

УДК 519.25

DOI: [10.26102/2310-6018/2023.41.2.022](https://doi.org/10.26102/2310-6018/2023.41.2.022)

## Формирование данных в фиксациях моделей нефтегазовых скважин на основе применения интеллектуального метода заполнения пропущенных значений

Н.В. Шарифьянов✉, В.А. Латыпова

*Уфимский университет науки и технологий, Уфа, Российская Федерация*  
*sharifyanov.niyaz@gmail.ru✉*

**Резюме.** В нефтегазовой сфере уделяется много внимания вопросу повышения качества данных, так как низкое качество может повлечь за собой неверное видение ситуации и в дальнейшем принятие неверного решения. Мониторинг нефтедобычи и профилактическое обслуживание предполагают сбор данных от самых разных датчиков, которые нужно корректно обработать и «упаковать». Поэтому особое внимание уделяется повышению качества формируемых данных в фиксациях моделей нефтегазовых скважин. Фиксация моделей нефтегазовых скважин – это процесс сбора, анализа и сохранения информации о параметрах скважинной эксплуатации, таких как дебит жидкости, газа и нефти, давление, температура, состав флюидов и другие параметры, используемые для оптимизации процессов добычи и повышения эффективности работы скважин. Наличие пропусков при формировании моделей скважин чревато заметным понижением качества данных моделей, что может привести к неполному представлению общей картины работы скважины и снизить точность прогнозирования ее производительности. В статье предлагается интеллектуальный метод заполнения пропущенных значений для формирования данных в фиксациях моделей нефтегазовых скважин для решения данной проблемы. Метод успешно апробирован в нефтегазовой компании ПАО «Газпром нефть» на данных по дебиту жидкости скважин Вынгапуровского месторождения.

**Ключевые слова:** интеллектуальный метод, заполнение пропущенных значений, метод ближайших соседей, качество данных, нефтегазовая скважина, фиксация модели скважины.

**Для цитирования:** Шарифьянов Н.В., Латыпова В.А. Формирование данных в фиксациях моделей нефтегазовых скважин на основе применения интеллектуального метода заполнения пропущенных значений. *Моделирование, оптимизация и информационные технологии.* 2023;11(2). URL: <https://moitvvt.ru/ru/journal/pdf?id=1381> DOI: 10.26102/2310-6018/2023.41.2.022

## Formation of data in fixations of oil and gas well models using an intelligent method for missing value completion

N.V. Sharifyanov✉, V.A. Latypova

*Ufa University of Science and Technology, Ufa, the Russian Federation*  
*sharifyanov.niyaz@gmail.ru✉*

**Abstract.** In the oil and gas sector, a lot of attention is paid to the issue of improving the quality of data because poor quality can distort the presentation of the situation and eventually cause making the wrong decision. Oil production monitoring and preventive maintenance involve the collection of data from a variety of sensors that need to be correctly processed and “packaged”. Therefore, particular emphasis is given to improving the quality of the generated data in oil and gas well model fixations. Fixing oil and gas well models is the process of collecting, analyzing, and storing information about well operation parameters such as fluid, gas and oil flow rates, pressure, temperature, fluid composition, and other parameters used to optimize production processes and improve well performance. The presence of gaps in the formation of well models can significantly reduce the quality of these models, which can lead to

an incomplete representation of the overall picture of well operation and decrease the accuracy of predicting its productivity. The article proposes an intelligent method of completing missing values for generating data in fixations of oil and gas well models to solve this problem. The method has been successfully tested at the oil and gas company Gazprom Neft PJSC using the data on the fluid flow rate of the wells in the Vyngapurovskoye field.

**Keywords:** intelligent method, completing missing values, nearest neighbor method, data quality, oil and gas well, well model fixation.

**For citation:** Sharifyanov N.V., Latypova V.A. Formation of data in fixations of oil and gas well models using an intelligent method for missing value completion. *Modeling, Optimization and Information Technology*. 2023;11(2). URL: <https://moitvvt.ru/ru/journal/pdf?id=1381> DOI: 10.26102/2310-6018/2023.41.2.022 (In Russ.).

## Введение

Работа с информацией в нефтегазовом секторе имеет свои особенности [1,2]. У предприятий отрасли традиционно сверхбольшой объем разнородных данных, часть из которых быстро устаревает или постоянно обновляется, а с учетом текущего перехода отрасли «в цифру» объем данных будет только расти. Цифровизация нефтяной промышленности может значительно улучшить качество данных и повысить эффективность производства.

Мониторинг нефтедобычи и профилактическое обслуживание предполагают сбор данных от самых разных датчиков, которые нужно корректно обработать и «упаковать». Поэтому особое внимание уделяется повышению качества формируемых данных в фиксациях моделей нефтегазовых скважин. Фиксация моделей нефтегазовых скважин – это процесс сбора, анализа и сохранения информации о параметрах скважинной эксплуатации, таких как дебит жидкости, газа и нефти, давление, температура, состав флюидов и другие параметры, используемые для оптимизации процессов добычи и повышения эффективности работы скважин.

Фиксация моделей является инструментом для автоматического отслеживания изменений в данных моделей скважины по одному или нескольким параметрам, что позволяет добиться упрощения моделирования, уменьшения количества переменных, снижения неопределенности и повышения точности прогнозов. Однако при фиксации параметров необходимо учитывать возможные изменения в реальных условиях добычи и геологических свойств пласта, которые могут повлиять на точность модели. В целом, фиксация моделей нефтегазовых скважин является важным инструментом для повышения точности прогнозов и упрощения моделирования, но требует тщательного анализа и обновления в соответствии с изменениями в реальных условиях добычи. Проблема низкого качества формируемых данных в фиксациях моделей нефтегазовых скважин заключается в том, что данные, полученные при эксплуатации скважин, могут быть неточными или неполными, содержать пропуски в данных. Это может привести к неверным выводам о состоянии скважины и ее производительности, а впоследствии увеличивается риск принятия неправильного решения при планировании и управлении производственными процессами.

Цель исследования – повысить качество формируемых данных в фиксациях моделей нефтегазовых скважин.

Задачи исследования:

- проанализировать существующие методы повышения качества формируемых данных в нефтегазовой области;
- разработать интеллектуальный метод заполнения пропущенных значений для использования при формировании данных в фиксациях моделей нефтегазовых скважин;

– провести апробацию предлагаемого метода в нефтегазовой компании ПАО «Газпром нефть».

### **Существующие методы повышения качества формируемых данных в нефтегазовой области**

Вопрос повышения качества мониторинга скважин поднимается в работе [3]. В рамках статьи рассматривается уникальная система удаленного мониторинга «Rosneft-WellView», обеспечивающая сбор информации со скважин и ее первичную обработку, структурирование и размещение в базах данных (БД), агрегацию данных из различных БД, выявление скважин с отклонениями от нормального режима работы, проведение экспресс-анализа режима работы скважин с учетом осложняющих факторов и истории работы, формирование аналитической отчетности. Особенностью системы является модуль визуализации архивов данных со станций управления, что позволяет анализировать работу установки и планировать качественные мероприятия в короткие сроки с минимальными потерями нефти.

В публикации [4] рассматривается архитектура типовой IWT-системы (intelligent well technology), которая позволяет в контексте обеспечения в режиме реального времени потоков и контуров информации о контроле, регулировании и диагностике скважин (кустов скважин) обеспечить непрерывный мониторинг месторождения и выйти на верхний уровень управления с помощью корпоративной ERP-системы. В работе [5] предлагается программно-технический комплекс СОКРАТ, который обеспечивает мониторинг, диагностику и адаптивное управление отдельными скважинами в нефтедобыче.

Для восстановления пропущенных значений при пространственно-панельном анализе данных нефтяного месторождения в работе [6] предлагается технология интерполяции на основе модели кригинга пористости по координатам пластопересечения. Кригинг – это улучшенный геостатистический метод, который позволяет строить предполагаемую поверхность из набора точек. По результатам моделирования получена восполненная карта исходной выборки с восстановленными пропущенными значениями показателя пористости, которая в дальнейшем может быть использована для построения панельных моделей.

В рамках работы [7] по подготовке исходных данных для углубленного анализа нефтегазового месторождения для восстановления пропущенных значений данных выбран алгоритм поиска ближайшего соседа (knearest algorithm imputation). Метод ближайших соседей – это алгоритм, который полезен для сопоставления точки с ее ближайшими к соседями в многомерном пространстве. Он может применяться для данных, которые являются непрерывными, дискретными, порядковыми и категориальными, что делает его особенно пригодным для работы со всеми видами недостающих данных. Предположение по использованию метода ближайших соседей для пропущенных значений состоит в том, что значение точки может быть аппроксимировано значениями ближайших к нему точек на основе других переменных, отвечающих за геолого-технические мероприятия, время работы скважины, простой, обводненность. Использование метода ближайших соседей в других областях рассмотрено в работах [8-10].

В ПАО «Газпром нефть» для восстановления пропущенных значений используется алгоритм пролонгации – это метод заполнения пропущенных значений во временных рядах данных путем использования имеющихся значений за предыдущие временные отрезки. Чаще всего на место пропуска вставляется предыдущее имеющееся значение, данные «протягиваются».

Восстановление пропущенных значений необходимо в различных сферах, таких как медицина, экономика, социология, статистика. Несмотря на то, что существует множество методов восстановления пропущенных значений, выбор конкретного метода зависит от многих факторов. Сюда относятся: тип данных, количество пропущенных значений, цель анализа и др. Например, используются следующие методы:

- простое восстановление, основанное на использовании среднего значения, медианы или моды вместо пропущенных значений [11];
- методы на основе применения алгоритма случайного леса;
- PMM (прогностическое сопоставление средних значений);
- MICE (многомерный способ заполнения пропущенных данных с помощью цепных уравнений);
- Rpart (рекурсивное разбиение и регрессионные деревья);
- Sequential Hot-Deck и Fractional Hot-Deck Imputation.

Более подробно перечисленные методы рассмотрены в работе [12].

Несмотря на активные исследования в области повышения качества данных в нефтегазовой сфере, восстановлению пропущенных значений в данных не уделяется достаточно внимания. А использование методов, разработанных вне контекста данной сферы, требует их дополнительного анализа на применимость и при необходимости доработки.

### **Интеллектуальный метод заполнения пропущенных значений для использования при формировании данных в фиксациях моделей нефтегазовых скважин**

В основе предлагаемого метода лежит применение метода ближайших соседей. Отличия от применения традиционного метода ближайших соседей заключаются в:

- выборе числа  $k$ -соседей;
- определении ближайших соседей.

Предлагаемый метод особенен тем, что при выявлении соседних дат замеров параметров по скважине дополнительно берется ближайшая последующая дата с непропущенным значением замера по конкретному параметру, которое тоже будет являться соседом без учета расстояния до дат с пропущенными значениями замеров. В ситуации незаполненной последней берется среднее  $k$  предыдущих дней с непустыми значениями. Алгоритм работы метода представлен на Рисунке 1.

Число соседей  $k$  рассчитывается по формуле:

$$k = \lg \left( \frac{f}{c} \right), \quad (1)$$

где  $f$  – количество дат замеров;  $c$  – случайное число, равномерно распределенное на диапазоне [1;10].

На входе алгоритм получает массив  $arr$  из  $M$  элементов, содержащий пропуски, а также переменные  $ind$  для хранения очередного индекса массива и  $T$  – количество соседей. На выходе алгоритм возвращает массив  $arr$ , содержащий восстановленные значения.

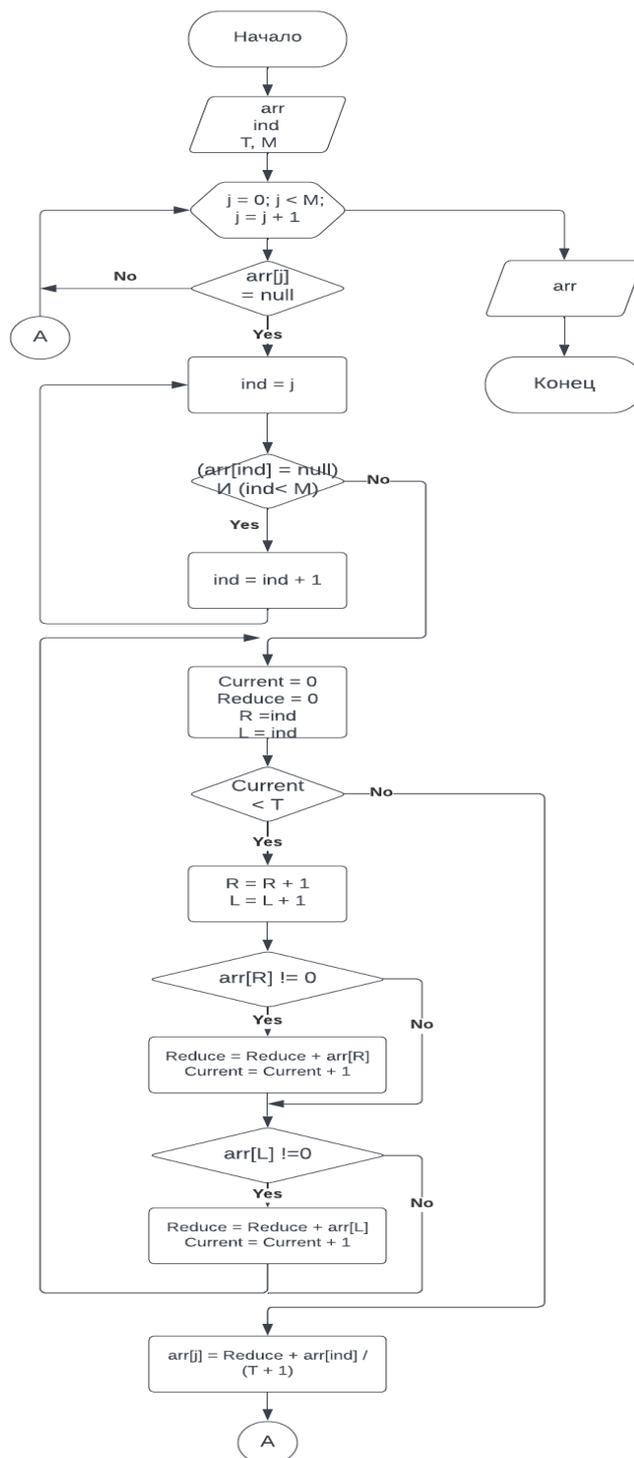


Рисунок 1 – Схема алгоритма работы предлагаемого метода  
Figure 1 – Scheme of the proposed method algorithm

### Описание эксперимента и используемое программное обеспечение

Предлагаемый метод прошел апробацию в ПАО «Газпром нефть». Проанализированы данные по 12 скважинам Вынгапурского месторождения. Выявлено, что в них содержатся пропущенные значения количеством в среднем в диапазоне от 5 до 23 %. Выделена скважина, не содержащая пропущенные значения за период 08.09.2022–29.12.2022 для проведения оценки эффективности предлагаемого метода.

Данные для эксперимента взяты из таблицы БД сервиса «мониторинг УВЧ», хранящей данные по фиксации с момента перевода модели скважины в статус «Реализовано». Фиксация за дату содержит в себе перечень параметров, таких как дебит жидкости, процент обводненности, давление буферное, затрубное давление, газовый фактор и др. Рассмотрим подробнее первый параметр.

Дебит жидкости в нефтяной отрасли – это объем жидкости (обычно нефти или газового конденсата), который может быть добыт из скважины за единицу времени. Он измеряется в баррелях в день или кубических метрах в день. Дебит жидкости является одним из ключевых показателей эффективности работы скважины и может быть использован для принятия решений о дальнейшем развитии нефтяного месторождения. В нашем случае дебит скважины хранится в кубических метрах в день.

Для эксперимента взяты данные параметра дебита жидкости (Liquid Rate) за период 08.09.2022–29.12.2022 (113 значений). Данные представляют собой вещественные значения в диапазоне 88–112 м<sup>3</sup> / д. В рамках эксперимента искусственно добавлены пустые значения в исходный массив данных с целью восстановления предлагаемым методом с последующим увеличением процента пропущенных значений.

Оценка эффективности восстановления данных  $I$  рассчитывается по следующей формуле:

$$I = \sqrt{\frac{1}{N} \sum_{i=1}^N (D_i - O_i)^2}, \quad (2)$$

где  $D_i$  –  $i$ -й элемент массива данных с пропусками  $D$ ;  $O_i$  –  $i$ -й элемент массива восстановленных данных  $O$ ;  $N$  – количество элементов в наборе данных, равное 113.

Разработано веб-приложение, реализующее предлагаемый метод, а также методы восстановления пропущенных значений, такие как простое восстановление, метод ближайших соседей и метод пролонгации, используемый в компании «Газпром нефть». Экранная форма приложения представлена на Рисунке 2.

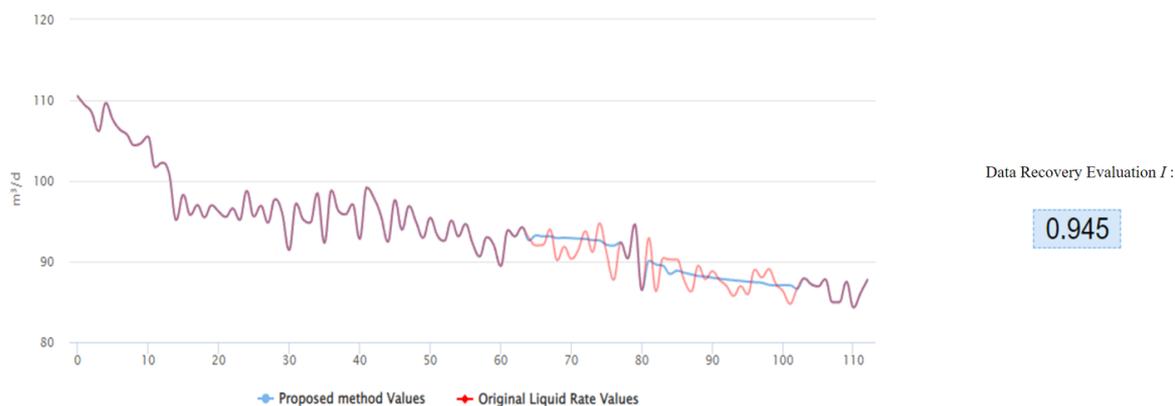


Рисунок 2 – Экранная форма приложения. Результат работы метода  
Figure 2 – Screen form of the application. The result of the method

### Результаты и обсуждение

На Рисунке 3 представлен результат сравнения методов посредством расчета оценки восстановления данных  $I$  для различного процента пропусков в данных значений замеров показателя дебита жидкости.

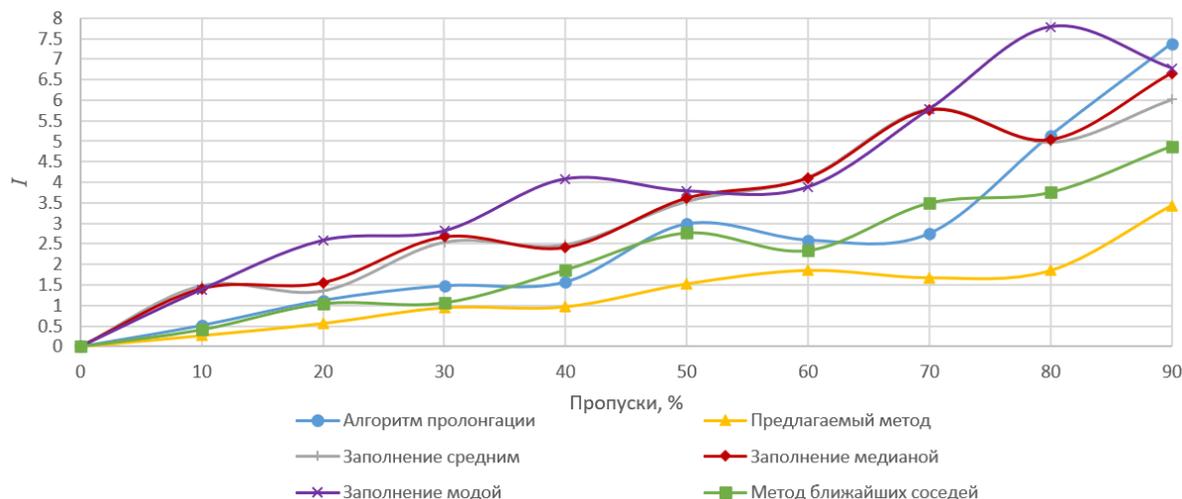


Рисунок 3 – Сравнение методов восстановления пропущенных значений  
Figure 3 – Comparison of the methods for recovering missing values

Анализируя график (Рисунок 3), можно прийти к выводу, что проверка предложенного метода показала его эффективность по сравнению с уже существующими, так как он имеет наименьшее значение  $I$ . Восстановление данных при помощи этого метода дает лучшие результаты, чем при использовании других методов.

### Заключение

В результате выполнения исследования получены следующие результаты:

- проанализированы методы повышения качества формируемых данных в нефтедобывающей области;
- разработан интеллектуальный метод заполнения пропущенных значений для использования при формировании данных в фиксациях моделей нефтегазовых скважин; он является модифицированным методом восстановления данных на основе метода ближайших соседей;
- проведена апробация предлагаемого метода в ПАО «Газпром нефть» на данных по дебиту жидкости 12 скважин Вынгапурского месторождения; результаты апробации метода показали его эффективность по сравнению с уже существующими решениями.

### СПИСОК ИСТОЧНИКОВ

1. Нургалиева З.Д., Латыпова В.А. Методика анализа ремонта нефтегазовых скважин при управлении на основе интеллектуального анализа данных. *Моделирование, оптимизация и информационные технологии*. 2022;10(2). Доступно по: <https://moitvvt.ru/ru/journal/pdf?id=1186> DOI: 10.26102/2310-6018/2022.37.2.017.
2. Хафизова А.У., Латыпова В.А. Оценка эффективности разработки нефтегазовых месторождений с помощью экспертных методов. *Моделирование, оптимизация и информационные технологии*. 2022;10(2). Доступно по: <https://moitvvt.ru/ru/journal/pdf?id=1183> DOI: 10.26102/2310-6018/2022.37.2.018.
3. Малышев А.С., Пашали А.А., Здольник С.Е., Волков М.Г. Удаленный мониторинг механизированного фонда скважин в ОАО «НК «Роснефть». *Научно-технический вестник ОАО «НК «Роснефть»*. 2009;1:23–28.

4. Жильцов В.В., Дударев А.В., Шитов Г.В., Чуви́кова В.В. Решения и развитие интеллектуальной технологии мониторинга и управления механизированным фондом скважин. *Нефтяное хозяйство*. 2006;10:128–130.
5. Жильцов В.В., Федотов А.В. Информационные технологии в проектировании «интеллектуальной» скважины. *Прикладная информатика. Энергетика и рациональное природопользование*. 2013;48(6):15–23.
6. Бахитова Р.Х., Султанов Б.Р. Подход к пространственно-панельному анализу данных нефтяного месторождения. *Инновации и инвестиции*. 2020;6:287–290.
7. Журбич Н.И. Подготовка исходных данных для углубленного анализа нефтегазового месторождения. *Информационные технологии в науке, управлении, социальной сфере и медицине*. 2019;13–19.
8. Qin J., Chen L., Liu Y., Liu C., Feng C., Chen B. A machine learning methodology for diagnosing chronic kidney disease. *IEEE*. 2020;8:4–7.
9. Поцыкайло А.А. Использование метода к-ближайших соседей при распознавании полутоновых изображений. *Известия ЮФУ. Технические науки*. 2011;118(5):258–260.
10. Галеев Э.Е. Кластерный анализ интегрированных корпоративных структур. *Инновации и инвестиции*, 2023;2:139–142.
11. Zliobaite I., Hollmen J. Optimizing regression models for data streams with missing values. *Machine Learning*. 2014;99(1):47–73. DOI: 10.1007/s10994-014-5450-3.
12. Sharifyanov N., Latypova V. A method of filling missing values in data using data mining. *2023 IX International Conference on Information Technology and Nanotechnology (ITNT)*. 2023;1–5. DOI: 10.1109/ITNT57377.2023.10139280.

## REFERENCES

1. Nurgalieva Z.D., Latypova V.A. Oil and gas well repair analysis technique based on data mining in management. *Modeling, Optimization and Information Technology*. 2022;10(2). Available from: <https://moitvvt.ru/ru/journal/pdf?id=1186> DOI: 10.26102/2310-6018/2022.37.2.017 (In Russ.).
2. Khafizova A.U., Latypova V.A. Evaluation of the efficiency of oil and gas field development using expert methods. *Modeling, Optimization and Information Technology*. 2022;10(2). Available from: <https://moitvvt.ru/ru/journal/pdf?id=1183> DOI: 10.26102/2310-6018/2022.37.2.018 (In Russ.).
3. Malyshev A.S., Pashali A.A., Zdol'nik S.E., Volkov M.G. Remote monitoring of mechanized well stock at OJSC Rosneft Oil Company. *Nauchno-tehnicheskij vestnik OJSC "NK Rosneft"*. 2009;1:23–28.
4. Zhil'cov V.V., Dudarev A.V., Shitov G. V., Chuvikova V. V., Solutions and development of intelligent technology for monitoring and managing mechanized well stock. *Neftjanoe hozjajstvo = Oil Industry Journal*. 2006;10:128–130.
5. Zhil'cov V.V., Fedotov A.V Information technologies in the design of "smart" wells. *Prikladnaja informatika. Energetika i racional'noe prirodopol'zovanie = Journal of Applied Informatics*. 2013;48(6):15–23.
6. Bahitova R.H., Sultanov B. R. An approach to spatial-panel analysis of oil field data. *Innovacii i investicii*. 2020;6:287–290.
7. Zhurbich N.I. Preparation of initial data for in-depth analysis of the oil and gas field. *Informacionnye tehnologii v nauke, upravlenii, social'noj sfere i medicine*. 2019;13–19.
8. Qin J., Chen L., Liu Y., Liu C., Feng C. and Chen B. A machine learning methodology for diagnosing chronic kidney disease. *IEEE*. 2020;8:4–7.

9. Посыкайло А.А. Using the method of k-nearest neighbors in the recognition of grayscale images. *Izvestiya YuFU. Tekhnicheskie nauki = Izvestija JuFU. Engineering Sciences*. 2011;118(5):258–260.
10. Galeev J.E. Cluster analysis of integrated corporate structures. *Innovacii i investicii*. 2023;2:139–142.
11. Zliobaite I., Hollmen J. Optimizing regression models for data streams with missing values. *Machine Learning*. 2014;99(1):47–73, DOI: 10.1007/s10994-014-5450-3.
12. Sharifyanov N., Latypova V. A method of filling missing values in data using data mining. *2023 IX International Conference on Information Technology and Nanotechnology (ITNT)*. 2023;1–5. DOI: 10.1109/ITNT57377.2023.10139280.

### ИНФОРМАЦИЯ ОБ АВТОРАХ / INFORMATION ABOUT THE AUTHORS

**Шарифьянов Нияз Вакилевич**, магистрант, Уфимский университет науки и технологий, разработчик, «Газпромнефть-Цифровые решения», Уфа, Российская Федерация.  
*e-mail*: [sharifyanov.nv@mail.ru](mailto:sharifyanov.nv@mail.ru)

**Niyaz Vakilevich Sharifyanov**, Master's Student, Ufa University of Science and Technology, Developer at Gazpromneft-Cifrovye resheniya, Ufa, the Russian Federation.

**Латыпова Виктория Александровна**, кандидат технических наук, доцент кафедры автоматизированных систем управления, Уфимский университет науки и технологий, Уфа, Российская Федерация.  
*e-mail*: [vikvaphoto@yandex.ru](mailto:vikvaphoto@yandex.ru)  
ORCID: [0000-0003-3063-105X](https://orcid.org/0000-0003-3063-105X)

**Viktoriya Aleksandrovna Latypova**, Candidate of Technical Sciences, Associate Professor at the Department of Automated Management Systems, Ufa University of Science and Technology, Ufa, the Russian Federation.

*Статья поступила в редакцию 19.05.2023; одобрена после рецензирования 13.06.2023; принята к публикации 16.06.2023.*

*The article was submitted 19.05.2023; approved after reviewing 13.06.2023; accepted for publication 16.06.2023.*