

# Research and analysis of intercalated meltblown nonwovens based on machine learning

Jingwen Liang<sup>\*</sup>, Xiner Huang, Huibing Wu

School of Economics and Finance, Zhanjiang University of Science and Technology, Zhanjiang, China

<sup>\*</sup> Corresponding Author Email: 1165765983@qq.com

**Abstract.** This paper explores the relationship and change law between process parameters, structural variables and product performance, and through seeking the optimal combination of process parameters, the filtration efficiency of the product is as high as possible and the filtration resistance is as small as possible. Firstly, the correlation between structural variables and product performance is analyzed by Spearman correlation coefficient method after the relevant data are processed by factor analysis method. Secondly, linear regression, random forest regression and lightGMB regression models are used to judge the relationship between process parameters and structural variables. Finally, the relationship between structural variables and product performance can be obtained by using multidimensional scaling transform (MDS). The results showed that the product filtration efficiency was the highest when the acceptance efficiency distance was 20m and the hot air velocity was 800km/h.

**Keywords:** factor analysis; Random forest regression; Multidimensional scaling transform (MDS)

## 1. Introduction

There are many process parameters for the preparation of intercalated meltblown nonwovens, and there are interactive effects between the parameters. In addition, the intercalated air flow is more complicated. Therefore, the structural variables (thickness, porosity and compression resilience) are determined by the process parameters (receiving distance and hot air velocity), and the research on the final product performance (filtration resistance, filtration efficiency and permeability) determined by the structural variables becomes more complicated. If the relationship models between process parameters and structural variables, structural variables and product performance can be established, it will help to provide a theoretical basis for the establishment of product performance control mechanism.

The factor analysis method is used to reduce the structural variables and product performance to one-dimensional variables. The non-intercalated samples and intercalated samples of structural variables and product performance are analyzed to obtain their change trends [1-3]. Then use Spearman correlation coefficient method to obtain the correlation between the intercalation value and the relevant parameters of structural variables and product performance, and it can be concluded that the intercalation rate has the greatest impact on the thickness, compressive resilience and porosity, and the others are slightly less, with more indirect effects.

Principal component analysis (PCA) was used to reduce the dimension of data indicators, and the relationship between process parameters and structural variables was judged according to the resolution rate. When the receiving distance is the same, the hot air velocity has a positive correlation with the thickness, porosity and compression resilience of the mask; When the hot air velocity is constant, the receiving distance has a negative correlation with the thickness and porosity of the mask. On this basis, three models of linear regression, random forest regression and lightGMB regression are used to predict "thickness", "porosity" and "compression resilience" respectively [4-7]. After comparing the errors of the three models, the random forest regression model is finally selected to calculate the predicted values of "thickness", "porosity" and "compression resilience" in the specified process parameter combination.

Multidimensional scaling transformation (MDS) is used to reduce the dimension of structural variables, product performance, process parameters and other data. By comparing the reduced dimension data, the relationship between structural variables and product performance, as well as the relationship between structural variables and product performance can be obtained. Then, the random forest regression method is used to obtain the prediction data, and the maximum value is found in the prediction data. When the acceptance distance is 20 and the hot air speed is 800, the filtration efficiency of the product will reach the highest.

## 2. Change law of structural variables and product performance

### 2.1. Factor analysis [8]

The factor analysis model of  $p$ -dimensional variable  $x = [x_1, \dots, x_i, \dots, x_p]^T$  is established as follows:

$$x = Af + \varepsilon \quad (1)$$

Where  $f = [f_1, f_2, \dots, f_m]^T$  is the extracted common factor vector, which represents  $m$  ( $m < p$ ) mutually independent common influencing factors that cannot be directly observed but exist objectively in the original variable;  $A = (aik)$  is the factor load matrix, and the matrix element  $aik$  is the load of the variable  $x_i$  on the common factor  $f_k$ , which reflects the correlation coefficient of the two. The greater the absolute value, the higher the correlation; The key to establishing factor analysis model for multi-dimensional variable  $x$  is to solve factor load matrix  $A$  and common factor vector  $f$ . the steps are as follows:

In order to eliminate the influence of different dimensions of variables, the sample  $X = [x_1, x_2, \dots, x_n]$  with  $p$ -dimensional variables was standardized. After normalization, the mean value of each variable is 0 and the variance is 1. Its elements are:

$$x_{ij} = \frac{(x_{ij} - \frac{1}{n} \sum_{j=1}^n x_{ij})}{\sqrt{\frac{1}{n-1} \sum_{j=1}^n (x_{ij} - \frac{1}{n} \sum_{j=1}^n x_{ij})^2}} \quad (2)$$

Find the covariance matrix  $S$  of the sample, and its elements are:

$$s_{ij} = \frac{1}{n-1} \sum_{k=1}^n x_{ik} x_{jk} \quad (3)$$

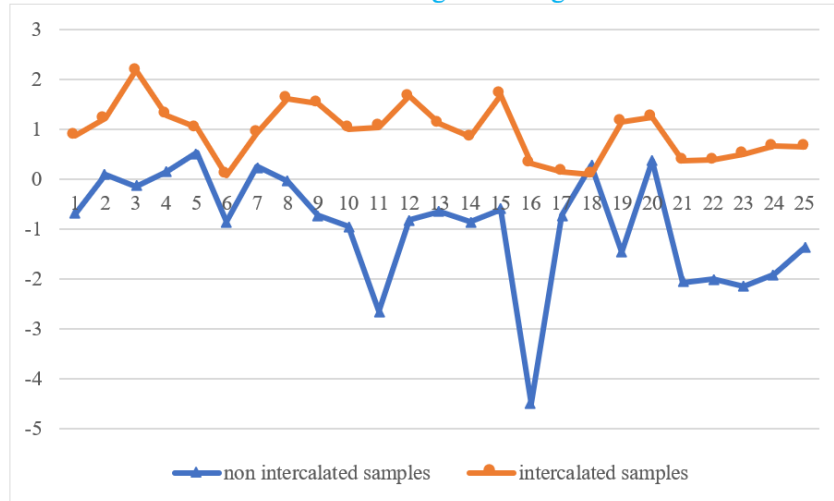
Perform eigenvalue decomposition on the sample covariance matrix  $S$  to obtain  $p$  eigenvalues:

$$\hat{A} = [\sqrt{\lambda_1} \gamma_1, \sqrt{\lambda_2} \gamma_2, \dots, \sqrt{\lambda_m} \gamma_m] \quad (4)$$

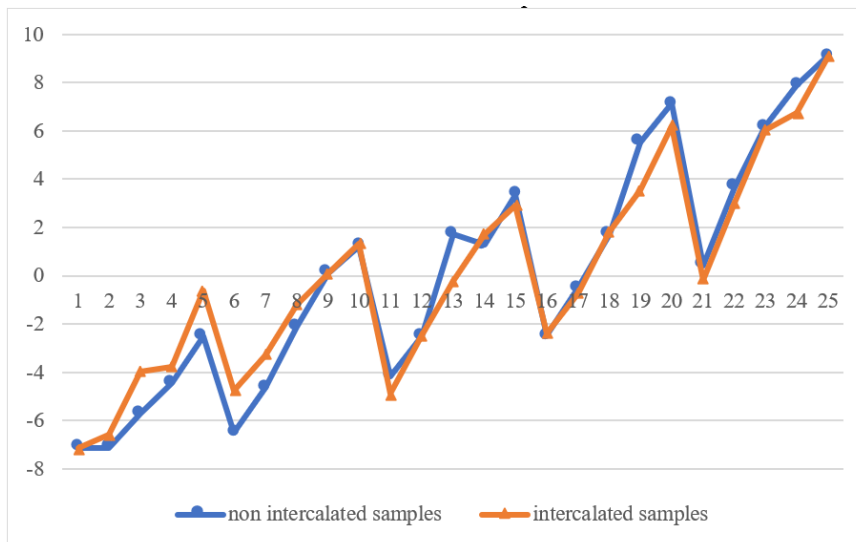
The parameter  $m$  is determined by the cumulative variance contribution rate of the common factor:

$$m = \arg \min_m \left( \frac{\sum_{i=1}^m \lambda_i}{\sum_{i=1}^p \lambda_i} \geq r \right) \quad (5)$$

The calculation results are shown in Fig. 1 and Fig. 2.



**Fig. 1** Comparison of change trends of structural variables between non intercalated samples and intercalated samples



**Fig. 2** Comparison of change trend of product performance between non intercalated samples and intercalated samples

**2.2. Effect of intercalation rate on process parameters and structural variables [9,10]**

Spearman correlation coefficient method (Table 1) was used.

X and Y are defined as two groups of data, and the Spearman correlation coefficient is:

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)} \tag{6}$$

Where  $d_i$  is the rank difference between  $y_i$  and  $x_i$ .

**Table 1.** Find the grade difference of X and Y and the square of the grade difference

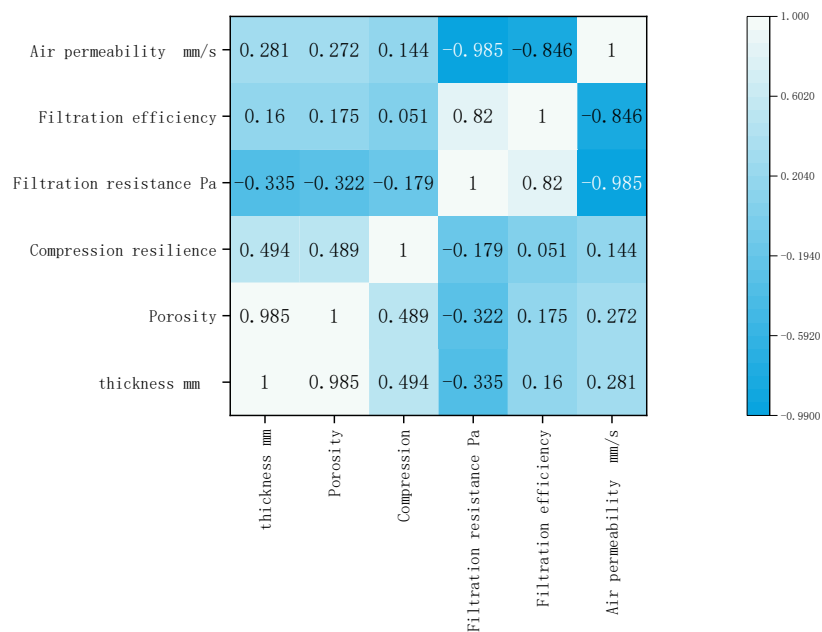
| X | Y  | Grade of X | Grade of Y | Grade difference | Square of grade difference |
|---|----|------------|------------|------------------|----------------------------|
| 3 | 5  | 2          | 1          | 1                | 1                          |
| 8 | 10 | 5          | 4.5        | 0.5              | 0.25                       |
| 4 | 8  | 3          | 3          | 0                | 0                          |
| 7 | 10 | 4          | 4.5        | -0.5             | 0.25                       |
| 2 | 6  | 1          | 2          | -1               | 1                          |

**2.3. Result analysis**

**Table 2.** Correlation of intercalation value with structural variables and relevant parameters of product performance

| Intercalation rate | thickness mm | Compression resilience | Porosity | Filtration resistance Pa | Filtration efficiency | Air permeability mm/s |
|--------------------|--------------|------------------------|----------|--------------------------|-----------------------|-----------------------|
| 1.000              | 0.782        | 0.375                  | 0.806    | -0.136                   | 0.330                 | 0.123                 |
| 0.782              | 1.000        | 0.494                  | 0.985    | 0.335                    | 0.160                 | 0.281                 |
| 0.375              | 0.494        | 1.000                  | 0.489    | -0.179                   | 0.051                 | 0.144                 |
| 0.806              | 0.985        | 0.489                  | 1.000    | 0.322                    | 0.175                 | 0.272                 |
| -0.136(0.346)      | 0.335        | 0.179                  | 0.322    | 1.000                    | 0.820                 | 0.985                 |
| 0.330              | 0.160        | 0.051                  | 0.175    | 0.820                    | 1.000                 | 0.846                 |
| 0.123              | 0.281        | 0.144                  | 0.272    | 0.985                    | 0.846                 | 1.000                 |

From the above table, it can be seen that the intercalation rate is not significantly correlated with the filtration assistance, filtration efficiency and permeability, and the filtration resistance is correlated with the thickness, the filtration efficiency is correlated with the intercalation rate, and the permeability is correlated with the thickness and porosity. Structural variables determine the final product performance, and the intercalation rate has a significant correlation with the relevant indicators of structural variables, which indirectly affects the relevant indicators of product performance. The thermodynamic diagram of correlation coefficient shows the value of correlation coefficient, which is mainly expressed by color depth (Fig. 3).



**Fig. 3** Thermodynamic diagram of correlation coefficient

**3. Analyze the relationship between the intercalation rate and structural variables and product performance**

In this paper, the principal component analysis method is used to reduce the "process parameters" and "structural variables" to one-dimensional variables, and then comparative analysis is carried out to explore the relationship between the two. Then, the data are analyzed by using linear regression model, random forest regression model and lightGBM regression model, and the model with the lowest error value is selected for solution.

### 3.1. Principal component analysis (PCA)

Standardize the range of structured indicators, and use standardization to convert data into comparable scales. The zero mean normalization formula is as follows:

$$z = \frac{x - \mu}{\sigma} \tag{7}$$

The eigenvectors and eigenvalues of the covariance matrix are calculated to identify the principal components. The principal component of the data is determined by calculating the eigenvector and eigenvalue of the covariance matrix.

$$\begin{cases} \max\{\omega^T \sum \omega\} \\ \text{s.t. } \omega^T \omega = 1 \end{cases} \tag{8}$$

$$D(x) = \omega^T \sum \omega^T = \lambda \tag{9}$$

Create eigenvectors to determine the retention of principal components.

$$\eta = \sqrt{\frac{\sum_{i=1}^d \lambda_i^2}{\sum_{i=1}^n \lambda_i^2}} \tag{10}$$

### 3.2. Comparison of process parameters and structural variables

According to the principal component analysis method, we can calculate that the analytical rates of process parameters and structural variables are 0.937 and 0.998 respectively. In general, the higher the resolution rate, the more important the principal component is, and the higher the weight proportion is (Fig. 4 and Fig. 5).

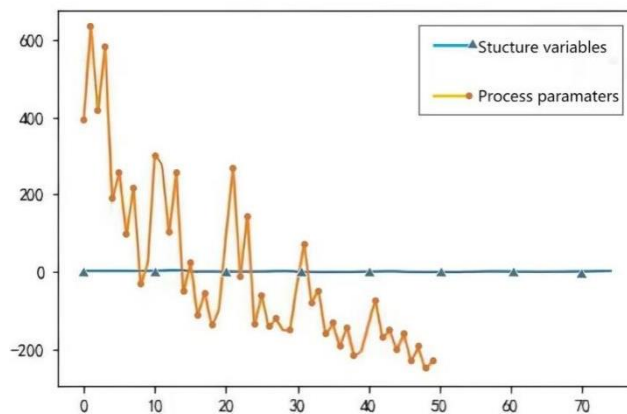


Fig. 4 Comparison of structural variables and process parameters

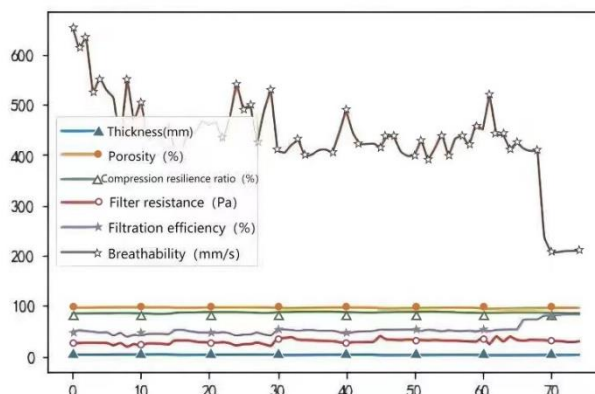


Fig. 5 Comparison of internal indexes of structural variables & process parameters

When the receiving distance is the same, the larger the hot air speed, the thicker the mask thickness, the larger the ratio of void to material volume in the meltblown nonwoven material, the better the resilience against compression, and the stronger the compression resistance. When the hot air velocity is constant, the closer the receiving distance, the thinner the thickness of the mask, the lower the porosity, and the compression resilience has no obvious change trend.

**3.3. Establishment of linear regression model**

$$y = \omega_0 + x_1\omega_1 + x_2\omega_2 + x_3\omega_3 + x_4\omega_4 + x_5\omega_5 + \dots + x_n\omega_n \tag{11}$$

**3.4. Random forest regression model**

$$Gini(D) = 1 - \sum_{k=1}^{|y|} p_k^2 \tag{12}$$

$$Gini(D, a) = \sum_{v=1}^v \frac{|D^v|}{|D|} Gini(D^v) \tag{13}$$

**3.5. LightGMB regression model [11]**

Linear regression model, random forest regression model and lightGMB regression model are used to analyze the data. The error results are shown in Table 3:

**Table 3.** Error comparison of regression model application results

|                                | Thickness (mm) | Porosity | Compression resilience |
|--------------------------------|----------------|----------|------------------------|
| Linear regression model        | 2.2352         | 0.3393   | 0.8523                 |
| Random forest regression model | 0.9565         | 0.2831   | 0.2337                 |
| LightGMB                       | 6.1874         | 0.3321   | 1.0887                 |

Model parameters are shown in Table 4:

**Table 4.** Parameters of random forest

|   | Thickness (mm) | Porosity | Compression resilience |
|---|----------------|----------|------------------------|
| Training time   | 0.096s         | 0.121s   | 0.129s                 |
| Data segmentation                                     | 0.7            | 0.7      | 0.7                    |
| Evaluation criteria for node splitting                | Mse            | Mse      | Mse                    |
| Maximum feature ratio considered during division      | None           | None     | None                   |
| Minimum number of samples for internal node splitting | 2              | 2        | 2                      |
| Minimum number of samples of leaf node                | 1              | 1        | 1                      |
| Minimum weight in leaf node samples                   | 0              | 0        | 0                      |
| Maximum depth of number                               | 10             | 10       | 10                     |
| Maximum number of leaf nodes                          | 50             | 50       | 50                     |
| Threshold of node partition purity                    | 0              | 0        | 0                      |

Therefore, we adopted the results of random forest regression model analysis, and the predicted structural variable data are shown in Table 5.

**Table 5.** Structural variables

| Receiving distance (cm) | Hot air speed (r/min) | thickness (mm) | Porosity | Compression resilience |
|-------------------------|-----------------------|----------------|----------|------------------------|
| 38                      | 850                   | 2.756          | 96.022   | 84.811                 |
| 33                      | 950                   | 2.701          | 96.414   | 87.263                 |
| 28                      | 1150                  | 2.771          | 96.384   | 87.358                 |
| 23                      | 1250                  | 2.609          | 95.998   | 86.175                 |
| 38                      | 1250                  | 3.513          | 96.755   | 83.793                 |
| 33                      | 1150                  | 3.068          | 96.420   | 87.082                 |
| 28                      | 950                   | 2.398          | 95.654   | 88.234                 |
| 23                      | 850                   | 1.892          | 94.362   | 87.490                 |

#### 4. Analysis of relationship between structural variables and product performance

In this section, multidimensional scaling transformation (MDS) is used to reduce the dimension of structural variables, product performance, process parameters and other data to ensure that the representation of these objects in the low dimensional space matches the original similarity as much as possible. Then, the random forest regression method is used to obtain the prediction data and find the maximum value in the prediction data.

##### 4.1. MDS dimension reduction algorithm

The distance matrix of the data points in the original space is calculated.

$$b_{ij} = -\frac{1}{2}(dist_{ij}^2 - dist_i^2 - dist_j^2 + dist_{..}^2) \tag{14}$$

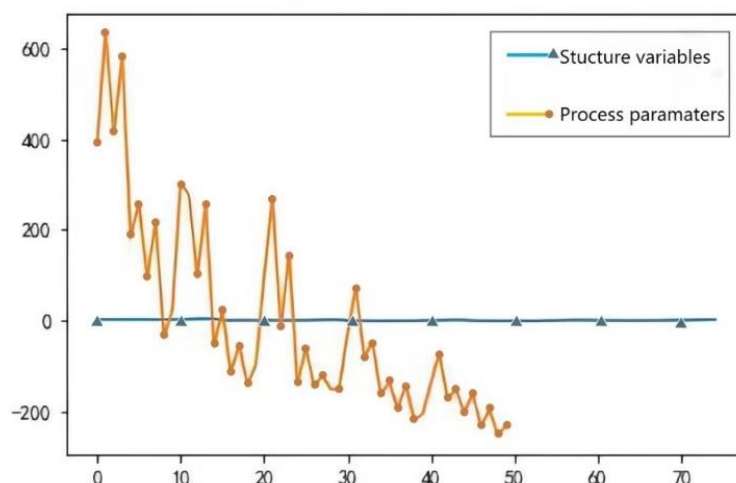
Calculate the inner product matrix.

$$B = V \Lambda V^T \tag{15}$$

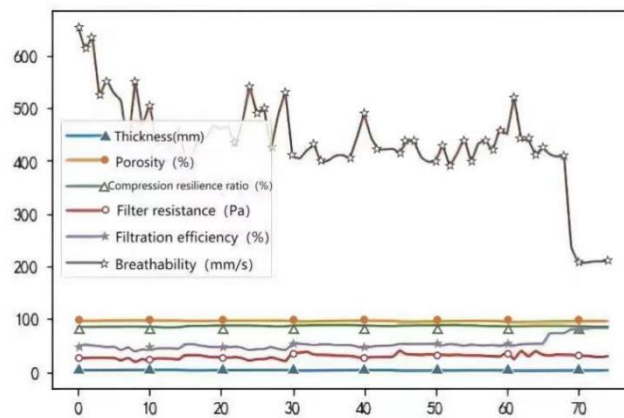
The eigenvalue decomposition is performed on the matrix  $B$  to obtain the eigenvalue matrix  $\Lambda$  and the eigenvector matrix  $V$ . Take the first  $Z$ -term of the largest eigenvalue matrix and its corresponding eigenvector.

$$Z = V_z A_z^{1/2} \tag{16}$$

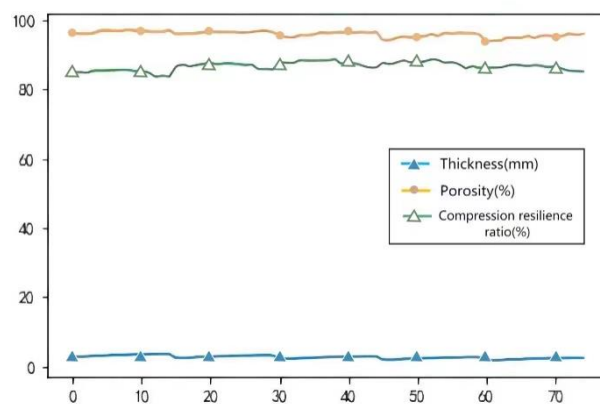
After dimension reduction, the data is analyzed (Fig. 6-9).



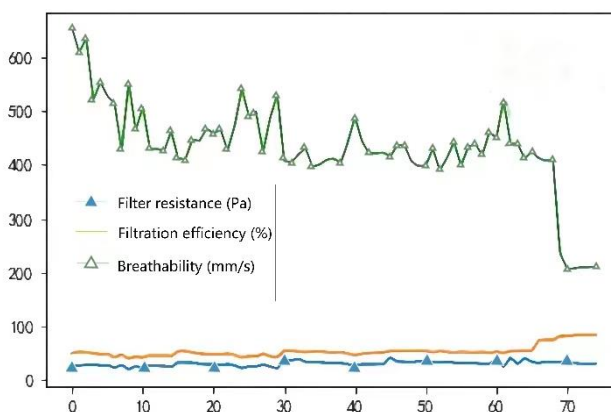
**Fig. 6** Comparison of structural variables and product performance



**Fig. 7** Comparison of structural variables and product performance



**Fig. 8** Product performance



**Fig. 9** Structural variables

**4.2. Calculate the maximum filtration efficiency**

In this paper, linear regression model, random forest regression model and lightGMB regression model are used to analyze the data. The error results are shown in Table 6:

**Table 6.** Regression Model Analysis

| Model name                     | Filtration efficiency |
|--------------------------------|-----------------------|
| Linear regression model        | 10.2161               |
| Random forest regression model | 5.1178                |
| LightGMB                       | 11.7069               |

Therefore, the results of random forest regression model analysis are adopted, and the model parameters are shown in Table 7:

**Table 7.** Model parameters

| Parameter name  | Parameter value |
|---|-----------------|
| Training time   | 0.146s          |
| Data segmentation                                     | 0.7             |
| Data shuffle  | No              |
| Cross validation                                      | No              |
| Evaluation criteria for node splitting                | mse             |
| Maximum feature ratio considered during division      | None            |
| Minimum number of samples for internal node splitting | 33              |
| Minimum number of samples of leaf node                | 1               |
| Minimum weight of samples in leaf nodes               | 0               |
| Maximum depth of the tree                             | 20              |
| Maximum number of leaf nodes                          | 50              |
| Threshold of node partition purity                    | 0               |
| Number of decision trees                              | 100             |
| Sublevel caving                                       | true            |
| Data test   | false           |

Select the maximum value of the predicted filtering efficiency data, as shown in Table 8.

**Table 8.** Result analysis

| Forecast results Y | Filtration efficiency(%) | Receiving distance (cm) | Hot air speed (r/min) |
|--------------------|--------------------------|-------------------------|-----------------------|
| 49.4059263         | 51.80447034              | 20                      | 800                   |
| 49.4059263         | 48.37860837              | 20                      | 800                   |
| 49.4059263         | 52.58110184              | 20                      | 800                   |

Since there are a plurality of the same highest values, the average value of the acceptance distance and the hot air speed is obtained as the final value. It can be seen that when the receiving distance is 20 and the hot air speed is 800, the filtration efficiency of the product will reach the highest.

## 5. Conclusions

This model can effectively calculate the maximum filtering efficiency. When the receiving distance is 20m and the hot air speed is 800km/h, the maximum throughput efficiency is generated.

## References

- [1] Fengcai Han, Zhaozhou Han, Haiming Lin. Comprehensive evaluation method of initial factor analysis and its application [J] Statistics and decision making, 2022, (14): 10-14.
- [2] Ailing Wang. Exploration on academic evaluation of university scholars based on factor analysis [J] Modern information technology, 2022,6 (01): 43-47.
- [3] Juwang Fu, Xinbing Kong. Research on causal inference based on factor analysis [J] Journal of Chongqing Industrial and Commercial University (NATURAL SCIENCE EDITION): 1-10.
- [4] Yunyan Shang, Shiqiang Zhu, Hao Sun. Analysis on heteroscedasticity of multiple linear regression model -- Discussion on teaching content of "applied regression analysis" [J] Science and technology wind, 2022, (15): 19-21 + 131.

- [5] Yingli Pan, Zhan Liu, Lingling Yan. Research on distributed computing method based on large-scale high-dimensional linear regression model [J] Journal of Applied Mathematics, 2022,45 (03): 339-354.
- [6] Chuanpeng He, Ling Yin, Bo Huang, Mingsheng Wang, Ruyan Guo, Shuai Zhang, Jiaji Ju. Text keyword extraction method based on Bert and lightgbm [J] Electronic technology: 1-8.
- [7] Xu Wen, Hao Wang, Gang Huang. Identification method of bus load abnormal data based on factor analysis [J] Journal of Chongqing University, 2021,44 (8): 91-102.
- [8] Zhihua Zhou. Machine learning. Beijing: Tsinghua University Press, 2016: PP 121-139, 298-300.
- [9] Hang Li. Statistical learning methods. Beijing: Tsinghua University Press, 2012: Chapter 7, pp.95-135.
- [10] Meng Q. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. 2018.
- [11] Yanan Gao, Wenqian Wang, Jianxin Wang. LightGBM food safety risk early warning model integrating fuzzy hierarchy: Taking meat products as an example [J] Food science, 2021 (1).