**Научная статья / Research Article**
Механика жидкости, газа и плазмы

https://doi.org/10.21684/2411-7978-2025-11-1-89-111
**УДК** 532.54; 550.832

# Прогнозирование накопления жидкости в промысловых газопроводах на основе машинного обучения

**Павел Александрович Крылов** [1]✉,
**Наиль Габсалямович Мусакаев** [1, 2]

[1] Тюменский государственный университет, Тюмень, Россия
[2] Тюменский филиал Института теоретической и прикладной механики им. С.А. Христиановича СО РАН, Тюмень, Россия
✉ E-mail для переписки: paul.kryloff@yandex.ru

**Аннотация.** Накопление жидкости в промысловых газопроводах является достаточно встречающейся проблемой, которая нарушает стабильность потока. В процессе эксплуатации, под действием рельефа местности и условий работы происходит постепенное отложение воды и газоконденсата с восходящих участков трубопровода на нисходящие, что, в свою очередь, снижает эффективность транспортировки, вызывает рост потерь давления, пульсации давления, способствует протеканию процессов коррозии и гидратообразования.

Из-за сложности многофазного течения механизм накопления жидкости до сих пор остается спорным. В настоящее время, большинство методик, предсказывающих накопление, являются полуэмпирическими и не обладают достаточной точностью. Развитие машинного обучения и технологий искусственного интеллекта предоставляют широкий спектр возможностей для анализа, выявления потенциальных зависимостей и предсказания поведения данных.

Целью данной работы является получение многофакторной модели прогнозирования накопления жидкости в промысловых газопроводах с высокой способностью к обобщению и прогнозу.

Основываясь на статистических данных эксплуатации трубопроводов нефтегазоконденсатных месторождений Западной и Восточной Сибири, был создан массив данных для расчета параметров в динамическом симуляторе неустановившихся многофазных потоков, необходимых для проведения машинного обучения. После предварительной обработки данных было проведено обучение модели по алгоритмам с учителем и проведено их дальнейшее сравнение,

включая методы: логистической регрессии (LR), линейного дискриминантного анализа (LDA), K-ближайших соседей (KNN), дерева принятия решений (CART), наивного байесовского классификатора (NB), линейных опорных векторов (LSVC), опорных векторов (SVC), «бэггинга» (BG), «случайного леса» (RF), классификатора экстремально рандомизированных деревьев (ET), адаптивного «бустинга» (AB), градиентного «бустинга» (GB), экстремального градиентного «бустинга» (XGB) и многослойного перцептрона (MLP), наиболее оптимальными из которых оказались алгоритмы «дерево принятия решений» и «K ближайших соседей». Данные модели были оптимизированы методами «кросс-валидации», затем обучены на тренировочных данных и протестированы.

В ходе работы было установлено множество различных комбинаций работы трубопровода без накопления, определена степень важности различных параметров на протекание процесса накопления. Разработанная модель может стать полезным средством для анализа и локализации процесса накопления жидкости, обеспечивая более упрощенное и всестороннее прогнозирование по сравнению с другими моделями.

**Ключевые слова:** газ, конденсат, газопровод, шлейф, сети сбора, накопление жидкости, машинное обучение.

**Благодарности:** Работа выполнена в рамках государственного задания (№ госрегистрации 124021500017-5).

**Цитирование:** Крылов П. А., Мусакаев Н. Г. 2025. Прогнозирование накопления жидкости в промысловых газопроводах на основе машинного обучения // Вестник Тюменского государственного университета. Физико-математическое моделирование. нефть, газ, энергетика. Том 11. № 1 (41). С. 89–111. https://doi.org/10.21684/ 2411-7978-2025-11-1-89-111

Поступила 07.04.2025; одобрена 19.04.2025; принята 21.04.2025

# Prediction of liquid accumulation in field gas pipelines based on machine learning

**Pavel A. Krylov**[1]✉,
**Nail G. Musakaev**[1, 2]

[1] University of Tyumen, Tyumen, Russia
[2] Tyumen Branch of the Khristianovich Institute of Theoretical and Applied Mechanics of the Siberian Branch of the Russian Academy of Sciences, Tyumen, Russia
✉ Corresponding author: paul.kryloff@yandex.ru

**Abstract.** Liquid accumulation in field gas pipelines is a common problem that disrupts flow stability. In the process of operation, under the influence of terrain and working conditions, there is a gradual deposition of water and gas condensate from upstream to downstream sections of the pipeline, which in turn reduces the efficiency of transportation, causes an increase in pressure losses, pressure pulsations, promotes corrosion and hydrate formation processes.

Due to the complexity of multiphase flow, the mechanism of fluid accumulation is still controversial. Currently, most of the techniques that predict accumulation are semi-empirical and do not have sufficient accuracy. Advances in machine learning and artificial intelligence technologies provide a wide range of possibilities for analyzing, identifying potential dependencies and predicting data behavior.

The aim of this paper is to obtain a multifactor prediction model for liquid accumulation in field gas pipelines with high generalization and prediction ability.

Based on the statistical data of pipeline operation in oil and gas condensate fields of Western and Eastern Siberia, a data set was created to calculate the parameters in the dynamic simulator of unsteady multiphase flows required for machine learning. After data preprocessing, the model was trained using teacher learning algorithms and further compared, including methods: logistic regression (LR), linear discriminant analysis (LDA), K-nearest neighbors (KNN), decision tree (CART), naive Bayesian classifier (NB), linear support vectors machines (LSVC), support vectors machines (SVC), bagging (BG), random forest (RF), extreme randomized trees classifier (ET), adaptive boosting (AB), gradient boosting (GB), extreme gradient boosting (XGB), and multilayer perceptron (MLP), the most optimal of which were found to be the «decision tree» and «K-nearest neighbors» algorithms. These models were optimized using «cross-validation» methods, then trained on training data and tested.

Many different combinations of non-accumulation pipeline operation were established, and the degree of importance of various parameters on the accumulation process was determined. The developed model can be a useful tool for analyzing and localizing the fluid accumulation process, providing a more simplified and comprehensive prediction than other models.

**Keywords:** gas, condensate, pipeline, plume, gathering networks, liquid accumulation, machine learning.

**Acknowledgements:** The research was carried out within the state assignment of Ministry of Science and Higher Education of the Russian Federation (project No. 124021500017-5).

**Citation:** Krylov, P. A., & Musakaev, N. G. (2025). Prediction of liquid accumulation in field gas pipelines based on machine learning.*Tyumen State University Herald. Physical and Mathematical Modeling. Oil, Gas, Energy*, *11*(1), 89-111. https://doi.org/10.21684/2411-7978-2025-11-1-89-111

Received Apr. 7, 2025; Reviewed Apr. 19, 2025; Accepted Apr. 21, 2025

## Введение.

С увеличением темпов развития газовой промышленности и объемов перекачиваемого газа возникает необходимость в повышении эффективности и безопасности работы газопроводов. При понижении температуры эксплуатации нередки случаи выпадения газоконденсата внутри газопроводов [Борисевич и др., 2022]. В свою очередь, накопление жидкости в газопроводах может повлечь за собой рост потерь давления, пульсации давления [Нуруллаев, Очилов, 2017], усугубление процессов коррозии [Ватузов и др., 2023] и гидратообразования [Бузников и др., 2016]. Кроме того, многофазный поток подвержен образованию пробкового режима при чрезмерном накоплении жидкости в трубопроводе, который способен нарушить целостность трубопровода [Лаурье, 2011]. По этой причине необходимо принимать действия по предупреждению накопления жидкости и изучать законы, по которым данный процесс происходит. Изучение данной проблематики осуществляется не только в рамках нашей страны, но и за рубежом [Rastogi, Fan, 2020].

В настоящий момент существует множество методик прогнозирования накопления жидкости, однако большинство из них являются полуэмпирическими и не располагают высокой точностью, также определяющим критерием для них, в подавляющем большинстве, является скорость газа [Алиев и др., 1978; Клапчук, Елин, 1979; Краснов, 2018; Кутателадзе, Накоряков, 1984; Одишария, Точигин, 1988]. Сложность протекания физического процесса затрудняет создание корректной математической модели. Современные коммерческие симуляторы неустановившегося многофазного потока позволяют моделировать большое количество сценариев, тем не менее, все они могут моделировать задачу под конкретно взятые условия и требуют постоянного перерасчета, больших временных и трудозатрат. По этой причине возникает необходимость в получении нового метода, способного быстро и эффективно прогнозировать накопление жидкости в газопроводах.

С появлением методов машинного обучения открылся широкий горизонт для анализа данных и предсказания их поведения. В последние годы наблюдается высокий интерес к применению алгоритмов машинного обучения для решения задач нефтегазовой отрасли [Цховребов, 2024]. В рамках работы было проведено обучение модели по предобработанным данным проведения симуляции по алгоритмам обучения с учителем и проведено их дальнейшее сравнение с выбором модели с наилучшими показателями качества, включая методы логистической регрессии, линейного дискриминантного анализа, K-ближайших соседей, дерева принятия решений, наивного байесовского классификатора и т.п.

## Постановка задачи и методология.

Накопление жидкости в трубопроводах происходит по причине действия силы тяжести. После того как слой жидкости становится плоским на восходящем участке трубопровода, он начинает перемещаться в нисходящую часть трубопровода. В данном процессе перепад давления увеличивается по мере расширения газа, вследствие чего увеличива-

ется объемный расход и скорость газовой фазы, поэтому увеличивается коэффициент скольжения между газовой и жидкой фазами, что снижает величину объемного расхода жидкой фазы. В свою очередь, на нисходящем участке трубопровода не может происходить накопление жидкости ввиду того, что скорость потока уменьшается под действием плавучести, что уменьшает коэффициент скольжения между двумя фазами и увеличивает способность газовой фазы пропускать через себя жидкую, поэтому следует рассматривать только случай восходящего участка. Предлагаемая методология, отраженная в данной работе, представлена на рисунке 1.


**Рис. 1.** Блок-схема методологии, представленной в данной работе
**Fig. 1.** Block diagram of the methodology presented in this article

Трубопроводы могут иметь различные геометрические характеристики, включая их диаметр и угол наклона к горизонтали. Трубопроводы сетей сбора имеют угол наклона к горизонтали порядка 2–5 градусов, однако колени узлов запорно-регулирующей арматуры (УЗРА) могут иметь «типичные» углы 30, 45, 60, 90 градусов, к тому же на линейной части могут быть использованы различные варианты отводов как и с «типичными», так и «нетипичными» углами, поэтому в данной работе освещены «типичные углы» с целью повышения вариативности использования.

По данным сопровождения эксплуатации нефтегазоконденсатных месторождений Западной и Восточной Сибири в течение двух лет, взятых с режимных листов работы промысла, был сформирован массив данных, представленный в таблице 1.

**Таблица 1** Исходная выборка данных для расчета
**Table 1** Initial data sample for the calculation

| **Диаметр, м** | **Расход газа, м³/с (ст. усл.)** | | **Расход жидкости, м³/с * 10⁶ (ст. усл.)** | | **Входная температура, °C** | | **Выходное давление, Па абс. / 10⁵** | |
|---|---|---|---|---|---|---|---|---|
| | **Мин.** | **Макс.** | **Мин.** | **Макс.** | **Мин.** | **Макс.** | **Мин.** | **Макс.** |
| **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** |
| 0.141 | 2.51 | 3.46 | 2.31 | 7.41 | 9.00 | 11.45 | 22.20 | 24.27 |
| 0.143 | 7.83 | 8.21 | 0.69 | 50.12 | 2.01 | 7.16 | 43.31 | 52.69 |
| 0.147 | 2.22 | 13.64 | 0.23 | 1056.25 | -8.80 | 18.09 | 18.58 | 53.73 |

**Окончание табл. 1**
**Table 1 (end)**

| Диаметр, м | Расход газа, м³/с (ст. усл.) | | Расход жидкости, м³/с * 10⁶ (ст. усл.) | | Входная температура, ⁰C | | Выходное давление, Па абс. / 10⁵ | |
|---|---|---|---|---|---|---|---|---|
| | Мин. | Макс. | Мин. | Макс. | Мин. | Макс. | Мин. | Макс. |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 0.148 | 3.94 | 12.80 | 0.00 | 636.81 | 8.58 | 15.62 | 25.21 | 53.73 |
| 0.150 | 1.88 | 9.35 | 2.20 | 737.62 | -1.36 | 18.09 | 20.34 | 76.44 |
| 0.152 | 1.46 | 12.59 | 0.23 | 2903.36 | 5.54 | 30.15 | 18.71 | 82.63 |
| 0.195 | 1.77 | 16.94 | 1.74 | 809.84 | 0.00 | 18.17 | 39.71 | 81.25 |
| 0.199 | 1.92 | 20.57 | 0.23 | 765.74 | -1.2 | 16.70 | 38.93 | 84.58 |
| 0.201 | 1.44 | 19.88 | 0.35 | 607.29 | 1.55 | 18.09 | 18.71 | 61.33 |
| 0.245 | 5.03 | 36.54 | 2.55 | 916.44 | 4.72 | 16.27 | 39.71 | 77.87 |
| 0.249 | 4.83 | 13.24 | 1.27 | 620.72 | 0.86 | 3.40 | 40.57 | 56.21 |
| 0.251 | 2.06 | 28.42 | 0.58 | 1241.09 | -9.25 | 15.59 | 17.32 | 30.78 |
| 0.259 | 3.59 | 11.01 | 0.23 | 159.03 | 0.73 | 12.00 | 18.32 | 32.43 |
| 0.297 | 2.13 | 45.52 | 0.23 | 2904.86 | 0.58 | 26.03 | 17.29 | 76.25 |
| 0.309 | 5.32 | 12.06 | 1.74 | 2.89 | -8.74 | 7.95 | 23.72 | 30.78 |
| 0.394 | 9.38 | 48.84 | 1.74 | 1241.20 | 1.66 | 15.44 | 17.32 | 24.42 |
| 0.408 | 20.74 | 31.51 | 6.94 | 163.89 | 8.39 | 12.52 | 19.75 | 23.91 |

Области варьируемых величин лежат в следующих диапазонах:

— внутренний диаметр трубопровода от 0.141 до 0.408 м;
— объемный расход газа от 1.44 до 48.84 м³/с (ст. усл.);
— объемный расход жидкости от 0.001 до 2904.86 м³/с (ст. усл.);
— давление на выходе из трубопровода от 17.29 до 84.58 Па абс./10⁵;
— температура на входе в трубопровод от -9.25 до 30.15 ⁰C.

Входными параметрами для расчета модели в “OLGA” являются температура на входе и на выходе трубопровода, давление на входе в трубопровод, обводненность, газожидкостный фактор, взятые при стандартных условиях плотности газа, газоконденсата и пластовой воды. В расчетах будем рассматривать участок трубопровода длиной 1 м. Конкретизируем диапазон параметров: внутренний диаметр трубопровода от 0.1 до 0.5 м с шагом 0.1 м; давление на входе (выходе) в трубопроводе от 0 до 105 Па с шагом 104 Па; температура во входном (выходном) сечении трубопровода от -13.3 до 40.0 0C с шагом 6.7 0C; обводненность от 0 до 1 д.ед. с шагом в 0.166 д. ед.; газожидкостный фактор от 333 до 225 000 000 м3/м3 с шагом в 37499945 м3/м3, взятых при стандартных условиях.

Ввиду того, что температура потока в исходной выборке данных представлена с учетом теплообмена с окружающей средой, в процессе моделирования в симуляторе опция

теплообмена отключена, так же по причине единичной длины трубопровода полагается, что градиент температуры и давления мал и температура на выходе равна температуре на входе в трубопровод, также как и давление. Мерой, по которой происходит решение о накоплении жидкости, является доля перекрытия проходного сечения трубопровода жидкой фазой. Из практических соображений и опыта считается, что в трубопроводе не происходит накопление, если эта величина имеет значение менее 3%, т.к. это в меньшей степени влияет на величину градиента давления. Так же была принята нулевая шероховатость внутренней стенки трубопровода, т.к. при этом определяющая величина принимает минимальное значение, что помогает определить оптимальные параметры эксплуатации в условиях отсутствия выноса. Данные принимаются в момент, когда режим потока стабилизировался, и параметры во времени не меняются.

## Предварительный анализ и обработка данных

При моделировании получено 617 413 различных наборов параметров, из которых в 347 778 случаев накопление не происходит, а в 269 635 происходит. Выходной массив данных был поделен на две категории, в качестве целевого значения принято «происходит ли накопление жидкости?»: если накопление происходит, то набору данных присваивается значение 1, в противном случае 0. Для достижения большей «простоты» обучения было принято решение о сокращении числа определяющих параметров (факторов); перечень величин с их частотным распределением для составления выборки для машинного обучения приведен на рисунке 2.

Перечень включает в себя: плотность газа в условиях потока $(\rho_g)$, плотность жидкости в условиях потока $(\rho_l)$, объемный расход газа в условиях потока $(Q_g)$, объемный расход жидкости в условиях потока $(Q_l)$, динамическую вязкость газа в условиях потока $(\mu_g)$, динамическую вязкость жидкости в условиях потока $(\mu_l)$, обводненность в условиях потока $(W)$, внутренний диаметр трубопровода $(D)$, угол наклона трубопровода к горизонтали $(a)$. При детальном рассмотрении гистограмм на рисунке 2 можно определить минимальные, максимальные, средние значения рассматриваемых величин, а также какой процент выборки соответствует определенном среднему результату (табл. 2). Аномалий на графиках при этом обнаружено не было.

Из рисунка 3 видно, что аномалии в данных все же присутствуют, визуальное отображение показывает сильное отклонение, которое выходит за квартиль и межквартильный размах для плотности жидкости в условиях потока. Самый большой разброс данных при этом наблюдается у плотности газа и угле наклона трубопровода к горизонтали.

Численные значения количества аномалий (выбросов) для каждого из определяющих параметров рассчитаны и представлены на рисунке 4. Из данных, представленных на этом рисунке, видно, что больше всего выбросов данных в таком случае наблюдается у обводненности, далее у плотности жидкости, объемного расхода жидкости, динамической вязкости жидкости и у расхода газа. Примечательно, что у плотности газа, динамической вязкости газа, угла наклона трубопровода к горизонтали и внутреннего диаметра трубопровода выбросов обнаружено не было.

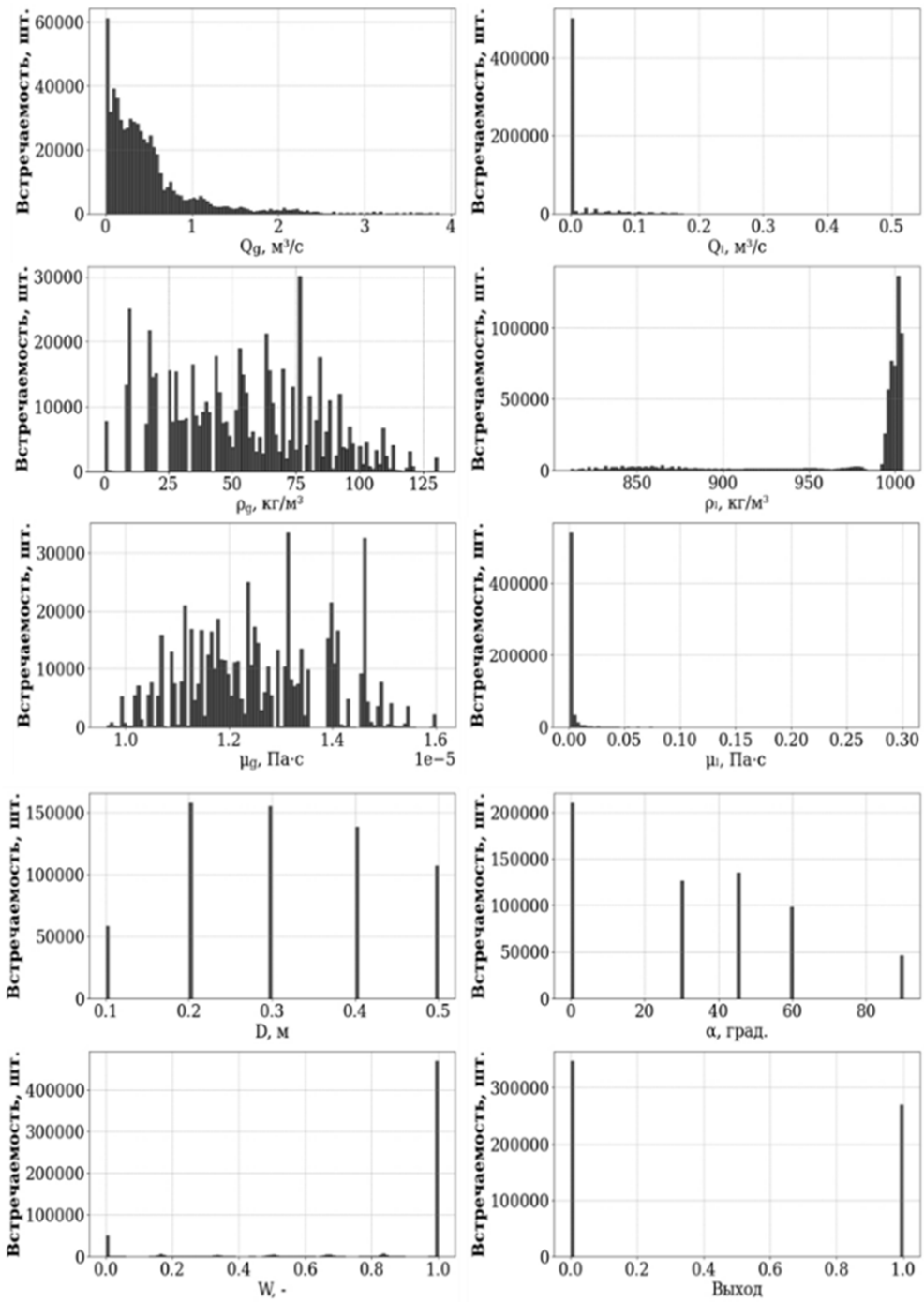


**Рис. 2.** Частота встречаемости определяющих параметров выборки

**Fig. 2.** Frequency occurrence of determining parameters (factors) of the data sample

**Таблица 2** Описательная статистика выборки параметров (факторов)
**Table 2** Descriptive statistics of the data sample parameters (factors)

| Параметр | Среднее | Мин. | 25% | 50% | 75% | Макс. |
|---|---|---|---|---|---|---|
| $\rho_g$, кг/м³ | 54.03 | 0.03 | 30.42 | 53.60 | 76.23 | 130.78 |
| $\rho_l$, кг/м³ | 975.19 | 811.39 | 993.25 | 998.81 | 1001.98 | 1004.73 |
| $\mu_g$, Па · с | $13 \cdot 10^{-6}$ | $10 \cdot 10^{-6}$ | $12 \cdot 10^{-6}$ | $12 \cdot 10^{-6}$ | $13 \cdot 10^{-6}$ | $16 \cdot 10^{-6}$ |
| $\mu$, Па · с | 0.00376 | 0.01451 | 0.00086 | 0.00143 | 0.00179 | 0.29982 |
| $W$, д. ед. | 0.84 | 0.00 | 0.99 | 1.00 | 1.00 | 1.00 |
| $Q_g$, м³/с | 0.49 | $4.69 \cdot 10^{-10}$ | 0.14 | 0.35 | 0.59 | 3.86 |
| $Q_l$, м³/с | 0.01 | $1.21 \cdot 10^{-10}$ | $1.6 \cdot 10^{-4}$ | $4.6 \cdot 10^{-4}$ | 0.3 · 10-4 | 0.52 |
| $\alpha$, º | 32.37 | 0.00 | 0.00 | 30.00 | 45.00 | 90.00 |
| $D$, м | 0.31 | 0.10 | 0.2 | 0.3 | 0.4 | 0.5 |

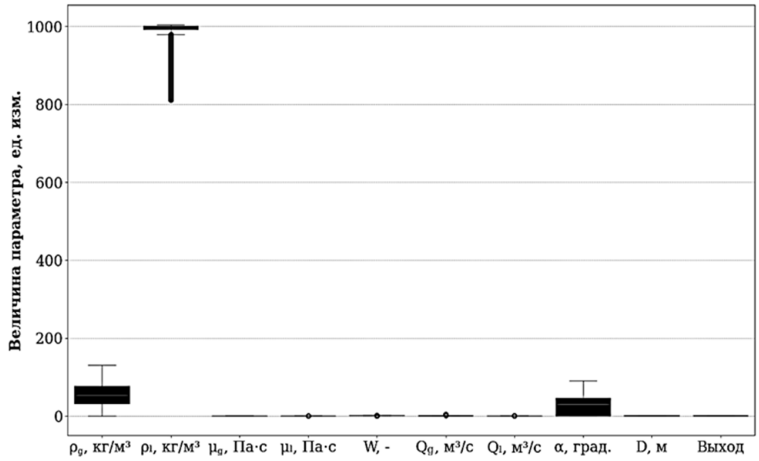


**Рис. 3.** Коробчатая диаграмма определяющих параметров (факторов) выборки
**Fig. 3.** Box plot of determining parameters (factors) of the data sample

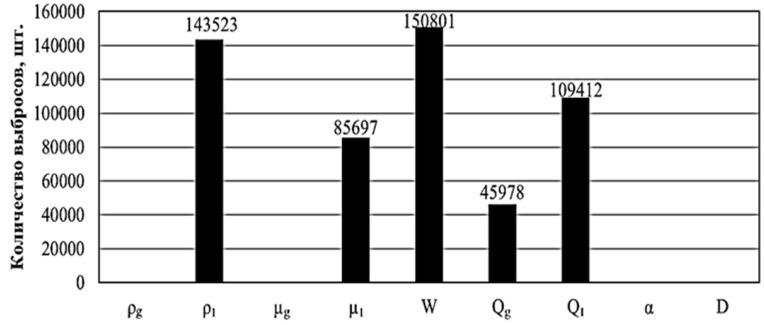


**Рис. 4.** Количество выявленных аномалий в определяющих параметрах (факторах) выборки
**Fig. 4.** Number of detected anomalies in determining parameters (factors) of the data sample

По корреляционной матрице (рис. 5) можно отметить связь между определяющими параметрами, так, например, плотность газа и его динамическая вязкость линейно коррелируют между собой с величиной 92%, плотность газа и его объемный расход с величиной -50%, динамическая вязкость газа с его объемным расходом с величиной -45%, плотность жидкости с её объемным расходом с величиной -43%, обводненность потока и объемным расходом жидкости с величиной -43% и т.д.

Эти связи подтверждаются при анализе рисунка 6. По диагональной линии отражены частоты встречаемости определяющих параметров (факторов), аналогичные были ранее представлены на рисунке 2, оставшиеся графики представляют собой точечные диаграммы рассеяния.

В соответствие с анализом ранее приведенных графиков можно заключить, что с данными расчета симулятора (пусть они и являются априори «чистыми», однако шум на практике неизбежен) необходимо провести процессы стандартизации, масштабирования и нормализации, с целью снижения влияния выбросов и исключения размерности параметров.

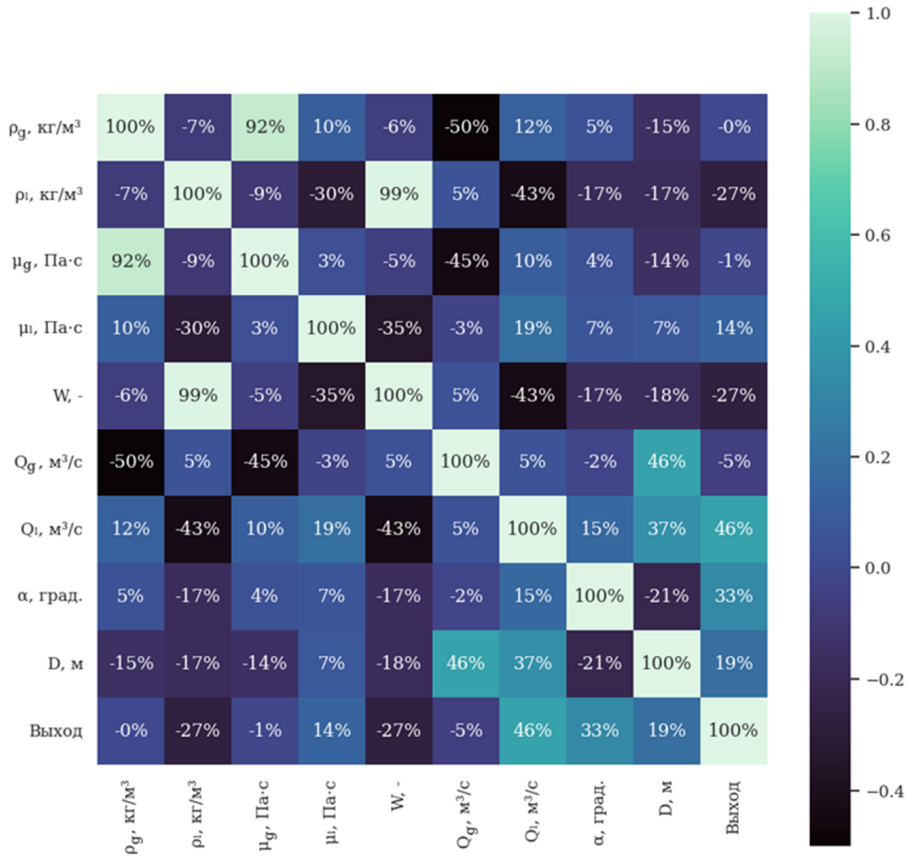


**Рис. 5.** Корреляционная матрица в определяющих параметрах выборки

**Fig. 5.** Correlation matrix in determining parameters (factors) of the data sample

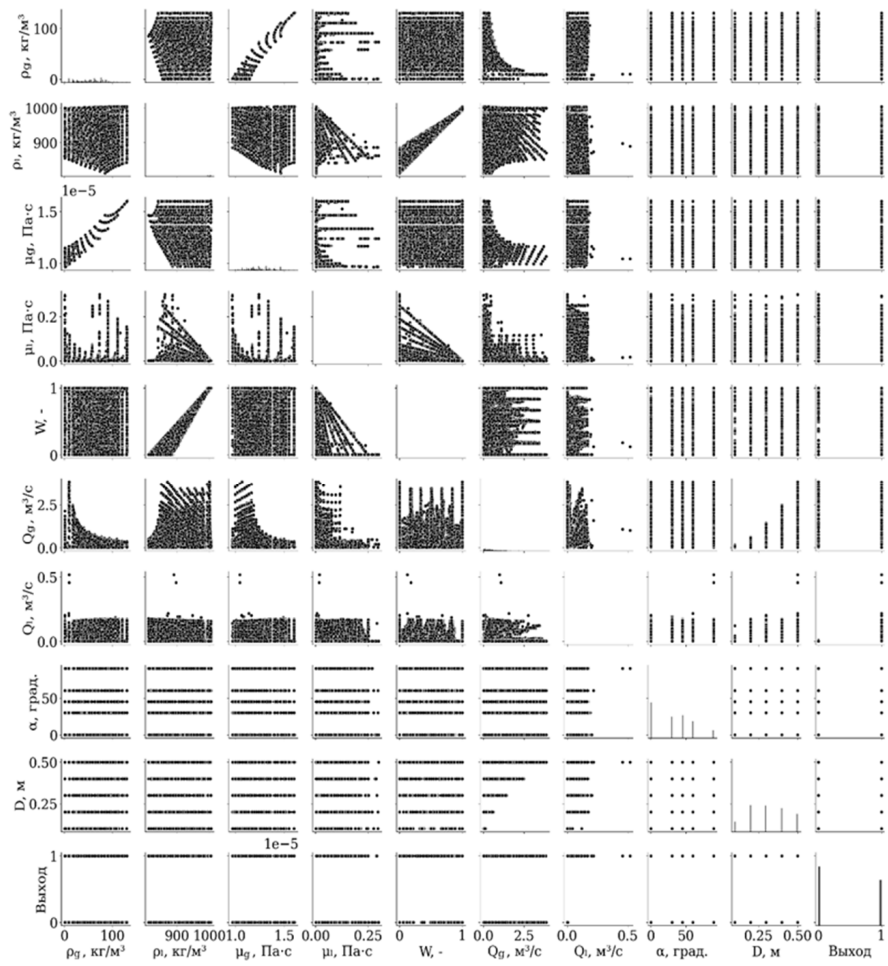


**Рис. 6.** Диаграмма рассеивания для определяющих параметров (факторов) выборки

**Fig. 6.** Scatter plot for determining parameters (factors) of the data sample

## Выбор алгоритма машинного обучения

На рисунке 7 отображены основные подходы машинного обучения. Задачей данной работы является бинарная классификация, для успешной реализации которой вполне подходят принципы классического обучения с учителем, так же возможно использование ансамблевых методов обучения и методов глубокого обучения (нейросетей).

В настоящее время широко используются следующие алгоритмы машинного обучения:

— логистическая регрессия (LR) [Пылов и др., 2024];

— линейный дискриминантный анализ (LDA) [Жангиров и др., 2018];
— K-ближайших соседей (KNN) [Родионов, Ищенко, 2024];
— дерево принятия решений (CART) [Усачев, 2018];
— наивный байесовский классификатор (NB) [Сабуров, 2024];
— линейных опорных векторов (LSVC) и опорных векторов (SVC) [Гефан, Иванов, 2012];
— «бэггинг» (BG) [Коргун, 2019];
— «случайный лес» (RF) [Ахикян, Данилюк, 2024] и классификатор экстремально рандомизированных деревьев (ET);
— адаптивный «бустинг» (AB), градиентный «бустинг» (GB) и экстремальный градиентный «бустинг» (XGB) [Ильичев и др., 2021];
— многослойный перцептрон (MLP) [Митрофанова, Комлев, 2019].


**Рис. 7.** Классификация основных алгоритмов машинного обучения
**Fig. 7.** Classification of basic machine learning algorithms

Для оценки качества моделей и последующим выбором оптимального алгоритма был использован метод кросс-валидации (или K-кратной перекрестной проверки). Принцип его работы заключается в разделении исходной выборки на величину K кратных частей и выбором одной в качестве тестовой, а остальных в качестве обучающих, далее после проведения нескольких итераций определяется среднее арифметическое между критериями качества, наглядно принцип продемонстрирован на рисунке 8.

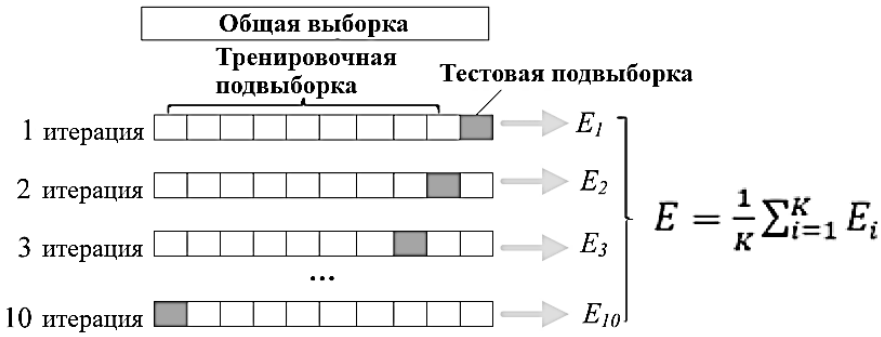


**Рис. 8.** Принцип работы метода кросс-валидации
**Fig. 8.** Principle of the cross-validation method

В последующем каждый из алгоритмов был протестирован, меткой качества определена точность, которая может быть вычислена по формуле 1.

$$A = \frac{TP+TN}{TP+TN+FP+FN}, \quad (1)$$

где $TP$ — истинно положительный результат, $TN$ — истинно отрицательный результат, $FP$ — ложно положительный результат, $FN$ — ложно отрицательный результат.

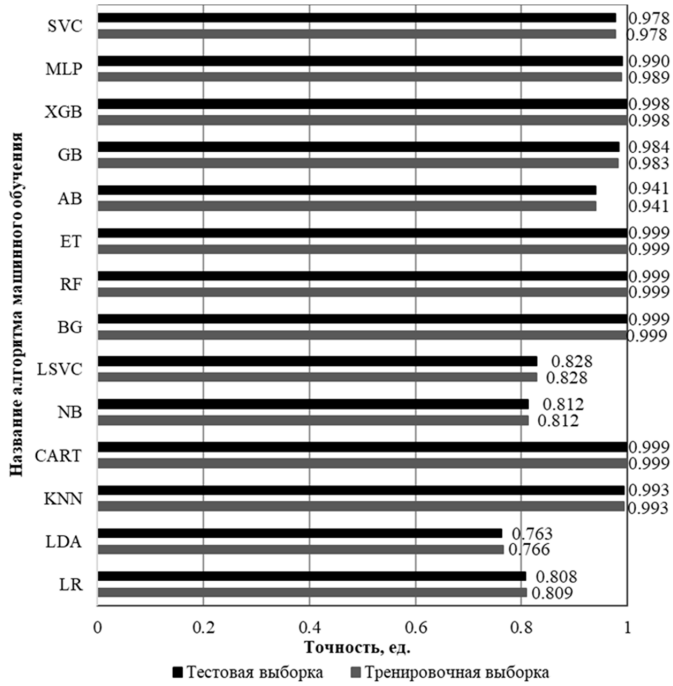


**Рис. 9.** Оценка точности различных алгоритмов машинного обучения
**Fig. 9.** Accuracy assessment of different machine learning algorithms

Из рисунка 9 видно, что наивысшей точностью из всех представленных обладают ансамблевые алгоритмы, однако и из простых алгоритмов лушие результаты показали «дерево принятия решений» (CART) и «K-ближайших соседей» (KNN), при условии того, что оптимизация гиперпараметров не была произведена и им было присвоено значение по умолчанию.

## Алгоритм «дерево принятия решений»

Как было отмечено ранее суть алгоритма «дерево принятия решений» заключается в построении моделью последовательности структуры, где каждый узел представляет собой условие, а каждое ребро — возможный результат условия, выходные узлы, содержащие ответы представляют собой «листья». Внешне собой эта структура напоминает дерево, от чего алгоритм и получил свое название (рис. 10). Корневой узел, принципиально отделяющий 2 класса между собой расходиться на узлы с условиями, называемые «потомками», для всех последующий узлов они являются «придатками».

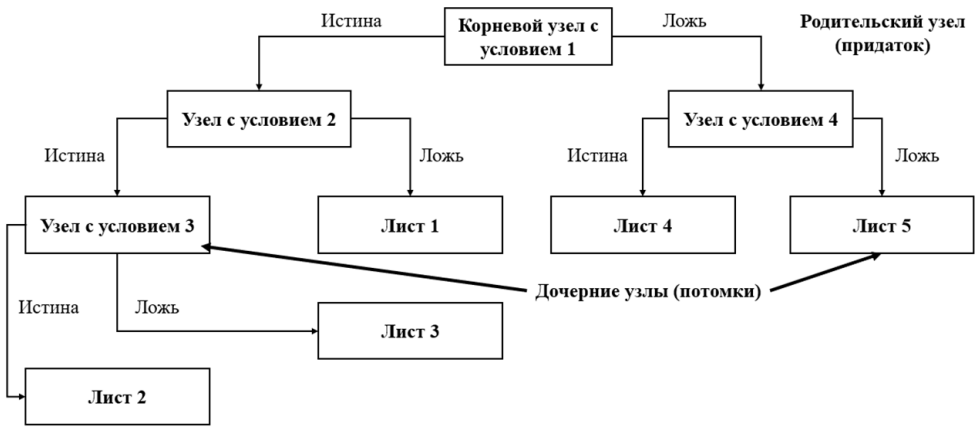


**Рис. 10.** Принципиальная схема работы алгоритма «дерево принятия решений»
**Fig. 10.** Principal scheme of the „decision tree" algorithm

Для оптимизации гиперпараметров «дерева принятия решений» используются те же принципы кросс-валидации, что и для сравнения алгоритмов между собой. На рисунке 11 по оси абсцисс показаны величины определяющих гиперпараметров в единицах, по оси ординат — мера точности в единицах. Как можно заметить с увеличением минимального числа образцов в узле точность падает, при увеличении максимального количества листовых узлов точность возрастает, при увеличении максимального количества определяющих параметров точность растет, при увеличении максимального глубины дерева точность возрастает, при увеличения максимального количества образцов точность так же возрастает, и наконец, при увеличении минимальной весовой доли узла точность падает. Общий тренд поведения графиков, что для обучающей выборки, что для кросс-валидации сохраняется.

Для данного алгоритма существуют и другие гиперпараметры вроде «критерия построения», однако в ходе их изменения существенных изменений точности обнаружено не было и в статье они не освещены.

В конечном итоге величины гиперпараметров для данного алгоритма могут быть определены как:

— минимальное количество образцов в узле — 1 ед.;
— максимальное количество листовых узлов — 2 ед.;
— максимальное количество параметров — 8 шт.;
— максимальная глубина дерева — 15 ед.;
— максимальное разделение образцов — 2 ед.;
— максимальная весовая доля узла — 0 ед.

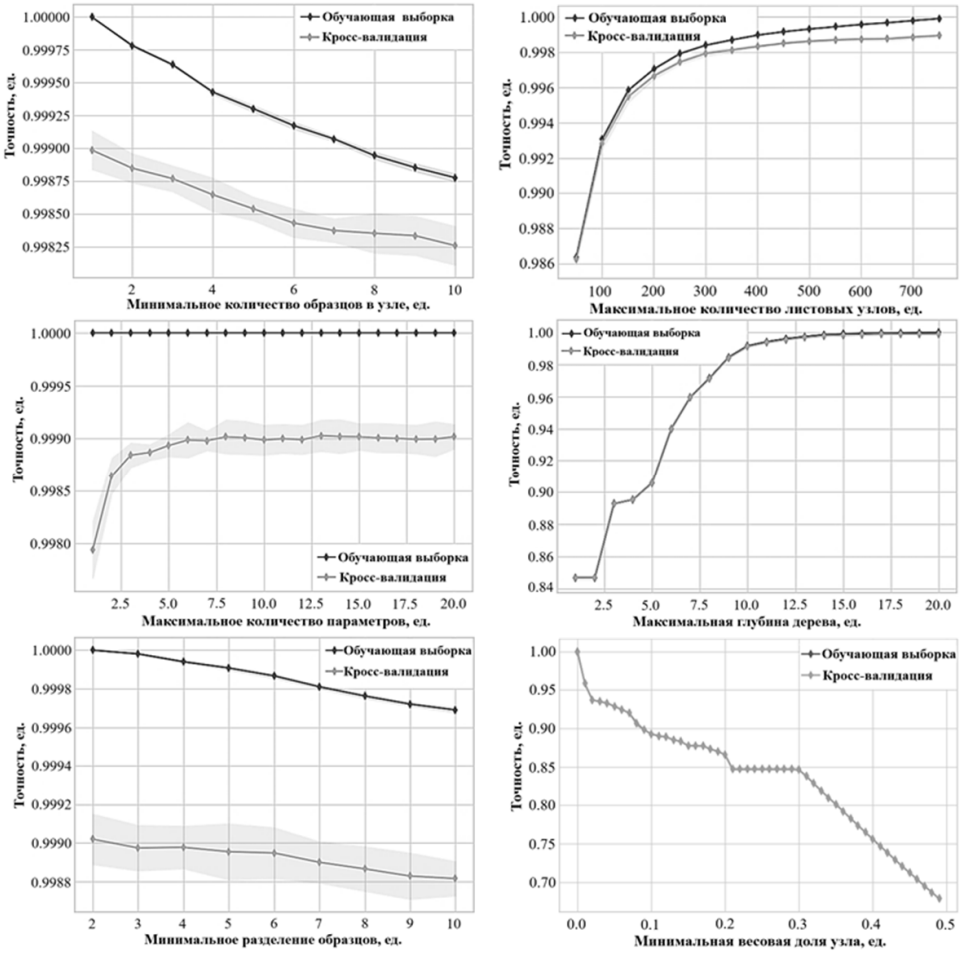


**Рис. 11.** Оптимизация гиперпараметров алгоритма «дерево принятия решений»
**Fig. 11.** Hyperparameters optimization of the “decision tree” algorithm

Далее с учетом оптимизированных гиперпараметров модель была обучена на 80% исходной выборки, в то время как оставшиеся 20% были использованы для тестирования. Как видно из рисунка 12 число истинных отрицательных результатов составляет 278296 ед., число ложных положительных результатов — 10 ед., число ложных отрицательных результатов — 5 ед. и число истинных положительных результатов — 215619 ед. Точность, чувствительность, специфичность, выпадение, арифметическое среднее, гармоническое среднее, геометрическое среднее, а так же F-мера оцениваются приблизительно в 100%.

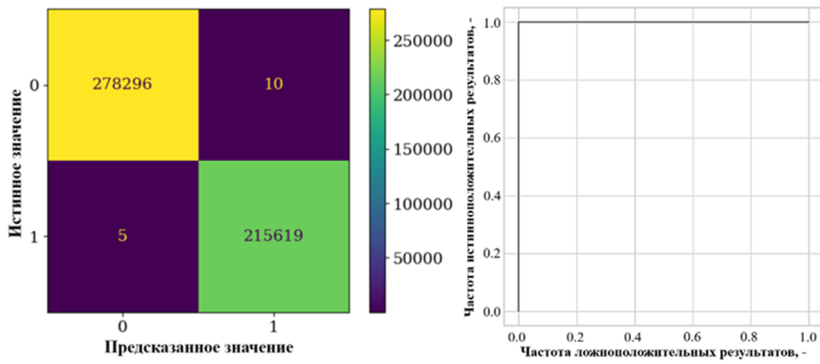


**Рис. 12.** Матрица неточностей (слева) и кривая ROC (справа) для алгоритма «дерево принятия решений» для обучающей выборки

**Fig. 12.** Inaccuracy matrix (left) and ROC curve (right) for the “decision tree” algorithm (training data sample)

Из рисунка 13 видно, что число истинных отрицательных результатов составляет 69407 ед., число ложных положительных результатов — 65 ед., число ложных отрицательных результатов — 55 ед. и число истинных положительных результатов — 53956 ед. Точность, чувствительность, специфичность, выпадение, арифметическое среднее, гармоническое среднее, геометрическое среднее, а так же F-мера приблизительно равны 99%.

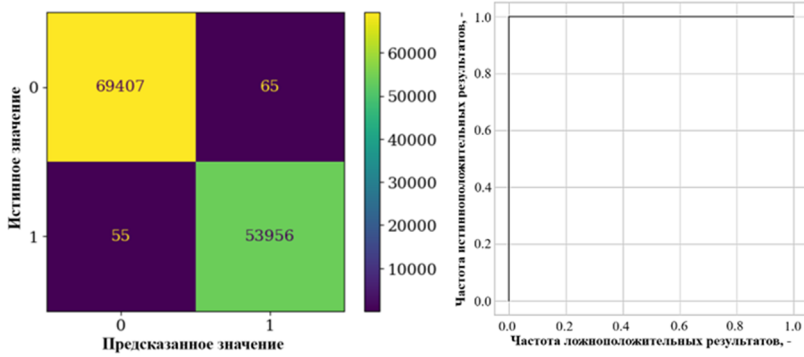


**Рис. 13.** Матрица неточностей (слева) и кривая ROC (справа) для алгоритма «дерево принятия решений» для тестовой выборки

**Fig. 13.** Inaccuracy matrix (left) and ROC curve (right) for the “decision tree” algorithm (test data sample)

Для разбиения узлов по определяющему условию в алгоритме «дерево принятия решений» модели необходимо выбрать параметр в качестве атрибута для разделения. Входной массив данных содержит в себе определенное количество признаков и каждый из них выбирается для разбиения в соответствии с величиной коэффициента Джини (рис. 14).

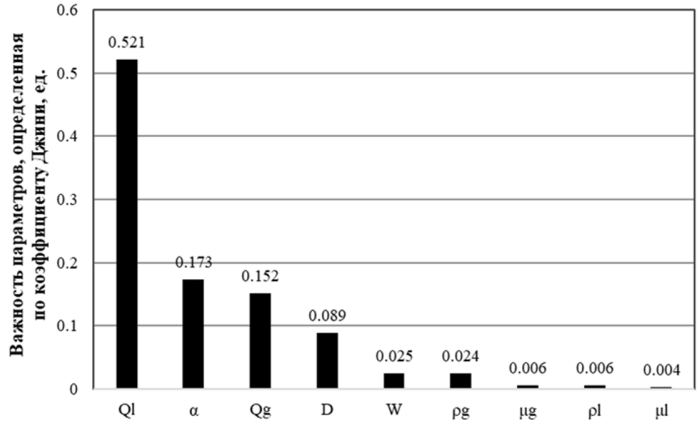


**Рис. 14.** Важность параметров, основанная на коэффициенте Джини
**Fig. 14.** Parameters importance based on "Gini coefficient"

Согласно проведенному разделению первостепенным параметром, влияющим на процесс накопления, является объемный расход жидкости с величиной 0.521, затем по значимости стоит угол наклона трубопровода к горизонтали — 0.173, объемный расход газа — 0.152, внутренний диаметр трубопровода — 0.089, обводненность потока — 0.025, плотность газа в условиях потока — 0.024, динамическая вязкость газа в условиях потока — 0.006, плотность жидкости в условиях потока — 0.006, динамическая вязкость жидкости в условиях потока — 0.004.

## Алгоритм «K-ближайших соседей»

Как было сказано ранее, в алгоритме «K-ближайших соседей», модель стремится определить объект к классу в зависимости от минимального расстояния в гиперпространстве. Последовательность действий алгоритма может быть представлена следующим образом:

- вычисляется расстояние между тестовыми и всеми обучающими наборами данных;
- из тестового набора выбирается количество K-ближайших, где число соседей (K) задается заранее;
- итоговым прогнозом среди выбранных наборов будет мода;
- для всех тестовых наборов вышеупомянутые шаги повторяются.

Как и у алгоритма «дерево принятия решений», для оптимизации гиперпараметров «K-ближайших соседей» используются те же принципы кросс-валидации, что и для сравнения алгоритмов между собой. На рисунке 15 по оси абсцисс показаны величины определяющих гиперпараметров в единицах, по оси ординат — мера точности в единицах.

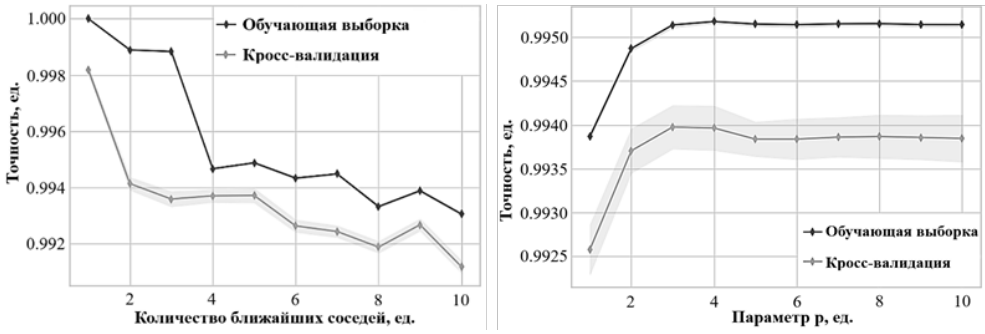


**Рис. 15.** Оптимизация гиперпараметров алгоритма «K-ближайших соседей»

**Fig. 15.** Hyperparameters optimization of the “K-nearest neighbors” algorithm

Можно заметить, что с ростом количества ближайших соседей точность модели падает, при увеличении параметра *p*, отвечающий за мощность по метрике Минковского, точность увеличивается. Общий тренд поведения графиков для обучающей выборки и для кросс-валидации сохраняется. Для данного алгоритма существуют и другие гиперпараметры вроде выбора подалгоритма («ball tree», «kd tree», «brute»), однако в ходе их варьирования существенных изменений точности обнаружено не было.

Далее с учетом оптимизированных гиперпараметров модель была обучена на 80% исходной выборки; оставшиеся 20% были использованы для тестирования (рис. 16). Как видно из рисунка число истинных отрицательных результатов составляет 278 306 ед., число ложных положительных результатов — 0 ед., число ложных отрицательных результатов — 0 ед. и число истинных положительных результатов — 215 624 ед. Точность, чувствительность, специфичность, выпадение, арифметическое среднее, гармоническое среднее, геометрическое среднее, а также F-мера оцениваются в 100%.

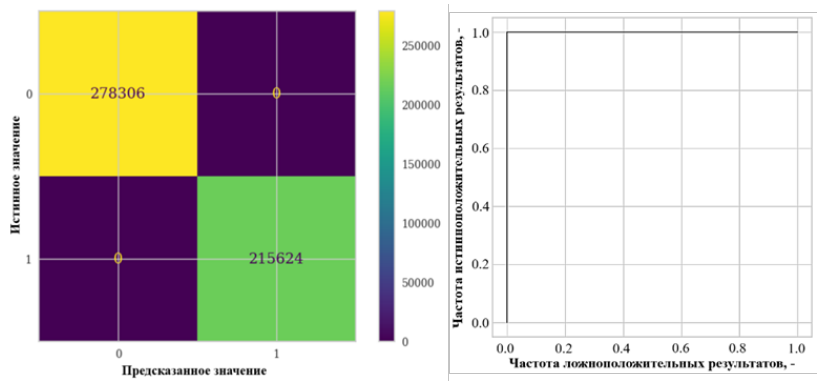


**Рис. 16.** Матрица неточностей (слева) и кривая ROC (справа) для алгоритма «K-ближайших соседей» для обучающей выборки

**Fig. 16.** Inaccuracy matrix (left) and ROC curve (right) for the “K-nearest neighbors” algorithm (training data sample)

Из рисунка 17 видно, что число истинных отрицательных результатов составляет 69 370 ед., число ложных положительных результатов — 102 ед., число ложных отрицательных результатов — 103 ед. и число истинных положительных результатов — 53 818 ед. Точность, чувствительность, специфичность, выпадение, арифметическое среднее, гармоническое среднее, геометрическое среднее, а так же F-мера приблизительно равны 99%.

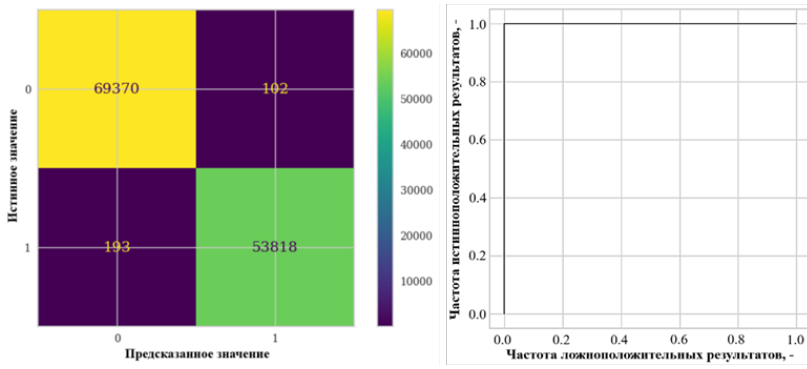


**Рис. 17.** Матрица неточностей (слева) и кривая ROC (справа) для алгоритма «K-ближайших соседей» для тестовой выборки

**Fig. 17.** Inaccuracy matrix (left) and ROC curve (right) for the "K-nearest neighbors" algorithm (test data sample)

После определения гиперпараметров моделей точность алгоритмов имеет максимально возможный уровень, более подробно меры оценки качества представлены в таблице 3.

**Таблица 3** Меры оценки качества полученных моделей машинного обучения

**Table 3** Quality measures of the obtained machine learning models

| Параметр | Дерево решений (тренировочная выборка) | Дерево решений (тестовая выборка) | Метод K-ближайших соседей (тренировочная выборка) | Метод K-ближайших соседей (тестовая выборка) |
|---|---|---|---|---|
| Истинный отрицательный | 278296 | 69407 | 278306 | 69370 |
| Ложный положительный | 10 | 65 | 0 | 102 |
| Ложный отрицательный | 5 | 55 | 0 | 193 |
| Истинный положительный | 215619 | 53956 | 215624 | 53818 |
| Точность | 1.0000 | 0.9990 | 1.0000 | 0.9976 |
| Чувствительность | 1.0000 | 0.9990 | 1.0000 | 0.9964 |
| Специфичность | 1.0000 | 0.9991 | 1.0000 | 0.9985 |
| Выпадение | 0.0000 | 0.0009 | 0.0000 | 0.0015 |
| Арифметическое среднее | 1.0000 | 0.9989 | 1.0000 | 0.9973 |
| Гармоническое среднее | 1.0000 | 0.9989 | 1.0000 | 0.9973 |
| Геометрическое среднее | 1.0000 | 0.9989 | 1.0000 | 0.9973 |
| F-мера | 1.0000 | 0.9989 | 1.0000 | 0.9973 |

## Заключение

По результатам моделирования в симуляторе было получено 617413 различных наборов параметров, из которых в 347778 случаев накопление не происходит, а в 269635 происходит, при этом, несмотря на то, что данные расчета симулятора являются «чистыми», в данных все же присутствуют аномальные выбросы, для устранения влияния шумов применены методы стандартизации, масштабирования и нормализации.

В процессе проведения машинного обучения по предобработанным данным методом кросс-валидации было выбрано два оптимальных алгоритма обучения: «дерево принятия решений», «K-ближайших соседей». Также установлено, что величиной, наиболее существенным образом влияющей на процесс накопления, является объемный расход жидкости, затем угол наклона трубопровода к горизонтали и объемный расход газа.

Для оценки накопления жидкости на практике в рамках данной методики предлагается рассчитать с использованием получившейся модели машинного обучения точки начала и окончания трубопровода, при необходимости с дальнейшим приближением к месту накопления жидкости при условиях потока и учетом его геометрических особенностей в данной точке.

Ввиду отсутствия некоторых фактически замеренных параметров работы трубопроводов, данные в этой статье взяты с результатов расчета симулятора и поэтому их достоверность может в некоторых случаях не отражать действительность. Для более корректного прогнозирования в дальнейшем рекомендуется проведение экспериментов на реальном оборудовании.

В настоящее время не один существующий симулятор многофазного потока не способен учитывать такие факторы как наличие механических примесей в потоке, объем скопления газовых гидратов в конкретном месте трубопровода и т.д., которые, в свою очередь, влияют на межфазные гидравлические сопротивления и сопротивления о стенку трубы и, в конечном итоге, на процесс накопления жидкости в газопроводах.

Разработанная модель имеет преимущество перед численным моделированием в различных симуляторах ввиду того, что повышает эффективность расчетов и снижает время и трудозатраты на моделирование. Данная модель может стать полезным средством для анализа и локализации процесса накопления жидкости, обеспечивая более упрощенное и всестороннее прогнозирование по сравнению с другими моделями, носящими чаще всего полуэмпирический характер.

## Список источников

Алиев З. С., Андреев С. А., Власенко А. П., Коротаев Ю. П. 1978. Технологический режим работы газовых скважин. М.: Недра. 279 с.

Ахикян А. И., Данилюк С. С. 2024. Адаптивный случайный лес и его применение // Вестник науки. Том 1. № 6 (75). С. 1393–1402.

Борисевич Ю. П., Голованова Ю. В., Краснова Г. З. 2022. Определение конденсато-опасных мест на газопроводах // Естественные и технические науки. № 10 (173). С. 197–200.

Бузников Н. А., Истомин В. А., Митницкий Р. А. 2016. Влияние накопленной в промысловом трубопроводе жидкости на движение ингибитора гидратообразования // Вести газовой науки. № 2 (26). С. 112–116.

Ватузов С. М., Ванчугов И. М., Резанов К. С., Автомонов П. Ю., Танасенко М. С., Шестаков Р. А. 2023. К вопросу о коррозии промысловых трубопроводов // Проблемы сбора, подготовки и транспорта нефти и нефтепродуктов. №5 (145). С. 106–122.

Гефан Г. Д., Иванов В. Б. 2012. Метод опорных векторов и альтернативный ему простой линейный классификатор // Информационные технологии и проблемы математического моделирования сложных систем. № 10. С. 84–94.

Жангиров Т. Р., Перков А. С., Иванова С. А., Лисс А. А., Григорьева Н. Ю., Чистякова Л. В. 2019. Сравнение эффективности решения задачи классификации методами линейного дискриминантного анализа и искусственных нейронных сетей // Известия СПбГЭТУ ЛЭТИ. № 5. С. 64–73.

Ильичев В. Ю., Жукова Ю. М., Шамов И. В. 2021. Использование технологии градиентного бустинга для создания аппроксимационных моделей // Заметки ученого. № 12-1. С. 62–67.

Клапчук О. В., Елин Н. Н. 1979. Истинные концентрации жидкости и газа в газопроводах системы промысел – ГПЗ // Газовая промышленность. Вып. 3. С. 18–28.

Коргун Д. А. 2019. Ансамблевые методы классификации в машинном обучении // Математические методы управления: сб. науч. тр. Тверь: Тверской гос. ун-т. С. 32–39.

Краснов А. Н. 2018. Особенности эксплуатации газосборной системы в условиях накопления жидкости // Нефтегазовое дело. Том 16. № 4. С. 118–126.

Кутателадзе С.С., Накоряков В. Е. 1984. Тепломассообмен и волны в газожидкостных системах. Новосибирск: Наука. 301 с.

Лурье М. В. 2011. Пробковое течение газожидкостной смеси в горизонтальном трубопроводе // Нефтяное хозяйство. № 7. С. 122–124.

Митрофанова А. С., Комлев Г. В. 2019. Обучение перцептрона // Тенденции развития науки и образования. № 49-12. С. 69-71.

Нуруллаев З. В., Очилов А. У. У. 2017. Исследование пульсации давления в промысловых трубопроводах // Научный аспект. № 4-1. С. 146–148.

Одишария Г. Э., Точигин А. А. 1998. Прикладная гидродинамика газожидкостных смесей. М.: Газпром ВНИИГАЗ. 400 с.

Пылов П. А., Майтак Р. В., Дягилева А. В., Салычева А. Д. 2024. Методы восстановления непараметрической регрессии в условиях несбалансированных данных. Вологда: Инфра-Инженерия. 192 с.

Родионов А. В., Ищенко К. Л. 2024. Исследование влияния параметров алгоритма k-ближайших соседей на метрики качества моделей // System Analysis and Mathematical Modeling. Том 6. № 2. С. 251–262.

Сабуров В. С. 2024. Байесовский классификатор в машинном обучении // Шаг в науку. № 1. С. 78–81.

Усачев П. 2018. Решение задачи классификации на основе алгоритмов дерева принятия решений // Современные технологии в науке и образовании — СТНО-2018: сб. тр. Междунар. науч.-техн. форума (28 февраля – 02 марта 2018 г., Рязань, Россия). Рязань: Рязанский гос. радиотехнический ун-т им. В.Ф. Уткина. Том 3. С. 73–76.

Цховребов А. С. 2024. Применение машинного обучения в нефтегазовой отрасли // Актуальные вопросы исследования нефтегазовых пластовых систем: сб. тез. докладов V Междунар. науч.-практ. конф. (03–04 октября 2024 г., Москва, Россия). М.: Газпром ВНИИГАЗ. С. 40.

Rastogi A., Fan Y. 2020. Experimental and modeling study of onset of liquid accumulation // Journal of Natural Gas Science and Engineering. V. 73. 103064.

## References

Aliev, Z.S., Andreev, S.A., Vlasenko, A.P., Korotayev, Yu.P. (1978). *Technological mode of operation of gas wells.* Nedra. Moscow. [In Russian]

Akhikyan, A.I., Danilyuk, S.S. (2024). Adaptive random forest and its application. *Bulletin of Science*, 1 (6). 1393–1402. [In Russian]

Borisevich, Yu.P., Golovanova, Yu.V., Krasnova, G.Z. (2022). Determination of condensate-hazardous places on gas pipelines, *Natural and Technical Sciences*, (10), 197–200. [In Russian]

Buznikov, N.A., Istomin, V.A., Mitnitsky, R.A. (2016). Influence of the accumulated liquid in the field pipeline on the movement of hydrate formation inhibitor. *Vesti gazovoy nauki*, (26), 112–116. [In Russian]

Vatuzov, S.M., Vanchugov, I.M., Rezanov, K.S., Avtomonov, P.Yu., Tanasenko, M.S., Shestakov, R.A. (2023).To the question of corrosion of field pipelines. *Problems of gathering, preparation and transportation of oil and petroleum products*, (5), 106–122. [In Russian]

Gefan, G.D., Ivanov, V.B. (2012). Method of support vectors and an alternative simple linear classifier. *Information technologies and problems of mathematical modeling of complex systems*, (10), 84–94. [In Russian]

Zhangirov, T.R., Perkov, A.S., Ivanova, S.A., Liss, A.A., Grigorieva, N.Y., Chistyakova, L.V. (2019). Comparison of the efficiency of solving the classification problem by methods of linear discriminant analysis and artificial neural networks. *Izvestiya SPbGETU LETI*, (5). 64–73. [In Russian]

Ilyichev, V.Yu., Zhukova, Yu.M., Shamov, I.V. (2021).Using the technology of gradient bousting to create approximation models. *Notes of the scientist*, (12-1), 62–67. [In Russian]

Klapchuk, O.V., Yelin, N.N. (1979).True concentrations of liquid and gas in gas pipelines of the field - gas processing plant system. *Gas Industry*, (3), 18–28. [In Russian]

Korgun, D.A. (2019). Ensemble methods of classification in machine learning. In *Mathematical methods of control*: *Collection of scientific papers* (pp. 32–39). Tver State University [In Russian]

Krasnov, A.N. (2018). Features of gas-collection system operation under liquid accumulation. *Petroleum Engineering*, 16 (4), 118–126. [In Russian]

Kutateladze, S.S., Nakoryakov, V.E. (1984*). Heat and mass transfer and waves in gas-liquid systems*. Nauka. Novosibirsk. [In Russian]

Lurie, M.V. (2011). Gas-liquid slug flow in horizontal pipeline. *Oil Industry*, (7), 122–124. [In Russian]

Mitrofanova, A.S., Komlev, G.V. (2019). Learning perceptron. *Trends in the development of science and education*. (49-12), 69–71. [In Russian]

Nurullaev, Z.V., Ochilov, A.U.U. (2017). Study of pressure pulsation in field pipelines. *Nauchny Aspect*, (4-1), 146–148. [In Russian]

Odisharia, G.E., Tochigin, A.A. (1998). *Applied hydrodynamics of gas-liquid mixtures.* Gazprom VNIIGAZ. [In Russian]

Pylov, P.A., Maytak, R.V., Dyagileva, A.V., Salycheva, A.D. (2024). *Methods for restoring nonparametric regression under conditions of unbalanced data*. Infra-Engineering. Vologda. [In Russian]

Rodionov, A.V., Ischenko, K.L. (2024). Investigation of the influence of the k-nearest neighbors algorithm parameters on the model quality metrics. *System Analysis and Mathematical Modeling*, 6 (2), 251–262. [In Russian]

Saburov, V.S. (2024). Bayesian classifier in machine learning. *Step in Science*, (1). 78–81. [In Russian]

Usachev, P. (2018).Solving the classification problem on the basis of decision tree algorithms. In *Modern technologies in science and education - STNO-2018: Proceedings of the International Scientific and Technical Forum* (vol.3, pp. 73–76). Ryazan State Radio Engineering University. [In Russian]

Tskhovrebov, A.S. (2024). Application of machine learning in oil and gas industry. In *Actual issues of oil and gas reservoir systems research: Abstracts of the V International Scientific and Practical Conference* (p.40). Gazprom VNIIGAZ LLC. [In Russian]

Rastogi, A., Fan, Y. (2020). Experimental and modeling study of onset of liquid accumulation. *Journal of natural gas science and engineering*, 73, 103064.

## Информация об авторах

*Павел Александрович Крылов,* магистрант, базовая кафедра ООО "ТННЦ", Тюменский государственный университет, Тюмень, Россия
paul.kryloff@yandex.ru

*Наиль Габсалямович Мусакаев,* доктор физико-математических наук, профессор, профессор кафедры прикладной и технической физики, Тюменский государственный университет; главный научный сотрудник, Тюменский филиал Института теоретической и прикладной механики им. С.А. Христиановича СО РАН, Тюмень, Россия
musakaev68@yandex.ru, https://orcid.org/0000-0002-8589-9793

## Information about the authors

*Pavel A. Krylov,* Master's student, basic department of LLC "TNNC", University of Tyumen, Tyumen, Russia
paul.kryloff@yandex.ru

*Nail G. Musakaev,* Dr. Sci. (Phys.-Math.), Professor, Professor of the Department of Applied and Technical Physics, University of Tyumen; Chief Researcher, Tyumen Branch of the Khristianovich Institute of Theoretical and Applied Mechanics of the Siberian Branch of the Russian Academy of Sciences, Tyumen, Russia
musakaev68@yandex.ru, https://orcid.org/0000-0002-8589-9793