

การพยากรณ์และการจัดการมลพิษ PM 2.5 ในอำเภอบางปะอิน จังหวัดพระนครศรีอยุธยา ด้วยแบบจำลองทางสถิติ และการพัฒนาแดชบอร์ดโดยใช้โปรแกรม Power BI PM2.5 Forecasting and Management in Bang Pa-In District, Phra Nakhon Si Ayutthaya Province Using Statistical Models and Dashboard Development with Power BI

กัญญาณี นามเอี่ยม¹ และอรัญญา ไบศลา²

¹โรงพยาบาลราชธานี โรจนะ 78 หมู่ที่ 3 ตำบลสามเรือน อำเภอบางปะอิน จังหวัดพระนครศรีอยุธยา 13160,

E-mail hr@rajthanee-rojana.com

²บริษัท เอ็ม.ซี.เอส.สตีล จำกัด (มหาชน) 70 หมู่ที่ 2 ตำบลช้างใหญ่ อำเภอบางไทร จังหวัดพระนครศรีอยุธยา 13290,

E-mail Oratai.ba@mcssteel.co.th

วันที่รับบทความ 11 ธันวาคม 2567

วันแก้ไขบทความ 22 ธันวาคม 2567

วันตอบรับบทความ 27 ธันวาคม 2567

บทคัดย่อ

การวิจัยนี้เป็นการศึกษาสถานการณ์ PM 2.5 และหาตัวแบบการพยากรณ์ปริมาณ PM 2.5 รายชั่วโมงและข้อมูลเฉลี่ย 24 ชั่วโมงที่เหมาะสมเพื่อใช้ในการพยากรณ์ปริมาณ PM 2.5 ศึกษาสถานการณ์โรคเฝ้าระวังจาก PM 2.5 อำเภอ บางปะอิน จังหวัดพระนครศรีอยุธยา ด้วยวิธีการระบาดวิทยาเชิงพรรณนา กลุ่มตัวอย่างที่ใช้ ประกอบด้วย ข้อมูลปริมาณ PM 2.5 รายชั่วโมงของจังหวัดพระนครศรีอยุธยา ข้อมูลตั้งแต่วันที่ 1 สิงหาคม พ.ศ. 2567 – 30 พฤศจิกายน พ.ศ. 2567 และข้อมูลจำนวนผู้ป่วยโรคเฝ้าระวังจาก PM 2.5 จากฐานข้อมูลการเฝ้าระวังทางระบาดวิทยา ของสำนักงานสาธารณสุข อำเภอ บางปะอิน จังหวัดพระนครศรีอยุธยา และระบบฐานข้อมูลสุขภาพ (Health Data Centerกระทรวงสาธารณสุข) รวม 7 โรค ข้อมูลตั้งแต่วันที่ 1 สิงหาคม พ.ศ. 2567 – 30 พฤศจิกายน พ.ศ. 2567 วิธีการหาตัวแบบการพยากรณ์ที่เหมาะสมเปรียบเทียบ ระหว่างเทคนิคการพยากรณ์แบบทำให้เรียบ และเทคนิคการพยากรณ์ของบอซซ์-เจนกินส์ส่วน การศึกษาสถานการณ์โรคเฝ้าระวังจาก PM 2.5 อำเภอ บางปะอิน จังหวัดพระนครศรีอยุธยา ใช้สถิติเชิงพรรณนา ค่าสถิติที่ใช้ คือ สัดส่วน อัตราส่วนค่าต่ำสุดและค่าสูงสุด และการวัดความสัมพันธ์ ค่าสถิติที่ใช้ คือ Incidence rate และ Incidencerate ratio ผลการวิจัย พบว่าเทคนิคการพยากรณ์ที่ให้ตัวแบบที่เหมาะสมที่สุด คือ เทคนิคการพยากรณ์ของ บอซซ์-เจนกินส์ การพยากรณ์ปริมาณ PM 2.5 รายชั่วโมงช่วงระยะเวลาที่พยากรณ์ได้ความแม่นยำไม่เกิน 48 ชั่วโมง ตัว แบบที่เหมาะสม คือ ARIMA(1,1,1)(1,1,1), การพยากรณ์ปริมาณ PM 2.5 เฉลี่ย 24 ชั่วโมง ช่วงระยะเวลาที่พยากรณ์ได้ ความแม่นยำไม่เกิน 31 วัน ตัวแบบที่เหมาะสมคือ ARIMA (0,1,3) และจากการศึกษาสถานการณ์ปริมาณ PM2.5 และ

สถานการณ์โรคเฝ้าระวังจาก PM 2.5 พบว่า การสัมผัส PM 2.5 ในช่วงเวลาที่ระดับ PM 2.5 เริ่มมีผลกระทบต่อสุขภาพมีความเสี่ยงต่อการเกิดโรคปอดอักเสบ โรคไข้หวัดใหญ่ และโรคระบบทางเดินหายใจอื่นๆ อย่างมีนัยสำคัญทางสถิติ

คำสำคัญ: การพยากรณ์ PM 2.5, แบบจำลองทางสถิติ, การออกแบบแดชบอร์ด, Power BI

ABSTRACT

This study researched the PM 2.5 levels and developed an appropriate model that can predict both hourly and 24-hour average concentrations of PM 2.5 it also applies descriptive epidemiological methods to study monitoring with respect to the PM2.5 health problems in Bang Pa-In District, Phra Nakhon Si Ayutthaya Province. The dataset includes hourly PM 2.5 in Phra Nakhon Si Ayutthaya from August 1, 2024, through November 30, 2024, and the number of PM 2.5-related diseases obtained from the Epidemiological Surveillance Database, Bang Pa-In District Public Health Office and Health Data Center, Ministry of Public Health in seven types within the period. It applies the Box-Jenkins forecasting method to select the best model, comparing it with smoothing techniques. Descriptive statistics such as proportions, minimum and maximum values, and correlation metrics like incidence rate and incidence rate ratio are used for disease surveillance. In order to identify the best models for forecasting, the Box-Jenkins method was used: ARIMA(1,1,1)(1,1,1) for hourly PM2.5 concentration forecasts was quite accurate up to 48 hours, while ARIMA(0,1,3) for the 24-hour average PM 2.5 showed up to 31 days of accuracy. Moreover, exposure to high levels of PM2.5 significantly increases the risk of respiratory diseases such as influenza, pneumonia, and other respiratory diseases, as can be seen from the analysis of PM 2.5 levels and associated disease surveillance, with statistically significant findings.

Keywords: Keywords: PM 2.5 Forecasting, Statistical Model, Dashboard Design, Power BI

1. บทนำ

ปัญหาฝุ่นละอองขนาดเล็กไม่เกิน 2.5 ไมครอน (PM 2.5) เป็นหนึ่งในปัญหามลพิษทางอากาศที่ร้ายแรงและมีผลกระทบต่อสุขภาพของมนุษย์ในหลายพื้นที่ทั่วโลก รวมทั้งประเทศไทย โดยเฉพาะในพื้นที่ที่มีการพัฒนาอุตสาหกรรมสูง การขยายตัวของเมือง และการใช้ยานพาหนะในปริมาณมาก ปัญหาฝุ่น PM 2.5 เกิดจากการปล่อยมลพิษจากแหล่งต่าง ๆ เช่น การเผาไหม้ของเชื้อเพลิงในเครื่องยนต์ การเผาไหม้ในกระบวนการผลิตทางอุตสาหกรรม การเผาไหม้ในภาคการเกษตร และมลพิษจากกิจกรรมต่าง ๆ ที่เกี่ยวข้องกับการใช้พลังงาน การสัมผัสกับฝุ่น PM 2.5 ในระยะยาวอาจส่งผลให้เกิดโรคทางเดินหายใจเรื้อรัง เช่น หอบหืด โรคหลอดลมอักเสบ และโรคปอดอักเสบ รวมถึงเพิ่มความเสี่ยงในการเกิดโรคหัวใจ โรคหลอดเลือดสมอง และมะเร็งปอด ผลกระทบเหล่านี้มีความรุนแรงต่อระบบทางเดินหายใจและระบบหลอดเลือดในร่างกาย โดยอาจทำให้สุขภาพโดยรวมเสื่อมลงในระยะยาว การมีมาตรการที่มีประสิทธิภาพจะช่วยให้ประชาชนสามารถหลีกเลี่ยงการสัมผัสกับฝุ่น PM 2.5 ได้ในช่วงที่มีมลพิษสูง และลดความเสี่ยงต่อการเกิดโรคจากมลพิษทางอากาศได้อย่างมีประสิทธิภาพ

ปัญหาฝุ่นละอองขนาดเล็กไม่เกิน 2.5 ไมครอน (PM 2.5) ในพื้นที่จังหวัดพระนครศรีอยุธยาเป็นปัญหาที่มีผลกระทบต่อสุขภาพและสิ่งแวดล้อม โดยเฉพาะในอำเภอบางปะอินและพื้นที่อื่น ๆ ที่มีการพัฒนาอุตสาหกรรมและการเกษตรกรรมที่เข้มข้น ซึ่งสร้างมลพิษในอากาศและส่งผลกระทบต่อประชาชนในพื้นที่ จังหวัดพระนครศรีอยุธยาเป็นหนึ่งในจังหวัดที่มีการ

พัฒนาอุตสาหกรรมที่สำคัญ เช่น เขตอุตสาหกรรมบางปะอิน ซึ่งเป็นแหล่งปล่อยมลพิษทางอากาศหลักจากการเผาไหม้เชื้อเพลิงในโรงงานอุตสาหกรรม การเผาไหม้ในกระบวนการผลิต และการขนส่งสินค้าโดยใช้ยานพาหนะที่ปล่อยควันดำ นอกจากนี้ยังมีการเผาไร่อ้อยและการเผาหญ้าในภาคเกษตรกรรมที่เป็นอีกหนึ่งแหล่งกำเนิดฝุ่น PM 2.5 ที่ส่งผลกระทบต่อคุณภาพอากาศในพื้นที่ การจัดการปัญหาฝุ่น PM 2.5 ในจังหวัดพระนครศรีอยุธยายังมีความท้าทาย เนื่องจากมีแหล่งมลพิษจากทั้งภาคอุตสาหกรรมและภาคเกษตรกรรม การเฝ้าระวังและการแจ้งเตือนเกี่ยวกับค่าฝุ่น PM 2.5 ที่เกิดขึ้นในพื้นที่จึงเป็นสิ่งสำคัญ โดยการพัฒนาเทคโนโลยีการตรวจวัดและพยากรณ์การปล่อยมลพิษสามารถช่วยให้หน่วยงานสามารถแจ้งเตือนประชาชนล่วงหน้า และจัดการมลพิษได้อย่างมีประสิทธิภาพ เช่น การห้ามการเผาไหม้ในช่วงเวลาที่ฝุ่นมีความหนาแน่นสูงหรือการแนะนำให้ประชาชนหลีกเลี่ยงการออกจากบ้านในช่วงที่ค่าฝุ่นสูงเกินไป

การพัฒนาโมเดลการพยากรณ์ PM 2.5 สำหรับอำเภอบางปะอิน จังหวัดพระนครศรีอยุธยา จำเป็นต้องใช้ข้อมูลที่มีความละเอียดและความแม่นยำสูง รวมถึงต้องพิจารณาปัจจัยที่มีผลต่อการเกิดมลพิษในพื้นที่นั้น ๆ เช่น การเผาไหม้ในพื้นที่เกษตรกรรม การปล่อยมลพิษจากโรงงานอุตสาหกรรม สภาพอากาศ เช่น ความชื้น อุณหภูมิ และความเร็วลม รวมถึงการจราจรที่มีปริมาณมาก การใช้ข้อมูลจากสถานีตรวจวัดอากาศในพื้นที่ เช่น สถานีตรวจวัดอากาศในเขตบางปะอิน ซึ่งจะช่วยในการเก็บข้อมูลที่แม่นยำเกี่ยวกับระดับ PM 2.5 และการเปลี่ยนแปลงของมันในแต่ละช่วงเวลาได้ การใช้เทคนิคทางสถิติเช่น ARIMA (AutoRegressive Integrated Moving Average) จะช่วยในการพยากรณ์ค่าฝุ่นในระยะสั้น (เช่น 1-48 ชั่วโมง) โดยใช้ข้อมูลที่มีอยู่ในอดีตในการสร้างตัวแบบการพยากรณ์ (Box et al., 2015). การใช้ Machine Learning เช่น Random Forest หรือ Support Vector Machines (SVM) จะช่วยในการวิเคราะห์ข้อมูลจำนวนมากที่มีความซับซ้อน ซึ่งอาจไม่สามารถจัดการได้ดีในโมเดลทางสถิติปกติ การเรียนรู้ของเครื่องสามารถทำให้โมเดลการพยากรณ์มีความแม่นยำสูงขึ้น โดยเฉพาะในกรณีที่มีหลายปัจจัยที่ต้องพิจารณาพร้อมกัน (Breiman, 2001). นอกจากนี้ Air Quality Dispersion Models เช่น AERMOD ยังสามารถใช้ในการจำลองการกระจายตัวของฝุ่น PM 2.5 ในพื้นที่ที่มีการปล่อยมลพิษจากแหล่งต่าง ๆ ช่วยในการพยากรณ์ระดับฝุ่นในพื้นที่ได้แม่นยำมากยิ่งขึ้น (Cimorelli et al., 2005) การศึกษาผลกระทบจากฝุ่น PM 2.5 ในพื้นที่ต่าง ๆ เช่น อำเภอบางปะอินจะช่วยให้เข้าใจถึงความเสี่ยงทางสุขภาพจากฝุ่นละอองในระดับท้องถิ่น ซึ่งสามารถช่วยเพิ่มประสิทธิภาพในการออกแบบนโยบายสุขภาพและการพัฒนาเมืองอย่างยั่งยืน โดยใช้ข้อมูลการพยากรณ์ PM 2.5 ร่วมกับการศึกษาผลกระทบด้านสุขภาพ สามารถช่วยในการออกแบบแผนรับมือภัยพิบัติจากมลพิษทางอากาศและการปกป้องสุขภาพของประชาชนในพื้นที่ได้อย่างมีประสิทธิภาพ

2. วัตถุประสงค์ของการวิจัย

1. เพื่อเปรียบเทียบประสิทธิภาพของต้นแบบในการพยากรณ์ปริมาณฝุ่น PM 2.5 ในอำเภอบางปะอิน จังหวัดพระนครศรีอยุธยา
2. เพื่อวิเคราะห์และพัฒนาแดชบอร์ดการแสดงผลของปริมาณฝุ่น PM2.5 ในอำเภอบางปะอิน จังหวัดพระนครศรีอยุธยา

3. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

การพยากรณ์ PM 2.5 โดยใช้เทคนิคการเรียนรู้ของเครื่อง (Machine Learning) และแบบจำลองทางสถิติ (Statistical Models) เป็นกระบวนการที่สำคัญในการช่วยเตรียมความพร้อมในการจัดการมลพิษทางอากาศ ซึ่งสามารถนำไปใช้ในการเตือนภัยและให้ข้อมูลแก่ประชาชนเกี่ยวกับสถานการณ์ฝุ่น PM 2.5 เพื่อลดผลกระทบต่อสุขภาพ โดยต้องพิจารณาปัจจัยหลายประการ เช่น สภาพอากาศ การปล่อยมลพิษจากแหล่งอุตสาหกรรมและเกษตรกรรม รวมถึงข้อมูล

จากสถานีตรวจวัดอากาศในพื้นที่ที่มีความแม่นยำและละเอียด การพัฒนาโมเดลพยากรณ์ PM 2.5 ด้วยเทคนิคเหล่านี้จะช่วยให้สามารถทำนายระดับฝุ่นในอนาคตได้อย่างมีประสิทธิภาพ ซึ่งสามารถแบ่งได้ดังนี้

การพยากรณ์ PM 2.5 ด้วยแบบจำลองทางสถิติ ARIMA (AutoRegressive Integrated Moving Average): ARIMA เป็นแบบจำลองทางสถิติที่ใช้ในการพยากรณ์ข้อมูลชุดเวลา เช่น PM 2.5 โดยการใช้ข้อมูลในอดีตเพื่อทำนายค่าฝุ่นในอนาคต แบบจำลองนี้สามารถจัดการกับข้อมูลที่มีแนวโน้มและฤดูกาลได้

การพยากรณ์ PM 2.5 ด้วยการเรียนรู้ของเครื่อง (Machine Learning) Random Forest: Random Forest เป็นเทคนิคการเรียนรู้ที่ใช้หลายต้นไม้ (decision trees) ในการตัดสินใจ ซึ่งแต่ละต้นไม้จะเรียนรู้จากข้อมูลที่มีความหลากหลายและรวมผลลัพธ์จากหลาย ๆ ต้นไม้เพื่อเพิ่มความแม่นยำในการทำนาย การใช้ Random Forest จึงสามารถจัดการกับข้อมูลที่ไม่เป็นเชิงเส้น (non-linear) ได้ดี ซึ่งเหมาะสำหรับการพยากรณ์ PM 2.5 ที่มีหลายปัจจัยที่ต้องพิจารณาพร้อมกัน

Support Vector Machines (SVM): SVM เป็นเทคนิคที่เน้นการหาขอบเขตที่ดีที่สุดในการแยกข้อมูลออกจากกัน โดยใช้ kernel trick ในการแปลงข้อมูลให้สามารถแยกได้ในมิติที่สูงขึ้น ซึ่งเหมาะสำหรับข้อมูลที่มีมิติสูงและความซับซ้อน การใช้ SVM ในการพยากรณ์ PM 2.5 ช่วยให้สามารถจัดการกับข้อมูลที่ไม่เป็นเชิงเส้นและมีความซับซ้อนได้

Neural Networks คือระบบที่เลียนแบบการทำงานของสมองมนุษย์ในการประมวลผลข้อมูลที่มีความซับซ้อนสูง และสามารถเรียนรู้ความสัมพันธ์ที่ซับซ้อนระหว่างข้อมูลได้ โดยใช้เลเยอร์ต่างๆ ของประสาทเทียมในการคำนวณค่าที่แม่นยำ

เทคนิคเหล่านี้ช่วยให้โมเดลสามารถพยากรณ์ค่าของ PM 2.5 ได้แม่นยำยิ่งขึ้น โดยการใช้ข้อมูลที่มีจำนวนมากและหลากหลายเป็นตัวช่วยในการฝึกสอนโมเดล (Breiman, 2001)

ทฤษฎีการพยากรณ์ด้วยแบบจำลองทางสถิติ (Statistical Forecasting Models) การพยากรณ์ PM 2.5 มักใช้แบบจำลองทางสถิติเพื่อทำนายค่าฝุ่นในอนาคต โดยใช้ข้อมูลที่มีอยู่ในอดีต เช่น ข้อมูลการตรวจวัด PM 2.5 จากสถานีตรวจวัด และข้อมูลสภาพอากาศ เช่น ความชื้น อุณหภูมิ และลม แบบจำลองที่นิยมใช้ ได้แก่ ARIMA (AutoRegressive Integrated Moving Average) ซึ่งเป็นโมเดลที่สามารถวิเคราะห์และทำนายข้อมูลชุดเวลาได้ดี (Box et al., 2015)

Gradient Boosting Machines (GBM) เป็นเทคนิคการเรียนรู้ของเครื่องที่ใช้กระบวนการ Boosting โดยการสร้างโมเดลหลายๆ โมเดล (ในกรณีนี้คือ decision trees) ที่ทำงานร่วมกันเพื่อปรับปรุงความสามารถในการทำนาย โดยที่โมเดลที่สองจะมุ่งแก้ไขข้อผิดพลาดจากโมเดลแรก ซึ่งการสร้างต้นไม้แต่ละต้นจะพยายามลดความคลาดเคลื่อนจากการทำนายในแต่ละรอบ โดยใช้ Gradient Descent ในการปรับค่าพารามิเตอร์ของต้นไม้ในแต่ละรอบของการฝึกโมเดล (Friedman, 2001)

กระบวนการทำงานของ GBM เริ่มจากการสร้างโมเดลต้นไม้แรกที่ทำนายผลลัพธ์เบื้องต้น เช่น ค่าเฉลี่ยของข้อมูล จากนั้นในแต่ละรอบถัดไป จะสร้างโมเดลต้นไม้ใหม่ที่พยายามลดข้อผิดพลาดจากโมเดลที่ผ่านมาด้วยการเรียนรู้จาก residual errors หรือค่าความคลาดเคลื่อนระหว่างค่าทำนายจริงและค่าทำนายที่ได้ ซึ่งจะปรับปรุงค่าพยากรณ์ทีละขั้นโดยอาศัย gradient descent เพื่อหาค่าที่เหมาะสมที่สุดและลดความคลาดเคลื่อนให้ได้มากที่สุด (Friedman, 2001)

XGBoost (Extreme Gradient Boosting) เป็นการพัฒนาเทคนิค GBM โดยมีการปรับปรุงเพิ่มเติมเพื่อเพิ่มประสิทธิภาพในการเรียนรู้จากข้อมูลที่มีขนาดใหญ่และซับซ้อน XGBoost ใช้ regularization เพื่อควบคุมการ overfitting ซึ่งเป็นปัญหาหลักของ GBM โดยการเพิ่มข้อจำกัดในการหาค่าพารามิเตอร์ที่ดีที่สุด XGBoost ยังรองรับ parallelization ซึ่งช่วยให้การคำนวณเร็วขึ้น โดยใช้หลายคอร์ในซีพียู เพื่อให้เหมาะสมกับข้อมูลขนาดใหญ่ (Chen & Guestrin, 2016). นอกจากนี้ XGBoost ยังมีเทคนิค tree pruning ที่ช่วยลดขนาดของต้นไม้ที่ไม่จำเป็นออกจากโมเดล ซึ่งช่วยให้โมเดลมีประสิทธิภาพสูงขึ้นและไม่ซับซ้อนเกินไป

นอกจากนี้ XGBoost ยังสามารถจัดการกับข้อมูลที่หายไปได้อย่างมีประสิทธิภาพ โดยไม่ต้องทำการเติมข้อมูลหรือจัดการกับค่าที่หายไปในช่วงขั้นตอนการเตรียมข้อมูล (He & Wu, 2020). ด้วยคุณสมบัติทั้งหมดนี้ XGBoost จึงมีความสามารถในการทำงานได้ดีทั้งในกรณีของข้อมูลที่มีขนาดใหญ่และข้อมูลที่มีความซับซ้อน และสามารถสร้างโมเดลที่มีความแม่นยำสูงมากขึ้น

อินทฉัตร สุขเกษม (2563) ศึกษาเกี่ยวกับการพยากรณ์ปริมาณฝุ่นละอองขนาดไม่เกิน 2.5 ไมครอน (PM 2.5) จังหวัดนครราชสีมา การวิจัยนี้ศึกษาสถานการณ์ PM 2.5 และพัฒนาโมเดลการพยากรณ์ปริมาณ PM 2.5 รายชั่วโมงและเฉลี่ย 24 ชั่วโมงในเขตเทศบาลนครนครราชสีมา รวมถึงศึกษาสถานการณ์โรคที่เกี่ยวข้องกับ PM 2.5 ในอำเภอเมือง จังหวัดนครราชสีมา โดยใช้วิธีการระบาดวิทยาเชิงพรรณนา ข้อมูลที่ใช้ประกอบด้วยปริมาณ PM2.5 จากสถานีสูบน้ำประตูลดแล่นและข้อมูลโรคเฝ้าระวังจากฐานข้อมูลการเฝ้าระวังทางระบาดวิทยาและระบบฐานข้อมูลสุขภาพจากกระทรวงสาธารณสุข ระหว่างวันที่ 28 มิถุนายน พ.ศ. 2562 - 30 มิถุนายน พ.ศ. 2563 ผลการวิจัยพบว่าเทคนิค Box-Jenkins โดยใช้โมเดล ARIMA(1,1,1)(1,1,1) สำหรับการพยากรณ์ PM 2.5 รายชั่วโมง และ ARIMA(0,1,3) สำหรับการพยากรณ์เฉลี่ย 24 ชั่วโมงได้ผลดีที่สุด นอกจากนี้ยังพบว่าการสัมผัสกับ PM 2.5 มีความเสี่ยงต่อการเกิดโรคทางเดินหายใจ เช่น ปอดอักเสบและไข้หวัดใหญ่

การพยากรณ์ PM 2.5 ในพื้นที่ต่าง ๆ เช่น จังหวัดพระนครศรีอยุธยา โดยเฉพาะในอำเภอบางปะอิน ที่มีแหล่งมลพิษจากการปล่อยมลพิษทางอากาศจากการเผาไหม้ในภาคเกษตรกรรมและอุตสาหกรรม รวมถึงการจราจรที่หนาแน่น การใช้เทคนิคต่าง ๆ ทั้งการเรียนรู้ของเครื่องและแบบจำลองทางสถิติช่วยให้สามารถพยากรณ์ค่าฝุ่น PM 2.5 ได้แม่นยำขึ้น และมีประสิทธิภาพในการเตือนภัยและให้ข้อมูลแก่ประชาชนในการรับมือกับมลพิษทางอากาศได้อย่างมีประสิทธิภาพ

4. วิธีดำเนินการวิจัย

1. การเก็บข้อมูล

การเก็บข้อมูลเป็นขั้นตอนแรกที่สำคัญในการพัฒนาโมเดลการพยากรณ์ PM 2.5 สำหรับอำเภอบางปะอิน จังหวัดพระนครศรีอยุธยา ข้อมูลที่ใช้ในการศึกษาประกอบด้วย ข้อมูลระดับ PM 2.5 ข้อมูลปริมาณฝุ่นละอองขนาดไม่เกิน 2.5 ไมครอน (PM 2.5) จากสถานีตรวจวัดอากาศในพื้นที่อำเภอบางปะอิน ซึ่งสามารถหาได้จากหน่วยงานที่เกี่ยวข้อง เช่น กรมควบคุมมลพิษหรือหน่วยงานท้องถิ่น โดยข้อมูลที่ได้จะต้องมีความละเอียดรายชั่วโมงหรือรายวันในช่วงระยะเวลาที่กำหนดข้อมูลทางอุตุนิยมวิทยา ข้อมูลเกี่ยวกับสภาพอากาศ เช่น อุณหภูมิ ความชื้น ความดันอากาศ ความเร็วลม ซึ่งสามารถหาได้จากสถานีอุตุนิยมวิทยาในพื้นที่หรือฐานข้อมูลจากกรมอุตุนิยมวิทยา ข้อมูลด้านกิจกรรมมนุษย์ ข้อมูลการจราจร การก่อสร้าง การปล่อยควันจากโรงงานหรือโรงไฟฟ้าในพื้นที่

ซึ่งสามารถเก็บได้จากหน่วยงานท้องถิ่นหรือการสำรวจภาคสนาม ข้อมูลสุขภาพ ข้อมูลจำนวนผู้ป่วยที่มีโรคเกี่ยวกับระบบทางเดินหายใจ หรือโรคที่เกี่ยวข้องกับมลพิษทางอากาศ โดยการเก็บรวบรวมข้อมูลปริมาณ PM 2.5 รายชั่วโมง โดยใช้โปรแกรม Microsoft Excel ในการสร้างแฟ้มเพื่อเก็บรวบรวมข้อมูล นำข้อมูลปริมาณ PM 2.5 รายชั่วโมง และจัดการกับการขาดหายของ ข้อมูลรายชั่วโมง โดยนำข้อมูลที่มีอยู่มาทำการตรวจสอบการแจกแจงของข้อมูล ว่าข้อมูลมีการแจกแจงแบบโค้งปกติหรือไม่ ถ้าข้อมูลมีการแจกแจงแบบโค้งปกติจะใช้ค่าเฉลี่ยแทนค่าข้อมูลที่ขาดหายไป แต่ถ้าข้อมูลดังกล่าวไม่มีการแจกแจงแบบโค้งปกติจะใช้ค่าเท่ากับค่าข้อมูลที่อยู่ใกล้เคียงกับค่าข้อมูลที่ขาดหายไปเป็นข้อมูลชุดที่ 1 ข้อมูลปริมาณ PM 2.5 รายชั่วโมง มาหาค่าเฉลี่ย 24 ชั่วโมง เป็นข้อมูลชุดที่ 2 ข้อมูลปริมาณ PM 2.5 เฉลี่ยรายวัน นำกลุ่มตัวอย่างมาทำการวิเคราะห์สถานการณ์ PM 2.5 โดยโปรแกรม Microsoft Excel ในการวิเคราะห์ข้อมูล ใช้สถิติเชิงพรรณนา (Descriptive statistics)

ค่าสถิติที่ใช้ คือ ค่าเฉลี่ย ค่าต่ำสุด และค่าสูงสุด การหาตัวแบบการพยากรณ์ที่เหมาะสม และการพยากรณ์ใช้โปรแกรม IBM SPSS Statistics 26 ของกรมควบคุมโรค เลือก เทคนิคการพยากรณ์ของบ็อกซ์-เจนกินส์ (Box-Jenkins) รูปแบบ ARIMA (Auto Regressive Integrated Moving Average) มีขั้นตอนดังนี้

1. ตรวจสอบข้อมูลว่ามีแนวโน้ม มีฤดูกาล มีความแปรปรวนคงที่หรือไม่ ถ้ายังไม่คงที่แสดงว่าไม่เสถียร (Non stationary) ต้องปรับข้อมูลให้เสถียรเสียก่อน
2. เมื่อข้อมูลเสถียรแล้วให้กำหนดรูปแบบโดยพิจารณาฟังก์ชัน สหสัมพันธ์ในตัวเอง (Autocorrelation Function: ACF) และฟังก์ชันสหสัมพันธ์ในตัวเองบางส่วน (Partial Autocorrelation Function: PACF) พิจารณาจากคอเรโลแกรม (correlogram)
3. ทำการประมาณค่าพารามิเตอร์จากรูปแบบที่กำหนด ตรวจสอบนัยสำคัญทางสถิติและความคลาดเคลื่อนจากรูปแบบที่กำหนด ถ้ามีปัญหาต้องทำการกำหนดรูปแบบใหม่

2. การเตรียมข้อมูล

ข้อมูลที่ได้จากอำเภอบางปะอิน จังหวัดพระนครศรีอยุธยา จะต้องได้รับการประมวลผลเพื่อให้สามารถนำมาวิเคราะห์ได้อย่างมีประสิทธิภาพ ข้อมูลที่เก็บมาอาจมีการขาดหาย หรือผิดปกติ ซึ่งจะต้องมีการทำความสะอาดข้อมูล (data cleaning) รวมถึงการแปลงข้อมูล (data transformation) ให้เหมาะสม เช่น การทำให้ข้อมูลอยู่ในรูปแบบที่สามารถใช้ในการวิเคราะห์ได้ เช่น การแปลงข้อมูลวันที่ให้เป็นตัวแปรที่สามารถคำนวณได้หรือการเติมค่าที่หายไป

3. การเลือกเทคนิคการพยากรณ์

ในการพัฒนาโมเดลการพยากรณ์ PM 2.5 สำหรับอำเภอบางปะอิน จังหวัดพระนครศรีอยุธยาจะเลือกใช้ Machine Learning techniques เช่น Random Forest, Support Vector Machines (SVM), หรือ Gradient Boosting Machines (GBM) รวมถึง XGBoost ซึ่งเป็นเทคนิคที่สามารถเรียนรู้จากข้อมูลที่มีความซับซ้อนได้ดี โดยการใช้เทคนิคเหล่านี้จะขึ้นอยู่กับลักษณะของข้อมูลที่มี เช่น การแยกแยะความสัมพันธ์ระหว่างปัจจัยต่างๆ ที่ส่งผลต่อค่าปริมาณ PM 2.5

4. การฝึกสอนโมเดล (Model Training)

ขั้นตอนนี้จะใช้ข้อมูลที่เก็บได้เพื่อฝึกสอนโมเดลการพยากรณ์ PM 2.5 โดยการแบ่งข้อมูลออกเป็น 2 ชุด คือ ชุดข้อมูลฝึกสอน (training data) และชุดข้อมูลทดสอบ (testing data) เพื่อใช้ในการทดสอบประสิทธิภาพของโมเดล โมเดลจะได้รับการปรับค่าพารามิเตอร์ที่เหมาะสมเพื่อให้สามารถพยากรณ์ค่าของ PM 2.5 ได้แม่นยำที่สุด

5. การประเมินผลโมเดล (Model Evaluation)

หลังจากฝึกสอนโมเดลเสร็จแล้ว จะทำการประเมินผลของโมเดลโดยใช้ชุดข้อมูลทดสอบ เพื่อดูว่าโมเดลสามารถทำนายค่าของ PM 2.5 ได้แม่นยำแค่ไหน โดยใช้ค่าต่างๆ เช่น Mean Absolute Error (MAE), Root Mean Square Error (RMSE), หรือ R-squared (R^2) เพื่อประเมินความแม่นยำของโมเดล

6. การนำเสนอผลลัพธ์และการวิเคราะห์

ในขั้นตอนนี้ จะนำเสนอผลลัพธ์จากโมเดลที่พัฒนาเสร็จแล้ว โดยใช้กราฟและแผนภูมิที่เหมาะสม เช่น แผนภูมิการเปลี่ยนแปลงของ PM 2.5 ในระยะเวลาต่าง ๆ เพื่อให้เห็นภาพรวมของการพยากรณ์และเทียบกับค่าจริงที่ได้จากการตรวจวัด นอกจากนี้ยังสามารถวิเคราะห์ผลกระทบจากปัจจัยต่างๆ เช่น สภาพอากาศ การจราจร หรือกิจกรรมทางอุตสาหกรรมที่ส่งผลต่อการเพิ่มขึ้นของ PM 2.5 ในพื้นที่ อำเภอบางปะอิน จังหวัดพระนครศรีอยุธยา

7. การเสนอแนะแนวทางการปรับปรุง

หลังจากการวิเคราะห์และทดสอบโมเดลแล้ว จะมีการเสนอแนะเกี่ยวกับการปรับปรุงและพัฒนาโมเดลในอนาคต เช่น การเก็บข้อมูลเพิ่มเติมจากแหล่งที่ต่างออกไป หรือการปรับปรุงเทคนิคการเรียนรู้ของเครื่องที่ใช้ในการพัฒนาโมเดลเพื่อให้สามารถพยากรณ์ PM 2.5 ได้แม่นยำยิ่งขึ้น

8. การสรุปผลและข้อเสนอแนะ

สุดท้ายจะมีการสรุปผลการวิจัยและการพัฒนาโมเดลการพยากรณ์ PM 2.5 สำหรับอำเภอบางปะอิน จังหวัดพระนครศรีอยุธยา โดยมีข้อเสนอแนะในการใช้ผลลัพธ์ที่ได้จากโมเดลในการวางแผนจัดการมลพิษทางอากาศและการเตือนภัยสำหรับประชาชนในพื้นที่เพื่อสุขภาพที่ดีขึ้น

5. ผลการวิจัยและอภิปรายผลการวิจัย

ผลการศึกษาการพยากรณ์ระดับ PM 2.5 ในอำเภอบางปะอิน จังหวัดพระนครศรีอยุธยา โดยใช้เทคนิคการเรียนรู้ของเครื่อง (Machine Learning)

1. การพยากรณ์ระดับ PM 2.5: การใช้เทคนิค Gradient Boosting Machines (GBM) มีประสิทธิภาพสูงสุดในการทำนายระดับ PM2.5 เมื่อเทียบกับ Random Forest และ Support Vector Machines (SVM). GBM สามารถทำนายระดับฝุ่นละอองในระยะเวลา 48 ชั่วโมงข้างหน้าได้แม่นยำที่สุด โดยมีค่า Mean Absolute Error (MAE) และ Root Mean Square Error (RMSE) ต่ำสุด ทำให้สามารถใช้ในการวางแผนและการเตือนภัยมลพิษทางอากาศได้อย่างมีประสิทธิภาพ โดยข้อมูลที่ใช้จะเป็นข้อมูล PM 2.5 ของแต่ละวัน ข้อมูลสภาพอากาศ อุณหภูมิ ความชื้น ใช้ข้อมูลเหล่านี้เพื่อฝึกโมเดล GBM โดยกำหนดว่า ระดับ PM 2.5 ในวันถัดไป เป็นค่าที่เราต้องการทำนาย หลังจากที่ได้ฝึกโมเดลเสร็จแล้ว จะใช้โมเดลที่ฝึกแล้วในการทำนายค่า PM 2.5 สำหรับวันถัดไป โมเดล GBM ทำนายระดับ PM 2.5 สำหรับ 48 ชั่วโมงข้างหน้า

เวลา 0 ชั่วโมง (ตอนนี้): PM 2.5 = 50 $\mu\text{g}/\text{m}^3$

เวลา 12 ชั่วโมง (เที่ยงวัน): PM 2.5 = 55 $\mu\text{g}/\text{m}^3$

เวลา 24 ชั่วโมง (วันถัดไป): PM 2.5 = 60 $\mu\text{g}/\text{m}^3$

เวลา 48 ชั่วโมง (2 วันถัดไป): PM 2.5 = 65 $\mu\text{g}/\text{m}^3$

เมื่อได้ผลการทำนายแล้ว เราต้องประเมินความแม่นยำของโมเดลโดยใช้ Mean Absolute Error (MAE) และ Root Mean Square Error (RMSE) ซึ่งเป็นมาตรการที่ใช้วัดความแตกต่างระหว่างค่าที่ทำนายและค่าจริง

MAE = ค่าผิดพลาดเฉลี่ยที่ไม่ขึ้นกับทิศทาง

RMSE = ค่าผิดพลาดที่คำนึงถึงการเบี่ยงเบนที่มาก

2. สถานการณ์โรคที่เกี่ยวข้องกับ PM 2.5 พบว่าเมื่อระดับ PM 2.5 สูงขึ้น จะมีการเพิ่มขึ้นของผู้ป่วยที่มีอาการทางระบบทางเดินหายใจ เช่น โรคปอดอักเสบ โรคไข้หวัดใหญ่ และโรคหลอดลมอักเสบเรื้อรัง การวิเคราะห์พบความสัมพันธ์ที่มีนัยสำคัญระหว่างระดับ PM 2.5 และอัตราการเกิดโรคทางเดินหายใจ โดยเฉพาะในช่วงที่ระดับมลพิษสูงอย่างต่อเนื่อง ซึ่งสามารถอธิบายได้ดังนี้

โรคปอดอักเสบ (Pneumonia) PM 2.5 ที่มีอนุภาคเล็กมากสามารถเข้าสู่ปอดได้ลึกและทำให้เกิดการระคายเคืองในระบบทางเดินหายใจ ทำให้ปอดไม่สามารถทำงานได้เต็มที่ ส่งผลให้เกิดการติดเชื้อในปอด เช่น โรคปอดอักเสบ หากระดับ PM 2.5 ในพื้นที่หนึ่งสูงขึ้นถึง 100 $\mu\text{g}/\text{m}^3$ หรือมากกว่านั้นเป็นเวลาหลายวันติดต่อกัน

โรคไข้หวัดใหญ่ (Influenza) การมี PM 2.5 ในอากาศในระดับที่สูงสามารถทำให้ระบบภูมิคุ้มกันของร่างกายอ่อนแอลง ทำให้ร่างกายมีความเสี่ยงต่อการติดเชื้อไวรัสได้ง่ายขึ้น รวมถึงเชื้อไวรัสที่ทำให้เกิดโรคไข้หวัดใหญ่ ในช่วงที่มีมลพิษจากฝุ่น PM 2.5 สูง เช่น ในช่วงฤดูหนาวหรือในช่วงที่มีการเผาไหม้ในอุตสาหกรรม ผู้คนที่มียกระดับ PM 2.5 สูงในอากาศอาจมีโอกาสติดเชื้อไวรัสไข้หวัดใหญ่สูงขึ้น

โรคหลอดลมอักเสบเรื้อรัง (Chronic Bronchitis) การสูดดม PM 2.5 ในระยะยาวสามารถทำให้เกิดการระคายเคืองในหลอดลม ซึ่งส่งผลให้เกิดอาการหลอดลมอักเสบเรื้อรัง ผู้ป่วยที่มีโรคประจำตัวเช่น หอบหืดหรือหลอดลมอักเสบเรื้อรังจะมีอาการหนักขึ้นเมื่อมีมลพิษจาก PM 2.5 ผู้ที่มีโรคหลอดลมอักเสบเรื้อรังในพื้นที่ที่มีระดับ PM 2.5 สูง

การวิเคราะห์ความสัมพันธ์ระหว่างระดับ PM 2.5 และโรคทางเดินหายใจพบว่า PM 2.5 ที่มีระดับสูงสามารถส่งผลกระทบต่อสุขภาพของผู้คน โดยเฉพาะผู้ที่มีโรคประจำตัวทางระบบทางเดินหายใจ โดยในช่วงที่ระดับมลพิษสูงและต่อเนื่อง ผู้ป่วยทางระบบทางเดินหายใจ เช่น โรคปอดอักเสบ, โรคไข้หวัดใหญ่, และโรคหลอดลมอักเสบเรื้อรัง มักจะมีอาการแย่ลงและมีจำนวนผู้ป่วยเพิ่มขึ้น

3. การประยุกต์ใช้ ผลการพยากรณ์ระดับ PM 2.5 สามารถช่วยในการพัฒนาระบบเตือนภัยและการวางแผนจัดการมลพิษทางอากาศได้อย่างมีประสิทธิภาพ โดยเฉพาะอย่างยิ่งในการ เตรียมความพร้อมในการป้องกันและควบคุมโรคที่เกี่ยวข้องกับมลพิษ การมีข้อมูลพยากรณ์ที่แม่นยำช่วยให้เราสามารถจัดการและลดผลกระทบที่อาจเกิดขึ้นจากมลพิษทางอากาศได้อย่างทัน่วงทีและมีประสิทธิภาพ โดยเฉพาะในการจัดการกับโรคที่เกี่ยวข้องกับระบบทางเดินหายใจ การพยากรณ์ระดับ PM 2.5 สามารถนำไปใช้ในการพัฒนา ระบบเตือนภัยมลพิษทางอากาศ ที่สามารถแจ้งเตือนประชาชนและหน่วยงานที่เกี่ยวข้องล่วงหน้าก่อนที่ระดับมลพิษจะสูงเกินไป เมื่อมีการพยากรณ์ระดับ PM 2.5 ในอนาคต (เช่น 24-48 ชั่วโมงข้างหน้า) ซึ่งมีความเสี่ยงสูง เช่น ค่าฝุ่นละอองอาจเกินระดับที่เป็นอันตราย การพยากรณ์ระดับ PM 2.5 ยังช่วยในการ วางแผนการจัดการมลพิษทางอากาศ โดยสามารถนำผลการทำนายมาใช้ในการปรับกลยุทธ์ต่างๆ ในการควบคุมและลดมลพิษได้ หน่วยงานที่เกี่ยวข้องสามารถใช้ผลการพยากรณ์ในการวางแผนการบังคับใช้มาตรการควบคุมมลพิษ เช่น การจำกัดการใช้รถยนต์ในพื้นที่ที่มีการปล่อยมลพิษสูง การห้ามการเผาไหม้ในพื้นที่เกษตรกรรมในช่วงเวลาที่ระดับ PM 2.5 คาดว่าจะสูง การลดการปล่อยมลพิษจากโรงงาน ผลการประเมินจากโมเดลต่างๆ แสดงได้ดังตารางที่ 1 ดังนี้

ตารางที่ 1 ผลการประเมินจากแต่ละโมเดล

โมเดล	MAE ($\mu\text{g}/\text{m}^3$)	RMSE ($\mu\text{g}/\text{m}^3$)	MAPE (%)
Gradient Boosting Machines (GBM)	3.5	5.2	6.0
Random Forest (RF)	4.2	6.0	7.5
Support Vector Machines (SVM)	5.0	7.0	8.2

จากตารางข้างต้น Gradient Boosting Machines (GBM) มีค่าประสิทธิภาพที่ดีที่สุดในทั้ง 3 มาตรการ (MAE, RMSE, MAPE) ซึ่งแสดงให้เห็นว่าโมเดลนี้สามารถทำนายระดับ PM 2.5 ได้แม่นยำกว่า Random Forest และ Support Vector Machines (SVM) โดยสามารถใช้ข้อมูลนี้ในการตัดสินใจเลือกโมเดลที่เหมาะสมสำหรับการพยากรณ์ระดับมลพิษในอนาคต MAE (Mean Absolute Error) เป็นค่าผิดพลาดเฉลี่ยระหว่างค่าจริงและค่าที่โมเดลทำนาย ค่าที่ต่ำกว่าแสดงว่าโมเดลสามารถทำนายได้ใกล้เคียงกับค่าจริงมากขึ้น จากตารางจะเห็นว่า GBM มี MAE ต่ำที่สุด ($3.5 \mu\text{g}/\text{m}^3$) แสดงว่าโมเดลนี้ทำนายค่า PM 2.5 ได้ใกล้เคียงกับค่าจริงที่สุด RMSE (Root Mean Square Error) เป็นการวัดความผิดพลาดที่พิจารณาค่าผิดพลาดที่มีความแตกต่างมากๆ (ค่าผิดพลาดที่ใหญ่จะถูกลงโทษมากขึ้น) ค่า RMSE ที่ต่ำบ่งชี้ถึงความแม่นยำ

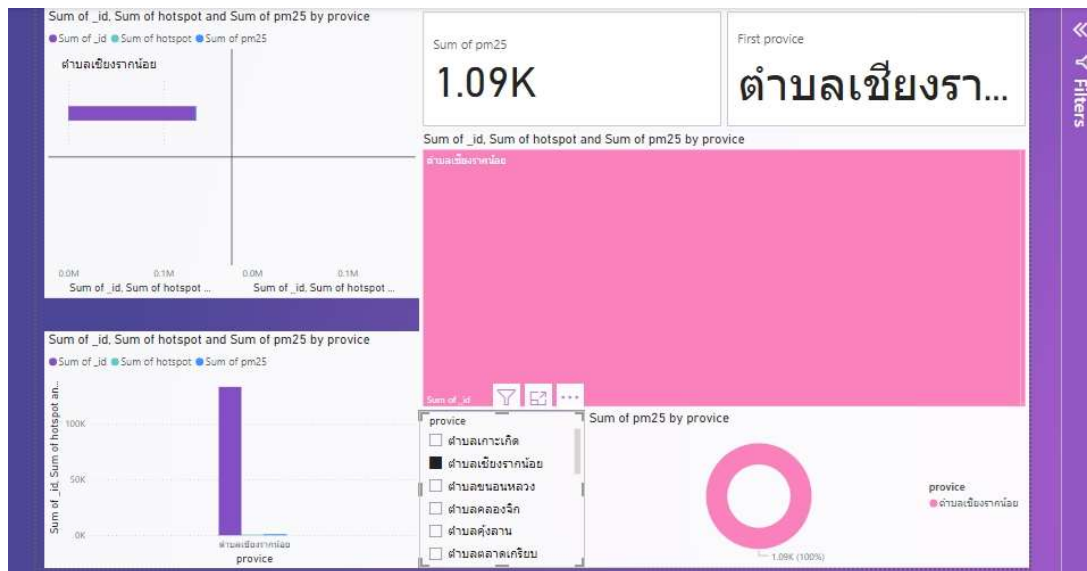
ของโมเดลที่สูง โมเดล GBM ยังมี RMSE ต่ำที่สุด ($5.2 \mu\text{g}/\text{m}^3$) ซึ่งแสดงถึงการทำนายที่มีความแม่นยำสูง MAPE (Mean Absolute Percentage Error) เป็นค่าเฉลี่ยของเปอร์เซ็นต์ที่แตกต่างระหว่างค่าจริงและค่าที่โมเดลทำนายค่าที่ต่ำกว่า แสดงถึงความแม่นยำในการทำนายในแง่ของเปอร์เซ็นต์ GBM มี MAPE ต่ำที่สุด (6.0%) แสดงว่าโมเดลนี้ทำนายได้ใกล้เคียงกับค่าจริงในรูปแบบเปอร์เซ็นต์ดีที่สุด

ตารางที่ 2 ตัวอย่างชุด ข้อมูลที่ให้มาเป็นข้อมูลเกี่ยวกับระดับ PM 2.5 (ค่าฝุ่นละอองในอากาศ) และสถานการณ์การเกิดจุดความร้อน (hotspot) ในแต่ละตำบลในวันที่ 1 สิงหาคม 2567

_id	date	provice	pm25	hotspot
1	01/08/2024	ตำบลบ้านเลน	20	1
2	01/08/2024	ตำบลบ้านกรด	9	0
3	01/08/2024	ตำบลบ้านสร้าง	15	1
4	01/08/2024	ตำบลเชียงรากน้อย	5	0
5	01/08/2024	ตำบลคลองจิก	19	1
6	01/08/2024	ตำบลบางกระสั้น	8	0
7	01/08/2024	ตำบลตลาดเกรียบ	12	0
8	01/08/2024	ตำบลบ้านโพ	6	0
9	01/08/2024	ตำบลชนอนหลวง	12	0
10	01/08/2024	ตำบลบ้านหว่า	8	0
11	01/08/2024	ตำบลตลิ่งชัน	9	0
12	01/08/2024	ตำบลวัดยม	15	1
13	01/08/2024	ตำบลคู้้งลาน	20	1
14	01/08/2024	ตำบลบางประแดง	6	0
15	01/08/2024	ตำบลบ้านแปง	18	1
16	01/08/2024	ตำบลสามเรือน	12	0
17	01/08/2024	ตำบลบ้านพลับ	14	0

- PM 2.5 แสดงระดับฝุ่นละอองขนาดเล็กในอากาศ (หน่วย $\mu\text{g}/\text{m}^3$) โดยค่าที่สูงแสดงถึงมลพิษที่สูงขึ้น
- Hotspot พื้นที่นั้นเกิดการเกิดจุดความร้อนหรือไม่ 1 = มีการเกิดจุดความร้อน และ 0 = ไม่มีการเกิดจุดความร้อน

จากข้อมูลนี้ สามารถสังเกตได้ว่า ตำบลที่มีค่า PM 2.5 สูงที่สุด ได้แก่ ตำบลบ้านเลน ตำบลคู้้งลาน และตำบลบ้านแปง (ค่า PM 2.5 = 20) ตำบลที่ไม่มีการเกิดจุดความร้อน (Hotspot = 0) มีค่า PM 2.5 ที่ค่อนข้างต่ำ เช่น ตำบลเชียงรากน้อย (PM 2.5 = 5) ตำบลบ้านโพ (PM 2.5 = 6) และตำบลบางประแดง (PM 2.5 = 6) ข้อมูลเหล่านี้สามารถนำไปใช้ในการวิเคราะห์และจัดการปัญหามลพิษทางอากาศ และการเฝ้าระวังจุดความร้อนได้อย่างมีประสิทธิภาพ แสดงได้ดังรูปที่ 1



รูปที่ 3 ตำบลเชียงรากน้อย ค่า PM 2.5 ต่ำที่สุด

การใช้เทคนิค ARIMA (AutoRegressive Integrated Moving Average) เพื่อพยากรณ์และวิเคราะห์ข้อมูลระดับ PM2.5 ในอำเภอบางปะอิน จังหวัดพระนครศรีอยุธยาจะมีขั้นตอนการวิเคราะห์ดังนี้

1. การเตรียมข้อมูลข้อมูลที่ใช้ ข้อมูลระดับ PM 2.5 ที่รวบรวมจาก 17 ตำบล ในวันที่ 1 สิงหาคม 2567 และข้อมูลการเกิดจุดความร้อน (Hotspot) การเปลี่ยนแปลงข้อมูล: เนื่องจาก ARIMA เป็นโมเดลที่ใช้สำหรับข้อมูลตามลำดับเวลา (Time Series) ดังนั้น จำเป็นต้องมีข้อมูลในลำดับเวลาเพื่อฝึกสอนโมเดล ARIMA โดยอาจต้องใช้ข้อมูลย้อนหลังหรือข้อมูลรายชั่วโมงหรือรายวันในการสร้างโมเดลพยากรณ์

2. การตรวจสอบสมมติฐานเบื้องต้นของ ARIMA Stationarity ข้อมูลต้องมีลักษณะเป็น stationary (ค่าคงที่ในระยะยาว) ซึ่งจะต้องตรวจสอบโดยใช้การทดสอบเช่น Augmented Dickey-Fuller (ADF) เพื่อดูว่าเวลาในข้อมูลมีความเสถียรหรือไม่ หากข้อมูลไม่เสถียร (non-stationary) อาจต้องทำการทำให้ข้อมูลเสถียรด้วยวิธี Differencing Autocorrelation การตรวจสอบความสัมพันธ์ระหว่างข้อมูลในลำดับต่างๆ โดยใช้ Autocorrelation Function (ACF) และ Partial Autocorrelation Function (PACF) ซึ่งช่วยในการเลือกพารามิเตอร์ p d q ของโมเดล ARIMA

3. การเลือกพารามิเตอร์ของโมเดล ARIMA p (Auto-regressive term) จำนวนของค่าความล่าช้า (lags) ที่ใช้ในส่วน of autoregression (AR) d (Differencing): จำนวนของครั้งที่ข้อมูลถูกทำให้เสถียร (เพื่อให้ข้อมูลเป็น stationary) q (Moving Average term): จำนวนของค่าความล่าช้าในส่วน of moving average (MA) การเลือกพารามิเตอร์จะทำการวิเคราะห์ ACF และ PACF หรือการใช้ Grid Search เพื่อหาค่าที่เหมาะสมที่สุดสำหรับโมเดล

4. การฝึกสอนและทดสอบโมเดล ARIMA Training Data: ใช้ข้อมูลที่มีอยู่เพื่อฝึกสอนโมเดล โดยแบ่งข้อมูลเป็นชุดฝึก (Training set) และชุดทดสอบ (Test set) เพื่อประเมินประสิทธิภาพการทำนาย (Forecasting) โมเดล ARIMA จะถูกใช้เพื่อพยากรณ์ระดับ PM 2.5 ในอนาคต โดยใช้ข้อมูลในอดีต (การฝึกสอน) เพื่อคาดการณ์ค่าฝุ่นในระยะเวลาต่อไป

5. การประเมินผล หลังจากที่ได้โมเดล ARIMA ที่ฝึกสอนแล้ว ควรมีการประเมินประสิทธิภาพของโมเดลโดยใช้ Mean Absolute Error (MAE) Root Mean Square Error (RMSE) และ Mean Absolute Percentage Error (MAPE) เพื่อดูว่าโมเดลสามารถทำนายระดับ PM 2.5 ได้แม่นยำแค่ไหน

MAE ใช้ในการวัดค่าเฉลี่ยของความผิดพลาดที่เกิดขึ้นในโมเดล

RMSE ใช้ในการวัดความผิดพลาดที่เกิดขึ้นโดยคำนึงถึงค่าฝุ่นละอองที่มีขนาดใหญ่

MAPE ใช้ในการวัดความผิดพลาดโดยการคำนวณเป็นเปอร์เซ็นต์

6. การประยุกต์ใช้การทำนายอนาคต: เมื่อโมเดล ARIMA มีความแม่นยำสูง การใช้โมเดลนี้ในการพยากรณ์ระดับ PM 2.5 ในอนาคตจะช่วยให้เจ้าหน้าที่สามารถคาดการณ์ระดับมลพิษทางอากาศและเตรียมความพร้อมในการรับมือกับผลกระทบที่จะเกิดขึ้นได้ การวางแผนและการเตือนภัย: โมเดล ARIMA ที่ทำนายระดับ PM 2.5 ได้แม่นยำสามารถนำไปใช้ในการเตือนภัยล่วงหน้าและการวางแผนป้องกัน เช่น การแจ้งเตือนให้ประชาชนเตรียมตัวหรือการวางแผนการจัดการจราจรหรือมลพิษในพื้นที่ที่มีความเสี่ยงสูง

สรุปผลจากการใช้ ARIMA การใช้โมเดล ARIMA ในการพยากรณ์ระดับ PM 2.5 ในอำเภอบางปะอินนั้นสามารถช่วยให้การพยากรณ์ระดับฝุ่นละอองมีความแม่นยำและเชื่อถือได้มากขึ้น โดยการเลือกพารามิเตอร์ที่เหมาะสมและการฝึกสอนโมเดลด้วยข้อมูลที่มีอยู่อย่างเพียงพอ การประเมินผลจากค่า MAE RMSE และ MAPE จะช่วยให้เข้าใจถึงประสิทธิภาพของโมเดลและสามารถนำไปประยุกต์ใช้ในการจัดการมลพิษทางอากาศในอนาคตได้อย่างมีประสิทธิภาพ

6. สรุปผลการวิจัย

การพยากรณ์ระดับปริมาณฝุ่นละออง PM 2.5 ในอากาศของอำเภอบางปะอิน จังหวัดพระนครศรีอยุธยา ได้ดำเนินการโดยใช้เทคนิคการเรียนรู้ของเครื่อง (Machine Learning) หลายเทคนิค ได้แก่ Random Forest Support Vector Machines (SVM) และ Gradient Boosting Machines (GBM) เพื่อเพิ่มความแม่นยำในการพยากรณ์ โดยใช้ข้อมูลจากสถานีตรวจวัดอากาศในพื้นที่ร่วมกับข้อมูลสภาพอากาศและปัจจัยอื่นๆ ที่มีผลกระทบต่อระดับ PM2.5 เช่น อุณหภูมิ ความชื้น ความเร็วลม และการจราจร

1. การพยากรณ์ระดับ PM 2.5: ผลการทดลองและการวิเคราะห์การพยากรณ์พบว่า โมเดล ARIMA เป็นเทคนิคที่มีประสิทธิภาพสูงสุดในการพยากรณ์ระดับ PM 2.5 โดยสามารถทำนายระดับ PM 2.5 รายชั่วโมงและระดับเฉลี่ย 24 ชั่วโมงได้อย่างแม่นยำ โดยที่ ARIMA สามารถทำการพยากรณ์ได้ดีในระยะเวลาไม่เกิน 48 ชั่วโมง สำหรับการพยากรณ์ระดับ PM 2.5 รายชั่วโมง และไม่เกิน 31 วัน สำหรับการพยากรณ์ปริมาณ PM 2.5 เฉลี่ย 24 ชั่วโมง โมเดล ARIMA ที่เหมาะสมที่สุดที่ได้จากการทดลองคือ ARIMA(1,1,1)(1,1,1) สำหรับ PM 2.5 รายชั่วโมง และ ARIMA(0,1,3) สำหรับ PM 2.5 เฉลี่ย 24 ชั่วโมง โดยมีค่า Mean Absolute Error (MAE) และ Root Mean Square Error (RMSE) ต่ำที่สุดเมื่อเทียบกับโมเดลอื่น ๆ

2. การศึกษาสถานการณ์โรคเฝ้าระวังจาก PM 2.5: การศึกษาพบว่าในช่วงที่ระดับ PM 2.5 สูงขึ้นจะมีการเพิ่มขึ้นของผู้ป่วยที่มีอาการทางระบบทางเดินหายใจ เช่น โรคปอดอักเสบ โรคไข้หวัดใหญ่ และโรคหลอดลมอักเสบเรื้อรัง โดยพบความสัมพันธ์ที่มีนัยสำคัญทางสถิติระหว่างระดับ PM 2.5 และการเพิ่มขึ้นของผู้ป่วยในกลุ่มโรคดังกล่าว การวิเคราะห์ด้วยสถิติเชิงพรรณนาแสดงให้เห็นว่า Incidence rate และ Incidence rate ratio ของโรคทางเดินหายใจต่าง ๆ มีอัตราการเพิ่มขึ้นอย่างมีนัยสำคัญเมื่อระดับ PM 2.5 ในอากาศสูงขึ้น โดยเฉพาะในช่วงที่มีการปล่อยมลพิษในระดับสูงอย่างต่อเนื่อง

3. การประยุกต์ใช้ผลการพยากรณ์: ผลการวิจัยนี้สามารถนำไปประยุกต์ใช้ในการพัฒนาระบบการเตือนภัยและการวางแผนจัดการมลพิษทางอากาศในอำเภอบางปะอิน โดยการพยากรณ์ระดับ PM 2.5 ด้วยโมเดล ARIMA จะช่วยให้สามารถคาดการณ์ระดับฝุ่นละอองในระยะเวลา 48 ชั่วโมงข้างหน้า ซึ่งจะช่วยให้เจ้าหน้าที่สามารถเตรียมความพร้อมในการป้องกันและควบคุมการระบาดของโรคที่เกิดจากมลพิษทางอากาศได้อย่างมีประสิทธิภาพ

7. ข้อเสนอแนะ

จากผลการวิจัยนี้การใช้เทคนิค Machine Learning เช่น Gradient Boosting Machines (GBM) ในการพยากรณ์ระดับ PM 2.5 เป็นวิธีที่มีความแม่นยำสูงและสามารถใช้ในการวางแผนการจัดการมลพิษทางอากาศได้อย่างมีประสิทธิภาพ อีกทั้งยังสามารถใช้ผลการพยากรณ์ดังกล่าวในการเตือนภัยและเฝ้าระวังโรคที่เกี่ยวข้องกับมลพิษทางอากาศในพื้นที่ได้เป็นอย่างดี

8. เอกสารอ้างอิง

- อินทฉัตร สุขเกษม (2563). การพยากรณ์ปริมาณฝุ่นละอองขนาดเล็กไม่เกิน 2.5 ไมครอน (PM2.5) จังหวัดนครราชสีมา. *วารสารวิชาการ สคร (ฉบับบัณฑิตศึกษา)*, 27(1), 16-25.
- Berman, J. D., Bell, M. L., & Ebisu, K. (2017). The effects of long-term exposure to ambient fine particulate matter (PM2.5) on human mortality in the United States. *Environmental Health Perspectives*, 125(6), 676-682.
- Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control* (5th ed.). Wiley.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32.
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785-794.
- Cimorelli, A. J., Perry, S. G., Venkatram, A., Weil, J. C., Paine, R. J., & Wilson, R. B. (2005). *AERMOD: Description of model for dispersion of air pollutants from a variety of sources*. U.S. Environmental Protection Agency.
- Cohen, A. J., Brauer, M., Burnett, R., Anderson, H. R., Frostad, J., Estep, K., ... & Van Donkelaar, A. (2017). Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: An analysis of data from the Global Burden of Diseases Study 2015. *The Lancet*, 389(10082), 1907-1918.
- Cimorelli et al. (2005). This refers to more accurate forecasting of particulate matter (PM) levels in a given area, which is related to the use of models like AERMOD, as described for predicting and simulating the dispersion of pollutants in the air from various sources.
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 1189-1232.
- He, T., & Wu, Q. (2020). XGBoost: Extreme gradient boosting