

Применение алгоритмов машинного обучения для прогнозирования отказов и адаптивного управления производственными системами

А.А. Скворцов, М.С.Анурьева, А.Н.

Солодовников

Тамбовский государственный университет имени Г.Р. Державина

Аннотация: Статья посвящена исследованию применения методов машинного обучения для прогнозирования отказов промышленного оборудования. Проведен обзор современных подходов, таких как метод случайного леса, метод опорных векторов и метод экстремального градиентного бустинга, с акцентом на их точность, устойчивость и применимость к инженерным задачам. На основе анализа реальных данных (температура, давление, вибрация, влажность) обучены и сравнены модели, из которых метод экстремального градиентного бустинга показал наилучшие результаты. Выявлены ключевые параметры, влияющие на поломки, и предложена система рекомендаций, основанная на статистическом анализе и предсказательной модели. Разработанное решение позволяет своевременно выявлять риски отказов и оптимизировать техническое обслуживание.

Ключевые слова: машинное обучение, предиктивное моделирование, управление оборудованием, прогнозирование отказов, анализ данных.

Современное промышленное производство активно использует автономные системы управления, от надежности которых зависит эффективность работы оборудования. Одной из ключевых задач становится прогнозирование отказов для предотвращения простоев и снижения затрат. С ростом объемов данных и сложности процессов традиционные подходы теряют актуальность, уступая место методам машинного обучения.

В рамках исследования были проанализированы публикации, в которых рассматриваются различные подходы к решению задачи прогнозирования отказов. Это позволило выделить наиболее эффективные алгоритмы, применимые к промышленным данным, и обосновать выбор конкретных моделей машинного обучения для дальнейшего практического использования.

В. В. Коржавых показал, что методы машинного обучения, включая случайный лес, позволяют точно прогнозировать энергопотребление и

выявлять аномалии, что применимо и к задачам предсказания отказов в автономных системах [1]. А. С. Горячев представил обзор алгоритмов (случайный лес, градиентный бустинг, нейронные сети и методы временных рядов), подчеркивая их пользу для диагностики и планирования обслуживания оборудования на основе данных с датчиков [2].

А. В. Чернухин и соавт. разработали предиктивную модель на основе мягкого голосования (нейросеть, логистическая регрессия, метод опорных векторов), подтвердив ее эффективность на данных с эксгаустеров. Модель адаптируема под разные типы оборудования [3]. В. А. Храмовских и соавт. создали Excel-инструмент с дашбордом для анализа отказов горной техники. Предложен подход, который может быть дополнен нейросетями для повышения точности прогнозов [4]. М. А. Золотухина и С. А. Золотухин применили глубокое обучение и кластеризацию для диагностики роторного оборудования, достигнув точности 95%, при этом разработанная система обеспечивает онлайн-распознавание неисправностей [5]. Н. И. Шаханов и соавт. предложили модель на основе метода случайный лес для условий с малым числом поломок. Алгоритм обучается на нормальных данных, выявляя аномалии по отклонениям прогноза, что важно для раннего обнаружения отказов [6].

Анализ представленных исследований показал, что методы машинного обучения эффективно применяются для диагностики и прогнозирования отказов оборудования в различных промышленных условиях. Наибольшую устойчивость и точность демонстрируют алгоритмы, способные работать с ограниченным количеством отказов, учитывать нелинейные зависимости и обрабатывать большие объемы сенсорных данных.

На основе проведенных тестов был получен набор данных, содержащий статистику по температуре, давлению, вибрации, влажности, вышло ли из строя оборудование. В данном датасете температура измеряется

в °С, давление – Бар (1 Бар = 100 кПа), вибрация – нормализованные единицы, влажность – % влажности. Поломка = 0, значит нет поломки, 1 – есть.

В ходе анализа попробуем определить, какой из этих факторов оказывает наибольшее влияние на возникновение поломок оборудования на производстве. После этого обучим искусственный интеллект для обнаружения поломок, затем предоставим рекомендации по их устранению.

Метод случайного леса активно используется в самых разных прикладных задачах – от государственного управления до технического обслуживания и кибербезопасности. Так, в работе С. А. Корчагина и соавт. случайный лес применялся для прогнозирования исполнения государственных контрактов, демонстрируя высокую точность благодаря учету нелинейных зависимостей и устойчивости к переобучению [7]. С. С. Панасенко и др. включили случайный лес в сравнение методов для прогноза индекса ровности дорожного покрытия, где он показал приемлемую точность и оказался полезен при составлении планов технического обслуживания [8]. В исследовании Д. М. Дудкина и соавт. метод использовался как базовая модель классификации вредоносного программного обеспечения по признакам, основанным на взвешенной частоте терминов, и дал неплохие результаты, уступив только нейросетевым языковым моделям. [9]. Наконец, в работе В. З. Ковалева и др. случайный лес входил в группу нейросетевых методов, применявшихся для краткосрочного прогноза потребления электроэнергии, где также показал высокую точность [10].

Метод опорных векторов применяется для диагностики и классификации технического состояния сложных инженерных систем. В работе С. У. Увайсова и др. данный метод был использован для тепловой диагностики печатных узлов радиоэлектронных устройств, где позволил

достоверно выявлять конструктивные дефекты на основе температурных карт, превосходя по точности ранее применявшиеся нейросетевые методы [11]. В другой работе метод продемонстрировал высокую эффективность в задачах классификации, обнаружения аномалий и диагностики деградации радиотехнических систем, обеспечивая адаптацию к изменяющимся условиям эксплуатации и высокую точность в реальном времени [12].

Алгоритм экстремального градиентного бустинга также показал высокую эффективность в инженерных задачах диагностики и прогноза. В работе Y. Wang и др. он использован для оценки скорости проходки тоннелепроходческого комплекса, где экстремальный градиентный бустинг интегрирован в онлайн-платформу с точностью $R^2 = 0.865$ [13]. Ch. Lee и Ya. Wind применили экстремальный градиентный бустинг с глубокой сверточной нейросетью, оптимизированной методом роя частиц с элитой, для классификации отказов в ветроустановках. Для балансировки данных автоматизированной системы управления и сбора данных использовались методы синтетического увеличения выборки и отбора ближайших противоположных примеров, а для визуализации – метод стохастического соседнего вложения с t -распределением. Достигнута точность 99.2% [14]. Другое исследование подтвердило преимущество связки одномерной сверточной нейронной сети и алгоритма экстремального градиентного бустинга над методами опорных векторов и случайного леса при анализе данных автоматизированной системы управления и сбора данных [15].

Все три метода – случайный лес, метод опорных векторов и алгоритм экстремального градиентного бустинга – эффективны для инженерной диагностики. Метод случайного леса удобен для устойчивых предсказаний на разнородных данных, метод опорных векторов – для точной классификации при ограниченном обучении, экстремальный градиентный

бустинг – наиболее мощный на больших и несбалансированных выборках, особенно в связке с нейросетями.

В качестве алгоритмов машинного обучения принято решение использовать три популярных метода: случайный лес, метод опорных векторов и экстремальный градиентный бустинг. Для системы рекомендаций дополнительно анализируются распределение параметров в данных. Эффективность модели оценивается с помощью метрик точности, площади под кривой операционной характеристики приемника (Receiver Operating Characteristic Area Under the Curve – ROC-AUC) и F1-метрики (гармоническое среднее точности и полноты).

1. Точность – простая метрика для оценки качества классификационных моделей. Она показывает долю правильно предсказанных меток классов относительно общего количества предсказаний.

$$\text{Точность} = \frac{\text{Количество_верных_предсказаний}}{\text{Общее_количество_предсказаний}}$$

2. ROC-AUC – это метрика, которая показывает, насколько хорошо модель различает два класса в задаче бинарной классификации. График операционной характеристики приемника строится на основе матрицы ошибок (путаницы) (рис. 1). На оси X расположена доля ложных срабатываний (False Positive Rate – FPR). Формула FPR:

$$FPR = \frac{FP}{(FP + TN)},$$

где *FP* (ложно положительное срабатывание – *False Positive*) – модель ошибочно предсказала наличие поломки, хотя ее нет; *TN* (истинно отрицательное срабатывание – *True Negative*) – модель правильно предсказала отсутствие поломки.

На оси Y расположена доля верных срабатываний (True Positive Rate – TPR). Формула TPR:

$$TPR = \frac{TP}{(TP + FN)},$$

где TP (истинно положительное срабатывание – *True Positive*) – модель правильно предсказала отсутствие поломки; FN (ложно отрицательное срабатывание – *False Negative*) – модель ложно предсказала наличие поломки, на самом деле ее нет.

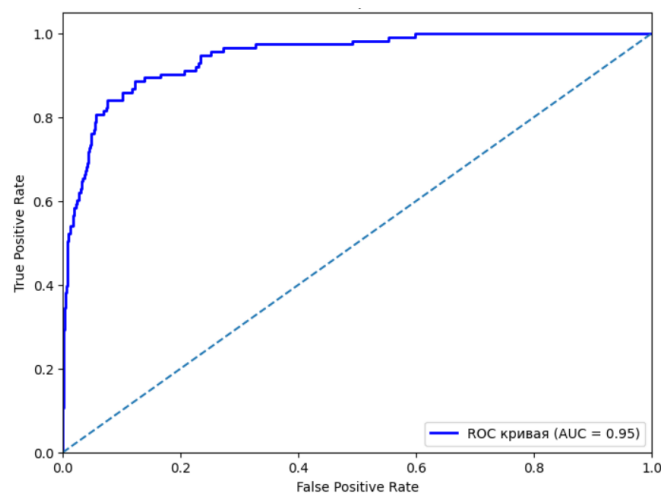


Рис. 1. Кривая ошибок: зависимость доли верных срабатываний от ложных срабатываний при площади под кривой, равной 0,95

Кривая отображает зависимость доли истинно положительных срабатываний от доли ложных срабатываний разных порогах классификации. Площадь под кривой, равная 0,95 свидетельствует о высокой способности модели различать классы – поломку и нормальное состояние оборудования.

ROC-AUC вычисляется как площадь под кривой ошибок первого и второго рода (Receiver Operating Characteristic, ROC-кривая), которая отображает зависимость TPR от FPR. В аналитическом виде ROC-AUC представляется как определенный интеграл:

$$ROC - AUC = \int_0^1 TPR(FPR)d(FPR)$$

Чем ближе значение ROC-AUC к 1, тем точнее модель предсказывает класс объекта. Значение около 0,5 говорит о случайных предсказаниях. Если ROC-AUC меньше 0,5, это означает, что модель систематически путает классы, предсказывая противоположное.

3. F1-метрика – метрика, отражающая баланс между точностью и полнотой, особенно полезна при несбалансированных классах. Высокое значение F1-метрики (ближе к 1) указывает на хорошее качество модели, низкое – на слабую способность корректно выявлять положительные случаи.

Всего собрано ~7000 данных с разных оборудований в разный момент их работы. Все данные сохранили в .csv файл. С помощью библиотеки Pandas и Matplotlib и языка программирования Python 3.12 были проанализированы полученные данные.

На рис. 2 представлено процентное распределение четырех параметров оборудования: температуры, влажности, вибрации и давления в двух разрезах: по типам оборудования (левый столбец) и по состоянию оборудования: исправное/неисправное (правый столбец).

Левые гистограммы показывают, как распределяются значения параметров среди различных типов оборудования (турбина, компрессор, насос). Правые гистограммы отражают различия в распределении тех же параметров в зависимости от наличия поломки.

Анализ графиков показывает, что при поломках возрастают значения температуры, давления и особенно вибрации, что подтверждает их важность для предсказания. Выделены критические диапазоны, при которых чаще фиксируются поломки: температура: ниже 40 °С и выше 100 °С; влажность: менее 25% и более 75%; вибрация: около 0 и выше 3 нормализованных единиц; давление: ниже 15 и выше 55 бар.



Эти значения будут использоваться при формировании рекомендаций для предупреждения возможных неисправностей оборудования. Выводы согласуются с данными таблицы №1.

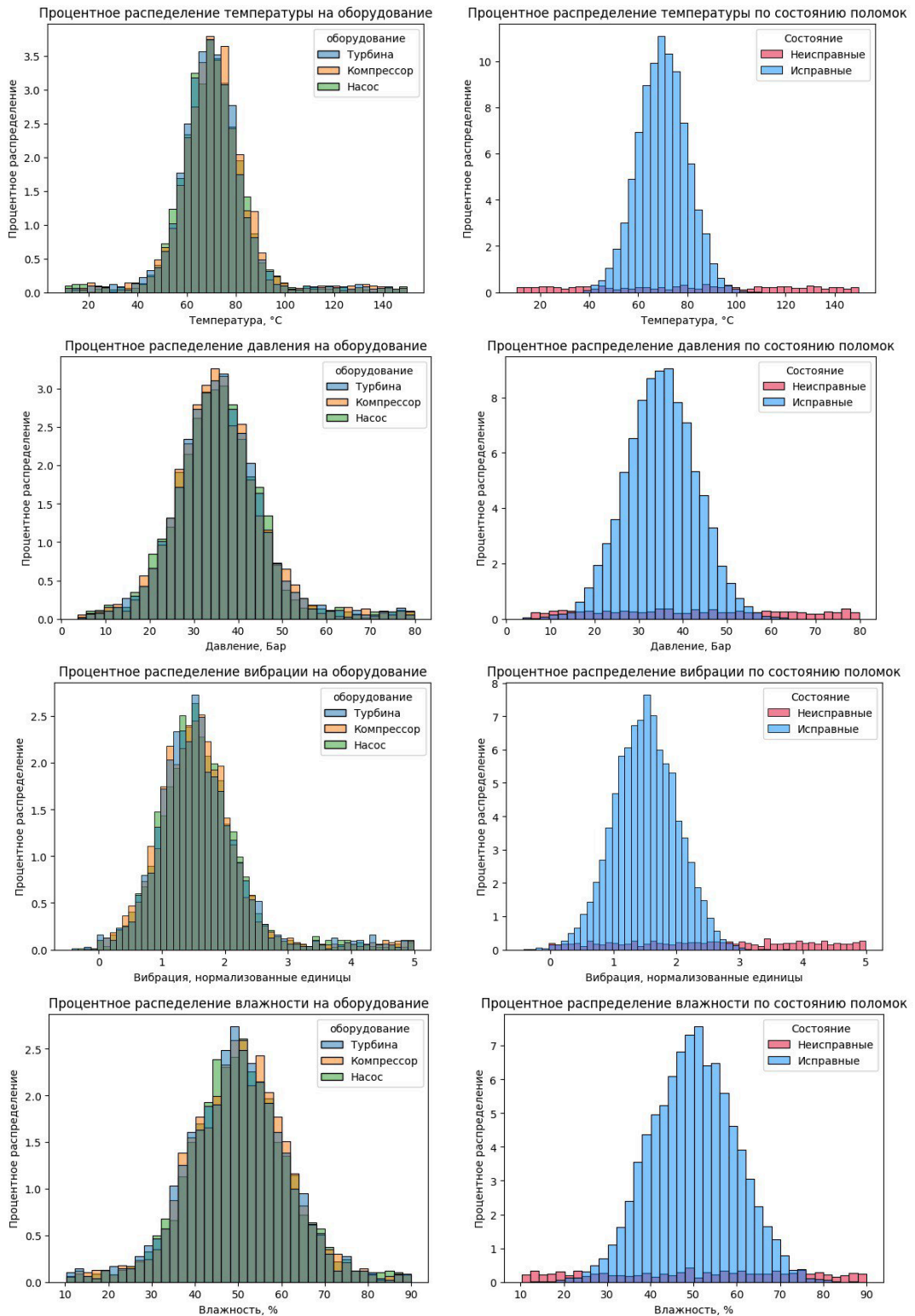


Рис. 2. Процентное распределение параметров по видам оборудования и состоянию

В таблице №1 приведены средние значения и стандартные отклонения параметров оборудования – давления, вибрации, влажности и температуры – как для всей выборки, так и отдельно для случаев с поломкой и без. Наиболее чувствительными к поломкам оказались вибрация, температура и давление, тогда как влажность почти не меняется. Эти данные используются для построения модели и рекомендаций по правилу трех сигм.

Таблица №1

Математическое ожидание и стандартное отклонение величин

	Математическое ожидание	Стандартное отклонение
Давление, общее	35.738	10.381
Давление, нет поломки	35.038	8.010
Давление, есть поломка	42.038	21.373
Вибрация, общее	1.611	0.728
Вибрация, нет поломки	1.507	0.502
Вибрация, есть поломка	2.554	1.430
Влажность, общее	50.016	11.841
Влажность, нет поломки	49.966	9.865
Влажность, есть поломка	50.466	22.952
Температура, общее	70.922	16.200
Температура, нет поломки	69.927	10.012
Температура, есть поломка	79.880	40.440

Анализ корреляционной матрицы (рис. 3) позволяет определить степень линейной зависимости между числовыми переменными. Значения коэффициента корреляции находятся в диапазоне от -1 до 1: 1 – прямая линейная связь (обе переменные растут), -1 – обратная связь (одна растет – другая падает), 0 – отсутствует линейная зависимость.

Анализ распределений и корреляций показал, что параметры оборудования в целом схожи, и наиболее значимым фактором поломок является вибрация. Температура и давление влияют умеренно, а влажность – минимально.

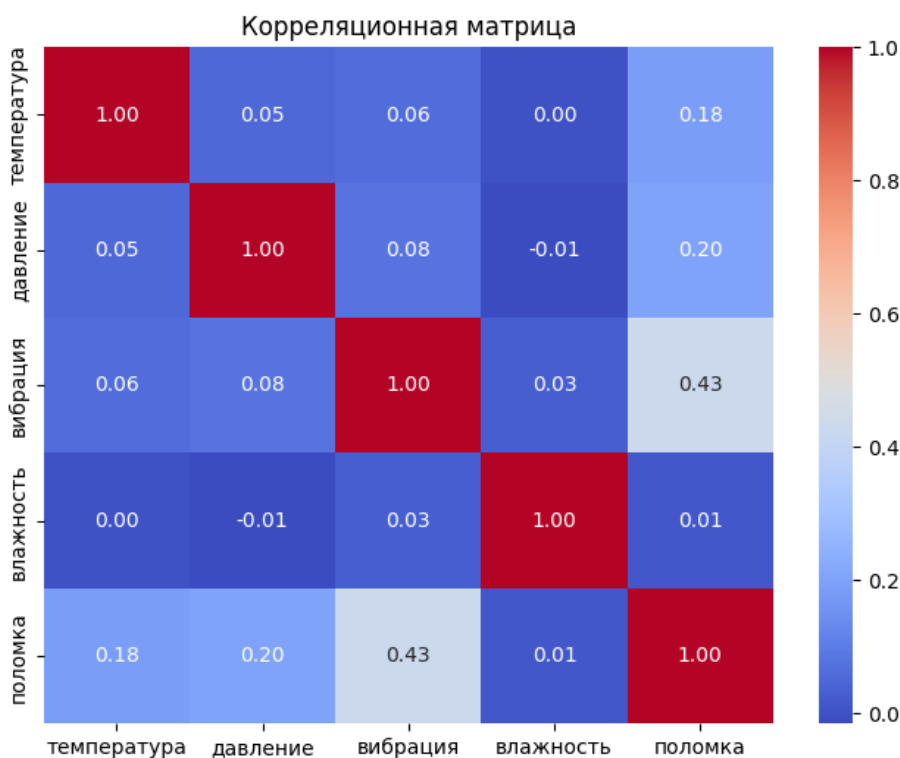


Рис. 3. Корреляционная матрица признаков датасета

Для оценки качества алгоритмов машинного обучения были построены и обучены три модели, результаты которых впоследствии сравниваются по ряду метрик. Реализация выполнена с использованием языка программирования Python 3.12 и следующих библиотек: pandas, numpy, scikit-learn, xgboost, seaborn, matplotlib и joblib. Исходный набор данных был предварительно разделён на обучающую и тестовую выборки в пропорции 80/20, что позволило обеспечить объективную проверку обобщающей способности моделей.

На рис. 4 показано, как изменяются метрики точности, ROC-AUC и F1-метрики в зависимости от количества деревьев в модели случайного леса. Анализ графиков показывает, что на начальном этапе увеличения числа деревьев наблюдается уверенный рост всех показателей, что свидетельствует об улучшении обобщающей способности модели. Однако, начиная примерно со 100 деревьев, темп роста метрик существенно снижается, а после 120 деревьев значения стабилизируются, демонстрируя эффект насыщения.

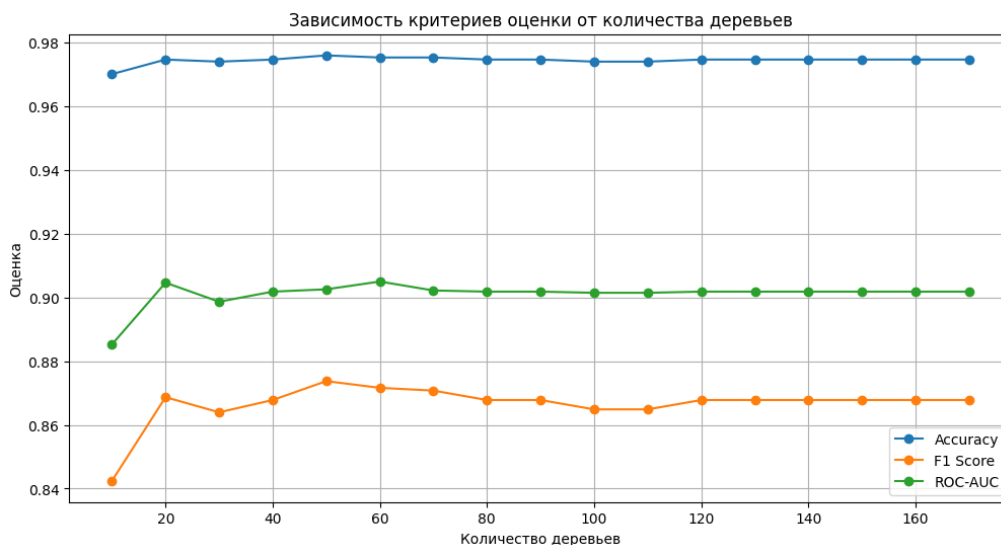


Рис. 4. Зависимость метрик качества от количества деревьев в случайном лесе

На рис. 5 представлено сравнение различных ядер, используемых в методе опорных векторов. Тестирование показало, что наилучшие значения метрик (точности, ROC-AUC и F1-метрики) достигаются при использовании радиальной базисной функции, в то время как линейное и сигмоидное ядра показали худшие результаты.

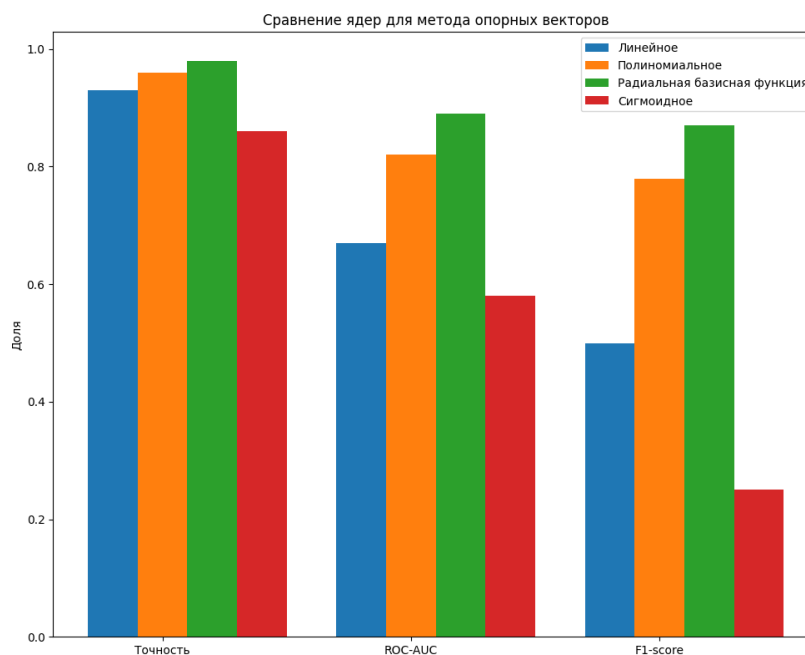


Рис. 5. Сравнение эффективности различных ядер в методе опорных векторов

На рис. 6 показано сравнение трех моделей – по метрикам точности, ROC-AUC и F1-метрики. Модель экстремального градиентного бустинга продемонстрировала наилучшие результаты по всем показателям, что указывает на ее высокую способность к точному и сбалансированному прогнозированию поломок.

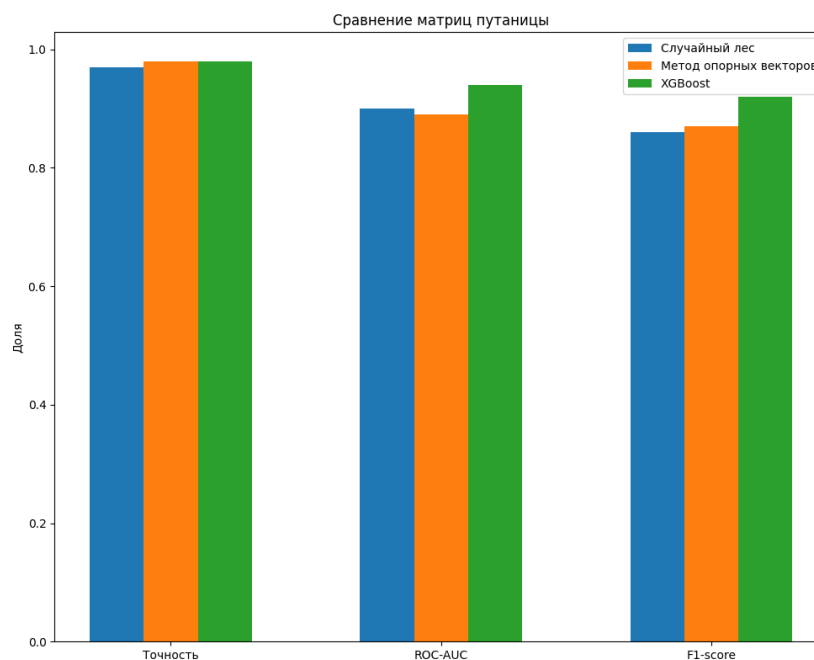


Рис. 6. Сравнение критериев эффективности трех моделей

Разработана предиктивная система рекомендаций, позволяющая не только оценить вероятность поломки оборудования, но и заблаговременно предложить корректирующие действия на основе текущих параметров. Система сочетает в себе модель машинного обучения экстремального градиентного бустинга и статистический подход на основе правила трех сигм.

Модель экстремального градиентного бустинга прогнозирует вероятность отказа на основе обученной зависимости между параметрами оборудования и историей поломок. Далее, если риск превышает допустимый порог, система активирует модуль интерпретации и рекомендаций.

Для анализа используется правило трех сигм, основанное на отклонении текущих значений параметров от их среднего по исправным случаям. Согласно нормальному распределению:

- ~68 % значений находятся в пределах $\pm 1\delta$,
- ~95 % – в пределах $\pm 2\delta$,
- ~99.7 % – в пределах $\pm 3\delta$,
- значения за пределами $\pm 3\delta$ считаются потенциально критичными.

На основе степени отклонения система классифицирует тревожность по четырем уровням: низкий, средний, высокий и критический, и выдает рекомендации по корректировке параметров.

Для демонстрации возможностей предиктивной системы рассмотрен пример с текущими значениями: температура – 90 °С, давление – 38 бар, вибрация – 2.6 ед., влажность – 45 %. Модель экстремального градиентного бустинга оценила риск поломки как 17.49 %.

Анализ по правилу трех сигм дал следующие рекомендации:

- температура – средний уровень тревоги, рекомендуется снизить до диапазона [54.72, 87.12];
- давление – низкий уровень тревоги, значение в норме;
- вибрация – средний уровень тревоги, рекомендуется снизить до диапазона [0.88, 2.34];
- влажность – низкий уровень тревоги, значение допустимо.

Проведенное исследование показало высокую эффективность применения методов машинного обучения для прогнозирования отказов промышленного оборудования. Среди протестированных моделей наилучшие результаты продемонстрировал алгоритм экстремального градиентного бустинга, обеспечивая высокие значения точности, ROC-AUC и F1-метрики. Наиболее значимыми параметрами для предсказания поломок стали вибрация, температура и давление, в то время как влияние влажности

оказалось минимальным. Разработанная предиктивная система не только точно определяет риск отказа, но и формирует интерпретируемые рекомендации по корректировке технических параметров, что способствует снижению простоев и повышению надежности оборудования.

Литература

1. Коржавых В. В. Сравнение эффективности алгоритмов машинного обучения на примере прогнозирования среднемесячного потребления электроэнергии интервальных приборов учета потребителей // Научный результат. Информационные технологии. 2024. Т. 9. №. 1. С. 58-72.

2. Горячев А. С. Обзор алгоритмов машинного обучения в задачах предиктивного анализа работы технологического оборудования // Инновационная наука. 2024. №. 10-1. С. 14-19.

3. Чернухин А. В., Богданова Е. А., Савицкая Т. В. Построение модели предиктивной аналитики неисправностей промышленного оборудования // Программные продукты и системы. 2024. Т. 37. №. 2. С. 254-261.

4. Храмовских В. А., Шевченко А. Н., Непомнящих К. А. Адаптивный интеллектуальный анализ данных как инструмент для прогнозирования ресурса узлов горных машин и оборудования // Науки о Земле и недропользование. 2023. Т. 46. №. 2 (83). С. 212-225.

5. Золотухина М. А., Золотухин С. А., Ивченко В. Д. Интеллектуальная система регулирования технической диагностики роторного оборудования с помощью машинного обучения // International journal of professional science. 2021. №. 6. С. 68-73.

6. Шаханов Н. И., Варфоломеев И.А., Ершов Е.В., Юдина О.В. Прогнозирование отказов оборудования в условиях малого количества поломок // Вестник Череповецкого государственного университета. 2016. №. 6 (75). С. 36-41.

7. Корчагин С. А., Рубцов Д. Ю., Беспалова Н. В., Сердечный Д. В. Применение моделей машинного обучения для прогнозирования исполнения государственных контрактов // Инженерный вестник Дона, 2024, № 9. URL: ivdon.ru/ru/magazine/archive/n9y2024/9507.

8. Панасенко С. С., Старков К. Н., Скоробогатченко Д. А. Разработка автоматизированной системы планирования работ по обслуживанию покрытия автомобильных дорог // Инженерный вестник Дона, 2024, № 6. URL: ivdon.ru/ru/magazine/archive/n6y2024/9295.

9. Дудкин Д. М., Кузнецов М. А., Авдосев Н. Г., Шабаловский В.А., Егунов В.А. Применение языковых нейросетевых моделей для обнаружения вредоносного программного обеспечения // Инженерный вестник Дона, 2024, № 7. URL: ivdon.ru/ru/magazine/archive/n7y2024/9332.

10. Ковалев В. З., Швецов С. Ю., Архипова О. В. Анализ методов прогнозирования потребления электрической энергии и мощности // Инженерный вестник Дона, 2023, № 4. URL: ivdon.ru/ru/magazine/archive/n4y2023/8324.

11. Увайсов С. У., Черноверская В. В., Куан Д. А., Туан Н. В., Тхань З. В. Метод опорных векторов в задаче тепловой диагностики конструктивных дефектов радиоэлектронных устройств // Вестник Концерна ВКО «Алмаз – Антей». 2022. № 1. – С. 58-70.

12. Микюляев Т. С. Метод опорных векторов (SVM) в идентификации технического состояния радиотехнической системы // Modern Science. 2023. № 11-2. С. 109-113.

13. Wang Y., Jiang P., Guo X., Xuechi G. An extreme gradient boosting technique to estimate TBM penetration rate and prediction platform // Bulletin of Engineering Geology and the Environment. 2022. Vol. 81, No. 1. P. 1-9.

14. Lee Ch. Ya., Maceren E. D. C. Wind energy system fault classification using deep CNN and improved PSO-tuned extreme gradient boosting // IET Renewable Power Generation. 2024. Vol. 18. No. 14. P. 2496-2511.

15. Chakraborty A., Tsokos Ch. P. An AI-driven Predictive Model for Pancreatic Cancer Patients Using Extreme Gradient Boosting // Journal of Statistical Theory and Applications. 2023. Vol. 22, No. 4. P. 262-282

References

1. Korzhavyh V. V. Nauchnyj rezul'tat. Informacionnye tehnologii. 2024. T. 9. №. 1. pp. 58-72.

2. Gorjachev A. S. Innovacionnaja nauka. 2024. №. 10-1. pp. 14-19.

3. Chernuhin A. V., Bogdanova E. A., Savickaja T. V. Programmnye produkty i sistemy. 2024. T. 37. №. 2. pp. 254-261.

4. Hramovskih V. A., Shevchenko A. N., Nepomnjashhih K. A. Nauki o Zemle i nedropol'zovanie. 2023. T. 46. №. 2 (83). pp. 212-225.

5. Zolotuhina M. A., Zolotuhin S. A., Ivchenko V. D. International journal of professional science. 2021. №. 6. pp. 68-73.

6. Shahanov N. I., Varfolomeev I.A., Ershov E.V., Judina O.V. Vestnik Cherepoveckogo gosudarstvennogo universiteta. 2016. №. 6 (75). pp. 36-41.

7. Korchagin S. A., Rubcov D. Ju., Bepalova N. V., Serdechnyj D. V. Inzhenernyj vestnik Dona, 2024, № 9. URL: ivdon.ru/ru/magazine/archive/n9y2024/9507.

8. Panasenko S. S, Starkov K. N., Skorobogatchenko D. A. Inzenernyj vestnik Dona, 2024, № 6. URL: ivdon.ru/ru/magazine/archive/n6y2024/9295.

9. Dudkin D. M., Kuznecov M. A., Avdosev N. G., Shabalovskij V.A., Egunov V.A. Inzhenernyj vestnik Dona, 2024, № 7. URL: ivdon.ru/ru/magazine/archive/n7y2024/9332.



10. Kovalev V. Z., Shvecov S. Ju., Arhipova O. V. Inzhenernyj vestnik Dona, 2023, № 4. URL: ivdon.ru/ru/magazine/archive/n4y2023/8324.
11. Uvajsov S. U., Chernoverskaja V. V., Kuan D. A., Tuan N. V., Than' Z. V. Vestnik Koncerna VKO «Almaz – Antej». 2022. № 1. pp. 58-70.
12. Mikjuljaev T. S. Modern Science 2023. № 11-2. pp. 109-113.
13. Wang Y., Jiang P., Guo X., Xuechi G. Bulletin of Engineering Geology and the Environment. 2022. Vol. 81, No. 1. pp. 1-9.
14. Lee Ch. Ya., Maceren E. D. C. IET Renewable Power Generation. 2024. Vol. 18. No. 14. pp. 2496-2511.
15. Chakraborty A., Tsokos Ch. P. Journal of Statistical Theory and Applications. 2023. Vol. 22, No. 4. pp. 262-282

Дата поступления: 24.02.2025

Дата публикации: 25.04 2025