Journal Pre-proof

The following article:

Assessment of the environmental impact of hard coal mining waste disposal sites using machine learning algorithms, with an indication of important features influencing the selection of learning algorithms

Joanna FAJFER and Michał ROLKA

is accepted, peer reviewed article assigned to issue 4 of volume 69 with article nr 65, that is not yet appropriately edited, but is citable using DOI:

https://doi.org/10.7306/gq.1838

This version will undergo additional copyediting, typesetting and review before it is published in its final form.

Assessment of the environmental impact of hard coal mining waste disposal sites using machine learning algorithms, with an indication of important features influencing the selection of learning algorithms

Joanna FAJFER[1,*] and Michał ROLKA[2]
[1]Polish Geological Institute - National Research Institute, Rakowiecka 4, 00-975 Warszawa, Poland ORCID ID: 0000-0002-7382-5383
[2] Polish Geological Institute - National Research Institute, Upper Silesian Branch, Królowej Jadwigi 1, 41-200 Sosnowiec, Poland
*Corresponding author, e-mail: joanna.fajfer@pgi.gov.pl

Key words: machine learning algorithms, hard coal mining waste, environmental impact, important features, disposal sites.

Received: October 20, 2025; accepted: December 23, 2025

Associate editor: Wojciech Granoszewski

Abstract
This paper presents the potential application of supervised machine learning algorithms to assess the environmental impact of mine disposal sites. Algorithms available for Python from the scikit-learn and pandas libraries were applied to a group of sites representing mine waste dumps used for the disposal of hard coal mining waste. Each disposal site was described with 11 attributes (site characteristics, waste characteristics, groundwater, surface water, air, soil, atmospheric factors, geology, geohazards, nature, and human environment) and 73 features (categorical, numerical, and descriptive) detailing the sites' environmental impact. As a result of applying the learning process to training data and verifying it on test data, prediction results of at least 80% were obtained for all algorithms tested. The results indicate that the best algorithm for determining the environmental impact of the waste dumps would be the BernouliNB algorithm (86% prediction accuracy), followed by the RidgeClassifier algorithm (87% prediction accuracy), with the currently available training dataset. Potential extension of the dataset could improve the results of the MLPClassifier, Support Vector Machine, and LogisticRegression algorithms.

Introduction
Hard coal mining has since the 15th century generated waste, deposited in heaps in former mining areas, near the then-existing mining plants (Jaros, 1975; Piątek, 1995; Rostański, 2006; Chudy et al., 2014; Świtała-Trybek and Świtała-Mastalarz, 2018) and mostly located in exploited open-pit deposits of sand and gravel without protection layers above the permeable bedrock (Szczepańska-Plewa et al., 2010). The location of these disposal sites directly in the environment (without the protections required by current legislation) has contributed to their negative impact on groundwater, surface water, air, soil and vegetation (Kostrz-Sikora et al., 2013). The extent and type of impacts result from the sum of physicochemical processes occurring in the mass of accumulated waste, the unfavorable hydrogeological and geological conditions in the surrounding area, and the lack of technical solutions to limit the release of potential contaminants into the environment. The potential negative impact of mining waste dumps on the environment is influenced by several factors, the most important of which include: the chemical composition of the accumulated waste, its compaction, selective or non-selective waste storage, rainwater infiltration through the structure, ventilation conditions (Konior, 2006; Czajkowska et al., 2018), hydrogeological and hydrological conditions in the area of the site, climatic factors, the time of waste deposition, and the extent and nature of geohazards (landslides, floods, and mining-induced deformation in the areas where the heap is located; Sroga et al., 2017; Baza Hałdy). The chemical composition of the accumulated waste is determined by the type of mineral extracted and the technological processes used during its processing, which in turn influence the diversity of impacts of the dumps on the soil and water environment, as well as on air pollution.

The impact of hard coal mining and processing waste disposal sites on groundwater and surface water is mainly related to the leaching of sulphates, chlorides and heavy metals from the accumulated waste (Twardowska, 1981; Szczepańska, 1987; Chudy and Marszałek, 2010: Czajkowska et al., 2018; Łaganowska, 2019). The following elements were found in these waters: Al, Mn, Fe, V, Zn, Li, Co, Cd, Be, Pb and Cu (Twardowska, 1981; Szczepańska, 1987; Szczepańska and Twardowska; 1999;

Stefaniak and Twardowska, 2009; Szczepańska-Plewa et al., 2010; Zając and Zarzycki, 2013; Chudy et al., 2014; Czajkowska et al., 2018; Łaganowska, 2019; Wolkersdorfer and Mugova, 2022),
The impact of hard coal waste heaps on soils may be related to their increased salinity and sulphur content as well as the concentration of heavy metals (Rusin et al., 2018; Piekut et al., 2018). Sulphur compounds and tar substances resulting from the self-heating of mining waste accumulated on burned heaps are washed out by rainwater from the dumps and permeate into the soil (Styrol, 2020). Heavy metals, including Zn, Pb, Cd, Cu, Ni, As, Cr and Hg, were found in the soils (Rusin et al., 2018, Piekut et al., 2018). In the area around the waste dumps, air pollution may occur through the release of fine dust from their surfaces, that lack or are only sparsely covered with vegetation. Furthermore, in the case of self-heating and spontaneous combustion of hard coal waste heaps, CO_2, CO, SO_2, H_2S, CH_4 are released into the atmosphere (Różański, 2019; Fabiańska et al., 2019; Styrol, 2020).

Current legal regulations specifically concern reduction of the negative impact of waste disposal facilities on the environment (Polish Mining Waste Act, 2008; Polish Waste Act, 2023), through the obligation to monitor environmental components (Regulation, 2014). As part of the monitoring of waste heaps, compounds and indicator parameters (specific to the type of waste deposited) are examined in surface water, leachate, and groundwater, and subsidence of the surface of the mining waste disposal facility is monitored (Regulation, 2014). Mining waste disposal sites have been adapted to applicable environmental protection law (i.e., monitoring is conducted) or closed if such technological possibilities are not justified (Polish Mining Waste Act, 2008). However, disposal sites that were once used by the mining industry, and for which it is impossible to identify their owners or current managers, are not environmentally monitored (Glubniak-Witwicka et al., 2012). The environmental impact of historical mining waste dumps has been the subject of many studies (Paszcza and Krogulski, 2006; Pikoń and Bugla, 2007; Zdechlik et al., 2011; Foltyn et al., 2011; Wróbel et al., 2012; Stefaniak et al., 2013; Chrzan and Mojza, 2018; Sołtysiak and Różkowski, 2025). Research on the environmental impact of waste dumps, dumping grounds and landfills where mining waste was deposited has also been conducted at the Polish Geological Institute - NRI (PGI-NRI) since 2012. As part of the work, an inventory of mining waste dumps, including historical ones, was conducted and their negative environmental impact was assessed. This assessment was based on archival and current research results, historical information from the literature on groundwater and surface water, air pollution (in some cases), soil and the chemical composition of waste deposited in the dumps. The inventory conducted constituted material for the diagnosis of environmental conditions in the area of individual mining waste disposal sites, as well as for the assessment and scope of their potential impact at a national scale (Fajfer et al., 2013; Kostrz-Sikora et al., 2013; Sroga et al., 2017; Baza Hałdy).

Assessing the environmental impact of historic waste dumps is a complex process requiring a thorough understanding of the chemical composition of the stored waste, as well as of the geological, hydrogeological and hydrological conditions in the area surrounding the dump. This involves conducting a series of tests of physicochemical parameters (including chemical leaching) in waste samples collected from various depths within the dump, monitoring groundwater (installing piezometers) and surface water, as well as soils in the area surrounding the site. This is a lengthy and capital-intensive process, as each dump is analysed individually.
An alternative to this assessment process can be the use of supervised machine learning algorithms. This type of learning utilizes algorithms that map input data onto output data based on sample datasets divided into test and training data, as well as classification (for data separation) and regression (for data matching) algorithms (Sarker, 2021). However, important conditions for using machine learning methods include the appropriate definition of the research goal, having a large dataset and proper preparation of the data for the learning process.

Machine learning (ML) algorithms are currently widely used in the field of environmental sciences (Kuźniar, 2016; Sun et al., 2020; Ma et al., 2021; Haupt et al., 2022; Xia et al., 2022: Uddin et al., 2023; Piotrowska and Dąbrowska, 2024; Pasa et al., 2025). One area of application of ML algorithms, although little explored, is the assessment of the environmental impact of mining waste dumps. Recurrent neural networks have been used to predict the rate of rainwater infiltration through the dump body and the chemical composition of leachate flowing from the dump depending on weather conditions, with particular emphasis on rainfall amounts over the years (Ma et al., 2021). Decision tree algorithms and the long-short-term memory (LSTM) algorithm were used to forecast the amount of leachate generated during spring in waste dumps (with particular emphasis on spring floods resulting from snow and ice melt) (Zhang et al., 2023). Machine learning models: multiple linear regression (MLR), support vector regression (SVR), random forest (RF), decision tree (DT), and extreme gradient boosting (XGB), were used to predict the stability of tailings pond slopes (Chand et al., 2025). Automatic machine learning (AutoML) combined with Bayesian modeling was used to predict the environmental impact assessment of mining activities (Gerassis et al., 2021). Support Vector Machine (SVM), Artificial Neural Network (ANN) and Random Forest (RF) methods were used to predict heavy metal concentrations (Zn, Pb, Mn, Cu and Cd) in soils near a mining waste disposal facility due to acidic drainage seepage into the soil (Trifi et al., 2022). The literature search indicated the application of machine learning methods in various areas of the dump's environmental impact (including soils, slope stability and leachate), but did not provide information regarding the use of such methods to assess comprehensive environmental impact. Modeling processes occurring within heaps, while taking into account external factors, is a very demanding task, both in terms of the proper selection of input data and knowledge of the processes occurring within the facility.

This paper attempts to assess the environmental impact of waste heaps using supervised machine learning algorithms. The goal was to determine whether supervised machine learning algorithms are a suitable tool for this type of task. If the answer is positive, the next step was to find the optimal machine learning algorithm whose prediction results would most accurately represent the environmental impact of mining disposal sites. Supervised machine learning algorithms are primarily used to forecast and predict outcomes based on previously defined patterns. Regression, classification, and clustering algorithms are most commonly used. Depending on the defined research goal and the size of the dataset planned for the study, the appropriate algorithm is selected, preceded by testing stages on as many algorithms as possible using scaling algorithms (reducing the ranges of features in the dataset) describing the object.

Characteristics of the Study Area

Geographical and geological setting. The study area is located in southern Poland in the Upper Silesian Coal Bassin (USCB). Geographically the USCB is sited in the Silesian Upland, the Kraków-Częstochowa Upland and the Oświęcim Basin.The USCB covers a total area of 7250 km^2. Most of the USCB (5 650 km^2) belongs to the Silesian and western Lesser Poland regions, an the smaller part to the Czech Republic (1 600 km^2 in the Moravian-Silesian region). The Carboniferous basement is composed of Precambrian (slates, gneisses), Cambrian (sandstones and mudstones) and Devonian (dark grey and black dolomites, as well as organic and detrital limestones) deposits. The Carboniferous succession begins with a carbonate association, which transitions into marine clastic deposits and then into the molasse-like coal-bearing strata of the Mississippian and Pennsylvanian (Kotas, 1995; Jureczka and Nowak, 2016). These deposits are divided into four lithostratigraphic units: the Paralic Series, the Upper Silesian Sandstone Series, the Mudstone Series, and the Kraków Sandstone Series. A characteristic feature of the Carboniferous coal-bearing deposits is their distinct division into a Paralic Series and a Limnic Series. Formations developed exclusively under terrestrial conditions overlie Paralic formations with a stratigraphic gap. The total thickness of coal-bearing deposits reaches 8,500 m (Jureczka and Nowak, 2016).

The Paralic Series is composed of mudstones, claystones, and sandstones with interbeds conglomerates, coal and carbonaceous shales. The thickness of the strata in this series ranges from ~200 m in the eastern part of the USCB to ~3,800 m in its western part. The numerous coal seams typically have a thickness of 1.0 to 1.5 m. The Limnic Series is represented by the Upper Silesian Sandstone Series, the Mudstone Series, and the Kraków Sandstone Series. The Upper Silesian Sandstone Series is composed of gravels and sandstones, which dominate over the mudstones and claystones. The coal seams in this series are usually thick, commonly with a thickness of 4–8 m, up to a maximum of 24 m (Jureczka et al., 2005). The maximum stratal thickness in this series (700 m) occurs in the western part of the USCB. It decreases towards the east, until it disappears at the north-eastern border. The Mudstone Series is represented by mudstones and claystones with sandstone interbeds. The maximum thickness of the series in the western part of the USCB reaches 2 km to several tens of metres in the eastern part. Coal seams in this series are numerous, reaching thicknesses of up to 1.5 m. The Krakow Sandstone Series is dominated by coarse- and medium-grained sandstones with interbeds of mudstones and claystones, as well as coal seams. The thickness of the coal seams is up to 6.0 m. The overburden consists of Triassic, Neogene (Miocene) and Quaternary deposits, less frequently Permian and Jurassic (Jureczka et al., 2005; Jureczka and Nowak, 2016).

Hydrogeological setting. The USCB is divided into two subregions: Subregion I (northeastern) and Subregion II (southwestern). Subregion I contains Quaternary, Jurassic and Triassic aquifers, hydraulically connected to Carboniferous strata. In Subregion II, the hydraulic connection between Carboniferous and Quaternary aquifers occurs only locally within hydrogeological windows (Różkowski et al., 2013).

The most important aquifers within the USCB are found in Quaternary, Triassic and Carboniferous formations. Quaternary formations have varied hydrogeological conditions. They are fed by infiltrating water from precipitation and surface water. Quaternary aquifers occur in sandy gravel deposits and their thickness varies from a few to several dozen metres (Różkowski et al., 2013). The Triassic hydrogeological profile has three main aquifers : the Muschelkalk, the Roet and the middle and lower Buntsandstein. The main Triassic aquifers occur in carbonate units of the Muschelkalk and the Roet. These are fissured-karst aquifers, highly permeable and strongly water-bearing. The hydrogeological profile of the Carboniferous System of this region includes sets of separate fissured-porous aquifers composed of sandstones and conglomerates. These aquifers, with thicknesses ranging from several to several tens of metres, are isolated from each other by impermeable claystone intercalations. The aquifer of the Upper Carboniferous reaches a thickness of up to 4,500 m (Różkowski, 2008; Różkowski and Różkowski, 2011).

Selection of sites for study. The basic condition for selecting the study sites was the assumption of the availability of as many heaps and facilities as possible and of data describing each site. This condition is important when using supervised machine learning methods.

Mining waste from hard coal mining constitutes the largest group of waste accumulated in heaps and dumps in Poland. Currently, over 436,154 million tonnes of mining and processing waste has been accumulated in sites in the USCB area (Environment, 2025). As part of the work carried out by PGI-NRI in 2012, 104 mining waste disposal sites were inventoried, where waste from hard coal mining and processing had accumulated (Kostrz-Sikora et al., 2013; Fajfer et al., 2025). The selection of study sites was made on the basis of previous work carried out at PGI-NRI (Kostrz-Sikora et al., 2013; Fajfer and Kostrz-Sikora, 2022; Baza Hałdy; Fajfer et al., 2025), available published data, archival materials and expert knowledge. The initial selection of sites was made at the data collection stage (among other things, it was analysed whether a given facility had been demolished and the area designated for other services) and the next selection was made at the stage of constructing the input dataset (the availability of data describing the facilities was analysed in available published data or archival materials). As a result, 48 heaps and facilities from hard coal mining located in the USCB area in the provinces of Silesia and Małopolska (in its western part) were selected for the environmental impact assessment study (Fig.1). These were active (3 sites), inactive (45) with some reclaimed and partially reclaimed (22 sites) where mainly hard coal mining and processing waste was accumulated in a non-selective manner, but also in some cases industrial waste (e.g. slag from heat and power plants).

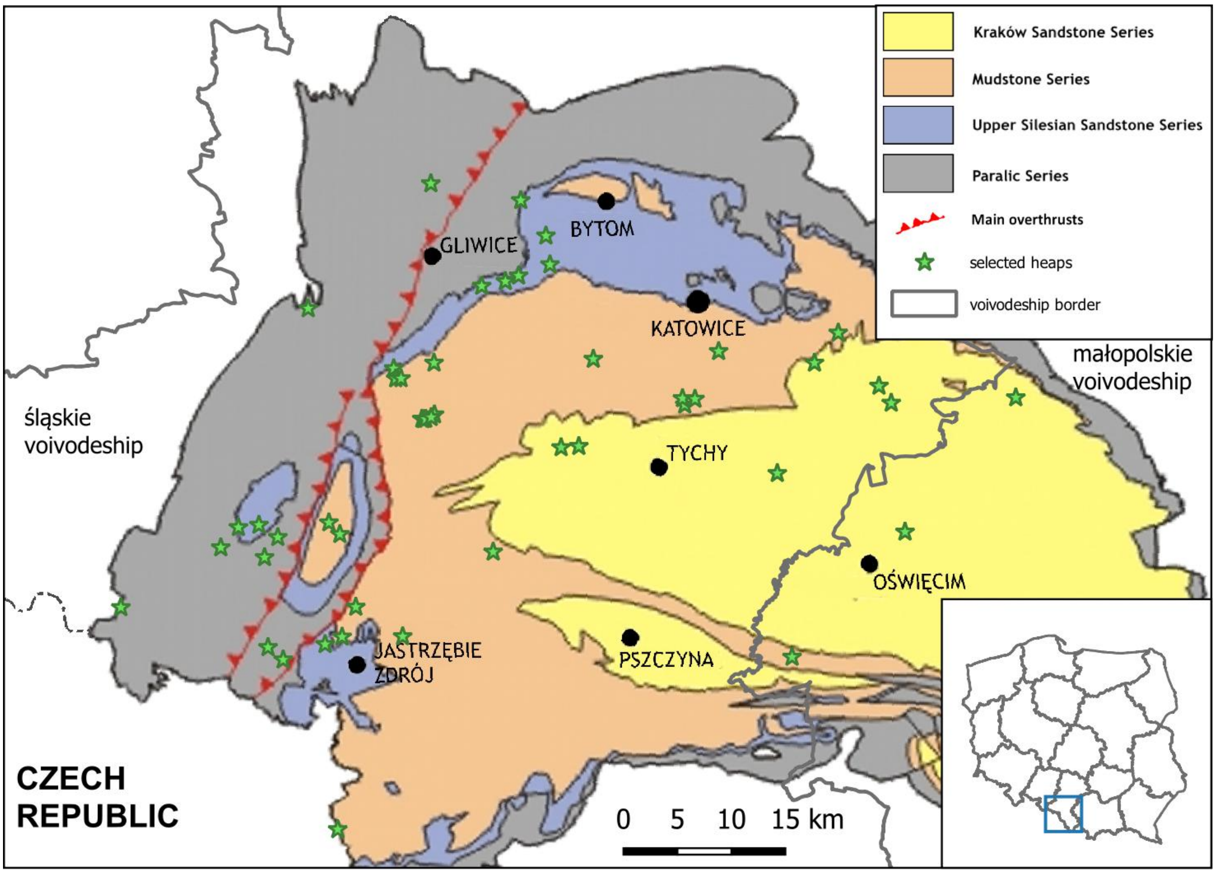


Fig. 1 Location of the sites analysed to test their environmental impact (simplified geological sketch map after Jureczka et al., 1995)

Methods

Attributes and features. Each heap was described with 11 attributes (site characteristics, waste characteristics, groundwater, surface water, air, soils, atmospheric factors, geology, geohazards, nature and human environment) and 73 features (categorical, numerical and descriptive) detailing the site's impact on the environment (Table 1), creating a dataset for further research (Baza Hałdy; Baza OPPI TPP 2.0; CBDG, 2025; Bank Danych o Lasach, 2025; GDOŚ, 2025; INMoTep, 2025; Meteoblue, 2025; Worldclim, 2025; Karty JCWPd, 2025; Karty JCWP, 2025; DTM, 2025). The set of input data characterizing attributes and features may tend to decrease, due to the recovery of waste accumulated on the heaps and no new ones being created, which will impact future analyses.

Table 1 Environmental attributes and features including type of feature

No.	Attribute	Feature	Type of feature
1.	Site characteristic	area, height	numerical
		status of the heap (operational, closed, reclaimed, unreclaimed, complete, exploited – waste recovery from the heap)	categorical
		heap id, heap name, municipality name	descriptive
2.	Waste characteristic	type of waste (mining, processing, mining and processing, other - mining waste not selectively deposited with energy waste)	categorical
		amount of accumulated waste, storage time	numerical
3.	Underground water	occurrence and susceptibility to anthropopressure on the major groundwater reservoir, ocurrence and characteristics of the groundwater bodies (including risk of exposure to pollution), protection zones of groundwater intakes (designated by the authorities)	categorical
		status of the groundwater bodies	descriptive
4.	Surface water	characteristics of the Surface bodies (risk of exposure to pollution, location of the disposal site in areas: waterlogged, marshy, dry, in river valleys, in close proximity to water bodies), risk of flooding	categorical
5.	Air	thermal condition of the heap (active or inactive), presence of vegetation on the site, waste recovery from the heap	categorical
6.	Atmospheric factors	precipitation amount in the area of the dump, average annual rainfall from 1970-2020	numerical
7.	Soils	occurrence of soils of quality class I-III and on organic soils in the area of the site	categorical
8.	Geology	the presence of permeable and impermeable deposits in the base of the facility	categorical
9.	Geohazards	occurrence of landslides and surface flows as a result of the large slope of heaps, operation of a mining plant in the vicinity, presence of deposits of mineral resources	categorical
10.	Nature	occurrence of protected area forms of nature (Natura 2000 areas, Protected Landscape Areas, National Parks, Landscape Parks, nature reserves, ecological areas, protected forest areas)	categorical
11.	Human environment	occurrence of spa area, residential and public buildings, prevailing wind direction, presence of groundwater protection zones (designated by the authorities)	categorical

Source: Fajfer et al., 2025 modified

To predict parameters affecting the assessment of the environmental impact of the waste dumps, the first stage of the study (aimed at determining whether supervised machine learning algorithms are a suitable tool for assessing the environmental impact of waste dumps) involved use of supervised machine learning algorithms from the scikit-learn (Scikit-learn, 2025) package available in Python libraries. Because the problem at hand boils down to the classification of yes/no (0/1) sites, this supervised learning only utilized classification algorithms representing the following classes (Géron, 2018):

- Linear models: RidgeClassifier, SGDClassifier and Perceptron Decision trees: DecisionTreeClassifier and ExtraTreeClassifier;
- Naive Bayesian classifiers: GaussianNB, ComplementNB and BernoulliNB Support vector machines: NuSVC and LinearSVC;
- Nearest neighbor algorithms: KneighborsClassifier;
- Neural network: MLPClassifier;
- Boosting algorithms (Ensemble): ExtraTreesClassifier, AdaBoostClassifier; RandomForestClassifier and GradientBoostingClassifier.

Expert Assessment. Each heap was analysed in terms of representative factors influencing its impact on individual elements of the environment, both natural and human. These factors included the site's impact on groundwater and surface water, air pollution, impact on soil and vegetation and on the human environment. The analyses were carried out taking into account the influence of the following elements: the amount and type of waste deposited, the surface area, the geological structure of the ground, as well as the distance from protected areas, residential buildings and public buildings. Each of the factors defined was assessed and measured by an expert for each of the 48 dumps selected for the study. The expert assessed each variable on an ordinal scale from 0 to 0.9, where 0 meant no impact and 0.9 meant a significant impact. In general, the lack of impact of a given variable on the environment ranged from 0 to 0.3, while the impact of a facility on the environment ranged from above 0.3 to 0.9 for a given variable. The number of values for each factor analysed was variable.

Data Preparation. To prepare the input data for the training process, a pre-prepared dataset consisting of 11 attributes and described by 73 features was used. The data preparation process involved creating new features from related features (e.g., relative height from height measurements). Categorical data was converted into binary features using One Hot Encoding. Descriptive data was converted into features containing word roots and word root pairs and their frequency. After the described data transformations into numerical values, the number of features was 183. Removing low-variability columns reduced the number of features to 139 and removing intercorrelated features lowered the number to 112. Reduction of the number of features is intended to improve the model's training quality, as low-variability features contribute little information, while highly intercorrelated features duplicate information.

Results and discussion

Finding the Optimal Algorithm and Its Hyperparameters. Choosing the optimal machine learning method involves not only selecting the algorithm but also selecting optimal parameter values (different for each algorithm), which influence its behavior and, ultimately, its results. Parameters define algorithm behaviors such as the initial partitioning of input data, the size of the algorithm's internal structures (e.g., number of neurons, tree size), learning sensitivity and data transformation functions. Parameters should be tailored to the input data. Because the number of algorithm and parameter combinations during training and testing reached ~30,000, the RandomizedSearchCV method was used to analyse them to make an acceptable quantity of repetitions in comparison to GridSearchCV, which does runs for all combinations of algorithms and parameters and which would take much more time. This method conducts a series of training and testing sessions using these machine learning algorithms on a training dataset and various randomly selected combinations of their parameter values from a given range. It then creates a ranking list (i.e., identifies the algorithms with the highest prediction confidence). This process required several iterations (including changing the list of algorithms analysed by removing both the lowest-performing algorithms from the analysis list and the best performing ones to better check algorithms that achieved average results) until acceptable results were achieved. RepeatedStratifiedKFold was used as the cross-validation algorithm, an algorithm for repeated splitting of the dataset into training and test sets, while maintaining the balanced representations of classes present in input data. The data number of splits was 4 and 5 and the count of repetitions was 3 to 5,

using 5 splits with 3 repetitions for final calculations. This was also used to prevent overfitting, together with enabling overfitting preventing hyperparameters in specific algorithms, for instance alpha parameter or early stopping.
After searching for the optimal algorithm and its hyperparameters using the method described above on a set of 112 features, only two algorithms achieved 80% accuracy. It was assumed that the reason for this might be a large number of features exceeding the number of objects in the set (the so-called curse of dimensionality (Géron, 2018: a large number of features causes significant variation in both the training samples to which the algorithm adapts and the validation samples, which can lead to erroneous predictions). Therefore, an initial feature reduction was performed using a combination of the SelectFromModel function with the RandomForestClassifier, LogisticRegression, LinearSVC algorithms, which give information about feature importance and also the SelectKBest algorithm with the f_classif function. The reduction aimed to select sets (10, 15, 20, 25, and 30) of the most significant features using selected algorithms, which were then combined into a single set. As a result, sets of 37, 46, 55, 68, and 86 features were selected (because some features were selected by more than one algorithm). Repeated tests on sets with reduced feature counts showed an increase in accuracy to ~88% on sets with 37, 46 and 55 features. Training results on sets with a larger number of retained features were less accurate (<85%). Further tests were conducted on this 55-feature set due to similar results to sets with fewer features and the desire to preserve as many of the original features as possible. The results of the search for optimal algorithms and their hyperparameters for assessing the environmental impact of spoil heaps using RandomizedSearchCV and cross-validation are shown in Table 2. The table shows the algorithms that achieved an average cross-validation accuracy of >85%. Additionally, the standard deviation of the cross-validation results is shown. The algorithms in this table, i.e., MLPClassifier, RidgeClassifier, BernoulliNB, SVC and LogisticRegression, were selected for further testing. Learning quality was verified by generating learning curves for the algorithms noted above (Fig. 2).

Table 2. Ranking of algorithms resulting from the procedure of selecting algorithms and their hyperparameters using RandomizedSearchCV.

ML algorithm	Mean of cross-validation results	Standard deviation of cross-validation results	Scaling algorithm	Number of features
MLPClassifier	89%	0.09	MaxAbsScaler	46
MLPClassifier	88%	0.11	Normalizer	37
MLPClassifier	87%	0.10	MaxAbsScaler	55
RidgeClassifier	87%	0.08	MaxAbsScaler	55
MLPClassifier	87%	0.09	MaxAbsScaler	46
BernoulliNB	86%	0.09	StandardScaler	46
BernoulliNB	86%	0.09	StandardScaler	55
BernoulliNB	86%	0.10	MinMaxScaler	37
RidgeClassifier	86%	0.08	MaxAbsScaler	46
MLPClassifier	86%	0.12	MaxAbsScaler	37
SVC	86%	0.07	RobustScaler	37
SVC	86%	0.09	MaxAbsScaler	55
LogisticRegression	86%	0.10	MaxAbsScaler	68
MLPClassifier	86%	0.11	Normalizer	37
MLPClassifier	86%	0.11	Normalizer	46
SVC	86%	0.08	MaxAbsScaler	46
LogisticRegression	86%	0.09	MaxAbsScaler	55
MLPClassifier	86%	0.09	MaxAbsScaler	55
MLPClassifier	86%	0.09	MaxAbsScaler	46
LogisticRegression	86%	0.09	MaxAbsScaler	46
LogisticRegression	86%	0.10	Normalizer	37
MLPClassifier	86%	0.10	Normalizer	37

MLPClassifier	86%	0.10	MaxAbsScaler	55

Source: this study

The table shows the algorithms that achieved an average cross-validation accuracy of over 85%. Additionally, the standard deviation of the cross-validation results is provided. The algorithms in this table, i.e., MLPClassifier, RidgeClassifier, BernoulliNB, SVC, and LogisticRegression, were selected for further testing. Learning quality was verified by generating learning curves for the algorithms noted above (Fig. 2).

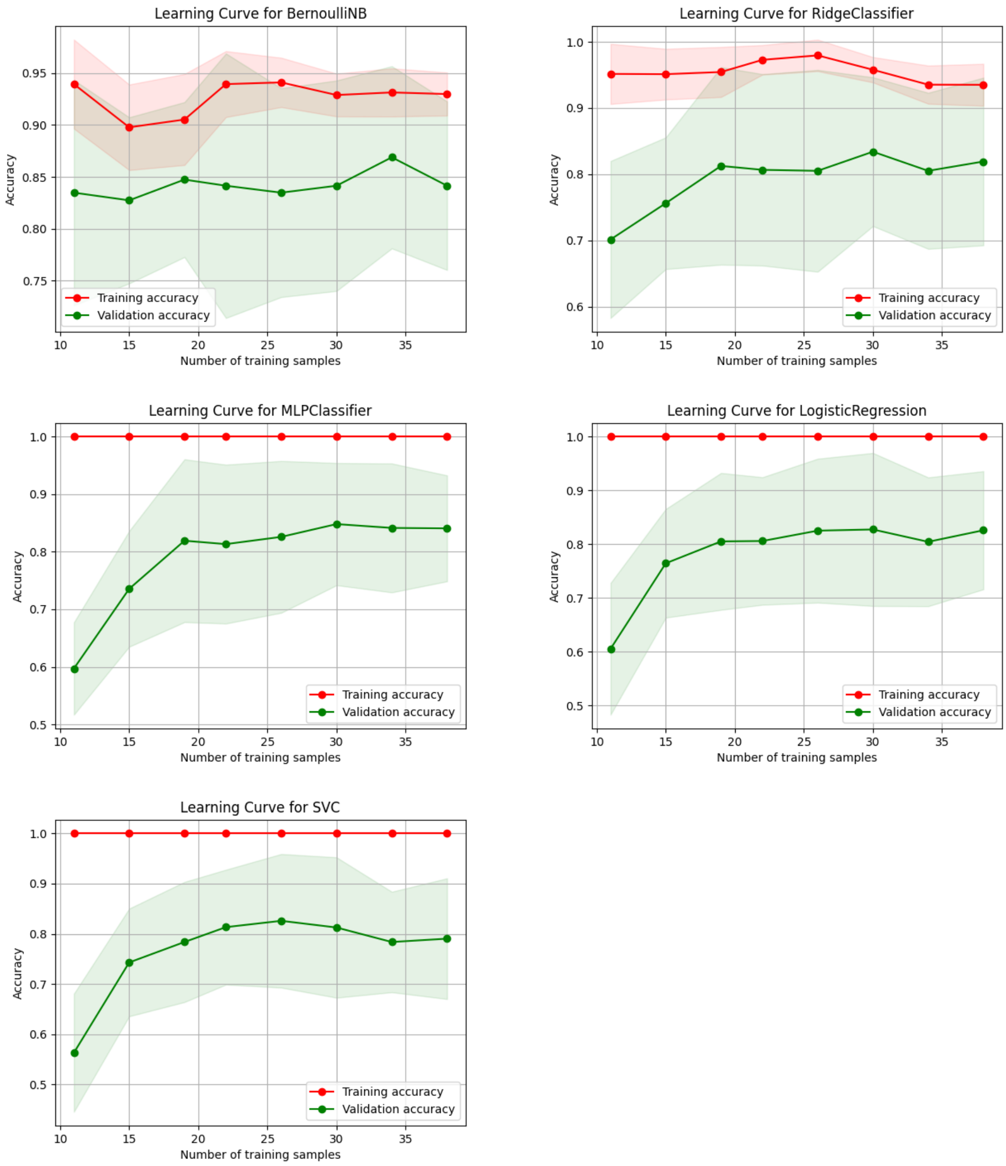
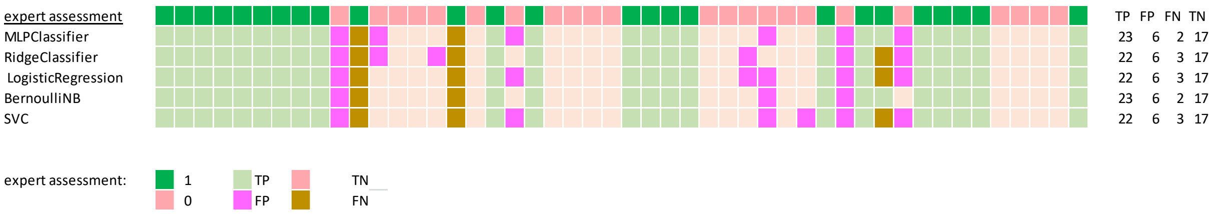


Fig. 2 Learning curves for selected algorithms

The LogisticRegression, MLPClassifier, and SVC algorithms, where the training accuracy is 1 from the initial training (from the smallest sample sizes), fit the training data strongly, but the validation results are lower, indicating poorer generalization of the trained model to new data. In contrast, the RidgeClassifier and BernoulliNB algorithms do not fit the test data closely, and the validation accuracy increases, indicating greater resistance to model overfitting. For the algorithms selected, the prediction results for the training set were tested using cross-validation (Table 3) to verify the model's performance (prediction stability) on specific data. Therefore, those algorithms are recommended for operating on a described set of data.

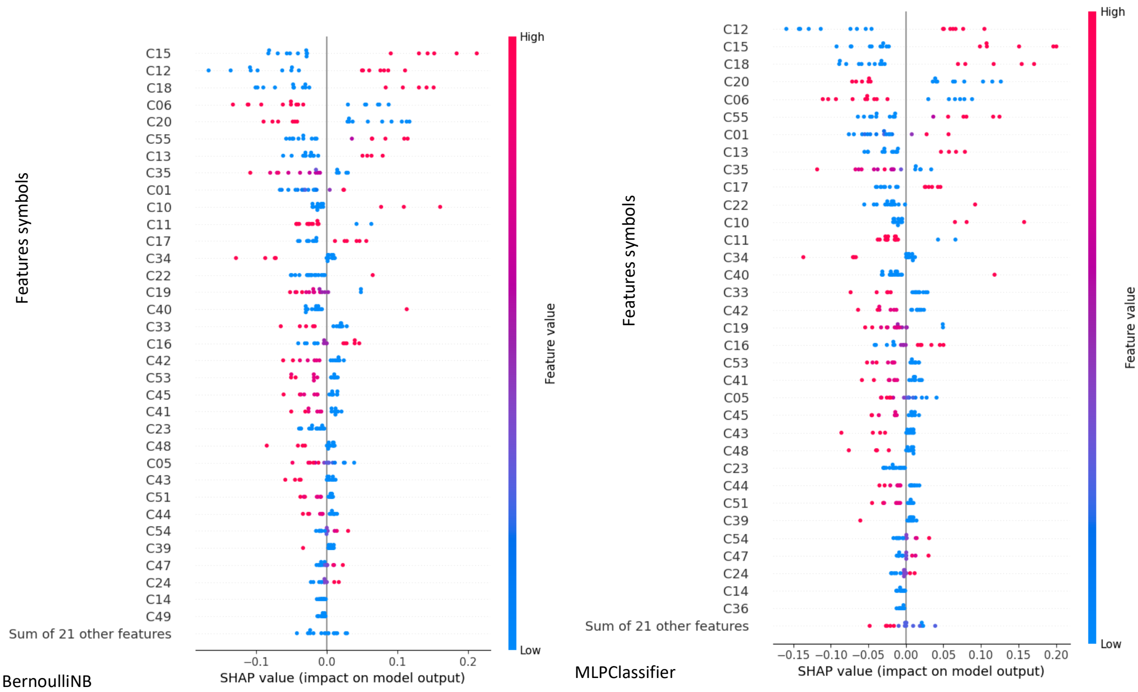
Table 3. Prediction results for the selected algorithms on the training dataset using cross-validation

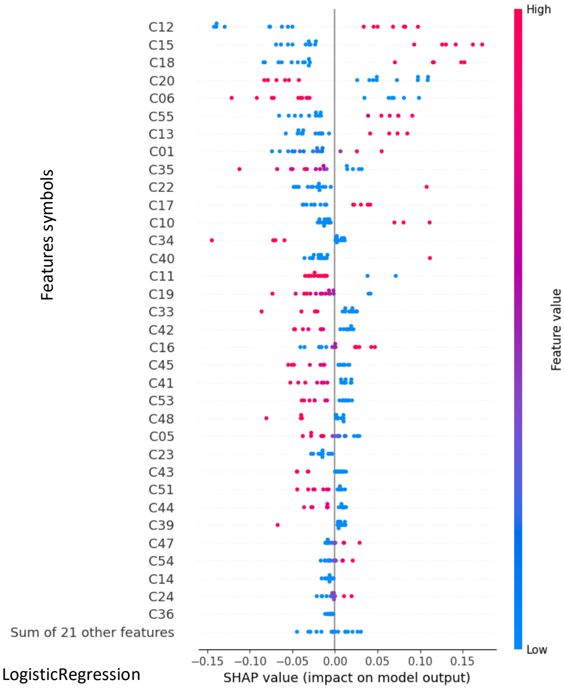


Source: this study
TP – True Positives, FP – False Positives, FN – False Negatives, TN – True Negatives

Each of the five algorithms tested correctly classified 17 heaps as having no impact on the environment (TN) and incorrectly classified 6 heaps as having an impact on the environment (FN) (out of 23 marked as having no impact in the training set). Similarly, 22 or 23 heaps were correctly identified as having an impact (TP), while 3 or 2 were incorrectly identified as having no impact (FP) (out of 25). Therefore, the accuracy for the entire set is 81.25-83.3%, 88-92% for objects marked as having an impact, and 73.9% for objects marked as having no impact (values differs from those stated in Table 1, as these are an effect of cross-validation without repeats, while the latter are effects of cross-validation with repeats).
As a result of the test prediction with cross-validation, 9 objects were classified inversely to the expert assessment by all or most of the algorithms tested. This may indicate the absence of significant features in the set, incorrect values assigned to these features, or an erroneous expert assessment.

Feature Analysis. Features that significantly influence classification results were also identified. Shapley values were used for this purpose. They represent the contribution of individual features to the final value predicted by the model by analysing possible feature combinations. The advantage of this method is that the model analysed is treated as a "black box," meaning that the analysis can be performed for various machine learning models regardless of whether and what information about feature importance is returned by the model. The disadvantage for larger datasets is the high computational complexity. Shapley values can be visualised in many ways helping analysis of feature importance, for instance in form of so-called beeswarm plots shown in Figure 3. More detailed plots are provided in Figure 4 (bar plots) and Figure 5 (waterfall plots).





LogisticRegression

Red indicates high feature values, blue indicates low feature values, and the distance of a point from the centreline indicates the feature's influence on the classification of a single object.

Description of features:

Feature symbol	Feature name
C01	Area [ha]
C02	Waste quantity [thousand Mg]
C03	Beginning of operation
C04	End of operation
C05	Deposition period
C06	Reclaimed facility
C07	Facility height - highest point
C08	Facility height - lowest point at the foot
C09	Height
C10	Occurrence of a major groundwater reservoir
C11	Risk of a threat to a groundwater body
C12	Location of the facility in river valleys
C13	Occurrence of inland water bodies in the vicinity of the facility
C14	Reliable areas in the vicinity of the facility
C15	Thermal condition of the facility
C16	Vegetation coverage of the facility
C17	Facility in operation
C18	Location of the facility on soils of quality classes I-III
C19	Geological structure of the subsoil on which the facility was located
C20	Geohazards - Mining areas
C21	Geohazards - Terrain deformation according to InMoTep
C22	Geohazards - Areas at risk of landslides
C23	Geohazards - Facility location in areas at risk of vertical movements
C24	Wind rose Dominant wind direction SW
C25	Wind rose - dominant wind direction W
C26	Wind rose - dominant wind direction WNW
C27	Wind rose - dominant wind direction NW
C28	Wind rose - dominant wind direction N
C29	Wind rose - dominant wind direction ESE
C30	Wind rose - dominant wind direction SE
C31	Wind rose - dominant wind direction SSW
C32	Occurrence of meadows in the area of the site
C33	Geological structure of the bedrock - valley floors
C34	Geological structure of the bedrock - valley floors and boulder clays
C35	Geological structure of the bedrock - boulder clays
C36	Geological structure of the bedrock - loess, sands;
C37	Geological structure of the bedrock - lacustrine-glacial and clays
C38	Geological structure of the bedrock - in places with sands and gravels
C39	Geological structure of the bedrock - Valley floor muds
C40	Geological structure of the bedrock - Sands and sands with gravels; Sandstones, conglomeratic sandstones
C41	Geological structure of the bedrock - Muds, sands, and river gravels of floodplain terraces
C42	Geological structure of the bedrock - rivers
C43	Geological structure of the bedrock - rivers and valley floors
C44	Geological structure of the bedrock - floodplain/supra-floodplain terraces
C45	Geological structure of the bedrock - floodplain terraces
C46	Geological structure of the bedrock - hard coal
C47	Geological structure of the bedrock - fluvial
C48	Geological structure of the bedrock - Sands and gravels [fluvial glacial; muds], sands, and gravels
C49	Geological structure of the bedrock - Sands and gravels [fluvial glacial; muds] of valley floors
C50	Geological structure of the bedrock - floodplain
C51	Geological structure of the bedrock - Sands and river gravels of floodplain terraces 0.5-2.5 m above sea level Rivers and boulder clays
C52	Geological structure of the bedrock - Gravels
C53	Geological structure of the bedrock - River sands and gravels
C54	Geological structure of the bedrock - Fluvioglacial gravels
C55	Agricultural and forestry use

Fig. 3. Shapley value graph for a subset of the data. The lack of the RidgeClassifier algorithm is due to incompatibility with the shap library.

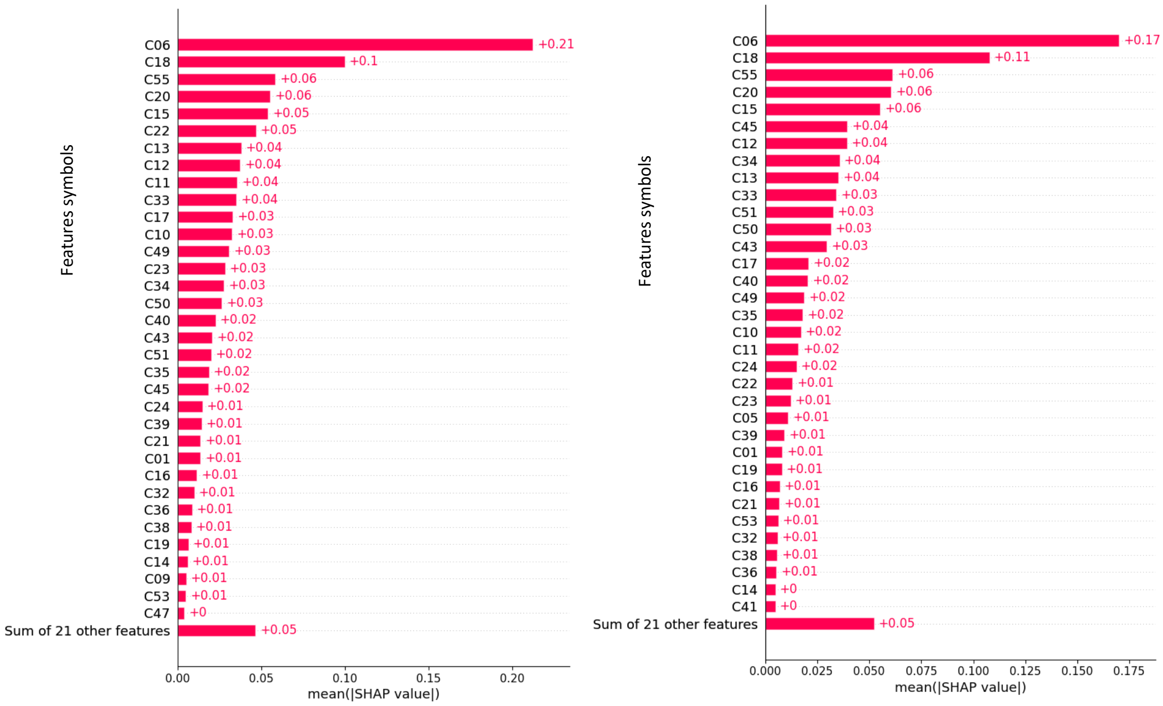


Fig. 4. Bar plots of Shapley values for classification of True Positives (left plot) and True Negatives (right plot) using LogisticRegression

The Shapley values on Figure 4 are means of absolute values of the influence of features. The exact influence of each feature on classification result of specific heap is varying, as shown on Figure 5. Analysing the features that affect the classification of heaps into groups that impact the environment and those that do not impact the environment can contribute to the identification of individual features that have the strongest influence on the classification result and those whose influence is negligible. In the case of sites classified as True Positives (Fig. 4, left plot) one feature has the strongest influence on the classification: C06 – status of heap reclamation, with an absolute value of +0.21. The weaker interaction in the absolute value range has 4 characteristics. These are: C18 - Location of the facility on soils of quality classes I-III (+0.1), C55 - Agricultural and forestry use (+0.06), C20 - Geohazards - Mining areas (+0.06), C15 – Thermal condition of the facility (+0.55) and C22 - Geohazards - Areas at risk of landslides. The remaining features indicate a very weak (14 features with an absolute value ranging from 0.04 to 0.02) to insignificant impact of the facility on the environment (13 features with an absolute value up to 0.01).The 21 remaining features with a total impact of 0.05 have a negligible impact on the assessment of the facilities.

In the case of classifying an heap into a group of disposal sites that do not have an impact on the environment, one feature has the strongest impact on the classification: C06 – status of heap reclamation, with an absolute value of +0.17. Weaker interaction in the absolute value range has 4 features, similarly as in the case of True Positives. These are: forestry use (+0.06), C20 - Geohazards - Mining areas (+0.06), C15 – Thermal condition of the facility (+0.06). The remaining features indicate a very weak (15 features with absolute values ranging from 0.04 to 0.02) to insignificant impact of the disposal site on the environment (14 features with absolute values up to 0.01), and they differ in the hierarchy of occurrence in relation to the group of features classified as True Positives. Twenty-one features have a negligible impact on the assessment of objects (totalling 0.05).

Analysing the absolute values of attributes in the environmental impact of facilities, it was observed that the attribute related to the facility's status as unreclaimed (in the group of disposal sites classified as True Positives) and reclaimed (in the group of disposal sites classified as True Negatives) is the leading attribute. In practice, when considering the environmental impact of dumps, this feature is also important because it directly affects the quality of surface and groundwater in the area of the waste dump. The site's location near agricultural and forest lands also has a direct impact on the environment. A waste dump fire also negatively impacts air quality, soil quality, water quality and vegetation. The presence of active mining areas can affect ground stability, cause mass movements of waste, and increase the impact on surface and groundwater.

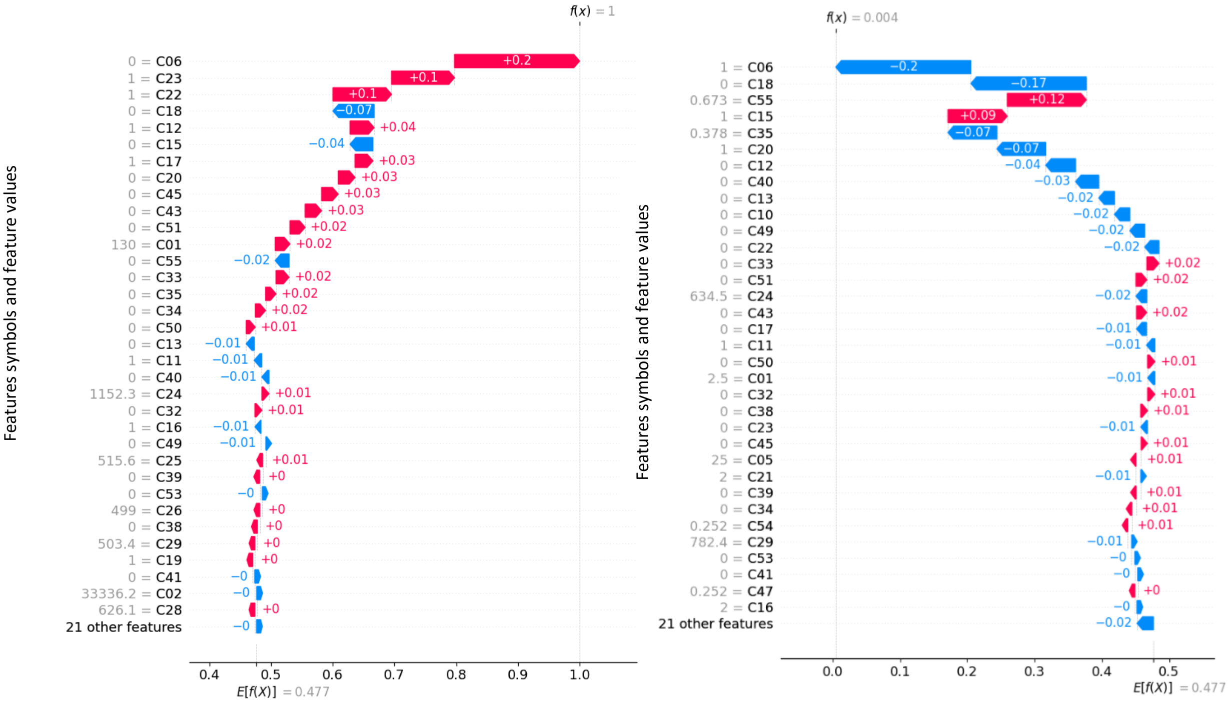


Fig. 5 Waterfall plots of Shapley values for classification of one sample of True Positive classification (left plot) and one sample of True Negative classification (right plot) using LogisticRegression

Analysing the occurrence of feature values in the process of the disposal site's impact on the environment, it was observed that the feature relating to the dump's status as unreclaimed (in the group of disposal sites classified as True Positives) is the leading feature. Other significant features include geohazards (surface flows and landslides) and the site's location in river valleys, as well as the presence of catchment terraces in the disposal site's subsoil and on its surface. All of the features indicated influence the heap's impact on the soil and water environment. In the case of the analysis of the disposal sites classified as True Negatives, the leading features also include the disposal sites' status as reclaimed, meaning it has no environmental impact (through, among other things, the elimination of the impact of atmospheric precipitation, minimizing dust on the surrounding environment and the absence of fires), as well as the presence of impermeable layers in the subsoil. The presence of these factors contributes to the dump's lack of environmental impact.

Feature importance analysis (Fig. 6) was also performed using the Permutation Importances method. This function changes the values of individual features in the input data and examines the impact of these changes on the resulting prediction results relative to those obtained using unmodified data (Scikit-learn, 2025).

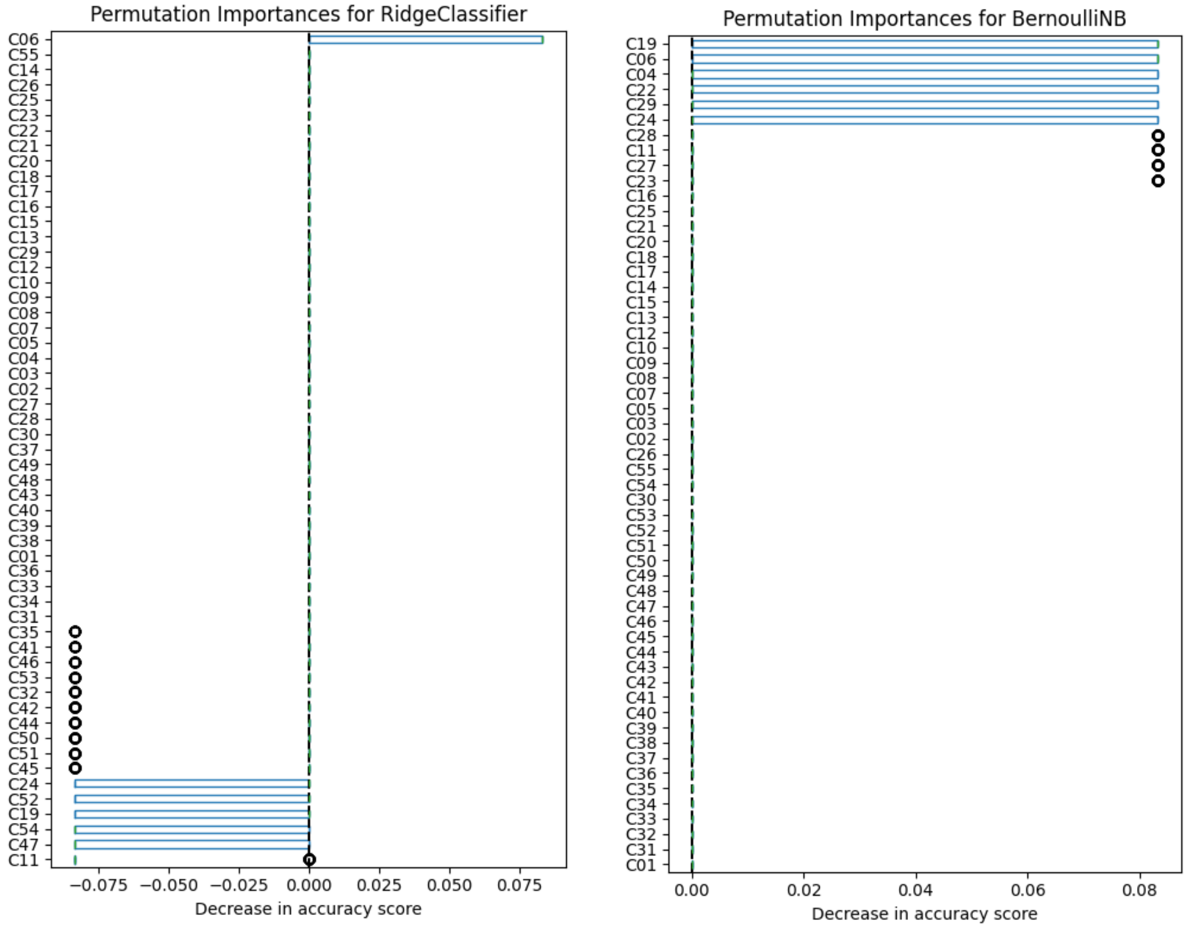


Fig. 6 Graphical representation of feature importance analysis for individual algorithms by means of Permutation importance algorithms. The description of features is shown in Figure 3

Permutation importances analysis shows the sensitivity of algorithms under consideration for changes in values of specific features. The features with negative values represents noise or overfitting. Combining results from Shapley values analysis and permutation importances allows further removal of unimportant features from the dataset.

Conclusions

Over the centuries, mining activities have contributed to the generation of waste, which was stored in waste dumps located near mining plants. The location of these facilities directly in the environment (without the safeguards required by current legislation) contributed to their negative impact on groundwater, surface water, air, soil and vegetation. Assessing the environmental impact of historical waste dumps is a complex, lengthy and capital-intensive process, as each facility is analysed individually according to a developed research plan. The use of machine learning algorithms can be a tool to streamline the process of assessing waste dumps.

The possibilities described in this article for applying machine learning algorithms to assess the environmental impact of coal mining waste dumps, using features characterizing both the object and its broader surroundings, demonstrated that this tool meets its intended objectives. Training conducted on the training set indicated that it is possible to achieve prediction accuracy of around 80%. The existence of objects simultaneously classified by multiple algorithms inconsistently with the training label may indicate the absence of a significant feature in the set or incorrect expert assessment. It should be noted that the training set is relatively small, and machine learning algorithms may show overfitting, i.e., excessive adaptation to the input data, resulting in worst prediction results for new data not seen during training. It is advisable to acquire data for new objects, if possible, characterizing new features other than those previously available, like changing the properties of the deposited material due to weathering, such as the impact of heavy rainfall, and have the results obtained verified by independent experts. The research indicates that the best algorithm for determining the environmental impact of spoil heaps would be the BernoulliNB algorithm, followed by the RidgeClassifier (both achieving up to 87% prediction accuracy), with the currently available training dataset. Its potential extension could improve the results of the MLPClassifier, SVM, and LogisticRegression algorithms.

Acknowledgements.
The authors thank the reviewers for their valuable comments and suggestions, which enriched the substantive content of the manuscript. The study was financed from funds allocated to the statutory activities of PGI-NRI (Project No. 61.4416.2400.00.0).

References:
Baza Hałdy, 2025. https://geologia.pgi.gov.pl/haldy/.
Baza OPPI TPP 2.0, 2025. geoportal.orsip.pl .
Bank Danych o Lasach, 2025. https://www.bdl.lasy.gov.pl/portal/mapy
CBDG, 2025. Centralna Baza Danych Geologicznych. https://baza.pgi.gov.pl/
Chand, K., Kumar, V., Raj, P., Sharma, N., Kumar Mankar, A., Koner, R., 2025. Artificial Intelligence Tool for Prediction of Mine Tailings Dam Slope Stability. Journal of Mining and Environment, **16**: 127-142. https://doi.org/10.22044/jme.2024.14602.2754
Chrzan A., Mojza K., 2018. Preliminary assessment of the heavy metal content on the post-mining waste dump in Czerwionka-Leszczyny (in Polish with English summary). Proceedings of ECOpole, https://doi.org/10.2429/proc.2018.12(2)044.
Chudy, K., Marszałek, H., 2010. Changes in the concentration of sulphates and iron in the vertical section of mine waste pile in Ludwikowice Kłodzkie (the middle Sudetes) (in Polish with English summary). Biuletyn Państwowego Instytutu Geologicznego, **440**: 49-54.
Chudy, K., Marszałek, H., Kierczak, J., 2014. Impact of hard-coal waste dump on water quality — A case study of Ludwikowice Kłodzkie (Nowa Ruda Coalfield, SW Poland). Journal of Geochemical Exploration, **146**: 127–135. https://doi.org/10.1016/j.gexplo.2014.08.011
Czajkowska, A., Gawor, Ł., Cieślok, P., 2018. The impact of the "Smolnica" mining waste disposal site in Trachy on the quality of surface and underground waters (in Polish with English summary). Quartely of Environmental Engineering and Design, **170:** 61-77. https://doi.org/10.5604/01.3001.0012.7463.
DTM, 2025. Digital Terrain Model. https://www.geoportal.gov.pl/
Environment 2025, 2025. Statistics Poland, Warsaw. https://stat.gov.pl/obszary-tematyczne/srodowisko-energia/srodowisko/ochrona-srodowiska-2025,1,26.html
[Accessed:15.12.2025 r.].
Fabiańska, M., Ciesielczuk, J., Nadudvari, A., Misz-Kennan, M., Kowalski, A., Kruszewski, Ł., 2019. Environmental influence of gaseous emissions from selfheating coal waste dumps in Silesia, Poland. Environmental Geochemistry and Health, **41:** 575–601. https://doi.org/10.1007/s10653-018-0153-5.
Fajfer, J., Kostrz-Sikora, P., 2022. Initial cost effectiveness of extracting waste accumulated on the rehabilitation of waste facilities and the concept of circular economy (in Polish with English summary). Przegląd Geologiczny, **70**: 190-201.
Fajfer, J., Krieger, W., Rolka, M., 2013. Closed and abandoned extractive waste facilities – methodology of inventory and database structure (in Polish with English summary). Zeszyty Naukowe IGSMiE PAN, **85**: 23-27. https://min-pan.krakow.pl/wp-content/uploads/sites/4/2017/12/02-11-fajer-krieger-rolka.pdf
Fajfer, J., Rolka, M., Kostrz-Sikora, P., 2025. Assessing the potential of secondary raw materials from hard coal and iron ore mining waste disposal sites using machine learning. Geological Quarterly, **69,** 14. https://doi.org/10.7306/gq.1787.
Foltyn, S., Bogda, A., Szopka, K., Karczewska, A., 2011. Properties of anthropogenic soils on a mine spoil "Kościelniok" in Pawłowice (Hard Coal Mine Pniówek) (in Polish with English summary). Roczniki Gleboznawcze, **62** (2): 79-85.
GDOŚ, 2025. Centralny rejestr form ochrony przyrody. https://crfop.gdos.gov.pl
Gerassis, S., Giráldez, E., Pazo-Rodríguez, M., Saavedra, Á., Taboada, J., 2021. AI Approaches to Environmental Impact Assessments (EIAs) in the Mining and Metals Sector Using AutoML and Bayesian Modeling. Applied Sciences, **11**, 7914. https://doi.org/10.3390/app11177914

Géron, A., 2018. Uczenie maszynowe z użyciem Scikit-Learn i TensorFlow: Pojęcia, techniki i narzędzia służące do tworzenia inteligentnych systemów (in Polish). Helion S.A. Publishing House, Gliwice.
Glubiak-Witwicka, E., Wdziekońska, D., Plewnia, B., Szczygieł, A. (eds.), 2012. Ocena stanu środowiska w rejonie obiektów objętych monitoringiem lokalnym, na terenie województwa śląskiego (in Polish). Biblioteka Monitoringu Środowiska, Katowice. Wyd. REMI-B, Bielsko-Biała.
Haupt, S.E., Gagne, D.J., Hsieh, W.W., Krasnopolsky, V., McGovern, A., Marzban, C., Moninger, W., Lakshmanan, V., Tissot, P., Williams, J.K., 2022. The History and Practice of AI in the Environmental Sciences. American Meteorological Society, **103**: 1351-1370. https://doi.org/10.1175/BAMS-D-20-0234.1.
InMoTeP, 2025. Interferometryczny Monitoring Powierzchni Terenu Polski. https://www.pgi.gov.pl/monitoring-osiadan.html
Jaros, J., 1975. An outline of the history of coal mining (in Poilsh with English summary). Śląski Instytut Naukowy w Katowicach. PWN Publishing House. Warszawa-Kraków.
Jureczka J., Aust J., Buła Z., Dopita M., Zdanowski A., 1995. Geological Map of the Upper Silesian Coal Basin (Carboniferous Subcrop), 1:200000. Państw. Inst. Geol.
Jureczka, J., Dopita, M., Gałka, M., Krieger, W., Kwarciński, J., Martinec, P., 2005. Geological Atlas of Coal Deposits of the Polish and Czech Parts of the Upper Silesian Coal Basin. Państw. Inst. Geol., Ministerstwo Środowiska.
Jureczka, J., Nowak, G.J., 2016. A short overview of data on geological investigation of the Polish bituminous coal basins (in Polish with English summary). Przegląd Geologiczny, **64:** 617-630.
Karty JCWPd, 2025. Karty charakterystyk Jednolitych Części Wód Podziemnych (in Polish). http://karty.apgw.gov.pl:4200/jcw-podziemne
Karty JCWP, 2025. Karty charakterystyk Jednolitych Części Wód Powierzchniowych (in Polish). http://karty.apgw.gov.pl:4200/jcw-powierzchniowe
Konior, J., 2006. The possibilities of limitation of unfavourable influence of mining dumping ground on the surrounding environment (in Polish with English summary). Zeszyty Naukowe Politechniki Śląskiej, Seria: Górnictwo, **271**: 71-82.
Kostrz-Sikora, P., Bliźniuk, A., Fajfer, J., Rolka, M., 2013. Inventory of closed and abandoned mining waste facilities (in Polish with English summary). Zeszyty Naukowe IGSMiE PAN, **85**:199–205. https://min-pan.krakow.pl/wp-content/uploads/sites/4/2017/12/15-12-kostrz-blizniuk-fajfer-.pdf
Kotas, A., 1995. Lithostratigraphy and sedimentologic – paleogeographic development. Moravian-Silesian-Cracovian region. Upper Silesian Coal Basin. Prace Państwowego Instytutu Geologicznego, **143:** 124–134.
Kuźniar, K., 2016. Artificial neural networks in earthquake engineering (in Polish with English summary). Annales Universitatis Paedagogicae Cracoviensis. Studia Technica, **9**: 109-118.. https://rep.up.krakow.pl/xmlui/bitstream/handle/11716/12928/AF204--12--Kuzniar--Sztuczne-sieci-neuronowe.pdf?sequence=1&isAllowed=y
Łaganowska, N., 2019. The influnce of mining waste landfill „Pochwacie" on the ground-water environment on the basis of monitoring tests (in Polish with English summary). Acta Geographica Silesiana, **13** (2): 75–94. WNoZ UŚ, Sosnowiec. https://ags.wnp.us.edu.pl/download/wydawnictwa/ags/ags_34_7.pdf
Ma, L., Huang, C., Liu, Z.-S., 2021. The Application of Artificial Neural Network to Predicting the Drainage from Waste Rock Storages. [in:] Artificial Neural Networks and Deep Learning - Applications and Perspective. IntechOpen. https://doi.org/10.5772/intechopen.96162.
Meteoblue, 2025. History & climate, https://www.meteoblue.com/en/weather/historyclimate/climatemodelled.
Pasa, L., Angelini, G., Ballarin, M., Fedrizzi, P., Sperduti, A., 2025. Enhancing door-to-door waste collection forecasting through ML. Waste Management, **194:** 36-44. https://doi.org/10.1016/j.wasman.2024.12.044.
Paszcza, H., Krogulski, K., 2007. Hard coal industry impact in environment in 2006 (in Polish with English summary). Zeszyty Naukowe Politechniki Śląskiej, Seria: Górnictwo, **276**: 121-137.
Piątek, Z., 1995. Górnictwo węgla kamiennego na Dolnym Śląsku (1434-1945-1994) (in Polish). Przegląd Górniczy, (1):11–13.
Piekut, A., Krzysztofik, L., Gut, K., 2018. Exposure of Zabrze residents to heavy metals emitted from post-industrial waste heaps (in Polish with English summary). Inżynieria Ekologiczna, **19** (4):30–36. https://doi.org/10.12912/23920629/93487.
Pikoń, K., Bugla, J., 2007. Emission from restored coal dumping grounds (in Polish with English summary). Archives of Waste management and Environmental Protection, **6**: 55-70.
Piotrowska, J., Dąbrowska, D., 2024. Artificial intelligence methods in water systems research – a literature review. Geological Quarterly, **68**, 19. https://doi.org/10.7306/gq.1747

Polish Mining Waste Act, 2008. The Act of July 2008, 10 on Mining Waste, Journal of Law 2022, item 2336 as amended (in Polish).
Polish Waste Act, 2012. The Act of December 2012,14 on Waste, Journal of Law 2023, item 1587 as amended (in Polish).
Regulation of the Minister of the Environment of May 2014, 29 on monitoring mining waste disposal facilities, Journal of Laws of 2014, item 875 (in Polish).
Rostański, A., 2006. Spontaniczne kształtowanie się pokrywy roślinnej na zwałowiskach po górnictwie węgla kamiennego na Górnym Śląsku (in Polish). Prace Naukowe Uniwersytetu Śląskiego, **2419**.
Różański, Z., 2019. Management of mining waste and the areas of its storage – environmental aspects. Mineral Resources Management, **35**: 119-142. https://doi.org/10.24425/gsm.2019.128525.
Różkowski, A., 2008. Hydrogeological environment of the Paleozoic formations beneath the Productive Carboniferous in the Upper Silesian Foredeep (in Polish with English summary). Przegląd Geologiczny, **56:** 490-494.
Różkowski, A., Różkowski, K., 2011. Impact of coal mining activity on ground and surface waters environment in the Upper Silesian Coal Basin in the multiyear period (in Polish with English summary). Biuletyn Państwowego Instytutu Geologicznego, **445:** 583-592.
Różkowski, A., Różkowski, K., Sołtysiak, M., 2013. Factors controlling the hydrogeological environment of the quaternary aquifer in the Upper Silesian Coal Basin (in Polish with English summary). Biuletyn Państwowego Instytutu Geologicznego, **445:** 513-518.
Rusin, M., Ćwieląg-Drabek, M., Dziubanek, G., Osmala, W., 2018. Secondary emission from heaps and post-industrial areas as the important source of exposure of Upper Silesia inhabitants to heavy metals. Environmental Medicine, **21** (2): 15-21. https://doi.org/10.19243/2018202.
Sarker, I.H., 2021. Machine Learning: Algorithms, Real-World Applications and Research Directions. SN Computer Science, **2**, 160. https://doi.org/10.1007/s42979-021-00592-x.
Scikit-learn, 2025. Scikit-learn: machine learning in Python - scikit-learn 1.4.1 documentation https://scikit-learn.org/stable/ (accessed 10.10.2025)
Sroga, C., Bobiński, W., Mikulski, S.Z., Adamski, M., Duliban, I., 2017. Mineralne surowce odpadowe na hałdach dawnego górnictwa i przetwórstwa kopalin Sudetów – baza danych wraz z mapami geochemicznymi wybranych rejonów w skali 1:10000 (in Polish). CAG, Warszawa, Nr Arch. 1551/2018
Stefaniak, S., Twardowska, I., 2009. Alteration of ground- and surface water quality resulted from the contact of infiltration and flood waters with the embankment made of re-disposed coal mining wastes (in Polish with English summary). Biuletyn Państwowego Instytutu Geologicznego, **436:** 483-487.
Stefaniak, S., Miszczyk, E., Kmiecik, E., Szczepańska-Plewa, J., Twardowska, I., 2013. interaction of coal mining wastes and powerplant coal ash and its effect on the pore solution chemistry in a disposal site (in Polish with English summary). Biuletyn Państwowego Instytutu Geologicznego, **456:** 555-562.
Styrol, D., 2020. Environmental impact of mine dumps. Example of mine dump „Marcel" in Radlin (in Polish with English summary). Systemy Wspomagania w Inżynierii Produkcji, **9** (2): 48-59.
Sun, T., Li, H., Wu, K., Chen, F., Zhu, Z., Hu, Z., 2020. Data-Driven Predictive Modelling of MineralProspectivity Using Machine Learning and Deep Learning Methods: A Case Study from Southern Jiangxi Province, China. Minerals, **10**, 102. https://doi.org/10.3390/min10020102
Szczepańska, J., 1987. Coal mine spoil tips as a source of the natural water environment pollution (in Polish with English summary). Zeszyty Naukowe AGH, **1135**, Geologia, **35**.
Szczepańska, J., Twardowska, I., 1999. Distribution and environmental impact of coal-mining wastes in Upper Silesia, Poland. Environmental Geology, **38**: 249-258. https://doi.org/10.1007/s002540050422
Szczepańska-Plewa, J., Stefaniak, S., Twardowska, I., 2010. Coal mining waste management and its impact on the groundwater chemical status exemplified in the Upper Silesia Coal Bassin (Poland). Biuletyn Państwowego Instytutu Geologicznego, **441:** 157-166.
Świtała-Trybek, D., Świtała-Mastalerz, J,. 2018. Post-industrial waste heaps – their cultural and tourist potential (exemplified by selected heaps in the silesian province) (in Polish with English summary). Zeszyty Naukowe, Turystyka i Rekreacja, **22** (2): 125–141. https://wstijo.edu.pl/wp-content/uploads/2023/05/zeszyt_naukowy_tir_nr_22.pdf. [Accessed: 2025-10-15].
Trifi, M., Gasmi, A., Carbone, C., Majzlan, J., Nasri, N., Dermach, M., Charef, A., Elfil, H., 2022. Machine learning-based prediction of toxic metals concentration in an acid mine drainage environment, Northern Tunisia. Environmental Science and Pollution Research, **29:** 87490–87508. https://doi.org/10.1007/s11356-022-21890-8
Twardowska, I., 1981. Mechanizm i dynamika ługowania odpadów karbońskich na zwałowiskach (in Polish). Prace i Studia PAN, **25**.
Uddin, M.G., Nash, S., Rahman, A., Olbert, A., 2023. Assessing optimization techniques for improving water quality model. Journal of Cleaner Production, **38**5, 135671. https://doi.org/10.1016/j.jclepro.2022.135671

Wolkersdorfer, C., Mugova, E., 2022. Human pressures and management of inland waters. Effects of Mining on Surface Water. Encyclopedia of Inland Waters, **2**nd edition, 170-188 https://doi.org/10.1016/B978-0-12-819166-8.00036-0.
Worldclim, 2025. Historical climate data. https://www.worldclim.org/data/worldclim21.html
Wróbel, Ł., Dołhańczuk-Śródka, A., Kłos, A., Wacławek, M., 2012. Gamma radiation in selected mine waste dumps at Upper Silesia. Proceedings of ECOpole, **6:** 799-803. https://doi.org/10.2429/proc.2012.6(2)111.
Xia, W., Jiang , Y., Chen, X., Zhao, R., 2022. Application of machine learning algorithms in municipal solid waste management: A mini review. Waste Management & Research, **40:** 609–624. https://doi.org/10.1177/0734242X211033716.
Zając, E., Zarzycki, J., 2013. The Effect of Thermal Activity of Colliery Waste Heap on Vegetation Development (in Polish with English summary). Annual Set The Environment Protection, **15**: 1862–1880.
Zdechlik, R., Gołębiowski, T., Tomecka-Suchoń, S., Żogała, B., 2011. Application of hydrogeochemical and geophysical methods in assessment of the influence of coal-mining waste dumps on hydrogeological environment (in Polish with English summary). Biuletyn Państwowego Instytutu Geologicznego, **445**: 725-736.
Zhang, C., Ma, L., Liu, W., 2023. A Machine Learning Approach for Prediction of the Quantity of Mine Waste Rock Drainage in Areas with Spring Freshet. Minerals, **13**, 376. https://doi.org/10.3390/min13030376.