

# Machine Learning in predicting the extent of gas and rock outburst

Maciej Bodlak<sup>1,\*</sup>, Jan Kudelko<sup>1</sup> and Andrzej Zibrow<sup>2</sup>

<sup>1</sup>Wrocław University of Science and Technology, Faculty of Geoengineering, Mining and Geology, 27 Wyb. Wyspiańskiego St., 50-370 Wrocław, Poland

<sup>2</sup>Coal Holding sp. z o.o., 4/8 Kopernik St., 40-064 Katowice, Poland

**Abstract.** In order to develop a method for forecasting the costs generated by rock and gas outbursts for hard coal deposit "Nowa Ruda Pole Piast Waclaw-Lech", the analyses presented in this paper focused on key factors influencing the discussed phenomenon. Part of this research consisted in developing a prediction model of the extent of rock and gas outbursts with regard to the most probable mass of rock [Mg] and volume of gas [m<sup>3</sup>] released in an outburst and to the length of collapsed and/or damaged workings [running meters, rm]. For this purpose, a machine learning method was used, i.e. a "random forests method" with the "XGBoost" machine learning algorithm. After performing the machine learning process with the cross-validation technique, with five iterations, the lowest possible values of the mean-square prediction error "RMSE" were achieved. The obtained model and the program written in the programming language "R" was verified on the basis of the "RMSE" values, prediction matching graphs, out of sample analysis, importance ranking of input parameters and the sensitivity of the model during the forecast for hypothetical conditions.

## 1 Introduction

Current methods for predicting the probability of rock and gas outbursts allow relatively safe exploitation in the analyzed conditions [1]. Owing to adequate regulations and preventive methods, no death due to rock and gas outburst was recorded in Australian mines since 1994 [2, 3]. As current prediction methods offer high accuracy, their further optimization seems unnecessary. However, until present no method allows predicting the most probable extent of the outburst in given conditions. History teaches that the hazard of outbursts remains the key consideration for economically viable mining exploitation. In 1930, a rock and gas outburst in the "Waclaw" mine (historically known as German "Wenceslaus" mine) significantly affected the company's financial base, already strained by the beginning economic crisis. As a result, the mine became bankrupt and was closed [4-6]. Therefore, it seems sensible to investigate the costs generated by the results of potential rock and gas outbursts. Firstly, however, a prediction model of outburst extent must be developed. The constructed model will allow more informed decisions on exploitation in hazardous conditions to be made already during the planning stage. Notably, until present the risk premium is calculated arbitrarily, on the basis of historical data or expert knowledge [7]. The developed method is meant to allow including additional costs resulting from the need to repair fixed assets of a mining

company and compensate for potential losses caused by rock and gas outbursts. The research results may prove vital for estimating the profitability of an exploitation project in the conditions characterized by high rock and gas outburst hazard.

This study presents the results from an attempt to use methods in the field of machine learning to predict the extent of gas and rock outbursts in the hard coal deposit "Nowa Ruda Pole Piast Waclaw-Lech". Tools of this type in forecasting the probability of occurrence of this phenomenon, in given geological and mining conditions they have already been used by many researchers. In particular, the work of Chinese scientists should be mentioned here. Their use of the described tools was possible thanks to the extensive database of the phenomenon characterized for machine learning algorithms. At the beginning of the 21st century, neural networks were used for this purpose with the "BP" back propagation algorithm [8, 9]. In 2013 [10] to develop a risk prediction method, a training set for neural networks was embedded in a database of gas and rock outbursts at the "Jiaozuo" mine. Data with high reliability was chosen, so that the number of variables was strongly reduced. In the training set they selected, among others geological structure and changes in the slope of coal seams, carbon structure, gas pressure.

As a result, it was possible to obtain a model for forecast of gas and rock outbursts, however, the results had binary character, i.e. conditions for ejection were/or

\* Corresponding author: [maciej.bodlak@pwr.edu.pl](mailto:maciej.bodlak@pwr.edu.pl)

not were. The added value of the model, as opposed to the Australian systems, is including many variables in the forecast. In addition, the extensive training set and reliability of predictors has contributed to the Chinese researchers' close to 100% prediction accuracy.

The author at work [11] has attempted to estimate the weight of predictive factors affecting the risk of gas and rock ejection. To this end, he consulted some Polish experts dealing with risk assessment and prevention of remissions, in relation to their knowledge and experience. On the basis of this knowledge, he developed an expert system based on fuzzy logic. The system allows automatic evaluation of the threat of reapproach.

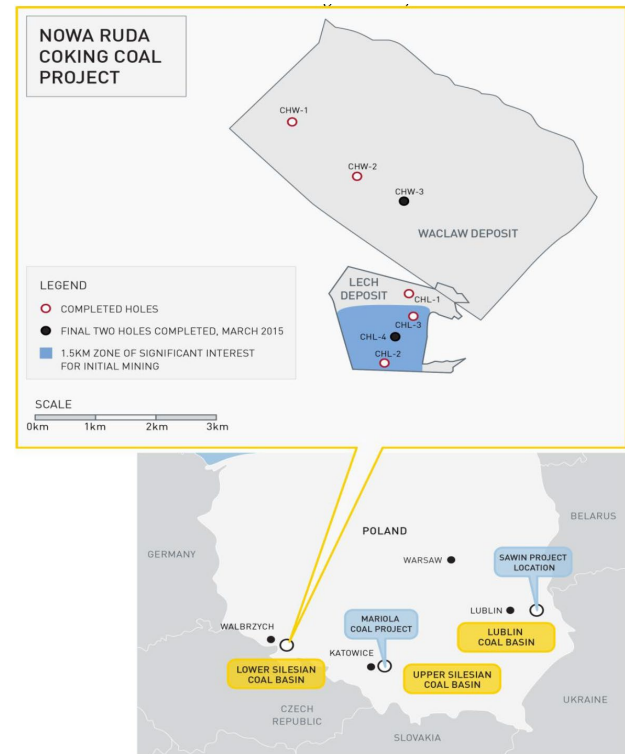
## 2 Selection of the test location

The selection of the test location was dictated by the availability of geological information and the degree of rock and gas outburst hazard.

Lower Silesian Coal Basin is a region associated with the hazard of rock and gas outbursts. In the mines of the Upper Silesian Coal Basin, this phenomenon has not occurred so far or has occurred incidentally. Therefore, research focused on the mines of the Lower Silesian Coal Basin. A total of 1733 rock and gas outbursts and 509 fatalities have been so far recorded in the Lower Silesian mines. During the restructuring of the mining industry in the 1990s, the region had four hard coal mines: "Julia," "Victoria," "Wałbrzych" and "Nowa Ruda." Of the above, the greatest outburst hazard was observed in the "Nowa Ruda" mine, and more specifically in the "Piaś" exploitation field, which is the region of the closed "Piaś" mine (German: "Ruben"). This field alone accounted for 1294 outbursts, i.e. 75 % of the total number of outbursts in the Lower Silesian mines. The outbursts claimed a total of 242 fatalities, i.e. 48 % of all fatalities recorded in the Lower Silesian mines. Unfortunately, a large part of the measurement data from the exploitation period is either lost or hardly accessible and therefore collecting all historical and survey information is problematic. However, in 2013 the Coal Holding company started to update the survey and prospecting information for the "Waław-Lech" hard coal deposit in the region of the "Piaś" field being part of the former "Nowa Ruda" mine. The prospecting works and tests of physical and mechanical parameters of rock and coal samples from the deposit allowed collecting the necessary geological information. The parameters of the coal seams used in the construction of the training dataset are mean values obtained from the results of physical and mechanical tests on rock samples from 7 exploration drill holes performed by the Coal Holding company as part of the deposit prospecting works – Fig. 1.

As the information on the parameters influencing the rock and gas outburst hazard in the "Waław-Lech" region was already available, a decision was made to collect a set of historical data on previous outbursts in the "Lech" field of the former "Nowa Rudna" mine.

This information was obtained from literature studies, inter alia from the 1971 work by Cis [6, 12] and from documents stored in the Kamieniec Ząbkowicki [13] branch of the National Archive in Wrocław, and in the archive of the Hard Coal Museum in Nowa Ruda (coll. of the Muzeum of Coal Mining in Nowa Ruda).



**Fig. 1.** Location of the deposit in the "Piaś – Nowa Ruda" field, region "Waław-Lech", and the locations of the exploration drill holes (materials provided by Coal Holding).

The resulting database comprised geological information and the description of gas conditions for six prospected coal seams in the "Lech" field. The preliminary research returned the description of 1288 out of 1294 rock and gas outbursts in the "Lech" field. Expressed in percent, this number accounts for 99.5% of the outbursts in the analyzed field and for 74% of all outbursts in Lower Silesian mines. It also represents the majority, i.e. 63% of all outbursts in Polish mines.

The "Nowa Ruda" mine has a rich history related to outburst hazards. The first outburst in the "Piaś" field was recorded in 1908. Since that time, the miners gained experience and learned how to work in changing geological and mining conditions. Mining was divided into stages with observance to strict exploitation rules and methods developed on the basis of research and observation. The history of the mine also shows episodes of coyoting without regard to health and safety regulations. The outburst of 1941, with 187 fatalities, may serve as an example here [14, 15]. This catastrophe could have been avoided or at least its results could have been mitigated if blasting works, which entailed a great outburst hazard, had not been performed when two shifts overlapped.

### 3 Research methodology

The choice of the random forests method was dictated by the possibilities of particular tools and a subjective evaluation of the results obtained. At the initial stage of the research, an attempt was made to use neural networks using the tensorflow library in the Python programming language. The modeling trials consisted in the use of a three-layer neural network with full connections, with Xavier initialization of the scales. However, the gaps and insufficient number of samples in the database caused the learning process to be distorted and the results obtained for all measurements were close to the average. This is typical behavior of this algorithm in the case of a poor database, because the algorithm predicts with the smallest possible error. Therefore, his tendency to predict values close to the average for a given set appears. In this way, incorrect prediction is at the same time fraught with the slightest mistake.

Then, as part of subsequent research trials, the logistic regression method was used. The application of this algorithm required entering any value in places of missing data. In this issue, the value "0" was chosen. The results obtained using this method were satisfactory, however, it was considered that other machine learning methods may enable more accurate predictions of the extent of gas and rock outbursts. The Mean Root Error (RMSE) was used as the prediction effectiveness tool.

RMSE error values for individual logistic regression models are shown in Tables 1-3.

**Table 1.** RMSE for logistic regression, prediction of the amount of mass of rock.

with data gaps	
training set	<b>6394.22</b>
validation set	-
with filling gaps with zero value	
training set	<b>398.68</b>
validation set	<b>492.33</b>

**Table 2.** RMSE for logistic regression, prediction of the of volume of gas.

with data gaps	
training set	<b>1 526 972.00</b>
validation set	-
with filling gaps with zero value	
training set	<b>66 796.99</b>
validation set	<b>158 048.80</b>

**Table 3.** RMSE for logistic regression, prediction of the released in an outburst and to the length of collapsed and/or damaged workings.

with data gaps	
training set	<b>507.14</b>
validation set	-
with filling gaps with zero value	
training set	<b>81.12</b>
validation set	<b>195.68</b>

Based on the knowledge of rock and gas outbursts in the Nowa Ruda coal basin and on machine learning techniques, a decision was made to use the random forests method. The choice of the method was aimed at using the latest, generally available and currently the most effective methods and algorithms, so attempts to use simpler methods were omitted.

The choice of the method based on popular random trees was dictated by previous experiences which prove that gaps in the database have a negative influence on the predictions of rock and gas outbursts obtained with neural networks and logistic regression. In the current method, modeling consists in "learning" more than one decision tree and building a prediction model based on the results from all of the trees.

In this research, the prediction model was developed with the XGBoost open source implementation in "R" and with xgboost library, as well as with the cross-validation and feature importance ranking methods based on caret library (also in "R"). "R" is an interpreted programming language and an environment (RStudio®) for statistical calculations and results visualization. "R" is the basic programming language in bioinformatics. So far "R" has been used with particular success in analyses related to clinical trials.

Importantly, the XGBoost model, as most other machine learning algorithms, is a black box method [16]. This means that the method will not produce obvious outburst extent predictions for individual new measurements. White box methods, e.g. logistic regression, stand in opposition to black box methods. After training such a model, a claim may be safely made that for instance the greater depth, the greater the chance that the outburst will produce more gases. In XGBoost, such claims are impossible, which is greatest disadvantage of this model. Therefore, XGBoost models are used only if white box methods prove too weak to allow predictions.

When predictions are made with the XGBoost model, the measurement with given features is analyzed in two decision trees, and the results are eventually added. This is the Gain value, and it is stored in leaves. The greatest advantage of this method over other algorithms is the possibility to use smaller databases with numerous gaps of missing data. However, its disadvantage consists in the inclination of the model to excessively adjust to the training data. This is known as overfitting. In order to avoid overfitting, the model was supplemented with regularization in the form of bootstrap cross-validation.

Teaching a particular model consisted in determining the best parameters of the decision tree in terms of the investigated rock and gas outbursts. The parameter in the decision tree is the cut-off value for each node. The decision tree included questions, e.g. whether the CO<sub>2</sub> content in the outburst gases "%CO<sub>2</sub>" is lower than 17.475%. In the next step, further questions are asked in accordance with the structure of the decision tree.

Full "machine learning" in the random forests method with XGBoost implementation consists in "teaching" the models for a given network of hyperparameters. Hyperparameter is a value which will

influence the choice of parameters in the decision tree, based on the gradient algorithm.

XGBoost models have six hyperparameters, all related to the structure of random forests. The values influence such properties as the learning step in the gradient boosting algorithm, the maximum number of levels in the tree etc. The hyperparameters used in the XGBoost model are listed and described below:

- Nrounds – number of trees
- Max\_depth – maximum tree depth
- Eta – learning rate, or the length of step in the gradient boosting algorithm
- Gamma – minimum loss reduction which allows creating the next node
- Colsample\_bytree – ratio of columns randomly selected for each tree
- Min\_child\_weight – parameter deciding on the feature cut-off and the number of leaves in the tree

In the first stage of works on the model, the data set was divided into one training and one validation subset, in the 90% to 10% ratio, respectively. The training set was used to determine the parameters of the tree, i.e. in the “model training” process, while the validation set contained samples which were not analyzed by the model during the training process. Thus, by comparing the prediction results (based on the RMSE – Root Mean Squared Error) from both sets, it is possible to determine whether the model sufficiently generalized the phenomenon (in this case – rock and gas outburst) and whether it avoided overfitting to the measurements found in the training set.

Predictions of the rock masses released in outbursts were performed by analyzing a set of 952 measurements (outbursts), of which 875 samples were in the training set and 95 samples were in the validation set. Predictions of the gas volumes released in outbursts were performed by analyzing a set of 245 measurements (outbursts), of which 220 samples were in the training set and 25 samples were in the validation set. Predictions of the length of workings collapsed and/or damaged due to a single outburst were performed by analyzing a set of 228 measurements (outbursts), of which 205 samples were in the training set and 23 samples were in the validation set.

In the next step, the training process was performed with the use of 5 bootstraps and in accordance with the cross-validation technique. The procedure was to perform five random divisions of the training data set into two subsets: the proper training set and the validation training set. Each of the five iterations included “training” of multiple models on the proper training sample, with the use of selected sets of hyperparameters. Simultaneously, the strength of the generated models was measured during each iteration by analyzing the results against the validation training sample. The model strength was determined by measuring the RMSE value.

Tables 4 to 6 present the RMSE error values for selected sets of hyperparameters.

**Table 4.** RMSE for random forests, prediction of the amount of mass of rock.

training set	<b>276.50</b>
validation set	<b>372.07</b>

**Table 5.** RMSE for random forests, prediction of the of volume of gas.

training set	<b>37 389.82</b>
validation set	<b>149 549.20</b>

**Table 6.** RMSE for random forests, prediction of the released in an outburst and to the length of collapsed and/or damaged workings.

training set	<b>38.37</b>
validation set	<b>188.58</b>

By measuring the RMSE, it was possible to estimate which of the hyperparameter sets allows best modeling of the proper validation group. Tables 7-9 present sets of hyperparameters for selected models.

**Table 7.** Hyperparameter sets, prediction of the amount of mass of rock.

Hyperparameter	value
Nrounds	<b>300</b>
Max_depth	<b>4</b>
Eta: - learning rate	<b>0.005</b>
Gamma	<b>7</b>
Colsample_bytree	<b>1</b>
Min_child_weight	<b>3</b>

**Table 8.** Hyperparameter sets, prediction of the of volume of gas.

Hyperparameter	value
Nrounds	<b>300</b>
Max_depth	<b>6</b>
Eta: - learning rate	<b>0.005</b>
Gamma	<b>8</b>
Colsample_bytree	<b>0.5</b>
Min_child_weight	<b>3</b>

**Table 9.** Hyperparameter sets, prediction of the released in an outburst and to the length of collapsed and/or damaged workings.

Hyperparameter	value
Nrounds	<b>500</b>
Max_depth	<b>5</b>
Eta: - learning rate	<b>0.005</b>
Gamma	<b>3</b>
Colsample_bytree	<b>0.5</b>
Min_child_weight	<b>3</b>

## 4 Model verification

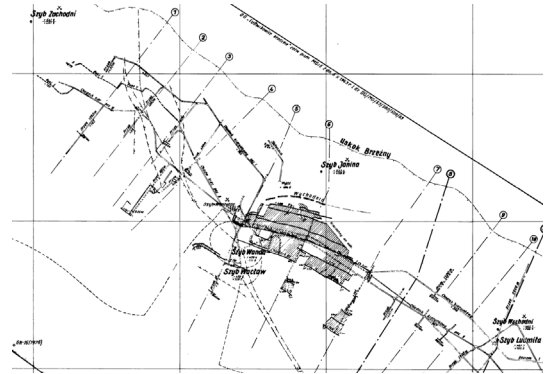
The prediction model based on the random forests method and on the XGBoost algorithm was verified in relation to: RMSE values; predictions for the “Józef” coal seam (301) (Fig. 2) in the “Wacław” mine (out of sample) (Table 10); evaluation of the plot of the predicted versus the observed values in the “Piaśt” field of the “Nowa Ruda” mine (Figs. 3 and 4); analysis of the importance ranking for the features which influence rock and gas outbursts (Table 11); and prediction test for hypothetical geological and mining conditions (Figs. 5, 6 and 7).

As the seam (the test seam for the purpose of this research) was mined, no rock or gas outbursts were observed. However, due to the properties of the seam, it was qualified in the highest, fourth class of rock and gas outburst hazard. The “Józef” (301) and the “Wacław” (304) seams are a link of the Gorce Mountains in the lithological profile of the upper Żaclér strata. It is part of the upper conglomerate-sandstone layers. It also includes layers of dark clay shale, clay-sand shale and mudstone. Importantly, the neighboring “Wacław” seam (304) was the location of the greatest number of rock and gas outbursts in the “Wacław” mine, including the most tragic event of July 9, 1930.

The seam (considered as a test seam for the purpose of this research) was mined directly at the line of bottom pillars protecting the “Wanda” and “Wacław” shafts, on their eastern side. Extraction works were mainly performed on small depths, not exceeding 240 m b.g.l. On the west side, the exploitation fields were protected by shaft pillars and on the north – by outcrops, and therefore the exploitation front advanced along the deposit length towards east and the “Ludmiła” shaft. Coal was extracted with electric longwall cutting machines. The mined cavern was hand filled mostly with rock obtained from interlayers and overbreaks. The workings map of the “Józef” seam (301) is shown in Fig. 3. The seam had the following features used in the prediction of the extent of rock and gas outburst possible in the given geological and mining conditions:

- Local mean gas bearing capacity of coal during outburst [ $\text{m}^3/\text{Mg}_{\text{CSW}}$ ] = 0
- Depth, [m b.g.l.] =  $40 \div 240$
- Mean  $\text{CO}_2$  content in outburst gases, [%] = 0
- Mean increase overpressure in borehole, before outburst [MPa] = 0
- $\text{CH}_4$  content in outburst gases, [%] = 0
- Coal density, [ $\text{kg}/\text{m}^3$ ] = 1577.67
- Coal strength to uniaxial compression, [MPa] = 4.3
- strain energy storage index of coal, [-] = 0.78
- RQD of rocks in the roof, [-] = 73.3
- RQD of rocks in the floor, [-] = 36.6
- Seam thickness, [m] = 1.3
- Mean gas bearing capacity for the seam, [ $\text{m}^3/\text{Mg}_{\text{CSW}}$ ] = 0.02
- Mean methane bearing capacity for the seam, [ $\text{m}^3/\text{Mg}_{\text{CSW}}$ ] = 0.106

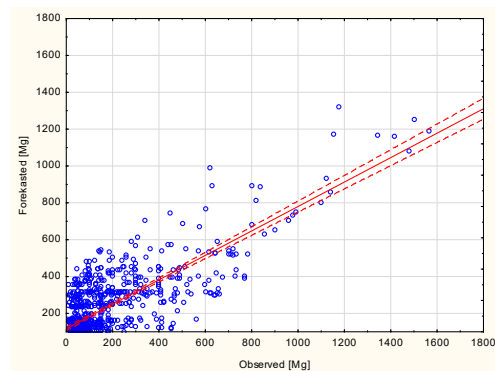
- Surface of mined out working, [thousands of  $\text{m}^2$ ] = 52
- Daily mine output, [ $\text{Mg}/\text{day}$ ] = 2897



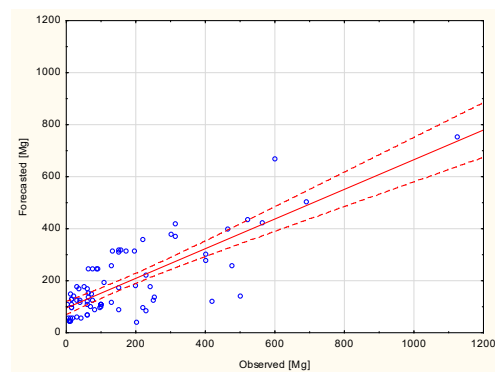
**Fig. 2.** Workings map of the “Józef” coal seam (301) in the “Wacław” mine, survey records of the “Nowa Ruda” mine (Kamieniec Żąbkowicki branch of the National Archive in Wrocław).

**Table 10.** Prediction of the potential extent of rock and gas outbursts, out of sample verification.

Predicted mass of rocks released in an outburst	
Depth of 240 m b.g.l.	<b>55.98 Mg</b>
Predicted volume of gases released in an outburst	
Depth of 240 m b.g.l.	<b>11 734.11 <math>\text{m}^3</math></b>



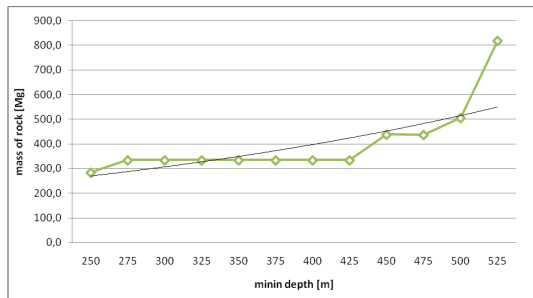
**Fig. 3.** The predicted versus the observed values of material released in an outburst in the “Piaśt” field of the former “Nowa Ruda” mine for the training set.



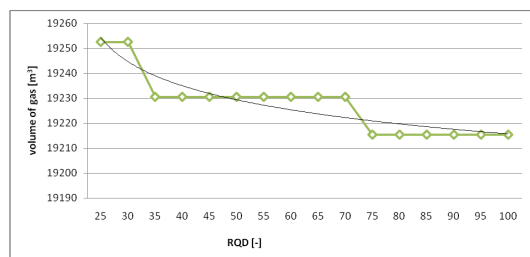
**Fig. 4.** The predicted versus the observed values of material released in an outburst in the “Piaśt” field of the former “Nowa Ruda” mine for the validation set.

**Table 11.** Feature importance ranking in the prediction model of the amount of material released in an outburst.

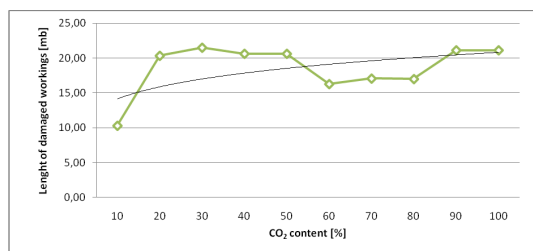
No.	Physical, mechanical and gaseous features of the seam	Importance [%]
1.	RQD of rocks in the floor	7.6177
2.	CO <sub>2</sub> / CH <sub>4</sub> [-] coefficient	5.2909
3.	RQD of rocks in the roof	2.6824
4.	Mean gas bearing capacity for the seam [m <sup>3</sup> /M <sub>g<sub>CSW</sub></sub> ]	1.8767
5.	Coal density [kg/m <sup>3</sup> ]	1.2916
6.	Mean methane bearing capacity for the seam	0.5480
No.	Anthropogenic features of the seam	Importance [%]
1.	Daily mine output [Mg/day]	12.8317
2.	Surface of mined-out workings [m <sup>2</sup> ]	6.6235
No.	Features of individual rock and gas outbursts	Importance [%]
1.	Depth [m b.g.l.]	100.0000
2.	Local gas bearing capacity of coal during outburst [m <sup>3</sup> /M <sub>g<sub>CSW</sub></sub> ]	10.6973
3.	CO <sub>2</sub> content in outburst gases [%]	9.2550



**Fig. 5.** Mass of rocks released in hypothetical geological and mining conditions, as a function of mining depth.



**Fig. 6.** Volume of gases released in hypothetical geological and mining conditions, as a function of mean RQD of rock mass surrounding the seam.



**Fig. 7.** Length of collapsed and/or damaged workings in hypothetical geological and mining conditions, as a function of mean CO<sub>2</sub> content in outburst gases.

Being approximately 56 Mg, the predicted mass of rocks released in an outburst is a small value in relation to the extent of releases recorded in the “Nowa Ruda” mine (max. 5 000 Mg, 274.2 Mg on average). Although no rock or gas outbursts were recorded during the exploitation of the “Józef” seam, their potential occurrence in the future cannot be ignored. Notably, the “Wacław-Lech” deposit also located in the “Józef” (301) test seam has the fourth – highest class of rock and gas outburst hazard. Rock and gas outbursts have a random character and occur in particular local gaseous, geological and dynamic conditions, most typically in locations where the seam is disturbed, either in a continuous or discontinuous manner.

Moreover, the predicted value is higher than the actual value of rock and gas release due to the outburst, which may be considered an advantage in that it allows for a safety margin. Similar procedures are used in predictions of seismic hazards in copper deposits in the KGHM Polska Miedź mines located in the Fore Sudetic monocline. They consist in increasing the energy classes of the predicted seismic events. Thus, adequate preventive actions may be taken during mining operations, while observing the priority of health and safety regulations over the economics of production.

The graphs above demonstrate that the constructed models allow predictions of the extent of rock and gas outbursts on the basis of historical data, as well as physical and mechanical parameters of the rocks forming the hard coal seam. However, there are significant deviations and this is an important fact that shows that further optimization of the model will require more measurements to build a larger set of geological parameters of the deposit. Although the XGBoost method used in this research is a black box method, it allows determining a feature importance ranking based on the learned model. This is possible with the varImp function from the caret package in “R.” The ranking is a combination of the number of references to a particular feature in relation to the number of all trees (specifically, what is the percentage of trees comprising the analyzed feature) and of the strength of the population sequence with the sole use of this feature. This may serve as a basis to select the most important features. Feature importance is normalized to 100% for the most important feature – the remaining values are scaled appropriately. Table 11 contain the results of feature importance analysis divided into physical, mechanical and gaseous features of the seam rocks, anthropogenic features and features of individual outbursts which occurred in the “Piast” field of the “Nowa Ruda” mine.

The analysis demonstrated that the outburst depth (and thus the mining depth) has the highest position in the feature importance ranking. Depending on the model, another distinctive feature located lower in the ranking is the mean daily mine output. Among mean physical and mechanical parameters of the rocks in the seam, the values of greatest importance for the extent of rock and gas outburst include rock strength to uniaxial compression and the Rock Quality Designation (RQD) for rocks in the roof and in the floor.

The ranking does not include one of the most important physical-mechanical parameters of coal – its impact strength. This fact, however, is probably caused by insufficient measurement diversification, as coal in the “Waclaw-Lech” deposit has very limited variability, in the range from 0.40 to 0.43. Therefore, this value does not have a sufficient potential to be used in the construction of the regression model. Similar is the case with mean gas bearing capacity of the deposit, which also shows limited variability between individual seams and as a result it was given low significance in the prediction model.

At the same time, the analysis demonstrated the importance of the mining depth, intensity and range, as well as of coal compressive strength and of the quality of rock mass surrounding the seam. Also, observations confirmed greater importance of CO<sub>2</sub> content over CH<sub>4</sub> content, which seems to support – if connected with RQD – a thesis proposed by experts in the field. In this theory, two types of CO<sub>2</sub> inflow are distinguished: transmission of CO<sub>2</sub> from porphyry veins into porous rocks such as sandstone and conglomerate, and subsequently its penetration through fault lines into coal seams; direct penetration of CO<sub>2</sub> into coal seams from inside the earth, through fault lines.

The above observation implies that the probability of an outburst in the Nowa Ruda coal basin increases together with mining depth and with decreasing distance to disturbed and dislocated zones, especially to locations where fault lines across the seam begin to disappear. Years of experience show that the physical and mechanical parameters of rocks are reduced in the vicinity of faults, and this fact, in connection with increased saturation of rocks with CO<sub>2</sub>, leads to an increase in the outburst probability and its extent.

In order to verify the proper functioning of the model, prediction was tested for sensitivity to changing input data. For this purpose, predictions were performed for a hypothetical seam with changing input features. Although the analysis covered all variables, space limits allow only some of the numerous graphs to be included in this paper.

The amount of material released in an outburst does not increase parabolically, exponentially or logarithmically, but takes a form similar to a moving average. This fact is caused by the operation of the employed algorithm, i.e. the random forests method which is based on individual, binary trees. The method uses recursive division of values into segments by minimizing outliers in each step. All divisions in the method are binary (each node in the tree may be split into no more than two branches).

Based on the results of the test and with regard to the above, the model was assumed to work properly and to offer logically coherent results which correspond to prior observations.

## 5 Conclusions

Research into a method for predicting the cost generated by rock and gas outbursts has produced a model for

predicting the extent of the phenomena based on the random forests method and on the XGBoost learning algorithm. The model was verified in relation to: RMSE values; predictions for the “Józef” coal seam (301) in the “Waclaw” mine (out of sample); evaluation of the plot of the predicted versus the observed values in the “Piaś” field of the “Nowa Ruda” mine; analysis of the importance ranking for the features which influence rock and gas outbursts; and prediction test for hypothetical geological and mining conditions.

The tests demonstrate that the model may be used for its primary purpose, i.e. to predict the costs induced by gaseous, geological and dynamic phenomena, already at the stage of designing a production plan in particular geological and mining conditions. However, further research is needed to improve the accuracy of the models based on machine learning. This improvement is possible by enlarging and enriching the training set. Developing a larger and more complete database of rock and gas outburst hazards will allow the XGBoost random forests model to be used in predictions of the extent of this phenomenon. On the other hand, increasing the quality of the input training set may in the future contribute to the use of methods from the white-box group. The works in this respect will be continued.

## References

1. C.R. Harvey, Int. Symp. Workshop on Management and Control of High Gas Emissions and Outbursts in Underground Coal Mines, Wollongong (1995)
2. D. Black, N. Aziz, M. Jurak, R. Florentin, The Coal Operators' Conference - University of Wollongong & the Australasian Institute of Mining and Metallurgy (2009)
3. N. Aziz, D. Black, T. Ren, First International Symposium on Mine Safety Science and Engineering (2011)
4. M. Bodlak and R. Borzęcki, *Hereditas Minariorum*, **1**, 151–167 (2014)
5. M. Bodlak., *Cuprum*, **1**, 39-56 (2016)
6. B. Bandurski, J. Bawecki, T. Kierzenkowski, L. Stalski, *Historia i kronika Kopalni Węgla Kamiennego „Nowa Ruda”*. *Zarys Dziejów* (KWK „Nowa Ruda”, 1989)
7. J. Kudęłko, *Strategie inwestycyjne przedsiębiorstw górniczych* (Wyd. KGHM Cuprum z o.o. - Centrum Badawczo-Rozwojowe, 2012)
8. F. Hua and S.L. Langshan, *Journal of Coal Science & Engineering*, **8**, 2, 80-83 (2002)
9. Z. Xusheng and W. Shanhai, *Coal Geology and exploration*, **30**, 2, 23–25 (2002)
10. W. Fuzhong and L. Weizhe, *Journal of Computers*, **8**, 1 (2013)
11. N. Skoczylas, *Arch. Min. Sci.*, **59**, 1, 41–52 (2014)
12. J. Cis, *Wyrzuty gazów i skał w dolnośląskim zagłębiu węgla kamiennego* (Wydawnictwo „Śląsk”, Katowice, 1971)

13. Documentation of gas and rock outbursts in the "Nowa Ruda" mine. No author, no year, KWK "Nowa Ruda", 150 technical drawings
14. M. Bodlak, R. Borzecki, A. Zibrow, *Hereditas Minariorum*, **3**, 53-81 (2016)
15. M. Bodlak, R. Borzecki, A. Zibrow, *Przegląd Górniczy*, **4**, 85-92 (2017)
16. T. Chen and C. Guestrin, KDD '16 Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785-794 (2016)