

УДК 004.8

Е.Т. Гарипов, И.Г. Боровской, В.В. Кручинин

Модель машинного обучения для прогнозирования аварий на технологических объектах добычи газа

Возникновение аварийных ситуаций на предприятиях приводит к прекращению или ослаблению рабочего потока, что несёт экономические убытки независимо от того, насколько быстро аварии устранены. Следовательно, возможность заранее предупредить возникновение подобных ситуаций будет крайне востребована во многих отраслях экономики. В работе реализован подход к подготовке данных из БД SCADA-системы, используя алгоритм вычисления средних значений параметров за единицу времени и корреляционный анализ. На подготовленных данных обучена модель машинного обучения, использующая реализацию метода случайного леса для классификации текущего состояния участка технологического объекта как предаварийной для симуляции работы реальной системы в потоковом режиме. Получившаяся модель без тонкой настройки и оптимизации гиперпараметров показала высокие значения точности предсказания и других метрик, что позволяет говорить о возможности использования моделей машинного обучения для решения задачи.

Ключевые слова: машинное обучение, подготовка данных, SCADA, прогнозирование аварий, технологический процесс, классификатор.

DOI: 10.21293/1818-0442-2025-28-3-53-58

Для разработки или обеспечения работы в реальном времени систем сбора, обработки, отображения и архивирования информации об объекте мониторинга или управления существуют такие решения, как SCADA-системы. Supervisory Control and Data Acquisition (диспетчерское управление и сбор данных) – это программный пакет, предназначенный для разработки прикладного ПО автоматизированного рабочего места (АРМ) оператора в области автоматизации технологических процессов, а также среда разработки для создания приложений верхнего уровня автоматизированной системы управления технологическим процессом (АСУТП) [1].

Традиционно для предсказания вероятности возникновения аварий в различных отраслях использовались такие математические методы, как теория вероятности [2], ретроспективный анализ [3], экспертно-статистический метод [4], дерево событий [5], дерево отказов [6] и др.

Однако такие алгоритмы даже будучи реализованными в современных SCADA-системах для простого анализа данных не способны устанавливать зависимости в сложносвязанных системах. В настоящее время всё больше технологических процессов автоматизируются благодаря технологии машинного обучения (МО). Модели МО позволяют прогнозировать изменение зависимых объектов, когда связь между ними не описывается простой математической зависимостью, например, как один участок технологического объекта влияет на совершенно другой. Далее представлен краткий обзор некоторых решений в предметной области.

Результаты исследования трех архитектур нейросетей: временных свёрточных сетей (TCN), многослойных перцептронов (MLP) и моделей с долгой краткосрочной памятью (LSTM) для опережающего выявления аварийных ситуаций на технологических объектах – представлены в [7]. С помощью алгоритма скользящего окна осуществлялось прогно-

зирование временного ряда путём экстраполяции данных.

В работе [8] дана методика, основанная на использовании математических моделей, получающих информацию в режиме реального времени из SCADA-системы. Разработанная методика позволяет расширить возможности по определению дефектов в узлах газотурбинных двигателей и установок.

Система предотвращения аварийных ситуаций на основе ситуационной математической модели с возможностью идентификации предаварийных ситуаций технологического процесса предложена в работе [9].

Для распознавания отказов оборудования предложено использовать нейронную сеть с LSTM-архитектурой, обученную на массиве данных смоделированных аварий [10]. Диагностика состояния технологического процесса по данным SCADA-системы происходит не с реального объекта, а с модели, точно соответствующей конкретному технологическому участку.

В статье [11] представлен обзор существующих методов обнаружения утечек в нефтегазопроводах, создана модель цифрового двойника, которая без единой ложной тревоги предсказывает нештатные ситуации на трубопроводах.

В [12] разработан автокодировщик на основе одномерной свёрточной сети с частичным применением обучения с учителем. Автокодировщик предназначен для обнаружения и локализации утечек в сетях распределения воды с помощью многомерных временных рядов.

Статья [13] предлагает систему структурного мониторинга состояния здоровья трубопровода на предмет трёх дефектов: расслабленность болтов, отложения, а также трещины у основания. Работа выполнена на данных, извлечённых из SCADA-системы, из которых отобраны признаки и построена модель искусственной нейронной сети с помощью встроенных функций в пакете MatLab.

В работе [14] применена модель на основе нейронной сети обратного распространения для предсказания состояния карбидного инструмента при высокоскоростной фрезеровке титановых пластин с помощью SCADA-системы в реальном времени.

В исследовании [15] изучен стационарный метод мониторинга рабочего состояния и обнаружения аномалий в ветряных турбинах. Метод полагается на алгоритм скользящего окна для данных в виде временного ряда. Анализ данных из SCADA-систем выполнен на основе аугментированного теста Дики-Фуллера.

Из рассмотренных статей видно, что в последние годы использованию МО для прогнозирования аварий на технологических объектах посвящается всё больше работ. Помимо этого, данный функционал по-прежнему не является встроенным во все современные SCADA-системы, что особенно актуально для относительно недавно появившихся российских аналогов.

Одна из таких систем – программный пакет APDAR от компании «Автоматизация производств». В нём на текущий момент отсутствует интеллектуальный анализ данных, позволяющий выявлять сложные зависимости и строить прогнозы за исключением простой математической экстраполяции данных. Данная работа посвящена реализации этого недостающего функционала и, соответственно, изучению возможности применения машинного обучения для решения поставленной задачи.

Целью работы является создание модели машинного обучения для прогнозирования аварий на технологических объектах добычи газа.

Для достижения поставленной цели требуется решить две задачи:

1. Провести предварительную обработку данных из БД SCADA-системы.

2. Обучить модель МО на подготовленных данных и оценить её точность предсказания. Если модель покажет точность предсказания не ниже 90%, то можно говорить о возможности использования машинного обучения для предупреждения возникновения аварий на технологических объектах.

Методы и алгоритмы решения

Специфика хранения данных в SCADA-системах заключается в том, что хотя измерения делаются в один промежуток времени, показания датчиков разных величин считываются с разной частотой. Это приводит к тому, что количество записей может отличаться в несколько раз при тех же начальной и конечной датах, что делает обучение моделей МО невозможным без предварительной обработки данных.

Если объединять исходные данные или равномерно их сокращать, то могут получиться несовпадения в значениях времён новых данных. Поэтому для решения данной задачи предлагается алгоритм вычисления среднего значения каждого параметра за каждую минуту. Порядок действий алгоритма: для каждого параметра разбить все записи на секции по минутно, в каждой секции вычислить среднее значе-

ние параметра и добавить его в новый набор данных. Длина интервала усреднения в минуту выбрана для удобства работы с данными и не имеет технологического обоснования, влияние длины этого интервала на результаты работы является предметом дальнейшего исследования.

Вторая задача подготовки исходных данных – исключение параметров, которые слабо коррелируют с целевым параметром. Решение данной задачи должно сократить время обучения модели, не оказывая существенного влияния на другие метрики. Для этого требуется провести корреляционный анализ каждого из параметров в обучающей выборке с целевым параметром по трём основным видам корреляции: Пирсона, Спирмена и Кендалла. Если коэффициенты корреляции близки к 1 или -1 (прямая и обратные корреляции, соответственно), то такие параметры следует включить в выборку; если же коэффициенты близки к нулю, то корреляция практически отсутствует. Этапы проведения анализа:

1. Для всех параметров вычислить коэффициенты корреляции между ними и целевым параметром по Пирсону, Спирмену и Кендаллу несколькими способами.

2. Если среднее значение всех коэффициентов для одного параметра больше 0,3 или меньше $-0,3$, добавить этот параметр в новый набор данных.

3. Обучить заново модель машинного обучения без исключённых коэффициентов и сравнить полученные метрики с моделью, обученной на данных со всеми параметрами.

4. Если получено приемлемое совпадение значений метрик, то анализ завершён. Иначе вернуться на шаг 2, сдвинуть границы интервала исключения параметров на 0,05 к нулю, повторить оставшиеся шаги.

Стоит отметить, что все полученные в работе результаты могут быть чувствительны к выбросам и шуму в данных, что, в свою очередь, может потребовать дополнительной предварительной обработки исходных данных. Зашумлённость исходных данных – одна из наиболее видных проблем предварительной подготовки данных, и ей посвящено немало работ. В данной работе этот аспект пропускается, так как она сосредоточена на подготовке исходных данных с учётом специфики их хранения в БД SCADA-системы, где значения параметров записываются с нерегулярными промежутками. Из-за этого, а также ввиду отсутствия подробной информации о технологическом процессе и измерительном оборудовании качество исходных данных следует принять как допущение в рамках данной работы.

Исходные данные и описание модели

Архивная копия проекта SCADA-системы содержит базу данных одного месторождения с 9 по 27 сентября 2022 г. В проекте содержатся десятки мнемосхем и сотни отслеживаемых параметров, архивные значения которых хранятся с различной частотой записи.

В системе для хранения архивных данных о событиях, авариях и значениях параметров использу-

ется СУБД временных рядов InfluxDB, для взаимодействия с которой применяется схожий с SQL язык запросов. В базе данных системы присутствуют верхние и нижние предупредительные и аварийные значения для каждого параметра. Созданная в работе модель МО имитирует данное поведение, считая аварийной ситуацию, когда значение параметра больше верхнего или меньше нижнего аварийного значений, заданных в системе.

В данной работе выбраны все изменяющиеся во времени параметры из узла коммерческого учёта газа (УКУГ), а также давления газа на УКУГ в теплообменниках двух модулей подготовки газа. В качестве целевой прогнозируемой величины выбран уровень дренажной ёмкости в одном из модулей подготовки газа. Количество записей у каждого из параметров отличается и может принимать значение от трёхсот тысяч до более чем девятисот тысяч. Уровень ёмкости не использовался напрямую: на его основе создан столбец, показывающий, была ли в данную единицу времени авария. Затем создан второй столбец со сдвигом вперёд, отвечающий за аварию в следующей единице времени.

В качестве ввода модели использованы все изменяющиеся во времени параметры, а также факт аварии в данную единицу времени, чтобы предсказать, будет ли в следующую единицу времени авария. Такая модель имитирует работу системы в потоковом режиме, когда на основании поступивших за определённый промежуток времени данных с датчиков определяется, была ли авария на целевом участке объекта. Затем на основании ряда параметров и факта аварии в текущую единицу времени делается предсказание на следующую единицу времени. Предсказывается только сам факт того, выйдет ли значение целевого параметра за верхнюю либо нижнюю границы, а не его конкретное числовое значение, т.е., двоичная величина. Таким образом, решается задача бинарной классификации текущего набора контролируемых параметров как предаварийного.

Подготовка данных, обучение и оценка модели

Алгоритм вычисления среднего значения параметров за каждую минуту уменьшил итоговое количество строк до 25 399 (что меньше в 13 раз для наименьшего и в 36 раз для наибольшего по числу записей параметра). Алгоритм гарантированно сокращает данные до одного уровня при любой частоте записи при условии одинаковых начальных и конечных времён и отсутствия пропусков минут в данных. Данный алгоритм быстрый, а также надёжный, поскольку каждое полученное значение привязано к строго определённой единице времени – минуте.

С помощью библиотеки scikit-learn языка Python создан классификатор на основе реализации метода случайного леса со значениями всех аргументов по умолчанию [16]. Эта ансамблевая модель машинного обучения обучает 100 деревьев решения для задачи классификации на различных подвыборках из набора данных и использует усреднение для улучшения точности предсказания и предотвращения переобучения.

В отличие от нейронных сетей, таких как многослойный перцептрон, этот метод не является итеративным и не использует (при значении параметров по умолчанию) весь набор данных для обучения. Исходя из этого, использование таких понятий, как количество эпох и шаг обучения, не совсем корректно в отношении данной модели. Тем не менее обучение всех деревьев требует одну итерацию алгоритма, а критерии ветвления определены параметрами модели.

Поскольку в исходном наборе данных аварии и штатные ситуации распределены неравномерно во времени, а их соотношение составляет примерно 1:4, был выбран меньший участок с более сбалансированными данными. 12 000 значений разделены почти поровну на обучающую (30% аварийных и 70% штатных ситуаций) и тестовую (33% аварий и 66% штатных ситуаций) выборки. Перемешивание данных не применялось, нормализация данных выполнена методом z-оценки [17]. С помощью получившегося набора данных построена модель машинного обучения, точность предсказания которой представлена в табл. 1.

Таблица 1

Ситуация	Показатели точности модели	
	Предсказаний	
	Правильных	Неправильных
Аварийная	2 005 (99,65%)	7 (0,35%)
Штатная	3 934 (99,85%)	6 (0,15%)
В общем	5 939 (99,78%)	13 (0,22%)

Точность вычислена двумя способами: собственным подсчётом, а также с использованием метрики `assu_gacy_score` [18], полученные значения точностей совпадают. Вычисление точности собственным подсчётом выполнено как отношение суммы верно распознанных аварий и штатных ситуаций к общему количеству тестовых значений. Количество верно распознанных аварий и штатных ситуаций, в свою очередь, вычислено поиндексным сравнением соответствующих элементов массивов тестовой выборки и массива предсказаний сначала для аварий, затем для штатных ситуаций.

В табл. 2 представлены коэффициенты корреляции, вычисленные для всех параметров из обучающей выборки относительно целевого параметра с помощью трёх разных библиотек языка Python. В сноске к таблице представлено разделение параметров на две группы по уровню корреляции, а также пояснения к параметрам в соответствии с их наименованиями в базе данных. В рамках одного типа коэффициента получены совпадения значений коэффициентов для всех использованных библиотек.

При обучении модели обнаружено, что при исключении параметров с коэффициентами корреляции в диапазоне от -0,3 до 0,3 метрики модели не претерпели изменений в пределах разрядов, как указано в табл. 3. При этом сократилось время обучения модели, что отражает коэффициент сокращения времени обучения, который равен отношению времени обучения модели с меньшим числом параметров ко времени обучения на всех данных.

Коэффициенты корреляции целевого параметра с остальными

Параметр	Numpy Пирсон	Scipy Пирсон	Scipy Спирмен	Scipy Кендалл	Pandas Пирсон	Pandas Спирмен	Pandas Кендалл
МИА-1.1 ^{*1}	-0,354	-0,354	-0,310	-0,197	-0,354	-0,310	-0,197
PIA 10.1 ^{*2}	-0,678	-0,678	-0,693	-0,572	-0,678	-0,693	-0,572
PIA 10.2 ^{**2}	-0,005	-0,005	0,028	-0,082	-0,005	0,028	-0,082
PIС T1-16a ^{**3}	-0,039	-0,039	-0,147	-0,084	-0,039	-0,147	-0,084
PIС T2-16a ^{*3}	0,653	0,653	0,710	0,532	0,653	0,710	0,532
ТТаметек ^{**4}	-0,071	-0,071	-0,018	-0,018	-0,071	-0,018	-0,018
ТТкип ^{**4}	-0,223	-0,223	-0,167	-0,068	-0,223	-0,167	-0,068
ТТхром ^{**4}	0,121	0,121	0,140	0,126	0,121	0,140	0,126

* Высокий уровень корреляции, ** низкий уровень корреляции.

¹ Влажность газа H₂O (ppmv) в хроматографе узла коммерческого учёта газа (УКУГ).

² Давления гелия (МПа) в двух разных баллонах в хроматографе, УКУГ.

³ Давления газа (МПа) на УКУГ в теплообменниках модулей подготовки газа МПГ-1 и МПГ-2 соответственно.

⁴ Температуры (°С) в помещении АМЕТЕК, щитовой КИП и помещении хроматографов соответственно, УКУГ.

Таблица 3

Метрики модели

Метрика	Значение
Простая точность (accuracy score)	99,7
Сбалансированная точность	99,7
F1-мера (положительный класс – авария)	99,6
Коэффициент фи (Matthews corcoeff)	0,995
Соотношения вероятности классов (LR+; LR-) (class likelihood ratios)	561; 0,002
Площадь под характеристической кривой (ROC AUC)	0,997
Средняя точность (Average Precision) (PR AUC)	0,994
Логарифмическая потеря (Log loss)	0,078
Потеря по Брайеру (Brier score loss)	0,002
Коэффициент сокращения времени обучения	0,555

Это позволяет говорить о том, что исключение из обучающей выборки параметров, слабо коррелирующих с целевым параметром, не оказывает существенного влияния на производительность модели, но позволяет обучать модель быстрее.

Сокращение количества параметров позволит повысить производительность на устройствах с малыми вычислительными мощностями и системными ресурсами, особенно при масштабировании количества параметров и их записей. Такое решение актуально для встроенной вычислительной техники, что часто встречается на технологических объектах.

Полученная точность предсказания вместе с другими метриками подтверждает предположение о правильности подхода к обработке данных, а также возможности применения технологии машинного обучения для прогнозирования возникновения аварий на технологических объектах.

Заключение

В результате выполнения работы решены задачи предварительной подготовки данных и обучения модели машинного обучения для классификации текущей ситуации на участке технологического объекта как предаварийной. Модель показала очень высокий показатель точности предсказания, что в сочетании с другими метриками даёт право говорить о возможности применения технологии машинного обучения для

предупреждающего выявления аварий на участках технологических объектов.

Научный вклад исследования заключается в формальном описании и практической реализации алгоритмов для решения задач подготовки исходных данных из БД конкретной SCADA-системы и построения классификационных моделей машинного обучения для выявления предаварийных ситуаций. Полученные результаты могут быть использованы для построения аналогичных моделей для других SCADA-систем, использующих схожие базы данных и структуры хранения данных, а также описывающих схожие технологические процессы в других отраслях экономики.

Дальнейшая работа предполагает два направления развития. Первое заключается в сравнении различных моделей машинного обучения для выявления наиболее точной для решения задачи, так как в данной работе применялась всего одна реализация метода случайного леса. Второе – в автоматизации описанных в работе шагов, т.е. реализации программного модуля для SCADA-системы, который самостоятельно строит такие модели. Предполагается, что пользователь выбирает узел на одном из участков объекта, после чего система самостоятельно решает описанные в работе задачи, задействуя большее количество параметров. Таким образом, создаётся инструмент для мониторинга состояния участка системы и поддержки принятия решения.

Литература

1. Жахина Р.У. SCADA системы и основные их компоненты / Р.У. Жахина, К.А. Костылев // Актуальные научные исследования в современном мире. – 2020. – № 4-1 (60). – С. 62–65.
2. Нычик Т.Ю. Обоснование методики оценки риска аварий судоходного шлюза // Вестник гос. ун-та морского и речного флота им. адмирала С.О. Макарова. – 2013. – № 1. – С. 15–21.
3. Стась Г.В. Оценка риска возникновения аварий при добыче бурых углей / Г.В. Стась, Л.Э. Шейнкман, Е.В. Смирнова // Известия Тульского гос. ун-та. Науки о Земле. – 2015. – № 3. – С. 31–41.
4. Морозова Т.С. О применении экспертно-статистического метода при оценке вероятности аварий и отказов

работы смесительно-зарядных машин // Горная промышленность. – 2020. – № 4. – С. 86–91.

5. Едаменко А.С. Анализ опасностей и оценка риска аварий на объектах хранения нефтепродуктов / А.С. Едаменко, А.В. Ястребинская // Технологии техноферной безопасности. – 2023. – № 3 (101). – С. 60–73.

6. Немцов В.С. Оценка риска возникновения аварии грузоподъемных сооружений при строительстве моста // Природные и техногенные риски (физико-математические и прикладные аспекты). – 2023. – № 1 (45). – С. 25–33.

7. Моделирование системы искусственного интеллекта для опережающего выявления аварийных ситуаций на жизненно важных объектах / И.Г. Боровской, Е.А. Шельмина, И.Г. Афанасьева, А.А. Матолыгин // Моделирование, оптимизация и информационные технологии. – 2022. – № 10 (3). – С. 1–16.

8. Суханов А.В. Параметрическая диагностика и оценка состояния газотурбинных энергетических установок с использованием SCADA-технологий и имитационного моделирования / А.В. Суханов, Д.А. Ахмедзянов // Вестник Уфим. гос. авиационного техн. ун-та. – 2021. – Т. 25, № 4(94). – С. 83–90.

9. Антонов О.В. Система предотвращения аварийных ситуаций производственных технологических объектов на основе ситуационной математической модели / О.В. Антонов, Е.Ф. Райкова, Р.Э. Муратов // Вестник Астрахан. гос. техн. ун-та. – Сер.: Управление, вычислительная техника и информатика. – 2022. – № 2. – С. 22–32.

10. Zubov D.V. Диагностика отказов технологического оборудования химических производств с помощью искусственного интеллекта / Д.В. Zubov, Д.А. Лебедев // Программные системы и вычислительные методы. – 2024. – № 2. – С. 30–40.

11. Digital twin for leak detection and fault diagnostics in gas pipelines: A systematic review, model development, and case study / W.A. Al-Ammari, A.K. Sleiti, M.A. Rahman, S. Rezaei-Gomari, I. Hassan, R. Hassan // Alexandria Engineering Journal. – 2025. – Vol. 123. – P. 91–111.

12. Tornyeviadzi H.M. Leakage detection in water distribution networks via 1D CNN deep autoencoder for multivariate SCADA data / H.M. Tornyeviadzi, R. Seidu // Engineering Applications of Artificial Intelligence. – 2023. – Vol. 122. – P. 106062.

13. Data-driven multi-fault detection in pipelines utilizing frequency response function and artificial neural networks / A.M.H. Hussein, B.A.R. Sharafiz, B.M. Faizal, S.K. Prajindra, A.B.A.J. Nawal // Journal of Pipeline Science and Engineering. – 2025. – Vol. 5, No. 1. – P. 100223.

14. A Data Driven Model for Predicting Tool Health Condition in High Speed Milling of Titanium Plates Using Real-Time SCADA / J. Bao, Y. Guangchao, X. Zheng, J. Zhang, X. Ji // Procedia CIRP. – 2017. – Vol. 61. – P. 317–322.

15. Phong B.D. Anomaly detection of wind turbines based on stationarity analysis of SCADA data / B.D. Phong, T. Barszcz, W.J. Staszewski // Renewable Energy. – 2024. – Vol. 232. – P. 121076.

16. RandomForestClassifier – документация scikit-learn 1.7.1 [Электронный ресурс]. – Режим доступа: <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html>, свободный (дата обращения: 04.09.2025).

17. StandardScaler – документация scikit-learn 1.7.1 [Электронный ресурс]. – Режим доступа: <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html>, свободный (дата обращения: 04.09.2025).

18. accuracy_score – документация scikit-learn 1.7.1 [Электронный ресурс]. – Режим доступа: [\[learn.org/stable/modules/generated/sklearn.metrics.accuracy_score.html\]\(https://scikit-learn.org/stable/modules/generated/sklearn.metrics.accuracy_score.html\), свободный \(дата обращения: 04.09.2025\).](https://scikit-</p></div><div data-bbox=)

Гарипов Егор Тимурович

Аспирант каф. экономической математики, информатики и статистики (ЭМИС) Томского государственного ун-та систем управления и радиоэлектроники (ТУСУР) Ленина пр-т, 40, г. Томск, Россия, 634050
Тел.: +7-952-899-15-98
Эл. почта: egor-garipov@mail.ru

Боровской Игорь Георгиевич

Науч. рук., д.ф.-м.н., проф., зав. каф. ЭМИС ТУСУРа Ленина пр-т, 40, г. Томск, Россия, 634050
ORCID: 0000-0003-2707-7203
Тел.: +7 (382-2) 90-01-85, доп. 28-19
Эл. почта: igor.g.borovskoi@tusur.ru

Кручинин Владимир Викторович

Д-р техн. наук, проф., зав. каф. технологий электронного обучения (ТЭО) ТУСУРа Ленина пр-т, 40, г. Томск, Россия, 634050
ORCID: 0000-0001-5564-2797
Тел.: +7 (382-2) 70-15-15
Эл. почта: kru@2i.tusur.ru

Поступила в редакцию: 08.07.2025.

Принята к публикации: 06.11.2025.

Garipov Y.T., Borovskoy I.G., Kruchinin V.V.

Machine learning model for predicting accidents at gas production facilities

Emergency situations lead to disruptions of workflows, regardless of how quickly these emergencies are resolved, therefore the ability to predict such situations would be extremely useful in many fields. The paper implements an approach to preprocessing data from a SCADA database using averaging and correlation analysis. A machine learning (ML) model is then trained on the preprocessed data to classify the current state of the node as an emergency precursor to simulate a real workflow. The model achieved high accuracy score and other metrics without fine tuning and hyperparameter optimization, thus confirming the possibility of using ML models for the task.

Keywords: machine learning, data preprocessing, SCADA, fault detection, technological process, classifier.

DOI: 10.21293/1818-0442-2025-28-3-53-58

References

1. Zhakhina R.U., Kostylev K.A. [SCADA systems and their main components]. *Actual scientific research in the modern world*, 2020, no. 4–1 (60), pp. 62–65 (in Russ.).

2. Nychik T.Yu. [Methodology justification of the accident risk assessment for the shipping lock]. *Vestnik gosudarstvennogo universiteta morskogo i rechnogo flota imeni admirala S.O. Makarova*, 2013, no. 1, pp. 15–21 (in Russ.).

3. Stas G.V., Shainkman L.E., Smirnova E.V. *Otsenka riska vzniknoveniya avarii pri dobyche burykh uglei* [Evaluating risk of arising accidents by underground mining brown coal]. *Izvestiya Tula State University. Earth sciences*, 2015, no. 3, pp. 31–41 (in Russ.).

4. Morozova T.S. [On implementation of expert statistical method in assessment of accident and failure probability for mixing and charging machines]. *Gornaya promyshlennost – Russian Mining Industry*, 2020, no. 4, pp. 86–91 (in Russ.).
5. Edamenko A.S., Yastrebinskaya A.V. [Hazard analysis and accident risk assessment at petroleum product storage facilities]. *Technosphere safety technologies*, no. 3(101), pp. 60–73 (in Russ.).
6. Nemtsov V.S. [Assessment of the risk of an accident of lifting structures during the construction of the bridge]. *Natural and man-made risks (physico-mathematical and applied aspects)*, 2023, no. 1(45), pp. 25–33 (in Russ.).
7. Borovskoy I.G., Shelmina E.A., Afanasyeva I.G., Matolygin A.A. [Modeling of artificial intelligence systems for early detection of emergency situations at vital facilities]. *Modeling, optimization and IT*, 2022, no. 10 (3), pp. 1–16 (in Russ.).
8. Sukhanov A.V., Akhmedzyanov D.A. *Parametricheskaya diagnostika i otsenka sostoyaniya gazoturbinnnykh energeticheskikh ustanovok s ispolzovaniem SCADA-tehnologii i imitatsionnogo modelirovaniya* [Parametric diagnostics and assessment of the state of gas turbine power plants using SCADA-technologies and imitation modelling]. *Vestnik USATU*, 2021, vol. 25, no. 4(94), pp. 83–90 (in Russ.).
9. Antonov O.V., Raykova E.F., Muratov R.E., [Emergency preventing system at production technological objects based on situational mathematical model]. *Vestnik of Astrakhan State Technical University. Series: Management, Computer Science and Informatics*, 2022, no. 2, pp. 22–32 (in Russ.).
10. Zubov D.V., Lebedev D.A. *Diagnostika otkazov tekhnologicheskogo oborudovaniya himicheskikh proizvodstv s pomoshchyu iskusstvennogo intellekta* [Diagnostics of failures of technological equipment of chemical industries using artificial intelligence]. *Software Systems and Computational Methods*, 2024, no. 2, pp. 30–40 (in Russ.).
11. Al-Ammari W.A., Sleiti A.K., Rahman M.A., Rezaei-Gomari S., Hassan I., Hassan R. Digital twin for leak detection and fault diagnostics in gas pipelines: A systematic review, model development, and case study. *Alexandria Engineering Journal*, 2025, vol. 123, pp. 91–111.
12. Tornyeviadzi H.M., Seidu R. Leakage detection in water distribution networks via 1D CNN deep autoencoder for multivariate SCADA data. *Engineering Applications of Artificial Intelligence*, 2023, vol. 122, 106062.
13. Hussein A.M.H., Sharafiz B.A.R., Faizal B.M., Prajindra S.K., Nawal A.B.A.J. Data-driven multi-fault detection in pipelines utilizing frequency response function and artificial neural networks. *Journal of Pipeline Science and Engineering*, 2025, vol. 5, no. 1, 100223.
14. Bao J., Guangchao Y., Zheng X., Zhang J., Ji X. A Data Driven Model for Predicting Tool Health Condition in High Speed Milling of Titanium Plates Using Real-Time SCADA. *Procedia CIRP*, 2017, vol. 61, pp. 317–322.
15. Phong B.D., Barszcz T., Staszewski W.J. Anomaly detection of wind turbines based on stationarity analysis of SCADA data. *Renewable Energy*, 2024, vol. 232, 121076.
16. RandomForestClassifier – scikit-learn 1.7.1 documentation [Electronic resource]. – Available at: <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html>, free (Accessed: 04.09.2025).
17. StandardScaler – scikit-learn 1.7.1 documentation [Electronic resource]. – Available at: <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html>, free (Accessed: 04.09.2025).
18. accuracy_score – scikit-learn 1.7.1 documentaion [Electronic resource]. – Available at: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.accuracy_score.html, free (Accessed: 04.09.2025).

Yegor T. Garipov

Postgraduate student, Department of Economic Mathematics, Informatics and Statistics, Tomsk State University of Control Systems and Radioelectronics (TUSUR)
40, Lenin pr., Tomsk, Russia, 634050
Phone: +7-952-899-15-98
Email: egor-garipov@mail.ru

Igor G. Borovskoy

Thesis Supervisor, Doctor of Physical and Mathematical Sciences, Professor, Head of the Department of Economic Mathematics, Informatics and Statistics, TUSUR
40, Lenin pr., Tomsk, Russia, 634050
ORCID: 0000-0003-2707-7203
Phone: +7 (382-2) 90-01-85, add. 28-19
Email: igor.g.borovskoi@tusur.ru

Vladimir V. Kruchinin

Doctor of Technical Sciences, Professor,
Head of the Department of Electronic Learning Technologies, TUSUR
40, Lenin pr., Tomsk, Russia, 634050
ORCID: 0000-0001-5564-2797
Phone.: +7 (382-2) 70-15-15
Email: kru@2i.tusur.ru

Received: 08.07.2025.

Accepted: 06.11.2025.