Министерство науки и высшего образования Российской Федерации

Федеральное государственное автономное образовательное учреждение
высшего образования
«Уральский федеральный университет имени первого Президента России Б.Н. Ельцина»

Уральский энергетический институт

Кафедра «Электротехника»

ДОПУСТИТЬ К ЗАЩИТЕ ПЕРЕД ГЭК

Зав. кафедрой электротехники

_____ В.Э. Фризен

«13» _____июня_____ 2024 г.

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА

ПРИМЕНЕНИЕ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПРОГНОЗИРОВАНИЯ ВЫРАБОТКИ СОЛНЕЧНОЙ ЭЛЕКТРИЧЕСКОЙ СТАНЦИИ

Научный руководитель: А.И. Хальясмаа
канд. техн. наук, доц. _____ подпись

Нормоконтролер: С.А. Ерошенко
доц., канд. техн. наук _____ подпись

Студент группы ЭНМ-221201 М.А. Харлашкина _____ подпись

Екатеринбург
2024

Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное учреждение высшего образования
«Уральский федеральный университет имени первого Президента России Б.Н. Ельцина»

Институт *Уральский энергетический*
Кафедра *«Электротехника»*
Направление подготовки *01.04.04 – Прикладная математика*
Образовательная программа *Искусственный интеллект в электроэнергетике*

УТВЕРЖДАЮ
Зав. кафедрой В.Э. Фризен
«15» января 2024 г.

ЗАДАНИЕ

на выполнение выпускной квалификационной работы

студента Харлашкиной Марии Андреевны **группы** *ЭНМ-221201*
(фамилия, имя, отчество)

1. Тема выпускной квалификационной работы *Применение алгоритмов машинного обучения для прогнозирования выработки солнечной электрической станции*
Утверждена распоряжением по институту от «10» января 2024 г. № 33.07-05/003
2. Руководитель Хальясмаа Александра Ильмаровна, заведующий научной лабораторией ЦДЭЭ, доцент, канд. техн. наук
(Ф.И.О., должность, ученое звание, ученая степень)
3. Исходные данные к работе
3.1. Метеорологические данные для прогнозирования генерации фотоэлектрической станции. 3.2 Измерения потока солнечного излучения, падающего на плоскость панелей фотоэлектрической станции
4. Перечень демонстрационных материалов Презентация (15 слайдов)
5. Календарный план

№ п/п	Наименование этапов выполнения работы	Срок выполнения этапов работы	Отметка о выполнении
1.	*Основы солнечных электростанций: концепция и архитектура*	до 01 марта 2024 г.	
2.	*Предварительная обработка данных для прогнозирования генерации солнечной электрической станции*	до 01 апреля 2024 г.	
3.	*Разработка алгоритмов машинного обучения для прогнозирования выработки солнечной электрической станции*	до 15 мая 2024 г.	
4.	*ВКР в целом*	до 06 июня 2024 г.	

Руководитель _____ А.И. Хальясмаа
(подпись) Ф.И.О.
Задание принял к исполнению 15.01.2024 _____
дата (подпись)

6. Выпускная квалификационная работа закончена «06» июня 2024 г. считаю возможным допустить Харлашкину Марию Андреевну к защите выпускной квалификационной работы закончена в Государственной экзаменационной комиссии.

Руководитель_____ А.И. Хальясмаа
(подпись) Ф.И.О.

7. Допустить Харлашкину Марию Андреевну к защите магистерской диссертации в Государственной экзаменационной комиссии (протокол заседания кафедры №___ от «___» _____ 202_г.).

Зав. кафедрой _____ В.Э. Фризен
(подпись) Ф.И.О.

РЕФЕРАТ

Отчет 70 с., 54 рис., 2 табл., 29 источн.

КЛЮЧЕВЫЕ СЛОВА: ВОЗОБНОВЛЯЕМЫЕ ИСТОЧНИКИ ЭНЕРГИИ, ФОТОЭЛЕКТРИЧЕСКАЯ СТАНЦИЯ, ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ, МАШИННОЕ ОБУЧЕНИЕ, ПРОГНОЗИРОВАНИЕ ГЕНЕРАЦИИ ЭЛЕКТРИЧЕСКОЙ ЭНЕРГИИ СОЛНЕЧНЫМИ ЭЛЕКТРОСТАНЦИЯМИ, АНСАМБЛИ ДЕРЕВЬЕВ РЕШЕНИЙ.

Объектом исследования является электроэнергетическая система с высокой долей генерации на основе возобновляемых источников энергии.

Предметом исследования являются методы краткосрочного прогнозирования генерации фотоэлектрических станций.

Цель работы – анализ существующих подходов к краткосрочному прогнозированию генерации фотоэлектрических станций с использованием методов машинного обучения.

Результатом работы стала разработка алгоритма для прогнозирования генерации солнечной электростанции с использованием технологий искусственного интеллекта.

Разработанные модели были успешно протестированы, показав наилучший результат с использованием случайного леса.

СОДЕРЖАНИЕ

РЕФЕРАТ 4
ТЕРМИНЫ И ОПРЕДЕЛЕНИЯ 6
УСЛОВНЫЕ ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ 7
ВВЕДЕНИЕ 8
1 ОСНОВЫ СОЛНЕЧНЫХ ЭЛЕКТРОСТАНЦИЙ: КОНЦЕПЦИЯ, АРХИТЕКТУРА, ОСОБЕННОСТИ РАБОТЫ 11
1.1 Виды солнечных электростанций 11
1.2 Фотоэлектрические СЭС 12
1.3 Влияние погодных условий на выработку СЭС 27
1.4 Зависимость выработки солнечной энергии от времени года и в течение суток 29
1.5 Существующие методы прогнозирования выработки солнечной электроэнергии 32
2 ПРЕДВАРИТЕЛЬНАЯ ОБРАБОТКА ДАННЫХ ДЛЯ ПРОГНОЗИРОВАНИЯ ГЕНЕРАЦИИ СОЛНЕЧНОЙ ЭЛЕКТРИЧЕСКОЙ СТАНЦИИ 34
2.1 Выбор инструментов для программной реализации 34
2.2 Описание задачи предварительной обработки данных 35
2.3 Реализация предварительной обработки данных 36
2.4 Заключение по главе 2 43
3 РАЗРАБОТКА АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПРОГНОЗИРОВАНИЯ ВЫРАБОТКИ СОЛНЕЧНОЙ ЭЛЕКТРИЧЕСКОЙ СТАНЦИИ 44
3.1 Описание регрессионных моделей машинного обучения 44
3.2 Описание выборки данных для прогнозирования генерации 48
3.3 Проверка корректности выборки данных 49
3.4 Обучение моделей 59
3.5 Заключение по главе 3 70
ЗАКЛЮЧЕНИЕ 71
СПИСОК ЛИТЕРАТУРЫ 72

ТЕРМИНЫ И ОПРЕДЕЛЕНИЯ

В настоящей работе применяют следующие термины с соответствующими определениями.

Машинное обучение	– это область искусственного интеллекта, которая изучает алгоритмы и статистические модели, позволяющие компьютерным системам автоматически учиться на основе данных и улучшать свои производительные характеристики без явного программирования для каждой конкретной задачи
Объяснимый искусственный интеллект	– это подход в разработке и применении систем искусственного интеллекта, который фокусируется на создании моделей и алгоритмов, способных предоставлять понятные и логически обоснованные объяснения своих решений и действий
Обучающая выборка данных	– выборка представляет собой набор данных, на которых модель обучается, т.е. на основе которых она настраивает свои веса и определяет зависимости между входными данными и целевой переменной.
Тестовая выборка данных	– представляет собой независимый набор данных, на которых модель не обучалась, но который используется для оценки ее качества и точности предсказаний.

УСЛОВНЫЕ ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ

В настоящем ВКР применяют следующие сокращения и обозначения.

ВКР – выпускная квалификационная работа

ФЭС – фотоэлектрическая станция

ВИЭ – возобновляемый источники энергии

СЭС – солнечная электростанция

МВт – мегаватт

кВт – киловатт

ВАХ – вольт-амперная характеристика

КПД – коэффициент полезного действия

АБ – аккумуляторная батарея

ВВЕДЕНИЕ

Актуальность темы. Солнечная энергетика является одним из лидеров в области использования возобновляемых источников энергии. С каждым годом в мире увеличивается количество солнечных электростанций. Однако эффективность их работы сильно зависит от солнечной радиации, которая может сильно варьироваться в зависимости от погодных условий, времени года, географического положения и других факторов.

Прогнозирование производства солнечной электроэнергии важно для оптимизации работы СЭС, учета потребления энергии и уменьшения затрат на ее хранение. Точные прогнозы позволяют энергетическим компаниям более эффективно управлять производством и распределением электроэнергии, а также минимизировать риски, связанные с неожиданными изменениями в производственных показателях.

В последние годы наблюдается прогресс в области машинного обучения и искусственного интеллекта. Алгоритмы машинного обучения позволяют эффективно анализировать большие объемы данных и выявлять сложные зависимости между переменными. Применение алгоритмов для прогнозирования производства солнечной энергии может значительно повысить точность прогнозов и обеспечить более эффективное использование солнечной энергии.

Таким образом, исследование применения алгоритмов машинного обучения для прогнозирования производства солнечной электроэнергии является актуальным и важным фактором с точки зрения развития солнечной энергетики и оптимизации энергетического баланса.

Объектом исследования является электроэнергетическая система с высокой долей генерации на основе возобновляемых источников энергии.

Предметом исследования являются методы краткосрочного прогнозирования генерации фотоэлектрических станций.

Цель работы – анализ существующих подходов к краткосрочному прогнозированию генерации фотоэлектрических станций с использованием методов машинного обучения.

Задачи, поставленные для достижения цели:

- анализ существующих подходов к прогнозированию генерации фотоэлектрических станций и данных, используемых в прогнозных моделях, с целью определения набора метеорологических параметров для прогнозирования генерации фотоэлектрических станций;
- определение степени влияния предварительной обработки исходных данных на результаты прогнозирования генерации фотоэлектрических станций с использованием методов машинного обучения с целью установления целесообразности применения алгоритма предварительной обработки данных для повышения точности результатов прогноза;
- адаптация ансамблевых методов машинного обучения с использованием деревьев решений для краткосрочного прогнозирования генерации фотоэлектрических станций с точностью выше 80%;
- адаптация генетического алгоритма для решения задачи оптимизации топологии электрической сети с высокой долей генерации на основе ВИЭ для снижения суммарных потерь электроэнергии при ее передаче на каждом часе суток.

Научная новизна работы заключается в следующем:

- разработан новый алгоритм предварительной обработки исходных данных, использующий методы машинного обучения, и позволяющий повысить точность краткосрочного прогнозирования генерации фотоэлектрических станций;
- проведена адаптация регрессионных моделей, основанных на ансамблевых методах машинного обучения, для краткосрочного

прогнозирования генерации фотоэлектрических станций на основе метеорологических данных из открытых источников;

- предложена адаптированная концепция генетического алгоритма для оптимизации (по критерию суммарных потерь электроэнергии) топологии электрической сети с высокой долей генерации на основе ВИЭ.

1 ОСНОВЫ СОЛНЕЧНЫХ ЭЛЕКТРОСТАНЦИЙ: КОНЦЕПЦИЯ, АРХИТЕКТУРА, ОСОБЕННОСТИ РАБОТЫ

1.1 Виды солнечных электростанций

Солнечная электростанция представляет собой различные технологии и конструкции, используемые для преобразования солнечной энергии в электричество. С развитием солнечной энергетики и увеличением интереса к возобновляемым источникам энергии, появилось множество различных типов СЭС, каждый из которых имеет свои особенности и преимущества [1].

Фотоэлектрические СЭС: наиболее распространенный тип солнечных электростанций, который использует фотоэлектрические панели для преобразования солнечной энергии в электрическую энергию. ФЭС могут быть как крупными коммерческими объектами, так и малыми домашними системами.

Концентрирующие СЭС (КСЭС): используют системы концентрации солнечного света для нагрева рабочего тела, которое затем используется для генерации электрической энергии. КСЭС могут быть установлены на земле или на плавучих платформах и могут использовать различные технологии концентрации, такие как зеркала, линзы или цилиндры.

Гибридные СЭС: комбинируют в себе фотоэлектрические и концентрирующие технологии для генерации электрической энергии. Гибридные СЭС могут быть более эффективными, чем отдельные ФЭС или КСЭС, но могут быть более сложными в реализации и требовать больших инвестиций.

Мобильные СЭС: малые солнечные электростанции, которые могут быть переносными и использоваться для питания мобильных устройств, таких как мобильные телефоны, ноутбуки и другие электронные устройства.

Солнечные батареи на космических аппаратах: используются для питания космических аппаратов, таких как спутники и космические корабли. Солнечные батареи на космических аппаратах могут быть установлены на

поверхности космического аппарата или на специальных панелях, которые могут поворачиваться для максимального использования солнечного света.

Каждый из этих видов солнечных электростанций имеет свои преимущества и недостатки, и выбор типа СЭС зависит от многих факторов, таких как местоположение, масштаб, доступность ресурсов и требования к эффективности и надежности.

1.2 Фотоэлектрические СЭС

Рассмотрим более подробно фотоэлектрические СЭС, их преимущества и недостатки и основные принципы работы.

1.2.1. Фотоэлектрический эффект

Фотоэлектрический эффект, открытый французским ученым А.Э. Беккерелем в 1839 году, заключается в способности токопроводящих материалов испускать электроны под воздействием электромагнитного излучения, включая свет.

Три основных закона фотоэффекта можно сформулировать следующим образом [2]:

1) сила фототока прямо пропорциональна плотности электромагнитного излучения;

2) максимальная кинетическая энергия вырываемых светом электронов линейно возрастает с частотой электромагнитного излучения и не зависит от его интенсивности;

3) для каждого вещества при определенном состоянии его поверхности существует граничная частота электромагнитного излучения, ниже которой фотоэффект не наблюдается. Эта частота и соответствующая длина волны называются красной границей фотоэффекта.

Фотоэффект проявляется в фотоэлектрических системах, которые напрямую преобразуют солнечную энергию в электричество. Для работы таких систем требуется дневной свет. ФЭС не обязательно должны быть под прямыми солнечными лучами, они могут генерировать электроэнергию даже в пасмурные дни.

Простейшая конструкция фотоэлектрического или солнечного элемента, основанного на монокристаллическом кремнии, показана на рисунке 1.1

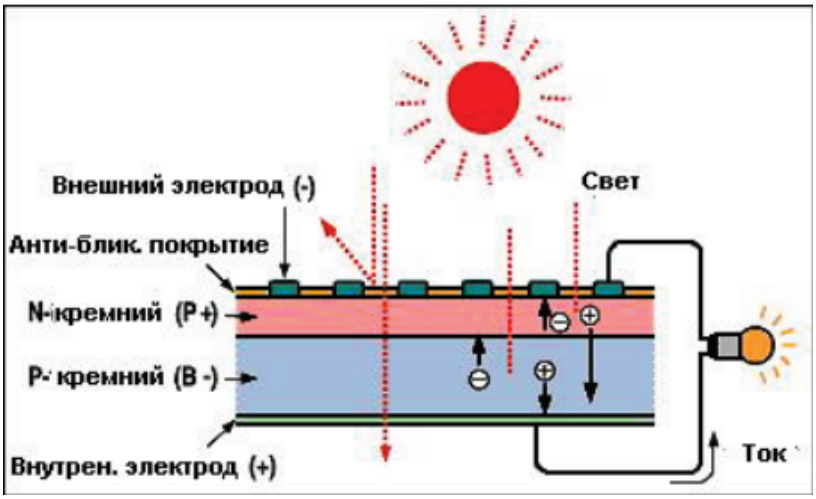

Рисунок 1.1 – Конструкция фотоэлектрического элемента

На небольшой глубине от поверхности кремниевой пластины типа p сформирован переход p-n с тонким металлическим контактом; на обратную сторону пластины нанесен сплошной металлический контакт.

Предположим, что переход p-n расположен близко к освещаемой поверхности полупроводника. Когда солнечный элемент используется как источник электроэнергии, к его выводам должно быть подключено сопротивление нагрузки R_H. Сначала рассмотрим два крайних случая: R_H =0 (режим короткого замыкания) и $R_H = \infty$ (режим холостого хода). Зонные диаграммы для этих режимов показаны на рисунке 1.2а, б [3].

В первом случае зонная диаграмма освещенного перехода p-n не отличается от зонной диаграммы при термодинамическом равновесии (без освещения и без приложенного напряжения смещения), поскольку внешнее закорачивание обеспечивает нулевую разность потенциалов между n- и p-областями. Однако через переход p-n и внешний проводник протекает ток, вызванный фотогенерацией электронно-дырочных пар в p-области. Фотоэлектроны, образовавшиеся в непосредственной близости от области

объемного заряда, увлекаются электрическим полем перехода p-n и попадают в n-область.

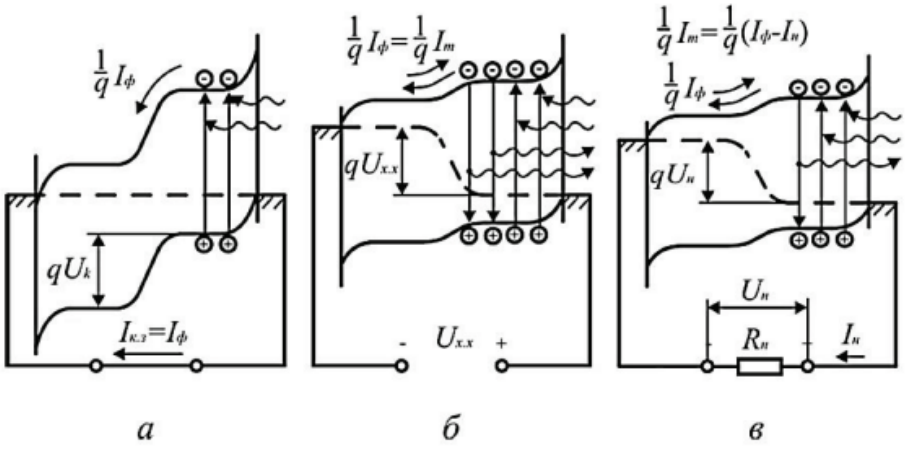

Рисунок 1.2 – Зонные энергетические диаграммы p–n-перехода при освещении в разных режимах: а – короткого замыкания; б – холостого хода; в – включения на сопротивление нагрузки

Остальные электроны диффундируют к p–n-переходу, стараясь восполнить их убыль, и в итоге также попадают в n-область. В n-области возникает направленное движение электронов к тыльному металлическому контакту, перетекание во внешнюю цепь и в контакт с p-областью. На границе контакта с p-областью происходит рекомбинация подошедших сюда электронов с фотогенерированными дырками.

При разомкнутой внешней цепи p–n-перехода (рисунок 1.2б) фотоэлектроны, попадая в n-область, накапливаются в ней и заряжают её отрицательно. Оставшиеся в p-области избыточные дырки заряжают p-область положительно. Возникающая таким образом разность потенциалов является напряжением холостого хода (Uxx), полярность которого соответствует прямому смещению p–n-перехода [4].

Поток света генерирует носители заряда, создавая фототок (Iф), который равен количеству фотогенерированных носителей, прошедших через p–n-переход за единицу времени. При отсутствии внутренних омических потерь в

солнечном элементе, режим короткого замыкания (рисунок 1.2а) соответствует нулевому напряжению смещения p–n-перехода, поэтому ток короткого замыкания (Iкз) равен фототоку (Iф).

В режиме холостого хода (рисунок 1.2б) фототок компенсируется током (Iт), который представляет собой прямой ток через p–n-переход при напряжении смещения (Uxx). Данный ток сопровождается рекомбинацией неосновных носителей тока (в данном случае электронов в p-области). При рекомбинации потенциальная энергия электронно-дырочных пар выделяется либо в виде излучения фотонов с энергией hv≈Eg, либо расходуется на нагревание кристаллической решетки (рисунок 1.2б). Таким образом, режим холостого хода солнечного элемента эквивалентен работе светодиодов и выпрямительных диодов в пропускном направлении.

Если к p–n-переходу подключить изменяемое сопротивление нагрузки (рисунок 1.2в), то направление тока в ней всегда совпадает с направлением фототока (Iф), а сам ток нагрузки (Iн) равен результирующему току через p–n-переход.

Максимальная мощность отбирается в том случае, когда солнечный элемент находится в режиме, отмеченном точкой а на рисунке 1.3.

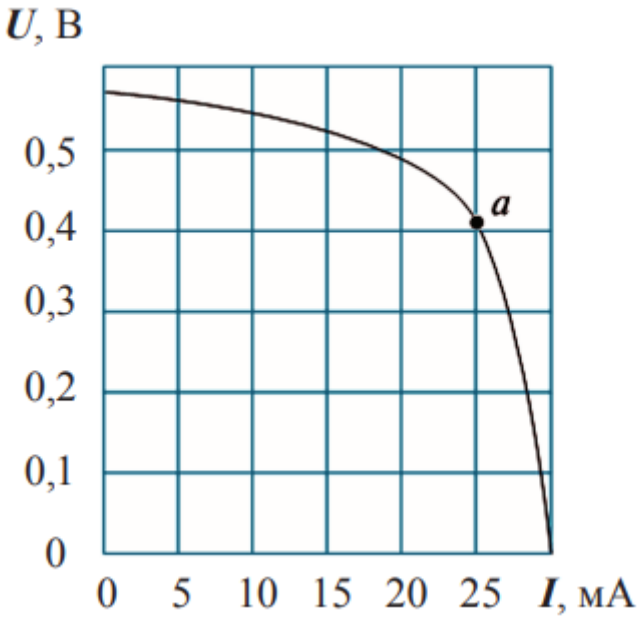

Рисунок 1.3 – Вольт-амперная характеристика солнечного элемента

1.2.2 Характеристики солнечного элемента

Напряжение холостого хода – это максимальное напряжение, которое может быть создано солнечным элементом при отсутствии тока (рисунок 1.4). Оно соответствует прямому смещению p–n-перехода, возникающему при появлении светового тока. Обозначается как U_{XX} или V_{OC}. У монокристаллических солнечных элементов высокого качества напряжение холостого хода может достигать 730 мВ при условии AM1.5, в то время как в коммерческих устройствах оно обычно составляет около 600 мВ. Напряжение холостого хода солнечного элемента мало изменяется при изменении уровня освещенности.



Рисунок 1.4 – Вольт-амперная характеристика солнечного элемента и напряжение холостого хода

Ток короткого замыкания (Iкз или Isc) представляет собой ток, протекающий через солнечный элемент, когда напряжение равно нулю (то есть когда солнечный элемент замкнут накоротко) (рисунок 1.5 [5]). Он возникает в результате генерации и разделения созданных носителей заряда. В идеальном солнечном элементе, при условии умеренных потерь от сопротивления, он эквивалентен световому току. Поэтому ток короткого замыкания можно рассматривать как максимально возможный ток, создаваемый солнечным элементом. Кроме того, он прямо пропорционален интенсивности света.

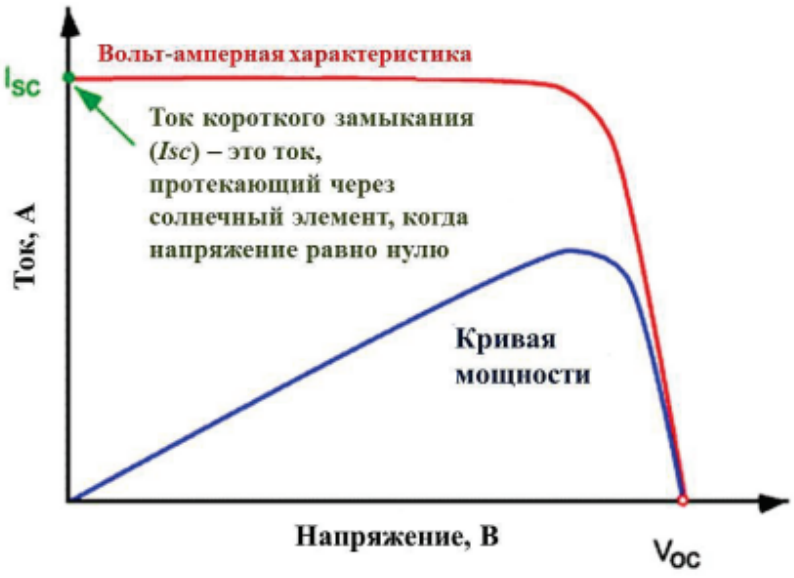

Рисунок 1.5 – Вольт-амперная характеристика солнечного элемента и ток короткого замыкания

Точка максимальной мощности (ТММ) – это лучшая комбинация тока и напряжения, позволяющая выработать достаточную мощность. Соответствующие напряжение и ток обозначаются *U*тмм и *I*тмм.

Коэффициент заполнения (fill factor, FF) вольт-амперной характеристики солнечного элемента является важным параметром, который определяет эффективность преобразования солнечной энергии в электрическую энергию. Ток короткого замыкания и напряжение холостого хода являются максимальными значениями тока и напряжения, которые можно получить от солнечного элемента. Однако, при этих значениях мощность солнечного элемента равна нулю.

Коэффициент заполнения определяется как отношение максимальной мощности солнечного элемента к произведению напряжения холостого хода и тока короткого замыкания. Он показывает, насколько близко реальная вольт-амперная характеристика солнечного элемента приближается к идеальной прямоугольной форме.

Графически коэффициент заполнения представляет собой меру квадратичности солнечного элемента и равен максимальной площади прямоугольника, который можно вписать в вольт-амперную кривую (рисунок

1.6). Чем выше коэффициент заполнения, тем более эффективно солнечный элемент преобразует солнечную энергию в электрическую.

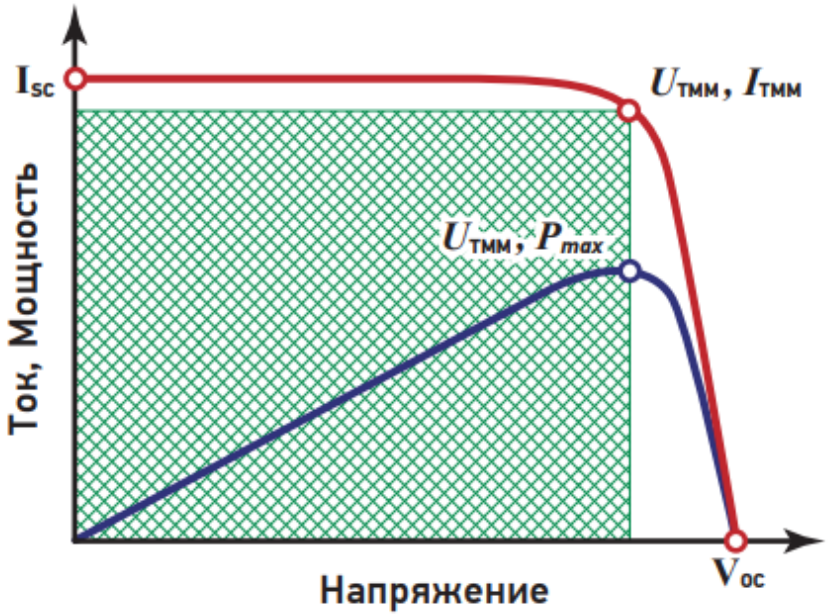

Рисунок 1.6 – Графическое представление коэффициента заполнения вольт-амперной характеристики солнечного элемента

Поскольку коэффициент заполнения является мерой квадратичности вольт-амперной кривой, солнечный элемент с более высоким напряжением будет иметь и более высокий потенциал для коэффициента заполнения, поскольку закругленная часть кривой занимает меньше места.

Коэффициент заполнения ВАХ является ключевым показателем, по которому можно оценить эффективность фотоэлектрического преобразователя. Стандартные серийно выпускаемые солнечные элементы имеют коэффициент заполнения ВАХ более 0,7. Дефектные элементы обычно имеют коэффициент заполнения ВАХ от 0,4 до 0,65. У аморфных и других тонкопленочных фотоэлектрических преобразователей коэффициент заполнения ВАХ 0,4–0,7. Чем выше коэффициент заполнения ВАХ, тем меньше потери в элементе из-за внутреннего сопротивления.

График на рисунке 1.7 демонстрирует ВАХ элементов с различным коэффициентом заполнения. Как видно, обе кривые имеют одинаковые значения тока короткого замыкания и напряжения холостого хода, однако

элемент с более низким коэффициентом заполнения ВАХ (нижний график) вырабатывает меньше мощности в точке максимальной мощности.

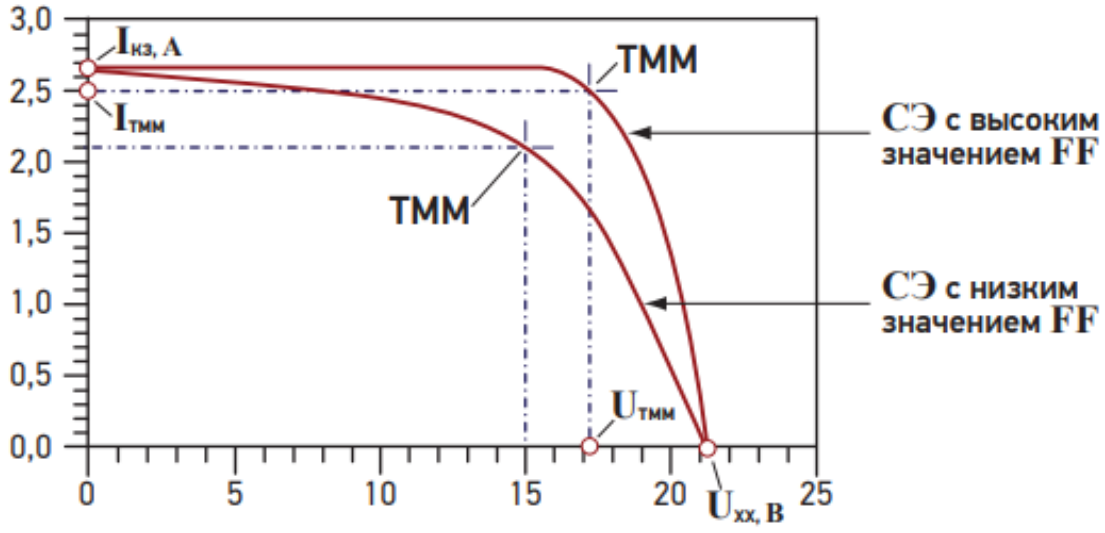

Рисунок 1.7 – Вольт-амперная характеристика солнечных элементов с различным коэффициентом заполнения

Коэффициент полезного действия является наиболее распространённым показателем для сравнения производительности двух фотоэлементов. Он рассчитывается как отношение электрической мощности, генерируемой фотоэлементом, к мощности солнечного излучения, падающего на него. Помимо собственной эффективности фотоэлемента, КПД также зависит от спектра и интенсивности падающего солнечного излучения, а также от температуры самого фотоэлемента. Поэтому для корректного сравнения двух фотоэлементов необходимо строго соблюдать установленные стандартные условия. Эффективность фотоэлемента определяется как доля падающей энергии, преобразованной в электричество.

1.2.3 Типы солнечных элементов

Наиболее распространены в фотоэлектрических установках элементы следующих видов (рисунок 1.8):

- на основе монокристаллического кремния (КПД до 21,5%),
- на основе поликристаллического кремния (КПД 14-17%)
- на основе аморфного кремния (КПД 5-8%).

Различие между этими видами солнечных элементов заключается в структуре кристаллической решетки кремния. Монокристаллические элементы изготавливаются из одного большого кристалла кремния с идеально упорядоченной структурой, что обеспечивает высокую эффективность преобразования солнечной энергии. Поликристаллические элементы состоят из множества мелких кристаллов с различной ориентацией, что немного снижает их КПД по сравнению с монокристаллическими. Аморфные элементы имеют совершенно другую структуру – они состоят из непрерывной сетки атомов кремния без четкой кристаллической решетки. Это делает их менее эффективными, но зато более дешевыми в производстве и гибкими в применении.

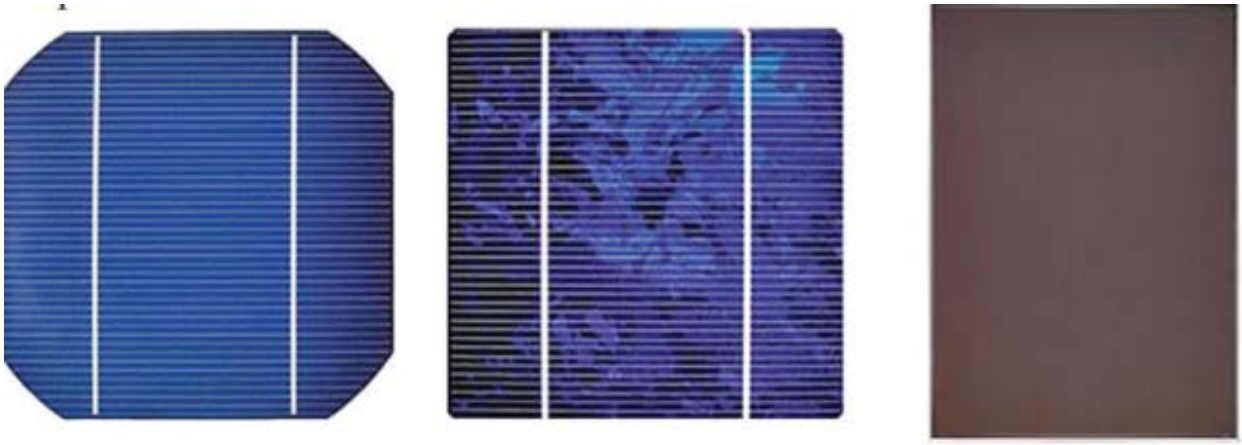

Рисунок 1.8 – Монокристаллические, поликристаллические и аморфные солнечные элементы

Стандартная структура фотоэлектрических элементов включает задний контакт и два слоя кремния разной проводимости. Сверху находится сетка из металлических контактов и антибликовое просветляющее покрытие, которое придает солнечному элементу характерный синий оттенок.

Размеры солнечных элементов могут варьироваться, наиболее часто встречающиеся параметры их электрических характеристик приведены в таблице 1.1 [6].

Таблица 1.1 – Типичные размеры солнечных элементов и их электрические параметры

Тип СЭ	Размер СЭ, мм	КПД, %	Максимальная мощность ($P_{тмм}$), Вт	Напряжение холостого хода ($U_{хх}$), В	Ток короткого замыкания ($I_{кз}$), А
	156×156	16,4	3,99	0,618	8,27
		16,2	3,94	0,616	8,21
		⋮	⋮	⋮	⋮
		15,0	3,65	0,607	7,77
	156×156	17,6	4,21	0,624	8,63
		17,4	4,16	0,623	8,57
		⋮	⋮	⋮	⋮
		16,0	3,82	0,613	8,08
	125×125	17,4	2,59	0,621	5,35
		17,2	2,56	0,619	5,31
		⋮	⋮	⋮	⋮
		16,0	2,38	0,607	5,08

В солнечной энергетике одним из перспективных материалов для создания высокоэффективных фотоэлектрических элементов является арсенид галлия (GaAs). Такие элементы обладают высоким КПД (для однопереходных элементов около 28 %). Отдельно можно выделить элементы, использующие органические материалы. Фотоэлектрические элементы на основе диоксида титана (TiO_2), покрытые органическим красителем, имеют КПД около 11 %. Принцип работы элемента основан на фотовозбуждении красителя и быстрой инжекции электрона в зону проводимости TiO_2.

В последние годы появились новые типы материалов для тонкопленочных фотоэлектрических элементов, такие как медь-индий-диселенид и теллурид кадмия (CdTe), которые широко используются в солнечных элементах. Технологии их производства постоянно развиваются, и за последнее десятилетие коэффициент преобразования энергии тонкопленочных элементов вырос примерно в 2 раза.

В настоящее время также активно исследуются гибридные методы, включающие как кристаллические, так и аморфные компоненты. Это позволяет повысить общий коэффициент преобразования солнечного элемента. В лабораториях уже получены элементы с коэффициентом преобразования энергии порядка 45%. Несмотря на то, что массовое использование этих технологий пока не близко, работа по удешевлению изготовления солнечных элементов активно ведется во всем мире.

1.2.4 Солнечные панели

Солнечные панели (они же фотоэлектрические или солнечные модули) состоят из солнечных элементов. Так как один солнечный элемент не производит достаточного количества электроэнергии, несколько таких элементов собирают в солнечные панели, чтобы производить больше электричества (рисунок 1.9).

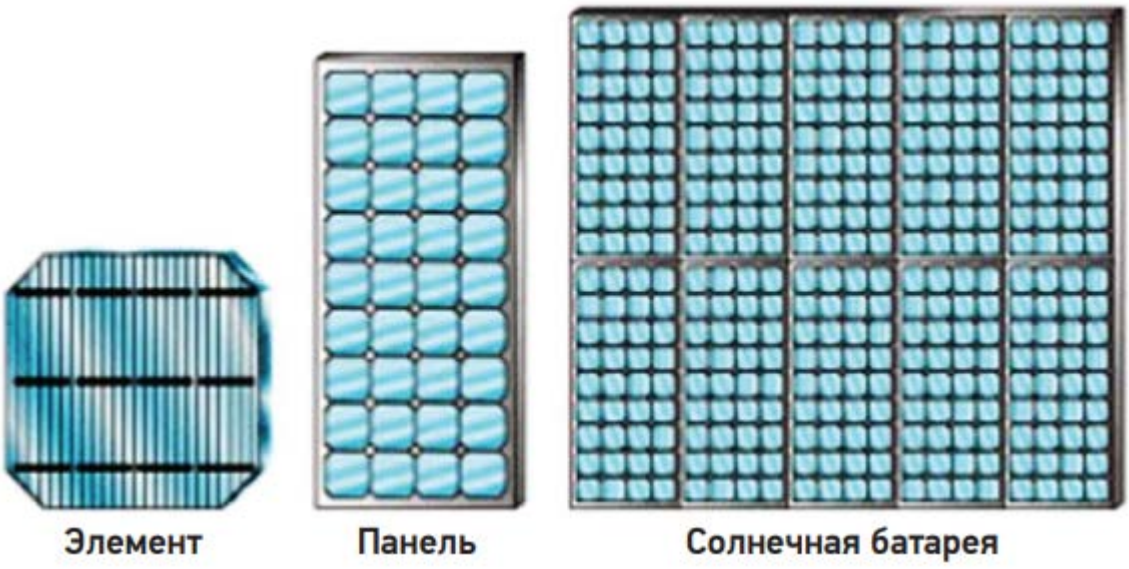

Рисунок 1.9 – Солнечный элемент, солнечная панель (фотоэлектрический модуль), солнечная батарея

Панель состоит из стеклянной плиты, на обратной стороне которой расположены солнечные элементы, электрически соединенные между собой металлическими шинами. Эти солнечные элементы находятся между двумя слоями герметизирующей пленки. Нижний слой герметизирующей пленки

защищен специальным защитным слоем, который обеспечивает защиту от внешних воздействий [7].

Кроме того, к внутренней стороне корпуса панели прикреплен блок терминалов. Под крышкой этого блока находятся электрические контакты для подключения. Эти контакты играют важную роль в обеспечении эффективного и безопасного подключения фотоэлектрической панели к другим устройствам или электрическим системам. (рисунок 1.10).

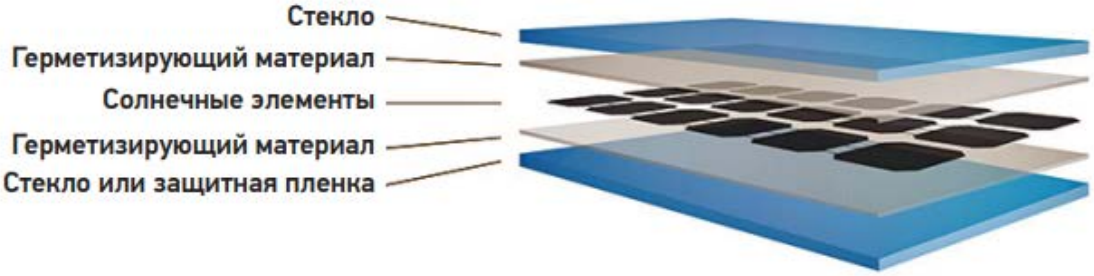

Рисунок 1.10 – Структура солнечной панели

Герметизирующий материал играет важную роль в обеспечении полной герметичности солнечных элементов, особенно при эксплуатации на открытом воздухе в течение всего года. Попадание воздуха или влаги внутрь солнечной панели может привести к окислению и разрушению контактов солнечных элементов, что в конечном итоге приводит к выходу панели из строя.

Наиболее распространенным герметизирующим материалом является этиленвинилацетатная (EVA) пленка. Однако, со временем она может потерять свою прозрачность, что является одним из факторов старения фотоэлектрических панелей. В настоящее время во всем мире ведутся исследования и разработки по замене EVA-пленки на другие материалы, но пока она остается основным герметизирующим материалом.

Солнечные панели производятся многих типов и размеров. Наиболее типичные – кремниевые солнечные панели мощностью 40-260 $Вт_{пик}$ (то есть с максимальной мощностью 40-260 Вт при ярком солнце) [8]. Такая солнечная панель имеет размер от 0,4 до 2,5 $м^2$. Панели могут соединяться между собой

в солнечные батареи для большей мощности (например, две панели по 50 $Вт_{пик}$, соединенные вместе, эквивалентны панели мощностью 100 $Вт_{пик}$).

КПД панелей колеблется от 5 до 20%, то есть именно такое количество солнечной энергии будет трансформировано в электричество.

1.2.5 Фотоэлектрические станции

Чтобы солнечные панели были надежным источником электроэнергии, необходимы дополнительные элементы в системе: кабели, поддерживающая структура, а также электронный инвертор и контроллер заряда с аккумуляторной батареей, в зависимости от типа системы (соединенная с сетью, автономная или резервная). Такая система в целом называется солнечной фотоэлектрической системой или солнечной станцией.

Есть три основных типа солнечных фотоэлектрических станций [9]:

- автономные, обычно применяемые для электроснабжения отдельных домов;
- соединенные с сетью;
- резервные.

Автономные фотоэлектрические станции

Автономные ФЭС применяются в случаях, когда отсутствуют централизованные сети электроснабжения. Для обеспечения энергией в темное время суток или в периоды с недостаточным солнечным светом требуется использование аккумуляторной батареи [10]. Автономные ФЭС часто используются для электроснабжения отдельных домов. Малые системы способны обеспечить базовую нагрузку, такую как освещение. Более мощные системы могут также питать предметы крупной бытовой техники (холодильник), обслуживающие устройства (водяной насос, электроинструмент) и другие устройства. Такая система включает в себя солнечную панель, контроллер, аккумуляторную батарею, кабели, электрическую нагрузку и поддерживающую структуру (рисунок 1.11).

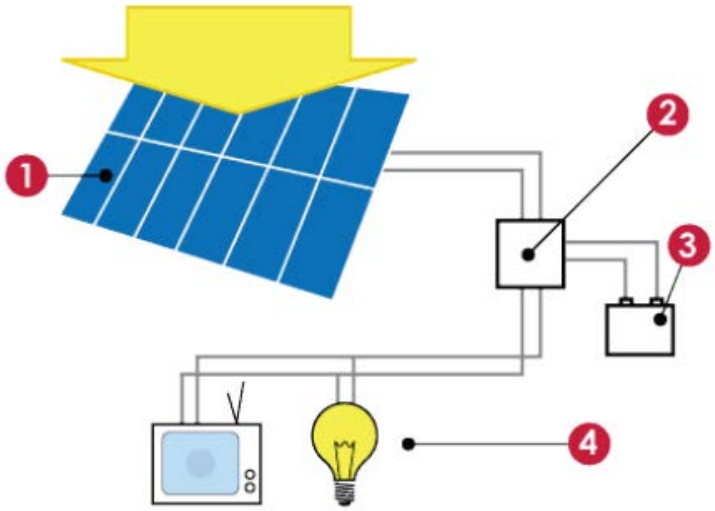

Рисунок 1.11 – Автономная фотоэлектрическая система: 1 – солнечные панели; 2 – контроллер; 3 – АБ; 4 – нагрузка

ФЭС, соединенные с сетью

При наличии централизованной сети электроснабжения, но желании получать электроэнергию от чистого источника, солнечные панели могут быть подключены к сети. При достаточном количестве солнечных панелей, определенная часть нагрузки в доме может быть обеспечена солнечным электричеством.

Системы, подключенные к сети, обычно состоят из одной или нескольких панелей, инвертора, кабелей, поддерживающей структуры и электрической нагрузки (рисунок 1.12).

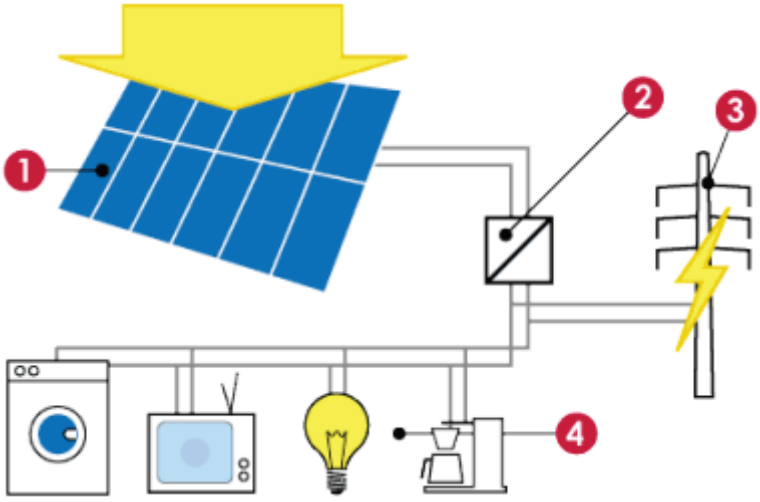

Рисунок 1.12 – Соединенная с сетью солнечная фотоэлектрическая система: 1 – солнечные панели; 2 – инвертор; 3 – сеть; 4 – нагрузка

Инвертор используется для соединения солнечных панелей с сетью. Также существуют так называемые АС-панели, в которых инвертор интегрирован в заднюю часть панели.

Избыточная электрическая энергия может продаваться электросетевым компаниям. При использовании отдельных тарифов для солнечного электричества устанавливают два счетчика – один на генерацию, другой на потребление. При этом вся электроэнергия, выработанная солнечной панелью, продается сетям по повышенной цене, а потребности дома в электроэнергии обеспечиваются покупкой электроэнергии у сетей по обычной цене. Такой подход позволяет не только обеспечить нулевые расходы на электроэнергию в течение года, но и достичь нулевого потребления электроэнергии за год. Летом, когда солнечная панель производит избыток энергии, эта энергия поставляется сетям, а зимой, когда солнечного света меньше, дом питается в основном от сетей.

Резервные системы

Резервные солнечные системы используются в случаях, когда существует связь с централизованной электросетью, но ее надежность оставляет желать лучшего. Эти системы могут обеспечить энергией в периоды отсутствия напряжения в сети. Небольшие резервные солнечные системы электроснабжения могут использоваться для освещения, питания компьютеров и средств связи (телефонов, радио, факсов и прочего). Более крупные системы способны обеспечивать энергией даже холодильники во время отключения сети. Чем больше мощность требуется для питания важной нагрузки и чем длительнее периоды отключения сети, тем больше мощности потребуется от фотоэлектрической системы. В случае наличия электросети система обычно работает в режиме подключения к ней.

Система состоит из солнечных панелей, контроллера, аккумуляторной батареи, кабелей, инвертора, нагрузки и поддерживающей структуры (рисунок 1.13).

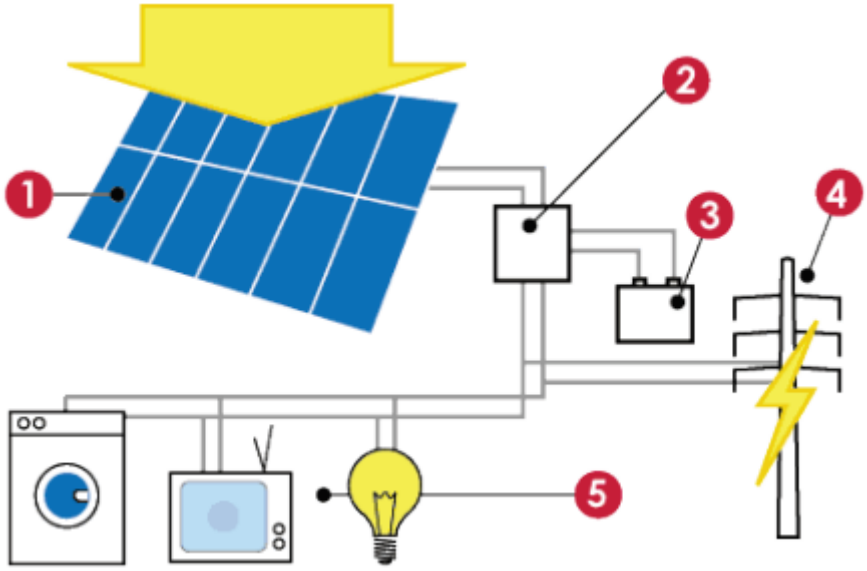

Рисунок 1.13 – Резервные системы: 1 – солнечные панели; 2 – инвертор; 3 – АБ; 4 – сеть; 5 – нагрузка

1.3 Влияние погодных условий на выработку СЭС

Погодные условия оказывают значительное влияние на выработку электроэнергии солнечными электростанциями. Различные погодные факторы, такие как солнечная радиация, облачность, осадки, ветер и температура, могут существенно влиять на производительность СЭС [11-14].

Солнечная радиация является основным источником энергии для СЭС. Интенсивность солнечной радиации влияет на количество электроэнергии, которое может быть произведено. В ясные и солнечные дни СЭС достигают максимальной производительности, в то время как в облачные или пасмурные дни выработка электроэнергии снижается. Исследования показывают, что облачность может снизить выработку СЭС на 50-80% по сравнению с ясными днями.

Осадки, такие как дождь, снег и град, могут временно снижать выработку солнечных электростанций по нескольким причинам. Во-первых, вода, снег или град могут покрыть поверхность солнечных панелей, создавая

препятствие для поглощения солнечной радиации. Это приводит к уменьшению эффективности панелей и снижению выработки электроэнергии. Особенно значительное влияние осадков наблюдается при плотном снеговом покрове или граде, которые могут полностью покрыть поверхность панелей. Во-вторых, осадки могут изменять характеристики солнечного света, отражая и рассеивая его. Это может привести к уменьшению интенсивности солнечной радиации, достигающей поверхности солнечных панелей. В результате, панели получают меньше энергии и выработка электроэнергии снижается. В-третьих, осадки, особенно сильные дожди или град, могут повредить электрические компоненты СЭС, такие как провода, соединения и инверторы. Это может привести к временному или постоянному снижению производительности системы.

Ветер может влиять на выработку СЭС, особенно если солнечные панели установлены на высоких мачтах или на открытых пространствах. Сильный ветер может вызывать колебания солнечных панелей, что может снизить их эффективность. Однако ветер также может охлаждать солнечные панели, что может улучшить их производительность в жаркую погоду.

Высокие температуры могут снижать производительность солнечных панелей. При высоких температурах электронная проводимость в солнечных панелях ухудшается, что приводит к снижению выработки электроэнергии. Исследования показывают, что в жаркую погоду, особенно в пустынных регионах, выработка СЭС может снижаться на 10-15% по сравнению с оптимальными условиями.

В целом, погодные условия играют важную роль в выработке электроэнергии солнечными электростанциями. Интенсивность солнечной радиации, облачность, осадки, ветер и температура могут существенно влиять на производительность СЭС. При проектировании и эксплуатации СЭС необходимо учитывать эти факторы для достижения максимальной эффективности и производительности системы.

1.4 Зависимость выработки солнечной энергии от времени года и в течение суток

Для анализа графика выработки солнечной энергии в течение года важно учитывать несколько ключевых факторов, влияющих на объем выработки СЭС.

1) Географическое положение. Широта местности сильно влияет на количество солнечного света, получаемого в течение года. Чем ближе местность к экватору, тем равномернее распределение солнечного излучения в течение года, в то время как в высоких широтах наблюдаются значительные сезонные колебания.

2) Сезонные изменения. Из-за наклона земной оси относительно плоскости её орбиты, в разные времена года солнце восходит и заходит под разными углами, что влияет на продолжительность светового дня и интенсивность солнечного света.

3) Атмосферные условия: Облачность, загрязнение воздуха и другие атмосферные факторы могут существенно снижать количество солнечной энергии, достигающей земной поверхности.

Подробнее рассмотрим типичный график выработки солнечной энергии в умеренных широтах в течение года:

Зима (декабрь – февраль):

- в этот период дни короткие, а угол восхода и захода солнца низкий, солнечный свет падает под малым углом, что уменьшает количество энергии, которое может быть захвачено солнечными панелями;
- выработка солнечной энергии находится на минимальном уровне, часто составляя менее 20% от максимального значения, достигаемого летом;
- в некоторых регионах с сильными снегопадами и облачностью выработка может быть еще ниже из-за дополнительного затенения и рассеивания солнечного света.

Весна (март – май):

- с увеличением продолжительности дней и повышением угла восхода солнца начинает расти и выработка солнечной энергии;
- в начале весны этот рост может быть довольно медленным, но к середине и концу весны выработка значительно увеличивается, приближаясь к летним значениям;
- в этот период солнечные панели начинают эффективнее работать, поскольку получают больше прямого солнечного света.

Лето (июнь – август):

- летом дни самые длинные, а угол восхода и захода солнца наиболее высокий, что обеспечивает максимальное количество солнечной энергии, которое может быть захвачено солнечными панелями;
- выработка солнечной энергии достигает пика, часто превышая 80-100% от среднегодового значения.
- в этот период солнечные системы могут генерировать достаточно энергии не только для покрытия потребностей дома или предприятия, но и для продажи излишков в электросеть.

Осень (сентябрь – ноябрь):

- с уменьшением продолжительности дней и понижением угла восхода солнца начинает снижаться и выработка солнечной энергии;
- этот спад может быть достаточно плавным, особенно в регионах с мягким климатом, или более резким в более холодных регионах;
- к концу осени выработка солнечной энергии снижается до уровней, сопоставимых с зимними месяцами.

На рисунке 1.14 показан график выработки электроэнергии четырьмя поликристаллическими солнечными панелями мощностью по 250Вт. Видно, что пик выработки приходится на период май-июль, в эти месяцы в сутки будет сгенерировано до 5кВт*час энергии. Минимум приходится на период ноябрь-январь. В зимние месяцы выработка вообще может снижаться в 10-15 раз по сравнению с летним периодом.



Рисунок 1.14 – Выработка в течение года

Если рассматривать график выработки солнечной энергии в течение дня, то также можно заметить цикличность данного процесса. Обычно утром, после восхода солнца, происходит постепенное увеличение выработки энергии. Пиковые значения достигаются примерно к середине дня, когда солнце находится на самом высоком уровне. В это время выработка солнечной энергии наиболее интенсивна и может достигать максимальных значений.

После полудня выработка энергии начинает постепенно снижаться по мере того, как солнце опускается к горизонту. Затем, к закату, происходит резкое снижение выработки солнечной энергии, и она стремительно приближается к нулю. Вечером солнечная энергия практически не производится из-за отсутствия прямого солнечного света.

Этот цикл повторяется каждый день в зависимости от времени года и погодных условий. В летний период дни длиннее, что позволяет солнечным батареям работать более продолжительное время и вырабатывать больше энергии. В зимний период дни короче, и выработка энергии может быть менее интенсивной из-за низкого стояния солнца.

1.5 Существующие методы прогнозирования выработки солнечной электроэнергии

Методы прогнозирования выработки солнечной электроэнергии играют важную роль в оптимизации работы солнечных электростанций, планировании энергопотребления и снижении затрат на хранение энергии. Рассмотрим основные методы прогнозирования, их преимущества и ограничения, а также их применение в реальных условиях [15-18].

Статистические методы: включают в себя методы временных рядов, авторегрессионные модели, скользящее среднее и другие. Эти методы основаны на анализе исторических данных о погодных условиях и выработке солнечной электроэнергии для предсказания будущей выработки. Они обычно требуют большого объема данных для обучения модели и могут быть достаточно точными при наличии достаточного количества данных.

Физические модели: основаны на физических законах и уравнениях, описывающих процесс преобразования солнечной энергии в электрическую энергию. Эти модели используют данные о погодных условиях, таких как облачность, температура, влажность и т.д., чтобы предсказать выработку солнечной электроэнергии. Физические модели могут быть достаточно точными, но требуют большого объема вычислений и могут быть сложными в реализации.

Методы машинного обучения: включают в себя алгоритмы машинного обучения, такие как линейная регрессия, деревья решений, случайный лес, нейронные сети и т.д. [19] Эти методы используют исторические данные о погодных условиях и выработке солнечной электроэнергии для обучения модели, которая затем может использоваться для прогнозирования будущей выработки. Методы машинного обучения могут быть достаточно точными и гибкими, но требуют большого объема данных для обучения модели.

Гибридные методы: комбинируют в себе статистические, физические и методы машинного обучения для прогнозирования выработки солнечной электроэнергии. Гибридные методы могут быть более точными, чем отдельные

методы, но требуют большего объема вычислений и могут быть сложными в реализации.

Каждый из этих методов имеет свои преимущества и недостатки, и выбор метода зависит от доступных данных, требуемой точности прогнозирования и доступных вычислительных ресурсов.

2 ПРЕДВАРИТЕЛЬНАЯ ОБРАБОТКА ДАННЫХ ДЛЯ ПРОГНОЗИРОВАНИЯ ГЕНЕРАЦИИ СОЛНЕЧНОЙ ЭЛЕКТРИЧЕСКОЙ СТАНЦИИ

2.1 Выбор инструментов для программной реализации

Для программной реализации выбран язык программирования Python по следующим причинам:

- наличие библиотек с широкими функциональными возможностями для обработки данных и машинного обучения;
- простота синтаксиса;
- возможность удобной интеграции с системами управления базами данных;
- язык, инструменты написания программного обеспечения на нем и библиотеки относятся к свободному программному обеспечению.

Для использования языка программирования нужно выбрать платформу (среду) для разработки. Для Python таких сред существует множество, в примерах данного пособия будет использована облачная среда Google Colab. Google Colab – облачная платформа для создания и распространения решений в области обработки и визуализации данных, машинного обучения и просто программирования на языке Python.

Во-первых, Google Colab требует установки и персонализированной настройки на компьютере, исключает необходимость копирования и пересылки данных. Это позволяет пользователю работать над проектами на любом рабочем месте (компьютере), имеющем подключение к сети Интернет.

Во-вторых, Google Colab имеет ряд достоинств, которые привлекают к нему и начинающих специалистов, и профессионалов в математическом моделировании и Data Science.

Работа в Google Colab построена на принципе блокнота Jupyter Notebook. Суть этого принципа в том, что программный код, пояснения к нему и результаты работы помещаются в одном пространстве, разбитом на ячейки. Такая логика особенно удобна для программирования в задачах

моделирования, анализа данных, научных исследованиях, потому что все оказывается на одном экране. В то время как классический подход к созданию программного обеспечения подразумевает написание кода в одних файлах, документации в других, а для получения результатов работы нужно запустить созданную программу.

2.2 Описание задачи предварительной обработки данных

Исходные данные, с которым нужно работать при решении задач машинного обучения, как правило, не позволяют сразу же применить к ним методы машинного обучения [20,21]. Основные причины приведены ниже:

- численные данные могут быть представлены в нечисловом формате, например, вместо числа «10» в поле таблицы указано «10 кВт» или ожидается число с десятичным разделителем «.», но в файле использован разделитель «,»;
- разные разделители столбцов, например, столбцы отделены друг от друга символом «;», но какие-то два разных параметра записаны в один столбец и разделены пробелом или запятой;
- пропущенные значения, когда вместо числа стоит прочерк «-», вопрос «?», пробел « » или другой символ, или просто пустое значение;
- выбросы, ошибочные значения, которые записаны в корректном формате, но из-за ошибки записи или других причин очень сильно выбивающиеся и искажающие статистические показатели, например, в столбце записаны мощности в кВт, но отдельные значения случайно были записаны в МВт.

Вручную находить и исправлять такие ошибки слишком долго, легко что-то пропустить. Поэтому необходимо уметь автоматизировать процесс очищения данных. В общем случае он состоит из следующих шагов:

1) чтение исходного файла;

2) приведение столбцов к нужным форматам;

3) обнаружение и удаление пропусков;

4) обнаружение и исправление выбросов;

5) запись полученного файла.

2.3 Реализация предварительной обработки данных

1. Чтение исходного файла, с данными о метеорологических факторах и генерации, фрагмент показан на рисунке 2.1.

	A	B	C	D	E	F	G	H	I	J
1	Data	Time	Temp	Lightness	Day_number	Solar_ang_1	Solar_ang_2	Humidity	Visibility	Wind_speed
2	11.02.2017	4:30:00	-21,240	0,212	42,000	-	0,000	71%	N/A	7.2 km/h / 2.0 m/s
3	11.02.2017	5:00:00	-21,170	0,271	42,000	-14,587	0,000	71%	10,0 км	10.8 km/h / 3.0 m/s
4	11.02.2017	5:30:00	-20,980	0,306	42,000		?	71%	10,0 км	10.8 km/h / 3.0 m/s
5	11.02.2017	6:00:00	-21,030	0,270	42,000	-14,587	0,000	71%	10,0 км	10.8 km/h / 3.0 m/s
6	11.02.2017	6:30:00	-21,240	0,239	42,000	-14,587	0,000	71%	10,0 км	10.8 km/h / 3.0 m/s

Рисунок 2.1 – Фрагмент исходного файла в Excel

Из фрагмента видны следующие проблемы:

- есть пропуски, причем пропущенные значения могут быть заменены различными символами (прочерк, вопрос), могут быть просто отсутствующие значения или N/A;
- есть столбцы с точкой в качестве десятичного разделителя, а есть с запятой;
- есть столбцы с численными данным, но в которых содержится еще и текст;
- дата и время разбиты по двум колонкам и между элементами времени и даты использованы свои разделители.

Для работы с табличными данными в Python обычно используют библиотеки NumPy и Pandas. Их нужно подключить:

```
import numpy as np
import pandas as pd
```

Прочитать файл и вывести на экран его начало можно следующим образом (результат на рисунке 2.2):

input_df = pd.read_excel('/content/solar_raw_data.xlsx')

input_df.head()

```
1 input_df = pd.read_excel('/content/solar_raw_data.xlsx')
2 input_df.head()
```

	Data	Time	Temp	Lightness	Day_number	Solar_ang_1	Solar_ang_2	Humidity	Visibility	Wind_speed
0	2017-02-11	4:30:00	-21.24	0.212	42	-	0	0.71	NaN	7.2 km/h / 2.0 m/s
1	2017-02-11	5:00:00	-21.17	0.271	42	-14.587	0	0.71	10,0 км	10.8 km/h / 3.0 m/s
2	2017-02-11	5:30:00	-20.98	0.306	42	NaN	?	0.71	10,0 км	10.8 km/h / 3.0 m/s
3	2017-02-11	6:00:00	-21.03	0.270	42	-14.587	0	0.71	10,0 км	10.8 km/h / 3.0 m/s
4	2017-02-11	6:30:00	-21.24	0.239	42	-14.587	0	0.71	10,0 км	10.8 km/h / 3.0 m/s

Рисунок 2.2 – Результат чтения файла

2. Приведение столбцов к нужному формату. Из рисунка 2.2 видно, что часть работы сделана автоматически: часть столбцов с численными данными прочитана сразу в нужном численном формате. Можно изменить функцию чтения, указав в ней варианты заполнения пропусков (видно, что есть прочерки и знаки вопроса). Проверить формат прочитанных данных можно с помощью поля dtypes (рисунок 2.3)

input_df = pd.read_excel('/content/solar_raw_data.xlsx',

na_values = ['?', '-'])

print(input_df.dtypes)

input_df.head()

Видно, что два самых правых столбца имеют тип данных «object», то есть не воспринимаются как численные. И столбец «Time» тоже. Как правило, такие ошибки формата исправляются путем разбиения столба на несколько и, при необходимости, заменой символов.

Создадим пустую таблицу для занесения в нее данных с исправлениями:

df_cleared = pd.DataFrame()

```
input_df = pd.read_excel('/content/solar_raw_data.xlsx',
  na_values = ['?', '-'])
print(input_df.dtypes)
input_df.head()
```

```
Data          datetime64[ns]
Time                  object
Temp                 float64
Lightness            float64
Day_number             int64
Solar_ang_1          float64
Solar_ang_2          float64
Humidity             float64
Visibility            object
Wind_speed            object
dtype: object
```

	Data	Time	Temp	Lightness	Day_number	Solar_ang_1	Solar_ang_2	Humidity	Visibility	Wind_speed
0	2017-02-11	4:30:00	-21.24	0.212	42	NaN	0.0	0.71	NaN	7.2 km/h / 2.0 m/s
1	2017-02-11	5:00:00	-21.17	0.271	42	-14.586996	0.0	0.71	10,0 км	10.8 km/h / 3.0 m/s
2	2017-02-11	5:30:00	-20.98	0.306	42	NaN	NaN	0.71	10,0 км	10.8 km/h / 3.0 m/s
3	2017-02-11	6:00:00	-21.03	0.270	42	-14.586996	0.0	0.71	10,0 км	10.8 km/h / 3.0 m/s
4	2017-02-11	6:30:00	-21.24	0.239	42	-14.586996	0.0	0.71	10,0 км	10.8 km/h / 3.0 m/s

Рисунок 2.3 – Результат чтения файла с указанием пропусков и выводом типов данных

Чтобы ничего не пропустить, будем анализировать столбец за столбцом. Столбец «Data» не требует преобразований, а столбец «Time» можно разбить на часы и минуты. Перед разбиением нужно перевести столбец в формат строки, затем разделить по символу «:» и перевести в целочисленный тип.

```
df_cleared['Data'] = input_df['Data']
tmp_df = input_df['Time'].str.split(':', expand = True)
tmp_df = tmp_df.astype('int32')
df_cleared['Hour'] = tmp_df[0]
df_cleared['Minute'] = tmp_df[1]
```

Далее идет ряд столбцов, которые не требуют действий.

```
df_cleared['Temp'] = input_df['Temp']
df_cleared['Lightness'] = input_df['Lightness']
df_cleared['Day_number'] = input_df['Day_number']
df_cleared['Solar_ang_1'] = input_df['Solar_ang_1']
df_cleared['Solar_ang_2'] = input_df['Solar_ang_2']
df_cleared['Humidity'] = input_df['Humidity']
```

Оставшиеся столбцы нужно обработать следующим образом (разбить, взять нужную часть, при необходимости заменить запятые на точки, кроме того, км/ч переведем в м/с делением на 3.6):

```
df_cleared['Visibility'] = input_df['Visibility'].\
 str.split(' ', expand = True)[0].str.replace(',', '.').\
 astype('float')
df_cleared['Wind_speed'] = input_df['Wind_speed'].\
 str.split('km', expand = True)[0].astype('float') / 3.6
```

Результат показан на рисунке 2.4

	Data	Hour	Minute	Temp	Lightness	Day_number	Solar_ang_1	Solar_ang_2	Humidity	Visibility	Wind_speed
0	2017-02-11	4	30	-21.24	0.212	42	NaN	0.0	0.71	NaN	2.0
1	2017-02-11	5	0	-21.17	0.271	42	-14.586996	0.0	0.71	10.0	3.0
2	2017-02-11	5	30	-20.98	0.306	42	NaN	NaN	0.71	10.0	3.0
3	2017-02-11	6	0	-21.03	0.270	42	-14.586996	0.0	0.71	10.0	3.0
4	2017-02-11	6	30	-21.24	0.239	42	-14.586996	0.0	0.71	10.0	3.0

Рисунок 2.4 – Таблица после преобразования форматов столбцов

Можно вывести размерность таблицы и проанализировать процент пропусков по столбцам (рисунок 2.5):

```
print(df_cleared.shape)
print('% NaN')
(df_cleared.isnull().sum() / len(df_cleared.index) * 100)
```

```
1 print('shape:', df_cleared.shape)
2 print('% NaN:')
3 (df_cleared.isnull().sum() / len(df_cleared.index) * 100)
```

```
shape: (179, 11)
% NaN:
Data            0.000000
Hour            0.000000
Minute          0.000000
Temp            0.000000
Lightness       0.000000
Day_number      0.000000
Solar_ang_1     1.117318
Solar_ang_2     0.558659
Humidity        7.821229
Visibility     34.636872
Wind_speed      7.821229
dtype: float64
```

Рисунок 2.5 – Размерность и проценты пропусков

3. Удаление пропусков. Чтобы удалить все строки, содержащие хотя бы один пропуск, можно использовать функцию dropna (также можно выполнить не удаление, а замену пропусков, на определенное значение, например, медианное):

```
df_cleared_2 = df_cleared.dropna()
print(df_cleared_2.shape)
df_cleared_2.head()
```

Результат показан на рисунке 2.6.

```
(116, 11)
```

	Data	Hour	Minute	Temp	Lightness	Day_number	Solar_ang_1	Solar_ang_2	Humidity	Visibility	Wind_speed
1	2017-02-11	5	0	-21.17	0.271	42	-14.586996	0.0	0.71	10.0	10.8
3	2017-02-11	6	0	-21.03	0.270	42	-14.586996	0.0	0.71	10.0	10.8
4	2017-02-11	6	30	-21.24	0.239	42	-14.586996	0.0	0.71	10.0	10.8
5	2017-02-11	7	0	-21.02	0.231	42	-14.586996	0.0	0.71	10.0	7.2
6	2017-02-11	7	30	-21.07	0.232	42	-14.586996	0.0	0.71	10.0	7.2

Рисунок 2.6 – Таблица без пропусков

4. Удаление выбросов. Для визуального анализа выбросов можно построить «ящики с усами» (рисунок 2.7):

```
df_cleared_2.boxplot(figsize=(16, 12))
```

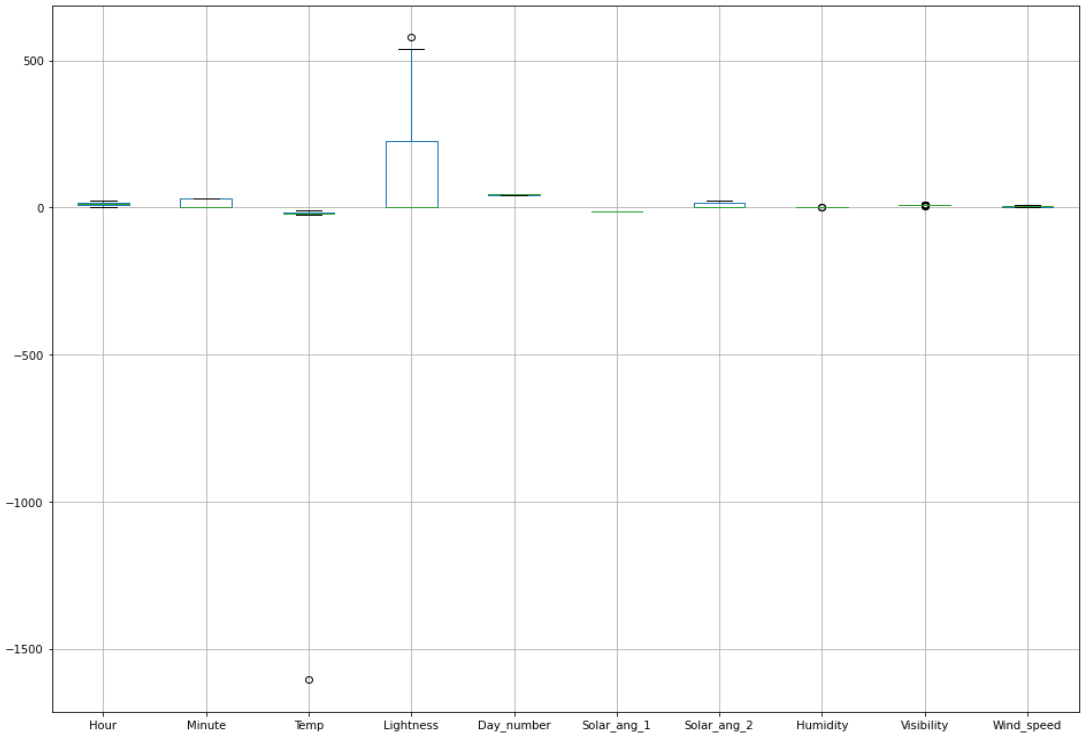

Рисунок 2.7 – Визуальный анализ выбросов

Можно построить такую диаграмму только для температуры:

```
df_cleared_2[['Temp']].boxplot(figsize=(20, 8))
```

Удалим найденный выброс (результат на рисунке 2.8):

```
df_cleared_3 = df_cleared_2.drop(
        df_cleared_2[df_cleared_2['Temp'] < -200].index)
df_cleared_3.boxplot(figsize=(16, 12))
```

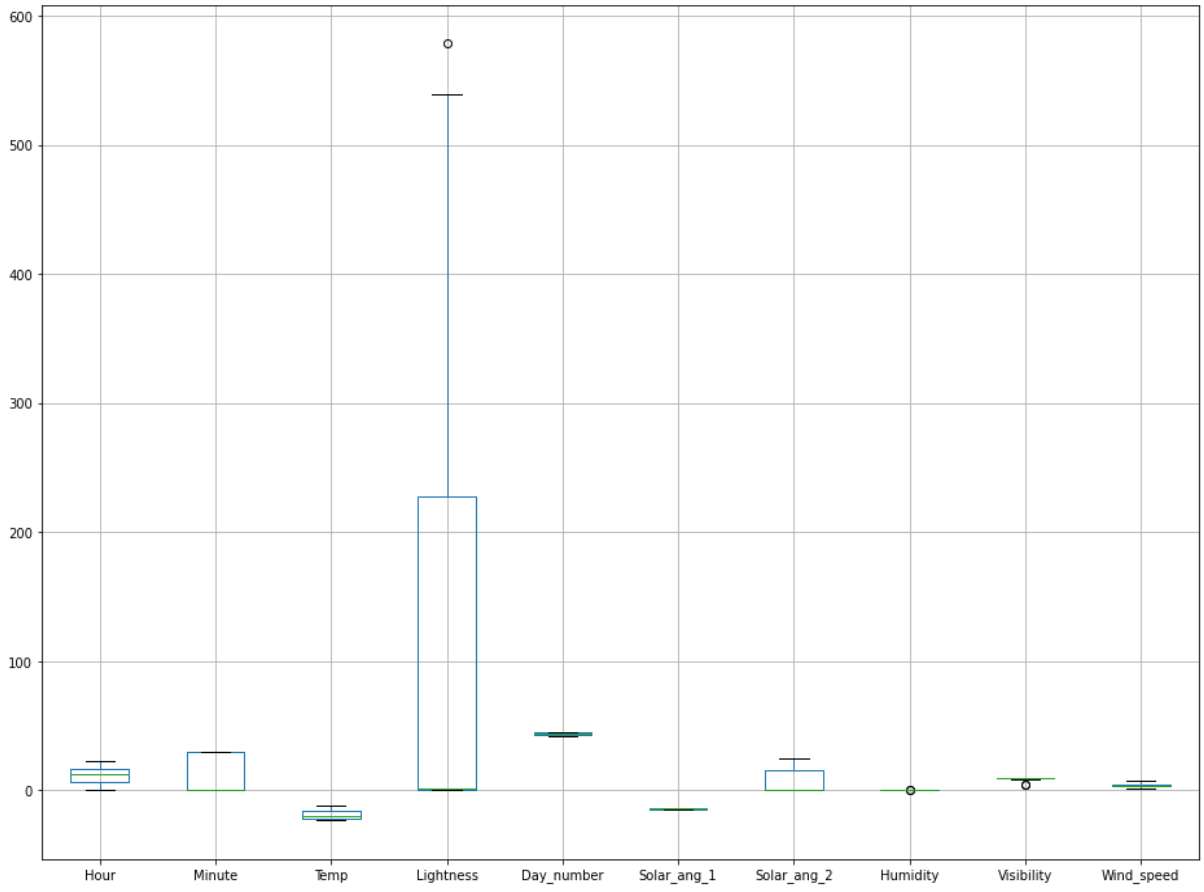

Рисунок 2.8 – Визуальный анализ выбросов

Pandas позволяет легко получать статистические показатели данных и визуализировать их (хотя для визуализации существуют более мощные библиотеки), пример показан на рисунках 2.9 и 2.10.

```
fig = df_cleared_3[['Visibility', 'Lightness', 'Wind_speed'
  , 'Humidity']].hist(figsize = (6, 6))
df_cleared_3.describe()
```

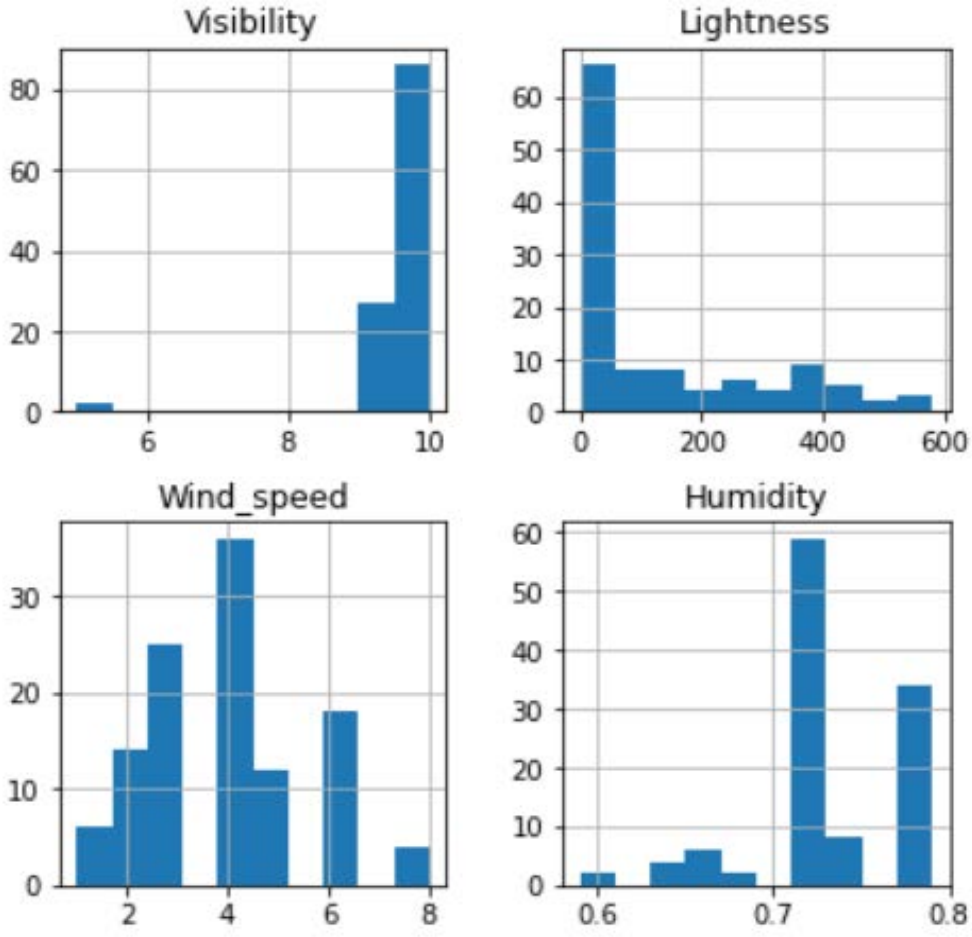

Рисунок 2.9 – Статистические показатели данных

```
1 df_cleared_3.describe()
```

	Hour	Minute	Temp	Lightness	Day_number	Solar_ang_1	Solar_ang_2	Humidity	Visibility	Wind_speed
count	115.000000	115.000000	115.000000	115.000000	115.000000	115.000000	115.000000	115.000000	115.000000	115.000000
mean	12.260870	14.869565	-18.332348	121.391200	43.626087	-14.064948	7.403057	0.727478	9.678261	3.939130
std	6.317737	15.065076	3.408252	169.517519	1.087899	0.351137	9.342989	0.043283	0.755655	1.574395
min	0.000000	0.000000	-23.130000	0.161000	42.000000	-14.586996	0.000000	0.590000	5.000000	1.000000
25%	7.000000	0.000000	-21.450000	0.247000	43.000000	-14.268783	0.000000	0.710000	9.500000	3.000000
50%	13.000000	0.000000	-19.430000	1.105000	44.000000	-13.946341	0.000000	0.720000	10.000000	4.000000
75%	17.000000	30.000000	-15.520000	227.924500	45.000000	-13.619766	15.687764	0.780000	10.000000	5.000000
max	23.000000	30.000000	-11.310000	578.659000	45.000000	-13.619766	25.080257	0.790000	10.000000	8.000000

Рисунок 2.10 – Визуализация очищенных данных в табличной форме

5. Запись полученной очищенной выборки в файл. Сохранение матрицы в XLSX файл делается одной функцией to_excel:

df_cleared_3.to_excel('solar_cleared_data.xlsx')

2.4 Заключение по главе 2

1. Выбраны инструментальные средства программной реализации: язык программирования Python и среда разработки Jupyter Notebook.

2. Описаны шаги предварительной обработки исходных данных для обучения и тестирования модели прогнозирования генерации солнечной электрической станции по метеорологическим параметрам.

3. Разработан код для реализации предварительной обработки и апробирован на выборке реальных данных.

3 РАЗРАБОТКА АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПРОГНОЗИРОВАНИЯ ВЫРАБОТКИ СОЛНЕЧНОЙ ЭЛЕКТРИЧЕСКОЙ СТАНЦИИ

3.1 Описание регрессионных моделей машинного обучения

Существует множество алгоритмов машинного обучения, применимых для рассматриваемой задаче прогнозирования выработки солнечной электрической станции. Ниже представлено описание наиболее часто используемых [22-26].

Линейная регрессия – линейная математическая модель зависимость вида:

$$y^*_{\square}(X) = a_1x_1 + a_2x_2 + ... + a_mx_m + b, \quad (3.1)$$

где $y^*_{\square}$ – прогнозируемое значение, X – вектор из m входных признаков, A – вектор коэффициентов регрессии, b – смещение.

Если линейная регрессия применяется вместе с формированием полиномиальных признаков, то ее называют полиномиальной. Например, полиномиальная регрессия 2-го порядка:

$$y^*_{\square}(X) = a_1x_1 + a_2x_2 + a_{11}x_1x_1 + a_{12}x_1x_2 + ... + a_{mm}x_mx_m + b, \quad (3.1)$$

Алгоритм kNN (k-nearest neighbors) заключается в том, что из обучающей выборки выбирается k объектов, наиболее похожих по своим признакам X на объект, для которого нужно определить y, затем происходит усреднение:

$$y^*_{\square}(X) = 1/k \sum_{j=1}^{k} sort(y(X), ||X - X_i||), \quad (3.2)$$

Метод опорных векторов (support vector machine, SVM) – метод, модифицирующий линейную регрессию. Он основан на поиске гиперплоскости, которая минимизирует сумму квадратов отклонений между предсказанными значениями и фактическими значениями обучающих примеров с наибольшим зазором между ними. Это позволяет достичь высокой обобщающей способности модели. Обучающие примеры, которые находятся на границе зазора или внутри него, называются опорными векторами. Кроме

того, SVM позволяет использовать функции ядра для нелинейного преобразования данных в пространство с более высокой размерностью, что подобно процессу генерации новых признаков, но для SVM этот процесс тесно связан с процессом самого обучения.

Модели на базе деревьев решений представляют собой иерархические модели в виде вложенных правил вида «ЕСЛИ … ТО … ИНАЧЕ». При обучении дерева решений как для классификации, так и для регрессии деревья строятся по принципу рекурсивного разбиения обучающей выборки на две части. В задачах классификации нижние узлы дерева решений – это классы, а в задаче регрессии – это значения искомой величины. Дерево строится сверху вниз. Для построения каждого очередного правила делается статистическая оценка, какое правило сильнее всего снизит уровень неопределенности.

Одно дерево можно представить в виде:

$$y^{*} = t(X), \tag{3.3}$$

где t – иерархическая система правил, каждое из которых сравнивает значение определенного признака с пороговым значением.

Деревья решений часто объединяют в ансамбли для повышения точности и устойчивости к переобучению отдельного регрессионного дерева.

Два наиболее распространенных ансамблевых метода: бэггинг (bagging) и бустинг (boosting) [27]. Бустинг, как правило, повышает точность базовой модели, снижает ошибку (bias). Бэггинг же, как правило, полезен, если нужно повысить стабильность базовой модели, снизить ее чувствительность к небольшим изменениям в данных (то, что в англоязычной литературе называется variance), иными словами, снизить склонность модели к переобучению (overfitting).

Линейная регрессия – это статистическая модель, используемая для аппроксимации зависимости между набором признаков (независимых переменных) и целевой переменной (зависимой переменной). Линейная регрессия представляет собой математическую модель, где зависимость между признаками и целевой переменной описывается линейной функцией:

взвешенной суммой входных признаков и свободного члена. Для определения параметров линейной регрессии используются различные методы, такие как метод наименьших квадратов, максимальное правдоподобие и другие.

Метод опорных векторов (SVM) является популярным алгоритмом машинного обучения, существенно модифицирующим модель логистической и линейной регрессии. SVM стремится найти гиперплоскость, которая максимально разделяет обучающие примеры разных классов с наибольшим зазором (маргинальным ядром) между ними. Это позволяет достичь высокой обобщающей способности модели. Обучающие примеры, которые находятся на границе маргинального ядра или внутри него, называются опорными векторами. Они играют ключевую роль в определении гиперплоскости разделения и определяют границы классов. SVM позволяет использовать ядерную функцию для нелинейного преобразования данных в пространство с более высокой размерностью, где линейное разделение может быть достигнуто. Это позволяет моделировать сложные нелинейные зависимости между признаками. Также SVM включает параметр регуляризации, который позволяет балансировать между точностью модели и ее обобщающей способностью. При переходе к задаче регрессии, вместо разделения классов в SVM стремится найти гиперплоскость, которая минимизирует сумму квадратов отклонений между предсказанными значениями и фактическими значениями обучающих примеров.

Отдельный большой класс моделей – нейросетевые. Они отличаются многообразием архитектур, среди которых можно выделить ряд наиболее подходящих для задач регрессии.

Одной из самых простых и базовых нейросетевых архитектур для регрессии является многослойный персептрон. Модель состоит из входного слоя, одного или нескольких скрытых слоев и выходного слоя. В каждом слое, кроме входного, применяются нелинейные активационные функции, такие как ReLU или гиперболический тангенс, которые позволяют модели обучаться сложным немонотонным зависимостям между входами и выходами.

Рекуррентные нейронные сети эффективны для работы с последовательностями данных, такими как временные ряды, где порядок данных важен. RNN используют циклические связи между нейронами, позволяя информации сохраняться и передаваться от одного шага к другому.

Регрессионное дерево решений – модель вида иерархической системы правил, каждое из которых делит подвыборку на две части, на основе сравнения одного признака с пороговым значением. Регрессионное дерево решений строится рекурсивно путем разделения данных на подмножества на основе значений признаков. На каждом узле дерева выбирается наилучший признак и значение разделения, которые минимизируют дисперсию целевой переменной в каждом подмножестве. Для выбора наилучшего разделения на каждом узле дерева используются различные критерии, такие как критерий Джини. Целью является минимизация дисперсии целевой переменной в каждом подмножестве.

Существует ряд методов, повышающих точность и стабильность (устойчивость к переобучению и выбросам) деревьев решений. Они основаны на объединении деревьев в ансамбль.

Случайный лес строится из деревьев, каждое из которых обучается на случайной подвыборке данных и случайном подмножестве признаков:

$$y_i^* = \frac{1}{m}\sum_{j=1}^{m} t_j(X_i). \quad (3.4)$$

Адаптивный бустинг последовательно обучает слабые модели (в данном случае деревья решений) на взвешенных выборках данных, где веса экземпляров обновляются на каждой итерации в зависимости от предыдущих ошибок модели.

Градиентный бустинг последовательно обучает слабые модели (в данном случае деревья решений) на взвешенных выборках данных, где веса экземпляров обновляются на каждой итерации в зависимости от градиента ошибки модели.

Обе последние в списке модели можно описать как:

$$y_i^* = \sum_{j=1}^{m} w_j t_j(X_i), \quad (3.5)$$

где w_j – вес j-й модели (в данном случае дерева решений).

3.2 Описание выборки данных для прогнозирования генерации

Используемая база данных представляется в виде таблицы и содержит информацию о выработанной электрической энергии солнечной электростанции на часовых интервалах времени [28]. Помимо фактически выработанной электрической энергии в таблице представлена информация о наблюдаемых погодных условиях.

Полный список параметров:

1. Час суток.

2. Номер дня в году.

3. Угол солнечного склонения (УСС) – это угол между плоскостью экватора Земли и прямой, соединяющей центры Солнца и Земли. Этот угол меняется в течение года из-за наклона оси вращения Земли относительно её орбиты вокруг Солнца.

4. Солнечное время. Это система времени, основанная на положении Солнца в небе, оно отличается от стандартного или координированного всемирного времени (UTC), которое использует среднее солнечное время в Гринвиче, так как это время, определяемое моментом, когда верхний край солнечного диска пересекает местный меридиан (линию, проходящую через географический полюс и данную точку на земной поверхности).

5. Плотность потока солнечного излучения (ППСИ) на границе с атмосферой.

6. Облачность, %.

7. Коэффициент прозрачности атмосферы (КПА), %.

8. Температура воздуха, °С.

9. Влажность воздуха, %.

10. Скорость ветра на высоте 2 м от уровня земной поверхности, м/с.

11. Фактическая выработка электрической энергии, кВтч.

3.3 Проверка корректности выборки данных

После предварительной обработки данных необходимо проверить выборку еще раз. Перед этим следует выполнить подключения библиотек. В работе использовано большое количество модулей из библиотеки sklearn, кроме того, библиотеки для визуализации данных. Подключение всех библиотек:

```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from bokeh.plotting import figure, output_file, show
from bokeh.io import output_notebook
output_notebook()

from sklearn import metrics
from sklearn.model_selection import cross_val_score
from sklearn.linear_model import LinearRegression
from sklearn.tree import DecisionTreeRegressor
from sklearn.ensemble import AdaBoostRegressor
from sklearn.ensemble import RandomForestRegressor
```

Чтение исходного файла:

```
data = pd.read_excel('/content/solar_data_regression.xlsx')
data.head()
```

Проверка типов данных показывает, что все столбцы в численном формате:

```
data.dtypes
```

Коэффициенты корреляции Пирсона и Спирмена между всеми парами признаков показаны на рисунках 3.1, 3.2. Ниже приведен код для его создания.

```
plt.figure(figsize=(9, 9))
sns.heatmap(data.corr(method = 'spearman')
```

, cmap = 'RdYlGn', vmax = 1, vmin = -1, square=True

, annot=True, fmt = '.2f', cbar=False)



Рисунок 3.1 – Матрица корреляций по Пирсону

Ниже приведен код для построения интерактивного графика генерации (рисунок 3.3):

p = figure(plot_width = 800, plot_height = 400)

p.scatter(np.arange(len(data.index)), data['Generation'].values

, line_width = 2, color = 'green', legend_label = 'G')

p.line(np.arange(len(data.index)), data['Generation'].values

, line_width = 2, color = 'green', legend_label = 'G')

show(p)



Рисунок 3.2 – Матрица корреляций по Спирмену

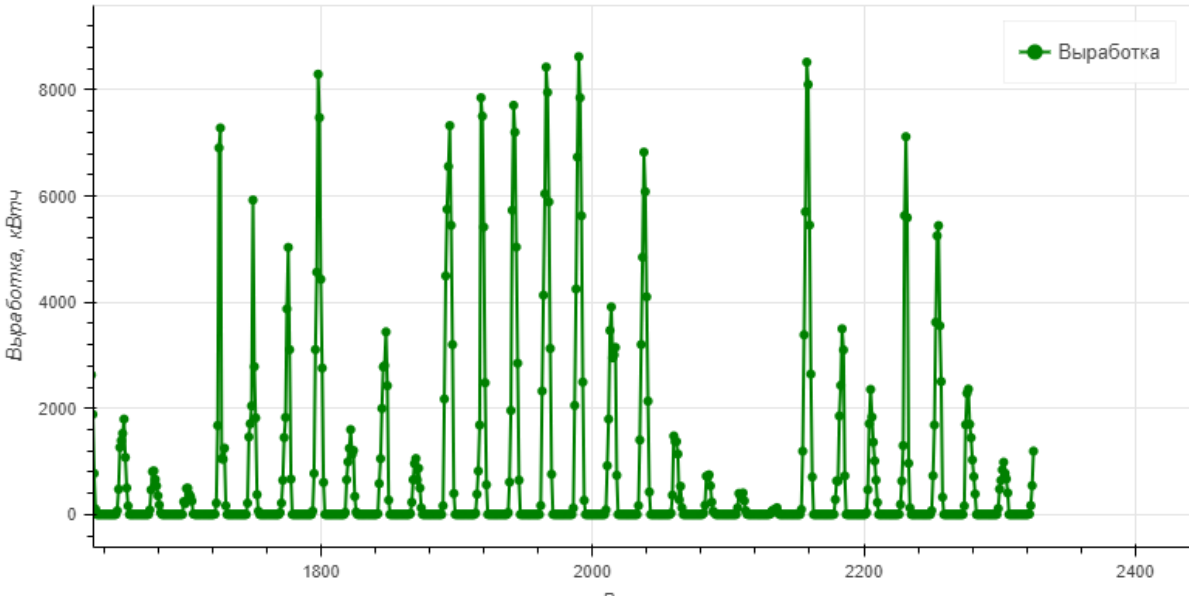

Рисунок 3.3 –График генерации (фрагмент)

На рисунках 3.4–3.18 показаны визуализации, полученные при анализе данных.

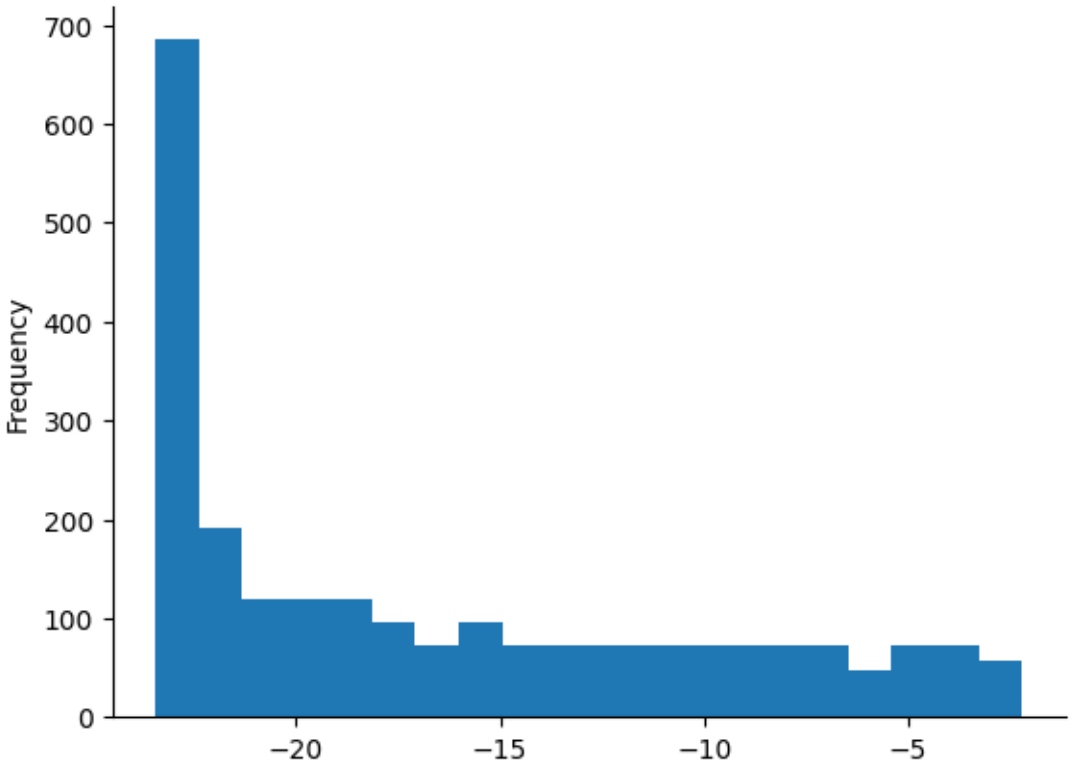

Рисунок 3.4 – Распределение значений угла солнечного склонения

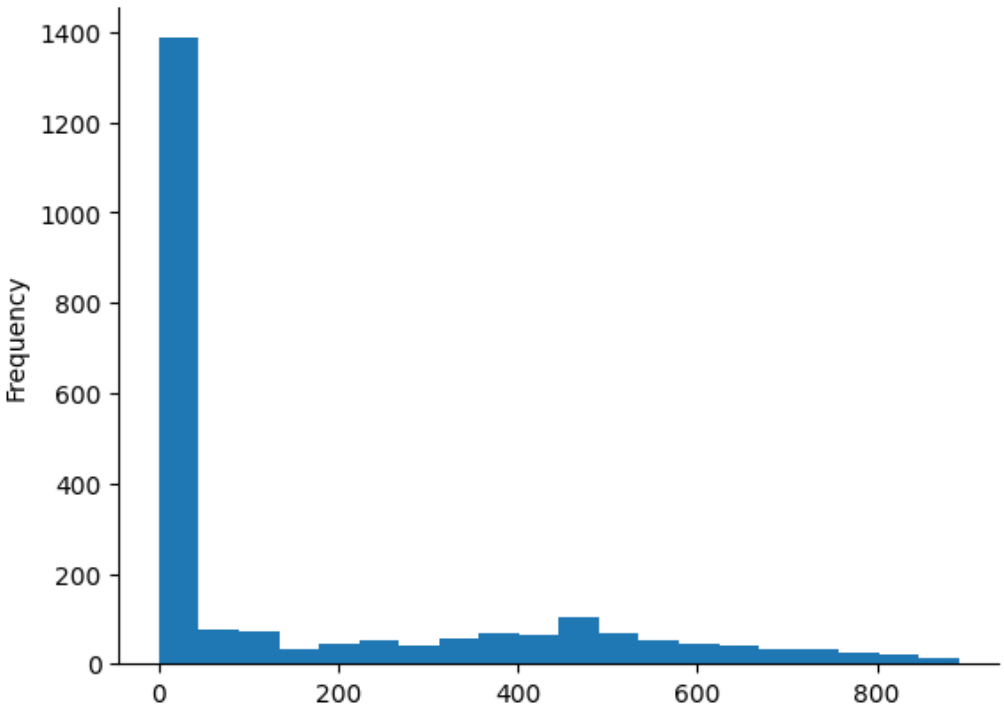

Рисунок 3.5 – Распределение значений ППСИ

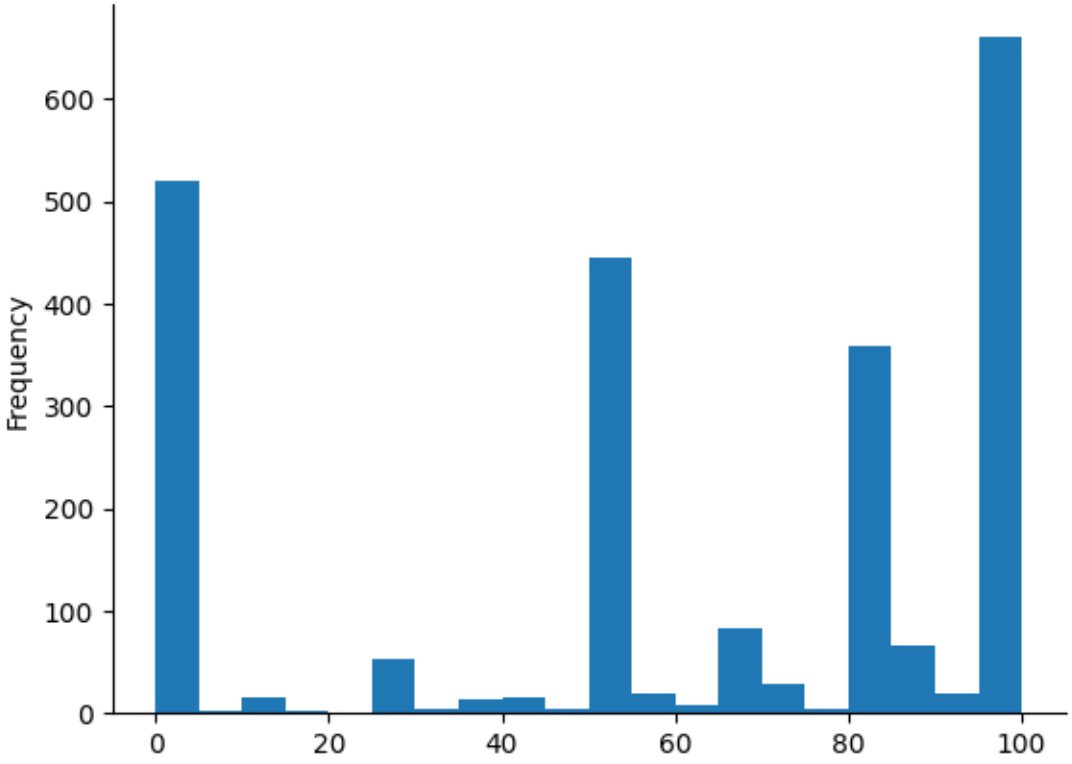

Рисунок 3.6 – Распределение значений облачности

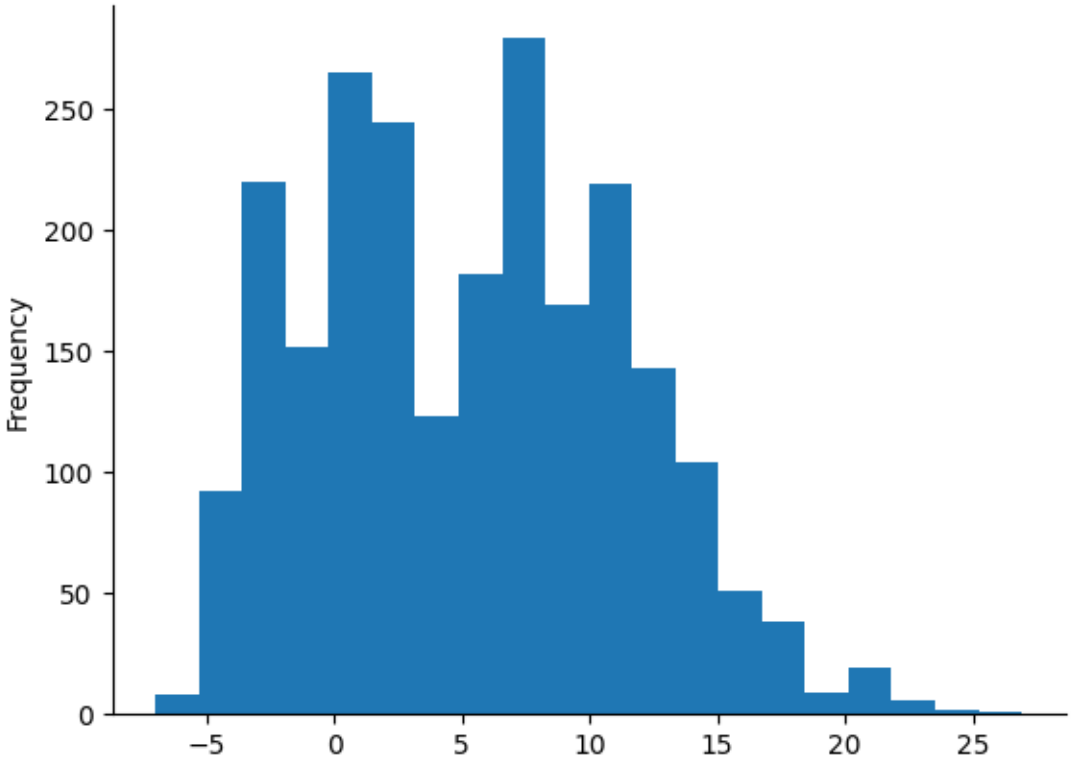

Рисунок 3.7 – Распределение значений температуры

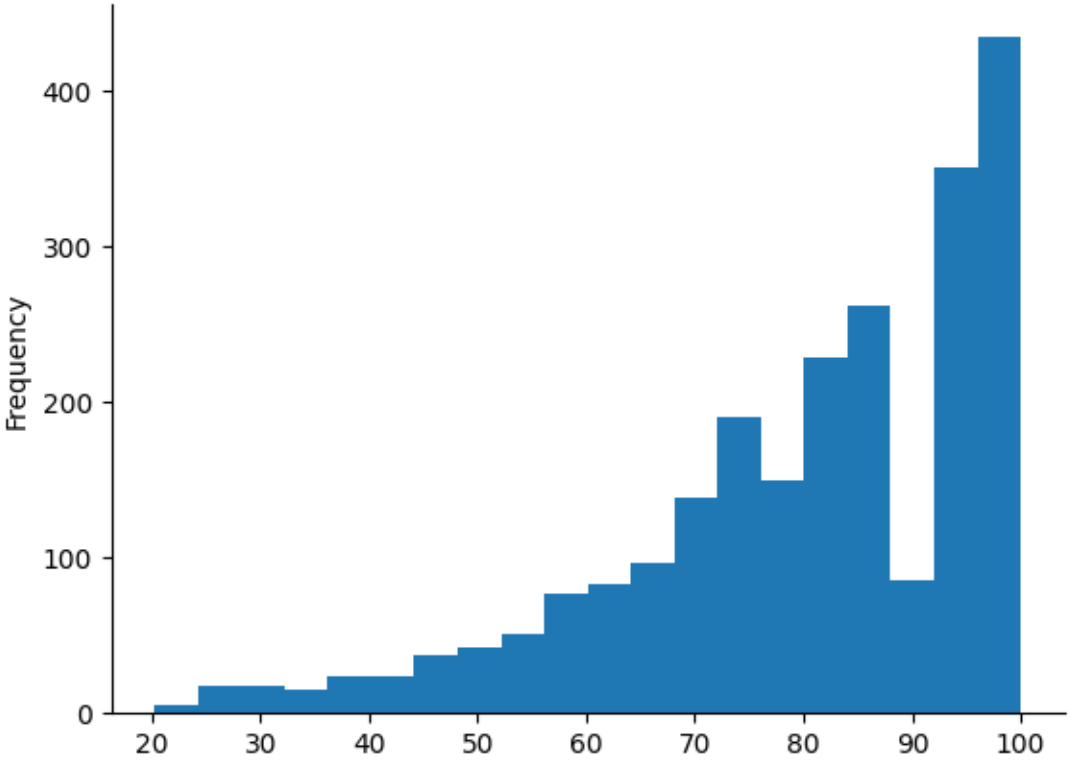

Рисунок 3.8 – Распределение значений влажности

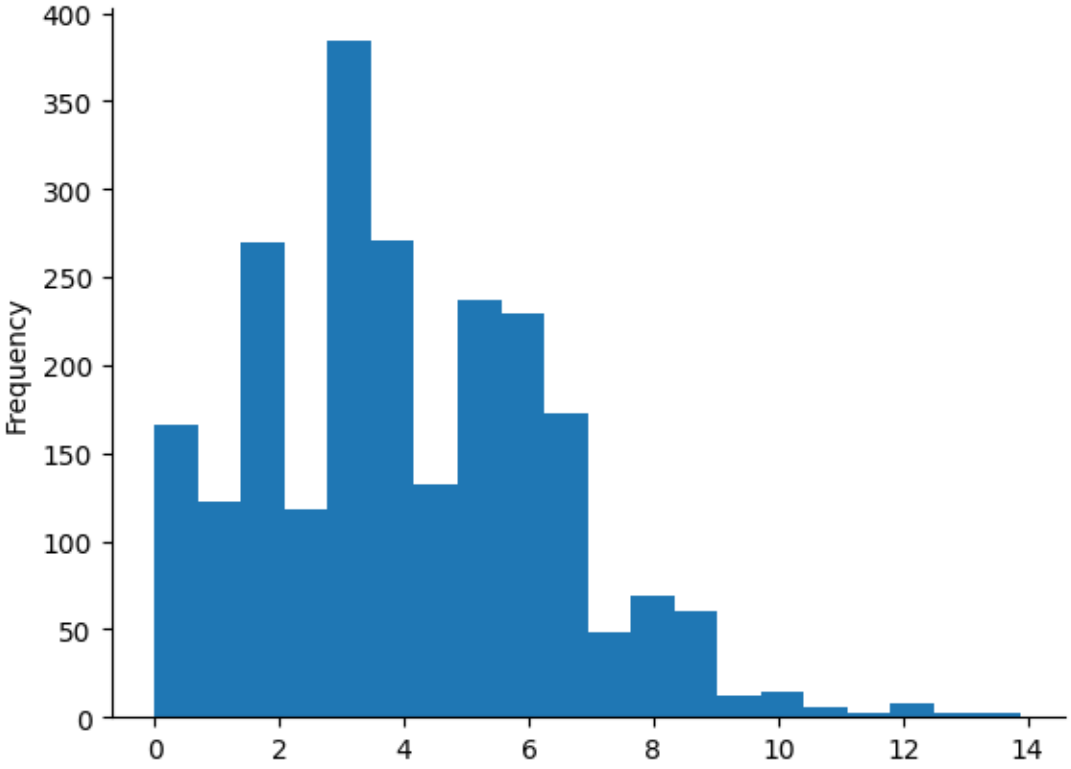

Рисунок 3.9 – Распределение скорости ветра

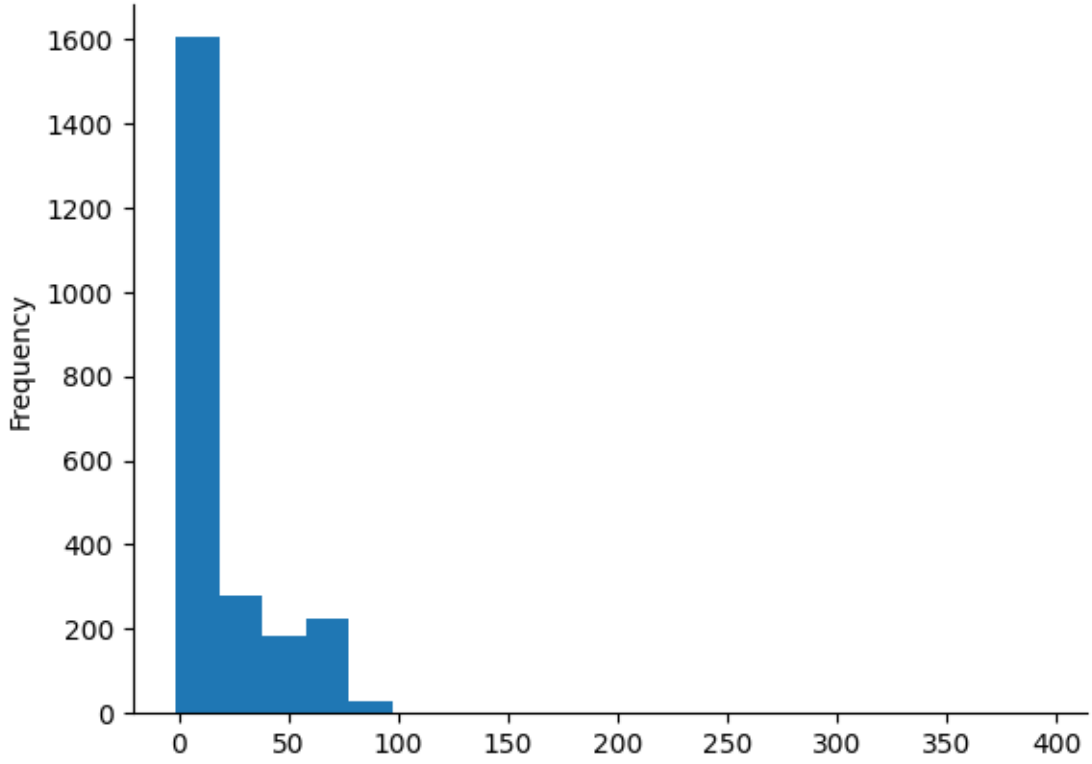

Рисунок 3.10 – Распределение КПА

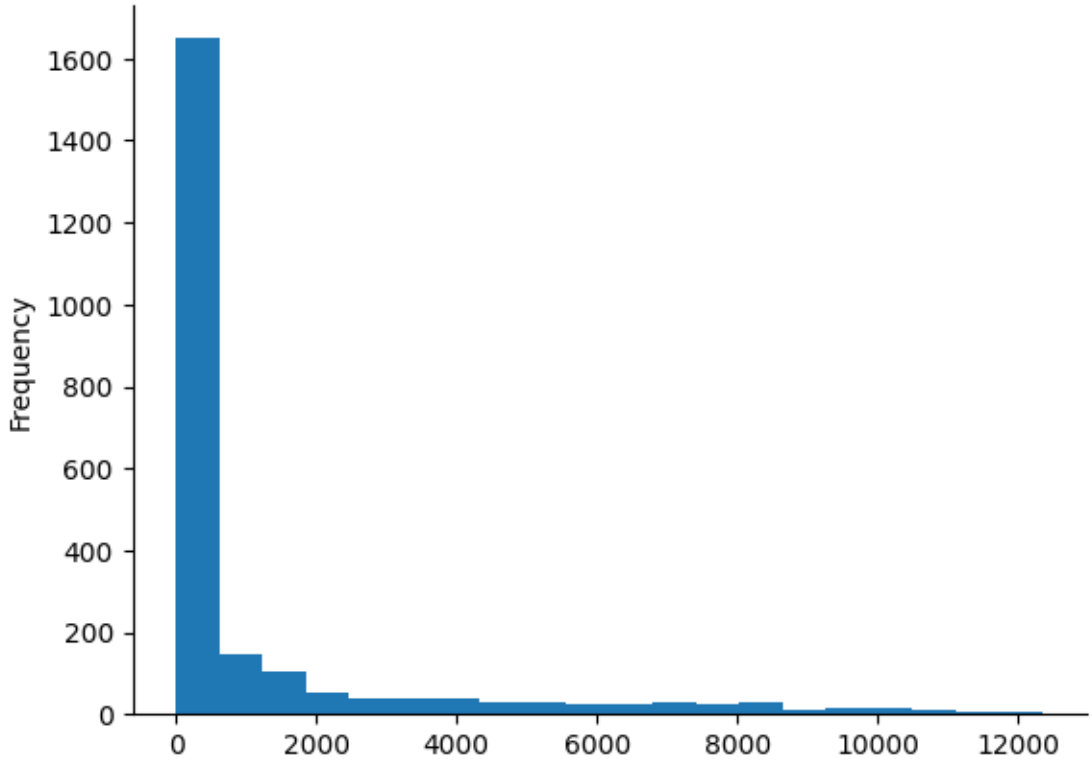

Рисунок 3.11 – Распределение выработки

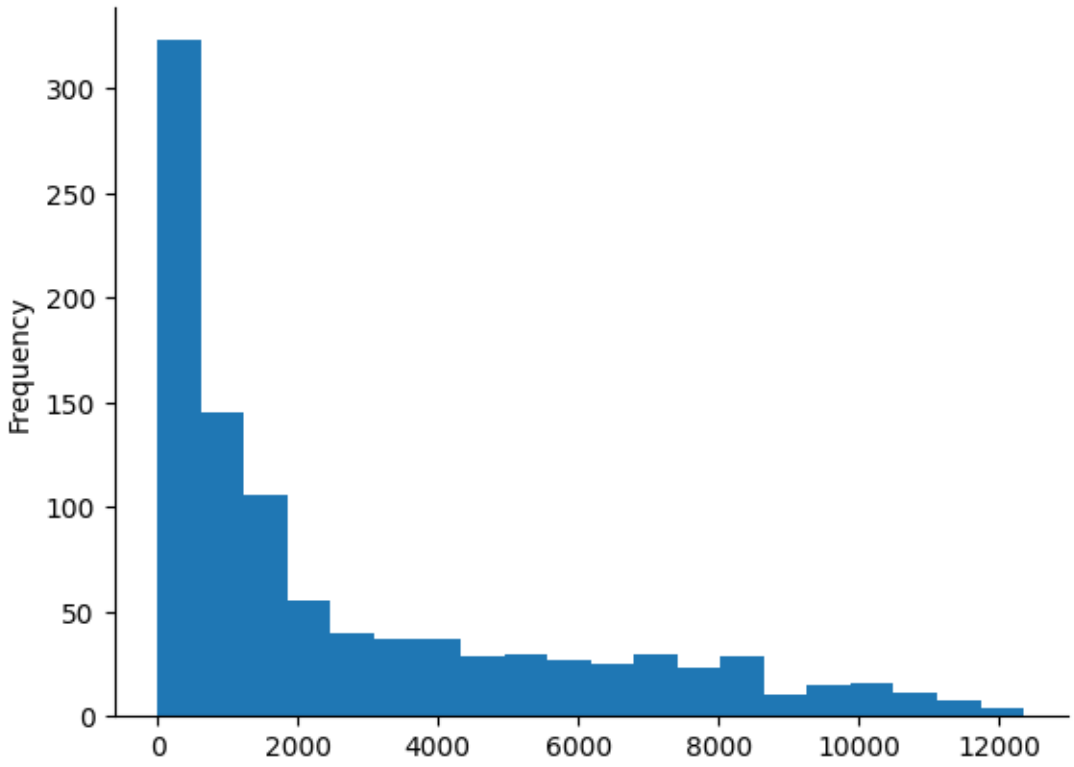

Рисунок 3.12 – Распределение выработки при значениях выше 0

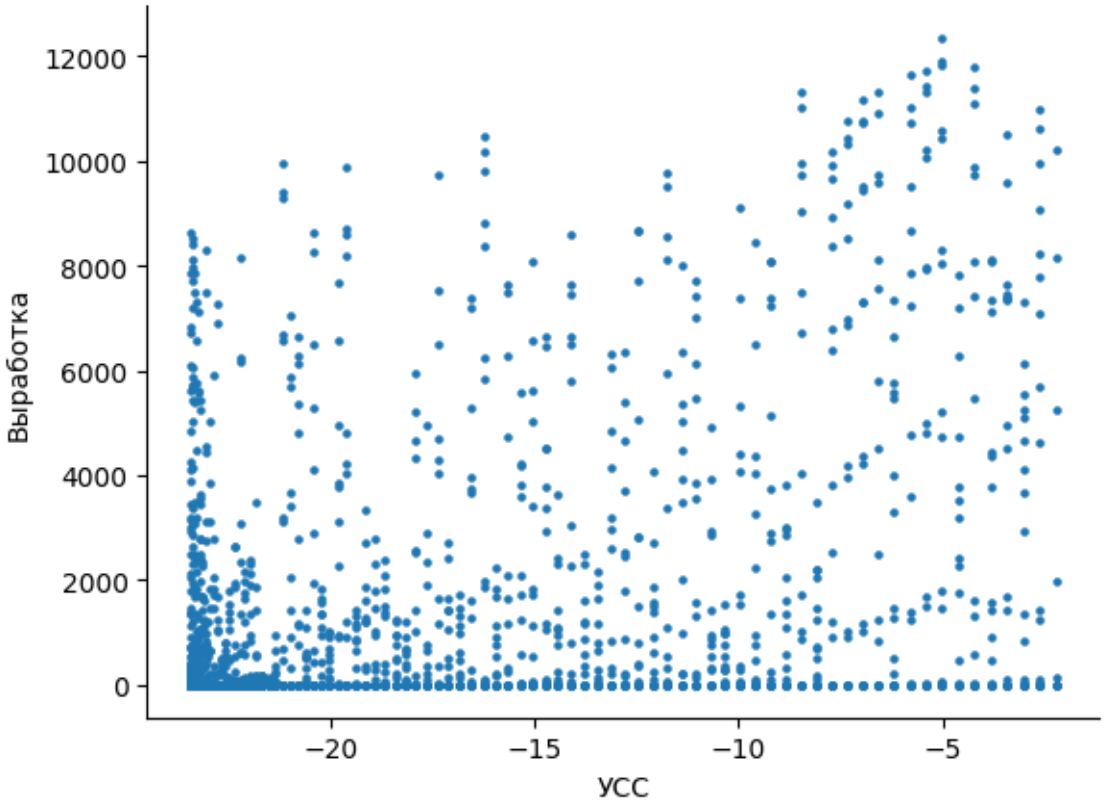

Рисунок 3.13 – Зависимость выработки от УСС

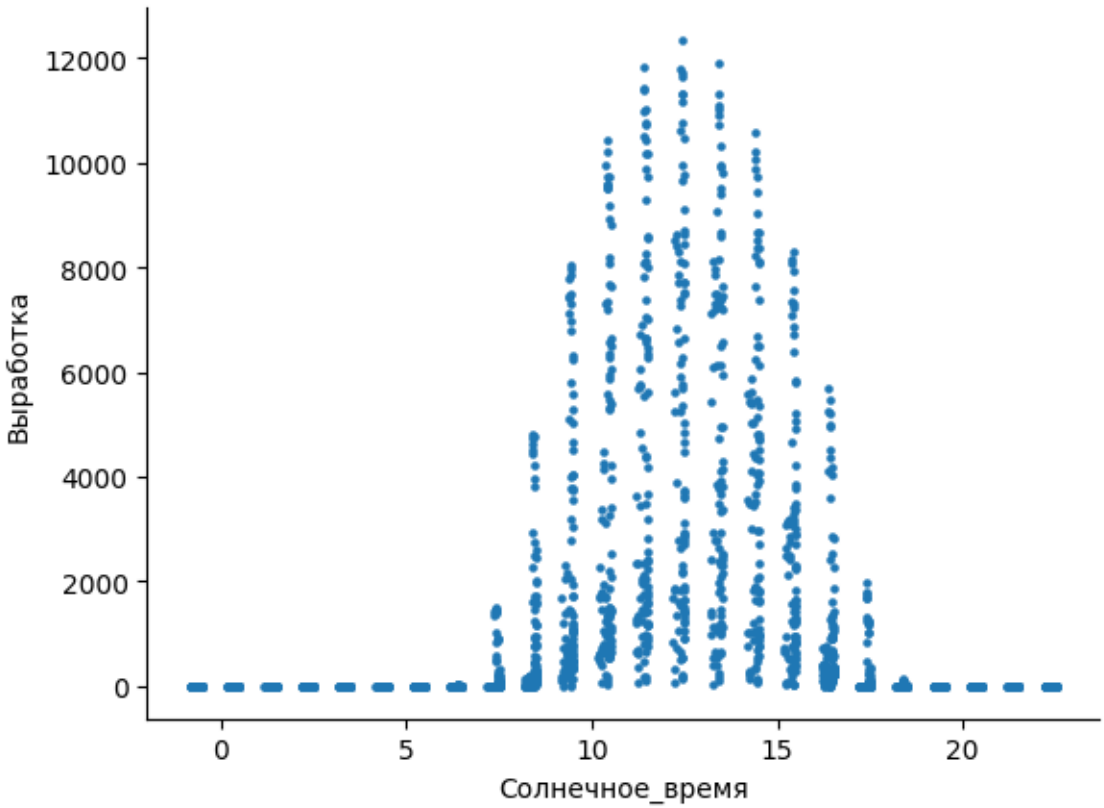

Рисунок 3.14 – Зависимость выработки от солнечного времени

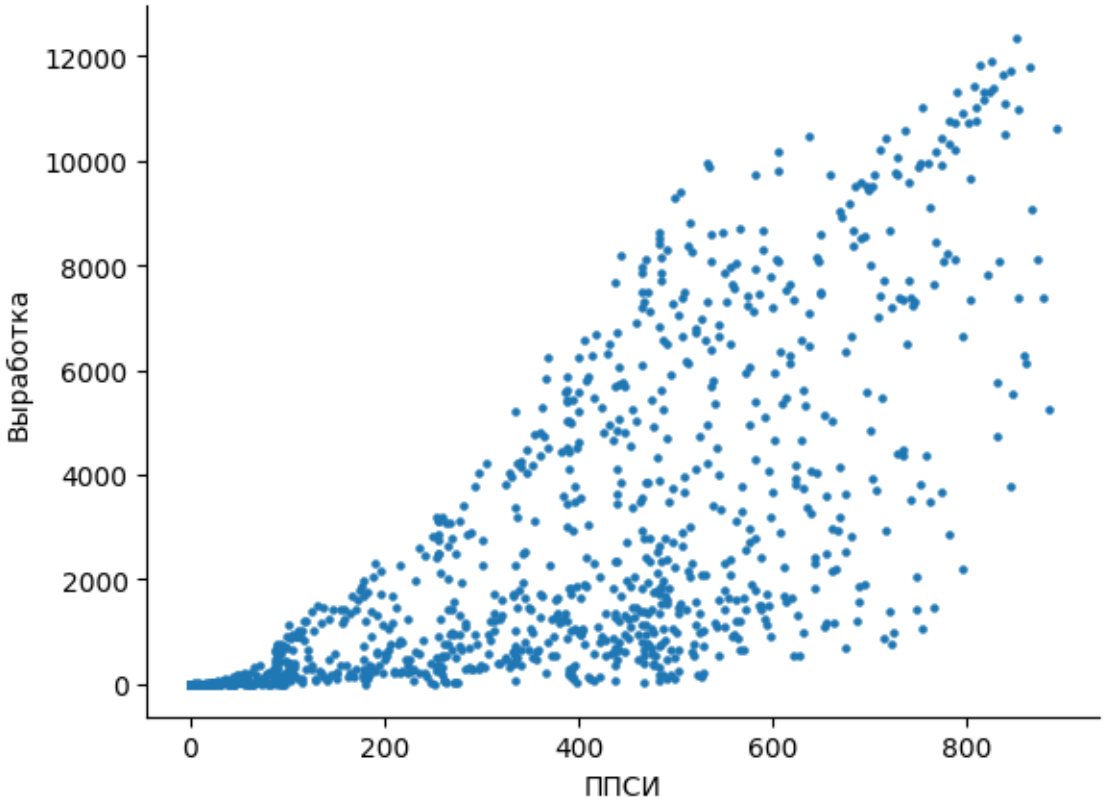

Рисунок 3.15 – Зависимость выработки от ППСИ

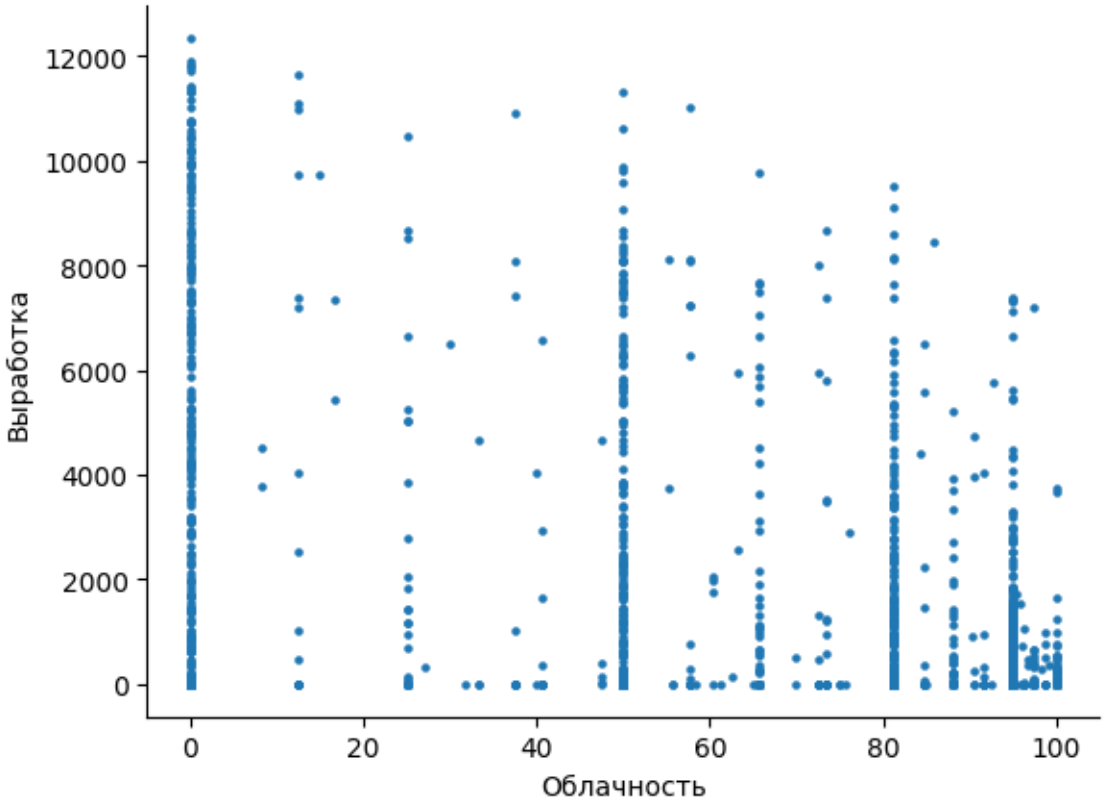

Рисунок 3.16 – Зависимость выработки от облачности

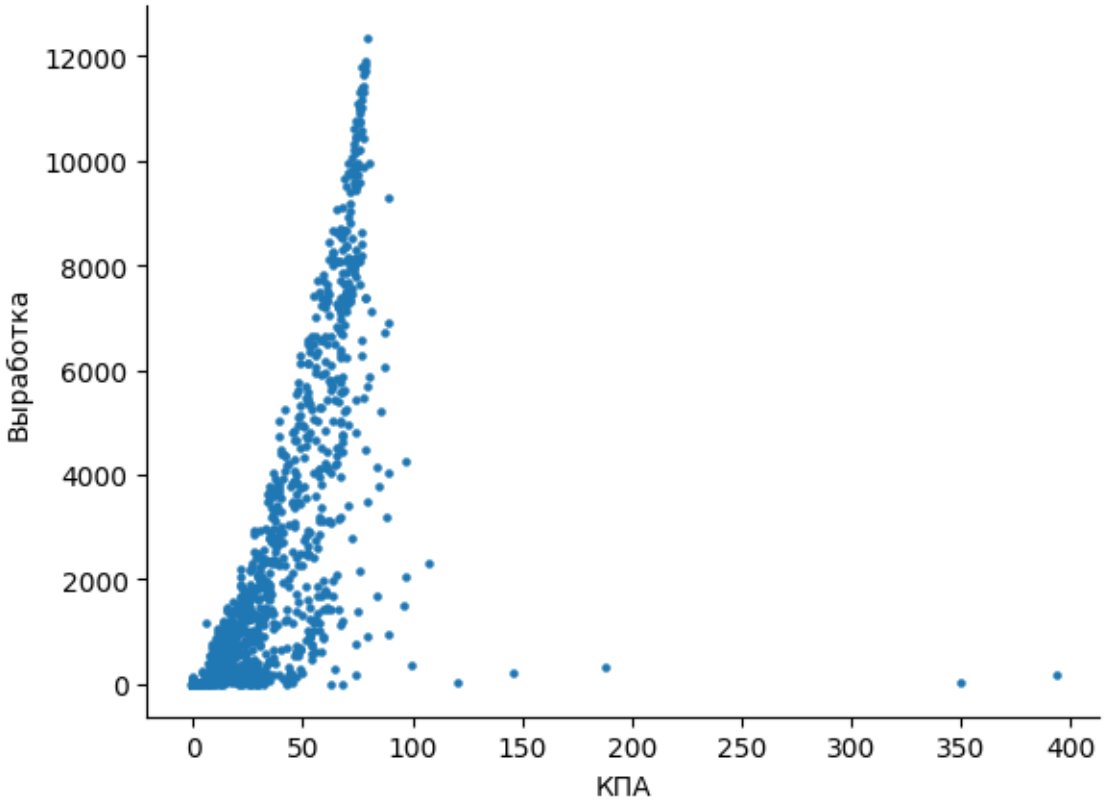

Рисунок 3.17 – Зависимость выработки от КПА

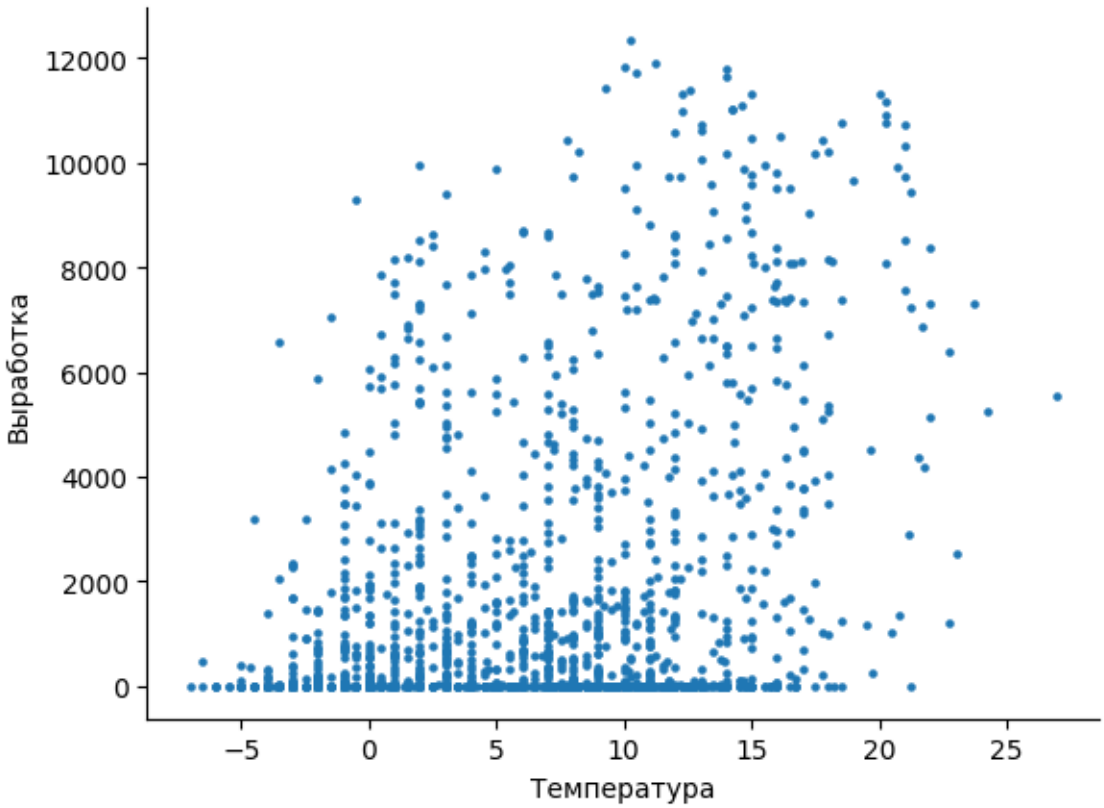

Рисунок 3.18 – Зависимость выработки от температуры

3.4 Обучение моделей

Разделим выборку на две части:

1) для обучения и валидации,

2) для итогового тестирования.

Чтобы задача была ближе к реальной, в тестирование берется не произвольная часть выборки, а 20-процентная часть, которая относится к более новым данным.

```
train_split_point = int(len(data.index) * 0.8)


X_trainval = data.iloc[: train_split_point, : -1]
y_trainval = data.iloc[ : train_split_point, -1]


X_test = data.iloc[train_split_point : , : -1]
y_test = data.iloc[train_split_point : , -1]
```

```
print(data.shape, X_trainval.shape, X_test.shape)
```

Поскольку далее одни и те же действия будут выполняться разными моделями, код оформлен в функции.

Первая функция выполняет процесс кросс-валидации для заданной модели и данных. Использовано разбиение на 5 разных наборов (cv = 5). Метрика качества – RMSE, корень из среднего квадрата ошибки:

Кроме того, использованы следующие показатели точности:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left|y_i^* - y_i\right|^2}, \quad . \tag{3.1}$$

В sklearn принято, что чем метрика больше, тем лучше, поэтому используется RMSE, умноженное на -1. И при выводе показателя на экран метрика тоже умножается на -1, чтобы получить положительный RMSE. При выводе ошибки округляются до целого.

```
def create_evaluate_model(model, X, y):
  scores = cross_val_score(model, X, y, cv=5
    , scoring='neg_root_mean_squared_error')
  print('RMSE on validation sets:', scores * -1, 'kWh')
  print('Mean RMSE on validation sets::'
    , int(scores.mean() * -1), 'kWh')

  model.fit(X, y)
  print('Mean RMSE on train set:'
     , int(calc_rmse(model, X, y)), 'kWh')

  return model
```

Функция для расчета RMSE модели:

```
def calc_rmse(model, X, y):
  return np.sqrt(metrics.mean_squared_error(
      model.predict(X), y))
```

Функция для визуализации работы модели:

```
def visualize(model, X, y):
  p = figure(plot_width = 800, plot_height = 400)

  times = np.arange(y.shape[0])
  p.scatter(times, y, line_width = 2, color = 'green'
  , legend_label = 'G')
  p.line(times, y, line_width = 2, color = 'green'
  , legend_label = 'G')
  p.scatter(times, model.predict(X), line_width = 2
          , color = 'red', legend_label = 'P')
  p.line(times, model.predict(X), line_width = 2
        , color = 'red', legend_label = 'P')

  show(p)
```

Применим функции для построения и оценки самой простой регрессионной модели – линейной (результаты на рисунке 3.19):

```
model = LinearRegression()
model = create_evaluate_model(
     model, X_trainval, y_trainval)
visualize(model, X_trainval, y_trainval)
```

После того, как модель и параметры модели выбраны с помощью кросс-валидации, ее нужно проверить на тестовой выборке:

```
print('Mean RMSE on test set:', int(calc_rmse(
   model, X_test, y_test)), 'kWh')
visualize(model, X_test, y_test)
```

Видно, что точность модели низкая, так как она не справляется тем, что ночные часы имеют нулевую выработку, а дневные – нет.

Использование дерева решений:

```
model = DecisionTreeRegressor(max_depth = 3)
model = create_evaluate_model(
```

model, X_trainval, y_trainval)

visualize(model, X_trainval, y_trainval)

Точность стала намного выше (рисунок 3.20). На рисунке 3.21 показан результат для дерева с глубиной 8. Заметен разрыв между точностью на обучающей выборке и результатами оценки с помощью кросс-валидации (на обучающей ошибка примерно в 4 раза ниже). Это говорит о переобучении. Результат при глубине 5 показан на рисунке 3.22.

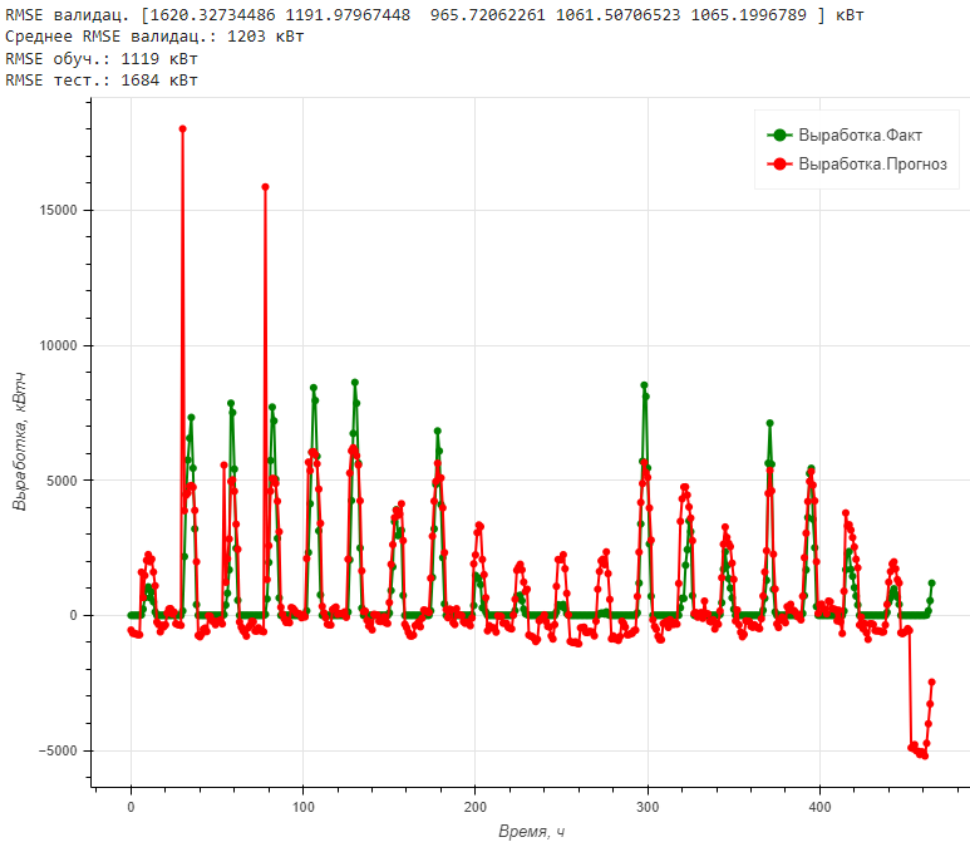

Рисунок 3.19 – Результат линейной модели

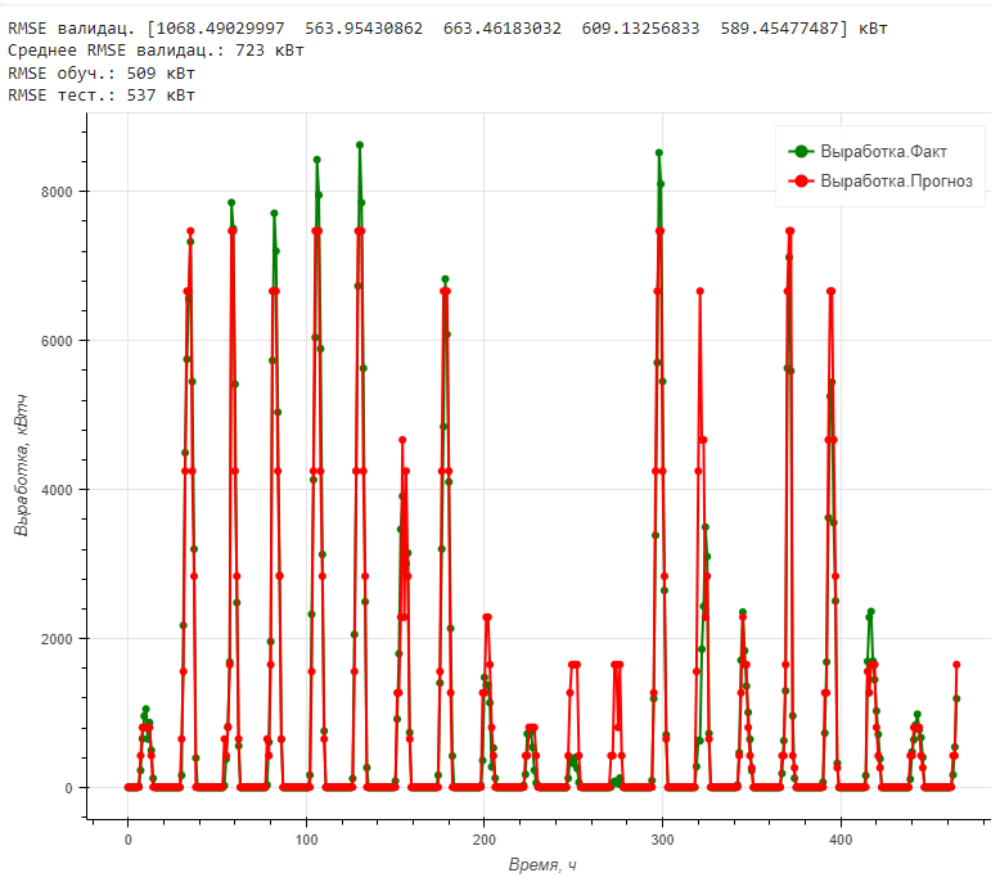

Рисунок 3.20 – Результат дерева решений с глубиной 3

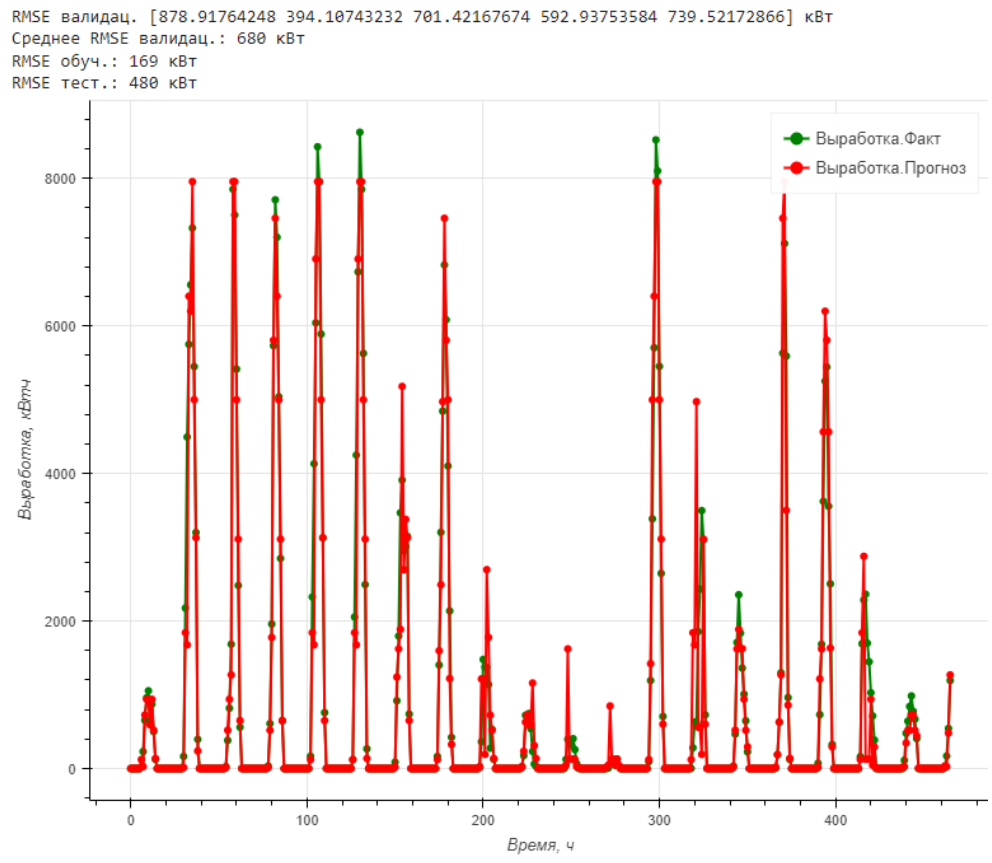

Рисунок 3.21 – Результат дерева решений с глубиной 8

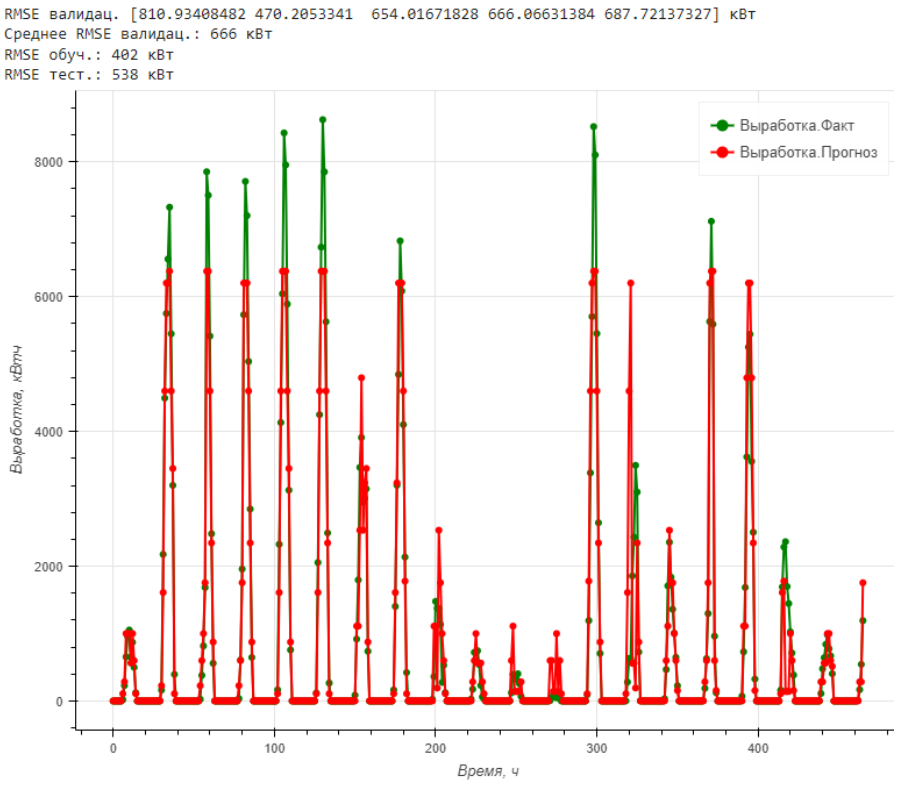

Рисунок 3.22 – Результат дерева решений с глубиной 5

Результат для дерева решений глубиной 4 показан на рисунке 3.23. Видно, что ошибка RMSE близка к той, которая получена в процессе кросс-валидации.

Применение случайного леса (результат на рисунке 3.24).

```
model = RandomForestRegressor()
model = create_evaluate_model(
    model, X_trainval, y_trainval)

visualize(model, X_trainval, y_trainval)
```

```
RMSE валидац. [1068.49029997  563.95430862  663.46183032  609.13256833  589.45477487] кВт
Среднее RMSE валидац.: 723 кВт
RMSE обуч.: 509 кВт
RMSE тест.: 537 кВт
```

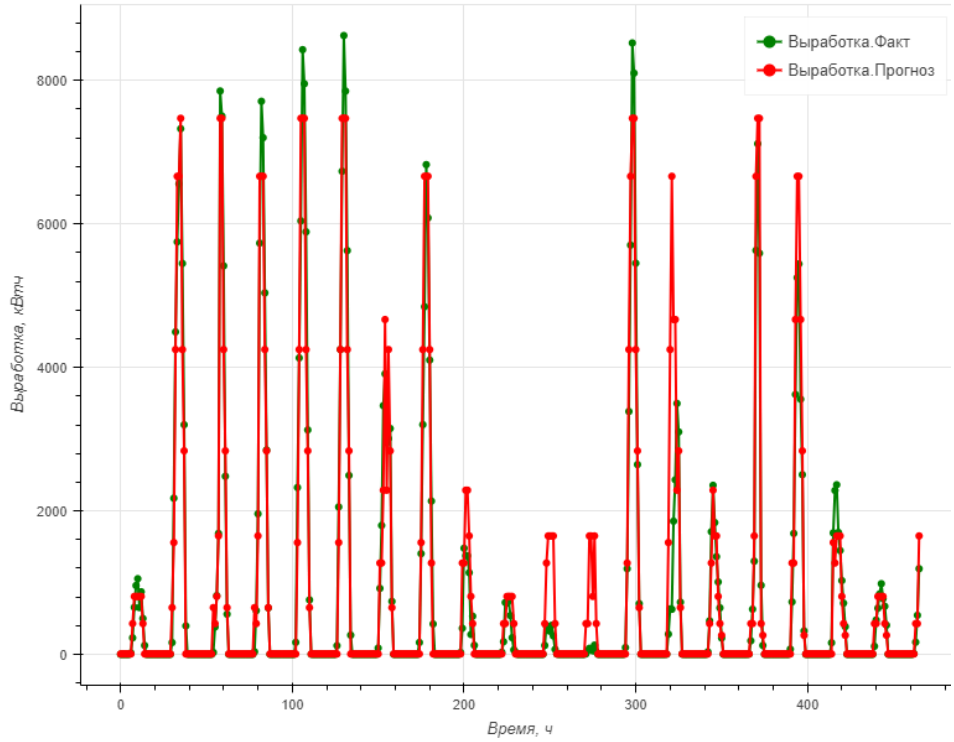

Рисунок 3.23 – Результат дерева решений на тестовой выборке

```
RMSE валидац. [694.35723474 329.57811434 503.73338572 471.47022128 556.29412304] кВт
Среднее RMSE валидац.: 524 кВт
RMSE обуч.: 180 кВт
RMSE тест.: 385 кВт
```

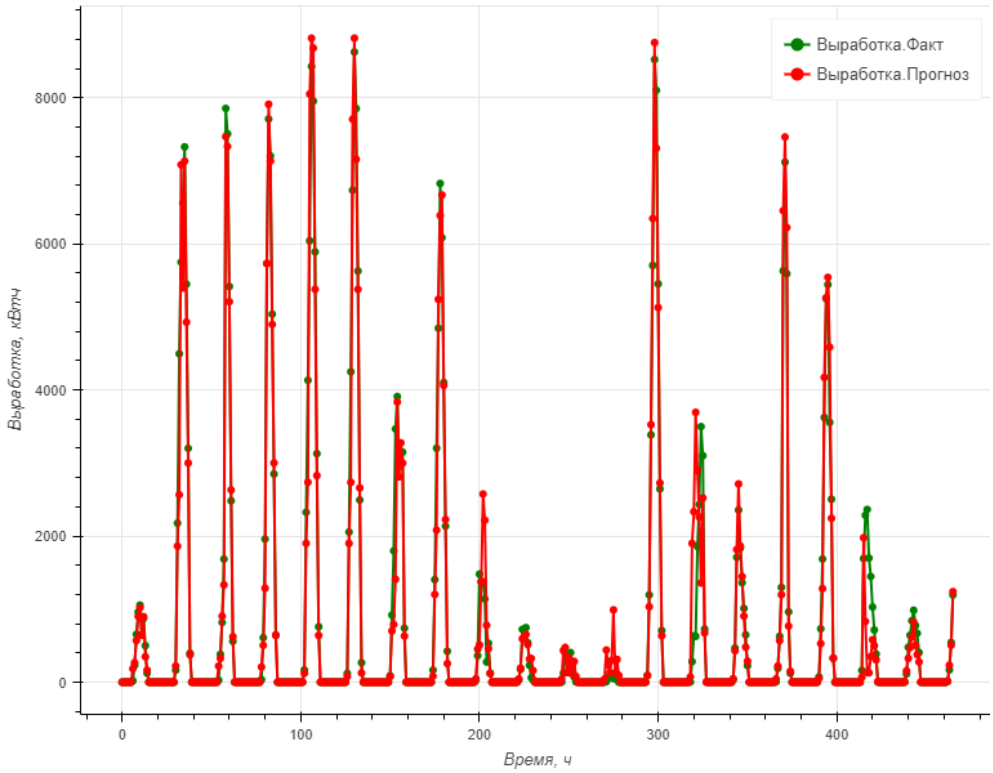

Рисунок 3.24– Результат случайного леса

Основные параметры, которые можно настраивать в Random Forest – число деревьев и максимальная глубина дерева:

model = RandomForestRegressor(n_estimators = 10, max_depth = 3)

В таблице 3.1 приведены полученные показатели точности различных моделей на тестовой выборке.

На рисунке 3.25 показана зависимость показателей точности дерева решений от максимальной глубины.

Результаты наилучшей модели (случайный лес с 10 деревьями максимальной глубины 5) показаны на рисунках 3.26–3.30.

Таблица 3.1 – Результаты моделей машинного обучения на тестовой выборке данных

Модель	Гиперпараметры	RMSE на обучающей выборке, кВт	RMSE на тестовой выборке, кВт
Линейная регрессия	-	1119	1684
Дерево решений	Глубина = 3	675	670
Дерево решений	Глубина = 4	509	537
Дерево решений	Глубина = 5	402	538
Дерево решений	Глубина = 6	304	496
Дерево решений	Глубина = 7	239	482
Дерево решений	Глубина = 8	169	478
Дерево решений	Глубина = 9	111	481
Дерево решений	Глубина = 10	77	507
Случайный лес	Глубина = 3 Кол-во деревьев = 10	548	542
Случайный лес	Глубина = 5 Кол-во деревьев = 10	332	398
Случайный лес	Глубина = 3 Кол-во деревьев = 20	537	509
Случайный лес	Глубина = 5 Кол-во деревьев = 20	330	409
Случайный лес	Глубина = 7 Кол-во деревьев = 10	229	394
Случайный лес	Глубина = 7 Кол-во деревьев = 20	193	393

Продолжение таблицы 3.1

Модель	Гиперпараметры	RMSE на обучающей выборке, кВт	RMSE на тестовой выборке, кВт
Случайный лес	Глубина = 9 Кол-во деревьев = 10	190	386
Случайный лес	Глубина = 9 Кол-во деревьев = 20	165	377
Случайный лес	Глубина = 9 Кол-во деревьев = 50	169	379
Случайный лес	Глубина = 12 Кол-во деревьев = 20	155	390
Случайный лес	Глубина = 12 Кол-во деревьев = 5	187	449
Случайный лес	Глубина = 20 Кол-во деревьев = 10	179	433
Случайный лес	Глубина = 2 Кол-во деревьев = 100	775	674

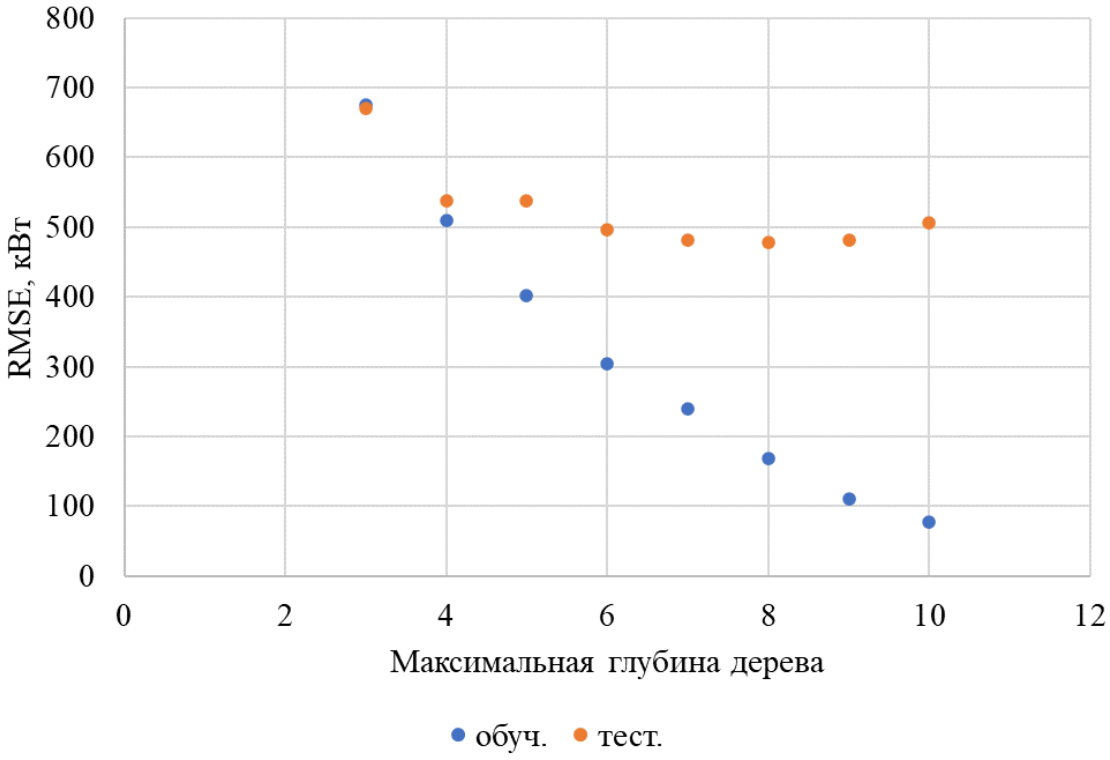

Рисунок 3.25– Зависимость ошибки дерева решений от его глубины

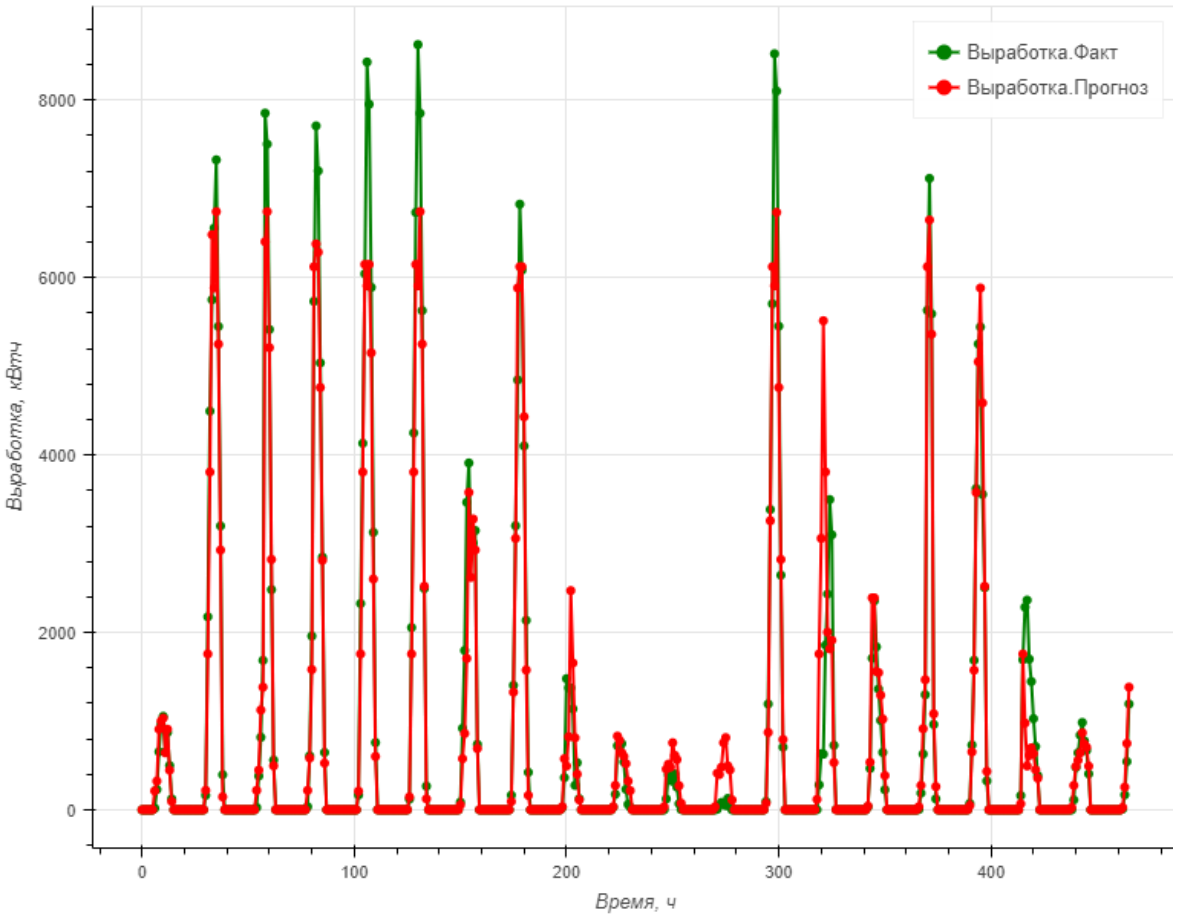

Рисунок 3.26– Результаты случайного леса на тестовой выборке

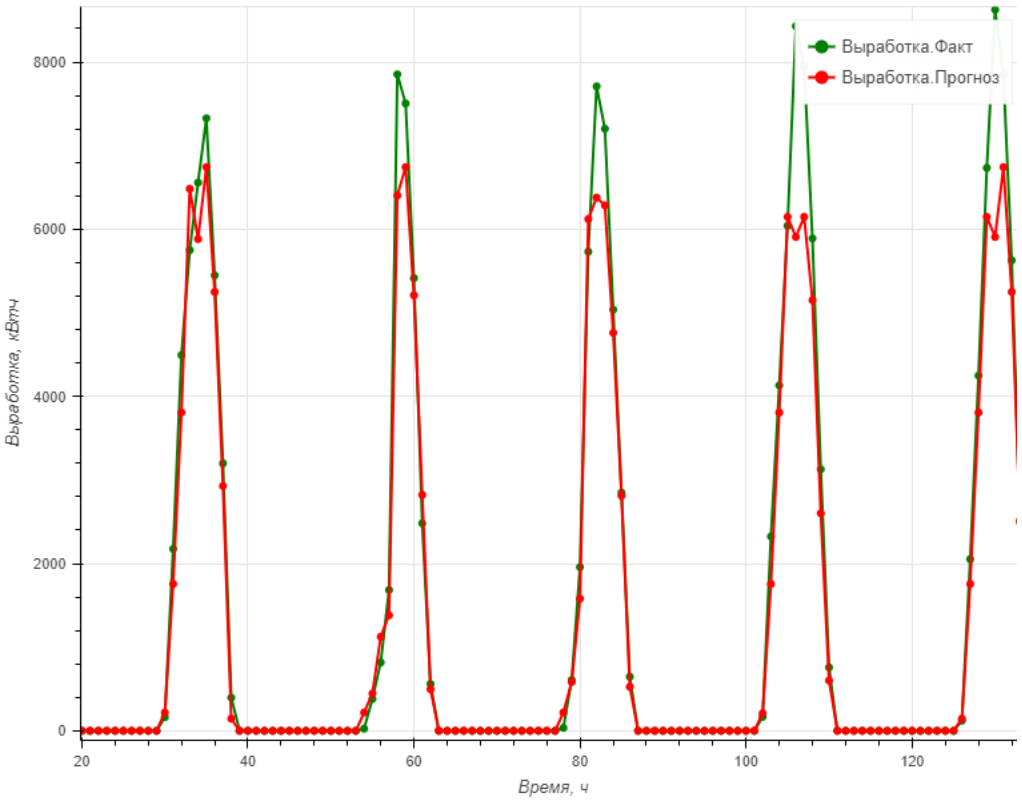

Рисунок 3.27 – Результаты случайного леса на тестовой выборке

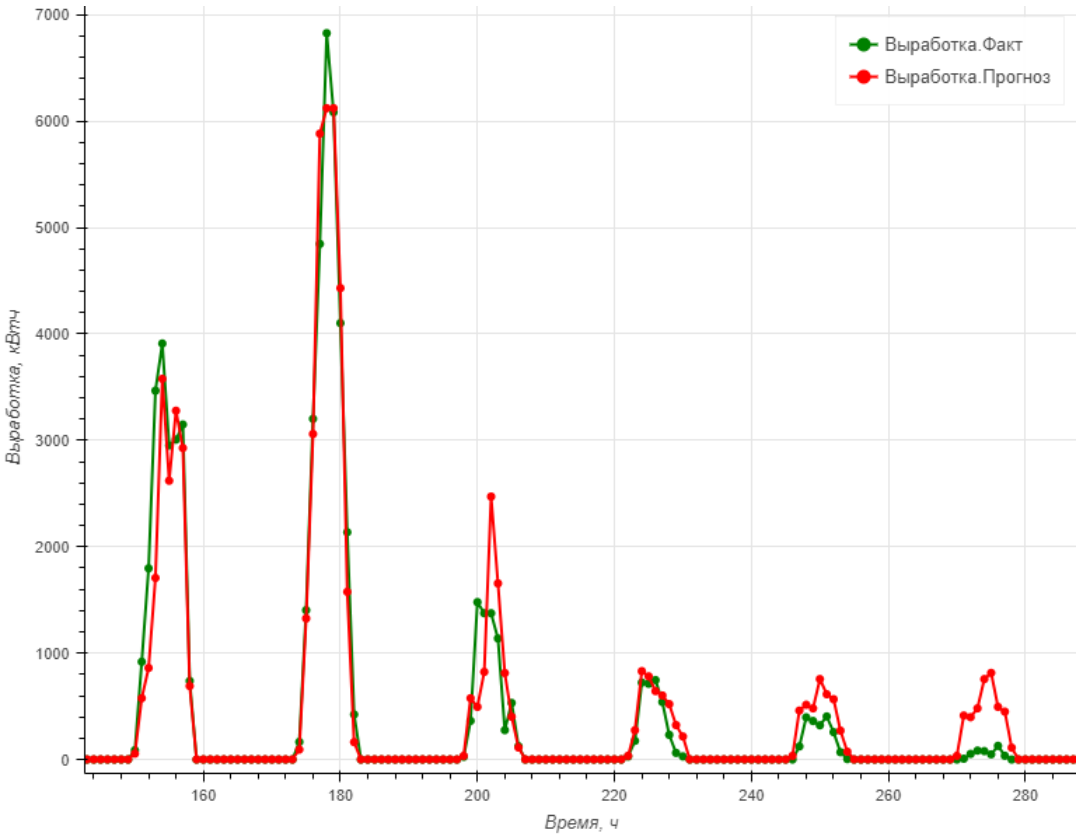

Рисунок 3.28 – Результаты случайного леса на тестовой выборке

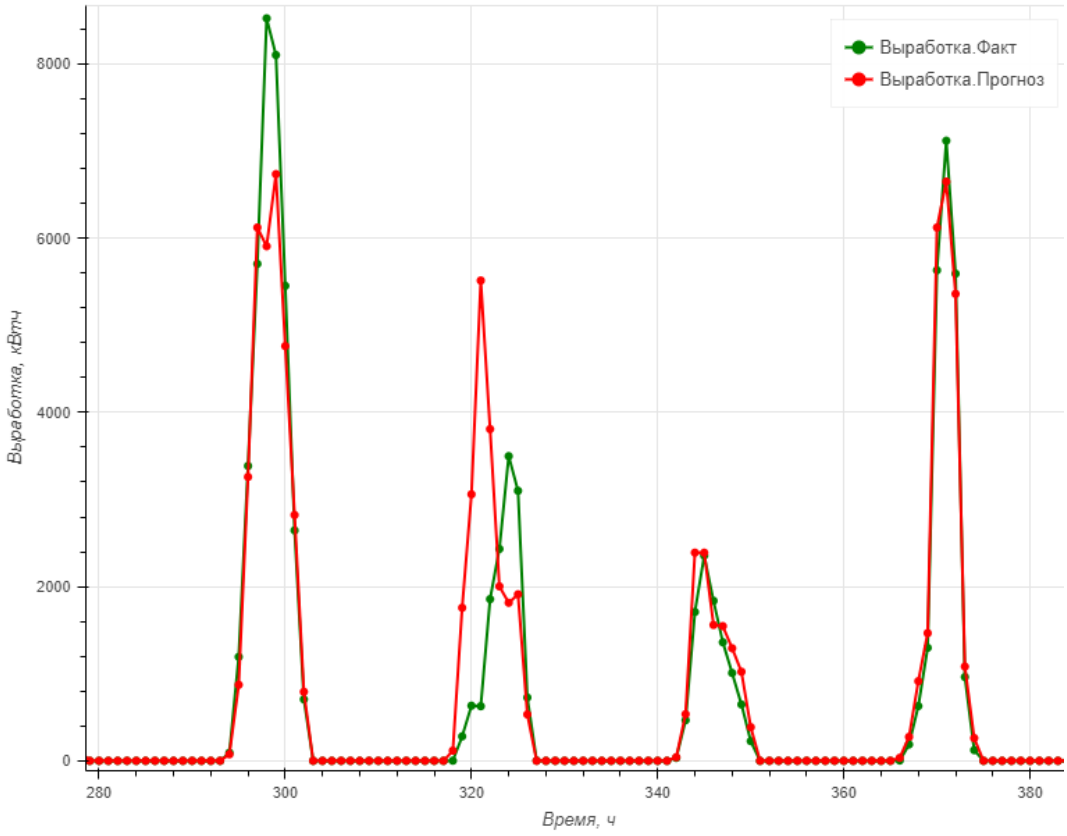

Рисунок 3.29 – Результаты случайного леса на тестовой выборке

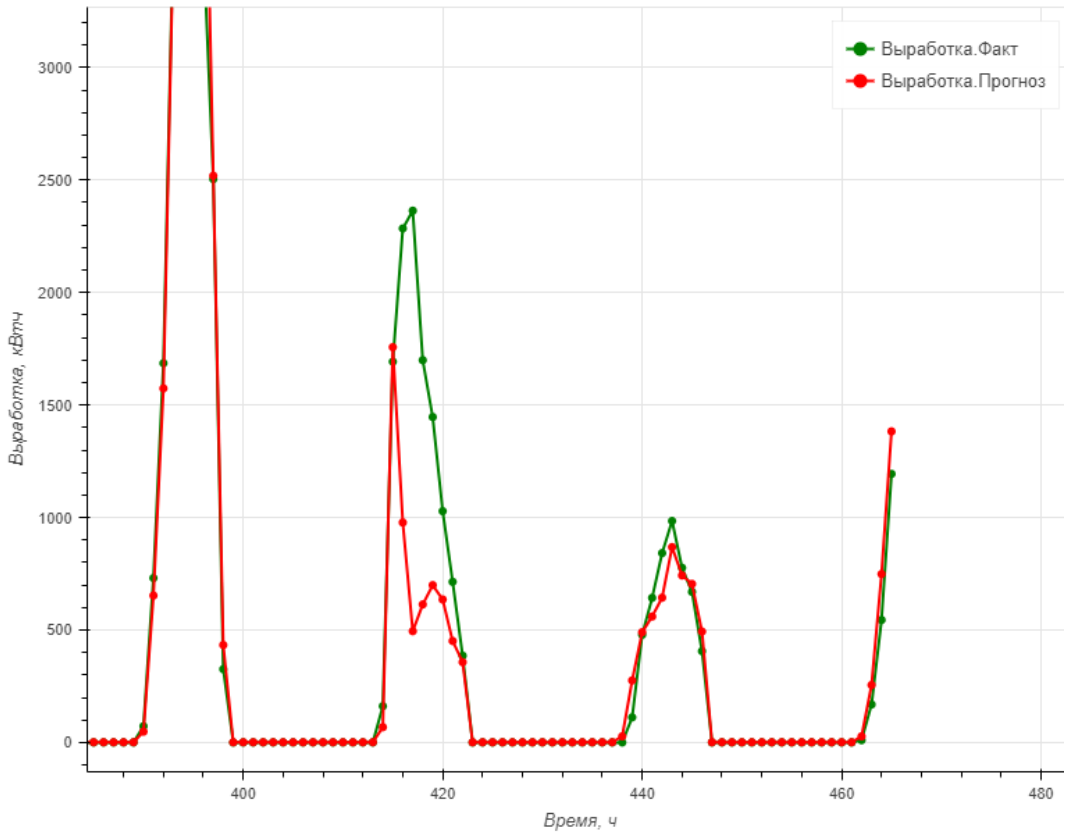

Рисунок 3.30 – Результаты случайного леса на тестовой выборке

3.5 Заключение по главе 3

1. Созданы модели на основе машинного обучения для прогнозирования генерации солнечных электрических станций. Проведено сравнение линейной регрессии, регрессионного дерева решений, случайного леса и градиентного бустинга.

2. Показана зависимость точности и устойчивости к переобучению моделей от гиперпараметров, в первую очередь – максимальной допустимой глубины дерева.

3. Наилучший результат получен с помощью случайного леса, среднеквадратическая ошибка на тестовой выборке составила 0,4 МВтч, что составляет лишь 4 % от установленной мощность рассматриваемой электрической станции.

ЗАКЛЮЧЕНИЕ

Основными результатами выпускной квалификационной работы магистра являются:

1. Разработаны модели на основе машинного обучения для прогнозирования генерации солнечных электрических станций.
2. Разработанные модели были успешно протестированы, показав наилучший результат с использованием случайного леса. Среднеквадратическая ошибка на тестовой выборке составила 0,4 МВтч, что составляет лишь 4 % от установленной мощность рассматриваемой электрической станции.

СПИСОК ЛИТЕРАТУРЫ

1. Григораш О.В., Евтушенко И.В., Попучиева М.А. Классификация и основные способы построения солнечных электростанций // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2016. № 124. С. 1453-1466
2. Основы возобновляемой энергетики. Компания «Ваш Солнечный Дом», 2016. [Электронный ресурс] – Режим доступа: http://www.solarhome.ru/ru/basics/pv/
3. Кейкиманова М.Т., Айтжанов Ш.Н. Фотоэлектрические свойства p- n перехода // Механика и технологии. 2015. № 2 (48). С. 21-30
4. Бессель В.В., Кучеров В.Г., Мингалеева Р.Д. Изучение солнечных фотоэлектрических элементов: Учебно-методическое пособие. – М.: Издательский центр РГУ нефти и газа (НИУ) имени И.М. Губкина, 2016. – 90 с
5. Федотикова М.В. Теоретические и прикладные аспекты повышения выработки энергии фотоэлектрическими преобразователями модификацией их поверхности нанокластерами серебра : диссертация канд. техн. наук, 2023, 158с
6. Research cell record efficiency chart. The National Center for Photovoltaics (NCPV) at the National Renewable Energy Laboratory (NREL), 2015. [Электронный ресурс] – Режим доступа: http://www.nrel.gov/ncpv/
7. Обоснование развития электроэнергетических систем: Методология, модели, методы, их использование / Н. И. Воропай, С. В. Подковальников, В. В. Труфанов [и др.] ; Отв. ред. Н. И. Воропай. – Новосибирск: Наука, 2015. – 448 с.
8. Городов Р.В, Губин В.Е., Матвеев А.С. Нетрадиционные и возобновляемые источники энергии: Учебное пособие. – Томск: Изд-во Томского политех. ун-та, 2009. – 294 с

9. Ali Sayigh. Comprehensive Renewable Energy. Volume One. Photovoltaic Solar Energy // Elsevier Ltd – 2012. – P. 746. [Электронный ресурс] – Режим доступа: http://www.sciencedirect.com.focus.lib.kth.se/
10. Michael Child, Dmitrii Bogdanov, Christian Breyer. The role of storage technologies for the transition to a 100% renewable energy system in Europe // Energy Procedia. 2018. Vol.155. P.44-60, https://doi.org/10.1016/j.egypro.2018.11.067.
11. Gopi, A., Sudhakar, K., Keng, N., Krishnan, A., & Priya, S. Performance Modeling of the Weather Impact on a Utility-Scale PV Power Plant in a Tropical Region // International Journal of Photoenergy. 2021. https://doi.org/10.1155/2021/5551014.
12. Atyia, T., Qasim, M. Evaluating the Impact of Weather Conditions on the Effectiveness and Performance of PV Solar Systems and Inverters // NTU Journal of Renewable Energy. 2023. https://doi.org/10.56286/ntujre.v5i1.551.
13. Voskanyan, A., Borden, D., Yakimovich, B., & Barmani, A. (2019). Studies of the Operation of a Solar Power Plant // Applied Solar Energy. 2019. Vol.55. P.235-241. https://doi.org/10.3103/S0003701X19040091.
14. Simeng Cheng, Yimiao Xie, Zhan Shu, Yulin Wang, Zhou Ning and Yue Yang. Effect of Different Solar Irradiance Parameters on Reliability Evaluation of the Grid // IOP Conference Series: Earth and Environmental Science. 2019. doi:10.1088/1755-1315/300/4/042115.
15. D. I. Jurj, D. D. Micu, A. Muresan. Overview of Electrical Energy Forecasting Methods and Models in Renewable Energy // 2018 International Conference and Exposition on Electrical And Power Engineering (EPE). 2018. pp. 0087-0090, doi: 10.1109/ICEPE.2018.8559807.
16. Kwok W. Cheung/ New forecasting and modeling tools for operating power systems with variable energy resources // 2014 IEEE Innovative Smart Grid Technologies - Asia (ISGT ASIA). 2014. pp.186-191.

17. Michael Negnevitsky, Paras Mandal, Anurag K. Srivastava. An Overview of Forecasting Problems and Techniques in Power Systems // 2009 IEEE Power & Energy Society General Meeting. 2009. pp.1-4.
18. G. Aburiyana, M.E.El-Hawary. An overview of forecasting techniques for load, wind and solar powers // 2017 IEEE Electrical Power and Energy Conference (EPEC), Saskatoon, SK. 2017. pp. 1-7, doi: 10.1109/EPEC.2017.8286192.
19. Kuzlu M., Cali U., Sharma V., Güler O. Gaining insight into solar photovoltaic power generation forecasting utilizing explainable artificial intelligence tools // IEEE Access. 2020. Vol. 8. P. 187814–187823.
20. H. Liu, F. Huang, H. Li, W. Liu and T. Wang, "A big data framework for electric power data quality assessment," 2017 14th Web Information Systems and Applications Conference (WISA), Liuzhou, China, 2017, pp. 289-292, doi: 10.1109/WISA.2017.29.
21. Improving of the generation accuracy forecasting of photovoltaic plants based on k-Means and k-Nearest neighbors algorithms / P. V. Matrenin, A. I. Khalyasmaa, V. V. Gamaley, S. A. Eroshenko [et al.] // ENERGETIKA. Proceedings of CIS higher education institutions and power engineering associations. 2023. Vol. 66(4). P. 305–321
22. T. Zufferey, A. Lepouze, G. Hug. Inadequacy of standard algorithms and metrics for short-term load forecasts in low-voltage grids // 2019 IEEE Milan PowerTech, Milan, Italy. 2019. pp. 1-6, doi: 10.1109/PTC.2019.8810430
23. Jehad Ali, Rehanullah Khan, Nasir Ahmad, Imran Maqsood. Random Forests and Decision Trees // International Journal of Computer Science Issues (IJCSI). 2012. 9.
24. Biau, Gérard. Analysis of a Random Forests Model // Journal of Machine Learning Research. 2010. 13.
25. Delgado, M., Sirsat, M., Cernadas, E., Alawadi, S., Barro, S., Febrero-Bande, M. An extensive experimental survey of regression methods. Neural networks

: the official journal of the International Neural Network Society. 2019. Vol.111, P.11-34. https://doi.org/10.1016/j.neunet.2018.12.010.

26. Gambella, C., Ghaddar, B., Naoum-Sawaya, J. Optimization Models for Machine Learning: A Survey. Eur. J. Oper. Res., 2019. Vol.290. P.807-828. https://doi.org/10.1016/J.EJOR.2020.08.045.
27. H. Liu, F. Huang, H. Li, W. Liu and T. Wang. A big data framework for electric power data quality assessment. 2017 14th Web Information Systems and Applications Conference (WISA), Liuzhou, China, 2017, pp. 289-292, doi: 10.1109/WISA.2017.29.
28. Machine learning methods for solar radiation forecasting: A review / C. Voyant, G. Notton, S. Kalogirou, M.-L. Nivet [et al.] // Renewable Energy. 2017. Vol.105. P.569-582.
29. A. Bramm, S. Eroshenko. Optimal Reconfiguration of Distribution Network with Solar Power Plants, 2021 Ural-Siberian Smart Energy Conference (USSEC), 2021, pp. 1-5, doi: 10.1109/USSEC53120.2021.9655718.