

EDN: HOUJHA

УДК 621.316.13:621.315.1:621.315.668: 621.3.019.3

## Machine Learning Application for Predicting 110 kV PTL Failures Based on PTL Parameters

Vadim E. Bolshev\*

*Federal Scientific Agroengineering Center VIM  
Moscow, Russian Federation*

Received 26.11.2024, received in revised form 02.12.2024, accepted 16.12.2024

**Abstract.** This study proposes the use of machine learning algorithms to predict 110 kV power line failures based on data on the parameters of the lines themselves. Five classifiers were tested as machine learning algorithms: support vector machine, logistic regression, random forest, gradient boosting LightGBM Classifier and CatBoostClassifier. For designed model a pipeline and a compositor of heterogeneous features were used to automate the process of data conversion and eliminate the possibility of data leakage. Data were prepared using hot coding method for categorical variables and standardization method for quantitative ones. The model was trained using the cross-validation method with stratified separation. Through the use of grid search and random parameter optimization techniques, the classifiers' hyperparameters were changed. The prediction quality of the trained models was compared using the metrics ROC-AUC, AUC-PR, Accuracy, accuracy, recall and F-1 measure. The best results in predicting outages were achieved by the logistic regression model with the class weighting method to combat class imbalance, the ROC-AUC metric of which reached 0.84 on the test sample. Thus, this study confirms the possibility of using data on power line parameters to predict 110 kV power line failures.

**Keywords:** electrical network, power transmission line, PTL, power supply reliability, power outage, power line failure, machine learning.

Citation: Bolshev V.E. Machine Learning Application for Predicting 110 kV PTL Failures Based on PTL Parameters. J. Sib. Fed. Univ. Eng. & Technol., 2025, 18(1), 19–32. EDN: HOUJHA



# Использование машинного обучения для прогнозирования отключений электроэнергии на ЛЭП 110 кВ на основе параметров самих ЛЭП

**В. Е. Большев**

*Федеральный научный агроинженерный центр ВИМ  
Российская Федерация, Москва*

**Аннотация.** В рамках данного исследования предложено использование алгоритмов машинного обучения для прогнозирования отключений электрической энергии на линиях электропередачи 110 кВ на основе данных по параметрам самих линий. В качестве алгоритмов машинного обучения были опробованы 5 классификаторов: машина опорных векторов, логистическая регрессия, случайный лес, градиентные бустинги LightGBM Classifier и CatBoostClassifier. Для автоматизации процесса преобразования данных и устранения возможности их утечки использовался пайплайн и компоновщик разнородных признаков, данные для прогнозирования подготавливались методами горячего кодирования для категориальных переменных и стандартизации для количественных. Обучение модели производилось методом кросс-валидации со стратифицированным разделением. Настройка гиперпараметров классификаторов осуществлена методами оптимизации случайных параметров и сеточного поиска. Сравнение качества предсказания обученных моделей производилось по метрикам ROC-AUC, AUC-PR, Accuracy, точность, полнота и F-1 мера. Наилучшие результаты прогнозирования отключений удалось достичь модели логистической регрессии с методом взвешивания классов в качестве борьбы с дисбалансом классов, метрика ROC-AUC которой на тестовой выборке достигла 0.84. Таким образом, данное исследование подтверждает возможность использования данных по параметрам ЛЭП для прогнозирования отключений электрической энергии на ЛЭП 110 кВ.

**Ключевые слова:** электрические сети, линии электропередачи, ЛЭП, надежность электроснабжения, отключения электроэнергии, отказы ЛЭП, машинное обучение.

Цитирование: Большев В. Е. Использование машинного обучения для прогнозирования отключений электроэнергии на ЛЭП 110 кВ на основе параметров самих ЛЭП / В. Е. Большев // Журн. Сиб. федер. ун-та. Техника и технологии, 2025, 18(1). С. 19–32. EDN: HOUJHA

## Введение

Снабжение потребителей электрической энергией осуществляется через сложную распределительную систему, которая включает в себя многочисленные воздушные и подземные ЛЭП различного напряжения, трансформаторы, опоры, линейную аппаратуру различных типов и другое оборудование. Надежность системы электроснабжения (СЭС), то есть вероятность бесперебойного снабжения потребителей электрической энергией, увеличивается с использованием новых, инновационных решений, включающих в себя внедрение современных средств защиты, мониторинга и управления электрическими сетями [1, 2]. Несмотря на внедрение современных технологий, отказы на линиях электропередачи являются относительно частым явлением и составляют 35–50 % от всех отказов в системах электроснабжения напряжением 35–750 кВ [3]. Такой объем отключений связан с большой территориальной протяженностью ЛЭП и их подверженностью влиянию климатическим воздействиям [4, 5]. Несмотря на то что неблагоприятные погодные условия являются самыми часто встречаемыми причинами неисправности, необходимо учитывать и другие причины, в том числе износ и устаревание инфраструктуры электрической сети [6].

Анализ данных об отключениях электрической энергии позволяет определить факторы, оказывающие наибольшее влияние на вероятность отказа. Существует большое количество работ, посвященных статистическому анализу отказов на ЛЭП, как например, в работах [7–9]. Несмотря на то что в некоторых работах по статистическому анализу аварийных отключений были определены причины отказов с указанием определенных элементов, вышедших из строя, в них не рассматривалось влияние типов этих элементов на вероятность отказов ЛЭП. Одним из эффективных методов определения вероятности отключения на основе таких факторов является прогностическое моделирование, построенное на методах машинного обучения (ML – machine learning) [10]. Так, в работе [11] предлагается применение трехмерного метода опорных векторов (Support Vector Machine – SVM) для прогнозирования отключений компонентов энергосистемы, а в работе [12] этот же метод используется для определения места повреждения в системе посредством измерения величины и угла падения напряжения на первичной подстанции распределительной системы. В статье [13] представлен подход к выявлению неисправностей оборудования в распределительных системах, в результате которого решается задача бинарной классификации: определение отказов, связанных с оборудованием и не связанных. В качестве классификаторов используются три алгоритма: дерево решений, логистическая регрессия и наивный байесовский классификатор. Применение искусственных нейронных сетей представлено в работе [14] для многоклассовой классификации неисправностей систем электроснабжения на основе значений тока и напряжения на всех линиях электропередачи (ЛЭП).

Большое количество работ посвящено применению методов машинного обучения для прогнозирования отключений в электрораспределительных сетях во время неблагоприятных погодных условий [15], в частности ураганов, тропических штормов, дождей и ветровых штормов, а также прибрежных наводнений. В случае возникновения тайфуна для определения количества отключений в работе [16] используется алгоритм случайного леса, решающий задачу многоклассовой классификации, в работе [17] – алгоритм градиентного бустинга, в работе [18] – ансамбль, состоящий из 2 уровней, причем последний – градиентный бустинг XGBoost. В статье [19] предлагается подход для прогнозирования отключений в распределительных системах, вызванных факторами окружающей среды, за счет использования глубоких нейронных сетей, объединенных в ансамбль.

Таким образом, проведя обзор существующей литературы по прогнозированию отключений электрической энергии, было установлено, что прогнозирование отказов ЛЭП на основе параметров самих линий не проводился ни в каком виде, что говорит об актуальности проводимого исследования, поэтому актуальность данного исследования подтверждена. Целью данной работы является разработка модели машинного обучения для прогнозирования возможных отключений электроэнергии на линиях электропередачи на основе характеристик самих ЛЭП.

## **1. Материалы и методология**

### *1.1. Материалы для исследования*

Настоящее исследование построено на данных по отключениям электрической энергии в электрических сетях Орловской области. В предыдущем исследовании [20] рассматриваемые данные были проанализированы методами математической статистики, в рамках которого были проведены обработка пропущенных значений, удаление дубликатов, создание синтети-

ческих параметров, включая целевой признак, выявление выбросов и аномалий, выбор наиболее подходящих признаков для обучения ML-модели. Итоговым результатом стала таблица из 9 признаков, включая целевой, и 395 объектов. Таблица содержит следующие признаки:

3 признака с категориальными значениями:

- факт отключения (целевой признак);
- проводник, тип, сечение;
- отношение ЛЭП к транзиту.

6 признаков с количественными значениями:

- индекс состояния,%;
- протяженность воздушных участков, км;
- переэксплуатация, бр.;
- ЖБ опоры,%;
- протяженность по лесу,%;
- протяженность по населенной местности,%.

### *1.2. Методология исследования*

Так как целью работы является прогнозирование вероятности отключения ЛЭП на основе её параметров, то алгоритмами машинного обучения решается задача бинарной классификации, поэтому в качестве классификаторов были выбраны 5 ML-моделей, основанных на следующих алгоритмах:

- машина опорных векторов (Support Vector Machine – SVM);
- логистическая регрессия (LogisticRegression – LR);
- случайный лес (RandomForestClassifier – RFC);
- градиентные бустинги, основанные на деревьях решений, LightGBM Classifier и CatBoostClassifier.

Правильная настройка гиперпараметров классификаторов осуществлена посредством метода оптимизации случайных параметров RandomizedSearchCV, а лучшие найденные параметры уточнены алгоритмом сеточного поиска GridSearchCV. Разбиение данных на обучающую и валидационную выборки выполнено через кросс-валидацию со стратифицированным разделением. Подготовка данных для ML-моделей осуществлена методом горячего кодирования One-Hot Encoding для категориальных переменных и методом стандартизации данных Standard Scaler для количественных. Для автоматизации процесса преобразования данных и обучения модели, а также для устранения возможности утечки данных использовался пайплайн (Pipeline) и компоновщик разнородных признаков Column Transformer. Выбор лучшей обученной ML-модели производился метрикой качества ROC-AUC. Кроме метрики ROC-AUC оценка качества моделей производилась дополнительными метриками, такими как AUC-PR, Accuracy, Precision, Recall, F1-score.

### *1.3 Применяемое программное обеспечение*

В данном исследовании обработка и анализ данных производились на языке программирования Python в среде разработки Jupyter notebook (ver. 7.0.6) программного комплекса Anaconda Python. Для работы с табличными данными применялась библиотека Pandas (ver. 2.1.4), для

математической обработки массивов данных – NumPy (ver. 1.26.3), для визуализации данных – Matplotlib (ver. 3.8.0) и Seaborn (ver. 0.12.2), для корреляционного анализа Phik (ver. 0.12.4). Для выстраивания алгоритмов обучения, преобразования данных, пригодных для ML-задач, а также работы с основными классическими моделями машинного обучения использовалась библиотека Scikit-learn (ver. 1.2.2). Дополнительно к ней применялись фреймворки градиентных бустингов из библиотек LightGBM (ver. 4.1.0) и CatBoost (ver. 1.2.2). Для работы с несбалансированным набором данных применялись инструменты из библиотеки Imbalanced-learn.

## 2. Подготовка и выбор ML-моделей

### 2.1. Алгоритм обучения и настройки гиперпараметров ML-моделей

Алгоритм обучения ML-моделей посредством метода оптимизации случайных параметров RandomizedSearchCV представлен на рис. 1. В соответствии с алгоритмом рандомизированный набор гиперпараметров модели выбирается из распределения по возможным значениям параметров (сетка параметров) и инициализируется для обучения (блок 4). Количество вариаций гиперпараметров моделей методом RandomizedSearchCV выбрано равным 100 итераций (значение “n\_iter” – блок 2 рис. 1). В этом же блоке выбирается количество фолдов для кросс-валидации (значение “cv” – блок 2 рис. 1). Значение “scor\_iter” (значение метрики качества модели) приравнивается к 0 и используется в дальнейшем для поиска лучшей модели (блоки 11, 12, 13). Как только количество итераций превысит значение “n\_iter” (блок 3), процесс подбора гиперпараметров остановится и будет выдана ML-модель с гиперпараметрами, соответствующими лучшей метрике качества (блок 14).

### 2.2. Кодирование и масштабирование признаков

Решающее значение в разработке и применении моделей машинного обучения, в том числе нейронных сетей, имеет целенаправленный и тщательный подход к подготовке данных. Успех процесса обучения ML-моделей во многом зависит от качества и пригодности данных, подаваемых в модели. На предыдущем этапе была выполнена важная часть обработки – это обработка пропущенных значений и дубликатов, а также определение подходящих признаков для поставленной задачи классификации [20]. Теперь требуется проведение как преобразования категориальных переменных в числовые представления, так и нормализация количественных переменных: масштабирование и центрирование числовых элементов для облегчения сходимости.

Как было отмечено выше, категориальные переменные имеют низкую кардинальность, поэтому решено воспользоваться кодированием этих переменных методом горячего кодирования One-Hot Encoding. Для нормализации количественных переменных был использован метод стандартизации данных Standard Scaler (блок 7 рис. 1). Для автоматизации процесса преобразования данных и обучения модели, а также для устранения возможности утечки данных применен пайплайн (Pipeline) и компоновщик разнородных признаков Column Transformer.

Для борьбы с дисбалансом классов целевого признака решено воспользоваться двумя методами: методом взвешивания классов, реализованных внутри моделей из библиотеки scikit-learn, и методом синтетической избыточной выборки меньшинства для номинальных и непрерывных значений (Synthetic Minority Over-sampling Technique for Nominal and Continuous – SMOTE-NC),

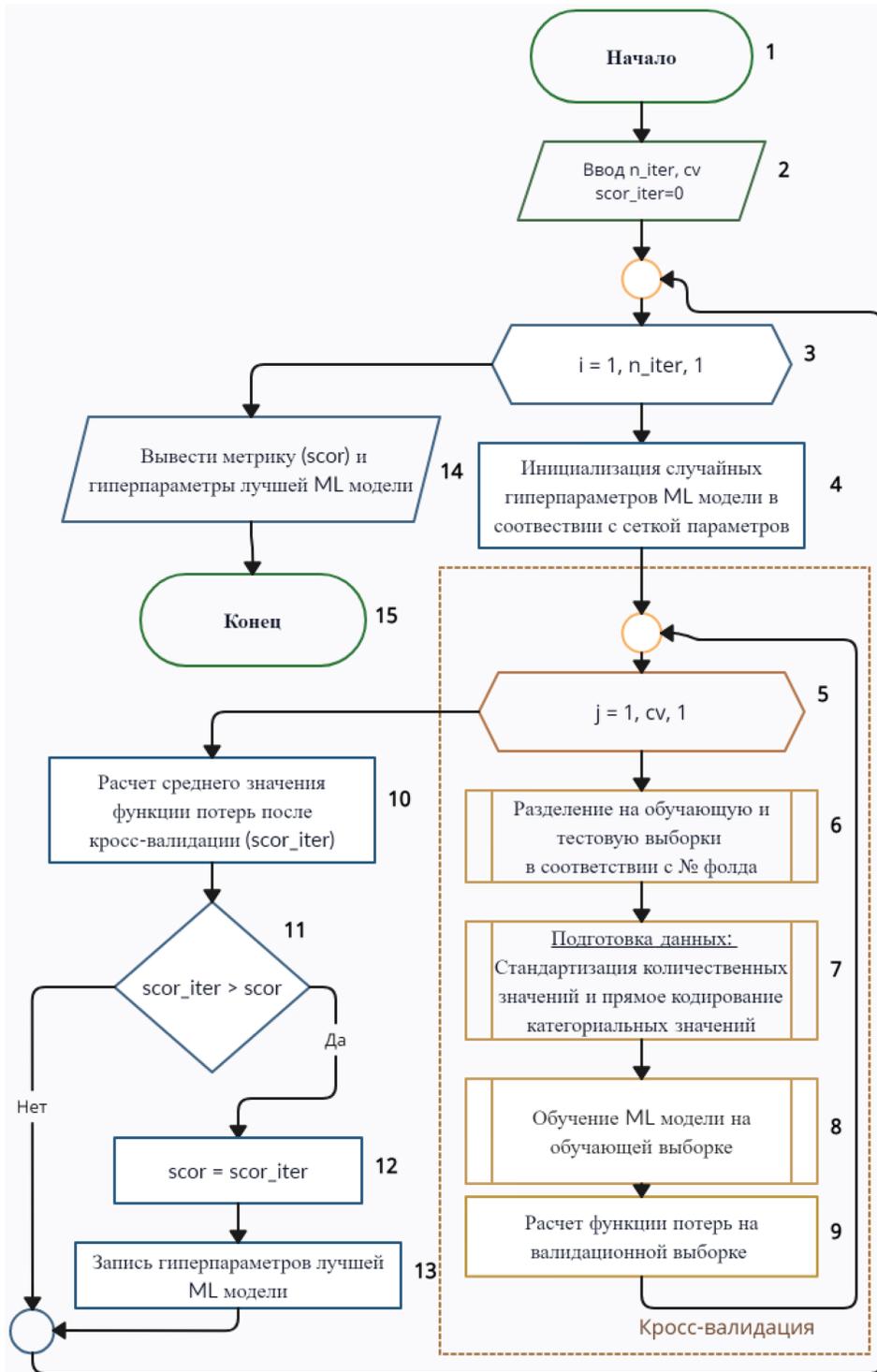


Рис. 1. Алгоритм оптимизации гиперпараметров ML-моделей

Fig. 1. Algorithm for optimizing ML model hyperparameters

реализованным через Pipeline библиотеки `imblearn` для автоматизации процесса и устранения проблемы утечки признаков. Оба метода применены при обучении каждой модели, и затем полученные результаты представлены в итоговой таблице.

### *2.3. Разбиение датасета на обучающую и тестовую выборки*

Еще одним важным моментом при построении моделей машинного обучения является проверка её качества на независимых данных, то есть на тех данных, которые модель не видела. С этой целью из исходных данных выделяется тестовая выборка обычно в размере 10–20 % от обучающей. Данный подход позволяет определить и, соответственно, устранить основную проблему обучения ML-моделей – переобучение, то есть явления, когда обученные ML-модели заучивают ответы на обучающей выборке, но плохо определяют закономерности на сторонних данных. Однако так как объем данных составляет 395 объектов, то было решено отделить тестовую выборку в размере 20 %, а к обучающей выборке применить метод кросс-валидации (на рис. 1 блоки 5–9,10), подразумевающий разделение данных на несколько частей (фолдов), при этом каждый фолд на своем этапе обучения должен выступать в качестве валидационной выборки, остальные – в качестве обучающих. В нашем случае была использована кросс-валидация на 5 фолдов со стратифицированным разделением, гарантирующим одинаковое соотношение классов на всех выборках, что особенно важно при несбалансированных данных. Таким образом, количество итераций обучения на одном наборе гиперпараметров модели зависит от количества фолдов (значение “cv” блок 5 рис. 1). Согласно алгоритму, как только обучение проведется на всех фолдах, будет произведен расчет среднего значения метрики качества (блок 10), которое далее будет сравнено с лучшей метрикой в блоке условия 11.

### *2.4. Обучаемые ML-модели и используемая сетка гиперпараметров*

#### ***Метод опорных векторов (Support Vector Machine – SVM)***

Метод опорных векторов (SVM) – это алгоритм обучения с учителем, используемый для классификации и регрессионного анализа [21]. Алгоритм SVM работает путем преобразования входных данных в многомерное пространство с помощью функции ядра, затем поиска гиперплоскости, которая лучше всего разделяет точки данных на разные классы. Основными гиперпараметрами, от которых зависит обучение SVM модели, являются коэффициент регуляризации  $C$ , тип используемого ядра и, соответственно, его коэффициент.

#### ***Логистическая регрессия (LogisticRegression – LR)***

Регрессионное моделирование – один из наиболее популярных статистических подходов, позволяющий определить взаимосвязи между целевой переменной и набором независимых предикторов. Для предсказания вероятности бинарного значения используется расширенная версия линейной регрессии – логистическая регрессия, предсказывающая вероятностное распределение события (да/нет или 1/0) через функцию логит-связи [22]. В процессе обучения были выбраны нерегулируемые гиперпараметры алгоритма оптимизации (“saga”) и вида регуляризации (“Elastic-Net”), позволяющего использовать оба вида регуляризации (l1 и l2) одновременно. Приемлемое соотношение между этими двумя видами регуляризации определялось изменением гиперпараметра “l1\_ratio”. Также регулировалась сила регуляризации  $C$ .

### ***Случайный лес (RandomForest)***

Случайный лес является одним из наиболее широко используемых алгоритмов машинного обучения, применяемого как в задачах регрессии, так и классификации [23]. Алгоритм случайного леса основан на методе ансамблевого обучения, результаты которого строятся за счет объединения независимых прогнозов леса – множества решающих деревьев, обученных на выборках, полученных с помощью метода бутстрап (англ. bootstrap) [24]. Как правило, чем больше решающих деревьев, тем выше качество модели, поэтому при обучении случайного леса использовалось до 1000 деревьев. Известно, что глубина решающих деревьев также влияет на качество предсказания модели – чем она больше, тем качество выше, поэтому обучение производилось на глубине от 1 до 21 объекта. В дополнение к этому также перебирались все критерии расщепления, ограничение на число объектов в листьях и минимальное число объектов, при котором выполняется расщепление.

### ***Алгоритмы градиентного бустинга LightGBM и CatBoost***

Также применен метод градиентного бустинга над решающими деревьями, суть которого заключается в обучении некоторого числа моделей (в нашем случае решающих деревьев) с учетом ошибок, полученных на предыдущих моделях. Градиентный бустинг позволяет строить аддитивную функцию в виде суммы решающих деревьев итерационно, по аналогии с методом градиентного спуска [25]. Таким образом, такой подход позволяет достичь более высокой точности предсказания. В качестве алгоритмов градиентного бустинга воспользуемся LightGBM Classifier и CatBoostClassifier, имеющими схожие функциональные возможности для поддержки автоматической обработки категориальных функций [26].

Алгоритмы градиентного бустинга достаточно быстро переобучаются, поэтому требуется производить точный подбор гиперпараметров для получения качественной модели. Так же как и со случайным лесом, важное значение имеет количество решающих деревьев, их глубина и количество листьев (терминальных узлов). Однако если для случайного леса большая глубина деревьев в большинстве случаев оказывает положительную роль на качество предсказаний, то для градиентных бустингов рекомендуется использовать среднюю глубину, что позволяет достичь баланса между обученностью и обобщенностью. Также важными параметрами являются скорость обучения, являющаяся степенью вклада каждого дерева в прогнозирование модели, тип алгоритма бустинга (gbdt, dart, goss) и степень регуляризации.

### ***2.5. Оценка качества ML-моделей***

Эффективность предсказания дихотомической переменной в ML-моделях оценивается различными метриками, рассчитываемыми на основе матрицы ошибок (рис. 2). Матрица ошибок (матрица неточностей) представляет собой классификацию результатов предсказания моделей на истинно-положительные TP (True Positive), истинно-отрицательные TN (True Negative), ложноположительные FP (False Positive) и ложноотрицательные FN (False Negative), как показано на рис. 2.

Так как в исследуемых данных имеется небольшой дисбаланс классов, то качество модели было оценено метрикой ROC-AUC, невосприимчивой к этой проблеме. Метрика ROC-AUC является площадью под кривой ошибок (ROC-кривой), отображающей соотношение между количеством верно и ошибочно классифицированных ответов при варьировании порога решающего правила [27]. Значения метрики находятся в пределах от 0 до 1, при этом значение 1 гово-

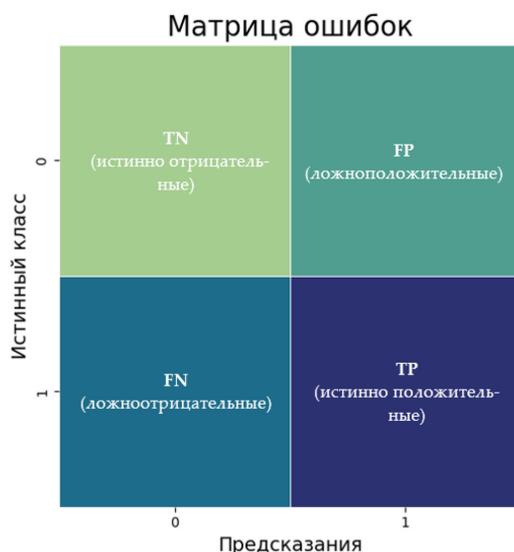


Рис. 2. Матрица ошибок

Fig. 2. Error Matrix

рит о высоком качестве модели, 0,5 – о случайном предсказании ответов модели, 0 – о ложном предсказании, когда предсказания модели противоположны истинным значениям. Верно классифицированные TPR (True Positive Rate) и ошибочно классифицированные FPR (False Positive Rate) находятся по формулам (1) и (2).

$$TPR = \frac{TP}{TP + FN}. \quad (1)$$

$$FPR = \frac{FP}{FP + TN}. \quad (2)$$

Кроме метрики ROC-AUC, взятой в качестве функции потерь при обучении моделей, рассчитаны дополнительные метрики качества для оценки работы моделей [28], такие как:

- Accuracy (правильность) – доля правильных ответов:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

- Precision (точность) – доля истинно положительных предсказаний среди всех положительных предсказаний модели:

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

- Recall (полнота) – доля положительных предсказаний среди всех положительных случаев:

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

- F1-score (F-1 мера) – среднее гармоническое полноты и точности, принимающее значение от 0 до 1 и позволяющее оценить качество модели при наличии несбалансированного набора данных:

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision * Recall} \quad (6)$$

– AUC-PR – площадь под PR-кривой, отображающей соотношение между метриками полноты (Precision) и точности (Recall). AUC-PR в отличие от ROC-AUC восприимчива к небалансу классов.

### Результаты обучения моделей

Результаты обучения моделей сведены в табл. 1 и отсортированы по метрике ROC-AUC. Для проверки на адекватность, в дополнение к рассмотренным моделям, была добавлена константная модель DummyClassifier со стратегией классификации “uniform”, генерирующей предсказания случайным образом с равной вероятностью для каждого класса.

Согласно табл. 1 все обученные ML-модели показали результаты на порядок выше предсказаний константной модели. Наилучшим методом борьбы с несбалансированной выборкой оказался метод взвешивания классов. По метрике ROC-AUC (0,78) лучшей моделью выбрана логистическая регрессия с гиперпараметрами penalty – 'elasticnet', solver – 'saga', l1\_ratio – 0,8, C – 1, что выражается в простоте самой модели и хорошей аппроксимации. Данная модель показала также лучшие результаты по метрике AUC-PR (0,68), что указывает на успешное прогнозирование целевого признака в условиях несбалансированной выборки. Модель была проверена на тестовой выборке и показала отличные результаты – ROC-AUC со значением 0,84. Значение метрики оказалось выше, чем на валидационной выборке, что указывает на отсутствие переобучения модели, однако большие расхождения между метриками на разных выборках сигнализируют о проблеме недостаточного количества данных (в тестовой выборке было всего 79 объектов). Проблема малой выборки наглядно видна на ROC и PR кривых, на которых имеются резкие «переломы» из-за недостаточности данных (рис. 3).

Таблица 1. Метрики качества обученных моделей

Table 1. Quality metrics of trained models

ML модель	ROC-AUC	AUC-PR	Accuracy	Recall	Precision	F1
Logistic Regression	0,779	0,683	0,69	0,685	0,611	0,644
CatBoost Classifier	0,776	0,655	0,693	0,708	0,618	0,657
Logistic Regression (SMOTE)	0,772	0,676	0,69	0,678	0,612	0,641
Random Forest Classifier	0,772	0,661	0,687	0,693	0,607	0,644
LightGBM Classifier (SMOTE)	0,772	0,652	0,69	0,686	0,609	0,642
LightGBM Classifier	0,771	0,647	0,715	0,762	0,631	0,687
CatBoost Classifier (SMOTE)	0,771	0,64	0,712	0,692	0,641	0,664
Support Vector Machine	0,768	0,672	0,687	0,685	0,606	0,642
Support Vector Machine (SMOTE)	0,761	0,671	0,687	0,67	0,61	0,637
Random Forest Classifier (SMOTE)	0,761	0,67	0,671	0,67	0,592	0,626
Dummy Model	0,5	0,411	0,519	0,523	0,431	0,472

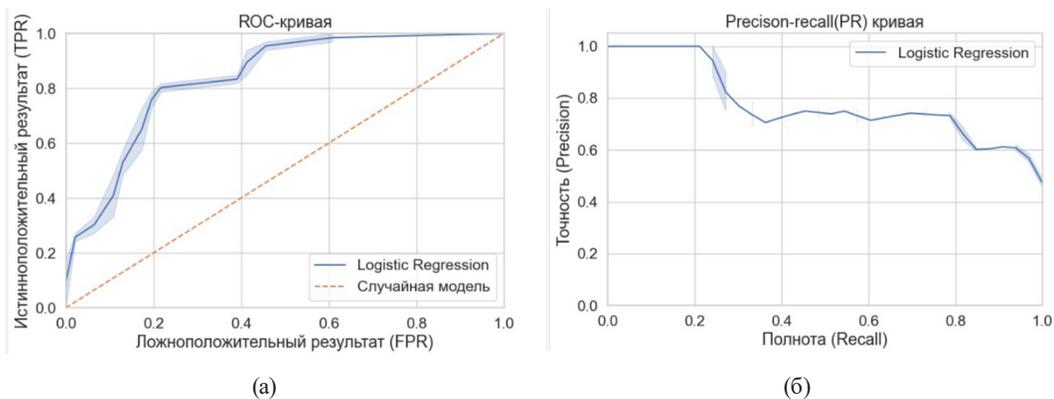


Рис. 3. ROC и PR кривые на тестовой выборке для модели логистической регрессии: (а) ROC кривая, (б) Precision-recall кривая

Fig. 3. ROC and PR curves on the test set for logistic regression model: (a) ROC curve, (b) Precision-recall curve

Несмотря на это, модель справляется с поставленной задачей, а графическое отображение ROC кривой (рис. 3а) позволяет утверждать, что при изменении порога классификации можно достичь высокого результата. Например, если необходимо, чтобы в предсказание модели попадали все 100 % ЛЭП, на которых наблюдаются отказы (отсутствие ошибок I рода), то согласно графику ROC кривой (рис. 3а), к этому классу будет отнесено до 60 % ЛЭП, на которых отказов не наблюдалось (ошибки II рода). И наоборот, если нужна точность предсказаний для обоих классов, то можно вычислить, что при пороге классификации приблизительно 0,4 истинно-положительные предсказания TPR равны 0,8, тогда как ложноположительные предсказания составят всего лишь 0,2. На PR кривой (рис. 3б) видно, что точность (Precision) стабилизируется на значении 0,72...0,75 при полноте (Recall) от 0,35 до 0,78 при соответствующих порогах классификации 0,45...0,58. Таким образом, знание рассмотренных выше метрик при определенных уровнях порога классификации позволяет регулировать количество ошибок I и II рода.

### Выводы

В рамках данного исследования было предложено использование моделей машинного обучения, а именно классификаторов SVM, LogisticRegression, RandomForestClassifier, LightGBM Classifier и CatBoostClassifier, для прогнозирования отключений электрической энергии на линиях электропередачи 110 кВ на основе данных по параметрам самих линий. Данные для моделей предложено подготавливать методом горячего кодирования One-Hot Encoding для категориальных переменных и методом стандартизации данных Standard Scaler для количественных. Для автоматизации процесса преобразования данных и устранения возможности их утечки используется пайплайн (Pipeline) и компоновщик разнородных признаков Column Transformer. Настройка гиперпараметров классификаторов осуществлена методами оптимизации случайных параметров RandomizedSearchCV и сеточного поиска GridSearchCV.

По метрике ROC-AUC лучшей моделью стала логистическая регрессия с методом взвешивания классов в качестве борьбы с их дисбалансом, показавшая результат в 0,78, что выражается в простоте самой модели и хорошей аппроксимации. Данная модель показала также лучшие результаты по метрике AUC-PR (0,68), что указывает на успешное прогнозирование целевого

признака даже в условиях несбалансированной выборки. На тестовой выборке логистическая регрессия также показала отличный результат – ROC-AUC со значением 0,84, что подтвердило отсутствие переобучения модели, но выявило влияние проблемы недостаточного количества данных, выраженного в значительном превышении метрик на тестовой выборке в сравнении с валидационной.

Результаты данного исследования показывают возможность прогнозирования отключений электрической энергии на линиях электропередачи 110 кВ на основе данных по параметрам самих линий, что видно из полученных метрик качества ML-моделей. Однако из-за ограниченного набора данных (395 объектов) не удалось достичь постоянного результата (наблюдаются значительные расхождения между метриками качества на тестовой и валидационной выборках), поэтому в рамках будущего исследования необходимо расширить набор данных за счет включения линий других регионов и/или анализа дополнительных периодов фиксации отказов на ЛЭП.

### Список литературы / References

[1] Long L. Research on status information monitoring of power equipment based on Internet of Things, *Energy Reports*, 2022, 8, 281–286. DOI: 10.1016/j.egy.2022.01.018

[2] Sun B., Jing R., Zeng Y., Li Y., Chen J., Liang G. Distributed optimal dispatching method for smart distribution network considering effective interaction of source-network-load-storage flexible resources, *Energy Reports*, 2023, 9, 148–162. DOI: 10.1016/j.egy.2022.11.178

[3] Базан Т.В., Галабурда Я.В., Иселёнок Е.Б. Анализ отключений воздушных линий 35–750 кВ, Актуальные проблемы энергетики. *Электроэнергетические системы*. Минск, республика Беларусь: БНТУ, 2020, 114–116. [Bazan T.V., Galaburda Y.V., Iselenok E.B. Analysis of outages of 35–750 kV overhead lines. Aktual'nyye problemy energetiki. *Elektroenergeticheskiye sistemy*. Minsk, Republic of Belarus: BNTU, 2020, 114–116. (In Rus.)]

[4] Yang L., Teh J. Review on vulnerability analysis of power distribution network, *Electric Power Systems Research*, 2023, 224, 109741. DOI: 10.1016/j.epsr.2023.109741

[5] Shakiba F.M., Shojaee M., Azizi S.M., Zhou M. Real-Time Sensing and Fault Diagnosis for Transmission Lines, *International Journal of Network Dynamics and Intelligence*. 2022, 36–47. DOI: 10.53941/ijndi0101004

[6] Latka M., Hadaj P. Technical and statistical analysis of the failure of overhead lines and its impact on evaluating the quality of the power supply, *Proceeding of 2016 Progress in Applied Electrical Engineering, PAEE 2016*, Koscielisko-Zakopane, Poland, 2016. DOI: 10.1109/PAEE.2016.7605117

[7] Виноградов А.В., Васильев А.Н., Семенов А.Е., Синяков А.Н., Большев В.Е. Анализ времени перерывов в электроснабжении сельских потребителей и методы его сокращения за счет мониторинга технического состояния линий электропередачи, *Вестник ВИЭСХ*, 2017, 2, 3–11. EDN: ZEOGKZ [Vinogradov, A.V., Vasiliev, A.N., Semenov, A.E., Sinyakov, A.N., Bolshev, V.E. Analysis of the time of interruptions in the power supply of rural consumers and methods for reducing it by monitoring the technical condition of power transmission lines, *Bulletin of VIESH*, 2017, 2, 3–11. EDN: ZEOGKZ (In Rus.)]

[8] Ланин А.В., Полковская М.Н., Якупов А.А. Статистический анализ аварийных отключений в электрических сетях 10 кВ, *Актуальные вопросы аграрной науки*, 2019, 30, 45–52.

EDN: ZBHPXF [Lanin A. V., Polkovskaya M. N., Yakupov A. A. Statistical analysis of emergency shutdowns in 10 kV electrical networks, *Aktual'nyye voprosy agrarnoy nauki*, 2019, 30, 45–52. EDN: ZBHPXF (In Rus.)]

[9] Сбитнев Е.А., Жужин М.С. Анализ аварийности сельских электрических сетей 0, 38 кВ Нижегородской энергосистемы, *Вестник НГИЭИ*, 2020, 11(114), 36–47. EDN: KEJHDX [Sbitnev E. A., Zhuzhin M. S. Analysis of accident rates of rural electrical networks 0.38 kV of the Nizhny Novgorod energy system, *Bulletin of NGIEI*, 2020, 11(114), 36–47. EDN: KEJHDX (In Rus.)]

[10] Sood S. Power Outage Prediction Using Machine Learning Technique, *2023 International Conference on Power Energy, Environment & Intelligent Control (PEEIC)*. IEEE, 2023, 78–80. DOI: 10.1109/PEEIC.59336.2023.10451753

[11] Eskandarpour R., Khodaei A. Leveraging accuracy-uncertainty tradeoff in SVM to achieve highly accurate outage predictions, *IEEE Transactions on Power Systems*, 2018, 33(1), 1139–1141. DOI: 10.1109/TPWRS.2017.2759061

[12] Gururajapathy S.S., Mokhlis H., Illias H. A., Abu Bakar A. H., Awalin L. J. Fault location in an unbalanced distribution system using support vector classification and regression analysis, *IEEE Transactions on Electrical and Electronic Engineering*, 2018, 13(2), 237–245. DOI: 10.1002/tee.22519

[13] Doostan M., Chowdhury B. H. Power distribution system equipment failure identification using machine learning algorithms, *Proceeding of IEEE Power and Energy Society General Meeting*, Chicago, USA, 2018, 1–5. DOI: 10.1109/PESGM.2017.8274109

[14] Warlyani P., Jain A., Thoke A. S., Patel R. N. Fault classification and faulty section identification in teed transmission circuits using ANN, *International Journal of Computer and Electrical Engineering*, 2011, 3(6), 807–811. DOI: 10.7763/IJCEE.2011.V3.424

[15] Alqudah M., Obradovic Z. Enhancing Weather-Related Outage Prediction and Precursor Discovery Through Attention-Based Multi-Level Modeling, *IEEE Access*, 2023, 11, 94840–94851. DOI: 10.1109/ACCESS.2023.3303110

[16] Li M., Hou H., Yu J., Geng H., Zhu L., Huang Y., Li X. Prediction of Power Outage Quantity of Distribution Network Users under Typhoon Disaster Based on Random Forest and Important Variables, *Mathematical Problems in Engineering*, 2021, 2021, 6682242. DOI: 10.1155/2021/6682242

[17] Taylor W. O., Cerrai D., Wanik D., Koukoula M., Anagnostou, E. N. Community power outage prediction modeling for the Eastern United, *Energy Reports*, 2023, 10, 4148–4169. DOI: 10.1016/j.egy.2023.10.073

[18] Hou H., Chen X., Li M., Zhu L., Huang Y., Yu J. Prediction of user outage under typhoon disaster based on multi-algorithm Stacking, *International Journal of Electrical Power & Energy Systems*, 2021, 131, 107123. DOI: 10.1016/j.ijepes.2021.107123

[19] Das S., Kankanala P., Pahwa A. Outage Estimation in Electric Power Distribution Systems Using a Neural Network Ensemble, *Energies*, 2021, 14(16), 4797. DOI: 10.3390/en14164797

[20] Большев В.Е., Виноградова А.В. Анализ влияния параметров ЛЭП 110 кВ на вероятность их отказов, *Журнал СФУ. Техника и технологии*, 2024 (Принято в печать) [Bolshev V. E., Vinogradova A. V. Analysis of the influence of 110 kV power transmission line parameters on the probability of their failures, *SFU Journal. Engineering and Technology*, 2024 (Accepted for publication)]

[21] Jang H.S., Bae K.Y., Park H.S., Sung, D.K. Solar Power Prediction Based on Satellite Images and Support Vector Machine. *IEEE Trans Sustain Energy*, 2016, 7(3), 1255–1263. DOI: 10.1109/TSTE.2016.2535466

[22] Das A. L Logistic Regression, *Encyclopedia of Quality of Life and Well-Being Research*. 2021, 3985–3986. DOI: 10.1007/978–3–031–17299–1

[23] Jogunuri S., Josh F. T., Stonier A. A., Peter G., Jayaraj J., Ganji, V. Random Forest machine learning algorithm based seasonal multi-step ahead short-term solar photovoltaic power output forecasting. *IET Renewable Power Generation*, 2024, 00, 1–16. <https://doi.org/10.1049/rpg2.12921>

[24] Villegas-Mier C.G., Rodriguez-Resendiz J., Álvarez-Alvarado J.M., Jiménez-Hernández H., Odry, Á., Optimized Random Forest for Solar Radiation Prediction Using Sunshine Hours. *Micromachines*, 2022, 13(9), 1406. DOI: 10.3390/mi13091406

[25] Салахутдинова К.И., Лебедев И. С., Кривцова И. Е. Алгоритм градиентного бустинга деревьев решений в задаче идентификации программного обеспечения, *Научно-технический вестник информационных технологий, механики и оптики*, 2018, 18(6), 1016–1022. EDN: YQZIE X [Salakhutdinova K. I., Lebedev I. S., Krivtsova I. E. Gradient boosting algorithm for decision trees in the problem of software identification, *Scientific and Technical Bulletin of Information Technologies, Mechanics and Optics*, 2018, 18(6), 1016–1022. EDN: YQZIE X]

[26] Hancock J.T., Khoshgoftaar T.M. CatBoost for big data: an interdisciplinary review, *Journal of Big Data*, 2020, 7(1), 1–45. DOI: 10.1186/s40537–020–00369–8

[27] Nusinovici S., Tham Y.C., Yan M.Y.C., Ting D.S.W., Li J., Sabanayagam C., Wong T.Y., Cheng C. Y. Logistic regression was as good as machine learning for predicting major chronic diseases, *Journal of Clinical Epidemiology*, 2020, 122, 56–69. DOI: 10.1016/j.jclinepi.2020.03.002.

[28] Chokr B., Chatti N., Charki A., Lemenand T., Hammoud M. Feature extraction-reduction and machine learning for fault diagnosis in PV panels, *Solar Energy*, 2023, 262, 111918. DOI: 10.1016/j.solener.2023.111918.