Article Risk Factors Evaluation for Monitoring of Well Drilling

Shamil Islamov ¹, Alexey Grigoriev ², Ilya Beloglazov ³, *, Sergey Savchenkov ⁴ and Ove T. Gudmestad ⁵

- 1 Department of Development and Operation of Oil and Gas Fields, Saint Petersburg Mining University, 199106 Saint Petersburg, Russia; <u>Islamov ShR@pers.spmi.ru</u>
- 2 Geological Drilling Support Department, JSC RN-Nyaganneftegaz Rosneft, 628183 Nyagan, Khanty-Mansi Autonomous Area, Russia; <u>ASGrigoryev@nng.rosneft.ru</u>
- ³ The Automation of Technological Processes and Production Department, Saint Petersburg Mining University, 199106 Saint Petersburg, Russia; <u>Beloglazov II@pers.spmi.ru</u>
- ⁴ Patent and Licensing Department, The Automation of Technological Processes and Production Department, Saint Petersburg Mining University, 199106 Saint Petersburg, Russia; <u>Savchenkov SA@pers.spmi.ru</u>
- ⁵ Faculty of Science and Technology, University of Stavanger, N-4036 Stavanger, Norway; <u>ove.t.gud-mestad@uis.no</u>
- * Correspondence: Beloglazov_II@pers.spmi.ru;

Abstract: Drilling of wells for oil and gas production is a highly complex and expensive part of reservoir development. Thus, together with injury prevention, there is a goal to save cost expenditures on downtime and repair of drilling equipment. Nowadays companies have begun to look for ways to improve the efficiency of drilling and minimize non-production time with the help of new technologies. To support decisions in a narrow time frame, it is valuable to have an early warning system. Such a decision support system will help an engineer to intervene in the drilling process and prevent high expenses of unproductive time and equipment repair due to a problem. This work is describing a comparison of machine learning algorithms for anomaly detection during well drilling. Tested models classify drilling problems based on historical data from previously drilled wells. To validate anomaly detection algorithms, we use historical logs of drilling problems for 67 wells at a large brownfield in Siberia, Russia. Wells with problems were selected and analyzed. It should be noted that out of the 67 wells, 20 wells were drilled without expenses for unproductive time. Experiential results illustrated that a model based on *gradient boosting* can classify the complications in the drilling process best of all.

Keywords: machine learning; drilling problems; artificial intelligence; risk factors evaluation; gradient boosting

1. Introduction

Today, the use of machine learning (ML) capabilities in the oil and gas industry is becoming a central topic in various research centers and universities in the modern world. ML algorithms can provide practical solutions for analyzing and leveraging big historical data. ML technology has long been successfully used in computer science, engineering, mathematics, physics and astronomy, neuroscience, and medicine [1-10].

However, for the oil and gas industry, the use of such technologies has significantly increased in recent years [11-18]. An important task of the development of the oil and gas industry in the coming years is to increase the efficiency of producing oil and gas and drilling of a well and the main impetus to the introduction of methods of ML was the fall in oil prices. Oil and gas companies have concentrated on resource efficiency, optimizing their production processes [19-30].

This challenge should be solved at the expense of the all-around development of fundamental and applied research and the rapid introduction of the results obtained. In drilling, one of the main issues in improving the quality of well construction is a reduction in the number and severity of problems, which is closely related to the use of modern computer-mathematical methods and computer technology.

(C) (D)

It is worth noting that the use of high-performance data analysis software is not a novelty for the oil industry. Since the 1990s, technologies for the collection and analysis of well data have been widely used. However, large capital expenditures on the implementation of these tools scared off many companies since their implementation could not be financially justified.

Currently one of the main challenges facing the oil and gas industry is to improve the efficiency of well drilling.

The requirements of the practice of drilling deep oil and gas wells require a wide range of requirements for the theory of machine learning. In this case, the theory should be defined as a normal process at the time of origin, and during development, considering any problem as an integral part of the drilling processes. It is desirable that a theoretical description of drilling problems (DP) allows judging them not only at a qualitative level, but also quantify the interrelation of their essential variables. Several years ago, these tasks seemed laborious.

At the moment the existing works, which were aimed at improving the drilling process using methods of artificial intelligence (AI).

Zhan and others [31] in their work used a nonparametric system of fuzzy inferences to predict the state of the rotary steerable system (RSS) by forecasting the state of the RSS is in real time based on the operating mode and drilling parameters. This method is allows reducing the cost of repair and maintenance of the drilling equipment.

Wang [32] presented an approach that uses multilayer neural network modeling to predict nonlinear optimization of DP. The proposed model can not only predict the pump pressure, as the desired parameter but also can ensure the impact of each input parameter in this model.

The mechanism of damage to drilling equipment is usually accompanied by several successive incidents that contribute to the loss of efficiency. Consequently, recognition, classification, elimination of breakdowns, and calculation of the remaining useful life are impossible without constant monitoring of the health of the system. So Camci and others [33] with the help of the hidden Markov model created a model capable of monitoring the current state of the mechanism, through signals sent by sensors. In particular, this model has shown excellent results for diagnosing the condition of drill bits.

At present, methods of programming neural networks for solving problems in various fields have been widely used. An artificial neural network is an interconnected group of nodes, similar to our brain system [34]. For example, Lind and Kabirova [35] used the neural programming method to predict possible problems that may arise when drilling wells, based on information about the oil field reserves. The results obtained showed the effectiveness of the neural network application for solving this problem.

The Bayesian neural network was used in work by Al-yami and Schubert [36]. The method used allowed to create a system for making expert decisions in drilling. This method can be used to train young engineers. The system can also provide advice during all stages of well construction. It can be a well completion, monitoring of drilling and cementing of wells, selection of drilling fluids, etc.

One of the most important parameters in the drilling process, the rate of penetration (ROP) is getting a lot of attention. Dedicated to many works aimed at predicting the rate of penetration, as this parameter affects the optimization of many processes during drilling, Jahanbakhshi and others used a neural network to predict ROP [37]. Type of rock, mechanical properties of the formation, hydraulics, type of bit and its features and rotor speed were chosen as input parameters. Monazami and others [38] in their article also used a neural network to estimate ROP. The authors consider this method as the most useful tool in forecasting in comparison with the currently available procedures. The model allows the drilling crew to assess the ROP not only at the planning stage but also during drilling. The results of this work have shown that neural programming for the quality of ROP prediction is superior to conventional methods. Amer and others [39] used the method of back propagation to predict the ROP, which showed its success in their work.

Gidh and others [40] also used an artificial neural network to develop a program to optimize drilling parameters. The result of this work was a model capable of choosing the optimal ROP and weight on bit to extend the life of the bit. This model selects the necessary drilling parameters based on the expected characteristics of the rock in which the drilling will take place. Further, all parameters were adjusted for the relevant conditions.

In another publication, the ROP, together with the specific mechanical energy of Rashidi and others [41], were used to calculate the bit wear in real time. Between the specific mechanical energy and the weight on bit, a linear relationship was obtained. Based on the analysis of a vast number of experiments, the authors believe that this model can become a valuable tool in the analysis of bit wear in real time.

Valisevich and others [42], using an artificial neural network, created a model optimizing the development of bits in real time. All this led to an increase in drilling speed, a decrease in bit wear during drilling.

Another application of the neural network was found by Dashevskiy and others [43]. This work allowed to simulate a nonlinear drilling system with the minimal error share by monitoring its dynamic behavior. The authors achieved the primary aim of the work – the use of neural networks for the intelligent control of drilling in dynamics.

GirirajKumar and others [44], for the improvement of drilling, suggest using the optimally tuned proportional-integral-differential (PID) controller employed in the highperformance drilling systems. The primary aim of the work was to obtain a stable, reliable and controlled system by tuning the PID controller, using the optimization algorithm for swarm intelligence. The results of the work showed that tuning the PID controller using RI (swarm intelligence) gives a smaller overshoot.

Using neural network, Lind and others created an algorithm for predicting the loss of drilling fluids [45]. This system allows you to receive a recommendation for the selection of drilling fluids.

Static training methods for predicting torque and friction in real time were applied by Hegde and others [46]. They considered algorithms such as regression, random forest, and the support vector method. These methods can be used to predict DP and take appropriate measures to eliminate them. For example, an unexpected change in the value of torque may be a sign of a complication.

Another common complication – the instability of the walls of the well with the help of a neural network was predicted by Okpo and others [47]. The program developed by the authors was used to predict the geomechanical parameters of the formation. The model was developed in a Neuroph Studio, and the platform of the neural network was Java and Netbeans IDE. The main advantage of the model is its simplicity and open-source code.

Unrau and others [48], using the ML method, improved the existing alarm system on the drilling rig. The standard alarm systems used for drilling can register too many false alarms that significantly affect the drilling process. The ML algorithm proposed by the authors can be used to reduce false alarms while maintaining the efficiency of the alarm system. The model successfully detects kicks and loss.

As noted above, the integration of AI methods in a drilling process has great practical importance.

A DP is a violation of the continuity of the technological process of the construction of a well, requiring for its liquidation the carrying out of special works not planned in the project. In the process of drilling of oil and gas wells due to the phenomena of geological nature, there are from time-to-time problems in the technological process. This could be loss of drilling mud and fluid, kick, stuck drill and casing columns [49].

The drilling crew constantly faces a lot of difficult situations, the exits from which can be very expensive, and even impossible. A drill string may be stuck, by pressing against the wall of the well during a draw-down or getting stuck as a result of key seating. To eliminate these problems, additional efforts will be required to free the drill string. Sometimes these efforts can fail. Then, drilling a side track is required [50].

Making a decision to eliminate these problems is a complex process. The damage from complications consists of the time spent for the elimination of DP, costs for materials and energy. To minimize the risks of drilling problems, work is being carried out to minimize vibrations of the bottom of the drilling assembly [51]; a mathematical model of the (screw downhole motors (SDM) - drilling string (DS) system is being developed, which allows predicting the range of DS self-oscillations and boundaries of rotational and translational wave disturbances for case of string modeling as a heterogeneous rod at drilling directionally straight sections of a well [52]. Thus, preventing problems and accordingly minimizing the risks of their occurrence is an actual problem today.

The aim of this work is to finding a learning algorithm to recognize and classify DP while drilling wells. Of the eight methods of ML, gradient boosting (GB) was chosen. This algorithm showed high performance precision, recall and F-score, see below. This learning algorithm, based on historical data from previously drilled wells, classifies the DP best of all. Such a decision support system will help the engineer to intervene in the drilling process and prevent high expenses due to unproductive time and equipment repair. Also, worth noting is another significant plus. An algorithm in addition to the classification of DP accurately determines the standard drilling mode. This minimizes the possibility of triggering false alarms, which will also save the drilling time. False alarms are also one of the problems when drilling wells, which takes up a significant amount of time and money.

2. Materials and Methods

In order to create a program that classifies the problems in the drilling process, the main methods of ML with which the calculation will be performed have been considered. These methods have shown successful applicability for solving problems in various industries.

Logistic regression

The method of logistic regression is based on fairly strong probabilistic assumptions, which have several interesting consequences. First, the linear classification algorithm turns out to be the optimal Bayesian classifier. Secondly, the form of the activation function (it is the sigmoid function) and the loss function are uniquely determined. Thirdly, an interesting additional possibility arises, along with the classification of the object, to obtain numerical estimates of the probability of problems belonging to each of the classes [53].

Naive Bayesian classifier

A naive Bayesian classifier is a simple probabilistic classifier based on the application of Bayes' theorem with strict (naive) assumptions about independence.

Depending on the exact nature of the probabilistic model, naive Bayesian classifiers can be trained very effectively. In many practical applications, the maximum likelihood method is used to estimate the parameters for naive Bayesian models. In other words, one can work with a naive Bayesian model, not believing in Bayesian probability and not using Bayesian methods [54].

Method k-nearest neighbors

The method of k-nearest neighbors is a metric algorithm for automatic classification of objects. The main principle of the method of the nearest neighbors is that the object is assigned to the class that is the most common among the neighbors of this element.

Neighbors are taken based on a set of objects whose classes are already known, and, based on the key value for this method, the value of k is calculated, to find which class is the most numerous among them. Each object has a finite number of attributes.

It is assumed that there is a certain set of objects with the already existing classification [55].

Decision tree

The decision tree is a decision support tool used in statistics and data analysis for predictive models. The structure of the tree consists of "leaves" and "branches". On the edges ("branches") of the decision tree, attributes are recorded, on which the objective function depends, in the "leaves" the values of the objective function are recorded, and in the remaining nodes - the attributes for which the cases differ. To classify a new case, you have to go down the tree to the sheet and give the corresponding value. Similar decision trees are widely used in intellectual data analysis. The goal is to create a model that predicts the value of the target variable, based on several variables at the input.

Support vector machine

The support vector method is a set of similar algorithms of the form "learning with the teacher", used for classification problems and regression analysis. This method belongs to the family of linear classifiers. A special property of the support vector method is a continuous decrease in the empirical classification error and an increase in the gap. Therefore, this method is also known as the classifier method with the maximum gap [56].

Random forest

A Random forest is a set of decision trees. In the regression problem, their answers are averaged, in the classification problem, a decision is made by voting on the majority.

The method is based on the construction of a large number (assembly) of decision trees, each of which is constructed from a sample obtained from the initial training sample using a sample with a return [57].

Gradient boosting

Boosting is a procedure for the sequential construction of a composition of ML algorithms, where each subsequent algorithm seeks to compensate for the shortcomings of the composition of all previous algorithms. Boosting is a «greedy» algorithm for composing the final algorithms (Figure 1).

Boosting over decision trees is considered one of the most effective methods in terms of quality classification. In many experiments, there was an almost unlimited reduction in the error rate on an independent test sample, as the composition was increased. Moreover, the quality of the test sample often continued to improve even after achieving an unmistakable recognition of the entire training sample. This overturned the ideas that existed for a sufficiently long time that it is necessary to limit the complexity of the algorithms in order to increase the generalizing ability [58-59].



Figure 1. Example of gradient boosting.

Neural network

An artificial neural network is a mathematical model, as well as its software or hardware implementation, built on the principle of the organization and functioning of biological neural networks – the nerve cell networks of a living organism. This concept arose when studying the processes occurring in the brain, and when trying to simulate these processes. The first such attempt was the neural networks of McCulloch and Pitts [60]. After the development of learning algorithms, the resulting models began to be used for practical purposes: in forecasting problems, for pattern recognition, in control tasks, etc.

A neural network is a system capable of changing its structure under the influence of external factors. An artificial network is trained on input data. During the training, the internal parameters of the artificial neural network are adjusted to the input data, which makes it possible to isolate patterns in the data or to solve problems of prediction, classification, and clustering. When using an artificial neural network for data analysis, the researcher solves several problems: what learning algorithm to use, what is the network configuration. The required internal parameters are found automatically, according to the chosen algorithm and configuration [60].

Evaluation of the quality of machine learning methods

Metrics are used to evaluate model quality and compare algorithms. Before moving to the metrics, we need to introduce an important concept for describing these metrics in terms of classification errors – the confusion matrix.

Having two classes and an algorithm that predicts the belonging of each object to one of the classes, then the classification error matrix will look like in Table 2.

	y=1	y=0
y′=1	True Positive (TP)	False Positive (FP)
y'=0	False Negative (FN)	True Negative (TN)

Table 2. Metrics by model.

In Table 2, "y'" is the answer of the algorithm on the object, and "y" is the true label of the class on this object.

Thus, classification errors are of two types: False Negative (FN) and False Positive (FP).

Precision, recall and F-score

Recall demonstrates the ability of the algorithm to detect a given class and precision demonstrates the ability to distinguish this class from other classes.

To assess the quality of the models used to classify the complications in the drilling process, the widely used precision, recall and F-score metrics were used.

$$precision = \frac{TP}{TP + FP'},\tag{1}$$

$$recall = \frac{TP}{TP + FN},$$
(2)

where TP – positive observation, and expected to be positive; FN – observation is positive, but it is predicted negatively; FP – observation is negative, but predicted positively.

There are several different ways to combine precision and recall in an aggregated quality criterion. The F-score is an average harmonic precision and recall:

$$F_{\beta} = (1 + \beta^2) \cdot \frac{\text{precision} \cdot \text{recall}}{(\beta^2 \cdot \text{precision}) + \text{recall}}$$
(3)

Where β – in this case determines the weight of accuracy in the metric, and for β = 1 this is the average harmonic (with a factor of 2, so that in the case of precision = 1 and recall = 1, we have F_{β} = 1); the F-score reaches a maximum for completeness and accuracy of one, and is close to zero if one of the arguments is close to zero.

The sklearn library in Python has a convenient function_metrics, classification_report, which returns recall, precision, and F-score for each of the classes, as well as the number of instances of each class [61].

Given data

As initial data, reports on drilling 67 wells were provided. Many of the wells have had DP that have led to rig downtime and loss of productive drilling time. The analysis of the total time spent on drilling all wells showed that about 10.33% of this time was unproductive operating time.

It is worth noting that 10.33% is an important value, considering that the average cost per hour of drilling varies from 15 to 55 thousand rubles. And in this data base, there is a well in which the unproductive time was 50% of the total operating time.

The main causes of unproductive drilling time at one specific field are shown in Figure 2. The greatest losses of time were due to rig downtime in waiting for contractors and equipment. Then, there is unproductive time due to the liquidation of penalty (unscheduled work, redrilling due to the fault of the contractor, etc.). In this project, we are interested in the trouble (DP) that arise during the drilling process. These include kick, loss of circulation, borehole instability, etc.



Figure 2. Distribution of non-productive time on field N.

Wells with problems were identified and analyzed. It should be noted that out of 67 wells, 20 wells were drilled without expenses for unproductive time. The most common problem is related to the seize that occurs when casing runs down. It is worth noting that in this project, calculations will be made for complications arising directly during drilling. For three wells, trouble arose during drilling and detailed records are available.

For further analysis, all drilling parameters that were recorded for each well were considered. The analysis of the data showed that not all the wells from the sample have the same number of corresponding recorded drilling parameters. Some wells recorded the minimum number of parameters. We would like to note that drilling reports were provided for 67 wells, but the files with the recorded drilling parameters were given for 78 wells. Therefore, data representing 78 wells were analyzed. For a wide analysis of the drilling parameters recorded on the wells, reports from 78 wells were taken into account. It can be seen from Figure 3 that only eight parameters are the most commonly reported for all wells; these are highlighted in red. And these parameters will be used as input parameters for classification of complications.



Figure 3. Drilling parameters analyzed for wells.

After the work was done, for the three wells in which the DP were plotted, the recorded drilling parameters were plotted. The graphs are constructed using the Python programming language, Figures 4 to 6.

Well 1. The DP is associated with borehole instability due to technical water entering at a depth of 2882 m (Figure 4). It can be noted that this problem is accompanied by steep changes in drilling parameters. In particular, the value of the hook load, rotary table torque, etc., steeply increases. To eliminate this problem, the drilling crew spent 231 hours.

Well 2. During well drilling in the interval 239-263 m, the drilling fluid was lost at a volume of 40 m³ (Figure 5). 7.1 hours of unproductive time was spent on solve this problem. It is worth noting that the graph clearly shows that during the loss, circulation significantly decreased the level of the fluid capacity to mud tank N^o 2. A mud tank is an open-top container, typically made of square steel tube and steel plate, to store drilling fluid on a drilling rig. They are also called mud pits, as they were once simple pits in the ground

Well 3. When drilling to 2493 m, the drilling fluid was lost. The total loss was 55 m³ (Figure 6). To eliminate the complication, colmatage fluid was injected. The total time taken to combat the DP was 27.9 hours. This well is one of those that did not record the complete list of required drilling parameters.



Figure 4. Drilling Problem at well 1 – borehole instability, drilling parameters versus time.



Figure 5. Drilling Problem at well 2 – circulation loss, drilling parameters versus time.



Figure 6. Drilling Problem at well 3 – circulation loss, drilling parameters versus time.

According to the algorithms of machine training given in the previous chapter, calculations were made to classify (forecast) the problems in the drilling process.

For calculations, the Python programming language and Scikit-Learn library was used. The percentage of training and test sample among data was set as 65/35%, respectively. The training sample is a sample based on which the chosen algorithm adjusts the dependency model. The test sample is the sample by which the accuracy of the model used is checked. The following drilling parameters were used as input parameters:

- Standpipe pressure;
- Tank level 02;
- Input flow rate;
- Hook load;
- Rotary table torque;
- Rate of penetration;
- Weight on bit;
- Gas content.

As a result of the calculations, the following metrics were obtained, for the subsequent detection of the most accurate model.

Table 2 shows that the following algorithms of the Machine Learning (ML) have the highest value of metric: Decision tree; Random forest; Gradient boosting (GB).

Next, we considered the number of correct and incorrect assumptions in the calculation of algorithms. Table 3 presents the case for situations where there are no problems while drilling, and in Table 4 the classification of problems while drilling. The goal is to see how the algorithm can misclassify the drilling process. "Right" is the number of correctly predicted values; "False" is the number of misplaced predictions when drilling process without problem recognized how drilling with problem. From the data presented, it can be seen that the greatest number of correct and accurate classifications of situations is obtained using the ML method Gradient Boosting (GB). GB allowed, with a minimum of errors, to classify the complication from the available data set.

Algorithm	Metrics (determination of drilling problems)		
	Precision	Recall	F-score
Logistic regression	0.00	0.00	0.00
Naive Bayesian classifier	0.03	1.00	0.06
Method of k-nearest neighbors	0.83	0.64	0.73
Decision tree	0.97	0.87	0.92
Support vector method	0.00	0.00	0.00
Random Forest	0.98	0.93	0.95
Gradient boosting	1.00	0.93	0.97
Neural network	1.00	0.53	0.70

Table 2. Metrics by model.

Table 3. Accuracy of prediction of a normal situation.

Algorithm	Situation	Right	False
Logistic regression	Normal	3916	1
Naive Bayesian classifier	Normal	2484	1433
Method of k-nearest neighbors	Normal	3911	6
Decision tree	Normal	3916	1
Support vector method	Normal	3917	0
Random Forest	Normal	3915	2
Gradient boosting	Normal	3917	0
Neural network	Normal	3917	0

Table 4. Accuracy of prediction of a problem situation.

Algorithm	Situation	Right	False
Logistic regression	Problem	0	45
Naive Bayesian classifier	Problem	45	0
Method of k-nearest neighbors	Problem	29	16
Decision tree	Problem	39	6
Support vector method	Problem	0	45
Random Forest	Problem	39	6
Gradient boosting	Problem	42	3
Neural network	Problem	27	18

Then, a sensitivity analysis was performed (Figure 7), when the drilling parameters were removed from the gradient boosting in turn by their weight coefficients from the smaller to the larger. This allowed to understand how many parameters at the input are needed in this situation for the correct operation of the gradient boosting. It has been established that when the parameters such as "Gas content", "Weight on bit" and "Rate of penetration" are removed from the model, the system classifies the drilling problems with the same accuracy. Accordingly, it can be concluded that this algorithm in the event of an emergency situation, can classify drilling problems according to the available 5 parameters without loss of accuracy: "Rotary table torque", "Standpipe pressure", "Hook load", "Tank level 02", and "Input flow rate".



Figure 7. Feature importance for gradient boosting.

4. Discussion

Based on the results of the work, a model is obtained capable of recognizing and classifying complications in the process of drilling wells. Such decision support system will help in time to assist with the drilling process and prevent high expenses for rig downtime and equipment repair. The program will signal a possible problem. It is worth noting that in this work, we used not too much initial data. Therefore, it is recommended to increase the efficiency of the model, to test it on more initial data.

Also, worth noting is another significant plus. An algorithm in addition to the classification of DP accurately determines the standard drilling mode (without problem). This minimizes the possibility of triggering false alarms, which will also save drilling time. False alarms are also one of the problems when drilling wells, which takes up a significant amount of time and money. And if new technologies are introduced by companies in oil and gas production, this will allow businesses to save their costs. For example, in the construction of a drilling rig that reaches hundreds of millions of dollars, even a 5% reduction in planning time can have a significant positive impact on the company's profits [62].

Nybø [63-64] solved a similar problem. In this work, a hybrid system is developed that includes a physical model and AI. Together, they allow one to recognize the problems when drilling much better than individually. Also in this paper, the problem of a small number of studies on the introduction of methods of ML in the drilling sector is addressed. The authors of the work are also convinced that this integration of machines and people will significantly increase the efficiency of drilling wells.

Based on the results of the analysis using eight algorithms, it can be seen that the logistic regression, support vector method shows metrics equal to zero for the recognition of complications. Perhaps these values are associated with a small number of initial data

of complications. And so, these algorithms show such poor results. As noted above for further work, it is recommended to experiment with a much larger number of initial data.

5. Conclusions

Avoiding the problems in the drilling process, its classification and timely elimination remain an urgent problem to date. The aim of this work was to create a program capable of recognizing and classifying Drilling Problems (DP). Following the results of the work the following achievements were made:

1. Based on the literature review, a wide application of AI in drilling is shown. From the creation of training programs to the prediction of the rate of penetration.

2. During the analysis of the initial data, wells with problems that were encountered during drilling were identified. To model the presented DP, a computer model was set up.

3. During the analysis of the drilling reports, a list of the main parameters was compiled, which will participate as input for the model: Standpipe pressure; Tank level; Input flow rate; Hook load; Rotary table torque; Rate of penetration; Weight on bit; Gas content.

4. Of the eight methods of Machine Learning (ML), the GB method was chosen. This algorithm showed high performance precision, recall and F-score.

5. For GB method, the parameters that make the greatest contribution to the operation of the algorithm are established using the feature importations parameter. These are the Rotary table torque, Standpipe pressure, Hook load.

6. During the GB analysis, it is established that in the case of removing parameters such as Gas content, the model continues to work without changing the accuracy of the classification of the DP.

Author Contributions: Conceptualization, S.G. and I.B.; formal analysis, S.S.; investigation, S.G.; methodology, Sh.I.; project administration, S.G. and Sh.I.; resources, V.M. and O.T.G.; software, S.G.; supervision, O.T.G.; validation, S.G.; visualization, S.S.; writing – original draft, Sh.I.; writing – review and editing, I.B. and O.T.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to its storage in private networks.

Acknowledgments: The authors thank Saint Petersburg Mining University for enabling the laboratory experiments. The investigations were carried out using the equipment of the Educational Research Center for Digital Technologies of Saint Petersburg Mining University.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

AI	Artificial intelligence
DP	Drilling problems
GB	Gradient boosting
ML	Machine learning
PID	Proportional-integral-differential
ROP	Process rate of penetration
RSS	Rotary steerable system

References

1. Hochreiter S., Schmidhuber J. Long Short-Term Memory. Neural Computation. 1997. 9(8):1735–1780.

- Jones DT. Protein Secondary Structure Prediction Based on Position-Specific Scoring Matrices. Journal of Molecular biology. 1999. 292(2):195–202.
- 3. LeCun Y., Bengio Y., Hinton G. Deep learning. Nature. 2015. 521(7553):436-444.
- 4. Milo R., Shen-Orr SS., Itzkovitz S., Kashtan N., Chklovskii DM., Alon U. Network Motifs: Simple Building Blocks of Complex Networks. Science. 2002. 298(5594):824–827.
- 5. Nielsen H., Engelbrecht J., Brunak S., Heijne G. Von. Identification of Prokaryotic and Eukaryotic Signal Peptides and Prediction of Their Cleavage Sites. Protein Engineering Design and Selection. 1997. 10(1):1–6.
- 6. Olden JD, Jackson DA. Illuminating the "Black Box": a Randomization Approach for Understanding Variable Contributions in Artificial Neural Networks. Ecological Modelling. 2002. 154(1-2):135–150.
- 7. Reichstein M., Camps-Valls G., Stevens B., Jung M., Denzler J., Carvalhais N., Prabhat. Deep learning and process understanding for data-driven Earth system science. Nature. 2019. 566(7743):195–204.
- Rubinov M., Sporns O. Complex Network Measures of Brain Connectivity: Uses and Interpretations. NeuroImage. 2010. 52(3):1059–1069.
- 9. Tu JV. Advantages and Disadvantages of Using Artificial Neural Networks versus Logistic Regression for Predicting Medical Outcomes. Journal of Clinical Epidemiology. 1996. 49(11):1225–1231.
- 10. Voyant C., Notton G., Kalogirou S., Nivet M-L., Paoli C., Motte F., Fouilloy A. Machine Learning methods for solar radiation forecasting. A review. Renewable Engergy. 2017. 105:569–582.
- 11. Thales Luiz Pinheiro de Almeida, Bruno Andrey Fonseca Passos, Jéssica Lia Santos da Costa, André José Neves Andrade. Identifying clay mineral using angular competitive neural network: A machine learning application for porosity estimative. Journal of Petroleum Science and Engineering. Volume 200, May 2021, 108303.
- 12. Yasin Hajizadeh. Machine learning in oil and gas; a SWOT analysis approach. Journal of Petroleum Science and Engineering Vol.176. 2019. P. 661-663.
- 13. Khadijah M.Hanga, Yevgeniya Kovalchuk. Machine learning and multi-agent systems in oil and gas industry applications: A survey. Computer Science Review. Vol. 34. 2019. 100191.
- 14. Nima Mohamadian, Hamzeh Ghorbani, David A.Wood, Mohammad Mehrad, Shadfar Davoodi, Sina Rashidi, Alireza Soleimanian, Amirafzal Kiani Shahvand. A geomechanical approach to casing collapse prediction in oil and gas wells aided by machine learning. Journal of Petroleum Science and Engineering. Vol. 196. 2021. 107811.
- 15. Mohamed Layouni, Mohamed Salah Hamdi, Sofiène Tahar. Detection and sizing of metal-loss defects in oil and gas pipelines using pattern-adapted wavelets and machine learning. Applied Soft Computing. Vol. 52. 2017. P. 247-261.
- 16. Sina Rashidi, Mohammad Mehrad, Hamzeh Ghorbani, David A.Wood, Nima Mohamadian, Jamshid Moghadasi, Shadfar Davoodi. Determination of bubble point pressure & oil formation volume factor of crude oils applying multiple hidden layers extreme learning machine algorithms. Journal of Petroleum Science and Engineering. Vol. 202. 2021. 108425.
- 17. Hao Chen, Chao Zhang, Ninghong Jia, Ian Duncan, Shenglai Yang, Yong Zhi Yang. A machine learning model for predicting the minimum miscibility pressure of CO2 and crude oil system based on a support vector machine algorithm approach. Fuel. Vol. 290. 2021. 120048.
- 18. Boikov A.V., Savelev R.V., Payor V.A., Potapov, A.V. Evaluation of bulk material behavior control method in technological units using dem. Part 2. CIS Iron and Steel Review. Volume 20. 2020. P. 3-6.
- 19. Litvinenko V.S., Tsvetkov P.S., Molodtsov K.V. The social and market mechanism of sustainable development of public companies in the mineral resource sector. Eurasian Mining. Vol. 2020. Is. 1. 2020. P. 36-41.
- Kamatov, K. A. and Buslaev, G. V. (2015) 'Solutions for drilling efficiency improvement in extreme geological conditions of Timano-Pechora region', in Society of Petroleum Engineers - SPE Russian Petroleum Technology Conference. Society of Petroleum Engineers. doi: 10.2118/176530-ms..
- 21. Lanouar Charfeddine, Karim Barkat. Short- and long-run asymmetric effect of oil prices and oil and gas revenues on the real GDP and economic diversification in oil-dependent economy. Energy Economics. Vol.86. 2020. 104680.
- 22. Aleksandrova T., Aleksandrov A., Nikolaeva N. An Investigation of the Possibility of Extraction of Metals from Heavy Oil. Mineral Processing and Extractive Metallurgy Review. Vol. 38. Is.2. 2017. P. 92-95.
- 23. Nevskaya M. A., Seleznev S. G., Masloboev V. A., Klyuchnikova E. M., Makarov D. V. Environmental and business challenges presented by mining and mineral processing waste in the Russian federation. Minerals. No 9, Vol. 7. 2019. P. 445 460.
- 24. Liu, T.; Leusheva, E.; Morenov, V.; Li, L.; Jiang, G.; Fang, C.; Zhang, L.; Zheng, S.; Yu, Y. Influence of Polymer Reagents in the Drilling Fluids on the Efficiency of Deviated and Horizontal Wells Drilling. Energies 2020, 13, 4704.
- 25. Gang Hua, Zhaoqiang Xu, Guorong Wang, Bin Zeng, Yubing Liu, Ye Lei. Forecasting energy consumption of long-distance oil products pipeline based on improved fruit fly optimization algorithm and support vector regression. Energy. Vol. 224. 2021. 120153.
- 26. Yurak V.V., Dushin A.V., Mochalova L.A. Vs sustainable development: scenarios for the future. Journal of Mining Institute. 2020. Vol. 242. P.242 247.
- Kondrasheva N.K., Rudko V.A., Kondrashev D.O., Gabdulkhakov R.R., Derkunskii I.O., Konoplin R.R. Effect of Delayed Coking Pressure on the Yield and Quality of Middle and Heavy Distillates Used as Components of Environmentally Friendly Marine Fuels. Energy and Fuels. Vol. 33. Is.1. 2019. P. 636-644.
- 28. Kondrasheva N.K., Rudko V.A., Ancheyta J. Thermogravimetric Determination of the Kinetics of Petroleum Needle Coke Formation by Decantoil Thermolysis. ACS Omega. Vol.5. Is. 45. 2020. P. 29570-29576.

- 29. Seçkin Karasu, Aytaç Altan, Stelios Bekiros, Wasim Ahmad. A new forecasting model with wrapper-based feature selection approach using multi-objective optimization technique for chaotic crude oil time series. Energy. Vol.212. 2020. 118750.
- 30. Hebert, D.; Misiti, A. The Growing Role of Artificial Intelligence in Oil and Gas. 2016. Available online: https://insights.globalspec.com/article/2772/the-growing-role-of-artificial-intelligence-in-oil-and-gas (accessed on 23 April 2021).
- 31. Zhan, S., Rodiek, J., Heuermann-Kuehn, L.E., Baumann, J. Prognostics Health Management for a Directional Drilling System. Proceedings of Prognostics and System Health Management Conference, Shenzhen, China, 24-25 May 2011, pp. 1-7.
- 32. Wang, Y. Drilling Hydraulics Optimization Using Neural Networks. Master of Science, University of Louisiana at Lafayette, Louisiana, USA, 2014.
- Camci, F., Chinnam, R.B. Dynamic Bayesian Networks for Machine Diagnostics: Hierarchical Hidden Markov Models vs. Competitive Learning. Proceedings of International Joint Conference on Neural Networks, 2005, Montreal, Canada, 31 July-4 August 2005, pp. 1-6.
- 34. Yang, Z.R., Yang, Z. Artificial Neural Networks. Comprehensive Biomedical Physics, volume 6, 2014, pp. 1-17.
- 35. Lind, Yu.B., Kabirova, A.R. Artificial Neural Networks in Drilling Troubles Prediction. SPE Russian Oil and Gas Exploration & Production Technical Conference and Exhibition, Moscow, Russia, 14-16 October 2014, pp. 1-7.
- Al-yami, A.S.H., Schubert, J. Systems and methods for expert systems for well completion using Bayesian decision models (BDNs), drilling fluids types, and well types. 2015. Available online: https://hdl.handle.net/1969.1/177120 (accessed on 23 April 2021).
- 37. Jahanbakhshi, R., Keshavarzi, R., Jafarnezhad, A. Real-time Prediction of Rate of Penetration During Drilling Operation In Oil And Gas Wells. Rock Mechanics/Geomechanics Symposium, Chicago, USA, 24-27 June 2012, pp. 1-9.
- 38. Monazami, M., Hashemi, A., Shahbazian, M. Drilling rate of penetration prediction using artificial neural network: a case study of one of Iranian southern oil fields. *Journal of Oil and Gas Business* **2012**, *6*(*6*), 21-31.
- 39. Amer, M.M., Dahab, A.S., El-Sayed, A.H. An ROP Predictive Model in Nile Delta Area Using Artificial Neural Networks. SPE Kingdom of Saudi Arabia Annual Technical Symposium and Exhibition, Dammam, Saudi Arabia, 24-27 April 2017, pp. 1-11.
- 40. Gidh, Y., Purwanto, A., Bits, S. Artificial Neural Network Drilling Parameter Optimization System Improves ROP by Predicting/Managing Bit Wear. SPE Intelligent Energy International, Utrecht, Netherlands, 27-29 March 2012, pp. 1-13.
- 41. Rashidi, B., Hareland, G., Nygaard, R. Real-Time Drill Bit Wear Prediction by Combining Rock Energy and Drilling Strength Concepts. Abu Dhabi International Petroleum Exhibition and Conference, Abu Dhabi, UAE, 3-6 November 2008, pp. 1-9.
- 42. Valisevich, A., Ruzhnikov, A., Bebeshko, I., Moreno, R., Zhentichka, M., Bits, S. Drillbit Optimization System: Real-Time Approach to Enhance Rate of Penetration and Bit Wear Monitoring. SPE Russian Petroleum Technology Conference, Moscow, Russia, 26-28 October 2015, pp. 1-14.
- 43. Dashevskiy, D., Dubinsky, V., Macpherson, J.D. Application of Neural Networks for Predictive Control in Drilling Dynamics. SPE Annual Technical Conference and Exhibition, Houston, Texas, USA, 3-6 October 1999, pp. 1-9.
- 44. GirirajKumar, S.M., Jayaraj, D., Kishan, A.R. PSO Based Tuning of a PID Controller for a High-Performance Drilling Machine. *International Journal of Computer Applications* **2010**, *1*(19), 12-18.
- 45. Lind, Yu.B., Samsykin, A.V., Galeev, S.R. Information and Analytical System for Prevention of Drilling Fluid Loss. SPE Russian Petroleum Technology Conference, Moscow, Russia, 26-28 October 2015, pp. 1-12.
- 46. Hegde, C., Wallace, S., Gray, K. Real Time Prediction and Classification of Torque and Drag During Drilling Using Statistical Learning Methods. SPE Eastern Regional Meeting, Morgantown, West Virginia, USA, 13-15 October 2015, pp. 1-13.
- 47. Okpo, E.E., Dosunmu, A., Odagme, B.S. Artificial Neural Network Model for Predicting Wellbore Instability. SPE Nigeria Annual International Conference and Exhibition, Lagos, Nigeria, 2-4 August 2016, pp. 1-10.
- Unrau, S., Torrione, P., Hibbard, M., Smith, R., Olesen, L., Watson, J. Machine Learning Algorithms Applied to Detection of Well Control Events. SPE Kingdom of Saudi Arabia Annual Technical Symposium and Exhibition, Dammam, Saudi Arabia, 24-27 April 2017, pp. 1-10.
- 49. Shchepetov, O.A. System classification of failures in drilling. Vestnik Astrakhan State Technical University. Ser.: Management, Computer Sciences and Informatics, Astrakhan, 2009, 2, pp. 36-42.
- 50. Aldred, W., Plumb, D., Bradford, I., Cook, J., Gholkar, V., Cousins, L., Minton, R., Fuller, J., Goraya, Sh., Tucker, D. Managing drilling risk. *Oilfield review* **1999**, 2-19.
- 51. Dvoynikov, M.V. Research on technical and technological parameters of inclined drilling. J. Min. Inst. 2017, 223, 86-92.
- 52. Litvinenko, V.S., Dvoynikov, M.V. Methodology for determining the parameters of drilling mode for directional straight sections of well using screw downhole motors. J. Min. Inst. 2020, 41, 105-112.
- 53. Vorontsov, K.V. Lectures on Linear Classification Algorithms. Moscow Institute of Physics and Technology, Moscow, Russia, 2009, pp. 13-16.
- 54. Ray, S. 6 Easy Steps to Learn Naive Bayes Algorithm with codes in Python and R. Available online: https://www.analyt-icsvidhya.com/blog/2017/09/naive-bayes-explained/ (accessed on 23 April 2021).
- 55. Piryonesi, S.M., Tamer E. El-Diraby. Role of Data Analytics in Infrastructure Asset Management: Overcoming Data Size and Quality Problems. Journal of Transportation Engineering, Part B: Pavements, June 2020, 146 (2).
- Vorontsov, K.V. Lectures on the Support Vector Machine. Moscow Institute of Physics and Technology, Moscow, Russia, 2007.
 Chistyakov, S.P. Random forest. Proceedings of the Karelian Research Center of the Russian Academy of Sciences, 2013, 1, pp.
- Chistyakov, S.F. Kaldom forest. Proceedings of the Karenan Research Center of the Russian Academy of Sciences, 2015, 1, pp. 117-136.
 Vorontsov, K.V. Mathematical Methods of Learning by Precedents: A Course of Lectures. Moscow Institute of Physics and
- 58. Vorontsov, K.V. Mathematical Methods of Learning by Precedents: A Course of Lectures. Moscow Institute of Physics and Technology, Moscow, Russia, 2009.

- 59. Matthew, M. More Steps to Mastering Machine Learning with Python. Available online: http://www.kdnug-gets.com/2017/03/seven-more-steps-machine-learning-python.html (accessed on 23 April 2021).
- 60. McCulloch, U.S., Pitts, V. Logical Calculus of Ideas Relating to Nervous Activity, Foreign Literature Publishing House: Moscow, Russia, 1956, pp. 363-384.
- 61. Labintcev, E. Metrics in the problems of machine learning. Available online: https://habrahabr.ru/company/ods/blog/328372/ (accessed on 23 April 2021).
- 62. Nelson, A. Driving efficiency in the Oil and Gas industry. Available online: https://biarri.com/driving-efficiency-oil-gas-industry/ (accessed on 23 April 2021).
- 63. Nybø, R. Efficient Drilling Problem Detection. Doctor of Philosophy, Norwegian University of Science and Technology, Trondheim, Norway, 2009.
- 64. Nybø, R., Sui, D. Closing the Integration Gap for the Next Generation of Drilling Decision Support Systems. SPE Intelligent Energy Conference & Exhibition, Utrecht, Netherlands, 1-3 April 2014, pp. 1-10.