INFORMATION TECHNOLOGIES
ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

Original Empirical Research

UDC 004.855.5

https://doi.org/10.23947/2587-8999-2024-8-4-19-26

Forecasting Drilling Mud Losses Using Python

Nikita V. Kornilaev[1] ✉, **Kamila F. Koledina**[1,2]

[1] Ufa State Petroleum Technological University, Ufa, Russian Federation

[2] Institute of Petrochemistry and Catalysis, Ufa Federal Research Center of the Russian Academy of Sciences, Ufa, Russian Federation

✉ nik-kornilaev@mail.ru

Abstract

Introduction. Drilling mud losses are among the most common complications encountered during well drilling. Forecasting these losses is a priority as it helps minimize drilling fluid wastage and prevent wellbore incidents. Mud loss events are primarily influenced by the geological properties of the formations being drilled. Understanding the relationship between mud loss occurrences and the geological characteristics of the formations has both fundamental and practical significance. Given the complexity of predicting mud loss probabilities using traditional mathematical models, this study aims to develop a machine-learning-based system to predict the probability of mud losses based on well location and stratigraphic description.

Materials and Methods. Experimental data from 735 wells at the Shkapovskoye oil field, including well location coordinates, geological layer indices, and mud loss intensities, were prepared for computational analysis. The dataset was divided into training and testing subsets. The classification problem was addressed using four intensity classes with the following machine learning models: Decision Tree, Random Forest, and Linear Discriminant Analysis.

Results. Predictions generated by the three models were compared against the experimental data in the test set. The evaluation metrics included accuracy and recall. All three models achieved an average prediction accuracy of 91%. Linear Discriminant Analysis was identified as the most accurate model.

Discussion and Conclusion. High-accuracy predictions enable reliable forecasting of the probability and intensity of mud losses based on the location and stratigraphic description of new wells. The study presents three machine learning methods that demonstrated superior results in solving this problem.

Keywords: Python, mud loss, drilling, machine learning methods, Decision Tree, Discriminant Analysis, Random Forest

For Citation. Kornilaev N.V., Koledina K.F. Forecasting Drilling Mud Losses Using Python. *Computational Mathematics and Information Technologies*. 2024;8(4):19–26. https://doi.org/10.23947/2587-8999-2024-8-4-19-26

Оригинальное эмпирическое исследование

Прогнозирование поглощений бурового раствора на Python

Н.В. Корнилаев[1] ✉, **К.Ф. Коледина**[1,2]

[1] Уфимский государственный нефтяной технический университет, г. Уфа, Российская Федерация

[2] Институт нефтехимии и катализа УФИЦ РАН, г. Уфа, Российская Федерация

✉ nik-kornilaev@mail.ru

Аннотация

Введение. Поглощения бурового раствора являются одним из наиболее распространённых видов осложнений в бурении скважин. Первостепенной задачей является прогнозирование процесса поглощения, так как предупреждение данного вида осложнения позволит минимизировать потери бурового раствора, а также предотвратить аварии в скважине. Возникновение поглощений обусловлено прежде всего геологическими свойствами пластов. Выяснение связи между возникновением поглощений бурового раствора и геологическими характеристиками разбуриваемых пластов представляет как фундаментальный, так и практический интерес. В связи со сложностью определения вероятности возникновений поглощений с помощью известных математических моделей была поставлена цель исследования — построить с помощью методов машинного обучения систему, прогнозирующую значения вероятности возникновения поглощений в зависимости от местоположения скважины и её стратиграфического описания.

Материалы и методы. Экспериментальные данные о 735 скважинах Шкаповского месторождения (координаты местоположения, геологический индекс пласта, значение интенсивности поглощений) были подготовлены авторами к вычислениям. Исходные данные были разделены на обучающую и тестовую выборки. Представлены варианты решения задачи классификации по четырем классам интенсивности поглощений с использованием следующих моделей машинного обучения: «дерево решений», «случайный лес», «линейный дискриминантный анализ».

Результаты исследования. Результаты прогнозирования по трём моделям сравнивались с экспериментальными данными тестовой выборки. Для оценки качества моделей использовались метрики «точность» и «полнота». По всем трём моделям была достигнута средняя точность предсказания значений — 91 %. Было установлено, что наиболее точной моделью является «линейный дискриминантный анализ».

Обсуждение и заключение. Прогнозы высокой точности позволяют предсказывать, с какой вероятностью будут возникать поглощения определённой интенсивности в зависимости от местоположения новой скважины и её стратиграфического описания. В работе представлено три метода решения задачи, показавших наилучшие результаты.

Ключевые слова: Python, поглощение, бурение, методы машинного обучения, дерево решений, дискриминантный анализ, случайный лес

Для цитирования. Корнилаев Н.В., Коледина К.Ф. Прогнозирование поглощений бурового раствора на Python. *Computational Mathematics and Information Technologies*. 2024;8(4):19–26. https://doi.org/10.23947/2587-8999-2024-8-4-19-26

Introduction. To enhance competitiveness and optimize drilling expenses, artificial intelligence methods are widely employed in managing drilling processes today. Preventing complications and accidents remains a primary objective, as it helps minimize or entirely avoid unexpected costs associated with their mitigation.

Mud losses represent the complete or partial loss of drilling fluid as it filtrates into the formation. This phenomenon is influenced by numerous factors grouped into two major categories: geological and technological. Geological factors, such as rock properties (porosity, fracturing, permeability), have a greater impact than technological factors (properties of the selected drilling fluid, flushing fluid pressure). This is because surface operations allow for fine-tuning of the drilling process and control of critical parameters, whereas obtaining precise rock characteristics is not always feasible. Under conditions of high subsurface pressures and temperatures, formations may exhibit unpredictable properties.

In [1], the patterns of mud loss occurrences were thoroughly studied at the Yuzhno-Orlovskoye field in the Samara region. Table 1 presents data on the presence and intensity of mud losses based on research conducted at specific drilling intervals, accompanied by stratigraphic descriptions of the underlying formations.

Table 1

Intensity of Mud Losses in Wells

Well Number	Loss Interval	Stratigraphy	Mud Loss, m³/h
16	2079–2087	$C_1^{\,t}$	10
	2174–2624	$D_3^{mn} + D_3^{fm}$	catastrophic
4	2005	$D_3^{mn} + D_3^{fm}$	0.4
5	2124–2181	$D_3^{mn} + D_3^{fm}$	6–20
	2188		full
	2245–2259		
12	1925–1964	$C_1^{\,t}$	2–3
	2064–2114	D_3^{fm}	4–18
	2150–2178	D_3^{mn}	full
19	2099–2103	D_3^{fm}	12–60
	2130–2236	D_3^{mn}	full

From Table 1, it is evident that wells with similar rock properties exhibit mud losses of varying intensity, highlighting the unpredictable nature of mud loss occurrences.

Numerous studies [2–5] have focused on forecasting various types of complications and developing recommendation systems to address potential issues, including mud losses. The core of such software solutions relies on artificial neural networks trained on large volumes of geological data collected from geophysical logging stations. However, these systems share common drawbacks, including the inability to promptly obtain comprehensive downhole data in real time. This limitation restricts the amount of input data, reducing the model's prediction accuracy. Furthermore, when developing such systems, it is essential to consider the protective policies of oil and gas companies, which often make it impossible to access sufficient initial data for model development. Therefore, there is a need to create effective algorithms and software solutions capable of operating under conditions of limited input data.

The aim of this study was to develop a machine learning-based system to predict the probability of drilling mud losses of a specified intensity, depending on the well's location and the stratigraphic description of the formations being drilled.

To achieve this goal, the following tasks were undertaken:

– prepare experimental data for calculations;

– analyze machine learning algorithms and develop a program using the most optimal methods.

Materials and Methods. The Shkapovskoye oil field, located in the Republic of Bashkortostan, was selected for studying mud losses. In [6], a map of the field was presented, where wells are marked with symbols indicating the intensity of mud losses for each well. The map of the Shkapovskoye oil field is shown in Fig. 1.

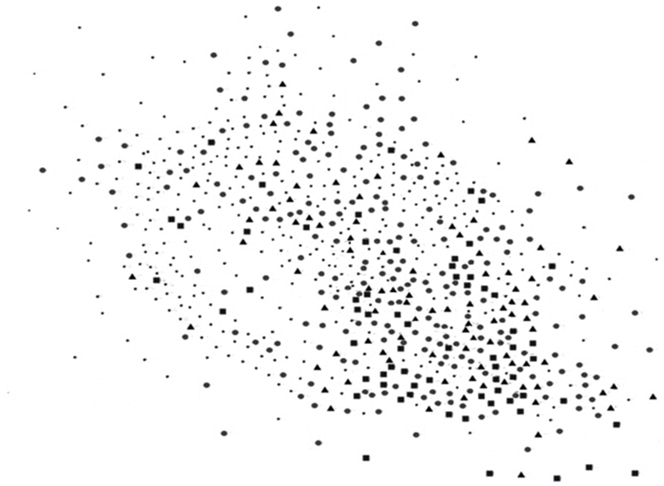

Fig. 1. Shkapovskoye Oil Field

Four classes of mud loss intensity were defined:

– 0 m³/h — no mud losses (dot, or a small circle);

– 0 to 40 m³/h — low-intensity mud losses (large circle);

– 40 to 80 m³/h — moderate-intensity mud losses (triangle);

– Over 80 m³/h — catastrophic mud losses (square).

Using the Yandex.Maps service, the length and width of the field were determined. Then, based on the map data, the coordinates of each well were calculated using the GeoGebra software package. A fragment of the calculation is shown in Fig. 2.

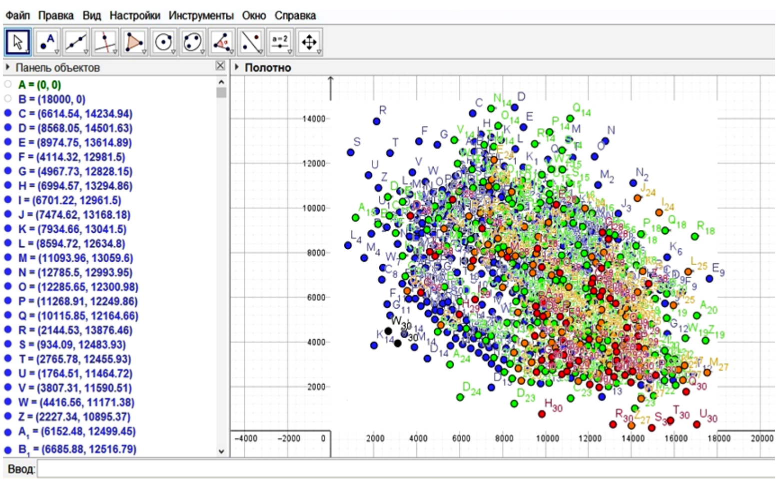

Fig. 2. Determination of Well Coordinates

Wells marked with differently colored dots correspond to mud loss intensity levels in increasing order:

– blue dots — 0 m³/h (no mud losses);

– green dots — 0 to 40 m³/h (low intensity);

– orange dots — 40 to 80 m³/h (moderate intensity);

– red dots — over 80 m³/h (catastrophic losses).

For stratigraphic descriptions, we utilized information indicating that the primary productive horizons are the Pashian (d3_p3), Kynovian (d3_kn), and Starooskolskiy (d2_st) horizons of the Devonian system. Additionally, the Bobrikovian horizon (c1_bb) has industrial significance, albeit to a lesser extent [7].

The data for the wells were categorized as follows: coordinates, stratigraphic description, and mud loss intensity. Various machine learning methods were tested on this dataset, and the most suitable ones were selected for further model optimization.

Decision Tree. This algorithm creates a tree-like structure based on "If ..., then ..." rules. These rules are generated during training on the dataset by generalizing multiple observations, making them easily interpretable. Mathematically, the decision rule can be expressed as a set of conjunctions:

$$R(x) = \wedge_{j \in J} \left[a_j \le f_i(x) \le b_j \right], \tag{1}$$

where J is the set of features selected for decision-making; $f_i(x)$ represents a real-valued feature, and a_j, b_j are the conditions. If all features satisfy the conditions, the rule returns 1; otherwise, it returns 0.

The advantages of decision trees include their simplicity of interpretation compared to neural networks and some other machine learning algorithms, as well as low requirements for data preprocessing. However, the disadvantages include a high likelihood of overfitting, as the algorithm can create excessively large trees, which may not generalize well to other datasets.

Random Forest. The Random Forest algorithm is a versatile machine learning method based on an ensemble of decision trees. Compared to other machine learning methods, the theoretical foundation of Random Forest is straightforward. The formula for the resulting classifier $a(x)$ is as follows:

$$a(x) = \frac{1}{N} \sum_{i=1}^{N} b_i(x), \tag{2}$$

where N is the number of trees; i is the tree index; b represents a decision tree, and $a(x)$ is the sample generated based on the input data.

Despite its versatility, this method has several significant drawbacks:

– difficulty in interpretation;

– inability to extrapolate;

– susceptibility to overfitting on highly noisy data;

– bias towards features with a larger number of levels when working with datasets containing categorical variables with varying levels [9].

Linear Discriminant Analysis (LDA). The main idea behind the selected algorithm is based on the assumption of a multivariate normal distribution within classes and the search for a linear transformation that maximizes the between-class variance while minimizing the within-class variance [10].

The proposed algorithm has the following advantages:

– lower tendency to overfit (compared to logistic regression), as LDA models the data distribution within each class and requires fewer parameters for estimation;

– it is more stable and efficient when there is a large number of classes with good linear separability.

The main disadvantage of LDA is its sensitivity to outliers and inefficiency when the number of features significantly exceeds the number of objects.

Results. The classification task was solved using the Python programming language, with the libraries sklearn, pandas, numpy, tkinter, and the MySQL DBMS. The program flowchart is shown in Fig. 3.

To visualize the modeling results, we compared the absorption intensity of wells in the test sample with the absorption intensities predicted by the models. The absorption intensity schemes for the test sample wells are shown in Fig. 4.

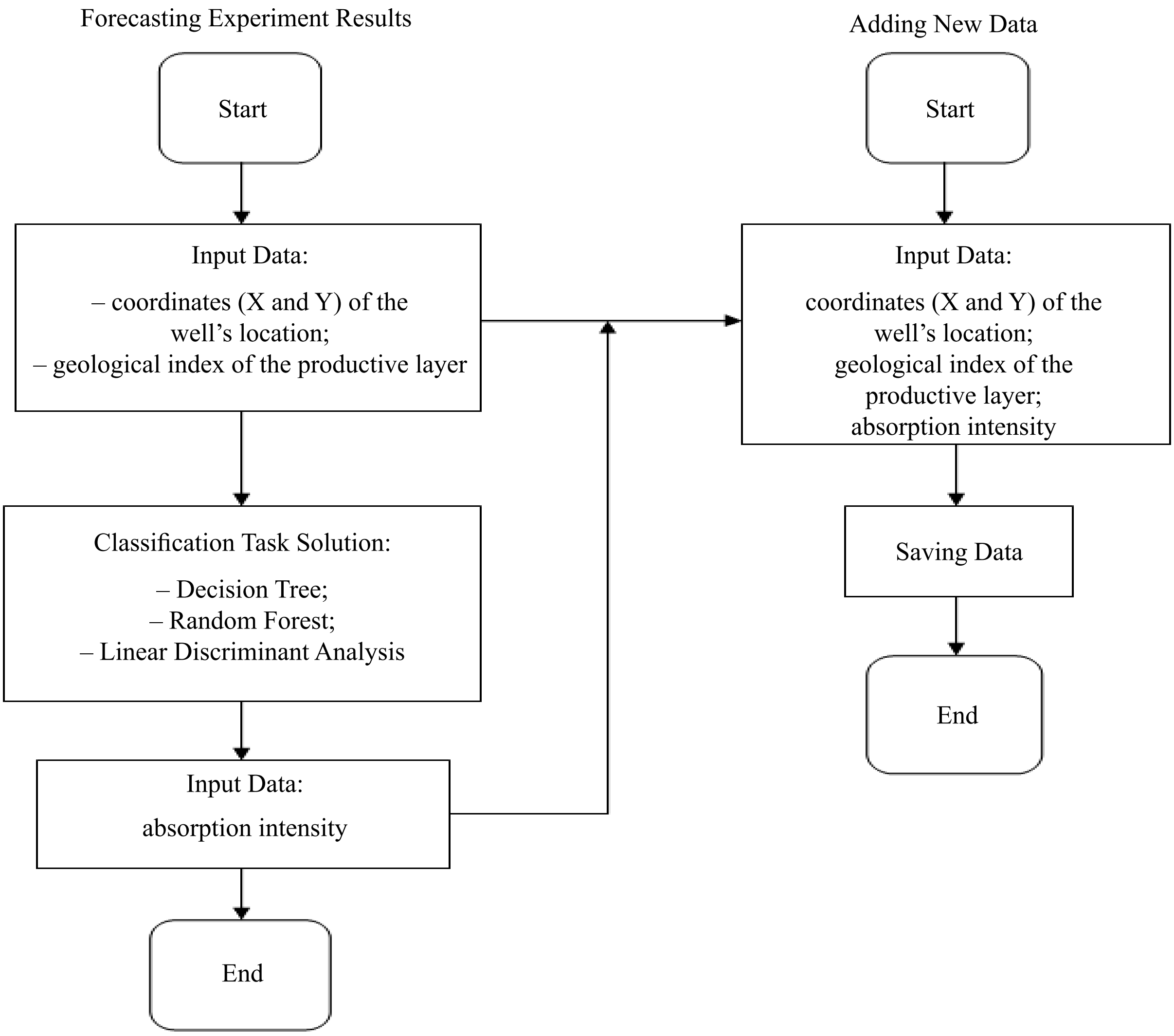

Fig. 3. The program flowchart

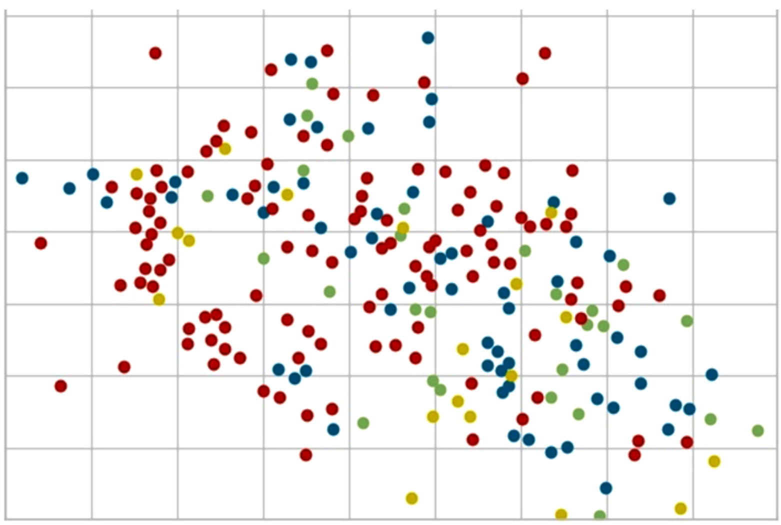

Fig. 4. Absorption intensity scheme (test sample wells)

Wells marked with differently colored dots correspond to mud loss intensity levels in increasing order:

– blue dots — 0 m³/h (no mud losses);
– green dots — 0 to 40 m³/h (low intensity);
– orange dots — 40 to 80 m³/h (moderate intensity);

– red dots — over 80 m³/h (catastrophic losses).

Fig. 5–7 show the schemes of the deposits with wells and predicted absorption intensities for the three machine learning models considered. Differences from the test sample are noted.

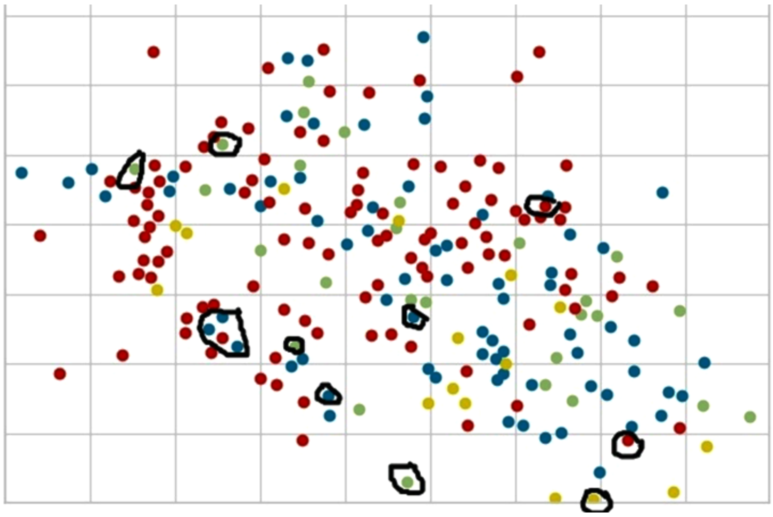

Fig. 5. Absorption intensity scheme (for the “Decision Tree” algorithm)

Wells marked with differently colored dots correspond to mud loss intensity levels in increasing order:

– blue dots — 0 m³/h (no mud losses);

– green dots — 0 to 40 m³/h (low intensity);

– orange dots — 40 to 80 m³/h (moderate intensity);

– red dots — over 80 m³/h (catastrophic losses).

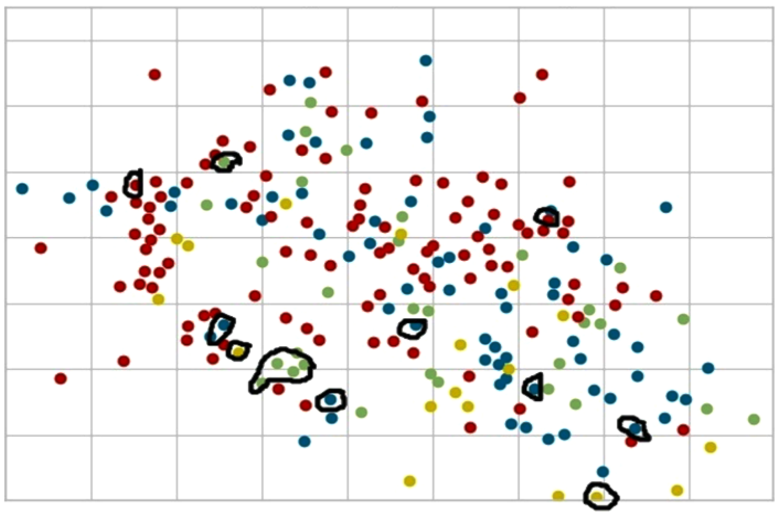

Fig. 6. Absorption intensity scheme (for the “Linear Discriminant Analysis” algorithm)

Wells marked with differently colored dots correspond to mud loss intensity levels in increasing order:

– blue dots — 0 m³/h (no mud losses);

– green dots — 0 to 40 m³/h (low intensity);

– orange dots — 40 to 80 m³/h (moderate intensity);

– red dots — over 80 m³/h (catastrophic losses).

The greatest number of “mismatches” was observed with the “Random Forest” algorithm model. This can be explained by the insufficient size and number of features in the training sample for constructing the ensemble of decision trees.

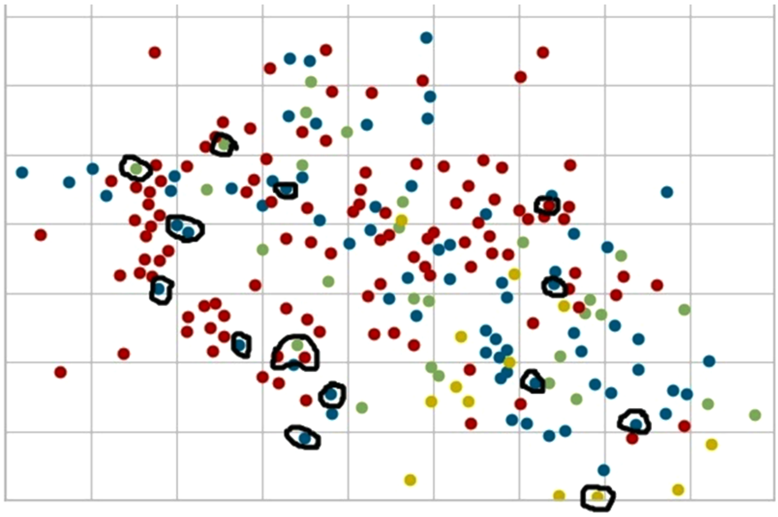

Fig. 7. Absorption intensity scheme (for the “Random Forest” algorithm)

For accuracy metrics, recall (sensitivity) was used, which characterizes the ability to identify the considered class, as well as precision, which allows distinguishing one class from another. These metrics are calculated using formulas (3, 4):

$$\text{precision} = \frac{TP}{TP + FP}, \tag{3}$$

$$\text{recall} = \frac{TP}{TP + FN}, \tag{4}$$

where TP — True Positives: correctly predicted values of the class under consideration; FP — False Positives: incorrectly predicted values of the class under consideration; FN — False Negatives: incorrectly predicted values of other classes.

The calculation results for each class are presented in Table 2.

Table 2

Metrics for evaluating the quality of machine learning models

Absorption Class	Precision max	Recall max
0 m^3	0.88 (LDA)	0.97 (Decision Tree)
0–40 m^3	0.89 (Random Forest)	0.93 (LDA)
40–80 m^3	0.93 (Decision Tree)	0.84 (LDA)
> 80 m^3	0.98 (LDA)	0.92 (Random Forest)

From Table 2, it can be concluded that all three models demonstrated high prediction performance. The most effective algorithm for the problem at hand is Linear Discriminant Analysis.

Discussion and Conclusion. The results obtained during the prediction of drilling fluid loss intensity in wells are relevant for practical application in assessing complications in the field. Despite the high predictive capability of the model, its main limitation is the lack of applicability to other fields. To achieve accurate classification for different fields, the model must be retrained on the corresponding operational data.

Thus, it is essential to develop solutions for the preliminary analysis of “raw” data provided by geological exploration and the subsequent transfer of processed data to machine learning algorithms.

References

1. Shipovskiy K.A., Tsirkova V.S., Koval M.E., Kozhin V.N. Regularities of the distribution of lost circulation areas in single reefs of the depression zone of the Kama-Kinel downfold system. *Exposition Oil & Gas*. 2021;1(80):34–39. (In Russ.) https://doi.org/10.24412/2076-6785-2021-1-34-39

2. Eremin N.A., Selenginsky D.A. On the possibilities of applying artificial intelligence methods in solving oil and gas problems. *Proceedings of Tula State University. Earth sciences*. 2023;1(1):201–211. (In Russ.) https://doi.org/10.46689/2218-5194-2023-1-1-201-211

3. Abu-Abed F.N., Borisov N.A., Khabarov A.R. The use of pattern recognition methods for the analysis of emergency situations. In: *Collection of articles of the XVI International scientific and practical conference "Mathematical methods and information technologies in economics, Sociology and Education"*. Penza: Autonomous non-profit scientific and educational organization "Volga House of Knowledge"; 2005. pp. 428–431. (In Russ.)

4. Kodirov Sh.Sh., Shestakov A.L. Development of an artificial neural network for predicting the tack of drill pipe columns. *Bulletin of SUSU. The series "Computer technology, control, radio electronics"*. 2019;19(3):20–32. (In Russ.) https://doi.org/10.14529/ctcr190302

5. Dmitrievsky A.N., Eremin N.A., Chernikov A.D., Zinatulina L.I. The use of artificial intelligence methods to prevent complications during well construction. *Proceedings of Tula State University. Earth sciences*. 2021;4:132–145. (In Russ.) https://doi.org/10.46689/2218-5194-2021-4-1-132-144

6. Lind Yu.B., Kabirova A.R., Nurislamova L.F. Predicting complications in the drilling process using parallel computing technology. In: *Proceedings of the international scientific conference "Parallel Computing Technologies"*. Novosibirsk: SUSU Publishing Center; 2012. pp. 571–576. (In Russ.)

7. Trebin G.F., Charygin N.V., Obukhova T.M. *Oil fields of the Soviet Union. Guide*. 2nd ed., additional and revised. Moscow: Nedra; 1980. 583 p. (In Russ.)

8. DecisionTrees — Geeksforgeeks. URL: https://www.geeksforgeeks.org/decision-tree-introduction-example/ (accessed: 21.04.2024).

9. Random Forest Algorithm in Machine Learning — Geekforgeeks. URL: https://www.geeksforgeeks.org/random-forest-algorithm-in-machine-learning/ (accessed: 21.04.2024).

10. Linear Discriminant Analysis in Machine Learning — GeeksforGeeks. URL: https://www.geeksforgeeks.org/ml-linear-discriminant-analysis/ (accessed: 21.04.2024).

About the Authors:

Nikita V. Kornilaev, student at the Department of Information Technologies and Applied Mathematics of the Ufa State Petroleum Technological University (1, Kosmonavtov str., Ufa, 450064, Russian Federation), ORCID, nik-kornilaev@mail.ru

Kamila F. Koledina, Dr. Sc. (Phys.-Math.), Associate Professor at the Department of Information Technologies and Applied Mathematics of the Ufa State Petroleum Technological University (1, Kosmonavtov str., 450064, Ufa, Russian Federation), Research Fellow of the Institute of Petrochemistry and Catalysis of the Russian Academy of Sciences (450075, October ave., 141, Ufa, Russian Federation), ORCID, koledinakamila@mail.ru

Claimed Contributorship:

N.V. Kornilaev: Program development and algorithm implementation.

K.F. Koledina: task formalization, testing, specification development.

Conflict of Interest Statement: **the authors declare no conflict of interest.**

All authors have read and approved the final manuscript.

Об авторах:

Никита Владимирович Корнилаев, студент кафедры информационных технологий и прикладной математики Уфимского государственного нефтяного технологического университета (450064, Российская Федерация, г. Уфа, ул. Космонавтов, 1), ORCID, nik-kornilaev@mail.ru

Камила Феликсовна Коледина, доктор физико-математических наук, доцент кафедры информационных технологий и прикладной математики Уфимского государственного нефтяного технологического университета (450064, Российская Федерация, г. Уфа, ул. Космонавтов, 1), научный сотрудник Института нефтехимии и катализа Российской академии наук (450075, Российская Федерация, г. Уфа, пр-т Октября, 141), ORCID, koledinakamila@mail.ru

Заявленный вклад авторов:

Н.В. Корнилаев: разработка программы и реализация алгоритма.

К.Ф. Коледина: формализация задачи, тестирование, разработка спецификации.

Конфликт интересов: **авторы заявляют об отсутствии конфликта интересов.**

Все авторы прочитали и одобрили окончательный вариант рукописи.

Received / Поступила в редакцию 21.10.2024

Reviewed / Поступила после рецензирования 07.11.2024

Accepted / Принята к публикации 15.11.2024