

## ИССЛЕДОВАНИЕ МОДЕЛЕЙ МАШИННОГО ОБУЧЕНИЯ ДЛЯ АНАЛИЗА ПРОИЗВОДСТВЕННЫХ ДАННЫХ

**Аннотация.** В работе рассматривается задача анализа производственных данных с целью выявления аномалий и закономерностей, влияющих на возникновение инцидентов. Проводится сравнительное исследование методов машинного обучения с оценкой их эффективности на реальных данных.

**Ключевые слова:** Машинное обучение, производственные данные, обнаружение аномалий, логистическая регрессия, метод опорных векторов, случайный лес.

Современные производственные системы генерируют большие объёмы данных, отражающих состояние оборудования и технологических процессов. Анализ этих данных позволяет выявлять отклонения и предотвращать возникновение инцидентов на ранних стадиях. В связи с этим возрастает интерес к методам машинного обучения, способным автоматически обнаруживать аномалии и закономерности в многомерных данных.

В работе рассматривается применение и сравнение пяти алгоритмов машинного обучения:

- логистическая регрессия,
- метод опорных векторов,
- случайный лес.

Основной целью является оценка их эффективности при выявлении отклонений в работе производственных систем.

Для анализа используется набор данных, состоящий из 12000 записей, включающий следующий набор признаков:

- температура (°C),
- вибрация (мм/с),
- давление (кПа),
- нагрузка (нормированный коэффициент от 0 до 1),
- влажность окружающей среды (%),
- Инцидент (да/нет).

Параметры охватывают основные физические процессы в оборудовании и позволяют фиксировать как нормальные состояния, так и отклонения, возникающие при нарушении режимов работы или деградации компонентов. Такой состав данных обеспечивает возможность построения алгоритмов, способных обнаруживать нестандартные ситуации, влияющие на надежность производственного процесса.

### Обучение модели на основе алгоритма логистической регрессии

Логистическая регрессия – это линейный алгоритм машинного обучения, предназначенный для решения задач бинарной классификации. Модель оценивает вероятность принадлежности объекта к одному из двух классов с помощью логистической функции, которая преобразует линейную комбинацию признаков в значение в диапазоне от 0 до 1 [1].

Алгоритм отличается простотой реализации, высокой интерпретируемостью и устойчивостью при работе с ограниченным числом признаков, что делает его удобным базовым методом для прогнозирования инцидентов, при этом логистическая регрессия плохо учитывает сложные нелинейные зависимости.



Логистическая регрессия обучается на подготовленных признаках с использованием L2-регуляризации. Это необходимо для автоматической балансировки весов классов для предотвращения переобучения.

В таблицах 1-2 представлены матрица ошибок и отчет классификации модели:

Таблица 1

Матрица ошибок

	Распознано	Нераспознано
Нормальные значения	3456	37
Аномалии	106	1

Таблица 2

Оценка выявления аномалий

	Точность	Полнота	F1-score	Кол-во записей
Нормальные значения	1.00	0.99	0.99	3493
Аномалии	0.74	0.99	0.85	107

Результаты показывают, что логистическая регрессия эффективно прогнозирует инциденты, особенно в задаче выявления аномалий. Для нормального состояния оборудования модель достигает почти идеальной точности (100%) и высокой полноты (0.99), что свидетельствует о минимальном числе ложных срабатываний. Для аномальных случаев точность составляет 74% при полноте 0.99, что означает выявление почти всех инцидентов при наличии части ложных тревог; F1-score = 0.85 отражает сбалансированное качество на редком классе. Высокая общая точность (99%) обусловлена дисбалансом данных, однако модель демонстрирует надежное обнаружение инцидентов, что важно для промышленного применения.

#### Обучение модели на основе алгоритма опорных векторов

Метод опорных векторов (SVM) – это алгоритм машинного обучения, предназначенный для решения задач бинарной классификации. Модель находит оптимальную разделяющую гиперплоскость, при которой различия между классами наиболее выражены. Благодаря возможности учитывать сложные взаимосвязи между параметрами метод способен выявлять скрытые отклонения в работе оборудования и использоваться для обнаружения потенциальных инцидентов [2].

Алгоритм отличается высокой устойчивостью к переобучению и хорошо работает в условиях несбалансированных классов при правильной настройке параметров, однако требует значительных вычислительных ресурсов при работе с большими объемами данных.

Модель SVM обучается на подготовленных признаках с использованием настроек, повышающих качество обнаружения аномалий. В качестве ядра применяется радиальная базисная функция (RBF), позволяющая учитывать нелинейные зависимости и точнее разделять нормальные и аварийные состояния. Параметр регуляризации равен 1.0 и обеспечивает баланс между точностью и устойчивостью модели без переобучения. Параметр  $\gamma$  задаётся автоматически (scale), что адаптирует модель к структуре данных без ручной настройки. Для учёта дисбаланса классов используется автоматическое взвешивание, благодаря чему модель лучше распознаёт редкие аварийные случаи.

В таблицах 3-4 представлены матрица ошибок и отчет классификации модели:



Таблица 3

Матрица ошибок

	Распознано	Нераспознано
Нормальные значения	3482	11
Аномалии	2	105

Таблица 4

Оценка выявления аномалий

	Точность	Полнота	F1-score	Кол-во записей
Нормальные значения	1.00	1.00	1.00	3493
Аномалии	0.91	0.98	0.94	107

Модель SVM демонстрирует высокую эффективность в прогнозировании инцидентов. Для нормальных состояний оборудования достигаются точность 100% и полнота 1.00, что свидетельствует о корректной классификации без ложных срабатываний. Для аномалий точность составляет 91% при полноте 0.98, что указывает на редкие пропуски инцидентов при высоком уровне их выявления. F1-score равен 0.94 и отражает сбалансированный результат на редком классе. Общая точность 100% подтверждает эффективность ядрового метода опорных векторов с балансировкой классов для задач промышленного мониторинга.

#### Обучение модели на основе алгоритма случайного леса

Ансамблевый метод случайного леса объединяет множество деревьев решений для повышения точности и устойчивости предсказаний. Каждое дерево обучается на случайной подвыборке признаков и объектов, что позволяет модели выявлять сложные нелинейные зависимости между параметрами оборудования [3].

Для подбора наилучших параметров леса используется инструмент «GridSearchCV». Он позволяет настраивать ключевые гиперпараметры случайного леса путём их перебора.

В результате обучения «GridSearchCV» подобрал следующие наилучшие параметры для модели случайного леса:

- Максимальная глубина – неограниченная.
- Количество признаков, используемых на каждом разбиении – квадратный корень от общего числа признаков.
- Количество деревьев – 200.

В таблицах 5-6 представлены матрица ошибок и отчет классификации модели:

Таблица 5

Матрица ошибок

	Распознано	Нераспознано
Нормальные значения	3491	2
Аномалии	105	2

Таблица 6

Оценка выявления аномалий

	Точность	Полнота	F1-score	Кол-во записей
Нормальные значения	1.00	1.00	1.00	3493
Аномалии	0.98	0.98	0.98	107



Модель случайного леса демонстрирует почти идеальное качество распознавания состояний оборудования. Для нормальных случаев достигаются точность 100% и полнота 1.00, что означает отсутствие ложных срабатываний и корректную классификацию нормальных точек. Для аномалий точность составляет 98% при полноте 0.98, что отражает высокую способность модели выявлять инциденты при минимальном числе пропусков. F1-score равен 0.98 и подтверждает стабильный баланс между точностью и полнотой, а общая точность 100% показывает высокую эффективность ансамблевого подхода.

### **Заключение**

Сравнение моделей показывает, что логистическая регрессия хорошо выявляет инциденты, но уступает в точности для редких событий, метод опорных векторов демонстрирует высокую точность и полноту за счёт использования нелинейного ядра, но требует более чувствительной настройки и хуже масштабируется на большие данные, а случайный лес объединяет сильные стороны подходов и обеспечивает наиболее стабильные результаты.

В рамках данной задачи именно случайный лес является наиболее предпочтительной моделью, так как он лучше всего балансирует точность, полноту и устойчивость предсказаний при анализе производственных данных.

### *Список литературы:*

1. Хасти Т., Тибширани Р., Фридман Дж. Элементы статистического обучения. М.: Вильямс, 2010. 764 с.
2. Метод опорных векторов (Support Vector Machines – SVM) // Scikit-learn URL: <http://scikit-learn.ru/stable/modules/svm.html> (дата обращения: 21.03.2026).
3. Random Forests // Springer Nature Link URL: <https://link.springer.com/article/10.1023/A:1010933404324> (дата обращения: 21.03.2026).

