Drill bit deterioration estimation with the Random Forest Regressor

Mateusz Góralczyk¹ *, Anna Michalak¹, Paweł Śliwiński²

¹ Faculty of Geoengineering, Mining and Geology, Wrocław University of Science and Technology, Na Grobli 15, 50-421 Wrocław, Poland
² KGHM Polska Miedź S.A., M. Skłodowskiej-Curie 48, 59-301 Lubin, Poland

* Corresponding author: mateusz.goralczyk@pwr.edu.pl

Abstract. Blastholes drilling performance is crucial for ensuring good performance of the whole excavation process, the correctness of which demands ‘healthy’ drill bit and appropriate behavior of an operator. Given the large volume of non-linear parameters describing the process, it appears reasonable to employ supervised learning methods to obtain drilling performance insights. Random Forest Regressor model has been trained on the dataset corresponding to correct performance of blastholes drilling and its hyperparameters have been tuned to obtain the highest possible accuracy. It has been later tested on three datasets corresponding to a good performance of drilling, and two cases of its non-optimal execution. Estimation errors are proposed to be used as bit technical state condition indicators (or more generally - process performance indicators). Root Mean Squared Error has been proven to differ significantly when compared estimation based on datasets corresponding to execution of drilling with ‘healthy’ drill bit, and its execution with worn-off one, however, it has been not sufficient to distinguish non-optimal drilling when additional feed pressure has been exerted by an operator to compensate the reduced pace of drilling. It has been, however, possible when the mean of absolute estimation errors has been used.

1. Introduction
Appropriate execution of blastholes preparation has a direct influence on the profitability of excavation of mineral resources and metal ores. It is especially demanding in the case of underground mining, in which performing the task in accordance to pre-defined blasting mesh, with high efficiency not compromising the accuracy is expected to ensure undisturbed production while maintaining the most favourable geometry of underground excavations. Such a superior performance can be obtained only when using a good quality drilling tool, and employing the correct technique by an operator. To aid operators and allow analysing the efficiency many of (most of the ones produced nowadays) machines are equipped in devices and systems supporting drilling throughout constant gathering the data regarding the performance of functional subsystem (hydraulic drifter and drill itself). Incorporation of on-board data acquisition systems to a machine undoubtedly increases its attractiveness, however in order to make the most of such technological enhancements and utilise the huge stream of data produced due to the employment of on-board monitoring appropriate analysis techniques have to be developed. Ones that are resistant to the non-linearity of interactions of various monitored variables are of major importance. Depending on the particular mine’s size, dozens or even more than a hundred mobile machines can perform work on a daily basis each of them contributing
data logs to be further analysed. Such a scale means that any non-highly-automated method of analysis of their work is an overwhelming task, therefore methods of objective and automatic evaluation are needed. Discovering patterns in the data corresponding to blastholes drilling, that could be informative in the context of estimation of drill’s deterioration level, machine damage, or applying incorrect technique by an operator is a foundation for the development of drill replacement strategies, predictive maintenance-, and operator assistance methods. Similar applies to the exploration or exploitation of mineral oil and gas deposits with the use of rotary steerable systems, which are employed in the execution of complex directional drilling. The mechanical systems used to penetrate valuable material bearing formations with changing trajectories are exposed to numerous failure factors [1, 2], thus the development of new methods allowing for determination of the drilling devices wear level in ore, mineral, and oil & gas deposits exploitation is of major importance.

A powerful, ubiquitous tool, which is employed in work with voluminous datasets containing various non-linear relations is machine learning. There are numerous methods to be possibly applied in a form of supervised, unsupervised, deep, and reinforcement learning, which, if applied on correctly preprocessed and interpreted data, may reveal information hidden in the large-scale datasets. Taking into consideration numerous combinations of parameters describing drilling performance and nonlinearity of the data, machine learning regression, and classification appear as legitimate approaches to the evaluation of multidimensional processes. One of such algorithms, namely Random Forest Regressor has been applied to evaluate the possibility of indicating the wear of a bit. Given the reference sets of data corresponding to a ‘healthy’ and damaged or worn-off tool a properly selected and tuned ensemble of decision trees is assumed to indicate occurring deterioration of the drill. Three sets of data corresponding to the drilling of blastholes with the use of: ‘healthy’ and worn-off drill bit have been preprocessed and analysed with the use of Random Forest Regressor and basic statistical measures to assess the possibility of employing this machine learning technique in analysing blastholes drilling performance.

2. State of the art

The widespread concept of using the so-called measurement while drilling (MWD) data involves various algorithms from machine learning [3] through fuzzy systems [4] to artificial neural networks [5]. The aforementioned methods found their usage in explorational geology and petrophysics as well - to enumerate the most common and obvious examples. One of the powerful, yet relatively simple methods used for classification problems is Boosting - a method combining a multitude of classifiers of guess accuracy only slightly higher than random guess (‘weak classifiers’) into iterative learning process resulting in high accuracy classification result. In its original version Boosting algorithm was suited to only two-class problems, however, its extensions like LogitBoost utilizing logistic regression exist, providing solutions to multi-class classification problems [6]. A comparison of multi-layer perceptron neural network, fuzzy inference system, and boosting in the context of distinguishing lithological types has been presented in [7]. The work provides a legitimate example of a method that is correctly adjusted to the problem’s nature, however, the least computationally complex whereas giving the most accurate, and easily interpretable predictions. Outstanding accuracy of ROP prediction has been achieved by the authors of [8], where three advanced intelligent systems based on Extreme Learning Machine (ELM), and multilayer perceptron (MLP) have been presented and compared. Significant improvement of the algorithms employed at the stage of preparation of the learning set for the ELM were two meta-heuristic algorithms - hybrid whale optimisation, and particle swarm optimisation applied. The use of these algorithms allowed to define the optimum weights of input variables, and optimize hidden layer biases. Improved ELM models were compared with the well-known MLP trained with the use of the Levenberg-Marquardt algorithm. Machine learning method has been used as an element of an autonomous percussive drilling architecture,
working principle, and experimental validation results of which have been presented in [9]. To optimise drilling in a layered rock mass with variable geomechanical characteristics, what has been achieved with the use of the so-called Golden Section Search algorithm, beforehand rock classes had to be distinguished. Assigning encountered rock to a particular type and distinguishing it from metal has been achieved with the use of the Linear Support Vector Machine algorithm - an easily implementable and reliable machine learning tool for the problems. The classifier has been experimentally proven to successfully detect transitions from granite to sandstone (in the opposite order as well), and from granite to steel allowing the drilling control system to adjust the weight on bit (WOB) and in turn to control the penetration rate. In the work [10] which is another example of successful implementation of ML technique in explorational drilling, the Random Forest model provided satisfactory results when given input from a large multivariate database with the aim of estimating continuous variable indicative for sulphide ore presence in volcanogenic deposits - sodium depletion. With regard to non-Gaussian distributed, multi-class data with weakly correlated variables the authors made a legitimate choice to select the RF algorithm, which contrary to methods like SVM or multiple regression performed well on non-strictly stationary data. In two separate, independent tests - one on standard lithogeochemistry data, and another on multiparameter data the RF approach has been proven to provide usable estimates of the variable of interest from the point of view of the deposit exploration. A successful approach to predicting the rate of penetration on the basis of multi-parameter input with the use of RF and Monotone Multi-Layer Perceptron has been presented in [11] Feature ranking was performed as a step preceding model training to shorten the calculations, reduce overfitting, and facilitate further interpretation. A reasonable step made by the authors before the feature ranking was performed, was the removal of constant or nearly constant observations from the input dataset. Potential model overfitting influence has been additionally decreased in the machine learning approach thanks to the use of ensembles of decision trees of two types, namely Random Forests and Cubist - [12]. In the ANN approach, in turn, the overfitting influence has been reduced by applying the so-called bagging (bootstrap aggregation). Whereas both of the approaches presented in the abovementioned work gave estimation results at a comparable, relatively good level of accuracy, the significant characteristic of the ML model - the possibility of extracting its general rules used during prediction has been underlined and utilised. In one of the most recent works [13] utilised data acquired real-time during vertical wells drilling containing standard variables describing the process performance such as flow of process water, hook height, hook load, bit depth, standpipe pressure, and differential pressure, rotation in RPM, or weight on bit. What is worth mentioning due to its value for further achievements after subjecting the data to a typical exploratory analysis, the author performed feature engineering, namely feature augmentation which consisted of appending adjacent raw observations of a given feature to the measurement at the current depth. Such an approach was justified by the uncertainty caused by measurements being taken at given intervals not always exactly corresponding to that depths. Some other applications of ANN, fuzzy logic, case-based reasoning, or even hybrids of the mentioned approaches in the petroleum sector, used for drilling problem detection, operational troubleshooting, well integrity assessment, decision making, drilling optimisation, and well planning can be found in [14].

3. Problem formulation
The condition of a drill bit as a function of time is strongly dependent on the mineralogical composition (especially quartz content), compressive strength of the rock mass, and its abrasiveness (the last being the product of the first two parameters mentioned). It may significantly reduce the penetration rate affecting the productivity of a drilling jumbo [15, 16]. In the least favourable cases, the gradual deterioration of a bit may remain unnoticed for a longer period for instance due to the compensation of decreased ROP by increasing feed pressure, which
ensures the same number of holes drilled during a shift - which is crucial from the point of view of an operator. Such under-optimal operation of the machine should be avoided but is currently difficult to notice in an automated way. What is more, in some cases, because of changes in local mining-geological conditions the performance of drilling may be sub-optimal due to the lack of adjustment of the operator’s technique (or the program’s - in the case of autonomous drilling) to the encountered rock-mass parameters. As it can be seen, the assessment of the blast holes drilling process based only on the number of holes drilled in a given time is very superficial, and taking into consideration the human aspect, might be unfair.

Moreover, to adjust the setting of drilling parameters to variable conditions encountered in underground mines, where the rock characteristics can practically never be defined accurately in advance, a flexible and easy-to-use indicator would be beneficial allowing the fleet management and work supervisory crews to react with no need of inspection in a particular place in person. Hence the authors’ proposition, to utilise machine learning techniques as a more versatile and insightful tool for the evaluation of the drilling process. The assumption is as follows:

- Firstly the selection of hyperparameter to a machine learning model is performed, such as to obtain the best possible accuracy of estimation on a dataset corresponding to the correct performance of drilling;
- Next the model is fit to a ‘healthy’ dataset and the estimation errors obtained are set as a benchmark for further comparison of analysed drilling datasets;
- The data corresponding to the drilling performance in real conditions is regularly compared to the benchmark, and at a given level of estimation errors, the process is defined as incorrect/unfavourable indicating the need for further analysis of possible causes.

4. Machine and data acquisition system

4.1. Machine

The machine, which performance has been subjected to the investigation presented in this article was the single boom drilling rig. The standard work performed by this diesel-powered vehicle is the drilling of 3.2 meter-long blast holes of the 41 to 76 mm diameter. The rotary-percussive movement of the jumbo’s rock drill is governed by a hydraulic drifter. When the vehicle is located in the mining face, its diesel engine is turned off and the power from an electric outlet in the mining face’s proximity is supplied through an extendable power line in which the machine is equipped. The electric energy drives the hydraulic pump transmitting the force to the drill, together with the extension arm and feeder which allow manoeuvring and aiming the working device along with the pre-defined blasting mesh.

![Figure 1: The drilling rig subjected to the study (taken from [17]).](image)

4.2. Data acquisition system

All of the mobile machinery in the underground mine operates on a high workload in harsh environmental conditions. To control its technical state on the large scale, and allow for diagnostics all of them have to be equipped with onboard data acquisition systems, compatible with a centralised management platform [18]. The data used in this study has been acquired
from report .csv files generated by the mine’s management crew having access to such a platform developed in the scope of the internal program of KGHM Polska Miedź SA, which is extending the functionality of supervisory control and data acquisition system (SCADA). Currently more than 200 drilling rigs, bolting rigs, trucks, and LHDs are feeding the system with the information from their electronic control units (ECU) accessed through CAN buses (and wirelessly in the case of some additional data). Since providing wireless communication devices in all of the workplaces, including dynamically reshaped mining faces and hundreds of intersecting corridors is extremely difficult, if not impossible, the information acquired during every shift is being stored locally in the machine’s modules and transferred from the wireless connection spots at the end of it to the common surface-located server [19]. The frequency of data sampling is 1 per second (1 Hz) for all of the features used in this study.

The raw data stored in the database server of the ore producer can be further subjected to the analysis in order to generate human-interpretable reports, create a foundation for predictive maintenance, or serve as a machine utilisation evaluation basis. From the machine learning point of view, however, it is the most valuable in its raw state saved to .csv files containing the variables presented in the further section.

5. Preprocessing
The penetration pace itself is a very indicative feature for the blastholes drilling performance evaluation, however, due to possible variations in geomechanical parameters of rock mass in which the drilling is performed all the features related to drill’s operation are considered to give a more universal tool for its evaluation. The pace will be estimated based on the other 3 pressure parameters to obtain the process correctness indicator.

There are only 4 features (presented in the figure 2 below) from the machinery operational data .csv file obtained from the data acquisition system that are related to drilling activity directly, namely:

- LHYDDFEDP [MPa] (‘feed pressure’)
- LHYDDIMPPA [MPa] (‘percussion pressure’)
- LHYDDRPMP [MPa] (‘rotational pressure’).
- Y [m] (drill slide-out distance at a given moment)

![Figure 2: Raw data subjected to further preprocessing.](image-url)
To obtain more informative, and model interpretative variable the selected features have been processed as follows:

- slide-out distances have been first averaged, taking the rolling mean of 3 samples,
- data-set has been cleaned from any rows containing non-numerical values,
- slide out distances have been differentiated in time to obtain the pace of the drilling (or penetration velocity),
- negative values of drilling pace have been replaced with zeros to obtain only actual drilling-related data to be further used in the ML model.

The pre-processed data to be further used as a model fitting set is presented in the figure 3 below. It can be seen that the relation of pressures responsible for the operation of the drill has a repeatable pattern, and (despite one very short-lasting drop in the data corresponding to the penultimate full hole) drilling pace pattern is depicting non-obstructed subsequent protrusions of the drill.

Figure 3: Preprocessed data corresponding to a correct performance of drilling used as the model fitting set.
Three datasets have been used to verify the possibility of performing the distinction between a properly performed drilling operation, and two different examples of its improper execution. The two non-optimal executions of drillings are 1) drilling with a deteriorated drill, hence not obtaining proper penetration pace; 2) performing the drilling with a deteriorated drill and 'compensating' for the decreased pace of the process by exerting additional force on the drill (increased feed pressure) that exposes machine to additional stress and increases the risk of abrupt major destruction. The pattern of drilling pace and hydraulic pressures corresponding to the three testing sets used has been presented in the figure 4.

![Drilling pace and hydraulic pressure data used as testing sets.](image-url)

Figure 4: Drilling pace and hydraulic pressure data used as testing sets.
6. Prediction of drilling pace with Random Forest Regressor

6.1. Random Forest Regressor

Because of the non-Gaussian distribution of the available variables to be used as training data and in order to omit over-fitting of the model the Random Forest Model has been selected. This ensemble ML tool utilises the bagging statistical technique [20] and many decision trees to achieve more accurate prediction by combining multiple predictions of the so-called ‘weak estimators’ (i.e. single decision trees). The algorithm performs a two-stage process whether predicting continuous outputs or making discrete predictions - first, it builds a given number of decision trees (regressors/estimators) and next averages the votes of each of them. Each estimator’s cost function is the Mean Squared Error (MSE), based on which a tree determines which particular value of a given attribute will ensure the best split. Such a feature selection is equivalent to variance reduction. Alternatively, the Mean Absolute Error can be used as an objective function. When the best split value is found the dataset is divided and the root node is created, which is thereafter repeated in other ranges. There are iterations of the aforementioned process performed as long as, either: a) each sample in the dataset has been reached, what happens if no stopping criterion has been established [21]; b) the stopping criterion has been reached. The stopping criteria can be: reaching maximum depth set preliminarily (it means that the tree was allowed to create only a given number of splits from the root node), getting to a certain predefined number of observations in the leaf node, or to a given number of leaf nodes [22].

Since single trees are prone to overfitting, as was mentioned before, especially when a tree’s depth is larger, the randomness allowing to alleviate this issue is being introduced by using ensembles. Two characteristic concepts used by this kind of models that are should be noted are:

- training samples are randomly sampled (drawn with replacement) in the learning phase, what is known as bootstrapping,
- using random feature subsets for node-splitting, which in other words means that when a tree 'makes a decision’ to split a node it bases only on a subset of all features - not all available ones.

6.2. Hyperparameter tuning and prediction

The cross-validation method has been employed in order to receive the best set of hyperparameters adjusted to the dataset corresponding to the correct performance of drilling. To facilitate the process of their selection the scikit-learn free Python machine learning library function has been used, namely GridSearchCV. Negative Mean Squared Error has been selected as the evaluation metric and the number of cross-validations equal to 3. The following hyperparameters have been tuned by trying out all the possible combinations with the corresponding pre-defined ranges of their values:

- max depth = [1, 2, 3, None],
- number of estimators = [2, 3, 5, 10, 50, 100, 200, 300],
- max features = [‘log2’, ‘auto’, ‘sqrt’],
- min samples per leaf = [1, 2, 3, 4, 5, 10],
- min samples per split = [2, 3, 4, 5, 6, 7, 8, 9, 10, 11].

The best parameters for the estimator, which have been obtained are following:

- max depth = 3,
- number of estimators = 5,
- max features = ‘auto’,
- min samples per leaf = 5,
- min samples per split = 2.

The Random Forest Regressor model with the hyperparameters selected as presented above has been used to predict the speed of rock mass penetration by the drill based on three pressures in the hydraulic system related to the blasthole drilling process. The results obtained are presented in the figures below.

![Test 1 - good drilling performance](image1)

(a) 'Healthy' drill bit.

![Test 2 - worn off bit](image2)

(b) Worn-off drill bit.

![Test 3 - worn off bit & increased feed pressure](image3)

(c) Worn-off bit and pressed drill.

Figure 5: Estimation results on three different datasets presenting various drill bit states and behaviours of an operator.

Additionally, the errors’ mean has been calculated based on differences between actual testing sets samples’ values and predictions made by the model. Boxplots presented on the next figure, serve for a better depiction of estimation errors. In each box, the central horizontal line depicts the mean of the prediction errors. The top and bottom edges of the boxes correspond to respectively the 75th and 25th percentile of error values. Top horizontal short lines visible at the ending of vertical ones coming out of the boxes - whiskers set a limit for the values, which are not considered as outliers. The outliers (diamonds above the whiskers) are the values exceeding $1.5 \times IQR$, where IQR is the interquartile range. The white star in the middle of the boxplot depicts the median of estimation errors.
Two types of estimation errors obtained on three different datasets are presented in the table 1 below. The root means squared error (RMSE) for the testing set corresponding to the worn-off component when no additional feed pressure has been used by the operator is near twice as big as the one related to the optimal state of the bit and correct execution of drilling. Unfortunately, when additional feed pressure is exerted this measure appears as not sufficient since the difference between the two estimations is only 0.00787. Mean of absolute estimation errors, nonetheless, seems to be more useful in such case, since it increases more than twice when comparing testing on 'healthy' bit-corresponding dataset, and worn-off, pressed bit - dataset.

Table 1: Estimation errors for 3 testing datasets.

| Test set | RMSE   | Mean of absolute estimation errors |
|----------|--------|-----------------------------------|
| Test set 1 ('healthy') | 0.02180 | 0.01010                           |
| Test set 2 (worn-off)   | 0.4408  | 0.04190                           |
| Test set 3 (worn-off and pressed) | 0.02967 | 0.02760                           |

7. Summary and conclusions
A proposition of drill-bit deterioration/drilling performance correctness indicating measures has been presented. Taking into consideration that the data from machine monitoring systems has a large volume, and contains numerous non-linearly interacting features a machine learning-based approach has been selected to create a proposition of component health/process performance correctness estimation method. Overfitting-resistant, Random Forest Regressor, together with basic statistical measures of its accuracy have been employed and tested on three different datasets.
The estimation error expressed as root mean squared error (RMSE) and mean of absolute error values, due to their significant difference when compared in three different cases has the potential to be used as an indicator of worsening performance, thus the health of the drill bit. Even though the RMSE values have been similar when compared testing set corresponding to ‘healthy’ bit and worn-off one with additional feed pressure exerted on it by the drilling rig’s operator, the second measure - the mean of absolute error values have been more than twice as big for non-optimal drilling performance. With such a big difference, it seems possible to establish a reference maximum error value, which could be used as an indicator of deterioration or deviation from the optimal performance caused by other factors.

The presented methodology bases on the estimation of single point-in-time process descriptive values, not blastholes drilled as a whole, which allows keeping insight into abrupt abnormalities lasting shorter than a single blasthole drilling. Nonetheless, such an approach demands further segmentation of blastholes and aggregation of estimation results to obtain directly interpretive information regarding the correctness of the drilling performance.

Acknowledgement
This research has received funding from European Institute of Innovation and Technology (EIT), a body of the European Union, under the Horizon 2020, the EU Framework Programme for Research and Innovation. This work is supported by EIT RawMaterials GmbH under framework partnership agreement number 19036 (SAFEME4MINE. Preventive Maintenance system on safety devices of Mining Machinery).

References
[1] Zhironkin V, Epikhin A, Novoseltsev D and Cehlár M 2020 Acta Montanistica Slovaca 25 393–405 URL https://doi.org/10.46544/ams.v25i13.11
[2] Epikhin A, Zhironkin V, Szurgacz D and Trzop K 2021 IOP Conference Series: Earth and Environmental Science 684 012001 URL https://doi.org/10.1088/1755-1315/684/1/012001
[3] Vezhapparambu V S, Eidsvik J and Ellefmo S L 2018 Minerals 8 ISSN 2075-163X URL https://www.mdpi.com/2075-163X/8/9/384
[4] Basarir H, Wesseloo J, Karrech A, Pasternak E and Dyskin A 2017 Proceedings of the Eighth International Conference on Deep and High Stress Mining 537–551
[5] Alkinani H H, Al-Hameedi A T T, Dunn-Norman S and Lian D 2020 Egyptian Journal of Petroleum 29 121–126 ISSN 1110-0621 URL https://www.sciencedirect.com/science/article/pii/S1110062119303174
[6] Friedman J, Hastie T and Tibshirani R 2000 The Annals of Statistics 28 337 – 407 URL https://doi.org/10.1214/aos/1012144092
[7] Kadkhodae-Ilkhchi A, Monteiro S T, Ramos F and Hatherly P 2010 IEEE Geoscience and Remote Sensing Letters 7 680–684
[8] Youcef M R, Hadjadj A, Bentriou A and Boukredera F S 2020 Earth Science Informatics 13 1351–1368 URL https://doi.org/10.1007/s12145-020-00524-y
[9] Mazumdar A, Su J C, Spencer S J and Buenger S P 2019 American Control Conference (ACC) URL https://doi.org/10.23919/acc.2019.8815280
[10] Schnitzler N, Ross P S and Gloaguen E 2019 Journal of Geochemical Exploration 205 106344 ISSN 0375-6742 URL https://www.sciencedirect.com/science/article/pii/S0375674218304047
[11] Eskandarian S, Bahrami P and Kazemi P 2017 Journal of Petroleum Science and Engineering 156 605–615 URL https://doi.org/10.1016/j.petrol.2017.06.039
[12] Quinlan J R 1992 Proceedings of Australian Joint Conference on Artificial Intelligence 343–348

[13] Chandrasekaran S and Kumar G S 2020 International Journal of Mathematical, Engineering and Management Sciences 5 381–394 URL https://doi.org/10.33889/ijmems.2020.5.3.032

[14] Bello O, Teodoriu C, Yaqoob T, Oppelt J, Holzmann J and Obinwanne A 2016 SPE Nigeria Annual International Conference and Exhibition

[15] Plinninger R and Ralf J 2008 Geomechanics and Tunnelling 1 38–46

[16] Rais K, Kara M, Hr Hadji, Riheb , Gadri L and Khochmen L 2017 Mining Science URL http://www.miningscience.pwr.edu.pl/original-approach-for-the-drilling-process-op-timization

[17] Wodecki J, Góralczyk M, Krot P, Ziętek B, Szrek J, Worsa-Kozak M, Zimroz R, Śliwiński P and Czajkowski A 2020 Energies 13 ISSN 1996-1073 URL https://www.mdpi.com/1996-1073/13/24/6748

[18] Wodecki J, Stefaniak P, Michalak A, Wyłomańska A and Zimroz R 2017 International Journal of Mining, Reclamation and Environment 32 392–400 URL https://doi.org/10.1080/17480930.2017.1388336

[19] Krot P, Sliwinski P, Zimroz R and Gomolla N 2020 Measurement 151 107111 URL https://doi.org/10.1016/j.measurement.2019.107111

[20] Breiman L 1996 Machine Learning 24 123–140 URL https://doi.org/10.1007/bf00058655

[21] Aggarwal C C 2015 Data mining: the textbook (Springer)

[22] Lipinski P, Brzychczy E and Zimroz R 2020 Sensors 20 5979 URL https://doi.org/10.3390/s20215979