

การพยากรณ์จำนวนผู้ป่วย COVID-19 ประเทศไทยโดยการใช้ข้อมูลจากหลายประเทศ

นายสิริธีร์ วรธรรมทองดี

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต  
สาขาวิชาวิทยาศาสตร์คอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์  
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย  
ปีการศึกษา 2565  
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

2927129371  
CT IThesis 6272091321 thesis / rev: 25122565 00:06:46 / seq: 30

6272091321\_2927129371

PREDICTIVE ANALYSIS OF COVID-19 PATIENTS IN THAILAND USING MULTIPLE  
COUNTRIES DATA

Mr. Siratee Vorathamthongdee

A Thesis Submitted in Partial Fulfillment of the Requirements  
for the Degree of Master of Science in Computer Science  
Department of Computer Engineering  
FACULTY OF ENGINEERING  
Chulalongkorn University  
Academic Year 2022  
Copyright of Chulalongkorn University



2927129371

CU Thesisis 6272091321 thesis / recv: 25122565 00:06:46 / seq: 30

หัวข้อวิทยานิพนธ์	การพยากรณ์จำนวนผู้ป่วย COVID-19 ประเทศไทยโดยการใช้ข้อมูลจากหลายประเทศ
โดย	นายสิริธีร์ วรรณมทองดี
สาขาวิชา	วิทยาศาสตร์คอมพิวเตอร์
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	ศาสตราจารย์ ดร.ประภาส จงสฤษดิ์วัฒนา

---

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้รับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่ง  
ของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

..... คณบดีคณะวิศวกรรมศาสตร์  
(ศาสตราจารย์ ดร.สุพจน์ เตชวรสินสกุล)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ  
(ผู้ช่วยศาสตราจารย์ ดร.สุกรี สินธุภิญโญ)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก  
(ศาสตราจารย์ ดร.ประภาส จงสฤษดิ์วัฒนา)

..... กรรมการภายนอกมหาวิทยาลัย  
(ผู้ช่วยศาสตราจารย์ ดร.ธนัสินี เพ็ชรตระกูล)

สิริธีร์ วรรณทองดี : การพยากรณ์จำนวนผู้ป่วย COVID-19 ประเทศไทยโดยใช้  
ข้อมูลจากหลายประเทศ. ( PREDICTIVE ANALYSIS OF COVID-19 PATIENTS IN  
THAILAND USING MULTIPLE COUNTRIES DATA) อ.ที่ปรึกษาหลัก : ศ. ดร.ประภาส  
จงสถิตย์วัฒนา

สถานการณ์โรคระบาดโควิด-19 ได้แพร่กระจายไปทั่วโลกตั้งแต่ปีพ.ศ.2562 การศึกษา  
ได้ทำการพยากรณ์ผู้ป่วยจากโรคโควิด-19ของประเทศไทย โดยใช้ข้อมูลตั้งแต่วันที่  
22 มกราคม พ.ศ.2563 จนถึงวันที่ 31 ธันวาคม พ.ศ.2564 ซึ่งมีจำนวนข้อมูลอยู่ 710 วัน โดย  
ข้อมูลสาธารณะจากมหาวิทยาลัยจอนส์ ฮอปคินส์ ผู้วิจัยได้ใช้วิธีการเรียนรู้ของเครื่อง (Machine  
Learning) ในการพยากรณ์จำนวนผู้ป่วยในประเทศเพื่อเป็นหนึ่งในข้อมูลในการช่วยรัฐบาลออก  
นโยบายการจัดการทรัพยากรเพื่อรับมือกับโรคระบาด การศึกษาได้ใช้วิธีการจัดกลุ่มเคมีน (K-  
Means) ในการจัดกลุ่มของประเทศที่มีรูปแบบของจำนวนผู้ป่วยจากโรคคล้ายกันกับประเทศไทย  
ผลจากการจัดกลุ่มมีประเทศที่อยู่ใน Cluster เดียวกันกับประเทศไทยทั้งหมด 8 ประเทศ ได้แก่  
ญี่ปุ่น (Japan), มาเลเซีย (Malaysia), ฟิลิปปินส์ (Philippines) ,บังกลาเทศ (Bangladesh) ,คิวบา  
(Cuba) ,อิรัก (Iraq) ,เม็กซิโก (Mexico) และ เวียดนาม (Vietnam) อยู่ในกลุ่มเดียวกับประเทศไทย  
จากนั้นทำการจับคู่ระหว่างประเทศไทยและประเทศที่อยู่ในกลุ่มเดียวกัน ใช้โมเดลหน่วยความจำ  
ระยะสั้นยาว (Long Short-Term Memory) เพื่อพยากรณ์จำนวนผู้ป่วยโควิด-19 ของประเทศไทย  
ผลจากโมเดลแสดงได้ว่าการใช้ข้อมูลคู่ประเทศไทยและบังกลาเทศ ญี่ปุ่น และเม็กซิโกมีค่าเฉลี่ยของ  
ร้อยละความผิดพลาดสัมบูรณ์ (Mean Absolute Percentage Error) น้อยที่สุดตามลำดับ เมื่อ  
เทียบกับการใช้ข้อมูลแค่ประเทศไทยอย่างเดียว

สาขาวิชา วิทยาศาสตร์คอมพิวเตอร์  
ปีการศึกษา 2565

ลายมือชื่อนิสิต .....  
ลายมือชื่อ อ.ที่ปรึกษาหลัก .....

# # 6272091321 : MAJOR COMPUTER SCIENCE

KEYWORD: COVID-19, Machine Learning, K-Means, LSTM

Siratee Vorathamthongdee : PREDICTIVE ANALYSIS OF COVID-19 PATIENTS  
IN THAILAND USING MULTIPLE COUNTRIES DATA. Advisor: Prof. PRABHAS  
CHONGSTITVATANA, Ph.D.

COVID-19 is a situation that has spread worldwide since 2019. This study forecast the number of patients with COVID-19 in Thailand. Using data between January 22, 2020, and December 31, 2021, we collect confirmed cases from John Hopskin Open Data. The total number of the data points are 710. Using the machine learning model to predict the number of patient cases in the country helps the government manage its policies and resources. In this study, the K-Means clustering algorithm is performed to group the countries that have similar patterns of confirmed cases to Thailand. Clustering results show that Japan, Malaysia, the Philippines, Bangladesh, Cuba, Iraq, Mexico, and Vietnam are all in the same cluster as Thailand. Using the Long Short-Term Memory (LSTM) to predict the confirmed case of Thailand by feeding the model pairs of countries with Thailand that are in the same cluster. The performance of LSTM shows that using pairs of countries with Thailand between Bangladesh, Japan, and Mexico has the lowest error on MAPE, respectively, when compared to using only Thailand data.

Field of Study: Computer Science

Student's Signature .....

Academic Year: 2022

Advisor's Signature .....

## กิตติกรรมประกาศ

ผู้วิจัยขอกราบขอบพระคุณอาจารย์ที่ปรึกษา ศาสตราจารย์ ดร.ประภาส จงสถิตย์วัฒนา ด้วยความช่วยเหลือของอาจารย์วิทยานิพนธ์ชั้นนี้จึงสำเร็จไปได้ด้วยดี ท่านได้ให้คำแนะนำและชี้แนะแนวทางการทำวิทยานิพนธ์ รวมถึงอุปสรรคต่างๆซึ่งได้บรรลุได้ดี ขอขอบพระคุณคณะกรรมการ สอบวิทยานิพนธ์ผู้ช่วยศาสตราจารย์ ดร.สุกรี สิ้นธุภิญโญ และผู้ช่วยศาสตราจารย์ ดร. ธนัสินี เพียรตระกูล ในการให้คำแนะนำสำหรับการทำวิทยานิพนธ์

สุดท้ายนี้ขอขอบพระคุณบิดามารดา ซึ่งเป็นกำลังใจและสนับสนุนการศึกษาปริญญาโท และขอบคุณตัวเองในการดำเนินการจนได้บรรลุผลสำเร็จจบการศึกษา

สิริธีร์ วรธรรมทองดี

## สารบัญ

	หน้า
.....	ค
บทคัดย่อภาษาไทย.....	ค
.....	ง
บทคัดย่อภาษาอังกฤษ.....	ง
กิตติกรรมประกาศ.....	จ
สารบัญ.....	ฉ
สารบัญตาราง.....	ญ
สารบัญรูปภาพ.....	ท
บทที่ 1 บทนำ.....	1
1.1 ความสำคัญและที่มาของงานวิจัย.....	1
1.2 คำถามทางงานวิจัย .....	2
1.3 วัตถุประสงค์ของงานวิจัย .....	2
1.4 ขอบเขตของงานวิจัย.....	2
1.5 ประโยชน์ที่คาดว่าจะได้รับ.....	3
1.6 ขั้นตอนศึกษาวิจัย .....	3
บทที่ 2 ทฤษฎีที่เกี่ยวข้อง.....	4
2.1 การเรียนรู้ของเครื่อง Machine Learning.....	4
2.2 การเรียนรู้เชิงลึก Deep Learning.....	4
2.3 K-Mean Clustering (K-means).....	4
2.4 Long Short-Term Memory (LSTM).....	6
2.1.4 Adam Optimizer .....	8

2.1.5 Reduce LRO Plateau ..... 9

2.1.6 Normalization ..... 10

2.1.7 Cross-correlation ..... 10

2.1.8 Evaluation Method ..... 11

บทที่ 3 งานวิจัยที่เกี่ยวข้อง ..... 12

บทที่ 4 วิธีการดำเนินการวิจัย ..... 15

4.1 ภาพรวมของข้อมูล ..... 15

4.2 เครื่องมือที่ใช้ในการวิเคราะห์งานวิจัย (Analysis Tools)..... 16

4.3 การเตรียมการของข้อมูล (Pre-processing)..... 17

4.3.1 การเปลี่ยนรูปแบบของข้อมูลจากรูปแบบ Wide Format ไปยัง รูปแบบ Long Format ..... 17

4.3.2 การเลือกข้อมูล (Data Selection)..... 18

4.3.3 การรวมข้อมูล (Data Aggregation)..... 18

4.3.4 การเปลี่ยนรูปแบบของข้อมูลจากรูปแบบ Long Format ไปยัง รูปแบบ Wide Format ..... 20

4.3.5 การแปลงข้อมูลจากข้อมูลยืนยันผู้ป่วยแบบสะสม (Cumulative Data) เป็นข้อมูลที่เพิ่มขึ้นตามจริง (Real Data)..... 21

4.3.6 การทำ Normalization ..... 22

4.4 การแบ่งกลุ่มข้อมูลประเทศ (Clustering) โดยใช้ K-Mean Clustering ..... 23

4.5 การปรับวางแนว (Alignment) ของผู้ป่วยยืนยันระหว่างคู่ประเทศไทยและประเทศในกลุ่มเดียวกัน ..... 23

4.6 การทำนายผู้ป่วย COVID-19 ด้วยวิธีการ Long short-term memory (LSTM) ..... 24

4.6.1 การทำ Normalization ..... 24

4.6.2 การรวมข้อมูล (Data Aggregation)..... 24



4.6.3 การแปลงข้อมูลจากข้อมูลยืนยันผู้ป่วยแบบสะสม (Cumulative Data) เป็นข้อมูลที่เพิ่มขึ้นตามจริง (Real Data)..... 24

4.6.4 การทำการเตรียมข้อมูลให้เหมาะกับการทำ Multiple Lag Timesteps ..... 25

4.6.5 การแบ่งข้อมูลเรียนรู้ (Training Set) และข้อมูลทดสอบ (Testing Set)..... 25

4.6.6 กำหนดโครงสร้าง LSTM ..... 25

4.6.7 Training Model LSTM..... 26

4.6.8 วัดผลการจาก Model (Evaluating Network)..... 26

    ทำการวัดผลจาก LSTM Model ด้วยการ ใช้ RMSE, MAE และ MAPE ..... 26

บทที่ 5 ผลการทดลองและอภิปรายผลการทดลอง ..... 27

5.1 การแบ่งกลุ่มข้อมูลประเทศ (Clustering) โดยใช้ K-Mean Clustering ..... 27

5.2 การปรับวางแนว (Alignment) ของผู้ป่วยยืนยันระหว่างคู่ประเทศไทยและประเทศในกลุ่มเดียวกัน ..... 31

    5.2.1 ประเทศไทยและคู่ประเทศญี่ปุ่น ..... 31

    5.2.2 ประเทศไทยและคู่ประเทศมาเลเซีย ..... 32

    5.2.3 ประเทศไทยและคู่ประเทศฟิลิปปินส์ ..... 34

    5.2.4 ประเทศไทยและคู่ประเทศบังกลาเทศ ..... 35

    5.2.5 ประเทศไทยและคู่ประเทศคิวบา..... 36

    5.2.6 ประเทศไทยและคู่ประเทศอิรัก..... 38

    5.2.7 ประเทศไทยและคู่ประเทศเม็กซิโก ..... 39

    5.2.8 ประเทศไทยและคู่ประเทศเวียดนาม..... 40

5.3 ผลจากการทำนายผู้ป่วยยืนยัน COVID-19 จาก Model LSTM..... 42

    5.3.1 ประเทศไทย..... 42

    5.3.2 ประเทศไทยและคู่ประเทศญี่ปุ่น ..... 44

    5.3.3 ประเทศไทยและคู่ประเทศมาเลเซีย ..... 47

5.3.4 ประเทศไทยและคู่ประเทศฟิลิปปินส์ .....	50
5.3.5 ประเทศไทยและคู่ประเทศบังกลาเทศ .....	53
5.3.6 ประเทศไทยและคู่ประเทศคิวบา.....	56
5.3.7 ประเทศไทยและคู่ประเทศอิรัก.....	59
5.3.8 ประเทศไทยและคู่ประเทศเม็กซิโก .....	62
5.3.9 ประเทศไทยและคู่ประเทศเวียดนาม.....	65
5.4 แนวทางวิจัยในอนาคต.....	69
บรรณานุกรม.....	2
ประวัติผู้เขียน.....	6



2927129371

## สารบัญตาราง

	หน้า
ตารางที่ 1 แสดงรายละเอียดแต่ละคอลัมน์และความหมาย.....	16
ตารางที่ 2 Inertia โดยเทียบ %ความแตกต่างของ Step ก่อนหน้า ตั้งแต่จำนวน Cluster 2 ถึง 50 .....	29
ตารางที่ 3 แสดงคู่ประเทศไทยในค่า Correlation ก่อนและหลังการทำ Pre-processing และทำ Alignment และค่า Lag Length.....	42
ตารางที่ 4 แสดง Metric จากการทำนาย LSTM โดยใช้ข้อมูลประเทศไทย.....	43
ตารางที่ 5 แสดง Metric จากการทำนาย LSTM โดยใช้ข้อมูลประเทศไทยและคู่ประเทศญี่ปุ่น .....	45
ตารางที่ 6 แสดงภาพรวมสถิติ RMSE ผลการทำนายระหว่างประเทศไทยและคู่ประเทศไทยญี่ปุ่น..	45
ตารางที่ 7 ผลจากทดสอบ T-Test .....	45
ตารางที่ 8 แสดงภาพรวมสถิติ MAE ผลการทำนายระหว่างประเทศไทยและคู่ประเทศไทยญี่ปุ่น ....	46
ตารางที่ 9 ผลจากทดสอบ T-Test .....	46
ตารางที่ 10 แสดงภาพรวมสถิติ MAPE ผลการทำนายระหว่างประเทศไทยและคู่ประเทศไทยญี่ปุ่น	46
ตารางที่ 11 ผลจากทดสอบ T-Test.....	46
ตารางที่ 12 แสดง Metric จากการทำนาย LSTM โดยใช้ข้อมูลประเทศไทยและคู่ประเทศมาเลเซีย	48
ตารางที่ 13 แสดงภาพรวมสถิติ RMSE ผลการทำนายระหว่างประเทศไทยและคู่ประเทศไทยมาเลเซีย .....	48
ตารางที่ 14 ผลจากทดสอบ T-Test.....	48
ตารางที่ 15 แสดงภาพรวมสถิติ MAE ผลการทำนายระหว่างประเทศไทยและคู่ประเทศไทยมาเลเซีย .....	49
ตารางที่ 16 ผลจากทดสอบ T-Test.....	49
ตารางที่ 17 แสดงภาพรวมสถิติ MAPE ผลการทำนายระหว่างประเทศไทยและคู่ประเทศไทย มาเลเซีย.....	49

ตารางที่ 18 ผลจากทดสอบ T-Test..... 49

ตารางที่ 19 แสดง Metric จากการทำนาย LSTM โดยใช้ข้อมูลประเทศไทยและคู่ประเทศฟิลิปปินส์  
..... 51

ตารางที่ 20 แสดงภาพรวมสถิติ RMSE แสดงภาพรวมสถิติ RMSE ผลการทำนายระหว่างประเทศไทย  
และคู่ประเทศไทยฟิลิปปินส์..... 51

ตารางที่ 21 ผลจากทดสอบ T-Test..... 51

ตารางที่ 22 แสดงภาพรวมสถิติ MAE ผลการทำนายระหว่างประเทศไทยและคู่ประเทศไทยฟิลิปปินส์  
..... 52

ตารางที่ 23 ผลจากทดสอบ T-Test..... 52

ตารางที่ 24 แสดงภาพรวมสถิติ MAPE ผลการทำนายระหว่างประเทศไทยและคู่ประเทศไทย  
ฟิลิปปินส์..... 52

ตารางที่ 25 ผลจากทดสอบ T-Test..... 52

ตารางที่ 26 แสดง Metric จากการทำนาย LSTM โดยใช้ข้อมูลประเทศไทยและคู่ประเทศบังกลาเทศ  
..... 54

ตารางที่ 27 แสดงภาพรวมสถิติ RMSE ผลการทำนายระหว่างประเทศไทยและคู่ประเทศไทย  
บังกลาเทศ ..... 54

ตารางที่ 28 ผลจากทดสอบ T-Test..... 54

ตารางที่ 29 แสดงภาพรวมสถิติ MAE แสดงภาพรวมสถิติ RMSE ผลการทำนายระหว่างประเทศไทย  
และคู่ประเทศไทยบังกลาเทศ..... 55

ตารางที่ 30 ผลจากทดสอบ T-Test..... 55

ตารางที่ 31 แสดงภาพรวมสถิติ MAPE ผลการทำนายระหว่างประเทศไทยและคู่ประเทศไทย  
บังกลาเทศ ..... 55

ตารางที่ 32 ผลจากทดสอบ T-Test..... 56

ตารางที่ 33 แสดง Metric จากการทำนาย LSTM โดยใช้ข้อมูลประเทศไทยและคู่ประเทศคิวบา.... 57

ตารางที่ 34 แสดงภาพรวมสถิติ RMSE ผลการทำนายระหว่างประเทศไทยและคู่ประเทศไทยคิวบา57

ตารางที่ 35 ผลจากทดสอบ T-Test..... 58

ตารางที่ 36 แสดงภาพรวมสถิติ MAE ผลการทำนายระหว่างประเทศไทยและคู่ประเทศไทยคิวบา . 58

ตารางที่ 37 ผลจากทดสอบ T-Test ..... 58

ตารางที่ 38 แสดงภาพรวมสถิติ MAPE ผลการทำนายระหว่างประเทศไทยและคู่ประเทศไทยคิวบา 58

ตารางที่ 39 ผลจากทดสอบ T-Test ..... 59

ตารางที่ 40 แสดง Metric จากการทำนาย LSTM โดยใช้ข้อมูลประเทศไทยและคู่ประเทศอิรัก ..... 60

ตารางที่ 41 แสดงภาพรวมสถิติ RMSE ผลการทำนายระหว่างประเทศไทยและคู่ประเทศไทยอิรัก . 60

ตารางที่ 42 ผลจากทดสอบ T-Test ..... 61

ตารางที่ 43 แสดงภาพรวมสถิติ MAE ผลการทำนายระหว่างประเทศไทยและคู่ประเทศไทยอิรัก ... 61

ตารางที่ 44 ผลจากทดสอบ T-Test ..... 61

ตารางที่ 45 แสดงภาพรวมสถิติ MAPE ผลการทำนายระหว่างประเทศไทยและคู่ประเทศไทยอิรัก . 61

ตารางที่ 46 ผลจากทดสอบ T-Test ..... 62

ตารางที่ 47 แสดง Metric จากการทำนาย LSTM โดยใช้ข้อมูลประเทศไทยและคู่ประเทศเม็กซิโก 63

ตารางที่ 48 แสดงภาพรวมสถิติ RMSE ผลการทำนายระหว่างประเทศไทยและคู่ประเทศไทยเม็กซิโก  
..... 63

ตารางที่ 49 ผลจากทดสอบ T-Test ..... 64

ตารางที่ 50 แสดงภาพรวมสถิติ MAE ผลการทำนายระหว่างประเทศไทยและคู่ประเทศไทยเม็กซิโก  
..... 64

ตารางที่ 51 ผลจากทดสอบ T-Test ..... 64

ตารางที่ 52 แสดงภาพรวมสถิติ MAPE ผลการทำนายระหว่างประเทศไทยและคู่ประเทศไทยเม็กซิโก  
..... 64

ตารางที่ 53 ผลจากทดสอบ T-Test ..... 65

ตารางที่ 54 แสดง Metric จากการทำนาย LSTM โดยใช้ข้อมูลประเทศไทยและคู่ประเทศเวียดนาม  
..... 66

ตารางที่ 55 แสดงภาพรวมสถิติ RMSE ผลการทำนายระหว่างประเทศไทยและคู่ประเทศไทย  
เวียดนาม ..... 66

ตารางที่ 56 ผลจากทดสอบ T-Test..... 67

ตารางที่ 57 แสดงภาพรวมสถิติ MAE ผลการทำนายระหว่างประเทศไทยและคู่ประเทศไทยเวียดนาม  
..... 67

ตารางที่ 58 ผลจากทดสอบ T-Test..... 67

ตารางที่ 59 แสดงภาพรวมสถิติ MAPE ผลการทำนายระหว่างประเทศไทยและคู่ประเทศไทย  
เวียดนาม..... 67

ตารางที่ 60 ผลจากทดสอบ T-Test..... 68

ตารางที่ 61 แสดงค่าเฉลี่ยของการทดลอง 30 ครั้งทุก Metric ทำนายผู้ป่วยประเทศไทย (\* หมายถึง  
แตกต่างกันอย่างมีนัยสำคัญเมื่อเทียบกับใช้ข้อมูลประเทศไทย)..... 68

## สารบัญรูปภาพ

	หน้า
รูปที่ 1 สมการ Euclidean Distance .....	5
รูปที่ 2 แสดงภาพ K-Mean.....	5
รูปที่ 3 แสดงสมการเลือก Centroid ของ K-Mean++ .....	5
รูปที่ 4 Elbow curve method .....	6
รูปที่ 5 โครงสร้าง Long Short-Term Memory .....	7
รูปที่ 6 แสดงสมการของ Adaptive Moment Estimation (Adam) Optimizer .....	9
รูปที่ 7 แสดงสมการ Max-Min Normalization.....	10
รูปที่ 8 แสดงสมการ ใช้ Pearson correlation coefficient (PCC).....	10
รูปที่ 9 แสดงสมการของ Metric RMSE, MAE และ MAPE ตามลำดับ.....	11
รูปที่ 10 ตัวอย่างของข้อมูลภาพรวม .....	15
รูปที่ 11 ตัวอย่างของข้อมูลของก่อนเปลี่ยนโครงสร้าง .....	17
รูปที่ 12 ตัวอย่างของข้อมูลของหลังทำ Unpivot Data.....	17
รูปที่ 13 รูปตัวอย่างของข้อมูลของ Feature ที่นำไปใช้วิจัย .....	18
รูปที่ 14 ตัวอย่างของข้อมูลของก่อนทำการรวมข้อมูล .....	19
รูปที่ 15 ตัวอย่างของข้อมูลของหลังทำการรวมข้อมูล .....	19
รูปที่ 16 ตัวอย่างของข้อมูลของเป็นรูปแบบ Long Format.....	20
รูปที่ 17 ตัวอย่างของข้อมูลเป็นรูปแบบ Wide Format.....	20
รูปที่ 18 ตัวอย่างของข้อมูลผู้ป่วยยืนยันแบบสะสม (Cumulative Data).....	21
รูปที่ 19 ตัวอย่างของข้อมูลผู้ป่วยยืนยันที่เพิ่มตามจริง (Real Data) .....	22
รูปที่ 20 ตัวอย่างของข้อมูลผู้ป่วยยืนยันที่เพิ่มขึ้นหลังทำ Normalization .....	22
รูปที่ 21 แสดงโครงสร้างของ LSTM.....	26

รูปที่ 22 Inertia ของ K-Means ตั้งแต่จำนวน Cluster 2 ถึง 50..... 27

รูปที่ 23 ค่า Inertia โดยเทียบ %ความแตกต่างของ Step ก่อนหน้า ตั้งแต่จำนวน Cluster 2 ถึง 50  
..... 27

รูปที่ 24 แสดงประเทศที่อยู่ในกลุ่ม Cluster เดียวกันประเทศไทย ..... 30

รูปที่ 25 แสดงคู่ประเทศไทยและญี่ปุ่นของผู้ป่วยก่อนการทำ Alignment..... 31

รูปที่ 26 แสดงคู่ประเทศไทยและญี่ปุ่นของหลังทำ Normalization และ Smoothing..... 31

รูปที่ 27 แสดงคู่ประเทศไทยและญี่ปุ่นของผู้ป่วยหลังการทำ Alignment ..... 32

รูปที่ 28 แสดงคู่ประเทศไทยและมาเลเซียของผู้ป่วยก่อนการทำ Alignment ..... 32

รูปที่ 29 แสดงคู่ประเทศไทยและมาเลเซียของหลังทำ Normalization และ Smoothing ..... 33

รูปที่ 30 แสดงคู่ประเทศไทยและมาเลเซียของผู้ป่วยหลังการทำ Alignment..... 33

รูปที่ 31 แสดงคู่ประเทศไทยและฟิลิปปินส์ของหลังทำ Normalization และ Smoothing ..... 34

รูปที่ 32 แสดงคู่ประเทศไทยและฟิลิปปินส์ของผู้ป่วยหลังการทำ Alignment ..... 34

รูปที่ 33 แสดงคู่ประเทศไทยและบังกลาเทศของผู้ป่วยก่อนการทำ Alignment..... 35

รูปภาพ 34 แสดงคู่ประเทศไทยและบังกลาเทศของหลังทำ Normalization และ Smoothing ..... 35

รูปที่ 35 แสดงคู่ประเทศไทยและบังกลาเทศของผู้ป่วยหลังการทำ Alignment ..... 36

รูปที่ 36 แสดงคู่ประเทศไทยและคิวบาของผู้ป่วยก่อนการทำ Alignment ..... 36

รูปที่ 37 แสดงคู่ประเทศไทยและคิวบาของหลังทำ Normalization และ Smoothing ..... 37

รูปที่ 38 แสดงคู่ประเทศไทยและคิวบาของผู้ป่วยหลังการทำ Alignment..... 37

รูปที่ 39 แสดงคู่ประเทศไทยและอิรักของผู้ป่วยก่อนการทำ Alignment ..... 38

รูปที่ 40 แสดงคู่ประเทศไทยและอิรักของหลังทำ Normalization และ Smoothing ..... 38

รูปที่ 41 แสดงคู่ประเทศไทยและอิรักของผู้ป่วยหลังการทำ Alignment..... 39

รูปที่ 42 แสดงคู่ประเทศไทยและเม็กซิโกของผู้ป่วยก่อนการทำ Alignment..... 39

รูปที่ 43 แสดงคู่ประเทศไทยและเม็กซิโกของหลังทำ Normalization และ Smoothing ..... 39

รูปที่ 44 แสดงคู่ประเทศไทยและเม็กซิโกของผู้ป่วยหลังการทำ Alignment ..... 40



2927129371



รูปที่ 45 แสดงคู่ประเทศไทยและเวียดนามของผู้ป่วยก่อนการทำ Alignment ..... 40

รูปที่ 46 แสดงคู่ประเทศไทยและเวียดนามของหลังทำ Normalization และ Smoothing ..... 41

รูปที่ 47 แสดงคู่ประเทศไทยและเวียดนามของผู้ป่วยหลังการทำ Alignment..... 41



2927129371

CU iThesis 6272091321 thesis / recv: 25122565 00:06:46 / seq: 30

# บทที่ 1

## บทนำ

### 1.1 ความสำคัญและที่มาของงานวิจัย

องค์การอนามัยโรค (WHO) รายงานการติดเชื้อรายแรกของโลกในเมืองอู่ฮั่นประเทศจีน ในวันที่ 31 ธันวาคม 2562 ถือเป็นจุดเริ่มต้นของการแพร่ระบาดของโรค Coronavirus Disease Starting in 2019 หรือที่รู้จักกันในชื่อย่อ COVID-19 เมื่อได้รับเชื้อจะทำให้มีอาการเช่นเป็นไข้หรือเกิดความผิดปกติทางเดินหายใจ หากอาการรุนแรงจะสามารถลุกลามลงปอดและทำลายเนื้อเยื่อจนกระทั่งผู้ป่วยไม่สามารถหายใจและเสียชีวิตในที่สุด เมื่อผู้ป่วยมีอาการไอหรือจามเชื้อสามารถแพร่ได้ถึง 2 เมตร หากผู้ป่วยไม่ได้สวมอุปกรณ์ป้องกันทางเดินหายใจที่ดีจะไม่สามารถช่วยลดการแพร่เชื้อได้ ไวรัสดังกล่าวสามารถแพร่เชื้อได้อย่างรวดเร็วและไม่แสดงอาการได้ จากนั้นวันที่ 12 มกราคม 2563 ว่ามีผู้ป่วยติด COVID-19 ในประเทศไทยซึ่งถือเป็นผู้ป่วยรายแรกที่ได้รับเชื้อจากนอกประเทศจีน [1] การแพร่ระบาดเกิดขึ้นรวดเร็วและทาง WHO ได้รายงานต่อมาในวันที่ 30 มกราคม 2563 ว่ามี การแพร่กระจายไวรัสถึง 18 ประเทศทั่วโลก และในวันที่ 11 มีนาคมจำนวนผู้ป่วยได้ขึ้นไปมากกว่าหนึ่งแสนรายใน 114 ประเทศทั่วโลก ถือว่าเป็นวิกฤติที่ทางรัฐบาลทั่วโลกต้องเผชิญ

การวิเคราะห์และพยากรณ์จำนวนผู้ป่วย COVID-19 รายวันเป็นประโยชน์อย่างมากต่อรัฐบาลในการคาดการณ์การแพร่ระบาดของโรค เพื่อให้สามารถเป็นหนึ่งในข้อมูลประกอบสำหรับวางแผนนโยบายทางด้านสาธารณสุขการรับมือผู้ป่วยที่จะเพิ่มขึ้น การจัดการทางด้านบุคลากรของภาครัฐและเอกชน จำนวนเตียงในโรงพยาบาล อุปกรณ์ทางการแพทย์เช่นอุปกรณ์ช่วยหายใจ แผนการนำเข้าหรือผลิตวัคซีนให้เพียงพอต่อประชาชน ยารักษาโรค เป็นต้น

การวิเคราะห์ผลกระทบที่จะส่งผลกระทบต่อทางด้านเศรษฐกิจซึ่งกระทบโดยตรงต่อประชาชน เช่น ภาคการท่องเที่ยวซึ่งคิดเป็น 11% ของ GDP ในปี 2562 ปีก่อนการระบาด COVID-19 [2] เรื่องความปกติใหม่ (New Normal) ของการดำรงชีวิตที่ถูกปรับให้ เข้ากับสถานการณ์ ส่งผลต่อกิจกรรมทางเศรษฐกิจที่เปลี่ยนไป เป็นต้น

การวิเคราะห์ผู้ป่วย COVID-19 ลักษณะข้อมูลของจำนวนผู้ป่วยที่เพิ่มขึ้นรายวันเป็นอนุกรมเวลา (Time-Series) โดยงานวิจัยจะใช้การจัดกลุ่ม Clustering Algorithm เป็นหนึ่งใน Unsupervised Learning ของประเทศต่างๆ เพื่อจัดกลุ่มพฤติกรรมของข้อมูลผู้ป่วย COVID-19

ในแต่ละประเทศที่มีความใกล้เคียงกันอยู่ในกลุ่มเดียวกัน เพื่อทราบพฤติกรรมของประเทศที่ใกล้เคียงกับประเทศไทย จากนั้นแนวทางการพยากรณ์จำนวนผู้ป่วย โรคระบาดสามารถใช้หลักการทางสถิติและการเรียนรู้ของเครื่อง เพื่อเรียนรู้ Pattern ของข้อมูลโรคระบาด โดยใช้ความรู้ทางสถิติสามารถใช้ เช่น Regression [3], Autoregressive Integrated Moving Average (ARIMA) [4] และการใช้ความรู้การเรียนรู้ของเครื่อง (Machine Learning) เช่น การใช้ Support Vector Regression, การเรียนรู้เชิงลึก (Deep Learning) ซึ่งมีหลาย Model ที่สามารถใช้ได้เช่น Recurrent Neural Network (RNN) หรือ Long Short-Term Memory (LSTM) [5] เป็นต้น ในงานวิจัยนี้ จึงมุ่งเน้นในการใช้ทั้งสถิติและการเรียนรู้ของเครื่องเพื่อวิเคราะห์และพยากรณ์จำนวนผู้ป่วย COVID-19

## 1.2 คำถามทางงานวิจัย

การใช้ข้อมูลของคู่ประเทศกับประเทศไทยช่วยให้ผลการทำนายจำนวนผู้ป่วย COVID-19 ดีขึ้นหรือไม่

## 1.3 วัตถุประสงค์ของงานวิจัย

1. เพื่อพยากรณ์จำนวนผู้ป่วยโรค COVID-19 ที่จะเกิดขึ้นในอนาคตของประเทศไทย โดยใช้ข้อมูลของ หลายประเทศที่มีความสัมพันธ์กับประเทศไทย
2. เพื่อศึกษาการประยุกต์ใช้หลักสถิติและการเรียนรู้ของเครื่องกับการพยากรณ์โรค COVID-19

## 1.4 ขอบเขตของงานวิจัย

1. ศึกษาข้อมูลจำนวนผู้ป่วยโรคโควิด-19 จากคลังข้อมูลของมหาวิทยาลัย John Hopkins
2. ศึกษาการใช้ Machine Learning เพื่อจัดกลุ่มแต่ละประเทศของข้อมูลผู้ป่วย และพยากรณ์จำนวนผู้ป่วย COVID-19



2927129371

## 1.5 ประโยชน์ที่คาดว่าจะได้รับ

1. เพื่อให้รัฐบาลเตรียมดำเนินการรองรับผู้ป่วย COVID-19 ที่จะเกิดขึ้นจากการพยากรณ์จำนวนผู้ป่วย
2. เพื่อคาดการณ์จำนวนผู้ป่วยโรค COVID-19 ของประเทศไทย โดยพิจารณาข้อมูลจากข้อมูลคู่ประเทศกับประเทศไทย

## 1.6 ขั้นตอนศึกษาวิจัย

1. ศึกษาขอบเขตและเป้าหมายในการทำวิจัย
2. ศึกษาข้อมูลเพื่อเก็บข้อมูลและเตรียมข้อมูลสำหรับการวิเคราะห์ข้อมูลในรูปแบบที่เหมาะสม
3. สร้างแบบจำลองการวิเคราะห์ข้อมูล เพื่อนำข้อมูลเข้าแบบจำลอง
4. วิเคราะห์ข้อมูล
5. เปรียบเทียบผลการทดลอง จากเทคนิคทางหลักสถิติและการเรียนรู้ของเครื่อง
6. สรุปและวิเคราะห์ผลการทดลอง
7. ทำบทความเชิงวิชาการ และทำรูปเล่มวิทยานิพนธ์

## บทที่ 2

### ทฤษฎีที่เกี่ยวข้อง

#### 2.1 การเรียนรู้ของเครื่อง Machine Learning

เป็นหนึ่งในศาสตร์ของปัญญาประดิษฐ์ (Artificial Intelligence) เป็นอัลกอริทึมที่สอนให้คอมพิวเตอร์ได้เรียนรู้ตัวเองและเข้าใจจากข้อมูลที่มนุษย์ป้อนให้คอมพิวเตอร์ ทั้งนี้วิธีการแบ่งใหญ่ๆ ออกเป็นการเรียนรู้แบบมีผู้สอน (Supervised Learning) ซึ่งการเรียนรู้แบบมีการป้อนข้อมูลแบบบอกคำตอบ (Label Data) ให้คอมพิวเตอร์ได้เรียนรู้จากคำตอบนั้น การเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) ตรงข้ามกับวิธีการก่อนหน้า ซึ่งจะให้คอมพิวเตอร์เรียนรู้จากข้อมูลโดยไม่มีคำตอบ [6]

#### 2.2 การเรียนรู้เชิงลึก Deep Learning

เป็นส่วนหนึ่งของอัลกอริทึมของการเรียนรู้ของเครื่อง โดยหลักการเป็นการใช้โครงข่ายประสาทเทียม (Neural Network) จำนวนหลายชั้นมาประกอบและอัลกอริทึมจะทำการคำนวณจากข้อมูลตัวอย่างที่ได้ เรียนรู้และจดจำเป็นหน่วยความจำในประสาทเทียมเหล่านี้ ทั้งนี้การเรียนรู้เชิงลึกสามารถประยุกต์ได้หลากหลายประเภท อาทิ การประมวลผลภาษาธรรมชาติ (Natural Language Processing) หรือ คอมพิวเตอร์วิทัศน์ (Computer Vision) เป็นต้น

#### 2.3 K-Mean Clustering (K-means)

K-Mean Clustering [7] เป็นหนึ่งใน Unsupervised Learning Algorithm ของ Machine Learning การทำ Clustering หลักการคือการกำหนดจุดศูนย์กลางของกลุ่ม (Centroid) มา k จุด และทำการคำนวณระยะห่างระหว่างจุดข้อมูล และจุดศูนย์กลาง (Centroid) ในหลักการของ Euclidean Distance จากนั้นกำหนดให้จุดข้อมูลที่มีระยะห่างน้อยที่สุดกับ Clustering เป็นข้อมูลกลุ่มนั้น และคำนวณหาจุดศูนย์กลางใหม่ ใน Cluster นั้น จากนั้นทำการคำนวณ Euclidean Distance ใหม่และวนทำแบบเดิม จนกระทั่งจุด Centroid ไม่มีการเปลี่ยนแปลง

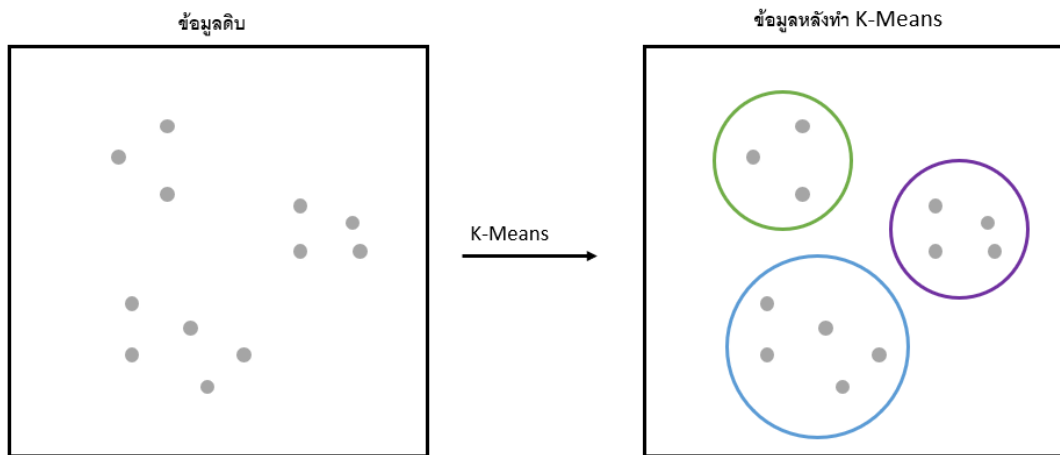


2927129371

CU Thesisis 6272091321 thesisis / revv: 25122565 00:06:46 / seq: 30

$$d = \sqrt{\sum_{i=1}^n (q_i - p_i)^2}$$

รูปที่ 1 สมการ Euclidean Distance



รูปที่ 2 แสดงภาพ K-Mean

K-Mean++ [8] เป็น Algorithm ที่ Upgrade Standard K-Mean โดยเพิ่ม Algorithm ในการเลือกค่าเริ่มต้น (Initial) ของ Centroid เนื่องจาก Standard K-Mean นั้นมีข้อด้อยที่มีความ Sensitive ต่อค่าเริ่มต้นของ Centroid ซึ่งจะส่งผลต่อการหาระยะห่างระหว่าง Centroid และ Datapoint และส่งผลต่อการจัดกลุ่มต่อไป สำหรับ K-Mean++ จะเริ่มจากการเลือก Centroid แรกอย่างสุ่มบน Datapoint ต่อมาทำการคำนวณระยะห่างระหว่างทุก Datapoint และ Centroid ตามสมการ

$$d_i = \max_{j:1 \rightarrow m} \|x_i - C_j\|^2$$

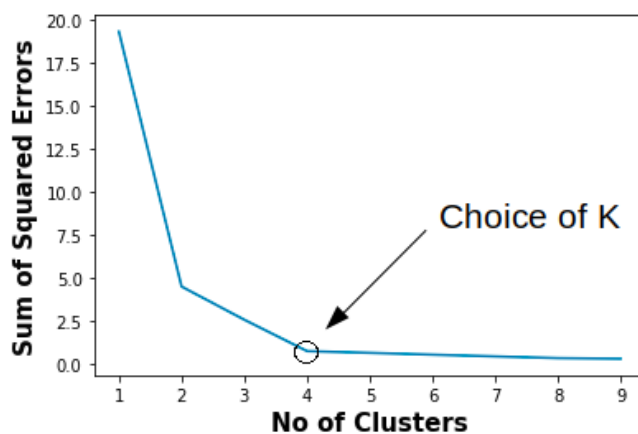
รูปที่ 3 แสดงสมการเลือก Centroid ของ K-Mean++

โดย  $d_i$  เป็นระยะห่างระหว่าง Data Point ( $x_i$ ) และ Centroid ( $C_j$ ) ที่ไกลที่สุด

m เป็นจำนวน Centroid ที่เลือก

จากนั้นทำการเลือกจุด Data Point ( $x_i$ ) ให้เป็น Centroid ใหม่ โดยเลือกจากค่า  $d_i$  หรือจุดที่มีระยะห่างจาก Centroid มากที่สุด จากนั้นทำวนจนกระทั่งได้ New Centroid ที่กำหนดครบ เมื่อจบการ Initial Centroid แล้วหลังจากนั้นทำขั้นตอนเหมือน Standard K-Means ปกติ

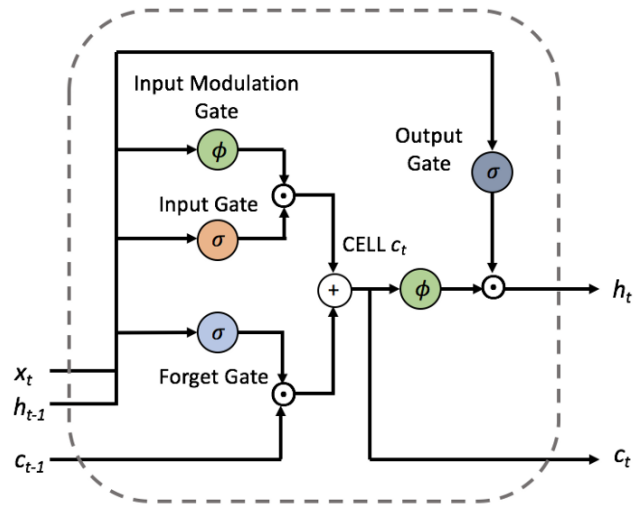
Elbow curve method เป็นวิธีการใช้ในการเลือกค่า K ที่เหมาะสม โดยทำการหาค่าเฉลี่ยค่า Inertia หรือเป็นค่าของ Error ระหว่างจุด Centroid และทุกจุดโดยเริ่มตั้งแต่ K ที่ 1 ไปจนถึง n ที่ต้องการ จากนั้นนำมาพิจารณาเลือก ค่า K ที่เหมาะสมตามค่า Error ที่อยู่ในจุดที่มีมุมหักเหลี่ยมศอก (Elbow) โดยเป็นการเลือกค่าความแตกต่างของ Error ที่มากที่สุดเมื่อเทียบกับค่า K ก่อนหน้า



รูปที่ 4 Elbow curve method

## 2.4 Long Short-Term Memory (LSTM)

เป็นรูปแบบโครงสร้างหนึ่งของ Neural Network ซึ่งอยู่ภายใต้ Deep Learning ทั้งนี้ Long Short-Term Memory เป็นหนึ่งในประเภทของ Recurrent Neural Network (RNN) โดยพัฒนา Gate ภายในที่ช่วยแก้ปัญหา Vanishing และ Exploding Gradient ของ RNN ได้ดียิ่งขึ้น



รูปที่ 5 โครงสร้าง Long Short-Term Memory

ประกอบขั้นตอนดังนี้

### 1. Forgetting

ใช้ตัวข้อมูลจาก Input data และ Hidden ของ Cell ก่อนหน้า ในการตัดสินใจว่าจะจำของข้อมูลเท่าไร และตามด้วย sigmoid function เป็นตัวตัดสินใจว่า

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f)$$

### 2. Writing

เมื่อมีข้อมูลเข้ามาจะทำการอัปเดต cell state ด้วย input data โดยใช้ข้อมูลจาก Input data และ Hidden ก่อนหน้า และ sigmoid function เป็นตัวตัดสินใจเช่นกัน

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i)$$

และนอกจากนี้ ข้อมูลที่เข้ามาใหม่ นำไปเป็น Input modulation gate หรือใส่ tanh function แทน

$$g_t = \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c)$$



2927129371



### 3. Updating cell

เป็นการนำข้อมูลจาก Forget Gate, Input Gate และ Input modulation gate นำมาคำนวณเพื่อ Update Cell stage โดยพจน์แรกของสมการนำ Forget Gate หากมีค่า 0 จะเป็นการไม่นำค่าของ Cell stage ก่อนหน้ามาคำนวณ แต่หากมีค่า 1 จะนำค่า Cell stage ก่อนหน้ามาคำนวณด้วย พจน์หลังของสมการ เป็นการนำค่า Input Gate และ Input modulation gate มาพิจารณา หาก Input Gate มีค่า 0 จะไม่นำ Input modulation gate แต่หากมีค่าเป็น 1 ก็ให้นำ Input modulation gate มา Update

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t$$

### 4. Reading

ค่าของ Hidden จะประกอบไปด้วยค่า Update Cell มาพิจารณาร่วมกับ Output Gate หาก Output มีค่า 0 จะทำให้ ค่าของ Hidden ไม่ส่งค่าต่อ หาก Output มีค่า 1 จะทำให้ Hidden มีค่าจากการคำนวณและส่งให้ LSTM Cell ถัดไป

$$h_t = o_t \odot \tanh(c_t)$$

ส่วนตัว Output Gate เป็นสมการเช่นเดียวกับ Input Gate และ Input Modulation Gate โดยพิจารณา  $h_{t-1}$  หากต้องการส่งค่า  $h_{t-1}$  เพื่อคำนวณจะใช้ Output Gate มาตัดสิน

$$o_t = \sigma(W_x \circ x_t + W_h \circ h_{t-1} + b_o)$$

#### 2.1.4 Adam Optimizer

Optimizer เป็นเทคนิคของการปรับค่าน้ำหนัก (Weight) และ Bias ของโครงข่ายประสาทเทียมในขณะที่มีการ Training ในแต่ละรอบของ Machine Learning หลักการจะเป็นการหา Local Minimum ของ Convex Function ซึ่งผลลัพธ์จากการใช้ Optimizer จะช่วยให้ผลการทำนายมีค่าดีขึ้น [9]

สำหรับ Adaptive Moment Estimation (Adam) Optimizer เป็น Algorithm ที่ได้รับความนิยมเพราะรวมข้อดีของแต่ละ Optimizer โดยหลักการในแต่ละรอบการ Training Data จะทำการปรับค่า Learning Rate ของตัว Weight และ Bias รวมถึงแก้ปัญหา Decaying ที่เกิดการจาก

Training ของรอบก่อนหน้าได้ดี และมีความไวไวกว่า Gradient Descent รวมถึงลดปัญหาการแกว่งของค่าพารามิเตอร์ [10]

$$w_{t+1} = w_t - \left( \frac{\alpha}{\sqrt{\hat{v}_t + \epsilon}} \right) \hat{m}_t$$

มาจากสมการ

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t}$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t}$$

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) \left[ \frac{\delta L}{\delta w_t} \right]$$

$$v_t = \beta_2 m_{t-1} + (1 - \beta_2) \left[ \frac{\delta L}{\delta w_t} \right]^2$$

รูปที่ 6 แสดงสมการของ Adaptive Moment Estimation (Adam) Optimizer

โดยค่า  $\epsilon$  ค่าคงที่ที่ใกล้ 0 เนื่องจากมีให้ส่วนเป็น 0

$\beta_1$  และ  $\beta_2$  เป็นค่า Decay Rate ของค่าเฉลี่ยจาก Gradient โดยมีค่า 0.9 และ 0.999 ตามลำดับ

$\alpha$  เป็นค่า Learning rate มีค่า 0.001

### 2.1.5 Reduce LRO Plateau

เป็นหนึ่งใน Algorithm ในการปรับ Learning Rate โดยเมื่อ Model ทำการ Training แล้ว Loss ของ Model ไม่มีการลดไปจนถึงรอบที่กำหนด จะทำการปรับค่า Learning Rate ให้ลดลง [11]

### 2.1.6 Normalization

ใช้เทคนิค Max-Min Normalization เป็นเทคนิคในการปรับให้ข้อมูลอยู่ในสเกลเดียวกันคือ ตั้งแต่ 0 ถึง 1 โดยมีสมการดังนี้

$$x_{std} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

$$x_{scaled} = x_{std} * (x_{max} - x_{min}) + x_{min}$$

รูปที่ 7 แสดงสมการ Max-Min Normalization

### 2.1.7 Cross-correlation

เป็นหนึ่งในวิธีการคำนวณเพื่อวัดความสัมพันธ์ระหว่างคู่ของข้อมูลที่เป็น Time series เพื่อหาความสัมพันธ์และตำแหน่งที่มีค่า correlation สูงสุด โดยมีค่าระหว่าง -1 จนถึง 1 หาก Time series ที่คำนวณได้ค่า cross-correlation ที่ใกล้ 1 จะหมายถึงทั้ง 2 ชุดข้อมูลนั้นมี correlation ใกล้เคียงกันมากที่สุด

ส่วนวิธีการคำนวณ Correlation ได้ใช้ Pearson correlation coefficient (PCC) โดยมีสมการดังนี้คือ

$$r_{xy} = \frac{n \sum xy - (\sum x)(\sum y)}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}}$$

รูปที่ 8 แสดงสมการ ใช้ Pearson correlation coefficient (PCC)

โดย r คือ Pearson correlation coefficient

X คือ ค่าของข้อมูลชุดแรก และ Y คือ ค่าของข้อมูลชุดที่สอง

## 2.1.8 Evaluation Method

การทดลองได้ใช้ Root mean square error (RMSE) , Mean absolute percentage error (MAE) และ Mean absolute percentage error (MAPE) ในการวัดค่าผลลัพธ์จาก Model

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (Actual_i - Predicted_i)^2}{N}}$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |Actual_i - Predicted_i|$$

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{Actual_i - Predicted_i}{Actual_i} \right| \times 100$$

รูปที่ 9 แสดงสมการของ Metric RMSE, MAE และ MAPE ตามลำดับ

โดยค่า *Actual<sub>i</sub>* เป็นค่าจริงของข้อมูล

ค่า *Predicted<sub>i</sub>* เป็นค่าที่ Model ได้ทำนายออกมา

### บทที่ 3

#### งานวิจัยที่เกี่ยวข้อง

การพยากรณ์ผู้ป่วย COVID-19 เพื่อที่จะสามารถรับมือกับจำนวนผู้ป่วยที่เพิ่มขึ้น การศึกษาเรื่องนี้ ได้มีหลายงานวิจัยที่มีการพยากรณ์จำนวนผู้ป่วย โดยใช้หลักสถิติหรือการเรียนรู้ของเครื่องหลากหลายวิธี

Punn, N. S., Sonbhadra, S. K., Agarwal, S. [12] ได้ศึกษาการวิเคราะห์ผู้ป่วย COVID-19 ด้วยข้อมูลจาก Johns Hopkins ในช่วงเวลา 22 มกราคม 2563 ถึง เมษายน 2563 โดยวิธีการเรียนรู้ของเครื่องได้แก่ Support Vector Regression (SVR), Polynomial Regression (PR), Deep Neural Network (DNN) และ Long Short-Term Memory (LSTM) ในการศึกษา ได้ผลการทดลองพบว่า Root Mean Square Error (RMSE) ของ Confirmed, Deaths และ Recovered Case ของวิธี PR มีค่าน้อยที่สุด งานวิจัยนี้ศึกษาการเรียนรู้ของเครื่องเพื่อความเข้าใจในจำนวนผู้ป่วย ที่จะเพิ่มขึ้นในอนาคต

Arora, P., Kumar, H. and Panigrahi, B. K. [13] ศึกษาและทำนายผู้ป่วย COVID-19 ด้วยข้อมูลจาก รัฐบาลอินเดียหน่วยงานกระทรวงสวัสดิการสุขภาพและครอบครัว (Ministry of Health and Family Welfare) โดยเป็นข้อมูลรายวันของผู้ป่วย COVID-19 ตั้งแต่วันที่ 14 มีนาคม 2563 จนถึง 14 เมษายน 2563 ของทั้ง 32 จังหวัดในประเทศอินเดีย จากนั้นได้ทำการวิเคราะห์ โดยใช้ Model Stacked LSTM, Bi-directional LSTM และ Convolutional LSTM เพื่อพยากรณ์จำนวนผู้ป่วย มีการ Tuning Hyper-parameter จากนั้นเมื่อได้ผลการ วิเคราะห์จาก Model แล้วจะทำการเลือก Model ที่มีความแม่นยำสูงสุดเพื่อใช้ในการพยากรณ์จำนวนผู้ป่วย ผลลัพธ์จากการทดลองสรุปได้ว่าการใช้ bi-directional LSTM สามารถให้ผลที่มีค่า Error น้อยที่สุด งานศึกษานี้มุ่งเน้นการนำเสนอใช้ Deep Learning Model สำหรับการพยากรณ์จำนวนผู้ป่วย COVID-19 ในประเทศอินเดีย และเพื่อให้ทางรัฐบาลสามารถวางแผน และจัดการด้านการแพทย์ สำหรับการเตรียมการรับมือจำนวนผู้ป่วยในอนาคต



2927129371

CT :Thesis 6272091321 thesis / recv: 25122565 00:06:46 / seq: 30

Javid, A. M., Liang, X., Venkitaraman, A., Chatterjee, S. [14] จัดทำวิเคราะห์แบบพยากรณ์ของการแพร่ระบาดของโรค COVID-19 จากการนำชุดข้อมูลออนไลน์มหาวิทยาลัย Johns Hopkins ของแต่ละประเทศใน 14 วันถัดไป การทำนายนี้ได้แปลงชุดข้อมูลให้อยู่บนสเกลของลอการิทึม ก่อนจะนำไปเข้าสู่โมเดล โดยการทดลองได้ศึกษา Model Polynomial Regression และ Neural Network และใช้ Technique Extreme Learning Machine (ELM) ซึ่งเป็น Neural Network 1 layer เพื่อลดการเกิด Over-fitting เนื่องจากข้อมูล ของ Model ในแต่ละประเทศมีจำนวนไม่มาก ผลการทดลองพบว่าค่า Error จาก Model ELM น้อยกว่า Polynomial Regression งานศึกษานี้ทำการประมาณผู้ป่วยเพื่อต่อยอดในการประมาณ Model การแพร่ระบาดของโรค COVID-19 และกำหนดนโยบายทางภาครัฐ

Bansal A. [15] ศึกษาและพยากรณ์ผู้ป่วย COVID-19 จากชุดข้อมูล Our World โดย Hannah Ritchie และ Max Roser ประเทศอินเดียจำนวน 190 วันนับจากวันแรกที่เกิดการติดเชื้อรายแรกในประเทศอินเดีย งานศึกษามีจุดประสงค์เพื่อศึกษาและใช้การเรียนรู้ของเครื่องในการพยากรณ์ และเพื่อศึกษา Feature ที่ส่ง ผลต่อการทำนายจำนวนผู้ป่วย งานวิจัยได้ใช้ Linear Regression และ Polynomial Regression ในการ ทดลองและพบว่า Polynomial Regression มีค่า Root Mean Square Error น้อยกว่า และการเพิ่มบาง Feature ก็จะทำให้เกิด Overfitting ส่งผลให้ RMS เพิ่มไปด้วย

Rahmad, K., Siti, N. H. S. A., Fitra, L., Mohd, Z. A. N., Akhmad, M., Noridayu, A. [16] ศึกษาและทำการจัดกลุ่มประเทศข้อมูลผู้ป่วย COVID-19 โดยใช้หลักการของ K-means ออกเป็น 5 กลุ่ม โดยมี Accuracy อยู่ที่ 97% และมีความสัมพันธ์ระหว่าง Total Death และ Critical Patient Attributes อยู่ที่ 0.85 งานวิจัยนี้มุ่งเน้นเพื่อศึกษาและจัดกลุ่มประเทศเพื่อศึกษาพฤติกรรมของผู้ป่วยที่ใกล้เคียงกัน

Hussein, B., Hassan, H., Ali, J., Chamseddine, Z., Chady A. J., Kifah, T., Layla, T. [17] ศึกษาและพยากรณ์ผู้ป่วย COVID-19 จากชุดข้อมูล WHO ในงานวิจัยได้ทำการสร้าง Model

COMAP โดยแบ่งออกเป็นระยะ Clustering และ Prediction โดย Clustering จัดกลุ่มประเทศที่มีพฤติกรรมผู้ป่วยคล้าย กันให้เป็นกลุ่มเดียวกันใน Cluster จากนั้นศึกษาและพยากรณ์จำนวนผู้ป่วย Covid-19 โดยใช้ Model LSTM และ Prophet ในระดับประเทศ และทวีป เพื่อศึกษาประสิทธิภาพของ Model COMAP งานวิจัยนี้มุ่งเน้นเพื่อ ศึกษาและจัดกลุ่มประเทศเพื่อศึกษาพฤติกรรมของผู้ป่วยที่ใกล้เคียงกันและพยากรณ์จำนวนผู้ป่วยเพื่อให้ภาครัฐได้รับมือในอนาคตได้ทัน

Ahmed, B. S., [18] ศึกษาและพยากรณ์ผู้ป่วย COVID-19 โดยการใช้ข้อมูลทางด้านข้อมูลประชากร, ภาวะเศรษฐกิจสังคม และ ตัวชี้วัดภาวะสุขภาพ เพื่อจัดกลุ่มของประเทศที่มีความใกล้เคียงกันด้วยวิธีการทำ Clustering วิธี K-means จากนั้นนำประเทศที่อยู่กลุ่มเดียวกันใน Cluster พยากรณ์จำนวนผู้ป่วย COVID-19 ด้วยวิธี Bidirectional LSTM (BI-LSTM) และใช้ข้อมูลเพิ่มเช่น การปิดโรงเรียน, การปิดการเดินทางขนส่งสาธารณะ, ข้อบังคับจำกัดจำนวนคนของการชุมนุมกัน, ข้อบังคับการเดินทางระหว่างพื้นที่ เป็นต้น ในการนำเข้า Model ผลที่ได้พบว่าได้ผลลัพธ์ที่ดีกว่าใน RMSE, MAE, CRM และ R-square งานวิจัยนี้มุ่งเน้นการศึกษานำข้อมูลของประเทศที่มีความใกล้เคียงกันและใช้ข้อมูลในกลุ่มเดียวกันเพื่อนำไปพยากรณ์จำนวนผู้ป่วย COVID-19

Dunfey, P. A., Emerson, V. O., Arthur, A. B., Davi, H. d. S., Andouglas, G. d. S. J., Lgor, G. P., Prisco, P., Alessandro, M., Cosimo, D., Jordan, S. C., Aura, C., Luiz, M.G. G. [19] ศึกษาพยากรณ์ผู้ป่วย COVID-19 โดยใช้ข้อมูลทางด้านสภาพอากาศเพื่อจัดกลุ่มประเทศที่มีความใกล้เคียงกันด้วยวิธีการทำ Clustering เพื่อวิเคราะห์ภาพรวมของข้อมูล โดยวิธี K-means จากนั้นพยากรณ์จำนวนผู้ป่วย COVID-19 ด้วยวิธี LSTM ของประเทศบราซิลในเมืองเซาเปาลู เปรียบเทียบระหว่างการทำ Univariate Approach หรือใช้แค่ข้อมูล Dead เทียบกับ Multivariate approach พบว่าได้ผลลัพธ์ที่ดีในแต่ละการ setting งานวิจัยนี้มุ่งเน้นเพื่อนำข้อมูลของประเทศที่มีความใกล้เคียงกันและเปรียบเทียบระหว่างการใช้อินพุตที่มี Feature เดียวและหลาย Feature เดียวกันเพื่อนำไปพยากรณ์จำนวนผู้ป่วย COVID-19

## บทที่ 4

### วิธีการดำเนินการวิจัย

#### 4.1 ภาพรวมของข้อมูล

ข้อมูลผู้ป่วย COVID-19 นำมาจากข้อมูล Open Data ของมหาวิทยาลัย John Hopkins โดยรวบรวมสถิติผู้ป่วยของประเทศทั่วโลกที่เกิดขึ้น ทางมหาวิทยาลัยได้เก็บข้อมูลเป็นรายระดับวัน และมีแบ่งออกมาเป็นระดับประเทศ 2 ระดับ คือ ได้แก่ระดับประเทศ และลงไประดับรัฐหรือภูมิภาค ขึ้นกับแต่ละประเทศ ประเทศที่แบ่งการเก็บข้อมูลลงไปถึง 2 ระดับเช่น ประเทศออสเตรเลีย เก็บข้อมูลลงระดับรัฐ เช่น นิวเซาท์เวลส์ ควีนส์แลนด์ เป็นต้น ส่วนประเทศอื่นๆ จะเก็บข้อมูลในระดับประเทศ เท่านั้น ผู้จัดทำได้นำข้อมูลผู้ป่วยยืนยัน (Confirmed case) นำช่วงที่มาใช้ตั้งแต่เริ่มต้นที่มีข้อมูลคือ วันที่ 22 มกราคม 2563 จนถึงวันที่ 31 ธันวาคม 2564 รวมแล้วเป็น 710 วัน และเลือกทุกประเทศที่มีการเก็บในช่วงนั้น

Province/State	Country/Region	Lat	Long	1/22/20	1/23/20	1/24/20	1/25/20	1/26/20	1/27/20	1/28/20
0	NaN	Afghanistan	33.93911	67.709953	0	0	0	0	0	0
1	NaN	Albania	41.15330	20.168300	0	0	0	0	0	0
2	NaN	Algeria	28.03390	1.659600	0	0	0	0	0	0
3	NaN	Andorra	42.50630	1.521800	0	0	0	0	0	0
4	NaN	Angola	-11.20270	17.873900	0	0	0	0	0	0
5	NaN	Antarctica	-71.94990	23.347000	0	0	0	0	0	0
6	NaN	Antigua and Barbuda	17.06080	-61.796400	0	0	0	0	0	0
7	NaN	Argentina	-38.41610	-63.616700	0	0	0	0	0	0
8	NaN	Armenia	40.06910	45.038200	0	0	0	0	0	0
9	Australian Capital Territory	Australia	-35.47350	149.012400	0	0	0	0	0	0
10	New South Wales	Australia	-33.86880	151.209300	0	0	0	0	3	4
11	Northern Territory	Australia	-12.46340	130.845600	0	0	0	0	0	0

รูปที่ 10 ตัวอย่างของข้อมูลภาพรวม



ชื่อคอลัมน์	ความหมาย
ประเทศ (Province/State)	ประเทศของข้อมูลผู้ป่วย
รัฐ/ภูมิภาค (Country/Region)	รัฐ/ภูมิภาคของข้อมูลผู้ป่วย
ละติจูด (Lat)	ละติจูดของระดับประเทศ หากมีข้อมูลในรัฐ/ภูมิภาคจะเป็นละติจูดของรัฐ/ภูมิภาคแทน
ลองจิจูด (Long)	ลองจิจูดของระดับประเทศ หากมีข้อมูลในรัฐ/ภูมิภาคจะเป็นลองจิจูดของรัฐ/ภูมิภาคแทน
ข้อมูลรายวัน (M/D/YYYY)	ข้อมูลผู้ป่วยยืนยันรายวัน

ตารางที่ 1 แสดงรายละเอียดแต่ละคอลัมน์และความหมาย

#### 4.2 เครื่องมือที่ใช้ในการวิเคราะห์งานวิจัย (Analysis Tools)

ในการวิเคราะห์และทำนายผู้ป่วย COVID-19 ผู้ทำการวิจัยได้ใช้ภาษา Python Version 3.8.15 ใช้ Library Numpy 1.21.6 เพื่อเป็นการเก็บข้อมูลให้อยู่ในรูปแบบพื้นฐาน Array ที่สามารถคำนวณได้อย่างเร็วและมีประสิทธิภาพ สามารถเก็บได้ในหลายมิติ (Multidimensional Array) ซึ่งเหมาะแก่การทำงานวิจัยในด้านการคำนวณทางด้านตัวเลข และวิทยาศาสตร์ ในการสร้างโครงสร้างข้อมูลที่เหมาะสมแก่การวิเคราะห์ข้อมูลเพื่อทำการเก็บข้อมูลของการวิจัย รวมไปถึงการเตรียมการข้อมูล Pre-processing ได้ใช้ Library Pandas Version 1.3.5 ได้ใช้ Library สำหรับการสร้าง Model Machine Learning แบ่งออกเป็นการทำ K-Mean เพื่อใช้ในการจัดกลุ่มของประเทศเดียวกัน ได้ใช้ Library ของ Tslearn ใน Version 0.5.2 ส่วน Model Long short-term memory (LSTM) เพื่อใช้ในการทำนายผู้ป่วย COVID-19 ได้ใช้ Keras ใน Version 2.9.0 และ Tensorflow ใน Version 2.9.2 สำหรับการสร้าง Metric เพื่อทำการวัดผลของการทำนายผู้ป่วย ได้ใช้ Sklearn Version 1.0.2

### 4.3 การเตรียมการของข้อมูล (Pre-processing)

#### 4.3.1 การเปลี่ยนรูปแบบของข้อมูลจากรูปแบบ Wide Format ไปยัง รูปแบบ Long Format

เปลี่ยนแปลงโครงสร้างข้อมูลให้เหมาะแก่การทำการรวมข้อมูล (Data Aggregation) ในขั้นตอนต่อไป ด้วยการทำ Unpivot Data โดยจากเดิม Data date จะเป็น Index ของโครงสร้างข้อมูล และข้อมูลประเทศ แทน Row ของข้อมูล หลังจากทำ Unpivot Data แล้วข้อมูลจะเปลี่ยนจาก Data ที่เป็น Index เป็นหนึ่งใน Column เช่นเดียวกับข้อมูลประเทศ

	Province/State	Country/Region	Lat	Long	1/22/20	1/23/20	1/24/20	1/25/20	1/26/20	1/27/20	1/28/20
0	NaN	Afghanistan	33.93911	67.709953	0	0	0	0	0	0	0
1	NaN	Albania	41.15330	20.168300	0	0	0	0	0	0	0
2	NaN	Algeria	28.03390	1.659600	0	0	0	0	0	0	0
3	NaN	Andorra	42.50630	1.521800	0	0	0	0	0	0	0
4	NaN	Angola	-11.20270	17.873900	0	0	0	0	0	0	0
5	NaN	Antarctica	-71.94990	23.347000	0	0	0	0	0	0	0
6	NaN	Antigua and Barbuda	17.06080	-61.796400	0	0	0	0	0	0	0
7	NaN	Argentina	-38.41610	-63.616700	0	0	0	0	0	0	0
8	NaN	Armenia	40.06910	45.038200	0	0	0	0	0	0	0
9	Australian Capital Territory	Australia	-35.47350	149.012400	0	0	0	0	0	0	0

รูปที่ 11 ตัวอย่างของข้อมูลของของก่อนเปลี่ยนโครงสร้าง

	Province/State	Country/Region	Lat	Long	Date	Confirmed
0	NaN	Afghanistan	33.93911	67.709953	1/22/20	0
1	NaN	Albania	41.15330	20.168300	1/22/20	0
2	NaN	Algeria	28.03390	1.659600	1/22/20	0
3	NaN	Andorra	42.50630	1.521800	1/22/20	0
4	NaN	Angola	-11.20270	17.873900	1/22/20	0
5	NaN	Antarctica	-71.94990	23.347000	1/22/20	0
6	NaN	Antigua and Barbuda	17.06080	-61.796400	1/22/20	0
7	NaN	Argentina	-38.41610	-63.616700	1/22/20	0
8	NaN	Armenia	40.06910	45.038200	1/22/20	0
9	Australian Capital Territory	Australia	-35.47350	149.012400	1/22/20	0

รูปที่ 12 ตัวอย่างของข้อมูลของของหลังทำ Unpivot Data

### 4.3.2 การเลือกข้อมูล (Data Selection)

เลือกข้อมูลเฉพาะ Feature ที่จะนำไปทดลองเท่านั้น ซึ่งจะเลือกเฉพาะข้อมูลประเทศ (Province/State), ข้อมูลระดับรัฐหรือภูมิภาค (Country/Region) และข้อมูลผู้ป่วยยืนยัน Confirm Cased รายวันตั้งแต่ 22 มกราคม 2563 จนถึงวันที่ 31 ธันวาคม 2564 รวมแล้วจำนวนประเทศทั้งหมด 199 ประเทศ มีจำนวนประเทศที่มีเก็บข้อมูลในระดับรัฐหรือภูมิภาคเป็นจำนวน 8 ประเทศ ได้แก่ ออสเตรเลีย (Australia), แคนาดา (Canada), จีน (China), เดนมาร์ก (Denmark), ฝรั่งเศส (France), เนเธอร์แลนด์ (Netherlands), นิวซีแลนด์ (New Zealand), อังกฤษ (United Kingdom) และประเทศที่เหลือ เป็นประเทศที่เก็บข้อมูลในระดับประเทศอย่างเดียว 191 ประเทศ

	Province/State	Country/Region	Date	Confirmed
0	NaN	Afghanistan	2020-01-22	0
1	NaN	Albania	2020-01-22	0
2	NaN	Algeria	2020-01-22	0
3	NaN	Andorra	2020-01-22	0
4	NaN	Angola	2020-01-22	0
5	NaN	Antarctica	2020-01-22	0
6	NaN	Antigua and Barbuda	2020-01-22	0
7	NaN	Argentina	2020-01-22	0
8	NaN	Armenia	2020-01-22	0
9	Australian Capital Territory	Australia	2020-01-22	0

รูปที่ 13 รูปตัวอย่างของข้อมูลของ Feature ที่นำไปใช้วิจัย

### 4.3.3 การรวมข้อมูล (Data Aggregation)

ทำการรวมข้อมูลของผู้ป่วยยืนยัน (Confirm Cased) จากระดับรัฐ/ภูมิภาค รวมเป็นระดับประเทศ เพื่อให้เหมาะแก่การทำวิจัยต่อในการเปรียบเทียบระดับของประเทศ เนื่องจากบางประเทศมีการแบ่งระดับผู้ป่วยออกเป็นระดับย่อยลงไป ซึ่งจะเหลือ Column ระดับประเทศและนำ Column ระดับรัฐ/ภูมิภาค (Province/State) ออก

	Province/State	Country/Region	Date	Confirmed
0	NaN	Afghanistan	2020-01-22	0
1	NaN	Albania	2020-01-22	0
2	NaN	Algeria	2020-01-22	0
3	NaN	Andorra	2020-01-22	0
4	NaN	Angola	2020-01-22	0
5	NaN	Antarctica	2020-01-22	0
6	NaN	Antigua and Barbuda	2020-01-22	0
7	NaN	Argentina	2020-01-22	0
8	NaN	Armenia	2020-01-22	0
9	Australian Capital Territory	Australia	2020-01-22	0

รูปที่ 14 ตัวอย่างของข้อมูลของของก่อนทำการรวมข้อมูล

	Date	Country/Region	Confirmed
0	2020-01-22	Afghanistan	0
1	2020-01-22	Albania	0
2	2020-01-22	Algeria	0
3	2020-01-22	Andorra	0
4	2020-01-22	Angola	0
5	2020-01-22	Antarctica	0
6	2020-01-22	Antigua and Barbuda	0
7	2020-01-22	Argentina	0
8	2020-01-22	Armenia	0
9	2020-01-22	Australia	0

รูปที่ 15 ตัวอย่างของข้อมูลของของหลังทำการรวมข้อมูล

#### 4.3.4 การเปลี่ยนรูปแบบของข้อมูลจากรูปแบบ Long Format ไปยัง รูปแบบ Wide Format

เปลี่ยนแปลงโครงสร้างข้อมูลให้เหมาะการแปลงข้อมูลในขั้นตอนต่อไป ด้วยการทำให้ Pivot Data โดยจากเดิมข้อมูลเก็บ Data Date ข้อมูลประเทศ และข้อมูลยืนยัน เป็น Column หลังจากทำ Pivot Data แล้วข้อมูลจะเปลี่ยน Data date จะเป็น Index ของโครงสร้างข้อมูล และข้อมูลประเทศเป็น Column คงเดิม

	Date	Country/Region	Confirmed
0	2020-01-22	Afghanistan	0
1	2020-01-22	Albania	0
2	2020-01-22	Algeria	0
3	2020-01-22	Andorra	0
4	2020-01-22	Angola	0
5	2020-01-22	Antarctica	0
6	2020-01-22	Antigua and Barbuda	0
7	2020-01-22	Argentina	0
8	2020-01-22	Armenia	0
9	2020-01-22	Australia	0

รูปที่ 16 ตัวอย่างของข้อมูลของเป็นรูปแบบ Long Format

Country/Region	Afghanistan	Albania	Algeria	Andorra	Angola	Antarctica	Antigua and Barbuda	Argentina	Armenia	Australia
Date										
2020-01-23	0	0	0	0	0	0	0	0	0	0
2020-01-24	0	0	0	0	0	0	0	0	0	0
2020-01-25	0	0	0	0	0	0	0	0	0	0
2020-01-26	0	0	0	0	0	0	0	0	0	4
2020-01-27	0	0	0	0	0	0	0	0	0	1
2020-01-28	0	0	0	0	0	0	0	0	0	0
2020-01-29	0	0	0	0	0	0	0	0	0	1
2020-01-30	0	0	0	0	0	0	0	0	0	3
2020-01-31	0	0	0	0	0	0	0	0	0	0
2020-02-01	0	0	0	0	0	0	0	0	0	3

รูปที่ 17 ตัวอย่างของข้อมูลเป็นรูปแบบ Wide Format

#### 4.3.5 การแปลงข้อมูลจากข้อมูลยืนยันผู้ป่วยแบบสะสม (Cumulative Data) เป็นข้อมูลที่เพิ่มขึ้นตามจริง (Real Data)

ทำการเปลี่ยนแปลงข้อมูลของข้อมูลผู้ป่วยยืนยันแบบสะสม (Cumulative Data) ไปเป็นข้อมูลผู้ป่วยยืนยันที่เพิ่มตามจริง (Real Data) รายวัน เพื่อต้องการให้ Model ทำการหารูปแบบและทำนายจำนวนผู้ป่วยที่เพิ่มตามจริง และนำข้อมูลของวันเริ่มต้นคือวันที่ 22 มกราคม 2563 ออก เนื่องจากไม่มีข้อมูลของวันก่อนหน้าเป็นข้อมูลอ้างอิง

Country/Region	Suriname	Sweden	Switzerland	Syria	Taiwan*	Tajikistan	Tanzania	Thailand	Timor-Leste	Togo	Tonga
Date											
2020-01-22	0	0	0	0	1	0	0	4	0	0	0
2020-01-23	0	0	0	0	1	0	0	4	0	0	0
2020-01-24	0	0	0	0	3	0	0	5	0	0	0
2020-01-25	0	0	0	0	3	0	0	6	0	0	0
2020-01-26	0	0	0	0	4	0	0	8	0	0	0
2020-01-27	0	0	0	0	5	0	0	8	0	0	0
2020-01-28	0	0	0	0	8	0	0	14	0	0	0
2020-01-29	0	0	0	0	8	0	0	14	0	0	0
2020-01-30	0	0	0	0	9	0	0	14	0	0	0
2020-01-31	0	0	0	0	10	0	0	19	0	0	0
2020-02-01	0	1	0	0	10	0	0	19	0	0	0
2020-02-02	0	1	0	0	10	0	0	19	0	0	0
2020-02-03	0	1	0	0	10	0	0	19	0	0	0
2020-02-04	0	1	0	0	11	0	0	25	0	0	0

รูปที่ 18 ตัวอย่างของข้อมูลผู้ป่วยยืนยันแบบสะสม (Cumulative Data)

Country/Region	Suriname	Sweden	Switzerland	Syria	Taiwan*	Tajikistan	Tanzania	Thailand	Timor-Leste	Togo	Tonga
Date											
2020-01-23	0	0	0	0	0	0	0	0	0	0	0
2020-01-24	0	0	0	0	2	0	0	1	0	0	0
2020-01-25	0	0	0	0	0	0	0	1	0	0	0
2020-01-26	0	0	0	0	1	0	0	2	0	0	0
2020-01-27	0	0	0	0	1	0	0	0	0	0	0
2020-01-28	0	0	0	0	3	0	0	6	0	0	0
2020-01-29	0	0	0	0	0	0	0	0	0	0	0
2020-01-30	0	0	0	0	1	0	0	0	0	0	0
2020-01-31	0	0	0	0	1	0	0	5	0	0	0
2020-02-01	0	1	0	0	0	0	0	0	0	0	0
2020-02-02	0	0	0	0	0	0	0	0	0	0	0
2020-02-03	0	0	0	0	0	0	0	0	0	0	0
2020-02-04	0	0	0	0	1	0	0	6	0	0	0

รูปที่ 19 ตัวอย่างของข้อมูลผู้ป่วยยืนยันที่เพิ่มตามจริง (Real Data)

#### 4.3.6 การทำ Normalization

ทำการปรับสเกลของข้อมูลให้จากข้อมูลดิบ เป็นข้อมูลที่มีค่าตั้งแต่ 0 ถึง 1 ใช้เทคนิค Max-Min Normalization [19]

Country/Region	Suriname	Sweden	Switzerland	Syria	Taiwan*	Tajikistan	Tanzania	Thailand	Timor-Leste	Togo	Tonga
Date											
2020-01-23	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
2020-01-24	0.000000	0.000000	0.000000	0.000000	0.007220	0.000000	0.000000	0.003610	0.000000	0.000000	0.000000
2020-01-25	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.002058	0.000000	0.000000	0.000000
2020-01-26	0.000000	0.000000	0.000000	0.000000	0.001495	0.000000	0.000000	0.002990	0.000000	0.000000	0.000000
2020-01-27	0.000000	0.000000	0.000000	0.000000	0.001247	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
2020-01-28	0.000000	0.000000	0.000000	0.000000	0.001140	0.000000	0.000000	0.002280	0.000000	0.000000	0.000000
2020-01-29	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
2020-01-30	0.000000	0.000000	0.000000	0.000000	0.000487	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
2020-01-31	0.000000	0.000000	0.000000	0.000000	0.000602	0.000000	0.000000	0.003010	0.000000	0.000000	0.000000
2020-02-01	0.000000	0.000479	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
2020-02-02	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
2020-02-03	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
2020-02-04	0.000000	0.000000	0.000000	0.000000	0.000251	0.000000	0.000000	0.001503	0.000000	0.000000	0.000000

รูปที่ 20 ตัวอย่างของข้อมูลผู้ป่วยยืนยันที่เพิ่มขึ้นหลังทำ Normalization

#### 4.4 การแบ่งกลุ่มข้อมูลประเทศ (Clustering) โดยใช้ K-Mean Clustering

สร้าง Model สำหรับการจัดกลุ่มประเทศโดยใช้ Library Tslern สำหรับการทำ Machine Learning ของข้อมูลที่เป็น Time Series โดยใช้ Algorithm K-Mean++ กำหนด iteration สำหรับการทำให้ Model 500 ครั้งต่อการ Run จากนั้นทำการหาจำนวนของเลขกลุ่ม Clustering ด้วยวิธี Elbow Method เพื่อประกอบการตัดสินใจเลือกจำนวนกลุ่ม Clustering

ทำการเลือกจำนวน Clustering โดยการสร้าง K-Mean กำหนดจำนวนกลุ่ม Clustering ตั้งแต่ 2 ไปจนถึง 50 กลุ่ม เพื่อเทียบ Inertia Measure หรือค่าที่ถูกคำนวณระยะทางระหว่าง Data point และ Centroid

#### 4.5 การปรับวางแนว (Alignment) ของผู้ป่วยยืนยันระหว่างคู่ประเทศไทยและประเทศในกลุ่มเดียวกัน

ผู้ทำการวิจัยได้จับคู่ระหว่างประเทศไทย และประเทศที่อยู่ในกลุ่ม Cluster เดียวกันครบทุกคู่ จากนั้นทำการปรับวางแนว (Alignment) ระหว่างข้อมูลประเทศไทยและคู่ประเทศ ด้วยการทำให้ Cross-Correlation เป็นการหาค่าความเหมือน (Similarity) โดยการ Shift ข้อมูลตามราย Date ของข้อมูลคู่ประเทศ และกำหนดให้ประเทศไทยคงที่ และใช้วิธีการของ Pearson Correlation เป็นวิธีการเปรียบเทียบระหว่างข้อมูลคู่ประเทศ เมื่อพบว่าค่า Pearson Correlation มีค่าสูงสุดในการ Shift ที่ตำแหน่งใด จะทำการ Shift ข้อมูลคู่ไปยังตำแหน่งที่มีค่าสูงสุด หลังจากนั้นจากนั้นจะทำการเติมค่าของข้อมูลในวันล่าสุดในต้นหรือปลายข้อมูลที่มีค่าว่างให้มีค่าเท่าเดิมค่าล่าสุด ทั้งนี้ข้อมูลก็นำไปทำได้ผ่านการทำให้ Normalize เพื่อปรับ Scaling Data ให้อยู่ในระหว่าง 0-1 และทำการ Smoothing Technique ด้วยการหาค่าเฉลี่ยย้อนหลัง 7 วัน (Moving Average)



#### 4.6 การทำนายผู้ป่วย COVID-19 ด้วยวิธีการ Long short-term memory (LSTM)

สร้าง Model สำหรับการทำนายผู้ป่วย COVID-19 โดยใช้ข้อมูลประเทศไทยและคู่ประเทศไทย เพื่อทำนายผู้ป่วยของประเทศไทย โดยใช้ Library Keras สำหรับการทำ Deep Learning ของข้อมูลที่เป็น Time Series โดยใช้ Long short-term memory (LSTM)

##### 4.6.1 การทำ Normalization

ทำการปรับสเกลของข้อมูลให้จากข้อมูลดิบ เป็นข้อมูลที่มีค่าตั้งแต่ 0 ถึง 1 ใช้เทคนิค Max-Min Normalization ก่อนนำเข้า Model LSTM

##### 4.6.2 การรวมข้อมูล (Data Aggregation)

ทำการรวมข้อมูลของผู้ป่วยยืนยัน (Confirm Cased) จากระดับรัฐ/ภูมิภาค รวมเป็นระดับประเทศ เพื่อให้เหมาะแก่การทำวิจัยต่อในการเปรียบเทียบระดับของประเทศ เนื่องจากบางประเทศมีการแบ่งระดับผู้ป่วยออกเป็นระดับย่อยลงไป ซึ่งจะเหลือ Column ระดับประเทศและนำ Column ระดับรัฐ/ภูมิภาค (Province/State) ออก

##### 4.6.3 การแปลงข้อมูลจากข้อมูลยืนยันผู้ป่วยแบบสะสม (Cumulative Data) เป็นข้อมูลที่เพิ่มขึ้นตามจริง (Real Data)

ทำการเปลี่ยนแปลงข้อมูลของข้อมูลผู้ป่วยยืนยันแบบสะสม (Cumulative Data) ไปเป็นข้อมูลผู้ป่วยยืนยันที่เพิ่มตามจริง (Real Data) รายวัน เพื่อต้องการให้ Model ทำการหารูปแบบและทำนายจำนวนผู้ป่วยที่เพิ่มตามจริง และนำข้อมูลของวันเริ่มต้นคือวันที่ 22 มกราคม 2563 ออก เนื่องจากไม่มีข้อมูลของวันก่อนหน้าเป็นข้อมูลอ้างอิง

#### 4.6.4 การทำการเตรียมข้อมูลให้เหมาะกับการทำ Multiple Lag Timesteps

ทำการปรับข้อมูลให้อยู่ในรูปแบบ Array ที่เหมาะสมกับการทำ Multiple Lag Timestep โดยกำหนดให้ Model จะได้เรียนรู้ข้อมูลใน 14 วันก่อนหน้า และทำนายวันที่ 15 ถัดมา เป็น Sliding Window สำหรับ LSTM ในการ Training เรียนรู้ข้อมูลผู้ป่วย

#### 4.6.5 การแบ่งข้อมูลเรียนรู้ (Training Set) และข้อมูลทดสอบ (Testing Set)

แบ่งข้อมูลออกเป็นข้อมูลสำหรับให้ข้อมูลเรียนรู้ (Training Set) และข้อมูลทดสอบ (Testing Set) ออกเป็น 90% และ 10% ตามลำดับ ข้อมูลของผู้ป่วยยืนยันที่เพิ่มขึ้นตามจริง เริ่มตั้งแต่วันที่ 23 มกราคม 2563 จนถึงวันที่ 31 ธันวาคม 2564 รวมแล้วเป็น 709 วัน ข้อมูล Training Set เริ่มตั้งแต่วันที่ 23 มกราคม 2563 จนถึง 22 ตุลาคม 2564 รวมทั้งหมด 638 วัน และข้อมูล Testing Set เริ่มตั้งแต่วันที่ 23 ตุลาคม 2564 จนถึง 31 ธันวาคม 2564 รวมทั้งหมด 71 วัน สำหรับ Validation set ได้แบ่ง 10% ของ Training Set ใน Library ของ Keras เพื่อทำการนำไปคำนวณและแสดงผลเป็นค่า Validation Loss ในการ Train แต่ละรอบ

#### 4.6.6 กำหนดโครงสร้าง LSTM

สร้าง Model สำหรับการจัดกลุ่มประเทศโดยใช้ Library Keras สำหรับการทำให้ Deep Learning ของข้อมูลที่เป็น Time Series โดยใช้ Model LSTM

กำหนดโครงสร้าง LSTM เริ่มจาก Input Layer โดยกำหนดให้ให้เป็น Vector 2 มิติ คือ 14 และ 2 เนื่องจาก Model จะได้เรียนรู้ข้อมูลใน 14 วันก่อนหน้า และ 2 คือข้อมูลยืนยันผู้ป่วยของประเทศไทย และคู่ประเทศไทย จากนั้นกำหนด LSTM Layer ที่ 1 เป็น 64 Node สำหรับให้ Model ได้เรียนรู้ข้อมูล ชั้นถัดมา กำหนด Dropout โดยคิดเป็น 20% ชั้นของ Layer ถัดมาเป็น LSTM 32 Node และสุดท้ายเป็น Fully Connected Layer เพื่อทำนายค่าของผู้ป่วยยืนยัน COVID-19 1 ค่า จึงมีค่า 1

Layer (type)	Output Shape	Param #
lstm_246 (LSTM)	(None, 14, 64)	17152
dropout_123 (Dropout)	(None, 14, 64)	0
lstm_247 (LSTM)	(None, 32)	12416
dense_123 (Dense)	(None, 1)	33

รูปที่ 21 แสดงโครงสร้างของ LSTM

#### 4.6.7 Training Model LSTM

กำหนด iteration สำหรับการทำ Model 50 ครั้งต่อการ Run ใช้ Optimization คือ Adaptive Moment Estimation (Adam) ใช้ ReduceLRonPlateau เพื่อเป็นการกำหนดค่า Learning Rate ให้มีค่าน้อยลงอย่าง Automatic และ Save Weight ของ Model เฉพาะ Validation Loss ที่ Best บน Validation Test Set

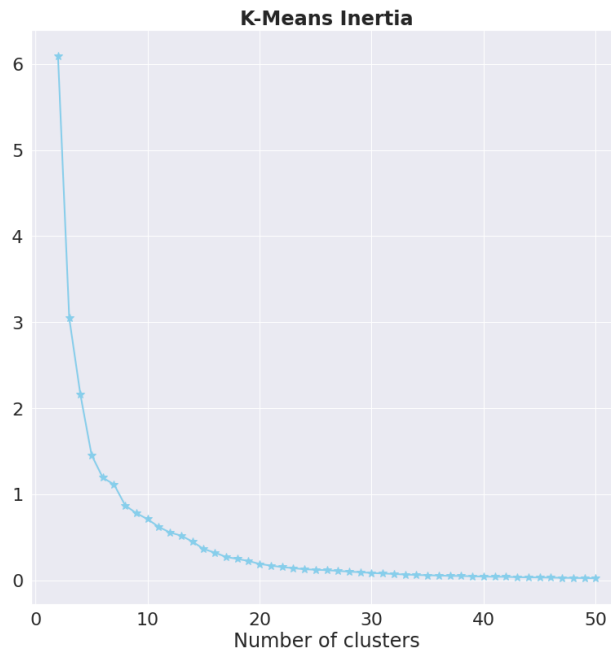
#### 4.6.8 วัดผลจาก Model (Evaluating Network)

ทำการวัดผลจาก LSTM Model ด้วยการใช้ RMSE, MAE และ MAPE

## บทที่ 5

### ผลการทดลองและอภิปรายผลการทดลอง

#### 5.1 การแบ่งกลุ่มข้อมูลประเภท (Clustering) โดยใช้ K-Mean Clustering



รูปที่ 22 Inertia ของ K-Means ตั้งแต่จำนวน Cluster 2 ถึง 50



รูปที่ 23 ค่า Inertia โดยเทียบ %ความแตกต่างของ Step ก่อนหน้า ตั้งแต่จำนวน Cluster 2 ถึง 50

No. Clustering	Inertia	% Difference Previous no. clustering
2	6.09577	-
3	3.04992	-49.967
4	2.15803	-29.243
5	1.45511	-32.572
6	1.19283	-18.025
7	1.11464	-6.5554
8	0.87069	-21.886
9	0.77944	-10.48
10	0.71143	-8.7255
11	0.61967	-12.898
12	0.55658	-10.181
13	0.5189	-6.7696
14	0.45037	-13.206
15	0.36432	-19.107
16	0.3221	-11.588
17	0.27203	-15.546
18	0.24885	-8.5206
19	0.22515	-9.5245
20	0.19025	-15.501
21	0.1712	-10.013
22	0.15399	-10.051
23	0.14172	-7.9692
24	0.12834	-9.4439
25	0.12259	-4.475
26	0.11807	-3.6923
27	0.10939	-7.3469
28	0.10092	-7.7464
29	0.09419	-6.6671
30	0.08527	-9.4658
31	0.0796	-6.6529
32	0.07356	-7.5896



2927129371

CU iThesis 6272091321 thesis / recv: 25122565 00:06:46 / seq: 30

33	0.06738	-8.4004
34	0.0624	-7.3852
35	0.05853	-6.2124
36	0.05584	-4.5922
37	0.05142	-7.9135
38	0.05051	-1.7786
39	0.04719	-6.5659
40	0.04334	-8.1616
41	0.04159	-4.0315
42	0.04083	-1.8326
43	0.03691	-9.6016
44	0.03443	-6.7219
45	0.03359	-2.4406
46	0.03142	-6.4495
47	0.02937	-6.5197
48	0.02799	-4.703
49	0.26204	-6.3859
50	0.023855	-8.9645

ตารางที่ 2 Inertia โดยเทียบ %ความแตกต่างของ Step ก่อนหน้า ตั้งแต่จำนวน Cluster 2 ถึง 50

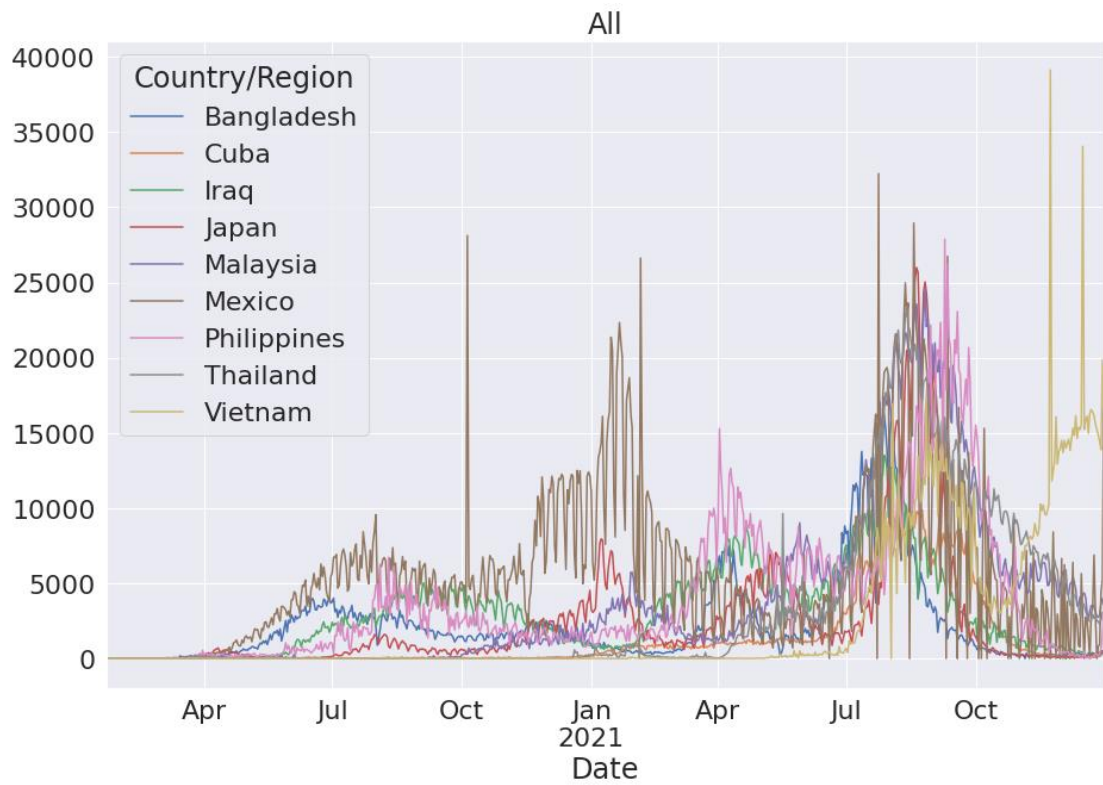
จากกราฟจะเห็นได้ว่า รูปที่ 22 แสดงถึง Inertia ของ K-Means ตั้งแต่จำนวน Cluster 2 ถึง 50 นั้น มีการลดลงอย่างรวดเร็วของ Inertia ตั้งแต่ Cluster 2 จนถึง 20 ทั้งนี้ผู้ทำการวิจัยเลือกจำนวน Cluster เป็น 8 แต่ผลลัพธ์หลังจากที่จัดกลุ่มพบว่าประเทศไทยอยู่ในกลุ่ม Cluster ที่ใหญ่ที่สุดซึ่งมีจำนวนประเทศทั้งหมดที่อยู่ใน Cluster เดียวกันเป็นจำนวน 183 ประเทศจากทั้งหมด 199 ประเทศ จึงเลือกใช้เป็น 15 แทน เนื่องจาก Percent Different ได้ลดลง - 19.107% ซึ่งเป็นจุด Elbow ที่ 2 ถัดมา

ผลลัพธ์จากการทำ K-Means จะได้ประเทศที่อยู่ใน Cluster เดียวกันกับประเทศไทยทั้งหมด 8 ประเทศ ได้แก่ ญี่ปุ่น (Japan), มาเลเซีย (Malaysia), ฟิลิปปินส์ (Philippines) ,บังกลาเทศ



2927129371

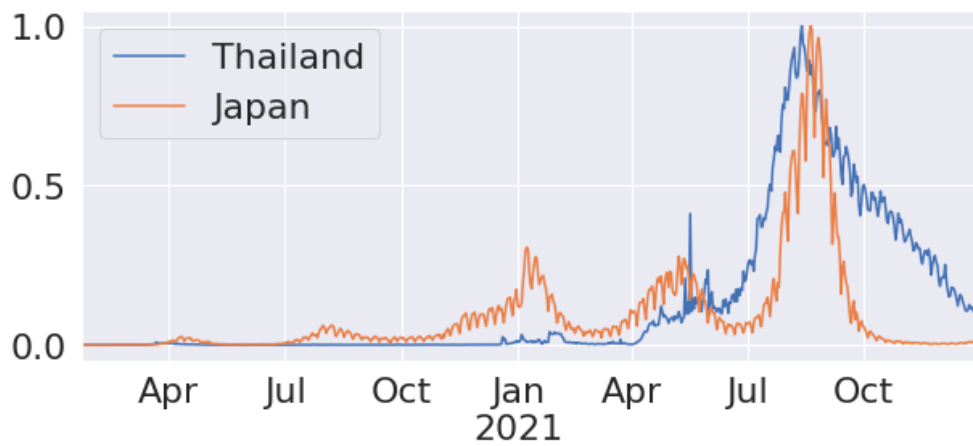
(Bangladesh) ,คิวบา (Cuba) ,อิรัก (Iraq) ,เม็กซิโก (Mexico) และ เวียดนาม (Vietnam)



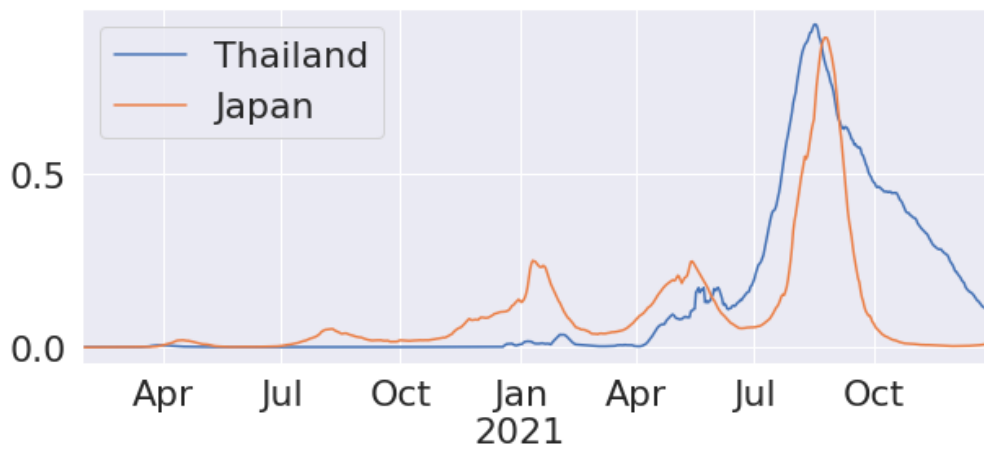
รูปที่ 24 แสดงประเทศที่อยู่ในกลุ่ม Cluster เดียวกันประเทศไทย

## 5.2 การปรับวางแนว (Alignment) ของผู้ป่วยยืนยันระหว่างคู่ประเทศไทยและประเทศในกลุ่มเดียวกัน

### 5.2.1 ประเทศไทยและคู่ประเทศญี่ปุ่น

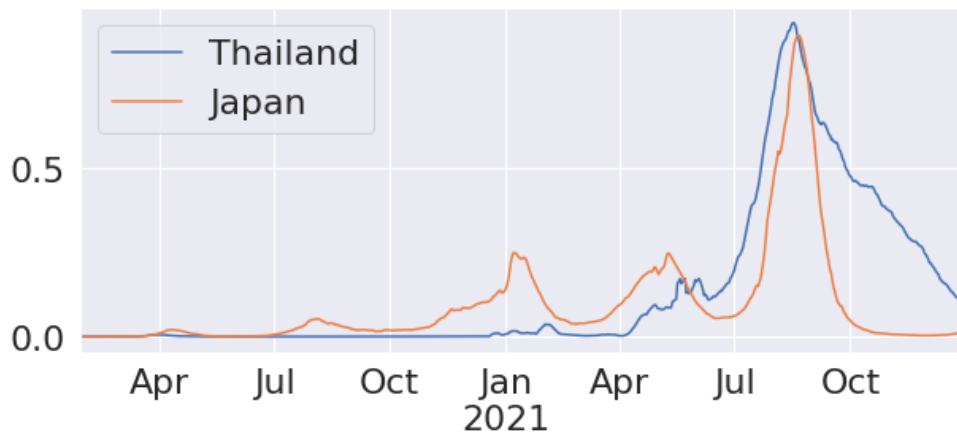


รูปที่ 25 แสดงคู่ประเทศไทยและญี่ปุ่นของผู้ป่วยก่อนการทำ Alignment



รูปที่ 26 แสดงคู่ประเทศไทยและญี่ปุ่นของหลังทำ Normalization และ Smoothing

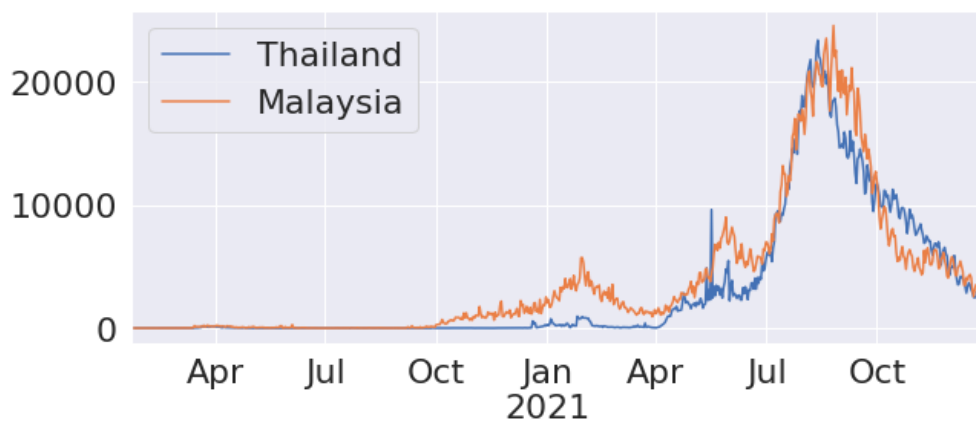




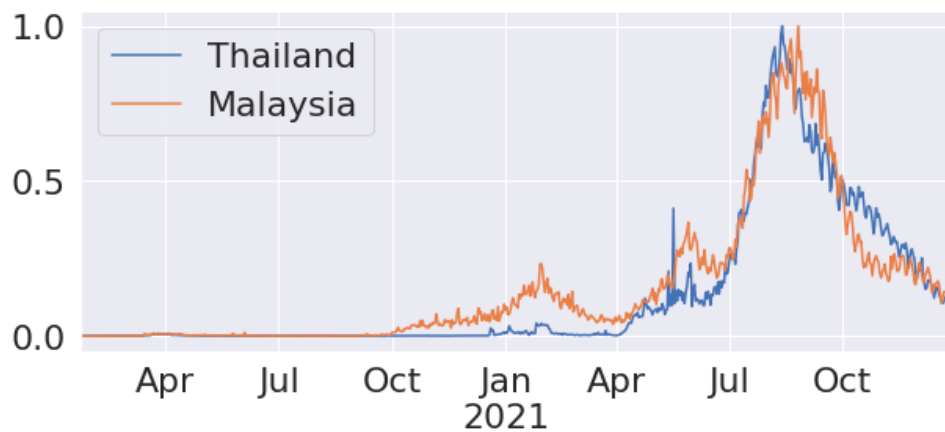
รูปที่ 27 แสดงคู่ประเทศไทยและญี่ปุ่นของผู้ป่วยหลังการทำ Alignment

หลังจากการทำ Normalization และ Smooth ของคู่ประเทศไทยและญี่ปุ่น พบว่าประเทศไทยญี่ปุ่นต้องการทำ Shift ไปยังข้อมูลก่อนหน้า 4 วัน เพื่อจะได้มีค่า Correlation ที่สูงสุด ซึ่งจะเห็นได้ว่าญี่ปุ่นมีการระบาดที่คล้ายกับไทยคือในช่วงเดือนมกราคม พฤษภาคม และช่วงเดือนกรกฎาคม

## 5.2.2 ประเทศไทยและคู่ประเทศมาเลเซีย



รูปที่ 28 แสดงคู่ประเทศไทยและมาเลเซียของผู้ป่วยก่อนการทำ Alignment



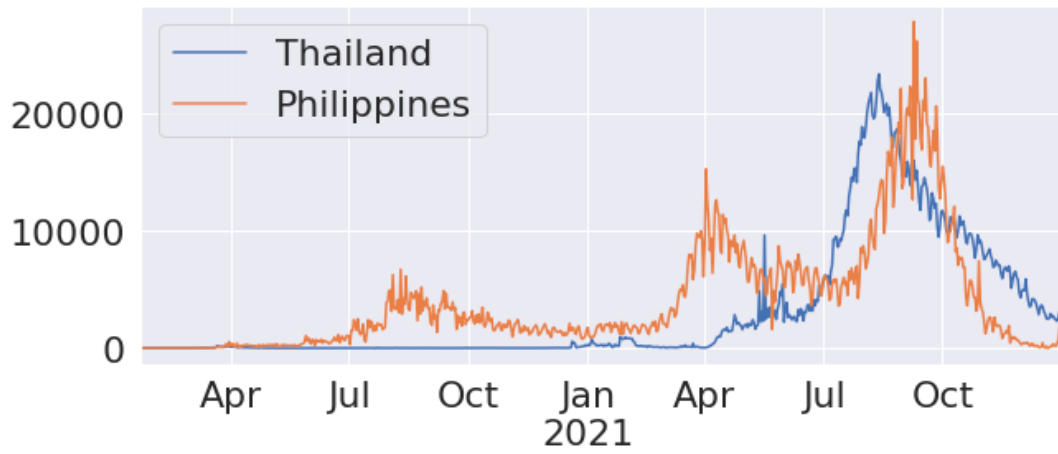
รูปที่ 29 แสดงคู่ประเทศไทยและมาเลเซียของหลังทำ Normalization และ Smoothing



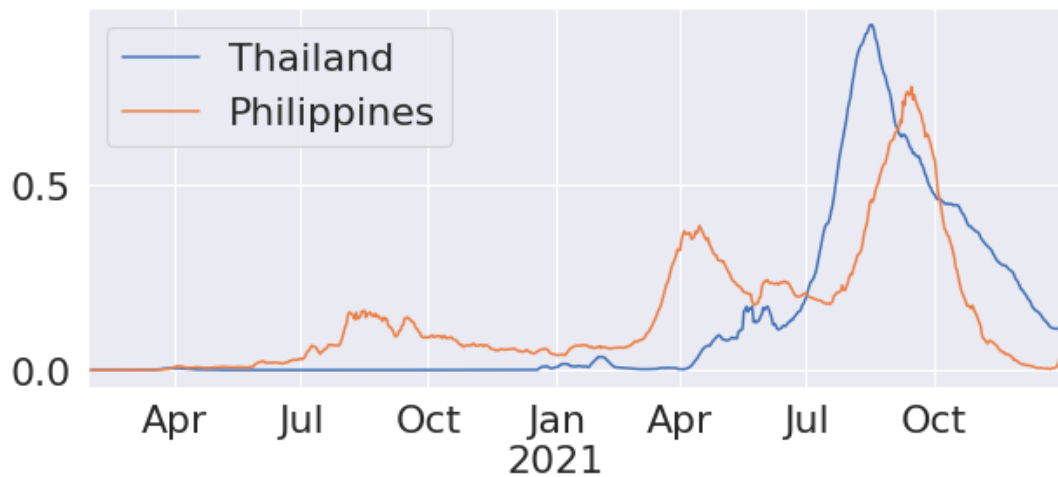
รูปที่ 30 แสดงคู่ประเทศไทยและมาเลเซียของผู้ป่วยหลังการทำ Alignment

หลังจากการทำ Normalization และ Smooth ของคู่ประเทศไทยและมาเลเซีย พบว่า ประเทศมาเลเซียต้องการทำ Shift ไปยังข้อมูลก่อนหน้า 1 วัน เพื่อจะได้มีค่า Correlation ที่สูงสุด ซึ่งจะเห็นได้ว่ามาเลเซียมีการระบาดที่คล้ายกับไทยคือในช่วงเดือนปลายปี 2563 เป็นต้นมา มีช่วงระบาดหนักคือ กุมภาพันธ์ พฤษภาคม และช่วงเดือนกรกฎาคม

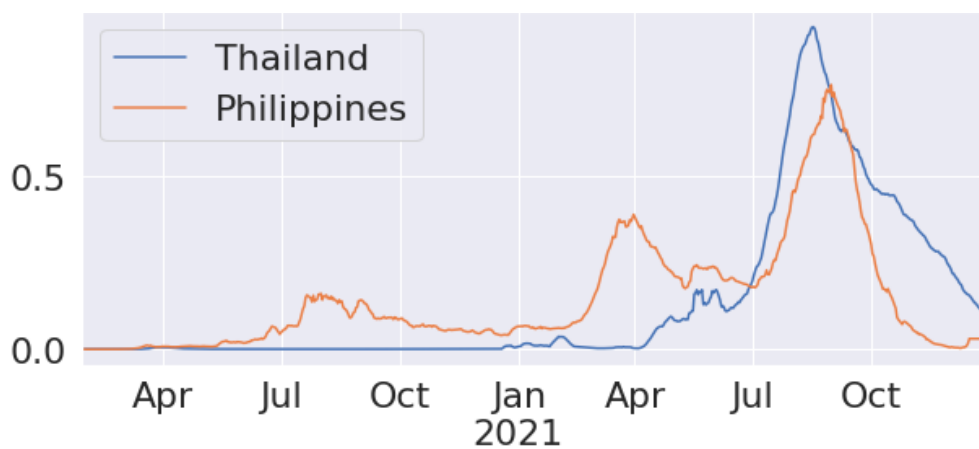
### 5.2.3 ประเทศไทยและคู่ประเทศฟิลิปปินส์



รูปที่ 30 แสดงคู่ประเทศไทยและฟิลิปปินส์ของผู้ป่วยก่อนการทำ Alignment



รูปที่ 31 แสดงคู่ประเทศไทยและฟิลิปปินส์ของหลังทำ Normalization และ Smoothing



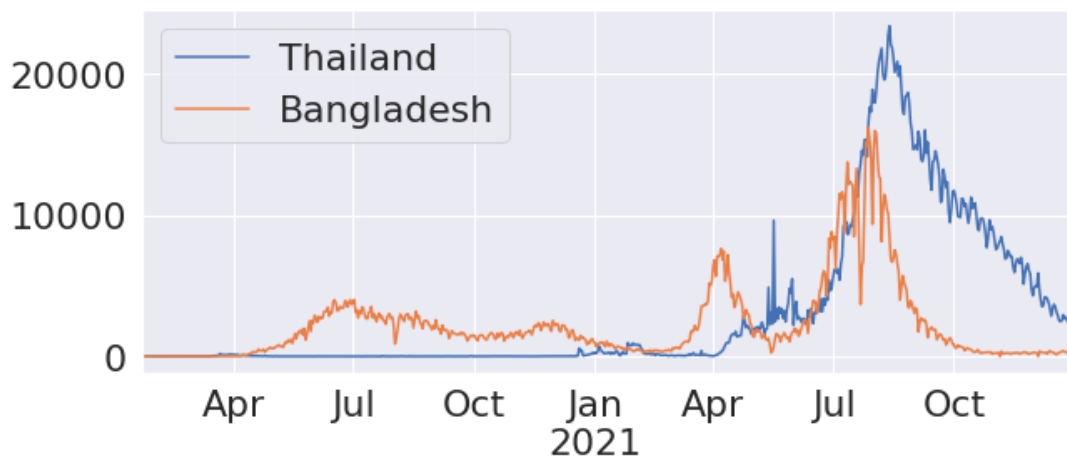
รูปที่ 32 แสดงคู่ประเทศไทยและฟิลิปปินส์ของผู้ป่วยหลังการทำ Alignment



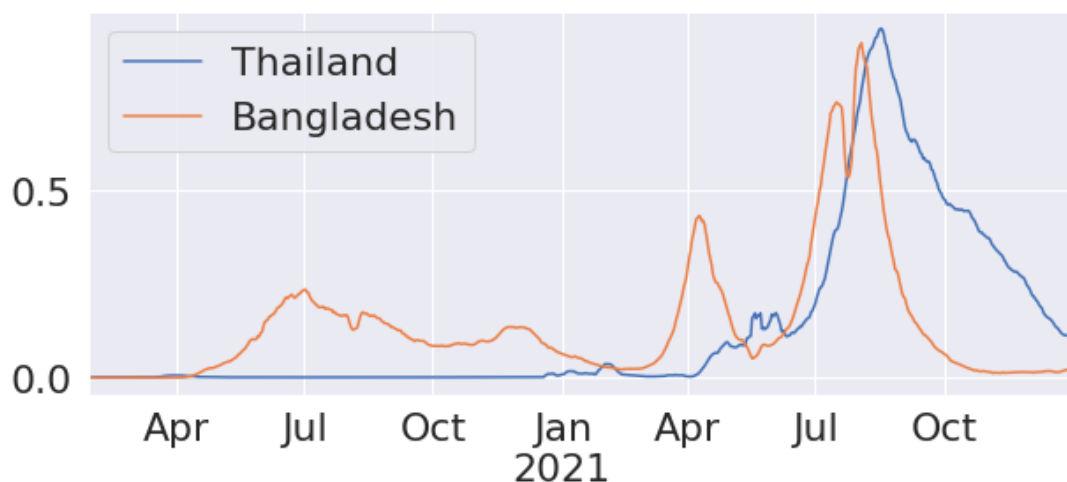
2927129371

หลังจากการทำ Normalization และ Smooth ของคู่ประเทศไทยและฟิลิปปินส์ พบว่า ประเทศมาฟิลิปปินส์ต้องการทำ Shift ไปยังข้อมูลก่อนหน้า 15 วัน เพื่อจะได้มีค่า Correlation ที่ สูงสุด ซึ่งจะเห็นได้ว่าฟิลิปปินส์มีการระบาดที่คือ ปี 2563 ในเดือนมิถุนายน ในปี 2564 คล้ายกับไทย คือในช่วงเดือนมกราคม มีนาคม และช่วงเดือนกรกฎาคม

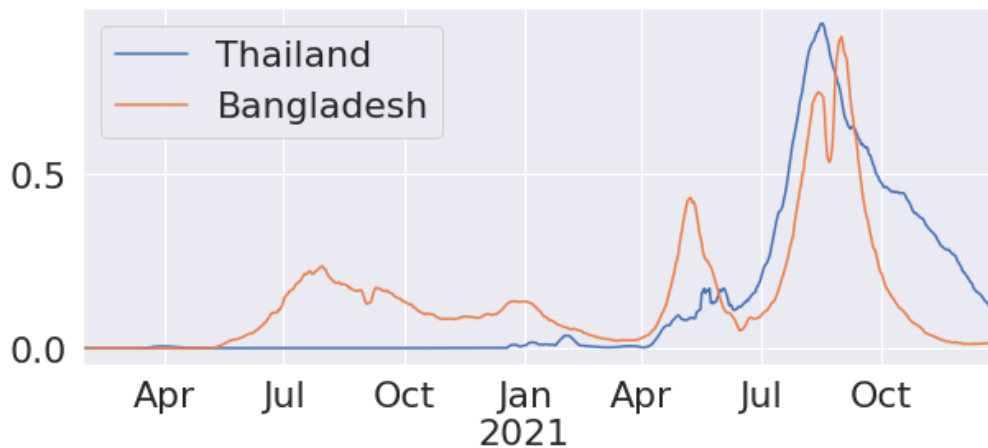
#### 5.2.4 ประเทศไทยและคู่ประเทศบังกลาเทศ



รูปที่ 33 แสดงคู่ประเทศไทยและบังกลาเทศของผู้ป่วยก่อนการทำ Alignment



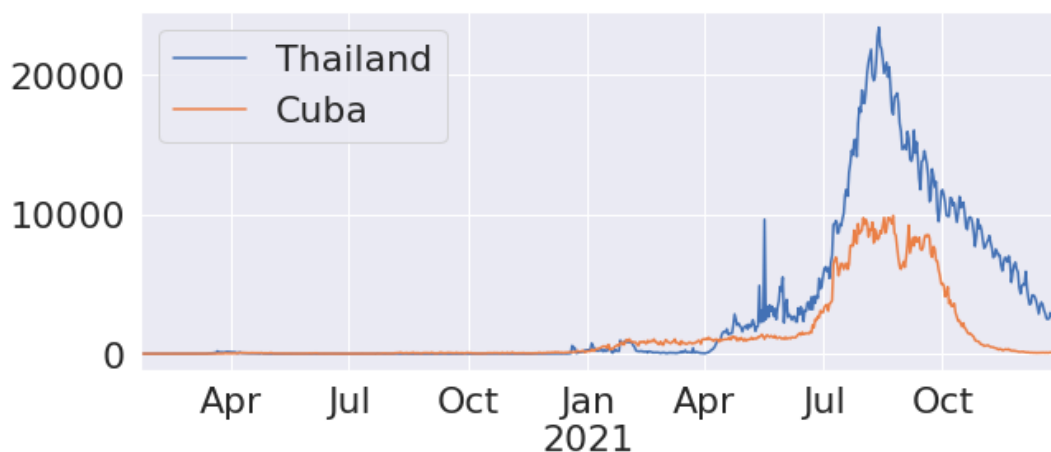
รูปภาพ 34 แสดงคู่ประเทศไทยและบังกลาเทศของหลังทำ Normalization และ Smoothing



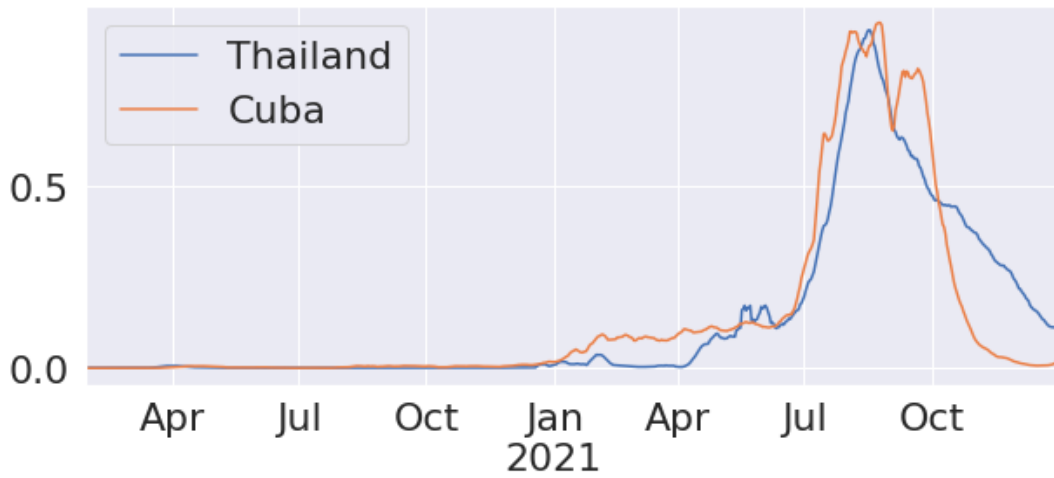
รูปที่ 35 แสดงคู่ประเทศไทยและบังกลาเทศของผู้ป่วยหลังการทำ Alignment

หลังจากการทำ Normalization และ Smooth ของคู่ประเทศไทยและบังกลาเทศ พบว่า ประเทศมาบังกลาเทศต้องการทำ Shift ไปยังข้อมูลถัดไป 29 วัน เพื่อจะได้มีค่า Correlation ที่สูงสุด ซึ่งจะเห็นได้ว่าบังกลาเทศมีการระบาดที่คือ ปี 2563 ในเดือนมิถุนายน ในปี 2564 คล้ายกับไทยคือ ในช่วงเดือนพฤษภาคม และช่วงเดือนกรกฎาคม

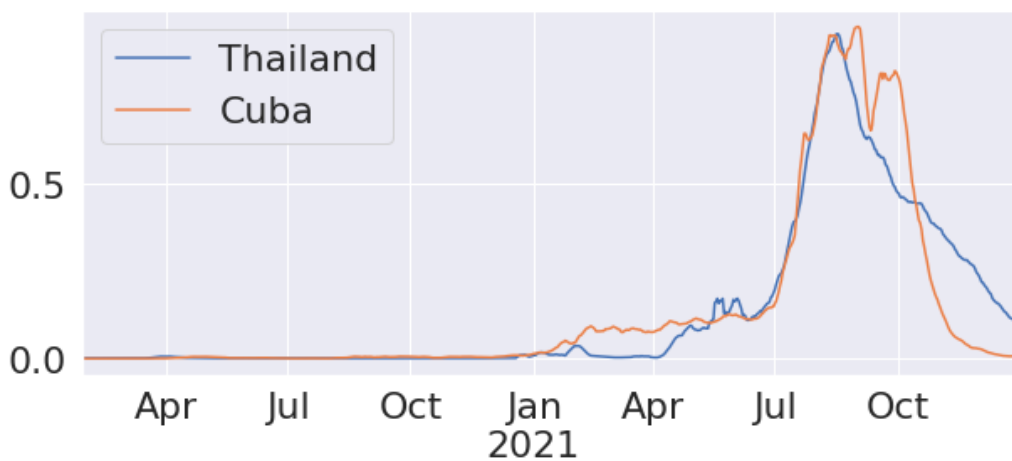
#### 5.2.5 ประเทศไทยและคู่ประเทศคิวบา



รูปที่ 36 แสดงคู่ประเทศไทยและคิวบาของผู้ป่วยก่อนการทำ Alignment



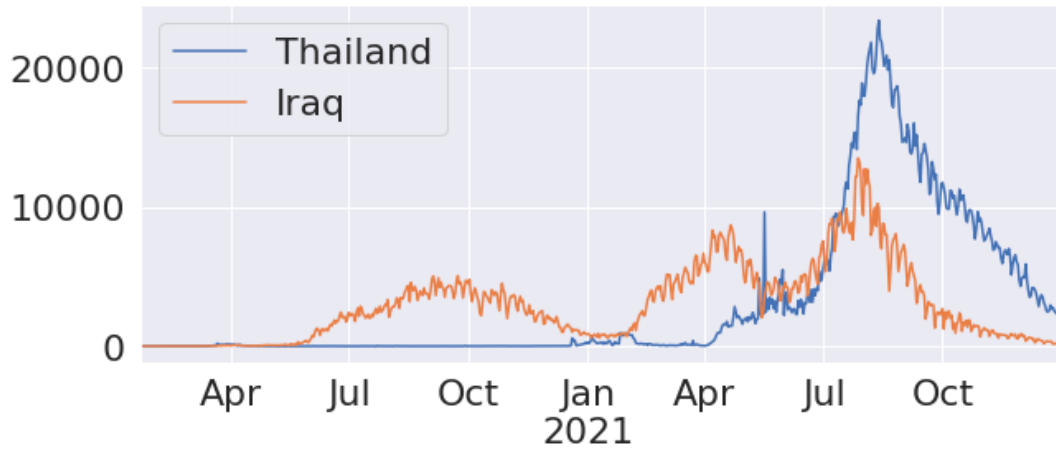
รูปที่ 37 แสดงคู่ประเทศไทยและคิวบาของหลังทำ Normalization และ Smoothing



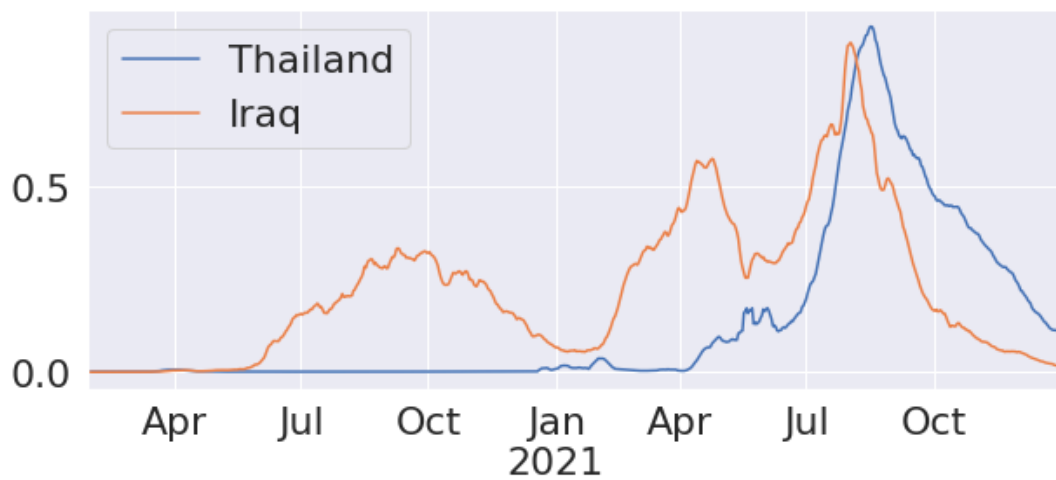
รูปที่ 38 แสดงคู่ประเทศไทยและคิวบาของผู้ป่วยหลังการทำ Alignment

หลังจากการทำ Normalization และ Smooth ของคู่ประเทศไทยและคิวบา พบว่าประเทศมาควิบาต้องการทำ Shift ไปยังข้อมูลถัดไป 8 วัน เพื่อจะได้มีค่า Correlation ที่สูงสุด ซึ่งจะเห็นว่าคิวบามีการระบาดที่คือ ในปี 2564 คล้ายกับไทยคือในช่วงเดือน กุมภาพันธ์ และช่วงเดือนมิถุนายน

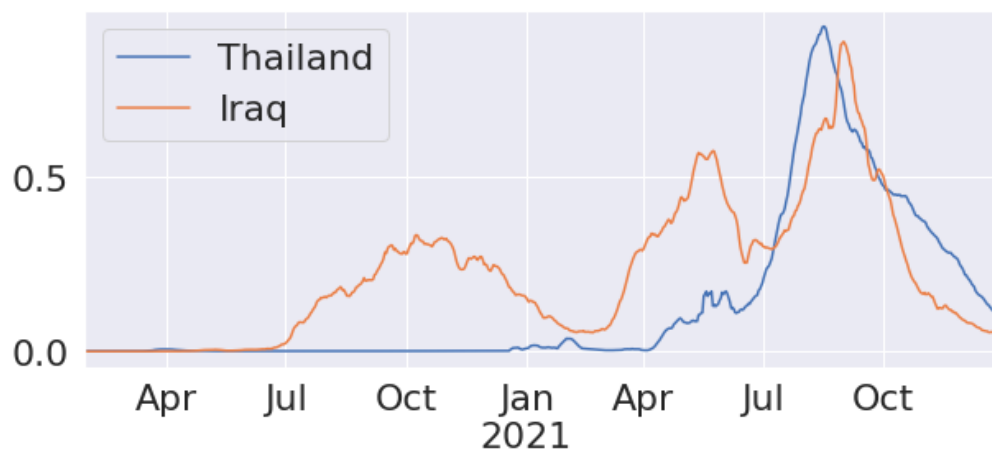
### 5.2.6 ประเทศไทยและคู่ประเทศอิรัก



รูปที่ 39 แสดงคู่ประเทศไทยและอิรักของผู้ป่วยก่อนการทำ Alignment



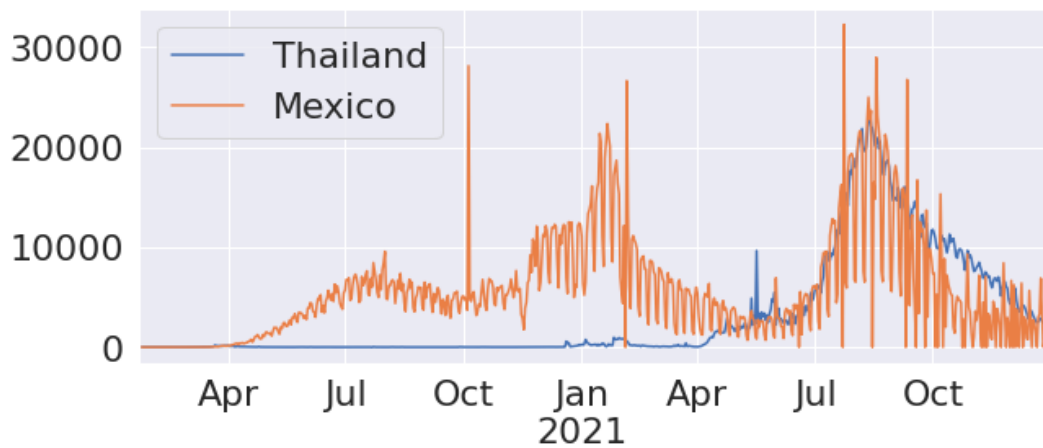
รูปที่ 40 แสดงคู่ประเทศไทยและอิรักของหลังทำ Normalization และ Smoothing



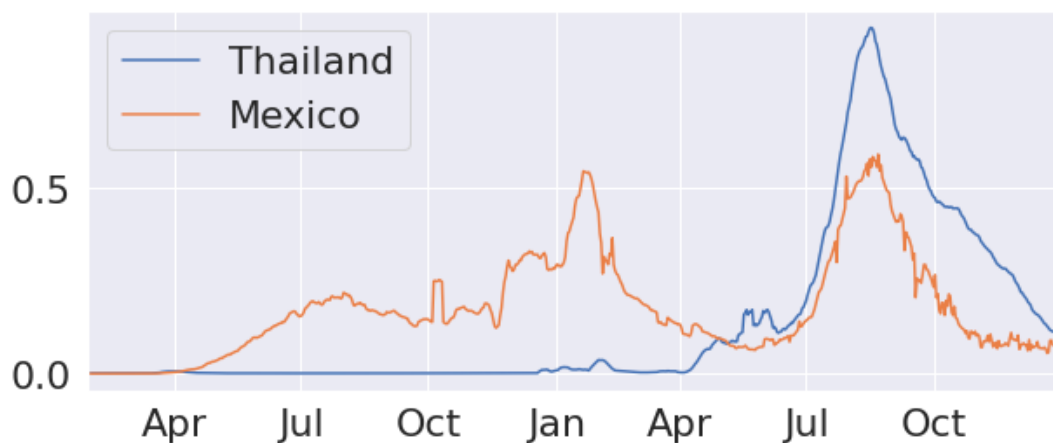
รูปที่ 41 แสดงคู่ประเทศไทยและอิรักของผู้ป่วยหลังการทำ Alignment

หลังจากการทำ Normalization และ Smooth ของคู่ประเทศไทยและอิรัก พบว่าประเทศมาอิรักต้องการทำ Shift ไปยังข้อมูลถัดไป 30 วัน เพื่อจะได้มีค่า Correlation ที่สูงสุด ซึ่งจะเห็นว่าอิรักการระบาดที่คือ ตั้งแต่ปี 2563 ช่วงเดือนกรกฎาคม ในปี 2564 คล้ายกับไทยคือในช่วงเดือนเมษายน และช่วงเดือนสิงหาคม

### 5.2.7 ประเทศไทยและคู่ประเทศเม็กซิโก

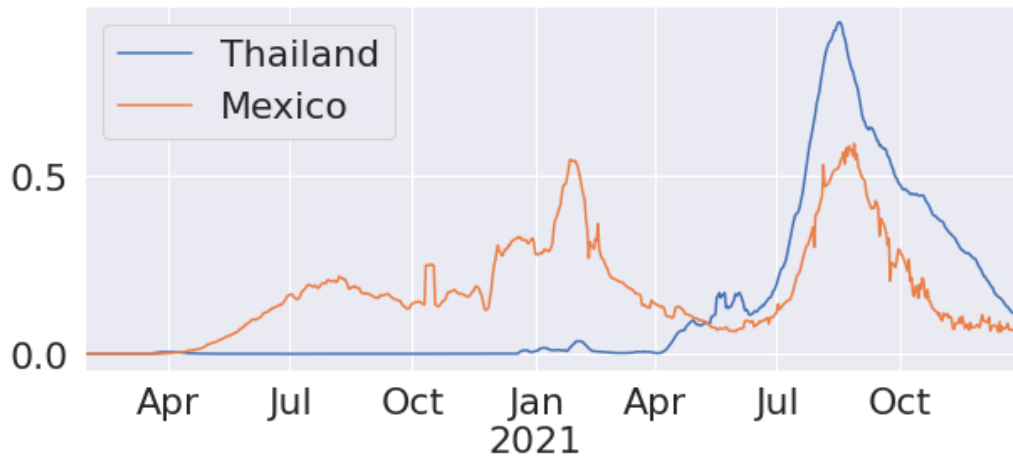


รูปที่ 42 แสดงคู่ประเทศไทยและเม็กซิโกของผู้ป่วยก่อนการทำ Alignment



รูปที่ 43 แสดงคู่ประเทศไทยและเม็กซิโกของหลังทำ Normalization และ Smoothing





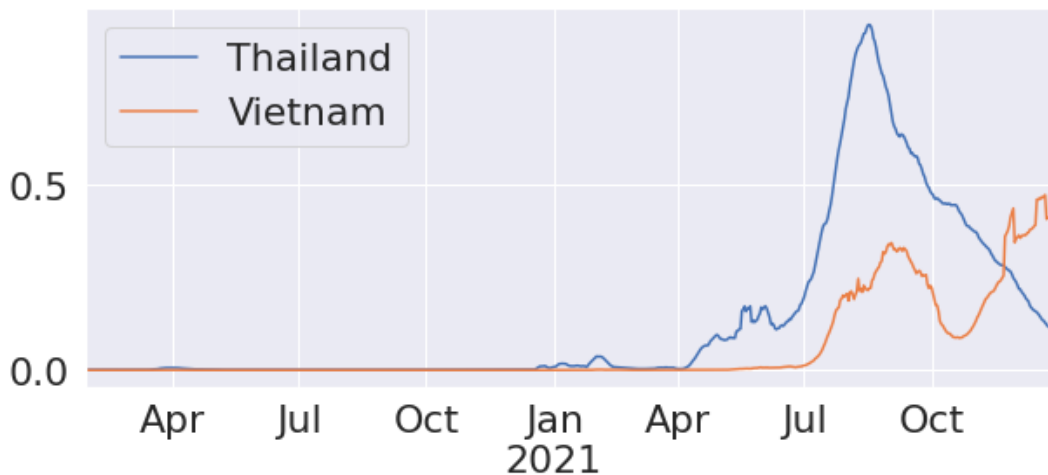
รูปที่ 44 แสดงคู่ประเทศไทยและเม็กซิโกของผู้ป่วยหลังการทำ Alignment

หลังจากการทำ Normalization และ Smooth ของคู่ประเทศไทยและเม็กซิโก พบว่า ประเทศเม็กซิโกต้องการทำ Shift ไปยังข้อมูลถัดไป 6 วัน เพื่อจะได้มีค่า Correlation ที่สูงสุด ซึ่งจะเห็นได้ว่าเม็กซิโกการระบาดมีตั้งแต่ปี 2563 ช่วงเดือนกรกฎาคมมาจนถึง ในปี 2564 คล้ายกับไทยคือ ในช่วงเดือนกรกฎาคมเป็นต้นมา

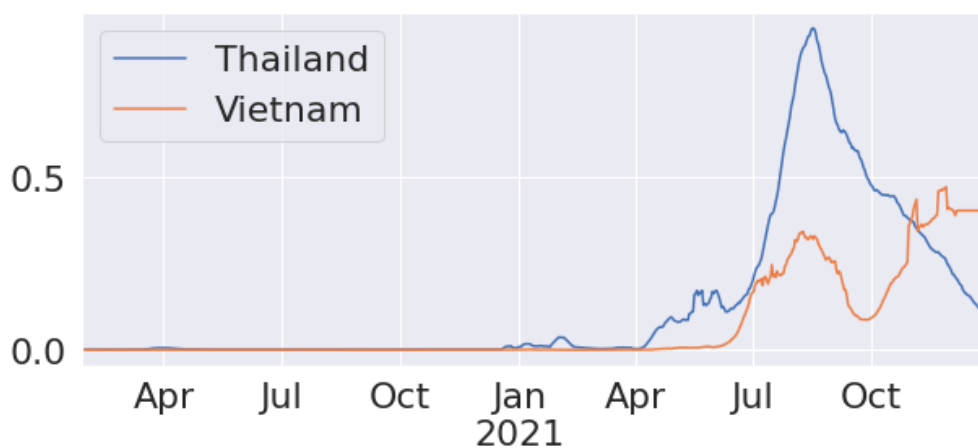
#### 5.2.8 ประเทศไทยและคู่ประเทศเวียดนาม



รูปที่ 45 แสดงคู่ประเทศไทยและเวียดนามของผู้ป่วยก่อนการทำ Alignment



รูปที่ 46 แสดงคู่ประเทศไทยและเวียดนามของหลังทำ Normalization และ Smoothing



รูปที่ 47 แสดงคู่ประเทศไทยและเวียดนามของผู้ป่วยหลังการทำ Alignment

หลังจากการทำ Normalization และ Smooth ของคู่ประเทศไทยและเวียดนาม พบว่า ประเทศเวียดนามต้องการทำ Shift ไปยังข้อมูลก่อนหน้า 24 วัน เพื่อจะได้มีค่า Correlation ที่สูงสุด ซึ่งจะเห็นได้ว่าเวียดนามการระบาดมีตั้งแต่ปี ค่ำกับไทยคือในช่วงเดือนกรกฎาคมเป็นระลอก ที่ 1 แต่มีระลอกที่ 2 เพิ่มอีกในประมาณเดือนตุลาคม ในขณะที่ไทยถือเป็นขาลง

No.	คู่ประเทศ ไทย	Correlation		Lag Length
		ก่อนทำ Normalize, Smooth และ Shifting	หลังทำ Normalize, Smooth และ Shifting	
1.	ญี่ปุ่น	0.71	0.73	-4
2.	มาเลเซีย	0.95	0.96	-1
3.	ฟิลิปปินส์	0.70	0.75	-15
4.	บังกลาเทศ	0.50	0.75	29
5.	คิวบา	0.92	0.94	8
6.	อิรัก	0.51	0.70	30
7.	เม็กซิโก	0.45	0.55	6
8.	เวียดนาม	0.64	0.69	-24

ตารางที่ 3 แสดงคู่ประเทศไทยในค่า Correlation ก่อนและหลังการทำ Pre-processing และทำ Alignment และค่า Lag Length

จากกราฟจะเห็นได้ว่าการ Shifting สูงที่สุดคือ อิรัก บังกลาเทศ และเวียดนาม ตามลำดับ และประเทศที่มี Shifting ที่ต่ำที่สุดคือ มาเลเซีย ญี่ปุ่น และเม็กซิโก ตามลำดับ

### 5.3 ผลจากการทำนายผู้ป่วยยืนยัน COVID-19 จาก Model LSTM

#### 5.3.1 ประเทศไทย

ผลการทำนายผู้ป่วยผู้ป่วยยืนยัน COVID-19 จาก Model LSTM โดยใช้ข้อมูลประเทศไทย ประเทศเดียว ได้ทำนายผู้ป่วยยืนยันใน Testing Set ตั้งแต่วันที่ 23 ตุลาคม 2564 จนถึง 31 ธันวาคม 2564 รวมทั้งหมด 71 วัน ได้ผลออกมาดังนี้

No.	RMSE	MAE	MAPE
1	1039.413986	899.95841	18.593995
2	935.814736	787.518097	18.993479
3	923.203207	779.182658	16.59515
4	823.355243	683.511234	14.879607

5	995.149982	850.568373	19.619551
6	951.128705	805.928866	19.223288
7	865.009831	723.437607	15.269111
8	586.967576	449.057143	9.516122
9	532.349649	420.510385	8.60616
10	1049.56154	901.743047	18.018055
11	1023.401852	883.45939	20.230192
12	1336.290971	1205.667508	24.118692
13	816.907632	695.273544	16.854167
14	893.420359	751.359296	17.144751
15	945.573535	809.699514	17.09405
16	567.996105	456.247253	8.596343
17	670.026294	520.789802	11.249351
18	1185.190285	1047.761664	24.917272
19	597.878087	485.967818	9.655942
20	649.132236	550.715549	12.509121
21	1209.361139	1081.949996	24.533336
22	687.230796	545.119079	12.029457
23	1169.907535	1036.001743	24.219285
24	887.62567	741.583293	18.396854
25	979.412341	837.26829	17.505856
26	612.925861	493.742029	9.559602
27	996.676817	853.851611	20.026779
28	1630.822301	1498.411917	29.313581
29	750.951909	643.155576	14.255349
30	1271.860482	1154.217959	24.799108

ตารางที่ 4 แสดง Metric จากการทำนาย LSTM โดยใช้ข้อมูลประเทศไทย

### 5.3.2 ประเทศไทยและคู่ประเทศญี่ปุ่น

ผลการทำนายผู้ป่วยผู้ป่วยยืนยัน COVID-19 จาก Model LSTM โดยใช้ข้อมูลประเทศไทย และคู่ประเทศญี่ปุ่น ได้ทำนายผู้ป่วยยืนยันใน Testing Set ตั้งแต่วันที่ 23 ตุลาคม 2564 จนถึง 31 ธันวาคม 2564 รวมทั้งหมด 71 วัน ได้ผลออกมาดังนี้

No.	RMSE	MAE	MAPE
1	955.264559	804.541023	20.254342
2	531.01152	420.947242	8.46412
3	623.855459	479.791701	10.310372
4	653.662234	535.173484	9.864366
5	568.825207	429.670716	8.881252
6	768.469118	627.695873	10.90915
7	697.087515	565.6501	10.065993
8	721.302767	604.709978	14.348393
9	1494.212851	1363.26465	26.842176
10	663.452723	563.648132	11.994629
11	977.728573	837.218812	17.64472
12	723.417648	581.373509	12.999702
13	608.116869	459.723175	9.594975
14	535.022914	428.64132	8.463754
15	896.069352	757.777108	13.67073
16	628.537373	523.288987	11.756548
17	548.489805	439.93316	8.571703
18	533.176101	430.49942	8.692452
19	1019.786306	874.265981	20.320231
20	562.909076	460.299221	8.886803
21	527.804687	417.222242	8.39286
22	734.437455	604.389277	10.901469

23	548.351008	422.061349	8.678462
24	898.501343	753.116947	17.918261
25	633.793401	487.271509	10.458995
26	825.278496	658.68664	13.056516
27	627.383823	480.276844	10.166034
28	762.098865	614.85701	13.012279
29	1137.587226	1006.904998	21.065438
30	1479.888399	1383.154078	30.58217

ตารางที่ 5 แสดง Metric จากการทำนาย LSTM โดยใช้ข้อมูลประเทศไทยและคู่ประเทศญี่ปุ่น

ผลจากการทำ Comparing Mean โดย T-Test Independent เมื่อเทียบกับการใช้ข้อมูลประเทศไทยอย่างเดียวจะได้ผลดังนี้

#### 1. RMSE

	country	N	Mean	Std. Deviation	Std. Error Mean
rmse	T	30	919.484889	256.2367342	46.7822131
	J	30	762.850756	255.1882504	46.5907870

ตารางที่ 6 แสดงภาพรวมสถิติ RMSE ผลการทำนายระหว่างประเทศไทยและคู่ประเทศไทยญี่ปุ่น

		Levene's Test for Equality of Variances		t-test for Equality of Means			95% Confidence Interval of the Difference			
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
rmse	Equal variances assumed	.028	.868	2.372	58	.021	156.6341330	66.0248203	24.4710957	288.7971702
	Equal variances not assumed			2.372	57.999	.021	156.6341330	66.0248203	24.4710484	288.7972176

ตารางที่ 7 ผลจากทดสอบ T-Test

2. MAE

Group Statistics

country		N	Mean	Std. Deviation	Std. Error Mean
mae	T	30	786.455288	252.3924178	46.0803402
	J	30	633.868483	253.0288391	46.1965343

ตารางที่ 8 แสดงภาพรวมสถิติ MAE ผลการทำนายระหว่างประเทศไทยและคู่ประเทศไทยญี่ปุ่น

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means					95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
mae	Equal variances assumed	.036	.850	2.339	58	.023	152.5868055	65.2496554	21.9754294	283.1981816
	Equal variances not assumed			2.339	58.000	.023	152.5868055	65.2496554	21.9754117	283.1981993

ตารางที่ 9 ผลจากทดสอบ T-Test

3. MAPE

Group Statistics

country		N	Mean	Std. Deviation	Std. Error Mean
mape	T	30	17.210787	5.4900224	1.0023364
	J	30	13.225630	5.6772314	1.0365159

ตารางที่ 10 แสดงภาพรวมสถิติ MAPE ผลการทำนายระหว่างประเทศไทยและคู่ประเทศไทยญี่ปุ่น

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means					95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
mape	Equal variances assumed	.003	.957	2.764	58	.008	3.9851570	1.4418888	1.0989029	6.8714111
	Equal variances not assumed			2.764	57.935	.008	3.9851570	1.4418888	1.0988339	6.8714802

ตารางที่ 11 ผลจากทดสอบ T-Test

### 5.3.3 ประเทศไทยและคู่ประเทศมาเลเซีย

ผลการทำนายผู้ป่วยผู้ป่วยยืนยัน COVID-19 จาก Model LSTM โดยใช้ข้อมูลประเทศไทย  
และคู่ประเทศมาเลเซีย

No.	RMSE	MAE	MAPE
1	902.912355	719.786542	11.828754
2	1608.319377	1316.334602	20.761051
3	1231.046823	962.38175	14.746947
4	630.78616	515.907973	11.063617
5	859.985384	681.185705	11.772876
6	925.550832	756.892454	13.603274
7	732.709835	591.758844	13.379476
8	664.765379	546.713547	11.843749
9	759.880494	614.35039	11.100847
10	700.192495	571.920035	11.058523
11	981.211807	783.608323	12.929917
12	656.075532	521.833654	9.556538
13	1241.795681	982.985733	15.170464
14	1015.806236	802.948992	12.764504
15	612.29068	495.924767	10.554615
16	1088.024764	957.497827	20.014899
17	1087.695414	852.058769	13.445911
18	921.80006	723.548625	11.819495
19	844.60285	663.999904	11.063675
20	2076.786137	1843.433984	31.762746
21	856.328192	672.029534	11.097561
22	1647.446469	1410.750889	23.501892
23	1614.633452	1327.863335	20.986702



24	1004.389764	785.133232	12.599934
25	719.320921	577.978571	10.505308
26	961.722936	749.671634	12.094843
27	607.754555	488.244344	9.649191
28	695.672745	549.44078	12.013258
29	627.50824	501.886519	9.31115
30	927.344337	792.222665	14.716705

ตารางที่ 12 แสดง Metric จากการทำนาย LSTM โดยใช้ข้อมูลประเทศไทยและคู่ประเทศมาเลเซีย

ผลจากการทำ Comparing Mean โดย T-Test Independent เมื่อเทียบกับการใช้ข้อมูลประเทศไทยอย่างเดียวจะได้ผลดังนี้

### 1. RMSE

	country	N	Mean	Std. Deviation	Std. Error Mean
rmse	T	30	919.484889	256.2367342	46.7822131
	Ma	30	973.478664	358.4354532	65.4410610

ตารางที่ 13 แสดงภาพรวมสถิติ RMSE ผลการทำนายระหว่างประเทศไทยและคู่ประเทศไทยมาเลเซีย

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
rmse	Equal variances assumed	1.438	.235	-.671	58	.505	-53.9937748	80.4431970	-215.0183286	107.0307790
	Equal variances not assumed			-.671	52.503	.505	-53.9937748	80.4431970	-215.3780870	107.3905374

ตารางที่ 14 ผลจากทดสอบ T-Test

### 2. MAE

## Group Statistics

	country	N	Mean	Std. Deviation	Std. Error Mean
mae	T	30	786.455288	252.3924178	46.0803402
	Ma	30	792.009797	316.3024809	57.7486679

ตารางที่ 15 แสดงภาพรวมสถิติ MAE ผลการทำนายระหว่างประเทศไทยและคู่ประเทศไทยมาเลเซีย

## Independent Samples Test

		Levene's Test for Equality of Variances					t-test for Equality of Means		95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
mae	Equal variances assumed	.393	.533	-.075	58	.940	-5.5545091	73.8803519	-153.4421013	142.3330831
	Equal variances not assumed			-.075	55.277	.940	-5.5545091	73.8803519	-153.5973807	142.4883626

ตารางที่ 16 ผลจากทดสอบ T-Test

## 3. MAPE

## Group Statistics

	country	N	Mean	Std. Deviation	Std. Error Mean
mape	T	30	17.210787	5.4900224	1.0023364
	Ma	30	13.890614	4.9066933	.8958355

ตารางที่ 17 แสดงภาพรวมสถิติ MAPE ผลการทำนายระหว่างประเทศไทยและคู่ประเทศไทยมาเลเซีย

## Independent Samples Test

		Levene's Test for Equality of Variances					t-test for Equality of Means		95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
mape	Equal variances assumed	1.085	.302	2.470	58	.016	3.3201728	1.3443212	.6292215	6.0111241
	Equal variances not assumed			2.470	57.283	.017	3.3201728	1.3443212	.6285044	6.0118412

ตารางที่ 18 ผลจากทดสอบ T-Test

### 5.3.4 ประเทศไทยและคู่ประเทศฟิลิปปินส์

ผลการทำนายผู้ป่วยผู้ป่วยยืนยัน COVID-19 จาก Model LSTM โดยใช้ข้อมูลประเทศไทย  
และคู่ประเทศฟิลิปปินส์

No.	RMSE	MAE	MAPE
1	1208.256832	1063.26242	24.594313
2	563.283168	468.50055	8.89327
3	834.127296	681.054278	11.541376
4	1020.927696	836.092742	13.618904
5	1560.030579	1324.517465	21.199383
6	1400.039878	1216.104994	23.744287
7	833.248966	692.496517	12.175543
8	695.739388	564.610207	10.036626
9	829.760987	692.233165	17.532911
10	790.730979	649.077007	11.312507
11	934.594677	782.573086	13.611235
12	714.674538	584.966821	10.957096
13	1118.979923	935.665777	15.742993
14	1635.10791	1403.699339	22.7066
15	1254.692	1084.669611	18.709698
16	532.492232	419.940017	8.486022
17	945.094709	782.822214	13.250513
18	592.398986	490.28246	9.219277
19	939.31966	781.925675	13.428928
20	636.278441	527.416641	10.187686
21	1284.816565	1061.32915	16.827943
22	1009.480281	834.062875	13.854737
23	562.024993	463.032539	8.731606
24	876.557869	735.424689	13.114726



2927129371

CU Thesais 6272091321 thesais / recv: 25122565 00:06:46 / seq: 30

25	681.381017	564.288567	13.305708
26	1212.309579	999.292989	15.962568
27	717.434177	587.663471	10.430806
28	554.579994	451.661862	8.687764
29	527.39762	415.375306	8.447497
30	805.782896	660.859169	11.411992

ตารางที่ 19 แสดง Metric จากการทำนาย LSTM โดยใช้ข้อมูลประเทศไทยและคู่ประเทศฟิลิปปินส์

ผลจากการทำ Comparing Mean โดย T-Test Independent เมื่อเทียบกับการใช้ข้อมูลประเทศไทยอย่างเดียวจะได้ผลดังนี้

#### 1. RMSE

	country	N	Mean	Std. Deviation	Std. Error Mean
rmse	T	30	919.484889	256.2367342	46.7822131
	Ph	30	909.051461	307.7539901	56.1879342

ตารางที่ 20 แสดงภาพรวมสถิติ RMSE แสดงภาพรวมสถิติ RMSE ผลการทำนายระหว่างประเทศไทยและคู่ประเทศไทยฟิลิปปินส์

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
rmse	Equal variances assumed	1.387	.244	.143	58	.887	10.4334275	73.1140165	-135.9201777	156.7870327
	Equal variances not assumed			.143	56.157	.887	10.4334275	73.1140165	-136.0225314	156.8893865

ตารางที่ 21 ผลจากทดสอบ T-Test

## 2. MAE

**Group Statistics**

	country	N	Mean	Std. Deviation	Std. Error Mean
mae	T	30	786.455288	252.3924178	46.0803402
	Ph	30	758.496720	271.2992409	49.5322380

ตารางที่ 22 แสดงภาพรวมสถิติ MAE ผลการทำนายระหว่างประเทศไทยและคู่ประเทศไทยฟิลิปปินส์

**Independent Samples Test**

		Levene's Test for Equality of Variances		t-test for Equality of Means			95% Confidence Interval of the Difference			
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
mae	Equal variances assumed	.392	.534	.413	58	.681	27.9585683	67.6523492	-107.4623220	163.3794586
	Equal variances not assumed			.413	57.700	.681	27.9585683	67.6523492	-107.4773171	163.3944537

ตารางที่ 23 ผลจากทดสอบ T-Test

## 3. MAPE

**Group Statistics**

	country	N	Mean	Std. Deviation	Std. Error Mean
mape	T	30	17.210787	5.4900224	1.0023364
	Ph	30	13.724151	4.6469975	.8484218

ตารางที่ 24 แสดงภาพรวมสถิติ MAPE ผลการทำนายระหว่างประเทศไทยและคู่ประเทศไทยฟิลิปปินส์

**Independent Samples Test**

		Levene's Test for Equality of Variances		t-test for Equality of Means			95% Confidence Interval of the Difference			
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
mape	Equal variances assumed	.786	.379	2.655	58	.010	3.4866364	1.3132013	.8579783	6.1152944
	Equal variances not assumed			2.655	56.459	.010	3.4866364	1.3132013	.8564503	6.1168224

ตารางที่ 25 ผลจากทดสอบ T-Test

### 5.3.5 ประเทศไทยและคู่ประเทศบังกลาเทศ

ผลการทำนายผู้ป่วยผู้ป่วยยืนยัน COVID-19 จาก Model LSTM โดยใช้ข้อมูลประเทศไทย  
และคู่ประเทศบังกลาเทศ

No.	RMSE	MAE	MAPE
1	539.119421	444.484289	8.792037
2	735.011766	618.092987	11.628374
3	610.206652	495.532615	11.337346
4	972.272844	829.5129	14.933732
5	708.081323	563.321134	12.008853
6	662.285729	539.766516	9.970086
7	772.287962	625.276869	10.7934
8	684.823587	553.422518	9.985021
9	922.008973	745.694759	12.165009
10	928.073905	785.466876	13.960482
11	591.594045	489.181592	9.268133
12	874.581838	745.588415	13.86878
13	946.901189	803.877242	14.350265
14	1133.643881	1002.285406	18.738763
15	1194.743748	1033.88348	18.194571
16	607.863378	460.330037	9.752955
17	838.515177	704.203545	13.672412
18	1178.15246	1017.175539	17.904061
19	722.397647	574.20371	12.082324
20	1177.22324	1039.829065	19.265051
21	929.877297	799.98509	14.91873
22	716.049478	598.986476	11.297062
23	623.438736	480.839723	10.473604
24	557.073367	417.894407	8.525851



2927129371

CU Thesais 6272091321 thesais / recv: 25122565 00:06:46 / seq: 30

25	1079.816521	937.374242	17.109032
26	978.983693	787.938628	12.69092
27	521.854201	416.32883	8.291887
28	668.554888	539.857119	9.644764
29	762.468725	619.530377	13.202104
30	1032.688242	903.760374	17.063726

ตารางที่ 26 แสดง Metric จากการทำนาย LSTM โดยใช้ข้อมูลประเทศไทยและคู่ประเทศบังกลาเทศ

ผลจากการทำ Comparing Mean โดย T-Test Independent เมื่อเทียบกับการใช้ข้อมูลประเทศไทยอย่างเดียวจะได้ผลดังนี้

### 1. RMSE

	country	N	Mean	Std. Deviation	Std. Error Mean
rmse	T	30	919.4849	256.23673	46.78221
	Ba	30	822.3531	206.59760	37.71939

ตารางที่ 27 แสดงภาพรวมสถิติ RMSE ผลการทำนายระหว่างประเทศไทยและคู่ประเทศไทยบังกลาเทศ

		Levene's Test for Equality of Variances					t-test for Equality of Means			
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
rmse	Equal variances assumed	.244	.623	1.616	58	.111	97.13176	60.09432	-23.16010	217.42362
	Equal variances not assumed			1.616	55.504	.112	97.13176	60.09432	-23.27538	217.53890

ตารางที่ 28 ผลจากทดสอบ T-Test

## 2. MAE

	country	N	Mean	Std. Deviation	Std. Error Mean
mae	T	30	786.455288	252.3924178	46.0803402
	Ba	30	685.787492	196.5981968	35.8937557

ตารางที่ 29 แสดงภาพรวมสถิติ MAE แสดงภาพรวมสถิติ RMSE ผลการทำนายระหว่างประเทศไทยและคู่ประเทศไทยบังกลาเทศ

		Levene's Test for Equality of Variances					t-test for Equality of Means		95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
mae	Equal variances assumed	.391	.534	1.723	58	.090	100.6677964	58.4102684	-16.2530591	217.5886518
	Equal variances not assumed			1.723	54.722	.090	100.6677964	58.4102684	-16.4023676	217.7379603

ตารางที่ 30 ผลจากทดสอบ T-Test

## 3. MAPE

	country	N	Mean	Std. Deviation	Std. Error Mean
mape	T	30	17.210787	5.4900224	1.0023364
	Ba	30	12.862978	3.2311578	.5899260

ตารางที่ 31 แสดงภาพรวมสถิติ MAPE ผลการทำนายระหว่างประเทศไทยและคู่ประเทศไทยบังกลาเทศ



		Levene's Test for Equality of Variances		Independent Samples Test			t-test for Equality of Means		95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
mape	Equal variances assumed	5.583	.022	3.738	58	.000	4.3478090	1.1630524	2.0197067	6.6759114
	Equal variances not assumed			3.738	46.938	.001	4.3478090	1.1630524	2.0079683	6.6876498

ตารางที่ 32 ผลจากทดสอบ T-Test

5.3.6 ประเทศไทยและคู่ประเทศคิวบา

ผลการทำนายผู้ป่วยผู้ป่วยยืนยัน COVID-19 จาก Model LSTM โดยใช้ข้อมูลประเทศไทย และคู่ประเทศคิวบา

No.	RMSE	MAE	MAPE
1	1543.82859	1408.524705	26.512553
2	2120.190774	1981.951421	36.893834
3	1149.788448	943.318621	15.302613
4	1202.437788	1061.736524	19.639639
5	885.730632	733.858312	13.83059
6	1078.510864	867.519524	14.5657
7	1409.672401	1206.133143	20.28062
8	1719.323205	1575.79558	29.56131
9	841.258021	659.706804	11.300593
10	1679.447735	1540.475397	28.757611
11	860.319576	699.922078	11.7561
12	746.253496	614.598365	11.444831
13	1485.25632	1337.205037	24.202235
14	1238.493216	1071.626008	19.095773
15	2167.422437	1921.196657	32.026311
16	1558.151593	1356.521359	23.114475
17	1566.14637	1413.323105	25.671022

18	2115.633402	1863.332	30.812364
19	1116.269588	906.365637	15.008793
20	2160.542447	1991.988895	35.734634
21	1149.745701	1016.168691	19.999201
22	1173.287919	1016.9957	19.348004
23	1866.26982	1732.750462	32.899344
24	1240.041689	1103.930879	20.7045
25	1239.702405	1086.383856	19.531312
26	1401.035456	1217.673555	21.009915
27	1228.039986	1056.158002	18.362058
28	1460.853036	1271.510957	21.898324
29	2158.237079	2024.066654	39.846594
30	1221.257384	1040.889213	18.666039

ตารางที่ 33 แสดง Metric จากการทำนาย LSTM โดยใช้ข้อมูลประเทศไทยและคู่ประเทศคิวบา

ผลจากการทำ Comparing Mean โดย T-Test Independent เมื่อเทียบกับการใช้ข้อมูลประเทศไทยอย่างเดียวจะได้ผลดังนี้

#### 1. RMSE

	country	N	Mean	Std. Deviation	Std. Error Mean
rmse	T	30	919.4849	256.23673	46.78221
	Cu	30	1426.1049	418.04941	76.32503

ตารางที่ 34 แสดงภาพรวมสถิติ RMSE ผลการทำนายระหว่างประเทศไทยและคู่ประเทศไทยคิวบา

Independent Samples Test										
		Levene's Test for Equality of Variances					t-test for Equality of Means		95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
rmse	Equal variances assumed	7.284	.009	-5.659	58	.000	-506.62002	89.52143	-686.81663	-327.42342
	Equal variances not assumed			-5.659	48.095	.000	-506.62002	89.52143	-686.60575	-326.63430

ตารางที่ 35 ผลจากทดสอบ T-Test

## 2. MAE

Group Statistics						
		country	N	Mean	Std. Deviation	Std. Error Mean
mae	T		30	786.455288	252.3924178	46.0803402
	Cu		30	1257.387571	414.9579858	75.7606164

ตารางที่ 36 แสดงภาพรวมสถิติ MAE ผลการทำนายระหว่างประเทศไทยและคู่ประเทศไทยคิวบา

Independent Samples Test										
		Levene's Test for Equality of Variances					t-test for Equality of Means		95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
mae	Equal variances assumed	8.175	.006	-5.311	58	.000	-470.9322830	88.6739463	-648.4324717	-293.4320943
	Equal variances not assumed			-5.311	47.874	.000	-470.9322830	88.6739463	-649.2353251	-292.6292409

ตารางที่ 37 ผลจากทดสอบ T-Test

## 3. MAPE

Group Statistics						
		country	N	Mean	Std. Deviation	Std. Error Mean
mape	T		30	17.2108	5.49002	1.00234
	Cu		30	22.5926	7.80332	1.42468

ตารางที่ 38 แสดงภาพรวมสถิติ MAPE ผลการทำนายระหว่างประเทศไทยและคู่ประเทศไทยคิวบา

		Levene's Test for Equality of Variances					t-test for Equality of Means		95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
mape	Equal variances assumed	4.077	.048	-3.090	58	.003	-5.38178	1.74195	-8.86868	-1.89488
	Equal variances not assumed			-3.090	52.059	.003	-5.38178	1.74195	-8.87717	-1.88638

ตารางที่ 39 ผลจากทดสอบ T-Test

5.3.7 ประเทศไทยและคู่ประเทศอิรัก

ผลการทำนายผู้ป่วยผู้ป่วยยืนยัน COVID-19 จาก Model LSTM โดยใช้ข้อมูลประเทศไทย และคู่ประเทศอิรัก

No.	RMSE	MAE	MAPE
1	598.526034	503.767152	9.827753
2	612.031477	499.578775	9.176634
3	1046.188282	921.839012	17.645438
4	572.180715	480.234347	9.435907
5	999.640892	859.284126	15.744999
6	649.462709	540.690447	10.206924
7	715.203165	600.569778	11.443292
8	703.422101	565.729764	10.257689
9	578.661005	434.070116	8.931896
10	1097.867436	948.578195	17.021711
11	1772.002757	1633.861307	29.807194
12	695.953639	561.631698	13.035763
13	746.052232	623.985154	11.562086
14	758.88666	615.913691	11.080804
15	1447.727338	1168.331079	17.813106

16	666.418637	558.294061	10.712538
17	593.857619	491.717079	10.927217
18	850.423448	715.148125	13.486281
19	1025.593315	892.724511	16.712661
20	1201.913301	1057.556916	19.351371
21	584.048169	493.354879	9.826169
22	1113.20354	960.449007	17.178131
23	626.59486	505.176909	9.58639
24	955.906035	821.90844	15.232235
25	761.832648	625.012049	10.926933
26	871.462535	724.547491	12.617433
27	645.04491	541.416824	10.494547
28	693.701129	572.746826	10.477331
29	1033.374934	912.296648	17.778356
30	1275.851358	1146.264277	21.561101

ตารางที่ 40 แสดง Metric จากการทำนาย LSTM โดยใช้ข้อมูลประเทศไทยและคู่ประเทศอิรัก

ผลจากการทำ Comparing Mean โดย T-Test Independent เมื่อเทียบกับการใช้ข้อมูลประเทศไทยอย่างเดียวจะได้ผลดังนี้

#### 1. RMSE

Group Statistics					
	country	N	Mean	Std. Deviation	Std. Error Mean
rmse	T	30	919.4849	256.23673	46.78221
	Ir	30	863.1011	291.41083	53.20410

ตารางที่ 41 แสดงภาพรวมสถิติ RMSE ผลการทำนายระหว่างประเทศไทยและคู่ประเทศไทยอิรัก

		Levene's Test for Equality of Variances		t-test for Equality of Means			95% Confidence Interval of the Difference			
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
rmse	Equal variances assumed	.736	.394	.796	58	.429	56.38379	70.84667	-85.43123	198.19882
	Equal variances not assumed			.796	57.066	.429	56.38379	70.84667	-85.48066	198.24825

ตารางที่ 42 ผลจากทดสอบ T-Test

## 2. MAE

		country	N	Mean	Std. Deviation	Std. Error Mean
mae	T		30	786.455288	252.3924178	46.0803402
	Ir		30	732.555956	273.0930407	49.8597396

ตารางที่ 43 แสดงภาพรวมสถิติ MAE ผลการทำนายระหว่างประเทศไทยและคู่ประเทศไทยอิรัก

		Levene's Test for Equality of Variances		t-test for Equality of Means			95% Confidence Interval of the Difference			
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
mae	Equal variances assumed	.429	.515	.794	58	.430	53.8993323	67.8924987	-82.0022695	189.8009340
	Equal variances not assumed			.794	57.643	.431	53.8993323	67.8924987	-82.0201772	189.8188417

ตารางที่ 44 ผลจากทดสอบ T-Test

## 3. MAPE

		country	N	Mean	Std. Deviation	Std. Error Mean
mape	T		30	17.2108	5.49002	1.00234
	Ir		30	13.6620	4.68485	.85533

ตารางที่ 45 แสดงภาพรวมสถิติ MAPE ผลการทำนายระหว่างประเทศไทยและคู่ประเทศไทยอิรัก

Independent Samples Test										
		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
mape	Equal variances assumed	.545	.463	2.693	58	.009	3.54879	1.31768	.91117	6.18641
	Equal variances not assumed			2.693	56.600	.009	3.54879	1.31768	.90979	6.18780

ตารางที่ 46 ผลจากทดสอบ T-Test

5.3.8 ประเทศไทยและคู่ประเทศเม็กซิโก

ผลการทำนายผู้ป่วยผู้ป่วยยืนยัน COVID-19 จาก Model LSTM โดยใช้ข้อมูลประเทศไทย และคู่ประเทศเม็กซิโก

No.	RMSE	MAE	MAPE
1	788.685612	642.10148	13.407093
2	1000.66335	850.689402	15.008083
3	525.584139	412.513737	8.341461
4	719.722698	569.29313	12.082845
5	966.929993	828.436462	19.217097
6	565.812859	444.146223	9.463547
7	1111.034542	966.590872	18.891574
8	526.903562	411.702819	8.49416
9	523.712113	409.62227	8.448081
10	630.712281	517.391963	9.597405
11	609.516525	488.615482	8.912513
12	756.27383	610.377558	10.417299
13	1536.833306	1401.91105	27.302758
14	739.969154	601.408052	13.719088
15	574.908963	473.559942	9.70052
16	543.059448	440.955635	8.900409

17	543.468757	434.56446	8.782266
18	720.452311	596.597202	10.70539
19	902.021641	768.026374	17.00045
20	833.073214	673.33646	11.152738
21	939.359876	803.873786	17.982229
22	558.748929	423.522203	8.766713
23	598.153097	499.890297	9.801784
24	937.644432	799.921414	17.188176
25	914.100597	765.431317	15.445126
26	527.203141	417.292635	8.504357
27	1037.337927	887.801184	21.220154
28	698.549912	569.354977	10.735905
29	530.636891	428.551359	8.194599
30	1626.622456	1539.39796	32.765911

ตารางที่ 47 แสดง Metric จากการทำนาย LSTM โดยใช้ข้อมูลประเทศไทยและคู่ประเทศเม็กซิโก

ผลจากการทำ Comparing Mean โดย T-Test Independent เมื่อเทียบกับการใช้ข้อมูลประเทศไทยอย่างเดียวจะได้ผลดังนี้

#### 1. RMSE

	country	N	Mean	Std. Deviation	Std. Error Mean
rmse	T	30	919.4849	256.23673	46.78221
	Me	30	782.9232	281.90836	51.46919

ตารางที่ 48 แสดงภาพรวมสถิติ RMSE ผลการทำนายระหว่างประเทศไทยและคู่ประเทศไทยเม็กซิโก



Independent Samples Test										
		Levene's Test for Equality of Variances					t-test for Equality of Means			
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
rmse	Equal variances assumed	.164	.687	1.963	58	.054	136.56170	69.55324	-2.66423	275.78763
	Equal variances not assumed			1.963	57.479	.054	136.56170	69.55324	-2.69109	275.81450

ตารางที่ 49 ผลจากทดสอบ T-Test

## 2. MAE

Group Statistics						
		country	N	Mean	Std. Deviation	Std. Error Mean
mae	T		30	786.455288	252.3924178	46.0803402
	Me		30	655.895924	277.5073314	50.6656751

ตารางที่ 50 แสดงภาพรวมสถิติ MAE ผลการทำนายระหว่างประเทศไทยและคู่ประเทศไทยเม็กซิโก

Independent Samples Test										
		Levene's Test for Equality of Variances					t-test for Equality of Means			
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
mae	Equal variances assumed	.101	.751	1.906	58	.062	130.5593649	68.4865562	-6.5313722	267.6501019
	Equal variances not assumed			1.906	57.486	.062	130.5593649	68.4865562	-6.5574864	267.6762161

ตารางที่ 51 ผลจากทดสอบ T-Test

## 3. MAPE

Group Statistics						
		country	N	Mean	Std. Deviation	Std. Error Mean
mape	T		30	17.2108	5.49002	1.00234
	Me		30	13.3383	5.99881	1.09523

ตารางที่ 52 แสดงภาพรวมสถิติ MAPE ผลการทำนายระหว่างประเทศไทยและคู่ประเทศไทยเม็กซิโก

		Levene's Test for Equality of Variances		Independent Samples Test			t-test for Equality of Means		95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
mape	Equal variances assumed	.119	.731	2.608	58	.012	3.87246	1.48466	.90060	6.84432
	Equal variances not assumed			2.608	57.550	.012	3.87246	1.48466	.90011	6.84482

ตารางที่ 53 ผลจากทดสอบ T-Test

5.3.9 ประเทศไทยและคู่ประเทศเวียดนาม

ผลการทำนายผู้ป่วยผู้ป่วยยืนยัน COVID-19 จาก Model LSTM โดยใช้ข้อมูลประเทศไทย และคู่ประเทศเวียดนาม

No.	RMSE	MAE	MAPE
1	3868.640489	3662.347718	86.728317
2	3575.361845	3353.637633	80.274681
3	3893.720878	3701.258603	87.032779
4	2198.42945	2029.814281	48.365699
5	2749.496879	2569.567562	61.447151
6	3326.958087	3162.645026	74.188772
7	2040.301614	1922.361026	44.51023
8	2867.070503	2681.136636	63.779667
9	2594.370784	2395.848138	57.903816
10	2059.197807	1869.484286	46.415426
11	2673.936709	2492.187218	60.069226
12	3580.178473	3371.94631	80.840765
13	3456.680693	3310.551744	77.48292
14	3628.035565	3342.425706	81.96666
15	4205.828645	4040.276099	94.489594
16	3202.699149	3000.876726	71.736658

17	4029.275485	3908.336419	89.077869
18	4463.965427	4289.27319	99.692233
19	4230.625104	4027.340807	95.44989
20	5034.469381	4832.923814	113.292718
21	1987.322348	1822.303993	43.934637
22	3802.465904	3607.079335	85.005172
23	3612.132452	3352.818373	81.347652
24	2924.786605	2777.293739	65.006132
25	3047.454966	2866.545905	68.450163
26	2288.731603	2105.531319	50.943879
27	2773.187426	2594.972897	62.494041
28	3530.16972	3315.075966	79.249913
29	2492.682392	2335.177088	55.19749
30	4073.75925	3866.163533	91.600598

ตารางที่ 54 แสดง Metric จากการทำนาย LSTM โดยใช้ข้อมูลประเทศไทยและคู่ประเทศเวียดนาม

ผลจากการทำ Comparing Mean โดย T-Test Independent เมื่อเทียบกับการใช้ข้อมูลประเทศไทยอย่างเดียวจะได้ผลดังนี้

#### 1. RMSE

	country	N	Mean	Std. Deviation	Std. Error Mean
rmse	T	30	919.4849	256.23673	46.78221
	Vi	30	3273.7312	790.78576	144.37707

ตารางที่ 55 แสดงภาพรวมสถิติ RMSE ผลการทำนายระหว่างประเทศไทยและคู่ประเทศไทยเวียดนาม

		Levene's Test for Equality of Variances					t-test for Equality of Means			
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
rmse	Equal variances assumed	33.026	.000	-15.512	58	.000	-2354.24630	151.76730	-2658.04155	-2050.45105
	Equal variances not assumed			-15.512	35.023	.000	-2354.24630	151.76730	-2662.34298	-2046.14962

ตารางที่ 56 ผลจากทดสอบ T-Test

## 2. MAE

Group Statistics						
		country	N	Mean	Std. Deviation	Std. Error Mean
mae	T		30	786.455288	252.3924178	46.0803402
	Vi		30	3086.906703	780.9240774	142.5765776

ตารางที่ 57 แสดงภาพรวมสถิติ MAE ผลการทำนายระหว่างประเทศไทยและคู่ประเทศไทยเวียดนาม

		Levene's Test for Equality of Variances					t-test for Equality of Means			
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
mae	Equal variances assumed	31.927	.000	-15.353	58	.000	-2300.45141	149.8381735	-2600.38511	-2000.51772
	Equal variances not assumed			-15.353	34.993	.000	-2300.45141	149.8381735	-2604.64123	-1996.26160

ตารางที่ 58 ผลจากทดสอบ T-Test

## 3. MAPE

Group Statistics						
		country	N	Mean	Std. Deviation	Std. Error Mean
mape	T		30	17.2108	5.49002	1.00234
	Vi		30	73.2658	17.94850	3.27693

ตารางที่ 59 แสดงภาพรวมสถิติ MAPE ผลการทำนายระหว่างประเทศไทยและคู่ประเทศไทยเวียดนาม

		Levene's Test for Equality of Variances		t-test for Equality of Means			95% Confidence Interval of the Difference			
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
mape	Equal variances assumed	34.778	.000	-16.358	58	.000	-56.05504	3.42680	-62.91453	-49.19555
	Equal variances not assumed			-16.358	34.379	.000	-56.05504	3.42680	-63.01631	-49.09377

ตารางที่ 60 ผลจากทดสอบ T-Test

ประเทศ	RMSE	MAE	MAPE
Thailand	919.4848887	786.4552884	17.21078687
Thailand + Japan	762.8507558*	633.8684829*	13.22562983*
Thailand + Malaysia	973.4786635	792.0097974	13.89061407*
Thailand + Philippines	909.0514612	758.4967201	13.7241505*
Thailand + Bangladesh	822.3531304	685.787492	12.86297783*
Thailand + Cuba	1426.104913*	1257.387571*	22.59256307*
Thailand + Iraq	863.101096	732.5559561	13.66199633*
Thailand + Mexico	782.9231852	655.8959235	13.33832437*
Thailand + Vietnam	3273.731188*	3086.906703*	73.26582493*

ตารางที่ 61 แสดงค่าเฉลี่ยของการทดลอง 30 ครั้งทุก Metric ทำนายผู้ป่วยประเทศไทย

(\* หมายถึงแตกต่างกันอย่างมีนัยสำคัญเมื่อเทียบกับใช้ข้อมูลประเทศไทย)

จากผลของ Model LSTM จะเห็นได้ว่าการใช้ข้อมูลคู่ประเทศที่มีค่า Error บน Metric ทั้ง 3 RMSE, MAE และ MAPE น้อยกว่าประเทศไทยอย่างเดียวได้แก่ คู่ประเทศญี่ปุ่น, ฟิลิปปินส์, บางกาทเทศ, อิรัก, เม็กซิโก ส่วนประเทศที่มากกว่าประเทศไทยทั้ง 3 Metrics ได้แก่ ประเทศคิวบา และเวียดนาม สำหรับมาเลเซียเป็นประเทศที่มีค่า MAPE น้อยกว่าประเทศไทย ส่วนประเทศคิวบา ในเดือนสิงหาคมและกันยายนกลับมาเป็นทรนขาขึ้นอีกครั้ง ซึ่งสวนทางกับประเทศไทยที่ประเทศไทยนั้นอยู่ในช่วงขาลงตั้งแต่ช่วงเดือนสิงหาคมตลอดมา จึงส่งผลต่อค่า Metrics โดยรวมทำให้ ในส่วนประเทศเวียดนามในเดือนตุลาคมจะมีการทรนขาขึ้นอีกครั้งนอก ซึ่งสวนทางกับประเทศ

ไทยที่กำลังอยู่ในช่วงทรนกลางอย่างชัดเจน จึงทำให้ Testing Set ของเวียดนามนั้นสวนทางกับประเทศไทยและส่งผลกระทบต่อ Metrics ให้ได้น้อยลงในทุก Metrics ในส่วนประเทศที่เหลือคือญี่ปุ่น, ฟิลิปปินส์, มาเลเซีย, บางกาทเทศ, อิรัก และ เม็กซิโก โดยรวมแล้วอยู่ในช่วงผู้ป่วยาลงจึงทำให้ Metrics มีค่าน้อยกว่าเมื่อเทียบกับการใช้ข้อมูลประเทศไทยอย่างเดียว สำหรับ Metrics MAPE ของประเทศญี่ปุ่น มาเลเซีย ฟิลิปปินส์ บางกาทเทศ อิรัก และเม็กซิโก ที่มีค่าน้อยกว่า การใช้ประเทศไทยอย่างเดียวนั้น เมื่อนำมาเปรียบเทียบกับตนเองแล้วพบว่า ไม่มีความแตกต่างอย่างมีนัยสำคัญ ส่วนหากนำมาเปรียบเทียบกับประเทศ คิวบา และเวียดนามแล้วพบว่าประเทศญี่ปุ่น มาเลเซีย ฟิลิปปินส์ บางกาทเทศ อิรัก และเม็กซิโก มีความแตกต่างกันอย่างมีนัยสำคัญ

#### 5.4 แนวทางวิจัยในอนาคต

เนื่องจากข้อมูลของทาง John Hopskin นั้นมีข้อมูลผู้ป่วย COVID-19 โดยไม่แยกแต่ละสายพันธุ์ของโรค จึงอาจมีข้อจำกัดในด้านชนิดของสายพันธุ์ของโรคที่ต่างกัน แนวทางการพัฒนาในด้านข้อมูลสามารถทำได้ด้วยการหาแหล่งข้อมูลโดยเพิ่มข้อมูลทางด้านสายพันธุ์ของผู้ที่ติดเชื้อจึงจะทำให้ Model สามารถเรียนรู้พฤติกรรมและความแตกต่างของแต่ละสายพันธุ์ได้ ซึ่งนำไปถึงการเพิ่มขนาด Data Set มากกว่า 2 ปีที่ใช้เนื่องด้วยข้อมูลมีปริมาณอย่างจำกัด การใช้เทคนิคเช่นการทำ Data Augmentation เพื่อเพิ่มปริมาณข้อมูลให้ Model เรียนรู้พฤติกรรมมากขึ้น ในการใช้ Model Machine Learning หรือ Deep learning ที่ใช้ในการทำด้าน Time Series อื่นๆ มากขึ้นเพื่อหาโครงสร้างของ Neural Network ที่เหมาะสมกับข้อมูล



2927129371



CU IThesis 6272091321 thesis / recv: 25122565 00:06:46 / seq: 30

2927129371

## บรรณานุกรม

1. Archived: WHO Timeline - COVID-19. 2020; Available from:  
<https://www.who.int/news/item/27-04-2020-who-timeline---covid-19>.
2. Warittha, P., Suchanan, C. *Tourism at a crossroad: อนาคตภาคการท่องเที่ยวไทย เดินต่ออย่างไรในฟ้าหลังฝน*. 2020; Available from:  
[https://www.bot.or.th/Thai/ResearchAndPublications/articles/Pages/Article\\_18Aug2021.aspx](https://www.bot.or.th/Thai/ResearchAndPublications/articles/Pages/Article_18Aug2021.aspx).
3. Ostertagová, E., *Modelling using polynomial regression*. Procedia Engineering, 2012. **48**: p. 500-506.
4. Fattah, J., et al., *Forecasting of demand using ARIMA model*. International Journal of Engineering Business Management, 2018. **10**: p. 1847979018808673.
5. Sherstinsky, A., *Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network*. Physica D: Nonlinear Phenomena, 2020. **404**: p. 132306.
6. Education, I.C. *Machine Learning*. 2020; Available from:  
<https://www.ibm.com/cloud/learn/machine-learning>.
7. MacQueen, J. *Some methods for classification and analysis of multivariate observations*. 1967.
8. Arthur, D. and S. Vassilvitskii, *k-means++: The Advantages of Careful Seeding*. 2006, Stanford InfoLab.
9. Le, Q.V., et al. *On optimization methods for deep learning*. in *International Conference on Machine Learning*. 2011.



2927129371

CU Thesisis 6272091321 thesisis / recv: 25122565 00:06:46 / seq: 30



10. Kingma, D.P. and J. Ba, *Adam: A Method for Stochastic Optimization*. CoRR, 2014. **abs/1412.6980**.
11. Keras. *ReduceLRonPlateau*. Available from: [https://keras.io/api/callbacks/reduce\\_lr\\_on\\_plateau/](https://keras.io/api/callbacks/reduce_lr_on_plateau/).
12. Punn, N.S., S.K. Sonbhadra, and S. Agarwal, *COVID-19 epidemic analysis using machine learning and deep learning algorithms*. MedRxiv, 2020.
13. Arora, P., H. Kumar, and B.K. Panigrahi, *Prediction and analysis of COVID-19 positive cases using deep learning models: A descriptive case study of India*. Chaos, Solitons & Fractals, 2020. **139**: p. 110017.
14. Javid, A.M., et al., *Predictive analysis of COVID-19 time-series data from Johns Hopkins University*. arXiv preprint arXiv:2005.05060, 2020.
15. Bansal, A. and U. Jayant. *Covid-19 Outbreak Modelling Using Regression Techniques*. in *2021 International Conference on Innovative Practices in Technology and Management (ICIPTM)*. 2021. IEEE.
16. Kurniawan, R., et al. *Clustering and correlation methods for predicting coronavirus COVID-19 risk analysis in pandemic countries*. in *2020 8th International Conference on Cyber and IT Service Management (CITSM)*. 2020. IEEE.
17. Baalbaki, H., et al., *Fighting against COVID-19: Who Failed and Who Succeeded?* Journal of Computer and Communications, 2022. **10(4)**: p. 32-50.
18. Said, A.B., et al., *Predicting COVID-19 cases using bidirectional LSTM on multivariate time series*. Environmental Science and Pollution Research, 2021. **28(40)**: p. 56043-56052.

19. Aragão, D.P., et al., *Multivariate data driven prediction of COVID-19 dynamics: Towards new results with temperature, humidity and air quality data.* Environmental Research, 2022. **204**: p. 112348.



2927129371

CU IThesis 6272091321 thesis / recv: 25122565 00:06:46 / seq: 30



CU IThesis 6272091321 thesis / recv: 25122565 00:06:46 / seq: 30

2927129371

## ประวัติผู้เขียน

ชื่อ-สกุล

สิริธีร์ วรรณทองดี

วัน เดือน ปี เกิด

2 กุมภาพันธ์ 2539



2927129371

CU Thesais 6272091321 thesis / recv: 25122565 00:06:46 / seq: 30