



การศึกษาประสิทธิภาพการพยากรณ์ฝุ่น PM2.5 ในเขตกรุงเทพมหานคร  
โดยแบบจำลองไฮบริดร่วมกับค่าถ่วงน้ำหนัก

ธัชกร พลเสนา

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิศวกรรมศาสตรมหาบัณฑิต  
สาขาวิชาวิศวกรรมข้อมูลขนาดใหญ่  
วิทยาลัยนวัตกรรมด้านเทคโนโลยีและวิศวกรรมศาสตร์  
มหาวิทยาลัยธุรกิจบัณฑิต  
ปีการศึกษา 2565

EFFICIENCY STUDY OF PM2.5 FORECASTING IN BANGKOK  
BY THE HYBRID MODEL WITH WEIGHTED VALUES

TATSHAKON POLSENA

A Thesis Submitted in Partial Fulfillment of the Requirements  
for the Degree of Master of Engineering  
Department of Big Data Engineering  
College of Innovative Technology and Engineering,  
Dhurakij Pundit University  
Academic Year 2022



หัวข้อวิทยานิพนธ์	การศึกษาประสิทธิภาพการพยากรณ์ฝุ่น PM2.5 ในเขตกรุงเทพมหานคร โดยแบบจำลองไฮบริดร่วมกับค่าถ่วงน้ำหนัก
ชื่อผู้เขียน	ธัชกร พลเสนา
อาจารย์ที่ปรึกษาวิทยานิพนธ์	ผู้ช่วยศาสตราจารย์ ดร. ดวงใจ จิตคงชื่น
สาขาวิชา	วิศวกรรมข้อมูลขนาดใหญ่
ปีการศึกษา	2565

### บทคัดย่อ

งานวิจัยนี้ทำการพยากรณ์ค่าฝุ่น PM2.5 กับ 4 สถานีในกรุงเทพมหานครประกอบด้วย สถานี 02T สถานี 11T สถานี 54T และสถานี 59T ซึ่งอยู่ภายใต้การดูแลของกรมควบคุมมลพิษ ประเทศไทย ด้วยแบบจำลองไฮบริด LSTM-ARIMA ร่วมกับแบบจำลองไฮบริด ARIMA-LSTM พร้อมทั้งมีการปรับค่าถ่วงน้ำหนักร่วมด้วย เพื่อพยากรณ์ค่าฝุ่น PM2.5 เนื่องจากการใช้แบบจำลองไฮบริด ARIMA-LSTM แต่เพียงอย่างเดียวมีประสิทธิภาพในการพยากรณ์ต่ำกว่าเมื่อเทียบกับค่าความคลาดเคลื่อนผลจากแบบจำลอง ARIMA

ผลการวิจัยพบว่าการพยากรณ์ด้วยแบบจำลอง LSTM แบบตัวแปรเดียว (Univariat) และ ARIMA ร่วมกับแบบจำลองไฮบริด ARIMA-LSTM พร้อมทั้งมีการปรับค่าถ่วงน้ำหนักร่วมด้วย ทั้ง 4 สถานีให้ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) ต่ำกว่า การพยากรณ์ด้วยแบบจำลอง ARIMA แต่หากพิจารณาจากค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Squared Error: RMSE) มีเพียงสถานี 02T เท่านั้นที่ให้ค่าความคลาดเคลื่อนต่ำกว่า นอกจากนี้ยังพบว่าสถานีที่ใกล้กันมีสหสัมพันธ์ในทิศทางบวก

**คำสำคัญ :** ฝุ่นละออง PM2.5, การพยากรณ์, แบบจำลองไฮบริด



(ผู้ช่วยศาสตราจารย์ ดร.ดวงใจ จิตคงชื่น)

อาจารย์ที่ปรึกษา

**Thesis Title** EFFICIENCY STUDY OF PM2.5 FORECASTING IN BANGKOK BY THE HYBRID MODEL WITH WEIGHTED VALUES.

**Author** TATSHAKON POLSENA

**Thesis Advisor** Asst. Prof. Duangjai Jitkongchuen, PhD.

**Department** Big Data Engineering

**Academic Year** 2022

### ABSTRACT

This research forecasts PM2.5 dust with 4 stations in Bangkok, consisting of stations 02T, 11T, 54T and 59T stations which are also under the supervision of the Pollution Control Department Thailand by LSTM-ARIMA hybrid model together with ARIMA-LSTM hybrid model with adjust the weight as well to predict the PM2.5 dust value because when usage only ARIMA-LSTM hybrid model is lower effective in compared to the error result from only ARIMA model.

The results showed that the predictions with the univariate LSTM- ARIMA hybrid Model combined with the ARIMA-LSTM hybrid model with weight adjustment. All 4 stations give the mean absolute error (MAPE) is lower than the prediction by the ARIMA model, but if considering the root mean squared error (RMSE), there is station 02T gave a lower RMSE. It was also found that the stations that are close to each other have positive correlation.

**Keywords** : PM2.5, Prediction, Hybrid Model



Asst.Prof.Dr. Duangjai Jitkongchuen

Advisor

## กิตติกรรมประกาศ

ขอขอบคุณอาจารย์ทุกท่านที่ทำหน้าที่การเป็นประธานและกรรมการในการจัดสอบวิทยานิพนธ์ฉบับนี้ โดยเริ่มจากที่ปรึกษาหลัก (อาจารย์ลูกไม้) ผู้ช่วยศาสตราจารย์ ดร.ดวงใจ จิตคงชื่น หากขาดบุคคลท่านนี้ งานวิจัยคงไม่สามารถผ่านกระบวนการจนเป็นรูปเป็นร่าง ขอคุณ อาจารย์เอก ดร.เอกสิทธิ์ พัชรวงศ์ศักดิ์ อาจารย์คนแรกที่เป็นผู้เปิดประตูโลกของวิศวกรรมข้อมูลขนาดใหญ่ ขอคุณอาจารย์ชาย ดร.ธนภัทร ชังคะจิตร ที่ร่วมสอนการวิเคราะห์อนุกรมเวลาด้วยภาษา LSTM ซึ่งเป็นจุดเริ่มต้นที่มาของวิจัยฉบับนี้ และขอคุณอาจารย์ปิ่น ดร.สรรพทุธิ มฤคทัต ที่สละเวลามาทำหน้าที่ประธานการสอบที่ใจดีมาก ก.ไก่ ล้านตัว

นอกจากอาจารย์ที่มีส่วนร่วมในการสอบแล้วต้องขอขอบคุณ พี่แดง คุณกุลธิตา รอดบุญ เลขาประจำภาควิชาที่คอยประสานงานและมีคำตอบทุกครั้งเมื่อเกิดข้อสงสัย ขอคุณ อาจารย์บอส ดร.รัฐศิลป์ ธานอกภานุวัชร ที่แม้จะไม่ได้เป็นกรรมการในการสอบ แต่ทุกวิชาที่อาจารย์สอน ทรงคุณค่าและมีความน่าสนใจ สมกับการเป็นวิศวกรรมข้อมูล และขอคุณ อาจารย์ณัฐ ผศ.ดร.ณัฐพัชร อารีรัชกุลกานต์ อาจารย์อีกท่านที่สอนการวิเคราะห์อนุกรมเวลาด้วยภาษา R และตอบไลน์ทุกครั้งที่มีข้อสงสัยเกี่ยวกับอนุกรมเวลา

ขอคุณมหาวิทยาลัยธุรกิจบัณฑิตที่เปิดสอนหลักสูตรที่มีความน่าสนใจและทันต่อความต้องการของตลาดงาน สุดท้ายขอขอบคุณตนเองที่อดทนและไม่ทิ้งการเรียนไปเสียก่อนที่จะสามารถเรียนรู้จนสำเร็จ

ธัชกร พลเสนา

## สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ง
บทคัดย่อภาษาอังกฤษ.....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญตาราง.....	ณ
สารบัญภาพ.....	ญ
บทที่	
1. บทนำ.....	1
1.1 ที่มาและความสำคัญของปัญหา.....	1
1.2 วัตถุประสงค์ของงานวิจัย.....	3
1.3 ขอบเขตงานวิจัย.....	3
1.4 ประโยชน์ที่คาดว่าจะได้รับ.....	3
2. แนวคิด ทฤษฎี และงานวิจัยที่เกี่ยวข้อง.....	4
2.1 ความรู้เบื้องต้นเกี่ยวกับฝุ่น PM2.5.....	4
2.2 การพยากรณ์ด้วยอนุกรมเวลา.....	7
2.3 แบบจำลอง ARIMA.....	9
2.4 แบบจำลอง LSTM.....	10
2.5 แบบจำลอง Hybrid.....	13
2.6 การเลือกแบบจำลองการเรียนรู้ของเครื่องที่ดีที่สุดสำหรับข้อมูลอนุกรม.....	15
2.7 การเข้ารหัสคุณลักษณะตามวัฏจักร.....	17
2.8 การคำนวณระยะทางระหว่างจุดละติจูดและลองจิจูด.....	18
2.9 การวัดแนวโน้มเข้าสู่ส่วนกลาง.....	19
2.10 งานวิจัยที่เกี่ยวข้อง.....	21
3. ระเบียบวิธีวิจัย.....	24
3.1 กำหนดกลุ่มตัวอย่าง.....	23
3.2 การออกแบบการทดลอง.....	25
3.3 ขั้นตอนการทดลอง.....	25
3.4 การเตรียมข้อมูล.....	29

สารบัญ (ต่อ)

	หน้า
3.5 เกณฑ์ที่ใช้ในการประเมินแบบจำลอง.....	31
3.6 เครื่องมือที่ใช้ในการวิจัย.....	31
4. ผลการวิจัย.....	33
4.1 ผลการเปรียบเทียบแนวโน้มของปริมาณฝุ่น PM2.5 ของแบบจำลอง.....	35
4.2 ผลการศึกษาความสัมพันธ์ระหว่างสถานีวัดปริมาณฝุ่น PM2.5.....	67
5. สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ.....	70
5.1 อภิปรายผล.....	70
5.2 ข้อเสนอแนะ.....	71
บรรณานุกรม.....	72
ภาคผนวก.....	75
ภาคผนวก ก หนังสือขอข้อมูลเพื่อประกอบการทำวิทยานิพนธ์.....	75
ภาคผนวก ข ผลงานตีพิมพ์ 1 .....	77
ภาคผนวก ค ผลงานตีพิมพ์ 2 .....	85
ประวัติผู้เขียน.....	101



## สารบัญตาราง

ตารางที่	หน้า
2.1	6
2.2	7
3.1	24
3.2	25
3.3	25
4.1	35
4.2	43
4.3	51
4.4	59
4.5	67

## สารบัญภาพ

ภาพที่	หน้า
1.1 เปรียบเทียบสถาปัตยกรรมโครงข่ายประสาทเทียม.....	2
2.1 เครือข่ายการติดตามและเฝ้าระวังคุณภาพอากาศของประเทศไทย.....	5
2.2 การแสดงอนุกรมเวลาเป็นเอาต์พุตจากตัวกรองเชิงเส้น.....	9
2.3 แผนภูมิองค์ประกอบ LSTM.....	11
2.4 การคำนวณโคไซน์.....	17
2.5 การคำนวณไซน์.....	18
2.6 กราฟไซน์และโคไซน์.....	18
2.7 ตัวอย่างการคำนวณระยะทางระหว่างจุดละติจูดและลองจิจูด.....	19
3.1 แนวคิดริเริ่มของแบบจำลอง.....	26
3.2 แนวคิดแบบจำลองในการวิจัย.....	26
3.3 การพยากรณ์ส่วนแบบจำลอง ARIMA-LSTM.....	27
3.4 การพยากรณ์ส่วนแบบจำลอง LSTM-ARIMA.....	28
3.5 ลอจิกสำหรับแบบจำลอง LSTM.....	28
3.6 ลอจิกสำหรับแบบจำลอง ARIMA.....	29
3.7 การสำรวจข้อมูลโดยใช้ฮิสโตแกรม.....	31
3.8 การตรวจสอบเพื่อทำการแก้ไขข้อมูลที่ผิดพลาด.....	31
3.9 เครื่องมือที่ใช้ในการวิจัย.....	34
4.1 กราฟการพยากรณ์ฝุ่น PM2.5 สถานี 02T.....	40
4.2 ผลลัพธ์การทำนายคุณภาพอากาศของสถานี 02T.....	41
4.3 กราฟการพยากรณ์ฝุ่น PM2.5 สถานี 11T.....	49
4.4 ผลลัพธ์การทำนายคุณภาพอากาศของสถานี 11T.....	50
4.5 กราฟการพยากรณ์ฝุ่น PM2.5 สถานี 54T.....	57
4.6 ผลลัพธ์การทำนายคุณภาพอากาศของสถานี 54T.....	58
4.7 กราฟการพยากรณ์ฝุ่น PM2.5 สถานี 59T.....	65
4.8 ผลลัพธ์การทำนายคุณภาพอากาศของสถานี 59T.....	66
4.9 ค่าสหสัมพันธ์ระหว่างสถานี.....	67

## บทที่ 1

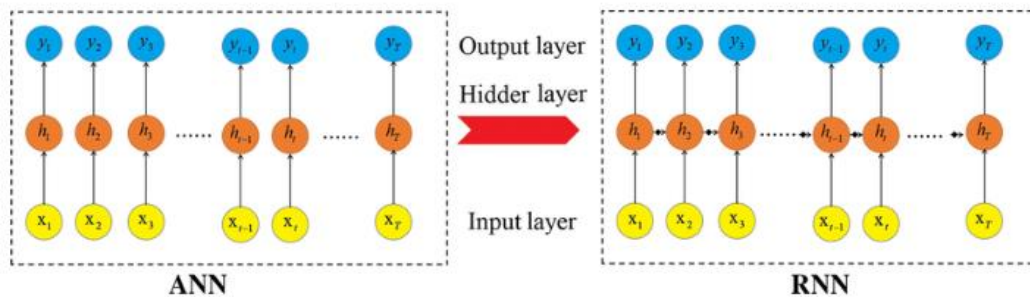
### บทนำ

#### 1.1 ที่มาและความสำคัญของปัญหา

คนไทยเริ่มตื่นตัวเกี่ยวเรื่องฝุ่น PM2.5 ในปี พ.ศ.2562 จากการนำเสนอของสื่ออันเนื่องมาจากปรากฏการณ์ฝุ่นปกคลุมอย่างหนาแน่นในประเทศไทย [1] ซึ่งฝุ่น PM2.5 หรือมีชื่อทางการที่เรียกว่า “Particulate matter with diameter of less than 2.5 micron” เป็นฝุ่นละอองที่มีขนาดไม่เกิน 2.5 ไมครอนและเป็นหนึ่งในตัววัดมาตรฐานคุณภาพอากาศในบรรยากาศ โดยฝุ่น PM2.5 ทำหน้าที่เป็นตัวกลางพาสารอันตรายต่าง ๆ เข้าสู่ปอด เช่น สารก่อมะเร็ง สารโลหะหนัก เป็นต้น ดังนั้นฝุ่น PM2.5 จึงเป็นปัญหาในระดับสากล อย่างเช่น คณะกรรมการเกี่ยวกับผลกระทบทางการแพทย์ของมลพิษทางอากาศ [2] เป็นกลุ่มผู้เชี่ยวชาญทางวิทยาศาสตร์ที่ให้คำแนะนำที่เป็นอิสระและเป็นผู้มีอำนาจแก่รัฐบาลสหราชอาณาจักรเกี่ยวกับผลกระทบต่อสุขภาพของมลพิษทางอากาศ ได้กล่าวถึงมาตรฐานที่องค์การอนามัยโลก ได้ออกมากำหนดมาตรฐานในเดือนกันยายน พ.ศ. 2564 และเผยแพร่แนวทางคุณภาพอากาศฉบับปรับปรุงสำหรับความเข้มข้นของ PM2.5 เฉลี่ยต่อปี คือ  $5 \mu\text{g}/\text{m}^3$  สำหรับในประเทศไทยหน่วยของรัฐทำหน้าที่เก็บรวบรวมข้อมูลและทำการเผยแพร่ข้อมูลฝุ่น PM2.5 คือ กองการจัดการคุณภาพอากาศและเสียง กรมควบคุมมลพิษ รับผิดชอบนำเสนอข้อมูลฝุ่น PM2.5 ของสถานีต่าง ๆ ทั่วประเทศไทย ซึ่งสามารถตรวจสอบข้อมูลผ่านเว็บไซต์ [air4thai.com](http://air4thai.com) และเมื่อพิจารณาเฉพาะสถานีในเมืองหลวงของประเทศไทย พบว่ามี 12 สถานี กระจายอยู่ทั่วกรุงเทพ

การเกิดฝุ่น PM2.5 นั้นมีลักษณะเป็นอนุกรมเวลา (Time series) จากการที่มีลำดับการเกิดขึ้นตามช่วงเวลา แนวคิดการพยากรณ์อนุกรมเวลาที่เรียกว่าด้วยแบบจำลองไฮบริด (Hybrid model) ถูกริเริ่มขึ้นในปี ค.ศ. 2003 โดย Peter Zhang [3] ซึ่งถือเป็นแนวคิดใหม่ในช่วงเวลานั้นสำหรับการสอนให้ระบบคอมพิวเตอร์ทำการเรียนรู้ได้ด้วยตนเองด้วยการใช้ข้อมูลในอดีต (Machine Learning) โดยเป็นการนำข้อดีของแบบจำลองอาร์มา (ARIMA model) มาใช้สำหรับการพยากรณ์ข้อมูลที่มีโครงสร้างเชิงเส้น (Linear Structure Data) หลังจากนั้นนำส่วนที่เหลือ (Residual) ไปพยากรณ์ร่วมกับโครงข่ายประสาทเทียม (ANN: Artificial Neural Networks) สำหรับการพยากรณ์ข้อมูลที่มีโครงสร้างที่ไม่เป็นเชิงเส้น (Non Linear Structure Data) ซึ่งการพยากรณ์อนุกรมเวลาด้วยแบบจำลองไฮบริดแบบใหม่ได้ถูกนำมาใช้ในปี ค.ศ. 2019 โดย Oussama FATHI [4] นับเป็นงานวิจัยที่มีการเปลี่ยนจากโครงข่ายประสาทเทียมจาก ANN เป็น LSTM หรือชื่ออย่างเป็นทางการคือ Long Short-Term Memory ซึ่งเป็นโครงข่ายประสาทเทียมแบบวนกลับ (RNN: Recurrent Neural Networks) เนื่องจาก RNN มีความสามารถในการนำข้อมูลในอดีต มาใช้ในการทำนายสิ่งที่อาจจะเกิดขึ้นในอนาคตได้ ดังภาพที่ 1.1 โดย LSTM ถูกพัฒนาขึ้นเพื่อแก้ปัญหา RNN ที่สามารถดูข้อมูลย้อนหลังได้แค่เพียงระยะสั้น และอาจก่อให้เกิดปัญหาในการทำ Backpropagation ที่จะเป็นสาเหตุให้

เกิดปัญหา Vanishing Gradient Problem ได้ โดยโครงสร้างเครือข่าย LSTM นี้ถูกเสนอโดย Sepp Hochreiter และ Jrgen Schmidhuber [5] ตั้งแต่ปี ค.ศ. 1997



ภาพที่ 1.1 เปรียบเทียบสถาปัตยกรรมโครงข่ายประสาทเทียม

ในปี ค.ศ. 2022 Junwei Xiao และ Qingfang Wang [6] ได้ทำงานวิจัยอนุกรมเวลาเกี่ยวกับฝุ่น PM2.5 เรื่อง Multi-feature PM2.5 Prediction with ARIMA-LSTM เพื่อเปรียบเทียบการพยากรณ์ค่าฝุ่น PM2.5 ระหว่างแบบจำลอง LSTM กับแบบจำลอง ARIMA-LSTM ซึ่งงานวิจัยสรุปว่าการพยากรณ์ด้วยแบบจำลอง ARIMA-LSTM ให้ความแม่นยำกว่า และเมื่อพิจารณาแบบจำลองไฮบริดจะพบว่าเกิดจาก 2 แบบจำลองประกอบไปด้วยแบบจำลอง ARIMA และแบบจำลอง LSTM โดยแบบจำลอง LSTM เป็นแบบจำลองที่เป็นที่ยอมรับและทำได้ดีด้วยการใช้หลายตัวแปรพยากรณ์ค่าเดียว (Global Univariate Models) และการใช้หลายตัวแปรพยากรณ์หลายค่า (Global Multivariate Models) จึงสร้างการพยากรณ์แบบจำลอง ARIMA แบบจำลอง LSTM แบบหลายตัวแปร (Multivariate) แบบจำลอง LSTM แบบตัวแปรเดียว (Univariate) แบบจำลองไฮบริด ARIMA-LSTM ซึ่งแบบจำลอง LSTM เป็นแบบหลายตัวแปร (Multivariate) ตามแนวคิดดั้งเดิมของ Peter Zhang และจากนำแนวคิดแบบจำลองไฮบริดการเรียนรู้เชิงลึก (Hybrid Deep Learning Models) มาประยุกต์ใช้ เกิดแบบจำลองไฮบริดที่มีการพยากรณ์ร่วมด้วยแบบจำลองไฮบริดสองชุดประยุกต์กับการปรับค่าถ่วงน้ำหนัก ได้แบบจำลองไฮบริดประยุกต์ ARIMA-LSTM ที่แบบจำลอง LSTM เป็นแบบหลายตัวแปร (Multivariate) และ LSTM-ARIMA ที่แบบจำลอง LSTM เป็นแบบหลายตัวแปร (Multivariate) กับแบบจำลองไฮบริดประยุกต์ ARIMA-LSTM ที่แบบจำลอง LSTM เป็นแบบหลายตัวแปร (Multivariate) และ LSTM-AIM m ที่แบบจำลอง LSTM เป็นแบบตัวแปรเดียว (Univariate) เพื่อนำมาทำการศึกษาเปรียบเทียบ แล้วทำการวัดผลโดยค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Squared Error: RMSE) และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE)

จากแนวคิดดังกล่าว ผู้วิจัยหวังว่า จากการศึกษาในครั้งนี้ นอกจากจะได้แบบจำลองในการพยากรณ์ค่าฝุ่น PM2.5 แล้ว ยังสามารถนำแบบจำลองไปใช้ในชุดข้อมูลอื่น ๆ ได้ในอนาคต

## 1.2 วัตถุประสงค์ของการศึกษาหรือวิจัย

1.2.1 เพื่อศึกษาแบบจำลองไฮบริด (Hybrid Model) ด้วยแบบจำลองโครงข่ายประสาทเทียมแบบลึก (Long Short-Term Memory: LSTM) และแบบจำลองอริมา (Autoregressive Integrated Moving Average Model: Arima) และบูรณาการจากแบบจำลองไฮบริดการเรียนรู้เชิงลึก (Hybrid Deep Learning Model)

1.2.2 เพื่อเปรียบเทียบแนวโน้มของปริมาณฝุ่น PM2.5 ในจากการพยากรณ์ฝุ่น PM2.5 จากแบบจำลองไฮบริดและแบบจำลองใหม่ที่พัฒนา

1.2.3 เพื่อศึกษาความสัมพันธ์ระหว่างสถานีวัดปริมาณฝุ่น PM2.5

## 1.3 ขอบเขตของงานวิจัย

1.3.1 ประชากรที่ใช้ในการวิจัย

เครือข่ายการติดตามและเฝ้าระวังคุณภาพอากาศของประเทศไทยในกรุงเทพมหานคร ประกอบด้วยกรมควบคุมมลพิษ จำนวน 12 สถานี

1.3.2 ตัวแปรที่ศึกษา

1. ตัวแปรอิสระ

1.1 Date (วันเดือนปี)

1.2 Hour (ชั่วโมง)

1.3 Wind speed (ความเร็วลม)

1.4 Wind direction (ทิศทางลม)

1.5 Temperature (อุณหภูมิ)

1.6 Relative Humidity (ความชื้นสัมพัทธ์)

1.7 Barometric Pressure (ความดันบรรยากาศ)

2. ตัวแปรตาม ได้แก่ ฝุ่น PM2.5

## 1.4 ประโยชน์ที่คาดว่าจะได้รับ

1. พัฒนาแบบจำลองใหม่ทางเพื่อการพยากรณ์ฝุ่น PM2.5 รวมถึงนำไปประยุกต์ใช้ในข้อมูลโดเมนอื่นได้

2. ทราบถึงความสัมพันธ์ของปริมาณฝุ่น PM2.5 ที่ส่งผลต่อสถานีอื่นๆ

3. เกิดแนวคิดในการต่อยอดแบบจำลองสำหรับนักวิจัยรุ่นใหม่

## บทที่ 2

### แนวคิด ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในการวิจัยครั้งนี้ได้ทำการศึกษาค้นคว้าข้อมูลและงานวิจัยที่เกี่ยวข้องกับการพยากรณ์ฝุ่น PM2.5 ดังต่อไปนี้

- 2.1 ความรู้เบื้องต้นเกี่ยวกับฝุ่น PM2.5
- 2.2 การพยากรณ์ด้วยอนุกรมเวลา
- 2.3 แบบจำลอง ARIMA
- 2.4 แบบจำลอง LSTM
- 2.5 แบบจำลอง Hybrid
- 2.6 การเลือกแบบจำลองการเรียนรู้ของเครื่องที่ดีที่สุดสำหรับข้อมูลอนุกรม
- 2.7 การเข้ารหัสคุณลักษณะตามวัฏจักร
- 2.8 การคำนวณระยะทางระหว่างจุดละติจูดและลองจิจูด
- 2.9 การวัดแนวโน้มเข้าสู่ส่วนกลาง
- 2.10 งานวิจัยที่เกี่ยวข้อง

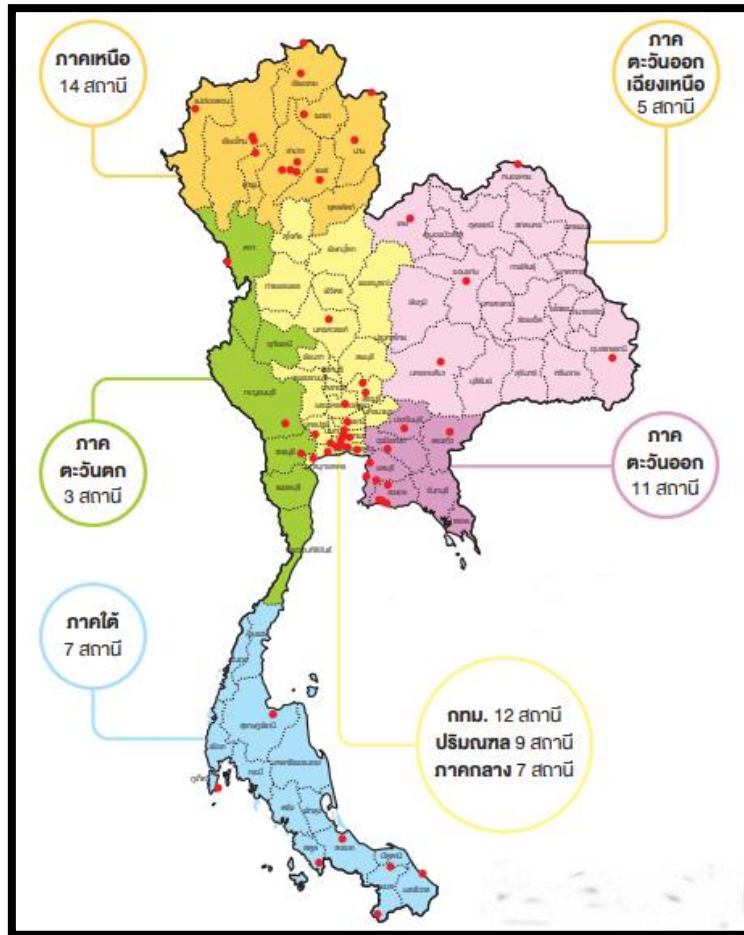
#### 2.1 ความรู้เบื้องต้นเกี่ยวกับฝุ่น PM2.5

ฝุ่นละออง หรือ Particulate Matter (PM) หมายถึง อนุภาคแขวนลอย ในบรรยากาศ การตรวจวัดฝุ่นละอองกระทำ โดยการแยกคัดขนาด เพื่อเป็นตัวแทนกลุ่มสารมลพิษ เช่น PM2.5 หมายถึง ฝุ่นละอองขนาดเล็กกว่า 2.5 ไมครอน และ PM10 หมายถึง ฝุ่นละอองขนาดเล็กกว่า 10 ไมครอน เป็นต้น [7]

ฝุ่นละออง PM2.5 หรือฝุ่นละอองขนาดเล็ก ประกอบด้วยอนุภาคสารเคมี ที่เกิดจากกระบวนการเผาไหม้ของเชื้อเพลิง มีแหล่งกำเนิดจาก ไฟป่า การเผา เศษวัสดุการเกษตรประเภทข้าว ข้าวโพด และอ้อย การคมนาคมขนส่ง โรงงานอุตสาหกรรม การใช้เชื้อเพลิงในครัวเรือน เป็นต้น และบางจังหวัดที่มี ขอบเขตติดประเทศเพื่อนบ้านอาจได้รับผลกระทบจากหมอกควันข้ามแดน

จากปัญหาหมอกควัน ไฟป่า และฝุ่นละออง PM2.5 ซึ่งส่งผลกระทบต่อ สุขภาพอนามัยของประชาชน สิ่งแวดล้อม และกระทบต่อเศรษฐกิจของชาติ ในวันที่ 12 กุมภาพันธ์ พ.ศ. 2562 รัฐบาลจึงกำหนดให้มีการแก้ไขปัญหาหมอกควันด้านฝุ่นละออง ส่งผลให้ในปี พ.ศ. 2563 เกิดการจัดตั้งเครือข่ายการติดตามและเฝ้าระวังคุณภาพอากาศของประเทศไทย โดย กรมควบคุมมลพิษ กระทรวงทรัพยากรธรรมชาติและสิ่งแวดล้อม เพื่อตรวจวัดด้วยเครื่องตรวจวัดฝุ่นละออง PM2.5 โดยสามารถทำ การตรวจวัด แบบอัตโนมัติอย่างต่อเนื่องตลอด 24 ชั่วโมง ตลอดทั้งปี มีระบบการ รับ-ส่งข้อมูลผ่านระบบอินเทอร์เน็ต และเผยแพร่ข้อมูลทางเว็บไซต์และ แอปพลิเคชัน Air4Thai เพื่อให้ทุกหน่วยงานที่เกี่ยวข้องและประชาชนได้รับข้อมูล อย่าง

รวดเร็ว ทันต่อเหตุการณ์ โดยกรมควบคุมมลพิษมีจำนวนเครื่องมือตรวจวัดคุณภาพอากาศ ภาคเหนือ 14 สถานี ภาคตะวันออกเฉียงเหนือ 5 สถานี ภาคตะวันตก 3 สถานี ภาคตะวันออก 11 สถานี ภาคใต้ 7 สถานี ภาคกลาง 7 สถานี ปริมณฑล 9 สถานี และกรุงเทพมหานคร 12 สถานี ครอบคลุม 37 จังหวัดทั่วประเทศ



ภาพที่ 2.1 เครือข่ายการติดตามและเฝ้าระวังคุณภาพอากาศ โดย กรมควบคุมมลพิษ

สำหรับเครือข่ายการติดตามและเฝ้าระวังคุณภาพอากาศของประเทศ เฉพาะในกรุงเทพ โดย กรมควบคุมมลพิษ มีดังต่อไปนี้

**ตารางที่ 2.1** เครือข่ายการติดตามและเฝ้าระวังคุณภาพอากาศในกรุงเทพฯ โดยกรมควบคุมมลพิษ

รหัสสถานี	ที่อยู่	ละติจูด	ลองจิจูด
02T	มหาวิทยาลัยราชภัฏบ้านสมเด็จเจ้าพระยา แขวงทรีรัฐจี้ เขตธนบุรี, กรุงเทพฯ	13.7328 N	100.4877 E
03T	ริมถนนทางหลวงหมายเลข 3902 ริมถนนกาญจนาภิเษก เขตบางขุนเทียน, กรุงเทพฯ	13.6365 N	100.4143 E
05T	กรมอุตุนิยมวิทยาบางนา แขวงบางนา เขต บางนา, กรุงเทพฯ	13.6662 N	100.6057 E
10T	เคหะชุมชนคลองจั่น แขวงคลองจั่น เขตบางกะปิ, กรุงเทพฯ	13.7799 N	100.6460 E
11T	การเคหะชุมชนห้วยขวาง แขวงดินแดง เขต ดินแดง, กรุงเทพฯ	13.7755 N	100.5692 E
12T	โรงเรียนนนทรีวิทยา แขวงช่องนนทรี เขตยานนาวา, กรุงเทพฯ	13.7081 N	100.5473 E
50T	โรงพยาบาลจุฬาลงกรณ์ ริมถนนพระราม 4 เขตปทุมวัน, กรุงเทพฯ	13.7299 N	100.5365 E
52T	การไฟฟ้าอโยธยาบุรี ริมถนนอินทรพิทักษ์ เขตธนบุรี, กรุงเทพฯ	13.7276 N	100.4866 E
53T	สถานีตำรวจนครบาลโชคชัย ริมถนนลาดพร้าว เขตวังทองหลาง, กรุงเทพฯ	13.7954 N	100.5930 E
54T	การเคหะชุมชนดินแดง ริมถนนดินแดง เขตดินแดง, กรุงเทพฯ	13.7925 N	100.5502 E
59T	กรมประชาสัมพันธ์ แขวงพญาไท เขตพญาไท, กรุงเทพฯ	13.7832 N	100.5405 E
61T	โรงเรียนบดินเดชา (สิงห์ สิงหเสนี) แขวงพลับพลา เขตวังทองหลาง, กรุงเทพฯ	13.7697 N	100.6146 E



จากปัญหาฝุ่นละออง PM2.5 รัฐบาลให้ความสำคัญและได้มีการประกาศปรับใช้มาตรฐานฝุ่น PM2.5 ใหม่ สำหรับค่ามาตรฐาน PM 2.5 ซึ่งมีผลบังคับใช้ตั้งแต่วันที่ 1 มิถุนายน 2566 ในการกำหนดโดยใช้มาตรฐาน WHO Guidline แบ่งระดับการแจ้งเตือนตามดัชนีคุณภาพอากาศ (Air Quality Index : AQI) ดังนี้ [8]

ตารางที่ 2.2 เปรียบเทียบดัชนีคุณภาพอากาศและมาตรฐาน PM 2.5 ค่าเฉลี่ย 24 ชั่วโมง

AQI	ความหมาย	ความเข้มข้นของ PM2.5 (มคก./ลบ.ม.)	
		ค่าเก่า	ค่าปรับใหม่
0 – 25	คุณภาพอากาศดีมาก	0 - 25	0 - 15
26 – 50	คุณภาพอากาศดี	25 - 37	15 - 25
51 – 100	คุณภาพอากาศปานกลาง	37 - 50	25 – 37.5
101 – 200	คุณภาพอากาศเริ่มมีผลต่อสุขภาพ	50 - 90	37.5 - 75
201 ขึ้นไป	คุณภาพอากาศมีผลต่อสุขภาพ	มากกว่า 90	มากกว่า 75

## 2.2 การพยากรณ์ด้วยอนุกรมเวลา

George E. P. Box, Gwilym M. Jenkins., Gregory C.Reinsel & Greta M., Ljung.[9] ได้นิยามอนุกรมเวลา คือ ลำดับของการสังเกตที่ดำเนินการตามลำดับเวลา ตัวอย่างเช่น ลำดับรายเดือนของปริมาณสินค้าที่จัดส่งจากโรงงาน ชุดข้อมูลรายสัปดาห์ของจำนวนอุบัติเหตุทางถนน ปริมาณน้ำฝนรายวัน เป็นต้น ตัวอย่างของอนุกรมเวลามีอยู่มากมาย เช่น เศรษฐศาสตร์ ธุรกิจ วิศวกรรมศาสตร์ วิทยาศาสตร์ธรรมชาติ (โดยเฉพาะธรณีฟิสิกส์และอุตุนิยมวิทยา) และสังคมศาสตร์

ลักษณะเฉพาะของอนุกรมเวลา คือ ลักษณะของการพึ่งพาระหว่างอนุกรมเวลาที่อยู่ติดกันซึ่งเป็นสิ่งที่น่าสนใจ การวิเคราะห์อนุกรมเวลาจึงเกี่ยวข้องกับเทคนิคสำหรับการวิเคราะห์การพึ่งพาที่ต้องการการพัฒนาแบบจำลองสุ่มและแบบจำลองไดนามิก

การพยากรณ์ด้วยอนุกรมเวลาจึงเป็นการใช้ค่าสังเกต ณ เวลา  $t$  จากอนุกรมเวลาเพื่อคาดการณ์มูลค่าในอนาคต  $t + 1$  ตัวอย่างเช่น ในปัญหาการพยากรณ์ยอดขาย ยอดขาย  $z_t$  ในเดือนปัจจุบัน  $t$  และ ยอดขาย  $z_{t-1}, z_{t-2}, z_{t-3}, \dots$  ในเดือนก่อนหน้า อาจถูกนำมาใช้เพื่อพยากรณ์การขายสำหรับ

$l = 1, 2, 3, \dots, 12$  เดือนข้างหน้า แสดงโดย  $\hat{z}_{t(l)}$  คือ การคาดการณ์ที่จุดเริ่มต้น  $t$  ของ ยอดขาย  $z_{t+1}$  ในอนาคต  $t + 1$  นั่น คือ ณ เวลา  $l$  ฟังก์ชัน  $\hat{z}_{t(l)}$  ซึ่งให้การพยากรณ์ที่จุดเริ่มต้น  $t$  สำหรับเวลานำในอนาคตทั้งหมด โดยอิงตามข้อมูลที่มีอยู่จากค่าปัจจุบันและค่าก่อนหน้า  $z, z_{t-1}, z_{t-2}, z_{t-3}, \dots$  เมื่อเวลาผ่านไป  $t$  จะถูกเรียกว่าฟังก์ชันพยากรณ์ที่จุดกำเนิด  $t$

ในบางสถานการณ์ อาจมีเหตุการณ์พิเศษภายนอกบางอย่าง เช่น การถูกแทรกแซงจาก เหตุการณ์ต่างๆ ที่อาจส่งผลกระทบต่ออนุกรมเวลา  $z_t$  ที่กำลังศึกษาอยู่ ตัวอย่างของเหตุการณ์การแทรกแซงดังกล่าว เช่น การรวมตัวกันของกฎระเบียบด้านสิ่งแวดล้อมใหม่ นโยบายเศรษฐกิจที่มีการเปลี่ยนแปลง การนัดหยุดงาน และแคมเปญส่งเสริมการขายพิเศษ เป็นต้น ความพยายามในการหาคำตอบสำหรับคำถามว่า "ผลของการเปลี่ยนแปลงประเภทนี้เกิดขึ้น" หรือความพยายามตอบคำถามว่า "ทำไมจึงเกิดขึ้น" เหล่านี้จึงเป็นแรงจูงใจในการพยากรณ์ด้วยอนุกรมเวลา

ตัวดำเนินการอย่างง่าย เราใช้ตัวดำเนินการถอยหลัง (Backward Shift) แทนด้วย B

กำหนดให้  $Bz_t = z_{t-1}$

ดังนั้น  $B^m z_t = z_{t-m}$

การดำเนินการผกผันดำเนินการโดยตัวดำเนินการเลื่อนไปข้างหน้า (Forward Shift) แทนด้วย F

$= B^{-1}$

กำหนดให้  $Fz_t = z_{t+1}$

ดังนั้น  $F^m z_t = z_{t+m}$

ตัวดำเนินการสำคัญ คือ ตัวดำเนินการผลต่างย้อนกลับ  $\nabla$

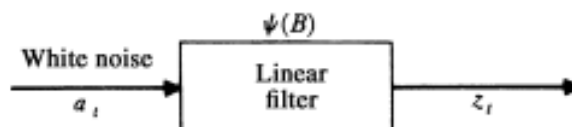
กำหนดให้  $\nabla z_t = z_t - z_{t-1}$

สามารถเขียนในรูปของ B ได้

ดังนั้น  $\nabla z_t = z_t - z_{t-1} = (1 - B)z_t$

โมเดลตัวกรองเชิงเส้น (Linear Filter Model) โมเดลแบบสุ่มที่ใช้กัน มีพื้นฐานมาจากแนวคิดเดิม เนื่องจาก Yule (1927) กล่าวว่าอนุกรมเวลาที่สังเกตได้  $z_t$  ซึ่งค่าที่ต่อเนื่องกันนั้นขึ้นต่อกันสูงมักจะถูกมองว่า สร้างจากชุดของ "แรงกระแทก (shocks)" ที่เป็นอิสระต่อกัน แรงกระแทกเหล่านี้เป็นการวาดแบบสุ่มจากการ แจกแจงแบบคงที่ โดยปกติจะถือว่าปกติและมีค่าเฉลี่ยเป็นศูนย์และความแปรปรวน  $\sigma_a^2$  โดยลำดับของตัวแปรสุ่มอิสระดังกล่าว  $a_t, a_{t-1}, a_{t-2}, \dots$  เรียกว่ากระบวนการไวท์นอยส์ (White Noise)

กระบวนการไวท์นอยส์ (White noise)  $a_t$  ควรจะเปลี่ยนเป็นกระบวนการ  $z_t$  โดยสิ่งที่เรียกว่าตัวกรองเชิงเส้น ดังรูปที่ 2.3 การดำเนินการกรองเชิงเส้นใช้เพียงการถ่วงน้ำหนัก



**ภาพที่ 2.2** การแสดงอนุกรมเวลาเป็นเอาต์พุตจากตัวกรองเชิงเส้น

ผลรวมของการกระแทกแบบสุ่มก่อนหน้านี้  $a_t$   
ดังนั้น

$$\begin{aligned} z_t &= \mu + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \dots \\ &= \mu + \psi(B)a_t \end{aligned} \tag{2.1}$$

โดยทั่วไป  $\mu$  เป็นพารามิเตอร์ที่กำหนด "ระดับ" ของกระบวนการ และ  $\psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \dots$  เป็นตัวดำเนินการเชิงเส้นที่แปลง  $a_t$  ให้เป็น  $z_t$  และเรียกว่าการถ่ายโอนฟังก์ชันของตัวกรอง แบบจำลอง (2.1) ที่แสดง สามารถให้ช่วงของรูปแบบมีระดับความยืดหยุ่นที่เป็นอิสระผ่านค่าของกระบวนการ  $\{z_t\}$  ที่แสดงในรูปแบบที่เป็นอิสระ (ไม่สามารถสังเกตได้) สุ่มแรงกระแทก  $a_t$

ลำดับของ  $\psi_1, \psi_2, \dots$  ที่เกิดขึ้นจากน้ำหนักนั้น ในทางทฤษฎีแล้ว อาจมีขอบเขตจำกัดหรือไม่จำกัด ถ้าลำดับนี้มีขอบเขตจำกัดหรือไม่จำกัดและสรุปได้ว่า  $\sum_{j=0}^{\infty} |\psi_j| < \infty$  ตัวกรองมีความเสถียรและกระบวนการ  $z_t$  หยุตนั้น พารามิเตอร์  $\mu$  จึงเป็นค่าเฉลี่ยที่กระบวนการแตกต่างกันไป มิฉะนั้น  $z_t$  จะไม่อยู่นิ่ง และ  $\mu$  เป็นเพียงจุดอ้างอิงสำหรับระดับของกระบวนการ

**2.3 แบบจำลอง ARIMA**

การวิเคราะห์อนุกรมเวลาที่เป็นการใช้ข้อมูลในอดีตเพื่อใช้เป็นแนวทางในการพยากรณ์อนาคตด้วยอนุกรมเวลาแบบ Box-Jenkins (1976) ที่เรียกว่า Autoregressive Integrated Moving Average Model หรือเป็นที่รู้จักว่า ARIMA ซึ่งเป็นแบบจำลองถดถอย ประกอบด้วยองค์ประกอบของ autoregression, integrated และ moving average เป็นแบบจำลองถดถอยใช้ในการพยากรณ์ข้อมูลอนุกรมเวลาที่อ้างอิง  $p, d, q$  [10]

Auto Regression, AR(p) เป็นส่วนหนึ่งของแบบจำลอง ARIMA ตามแนวคิดที่ว่าจะใช้ความล่าช้าของตัวเอง (ค่าในอดีต) เป็นตัวทำนาย โดยที่พารามิเตอร์  $p$  คือ ขอบเขตของจำนวนความล่าช้าที่เกิดขึ้น

$$y_t = \alpha + \sum_{i=0}^p \beta_i Y_{t-i} + \epsilon_t \tag{2.2}$$

เมื่อ  $\alpha$  คือ ค่าจุดตัดแกนตั้ง (intercept) ในสมการถดถอย

$\epsilon_t$  คือ ค่าไวท์นอยส์ (White noise)

$\sum_{i=0}^p \beta_i$  คือ ค่าสัมประสิทธิ์ (Coefficients) ของค่าที่ผ่านมา ที่กำหนดโดย  $\sum_{i=0}^p Y_{t-i}$

Moving Average, MA(q) ใช้ข้อผิดพลาดที่เหลือ (residual) ของเวลาที่ผ่านมาชี้ให้เห็นการคาดการณ์ในปัจจุบันและอนาคต โดยที่พารามิเตอร์ q คือ จำนวนของข้อผิดพลาดการคาดการณ์ความล่าช้าที่ใช้ในการคำนวณค่าปัจจุบัน

$$y_t = \alpha + \sum_{i=0}^q \phi_i \varepsilon_{t-i} + \varepsilon_t \quad (2.3)$$

เมื่อ  $\sum_{i=0}^q \varepsilon_{t-i}$  คือ เงื่อนไขข้อผิดพลาดของการล่าช้าที่เกี่ยวข้อง (error terms of the respective lags)

Integrated I(d) ช่วยทำให้ข้อมูลอนุกรมเวลายุ่งยากขึ้นเพื่อลดการขึ้นต่อกันของเวลาและแนวโน้ม โดยที่พารามิเตอร์ d คือ ระดับความแตกต่างซึ่งหมายถึงจำนวนครั้งที่ข้อมูลมีความแตกต่างกัน ถ้าอนุกรมเวลาหยุดนิ่ง (stationary) ระดับความแตกต่างของมันคือศูนย์

แบบจำลอง ARIMA(p,d,q) จะสามารถเขียนได้ดังนี้:

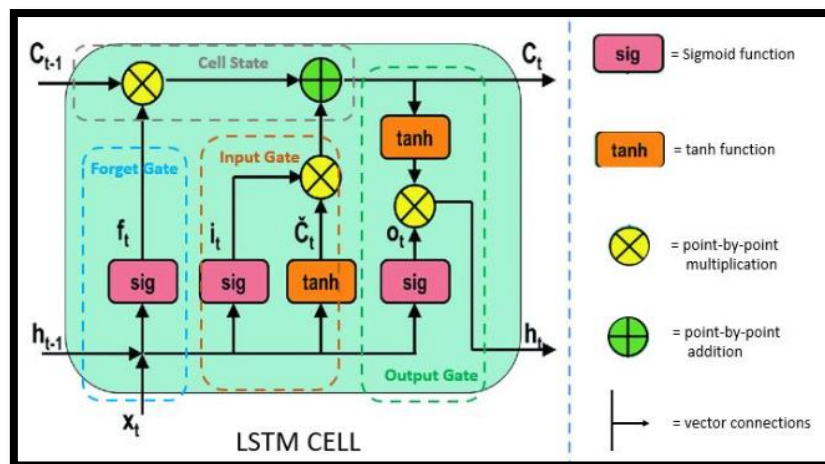
$$y_t = \alpha + \sum_{i=0}^p \beta_i Y_{t-i} + \sum_{i=0}^q \phi_i \varepsilon_{t-i} + \varepsilon_t \quad (2.4)$$

แบบจำลอง ARIMA มาตรฐาน สามารถรับค่าอินพุตด้วยตนเอง (p,d,q) โดยใช้ความสัมพันธ์อัตโนมัติและการทดสอบทางสถิติอื่นๆ โดยใช้ Auto Arima จากแพ็คเกจ pmdarima ซึ่งจะประเมินค่า p, d, q ด้วยตัวเองโดยอัตโนมัติ

## 2.4 แบบจำลอง LSTM

Sepp Hochreiter และ Jürgen Schmidhuber [5] นำเสนอ Long Short-Term Memory (LSTM) เป็นครั้งแรก ตั้งแต่ปี ค.ศ. 1997 ซึ่งได้รับการออกแบบอย่างระมัดระวังและเพิ่มหน่วยความจำ และหลีกเลี่ยงปัญหาที่เกิดขึ้นมานานตามลักษณะ Recurrent Neural Network (RNN) โดย LSTM สามารถรักษาหน่วยความจำระยะยาวของโครงข่ายประสาทเทียม อันเนื่องมาจาก RNN เครือข่ายเลเยอร์จะอัปเดตข้อมูลโดยไม่มีข้อจำกัด ข้อมูลจึงมีความยุ่งเหยิงและหายไปได้ง่าย ซึ่งนำไปสู่ปัญหาในระหว่างการเทรนเกรเดียนต์ (Gradient) มีขนาดเล็กลงเรื่อย ๆ จนเท่ากับ 0 ทำให้ Weight ไม่ถูกอัปเดตอีกต่อไป ทำให้โมเดลเทรนต่อไม่ได้ (Vanishing Gradient Problem)

อย่างไรก็ตาม เครือข่าย LSTM จะเพิ่มหน่วยการลืมและหน่วยหน่วยความจำให้กับเลเยอร์ที่ซ่อนอยู่ โดยจะละทิ้งข้อมูลเล็กน้อยเมื่อมีการป้อนข้อมูลใหม่ สำรองข้อมูลสำคัญไว้ในหน่วยความจำระยะยาว หน่วยเหล่านี้เรียกว่าเกต (Gate) ใน LSTM และแกนหลักของกลไกการเกตประกอบด้วยหนึ่งเซลล์และสามเกต เช่น เกตอินพุต (Input Gate) เกตเอาต์พุต (Output Gate) และเกตลืม (Forget Gate)



ภาพที่ 2.3 แผนภูมิองค์ประกอบ LSTM

เครือข่ายประสาท LSTM ประกอบด้วยเกตดังต่อไปนี้:

- (1) อินพุตเกต (Input Gate:  $i$ ) กำหนดจำนวนอินพุต ณ ปัจจุบัน บันทึกขึ้นตอนเวลาในสถานะเซลล์ บางครั้งเรียกอินพุตเกตว่าเกตปรับปรุง
- (2) เกตลืม (Forget Gate:  $f$ ) กำหนดจำนวนสถานะของเซลล์ที่ขึ้นตอนของเวลาก่อนหน้าจะยังคงอยู่ในขั้นตอนของเวลาปัจจุบัน และอาจไม่เก็บไว้เลยก็ได้
- (3) เกตเอาต์พุต (Output Gate:  $o$ ) กำหนดจำนวนสถานะของเซลล์ที่ส่งออก เป็นค่าเอาต์พุตปัจจุบันของสถานะที่ซ่อนอยู่

จากภาพที่ 2.4 เส้นแต่ละเส้นจะส่งเวกเตอร์ทั้งหมด วงกลมแสดงถึงการดำเนินการตามจุด และกล่องสี่เหลี่ยม คือ เลเยอร์เครือข่าย เส้นที่รวมกันแสดงถึงการเชื่อมต่อของเวกเตอร์และเส้นประคั่นระบุว่า เนื้อหาถูกคัดลอกแล้วกระจายไปตามสถานที่ต่างๆ ตัวห้อย  $t$  ดัชนีเวลา เมื่อเปรียบเทียบกับ RNN ที่มีสถานะการส่งสัญญาณเพียงสถานะเดียว LSTM มีสองสถานะการส่งข้อมูล นั่นคือสถานะเซลล์ (Cell State:  $c_t$ ) และสถานะที่ซ่อนอยู่ (Hidden State:  $h_t$ ) กุญแจสำคัญของ LSTM คือ สถานะเซลล์ ( $c_t$ ) ซึ่งคล้ายกับสายพานลำเลียงที่วิ่งโดยตรงบองค์ประกอบของโซ่ทั้งหมดโดยมีปฏิสัมพันธ์เชิงเส้นเพียงเล็กน้อย ทำให้ข้อมูลของแต่ละเซลล์หมุนเวียนบนเครือข่ายได้ง่ายและยังคงเหมือนเดิม นอกจากนี้ เพื่อแก้ปัญหาการหายไปของเกร

เดียนต์ LSTM ไม่ได้ใช้การแปลงค่าสัมพัทธ์แบบไม่เชิงเส้นแบบเชิงองค์ประกอบสำหรับการวนซ้ำภายใน แต่ตั้งค่ากลไกการเกทเพื่อควบคุมการไหลของข้อมูลใน LSTM โดยทั่วไป Gates มีการคูณแบบ pointwise สำหรับการเพิ่มและเลเยอร์โครงข่ายประสาทเทียมที่มีฟังก์ชัน sigmoid  $\sigma(x) = 1 / (1 + e^{-x})$  ไฮเปอร์โบลิกแทนเจนต์ฟังก์ชัน (hyperbolic tangent function)  $\tanh(x) = (e^x - e^{-x}) / (e^x + e^{-x})$  ทำการกำหนดจำนวนข้อมูลที่สามารถผ่านได้

ในสูตรการคำนวณโดยละเอียดของ LSTM จากสูตร (2.5) ถึง (2.10) โดยตัวแปรตัวพิมพ์เล็กแทนเวกเตอร์ในขณะที่ตัวพิมพ์ใหญ่ตัวแปรเป็นตัวแทนของเมทริกซ์ กระบวนการอัปเดตแสดงในภาพที่ 2.3:

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \quad (2.5)$$

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \quad (2.6)$$

$$\tilde{c}_t = \tanh(W_c x_t + U_c h_{t-1} + b_c) \quad (2.7)$$

$$c_t = f_t \circ c_{t-1} + i_t \circ \tilde{c}_t \quad (2.8)$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \quad (2.9)$$

$$h_t = o_t + \tanh(c_t) \quad (2.10)$$

ตัวแปร คือ  $x_t \in \mathbb{R}^m$  เป็นเวกเตอร์อินพุตของเซลล์

$h_t \in (-1, 1)^n$  เวกเตอร์สถานะที่ซ่อนอยู่ของเซลล์

$f_t \in (0, 1)^n$  เกทลืมเวกเตอร์เปิดใช้การทำงานของเกท

$i_t \in (0, 1)^n$  เวกเตอร์การเปิดใช้งานของเกทอินพุต

$\tilde{c}_t \in (-1, 1)^n$  เวกเตอร์การเปิดใช้งานอินพุตเซลล์

$c_t \in \mathbb{R}^n$  เป็นเวกเตอร์สถานะของเซลล์

$o_t \in (0, 1)^n$  เวกเตอร์การปิดใช้งานของเกทเอาต์พุต

ตัวยกกำลัง  $m$  และ  $n$  หมายถึง จำนวนของตัวแปรที่ถูกป้อนข้อมูลและจำนวนของตัวแปรของเซลล์ที่ซ่อนอยู่ตามลำดับ

◦ ย่อมาจากผลิตภัณฑ์ Hadamard สำหรับ  $W_t \in \mathbb{R}^{n \times m}$ ,  $U_k \in \mathbb{R}^{n \times n}$ ,  $b_k \in \mathbb{R}^n$

$k = f, i, c, o$  เป็นตัวแทนของเมทริกซ์น้ำหนักสำหรับเวกเตอร์อินพุต  $X_t$  เมทริกซ์

น้ำหนักสำหรับเวกเตอร์สถานะที่ซ่อนอยู่  $h_t$  และพารามิเตอร์เวกเตอร์อคติสำหรับเกทหรือสถานะเซลล์ต่างๆ

นอกจากนี้ยังใช้เวกเตอร์พิเศษเพื่อให้ดีขึ้น ตัวอย่างเช่น เวกเตอร์สถานะของเซลล์  $c_t \in \mathbb{R}^n$  ไม่เพียงมีข้อมูลของเซลล์หนึ่งในเครือข่ายประสาท LSTM แต่มีข้อมูลจำนวน  $n$  ของเซลล์ LSTM

## 2.5 แบบจำลอง Hybrid

การพยากรณ์อนุกรมเวลาเป็นพื้นที่สำคัญของการพยากรณ์ ซึ่งการสังเกตในอดีตของตัวแปรเดียวกันจะถูกรวบรวมและวิเคราะห์เพื่อพัฒนาแบบจำลองที่อธิบายความสัมพันธ์พื้นฐาน ซึ่งในช่วงในหลายทศวรรษที่ผ่านมาได้ทุ่มเทความพยายามอย่างมากในการพัฒนาและปรับปรุงแบบจำลองการคาดการณ์อนุกรมเวลา โดยโมเดลอนุกรมเวลาที่สำคัญและใช้กันอย่างแพร่หลายรูปแบบหนึ่งคือแบบจำลอง ARIMA เกิดจากคุณสมบัติทางสถิติเช่นเดียวกับวิธีการของ Box-Jenkins ที่รู้จักกันดีในกระบวนการสร้างแบบจำลอง [3]

แม้ว่าแบบจำลอง ARIMA ค่อนข้างจะมีความคล่องแคล่วเนื่องจากสามารถแสดงอนุกรมเวลาได้หลายประเภท เช่น ค่าเฉลี่ยเคลื่อนที่อัตโนมัติ (AR) ค่าเฉลี่ยเคลื่อนที่เฉลี่ย (MA) และอนุกรม AR และ MA (ARMA) ที่รวมกัน ข้อจำกัดที่สำคัญของแบบจำลองคือ รูปแบบเชิงเส้นของแบบจำลอง นั่นคือ โครงสร้างความสัมพันธ์เชิงเส้นจะถือว่าอยู่ในค่าอนุกรมเวลา ดังนั้นโมเดล ARIMA จึงไม่สามารถจับรูปแบบที่ไม่เชิงเส้นได้

ในช่วงปี ค.ศ. 2003 เครือข่ายประสาทเทียม/สังคม (artificial neural networks : ANNs) ได้รับความสนใจอย่างกว้างขวางและใช้ในการพยากรณ์อนุกรมเวลา Zhang และคณะได้ทำการศึกษาเกี่ยวกับเรื่องนี้ในขณะนั้น ซึ่งพบว่าข้อได้เปรียบที่สำคัญของโครงข่ายประสาทเทียมคือความสามารถในการสร้างแบบจำลองแบบไม่เชิงเส้นที่ยืดหยุ่นด้วย ANN ไม่จำเป็นต้องระบุรูปแบบเฉพาะ แต่แบบจำลองนั้นถูกสร้างขึ้นแบบปรับเปลี่ยนตามคุณสมบัติที่นำเสนอจากข้อมูล ทั้งโมเดล ARIMA และ ANN ต่างประสบความสำเร็จในโดเมนเชิงเส้นหรือไม่เชิงเส้นของตนเอง อย่างไรก็ตาม ไม่มีแบบจำลองที่เหมาะสมกับทุกสถานการณ์ การประมาณแบบจำลอง ARIMA กับปัญหาไม่เชิงเส้นที่ซับซ้อนอาจไม่เพียงพอ ในทางกลับกัน การใช้ ANN เพื่อสร้างแบบจำลองปัญหาเชิงเส้นให้ผลลัพธ์ที่หลากหลาย จึงเกิดแนวคิดวิธีการแบบผสมผสานที่มีความสามารถในการสร้างแบบจำลองทั้งเชิงเส้นและไม่เชิงเส้นอาจที่เป็นกลยุทธ์ที่ดีสำหรับการใช้งานจริง โดยการรวมโมเดลที่แตกต่างกัน

ดังนั้น จึงมีเหตุผลที่จะพิจารณาอนุกรมเวลาที่ประกอบด้วยโครงสร้างความสัมพันธ์เชิงเส้นอัตโนมัติและองค์ประกอบที่ไม่เชิงเส้น นั่นคือ,

$$y_t = L_t + N_t \quad (2.11)$$

โดยที่  $L_t$  หมายถึงส่วนประกอบเชิงเส้น  
 $N_t$  หมายถึงส่วนประกอบที่ไม่เชิงเส้น

กระบวนการต้องมีการประมาณสององค์ประกอบจากข้อมูล อันดับแรก ให้ ARIMA เป็นแบบจำลองส่วนประกอบเชิงเส้น จากนั้นส่วนที่เหลือจากแบบจำลองเชิงเส้นจะมีเฉพาะความสัมพันธ์แบบไม่เชิงเส้น ให้  $e_t$  แทนค่าที่เหลือ (residuals) ณ เวลา  $t$  จากโมเดลเชิงเส้นแล้ว

$$e_t = y_t - L_t \quad (2.12)$$

โดยที่

$L_t$  หมายถึงค่าพยากรณ์สำหรับเวลา  $t$  จากความสัมพันธ์ที่ประมาณไว้ ค่าที่เหลือมีความสำคัญในการวินิจฉัยความสมบูรณ์ของโมเดลเชิงเส้น โมเดลเชิงเส้นไม่ใช่เพียงพอหากยังมีโครงสร้างความสัมพันธ์เชิงเส้นเหลืออยู่ในส่วนที่เหลือ อย่างไรก็ตาม การวิเคราะห์สามารถตรวจจบบรูปแบบที่ไม่เชิงเส้นในข้อมูลได้ ซึ่งส่วนที่เหลือจะระบุข้อจำกัดของ ARIMA ดังนั้นการสร้างแบบจำลองส่วนที่เหลือโดยใช้ ANN จะสามารถค้นพบความสัมพันธ์แบบไม่เชิงเส้นได้ ด้วย  $n$  โหนดอินพุต รูปแบบ ANN สำหรับส่วนที่เหลือจะเป็น

$$e_t = f(e_{t-1}; e_{t-2}; \dots; e_{t-n}) + \epsilon_t \quad (2.13)$$

โดยที่

$f$  หมายถึง ฟังก์ชันไม่เชิงเส้นที่กำหนดโดยโครงข่ายประสาทเทียม

$\epsilon_t$  หมายถึง ข้อผิดพลาดของการสุ่ม

ดังนั้นการ model ที่ถูกต้องจึงมีความสำคัญ

จากสมการ (2.13)  $\hat{N}_t$  การคาดการณ์แบบรวมจะเป็น

$$\hat{y}_t = \hat{L}_t + \hat{N}_t \quad (2.14)$$

โดยสรุป วิธีการที่นำเสนอของระบบไฮบริดประกอบด้วยสองขั้นตอน

ขั้นแรก จะใช้แบบจำลอง ARIMA เพื่อวิเคราะห์ส่วนเชิงเส้นของปัญหา

ขั้นตอนที่สอง มีการพัฒนาแบบจำลองโครงข่ายประสาทเทียมเพื่อจำลองสิ่งตกค้างจากโมเดล

ARIMA เนื่องจากโมเดล ARIMA ไม่สามารถจับโครงสร้างแบบไม่เชิงเส้นของ

ข้อมูลส่วนที่เหลือของโมเดลเชิงเส้นจะมีข้อมูลเกี่ยวกับความไม่เชิงเส้น ผลลัพธ์จากโครงข่ายประสาทเทียมสามารถใช้เป็นการคาดคะเนเงื่อนไขข้อผิดพลาดได้สำหรับแบบจำลอง ARIMA แบบจำลองไฮบริดใช้ประโยชน์จากคุณลักษณะเฉพาะและความแข็งแกร่งของโมเดล ARIMA และโมเดล ANN ในการกำหนดรูปแบบที่แตกต่างกัน ดังนั้นจึงสามารถเป็นประโยชน์ในการสร้างแบบจำลองเชิงเส้นและไม่เชิงเส้นแยกจากกันโดยใช้ความแตกต่างแบบจำลองแล้วรวมการคาดการณ์เพื่อปรับปรุงการสร้างแบบจำลองและการพยากรณ์โดยรวม

## 2.6 การเลือกแบบจำลองการเรียนรู้ของเครื่องที่ดีที่สุดสำหรับข้อมูลอนุกรม

Jeremy DiBattista [11] ได้ให้ข้อสังเกตการตัดสินใจเลือกแบบจำลองในการพยากรณ์ข้อมูลอนุกรม โดยการเลือกแบบจำลองโลคอล (Local Model) หรือแบบจำลองโกลบอล (Global Model) เพื่อทำการพยากรณ์ตัวแปรเดียวหรือหลายตัวแปร



แบบจำลองโลคอล (Local Model) หรือบางครั้งเรียกว่าแบบจำลองแบบวนซ้ำ (Iterative Model) หรือแบบจำลองแบบดั้งเดิม (Traditional Model) เป็นการพยากรณ์โดยใช้เฉพาะค่าก่อนหน้าของข้อมูลคอลัมน์เดียวในการทำนายค่าในอนาคต เนื่องจากแบบจำลองเฉพาะที่อาศัยเพียงคอลัมน์ข้อมูลเดียวพวกมันจึงต้องเป็นตัวแปรเดี่ยวด้วย

ในทางตรงกันข้าม แบบจำลองโกลบอล (Global Model) ใช้คอลัมน์ข้อมูลหลายตัวแปรเพื่อพยากรณ์มูลค่าในอนาคต โดยทั่วไปจะเป็นตัวแปรที่ไม่ขึ้นกับเวลา แบบจำลองส่วนกลางมีอยู่สองประเภท ได้แก่ แบบจำลองหลายตัวแปรเพื่อทำการพยากรณ์ค่าเดียว (Global Univariate Model) และแบบจำลองหลายตัวแปรเพื่อทำการพยากรณ์หลายค่า (Global Multivariate Model) โดยการเลือกแบบจำลองการเรียนรู้ของเครื่องที่ดีที่สุดสำหรับข้อมูลอนุกรม มีรายละเอียดดังนี้

### 1. การใช้แบบจำลองโลคอลเพื่อพยากรณ์ตัวแปรเดี่ยว (Local Univariate Models)

#### 1) ข้อมูลมีจำนวนไม่มากหรือมีความสำคัญไม่มาก

เหมาะกับลักษณะของข้อมูลมีความเรียบง่าย ไม่แปรผัน และคาดเดาได้ง่าย สามารถฝึกฝนได้ทันที ใช้ทรัพยากรในการประมวลผลเพียงเล็กน้อย ซึ่งการใช้แบบจำลองโมเดลที่ซับซ้อนมากขึ้นอาจใช้งานเกินความจำเป็นและเกินพอดีข้อมูลของคุณ

#### 2) ต้องการผลพยากรณ์ที่แม่นยำ แต่มี 1 ตัวแปรเท่านั้น

มีตัวแปรที่ใช้งานได้เพียง 1 ตัว การพยากรณ์ซ้ำ/แบบจำลองโลคอลทั้งชุด ซึ่งการพิสูจน์แสดงผลว่าสามารถทำนายข้อมูลได้แม่นยำกว่า

#### 3) แบบจำลองที่คาดเดาได้ตามฤดูกาล

ข้อมูลของเป็นไปตามรูปแบบตามฤดูกาลที่คาดเดาได้

#### ตัวอย่างการเลือกแบบจำลองโลคอลพยากรณ์ตัวแปรเดี่ยว (Local Univariate Models)

- Moving Average เป็นวิธีที่ง่ายที่สุดและคำนวณได้ด้วยคำสั่ง python ด้วย pandas 1 บรรทัดเท่านั้น

- Exponential Smoothing/Holt-Winters เป็นวิธีคาดการณ์ค่าโดยใช้ค่าเฉลี่ยถ่วงน้ำหนักของค่าก่อนหน้าทั้งหมด โดยที่ค่าล่าสุดจะถ่วงน้ำหนักให้สูงกว่า ใน Holt-Winters ฤดูกาลและแนวโน้มจะถูกนำมาเป็นพารามิเตอร์ในสมการ

- ARIMA/SARIMA/Auto-ARIMA โดย ARIMA เป็นวิธีใช้อนุพันธ์ของค่าเฉลี่ยเคลื่อนที่บวกกับระยะถอยหลังอัตโนมัติ (ใช้ค่าในอดีตที่มีสัญญาณรบกวน) เพื่อทำนายค่าในอนาคต SARIMA เพิ่มส่วนการคำนวณที่คำนึงถึงฤดูกาลและ Auto-ARIMA จะทำการค้นหาเพื่อลองและค้นหาพารามิเตอร์รุ่นที่เหมาะสมที่สุด

- Prophet เป็นวิธีที่ถูกพัฒนาโดย Facebook และตั้งแต่เปิดแหล่งที่มา Prophet เป็นแบบจำลองถดถอยที่รวมเอาแนวโน้มการเติบโตเชิงเส้นหรือโลจิสติก ส่วนประกอบตามฤดูกาล และการตรวจจับจุดเปลี่ยน

### 2. การใช้แบบจำลองหลายตัวแปรเพื่อพยากรณ์ค่าเดียว (Global Univariate Models)

1) มีตัวแปรเสริมจำนวนมากและกำลังทำการพยากรณ์ค่าเดียว  
2) ไม่ทราบถึงฤดูกาลหรือแนวโน้มของโมเดล ข้อดีของแบบจำลองการเรียนรู้ของเครื่อง (Machine Learning) คือ การออกแบบสามารถตรวจจับรูปแบบในข้อมูลที่ผู้สังเกตการณ์มองไม่เห็นในทันที แบบจำลองโกลบอลเหล่านี้ประสบความสำเร็จมากกว่า

3) ต้องทำการฝึกอนุกรมเวลาจำนวนมากในแบบจำลองเดียว สำหรับการใช้งานการเรียนรู้เชิงลึก (Deep Learning) หลายๆ รูปแบบ แบบจำลองสามารถเรียนรู้แบบจำลองอนุกรมเวลาหลายตัวพร้อมกันได้

### ตัวอย่างการเลือกแบบจำลองโกลบอลพยากรณ์ตัวแปรเดียว (Global Univariate Models)

- SARIMAX SARMAX เป็นวิธีที่คล้ายแบบจำลอง SARIMA เพิ่มการพิจารณาตัวแปรภายนอก เพื่อให้อนุกรมเวลาปรับตัวเข้ากับตัวแปรที่เปลี่ยนแปลงได้เร็วขึ้น

- Tree-Based Methods เป็นวิธีที่สามารถแก้ปัญหาได้เกือบทุกปัญหาด้วยความสำเร็จบางอย่าง รวมถึงอนุกรมเวลา วิธีการนี้มีแนวโน้มที่จะถูกต้อง และค่อนข้างเร็วในการฝึกอบรมเมื่อเปรียบเทียบกับวิธีการเชิงลึก การใช้งานปัจจุบันที่ได้รับความนิยมมากที่สุดคือ lightgbm และ xgboost

- MLP-Based Methods เป็นวิธีการใช้โครงข่ายประสาทเทียมแบบคลาสสิกที่เชื่อมต่ออย่างสมบูรณ์ การใช้งานที่ได้รับความนิยมมากที่สุดในปัจจุบันคือ N-BEATS และ GP Forecaster

- CNN-Based Methods เป็นวิธีการคล้ายกับ MLP ยกเว้นว่าไม่ได้เชื่อมต่ออย่างสมบูรณ์ CNN ถูกใช้อย่างแพร่หลายเพราะมีขนาดเล็กกว่า สั้นเปลืองน้อยกว่า และฝึกง่ายกว่า

- RNN/LSTM-Based Methods โดย RNN คือ โครงข่ายประสาทเทียมวนซ้ำช่วยจดจำข้อมูลที่ตามมา ทำให้สามารถคาดการณ์แบบไดนามิกได้มากขึ้น เนื่องจากอนุกรมเวลาขึ้นอยู่กับค่าก่อนหน้านี้ตามธรรมชาติ ส่วน LSTMs เป็น RNN ประเภทหนึ่งที่เฉพาะเจาะจงและเป็นที่ยอมรับมากกว่า

### 3. การใช้แบบจำลองโกลบอลเพื่อพยากรณ์หลายตัวแปร (Global Multivariate Models)

1) มีตัวแปรเสริมจำนวนมากและกำลังทำการพยากรณ์หลายค่า

2) ไม่ทราบถึงฤดูกาลหรือแนวโน้มของโมเดล

3) มีความจำเป็นที่ต้องฝึกอนุกรมเวลาหลายชุดสำหรับตัวแปรต่างๆ มากมาย

### ตัวอย่างการเลือกแบบจำลองโกลบอลพยากรณ์หลายตัวแปร (Global Univariate Models)

- RNN/LSTM-Based Methods เป็นวิธีการที่เกือบทุกการใช้งานแบบจำลองโกลบอลหลายตัวแปรเป็นรูปแบบหนึ่งของโมเดล RNN/LSTM ซึ่งมีความแตกต่างเล็กน้อยระหว่างกัน

## 2.7 การเข้ารหัสคุณลักษณะตามวัฏจักร

Pierre-Louis Bescond [12] นำเสนอเทคนิคง่ายๆ ในการแปลงลักษณะต่างๆ เช่น เวลา สัปดาห์ เดือน หรือฤดูกาล และยังคงรักษานัยสำคัญตามวัฏจักรไว้

เมื่อจัดการกับข้อมูลอนุกรมเวลา โดยทั่วไปจะอยู่ในรูปแบบวันที่และเวลา (YYYY-MM-DD HH:MM:SS) และมักไม่ได้ถูกนำมาใช้ในการพยากรณ์ นอกจากใช้สำหรับการเรียงข้อมูลจากน้อยไปหามาก อย่างไรก็ตาม อาจมีรูปแบบที่ซ่อนอยู่ในชุดข้อมูลของคุณ ซึ่งคุณสมบัติทั่วไปจะไม่เปิดเผย เช่น ชั่วโมงของวัน วันในสัปดาห์ เดือน ฤดูกาล ฯลฯ ซึ่งมีความท้าทายในการแปลงข้อมูลนี้ให้เป็นคุณสมบัติที่ดีความได้

ในการเข้ารหัสเวลาตั้งแต่ 00:00 น. ถึง 23:59 น. อาจแปลงเป็นจำนวนนาฬิกาที่ผ่านไปตั้งแต่ 00:00 น. จาก 0 ถึง 1439 แต่มีฟังก์ชันบางอย่างที่มีพฤติกรรมเป็นวัฏจักร นั่นคือ โคไซน์ (cosine) เพราะมันแปรผันระหว่าง -1 และ 1 ตามวัฏจักร  $0-2\pi$

```

1 # We normalize x values to match with the 0-2π cycle
2 df["x_norm"] = 2 * math.pi * df["x"] / df["x"].max()
3
4 df["cos_x"] = np.cos(df["x_norm"])
5
6 display(df)
7
8 import plotly.graph_objects as go
9
10 fig = go.Figure()
11
12 fig.add_trace(go.Scatter(x=df.x_norm, y=df.cos_x))
13
14 fig.update_layout(yaxis=dict(scaleanchor="x", scaleratio=1))
15
16 fig.show()

```

### ภาพที่ 2.4 การคำนวณโคไซน์

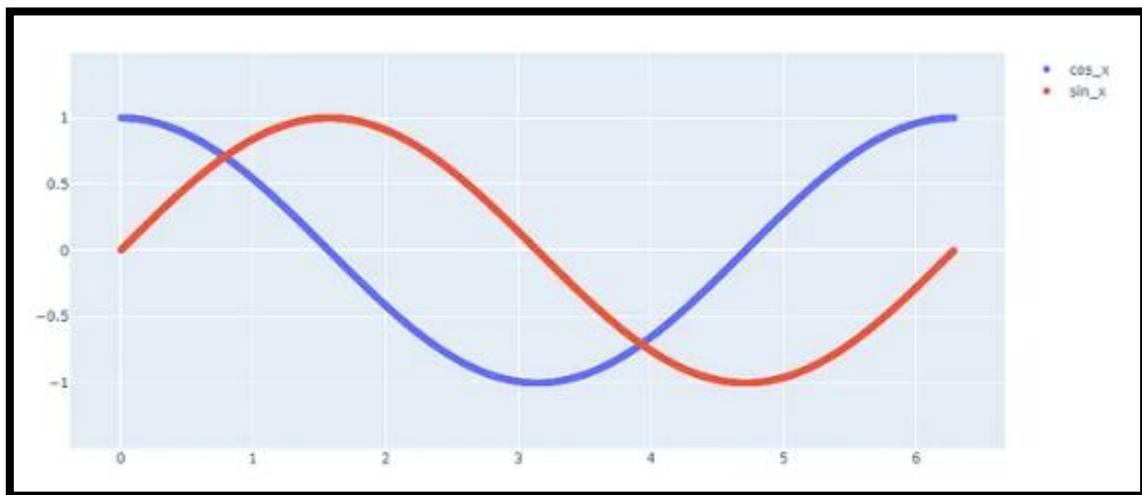
แต่การเรียกใช้ฟังก์ชันโคไซน์หลังจากทำให้  $x$  เป็นมาตรฐานระหว่าง 0 ถึง  $2\pi$  ซึ่งสอดคล้องกับหนึ่งรอบโคไซน์ (cosine) แต่เวลา 00:00 และ 23:59 ทั้งคู่มีค่าประมาณ 1! ซึ่งวิธีที่ดีที่สุดในการแก้ปัญหาใหม่นี้คือการเพิ่มข้อมูลวัฏจักรอื่นเพื่อแยกแยะสองครั้งที่มีค่าโคไซน์ (cosine) เหมือนกัน เราอาจคิดว่ามันเป็นระบบพิกัดสองแกน ซึ่งเรียกรูปอีกตัวว่าไซน์ (sine)

```

1 df["sin_x"] = np.sin(df["x_norm"])
2
3 import plotly.graph_objects as go
4
5 fig = go.Figure()
6
7 fig.add_trace(go.Scatter(x=df.x_norm, y=df.cos_x, name='cos_x'))
8 fig.add_trace(go.Scatter(x=df.x_norm, y=df.sin_x, name='sin_x'))
9
10 fig.update_layout(yaxis = dict(scaleanchor = "x", scaleratio = 1))
11
12 fig.show()

```

ภาพที่ 2.5 การคำนวณไซน์



ภาพที่ 2.6 กราฟไซน์และโคไซน์

จากการแปลงเวลาดังเดิมเป็นค่าโคไซน์และไซน์ที่สอดคล้องกัน ทำให้สามารถกำหนดพิกัดที่แตกต่างกันสำหรับทุกช่วงเวลาระหว่าง 00:00 น. ถึง 23:59 น. ทำให้ไม่ซ้ำกันได้

## 2.8 การคำนวณหรือค้นหาระยะทางระหว่างจุดละติจูดและลองจิจูด

Douglas Karr [13] นำเสนอการคำนวณทางภูมิศาสตร์ เพื่อหาระยะทางระหว่างสถานที่สองแห่ง โดยใช้สูตร Haversine ซึ่งใช้ตรีโกณมิติเพื่อให้ได้ความโค้งของโลก เมื่อคุณหาระยะทางระหว่างสถานที่ 2 แห่งบนโลก เส้นตรงคือส่วนโค้งเหมือนแผ่นที่จริงของทีวบินและสังเกตเห็นว่ามันโค้งเป็นเพราะการบินเป็นโค้งระหว่างสองจุดนั้นสั้นกว่าตรงไปยิ่ง โดยสามารถใช้ Python คำนวณได้ดังนี้

```

from numpy import sin, cos, arccos, pi, round

def rad2deg(radians):
    degrees = radians * 180 / pi
    return degrees

def deg2rad(degrees):
    radians = degrees * pi / 180
    return radians

def getDistanceBetweenPointsNew(latitude1, longitude1, latitude2, longitude2, unit = 'miles'):

    theta = longitude1 - longitude2

    distance = 60 * 1.1515 * rad2deg(
        arccos(
            (sin(deg2rad(latitude1)) * sin(deg2rad(latitude2))) +
            (cos(deg2rad(latitude1)) * cos(deg2rad(latitude2)) * cos(deg2rad(theta)))
        )
    )

    if unit == 'miles':
        return round(distance, 2)
    if unit == 'kilometers':
        return round(distance * 1.609344, 2)
    
```

ภาพที่ 2.7 ตัวอย่างการคำนวณระยะทางด้วยภาษา Python

## 2.9 การวัดแนวโน้มเข้าสู่ส่วนกลาง

ออร์ทอ ซัวเจริญ [14] ให้ความหมายการวัดแนวโน้มเข้าสู่ส่วนกลาง คือ การหาเลขจำนวนเดียว ๆ จำนวนหนึ่ง ซึ่งใช้แทนค่ากลาง ๆ ของข้อมูล หรือที่เรียกกันทั่ว ๆ ไปว่าค่าเฉลี่ย (Average)

ค่าเฉลี่ยเป็นคุณสมบัติอย่างหนึ่งที่สำคัญของข้อมูลซึ่งแสดงถึงการโน้มเข้าหาส่วนกลางของข้อมูล ซึ่งตามวิธีการวัดแนวโน้มเข้าสู่ส่วนกลางที่ดีควรมีคุณสมบัติดังนี้

- ก. เป็นค่าไม่ลำเอียง คือ เป็นค่ากลาง ๆ
- ข. เป็นศูนย์กลางของการแจกแจง
- ค. เข้าใจง่ายและสื่อความหมายได้ดี
- ง. เป็นวิธีที่ใช้กันอย่างกว้างขวาง และใช้ประโยชน์ในการเปรียบเทียบข้อมูล

วิธีวัดแนวโน้มเข้าสู่ส่วนกลาง เช่น มัชฌิมเลขคณิต สามารถคำนวณได้ดังนี้

1. มัชฌิมเลขคณิต (Arithmetic Mean)

เป็นวิธีที่รู้จักกันอย่างแพร่หลาย หรือบางครั้งเรียกสั้น ๆ ว่าค่าเฉลี่ย (Average หรือ Mean)

มัชฌิมเลขคณิต คือ ผลรวมของค่าสังเกตทุกค่าหารด้วยจำนวนค่าสังเกตทั้งหมด การวัดโดยวิธีนี้มีทั้งข้อดีและข้อเสีย ดังนี้คือ

ข้อดี

1. เข้าใจและคำนวณได้ง่าย
2. การคำนวณใช้ค่าสังเกตทุกค่าที่รวบรวมได้
3. สามารถหาค่าของมัชฌิมเลขคณิตได้เสมอ และเป็นค่าที่แน่นอน
4. เหมาะสำหรับการนำไปใช้คำนวณค่าทางค่าต่าง ๆ ทางสถิติ
5. ส่วนเบี่ยงเบนของค่าสังเกตจากมัชฌิมเลขคณิตจะมีค่าน้อยที่สุด

### ข้อเสีย

1. เนื่องจากมัชฌิมเลขคณิต ใช้ค่าสังเกตทุกค่า ดังนั้นจึงเปลี่ยนแปลงได้ง่าย ถ้าค่าสังเกตบางค่าที่รวบรวมได้มีค่าผิดปกติ ก็จะทำให้มัชฌิมเลขคณิตผิดปกติไปด้วย
2. ค่ามัชฌิมเลขคณิตที่คำนวณได้จะตรงกับค่าสังเกตที่มีอยู่จริง ๆ น้อยมากหรือไม่มีเลย หลักในการพิจารณาว่าจะนำมัชฌิมเลขคณิตไปใช้ในการหาแนวโน้มเข้าสู่ส่วนกลางมีหลักในการพิจารณาดังนี้

1. เมื่อค่าสังเกต แต่ละค่ามีค่าใกล้เคียงกัน
2. เมื่อต้องการวัดการกระจายที่น้อยที่สุด
3. เมื่อต้องการมีมัชฌิมที่เชื่อถือได้มากที่สุด
4. เมื่อต้องการมัชฌิมไปใช้ในการคำนวณค่าต่าง ๆ ในทางสถิติต่อไป

วิธีคำนวณหามัชฌิมเลขคณิต มีวิธีหาดังต่อไปนี้

การหามัชฌิมเลขคณิตสำหรับข้อมูลประเภทนี้หาได้จากสูตร

$$\bar{X} = \frac{x_1 + x_2 + \dots + x_N}{N} \quad (2.15)$$

หรือ 
$$\bar{X} = \frac{\sum_{i=1}^N x_i}{N} \quad (2.16)$$

หรือ 
$$\bar{X} = \frac{\Sigma X}{N} \quad (2.17)$$

เมื่อ  $\bar{X}$  = มัชฌิมเลขคณิต

N = จำนวนข้อมูลทั้งหมด

$x_1 + x_2 + \dots + x_n$  = ข้อมูลแต่ละข้อมูลที่รวบรวมได้

2. มัชฌิมเลขคณิตแบบถ่วงน้ำหนัก (Weight Arithmetic Mean)

ถ้าข้อมูลมี  $x_1 + x_2 + \dots + x_n$  มีน้ำหนักของแต่ละข้อมูลดังนี้คือ  $w_1 + w_2 + \dots + w_n$  ตามลำดับ มัชฌิมเลขคณิตแบบถ่วงน้ำหนัก ( $\bar{X}_w$ )

วิธีคำนวณหามัชฌิมเลขคณิตแบบถ่วงน้ำหนัก (Weight Arithmetic Mean) มีวิธีหาดังต่อไปนี้

$$\bar{X}_w = \frac{x_1 w_1 + x_2 w_2 + \dots + x_N w_N}{w_1 + w_2 + \dots + w_N} \quad (2.18)$$

## 2.10 งานวิจัยที่เกี่ยวข้อง

สำหรับงานวิจัยที่เกี่ยวกับการพยากรณ์ฝุ่น PM2.5 ซึ่งฝุ่น PM2.5 สามารถส่งผลกระทบต่อ สุขภาพอนามัยของประชาชน สิ่งแวดล้อม โดยผู้วิจัยได้ทำการศึกษาซึ่งสามารถสรุปได้ดังต่อไปนี้

งานวิจัยเรื่อง “Time series forecasting using a hybrid ARIMA and neural network model” (Peter Zhang, 2003) [3] ถือเป็นงานวิจัยที่มีจุดเริ่มต้นและถูกกล่าวถึงแนวคิดแบบจำลองไฮบริด (Hybrid Model) มาจนถึงปัจจุบัน ทำการวิจัยเพื่อที่เป็นการนำแบบจำลองมาใช้ในการพยากรณ์รวมกัน ประกอบไปด้วยแบบจำลอง ARIMA และ ANN เพื่อใช้ประโยชน์จากความแข็งแกร่งที่เป็นเอกลักษณ์ของแบบจำลอง โดยให้นำแบบจำลอง ARIMA พยากรณ์ข้อมูลที่มีโครงสร้างเชิงเส้นและพยากรณ์ข้อมูลที่มีโครงสร้างที่ไม่เป็นเชิงเส้นซึ่งเป็นส่วนที่เหลือ (Residual) ด้วยแบบจำลอง ANN โดยใช้ข้อมูล 3 ชุด ข้อมูลชุดแรกเป็นข้อมูลจุดดับบนดวงอาทิตย์ (Sunspot) ระหว่างปี ค.ศ. 1700 ถึง ค.ศ. 1987 จำนวน 288 กลุ่มตัวอย่าง ข้อมูลชุดที่สองเป็นข้อมูลจำนวนแมวป่าชนิดหนึ่ง ทางตอนเหนือของแคนาดา (Lynx) ระหว่างปี ค.ศ. 1821 ถึง ค.ศ. 1934 จำนวน 114 กลุ่มตัวอย่าง ข้อมูลชุดสามเป็นข้อมูลอัตราแลกเปลี่ยนระหว่างเงินปอนด์อังกฤษกับดอลลาร์สหรัฐ (Exchange rate) ระหว่างปี ค.ศ. 1980 ถึง ค.ศ. 1993 จำนวน 731 กลุ่มตัวอย่าง

ผลการทดลองพบว่าค่าเฉลี่ยความผิดพลาดกำลังสอง (Mean Square Error: MSE) และค่าเฉลี่ยความผิดพลาดสัมบูรณ์ (Mean Absolute Deviation: MAD) ของแบบจำลองไฮบริด ได้ค่าเฉลี่ยความผิดพลาดน้อยที่สุด จึงสรุปได้ว่าแบบจำลองไฮบริดสามารถให้ค่าพยากรณ์ที่ถูกต้องแม่นยำกว่าเมื่อเปรียบเทียบกับการพยากรณ์ด้วยแบบจำลอง ARIMA หรือแบบจำลอง LSTM แบบจำลองเดียว

งานวิจัยเรื่อง “Time series forecasting using a hybrid ARIMA and LSTM model” (Fathi, 2019) [4] ทำการวิจัยเพื่อพยากรณ์ข้อมูลอนุกรมโดยอาศัยแนวคิดแบบจำลองไฮบริด (Hybrid Model) ที่ยังคงใช้แบบจำลอง ARIMA พยากรณ์ข้อมูลที่มีโครงสร้างเชิงเส้นโดยเพิ่มองค์ประกอบเรื่องฤดูกาล (seasonal) เข้าไปในแบบจำลอง และเปลี่ยนจากแบบจำลอง ANN เป็นแบบจำลอง LSTM ที่ยังคงแนวคิดเครือข่ายประสาทเทียม (Neural networks) โดยใช้ข้อมูล 2 ชุด ข้อมูลชุดแรกเป็นข้อมูลจุดดับบนดวงอาทิตย์ (Wolf's Sunspot) ระหว่างปี ค.ศ. 1700 ถึง ค.ศ. 1988 จำนวน 289 กลุ่มตัวอย่าง ข้อมูลชุดที่สองเป็นข้อมูลอัตราแลกเปลี่ยนระหว่างเงินปอนด์อังกฤษกับดอลลาร์สหรัฐ (Exchange rate) ระหว่างปี ค.ศ. 1971 ถึง ค.ศ. 2018 จำนวน 11,939 กลุ่มตัวอย่าง

ผลการทดลองพบว่า มีเพียงรากที่สองของค่าเฉลี่ยความผิดพลาดลอการิทึมกำลังสอง (Root Mean Squared Logarithmic Error: RMSLE) จากแบบจำลอง ARIMA เท่านั้นที่ผลการพยากรณ์ดีกว่าค่าของแบบจำลอง LSTM และแบบจำลองไฮบริด นอกนั้นค่าเฉลี่ยสเกลความคลาดเคลื่อนสัมบูรณ์ (Mean Absolute Scaled Error: MASE) และค่าเฉลี่ยของเปอร์เซ็นต์ความผิดพลาดสัมบูรณ์ (Mean Absolute Percentage Error : MAPE) แบบจำลองไฮบริดได้ค่าเฉลี่ยความผิดพลาดน้อยที่สุดเมื่อเปรียบเทียบกับการพยากรณ์ด้วยแบบจำลอง ARIMA หรือแบบจำลอง LSTM แบบจำลองเดียว

งานวิจัยเรื่อง “Prediction of Satellite Time Series Data Based on Long Short Term Memory-Autoregressive Integrated Moving Average Model (LSTM-ARIM)” (Yuwei, C. & Kaizhi, W., 2019) [15] ทำการวิจัยเพื่อพยากรณ์การดำเนินการบนวงโคจรของดาวเทียมด้วยข้อมูลอนุกรมเวลาตัวแปรการวัดและส่งข้อมูลทางไกลจำนวนมากด้วยแนวคิดแบบจำลอง LSTM-ARIMA โดยเป็นการคำนวณจากทั้ง 2 แบบจำลองแล้วจึงหาค่าน้ำหนัก (Weight) ที่เหมาะสมจากการคำนวณด้วยทั้งสองแบบจำลอง

ผลการทดลองพบว่า การคำนวณด้วยแบบจำลอง LSTM ด้วยค่าน้ำหนัก (Weight) เท่ากับ 0.5 และแบบจำลอง ARIMA ด้วยค่าน้ำหนัก (Weight) เท่ากับ 0.5 ให้ค่ารากที่สองของค่าเฉลี่ยความผิดพลาดกำลังสอง (Root Mean Squared Error: RMSE) ที่น้อยที่สุด จึงเป็นค่าน้ำหนักที่เหมาะสมของการพยากรณ์

งานวิจัยเรื่อง “Multi-feature PM2.5 Prediction with ARIMA-LSTM” (J. Xiao, Q. Wang, J. Cui, & J. Yu., 2022) [6] ทำการวิจัยเพื่อพยากรณ์ข้อมูลค่าฝุ่น PM2.5 โดยใช้ข้อมูลรายชั่วโมงจากสถานที่แห่งหนึ่งประกอบไปด้วยความชื้น (Humidity) ความกดอากาศ (Pressure) อุณหภูมิ (Air Temperature) ปริมาณน้ำฝน (Rainfall) หิมะ (Snow) และภูมิภาค (Region) ด้วยแบบจำลอง Arima-LSTM จำนวน 17,500 กลุ่มตัวอย่าง

ผลการทดลองพบว่า รากที่สองของค่าเฉลี่ยความผิดพลาดกำลังสอง (Root Mean Squared Error: RMSE) และรากที่สองของค่าเฉลี่ยความผิดพลาดเปอร์เซ็นต์กำลังสอง (Root Mean Squared Percent Error: RMSPE) ของแบบจำลองไฮบริด ได้ค่าเฉลี่ยความผิดพลาดน้อยที่สุด จึงสรุปได้ว่าแบบจำลองไฮบริดสามารถให้ค่าพยากรณ์ที่ถูกต้องแม่นยำกว่าเมื่อเปรียบเทียบกับการพยากรณ์ด้วยแบบจำลอง LSTM อย่างเดียว

งานวิจัยเรื่อง “A Hybrid Deep Learning Model for Predicting PM2.5” (Yuan, P., Mei, YF., Zhong, YL. & Xia, Y, 2022) [16] ทำการวิจัยเพื่อพยากรณ์ข้อมูลค่าฝุ่น PM2.5 โดยใช้ข้อมูลคุณภาพอากาศและข้อมูลอุตุนิยมวิทยาของเมืองฉางซาตั้งแต่วันที่ 1 มกราคม ค.ศ. 2020 ถึง 31 ธันวาคม ค.ศ. 2021 ประกอบไปด้วยความเข้มข้นของสารมลพิษ 6 ชนิดประกอบไปด้วย ฝุ่น PM2.5 ฝุ่น PM10 โอโซน (O<sub>3</sub>) ซัลเฟอร์ไดออกไซด์ (SO<sub>2</sub>) คาร์บอนมอนอกไซด์ (CO) และ ไนโตรเจนไดออกไซด์ (NO<sub>2</sub>) ด้วยแบบจำลองการเรียนรู้เชิงลึก (Hybrid Deep Learning Model: HDL) สำหรับการทำนายความเข้มข้นของ PM2.5 เทียบกับอีก 4 แบบจำลองประกอบไปด้วย แบบจำลองโครงข่ายประสาทเทียมแบบวนกลับ (Simple RNN Model: RNN) แบบจำลองโครงข่ายประสาทเทียมแบบลึก (Long Short - Term Memory: LSTM) แบบจำลองโครงข่ายประสาทเทียมแบบลึก (Long Short - Term Memory: LSTM) แบบจำลองโครงข่ายประสาทเทียมแบบลึก (Long Short - Term Memory: LSTM) แบบจำลองโครงข่ายประสาทเทียมแบบลึก (Long Short - Term Memory: LSTM) แบบจำลองโครงข่ายประสาทเทียมแบบลึก (Long Short - Term Memory: LSTM) โดยแบ่งระยะเวลาการพยากรณ์เป็นล่วงหน้า 1 ชั่วโมง 6 ชั่วโมง 12 ชั่วโมง และ 24 ชั่วโมง

ผลการทดลองพบว่า ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Error: MAE) ค่าเฉลี่ย



ของร้อยละความผิดพลาดสัมบูรณ์ (Mean Absolute Percentage Error: MAPE) ค่าเฉลี่ยความผิดพลาดกำลังสอง (Mean Square Error: MSE) รากที่สองของค่าเฉลี่ยความผิดพลาดกำลังสอง (Root Mean Squared Error: RMSE) ทุกค่าคลาดเคลื่อนของแบบจำลองการเรียนรู้เชิงลึก (Hybrid Deep Learning Model: HDL) ได้ค่าเฉลี่ยความผิดพลาดน้อยที่สุด และได้ค่าและค่าความผันแปรของตัวแปรตอบสนอง (R-Squared:  $R^2$ ) มากที่สุด ในทุกระยะเวลาการพยากรณ์ จึงสรุปได้ว่าแบบจำลองการเรียนรู้เชิงลึกสามารถให้ค่าพยากรณ์ที่ถูกต้องแม่นยำกว่าเมื่อเปรียบเทียบกับการพยากรณ์ด้วยแบบจำลองอื่น

### บทที่ 3 ระเบียบวิธีวิจัย

งานวิจัยครั้งนี้มีวัตถุประสงค์เพื่อการศึกษาประสิทธิภาพการพยากรณ์ฝุ่น PM2.5 ในเขตกรุงเทพมหานคร โดยแบบจำลองไฮบริดร่วมกับค่าถ่วงน้ำหนัก ซึ่งมีรายละเอียดดังต่อไปนี้

- 3.1 กำหนดกลุ่มตัวอย่าง
- 3.2 การออกแบบการทดลอง
- 3.3 ขั้นตอนการทดลอง
- 3.4 การเตรียมข้อมูล
- 3.5 เกณฑ์ที่ใช้ในการประเมินแบบจำลอง
- 3.6 เครื่องมือที่ใช้ในการวิจัย

#### 3.1 กำหนดกลุ่มตัวอย่าง

ผู้วิจัยทำหนังสือขอความอนุเคราะห์จากกรมควบคุมมลพิษ จำนวน 12 สถานี โดยใช้ข้อมูลรายชั่วโมงระหว่าง 1 พฤศจิกายน 2565 เวลา 0:00 น. ถึง 31 มกราคม 2566 เวลา 23:00 น. จำนวน 2208 รายการต่อ 1 สถานี โดยช่วงเวลาดังกล่าวเป็นช่วงเวลา สำหรับกรุงเทพมหานครและปริมณฑลที่ปริมาณ PM2.5 มีค่าเกินเกณฑ์มาตรฐานในปี 2562 [17]

#### ตารางที่ 3.1 พิจารณาตัวแปรที่ใช้ในการวิจัยเบื้องต้น

ตัวแปร	สถานี											
	02T	03T	05T	10T	11T	12T	50T	52T	53T	54T	59T	61T
วัน เดือน ปี												
เวลา												
ความเร็วลม							X					
ทิศทางลม							X					
อุณหภูมิ												
ความชื้นสัมพัทธ์		X		X								X
ความดันบรรยากาศ		X	X	X				X			X	X

หมายเหตุ: X หมายถึง ไม่มีการจัดเก็บค่าของตัวแปร

จากตารางที่ 3.1 พิจารณาตัดสถานีที่ 50T คือ โรงพยาบาลจุฬาลงกรณ์ ริมถนนพระราม 4 เขตปทุมวัน กรุงเทพฯ ออกจากกลุ่มตัวอย่างในการทดลอง เนื่องจากไม่มีตัวแปร ความเร็วลมและทิศทางลม

หลังจากที่เลือกตัวแปรสำหรับการพยากรณ์ได้แล้ว ต้องพิจารณาความสมบูรณ์ของข้อมูลในแต่ละสถานี ซึ่งสามารถแสดงผลได้ดังนี้

### ตารางที่ 3.2 อัตราความสมบูรณ์ของข้อมูล

อัตราความสมบูรณ์ของข้อมูล										
02T	03T	05T	10T	11T	12T	52T	53T	54T	59T	61T
99.50%	0.00%	69.02%	84.24%	100%	0.00%	0.00%	40.62%	98.19%	98.68%	97.10%

จากตารางที่ 3.2 พบว่า หากพิจารณาความสมบูรณ์ของข้อมูลที่ระดับ 95% จะเหลือเพียง 5 สถานี ในกรุงเทพฯ ประกอบด้วย ดังนี้

- 1) สถานี 02T คือ มหาวิทยาลัยราชภัฏบ้านสมเด็จเจ้าพระยา แขวงหิรัญรูจี เขตธนบุรี
- 2) สถานี 11T คือ การเคหะชุมชนห้วยขวาง แขวงดินแดง เขตดินแดง
- 3) สถานี 54T คือ การเคหะชุมชนดินแดง ริมถนนดินแดง เขตดินแดง
- 4) สถานี 59T คือ กรมประชาสัมพันธ์ แขวงพญาไท เขตพญาไท
- 5) สถานี 61T คือ โรงเรียนบดินเดชา (สิงห์ สิงหเสนี) แขวงพลับพลา เขตวังทองหลาง

หลังจากพิจารณาอัตราความสมบูรณ์ของข้อมูล พิจารณาตัวแปรที่ใช้ในการวิจัยซ้ำ ซึ่งสามารถแสดงผลได้ดังนี้

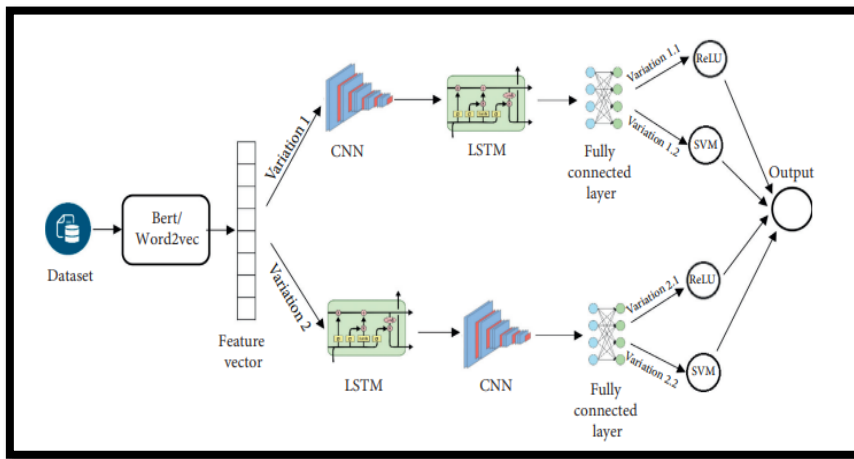
### ตารางที่ 3.3 ตัวแปรที่ใช้ในการวิจัย

ตัวแปร	02T	11T	54T	59T	61T
วัน เดือน ปี					
เวลา					
ความเร็วลม					
ทิศทางลม					
อุณหภูมิ					
ความชื้นสัมพัทธ์					X
ความดันบรรยากาศ				X	X

จากตารางที่ 3.3 พิจารณาตัดสถานที่ที่ 61T คือ โรงเรียนบดินเดชา (สิงห์ สิงหเสนี) แขวงพลับพลา เขตวังทองหลาง กรุงเทพฯ ออกจากกลุ่มตัวอย่างในการทดลอง เนื่องจากไม่มีตัวแปร ความชื้นสัมพัทธ์

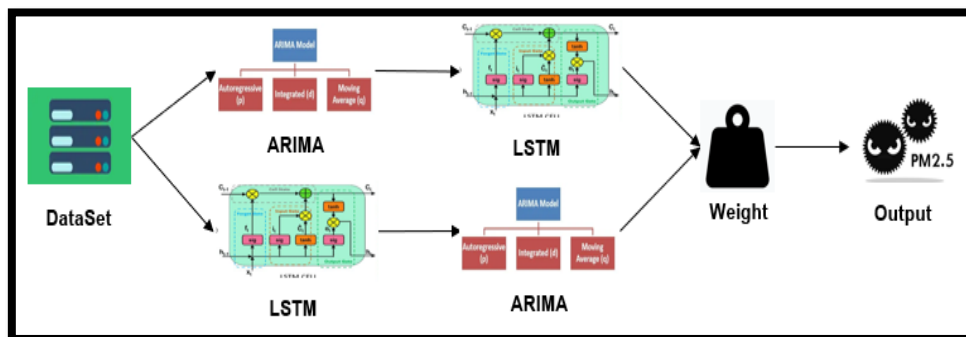
### 3.2 การออกแบบการทดลอง

จากการทบทวนศึกษางานวิจัย “Hybrid Deep Learning Models for Sentiment Analysis” (Dang, Moreno-García & Prieta, 2021) [18] ได้แนวคิดแบบจำลองดังภาพที่ 3.1



ภาพที่ 3.1 แนวคิดริเริ่มของแบบจำลอง

เมื่อพิจารณาแนวคิดแบบจำลองดังภาพที่ 3.1 พบว่า แบบจำลอง ARIMA มีข้อจำกัดที่ไม่สามารถทำ Fully connected layer ได้ จึงเกิดแนวคิดในการทำการปรับปรุงแบบจำลองโดยใช้วิธีการการปรับค่าถ่วงน้ำหนัก (Weight) แทนซึ่งได้แบบจำลองดังนี้



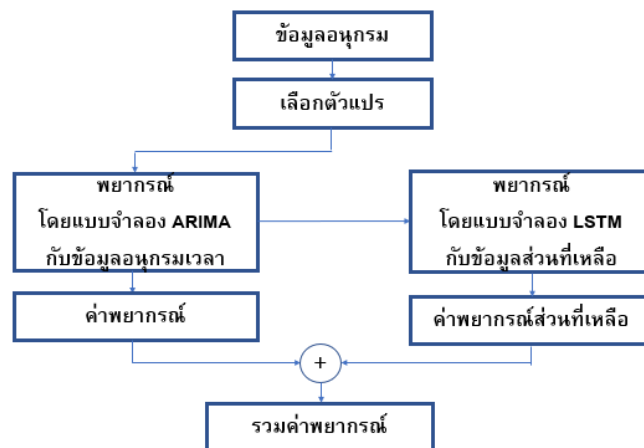
ภาพที่ 3.2 แนวคิดแบบจำลองในการวิจัย

### 3.3 ขั้นตอนการทดลอง

1) ทำการสร้างแบบจำลองไฮบริด ARIMA-LSTM โดยมีขั้นตอนดังนี้

1.1) ทำการพยากรณ์ฝุ่น PM2.5 ด้วยแบบจำลอง ARIMA โดยใช้ข้อมูลตั้งแต่วันที่ 1 พฤศจิกายน พ.ศ. 2565 เวลา 0:00 น. ถึง 31 ธันวาคม พ.ศ. 2566 เวลา 23:00 น. เป็นชุดข้อมูลการเรียนรู้ และใช้ข้อมูลตั้งแต่วันที่ 1 มกราคม พ.ศ. 2566 เวลา 0:00 น. ถึง 31 มกราคม พ.ศ. 2566 เวลา 23:00 น. เป็นชุดข้อมูลการทดสอบ สำหรับการพยากรณ์ ฝุ่น PM2.5 ส่วนแรก ซึ่งแบบจำลอง ARIMA สามารถคำนวณส่วนที่เหลือ (Residual) ได้ทันที

1.2) ทำการพยากรณ์ส่วนที่เหลือ (Residual) โดยใช้ข้อมูลตั้งแต่วันที่ 1 พฤศจิกายน พ.ศ. 2565 เวลา 0:00 น. ถึง 31 ธันวาคม พ.ศ. 2566 เวลา 23:00 น. เป็นชุดข้อมูลการเรียนรู้ และใช้ข้อมูลตั้งแต่วันที่ 1 มกราคม พ.ศ. 2566 เวลา 0:00 น. ถึง 31 มกราคม พ.ศ. 2566 เวลา 23:00 น. เป็นชุดข้อมูลการทดสอบ มาพยากรณ์ด้วยแบบจำลอง LSTM ย้อนหลังหกช่วงเวลา ( $t-6$ ) จากนั้นจึงนำจากค่าฝุ่น PM2.5 ส่วนแรกรวมกับส่วนที่เหลือ

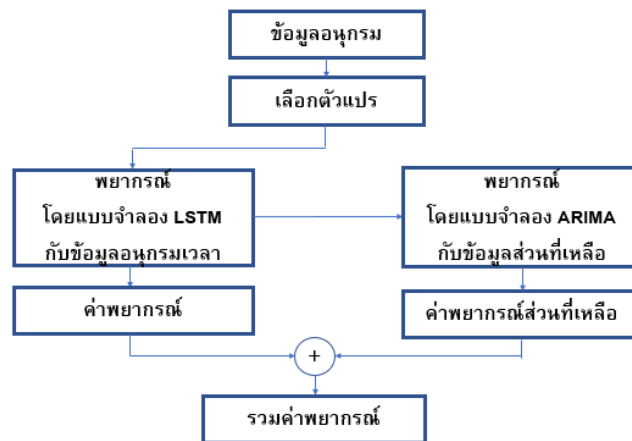


ภาพที่ 3.3 การพยากรณ์ส่วนแบบจำลอง ARIMA-LSTM

2) ทำการสร้างแบบจำลองไฮบริด LSTM-ARIMA โดยมีขั้นตอนดังนี้

2.1) ทำการพยากรณ์ฝุ่น PM2.5 ด้วยแบบจำลอง LSTM โดยใช้ข้อมูลตั้งแต่วันที่ 1 พฤศจิกายน พ.ศ. 2565 เวลา 0:00 น. ถึง 31 ธันวาคม พ.ศ. 2566 เวลา 23:00 น. ย้อนหลังหกช่วงเวลา ( $t-6$ ) เป็นชุดข้อมูลการเรียนรู้ และใช้ข้อมูลตั้งแต่วันที่ 1 มกราคม พ.ศ. 2566 เวลา 0:00 น. ถึง 31 มกราคม พ.ศ. 2566 เวลา 23:00 น. เป็นชุดข้อมูลการทดสอบ สำหรับการพยากรณ์ ฝุ่น PM2.5 ส่วนแรก

2.2) ทำการพยากรณ์ส่วนที่เหลือ (Residual) โดยใช้ข้อมูลตั้งแต่วันที่ 1 พฤศจิกายน พ.ศ. 2565 เวลา 0:00 น. ถึง 31 ธันวาคม พ.ศ. 2566 เวลา 23:00 น. เป็นชุดข้อมูลการเรียนรู้ และใช้ข้อมูล ตั้งแต่วันที่ 1 มกราคม พ.ศ. 2566 เวลา 0:00 น. ถึง 31 มกราคม พ.ศ. 2566 เวลา 23:00 น. เป็นชุดข้อมูลการทดสอบ จากนั้นจึงนำจากค่าฝุ่น PM2.5 ส่วนแรกรวมกับส่วนที่เหลือ



ภาพที่ 3.4 การพยากรณ์ส่วนแบบจำลอง LSTM-ARIMA

แบบจำลอง LSTM ใช้ Vanilla LSTM algorithm

โดย

กำหนด Epoch ในแบบจำลอง LSTM สำหรับแบบจำลอง ARIMA-LSTM เท่ากับ 50

กำหนด Epoch ในแบบจำลอง LSTM สำหรับแบบจำลอง LSTM-ARIMA เท่ากับ 500

```
#LSTM model
model = Sequential()
model.add(LSTM(32, activation='relu', input_shape=(train_X.shape[1], train_X.shape[2])))
model.add(Dense(1))

# summary display
model.summary()

model.compile(optimizer="adam", loss="mean_squared_error", metrics=["acc"], run_eagerly=True)

# fit network
#history = model.fit(train_X, train_y, epochs = 200, batch_size = 72, validation_split = 0.05, verbose = 2, shuffle=False)
history = model.fit(train_X, train_y, epochs = epochs, batch_size = 128, validation_split = 0.05, verbose = 2, shuffle=False)
```

ภาพที่ 3.5 ลอจิกสำหรับแบบจำลอง LSTM

แบบจำลอง Arima ใช้วิธีการคำนวณแบบไม่มี Seasonal และกำหนดให้ p และ q มีค่าไม่เกิน 5 โดยคำสั่ง auto\_arima

```
#Find Parameter by Auto Arima
model = auto_arima(training_set[fieldname], start_p = 1, start_q = 1,
                  test = 'adf',
                  max_p = 5, max_q = 5,
                  m = 1,
                  d = 1,
                  seasonal = False,
                  start_P = 0,
                  D = None,
                  trace = True,
                  error_action = 'ignore',
                  suppress_warnings = True,
                  stepwise = True)
```

**ภาพที่ 3.6** ลอจิกสำหรับแบบจำลอง ARIMA

3) นำผลการพยากรณ์ที่ได้จากแบบจำลอง ARIMA-LSTM และแบบจำลอง LSTM-ARIMA มาทำการคำนวณเพื่อถ่วงน้ำหนัก และเพิ่มประสิทธิภาพในการคำนวณ ด้วยสองวิธีการ เพื่อนำค่าที่ได้จากการคำนวณไปเปรียบเทียบประสิทธิภาพด้วยค่าความคลาดเคลื่อน โดยสูตรการคำนวณเพื่อปรับค่าถ่วงน้ำหนัก (Weight) ที่จะใช้ในการวิจัยประกอบไปด้วย 2 วิธีดังนี้

3.1 วิธีที่ 1 สามารถคำนวณได้ตามสูตรดังนี้

คำนวณจาก

$$\text{NewPredict}_{[i]} = (((\text{ARIMA-LSTM}_{\text{Weight}} \times \text{ARIMA-LSTM}_{\text{Predict}[i]}) + (\text{LSTM-ARIMA}_{\text{Weight}} \times \text{LSTM-ARIMA}_{\text{Predict}[i]})) / 2)$$

โดยที่

$\text{ARIMA-LSTM}_{\text{Predict}[i]}$  คือ ค่าเดิม ณ ลำดับที่  $i$  ที่ได้จากแบบจำลอง ARIMA-LSTM

$\text{LSTM-ARIMA}_{\text{Predict}[i]}$  คือ ค่าเดิม ณ ลำดับที่  $i$  ที่ได้จากแบบจำลอง LSTM-ARIMA

$\text{ARIMA-LSTM}_{\text{Error}}$  คือ ค่า MAPE จากแบบจำลอง ARIMA-LSTM

$\text{LSTM-ARIMA}_{\text{Error}}$  คือ ค่า MAPE จากแบบจำลอง LSTM-ARIMA

$\text{ARIMA-LSTM}_{\text{Weight}}$  คือ ค่าที่คำนวณจาก

$$(1 - (\text{ARIMA-LSTM}_{\text{Error}} / (\text{ARIMA-LSTM}_{\text{Error}} + \text{LSTM-ARIMA}_{\text{Error}}))) \times 2$$

$\text{LSTM-ARIMA}_{\text{Weight}}$  คือ ค่าที่คำนวณจาก  $2 - \text{ARIMA-LSTM}_{\text{Weight}}$

3.2 วิธีที่ 2 สามารถคำนวณตาม Pseudocode ดังนี้

for  $i = 1$  to 10

if  $i = 1$

$$\text{ARIMA-LSTM}_{\text{Ratio}} = 1$$

else

$$\text{ARIMA-LSTM}_{\text{Ratio}} = \text{ARIMA-LSTM}_{\text{Ratio}} - 0.1$$

end if

$$\text{LSTM-ARIMA}_{\text{Ratio}} = 1 - \text{ARIMA-LSTM}_{\text{Ratio}}$$

$$\text{NewPredict}_{[i]} = (\text{ARIMA-LSTM}_{\text{Predict}[i]} \times \text{ARIMA-LSTM}_{\text{Ratio}}) + (\text{LSTM-ARIMA}_{\text{Predict}[i]} \times \text{LSTM-ARIMA}_{\text{Ratio}})$$

end for

โดยที่

$\text{ARIMA-LSTM}_{\text{Predict}[i]}$  คือ ค่าเดิม ณ ลำดับที่  $i$  ที่ได้จากแบบจำลอง ARIMA-LSTM

$\text{LSTM-ARIMA}_{\text{Predict}[i]}$  คือ ค่าเดิม ณ ลำดับที่  $i$  ที่ได้จากแบบจำลอง LSTM-ARIMA

$\text{ARIMA-LSTM}_{\text{Ratio}}$  คือ ค่า Ratio สำหรับแบบจำลอง ARIMA-LSTM

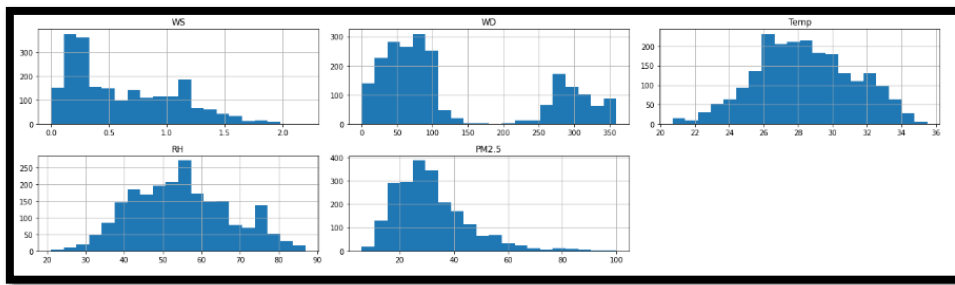
$\text{LSTM-ARIMA}_{\text{Ratio}}$  คือ ค่า MAPE สำหรับแบบจำลอง LSTM-ARIMA

### 3.4 การเตรียมข้อมูล

1) การทำความสะอาดข้อมูล (Data Cleaning) การทำความสะอาดข้อมูล ถือเป็นกระบวนการที่สำคัญในการตรวจสอบและการแก้ไข คลอบคลุมถึงการลบรายการข้อมูลที่ไม่ถูกต้องออกไปจากชุดข้อมูล จึงต้องมีการ การปรับปรุง เพื่อให้ข้อมูลที่นำไปใช้วิเคราะห์ข้อมูลมีคุณภาพ ประกอบไปด้วยการสำรวจข้อมูล (Explore Data) และการแก้ไขข้อมูลที่ผิดพลาด (Correcting Data)

1.1) การสำรวจข้อมูล (Explore Data) โดยใช้ฮิสโตแกรมเป็นเครื่องมือในการสรุปและแปลผลข้อมูลด้วยการแจกแจงความถี่ตามปริมาณที่เกิดขึ้น เพื่อทำให้เกิดความเข้าใจภาพรวมของข้อมูลจากลักษณะการกระจายตัวของกราฟ และยังสามารถตรวจสอบค่าที่ผิดปกติเพื่อวิเคราะห์กระบวนการได้อย่างรวดเร็ว จากการสำรวจข้อมูล ไม่พบข้อมูลที่มีความผิดปกติ





ภาพที่ 3.7 การสำรวจข้อมูลโดยใช้ฮิสโตแกรม

1.2) การแก้ไขข้อมูลที่ผิดพลาด (Correcting Data) เป็นการจัดการข้อมูลที่มีค่าสูญหายในบางส่วน หรือมีค่าข้อมูลผิดปกติเกิดขึ้น รวมถึงข้อมูลที่ซ้ำซ้อน ก่อนนำไปวิเคราะห์ซึ่งประกอบด้วยเทคนิค การลบข้อมูลออกไปจากชุดข้อมูล (Removal) การปรับปรุงข้อมูลให้เป็นค่าที่ถูกต้อง (Correction) และการแทนค่า (Replacement)

```
In [24]: # Look at rows with NAs
df[df.isna().any(axis=1)]

Out[24]:
   Timestamp  WS  WD  Temp  RH  PM2.5  Hour
277 2022-11-12 13:00:00  NaN  NaN  NaN  NaN  NaN  13
278 2022-11-12 14:00:00  NaN  NaN  NaN  NaN  NaN  14
326 2022-11-14 14:00:00  NaN  NaN  NaN  NaN  NaN  14
457 2022-11-20 01:00:00  NaN  NaN  NaN  NaN  NaN  1
458 2022-11-20 02:00:00  NaN  NaN  NaN  NaN  NaN  2
459 2022-11-20 03:00:00  NaN  NaN  NaN  NaN  NaN  3
460 2022-11-20 04:00:00  NaN  NaN  NaN  NaN  NaN  4
461 2022-11-20 05:00:00  NaN  NaN  NaN  NaN  NaN  5
462 2022-11-20 06:00:00  NaN  NaN  NaN  NaN  NaN  6
463 2022-11-20 07:00:00  NaN  NaN  NaN  NaN  NaN  7
464 2022-11-20 08:00:00  NaN  NaN  NaN  NaN  NaN  8

In [25]: #Filling null values with the previous ones
df.fillna(method='pad', inplace=True)

In [26]: # Look at rows with NAs
df[df.isna().any(axis=1)]

Out[26]:
   Timestamp  WS  WD  Temp  RH  PM2.5  Hour

In [27]: duplicateRows = df[df.duplicated()]
duplicateRows

Out[27]:
   Timestamp  WS  WD  Temp  RH  PM2.5  Hour

In [28]: df = df.drop_duplicates(keep='first')

In [29]: duplicateRows = df[df.duplicated()]
duplicateRows

Out[29]:
   Timestamp  WS  WD  Temp  RH  PM2.5  Hour
```

ภาพที่ 3.8 การตรวจสอบเพื่อทำการแก้ไขข้อมูลที่ผิดพลาด

จากการตรวจสอบข้อมูลที่มีค่าสูญหาย ในกรณีที่พบข้อมูลที่ผิดปกติ จะใช้ค่าก่อนหน้ามาแทนค่า (Replacement) และเมื่อทำการตรวจสอบข้อมูลที่ซ้ำซ้อน จะใช้การลบข้อมูลออกไปจากชุดข้อมูล (Removal) โดยเลือกเก็บข้อมูลตัวแรกหลังจากที่พบว่าซ้ำ แต่ไม่พบข้อมูลที่ซ้ำซ้อน

## 2) การสร้างตัวแปรใหม่

2.1) จากสาเหตุหลักของการเกิดฝุ่น PM2.5 ของกรุงเทพมหานคร มาจากมลพิษทางอากาศ ดังนั้นจึงน่าจะมีความแตกต่างระหว่างวันทำงานและวันหยุดสุดสัปดาห์ เนื่องจากปริมาณการสัญจรของรถยนต์ลดลง ซึ่งน่าจะส่งอิทธิพลถึงปริมาณฝุ่น PM2.5 จึงทำการกำหนดให้เกิดความแตกต่างระหว่างวันหยุดสุดสัปดาห์และวันทำงาน โดยกำหนดให้วันหยุดสุดสัปดาห์มีค่าเป็น 1 และกำหนดให้วันทำงานมีค่าเป็น 0

2.2) นำหลักการเข้ารหัสคุณลักษณะตามวัฏจักร สร้างตัวแปรเพิ่มอีก 6 ตัวแปร จากตัวแปรวันเดือนปี (Date) และ เวลา (Hour) Month\_Sin (เดือนไซน์) Month\_Cos (เดือนโคไซน์) Day\_Sin (เดือนไซน์) Day\_Cos (เดือนโคไซน์) Hour\_Sin (เดือนไซน์) Hour\_Cos (เดือนโคไซน์)

หลังจากดำเนินการเตรียมข้อมูลตัวแปรที่ใช้ในการศึกษา สามารถสรุปได้ดังนี้

### 1. ตัวแปรอิสระ

- 1.1 Date (วันเดือนปี) สำหรับสร้าง Index
- 1.2 Month Sin (เดือนไซน์)
- 1.3 Month Cosin (เดือนโคไซน์)
- 1.4 Day Sin (วันไซน์)
- 1.5 Day Cosin (วันโคไซน์)
- 1.6 Hour Sin (ชั่วโมงไซน์)
- 1.7 Hour Cosin (ชั่วโมงโคไซน์)
- 1.8 Weekend (วันสุดสัปดาห์)
- 1.9 Wind speed (ความเร็วลม)
- 1.10 Wind direction (ทิศทางลม)
- 1.11 Temperature (อุณหภูมิ)
- 1.12 Relative Humidity (ความชื้นสัมพัทธ์)

### 2. ตัวแปรตาม ได้แก่ ฝุ่น PM2.5

### 3.5 เกณฑ์ที่ใช้ในการประเมินแบบจำลอง

เกณฑ์ที่ใช้ในการวัดค่าความแม่นยำของโมเดลมีด้วยกันหลากหลายวิธี โดยในการอธิบาย วิธีคำนวณค่าความคลาดเคลื่อนที่จะบ่งชี้ถึงความแม่นยำของโมเดล ซึ่งเกณฑ์ที่ใช้ในการวิจัยใช้ในการประเมินแบบจำลองครั้งนี้ มีดังต่อไปนี้

1) ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Squared Error: RMSE) เป็นวิธีการวัดค่าความคลาดเคลื่อนแบบมาตรฐาน ซึ่งนิยมใช้กันอย่างแพร่หลาย โดยมีสมการดังสมการที่ 3.1 ในการวัดค่าความแม่นยำจากวิธีการนี้ ยิ่งค่า RMSE ที่ได้มีค่าน้อยแสดงว่าโมเดลที่ได้จะมีความแม่นยำมาก

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^n (y_t - \hat{y}_t)^2} \quad (3.1)$$

โดยที่

RMSE คือค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย

$n$  คือ จำนวนข้อมูลที่ใช้

$y_t$  คือ ค่าจริงที่เวลา  $t$  ใด ๆ

$\hat{y}_t$  คือค่าที่ได้จากการพยากรณ์ที่เวลา  $t$  ใด ๆ

2) ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เป็นวิธีที่การวัดค่าความคลาดเคลื่อนที่นิยมอีกวิธีหนึ่ง ซึ่งวิธีนี้จะ ช่วยบอกถึงขนาดของความคลาดเคลื่อนรวมได้ โดยมีสมการดังสมการที่ 3.2 ในการวัดค่าความแม่นยำจากวิธีการนี้ ยิ่งค่าที่ได้มีค่าน้อยแสดงว่าโมเดลที่ได้จะมีความแม่นยำมาก

$$MAPE = \left( \frac{1}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right| \right) \quad (3.2)$$

โดยที่

MAPE คือ ค่าสัมบูรณ์ของเปอร์เซ็นต์ของความคลาดเคลื่อน

$n$  คือ จำนวนข้อมูลที่ใช้

$y_t$  คือ ค่าจริงที่เวลา  $t$  ใด ๆ

$\hat{y}_t$  คือ ค่าที่ได้จากการพยากรณ์ที่เวลา  $t$  ใด ๆ

### 3.6 เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการวิจัยประกอบด้วย

- 1) ฮาร์ดแวร์ (Hardware) ใช้คอมพิวเตอร์โน้ตบุ๊ก 1 เครื่อง
- 2) ภาษาที่ใช้โปรแกรม (Programming Language) ใช้ภาษาไพธอน (Python) เนื่องจากลักษณะของภาษา สามารถเรียนรู้การใช้งานได้ง่าย และรองรับงานวิจัยทางด้านวิทยาศาสตร์ข้อมูล (Data Science)
- 3) โฮสต์โปรแกรม (Host) พัฒนาและทดสอบบน Google Collaboratory ซึ่งเป็นบริการ Software as a Service (Saas) ของโปรแกรม Jupyter Notebook บน Cloud โด Google



ภาพที่ 3.9 เครื่องมือที่ใช้ในการวิจัย

## บทที่ 4

### ผลการทดลอง

#### 4.1 ผลการพยากรณ์เปรียบเทียบปริมาณฝุ่น PM2.5

ผลการพยากรณ์ค่าฝุ่น PM 2.5 สถานีมหาวิทยาลัยราชภัฏบ้านสมเด็จเจ้าพระยา แขวงทิวสุรภูมิ เขตธนบุรี กรุงเทพฯ (O2T) สามารถสรุปได้ดังนี้

#### ตารางที่ 4.1 ผลการพยากรณ์เปรียบเทียบสถานี O2T

แบบจำลอง	ค่าความคลาดเคลื่อน				
	RMSE		MAPE		
1. ARIMA	4.93553		0.11366		
2. LSTM-Multivariate	6.54864		0.15051		
3. LSTM-Univariate	20.65403		0.38795		
แบบจำลองไฮบริด	LSTM-Multivariate		LSTM-Univariate		
	RMSE	MAPE	RMSE	MAPE	
4. ARIMA-LSTM	5.0691	0.11584			
5. LSTM-ARIMA	6.10633	0.13568	5.3158	0.12017	
6. ARIMA-LSTM & LSTM-ARIMA					
6.1 ปรับค่าถ่วงน้ำหนักวิธีที่ 1	5.00367	0.11475	4.81928	0.11154	
6.2 ปรับค่าถ่วงน้ำหนักวิธีที่ 2					
ค่าถ่วงน้ำหนัก (ARIMA-LSTM)	ค่าถ่วงน้ำหนัก (LSTM-ARIMA)				
0.9	0.1	5.0691	0.11584	5.0691	0.11584
0.8	0.2	6.10633	0.13568	5.3158	0.12017
0.7	0.3	4.90954	0.11301	4.83031	0.1115
0.6	0.4	4.95364	0.11376	4.81047	0.11132
0.5	0.5	5.04542	0.11555	4.82158	0.1116

ตารางที่ 4.1 (ต่อ)

แบบจำลองไฮบริด		ค่าความคลาดเคลื่อน			
		LSTM-Multivariate		LSTM-Univariate	
		RMSE	MAPE	RMSE	MAPE
0.4	0.6	5.18236	0.11795	4.86344	0.11233
0.3	0.7	5.36099	0.12128	4.93525	0.11354
6. ARIMA-LSTM & LSTM-ARIMA					
6.2 ปรับค่าถ่วงน้ำหนักวิธีที่ 2					
ค่าถ่วงน้ำหนัก (ARIMA-LSTM)	ค่าถ่วงน้ำหนัก (LSTM-ARIMA)				
0.2	0.8	5.57731	0.1254	5.03574	0.11519
0.1	0.9	5.82713	0.13018	5.16324	0.11748

จากตารางที่ 4.1 สามารถสรุปผลการพยากรณ์ได้ดังนี้

1) การพยากรณ์จากแบบจำลองเดี่ยว

1.1) จากการพยากรณ์โดยแบบจำลอง ARIMA ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 4.93553 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.11366

1.2) จากการพยากรณ์โดยแบบจำลอง LSTM (Multivariate) ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 6.54864 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.15051

1.3) จากการพยากรณ์โดยแบบจำลอง LSTM (Univariate) ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 20.65403 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.38795

2) การพยากรณ์จากแบบจำลองเดี่ยวไฮบริดด้วย ARIMA-LSTM & LSTM (Multivariate)-ARIMA

2.1) จากการพยากรณ์โดยแบบจำลองไฮบริด ARIMA-LSTM ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 5.0691 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.11584



2.9) จากการพยากรณ์โดยแบบจำลองไฮบริด ARIMA-LSTM & LSTM-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 2 โดย ผลการพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.4 และ การพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.6 ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 5.18236 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.11795

2.10) จากการพยากรณ์โดยแบบจำลองไฮบริด ARIMA-LSTM & LSTM-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 2 โดย ผลการพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.3 และ การพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.7 ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 5.36099 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.12128

2.11) จากการพยากรณ์โดยแบบจำลองไฮบริด ARIMA-LSTM & LSTM-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 2 โดย ผลการพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.2 และ การพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.8 ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 5.57731 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.1254

2.12) จากการพยากรณ์โดยแบบจำลองไฮบริด ARIMA-LSTM & LSTM-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 2 โดย ผลการพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.1 และ การพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.9 ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 5.82713 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.13018

3) การพยากรณ์จากแบบจำลองเดี่ยวไฮบริดด้วย ARIMA-LSTM & LSTM (Univariate)-ARIMA

3.1) จากการพยากรณ์โดยแบบจำลองไฮบริด LSTM-ARIMA ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 5.3158 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.12017

3.2) จากการพยากรณ์โดยแบบจำลองไฮบริด ARIMA-LSTM & LSTM-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 1 ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 4.81928 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.11154

3.3) จากการพยากรณ์โดยแบบจำลองไฮบริด ARIMA-LSTM & LSTM-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 2 โดย ผลการพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.9 และ การพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.1 ได้ค่ารากที่สองของค่าความคลาดเคลื่อน



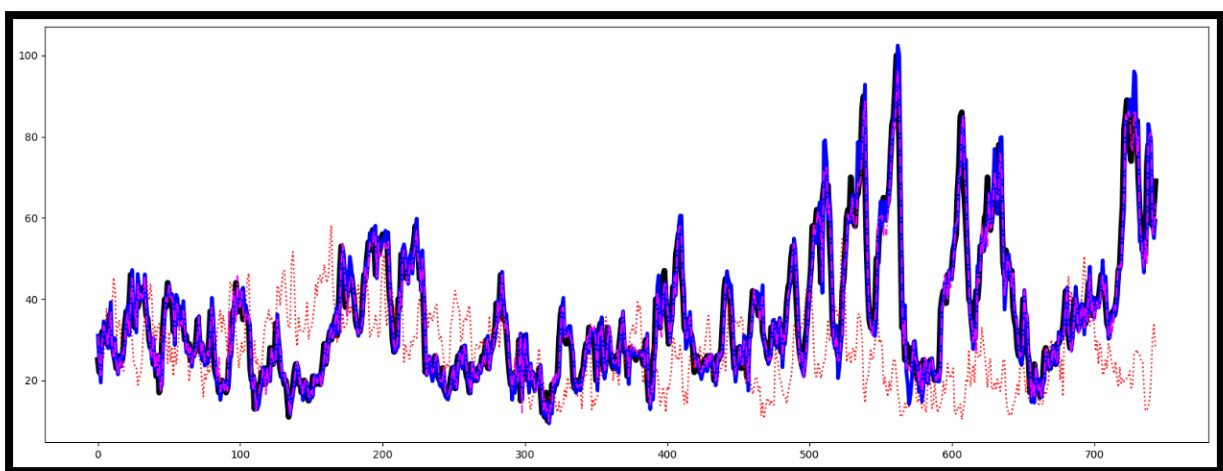


กำลังสองเฉลี่ย (RMSE) เท่ากับ 4.93525 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.11354

3.10) จากการพยากรณ์โดยแบบจำลอง ไฮบริด ARIMA-LSTM & LSTM-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 2 โดย ผลการพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.2 และ การพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.8 ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 5.03574 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.11519

3.11) จากการพยากรณ์โดยแบบจำลอง ไฮบริด ARIMA-LSTM & LSTM-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 2 โดย ผลการพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.1 และ การพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.9 ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 5.16324 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.11748

พิจารณาผลการพยากรณ์ค่าฝุ่น PM 2.5 สถานีมหาวิทยาลัยราชภัฏบ้านสมเด็จเจ้าพระยา แขวง ทรัพยากรูจี เขตธนบุรี กรุงเทพฯ (O2T) พบว่า วิธีการพยากรณ์แบบจำลองไฮบริด ARIMA-LSTM & LSTM-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 2 โดย ผลการพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.6 และ การพยากรณ์จากแบบจำลอง LSTM(Univariate)-ARIMA มีค่าถ่วงน้ำหนักเท่ากับ 0.4 ให้ค่าพยากรณ์ดีที่สุด เมื่อพิจารณาจากค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) และค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) ซึ่งจากผลการพยากรณ์ สามารถแสดงเป็นกราฟได้ ดังภาพที่ 4.1

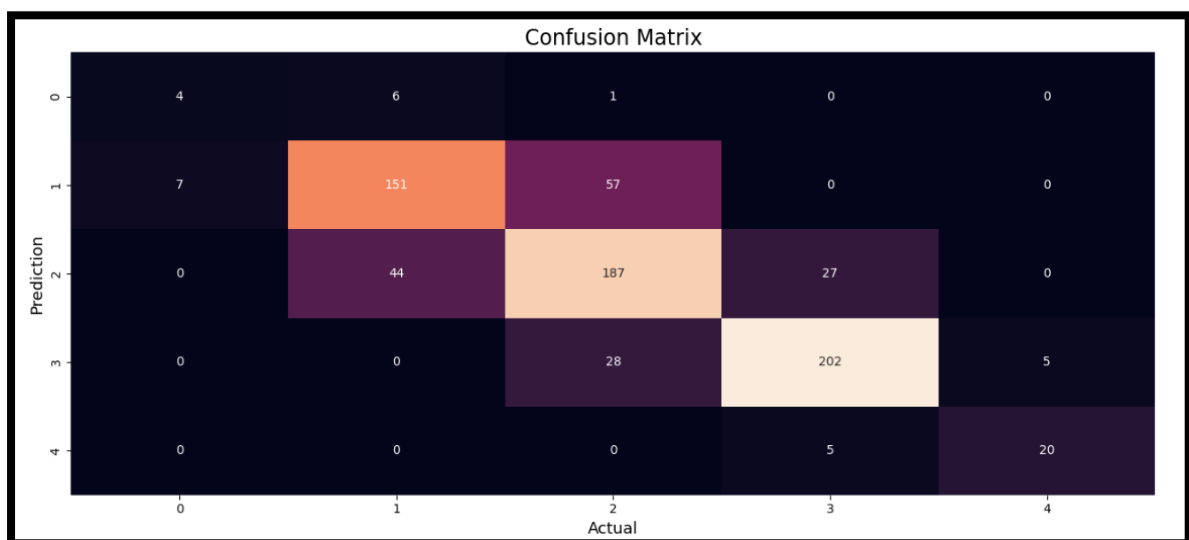


ภาพที่ 4.1 กราฟการพยากรณ์ฝุ่น PM2.5 สถานี O2T

จากภาพที่ 4.1 แสดงกราฟพยากรณ์ฝุ่น PM2.5 สถานี O2T โดยมีรายละเอียดดังนี้

- กราฟเส้นสีดำ แทน ค่าฝุ่น PM2.5
- กราฟเส้นประสีแดงแทนค่าฝุ่น PM2.5 ที่ถูกพยากรณ์โดยแบบจำลอง LSTM (Univariate) ซึ่งจะเห็นว่ามีความแตกต่างจากค่าฝุ่น PM2.5
- กราฟเส้นสีน้ำเงินแทนค่าฝุ่น PM2.5 ที่ถูกพยากรณ์โดยแบบจำลองไฮบริด LSTM (Univariate)-ARIMA ซึ่งผลการพยากรณ์มีแนวโน้มไปในทิศทางเดียวกับค่าฝุ่น PM2.5
- กราฟเส้นประสีชมพูแทนค่าฝุ่น PM2.5 ที่ถูกพยากรณ์โดยแบบจำลองไฮบริด ARIMA-LSTM & LSTM-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 2 โดย ผลการพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.6 และ การพยากรณ์จากแบบจำลอง LSTM (Univariate)-ARIMA มีค่าถ่วงน้ำหนักเท่ากับ 0.4 ซึ่งผลการพยากรณ์ปรับปรุงได้ดีขึ้น

จากผลการพยากรณ์จากค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) จะได้แบบจำลองไฮบริด ARIMA-LSTM & LSTM (Univariate)-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 2 โดยการพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.6 และ การพยากรณ์จากแบบจำลอง LSTM (Univariate)-ARIMA มีค่าถ่วงน้ำหนักเท่ากับ 0.4 เมื่อนำไปพยากรณ์ตามคุณภาพอากาศ โดยกำหนดให้ คุณภาพอากาศดีมากมีค่าเท่ากับ 0 คุณภาพอากาศดีมีค่าเท่ากับ 1 คุณภาพอากาศปานกลางมีค่าเท่ากับ 2 คุณภาพอากาศเริ่มมีผลต่อสุขภาพมีค่าเท่ากับ 3 และคุณภาพอากาศมีผลต่อสุขภาพมีค่าเท่ากับ 4 สามารถแสดงผลได้ดัง ภาพที่ 4.2



ภาพที่ 4.2 ผลลัพธ์การทำนายคุณภาพอากาศของสถานี O2T

จากภาพที่ 4.2 ผลลัพธ์การทำนาย PM2.5 ของสถานี O2T สามารถสรุปผลได้ดังนี้

1) คุณภาพอากาศมีค่าเท่ากับ 0 มีทั้งหมด 11 รายการ ผลการทำนายคุณภาพอากาศมีค่าเท่ากับ 0 มี 4 รายการ คิดเป็นทำนายถูก 36.36% โดยทำนายผิดรวมเป็นการพยากรณ์ค่าสูงกว่าค่าจริง 63.63% โดยผลการทำนายคุณภาพอากาศมีค่าเท่ากับ 1 มี 6 รายการ คิดเป็นทำนายผิด 54.55% ทำนายคุณภาพอากาศมีค่าเท่ากับ 2 มี 1 รายการ คิดเป็นทำนายผิด 9.09%

2) คุณภาพอากาศมีค่าเท่ากับ 1 มีทั้งหมด 215 รายการ ผลการทำนายคุณภาพอากาศมีค่าเท่ากับ 1 มี 151 รายการ คิดเป็นทำนายถูก 70.23% โดยทำนายผิดเป็นการพยากรณ์ค่าต่ำกว่าค่าจริง 3.26% โดยผลการทำนายคุณภาพอากาศมีค่าเท่ากับ 0 มี 7 รายการ ทำนายผิดเป็นการพยากรณ์ค่าสูงกว่าค่าจริง 26.51% โดยผลการทำนายคุณภาพอากาศมีค่าเท่ากับ 2 มี 57 รายการ

3) คุณภาพอากาศมีค่าเท่ากับ 2 มีทั้งหมด 258 รายการ ผลการทำนายคุณภาพอากาศมีค่าเท่ากับ 2 มี 202 รายการ คิดเป็นทำนายถูก 72.48% โดยทำนายผิดเป็นการพยากรณ์ค่าต่ำกว่าค่าจริง 17.05% โดยผลการทำนายคุณภาพอากาศมีค่าเท่ากับ 1 มี 44 รายการ ทำนายผิดเป็นการพยากรณ์ค่าสูงกว่าค่าจริง 10.47% โดยผลการทำนายคุณภาพอากาศมีค่าเท่ากับ 3 มี 27 รายการ

4) คุณภาพอากาศมีค่าเท่ากับ 3 มีทั้งหมด 235 รายการ ผลการทำนายคุณภาพอากาศมีค่าเท่ากับ 3 มี 202 รายการ คิดเป็นทำนายถูก 85.96% โดยทำนายผิดเป็นการพยากรณ์ค่าต่ำกว่าค่าจริง 11.91% โดยผลการทำนายคุณภาพอากาศมีค่าเท่ากับ 2 มี 28 รายการ ทำนายผิดเป็นการพยากรณ์ค่าสูงกว่าค่าจริง 2.13% โดยผลการทำนายคุณภาพอากาศมีค่าเท่ากับ 4 มี 5 รายการ

5) คุณภาพอากาศมีค่าเท่ากับ 4 มีทั้งหมด 25 รายการ ผลการทำนายคุณภาพอากาศมีค่าเท่ากับ 4 มี 20 รายการ คิดเป็นทำนายถูก 80.00% โดยทำนายผิดรวมเป็นการพยากรณ์ค่าต่ำกว่าค่าจริง 20.00% โดยผลการทำนายคุณภาพอากาศมีค่าเท่ากับ 3 มี 5 รายการ

ผลการพยากรณ์ค่าฝุ่น PM 2.5 สถานีการเคหะชุมชนห้วยขวาง แขวงดินแดง เขตดินแดง กรุงเทพฯ (11T) สามารถสรุปได้ดังนี้

ตารางที่ 4.2 ผลการพยากรณ์เปรียบเทียบสถานี 11T

แบบจำลอง	ค่าความคลาดเคลื่อน			
	RMSE		MAPE	
1. ARIMA	2.78227		0.08284	
2. LSTM-Multivariate	4.51227		0.13201	
3. LSTM-Univariate	20.96003		0.53691	
แบบจำลองไฮบริด	LSTM-Multivariate		LSTM-Univariate	
	RMSE	MAPE	RMSE	MAPE
4. ARIMA-LSTM	2.97699	0.08929		
5. LSTM-ARIMA	3.64696	0.10638	3.29844	0.09238
6. ARIMA-LSTM & LSTM-ARIMA				
6.1 ปรับค่าถ่วงน้ำหนักวิธีที่ 1	2.91042	0.08703	2.81365	0.0829
6.2 ปรับค่าถ่วงน้ำหนักวิธีที่ 2				
ค่าถ่วงน้ำหนัก (ARIMA-LSTM)				
ค่าถ่วงน้ำหนัก (LSTM-ARIMA)				
0.9	2.90335	0.08664	2.8925	0.08695
0.8	2.86204	0.08538	2.83296	0.08499

ตารางที่ 4.2 (ต่อ)

แบบจำลองไฮบริด	ค่าความคลาดเคลื่อน			
	LSTM-Multivariate		LSTM-Univariate	
	RMSE	MAPE	RMSE	MAPE
6. ARIMA-LSTM & LSTM-ARIMA				
6.1 ปรับค่าถ่วงน้ำหนักวิธีที่ 1	2.91042	0.08703	2.81365	0.0829
6.2 ปรับค่าถ่วงน้ำหนักวิธีที่ 2				
ค่าถ่วงน้ำหนัก (ARIMA-LSTM)				
ค่าถ่วงน้ำหนัก (LSTM-ARIMA)				
0.7	2.85446	0.0854	2.79997	0.08348
0.6	2.88087	0.08617	2.79446	0.08279
0.5	2.94037	0.08781	2.81659	0.08295
0.4	3.03101	0.09021	2.86573	0.08369
0.3	3.15009	0.09333	2.94053	0.08508
0.2	3.29453	0.09705	3.03908	0.08711
0.1	3.46117	0.10136	3.15917	0.08952

จากตารางที่ 4.2 สามารถสรุปผลการพยากรณ์ได้ดังนี้

1) การพยากรณ์จากแบบจำลองเดี่ยว

1.1) จากการพยากรณ์โดยแบบจำลอง ARIMA ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 2.78227 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.08284

1.2) จากการพยากรณ์โดยแบบจำลอง LSTM (Multivariate) ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 4.51227 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.13201

1.3) จากการพยากรณ์โดยแบบจำลอง LSTM (Univariate) ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 20.96003 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.53691

2) การพยากรณ์จากแบบจำลองเดี่ยวไฮบริดด้วย ARIMA-LSTM & LSTM (Multivariate)-

## ARIMA

2.1) จากการพยากรณ์โดยแบบจำลองไฮบริด ARIMA-LSTM ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 2.97699 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.08929

2.2) จากการพยากรณ์โดยแบบจำลองไฮบริด LSTM-ARIMA ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 3.64696 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.10638

2.3) จากการพยากรณ์โดยแบบจำลองไฮบริด ARIMA-LSTM & LSTM-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 1 ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 2.91042 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.08703

2.4) จากการพยากรณ์โดยแบบจำลองไฮบริด ARIMA-LSTM & LSTM-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 2 โดย ผลการพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.9 และ การพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.1 ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 2.90335 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.08664

2.5) จากการพยากรณ์โดยแบบจำลองไฮบริด ARIMA-LSTM & LSTM-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 2 โดย ผลการพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.8 และ การพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.2 ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 2.86204 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.08538

2.6) จากการพยากรณ์โดยแบบจำลองไฮบริด ARIMA-LSTM & LSTM-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 2 โดย ผลการพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.7 และ การพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.3 ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 2.85446 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.0854

2.7) จากการพยากรณ์โดยแบบจำลองไฮบริด ARIMA-LSTM & LSTM-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 2 โดย ผลการพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.6 และ การพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.4 ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 2.88087 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.08617

2.8) จากการพยากรณ์โดยแบบจำลองไฮบริด ARIMA-LSTM & LSTM-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 2 โดย ผลการพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.5 และ การพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.5 ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 2.94037 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.08781

2.9) จากการพยากรณ์โดยแบบจำลองไฮบริด ARIMA-LSTM & LSTM-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 2 โดย ผลการพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.4 และ การพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.6 ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 3.03101 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.09021

2.10) จากการพยากรณ์โดยแบบจำลองไฮบริด ARIMA-LSTM & LSTM-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 2 โดย ผลการพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.3 และ การพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.7 ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 3.15009 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.09333

2.11) จากการพยากรณ์โดยแบบจำลองไฮบริด ARIMA-LSTM & LSTM-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 2 โดย ผลการพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.2 และ การพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.8 ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 3.29453 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.09705

2.12) จากการพยากรณ์โดยแบบจำลองไฮบริด ARIMA-LSTM & LSTM-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 2 โดย ผลการพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.1 และ การพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.9 ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 3.46117 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.10136

3) การพยากรณ์จากแบบจำลองเดี่ยวไฮบริดด้วย ARIMA-LSTM & LSTM (Univariate)-ARIMA

3.1) จากการพยากรณ์โดยแบบจำลองไฮบริด LSTM-ARIMA ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 3.29844 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.09238





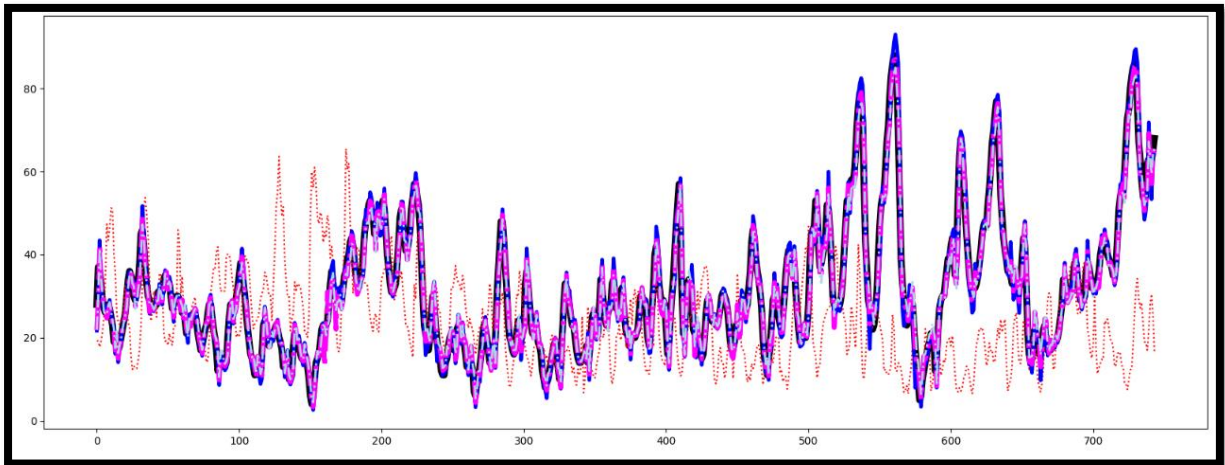
กำลังสองเฉลี่ย (RMSE) เท่ากับ 2.86573 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.08369

3.9) จากการพยากรณ์โดยแบบจำลองไฮบริด ARIMA-LSTM & LSTM-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 2 โดย ผลการพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.3 และ การพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.7 ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 2.94053 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.08508

3.10) จากการพยากรณ์โดยแบบจำลองไฮบริด ARIMA-LSTM & LSTM-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 2 โดย ผลการพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.2 และ การพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.8 ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 3.03908 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.08711

3.11) จากการพยากรณ์โดยแบบจำลองไฮบริด ARIMA-LSTM & LSTM-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 2 โดย ผลการพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.1 และ การพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.9 ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 3.15917 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.08952

พิจารณาผลการพยากรณ์ค่าฝุ่น PM 2.5 สถานีการเคหะชุมชนห้วยขวาง แขวงดินแดง เขตดินแดง กรุงเทพฯ (11T) พบว่า วิธีการพยากรณ์แบบจำลองไฮบริด ARIMA-LSTM & LSTM-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 2 โดย ผลการพยากรณ์จากแบบจำลอง ARIMA-LSTM(Univariate) มีค่าถ่วงน้ำหนักเท่ากับ 0.6 และ การพยากรณ์จากแบบจำลอง LSTM-ARIMA มีค่าถ่วงน้ำหนักเท่ากับ 0.4 ให้ค่าพยากรณ์ดีที่สุดเมื่อพิจารณาจากค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) แต่หากพิจารณาค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) วิธีการพยากรณ์แบบจำลอง ARIMA ให้ค่าพยากรณ์ดีที่สุดซึ่งจากผลการพยากรณ์ สามารถแสดงเป็นกราฟได้ดังภาพที่ 4.3



ภาพที่ 4.3 กราฟการพยากรณ์ฝุ่น PM2.5 สถานี 11T

จากภาพที่ 4.3 แสดงกราฟพยากรณ์ฝุ่น PM2.5 สถานี 11T โดยมีรายละเอียดดังนี้

กราฟเส้นสีดำ แทน ค่าฝุ่น PM2.5

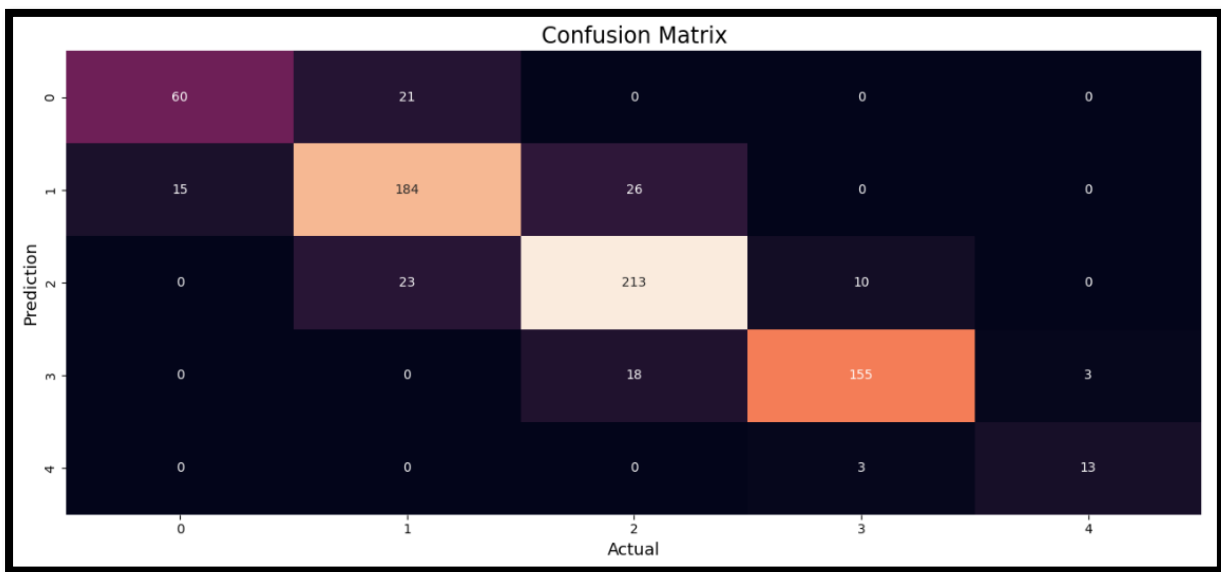
- กราฟเส้นประสีแดงแทนค่าฝุ่น PM2.5 ที่ถูกพยากรณ์โดยแบบจำลอง LSTM (Univariate) ซึ่งจะเห็นว่ามีความแตกต่างจากค่าฝุ่น PM2.5

- กราฟเส้นสีน้ำเงินแทนค่าฝุ่น PM2.5 ที่ถูกพยากรณ์โดยแบบจำลองไฮบริด LSTM (Univariate)-ARIMA ซึ่งผลการพยากรณ์มีแนวโน้มไปในทิศทางเดียวกับค่าฝุ่น PM2.5

- กราฟเส้นประสีชมพูแทนค่าฝุ่น PM2.5 ที่ถูกพยากรณ์โดยแบบจำลองไฮบริด ARIMA-LSTM & LSTM (Univariate)-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 2 โดย ผลการพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.6 และ การพยากรณ์จากแบบจำลอง LSTM (Univariate)-ARIMA มีค่าถ่วงน้ำหนักเท่ากับ 0.4 ซึ่งผลการพยากรณ์ปรับปรุงได้ดีขึ้น

- กราฟเส้นสีฟ้าแทนค่าฝุ่น PM2.5 ที่ถูกพยากรณ์โดยแบบจำลองไฮบริด ARIMA ซึ่งมีความใกล้เคียงกับกราฟเส้นสีชมพู

จากผลการพยากรณ์จากค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error:MAPE) จะได้แบบจำลองไฮบริด ARIMA-LSTM & LSTM (Univariate)-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 2 โดย การพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.6 และ การพยากรณ์จากแบบจำลอง LSTM (Univariate)-ARIMA มีค่าถ่วงน้ำหนักเท่ากับ 0.4 เมื่อนำไปพยากรณ์ตามคุณภาพอากาศ โดยกำหนดให้คุณภาพอากาศดีมากมีค่าเท่ากับ 0 คุณภาพอากาศดีมีค่าเท่ากับ 1 คุณภาพอากาศปานกลางมีค่าเท่ากับคุณภาพอากาศเริ่มมีผลต่อสุขภาพมีค่าเท่ากับ 3 และคุณภาพอากาศมีผลต่อสุขภาพมีค่าเท่ากับ 4 สามารถแสดงผลได้ดัง ภาพที่ 4.4



ภาพที่ 4.4 ผลลัพธ์การทำนายคุณภาพอากาศของสถานี 11T

จากภาพที่ 4.4 ผลลัพธ์การทำนาย PM2.5 ของสถานี 11T สามารถสรุปผลได้ดังนี้

1) คุณภาพอากาศมีค่าเท่ากับ 0 มีทั้งหมด 81 รายการ ผลการทำนายคุณภาพอากาศมีค่าเท่ากับ 0 มี 60 รายการ คิดเป็นทำนายถูก 74.07% โดยทำนายผิดเป็นการพยากรณ์ค่าสูงกว่าค่าจริง 25.03% โดยผลการทำนายคุณภาพอากาศมีค่าเท่ากับ 1 มี 21 รายการ

2) คุณภาพอากาศมีค่าเท่ากับ 1 มีทั้งหมด 225 รายการ ผลการทำนายคุณภาพอากาศมีค่าเท่ากับ 1 มี 184 รายการ คิดเป็นทำนายถูก 81.78% โดยทำนายผิดเป็นการพยากรณ์ค่าต่ำกว่าค่าจริง 6.67% โดยผลการทำนายคุณภาพอากาศมีค่าเท่ากับ 0 มี 15 รายการ ทำนายผิดเป็นการพยากรณ์ค่าสูงกว่าค่าจริง 11.55% โดยผลการทำนายคุณภาพอากาศมีค่าเท่ากับ 2 มี 26 รายการ

3) คุณภาพอากาศมีค่าเท่ากับ 2 มีทั้งหมด 246 รายการ ผลการทำนายคุณภาพอากาศมีค่าเท่ากับ 2 มี 213 รายการ คิดเป็นทำนายถูก 86.59% โดยทำนายผิดเป็นการพยากรณ์ค่าต่ำกว่าค่าจริง 9.35% โดยผลการทำนายคุณภาพอากาศมีค่าเท่ากับ 1 มี 23 รายการ ทำนายผิดเป็นการพยากรณ์ค่าสูงกว่าค่าจริง 4.06% โดยผลการทำนายคุณภาพอากาศมีค่าเท่ากับ 3 มี 10 รายการ

4) คุณภาพอากาศมีค่าเท่ากับ 3 มีทั้งหมด 176 รายการ ผลการทำนายคุณภาพอากาศมีค่าเท่ากับ 3 มี 155 รายการ คิดเป็นทำนายถูก 88.07% โดยทำนายผิดเป็นการพยากรณ์ค่าต่ำกว่าค่าจริง 10.23% โดยผลการทำนายคุณภาพอากาศมีค่าเท่ากับ 2 มี 18 รายการ ทำนายผิดเป็นการพยากรณ์ค่าสูงกว่าค่าจริง 1.70% โดยผลการทำนายคุณภาพอากาศมีค่าเท่ากับ 4 มี 3 รายการ

5) คุณภาพอากาศมีค่าเท่ากับ 4 มีทั้งหมด 16 รายการ ผลการทำนายคุณภาพอากาศมีค่า

เท่ากับ 4 มี 13 รายการ คิดเป็นทำนายถูก 81.25% โดยทำนายผิดรวมเป็นการพยากรณ์ค่าต่ำกว่าค่าจริง 18.75% โดยผลการทำนายคุณภาพอากาศมีค่าเท่ากับ 3 มี 3 รายการ

ผลการพยากรณ์ค่าฝุ่น PM 2.5 สถานีการเคหะชุมชนดินแดง ริมถนนดินแดง เขตดินแดง กรุงเทพฯ (54T) สามารถสรุปได้ดังนี้

**ตารางที่ 4.3** ผลการพยากรณ์เปรียบเทียบสถานี 54T

แบบจำลอง	ค่าความคลาดเคลื่อน			
	RMSE		MAPE	
1. ARIMA	5.93144		0.13543	
2. LSTM-Multivariate	7.79483		0.15834	
3. LSTM-Univariate	18.14426		0.36932	
แบบจำลองไฮบริด	LSTM-Multivariate		LSTM-Univariate	
	RMSE	MAPE	RMSE	MAPE
4. ARIMA-LSTM	6.1834	0.13983		
5. LSTM-ARIMA	7.44563	0.16333	6.50537	0.14361

ตารางที่ 4.3 (ต่อ)

แบบจำลองไฮบริด	ค่าความคลาดเคลื่อน			
	LSTM-Multivariate		LSTM-Univariate	
	RMSE	MAPE	RMSE	MAPE
6. ARIMA-LSTM & LSTM-ARIMA				
6.1 ปรับค่าถ่วงน้ำหนักวิธีที่ 1	6.53662	0.14301	6.14587	0.13686
6.2 ปรับค่าถ่วงน้ำหนักวิธีที่ 2				
ค่าถ่วงน้ำหนัก (ARIMA-LSTM)				
ค่าถ่วงน้ำหนัก (LSTM-ARIMA)				
0.9	6.08376	0.13718	6.08906	0.13729
0.8	6.03687	0.13554	6.0222	0.13552
0.7	6.04395	0.135	5.98374	0.13429
0.6	6.10481	0.13571	5.97424	0.13374
0.5	6.21788	0.13748	5.99384	0.13362
0.4	6.38038	0.14056	6.04224	0.13417
0.3	6.58866	0.14481	6.11877	0.13541
0.2	6.83853	0.1502	6.22238	0.13748
0.1	7.12562	0.15652	6.35176	0.14026

จากตารางที่ 4.3 สามารถสรุปผลการพยากรณ์ได้ดังนี้

1) การพยากรณ์จากแบบจำลองเดียว

1.1) จากการพยากรณ์โดยแบบจำลอง ARIMA ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 5.93144 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.13543

1.2) จากการพยากรณ์โดยแบบจำลอง LSTM (Multivariate) ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 7.79483 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.15834

1.3) จากการพยากรณ์โดยแบบจำลอง LSTM (Univariate) ได้ค่ารากที่สองของค่าความ

คลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 18.14426 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.36932

2) การพยากรณ์จากแบบจำลองเดี่ยวไฮบริดด้วย ARIMA-LSTM & LSTM (Multivariate)

ARIMA

2.1) จากการพยากรณ์โดยแบบจำลองไฮบริด ARIMA-LSTM ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 6.1834 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.13983

2.2) จากการพยากรณ์โดยแบบจำลองไฮบริด LSTM-ARIMA ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 7.44563 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.16333

2.3) จากการพยากรณ์โดยแบบจำลองไฮบริด ARIMA-LSTM & LSTM-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 1 ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 6.53662 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.14301

2.4) จากการพยากรณ์โดยแบบจำลองไฮบริด ARIMA-LSTM & LSTM-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 2 โดย ผลการพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.9 และ การพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.1 ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 6.08376 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.13718

2.5) จากการพยากรณ์โดยแบบจำลองไฮบริด ARIMA-LSTM & LSTM-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 2 โดย ผลการพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.8 และ การพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.2 ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 6.03687 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.13554

2.6) จากการพยากรณ์โดยแบบจำลองไฮบริด ARIMA-LSTM & LSTM-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 2 โดย ผลการพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.7 และ การพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.3 ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 6.04395 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.135

2.7) จากการพยากรณ์โดยแบบจำลองไฮบริด ARIMA-LSTM & LSTM-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 2 โดย ผลการพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.6 และ การพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.4 ได้ค่ารากที่สองของค่าความคลาดเคลื่อน

กำลังสองเฉลี่ย (RMSE) เท่ากับ 6.10481 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.13571

2.8) จากการพยากรณ์โดยแบบจำลองไฮบริด ARIMA-LSTM & LSTM-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 2 โดย ผลการพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.5 และ การพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.5 ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 6.21788 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.13748

2.9) จากการพยากรณ์โดยแบบจำลองไฮบริด ARIMA-LSTM & LSTM-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 2 โดย ผลการพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.4 และ การพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.6 ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 6.38038 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.14056

2.10) จากการพยากรณ์โดยแบบจำลองไฮบริด ARIMA-LSTM & LSTM-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 2 โดย ผลการพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.3 และ การพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.7 ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 6.58866 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.14481

2.11) จากการพยากรณ์โดยแบบจำลองไฮบริด ARIMA-LSTM & LSTM-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 2 โดย ผลการพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.2 และ การพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.8 ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 6.83853 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.1502

2.12) จากการพยากรณ์โดยแบบจำลองไฮบริด ARIMA-LSTM & LSTM-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 2 โดย ผลการพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.1 และ การพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.9 ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 7.12562 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.15652



3) การพยากรณ์จากแบบจำลองเดี่ยวไฮบริดด้วย ARIMA-LSTM & LSTM (Univariate)

ARIMA

3.1) จากการพยากรณ์โดยแบบจำลองไฮบริด LSTM-ARIMA ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 6.50537 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.14361

3.2) จากการพยากรณ์โดยแบบจำลองไฮบริด ARIMA-LSTM & LSTM-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 1 ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 6.14587 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.13686

3.3) จากการพยากรณ์โดยแบบจำลองไฮบริด ARIMA-LSTM & LSTM-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 2 โดย ผลการพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.9 และ การพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.1 ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 6.08906 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.13729

3.4) จากการพยากรณ์โดยแบบจำลองไฮบริด ARIMA-LSTM & LSTM-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 2 โดย ผลการพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.8 และ การพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.2 ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 6.0222 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.13552

3.5) จากการพยากรณ์โดยแบบจำลองไฮบริด ARIMA-LSTM & LSTM-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 2 โดย ผลการพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.7 และ การพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.3 ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 5.98374 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.13429

3.6) จากการพยากรณ์โดยแบบจำลองไฮบริด ARIMA-LSTM & LSTM-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 2 โดย ผลการพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.6 และ การพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.4 ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 5.97424 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.13374

3.7) จากการพยากรณ์โดยแบบจำลองไฮบริด ARIMA-LSTM & LSTM-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 2 โดย ผลการพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.5 และ การพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.5 ได้ค่ารากที่สองของค่าความคลาดเคลื่อน

กำลังสองเฉลี่ย (RMSE) เท่ากับ 5.99384 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.13362

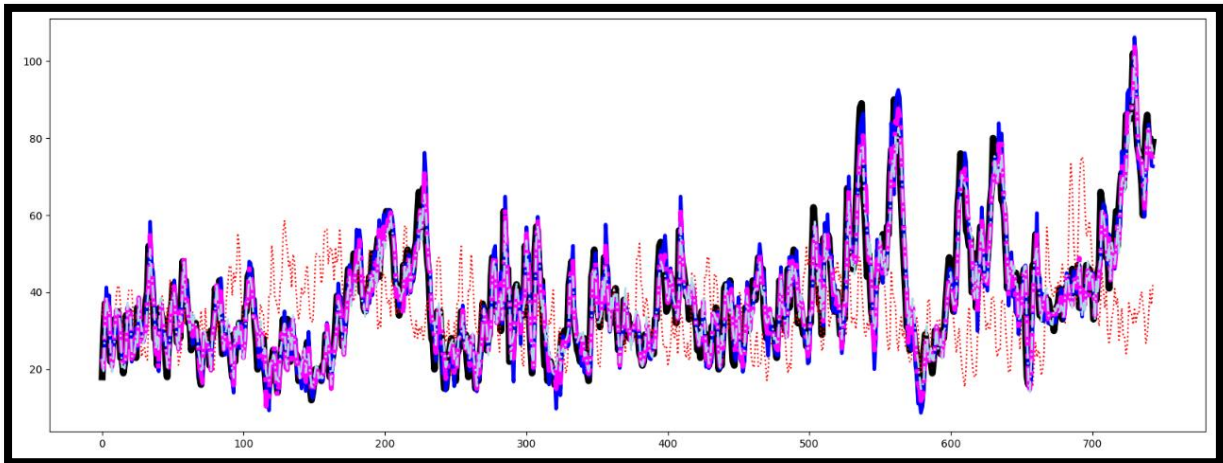
3.8) จากการพยากรณ์โดยแบบจำลองไฮบริด ARIMA-LSTM & LSTM-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 2 โดย ผลการพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.4 และ การพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.6 ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 6.04224 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.13417

3.9) จากการพยากรณ์โดยแบบจำลองไฮบริด ARIMA-LSTM & LSTM-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 2 โดย ผลการพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.3 และ การพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.7 ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 6.11877 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.13541

3.10) จากการพยากรณ์โดยแบบจำลองไฮบริด ARIMA-LSTM & LSTM-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 2 โดย ผลการพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.2 และ การพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.8 ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 6.22238 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.13748

3.11) จากการพยากรณ์โดยแบบจำลองไฮบริด ARIMA-LSTM & LSTM-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 2 โดย ผลการพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.1 และ การพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.9 ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 6.35176 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.14026

พิจารณาผลการพยากรณ์ค่าฝุ่น PM 2.5 สถานีการเคหะชุมชนดินแดง ริมถนนดินแดง เขตดินแดง กรุงเทพฯ (54T) พบว่า วิธีการพยากรณ์แบบจำลองไฮบริด ARIMA-LSTM & LSTM-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 2 โดย ผลการพยากรณ์จากแบบจำลอง ARIMA-LSTM (Univariate) มีค่าถ่วงน้ำหนักเท่ากับ 0.5 และ การพยากรณ์จากแบบจำลอง LSTM-ARIMA มีค่าถ่วงน้ำหนักเท่ากับ 0.5 ให้ค่าพยากรณ์ดีที่สุดเมื่อพิจารณาจากค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) แต่หากพิจารณาค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) วิธีการพยากรณ์แบบจำลอง ARIMA ให้ค่าพยากรณ์ดีที่สุดซึ่งจากผลการพยากรณ์ สามารถแสดงเป็นกราฟได้ดังภาพที่ 4.5

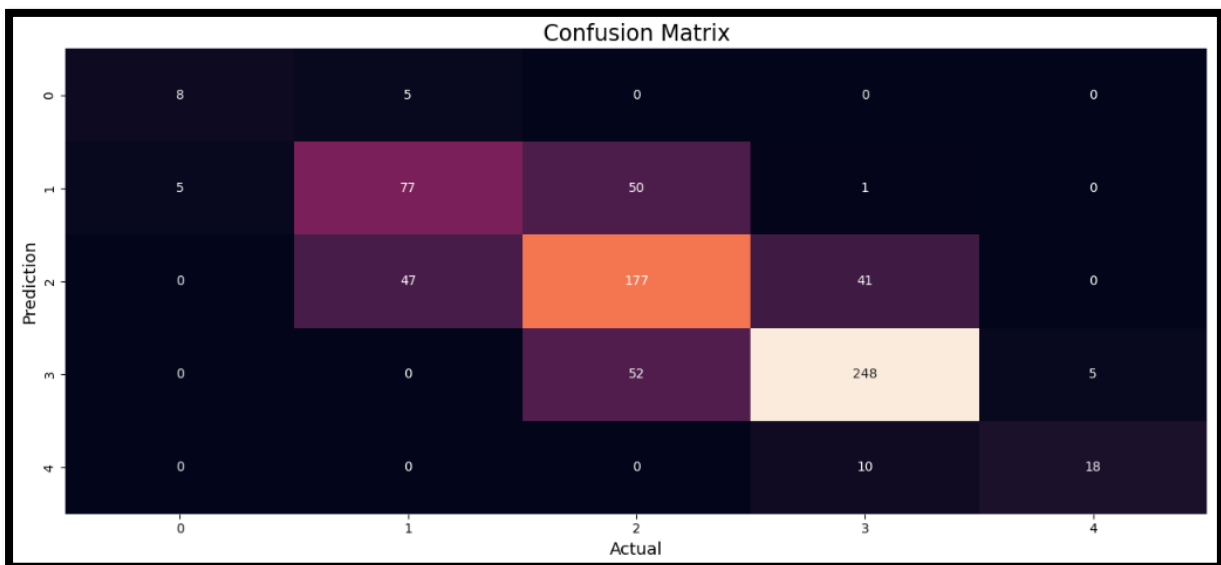


ภาพที่ 4.5 กราฟการพยากรณ์ฝุ่น PM2.5 สถานี 54T

จากภาพที่ 4.5 แสดงกราฟพยากรณ์ฝุ่น PM2.5 สถานี 54T โดยมีรายละเอียดดังนี้

- กราฟเส้นสีดำ แทน ค่าฝุ่น PM2.5
- กราฟเส้นประสีแดงแทนค่าฝุ่น PM2.5 ที่ถูกพยากรณ์โดยแบบจำลอง LSTM (Univariate) ซึ่งจะเห็นว่ามีความแตกต่างจากค่าฝุ่น PM2.5
- กราฟเส้นสีน้ำเงินแทนค่าฝุ่น PM2.5 ที่ถูกพยากรณ์โดยแบบจำลองไฮบริด LSTM (Univariate)-ARIMA ซึ่งผลการพยากรณ์มีแนวโน้มไปในทิศทางเดียวกับค่าฝุ่น PM2.5
- กราฟเส้นประสีชมพูแทนค่าฝุ่น PM2.5 ที่ถูกพยากรณ์โดยแบบจำลองไฮบริด ARIMA-LSTM & LSTM (Univariate)-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 2 โดย ผลการพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.6 และ การพยากรณ์จากแบบจำลอง LSTM (Univariate)-ARIMA มีค่าถ่วงน้ำหนักเท่ากับ 0.4 ซึ่งผลการพยากรณ์ปรับปรุงได้ดีขึ้น
- กราฟเส้นสีฟ้าแทนค่าฝุ่น PM2.5 ที่ถูกพยากรณ์โดยแบบจำลองไฮบริด ARIMA ซึ่งมีความใกล้เคียงกับกราฟเส้นสีชมพู

จากผลการพยากรณ์จากค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) จะได้แบบจำลองไฮบริด ARIMA-LSTM & LSTM (Univariate)-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 2 โดยการพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.5 และ การพยากรณ์จากแบบจำลอง LSTM (Univariate)-ARIMA มีค่าถ่วงน้ำหนักเท่ากับ 0.5 เมื่อนำไปพยากรณ์ตามคุณภาพอากาศ โดยกำหนดให้ คุณภาพอากาศดีมีค่าเท่ากับ 0 คุณภาพอากาศดีมีค่าเท่ากับ 1 คุณภาพอากาศปานกลางมีค่าเท่ากับ 2 คุณภาพอากาศเริ่มมีผลต่อสุขภาพมีค่าเท่ากับ 3 และคุณภาพอากาศมีผลต่อสุขภาพมีค่าเท่ากับ 4 สามารถแสดงผลได้ดัง ภาพที่ 4.6



ภาพที่ 4.6 ผลลัพธ์การทำนายคุณภาพอากาศของสถานี 54T

จากภาพที่ 4.6 ผลลัพธ์การทำนาย PM2.5 ของสถานี 54T สามารถสรุปผลได้ดังนี้

- 1) คุณภาพอากาศมีค่าเท่ากับ 0 มีทั้งหมด 13 รายการ ผลการทำนายคุณภาพอากาศมีค่าเท่ากับ 0 มี 8 รายการ คิดเป็นทำนายถูก 61.54% โดยทำนายผิดเป็นการพยากรณ์ค่าสูงกว่าค่าจริง 38.46% โดยผลการทำนายคุณภาพอากาศมีค่าเท่ากับ 1 มี 5 รายการ
- 2) คุณภาพอากาศมีค่าเท่ากับ 1 มีทั้งหมด 133 รายการ ผลการทำนายคุณภาพอากาศมีค่าเท่ากับ 1 มี 77 รายการ คิดเป็นทำนายถูก 57.89% โดยทำนายผิดเป็นการพยากรณ์ค่าต่ำกว่าค่าจริง 3.76% โดยผลการทำนายคุณภาพอากาศมีค่าเท่ากับ 0 มี 5 รายการ ทำนายผิดเป็นการพยากรณ์ค่าสูงกว่าค่าจริง 38.35% โดยผลการทำนายคุณภาพอากาศมีค่าเท่ากับ 2 มี 50 รายการ คิดเป็นทำนายผิด 37.59% ทำนายคุณภาพอากาศมีค่าเท่ากับ 3 มี 1 รายการ คิดเป็นทำนายผิด 0.76%
- 3) คุณภาพอากาศมีค่าเท่ากับ 2 มีทั้งหมด 265 รายการ ผลการทำนายคุณภาพอากาศมีค่าเท่ากับ 2 มี 177 รายการ คิดเป็นทำนายถูก 66.79% โดยทำนายผิดเป็นการพยากรณ์ค่าต่ำกว่าค่าจริง 17.74% โดยผลการทำนายคุณภาพอากาศมีค่าเท่ากับ 1 มี 47 รายการ ทำนายผิดเป็นการพยากรณ์ค่าสูงกว่าค่าจริง 15.47% โดยผลการทำนายคุณภาพอากาศมีค่าเท่ากับ 3 มี 41 รายการ
- 4) คุณภาพอากาศมีค่าเท่ากับ 3 มีทั้งหมด 305 รายการ ผลการทำนายคุณภาพอากาศมีค่าเท่ากับ 3 มี 248 รายการ คิดเป็นทำนายถูก 81.31% โดยทำนายผิดเป็นการพยากรณ์ค่าต่ำกว่าค่าจริง 17.05% โดยผลการทำนายคุณภาพอากาศมีค่าเท่ากับ 2 มี 52 รายการ ทำนายผิดเป็นการพยากรณ์ค่าสูงกว่าค่าจริง 1.64% โดยผลการทำนายคุณภาพอากาศมีค่าเท่ากับ 4 มี 5 รายการ

5) คุณภาพอากาศมีค่าเท่ากับ 4 มีทั้งหมด 28 รายการ ผลการทำนายคุณภาพอากาศมีค่าเท่ากับ 4 มี 18 รายการ คิดเป็นทำนายถูก 64.29% โดยทำนายผิดรวมเป็นการพยากรณ์ค่าต่ำกว่าค่าจริง 35.71% โดยผลการทำนายคุณภาพอากาศมีค่าเท่ากับ 3 มี 10 รายการ

ผลการพยากรณ์ค่าฝุ่น PM 2.5 สถานีกรมประชาสัมพันธ์ แขวงพญาไท เขตพญาไท กรุงเทพฯ (59T) สามารถสรุปได้ดังนี้

**ตารางที่ 4.4** ผลการพยากรณ์สถานี 59T

แบบจำลอง	ค่าความคลาดเคลื่อน	
	RMSE	MAPE
1. ARIMA	6.15301	0.26127
2. LSTM-Multivariate	8.40077	0.33413
3. LSTM-Univariate	17.47768	0.61475

ตารางที่ 4.4 (ต่อ)

แบบจำลองไฮบริด	ค่าความคลาดเคลื่อน				
	LSTM-Multivariate		LSTM-Univariate		
	RMSE	MAPE	RMSE	MAPE	
4. ARIMA-LSTM	6.50396	0.27409			
5. LSTM-ARIMA	7.70044	0.3207	7.23823	0.28992	
6. ARIMA-LSTM & LSTM-ARIMA					
6.1 ปรับค่าถ่วงน้ำหนักวิธีที่ 1	6.57654	0.28121	6.2363	0.25806	
6.2 ปรับค่าถ่วงน้ำหนักวิธีที่ 2					
ค่าถ่วงน้ำหนัก (ARIMA-LSTM)	ค่าถ่วงน้ำหนัก (LSTM-ARIMA)				
0.9	0.1	6.44267	0.27276	6.34943	0.26657
0.8	0.2	6.42407	0.27312	6.24457	0.26133
0.7	0.3	6.4485	0.27526	6.1919	0.25834
0.6	0.4	6.5155	0.27867	6.19275	0.25752
0.5	0.5	6.62377	0.28315	6.24711	0.25837
0.4	0.6	6.77132	0.28868	6.35361	0.26143
0.3	0.7	6.95567	0.29525	6.50967	0.26591
0.2	0.8	7.17397	0.30264	6.71185	0.27222
0.1	0.9	7.42323	0.31109	6.95612	0.27991

จากตารางที่ 4.4 สามารถสรุปผลการพยากรณ์ได้ดังนี้

1) การพยากรณ์จากแบบจำลองเดี่ยว

1.1) จากการพยากรณ์โดยแบบจำลอง ARIMA ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 6.15301 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.26127

1.2) จากการพยากรณ์โดยแบบจำลอง LSTM (Multivariate) ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 8.40077 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.33413

1.3) จากการพยากรณ์โดยแบบจำลอง LSTM (Univariate) ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 17.47768 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.61475

2) การพยากรณ์จากแบบจำลองเดี่ยวไฮบริดด้วย ARIMA-LSTM & LSTM (Multivariate) ARIMA

2.1) จากการพยากรณ์โดยแบบจำลองไฮบริด ARIMA-LSTM ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 6.50396 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.27409

2.2) จากการพยากรณ์โดยแบบจำลองไฮบริด LSTM-ARIMA ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 7.70044 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.3207

2.3) จากการพยากรณ์โดยแบบจำลองไฮบริด ARIMA-LSTM & LSTM-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 1 ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 6.57654 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.28121

2.4) จากการพยากรณ์โดยแบบจำลองไฮบริด ARIMA-LSTM & LSTM-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 2 โดย ผลการพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.9 และ การพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.1 ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 6.44267 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.27276

2.5) จากการพยากรณ์โดยแบบจำลองไฮบริด ARIMA-LSTM & LSTM-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 2 โดย ผลการพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.8 และ การพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.2 ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 6.42407 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.27312

2.6) จากการพยากรณ์โดยแบบจำลองไฮบริด ARIMA-LSTM & LSTM-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 2 โดย ผลการพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.7 และ การพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.3 ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 6.4485 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.27526

2.7) จากการพยากรณ์โดยแบบจำลองไฮบริด ARIMA-LSTM & LSTM-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 2 โดย ผลการพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.6 และ การ

พยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.4 ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 6.5155 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.27867

2.8) จากการพยากรณ์โดยแบบจำลองไฮบริด ARIMA-LSTM & LSTM-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 2 โดย ผลการพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.5 และ การพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.5 ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 6.62377 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.28315

2.9) จากการพยากรณ์โดยแบบจำลองไฮบริด ARIMA-LSTM & LSTM-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 2 โดย ผลการพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.4 และ การพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.6 ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 6.77132 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.28868

2.10) จากการพยากรณ์โดยแบบจำลองไฮบริด ARIMA-LSTM & LSTM-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 2 โดย ผลการพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.3 และ การพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.7 ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 6.95567 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.29525

2.11) จากการพยากรณ์โดยแบบจำลองไฮบริด ARIMA-LSTM & LSTM-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 2 โดย ผลการพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.2 และ การพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.8 ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 7.17397 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.30264

2.12) จากการพยากรณ์โดยแบบจำลองไฮบริด ARIMA-LSTM & LSTM-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 2 โดย ผลการพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.1 และ การพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.9 ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 7.42323 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.31109

3) การพยากรณ์จากแบบจำลองเดี่ยวไฮบริดด้วย ARIMA-LSTM & LSTM (Univariate)

ARIMA





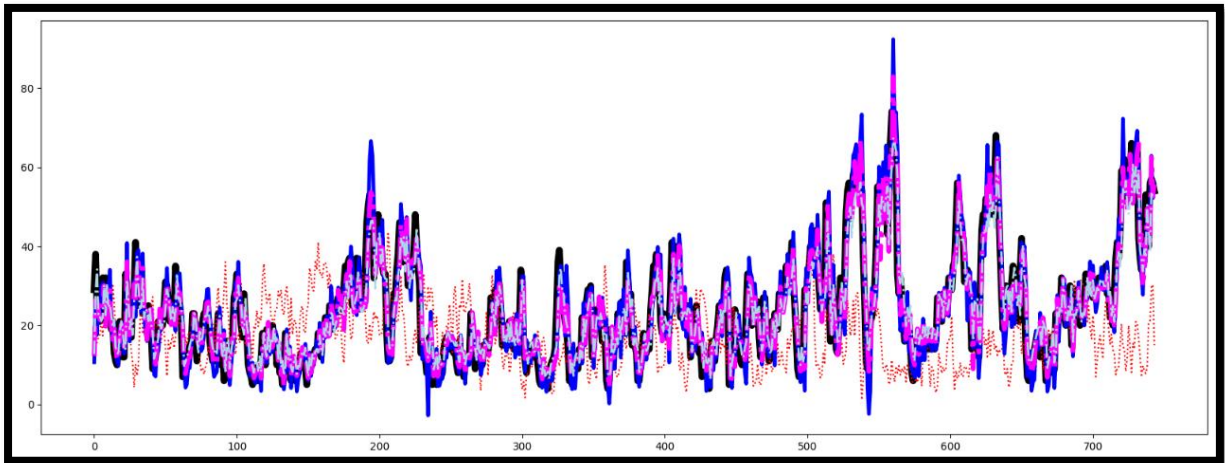
3.8) จากการพยากรณ์โดยแบบจำลองไฮบริด ARIMA-LSTM & LSTM-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 2 โดย ผลการพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.4 และ การพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.6 ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 6.35361 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.26143

3.9) จากการพยากรณ์โดยแบบจำลองไฮบริด ARIMA-LSTM & LSTM-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 2 โดย ผลการพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.3 และ การพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.7 ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 6.50967 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.26591

3.10) จากการพยากรณ์โดยแบบจำลองไฮบริด ARIMA-LSTM & LSTM-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 2 โดย ผลการพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.2 และ การพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.8 ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 6.71185 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.27222

3.11) จากการพยากรณ์โดยแบบจำลองไฮบริด ARIMA-LSTM & LSTM-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 2 โดย ผลการพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.1 และ การพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.9 ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 6.95612 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เท่ากับ 0.27991

พิจารณาผลการพยากรณ์ค่าฝุ่น PM 2.5 สถานีกรมประชาสัมพันธ์ แขวงพญาไท เขตพญาไท กรุงเทพฯ (59T) พบว่า วิธีการพยากรณ์แบบจำลองไฮบริด ARIMA-LSTM & LSTM-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 2 โดย ผลการพยากรณ์จากแบบจำลอง ARIMA-LSTM (Univariate) มีค่าถ่วงน้ำหนักเท่ากับ 0.6 และ การพยากรณ์จากแบบจำลอง LSTM-ARIMA มีค่าถ่วงน้ำหนักเท่ากับ 0.4 ให้ค่าพยากรณ์ดีที่สุดเมื่อพิจารณาจากค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) แต่หากพิจารณาค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) วิธีการพยากรณ์แบบจำลอง ARIMA ให้ค่าพยากรณ์ดีที่สุดซึ่งจากผลการพยากรณ์ สามารถแสดงเป็นกราฟได้ดังภาพที่ 4.7

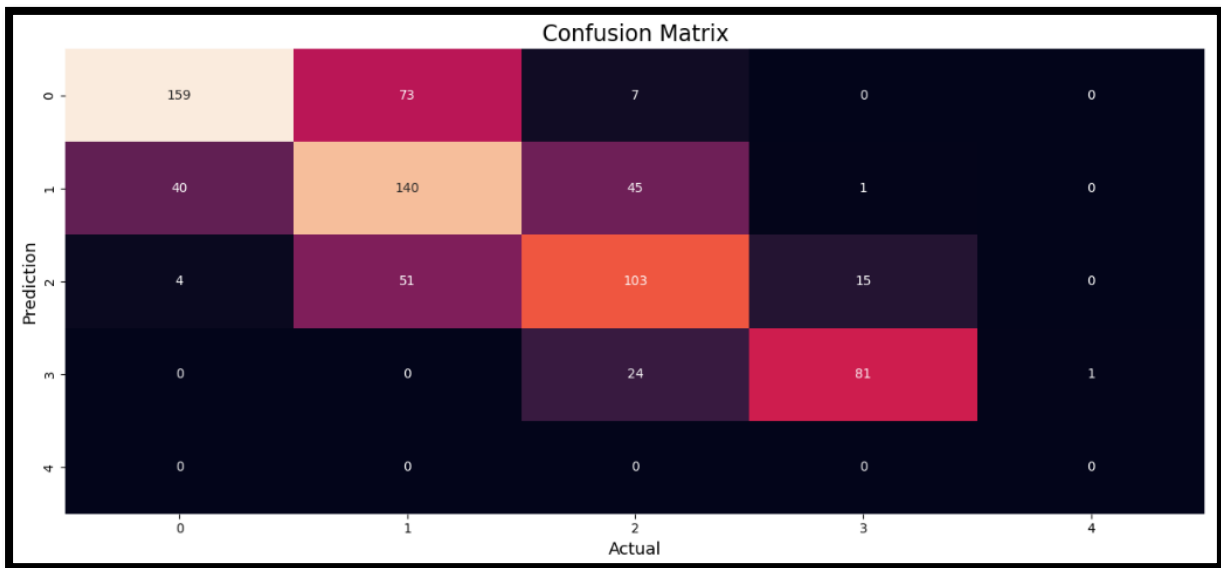


ภาพที่ 4.7 กราฟการพยากรณ์ฝุ่น PM2.5 สถานี 59T

จากภาพที่ 4.7 แสดงกราฟพยากรณ์ฝุ่น PM2.5 สถานี 59T โดยมีรายละเอียดดังนี้

- กราฟเส้นสีดำ แทน ค่าฝุ่น PM2.5
- กราฟเส้นประสีแดงแทนค่าฝุ่น PM2.5 ที่ถูกพยากรณ์โดยแบบจำลอง LSTM (Univariate) ซึ่งจะเห็นว่ามีความแตกต่างจากค่าฝุ่น PM2.5
- กราฟเส้นสีน้ำเงินแทนค่าฝุ่น PM2.5 ที่ถูกพยากรณ์โดยแบบจำลองไฮบริด LSTM (Univariate)-ARIMA ซึ่งผลการพยากรณ์มีแนวโน้มไปในทิศทางเดียวกับค่าฝุ่น PM2.5
- กราฟเส้นประสีชมพูแทนค่าฝุ่น PM2.5 ที่ถูกพยากรณ์โดยแบบจำลองไฮบริด ARIMA-LSTM & LSTM (Univariate)-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 2 โดย ผลการพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.6 และ การพยากรณ์จากแบบจำลอง LSTM (Univariate)-ARIMA มีค่าถ่วงน้ำหนักเท่ากับ 0.4 ซึ่งผลการพยากรณ์ปรับปรุงได้ดีขึ้น
- กราฟเส้นสีฟ้าแทนค่าฝุ่น PM2.5 ที่ถูกพยากรณ์โดยแบบจำลองไฮบริด ARIMA ซึ่งมีความใกล้เคียงกับกราฟเส้นสีชมพู

จากผลการพยากรณ์จากค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error:MAPE) จะได้แบบจำลองไฮบริด ARIMA-LSTM & LSTM (Univariate)-ARIMA ปรับค่าถ่วงน้ำหนักวิธีที่ 2 โดย การพยากรณ์จากแบบจำลอง ARIMA-LSTM มีค่าถ่วงน้ำหนักเท่ากับ 0.6 และ การพยากรณ์จากแบบจำลอง LSTM (Univariate)-ARIMA มีค่าถ่วงน้ำหนักเท่ากับ 0.4 เมื่อนำไปพยากรณ์ตามคุณภาพอากาศ โดยกำหนดให้คุณภาพอากาศดีมากมีค่าเท่ากับ 0 คุณภาพอากาศดีมีค่าเท่ากับ 1 คุณภาพอากาศปานกลางมีค่าเท่ากับ 2 คุณภาพอากาศเริ่มมีผลต่อสุขภาพมีค่าเท่ากับ 3 และคุณภาพอากาศมีผลต่อสุขภาพมีค่าเท่ากับ 4 สามารถแสดงผลได้ดัง ภาพที่ 4.8



ภาพที่ 4.8 ผลลัพธ์การทำนายคุณภาพอากาศของสถานี 59T

จากภาพที่ 4.8 ผลลัพธ์การทำนาย PM2.5 ของสถานี 59T สามารถสรุปผลได้ดังนี้

1) คุณภาพอากาศมีค่าเท่ากับ 0 มีทั้งหมด 239 รายการ ผลการทำนายคุณภาพอากาศมีค่าเท่ากับ 0 มี 159 รายการ คิดเป็นทำนายถูก 66.53% โดยทำนายผิดเป็นการพยากรณ์ค่าสูงกว่าค่าจริง 33.47% โดยผลการทำนายคุณภาพอากาศมีค่าเท่ากับ 1 มี 73 รายการ คิดเป็นทำนายผิด 30.54% ทำนายคุณภาพอากาศมีค่าเท่ากับ 2 มี 7 รายการ คิดเป็นทำนายผิด 2.93%

2) คุณภาพอากาศมีค่าเท่ากับ 1 มีทั้งหมด 226 รายการ ผลการทำนายคุณภาพอากาศมีค่าเท่ากับ 1 มี 140 รายการ คิดเป็นทำนายถูก 61.95% โดยทำนายผิดเป็นการพยากรณ์ค่าต่ำกว่าค่าจริง 17.70% โดยผลการทำนายคุณภาพอากาศมีค่าเท่ากับ 0 มี 40 รายการ ทำนายผิดเป็นการพยากรณ์ค่าสูงกว่าค่าจริง 20.35% โดยผลการทำนายคุณภาพอากาศมีค่าเท่ากับ 2 มี 45 รายการ คิดเป็นทำนายผิด 19.91% ทำนายคุณภาพอากาศมีค่าเท่ากับ 3 มี 1 รายการ คิดเป็นทำนายผิด 0.44%

3) คุณภาพอากาศมีค่าเท่ากับ 2 มีทั้งหมด 173 รายการ ผลการทำนายคุณภาพอากาศมีค่าเท่ากับ 2 มี 103 รายการ คิดเป็นทำนายถูก 59.54% โดยทำนายผิดเป็นการพยากรณ์ค่าต่ำกว่าค่าจริง 31.79% โดยผลการทำนายคุณภาพอากาศมีค่าเท่ากับ 0 มี 4 รายการ คิดเป็นทำนายผิด 2.31% ทำนายคุณภาพอากาศมีค่าเท่ากับ 1 มี 51 รายการ คิดเป็นทำนายผิด 29.48% ทำนายผิดเป็นการพยากรณ์ค่าสูงกว่าค่าจริง 8.67% โดยผลการทำนายคุณภาพอากาศมีค่าเท่ากับ 3 มี 15 รายการ

4) คุณภาพอากาศมีค่าเท่ากับ 3 มีทั้งหมด 106 รายการ ผลการทำนายคุณภาพอากาศมีค่าเท่ากับ 3 มี 81 รายการ คิดเป็นทำนายถูก 76.42% โดยทำนายผิดเป็นการพยากรณ์ค่าต่ำกว่าค่าจริง 22.64%

โดยผลการทำนายคุณภาพอากาศมีค่าเท่ากับ 2 มี 24 รายการ ทำนายผิดเป็นการพยากรณ์ค่าสูงกว่าค่าจริง 0.94% โดยผลการทำนายคุณภาพอากาศมีค่าเท่ากับ 4 มี 1 รายการ

#### 4.2 ผลการศึกษาความสัมพันธ์ระหว่างสถานีวัดปริมาณฝุ่น PM2.5

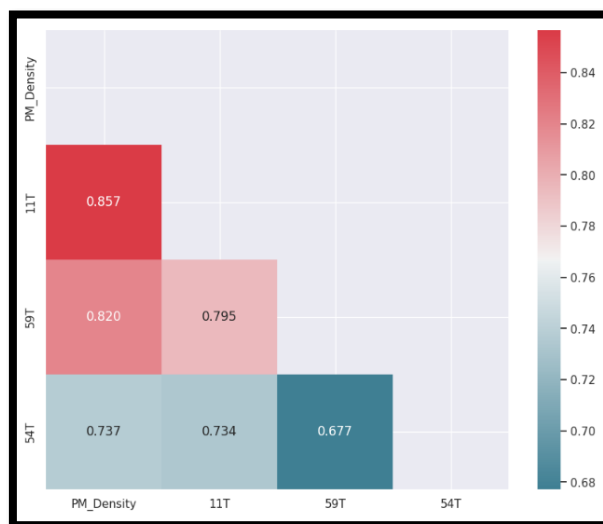
การคำนวณระยะห่างระหว่างสถานีโดยใช้กลุ่มตัวอย่างจำนวน 4 สถานีของ สามารถคำนวณได้ดังต่อไปนี้

#### ตารางที่ 4.5 ระยะห่างระหว่างสถานีตัวอย่างในกรุงเทพ

(กิโลเมตร)

สถานี	02T	11T	54T	59T
02T		10.00	9.45	7.98
11T			2.79	3.22
54T				1.47

จากข้อมูลระหว่าง 1 พฤศจิกายน 2565 เวลา 0:00 น. ถึง 31 มกราคม 2566 เวลา 23:00 น. ของสถานีกลุ่มตัวอย่าง สามารถหาค่าสหสัมพันธ์ได้ดังรูป



ภาพที่ 4.9 ค่าสหสัมพันธ์ระหว่างสถานี

หมายเหตุ: PM\_Density คือ สถานี 02T

จากข้อมูลสัมพันธและข้อมูลระยะทางของสถานีตัวอย่าง สามารถสรุปได้ดังนี้

- 1) ค่าสหสัมพันธ์ของปริมาณฝุ่น PM 2.5 ของสถานีมหาวิทยาลัยราชภัฏบ้านสมเด็จเจ้าพระยา แขวงหิรัญรูจี เขตธนบุรี (02T) ระหว่างสถานี สามารถสรุปได้ดังนี้
  - 1.1) มีค่าสหสัมพันธ์ในระดับสูงมากกับสถานีตัวอย่างอื่น ประกอบด้วย
    - สถานีการเคหะชุมชนห้วยขวาง แขวงดินแดง เขตดินแดง (11T) ซึ่งมีระยะห่างระหว่างสถานี 10.00 กิโลเมตร
    - สถานีกรมประชาสัมพันธ์ แขวงพญาไท เขตพญาไท (59T) ซึ่งมีระยะห่างระหว่างสถานี 7.98 กิโลเมตร
  - 1.2) มีค่าสหสัมพันธ์ในระดับสูงกับสถานีตัวอย่างอื่น ประกอบด้วย
    - สถานีการเคหะชุมชนดินแดง ริมถนนดินแดง เขตดินแดง (54T) ซึ่งมีระยะห่างระหว่างสถานี 9.45 กิโลเมตร
- 2) ค่าสหสัมพันธ์ของปริมาณฝุ่น PM 2.5 ของสถานีการเคหะชุมชนห้วยขวาง แขวงดินแดง เขตดินแดง (11T) ระหว่างสถานี สามารถสรุปได้ดังนี้
  - 2.1) มีค่าสหสัมพันธ์ระดับสูงในทิศทางบวกกับสถานีตัวอย่างอื่น ประกอบด้วย
    - สถานีการเคหะชุมชนดินแดง ริมถนนดินแดง เขตดินแดง (54T) ซึ่งมีระยะห่างระหว่างสถานี 2.79 กิโลเมตร
    - สถานีกรมประชาสัมพันธ์ แขวงพญาไท เขตพญาไท (59T) ซึ่งมีระยะห่างระหว่างสถานี 3.22 กิโลเมตร
- 3) ค่าสหสัมพันธ์ของปริมาณฝุ่น PM 2.5 ของสถานีกรมประชาสัมพันธ์ แขวงพญาไท เขตพญาไท (59T) ระหว่างสถานี สามารถสรุปได้ดังนี้
  - 3.1) มีค่าสหสัมพันธ์ระดับปานกลางในทิศทางบวกกับสถานีตัวอย่างอื่น ประกอบด้วย
    - สถานีการเคหะชุมชนดินแดง ริมถนนดินแดง เขตดินแดง (54T) ซึ่งมีระยะห่างระหว่างสถานี 1.47 กิโลเมตร

## บทที่ 5

### บทสรุปและข้อเสนอแนะ

#### 5.1 อภิปรายผลการศึกษา

1. แบบจำลองไฮบริด ARIMA-LSTM เป็นแนวคิดที่ถูกพัฒนาโดย Peter Zhang ตั้งแต่ ค.ศ. 2003 ซึ่งเป็นแนวคิดที่ผ่านมากกว่า 20 ปี และเป็นแบบจำลองที่ถูกใช้ในการพยากรณ์ฝุ่น PM2.5 หลายแบบจำลอง แนวคิดพื้นฐานคิดการพยากรณ์ค่าฝุ่น PM2.5 ส่วนแรก ด้วยแบบจำลอง ARIMA เพื่อทำการพยากรณ์เชิงเส้น ข้อมูลที่มีโครงสร้างเชิงเส้น (Linear Structure Data) และพยากรณ์ส่วนที่เหลือ (Residual) ด้วยโครงข่ายประสาทเทียม (ANN: Artificial Neural Networks) จนมีการพัฒนาจากแบบจำลอง ANN เป็นแบบจำลอง (LSTM: Long Short-Term Memory) จากแนวคิดดังกล่าวเมื่อนำค่าพยากรณ์ที่ได้มารวมกัน จะควรได้ค่าที่มีความคลาดเคลื่อน (Error) ลดลงอันแสดงถึงประสิทธิภาพในการพยากรณ์ที่ดีขึ้น แต่จากข้อมูลปริมาณฝุ่น PM2.5 ของกรมมลพิษ ในกรุงเทพมหานครจำนวน 4 สถานีซึ่งประกอบไปด้วย สถานี 02T คือ มหาวิทยาลัยราชภัฏบ้านสมเด็จเจ้าพระยา แขวงหิรัญรูจี เขตธนบุรี สถานี 11T คือ การเคหะชุมชนห้วยขวาง แขวงดินแดง เขตดินแดง สถานี 54T คือ การเคหะชุมชนดินแดง ริมถนนดินแดง เขตดินแดง สถานี 59T คือ กรมประชาสัมพันธ์ แขวงพญาไท เขตพญาไท มาพยากรณ์ด้วยแบบจำลองไฮบริด ARIMA-LSTM แล้วพบว่าผลการพยากรณ์มีความคลาดเคลื่อน (Error) สูงกว่าการพยากรณ์ด้วยแบบจำลองไฮบริด ARIMA เพียงอย่างเดียว จึงเกิดความคิดริเริ่มในการพยากรณ์ด้วยแบบจำลองไฮบริด LSTM-ARIMA ซึ่งการวิจัยครั้งนี้ทำการ Lag แบบจำลอง LSTM จำนวน 12 ชั่วโมง ซึ่งผลของการทดลองพยากรณ์ด้วยแบบจำลอง LSTM แบบหลายตัวแปร (Multivariate) แล้วจึงพยากรณ์ด้วยแบบจำลอง ARIMA ผลการพยากรณ์ แบบจำลองไฮบริด LSTM (Multivariate)-ARIMA มีความคลาดเคลื่อน (Error) สูงกว่าขึ้น เมื่อเทียบกับการพยากรณ์ด้วยแบบจำลอง LSTM แบบหลายตัวแปร (Multivariate) อย่างเดียว แต่เมื่อทำการทดลองพยากรณ์ด้วยแบบจำลอง LSTM แบบตัวเดียว (Univariate) แล้วจึงพยากรณ์ด้วยแบบจำลอง ARIMA พบว่าแบบจำลองไฮบริด LSTM (Univariate)-ARIMA มีความคลาดเคลื่อน (Error) ต่ำกว่า เมื่อเทียบกับการพยากรณ์ด้วยแบบจำลอง LSTM (Univariate) อย่างเดียว แต่อย่างไรก็ดี แบบจำลองไฮบริด ARIMA-LSTM แบบจำลองไฮบริด LSTM (Multivariate)-ARIMA และแบบจำลองไฮบริด LSTM (Univariate)-ARIMA ยังคงไม่สามารถให้ค่าความคลาดเคลื่อน (Error) ที่ต่ำกว่าแบบจำลอง ARIMA แต่เพียงอย่างเดียว

แต่เมื่อทำการศึกษาเพิ่มเติมพบแนวคิดแบบจำลองไฮบริดการเรียนรู้เชิงลึก (Hybrid Deep Learning Models) ทำให้เกิดแนวคิดการคำนวณร่วมระหว่างแบบจำลองไฮบริด ARIMA-LSTM ร่วมกับแบบจำลองไฮบริด LSTM-ARIMA ร่วมกับการปรับค่าถ่วงน้ำหนัก ซึ่งพบว่า แบบจำลองไฮบริด ARIMA-LSTM ร่วมกับแบบจำลองไฮบริด LSTM (Univariate)-ARIMA ที่มีการปรับค่าถ่วงน้ำหนัก สามารถให้ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) ที่ต่ำกว่าแบบจำลอง ARIAM แต่เพียงอย่างเดียว แต่มี

สถานี O2T ที่มีค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Squared Error: RMSE) ต่ำกว่าแบบจำลอง ARIAM แต่เพียงอย่างเดียว ทั้งนี้ค่าความคลาดเคลื่อนที่ได้ลดลงเพียงเล็กน้อยในบางสถานี เกือบเทียบเท่าการพยากรณ์ด้วยแบบจำลอง ARIMA แต่เพียงอย่างเดียว ซึ่งต้องแลกมากับเวลาในการพยากรณ์ที่เพิ่มขึ้น ซึ่งอาจต้องพิจารณากับความคุ้มค่าของประสิทธิภาพที่เพิ่มขึ้น

2. ผลการศึกษาความสัมพันธ์กับระยะทางระหว่างสถานีพบว่า ค่าปริมาณฝุ่น PM2.5 ระหว่างสถานี สามารถส่งอิทธิพลต่อกันได้ แต่ยังไม่สามารถระบุชัดได้ว่าเกิดจากปัจจัยใด เนื่องจากระยะห่างระหว่างสถานีที่มีระยะห่างที่ไกลกว่า สามารถมีค่าสหสัมพันธ์ที่สูงกว่าได้ ในการวิจัยครั้งนี้ตัวแปรอิสระที่ใช้ประกอบด้วย ความเร็วลม ทิศทางลม และอุณหภูมิ ซึ่งทิศทางของกระแสลมอาจเป็นตัวแปรสำคัญที่ส่งผลต่ออิทธิพลระหว่างสถานี ซึ่งต้องทำการศึกษาต่อไป

## 5.2 ข้อเสนอแนะ

เมื่อพิจารณาแบบจำลอง LSTM-ARIMA สามารถพัฒนาเพื่อให้มีประสิทธิภาพเพิ่มขึ้นได้โดยอาศัยแนวทางพื้นฐานดังต่อไปนี้

- 1) ส่วนข้อมูลเรียนรู้และข้อมูลทดสอบ สามารถเพิ่มสัดส่วนของข้อมูลการเรียนรู้เพิ่มขึ้น
- 2) ส่วนแบบจำลอง ARIMA ในการทดลองครั้งนี้ไม่ได้พิจารณาเรื่องการเคลื่อนไหวตามฤดูกาล รวมถึง การคำนวณค่า  $p$ ,  $d$ ,  $q$  ในการวิจัยครั้งนี้ใช้คำสั่ง auto arima ซึ่งกำหนดค่าสูงสุดของค่า  $p$  และ  $q$  เท่ากับ 5 ซึ่งอาจเลือกการกำหนดค่า  $p$ ,  $d$ ,  $q$  โดยไม่ใช่คำสั่ง auto arima และหรือกำหนดค่าสูงสุดเป็นค่าอื่น
- 3) ส่วนแบบจำลอง LSTM สามารถกำหนดค่า Neuron ค่า Batch size หรือค่า Epoch ที่เหมาะสม ซึ่งค่าเหล่านี้ เมื่อคำนวณด้วยแบบจำลองไฮบริด ARIMA-LSTM และแบบจำลอง LSTM-ARIM รวมทั้งทดลองใช้ BiLSTM หรือเลือกใช้ Algorithm อื่น



## บรรณานุกรม

### บรรณานุกรม

- [1] จุฬาลงกรณ์มหาวิทยาลัย (2562). เรียนรู้ อยู่กับฝุ่น PM2.5. ค้นหาเมื่อ 2 มกราคม 2566 สืบค้นจาก <https://www.chula.ac.th/wp-content/uploads/2019/10/Chula-PM25.pdf>
- [2] Committee on the Medical Effects of Air Pollutants. (2022). Advice on health evidence relevant to setting PM2.5 targets – update. Retrieved January 15, 2023, from [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/1060744/COMEAP\\_WHO\\_AQG\\_-\\_Defra\\_PM2.5\\_targets\\_advice\\_\\_2\\_.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1060744/COMEAP_WHO_AQG_-_Defra_PM2.5_targets_advice__2_.pdf)
- [3] Peter Zhang, G. (2003). Time series forecasting using a hybrid ARIMA and neural network model, Neurocomputing, vol. 50, pp. 159 – 175.
- [4] Fathi, O. (2019). Time series forecasting using a hybrid ARIMA and LSTM model. France: Velvet Consulting.
- [5] Hochreiter, S. & Schmidhuber, J. (1997). Long-Short term memory, Neurocomputing, vol. 9(8), pp. 1735 – 1780.
- [6] Xiao, J., Wang, Q., Cui, J. & Yu, J. (2022). Multi-feature PM2.5 Prediction with ARIMA-LSTM. International Conference on Machine Learning, Cloud Computing and Intelligent Mining.
- [7] กรมควบคุมมลพิษ, กระทรวงทรัพยากรธรรมชาติและสิ่งแวดล้อม (2564). คู่มือปฏิบัติการในการป้องกันและแก้ไขปัญหาฝุ่นละออง PM2.5. สืบค้นจาก <https://www.pcd.go.th/publication/13479>
- [8] ไทยพีบีเอส (2566). ดีเดย์ 1 มิ.ย.นี้ ปรับเกณฑ์ PM2.5 ใหม่ 37.5 มคก.ต่อลบม. ค้นหาเมื่อ 1 กรกฎาคม 2566 สืบค้นจาก <https://www.thaipbs.or.th/news/content/327288>
- [9] George E. P. Box, Gwilym M. Jenkins., Gregory C.Reinsel & Greta M., Ljung. (2016). Time Series Analysis: Forecasting and Control. John Wiley & Sons, Inc.
- [10] Shaun, W. (2017). Time Series Analysis: Forecasting and Control, 5th edition. Journal of Quality Technology. 49. pp. 418-419.
- [11] DiBattista, J. (2022). Choosing the Best ML Time Series Model for Your Data. Retrieved January 15, 2023, from <https://towardsdatascience.com/choosing-the-best-ml-time-series-model-for-your-data-664a7062f418>

- [12] Bescond, P-L. Cyclical features encoding, it's about time!. Retrieved January 15, 2023, from <https://towardsdatascience.com/cyclical-features-encoding-its-about-time-ce23581845ca>
- [13] Karr, D. (2022). Calculate or Query Great Circle Distance Between Points of Latitude and Longitude Using the Haversine Formula (PHP, JavaScript, Java, Python, MySQL, MSSQL Examples). Retrieved January 15, 2023, from <https://martech.zone/calculate-great-circle-distance/>.
- [14] อรไท ชั่วเจริญ. (2558). สถิติเบื้องต้น Introduction to Statistics. สำนักพิมพ์ มหาวิทยาลัยรามคำแหง.
- [15] Yuwei, C. & Kaizhi, W. (2019). Prediction of Satellite Time Series Data Based on Long Short Term Memory Autoregressive Integrated Moving Average Model (LSTM-ARIMA). IEEE International Conference on Signal and Image Processing.
- [16] Yuan, P., Mei, YF., Zhong, YL. & Xia, Y. A Hybrid Deep Learning Model for Predicting PM2.5. Liang Fang College of Computer Science and Electronic Engineering, Hunan University, Changsha, China.
- [17] กระทรวงสาธารณสุข (2562). คู่มือการดำเนินงานด้านการแพทย์และสาธารณสุข กรณี ฝุ่นละอองขนาดไม่เกิน 2.5 ไมครอน (PM2.5) ปี 2563 ค้นหาค้นหาเมื่อ 14 มีนาคม 2566 สืบค้นจาก <https://multimedia.anamai.moph.go.th/associates/ebook/>
- [18] Cach N. Dang, Maria N. Moreno-García, & Fernando De la Prieta. (2021). Hybrid Deep Learning Models for Sentiment Analysis, Complexity. Vol. 2021.

## ภาคผนวก

## ภาคผนวก ก

หนังสือขอข้อมูลเพื่อประกอบการทำวิทยานิพนธ์



ที่ วทว.ม.0408(3)/0601

1 มิถุนายน 2566

เรื่อง ขอข้อมูลเพื่อประกอบการทำวิทยานิพนธ์

เรียน กรมควบคุมมลพิษ, กองการจัดการคุณภาพอากาศและเสียง

ด้วย นายธจักร พลเสนา เลขทะเบียนนักศึกษา 6451-6202-0017 นักศึกษาระดับบัณฑิตศึกษา สาขาวิชาวิศวกรรมข้อมูลขนาดใหญ่ วิทยาลัยนวัตกรรมด้านเทคโนโลยีและวิศวกรรมศาสตร์ อยู่ระหว่างทำการค้นคว้าหาข้อมูลประกอบการทำวิทยานิพนธ์ เรื่อง “การศึกษาประสิทธิภาพการพยากรณ์ฝุ่น PM2.5 ในเขตกรุงเทพมหานคร โดยแบบจำลองไฮบริดร่วมกับค่าดวงน้ำหนัก”

วิทยาลัยนวัตกรรมด้านเทคโนโลยีและวิศวกรรมศาสตร์ มหาวิทยาลัยธุรกิจบัณฑิต จึงเรียนมาเพื่อขอความอนุเคราะห์ข้อมูลเกี่ยวกับ PM2.5 ความเร็วลม ทิศทางลม อุณหภูมิ ความชื้นสัมพัทธ์ ความกดอากาศ ปริมาณน้ำฝน ปี 2565 ทุกสถานีในจังหวัดกรุงเทพมหานคร เพื่อใช้ในการทดสอบงานวิจัย และเข้าศึกษาค้นคว้าข้อมูล โดยศึกษาและการสร้างแบบจำลองเพื่อพยากรณ์ฝุ่นละอองที่มีขนาดไม่เกิน 2.5 ไมครอน (PM2.5) รายชั่วโมง ระหว่างวันที่ 1 มกราคม 2565 – 30 เมษายน 2566 เวลา 0.00 – 23.00 น.

ทั้งนี้ ข้อมูลที่ได้มาจะได้ใช้สำหรับการทำวิทยานิพนธ์เท่านั้น

จึงเรียนมาเพื่อโปรดพิจารณาและหวังเป็นอย่างยิ่งว่าจะได้รับความอนุเคราะห์จากท่านด้วยดี ขอขอบคุณ มา ณ โอกาสนี้

ขอแสดงความนับถือ

(อาจารย์ ดร.ชัยพร เชมะภาคะพันธ์)

คณบดีวิทยาลัยนวัตกรรมด้านเทคโนโลยีและวิศวกรรมศาสตร์  
ปฏิบัติการแทนอธิการบดี

นางสาวกุลธิดา รอดบุญ ผู้ประสานงาน  
บัณฑิตศึกษา สาขาวิชาวิศวกรรมข้อมูลขนาดใหญ่ วิทยาลัยนวัตกรรมด้านเทคโนโลยีและวิศวกรรมศาสตร์  
โทร. 02-9547300 ต่อ 681  
Email: kulthida.rod@dpu.ac.th

**ภาคผนวก ข**  
**ผลงานตีพิมพ์ ชั้นที่ 1**



## The 7<sup>th</sup> RMUTP Conference on Engineering and Technology 2023

การประชุมวิชาการวิศวกรรมศาสตร์และเทคโนโลยี มทร.พระนคร ครั้งที่ 7

“การวิจัยเชิงนวัตกรรมเพื่อการพัฒนา  
เทคโนโลยีพลังงานและสิ่งแวดล้อมที่ยั่งยืน”

INNOVATION RESEARCH FOR  
DEVELOPMENT IN SUSTAINABLE ENERGY  
AND ENVIRONMENTAL TECHNOLOGY

19 พฤษภาคม 2566

คณะวิศวกรรมศาสตร์

มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร





รหัสบทความ	ชื่อเรื่อง	หน้า
<b>คอมพิวเตอร์และเทคโนโลยีสารสนเทศ</b>		
CP-501	การแสดงผลข้อมูลด้วยภาพสำหรับวัดไทยในจังหวัดนนทบุรี วิวรรณ จันทร์ทรัพย์ นริศรา นาคเมธี และ สยาม สางคุณเสน	208
CP-502	ระบบสืบค้นผ้าขึ้นสำหรับศูนย์วัฒนธรรมศึกษาจังหวัดน่าน สุพรรณศรี คำวงศ์ เกวรินทร์ จันทร์ดำ และสิทธิเดช วิชาศิริศิริกุล	214
CP-503	ระบบจัดการต้นทุนสำหรับเช่ารถทัวร์ ชนพันธ์ นวรวงศ์นันต์ และ นันทิกา ปริญญาพร	222
CP-504	"Dischool" บอทติสคอร์ดสำหรับช่วยจัดการการเรียนการสอนออนไลน์ ศุวิล ชมชัยยา บุรีสกร อยู่สุข กิตติศักดิ์ สังฆะรักษ์ และ ศิริพงษ์ เอี่ยมอาษา	227
CP-505	การพัฒนาเว็บไซต์เผยแพร่ทวิตส์เสี่ยงที่ต้องการการอุปถัมภ์ ศิวกร ศิวกร สิริภัสสร สุขชนาน และ ณสิทธิ์ เวฬุาเสิน	233
CP-506	การศึกษาเปรียบเทียบความเร็วของอินเทอร์เน็ตที่ให้บริการโดยโครงข่าย 5G : กรณีแหล่งท่องเที่ยวสำคัญในเขตอำเภอเมือง จังหวัดเพชรบุรี จุฑาทันต์ อินทรีย์นันท์ พงษ์พิสิฐ วุฒิสุโขทัย และ เทอดพงษ์ แดงสี	238
CP-507	การศึกษาชุดคำสั่งสำหรับค้นหาแฮชแท็กแรกในทวีตเตอร์ สิริพงษ์ วรประชา และ พงษ์พิสิฐ วุฒิสุโขทัย	243
CP-508	การพัฒนาระบบสารสนเทศการให้บริการรถโดยสารสาธารณะในจังหวัดภูเก็ต ด้วยเทคโนโลยีเสมือนจริง ภัทรวรรณ ศิริภักษ์ ณ อานาวิน บุญเทียม และ อุไรรัตน์ มากจันทร์	249
CP-510	การพยากรณ์การตัดสินใจในใช้ e-wallet ด้วยเหมืองข้อมูล รวมนัด เจริญรักษา และ สุวรรณีย์ อิศกุลชัย	253
CP-511	การพัฒนาแอปพลิเคชันแนะนำแหล่งท่องเที่ยววิสาหกิจชุมชนใน 3 จังหวัด ฝั่งอันดามัน ศักดิ์ชัย ไตรศรี วชิษฐ์พล ศิวรัตน์ อรรถพล ปิธโลย์ และ อุไรรัตน์ มากจันทร์	258
CP-512	สหกิจศึกษาบนแพลตฟอร์มออนไลน์ ณัฐพร นิมิตต์ และ ณัฐพร นิมิตต์	262
CP-513	การศึกษาประสิทธิภาพการพยากรณ์ฝุ่น PM2.5 ในเขตกรุงเทพมหานคร โดยแบบจำลองไฮบริดร่วมกับค่าอ่งน้ำหนัก อัคร พงเสนา และ ดวงใจ จิตคงชื่น	268

บทความวิจัย

การประชุมวิชาการวิศวกรรมศาสตร์และเทคโนโลยี มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี ครั้งที่ 7  
 Proceedings of the 7<sup>th</sup> RMUTP Conference on Engineering and Technology

การศึกษาประสิทธิภาพการพยากรณ์ฝุ่น PM2.5 ในเขตกรุงเทพมหานคร  
 โดยแบบจำลองไฮบริดร่วมกับค่าถ่วงน้ำหนัก  
 Efficiency study of PM2.5 forecasting in Bangkok  
 by the Hybrid Model with Weighted values.

จักร พงษ์สนา<sup>1</sup> และ ดวงใจ ฉัตรงษ์สิน, ผู้ช่วยศาสตราจารย์ ดร.<sup>2</sup>  
 มหาวิทยาลัยธุรกิจบัณฑิต คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี  
 110/1-4 ประชาชื่น หลักสี่ กรุงเทพฯ 10210  
 645162020017@dpu.ac.th | 082 4889961@95dpu.ac.th

บทคัดย่อ

กรุงเทพมหานคร ซึ่งเป็นเมืองหลวงของประเทศไทยนั้น มักประสบปัญหาฝุ่นละออง PM2.5 ในช่วงฤดูหนาว โดยเดือนมกราคม พ.ศ. 2566 ที่ผ่านมามีแนวโน้มของค่าเฉลี่ยของฝุ่นละออง PM2.5 เกินกว่ามาตรฐานที่กรมควบคุมมลพิษกำหนด ฝุ่นละออง PM2.5 เป็นปัญหาที่ควรให้ความสำคัญ อันเนื่องมาจากฝุ่นละออง PM2.5 เป็นตัวกลางพาหะอันตรายเข้าสู่ปอด ผู้วิจัยจึงเกิดแนวคิดในการพัฒนาประสิทธิภาพการพยากรณ์ฝุ่น PM2.5 ด้วยแบบจำลองไฮบริดตามแนวคิดของรุ่นต้นแบบที่ชื่อค่าถ่วงน้ำหนักในปี พ.ศ. 2003<sup>[1]</sup> ซึ่งนอกจากใช้แบบจำลอง Arima-LSTM แล้ว ผู้วิจัยทำการเพิ่มแบบจำลอง LSTM-Arima เพื่อทำการถ่วงน้ำหนักและเพิ่มประสิทธิภาพในการพยากรณ์

คำสำคัญ : ฝุ่นละออง PM2.5, การพยากรณ์, แบบจำลองไฮบริด

Abstract

Bangkok which is the capital of Thailand found PM2.5 problems during the winter. In January 2023, there were news reports about areas where PM2.5 exceeded the standard. PM2.5 is a problem that should be addressed. Due to PM2.5 dust, it is a medium that carries harmful substances into the lungs. Therefore, the researcher has the idea of improving the PM2.5 forecasting efficiency with a hybrid model based on the original concept that was born in 2003<sup>[1]</sup> by using the Arima-LSTM model, the researcher added the LSTM-Arima model for weighting, and increase the efficiency of forecasting.

Keywords: PM2.5; Prediction; Hybrid Model

1. บทนำ

ความเสียหายมีสุขภาพที่เข้มแข็งเป็นความท้าทายพื้นฐานที่ตรงกันของมนุษยชาติ โดยประเทศไทยเริ่มให้ความสำคัญกับปัญหาฝุ่นละออง PM2.5 ในปี พ.ศ. 2562 ซึ่งฝุ่นละออง PM2.5 หรือมีชื่อทางการเรียกว่า “Particulate matter with diameter of less than 2.5 micron” เป็นฝุ่นละอองที่มีขนาดไม่เกิน 2.5 ไมครอน ทำหน้าที่เป็นตัวกลางพาหะอันตรายต่าง ๆ เข้าสู่ปอด เช่น สารก่อมะเร็ง สารโลหะหนัก เป็นต้น<sup>[1]</sup>

ความเสียหายในการพยากรณ์ฝุ่นละออง PM2.5 ซึ่งถือเป็นแนวทางการหนึ่งในการป้องกันปัญหาสุขภาพที่อาจเกิดขึ้นจากฝุ่นละออง PM2.5

แบบจำลองที่มักถูกใช้ในการพยากรณ์ฝุ่นละออง PM2.5 คือแบบจำลองไฮบริด (Hybrid) ตามแนวคิดของ Peter Zhang ในปี พ.ศ. 2013 โดยให้แบบจำลอง (Auto Regressive Integrated Moving Average: ARIMA) พยากรณ์ข้อมูลที่มีโครงสร้างเชิงเส้น หลังจากนั้นจึงพยากรณ์ค่าส่วนที่เหลือ (Residual) ด้วยแบบจำลองโครงข่ายประสาทเทียม ซึ่งในขณะนั้น คือ แบบจำลองโครงข่ายประสาทเทียม (Artificial Neural Networks: ANN)

ในปัจจุบันแบบจำลองไฮบริด ได้รับการพัฒนาจากแบบจำลอง ANN เป็นแบบจำลองหน่วยความจำระยะสั้นระยะยาว (Long Short-Term Memory: LSTM) ซึ่งนอกจากแบบจำลองไฮบริดแบบ ARIMA-LSTM แล้ว ผู้วิจัยเกิดแนวคิดในการสร้างแบบจำลองไฮบริดแบบ LSTM-ARIMA เพื่อการนำสองแบบจำลองมาหาค่าถ่วงน้ำหนัก (Weight) เพื่อแก้ปัญหาค่าความคลาดเคลื่อนที่เพิ่มขึ้น หลังจากการพยากรณ์ส่วนที่เหลือ โดยมีแนวทางการถ่วงน้ำหนักสองแบบ<sup>[10]</sup> เพื่อสร้างความสมมาตรของแบบจำลองและเพิ่มประสิทธิภาพในการพยากรณ์

2. วัตถุประสงค์

2.1 วัตถุประสงค์ที่เกี่ยวข้อง

C. Yawei & W. Kaihui<sup>[2]</sup> (2019) ทำการวิจัยเพื่อพยากรณ์การเปลี่ยนแปลงของโครงข่ายของดาวเทียมด้วยชุดข้อมูลอนุกรมเวลาตัวแปรวัดและดัชนีคุณภาพอากาศจำนวนมากด้วยเทคนิคแบบจำลอง LSTM-ARIMA โดยเป็นการคำนวณจากทั้ง 2 แบบจำลองแล้วหาค่าถ่วงน้ำหนัก (Weight) ที่เหมาะสมจากการคำนวณค่าของทั้งสองแบบจำลอง ผลการทดลองพบว่า การคำนวณด้วยแบบจำลอง LSTM ด้วยค่าถ่วงน้ำหนัก (Weight) เท่ากับ 0.5 และแบบจำลอง ARIMA ด้วยค่าถ่วงน้ำหนัก (Weight) เท่ากับ 0.5 ได้ค่ารากที่สองของค่าเฉลี่ยความผิดพลาดกำลังสอง (Root Mean Squared Error: RMSE) ที่น้อยที่สุด จึงเป็นค่าถ่วงน้ำหนักที่เหมาะสมของดาวพยากรณ์

**บทความวิจัย**

การประชุมวิชาการวิศวกรรมศาสตร์และเทคโนโลยี มห.บูรพาฯ ครั้งที่ 7  
*Proceedings of the 7<sup>th</sup> RMUTP Conference on Engineering and Technology*

P. Kanokrot & K. Chompeesool<sup>1</sup> (2022) พัฒนาแบบจำลอง การพยากรณ์ค่ามลพิษทางอากาศโดยใช้เทคนิคผสมผสานระหว่าง ARIMA และ LSTM (ARIMA-LSTM) แบบจำลอง ARIMA ใช้ทำนาย ผลกระทบของพารามิเตอร์ตัวอนุกรมเวลาและความสัมพันธ์เชิงเส้น (Linear) ในขณะที่ใช้ LSTM เพื่อทำนายผลกระทบจากพารามิเตอร์ที่มีความสัมพันธ์แบบไม่เชิงเส้น (Non-Linear) โดยใช้อัตรา 84 เดือน ระหว่างเดือนมกราคม พ.ศ. 2557 ถึงเดือนธันวาคม พ.ศ. 2562 นำมาใช้ในงานวิจัยนี้ ประเมินโดยใช้ค่าเฉลี่ยสัมบูรณ์ ได้แก่ ปรากฏการณ์ของ ค่าเฉลี่ยของความผิดพลาดกำลังสอง (Root Mean Squared Error: RMSE) เท่ากับ 10.275 บาท ค่าเฉลี่ยของผลต่างสัมบูรณ์ (Mean Absolute Error: MAE) เท่ากับ 8.512 บาท และค่าเฉลี่ยของเปอร์เซ็นต์ความผิดพลาด สัมบูรณ์ (Mean Absolute Percentage Error : MAPE) เท่ากับ 13.618% ตามลำดับ

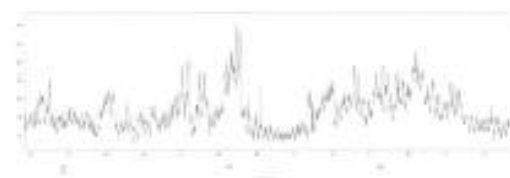
J. Xiao, Q. Wang, J. Cai, & J. Yu<sup>11</sup> (2022) พยากรณ์ข้อมูลค่าฝุ่น PM2.5 โดยใช้ข้อมูลรายชั่วโมงจากสถานีที่ต่างหากหนึ่ง ผลการทดลองพบว่า เราที่ผสมของค่าเฉลี่ยของความผิดพลาดกำลังสอง (Root Mean Squared Error: RMSE) และค่าเฉลี่ยของค่าเปอร์เซ็นต์ความผิดพลาด เปอร์เซ็นต์กำลังสอง (Root Mean Squared Percent Error: RMSPE) ของแบบจำลองโสมบิวค ได้ค่าเฉลี่ยของความผิดพลาดที่ต่ำกว่าที่สุด จึงสรุปได้ว่าแบบจำลองโสมบิวคสามารถให้ค่าพยากรณ์ที่ถูกต้องแม่นยำกว่าเมื่อเปรียบเทียบกับผลการพยากรณ์ด้วยแบบจำลอง LSTM อย่างเดียว

**2.2 เครื่องมือที่ใช้ในการวิจัย**

ในส่วนของการเตรียมตัวข้อมูลที่ใช้ก่อนทำการรันโค้ด 1 เครื่องในการ โปรแกรมด้วยภาษาไพทอน (Python) โดยลักษณะของภาษา สามารถเขียนโค้ดได้ง่าย และรองรับงานวิจัยทางด้านวิทยาศาสตร์ข้อมูล (Data Science) พัฒนา และทดสอบบน Google Colaboratory ซึ่งเป็น บริการคลาวด์ของ Google (2020) โดยทั่วไปโปรแกรม Jupyter ของระบบ บน Cloud 910 Google

**2.3 ข้อมูลที่ใช้ในการวิจัย**

ใช้ข้อมูลรายชั่วโมง จากสถานีวัดสิ่งแวดล้อมที่ 25 ขึ้นวาทน พ.ศ. 2556 เวลา 0:00 น. ถึง 24 มีนาคม พ.ศ. 2566 เวลา 23:00 น. รวม 2,160 รายการ จากข้อมูลจากกองจัดการคุณภาพอากาศและเสียง ถึงกัก คุมคุณภาพอากาศ<sup>14</sup>



รูปที่ 1 ข้อมูล PM2.5 รายชั่วโมง จากสถานีวัด

ตารางที่ 1 ตัวแปร ที่ใช้ในการพยากรณ์

ลำดับ	ตัวแปร	สิ่งตีพิมพ์
1	WS (m/s)	ความเร็วลม (Wind Speed)
2	WD (Deg)	ทิศทางลม (Wind Direction)
3	Temp (°C)	อุณหภูมิของอากาศ (Temperature)
4	RH (%)	ความชื้นสัมพัทธ์ (Relative Humidity)
5	BP (mBar)	ความกดอากาศ (Barometric Pressure)
6	PM2.5 (ug/m3)	ฝุ่นขนาดเล็กที่มีขนาดไม่เกิน 2.5 ไมครอน

**2.4 แบบจำลองที่ใช้ในการวิจัย**

**2.4.1 แนวคิดแบบจำลองออริมา**

แบบจำลอง ARIMA เป็นแบบจำลองของชุดข้อมูล ประกอบด้วย 84 ค่าประกอบของ Autoregression, Integrated I(0) Moving average เป็นแบบจำลองของค่าที่ใช้ในการพยากรณ์ข้อมูลอนุกรมเวลาที่มีช่วง  $p, d, q$  ซึ่งสามารถอธิบายแบบจำลอง ARIMA ได้ดังนี้<sup>15</sup>

Auto Regression, AR(p) เป็นส่วนหนึ่งของแบบจำลอง ARIMA ตามแนวคิดที่จะใช้ความล่าช้าของตัวเอง (ค่าในอดีต) เป็นตัวทำนาย โดยที่พารามิเตอร์  $p$  คือ ระดับของจำนวนความล่าช้าที่เกิดขึ้น

Integrated, I(d) ช่วงเวลาที่ข้อมูลอนุกรมเวลาอยู่กับที่เพื่อลด การขึ้นต่อกันของเวลาจะลบ  $d$  โดยที่  $d$  มีค่า 0 คือ ระดับความแตกต่างซึ่งหมายถึงจำนวนครั้งที่ข้อมูลมีความแตกต่างกัน ถ้าอนุกรม เวลาเป็น stationary) ระดับความแตกต่างของ  $d$  มีค่า 0

Moving Average, MA(q) ใช้ข้อมูลผลตกที่เหลือน (residual) ของเวลาที่ผ่านมายี่ใช้ในการคาดการณ์ในปัจจุบันและอนาคต โดยที่ พารามิเตอร์  $q$  คือ จำนวนของข้อผิดพลาดการคาดการณ์ความล่าช้าที่ใช้ ในการคำนวณค่าปัจจุบัน

**2.4.2 แนวคิดแบบจำลองหน่วยความจำระยะสั้นระยะยาว**

หน่วยความจำระยะสั้นระยะยาว เป็นโครงข่ายประสาทเทียม แบบวนกลับ (Recurrent Neural Networks: RNN) โดย LSTM มีสูตรการ คำนวณดังนี้<sup>16</sup>

$$\begin{aligned}
 i_t &= \sigma(\mathbf{W}_i \cdot [h_{t-1}, \mathbf{x}_t] + b_i) & (1) \\
 f_t &= \sigma(\mathbf{W}_f \cdot [h_{t-1}, \mathbf{x}_t] + b_f) & (2) \\
 c_t &= \tanh(\mathbf{W}_c \cdot [h_{t-1}, \mathbf{x}_t] + b_c) & (3) \\
 o_t &= \sigma(\mathbf{W}_o \cdot [h_{t-1}, \mathbf{x}_t] + b_o) & (4) \\
 h_t &= f_t \cdot c_t + (1 - f_t) \cdot c_{t-1} & (5) \\
 y_t &= h_t = \tanh(c_t) o_t & (6)
 \end{aligned}$$

**บทความวิจัย**

การประยุกต์ใช้การวิเคราะห์อนุกรมเวลาและเทคนิค โมเดล แนวคิด ครั้งที่ 7  
Proceedings of the 7<sup>th</sup> RMUTP Conference on Engineering and Technology

โดยที่

- $X_t$  แทน จำนวนของชิ้นของเวลาที่ปัจจุบัน
- $f_t$  แทน ผลลัพธ์ที่คาดการณ์
- $I_t$  แทน ผลลัพธ์ของอนุกรมเวลา
- $O_t$  แทน ผลลัพธ์ของอนุกรมเวลาชุด
- $C_t$  แทน สถานะของผลผลิตที่ส่งไปยังเซลล์ถัดไป
- $h_t$  แทน สถานะที่ซ่อนอยู่ที่ไม่ได้สังเกตได้
- $y_t$  แทน ผลลัพธ์ของเวลาที่ปัจจุบัน
- $w_f, w_i, w_c, w_o, b_f, b_i, b_c$  และ  $b_o$  เป็นพารามิเตอร์

สมการที่ 1

**2.4.3 แนวคิดแบบจำลองไฮบริด**

แบบจำลองที่ประกอบด้วยวิธีการแบบผสมผสานที่มีความสามารถในการสร้างแบบจำลองที่มีลักษณะไม่เชิงเส้นของความเป็นกลุ่มที่สัมพันธ์กันในการใช้งานจริง โดยที่รวมโมเดลที่แตกต่างกันซึ่งสามารถเขียนเป็นสมการได้ดังนี้<sup>[1]</sup>

$$Y_t = L_t + N_t \tag{7}$$

โดยที่

- $L_t$  หมายถึง ส่วนข้อมูลที่เป็นโครงสร้างเชิงเส้น
- $N_t$  หมายถึง ส่วนข้อมูลที่ไม่เป็นโครงสร้างเชิงเส้น

**2.5 เทคนิคที่ใช้ในการประเมินแบบจำลอง**

ตารางที่ 2 สูตรคำนวณค่าความคลาดเคลื่อน

ค่าความคลาดเคลื่อน	สูตรคำนวณ	
RMSE	$\sqrt{\frac{1}{N} \sum_{t=1}^N (y_t - \hat{y}_t)^2}$	(8)

MAE	$\frac{1}{N} \sum_{t=1}^N  y_t - \hat{y}_t $	(9)
-----	--	-----

MAPE	$\left( \frac{1}{N} \sum_{t=1}^N \left  \frac{y_t - \hat{y}_t}{y_t} \right  \right)$	(10)
------	--	------

โดยที่

RMSE คือ ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสอง

เฉลี่ย

MAE คือ ค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย

MAPE คือ ค่าสัมบูรณ์ของค่าเปอร์เซ็นต์ของความคลาดเคลื่อน

- $N$  คือ จำนวนข้อมูลที่ใช้
- $y_t$  คือ ค่าจริงที่เวลา  $t$  ใดๆ
- $\hat{y}_t$  คือ ค่าที่ได้จากการพยากรณ์ที่เวลา  $t$  ใดๆ

**3. วิธีดำเนินการวิจัย**

1) ทำการพยากรณ์ฝุ่น PM2.5 ด้วยแบบจำลอง ARIMA โดยใช้ข้อมูลตั้งแต่วันที่ 25 ธันวาคม พ.ศ. 2565 เวลา 0:00 น. ถึง 24 กุมภาพันธ์ พ.ศ. 2566 เวลา 23:00 น. เป็นจุดข้อมูลการเรียนรู้ และใช้ข้อมูลตั้งแต่วันที่ 25 กุมภาพันธ์ พ.ศ. 2565 เวลา 0:00 น. ถึง 24 มีนาคม พ.ศ. 2566 เวลา 23:00 น. เป็นชุดข้อมูลทดสอบ สำหรับการพยากรณ์ฝุ่น PM2.5 ส่วนแรก ซึ่งแบบจำลอง ARIMA สามารถคำนวณส่วนที่เหลือ (Residual) ได้ทันที จากนั้นนำส่วนที่เหลือตั้งแต่วันที่ 25 กุมภาพันธ์ พ.ศ. 2565 เวลา 0:00 น. ถึง 24 มีนาคม พ.ศ. 2565 เวลา 23:00 น. มาพยากรณ์ด้วยแบบจำลอง LSTM ซึ่งผลลัพธ์ในช่วงเวลา (t) ภายหลังจะคำนวณฝุ่น PM2.5 ส่วนแรกรวมกับส่วนที่เหลือ จะได้ค่าพยากรณ์ PM2.5 หรือขั้นตอนนี้คือการสร้างแบบจำลองไฮบริดที่ 1 (Hybrid1)

2) ทำการพยากรณ์ฝุ่น PM2.5 ด้วยแบบจำลอง LSTM โดยใช้ข้อมูลตั้งแต่วันที่ 25 ธันวาคม พ.ศ. 2565 เวลา 0:00 น. ถึง 24 กุมภาพันธ์ พ.ศ. 2566 เวลา 23:00 น. เป็นชุดข้อมูลการเรียนรู้ และใช้ข้อมูลตั้งแต่วันที่ 25 กุมภาพันธ์ พ.ศ. 2565 เวลา 0:00 น. ถึง 24 มีนาคม พ.ศ. 2566 เวลา 23:00 น. เป็นชุดข้อมูลทดสอบ สำหรับการพยากรณ์ ฝุ่น PM2.5 ส่วนแรก จากนั้นใช้ข้อมูลตั้งแต่วันที่ 23 กุมภาพันธ์ พ.ศ. 2565 เวลา 0:00 น. ถึง 10 มีนาคม พ.ศ. 2566 เพื่อใช้เป็นจุดข้อมูลการเรียนรู้สำหรับแบบจำลอง ARIMA จากนั้นใช้ข้อมูลตั้งแต่วันที่ 11 มีนาคม พ.ศ. 2565 เวลา 0:00 น. ถึง 24 มีนาคม พ.ศ. 2566 เป็นข้อมูลทดสอบ ใช้ค่าส่วนที่เหลือจากแบบจำลอง ARIMA มาคำนวณเชื่อมกับเป็นค่าส่วนที่เหลือของการพยากรณ์ จากนั้นจึงนำจากค่าฝุ่น PM2.5 ส่วนแรกรวมกับส่วนที่เหลือ จะได้ค่าพยากรณ์ PM2.5 หรือขั้นตอนนี้คือการสร้างแบบจำลองไฮบริดที่ 2 (Hybrid2)

3) นำแบบจำลองไฮบริดที่ 1 และแบบจำลองไฮบริดที่ 2 มาทำการคำนวณเพื่อช่วงนี้หนัก และเพิ่มประสิทธิภาพในการคำนวณ ด้วยสองวิธีการ เพื่อนำค่าที่ได้จากการคำนวณ ไปเปรียบเทียบกับประสิทธิภาพด้วยค่าความคลาดเคลื่อน

**3.1 วิธีคำนวณแบบที่ 1**

ถ้าหากให้

$$Hybrid1_{RMSE} = Hybrid1_{MAE} \tag{11}$$

$$Hybrid2_{RMSE} = Hybrid2_{MAE} \tag{12}$$

$$Hybrid1_{MAPE} = \frac{(1 - (Hybrid1_{RMSE} / (Hybrid1_{RMSE} + Hybrid2_{RMSE}))^2) \times 2}{2} \tag{13}$$

$$Hybrid2_{MAPE} = 2 - Hybrid1_{MAPE} \tag{14}$$

ถ้าหาก

$$Hybrid_{RMSE} = \frac{((Hybrid1_{MAPE} \times Hybrid_{RMSE1}) + (Hybrid2_{MAPE} \times Hybrid_{RMSE2}))}{2} \tag{15}$$

**บทความวิจัย**

การประชุมวิชาการวิศวกรรมศาสตร์และเทคโนโลยี มจร. ครอบคลุม ครั้งที่ 7  
Proceedings of the 7<sup>th</sup> RMUTP Conference on Engineering and Technology

**3.2 วิธีคำนวณแบบที่ 2**

คำนวณ

$$Hybrid_{PREDICT} = (Hybrid_{PREDICT1}) \times WEIGHT1 + (Hybrid_{PREDICT2}) \times WEIGHT2 \quad [16]$$

โดยที่

$$WEIGHT1 = [0.1, 0.2, \dots, 0.9]$$

$$WEIGHT2 = [0.9, 0.8, \dots, 0.1]$$

**4. ผลการวิจัย**

ตารางที่ 3 ค่าความคลาดเคลื่อน

แบบจำลอง	RMSE	MAE	MAPE
ARIMA-LSTM	5.5228	4.27186	0.09397
LSTM-ARIMA	6.96943	5.26854	0.10816

จากตารางที่ 3 เมื่อพิจารณาแบบจำลอง ARIMA-LSTM ได้ค่าความคลาดเคลื่อนค่าเฉลี่ยสองเฉลี่ย (RMSE) เท่ากับ 5.5228 ค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) เท่ากับ 4.27186 และค่าสัมบูรณ์ของเปอร์เซ็นต์ของความคลาดเคลื่อน (MAPE) เท่ากับ 0.09397 ซึ่งแบบจำลอง LSTM-ARIMA ได้ค่าความคลาดเคลื่อนค่าเฉลี่ยสองเฉลี่ย (RMSE) เท่ากับ 6.96943 ค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) เท่ากับ 5.26854 และค่าสัมบูรณ์ของเปอร์เซ็นต์ของความคลาดเคลื่อน (MAPE) เท่ากับ 0.10816 เมื่อพิจารณาเปรียบเทียบค่าความคลาดเคลื่อนของแบบจำลองไฮบริด 2 แบบจำลอง พบว่าแบบจำลอง ARIMA-LSTM ได้ผลการพยากรณ์ที่ดีกว่า แบบจำลอง LSTM-ARIMA

พบที่ 4 ค่าความคลาดเคลื่อนเปรียบเทียบแบบวิธีที่ 1

แบบจำลอง	RMSE	MAE	MAPE
ARIMA-LSTM x LSTM-ARIMA	5.82953	4.4395	0.09469

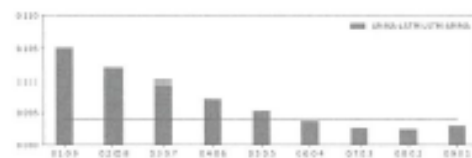
จากตารางที่ 4 นำแบบจำลอง ARIMA-LSTM และแบบจำลอง LSTM-ARIMA คำนวณค่าตัวนำเทคนิคการคำนวณแบบที่ 1 ได้ค่าความคลาดเคลื่อนค่าเฉลี่ยสองเฉลี่ย (RMSE) เท่ากับ 5.82953 ค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) เท่ากับ 4.4395 และค่าสัมบูรณ์ของเปอร์เซ็นต์ของความคลาดเคลื่อน (MAPE) เท่ากับ 0.09469 ซึ่งแบบจำลอง ARIMA-LSTM และแบบจำลอง LSTM-ARIMA มีความคลาดเคลื่อนสูงกว่าแบบจำลอง ARIMA-LSTM จากตารางที่ 3

ตารางที่ 5 ค่าความคลาดเคลื่อนเปรียบเทียบเทคนิควิธีที่ 2

Weight <sub>ARIMA</sub>	Weight <sub>LSTM</sub>	RMSE	MAE	MAPE
0.1	0.9	6.7072	5.08385	0.10505

0.2	0.8	6.46494	4.90525	0.10207
0.3	0.7	6.24497	4.74427	0.09944
0.4	0.6	6.04971	4.59796	0.09712
0.5	0.5	5.88164	4.47743	0.09528
0.6	0.4	5.74313	4.37651	0.09375
0.7	0.3	5.62637	4.30715	0.09271
0.8	0.2	5.56319	4.26403	0.09255
0.9	0.1	5.52491	4.25693	0.09304

จากตารางที่ 5 นำแบบจำลอง ARIMA-LSTM และแบบจำลอง LSTM-ARIMA แล้วมีการคำนวณค่าตัวนำเทคนิคการคำนวณแบบที่ 2 เปรียบเทียบค่าสัมบูรณ์ของเปอร์เซ็นต์ของความคลาดเคลื่อน (MAPE) กับแบบจำลอง ARIMA-LSTM สามารถแสดงได้ดังรูปที่ 2



รูปที่ 2 ค่าสัมบูรณ์ของเปอร์เซ็นต์ของความคลาดเคลื่อน (MAPE)

จากรูปที่ 2 เห็นชัดเจน แขน ค่าสัมบูรณ์ของเปอร์เซ็นต์ของความคลาดเคลื่อน (MAPE) จากแบบจำลอง ARIMA-LSTM แห่งนี้ที่แสดงค่าสัมบูรณ์ของเปอร์เซ็นต์ของความคลาดเคลื่อน (MAPE) ของแบบจำลอง ARIMA-LSTM และแบบจำลอง LSTM-ARIMA เมื่อเปรียบเทียบกัน พบว่า มีเพียงค่าตัวนำเทคนิคการคำนวณแบบจำลอง ARIMA-LSTM ตั้งแต่ 0.6 ขึ้นไปที่ได้ค่าสัมบูรณ์ของเปอร์เซ็นต์ของความคลาดเคลื่อน (MAPE) น้อยกว่า

จากตารางที่ 5 นำแบบจำลอง ARIMA-LSTM และแบบจำลอง LSTM-ARIMA แล้วมีการคำนวณค่าตัวนำเทคนิคการคำนวณแบบที่ 2 เปรียบเทียบค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) กับแบบจำลอง ARIMA-LSTM สามารถแสดงได้ดังรูปที่ 3



รูปที่ 3 ค่าสัมบูรณ์ของเปอร์เซ็นต์ของความคลาดเคลื่อน (MAPE)

จากรูปที่ 3 เห็นชัดเจน แขน ค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) จากแบบจำลอง ARIMA-LSTM แห่งนี้ที่แสดงค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) จากแบบจำลอง ARIMA-LSTM และแบบจำลอง LSTM-ARIMA เมื่อเปรียบเทียบกัน เพียงค่าตัวนำเทคนิคการคำนวณแบบจำลอง

**บทความวิจัย**

การประชุมวิชาการวิศวกรรมศาสตร์และเทคโนโลยี มหาวิทยาลัยธุรกิจบัณฑิต ครั้งที่ 7  
*Proceedings of the 7<sup>th</sup> RMUTP Conference on Engineering and Technology*

ARIMA-LSTM ตั้งแต่ 0.8 ขึ้นไปที่ได้ค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) น้อยกว่า

จากตารางที่ 5 จำนวนข้อผิดพลาด ARIMA-LSTM และแบบจำลอง LSTM-ARIMA ยังมีการคำนวณค่ากว่าโมเดลการเรียนรู้การคำนวณแบบที่ 2 เปรียบเทียบค่าราคาที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) สามารถลดลงได้ดังรูปที่ 4



รูปที่ 4 ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE)

จากรูปที่ 5 เปรียบเทียบ ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) จากแบบจำลอง ARIMA-LSTM และแบบจำลอง LSTM-ARIMA โดยค่าเป้าหมายที่ดีที่สุด คือ ค่าเป้าหมายสำหรับแบบจำลอง ARIMA-LSTM เท่ากับ 0.9 และค่าเป้าหมายที่ดีที่สุดสำหรับแบบจำลอง LSTM-ARIMA เท่ากับ 0.1 โดยค่าความคลาดเคลื่อนค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 5.52491 ค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) เท่ากับ 4.25693 และค่าสัมบูรณ์ของเปอร์เซ็นต์ของความคลาดเคลื่อน (MAPE) เท่ากับ 0.69304

**5. สรุป**

จากผลวิเคราะห์ของพบว่า การนำผลการพยากรณ์คำนวณร่วมกับค่าเฉลี่ยเคลื่อนที่ของแบบจำลอง เพื่อให้ได้ค่าพยากรณ์ที่เหมาะสมสามารถให้ค่าพยากรณ์ที่คำนวณจากของโมเดลที่คิด และเมื่อพิจารณาแบบจำลอง ARIMA-LSTM และแบบจำลอง LSTM-ARIMA ที่มีการคำนวณปรับค่าอย่างมีประสิทธิภาพ การคำนวณปรับค่าอย่างมีประสิทธิภาพวิธีแบบยุคลิด แบบที่ 2 ได้ดีที่สุดของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) น้อยกว่าการคำนวณปรับค่าอย่างมีประสิทธิภาพวิธีการคำนวณแบบที่ 1 จึงสรุปว่าการคำนวณปรับค่าอย่างมีประสิทธิภาพวิธีการคำนวณแบบที่ 2 ให้ผลการพยากรณ์ที่ดีกว่า จากผลของการทดสอบแบบจำลองสามารถนำมาใช้ประโยชน์ได้โดยใช้ในการพยากรณ์ค่าฝุ่น PM2.5 ในเขตกรุงเทพมหานครช่วงชั่วโมง เพื่อพยากรณ์ค่าฝุ่นรายสัปดาห์ เป็นต้น ทั้งนี้ข้อมูลที่ใช้ในการทดลองเป็นข้อมูลระยะสั้น ในกรณีที่ทำการทดลองที่ใช้ข้อมูลระยะยาวจึงมีการเคลื่อนไหวตามฤดูกาล (Seasonal หรือ Periodic Movement) ของแบบการคำนวณปรับค่าอย่างมีประสิทธิภาพได้ผลลัพธ์ที่แตกต่างออกไป

**เอกสารอ้างอิง**

[1] G. Peter Zhang, "Time series forecasting using a hybrid ARIMA and neural network model", *Neurocomputing*, vol. 50, pp. 159 –

[2] กระทรวงสาธารณสุข, "คู่มือการดำเนินงานด้านการแพทย์และสาธารณสุข กรณี ศูนย์ตรวจเฝ้าระวังโรค 2.5 โทरण (PM2.5) ปี 2561", หน้า 1-7, 2562.

[3] C. Yewei & W. Kaizhi, "Prediction of Satellite Time Series Data Based on Long Short Term Memory Autoregressive Integrated Moving Average Model (LSTM-ARIMA)," *IEEE International Conference on Signal and Image Processing*, 2019.

[4] P. Kasikrot & K. Chompoonot, "Shallow Price Forecasting Models", *Data Science and Engineering*, vol. 3(1), pp. 35-43, 2022.

[5] J. Xiao, Q. Wang, J. Cui, & J. Yu, "Multi-feature PM2.5 Prediction with ARIMA-LSTM", *International Conference on Machine Learning, Cloud Computing and Intelligent Mining*, 2022.

[6] กองจัดการคุณภาพอากาศและเสียง, (2566). "รายงาน". เก็บถาวรเมื่อ 25 มีนาคม 2566. จาก <https://bangkokairquality.com/bma/report/> [Lang-th]

[7] S. W. Shuan, "Time Series Analysis: Forecasting and Control", *Journal of Quality Technology*, vol. 49(5), pp. 418-419, 2017.

[8] F.A. Gers, J. Schmidhuber and F. Cummins, "Learning to forget: continuous prediction with LSTM," *Neural Networks*, vol.12(f0), pp. 2451-2471, October 2000.



**จิรพร พลธนา** ได้รับปริญญาวิทยาศาสตรบัณฑิต (เทคโนโลยีสารสนเทศ) เกียรตินิยมอันดับสอง จาก มหาวิทยาลัยอัสสัมชัญเมื่อปี พ.ศ. 2544 ปริญญาตรุษศาสตร์มหาบัณฑิต จาก มหาวิทยาลัยรามคำแหง ปี พ.ศ. 2549 ปริญญาวิทยาศาสตรมหาบัณฑิต (ไอที) ปี พ.ศ. 2556 จาก มหาวิทยาลัยรามคำแหง ปี พ.ศ. 2556 ปัจจุบันดำรงตำแหน่งนักวิเคราะห์ ความคุม รางและระบบ ธนาคารทหารไทยธนชาต จำกัด มีความสนใจในเทคโนโลยีปัญญาประดิษฐ์ (Artificial Intelligence) วิศวกรรมข้อมูล (Data Engineer) และการเรียนรู้ของเครื่อง (Machine Learning)



**ดวงใจ ชัยคงชื่น** ผู้ช่วยศาสตราจารย์ ดร. ได้รับปริญญาบริหารธุรกิจบัณฑิต (คอมพิวเตอร์ธุรกิจ) จาก มหาวิทยาลัยธุรกิจบัณฑิตย์ ปี พ.ศ. 2544 ปริญญาวิทยาศาสตรมหาบัณฑิต เทคโนโลยีสารสนเทศ จาก มหาวิทยาลัยอัสสัมชัญ ปี พ.ศ. 2547 ปริญญาคุณวุฒิบัณฑิต (เทคโนโลยีสารสนเทศ) ปี พ.ศ. 2557 ปัจจุบันดำรงตำแหน่ง ผู้อำนวยการศูนย์วิศวกรรมข้อมูลขนาดใหญ่ วิทยาลัยวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี มีประสบการณ์สอนในสาขาวิชาวิศวกรรมศาสตร์ มหาวิทยาลัยธุรกิจบัณฑิตย์ มีประสบการณ์งานวิจัยทางด้านปัญญาประดิษฐ์ (Artificial Intelligence) วิทยาศาสตร์ข้อมูล (Data Science) ที่ 807111เชิงวิวัฒนาการ (Evolutionary Algorithms) และการเรียนรู้ของเครื่อง (Machine Learning)

**ภาคผนวก ค**  
**ผลงานตีพิมพ์ ชั้นที่ 2**

ICBISS  
2023

# Certificate of Presentation

DPU International Conference on Business, Innovation, and  
Social Sciences 2023  
25-26 May, 2023, Bangkok, Thailand

**Paper Title:** การศึกษาประสิทธิภาพการพยากรณ์ฝุ่น PM2.5 ในเขตกรุงเทพมหานคร  
ด้วยตัวแปรเดียวและหลายตัวแปร โดยแบบจำลองไฮบริด

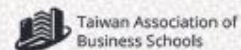
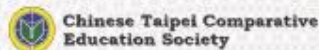
**Presenter' s Name:** ชัชกร พลเสนา      ดวงใจ จิตคงชื่น

**Presenter' s Affiliation:** มหาวิทยาลัยธุรกิจบัณฑิต



(Dr. Darika Lathapipat)

President of Dhurakij Pundit University





ICBISS  
2023

55  
55th Anniversary  
1968-2023

DPU 傳仁大學  
DHURAKIJ PUNDIT UNIVERSITY

Chinese Taipei Comparative  
Education Society

Taiwan Association of  
Business Schools

# DPU International Conference on Business, Innovation, and Social Sciences 2023



Bangkok, Thailand | 25 - 26 May 2023

*DPU International Conference on Business Innovation and Social Sciences 2023 (DPU ICBISS 2023)*  
<https://www.dpu.ac.th/icbiss2023/>

แนวทางการพัฒนาพัฒนาการเก็บรวบรวมพยานหลักฐานทางนิติวิทยาศาสตร์และการป้องกันสถานที่เกิดเหตุของพนักงานสอบสวนเขตพื้นที่จังหวัดนครปฐม 10 สถานี.....	1452-1462
ชยกร มีสาวงษ์	
บัณฑิตยประชากรศาสตร์และส่วนประสมทางการตลาดที่มีต่อความเชื่อมั่นใช้บริการทางการเงินผ่าน Mobile Banking Application ของธนาคารแห่งหนึ่ง.....	1463-1472
ชญานิน ไชยรักษ์ ปิยะวิทย์ ทิพรส นวพงศ์ ตันนทติลล	
บัณฑิตยที่มีผลต่อการตัดสินใจซื้อเฟอร์นิเจอร์ออนไลน์ ของคนวัยทำงาน ในกรุงเทพมหานคร.....	1473-1482
Feng Lu คุณากร วิวัฒนากรวงศ์	
บัณฑิตยส่วนบุคคลของนักออกแบบศิลปะสร้างสรรค์ (Creative art) ที่มีผลกระทบต่อพัฒนาอาชีพนักออกแบบในธุรกิจการพัฒนาศิลปะเป็นสินทรัพย์ดิจิทัล (NFT).....	1483-1493
สุริยนต์ สติณ ธนิตา จิตรน้อมรัตน์	
บัณฑิตยประชากรศาสตร์ ความรู้ และทัศนคติที่มีผลต่อพฤติกรรมยอมรับปฏิบัติการใช้ยาปฏิชีวนะของบุคลากรประจำการไฟฟ้าส่วนภูมิภาค สำนักงานใหญ่ กรุงเทพมหานคร.....	1494-1504
วินิจ เกิดเพ็ง ปิยะวิทย์ ทิพรส อาริน พงษ์ชัยภูมิ วิศรา แหมมทอง	
การเพิ่มประสิทธิภาพของการจัดการคลังสินค้าด้วยเทคโนโลยีบาร์โค้ด: กรณีศึกษา บริษัท ดีเอทีฟิยูนิฟอร์ม จำกัด.....	1505-1514
กรภัทร์ บุญเรืองพานิชย์ ณัฐธยาน์ ไสกุล สุภรัชชัย วรรัตน์	
การต้านทานจากการโจมตี DDoS ของบอร์ด Raspberry Pi ที่ทำงานเป็น IoT.....	1515-1523
อินทพงษ์ รัตนติลล ณ กุเกิด ชัยพร เขมะภาคะพันธ์	
การศึกษาประสิทธิภาพการพยากรณ์ฝุ่น PM2.5 ในเขตกรุงเทพมหานคร ด้วยตัวแปรเดียวและหลายตัวแปรโดยแบบจำลองไฮบริด.....	1524-1535
อัคร พลเสนา ดวงใจ จิตคงชื่น	
การประเมินการปล่อยก๊าซเรือนกระจกจากอาคารสำนักงานโดยการประเมินวัฏจักรชีวิต.....	1536-1545
ชัยวัฒน์ มั่นคงดี สุภรัชชัย วรรัตน์ อานาจ ผดุงศิลป์ วิชรพงษ์ ตันธวัฒน์ ประยุทธ์ ฤทธิเดช	

## การศึกษาประสิทธิภาพการพยากรณ์ฝุ่น PM2.5 ในเขตกรุงเทพมหานคร ด้วยตัวแปรเดียวและ หลายตัวแปรโดยแบบจำลองไฮบริด

ธัชกร พลเสนา<sup>1\*</sup> ดวงใจ จิตคงชื่น, ผู้ช่วยศาสตราจารย์ ดร.<sup>2</sup>

<sup>1</sup>มหาวิทยาลัยธุรกิจบัณฑิต; <sup>2</sup>มหาวิทยาลัยธุรกิจบัณฑิต

\*645162020017@dpu.ac.th

### บทคัดย่อ

ความต้องการมีสุขภาพที่แข็งแรงเป็นความต้องการพื้นฐานที่ตรงกันของมนุษยชาติ ซึ่งองค์การอนามัยโลกทำหน้าที่ดูแลและประสานงานด้านการสาธารณสุขเพื่อคุ้มครองและส่งเสริมสุขภาพอนามัยของประชาชนทั่วโลก โดยองค์การอนามัยโลกได้มีประกาศปรับปรุงมาตรฐานฝุ่นละออง PM2.5 ใหม่ แสดงถึงฝุ่นละออง PM2.5 เป็นปัญหาที่ควรให้ความสำคัญ อันเนื่องมาจากฝุ่นละออง PM2.5 เป็นตัวกลางพาสารอันตรายเข้าสู่ปอด โดยประเทศไทยเริ่มให้ความสำคัญกับปัญหาฝุ่นละออง PM2.5 ในปี พ.ศ. 2562 ความสามารถในการพยากรณ์ฝุ่นละออง PM2.5 จึงถือเป็นแนวทางหนึ่งในการป้องกันปัญหาสุขภาพที่อาจเกิดขึ้นจากฝุ่นละออง PM2.5 แบบจำลองที่มีถูกใช้ในการพยากรณ์ฝุ่นละออง PM2.5 คือ แบบจำลองไฮบริด ตามแนวคิดของ Peter Zhang ในปี ค.ศ. 2003 โดยใช้แบบจำลอง Arima พยากรณ์ข้อมูลที่มีโครงสร้างเชิงเส้น หลังจากพยากรณ์ค่าส่วนที่เหลือด้วยแบบจำลองโครงข่ายประสาทเทียม ซึ่งขณะนั้นคือแบบจำลอง ANN แต่ในปัจจุบันถูกพัฒนาเป็นแบบจำลอง LSTM โดยแบบจำลอง LSTM นั้นมีประสิทธิภาพในการพยากรณ์ทั้งตัวแปรเดียวและหลายตัวแปร

สำหรับกรุงเทพมหานคร เมืองหลวงของประเทศไทยนั้น มักประสบปัญหาฝุ่นละออง PM2.5 ในช่วงฤดูหนาว ซึ่งตั้งแต่เดือนมกราคม พ.ศ. 2566 ที่ผ่านมา เริ่มมีรายงานข่าวถึงเขตที่มีค่าฝุ่นละออง PM2.5 เกินกว่ามาตรฐานที่กรมควบคุมมลพิษกำหนด ผู้วิจัยจึงเกิดแนวคิดในการเปรียบเทียบประสิทธิภาพการพยากรณ์ด้วยการเปรียบเทียบค่าความคลาดเคลื่อน โดยการพยากรณ์ด้วยแบบจำลอง Arima กับแบบจำลอง LSTM และแบบจำลอง BiLSTM แบบตัวแปรเดียว แบบจำลอง Arima กับแบบจำลอง LSTM และแบบจำลอง BiLSTM แบบหลายตัวแปร รวมถึงแบบจำลอง Arima กับแบบจำลอง LSTM และแบบจำลอง BiLSTM แบบเลือกตัวแปรที่ไม่มีค่าสหสัมพันธ์กับค่าส่วนที่เหลือ

ผลการทดสอบพบว่า มีเพียงแบบจำลอง Arima กับแบบจำลอง LSTM และแบบจำลอง BiLSTM แบบเลือกตัวแปรให้ค่าดีกว่าการพยากรณ์ด้วยแบบจำลอง Arima อย่างเดียว โดยแบบจำลอง Arima-LSTM ให้ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Error: MAE) เท่ากับ 4.3073 ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Squared Error: RMSE) เท่ากับ 5.5122 ค่าสัมบูรณ์ของเปอร์เซ็นต์ของความคลาดเคลื่อน (Mean Absolute Percentage Error : MAPE) เท่ากับ 0.0928 ซึ่งทุกค่าความผิดพลาดต่ำที่สุดเมื่อเทียบกับแบบจำลองอื่น

**คำสำคัญ:** PM2.5, การพยากรณ์, แบบจำลองไฮบริด, Univariate, Multivariate

### 1. บทนำ

ความต้องการมีสุขภาพที่แข็งแรง เป็นความต้องการความมั่นคงปลอดภัย (Safety Needs) ที่พื้นฐานที่สุดและ

ตรงกันของมนุษยชาติ ตามทฤษฎีลำดับขั้นความต้องการของมาสโลว์ (Maslow's Hierarchy of Needs) ของ อับราฮัม เอช. มาสโลว์ โดยเป็นทฤษฎีจิตวิทยาที่เกิดขึ้นเมื่อปี ค.ศ. 1943 (Maslow, 1943) ซึ่งเป็นแนวคิดที่แพร่หลายและเป็นที่ยอมรับจนถึงปัจจุบัน จากความต้องการพื้นฐานที่ตรงกันของมนุษยชาตินี้ จึงมีการจัดตั้งองค์การอนามัยโลก (World Health Organization: WHO) ซึ่งเป็นหน่วยงานระหว่างประเทศ ทำหน้าที่ดูแลและประสานงานด้านการสาธารณสุขเพื่อคุ้มครองและส่งเสริมสุขภาพอนามัยของประชาชนทั่วโลก และเมื่อเร็วๆ นี้ องค์การอนามัยโลกได้มีประกาศมาตรฐานฝุ่นละออง PM2.5 ใหม่ ส่งผลให้มีราชกิจจานุเบกษาเผยแพร่ประกาศคณะกรรมการสิ่งแวดล้อมแห่งชาติ เมื่อวันที่ 8 กรกฎาคม พ.ศ. 2565 ได้กำหนดมาตรฐานฝุ่นละออง PM2.5 ซึ่งค่าเฉลี่ยในเวลา 24 ชั่วโมง จะต้องไม่เกิน 50 ไมโครกรัมต่อลูกบาศก์เมตรโดยให้มีผลจนถึงวันที่ 31 พฤษภาคม พ.ศ. 2566 (กรมควบคุมมลพิษ, 2565) ซึ่งฝุ่นละออง PM2.5 หรือมีชื่อทางการเรียกว่า "Particulate matter with diameter of less than 2.5 micron" เป็นฝุ่นละอองที่มีขนาดไม่เกิน 2.5 ไมครอน ทำหน้าที่เป็นตัวกลางพาสารอันตรายต่าง ๆ เข้าสู่ปอด เช่น สารก่อมะเร็ง สารโลหะหนัก เป็นต้น (กระทรวงสาธารณสุข, 2562)

ข้อมูลฝุ่นละออง PM2.5 เป็นข้อมูลคุณภาพอากาศมีลักษณะเป็นข้อมูลอนุกรมเวลา (Time series) ไม่คงที่และไม่เป็นข้อมูลเชิงเส้น หากสามารถพยากรณ์ข้อมูลฝุ่นละออง PM2.5 ได้ จะเป็นประโยชน์ในการหาแนวทางในการป้องกันอันตรายที่จะส่งผลต่อสุขภาพ ซึ่งแบบจำลองไฮบริด (Hybrid model) เช่น แบบจำลอง Arima-LSTM มักถูกนำมาใช้ในการพยากรณ์โดยการสอนให้ระบบคอมพิวเตอร์ทำการเรียนรู้ได้ด้วยตนเองจากการใช้ข้อมูลในอดีต (Machine Learning) ตามแนวคิดของเริ่มต้นของทีโอเกอเนตขึ้นในปี ค.ศ. 2003 โดย Peter Zhang เนื่องจากข้อดีของแบบจำลองอริมา (Auto Regressive Integrated Moving Average: Arima) ที่ถูกนำมาใช้สำหรับการพยากรณ์ข้อมูลที่มีโครงสร้างเชิงเส้น หลังจากนั้นนำค่าส่วนที่เหลือ (Residual) ไปพยากรณ์ร่วมกับโครงข่ายประสาทเทียม (Artificial Neural Networks: ANN) (Zhang, 2003) ซึ่งในปัจจุบันค่าส่วนที่เหลือ (Residual) มักพยากรณ์ด้วยแบบจำลองหน่วยความจำสั้นยาว (Long Short-Term Memory: LSTM) เพื่อแก้ปัญหาการสูญหายของเกรเดียนต์ (Vanishing Gradient Problem: VGP) นอกจากนี้แบบจำลอง LSTM ยังเป็นแบบจำลองที่เป็นที่ยอมรับในการพยากรณ์ได้ดีทั้ง Global Univariate Models และ Global Multivariate Models อีกด้วย (DiBattista, 2022)

กรุงเทพมหานคร ประกอบไปด้วย 50 เขต และเป็นเมืองหลวงของประเทศไทยมักประสบปัญหาฝุ่นละออง PM2.5 ในระหว่างช่วงเดือนธันวาคม ถึง เดือนเมษายน ปีถัดไป (กระทรวงสาธารณสุข, 2562) ด้วยคุณสมบัติของแบบจำลอง LSTM ผู้วิจัยจึงเกิดแนวคิดในการเปรียบเทียบประสิทธิภาพการพยากรณ์ด้วยการเปรียบเทียบค่าความคลาดเคลื่อน โดยการพยากรณ์ด้วยแบบจำลอง Arima กับแบบจำลอง LSTM และแบบจำลอง BiLSTM แบบตัวแปรเดียว แบบจำลอง Arima กับแบบจำลอง LSTM และแบบจำลอง BiLSTM แบบหลายตัวแปร รวมถึงแบบจำลอง Arima กับแบบจำลอง LSTM และแบบจำลอง BiLSTM แบบเลือกตัวแปรที่ไม่มีค่าสหสัมพันธ์กับค่าส่วนที่เหลือ (Residual) เพื่อป้องกันความโน้มเอียง (Bias) จากการพยากรณ์ ซึ่งผลจากการศึกษาประสิทธิภาพของแบบจำลองในครั้งนี้ สามารถเป็นแนวทางในการการพยากรณ์ค่าฝุ่น PM2.5 ได้อีกด้วย

#### ภาพที่ 1

แสดงกรอบแนวคิดในการวิจัย



## 2. การทบทวนวรรณกรรม

### 2.1 แนวคิด ทฤษฎี และเครื่องมือที่ใช้ในการวิเคราะห์

บทความวิจัยเรื่อง “การศึกษาประสิทธิภาพการพยากรณ์ฝุ่น PM2.5 ในเขตกรุงเทพมหานคร ด้วยตัวแปรเดียวและหลายตัวแปรโดยแบบจำลองไฮบริด” นี้ใช้แนวคิด 3 แบบจำลอง ดังนี้

#### 2.1.1 แนวคิดแบบจำลอง Hybrid

ในช่วงปี ค.ศ. 2003 เครือข่ายประสาทเทียม/สังคม (artificial neural networks : ANNs) ได้มีการศึกษาอย่างกว้างขวางและใช้ในการพยากรณ์อนุกรมเวลา ข้อได้เปรียบที่สำคัญของโครงข่ายประสาทเทียมคือความสามารถในการสร้างแบบจำลองแบบไม่เชิงเส้นที่ยืดหยุ่นด้วย ANN ไม่จำเป็นต้องระบุรูปแบบเฉพาะ แต่แบบจำลองนั้นถูกสร้างขึ้นแบบปรับเปลี่ยนตามคุณสมบัติที่นำเสนอจากข้อมูล ทั้งโมเดล ARIMA และ ANN ต่างประสบความสำเร็จในโดเมนเชิงเส้นหรือไม่เชิงเส้นของตนเอง อย่างไรก็ตาม ไม่มีแบบจำลองที่เหมาะสมกับทุกสถานการณ์ การประมาณแบบจำลอง ARIMA กับปัญหาไม่เชิงเส้นที่ซับซ้อนอาจไม่เพียงพอ ในทางกลับกัน การใช้ ANN เพื่อสร้างแบบจำลองปัญหาเชิงเส้นให้ผลลัพธ์ที่หลากหลาย จึงเกิดแนวคิดวิธีการแบบผสมผสานที่มีความสามารถในการสร้างแบบจำลองทั้งเชิงเส้นและไม่เชิงเส้นอาจที่เป็นกลยุทธ์ที่ดีสำหรับการใช้งานจริง โดยการรวมโมเดลที่แตกต่างกัน (Zhang, 2003)

ดังนั้น จึงมีเหตุผลที่จะพิจารณาอนุกรมเวลาที่ประกอบด้วยโครงสร้างความสัมพันธ์เชิงเส้นอัตโนมัติและองค์ประกอบที่ไม่เชิงเส้น ตามสมการ (2.1)

$$Y_t = L_t + N_t \quad (2.1)$$

โดยที่  $L_t$  หมายถึง ส่วนข้อมูลที่เป็นโครงสร้างเชิงเส้น

$N_t$  หมายถึง ส่วนข้อมูลที่ไม่เป็นโครงสร้างเชิงเส้น

กระบวนการต้องมีการประมาณสององค์ประกอบจากข้อมูล อันดับแรก ให้ ARIMA เป็นแบบจำลองส่วนประกอบเชิงเส้น จากนั้นส่วนที่เหลือจากแบบจำลองเชิงเส้นจะมีเฉพาะความสัมพันธ์แบบไม่เชิงเส้น ให้  $e_t$

แทนค่าที่เหลือ (residuals) ณ เวลา  $t$  จากโมเดลเชิงเส้น ตามสมการ (2.2)

$$e_t = Y_t - \hat{C}_t \quad (2.2)$$

โดยที่  $\hat{C}_t$  หมายถึง ค่าพยากรณ์สำหรับเวลา  $t$

จากความสัมพันธ์ที่ประมาณไว้ ค่าที่เหลือมีความสำคัญใน การวินิจฉัยความสมบูรณ์ของโมเดลเชิงเส้น โมเดลเชิงเส้นไม่ใช่เพียงพอหากยังมีโครงสร้างความสัมพันธ์เชิงเส้นเหลืออยู่ในส่วนที่เหลือ อย่างไรก็ตาม การวิเคราะห์สามารถตรวจจบบรูปแบบที่ไม่เชิงเส้นในข้อมูลได้ ซึ่งส่วนที่เหลือจะระบุข้อจำกัดของ ARIMA ดังนั้นการสร้างแบบจำลองส่วนที่เหลือโดยใช้ ANN จะสามารถค้นพบความสัมพันธ์แบบไม่เชิงเส้นได้ ด้วย  $n$  โหนดอินพุต รูปแบบ ANN สำหรับส่วนที่เหลือจะแสดงได้ตามสมการ (2.3)

$$e_t = f(e_{t-1}; e_{t-2}; \dots; e_{t-n}) + \mathcal{E}_t \quad (2.3)$$

โดยที่  $f$  หมายถึง ฟังก์ชันไม่เชิงเส้นที่กำหนดโดยโครงข่ายประสาทเทียม

$\mathcal{E}_t$  หมายถึง ข้อผิดพลาดของการสุ่ม

จากสมการ (2.3)  $N_t$  การคาดการณ์แบบรวมจะแสดงได้ตามสมการ (2.4)

$$Y_t = \hat{C}_t + N_t \quad (2.4)$$

โดยสรุป วิธีการที่นำเสนอของระบบไฮบริดประกอบด้วยสองขั้นตอน ขั้นแรกจะใช้แบบจำลอง ARIMA เพื่อวิเคราะห์ส่วนเชิงเส้นของปัญหา ส่วนขั้นตอนที่สอง มีการพัฒนาแบบจำลองโครงข่ายประสาทเทียมเพื่อจำลองสิ่งตกค้างจากโมเดล ARIMA เนื่องจากโมเดล ARIMA ไม่สามารถจับโครงสร้างแบบไม่เชิงเส้นของ ข้อมูลส่วนที่เหลือของโมเดลเชิงเส้นจะมีข้อมูลเกี่ยวกับความไม่เชิงเส้น

ผลลัพธ์จากโครงข่ายประสาทเทียมสามารถใช้เป็นการคาดคะเนเงื่อนไขข้อผิดพลาดได้สำหรับแบบจำลอง ARIMA แบบจำลองไฮบริดใช้ประโยชน์จากคุณลักษณะเฉพาะและความแข็งแกร่งของโมเดล ARIMA และโมเดล ANN ในการกำหนดรูปแบบที่แตกต่างกัน ดังนั้นจึงสามารถเป็นประโยชน์ในการสร้างแบบจำลองเชิงเส้นและไม่เชิงเส้นแยกจากกันโดยใช้ความแตกต่างแบบจำลองแล้วรวมการคาดการณ์เพื่อปรับปรุงการสร้างแบบจำลองและการพยากรณ์โดยรวม

### 2.1.2 แนวคิดแบบจำลอง Arima

การวิเคราะห์อนุกรมเวลาที่เป็นการใช้ข้อมูลในอดีตเพื่อใช้เป็นแนวทางในการพยากรณ์อนาคตด้วยอนุกรมเวลาแบบ Box-Jenkins (1976) ที่เรียกว่า Autoregressive Integrated Moving Average Model หรือเป็นที่รู้จักว่าแบบจำลอง ARIMA ซึ่งเป็นแบบจำลองถดถอย ประกอบด้วยองค์ประกอบของ autoregression, integrated และ moving average เป็นแบบจำลองถดถอยใช้ในการพยากรณ์ข้อมูลอนุกรมเวลาที่ยังอิง  $p, d, q$  .Shaun (2017) ได้อธิบายแบบจำลอง ARIMA ไว้ดังนี้

Auto Regression, AR( $p$ ) เป็นส่วนหนึ่งของแบบจำลอง ARIMA ตามแนวคิดที่ว่าจะใช้ความล่าช้าของตัวเอง (ค่าในอดีต) เป็นตัวทำนาย โดยที่พารามิเตอร์  $p$  คือ ขอบเขตของจำนวนความล่าช้าที่เกิดขึ้น

Integrated I( $d$ ) ช่วยทำให้ข้อมูลอนุกรมเวลาอยู่กับที่เพื่อลดการขึ้นต่อกันของเวลาและแนวโน้ม โดยที่พารามิเตอร์  $d$  คือ ระดับความแตกต่างซึ่งหมายถึงจำนวนครั้งที่ข้อมูลมีความแตกต่างกัน ถ้าอนุกรมเวลาหยุดนิ่ง (stationary) ระดับความแตกต่างของมันคือศูนย์

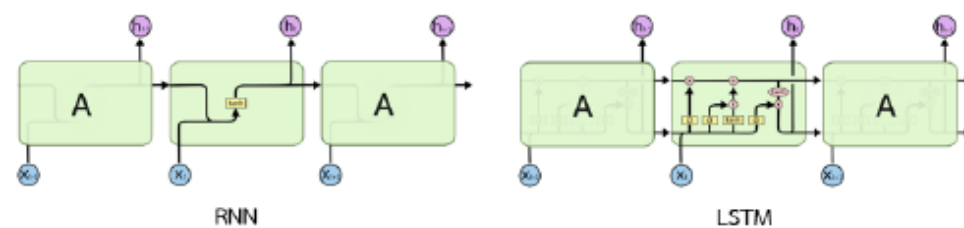
Moving Average, MA( $q$ ) ใช้ข้อผิดพลาดที่เหลือ (residual) ของเวลาที่ผ่านมาซึ่งให้เห็นการคาดการณ์ในปัจจุบันและอนาคต โดยที่พารามิเตอร์  $q$  คือ จำนวนของข้อผิดพลาดการคาดการณ์ความล่าช้าที่ใช้ในการคำนวณค่าปัจจุบัน

### 2.1.3 แนวคิดแบบจำลอง LSTM

หน่วยความจำระยะสั้นระยะยาว (Long Short Term Memory: LSTM) เป็นโครงข่ายประสาทเทียมแบบวนกลับ (Recurrent Neural Networks: RNN) ที่สามารถเรียนรู้การพึ่งพาระยะยาวได้ มีโครงสร้างที่เชื่อมต่อกันเช่นกัน แต่ประกอบด้วยโครงสร้างที่แตกต่างและมีความซับซ้อนกว่ามากที่โต้ตอบซึ่งกันและกันดังภาพที่ 2 (Dagshub, 2023)

ภาพที่ 2

แสดงโมดูลโครงข่ายประสาทเทียมแบบวนกลับ (RNN) และโมดูลหน่วยความจำระยะสั้นระยะยาว (LSTM)



ภายในโมดูล LSTM หนึ่งโมดูล ส่วนประกอบสำคัญที่อนุญาตให้ถ่ายโอนข้อมูลผ่านโมดูลทั้งหมดเรียกว่าสถานะของเซลล์ (cell state) นอกจากการถ่ายโอนข้อมูลแล้ว โมดูลยังมีความสามารถในการเพิ่มหรือลบข้อมูลไปยังสถานะของเซลล์ ซึ่งควบคุมโดยโครงสร้างที่เรียกว่าเกต (gate) ในโมดูล LSTM มาตรฐาน มีทั้งหมด 3 เกต ดังนี้

- 1) เกตลืม (forget gate) ทำหน้าที่กำหนดจำนวนสถานะของเซลล์ที่ขั้นตอนของเวลาก่อนหน้าจะยังคงอยู่ในขั้นตอนของเวลาปัจจุบัน และอาจไม่เก็บไว้เลยก็ได้
- 2) อินพุตเกต (input gate) ทำหน้าที่กำหนดจำนวนอินพุต  $u$  ปัจจุบัน บันทึกลงขั้นตอนเวลาในสถานะ

เซลล์ บางครั้งเรียกอินพุตเกทว่าเกทปรับปรุง

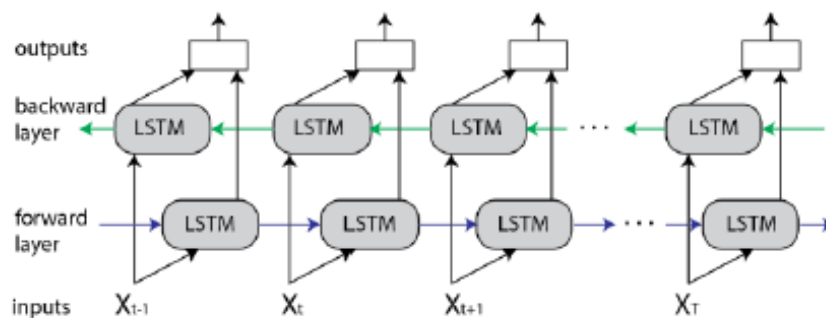
3) เกทเอาต์พุต (Output Gate:  $o$ ) กำหนดจำนวนสถานะของเซลล์ที่ส่งออก เป็นค่าเอาต์พุตปัจจุบันของสถานะที่ซ่อนอยู่

กล่าวโดยสรุป LSTM ใช้เส้นทางที่แยกจากกันสำหรับหน่วยความจำระยะยาวและหน่วยความจำระยะสั้น และใช้ 3 เกท ประกอบด้วย เกทลืม (forget gate) อินพุตเกท (input gate) และเกทออก (output gate) เพื่อควบคุมสถานะของเซลล์ และเพื่อจัดการกับปัญหาการพังพารระยะยาวของ RNN โดยการกรองข้อมูลที่ไร้ประโยชน์ออกจากเครือข่าย

LSTM แบบสองทิศทาง (A Bidirectional LSTM: BiLSTM) เป็นโครงข่ายประสาทแบบเกิดซ้ำ ซึ่งแตกต่างจาก LSTM มาตรฐานตรงที่อินพุตไหลในทั้งสองทิศทาง และสามารถใช้อินพุตจากทั้งสองด้านได้ BiLSTM ทำการเพิ่มเลเยอร์ LSTM อีกหนึ่งชั้น ซึ่งจะกลับทิศทางการไหลของข้อมูล ซึ่งหมายถึงลำดับอินพุตไหลย้อนกลับในเลเยอร์ LSTM เพิ่มเติม ตามด้วยการรวมเอาต์พุตจากเลเยอร์ LSTM ดังภาพที่ 3

ภาพที่ 3

แสดงแนวคิดหน่วยความจำระยะสั้นระยะยาวแบบสองทิศทาง (BiLSTM)



## 2.2 ทบทวนวรรณกรรม

จากการศึกษาพบว่ามโนทัศน์ที่เกี่ยวข้อง ดังนี้

งานวิจัยเรื่อง “Time series forecasting using a hybrid ARIMA and neural network model” (Zhang, G. P., 2003) กล่าวถึงแนวคิดแบบจำลองไฮบริด (Hybrid Model) เป็นการนำแบบจำลองมาใช้ในการพยากรณ์ร่วมกัน ประกอบไปด้วยแบบจำลอง ARIMA และ ANN ผลการทดลองพบว่าค่าเฉลี่ยความผิดพลาดกำลังสอง (Mean Square Error: MSE) และค่าเฉลี่ยความผิดพลาดสัมบูรณ์ (Mean Absolute Deviation: MAD) ของแบบจำลองไฮบริด ได้ค่าเฉลี่ยความผิดพลาดน้อยที่สุด จึงสรุปได้ว่าแบบจำลองไฮบริดสามารถให้ค่าพยากรณ์ที่ถูกต้องแม่นยำกว่าเมื่อเปรียบเทียบกับการพยากรณ์ด้วยแบบจำลอง ARIMA หรือแบบจำลอง LSTM แบบจำลองเดียว



งานวิจัยเรื่อง “Time series forecasting using a hybrid ARIMA and LSTM model” (Fathi, 2019) ทำการวิจัยเพื่อพยากรณ์ข้อมูลอนุกรมโดยอาศัยแนวคิดแบบจำลองไฮบริด (Hybrid Model) ที่ยังคงใช้แบบจำลอง ARIMA พยากรณ์ข้อมูลที่มีโครงสร้างเชิงเส้นโดยเพิ่มองค์ประกอบเรื่องฤดูกาล (seasonal) เข้าไปในแบบจำลอง และเปลี่ยนจากแบบจำลอง ANN เป็นแบบจำลอง LSTM ที่ยังคงแนวคิดเครือข่ายประสาทเทียม (Neural networks) ผลการทดลองพบว่า มีเพียงรากที่สองของค่าเฉลี่ยความผิดพลาดลอการิทึมกำลังสอง (Root Mean Squared Logarithmic Error: RMSLE) จากแบบจำลอง ARIMA เท่านั้นที่ผลการพยากรณ์ดีกว่าค่าของแบบจำลอง LSTM และแบบจำลองไฮบริด นอกนั้นค่าเฉลี่ยสเกลความคลาดเคลื่อนสัมบูรณ์ (Mean Absolute Scaled Error: MASE) และค่าเฉลี่ยของเปอร์เซ็นต์ความผิดพลาดสัมบูรณ์ (Mean Absolute Percentage Error : MAPE) แบบจำลองไฮบริดได้ค่าเฉลี่ยความผิดพลาดน้อยที่สุดเมื่อเปรียบเทียบกับผลการพยากรณ์ด้วยแบบจำลอง ARIMA หรือแบบจำลอง LSTM แบบจำลองเดียว

งานวิจัยเรื่อง “Multi-feature PM2.5 Prediction with ARIMA-LSTM” (Xiao, Wang, Cui & Yu, 2022) ทำการวิจัยเพื่อพยากรณ์ข้อมูลค่าฝุ่น PM2.5 โดยใช้ข้อมูลรายชั่วโมงจากสถานที่แห่งหนึ่ง ผลการทดลองพบว่า รากที่สองของค่าเฉลี่ยความผิดพลาดกำลังสอง (Root Mean Squared Error: RMSE) และรากที่สองของค่าเฉลี่ยความผิดพลาดเปอร์เซ็นต์กำลังสอง (Root Mean Squared Percent Error: RMSPE) ของแบบจำลองไฮบริด ได้ค่าเฉลี่ยความผิดพลาดน้อยที่สุด จึงสรุปได้ว่าแบบจำลองไฮบริดสามารถให้ค่าพยากรณ์ที่ถูกต้องแม่นยำกว่าเมื่อเปรียบเทียบกับผลการพยากรณ์ด้วยแบบจำลอง LSTM อย่างเดียว

### 3. วิธีดำเนินการวิจัย

#### 3.1 กลุ่มตัวอย่าง

สำนักข่าวผู้จัดการออนไลน์ (MRG Online, 2566) ได้เสนอข่าวในวันที่ 10 มกราคม พ.ศ.2566 ที่แสดงให้เห็นว่าในช่วงเวลาระหว่าง 05.00-07.00 มี 19 เขตพื้นที่ ที่มีค่าฝุ่นละออง PM2.5 เกินเกณฑ์มาตรฐาน ซึ่งถือว่าอยู่ในระดับที่เริ่มส่งผลกระทบต่อสุขภาพ ประกอบไปด้วย 1) เขตคลองสามวา 2) เขตปทุมวัน 3) เขตธนบุรี 4) เขตวังทองหลาง 5) เขตหนองแขม 6) เขตลาดกระบัง 7) เขตคลองสาน 8) เขตบางซื่อ 9) เขตบางพลัด 10) เขตประเวศ 11) เขตยานนาวา 12) เขตสัมพันธวงศ์ 13) เขตบางกอกใหญ่ 14) เขตบึงกุ่ม 15) เขตหลักสี่ 16) เขตสาทร 17) เขตภาษีเจริญ 18) เขตหนองจอก 19) เขตบางขุนเทียน จากทั้งหมด 50 เขตทั่วกรุงเทพฯ

#### 3.2 จำนวนตัวอย่างและวิธีการเลือกตัวอย่าง

จากจำนวน 19 เขตทำการเลือกเขตปทุมวัน นอกจากจะเป็นเขตที่มีค่าฝุ่น PM2.5 เป็นอันดับหนึ่งในวันที่ 10 มกราคม พ.ศ.2566 แล้ว ข้อมูลจากกองจัดการคุณภาพอากาศและเสียง สังกัดกรุงเทพมหานคร (กองจัดการคุณภาพอากาศและเสียง, 2566) ยังมีความสมบูรณ์ของข้อมูลที่จะนำมาใช้ในการวิเคราะห์อีกด้วย สำหรับจำนวนข้อมูลที่ใช้ในการวิจัยสามารถพิจารณาได้ดังตารางที่ 1 และตัวแปรที่ใช้ในการวิจัยสามารถพิจารณาได้ดังตารางที่ 2

**ตารางที่ 1**

**รายละเอียดข้อมูลที่ใช้ในการวิจัย**

2,160 รายการ

ประเภทข้อมูล	ระยะเวลา	รายการ
ข้อมูลเรียนรู้ (train)	25 ธันวาคม พ.ศ. 2565 เวลา 0:00 น. ถึง 24 กุมภาพันธ์ พ.ศ. 2566 เวลา 23:00 น.	1,488
ข้อมูลทดสอบ (test)	25 กุมภาพันธ์ พ.ศ. 2566 เวลา 0:00 น. ถึง 24 มีนาคม พ.ศ. 2566 เวลา 23:00 น.	672

**ตารางที่ 2**

**ตัวแปรที่ใช้ในการวิจัย**

ลำดับ	ตัวแปร	คำอธิบาย	ประเภท
1.	Datetime	ปี/เดือน/วัน เวลา ที่ทำการวัดค่า	ตัวแปรอนุกรม
2.	WS (m/s)	ความเร็วลม (Wind Speed)	ตัวแปรต้น
3.	WD (Deg)	ทิศทางลม (Wind Direction)	ตัวแปรต้น
4.	Temp (°C)	อุณหภูมิองศาเซลเซียส (Temperature)	ตัวแปรต้น
5.	RH (%)	ความชื้นสัมพัทธ์ (Relative Humidity)	ตัวแปรต้น
6.	BP (mBar)	ความกดอากาศ (Barometric Pressure)	ตัวแปรต้น
7.	PM2.5 (ug/m3)	ฝุ่นขนาดเล็กที่มีขนาดไม่เกิน 2.5 ไมครอน (Particulate Matter)	ตัวแปรตาม

**3.3 การวิเคราะห์ข้อมูล**

1) พยากรณ์ค่าฝุ่นขนาดเล็กที่มีขนาดไม่เกิน 2.5 ไมครอน (PM2.5) ซึ่งเป็น PM2.5 ส่วนแรก ด้วยแบบจำลอง Arima โดยใช้เพียงตัวแปร  $PM_{2.5(t)}$  ได้ผลพยากรณ์ PM2.5 ส่วนแรก และส่วนที่เหลือ (Res)

2) พยากรณ์ค่าส่วนที่เหลือ (Res) ด้วยแบบจำลอง LSTM และ BiLSTM โดยมีรูปแบบดังนี้

2.1) พยากรณ์ค่าส่วนที่เหลือ โดยตัวแปรหลายตัวแปร (Multivariate) แบ่งออกเป็น 2 วิธี

2.1.1) ใช้ตัวแปรฝุ่นขนาดเล็กที่มีขนาดไม่เกิน 2.5 ไมครอน ( $PM_{2.5(t-1)}$ ) ความเร็วลม ( $WS_{(t-1)}$ )

ทิศทางลม ( $WD_{(t-1)}$ ) อุณหภูมิ ( $Temp_{(t-1)}$ ) ความชื้นสัมพัทธ์ ( $RH_{(t-1)}$ ) ความกดอากาศ ( $BP_{(t-1)}$ ) และส่วนที่เหลือ ( $Res_{(t-1)}$ ) โดยตัวแปรทุกตัวย้อนหลังหนึ่งช่วงเวลา

2.1.2) ใช้ตัวแปรเหมือนข้อ 2.1.1) โดยพิจารณาเลือกเฉพาะตัวแปรที่ไม่มีค่าสหสัมพันธ์ (Correlation)

ในช่วงเวลาปัจจุบัน (t) ทุกตัวกับค่าส่วนที่เหลือ

2.2) พยากรณ์โดยตัวแปรเดียว (Univariate) โดยตัวแปรส่วนที่เหลือ ( $Res_{(t-1)}$ ) ย้อนหลังหนึ่งช่วงเวลา

3) เมื่อได้ผลการพยากรณ์ทั้ง 2 ข้อแล้วนำผลลัพธ์มารวมกัน และนำไปวิเคราะห์ค่าความคลาดเคลื่อนต่างๆ ประกอบไปด้วย ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Error : MAE) ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Squared Error: RMSE) และ ค่าสัมบูรณ์ของเปอร์เซ็นต์ของความ

คลาดเคลื่อน (Mean Absolute Percentage Error: MAPE) ทุกค่าความคลาดเคลื่อนที่คำนวณได้ ยิ่งค่าที่ได้มีค่าน้อยแสดงว่าโมเดลที่ได้จะมีความแม่นยำมาก

#### 4 ผลการวิจัย

4.1 เมื่อทำการพยากรณ์เฉพาะแบบจำลอง Arima ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) เท่ากับ 4.3631 ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 5.6713 และ ค่าสัมบูรณ์ของเปอร์เซ็นต์ของความคลาดเคลื่อน (MAPE) เท่ากับ 0.0941

4.2 เมื่อทำการพยากรณ์แบบจำลอง Arima-LSTM และ Arima-BiLSTM ได้ผลการพยากรณ์ ดังตารางที่ 3

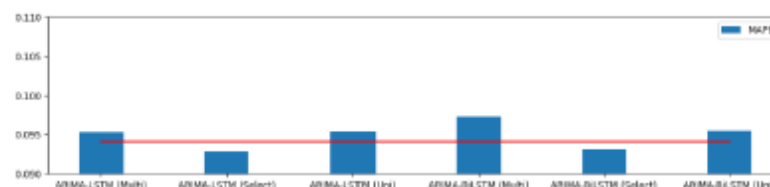
ตารางที่ 3

ผลการวิจัย

แบบจำลอง	รูปแบบ	MAE	RMSE	MAPE
ARIMA-LSTM	หลายตัวแปร	4.3440	5.5782	0.0953
	เลือกตัวแปร	4.3073	5.5122	0.0928
	ตัวแปรเดียว	4.4002	5.7013	0.0954
ARIMA-BiLSTM	หลายตัวแปร	4.6606	6.0001	0.0973
	เลือกตัวแปร	4.3544	5.5968	0.0931
	ตัวแปรเดียว	4.4045	5.7080	0.0955

ภาพที่ 4

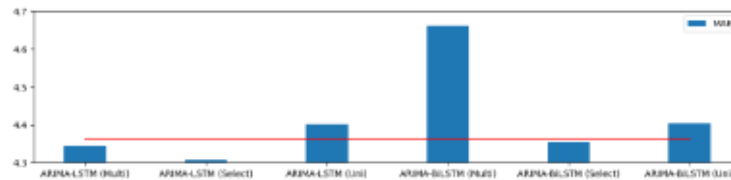
แสดงค่าสัมบูรณ์ของเปอร์เซ็นต์ของความคลาดเคลื่อน (MAPE)



เมื่อพิจารณาค่าสัมบูรณ์ของเปอร์เซ็นต์ของความคลาดเคลื่อน (MAPE) แบบจำลอง Arima-LSTM และ Arima-BiLSTM เปรียบเทียบกับแบบจำลอง Arima อย่างเดียว พบว่า ค่าสัมบูรณ์ของเปอร์เซ็นต์ของความคลาดเคลื่อน (MAPE) ของแบบจำลอง Arima-LSTM แบบเลือกตัวแปร เท่ากับ 0.0928 ค่าสัมบูรณ์ของเปอร์เซ็นต์ของความคลาดเคลื่อน (MAPE) ของแบบจำลอง Arima-BiLSTM แบบเลือกตัวแปร เท่ากับ 0.0931 ได้ค่าต่ำกว่าแบบจำลอง Arima อย่างเดียว ซึ่งค่าสัมบูรณ์ของเปอร์เซ็นต์ของความคลาดเคลื่อน (MAPE) เท่ากับ 0.0941

ภาพที่ 5

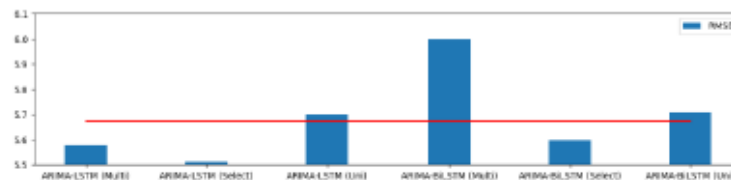
แสดงค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE)



เมื่อพิจารณาค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) แบบจำลอง Arima-LSTM และ Arima-BiLSTM เปรียบเทียบกับแบบจำลอง Arima อย่างเดียว พบว่า ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) ของแบบจำลอง Arima-LSTM แบบหลายตัวแปร เท่ากับ 4.3440 ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) ของแบบจำลอง Arima-LSTM แบบเลือกตัวแปร เท่ากับ 4.3073 ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) ของแบบจำลอง Arima-BiLSTM แบบเลือกตัวแปร เท่ากับ 4.3544 ได้ค่าต่ำกว่าแบบจำลอง Arima อย่างเดียว ซึ่งค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) เท่ากับ 4.3631

ภาพที่ 6

แสดงค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE)



เมื่อพิจารณาค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) แบบจำลอง Arima-LSTM และ Arima-BiLSTM เปรียบเทียบกับแบบจำลอง Arima อย่างเดียว พบว่า ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) ของแบบจำลอง Arima-LSTM แบบหลายตัวแปร เท่ากับ 5.5782 ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) ของแบบจำลอง Arima-LSTM แบบเลือกตัวแปร เท่ากับ 5.5122 ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) ของแบบจำลอง Arima-BiLSTM แบบเลือกตัวแปร เท่ากับ 5.5968 ได้ค่าต่ำกว่าแบบจำลอง Arima อย่างเดียว ซึ่งค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 5.6713

## 5 สรุป

### 5.1 ผลการวิจัยโดยรวม

เมื่อพิจารณาค่าสัมบูรณ์ของเปอร์เซ็นต์ของความคลาดเคลื่อน (MAPE) ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) ของแบบจำลอง Arima-LSTM แบบตัวแปรเดียว แบบจำลอง Arima-BiLSTM แบบตัวแปรเดียว แบบจำลอง Arima-LSTM แบบหลายตัวแปร แบบจำลอง Arima-BiLSTM แบบหลายตัวแปร แบบจำลอง Arima-LSTM แบบเลือกตัวแปร แบบจำลอง Arima-BiLSTM แบบเลือกตัวแปร พบว่าแบบจำลองที่มีการเลือกตัวแปร ทั้งแบบจำลอง Arima-LSTM และ แบบจำลอง Arima-BiLSTM ให้ค่าความคลาดเคลื่อนต่ำกว่า แบบจำลอง Arima อย่างเดียวเนื่องจาก กระบวนการเลือกข้อมูล (Data Selection) สามารถเพิ่มประสิทธิภาพของโมเดล (Improves accuracy) เนื่องจากการข้อมูลรบกวน (Noise) น้อยลง จึงส่งผลให้โมเดลมีประสิทธิภาพ การวิเคราะห์หรือพยากรณ์ที่ถูกต้องมากยิ่งขึ้น ซึ่งแบบจำลอง Arima-LSTM แบบเลือกตัวแปร ให้ประสิทธิภาพได้ดีที่สุด รองลงมาคือ แบบจำลอง Arima-BiLSTM แบบเลือกตัวแปร

### 5.2 ข้อเสนอแนะสำหรับการวิจัยในครั้งต่อไป

1. กรณีแบบจำลอง Arima-LSTM แบบหลายตัวแปร สามารถนำความรู้เกี่ยวกับการจำแนกประเภทแบบหลายคลาส (multi-label classification) มาประยุกต์ใช้ ในการจัดลำดับตัวแปร แล้วพยากรณ์ใหม่อีกครั้ง อาจทำให้ผลการพยากรณ์ที่ดีขึ้น
2. กรณีแบบจำลอง Arima-LSTM แบบตัวแปรเดียว สามารถนำตัวแปรค่าส่วนที่เหลือย้อนหลังหนึ่งช่วงเวลา (t-1) ทำนายร่วมกับตัวแปรที่เหลือที่ไม่มีค่าสหสัมพันธ์กับค่าส่วนที่เหลือ ณ เวลาปัจจุบัน (t) เพื่อทดสอบแบบจำลองตัวแปรคู่ (bivariate variable) แล้วพยากรณ์ใหม่อีกครั้ง อาจทำให้ผลการพยากรณ์ที่ดีขึ้น
3. สามารถทดลองแบบจำลองที่เลือกตัวแปรได้โดยใช้เทคนิคอื่น ๆ

## 6 การอ้างอิง

- กรมควบคุมมลพิษ. (2565). กค.วล. ยกระดับค่ามาตรฐาน PM2.5 ราชกิจจานุเบกษาบังคับใช้แล้ว. ค้นหาค้นหาเมื่อ 25 มีนาคม 2566. จาก [https://www.pcd.go.th/pcd\\_news/26385](https://www.pcd.go.th/pcd_news/26385).
- กระทรวงสาธารณสุข. (2562). คู่มือการดำเนินงานด้านการแพทย์และสาธารณสุข กรณี ฝุ่นละอองขนาดเล็กเกิน 2.5 ไมครอน (PM2.5) ปี 2563. หน้า 1-7.
- กองจัดการคุณภาพอากาศและเสียง. (2566). รายงาน. ค้นหาค้นหาเมื่อ 25 มีนาคม 2566. จาก <https://bangkokairquality.com/bma/report?lang=th>
- Dagshub. (2023). Complete Guide to RNN, LSTM, and Bidirectional LSTM. ค้นหาค้นหาเมื่อ 25 มีนาคม 2566. จาก <https://dagshub.com/blog/rnn-lstm-bidirectional-lstm/>
- DiBattista, J. (2022). Choosing the Best ML Time Series Model for Your Data. ค้นหาค้นหาเมื่อ 25 มีนาคม 2566. จาก <https://towardsdatascience.com/choosing-the-best-ml-time-series-model-for-your-data-664a7062f418>

Fathi, O. (2019). Time series forecasting using a hybrid ARIMA and LSTM model. France: Velvet Consulting.

Maslow, A. H. (1943). A theory of human motivation, *Psychological Review*, 50(4), pp. 370–396.

MRG Online. (2566). เข้านี้ กทม. ค่าฝุ่น PM 2.5 เกินมาตรฐาน 19 พื้นที่ อยู่ในระดับเริ่มมีผลกระทบต่อสุขภาพ. ค้นหามือ 25 มีนาคม 2566. จาก <https://mgronline.com/qpl/detail/9660000002510>.

Shaun S. W. (2017). Time Series Analysis: Forecasting and Control, *Journal of Quality Technology*. 49(5), 418-419.

Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model, *Neurocomputing*, 50, pp. 159 – 175.

Xiao, J., Wang, Q., Cui, J. & Yu, J. (2022). Multi-feature PM2.5 Prediction with ARIMA-LSTM. International Conference on Machine Learning, Cloud Computing and Intelligent Mining.

### ประวัติผู้เขียน

ชื่อ - นามสกุล

ธัชกร พลเสนา

#### ประวัติการศึกษา

พ.ศ. 2556

ปริญญาโท วิทยาศาสตร์มหาบัณฑิต สาขาจิตวิทยาการปรึกษา  
มหาวิทยาลัยรามคำแหง

พ.ศ. 2549

ปริญญาโท เศรษฐศาสตรมหาบัณฑิต มหาวิทยาลัยรามคำแหง

พ.ศ. 2544

ปริญญาตรี วิทยาศาสตร์บัณฑิต สาขาเทคโนโลยีสารสนเทศ  
เกียรตินิยมอันดับสอง มหาวิทยาลัยหัวเฉียวเฉลิมพระเกียรติ

#### ประสบการณ์ทำงาน

พ.ศ. 2566

System Analyst ธนาคารทหารไทยธนชาติ จำกัด (มหาชน)

พ.ศ. 2565

Team Lead of Information Planning ธนาคารมิซูโฮ จำกัด  
สาขา กรุงเทพ