

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

INFORMATION TECHNOLOGIES



УДК 004.855.5

Оригинальное эмпирическое исследование

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


Прогнозирование поглощений бурового раствора на Python

Н.В. Корнилаев¹ , К.Ф. Коледина^{1,2} ¹ Уфимский государственный нефтяной технический университет, г. Уфа, Российская Федерация² 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 возникновения 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>

Original Empirical Research

Forecasting Drilling Mud Losses Using Python

Nikita V. Kornilaev¹ , Kamila F. Koledina^{1,2} ¹ Ufa State Petroleum Technological University, Ufa, Russian Federation² 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>

Введение. На сегодняшний день с целью повышения конкурентоспособности и оптимизации расходов на бурение скважин широко применяются методы искусственного интеллекта для управления технологическим процессом. Предупреждение осложнений и аварий является основной задачей, так как это позволяет сократить или вовсе избежать непредвиденных расходов на ликвидацию последствий.

Поглощения представляют собой полную или частичную потерю бурового раствора при её фильтрации в пласт. Данное явление зависит от многих факторов, объединённых в две большие группы: геологические и технологические. Геологические факторы, т. е. характеристики горных пород (пористость, трещиноватость, проницаемость) имеют наибольшее влияние, чем технологические (свойства подобранного бурового раствора, давление промывочной жидкости). Это объясняется тем, что на поверхности земли возможно подобрать заданный режим бурения, контролируя все необходимые параметры, однако получить точные характеристики горных пород не всегда удастся. В условиях высоких горных давлений и температур пласты могут обладать непредсказуемыми характеристиками. В работе [1] подробно исследовалась закономерность возникновения поглощений на Южно-Орловском месторождении в Самарской области. В таблице 1 представлены данные о наличии и интенсивности поглощений в соответствии с исследованиями на определённых интервалах бурения со стратиграфическим описанием залегающих пластов.

Таблица 1

Интенсивность поглощений в скважинах

№ скважины	Интервал поглощения	Стратиграфия	Поглощение, м ³ /ч
16	2079–2087	C_1^t	10
	2174–2624	$D_3^{mn} + D_3^{fm}$	катастрофическое
4	2005	$D_3^{mn} + D_3^{fm}$	0,4
5	2124–2181	$D_3^{mn} + D_3^{fm}$	6–20
	2188		полное
	2245–2259		
12	1925–1964	C_1^t	2–3
	2064–2114	D_3^{fm}	4–18
	2150–2178	D_3^{mn}	полное
19	2099–2103	D_3^{fm}	12–60
	2130–2236	D_3^{mn}	полное

Из таблицы 1 видно, что в скважинах со схожими свойствами горных пород наблюдаются поглощения с различной интенсивностью, что говорит о непредсказуемом характере возникновения поглощений.

Множество исследований [2–5] было направлено на прогнозирование различных видов осложнений и разработку рекомендательных систем при возможном их возникновении, в том числе и для поглощений. Основой работы программных продуктов являются искусственные нейронные сети, обученные на больших объемах геоданных со станций геолого-технологических измерений. К общим недостаткам данных программных комплексов относится невозможность оперативного получения полного объема информации с забоя в режиме

реального времени, из-за чего количество входных данных ограничивается, что снижает точность прогноза модели. Помимо этого, при разработке необходимо учитывать протекционизм нефтегазовых компаний в современных условиях, что делает невозможным получить множество исходных данных для создания модели. Следовательно, необходимо разрабатывать эффективные алгоритмы и программные комплексы, работающие в условиях ограничения исходных данных.

Таким образом, была поставлена цель исследования — построить с помощью методов машинного обучения систему, прогнозирующую вероятность возникновения поглощений бурового раствора заданной интенсивности в зависимости от местоположения и стратиграфического описания вскрываемых пластов.

Для реализации цели были поставлены следующие задачи:

- подготовить экспериментальные данные для расчётов;
- проанализировать алгоритмы машинного обучения и создать программу с применением наиболее оптимальных методов.

Материалы и методы. Для исследования поглощений было выбрано Шкаповское месторождение республики Башкортостан. В работе [6] была представлена карта месторождения, на которой скважины обозначены условными знаками, передающими информацию об интенсивности поглощений в каждой скважине. Карта Шкаповского месторождения представлена на рис. 1.

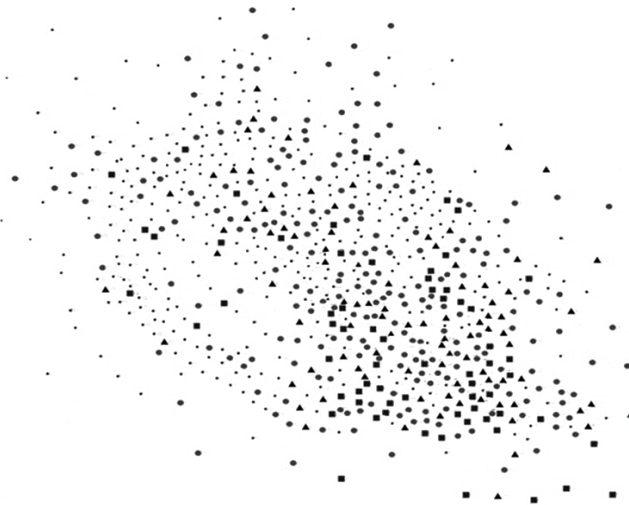


Рис. 1. Шкаповское месторождение

Было выделено 4 класса поглощений по их интенсивности, на рис. 1 они отмечены следующим образом:

- $0 \text{ м}^3/\text{час}$ — поглощения отсутствуют (точки);
- от 0 до $40 \text{ м}^3/\text{час}$ — поглощения с небольшой интенсивностью (круги);
- от 40 до $80 \text{ м}^3/\text{час}$ — поглощения средней интенсивности (треугольники);
- более $80 \text{ м}^3/\text{час}$ — катастрофические поглощения (квадраты).

С помощью сервиса Яндекс.Карты были определены длина и ширина месторождения, затем с помощью программного комплекса GeoGebra на основе данных с карты были определены координаты каждой скважины. Фрагмент расчёта представлен на рис. 2.

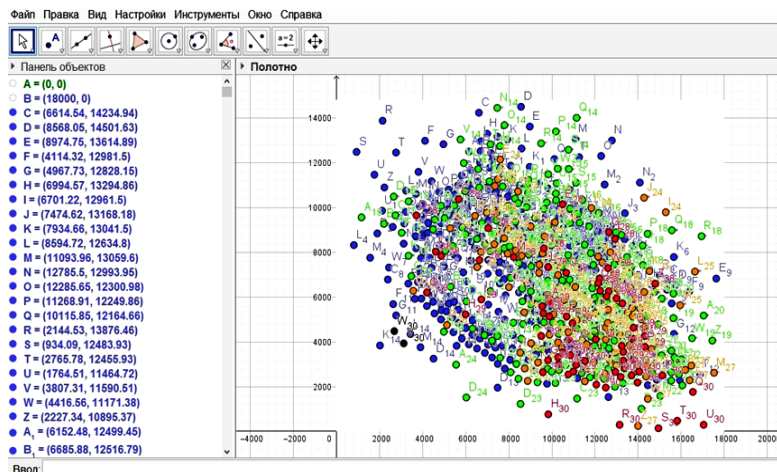


Рис. 2. Определение координат скважин

Скважины, помеченные разноцветными точками, здесь и далее соответствуют интенсивности поглощений в порядке возрастания интенсивности: синие точки — 0 м³/час; зелёные точки — от 0 до 40 м³/час; оранжевые точки — 40 до 80 м³/час; красные точки — более 80 м³/час.

Для определения стратиграфического описания воспользуемся информацией о том, что основным продуктивным горизонтом является пласт пашийского (d3_p3), кыновского (d3_kn) и старооскольского (d2_st) горизонтов девонской системы. Также имеют промышленное значение, но в меньшей мере, отложения бобриковского горизонта (c1_bb) [7]. Данные по скважинам представлены следующими категориями: координаты, стратиграфическое описание, интенсивность поглощений. На полученных данных были опробованы различные методы машинного обучения и выбраны наиболее подходящие для последующего улучшения модели.

Дерево решений. Данный алгоритм представляет собой создание древовидной структуры, в которой реализованы правила: «Если ..., то...». Данные правила формируются в процессе решения на обучающей выборке за счёт обобщений множества наблюдений, причем они являются легко интерпретируемыми. С математической точки зрения решающее правило представляет собой набор конъюнкций:

$$R(x) = \bigwedge_{j \in J} [a_j \leq f_i(x) \leq b_j], \quad (1)$$

где J — множество признаков, отбираемые для формирования вывода; $f_i(x)$ — вещественный признак; a_j, b_j — набор условий. Если вся совокупность признаков удовлетворяет условиям, то возвращается 1, иначе 0.

К преимуществам алгоритма можно отнести простоту интерпретации относительно нейронных сетей и некоторых других алгоритмов машинного обучения, невысокие требования к предварительной обработке данных. Недостатками решающих деревьев являются высокая вероятность переобучения, так как алгоритм может построить слишком большое дерево, которое может быть неприменимо к другому набору исходных данных.

Случайный лес. Алгоритм случайного леса (Random Forest) — универсальный алгоритм машинного обучения, суть которого состоит в использовании ансамбля решающих деревьев.

По сравнению с другими методами машинного обучения теоретическая часть алгоритма Random Forest проста, формула итогового классификатора $a(x)$ выглядит следующим образом:

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

где N — количество деревьев; i — счетчик для деревьев; b — решающее дерево; $a(x)$ — сгенерированная на основе данных выборка.

Однако, несмотря на универсальность, у данного метода наблюдается ряд существенных недостатков:

- сложность интерпретации;
- случайный лес не умеет экстраполировать;
- алгоритм склонен к переобучению на сильно зашумленных данных;
- для данных, включающих категориальные переменные с различным количеством уровней, случайные леса предвзяты в пользу признаков с большим количеством уровней [9].

Линейный дискриминантный анализ (LDA). Основная идея выбранного алгоритма заключается в предположении о многомерном нормальном распределении внутри классов и поиске линейного преобразования для максимизации межклассовой дисперсии и минимизации внутриклассовой [10].

Предложенный алгоритм обладает следующими преимуществами:

- меньшая склонность к переобучению (в отличие от алгоритма логистической регрессии), поскольку LDA основан на моделировании распределения данных в каждом классе и ему требуется меньше параметров для оценки;
- при большом количестве классов и их хорошем линейном разделении является более стабильным и эффективным.

Основным недостатком LDA является чувствительность к выбросам и неэффективность к сильно превышающему числу признаков над числом объектов.

Результаты исследования. Решение задачи классификации осуществлялось с помощью языка программирования Python, использовались библиотеки sklearn, pandas, numpy, tkinter, СУБД MySQL. Блок-схема программы представлена на рис. 3.

Для визуализации результатов моделирования сравним интенсивность поглощений скважин тестовой выборки и интенсивности поглощений, предсказанные моделями. Схемы интенсивности скважин тестовой выборки представлены на рис. 4.

На рис. 5–7 представлены схемы месторождений со скважинами с предсказанными интенсивностями поглощений для трёх рассмотренных моделей машинного обучения. Отмечены отличия от тестовой выборки.

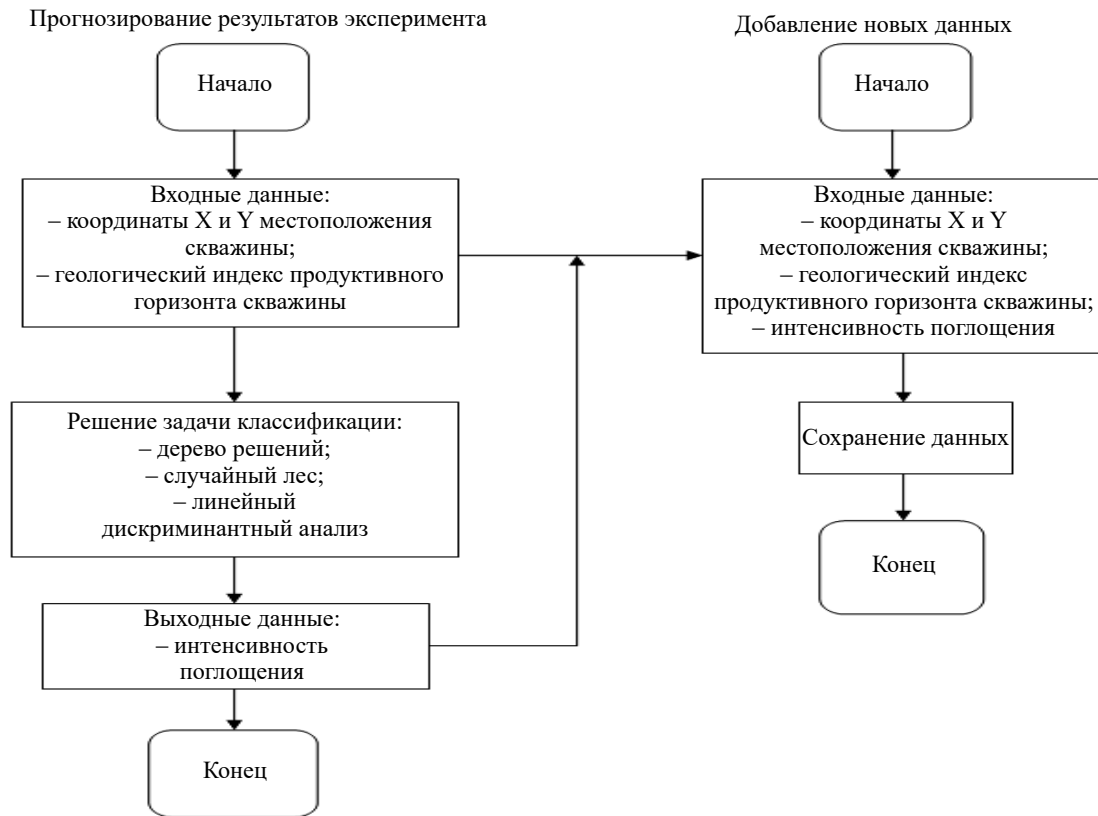


Рис. 3. Блок-схема программы

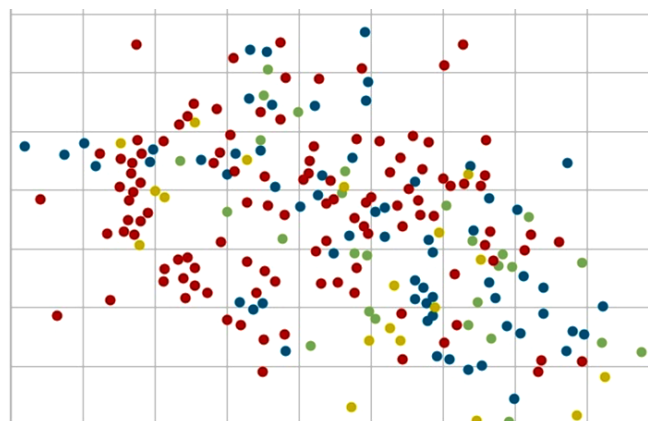


Рис. 4. Схема интенсивности поглощений (скважины тестовой выборки)

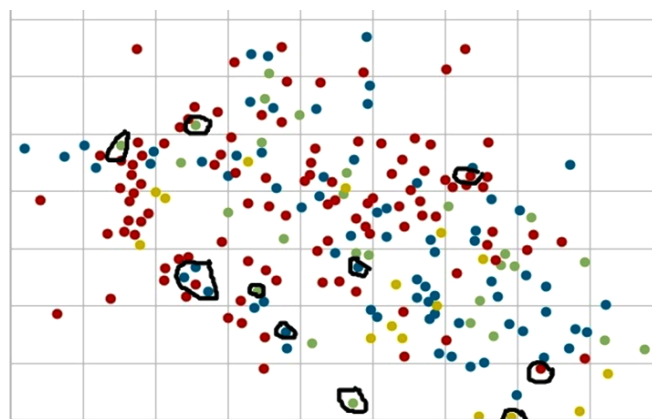


Рис. 5. Схема интенсивности поглощений (для алгоритма «Дерево решений»)

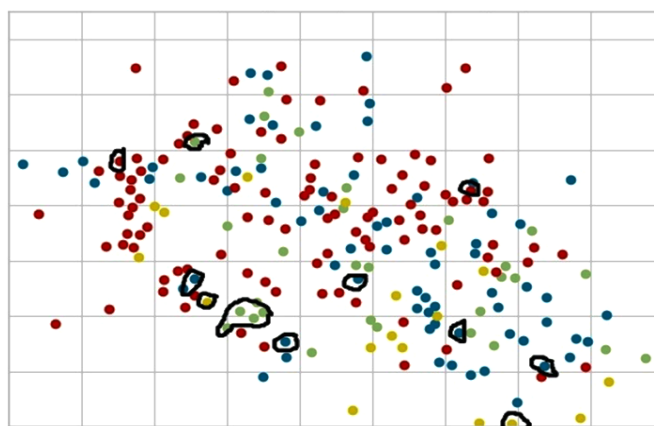


Рис. 6. Схема интенсивности поглощений (для алгоритма «Линейный дискриминантный анализ»)

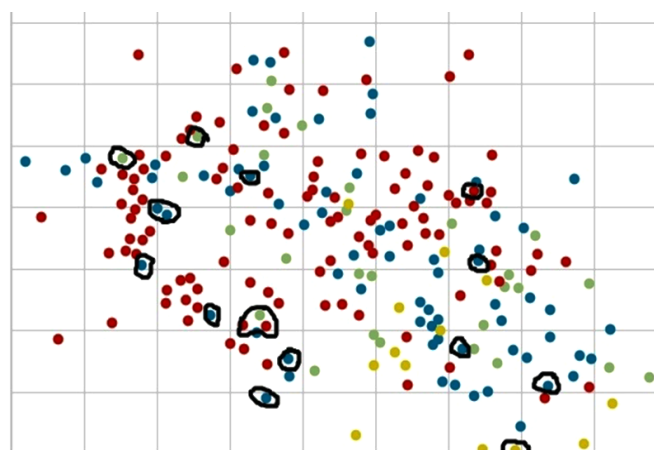


Рис. 7. Схема интенсивности поглощений (для алгоритма «Случайный лес»)

Наибольшее количество «несовпадений» наблюдается у модели алгоритма «Случайный лес». Это можно объяснить недостаточным размером и количеством признаков обучающей выборки для построения ансамбля решающих деревьев.

В качестве метрик точности использовались метрики *recall* (полнота), характеризующие способность находить рассматриваемый класс, а также показатель *precision* (точность), позволяющий отличать один класс от другого. Данные метрики рассчитываются по формулам (3, 4):

$$\text{precision} = \frac{TP}{TP + FP}, \quad (3)$$

$$\text{recall} = \frac{TP}{TP + FN}, \quad (4)$$

где *TP* — верно предсказанные значения рассматриваемого класса; *FP* — неверно предсказанные значения рассматриваемого класса; *FN* — неверно предсказанные значения других классов.

Результаты расчётов для каждого из классов приведены в таблице 2.

Таблица 2

Метрики оценки качества моделей машинного обучения

Класс поглощений	Максимальная точность	Максимальная полнота
0 м ³	0,88 (ЛДА)	0,97 (Дерево решений)
0–40 м ³	0,89 (Случайный лес)	0,93 (ЛДА)
40–80 м ³	0,93 (Дерево решений)	0,84 (ЛДА)
> 80 м ³	0,98 (ЛДА)	0,92 (Случайный лес)

Из таблицы 2 можно сделать вывод, что все три модели показали высокие результаты прогнозирования. Наиболее эффективным алгоритмом в рассматриваемой задаче является линейный дискриминантный анализ.

Обсуждение и заключение. Результаты, полученные в ходе решения задачи прогнозирования интенсивности поглощений бурового раствора в скважинах, актуальны для применения в реальной оценке возникновения осложнений на промысле. Несмотря на высокую предсказательную способность данной модели, главным ее недостатком является неприменимость на других месторождениях. Для получения актуального решения задачи классификации на других месторождениях необходимо переобучать модель на соответствующих промысловых данных. Поэтому необходимо разрабатывать решения для предварительного анализа «сырых» данных, предоставленных геологоразведкой, и последующей передачи данных для алгоритмов машинного обучения.

Список литературы / References

1. Шиповский К.А., Циркова В.С., Коваль М.Е., Кожин В.Н. Закономерности распределения зон поглощений в одиночных рифах депрессионной зоны Камско-Кинельской системы прогибов. *Экспозиция Нефть Газ*. 2021;1(80):34–39. <https://doi.org/10.24412/2076-6785-2021-1-34-39>
- 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. Ерёмин Н.А., Селенгинский Д.А. О возможностях применения методов искусственного интеллекта в решении нефтегазовых задач. *Известия Тульского государственного университета. Науки о земле*. 2023;1(1):201–211. <https://doi.org/10.46689/2218-5194-2023-1-1-201-211>
- 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. Абу-Абед Ф.Н., Борисов Н.А., Хабаров А.Р. Использование методов распознавания образов для анализа аварийных ситуаций. В: *Сборник статей XVI международной научно-практической конференции «Математические методы и информационные технологии в экономике, социологии и образовании»*. Пенза: Автономная некоммерческая научно-образовательная организация «Приволжский Дом знаний»; 2005. С. 428–431.
- 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. Кодиров Ш.Ш., Шестаков А.Л. Разработка искусственной нейронной сети для прогнозирования прихватов колонн буровых труб. *Вестник ЮУрГУ. Серия «Компьютерные технологии, управление, радиоэлектроника»*. 2019;19(3):20–32. <https://doi.org/10.14529/ctcr190302>
- 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. Дмитриевский А.Н., Еремин Н.А., Черников А.Д., Зинатулина Л.И. Использование методов искусственного интеллекта для предотвращения осложнений при строительстве скважин. *Известия тульского государственного университета. Науки о земле*. 2021;4:132–145. <https://doi.org/10.46689/2218-5194-2021-4-1-132-144>
- 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. Линд Ю.Б., Кабирова А.Р., Нурисламова Л.Ф. Прогнозирование осложнений в процессе бурения с использованием технологии параллельных вычислений. В: *Труды международной научной конференции «Параллельные вычислительные технологии»*. Новосибирск: Издательский центр ЮУрГУ; 2012. С. 571–576.
- 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. Требин Г.Ф., Чарыгин Н.В., Обухова Т.М. *Нефти месторождений Советского Союза. Справочник*. 2-е изд., доп. и перераб. Москва: Недра; 1980. 583 с.
- 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).

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.

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.

Поступила в редакцию / Received 21.10.2024

Поступила после рецензирования / Reviewed 07.11.2024

Принята к публикации / Accepted 15.11.2024