Istraživanja i projektovanja za privredu
ISSN 1451-4117
DOI:10.5937/jaes0-38729
www.engineeringscience.rs

Journal of Applied Engineering Science
Vol. 21, No. 1, 2023
Original Scientific Paper
Paper number: 21(2023)1, 1052, 76-86

APPLICATION OF MACHINE LEARNING METHODS FOR PREDICTING WELL DISTURBANCES

Ekaterina Kusimova[1]*, Liliyan Saychenko[1], Nelli Islamova[2], Pavel Drofa[2], Elena Safiullina[1], Alexey Dengaev[3]

[1] *Saint Petersburg Mining University; 2, 21st Line, St Petersburg*
[2] *Gazpromneft STC*
[3] *National University of Oil and Gas «Gubkin University»; 65 Leninsky Prospekt, Moscow*
** katerinakusimova@gmail.com*

In the process of field exploration, along with regular flooding, a significant part of the wells is flooded prematurely due to leakage of the string and outer annulus. In an effort to intensify the flow of oil to the bottom of wells in field conditions, specialists often try to solve this problem by using various technologies that change the reservoir characteristics of the formation. Any increase in pressure that exceeds the strength of the rocks in compression or tension leads to rock deformation (destruction of the cement stone, creation of new cracks). Moreover, repeated operations under pressure, as a rule, lead to an increase in water cut and the appearance of behind-the-casing circulations. For that reason, an important condition for maintaining their efficient operation is the timely forecasting of such negative phenomena as behind-casing cross flow and casing leakage. The purpose of the work is to increase the efficiency of well interventions and workover operations by using machine learning algorithms for predicting well disturbances. Prediction based on machine learning methods, regression analysis, identifying outliers in the data, visualization and interactive processing. The algorithms based on oil wells operation data allow training the forecasting model and, on its basis, determine the presence or absence of disturbances in the wells. As a result, the machine forecast showed high accuracy in identifying wells with disturbances. Based on this, candidate wells can be selected for further work. For each specific well, an optimal set of studies can be planned, as well as candidate wells can be selected for further repair and isolation work. In addition, in the course of this work, a set of scientific and technical solutions was developed using machine learning algorithms. This approach will allow predicting disturbances in the well without stopping it.

Keywords: machine learning, behind-casing crossflow, artificial intelligence, classification

1 INTRODUCTION

Around the world, there is now a deterioration in the structure of hydrocarbon reserves and a reduction in the number of new discovered fields. Consequently, the use of modern innovative technologies is extremely important for the further development of the sector. According to experts, only this will allow to reduce production costs and accidents [1, 2, 3, 4]. It will also increase the efficiency of geological exploration and enhance oil recovery. To implement this, the introduction of such information technologies as, for example, big data analysis, the Internet of things, artificial intelligence is required [5, 6, 7]. In particular, machine learning is beginning to be applied at a fairly rapid pace. It has received great development in recent years as a branch of science that originated at the intersection of artificial intelligence and data science technologies [8, 9].

Machine learning is a subsection of science dedicated to the development and study of artificial intelligence. It, in turn, includes various computer systems that can simulate human thinking. These concepts are often used in the same context and sometimes as interchangeable, but they carry different meanings. Thus, machine learning is always based on the use of artificial intelligence, while artificial intelligence does not always involve machine learning [10].

Machine learning is about building mathematical models to explore some kind of data. First, the model is trained on the initial data, identifies key patterns and relationships. And then, after completing the training, it can use the received information to predict the data of the next observations.

There are a sufficient number of successful examples of the implementation of machine learning methods both in Russia and abroad. Among domestic companies, good results in the application of this technology are shown by the divisions of the companies Gazprom, Lukoil, Rosneft, Tatneft and others.

Among the latest projects of Gazprom Neft is the “Cognitive Geologist” system. This technology makes it possible to reduce the duration of geological analysis, determines the probabilities of possible geological scenarios, and also gives recommendations on optimal additional study measures. The developed solution helps to better consider the risks in the early stages of exploration and quickly determine which geological scenario is true. The potential economic effect of this development is estimated at billions of rubles [11, 12].

In February 2019, work began on the construction of six wells using the “Digital Drilling” project at the assets of Gazpromneft-Noyabrskneftegaz. The use of automation and machine learning technologies made it possible to increase the rate of penetration of the well by 28.5% on average [13].

“PermNIPIneft” (branch of "LUKOIL-Engineering") has developed and implemented technologies for automatic description of thin sections and analysis of the results of core studies of the Usinskfield. A trained neural network allows building three-dimensional geological models and interpreting research data faster and better. The use of this machine learning technology can help to reduce the total time spent on modeling work by more than 2 times [14].

Many international companies, such as British Petroleum, Shell, Chevron, Schlumberger, General Electric Oil & Gas, began to actively use digital solutions at all stages of hydrocarbon production even earlier than their Russian colleagues. They have implemented a sufficient number of successful projects. for example, the creation of systems for predicting and preventing complications during well operation [15, 16, 17], and the widespread implementation of the concept of a "smart field" [18, 19]. These solutions have contributed to a significant reduction in the cost of extracting petroleum products. In addition, the equipment utilization efficiency has improved.

Artificial neural networks (ANN) are successfully used in solving geophysical and geotechnical problems. Involvement of ANN allows to reduce the required number of wells and tests to determine the characteristics of soils and properties of reservoirs, leading to significant savings in money and time. For example, the use of neural networks in mapping soil layers in northern Iran has shown a high degree of prediction accuracy of trained ANN-based models — about 90% (when compared with test well data) [20]. The use of neural networks reduces the cost of research, improves the quality of geological assessment, and facilitates the interpretation of the structure of underground layers. The usefulness of ANNs is due to their ability to process a large amount of data, work with non-linear relationships, adapt to changing conditions, generalize and learn. Artificial neural networks, along with linear regressions, are used to predict various geophysical parameters (for example, reservoir properties such as porosity and effective formation thickness), to build well logging curves. ANNs are also used in geophysics to interpret log data [21], interpret seismic data [22], determine the lithological structure [23] and the boundaries of geological objects [24], water saturation analysis [25] and permeability [26]. The analysis of geological data is extremely important for assessing the oil and gas content of the studied areas since due to various deposits, such a significant component as permeability can reach only a third of the initial permeability [27]. Artificial neural networks make it possible to analyze a geological section based on seismic data (the most effective geophysical method for finding hydrocarbons) [28]. The use of artificial intelligence in this direction increases the efficiency of geological exploration, increasing their speed and accuracy and reducing costs [29, 30].

This experience of using machine learning methods proves the high efficiency of their application in practice. They are suitable for solving a large number of problems that enterprises face on a daily basis when developing oil and gas fields [31, 32].

One of such important tasks is the preventive detection of such a negative phenomenon as the behind-casing crossflow. It impedes oil production from wells due to an abnormal increase in their water cut. Behind-casing circulation is the communication of the perforations of the main development object (oil or gas reservoir) with a water or gas layer lying above or below [33, 34].

The appearance of behind-the-casing flows in wells occurs for various reasons:

- carrying out various geological and technical measures - acid treatments, perforation works;
- injection of fluids during hydraulic fracturing;
- poor quality of primary cementing of the production casing;
- high bottomhole injection pressure in injection wells.

An analysis of the results of special well surveys, which consist in the systemic control over the development of an oil field, showed that technogenic changes from the use of hydraulic fracturing and bottom-hole zone treatment can significantly change the hydrodynamic characteristics of the reservoir and the integrity of the production string, especially due to technologies carried out at high pressures. So, this type of complications is one of the most frequent in practice and difficult to diagnose [35, 36].

The aim of the project is to develop algorithms, that allow determining the presence of behind-casing crossflow with using oil well operation data, based on machine learning methods. The object of research, respectively, are methods, algorithms, and tools for predicting the presence of behind-casing circulation. As the main part of the work, it is required to develop a machine learning model that makes it possible to determine the probability of the presence of a crossflow based on the history of well operation data.

The developed algorithms will allow:

a) improve the responsiveness in conditions of an increased probability of the presence of a behind-casing crossflow.
b) reduce the cost of conducting geological field research.

The solution to this problem will make it possible to qualitatively identify candidate wells for geological and technological measures, both in producing and injection wells. In particular, this will help in identifying wells for repair and insulation works based on operational data. Such data include measurements of water cut, fluid flow rate, data on bottomhole and reservoir pressures, information on the presence of fractures and their characteristics.

Thus, the novelty of the presented work lies in the study of the problem at the intersection of the sciences of oil and gas engineering and machine learning. Such a combination of scientific fields makes it possible to improve the accuracy of identifying complications and do it in a timely manner, while avoiding significant costs for restoration work.

2 MATERIALS AND METHODS

Models for predicting well disturbances were built on the basis of the following data:

- Water cut (fractions of units or percentages)
- Liquid flow rate (m3 / day)
- Data on bottomhole and reservoir pressures (atm.)
- The presence of disturbances in the investigated well (Boolean variable)

For instance, the water cut of various objects in the data could range from 0 to 100 percent and the fluid flow rate from 0 to 86 m3/day. The feature of the presence of violations as a Boolean variable has only values interpreted as "true" and "false".

2.1 Temporal data classification methodology

According to the terms of the formulation of the research problem, it is necessary to determine the serviceability of oil producing wells using machine learning methods. To solve it, production data containing well characteristics in the form of time series were provided. A time series can be explained as an ordered sequence of values, each of which corresponds to a specific date.

In addition, the dates of the inspections were also obtained, as a result of the analysis of which it is possible to draw conclusions about the disturbances in the operation of the wells. Based on these data, it is proposed to consider the binary classification of the technical condition of the well based on its characteristics. In this situation, classification means determining the probability of the presence or absence of complications in the well.

Considering these aspects, the problem can be defined as Time Series Classification (TSC). The difference between TSC and conventional classification tasks is that all characteristic values in it have a strictly defined sequence. This method has been known for a long time and has wide application in many industries such as medicine, biology, finance, and engineering [37, 38].

2.2 Data preparation

The first step in data processing is to remove outliers and duplicate records. The three-sigma rule was applied to determine outliers for each parameter. It is almost certain that a random variable will not deviate from the mathematical expectation in absolute value by more than three times the standard deviation. Thus, the upper and lower limits of the values were determined. Values that do not fall within a certain interval were considered outliers and cut out from the data (Figure 1). Figure 2 shows a visualization of the fluid rate and water cut data after removing outliers and duplicates.

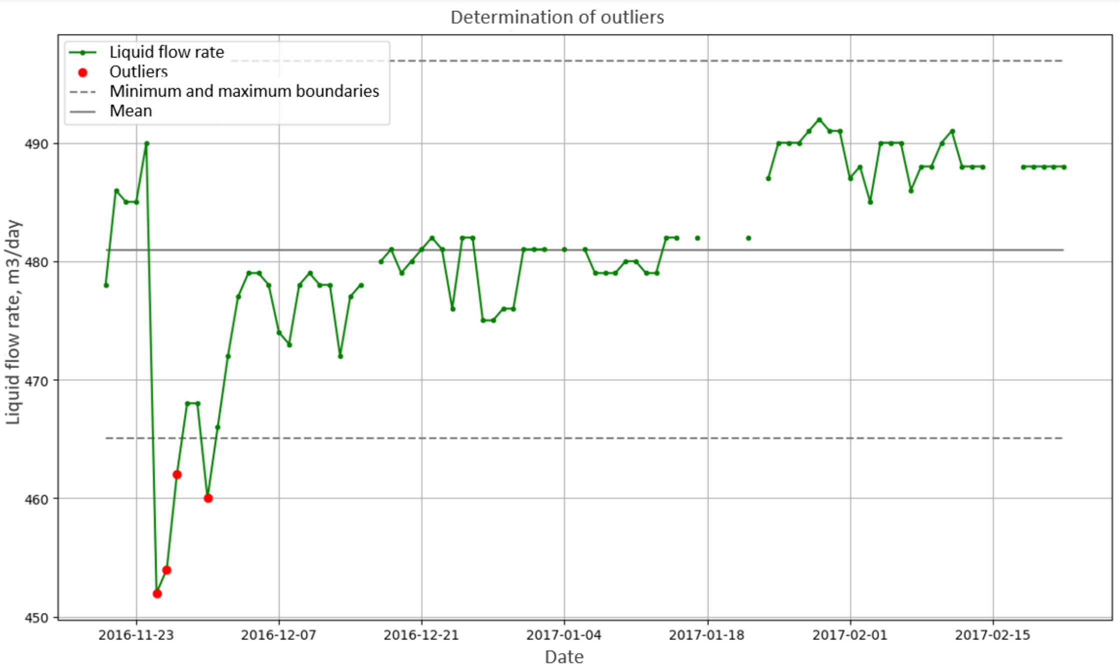

Figure 1. Illustration of defining outliers.

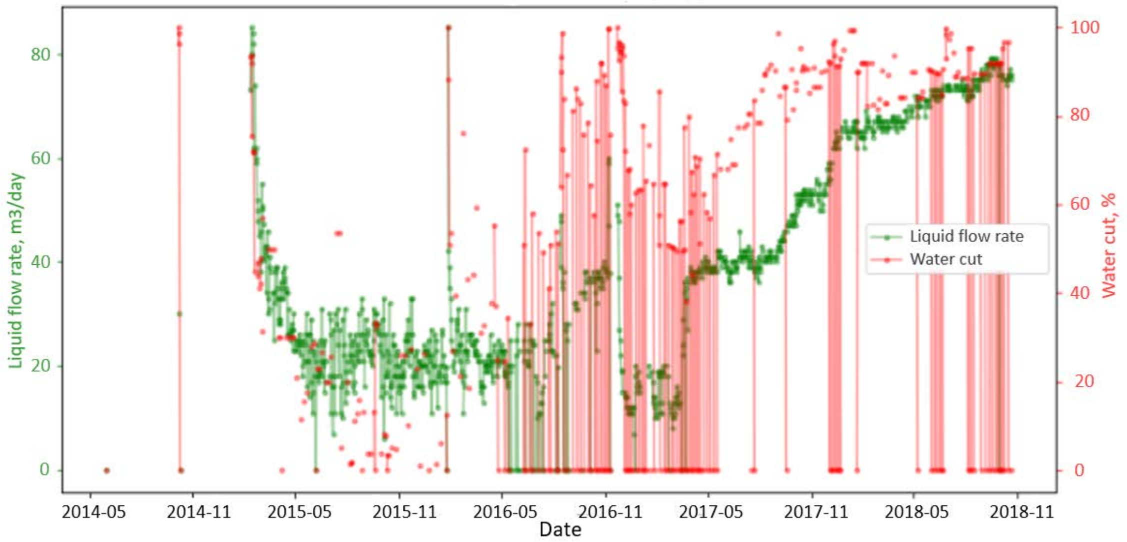

Figure 2. Initial data after cleaning.

2.3 Creation of a synthetic dataset.

Since the number of missing values is not large - about 14% of the available data, it was reasonable to fill them in by interpolation. For this, 4 different interpolation methods were tested: cubic, quadratic, monotone cubic, and Akima interpolation. With cubic interpolation, the tails between points turned out to be the most pronounced among all the listed approaches. Interpolation using other methods is shown in Figure 3.

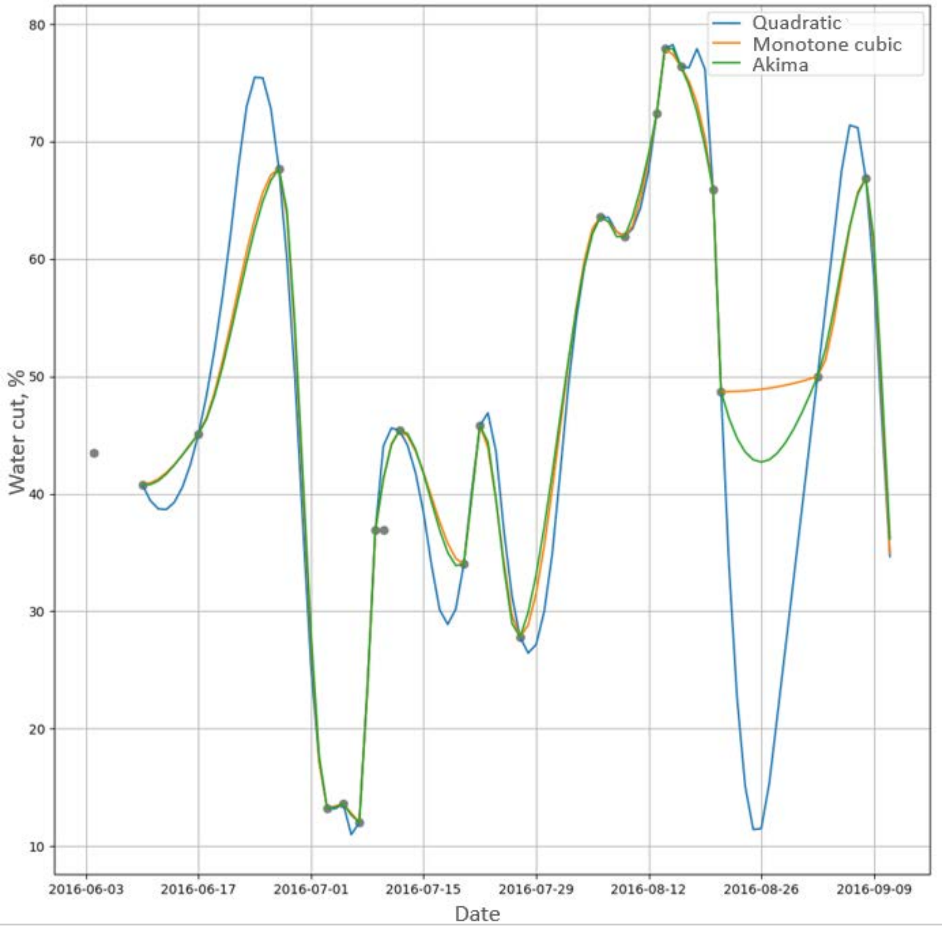

Figure 3. Various interpolation methods.

Based on the results obtained, Akima interpolation was chosen for application on the dataset, since, unlike the quadratic, it does not leave tails and is smoother than the monotone cubic one. The interpolation method by Akima uses a continuously differentiable sub-spline built from piecewise cubic polynomials. The resultant curve passes through the given data points and will appear smooth and natural.

Then the data were divided into training sets, where each time series contains the date of the production logging. Its results were used for data labeling and subsequent training. In this way, since labeled datasets were used to train algorithms that classify data, supervised learning was implemented in that work. And it's important that the data were filtered for the presence of actual values of the characteristics previously selected for prediction.

Initially there were measured data for almost 200 wells. As a result of data processing, the number of wells decreased to 44. Since these objects had the most relevant and clean data that is also quite balanced by class.

After segmentation, the following parameters were generated for each segment:

- mathematical expectation for each characteristic;
- standard deviation for each characteristic;
- ratio of water cut to liquid flow rate;
- ratio of water cut to bottomhole pressure;
- ratio of fluid flow rate to bottomhole pressure;
- trend;
- coefficient of determination, calculated on the basis of linear regression.

The trend function is used to represent the behavior of the features in order to determine if a certain pattern exists by the slope of the approximating line. The coefficient of determination was defined for this line for each parameter. These new features further allowed the model to be trained more qualitatively. Since with them it is possible to capture patterns in data changes more efficiently.

2.4 Classification models

To select a suitable model, 5 classification models were tested using cross-validation: Logistic Regression (LR), Support Vector Machine (SVM), K-Nearest Neighbors algorithm (KNN), Random Forest (RF), and Naive Bayes (NB).

For modeling, a Scikit-learn library with the implementation of a number of algorithms for the Python programming language was used. The results of applying various models on the test set depending on the number of wells are shown in Figure 4. In each case, a ratio of 80% training set to 20% test set was used to split the data.

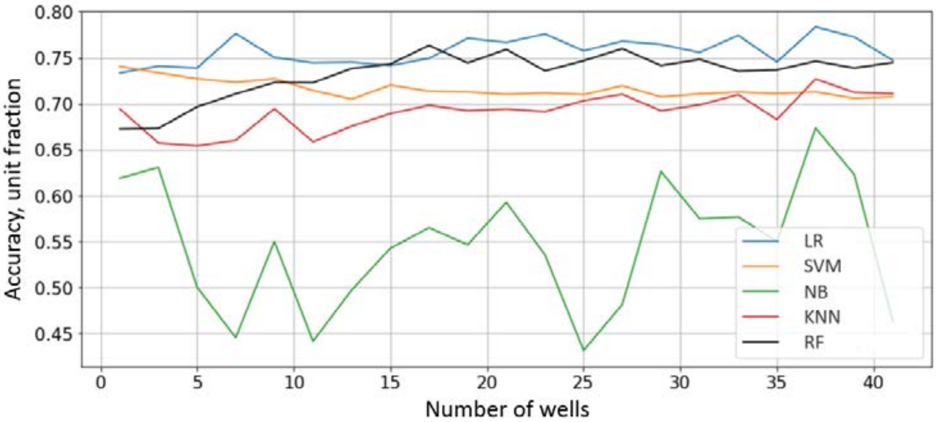

Figure 4. Results of testing models.

As can be seen from the graph, Naive Bayes model gives extremely unstable results, and SVM model results behave irregularly, so the best and most reliable results are shown by Logistic Regression and Random Forest models. Comparison of these models is quite common in the literature [39]. Based on this, it was decided to check the quality of the selected models by the number of correctly classified wells.

Thus, the predictions of these models with the default parameters were compared. The Logistic Regression model correctly identified more wells with disturbances than Random Forest. Consequently, it coped with the task better, and therefore this particular model was used in further research.

Logistic regression has a curve towards the line of best fit, which makes it much more suitable for categorical data predictions. Unlike the linear model, it predicts not only values of 0 or 1, but also a probability score that ranges between them. This is realized by taking the natural logarithm of the equation with probability of event P.

$$Z_i = ln\left(\frac{P_i}{1 - P_i}\right) = \alpha + \beta_1 x_1 + \cdots + \beta_n x_n$$

Taking exponent gives:

$$P_i = E(y = 1|x_i) = \frac{e^z}{1 + e^z} = \frac{e^{\alpha+\beta_i x_i}}{1 + e^{\alpha+\beta_i x_i}}$$

Alpha and beta are weights that allow to get the output value y through the input values x. Based on the predicted probability values, it is possible to classify all observations into two groups. For this, a cut-off point is used, which is usually taken to be 0.5.

The next step after choosing a model was to select the most suitable parameters for it. In this case, three main parameters were considered: “penalty”, “C” and “class_weight”. They are responsible for the norm of the penalty, the inverse of regularization strength and weights for classes, respectively. The Random Search method was used to optimize them. It is a method in which random combinations of hyperparameters are selected and used to train a model. It is faster than Grid Search and, since it doesn’t reach the best point in the grid, it avoids overfitting and is more able to generalize. As a result of Random Search, L2 penalty term, C = 10 and “balanced” mode for classes were determined. This “class_weight” value automatically adjusts the weights in inverse proportion to the class frequencies, which helps to support learning when the ratio of data in the original dataset changes. However, it is important to understand that the setting of hyperparameters is very dependent on the conditions of the problem and the data supplied, so their values are selected for each specific case.

Further, after selecting and applying the optimal parameters on the model, it was necessary to choose the appropriate window size that determines. A sliding window is a time interval that allows to form a data set from the members of the time series, which can serve as a training set for building a forecasting model. Using a sliding window, it is possible to noticeably increase the number of available observations in the original dataset. For selection, graphs of the dependence of determining the correct number of complicated wells on the window size were built (Figure 5). They show curves for two cases: when the model was trained on all features, and when only data on water cut and fluid flow rate were used for training. These cases were considered in order to take into account the issue of the importance of features from a practical point of view. It can be seen that by removing the pressure data, the prediction worsens for almost all values of the window length. However, in this case, training of the model is still possible, while the data on the flow rate and water cut are strictly necessary for building the model. The y-axis shows the average number of correctly identified wells in each of the folds, or in other words, the true positive rate.

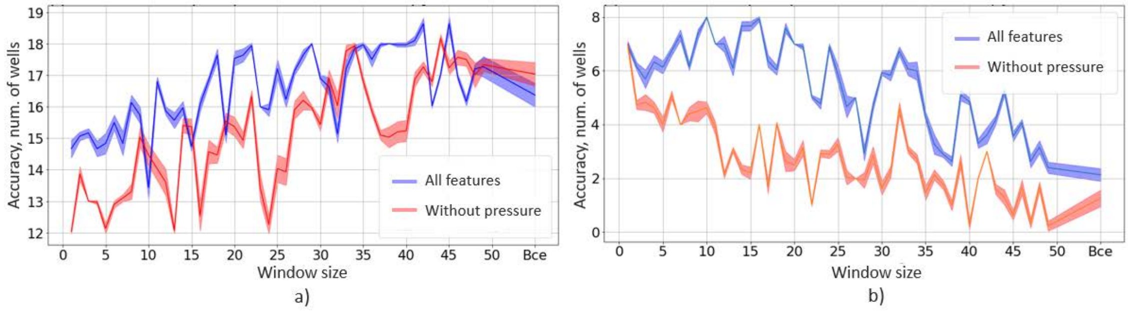

Figure 5. Dependence of the accuracy of predictions on the window size and the number of parameters for wells with disturbances (a) and without disturbances (b).

Based on this, the optimal window size was chosen equal to 20, since it achieves the highest average accuracy of the results. Further, for the analysis of errors, an additional graph was built (Figure 6). It shows the probability of the presence of a behind-casing crossflow in the well, depending on the number of selected segments in the time series of this well.

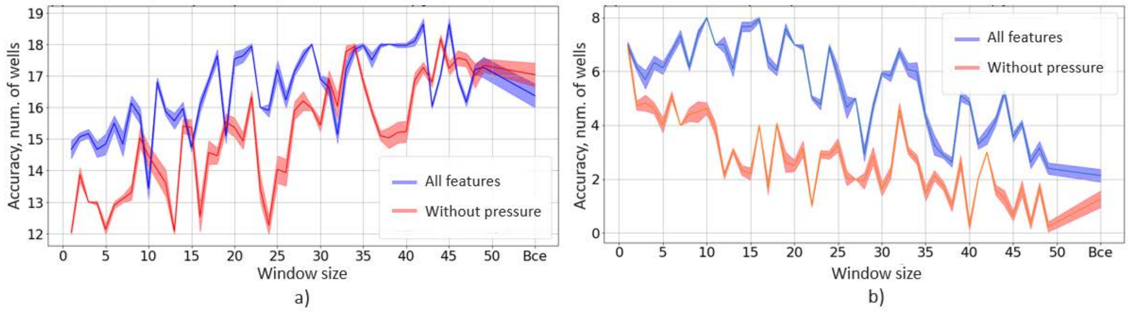

Figure 6. Dependence of the probability of a disturbance on the number of selected segments in the well.

Analyzing this graph, 4 errors can be distinguished, two of which are type I errors (the correct null hypothesis is rejected), and two more are type II errors (an incorrect null hypothesis is accepted). It can be seen that the developed algorithm in this case allowed a large number of errors with less than 30 segments in the well. And with a larger number of segments, more correct results are observed. It is also clearly seen that the cut-off threshold between wells with and without disturbances is equal to 0.4.

2.5 Ensemble method

To improve the quality of model predictions, it was decided to use bootstrap aggregation (bagging) - one of the main classes of ensemble learning methods [40]. Bagging is composed of two parts: bootstrapping and aggregation. Bootstrapping is the method of creating different subsets of the training dataset by selecting data points with replacement. Next, training occurs in parallel on each subset, and the results are aggregated. That is, in the case of the considered classification problem, the value of the class that was chosen by the majority is selected (majority voting).

In this way, ten thousand random permutations were made to create different samples of the same training dataset, then each model was trained on such sample of 35 wells. The results of applying this method are shown on bar graphs with the number of correctly identified wells from the train, test, and full samples (Figure 7).

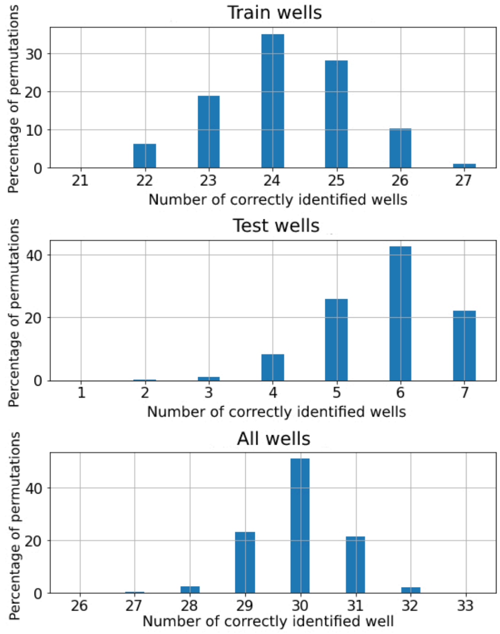

Figure 7. Results of model prediction after applying the ensemble method.

The relationship between the number of permutations and correctly classified wells is not direct, and with fewer permutations, better results are possible. Accordingly, it is important to strike a balance, as choosing the right percentage affects the performance of bagging.

It can be seen that, on the whole, good final results were obtained. On the other side, they were most likely influenced by some anomalies in the data. They persist, despite the preparation of the data, and slightly deteriorate the quality of the predictions.

Further, to analyze the trained models, histograms were built displaying the estimates of segments by different models (Figure 8). Kernel smoothing was applied to each histogram and the following characteristics were calculated: mean, median, mode. It can be seen from the graphs that the mode best reflects the situation for each segment, as it shows the most repetitive assessment.

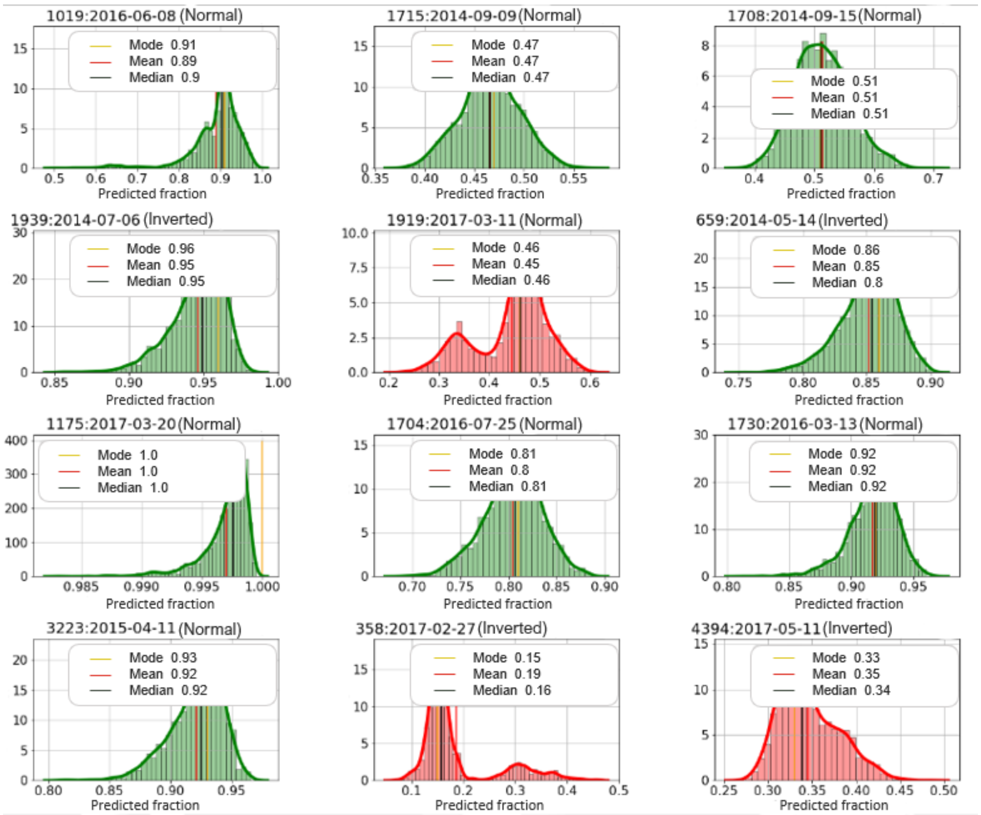

Figure 8. Graphs of evaluation of segments by an ensemble of models.

Based on a more detailed contemplation of the distribution of estimates over the ensemble of models, it was noticed that the cut-off threshold is higher than previously considered. The analysis of the values showed that the highest probability of the presence of a behind-casing crossflow in wells where it does not exist is equal to 0.52. At the same time, the minimum probability of the presence of a crossflow in wells where it is present exceeds 0.8. The new cut-off threshold value was taken equal to 0.6, as it allows to minimize type I and type II errors.

3 RESULTS

Thus, the statistical model in our case was represented by logistic regression. The mathematical equation of this approach is given above. The model selects weights for each of the input parameters. At the output, we get a value compared with the cut-off threshold value, and the final binary result is determined. For instance, if a well has a probability value of 0.5 and a threshold was set to 0.6, then the predicted value is 0, which classifies it as not leaking.

Upon detailed consideration of the classifier, it was found that the signs related to saturation pressure had the greatest effect on the shift in the predicted value, and the oil production feature had the most weight in the case of a problem well.

To analyze the coefficients of the models, the results of their decomposition were visualized. For this, Principal Components Analysis (PCA) (Figure 9) and Uniform Manifold Approximation and Projection (UMAP) (Figure 10) were applied. These dimensionality reduction algorithms differ in that the PCA tries to preserve the global data structure, while the UMAP tries to preserve the local distances between the existing points [41, 42].

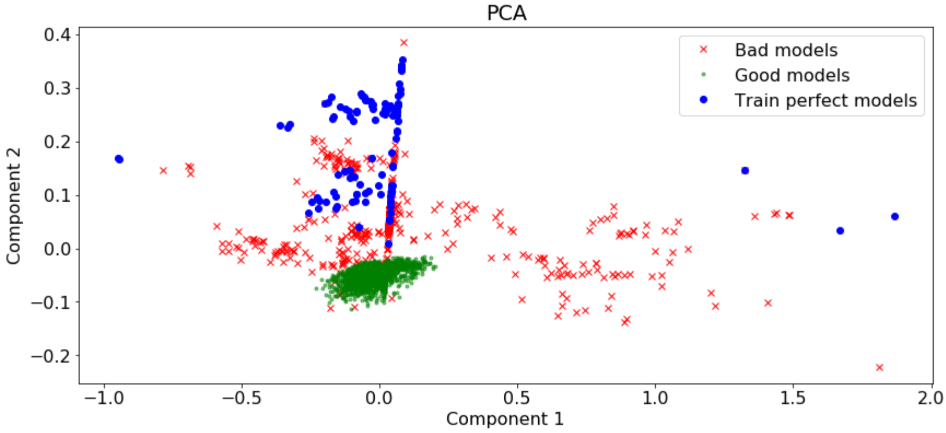

Figure 9. The result of decomposition of the model coefficients using the PCA.

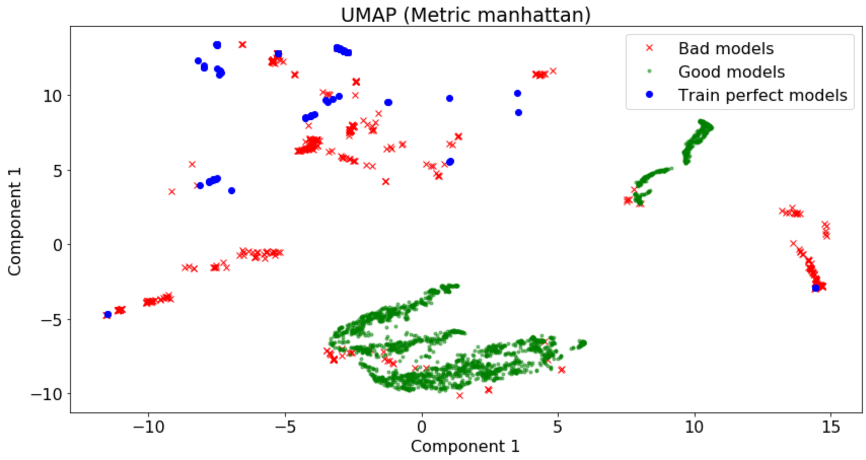

Figure 10. The result of decomposition of the model coefficients using the UMAP.

The PCA looks for the hyperplane that is closest in space to the available data and projects the data onto it. The principal component is the axis that explains the maximum amount of variance in the training set. And the second principal component is the axis orthogonal to it. The UMAP is a nonlinear dimensionality reduction method whose main idea is also to represent components. As can be seen from these pictures, good models stand out in a separate cluster both with the use of PCA and UMAP.

As a result, the trained model was used to predict complications in three mature fields with the number of wells exceeding 1000. Testing the forecasting model on the entire operating stock of injection wells of the selected fields showed that for the first field the prediction accuracy was 92%, for the second - 82%, and for the third - 90%. For production wells, machine learning results showed an accuracy of 92%, 86%, and 80%, respectively. Then the final analysis and comparison of the convergence of the machine learning predictions with the field geophysical survey results were carried out. On average, for all three cases, the convergence with these works was about 90%. Cross-validation was used for the data, and then the results were averaged over the folds. Therefore, the obtained results can be trusted.

Thus, machine learning has shown a sufficiently high accuracy of predictions. Based on this, it is possible to select candidate wells for geological and technical measures and production logging. This technology allows planning the optimal production logging complex for each specific well, depending on the problem being solved, and removing uncertainties. It is also able to assist in the selection of candidate wells for further exploration and survey work.

4 DISCUSSION

The developed algorithms made it possible to build a machine learning model, which showed good results not only on the training sample but also on the test sample. And this indicates the possibility of its application on new, yet unknown for the model data. This model can be used both in new and mature fields, as an analytical assessment of the technical condition of wells and for planning an optimal set of surveys.

This model has some limitations for its correct operation. In order to correctly train the model on the data, it is necessary that these data have as few missing values and outliers as possible, the frequency of measurement of the main input indicators is also important. The original dataset must be carefully and accurately formed in a uniform format. And it should be kept in mind that the results of machine learning can be used as an additional tool in the complex of decision-making and issuing conclusions, and not as a fundamental one.

When working with the model, it was noted that for some wells there was no correct and regular measurement data, which complicates the process of training the model. Therefore, in order to further improve the existing solutions, it is proposed to increase the well database. Additionally, in one of the fields, the number of segments with disturbances was three times less than without them. So, it is worth trying to train the model using the most equal partitioning according to this criterion.

In the course of the work, it was possible to develop a set of scientific and technical solutions using machine learning algorithms. This technological solution makes it possible to predict the presence or absence of disturbances in the operation of the well without shutting it down. Such a forecasting approach will help to promptly select candidates for production logging in conditions of the probability of disturbances in the production casing, plan candidates for further geological and technical actions, and in particular, identify wells for repair and insulation works. In addition, this reduces the cost of conducting uncertain, uninformative geological surveys and extends the period of trouble-free well operation.

5 CONCLUSIONS

As a result of the work, a study of approaches that allow, based on machine learning methods using oil well operation data, to determine the presence of behind-the-casing crossflow was carried out. Thus, in the course of this study, the following tasks were completed:

- comparative analysis of the data from different sources to determine the quantitative and qualitative characteristics of the parameters for subsequent analysis of the presence of disturbances;
- analysis of workover operations and geophysical surveys results for the presence of the data on well disturbances in the relevant periods;
- preparation of initial data and the formation of a synthetic dataset, compiled from various sources;
- implementation of a model for predicting the presence of behind-the-casing circulation based on the data from oil producing wells.

In the future, to improve the performance of the machine learning model, it is planned, in addition to expanding the existing database, to also add the ability to predict a number of other disturbances in wells. It will help not only identify problem wells but also more accurately differentiate the source of these disturbances. The application of this approach will make it possible to significantly reduce time and money costs, which emphasizes the feasibility of using innovative methods in field development.

6 REFERENCES

[1] Mittal, A., Slaughter, A. and Bansa, V., 2017. From bytes to barrels. The digital transformation in upstream oil and gas. Deloitte Insights. Available at:
<https://www2.deloitte.com/content/dam/Deloitte/global/Documents/Energy-and-Resources/gx-online-from-bytes-to-barrels.pdf> [Accessed 17 September 2021].

[2] Baker, D., 2021. Digital Technology Trends in the Oil and Gas Industry. GEO ExPro. Available at: <https://www.geoexpro.com/articles/2019/07/digital-technology-trends-in-the-oil-and-gas-industry> [Accessed 19 September 2021].

[3] Legkokonets, V.A., Islamov, S.R., Mardashov, D.V., 2019. Multifactor analysis of well killing operations on oil and gas condensate field with a fractured reservoir. Proceedings of the International Forum-Contest of Young Researchers: Topical Issues of Rational Use of Mineral Resources, Taylor & Francis: London, UK, pp. 111-118.

[4] Stroykov, G., Babyr, N., Ilin, I., Marchenko, R., 2021. System of Comprehensive Assessment of Project Risks in the Energy Industry, International Journal of Engineering, 34(7), pp. 1778-1784. DOI: 10.5829/ije.2021.34.07a.22.

[5] Slaughter A., Bean G., andMittal A., August 14, 2015. Connected barrels: Transforming oil and gas strategies with the Internet of Things, Deloitte University Press. Available at: <https://dupress.deloitte.com/dup-us-en/focus/internet-of-things/iot-in-oil-and-gas-industry.html> [Accessed 10 September 2021].

[6] Alsaihati, A., Elkatatny, S., Mahmoud, A. and Abdulraheem, A., 2021. Use of Machine Learning and Data Analytics to Detect Downhole Abnormalities While Drilling Horizontal Wells, With Real Case Study. Journal of Energy Resources Technology, 143(4). DOI:10.1115/1.4048070.

[7] Bello, O., Holzmann, J.,Yaqoob, T. and Teodoriu, C., 2015. Application Of Artificial Intelligence Methods In Drilling System Design And Operations: A Review Of The State Of The Art. Journal of Artificial Intelligence and Soft Computing Research, Vol.5 (Issue 2), pp. 121-139. DOI: 10.1515/jaiscr-2015-0024.

[8] Forsyth, D., 2020. Applied machine learning. [s.l.]: Springer Nature.

[9] Babyr, N., & Babyr, K., 2021. To improve the contact adaptability of mechanical roof support. Paper presented at the E3S Web of Conferences, 266. DOI:10.1051/e3sconf/202126603015.

[10] Turner, R., 2019. Machine Learning: The Ultimate Beginner's Guide to Learn Machine Learning, Artificial Intelligence & Neural Networks Step by Step. N.p., nelly B.L. International Consulting Limited.

[11] Gazprom-neft.com. 2021. Geological prospecting. Available at: <https://www.gazprom-neft.com/company/exploration-and-production/exploration/> [Accessed 2 October 2021].

[12] Offshore-technology.com. 2021. Gazprom Neft and IBM Research Brazil work on geological processing AI. Available at: <https://www.offshore-technology.com/news/gazprom-neft-ibm-ai/> [Accessed 2 October 2021].

[13] Ir.gazprom-neft.com. 2021. Gazprom Neft and Nefteservisholding to cooperate in technological development - Investors - «Gazprom Neft» PJSC. Available at: <https://ir.gazprom-neft.com/news-and-events/news/2020/gazprom_neft_and_nefteservisholding_to_cooperate_in_technological_development_5203404/> [Accessed 13 September 2021].

[14] Martyushev, D., Ponomareva, I., Zakharov, L. and Shadrov, T., 2021. Application Of Machine Learning For Forecasting Formation Pressure In Oil Field Development. Izvestiya Tomskogo Politekhnicheskogo Universiteta Inziniring Georesursov, 332(10), pp.140-149. DOI: 10.18799/24131830/2021/10/3401.

[15] Parviainen, P., Kaariainen, J., Tihinen, M. &Teppola, S., 2017. Tackling the digitalization challenge: how to benefit from digitalization in practice. International Journal of Information Systems and Project Management,5(1), pp. 63-77. DOI: 10.12821/ijispm050104.

[16] Wei, Z., Zhu, S., Dai, X., Wang, X., Yapanto, L. M., & Raupov, I. R., 2021. Multi-criteria decision making approaches to select appropriate enhanced oil recovery techniques in petroleum industries. Energy Reports, 7(3), pp. 2751-2758. DOI: 10.1016/j.egyr.2021.05.002.

[17] Rogatchev, M. K., & Kuznetsova, A. N., 2019. Technology of low-permeable polimictic reservoirs water-flooding with surfactant solutions. Paper presented at the Innovation-Based Development of the Mineral Resources Sector: Challenges and Prospects - 11th Conference of the Russian-German Raw Materials, 2018, 161-166.

[18] Temizel, C., Canbaz, C., Palabiyik, Y., Putra, D., Asena, A., Ranjith, R. and Jongkittinarukorn, K., 2019. A Comprehensive Review of Smart/Intelligent Oilfield Technologies and Applications in the Oil and Gas Industry. SPE Middle East Oil and Gas Show and Conference. DOI: 10.2118/195095-MS.

[19] Bourgeois, B. et al., 2015. A Framework for Sustainable Digital Oilfield Solutions. SPE Digital Energy Conference and Exhibition. The Woodlands, Texas, USA, Society of Petroleum Engineers. DOI: 10.2118/173436-MS.

[20] Choobbasti A. J., Farrokhzad F., Mashaie E.R., Azar P.H. Mapping of soil layers using artificial neural network (case study of Babol, northern Iran). Journal of the South African Institution of Civil Engineering, March 2015, vol. 57, no. 1, pp. 59-66.

[21] Rodina S. N., Silkin K. Yu. Primenenie nejrosetevogo podhoda pri interpretacii karotazhnyh dannyh [Application of neural network approach in interpreting well logging data]. Vestnik VGU, Geologiya [Proceedings of Voronezh State University. Geology], 2007, no. 2, pp. 184-188.

[22] Kuroda M. C., Vidal A. C., Almeida de Carvalho A. M. Interpretation of seismic multiattributes using a neural network. Journal of Applied Geophysics, 2012, vol. 85, pp. 15-24.

[23] Akinyokun O.C., Enikanselu P.A., Adeyemo A.B., Adesida A. Well log interpretation model for the determination of lithology and fluid contents. The Pacific Journal of Science and Technology, 2009, vol. 10, no. 1, pp. 507-517.

[24] Shajbakov R. A. Ispol’zovanie nejrosetevogo apparata dlya identifikacii granic geologicheskih ob”ektov [The use of neural network mechanism for identifying of geological objects boundaries]. Materials of the international scientific conference “Tekhnicheskie nauki: tradicii i innovacii” [“Technical Sciences: Traditions and Innovations”], Chelyabinsk: Dva komsomol’ca, 2012, pp. 8-11.

[25] Nakutnyy P., Asghari K., Torn A. Analysis of waterflooding through application of neural networks. Conference Paper, Canadian International Petroleum Conference, Calgary, Alberta, 17-19 June, 2008.

[26] Huang Z., Shimeld J., Williamson M., Katsube J. Permeability prediction with artificial neural network modeling in the Venture gas field, offshore eastern Canada, Geophysics, 1996, vol. 61, pp. 422-436.

[27] Khormali A., Moghadasi R., Kazemzadeh Y., Struchkov I., 2021. Development of a new chemical solvent package for increasing the asphaltene removal performance under static and dynamic conditions, Journal of Petroleum Science and Engineering, vol. 206, DOI: 10.1016/j.petrol.2021.109066.

[28] Borisov A.S., Kulikov S.A. Iskusstvennye nejronnye seti v prognozirovanii neftegazonosnosti po dannym sejsmorazvedki [Artificial neural networks in the forecasting of petroleum potential from seismic data]. Kazanskij (Privolzhskij) federal’nyj universitet, Institut geologii i neftegazovyh tekhnologij [Kazan Federal University, Institute of Geology and Petroleum Technologies]. 2012.

[29] Mutalimov, V., Kovaleva, I., Mikhaylov, A., & Stepanova, D. (2021). Assessing regional growth of small business in Russia. Entrepreneurial Business and Economics Review, 9(3), 119-133. https://doi.org/10.15678/EBER.2021.090308

[30] Kaznacheev P. F., Samoylova R. V., Kurchiski N. V. Application of artificial intelligence methods to improve efficiency in the oil and gas and other raw materials industries. Economic Policy, 2016, vol. 11, no. 5, pp. 188-197.

[31] Sultanbekov, R., Beloglazov, I.; Islamov, S., Ong, M.C., 2021. Exploring of the incompatibility of marine residual fuel: A case study using machine learning methods. Energies 2021, 14, 8422. (20)

[32] An, J., Mikhaylov, A., Jung, S.-U. (2020). The Strategy of South Korea in the Global Oil Market. Energies, 13(10), 2491. https://doi.org/10.3390/en13102491

[33] Mardashov D. V., 2006. Improving the technology of casing the near-wellbore zone in order to eliminate crossflows in gas condensate wells. Journal of Mining Institute, 167(2), p. 35. (21)

[34] Nikolaev N. I., & Khaoya L., 2017. Results of cement-to-rock contact study. Journal of Mining Institute, 226, 428.

[35] Syah, R., Alizadeh, S. M., Nurgalieva, K. S., Guerrero, J. W. G., Nasution, M. K. M., Davarpanah, A., Metwally, A. S. M., 2021. A laboratory approach to measure enhanced gas recovery from a tight gas reservoir during supercritical carbon dioxide injection. Sustainability (Switzerland), 13(21). DOI:10.3390/su132111606.

[36] Dvoynikov M. V., Kuchin V. N., & Mintzaev M. S., 2021. Development of viscoelastic systems and technologies for isolating water-bearing horizons with abnormal formation pressures during oil and gas wells drilling. Journal of Mining Institute, 247(1), 1-9.

[37] Buza, K., 2018. Time Series Classification and its Applications. Proceedings of the 8th International Conference on Web Intelligence, Mining and Semantics. DOI: 10.1145/3227609.3227690.

[38] Islamov, S., Grigoriev, A., Beloglazov, I., Savchenkov, S., Gudmestad, O.T., 2021. Research risk factors in monitoring well drilling – A case study using machine learning methods. Symmetry 2021, 13, 1293.

[39] Kirasich, K., Smith, T., and Sadler, B., 2018. Random Forest vs Logistic Regression: Binary Classification for Heterogeneous Datasets. SMU Data Science Review: Vol. 1: No. 3, Article 9. Available at: <https://scholar.smu.edu/datasciencereview/vol1/iss3/9> [Accessed 29 October 2021].

[40] Brownlee, J., 2016. Machine Learning Algorithms From Scratch with Python. Machine Learning Mastery.

[41] Raupov, I. and Milic J., 2022. Improvement of operational efficiency of high water-cut oil wells. IOP Conf. Ser.: Earth Environ. Sci. 1021 012077

[42] Tetko, I., Kurkova, V., Karpov, P. and Theis, F., 2019. Artificial Neural Networks and Machine Learning – ICANN 2019: Text and Time Series. 28th International Conference on Artificial Neural Networks, Munich, Germany, September 17–19, 2019, Proceedings, Part IV. Springer International Publishing.

Paper submitted: 02.07.2022.

Paper accepted: 10.09.2022.