

Article

Research on Sinter Quality Prediction System Based on Granger Causality Analysis and Stacking Integration Algorithm

Xin Li, Xiaojie Liu, Hongyang Li, Ran Liu *, Zhifeng Zhang, Hongwei Li, Qing Lyu and Liangyixin Wen

College of Metallurgy and Energy, North China University of Science and Technology, Tangshan 063210, China

* Correspondence: lr819@163.com

Abstract: Sinter ore quality directly affects the stability of blast furnace production. In terms of both physical and chemical properties, the main indicators of sinter quality are the TFe content, alkalinity, and drum index. By analyzing the massive historical data on the sinter production of a steel company, this study proposes a sinter quality prediction system based on Granger causality analysis and a stacking integration algorithm. First, based on real historical data of sintering production in steel enterprises (including coal gas pressure, ignition temperature, combustion air pressure, etc.), data preprocessing of raw data was carried out using a combination of feature engineering and the sintering process. Second, Pearson correlation analysis, Spearman correlation analysis, and Granger causality analysis were used to screen out the characteristic parameters with a strong influence on the target variable of sinter quality (drum Index, TFe, alkalinity). Third, a prediction model for sinter quality parameters was developed using a stacking integration algorithm pair for training. Finally, a program development tool was used to realize the establishment and online operation of a sinter ore quality prediction system. The test results showed that the predicted goodness of fit of the model for the TFe content, alkalinity (R), and drum index were 0.942, 0.958, and 0.987, respectively, and the model calculation time met the actual production requirements. By establishing a suitable model and running the program online, the real-time prediction of the main indicators of sinter quality was realized to guide production promptly.



Citation: Li, X.; Liu, X.; Li, H.; Liu, R.; Zhang, Z.; Li, H.; Lyu, Q.; Wen, L. Research on Sinter Quality Prediction System Based on Granger Causality Analysis and Stacking Integration Algorithm. *Metals* **2023**, *13*, 419. <https://doi.org/10.3390/met13020419>

Academic Editor: Alain Pasturel

Received: 21 November 2022

Revised: 9 February 2023

Accepted: 12 February 2023

Published: 17 February 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: sintering quality; Granger causality analysis; stacking integration algorithm

1. Introduction

At this stage, China's black smelting technology is still dominated by the "sintering-blast furnace-converter" process. The CO₂ emissions of sintering and blast furnace (BF), which consume the most energy, account for about 80% of the CO₂ emissions of the iron and steel industry [1]. In the main three-step process, the former always provides raw materials for the latter, so sintering production is crucial. Although some Chinese steel enterprises have vigorously increased the proportion of pellets in the charge structure of BF, the proportion of sinter is still as high as about 75% [2]. Therefore, the quality of the sinter is still the key to the overall production level and emission level: high-quality sinter can effectively reduce the energy consumption of each part and play a positive role in the heat supply and fuel consumption in the furnace. In the sinter-production process, the staff mainly pays attention to the location of the burning through point (BTP). Stable control of BTP requires improving the utilization efficiency of the sintering machine and the quality of the sinter minerals, which can effectively reflect the sintering thermal state, which is one of the important signs when judging whether the sintering process is normal. In addition, the physicochemical properties of sinter quality are also of concern at the production site.

The sintering process has the characteristics of time delay and nonlinearity [3–6]. Although some mechanistic models simulate part of the sintering process, they cannot be adapted to the production environment due to the instability of the sintered material and the frequent changes in the production pattern. The data-driven prediction model regards

the sintering system as a black box model and has achieved good results in the monitoring of BTP and the prediction of sinter quality. Wang Qingyao established an artificial-neural-network-based online prediction model for the sintered ore drum index using the predictive power of the Elman model to improve its prediction model's guidance [7,8]. Xin Zicheng applied the BP neural network algorithm to the prediction of low-temperature reduction pulverization performance of vanadium and titanium sinter ore to explore the relationship between the input samples and the output samples. The results showed that the BP neural network model applied to the study of the reduction pulverization performance of sintered ore, and the average relative error was 5.7%, which satisfied the requirement of prediction accuracy in production [9]. Yi Zhengming proposed a quality prediction model applying momentum term and change learning rate to improve the BP neural network to achieve the prediction and output of the TFe index of sintered ore [10]. Li developed a prediction model for BF production using the superposition algorithm and demonstrated that the process parameters of different regions of the BF were predicted using the superposition algorithm to achieve better results in both classification and regression models [11]. Chen analyzed the correlation of a telemetry data time series using the Granger causality model and established a causality model to detect anomalies and determine the causes of anomalies through causality under normal conditions [12]. Some other scholars have used an intelligent approach to achieve the development and application of predictive models with different parameters [13–15].

The existing technology of the prediction model for sinter quality can only predict a single project and has not yet realized a model-establishment mechanism that can meet the prediction of multiple sinter physicochemical properties. Moreover, the sinter quality contains many parameters, and the dimensions of different parameters are different, resulting in the model's prediction accuracy.

Therefore, this study combined the advantages of mechanistic analysis and the black box model to obtain the causal relationships and lag time windows between sintering variables and BTP, sintering variables, and sintering quality using correlation analysis and causality analysis. This result is then applied to the integrated learning model to build a quality state prediction model for the sintering system.

2. Data Foundation and Model Algorithm

2.1. Production Data

Taking a 360 m² sintering machine in a domestic steel company as the research object, the time range included the actual production data of the sintering process from January 2018 to December 2019. The sintering data were divided into four parts: raw material data, operation data, status data, and quality-inspection data. The specific names of some parameters are shown in Table 1. It can be seen in Table 1 that the raw material data include the amount of sintering raw materials used and the chemical composition content of raw materials. Among them, the composition of SiO₂ and MgO in raw materials all have a great influence on the quality of sintered ore. The operation data are human-operated sintering machine parameters, which cover important operating parameters of the whole sintering production process, such as trolley material thickness, coal gas pressure, exhaust gas temperature of north bellows, ring cooler speed, etc. The status data indicate the operating and working status of the sintering machine, including significant status parameters of the whole sintering production process such as BTP, air box exhaust gas temperatures, air box vacuum degree, etc. The quality-inspection data refer to the physical and chemical properties of the sinter ore, including alkalinity, drum Index, FeO, etc. Based on the characteristics of different data types, sintering production data are stored in different databases. The quality-inspection data are stored in the Oracle database, and other data are stored in the SQL Server database. Based on the difference in data sources and production processes, the generation frequency of sintering production data is different. The generation frequency of raw material data and quality-inspection data is related to the detection frequency of the production site, and there is no obvious rule.

The operation data and status data are automatically generated by the equipment in the sintering production process, and the frequency is 1 second/iteration, but to facilitate subsequent data processing and analysis, the collection frequency of these two types of data is set at 1 hour/iteration.

Table 1. Process parameters of the sintering machine.

Parameter Type	Parameter Name	Abbreviations	Parameter Name	Abbreviations
Raw material parameters	63.5%Vanadium powder	VP63.5	Calcium lime powder/(t·h ⁻¹)	GSHF
	Steelmaking dust ash/(t·h ⁻¹)	SDA	Magnesium lime powder/(t·h ⁻¹)	MSHF
	Sintering return mines/(t·h ⁻¹)	SRM	Blast furnace return mines/(t·h ⁻¹)	BFRM
Mixes Composition parameters	Iron ore powder with vanadium_SiO ₂	IOPV_SiO ₂	Iron ore powder with vanadium_MgO	IOPV_MgO
	Mixes_(water)/%	M-H ₂ O	Mixes_(FeO)/%	M-FeO
	Mixes_(SiO ₂)/%	M-SiO ₂	Mixes_(CaO)/%	M-CaO
	Trolley material thickness/mm	TMT	Coal gas pressure/kPa	CGP
Operating parameters	Coal gas flow/(m ³ ·h ⁻¹)	GGF	Ignition temperature/°C	IT
	Combustion air flow/(m ³ ·h ⁻¹)	CAF	Combustion air pressure/kPa	CAP
	No. 1Damper opening/%	1DO	No. 2Damper opening/%	2DO
	Sintering machine speed/(m·min ⁻¹)	SMS	Ring cooler speed/(m·min ⁻¹)	RCS
Status Parameters	Exhaust gas temperature of north bellows/°C	EGTNB	Negative pressure of south pipe/kPa	NPSP
	No. 2Blast volume/(m ³ ·h ⁻¹)	2BV	Round roll speed	RRS
	Burning through point/No.	BTP	Burn through temperature/°C	BTT
	No. 1Airbox exhaust gas temperatures/°C	1AEGT	No. 2Airbox exhaust gas temperatures/°C	2AEGT
	No. 3Airbox exhaust gas temperatures/°C	3AEGT	No. 5Airbox exhaust gas temperatures/°C	5AEGT
	No. 7Airbox exhaust gas temperatures/°C	7AEGT	No. 22Airbox exhaust gas temperatures/°C	22AEGT
	No. 1Air box vacuum degree/kPa	1ABVD	No. 2Air box vacuum degree/kPa	3ABVD
	No. 3Air box vacuum degree/kPa	3ABVD	No. 5Air box vacuum degree/kPa	5ABVD
	No. 7Air box vacuum degree/kPa	7ABVD	No. 22Air box vacuum degree/kPa	22ABVD
	Drum index/%	DI	Screening index/%	SI
	Particle size less than 10 mm/%	PSLT10mm	Fe/%	TFe
	FeO/%	FeO	Alkalinity	R
Sintered Ore Quality parameters	SiO ₂ /%	SiO ₂	CaO/%	CaO

2.2. Data Processing

Sintering production data are a multivariate and heterogeneous type of data with different characteristics and varying quality due to various production conditions. It is necessary to preprocess the original sintering production data to meet the data quality requirements of subsequent causal analysis and predictive model input function.

2.2.1. Abnormal Data Processing

The relative size of the missing value capacity of the data is related to the historical period of subsequent data utilization, so the percentage of missing data is determined by the total amount of data for subsequent analysis. Separate treatment is required according to different missing categories.

Missing Data Values

The relative size of the missing value capacity of the data is related to the historical period of subsequent data utilization, so the percentage of missing data is determined by the total amount of data for subsequent analysis. Separate treatment is carried out according to different missing categories.

First, the missing proportion of all the data is counted, and the feature parameters with a missing proportion greater than 50% are directly deleted.

Second, the device generates class data divided into short-term deletion and long-term deletion according to the length of the deletion time. The short-term deletion state is defined as the missing data in the interval of 1 to 8 h, which is filled by linear interpolation and multiple regression. The multiple regression method uses the Pearson correlation method to find the top n feature parameters with high correlation with the target feature (n is an artificially set critical value, set to the value of 5 in this study). The Pearson correlation coefficient is calculated as follows.

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

In the formula:

r —Pearson correlation coefficient;

x_i and y_i —two characteristic parameters;

$\bar{x}$ and $\bar{y}$ —the mean of the two characteristic parameters, respectively.

The value range of r is $[-1, 1]$, and the larger the absolute value, the stronger the linear correlation between the feature parameters. Using the non-missing data information, the linear relationship between the target feature and all related variables is obtained through the multiple linear regression equation to fill in the short-term missing values of the target feature.

Quality-inspection data mainly include physical and chemical testing data of raw materials, fuels, and finished products in the BF production process. Because the distribution of such data has corresponding periodicity (the fluctuation range is limited in a period of time), the principle of proximity is adopted to fill the data with real values.

Data Noise Value

The data are screened using the box plot method. This method has better resistance for screening abnormal data by depicting the discrete distribution state of the data. The edges of the box plot are defined as follows.

$$\text{Upper limit} = Q_3 + 1.5IQR \quad (2)$$

$$\text{Lower limit} = Q_1 - 1.5IQR \quad (3)$$

$$IQR = Q_3 - Q_1 \quad (4)$$

In the formula:

Q_1 and Q_3 —the first and third quartiles of the dataset, respectively;

IQR . —the interquartile distance, which represents the distance between the upper quartile and the lower quartile.

The intervals in the upper and lower limits belong to the normal data areas, and all data beyond the upper and lower limits are deleted if there is no special demand. The results of the box plot calculation for some parameters are shown in Figure 1. The distribution state of different parameters is different, and the outlier range of parameters is slightly different according to the actual production characteristics.

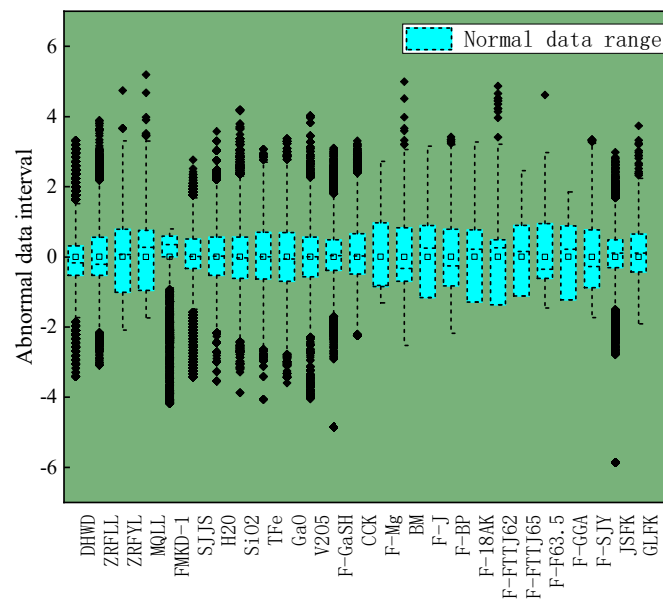


Figure 1. Box diagram of some parameters of BF.

Data Standardization

The sintering production process contains a variety of data parameters, the dimensional units between different data parameters vary greatly, and different dimensional and dimensional units will adversely affect the results of data analysis. The minimum and maximum values in the actual production data are not all values with practical effects, and to eliminate the influence of dimensions between parameters, the data are standardized using the Z-Score. The logic of Z-Score normalization is to normalize the original value x_i of the covariate a to the target value y_i so that it conforms to the standard normal distribution. The Z-Score expression is as follows.

$$y_i = (x_i - \mu) / \sigma \quad (5)$$

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i \quad (6)$$

$$\sigma = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (7)$$

The raw sintering production data, after the above-mentioned missing value processing, outlier processing, and normalization, are used for the data preprocessing of BF production data and can obtain the initial feature dataset suitable for data analysis and model prediction.

2.2.2. Data Frequency Unification

Since the second-level and minute-level data in the sintering production process are not much more useful for analyzing the prediction of sintering parameters than the hourly level, and to facilitate the later data processing and analysis and reduce the pressure on the model calculation, the frequency of data (characteristic parameters and target parameters) collation was set as an hourly frequency in this study. The recording of equipment information was characterized by storing equipment signal data at a frequency of seconds based on the PLC control system. As required, the frequency of the data needs to be reduced to the hourly level. This approach averages the data within the whole hourly interval, which smooths the data overall while reducing the data volume. The record of sintering feeding information is determined by the number of fabrics, and the amount of sintering feeding in the hours is accumulated to obtain the feeding information with a uniform frequency. The quality-inspection data mainly include the physical and chemical information of raw

materials and fuels in the sintering production process, and the frequency of recording is 2–4 times for each shift. The time of entering into the data system is not fixed, and the method of processing the whole hourly frequency involves updating the data according to the data entry time; that is, the latest data are determined as the next hourly value.

According to the above processing method, all sintering data points are sorted using Python tools to obtain hour-level frequency production datasets.

2.2.3. Data Derivation

In the daily production process of sintering, a large number of processed derivative data are required to guide production, by including the vertical sintering speed and air permeability index. Although some derived data exist in the database records, they are all recorded by field staff after manual calculation based on transient data, with poor quality and irregular frequency, which cannot guarantee the needs of data analysis. The hour-level derivative characteristic parameters are established by using the real-time sintering data and sintering ironmaking metallurgical theory to facilitate subsequent data analysis and application.

2.3. Feature Selection

Feature selection involves selecting an optimal subset from a raw data feature space, which can improve practicability, stability, and accuracy. The filtered mode is independent of subsequent learners, and the dataset is filtered before training the learner. In this study, the correlation and causality of the data were used to analyze the influencing factors of sinter quality to screen the characteristic parameters of the predictive model.

For correlation analysis, Pearson's coefficient was utilized as a measure of the linear correlation between two parameters, and Spearman's coefficient was utilized as a measure of the nonlinear correlation between two parameters.

X and Y are two sets of independent and identically distributed data, the number of its elements is N , and the two sets of random variables taken in the i ($1 \leq i \leq N$) values are X_i, Y_i . By sorting X and Y simultaneously in descending or ascending order, two sets of elemental rows, x and y , are obtained, where elements x_i and y_i are the rows of X_i in X and the rows of Y_i in Y , respectively. The differences between the corresponding elements in the set x, y form the ranked difference set d , where $d_i = x_i - y_i$, ($1 \leq i \leq N$). The Spearman correlation coefficient between X and Y can be calculated from x, y , or d . r_s is calculated as shown below.

$$r_s = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \sum_{i=1}^N (y_i - \bar{y})^2}} \quad (8)$$

The Spearman's correlation coefficient takes values in the range of $[-1, 1]$, and the larger the absolute value of r_s , the stronger the correlation. When the Spearman correlation coefficient $r_s > 0$, the two groups of variables under discussion are considered to have a positive correlation. When the Spearman correlation coefficient $r_s < 0$, the two groups of variables under discussion are considered to have a negative correlation.

For causal analysis, Granger causal analysis was used to analyze the sinter quality parameters. The Granger causality test is used to test whether one set of time series is the cause of another set of time series. If A is said to be the Granger cause of B , it means that the change in A is one of the causes of the change in B .

If the prediction of variable Y is better than the prediction of Y using the past information of Y alone, i.e., variable X helps to explain the future change in variable Y , then variable X is considered to be the Granger cause of variable Y . Sintering production is a process of continuous production, and its production parameters are characterized by nonlinearity, coupling, and time lag, which meet the conditions of the Granger causality test. It is worth noting that the conclusion that the Granger causality test does not hold

does not mean that there is no causal relationship between X and Y . Therefore, if the same parameter is selected by these multiple methods, it is more likely to have a good interpretation. The steps of Granger's causality test are as follows [16,17].

- (1) Regress the current y on all lags of y and whatever other variables (if such a variable exists), i.e., y on the lags of $y_{t-1}, y_{t-2}, \dots, y_{t-q}$ and other variables, but do not include the lag x in this regression, which is a constrained regression. This regression then yields the constrained residual sum of squares RSS_R .
- (2) Completing a regression with a lag term x , i.e., adding the lag term x to the previous regression equation, is an unconstrained regression, from which the regression yields the unconstrained residual sum of squares RSS_{UR} .
- (3) The null hypothesis is that $H_0: \alpha_1 = \alpha_2 = \dots = \alpha_q = 0$, i.e., the lag term x is not part of this regression.
- (4) To test this hypothesis, an F-test was used.

$$F = \frac{(RSS_R - RSS_{UR})/q}{RSS_{UR}/(n - k)} \quad (9)$$

This test follows an F-distribution with degrees of freedom q and $(n - k)$. n is the sample size; q is equal to the number of lagged terms x , i.e., the number of parameters to be estimated in the regression equation with constraints; and k is the number of parameters to be estimated in the unconstrained regression.

- (5) The null hypothesis is rejected if the critical $F \alpha$ value of F-value speculation is calculated at the selected level of significance α , such that the lagged x term belongs to this regression, indicating that x is the cause of y .

The above three commonly selected parameters are used as input parameters to satisfy the necessity of multi-angle relationship analysis while adding some manual empirical data.

2.4. Model Principle

The sintering production process is complex, and the feature-selection results are selected based on multiple methods, so it is unknown which method is more efficient for model prediction. Additionally, in the data-mining process, the generalization of a single model is often weak, while the model fusion method can combine the advantages of multiple models and improve prediction accuracy. Therefore, we selected multiple base models to make preliminary predictions of the selected parameters and used the stacking learning method as the combination strategy of the base model to make a secondary prediction of the prediction results of the previous stage and obtain the optimal prediction results. In this study, support vector machines, random forests, GBDT, and neural networks were selected as the base models as the first layer learners. A linear regression algorithm was used as the second layer learner. To illustrate the superiority of stacking models, this paper used support vector machines, random forests, GBDT, and neural networks as control models to compare the prediction results of the above single models and integrated models for validation.

As a stacking model that relies on the results of multiple base models, the stacking model generally outperforms a single strong model [18,19]. The main principles are shown in Figure 2.

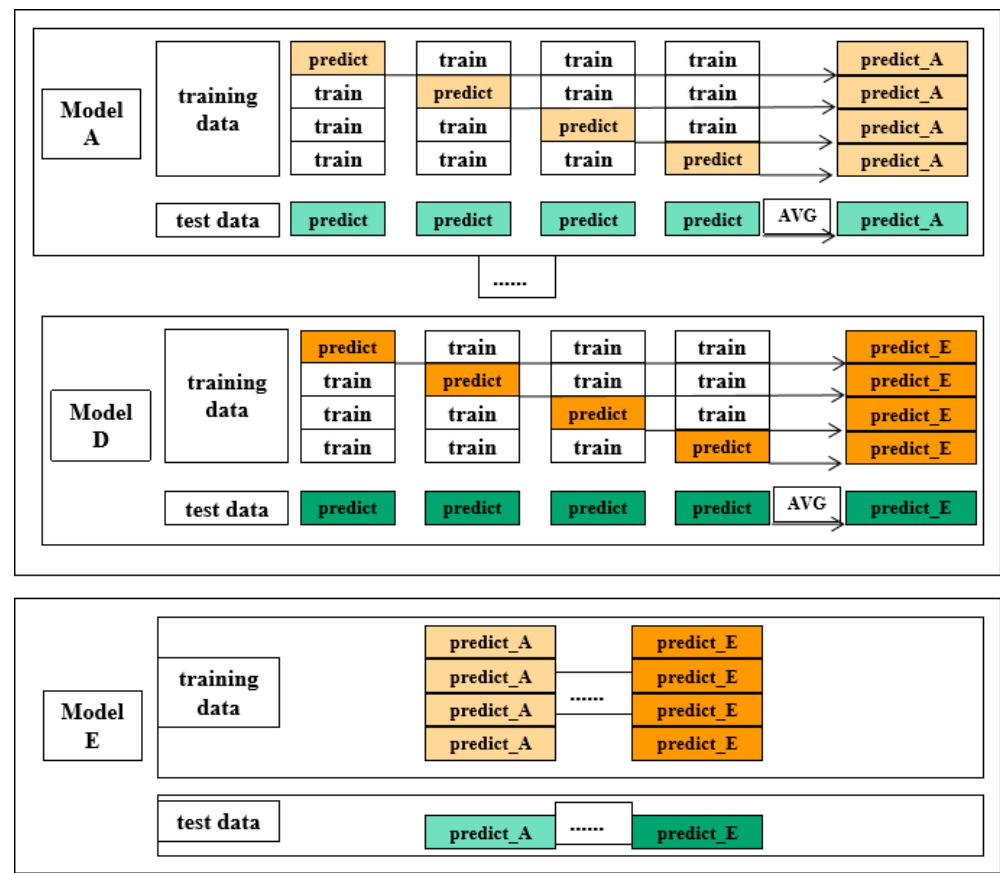


Figure 2. Stacking model flow chart.

- (1) Divide the training set data into five parts; each part should include a validation set and a test set. Use one as the validation set and the rest as the training set.
- (2) Iterating through the five base models proposed will produce one prediction for each part of the test set; each sample will produce five prediction results, and the average of the five results can be taken. The prediction results of each model on the training set can be combined as the new features of the next layer.
- (3) The learning model can fit the dataset after combining new features and target parameters, and the learned model can be used for prediction tests to obtain the best results.

3. Model Building

3.1. Sintering State Division

The sintering process detection parameter data can be regarded as containing only a certain data point, so the problem of dividing and identifying the sintering state is the classification of data points. On short time scales, only using the value of BTP as the basis for the condition classification will omit some important information. Therefore, the burning through temperature (BTT) is used as a supplementary parameter to participate in the division of the sintering state. The BTT is the highest point temperature of the fitted curve for different airbox exhaust gas temperatures. As the mixture is completely burned through, it brings a significant amount of heat to the exhaust gas, leading to an increase in the exhaust gas temperature of the blast box. Then, the level of the maximum temperature (BTT) generated when the mixture is burned through can also reflect the sintering process's working condition. In normal working conditions, the BTP is in the desired range, and the BTT is greater than the expected threshold. In the underburning condition, the BTP is greater than the desired upper bound, or the BTT is less than the desired threshold. Under the overburning condition, the BTP is less than the expected lower bound, and the BTT is greater than the expected threshold. Considering the BTP as an important factor affecting

the sinter quality, a method for predicting the sinter quality based on the BTP is proposed. The desired range of BTP is determined by the sintering machine design process and ranges from the location of the No. 20 to No. 21 bellows. The flue gas temperature threshold of BTP is determined by the production process experience and is in the range of 320 °C to 360 °C. The specific definitions of the sintering states are shown in Table 2.

Table 2. The specific definitions of the sintering states.

Definition	BTP	Sintering Temperature	Relationship
overburning	$BTP < \text{No. 20}$	$360\text{ }^{\circ}\text{C} < T$	AND
normal	$\text{No. 20} \leq BTP \leq \text{No. 21}$	$320\text{ }^{\circ}\text{C} \leq T \leq 360\text{ }^{\circ}\text{C}$	AND
underburning	$\text{No. 21} < BTP$	$T < 320\text{ }^{\circ}\text{C}$	OR

Taking the drum index and TFe as examples, the distribution of the drum index in different sintering states is shown in Table 3. Taking the drum index as an example, the fine adjustment of the proportion of raw materials, the change in the working parameters of the sintering machine, and various physical and chemical reactions in the sintering process lead to differences in the quality of the sinter. In the range of parameters covered in this study, the three sintering states account for different proportions, and most of them belong to the normal sintering state. In addition, the degree of adaptation of different sinter quality parameters to the sintering state is different. By comparing the amount of data for the quality parameters in the same state, it can be seen that the transfer index is more sensitive to the change in the sintering state. Meanwhile, by analyzing the distribution ranges of sintered ore quality parameters in different sintering states, it can be seen that the distribution ranges of drum index and total iron content are in an abnormal state and far from the normal distribution range in the abnormal sintering state. Therefore, underburning and overburning can seriously affect the quality of the sintered ore. The accuracy of sinter quality prediction cannot be guaranteed due to the small number of parameters and the wide parameter distribution in the abnormal sintering condition. Therefore, the prediction of sinter quality should be kept within the range of normal data. In this study, only the sintered ore quality under normal sintering conditions is predicted.

Table 3. Distribution of drum index in different sintering states.

	Data Volume Share of Drum Index	Drum Index Distribution Range	TFe Distribution Range
Overburning	7.25%	[75.7, 78.7]	[53.5, 56.5]
Normal	86.30%	[76.8, 77.5]	[54.5, 55.8]
Underburning	6.45%	[74.8, 78.5]	[54.5, 55.8]

3.2. Feature Selection Results

Take the drum index as an example, the top 10 parameters with numerical absolute value ranking were presented by Pearson correlation analysis, Spearman correlation analysis, and Granger causality analysis, as shown in Table 4. In the Granger causality test results, the smoothness test and cointegration relationship test were completed for the final selected data, which are consistent with the operation of the Granger causality test.

Table 4. Results obtained by different feature selection methods.

	NO.	Pearson	Value	Spearman	Value	Granger Causality	Value
Drum Index	1	Screening index	−0.437	Screening index	−0.500	Blast volume	0.621
	2	No. 1Air box vacuum degree	+0.287	1Air box vacuum degree	+0.344	No. 5Air box vacuum degree	0.589
	3	CaO	+0.279	Particle size less than 10 mm	−0.323	Titanium iron concentrates	0.531
	4	BTP	−0.270	Sintering return mines	+0.287	No. 1Damper opening	0.498
	5	Magnesium lime powder	−0.270	No. 5Air box vacuum degree	−0.278	Vanadium titanium iron powder	0.456
	6	Mixed material CaO content	+0.255	No. 11Air box vacuum degree	−0.278	BTP	0.448
	7	No. 5Air box vacuum degree	−0.254	Titanium iron concentrates	+0.278	Sintering return mines	0.431
	8	No. 3Air box vacuum degree	+0.243	BTP	−0.266	Screening index	0.402
	9	lime powder	+0.243	Mixed material CaO content	−0.265	Lime powder	0.387
	10	Particle size less than 10 mm	−0.225	No. 2Damper opening	0.243	Mixed material CaO content	0.377

As seen in Table 4, there are differences in the feature parameters obtained by the three different feature selection methods, of which there are four identical parameters, accounting for 40%. Different parameters were selected by different feature-selection methods, and the parameters selected by two or more methods were used as input parameters for the prediction model. Meanwhile, as seen in the sintering mechanism analysis, the influencing factors with a strong correlation with the drum index are mainly vanadium and titanium iron concentrates, lime powder, fuel, return mines, blast volume, airbox exhaust gas temperatures, BTP, and mixed material CaO content. The characteristic importance score of the nine-roller velocity on the drum index is also relatively high, indicating that the fabric speed of the mixture on the sinter also influences the drum index. In addition, an important influence of the drum index is the amount of carbon assigned, but this is not provided in the original dataset.

The final feature-selection results for TFe and alkalinity are shown in Table 5. As can be seen in Table 5, there are differences in the input characteristic parameters when predicting different sinter quality parameters.

Table 5. Final feature selection results for TFe and alkalinity.

TFe				Alkalinity			
1	Mixes_(SiO ₂)	7	Mixed material SiO ₂ content	1	Trolley material thickness	7	Water addition rate of second mix
2	Mixes_(CaO)	8	Fuel ratio	2	Sintering machine speed	8	FeO
3	Mixes_(MgO)	9	No. 2Damper opening	3	Water addition rate of first mix	9	CaO
4	Mixes_(Al ₂ O ₃)	10	Ring cooler speed	4	Mixture temperature	10	Coal powder proportion
5	Percentage of returned mine	11	Blast volume	5	Mixed material SiO ₂ content	11	BTP
6	BTP	12	Moisture rate	6	Mixed material CaO content	12	Blast volume

3.3. Evaluation Function

Model prediction effects were demonstrated using evaluation metrics, such as mean absolute error (MAE) and mean square error (MSE). Indicators are defined as shown in Equations (10) and (11).

$$\text{MAE} = \frac{1}{m} \sum_{i=1}^m |y_i - \hat{y}_i| \quad (10)$$

$$\text{MSE} = \frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2 \quad (11)$$

In the above formula, m is the sample size, y_i is the true value, $\hat{y}_i$ is the predicted value, and $\bar{y}_i$ is the mean value.

MAE indicates the mean of the absolute error between the predicted and true values. MSE indicates the degree of deviation between the true and predicted values of the sample as a whole.

The following two criteria were used to evaluate the drum index forecasting model in this study: the root mean square error (RMSE) and the coefficient of determination (R^2). The evaluation methods are Equation (12) and Equation (13), respectively.

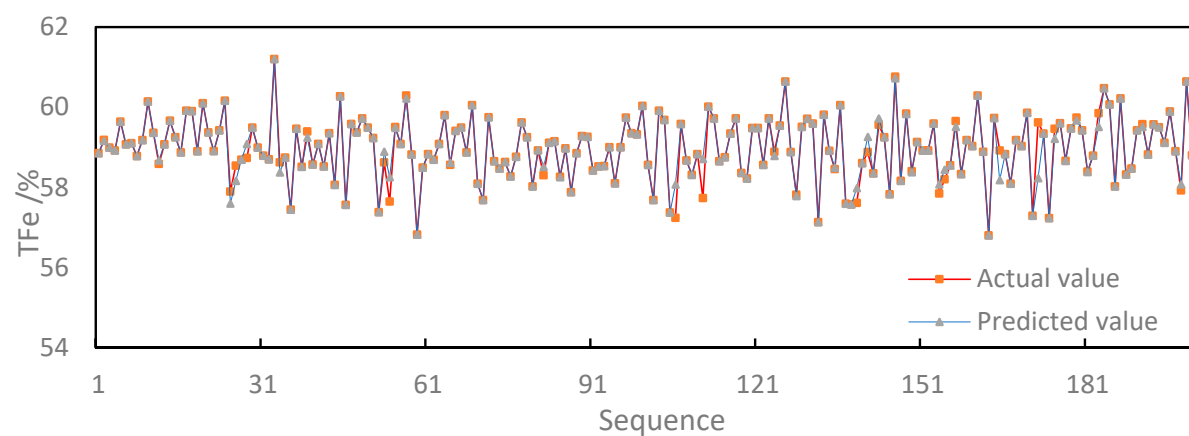
$$\text{RMSE} = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2} \quad (12)$$

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (\bar{y}_i - \hat{y}_i)^2} \quad (13)$$

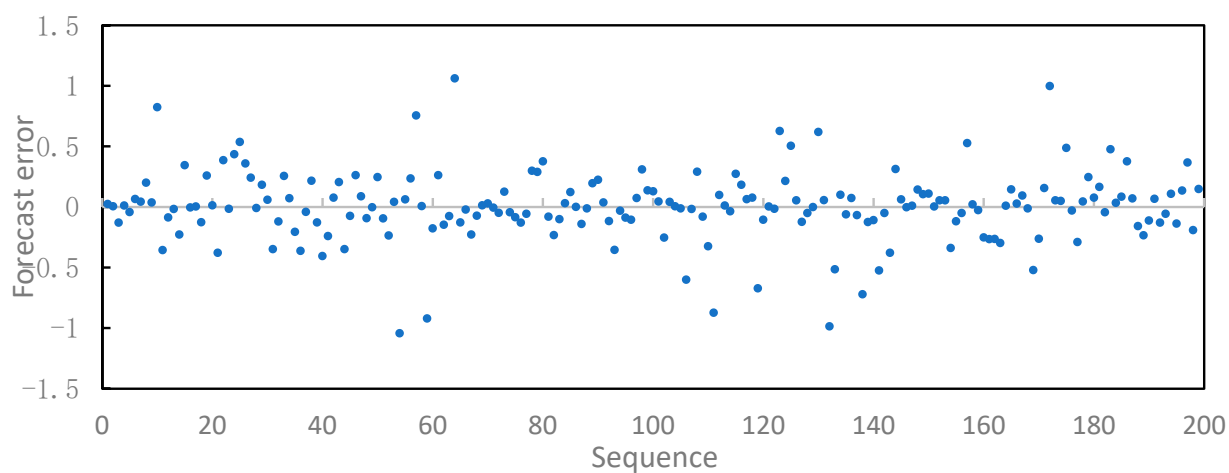
RMSE indicates the deviation between the predicted and true values. The smaller its value is, the better the fit is. The R^2 coefficient indicates the quotient of the sum of squared residuals and the total sum of squares subtracted from 1. The range of values is $[0, 1]$, and the larger the R^2 , the better the model fit.

4. Prediction Model Results

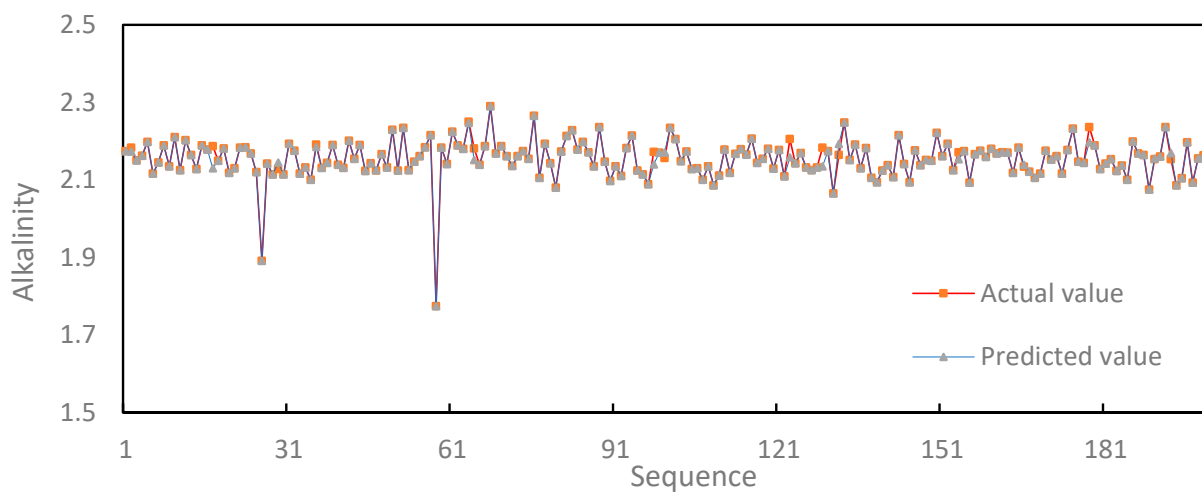
Clean input datasets were obtained using the above data preprocessing methods and feature-selection methods. The input datasets were arranged in chronological order, and the top 90% of the data for each of the three target parameters were selected as the training set for model optimization and training, while the remaining 10% of the data was used as the test set for model prediction result analysis. The prediction model for different sinter quality parameters was based on the stacking algorithm. The optimal combination of hyperparameters adapted to different prediction models was determined by grid search and five-fold cross-validation, then each prediction model was adjusted to the optimal state, and RMSE was applied to optimize the loss function. The predicted results for the three different sinter quality parameters are shown in Figure 3. As can be seen in Figure 3, an accurate prediction of sinter quality parameters can be achieved using the stacking algorithm. The prediction error of TFe is distributed within ± 1.5 , the prediction error of alkalinity is distributed within ± 0.15 , and the prediction error of drum index is distributed within ± 0.6 . The error range of each parameter can meet the actual production requirements. At the same time, the prediction error distribution of different parameters shows that the prediction error of TFe is higher than that of alkalinity and the drum index, which is mainly due to the influence of the production environment during the production of sintered ore. The fluctuation range of TFe of sintered ore is relatively large, resulting in a slightly larger error distribution of the model than the other two parameters, but it can still meet the practical needs of guiding sinter production.



(a)



(b)



(c)

Figure 3. *Cont.*

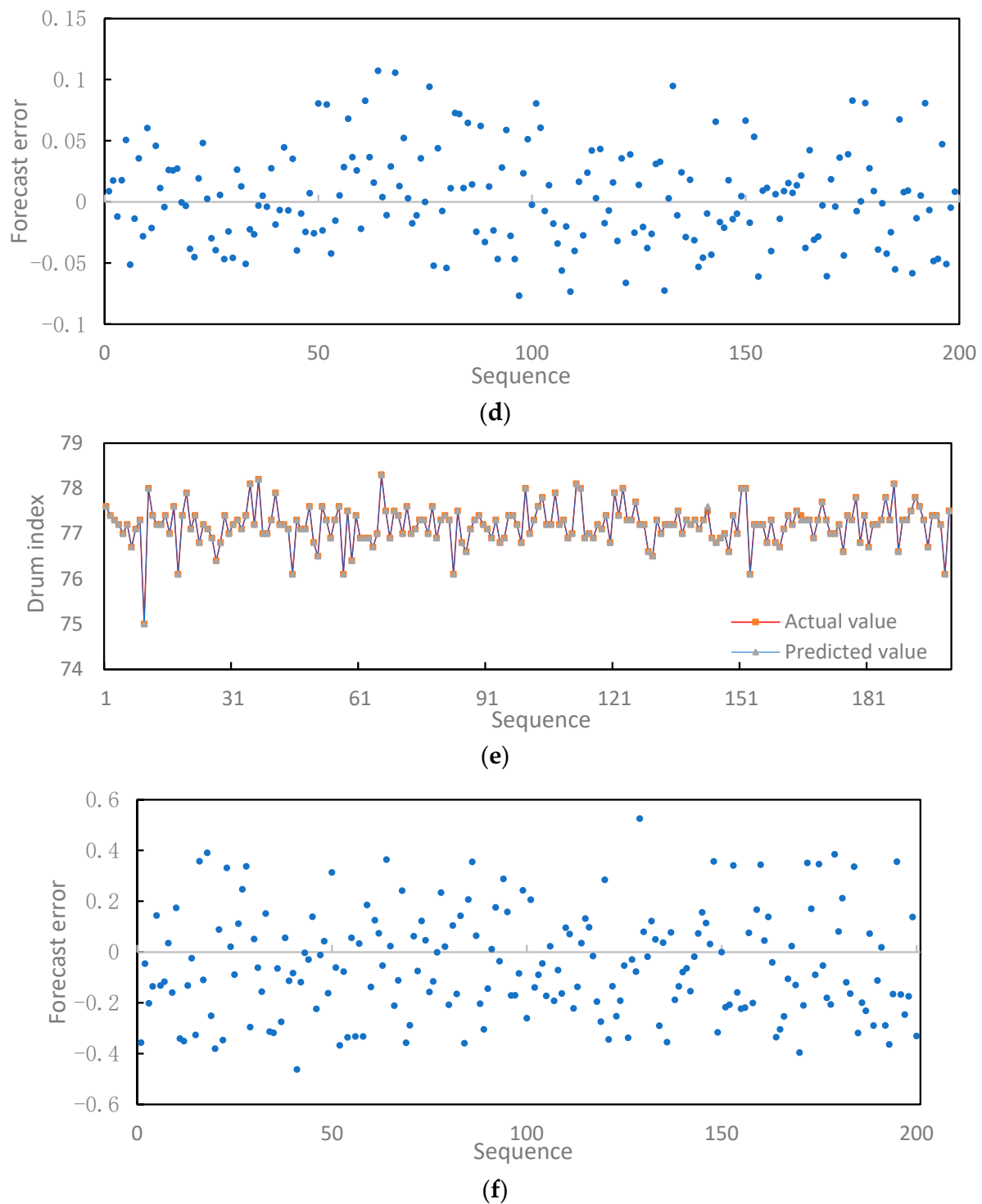


Figure 3. Prediction results of three different sinter quality parameters. (a) TFe prediction result. (b) TFe prediction error distribution. (c) Alkalinity prediction result. (d) Alkalinity prediction error distribution. (e) Drum index forecast results. (f) Drum index prediction error distribution.

To further validate the superiority of the stacking integrated model, different prediction models were developed for the three sinter quality parameters using the base learner as a control group. The evaluation parameters of each model are shown in Table 6. The R^2 of the Stacking model is much higher than that of the other models. Although the modeling time for the Stacking model is slightly longer than the other models, it is acceptable because

it is a stacked model. Therefore, using the Stacking model achieves better performance than other models and can be used as an online sinter ore quality prediction model.

Table 6. Evaluation results of prediction models for three sinter quality parameters.

Sinter Quality Parameters	Evaluation Indicators	SVM	Random Forest	GBDT	Neural Network	Stacking
TFe	MSE	0.332	0.075	0.368	0.068	0.038
	RMSE	0.576	0.274	0.607	0.261	0.195
	MAE	0.447	0.177	0.470	0.168	0.058
	R ²	0.498	0.827	0.443	0.827	0.942
	Test time/s	13.866	20.145	28.508	8.387	35.138
Alkalinity	MSE	0.003	0.002	0.003	0.002	0.001
	RMSE	0.052	0.047	0.059	0.030	0.010
	MAE	0.046	0.032	0.035	0.021	0.002
	R ²	0.512	0.621	0.487	0.583	0.958
	Test time/s	2.221	5.095	1.445	0.154	6.081
Drum index	MSE	0.170	0.007	0.065	0.012	0.003
	RMSE	0.412	0.084	0.255	0.095	0.054
	MAE	0.277	0.031	0.198	0.045	0.005
	R ²	0.268	0.870	0.720	0.821	0.987
	Test time/s	16.969	14.326	5.077	6.016	34.441

5. Sinter Quality Prediction System Implementation

The sinter quality prediction system mainly relies on a Linux server, Webstorm2021, Xshell7, and other software as the development environment, and its establishment can mainly be divided into the following steps.

- (1) Front-end design. The front-end page was developed using the react open-source framework released by Facebook, which helps to create an interactive UI. During the development phase, the framework supports many developers in receiving adequate technical support, and the design process components can be reused. Once the design is complete, if a panel needs to be added or deleted, it can be easily modified in a widget and indexed simply. The framework has sufficient superiority both from the design and maintenance points of view.
- (2) Back-end design. The back end is developed using an Express framework based on the Node platform. Using Express greatly reduces the number of code functions, and the logic is more concise, improving development efficiency and reducing engineering maintenance costs. The development process registers routes with web pages to provide path requests to different modules, avoiding the massive path problem. In addition, middleware modules developed for specific routes can be reused, solving the problem of interleaved references to complex logic. In front- and back-end interaction, we can combine redux to globally control the state in react and use ajax to read data from the server and store data in the action of redux
- (3) Database design. The database is MYSQL, where multiple tables, such as user tables, are connected by foreign keys, and suitable indexes are also created within the tables to speed up the table query process.
- (4) Deployment. The project is deployed on a Linux server; the front-end is deployed through the nginx proxy server while being able to ensure load balancing. The back end is deployed on the server by packaging the project into a jar package, so the front and back ends can be deployed separately.

The sinter quality prediction system built according to the above process is shown in Figure 4. The system interface mainly includes sinter quality parameter display, sinter quality index prediction, and core parameter monitoring. The parameter display function provides manual monitoring results and the timing of sinter quality. The parameter prediction function provides the predicted results of the sinter quality for the next two

hours and displays the historical monitoring results and predicted results in the form of line graphs, making it easy for the operator to intuitively grasp the trend of sinter quality changes and the accuracy of prediction. The status parameter monitoring module is used to display most of the parameters of the BF ironmaking process that are of interest to the operator, such as the hot air pressure and top gas pressure. The creation of this module has greatly improved work efficiency. On the one hand, it is convenient for the field operator to view the status of each parameter in real-time. On the other hand, it enables timely adjustment of the solution strategy according to the trend of the forecast results.

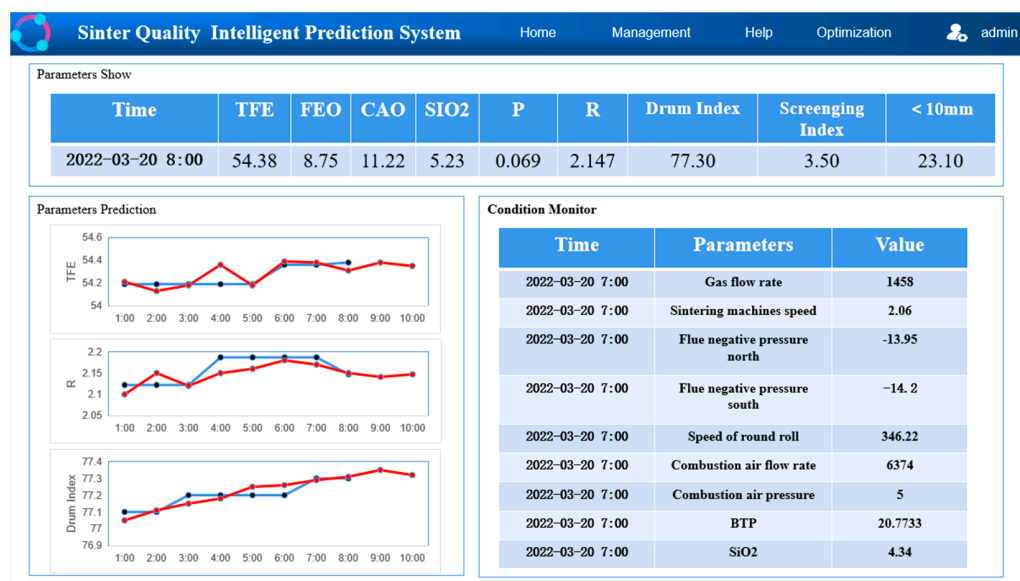


Figure 4. Sinter quality prediction system interface function.

6. Conclusions

To fully utilize the production data of the sintering system, deeply explore the value of the data, and guide production on time, this study combined big data and machine learning to establish an intelligent prediction system for sintered ore quality. The following conclusions were obtained.

- (1) To address the problems of different storage methods, large differences in scale levels and outliers in sintering data, data collection, extraction, integration, standardization, and outlier processing for the sintering system were completed by using a database, Excel, and Python language under the principles of close integration with the sintering process and ensuring data reliability, comprehensiveness, and timeliness. The data collection, extraction, integration, standardization, and outlier processing for the sintering system were completed by using a database, Excel, and Python language. The sintering sample set of Chenggang was constructed in a time-based and standardized format spanning 2 years.
- (2) The Pearson correlation coefficient, Spearman correlation coefficient, and Granger causality coefficient were used to analyze the correlation between the sinter ore quality and sintering process production parameters from different perspectives and realize the feature selection of the prediction model by combining the experience of sintering experts. The selected characteristic parameters are the actual data of the sintering process, including airbox exhaust gas temperatures, fuel ratio, Sintering machine speed, etc. And the selected input parameters can meet the requirements of online calculation of the prediction model.
- (3) A sinter quality prediction model was established using a stacking integrated learning algorithm, which could accurately predict the indices of TFe, R, and the drum index of sinter ore. The prediction errors of different parameters could meet the accuracy requirements for guiding production.

- (4) Based on Webstorm2021 and Xshell7 as the development environment, the sinter quality prediction system was built and tested online using the existing platform of the plant. The results showed that the operation status was stable during the test period, and the prediction results met the expected results.

Author Contributions: Conceptualization, X.L. (Xin Li) and R.L.; methodology, X.L. (Xin Li); software, X.L. (Xin Li), X.L. (Xiaojie Liu) and H.L. (Hongyang Li); validation, X.L. (Xin Li), X.L. (Xiaojie Liu), H.L. (Hongwei Li) and Z.Z.; formal analysis, X.L. (Xin Li); investigation, X.L. (Xin Li); resources, X.L. (Xin Li); data curation, X.L. (Xin Li); writing—original draft preparation, X.L. (Xin Li); writing—review and editing, X.L. (Xin Li); visualization, X.L. (Xin Li); supervision, Z.Z., Q.L. and L.W.; project administration, X.L. (Xin Li); funding acquisition, X.L. (Xin Li). All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Hebei Basic Research Projects of Higher Education Institutions (NO. JQN2020032), National Nature Science Foundation of China (NO. 52004096), Hebei Province High-end Iron and Steel Metallurgical Joint Research Fund Project (NO. E2020209208).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: Thanks to the North China University of Science and Technology for training and educating, thanks to my teacher Liu Ran for his careful guidance. And thanks to classmates in my team for their companionship.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zhang, J.W.; Niu, L.Q. The practical exploration and enlightenment of Shougang Group's green and low-carbon development in the century. *N. Econ. Guide* **2021**, *2*, 41.
2. Wu, W.N.; Lu, Q.; Wan, X.Y. Effect of base characteristics of iron ore fines on the strength of sintering drums in Chengde Steel. *Iron Steel* **2013**, *48*, 12.
3. Kawaguchi, T.; Yoshinaga, M. Development and application of an integrate simulation model for iron ore Sintering. *Ironmak. Proceeding* **1987**, *12*, 99–106.
4. Hamada, K.; Matoba, Y.; Murai, T.; Ueno, Y.; Sato, K. Control system of chemical composition of iron ore sinter. *Trans. Iron Steel Inst. Jpn.* **1986**, *84*, 409–419.
5. Peng, Q.K. *Simulation Model of Sinter Bed Temperature Field and Experts System of Sinter Quality Optimization*; Central South University: Changsha, China, 2011.
6. Zhao, D.K.; Wang, J.; Wang, H. Control system of quality supervision and operating guidance in sintering production based on expert system. *Metall. Ind. Autom.* **2004**, *2*, 26–29.
7. Wang, Q.Y. *Research on Prediction Model of Tumbler Strength Based on the Online Inspection of Sinter Composition*; Wuhan University of Science and Technology: Wuhan, China, 2017.
8. Wang, Q.Y.; Liu, Q. Research on online prediction of sinter quality based on Elman Neural Network. *Instrum. Tech. Sens.* **2017**, *98–100*, 104.
9. Xin, Z.C.; Li, J.; Liu, W.X. Forecast for low temperature reduction disintegration properties of vanadium-titanium sinter based on BP Neural Network. *Iron Steel Vanadium Titan.* **2017**, *38*, 94–99.
10. Wang, D. *Application Research Based on GA-FWA in Prediction of Sintering Burning through Point*; Liaoning University of Science and Technology: Anshan, China, 2019.
11. Li, H.Y.; Li, X.; Liu, X.J.; Bu, X.; Li, H.; Lyu, Q. Prediction of blast furnace parameters using feature engineering and Stacking algorithm. *Ironmak. Steelmak.* **2022**, *49*, 283–296. [[CrossRef](#)]
12. Chen, S.; Jin, G.; Peng, S.; Zhang, L. Anomaly Detection Method Based on Granger Causality Modeling. In *Proceedings of the International Conference on Wireless and Satellite Systems, Nanjing, China, 17–18 September 2020*; Springer: Cham, Switzerland, 2021; pp. 145–151.
13. Najjar, I.R.; Sadoun, A.M.; Fathy, A.; Abdallah, A.W.; Elaziz, M.A.; Elmahdy, M. Prediction of Tribological Properties of Alumina-Coated, Silver-Reinforced Copper Nanocomposites Using Long Short-Term Model Combined with Golden Jackal Optimization. *Lubricants* **2022**, *10*, 277. [[CrossRef](#)]
14. Sadoun, A.M.; Najjar, I.M.R.; Fathy, A.; Abd Elaziz, M.; Al-qaness, M.A.; Abdallah, A.W.; Elmahdy, M. An enhanced Dendritic Neural Algorithm to predict the wear behavior of alumina coated silver reinforced copper nanocomposites. *Alex. Eng. J.* **2023**, *65*, 809–823. [[CrossRef](#)]

15. Sadoun, A.M.; Wagih, A.; Fathy, A.; Essa, A.R.S. Effect of tool pin side area ratio on temperature distribution in friction stir welding. *Results Phys.* **2019**, *15*, 102814. [[CrossRef](#)]
16. Cao, Y.F. A Comment on Granger Causality Test. *J. Quant. Tech. Econ.* **2006**, *1*, 155–160.
17. Liu, J.; Zhao, H.Y.; Liu, J.C.; Pan, L.; Wang, K. Medium-term load forecasting based on cointegration-Granger causality test and seasonal decomposition. *Autom. Electr. Power Syst.* **2019**, *43*, 73–80. [[CrossRef](#)]
18. Cui, G.Q. *Feature Selection and Regression Prediction in Complex Dataset*; Shanghai Jiao Tong University: Shanghai, China, 2017.
19. Song, J.; Wang, W.L.; Li, D. Injection molding part size prediction method based on Stacking ensemble learning. *J. South China Univ. Technol. (Nat. Sci. Ed.)* **2022**, *50*, 19–26.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.