ISSN: 2688-9323

Planning-oriented Protected Area Construction Model -- The Case of China and India

Panzhen Wu[1], Huida Chen[2], Shun Zhang[3], Liangliang Yang[2], Yonghan Li[4] and Jingqi Dai[2, *]

[1]Henan University of Science and Technology, Luoyang, China

[2]Nanjing Institute of Technology, Nanjing, China

[3]Henan Agricultural University, Zhengzhou, China

[4]Anyang Institute of Technology, Anyang, China

Abstract

This paper examines the construction of planned ecological reserves and the evaluation of their environmental impact. It is of great practical importance for the world to achieve the goals of carbon neutrality, absorption of greenhouse gases and reduction of carbon emissions. The paper makes an urban plan of the proposed ecological reserve and rationalizes the evaluation of the positive impact of the construction of the reserve on the national and world environment. Based on the construction of the Seyhanba Ecological Reserve and the evaluation of its environmental impact, this paper is extended to cities in the country where ecological reserves have been established. Three machine learning algorithms, SVM, decision tree and KNN, were used to train the data and compare the accuracy, and then a model based on SVM algorithm for the proposed ecological reserve was established, and 29 areas where ecological reserves need to be established were finally identified through tuning and optimization. And according to the characteristics of ecological reserves, it is estimated that 196,000 hectares of reserves need to be established nationwide, and its expected to reduce the PM2.5 index of the country by 2548 units in ten years. The model was then extended to the world environment, using India as the study target and following the characteristics of ecological reserves in China. Using the same model, 37 areas requiring ecological reserves were finally identified through training of data and optimization of model tuning parameters, and it was estimated that 707,600 hectares of reserves need to be established in the country, which is expected to reduce the national PM2.5 index by 9,198.8 units in ten years.

Keywords

SVM Algorithm; Decision Tree; Scheme Planning.

1. Restatement of the Problem

1.1. Background of the Problem

With the impact of human activities on the global climate, the climate crisis is becoming increasingly widespread and accelerating global warming is an urgent issue for every country to address. The global average temperature is rising at an unprecedented rate, and China and the world are actively addressing climate change. China has announced to the world the goal of achieving carbon peaking and carbon neutrality, and has already achieved remarkable results through the principles of scientific and unified planning, local adaptation and rational utilization.

2. Basic Assumptions

1. The data collected is true and valid or within a reasonable range.
2. The characteristics of the established ecological reserves are fully consistent.

3. Establishment of the Proposed Ecological Reserve Model based on SVM Algorithm

3.1. Data Collection and Organization

We have collected monthly AQI, PM2.5 concentration and PM10 concentration indicators for some cities in the country in 2020, and the air quality levels can be classified as excellent, as well as mild, moderate, severe and serious pollution according to the size of the AQI value. One of the criteria for severe pollution is AQI value over 300, when it will cause low visibility and endanger the health of plants and animals. We first proposed cities that have experienced severe pollution as protected areas, with Anhui Maanshan, Xinjiang Hotan region and Xinjiang Kulle region.

We also collected cities that have established ecological reserves for wind and sand control, which were recorded as 1, and recorded as 0 for other cities in their provinces, indicating that other cities have not established reserves. The distribution of all types of ecological reserves that have been established in China is shown in the figure below. To facilitate the analysis of the data, the indicators of cities that have not experienced serious pollution are averaged over the whole year as the sample evaluation indicators.

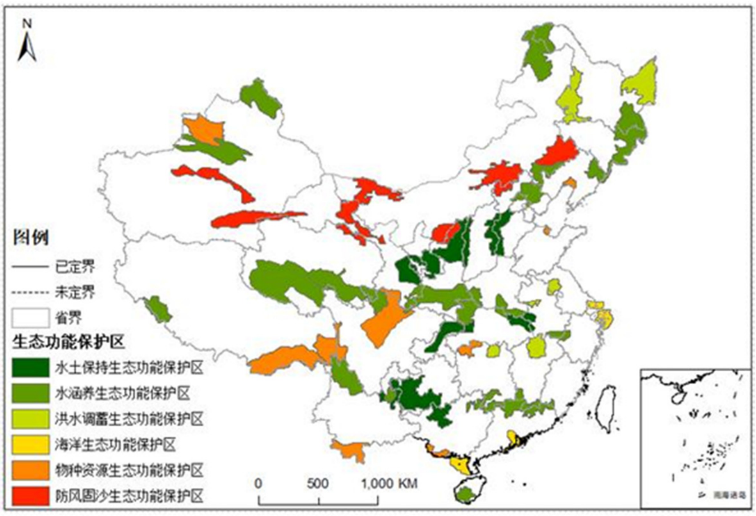

Figure 1. Distribution of ecological function reserves in China

The figure contains all types of ecological function reserves, and according to the question, we only take the ecological function reserve for wind and sand control as the object of study.

3.2. Modeling of the Proposed Ecological Reserve

Based on the data we already have, we can use the indicators of these cities to make predictions for cities in other provinces. We first build SVM, decision tree and KNN models to train the data separately and make predictions for other cities. By comparing the prediction results, the model with the highest accuracy rate can be used. Then, we calculate the F1 score and AUC value by the model and use grid search to adjust the reference, so as to optimize the model, and finally, we can predict the areas that need to establish ecological zones more accurately.

ISSN: 2688-9323

3.2.1. SVM Prediction Model Building

Support vector machine is a two-class classification model, the basic model is defined as a linear classifier with maximum interval on the feature space, and its learning strategy is interval maximization, which can eventually be translated into the solution of a convex quadratic programming problem.

step1: Hyperplane equation. Given some data points which belong to two different classes, a linear classifier needs to be found to classify these data into two classes. The learning goal of a linear classifier is to find a hyperplane in the n-dimensional data space, and the equation of this hyperplane can be expressed as follows.

$$\omega^T x + b = 0 \tag{1}$$

where x denotes the data point and y denotes the class (y takes 1 or -1, representing two different classes, respectively), so that the geometric interval from any sample point to the plane is $|\omega^T x + b|/||\omega||$. To find the partitioned hyperplane with the maximum interval, which is minimized to $|\omega||/2$, the model is then transformed into a convex quadratic programming problem with a constraint on the number of samples N as follows.

$$\min_{\omega,b} ||\omega||^2/2 \tag{2}$$

$$s.t.\, y_i(\omega x_i + b) - 1 \geq 0, i = 1,2, \dots, N \tag{3}$$

step2: Use the Lagrange multiplier method to transform the above problem into its dual problem, i.e., add N Lagrange multipliers. The relationship between the parameters in the original problemω and α in the dual problem are related as follows.

$$\min_{\alpha}\left(\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i\, y_j\left(x_i g x_j\right) - \sum_{i=1}^{N}\alpha_i\right) \tag{4}$$

$$s.t.\sum_{i=1}^{N}\alpha_i y_i = 0, \alpha_i \geq 0, i = 1,2, \dots, N \tag{5}$$

step3: Its required to meet the following KKT conditions.

$$\alpha_i \geq 0 \tag{6}$$

$$y_i f(x_i) - 1 \geq 0 \tag{7}$$

$$\alpha_i(y_i f(x_i) - 1) \geq 0 \tag{8}$$

Then it means that the sample points will not have any effect on the parameters ω and hyperplane of the final divided hyperplane, which itself is a point of the support vector and lies

ISSN: 2688-9323

on the maximum interval boundary. Therefore the final division hyperplane is only related to the support vector [1].

3.2.2. Decision Tree Prediction Modeling

The decision tree contains a root node, several internal nodes and several leaf nodes. The leaf nodes correspond to decision results, and each other node corresponds to an attribute test; the set of samples contained in each node is divided into sub-nodes according to the results of the attribute tests; the root node contains the full set of samples, and the path from the root node to each leaf node corresponds to a sequence of decision tests.

Step1: The goal of decision tree learning is to choose the optimal division attribute, for binary classification, we need to try to make the divided samples belong to the same category. Information entropy" is used to measure the purity of the features, and the information entropy of the sample set is.

$$Ent(D) = -\sum_{k=1}^{K} p_k \log_2 p_k \tag{9}$$

where D denotes the sample set,p_k denotes the proportion of samples in the kth category, and K is the total number of categories, i.e., 2. It can be found that the smaller the value of $Ent(D)$, the higher the purity of D.

Step2: Considering the different number of samples contained in different branch nodes, the branch nodes are given weights $|D^v|/|D|$, i.e., the more samples the greater the influence of the branch nodes, so that the "information gain" obtained from the division of the sample set D by the featureα can be calculated.

$$Gain(D, a) = Ent(D) - \sum_{v=1}^{V} (|D^v| / |D|) Ent(D^v) \tag{10}$$

The greater the information gain, the greater the "purity gain" obtained by using the featuresα to partition the dataset. Therefore, the information gain can be used for the selection of attributes for decision tree partitioning, i.e., for selecting the attributes with the highest information gain.

Step3: Divide each branch and continue recursively until each branch contains only one sample, at which time the current node is marked as a leaf node with the same category as the current sample.

3.2.3. KNN Model Building

In order to judge the category of the unknown sample, the distance between the unknown sample and all known samples is calculated using all samples of known categories as reference, from which the K known samples with the closest distance to the unknown sample are selected, and the unknown sample is grouped with the K nearest samples belonging to more categories according to the voting rule of majority rule of minority. It consists of the following three main steps.

Step1: Calculate the distance. (a) For a given sample to be classified, calculate its distance from each of the samples already classified.

Step2: Find neighbors. Circle the K classified samples that are closest to the sample to be classified as the nearest neighbors of the sample to be classified.

Step3: Autonomous classification. Decide which classification the sample to be classified should belong to based on the class to which most of the samples in these K nearest neighbors belong[2].

ISSN: 2688-9323

3.3. Selection of the Proposed Ecological Reserve Model

We have trained the samples by SVM, Tree and KNN algorithms respectively, and the results of the initial training will be represented by the accuracy, scatter plot and confusion matrix as follows.

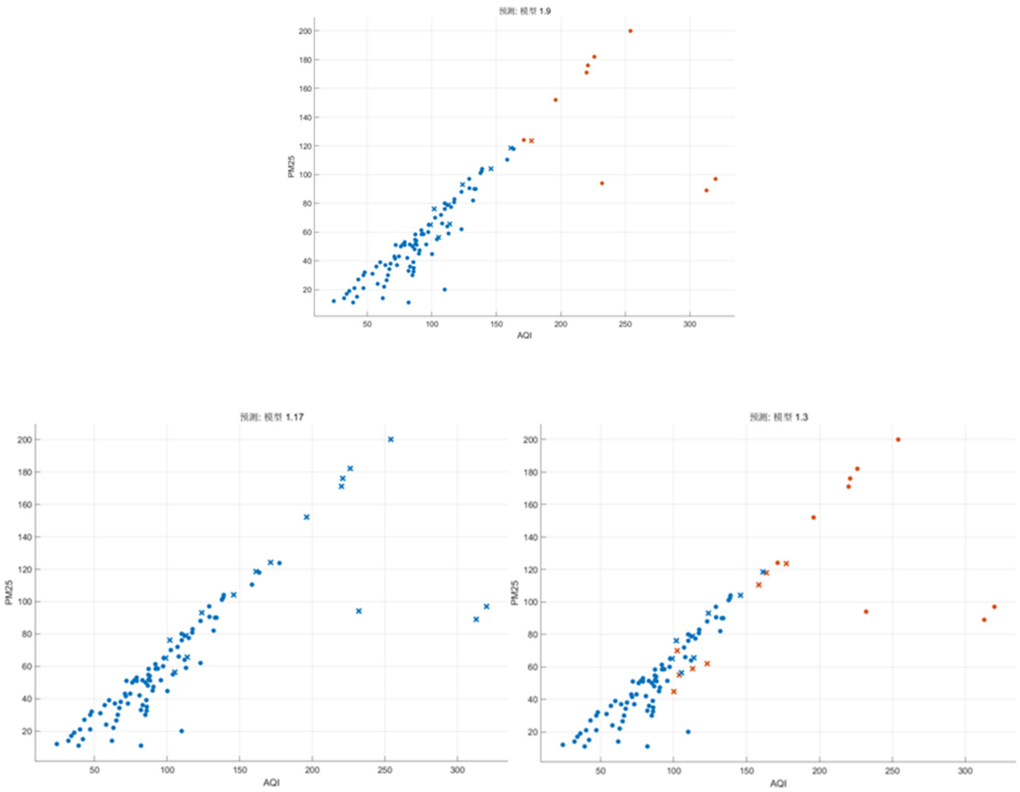

Figure 2. Scatter plot of SVM, Tree and KNN training results

The scatter plot shows the relationship between AQI and PM2.5 concentration index under different models. The blue scatter points indicate that no eco-region is proposed, and the dot and fork symbols represent the correct and incorrect training results, respectively. We observe that the training results of the KNN model do not have cities that need to be proposed as ecological zones, which is against the reality and the requirements of the question, so this model is discarded. The training results of the SVM and Tree models include whether or not an ecological zone is proposed, but the accuracy of the Tree model is relatively low.

The confusion matrix contains elements: the number of positive classes predicted as positive classes TP, the number of positive classes predicted as negative classes FN, the number of negative classes predicted as positive classes FP, and the number of negative classes predicted as negative classes FN. From this confusion matrix, the classification accuracy rate, the check-all rate, the check-accuracy rate and the F1 score can be calculated. The figure shows the true rate TPR and the false negative rate FNR, and the closer the TPR is to 1, the higher the training accuracy and the better the model effect. The comparison shows that the Tree model is trained with a lower TPR, which indicates a certain error rate[3].

And the accuracy of training by SVM, Tree and KNN models were 91.0%, 84.0% and 83.0%, respectively. In summary, we use the SVM model as the proposed ecological reserve model.

ISSN: 2688-9323

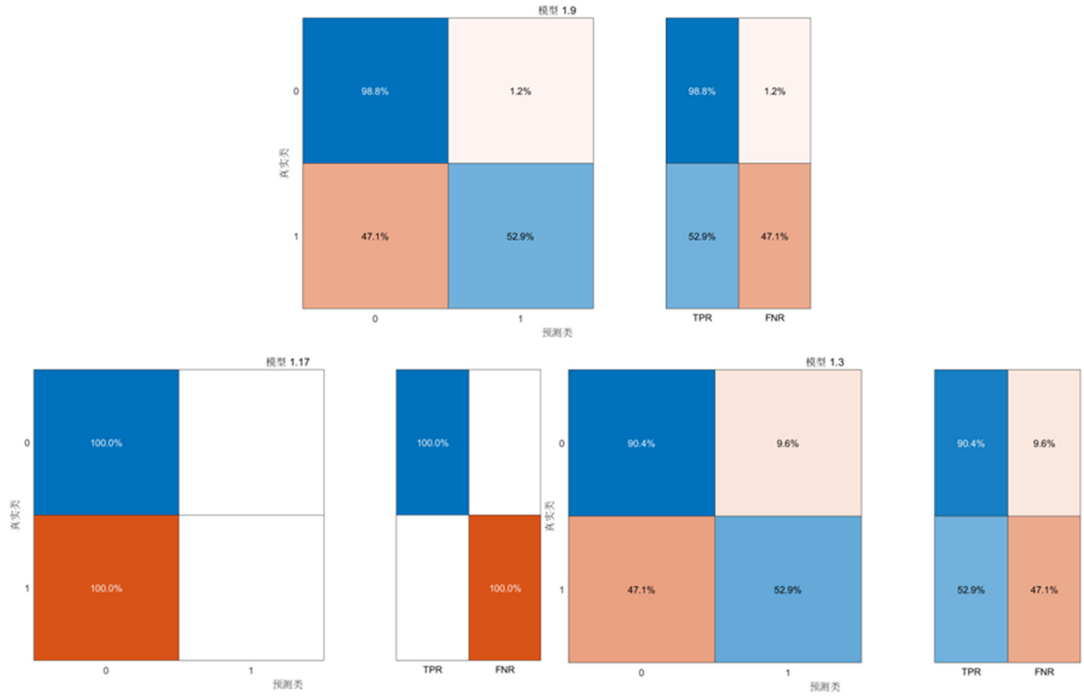

Figure 3. Confusion matrix of SVM, Tree and KNN training results

3.4. Solution and Optimization of the Proposed Ecological Reserve Model

We started with preliminary forecasts for cities in other provinces using the SVM model, and some of the results of the preliminary forecasts are shown in the following table (see Appendix for the complete table).

Table 1. Preliminary training results of SVM model (partial)

Province	City	Protected Areas	Province	City	Protected Areas
Hubei	Enshi Prefecture	2	Sichuan	Suining	0
Sichuan	Meishan	0	Sichuan	Ziyang	0
Sichuan	Neijiang	0	Sichuan	Guang'an	2
Guizhou	Bijie	2	Shandong	Heze	0
Guizhou	Liupanshui	2	Shandong	Texas	0
Sichuan	Dazhou	0	Shandong	Liaocheng	2
Tibet	Chamdo region	0	Sichuan	Bazhong	2

The number "0" indicates that the city is not proposed as an ecological zone, and the number "2" indicates that the city will be proposed as an ecological reserve for wind and sand control.

Next, we need to perform parametric optimization of the model. According to the basic principle of SVM model, the parameter KernelFunction can be chosen as linear, polynomial and gaussian kernel functions. We use the grid search method to train the model several times, and the results are measured by ROC curves and AUC values to measure the correctness of the parameters. The ROC curve, which is based on a series of different classification thresholds, is plotted with the true class rate TPR as the vertical coordinate and the false positive class rate FPR as the horizontal coordinate. the AUC is defined as the area enclosed by the ROC curve and the lower axis, and the range of the AUC lies between [0,1], the larger the AUC the better the classification of the model. The ROC curves of different models are plotted in the same graph, and the curve closest to the top left corner has the highest AUC value, which represents the best classification of the model.

ISSN: 2688-9323

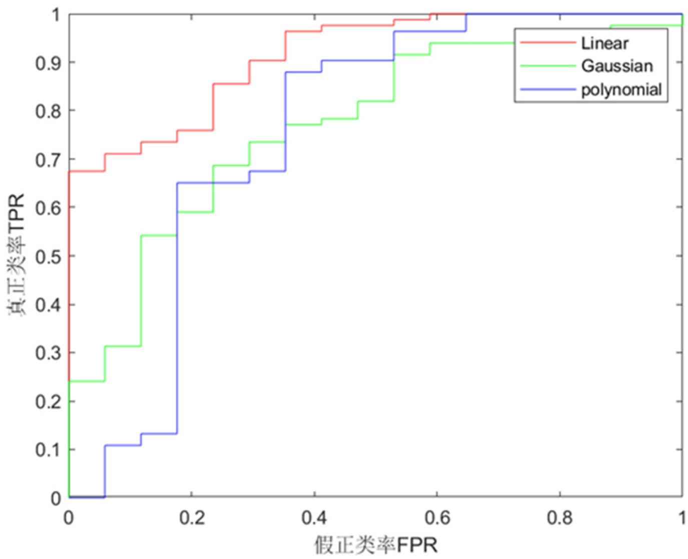

Figure 4. ROC curves of training results with three parameters

We can see that the curve represented by the linear kernel function is closest to the upper left corner of the plane, while the curves represented by the Gaussian kernel function and the polynomial kernel function are entangled with each other and located at the lower right. The calculated AUC values are 0.9142, 0.7690, and 0.7548, respectively, which shows that the linear kernel function works best. We then re-performed the prediction using this parameter, and the partial results of the prediction are shown in the following table[4] (see Appendix for the complete table).

Table 2. Training results after tuning (partial)

Province	City	Protected Areas	Province	City	Protected Areas
Heilongjiang	Hegang	2	Shandong	Linyi	0
Hebei	Shijiazhuang	0	Ningxia	Zhongwei	0
Sichuan	Suining	0	Shandong	Zaozhuang	0
Hebei	Hengshui	0	Hebei	Tangshan	0
Sichuan	Ziyang	0	Shandong	Zibo	0
Sichuan	Guang'an	2	Ningxia	Guyuan	0
Shandong	Heze	0	Heilongjiang	Qitaihe	0

The result of this prediction is whether the cities will be proposed as ecological function protection zones for wind and sand control, "0" means not to be proposed for the time being, "2" means to be built soon. Therefore, we finally identified 29 cities as proposed ecological zones.

By consulting the information, we know that each 10,000 hectares of ecological function reserve for wind and sand control can reduce the PM2.5 concentration by 9 to 17 units in a year, and for the sake of calculation, the average value of 13 can be taken. If we need to control the AQI value of the above cities in the best range within ten years, i.e., the AQI value is below 50, so we can estimate the scale of ecological reserve to be established in these cities, the cities and their estimates The results are shown in the following table.

Where the area unit is 10,000 hectares. The total size of the proposed eco-district plan cities amounts to 196,000 hectares, which is expected to reduce the PM2.5 index of the country by 2,548 units in ten years, and is of great significance to China's efforts and goals to achieve carbon neutrality and carbon peaking.

ISSN: 2688-9323

Table 3. Cities of the proposed eco-district plan and their sizes

Province	City	Area	Province	City	Area	Province	City	Area
Anhui	Ma On Shan	3.462	Hunan	Yiyang	0.538	Guangxi	Hezhou	0.423
Hubei	Enshi Prefecture	1.454	Heilongjiang	Jiamusi	0.531	Hunan	Changsha	0.418
Guizhou	Bijie	0.992	Hunan	Yongzhou	0.531	Guizhou	Qianxinan	0.385
Guizhou	Liupanshui	0.992	Hubei	Suizhou	0.508	Zhejiang	Jinhua	0.385
Heilongjiang	Hegang	0.900	Hunan	Huaihua	0.485	Yunnan	Nujiang Prefecture	0.377
Sichuan	Guang'an	0.800	Yunnan	Baoshan	0.477	Yunnan	Wenshan Prefecture	0.377
Shandong	Liaocheng	0.733	Hunan	Loudi	0.462	Guangxi	Hechi	0.377
Sichuan	Bazhong	0.723	Jiangsu	Taizhou	0.448	Yunnan	Nujiang Prefecture	0.377
Yunnan	Zhaotong	0.615	Heilongjiang	Shuangyashan	0.438	Yunnan	Wenshan Prefecture	0.377
Jiangxi	Yichun	0.585	Guangxi	Baise	0.431			

4. Re-application of the Proposed Ecological Reserve Model based on SVM Algorithm

4.1. Data Collection and Organization

In this question, we have collected the average AQI, PM2.5 concentration and PM10 concentration indicators for the year 2020 for 40 cities in the country of India as the study subject. Since the indicators collected in this question are the same as the national indicators collected in the third question, the two sets of data can be compared. To facilitate the analysis of air quality in India, we also collected the version of India that visualizes the AQI values, as follows.

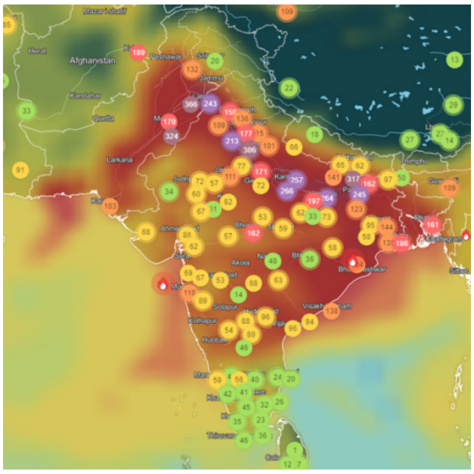

Figure 5. AQI visualization of the Indian version

The values in the graph represent the average AQI values for 2020, and the location of the values indicates the AQI values for the corresponding cities in India. The larger the value, the darker the color of the disk it is located in, indicating poorer air quality. Therefore, the figure shows that the air quality in the south of India is good, but the air quality in its central, northern, eastern and most of the southeast is extremely poor and widely distributed, so the Indian country needs to address the air quality issue urgently.

4.2. Modeling of Proposed Ecological Reserves in Indian Countries

Since the collected indicators for some parts of China and India are the same, they can be compared and unified for analysis. The AQI is the main measure of air quality for both countries, and the topography and climate of some parts of the Indian country are similar to those of China, so we can follow the characteristics of Chinese ecological reserves for India, i.e., the same amount of air purified per unit in ecological reserves and the same density of reserves. Therefore, we can embed this question into the model of the third question, and train the SVM model on the sample of established ecological reserves in China, and then predict the sample of Indian regions, and then we can derive the number and scale of ecological reserves to be proposed in the Indian country.

4.3. Solution of the Proposed Ecological Reserve Model for Indian Countries

We embedded this question into the model in the third question by training a sample of established ecological reserves in China and then applying the training results to the predictions for this sample of 40 Indian regions. We measure the precision of the training results by means of ROC curves.

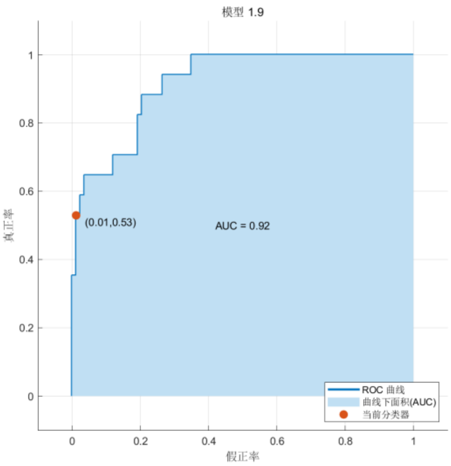

Figure 6. ROC curve of training under SVM model

It can be seen that the ROC curve is close to the upper left corner of the coordinates and the area enclosed with the axis below is the AUC value of 0.92 and the accuracy of training is 89%, then it means that the model has a good classification effect. The predicted results are now put into the following table.

Table 4. Results of model predictions

Indian Cities	Protected Area	Indian Cities	Protected Area	Indian Cities	Protected Area	Indian Cities	Protected Area
Agra	1.185	Faridabad	5.792	Jaipur	0.892	Manesar	1.477
Amritsar	0.715	Fatehabad	0.985	Jind	1.762	Meerut	1.146
Baghpat	1.562	Ghaziabad	2.115	Jodhpur	0.815	Narnaul	1.208
Bahadurgrah	1.938	Greater Noida	1.946	Kaithal	0.900	Noida	1.915
Ballabgarh	1.077	Gurgaon	1.423	karnal	7.300	Pali	0.669
Bhiwani	1.269	Gurugram	2.000	Kurukshetra	1.292	Palwal	0.600
Bulandshahr	2.100	Haldia	6.031	Lucknow	2.008	Rohtak	1.723
Charkhi Dadri	2.846	Hapur	1.246	Mandideep	2.100	Rupnagar	3.423
Delhi	3.046	Hisar	1.800	Mandikhear	0.231	Sirsa	0.862
Dharuhear	1.362						

ISSN: 2688-9323

The table shows the predicted results for a sample of 40 cities in India, where the number "0" indicates that the city does not need to be built as an ecological zone, and the number "2" indicates that the city will be built as an ecological reserve for wind and sand control. The prediction results are basically consistent with the distribution of AQI values in the Indian landscape. It can be seen that cities in the central, northern, eastern and southeastern regions of India have higher AQI values and almost all of them need to build protected areas, while only a few cities in the west and south do not need to build protected areas for the time being. We continue to use the characteristics of protected areas in China, i.e., each 10,000 hectares of protected area for wind and sand control can reduce PM2.5 concentrations by an average of 13 units in a year. Again, if the AQI values of the above cities need to be controlled in the optimal range of less than 50 within 10 years, we can estimate the size of the ecological reserves needed in these cities, and the cities and their estimates are shown in the following table.

Table 5. Cities of the proposed eco-district plan and their sizes

Indian Cities	Protected Area Area	Indian Cities	Protected Area Area	Indian Cities	Protected Area Area
Agra	1.185	Faridabad	5.792	Jaipur	0.892
Amritsar	0.715	Fatehabad	0.985	Jind	1.762
Baghpat	1.562	Ghaziabad	2.115	Jodhpur	0.815
Bahadurgrah	1.938	Greater Noida	1.946	Kaithal	0.900
Ballabgarh	1.077	Gurgaon	1.423	karnal	7.300
Bhiwani	1.269	Gurugram	2.000	Kurukshetra	1.292
Bulandshahr	2.100	Haldia	6.031	Lucknow	2.008
Charkhi Dadri	2.846	Hapur	1.246	Mandideep	2.100
Delhi	3.046	Hisar	1.800	Mandikhear	0.231
Dharuhear	1.362				

Where the area unit is million hectares. As can be seen from the table, the Indian state has more cities that need to build eco-districts and their scale is relatively large, some cities even need to build more than 50,000 hectares, which shows that their air quality is very poor. The total size of the cities of the proposed eco-district program of the Indian state amounts to 707,600 hectares, which is expected to reduce the PM2.5 index of the country by 9,198.8 units in ten years, and has a profound impact on the efforts and goals of carbon neutrality, greenhouse gas absorption and carbon emission reduction worldwide.

References

[1] ZHANG Zhenyu, PAN Ziyu, ZHANG Pengfei. A method of selecting coal price impact indicators based on game theory synthesis power[J]. Information Technology Research, 47(1):5.

[2] Zhang M.X., Han D., Zhao T. M.. A comprehensive evaluation method of ship type based on game theory-TOPSIS method [J/OL]. Applied Science and Technology:1-7 [2020-11-06].

[3] Mitchell R, Frank E. Accelerating the XGBoost algorithm using GPU computing[J]. PeerJ Computer Science, 2017, 3(1): e127.

[4] Jiang Jinwen, Liu Weiguang. Application of XGBoost algorithm in manufacturing quality prediction [J]. Intelligent computers and applications, 2017, 7(6):3.