

รายงานโครงงานวิทยาการคอมพิวเตอร์

การพยากรณ์ปริมาณฝุ่นละอองขนาดเล็กในอากาศ (PM2.5)

ในพื้นที่กรุงเทพมหานคร ด้วยการเรียนรู้ของเครื่อง

Forecasting the Amount of Small Dust

in the Air (PM 2.5) with the Machine Learning

ผู้จัดทำ

นางสาวณัฏฐพัชร์ ปุณณหิรัญย์ 6210450806

นางสาวนิภาวรรณ ประสพดี 62104512184

อาจารย์ที่ปรึกษาโครงงาน

ผศ.ดร.อรรณณ อิมสมบัติ

โครงงานวิทยาการคอมพิวเตอร์ (01418499)

ภาควิชาการคอมพิวเตอร์ คณะวิทยาศาสตร์

มหาวิทยาลัยเกษตรศาสตร์ ภาควิชา ศึกษาศาสตร์ 2565

โครงการวิทยการคอมพิวเตอร์

หัวข้อ การพยากรณ์ปริมาณฝุ่นละอองขนาดเล็กในอากาศ (PM2.5) ในพื้นที่กรุงเทพมหานคร
ด้วยการเรียนรู้ของเครื่อง

โดย นางสาวณัฏฐพัชร์ ปุณณศิริณย์ 6210450806
นางสาวนิภาวรรณ ประสพดี 6210451284

อาจารย์ที่ปรึกษาหลัก



(ผศ.ดร.อรรวรรณ อิมสมบัติ)

20 / 4 / 66

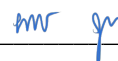
อาจารย์ที่ปรึกษาร่วม



(ดร.พิศุทธิ์ วิเศษสิงห์)

20 / 4 / 66

หัวหน้าภาควิทยาการคอมพิวเตอร์




(ผศ.ดร.ผกาเกษ วัฒนา)

20 / 04 / 66

ชื่อหัวข้อโครงการ การพยากรณ์ปริมาณฝุ่นละอองขนาดเล็กในอากาศ (PM2.5) ในพื้นที่กรุงเทพมหานคร
ด้วยการเรียนรู้ของเครื่อง

โดย นางสาวณัฏฐพัชร ปุณณศิริชัย 6210450806
 นางสาวนิภาวรรณ ประสพดี 6210451284

คำสำคัญ Long short-term memory, Light Gradient Boosting Machine, XGBoost,
Prophets, PM 2.5, Forecasting

อาจารย์ที่ปรึกษาหลัก  _____
(Asst. Prof. Dr. Aurawan imsombut)

บทคัดย่อ

อันตรายจากฝุ่นละอองขนาดเล็กในอากาศ (PM 2.5) มีได้จากหลายสาเหตุ เช่น ไอเสียจากรถยนต์ หรือจากการจราจร การปล่อยอากาศพิษจากโรงงานอุตสาหกรรมและโรงไฟฟ้า การเผาไหม้และวัสดุต่างๆ เป็นต้น ซึ่งสาเหตุที่กล่าวมานี้ ล้วนส่งผลต่อค่าฝุ่นที่มีแนวโน้มที่เพิ่มขึ้นเรื่อยๆของประเทศไทยอย่างมาก และโดยเฉพาะเขตพื้นที่กรุงเทพมหานคร ที่มีจำนวนผู้อาศัยเยอะ และแออัด ทำให้เกิดค่าฝุ่นที่พบอยู่ได้ตลอด ในช่วงของฤดูหนาว โครงการการพยากรณ์ปริมาณฝุ่นละอองขนาดเล็กในอากาศ (PM2.5) ในพื้นที่ กรุงเทพมหานครนี้จึงพัฒนาระบบที่คาดการณ์ค่าฝุ่นละอองขนาดเล็กนี้ โดยพิจารณาจากปัจจัยต่างๆ ได้แก่ ฝุ่นละออง PM2.5 ในอดีต, CO (คาร์บอนมอนอกไซด์), NO₂(ไนโตรเจนออกไซด์), NO (ไนตริกออกไซด์), NO_x (ไนโตรเจนออกไซด์), O₃ (โอโซน), ฝุ่นละออง PM10 ในอดีต, Wind speed (ความเร็วลม), Wind direction (ทิศทางลม, Average Temperature (อุณหภูมิเฉลี่ย), humidity (ค่าความชื้นสัมพัทธ์ในอากาศ), Pressure (ความกดอากาศ) โดยมีการศึกษาเปรียบเทียบการเรียนรู้ของเครื่องต่างๆ ได้แก่ Long short-term memory (LSTM), Prophet, XGBoost และ Light Gradient Boosting Machine (LightGBM)

การประเมินผลของการเรียนรู้ของเครื่อง ของโครงการการพยากรณ์ปริมาณฝุ่นละอองขนาดเล็กในอากาศ (PM2.5) ในพื้นที่กรุงเทพมหานคร พบว่าแบบจำลอง XGBoost มีค่า RMSE เป็น 0.05931 และค่า MSE เป็น 0.00352 จึงเป็นแบบจำลองที่มีความแม่นยำ ที่สามารถนำมาใช้เพื่อคาดการณ์ค่าฝุ่นล่วงหน้าที่ดีที่สุด

Title Forecasting the Amount of Small Dust in the Air (PM 2.5) with the Machine Learning

By Natthapat Punnahiran ID 6210450806
 Nipawan Prasopdee ID 6210451284

Keywords Long short-term memory, Light Gradient Boosting Machine, XGBoost, Prophets, PM 2.5, Forecasting

Advisor Ms. Aurawan Imsombut

Abstract

This study focuses on forecasting the amount of small dust in the air (PM 2.5) using machine learning techniques. Air pollution caused by PM 2.5 has become a severe issue in many parts of the world. In this research, we propose a novel approach to predict PM 2.5 concentrations using machine learning algorithms such Long short-term memory, Light Gradient Boosting Machine, XGBoost and Prophets. We collected PM 2.5 data from various monitoring stations in a specific geographic region over a period of several years. We then used this data to train and evaluate our machine learning models. Our results demonstrate that the proposed approach can provide accurate predictions of PM 2.5 concentrations, with a high degree of accuracy and reliability. This research has significant implications for air quality management and public health policies, as it can help in the early detection of high PM 2.5 concentrations and aid in the development of effective mitigation strategies.

Assessment of machine learning results of the small dust forecast project in the air (PM2.5) in Bangkok found that the XGBoost model has a RMSE value of 0.05931 and the MSE value is 0.00352, so it is a precision model that can be used to predict the best dust value in advance.

คำนำ

เอกสารฉบับนี้เป็นเอกสารประกอบโครงการวิทยการคอมพิวเตอร์ในหัวข้อ **โครงการพยากรณ์ปริมาณฝุ่นละอองขนาดเล็กในอากาศ (PM2.5) ในพื้นที่กรุงเทพมหานคร** จัดทำขึ้นเพื่ออธิบายโครงสร้างขั้นตอน การดำเนินงาน และเทคโนโลยีที่ใช้ในการพัฒนาโครงการ ไปจนถึงทดสอบแบบจำลองของระบบ ให้มีความผิดพลาดน้อยที่สุด โดยโครงการนี้มีจุดประสงค์ เพื่อให้เปรียบเทียบจากแบบจำลองที่จะนำมาใช้คาดการณ์ขนาดอนุภาคฝุ่นละอองขนาดเล็ก (PM 2.5) ในอากาศล่วงหน้าเพื่อให้เกิดความแม่นยำที่สุด

คณะผู้จัดทำ

นางสาวณัฐพัชร์ ปุณณศิริชัย รหัสนิสิต 6210450806

นางสาวนิภาวรรณ ประสพดี รหัสนิสิต 6210451284

กิตติกรรมประกาศ

จากการทำโครงการการพยากรณ์ปริมาณฝุ่นละอองขนาดเล็กในอากาศ (PM2.5) ในพื้นที่กรุงเทพมหานคร ทางคณะผู้จัดทำสามารถทำโครงการประสบความสำเร็จลุล่วงไปได้ด้วยดี ทั้งนี้ระหว่างการพัฒนาโครงการ ได้รับคำแนะนำ และชี้แนะแนวทางต่าง ๆ จึงขอขอบคุณอาจารย์ และแหล่งข้อมูล ต่อไปนี้

ขอขอบคุณอาจารย์ ผศ.ดร. อรรวรรณ อิ่มสมบัติ อาจารย์ที่ปรึกษาโครงการ ซึ่งเป็นผู้ให้คำปรึกษา ชี้แนะเกี่ยวกับการพัฒนาโครงการ การวางแผนแต่ละอาทิตย์ และการทำเรื่องขอข้อมูลปัจจัยภูมิอากาศ กับมลพิษทางอากาศ

ขอขอบคุณผู้ให้ข้อมูลปัจจัยในการนำมาพัฒนาแบบจำลอง ได้แก่ กรมอุตุนิยมวิทยา กรมควบคุมมลพิษ และสำนักสิ่งแวดล้อม ซึ่งทางคณะผู้จัดทำนั้น ขอขอบคุณเป็นอย่างสูงยิ่ง

และขอขอบคุณอาจารย์ที่ปรึกษาร่วม คือ ดร.พิศุทธิ์ วิเศษสิงห์ ที่ให้คำแนะนำกับการจัดทำโครงการ ให้สำเร็จไปได้ด้วยดี

คณะผู้จัดทำ

นางสาวณัฏฐพัชร์ ปุณณหิรัญย์ รหัสนิสิต 6210450806

นางสาวนิภาวรรณ ประสพดี รหัสนิสิต 6210451284

สารบัญ

โครงการวิทยาการคอมพิวเตอร์.....	ก
บทคัดย่อ.....	ข
Abstract.....	ค
คำนำ.....	ค
กิตติกรรมประกาศ.....	จ
สารบัญ.....	ฉ
สารบัญ(ต่อ).....	ช
สารบัญรูปภาพ.....	ช
สารบัญรูปภาพ(ต่อ).....	ฅ
สารบัญตาราง.....	ญ
บทที่ 1.....	1
ที่มาและความสำคัญ.....	1
วัตถุประสงค์.....	1
แนวทางการพัฒนา.....	1
ขอบเขตของโครงการ.....	2
ทรัพยากรที่ใช้ในการดำเนินโครงการ.....	2
ประโยชน์ที่คาดว่าจะได้รับ.....	2
แผนการดำเนินงาน.....	3
บทที่ 2.....	4
ทฤษฎีที่เกี่ยวข้อง.....	4
งานวิจัยที่เกี่ยวข้อง.....	8
บทที่ 3.....	11
ศึกษาและรวบรวมข้อมูล.....	11
ประมวลผลข้อมูล.....	12
วิเคราะห์ข้อมูล.....	13
ประเมินผลประสิทธิภาพของแบบจำลอง.....	14
การออกแบบแบบจำลอง.....	16

สารบัญ (ต่อ)

บทที่ 4.....	17
ผลลัพธ์ประสิทธิภาพของแบบจำลองในเขตราชเทวี.....	17
ผลลัพธ์ประสิทธิภาพของแบบจำลองในเขตพระโขนง.....	18
ผลลัพธ์ประสิทธิภาพของแบบจำลองในเขตดินแดง.....	20
ผลลัพธ์ประสิทธิภาพของแบบจำลองในการเฉลี่ยทุกเขต.....	21
บทที่ 5.....	33
สรุปผลการวิจัย.....	33
เปรียบเทียบผลกับงานวิจัย LSTM.....	34
เปรียบเทียบผลกับงานวิจัย LightGBM.....	34
อภิปรายผล.....	35
ข้อเสนอแนะ.....	35
บรรณานุกรม.....	36

สารบัญรูปภาพ

รูปที่ 2.1 แสดงโครงสร้างของ RNN.....	4
รูปที่ 2.2 ภาพโครงสร้างต้นไม้หลายๆต้นในการเรียนรู้.....	5
รูปที่ 2.3 ขั้นตอนการทำงานของ Prophet.....	6
รูปที่ 2.4 ขั้นตอนการทำงานของ XGBoost.....	7
รูปที่ 2.5 ขั้นตอนการ Pre-processing data.....	8
รูปที่ 2.6 ขั้นตอนการดำเนินการวิจัย.....	9
รูปที่ 2.7 ตัวแปรในแบบจำลอง และผลเปรียบเทียบความแม่นยำ.....	10
ของแบบจำลองพยากรณ์	
รูปที่ 3.1 การทำงานของของแบบจำลอง LSTM, LightGBM, Prophet, XGBoost.....	13
รูปที่ 3.2 System Architecture.....	15
รูปที่ 4.1 กราฟข้อมูล Train เทียบกับ Actual ทุกแอลกอริทึม เขตราชเทวี.....	18
รูปที่ 4.2 กราฟข้อมูล Train เทียบกับ Actual ทุกแอลกอริทึม เขตพระโขนง.....	20
รูปที่ 4.3 กราฟข้อมูล Train เทียบกับ Actual ทุกแอลกอริทึม เขตดินแดง.....	21
รูปที่ 4.4 กราฟข้อมูล Train เทียบกับ Actual ทุกแอลกอริทึม รวมทุกเขต.....	23
รูปที่ 4.5 กราฟเปรียบเทียบค่าความแม่นยำของเขตราชเทวี.....	24
รูปที่ 4.6 กราฟเปรียบเทียบค่าความแม่นยำของเขตพระโขนง.....	24
รูปที่ 4.7 กราฟเปรียบเทียบค่าความแม่นยำของเขตดินแดง.....	25
รูปที่ 4.8 กราฟเปรียบเทียบค่าความแม่นยำของเฉลี่ยรวมทุกเขต.....	25
รูปที่ 4.9 ผลการทำนายค่าฝุ่นละอองล่วงหน้า 7 วัน ด้วยอัลกอริทึม LSTM เขตราชเทวี.....	26
รูปที่ 4.10 ผลการทำนายค่าฝุ่นละอองล่วงหน้า 7 วัน ด้วยอัลกอริทึม LSTM พระโขนง.....	26
รูปที่ 4.11 ผลการทำนายค่าฝุ่นละอองล่วงหน้า 7 วัน ด้วยอัลกอริทึม LSTM ดินแดง.....	26
รูปที่ 4.12 ผลการทำนายค่าฝุ่นละอองล่วงหน้า 7 วัน ด้วยอัลกอริทึม LSTM เฉลี่ยรวมทุกเขต.....	27
รูปที่ 4.13 ผลการทำนายค่าฝุ่นละอองล่วงหน้า 7 วัน ด้วยอัลกอริทึม LightGBM ราชเทวี.....	27
รูปที่ 4.14 ผลการทำนายค่าฝุ่นละอองล่วงหน้า 7 วัน ด้วยอัลกอริทึม LightGBM พระโขนง.....	27
รูปที่ 4.15 ผลการทำนายค่าฝุ่นละอองล่วงหน้า 7 วัน ด้วยอัลกอริทึม LightGBM ดินแดง.....	28
รูปที่ 4.16 ผลการทำนายค่าฝุ่นละอองล่วงหน้า 7 วัน ด้วยอัลกอริทึม LightGBM เฉลี่ยรวมทุกเขต.....	28
รูปที่ 4.17 ผลการทำนายค่าฝุ่นละอองล่วงหน้า 7 วัน ด้วยอัลกอริทึม Prophet ราชเทวี.....	29

สารบัญรูปภาพ (ต่อ)

รูปที่ 4.18 ผลการทำนายค่าฝุ่นละอองล่วงหน้า 7 วัน ด้วยอัลกอริทึม Prophet พระโขนง.....	29
รูปที่ 4.19 ผลการทำนายค่าฝุ่นละอองล่วงหน้า 7 วัน ด้วยอัลกอริทึม Prophet ดินแดง.....	30
รูปที่ 4.20 ผลการทำนายค่าฝุ่นละอองล่วงหน้า 7 วัน ด้วยอัลกอริทึม Prophet เฉลี่ยรวมทุกเขต.....	30
รูปที่ 4.21 ผลการทำนายค่าฝุ่นละอองล่วงหน้า 7 วัน ด้วยอัลกอริทึม XGBoost ราชเทวี.....	31
รูปที่ 4.22 ผลการทำนายค่าฝุ่นละอองล่วงหน้า 7 วัน ด้วยอัลกอริทึม XGBoost พระโขนง.....	31
รูปที่ 4.23 ผลการทำนายค่าฝุ่นละอองล่วงหน้า 7 วัน ด้วยอัลกอริทึม XGBoost ดินแดง.....	31
รูปที่ 4.24 ผลการทำนายค่าฝุ่นละอองล่วงหน้า 7 วัน ด้วยอัลกอริทึม XGBoost เฉลี่ยรวมทุกเขต.....	32

สารบัญตาราง

ตารางที่ 1.1 ตารางแผนการดำเนินงานของโครงการวิทยาการคอมพิวเตอร์.....	3
ตารางที่ 3.1 แสดงรายละเอียดของปัจจัยที่มีความสัมพันธ์กับค่าฝุ่น PM 2.5 และ feature.....	11
ตารางที่ 3.2 data set ที่พร้อมนำไปเข้า model.....	12
ตารางที่ 4.1. แสดงผลลัพธ์ของการวัดประสิทธิภาพของแต่ละแบบจำลอง.....	23
ในข้อมูลของเขตราชเทวี เขตพระโขนง เขตดินแดง และการเฉลี่ยรวมทุกเขต	
ตารางที่ 5.1 เปรียบเทียบงานวิจัยนี้กับงานวิจัยอื่นของ LSTM.....	34
ตารางที่ 5.2 เปรียบเทียบงานวิจัยนี้กับงานวิจัยอื่นของ LighGBM.....	34

บทที่ 1

ที่มา และความสำคัญ

1.1. ที่มาและความสำคัญ

ฝุ่นละอองขนาดเล็กในอากาศ (PM2.5) ซึ่ง PM2.5 มีชื่อเต็มว่า Particulate matter with diameter of less than 2.5 micron โดยเป็นฝุ่นละอองขนาดเล็กซึ่งไม่เกิน 2.5 ไมครอน ทำให้เป็นปัญหาทางมลพิษทางอากาศที่พบได้ทั่วทุกมุมโลก เนื่องจากแต่ละประเทศมีการพัฒนาด้านอุตสาหกรรมที่รวดเร็ว และปัจจัยในด้านอื่นๆ ทำให้เกิดมลพิษต่อสิ่งแวดล้อม และผู้คน

ในด้านสุขภาพที่เป็นปัจจัยหลักในการดำเนินชีวิต ฝุ่นละอองขนาดเล็กนี้ มีผลอย่างมากต่อระบบหายใจและปอดโดยตรง เกิดการระคายเคือง แสบจูก ไอ เป็นต้น และในระดับที่ร้ายแรง จะส่งผลให้เกิดโรคหัวใจวายเฉียบพลัน หลอดเลือดสมองตีบ และอันตรายมากที่สุด ก็คือ มะเร็งปอด

เนื่องจากกรุงเทพมหานคร เป็นพื้นที่ที่มีผู้คนอาศัยจำนวนมาก และยังมีค่าฝุ่นละอองขนาดเล็กที่สูงขึ้นเรื่อยๆ และสูงกว่าจังหวัดอื่นๆโดยรอบ การคาดการณ์แนวโน้มมลพิษของฝุ่นของวันถัดไป จะช่วยแจ้งเตือนประชาชนล่วงหน้าให้ป้องกันตนเองได้ โครงการนี้จึงพัฒนาระบบที่คาดการณ์ค่าฝุ่นละอองขนาดเล็ก โดยพิจารณาจากปัจจัยต่างๆได้แก่ ฝุ่นละออง PM2.5, CO (คาร์บอนไดออกไซด์), NO_2 (ไนโตรเจนไดออกไซด์), NO (ไนตริกออกไซด์), NO_x (ไนโตรเจนออกไซด์), O_3 (โอโซน), ฝุ่นละออง PM10, Wind speed (ความเร็วลม), Wind direction (ทิศทางลม), Average Temperature (อุณหภูมิเฉลี่ย), humidity (ค่าความชื้นสัมพัทธ์ในอากาศ), Pressure (ความกดอากาศ) ในการเปรียบเทียบการเรียนรู้ของเครื่องต่างๆ ได้แก่ Long short-term memory (LSTM), Prophet, XGBoost และ Light Gradient Boosting Machine (LightGBM) เพื่อให้ได้ผลทำนายที่มีค่าแม่นยำมากที่สุด

1.2 วัตถุประสงค์

1. เพื่อศึกษาแอลกอริทึมในการทำนายค่าฝุ่น PM2.5 จากปัจจัยที่เกี่ยวข้อง
2. เพื่อศึกษากราฟว่ามีแนวโน้มเพิ่มหรือลดลงอย่างไร

1.3 แนวทางการพัฒนาโปรแกรม

1. ศึกษา และรวบรวมข้อมูลตัวแปรค่าสภาพอากาศ และค่าฝุ่น มลพิษต่างๆ
2. ศึกษา และพัฒนาแอลกอริทึมในการทำนายค่าฝุ่น PM 2.5
3. วิเคราะห์และประเมินผล
4. พัฒนา และตรวจสอบระบบ แก้ไขข้อผิดพลาด ให้มีความถูกต้องและตรงตามวัตถุประสงค์

1.4 ขอบเขตของโครงการงาน

ใช้ Machine Learning ในการคาดการณ์ค่าฝุ่น PM2.5 ข้อมูลมลพิษทางอากาศ ได้แก่ PM10 , O_3 , NO_2 , NO , NO_x , CO , SO_2 โดยมีข้อมูลย้อนหลังของทุกวันเป็นเวลา 5 ปี โดยใช้เวลาในการเทรน 3 ปี และใช้เวลาทดสอบ 2 ปี และใช้ Machine Learning คาดการณ์ ค่าฝุ่น PM 2.5 ของปีหน้าออกมาเป็นรายวันล่วงหน้า 7 วัน โดยใช้แอลกอริทึม 4 ตัวมาเปรียบเทียบความแม่นยำคือ Long short-term memory (LSTM) , Light Gradient Boosting Machine (Light GBM) , Prophet และ XGBoost

1.5 ทรัพยากรที่ใช้ในการดำเนินโครงการงาน

1.5.1 ด้านซอฟต์แวร์ (Software)

ไลบรารีที่ใช้ ได้แก่ sklearn, numpy, pandas, matplotlib, tensorflow และ fbprophet

1.5.2.ด้านฮาร์ดแวร์ (Hardware)

- แล็ปท็อปคอมพิวเตอร์ (Laptop computer) โดยมีสเปคเป็น Windows11 และมี Processor เป็น Intel(R) Core(TM) i5-8265U CPU @ 1.60GHz 1.80 GHz และ RAM ใช้เป็น 8.00 GB

1.6 ประโยชน์ที่คาดว่าจะได้รับ

สามารถรู้ค่าการทำนายฝุ่นละอองขนาดเล็กในวันถัดไปที่จะมาถึงได้ และฝึกฝนเรียนรู้วิธีการเขียน แอลกอริทึมเพื่อนำมาเปรียบเทียบประสิทธิภาพได้

1.7 แผนการดำเนินงาน

ตารางที่ 1.1 ตารางแผนการดำเนินงานของโครงการวิทยาการคอมพิวเตอร์

กิจกรรม	ระยะเวลาวางแผนการดำเนินงานและการดำเนินงาน			
	ธ.ค.	ม.ค.	ก.พ.	มี.ค.
ค้นคว้าหัวข้อโครงการ	/			
กำหนดขอบเขตโครงการ	/			
ศึกษาข้อมูลและเครื่องมือที่ใช้	/	/		
พัฒนาโมเดล		/	/	
ทดสอบ, แก้ไขข้อผิดพลาด				/
จัดทำเล่มโครงการ				/

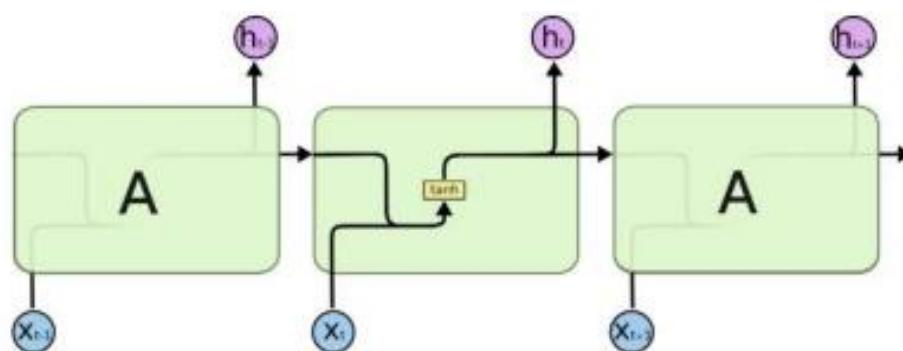
บทที่ 2

ทฤษฎี และงานวิจัยที่เกี่ยวข้อง

เนื้อหาภายในบทที่ 2 นี้กล่าวถึงทฤษฎีที่เกี่ยวข้องในการพัฒนาโครงงานวิทยาการคอมพิวเตอร์ฉบับนี้ เพื่อเพิ่มความเข้าใจในเทคโนโลยีที่ผู้จัดทำได้เลือกเข้ามาใช้ในการทำงาน ดังนี้

2.1 ทฤษฎีที่เกี่ยวข้อง

2.1.1. Long short-term memory (LSTM)

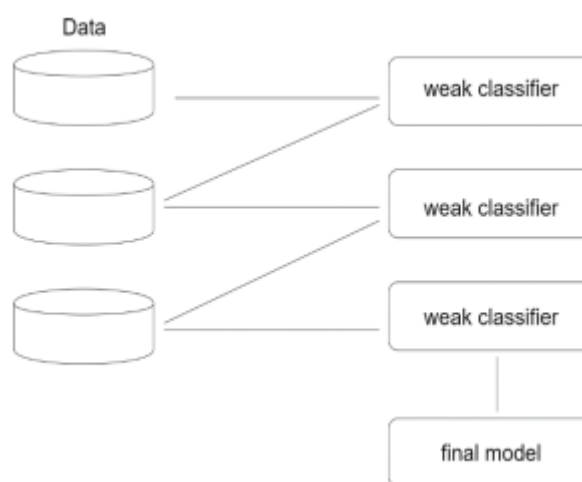


รูปที่ 2.1 แสดงโครงสร้างของ RNN

ที่มา : Olah, C. (2015). Understanding LSTM Networks -- colah's blog. Retrieved from <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>

LSTM [4] นั้นจัดว่าเป็นโครงข่ายประเภท Recurrent Neural Network (RNN) และเหมาะกับการประมวลผลลำดับ (sequence) ข้อดีของ RNN คือ สามารถนำข้อมูลในอดีต มาใช้ในการทำนายสิ่งที่จะเกิดขึ้นในอนาคตได้ ส่วนข้อเสียของ RNN คือ จะสามารถดูข้อมูลย้อนหลังได้แค่เพียงระยะสั้นๆ เท่านั้น โดยมีหลักการทำงาน คือ เก็บ 'สถานะ' แต่ละโหนดเอาไว้เพื่อที่เวลาย้อนกลับไปดูจะได้ทราบถึงที่มาของข้อมูลค่าดังกล่าวว่า ค่าเก่าเป็นค่าอะไร และยังมีฟังก์ชันพิเศษที่เรียกว่า Gate จะเป็นตัวคอยควบคุมข้อมูลที่จะเข้ามาในแต่ละโหนด ซึ่งประกอบไปด้วย Forget gate layer, Input gate layer และ Output gate layer

2.1.2. Light Gradient Boosting Machine (Light GBM)

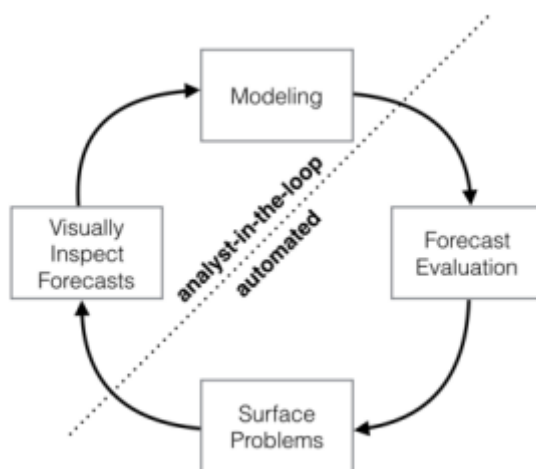


รูปที่ 2.2 ภาพโครงสร้างต้นไม้หลายๆต้นในการเรียนรู้

ที่มา : ชนะวิชญ์ พชรเจริญวงศ์, กันต์กวี เหิรเมฆ, วราภคณา กิมปาน. (2563). แบบจำลองการทำนาย
ระยะเวลาในการเข้าเทียบท่าของเรือโดยสารสาธารณะ จาก [https://li01.tci-](https://li01.tci-thaijo.org/index.php/science_kmitl/article/download/241105/169801/862300)
[thaijo.org/index.php/science_kmitl/article/download/241105/169801/862300](https://li01.tci-thaijo.org/index.php/science_kmitl/article/download/241105/169801/862300)

LightGBM [5] เป็นโครงสร้างแบบต้นไม้หลาย ๆ ต้น (trees) โดย Gradient Boosting เป็นเทคนิคการเรียนรู้ของเครื่องสำหรับแก้ปัญหาการถดถอย (Regression) และการจำแนกประเภท (Classification) ดังนั้น GBM จะสร้างโครงสร้างการถดถอยตามลำดับ ซึ่ง GBM ใช้เทคนิคการเพิ่มการรวมจำนวน classifier ที่มีความแม่นยำต่ำ เพื่อสร้างเป็น classifier ใหม่โดยต้นไม้ในลำดับต่อไปจะถูกสร้างจากข้อผิดพลาดจากการคำนวณต้นไม้ก่อนหน้าโดยใช้อัลกอริทึม Levelwise ในการสร้างต้นไม้

2.1.3. Prophet

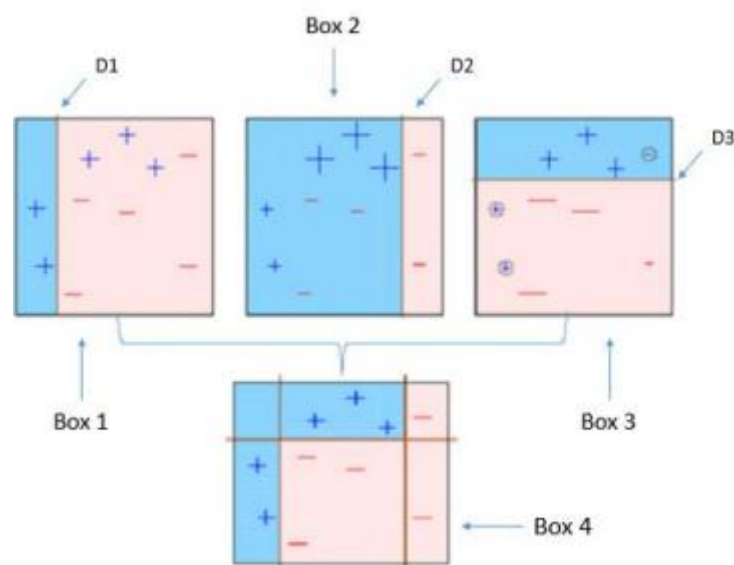


รูปที่ 2.3 ขั้นตอนการทำงานของ Prophets

ที่มา : Sean J Taylor, Benjamin Letham. (2017). Forecasting at scale. Retrieved From
<https://peerj.com/preprints/3190/>

Prophet [6] เป็น open source ที่ถูกพัฒนาโดย Facebook ทางทีมของ Data Science จะเป็นโมเดลที่สำเร็จรูปในการทำนายข้อมูล และยังเรียกใช้งานง่าย เหมาะสมกับการสร้างแบบจำลอง time-series ซึ่งคือ ผลรวมของสามฟังก์ชันของเวลาบวกกับระยะข้อผิดพลาด $growth(t)$, $seasonality\ s(t)$, $holidays\ h(t)$ และ $error\ e_t$ ในที่นี้ $g(t)$ คือฟังก์ชันแนวโน้มซึ่งจำลองการเปลี่ยนแปลงในค่าของอนุกรมเวลา $s(t)$ แสดงถึงการเปลี่ยนแปลง เช่น ฤดูกาลรายสัปดาห์และรายปี และ $h(t)$ แสดงถึงผลกระทบของวันหยุดที่เกิดขึ้นตามตารางเวลาที่อาจไม่เหมือนเดิม และข้อผิดพลาด e_t แสดงถึงการเปลี่ยนแปลงที่มีลักษณะเฉพาะ ซึ่งเป็นคลาสของแบบจำลองการถดถอยที่มีตัวปรับที่ไม่ใช่เชิงเส้นซึ่งนำไปใช้กับการถดถอย โดยใช้เวลาเป็นตัวถดถอยเท่านั้น แต่อาจเป็นแบบเชิงเส้นและไม่เป็นเส้นตรงหลายตัว

2.1.4. XGBoost



รูปที่ 2.4 ขั้นตอนการทำงานของ XGBoost

ที่มา : Manish Pathak. (2019). Using XGBoost in Python. Retrieved from <https://www.datacamp.com/community/tutorials/xgboost-in-python/>

XGBoost [7] เป็นเทคนิคที่พัฒนาจากเทคนิค Gradient boosting เพื่อเพิ่มความแม่นยำ และความยืดหยุ่นให้กับแบบจำลอง โดยใช้หลักการของ Ensemble Learning Method ในการ Boosting เพื่อสร้างตัวเรียนรู้หลายๆตัว (Multiple Learner) หรือเรียกได้ว่าเป็นการรวม Weak Learners หลายๆตัวเข้าด้วยกัน โดยการทำงานจะเป็นนำเอาต้นไม้ตัดสินใจมาฝึกสอนต่อกันหลาย ๆ ต้น โดยที่ต้นไม้ตัดสินใจแต่ละต้นจะเรียนรู้จากค่าความผิดพลาดของต้นก่อนหน้านี้ ซึ่งทำให้ความแม่นยำในการทำนายจะมากขึ้นเรื่อย ๆ เมื่อมีการเรียนรู้ของต้นไม้ตัดสินใจต่อเนื่องกันจนมีความลึกมากพอ แบบจำลองจะหยุดเรียนรู้เมื่อไม่เหลือค่าความผิดพลาดจากต้นไม้ตัดสินใจต้นก่อนหน้านี้ให้เรียนรู้แล้ว

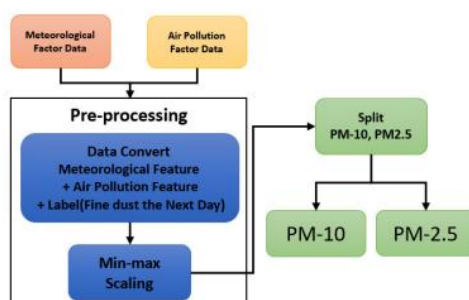
2.2. งานวิจัยที่เกี่ยวข้อง

2.2.1. Prediction System for Fine Particulate Matter Concentration Index by Meteorological and Air Pollution Material Factors Based on Machine Learning

โดย Jihoon Yoo, Dongil Shin และ Dongkyoo Shin (2019)

Jihoon Yoo และคณะ[1] ทำนายอนุภาคฝุ่นละอองขนาดเล็ก PM2.5 ในกรุงโซล โดยต้องการวัดความแม่นยำของอัลกอริทึมหลายๆตัว ข้อมูลทางอุตุนิยมวิทยาที่งานวิจัยนี้ใช้ คือข้อมูลมาตั้งแต่เดือน 2013 ถึงเดือน ธันวาคม 2016 โดย Seoul City ถูกนำมาใช้ ทิศทางลม (avg-wd), ความเร็วลมเฉลี่ย (avg-ws), อุณหภูมิเฉลี่ย (avg-t), อุณหภูมิต่ำสุด (min-t) , อุณหภูมิสูงสุด (max-t), ปริมาณน้ำฝน (rain) และค่าความชื้นสัมพัทธ์ในอากาศ (avg-h) และข้อมูลมลพิษอากาศที่งานวิจัยนี้นำมา ใช้ จะเป็นข้อมูลเฉพาะของกรุงโซลที่จัดทำ โดย Air Korea หกองค์ประกอบที่เกี่ยวข้องกับฝุ่นละอองถูกนำมา ใช้ ได้แก่ PM10, PM2.5, O3, NO2, CO และ SO2 ส่วนอัลกอริทึมที่ใช้ในการทำนาย Linear Regression, Multi-Layer Perceptron, Gradient Boosting, Random Forest

$$z_i = \frac{x_i - \min(x)}{\max(x) - \min(x)} \quad (1)$$



รูปที่ 2.5 ขั้นตอนการ Pre-processing data

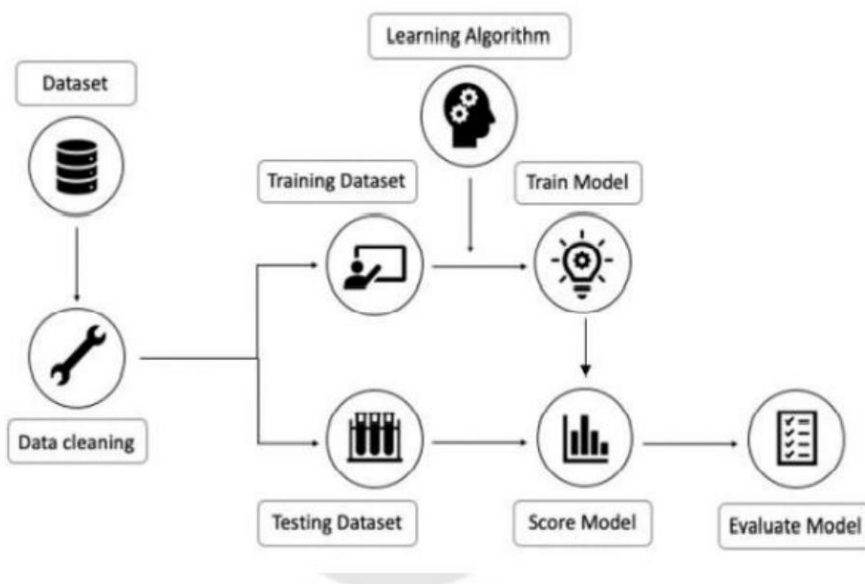
ขั้นตอนการ Pre-processing จะนำข้อมูลจากทางอุตุนิยม และมลพิษมารวมกัน แล้วจากนั้นทำการปรับค่าเป็น Min-max Scaling เนื่องจากถ้ามี feature ตัวหนึ่งมีขนาดที่ใหญ่มากกว่า feature ตัวอื่นๆ จะทำให้การทำงานของอัลกอริทึมมีความคลาดเคลื่อนได้ แล้วหลังจากนั้นก็นำข้อมูลมาแยกเป็นระหว่าง PM10 กับ PM 2.5 เพื่อนำข้อมูลที่ได้มาให้อัลกอริทึมทำการทำนาย

ซึ่งผลการทดลองของวิจัยนี้ แบบจำลอง neural network แสดงการคาดการณ์ที่ดีกว่าแบบจำลอง linear regression นอกจากนี้แบบจำลอง tree แบบจำลอง random forest และแบบจำลอง gradient boosting ยังแสดงประสิทธิภาพการทำนายที่ดีกว่าอัลกอริทึมอื่นๆ มาก สิ่งนี้ทำให้แบบจำลอง tree เหมาะสมกว่าอัลกอริทึมอื่นๆ เมื่อเพิ่มปัจจัยที่มีอิทธิพลต่างๆ เช่น สภาพอากาศและมลพิษทางอากาศ

2.2.2. การพยากรณ์ฝุ่น PM2.5 ล่วงหน้าในจังหวัดเชียงใหม่ โดยใช้แบบจำลองหน่วยความจำ ระยะสั้นแบบยาว

โดย ปรัชญา สิงหรวงศ์ บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ (2563)

ปรัชญา สิงหรวงศ์ [2] ทำนายอนุภาคฝุ่นละอองขนาดเล็กในจังหวัดเชียงใหม่ โดยต้องการวัดความแม่นยำของ LSTM ด้วย Optimizer ที่แตกต่างกัน โดยใช้ Adam กับ RMSprop และ AdaGrad เพื่อหา RMSE/MSE โดยจะใช้ชุดข้อมูลสถานีศูนย์ราชการจังหวัดเชียงใหม่ ข้อมูลที่ใช้ระหว่างวันที่ 1 มกราคม 2018 ถึง 31 พฤษภาคม 2019 ทั้งหมด 12 ตัวแปร ได้แก่ NO, NOX, NO2, SO2, O3, PM10, Wind dir, Temp, Rel hum, Rain, Wind speed, PM2.5 และชุดข้อมูลสถานีโรงเรียนยุพราชวิทยาลัย ข้อมูลที่ใช้ระหว่างวันที่ 1 กรกฎาคม 2018 ถึง 31 พฤษภาคม 2019 12 ตัวแปร ได้แก่ CO, NO, NOX, NO2, SO2, Wind dir, Temp, Rel hum, Wind speed, PM2.5 หลังจากได้ข้อมูลมาก็ทำการลบค่า missing value หรือ NAN และหา Min-max scaling แล้วนำเข้าสู่อัลกอริทึมเพื่อทำนาย และวัดผลด้วย optimizer ที่แตกต่างกัน



รูปที่ 2.6 ขั้นตอนการดำเนินการวิจัย

ผลการทดลองของสถานีศูนย์ราชการจังหวัดเชียงใหม่ เป็นชุดข้อมูลที่มีค่าข้อมูลสูญหายไม่มากพยากรณ์ล่วงหน้า 24 ชั่วโมงได้ดีที่สุด Adam มีค่า RMSE เท่ากับ 11.48 $\mu\text{g}/\text{m}^3$,ค่า MAE เท่ากับ 8.78 $\mu\text{g}/\text{m}^3$ และค่า MAPE เท่ากับ 21.65 เปอร์เซ็นต์ AdaGrad มีค่า RMSE เท่ากับ 11.83 $\mu\text{g}/\text{m}^3$, ค่า MAE เท่ากับ 9.05 $\mu\text{g}/\text{m}^3$ และค่า MAPE เท่ากับ 20.42 เปอร์เซ็นต์ และสถานีโรงเรียนยุพราชวิทยาลัยเป็นชุดข้อมูลที่มีค่าข้อมูลสูญหายมาก จะทำให้สามารถพยากรณ์ล่วงหน้า 24 ชั่วโมงได้ดีที่สุด RMSprop มีค่า RMSE เท่ากับ 10.38 $\mu\text{g}/\text{m}^3$, ค่า MAE เท่ากับ 8.01 $\mu\text{g}/\text{m}^3$ และ MAPE เท่ากับ 46.40 เปอร์เซ็นต์

2.2.3. แบบจำลองการพยากรณ์ค่า PM2.5 โดยใช้โครงข่ายประสาทเทียมแบบ LSTM ในพื้นที่

กรุงเทพมหานคร

โดย ศิริรักษ์ ศรีทองชัย ภาควิชาวิศวกรรมอุตสาหการและการจัดการ คณะวิศวกรรมศาสตร์และเทคโนโลยี
อุตสาหกรรม มหาวิทยาลัยศิลปากร นครปฐม (2564)

ศิริรักษ์ ศรีทองชัย [3] ทำนายอนุภาคฝุ่นละอองขนาดเล็กในจังหวัดกรุงเทพมหานคร โดยต้องการวัดความแม่นยำของอัลกอริทึม LSTM ที่มีการใส่ตัวแปรในแต่ละแบบจำลองที่แตกต่างกัน เพื่อต้องการที่จะรู้ว่าการเพิ่มตัวแปรอะไร ที่จะทำให้อัลกอริทึมมีการทำนายที่คลาดเคลื่อนน้อยที่สุด โดยจะใช้ข้อมูลข้อมูลสารก่อกมลพิษทางอากาศ ข้อมูลพื้นที่ทั่วไปจากสถานีตรวจวัดคุณภาพอากาศที่บางนา และข้อมูลพื้นที่ริมถนนมาจากสถานีโรงพยาบาลจุฬาลงกรณ์ และใช้ข้อมูลอุตุนิยมวิทยาจากสถานีอุตุนิยมวิทยาศูนย์ประชุมแห่งชาติสิริกิติ์ ซึ่งจะใช้ข้อมูลย้อนหลัง 4 ปีล่าสุด (พ.ศ.2560-2563) โดยเป็นข้อมูลรายวัน ข้อมูลที่ไม่ใช่รายวันจะทำการเฉลี่ยต่อวันแทนข้อมูลรายวัน ตั้งแต่เดือนมกราคม พ.ศ. 2560 ถึง เดือน ธันวาคม พ.ศ. 2563

ตารางที่ 3 : ตัวแปรในแบบจำลอง

แบบจำลอง	ตัวแปรป้อนเข้า	
1	พื้นที่ ทั่วไป	PM10 NO2 CO O3 SO2
2		PM10 NO2 CO O3 SO2 LAGPM2.5
3		PM10 NO2 CO O3 SO2 Temp Hum Press WindSp WindDir
4		PM10 NO2 CO O3 SO2 Temp Hum Press WindSp WindDir LAGPM2.5
5	พื้นที่ ริมถนน	PM10 Press Hum NO2 CO
6		PM10 Press Hum NO2 CO LAGPM2.5
7		PM10 NO2 CO Temp Hum Press WindSp WindDir
8		PM10 NO2 CO Temp Hum Press WindSp WindDir LAGPM2.5

ตารางที่ 4 : ผลเปรียบเทียบความแม่นยำของแบบจำลองพยากรณ์

แบบจำลอง	RMSE			MAPE		
	1 วัน	3 วัน	5 วัน	1 วัน	3 วัน	5 วัน
1	7.98	8.29	9.76	32.68	34.22	33.58
2	7.73	7.89	8.52	27.81	26.28	27.57
3	8.31	8.49	9.36	32.02	33.75	37.28
4	7.82	7.66	8.69	31.43	29.67	29.78
5	5.09	5.71	5.90	24.45	28.42	28.88
6	4.82	5.05	5.36	22.88	24.27	25.57
7	5.02	4.84	6.58	22.96	21.96	30.08
8	4.55	4.75	5.18	20.47	23.11	24.13

รูปที่ 2.7 ตัวแปรในแบบจำลอง และผลเปรียบเทียบความแม่นยำของแบบจำลองพยากรณ์

ผลการทดลอง ในของพื้นที่ทั่วไป ตัวที่ดีที่สุดจะเป็นแบบจำลองที่ 2 เพราะมีการเพิ่มตัวแปร LAGPM2.5 ซึ่งคือค่า PM2.5 ของวันถัดไป โดยจะมีค่า RMSE ตามตารางที่ 4 ทำนายล่วงหน้า 1 วัน ได้ 7.73, 3 วัน เป็น 7.89 และ 5 วัน 8.52 ส่วนในของพื้นที่ริมถนน ตัวที่ดีที่สุดจะเป็นแบบจำลองที่ 8 เพราะมีตัวแปรที่เพิ่มเข้าไปครบแล้วยังเพิ่มตัวแปร LAGPM2.5 โดยจะมีค่า RMSE ตามตารางที่ 4 ทำนายล่วงหน้า 1 วัน ได้ 4.55, 3 วัน เป็น 4.75 และ 5 วัน เป็น 5.18

บทที่ 3

การวิเคราะห์ และการออกแบบ

สำหรับเนื้อหาในบทนี้ จะอธิบายการวิเคราะห์และการออกแบบ แอลกอริทึม ที่ใช้ในการทำนายค่าฝุ่น PM2.5

ในการวิจัยครั้งนี้ ผู้วิจัยได้ดำเนินการตามขั้นตอนดังนี้

1. ศึกษาและรวบรวมข้อมูล
2. ประมวลผลข้อมูล
3. วิเคราะห์ข้อมูล
4. ประเมินผลประสิทธิภาพของแบบจำลอง

3.1 ศึกษาและรวบรวมข้อมูล

จากการศึกษางานวิจัยที่เกี่ยวข้องพบว่าปัจจัยที่มีความสัมพันธ์กับค่าฝุ่น PM2.5 มีหลายปัจจัยเช่น CO , NO_2 , NO , NO_X , O_3 , PM10, Wind speed, Wind direction, Average Temperature, humidity, Pressure ทางผู้จัดทำจึงได้ขอข้อมูลจากหน่วยงานที่เกี่ยวข้อง ดังแสดงในตาราง 3.1

ตารางที่ 3.1 แสดงรายละเอียดของปัจจัยที่มีความสัมพันธ์กับค่าฝุ่น PM 2.5 และ feature

ที่มาของข้อมูล	Variable	Description
สำนักสิ่งแวดล้อม กรมควบคุมมลพิษ	CO (ppm)	คาร์บอนมอนนอกไซด์
	NO_2 (ppb)	ไนโตรเจนไดออกไซด์
	O_3 (ppb)	โอโซน
	NO (ppb)	ไนตริกออกไซด์
	NO_X (ppb)	ไนโตรเจนออกไซด์
	PM10 ($\mu\text{g}/\text{m}^3$)	ฝุ่นละออง PM10
	PM2.5 ($\mu\text{g}/\text{m}^3$)	ฝุ่นละออง PM2.5
กรมอุตุนิยมวิทยา	Wind speed (m/s)	ความเร็วลม
	Wind direction (degrees)	ทิศทางลม
	Average Temperature	อุณหภูมิ
	Humidity (%RH)	ค่าความชื้นสัมพัทธ์ในอากาศ
	Pressure (mmHg)	ความกดอากาศ
	date	วัน/ เดือน/ ปี

โดยข้อมูลที่ได้เป็นข้อมูลบริเวณ พื้นที่ เขตราชเทวี พระโขนง ดินแดง ตั้งแต่วันที่ 1 มกราคม 2561 ถึง

31 ตุลาคม 2565

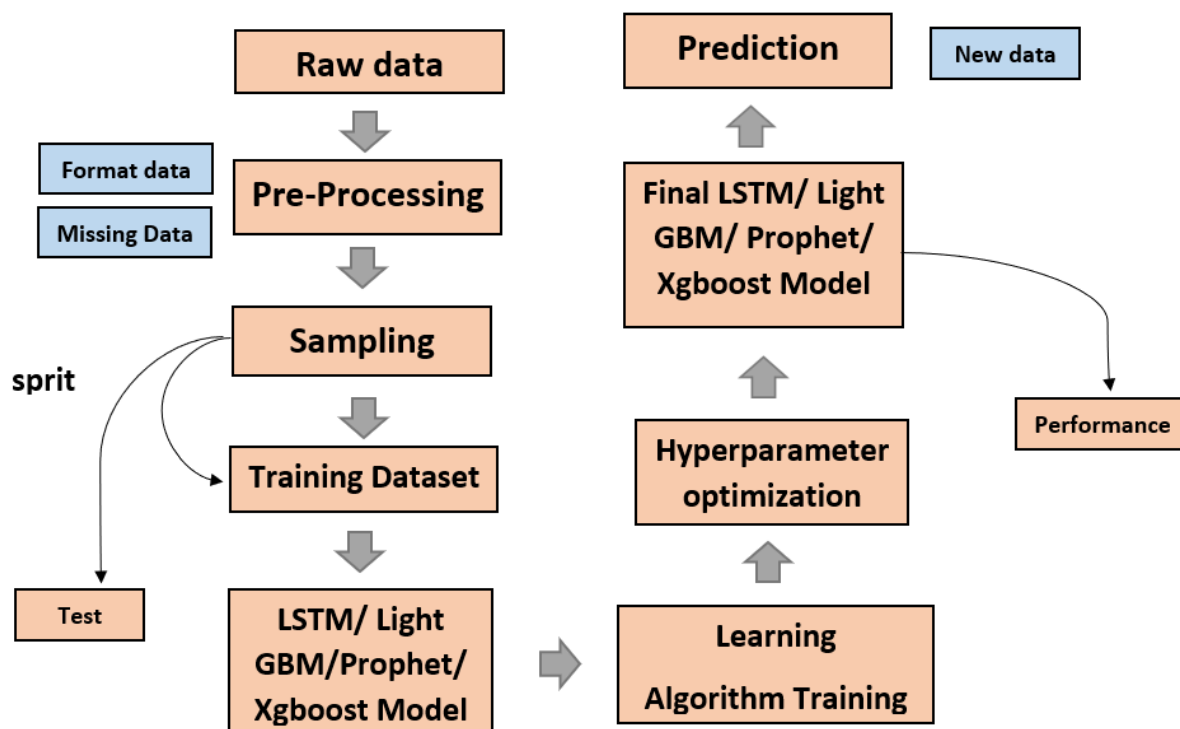
3.2 ประมวลผลข้อมูล

1. ทำการเลือกข้อมูล Dataset บริเวณพื้นที่กรุงเทพมหานคร มีสามเขต ได้แก่ เขตราชเทวี เขตพระโขนง เขตดินแดง และ Dataset ที่เฉลี่ยรวมทั้งสามเขต จัดรูปแบบข้อมูลให้เหมาะสม เพื่อให้สามารถนำมาใช้งานได้
2. ศึกษาและทำการวิเคราะห์ feature ใน Dataset โดยทำการ clean ข้อมูลที่มาจากกรมมลพิษ และ สำนักสิ่งแวดล้อม โดยนำข้อมูลที่เป็นรายชั่วโมงมาเฉลี่ยรวมให้เป็นข้อมูลรายวันและนำข้อมูลจากกรมอุตุนิยมวิทยามาใส่เพิ่มของแต่ละวัน ดังตาราง 3.2
3. ถ้าข้อมูลหนึ่งมีตัวแปรไม่เท่ากับหรือเหมือนกับเขตอื่นๆ จะมีการลบตัวแปรที่แตกต่างนั้นออก เพื่อให้แต่ละข้อมูลแต่ละเขตมีตัวแปรที่เหมือนกัน เพราะเมื่อนำข้อมูลของทุกเขตมาเฉลี่ยรวมกันแล้ว จะได้ไม่คลาดเคลื่อนมากเกินไป

ตารางที่ 3.2 ชุดข้อมูลที่พร้อมนำไปเข้าโมเดล

date	pm25	pm10	o3	no2	co	avgT	humid	pressure-n	windDir	windSpd	
1/1/2018	21	32	37.91	13.26	0.64	27.1	70	1009.4	100	9	
1/2/2018	26	37	38.22	18.96	0.68	27.5	67	1009.4	120	4	
1/3/2018	38	51	29.5	29.08	0.85	28.2	66	1009.4	60	4	
1/4/2018	39	55	45.09	22.22	0.86	29.3	64	1009.4	330	6	
1/5/2018	51	67	49.26	36.7	1	30.4	63	1009.4			
1/6/2018	42	54	45.09	34.35	0.96	29.8	70	1009.4	120	2	
1/7/2018	26	29	23.26	13.48	0.57	29.2	78	1009.4	140	2	
1/8/2018	10	17	13.74	16.52	0.64	29.8	77	1009.4	140	3	
1/9/2018	22	33	16.3	31.83	0.92	30.7	75	1009.4	140	6	
1/10/2018	10	18	14.96	19.65	0.89	26.6	88	1009.4	120	4	
1/11/2018	25	50	16.09	32.96	1.04	25.6	74	1009.4	120	5	
1/12/2018	34	49	26.43	27.39	0.84	25	65	1009.4	120	8	
1/13/2018	36	60	34.43	26.87	0.75	24	65	1009.4	330	6	
1/14/2018	38	56	33.09	25.83	0.87	25.2	63	1009.4	320	5	
1/15/2018	52	69	28.26	34	1	25.9	64	1009.4	300	6	
1/16/2018	55	74	27.96	44.7	1.29	27.6	68	1009.4	330	7	
1/17/2018	69	93	36.35	44.3	1.51	28.5	69	1009.4	300	5	
1/18/2018	54	79	29.17	36.04	1.15	29.2	63	1009.4	310	6	
1/19/2018	57	74	22.3	38.04	1.16	29.3	65	1009.4	100	4	
1/20/2018	45	73	18.09	34.61	1.07	29	64	1009.4	80	6	
1/21/2018	62	86	23.39	42.35	1.26	29.2	67	1009.4	70	4	
1/22/2018	97	121	24.22	45.96	1.5	29.6	69	1009.4	330	4	
1/23/2018	66	77	26.57	38.26	1.36	29.7	74	1009.4	120	6	
1/24/2018	86	77	11.87	38.65	1.38	28.9	79	1009.4	120	4	
1/25/2018	25	40	10.86	21.83	1.05	29.2	81	1009.4	270	4	
1/26/2018	17	28	13.04	15.52	0.92	29.8	80	1009.4	120	5	

3.3 วิเคราะห์ข้อมูล



รูปที่ 3.1 การทำงานของของแบบจำลอง LSTM, LightGBM, Prophet, Xgboost

จากรูป 3.1 Flowchart ด้านบนได้อธิบายถึงกระบวนการทำงานของแบบจำลอง LSTM, LightGBM, Prophet, XGBoost โดยเริ่มตั้งแต่ขั้นตอนแรกที่น่าข้อมูลเข้ามาทำความสะอาดและแบ่งข้อมูลออกเป็น Train, Test แล้วจึงนำข้อมูล Train ใส่เข้าแบบจำลอง เพื่อให้แบบจำลองได้ทำการเรียนรู้ข้อมูล และ test ข้อมูลออกมาเพื่อประเมินความแม่นยำของแบบจำลองแต่ละชนิดและวิเคราะห์ผลลัพธ์ว่าแบบจำลองชนิดไหนมีประสิทธิภาพมากที่สุด

3.4 ประเมินผลประสิทธิภาพของแบบจำลอง

ในการประเมินประสิทธิภาพของแบบจำลองเราจะใช้ค่า MSE และ RMSE ในการวัดประสิทธิภาพ

1. MSE (Mean Square Error)

$$\text{สูตรการคำนวณ MSE} = \Sigma(\hat{y}_i - y_i)^2 / n$$

Σ เป็นสัญลักษณ์ที่หมายถึง "ผลรวม"

\hat{y}_i คือค่าที่ทำนายไว้สำหรับการสังเกตที่ i^{th}

y_i คือค่าที่สังเกตได้สำหรับการสังเกตที่ i^{th}

n คือขนาดตัวอย่าง

MSE : บอกความแตกต่างของค่าเฉลี่ยกำลังสองระหว่างค่าที่คาดการณ์และค่าจริงในชุดข้อมูล ยิ่ง MSE ต่ำเท่าใด โมเดลก็จะยิ่งเหมาะสมกับชุดข้อมูลมากขึ้นเท่านั้น

2. RMSE (Root Mean Square Error)

สูตรการคำนวณ

$$\text{RMSE} = \sqrt{\Sigma(\hat{y}_i - y_i)^2 / n}$$

Σ เป็นสัญลักษณ์ที่หมายถึง "ผลรวม"

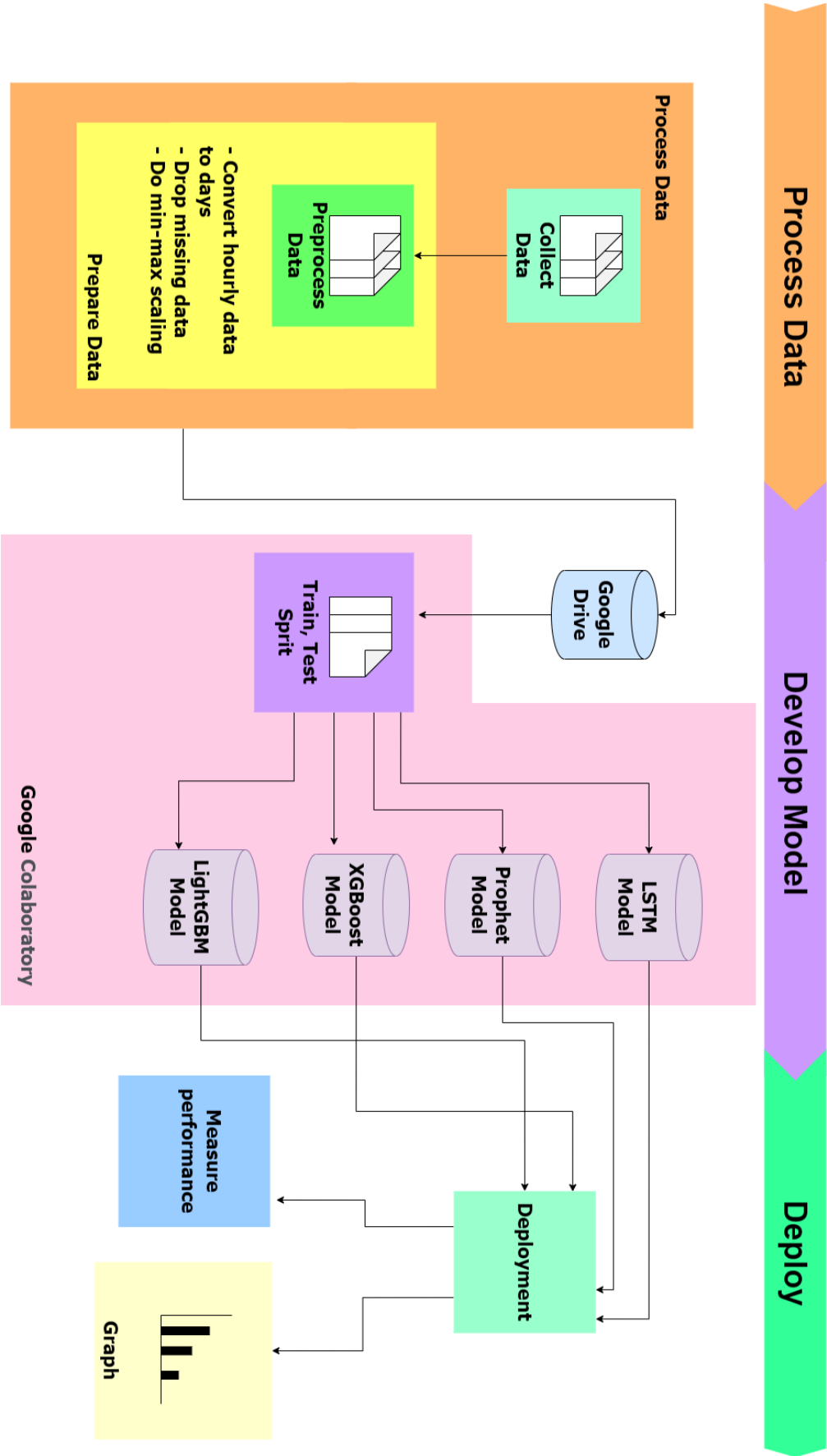
\hat{y}_i คือค่าที่ทำนายไว้สำหรับการสังเกตที่ i^{th}

y_i คือค่าที่สังเกตได้สำหรับการสังเกตที่ i^{th}

n คือขนาดตัวอย่าง

RMSE : บอกถึงรากที่สองของผลต่างกำลังสองเฉลี่ยระหว่างค่าที่คาดการณ์และค่าจริงในชุดข้อมูล ยิ่ง RMSE ต่ำเท่าใด โมเดลก็จะยิ่งเหมาะสมกับชุดข้อมูลมากขึ้นเท่านั้น

รูป 3.2 System Architecture



3.5 การออกแบบแบบจำลอง

ขณะนี้ผู้จัดทำไม่ได้เปลี่ยนแปลงแบบจำลอง ซึ่ง Model Parameters ของทุกแบบจำลอง จะใช้ค่า default หมดทั้งสิ้น ความแตกต่างของแต่ละแบบจำลองนั้น จะอยู่ที่การแบ่งข้อมูล test และข้อมูล train เพราะแต่ละข้อมูลแต่ละเซตจะมีจำนวนข้อมูลที่แตกต่างกัน และในแบบจำลอง Prophet นั้นไม่สามารถที่จะกำหนดข้อมูลการแบ่งข้อมูลได้ เนื่องจากเป็นแบบจำลองสำเร็จรูป ทำเพียงแคใส่ข้อมูลลงไป แล้วให้แบบจำลองทำการทำนายขึ้นมา แล้วหลังจากนั้นแบบจำลองทั้งหมด เราจะทำการประเมินผล พร้อมแสดงค่าฝุ่นที่ทางแบบจำลองได้ทำการคาดการณ์ออกมา

บทที่ 4

ผลการศึกษา

ในการวิจัยการพยากรณ์ขนาดอนุภาคฝุ่นละอองขนาดเล็ก (PM2.5) ในอากาศ ใน 7 วันล่วงหน้า ซึ่งจะใช้ข้อมูลมลพิษทางอากาศ จากกรมมลพิษ และสำนักสิ่งแวดล้อม และใช้ข้อมูลภูมิอากาศ จากกรมอุตุนิยมวิทยา โดยจะคัดเลือกมาจำนวน 3 เขต ในจังหวัดกรุงเทพมหานคร ได้แก่ เขตราชเทวี เขตพระโขนง เขตดินแดง โดยจะใช้เทคนิคของการเรียนรู้ของเครื่อง ซึ่งได้ดำเนินการพัฒนาโดยการศึกษาตามขั้นตอนต่าง ๆ ตลอดจนการวัดประสิทธิภาพของแบบจำลอง เพื่อให้บรรลุจุดประสงค์ของโครงการไว้ ได้ดังนี้

1. ผลลัพธ์การวัดประสิทธิภาพของแต่ละแบบจำลอง ในข้อมูลของเขตราชเทวี เขตพระโขนง เขตดินแดง และการเฉลี่ยรวมทุกเขต
2. ผลลัพธ์แนวโน้มที่เป็นกราฟของฝุ่นละอองขนาดเล็กจากการทดสอบแบบจำลอง

4.1. ผลลัพธ์ประสิทธิภาพของแต่ละแบบจำลอง ในข้อมูลของเขตราชเทวี เขตพระโขนง เขตดินแดง และการเฉลี่ยรวมทุกเขต

4.1.1. ผลลัพธ์ประสิทธิภาพของแบบจำลองในเขตของราชเทวี

ราชเทวี มี 10 ตัวแปร ได้แก่ pm2, pm10, o3, no2, co, avgT, humid, pressure-monthly, windDir, windSpd มีข้อมูลทั้งหมด 1706 ข้อมูล

1. Long short-term memory (LSTM)

จากการสร้างแบบจำลอง LSTM โดยใช้ข้อมูลภูมิอากาศจากกรมอุตุนิยมวิทยา และข้อมูลมลพิษของเขตราชเทวี จากกรมควบคุมมลพิษ เพื่อทำนาย 7 วันล่วงหน้า โดยใช้ตัววัดประสิทธิภาพแบบจำลอง คือ ค่า RMSE กับ MSE โดยกำหนดให้ ข้อมูล train เป็น 1,364 ข้อมูล หรือ 80% และ ข้อมูล test เป็น 342 หรือ 20% และกำหนด Epochs เป็น 60 Batch size เป็น 256 ซึ่งค่าจากการวัดประสิทธิภาพออกมา จะได้ ค่า RMSE เป็น 2.17206 กับ MSE 4.71785 ซึ่งถือเป็นอันดับที่ 2 ของในทั้ง 4 แบบจำลอง

2. Light Gradient Boosting Machine (Light GBM)

จากการสร้างแบบจำลอง LightGBM โดยใช้ข้อมูลภูมิอากาศจากกรมอุตุนิยมวิทยา และข้อมูลมลพิษของเขตราชเทวี จากกรมควบคุมมลพิษ เพื่อทำนาย 7 วันล่วงหน้า โดยใช้ตัววัดประสิทธิภาพแบบจำลอง คือ ค่า RMSE กับ MSE โดยกำหนดให้ ข้อมูล train เป็น 1,500 ข้อมูล หรือ 80% และ ข้อมูล test เป็น 206

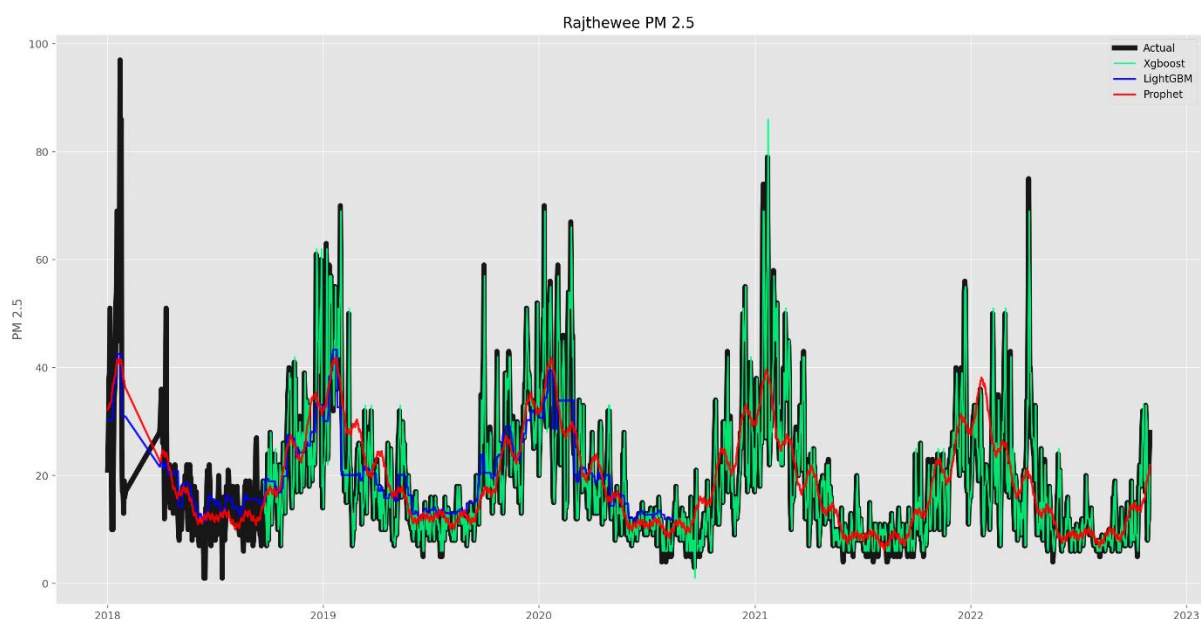
หรือ 20% และกำหนด Early stopping rounds เป็น 200 verbose เป็น 10 ซึ่งค่าจากการวัดประสิทธิภาพออกมา จะได้ค่า RMSE เป็น 7.12309 กับ MSE เป็น 50.73837 ซึ่งถือเป็นอันดับที่ 3 ของในทั้ง 4 แบบจำลอง

3. Prophet

จากการสร้างแบบจำลอง Prophet โดยใช้ข้อมูลภูมิอากาศจากกรมอุตุนิยมวิทยา และข้อมูลมลพิษของเขตราชเทวี จากกรมควบคุมมลพิษ เพื่อทำนาย 7 วันล่วงหน้า โดยใช้ตัววัดประสิทธิภาพแบบจำลอง คือ ค่า RMSE กับ MSE โดยกำหนดให้ periods = 180 จะได้ค่า RMSE เป็น 8.54326 กับ MSE เป็น 72.9872 ซึ่งถือเป็นอันดับที่ 4 ของในทั้ง 4 แบบจำลอง

4. XGBoost

จากการสร้างแบบจำลอง XGBoost โดยใช้ข้อมูลภูมิอากาศจากกรมอุตุนิยมวิทยา และข้อมูลมลพิษของเขตราชเทวี จากกรมควบคุมมลพิษ เพื่อทำนาย 7 วันล่วงหน้า โดยใช้ตัววัดประสิทธิภาพแบบจำลอง คือ ค่า RMSE กับ MSE ซึ่งค่าจากการวัดประสิทธิภาพออกมา จะได้ค่า RMSE เป็น 0.00619 กับ MSE เป็น $3.8316e-5$ ซึ่งถือเป็นอันดับที่ 1 ของในทั้ง 4 แบบจำลอง



รูปที่ 4.1 กราฟแสดงข้อมูลการทำนายจากโมเดลเปรียบเทียบกับค่าจริง ของทุกอัลกอริทึม ในเขตราชเทวี

4.1.2. ผลลัพธ์ประสิทธิภาพของแบบจำลองในเขตของพระโขนง

พระโขนง มี 11 ตัวแปร ได้แก่ pm10, co, no, no2, nox, pm25, avgT, humid, pressure-monthly, windDir, windSpd มีข้อมูลทั้งหมด 1,016 ข้อมูล

1. Long short-term memory (LSTM)

จากการสร้างแบบจำลอง LSTM โดยใช้ข้อมูลภูมิอากาศจากกรมอุตุนิยมวิทยา และข้อมูลมลพิษของเขตพระโขนง จากสำนักสิ่งแวดล้อม กรุงเทพมหานคร เพื่อทำนาย 7 วันล่วงหน้า โดยใช้ตัววัดประสิทธิภาพแบบจำลอง คือ ค่า RMSE กับ MSE โดยกำหนดให้ ข้อมูล train เป็น 822 ข้อมูล หรือ 80% และ ข้อมูล test เป็น 206 หรือ 20% และกำหนด Epochs เป็น 60 Batch size เป็น 256 ซึ่งค่าจากการวัดประสิทธิภาพออกมา จะได้ค่า RMSE เป็น 2.30964 กับ MSE 5.33445 ซึ่งถือเป็นอันดับที่ 2 ของในทั้ง 4 แบบจำลอง

2. Light Gradient Boosting Machine (Light GBM)

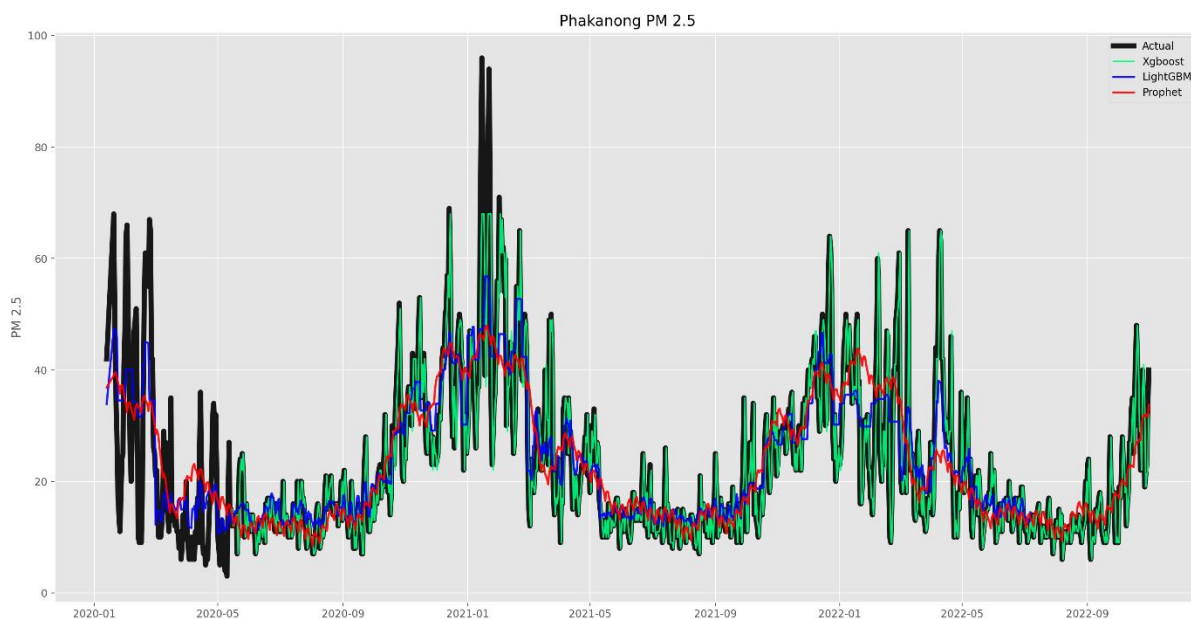
จากการสร้างแบบจำลอง LightGBM โดยใช้ข้อมูลภูมิอากาศจากกรมอุตุนิยมวิทยา และข้อมูลมลพิษของเขตพระโขนง จากสำนักสิ่งแวดล้อม กรุงเทพมหานคร เพื่อทำนาย 7 วันล่วงหน้า โดยใช้ตัววัดประสิทธิภาพแบบจำลอง คือ ค่า RMSE กับ MSE โดยกำหนดให้ ข้อมูล train เป็น 900 ข้อมูล หรือ 90% และ ข้อมูล test เป็น 116 หรือ 10% และกำหนด Early stopping rounds เป็น 200 verbose เป็น 10 ซึ่งค่าจากการวัดประสิทธิภาพออกมา จะได้ค่า RMSE เป็น 6.30152 กับ MSE เป็น 39.70917 ซึ่งถือเป็นอันดับที่ 3 ของในทั้ง 4 แบบจำลอง

3. Prophet

จากการสร้างแบบจำลอง Prophet โดยใช้ข้อมูลภูมิอากาศจากกรมอุตุนิยมวิทยา และข้อมูลมลพิษของเขตพระโขนง จากสำนักสิ่งแวดล้อม กรุงเทพมหานคร เพื่อทำนาย 7 วันล่วงหน้า โดยใช้ตัววัดประสิทธิภาพแบบจำลอง คือ ค่า RMSE กับ MSE โดยกำหนดให้ periods = 180 ซึ่งค่าจากการวัดประสิทธิภาพออกมา จะได้ค่า RMSE เป็น 9.45775 กับ MSE เป็น 89.44899 ซึ่งถือเป็นอันดับที่ 4 ของในทั้ง 4 แบบจำลอง

4. XGBoost

จากการสร้างแบบจำลอง XGBoost โดยใช้ข้อมูลภูมิอากาศจากกรมอุตุนิยมวิทยา และข้อมูลมลพิษของเขตพระโขนง จากสำนักสิ่งแวดล้อม กรุงเทพมหานคร เพื่อทำนาย 7 วันล่วงหน้า โดยใช้ตัววัดประสิทธิภาพแบบจำลอง คือ ค่า RMSE กับ MSE ซึ่งค่าจากการวัดประสิทธิภาพออกมา จะได้ค่า RMSE เป็น 0.00420 กับ MSE เป็น $1.772e-5$ ซึ่งถือเป็นอันดับที่ 1 ของในทั้ง 4 แบบจำลอง



รูปที่ 4.2 กราฟแสดงข้อมูลการทำนายจากโมเดลเปรียบเทียบกับค่าจริง ของทุกอัลกอริทึม ในเขตพระโขนง

4.1.3. ผลลัพธ์ประสิทธิภาพของแบบจำลองในเขตของดินแดง

ดินแดง มี 11 ตัวแปร ได้แก่ pm10, co, no, no2, nox, pm25, avgT, humid, pressure-monthly, windDir, windSpd มีข้อมูลทั้งหมด 1,028 ข้อมูล

1. Long short-term memory (LSTM)

จากการสร้างแบบจำลอง LSTM โดยใช้ข้อมูลภูมิอากาศจากกรมอุตุนิยมวิทยา และข้อมูลมลพิษของเขตดินแดง จากสำนักสิ่งแวดล้อม กรุงเทพมหานคร เพื่อทำนาย 7 วันล่วงหน้า โดยใช้ตัววัดประสิทธิภาพแบบจำลอง คือ ค่า RMSE กับ MSE โดยกำหนดให้ ข้อมูล train เป็น 822 ข้อมูล หรือ 80% และ ข้อมูล test เป็น 206 หรือ 20% และกำหนด Epochs เป็น 60 Batch size เป็น 256 ซึ่งค่าจากการวัดประสิทธิภาพออกมา จะได้ค่า RMSE เป็น 1.71952 กับ MSE 2.95676 ซึ่งถือเป็นอันดับที่ 2 ของในทั้ง 4 แบบจำลอง

2. Light Gradient Boosting Machine (Light GBM)

จากการสร้างแบบจำลอง LightGBM โดยใช้ข้อมูลภูมิอากาศจากกรมอุตุนิยมวิทยา และข้อมูลมลพิษของเขตดินแดง จากสำนักสิ่งแวดล้อม กรุงเทพมหานคร เพื่อทำนาย 7 วันล่วงหน้า โดยใช้ตัววัดประสิทธิภาพแบบจำลอง คือ ค่า RMSE กับ MSE โดยกำหนดให้ ข้อมูล train เป็น 800 ข้อมูล หรือ 80% และ ข้อมูล test เป็น 228 หรือ 20% และกำหนด Early stopping rounds เป็น 200 verbose เป็น 10 ซึ่ง

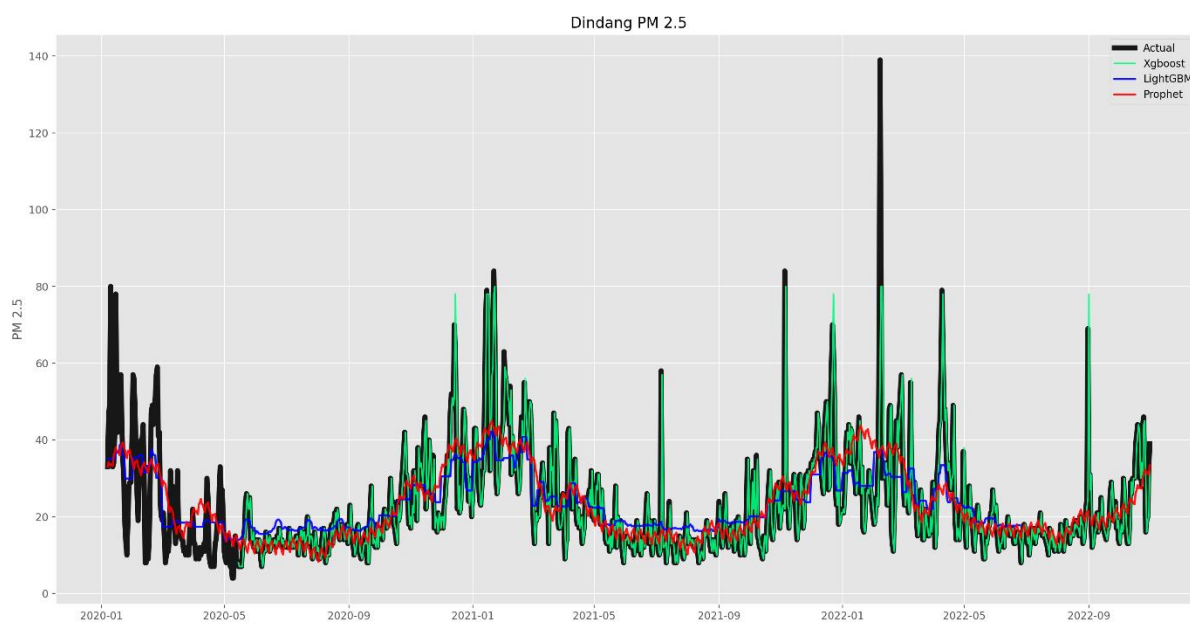
ค่าจากการวัดประสิทธิภาพออกมา จะได้ค่า RMSE เป็น 9.24842 กับ MSE เป็น 85.3335 ซึ่งถือเป็นอันดับที่ 3 ของในทั้ง 4 แบบจำลอง

3. Prophet

จากการสร้างแบบจำลอง Prophet โดยใช้ข้อมูลภูมิอากาศจากกรมอุตุนิยมวิทยา และข้อมูลมลพิษของเขตดินแดง จากสำนักสิ่งแวดล้อม กรุงเทพมหานคร เพื่อทำนาย 7 วันล่วงหน้า โดยใช้ตัววัดประสิทธิภาพแบบจำลอง คือ ค่า RMSE กับ MSE โดยกำหนดให้ periods = 180 จะได้ค่า RMSE เป็น 10.01462 กับ MSE เป็น 100.29266 ซึ่งถือเป็นอันดับที่ 4 ของในทั้ง 4 แบบจำลอง

4. XGBoost

จากการสร้างแบบจำลอง XGBoost โดยใช้ข้อมูลภูมิอากาศจากกรมอุตุนิยมวิทยา และข้อมูลมลพิษของเขตดินแดง จากสำนักสิ่งแวดล้อม กรุงเทพมหานคร เพื่อทำนาย 7 วันล่วงหน้า โดยใช้ตัววัดประสิทธิภาพแบบจำลอง คือ ค่า RMSE กับ MSE ซึ่งค่าจากการวัดประสิทธิภาพออกมา จะได้ค่า RMSE เป็น 0.00214 กับ MSE เป็น 4.5845e-6 ซึ่งถือเป็นอันดับที่ 1 ของในทั้ง 4 แบบจำลอง



รูปที่ 4.3 กราฟแสดงข้อมูลการทำนายจากโมเดลเปรียบเทียบกับค่าจริง ของทุกอัลกอริทึม ในเขตดินแดง

4.1.4. ผลลัพธ์ประสิทธิภาพของแบบจำลองในการเฉลี่ยทุกเขต

เฉลี่ยรวมทุกเขต มี 9 ตัวแปร ได้แก่ pm10, co, no2, pm25, avgT, humid, pressure-monthly, windDir, windSpd มีข้อมูลทั้งหมด 1,765 ข้อมูล

1. Long short-term memory (LSTM)

จากการสร้างแบบจำลอง LSTM โดยใช้ข้อมูลภูมิอากาศจากกรมอุตุนิยมวิทยา และข้อมูลมลพิษของทุกเขต จากสำนักสิ่งแวดล้อม กรุงเทพมหานคร และกรมควบคุมมลพิษ เพื่อทำนาย 7 วันล่วงหน้า โดยใช้ตัววัดประสิทธิภาพแบบจำลอง คือ ค่า RMSE กับ MSE โดยกำหนดให้ ข้อมูล train เป็น 1,366 ข้อมูล หรือ 80% และ ข้อมูล test เป็น 342 หรือ 20% และกำหนด Epochs เป็น 60 Batch size เป็น 256 ซึ่งค่าจากการวัดประสิทธิภาพออกมา จะได้ค่า RMSE เป็น 2.24704 กับ MSE 5.04921 ซึ่งถือเป็นอันดับที่ 2 ของในทั้ง 4 แบบจำลอง

2. Light Gradient Boosting Machine (Light GBM)

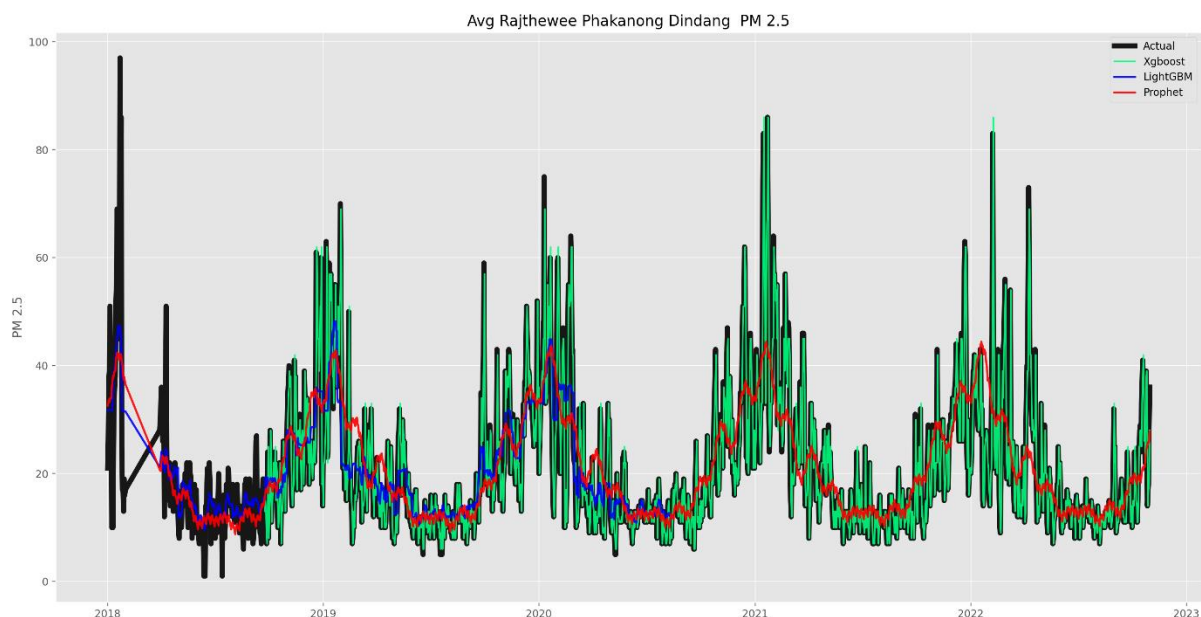
จากการสร้างแบบจำลอง LightGBM โดยใช้ข้อมูลภูมิอากาศจากกรมอุตุนิยมวิทยา และข้อมูลมลพิษของทุกเขต จากสำนักสิ่งแวดล้อม กรุงเทพมหานคร และกรมควบคุมมลพิษ เพื่อทำนาย 7 วันล่วงหน้า โดยใช้ตัววัดประสิทธิภาพแบบจำลอง คือ ค่า RMSE กับ MSE โดยกำหนดให้ ข้อมูล train เป็น 900 ข้อมูล หรือ 90% และ ข้อมูล test เป็น 116 หรือ 10% และกำหนด Early stopping rounds เป็น 200 verbose เป็น 10 ซึ่งค่าจากการวัดประสิทธิภาพออกมา จะได้ค่า RMSE เป็น 9.47088 กับ MSE เป็น 89.69760 ซึ่งถือเป็นอันดับที่ 3 ของในทั้ง 4 แบบจำลอง

3. Prophet

จากการสร้างแบบจำลอง Prophet โดยใช้ข้อมูลภูมิอากาศจากกรมอุตุนิยมวิทยา และข้อมูลมลพิษของทุกเขต จากสำนักสิ่งแวดล้อม กรุงเทพมหานคร และกรมควบคุมมลพิษ เพื่อทำนาย 7 วันล่วงหน้า โดยใช้ตัววัดประสิทธิภาพแบบจำลอง คือ ค่า RMSE กับ MSE โดยกำหนดให้ periods = 180 ซึ่งค่าจากการวัดประสิทธิภาพออกมา จะได้ค่า RMSE เป็น 8.98196 กับ MSE เป็น 80.67570 ซึ่งถือเป็นอันดับที่ 4 ของในทั้ง 4 แบบจำลอง

4. XGBoost

จากการสร้างแบบจำลอง XGBoost โดยใช้ข้อมูลภูมิอากาศจากกรมอุตุนิยมวิทยา และข้อมูลมลพิษของทุกเขต จากสำนักสิ่งแวดล้อม กรุงเทพมหานคร และกรมควบคุมมลพิษ เพื่อทำนาย 7 วันล่วงหน้า โดยใช้ตัววัดประสิทธิภาพแบบจำลอง คือ ค่า RMSE กับ MSE ซึ่งค่าจากการวัดประสิทธิภาพออกมา จะได้ค่า RMSE เป็น 0.05931 กับ MSE เป็น 0.00351 ซึ่งถือเป็นอันดับที่ 1 ของในทั้ง 4 แบบจำลอง



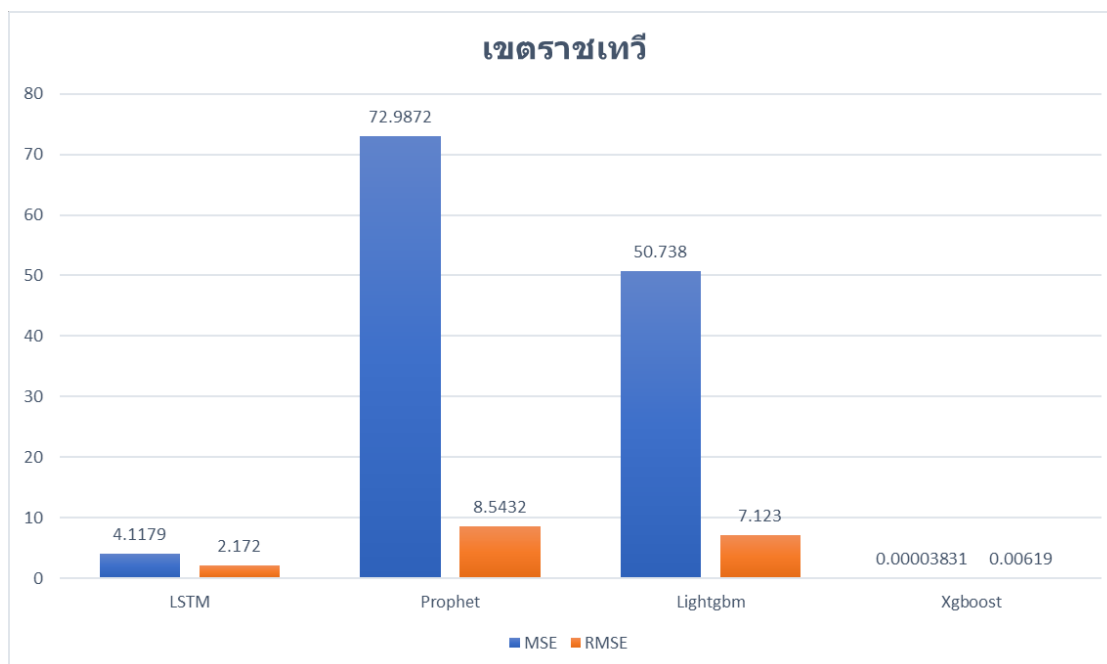
รูปที่ 4.4 กราฟแสดงข้อมูลการทำนายจากโมเดลเปรียบเทียบกับค่าจริง ของทุกอัลกอริทึม ในเฉลี่ยรวมทุกเขต

จากกราฟทั้ง 4 รูปทุกเขตไม่ว่าจะเป็นเขตดินแดง เขตพระโขนง เขตราชเทวีหรือเฉลี่ยรวมทุกเขต อัลกอริทึมที่มีค่าในการเทรนเข้าใกล้ค่าจริงมากที่สุดคืออัลกอริทึม XGBoost

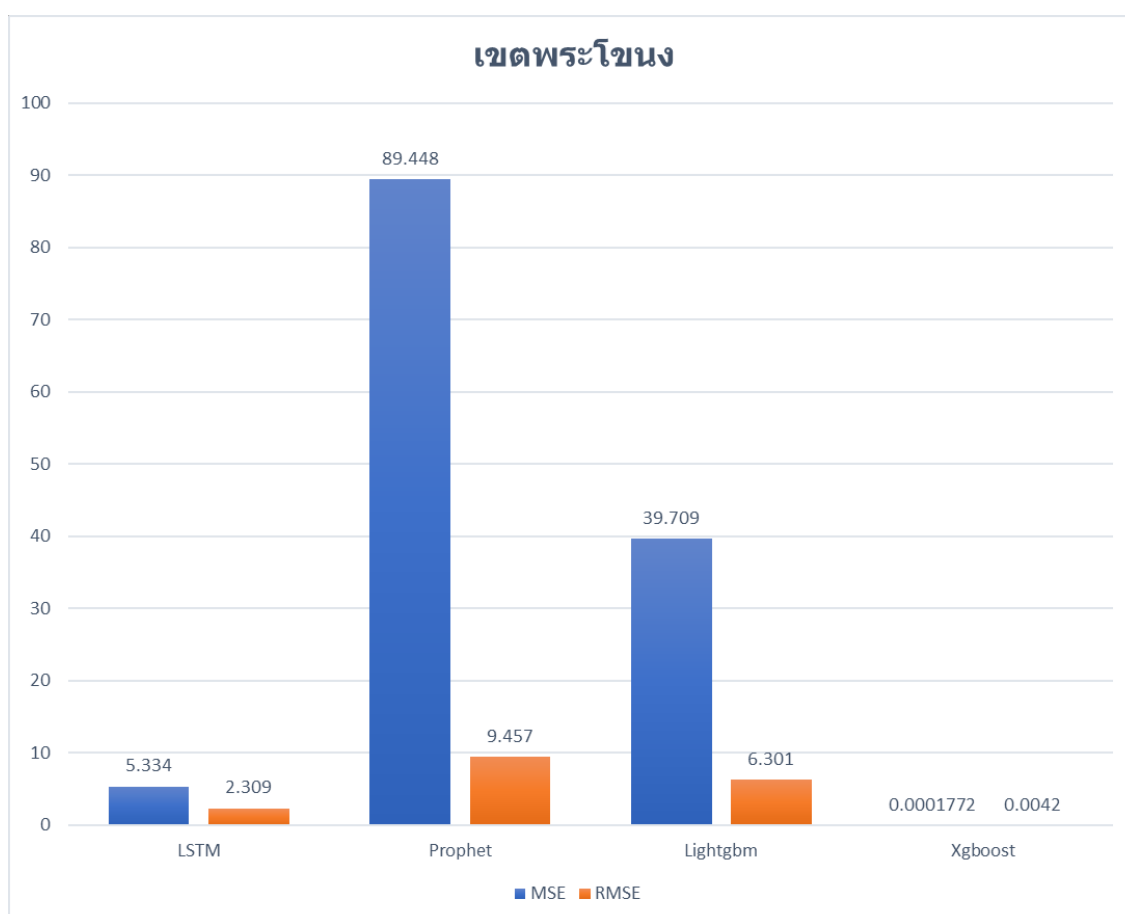
แบบจำลอง	ราชเทวี		พระโขนง		ดินแดง		เฉลี่ยรวมทุกเขต	
	MSE	RMSE	MSE	RMSE	MSE	RMSE	MSE	RMSE
XGBoost	3.8316e-5	0.00619	1.772e-5	0.00420	4.5845e-6	0.00214	0.00351	0.05931
LSTM	4.71785	2.17206	5.33445	2.30964	2.95676	1.71952	5.04921	2.24704
LightGBM	50.73837	7.12309	39.70917	6.30152	85.3335	9.24842	89.69760	9.47088
Prophets	72.9872	8.54326	89.44899	9.45775	100.29266	10.01462	80.67570	8.98196

ตารางที่ 4.1. แสดงผลลัพธ์ของการวัดประสิทธิภาพของแต่ละแบบจำลอง ในข้อมูลของเขตราชเทวี

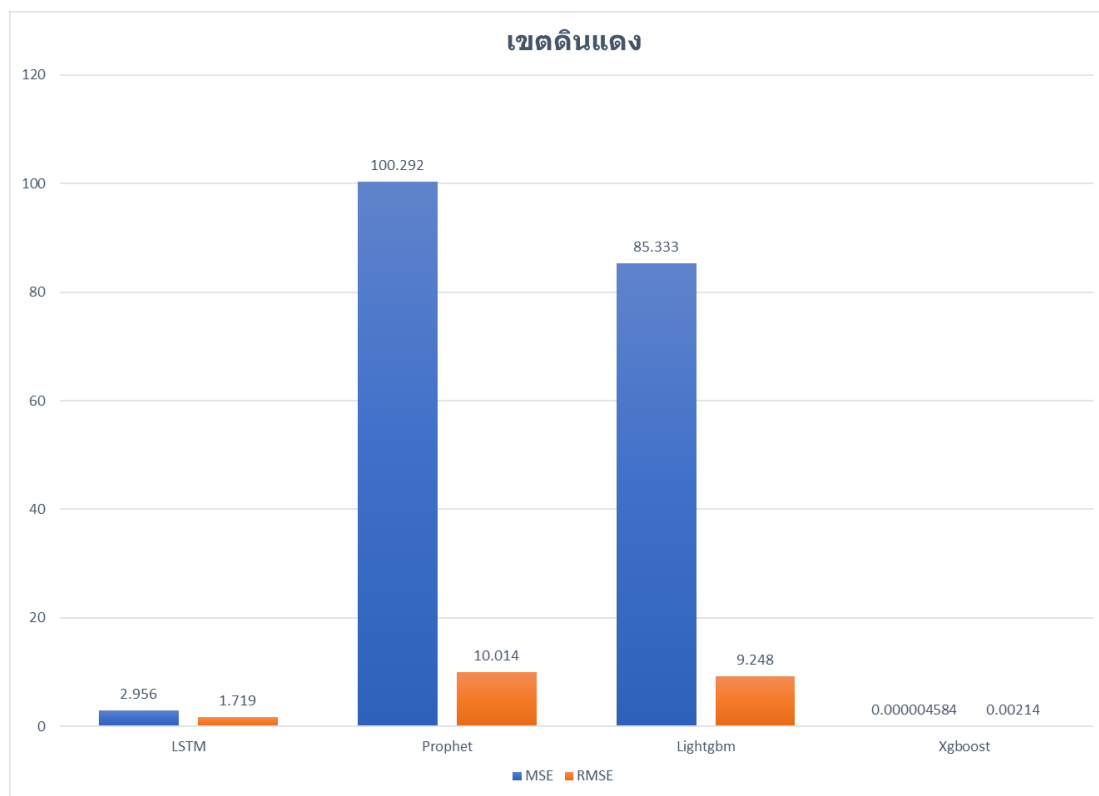
เขตพระโขนง เขตดินแดง และการเฉลี่ยรวมทุกเขต



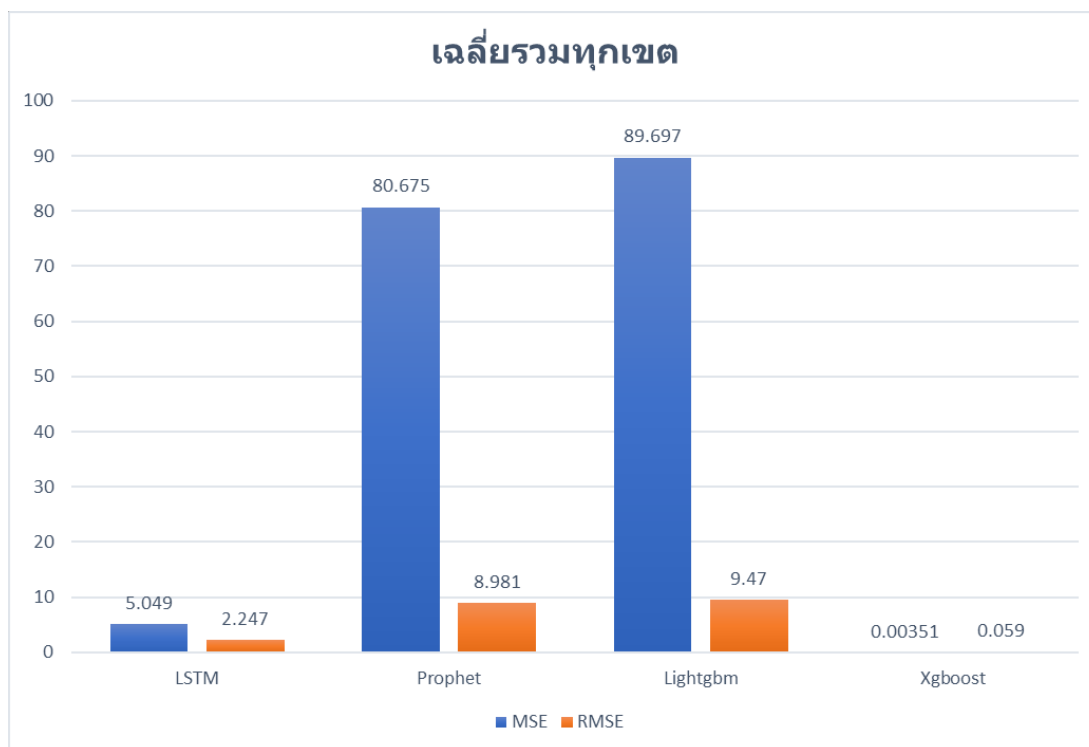
รูปที่ 4.5 กราฟเปรียบเทียบค่าความแม่นยำของเขตราชเทวี



รูปที่ 4.6 กราฟเปรียบเทียบค่าความแม่นยำของเขตพระโขนง

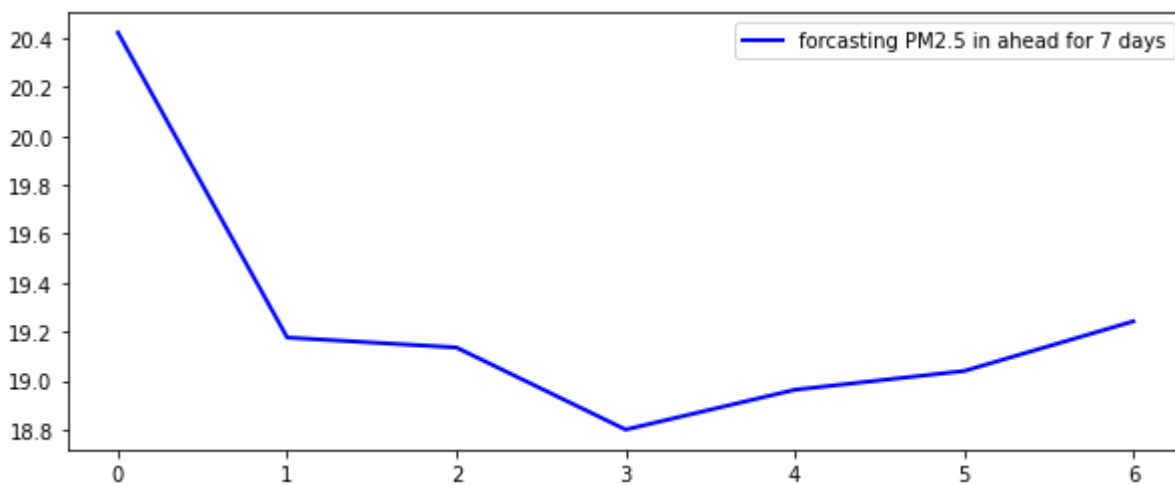


รูปที่ 4.7 กราฟเปรียบเทียบค่าความแม่นยำของเขตดินแดง

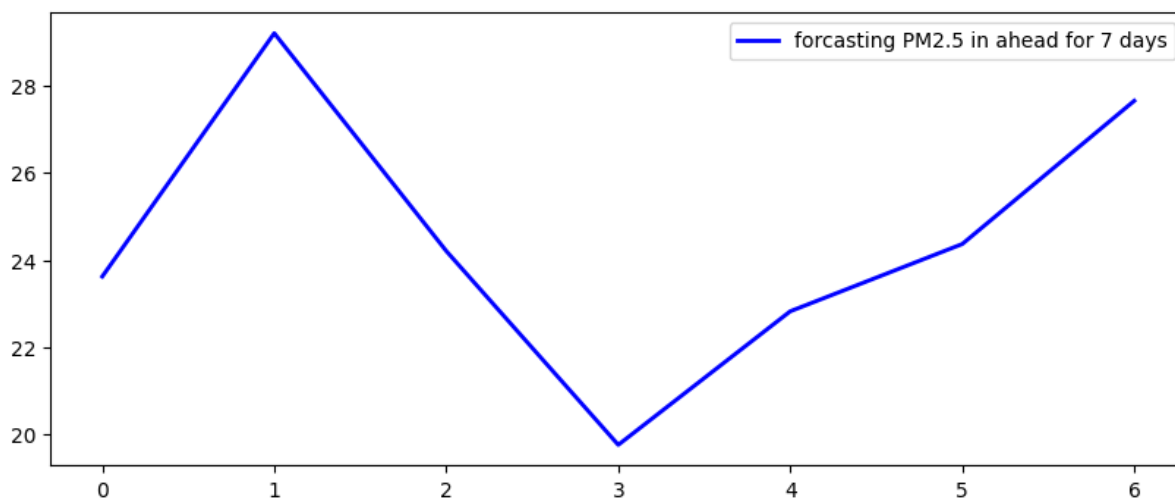


รูปที่ 4.8 กราฟเปรียบเทียบค่าความแม่นยำของเฉลี่ยรวมทุกเขต

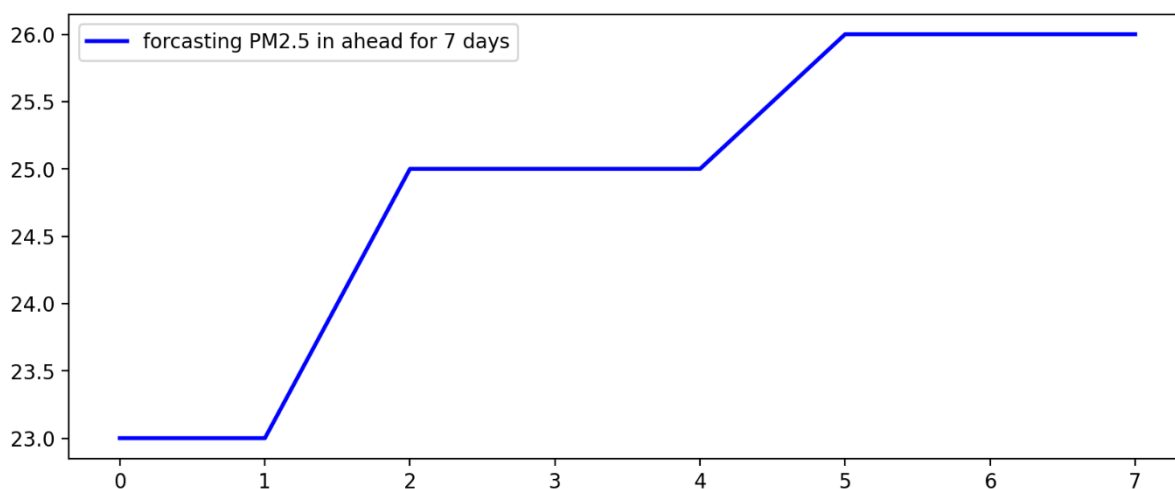
4.2. ผลลัพธ์แนวโน้มที่เป็นกราฟของฝุ่นละอองขนาดเล็กจากการทดสอบแบบจำลอง



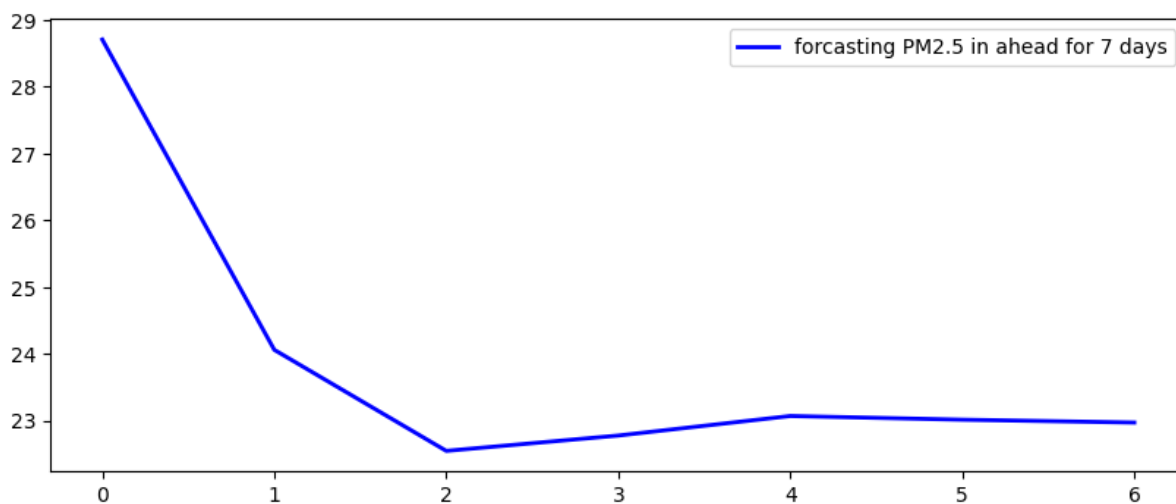
รูปที่ 4.9 ผลการทำนายค่าฝุ่นละอองล่วงหน้า 7 วัน ด้วยอัลกอริทึม LSTM เขตราชเทวี



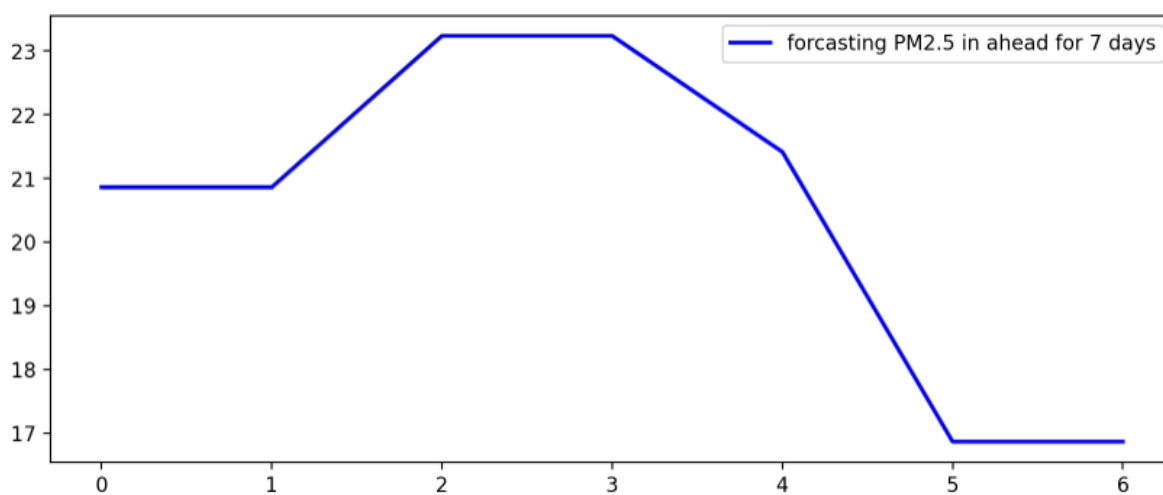
รูปที่ 4.10 ผลการทำนายค่าฝุ่นละอองล่วงหน้า 7 วัน ด้วยอัลกอริทึม LSTM เขตพระโขนง



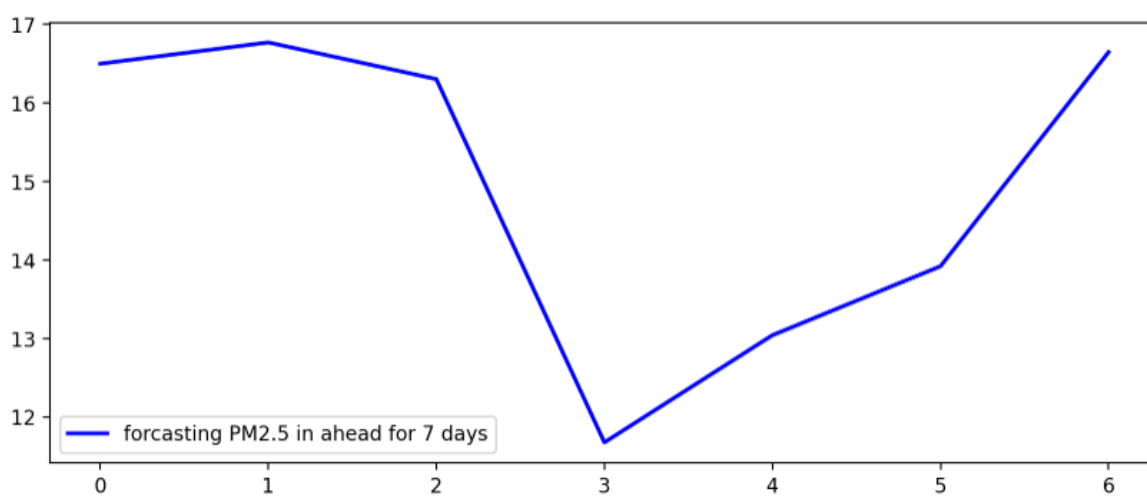
รูปที่ 4.11 ผลการทำนายค่าฝุ่นละอองล่วงหน้า 7 วัน ด้วยอัลกอริทึม LSTM เขตดินแดง



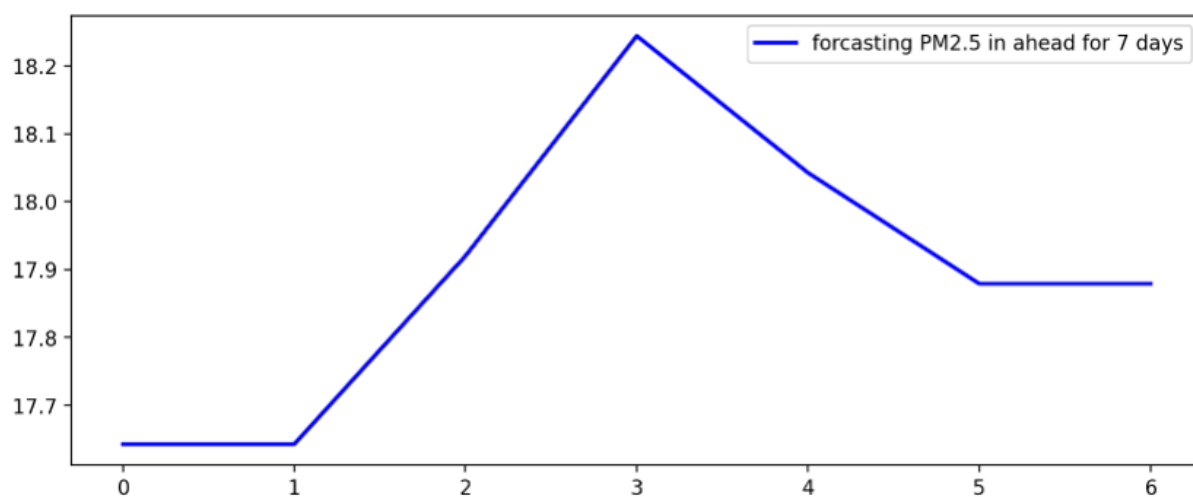
รูปที่ 4.12 ผลการทำนายค่าฝุ่นละอองล่วงหน้า 7 วัน ด้วยอัลกอริทึม LSTM เฉลี่ยรวมทุกเขต



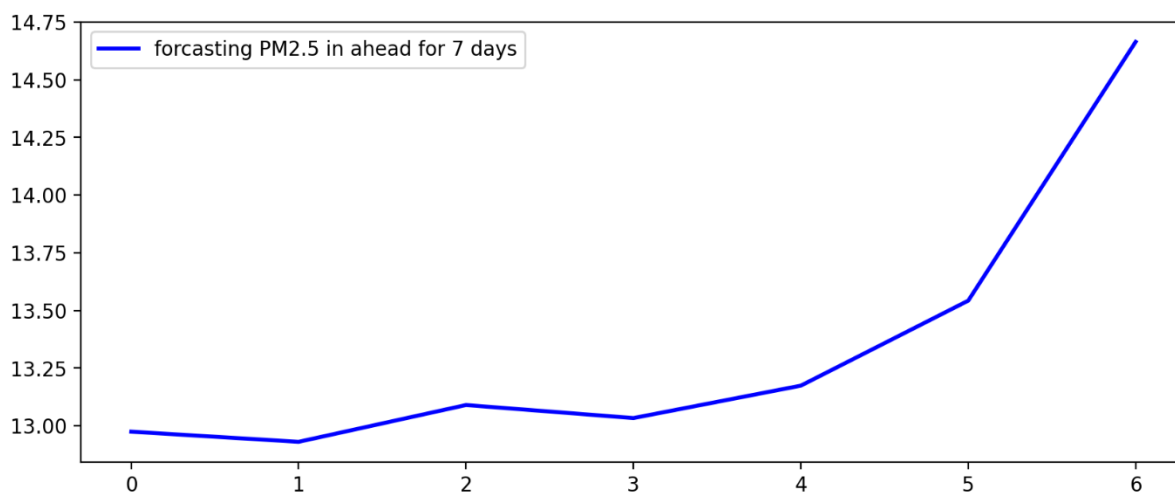
รูปที่ 4.13 ผลการทำนายค่าฝุ่นละอองล่วงหน้า 7 วัน ด้วยอัลกอริทึม LightGBM ราชเทวี



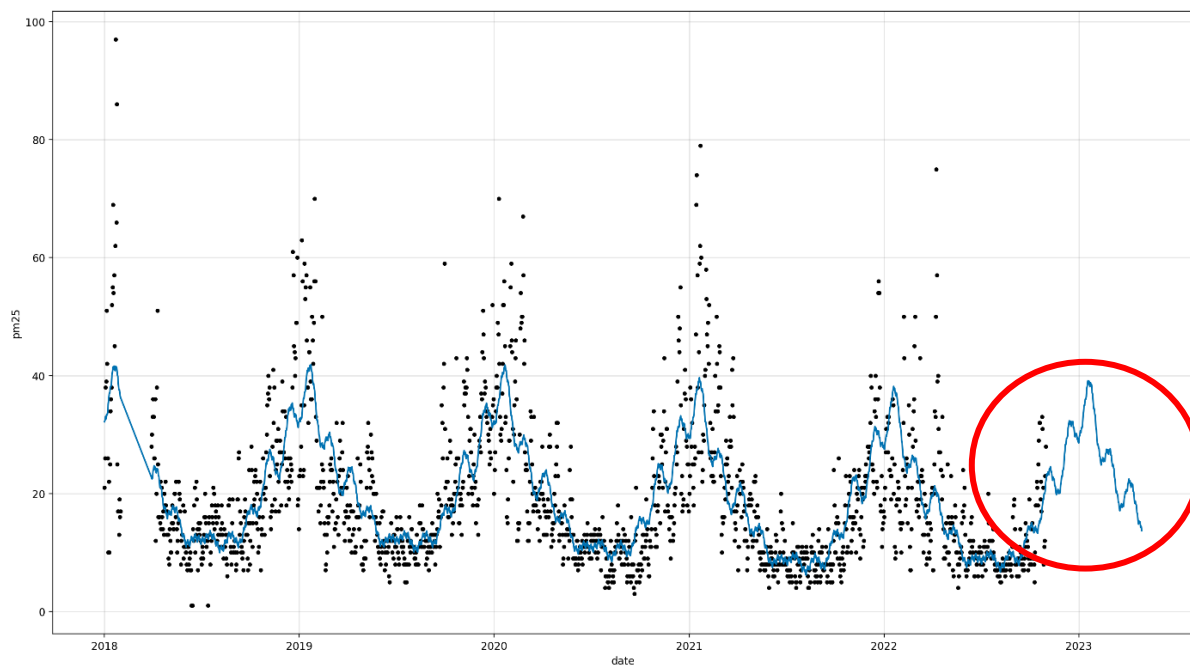
รูปที่ 4.14 ผลการทำนายค่าฝุ่นละอองล่วงหน้า 7 วัน ด้วยอัลกอริทึม LightGBM พระโขนง



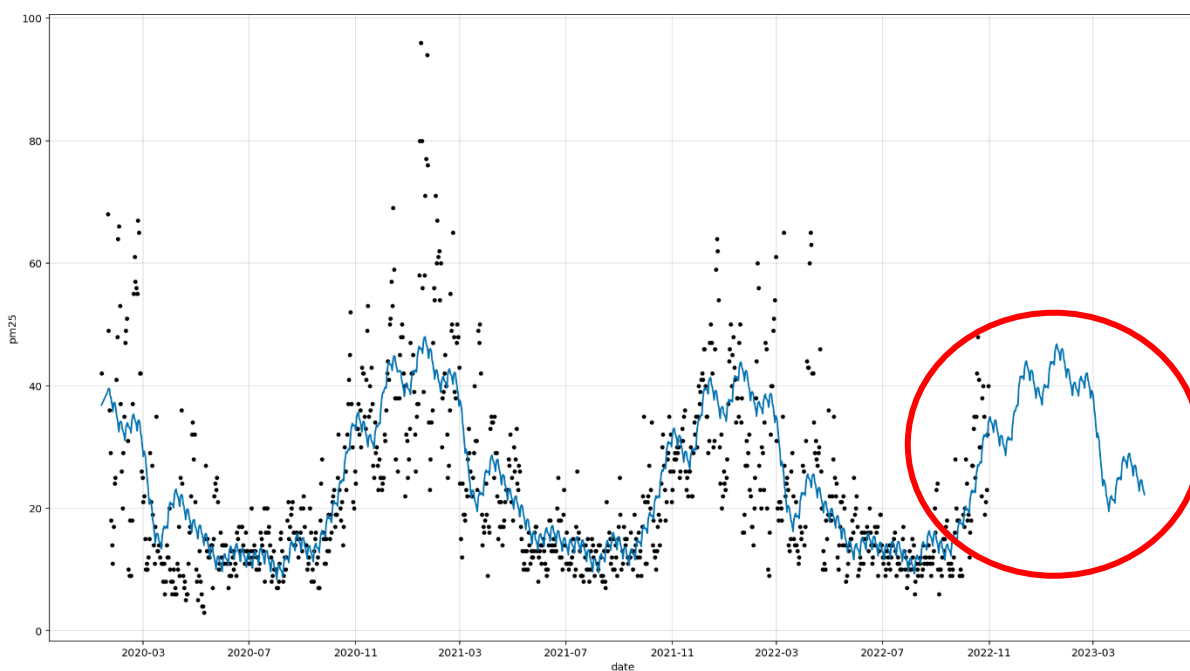
รูปที่ 4.15 ผลการทำนายค่าฝุ่นละอองล่วงหน้า 7 วัน ด้วยอัลกอริทึม LightGBM ดินแดง



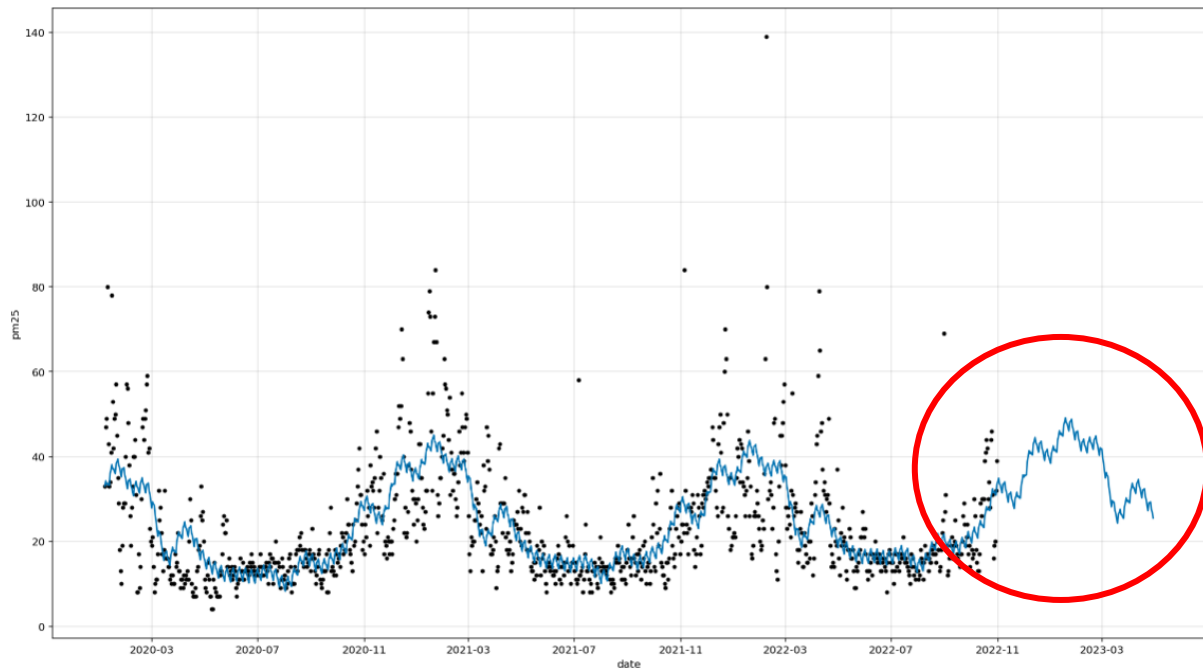
รูปที่ 4.16 ผลการทำนายค่าฝุ่นละอองล่วงหน้า 7 วัน ด้วยอัลกอริทึม LightGBM เฉลี่ยรวมทุกเขต



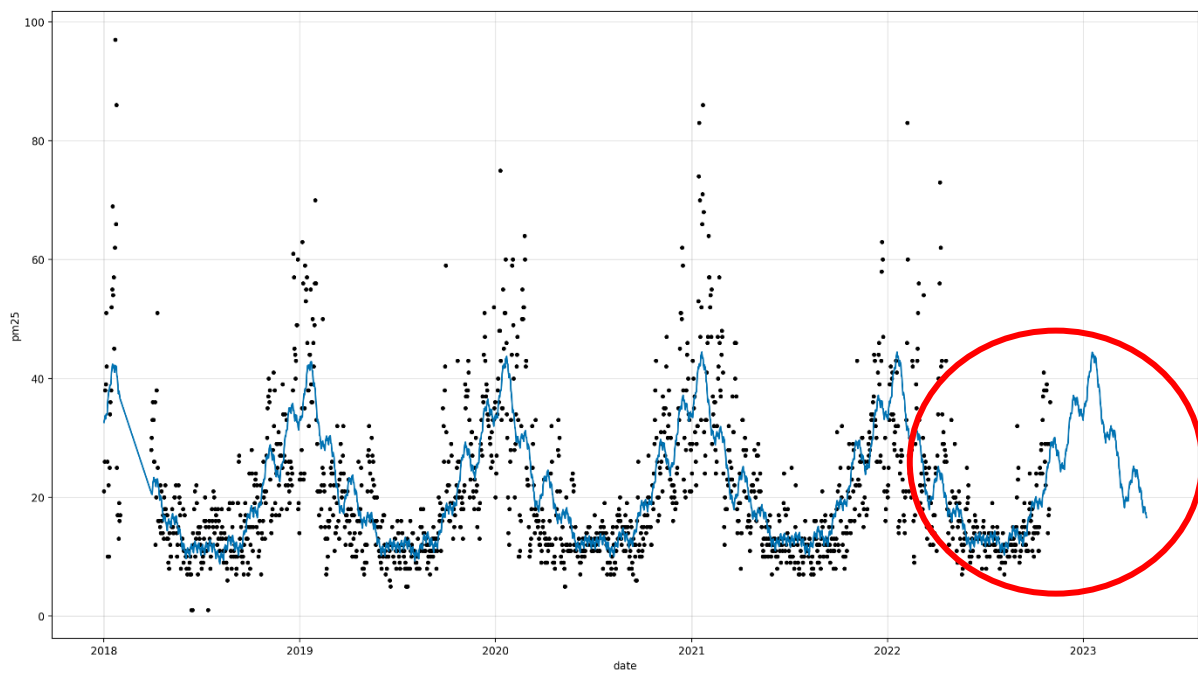
รูปที่ 4.17 ผลการทำนายค่าฝุ่นละอองล่วงหน้า 7 วัน ด้วยอัลกอริทึม Prophet ราชเทวี



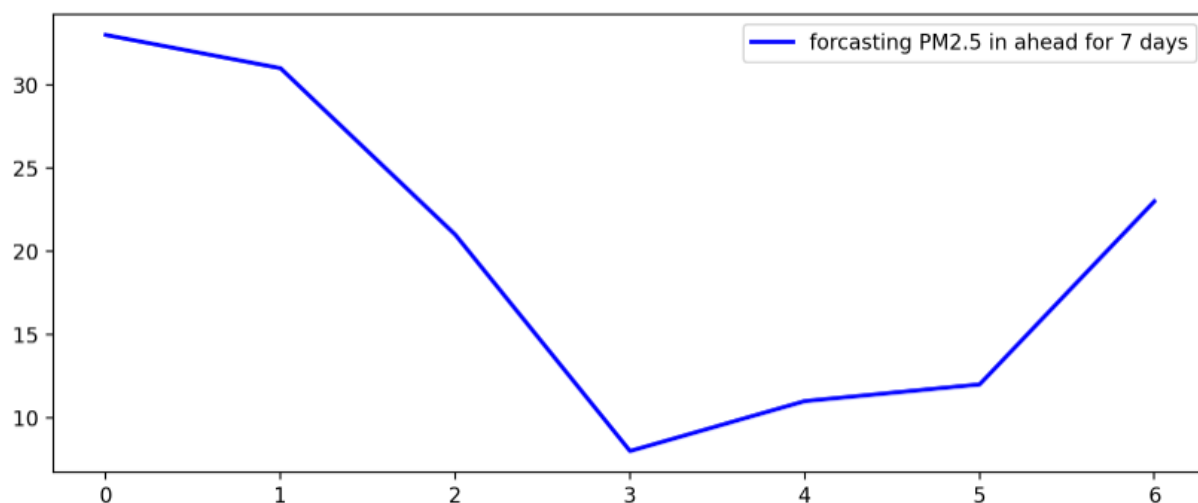
รูปที่ 4.18 ผลการทำนายค่าฝุ่นละอองล่วงหน้า 7 วัน ด้วยอัลกอริทึม Prophet พระโขนง



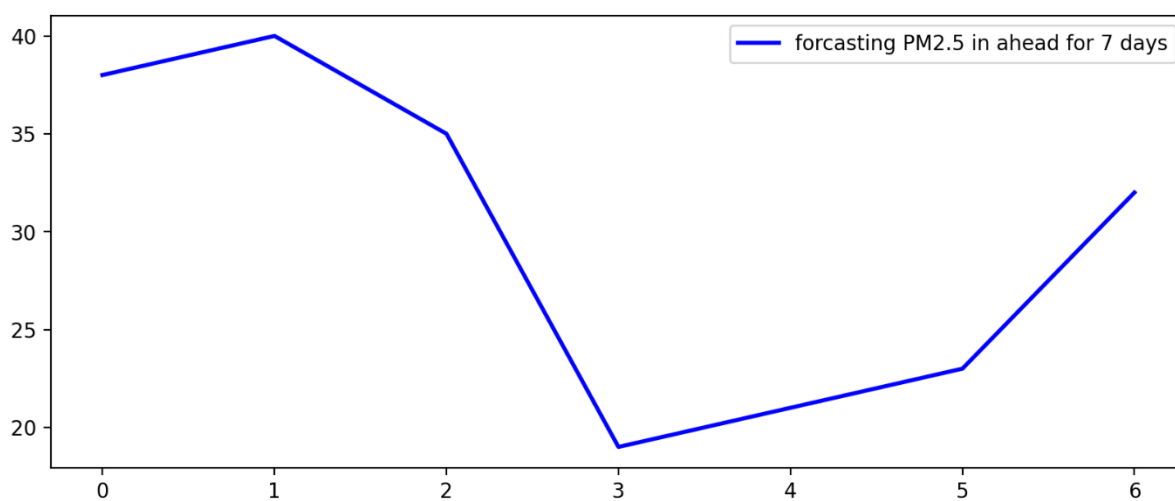
รูปที่ 4.19 ผลการทำนายค่าฝุ่นละอองล่วงหน้า 7 วัน ด้วยอัลกอริทึม Prophet ดินแดง



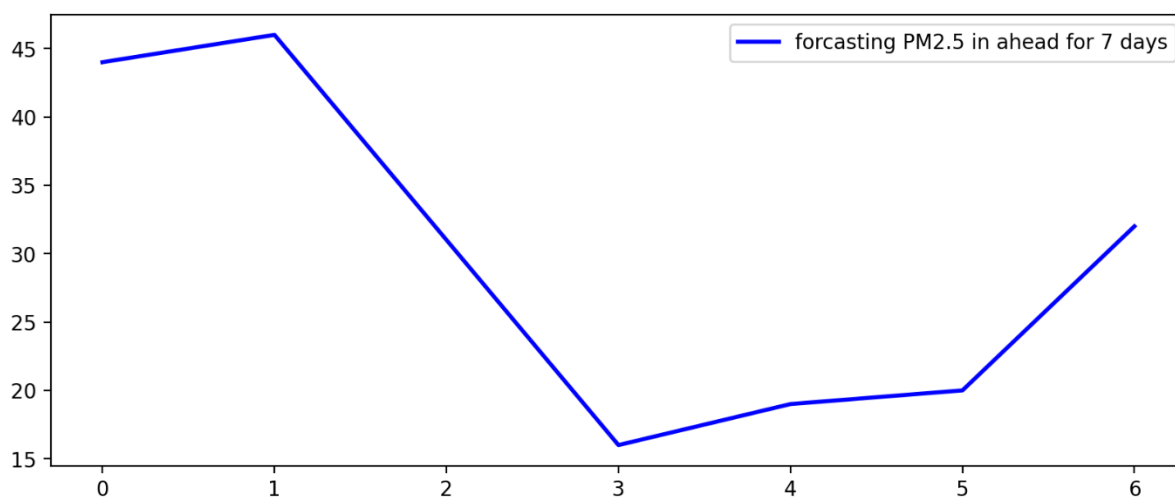
รูปที่ 4.20 ผลการทำนายค่าฝุ่นละอองล่วงหน้า 7 วัน ด้วยอัลกอริทึม Prophet เฉลี่ยรวมทุกเขต



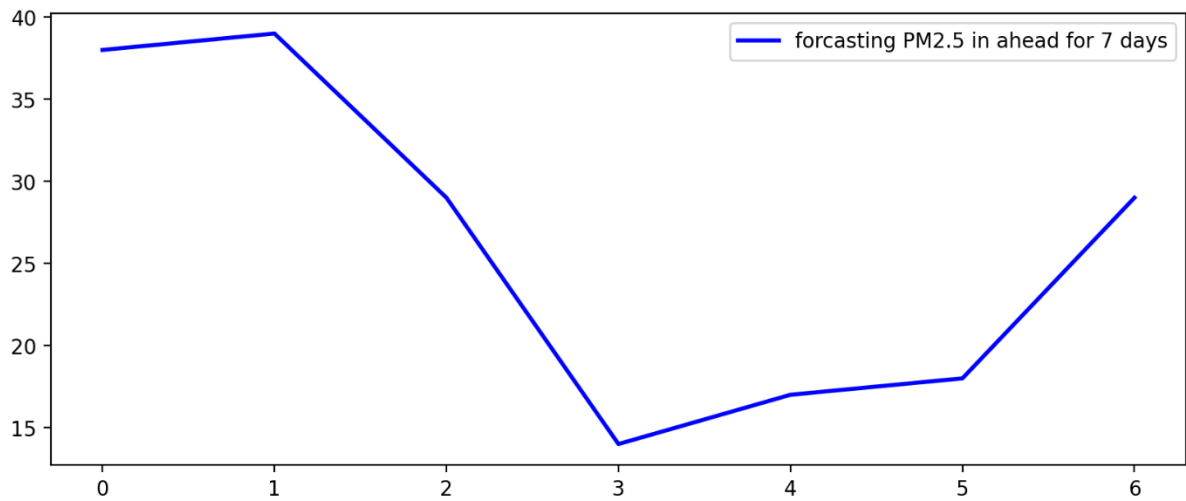
รูปที่ 4.21 ผลการทำนายค่าฝุ่นละอองล่วงหน้า 7 วัน ด้วยอัลกอริทึม XGBoost ราชบุรี



4.22 ผลการทำนายค่าฝุ่นละอองล่วงหน้า 7 วัน ด้วยอัลกอริทึม XGBoost พระโขง



4.23 ผลการทำนายค่าฝุ่นละอองล่วงหน้า 7 วัน ด้วยอัลกอริทึม XGBoost ดินแดง



รูปที่ 4.24 ผลการทำนายค่าฝุ่นละอองล่วงหน้า 7 วัน ด้วยอัลกอริทึม XGBoost เฉลี่ยรวมทุกเขต

บทที่ 5

สรุป อภิปรายผล และข้อเสนอแนะ

ในการวิจัยการพยากรณ์ปริมาณฝุ่นละออง PM2.5 ในเวลา 7 วันล่วงหน้าซึ่งใช้ข้อมูลปริมาณฝุ่นละออง PM2.5 บริเวณเขตราชเทวี เขตพระโขนง และเขตดินแดงโดยใช้อัลกอริทึม LSTM, LightGBM, XGBoost, Prophet ผู้วิจัยได้ทำการทดลองใช้อัลกอริทึมทั้งสี่แบบเพื่อเปรียบเทียบประสิทธิภาพว่าอัลกอริทึมใดสามารถทำงานได้ดีที่สุด

5.1 สรุปผลการวิจัย

โดยการประเมินประสิทธิภาพผู้วิจัยใช้ค่า MSE และ RMSE อัลกอริทึมที่มีประสิทธิภาพมากที่สุด คืออัลกอริทึม XGBoost เนื่องจากมีค่า MSE และ RMSE น้อยที่สุดจากการชุดข้อมูล เขตราชเทวี เขตพระโขนง เขตดินแดง และเฉลี่ยรวมทั้งสามเขต ถัดมาลำดับที่สองคืออัลกอริทึม LSTM ถัดมาลำดับที่สามคือ อัลกอริทึม LightGBM และลำดับสุดท้ายเป็นอัลกอริทึม Prophet

5.2 เปรียบเทียบผลกับงานวิจัยที่เกี่ยวข้อง

This paper : การพยากรณ์ปริมาณฝุ่นละอองขนาดเล็กในอากาศ (PM2.5) ในพื้นที่กรุงเทพมหานคร ด้วยการเรียนรู้ของเครื่อง

ตัวแปรที่ใช้ PM2.5, CO , NO_2 , NO , NO_x , O_3 , PM10, Wind speed, Wind direction, Average Temperature, humidity, Pressure

Paper1 : Prediction System for Fine Particulate Matter Concentration Index by Meteorological and Air Pollution Material Factors Based on Machine Learning

ตัวแปรที่ใช้ PM10, PM2.5, O_3 , NO_2 , CO , SO_2 (ของประเทศเกาหลี)

Paper2 : การพยากรณ์ฝุ่น PM2.5 ล่วงหน้าในจังหวัดเชียงใหม่ โดยใช้แบบจำลองหน่วยความจำ ระยะสั้นแบบยาว

ตัวแปรที่ใช้ NO_2 , NO_x , SO_2 , NO , O_3 , PM10, Wind dir, Temp, Rel hum, Rain, Wind speed, PM2.5

Paper3 : แบบจำลองการพยากรณ์ค่า PM2.5 โดยใช้โครงข่ายประสาทเทียมแบบ LSTM ในพื้นที่กรุงเทพมหานคร

ตัวแปรที่ใช้ได้แก่ NO_2 , CO , PM10, Wind dir, Temp, Rel hum, Press, Wind speed, PM2.5

เปรียบเทียบ This paper, Paper1, Paper2 โดยใช้อัลกอริทึม LSTM

ตารางที่ 5.1 เปรียบเทียบงานวิจัยนี้กับงานวิจัยอื่นของ LSTM

Paper	RMSE
This paper	1.719
Paper2	11.48
Paper3	5.18

สรุปผล จากการใช้อัลกอริทึม LSTM : This paper มีประสิทธิภาพที่ดีมากกว่า Paper1 และ Paper2

เปรียบเทียบ This paper, Paper1 โดยใช้อัลกอริทึม LightGBM

ตารางที่ 5.2 เปรียบเทียบงานวิจัยนี้กับงานวิจัยอื่นของ LightGBM

Paper	RMSE
This paper	7.123
Paper1	6.00

สรุปผล จากการใช้อัลกอริทึม LightGBM : Paper1 ที่มีประสิทธิภาพที่ดีกว่า This paper

5.3 อภิปรายผล

งานวิจัยได้แสดงให้เห็นว่าการสร้างแบบจำลองพยากรณ์ค่าความเข้มข้นของ PM2.5 นั้นต้องพิจารณาช่วงเวลา เนื่องจากในเวลา 1 วัน เป็นช่วงเวลาที่มีความเข้มข้นของฝุ่นละออง PM2.5 ที่มีรูปแบบของรอบการขึ้นลงค่อนข้างชัดเจนทำให้แบบจำลองสามารถเรียนรู้ได้ดีในระยะเวลา 1 วัน ดังนั้นผู้วิจัยจึงเลือกการเรียนรู้ของอัลกอริทึมในรอบ 1 วัน เพื่อพยากรณ์ค่าความเข้มข้นของ PM2.5 ล่วงหน้าเป็นระยะเวลา 7 วัน นอกจากนี้ก็ต้องมีการพิจารณาอัลกอริทึมด้วยอัลกอริทึมไหนเหมาะที่จะนำมาใช้ในการพยากรณ์ค่าความเข้มข้นของ PM2.5 เมื่อพยากรณ์ออกมาแล้วนั้นจะต้องมีประสิทธิภาพมากที่สุด ผู้วิจัยจึงต้องมีการสร้างแบบจำลอง LSTM, LightGBM, XGBoost, Prophet เพื่อพิสูจน์ว่าแบบจำลองชนิดใดเหมาะสมมากที่สุดกับการนำมาใช้พยากรณ์

5.4 ข้อเสนอแนะ

ผู้วิจัยเล็งเห็นว่าการทำการทดลองควรมีปัจจัยเกี่ยวกับสถานการณ์มาใช้ร่วมด้วย เช่น การมีฤดูหนาว วัสดุวัสดุการเกษตรนั้นทำให้ค่าความเข้มข้นของ PM2.5 เพิ่มขึ้นหรือไม่ และข้อมูลที่ได้รับมาบางเขตนั้นมีไม่ครบ ข้อมูลขาดหายไปเป็นจำนวนมากทำให้ข้อมูลของแต่ละเขตมีจำนวนไม่เท่ากันซึ่งทำให้การแบ่งข้อมูลในการ train และ test แตกต่างกันจึงอาจทำให้ประสิทธิภาพนั้นไม่ดีเท่าที่ควร

บรรณานุกรม

- [1] ปรัชญา สิงหรวงศ์. (2563). การพยากรณ์ค่าฝุ่น PM2.5 ล่วงหน้า ในจังหวัดเชียงใหม่โดยใช้หน่วยความจำระยะสั้นแบบยาว(FORECASTING PM2.5 IN CHIANGMAI USING LONG SHOT-TERM MEMORY MODELS). [ออนไลน์]. เข้าถึงได้จาก:
<http://irithesis.swu.ac.th/dspace/bitstream/123456789/960/1/g611130435.pdf>
 (วันที่สืบค้น : 2 ธันวาคม 2565)
- [2] ศรียักษ์ ศรีทองชัย. (2564). แบบจำลองการพยากรณ์ค่า PM2.5 โดยใช้โครงข่ายประสาทเทียมแบบ LSTM ในพื้นที่กรุงเทพมหานคร. [ออนไลน์]. เข้าถึงได้จาก :
<https://ph01.tci-thaijo.org/index.php/TNIJournal/article/download/246987/168721/892399>
 วันที่สืบค้น : (5 ธันวาคม 2565)
- [3] Jihoon Yoo, Dongil Shin, Dongkyoo Shin. (2019). Prediction System for Fine Particulate Matter Concentration Index by Meteorological and Air Pollution Material Factors Based on Machine Learning. [ออนไลน์]. เข้าถึงได้จาก : <https://dl.acm.org/doi/10.1145/3368926.3369684>
 วันที่สืบค้น : (10 ธันวาคม 2565)
- [4] Olah, C. (2015). Understanding LSTM Networks -- colah's blog. Retrieved from <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>
 วันที่สืบค้น : (27 มีนาคม 2566)
- [5] ชนะวิชญ์ พัทธเจริญวงศ์, กันต์กวี เจริญเมฆ, วราภรณ์ กิมปาน. (2563). แบบจำลองการทำนายระยะเวลาในการเข้าเทียบท่าของเรือโดยสารสาธารณะ เข้าถึงได้จาก :
https://li01.tci-thaijo.org/index.php/science_kmitl/article/download/241105/169801/862300
 วันที่สืบค้น : (27 มีนาคม 2566)
- [6] Sean J Taylor, Benjamin Letham. (2017). Forecasting at scale. Retrieved From <https://peerj.com/preprints/3190/>
 วันที่สืบค้น : (27 มีนาคม 2566)
- [7] Manish Pathak. (2019). Using XGBoost in Python. Retrieved from <https://www.datacamp.com/community/tutorials/xgboost-in-python/>
 วันที่สืบค้น : (27 มีนาคม 2566)