Academic Journal of Science and Technology
ISSN: 2771-3032 | Vol. 4, No. 3, 2022

Study on Ceramic Sintering Process Based on Random Forest

Runhui He[1], Yuna Sheng[1], Jianxin Zhang[1], Zihan Xu[1], Chengrui Li[1], Liya Wang[1, 2, *]

[1]North China University of Technology, Tangshan, 063210, China
[2]Tangshan Sanitary Ceramics Quality Intelligent Monitoring Technology Basic Innovation Team, Tangshan, 063210, China
*** Corresponding author**

Abstract: While the manufacturing process of ceramics is complex and each link is important to the ceramic, we focus on the sintering process of ceramics and construct models and algorithms to investigate the temperature of the ceramic sintering process in order to improve the yield of the ceramic sinter. We make full use of the ceramic sintering process node temperature and ceramic product problem data, adopt factor analysis for data dimensionality reduction, and use principal component analysis to extract factors to improve efficiency for subsequent optimisation. The random forest algorithm of integrated machine learning is used to investigate the effect of different nodal temperature rates on the ceramic product, and a decision tree is derived to identify the more influential temperature rates. These obtained important temperature factors affecting ceramic sintering can be used as a reference for the ceramic sintering process, thus achieving a study of the temperature of the ceramic sintering process.

Keywords: Ceramic sintering, Factor analysis method, Random forest.

1. Introduction

The ceramics industry plays an important role in modern construction and everyday life. The variety of forms and complex structures of modern ceramic products, as well as the increasing demands on product quality, have placed new demands on ceramic materials and manufacturing processes. The separation of the product design and manufacturing processes in traditional ceramic manufacturing technology has led to a large number of iterations from product design to the finalisation of the production process; relying on the "experience + experiment" method to develop the process. These have become bottlenecks in the development of ceramic technology, and there is an urgent need to adopt new product development models and technologies to change this situation. With the development of computer technology, ceramic manufacturing technology has changed greatly. The use of computer technology to simulate the ceramic production process can verify the rationality of the production process selection, but also optimize the production process, reduce losses, shorten the development cycle, improve product quality, etc.

Random forests are a type of semi-supervised learning, which are efficient to run and easier to use, but decision trees are more likely to cause overfitting, making the trained learning models less generalisable. In order to obtain the correctness of the combination and the diversity of the single-root tree, many scholars have made many researches. Liu abet proposed a random forest algorithm based on multiple fuzzy kernel constraints; Wu Chenwen et al. improved the prediction accuracy through variable prediction and selection; Xla et al. solved the classification problem of incomplete data without attribution by proposing a random forest with surrogate algorithm; Shi Jinyu et al. proposed an improved random forest algorithm to improve the classification accuracy by constructing fuzzy decision trees.

2. Data Pre-processing

The data set contains the temperatures of 27 nodes in the ceramic firing process, as well as the five problems of cracking, oxidation, bulging, soiling and wind shock in the production of ceramic products. From the 800,000 data, 560,000 data were randomly selected as the training set and 240,000 data as the test set to investigate the effect of temperature rate on the quality of ceramic products. The processed data are shown in Table 1.

Where a positive sign indicates the rate of temperature rise and a negative sign indicates the rate of temperature fall, 0 means the product does not have this problem and 1 means the product has this problem.

3. Factor Analysis Method

Factor analysis is an important method of data dimensionality reduction. The principle is to combine some variables with complex relationships (X_i) into a small number of factors (F_k) according to the dependencies within the correlation matrix of various indicators, and to use the resulting factors to reflect the main information of the original variables. The mathematical model of factor analysis is used to find the original variables

$$X = AF + \epsilon \tag{1}$$

where F is the extracted common factor; A is the factor loading matrix.ϵ is the special factor.

3.1. Indicator correlation analysis

When performing factor analysis on data, we first test whether the data are suitable for factor analysis. The basic logic of factor analysis is to construct a few representative factor variables from the original variables, which requires a strong correlation between the original variables, otherwise factor analysis will not be able to extract the 'common features' between the variables.

Table 1. Partial data presentation

Kiln car number	V1	V2	V3	V4
06080	2.227124464	1.352734610	1.695913540	1.557662630
06081	2.158663761	1.426641993	1.581061737	1.588638055
06082	2.034577636	1.359453215	1.644596840	1.767858508
06083	1.779988723	1.764824962	1.702132689	1.685992475
06084	1.942583953	1.583415125	1.722795668	1.398354660
	V5	V6	V7	V8
06080	1.707215787	3.818557288	3.703093038	2.980067668
06081	1.682474525	3.725085829	3.821305758	3.004810595
06082	1.418555999	3.821305758	3.989004355	2.903092049
06083	1.118900655	4.137456954	4.379381012	2.243298260
06084	1.731958725	4.398625333	3.384193372	2.905842185
	V9	V10	V11	V12
06080	1.190379300	0.591065279	0.390376679	0.533333981
06081	1.118898972	0.637802680	0.384879717	0.585564991
06082	1.173881789	0.492096896	0.467353942	0.654296838
06083	1.250859079	0.536082463	0.409622644	0.574571084
06084	1.017182108	0.657045346	0.390376651	0.602062502
	V13	V14	V15	V16
06080	0.335393835	0.079725771	0.134708561	-0.175945673
06081	0.288659788	0.043988942	0.170445385	-0.065980040
06082	0.203437109	0.085222679	0.148454292	-0.137457042
06083	0.373882494	0.008248770	0.148450911	-0.082474225
06084	0.302405492	0.082474225	0.126458919	-0.148454265
	V17	V18	V19	V20
06080	-4.248816697	-5.052194668	-2.083349152	-0.606767866
06081	-4.376313083	-5.002398403	-2.191315898	-0.459079611
06082	-4.076697754	-5.155935819	-2.045361756	-0.574739258
06083	-4.503807525	-4.817739879	-2.112340335	-0.741110777
06084	-4.035259867	-5.155935827	-2.147329401	-0.683280549
	V21	V22	V23	V24
06080	-1.887152592	-0.748918179	-1.723222388	-3.852839835
06081	-1.981155256	-0.748918179	-1.723222388	-3.852819835
06082	-1.839264116	-1.008158988	-1.647881796	-3.753789724
06083	-1.754129648	-0.926829196	-1.636660128	-3.577926247
06084	-1.947455677	-0.596423981	-1.667117473	-3.658783330
	V25	V26	Burning crack	Oxidation
06080	-1.834593429	-1.929634962	0	0
06081	-1.719129458	-2.097201731	0	0
06082	-1.746620865	-2.120765885	1	0
06083	-1.984879597	-2.107674689	0	1
06084	-1.966551991	-2.171304233	0	1
	Dirty	Wind shock	Bulge	
06080	0	0	0	
06081	0	0	0	
06082	0	0	0	
06083	0	0	0	
06084	0	0	0	

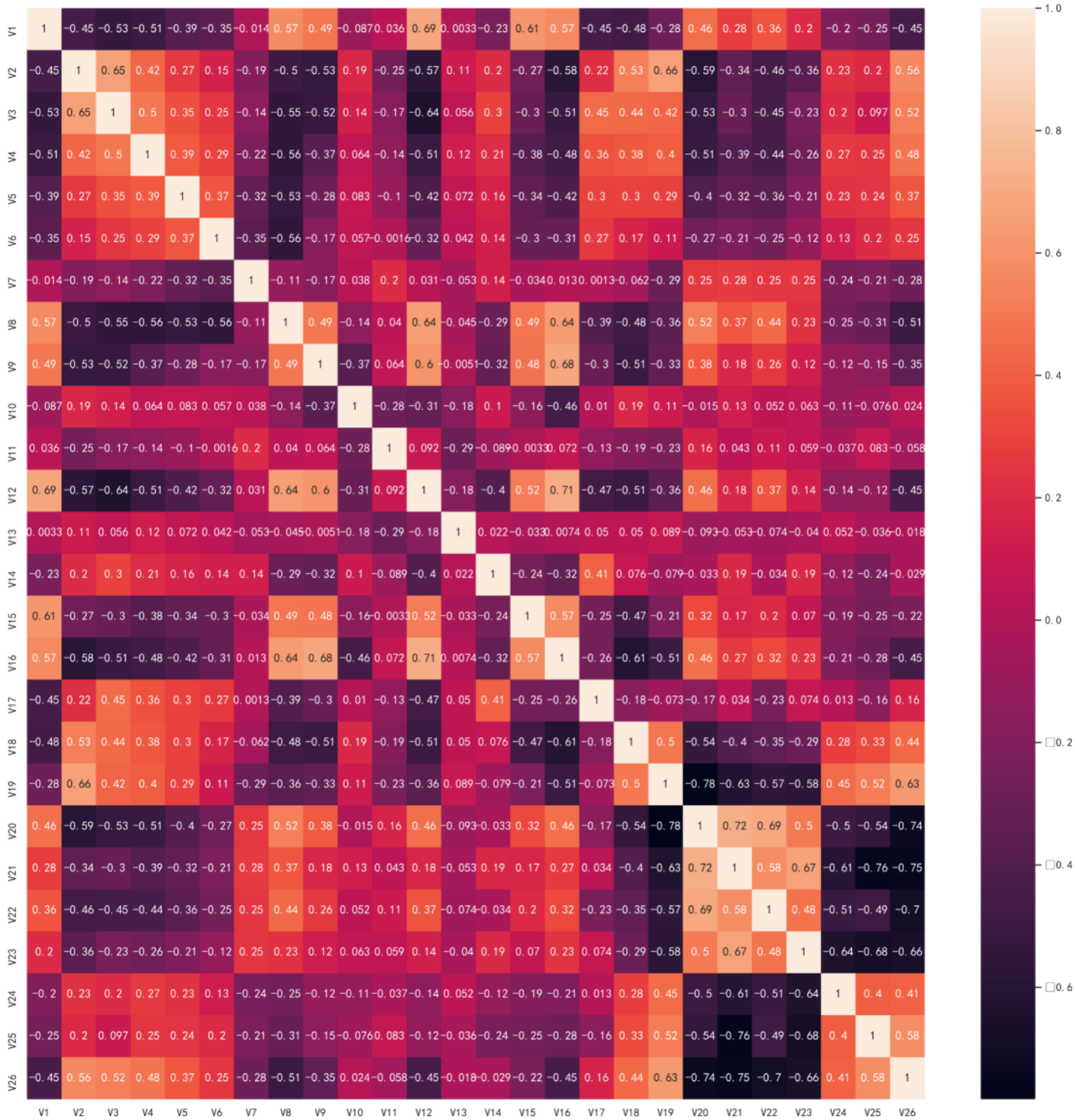

Figure 1. Correlation heat map

A heat map of the correlation coefficients of the variables is shown in Figure 2. The correlation between the variables is strong, with a value of 0.586 for KMO and a significance of 0.00, so factor analysis can be used.

3.2. Factor extraction

Factor extraction methods commonly used in factor analysis include the principal component method, the maximum likelihood method and the principal axis factor decomposition method. This method assumes that the variables are linear combinations of factors, and the variance of the original variables is explained by the principal components as much as possible from the overall variance of the original variables, and the proportion of the variance of the original variables explained by each common factor is reduced in turn.

The influence factors in the project, the 26 temperature variations affecting ceramic sintering, were brought in to obtain the eigenvalues and the component matrix. The number of principal components was determined based on the total variance explained, and the top 6 factors were selected in order of eigenvalue greater than one, with a cumulative variance explained of 71.230%.

Table 2. Initial eigenvalue

Composition	Total	Percentage of variance	Cumulative (%)
1	9.107	35.029	35.029
2	3.734	14.36	49.389
3	1.801	6.929	56.318
4	1.548	5.955	62.273
5	1.222	4.701	66.974
6	1.107	4.257	71.23

Table 3. Extract the sum of squares of loads

Composition	Total	Percentage of variance	Cumulative (%)
1	9.107	35.029	35.029
2	3.734	14.36	49.389
3	1.801	6.929	56.318
4	1.548	5.955	62.273
5	1.222	4.701	66.974
6	1.107	4.257	71.23

Table 4. Sum of squares of rotational loads

Composition	Total	Percentage of variance	Cumulative (%)
1	6.175	23.751	23.751
2	4.584	17.631	41.382
3	2.993	11.511	52.893
4	1.87	7.194	60.087
5	1.688	6.492	66.578
6	1.21	4.652	71.23

A reference line with a value of 1 is set in the gravel plot to obtain 6 components with eigenvalues over 1. It is sufficient to consider only these 6 components.

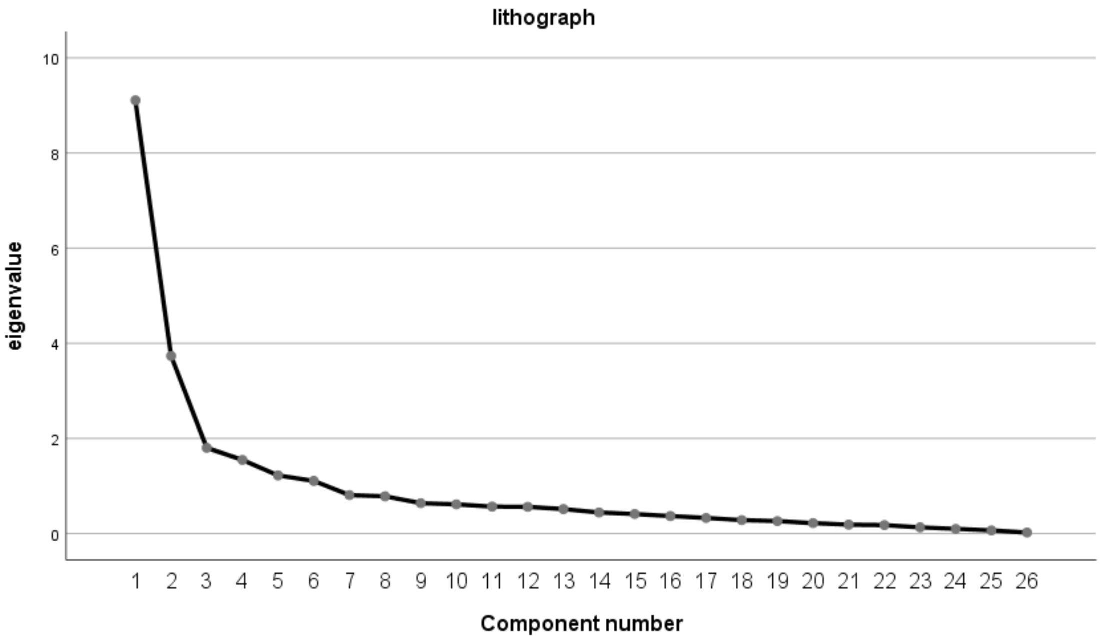

Figure 2. Gravel map

3.3. Factor rotation

In order to obtain a clearer meaning of each factor, the Kaiser normalized maximum variance method was used to rotate the factors and to summarize the indicators corresponding to the six factors, and the rotated component loading matrix and component score coefficient matrix are shown in Table.

Table 5. Rotated factor component loading factors and score coefficients

Indicators	Factor component loadings after rotation						Factor score coefficient					
	F1	F2	F3	F4	F5	F6	F1	F2	F3	F4	F5	F6
V01	0.243	0.648	-0.388	-0.208	0.116	-0.079	-.016	.147	-.042	-.031	.132	-.081
V02	-0.503	-0.361	0.386	-0.109	0.42	-0.038	-.086	.018	.138	-.180	.203	-.036
V03	-0.357	-0.327	0.673	0.051	0.211	-0.002	-.057	.068	.272	-.099	.072	-.030
V04	-0.339	-0.364	0.386	0.314	0.099	0.115	-.010	-.015	.079	.113	.023	.081
V05	-0.254	-0.302	0.212	0.596	0.049	0.01	.037	-.020	-.029	.346	.018	-.010
V06	-0.082	-0.228	0.193	0.76	-0.103	-0.014	.080	-.013	-.043	.480	-.060	-.038
V07	0.325	-0.277	0.086	-0.637	-0.393	0.023	.006	-.171	.073	-.444	-.284	.072
V08	0.299	0.615	-0.41	-0.297	0.1	-0.013	-.009	.114	-.054	-.087	.120	-.016
V09	0.138	0.73	-0.311	0.126	-0.11	0.149	-.015	.185	-.030	.163	-.001	.084
V10	0.213	-0.372	-0.023	0.027	0.541	-0.495	.097	-.093	-.095	.074	.335	-.382
V11	0.019	0.009	-0.053	-0.053	-0.783	-0.233	-.043	-.040	.030	-.050	-.500	-.194
V12	0.147	0.67	-0.51	-0.118	-0.149	-0.053	-.036	.124	-.098	.025	-.031	-.061
V13	0.018	-0.045	0.031	-0.002	0.197	0.873	.042	-.074	-.058	-.032	.124	.745
V14	0.278	-0.206	0.626	0.017	0.014	-0.068	.059	.034	.267	-.045	-.018	-.072
V15	0.041	0.778	-0.023	-0.228	0.16	-0.121	-.090	.300	.184	-.108	.143	-.157
V16	0.216	0.794	-0.207	-0.118	-0.18	0.189	-.038	.207	.066	-.013	-.053	.117
V17	0.062	-0.083	0.83	0.24	-0.074	0.079	.017	.137	.371	.054	-.075	.014
V18	-0.365	-0.717	-0.159	-0.025	0.229	0.028	.004	-.261	-.228	-.058	.093	.084
V19	-0.774	-0.27	-0.01	-0.052	0.334	-0.042	-.138	-.007	-.038	-.119	.161	-.034
V20	0.794	0.303	-0.215	-0.107	-0.117	-0.046	.136	-.016	-.053	.055	-.017	-.023
V21	0.872	0.103	0.073	-0.176	0.075	-0.052	.163	-.029	.050	-.022	.082	-.024
V22	0.736	0.078	-0.328	-0.129	-0.017	-0.043	.152	-.115	-.158	.043	.035	.005
V23	0.86	0.034	0.096	-0.055	-0.021	0.032	.178	-.058	.030	.047	.020	.047
V24	-0.691	-0.097	-0.057	0.122	-0.057	0.127	-.127	.010	-.046	.002	-.065	.094
V25	-0.756	-0.198	-0.299	0.176	-0.183	-0.064	-.127	-.064	-.178	.062	-.143	-.049
V26	-0.829	-0.209	0.248	0.135	0.049	-0.096	-.158	.071	.100	-.037	-.024	-.110

Based on the matrix of coefficients for the table component scores, the one-way function can be derived as follows:

$$F1 = -0.016 * V1 - 0.086 * V2 - 0.057 * V3 - 0.01 * V4 + 0.037 * V5 + 0.08 * V6 + 0.006 * V7 - 0.009 * V8 - 0.015 * V9 + 0.097 * V10 - 0.043 * V11 - 0.036 * V12 + 0.042 * V13 + 0.059 * V14 - 0.09 * V15 - 0.038 * V16 + 0.017 * V17 + 0.004 * V18 - 0.138 * V19 + 0.136 * V20 + 0.163 * V21 + 0.152 * V22 + 0.178 * V23 - 0.127 * V24 - 0.127 * V25 - 0.158 * V26 \quad (2)$$

$$F2 = 0.147 * V1 + 0.018 * V2 + 0.068 * V3 - 0.015 * V4 - 0.02 * V5 - 0.013 * V6 - 0.171 * V7 + 0.114 * V8 + 0.185 * V9 - 0.093 * V10 - 0.04 * V11 + 0.124 * V12 - 0.074 * V13 + 0.034 * V14 + 0.3 * V15 + 0.207 * V16 + 0.137 * V17 - 0.261 * V18 - 0.007 * V19 - 0.016 * V20 - 0.029 * V21 - 0.115 * V22 - 0.058 * V23 + 0.01 * V24 - 0.064 * V25 + 0.071 * V26 \quad (3)$$

$$F3 = -0.042 * V1 + 0.138 * V2 + 0.272 * V3 + 0.079 * V4 - 0.029 * V5 - 0.043 * V6 + 0.073 * V7 - 0.054 * V8 - 0.03 * V9 - 0.095 * V10 + 0.03 * V11 - 0.098 * V12 - 0.058 * V13 + 0.267 * V14 + 0.184 * V15 + 0.066 * V16 + 0.371 * V17 - 0.228 * V18 - 0.038 * V19 - 0.053 * V20 + 0.05 * V21 - 0.158 * V22 + 0.03 * V23 - 0.046 * V24 - 0.178 * V25 + 0.1 * V26 \quad (4)$$

$$F4 = -0.031 * V1 + 0.180 * V2 - 0.099 * V30.113 * V4 + 0.346 * V5 + 0.480 * V6 + -0.444 * V7 - 0.087 * V8 + 0.163 * V9 + 0.074 * V10 - 0.05 * V11 + 0.025 * V12 - 0.032 * V13 - 0.045 * V14 - 0.108 * V15 - 0.013 * V16 + 0.054 * V17 - 0.058 * V18 - 0.119 * V19 + 0.055 * V20 - 0.022 * V21 + 0.043 * V22 + 0.047 * V23 + 0.002 * V24 + 0.062 * V25 - 0.037 * V26 \quad (5)$$

$$F5 = 0.132 * V1 + 0.203 * V2 + 0.072 * V3 + 0.023 * V4 + 0.018 * V5 - 0.06 * V6 - 0.284 * V7 + 0.12 * V8 - 0.001 * V9 + 0.335 * V10 - 0.5 * V11 - 0.031 * V12 + 0124 * V13 - 0.018 * V14 + 0.143 * V15 - 0.053 * V16 - 0.075 * V17 + 0.093 * V18 + 0.161 * V19 - 0.017 * V20 + 0.082 * V21 + 0.035 * V22 + 0.02 * V23 - 0.065 * V24 - 0.143 * V25 - 0.024 * V26 \quad (6)$$

$$F6 = -0.081 * V1 - 0.036 * V2 - 0.03 * V3 + 0.081 * V4 - 0.01 * V5 - 0.038 * V6 + 0.072 * V7 - 0.016 * V8 + 0.084 * V9 - 0.382 * V10 - 0.194 * V11 - 0.061 * V12 + 0.745 * V13 - 0.072 * V14 - 0.157 * V15 + 0.117 * V16 + 0.014 * V17 + 0.084 * V18 - 0.034 * V19 - 0.023 * V20 - 0.024 * V21 + 0.005 * V22 + 0.047 * V23 + 0.094 * V24 - 0.049 * V25 - 0.11 * V26 \quad (7)$$

4. Random Forests

4.1. Random forests

Each classification tree in a random forest is a binary tree, which is generated based on the principle of top-down recursive splitting, where the training set is partitioned one by one, starting from the root node. In a binary tree, all the training data are concentrated at the root node, and they are divided into left and right nodes, each containing a subset of the training data, according to the principle of the lowest node purity. The nodes continue to be split according to the same rules until the rules are satisfied before stopping further subdivision. If all the classified data on a node are from the same class, then the impurity of this nodel(n) = 0 , in this method, impurity is measured according to the Gini criterion. It is assumed that$P(X_j)$ is the number of samples on node n belonging toX_j the frequency of the number of samples of class to the total number of training samples, the specific implementation of the random forest training process is as follows.

(1) Assuming an initial training set of N, a set of k self-help samples is drawn randomly and with put-back using the bootstrap method, so that k classification trees are constructed, with the samples not drawn each time forming the out-of-bag data.

(2) Let there be a total of M feature variables, and at each node of the tree, m features are randomly selected from the overall M feature variables ($m < M$) at each node, and then select the feature with the best classification ability among the m features based on the criterion of minimum impurity of each node to continue the branching operation.

(3) Each tree is allowed to grow to its maximum extent and no pruning is done to them with the aim of minimizing the impurity of the nodes.

(4) The resulting multiple trees as a whole form a random forest, and the resulting random forest classifier can then be used to identify and categorise the new data.

4.2. Random forest use

The random forest model was implemented in the program to analyse the available data using the random forest algorithm, and the results were tabulated to show the degree of influence of the ceramic products, see Table 3.

Table 6. Table of the degree of influence of temperature rate on ceramic products

	Burning crack	Oxidation	bulging	dirty	Wind Surprise
F1	0.16714	0.16637	0.17323	0.16675	0.16817
F2	0.17084	0.16703	0.16555	0.17543	0.16680
F3	0.16597	0.16688	0.16254	0.16785	0.16610
F4	0.16642	0.16697	0.16494	0.15997	0.16541
F5	0.16454	0.16658	0.17043	0.16281	0.16626
F6	0.16510	0.16616	0.16331	0.16720	0.16726

5. Experimental Analysis

The final results derived from the Random Forest algorithm show that among the five questions for ceramic products, F1 and F2 have a greater weighting, and according to the characteristic weights obtained from the factor analysis method, it can be seen that V1, V9, V15, V16, V20, V21, V22 and V23, the temperature rates of these 8 nodes have a greater degree of influence on the ceramic products.

6. Conclusion

In this experiment, by collecting data on 27 nodes of temperature and ceramic product issues during the ceramic firing kiln process, after data pre-processing, checking the correlation of variables, factor analysis method to reduce the dimensionality, the random forest algorithm model was adopted for scientific and objective analysis of the data. Finally the weights of the degree of influence of each temperature node rate on the ceramic product were obtained. It is a reference value for the control of the temperature rate of ceramic firing process.

Acknowledgment

The authors would like to acknowledge the support of the Innovation and Entrepreneurship Program for Students of North China University of Technology (X2022232) for writing this article.

References

[1] Zhang Changlu. Factor analysis-based dimensionality reduction of coal quality indicators and its visualization application[J]. Coal Technology,2017,36(01):310-313.

[2] HU Xinxiang,ZHAO Xia,ZHANG Qian,LI Yinglan,KONG Xiangsheng,MA Yuting. Weight analysis of factors influencing agricultural yield based on random forest algorithm[J]. Gansu Science and Technology,2020,36(09):16-19+40.

[3] Wang Fubin,Wang Rui,Wu Chen. Short-term prediction of sintering state based on improved random forest algorithm[J]. Advances in Lasers and Optoelectronics,2022,59(18):382-388.

[4] LI Jianming,WANG Su,YUAN Shenghua. Numerical simulation and simulation of the temperature field of ceramic sintering process [J]. China Ceramics, 2003 (03): 14-16.DOI:10.16521.

[5] Zhang Xiaofeng,Lin Bin,Chen Shihuan. Temperature field analysis of ceramic sintering process[J]. Rare Metal Materials and Engineering,2008,37(S1):314-318.

[6] Zhang Cui, Ai Chunting, Chen Zuobing, Niu Jun, Zheng Hongqin. Research on computer simulation of bioceramic sintering[J]. Journal of Wuhan University of Technology, 2002 (05): 26-28.

[7] Kong Weijing, Sana, Ma Yulan, Li Zhipeng, Zhou Hengwei, Dong Xiao. Effect of different sintering temperatures on the structure and morphology of LaOCl-MgO-SiC ceramics[J]. Journal of Yili Normal College (Natural Science Edition), 2018, 12(01): 39-42.

[8] Zhang Hao, Guo Jun, Dang Junjie, Zhan Jun, Geng Chunlei, Cui Song, Tang Wenming. Effect of sintering temperature on the structure and properties of aluminum nitride ceramics[J]. China Ceramics,2015,51(07):54-58.DOI:10.16521.