

**ИНТЕГРАЦИЯ МЕТОДОВ ЛИНЕЙНОЙ РЕГРЕССИИ И СЛУЧАЙНОГО ЛЕСА
В СИСТЕМЫ ИНТЕЛЛЕКТУАЛЬНОЙ ПОДДЕРЖКИ УПРАВЛЕНИЯ
КРИОГЕННЫМИ ТРАНСПОРТНЫМИ СИСТЕМАМИ****И.А. Астраханцева, Т.Е. Котенев, С.В. Горев, Р.Г. Астраханцев, П.Н. Грименицкий**

Ирина Александровна Астраханцева* (ORCID 0000-0003-2841-8639), Тимофей Евгеньевич Котенев (ORCID 0009-0003-4963-2156), Сергей Владимирович Горев (ORCID 0000-0002-4370-9533), Роман Геннадьевич Астраханцев (ORCID 0000-0001-9880-2826), Павел Николаевич Грименицкий (ORCID 0009-0002-3379-8760)

Ивановский государственный химико-технологический университет, Шереметевский проспект, 7, Иваново, 153000, Россия

E-mail: i.astrakhantseva@mail.ru*, kotenev@me.com, gorev@srosovnet.ru, rgastrakhantsev@gmail.com, grim@isuct.ru

В научной статье исследуется эффективность использования моделей логистической регрессии и случайного леса для прогнозирования управленческих решений в многослойных криогенных системах. Исследование проводилось с использованием датасета, содержащего 730 наблюдений. Вводятся дополнительные нелинейные признаки, такие как произведение уровней и давлений газов (He и N_2) и уровень азота, равный нулю, что позволило улучшить качество моделирования за счет более корректного учета взаимодействий переменных.

Модель логистической регрессии показала значение $F1$ -меры 0,653, указывая на высокую точность в предсказании класса, не требующего действий, но испытывала трудности с предсказанием класса, требующего вмешательства, из-за несбалансированности классов в данных.

Модель случайного леса продемонстрировала высокую производительность, что подтверждается значением $F1$ -меры 0,925, указывающей на эффективное сочетание точности и полноты предсказаний модели. Анализ матрицы ошибок показал надежную способность модели определять состояния, не требующие вмешательства, при минимальном допущении ошибок второго типа, что критически важно для поддержания стабильности системы.

В работе была проведена оценка значимости признаков, которая подтвердила, что уровни и давления газов являются основными драйверами предсказаний модели случайного леса. Доказывается эффективность использования моделей логистической регрессии и случайного леса для прогнозирования управленческих решений в криогенных системах.

Ключевые слова: многослойная криогенная система, криогенный контейнер-цистерна, вакуумный экран, жидкий азот, жидкий гелий, теплозащитный экран, вакуумная изоляция, машинное обучение, обучение с учителем, классификация, случайный лес, логистическая регрессия, значимость факторов, полиномиальные признаки.

INTEGRATION OF LINEAR REGRESSION AND RANDOM FOREST METHODS INTO INTELLIGENT SUPPORT SYSTEMS FOR MANAGING CRYOGENIC TRANSPORT SYSTEMS**I.A. Astrakhantseva, T.E. Kotenev, S.V. Gorev, R.G. Astrakhantsev, P.N. Grimenitsky**

Irina A. Astrakhantseva*(ORCID 0000-0003-2841-8639), Timofey E. Kotenev (ORCID 0009-0003-4963-2156), Sergey V. Gorev (ORCID 0000-0002-4370-9533), Roman G. Astrakhantsev (ORCID 0000-0001-9880-2826), Pavel N. Grimenitsky (ORCID 0009-0002-3379-8760)

Ivanovo State University of Chemistry and Technology, Sheremetevsky avenue, 7, Ivanovo, 153000, Russia
E-mail: i.astrakhantseva@mail.ru*, kotenev@me.com, gorev@srosovnet.ru, rgastrakhantsev@gmail.com, grim@isuct.ru

This scientific article investigates the effectiveness of using logistic regression and random forest models for predicting management decisions in multilayer cryogenic systems. The study was conducted using a dataset containing 730 observations. Additional nonlinear features were introduced, such as the product of gas levels and pressures (He and N₂) and the nitrogen level being zero, which improved the modeling quality by more accurately accounting for variable interactions.

The logistic regression model demonstrated an F1-score of 0.653, indicating high accuracy in predicting the class that does not require action, but faced difficulties in predicting the class that requires intervention due to class imbalance in the data.

The random forest model showed high performance, as evidenced by an F1-score of 0.925, indicating an effective combination of accuracy and recall in the model's predictions. The confusion matrix analysis demonstrated the model's reliable ability to identify states that do not require intervention, with minimal type II errors, which is critically important for maintaining system stability.

The study included an assessment of feature importance, which confirmed that gas levels and pressures are the primary drivers of the random forest model's predictions. This proves the effectiveness of using logistic regression and random forest models for predicting management decisions in cryogenic systems.

Keywords: multilayer cryogenic system, cryogenic tank container, vacuum screen, liquid nitrogen, liquid helium, heat shield, vacuum insulation, machine learning, supervised learning, classification, random forest, logistic regression, significance of factors, polynomial features.

Для цитирования:

Астраханцева И.А., Котенев Т.Е., Горев С.В., Астраханцев Р.Г., Грименицкий П.Н. Интеграция методов линейной регрессии и случайного леса в системы интеллектуальной поддержки управления криогенными транспортными системами. *Известия высших учебных заведений. Серия «Экономика, финансы и управление производством» [Ивэкофин]*. 2024. № 03(61). С.81-90. DOI: 10.6060/ivecofin.2024613.692

For citation:

Astrakhantseva I.A., Kotenev T.E., Gorev S.V., Astrakhantsev R.G., Grimenitsky P.N. Integration of linear regression and random forest methods into intelligent support systems for managing cryogenic transport systems. *Ivecofin*. 2024. N 03(61). С.81-90. DOI: 10.6060/ivecofin.2024613.692 (in Russian)

ВВЕДЕНИЕ

Управление криогенными транспортными системами сопряжено с многочисленными проблемами. В первую очередь они связаны с чувствительными характеристиками криогенных веществ и настоятельной необходимостью точного регулирования температуры во время транспортировки. Эти системы имеют важное значение в таких секторах, как исследование космоса, медицинское применение и логистика промышленных газов, где безопасная и эффективная транспортировка криогенных материалов является обязательной. Интеграция передовых аналитических методологий в интеллектуальные системы поддержки является актуальным подходом повышения надежности принятия решений и оперативной эффективности управления этими сложными системами [1, 2].

Управление криогенными транспортными системами требует более ювелирных технологий из-за сложного характера криогенных операций, которые включают поддержание чрезвычайно низких температур и обращение с летучими веществами. В этом исследовании представлена интеграция алгоритмов линейной регрессии и случайного леса в интеллектуальные системы поддержки,

которые предназначены для оптимизации обработки и транспортировки криогенных материалов в различных отраслях промышленности.

Алгоритмы линейной регрессии и случайного леса являются надежными инструментами благодаря своим прогностическим возможностям и адаптируемости к разным сценариям обработки данных. Линейная регрессия предлагает простую модель, объясняющую взаимосвязь между зависимыми и независимыми переменными, обеспечивая четкое представление о том, как входные данные системы влияют на выходные данные. Этот традиционный статистический метод обеспечивает ясность и простоту интерпретации, позволяя операторам понять, как изменения во входных переменных, таких как качество изоляции, температура окружающей среды и свойства материалов, напрямую влияют на такие результаты, как стабильность температуры и потребление энергии.

Случайный лес, метод ансамблевого обучения, превосходно справляется с управлением нелинейными взаимодействиями и многомерными пространствами. Этот алгоритм устраняет сложности, возникающие из-за нелинейных взаимосвязей и многомерных данных, которые невозможно адекватно смоделировать с помощью ли-

нейной регрессии. Благодаря построению множества деревьев решений и объединению их прогнозов, метод случайного леса обеспечивает надежную прогностическую модель, которая отражает сложные взаимодействия между рабочими параметрами. Этот метод особенно эффективен для повышения точности прогнозов в условиях неопределенности и изменчивости, которые характерны для криогенной транспортировки из-за воздействия внешней среды и специфических характеристик материала.

Интеграции этих двух методологий способствует сложная система машинного обучения, которая включает мониторинг данных в режиме реального времени, адаптивные алгоритмы обучения и непрерывное уточнение модели. Этот подход позволяет системе динамически адаптироваться к новым данным и условиям эксплуатации, тем самым повышая точность прогнозирования и оперативность реагирования системы.

МЕТОДЫ И МАТЕРИАЛЫ

В настоящем исследовании этот метод машинного обучения применяется для разработки модели прогнозирования управленческих решений в многослойной криогенной системе с вакуумной изоляцией и охлаждением жидким азотом. Она представляет собой инженерно-техническое транспортное сооружение, состоящее из нескольких компонентов, способствующих эффективной изоляции и охлаждению [3,4]. Основная структура системы включает внутренний сосуд, содержащий криогенный материал (жидкий гелий), окруженный вакуумной камерой для минимизации теплопередачи конвекцией и проводимостью. Между внутренним сосудом и вакуумной камерой располагаются многослойные изоляционные материалы, состоящие из чередующихся слоев отражающего и непрозрачного материалов, что существенно снижает теплопередачу излучением. Внешний сосуд системы содержит жидкий азот, который используется для дополнительного охлаждения и обладает температурой кипения около -196°C .

В работе многослойных криогенных систем неконтролируемое преобразование этой энергии может привести к повышению внутренних температур, что негативно сказывается на эффективности и безопасности системы. Управление процессом преобразования кинетической энергии в тепло является ключевым аспектом в обеспечении стабильной работы и поддержания низких температур, необходимых для эффективного функционирования криогенных систем. Поэтому понимание и снижение влияния преобразования кинетической энергии в тепловую является критически важным фактором при разработке и эксплуатации транспортных систем для криогенных материалов [5, 6].

В данном исследовании был использован ряд библиотек для анализа данных, построения моделей машинного обучения и визуализации результатов. Для обработки и анализа данных были использованы библиотеки `pandas` и `pumpy` - инструменты для работы с табличными данными и многомерными массивами соответственно. Для предобработки данных и кодирования категориальных признаков была использована библиотека `sklearn.preprocessing` [7, 8, 20]. Она содержит классы, такие как `OneHotEncoder`, который преобразует категориальные признаки в бинарные. Для построения моделей машинного обучения были использованы библиотеки `sklearn.ensemble` и `sklearn.linear_model`. Класс `GradientBoostingClassifier` использовался для реализации модели градиентного бустинга. Для оценки производительности моделей были использованы метрики из модуля `sklearn.metrics`, такие как `confusion_matrix` и `f1_score`. Для визуализации результатов и объяснения прогнозов моделей были использованы библиотеки `shap` и `matplotlib.pyplot`. Для создания новых полиномиальных признаков, включая взаимодействия между признаками использовалась функция `Polynomial Features` [9, 10].

АНАЛИЗ ПРЕДМЕТНОЙ ОБЛАСТИ

Криогенные транспортные системы разработаны таким образом, чтобы жидкий азот, гелий или кислород поддерживались ниже точки кипения, оставались в жидком состоянии при чрезвычайно низких температурах. Точное управление такими системами необходима как для безопасности и эффективности операций, так и для сохранения криогенных материалов. Использование алгоритмов машинного обучения для управления криогенными системами представляет собой сложную задачу из-за сложности и критичности ошибок [11, 12, 19, 21].

Для управления и минимизации рисков при управлении криосистемами одним из алгоритмов прогнозирования управляющей функции системы являются логистическая регрессия и случайный лес [13, 14, 15].

Логистическая регрессия - статистическая модель, которая используется для предсказания вероятности наступления определенного события путем применения логистической функции. Она используется для бинарной классификации для предсказания управляющих действий с системой.

Эта функция также известна как сигмоидная функция для преобразования линейной комбинации входных переменных в вероятность. Модель предсказывает вероятность $P(Y=1)$, что

событие Y принимает значение 1. Математически модель логистической регрессии можно выразить следующим образом.

$$P(y_i = 1|X) = \frac{1}{1+e^{-(\beta_0+\beta_1x_1+\dots+\beta_nx_n)}} \quad (1)$$

где $P(y = 1|x)$ - вероятность зависимой переменной y_i (например, состояние отказа или успеха компонента) при заданных предикторах X ;

$\beta_0, \beta_1, \dots, \beta_n$ - коэффициенты;

x_1, \dots, x_n - независимые переменные, например такие как давление и уровень гелия и азота;

e - основание натурального логарифма.

Эта модель показывает влияния каждого предиктора на вероятность различных состояний системы.

Одним из преимуществ алгоритма логистической регрессии является его интерпретируемость, изменение x_i на одну единицу приводит к изменению логарифма шансов на β_i единиц. Коэффициенты логистической регрессии представляют собой количественную меру воздействия изменений независимых переменных на вероятность исхода события. Это показывает понятную взаимосвязь между переменными и результатами. Логистическая регрессия имеет высокую вычислительную эффективность, что позволяет использовать его в условиях, ограниченных вычислительных ресурсах.

Однако логистическая регрессия имеет ключевое ограничение - это предположение о линейности взаимосвязей между зависимой переменной и независимыми переменными, что не всегда отражено в реальных данных. Алгоритм чувствителен к выбросам и шуму в данных, что искажает оценку параметров и делает предсказания менее надежными. Еще одним из недостатков логистической регрессии является то, что модель предполагает равномерное распределение по классам и показывает снижение эффективности при несбалансированности классов.

Алгоритм машинного обучения случайный лес представляет более глубокое понимание сложных взаимодействий между множественными переменными для понимания динамики криогенных систем. Выбор этого алгоритма случайного леса в дополнении к логистической регрессии обусловлен недостатками последнего, в частности чувствительности к выбросам и несбалансированности классов. Случайный лес использует ансамблевый подход с множеством деревьев и позволяет проводить более глубокий анализ данных.

Случайный лес относится к ансамблевым методам машинного обучения и строится на основе множества деревьев решений, каждое из которых обучается на случайно выбранной подвыборке тренировочного набора данных с применением бутстрэппинга (bootstrap aggregating или bagging).

Алгоритм случайного леса строит множество деревьев решений. Каждое дерево обучается на случайно выбранной подвыборке из тренировочной выборки. Поэтому прогноз для нового объекта получается путем голосования результатов всех деревьев.

$$Y = \text{мода}(\hat{y}_1, \hat{y}_2, \dots, \hat{y}_T) \quad (2)$$

где T — количество деревьев,

\hat{y}_i - прогноз от i -го дерева.

В конструкции каждого дерева в ансамбле случайного леса используется алгоритм, который начинается с выбора подвыборки из исходного набора данных методом бутстрэп, то есть выборкой с возвращением. На каждом этапе разделения узла в дереве происходит случайный отбор подмножества признаков, среди которых далее определяется наилучший признак для разделения. Этот процесс разделения продолжается либо до достижения максимально установленной глубины дерева, либо до достижения слишком малого количества объектов, которое делает невозможным дальнейшее разделение. Случайный лес интегрирует прогнозы отдельных деревьев. Это снижает риск переобучения. Такая модель эффективна при обработке многомерных и неоднородных данных от криогенных систем.

Основным преимуществом этой модели является его высокая точность прогнозирования и устойчивость к переобучению. Поэтому модель эффективна в обработке данных с большим количеством признаков и не требует масштабирования или трансформации переменных. Случайный лес автоматически оценивает важность признаков в виде показателя важности. Основным недостатком этого алгоритма является высокая вычислительная сложность и потребность в ресурсах.

Совместное использование логистической регрессии и случайного леса при управлении многослойной криогенной системой повышают точность прогнозов и управления системой в связи с использованием возможностей логистической регрессии моделирования прямых связей и алгоритма случайного леса для обработки сложных нелинейных взаимосвязей признаков [16, 17, 18].

Для анализа многослойной криогенной системы с использованием линейной регрессии и метода случайного леса использовался датасет из 730 наблюдений. Каждая запись включала в себя информацию о характеристиках многослойной криогенной системы, в том числе дату наблюдения, текущее состояние контейнера (в движении ли система или находится в состоянии покоя), уровень и давление азота и гелия. Дополнительно для каждого наблюдения была определена целевая переменная y . Управленческое решение является бинарным - оставить систему без изменений или выполнить определенные действия (рис. 1).

№ п/п	Дата	Статус контейнера	LI - 01 уровень He	PI-01/PI-03 давление He (psi)	LI-02 уровень N2	PI-02 давление N2 (psi)	Управленческое решение
0	1 14.02.2023	В пути	0	64	0	1	Требуется действие
1	2 14.02.2023	В пути	0	60	0	1	Требуется действие
2	3 14.02.2023	В пути	0,1	2	0	0	Требуется действие
3	4 15.02.2023	В пути	0	67	0	0	Требуется действие
4	5 15.02.2023	В пути	0	60	0	2	Требуется действие
...
722	726 18.12.2023	В пути	8,6	22	44	3,9	Оставить без действий
723	727 18.12.2023	В пути	7,9	6	30	4	Оставить без действий
724	728 18.12.2023	В пути	0,9	94	33	1,5	Оставить без действий
725	729 18.12.2023	В пути	0	60	30	6,9	Оставить без действий
726	730 18.12.2023	В пути	7,5	70	55	9,5	Оставить без действий

Рисунок 1. Дата сет
Figure 1. Data set

Для реализации алгоритмов линейной регрессии и случайного леса при анализе многослойной криогенной системы используются данные об уровне азота и давления в контейнере. Уровень азота напрямую влияет на эффективность теплоизоляции и поддержание температурного баланса. Изменения в уровне азота могут указывать на потери из-за испарения или неэффективности изоляционных материалов, что в свою очередь может привести к снижению охлаждения. Давление азота отражает термодинамическое состояние системы и требует контроля для предотвращения рисков образования конденсата, обледенения или ледяных пробок.

Гелий обеспечивает необходимые условия для поддержания сверхнизких температур. Изменения в их уровне или давлении указывают на потери или неэффективность теплообмена. Поддержание определенного давления гелия необходимо для гарантированной безопасности и эффективности работы системы. Информация о статусе контейнера позволяет адаптировать параметры системы для обеспечения оптималь-

ных условий покоя или транспортировке криогенных материалов. Это не только снижает риск потери материала из-за утечек, но и обеспечивает безопасность и эффективность работы всей криогенной системы.

В рамках предварительного анализа данных был проведено исследование на сбалансированность классов для задачи логистической регрессии с помощью функции `groupby` из библиотеки Pandas. Данные были сгруппированы по категории 'Управленческое решение для задачи логистической регрессии', и с помощью метода `count()` было определено количество записей для каждой категории решений. Результаты данной операции указывают на значительную несбалансированность классов. Большинство наблюдений (688 из 727) относятся к управленческому решению "Оставить без действий", в то время, как только 39 записей классифицируются как "Требуется действие". Именно поэтому в рамках анализа данных в дополнении к логистической регрессии используется случайный лес, который не чувствителен к несбалансированности классов.

`data.groupby('Управленческое решение для задачи логистической регрессии').count()`

	№ п/п	Дата	Номер контейнера	Статус контейнера	LI - 01 уровень He	PI-01/PI-03 давление He (psi)	LI-02 уровень N2	PI-02 давление N2 (psi)	Управленческое решение для задачи классификации
Управленческое решение для задачи логистической регрессии									
Оставить без действий	688	688	688	688	688	688	688	688	688
Требуется действие	39	39	39	39	39	39	39	39	39

Рисунок 2. Распределение данных в дата сете по классам
Figure 2. Distribution of data in the data set by class

Анализируемый набор данных также по уровню азота (N₂) в системе характеризуется ненормальным распределением. Значительное количество наблюдений показывает нулевой уровень азота, что указывает на данные, собранные при запуске криосистем (рис. 3).

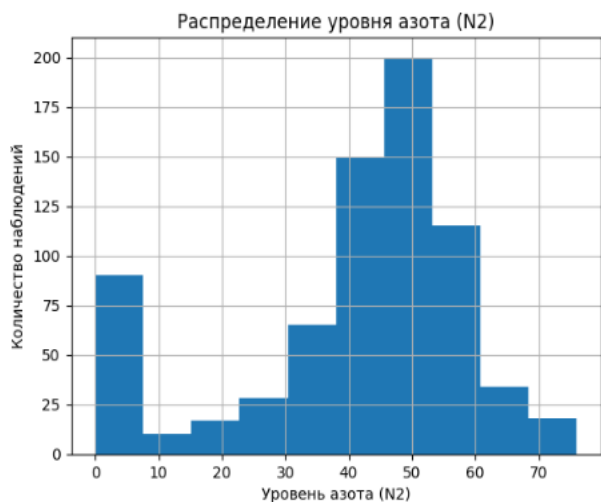


Рисунок 3. Распределение данных в дата сете по уровню азота

Figure 3. Distribution of data in the data set by nitrogen level

Такое распределение существенно влияет на эффективность применения алгоритмов машинного обучения, особенно тех, которые предполагают нормальность данных, в частности алгоритмы классификации. Ненормальное распределение приводит к искажению результатов прогноза, если не применяются специальные методы предварительной обработки данных, такие как нормализация или стандартизация. В результате этого было принято решение в исходном наборе данных DataFrame `data` реализовать процесс добавления нового столбца "LI-02 уровень N₂ == 0", который предназначен для выделения нулевых значений в столбце "LI-02 уровень N₂". Для каждого элемента этого столбца применяется функция, которая возвращает значение 1, если исходное значение равно нулю, и значение 0 в противном случае. Такая трансформация помогает более точной настройке моделей машинного обучения, поскольку такие записи влияют на обучение и качество моделирования.

В рамках предварительной обработки данных была проведена операция генерации новых признаков на основе взаимодействий между параметрами, существующими в наборе данных, с использованием метода полиномиальной регрессии. Для этого применим класс `PolynomialFeatures` из библиотеки `scikit-learn` с активированным параметром `interaction_only=True`. Это позволяет ге-

нерировать признаки, которые представляют собой произведения исходных признаков без включения их степеней. Были созданы новые признаки путем вычисления произведений между уровнем гелия ("LI - 01 уровень He") и его давлением ("PI-01/PI-03 давление He (psi)"), а также аналогичных взаимодействий для азота ("LI-02 уровень N₂" и "PI-02 давление N₂ (psi)"). Полученные признаки были дополнительно трансформированы путем извлечения квадратного корня. Такие операции способствуют улучшению линейности взаимосвязей и повышению эффективности последующих аналитических моделей.

Далее эти новые признаки были интегрированы в исходный набор данных путем соединения двух и более последовательностей или структур данных в одну. После этого выполнена операция очистки данных от пропущенных значений и сброса индексации для обеспечения целостности, логичности и однородности данных. Эти операции улучшили информативность дата сета за счет включения взаимодействий между параметрами, а также подготовили его к более эффективному применению в методах машинного обучения. Учет взаимодействий между признаками влияет на точность прогнозных моделей.

Таким образом, в рамках предобработки данных с целью улучшения предсказательной способности модели были созданы дополнительные признаки на основе существующего дата сета. Этот процесс направлен на выявление потенциальных нелинейных взаимосвязей между различными предикторами и целевой переменной - управленческим решением для многослойной криосистемы. В результате такого нелинейного преобразования предикторов были добавлены следующие новые признаки: квадратный корень из произведения давления и уровня гелия, квадратный корень из произведения уровня и давления азота, уровень азота равный нулю.

РЕЗУЛЬТАТЫ И ДИСКУССИИ

Для подготовки данных к моделированию был применен OneHotEncoder из библиотеки `sklearn.preprocessing`, который позволил преобразовать категориальные переменные в бинарный формат. Для базового сравнения и оценки эффективности разработанных моделей был использован `DummyClassifier` из `sklearn.dummy`, который предоставляет классификатор для выполнения минимальных базовых проверок эффективности модели.

Далее дата сет был разделен на обучающую (80%) и тестовую (20%) выборки с помощью функции `train_test_split` из `sklearn.model_selection`. Стратификация по целевой переменной у обеспечивал равномерное распределение классов

в разделенных выборках. Стратификация в данном случае являлась важным шагом, чтобы предотвратить наличие одного класса в обучающей выборке, а другого в тестовой.

После построения модели логистической регрессии из `sklearn.linear_model`, проводилась оценка её производительности с помощью метрик, таких как матрица ошибок (`confusion matrix`) и F1-скор (`f1_score`) из `sklearn.metrics`. Матрица ошибок позволяет визуально оценить, как модель классифицирует различные классы, тогда как F1-скор предоставляет гармоническое среднее точности и полноты, что является необходимой метрикой в случае несбалансированных классов.

Матрица ошибок показала следующее распределение предсказаний (рис. 4):

- True Negative (TN): 123 — правильные предсказания класса 0 (управленческое решение по криосистеме "Не требует действий").
- False Positive (FP): 14 — ошибочные предсказания класса 1, когда истинный класс был 0.
- False Negative (FN): 3 — ошибочные предсказания класса 0, когда истинный класс был 1.
- True Positive (TP): 5 — правильные предсказания класса 1 (управленческое решение по системе "Требуется действия").

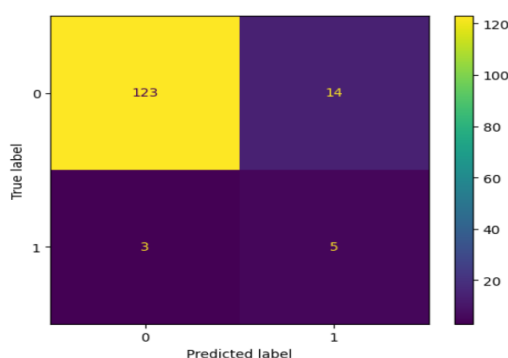


Рисунок 4. Матрица ошибок для логистической регрессии
Figure 4. Confusion Matrix for Logistic Regression

Логистическая регрессия показывает высокую точность в предсказании класса 0, с 123 правильными предсказаниями из общего числа предсказаний для этого класса. Однако модель показывает сравнительно высокое число ошибок (14) в предсказаниях класса 1, когда истинный класс был 0, и пропускает 3 случая класса 1, предсказывая их как класс 0. Несмотря на это, модель правильно идентифицирует 5 случаев класса 1.

Для более точной оценки производительности модели была вычислена метрика F1-score, которая является гармоническим средним точности и полноты. Использование усреднения по

классам ('macro' average) позволяет учитывать несбалансированность классов в данных. Рассчитанное значение F1-score составило 0,653, что отражает среднюю производительность модели для обоих классов.

Выбор параметров модели, а именно использование штрафа L2 и заданных весов классов (`class_weight={0:0.1, 1:0.9}`), был направлен на улучшение предсказательной способности модели для менее представленного класса. Тем не менее, полученные результаты свидетельствуют о необходимости дальнейшего улучшения модели для повышения её точности в отношении класса 1.

Эти результаты подчеркивают проблему существенного дисбаланса классов в обучающем наборе данных, который привел к предвзятости модели в пользу доминирующего класса. Такая модель, хоть и показывает высокую точность в определении одного типа ситуаций, оказывается бесполезной для целей предсказания критических событий, что может иметь серьезные последствия в практическом применении, особенно в областях, где своевременное реагирование критически важно.

В дополнении к модели логистической регрессии была разработана модель классификации с использованием алгоритма случайного леса. Она была реализована с помощью класса `Random ForestClassifier` из библиотеки `scikit-learn`. Модель была сконфигурирована с параметрами 200 деревьев решений и максимальной глубиной дерева 9.

В процессе оптимизации модели случайного леса выявлено, что количество деревьев и их глубина влияют на количество ошибок. Параметры, отличные от 200 деревьев и максимальной глубины 9, приводили к увеличению ошибок. Меньшее число деревьев и меньшая глубина не захватывают сложность данных, в то время как избыточное увеличение этих параметров делает модель чрезмерно адаптированной к обучающей выборке. Оптимальные параметры выбраны для достижения баланса между точностью и вычислительной эффективностью. При этом снижается риск недообучения и переобучения, что обеспечивает стабильность и надежность системы. Первоначально модель подверглась обучению на тренировочном наборе данных. После чего была оценена на тестовом наборе.

Для оценки качества классификации использовалась метрика F1-меры, которая рассчитывается усреднением по всем классам ('macro average'). Это позволяет учесть влияние несбалансированности классов. Значение F1-меры равной 0,925 говорит о высокой эффективности бинарной классификационной модели в управлении многослойной криосистемой. Макро-усредненная

F1-мера рассчитывается путем отдельного вычисления F1-меры для каждого класса с последующим их усреднением. Такой подход обеспечивает равное внимание к каждому классу, что критически важно в условиях несбалансированных классов, представленных в исследуемой дата-сети.

Высокий показатель F1-меры показывает, что модель обладает как высокой точностью, так и высокой полнотой в процессе идентификации состояний многослойной криосистемы. Точность модели означает её способность корректно определять только релевантные случаи в каждом классе, в то время как полнота отражает способность обнаруживать все релевантные случаи в каждом классе. Модель случайного леса успешно минимизирует количество ложных срабатываний и обеспечивает высокую чувствительность к действительно требующим внимания состояниям в криосистеме.

Для визуализации результатов классификации была использована матрица ошибок, созданная с помощью функции `'confusion_matrix'` библиотеки `'scikit-learn'` (рис. 5).

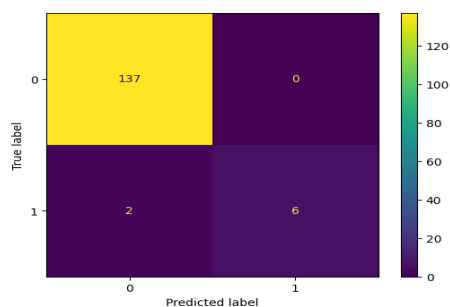


Рисунок 5. Матрица ошибок для модели бинарной классификации
Figure 5. Confusion Matrix for Binary Classification Model

Матрица ошибок показывает результаты классификации модели машинного обучения для двух классов. Из 137 случаев, когда истинная метка была 0 (управленческое решение по криосистеме "Не требует действий"), модель правильно предсказала все 137 случаев, не допустив ни одной ошибки ложноположительного результата. Из 8 случаев, когда истинная метка была 1 (управленческое решение по системе "Требуется действия"), модель правильно предсказала 6 случаев, допустив 2 ошибки ложноотрицательного результата. Эти результаты показывают высокую специфичность модели. Она эффективно идентифицировала большинство случаев класса "не требует действий". Это указывает на надежность модели в предотвращении ложноположительных решений.

В рамках исследования была проведена оценка важности признаков модели случайного леса с использованием библиотеки SHAP

(SHapley Additive exPlanations), которая позволяет интерпретировать предсказания модели. Для этого использовался класс `TreeExplainer` из библиотеки `shap`, предназначенный для деревьев решений и ансамблей деревьев. Объект `'TreeExplainer'` был инициализирован с параметрами обученной модели градиентного бустинга и названиями признаков, что обеспечило точное отображение данных в анализе.

Объект `TreeExplainer` был инициализирован с параметрами обученной модели случайного леса и названиями признаков, что обеспечило точное отображение данных в анализе. Далее для тестового набора данных были вычислены значения SHAP с помощью метода `shap_values`, которые представляют величину влияния каждого признака на каждое предсказание модели.

Для визуализации и сравнения вклада признаков была создана столбчатая диаграмма, отображающая нормализованные значения SHAP. Нормализация выполнена путем деления значений SHAP на сумму средних абсолютных значений по всем признакам, что позволило представить относительный вклад каждого признака в магнитуду выходных значений модели (рис. 6).

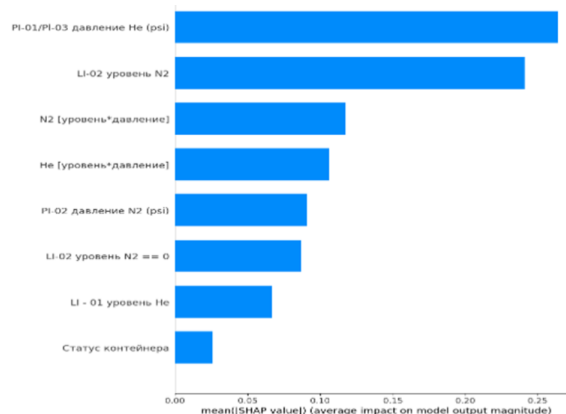


Рисунок 6. Значимость признаков в модели случайного леса
Figure 6. Feature Importance in Random Forest Model

На диаграмме показаны следующие признаки, упорядоченные по степени их влияния на модель:

1. **PI-01/PI-03 давление He (psi)** - признак имеет наибольшее среднее влияние на предсказания модели.
2. **LI-02 уровень N2** - значимый признак, следующий по важности после давления He.
3. **N2 [уровень*давление]** - взаимодействие между уровнем и давлением N2 также играет важную роль.
4. **He [уровень*давление]** - взаимодействие между уровнем и давлением He.

5. **PI-02 давление N2 (psi)** - важный показатель, связанный с давлением N2.
6. **LI-02 уровень N2 == 0** - наличие или отсутствие уровня N2.
7. **LI - 01 уровень He** - уровень He также значим для модели.
8. **Статус контейнера** - признак, имеющий наименьшее влияние среди представленных.

Анализ важности признаков позволяет лучше понять, какие факторы наиболее сильно влияют на предсказания модели. Высокие значения SHAP показывают, что соответствующие признаки оказывают значительное влияние на результативность модели. Таким образом, полученные данные могут быть использованы для дальнейшего улучшения модели и оптимизации управления системой.

В частности, наиболее значимые признаки, такие как давление гелия и уровень азота He и уровень N2, следует учитывать при разработке стратегий управления многослойной криосистемой. Эти признаки могут быть критически важными для повышения точности и надежности предсказаний модели, а также для минимизации рисков, связанных с эксплуатацией криосистемы.

Таким образом, обученная модель случайного леса является надежным инструментом мониторинга и контроля многослойных криосистем, что важно в условиях точного и эффективного реагирования на аномалии для предотвращения нежелательных последствий в критически важных операциях.

ЗАКЛЮЧЕНИЕ

В рамках научного исследования были разработаны и оценены две модели классификации - логистическая регрессия и случайный лес.

Обе модели показали свою пригодность для мониторинга и контроля многослойных криосистем.

Модель логистической регрессии показывает значение F1-меры 0.653, модель случайного леса - 0.925. Высокие значения F1-меры случайного леса и анализ важности признаков с помощью SHAP показывают, что эта модель обладает как высокой точностью, так и полнотой. Модель случайного леса надежно идентифицирует значимые факторы, влияющие на состояние криосистемы, и минимизирует количество ложных срабатываний.

Таким образом, обученная модель случайного леса является надежным инструментом мониторинга и контроля многослойных криосистем, что критически важно в условиях точного и эффективного реагирования на аномалии для предотвращения нежелательных последствий в критически важных операциях. Модель случайного леса может эффективно минимизировать ложные срабатывания и точно обнаруживать состояния, требующие вмешательства, что критически важно для поддержания стабильности и безопасности в управлении криосистемами.

Обе модели классификации, логистическая регрессия и случайный лес, предоставляют ценные инструменты для мониторинга и контроля криосистем. Логистическая регрессия показывает простоту и интерпретируемость, хотя менее эффективна в условиях несбалансированности классов. Модель случайного леса обеспечивает более высокую точность и полноту, что делает её предпочтительным выбором для критически важных операций.

Авторы заявляют об отсутствии конфликта интересов.

The authors declare no conflict of interest.

ЛИТЕРАТУРА

1. **Филин Н.В., Буланов А.Б.** Жидкостные криогенные системы. Л.: Машиностроение. 1985. 247 с.
2. **Архаров А.М., Беляков В.П., Микулин Е.И. и др.** Криогенные системы: Учебник для вузов по курсу «Криогенная техника». М.: Машиностроение. 1987. 536 с.
3. **Микулин Е.И.** Криогенная техника. М.: Машиностроение. 1969. 272 с.
4. **Баранов А.Ю., Соколова Е.В.** Хранение и транспортировка криогенных жидкостей. Часть 1: Учебное пособие. - СПб.: Университет ИТМО. 2017. 95 с.
5. **Van Sciver S.W.** Helium Cryogenics. New York: Springer Science. 2012. 470 p. DOI: 10.1007/978-1-4419-9979-5.
6. **Гордина Н.Е., Мельников А.А., Гусев Г.И., Гушин А.А., Румянцев Р.Н., Астраханцева И.А.** Использование механохимической и плазмохимической обработок при синтезе каталитических систем на основе вермикулита и оксихлорида циркония. *Изв. вузов. Химия и хим. технология.* 2022. Т. 65. Вып. 5. С. 43-57. DOI: 10.6060/ivkkt.20226505.6612. EDN KDQEVVD.
7. Учебник по машинному обучению. <https://education.yandex.ru/handbook/ml>.

REFERENCES

1. **Filin N.V., Bulanov A.B.** Liquid Cryogenic Systems. Leningrad: Mashinostroenie. 1985. 247 p. (in Russian).
2. **Arkharov A.M., Belyakov V.P., Mikulin E.I. et al.** Cryogenic Systems: Textbook for Universities on the Course "Cryogenic Engineering". Moscow: Mashinostroenie. 1987. 536 p. (in Russian).
3. **Mikulin E.I.** Cryogenic Engineering. Moscow: Mashinostroenie. 1969. 272 p. (in Russian).
4. **Baranov A.Yu., Sokolova E.V.** Storage and Transportation of Cryogenic Liquids. Part 1: Study Guide. St. Petersburg: ITMO University. 2017. 95 p. (in Russian).
5. **Van Sciver S.W.** Helium Cryogenics. New York: Springer Science. 2012. 470 p. DOI: 10.1007/978-1-4419-9979-5.
6. **Gordina N.E., Melnikov A.A., Gusev G.I., Gushchin A.A., Rumyantsev R.N., Astrakhantseva I.A.** Mechanochemical and plasmachemical processing in the synthesis of catalytic systems based on vermiculite and zirconium oxychloride. *ChemChemTech [Izv. Vyssh. Uchebn. Zaved. Khim. Khim. Tekhnol.].* 2022. V. 65. N 5. P. 43-57. DOI: 10.6060/ivkkt.20226505.6612. EDN KDQEVVD.
7. Textbook on Machine Learning. <https://education.yandex.ru/handbook/ml>. (in Russian).

8. **Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A.** Classification and Regression Trees. New York: CRC press. 1984. 368 p.
9. **Murphy K.P.** Machine Learning: A Probabilistic Perspective. London: MIT Press. 2012. 29 p.
10. **Ho T.K.** Random Decision Forests. *Materials of the 3rd International Conference on Document Analysis and Recognition*. Montreal: QC. 1995. P. 14-16.
11. **Зимнуров М.Ф., Астраханцева И.А., Грименицкий П.Н.** Системный анализ и оптимизация количественных показателей эффективности в технологических проектах на основе гибких методологий. *Современные наукоемкие технологии. Региональное приложение*. 2023. № 3(75). С. 61-68. DOI: 10.6060/snt.20237503.0008. EDN OYNXHV.
12. **Астраханцева И.А., Горев С.В., Астраханцев Р.Г.** Системный подход к анализу фрактальной природы сложных технических систем. *Известия высших учебных заведений. Серия «Экономика, финансы и управление производством» [Ивэкофин]*. 2023. № 3(57). С. 89-97. DOI: 10.6060/ivecofin.2023573.657. EDN PSPGBG.
13. **Астраханцева И.А., Бобков С.П.** Дискретная стохастическая модель гидродинамики потока. *Моделирование систем и процессов*. 2023. Т. 16. № 2. С. 7-14. DOI: 10.12737/2219-0767-2023-16-2-7-14. EDN BYGGKR.
14. **Астраханцева И.А., Горев С.В., Астраханцев Р.Г.** Фрактальный анализ в оценке эффективности и надежности сложных технических систем. *Современные наукоемкие технологии. Региональное приложение*. 2023. № 4(76). С. 60-68. DOI: 10.6060/snt.20237604.0008. EDN NBDYHR.
15. **Морозов Е.Н., Горев С.В.** Математические модели для оптимизации машиночитаемых регулятивных систем. *Известия высших учебных заведений. Серия «Экономика, финансы и управление производством» [Ивэкофин]*. 2023. № 4(58). С. 71-78. DOI: 10.6060/ivecofin.2023584.666. EDN LLSUGG.
16. **Мамедова Л.Э., Иванова Л.Н., Алтаев Е.С.** Основные аспекты технологии искусственного интеллекта. *Известия высших учебных заведений. Серия «Экономика, финансы и управление производством» [Ивэкофин]*. 2023. № 3(57). С. 78-88. DOI: 10.6060/ivecofin.2023573.656. EDN QCPKWV.
17. **Astrakhantseva I.A., Astrakhantsev R.G., Mitin A.V.** Randomized C/C++ dynamic memory allocator. *Journal of Physics: Conference Series: 2*. Moscow. 2021. P. 012006. DOI 10.1088/1742-6596/2001/1/012006.
18. **Бобков С.П., Астраханцева И.А., Галиаскаров Э.Г.** Применение системного подхода при разработке математических моделей. *Современные наукоемкие технологии. Региональное приложение*. 2021. № 1(65). С. 66-71. DOI: 10.6060/snt.20216501.0008. EDN KOXZWY.
19. **Бобков С.П., Астраханцева И.А.** Применение агентного подхода для моделирования процессов теплопроводности. *Вестник Ивановского государственного энергетического университета*. 2022. № 2. С. 58-66. DOI: 10.17588/2072-2672.2022.2.058-066. EDN RWGSSW.
20. **Морозов Е.Н., Горев С.В.** Математические модели для оптимизации машиночитаемых регулятивных систем. *Известия высших учебных заведений. Серия «Экономика, финансы и управление производством» [Ивэкофин]*. 2023. № 4(58). С. 71-78. DOI: 10.6060/ivecofin.2023584.666. EDN LLSUGG.
21. **Bobkov S.P., Astrakhantseva I.A.** The use of multi-agent systems for modeling technological processes. *J. Phys.: Conf. Ser.* 2001 012002. DOI: 10.1088/1742-6596/2001/1/012002. EDN ZXPPKV.
8. **Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A.** Classification and Regression Trees. New York: CRC press. 1984. 368 p.
9. **Murphy K.P.** Machine Learning: A Probabilistic Perspective. London: MIT Press. 2012. 29 p.
10. **Ho T.K.** Random Decision Forests. *Materials of the 3rd International Conference on Document Analysis and Recognition*. Montreal: QC. 1995. P. 14-16.
11. **Zimnurov M.F., Astrakhantseva I.A., Grimenitsky P.N.** System Analysis and Optimization of Quantitative Performance Indicators in Technological Projects Based on Flexible Methodologies. *Modern High Technologies. Regional Application*. 2023. N 3(75). P. 61-68. DOI: 10.6060/snt.20237503.0008. EDN OYNXHV. (in Russian).
12. **Astrakhantseva I.A., Gorev S.V., Astrakhantsev R.G.** Systems Approach to the Analysis of the Complex Technical Systems Fractal Nature. *Ivecofin*. 2023. N 3(57). P. 89-97. DOI: 10.6060/ivecofin.2023573.657. EDN PSPGBG. (in Russian).
13. **Astrakhantseva I.A., Bobkov S.P.** Discrete Stochastic Model of Flow Hydrodynamics. *Modeling of Systems and Processes*. 2023. Vol. 16. N 2. P. 7-14. DOI: 10.12737/2219-0767-2023-16-2-7-14. EDN BYGGKR. (in Russian).
14. **Astrakhantseva I.A., Gorev S.V., Astrakhantsev R.G.** Fractal Analysis in Assessing the Efficiency and Reliability of Complex Technical Systems. *Modern High Technologies. Regional Application*. 2023. N 4(76). P. 60-68. DOI: 10.6060/snt.20237604.0008. EDN NBDYHR. (in Russian).
15. **Morozov E.N., Gorev S.V.** Mathematical Models for Optimization of Machine-Readable Regulation. *Ivecofin*. 2023. N 4(58). P. 71-78. DOI: 10.6060/ivecofin.2023584.666. EDN LLSUGG. (in Russian).
16. **Mamedova L.E., Ivanova L.N., Altaev E.S.** The Main Aspects of Artificial Intelligence Technology. *Ivecofin*. 2023. N 3(57). P. 78-88. DOI: 10.6060/ivecofin.2023573.656. EDN QCPKWV. (in Russian).
17. **Astrakhantseva I.A., Astrakhantsev R.G., Mitin A.V.** Randomized C/C++ dynamic memory allocator. *Journal of Physics: Conference Series: 2*. Moscow. 2021. P. 012006. DOI 10.1088/1742-6596/2001/1/012006.
18. **Bobkov S.P., Astrakhantseva I.A., Galiaskarov E.G.** Application of a System Approach in Developing Mathematical Models. *Modern High Technologies. Regional Application*. 2021. N 1(65). P. 66-71. DOI: 10.6060/snt.20216501.0008. EDN KOXZWY. (in Russian).
19. **Bobkov S.P., Astrakhantseva I.A.** Application of Agent-Based Approach for Heat Conduction Processes Simulation. *Bulletin of the Ivanovo State Power Engineering University*. 2022. N 2. P. 58-66. DOI: 10.17588/2072-2672.2022.2.058-066. EDN RWGSSW. (in Russian).
20. **Morozov E.N., Gorev S.V.** Mathematical Models for Optimization of Machine-Readable Regulation. *Ivecofin*. 2023. N 4(58). P. 71-78. DOI: 10.6060/ivecofin.2023584.666. EDN LLSUGG. (in Russian).
21. **Bobkov S.P., Astrakhantseva I.A.** The use of multi-agent systems for modeling technological processes. *J. Phys.: Conf. Ser.* 2001 012002. DOI: 10.1088/1742-6596/2001/1/012002. EDN ZXPPKV.

Поступила в редакцию 25.07.2024
Принята к опубликованию 08.08.2024

Received 25.07.2024
Accepted 08.08.2024