การวิเคราะห์แนวโน้มการปล่อยก๊าซคาร์บอนไดออกไซด์ (CO_2) ในภาคขนส่งโดยใช้อัลกอริทึมเหมืองข้อมูล

Trend Analysis of CO_2 Emissions in the Transport Sector Using Data Mining Algorithms

ปรีชาเกียรติ กอนแก้ว[1*] วันชัย ใจศร[1] รุจิพันธุ์ โกษารัตน์[1] และธนิต เกตุแก้ว[1]

Preechakiat Konkaew[1*] Wanchai Jaisorn[1] Rujipan Kosarat[1] and Thanit Keatkaew[1]

Received: *March 13, 2025;* ***Revised:*** *April 16, 2025;* ***Accepted:*** *April 17, 2025*

บทคัดย่อ

การวิเคราะห์แนวโน้มการปล่อยก๊าซคาร์บอนไดออกไซด์ในภาคขนส่งเป็นประเด็นสำคัญที่ส่งผลต่อสิ่งแวดล้อมและสุขภาพมนุษย์ งานวิจัยนี้เน้นการประยุกต์ใช้เทคนิคเหมืองข้อมูล ได้แก่ การแบ่งกลุ่มข้อมูล การวิเคราะห์ถดถอยเชิงเส้น และการสุ่มป่าไม้ เพื่อจัดกลุ่มและทำนายแนวโน้มการปล่อย CO_2 โดยใช้ข้อมูลจากการใช้เชื้อเพลิงน้ำมันในอดีต พร้อมประเมินประสิทธิภาพของแต่ละเทคนิคด้วยตัวชี้วัด เช่น ค่าความคลาดเคลื่อนเฉลี่ยเชิงสัมบูรณ์ ค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย และค่าสัมประสิทธิ์การกำหนด จากผลการวิจัยพบว่า เทคนิคการสุ่มป่าไม้มีความแม่นยำสูงสุด โดยให้ค่าสัมประสิทธิ์การกำหนดถึงร้อยละ 97.14 ในขณะที่เทคนิคถดถอยเชิงเส้นและการแบ่งกลุ่มข้อมูลมีค่าอยู่ที่ร้อยละ 86.58 และ 51.14 แสดงให้เห็นถึงศักยภาพของการสุ่มป่าไม้ในการพยากรณ์การปล่อยก๊าซในอนาคตเพื่อใช้ในการวางแผนลดก๊าซเรือนกระจก

คำสำคัญ : การปล่อยก๊าซคาร์บอนไดออกไซด์; การเหมืองข้อมูล; เทคนิคการแบ่งกลุ่มข้อมูล; การวิเคราะห์ข้อมูลถดถอยเชิงเส้น; เทคนิคการสุ่มป่าไม้

[1] *คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา*

[1] *Faculty of Engineering, Rajamangala University of Technology Lanna*

[*] *Corresponding Author, Tel. 09 6501 2964, E - mail: preechakiat_ko66@live.rmutl.ac.th*

Abstract

Analyzing trends in carbon dioxide (CO_2) emissions in the transportation sector is a critical issue that impacts both the environment and human health. This research focuses on the application of data mining techniques—namely clustering, linear regression analysis, and random forest—to group and predict future CO_2 emission trends. Historical fuel consumption data was used as the primary input for analysis. The performance of each technique was evaluated using metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and the coefficient of determination (R^2). The results revealed that the random forest technique provided the highest accuracy, achieving an R^2 of 97.14 %. In comparison, linear regression and clustering yielded R^2 values of 86.58 % and 51.14 %. These findings highlight the potential of the random forest algorithm as an effective tool for forecasting carbon emissions to support greenhouse gas reduction planning efforts.

Keywords: CO_2 Emissions; Data Mining; K-means; Linear Regression; Random Forest

บทนำ

การเปลี่ยนแปลงสภาพภูมิอากาศและปัญหาสิ่งแวดล้อมจากการปล่อยก๊าซคาร์บอนไดออกไซด์ (Aekrattanawat et al., 2024; Palason et al., 2024) ส่งผลกระทบต่อสิ่งแวดล้อม เศรษฐกิจ และสุขภาพของประชากร โดยเฉพาะในภูมิภาคที่มีการเติบโตทางเศรษฐกิจและอุตสาหกรรมอย่างรวดเร็ว ก๊าซคาร์บอนไดออกไซด์ซึ่งมาจากการเผาไหม้เชื้อเพลิงฟอสซิลในหลายภาคส่วน โดยเฉพาะภาคขนส่ง มีบทบาทสำคัญต่อภาวะเรือนกระจกและภาวะโลกร้อน การปล่อยก๊าซจากภาคขนส่งยังส่งผลให้เกิดปัญหาฝุ่น PM2.5 (Kanabkaew, 2024) ซึ่งเป็นฝุ่นละอองขนาดเล็กที่กระทบต่อระบบทางเดินหายใจ หัวใจ และหลอดเลือด รวมถึงระบบอื่น ๆ ในร่างกาย (Namcome and Tansuchat, 2021; Wongchan, 2023; Maijaroensri and Hirunrueng, 2024) โดยเฉพาะในพื้นที่ที่มีการจราจรหนาแน่นหรือกิจกรรมอุตสาหกรรมสูง ปัจจุบันมีการใช้เทคนิคเหมืองข้อมูล (Data Mining) (Khruachalee et al., 2024) ในการวิเคราะห์ข้อมูลขนาดใหญ่เพื่อค้นหารูปแบบและแนวโน้มที่ซ่อนอยู่ การศึกษานี้จึงประยุกต์ใช้ K-means Clustering (Khatbanjong and Mawan, 2021; Posing et al., 2024; Phounsataphorn and Bamroongcheep, 2025) ในการจัดกลุ่มข้อมูลการปล่อยก๊าซตามประเภทเชื้อเพลิงและช่วงเวลา รวมถึงอัลกอริทึม Linear Regression (Patcharacharoenwong et al., 2020; Settakhumpoo and Benjaoran, 2022; Karseewong and Boonlha, 2024) และ Random Forest (Ongrungrueng et al., 2020; Rattanachoung, 2023; Sangthong et al., 2023; Lekdee et al., 2024) เพื่อพยากรณ์แนวโน้มการปล่อยก๊าซในอนาคต เทคนิคเหล่านี้ช่วยให้เข้าใจพฤติกรรมการปล่อยก๊าซของภาคขนส่ง และสามารถประเมินผลกระทบต่อสิ่งแวดล้อมและสุขภาพได้อย่างแม่นยำ

การวิจัยนี้มีวัตถุประสงค์เพื่อวิเคราะห์แนวโน้มการปล่อยก๊าซคาร์บอนไดออกไซด์ในภาคขนส่ง และเปรียบเทียบประสิทธิภาพของอัลกอริทึมต่าง ๆ ในการพยากรณ์การปล่อยก๊าซ โดยใช้ตัวชี้วัด เช่น Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) และ R-squared ในการประเมินความแม่นยำของแบบจำลอง ผลการศึกษานี้สามารถใช้พยากรณ์การปล่อยก๊าซในอนาคต รวมถึงใช้เป็นแนวทางในการพัฒนาและวางแผนกลยุทธ์ลดการปล่อยก๊าซเรือนกระจกและมลพิษทางอากาศ เช่น PM2.5 ได้อย่างมีประสิทธิภาพ พร้อมกำหนดมาตรการเพื่อปรับปรุงคุณภาพอากาศและลดผลกระทบต่อสุขภาพ การศึกษานี้ยังเน้นการใช้เทคนิคเหมืองข้อมูลเพื่อให้ได้ข้อมูลเชิงลึกที่สนับสนุนการตัดสินใจด้านนโยบายและกลยุทธ์ในการลดการปล่อยมลพิษจากภาคขนส่ง ซึ่งเป็นปัจจัยสำคัญในการรักษาสิ่งแวดล้อมและส่งเสริมความยั่งยืนในอนาคต

วิธีการดำเนินงานวิจัย

เริ่มต้นจากการจัดเตรียมและทำความสะอาดข้อมูลให้พร้อมสำหรับการวิเคราะห์ จากนั้นใช้วิธี K-means Clustering เพื่อแบ่งข้อมูลออกเป็นกลุ่มน้ำมันและก๊าซธรรมชาติ ข้อมูลแต่ละกลุ่มถูกนำไปวิเคราะห์ด้วยโมเดล Linear Regression และ Random Forest เพื่อประเมินประสิทธิภาพในการพยากรณ์ สุดท้ายเปรียบเทียบผลลัพธ์ของแต่ละโมเดลเพื่อสรุปว่าโมเดลใดมีความเหมาะสมสูงสุด โดยขั้นตอนทั้งหมดแสดงดังรูปที่ 1

รูปที่ 1 ขั้นตอนการวิเคราะห์ข้อมูล

1. การรวบรวมข้อมูลการปล่อยก๊าซคาร์บอนไดออกไซด์ (CO_2) ในภาคขนส่ง

ดำเนินการโดยใช้ข้อมูลจากศูนย์เทคโนโลยีสารสนเทศและการสื่อสาร สำนักงานนโยบายและแผนพลังงาน (สนพ.) ที่เผยแพร่ผ่านคลังข้อมูลเปิดภาครัฐ (Open Data) ซึ่งจัดทำจากรายงานของกรมธุรกิจพลังงาน โดยมีรายละเอียดการใช้พลังงานในภาคขนส่ง แยกตามประเภทเชื้อเพลิง พร้อมการคำนวณการปล่อยก๊าซคาร์บอนไดออกไซด์ (CO_2) รายเดือน ข้อมูลครอบคลุม 37 ปี 9 เดือน ตั้งแต่มกราคม 2530 ถึงตุลาคม 2567 รวม 740 รายการ ครอบคลุมทั่วประเทศไทย https://catalog.eppo.go.th/dataset/dataset_11_63

2. การเตรียมข้อมูล

ทำความสะอาดข้อมูล (Data Cleaning) เช่น จัดการค่าขาดหาย (Missing Value) พร้อมแปลงข้อมูล (Data Transform) ให้อยู่ในรูปแบบที่วิเคราะห์ได้

3. การวิเคราะห์ด้วยอัลกอริทึมเหมืองข้อมูล

3.1 K-means Clustering เป็นอัลกอริทึมการจัดกลุ่มข้อมูลที่ได้รับความนิยมอย่างแพร่หลายในด้าน Machine Learning และ Data Mining โดยเป็นอัลกอริทึมประเภท Unsupervised Learning ซึ่งไม่จำเป็นต้องมีฉลาก (Labels) กำกับข้อมูลล่วงหน้า การทำงานของ K-means เริ่มจากการกำหนดจำนวนคลัสเตอร์ที่ต้องการ (K) แล้วสุ่มเลือกจุดศูนย์กลาง (Centroids) ให้แต่ละคลัสเตอร์ จากนั้นระบบจะทำการจัดกลุ่มข้อมูลโดยพิจารณาระยะห่างระหว่างข้อมูลแต่ละจุดกับ Centroid และกำหนดให้ข้อมูลนั้นอยู่ในกลุ่มที่ใกล้ที่สุด ต่อมาจะมีการคำนวณหาตำแหน่งใหม่ของ Centroid โดยใช้ค่าเฉลี่ยของจุดข้อมูลภายในแต่ละกลุ่ม กระบวนการนี้จะทำซ้ำไปเรื่อย ๆ จนกว่าตำแหน่งของ Centroid จะคงที่ หรือเมื่อค่าความคลาดเคลื่อนลดลงเพียงเล็กน้อย ซึ่งถือว่าอัลกอริทึมได้เข้าสู่จุดสมดุล

สมการของ K-means Clustering ใช้สมการระยะห่าง (Euclidean Distance) เพื่อจัดกลุ่มข้อมูล ดังสมการที่ (1)

$$d(x,y)=\sqrt{\sum_{i=1}^{n}(x_i-y_i)^2} \quad (1)$$

โดย x, y คือ จุดข้อมูล
n คือ จำนวนมิติของข้อมูล

การคำนวณ Centroid C_k ของแต่ละกลุ่มดังสมการที่ (2)

$$C_k = \frac{1}{N_k}\sum_{i=1}^{N_k} x_i \tag{2}$$

โดย N_k คือ จำนวนข้อมูลในกลุ่ม k

3.2 Linear Regression (การถดถอยเชิงเส้น) เป็นอัลกอริทึม Supervised Learning ที่ใช้ในการ พยากรณ์ค่าต่อเนื่อง หรือหาความสัมพันธ์ระหว่างตัวแปร โดยมีลักษณะเป็นเส้นตรง (Linear Relationship) ระหว่างตัวแปรอิสระ (Independent Variable, X) และตัวแปรตาม (Dependent Variable, Y) สมการทั่วไปของ Simple Linear Regression (มีตัวแปรอิสระ 1 ตัว) ดังสมการที่ (3)

$$Y = mX + b \tag{3}$$

โดยที่

Y = ค่าที่ต้องการพยากรณ์ (ตัวแปรตาม)
X = ตัวแปรอิสระ (Feature)
m = ค่าสัมประสิทธิ์ (Coefficient) หรือความชันของเส้น
b = ค่าคงที่ (Intercept)

ถ้ามีตัวแปรอิสระหลายตัว (Multiple Linear Regression) ดังสมการที่ (4)

$$Y = b_0 + b_1X_1 + b_2X_2 + \cdots + b_nX_n \tag{4}$$

โดยที่ b_0 เป็นค่าคงที่ และ $b_1, b_2, \ldots, b_n$ เป็นค่าสัมประสิทธิ์ของแต่ละตัวแปร

หลักการทำงานของ Linear Regression

1. กำหนดตัวแปร X และ Y
2. Fit เส้นตรงที่เหมาะสมกับข้อมูล โดยใช้วิธี Least Squares เพื่อลดค่า Loss Function สมการที่ (5)

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 \tag{5}$$

Loss Function Mean Squared Error (MSE) ซึ่งเป็นค่าเฉลี่ยของผลต่างระหว่างค่าจริง Y_i และค่าพยากรณ์ $\hat{Y}_i$

3. อัปเดตค่าสัมประสิทธิ์ m และ b โดยใช้ *Gradient Descent* หรือ *Ordinary Least Squares* (OLS)
4. เมื่อ Loss ลดลงจนคงที่ แสดงว่าได้เส้นที่เหมาะสมที่สุดแล้ว

3.3 Random Forest เป็นอัลกอริทึม Machine Learning ประเภท Supervised Learning ที่ใช้สำหรับ Regression (การพยากรณ์ค่าต่อเนื่อง) โดยเป็นการรวม Decision Trees หลายต้น (Ensemble Learning) เข้าด้วยกัน เพื่อให้การพยากรณ์แม่นยำขึ้นและลดปัญหา Overfitting

หลักการทำงานของ Random Forest

1. สร้างชุดข้อมูลสุ่ม (Bootstrapping) และเลือกตัวแปรแบบสุ่ม
   1.1 ระบบจะสุ่มเลือกข้อมูลบางส่วนจาก Training Set และเลือกตัวแปรบางตัวเพื่อสร้าง Decision Tree

1.2 แต่ละต้นไม้จะถูกฝึกด้วยชุดข้อมูลที่แตกต่างกัน

2. สร้างและฝึก Decision Trees หลายต้น

2.1 แต่ละ Decision Tree จะทำการพยากรณ์ค่า y โดยอิงจากตัวแปรที่เลือกมา

2.2 แต่ละต้นอาจให้ค่าพยากรณ์ที่แตกต่างกัน

3. รวมผลลัพธ์จากทุกต้นไม้

3.1 นำค่าที่ได้จากทุก Decision Tree มาเฉลี่ยกัน

3.2 ผลลัพธ์สุดท้ายจะเป็นค่าพยากรณ์เฉลี่ยของทุกต้นไม้

สมการของ Random Forest ดังสมการที่ (6)

$$\hat{y} = \frac{1}{N}\sum_{i=1}^{N} f_i(X) \tag{6}$$

โดยที่

$\hat{y}$ = ค่าพยากรณ์สุดท้าย

N = จำนวนต้นไม้ในป่า (Number of Trees)

$f_i(X)$ = ค่าพยากรณ์จาก Decision Tree ต้นที่ i

การประเมินผลโมเดล คือกระบวนการวัดและวิเคราะห์ประสิทธิภาพของโมเดลว่าทำงานได้ดีเพียงใดกับข้อมูลใหม่ ช่วยให้เลือกโมเดลที่เหมาะสมและปรับปรุงผลลัพธ์ให้ดียิ่งขึ้น ซึ่งเป็นขั้นตอนสำคัญใน Machine Learning และ Data Science

1. Mean Absolute Error (MAE) เป็นตัวชี้วัดที่ใช้วัดความคลาดเคลื่อนเฉลี่ยระหว่างค่าที่ทำนาย ($\hat{y}_i$) กับค่าจริง (y_i) โดยวัดจากค่าเฉลี่ยของความแตกต่างแบบสัมบูรณ์ (Absolute) ของแต่ละจุดข้อมูล ดังสมการที่ (7)

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{7}$$

2. Root Mean Squared Error (RMSE) เป็นตัวชี้วัดที่ใช้วัดความคลาดเคลื่อนแบบกำลังสอง และคำนวณค่าเฉลี่ยของความคลาดเคลื่อนก่อนที่จะนำมาถอดรากที่สอง เน้นให้ค่าความผิดพลาดที่สูงมีผลต่อค่า RMSE มากขึ้น ดังสมการที่ (8)

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{8}$$

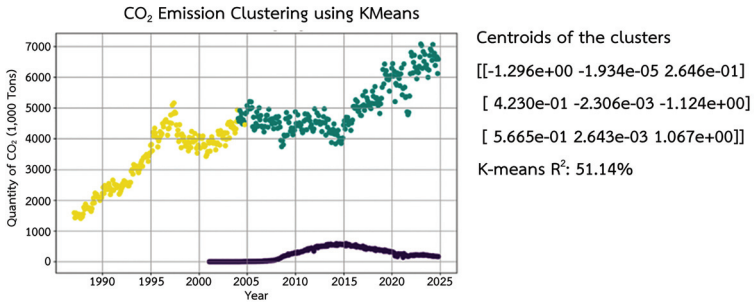
3. R^2 Score (Coefficient of Determination) เป็นตัววัดที่ใช้บอกว่าโมเดลของเราสามารถอธิบายความแปรปรวนของข้อมูลได้ดีแค่ไหน เมื่อเทียบกับค่าเฉลี่ยของข้อมูล ดังสมการที่ 9

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{9}$$

ผลการทดลอง

1. ผลการพยากรณ์การปล่อย CO_2 ที่เกิดจากการใช้ก๊าซธรรมชาติด้วยอัลกอริทึม K-means Clustering โดยข้อมูลการปล่อยก๊าซคาร์บอนไดออกไซด์ (CO_2) ในภาคขนส่งถูกจัดกลุ่มออกเป็น 3 กลุ่มหลัก ได้แก่ กลุ่มที่มีการปล่อยต่ำ แสดงถึงการปล่อย CO_2 อยู่ในระดับต่ำจากการใช้ก๊าซเป็นเชื้อเพลิงเพื่อเป็นพลังงานทางเลือกกลุ่มที่มีการปล่อยปานกลาง

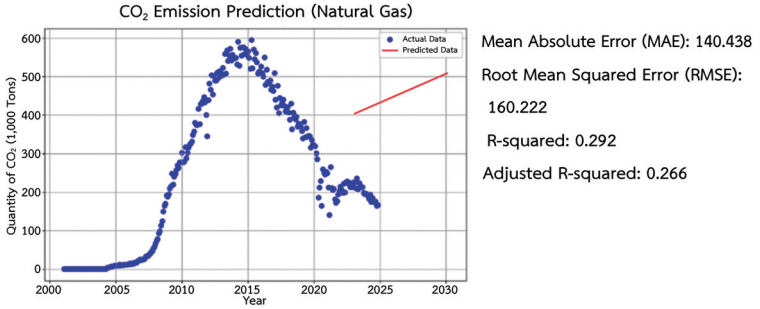
แสดงถึงการปล่อย CO_2 ในระดับปานกลาง ซึ่งพบในช่วงเวลาที่ผ่านมาและกลุ่มที่มีการปล่อยสูง สะท้อนการปล่อย CO_2 สูงในช่วงเวลาที่มีการขยายตัวของอุตสาหกรรมจนถึงปัจจุบัน เมื่อทำการเปรียบเทียบพฤติกรรมของการปล่อย CO_2 จากแต่ละกลุ่ม พบว่าแนวโน้มที่ชัดเจนคือ การเพิ่มขึ้นของการปล่อย CO_2 ต่อเนื่องในปัจจุบัน เมื่อเทียบกับอดีต การเปลี่ยนแปลงนี้สะท้อนถึงการใช้เชื้อเพลิงที่มีผลกระทบสูงต่อสิ่งแวดล้อมและการปล่อย CO_2 จากกิจกรรมทางเศรษฐกิจที่ขยายตัว ทั้งนี้การศึกษาดังกล่าวได้เน้นย้ำถึงความสำคัญในการใช้พลังงานที่มีความยั่งยืนเพื่อลดการปล่อย CO_2 และช่วยในการต่อสู้กับการเปลี่ยนแปลงสภาพภูมิอากาศในอนาคต



รูปที่ 2 การปล่อยก๊าซคาร์บอนไดออกไซด์ (CO_2) โดยอัลกอริทึม K-means Clustering

จากรูปที่ 2 แสดงให้เห็นโดยชัดเจนว่ามีการแบ่งกลุ่มของการปล่อยก๊าซคาร์บอนไดออกไซด์ (CO_2) ถูกจัดกลุ่มอย่างชัดเจนด้วยเทคนิค K-means ได้ 3 กลุ่ม คือ กลุ่มที่ 1 (สีเหลือง) กลุ่มนี้คือ การปล่อยก๊าซคาร์บอนไดออกไซด์ (CO_2) จากน้ำมันเชื้อเพลิงในอดีต ก่อนปี ค.ศ. 2005 ซึ่งมักจะมีปริมาณการปล่อยที่สูงในช่วงแรก ๆ และอาจมีลักษณะที่กระจุกตัวเนื่องจากการพึ่งพาพลังงานจากน้ำมันเป็นหลักในช่วงเวลานั้น โดยกลุ่มนี้จะมีลักษณะการกระจายที่ชัดเจนในช่วงเวลานี้ กลุ่มที่ 2 (สีคราม) การปล่อยก๊าซคาร์บอนไดออกไซด์ (CO_2) จากน้ำมันเชื้อเพลิงในช่วงปี ค.ศ. 2005 ถึงปัจจุบัน แม้ว่าจะมีการพยายามใช้พลังงานทางเลือกเพื่อทดแทน แต่การใช้พลังงานจากน้ำมันยังคงเพิ่มขึ้นอย่างต่อเนื่อง กลุ่มนี้จะแสดงให้เห็นถึงการขยายตัวของภาคอุตสาหกรรมและการขนส่งที่ยังคงใช้เชื้อเพลิงฟอสซิลเป็นหลัก กลุ่มที่ 3 (สีม่วง) กลุ่มนี้จะเกี่ยวข้องกับการปล่อยก๊าซคาร์บอนไดออกไซด์ (CO_2) จากก๊าซธรรมชาติ ซึ่งมักจะมีการปล่อยที่เพิ่มขึ้นในช่วงเวลาปัจจุบัน โดยอาจจะมีลักษณะการกระจายที่ต่ำกว่าและแตกต่างจากกลุ่มน้ำมัน เนื่องจากก๊าซธรรมชาติมักถูกมองว่าเป็นพลังงานสะอาดมากกว่า แต่ก็ยังมีการปล่อยมลพิษที่ไม่สามารถหลีกเลี่ยงได้ทั้งหมด การปล่อยก๊าซคาร์บอนไดออกไซด์ (CO_2) โดยเฉพาะจากน้ำมันเชื้อเพลิงยังคงเพิ่มสูงขึ้นอย่างต่อเนื่อง แม้ว่าจะมีความพยายามในการนำพลังงานทดแทนมาใช้ในการขนส่ง แต่การปล่อยก๊าซ CO_2 ก็ยังไม่ลดลงตามที่คาดหวัง โดยปริมาณการปล่อยจากน้ำมันเชื้อเพลิงยังคงมีความสำคัญและส่งผลกระทบต่อสิ่งแวดล้อมอย่างต่อเนื่อง

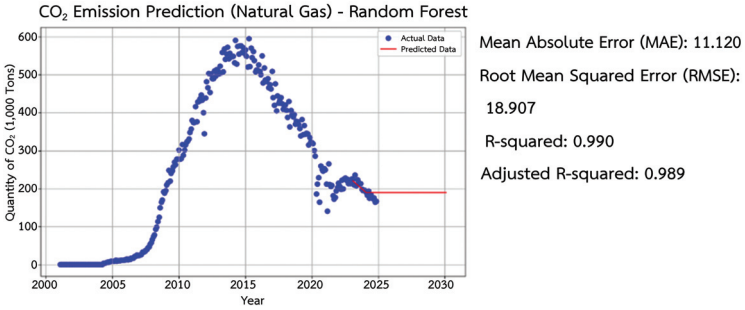
2. การพยากรณ์การปล่อยก๊าซคาร์บอนไดออกไซด์ (CO_2) จากการใช้ก๊าซธรรมชาติด้วยอัลกอริทึม Linear Regression



รูปที่ 3 การปล่อยก๊าซคาร์บอนไดออกไซด์ (CO_2) ที่เกิดจากการใช้ก๊าซธรรมชาติเป็นเชื้อเพลิงโดยอัลกอริทึม Linear Regression

จากรูปที่ 3 แสดงให้เห็นว่าค่าที่ได้จากการที่จุดทำนาย (สีแดง) แตกต่างจากจุดจริง (สีน้ำเงิน) มาก แสดงว่าโมเดลไม่สามารถทำนายค่าได้อย่างแม่นยำ ซึ่งอาจเกิดจากปัจจัยที่โมเดลไม่ได้คำนึงถึง เช่น ความผันผวนของราคาก๊าซธรรมชาติ (Natural Gas) ที่มีการเปลี่ยนแปลงอย่างรวดเร็วและค่าใช้จ่ายที่สูง ทำให้ทำนายยากขึ้น เพราะราคาก๊าซธรรมชาติมีผลต่อปัจจัยหลายอย่างในการคำนวณต้นทุนที่อาจไม่ได้รับการพิจารณาในโมเดลนี้ การใช้โมเดลที่ซับซ้อนมากขึ้นอาจช่วยจับความสัมพันธ์ที่ลึกซึ้งระหว่างปัจจัยเหล่านี้และปรับปรุงความแม่นยำในการทำนายได้

3. การพยากรณ์การปล่อยก๊าซคาร์บอนไดออกไซด์ (CO_2) จากการใช้ก๊าซธรรมชาติด้วยอัลกอริทึม Random Forest



รูปที่ 4 การปล่อยก๊าซคาร์บอนไดออกไซด์ (CO_2) ที่เกิดจากการใช้ก๊าซเป็นเชื้อเพลิงโดยใช้แบบจำลอง Random Forest

จากรูปที่ 4 พบว่าโมเดล Random Forest ให้ค่า MAE และ RMSE ที่ต่ำแสดงว่าโมเดล Random Forest สามารถทำนายได้แม่นยำขึ้น โดยค่าผิดพลาดเฉลี่ยและความผิดพลาดที่ยกกำลังสองมีค่าน้อยมาก ซึ่งหมายความว่าโมเดลทำงานได้ดีในการทำนายข้อมูลจริง ส่วนค่า R-squared ที่สูงถึง 0.990 และ Adjusted R-squared ที่ 0.989 แสดงว่าโมเดลสามารถอธิบายความแปรปรวนของข้อมูลได้เกือบทั้งหมด การใช้ Natural Gas ด้วย Random Forest มีความแม่นยำสูง เพราะโมเดลนี้สามารถจับความสัมพันธ์ที่ซับซ้อนได้ดีกว่าโมเดลอื่น ๆ และเหมาะสมกับข้อมูลที่มีความหลากหลาย

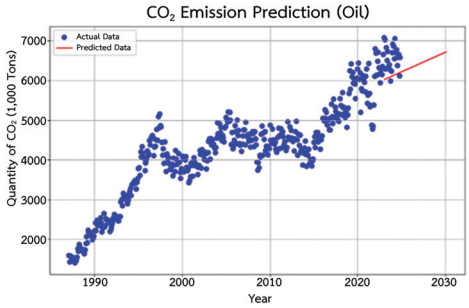
ตารางที่ 1 เปรียบเทียบค่าตัวชี้วัดที่ได้การปล่อยก๊าซคาร์บอนไดออกไซด์จากการใช้ก๊าซธรรมชาติแต่ละโมเดล (Linear Regression, Random Forest, K-means)

Metric	Linear Regression	Random Forest	K-means
MAE	140.44	11.12	-
RMSE	160.22	18.91	-
R-squared	29.00 %	99.02 %	51.14 %
Adjusted R-squared	26.70 %	98.98 %	-

จากตารางที่ 1 จากการเปรียบเทียบค่าตัวชี้วัดของโมเดลที่ใช้ในการพยากรณ์ข้อมูลการปล่อยก๊าซคาร์บอนไดออกไซด์จากก๊าซธรรมชาติ พบว่าโมเดล Linear Regression แม้จะสามารถแสดงแนวโน้มเชิงเส้นของข้อมูลได้ดี และมีความเรียบง่ายในการตีความ แต่ให้ผลลัพธ์ที่แม่นยำไม่มากนัก โดยมีค่า MAE และ RMSE สูงถึง 140.44 และ 160.22 ขณะที่ค่า R-squared อยู่ที่เพียงร้อยละ 29 สะท้อนว่าโมเดลยังไม่สามารถจับความซับซ้อนของข้อมูลได้ดีนัก โดยเฉพาะในกรณีที่ความสัมพันธ์ระหว่างตัวแปรไม่เป็นเชิงเส้นอย่างชัดเจน ในทางตรงกันข้ามโมเดล Random Forest ซึ่งเป็นอัลกอริทึมแบบ Ensemble Learning ที่รวมผลของหลายต้นไม้ตัดสินใจ (Decision Trees) สามารถเรียนรู้ความสัมพันธ์ที่ซับซ้อนและไม่เป็นเชิงเส้นของข้อมูลได้อย่างมีประสิทธิภาพ ส่งผลให้ค่าความผิดพลาดลดลงอย่างมาก โดย MAE เท่ากับ 11.12 และ RMSE เท่ากับ 18.91 และค่า R-squared สูงถึงร้อยละ 99.02 ซึ่งชี้ว่าโมเดลสามารถอธิบายความแปรปรวนของข้อมูลได้เกือบทั้งหมด จึงถือว่าเป็นโมเดลที่เหมาะสมและแม่นยำที่สุดในบริบทของการพยากรณ์

การปล่อย CO_2 จากก๊าซธรรมชาติ ส่วนโมเดล K-means ซึ่งใช้สำหรับการจัดกลุ่มข้อมูล ไม่ใช่การพยากรณ์โดยตรง จึงไม่มีค่าชี้วัด MAE หรือ RMSE อย่างไรก็ตาม ค่า R-squared ที่ได้จากการจัดกลุ่มอยู่ที่ร้อยละ 51.14 แสดงให้เห็นว่าการจัดกลุ่มสามารถอธิบายความแปรปรวนของข้อมูลได้ในระดับหนึ่ง อาจเป็นประโยชน์สำหรับการจำแนกประเภทของกิจกรรมที่ส่งผลต่อการปล่อย CO_2 เพื่อการวิเคราะห์เชิงสำรวจ (Exploratory Analysis) มากกว่าการพยากรณ์

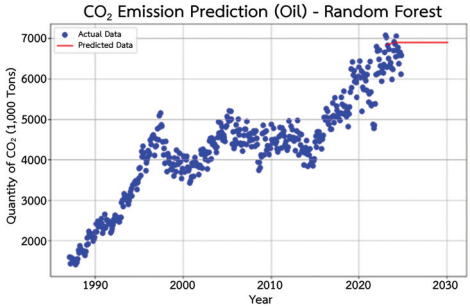
4. การพยากรณ์การปล่อยก๊าซคาร์บอนไดออกไซด์ (CO_2) จากการใช้น้ำมันเป็นเชื้อเพลิงด้วยอัลกอริทึม Linear Regression



รูปที่ 5 การปล่อยก๊าซคาร์บอนไดออกไซด์ (CO_2) ที่เกิดจากการใช้น้ำมันเป็นเชื้อเพลิงโดยใช้แบบจำลอง Linear Regression

จากรูปที่ 5 พบว่าค่าทำนาย (สีแดง) แตกต่างจากจุดจริง (สีน้ำเงิน) น้อย แสดงว่าโมเดลทำนายได้ดีขึ้น โดยมีค่า MAE และ RMSE ที่ต่ำกว่า ซึ่งหมายความว่าโมเดลมีความผิดพลาดน้อยลง และสามารถทำนายได้แม่นยำขึ้น นอกจากนี้ ค่า R-squared ที่สูงถึง 0.865 และ Adjusted R-squared ที่ 0.862 แสดงว่าโมเดลสามารถอธิบายความแปรปรวนของข้อมูลได้มากกว่าร้อยละ 86 การใช้น้ำมันเป็นตัวแปรที่ค่อนข้างคาดการณ์ได้ เพราะมีการใช้งานอย่างต่อเนื่องและสามารถตรวจสอบปัจจัยต่าง ๆ ได้ง่ายกว่าการใช้งานของพลังงานประเภทอื่น ๆ

5. การพยากรณ์การปล่อยก๊าซคาร์บอนไดออกไซด์ (CO_2) จากการใช้น้ำมันเป็นเชื้อเพลิงด้วยอัลกอริทึม Random Forest



รูปที่ 6 การปล่อยก๊าซคาร์บอนไดออกไซด์ (CO_2) ที่เกิดจากการใช้ น้ำมันเป็นเชื้อเพลิงโดยใช้แบบจำลอง Random Forest

จากรูปที่ 6 พบว่าค่าทำนาย (สีแดง) ใกล้เคียงกับจุดจริง (สีน้ำเงิน) แสดงว่าโมเดล Random Forest ทำงานได้ดี โดยมีค่า MAE และ RMSE ที่ต่ำ แสดงถึงความผิดพลาดที่ไม่มากนัก ค่า R-squared และ Adjusted R-squared ที่สูงถึง 0.971 และ 0.970 หมายความว่าโมเดลสามารถอธิบายข้อมูลได้เกือบร้อยละ 97 โมเดลนี้เหมาะสมในการทำนายข้อมูลที่มีความซับซ้อนและความแปรปรวนสูง ผลลัพธ์นี้บ่งชี้ว่าโมเดลสามารถจับความสัมพันธ์ระหว่างตัวแปรต่าง ๆ ได้อย่างมีประสิทธิภาพ

ตารางที่ 2 เปรียบเทียบค่าตัวชี้วัดที่ได้จากน้ำมันเชื้อเพลิงแต่ละโมเดล (Linear Regression, Random Forest, K-means)

Metric	Linear Regression	Random Forest	K-means
MAE	418.259	154.260	-
RMSE	524.337	242.136	-
R-squared	86.50 %	97.10 %	51.14 %
Adjusted R-squared	86.20 %	97.00 %	-

จากตารางที่ 2 พบว่าโมเดล Random Forest มีประสิทธิภาพในการพยากรณ์สูงที่สุดเมื่อเทียบกับ Linear Regression และ K-means โดยมีค่า MAE และ RMSE ต่ำที่สุด (154.260 และ 242.136) ซึ่งแสดงถึงความแม่นยำในการทำนายสูง และค่าความแปรปรวนที่โมเดลสามารถอธิบายได้ (R-squared) อยู่ที่ 97.1 % ถือว่าสูงมากใกล้เคียงกับโมเดลสมบูรณ์แบบ ขณะที่โมเดล Linear Regression แม้จะมีค่า R-squared สูงถึง 86.5 % แต่ค่าความผิดพลาด MAE และ RMSE ยังค่อนข้างสูง (418.259 และ 524.337) ซึ่งอาจเกิดจากสมมติฐานเชิงเส้นของโมเดลที่ไม่สามารถจับความสัมพันธ์ที่ซับซ้อนในข้อมูลได้อย่างมีประสิทธิภาพ สำหรับ K-means Clustering แม้จะไม่ใช่โมเดลสำหรับการพยากรณ์โดยตรง จึงไม่มีค่า MAE และ RMSE แต่ยังสามารถใช้ประโยชน์ในการจำแนกกลุ่มของข้อมูลการปล่อย CO_2 ได้ระดับหนึ่ง โดยมีค่า R-squared อยู่ที่ 51.14 % ซึ่งสามารถนำไปใช้วิเคราะห์โครงสร้างของข้อมูลเบื้องต้น หรือใช้ร่วมกับโมเดลอื่น ๆ ในการวางกลยุทธ์เชิงกลุ่มได้

สรุปการเปรียบเทียบ

จากการเปรียบเทียบผลลัพธ์ของโมเดล Linear Regression, Random Forest และ K-means โดยอิงจากข้อมูลการใช้เชื้อเพลิงธรรมชาติที่มีความผันผวน พบว่า Random Forest ให้ผลแม่นยำที่สุด โดยมีค่า MAE และ RMSE ต่ำกว่า Linear Regression อย่างมีนัยสำคัญ สะท้อนถึงความสามารถในการพยากรณ์ CO_2 ได้ดีกว่า อีกทั้งค่า R-squared และ Adjusted R-squared ของ Random Forest อยู่ที่ 99.02 % และ 98.98 % แสดงให้เห็นถึงความสามารถในการอธิบายความแปรปรวนของข้อมูลได้ดีกว่า Linear Regression ที่มีเพียง 29.00 % ขณะที่ K-means ซึ่งเน้นการจัดกลุ่มข้อมูลมีค่า R-squared 51.14 % สูงกว่า Linear Regression แต่ยังต่ำกว่า Random Forest อย่างชัดเจน โดยสรุป Random Forest เหมาะสำหรับการพยากรณ์ CO_2 ที่ซับซ้อนและไม่เป็นเชิงเส้น ในขณะที่ Linear Regression อธิบายความแปรปรวนได้จำกัด และ K-means เหมาะกับการจัดกลุ่มมากกว่าพยากรณ์ เมื่อทดสอบกับข้อมูลการใช้น้ำมันที่มีความผันผวนน้อย Random Forest ยังคงให้ผลลัพธ์แม่นยำสูงสุดเช่นเดิม

สาเหตุที่ Random Forest มีประสิทธิภาพเหนือกว่า Linear Regression และ K-means มาจากลักษณะข้อมูลการใช้เชื้อเพลิงที่มีความผันผวนสูงและความสัมพันธ์ที่ไม่เป็นเชิงเส้น อีกทั้งอาจมีปฏิสัมพันธ์ระหว่างตัวแปรที่ซับซ้อน ซึ่ง Linear Regression ซึ่งเป็นโมเดลเชิงเส้นไม่สามารถจัดการกับความซับซ้อนได้ดี ในขณะที่ Random Forest ซึ่งเป็นโมเดลแบบ Tree-Based Ensemble สามารถจับความสัมพันธ์ที่ไม่เป็นเชิงเส้นได้ดี โดยอาศัยการรวมผลของหลายต้นไม้เพื่อลดความเอนเอียงและเพิ่มความแม่นยำในการพยากรณ์ ส่วน K-means แม้จะช่วยจัดกลุ่มข้อมูลตามลักษณะการใช้เชื้อเพลิงได้ แต่ไม่เหมาะกับการพยากรณ์เพราะไม่ได้ออกแบบมาเพื่อหาแนวโน้มของตัวแปรเป้าหมายและไม่พิจารณาความสัมพันธ์เชิงสาเหตุ นอกจากนี้เมื่อทดสอบด้วยข้อมูลการใช้น้ำมันที่มีความผันผวนน้อย ผลลัพธ์ยังแสดงว่า Random Forest มีความแม่นยำสูงสุด สะท้อนถึงความยืดหยุ่นและความแข็งแกร่งของโมเดลในการพยากรณ์ CO_2 จากข้อมูลที่หลากหลาย

สรุปผล

การวิจัยนี้มุ่งเน้นการวิเคราะห์และพยากรณ์การปล่อยก๊าซคาร์บอนไดออกไซด์ (CO_2) ในภาคขนส่ง โดยใช้อัลกอริทึมเหมืองข้อมูล (Data Mining Algorithms) เช่น K-means Clustering, Linear Regression และ Random Forest

เพื่อตรวจสอบแนวโน้มการปล่อย CO_2 และคาดการณ์ในอนาคต รวมทั้งเปรียบเทียบประสิทธิภาพของอัลกอริทึมที่ใช้การพยากรณ์ด้วย Linear Regression แบบจำลอง Linear Regression ช่วยแสดงแนวโน้มการเพิ่มขึ้นของการปล่อย CO_2 โดยพยากรณ์การปล่อย CO_2 ในระยะยาว ผลลัพธ์ที่ได้มีความแม่นยำพอสมควร โดยค่าตัวชี้วัด MAE, RMSE และ R-squared ชี้ให้เห็นว่าแม้จะมีความสัมพันธ์เชิงบวกกับปริมาณเชื้อเพลิงที่ใช้ แต่ค่าความแม่นยำยังอยู่ในระดับที่สามารถปรับปรุงได้ การพยากรณ์ด้วย Random Forest การใช้ Random Forest ช่วยเพิ่มความแม่นยำในการพยากรณ์การปล่อย CO_2 โดยค่าตัวชี้วัด MAE, RMSE และ R-squared ชี้ให้เห็นว่า Random Forest ให้ผลลัพธ์ที่แม่นยำและเสถียรกว่า Linear Regression มาก โดยมี R-squared ใกล้เคียงกับ 1 ซึ่งบ่งชี้ถึงความสัมพันธ์ที่แข็งแกร่งระหว่างข้อมูลที่ใช้ จากการเปรียบเทียบพบว่า Random Forest ให้ผลลัพธ์ที่แม่นยำและเสถียรกว่า Linear Regression อย่างมีนัยสำคัญ จึงถือเป็นโมเดลที่เหมาะสมสำหรับการพยากรณ์การปล่อย CO_2 ในบริบทของข้อมูลที่มีความผันผวนและซับซ้อน นอกจากนี้แบบจำลอง Random Forest ยังสามารถนำไปประยุกต์ใช้ได้อย่างมีประสิทธิภาพทั้งในภาครัฐและภาคเอกชน โดยในภาครัฐ เช่น สำนักงานนโยบายและแผนพลังงาน (สนพ.) หรือกระทรวงพลังงาน สามารถใช้เพื่อวางแผนเชิงนโยบายด้านการใช้เชื้อเพลิงอย่างแม่นยำยิ่งขึ้น และคาดการณ์ปริมาณการปล่อย CO_2 เพื่อสนับสนุนการกำหนดมาตรการลดก๊าซเรือนกระจกได้อย่างมีประสิทธิภาพ ส่วนในภาคเอกชน โดยเฉพาะกลุ่มธุรกิจโลจิสติกส์หรือขนส่ง โมเดลนี้สามารถนำไปใช้ในการประเมินคาร์บอนฟุตพริ้นท์จากกิจกรรมขนส่ง ตลอดจนวางกลยุทธ์ลดการใช้เชื้อเพลิงฟอสซิล เช่น การปรับเปลี่ยนเส้นทางการเดินรถ หรือการเลือกใช้ยานพาหนะที่มีประสิทธิภาพด้านพลังงาน เพื่อส่งเสริมการดำเนินงานที่ยั่งยืนและเป็นมิตรต่อสิ่งแวดล้อม

References

Aekrattanawat, S., Sangkhiew, N., Karot, T. and Jomtong, P. (2024). The Analysis of Carbon Dioxide Emissions from Driving Behavior Using Multiple Linear Regression: A Case Study of Sample Company. *The Journal of Industrial Technology: Suan Sunandha Rajabhat University*, 12(2), 87-95. https://ph01.tci-thaijo.org/index.php/fit-ssru/article/view/257827/173941 (in Thai)

Kanabkaew, T. (2024). Thailand and Overcoming the PM2.5 Crisis. *Journal of Multidisciplinary Academic Research and Development (JMARD)*, 6(4), 1-9. https://he02.tci-thaijo.org/index.php/JMARD/article/view/272524/185467 (in Thai)

Karseewong, S. and Boonlha, K. (2024). Prediction of Carbon Dioxide Emission from Energy Consumption in Thailand with Sarima-Ann-Reg Model. *Srinakharinwirot University Journal of Sciences and Technology*, 16(32), 1-11, https://ph02.tci-thaijo.org/index.php/swujournal/article/view/251667/171410 (in Thai)

Khruachalee, K., Kitkamhang, O. and Apinawin, W. (2024). An Assessment and Forecasting of Carbon Dioxide (CO_2) Emissions from the Transportation Sector Using Regression Analysis and Artificial Neural Network Methods. *Journal of Energy and Environment Technology*, 11(2), 50-65. https://ph01.tci-thaijo.org/index.php/JEET/article/view/258090/173778 (in Thai)

Khatbanjong, S. and Mawan, V. (2021). Artificial Intelligence (AI) Technology and the Categorization of Fifth Grade Students' Academic Achievements Using K-means Clustering Learning Algorithm. *Journal of Education Studies*, 49(2), EDUCU4902015. https://doi.org/10.14456/educu.2021.36

Lekdee, T., Santianotai, R. and Udompittayason, J. (2024). Efficiency Comparison of Lung Cancer Risk Prediction Models using Data-mining Techniques. *Science and Technology to Community*, 2(1), 22-35. https://li02.tci-thaijo.org/index.php/STC/article/view/705/449 (in Thai)

Maijaroensri, P. and Hirunrueng, T. (2024). A Study of Method for Solving Air Pollution from Particulate Matter and Impacts of Health: Case Study in Tapioca Starch Business and Cement Grinding and Mix Business. *Regional Health Promotion Center 9 Journal*, 18(3), 1074-1090. https://he02.tci-thaijo.org/index.php/RHPC9Journal/article/view/269434/184443 (in Thai)

Namcome, T. and Tansuchat, R. (2021). The Impact of Fine Particulate Matter (PM 2.5) Pollution on the Number of Foreign Tourists in Chiang Mai and Bangkok. *Rajabhat Chiang Mai Research Journal*, 22(3), 19-35. https://so05.tci-thaijo.org/index.php/cmruresearch/article/view/247437/173176 (in Thai)

Ongrungrueng, D., Thaiwirach, S. and Prachuabsupakij, W. (2020). Forecasting Model for the Amount of Spare Parts Storage in the Warehouse Using Machine Learning Techniques: Case Study of a Cement Plant. *PSRU Journal of Science and Technology*, 5(1), 81-92. https://ph01.tci-thaijo.org/index.php/Scipsru/article/view/239979/163864 (in Thai)

Palason, K., Pol-ard, P. and Rattanasamakarn, T. (2024). Analyzing the Relationship between Economic Growth and Carbon Dioxide (CO_2) Emissions in Thailand: An Application of the VAR Model. *Journal of Management Science and Communication*, 3(2), 1-28. https://so07.tci-thaijo.org/index.php/jmsc_journal/article/view/5587/3885 (in Thai)

Patcharacharoenwong, C., Hernmek, K. and Kimpan, W. (2020). Arrival Time Prediction Model to a Pier for Public Transportation Boats. *Journal of Science Ladkrabang*, 29(2), 31-44. https://li01.tci-thaijo.org/index.php/science_kmitl/article/view/241105/169801 (in Thai)

Phounsataphorn, A. and Bamroongcheep, U. (2025). Intelligent Student Grouping with "Altair Studio K-means AI" The Helper for Teachers Digital Era. *Journal of Innovation in Administration and Educational Management*, 3(1), 106-122. https://so11.tci-thaijo.org/index.php/IAEM/article/view/1287/596 (in Thai)

Posing, N., Luangsiriwan, A., Boonpok, W. and Moonjat, T. (2024). An Analysis and Clustering of Relationships of Digital Literacy Instructors' data Using Unsupervised Learning of Data Mining Techniques. *Journal of Technology Management Rajabhat Maha Sarakham University*, 11(1), 144-158, https://ph02.tci-thaijo.org/index.php/itm-journal/article/view/252989/170699 (in Thai)

Rattanachoung, N. (2023). Rapid Prediction of Melon Sweetness Using Image Processing Techniques and Algorithmic Models. *Journal of Applied Research on Science and Technology (JARST)*, 22(1), 117-127. https://ph01.tci-thaijo.org/index.php/rmutt-journal/article/view/250908/170622 (in Thai)

Sangthong, K., Chotchara, P., Jantawong, L. and Aiadrit, A. (2023). The Comparison of Data Classification Efficiency to Predict the Decision-Making in Future Elections Among Thai Businesspersons Using Data Mining Techniques. *Local Administration Journal*, 16(4), 559-578. https://so04.tci-thaijo.org/index.php/colakkujournals/article/view/265610/182405 (in Thai)

Settakhumpoo, J. and Benjaoran, V. (2022). The Comparative Accuracy Evaluation of Water Evaporation using Multiple Linear Regression and Simplified Penman's Equation. *Ladkrabang Engineering Journal*, 39(4), 138-148. https://ph01.tci-thaijo.org/index.php/lej/article/view/248401/170141 (in Thai)

Wongchan, W. (2023). Health Promotion and Prevention Diseases of Particulate Matter (PM2.5) in School-Age Children. *Ramathibodi Medical Journal*, 46(4), 52-65. https://he02.tci-thaijo.org/index.php/ramajournal/article/view/266309/181836 (in Thai)