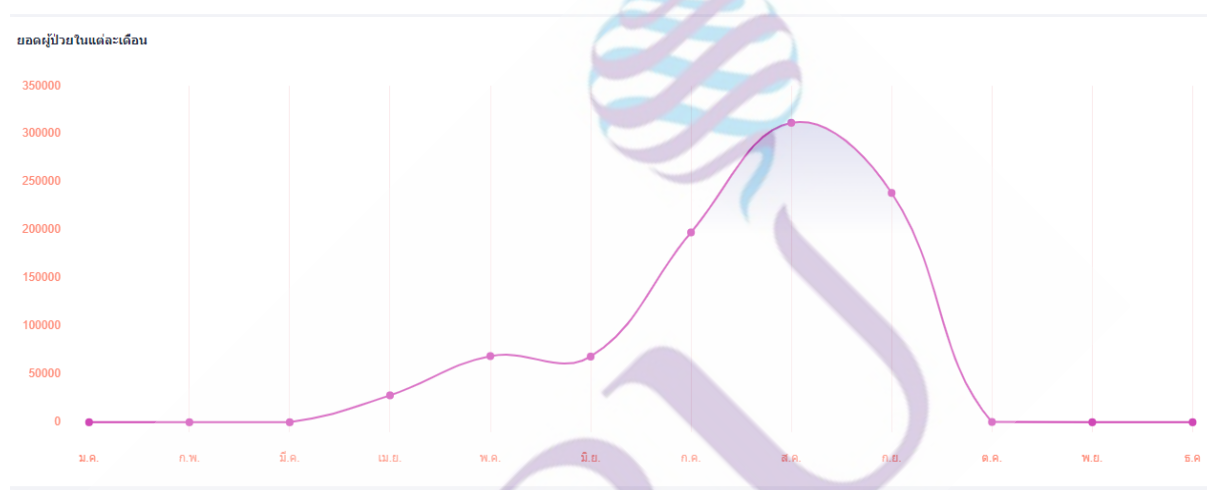
ภาพที่ ข. 6 ข้อมูลผู้ป่วยเดิม

จากภาพที่ ข.5 - ข.6 จะเป็นส่วนตารางข้อมูลของผู้ป่วย โดยจะมี 2 ส่วนในการเก็บคือ ข้อมูลผู้ป่วยใหม่ และ ข้อมูลผู้ป่วยเดิม สามารถอธิบายได้ดังนี้

1. ข้อมูลผู้ป่วยใหม่ คือ ข้อมูลที่บันทึกใหม่ ไม่ได้รวบรวมจากกรมควบคุมโรค
2. ข้อมูลผู้ป่วยเดิม คือ ข้อมูลที่รวบรวมมาจากกรมควบคุมโรค

## 5. แดชบอร์ด

กราฟเส้นแสดงแนวโน้มการติดเชื้อของผู้ป่วยในแต่ละเดือน



ภาพที่ ข. 7 กราฟแสดงยอดผู้ป่วยในแต่ละเดือน

สำรวจข้อมูล

เพศ: ชาย | ช่วงอายุ: 10 ถึง 19 ปี | สัญชาติ: ไทย

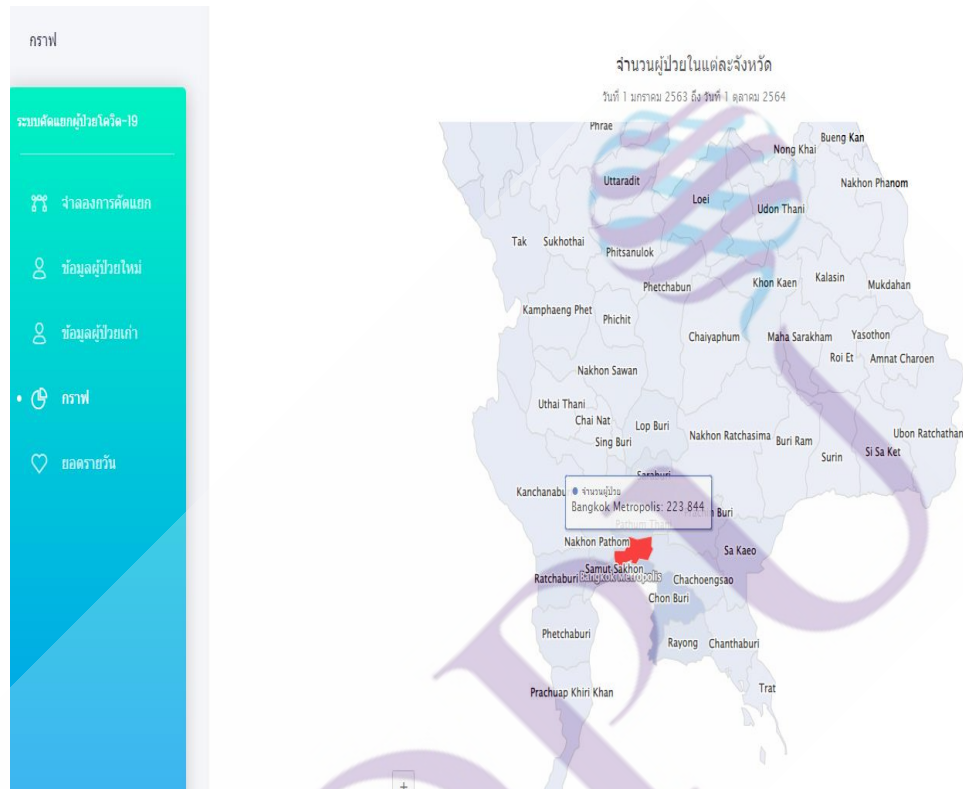
ความเสี่ยง: พื้นที่เสี่ยง | จังหวัด: กรุงเทพมหานคร

[ค้นหา](#)

CLASS	TOTAL
การเฝ้าระวังในกลุ่มผู้ป่วย หรือผู้เฝ้าระวังการเข้าถึงกับนิยามผู้สงสัยติดเชื้อไวรัสโคโรนา 2019 ที่เข้าเกณฑ์สอบสวนโรค (patient under investigation : PUI)	21
การตรวจคัดกรองในประชากรเสี่ยงตามจุดคัดกรองและด่านเข้าออกระหว่างประเทศ(screening)	0
การเฝ้าระวังในกลุ่มเป้าหมายเฉพาะหรือพื้นที่เฉพาะ (sentinel surveillance)	142
การเฝ้าระวังเหตุการณ์ในสถานที่เสี่ยง เก็บตัวอย่างส่งตรวจเมื่อเข้านิยาม PUI หรือเป็นกลุ่มก่อนของผู้ป่วยทางเดินหายใจ	186

ภาพที่ ข. 8 จำนวน Instance ในแต่ละคลาส

จากภาพที่ ข.8 เป็นการสำรวจจำนวนข้อมูลโดยมีเงื่อนไขตามแบบฟอร์ม เช่น เพศ ช่วงอายุ สัญชาติ ความเสี่ยง จังหวัด เพื่อดูกลาสข้อมูลของประเภทผู้ป่วยว่ามีจำนวนเท่าไร



ภาพที่ ข. 9 พื้นที่การติดเชื้อ

ส่วนเสริมของรายงานยอดผู้ติดเชื้อรายวัน โดยจะแสดงผลข้อมูล เช่น จำนวนผู้ติดเชื้อวันนี้ จำนวนผู้ติดเชื้อทั้งหมด จำนวนผู้เสียชีวิตวันนี้ จำนวนผู้ติดเชื้อทั้งหมด รักษาหายวันนี้ และ รักษาหายแล้วทั้งหมด โดยข้อมูลจะ ได้มาจาก <https://static.easysunday.com/covid-19/getTodayCases.json> ซึ่งเป็นชุดข้อมูลแบบเปิดมีให้บริการข้อมูลผู้ติดเชื้อ covid-19 ทั่วโลก ซึ่งข้อมูลมีการอัปเดตทุกวันดังภาพที่

ข.10

ข้อมูลเมื่อ 06/07/2022 18:09	
ข้อมูลจาก Sunday Thailand	
จำนวนผู้ติดเชื้อวันนี้	จำนวนผู้ติดเชื้อทั้งหมด
2,428 ราย	4,536,445 ราย
จำนวนผู้เสียชีวิตวันนี้	จำนวนผู้เสียชีวิตทั้งหมด
22 ราย	30,761 ราย
รักษาหายแล้ววันนี้	รักษาหายแล้วทั้งหมด
2,049 ราย	4,480,895 ราย

ภาพที่ ข. 10 ยอดผู้ติดเชื้อ โควิด-19 รายวัน



## ประวัติผู้เขียน

ชื่อ-นามสกุล	ศิวกร บรรลือทรัพย์
ประวัติการศึกษา	วิศวกรรมศาสตรบัณฑิต สาขาวิศวกรรมคอมพิวเตอร์ ระดับปริญญาตรี มหาวิทยาลัยกาฬสินธุ์ ปี 2559
ตำแหน่งและสถานที่ทำงานปัจจุบัน	นักพัฒนาซอฟต์แวร์ บริษัท ไอดีไครฟ์
ประสบการณ์ ผลงานทางวิชาการ	The 18th National Conference on Computing and Information Technology A Comparison of Machine Learning Models for COVID-19 Patients Screening Journal of Information Science and Technology (JIST) Towards Machine Learning Algorithm for Screening Prediction of COVID-19 Patients