

ALINA ŻOGAŁA<sup>1</sup>, MACIEJ RZYCHON<sup>2</sup>

## Prediction of quality parameters of a dry air separation product using machine learning methods

### Introduction

The process of coal enrichment is one of the most important stages of the process of obtaining this raw material. The effectiveness of this step depends on the quantitative and qualitative parameters of the product obtained, as well as the characteristics of the resulting wastes (Michalak et al. 2012).

At present, separation processes in industrial conditions utilize mostly heavy liquid or water based gravity separation methods (Blaschke 1976, 2009; Baic and Blaschke 2013). However, the coal market is continuously changing. An increase in demand for coal as a fuel, despite the fact that its production has decreased, is causing a deficit that must be handled. The largest coal recipients, e.g. power plants, are able to ecologically use even lower-quality coal thanks to well-developed combustion technology. Therefore, it is not always necessarily justified to process coal by means of wet methods, that are known for their precision, but high costs and technological complexity as well. There is a reliable and less complex alternative to substitute them – dry air separation.

---

✉ Corresponding Author: Alina Żogała; e-mail: azogala@gig.eu

<sup>1</sup> Central Mining Institute, Poland; ORCID iD: 0000-0002-3711-6977; e-mail: azogala@gig.eu

<sup>2</sup> Central Mining Institute, Poland; ORCID iD: 0000-0003-2679-2563; e-mail: mrzychon@gig.eu



© 2019. The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution-ShareAlike International License (CC BY-SA 4.0, <http://creativecommons.org/licenses/by-sa/4.0/>), which permits use, distribution, and reproduction in any medium, provided that the Article is properly cited.

The dry air separation is the process based on the separation of the material, called the feed, under the influence of the pressure difference potential over the suction (vacuum) layer to separate certain types of grains on two or more products that differ from the starting material quantitatively or qualitatively in terms of at least one characteristic (Drzymała 2009). Research has been undertaken, aimed at incorporating the suction method into the process of dust coal cleaning. One such machine used in air-based coal separation is the pneumatic sorter, built as part of the KIC InnoEnergy project entitled *Novel dry sorter for coal processing and coal recovery from mine originating wastes* (acronym AMSEP).

A very important task in the coal enrichment process is the proper design of the technological line, adapting it to the characteristics of the delivered raw material and ensuring stable and expected parameters of the output products. This is an extremely complex issue (Bergh 2016). One of the ways to deal with this complexity is the use of mathematical modeling. Modeling is usually about describing working devices (or the entire technological line) using mathematical equations. The input values in the equations are usually the feed characteristic and the device operating parameters. The output values are the quantitative and qualitative characteristics of the product and waste obtained (Michalak et al. 2012). From the perspective of application of coal for energy purposes, the determination of parameters of the output as: ash content, moisture content, sulfur content, calorific value is essential (King 2001). The development of computational methods allowed for the use of machine learning methods in issues of mathematical modeling of technological systems and forecasting output of technological processes (Tumidajski 2010).

Machine learning is a field of artificial intelligence dealing with the study of algorithms and systems that improve their performance along with the experience gained. The experience is the information from the learning data used to train the system. Therefore the performance of machine learning algorithms depends heavily on the representation of the data they are given (Boser et al. 1992). Machine learning solves the real world problems by building a model that is good and useful approximation to the data (Raina 2016). The described process is reminiscent of the human cognitive process.

The use of machine learning methods and artificial intelligence to describe the processing of coal has been the subject of various papers. Ali et al. (Ali et al. 2018) used five different machine learning models (random forest, artificial neural networks, the adaptive neuro-fuzzy inference system, Mamdani fuzzy logic and a hybrid neural fuzzy inference system) to predict the flotation behavior of fine high ash coal. Jorjani et al. (Jorjani et al. 2009) predicted the combustible value and combustible recovery of coal flotation concentrate by regression and an artificial neural network based on proximate and group macerals analysis. Chelgani et al. (Chelgani et al. 2018) used support vector regression (SVR) modeling to predict the coal flotation recovery and flotation rate constant. Oruç et al. (Oruç et al. 2010) applied regression analysis to predict the ash content and the recovery of the clean coal after processing in the hydrocyclone and Falcon concentrator under different operating conditions.

The purpose of the work was to predict the selected product parameters of the dry separation process using a AMSEP pneumatic sorter. The prediction was carried out using chosen machine learning techniques. A brief description of these techniques is available below.

## 1. Machine learning techniques

### 1.1. Multiple Linear Regression (MLR)

Regression analysis is performed in order to determine the correlation between two or more variables having cause-effect relationships and on the basis of which the prediction of a given phenomenon is made (Uyanik and Guler 2013).

The multiple linear regression model is an extension of a simple linear regression model to incorporate two or more explanatory variable in a prediction equation for a response variable.

A multiple linear regression model (for the  $n$ -element sample) with  $p$  predictor variables  $x_1, x_2, \dots, x_p$  and a response  $y$ , can be written as:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi} + e_i \quad \text{for } i = 1, 2, \dots, n \quad (1)$$

- $\beta_0$  – the constant term,
- $\beta_1$  to  $\beta_p$  – the coefficients relating, to the  $p$  explanatory variables to the variables of interest,
- $e_i$  – the random error in prediction, that is variance that cannot accurately be predicted by the model.

In multiple linear regression it is assumed that  $y$  is directly related to a linear combination of the explanatory variables.

### 1.2. Support Vector Regression (SVR)

The support vector machine (SVM) introduced by Vapnik (Boser et al. 1992; Cortes and Vapnik 1995; Vapnik 1995) has widely been applied in many areas including pattern recognition, image classification, regression analysis, time series prediction and other fields (Jonsson et al. 2002; Wang et al. 2012; Zhou et al. 2014; Chelgani et al. 2018; Zhengwei 2007).

In this algorithm, each data item is plotted as a point in multidimensional space. Particular points have a different class membership. Classification is performed by finding the hyperplane that differentiates particular classes. There are many hyperplanes that might classify the data. The best hyperplane represents the largest separation – the distance from

that hyperplane to the nearest data point on each side is maximized. For non-linear SVM, the data is mapped to a higher dimensional space via a user-specified kernel, such as a polynomial kernel or a radial basis function.

The support vector machine (SVM) is used for the classification and support vector regression machine (SVR) can be used to make predictions. The training of the SVR model is based on minimization of the error function (Osowski 2013):

$$\frac{1}{2} w^T w + C \sum_{i=1}^n \xi_i + C \sum_{i=1}^n \xi'_i \quad (2)$$

subject to the constrains:

$$w^T \phi(x_i) + b - y_i \leq \varepsilon + \xi_i$$

$$y_i - w^T \phi(x_i) - b \leq \varepsilon + \xi'_i$$

$$\xi'_i > 0, \quad \xi_i \geq 0, \quad i = 1, \dots, N$$

- ↪  $C$  – capacity constant (chosen individually; the larger the  $C$ , the more the error is penalized),
- $\varepsilon$  – tolerance,
- $w$  – vector of weights,
- $b$  – constant,
- $\xi_i, \xi'_i$  – parameters for handling non-separable data (inputs),
- $n$  – number of training cases,
- $x_i$  – independent variables,
- $y_i$  – dependent variables,
- $\phi$  – kernel used to transform data from the input (independent) to the feature space.

### 1.3. Multilayer Perceptron neural network (MLP)

An artificial neural network (ANN) is an information-processing system that is based on generalizations of human cognition (Lippmann 1987). ANNs can identify and learn correlated patterns between input data sets and corresponding actual target values.

A neuron is a basic information processing unit which forms the basis for designing the artificial neural network. The neurons are logically arranged into two or more layers, and interact with each other via weighted connections (Liu et al. 2006). The weight represents the connection strength for a particular input for each connection. Each neuron is a summing element followed by a transfer function. The output of each neuron is fed as the input to all of the neurons in the next layer (Allahkarami 2017). The data is presented to the neural

network by the input layer, and an output layer holds the response of the network to the input. The hidden layers enable these networks to represent and compute complicated associations between patterns (Liu et al. 2006).

The most basic and commonly used neural network is the multilayer perceptron (MLP) network that consists of one input layer, one or more hidden layers and one output layer (Feng 2015). After establishing the network structure, weights and thresholds must be set so as to minimize the prediction error made by the network. This is the role of the training algorithms (Osowski 2013). It is not possible to analytically determine where the global minimum of the error surface is. The gradient descent used in an article is an iterative minimization method. The gradient of the error function always shows the direction of the steepest ascent of the error function. Thus, the algorithm can start with a random weight vector and subsequently follow the negative gradient.

#### 1.4. Clustering k-means

K-means is one of the simplest unsupervised learning algorithms that solve the well-known clustering problem. The procedure follows a simple way to classify a given data set through a certain number of clusters.

The first step of the procedure (Liu 2018) is a random selection of  $k$  data points in the raw data set, the initial values of these points are taken as the center of each cluster. Then, the distance between non-central data points and the center of each cluster is calculated, and the data points are allocated to the cluster nearest to them. After these points are clustered, the mean of each cluster is calculated and the center point of each cluster is selected once again. This process is repeated until the objective criterion function converges (Gerhard 2015).

An algorithm for clustering  $n$  input data points  $x_1, x_2, \dots, x_n$  into  $k$  disjoint subsets  $c_i$  ( $i = 1, \dots, k$ ) minimizes the following objective criterion function:

$$J = \sum_{i=1}^k \sum_{x_j \in c_i} \|x_j - c_i\|^2 \quad (3)$$

- ↪  $x_j$  – vector representing the  $j$ -th data point in cluster  $c_i$ ,
- $\|x_j - c_i\|^2$  – chosen distance measurement (norm) between data point  $x_j$  and the cluster centre  $c_i$ .

This criterion tries to make the resulting  $k$  clusters as compact and as separate as possible (Sahu 2009).

## 2. Materials and methods

### 2.1. Description of AMSEP sorter

The sorting method is based on separation of the grains of feed material with use of negative pressure, generated by a nozzle located above the layer of enriched material, thus setting particles in motion. The implementation of the perforated working surface in the form of a rotating cylinder allows overpressure to be generated under the layer of grains of enriched materials. Overpressure under the layer and curvature of the cylindrical working surface improves the efficiency of the sorting process through: initiating grain vibrations and their mutual displacement, as well as the drying of grains, therefore reducing the possibility of grain sticking. The experiments were conducted using the installation for dry air separation presented in Figure 1.

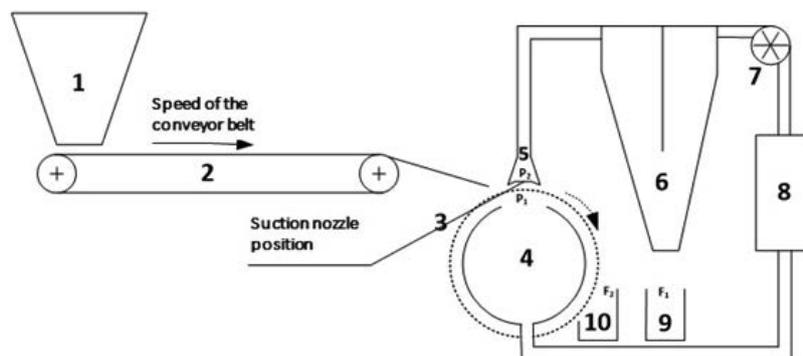


Fig. 1. Installation for dry air separation process

Rys. 1. Instalacja do procesu suchej separacji

The installation for carrying out the dry air separation process consists of the basket (1) as a tank to prepare the material to the sorting process. The material from the basket is loaded on the belt conveyor (2). The next operation is loading the material on a rotating cylinder made of a perforated surface (3). The directional nozzle (4) is located inside the rotating cylinder. The outlet of the directional nozzle is directed perpendicular to the surface of the rotating cylinder. The suction nozzle (5) is located above the rotating cylinder. The sucked material grains are transported to a particle separator (6), where the grains are separated from the air stream. An air flow is regulated by a suction fan (7). The air stream is purified from the dust that was not separated by the particle separator in the dust collector (8). The air flow after cleaning in dust collector is fed to the directional nozzle.

## 2.2. Methodology

The main objective of the presented paper was to predict chosen parameters of the upper fraction of the dry separation process: ash content, sulfur content, moisture content and calorific value, based on characteristic of feed and selected sorter settings. Output parameters were analyzed separately – independent groups of predictive models were created for each of them.

The construction and implementation of predictive models was based on the knowledge of the basic and most often determined parameters of coal, such as calorific value or sulfur content. The aim of the authors was to create a methodology that allows for the prediction of the parameters of a dry separation product, without the need for complex and costly laboratory tests of the feed material.

Before starting to develop models, the input variables, influencing particular product parameters, were selected by the analysis of the correlation coefficient. It was found that the individual parameters of the upper fraction depend on the corresponding value of this parameter in the feed and frequency of the fan.

Multiple linear regression (MLR) was used as the baseline predictive technique. If the results of the regression model were unsatisfactory (taking the values of the correlation coefficient between observed and predicted output into account), the more complex machine learning algorithms was used and analyzed: support vector regression machine (SVR) and multilayer perceptron neural network (MLP).

Additionally, the k-means clustering algorithm was implemented in order to extract similar coal samples in the data set. This approach provided additional information to develop a more accurate predictive model.

The research methodology has the following steps:

- ◆ data preparation,
- ◆ the determination of the structure of MLR models for each of the analyzed parameters,
- ◆ evaluation of forecasting results of MLR models,
- ◆ the determination of the structure of machine learning models (SVR, MLP),
- ◆ cluster analysis (k-means method),
- ◆ development of a hybrid model, based on the results of a cluster analysis and machine learning techniques,
- ◆ evaluation of forecasting results.

The Statistica package and python programming language together with the scikit-learn library were used for the analysis.

## 2.3. Data preparation for prediction

A source of data used in the work were experiments of dry separation of coal samples. The experiments aimed to evaluate the process efficiency, based on parameters of feed and selected sorter parameters. Coal samples were acquired from 8 Polish hard coal mines:

KWK Ziemowit, KWK Chwałowice, KWK Jankowice, KWK Bielszowice Ruch Bielszowice, KWK Bielszowice Ruch Pokój, KWK Bolesław Śmiały, KWK Halemba, KWK Sośnica. All of them were received as a grain size class of 0–50 mm.

The preparation of raw coals for the separation process included: quating the materials to create uniform samples and dividing the averaged samples into different grain size classes: 0–50 mm, 20–50 mm, 0–20 mm, 6–20 mm, 6–12 mm, 6–25 mm, 12–25 mm, 25–50 mm, 8–25 mm. Experiments of dry separation were carried out for the frequency of the fan being changed in range from 26 to 50 Hz.

The feed materials and separation products were analyzed in the laboratory to determine their physicochemical parameters:

- ◆ the free moisture ( $fm$ ),
- ◆ ash content ( $a$ ),
- ◆ moisture content ( $m$ ),
- ◆ sulfur content ( $s$ ),
- ◆ higher heating value ( $hhv$ ),
- ◆ lower heating value ( $lhv$ ).

The following variables were applied as input data for the predictive models:

- ◆ the moisture content of the feed ( $m_f$ ) and the frequency of the fan ( $f$ ) for the prediction moisture content in the upper fraction ( $m_u$ ),
- ◆ the sulfur content of the feed ( $s_f$ ) and the frequency of the fan ( $f$ ) for the prediction sulfur content in the upper fraction ( $s_u$ ),
- ◆ the ash content of the feed ( $a_f$ ) and the frequency of the fan ( $f$ ) for the prediction ash content in the upper fraction ( $a_u$ ),
- ◆ the lower heating of the feed ( $lhv_f$ ) and the frequency of the fan ( $f$ ) for lower heating value the product ( $lhv_u$ ).

The measured parameters were significantly different in order of magnitude. Therefore a normalization procedure was required, according to the equation below:

$$X' = \frac{X - \min(X)}{\max(X) - \min(X)} \quad (4)$$

Both input and output data were subjected to the normalization procedure.

The data set was divided in a random manner into a training set (80% of samples) and a test set (20% of samples). The analyzed models were established based on the training set. The test set was used to evaluate the models.

## 2.4. Evaluation of models

The performance of the predictive models was evaluated through the following parameters (Osowski 2013):

- ◆ mean absolute error (*MAE*)

$$MAE = \frac{1}{n} \left( \sum_{i=1}^n |d_i - y_i| \right) \quad (5)$$

- ◆ root mean squared error (*RMSE*)

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n |d_i - y_i|^2} \quad (6)$$

- ◆ coefficient of correlation (*r*)

$$r = \frac{\sum_{i=1}^n (d_i - \bar{d})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (d_i - \bar{d})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (7)$$

- ◆ agreement index (*AI*)

$$AI = 1 - \frac{\sum_{i=1}^n (d_i - y_i)^2}{\sum_{i=1}^n (|d_i - \bar{d}| + |y_i - \bar{d}|)^2} \quad (8)$$

- ↻  $n$  – number of samples,
- $d_i$  – observed value,
- $y_i$  – predicted value.

### 3. Results and discussion

#### 3.1. Multiple linear regression

Coefficients of the regression equation were determined by the least squares procedure. Only statistically important variables were included in models.

Following equations were determined:

- ◆ for moisture content:

$$m_u = -0.057 + 0.053 \cdot f + 1.061 \cdot m_f \quad (9)$$

- ◆ for sulfur content:

$$s_u = 0.04 + 1.092 \cdot s_f \quad (10)$$

- ◆ for ash content:

$$a_u = -0.148 \cdot f + 0.701 \cdot a_f \quad (11)$$

- ◆ for lower heating value:

$$lhv_u = 0.348 + 0.154 \cdot f + 0.554 \cdot lhv_f \quad (12)$$

### 3.2. Results of regression models

The performance of multiple linear regression models, examined on the test set, are presented in Table 1. Taking the index  $r$  (correlation coefficient between observed and predicted output) into account, the efficiency of the models determining the moisture and sulfur content in the product is very high. In both cases the value of coefficient  $r$  is close to 1. Very high values are also achieved by the index of agreement  $AI$  of these two models. It was therefore decided that the linear regression model is a sufficient tool to predict the moisture and sulfur content in the upper fraction.

Table 1. Results obtained with MLR model

Tabela 1. Wyniki uzyskane z wykorzystaniem modelu MLR

	<i>MAE</i>	<i>RMSE</i>	<i>r</i>	<i>AI</i>
Moisture	0.0574	0.0598	0.9908	0.9946
Sulfur	0.0871	0.0840	0.9732	0.9866
Ash	0.1684	0.1601	0.6789	0.7847
Calorific value	0.1847	0.1658	0.5427	0.7239

On the other hand, in the case of models predicting the ash content and calorificity of the upper fraction, values of the  $r$ -coefficient differed significantly from 1. It was therefore decided that additional models, applying more complex machine learning algorithms will be developed for these two parameters.

### 3.3. Support Vector Regression

Many SVR network structures with different kernel functions have been tested. Finally, two models were selected – one for the ash content, one for the calorific value. Details about the structures of particular models are given in Table 2. Model constants (capacity and epsilon) were determined by a cross-validation procedure (with the number of ten folds).

Table 2. Parameters of SVR models

Tabela 2. Parametry modeli SVR

Output variable	Model constants	Kernel type	Number of support vectors
Ash	Capacity = 9.0 Epsilon = 0.1	Radial Basis Function (gamma = 0.5)	32
Calorific value	Capacity = 3.0 Epsilon = 0.1	Polynomial (degree = 3, gamma = 0.5, coefficient = 1.0)	28

### 3.4. Multilayer Perceptron Neural network

After testing with multiple MLP network structures, two final models were selected – one for ash content, one for calorific value. The selected structures had one output and one hidden layer with two neurons (Fig. 2). The logistic function, which takes values in the range of (0,1), was used as an activation function in the created neural networks.

The learning process was carried out for the following conditions and parameters:

- ◆ learning algorithm: gradient descent,
- ◆ number of epoch: 10000,

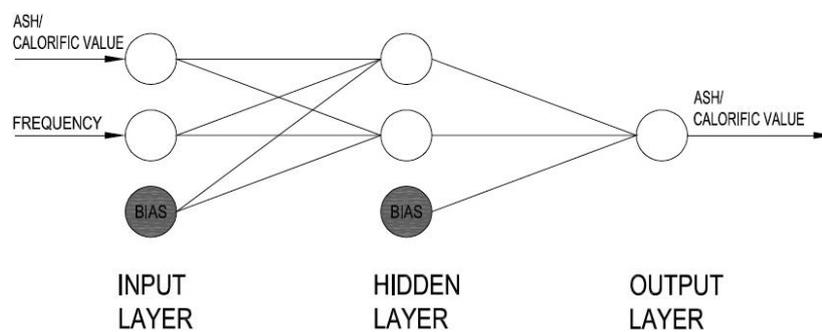


Fig. 2. Architecture of neural network selected for prediction ash content and calorific value

Rys. 2. Architektura sieci neuronowej wybrana do prognozowania zawartości popiołu i wartości opałowej

- ◆ learning rate: 0.1,
- ◆ momentum: 0.9,
- ◆ weights initialization: normal randomization (the weights are initialized using normally distributed values, within a range whose mean is zero and standard deviation are equal to one),
- ◆ error function: sum of squares error function (given by the sum of differences between the observed values and predicted values).

Learning curves of the described network structures are presented in Figure 3. Details about the structures of selected networks and sensitivity analysis of particular variables are given in Table 3.

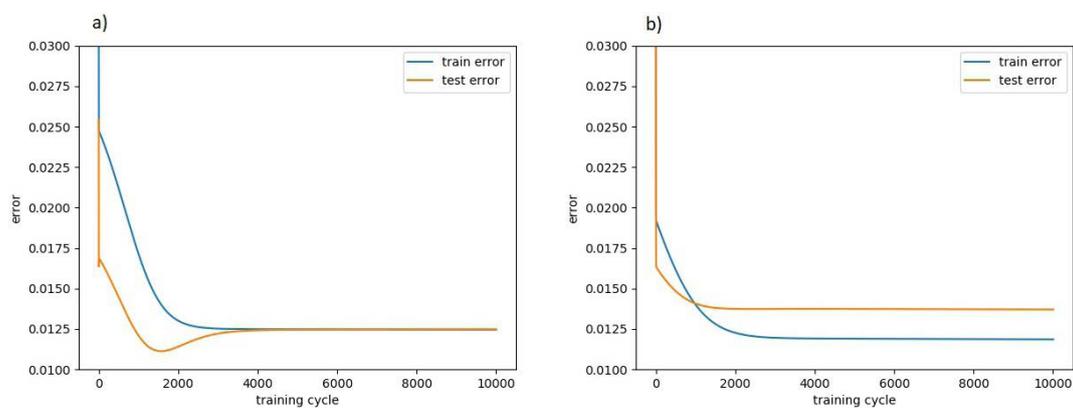


Fig. 3. Error curve of neural network  
a) model predicting ash content, b) model predicting calorific value

Rys. 3. Krzywa uczenia sieci neuronowej  
a) model prognozujący zawartość popiołu, b) model prognozujący wartość opałową

Table 3. Parameters of MLP models

Tabela 3. Parametry modeli MLP

Output variable	Network architecture	Inputs	Sensitivity analysis
Ash	2-2-1	$a_f$	2.3689
		$f$	1.2493
Calorific value	2-2-1	$l_{hv}$	1.8063
		$f$	1.3680

### 3.5. Results of SVR and MLP models

The performance of SVR and MLP models, examined on test set, are presented in Tables 4–5. Taking the index  $r$  into account, the efficiency of both models determining the calorific value of the upper fraction was higher in comparison to the corresponding linear regression model. Values of the index of agreement  $AI$  were also closer to 1 for MLP and SVR, as well as the values of  $MAE$  and  $RMSE$  statistics were lower compared to the linear regression model.

Table 4. Results obtained with SVR model

Tabela 4. Wyniki uzyskane z wykorzystaniem modelu SVR

	$MAE$	$RMSE$	$r$	$AI$
Ash	0.1717	0.1585	0.6786	0.7935
Calorific value	0.1735	0.1531	0.6143	0.7558

Table 5. Results obtained with MLP model

Tabela 5. Wyniki uzyskane z wykorzystaniem modelu MLP

	$MAE$	$RMSE$	$r$	$AI$
Ash	0.0882	0.1134	0.6850	0.8988
Calorific value	0.0890	0.1146	0.6113	0.8821

In the case of SVR and MLP models predicting ash content in the upper fraction, the values of  $MAE$ ,  $RMSE$  and  $AI$  measures improved in comparison to the corresponding linear regression model. However, taking the coefficient  $r$  into account, the efficiency of prediction with SVR and MLP is similar to the efficiency of the linear regression.

Taking the  $r$ -coefficient into account, the efficiency of SVR and MLP models, predicting ash content and calorific value is comparable. However, for both parameters, the  $AI$  index is higher for the MLP model and the values of  $MAE$  and  $RSME$  statistics are lower compared to the SVR model. In general, the MLP model showed better predictive efficiency.

Since the values of the  $r$ -coefficient for models predicting ash content and calorific value of the upper fraction (both MLP and SVR) were still different significantly from 1, it was decided to improve these tools by applying the clustering algorithm.

### 3.6. Cluster analysis

A cluster analysis was implemented in order to extract additional information about coal samples used as feed material. It was assumed that samples similar in terms of physicochemical properties (such as moisture or ash content) may behave similarly during processing on a pneumatic sorter. Information about the place of sample origin may be insufficient. Coals from the same mine may differ significantly from each other, but at the same time may be physicochemically similar to coals from other origin. Therefore, the use of clustering algorithms was decided.

The k-means algorithm was chosen and applied as a clustering technique in the presented work. The assumptions used for k-means implementation were as follows:

- ◆ initial cluster centers were initialized with random values,
- ◆ six variables (parameters of coal feed) were taken into account in calculations: the free moisture, ash content, moisture content, sulfur content, higher heating value, lower heating value,
- ◆ Euclidean distance between cluster centers was applied as a distance measure,
- ◆ the number of clusters were determined by the cross-validation procedure (with the number of ten folds).

The general idea of cross-validation algorithm is to divide the total data set into a number of  $\nu$  folds (sub-sets). Then, the same type of analysis (for instance, determination of cluster centers) is successively applied to the samples belonging to the  $\nu-1$  folds (sub-sets). After that the results of the analysis are applied to fold  $\nu$  to perform validation. Finally, the results of the calculations for consecutive sub-sets  $\nu$  are averaged (Osowski 2013).

### 3.7. Results of the cluster analysis

The cross-validation procedure extracted 5 clusters from the data. The normalized mean values of the analyzed variables in particular clusters are presented in Figure 4.

According to the obtained results, the following profiles of the studied coals may be created:

- ◆ group 1 – high calorific coals with the average content of ash, moisture and sulfur and high content of free moisture,
- ◆ group 2 – high calorific coals with the average content of ash and sulfur and low moisture content,
- ◆ group 3 – high calorific coals with low moisture and sulfur content and the average content of ash,
- ◆ group 4 – low calorific coals with low moisture and sulfur content and high content of ash,
- ◆ group 5 – high calorific coals with average content of ash and very high content of sulfur and moisture.

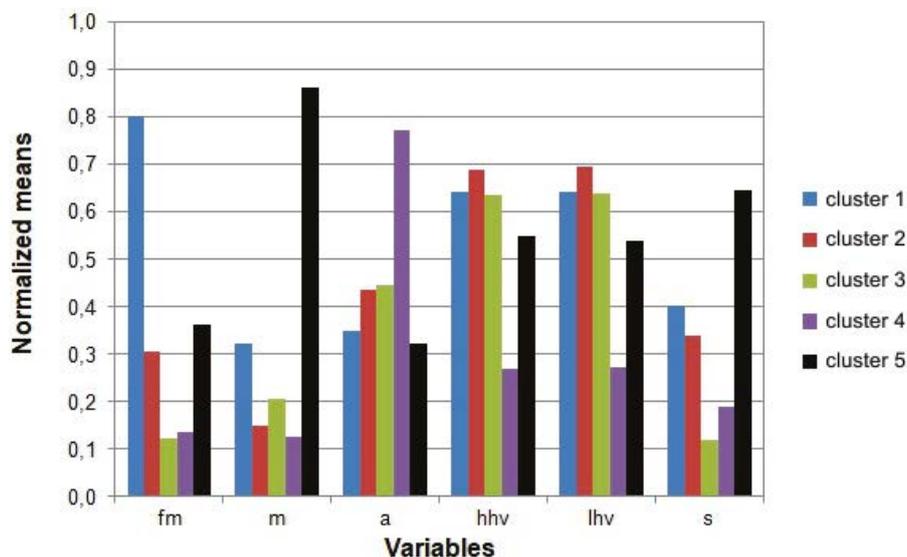


Fig. 4. The normalized mean values of the analyzed variables in particular clusters

Rys. 4. Znormalizowane średnie wartości analizowanych zmiennych w poszczególnych klastrach

### 3.8. Hybrid algorithm

The hybrid algorithm was built on the basis of the MLP and SVR models and the results of the k-means algorithm. Classification into one of the particular coal groups (which were extracted by the clustering algorithm) was inserted into the structures of the above mentioned algorithms as an additional categorical variable (the structure of the neural network including the results of the cluster analysis are presented in Fig. 5). This procedure was carried out for ash content and the calorific value of the upper fraction. The results of hybrid models are presented in Tables 6–7.

The insertion of an additional variable, describing the category of the coal group, did not improve the quality of the model predicting ash content in the upper fraction into the structure of SVR (taking the  $r$ -coefficient into account). However, in the case of determining the calorific value of the upper fraction, the improvement of the prediction capabilities of the model may be observed (from  $r = 0.61$  to  $r = 0.65$ ).

The insertion of information about the coal category into the structure of MLP improved the quality of model predicting ash content in the upper fraction (from  $r = 0.69$  to  $r = 0.75$ ) as well as the quality of the model determining the calorific value of the product (from  $r = 0.61$  to  $r = 0.65$ ).

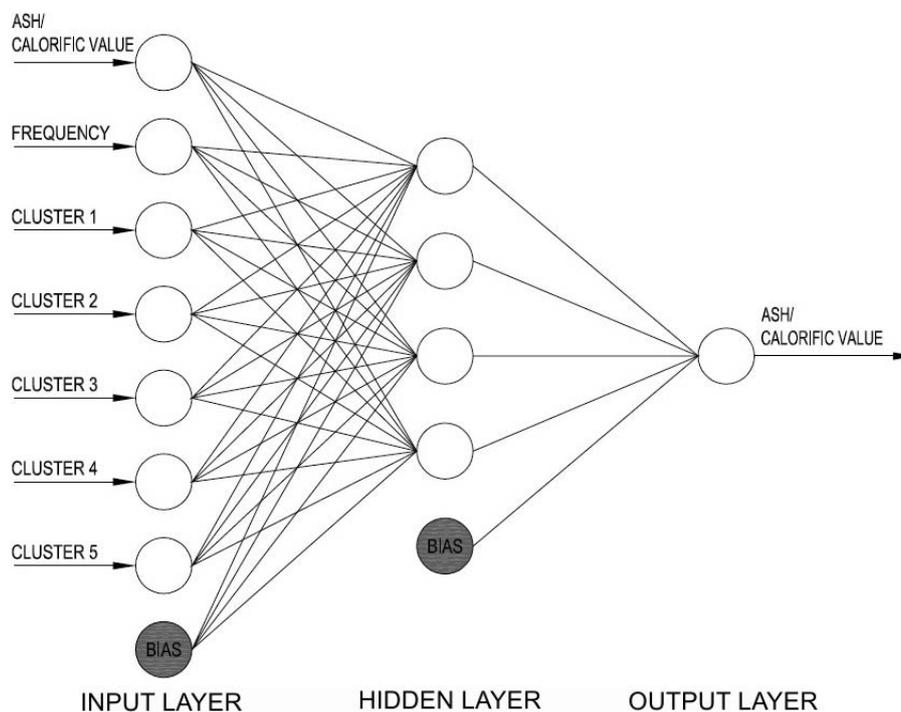


Fig. 5. Structure of a neural network with a categorical variable (classification into one particular coal group)

Rys. 5. Struktura sieci neuronowej ze zmienną kategorię (przynależność do określonej grupy węglowej)

Table 6. Results obtained with SVR model combined with k-means clustering

Tabela 6. Wyniki uzyskane z wykorzystaniem modelu SVR połączonego z grupowaniem

	<i>MAE</i>	<i>RMSE</i>	<i>r</i>	<i>AI</i>
Ash	0.1619	0.1499	0.6825	0.7991
Calorific value	0.1512	0.1505	0.6518	0.7865

Table 7. Results obtained with MLP model combined with k-means clustering

Tabela 7. Wyniki uzyskane z wykorzystaniem modelu MLP połączonego z grupowaniem

	<i>MAE</i>	<i>RMSE</i>	<i>r</i>	<i>AI</i>
Ash	0.0906	0.1112	0.7462	0.9172
Calorific value	0.0895	0.1117	0.6506	0.9970

In general, the hybrid MLP model has a better predictive efficiency than the hybrid SVR model for both parameters (ash content and calorific value). Values of  $r$  and AI coefficients are higher for the neural structure and values of MAE and RSME statistics are lower compared to the SVR algorithm.

## Conclusions

1. The aim of presented paper was to develop, apply and evaluate prediction models for coal enrichment using the AMSEP innovative pneumatic sorter. These models are based on selected machine learning algorithms – multiple linear regression (MLR), support vector regression machine (SVR), multilayer perceptron neural network (MLP). The chosen parameters of the upper fraction (ash content, sulfur content, moisture content and calorific value) of a dry separation process were determined with the above-mentioned models.
2. The mechanical processing on the AMSEP pneumatic sorter has an impact on improving the parameters of the samples – the calorific value increases, and the moisture, sulfur and ash content is reduced.
3. The prediction of the characteristics of mineral processing products is a very complex issue. These parameters are dependent on the physicochemical properties of the feed material, its preparation and grain size as well as the working conditions of the chosen separator.
4. Machine learning techniques enable the processing of a very large amount of data and may be used in situations when the relationships between variables are difficult to present by strict mathematical equations. All these features make machine learning an excellent tool for simulating complex processes of coal separation.
5. Efficiency of the linear regression models determining the moisture and sulfur content in the product was very high (values of coefficient  $r$  were close to 1). The MLR technique is a sufficient tool to predict the values of these parameters in product of dry separation.
6. Taking the index  $r$  into account, the efficiency of the SVR and MLP algorithms determining the calorific value of the upper fraction was higher in comparison to the corresponding linear regression model. The efficiency of SVR and MLP, used to predict ash content, are similar to performance of corresponding linear regression models.
7. In order to improve the efficiency of SVR and MLP models, the k-means clustering technique has been implemented. The role of this algorithm was to extract similar coal samples in the data set. Finally, 5 clusters (different coal groups) were determined. Insertion information about these groups into the SVR model improved its quality in case of the prediction of the calorific value of the upper fraction. The insertion of information about the coal category into the MLP structure of improves the efficiency of the model dedicated to ash prediction as well as to calorific value prediction.
8. The methodology presented in the work allows for the prediction of the parameters of a dry separation product based on the basic and most frequently determined characteristics of the coal material with satisfactory efficiency. However, in the future, the presented

approach is planned to be expanded by additional characteristics of the feed, which may have a potential impact on the product parameters – such as petrographic composition, densymetric analysis and selectivity curves.

*This paper presents the results of Statutory Research No 11326018-173 and KIC InnoEnergy project “Novel dry sorter for coal processing and coal recovery from mine originating wastes” (acronym AMSEP).*

## REFERENCES

- Allahkarami et al. 2017 – Allahkarami, E., Salmani Nuri, O., Abdollahzadeh, A., Rezai, B. and Maghsoudi, B. 2017. Improving estimation accuracy of metallurgical performance of industrial flotation process by using hybrid genetic algorithm–artificial neural network (GA-ANN). *Physicochemical Problems of Mineral Processing* 53(1), pp. 366–378.
- Ali et al. 2018 – Ali, D., Hayat, M. B., Alagha, L. and Molatlhegi, O.K. 2018. An evaluation of machine learning and artificial intelligence models for predicting the flotation behavior of fine high-ash coal. *Advanced Powder Technology* (in press).
- Baic, I. and Blaschke, W. 2013. Analysis of the possibility of using air concentrating tables in order to obtain clean coal fuels and substitute natural aggregates (*Analiza możliwości wykorzystania powietrznych stolów koncentracyjnych do otrzymywania węglowych paliw kwalifikowanych i substytutów kruszyw*). *Polityka Energetyczna – Energy Policy Journal* 16(3), pp. 247–260 (in Polish).
- Bergh, L. 2016. Artificial Intelligence in Mineral Processing Plants: An Overview [In:] Petrillo, A. et al. red. *Proceedings of the 2016 International Conference on Artificial Intelligence: Technologies and Applications*. Bangkok, 24–25 January 2016. Paryż: Atlantis Press, pp. 278–281.
- Blaschke, W. 1976. *Wzbogacanie powietrzne. Poradnik Górnika t. V*. Katowice: Wyd. Śląsk, pp. 556–563.
- Blaschke, W. 2009. *Przeróbka węgla kamiennego – wzbogacanie grawitacyjne*. Kraków: Wyd. IGSMiE PAN, 234 pp.
- Boser et al. 1992 – Boser, B.E., Guyon, I.M. and Vapnik, V.N. 1992. A training algorithm for optimal margin classifiers [In:] Haussler, D. red. *COLT'92 Proceedings of the fifth annual workshop on Computational learning theory*. Pittsburg, 27–29 July 1992. NY: ACM, pp.144–152.
- Chelgani et al. 2018 – Chelgani S.C., Shahbazi B. and Hadavandi E. 2018. Support vector regression modeling of coal flotation based on variable importance measurements by mutual information method. *Measurement* 114, pp. 102–108.
- Cortes, C. and Vapnik, V. 1995. Support vector networks. *Machine Learning* 20(3), pp. 273–297.
- Drzymała, J. 2001. Basics of mineral processing (*Podstawy mineralurgii*). Wrocław: Oficyna Wydawnicza Politechniki Wrocławskiej, 478 pp. (in Polish).
- Feng 2015 – Feng, Q., Zhang, J., Zhang, X. and Wen, S. 2015. Proximate analysis based prediction of gross calorific value of coals: a comparison of support vector machine, alternating conditional expectation and artificial neural network. *Fuel Processing Technology* 129, pp. 120–129.
- Goodfellow et al. 2016 – Goodfellow, I., Bengio, Y. and Courville, A. 2016. *Deep Learning*. Cambridge: MIT Press, 775 pp.
- Jonsson et al. 2002 – Jonsson, K., Kittler, J., Li, Y.P. and Matas, J. 2002. Support vector machines for face authentication. *Image and Vision Computing* 20(5–6), pp. 369–375.
- Jorjani et al. 2009 – Jorjani, E., Poorali, H.A., Sam, A., Chelgani, S.C., Mesroghli, S. and Shayestehfar, M.R. 2009. Prediction of coal response to froth flotation based on coal analysis using regression and artificial neural network. *Minerals Engineering* 22(11), pp. 970–976.
- King, R.P. 2001. *Modeling and simulation of mineral processing systems*. Oxford: Butterworth-Heinemann, 416 pp.
- Lippmann, R.P. 1987. An introduction to computing with neural nets. *IEEE Acoustical Speech and Signal Processing Magazine* 4(2), pp. 4–22.

- Liu et al. 2018 – Liu, B., Fu, Z., Wang, P., Liu, L., Gao, M., and Liu, J. 2018. Big-Data-Mining-Based Improved K-Means Algorithm for Energy Use Analysis of Coal-Fired Power Plant Units: A Case Study. *Entropy* 20(9), pp. 702.
- Liu et al. 2006 – Liu, Y., Wu, M. and Qian, J. 2006. Predicting Coal Ash Fusion Temperature Using Hybrid of Ant Colony Algorithm and BP Neural Network [In:] *4<sup>th</sup> IEEE International Conference on Industrial Informatics In Industrial Informatics*, Singapore 16–18 August 2006 IEEE pp. 805–809.
- Michalak et al. 2012 – Michalak, M., Iwaszenko, S. and Wierzychowski, K. 2012. Coal washing modelling in a jig (*Coal analyzer – modelowanie procesu wzbogacania węgla w osadzarkie*). *Mechanizacja i Automatyza Górnictwa* 50(9), pp. 18–22 (in Polish).
- Oruç et al. 2010 – Oruç, F., Özgen, S. and Sabah, E. 2010. An enhanced-gravity method to recover ultra-fine coal from tailings: Falcon concentrator. *Fuel* 89(9), pp. 2433–2437.
- Osowski, S. 2013. *Data mining methods and tools (Metody i narzędzia eksploracji danych)*. Legionowo: Wyd. BTC, 388 pp. (in Polish).
- Raina, C.K. 2016. A review on machine learning techniques. *International Journal on Recent and Innovation Trends in Computing and Communication* 4(3), pp. 395–399.
- Sahu et al. 2009 – Sahu, H., Panigrahi, D.C. and Mohapatra, S.S. 2009. Classification Of Coal Seams With Respect To Their Spontaneous Heating Susceptibility Using K-Means Clustering [In:] *Ninth International Mine Ventilation Congress*, New Delhi, 10–13 November 2009, Oxford: IBH Publishing Company, pp. 273–287.
- Tumidajski, T. 2010. Actual tendencies in description and mathematical modeling of mineral processing (*Aktualne tendencje w opisie i modelowaniu matematycznym procesów przeróbki materiałów uziarnionych*). *Gospodarka Surowcami Mineralnymi – Mineral Resources Management* 26(3), pp. 111–123 (in Polish).
- Uyanik, G. and Guler, N. 2013. A Study on Multiple Linear Regression Analysis. *Procedia – Social and Behavioral Sciences* 106, pp. 234–240.
- Vapnik, V.N. 1995. *The Nature of Statistical Learning Theory*. NY: Springer, pp. 314.
- Wang et al. 2012 – Wang, J., Li, L., Niu, D. and Tan, Z. 2012. An annual load forecasting model based on support vector regression with differential evolution algorithm. *Applied Energy* 94, pp. 65–70.
- Zhengwei et al. 2007 – Zhengwei, L., Shixiong, X. and Niuqiang, Zhanguo, X. 2007. Coal Thickness Prediction Based on Support Vector Machine Regression [In:] Feng W. red. *Eighth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing, Qingdao*, 30 July–1 August 2007. IEEE pp. 379–383.
- Zhou et al. 2014. – Zhou, H., Tang, Q., Yang, L., Yan, Y., Lu, G. and Cen, K. 2014. Support vector machine based online coal identification through advanced flame monitoring. *Fuel* 117, pp. 944–951.
- Zucker et al. 2015 – Zucker, G., Habib, U., Blöchle, M., Judex, F. and Leber, T. 2015. Sanitation and analysis of operation data in energy systems. *Energies* 8(11), pp. 12776–12794.

**PREDICTION OF QUALITY PARAMETERS OF A DRY AIR SEPARATION  
PRODUCT USING MACHINE LEARNING METHODS**

Keywords

artificial neural networks, multiple linear regression,  
support vector machine (SVM), dry coal separation

Abstract

The purpose of the work was to predict the selected product parameters of the dry separation process using a pneumatic sorter. From the perspective of application of coal for energy purposes,

determination of process parameters of the output as: ash content, moisture content, sulfur content, calorific value is essential. Prediction was carried out using chosen machine learning algorithms that proved to be effective in forecasting output of various technological processes in which the relationships between process parameters are non-linear. The source of data used in the work were experiments of dry separation of coal samples. Multiple linear regression was used as the baseline predictive technique. The results showed that in the case of predicting moisture and sulfur content this technique was sufficient. The more complex machine learning algorithms like support vector machine (SVM) and multilayer perceptron neural network (MPL) were used and analyzed in the case of ash content and calorific value. In addition, k-means clustering technique was applied. The role of cluster analysis was to obtain additional information about coal samples used as feed material. The combination of techniques such as multilayer perceptron neural network (MPL) or support vector machine (SVM) with k-means allowed for the development of a hybrid algorithm. This approach has significantly increased the effectiveness of the predictive models and proved to be a useful tool in the modeling of the coal enrichment process.

#### PRZEWIDYWANIE PARAMETRÓW JAKOŚCIOWYCH PRODUKTU SUCHEJ SEPARACJI WĘGLA METODAMI UCZENIA MASZYNOWEGO

##### Słowa kluczowe

sztuczne sieci neuronowe, sucha separacja węgla,  
wielokrotna regresja liniowa, maszyna wektorów nośnych (SVM)

##### Streszczenie

Celem pracy było prognozowanie wybranych parametrów produktu procesu suchej separacji za pomocą sortera pneumatycznego. Z punktu widzenia zastosowania węgla do celów energetycznych niezbędne jest określenie parametrów procesowych wydobycia, takich jak: zawartość popiołu, zawartość wilgoci, zawartość siarki czy wartość kaloryczna. Prognozowanie przeprowadzono przy użyciu wybranych algorytmów uczenia maszynowego, które okazały się skuteczne w prognozowaniu wyjścia różnych procesów technologicznych, w których zależności między parametrami procesu są nieliniowe. Źródłem danych wykorzystanych w pracy były eksperymenty procesu suchej separacji węgla. Zastosowano wieloraką regresję liniową jako bazową metodę predykcyjną. Wyniki pokazały, że w przypadku przewidywania zawartości wilgoci i siarki technika ta była wystarczająca. Bardziej złożone algorytmy uczenia maszynowego, takie jak maszyna wektorów nośnych (SVM) i perceptron wielowarstwowy (MLP) zostały wykorzystane i przeanalizowane w przypadku zawartości popiołu i wartości opałowej. Ponadto wdrożono technikę k-średnich. Rolą analizy skupień było uzyskanie dodatkowych informacji na temat próbek węgla będących wejściem procesu. Połączenie technik, takich jak perceptron wielowarstwowy (MLP) lub maszyna wektorów nośnych (SVM) z metodą k-średnich pozwoliło na opracowanie hybrydowego algorytmu. Takie podejście znacznie zwiększyło efektywność modeli predykcyjnych i okazało się użytecznym narzędziem w modelowaniu procesu wzbogacania węgla.