

การทำนายค่าความเข้มข้นของฝุ่นละออง (PM2.5) ด้วยวิธี Machine Learning

Prediction of PM2.5 Dust Concentration Using Machine Learning Techniques

*Pradthana Minsan and Panida mahahing**

ภาควิชาคณิตศาสตร์และสถิติ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏเชียงใหม่

Department of Mathematics and Statistics, Faculty of Science and Technology,
Chiang Mai Rajabhat University

Received: April 25, 2025; Revised: June 14, 2025; Accepted: June 24, 2025; Published: June 28, 2025

ABSTRACT – Fine particulate matter (PM2.5) is a significant air pollution issue, particularly in urban areas with high emissions, such as Bangkok, Thailand, and Guangzhou, China—one of the most populous cities in the world. This research focuses on developing a predictive system for PM2.5 concentration using machine learning techniques, including Linear Regression (LR), Support Vector Regression (SVR), and XGBoost, to aid in air pollution monitoring and management. The dataset used in this study is a secondary source containing recorded PM2.5 values from Guangzhou, China, between 2010 and 2015. Experimental comparisons of the three models reveal that XGBoost demonstrates the highest predictive performance across all timeframes. Specifically, for the 1-hour ahead prediction, the XGBoost model incorporating historical PM2.5 averages and seasonal data achieved an R^2 of 0.6728, MAE of 12.06, and RMSE of 17.87, outperforming both LR and SVR. Furthermore, the predictive performance of all models declined as the forecasting timeframe increased, but XGBoost consistently outperformed the other methods in every scenario. The inclusion of seasonal information and historical PM2.5 averages significantly enhanced the model's ability to predict future PM2.5 concentrations.

KEY WORDS: PM2.5, Machine Learning, Linear Regression, XGBoost, Support Vector Regression

บทคัดย่อ -- ฝุ่นละอองขนาดเล็ก PM 2.5 เป็นปัญหามลพิษทางอากาศที่สำคัญ โดยเฉพาะในพื้นที่เมืองที่มีการปล่อยมลพิษสูง โดยเฉพาะเมืองใหญ่ เช่น กรุงเทพมหานคร ประเทศไทย หรือ นครกวางโจว เมืองใหญ่ในสาธารณรัฐประชาชนจีน ซึ่งได้ชื่อว่าเป็นประเทศที่มีประชากรมากที่สุด งานวิจัยนี้จึงมุ่งเน้นการพัฒนาระบบทำนายค่าความเข้มข้นของฝุ่น PM 2.5 ล่วงหน้า โดยใช้เทคนิคการเรียนรู้ของเครื่อง (Machine Learning) ได้แก่ การวิเคราะห์การถดถอยเชิงเส้น Linear Regression (LR) , Support Vector Regression (SVR) และ Extreme Gradient Boosting (XGBoost) เพื่อนำไปใช้ในการเฝ้าระวังและจัดการมลพิษทางอากาศ ข้อมูลในงานวิจัยครั้งนี้ เป็นข้อมูลทุติยภูมิซึ่งบันทึกค่าฝุ่น PM 2.5 ระหว่างปี 2010-2015 ของนครกวางโจว สาธารณรัฐประชาชนจีน จากการทดลองเปรียบเทียบแบบจำลองทั้งสามแบบ พบว่า XGBoost มีประสิทธิภาพสูงสุดในการทำนายค่าฝุ่น PM 2.5 ในทุกช่วงเวลา โดยในกรณีทำนาย 1 ชั่วโมงล่วงหน้า XGBoost ที่ใช้คุณลักษณะค่าเฉลี่ยฝุ่น PM 2.5 ย้อนหลังและข้อมูลฤดูกาล โดยให้ค่า R^2 , MAE และ RMSE เท่ากับ 0.6728 ,12.06 และ 17.87 ตามลำดับ ซึ่งแน่นอนว่า LR และ SVR นอกจากนั้น ประสิทธิภาพของแบบจำลองจะลดลงเมื่อพยากรณ์ช่วงเวลาล่วงหน้าเพิ่มขึ้น ในขณะที่ XGBoost ยังคงให้ผลลัพธ์ที่ดีที่สุดในทุกกรณี โดยการเพิ่มคุณลักษณะของข้อมูลฤดูกาลและค่าเฉลี่ยของฝุ่น PM 2.5 ในอดีตช่วยปรับปรุงความสามารถของแบบจำลองในการทำนายค่าฝุ่น PM 2.5 ได้อย่างมีนัยสำคัญ

คำสำคัญ: PM 2.5 ,การเรียนรู้ของเครื่อง ,การวิเคราะห์การถดถอยเชิงเส้น , Extreme Gradient Boosting ,Support Vector Regression

1. บทนำ

ฝุ่น PM 2.5 เป็นหนึ่งในปัญหามลพิษทางอากาศที่รุนแรงที่สุดในเมืองใหญ่ทั่วโลก โดยเฉพาะ เมืองศูนย์กลางทางเศรษฐกิจและการค้าในเมืองใหญ่ของประเทศต่าง ๆ ปัญหาที่ได้รับ ความสนใจเป็นอย่างมาก ด้วยในเขตเมืองที่มีประชากรหนาแน่น และมีกิจกรรมทางเศรษฐกิจสูง การปล่อยมลพิษจากภาคอุตสาหกรรม การคมนาคมขนส่ง และการใช้พลังงานในครัวเรือนต่างล้วนเป็นปัจจัยที่ก่อให้เกิดฝุ่น PM 2.5 ในปริมาณมาก ฝุ่น PM 2.5 มีขนาดเล็กกว่า 2.5 ไมครอน สามารถเข้าสู่ระบบทางเดินหายใจส่วนลึกและกระแสเลือด ส่งผลกระทบต่อสุขภาพอย่างรุนแรง ไม่ว่าจะเป็นการเพิ่มความเสี่ยงของโรคทางเดินหายใจ เช่น โรคหอบหืด และโรคถุงลมโป่งพอง รวมถึงโรคเรื้อรัง เช่น โรคหัวใจและมะเร็งปอด ดังนั้นความหนาแน่นของประชากรและการใช้ชีวิตในเมืองยิ่งทำให้กลุ่มเสี่ยง เช่น ผู้สูงอายุและเด็ก ได้รับผลกระทบมากขึ้น (กระทรวงสาธารณสุข , 2566)

ข้อมูลการเฝ้าระวังคุณภาพอากาศพบว่า ระดับฝุ่น PM 2.5 ในหลาย ๆ พื้นที่ของเมืองมักเกินค่ามาตรฐานที่กำหนดโดยองค์การอนามัยโลก (WHO) โดยเฉพาะในช่วงฤดูหนาวและฤดูใบไม้ร่วง ซึ่งมีการสะสมของมลพิษสูงขึ้นเนื่องจากลักษณะภูมิอากาศแบบลมสงบ สาเหตุหลักของการเกิดฝุ่น PM 2.5 ในส่วนหนึ่งมาจากการปล่อยก๊าซเสียจากรถยนต์ การเผาไหม้เชื้อเพลิงในโรงงานอุตสาหกรรม การก่อสร้างที่ขาดการควบคุมฝุ่น และการใช้ถ่านหินในกระบวนการผลิตพลังงาน นอกจากนี้ การเปลี่ยนแปลงสภาพภูมิอากาศยังทำให้ฝุ่นละอองลอยตัวในชั้นบรรยากาศเป็นเวลานานขึ้น ส่งผลให้สถานการณ์มลพิษยังซับซ้อน (Carlos Ordóñez , Ruth M. Doherty, 2566)

ผลกระทบจากมลพิษทางอากาศไม่ได้ส่งผลแค่ด้านสุขภาพเท่านั้น แต่ยังส่งผลกระทบต่อเศรษฐกิจและภาพลักษณ์ของเมืองในฐานะศูนย์กลางการท่องเที่ยวและการลงทุนในภูมิภาค การที่คุณภาพอากาศต่ำลงทำให้ค่าใช้จ่ายด้านสุขภาพเพิ่มสูงขึ้น และส่งผลต่อผลิตภาพของแรงงาน รวมถึงการสูญเสียความเชื่อมั่นของนักลงทุนและนักท่องเที่ยว การแก้ไขปัญหฝุ่น PM 2.5 จำเป็นต้องอาศัยความร่วมมือของหลายๆฝ่าย ไม่เพียงเพื่อบรรเทาผลกระทบด้านสุขภาพในระยะสั้น แต่ยังเพื่อสร้างความยั่งยืนในการพัฒนาเศรษฐกิจและคุณภาพชีวิตของประชาชนในระยะยาว การมีข้อมูลที่ถูกต้อง การพัฒนานโยบายที่เข้มงวด และการสร้างความตระหนักรู้ให้กับประชาชนล้วนเป็นกุญแจสำคัญในการเปลี่ยนแปลงอนาคตของเมืองนี้ให้ดีขึ้น จึงได้มีงานวิจัย เพื่อศึกษาการพยากรณ์แนวโน้มฝุ่น PM 2.5 อาทิ เช่น เทคนิคการแยกส่วนประกอบและโหนด-วินเทอร์ ที่ได้รับการปรับปรุงด้วยขั้นตอนวิธีการหา

ค่าที่เหมาะสมที่สุดแบบวาฟ : กรณีศึกษาการพยากรณ์ PM2.5 ใน 8 จังหวัดภาคเหนือของประเทศไทย (ปรารธนา มินเสน , วฐา มินเสน , 2567) และ การเพิ่มประสิทธิภาพการแยกองค์ประกอบและการพยากรณ์รายสัปดาห์ของความเข้มข้น PM2.5 ในแปดจังหวัดภาคเหนือของประเทศไทยโดยใช้อัลกอริทึม Cuckoo Search (Watha Minsan, Pradthana Minsan and Wararit Panichkitkosolkul , 2567) อย่างไรก็ตาม ปัจจุบันเทคโนโลยีทางด้านคอมพิวเตอร์มีการพัฒนาไปอย่างรวดเร็ว โดยเฉพาะการวิเคราะห์ข้อมูล ได้มีการนำ Machine Learning มาใช้ในการวิเคราะห์ข้อมูลที่มีจำนวนมากมหาศาล เพื่อช่วยในการพยากรณ์ หรือตัดสินใจ ดังนั้นในการวิจัยครั้งนี้ ผู้วิจัยจึงใช้เทคโนโลยีขั้นสูง Machine Learning เพื่อช่วยในการศึกษาการพยากรณ์ค่าฝุ่น PM 2.5 ล่วงหน้า โดยใช้ชุดข้อมูลตัวอย่างปริมาณค่าฝุ่น PM 2.5 ที่นครกวังโจว สาธารณรัฐประชาชนจีนเป็นกรณี ศึกษาวิจัยครั้งนี้

2. วิธีดำเนินการวิจัย

ขั้นตอนที่ 1: ชุดข้อมูลที่ใช้ในการศึกษา นำเข้าข้อมูลทุติยภูมิผ่านหน้าเว็บไซต์ www.kaggle.com โดยเป็นข้อมูลปริมาณฝุ่น PM 2.5 ที่นำมาอยู่ในช่วง 1 มกราคม 2010 – 31 ธันวาคม 2015 ค่าข้อมูลที่เก็บมีระยะห่าง 1 ชั่วโมง ที่นครกวังโจว สาธารณรัฐประชาชนจีน จำนวนข้อมูล 52,584 แถว โดยข้อมูลมีความต่อเนื่องและเพียงพอสำหรับสร้างแบบจำลองระยะยาว ข้อมูลครอบคลุมฤดูกาลหลากหลาย

	date	time	PM_2.5	DEWP_C	HUMI_%	PRES_hPa	TEMP_C	Wind	Iws_m/s	precipitation_mm/h
0	1/1/2010	0:00	55	9.4	76	1015.100000	13.5	NW	0.8	0.0
1	1/1/2010	1:00	54	10.2	83	1015.200000	13.0	NW	0.5	0.0
2	1/1/2010	2:00	51	10.4	87	1015.000000	12.5	NW	0.6	0.3
3	1/1/2010	3:00	48	10.2	89	1014.900000	12.0	NW	1.4	0.6
4	1/1/2010	4:00	49	10.4	91	1014.600000	11.8	NE	0.6	0.7
...	...	...	...	...	...	...	...	...	...	...
52579	31/12/2015	19:00	59	11.0	80	1018.299988	14.4	NW	10.9	0.0
52580	31/12/2015	20:00	58	11.4	82	1018.700012	14.5	NW	12.9	0.0
52581	31/12/2015	21:00	60	11.1	78	1019.300000	14.9	NW	14.7	0.0
52582	31/12/2015	22:00	60	10.4	75	1019.500000	14.8	NW	15.9	0.0
52583	31/12/2015	23:00	63	10.4	75	1019.500000	14.8	NW	15.9	0.0

รูปที่ 1. ข้อมูลฝุ่น PM 2.5 ที่นครกวังโจว สาธารณรัฐประชาชนจีน

ขั้นตอนที่ 2: การจัดการข้อมูล

1.สร้าง คอลัมน์ใหม่ ชื่อว่า ‘Season’ เป็นคอลัมน์บอกถึงฤดูกาลของข้อมูลแต่ละแถว ซึ่งฤดูกาลในกวังโจว สาธารณรัฐประชาชนจีน สามารถแบ่งออกได้เป็น 4 ฤดู การเริ่มต้นและสิ้นสุดของฤดูกาล อาจผันแปรไปจากปกติได้ในแต่ละปี ซึ่งในที่นี้ใช้เงื่อนไขดังนี้ ฤดูใบไม้ผลิ (Spring) ระหว่างเดือนกุมภาพันธ์ถึงเดือนมีนาคม ฤดูร้อน (Summer) ระหว่างเดือนเมษายนถึงเดือนกันยายน ฤดูใบไม้ร่วง (Autumn) ระหว่าง

เดือนตุลาคมถึงเดือนพฤศจิกายน และ ฤดูหนาว (Winter) ระหว่างเดือนธันวาคมถึงเดือนมกราคม

2. ทำการ One-Hot Encoding ในคอลัมน์ “Season” โดยแต่ละตัวแปรจะแทนค่าของหมวดหมู่ด้วยตัวเลข 0 หรือ 1 เพื่อให้แบบจำลองเรียนรู้และวิเคราะห์ข้อมูลได้ถูกต้อง ผลลัพธ์ที่ได้จะเป็นตารางใหม่มี 4 คอลัมน์ ดังนี้ Season_Spring, Season_Summer, Season_Autumn, Season_Winter

รูปที่ 2. การ One-Hot Encoding ในคอลัมน์ “Season”

season	Season_Spring	Season_Summer	Season_Autumn	Season_Winter
Winter	0	0	0	1
Winter	0	0	0	1
Winter	0	0	0	1
Winter	0	0	0	1
Winter	0	0	0	1
...	...	...	...	...
Winter	0	0	0	1
Winter	0	0	0	1
Winter	0	0	0	1
Winter	0	0	0	1
Winter	0	0	0	1

3. ทำการ One-Hot Encoding ในคอลัมน์ “Wind” เนื่องจากการประมวลผลของแบบจำลอง ต้องใช้ข้อมูลที่เป็นตัวเลขเท่านั้น ไม่สามารถอ่านค่าที่เป็นตัวอักษรในการคำนวณได้ โดยในคอลัมน์ “Wind” แต่ละตัวแปรจะแทนค่าของหมวดหมู่ด้วยตัวเลข 0 หรือ 1 เพื่อให้วิเคราะห์ข้อมูลได้ถูกต้อง

4. ทำการ Rolling Mean ข้อมูล 24 48 72 ชั่วโมง เพื่อสร้างคอลัมน์ใหม่ เป็นคอลัมน์ Last24hrs_mean, Last48hrs_mean, Last72hrs_mean การ Rolling เป็นการนำข้อมูลในช่วงเวลาหนึ่งมาประมวลผล เช่น หาค่าเฉลี่ย ผลรวมเฉลี่ยรายชั่วโมง

5. สร้างคอลัมน์ใหม่ จากข้อมูลคอลัมน์ “PM 2.5” โดยการ shift (1),(6),(12),(24) ตามลำดับ เพื่อเลื่อนข้อมูลมาใช้ในชั่วโมงย้อนหลัง ซึ่งใช้เป็นค่า PM2.5 ของ 1,6,12,24 ชั่วโมงย้อนหลัง

ได้คอลัมน์ ดังนี้ PM2.5(h-1), PM2.5(h-6), PM2.5(h-12), PM2.5(h-24)

6. สร้างคอลัมน์ใหม่ จากข้อมูลคอลัมน์ “PM2.5” ด้วยเช่นกัน โดยการ shift (-1), (-6), (-12), (-24) ตามลำดับ เพื่อเลื่อนข้อมูลมาใช้ในชั่วโมงล่วงหน้า 1,6,12,24 ชั่วโมงตามลำดับ ซึ่งคอลัมน์ใหม่ที่ได้จากการขยับคอลัมน์ “PM 2.5” ดังนี้ PM2.5(h+1), PM2.5(h+6), PM2.5(h+12), PM2.5(h+24) เพื่อใช้เป็นข้อมูล Target ที่จะวิเคราะห์ PM 2.5 ในชั่วโมงล่วงหน้าถัดไป

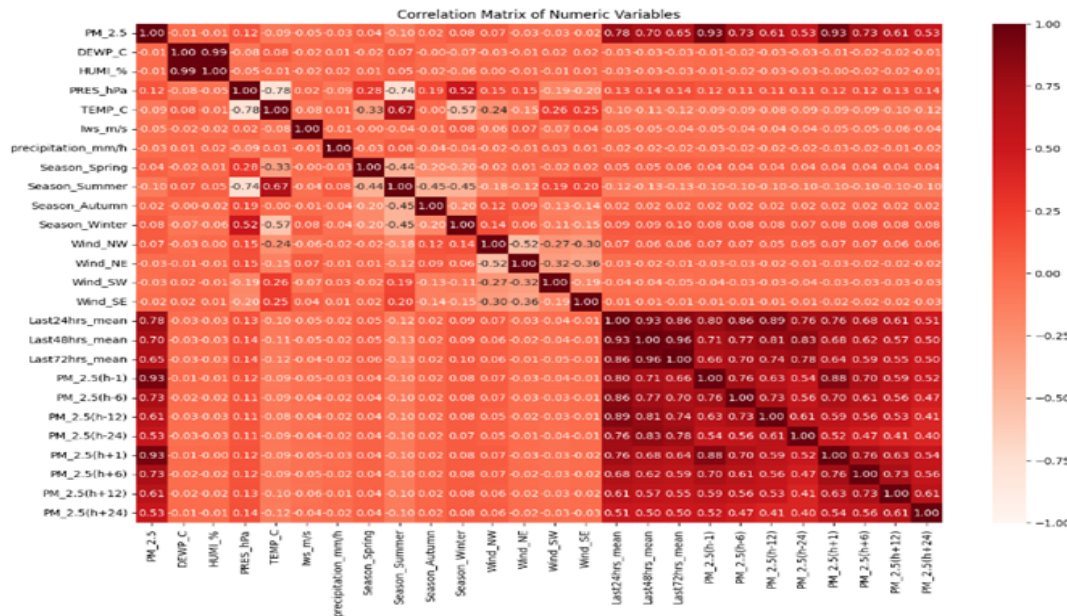
7. จัดการข้อมูลสูญหาย โดยการ Fill ซึ่งเป็นเทคนิคในการจัดการข้อมูลที่สูญหายในชุดข้อมูลการเติมค่าข้อมูลที่ขาดหายไป ในที่นี้ใช้ 2 วิธี คือ การ Backward Fill จะเป็นการกรอกค่าสูญหายด้วยค่าที่อยู่ในแถวถัดไป ย้อนกลับมาเติม และ แบบ Forward Fill จะเป็นการกรอกค่าสูญหายด้วยค่าที่อยู่ในแถวก่อนหน้า

3. การสำรวจและวิเคราะห์ข้อมูลเบื้องต้น

ตารางที่ 1. แสดงรายละเอียดคำอธิบายของข้อมูล

ตัวแปร	ความหมาย
PM_2.5	ค่า pm2.5 ($\mu\text{g}/\text{m}^3$)
DEWP_C	จุดน้ำค้าง (c)
TEMP_C	อุณหภูมิ (c)
HUMI_%	ความชื้นสัมพัทธ์ (%)
PRES_hPa	ความกดอากาศ (hPa)
Iws_m/s	ความเร็วลมสะสม (m/s)
Precipitation_mm/h	ปริมาณฝนสะสม (mm/h)
Wind	ทิศทางลม
Season	ฤดูกาล
PM_2.5(h-1),(h-6),(h-12),(h-24)	ค่าฝุ่น PM2.5 1,6,12,24 ชม.ย้อนหลัง
PM_2.5(h+1),(h+6),(h+12),(h+24)	ค่าฝุ่น PM2.5 1,6,12,24 ชม. ล่วงหน้า

วัดค่าความสัมพันธ์ระหว่างตัวแปร ว่ามีการเคลื่อนไหวไปในทิศทางเดียวกัน (Positive Correlation) ทิศทางตรงข้าม (Negative Correlation) หรือไม่มีความสัมพันธ์ (No Correlation)

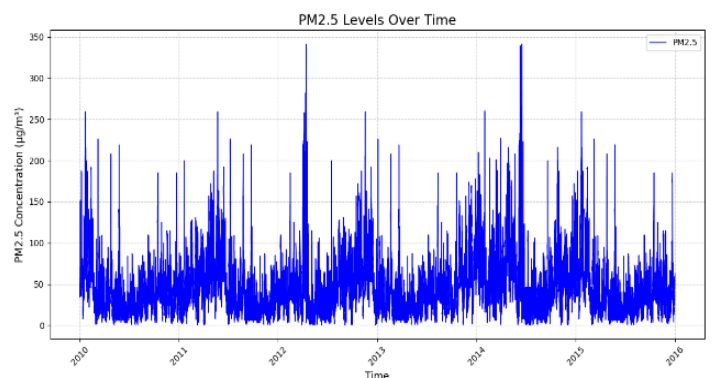


รูปที่ 3. ค่าเมทริกซ์ความสัมพันธ์ระหว่างตัวแปรต่างๆ ในชุดข้อมูล

แผนภาพ Matrix สหสัมพันธ์ (Correlation Matrix) ใช้ในการตรวจสอบความสัมพันธ์ระหว่างตัวแปรเชิงปริมาณทั้งหมดในชุดข้อมูล ซึ่งค่าสหสัมพันธ์ (Correlation Coefficient) มีค่าตั้งแต่ -1 ถึง 1 โดยแสดงเป็นสีเข้มตามระดับความสัมพันธ์ ผลการวิเคราะห์พบว่า ตัวแปร PM2.5 มีความสัมพันธ์ทางบวกในระดับสูงกับค่า PM2.5 จากช่วงเวลาก่อนหน้านี้ เช่น PM2.5(h-1), PM2.5(h-2), PM2.5(h-3) และค่าเฉลี่ยย้อนหลัง 24–72 ชั่วโมง แสดงถึงลักษณะของความต่อเนื่องตามเวลา (temporal dependence) ซึ่งสามารถนำมาใช้เป็น feature สำคัญในการพยากรณ์ค่า PM2.5 ในอนาคต

นอกจากนี้ พบว่าตัวแปรด้านสภาพอากาศ เช่น อุณหภูมิ (TEMP_C), ความชื้นสัมพัทธ์ (HUM_RH), และจุดน้ำค้าง (DEWP_C) มีความสัมพันธ์กับค่า PM2.5 ในระดับปานกลาง ซึ่งสามารถใช้ร่วมในการปรับปรุงประสิทธิภาพของโมเดลได้ ในทางกลับกัน ตัวแปรฤดูกาล และทิศทางลม มีค่าสหสัมพันธ์ต่อกับค่า PM2.5 ซึ่งแสดงว่าอาจไม่ได้มีผลกระทบโดยตรง แต่สามารถนำมาใช้เป็นตัวแทนของบริบทเชิงสิ่งแวดล้อมหรือวงจรมลพิษตามฤดูกาลได้

จากการวิเคราะห์นี้จึงสามารถคัดเลือกตัวแปรที่มีความสัมพันธ์สูงมาใช้เป็น feature ในการสร้างโมเดลพยากรณ์ค่าฝุ่น PM2.5 ได้อย่างมีประสิทธิภาพ



รูปที่ 4. ปริมาณฝุ่นละอองขนาดเล็ก PM2.5 ในปี 2010 – 2015

จากการวิเคราะห์ข้อมูลจากกราฟ ค่าฝุ่น PM2.5 มีแนวโน้มผันผวนสูง มีการขึ้นลงของค่า PM2.5 อย่างชัดเจนในแต่ละช่วงเวลา และพบช่วงที่ค่าฝุ่นพุ่งสูงเป็นพิเศษในบางปี เช่น 2012 และ 2014 นอกจากนี้

ค่าฝุ่น PM2.5 มีลักษณะเป็นคลื่นซ้ำๆ ตามฤดูกาล ช่วงปลายปีถึงต้นปีถัดไปมักมีค่าฝุ่นสูงขึ้น ซึ่งเกิดจากสภาพอากาศในฤดูหนาว ที่มีลมพัดมา ทำให้ฝุ่นสะสมได้มากขึ้น ฤดูร้อนอาจมีค่าฝุ่นต่ำกว่าเล็กน้อย เนื่องจากการกระจายตัวของฝุ่นดีขึ้นจากลมแรงและฝนตก ช่วงที่มีค่าฝุ่นสูงมากผิดปกติ (Peak Events) จุดพุ่งสูงที่สังเกตได้ในบางปี เช่น ปี 2012 และ ปี 2014 ที่ค่าฝุ่นพุ่งเกิน 300 $\mu\text{g}/\text{m}^3$ อาจเกิดจากเหตุการณ์เฉพาะ เช่น ไฟป่า, การเผาไหม้เชื้อเพลิง, หรือสภาวะอากาศที่ทำให้ฝุ่นไม่กระจาย

4. การสร้างแบบจำลอง

ในการวิจัยนี้ ได้ใช้แบบจำลอง ทั้งหมด 3 แบบ ได้แก่ LR (Linear Regression), SVR (Support Vector Regression) และ XGBoost (eXtreme Gradient Boosting) จากนั้นเริ่มสร้างแบบจำลองในการทำนายค่า ในการสร้างแบบจำลองเพื่อทำนายค่าฝุ่นละออง PM2.5 ส่วนหน้า งานวิจัยนี้ได้กำหนด ตัวแปรตาม (Dependent Variable) ให้เป็นค่าฝุ่น PM2.5 ($\mu\text{g}/\text{m}^3$) ในช่วงเวลาช่วงหน้า โดยใช้ค่าจากคอลัมน์ PM2.5(h+1), PM2.5(h+6), PM2.5(h+12) และ PM2.5(h+24) ซึ่งแสดงถึงค่าที่ต้องการทำนายในอนาคตอีก 1, 6, 12 และ 24 ชั่วโมงตามลำดับ

ตัวแปรอิสระ (Independent Variables) คือชุดข้อมูลคุณลักษณะที่มีผลต่อการเปลี่ยนแปลงของค่าฝุ่น PM2.5 ในอนาคต โดยประกอบด้วยข้อมูลด้านอุตุนิยมวิทยา เช่น อุณหภูมิ (c) จุดน้ำค้าง (c) ความชื้นสัมพัทธ์ (%) ความกดอากาศ (hPa) ความเร็วลมสะสม (m/s) ปริมาณฝนสะสม (mm/h) ค่าฝุ่น PM2.5 1,6,12 และ 24 ชม.ย้อนหลัง และ ตัวแปรฤดูกาล 4 ค่า ได้แก่ ฤดูใบไม้ผลิ (Spring) ระหว่างเดือนกุมภาพันธ์ถึงเดือนมีนาคม ฤดูร้อน (Summer) ระหว่างเดือนเมษายนถึงเดือนกันยายน ฤดูใบไม้ร่วง (Autumn) ระหว่างเดือนตุลาคมถึงเดือนพฤศจิกายน และ ฤดูหนาว (Winter) ระหว่างเดือนธันวาคมถึงเดือนมกราคม เพื่อให้แบบจำลองสามารถเรียนรู้และประมวลผลได้อย่างมีประสิทธิภาพ มีการทำ Normalization กับข้อมูลชุดนี้ โดยใช้ทั้ง Standard Scaling และ Min-Max Scaling กับคุณลักษณะเชิงตัวเลข เพื่อให้อยู่ในช่วงที่เหมาะสม ลดปัญหาอคติจากขนาดของหน่วยวัดที่แตกต่างกัน และช่วยเพิ่มประสิทธิภาพในการฝึกแบบจำลองในการวิเคราะห์ ข้อมูลถูกแบ่งออกเป็น 2 ชุด

ได้แก่ Train Dataset คือชุดที่ใช้สำหรับให้แบบจำลองเรียนรู้ และอีกชุด Test Dataset เป็นชุดข้อมูลสำหรับทดสอบประสิทธิภาพของแบบจำลอง ส่วนที่ใช้ในการจำลองคือ Train 80% และ Test 20% ใช้เทคนิค Min-Max Scaling และ Standard Scaling กับ features ต่างๆ

1. Linear Regression (LR)

คือ อัลกอริทึมหนึ่งใน Supervised Learning ใช้สำหรับการวิเคราะห์และคาดการณ์ ทำนายค่าของตัวแปรตาม จากตัวแปรอิสระ โดย Linear Regression จะใช้แบบจำลองการทำนายซึ่งแทนด้วย สมการเชิงเส้น เป็นการวาดเส้นตรง (line) บนกราฟ ที่เชื่อมโยงความสัมพันธ์ระหว่างตัวแปรอิสระ X และตัวแปรตาม Y โดยใช้ข้อมูล Training Data พยากรณ์ค่า Y จากค่า X ที่มีอยู่แล้วในการเรียนรู้และปรับค่าพารามิเตอร์ให้เหมาะสม เพื่อทำนายค่าตัวแปรที่ต้องการในข้อมูลอื่น ๆ (Jay Chugh, 2018) โดยสมการของ Linear Regression มีดังนี้

$$y = b_0 + b_1 * x \quad (1)$$

โดยที่ :

y = ตัวแปรตาม (Dependent variable)

x = ตัวแปรอิสระ (Independent variable)

b_0 = ค่าคงที่ (Intercept)

b_1 = ค่า Slope หรือความชันของเส้นตรง

การเลือกใช้ Linear Regression ในงานวิจัยนี้มีวัตถุประสงค์เพื่อใช้เป็น baseline สำหรับเปรียบเทียบกับแบบจำลองที่มีความซับซ้อนมากกว่า แม้ว่า Linear Regression จะเหมาะสมกับข้อมูลที่มีความสัมพันธ์เชิงเส้นและไม่ซับซ้อน แต่ในเชิงปฏิบัติยังสามารถให้ข้อมูลเบื้องต้นที่มีคุณค่าเกี่ยวกับโครงสร้างของข้อมูล PM2.5 ได้

2. Support Vector Regression (SVR) คือเทคนิคการสร้างแบบจำลองสำหรับการทำนายค่าต่อเนื่อง โดยใช้หลักการคล้ายกับ Support Vector Machine (SVM) ที่ใช้ในงานจำแนกประเภท แต่ปรับให้เหมาะกับการพยากรณ์ ตัวแบบจะพยายามหาฟังก์ชันที่ใกล้เคียงกับข้อมูลมากที่สุด โดยกำหนดขอบเขตความคลาดเคลื่อน (epsilon) ที่ยอมรับได้ และใช้เฉพาะข้อมูลที่สำคัญ (support vectors) เพื่อสร้างแบบจำลองที่มีความซับซ้อนน้อยแต่แม่นยำ

โดยสมการของ SVR มีดังนี้:

$$y = wTx + b = \sum_{i=1}^n \alpha_i k(x_i, x) + b \quad (2)$$

โดยที่:

y คือ ค่าที่ต้องการทำนาย, w คือ เวกเตอร์น้ำหนัก (Weight Vector), x คือ เวกเตอร์ข้อมูลเข้า (Input Vector)

b คือ ค่าไบแอส (Bias), n คือ จำนวนข้อมูล (Data Points)

α_i คือ ค่าเวกเตอร์ของ Lagrange Multiplier

$k(x_i, x)$ คือ ฟังก์ชันความสัมพันธ์ระหว่างเวกเตอร์ข้อมูลเข้า ฟังก์ชันความสัมพันธ์ k ที่ใช้กับ SVR สามารถเลือกได้หลากหลาย โดยที่ Gaussian Kernel (RBF Kernel) ถือว่าเป็นที่นิยมมากที่สุด เช่น ข้อมูลที่ไม่เป็นเชิง โดยมีสมการดังนี้:

$$K(x, y) = \exp(-\gamma \|x - y\|^2) \quad (3)$$

โดยที่: x และ y คือ ข้อมูลตัวอย่างแต่ละตัวที่จะนำมาทำการวิเคราะห์ γ คือ พารามิเตอร์ที่ควบคุมความชันของค่าพารามิเตอร์ γ นั้นจะต้องพิจารณาตามปริมาณข้อมูลที่มีอยู่ เนื่องจากการเลือก γ ไม่เหมือนกัน อาจทำให้ผลลัพธ์ที่ต่างกัน SVR เป็นแบบจำลองที่นิยมใช้ในการทำนายข้อมูลที่มีความซับซ้อน ใช้กับตัวแปรตาม ที่เป็น Non-Linear ได้ สามารถใช้งานได้ดีกับชุดข้อมูลที่มีจำนวน ส่วนการใช้ SVR ก็ต้องมีการแยกชุดข้อมูลออกเป็นชุด Train และ Test เพื่อป้องกัน Overfitting และ ไม่สามารถจัดการ Outlier ได้ดี การทำงานของ SVR อาจช้ากว่าแบบอื่นๆ เนื่องจากการเพิ่ม Kernel Function เข้าไปในการประมวลผล

3. Extreme Gradient Boosting (XGBoost) คือการเรียนรู้ในกลุ่ม Ensemble การพัฒนาและเปิดตัวครั้งแรกของ XGBoost ที่การแข่งขัน Machine learning ของ Kaggle ในช่วงปี 2015 ในบรรดา 29 โซลูชันที่ชนะนั้น มีถึง 17 โซลูชัน ที่ใช้ XGBoost (Jason Brownlee, 2021) อัลกอริทึมถูกพัฒนาขึ้นมา เพื่อแก้ปัญหาการ Regression Predict หรือ Classification ที่มี Dataset ขนาดใหญ่ ซึ่งพัฒนาขึ้นจากแนวคิด พื้นฐานเดียวกับ Gradient Boosting โดยใช้แบบจำลอง Boosting ที่เป็น Ensemble Learning Algorithm ที่รวมแบบจำลองหลายๆ แบบจำลองเข้าด้วยกัน เพื่อสร้างแบบจำลองที่มีความแม่นยำสูงกว่าแบบจำลองเดี่ยว ซึ่ง XGBoost จะทำการเรียนรู้จากข้อมูลที่มีอยู่แล้ว และทำการ Optimize Hyperparameter เพื่อให้แบบจำลองทำนายได้อย่างแม่นยำ (บัญชา ปะสิละเตสัง, 2564) XGBoost เป็น Gradient Boosting ที่มีประสิทธิภาพมากขึ้น โดยเพิ่มการประมวลผลแบบขนาน (Parallel Processing)

ป้องกันการเกิด Overfitting ด้วย Regularization แต่สามารถเกิด Overfitting หรือ Underfitting ได้ง่าย หากมีการเลือก Hyperparameter ที่ไม่เหมาะสม มีการจัดการค่า Missing Values ได้อย่างดี และสามารถประมวลผลข้อมูลหลายมิติได้ เช่น ข้อมูลแบบ ช่วงเวลา (Time Series) และข้อมูลแบบที่มีการแบ่งกลุ่ม (Clustering) สำหรับ สมการของ XGBoost ไม่ได้มีสมการเดียวกันที่ใช้ในทุกๆ กรณี เนื่องจากการประมวลผลของ XGBoost เป็นการเพิ่มความซับซ้อนของแบบจำลองโดยใช้ค่าความคลาดเคลื่อน (Loss Function) ที่ต้องการลดลงในแต่ละรอบการเรียนรู้สามารถกำหนด Loss Function และ สามารถกำหนดพารามิเตอร์ที่เหมาะสมเพื่อปรับความซับซ้อนของแบบจำลอง ดังนั้น สมการของ XGBoost จะไม่เท่ากัน ขึ้นอยู่กับแต่ละปัญหาและการกำหนดพารามิเตอร์ของแต่ละแบบจำลอง (Chen & Guestrin, 2016) ในงานวิจัยนี้ XGBoost แสดงให้เห็นถึงความสามารถในการทำนายค่าฝุ่น PM2.5 ล่วงหน้าได้อย่างแม่นยำที่สุด โดยเฉพาะเมื่อมีการเพิ่มคุณลักษณะเชิงเวลา เช่น ค่าเฉลี่ยย้อนหลังของ PM2.5 และ ข้อมูลฤดูกาล

การสร้างโมเดล จะมีการเพิ่มตัวแปรด้านค่าเฉลี่ยฝุ่น PM 2.5 ย้อนหลัง และตัวแปรด้านฤดูกาล (Season) ในข้อมูลนำเข้า Input Data เพื่อใช้ในการสร้างแบบจำลอง

1. Without Mean_PM and Season การสร้างแบบจำลองโดยใช้คุณลักษณะพื้นฐาน ไม่เพิ่มคุณลักษณะค่าเฉลี่ยฝุ่น PM 2.5 ย้อนหลัง และ ไม่เพิ่มตัวแปรของข้อมูลฤดูกาล (Season)

จะมี 9 ตัวแปรได้แก่ DEWP_C, TEMP_C, HUMI_%, PRES_hPa, Wind_NE, Wind_NW, Wind_SE, Wind_SW, Iws_m/s, Precipitation_mm/h

2. Mean_PM Without Season การสร้างแบบจำลองโดยการเพิ่มคุณลักษณะค่าเฉลี่ยฝุ่น PM 2.5 ย้อนหลัง 24, 48, 72 ชั่วโมง ย้อนหลัง จะมีการเพิ่มตัวแปรจาก Without Mean_PM and Season อีก 3 ตัว ได้แก่ Last24hrs_mean, Last48hrs_mean, Last72hrs_mean เป็น 12 ตัวแปร

3. Only Season การสร้างแบบจำลองโดยการเพิ่มคุณลักษณะของข้อมูลฤดูกาล (Season) จะมีการเพิ่มตัวแปรจาก Without Mean_PM and Season อีก 4 ตัว ได้แก่ Season_Winter, Season_Summer, Season_Spring, Season_Autumn เป็น 13 ตัวแปร

4. Mean_PM and Season การสร้างแบบจำลองโดยการเพิ่มคุณลักษณะค่าเฉลี่ยฝุ่นPM 2.5 ย้อนหลัง และข้อมูลฤดูกาล (Season) จะมีการเพิ่มตัวแปรจาก Without Mean_PM and Season อีก 7 ตัว ตัว ได้แก่ Last24hrs_mean , Last48hrs_mean , Last72hrs_mean , Season_Winter , Season_Summer , Season_Spring , Season_Autumn เป็น 16 ตัวแปร

5. การประเมินผล

ในแบบจำลอง Regression มีวิธีการประเมินผลหลักๆ 3 วิธี

1. **R-squared (R^2)** คำนวณระดับความใกล้เคียงระหว่างผลการทำนายกับข้อมูลจริง หรือ การบ่งชี้ระดับความถูกต้องแม่นยำ Accuracy โดยค่า R^2 จะอยู่ในช่วง 0-1 โดยค่าที่ใกล้เคียง 1 หมายถึงแบบจำลองทำนายได้ดี มีสมการ ดังนี้ :

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (4)$$

โดยที่ y_i = ค่าจริง (Actual Value) $\hat{y}_i$ = ค่าที่โมเดลพยากรณ์ได้ (Predicted Value) $\bar{y}$ = ค่าเฉลี่ยของค่าจริง และ n = จำนวนตัวอย่างทั้งหมด

2. **Mean Absolute Error (MAE)** ค่าเฉลี่ยของความคลาดเคลื่อนในการทำนายกับข้อมูลจริง ทุกรายการในชุดข้อมูล ทดสอบ หรือ Test Set ยิ่งค่า MAE น้อยแสดงว่าคลาดเคลื่อนน้อยมีแม่นยำสูง โดยมีสมการ ดังนี้

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (5)$$

โดยที่ y_i = ค่าจริง (Actual Value) $\hat{y}_i$ = ค่าที่โมเดลพยากรณ์ได้ (Predicted Value) และ n = จำนวนตัวอย่างทั้งหมด

3. **Root Mean Squared Error (RMSE)** นำ MSE มาถอดรากที่สอง Square root โดยค่าที่ได้ จะเป็นหน่วยเดียวกันกับค่า y ยิ่งค่า RMSE น้อย แสดงว่า คลาดเคลื่อนน้อย มีแม่นยำสูง โดยมีสมการ ดังนี้

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (6)$$

โดยที่ y_i = ค่าจริง (Actual Value) $\hat{y}_i$ = ค่าที่โมเดลพยากรณ์ได้ (Predicted Value) และ n = จำนวนตัวอย่างทั้งหมด

6.อภิปรายผลการวิจัย

การเปรียบเทียบผลการทำนายค่าฝุ่นPM 2.5 ล่วงหน้าที่ช่วงเวลาต่างๆ ผลเปรียบเทียบการทำนายระดับค่า

ฝุ่นPM 2.5 เป็นเวลา 1,6,12 และ 24 ชั่วโมงล่วงหน้าตามลำดับ

1. ผลการทำนายระดับค่าฝุ่นPM 2.5 เป็นเวลา 1 ชั่วโมงล่วงหน้า

จากผลลัพธ์ประสิทธิภาพของแบบจำลอง 1 ชั่วโมงล่วงหน้า จะเห็นได้ว่าแบบจำลอง XGBoost โดยการเพิ่มคุณลักษณะค่าเฉลี่ยฝุ่นPM 2.5 ย้อนหลัง และข้อมูลฤดูกาล (Mean_PM and Season) ให้ประสิทธิภาพที่ดีที่สุดในการทำนายฝุ่นPM 2.5 โดยมีค่า R^2 Score สูงที่สุดและค่า MAE และ RMSE ต่ำที่สุด ด้วยค่า R^2 Score 0.71, MAE: 11.83, RMSE: 17.46 แบบจำลอง XGB+1 ให้ผลลัพธ์สูงสุด รองลงมาเป็น LR+1 และ SVR+1 ตามลำดับ สังเกตได้ว่าใช้ Without Mean_PM and Season และ Only Season ในการทำนายของทุกแบบจำลอง ได้ผลลัพธ์น้อยที่สุด แสดงให้เห็นการใช้ข้อมูลค่าเฉลี่ยฝุ่นPM 2.5 ย้อนหลัง เป็นองค์ประกอบที่สำคัญในการทำนายค่าฝุ่นPM 2.5 ในชุดข้อมูลนี้ แบบจำลองที่ใช้ข้อมูลค่าเฉลี่ยฝุ่นPM 2.5 ย้อนหลัง เข้ามาเกี่ยวข้อง มีประสิทธิภาพที่ดีกว่าในการทำนาย ดังตารางผลแบบจำลองค่าฝุ่นPM 2.5 ใน 1 ชั่วโมงล่วงหน้า

ตารางที่ 2. ผลการทำนายระดับค่าฝุ่นPM 2.5 ใน 1 ชั่วโมงล่วงหน้า

แบบจำลอง	การประเมิน	Without Mean_PM and Season	Mean_PM Without Season	Only Season	Mean_PM and Season
LR+1	R^2	0.03	0.62	0.03	0.62
	MAE	23.41	13.46	23.35	13.45
	RMSE	32.19	19.98	32.12	19.98
SVR+1	R^2	0.01	0.61	0.02	0.61
	MAE	22.05	13.31	21.91	13.29
	RMSE	32.50	20.11	32.32	20.10
XGB+1	R^2	0.12	0.70	0.14	0.71
	MAE	22.01	11.89	21.70	11.83
	RMSE	30.66	17.52	30.21	17.46

2. ผลการทำนายระดับค่าฝุ่นPM 2.5 เป็นเวลา 6 ชั่วโมงล่วงหน้า

จากผลลัพธ์ประสิทธิภาพของแบบจำลอง 6 ชั่วโมงล่วงหน้า จะเห็นได้ว่าแบบจำลอง XGBoost โดยการเพิ่มคุณลักษณะค่าเฉลี่ยฝุ่นPM 2.5 ย้อนหลัง และข้อมูลฤดูกาล (Mean_PM and Season) ให้ประสิทธิภาพที่ดีที่สุดในการทำนายฝุ่นPM 2.5 โดยมีค่า R^2 Score สูงที่สุดและค่า MAE และ RMSE ต่ำที่สุด ด้วยค่า R^2 Score 0.64, MAE:

13.37, RMSE: 19.27 แบบจำลอง XGB+6 ให้ผลลัพธ์สูงที่สุด รองลงมาเป็น LR+6 และ SVR+6 ตามลำดับ

ตารางที่ 3. ผลการทำนายระดับค่าฝุ่นPM 2.5 ใน 6 ชั่วโมงล่วงหน้า

แบบจำลอง	การประเมิน	Without Mean_PM and Season	Mean_PM Without Season	Only Season	Mean_PM and Season
LR+6	R ²	0.03	0.51	0.04	0.50
	MAE	23.39	15.63	23.32	15.62
	RMSE	32.09	22.75	31.99	22.75
SVR+6	R ²	0.008	0.50	0.02	0.51
	MAE	22.17	15.16	22.05	15.15
	RMSE	32.48	22.64	32.30	22.62
XGB+6	R ²	0.13	0.64	0.15	0.64
	MAE	22.09	13.39	21.73	13.37
	RMSE	30.51	19.33	30.06	19.27

3. ผลการทำนายระดับค่าฝุ่นPM 2.5 เป็นเวลา 12 ชั่วโมงล่วงหน้า

จากผลลัพธ์ประสิทธิภาพของแบบจำลอง 12 ชั่วโมงล่วงหน้า จะเห็นได้ว่าแบบจำลอง XGBoost โดยการเพิ่มคุณลักษณะค่าเฉลี่ยฝุ่นPM 2.5 ย้อนหลัง และข้อมูลฤดูกาล (Mean_PM and Season) ให้ประสิทธิภาพที่ดีที่สุดในการทำนายฝุ่นPM 2.5 โดยมีค่า R² Score สูงที่สุดและค่า MAE และ RMSE ต่ำที่สุด ด้วยค่า R² Score 0.59, MAE: 14.62, RMSE: 20.68 แบบจำลอง XGB+12 ให้ผลลัพธ์สูงที่สุด รองลงมาเป็น LR+12 และ SVR+12 ตามลำดับ

ตารางที่ 4. ผลการทำนายระดับค่าฝุ่นPM 2.5 ใน 12 ชั่วโมงล่วงหน้า

แบบจำลอง	การประเมิน	Without Mean_PM and Season	Mean_PM Without Season	Only Season	Mean_PM and Season
LR+12	R ²	0.0374	0.41	0.04	0.41
	MAE	23.28	17.22	23.24	17.22
	RMSE	31.55	24.66	31.49	24.66
SVR+12	R ²	0.0214	0.42	0.03	0.42
	MAE	22.03	16.59	21.92	16.59
	RMSE	31.81	24.53	31.70	24.51
XGB+12	R ²	0.1288	0.59	0.16	0.59
	MAE	22.00	14.71	21.58	14.62
	RMSE	30.01	20.78	29.56	20.68

4. ผลการทำนายระดับค่าฝุ่นPM 2.5 เป็นเวลา 24 ชั่วโมงล่วงหน้า

จากผลลัพธ์ประสิทธิภาพของแบบจำลอง 24 ชั่วโมงล่วงหน้า จะเห็นได้ว่าแบบจำลอง XGBoost โดยการเพิ่มคุณลักษณะค่าเฉลี่ยฝุ่นPM 2.5 ย้อนหลัง และข้อมูลฤดูกาล (Mean_PM and Season) ให้ประสิทธิภาพที่ดีที่สุดในการทำนายฝุ่นละอองขนาดเล็ก PM2.5 โดยมีค่า R² Score สูงที่สุดและค่า MAE และ RMSE ต่ำที่สุด ด้วยค่า R² Score 0.51, MAE: 15.91, RMSE: 22.22 แบบจำลอง XGB+24 ให้ผลลัพธ์สูงที่สุด รองลงมาเป็น LR+24 และ SVR+24 ตามลำดับ

ตารางที่ 5. ผลการทำนายระดับค่าฝุ่นPM 2.5 ใน 24 ชั่วโมงล่วงหน้า

แบบจำลอง	การประเมิน	Without Mean_PM and Season	Mean_PM Without Season	Only Season	Mean_PM and Season
LR+24	R ²	0.04	0.31	0.04	0.31
	MAE	23.24	18.97	23.22	18.97
	RMSE	31.79	26.49	31.76	26.49
SVR+24	R ²	0.01	0.31	0.02	0.31
	MAE	22.05	18.12	21.94	18.06
	RMSE	32.16	26.44	32.06	26.36
XGB+24	R ²	0.12	0.50	0.15	0.51
	MAE	22.09	16.08	21.69	15.91
	RMSE	30.33	22.47	29.86	22.22

7.สรุปผลการวิจัย

ผลเปรียบเทียบประสิทธิภาพของแบบจำลองทั้งหมด ใน 4 ช่วงเวลา ได้แก่ +1, +6, +12, +24 ชั่วโมงล่วงหน้า พบว่าแบบจำลอง XGBoost ให้ผลลัพธ์ที่ดีที่สุดทั้งในแง่ของความถูกต้องแม่นยำ และความคลาดเคลื่อนที่ต่ำลดลง รองลงมาเป็น แบบจำลอง LR และ SVR ตามลำดับแบบจำลองทุกแบบ ผลลัพธ์จะลดลงตามช่วงเวลาที่มากขึ้นจากการสร้างแบบจำลอง 4 กรณี โดยมีการเพิ่มตัวแปรด้านค่าเฉลี่ยฝุ่นPM 2.5 ย้อนหลัง และตัวแปร ด้านฤดูกาล ข้อมูลนำเข้าเพื่อใช้ในการสร้างแบบจำลอง พบว่าแบบจำลอง Mean_PM and Season ที่มีการเพิ่มคุณลักษณะค่าเฉลี่ยฝุ่นPM 2.5 ย้อนหลัง และข้อมูลฤดูกาลให้ผลลัพธ์ที่ดีที่สุด ทั้งในแง่ของความถูกต้องแม่นยำ และความคลาดเคลื่อนที่ต่ำลดลง จาก ในทุกช่วงเวลาและจาก 4 แบบจำลองที่เพิ่มคุณลักษณะของข้อมูล การใช้ Mean_PM Without Season และ Mean_PM and Season แบบจำลองที่มีค่าเฉลี่ยของฝุ่นPM 2.5 ย้อนหลัง เข้ามาเกี่ยวข้อง จะอยู่ในกลุ่มที่ให้ผลลัพธ์สูงกว่า แบบ Only

Season และ Without Mean_PM and Season ที่ไม่ใช่
ค่าเฉลี่ยของฝุ่นPM 2.5 ย้อนหลัง เข้ามาเกี่ยวข้อง

เอกสารอ้างอิง

- [1] C. Ordóñez and R. M. Doherty, "Modulation of daily PM2.5 concentrations over China in winter by large-scale circulation and climate change," vol. 23, no. 4, 2023.
- [2] W. Fan, L. Xu, and H. Zheng, "Using Multisource Data to Assess PM2.5 Exposure and Spatial Analysis of Lung Cancer in Guangzhou, China," Int. J. Environ. Res. Public Health, vol. 19, no. 5, p. 2629, Feb. 2022.
- [3] J. Chugh, "Types of Machine Learning and Top 10 Algorithms Everyone Should Know," Dec. 13, 2018.
- [4] P. Minsan and W. Panichkitkosolkul, "Enhancing decomposition and Holt-Winters weekly forecasting of PM2.5 concentrations in Thailand's eight northern provinces using the cuckoo search algorithm," Thailand Statistician, vol. 22, no. 4, 2024. pp. 963–985
- [5] P. Minsan and W. Minsan, "Decomposition and Holt-Winters techniques enhanced by whale optimization algorithm: case study of PM2.5 forecasting in 8 northern provinces of Thailand," Thai Sci. Technol. J., vol. 32, no. 6, 2024. pp. 12–34
- [6] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," University of Washington, 2016. [Online]. Available:<https://www.kdd.org/kdd2016/papers/files/rfp0697-chenAemb.pdf>
- [7] กษรรัตน์ นฤพัฒน์ผจง และ นภา แซ่เบ๊., "การเรียนรู้ของเครื่องสำหรับการทำนายค่าความเข้มข้นของฝุ่นละอองขนาดเล็ก (PM2.5)," ปริญญานิพนธ์ วท.ม., คณะวิทยาศาสตร์, มหาวิทยาลัยศรีนครินทรวิโรฒ, 2566.
- [8] กระทรวงสาธารณสุข, "รายงานสรุปสถานการณ์และผลการดำเนินงานด้านการแพทย์และสาธารณสุข กรณีหมอกควันและฝุ่นละอองขนาดเล็ก ปี 2567," กองประเมินผลกระทบต่อสุขภาพ, กรมอนามัย, กระทรวงสาธารณสุข, 2566.