ISSN 2616-5775 Vol. 7, Issue 12: 24-35, DOI: 10.25236/AJCIS.2024.071204

Machine learning-based PM concentration prediction and model interpretability analysis

Junxi Li*

School of Mathematical Sciences & College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, 518060, China
**Corresponding author*

Abstract: *Accurate $PM_{2.5}$ concentration forecasting is pivotal for environmental health and sustainable development. This study introduces a machine learning model leveraging the SHAP framework for enhanced interpretability and prediction accuracy. Utilizing 2023 meteorological data from Beijing's Wanshou Xigong meteorological station, we initially explored all characteristics and selected three algorithms, RF, SVR, and LightGBM, to construct machine learning models for $PM_{2.5}$ concentration prediction. The R^2 of each model on the validation set reached 0.9334, 0.9185, and 0.9472. Ultimately, we conducted SHAP framework interpretability analysis on the LightGBM model, removing features with minimal predictive impact. The R^2 of the final prediction model reaches 0.9501. This advancement significantly aids in precisely predicting $PM_{2.5}$ concentration, supporting proactive environmental and health policies.*

Keywords: *RF, SVR, LightGBM, SHAP Framework*

1. Background And Introduction

$PM_{2.5}$, atmospheric particles under 2.5 μm that penetrate deep into the lungs, are laden with toxins and pose significant health risks, including respiratory ailments and lung cancer[1]. Recognized globally as a critical environmental pollutant, precise and timely $PM_{2.5}$ forecasting is essential for crafting air quality policies. Such predictions empower authorities to adjust emissions, enhance transportation, and establish warning systems, thereby combating climate change, safeguarding health, and fostering sustainable development.

Predicting $PM_{2.5}$ concentrations is a complex challenge integral to air quality management. Traditional methods such as statistical, numerical, and integrated models have been pivotal but face limitations in real-time data updating and predictive accuracy. The WRF-Chem model[2], while influential, struggles with timely data synchronization, leading to delayed forecasts. To address this, satellite remote sensing has emerged as a promising approach, yet its complexity in capturing spatial-temporal dynamics hinders real-time predictive capabilities[3]. As data volume swells, traditional machine learning models[4] encounter drawbacks like high computational costs and uncertainties. This paper introduces the LightGBM model, which offers efficient, scalability, and multi-thread parallelization, coupled with the SHAP framework for model interpretability and optimization. Our technical approach, depicted in Fig. 1, which focuses on training the model through data processing, model training, model comparison, and SHAP interpretation modification. Finally, the SHAP-LightGBM model is utilized to predict $PM_{2.5}$ concentration.

ISSN 2616-5775 Vol. 7, Issue 12: 24-35, DOI: 10.25236/AJCIS.2024.071204

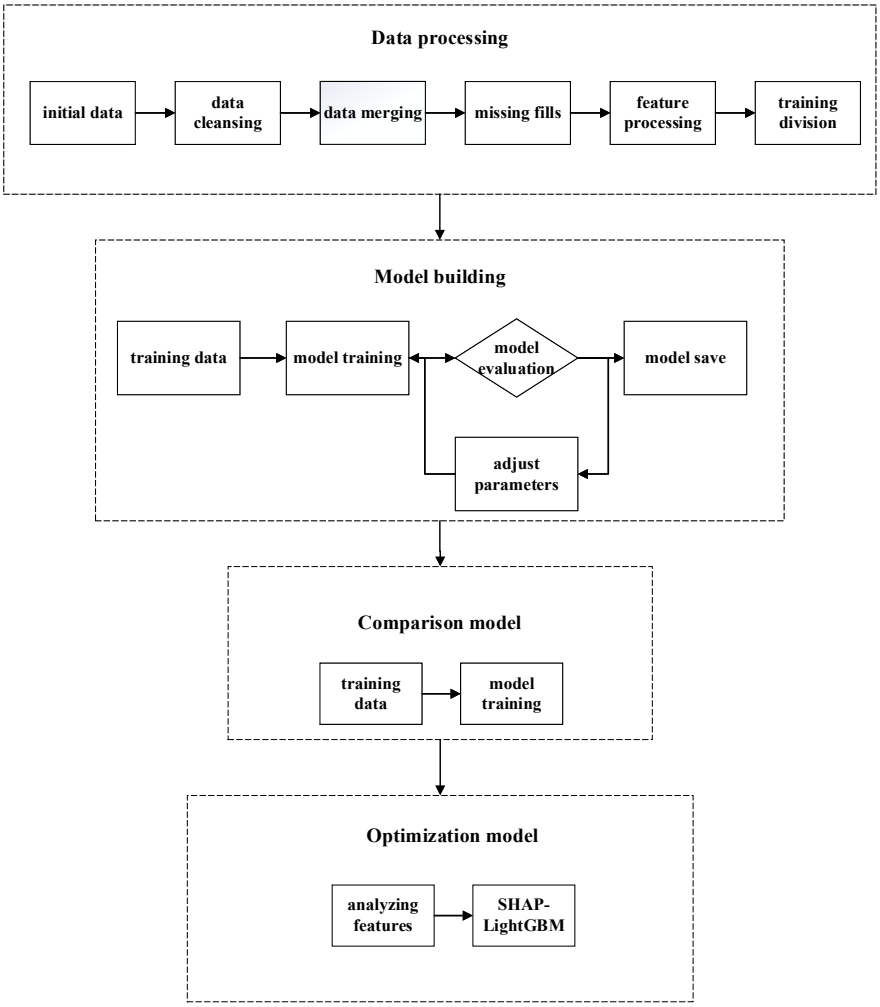

Figure 1: Flow chart of the experiment

2. Related Work

2.1 Data set sources and analysis

The dataset collects a total of 8760 data for each hourly period of the whole year 2023 from the WanshouXigong meteorological station. In addition to the four temporal features, there are 11 main data features as follows: $PM_{2.5}$, PM_{10}, SO_2, NO_2, CO, O_3, TEMP, P, Precipitation, Direction, and Velocity.

Weather station transmission as well as logging problems can result in some data being missing and unrecorded. Given the strong autocorrelation within the data, Lagrange interpolation is employed to fill in missing values, facilitating the model training process. Table I shows the number of missing values for each feature.

Table 1: Number of missing values for each feature

Features	Missing values
$PM_{2.5}$	105
PM_{10}	65
SO_2	184
NO_2	262
CO	719
O^3	625
Direction	1

The Lagrange interpolation method[5] is as follows:

ISSN 2616-5775 Vol. 7, Issue 12: 24-35, DOI: 10.25236/AJCIS.2024.071204

$$P(\mathrm{x}) = \sum_{i=0}^{n} y_i \times \prod_{\substack{j=0 \\ j \neq i}}^{n} \frac{x - x_j}{x_i - x_j} \tag{1}$$

(x_i, y_i) is the data points;

x is the index value of the missing value;

$P(\mathrm{x})$ is the fill value of the missing value.

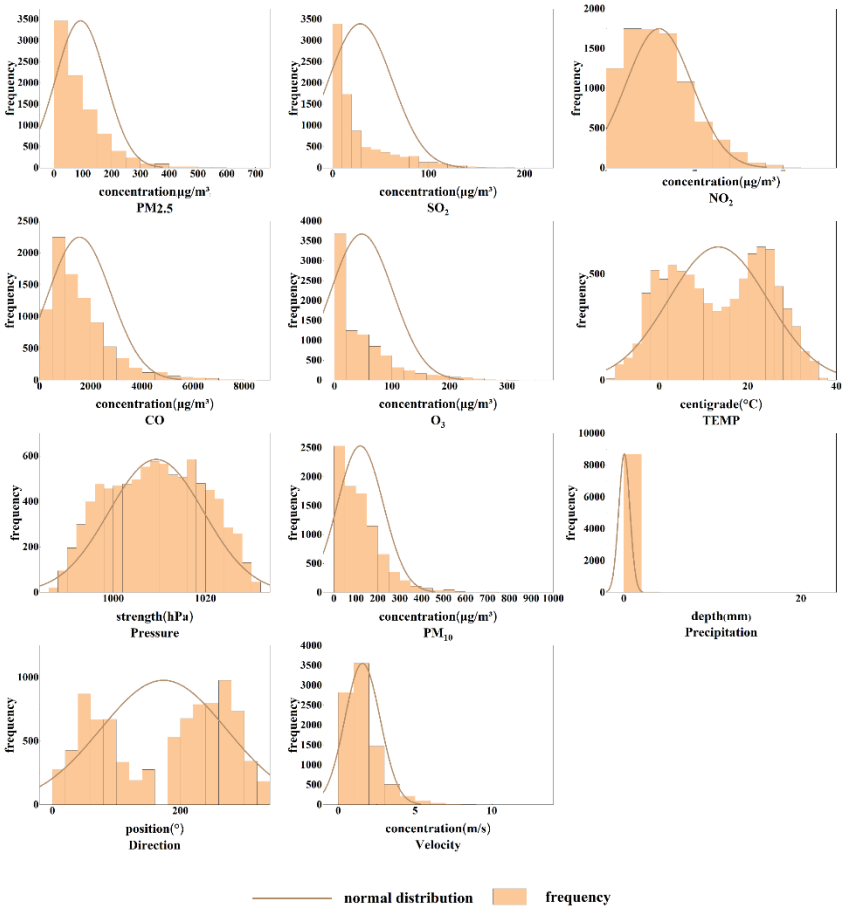

Figure 2: Data Distribution Chart

Figure 2 shows the distribution of the data plotted after data processing, it is obvious that these eleven features are more evenly distributed except for the feature Precipitation. In order to train the model, 8760 data were divided into five parts, four as the training set and one as the validation set for prediction and to test the prediction effect of the model. At the same time, the ten features PM_{10}, SO_2, NO_2, CO, O_3, TEMP, P, Precipitation, Direction, and Velocity, as well as the temporal features, are set as inputs to the prediction model. The concentration of $PM_{2.5}$ is set as the output of the prediction model. Normalize the data to improve the numerical stability of the model. The normalized data is first used for model training and the predicted outputs for the validation set. The output is then back-normalized to obtain the final prediction of $PM_{2.5}$ concentration.

2.2 Model preparation

2.2.1 Random Forest

Random forest regression[6] is one of the more widely used methods in machine learning regression methods. Random forest regression method is mainly based on decision tree theory, through the integration of multiple decision tree regression models to predict the target variable, and finally

ISSN 2616-5775 Vol. 7, Issue 12: 24-35, DOI: 10.25236/AJCIS.2024.071204

summarize the results of a large number of decision tree regression models, so as to improve the accuracy of the regression model. Firstly, random forest regression constructs decision trees by randomly obtaining multiple sample subsets in the training set through bootstrap resampling; secondly, when training the decision trees, randomly selecting features from the feature set to train each decision tree, so that the prediction structure of each decision tree is not the same; then, regression prediction is performed on each decision tree without pruning, and all the regression predictions are averaged; furthermore, the regression model results are summarized to improve the accuracy of the regression model. In addition, Random Forest regression can also evaluate the importance of feature variables. Numerous studies have shown that the Random Forest machine learning algorithm can balance the error of unevenly distributed samples, provide a better fit to nonlinear data, and have a better tolerance for outliers and noise. The most important parameters affecting the effectiveness of the random forest regression model are the number of decision trees in the forest, the minimum number of samples of leaf nodes, the minimum number of samples of non-leaf nodes, the number of features of optimal splitting, the proportion of randomly selected samples, and the maximum depth. So this paper will be traversing the optimization of these six parameters, and the rest of the parameters are uniform default values. The specific algorithm of random forest regression(2) is as follows:

$$H(x)=\frac{1}{N}\sum_{i=1}^{N}\{H(x,\Phi_i)\} \tag{2}$$

$H(\mathrm{x})$ is the random forest regression result;

x is the independent variable;

Φ_{i} is an independent and identically distributed random vector for output based on x and Φ_{i};

N is the number of regression decision trees.

2.2.2 Support vector machine regression

The basic idea of the SVR model[7] is that the original input space, which is nonlinearly correlated with the predictor variables, is mapped onto a high-dimensional feature space by a nonlinear mapping function (kernel function) to obtain a model that is as suitable as possible to fit the samples of the training set. A common approach is to construct a loss function between the sample labels and the model predictions and determine the function model by minimizing the loss function. We can create a dataset considering output vectors. The goal of SVR is to find a multiple regression function to predict the desired output properties of an unknown object based on a given dataset S. The SVR model (3) is as follows:

$$f(x)=\sum_{i,j=1}^{N}(a_i-a_i^*)\left\langle\varphi(x_i),\varphi(x_j)\right\rangle+b \tag{3}$$

Lagrange multipliers for which a_{i} and $\mathrm{a}_{\mathrm{i}}^{*}$ satisfy the constraints;

$\varphi(x_i),\varphi(x_j)$ are nonlinear mapping functions;

b is the offset for the regression function $f(x)$.

2.2.3 LightGBM model

The LightGBM model[8], first proposed by the Microsoft team in 2017, is an improved gradient-boosting decision tree framework. Its basic idea is to linearly combine M weak regression trees into a strong regression tree. The combination formula (4) is as follows:

$$F(x)=\sum_{m=1}^{M}f_m(x) \tag{4}$$

$F(x)$ is the final output value;

$f_m(x)$ is the output value of the m^{th} weak regression tree.

ISSN 2616-5775 Vol. 7, Issue 12: 24-35, DOI: 10.25236/AJCIS.2024.071204

The main improvements to the Light GBM model include the histogram algorithm and the leaf-growth (leaf-wise) strategy with depth restriction. The histogram algorithm divides the continuous data into K integers and constructs a histogram of width K. The histogram is then traversed and the discretized values are accumulated as indexes. The discretized values are accumulated as indexes in the histogram during traversal, which in turn searches for the optimal decision tree split points. The leaf-wise strategy with depth restriction means that at each split, the leaf with the maximum gain is found for splitting and the cycle continues. At the same time, by limiting the depth of the tree as well as the number of leaves, the complexity of the model is reduced, and overfitting is prevented.

2.2.4 Model Evaluation Parameters

Mean Absolute Error (5) is a measure of the accuracy of a predictive model or estimation method. It indicates the average size of the difference between the predicted values and the actual observed values.

$$MAE = \frac{\sum_{i=1}^{n} | y_i - \widehat{y_i} |}{n} \tag{5}$$

$\widehat{y_i}$ is the $Predicted\ value$;

y_i is the Observed value;

n is the total number of observations.

Mean Squared Error (MSE) (6) is a commonly used measure of the difference between the predicted values of a model and the actual observed values to assess how well the model fits the given data. MSE is obtained by calculating the average of the squares of the differences between the predicted values and the actual observed values. It provides an indication of how far the expected value differs from the original value.

$$MSE = \frac{\sum_{i=1}^{n} (\widehat{y_i} - y_i)^2}{n} \tag{6}$$

$\widehat{y_i}$ is the $Predicted\ value$;

y_i is the Observed value;

n is the total number of observations.

R^2(7) is a statistic used in regression analysis to measure how well the model fits the data. The value of R^2 ranges between 0 and 1, with 0 indicating that the model does not account for any of the variability and 1 indicating that the model fits the data perfectly.

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \widehat{y_i})^2}{\sum_{i=1}^{n} (y_i - \overline{y})^2} \tag{7}$$

$\widehat{y_i}$ is the $Predicted\ value$;

y_i is the Observed value;

$\overline{y}$ is the average value;

n is the total number of observations.

ISSN 2616-5775 Vol. 7, Issue 12: 24-35, DOI: 10.25236/AJCIS.2024.071204

3. Research process and results

3.1 Model construction process and analysis

The construction of the three models needs to establish the corresponding parameters. To train the model parameters optimal can predict the corresponding optimal effect. We use the grid search to go to the parameters that affect the model effect of the grid search traversal to find the model optimal parameters.

The construction of the model I random forest model needs to adjust the number of decision trees in the forest, the minimum number of samples of leaf nodes, the minimum number of samples of non-leaf nodes, the number of features of optimal splitting, the proportion of randomly selected samples, and the maximum depth, which are the six parameters that mainly affect the prediction effect of the model. The optimal parameters of the model are obtained by performing a grid search on the training set as shown in Table 2 below:

Table 2: Random forest model parameterization

parameters	value
number of decision trees in the forest	700
minimum number of samples of leaf nodes	1
minimum number of samples of non-leaf nodes	4
the number of features of optimal splitting	4
proportion of randomly selected samples	100%
maximum depth	34

The construction of the model II support vector machine regression model requires the adjustment of the regularization parameter (C), the threshold of the insensitive loss function, the kernel function selection, and the kernel coefficients, which are the four parameters that mainly affect the prediction effect of the model. The optimal parameters of the model are obtained by performing a grid search on the training set as shown in Table 3 below:

Table 3: Support vector machine regression model parameterization

parameters	value
C	150
epsilon	10
kernel	RBF
Gamma	0.1

The construction of the model III LightGBM model requires adjusting five parameters that mainly affect the prediction effect of the model: the number of trees, the learning rate, the number of leaf nodes, the maximum depth of the tree, and the minimum number of samples per leaf node.

The optimal parameters of the model are obtained by performing a grid search on the training set as shown in Table 4 below:

Table 4: LightGBM model parameterization

parameters	value
number of trees	750
learning rate	0.09
number of leaf nodes	40
maximum depth of the tree	-1(limitless)
minimum number of samples per leaf node	20

3.2 Analysis of forecast results

The best model trained by grid search is used to predict the $PM_{2.5}$ concentration on the validation set and the results are obtained as shown below.

Observation of Figure 3 shows that the LightGBM model has the best prediction effect, the true value and the predicted value almost match, and the two line graphs have high overlap, while the SVR model and the RF model do not have much difference, but they are obviously weaker than the prediction effect of LightGBM model. The difference between the normalized data of the true value and the predicted

ISSN 2616-5775 Vol. 7, Issue 12: 24-35, DOI: 10.25236/AJCIS.2024.071204

value is relatively large.

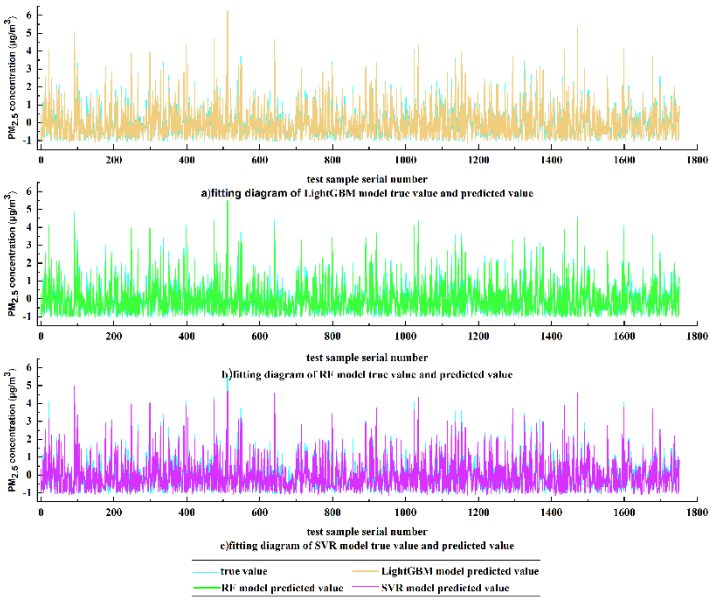

Figure 3: Fitting diagram of three models' true value and predicted value

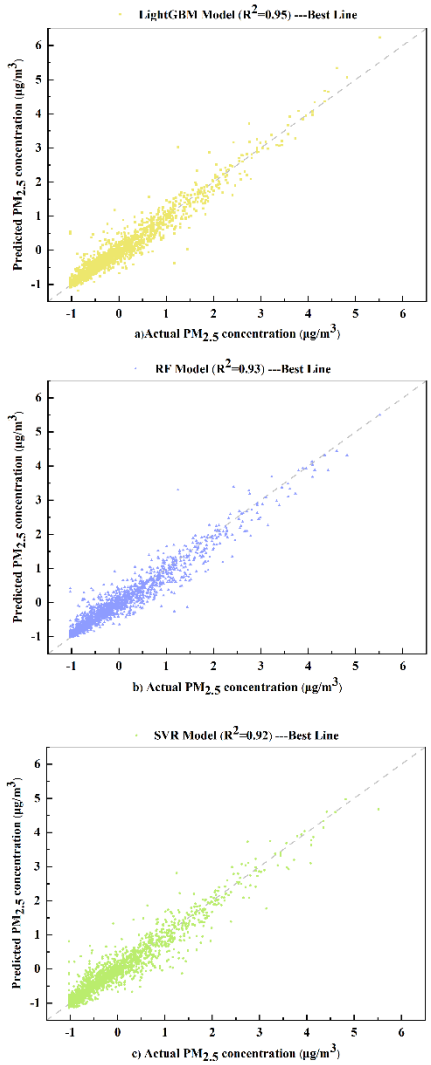

Figure 4: Scatter plot of $PM_{2.5}$ concentration

ISSN 2616-5775 Vol. 7, Issue 12: 24-35, DOI: 10.25236/AJCIS.2024.071204

As depicted in Figure 4, the scatter plots for all three models demonstrate a strong predictive performance. The horizontal axis represents the actual values, while the vertical axis represents the predicted values. Notably, the scatter plot for the LightGBM model aligns more closely with the line y=x, suggesting a superior fit compared to the other models.

The error line graph in Figure 5 clearly shows that the LightGBM model has superior predictive performance on the validation set compared to the RF and SVR models. The LightGBM model demonstrates a notably lower ratio of prediction error to the predicted value, suggesting a higher accuracy.

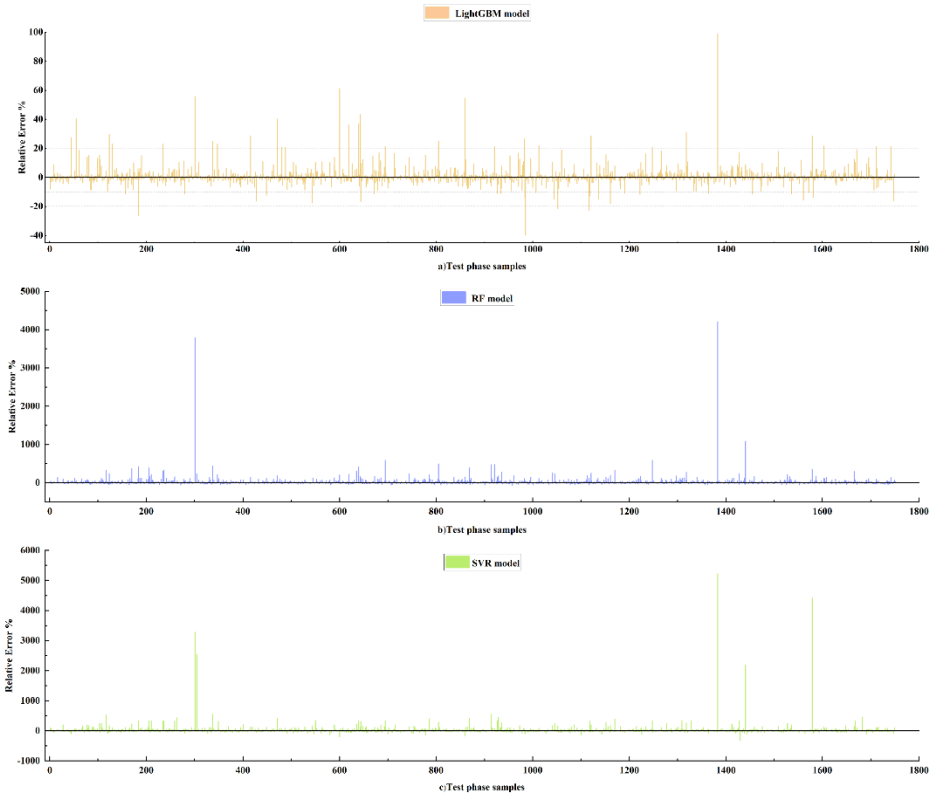

Figure 5: Error line graph

Table 5 counts the model evaluation parameters of each model prediction. It is obvious that the LightGBM model evaluation parameter R^2 is 0.9472, MAE is 0.14, MSE is 0.05, and the model evaluation parameters[9] are all significantly better than RF and SVR prediction models.

Table 5: Model evaluation parameters

Model	MAE	MSE	R^2
RF	0.16	0.06	0.9334
SVR	0.18	0.08	0.9185
Lightgbm	0.14	0.05	0.9472

Through the above comparison, it can be found that LightGBM as a modern machine learning integrated learning algorithm is significantly better than the traditional machine learning models RF, and SVR. The LightGBM model was chosen to predict $PM_{2.5}$ concentrations with the best results.

4. SHAP Framework interpretability analysis

4.1 Introduction to the SHAP Framework

SHAP[10] (Shapley Additive Explanations) is a method for interpreting the predictions of machine learning models, and its core principle is based on the concept of Shapley Value in game theory. The SHAP framework can be applied to the interpretation of various black-box models. It identifies the impact of each feature on the model prediction by decomposing the model prediction into a weighted sum of each feature value. The SHAP framework provides a comprehensive feature importance analysis and prediction interpretation by considering the combination and interaction of features, which is

ISSN 2616-5775 Vol. 7, Issue 12: 24-35, DOI: 10.25236/AJCIS.2024.071204

calculated as follows:

$$y_i = y_{base} + f(\lambda_{i1}) + f(\lambda_{i2}) + \cdots f(\lambda_{ik}) \tag{8}$$

Where λ_{ij} denotes the jth feature of the ith sample, y_i denotes the predicted value of the model for the sample, and y_{base} denotes the mean value of the predicted value of the model for all samples (i.e., the baseline value of the model). $f(\lambda_{ij})$ denotes the contribution of the jth feature in the ith sample to the final predicted result of the model, y_i. When $f(\lambda_{ij}) > 0$, it means that the feature has an enhancing effect on the prediction value, i.e., it makes a positive contribution; conversely, it means that the feature has a decreasing effect on the prediction value, i.e., it makes a negative contribution.

4.2 Feature Importance Visualization Analysis

Since LightGBM shows excellent prediction ability, this paper introduces the DeepExplainer explanatory framework in SHAP during the training of the $PM_{2.5}$ concentration prediction model in order to further improve the model accuracy. To study the correlation between $PM_{2.5}$ concentration variables and different features, the importance of different feature quantities is analyzed by calculating the absolute value of the SHAP value of each input feature, and sorting the mean value of the impact of $PM_{2.5}$ concentration prediction, as shown in Figure 6. The horizontal axis of the figure represents the mean of the absolute value of SHAP for the overall sample. According to the figure, it can be found that the concentration of PM_{10} is the feature that has the greatest influence on the results of $PM_{2.5}$ concentration predicted by the model, and the rest are CO, and No. The SHAP value of Precipitation is the smallest, and its contribution to the model decision is also the smallest, which is also related to the relatively concentrated data distribution of Precipitation.

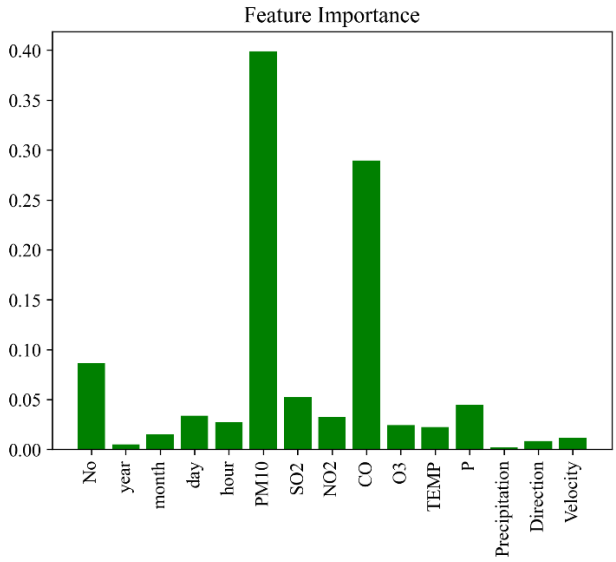

Figure 6: Contribution map

Figure 7 shows a summary plot of the SHAP values for the 15 features, where the horizontal coordinates indicate the SHAP values. The SHAP values of different features are distributed on both sides of the baseline in the center. The left-hand area represents the influencing factor that negatively affects the outcome, while the right-hand area represents the influencing factor that positively affects the outcome. Each sample in the dataset is run through the model and a point result is created for each feature. Each point color varies depending on the size of the feature value. The right eigenvalue fades from blue to red, indicating a gradual increase in the value of the independent variable.

By analyzing the feature SHAP value plot, the effects of PM_{10}, CO, and No on the model are more significant. Further analysis of the influence of each feature on the model output in the figure shows that for the concentration of PM_{10}, as its eigenvalue increases, it makes a positive contribution to the predicted results of $PM_{2.5}$ concentration, i.e., the larger the concentration of PM_{10}, the higher the concentration of $PM_{2.5}$. As the concentration of PM_{10} decreases, it basically makes a reverse contribution to the predicted results of $PM_{2.5}$ concentration, i.e., the smaller the concentration of PM_{10}, the lower the concentration of $PM_{2.5}$. This indicates that generally the higher the concentration of PM_{10} the higher the corresponding concentration of $PM_{2.5}$. The same is true for CO concentrations with large SHAP values. Therefore, in

ISSN 2616-5775 Vol. 7, Issue 12: 24-35, DOI: 10.25236/AJCIS.2024.071204

environmental management, the concentration of $PM_{2.5}$ can be reduced by encouraging the use of clean energy to reduce particulate matter PM_{10} as well as CO emissions from inadequate combustion treatment.

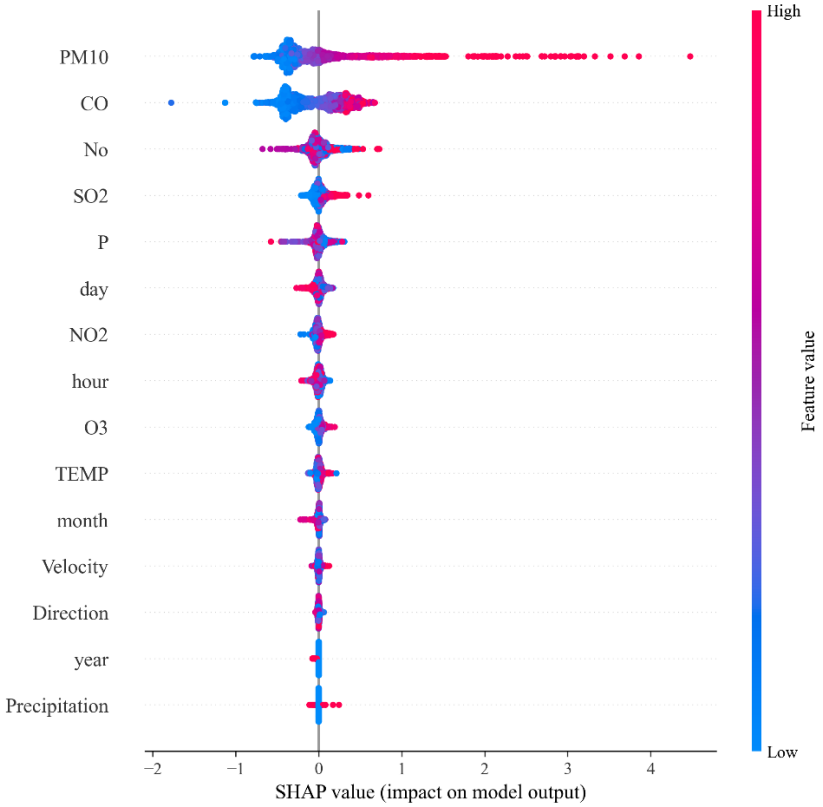

Figure7: SHAP value

The quantitative visualization shown in Figure 7 shows the contribution of each feature to the final score and ranks the most important positive and negative predicted impacts. The key insights into this prediction transform the complex decision problem, including all components of the LightGBM model, into a simple readable, and informative diagram where the unit of measurement is the target unit. A visualization of the model's first prediction is shown in Figure 8.

It is clear that the first $PM_{2.5}$ concentration predicted by the model is labeled 118.447μg/m³ and provides information about all the features (and their values) that influence this prediction. The main features that increase the $PM_{2.5}$ concentration: O_3, TEMP, CO, and PM_{10} are shown in red. While comparing this result with the initial feature importance analysis (Figure 6), it can be concluded that Precipitation, Direction, and Velocity do not have high enough values.

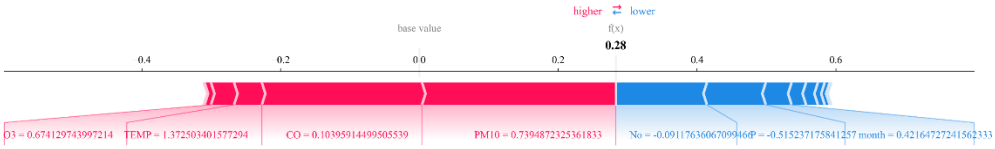

Figure 8: First $PM_{2.5}$ concentration prediction

Figure 9 presents a SHAP heatmap where darker colors indicate larger absolute SHAP values, signifying a greater impact on model predictions. The top section visualizes the model's predictions based on these values for 1752 data points from the validation set. It is observed that features such as PM_{10}, CO, No, SO_2, P, and day have both positive and negative contributions to the predictions. In contrast, the predictions for the validation set involving features like hour, O_3, and six other features nearly converge to zero. The SHAP values for each sample in the heatmap predominantly cluster in the white region near the 0 threshold, indicating minimal impact.

ISSN 2616-5775 Vol. 7, Issue 12: 24-35, DOI: 10.25236/AJCIS.2024.071204

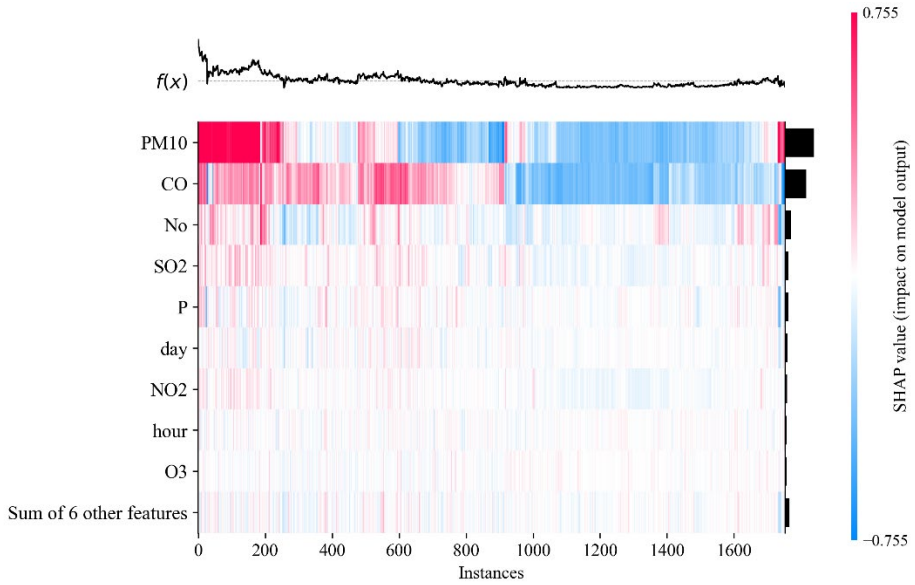

Figure 9: SHAP Heatmap

Figure 10 is a summary plot of the SHAP interaction values, where the seven features with the highest contribution are summarized in a matrix of SHAP interaction values. It can be found that the SHAP interaction value between each feature outside the diagonal is close to 0, which means that the marginal contribution of each of these seven features interacting with each other in the model's prediction is very small and the dependency is not very strong, indicating that their effects on the model's prediction are independent of each other.

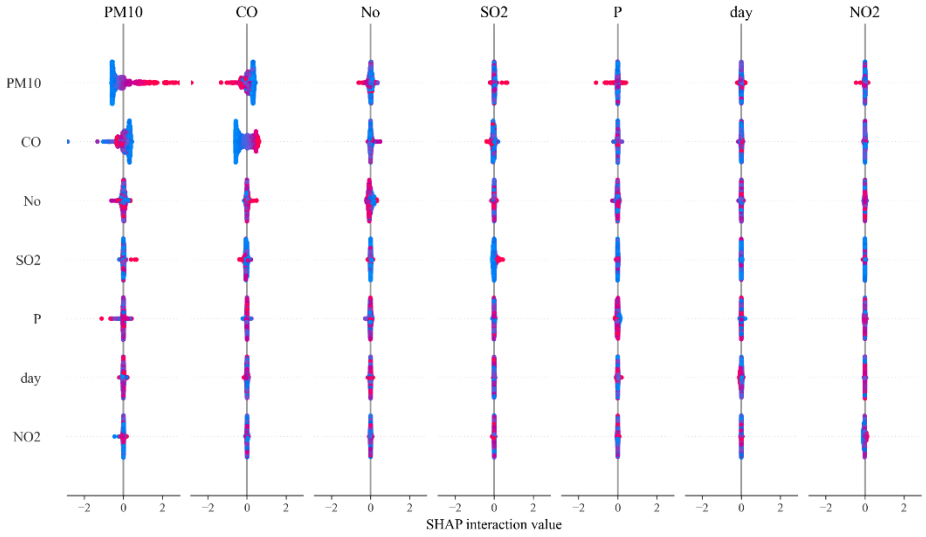

Figure 10: Summary of SHAP interaction value matrix

Through the above experiments, the SHAP-LightGBM model was established, and three features Precipitation, Direction, and Velocity were eliminated for re-prediction. It was found that the R^2 of the prediction model was improved from 0.9472 to 0.9501, and the elimination of the redundant features had a facilitating effect on the prediction of the model.

5. Conclusions and outlook

5.1 Conclusion

In this study, by applying the LightGBM model, the $PM_{2.5}$ concentration was successfully and effectively predicted. The experimental results show that the model performs well in predicting $PM_{2.5}$ concentration, especially with an R^2 value of 0.9472 on the validation set, which verifies the accuracy

ISSN 2616-5775 Vol. 7, Issue 12: 24-35, DOI: 10.25236/AJCIS.2024.071204

and stability of the model. Analyzing prediction models through SHAP framework, we found that the concentrations of CO, PM_{10}, and SO_2 were positively correlated with $PM_{2.5}$ concentration in most data points, while air pressure (P) and month were negatively correlated with $PM_{2.5}$ concentration. These findings provide a new perspective for $PM_{2.5}$ management. To effectively reduce PM2.5 concentration, it is necessary to control CO emissions. In addition, through feature selection, we identified three features, Precipitation, Direction, and Velocity, which contribute less to $PM_{2.5}$ prediction, and removed them from the data set to further optimize the model performance. The R^2 value on the validation set is improved to 0.9501.

5.2 Outlook

Although this study has achieved significant results in $PM_{2.5}$ concentration prediction, there are still some limitations and future research directions. First, the size of the current dataset limits the real-time nature of the model predictions. In order to improve the usefulness and real-time performance of the model, data collection capabilities need to be enhanced to support more frequent and extensive data collection. Second, future research could explore more features and variables to further improve the prediction accuracy and generalization ability of the model. In addition, the methodology of this study can be considered to be applied to the prediction of $PM_{2.5}$ concentration in other regions to verify the generalizability of the model. Finally, with the continuous advancement of machine learning technology, more advanced algorithms can be explored to further improve the performance of the prediction model. With these improvements and extensions, we expect to be able to provide a more accurate and real-time scientific basis for air quality management and $PM_{2.5}$ pollution control and make greater contributions to environmental protection and public health.

References

[1] Gehrmann S , Dernoncourt F , Li Y ,et al.Comparing Rule-Based and Deep Learning Models for Patient Phenotyping[J]. 2017.DOI:10.48550/arXiv.1703.08705.

[2] Aboonq M S , Alqahtani S A .Leveraging multivariate analysis and adjusted mutual information to improve stroke prediction and interpretability[J].Neurosciences, 2024, 29(3).DOI:10.17712/nsj. 2024.3. 20230100.

[3] Tao C , Jia M , Wang G ,et al.Time-sensitive prediction of NO_(2) concentration in China using an ensemble machine learning model from multi-source data[J].Journal of Environmental Sciences, 2024(3):30-40.DOI:10.1016/j.jes.2023.02.026.

[4] Luo X , Cheng Y , Wu C ,et al.[An interpretable machine learning-based prediction model for risk of death for patients with ischemic stroke in intensive care unit].[J]. Journal of Southern Medical University, 2023, 43 7:, 1241-1247. DOI:10.12122/j.issn.1673-4254.2023.07.21.

[5] Li X , Sun J , Chen X .Machine Learning-Based Prediction of High-Entropy Alloy Hardness: Design and Experimental Validation of Superior Hardness[J].Transactions of the Indian Institute of Metals, 2024, 77(11):3973-3981.DOI:10.1007/s12666-024-03450-5.

[6] Jessica S , Whitney W , Erin H ,et al.Machine learning-based prediction models in medical decision-making in kidney disease: patient, caregiver, and clinician perspectives on trust and appropriate use[J].Journal of the American Medical Informatics Association, 2024.DOI:10.1093/jamia/ocae255.

[7] Pan Y , Zhao D , Zhang X ,et al.Machine learning-Based model for prediction of Narcolepsy Type 1 in Patients with Obstructive Sleep Apnea with Excessive Daytime Sleepiness[J].Nature & Science of Sleep, 2024, 16.DOI:10.2147/NSS.S456903.

[8] Yan W , Shen Y , Chen S ,et al.Viscosity and melting temperature prediction of mold fluxes based on explainable machine learning and SHapley additive exPlanations[J].Journal of Non-Crystalline Solids: A Journal Devoted to Oxide, Halide, Chalcogenide and Metallic Glasses, Amorphous Semiconductors, Non-Crystalline Films, Glass-Ceramics and Glassy Composites, 2024:636.

[9] Shang Z , Chen Y , Lai D ,et al.A novel interpretability machine learning model for wind speed forecasting based on feature and sub-model selection[J].Expert Systems With Applications, 2024, 255.DOI:10.1016/j.eswa.2024.124560.

[10] Liu C , Lu Y , Feng J ,et al.Prediction and customized design of Curie temperature of Fe-based amorphous alloys based on interpretable machine learning[J].Materials Today Communications, 2024, 38.DOI:10.1016/j.mtcomm.2023.107667.