บทความวิจัย (Research Article)

ประสิทธิภาพของเทคนิคการเรียนรู้ของเครื่อง สำหรับการพยากรณ์ฝุ่นละอองขนาดเล็กในอากาศ

The Effectiveness of Machine Learning Forecasting of $PM_{2.5}$

กฤติกา ทิพย์คำมี[1] อนุพงศ์ สุขประเสริฐ[2*] สุพัตรา กอผจญ[1] และณัฐกานต์ ชุติมารังสรรค์[2]
Kittika Thipkummee[1], Anupong Sukprasert[2*], Supattra Kopajone[1] and Nattakarn Shutimarrungson[2]

วันที่รับบทความ (Received) วันที่ได้รับบทความฉบับแก้ไข (Revised) วันที่ตอบรับบทความ (Accepted)
4 มีนาคม 2566 23 เมษายน 2566 24 พฤษภาคม 2566

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของเทคนิคการเรียนรู้ของเครื่องสำหรับการพยากรณ์ฝุ่นละอองขนาดเล็กในอากาศ ($PM_{2.5}$) โดยตัวแปรที่ใช้ในการสร้างตัวแบบฝุ่น $PM_{2.5}$ จำนวน 16 ตัวแปร จากสถานีตรวจวัดคุณภาพอากาศและอุตุนิยมวิทยา ของกรมควบคุมมลพิษ จังหวัดอุดรธานี ที่ถูกจัดเก็บในรูปแบบรายชั่วโมง ตั้งแต่เดือนมิถุนายน ถึงเดือนธันวาคม ปี พ.ศ. 2564 รวมทั้งสิ้น 4,608 แถว โดยนำมาวิเคราะห์ตามกระบวนการวิเคราะห์ข้อมูล เพื่อสร้างตัวแบบพยากรณ์ฝุ่นละอองขนาดเล็กในอากาศ ด้วยเทคนิคการประมาณค่าข้อมูล (Regression Model) โดยในงานวิจัยในครั้งนี้ได้เลือกใช้เทคนิคการเรียนรู้ของเครื่องทั้งหมด 4 เทคนิค ได้แก่ เทคนิคการถดถอยเชิงเส้น (Linear Regression) เทคนิคโครงข่ายประสาทเทียม (Neural Network) เทคนิคซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machines) และเทคนิคการเรียนรู้เชิงลึก (Deep Learning) จากผลการวิจัยพบว่าเทคนิคที่มีความเหมาะสมที่สุดสำหรับการสร้างตัวแบบการพยากรณ์ฝุ่นละอองขนาดเล็กในอากาศ คือ เทคนิคโครงข่ายประสาทเทียม (Neural Network) โดยให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) น้อยที่สุดเท่ากับ 8.673 ค่ารากที่สองของความคลาดเคลื่อนเฉลี่ย

[1,2] คณะการบัญชีและการจัดการ มหาวิทยาลัยมหาสารคาม
[1,2] Mahasarakham Business School, Mahasarakham University
* Corresponding author; email: anupong.s@acc.msu.ac.th

กำลังสอง (RMSE) เท่ากับ 2.945 ค่าความคลาดเคลื่อนสัมบูรณ์ (AE) เท่ากับ 2.126 ค่าความคลาดเคลื่อนกำลังสอง (SE) เท่ากับ 8.709 และค่าสัมประสิทธิ์การตัดสินใจพหุคูณ (R^2) เท่ากับ 0.948 ผลลัพธ์จากการวิเคราะห์ข้อมูลนี้สามารถนำไปสร้างตัวแบบที่เหมาะสมไปใช้สำหรับการพยากรณ์ฝุ่นละอองขนาดเล็กในอากาศ เพื่อให้หน่วยงานที่เกี่ยวข้องสามารถเข้าถึงข้อมูลและนำไปใช้ในการกำหนดมาตรการการเฝ้าระวังการเกิดฝุ่นละอองขนาดเล็ก อีกทั้งสามารถพัฒนาต่อยอดไปเป็นระบบสารสนเทศสำหรับการการพยากรณ์ฝุ่นละอองขนาดเล็กที่จะเกิดขึ้นในอนาคต

คำสำคัญ : ฝุ่นละอองขนาดเล็ก, เทคนิคการถดถอยเชิงเส้น, เทคนิคโครงข่ายประสาทเทียม, เทคนิคซัพพอร์ตเวกเตอร์แมชชีน, เทคนิคการเรียนรู้เชิงลึก

Abstract

The purpose of this research was to compare the efficiency of machine learning techniques to predict small dust particles in the air ($PM_{2.5}$). The 16 variables were used to construct PM2.5 dust models from air quality and meteorological measurement stations of the Pollution Control Department that was collected an hourly format during June - December 2021, a total of 4,608 rows. The data analysis process is to build a trend prediction model of small dust particles in the air with data estimation techniques (Regression Model). Four different machine learning techniques were used as linear regression, neural networks, support vector machines, and deep learning. The results show that the most suitable technique for modeling dust particles in the air prediction is the neural network technique with the least mean squared error (MSE) of 8.673, the mean squared error (RMSE) of 2.945, the absolute error (AE) of 2.126, the doubled squared error (SE) of 8.709, and the multiple decision coefficient (R^2) of 0.948. The results can be used to build a suitable model for trend prediction of small dust particles in the air and set a standard scale for fine particulate matter in the air. Including, the relevant organizations can access the information and used to determine measures to monitor the occurrence of small dust particles. The results can also be further developed an information system for predicting small dust particles to appear in the air more accuracy.

Keywords: $PM_{2.5}$, Linear Regression, Neural Network, Support Vector Machines, Deep Learning

บทนำ

ในปัจจุบันปัญหามลพิษทางอากาศถือเป็นปัญหาหลักที่มักเกิดขึ้นในเมืองใหญ่โดยมีผลมาจากปัจจัยหลาย ๆ อย่างทั้งที่เกิดขึ้นเองโดยธรรมชาติและเกิดจากมนุษย์สร้างขึ้น ซึ่งปัญหาหลักมักเกิดจากการที่มนุษย์ก่อขึ้นไม่ว่าจะเป็นการขยายตัวของเศรษฐกิจที่มาจากการก่อสร้าง การคมนาคมขนส่ง การจราจรและการประกอบกิจกรรมหลาย ๆ ด้าน จึงเป็นสาเหตุของการเกิดมลพิษทางอากาศ

ฝุ่นละอองขนาดไม่เกิน 2.5 ไมครอน หรือ $PM_{2.5}$ เป็นสารเจือปนในอากาศอย่างหนึ่งที่ก่อให้เกิดมลพิษทางอากาศ สามารถแพร่กระจายเข้าสู่ทางเดินหายใจ กระแสเลือดและแทรกซึมสู่การทำงานของอวัยวะต่าง ๆ ในร่างกายได้ซึ่งเพิ่มความเสี่ยงเป็นโรคระบบทางเดินหายใจแบบเรื้อรังและมะเร็งได้ ตัวฝุ่นเป็นพาหะนำสารอื่นเข้ามาด้วย เช่น แคดเมียม ปรอท โลหะหนักและสารก่อมะเร็งจำนวนมาก [1] หากในอากาศมีฝุ่นละอองขนาดเล็กปริมาณมากจะมีผลเสียต่อระบบทางเดินหายใจและอาจทำให้เกิดโรคที่เกี่ยวข้องต่อระบบทางเดินหายใจร้ายแรง ถึงขั้นเสียชีวิตได้โดยผลกระทบต่อสุขภาพจากฝุ่นละอองขนาดเล็กสามารถก่อให้เกิดผลกระทบต่อสุขภาพได้หลายระบบ เช่น ระบบทางเดินหายใจ (การไอและอาการของระบบทางเดินหายใจส่วนล่าง) ระบบหัวใจ และหลอดเลือด (กล้ามเนื้อหัวใจขาดเลือด หัวใจเต้นไม่สม่ำเสมอ หัวใจวาย) ระบบตา ระบบผิวหนัง ฝุ่นขนาดเล็กยังเพิ่มความเสี่ยงของอัตราการตายจากภาวะเส้นเลือดอุดตันในสมองและทำให้น้ำหนักของทารกในครรภ์ลดลงอีกด้วย ทำให้อัตราป่วยและอัตราตายด้วยโรคระบบทางเดินหายใจและระบบหัวใจและหลอดเลือดเพิ่มขึ้น และอัตราดังกล่าวจะเพิ่มขึ้นตามปริมาณความเข้มข้นของฝุ่น ในอากาศ [2]

ประเทศไทยกำลังเผชิญกับปัญหาฝุ่น $PM_{2.5}$ ส่งผลต่อสุขภาพของประชาชนและทำให้เกิดการเจ็บป่วยจากโรคหลายชนิด มีงานวิจัยจำนวนมากที่ได้ศึกษาผลของ $PM_{2.5}$ ที่มีความเสี่ยงต่อสุขภาพของประชาชนส่งผลต่อการทำงานของระบบต่าง ๆ ในร่างกาย เช่น การทำงานของตับ ระบบหัวใจหลอดเลือด ระบบทางเดินหายใจ การเกิดเบาหวาน ภาวะไขมันพอกตับ โรคข้ออักเสบรูมาตอยด์ มะเร็งปอด และมะเร็งชนิดอื่น ๆ เป็นต้น [3] จากผลกระทบของ $PM_{2.5}$ ดังกล่าวจึงมีความจำเป็นอย่างยิ่งที่จะมีแนวทางในการพยากรณ์เพื่อให้รับรู้ปริมาณฝุ่นที่จะเกิดขึ้นในอนาคต เพื่อแจ้งให้ประชาชนในพื้นที่ได้เตรียมรับมือกับฝุ่นละอองขนาดเล็กที่จะเกิดขึ้น มีงานวิจัยหลายงานวิจัยที่ได้พัฒนาตัวแบบสำหรับการพยากรณ์ปริมาณฝุ่น $PM_{2.5}$ โดยใช้เทคนิคสำหรับการพยากรณ์ตัวแบบ เช่น งานวิจัยของอินท์ฉัตร สุขเกษม [4] ได้ทำพยากรณ์ปริมาณ $PM_{2.5}$ รายชั่วโมงโดยการหาตัวแบบการพยากรณ์ที่เหมาะสมโดยทำการเปรียบเทียบระหว่างเทคนิคการพยากรณ์แบบทำให้เรียบและเทคนิคการพยากรณ์ของบอกซ์-เจนกินส์ งานวิจัยของธรณินทร์ สัจวิริยทรัพย์ [5] ได้ทำการศึกษาตัวแบบพยากรณ์ปริมาณฝุ่น $PM_{2.5}$ โดยใช้เทคนิคผสมอารีมา-ซัพพอร์ตเวกเตอร์แมชชีน งานวิจัยของพรนภา แสงศรี, พรพิมล ชัยวุฒิศักดิ์ [6] ได้ทำการเปรียบเทียบการพยากรณ์ฝุ่นละอองขนาดเล็กด้วยเทคนิค

โครงข่ายประสาทเทียมและวิธีผลมของโครงข่ายประสาทเทียมและห่วงโซ่มาร์คอฟ งานวิจัยชญานนท์ เทพแสงพราว, ธงทิศ ฉายากุล, ศิริมา ปัญญาเมธีกุล [7] ได้สร้างสมการทางคณิตศาสตร์เพื่อใช้ในการหาปริมาณความเข้มข้นของ $PM_{2.5}$ ด้วยวิธีการวิเคราะห์การถดถอยเชิงพหุคูณ (Multiple Linear Regression; MLR) งานวิจัยของปรัชญา สิงหวรวงศ์ [8] ได้ทำการพยากรณ์ความเข้มข้นฝุ่นละออง PM2.5 ด้วยเทคนิคการเรียนรู้เชิงลึก (Deep Learning) ซึ่งจากการศึกษาพบว่างานวิจัยดังกล่าวยังไม่ได้มีการเปรียบเทียบเทคนิคการพยากรณ์หลายเทคนิคร่วมกันจากชุดข้อมูลชุดเดียวกัน เพื่อให้การพัฒนาตัวแบบสำหรับการพยากรณ์ปริมาณฝุ่น $PM_{2.5}$ ที่มีประสิทธิภาพมากยิ่งขึ้น

ผู้วิจัยได้สนใจศึกษาการสร้างตัวแบบและเปรียบเทียบประสิทธิภาพของตัวแบบสำหรับการพยากรณ์ฝุ่นละอองขนาดเล็กในอากาศ โดยใช้เทคนิคการประมาณค่า 4 เทคนิคมาทำการสร้างตัวแบบการพยากรณ์ที่ได้มาจากการศึกษางานวิจัยที่เกี่ยวข้องกับงานวิจัยในครั้งนี้ ซึ่งได้แก่ เทคนิคการถดถอยเชิงเส้น (Linear Regression) เทคนิคโครงข่ายประสาทเทียม (Neural Network) เทคนิคการซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machines) และเทคนิคการเรียนรู้เชิงลึก (Deep Learning) เพื่อหาตัวแบบที่เหมาะสมสำหรับการนำไปใช้ในการสร้างตัวแบบการพยากรณ์ฝุ่นละอองขนาดเล็กในอากาศที่จะเกิดขึ้นในอนาคต อีกทั้งยังสามารถนำผลการวิจัยครั้งนี้ไปพัฒนาต่อยอดในการสร้างแบบจำลองสำหรับการพยากรณ์เกี่ยวกับฝุ่นละอองขนาดเล็กในอากาศต่อไป

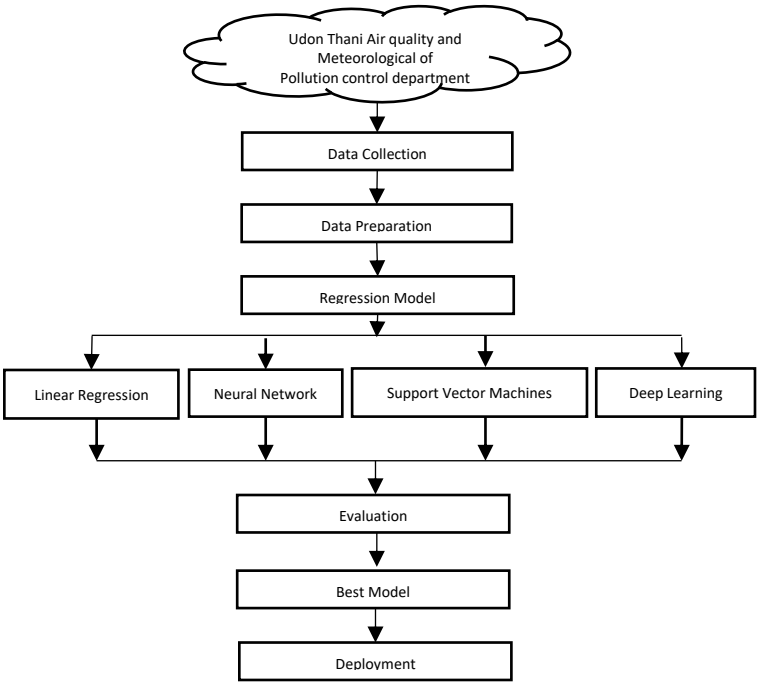
วัตถุประสงค์ของการวิจัย

1. เพื่อสร้างตัวแบบสำหรับการพยากรณ์ฝุ่นละอองขนาดเล็กในอากาศ
2. เพื่อเปรียบเทียบประสิทธิภาพของตัวแบบสำหรับการพยากรณ์ฝุ่นละอองขนาดเล็กในอากาศ

วิธีดำเนินการวิจัย

ผู้วิจัยได้ศึกษาเกี่ยวกับเทคนิคการเรียนรู้ของเครื่องสำหรับนำมาสร้างตัวแบบการพยากรณ์ฝุ่นละอองขนาดเล็กในอากาศ โดยผ่านกระบวนการการวิเคราะห์ข้อมูล (Data Analysis Process) โดยการแบ่งข้อมูลออกเป็น 2 ส่วนเพื่อใช้สำหรับการสร้างตัวแบบ (Training Set) และการทดสอบประสิทธิภาพของตัวแบบ (Testing Set) ด้วยเทคนิค Cross Validation โดยผู้วิจัยได้ทำการแบ่งข้อมูลออกเป็น 10 ส่วนเท่า ๆ กัน (10-fold cross validation) [9] ซึ่งเทคนิคที่ใช้สำหรับการสร้างตัวแบบการพยากรณ์ผู้วิจัยได้ใช้เทคนิคในการประมาณค่าข้อมูล (Regression Model) ทั้งหมด 4 เทคนิค ได้แก่ เทคนิคการถดถอยเชิงเส้น (Linear Regression) เทคนิคโครงข่ายประสาทเทียม (Neural Network) เทคนิคการจำแนกประเภทข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน (Support

Vector Machines) และเทคนิคการเรียนรู้เชิงลึก (Deep Learning) จากนั้นทำการทดสอบประสิทธิภาพกับตัวแบบพยากรณ์เพื่อหาตัวแบบที่มีความเหมาะสมสำหรับการนำไปใช้ในการพยากรณ์ฝุ่นละอองขนาดเล็ก ดังภาพที่ 1



ภาพที่ 1 ขั้นตอนการดำเนินงาน

ในการสร้างตัวแบบสำหรับการพยากรณ์ฝุ่นละอองขนาดเล็ก และการเปรียบเทียบประสิทธิภาพสำหรับการพยากรณ์ฝุ่นละอองขนาดเล็กในอากาศ โดยใช้เทคนิคการเรียนรู้ของเครื่อง ผู้วิจัยได้ดำเนินการตามกระบวนการวิเคราะห์ข้อมูล (Data Analysis Process) ซึ่งประกอบด้วย 6 ขั้นตอน ดังนี้

1. การทำความเข้าใจปัญหา (Business Understanding)

จังหวัดอุดรธานี นับเป็นพื้นที่เขตเมืองใหญ่เมืองหนึ่งในปัจจุบัน มีประชากรหนาแน่น และยังมีการพัฒนาในด้านต่าง ๆ ไม่ว่าด้านเศรษฐกิจที่พบการขยายตัวในกิจกรรม การก่อสร้างอาคารพานิชย์ สร้างอาคารหอพัก การเผาไหม้เชื้อเพลิงการจราจร การเผาป่าไม้ และโรงงานอุตสาหกรรมที่ใช้เชื้อเพลิงฟอสซิล ซึ่งเป็นปัญหาหลักที่ก่อให้เกิดมลพิษในอากาศ ซึ่งมีรายงานจากองค์การอนามัยโลกว่ามลพิษทางอากาศส่งผลกระทบต่อสุขภาพของมนุษย์ และพบว่าประชากรทั่วโลกมากกว่า 2.7

ล้านคน ที่เสียชีวิตจากมลพิษทางอากาศ (Word Health Organization) [10] จากรายงานสถานการณ์มลพิษของประเทศไทยของกรมควบคุมมลพิษ พบว่าในปี พ.ศ.2554 ปัญหามลพิษทางอากาศ (กลิ่นเหม็น ฝุ่นละออง เขม่าควัน) และเสียงเป็นปัญหามากที่สุด คิดเป็นร้อยละ 85 จากการศึกษางานวิจัยในอดีตพบว่างานวิจัยในการสร้างตัวแบบพยากรณ์สำหรับการพยากรณ์ฝุ่นละอองขนาดเล็กในอากาศ ด้วยเทคนิคการเรียนรู้ของเครื่องยังไม่มีการศึกษาในเรื่องนี้อย่างจริงจัง ดังนั้นผู้วิจัยจึงสนใจศึกษาสร้างตัวแบบและเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์ฝุ่นละอองขนาดเล็กในอากาศ ด้วยเทคนิคการเรียนรู้ของเครื่อง

2. การทำความเข้าใจข้อมูล (Data Understanding)

ผู้วิจัยได้ดำเนินการรวบรวมข้อมูลที่ได้จากสถานีตรวจวัดคุณภาพอากาศและอุตุนิยมวิทยาของกรมควบคุมมลพิษ ในพื้นที่ตำบลหมากแข้ง อำเภอเมือง จังหวัดอุดรธานี ที่ถูกจัดเก็บในรูปแบบรายชั่วโมง ตั้งแต่วันที่ 25 เดือนมิถุนายน ถึงวันที่ 31 เดือนธันวาคม พ.ศ. 2564 จำนวนทั้งสิ้น 4,608 แถว ที่ถูกจัดเก็บไว้ในรูปแบบไฟล์ Excel สำหรับข้อมูลที่นำใช้ในการวิเคราะห์นี้ คือปัจจัยที่ทำให้เกิดฝุ่นละอองขนาดเล็กในอากาศ จำนวน 16 ตัวแปร ประกอบไปด้วย วัน (DATE), เวลา (TIME), PM_{10}, $PM_{2.5}$, ก๊าซคาร์บอนไดออกไซด์ (CO), ไนตริออกไซด์ (NO), ไนโตรเจนออกไซด์ (NO_X), ไนโตรเจนไดออกไซด์ (NO_2), ก๊าซซัลฟอร์ไดออกไซด์ (SO_2) , ก๊าซโอโซน (O_3) , ความเร็วลม (WS), ทิศทางลม (WD), อุณหภูมิ (TEMP), ความชื้นสัมพัทธ์ (RH), ความกดอากาศ (BP) และ ฝน (RAIN) ดังแสดงในตารางที่ 1

ตารางที่ 1 ข้อมูลที่ใช้ในการสร้างตัวแบบการพยากรณ์

ลำดับ	ตัวแปร	คำอธิบาย	ประเภทข้อมูล
1	DATE	วัน/เดือน/ปี	Integer
2	TIME	ชั่วโมง	Integer
3	PM_{10}	ฝุ่นละอองขนาดไม่เกิน10ไมครอน	Integer
4	$PM_{2.5}$	ฝุ่นละอองขนาดเล็กกว่า 2.5	Integer
5	CO	ก๊าซคาร์บอนไดออกไซด์	Real
6	NO	ไนตริออกไซด์	Real
7	NO_x	ไนโตรเจนออกไซด์	Real
8	NO_2	ไนโตรเจนไดออกไซด์	Real
9	SO_2	ก๊าซซัลฟอร์ไดออกไซด์	Integer
10	O_3	ก๊าสโอโซน	Real
11	WS	ความเร็วลม	Real

ลำดับ	ตัวแปร	คำอธิบาย	ประเภทข้อมูล
12	WD	ทิศทางลม	Integer
13	TEMP	อุณหภูมิ	Real
14	RH	ความชื้นสัมพัทธ์	Integer
15	BP	ความกดอากาศ	Integer
16	RAIN	ฝน	Real

3. การเตรียมข้อมูล (Data Preparation)

การเตรียมข้อมูลเป็นขั้นตอนการทำความสะอาดข้อมูลให้เกิดความมั่นใจในคุณภาพของข้อมูลก่อนที่จะนำมาวิเคราะห์ซึ่งแสดงให้เห็นถึงความน่าเชื่อถือก่อนนำไปวิเคราะห์ข้อมูล สำหรับงานวิจัยในครั้งนี้ผู้วิจัยได้ดำเนินเตรียมข้อมูล ดังนี้

3.1 การทำความสะอาดข้อมูล (Data Cleaning)

เนื่องจากข้อมูลของการพยากรณ์ฝุ่นละอองขนาดเล็ก ในทุกตัวแปรมีค่าสูญหาย (Missing Value) ผู้วิจัยจึงได้ทำการแทนค่าสูญหาย (Replace Missing Values) ด้วยการหาค่าเฉลี่ยจากข้อมูล 5 วันก่อนหน้า ณ ชั่วโมงที่เกิดค่าสูญหายเกิดขึ้น และ 5 วันถัดไปของวันที่เกิดค่าสูญหาย [11] เนื่องจากข้อมูลที่เกิดค่าสูญหายเหล่านี้เป็นข้อมูลที่อยู่ในรูปแบบของข้อมูลเชิงปริมาณ [9] โดยจะแทนค่าสูญหายนี้ด้วยค่าเฉลี่ยกับค่าในชั่วโมงที่เกิดค่าสูญหายเกิดขึ้น ณ ขณะนั้น

3.2 การแปลงข้อมูล (Data Transformation)

ผู้วิจัยได้ทำการแปลงข้อมูลในแอตทริบิวต์ของชั่วโมง (TIME) ที่ถูกจัดเก็บข้อมูลในรูปแบบเวลาตั้งแต่ 00:00-23:59 โดยผู้วิจัยได้ทำการแปลงข้อมูลให้ออกเป็น 24 ชั่วโมง โดยกำหนดให้ชั่วโมงที่ 00:00-01:00 แปลงข้อมูลเป็นชั่วโมงที่ 1 ชั่วโมงที่ 01:01-02:00 แปลงข้อมูลเป็นชั่วโมงที่ 2 และชั่วโมงที่ 02:01-03:00 แปลงข้อมูลเป็นชั่วโมงที่ 3 เป็นต้น

3.3 การกำหนดหน้าที่ของตัวแปร (Set Role)

การกำหนดหน้าที่ให้แต่ละแอตทริบิวต์เพื่อไปใช้ในการวิเคราะห์ให้แต่ละแอตทริบิวต์ โดยกำหนดแอตทริบิวต์วันที่เก็บข้อมูล (Date) ให้ทำหน้าที่เป็นไอดี (ID) และกำหนดแอตทริบิวต์ $PM_{2.5}$ ให้ทำหน้าที่เป็นตัวแปรตาม (Label) ที่ใช้สำหรับการพยากรณ์ สำหรับแอตทริบิวต์ที่เหลือจะมีการกำหนดหน้าที่ให้เป็นตัวแปรอิสระ

4. การสร้างตัวแบบ (Modeling)

ขั้นตอนนี้จะนำข้อมูลมาวิเคราะห์ตามอัลกอริทึมด้วยเทคนิคการการเรียนรู้ของเครื่องสำหรับการการประมาณค่าข้อมูล (Regression Model) โดยในงานวิจัยในครั้งนี้ ผู้วิจัยได้เลือกใช้เทคนิคการเรียนรู้ของเครื่องทั้งหมด 4 เทคนิค ได้แก่ เทคนิคการถดถอยเชิงเส้น (Linear Regression) เทคนิคโครงข่ายประสาทเทียม (Neural Network) เทคนิคซัพพอร์ตเวกเตอร์-แมชชีน (Support Vector Machines) และเทคนิคการเรียนรู้เชิงลึก (Deep Learning) โดยใช้โปรแกรม RapidMiner Studio Version 10 [9] ในการสร้างตัวแบบในครั้งนี้

4.1 เทคนิคการถดถอยเชิงเส้น (Linear Regression)

การวิเคราะห์การถดถอยเป็นวิธีการทางสถิติที่ใช้ศึกษาความสัมพันธ์ระหว่างตัวแปรอิสระ (Independent Variable) กับตัวแปรตาม (Dependent Variable) โดยตัวแปรทั้งสองมีความสัมพันธ์เชิงเส้นตรงกัน อาจเป็นความสัมพันธ์ตามกันหรือผกผันก็ได้ รูปแบบที่ใช้ในการวิเคราะห์เป็นรูปแบบพื้นฐานอย่างง่ายที่สุดของการวิเคราะห์การถดถอยโดยมีตัวแบบการถดถอยมาสร้างเป็นสมการเชิงเส้น [9] โดยมีรูปแบบของสมการ ดังนี้

$$y=b_0+b_1x_1+b_2x_2+...+b_kx_k+e \quad (1)$$

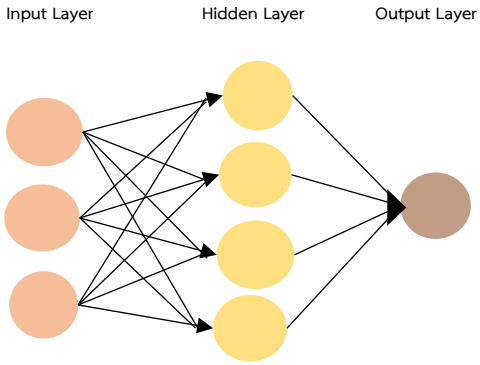
โดยที่ y คือ ตัวแปรตาม (Dependent Variable)

b_0 คือ ค่าคงที่ของสมการถดถอยเป็นค่าจุดตัด แกน y ของสมการ

$b_1, b_2,..., b_k$ คือ ค่าสัมประสิทธิ์การถดถอยของตัวแปรต้น $x_1, x_2,.., x_k$

4.2 เทคนิคโครงข่ายประสาทเทียม (Neural Network)

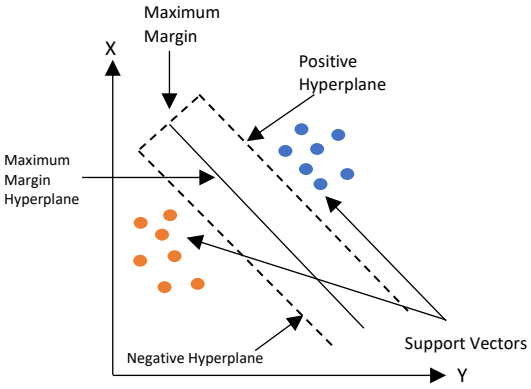
เป็นวิธีที่พัฒนาจากการทำงานของเซลล์ประสาททางชีววิทยา (Biological Neuron) ซึ่งมีความสามารถในการปรับตัวเมื่อข้อมูลไม่เป็นเชิงเส้น ความสามารถในการเรียนรู้การจดจำรูปแบบ และทำนายอนาคต หลังจากที่ได้รับการกระตุ้นด้วยสิ่งเร้าภายนอกหรือกระตุ้นด้วยเซลล์ด้วยกัน กระแสประสาทจะวิ่งผ่านเดนไดรท์เข้าสู่นิวเคลียสโดยจะเป็นตัวตัดสินว่าต้องกระตุ้นเซลล์อื่น ๆ ต่อหรือไม่ [9] ดังภาพที่ 2



ภาพที่ 2 สถาปัตยกรรมเทคนิคโครงข่ายประสาทเทียม

4.3 เทคนิคซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machines)

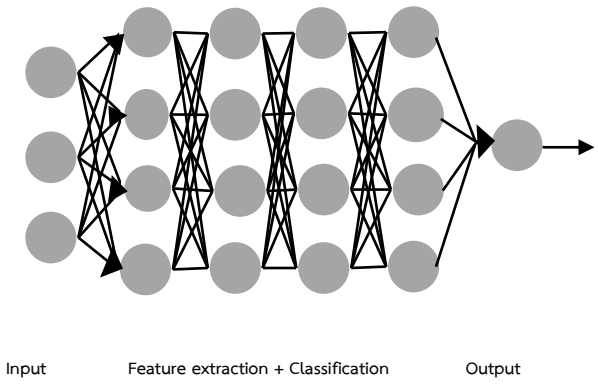
เทคนิคซัพพอร์ตเวกเตอร์แมชชีนเป็นแบบจำลองการเรียนรู้ของเครื่องแบบผู้สอนที่ใช้ในการแบ่งข้อมูลออกเป็นประเภทต่าง ๆ ที่ได้กำหนดไว้ซึ่งตัวแบบจำลองนั้นมีรูปแบบการทำงานคือสร้างตัวจำแนกประเภท (Classifier) โดยตัวจำแนกประเภทของแบบจำลองซัพพอร์ตเวกเตอร์แมชชีนนั้นโดยลักษณะเป็นระนาบเกิน (Hyperplane) ซึ่งการหาระนาบเกินอาจเจอปัญหา Overfitting ได้ดังนั้น เพื่อที่จะเลือกระนาบเกินที่มีความเหมาะสม (Optimal Hyperplane) ที่ใช้ระยะขอบที่มากที่สุด (Maximized Margin) เป็นเกณฑ์ในการเลือกระนาบเกินให้มีความเหมาะสมโดยปัญหาการหาระนาบเกินที่มีความเหมาะสมนั้นสามารถเขียนให้อยู่ในรูปปัญหาการหาค่าสูงสุดได้ ดังภาพที่ 3



ภาพที่ 3 เทคนิคซัพพอร์ตเวกเตอร์แมชชีน (SVM)

4.4 เทคนิคการเรียนรู้เชิงลึก (Deep Learning)

เทคนิคการเรียนรู้เชิงลึก (Deep Learning) เป็นการพัฒนาให้คอมพิวเตอร์สามารถเลียนแบบการทำงานระบบโครงข่ายประสาท (Neurons) เหมือนกับสมองมนุษย์ เรียกว่าโครงข่ายประสาทเทียม (Neural Network: NN) โดย Deep Learning ถูกสร้างขึ้นจากการนำ Neural Network หลาย ๆ Layer มาใช้วิเคราะห์และหาคำตอบซึ่งคำ Deep Learning ก็มาจากการใช้ Neural Network มากกว่า 2 Layer เพื่อให้คอมพิวเตอร์เรียนรู้และสร้างแบบจำลองจึงเปรียบเทียบได้ว่า Layer ของ Neural Network ถูกใช้จำนวนมากสำหรับขั้นตอนการประมวลผลทำให้มีโครงสร้างการเรียนรู้ที่ลึก (Deep) ยิ่งขึ้น ดังแสดงในภาพ 4 โดยวิธีการเรียนรู้เชิงลึกนี้ได้ถูกนำเสนอหลายแบบ เช่น โครงข่ายประสาทเทียมแบบสังวัตนาการ (Convolutional Neural Network: CNN) โครงข่ายประสาทเทียมแบบวนซ้ำ (Recurrent Neural Network: RNN) เป็นต้น ซึ่งในงานวิจัยนี้จะใช้เทคนิค H2O Deep Learning เพื่อใช้สำหรับการสร้างตัวแบบการพยากรณ์ฝุ่นละอองในอากาศขนาดเล็ก โดยเทคนิค H2O มีลักษณะหลายชั้นแบบป้อนไปข้างหน้าซึ่งได้รับการฝึกฝนด้วยการไล่ระดับสีแบบสุ่มโดยใช้การเผยแพร่กลับ และสามารถมี Layer ที่ซ่อนอยู่จำนวนมาก ซึ่งประกอบด้วยเซลล์ประสาทที่มีฟังก์ชัน tanh, rectifier และ maxout activation ซึ่งการทำให้เป็นปกติของ L1 หรือ L2 การเช็คพอยต์ และการค้นหากริดช่วยให้สามารถคาดการณ์ได้อย่างแม่นยำสูง [12] ดังภาพที่ 4



ภาพที่ 4 สถาปัตยกรรมเทคนิคการเรียนรู้เชิงลึก (Deep Learning)

5. การทดสอบประสิทธิภาพของตัวแบบ (Evaluation)

หลังจากที่ได้สร้างตัวแบบขึ้นมาได้แล้ว ขั้นตอนถัดมาจะต้องทำการวัดประสิทธิภาพการจำแนกประเภทข้อมูลของตัวแบบที่สร้างได้ ก่อนที่จะนำตัวแบบไปใช้ในการพยากรณ์ สำหรับงานวิจัยนี้ ผู้วิจัยได้ใช้เกณฑ์ในการวัดค่าความแม่นยำของตัวแบบซึ่งมี 5 วิธี [9] ประกอบด้วย

5.1 ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย

ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Squared Error: MSE) เป็นการวัดค่าความคลาดเคลื่อน โดยการนำค่าความคลาดเคลื่อนมายกกำลังแล้วนำไปหาค่าเฉลี่ย ในการวัดค่าความแม่นยำจากวิธีการนี้ยิ่งค่าที่ได้มีค่าน้อยแสดงว่าตัวแบบที่ได้จะมีความแม่นยำมาก โดยมีสูตรดังนี้

$$MSE=\frac{1}{n}\sum_{t=1}^{n}\left(Y_t-\hat{Y}_t\right)^2 \qquad (2)$$

โดยที่ n คือ จำนวนข้อมูลที่ใช้

Y_t คือ ค่าจริงที่เวลา t ใด ๆ

$\hat{Y}_t$ คือ ค่าที่ได้จากการพยากรณ์ที่เวลา t ใด

5.2 ค่ารากที่สองของความคลาดเคลื่อนเฉลี่ยกำลังสอง

ค่ารากที่สองของความคลาดเคลื่อนเฉลี่ยกำลังสอง (Root Mean Squared Error: RMSE) เป็นวิธีการวัดค่าความคลาดเคลื่อนแบบมาตรฐาน ซึ่งนิยมใช้กันอย่างแพร่หลาย โดยมีสมการที่ใช้ในการวัดค่าความแม่นยำจากวิธีการนี้ยิ่งค่า RMSE ที่ได้ค่าน้อยแสดงว่าตัวแบบที่ได้จะมีความแม่นยำมาก โดยมีสูตรดังนี้

$$RMSE=\sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(Y_t-\hat{Y}_t\right)^2} \qquad (3)$$

โดยที่ n คือ จำนวนข้อมูลที่ใช้

Y_t คือ ค่าจริงที่เวลา t ใด ๆ

$\hat{Y}_t$ คือ ค่าที่ได้จากการพยากรณ์ที่เวลา t ใด

5.3 ค่าคลาดเคลื่อนสัมบูรณ์ (Absolute Error)

ค่าคลาดเคลื่อนสัมบูรณ์ (Absolute Error: AE) การนำค่าความคลาดเคลื่อนมาหาค่าสัมบูรณ์ โดยมีสูตรดังนี้

$$AE = \left|Y_t - \hat{Y}_t\right| \quad (4)$$

โดยที่ n คือ จำนวนข้อมูลที่ใช้

Y_t คือ ค่าจริงที่เวลา t ใด ๆ

$\hat{Y}_t$ คือ ค่าที่ได้จากการพยากรณ์ที่เวลา t ใด

5.4 ค่าความคลาดเคลื่อนยกกำลังสอง (Square Error)

ค่าคลาดเคลื่อนยกกำลังสอง (Square Error: SE) คือการนำค่าความคลาดเคลื่อนมายกกำลังสอง โดยมีสูตรดังนี้

$$SE = \left(\left|Y_t - \hat{Y}_t\right|\right)^2 \quad (5)$$

โดยที่ n คือ จำนวนข้อมูลที่ใช้

Y_t คือ ค่าจริงที่เวลา t ใด ๆ

$\hat{Y}_t$ คือ ค่าที่ได้จากการพยากรณ์ที่เวลา t ใด

5.5 สัมประสิทธิ์การตัดสินใจ (Coefficient of Determination: R^2)

สัมประสิทธิ์การตัดสินใจ (Coefficient of Determination: R^2) คือ สัดส่วนที่ตัวแปรอิสระสามารถอธิบายการเปลี่ยนแปลงของตัวแปรตามได้ ดังนั้นถ้า R^2 มีค่ามากแสดงว่าตัวแปรตามและตัวแปรอิสระ มีความสัมพันธ์กันมากหรือแสดงว่าตัวแปรอิสระของสมการความถดถอยนั้นสามารถอธิบายการเปลี่ยนแปลงของตัวแปรตามได้มาก โดยมีสูตรดังนี้

$$R^2 = 1 - \left(\frac{\sum(Y-\hat{Y})^2}{\sum(Y-\bar{Y})^2}\right) \quad (6)$$

โดยที่ n คือ จำนวนข้อมูลที่ใช้

Y_t คือ ค่าจริงที่เวลา t ใด ๆ

$\hat{Y}_t$ คือ ค่าที่ได้จากการพยากรณ์ที่เวลา t ใด

6. การนำไปใช้ (Deployment)

ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลในครั้งนี้จะได้ตัวแบบที่เหมาะสมสำหรับการสร้างตัวแบบการพยากรณ์ฝุ่นละอองขนาด-เล็กในอากาศ เพื่อสนับสนุนให้หน่วยงานที่เกี่ยวข้องสามารถนำไปใช้ในการวิเคราะห์ข้อมูลการเกิดฝุ่นละอองขนาดเล็กในอากาศ และสามารถวางแผนนโยบายอีกทั้งมาตรการต่าง ๆ สำหรับการป้องกันการเกิดฝุ่นละออง $PM_{2.5}$ ที่จะเกิดขึ้นในอนาคต และนอกจากนี้ยังสามารถนำผลการวิจัยในครั้งนี้ไปพัฒนาต่อยอดในการสร้างระบบสารสนเทศสำหรับการพยากรณ์เกี่ยวกับฝุ่น โดยอาจเพิ่มปัจจัยในส่วนอื่น ที่ยังไม่ได้นำมาศึกษาในครั้งนี้ เพื่อให้มีความแม่นยำสำหรับการพยากรณ์มากยิ่งขึ้น

ผลการวิจัย

ผู้วิจัยได้ทำการเปรียบเทียบค่าที่ทดสอบประสิทธิภาพของตัวแบบการพยากรณ์ฝุ่นละอองขนาดเล็ก ด้วยเทคนิคการเรียนรู้ของเครื่องทั้ง 4 เทคนิค ได้แก่ เทคนิคการถดถอยเชิงเส้น (Linear Regression) เทคนิคโครงข่ายประสาทเทียม (Neural Network) เทคนิคซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machines) และเทคนิคการเรียนรู้เชิงลึก (Deep Learning) และทำการทดสอบประสิทธิภาพของการประมาณค่าด้วยค่า ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Squared Error) ค่ารากที่สองของความคลาดเคลื่อนเฉลี่ยกำลังสอง (Root Mean Squared Error) ค่าความคลาดเคลื่อนสัมบูรณ์ (Absolute Error) ค่าความคลาดเคลื่อนกำลังสอง (Squared Error) และให้ค่าสัมประสิทธิ์การตัดสินใจพหุคูณ (R^2) ซึ่งผลการวิเคราะห์ข้อมูลได้ผลแสดงดังตารางที่ 2

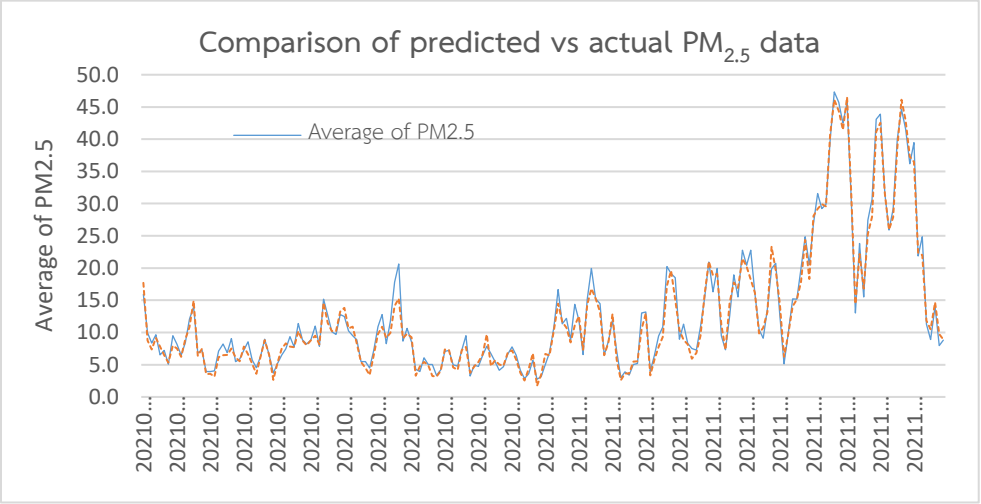
ตารางที่ 2 การเปรียบเทียบประสิทธิภาพของตัวแบบการประมาณค่าข้อมูลทั้ง 4 เทคนิค

No.	Regression Model	Mean Square Error (MSE)	Root Mean Squared Error (RMSE)	Absolute Error (AE)	Squared Error (SE)	R-Squared (R^2)
1	Linear Regression	10.850436	3.294	2.322	10.877	0.929
2	Neural Network*	8.673025	2.945	2.126	8.709	0.948
3	Support Vector Machines	11.390625	3.375	2.313	11.418	0.928
4	Deep Learning	12.852225	3.585	2.559	13.055	0.935

* คือ เทคนิคที่มีความเหมาะสมสำหรับนำมาสร้างตัวแบบการพยากรณ์ฝุ่นละอองขนาดเล็กในอากาศ ($PM_{2.5}$)

จากตารางที่ 2 พบว่าเทคนิคที่มีความหมาะสมที่สุดสำหรับการสร้างตัวแบบการพยากรณ์ฝุ่นละอองขนาดเล็กในอากาศ ($PM_{2.5}$) คือเทคนิคโครงข่ายประสาทเทียม โดยให้ค่าความคลาดเคลื่อน

กำลังสองเฉลี่ย เท่ากับ 8.673 ค่ารากที่สองของความคลาดเคลื่อนเฉลี่ยกำลังสอง เท่ากับ 2.945 ค่าความคลาดเคลื่อนสัมบูรณ์ เท่ากับ 2.126 ค่าความคลาดเคลื่อนกำลังสองเท่ากับ 8.709 ซึ่งมีค่าความคลาดเคลื่อนน้อยที่สุดเมื่อเทียบกับเทคนิคอื่น ๆ และให้ค่าสัมประสิทธิ์การตัดสินใจพหุคูณ สูงที่สุดเมื่อเทียบกับเทคนิคอื่น ๆ ซึ่งมีค่าเท่ากับ 0.948 หรือ 94.8% และผู้วิจัยได้ทำการเปรียบเทียบข้อมูลระหว่างค่าที่ได้จากข้อมูลจริงของฝุ่นละอองขนาดเล็กในอากาศ และค่าจากการพยากรณ์ของฝุ่นละอองขนาดเล็กในอากาศ ด้วยเทคนิควิธี (Neural Network) ดังภาพที่ 5



ภาพที่ 5 การเปรียบเทียบข้อมูลระหว่างค่าจากข้อมูลจริงและค่าจากการพยากรณ์

จากภาพที่ 5 แสดงให้เห็นว่าข้อมูลของค่าเฉลี่ยของฝุ่นละอองขนาดเล็กในอากาศ และข้อมูลพยากรณ์ของค่าเฉลี่ยของฝุ่นละอองขนาดเล็กในอากาศ มีค่าใกล้เคียงกันมาก ซึ่งแสดงให้เห็นว่าเทคนิคที่มีประสิทธิภาพในการนำไปใช้ในการพยากรณ์การเกิดฝุ่นละอองขนาดเล็กในอากาศ ($PM_{2.5}$) มากที่สุดคือ เทคนิคโครงข่ายประสาทเทียม (Neural Network) ที่ให้ค่าประสิทธิภาพแม่นยำมากที่สุด

อภิปรายผลการวิจัย

จากวัตุประสงค์ของการวิจัยการสร้างตัวแบบและเปรียบเทียบประสิทธิภาพตัวแบบสำหรับการพยากรณ์ฝุ่นละอองขนาดเล็กในอากาศด้วยเทคนิคการเรียนรู้ของเครื่อง จากผลการวิจัยสามารถสรุปได้ว่าเทคนิคที่มีความเหมาะสมที่สุดสำหรับการสร้างตัวแบบการพยากรณ์ฝุ่นละอองขนาดเล็กในอากาศ คือ เทคนิคโครงข่ายประสาทเทียม (Neural Network) โดยให้ค่าความคลาดเคลื่อนน้อยที่สุดเมื่อเทียบเทคนิคอื่น ๆ และให้ค่าสัมประสิทธิ์การตัดสินใจพหุคูณสูงที่สุดเมื่อเทียบกับเทคนิคอื่น ๆ

โดยมีค่าเท่ากับ 0.948 ซึ่งตัวแปรปัจจัยที่ทำให้เกิดฝุ่นที่ใช้ในการสร้างตัวแบบพยากรณ์มีความสอดคล้องกับเวชกร ไกรยราช [2] ที่ได้ศึกษางานวิจัยเรื่องการพยากรณ์ฝุ่นละอองขนาดเล็กในอากาศ PM2.5 ด้วยเครือข่ายประสาทเทียมโดยได้ใช้ตัวแปรที่สอดคล้องกัน มีดังนี้ วัน/เดือน/ปี (DATE) ชั่วโมง (TIME) ฝุ่นละอองขนาดเล็กกว่า 2.5 ($PM_{2.5}$) อุณหภูมิ (TEMP) ทิศทางลม (WD) ความชื้นสัมพัทธ์ (RS) และฝน (RAIN) และ เทคนิคที่ใช้ในการสร้างตัวแบบพยากรณ์มีความสอดคล้องกับงานวิจัยของรณชัย ชื่นธวัช [13] ที่ได้ศึกษางานวิจัยเรื่องการพยากรณ์ค่าความเข้มข้นของฝุ่นละอองขนาดเล็กไม่เกิน 10 ไมครอน ด้วยเครือข่ายประสาทเทยีมร่วมกับรูปแบบออโตรีเกรสซีพ ซึ่งได้มีการใช้เทคนิคโครงข่ายประสาทเทียม (Neural Network) ร่วมกับรูปแบบออโตรีเกรสซีพ (Autoregressive) พบว่า ตัวแบบผสมผสาน AR-ANN สามารถนำมาใช้พยากรณ์ค่าความ เข้มข้นเฉลี่ยรายวันของ PM_{10} ในเขตพื้นที่ภาคเหนือของประเทศไทย ได้อย่างเหมาะสมสำหรับการสร้างตัวแบบในการพยากรณ์ ดังนั้นผลการวิจัยนี้สามารถนำไปเปรียบเทียบเทคนิคอื่น ๆ ที่ไม่นำมาศึกษาในครั้งนี้ได้

สรุปผลการวิจัย

จากการศึกษาในครั้งนี้มีวัตถุประสงค์หลักเพื่อสร้างตัวแบบและเปรียบเทียบประสิทธิภาพตัวแบบสำหรับการพยากรณ์ฝุ่นละอองขนาดเล็กในอากาศด้วยเทคนิคการเรียนรู้ของเครื่อง โดยใช้ข้อมูลปัจจัยที่ทำให้เกิดฝุ่นละอองขนาดเล็ก จำนวน 16 ตัวแปร จากสถานีตรวจวัดคุณภาพอากาศและอุตุนิยมวิทยา ของกรมควบคุมมลพิษ ในพื้นที่ ตำบลหมากแข้ง อำเภอเมือง จังหวัดอุดรธานี ผลการวิจัยสามารถสรุปได้ว่าเทคนิคที่มีความเหมาะสมที่สุดสำหรับการสร้างตัวแบบการพยากรณ์ฝุ่นละอองขนาดเล็กในอากาศ คือ เทคนิคโครงข่ายประสาทเทียม (Neural Network) โดยให้ค่าความคลาดเคลื่อนน้อยที่สุดเมื่อเทียบเทคนิคอื่น ๆ โดยให้ค่ารากที่สองของความคลาดเคลื่อนเฉลี่ยกำลังสอง เท่ากับ 2.945 ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย เท่ากับ 8.673 ค่าความคลาดเคลื่อนสัมบูรณ์เท่ากับ 2.126 ค่าความคลาดเคลื่อนกำลังสอง เท่ากับ 8.709 และให้ค่าสัมประสิทธิ์การตัดสินใจพหุคูณ สูงที่สุดเมื่อเทียบกับเทคนิคอื่น ๆ โดยมีค่าเท่ากับ 0.948 ซึ่งหมายความว่า ตัวแปรอิสระที่ใช้ศึกษาในครั้งนี้สามารถอธิบายการเปลี่ยนแปลงของการเกิดฝุ่นละอองขนาดเล็กในอากาศ ได้ถึง 94.8% นอกนั้นเป็นอิทธิพลอื่น ๆ ที่ไม่ได้นำมาศึกษาในครั้งนี้

ข้อเสนอแนะ

ผลการวิจัยครั้งนี้สามารถใช้สำหรับการพยากรณ์ฝุ่นละอองขนาดเล็กในอากาศ $PM_{2.5}$ ในพื้นที่ ตำบลหมากแข้ง อำเภอเมือง จังหวัดอุดรธานี เท่านั้น หากมีผู้ที่สนใจศึกษาการสร้างตัวแบบ

สำหรับการพยากรณ์ฝุ่นละอองขนาดเล็กในอากาศ ในอนาคตสามารถใช้เทคนิคการเรียนรู้ของเครื่องเทคนิคอื่น ๆ รวมถึงพื้นที่ที่ต้องการจะศึกษา และช่วงเวลาในการเกิดปัญหาของการเกิดฝุ่นละอองขนาดเล็กในอากาศที่ยังไม่ได้ใช้ในการศึกษาครั้งนี้มาวิเคราะห์ เพื่อให้เกิดความแม่นยำของการพยากรณ์มากยิ่งขึ้น นอกจากนี้ยังสามารถเพิ่มตัวแปรหรือปัจจัยที่ส่งผลทำให้เกิดฝนละออง ตัวอื่น ๆ ที่ไม่ได้นำมาศึกษาในครั้งนี้มาทำการวิเคราะห์ในอนาคต

กิตติกรรมประกาศ

งานวิจัยในครั้งนี้ได้รับการสนับสนุนภายใต้เครือข่ายนักวิจัยสหวิทยาการ หน่วยปฏิบัติการวิจัยการเปลี่ยนแปลงสภาพภูมิอากาศ การบรรเทา และการปรับตัว (CMARE) สังกัดคณะสิ่งแวดล้อมและทรัพยากรศาสตร์ มหาวิทยาลัยมหาสารคาม ที่ให้ข้อมูลปัจจัยที่ทำให้เกิดฝุ่นละอองในอากาศขนาดเล็ก เพื่อใช้ในการทำวิจัย และขอขอบคุณคณะการบัญชีและการจัดการ มหาวิทยาลัยมหาสารคาม ที่ให้การสนับสนุนในการทำวิจัยครั้งนี้

เอกสารอ้างอิง

1. DAIKIN THAILAND. $PM_{2.5}$ คืออะไร? อันตรายและการป้องกันฝุ่นละอองขนาดเล็ก [อินเทอร์เน็ต] 2563. [เข้าถึงเมื่อ 25 มีนาคม 2566]. เข้าถึงได้จาก: https://www.daikin.co.th/service- knowledge/pm-2-5/
2. เวชกร ไกรยราช, ภัทรพงศ์ บัณฑิตา. แอปพลิเคชันพยากรณ์ปริมาณฝุ่นละอองขนาดเล็กในอากาศ $PM_{2.5}$ ด้วยโครงข่ายประสาทเทียม (ปริญญานิพนธ์ปริญญาวิทยาศาสตรบัณฑิต). คณะวิทยาการสารสนเทศ. มหาสารคาม:มหาวิทยาลัยมหาสารคาม; 2564.
3. บรรจบ ชุณหสวัสดิกุล, พยงค์ วณิเกียรติ, อัมพร กรอบทอง, กมล ไชยสิทธิ. ผลต่อสุขภาพของฝุ่นละอองในอากาศขนาดไม่เกิน 2.5 ไมครอน กลไกก่อให้เกิดโรค และการรักษาด้วยการแพทย์ทางเลือก. วารสารการแพทย์แผนไทยและการแพทย์ทางเลือก. 2563; 18(1): 187-202.
4. อินท์ฉัตร สุขเกษม. การพยากรณ์ปริมาณฝุ่นละอองขนาดไม่เกิน 2.5 ไมครอน (PM2. 5) จังหวัดนครราชสีมา. The office of disease prevention and control 9th Nakhon Ratchasima Journal. 2021; 27(1): 16-25.
5. ธรณินทร์ สัจวิริยทรัพย์. ตัวแบบผสมอารีมา-ซัพพอร์ตเวกเตอร์แมชชีนเน้น องค์ประกอบเชิงบวกสำหรับการพยากรณ์ปริมาณฝุ่นละอองขนาดไม่เกิน 10 ไมครอน. Thai Science and Technology Journal. 2019; 27(3): 526-538.

6. พรนภา แสงศรี, พรพิมล ชัยวุฒิศักดิ์. วิธีผสมของโครงข่ายประสาทเทียมและห่วงโซ่มาร์คอฟสำหรับพยากรณ์ปริมาณ ความเข้มข้นของฝุ่นละอองขนาด 2.5 ไมครอน. Thai Science and Technology Journal. 2021; 29(5): 752-765. doi: 10.14456/tstj.2021.63.
7. ชญานนท์ เทพแสงพราว, ธงทิศ ฉายากุล, ศิริมา ปัญญาเมธีกุล. การทำนายปริมาณฝุ่นละอองขนาดไม่เกิน 2.5 ไมครอนจากข้อมูลความลึกเชิงแสงของอนุภาคแขวนลอยในอากาศและข้อมูลอุตุนิยมวิทยาในพื้นที่กรุงเทพมหานครและปริมณฑล. ใน: รายงานการประชุมวิชาการวิศวกรรมโยธาแห่งชาติ ปีที่ 26, 23-25 มิถุนายน 2564 การประชุมรูปแบบออนไลน์: หน้า SGI-02-1 - SGI-02-8.
8. ปรัชญา สิงหวรวงศ์. การพยากรณ์ฝุ่น $PM_{2.5}$ ล่วงหน้าในจังหวัดเชียงใหม่ โดยใช้แบบจำลองหน่วยความจำระยะสั้นแบบยาว (สารนิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต). คณะวิทยาศาสตร์. นครนายก: มหาวิทยาลัยศรีนครินทรวิโรฒ; 2563.
9. อนุพงศ์ สุขประเสริฐ, คู่มือการทำเหมืองข้อมูลด้วยโปรแกรม RAPIDMINER STUDIO. พิมพ์ครั้งที่ 4. มหาสารคาม: คณะการบัญชีและการจัดการ มหาวิทยาลัยมหาสารคาม; 2564.
10. World Health organization. WHO guidelines for air quality [อินเทอร์เน็ต]. 2565. [เข้าถึงเมื่อ 25 พฤศจิกายน 2565]. เข้าถึงได้จาก: https://www.who.int/news-room/feature-stories/detail/what-are-the-who-air-quality-guidelines
11. จิรายุทธ เจริญ, สิทธิเดช สารจันทร์, อนุพงศ์ สุขประเสริฐ, ธัญญธร ศรีวิเชียร. การพยากรณ์ราคาสกุลเงินอีเธอเรียมด้วยวิธีการเรียนรู้ของเครื่อง. ใน: รายงานการประชุมวิชาการระดับชาติเครือข่ายวิจัยสถาบันอุดมศึกษาทั่วประเทศ ครั้งที่ 15, 26-28 เมษายน 2565 กรุงเทพฯ: หน้า 196–208.
12. กิตติญา หมุนไหม. โปรแกรมแปลงภาพเป็นคำบรรยายภาษาไทยอัตโนมัติ [อินเทอร์เน็ต]. 2563. [เข้าถึงเมื่อ 25 มกราคม 2566]. เข้าถึงได้จาก: https://shorturl.asia/8gROl.
13. รณชัย ชื่นธวัช. “การพยากรณ์ค่าความเข้มข้นของฝุ่นละอองขนาดเล็กไม่เกิน 10 ไมครอนด้วยเครือข่ายประสาทเทียมร่วมกับรูปแบบออโตรีเกรสซีพ,”.วารสารวิทยาศาสตร์บูรพา ปีที่ 22, ฉบับที่ 3 (กันยายน–ธันวาคม พ.ศ. 2560), 519-537.