

# Определение зон стабильности газовых гидратов при помощи методов машинного обучения

Гулиев Р.З.<sup>1</sup>, Еремин Н.А.<sup>1,2</sup>, Зиганшин А.Р.<sup>1</sup>

<sup>1</sup>Северный (Арктический) федеральный университет имени М.В. Ломоносова, Архангельск, Россия, <sup>2</sup>РГУ нефти и газа (НИУ) имени И.М. Губкина, Москва, Россия  
r.guliev@narfu.ru

## Аннотация

В настоящее время возрастает интерес к газогидратам как к альтернативным источникам энергии, способу транспортировки и хранения природного газа. Актуальной проблемой является прогнозирование образования газогидратов, следовательно, существует потребность в разработке эффективной технологии, способной прогнозировать термобарические условия гидратообразования.

Целью данной работы является создание алгоритма машинного обучения для прогнозирования условий гидратообразования. Задачи, которые были поставлены: сбор эмпирических данных о зонах стабильности газогидратов, анализ данных алгоритмами машинного обучения, построение алгоритма, способного достаточно точно прогнозировать зоны стабильности газогидратов.

## Материалы и методы

В ходе данной работы будет построен алгоритм машинного обучения для прогнозирования температуры и структуры гидратообразования, также модель машинного обучения проведет анализ исходных данных. В основе модели будет лежать метод случайного леса.

Исходные данные для обучения и тестирования модели будут взяты из открытых источников.

## Ключевые слова

газовые гидраты, машинное обучение, прогнозирование, модель, термобарические условия

## Для цитирования

Гулиев Р.З., Еремин Н.А., Зиганшин А.Р. Определение зон стабильности газовых гидратов при помощи методов машинного обучения // Экспозиция Нефть Газ. 2023. № 6. С. 57–61. DOI: 10.24412/2076-6785-2023-6-57-61

Поступила в редакцию: 06.09.2023

AUTOMATION

UDC 622.279.51 | Original Paper

## Determination of stability zones of gas hydrates using machine learning methods

Guliev R.Z.<sup>1</sup>, Eremin N.A.<sup>1,2</sup>, Ziganshin A.R.<sup>1</sup>

<sup>1</sup>Northern (Arctic) federal university, Arkhangelsk, Russia, <sup>2</sup>Gubkin University, Moscow, Russia  
r.guliev@narfu.ru

## Abstract

Currently, there is a growing interest in gas hydrates as an alternative energy source, a method of transporting and natural gas storage. An urgent problem is the prediction of the gas hydrates formation, therefore, there is a need to develop an effective technology capable of predicting the thermobaric conditions of hydrate formation.

The purpose of this work is to create a machine learning algorithm for predicting hydrate formation conditions. The tasks that were set are the collection of empirical data on the stability zones of gas hydrates, the analysis of data by machine learning algorithms, the developing of the model capable of accurately predicting the stability zones of gas hydrates.

## Materials and methods

In the course of this work, a machine learning algorithm will be built to predict the temperature and structure of hydrate formation and initial data will be analyzed by the model. The model will be based on the

Random Forest method. The initial data for training and testing the algorithm will be taken from open sources.

## Keywords

gas hydrates, machine learning, forecasting, thermobaric conditions

## For citation

Guliev R.Z., Eremin N.A., Ziganshin A.R. Determination of stability zones of gas hydrates using machine learning methods. Exposition Oil Gas, 2023, issue 6, P. 57–61. (In Russ). DOI: 10.24412/2076-6785-2023-6-57-61

Received: 06.09.2023

## Введение

При проведении глубоководных буровых работ образование газогидратов является угрозой для безопасности на производстве, потому что диссоциация гидрата метана может привести к выбросу оборудования или взрыву [3]. Кроме этого, метан является парниковым газом, который

способен выделиться с морского дна при температуре морской воды или подводного оползня. Поэтому точная оценка запасов газогидратов на морском дне и определение зон их стабильности позволят вовремя замечать необратимые изменения морского дна. Следует также учесть главный интерес к изучению газогидратов — это

возможность использовать их в качестве источника энергии.

Газовые гидраты представляют собой группу нестехиометрических ледоподобных кристаллических соединений, образующихся в результате сочетания воды и гостевых молекул подходящего размера при низких температурах и повышенных давлениях.

В газогидратной решетке молекулы воды образуют водородно-связанные клеточные структуры [3].

Общая формула для всех газовых гидратов:

$$M \times nH_2O, \quad (1)$$

где  $M$  — молекула гидратообразующего газа;  $n$  — число молекул воды, приходящихся на одну включенную молекулу газа, ввиду того, что эти вещества имеют переменный состав, принимает значения от 5,75 до 17.

Полости газогидратов имеют правильную геометрическую форму, которая встречается в трех видах (рис. 1):

- первый вид ( $5^{12}$ ) представляет собой многогранник, содержащий 12 граней по 5 ребер;
- второй вид ( $5^{12}6^2$ ) является многогранником, который содержит 12 граней по 5 ребер и 2 грани по 6 ребер;
- третий вид ( $5^{12}6^4$ ) построен из 12 граней по 5 ребер и 4 граней по 6 ребер.

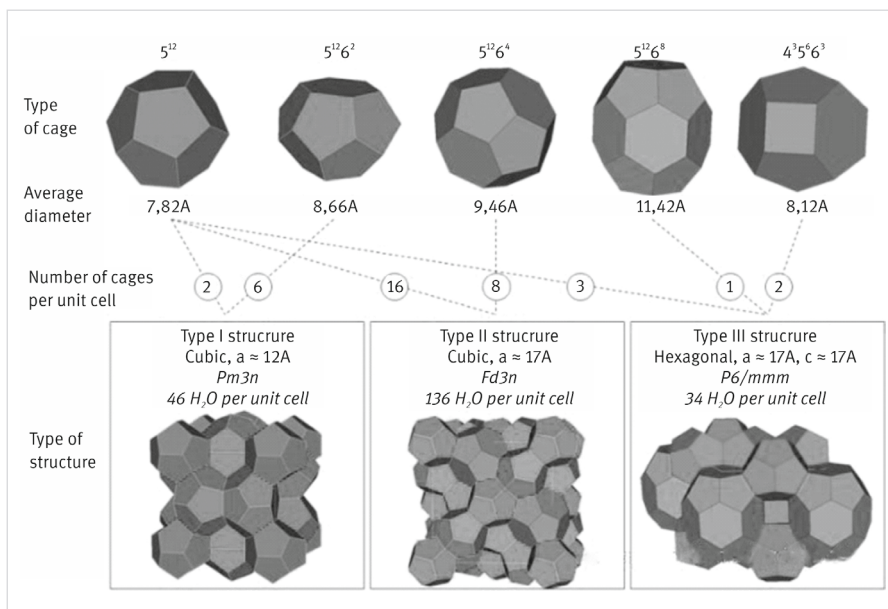


Рис. 1. Схема образования различных структур газогидратов  
Fig. 1. Scheme of formation of various structures of gas hydrate

### Методология исследования

Для прогнозирования зон стабильности газогидратов в данной работе применяется алгоритм, который получает на вход значение температуры и предсказывает давление.

В качестве исходных данных принимались экспериментальные термобарические условия из литературы [1, 2, 5]. Обучение машинного алгоритма осуществляется на данных, в которых содержатся все фазовые состояния газогидратов.

Затем модель тестируется и проверяется на данных, основанных на одной из фаз газового гидрата.

Компьютерная программа обучается на основе опыта  $E$  по отношению к некоторой задаче  $T$  и некоторой оценке производительности  $R$ , если ее производительность на  $T$ , измеренная посредством  $R$ , улучшается с опытом  $E$ .

#### Методы машинного обучения

Методы машинного обучения способны справляться с широким спектром задач, но в данной работе рассматривается задача восстановления регрессии или же прогнозирования, поэтому будут рассматриваться и применяться методы, способные решить поставленную задачу.

Задачу обучения по прецедентам при  $Y = R$  принято называть задачей восстановления регрессии. Основные обозначения

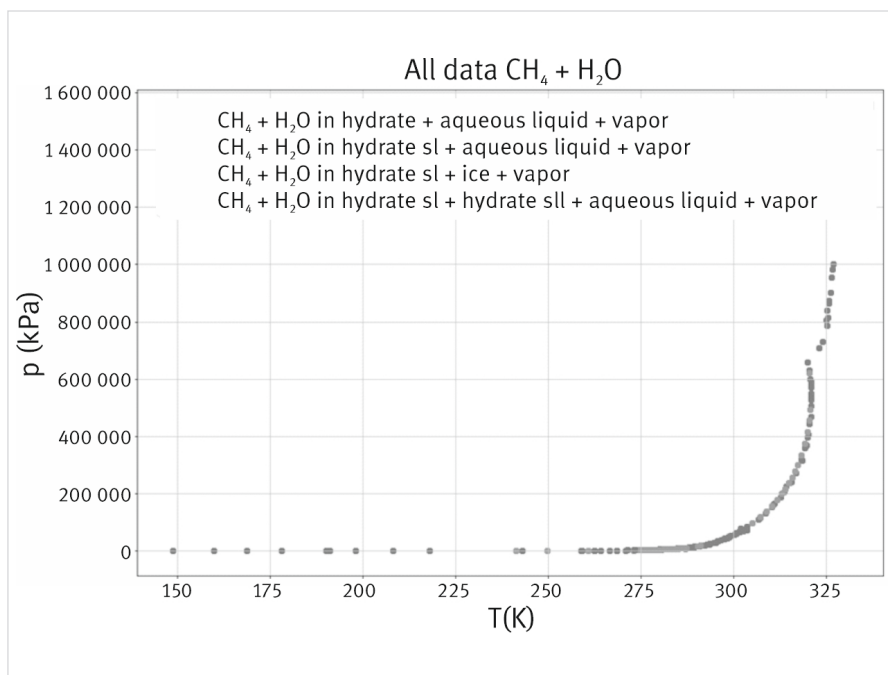


Рис. 2. Зависимость давления от температуры в различных структурах газогидратов  
Fig. 2. Dependence of pressure on temperature in various structures of gas hydrates

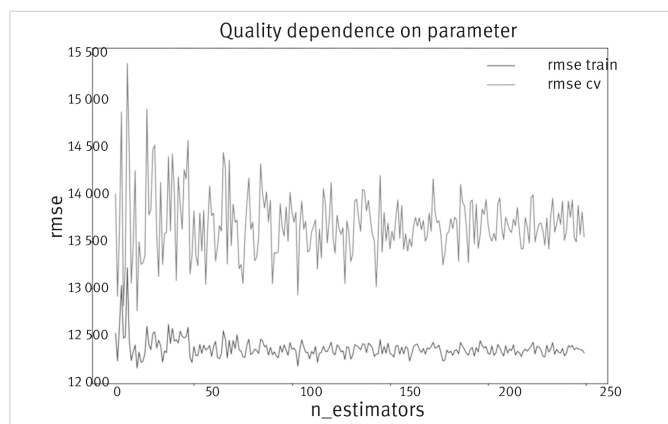


Рис. 3. Зависимость значения RMSE от параметра n\_estimators  
Fig. 3. Dependence of the RMSE value on the n\_estimators parameter

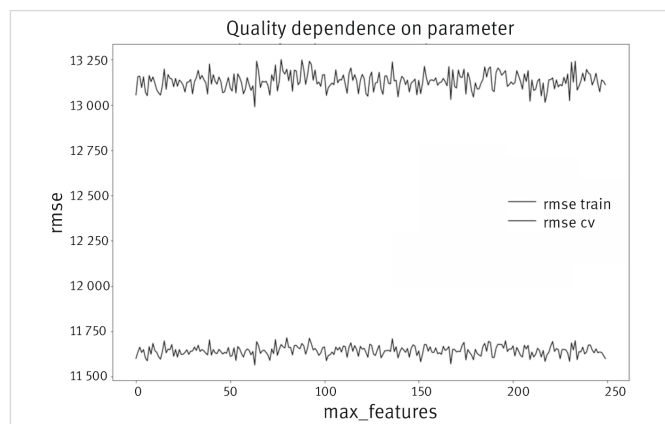


Рис. 4. Зависимость RMSE от параметра max\_features  
Fig. 4. Dependence of RMSE on the max\_features parameter

остаются прежними. Задано пространство объектов  $X$  и множество возможных ответов  $Y$ . Существует неизвестная целевая зависимость:

$$y^* = X \rightarrow Y, \quad (2)$$

значения которой известны только на объектах обучающей выборки  $X_l = (x_p, y_p)^{l=1, \dots, n}$ ,  $y_i = y(x_i)$ .

Требуется построить алгоритм, который в данной задаче принято называть функцией регрессии  $X \rightarrow Y$ , аппроксимирующий целевую зависимость  $y^*$ .

К методам восстановления регрессии можно отнести: линейную (полиномиальную) регрессию, метод К-ближних соседей и лес случайных решений.

В данной работе рассмотрен алгоритм леса случайных решений.

#### Случайный лес

Случайный лес — один из самых эффективных алгоритмов машинного обучения и является одним из немногих универсальных алгоритмов. Универсальность заключается, во-первых, в том, что он хорош во многих задачах, во-вторых, в том, что есть случайные леса для решения задач классификации, регрессии, кластеризации, поиска аномалий, селекции признаков и т.д. [4, 6].

#### Метрики для оценки алгоритмов машинного обучения

Метрики необходимы для измерения и сравнения производительности алгоритмов машинного обучения. Они отражают оценку различных характеристик в результатах работы алгоритмов.

#### Средняя абсолютная ошибка

Средняя абсолютная ошибка (или MAE) представляет собой сумму абсолютных различий между прогнозами и фактическими значениями:

$$MAE = \frac{\sum_{i=1}^n |y_i - x_i|}{n},$$

где  $y_i$  — это предсказанное значение,  $x_i$  — это истинное значение,  $n$  — количество примеров для обучения.

#### Среднеквадратичная ошибка

Среднеквадратичная ошибка (или RMSE) представляет собой квадратный корень из среднеквадратической ошибки (MSE). Среднеквадратическая ошибка похожа на абсолютную ошибку в том, что она дает общее представление о величине ошибки. MSE находится по формуле:

$$MSE = \frac{1}{n} \sqrt{\sum_{i=1}^n (y_i - x_i)^2},$$

где  $y_i$  — это предсказанное значение,  $x_i$  — это истинное значение,  $n$  — количество примеров для обучения.

#### Коэффициент детерминации

Коэффициент детерминации (R Squared) указывает на точность соответствия спрогнозированных значений и фактических. Это значение изменяется от 0 до 1, что характеризует в долях эффективность работы алгоритма. R Squared вычисляется по формуле:

$$R^2 = 1 - \frac{SSE}{SST},$$

где  $SSE$  — это сумма значений квадратичной функции потерь;  $SST$  — полная сумма квадратов, или, другими словами, это дисперсия переменной отклика.

#### Результаты исследований

Для качественной работы алгоритма необходимо произвести кросс-валидацию, то есть разбить исходные данные на обучающую и на тестовую выборки. Также для работы алгоритма необходимо правильно подобрать значения переменных, которые влияют на процесс обучения.

Показателями эффективности работы алгоритма будут являться следующие метрики:

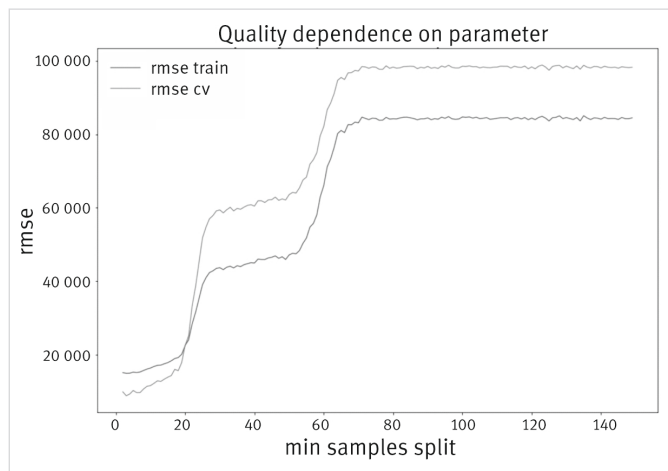


Рис. 5. Зависимость RMSE от параметра min\_samples\_split  
Fig. 5. RMSE dependency on the min\_samples\_split parameter

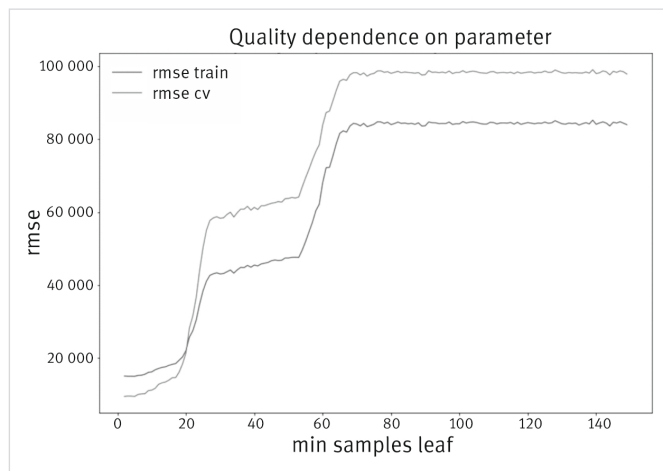


Рис. 6. Зависимость RMSE от параметра min\_samples\_leaf  
Fig. 6. RMSE dependency on the min\_samples\_leaf parameter

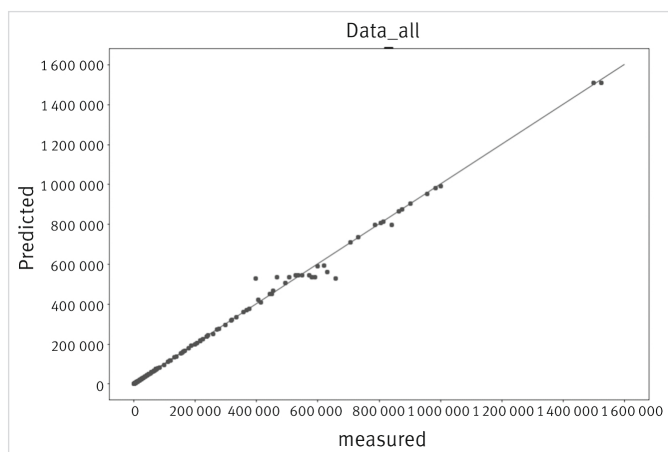


Рис. 7. Сравнение реального и прогнозируемого значения давления для всех данных в выборке  
Fig. 7. Comparison of the real and predicted pressure value for all data in the sample

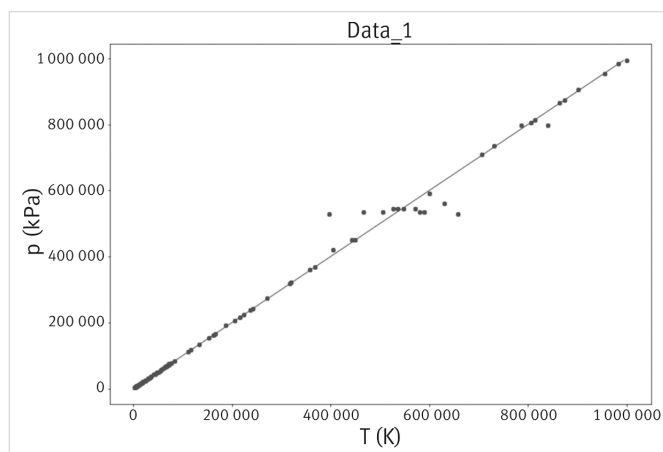


Рис. 8. Сравнение реального и прогнозируемого значения давления для состава  $CH_4 + H_2O$  in Hydrate + Aqueous Liquid + Vapor  
Fig. 8. Comparison of the actual and predicted pressure value for  $CH_4 + H_2O$  in Hydrate + Aqueous Liquid + Vapor

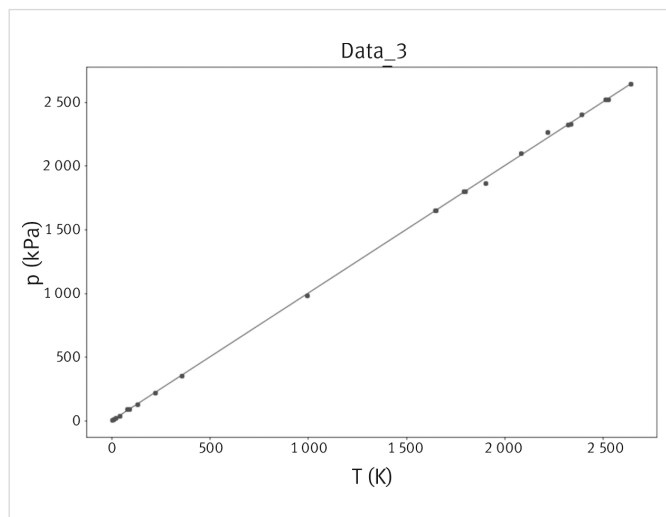


Рис. 9. Сравнение реального и прогнозируемого значения давления для состава  $CH_4 + H_2O$  in Hydrate sl + Aqueous Liquid + Vapor  
Fig. 9. Comparison of the real and predicted pressure value for the composition  $CH_4 + H_2O$  in Hydrate sl + Aqueous Liquid + Vapor

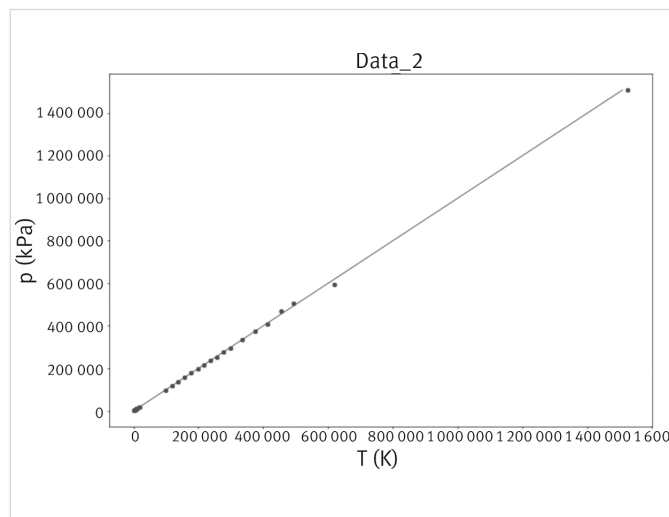


Рис. 10. Сравнение реального и прогнозируемого значения давления для состава  $CH_4 + H_2O$  in Hydrate sl + Ice + Vapor  
Fig. 10. Comparison of the actual and predicted pressure value for the composition  $CH_4 + H_2O$  in Hydrate sl + Ice + Vapor

среднеквадратичная ошибка и коэффициент детерминации.

Число «деревьев» оказывает значительное влияние на качество алгоритма. Наименьшее значение ошибки мы получаем при значении количества «деревьев» — 20.

Число признаков для расщепления отвечают за однообразие «деревьев», чем больше значение, тем больше время работы алгоритма и тем однообразнее «деревья». Наименьшее значение ошибки мы получаем при значении количества признаков расщепления — 18.

Минимальное число объектов, при котором выполняется расщепление, влияет на время выполнения алгоритма и на качество обучения.

Ограничение на число объектов в листьях оказывает такое же влияние на алгоритм обучения, как и предыдущий параметр. Значение было выставлено значением по умолчанию и равнялось пяти.

Таким образом были подобраны значения параметров для корректной работы алгоритма и его качественного обучения и тестов.

Далее модель была обучена на общей выборке для последующего применения на данных, сгруппированных по составу. На графиках ниже отображено соответствие предсказанных значений к целевым, то есть чем ближе точки располагаются к функции  $y = x$ , тем точнее алгоритм выполнил предсказание.

### Итоги

Применение модели машинного обучения показало высокую эффективность. В рамках текущего исследования были достигнуты точности в 75,718 % при прогнозировании температуры образования гидрата и 84,865 % при прогнозировании структуры образующегося гидрата различных газов.

### Выводы

В настоящем исследовании построены математические модели, основанные на алгоритмах машинного обучения. Построение алгоритмов и их тестирование осуществлялось на экспериментальных данных термобарических свойств газовых гидратов, собранных из литературы. Простота и допустимая точность результатов прогнозирования свидетельствуют о том, что данная технология может применяться для определения зон стабильности газогидратов. Но были выявлены недостатки в исходных данных, которые учитывают не все параметры условия образования гидратов, содержат ошибочные значения и не обладают исчерпывающей информацией о происходящих процессах.

Дальнейший сбор данных и совершенствования алгоритмов могут позволить эффективно оценивать влияние различных ингибиторов на смещение зон стабильности и их концентрации в заданных условиях. Такая гибкость модели позволит эффективно эксплуатировать трубопроводы, технологические линии и с доработками оценивать диссоциацию гидратов газа в газогидратных месторождениях.

### Литература

1. Chapoy A., Mohammadi A.H., Richon D. Predicting the hydrate stability zones of natural gases using artificial neural networks. Oil & Gas science and technology. 2007. Vol. 62, issue 5, P. 701–706. (In Eng).
2. Abooali D., Khomehchi E. New predictive method for estimation of natural gas hydrate formation temperature using genetic programming. Neural Computing and Applications, 2019, Vol. 31, issue 7, P. 2485–2494. (In Eng).
3. Carroll J.J. Natural gas hydrates: a guide for engineers 2nd ed. Amsterdam; Boston: Elsevier, 2009, 276 p. (In Eng).
4. Clayton C.R.I., Priest J.A., Best A.I. The effects of disseminated methane hydrate on the dynamic stiffness and damping of a sand. Geotechnique, 2005, Vol. 55, issue 6, P. 423–434. (In Eng).
5. Elgibaly A.A., Elkamel A.M. A new correlation for predicting hydrate formation conditions for various gas mixtures and inhibitors. Fluid phase equilibria, 1998, Vol. 152, issue 1, P. 23–42. (In Eng).
6. Eymold W.K., Frederick J.M., Nole M., Phrampus B.J., Wood W.T. Prediction of gas hydrate formation at Blake Ridge using machine learning and probabilistic reservoir simulation. Geochemistry, Geophysics, Geosystems, Vol. 22, issue 4, 22 p. (In Eng).

## ENGLISH

### Results

The use of a machine learning model has shown high efficiency. Within the framework of the current study, an accuracy of 75,718 % was achieved in predicting the temperature of hydrate formation and 84,865 % in predicting the structure of the resulting hydrate of various gases.

### Conclusions

In this study, mathematical models based on machine learning algorithms are constructed. The construction of algorithms and their testing was carried out on experimental data on the thermobaric properties of gas hydrates collected from the literature. The simplicity and permissible accuracy of the forecasting results indicate that this technology can be

used to determine the stability zones of gas hydrates. But shortcomings were identified in the initial data, which do not take into account all the parameters of the conditions for the formation of hydrates, contain erroneous values and do not have comprehensive information about the processes taking place.

Further data collection and improvements in algorithms can make it possible to effectively assess the effect of various inhibitors on the displacement of stability zones and from concentration under given conditions. Such flexibility of the model will make it possible to efficiently operate pipelines, technological lines and, with modifications, assess the dissociation of gas hydrates in gas hydrate fields.

## References

1. Chapoy A., Mohammadi A.H., Richon D. Predicting the hydrate stability zones of natural gases using artificial neural networks. Oil & Gas science and technology, 2007, Vol. 62, issue 5, P. 701–706. (In Eng).
2. Abooli D., Khomechi E. New predictive method for estimation of natural gas hydrate formation temperature using genetic programming. Neural Computing and Applications, 2019, Vol. 31, issue 7, P. 2485–2494. (In Eng).
3. Carroll J.J. Natural gas hydrates: a guide for engineers 2nd ed. Amsterdam; Boston: Elsevier, 2009, 276 p. (In Eng).
4. Clayton C.R.I., Priest J.A., Best A.I. The effects of disseminated methane hydrate on the dynamic stiffness and damping of a sand. Geotechnique, 2005, Vol. 55, issue 6, P. 423–434. (In Eng).
5. Elgibaly A.A., Elkamel A.M. A new correlation for predicting hydrate formation conditions for various gas mixtures and inhibitors. Fluid phase equilibria, 1998, Vol. 152, issue 1, P. 23–42. (In Eng).
6. Eymold W.K., Frederick J.M., Nole M., Phrampus B.J., Wood W.T. Prediction of gas hydrate formation at Blake Ridge using machine learning and probabilistic reservoir simulation. Geochemistry, geophysics, geosystems, Vol. 22, issue 4, 22 p. (In Eng).

## ИНФОРМАЦИЯ ОБ АВТОРАХ | INFORMATION ABOUT THE AUTHORS

**Гулиев Рамиль Зафарович**, старший преподаватель кафедры бурения скважин, разработки нефтяных и газовых месторождений, Северный (Арктический) федеральный университет имени М.В. Ломоносова, Архангельск, Россия  
Для контактов: [r.guliev@narfu.ru](mailto:r.guliev@narfu.ru)

**Еремин Николай Александрович**, д.т.н., профессор, научный руководитель кафедры бурения скважин, разработки нефтяных и газовых месторождений, Северный (Арктический) федеральный университет имени М.В. Ломоносова, Архангельск, Россия; профессор, РГУ нефти и газа (НИУ) имени И.М. Губкина, Москва, Россия

**Зиганшин Артур Радикович**, студент 4 курса кафедры бурения скважин, разработки нефтяных и газовых месторождений, Северный (Арктический) федеральный университет имени М.В. Ломоносова, Архангельск, Россия

**Guliev Ramil Zafarovich**, senior lecturer, department of well drilling, oil and gas field development, Northern (Arctic) federal university, Arkhangelsk, Russia

Corresponding author: [r.guliev@narfu.ru](mailto:r.guliev@narfu.ru)

**Eremin Nikolai Alexandrovich**, doctor of technical sciences, professor, scientific director of the department of well drilling, oil and gas field development, Northern (Arctic) federal university, Arkhangelsk, Russia; professor, Gubkin University, Moscow, Russia

**Ziganshin Artur Radikovich**, 4th year bachelor's student, department of well drilling, oil and gas field development, Northern (Arctic) federal university, Arkhangelsk, Russia



**03–07 ИЮНЯ 2024, ТЮМЕНЬ,**  
отель «DOUBLE TREE BY HILTON TYUMEN»,  
конференц-зал «Diamond», конференц-зал «Golden»

«Инновационные решения в области КРС, ПНП, ГНКТ, внутрискважинные работы и супервайзинг в горизонтальных и разветвленных скважинах. Контроль скважин. Управление скважиной при ГНВП. Противофонтанная безопасность»

**15–19 ИЮЛЯ 2024, ТЮМЕНЬ,**  
отель «ЛЕТОЛЕТО», конференц-зал «ИюньИюль»

«Инновационные технологические решения при эксплуатации и ремонте скважин»

**02–06 СЕНТЯБРЯ 2024, МИНЕРАЛЬНЫЕ ВОДЫ, БЕЛОКУРИХА, АЛТАЙСКИЙ КРАЙ**

«Бурение, освоение, испытания, ремонт и эксплуатация горизонтальных скважин. Инновации в области добычи нефти и газа. Промышленная безопасность на ОПО нефтегазовой отрасли»

**07–11 ОКТЯБРЯ 2024, ТЮМЕНЬ,** отель «DOUBLE TREE BY HILTON TYUMEN»,  
конференц-зал «Diamond», конференц-зал «Golden»

«Кадровый ресурс — потенциал повышения эффективности и безопасности компании. Оценка квалификаций и развитие персонала»

**+7 3452 520-958**

бронирование участия в конференциях  
[academy.intechnol.com](http://academy.intechnol.com)

Генеральный информационный партнер

**ЭКСПОЗИЦИЯ  
НЕФТЬ ГАЗ**