On increasing the productive time of drilling oil and gas wells using machine learning methods

A.N. Dmitrievsky[1], A.G. Sboev[2], N.A. Eremin[1,3], A.D. Chernikov[1], A.V. Naumov[1], A.V. Gryaznov[1], I.A. Moloshnikov[1], S.O. Borozdin[3], E.A. Safarova[1]*

[1]*Oil and Gas Research Institute of the Russian Academy of Sciences, Moscow, Russian Federation*
[2]*National Research Center «Kurchatov Institute», Moscow, Russian Federation*
[3]*National University of Oil and Gas «Gubkin University», Moscow, Russian Federation*

Abstract. The article is devoted to the development of a hybrid method for predicting and preventing the development of troubles in the process of drilling wells based on machine learning methods and modern neural network models. Troubles during the drilling process, such as filtrate leakoff; gas, oil and water shows and sticking, lead to an increase in unproductive time, i.e. time that is not technically necessary for well construction and is caused by various violations of the production process. Several different approaches have been considered, including based on the regression model for predicting the indicator function, which reflects an approach to a developing trouble, as well as anomaly extraction models built both on basic machine learning algorithms and using the neural network model of deep learning. Showing visualized examples of the work of the developed methods on simulation and real data. Intelligent analysis of Big Geodata from geological and technological measurement stations is based on well-proven machine learning algorithms. Based on these data, a neural network model was proposed to prevent troubles and emergencies during the construction of wells. The use of this method will minimize unproductive drilling time.

Keywords: machine learning, neural networks, detection of anomalies, prediction of troubles, hybrid simulation, drilling of oil and gas wells, geological and technological information, prevention of accidents and complications, artificial intelligence, automated system, neural network modeling

1. Introduction

Modern world science is characterized by significant progress in the development of modern methods of data analysis (Data Driven methods) and mathematical models, including those based on machine learning technologies and neural networks. Thanks to these technologies, the most modern (state-of-the-art) algorithms began to appear that make it possible to effectively help in solving complex problems in the oil and gas field (Arkhipov et al., 2020; Borozdin et al., 2020; Dmitrievsky et al., 2020; Kaznacheev et al., 2016).

Drilling oil and gas wells is an essential aspect of oil and gas production. Improving safety in the course of this complex technological process is an urgent and important task. One of the options for solving this problem is to prevent drilling troubles and emergencies by timely warning the drilling crew about the beginning of their development.

In the presence of several sources of Big geodata during drilling (geosteering system, geological and technological information (GTI), drilling simulator), it is effective to use a new type of modeling – hybrid. A hybrid model is a set of models consisting of a basic 4D wellbore model, a probabilistic (or fuzzy) uncertainty model, and a machine learning model. The hybrid model is continuously refined during the drilling process as heterogeneous big volumes of geological and technological data are received and is used for automated prediction of complications and emergencies. In particular, in the study (Diakonov, Golovina, 2017), the problem of automatic detection of breakdowns of mechanisms and determination of their types based on the collected historical data was considered, which was reduced to the classic problem of machine learning – the detection of anomalies. The paper provided an

*Corresponding author: Nikolai A. Eremin
E-mail: ermn@mail.ru*

extensive review of methods for solving this problem, and also presented the results of their testing on real data, where the best result was shown by the "learning without a teacher" approach, namely the "Isolation Forest" model (Liu et al., 2008). In work (Gurina et al., 2020), the problem of detecting complications and determining their types during drilling was solved by building a machine learning model to identify anomalies in the data. In contrast to previous work, this study took a "supervised learning" approach. In the work described above, real-time logging data were compared with similar data collected previously in the database, in which various types of complications were present. The search results were ranked and the most appropriate complication was selected. For such a comparison, ranking and determination of complications, a gradient boosting classification model was trained (Chen, Guestrin, 2016), which made it possible to achieve an accuracy of determining complications of 0.908 according to the ROC AUC metric, i.e. calculating the area under the performance curve (the ROC AUC metric is one of the popular metrics used in the industry, where AUC (Area Under the Curve) is the area under the curve and the curve is the Receiver Operating Characteristic (ROC) path). In practice, depending on the AUC value, the model's efficiency is classified as follows: $0.8 \leq AUC \leq 1.0$ – the model works perfectly; $0.6 \leq AUC < 0.8$ – the model works well; $0.5 < AUC < 0.6$ – the model works satisfactorily and $AUC \leq 0.5$ – the model does not work. In the study (Kodirov, Shestakov, 2019), a method was developed for detecting stuck drill pipe string based on a neural network. The authors built a multilayer fully connected neural network (MLP – multilayer perceptron), which determined the occurrence of sticking and its type with an accuracy of 93% according to the basic metric "Accuracy" (measures the number of correctly classified objects relative to the total number of all objects).

In the works discussed above, methods were described for determining the events that have already occurred, which do not allow timely response to rapidly developing pre-emergency situations in drilling. They are often successfully used for preliminary marking of a large volume of unmarked GTI from geological and technological measurement stations.

A more difficult task is not only to determine the complications type, but also to predict the likelihood of their origin when drilling in the future. In the work (Pichugin et al., 2013), it is shown that by training a model of decision trees (Decision Tree) on various geo-information obtained from previously drilled wells, it is possible to assess the risk of drilling and occurrence of undesirable situations, as well as to increase the success of new production wells into operation by 15–25%. In another study (Lind et al., 2013), the authors solved the problem of predicting the amount of lost circulation while drilling a new well. For this, a self-learning neural network model was built, called the Kohonen map (Kohonen, 1990), which was trained on information collected from previously drilled wells. The resulting model, according to the authors, will reduce the cost of drilling by up to 4%. The tasks of predicting the values of various drilling parameters in real time are even more complex (Eremin, Stolyarov, 2020; Noshi, Schubert, 2018) and at this stage of the development of machine learning methods are poorly developed. The most successful of these methods are based on deep neural networks with recurrent and convolutional layers (Kanfar et al., 2020; Li et al., 2019).

The forecasting models discussed above allow to assess the risks of future drilling and prepare in advance for possible complications in the drilling process, but they do not allow to predict possible complications so that the drilling team can take timely actions in real time to completely prevent them or minimize possible consequences. To make such forecasts, it is necessary to use GTI obtained in real time. This paper discusses various approaches for predicting the occurrence of three types of pre-accident situations during drilling, using marked and unmarked data sets:
1. Show of gas, oil and water;
2. Drill string sticking;
3. Mud loss.

2. Data

The data used in the process of developing methods and for carrying out planned experiments were provided by partners from the Gubkin Russian State University of Oil and Gas (Gubkin University), in the form of simulation data obtained from a drilling simulator, as well as data from GTI stations when drilling wells in real fields. Both datasets consist of readings taken by various sensors while drilling a well (or simulating it) installed on the equipment. The number of monitored parameters, their completeness and recording frequency in the above datasets differ, which is additional complexity for analysis.

Simulation data

The simulation dataset was obtained from the drilling simulation experiments on the DrillSim-5000 simulator. 79 simulation records were received, of which 33 refer to drilling with reservoir fluids show complications, 27 with cuttings accumulation (drill string sticking), 9 with lost circulation and 10 simulations of accident-free drilling. An example of data from one simulation is shown in Figure 1.

Simulation records data from the simulator are presented in the form of tables with 16 parameters (including hook weight, ROP, bit rotation speed, etc.).
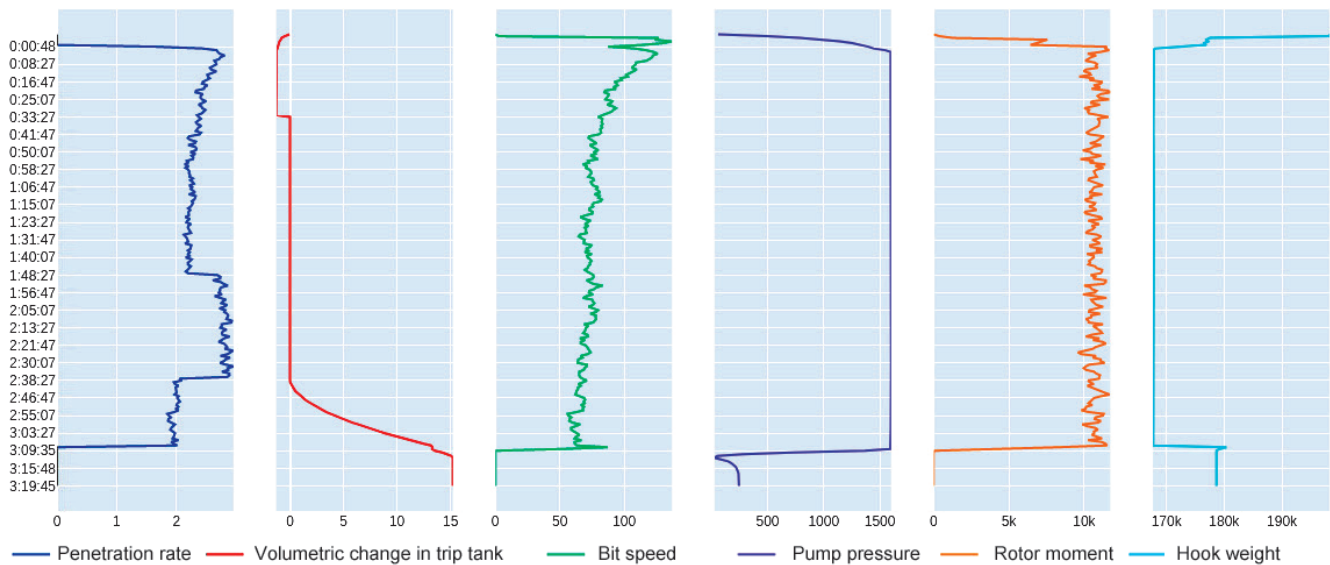
Fig. 1. An example of simulation data with some parameters, with the complication of the type of "gas, oil, and water show"

Each simulation recording was time stamped to indicate the onset of each of the simulated complications. The time stamp was hand-stamped by drilling experts.

Real field data

Real-time records of monitored parameters during drilling from 25 different wells were presented as data from the fields being drilled, where regular drilling records were indicated for 23 wells; one well contains records with a "Drill string sticking" complication, the other with a "Loss of circulation" complication. Examples of some monitored parameters with their ranges of values are shown in Table 1.

Data tagging was carried out by drilling specialists and included the indication of the start time of the complication for each of the 2 types of complications (sticking and mud loss).

Data preparation

In order for the data to be used in training machine learning models and to conduct a qualitative analysis, it was necessary to perform preliminary preparation. For the analysis, segments of continuous drilling were selected (observations with a non-zero rate of penetration), in which all the values of the parameters corresponding to the points in time after the onset of a complication of a given type were discarded, since such observations

are not of interest for the process of predicting its occurrence. Due to the fact that the frequency of taking parameters in the simulation data varied from experiment to experiment, linear interpolation over two neighboring points was used to align them. The final time step between points is 2 seconds. Of all the parameters, the main ones were selected, which are of the greatest interest for determining the considered complications, and are also present in both of the considered datasets. Table 2 presents the final list of the initial parameters used that were used in training machine learning and neural network models.

To expand the space of features and their normalization, a number of additional derived parameters were used,

Parameter	Measurement units	Data type
Hook load	pounds	numerical
Mechanical penetration rate	ft / h	numerical
Rotor speed	rpm	numerical
The difference between the value of the mud flow at the outlet and inlet	US oil barrel	numerical
Pump pressure	psi	numerical

Tab. 2. List of parameters used in models

Parameter name	Minimum value	Average value	Maximum value	Number of unique values
Hook weight (10 kN)	0.0	3.84	46.82	4513
Pressure (kPa)	0.0	1840.83	23781.44	87313
Talblock (m)	0.0	10.83	25.27	2520
Sandface (m)	0.0	625.52	849.01	20745
Bit load (10 kN)	0.0	1.17	19.98	1865

Tab. 1. Some parameters and their range of values for well No. 1

obtained from the readings of the sensors selected after the preliminary data preparation procedure, including: the difference between the current and previous values, the calculation of the moving trend and the decomposition of the parameter values into the slope of the trend and deviation from it, the calculation of the percentiles of the parameter values and their normalization within the percentiles. The data was split into training and test datasets. Since there were no labeled examples for real data, except for two test ones, all wells were included in the training set, except for two, which became control and were also added to the test set. In the case of simulation data, for each of the considered complications, a separate test set of wells was allocated, which includes 20% of all well records containing the considered complication, as well as 20% of randomly selected accident-free wells.

3. Methods and approaches

To solve the problem of predicting the beginning of various pre-emergency situations (complications), in particular, the drill string sticking, mud loss and gas kick considered in this work, the following approaches were implemented, based on:

1. Allocation of anomalies, with construction:
    a. Single-class machine learning model;
    b. Regression neural network model.

2. Construction of a regression function of the indicator, reflecting the approximation to a probable complication.

Approaches 1a and 2 were tested on simulation data obtained from a drilling simulator, as they contained a set of labeled examples for various complications. Approach 1b was applied to real data (two test cases for complications such as "Sticking" and "Mud loss"), there were no labeled training examples.

Anomaly detection model

This approach is based on the task of identifying abnormal situations in the readings from the observed drilling parameters. The main idea here is the following: the closer to complication the values of the observed drilling parameters are considered, the more they differ from those that are typical for accident-free routine drilling under the same conditions. This approach makes it possible to use a large amount of unmarked data, highlighting abnormal deviations of various drilling parameters, as well as an unusual combination of their values. To build such a model, the *Isolation Forest* method was used with the *n_estimators=500* parameter. This method is taken from the *sklearn* open source library and consists in building a random binary decision tree that can recognize anomalies of various types: both isolated points with a low local density and clusters of small anomalies.

The results of the trained model on the test case can be seen in Figure 2. The end of the example (right border of the graph) means the beginning of the development of a complication of the type of gas, oil, and water show. The graph shows that the method has highlighted the anomalous behavior of the parameters towards the end of the example. The disadvantage of this approach is that the method determines not only the approach to a complication, but also other possible deviations of parameters, such as sensor failure, possible other complications, abnormal control, etc.

Prediction model of indicator function

This model is based on the introduction of an indicator function that has a zero value on a time interval that is sufficiently distant from the pre-emergency situation, and increases as it approaches (Figure 3). When training the model, the indicator function was set in the form of a sigmoidal function, taking a value equal 0.5 7 minutes before the accident and close to 1 at the point of the start of the accident. The model starts signaling an impending complication when the specified threshold is exceeded, calculated from the training sample of examples.
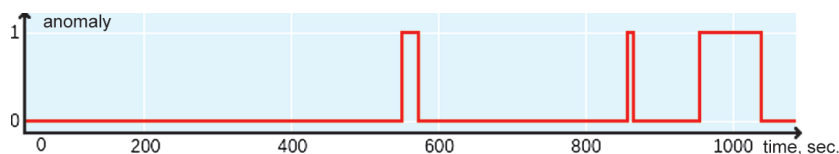

Fig. 2. Highlighted anomalies for a test case with a complication type of "gas, oil, and water show"
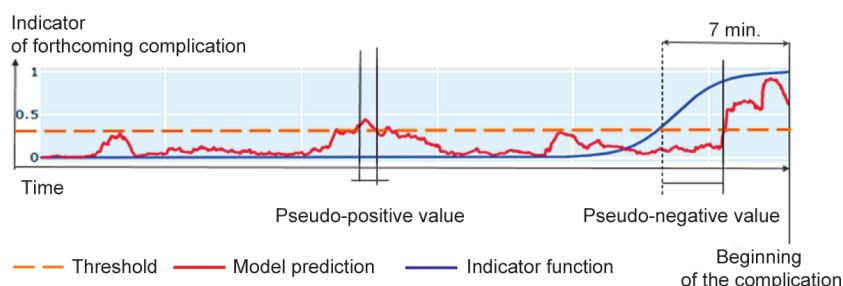

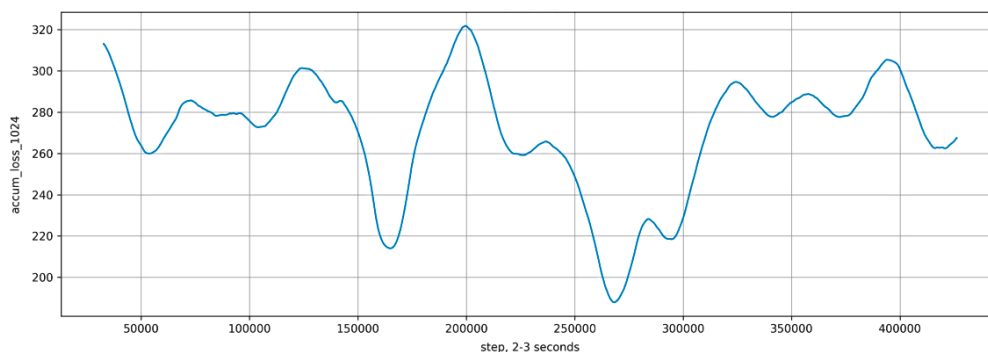*Fig. 3. Regression prediction of the indicator function*

Fig. 4. Error in predicting parameters in a well without complications that did not participate in the model training process
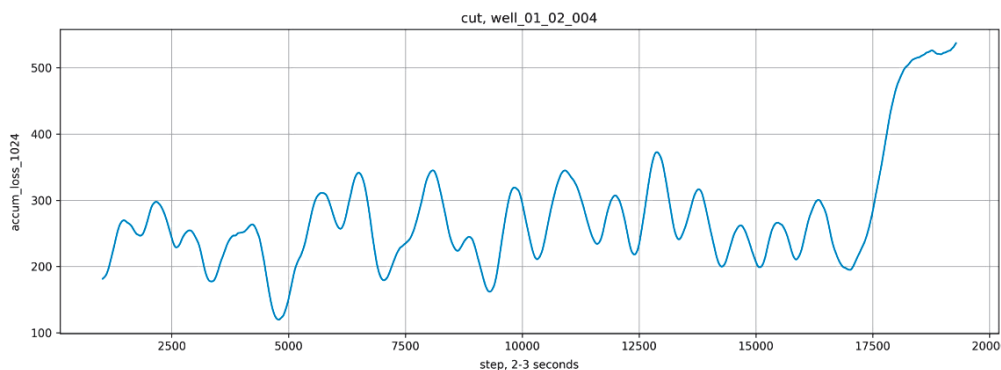

Fig. 5. Well with complication of the "Drill string sticking" type
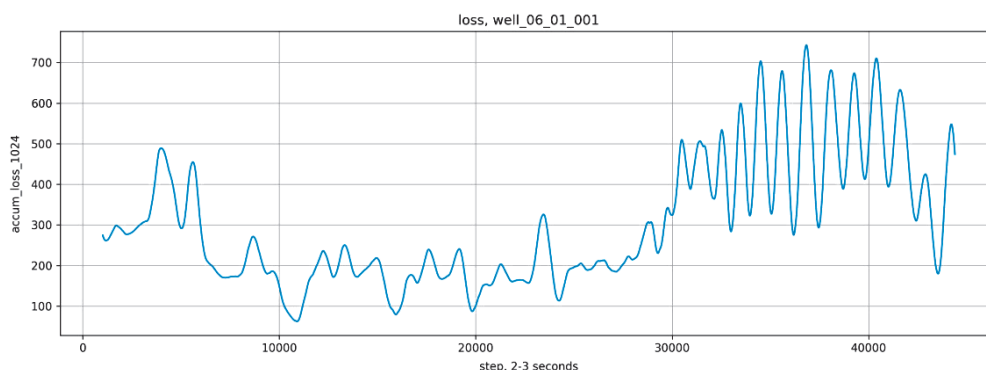

Fig. 6. Well with "Mud loss" complication

On average, the model begins to exceed the threshold 4–5 minutes before a complication event occurs.

The main algorithm used was a *Random forest* method with 100 trees. A *Random forest* is an ensemble regression method that employs a series of regression trees for different randomly selected subsamples of a dataset (parameters) and uses averaging to improve prediction accuracy and overtraining control.

Neural network model for real data

To experiment with real-world drilling data, it is not possible to use supervised learning approaches. Such approaches require a sufficient set of labeled data that is not represented in the given data source. Therefore, it was decided to use an approach similar to that described in the "anomaly detection" approach and solve a similar problem, evaluating the work of the developed method on the existing marked examples of complications such as "Sticking" and "Mud loss". Since there is a lot of unlabeled data in the specified source, an autoregressive approach was tested using convolutional, recurrent and fully connected layers of a neural network.

To process real data and feed them into a neural network model, an overlapping sliding window approach is used (Dmitrievsky et al., 2020). The model uses 5 selected parameters (Table 2) for which a window of 1024 consecutive values is formed. The resulting matrix is fed to the input of the model, which is trained to predict five parameters in the next step. To search for anomalous dynamics of parameters, the difference between the predicted values of the parameters and the observed ones was estimated. After training on 20 trouble-free wells, the model was run on three trouble-free wells, and two examples, one of which was "Mud loss" and the other was "Sticking". The results showed that in the absence of complications, the total error of

the predicted values does not exceed 500 (for example, Figure 4). However, when approaching the "Sticking" complication and the beginning of "Mud loss" – this error started to grow. In Figures 5–6 the last point (the right border of the graph) means the beginning of the pre-emergency situation development. It can be seen that in the example with a "Sticking" complication (Figure 5), more than 5000 seconds (one and a half hours), the error begins to increase and does not go down to the very beginning of the "Sticking" complication. In the example with the complication "Mud loss" (Figure 6), more than 6 hours, large fluctuations in the prediction error begin, at which it exceeds the threshold value.

4. Conclusion

Within the framework of the system being created to predict the main types of complications in the drilling process a number of modern hybrid methods of data mining (Data Driven methods) were developed and tested. Those methods included machine learning technologies and Data Driven neural networks, which have demonstrated their effectiveness on the available small amounts of simulation and real data of well drilling (Eremin et al., 2020a-2020e). Further work will be aimed at expanding data sets from geological and technological measurement stations, accordingly assessing the accuracy of the proposed models and their refinement.

References

Arkhipov A.I., Dmitrievsky A.N., Eremin N.A. et al. (2020). Data quality analysis of the station of geological and technological researches in recognizing losses and kicks to improve the prediction accuracy of neural network algorithms. *Neftyanoe khozyaystvo = Oil industry*, 8, pp. 63–67. DOI: 10.24887/0028-2448-2020-8-63-67

Borozdin, S., Dmitrievsky, A., Eremin, N., Arkhipov, A., Sboev, A., Chashchina-Semenova, O., Fitzner L., Safarova, E. (2020). Drilling Problems Forecast System Based on Neural Network. *Society of Petroleum Engineer*s. https://doi.org/10.2118/202546-MS

Chen T., Guestrin C. (2016). Xgboost: A scalable tree boosting system. *Proc. 22nd ASM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*. ACM, pp. 785–794. https://doi.org/10.1145/2939672.2939785

Diakonov A.G., Golovina A.M. (2017). Detection of anomalies in the work of mechanisms by machine learning methods. *Proc. XIX Int. Conf.:*

Analytics and data management in areas with intensive data use. Moscow, pp. 469–476. (In Russ.)

Dmitrievsky A.N., Eremin N.A., Filippova D.S.et al. (2020). Qualitative Analysis of Time Series GeoData to Prevent Complications and Emergencies During Drilling of Oil and Gas Wells. *Socar Proceedings,* 3, pp. 31–37. http://dx.doi.org/10.5510/OGP20200300442

Dmitrievsky A.N., Eremin N.A. Filippova D.S., Safarova E.A. (2020). Digital oil and gas complex of Russia. *Georesursy = Georesources, Special issue*, pp. 32–35. DOI: https://doi.org/10.18599/grs.2020.SI.32-35

Eremin N.A., Stolyarov V.E. (2020). On the digitalization of gas production processes at the late stages of field development. *Socar Proceedings*, 1, pp. 59–69. (In Russ.). DOI: 10.5510/ogp20200100424

Eremin N.A., Vodopiyan A.O., Duplyakin V.O., Chernikov A.D., Kosmos S.A. (2020a). Software component «Oil and Gas Blockchain». Software Registration Certificate RU 2020614626. Application form 2020613699.

Eremin N.A., Dmitrievsky A.N., Chashchina-Semenova O.K., Fitsner L.K., Chernikov A.D. (2020b). Software component «Adaptation of generalized neural network models for predicting complications and emergencies to geophysical parameters when drilling a specific well». Software Registration Certificate RU 2020660890. Application form 2020660179.

Eremin N.A., Dmitrievsky A.N., Chashchina-Semenova O.K., Fitsner L.K., Chernikov A.D. (2020c). Software component «Orchestration - integration of modules of the system for predicting complications and emergencies during drilling and well construction». Software Registration Certificate RU 2020660891. Application form 2020660181.

Eremin N.A., Dmitrievsky A.N., Chashchina-Semenova O. K., Fitsner L.K., Chernikov A.D. (2020d). Software component «Neural network calculations - construction of models for forecasting complications and emergencies during drilling and well construction». Software Registration Certificate RU 2020660892. Application form 2020660182.

Eremin N.A., Dmitrievsky A.N., Chashchina-Semenova O.K., Fitsner L.K., Chernikov A.D. (2020e). Software component «Indication of forecast of complications and emergencies during drilling and well construction» (SC «Indication»). Software Registration Certificate RU 2020661356. Application form 2020660450.

Gurina E. et al. (2020). Application of machine learning to accidents detection at directional drilling. *Journal of Petroleum Science and Engineering,* 184, 106519. https://doi.org/10.1016/j.petrol.2019.106519

Kanfar R. et al. (2020). Real-Time Well Log Prediction From Drilling Data Using Deep Learning. arXiv preprint arXiv: 2001.10156. https://doi.org/10.2523/IPTC-19693-MS

Kaznacheev P.F., Samoilova R.V., Kjurchinski N.V. (2016). Improving Efficiency of the Oil and Gas Sector and Other Extractive Industries by Applying Methods of Artificial Intelligence. *Ekonomicheskaya Politika = Economic Policy*, 11(5). (In Russ.)

Kodirov S.S., Shestakov A.L. (2019). The development of an artificial neural network for predicting stuck pipe columns of drill pipes. *Vestnik Yuzhno-Ural'skogo gosudarstvennogo universiteta. Seriya: Komp'yuternye tekhnologii, upravlenie, radioelektronika = Bulletin of the South Ural State University. Series Computer Technology, Aotimatic Control, Radio Electronics*, 19(3). (In Russ.)

Kohonen T. (1990). The self-organizing map. *Proc. IEEE,* 78(9), pp. 1464–1480. https://doi.org/10.1109/5.58325

Li Y. et al. (2019). Deep learning for well data history analysis. *Society of Petroleum Engineers.* https://doi.org/10.2118/196011-MS

Liu F.T., Tony T.K.M., Zhou Z.H. (2008). Isolation forest. *Proc. 8th IEEE Int. Conf. on Data Mining,* pp. 413–422. https://doi.org/10.1109/ICDM.2008.17

Noshi C.I., Schubert J.J. (2018). The Role of Machine Learning in Drilling Operations. A Review. *Society of Petroleum Engineers*. https://doi.org/10.2118/191823-18ERM-MS

Pichugin O.N., Prokofiev Y.Z., Alexandrov D.M. (2013). Decision Trees as an effective method for analysis and forecasting. *Neftepromyslovoe delo,* 11, pp. 69–75. (In Russ.)

About the Authors

Anatoliy N. Dmitrievsky – Dr. Sci. (Engineering), Professor, member of the Russian Academy of Sciences, Research advisor, Oil and Gas Research Institute of the Russian Academy of Sciences

3, Gubkin st., Moscow, 119333, Russian Federation

Alexander G. Sboev – Leading Researcher, Cand. Sci. (Engineering), National Research Center «Kurchatov Institute»

1, Ak. Kurchatov pl., Moscow, 123098, Russian Federation

Nikolai A. Eremin – Dr. Sci. (Engineering), Professor, Deputy Director for Innovation, Oil and Gas Research Institute of the Russian Academy of Sciences; Professor, National University of Oil and Gas «Gubkin University»

3, Gubkin st., Moscow, 119333, Russian Federation

Alexander D. Chernikov – Leading Researcher, Cand. Sci. (Engineering), Oil and Gas Research Institute of the Russian Academy of Sciences

3, Gubkin st., Moscow, 119333, Russian Federation

Aleksandr V. Naumov – Engineer, Oil and Gas Research Institute of the Russian Academy of Sciences

3, Gubkin st., Moscow, 119333, Russian Federation

Artem V. Gryaznov – Engineer, Oil and Gas Research Institute of the Russian Academy of Sciences

3, Gubkin st., Moscow, 119333, Russian Federation

Ivan A. Moloshnikov – Engineer, Oil and Gas Research Institute of the Russian Academy of Sciences

3, Gubkin st., Moscow, 119333, Russian Federation

Sergei O. Borozdin – Senior Lecturer, Department of Oil and Gas Wells Drilling, National University of Oil and Gas «Gubkin University»

65, Leninsky ave, Moscow, 119991, Russian Federation

Elizaveta A. Safarova – Junior Researcher, Oil and Gas Research Institute of the Russian Academy of Sciences

3, Gubkin st., Moscow, 119333, Russian Federation