

Е.С. Подоплелова**ПРОГНОЗИРОВАНИЕ ОТКАЗОВ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ
ФАКТОРНОГО АНАЛИЗА**

Рассматривается применение метода оценки рисков, основанного на объединении методологии FMEA (failure mode and effect analysis) – анализ рисков и последствий отказа и методов многокритериального принятия решений MCDM (Multiple Criteria Decision Making). Такой подход позволяет учитывать как экспертные знания, так и исторические данные о работе оборудования. Методы MCDM обрабатывают оценку более гибко в сравнении со стандартным способом расчета приоритетного числа риска (ПЧР), что помогает качественнее оценить риски по трем критериям: вероятность возникновения, сложность обнаружения и тяжесть последствий. Один из критериев возможно получить не только через оценку экспертом, но и на основе данных, фиксирующих работу оборудования. На примере синтетических данных из открытого доступа о режимах работы производственного оборудования был опробован данный подход. Задача заключалась в прогнозировании как самого отказа, так и его вида, а также выявлении факторов, сильнее всех оказывающих влияние на отказ. Для этого проводилась преобработка данных, в ходе которой потребовалось устранить дисбаланс классов. Существует несколько подходов к решению этой проблемы, направленные на сокращение преобладающего класса, либо генерацию экземпляров слабо представленных классов. В этом примере использовалось сокращение количества записей не имеющих ошибок случайным образом. Далее, в качестве алгоритмов классификации сравнивались AdaBoost, Random Forest и LinearSVC. Так как требовалась многоклассовая классификация, было решено использовать стратегию «one-vs-the-rest» (один против всех). В итоге удалось добиться точности прогнозирования по F-мере в 86% алгоритмами AdaBoost и Random Forest. LinearSVC оказался неэффективным. Таким образом, полученная модель прогнозирования распознает разные виды ошибок, но существует перспектива к улучшению, для чего требуется более объемная выборка, включающая больше примеров с разными видами отказа. Исходя из этого, такой подход как альтернатива экспертной оценки является перспективным, улучшая объективность, а также давая возможность предвидеть риски и не допустить реального отказа или инцидента, связанного с риском.

Прогнозирование; многоклассовая классификация; машинное обучение; факторный анализ; оценка рисков.

E.S. Podoplelova**FAILURE PREDICTION USING FACTOR ANALYSIS METHODS**

This article discusses the application of a risk assessment method based on the combination of the FMEA (failure mode and effect analysis) methodology and the MCDM (Multiple Criteria Decision Making) methods. This approach allows taking into account both expert knowledge and historical data on the operation of the equipment. MCDM methods process the assessment more flexibly in comparison with the standard method of calculating the priority number of risks (PRN), which helps to better assess the risks by three criteria: the probability of occurrence, the complexity of detection and the severity of the consequences. One of the criteria can be obtained not only through an expert assessment, but also on the basis of data recording the operation of the equipment. This approach was tested using the example of synthetic open-source data on the operating modes of production equipment. The task was to predict both the failure itself and its type, as well as to identify the factors that have the greatest impact on the failure. For this purpose, data preprocessing was carried out, during which it was necessary to eliminate the imbalance of classes. There are several approaches to solving this problem, aimed at reducing the dominant class or generating instances of poorly represented classes. In this example, random reduction of the number of records without errors was used. Then, AdaBoost, Random Forest and LinearSVC were compared as classification algorithms. Since multi-class classification was required, it was decided to use the one-vs-the-rest strategy. As a result, it was possible to achieve 86% forecasting accuracy by F-measure using the AdaBoost and Random Forest algorithms. LinearSVC turned out to be ineffective. Thus, the resulting forecasting model recognizes different types of errors, but there is room for improvement, which requires a larger sample, including more examples with different types of failure. Based on this, this approach as an alternative to expert assessment is promising, improving objectivity, and also making it possible to foresee risks and prevent a real failure or risk-related incident.

Forecasting; multi-class classification; machine learning; factor analysis; risk assessment.

Введение. Оценка рисков является неотъемлемой частью управления предприятием, особенно важна и разнообразна оценка на производстве ввиду большого количества оборудования. Всегда существует риск производственных травм, отказов, поломок дорогостоящего оборудования, которые могут привести к более серьезным последствиям. Для повышения безопасности на предприятиях разрабатываются меры по предупреждению негативных событий. Существуют разные подходы и методы к риск менеджменту, один из них – анализ рисков и последствий отказов (FMEA) [1]. Подробно данный подход описывается в ГОСТ Р 51814.2–2001 [2], а идея его модификации путем замены экспертной оценки по критерию «вероятность возникновения» исследована в работе [3].

В данном исследовании рассмотрим применение метода оценки рисков при наличии исторических данных с целью спрогнозировать не только отказ производственного оборудования, но и его вид в зависимости от рабочих параметров.

Анализ исследований. Как говорилось выше, подходы к оценке рисков существуют разные. Например, в работе [4] говорится о системе управления рисками усталости (FRM), которая представляет собой набор методов управления для выявления и управления рисками безопасности, связанных с усталостью. Этот подход учитывает время сна и длительность рабочей смены. Нацелен на постоянную оценку риска и мониторинг состояния рабочих. В этом исследовании рассматривается также эффективность FRM, а также барьеры и факторы в реализации FRM.

Работа [5] посвящена улучшению методики FMEA, где авторы выделяют следующие недостатки:

- ◆ сложность выражения и получения оценок;
- ◆ неточность в агрегировании оценок;
- ◆ и отсутствие взаимосвязей между факторами риска.

Для устранения этих недостатков авторами работы применяется теория нечетких множеств изображений (PFS), что позволяет экспертам выражать оценки более эффективно и получать высокую точность. Для улучшения ими были разработаны следующие шаги:

- ◆ для упрощения процесса экспертной оценки создается гибкая система приобретения знаний (FKAF), позволяющая экспертам выражать нечеткую информацию с помощью различных форм нечетких оценок;
- ◆ разрабатывается метод нечеткого преобразования изображений (PFC) для стандартизации нечетких значений для изображения нечетких чисел (PFN);
- ◆ для повышения точности агрегирования оценок в неопределенных условиях предлагается метод нечеткого доказательного рассуждения (PFER), который расширяет существующие методы нечеткого доказательного рассуждения (FER);
- ◆ для описания параллельных и причинно-следственных связей между факторами риска создаются четыре альтернативные модели с использованием нечетких сетей Петри (PFPN). Ранжирование приоритетов рисков определяется путем умозаключений.

Работа [6] также использует в своей основе FMEA, применяя методы глубокого обучения для интеллектуального анализа данных различных аспектов предметной области. На примере прогнозирования ремонта самолета рассматривается четыре области: физический мир, получение данных, кибер-мир и поддержка принятия решений. Авторы пишут об использовании глубокого обучения при наличии достаточного количества записей о работе оборудования, его сбоях и ремонтах, результаты которого можно использовать для прогнозирования возникновения.

Помимо описанных выше, работы [7–9] также описывают анализ рисков по методологии FMEA, используя их при разработке систем поддержки принятия решений в различных сферах. В работах [10, 11] описываются методы прогнозирования рисков в медицинских задачах.

В своем исследовании я использую идею прогнозирования отказа, но на примере синтетических данных о работе производственного оборудования.

Описание задачи и данных. Набор данных состоит из 10 000 строк, имеющих 14 признаков, 5 из которых – вид отказа:

1. UID: уникальный идентификатор в диапазоне от 1 до 10000.
2. Идентификатор продукта.
3. Тип продукта [L, M или H].
4. Температура воздуха [K].
5. Температура процесса [K].
6. Скорость вращения [об/мин].
7. Крутящий момент [Нм].
8. Износ инструмента [мин].
9. Целевая переменная «отказ машины» по любому из режимов.

Отказ машины включает 5 видов ошибок оборудования:

1. Отказ из-за износа комплектующих (TWF).
2. Нарушение отвода тепла (HDF).
3. Сбой мощности (PWF).
4. Перенапряжение (OSF).
5. Случайные ошибки (RNF).

В работе [12] приведено подробное описание этого датасета, для исследования я взяла только необходимую информацию.

Теперь опишем целевую переменную. В данном наборе при одном из вышеперечисленных режимов сбоя стоит единица, процесс завершается отказом, и в параметре «отказ машины» присваивается значение 1.

Исходя из всего вышеперечисленного, можно выделить 5 видов ошибок:

1. Отказ из-за износа.
2. Нарушение отвода тепла.
3. Сбой мощности.
4. Перегрузка.
5. Случайные сбои.

Возникновение каждого вила и будем прогнозировать. Также, добавим прогноз отказа оборудования в целом, по параметру «Machine Failure» (отказ оборудования).

В итоге, переходим к задаче бинарной классификации в случае, когда мы прогнозируем отказ в целом по параметру «отказ оборудования». При загрузке датасета в систему и определенной настройке модель способна спрогнозировать отказ на основании данных об оборудовании еще до его реализации, тем самым предотвратив реализацию угрозы. В контексте типов ошибок мы сводим все к задаче многоклассовой классификации с перекрестными классами (так как возможен вариант реализации нескольких ошибок сразу), чтобы определить, какой вид отказа наиболее вероятен при текущих значениях признаков системы.

Факторный анализ. Для того, чтобы определить наиболее значимые признаки была создана корреляционная матрица для количественных признаков, и по ней отрисована тепловая карта на рис. 1.

Фиксируется сильная корреляция температуры воздуха от температуры процесса, а также обратная связь между скоростью вращения и крутящим моментом. Однако, обычной матрицы корреляции недостаточно, потому используем факторный анализ. Факторный анализ позволяет решить две важные проблемы исследователя: описать объект измерения всесторонне и в то же время компактно. С помощью факторного анализа возможно выявление скрытых переменных факторов, отвечающих за наличие линейных статистических корреляций между наблюдаемыми переменными.

Две основных цели факторного анализа:

- ◆ определение взаимосвязей между переменными [13, 14];
- ◆ сокращение числа переменных необходимых для описания данных.

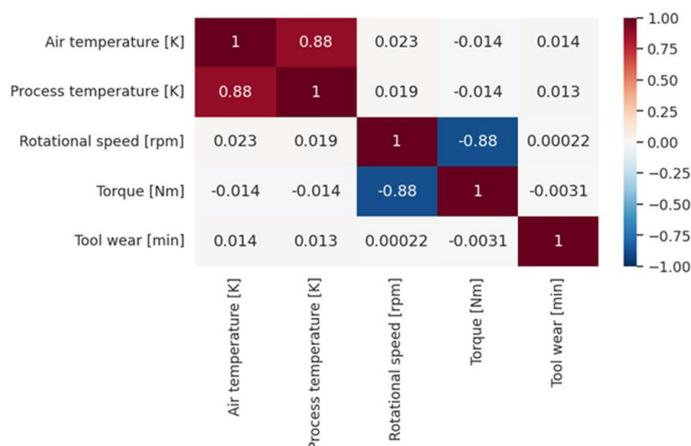


Рис. 1. Тепловая карта

Воспользуемся двумя методами: факторный анализ на основе метода главных компонент (PCA)[15] и метод `feature_importances_` [16] в классификаторе Случайного леса библиотеки `ScikitLearn`. На рис. 2 отображено влияние признаков.

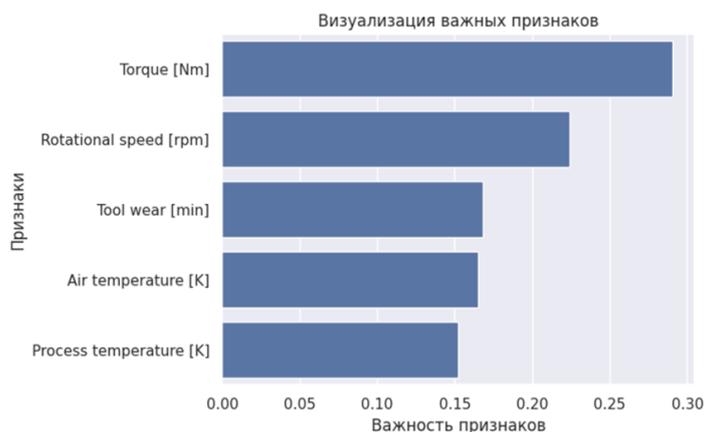


Рис. 2. Результаты влияния признаков методом Случайного леса

Ниже, в табл. 1 представлены результаты факторного анализа методом главных компонент и методом `feature importances_`.

Таблица 1

Результаты факторного анализа

	<code>feature importances</code>	PCA. Фактор 1	PCA. Фактор 2
Крутящий момент	0.290333	-0.693762	-0.628608
Скорость вращения	0.224212	0.705323	0.621208
Износ	0.167999	-0.011478	-0.020076
Температура воздуха	0.165356	0.642	-0.6804
Температура рабочего процесса	0.152099	0.643	-0.6802

Как видим из таблицы, методы показали, что наиболее влиятельными являются крутящий момент и скорость вращения по обоим подходам, факторный анализ методом главных компонент же выделил еще обе температуры, износ был определен самым слабым. Оба метода не противоречат друг другу, а скорее, дополняют.

Предобработка и обучение модели. Для лучшего понимания данных проведем разведочный анализ, чтобы лучше понять закономерности в датасете. На рис. 3 представлена визуализация распределения видов ошибок.

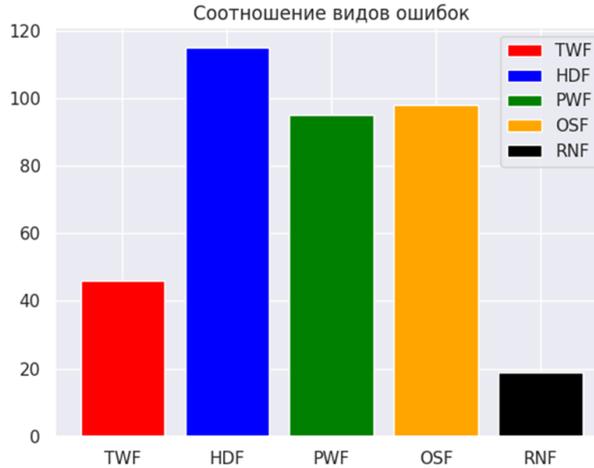


Рис. 3. Распределение видов ошибок

Исходя из этого, можно сделать вывод, что наибольшее количество ошибок более 100 это нарушение обмена тепла (HDF), процент случайных ошибок самый маленький.

Всего отказов в датасете зафиксировано 339 по параметру «Ошибка оборудования». Это меньше, чем ошибок по всем видам суммарно, т. к. есть записи, где возникали сразу две ошибки. В 23 записях ошибок две одновременно, в 1 записи три ошибки - суммарно 0,24% от всех записей или 7% от всех ошибок

Рис. 4 отражает распределение ошибок по качеству.

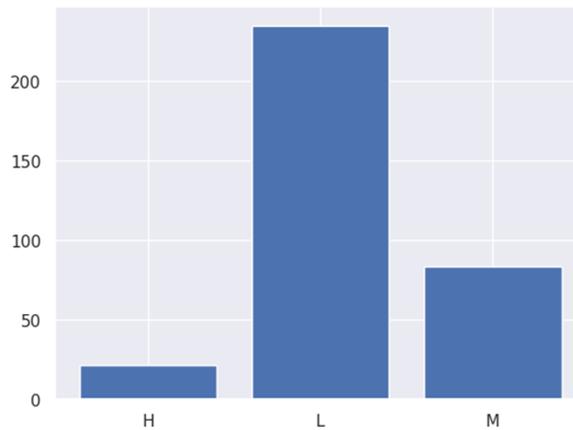


Рис. 4. Распределение ошибок по качеству

Распределение количества ошибок по типу показывает, что наибольшей вероятности подвержены детали низкого качества, однако это объясняется еще и тем, что их больше остальных.

Из-за проблемы дисбаланса классов было принято решение сократить количество записей без ошибок, чтобы исключить превосходство одного класса, так как в разрезе определения вида ошибки это значительно влияет на точность. Так как данные генерировались на основе реальных настроек, в случае прогнозирования аномалий всегда будет присутствовать превосходство «рабочих» состояний над «рисковыми». Для многоклассо-

вой классификации были выбраны алгоритмы из библиотеки Scikit-learn: AdaBoost [17, 18] и Random Forest (RF) [19, 20], а также использовалась стратегия «Один против всех» [21, 22]. На рис. 5 представлены матрицы без изменения количества строк в датасете.

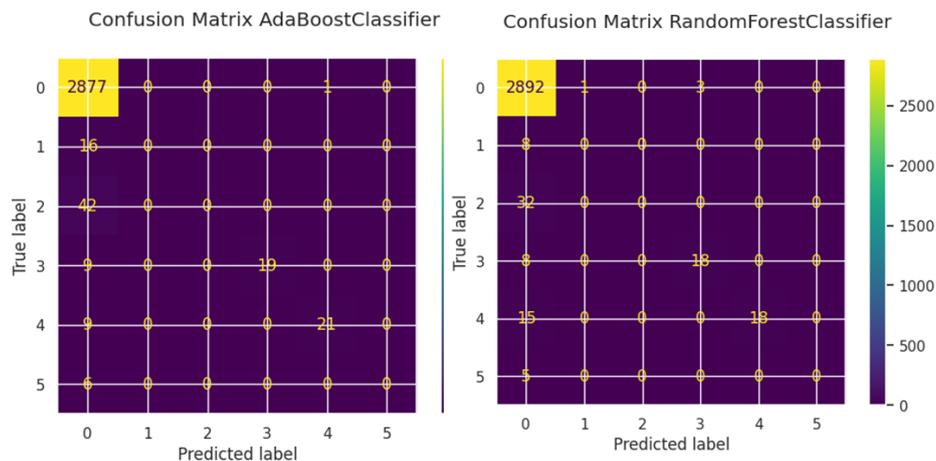


Рис. 5. Матрица ошибок

На рис. 6 приведены подробные описания моделей.

```

Classification report for classifier OneVsRestClassifier(estimator=AdaBoostClassifier(algorithm='SAMME',
n_estimators=200,
random_state=45)):

```

	precision	recall	f1-score	support
0	0.98	1.00	0.99	2896
1	0.00	0.00	0.00	8
2	0.00	0.00	0.00	32
3	0.87	0.50	0.63	26
4	0.95	0.58	0.72	33
5	0.00	0.00	0.00	5
accuracy			0.98	3000
macro avg	0.47	0.35	0.39	3000
weighted avg	0.96	0.98	0.97	3000

```

Classification report for classifier OneVsRestClassifier(estimator=RandomForestClassifier(random_state=0)):

```

	precision	recall	f1-score	support
0	0.98	1.00	0.99	2896
1	0.00	0.00	0.00	8
2	0.00	0.00	0.00	32
3	0.86	0.69	0.77	26
4	1.00	0.55	0.71	33
5	0.00	0.00	0.00	5
accuracy			0.98	3000
macro avg	0.47	0.37	0.41	3000
weighted avg	0.96	0.98	0.97	3000

Рис. 6. Подробное описание результата применения модели

Как видим, в разрезе классов модель сильно ошибается и не определяет вообще некоторые виды ошибок.

Чтобы улучшить качество, я случайным образом удаляю из датасета ~9200 значений, так как удаляемые всегда выбираются случайно, размер варьируется в пределах 20 строк.

Итоговый датасет содержит ~760 строк, где ошибки занимают около 45% от всего размера. Из-за сокращения наблюдений тестовую выборку сократили до 20%.

Далее, для получения наилучшего результата подберем показатель estimator для алгоритма AdaBoost так, чтобы точность модели на тестовой выборке была наилучшей. В качестве критерия качества я выбрала F1-меру (рис. 7).

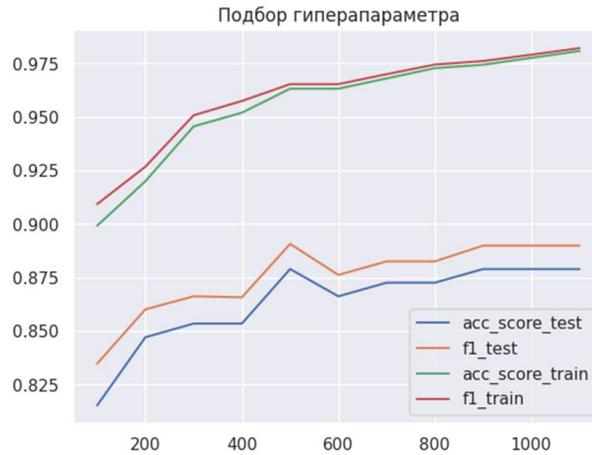


Рис. 7. Подбор гиперпараметров

Наилучшим вариантом оказался $n_estimators$ равный 500. Видим в сравнении с тестовой и тренировочной выборками, что переобучения нет на этом этапе, а полученная точность по ассигасу и F1 достигает 88%. Рассмотрим результат в разрезе классов на рис. 8.

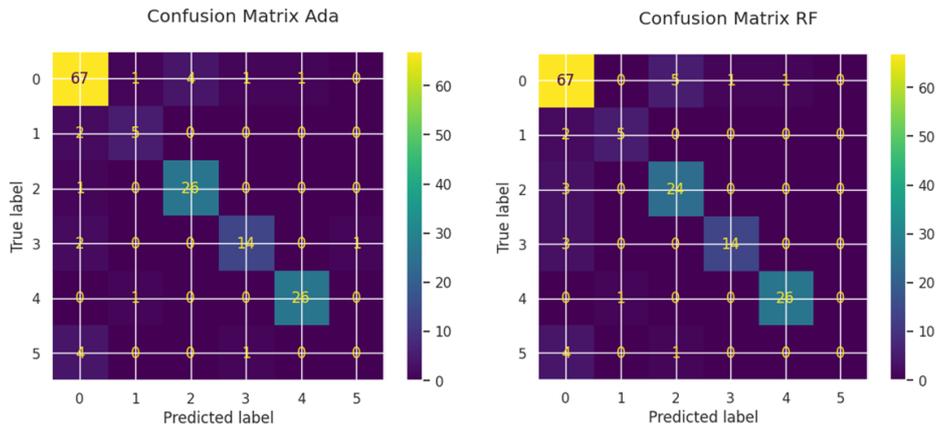


Рис. 8. Матрица ошибок по классам

Как видим, оба алгоритма неплохо справляются с классификациями ошибок, однако не определяется никак последний класс. Это можно объяснить следующим образом:

- ◆ количество этого вида очень маленькое;
- ◆ сам вид ошибки называется случайным и не маркируется как ошибка оборудования в исходном датасете, потому то, что наша модель этот вид ошибки определяет как ее отсутствие – наилучшее решение. 1-й вид ошибки определяется также не очень стабильно, потому как соответствующих наблюдений этой ошибки меньше всего.

В целом, модель имеет 85–88% точности, что достаточно неплохой показатель с учетом маленькой итоговой выборки. При наличии более качественных данных можно получить точность выше.

Полученная имитационная модель прогнозирования позволит определять вид отказа при входных параметрах оборудования до поломки, что позволит:

- ◆ определить диапазоны «рабочих» характеристик оборудования;
- ◆ протестировать сценарии и комбинации этих характеристик;
- ◆ заблаговременно принимать меры или сигнализировать о рисках.

В методологии указано, что критерий «вероятность возникновения» является относительным, а не абсолютным значением. Эксперт делает оценку на опыте, если не имеется иных указаний и шкал. В данном эксперименте была взята шкала преобразований из ГОСТ Р 51814.2-2001, описывающая соотношение количества возникновений с баллами от 1 до 10.

Итоговая оценка риска. Помимо этого, нам необходимо оценить еще два фактора: сложность обнаружения и тяжесть последствий. Для этого была привлечена экспертная оценка.

В табл. 2 представлен расчёт вероятности возникновения сходя из частоты появления каждого вида ошибки во всех 10000 значениях. Как видим, ранжирование сильно изменилось. Однако, мы здесь не учитываем важность критериев между собой.

Таблица 2

Расчет риска на основе прогноза

Ранг	Наименование ошибки	Вероятность возникновения	Сложность обнаружения	Тяжесть последствий	ПЧР
2	OSF. Отказ из-за перегрузки	5	6	9	270
3	PWF. Сбой мощности	5	6	6	180
1	RNF. Случайные сбои	4	10	7	280
5	HDF. Нарушение отвода тепла	5	1	5	25
4	TWF. Отказ из-за износа	5	3	8	120

Методы машинного обучения помогут обрабатывать и прогнозировать вид отказа методами многоклассовой классификации RandomForest, LinearSVC, AdaBoost, а также выявлять наиболее влияющие признаки методом главных компонент. Сравнение точности прогнозирования ошибок предоставлены в табл. 3.

Таблица 3

Точность прогнозирования разными методами

	LinearSVC	RandomForestClassifier	AdaBoost
Accuracy_Score	0.962	0.98175	0.973
F1_weighted	0.98	0.986	0.982
Accuracy_Score_сокращенный	0.53	0.88	0.84
F1_weighted_сокращенный	0.67	0.86	0.86

Как видим, Случайный лес показывает лучшую классификацию на нашей задаче. Это позволит применять систему для оценки рисков с учетом прогнозирования видов отказа.

Таким образом, пользователь получит следующие рекомендации, показанные в табл. 4.

Таблица 4

Итоговые рекомендации для ЛПП

Вид ошибки	Расшифровка	ПЧР Эксперта	ПЧР на основе прогноза
OSF	Отказ из-за перегрузки	1	2
PWF	Сбой мощности	2	3
RNF	Случайные сбои	3	1
HDF	Нарушение отвода тепла	4	5
TWF	Отказ из-за износа	5	4
Степень влияния признаков	Наиболее влиятельными признаками являются скорость вращения и крутящий момент, наименее – степень износа.		

Заключение. На текущем этапе исследования была протестирована часть гибридного подхода оценки рисков, включающая модель прогнозирования видов отказа, а также получение значения по критерию «вероятность возникновения» на основе данных о рабочих характеристиках оборудования и статистики появления. Помимо этого, были получены пояснения для ЛПП о влиянии признаков на возникновения отказа, выявленные при помощи факторного анализа. Для сравнения были использованы два метода: метод главных компонент и встроенный параметр алгоритма Random Forest. Метод главных компонент показал существенную разницу между влиянием признаков (разница в коэффициентах больше, чем у другого метода), потому в дальнейшем использовании в рамках разработки системы поддержки принятия решений планируется использовать именно его.

Результаты прогнозирования алгоритмами RF и AdaBoost показали приемлемые результаты в 85-88% распознавания по классам по F-мере. Если расширить выборку, можно улучшить показатели.

Использование модели прогнозирования для оценке рисков позволит меньше зависеть от эксперта, а в случае отсутствия достаточно квалифицированного эксперта заменить эту оценку, но при этом возникнет потребность в качественных данных.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *McDermott R.E., Mikulak, Raymond J., Beauregard Michael R.* The Basics of FMEA. – Productivity Press, 1996. – 80 p. – ISBN 9780527763206.
2. ГОСТ Р 51814.2–2001. Системы качества в автомобилестроении. Метод анализа видов и последствий потенциальных дефектов.
3. *Подоплелова Е.С., Князев И.И.* Модификация метода FMEA при помощи алгоритмов машинного обучения // Известия ЮФУ. Технические науки. – 2024. – № 6 (2023). – С. 88-95.
4. *Sprajcer M., Matthew J.W. Thomas, Charli Sargent, Meagan E. Crowther, Diane B. Boivin, Imelda S. Wong, Alison Smiley, Drew Dawson.* How effective are Fatigue Risk Management Systems (FRMS)? A review // Accident Analysis & Prevention. – 2022. – No. 165. – P. 106398.
5. *Jin C., Ran Y., Zhang G.* An improving failure mode and effect analysis method for pallet exchange rack risk analysis // Soft Comput. – 2021. – 25. – P. 15221-15241. – <https://doi.org/10.1007/s00500-021-06359-z>.
6. *Filz M., Langner J.E., Herrmann C., Thiede S.* Data-driven failure mode and effect analysis (FMEA) to enhance maintenance planning // Comput. Ind. – 2021. – 129. – 103451.
7. *Aboozar Jamalnia, Yu Gong, Kannan Govindan, Michael Bourlakis, Sachin Kumar Mangla.* A decision support system for selection and risk management of sustainability governance approaches in multi-tier supply chain // International Journal of Production Economics. – 2023. – No. 264. – P. 108960.
8. *Wan Suzila Wan Husin, Yazriwati Yahya, Nurulhuda Firdaus Mohd Azmi, Nilam Nur Amir Sjarif, Suriyati Chuprat, Azri Azmi.* Risk Management Framework for Distributed Software Team: A Case Study of Telecommunication Company // Procedia Computer Science. – 2019. – 161. – P. 178-186.
9. *Nabil K., Dkhissi. P.B.* A decision support system for evaluating the logistical risks in Supply chains based on RPN factors and multi criteria decision making approach // 2022 IEEE 6th International Conference on Logistics Operations Management (GOL). – Strasbourg, France, 2022. – P. 1-6.
10. *Dent T.H.S. et al.* Risk prediction models: a framework for assessment // Public health genomics. – 2012. – Vol. 15 (2). – P. 98-105. – DOI: 10.1159/000334436.

11. *Cai Y., Cai YQ., Tang, LY. Et al.* Artificial intelligence in the risk prediction models of cardiovascular disease and development of an independent validation screening tool: a systematic review // *BMC Med.* – 2024. – 22. – 56. – <https://doi.org/10.1186/s12916-024-03273-7>.
12. *Matzka S.* AI4I 2020 Predictive Maintenance Dataset, submitted to UCI Machine Learning Repository. – 2020.
13. *Ким Дж.-О., Мьюллер Ч.У.* Факторный анализ: статистические методы и практические вопросы // Сб. работ «Факторный, дискриминантный и кластерный анализ»: пер. с англ. / под. ред. И.С. Енюкова. – М.: Финансы и статистика, 1989. – 215 с.
14. *Бююль А., Цёфель П.* SPSS: Искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей. – СПб.: ООО «ДиаСофтЮП», 2002. – 603 с.
15. *Pearson K.* On lines and planes of closest fit to systems of points in space // *Philosophical Magazine.* – 1901. – Vol. 2. – P. 559-572.
16. Официальный сайт Scikit-learn. Feature importances with a forest of trees. – URL: https://scikit-learn.org/stable/auto_examples/ensemble/plot_forest_importances.html.
17. *Freund Y., Schapire R.* A Decision-Theoretic Generalization of on-Line Learning and an Application to Boosting. – 1995.
18. *Zhu J., Zou H., Rosset S., Hastie T.* Multi-class adaboost // *Statistics and its Interface.* – 2009. – Vol. 2. – P. 349-360.
19. *Prinzie A, Poel D.* Random Multiclass Classification: Generalizing Random Forests to Random MNL and Random NB // *Database and Expert Systems Applications. Lecture Notes in Computer Science.* – 2007. – Vol. 4653. – P. 349. – DOI: 10.1007/978-3-540-74469-6_35. ISBN 978-3-540-74467-2.
20. *Denisko D, Hoffman M.M.* Classification and interaction in random forests // *Proceedings of the National Academy of Sciences of the United States of America.* – 2018. – 115 (8). – P. 1690-1692. – Bibcode: 2018PNAS..115.1690D. – DOI: 10.1073/pnas.1800256115. PMC 5828645. – PMID 29440440.
21. *Bishop Christopher M.* Pattern Recognition and Machine Learning. – Springer, 2006.
22. Официальный сайт Scikit-learn. User Guide. OneVsRestClassifier. – URL: <https://scikit-learn.org/stable/modules/multiclass.html#ovr-classification>.

REFERENCES

1. *McDermott R.E., Mikulak, Raymond J., Beauregard Michael R.* The Basics of FMEA. Productivity Press, 1996, 80 p. ISBN 9780527763206.
2. GOST R 51814.2–2001. Sistemy kachestva v avtomobilestroenii. Metod analiza vidov I posledstviy potentsial'nykh defektov [GOST R 51814.2–2001. Quality systems in the automotive industry. Method for analyzing the types and consequences of potential defects].
3. *Podoplelova E.S., Knyazev I.I.* Modifikatsiya metoda FMEA pri pomoshchi algoritmov mashinnogo obucheniya [Modification of the FMEA method using machine learning algorithms], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SfedU. Engineering Sciences], 2024, No. 6 (2023), pp. 88-95.
4. *Sprajcer M., Matthew J.W. Thomas, Charli Sargent, Meagan E. Crowther, Diane B. Boivin, Imelda S. Wong, Alison Smiley, Drew Dawson.* How effective are Fatigue Risk Management Systems (FRMS)? A review, *Accident Analysis & Prevention*, 2022, No. 165, pp. 106398.
5. *Jin C., Ran Y., Zhang G.* An improving failure mode and effect analysis method for pallet exchange rack risk analysis, *Soft Comput*, 2021, 25, pp. 15221-15241. Available at: <https://doi.org/10.1007/s00500-021-06359-z>.
6. *Filz M., Langner J.E., Herrmann C., Thiede S.* Data-driven failure mode and effect analysis (FMEA) to enhance maintenance planning, *Comput. Ind.*, 2021, 129, 103451.
7. *Aboozar Jamalnia, Yu Gong, Kannan Govindan, Michael Bourlakis, Sachin Kumar Mangla.* A decision support system for selection and risk management of sustainability governance approaches in multi-tier supply chain, *International Journal of Production Economics*, 2023, No. 264, pp. 108960.
8. *Wan Suzila Wan Husin, Yazriwati Yahya, Nurulhuda Firdaus Mohd Azmi, Nilam Nur Amir Sjarif, Suriyati Chuprat, Azri Azmi.* Risk Management Framework for Distributed Software Team: A Case Study of Telecommunication Company, *Procedia Computer Science*, 2019, 161, pp. 178-186.
9. *Nabil K., Dkhissi. P.B.* A decision support system for evaluating the logistical risks in Supply chains based on RPN factors and multi criteria decision making approach, *2022 IEEE 6th International Conference on Logistics Operations Management (GOL)*. Strasbourg, France, 2022, pp. 1-6.
10. *Dent T.H.S. et al.* Risk prediction models: a framework for assessment, *Public health genomics*, 2012, Vol. 15 (2), pp. 98-105. DOI: 10.1159/000334436.

11. Cai Y., Cai YQ., Tang, LY. et al. Artificial intelligence in the risk prediction models of cardiovascular disease and development of an independent validation screening tool: a systematic review, *BMC Med.*, 2024, 22, 56. Available at: <https://doi.org/10.1186/s12916-024-03273-7>.
12. Matzka S. AI4I 2020 Predictive Maintenance Dataset, submitted to UCI Machine Learning Repository, 2020.
13. Kim Dzh.-O., M'yuller Ch.U. Faktornyy analiz: statisticheskie metody i prakticheskie voprosy [Factor analysis: statistical methods and practical issues], *Sb. rabot «Faktornyy, diskriminantnyy i klasternyy analiz»* [Collection of works "Factor, discriminant and cluster analysis"]: transl. from engl., ed. by I.S. Enyukova. Moscow: Finansy i statistika, 1989, 215 p.
14. Byuyul' A., Tsefel' P. SPSS: Iskustvo obrabotki informatsii. Analiz statisticheskikh dannykh i vosstanovlenie skrytykh zakonornostey [SPSS: The art of information processing. Analysis of statistical data and reconstruction of hidden patterns]. Saint Petersburg: OOO «DiaSoftYUP», 2002, 603 p.
15. Pearson K. On lines and planes of closest fit to systems of points in space, *Philosophical Magazine*, 1901, Vol. 2, pp. 559-572.
16. Scikit-learn. Feature importances with a forest of trees. Available at: https://scikit-learn.org/stable/auto_examples/ensemble/plot_forest_importances.html.
17. Freund Y., Schapire R. A Decision-Theoretic Generalization of on-Line Learning and an Application to Boosting, 1995.
18. Zhu J., Zou H., Rosset S., Hastie T. Multi-class adaboost, *Statistics and its Interface*, 2009, Vol. 2, pp. 349-360.
19. Prinzie A, Poel D. Random Multiclass Classification: Generalizing Random Forests to Random MNL and Random NB, *Database and Expert Systems Applications. Lecture Notes in Computer Science*, 2007, Vol. 4653, pp. 349. DOI: 10.1007/978-3-540-74469-6_35. ISBN 978-3-540-74467-2.
20. Denisko D, Hoffman M.M. Classification and interaction in random forests, *Proceedings of the National Academy of Sciences of the United States of America*, 2018, 115 (8), pp. 1690-1692. Bibcode: 2018PNAS..115.1690D. DOI: 10.1073/pnas.1800256115. PMC 5828645. PMID 29440440.
21. Bishop Christopher M. Pattern Recognition and Machine Learning. Springer, 2006.
22. Scikit-learn. User Guide. OneVsRestClassifier. Available at: <https://scikit-learn.org/stable/modules/multiclass.html#ovr-classification>.

Подоплелова Елизавета Сергеевна – Южный федеральный университет; e-mail: chuzhinova@sfedu.ru; г. Таганрог, Россия; тел.: +79525844188; кафедра информационно-аналитических систем безопасности им. Л.С. Берштейна; старший преподаватель.

Podoplelova Elizaveta Sergeevna – Southern Federal University; e-mail: chuzhinova@sfedu.ru; Taganrog, Russia; phone: +79525844188; Department of Information and Analytical Security Systems named after L.S. Bershtein; senior lecturer.