

Article

Big Data as a Tool for Building a Predictive Model of Mill Roll Wear

Natalia Vasilyeva ^{1,*}, Elmira Fedorova ¹ and Alexandr Kolesnikov ²

¹ Department of Economics, Organization and Management, Saint Petersburg Mining University, 199106 St. Petersburg, Russia; fedorova_er@pers.spmi.ru

² Non-Profit Joint Stock Company «M. Auezov South Kazakhstan University», Shymkent 160012, Kazakhstan; kas164@yandex.kz

* Correspondence: vasileva_nv@pers.spmi.ru; Tel.: +7-921-441-51-78

Abstract: Big data analysis is becoming a daily task for companies all over the world as well as for Russian companies. With advances in technology and reduced storage costs, companies today can collect and store large amounts of heterogeneous data. The important step of extracting knowledge and value from such data is a challenge that will ultimately be faced by all companies seeking to maintain their competitiveness and place in the market. An approach to the study of metallurgical processes using the analysis of a large array of operational control data is considered. Using the example of steel rolling production, the development of a predictive model based on processing a large array of operational control data is considered. The aim of the work is to develop a predictive model of rolling mill roll wear based on a large array of operational control data containing information about the time of filling and unloading of rolls, rolled assortment, roll material, and time during which the roll is in operation. Preliminary preparation of data for modeling was carried out, which includes the removal of outliers, uncharacteristic and random measurement results (misses), as well as data gaps. Correlation analysis of the data showed that the dimensions and grades of rolled steel sheets, as well as the material from which the rolls are made, have the greatest influence on the wear of rolling mill rolls. Based on the processing of a large array of operational control data, various predictive models of the technological process were designed. The adequacy of the models was assessed by the value of the mean square error (MSE), the coefficient of determination (R^2), and the value of the Pearson correlation coefficient (R) between the calculated and experimental values of the mill roll wear. In addition, the adequacy of the models was assessed by the symmetry of the values predicted by the model relative to the straight line $\hat{Y}_{\text{predicted}} = Y_{\text{actual}}$. Linear models constructed using the least squares method and cross-validation turned out to be inadequate (the coefficient of determination R^2 does not exceed 0.3) to the research object. The following regressions were built on the basis of the same operational control database: Linear Regression multivariate, Lasso multivariate, Ridge multivariate, and ElasticNet multivariate. However, these models also turned out to be inadequate to the object of the research. Testing these models for symmetry showed that, in all cases, there is an underestimation of the predicted values. Models using algorithm composition have also been built. The methods of random forest and gradient boosting are considered. Both methods were found to be adequate for the object of the research (for the random forest model, the coefficient of determination is $R^2 = 0.798$; for the gradient boosting model, the coefficient of determination is $R^2 = 0.847$). However, the gradient boosting algorithm is recognized as preferable thanks to its high accuracy compared with the random forest algorithm. Control data for symmetry in reference to the straight line $\hat{Y}_{\text{predicted}} = Y_{\text{actual}}$ showed that, in the case of developing the random forest model, there is a tendency to underestimate the predicted values (the calculated values are located below the straight line). In the case of developing a gradient boosting model, the predicted values are located symmetrically regarding the straight line $\hat{Y}_{\text{predicted}} = Y_{\text{actual}}$. Therefore, the gradient boosting model is preferred. The predictive model of mill roll wear will allow rational use of rolls in terms of minimizing overall roll wear. Thus, the proposed model will make it possible to redistribute the existing work rolls between the stands in order to reduce the total wear of the rolls.



Citation: Vasilyeva, N.; Fedorova, E.; Kolesnikov, A. Big Data as a Tool for Building a Predictive Model of Mill Roll Wear. *Symmetry* **2021**, *13*, 859. <https://doi.org/10.3390/sym13050859>

Academic Editors: Rudolf Kawalla, Beloglazov Ilya and Basil Papadopoulos

Received: 23 April 2021

Accepted: 9 May 2021

Published: 12 May 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 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: big data; rolling mill; rolled steel; rolling mill roll wear; mathematical model; correlation coefficient

1. Introduction

The metallurgical industry is one of the leading sectors of the Russian economy. The products manufactured by this industry are used in construction, mechanical engineering, the chemical industry, and many other industries [1–3].

Rolled steel production is one of the most important items of Russian export. By deforming the metal in the space between the rotating rolls, you can get almost any kind of metal product from steel and other alloys. This process is called metal rolling. One of the major problems of rolled products is the wear of rolls that deform the metal.

In this work, wear refers to qualitative and quantitative changes in the roll surface caused by physical and chemical processes, as well as mechanical effects of one body on another [4–6].

Current trends in the development of metallurgy are characterized by the development and implementation of information systems and technologies, which are based on computers and computer networks with the richest software, as well as database management systems and computer decision support systems, the methodological basis of which is systems theory and systems analysis.

Scientific and technological progress creates prerequisites for improving the quality of management through the use of computer technology, mathematical methods of data processing, control theory, and control automation. All this has found concrete implementation in automated control systems. Owing to the development of information technology (IT), there are modern software products and database management systems (DBMS) for solving production management problems. Modern software and microprocessor technology makes it possible to create high-level control systems with the inclusion of powerful control algorithms.

The relevance of the work is thanks to the fact that the construction of linear and multidimensional regression models based on a large data set does not provide a high-quality result, as it does not allow taking into account complex and multi-connected dependencies between the input variables. In this case, compositional models that are resistant to overtraining, noise, and outliers show themselves in the best way. However, with less data that can be described by a simple model, it makes more sense to use multivariate regression.

The aim of the work is to develop a predictive model of rolling mill roll wear based on a large array of operational control data containing information about the time of filling and unloading of rolls, rolled assortment, roll material, and the time during which the roll is in operation.

To achieve the set objective, it is necessary to solve the following tasks:

1. Prepare data for modeling (filter and aggregate data).
2. Conduct a correlation analysis of the data to identify the factors that have the greatest impact on the wear of the mill rolls.
3. Build various models for predicting mill roll wear (linear models, multidimensional models, and intelligent models). Test their adequacy and identify the most accurate one.

The predictive model of mill roll wear will allow rational use of rolls in terms of minimizing overall roll wear. Thus, the proposed model will make it possible to redistribute the existing work rolls between the stands in order to reduce the total wear of the rolls.

2. Theoretical Basis

In the technical literature, data on the durability and wear of mill rolls are extremely rare. The amount and nature of work roll wear depend on many factors. The main factors are as follows: force, temperature and speed conditions of rolling, properties and amount

of rolled metal, hardness, and diameter of rolls. However, it is extremely difficult to study the individual influence of each factor on roll wear [7,8].

The presence of a large number of factors makes it difficult to obtain dependencies that would take them into account and makes it possible to calculate the wear of the rolls.

Based on the literature review, wear is associated with the number (length) of rolled strips and this dependence is described using empirical equations, the coefficients of which are determined experimentally at each rolling mill. The main disadvantage of these dependencies is that they take into account the influence of a small number of factors and cannot be used when changing the rolling conditions.

The existing theoretical methods are based on determining the path of friction in the deformation zone and contact stresses or on calculating the work of deformation. They are quite complex and lengthy, and often give a high error [9].

Therefore, to assess the wear of mill rolls, it is more convenient to use the methods of statistical analysis and mathematical modeling, which make it possible to use statistical data accumulated during operation to assess the condition and predict further roll behavior. Here, the methods of statistical analysis and mathematical modeling are understood as a certain computational algorithm implemented on computers and simplified simulating of the functioning of objects.

Statistical analysis is divided into three sequential stages [10]:

- Statistical observation, i.e., collection of primary statistical material;
- Summary and development of observation results, i.e., their processing;
- Analysis of the received overall materials.

With the development of Big Data and IIoT technologies, finding dependencies between the parameters of the technological process can provide a company with a greater effect than just methods of statistical analysis.

Big Data and data analysis technologies allow the following [11–13]:

- To find patterns that appear in mass phenomena under the influence of the law of large numbers;
- To systematize and classify data based on similarities and differences;
- To analyze the overall material, identify patterns and relationships in the studied facts, and calculate generalizing indicators (total, relative, and average values, as well as statistical coefficients).

3. Object and Problem Statement

The data of the operational control of the technological process are characterized by a different origin and are measured in different quantitative and qualitative scales. Bringing operational control data to a form suitable for developing a model of a technological process is a prerequisite for the effectiveness of the modeling process [14].

Initial data are presented in five sheets (Figure 1) in a Microsoft Office Excel file. The data contains information about roll material (500 lines), roll workflow for 9 months of rolling mill operation (18,080 lines), roll suppliers (25 lines), and rolled assortment (269,968 lines).

	A	B	C	D	E	F
1	дата_завалки	дата_вывалки	номер_клетки	положение_в_клетки	номер_валка	
2	01.10.2018 18:15	01.10.2018 21:46	12	низ	199	
3	01.10.2018 18:15	01.10.2018 21:46	12	верх	445	
4	01.10.2018 18:15	01.10.2018 21:46	11	верх	284	
5	01.10.2018 18:15	01.10.2018 21:46	10	низ	172	
6	01.10.2018 18:15	01.10.2018 21:46	11	низ	275	

(a) Sheet 5

	A	B	C	D	E	F	G
1	дата_завалки	дата_вывалки	номер_клетки	положение_в_клетки	номер_валка	износ	
2	01.01.2018 0:14	01.01.2018 3:32	8	верх	222	0,684804912	
3	01.01.2018 0:14	01.01.2018 3:32	8	низ	260	0,614913529	
4	01.01.2018 0:14	01.01.2018 3:32	9	верх	410	1,135412151	
5	01.01.2018 0:14	01.01.2018 3:32	9	низ	57	0,516086546	
6	01.01.2018 0:14	01.01.2018 3:32	10	верх	273	0,657682871	

(b) Sheet 2

Figure 1. Fragment of a file with initial data.

The following were considered as initial data for modeling: minutes (time of rolling of a batch of products); stand number (set by a number); mill stand position (top or bottom); number and material of the roll (in coded form, each of the parameters); the number

of sheets rolled by a certain roll; gauge, width, and weight of the sheet; grade of rolled products; and roll wear.

The column «mill stand position» is problematic, as it contains text data («top»–«bottom»). For convenience, they are encoded with numbers 0 and 1.

To correctly prepare data for the development of a predictive model, you first need to find out the data types presented in the source file and check them for integrity. It is easiest to delete «empty» values, but if there are a lot of them, it makes sense to replace the missing data with some number, for example, the arithmetic average of the entire column.

As a result of the check, it was found that there are no gaps in the columns. In addition, some lines were found to contain zero roll wear after rolling steel. Such records should be disregarded, because, even if such «outliers» are not errors, but are rare exceptional situations, they can still hardly be used [15–17].

Calculation of the difference between filling up and unloading times allows to obtain the roll operating time for one rolled batch. By analyzing the rolling time of coils with the ranges of filling up and unloading of rolls indicated in the «rolls» sheet, it is possible to calculate the average weight, width, gauge, and number of coils rolled through these rolls. The resulting features can be used to build models.

To determine the influence of each investigated factor on roll wear, the Pearson correlation criteria (R) were calculated, characterizing the linear effects of the factors, and a cross-correlation matrix was constructed. With an insignificant value of the coefficient, certain features can be ignored when building models (Table 1).

Table 1. Feature correlation diagram.

	Minutes	Stand Number	Mill Stand Position	Roll Number	Roll Material	Sheets	Gauge	Width	Steel Grades	Weight	Wear
minutes	1										
stand number	0.0037	1									
mill stand position	0.00012	-4×10^{-5}	1								
roll number	0.0089	0.0086	0.0041	1							
roll material	0.009	0.0021	3×10^{-5}	−0.014	1						
sheets	0.87	0.0034	-9×10^{-5}	0.0033	0.0046	1					
gauge	− 0.34	−0.0011	4.2×10^{-5}	−0.003	−0.0069	−0.025	1				
width	− 0.27	0.00049	−0.00085	−0.0026	−0.0035	−0.016	0.35	1			
steel grades	0.12	0.002	−0.00046	0.00047	−0.012	0.16	0.1	−0.032	1		
weight	− 0.16	−0.00079	-4.1×10^{-5}	−0.0091	−0.0045	− 0.23	−0.0021	0.14	−0.087	1	
wear	0.28	− 0.35	0.0011	−0.0029	0.063	0.17	− 0.23	−0.037	0.094	−0.021	1

(significant coefficients are in bold).

Checking the significance of the correlation coefficients according to the Student's test showed that the correlation coefficients are significant, the absolute value of which exceeds 0.1; that is, the condition $|R| \geq 0.1$ must be satisfied.

From the data obtained, it follows that the position of the roll in the stand ($R = 0.0011$) and the serial number of the roll ($R = -0.0029$) do not have a linear effect on the wear of the rolls. In addition, the serial number of the roll (from 1 to 500) is not a technological parameter and is only for informational purposes. The position of the roll in the stand (top or bottom) is also for informational purposes only. These signs will not be taken into account in the construction of the future model.

Despite the fact that such operational parameters as the roll material, width, weight, and grade of rolled steel also do not satisfy the condition $|R| \geq 0.1$, it was decided not to exclude these parameters from consideration.

Thus, the next stage of the study is to develop a predictive model of rolling mill roll wear based on a large array of operational control data containing information about the time of filling and unloading of rolls, rolled assortment, roll material, and time during which the roll is in operation [18].

4. Algorithm

The algorithm for the development of a predictive model of mill roll wear based on a large array of operational control data is presented in Figure 2.

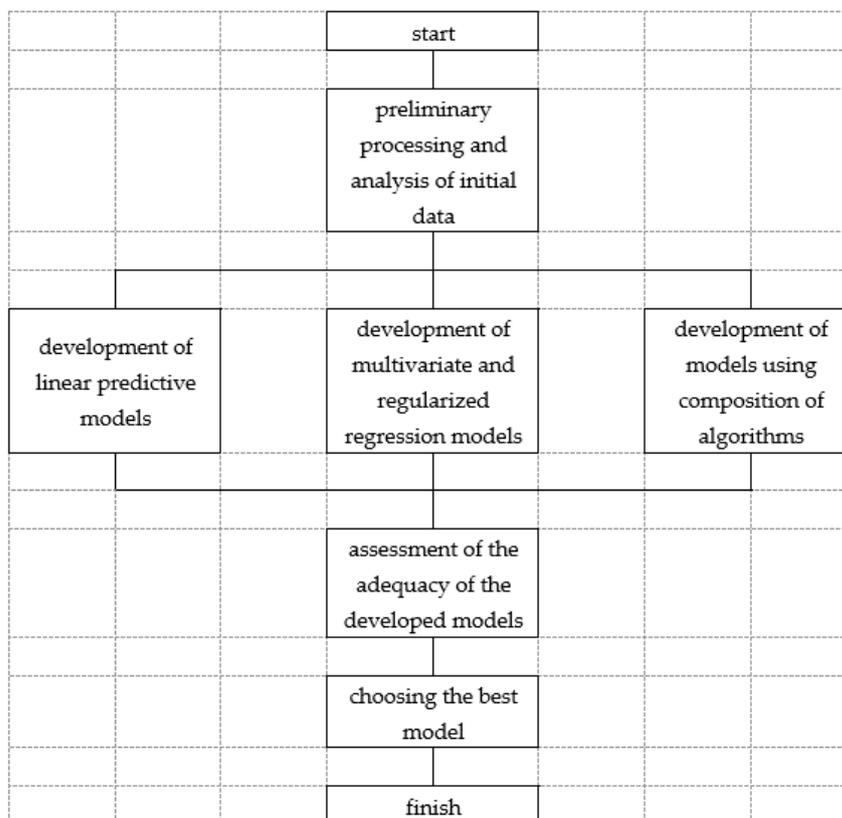


Figure 2. Algorithm for developing a predictive model of mill roll wear.

4.1. Using Big Data to Develop Linear Predictive Models

Cross-validation (CV) and least squares are used to develop a linear predictive model.

The essence of the least squares method is that the sum of the squares of deviations of the experimental values from the smoothing curve is reduced to a minimum:

$$\sum_{i=1}^N [y_i - \varphi(x_i)]^2 = \min$$

where y_i and x_i —experimental data values in the i -th experiment, N —number of experiments, $\varphi(x)$ —desired linear regression y of x of the form $\varphi(x) = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \dots + b_kx_k$, and k —number of factors.

The essence of the CV method is that the entire array of operational control data is divided into a certain number of subsamples (blocks). One of the blocks is used to test the model (check the model for adequacy to the process under study), while the others are used for training. Then, the test block is used for training, and the next block is selected for the test. The cross-validation scheme is shown in Figure 3 (open blocks are model training blocks, filled block is a test subsample). This method allows you to obtain an unbiased estimate of the probability of error in the predictive model and to prevent optimistic overestimation of the quality of the above-mentioned.

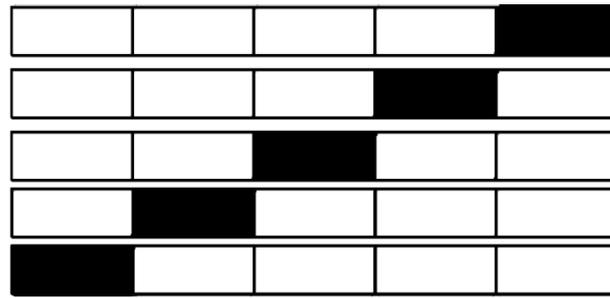


Figure 3. «Cross-validation» operation scheme.

4.2. Using Big Data to Develop Multi-Dimensional and Regularized Regression Models

The essence of regularization is to impose additional constraints on various parameters or to add a priori information, thus reducing the model error as its complexity increases [19,20].

Based on the same operational control database, the following were built: multivariate regression with L1 regulator (Lasso), multivariate regression with L2 regulator (Ridge), and multivariate regression with mixed regulator (ElasticNet).

Regularization is a way to reduce the complexity of a model in order to prevent overtraining or to fix an incorrectly posed problem. This is usually achieved by adding some a priori information to the problem statement.

The essence of L1 regularization is to select from the entire array of factors only a small number of the most important ones that set the trend, and to remove all the rest, which are just noise. Thus, L1 regularization is aimed at decreasing the dimension of the model.

L2 regularization is aimed at reducing the dimension of space by prohibiting disproportionately large weight coefficients, which prevents overtraining of the model.

The development of multivariate regression using both L1 and L2 regularization is called a mixed regulator (ElasticNet) and takes into account the effectiveness of both methods: decreasing the model dimension and decreasing the dimension of the factor space.

4.3. Algorithm Composition for Model Development Based on Big Data

The main method of composing algorithms is to combine a large number of models into one composition. The final quality of the resulting model will be significantly improved owing to the fact that the individual ones will correct the errors of each other.

This study explores such methods as random forest and gradient boosting [21–23].

The random forest method is one of the most professional and high-quality machine learning methods. The key idea of this method for finding regression dependencies is averaging the result of several models built independently of each other on random subsamples of one data array. Thus, a set of low-precision algorithms when combined into one composition give an impressive result, despite the significant amount of randomness represented in this method.

The advantage of the random forest method is its resistance to overfitting. As all algorithms are developed independently of each other, an increase in their number in a composition does not complicate the final model [24,25].

In this study, the random forest algorithm uses feature space dimensionality reduction using principal component analysis (PCA). Using the technique of reducing the dimensionality of the feature space, it is possible to represent the initial data set in terms of fewer variables and, at the same time, reduce the amount of computing resources required to ensure the operation of the model.

Gradient boosting method. The difference between this method and the previous one is that, in this algorithm, when building a composition, all models are not independent, but follow each other. Moreover, each subsequent algorithm tries to correct and compensate for the errors of the previous one. So it takes less time to get the correct answer.

In this study, gradient boosting uses a gradient descent technique to minimize the error function right in these sequential models. This approach makes it possible to expand the range of problems solved by this algorithm, as well as often leading to a gain in prediction accuracy.

4.4. Assessment of the Model Quality

Model quality is assessed using the mean squared error (MSE) between the predicted and actual roll wear, the correlation coefficient (R) between the actual and predicted mill roll wear values, and the determination coefficient (R^2) between the actual and the predicted values of rolling mill roll wear.

The coefficient of determination clearly shows how the constructed model is more accurate than the mean value of the target variable, and is in accordance with the following expression:

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} \approx 1 - \frac{MSE}{VAR(y)}$$

where y_i —actual value of roll wear, \hat{y}_i —model predicted roll wear, and \bar{y} —average roll wear according to the initial data. If the coefficient of determination R^2 is equal to 1, then the values of the rolling mill roll wear calculated by the model exactly repeat the actual values, which indicate the adequacy of the mathematical model to the object of the research. If the coefficient of determination R^2 is close to zero, then this means that the model is imperfect and it would be better to take the average value \bar{y} . Models are recognized as adequate if the coefficient of determination is $R^2 \geq 0.7$.

5. Results

Figure 4 shows the results of comparing the actual and predicted roll wear for different models.

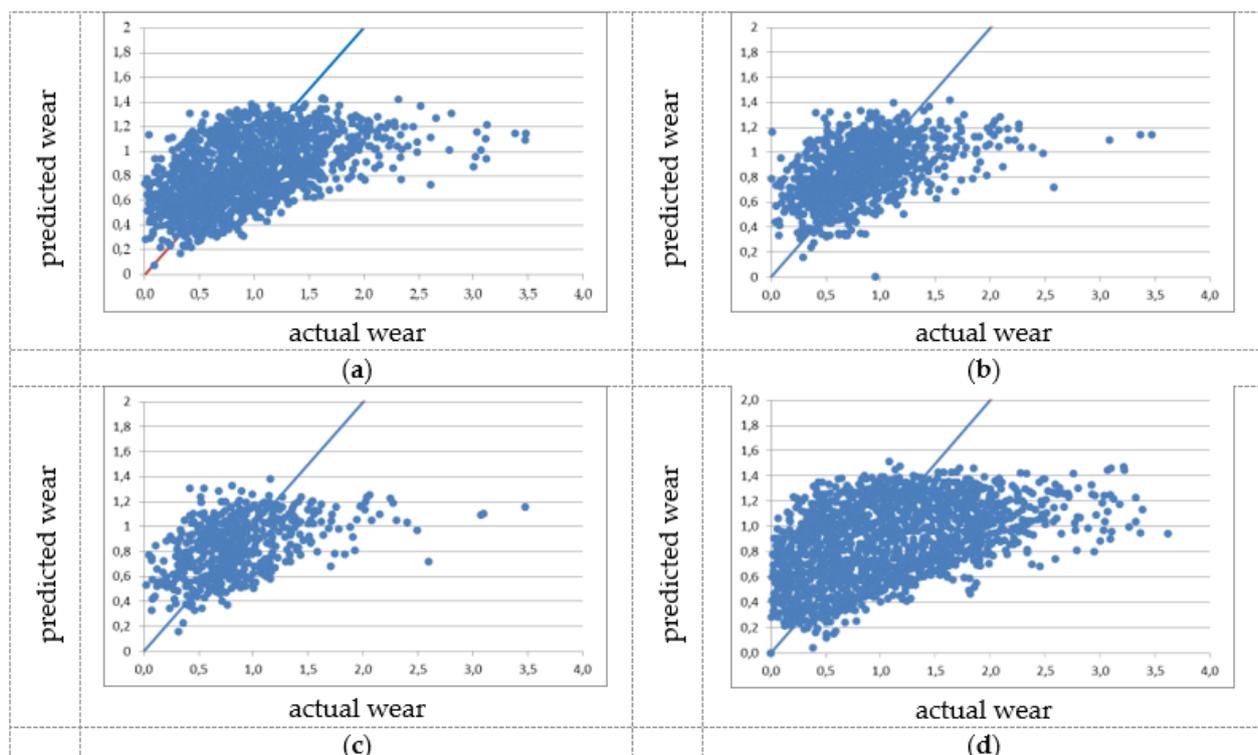


Figure 4. Comparison of predicted and actual rolling mill roll wear: (a) 30% test sample; (b) 10% test sample; (c) 5% test sample; (d) cross-validation.

For clarity, you can compare the models built with and without cross-validation.

Instead of cross-validation, the entire array of operational control data is divided into training and test samples by mixing all the features and choosing a certain percentage between the training and test samples. Linear regression, found by the method of least squares, is used as a model.

Analysis of the graphs (Figure 4) for symmetry regarding the straight line $Y_{\text{predicted}} = Y_{\text{actual}}$ shows that, in all cases, there is an underestimation of the predicted values. With real wear values of 0–4, the predicted values do not exceed 0–1.6.

In this case, the quality of the model changes depending on the amount of data selected for training the model and test validation. More data per test reduces the amount of training data and leads to a decrease in model accuracy, and vice versa [26,27].

The results of assessing the adequacy of the obtained models are shown in Table 2.

Table 2. Assessment of the adequacy of multivariate and regularized regression models. MSE, mean square error.

Model	Assessment	MSE	R ²	R
	30% test sample	0.109	0.257	0.513
	10% test sample	0.112	0.257	0.505
	5% test sample	0.108	0.253	0.503
	Cross-validation	0.113	0.256	0.507
	Lasso	0.113	0.253	0.504
	Ridge	0.113	0.256	0.506
	ElasticNet	0.113	0.256	0.507
	Random forest	0.021	0.798	0.933
	Gradient boosting	0.021	0.847	0.927

Thus, the results of this analysis indicate insignificant differences in the simulation results. All models cannot be considered suitable for predicting the amount of roll wear in a rolling mill. Therefore, it is necessary to choose another type of dependence [28,29].

The introduction of a regularizer into a linear or multidimensional model did not lead to an increase in the accuracy of predicting the wear of the rolling mill rolls. It can be clearly seen that the proposed models predict the value of the target parameter not more accurately than the arithmetic mean of the wear of the rolling mill roll.

Based on the data obtained, it can be stated that, in this case, either rethinking or intellectualization of the initial data is required, or the use of more complex models [30].

A comparison of the predicted by the random forest method and the actual values of rolling mill roll wear is shown in Figure 5a. A comparison of the predicted by the gradient boosting method and the actual values of the rolling mill roll wear is shown in Figure 5b.

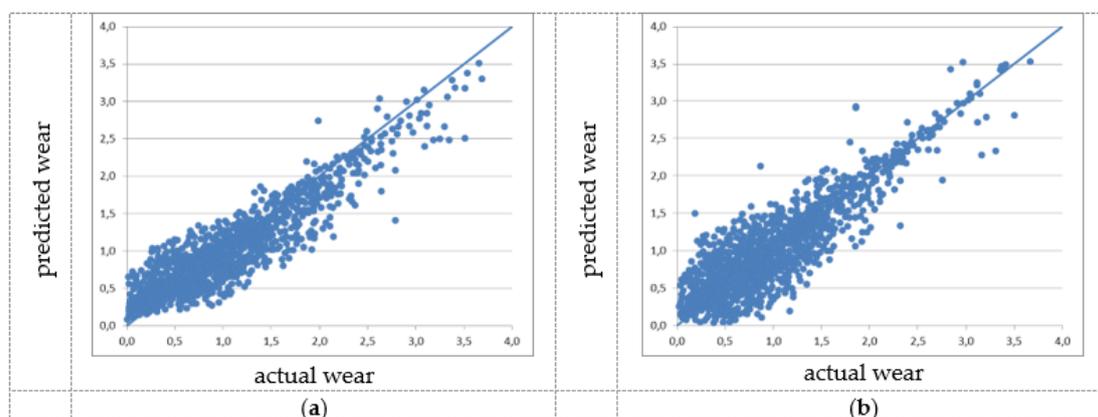


Figure 5. Comparison of predicted and actual rolling mill roll wear: (a) by the random forest model; (b) by the gradient boosting model.

The results of assessing the adequacy of the random forest model and gradient boosting model are far superior to previous models (Table 2).

Compared with linear, multivariate, and regularized models, the root mean square error (MSE) has decreased by about five times, and the coefficients of determination and correlation approximated to unity. That is to say that the random forest model can be recognized as adequate to the object of research and can be used to predict the degree of wear of the rolls of a rolling mill in the steel industry.

In terms of the coefficient of determination R^2 , gradient boosting is a more accurate model compared with the random forest model (the coefficient of determination is closer to unity). The root mean square errors of both models are equal, but, according to Figure 5, it can be seen that, when using the gradient boosting method, there is a greater number of coincidences of predicted and actual wear than when using the random forest method.

Analysis of the graphs (Figure 5) for symmetry regarding the straight line $Y_{\text{predicted}} = Y_{\text{actual}}$ shows that, in the case of developing the random forest model, there is a tendency to underestimate the predicted values. It is apparent that most of the values are located below the straight line (Figure 5a). In the case of developing the gradient boosting model, the predicted values are located symmetrically in reference to the straight line $Y_{\text{predicted}} = Y_{\text{actual}}$ (Figure 5b). Therefore, the gradient boosting model is preferred.

If necessary, carrying out additional optimization of the model, it is possible to achieve an even greater decrease in the forecast error [31]. Thus, the gradient boosting forecast model is preferable.

6. Conclusions

Based on the above study, the following conclusions can be drawn.

1. The hypothesis of using a large volume of production data (Big Data) to find statistically significant dependencies turned out to be completely consistent [32]. Operational control data are an inexhaustible source of information. Extracting useful information from Big Data is an important production task [33].
2. To improve the accuracy of the models, it is necessary to prepare statistical material in advance (remove outliers, «odd», and random measurement results; filter the data; identify different modes of operation; and consider them separately) and select the appropriate type of mathematical dependence. The quality of the developed models directly depends on the quality of training material preparation [34].
3. The analysis of the correlation dependences of the data showed that the most significant factors affecting the wear of the rolls are the dimensions and brands of rolled steel sheets. In addition, not least important is the material from which the rolls are made.
4. The construction of linear and multivariate regression models based on a large data set does not provide a qualitative result, as it does not allow taking into account complex and multi-connected dependencies between the input variables. Compositional models that are resistant to overfitting, noise, and outliers perform best. However, with a smaller amount of data that can be described by a simple model, it makes more sense to use multivariate regression.
5. Thus, a predictive model of rolling mill roll wear will allow rational use of rolls in terms of minimizing overall roll wear. The proposed model will make it possible to redistribute the existing work rolls between the stands in order to reduce the total wear of the rolls.

Author Contributions: Conceptualization, N.V.; methodology, N.V.; software, E.F.; validation, N.V. and E.F. data curation, A.K.; visualization, A.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

1. Abrosimov, A.A.; Shelyago, E.V.; Yazynina, I.V. Justification of Representative Data Volume of Porosity and Permeability Properties for Obtaining Statistically Reliable Petrophysical Connections. *J. Min. Inst.* **2018**, *233*, 487–491. [[CrossRef](#)]
2. Sharikov, Y.V.; Snegirev, N.V.; Tkachev, I.V. development of a control system based on predictive mathematical model of the C5–C6 isomerization process. *J. Chem. Technol. Metall.* **2020**, *55*, 335–344.
3. Kadyrov, E.D.; Koteleva, N.I. Introducing neural-network algorithms into an automated system designed to control metallurgical processes. *Metallurgist* **2011**, *54*, 799–802. [[CrossRef](#)]
4. Demidovich, V.B.; Chmilenko, F.V.; Rastvorova, I.I. Utilization of induction heating in the line of continuous casting-continuous rolling of steel. *Acta Technica CSAV* **2015**, *60*, 107–118.
5. Zhukovskiy, Y.L.; Korolev, N.A.; Babanova, I.S.; Boikov, A.V. The prediction of the residual life of electromechanical equipment based on the artificial neural network. *IOP Conf. Ser. Earth Environ. Sci.* **2017**, *87*, 032056. [[CrossRef](#)]
6. Oprea, G.; Andrei, H. Power quality analysis of industrial company based on data acquisition system, numerical algorithms and compensation results. In Proceedings of the 2016 International Symposium on Fundamentals of Electrical Engineering, ISFEE, Bucharest, Romania, 30 June–2 July 2016; p. 7803232.
7. Galkin, V.; Koltyrin, A. Investigation of probabilistic models for forecasting the efficiency of proppant hydraulic fracturing technology. *J. Min. Inst.* **2021**, *246*, 650–659. [[CrossRef](#)]
8. Vasilyeva, N.V.; Koteleva, N.I.; Fedorova, E.R. Real-time control data wrangling for development of mathematical control models of technological processes. *J. Phys. Conf. Ser.* **2018**, *1015*, 032067. [[CrossRef](#)]
9. Bazhin, V.Y.; Kulchitskiy, A.A.; Kadrov, D.N. Complex control of the state of steel pins in Soderberg electrolytic cells by using computer vision systems. *Tsvetnye Met.* **2018**, *27–32*. [[CrossRef](#)]
10. Utekhin, G. Use of statistical techniques in quality management systems. In Proceedings of the 8 International Conference Reliability and Statistics in Transportation and Communication–2008, Riga, Latvia, 17–20 October 2008; pp. 329–334.
11. Boikov, A.V.; Savelev, R.V.; Payor, V.A.; Erokhina, O.O. The control method concept of bulk material behaviour in the pelletizing drum for improving the results of DEM-modeling. *CIS Iron Steel Rev.* **2019**, *17*, 10–13. [[CrossRef](#)]
12. Leonidovich, Z.Y.; Urievich, V.B. The development and use of diagnostic systems and estimation of residual life in industrial electrical equipment. *Int. J. Appl. Eng. Res.* **2015**, *10*, 41150–41155.
13. Milyuts, V.G.; Tsukanov, V.V.; Pryakhin, E.I.; Nikitina, L.B. Saint Petersburg Mining University Development of Manufacturing Technology for High-Strength Hull Steel Reducing Production Cycle and Providing High-Quality Sheets. *J. Min. Inst.* **2019**, *239*, 536–543. [[CrossRef](#)]
14. Thombansen, U.; Purrio, M.; Buchholz, G.; Hermanns, T.; Molitor, T.; Willms, K.; Schulz, W.; Reisgen, U. Determination of process variables in melt-based manufacturing processes. *Int. J. Comput. Integr. Manuf.* **2016**, *29*, 1147–1158. [[CrossRef](#)]
15. Servin, R.; Arreola, S.A.; Calderón, I.; Perez, A.; Miguel, S.M.S. Effect of Crown Shape of Rolls on the Distribution of Stress and Elastic Deformation for Rolling Processes. *Metals* **2019**, *9*, 1222. [[CrossRef](#)]
16. Li, H.-J.; Xu, J.-Z.; Wang, G.-D.; Shi, L.-J.; Xiao, Y. Development of strip flatness and crown control model for hot strip mills. *J. Iron Steel Res. Int.* **2010**, *17*, 21–27. [[CrossRef](#)]
17. Zhao, N.; Cao, J.; Zhang, J.; Su, Y.; Yan, T.; Rao, K. Work roll thermal contour prediction model of nonoriented electrical steel sheets in hot strip mills. *J. Univ. Sci. Technol. Beijing Miner. Met. Mater.* **2008**, *15*, 352–356. [[CrossRef](#)]
18. Turk, R.; Fajfar, P.; Robic, R.; Perus, I. Prediction of hot strip mill roll wear. *Metalurgija* **2002**, *41*, 47–51.
19. Taimasov, B.T.; Sarsenbayev, B.K.; Khudyakova, T.M.; Kolesnikov, A.S.; Zhanikulov, N.N. Development and testing of low-energy intensive technology of receiving sulfate-resistant and road Portland cement. *Eurasian Chem. Technol. J.* **2017**, *19*, 347–355. [[CrossRef](#)]
20. Abdulaev, E.K.; Makharatkin, P.N.; Kuzhelev, A.I.; Grudinin, N.N. Assessment of technical condition of gearbox-motor-wheels and tires according to heating wear criterion when transporting building materials. *IOP Conf. Series Mater. Sci. Eng.* **2020**, *775*, 012001. [[CrossRef](#)]
21. Bolobov, V.; Chupin, S.; Binh, L.T. On the Wear Intensity Ratio of a Striker under Dynamic and Static Conditions. *IOP Conf. Ser. Earth Environ. Sci.* **2020**, *459*, 062085. [[CrossRef](#)]
22. Krasnyy, V.; Maksarov, V.V.; Maksimov, D. Improving the Wear Resistance of Piston Rings of Internal Combustion Engines when Using Ion-Plasma Coatings. *Key Eng. Mater.* **2020**, *854*, 133–139. [[CrossRef](#)]
23. Ratra, R.; Gulia, P. Big Data Tools and Techniques: A Roadmap for Predictive Analytics. *Int. J. Eng. Adv. Technol.* **2019**, *9*, 4986–4992.
24. Thillaieswari, B. Comparative Study on Tools and Techniques of Big Data Analysis. *Int. J. Adv. Netw. Appl.* **2017**, *8*, 61–66.
25. George, G.; Lavie, D. Big data and data science methods for management research. *Acad. Manag. J.* **2016**, *59*, 1493–1507. [[CrossRef](#)]
26. Maratea, A.; Petrosino, A.; Manzo, M. Extended Graph Backbone for Motif Analysis. In Proceedings of the 18th International Conference on Hybrid Systems: Computation and Control, Seattle, WA, USA, 14–16 April 2015; pp. 36–43.

27. Nguyen, T.L. A Framework for Five Big V's of Big Data and Organizational Culture in Firms. In Proceedings of the 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 10–13 December 2018; pp. 5411–5413.
28. Kaur, N.; Singh, G. A Review Paper On Data Mining And Big Data. *Int. J. Adv. Res. Comput. Sci.* **2017**, *8*, 407–409.
29. Rao, J.N.; Ramesh, M. A Review on Data Mining & Big Data, Machine Learning Techniques. *Int. J. Recent Technol. Eng.* **2019**, *7*, 914–916.
30. Kaisler, S.; Armour, F.; Espinosa, J.A.; Money, W. Big data: Issues and challenges moving forward. System sciences (HICSS). In Proceedings of the 2013 46th Hawaii International Conference on System Sciences, Maui, HI, USA, 7–10 January 2013; pp. 995–1004.
31. Dean, J.; Ghemawat, S. Mapreduce: Simplified data processing on large clusters. *Commun. ACM* **2008**, *51*, 107–113. [[CrossRef](#)]
32. Katal, A.; Wazid, M.; Goudar, R.H. Big data: Issues, challenges, tools and Good practices. In Proceedings of the 2013 Sixth International Conference on Contemporary Computing (IC3), Noida, India, 8–10 August 2013; pp. 404–409.
33. Wu, X.; Zhu, X.; Wu, G.-Q.; Ding, W. Data mining with big data. *IEEE Trans. Knowl. Data Eng.* **2014**, *26*, 97–107. [[CrossRef](#)]
34. Lindell, Y.; Pinkas, B. Privacy Preserving Data Mining. *J. Cryptol.* **2002**, *15*, 177–206. [[CrossRef](#)]