

УДК 004.855.5

DOI: 10.18413/2518-1092-2024-9-1-0-7

Коржавых В.В.

**СРАВНЕНИЕ ЭФФЕКТИВНОСТИ АЛГОРИТМОВ  
МАШИННОГО ОБУЧЕНИЯ НА ПРИМЕРЕ  
ПРОГНОЗИРОВАНИЯ СРЕДНЕМЕСЯЧНОГО ПОТРЕБЛЕНИЯ  
ЭЛЕКТРОЭНЕРГИИ ИНТЕРВАЛЬНЫХ ПРИБОРОВ УЧЕТА  
ПОТРЕБИТЕЛЕЙ**

Филиал Публичного акционерного общества «Россети Центр» – «Белгородэнерго»,  
Валуйский район электрических сетей,  
ул. Суржикова, 114, г. Валуйки, Белгородская обл., 309990, Россия

*e-mail: Korzhavyh.VV@mrsk-1.ru*

**Аннотация**

Поиск и снижение потерь электроэнергии – одно из ключевых направлений деятельности сетевых организаций для улучшения финансовых результатов. Прогнозирование потребления электроэнергии на основе большого количества критериев и сравнение с фактическими данными является преимущественным способом обнаружения потерь. Однако, данный процесс требует высокой доли автоматизации. Поэтому, для решения этой задачи в настоящей работе рассмотрено применение трех алгоритмов машинного обучения, а также выполнено сравнение их эффективности. Автором сформирована обучающая выборка из базы данных Валуйского района электрических сетей на основе данных приборов учета, входящих в систему АИИСКУЭ, а также проведены эксперименты по реализации на ней следующий алгоритмов: k-ближайших соседей, линейной регрессии и случайного леса. Для сравнения полученных моделей автором были использованы такие показатели эффективности как среднеквадратичная ошибка (MSE), абсолютная средняя ошибка (MAE) и коэффициент детерминации ( $R^2$ ). Результаты эксперимента показали наибольшую эффективность метода случайного леса в сравнении с остальными рассматриваемыми алгоритмами.

**Ключевые слова:** машинное обучение; потери электроэнергии; алгоритм k-ближайших соседей; линейная регрессия; случайный лес; среднеквадратичная ошибка; средняя абсолютная ошибка; коэффициент детерминации

**Для цитирования:** Коржавых В.В. Сравнение эффективности алгоритмов машинного обучения на примере прогнозирования среднемесячного потребления электроэнергии интервальных приборов учета потребителей // Научный результат. Информационные технологии. – Т.9, №1, 2024. – С. 58-73. DOI: 10.18413/2518-1092-2024-9-1-0-7

Korzhavykh V.V.

**COMPARISON OF THE EFFICIENCY OF MACHINE LEARNING  
ALGORITHMS BY THE EXAMPLE OF FORECASTING  
THE AVERAGE ELECTRICITY CONSUMPTION  
OF INTEGRATED CONSUMER METERING DEVICES**

Branch of the Public Joint Stock Company “Rosseti Center” – “Belgorodenergo”,  
Valuysky district of electrical networks,  
114 Surzhikova st., Valuiki, Belgorod region, 309990, Russia

*e-mail: Korzhavyh.VV@mrsk-1.ru*

**Abstract**

Finding and reducing electricity losses is one of the key activities of network organizations to improve financial results. Forecasting based on a large number of criteria and comparing with actual electricity consumption is the preferred way to detect losses. However, this process requires a high degree of automation. Therefore, to solve this problem, this paper considers the use of three machine learning algorithms, as well as a comparison of their effectiveness. The author formed a

training sample from the database of one of the districts of electrical networks, and also conducted experiments on the implementation of the following algorithms on it: k-nearest neighbors, linear regression and random forest. To compare the resulting models, the author used such performance indicators as mean square error (MSE), absolute mean error (MAE) and coefficient of determination ( $R^2$ ). The results of the experiment showed the greatest efficiency of the random forest method in comparison with other considered algorithms.

**Keywords:** machine learning; power loss; k-nearest neighbors' algorithm; linear regression; random forest; mean square error; mean absolute error; coefficient of determination

**For citation:** Korzhavykh V.V. Comparison of the efficiency of machine learning algorithms by the example of forecasting the average electricity consumption of integrated consumer metering devices // Research result. Information technologies. – Т.9, №1, 2024. – P. 58-73. DOI: 10.18413/2518-1092-2024-9-1-0-7

## ВВЕДЕНИЕ

Фактические (отчетные) потери электроэнергии – разность между электроэнергией, поступившей в сеть, и электроэнергией, отпущенной потребителям, определяемая по данным системы учета поступления и полезного отпуска электроэнергии [5]. Фактические потери разделяются на техническую и коммерческую составляющую. Снижение технической составляющей – модернизация оборудования, а коммерческой – повышение “платежной дисциплины” контрагентов, пресечение “хищений” электроэнергии посредством анализа и контроля за потреблением. Решения задачи поиска и оптимизации потерь электрической энергии является наиболее актуальной проблемой не только в России, но и во всем мире. Статистика Международного энергетического агентства (International Energy Agency, IEA) говорит о том, что страны с развитой экономикой имеют уровень потерь не более 8–10%, а развивающиеся страны – от 10% до 30% [9]. По данным ежегодного исследования Emerging Markets Smart Grid: Outlook, проводимого компанией Northeast Group, LLC, 30% от \$89 млрд потерь приходилось на три страны: Индия (\$16,2 млрд), Бразилия (\$10,5 млрд) и Россия (\$5,1 млрд) [10]. Если снижение технической составляющие потерь – это дорогостоящий процесс, требующий значительных инвестиций и дальний горизонт окупаемости, то работа над коммерческими потерями в большей части сводится к анализу различных критериев потребления электроэнергии из баз данных сетевой организации и формированию адресных списков для проведения проверок системы учета, снятия контрольных показаний и т.д. Качественное проведение анализа и непосредственной работы приводит к существенному снижению коммерческой составляющей потерь. Однако особенности законодательства Российской Федерации в отрасли электроэнергетики, ограниченное количество человеческих ресурсов сетевых компаний и растущее число аналитических критериев способствует поиску новых подходов и механизмов для решения вышеуказанной задачи. Внедрение информационных технологий - процессов, использующих совокупность средств и методов сбора, обработки, накопления и передачи данных (первичной информации) для получения информации нового качества о состоянии объекта, процесса, явления, информационного продукта [6] становится необходимостью в современных условиях работы. Часто анализ потребления электроэнергии сводится к установлению объемов в кВтч (киловатт-часах) и сравнению его с текущим потреблением. Другими словами – составление прогноза потребления электроэнергии. Технологией, успешно справляющийся с решением такого рода задач является подраздел искусственного интеллекта, называющийся машинным обучением. Смысл машинного обучения состоит в использовании нужных признаков для построения моделей, подходящих для решения правильно поставленных задач. Признаки определяют “язык”, на котором описываются объекты предметной области. Задача – это абстрактное представление проблемы с участием объектов предметной области, которую необходимо решить. Модель – результат машинного обучения, примененного к обучающим данным. Можно сказать, что модели обеспечивают разнообразие предмета машинного обучения, тогда как задачи и признаки придают ему единство [15]. Методы обучения можно разделить на обучение с учителем и без учителя. Обучение без учителя подходит для разделения данных на группы имеющие схожие характеристики или свойства (задача

кластеризации). Обучение с учителем предполагает какой-либо обучающий набор размеченных данных, например среднемесячное потребление приборов учета входящих в АИИСКУЭ к площади помещения домовладения. На его основе формируется прогноз потребления электроэнергии. (задача регрессии). Полученную модель можно использовать для прогноза потребления приборов учета, не входящих в АИИСКУЭ, для выявления наибольшей разницы прогнозных-фактических значений среднемесячного потребления, по результатам которых бригады направляются для проведения проверок по адресной части. Разумеется, модель, основанная только на одном критерии, не сможет претендовать на высокую точность прогноза, поэтому для ее формирования нужно большее их количество. Для оценки качества модели и сравнения ее с другими моделями машинного обучения используются специальные метрики. В рамках данного исследования выбраны такие метрики как среднеквадратичная ошибка (MSE), формула которой приведена ниже:

$$MSE(y_i, \hat{y}_i) = \frac{1}{N} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

где,  $y_i$  – фактическое значение переменной,  $\hat{y}_i$  – прогнозируемое значение переменной,  $n$  – размер выборки данных. Если не возводить в квадрат разницу между прогнозируемым и фактическим значением переменной, а взять ее по модулю, то получим метрику абсолютной средней ошибки (MAE):

$$MAE(y_i, \hat{y}_i) = \frac{1}{N} \sum_{i=1}^n |y_i - \hat{y}_i|$$

MAE является более наглядной из-за линейности шкалы оценки. В целом, данные виды метрик зависимы от шкалы измерений [18] и обладают чувствительностью к выбросам данных [19]. Таким образом, такой способ подготовки данных перед моделированием как нормализация, является обязательным. Для оценки “адекватности” модели, соответствии прогнозных данных модели фактическим, используется коэффициент детерминации:

$$R^2 = 1 - \frac{(RMSE)^2}{D}$$

где,  $RMSE$ – метрика, равная квадратному корню из MSE,  $D$ – дисперсия. После раскрытия формул:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \langle y \rangle)^2}$$

где,  $\langle y \rangle$ – среднее, прогнозируемое значение.

Подытожив все вышесказанное, стало возможным определить цель данного исследования – подготовить набор данных сетевой организации, описать и применить к ней алгоритмы машинного обучения с учителем, сравнить полученные результаты с помощью метрик качества MSE, MAE и коэффициента детерминации.

### **ИССЛЕДУЕМЫЕ ДАННЫЕ**

В качестве исходных взяты данные из базы SAP RT2 Валуйского района электрических сетей о приборах учета потребителей – физических лиц. Формат данных – файл MS Excel расширения *xlsx*. Внешний представлен на рисунке 1.

Договор	Статус "сезонного потребителя"	Место прибора	Количество проживающих	Наличие электроотопительной установки	Наличие электроводонагревателя	Количество комнат	Класс точности	Фазность	Разрядность	Максимальная мощность точки поставки	Площадь	Количество КП	Среднее 2019	Среднее 2020	Среднее 2021	Среднее 2022
486226	Не определен	в квартире (доме)	2	0	0	4	1	1	5,1	3,000	90	9	66	64	61	18
486242	Не определен	в квартире (доме)	4	0	0	4	2	1	5,1	3,000	89	12	143	192	300	243
486251	Не определен	в квартире (доме)	2	0	0	4	2	1	5,1	3,000	58	10	183	177	181	175
486264	Не определен	в квартире (доме)	2	0	0	3	1	1	5,1	3,000	42	10	115	93	189	206
486275	Дачники	в квартире (доме)	1	0	0	4	1	1	6,1	3,000	72	9	45	62	99	82
486296	Не определен	на фасаде	5	0	0	5	1	1	5,1	3,000	88	9	99	84	186	204
486309	Не определен	в квартире (доме)	1	0	0	3	1	1	6,1	3,000	37	10	94	99	82	97
486321	Не определен	в квартире (доме)	4	0	0	4	1	1	5,1	3,000	103	9	58	74	399	381
486326	Не определен	в квартире (доме)	1	0	0	4	1	1	5,1	3,000	49	9	99	94	73	59
486356	Не определен	в квартире (доме)	5	0	0	4	1	1	5,1	3,000	100	9	59	112	259	188
486363	Не определен	в квартире (доме)	3	0	0	4	1	1	6,1	3,000	56	10	421	471	212	210
486423	Не определен	в квартире (доме)	3	0	0	4	2	1	5,1	3,000	56	9	224	273	307	234
486431	Не жилой	в хозпостройке (гараже)	2	0	0	6	1	3	6,0	0,000	250	11	372	404	1940	1713
486526	Хоз. постр.	в квартире (доме)	2	0	0	4	1	3	5,1	3,000	92	10	91	88	94	98
486553	Не определен	в квартире (доме)	4	0	0	4	1	1	5,1	3,000	56	12	214	196	230	204
486579	Не определен	в хозпостройке (гараже)	4	0	0	4	1	3	5,1	3,000	67,4	11	215	465	223	157

Рис. 1. Набор исследуемых данных

Fig. 1. Data set under study

Критерий Договор является уникальным идентификатором для последующего объединения прогноза с адресной частью контрагента. Статус сезонности потребителя является категориальной переменной, определяющей статус домовладения. К примеру, статус «Хоз. постр.» говорит о том, что домовладение является гаражом, а не жилым домом. Место прибора – определяет физическое местоположение прибора учета. Наличие электроотопительной установки и электроводонагревателя являются булевыми переменными. Остальные критерии достаточно наглядны. Первоначальный анализ набора данных представлен в таблице 1.

Таблица 1

Набор исследуемых данных

Table 1

Data set under study

№	Критерий	Тип данных	Диапазон значений
1	Статус "сезонного потребителя"	Текстовый	
2	Место прибора	Текстовый	
3	Количество проживающих	Числовой	0...9
4	Наличие электроотопительной установки	Булев	0,1
5	Наличие электроводонагревателя	Булев	0,1
6	Количество комнат	Числовой	0...96
7	Класс точности	Числовой	0...2,5
8	Фазность	Булев	1,3
9	Разрядность	Числовой	0...7
10	Максимальная мощность точки поставки	Числовой	3...15
11	Площадь	Числовой	3..327
12	Количество КП	Числовой	0...26
13	Среднее 2019	Числовой	0...2452
14	Среднее 2020	Числовой	0...1940
15	Среднее 2021	Числовой	0...3682
16	Среднее 2022	Числовой	0...1713

Так как критерии 1,2 имеют текстовый тип данных, то необходимо категориальное преобразование. Для данного исследования применено фиктивное кодирование, представляющее собой метод, используемый для преобразования категориальных переменных в числовые значения путем создания двоичных столбцов для каждой категории [13]. Критерий 6 вызывает сомнения ввиду максимального количества комнат. Его следует проверить на предмет выбросов данных и удалить ошибочные. Критерии 13-16 необходимо проверить на мультиколлинеарность и удалить

коррелирующие критерии, в противном случае это снизит точность прогноза. Для всех числовых данных в последующем необходимо применить нормализацию из-за большого разброса значений. Всего для моделирования выбраны 4000 точек учета. Там, где это необходимо обучающая и тестовая выборки разделены в соотношении 70 и 30% [8].

## **МАТЕРИАЛЫ И МЕТОДЫ ИССЛЕДОВАНИЯ**

В качестве алгоритмов машинного обучения выбраны три популярных и наиболее подходящих к исследуемой задаче. Это метод  $k$ - ближайших соседей, случайный лес и линейная регрессия.

### **1. Метод $k$ – ближайших соседей**

Алгоритм использует весь набор данных в качестве обучающей выборки, а не разделяет данные на набор данных для обучения и теста. Когда для нового набора данных требуется определить результат, алгоритм проходит весь набор данных, чтобы найти  $k$ -ближайших соседей для нового экземпляра, то есть  $k$  экземпляров, наиболее похожих на новую точку, а затем решает, к какой группе эта точка относится. Сходство между экземплярами рассчитывается с использованием таких мер, как евклидово расстояние и расстояние Хемминга [7].

### **2. Случайный лес**

Данный алгоритм может быть использован для задач классификации, регрессии и кластеризации. Он основан на концепции обучения ансамбля, которая представляет собой процесс объединения нескольких классификаторов для решения сложной задачи и повышения производительности модели. Как следует из названия, случайный лес — это классификатор, который содержит несколько деревьев решений в различных подмножествах заданного набора данных и использует среднее значение для повышения точности прогнозирования этого набора данных. Вместо того, чтобы полагаться на одно дерево решений, случайный лес берет прогноз от каждого дерева и основывается на большинстве голосов прогнозов, и далее предсказывает окончательный результат [20].

### **3. Линейная регрессия**

Линейная регрессия — это алгоритм машинного обучения, позволяющий аппроксимировать некую зависимость линейной функцией с использованием метода наименьших квадратов. Несмотря на кажущуюся простоту алгоритма, он является базовым алгоритмом в анализе данных, позволяющим грубо, но достоверно оценить тренды, характер взаимосвязей переменных. Кроме того, алгоритм развивается и обрастает дополнениями, получает новые приложения, так, например нечеткая линейная регрессия или полиномиальная регрессия, построенная на базе линейной, позволяет аппроксимировать нелинейные зависимости [11]. Визуализация работы алгоритма для зависимости двух переменных представлена на Рис. 2. Задача заключается в том, чтобы подобрать такую линейную функцию, которая бы наиболее точно, с минимальной ошибкой описывала существующую зависимость.

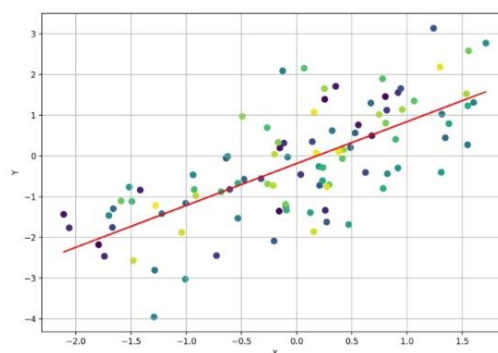


Рис. 2. Линейная регрессия  
Fig. 2. Linear regression

Программной средой для реализации вышеуказанных алгоритмов выбраны библиотеки sklearn, NumPy, pandas, seaborn языка программирования Python. Перед реализацией проведены мероприятия по подготовке данных. При помощи тепловой карты (heatmap) выполнен поиск корреляций в исследуемом наборе данных. Представленная на Рис. 3 карта говорит нам о высокой корреляции между критериями Среднее 2019-2022. Так как прогнозируемым параметром будет Среднее 2022, то необходимо удалить из набора данных Среднее за 2019 и 2020 год.

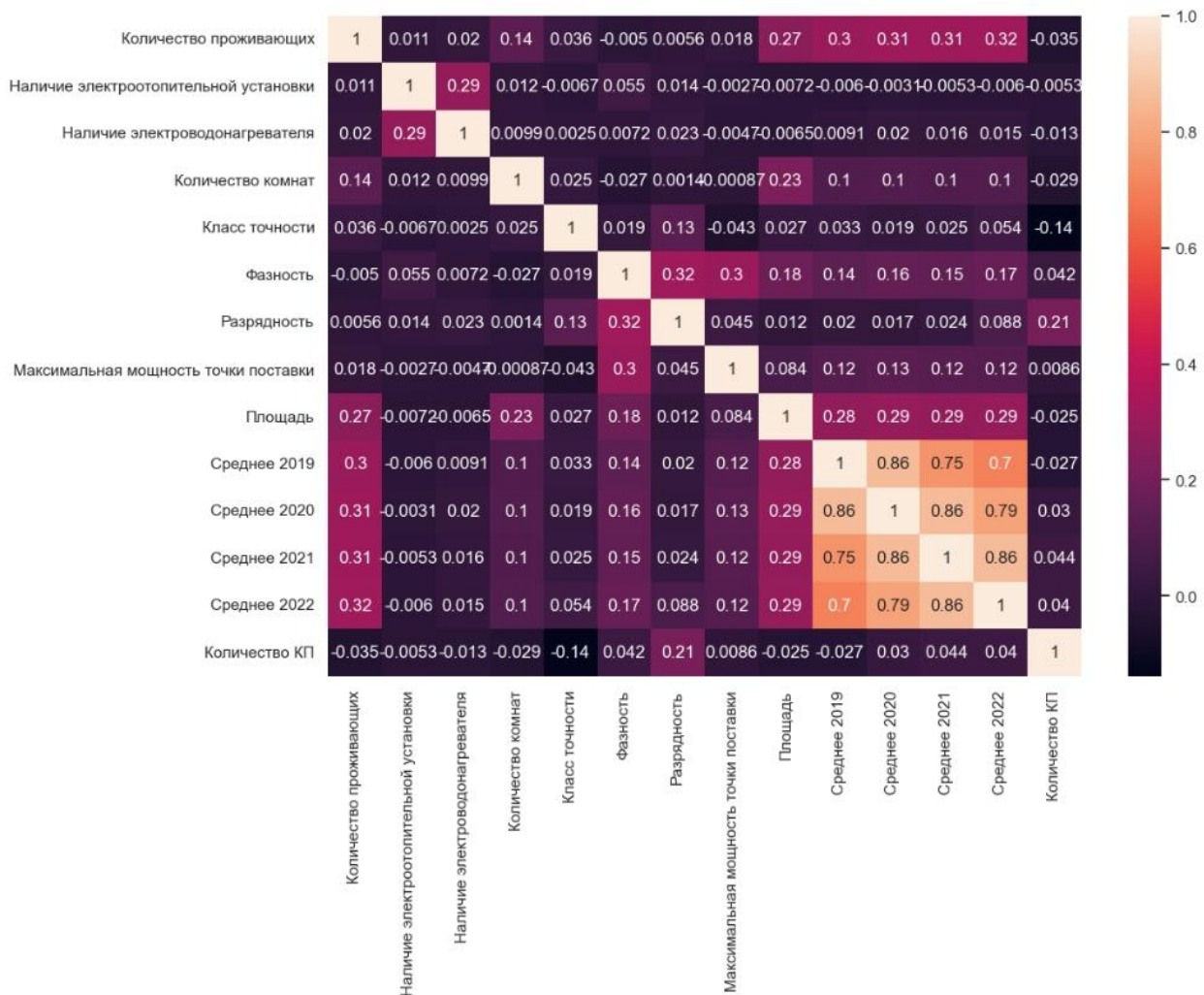


Рис. 3. Тепловая карта  
Fig. 3. Heatmap

Категориальные переменные закодированы с помощью добавления фиктивной переменной, пример по критерию “Место прибора” представлен в таблице 2.

Таблица 2

Кодирование категориальных переменных

Table 2

Coding of categorical variables

id	в доме	в хозпостройке	на опоре	на фасаде
0	0	0	1	0
1	0	0	0	1
2	1	0	0	0
3	0	0	1	0
4	0	1	0	0

Далее, для “выравнивания” данных выбрана минимакс нормализация – линейное преобразование данных в диапазоне [0...1], где минимальное и максимальное масштабируемые значения соответствуют 0 и 1 соответственно. Формальное представление:

$$X_n = \frac{X - X_{min}}{X_{max} - X_{min}}$$

где, X – текущее значение данных, Xmin – минимальное значение данных выборки, Xmax – максимально значение данных выборки. После проведенных выше мероприятий набор данных подготовлен для моделирования.

### РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

В качестве результатов моделирования представлены кривые показателей MSE, MAE, R2 в зависимости от: числа соседей для метода k-ближайших соседей, числа деревьев в лесу для метода “Случайный лес”, графики частотных зависимостей для всех методов для определения наиболее влияющих критериев, графики сравнения прогнозных-фактических данных для всех методов и итоговая таблица сравнения метрик по каждому методу.

#### 1. Метод k – ближайших соседей

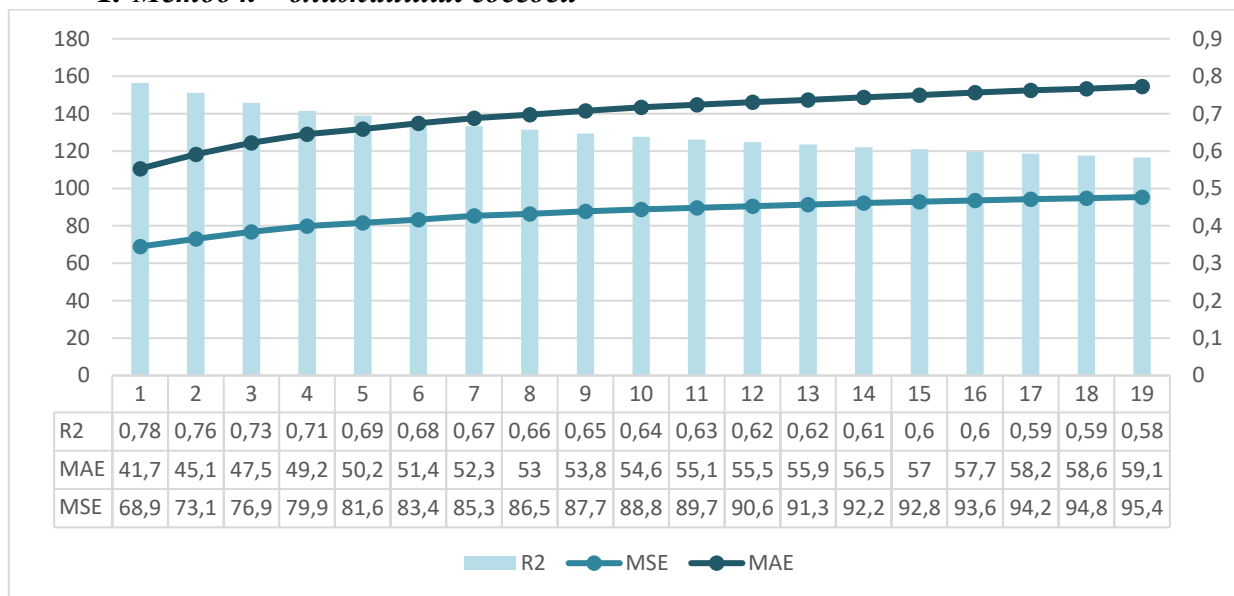


Рис. 4. Метрики качества модели в зависимости от числа соседей  
Fig. 4. Model quality metrics depending on the number of neighbors

Итерации производились в диапазоне от 2 до 20 соседей. Как видно из графика на Рис. 4, наибольшие показатели MSE, MAE и R2 достигаются при количестве соседей равном двум. Далее на Рис. 5 представлена степень корреляции между прогнозными и фактическими значениями потребления электроэнергии. Данные коррелируют на 88%.

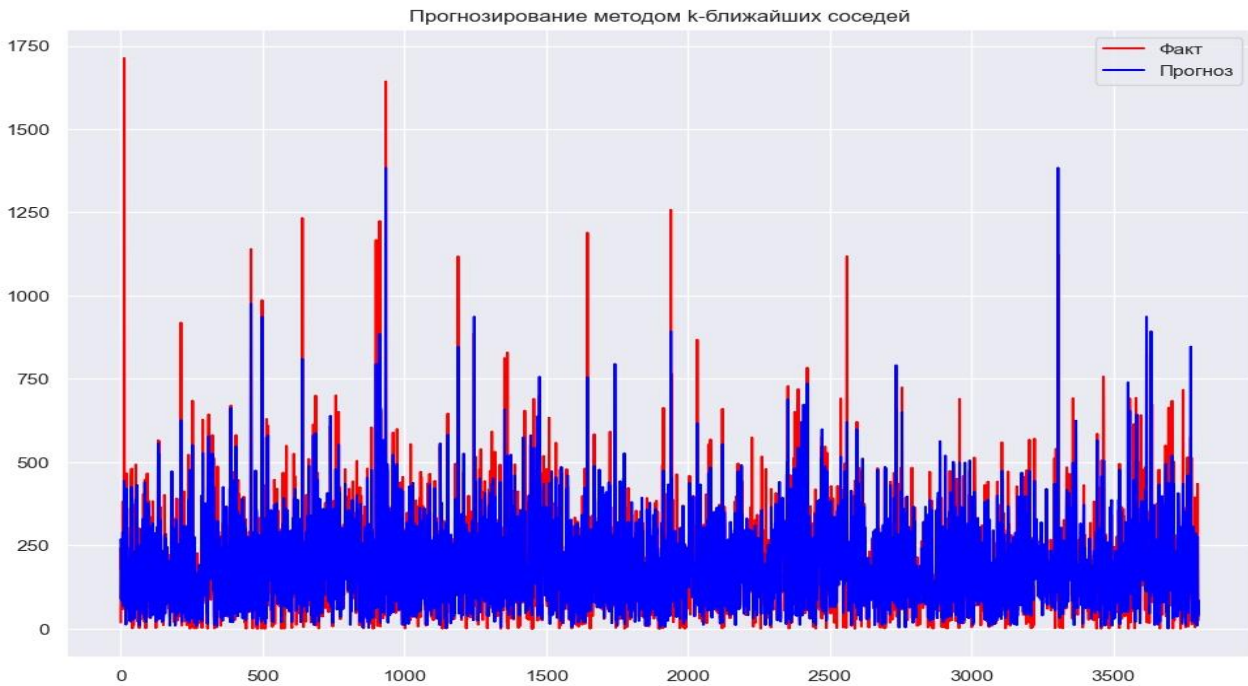
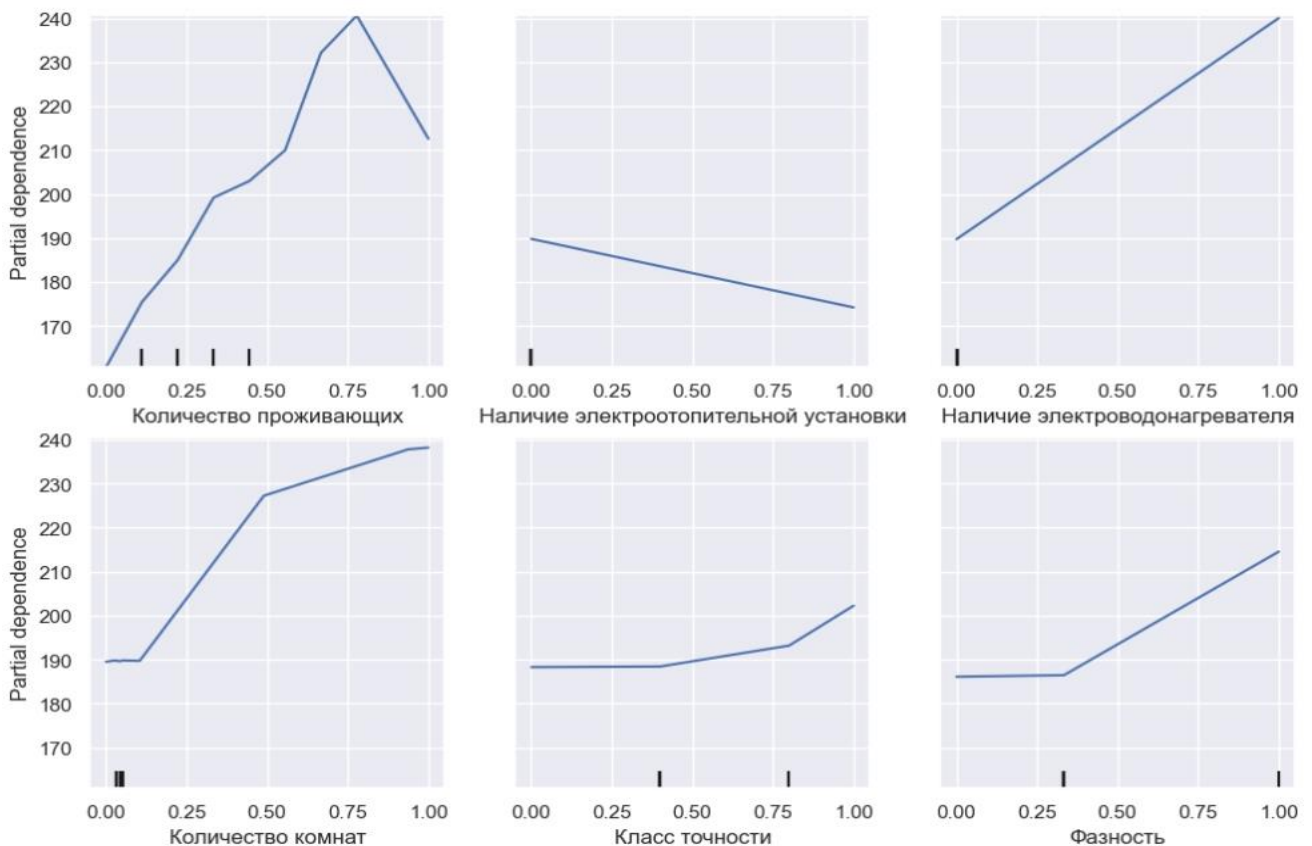


Рис. 5. Корреляция прогнозных и фактических данных  
Fig. 5. Correlation of forecast and actual data

Рассмотрим какие критерии из набора данных наиболее влияют на модель. Представлены только первые 12 критериев, так как остальные получены с помощью добавления фиктивной переменной





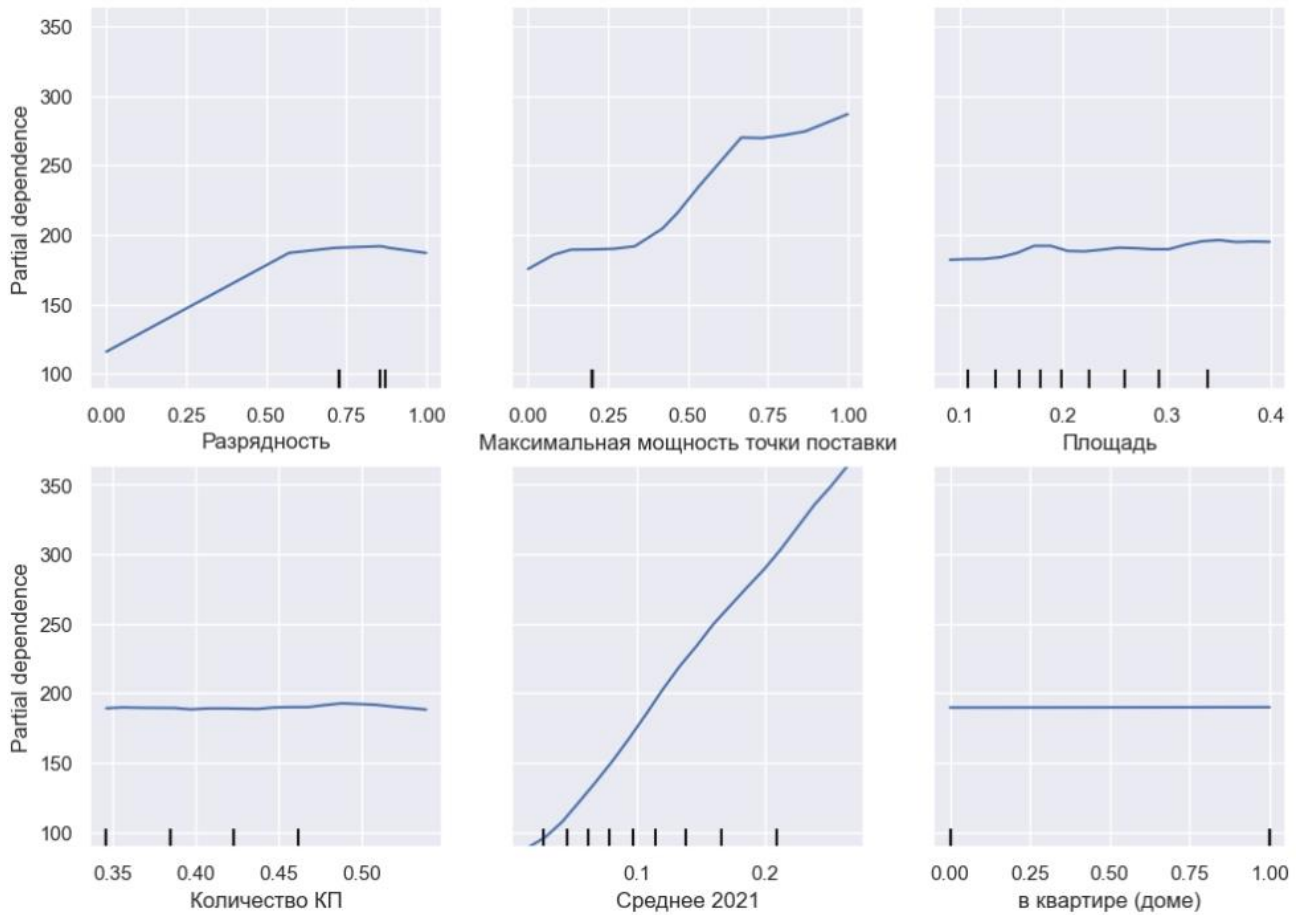


Рис. 6. Степень влияния критериев из набора данных  
Fig. 6. The degree of influence of criteria from the data set

Наиболее влияющими критериями, согласно Рис. 6 являются: наличие электроводонагревателя, количество комнат и проживающих, максимальная мощность точки поставки и среднее 2021.

## 2. Линейная регрессия

Степень корреляции между фактическими и прогнозными значениями составила 87% (Рис.6).



Рис. 7. Корреляция прогнозных и фактических данных  
Fig. 7. Correlation of forecast and actual data

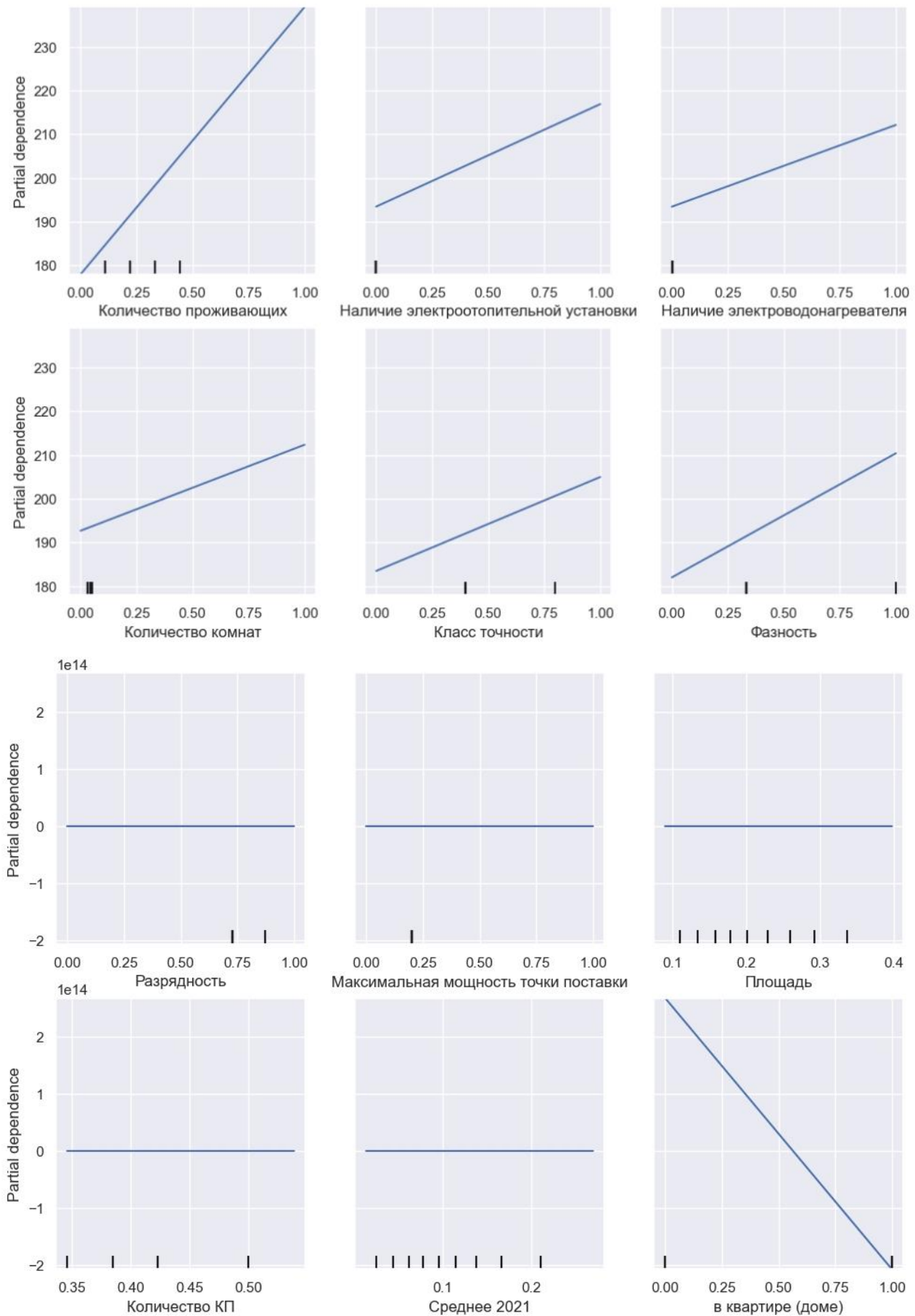


Рис. 8. Степень влияния критериев из набора данных  
Fig. 8. The degree of influence of criteria from the data set

Наиболее влияющими критериями, согласно Рис. 9 являются: количество комнат и проживающих, фазность, класс точности, наличие электроотопления и электроводонагрева.

### 3. Случайный лес

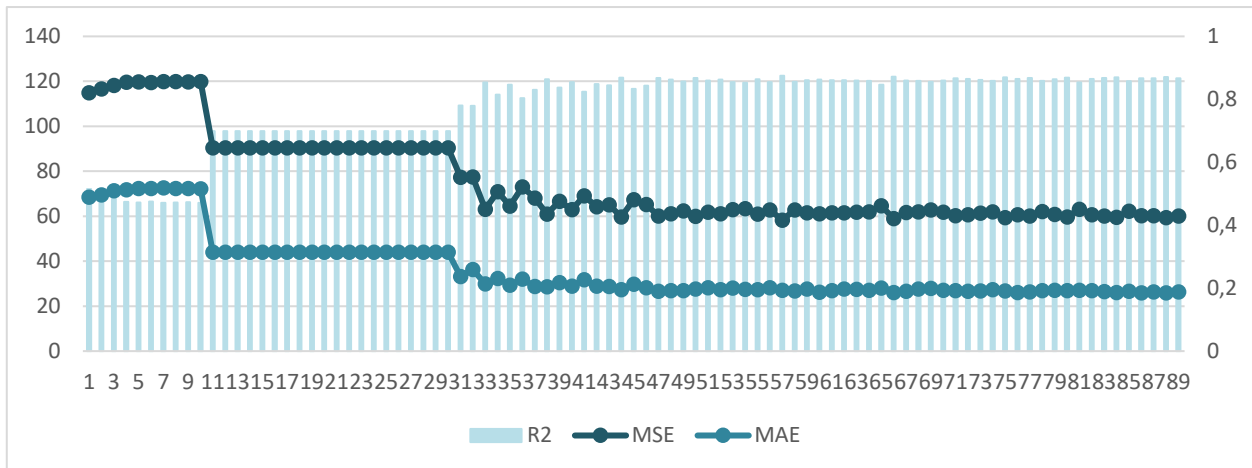


Рис. 9. Метрики качества модели в зависимости от числа деревьев  
Fig. 9. Model quality metrics depending on the number of trees

Итерации производились в диапазоне от 10 до 1000 деревьев. На графике из Рис. 9 представлены только первые 89 итераций. Модель демонстрирует наилучшие метрики при 258 деревьях. Степень корреляции 95%. (Рис. 10).

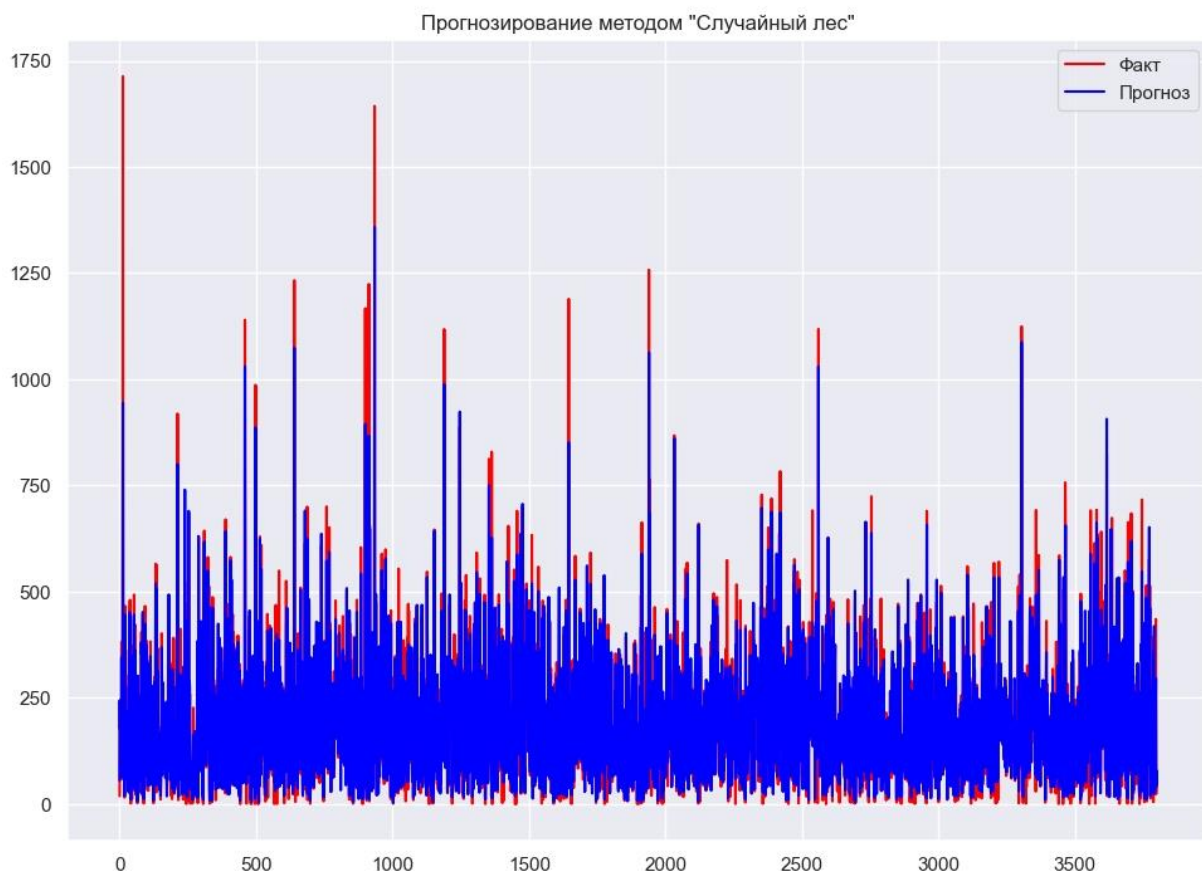


Рис. 10. Корреляция прогнозных и фактических данных  
Fig. 10. Correlation of forecast and actual data

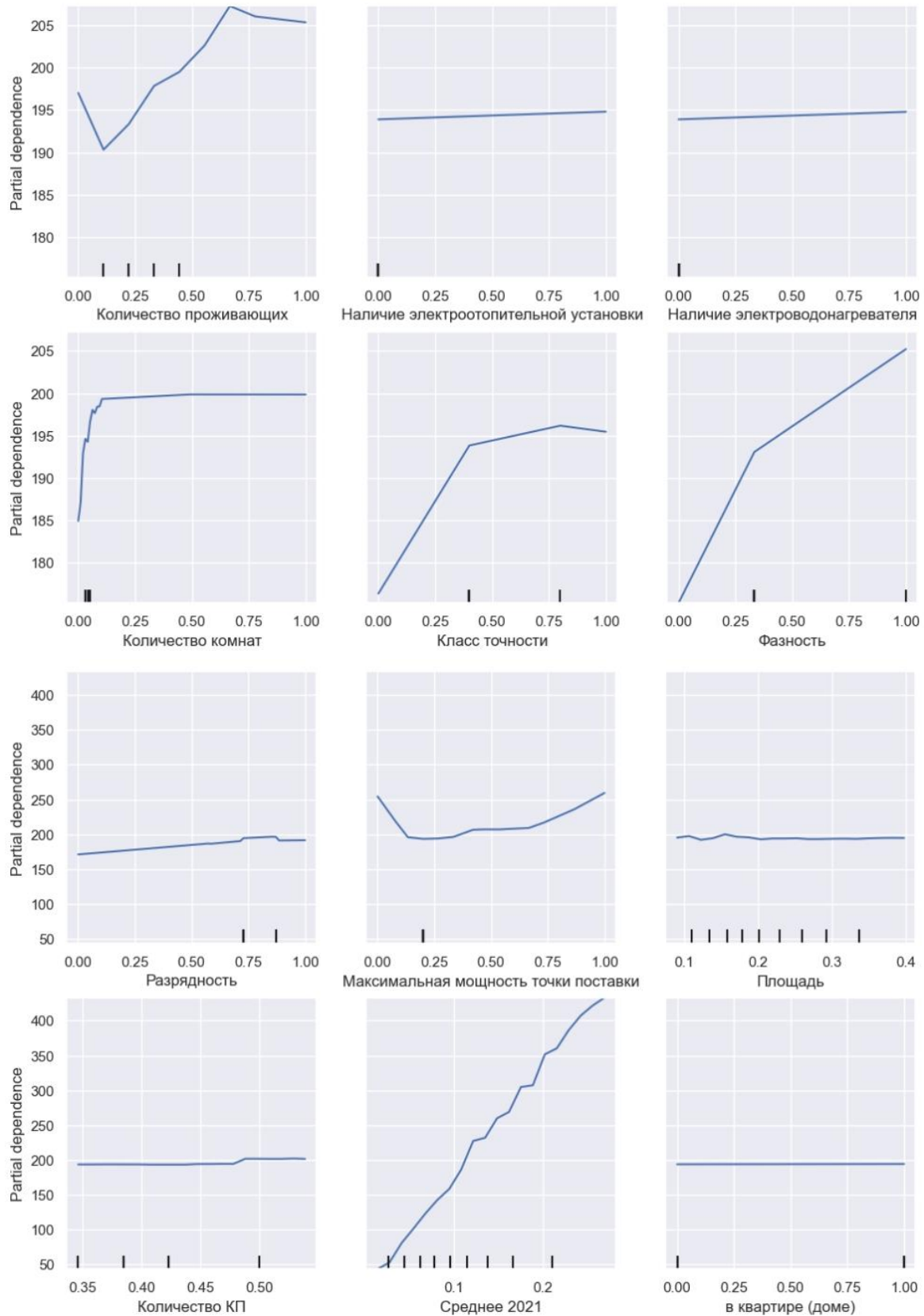


Рис. 11. Степень влияния критериев из набора данных  
Fig. 11. The degree of influence of criteria from the data set

Наиболее влияющими критериями, согласно Рис. 10 являются: количество комнат и проживающих, фазность, класс точности, наличие электроотопления и электроводонагрева и среднее 2021. Итоговая таблица сравнения моделей:

Таблица 3

Сравнение метрик качества моделей

Table 3

Comparison of model quality metrics

Метрика	к-ближайших соседей	Линейная регрессия	Случайный лес
<b>MSE</b>	68,921	70,902	42,923
<b>MAE</b>	41,662	37,687	21,586
<b>R2</b>	0,782	0,769	0,915

Таким образом, наилучшие показатели у алгоритма Случайный лес, алгоритмы к-ближайших соседей и Линейной регрессии показали практически одинаковые результаты.

### ЗАКЛЮЧЕНИЕ

В представленной работе поднят вопрос о применении алгоритмов машинного обучения к решению проблем поиска потерь электроэнергии. Обучив модель на данных приборов учета, включенных в АИИСКУЭ, можно проверить на отклонения в потреблении контрагентов, не имеющих данные интеллектуальные приборы учета. Для решения этой задачи изучена база данных организации проведена подготовка данных для моделирования. Для моделей использовались три наиболее популярных алгоритма. Наилучшим образом зарекомендовал себя алгоритм “Случайный лес”. Причиной лучшей работы данного алгоритма является отсутствие экстраполяции, то есть выхода за пределы обучающей выборки.

Помимо перечисленных, существует множество других алгоритмов машинного обучения, это ответвление искусственного интеллекта постоянно развивается и разрабатывает все более эффективные методики.

### Список литературы

1. Бокс Дж. Анализ временных рядов. Прогноз и управление / Бокс Дж, Дженкинс Г. М.: Мир, Вып.1, 1974. – 406 с.
2. Гаврилова Т.А. Базы знаний интеллектуальных систем. Учебник / Гаврилова Т.А., Хорошевский В.Ф. — СПб.: Питер, 2000. – 384 с.
3. Галушкин А.И. Нейроматематика (проблемы развития) / М.: Радиотехника, 2003.40с.
4. Донской, Д. А. Применение аналитических технологий в системах управления и информатике/ Донской Д.А., Слепцов Н.В., Щербаков М.А.– Пенза, 2005.
5. Железко, Ю.С. Расчет, анализ и нормирование потерь электроэнергии в электрических сетях / Ю.С. Железко. // М.: НУ ЭНАС, 2002. – 280с.
6. Иванов В.Л. Электронный учебник: системы контроля знаний // Информатика и образование. – 2002. – № 1.
7. Казанская А.А. Использование машинного обучения в инвестиционной деятельности / А.А. Казанская, Л.Г. Мишура // Научный журнал НИУ ИТМО. Серия: Экономика и экологический менеджмент. – 2020. – № 2. – С. 23-34. – DOI 10.17586/2310-1172-2020-13-2-23-34. – EDN MUJXYZ.
8. Кафтанников, И.Л. Проблемы формирования обучающей выборки в задачах машинного обучения / И.Л. Кафтанников, А.В. Парасич // Вестник ЮУрГУ. Серия «Компьютерные технологии, управление, радиоэлектроника». – 2016. – Т. 16, № 3. – С. 15–24. DOI: 10.14529/ctcr160302

9. Кудашев К., Коммерческие потери электроэнергии без границ, 2017 г. URL: <http://www.bigpowernews.ru/interview/document76022.phtml> (дата обращения: 23.11.2023)
10. Найти утечку, 2021 г. URL: <https://www.kommersant.ru/doc/4877601> (дата обращения: 23.11.2023)
11. Нечеткая линейная регрессия в задачах оценки / Е.В. Вишнякова, Е.В. Иванова, С.М. Камалов [и др.] // Научные записки молодых исследователей. – 2015. – № 5. – С. 14-29.
12. Джонс Т. Программирование искусственного интеллекта в приложениях / Пер. с англ. Осипов А.И. – М.: ДМК Пресс, 2011. – 312 с.
13. Тоуди Т. Преобразование категориальных данных: Практическое руководство по обработке нечисловых переменных для алгоритмов машинного обучения, 2023 г. URL: <https://dev-gang.ru/article/preobrazovanie-kategorialnyh-dannyh-prakticheskoe-rukovodstvo-po-obrabotke-necislovyh-peremennyh-dlja-algoritmov-mashinnogo-obuczenija-buyh1q4ttt/>
14. Трикоз Д.В. Нейронные сети: как это делается? Компьютеры + программы N 4(5). – 1993. – С. 14-20.
15. Флах П. Машинное обучение / П. Флах // М.: ДМК Пресс, 2015. с. 25.
16. Хайкин С. Нейронные сети: полный курс / С. Хайкин. – М.: Диалектика, 2019. – 1104 с.
17. Andrzej C. Neural Networks for Optimization and Signal Processing [Текст] / C. Andrzej, R. Unbehauen, J. Wiley and Sons Ltd, 1993. – 526 с.
18. Hyndman R.J., Koehler A.B. Another look at measures of forecast accuracy // International Journal of Forecasting. –2006. – № 22(4). – P. 679-688.
19. Shcherbakov M.V., Brebels A. Outliers and anomalies detection based on neural networks forecast procedure: Proceedings of the 31st Annual International Symposium on Forecasting (ISF 2011) / Prague: International Institute of Forecasters, 2011. – pp. 21-22. URL: [http://www.forecasters.org/isf/pdfs/ISF11\\_Proceedings.pdf](http://www.forecasters.org/isf/pdfs/ISF11_Proceedings.pdf)
20. Yu, Chong Ho. Exploratory data analysis in the context of data mining and resampling // International Journal of Psychological Research. 3. 2010.

### References

1. Box J. Time series analysis. Forecasting and management / Box J, Jenkins G. M.: Mir, V.1, 1974. – 406 p.
2. Gavrilova T.A. Knowledge bases of intellectual systems. Textbook / Gavrilova T.A., Khoroshevsky V.F. – SPb.: Piter, 2000. – 384 p.
3. Galushkin A.I. Neuromathematics (problems of development) / М.: Radiotekhnika, 2003. 40 p.
4. Donskoy D.A. Application of analytical technologies in control systems and informatics / Donskoy D.A., Sleptsov N.V., Shcherbakov M.A. – Penza, 2005.
5. Zhelezko Yu.S. Calculation, analysis and rationing of the electric power losses in the electric networks / Yu.S. Zhelezko // М.: NU ENAS, 2002. – 280 p.
6. Ivanov V.L. Electronic textbook: knowledge control systems (in Russian) // Informatics and Education. – 2002. – № 1.
7. Kazanskaya A.A. The use of machine learning in investment activity / A.A. Kazanskaya, L.G. Mishura // Scientific Journal of NIU ITMO. Series: Economics and Environmental Management. – 2020. – № 2. – P. 23-34. – DOI: 10.17586/2310-1172-2020-13-2-23-34. – EDN MUJXYZ.
8. Kaftannikov I.L. Problems of training sample formation in machine learning tasks / I.L. Kaftannikov, A.V. Parasich // Vestnik SUSU. Series "Computer technologies, management, radio electronics". – 2016. – Т. 16, № 3. – P. 15-24. DOI: 10.14529/ctcr160302
9. Kudashev K. Commercial electricity losses without borders, 2017. URL: <http://www.bigpowernews.ru/interview/document76022.phtml> (date of reference: 23.11.2023)
10. Find Leakage, 2021 URL: <https://www.kommersant.ru/doc/4877601> (date access: 23.11.2023)
11. Fuzzy linear regression in estimation problems / E.V. Vishnyakova, E.V. Ivanova, S.M. Kamalov [et al.] // Scientific Notes of Young Researchers. – 2015. – № 5. – P. 14-29.
12. Jones T. Programming of Artificial Intelligence in Applications / Per. from Engl. Osipov A.I. – М.: ДМК Пресс, 2011. – 312 p.
13. Toady T. Transforming categorical data: A practical guide to handling non-numeric variables for machine learning algorithms, 2023 URL: <https://dev-gang.ru/article/preobrazovanie-kategorialnyh-dannyh-prakticheskoe-rukovodstvo-po-obrabotke-necislovyh-peremennyh-dlja-algoritmov-mashinnogo-obuczenija-buyh1q4ttt/>
14. Tricoz D.V. Neural networks: how to do it? Computers + Programs N 4(5). 1993. – 14-20 p.

15. Flach P. Machine learning / P. Flach // М.: DMK Press, 2015. p. 25
16. Haykin S. Neural networks: a complete course / S. Haykin. – М.: Dialectics, 2019. – 1104 p.
17. Andrzej C. Neural Networks for Optimization and Signal Processing [Text] / C. Andrzej, R. Unbehauen, J. Wiley and Sons Ltd, 1993. – 526 p.
18. Hyndman R.J., Koehler A.B. Another look at measures of forecast accuracy // International Journal of Forecasting. –2006. – № 22(4). – P. 679-688.
19. Shcherbakov M.V., Brebels A. Outliers and anomalies detection based on neural networks forecast procedure: Proceedings of the 31st Annual International Symposium on Forecasting (ISF 2011) / Prague: International Institute of Forecasters, 2011. – pp. 21-22. URL: [http://www.forecasters.org/isf/pdfs/ISF11\\_Proceedings.pdf](http://www.forecasters.org/isf/pdfs/ISF11_Proceedings.pdf)
20. Yu, Chong Ho. Exploratory data analysis in the context of data mining and resampling // International Journal of Psychological Research. 3. 2010.

**Коржавых Владислав Валерьевич**, заместитель начальника Валуйского района электрических сетей по реализации услуг

**Korzhavykh Vladislav Valerievich**, Deputy Head of the Valuysky Sistrict of Electric Grids for Sales of Services