

แบบจำลองการพยากรณ์ปริมาณฝุ่นละออง PM_{2.5} ของจังหวัดพะเยา

คมกฤต เมฆสกุล^{1*} ทรงวุฒิ ประกายวิเชียร² อธิคม บุญเชื้อ³

^{1*,2,3}สาขาวิชาวิศวกรรมอุตสาหการ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยพะเยา, พะเยา, ประเทศไทย

*ผู้ประพันธ์บรรณกิจ อีเมล : kit_9m@hotmail.com

รับต้นฉบับ : 3 กุมภาพันธ์ 2566; รับบทความฉบับแก้ไข : 25 เมษายน 2566; ตอรับบทความ : 26 เมษายน 2566

เผยแพร่ออนไลน์ : 27 ธันวาคม 2566

บทคัดย่อ

ฝุ่นละอองในบรรยากาศขนาดเล็กไม่เกิน 2.5 ไมครอน (PM_{2.5}) เป็นปัญหามลพิษทางอากาศที่สำคัญในพื้นที่ภาคเหนือตอนบนของประเทศไทยซึ่งส่งผลกระทบต่อสุขภาพของประชาชนทั้งทางตรงและทางอ้อม โดยจังหวัดพะเยามีมลภาวะนี้จัดอยู่ 3 ลำดับแรกของภาคเหนือ การพยากรณ์ระดับความเข้มข้นของปริมาณฝุ่นละอองในบรรยากาศขนาดเล็กไม่เกิน 2.5 ไมครอน โดยใช้แบบจำลองทางคณิตศาสตร์ เป็นทางเลือกหนึ่งที่สามารถนำไปใช้ประกอบเป็นแนวทางในการวางแผนการแก้ไขหรือป้องกันปัญหามลพิษทางอากาศ การสร้างแบบจำลองการพยากรณ์ด้วยการประยุกต์ใช้วิธีการวิเคราะห์การถดถอยโลจิสติก โดยใช้ชุดข้อมูลจากกรมควบคุมมลพิษทางอากาศ และกรมอุตุนิยมวิทยาที่ตรวจวัดสภาพอากาศในจังหวัดพะเยา ซึ่งนำข้อมูลสถิติย้อนหลัง 5 ปี ในช่วงปี พ.ศ. 2558–2564 ไปสร้างแบบจำลองการพยากรณ์ด้วยโปรแกรมภาษาไพทอน ผลการทดสอบ พบว่า ตัวแปรอิสระที่นำไปสร้างสมการจำลองทางคณิตศาสตร์สำหรับการพยากรณ์ค่า PM_{2.5} ของงานวิจัยนี้ได้แก่ ค่า PM₁₀, NO₂, O₃ และ CO เมื่อนำมาตรวจสอบความถูกต้องของแบบจำลองการพยากรณ์ โดยให้ค่าความไวและค่าความจำเพาะเท่ากับ 0.96 และ 0.72 ตามลำดับ โดยมีค่าพื้นที่ใต้เส้นโค้งเท่ากับ 0.98

คำสำคัญ : มลพิษทางอากาศ การถดถอยโลจิสติก สมการพยากรณ์

The Prediction Model of PM_{2.5} in Phayao Province

Khomkit Meksagul^{1*} Songwut Prakaiwichien² Athikom Boonsue³

^{1*,2,3}*Industrial Engineering Program, School of Engineering, University of Phayao, Phayao, Thailand*

*Corresponding Author. E-mail address: kit_9m@hotmail.com

Received: 3 February 2023; Revised: 25 April 2023; Accepted: 26 April 2023

Published online: 27 December 2023

Abstract

The dust in the atmosphere (Particulate Matter with a diameter less than or equal to 2.5 micrometers; PM_{2.5}) is the most significant air pollution in the northern region of Thailand, directly and indirectly affecting the health of the people. This pollution in Phayao province has consistently ranked in the top 3 of Thailand's northern regions. Forecasting the concentration of PM_{2.5} in the atmosphere using a mathematical model is an alternative that can be employed as a guideline for planning solutions or preventing air pollution problems. Logistic regression analysis and data from the Thai Meteorological Department and Pollution Control Department were applied to this prediction model. The 5 years of data between 2015–2021 were utilized for this model and created using Python programming. The results demonstrate that the independent variables of this model are PM₁₀, NO₂, O₃ and CO. The sensitivity and specificity are 0.96 and 0.72, respectively. The area under the curve (AUC) is 0.98.

Keywords: Air pollution, Logistic regression, Predict model

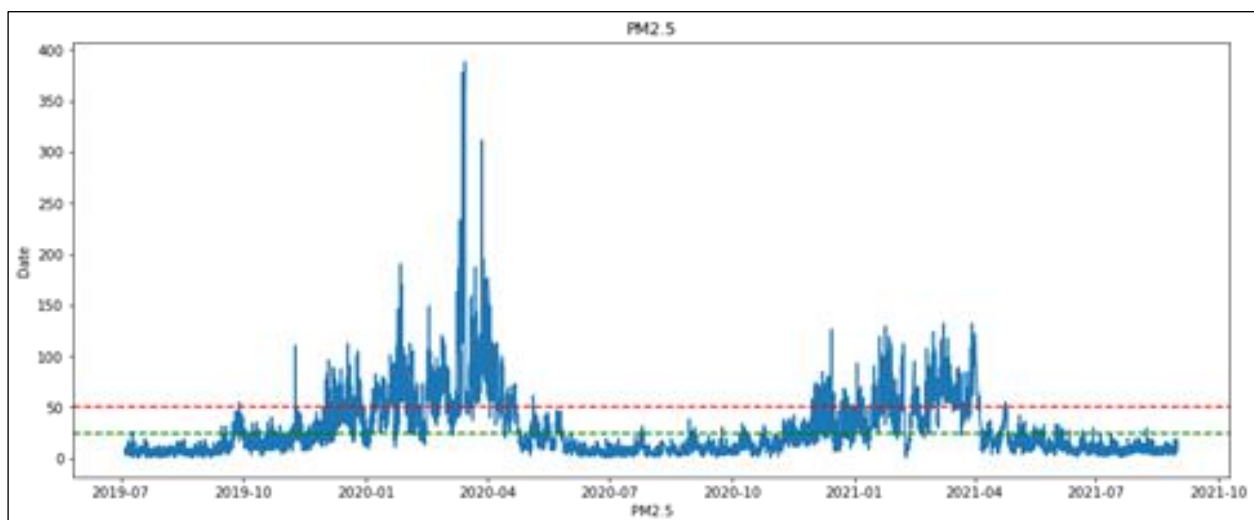
1) บทนำ

ปัญหามลพิษทางอากาศเป็นปัญหาที่สำคัญที่ส่งผลกระทบต่อสุขภาพของประชาชนทั้งทางตรงและทางอ้อม ซึ่งเกิดจากฝุ่นละอองในบรรยากาศ เป็นอนุภาคของแข็งขนาดเล็กที่ลอยลอยอยู่ในอากาศ ซึ่งเกิดจากวัตถุที่ถูกทุบ ตี บด กระแทก จนแตกเป็นชิ้นส่วนเล็ก ๆ หรือปฏิกิริยาทางเคมี เมื่อถูกกระแสนลมพัด ก็จะปลิวกระจายตัวอยู่ในอากาศ และอนุภาคบางส่วนตกลงสู่พื้น [1] ซึ่งปัญหานี้ส่งผลกระทบต่อความเป็นอยู่ของประชาชนในพื้นที่ โดยเฉพาะผลกระทบต่อสุขภาพ ระบบทางเดินหายใจ สภาพเศรษฐกิจและการท่องเที่ยว การเดินทางและการขนส่ง [2] จากการศึกษาผลกระทบจากฝุ่นละอองขนาดเล็กในบรรยากาศ กับจำนวนผู้ป่วยที่มีอาการของโรคที่เกิดจากการได้รับปริมาณฝุ่นเข้าสู่ร่างกาย มีความสอดคล้องกันตามลำดับเวลา เช่น ช่วงเดือนกุมภาพันธ์ถึงกลางเดือนมีนาคม ซึ่งมีระดับฝุ่นละอองขนาดเล็กปริมาณสูง ก็พบว่า มีปริมาณจำนวนผู้ป่วยเพิ่มมากขึ้น และช่วงเดือนเมษายนถึงเดือนพฤษภาคม ระดับฝุ่นละอองขนาดเล็กมีปริมาณลดต่ำลง จำนวนผู้ป่วยก็มีแนวโน้มลดลง ซึ่งมีสัดส่วนในการเจ็บป่วยในกลุ่มคอหอยอักเสบเฉียบพลัน เจ็บคอมากที่สุด รองลงมาคือ ไข้หวัด เยื่อจมูกอักเสบเฉียบพลัน เยื่อตาอักเสบเฉียบพลัน ต้อเนื้อ และการติดเชื้อทางเดินอาหาร [3]

จังหวัดพะเยาเป็นจังหวัดที่มีฝุ่น $PM_{2.5}$ สูงติด 1 ใน 3 ของภาคเหนือ และสาเหตุหลักของฝุ่นในพะเยาเกิดจากการเผาชีวมวล

ในที่โล่ง [4] โดยข้อมูลของฝุ่นละออง $PM_{2.5}$ ระยะเวลาตั้งแต่ กรกฎาคม พ.ศ. 2562 ถึง สิงหาคม พ.ศ. 2564 จากกรมควบคุมมลพิษของจังหวัดพะเยา แสดงให้เห็นว่าจังหวัดพะเยามีค่า $PM_{2.5}$ เกินค่ามาตรฐาน ซึ่งในบางวันสูงเกือบ 400 ไมโครกรัม ($\mu g/m^3$) ดังรูปที่ 1 โดยค่า $PM_{2.5}$ ในบรรยากาศทั่วไป เฉลี่ย 24 ชั่วโมง ตามมาตรฐานของประเทศไทย อยู่ที่ $50 \mu g/m^3$ [5]

ดังนั้น ปัญหาฝุ่นละอองในพื้นที่จึงมีความจำเป็นต้องได้รับการแก้ไขอย่างทันด่วนที่ ซึ่งการนำข้อมูลด้านคุณภาพอากาศที่ตรวจวัดได้ มาทำนายค่าระดับ $PM_{2.5}$ ในอนาคต เพื่อป้องกันสถานการณ์คุณภาพอากาศ และเตรียมการล่วงหน้าได้มีประสิทธิภาพมากยิ่งขึ้น [6] จึงมีความจำเป็นต้องใช้คอมพิวเตอร์เป็นเครื่องมือช่วยในการหาคำตอบที่ต้องการในรูปแบบของเครื่องจักรการเรียนรู้ (machine learning) ที่เกิดจากการพัฒนาโปรแกรมทางคอมพิวเตอร์ ซึ่งสามารถเรียนรู้ได้จากชุดข้อมูลที่จัดเตรียมไว้ให้ [6] โดยการใช้การวิเคราะห์สมการถดถอยโลจิสติก (logistic regression) ซึ่งจะไม่มีการกำหนดที่เกี่ยวกับตัวแปรตาม (dependent variable) และความความคลาดเคลื่อนที่จะต้องมีการแจกแจงแบบปกติ (normal distribution) รวมไปถึงความสัมพันธ์เชิงเส้นตรง (linear function) ระหว่างตัวแปรอิสระและตัวแปรตาม นอกจากนี้ การถดถอยโลจิสติกยังสามารถใช้วิเคราะห์ข้อมูลได้ทุกระดับการวัดของตัวแปรอีกด้วย ซึ่งถือว่าเป็นจุดแข็งของเทคนิคทางสถิตินี้ [7]



รูปที่ 1 : ผลการตรวจวัดฝุ่นละออง $PM_{2.5}$ ของ จ.พะเยา

2) ทบทวนวรรณกรรม

2.1) ดัชนีคุณภาพอากาศ (Air Quality Index หรือ AQI)

เป็นดัชนีข้อมูลคุณภาพอากาศในรูปแบบที่ง่ายต่อความเข้าใจ เป็นเผยแพร่ให้ได้รับทราบถึงสถานการณ์มลพิษทางอากาศ ซึ่งดัชนีคุณภาพอากาศ 1 ค่า ใช้เป็นตัวแทนค่าความเข้มข้นของสารมลพิษทางอากาศ 6 ชนิด ได้แก่ ก๊าซโอโซน (O_3) ก๊าซคาร์บอนมอนอกไซด์ (CO) ก๊าซไนโตรเจนไดออกไซด์ (NO_2) ก๊าซซัลเฟอร์ไดออกไซด์ (SO_2) ฝุ่นละอองขนาดไม่เกิน 10 ไมครอน (PM_{10}) และฝุ่นละอองขนาดไม่เกิน 2.5 ไมครอน ($PM_{2.5}$)

โดยการคำนวณค่าความเข้มข้นของสารมลพิษทางอากาศ ที่ได้มาจากข้อมูลผลการตรวจวัดคุณภาพอากาศ โดยการคำนวณดัชนีคุณภาพอากาศภายในช่วงระดับเป็นสมการเส้นตรง ดังนี้

$$I = \frac{I_j - I_i}{x_j - x_i}(x - x_i) + I_i \quad (1)$$

โดย I คือ ค่าดัชนีย่อยคุณภาพอากาศ

x คือ ความเข้มข้นของมลพิษทางอากาศจากการตรวจวัด

x_i, x_j คือ ค่าต่ำสุด, ค่าสูงสุด ของช่วงความเข้มข้นมลพิษ

I_i, I_j คือ ค่าต่ำสุด, ค่าสูงสุด ของช่วงดัชนีคุณภาพอากาศ ที่ตรงกับช่วงความเข้มข้นของ x

2.2) การวิเคราะห์การถดถอยเชิงเส้น (Linear Regression)

เป็นกระบวนการทางสถิติ เพื่อให้ได้สมการถดถอย สำหรับทำนาย ปรากฏการณ์ต่าง ๆ โดยสิ่งที่ถูกทำนายก็จะถูกเรียกว่า ตัวแปรตาม ส่วนตัวแปรทำนายก็คือ ตัวแปรอิสระ ในกระบวนการทำนายนี้ ตัวแปรตามจะมีเพียงตัวเดียวเท่านั้น ส่วนตัวแปรอิสระอาจจะมีหลายตัวแปร โดยแบ่งออกเป็น

2.2.1) การวิเคราะห์การถดถอยอย่างง่าย (Simple Linear Regression) สมการถดถอยอย่างง่ายเป็นความสัมพันธ์ที่มีตัวแปรตาม และตัวแปรอิสระ อย่างละ 1 ตัว เท่านั้น โดยมีสมการถดถอยอย่างง่าย ดังนี้

$$z = a + b_0x_0 + e \quad (2)$$

2.2.2) การวิเคราะห์การถดถอยพหุคูณ (Multiple Linear Regression) สมการถดถอยพหุคูณ ความสัมพันธ์ที่มีตัวแปรตาม เพียง 1 ตัว และมีตัวแปรอิสระ ตั้งแต่ 2 ตัวขึ้นไป ซึ่งสมการถดถอยพหุคูณสามารถแสดงได้ในสมการ (3) ดังต่อไปนี้

$$z = a + b_1x_1 + b_2x_2 + \dots + b_nx_n + e \quad (3)$$

โดยที่ z คือ ค่าของตัวแปรตาม

e คือ ค่าความคลาดเคลื่อน (Error or Residual)

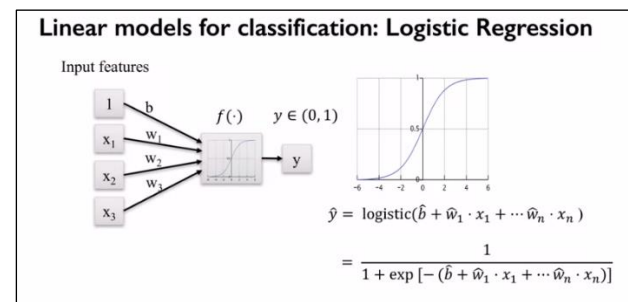
a คือ ค่าคงที่ (Constant) ของสมการถดถอย

$x_0, x_1, x_2, \dots, x_n$ คือ ค่าของตัวแปรอิสระ

$b_0, b_1, b_2, \dots, b_n$ คือ ค่าสัมประสิทธิ์ของสมการถดถอย

2.2.3) การวิเคราะห์การถดถอยโลจิสติก (Logistic Regression)

เป็นการวิเคราะห์ข้อมูลเพื่อทำนาย โอกาสการเกิดเหตุการณ์หนึ่ง ๆ หรือโอกาสเกิดเหตุการณ์มากหรือน้อย โดยมีตัวแปรตัวหนึ่ง หรือหลายตัวที่ส่งผลต่อการเกิดเหตุการณ์นั้น ๆ โดยตัวแปรอิสระ จะต้องเป็นตัวแปรข้อมูลในระดับอันตรภาค (interval scale) เป็นอย่างต่ำ กรณีที่เป็นข้อมูลเชิงกลุ่มให้แปลงเป็นตัวแปรหุ่น (dichotomous variable) โดยที่มีค่าเป็น 0 กับ 1 เท่านั้น ส่วนตัวแปรตามจะเป็นการถดถอยโลจิสติกแบบทวิ (Binary Logistic Regression) ซึ่งกำหนดเพียง 2 ค่าเท่านั้น โดยอาจเป็นคำว่า ใช่ (Yes) หรือ ไม่ใช่ (No) ซึ่งจะแทนด้วยค่าเป็น 1 กับ 0 ตามลำดับ ซึ่งสามารถแสดงความสัมพันธ์จากสมการ (3) ได้ดังรูป



รูปที่ 2 : ความสัมพันธ์ของการถดถอยเชิงเส้นพหุคูณ และการถดถอยโลจิสติก [8]

ดังนั้น การถดถอยโลจิสติกจึงสามารถเขียนเป็นสมการ (4) ได้ดังนี้

$$Prob(event) = \frac{1}{1 + \exp[-(a + b_1x_1 + \dots + b_nx_n)]} \quad (4)$$

2.3) สหสัมพันธ์แบบเพียร์สัน (Pearson Correlation)

เป็นการศึกษาความสัมพันธ์เชิงเส้น (linear relationship) ระหว่างตัวแปร 2 ตัวหรือข้อมูล 2 ชุด โดยมีข้อตกลงเบื้องต้น คือ ตัวแปรทั้งสองจะต้องเป็นตัวแปรข้อมูลระดับอันตรภาค (interval scale) ขึ้นไป และมีการแจกแจงปกติสองตัวแปร (bivariate normal distribution) โดยค่าสัมประสิทธิ์สหสัมพันธ์

แบบเพียร์สัน จะมีค่าอยู่ระหว่าง -1 ถึง +1 ถ้า เข้าใกล้ ± 1 หมายความว่า ตัวแปรทั้งสองมีความสัมพันธ์กันอยู่ในระดับมาก ถ้าสัมประสิทธิ์สหสัมพันธ์มีค่าเข้าใกล้ 0 หมายความว่า ตัวแปรทั้งสองมีความสัมพันธ์กันในระดับน้อยหรือไม่สัมพันธ์กันเลย เครื่องหมายบวกลบ ($\pm$) จะแสดงทิศทางของความสัมพันธ์ว่ามีทิศทางเป็นแบบใด โดยถ้าสัมประสิทธิ์สหสัมพันธ์มีเครื่องหมายบวก (+) หมายความว่าตัวแปรทั้งสองมีความสัมพันธ์ไปในทิศทางเดียวกัน หากมีเครื่องหมายลบ (-) หมายความว่าตัวแปร ทั้งสองมีความสัมพันธ์ไม่ไปในทิศทางเดียวกันหรือตรงกันข้าม สัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันสามารถคำนวณจากสมการ (5) ดังนี้ [9]

$$r_{xy} = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{[n(\Sigma x^2) - (\Sigma x)^2][n(\Sigma y^2) - (\Sigma y)^2]}} \quad (5)$$

โดยให้ r_{xy} คือ สัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน

Σx คือ ผลรวมข้อมูลที่วัดจากตัวแปร x

Σy คือ ผลรวมข้อมูลที่วัดจากตัวแปร y

Σxy คือ ผลรวมผลคูณของตัวแปร x และ ตัวแปร y

Σx^2 คือ ผลรวมกำลังสองข้อมูลที่วัดจากตัวแปร x

Σy^2 คือ ผลรวมกำลังสองข้อมูลที่วัดจากตัวแปร y

n คือ ขนาดตัวอย่าง

2.4) การแปลผลการวิเคราะห์ข้อมูลสหสัมพันธ์และการถดถอย

เป็นการแปลผลการวิเคราะห์ความสัมพันธ์ของข้อมูลสหสัมพันธ์และสมการถดถอย เมื่อพบว่ามีความสำคัญทางสถิติที่ .05 หรือ .01 ให้แปลว่า มีความสัมพันธ์กัน โดยทั่วไป การแปลความหมายของขนาดความสัมพันธ์จะพิจารณาจาก ค่าสัมประสิทธิ์สหสัมพันธ์ (ใช้สัญลักษณ์ r) ซึ่งมีค่าอยู่ระหว่าง -1 ถึง +1 ดังนี้ [10]

ถ้า ค่า r มีค่าเท่ากับ 0.00 แปลว่า ไม่มีความสัมพันธ์กัน

ค่า r มีค่าต่ำกว่า 0.40 แปลว่า มีความสัมพันธ์น้อย

ค่า r มีค่า 0.40 ถึง 0.60 แปลว่า มีความสัมพันธ์ปานกลาง

ค่า r มีค่า 0.60 ขึ้นไป แปลว่า มีความสัมพันธ์มาก

การแปลความหมายทิศทางของความสัมพันธ์ คือ หากค่าที่คำนวณได้เป็น บวก หมายถึง มีสัมพันธ์ในทิศทางเดียวกัน แต่ถ้าค่าที่คำนวณได้มีเครื่องหมายเป็น ลบ หมายถึง มีสัมพันธ์ในทิศทางตรงกันข้าม

สำหรับการแปลความหมายของสมการถดถอย จะพิจารณาจากค่าสัมประสิทธิ์การถดถอย ซึ่งใช้สัญลักษณ์ b (หรือ Beta) จากสมการ (2) และ (3) ซึ่งแปลความหมายเช่นเดียวกับค่าสหสัมพันธ์ แต่เปลี่ยนเป็นการพยากรณ์ หรือเป็นการทำนาย

ค่าตัวแปรตาม ส่วนการแปลผลการถดถอยพหุคูณ ซึ่งจะต้องแปลงเป็น ค่า R-Square หรือ R^2 [9] ซึ่งหมายถึง ความสามารถของตัวแปรอิสระในการทำนายตัวแปรตาม

3) ผลการจำลอง

การสร้างสมการรูปแบบการพยากรณ์ได้ถูกพัฒนาขึ้นบนโปรแกรมไพทอน (Python) โดยใช้คอมพิวเตอร์ที่มีหน่วยประมวลผลกลาง Intel® Core™ i5 Gen.11 2.40 GHz และมีหน่วยความจำขนาด 8 GB ซึ่งรวบรวมข้อมูลผ่านโปรแกรม Microsoft Excel

3.1) การเก็บรวบรวมข้อมูลทางอากาศจากสถานีตรวจวัดอากาศแบบอัตโนมัติบริเวณพื้นที่ทั่วไป

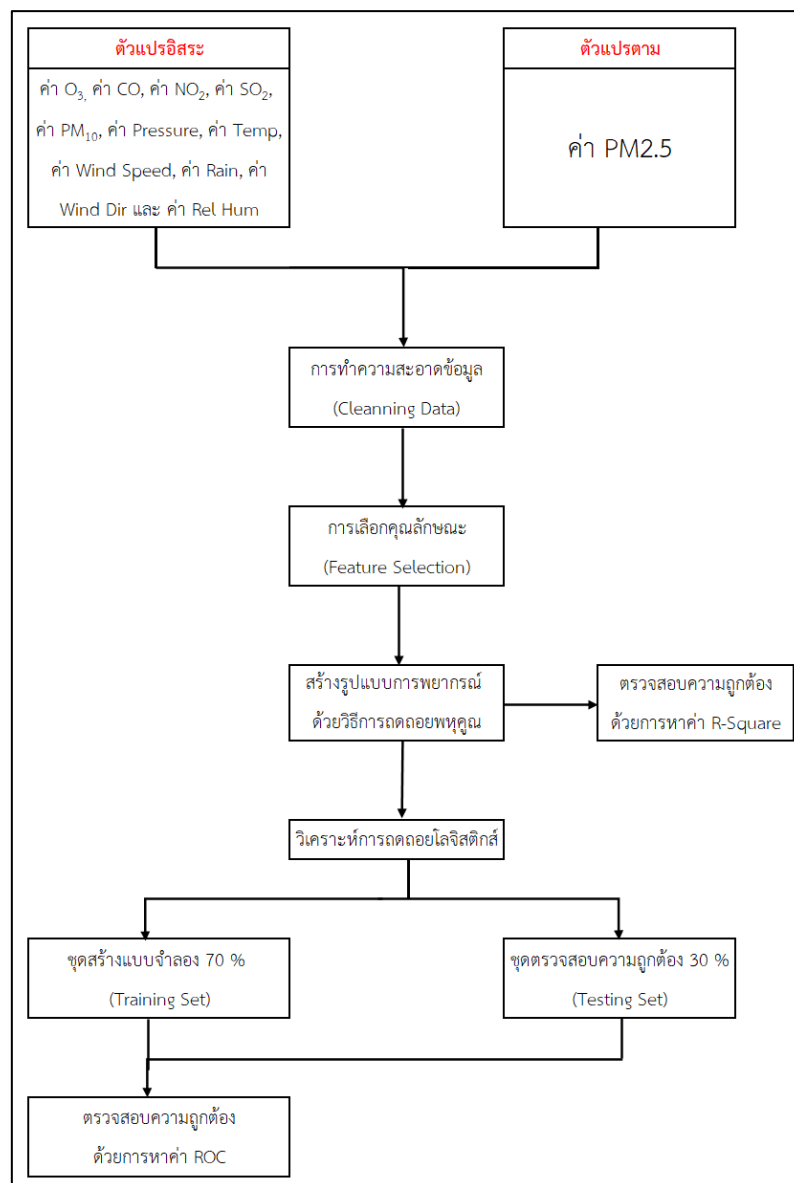
เป็นการข้อมูลย้อนหลังเป็นรายชั่วโมง ตั้งแต่วันที่ 1 มกราคม พ.ศ. 2558 จนถึง วันที่ 31 สิงหาคม พ.ศ. 2564 รวม 5 ปี จากสถานีตรวจวัดอากาศของจังหวัดพะเยา ทั้งสิ้น 12 ชุดข้อมูล โดยชุดข้อมูลที่ถูกรวบรวมคือ ฝุ่นละอองขนาดเล็กไม่เกิน 2.5 ไมครอน มีหน่วยเป็น $\mu\text{g}/\text{m}^3$ แทนด้วย “PM2.5” โดยกำหนดให้เป็นตัวแปรตาม ส่วนการกำหนดตัวแปรอิสระ คือ ชุดข้อมูลที่เหลือทั้ง 11 ชุดข้อมูล อันได้แก่ 1) ก๊าซโอโซน มีหน่วยเป็น ppm แทนด้วย “O3” 2) ก๊าซคาร์บอนมอนอกไซด์ มีหน่วยเป็น ppm แทนด้วย “CO” 3) ก๊าซไนโตรเจนไดออกไซด์ มีหน่วยเป็น ppb แทนด้วย “NO2” 4) ก๊าซซัลเฟอร์ไดออกไซด์ มีหน่วยเป็น ppb แทนด้วย “SO2” 5) ฝุ่นละอองขนาดเล็กไม่เกิน 10 ไมครอน มีหน่วยเป็น $\mu\text{g}/\text{m}^3$ แทนด้วย “PM10” 6) อุณหภูมิ มีหน่วยเป็น $^{\circ}\text{C}$ แทนด้วย “Temp” 7) ปริมาณน้ำฝน มีหน่วยเป็น mm แทนด้วย “Rain” 8) ความเร็วลม มีหน่วยเป็น km/h แทนด้วย “Wind Speed” 9) ทิศทางลม มีหน่วยเป็น องศา แทนด้วย “Wind Dir” 10) ความชื้นสัมพัทธ์ มีหน่วยเป็น % แทนด้วย “Rel Hum” และ 11) ความกดอากาศ มีหน่วยเป็น mb แทนด้วย “Pressure” โดยนำค่าตัวแปรต่าง ๆ มาสร้างแบบจำลองทางคณิตศาสตร์ ด้วยสมการ (3) เป็นสมการรูปแบบการพยากรณ์ โดยนำข้อมูลจากโปรแกรม Microsoft Excel ไปทำการแปลงด้วยโปรแกรมไพทอน (Python) ซึ่งสามารถแสดงผล ดังรูปที่ 3

3.2) การสร้างสมการรูปแบบการพยากรณ์

เป็นการนำชุดข้อมูลที่แปลได้จากโปรแกรมไพทอน มาสร้างแบบจำลองการพยากรณ์ (Prediction Model) โดยมีขั้นตอนและกระบวนการ ดังรูปที่ 4

	Year_Month_Day	Hour	CO	NO2	SO2	O3	PM10	Wind speed	Wind dir	Temp	Rel hum	Pressure	Rain	PM2.5
0	150101	100	1.50	12.6	1.0	11.0	36.0	0.1	106.0	20.9	81.4	725.9	0.0	NaN
1	150101	200	1.40	12.7	1.0	11.0	50.0	0.2	68.0	20.1	81.5	725.8	0.0	NaN
2	150101	300	0.60	10.3	1.0	11.0	46.0	0.1	79.0	19.1	86.0	725.7	0.0	NaN
3	150101	400	0.50	9.1	1.0	9.0	42.0	0.2	360.0	18.2	92.9	725.7	0.0	NaN
4	150101	500	0.50	7.2	0.0	9.0	41.0	0.2	314.0	18.0	97.6	725.9	0.0	NaN
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
56251	210831	2000	0.49	5.0	0.0	8.0	16.0	0.5	136.0	25.3	84.0	721.0	0.0	11.0
56252	210831	2100	0.57	4.0	0.0	4.0	20.0	0.2	90.0	24.6	85.0	721.0	0.0	10.0
56253	210831	2200	0.52	4.0	0.0	4.0	19.0	1.2	65.0	24.5	86.0	721.0	0.0	7.0
56254	210831	2300	0.40	2.0	0.0	5.0	17.0	1.3	67.0	24.4	86.0	722.0	0.0	9.0
56255	210831	2400	0.44	2.0	0.0	3.0	18.0	0.8	69.0	24.2	87.0	722.0	0.0	9.0

รูปที่ 3 : การแปลงข้อมูลที่ได้จากการเก็บรวบรวมด้วยโปรแกรมไพทอน (Python)



รูปที่ 4 : ขั้นตอนและกระบวนการสร้างแบบจำลองการพยากรณ์

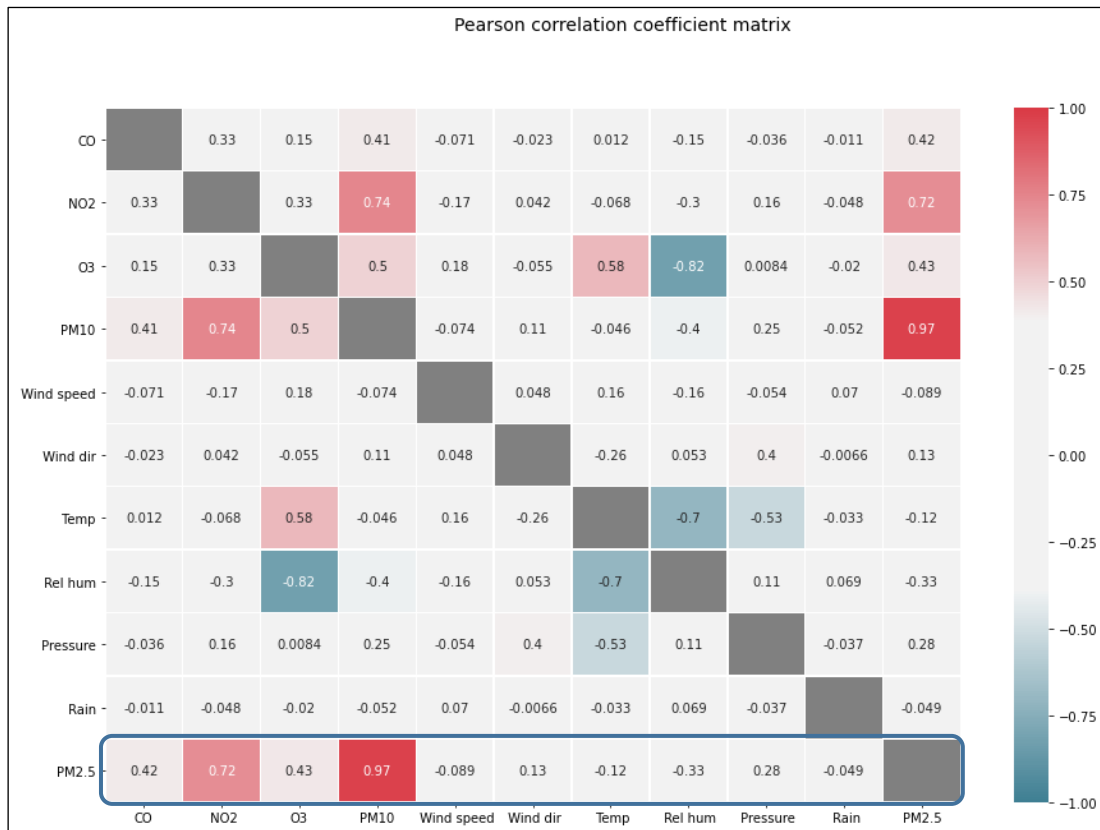
3.2.1) การทำความสะอาดข้อมูล (Cleaning Data) เป็นขั้นตอนกระบวนการตรวจสอบ และแก้ไข (หรือลบ) รายการชุดข้อมูลที่ไม่ถูกต้องหรือไม่สมบูรณ์ ออกไปจากชุดข้อมูลในการกำหนดตัวแปรอิสระเพื่อนำมาสร้างแบบจำลองการพยากรณ์ เนื่องจาก ชุดข้อมูลอาจจะมีลักษณะของข้อมูลที่ไม่สมบูรณ์ (incomplete) มีข้อมูลรบกวน (noise) หรือมีข้อมูลที่ไม่สอดคล้อง (inconsistent) จึงต้องมีการประมวลผลชุดข้อมูลก่อน (pre-processing) ซึ่งใช้วิธีการทำความสะอาด แต่เนื่องจากข้อมูลไม่ครบถ้วนสมบูรณ์ ข้อมูลที่ขาดหายไปอย่างต่อเนื่องติดต่อกันนานจะถูกตัดทิ้ง และทดแทนค่าสูญหายกรณีอื่นด้วยค่าเฉลี่ย (Mean) ซึ่งพบว่าข้อมูลสถานีตรวจวัดอากาศสนามกีฬา จ.พะเยา ไม่มีการตรวจจับค่าก๊าซซัลเฟอร์ไดออกไซด์ (SO₂) ในช่วงวันที่ 5 กรกฎาคม พ.ศ. 2562 ถึง วันที่ 31 สิงหาคม พ.ศ. 2564 มีค่าที่ขาดหายไปกว่า 25% จึงมีความจำเป็นต้องนำชุดข้อมูลนี้ ออกจากการสร้างแบบจำลองการพยากรณ์ [11] ดังรูปที่ 5

3.2.2) การเลือกคุณลักษณะ (Feature Selection) เป็นวิธีการลดมิติ โดยจะเลือกคุณลักษณะ หรือข้อมูลที่เหมาะสมเนื่องจาก ข้อมูลที่นำมาสร้างรูปแบบการพยากรณ์มีจำนวนมาก

และมีลักษณะข้อมูลที่กระจัดกระจาย และบางข้อมูลไม่มีประโยชน์ในการวิเคราะห์ ซึ่งจะส่งผลกระทบต่อความถูกต้องของการประมวลผล จึงต้องมีกระบวนการเลือกคุณลักษณะ ซึ่งในกรณีนี้ใช้วิธีการฟิเตอร์ (filter) ข้อดีของวิธีเลือกคุณลักษณะแบบฟิเตอร์นี้ คือ เป็นกระบวนการประเมินประสิทธิภาพของคุณลักษณะของข้อมูลแต่ละตัวว่ามีความเหมาะสมกับการวิเคราะห์ข้อมูลมากน้อยเพียงใด โดยไม่ขึ้นกับขั้นตอนวิธีของการเรียนรู้แบบใดแบบหนึ่ง การเลือกคุณลักษณะแบบนี้ จะทำการจัดลำดับ (Ranking) ตามความสำคัญของคุณลักษณะแต่ละตัว และเลือกคุณลักษณะที่มีระดับความสำคัญสูงสุดตามจำนวนที่ผู้ใช้ระบุ ซึ่งเป็นเทคนิคที่คำนวณได้ง่าย รวดเร็ว และหลีกเลี่ยงการเกิด Overfitting [12] ตัวแปรที่ได้ทั้งหมดที่นำมาทำการวิเคราะห์สหสัมพันธ์แบบเพียร์สัน โดยจะคัดเลือกรวมตัวแปรที่มีความสัมพันธ์กันมากกว่า 0.40 ซึ่งได้รับการทดสอบนัยสำคัญของสหสัมพันธ์ มีแนวโน้มควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ภายใต้ระดับนัยสำคัญที่ .05 [13] ซึ่งขนาดของตัวแปรที่มากขึ้นส่งผลให้การวิจัยและการวิเคราะห์ข้อมูลมีความคลาดเคลื่อนลดลงด้วย [14] และได้ตารางเมทริกซ์ค่าสหสัมพันธ์ ดังรูปที่ 6

	CO	NO2	O3	PM10	Wind speed	Wind dir	Temp	Rel hum	Pressure	Rain	PM2.5
Date											
2019-07-05	0.37	5.0	20.0	23.0	1.3	134.0	26.1	82.0	717.0	0.0	9.0
2019-07-05	0.35	5.0	20.0	20.0	1.4	136.0	25.6	81.0	717.0	0.0	8.0
2019-07-05	0.38	5.0	19.0	10.0	1.4	136.0	25.2	81.0	717.0	0.0	7.0
2019-07-05	0.41	6.0	18.0	9.0	1.5	136.0	25.0	81.0	717.0	0.0	6.0
2019-07-05	0.43	7.0	15.0	10.0	1.4	135.0	24.8	82.0	717.0	0.0	7.0
...	...	...	...	...	...	...	...	...	...	...	...
2021-08-31	0.49	5.0	8.0	16.0	0.5	136.0	25.3	84.0	721.0	0.0	11.0
2021-08-31	0.57	4.0	4.0	20.0	0.2	90.0	24.6	85.0	721.0	0.0	10.0
2021-08-31	0.52	4.0	4.0	19.0	1.2	65.0	24.5	86.0	721.0	0.0	7.0
2021-08-31	0.40	2.0	5.0	17.0	1.3	67.0	24.4	86.0	722.0	0.0	9.0
2021-08-31	0.44	2.0	3.0	18.0	0.8	69.0	24.2	87.0	722.0	0.0	9.0

รูปที่ 5 : การแปลงผลหลังจากการทำความสะอาดข้อมูล



รูปที่ 6 : ตารางเมทริกซ์สัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน

3.2.3) การสร้างสมการพยากรณ์ด้วยการวิเคราะห์การถดถอยเชิงเส้นพหุคูณ โดยการคัดเลือกตัวแปรที่มีความสัมพันธ์กับค่า $PM_{2.5}$ ที่มีความเหมาะสมที่ได้จากการเลือกคุณลักษณะ เพื่อกำหนดเป็นตัวแปรอิสระ โดยในการวิเคราะห์เพื่อสร้างสมการจำลองทางคณิตศาสตร์ สำหรับการพยากรณ์ค่า $PM_{2.5}$ อันได้แก่ค่า PM_{10} , NO_2 , O_3 และ CO ตามรูปแบบสมการ (3) ได้เป็นสมการพยากรณ์ของแบบจำลอง (6) ดังนี้

$$Z = -3.412 + 1.1819X_1 - 0.1392X_2 - 0.1149X_3 + 0.8260X_4 \quad (6)$$

โดยที่ Z คือ ฝุ่นละอองขนาดเล็กไม่เกิน 2.5 ไมครอน ($PM_{2.5}$)

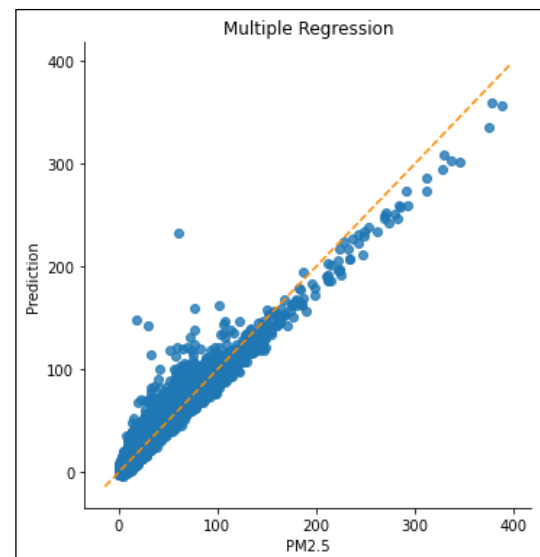
X_1 คือ ก๊าซคาร์บอนมอนอกไซด์ (CO)

X_2 คือ ก๊าซไนโตรเจนไดออกไซด์ (NO_2)

X_3 คือ ก๊าซโอโซน (O_3)

X_4 คือ ฝุ่นละอองขนาดเล็กไม่เกิน 10 ไมครอน (PM_{10})

โดยเราสามารถนำสมการถดถอยพหุคูณที่ได้มาสร้างกราฟความสัมพันธ์เชิงเส้น เพื่อวิเคราะห์ความสัมพันธ์ระหว่าง ค่าฝุ่นละอองขนาดเล็กไม่เกิน 2.5 ไมครอน ($PM_{2.5}$) กับ ค่าการพยากรณ์หรือตัวแปรอิสระ ดังรูปที่ 7



รูปที่ 7 : กราฟความสัมพันธ์เชิงเส้นพหุคูณ

เมื่อนำค่าจากสมการ (6) แปลผลด้วยวิธีการหาค่า R-Square หรือ R^2 มีค่าเท่ากับ 0.9557 ซึ่งหมายความว่า ตัวแปรอิสระทั้ง 4 ชุดข้อมูล อันได้แก่ ค่า PM_{10} , NO_2 , O_3 และ CO ร่วมกันทำนายค่าตัวแปรตาม หรือค่าฝุ่นละอองขนาดเล็กไม่เกิน 2.5 ไมครอน ($PM_{2.5}$) ได้ร้อยละ 95.57

3.2.4) การวิเคราะห์การถดถอยโลจิสติก (Logistic Regression Analysis) เพื่อสร้างรูปแบบการเรียนรู้การจำลองการถดถอยเพื่อการพยากรณ์ โดยการแบ่งข้อมูลออกเป็น 2 ชุดข้อมูล คือ ชุดการสร้างแบบจำลอง (Training Set) โดยวิธีการวิเคราะห์การถดถอยโลจิสติก ซึ่งจะใช้ข้อมูล 70% จากชุดข้อมูลทั้งหมด ในการสร้างแบบจำลองพยากรณ์ จากนั้นนำรูปแบบสมการพยากรณ์ที่สร้างได้มาทดสอบด้วยการทำนายในชุดข้อมูลที่เหลืออีก 30% ซึ่งจะใช้สำหรับการตรวจสอบความถูกต้อง (Testing Set) และการวิเคราะห์จำแนกประเภท (Discriminant Analysis) คือ การศึกษาความสัมพันธ์ระหว่างตัวแปรตาม และตัวแปรอิสระ โดยกำหนดค่า $PM_{2.5}$ เกินกว่า $50 \mu g/m^3$ ให้เป็นค่าเกินมาตรฐาน หรือ จริง (Yes) หากต่ำกว่า $50 \mu g/m^3$ ถือว่าเป็นค่าที่ไม่เกินมาตรฐาน หรือ ไม่จริง (No) ซึ่งจะแทนด้วยค่าเป็น 1 กับ 0 ตามลำดับ แล้วทำการสร้างแบบจำลองการพยากรณ์ด้วยสมการถดถอยโลจิสติก (4) ได้ดังนี้

$$Prob(Y) = \frac{1}{1 + e^{-(8.6581 + 0.4365X_1 - 0.706X_2 - 0.0212X_3 + 0.0598X_4)}} \quad (7)$$

โดยที่ $Prob(Y)$ คือ ฝุ่นละอองขนาดไม่เกิน 2.5 ไมครอน ($PM_{2.5}$) เกินมาตรฐาน หรือไม่เกินมาตรฐานที่ 50 ไมโครกรัม

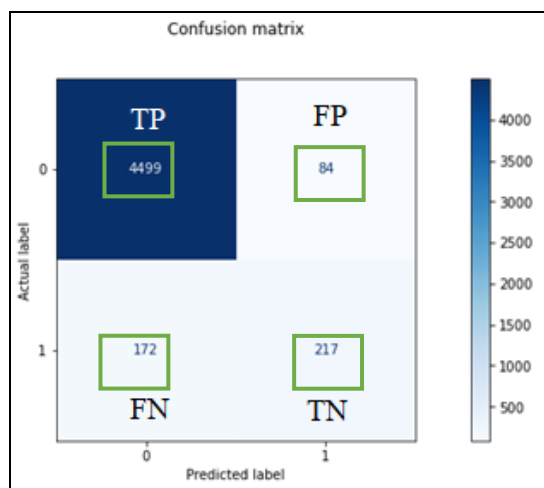
X_1 คือ ก๊าซคาร์บอนมอนอกไซด์ (CO)

X_2 คือ ก๊าซไนโตรเจนไดออกไซด์ (NO_2)

X_3 คือ ก๊าซโอโซน (O_3)

X_4 คือ ฝุ่นละอองขนาดไม่เกิน 10 ไมครอน (PM_{10})

3.2.5) การตรวจสอบความถูกต้อง (Accuracy) โดยสร้างตารางสำคัญในการวัดความสามารถ (Confusion Matrix) จากสมการ (7) ได้ดังรูปที่ 8



รูปที่ 8 : ตารางสำคัญในการวัดความสามารถ (Confusion Matrix)

จากนั้น นำค่าที่ได้จากตารางสำคัญในการวัดความสามารถ โดยหาความถูกต้อง (Accuracy) ได้ตามสมการ (8)

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (8)$$

โดยที่ TP คือ ทำนายว่า “จริง” ตรงกับสิ่งที่เกิดขึ้น “จริง”

TN คือ ทำนายว่า “ไม่จริง” ตรงกับสิ่งที่เกิดขึ้น “ไม่จริง”

FP คือ ทำนายว่า “จริง” แต่สิ่งที่เกิดขึ้น “ไม่จริง”

FN คือ ทำนายว่า “ไม่จริง” แต่สิ่งที่เกิดขึ้น “จริง”

ค่าความไว (Sensitivity) คือ สัดส่วนของผลบวกที่เป็นจริง สำหรับการทดลองนั้น ๆ เช่น การเกิดมลพิษอากาศที่เป็นจริง (true positives) จึงช่วยในการแยกการทดลองที่เป็นผลลบปลอม (false negative) เพราะว่าการทดสอบยิ่งไวเท่าไร โอกาสการได้ผลลบ (เช่นพบว่า ไม่เกิดมลพิษอากาศ) ที่ไม่เป็นจริง (เช่น ผลที่เกิดขึ้นคือ เกิดมลพิษอากาศ) ก็น้อยลงเท่านั้น สามารถคำนวณได้จากสมการ (9) ดังนี้

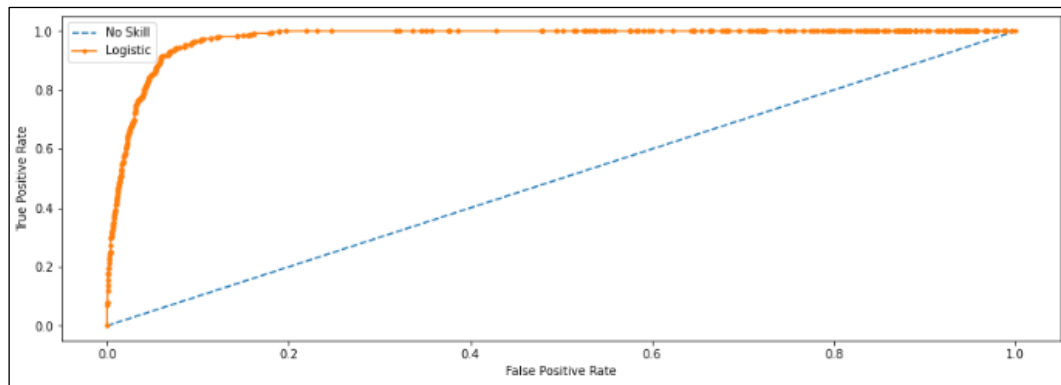
$$Sensitivity = \frac{TP}{TP+FN} \quad (9)$$

ค่าความจำเพาะ (Specificity) คือ สัดส่วนของผลลบที่เป็นจริง สำหรับการทดลองนั้น ๆ เช่น การไม่เกิดมลพิษอากาศ ซึ่งจะช่วยในการยืนยันสถานะที่มีผลบวกปลอม (false positive) เพราะว่าการทดสอบยิ่งจำเพาะเท่าไร โอกาสที่ได้ผลบวก (เช่นพบว่า เกิดมลพิษอากาศ) ที่ไม่เป็นจริง (ผลที่เกิดขึ้นคือ ไม่เกิดมลพิษอากาศ) ก็น้อยลงเท่านั้น สามารถคำนวณได้จากสมการ (10) ดังนี้

$$Specificity = \frac{TN}{TN+FP} \quad (10)$$

เมื่อคำนวณค่าความไว (Sensitivity) และค่าความจำเพาะ (Specificity) จากสมการ (9) และ (10) ก็จะได้ผลการคำนวณอยู่ที่ 0.9632 และ 0.7209 ตามลำดับ หรือคิดเป็นร้อยละ 96.32 และ 72.09 โดยมีค่าความถูกต้อง (Accuracy) ซึ่งคำนวณได้จากสมการ (8) อยู่ที่ 0.9485 หรือร้อยละ 94.85

3.2.6) การทดสอบประสิทธิภาพของแบบจำลองการพยากรณ์ ด้วยค่า Receiver Operating Characteristic Curve (ROC Curve) และค่าพื้นที่ใต้เส้นโค้ง หรือ Area under the curve (AUC) เป็นการหาสมการสหสัมพันธ์ระหว่างค่าจริงที่เกิดจากการวัด และค่าที่ได้จากการทำนายที่ระดับความเชื่อมั่นร้อยละ 95 เพื่อให้การทดสอบประสิทธิภาพการพยากรณ์ที่จะเกิดมลพิษทางอากาศหรือไม่เกิดมลพิษอากาศ ซึ่งเป็นวิธีการที่สามารถนำมาใช้ในการเลือกจุดตัดที่เหมาะสมได้ คือ การสร้างความสัมพันธ์ระหว่าง true positive (Sensitivity) กับ false positive rate (1-Specificity) ด้วยการแปรค่าจุดตัด (Cut-off Point)



รูปที่ 9 : กราฟ Receiver Operating Characteristic

โดยการเลือก Optimal cut-off point หรือจุดที่อยู่บนกราฟที่จะได้ค่า AUC มากที่สุด (ให้ผลดีที่สุด) คือ จุดที่อยู่ใกล้มุมซ้ายบนมากที่สุด ซึ่งจะเป็นจุดที่ Sensitivity สูง และ Specificity สูงด้วยเช่นกัน ดังนั้น หากค่า AUC ยิ่งเข้าใกล้ 1 ก็แสดงว่า แบบจำลองการพยากรณ์จะมีประสิทธิภาพมาก

เมื่อทำการทดสอบความถูกต้องจากค่าความไว (sensitivity) ที่ได้ คือ 0.9632 และค่าความจำเพาะ (specificity) ที่ได้ คือ 0.7209 มาคำนวณลงบนค่าพื้นที่ใต้เส้นโค้ง หรือค่า AUC ก็จะได้เท่ากับ 0.9757 หรือ 0.98 ดังรูปที่ 9

4) อภิปรายผลการวิจัย

การศึกษานี้เป็นการวิเคราะห์สหสัมพันธ์เพื่อคัดเลือกตัวแปรที่เหมาะสมในการสร้างสมการแบบจำลองการพยากรณ์ โดยใช้ข้อมูลสภาพอากาศย้อนหลัง 5 ปี พบว่า ตัวแปรที่มีสหสัมพันธ์กับความเข้มข้นของ $PM_{2.5}$ โดยเรียงจากตัวแปรที่มีความสัมพันธ์มากไปหาน้อย จากค่าสัมประสิทธิ์ของสมการถดถอย คือ ก๊าซคาร์บอนมอนอกไซด์ (CO), ฝุ่นละอองขนาดเล็กไม่เกิน 10 ไมครอน (PM_{10}), ก๊าซไนโตรเจนไดออกไซด์ (NO_2) และ ก๊าซโอโซน (O_3) ตามลำดับ ซึ่งพบว่า มีค่าความไว (sensitivity) อยู่ในระดับสูงที่ 96.32% แต่ค่าความจำเพาะ (specificity) มีระดับเพียง 72.09% เนื่องจาก ปัจจัยของก๊าซซัลเฟอร์ไดออกไซด์ (SO_2) ถูกทำความสะอาด (cleaning) ออกไป ซึ่งค่าความเข้มข้นของ SO_2 เกิดมาจากการเผาไหม้เชื้อเพลิงฟอสซิลที่มีกำมะถันสูง เช่น น้ำมันเตาและถ่านหิน โดยจะค่าสูงขึ้นในเขตอุตสาหกรรม และโรงผลิตไฟฟ้า [6] สอดคล้องกับผลการวิจัยของ ศรีรักษ์ ศรีทองชัย [14] ซึ่งพบว่า ตัวแปร PM_{10} , NO_2 , CO, O_3 , และ SO_2 เป็นแบบจำลองที่ให้ความแม่นยำสูงสุดในการพยากรณ์ $PM_{2.5}$ และผลการศึกษาของ Zhai *et al.* [15] ก็ยังพบว่า ค่า SO_2 , CO และ NO_2 มีความ

สัมพันธ์กับค่า $PM_{2.5}$ นอกจากนี้ ผลการศึกษาของ Lachatre *et al.* [16] ยังพบอีกว่า ค่า $PM_{2.5}$ มีความสัมพันธ์กับค่า SO_2 , NO_x และ NH_3 อีกด้วย

5) สรุปผล

แบบจำลองการพยากรณ์นี้ มีประสิทธิภาพของแบบจำลองการพยากรณ์ด้วยค่า Receiver Operating Characteristic Curve (ROC Curve) และค่าพื้นที่ใต้เส้นโค้ง หรือ Area Under the Curve (AUC) อยู่ที่ 0.9757 หรือ 0.98 ซึ่งมีค่าอยู่ในระดับสูงแสดงว่า แบบจำลองสามารถพยากรณ์ได้ค่อนข้างแม่นยำ แต่ค่าความจำเพาะ (Specificity) ไม่ดีเท่าที่ควร

หากมีชุดข้อมูลตัวแปรอิสระของก๊าซซัลเฟอร์ไดออกไซด์ (SO_2) ซึ่งถูกขจัดออกไปในขั้นตอนการทำความสะอาด (cleaning) โดยสามารถนำชุดข้อมูลนี้ มาสร้างแบบจำลองการพยากรณ์ ก็จะทำให้การพยากรณ์มีประสิทธิภาพที่สูงขึ้น และมีความแม่นยำเพิ่มมากขึ้น

6) ข้อเสนอแนะ

ปัญหาการวิจัยแบบจำลองการพยากรณ์นี้ พบว่า มีข้อมูลชุดขาดหายเป็นจำนวนมาก จึงมีข้อเสนอแนะให้ ขั้นตอนการเก็บข้อมูลต่าง ๆ ควรมีค่าที่ครบถ้วน ไม่ตกหล่น หรือสูญหาย ซึ่งจะทำให้ชุดข้อมูลถูกคัดกรองออกน้อย โดยเฉพาะชุดข้อมูลตัวแปรอิสระของก๊าซซัลเฟอร์ไดออกไซด์ (SO_2) หรือการเก็บข้อมูลที่มีปัจจัยความสัมพันธ์เพิ่มเติม อาทิเช่น ค่าแอมโมเนีย (NH_3) ซึ่งจะทำให้สมการแบบจำลองการพยากรณ์จะมีความถูกต้องมากขึ้น และแม่นยำเพิ่มขึ้น ซึ่งจะช่วยพัฒนาอัลกอริทึมให้ฉลาดเพิ่มมากขึ้นอีกด้วย

REFERENCES

- [1] S. Pansripong, "Development of artificial neural network model and multiple regression models for prediction of PM10 concentrations in Bangkok and Nakhon Ratchasima municipality area," (in Thai), M.S. thesis, Dept. Environ. Eng., Suranaree Univ. Technol., Nakhon Ratchasima, Thailand, 2007.
- [2] W. Simachaya. "Smog Pollution in the Northern: Problems and Solutions." (in Thai), PCD.go.th. https://www.pcd.go.th/wp-content/uploads/2020/06/pcdnew-2020-06-05_07-33-31_040770.pdf (accessed Feb. 1, 2023).
- [3] S. Pimonsree and P. Arrin, "Situation of Particulate Matter in Ambient Air and Impacts on Healthy," (in Thai), in *Proc. Naresuan Res. Conf.: Sustain. Ways of Living Based on Sufficiency Economy*, Phitsanulok, Thailand, Jul. 2010, pp. 670–681.
- [4] T. Rattanapong, "University of Phayao joins forces with RISC by MQDC to move forward with Fasai 2 to Phayao, expanding urban air purification towers for people in the region to further research towards understanding PM 2.5 dust in the northern region." (in Thai), UP.ac.th. <https://www.up.ac.th/th/NewsRead.aspx?itemID=23918> (accessed Jul. 2, 2021).
- [5] Greenpeace Thailand. "The different air quality measurement standards." (in Thai), GREENPEACE.org. <https://www.greenpeace.org/thailandexplore/protect/cleanair/air-standard/> (accessed Nov. 23, 2022).
- [6] K. Meksagul and A. Phantong, "The prediction model of climate in Phayao," (in Thai), in *Proc. Operations Res. Netw. (OR-net) Conf.*, Chiang Mai, Thailand, Feb. 2019, pp. 30–40.
- [7] U. Amornnimit, "Logistic regression: Choice of risk analysis," (in Thai), *Univ. Thai Chamber of Commerce J.*, vol. 23, no. 2, pp. 21–35, 2003.
- [8] V. Kovenko and V. Shevchuk. "Machine Learning with Python: Differences between classification and regression." READ THEDOCS.io. https://machine-learning-and-data-science-with-python.readthedocs.io/en/latest/assignment1_sup_ml_cls.html (accessed July. 2, 2021).
- [9] K. Pearson, "Notes on the history of correlation," *Biometrika*, vol. 13, no. 1, pp. 25–45, 1920.
- [10] B. Pommapun, "Techniques for data interpreting for using correlation and regression in research," (in Thai), *STOU Educ. J.*, vol. 11, no. 1, pp. 32–45, 2018.
- [11] S. Srithongchai, "A PM2.5 prediction model using LSTM neural network in Bangkok area," (in Thai), *J. Eng. Digit. Technol. (JEDT)*, vol. 10, no. 1, pp. 1–9, 2022.
- [12] R. Kladchuen and C. Sanrach, "An efficiency comparison of algorithms and feature selection methods to predict the learning achievement of vocational students," (in Thai), *Res. J. Rajamangala Univ. Technol. Thanyaburi*, vol. 17, no. 1, pp. 1–10, 2018.
- [13] C. Pruekpramool, N. Jaroentaku, and S. Srisuttiyakorn, "Efficiency of Pearson, Spearman and Kendall's correlation coefficients when data is non-normal distributed," (in Thai), *OJED*, vol. 15, no. 2, 2020, Art. no. OJED1502040.
- [14] B. Srisa-ard, *Basic Research*, 10th ed. Bangkok, Thailand: SUWEERIVASARN (in Thai), 2017.
- [15] S. Zhai *et al.*, "Fine particulate matter (PM_{2.5}) trends in China, 2013–2018: separating contributions from anthropogenic emissions and meteorology," *Atmos. Chem. Phys.*, vol. 19, no. 16, pp. 11031–11041, Aug. 2019, doi: 10.5194/acp-19-11031-2019.
- [16] M. Lachatre *et al.*, "The unintended consequence of SO₂ and NO₂ regulations over China: Increase of ammonia levels and impact on PM_{2.5} concentrations," *Atmos. Chem. Phys.*, vol. 19, no. 10, pp. 6701–6716, May 2019, doi: 10.5194/acp-19-6701-2019.