

Влияние настройки гиперпараметров на производительность солнечных фотоэлектрических систем с использованием машинного обучения для Москвы, Россия

Кингсли Околи¹, Она Уче¹, Хааструп Адебайо Ибукун², Виктор Ачиргбенда³, Инносент Джозеф¹

¹ Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина)

² Федеральный технологический университет Акуре, Нигерия

³ Университет сельского хозяйства Макурди, Нигерия

kokoli@stud.etu.ru, onah9682@gmail.com, haastrupbay012@gmail.com, achigbenda.victor@uam.edu.ng, idzhofez@stud.etu.ru

Аннотация. Использование солнечных фотоэлектрических (PV) систем стало жизнеспособным вариантом для обеспечения растущих мировых потребностей в энергии при одновременном снижении выбросов парниковых газов. Эффективность и рентабельность системы должны быть спроектированы и оптимизированы, что требует точного моделирования производительности солнечной фотоэлектрической системы. Чтобы выбрать лучшие гиперпараметры для конкретной ситуации, важно иметь более глубокое представление о том, как настройка гиперпараметров влияет на производительность модели солнечной фотоэлектрической системы. В этой статье исследуется влияние настройки гиперпараметров на эффективность моделирования солнечных фотоэлектрических систем для Москвы, с использованием подходов машинного обучения. С целью прогнозирования производительности солнечной фотоэлектрической системы анализируются данные и различные модели машинного обучения. Производительность этих моделей оценивается с использованием различных критериев при сопоставлении результатов во время гипернастройки. Согласно исследованию, точность и надежность моделей производительности солнечных фотоэлектрических систем можно значительно повысить за счет настройки гиперпараметров. Мы обрисовываем последствия и предполагаемые последствия нашего исследования для проектирования и оптимизации солнечных фотоэлектрических систем в Москве и вносим предложения для дальнейшего изучения. Результаты этого исследования помогут России и другим регионам разработать более эффективные и устойчивые солнечные фотоэлектрические системы.

Ключевые слова: машинное обучение, Solar Pv, гиперпараметры, оптимизация, модель

I. ВВЕДЕНИЕ

Необходимость удовлетворения растущих потребностей в энергии в результате роста населения мира, а также сдерживания воздействия изменения климата на окружающую среду, вызванного использованием ископаемого топлива, вызвала необходимость изучения возобновляемых источников энергии. Ископаемые виды топлива имеют серьезные последствия для глобального климата, а также для

здоровья человека. По данным Международного агентства по возобновляемым источникам энергии (IRENA), до 30% мирового потребления электроэнергии обеспечивается за счет возобновляемых источников энергии, таких как солнечная, ветровая и гидроэнергетика, но солнечная энергия обеспечивает лишь около 1% мирового производства электроэнергии [1, 2]. Тем не менее, солнечная энергия сегодня заслуживает серьезного внимания, поскольку ожидается, что к середине века она будет играть гораздо более важную роль в глобальной энергетической системе, а устранение ряда важных препятствий в течение следующих нескольких десятилетий значительно повысит вероятность того, что солнечная энергия станет лучшим вариантом. Важность солнечной энергии, в конечном счете, проистекает из серьезной долгосрочной угрозы, которую представляет глобальное изменение климата [3]. Выбросы углекислого газа (CO₂) при сжигании ископаемого топлива составляют наибольшую долю парниковых газов, вызывающих изменение климата [4]. При прочих равных условиях, если к середине века солнечная энергия станет альтернативным источником электроэнергии, солнечная промышленность и ее цепочка поставок должны резко возрасти. По прогнозам Международного энергетического агентства (МЭА), к 2050 году производство солнечной энергии увеличится в 50 раз по сравнению с 2013 годом [5]. Основным препятствием для крупномасштабного использования солнечной генерации во многих регионах являются перебои в подаче электроэнергии. Современные электроэнергетические системы должны почти мгновенно приводить выработку в соответствие со спросом. Хотя колебания спроса также плохо предсказуемы, добавление небольшого количества солнечной энергии не создает заметных проблем. Однако в энергосистеме, которая сильно зависит от солнечной энергии, прерывистость солнечного ресурса сделает чистую нагрузку более изменчивой и, следовательно, менее предсказуемой.

Для максимизации эффективности систем ВИЭ и систем ВИЭ на основе солнечной энергии важно учитывать различные многоуровневые факторы, такие, как единовременный баланс между генерирующими и потребляющими/аккумулирующими мощностями,

работой контроллера, местными климатическими условиями, региональные электросетевые условия и условия эксплуатации в дополнение к таким глобальным характеристикам, как учетная ставка ЦБ и тарифы производителей солнечных панелей. Имитационное программное обеспечение позволяет решить большинство вопросов с моделированием условий работы потенциальных систем ВИЭ [6]. Машинное обучение использовалось для определения потенциала возобновляемых источников энергии [7], и поэтому в этой работе будет дополнительно обсуждаться, как гиперпараметры влияют на производительность солнечных фотоэлектрических систем при настройке для оптимизации.

II. ЛИТЕРАТУРНЫЙ ОБЗОР

В этой работе мы исследуем влияние настройки гиперпараметров на производительность солнечных фотоэлектрических (PV) систем с использованием машинного обучения (ML). Солнечные фотоэлектрические панели стали важнейшим источником возобновляемой энергии, но такие факторы, как климат, затенение и технологии, могут влиять на эффективность и действенность этих систем. Использование машинного обучения было предложено в качестве мощного подхода к оптимизации производительности фотоэлектрических систем. Балансировка и интеграция прерывистых возобновляемых источников энергии, таких как фотоэлектрические солнечные батареи, в обычную сеть – сложная задача, требующая специальных навыков и опыта. Непредсказуемость производства солнечной энергии из-за изменений погодных условий может привести к внезапным колебаниям в поставках, что создает проблемы для специалистов оператора системы передачи и распределения (TDSO) в поддержании стабильности сети и предотвращении отключений. В результате растет потребность в передовых технологиях и инновационных решениях для управления растущим проникновением солнечной энергии в сеть и обеспечения надежного и устойчивого электроснабжения конечных пользователей. Поэтому точные методы прогнозирования приобретают все большее значение. Журнал возобновляемых и устойчивых источников энергии [8] отмечает, что некоторые исследования по солнечному прогнозированию страдают от определенных ограничений, которые снижают их общую ценность, таких как потенциальное повторение, ограниченная воспроизводимость и проблемы сопоставимости. Несмотря на множество подходов, доступных для различных типов задач прогнозирования солнечной активности, по-прежнему сложно определить наиболее подходящий метод для данного практического применения. В результате существует потребность в усовершенствованных методах исследований, которые могут обеспечить надежность, воспроизводимость и сопоставимость исследований по прогнозированию солнечной активности, что позволит более эффективно выбирать подходящие методы для практического применения. [9] В их исследовании был изучен значительный объем исследований по прогнозированию для анализа влияния горизонта прогноза и шести основных классов моделей на точность прогноза. Для сравнения точности различных моделей прогнозирования в исследовании использовались две

основные метрики, а именно средненормализованная средняя ошибка смещения и среднеквадратическая ошибка. Проведя всесторонний анализ большой выборки прогностических исследований, авторы стремились пролить свет на эффективность различных моделей на разных горизонтах прогнозирования и определить наиболее эффективные подходы к моделированию для точного прогнозирования солнечной активности.

Авторы статьи «Проверка детерминированных солнечных прогнозов» [10] отмечают, что предсказуемость солнечного излучения может значительно различаться, что затрудняет сравнение методов прогнозирования, которые были проверены для разных мест и временных масштабов, даже когда используются нормализованные формы метрик общих ошибок. В результате крайне важно учитывать уникальные характеристики каждого местоположения и задачи прогнозирования при оценке точности методов прогнозирования солнечной радиации. Авторы подчеркивают необходимость более тонкого и контекстно-зависимого подхода к оценке солнечного прогнозирования, чтобы учесть влияние различных уровней предсказуемости на производительность различных моделей [10]. В то время как машинное обучение было популярным инструментом для моделирования фотоэлектрических установок, статистические методы преобразования солнечной радиации в мощность также оказались полезными для практических приложений, таких как планирование выходной мощности фотоэлектрических установок для рынка на сутки вперед. Эти методы используют статистические модели для преобразования измерений энергетической освещенности в прогнозы выходной мощности, тем самым облегчая интеграцию солнечной энергии в энергосистему. Хотя подходы машинного обучения продемонстрировали высокую точность прогнозирования солнечной радиации и выходной мощности, статистические методы могут быть более простой и эффективной альтернативой в определенных сценариях. Таким образом, как машинное обучение, так и статистические методы обладают уникальными преимуществами и могут дополнять друг друга при оптимизации производительности фотоэлектрических установок для различных практических приложений [11]. Путем явного моделирования неопределенности в значениях параметров и включения вероятностных ограничений и целей в модели планирования стохастическое программирование/оптимизация могут генерировать все более надежные решения планирования, учитывающие потенциальные отклонения от ожидаемых значений. Недавние исследования показали, что подходы стохастической оптимизации могут значительно улучшить работу энергетических систем в условиях неопределенности, особенно в контексте возобновляемых источников энергии. Таким образом, стохастическое программирование/оптимизация стало важным инструментом для операторов энергосистем, стремящихся оптимизировать использование возобновляемых источников энергии и повысить надежность и устойчивость энергетических систем [12].

Проблемы прогнозирования последовательности относятся к задаче использования информации о прошлой последовательности для прогнозирования последующего значения или значений в

последовательности. Эти типы проблем можно найти в различных приложениях, включая прогнозирование погоды, прогнозирование цен на акции и обработку естественного языка. Цель прогнозирования последовательности состоит в том, чтобы точно зафиксировать временные зависимости в данных и сделать надежные прогнозы будущих значений. Для решения этой проблемы были разработаны и применены к задачам прогнозирования последовательности различные алгоритмы машинного обучения, такие как рекуррентные нейронные сети (RNN) и сети с долговременной кратковременной памятью (LSTM) с многообещающими результатами. Эти алгоритмы способны моделировать сложные временные зависимости и могут быть обучены с использованием больших объемов исторических данных для повышения точности прогнозов [13]. Методы автоматической настройки гиперпараметров необходимы для эффективной оптимизации моделей машинного обучения. В этих методах используются алгоритмы для систематического и эффективного поиска наилучших значений гиперпараметров. Одним из таких методов является байесовская оптимизация, которая моделирует взаимосвязь между гиперпараметрами и производительностью модели с использованием вероятностной модели и последовательно выбирает следующий набор гиперпараметров для оценки на основе неопределенности модели. Этот подход позволяет эффективно исследовать пространство гиперпараметров и дает оптимальные гиперпараметры с меньшим количеством оценок, чем поиск по сетке. Кроме того, байесовскую оптимизацию можно легко расширить для обработки нескольких целей и ограничений. Байесовскую оптимизацию можно использовать как более эффективную альтернативу поиску по сетке при работе с большим пространством поиска или большим количеством гиперпараметров. Однако, не всегда возможно использовать только байесовскую оптимизацию, так как это может быть дорогостоящим в вычислительном отношении для очень больших пространств поиска. В таких случаях можно использовать комбинацию обоих методов, где байесовская оптимизация используется для сужения пространства поиска, а затем поиск по сетке используется для нахождения оптимальных гиперпараметров в этом пространстве [14].

Настройка гиперпараметров имеет решающее значение для достижения наилучшей производительности моделей ML для прогнозирования производительности PV. Без надлежащей настройки гиперпараметров модель может переоценивать или недооценивать данные, что приводит к снижению производительности при работе с новыми данными. Настраивая гиперпараметры, мы можем найти наилучший набор значений, которые помогут модели хорошо обобщать новые данные. Исследования показали, что правильная настройка гиперпараметров может повысить точность и надежность моделей ML, делая их более эффективными для задач прогнозирования производительности PV.

III. МЕТОДОЛОГИЯ

Данные, используемые для обучения моделей, были собраны из Kaggle и предварительно обработаны с помощью Python. Различные собранные параметры

включают в себя; температуру воздуха, скорость ветра, давление, влажность, направление ветра и солнечную радиацию. Для настройки параметров Solar PV Performance использовались две платформы машинного обучения, т. е. комбинация Xgboost и Optuna. В машинном обучении проблема выбора набора идеальных гиперпараметров для алгоритма обучения известна как оптимизация или настройка гиперпараметров. Гиперпараметр – это параметр, значение которого влияет на то, как происходит обучение. Напротив, значения других параметров (обычно веса узлов) изучаются. Чтобы обобщить различные шаблоны данных, одной и той же модели машинного обучения могут потребоваться различные ограничения веса или скорости обучения. Гиперпараметры – это переменные, которые необходимо настроить, чтобы модель работала лучше всего при решении задачи машинного обучения. Оптимальная модель создается путем оптимизации гиперпараметров, которая идентифицирует набор гиперпараметров, который минимизирует заданную функцию потерь для заданных независимых данных. При задании кортежа гиперпараметров целевая функция возвращает соответствующую потерю.

Optuna, платформа оптимизации гиперпараметров для платформ машинного обучения и решателей оптимизации черного ящика, использует целевую функцию, которая дает числовое значение для оценки производительности гиперпараметров и определения того, где брать образцы в будущих испытаниях. Выбирая точки ближе к более ранним успешным результатам, Optuna использует древовидную структуру Parzen Estimator (TPE) для более эффективного поиска, чем случайный поиск. Каждое испытание выбирается после рассмотрения всех предыдущих испытаний и использования пробоотборника TPE для обоснованных предположений о том, где могут быть расположены наилучшие значения гиперпараметров. На рис. 1 показана структура настройки гиперпараметров Optuna.

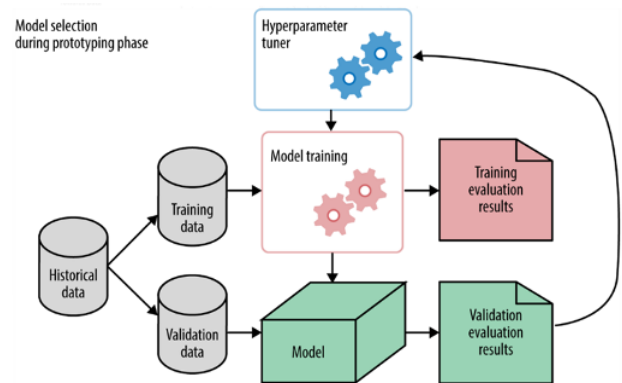


Рис. 1. Настройка гиперпараметров с помощью Optuna

Одним из самых быстрых и эффективных доступных алгоритмов машинного обучения является XGBoost, который стал возможен благодаря ряду методов, включая распараллеливание, оптимизацию кеша, вычисления и регуляризацию нехватки памяти, автоматическую обрезку и обработку пропущенных значений. С помощью выпуклой функции потерь (основанной на разнице между прогнозируемыми и целевыми выходными данными) и штрафного коэффициента сложности модели, XGBoost минимизирует регуляризованную (L1 и L2) целевую

функцию (другими словами, функции дерева регрессии). Итеративно процесс обучения добавляет новые деревья, которые прогнозируют остатки или ошибки более ранних деревьев, которые затем объединяются с более ранними деревьями для получения окончательного прогноза.

Комбинируя выпуклую функцию потерь (основанную на разнице между прогнозируемыми и целевыми выходными данными) и штрафной коэффициент сложности модели, XGBoost минимизирует регуляризованную (L1 и L2) целевую функцию. Итеративно процесс обучения добавляет новые деревья, которые прогнозируют остатки или ошибки более ранних деревьев, которые затем объединяются с более ранними деревьями для получения конечного прогноза.

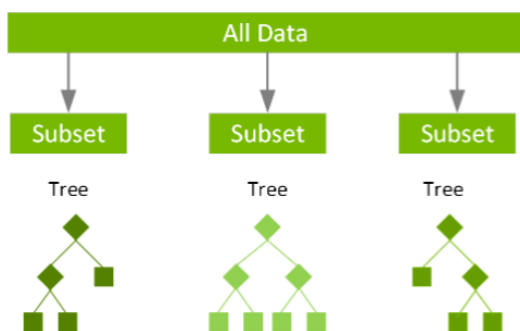


Рис. 2. Архитектура XGboost

Использовались данные из московской статистики PV Solar, а среди факторов, влияющих на излучение, были температура, влажность, давление и скорость ветра. Эти параметры можно использовать для прогнозирования идеального излучения посредством настройки гиперпараметров.

С помощью этой методологии настройки параметров Xgboost использует регрессию среднеквадратичной ошибки (RMSE) для оценки качества прогнозирования излучения PV. Чтобы вычислить RMSE, была рассчитана невязка (разница между предсказанием и истиной) для каждой точки данных, вычислена норма невязки для каждой точки данных, вычислено среднее значение невязок и взят квадратный корень из этого среднего значения. RMSE обычно используется в приложениях для обучения с учителем, поскольку RMSE использует и нуждается в истинных измерениях в каждой прогнозируемой точке данных.

Среднеквадратическая ошибка может быть выражена как в уравнении (1);

$$RMSE = \sqrt{\frac{\sum_{i=1}^N |y(i) - \hat{y}(i)|^2}{N}} \quad (1)$$

где: N – количество точек данных; $y(i)$ – i-е измерение; $\hat{y}(i)$ – соответствующее предсказание.

IV. РЕЗУЛЬТАТ И ОБСУЖДЕНИЕ

Optuna, являющаяся платформой параллельной оптимизации, использовала пробный объект, чтобы предложить лучшее значение гиперпараметра, которое будет использоваться при обучении данных PV (эти значения включают скорость обучения, максимальную глубину, регуляризацию L1 и L2). Поскольку целью этой

модели является точность необученных данных, испытание Optuna было выполнено для трех разных значений; 80, 100 и 120.

ТАБЛИЦА I. ПАРАМЕТРЫ, ПОЛУЧЕННЫЕ ПРИ НАСТРОЙКЕ

Серийный номер	No_of Trials	Скорость обучения	Максимальная глубина	L1 Регуляризация	L2 Регуляризация
1	120	0,13846347 425358296	8	0,138463474 25358296	0,00740369 8758835623
2	100	0,18003871 03747378	7	0,000542461 2273460876	7,47973131 7146002
3	80	0,14966140 63756982	8	0,045492353 08933801	0,06538092 173802672

а. Испытание Optuna 80: при значениях гиперпараметров оптимальная среднеквадратическая ошибка (RMSE), полученная на данных обучения PV, была уменьшена с 326,75082 до 40,81769, и это имело аналогичные характеристики на наборе проверочных данных, т. е. RMSE был оптимизирован с 325,12671 до 325,12671.

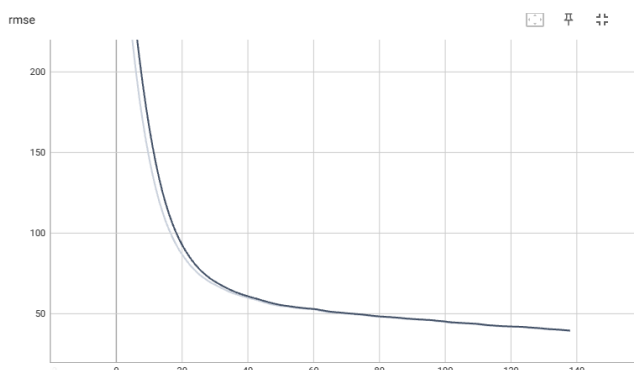


Рис. 3. Эффективность тренировок с Trial 80

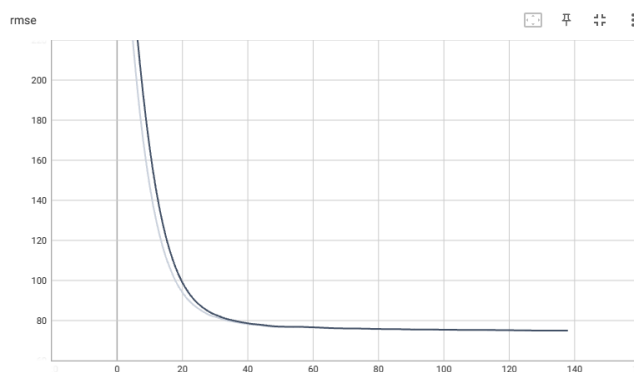


Рис. 4. Проверка производительности с Trial 80

После оценки ошибки между фактическими значениями и прогнозируемыми значениями была получена точность оценки 0,93 или 93 %.

б. Optuna Trial 100: при значениях гиперпараметров оптимальная среднеквадратическая ошибка (RMSE), полученная для данных обучения PV, была снижена с 304,32139 до 42,56662, и это имело сопоставимую производительность на наборе проверочных данных, т. е. RMSE был оптимизирован с 302,77990 до 76,08731.

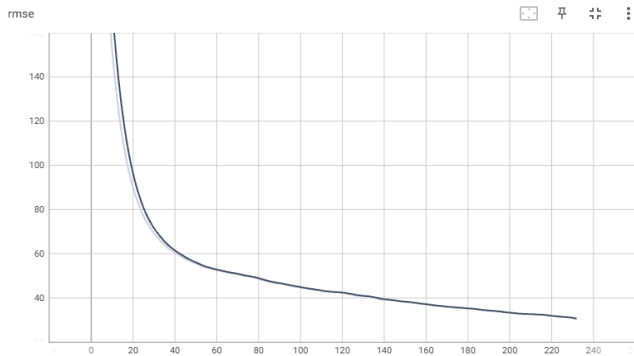


Рис. 5. Эффективность тренировок с Trial 100

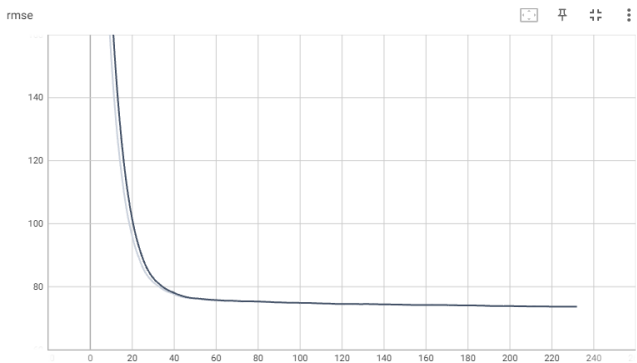


Рис. 6. Проверка производительности с Trial 100

После оценки ошибки между фактическими значениями и прогнозируемыми значениями была получена точность оценки 0,92 или 92 %.

Optuna Trial 120: Оптимальная среднеквадратическая ошибка (RMSE), полученная с использованием значений гиперпараметров в обучающих данных PV, была уменьшена с 332,54634 до 36,10975, и это имело сопоставимую производительность в наборе проверочных данных, т. е. RMSE был оптимизирован с 330,90083 до 75.02654.

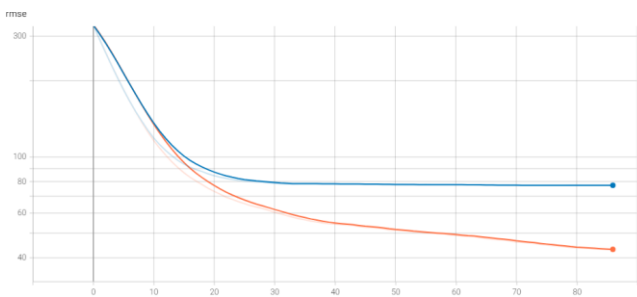


Рис. 7. Обучение и проверка производительности с помощью Trial 120

После оценки ошибки между фактическими значениями и прогнозируемыми значениями была получена точность оценки 0,93 или 93 %.

V. ЗАКЛЮЧЕНИЕ

Судя по результатам настройки, модель показала лучшие результаты при испытаниях 80 и 120 с точностью 93 % в каждом случае. Это означает, что солнечный P_v будет давать наилучшие показатели среднеквадратичной ошибки в Москве при настройке с конфигурациями 80 и 120 испытаний, полученных на

основе данных проверки. Со скоростью обучения 0,13846347425358296, максимальной глубиной 8, регуляризации L1 0,13846347425358296 и регуляризации L2 0,007403698758835623. Модель будет выполнена лучше всего при испытании 120. Также с уровнем обучения от 0,14966114066377, 206663777, 206666377, 206666377, 206666377, 2066166637, 268, 20666637, 268, 2666637, 3666637, 26, 26, 206666, 36, 20661, 26, 2066123. регуляризации 0,06538092173802672, модель будет работать лучше всего при испытании 80.

СПИСОК ЛИТЕРАТУРЫ

- [1] IRENA, Затраты на производство возобновляемой энергии в 2021 году, Международное агентство по возобновляемым источникам энергии, Абу-Даби. 2022. ISBN 978-92-9260-452-3.
- [2] НАСА. Прогноз мировых энергетических ресурсов (МОЩНОСТЬ): <https://power.larc.nasa.gov/> 2023 . По состоянию на 1 марта 2023 г.
- [3] СОЛАРГИС. Данные о погоде и программные услуги. <https://solargis.com/> 2023. По состоянию на 1 марта 2023 г.
- [4] Климатические данные онлайн от Национальной службы экологической информации Национального управления океанических и атмосферных исследований (NOAA), URL: <https://www.ncei.noaa.gov/cdo-web/>
- [5] NSRDB: Национальная база данных солнечной радиации Национальной лаборатории возобновляемых источников энергии (NREL), URL: <https://nsrdb.nrel.gov/>
- [6] Карлссон М., Палм Дж., Уиден Дж. «Методология междисциплинарных энергетических систем – подборка методов исследования, используемых в Программе энергетических систем». 2012. Арбетснотат, №. 45, версия 1.0.
- [7] К. Мадуабучи и др., «Оценка потенциала возобновляемых источников энергии с использованием алгоритмов машинного обучения прогнозирования климата и погоды» // Энергетика (Базель), том. 16, №4, с. 1603, февраль 2023 г., doi: 10.3390/en16041603.
- [8] Ян Д. Руководство по исследовательской практике прогнозирования солнечной активности: воспроизводимые, оперативные, вероятностные или физические, ансамблевые и профессиональные (ROPES) // Журнал возобновляемых и устойчивых источников энергии, 11(2), 022701. 2019 г. <https://doi.org/10.1063/1.5087462>
- [9] Блага Р., Сабадус А., Стефу Н., Дугир С., Паулеску М., Бадеску В. Современный взгляд на точность прогнозирования поступления солнечной энергии. Прогресс в области энергетики и науки о горении, 70, 119–144. 2019. <https://doi.org/10.1016/j.pecs.2018.10.003>.
- [10] Ян Д., Алессандрини С., Антонансас Дж., Антонансас-Торрес Ф., Бадеску В., Бейер Х.Г., Блага Р., Боланд Дж., Брайт Дж.М., Коимбра CFM, Дэвид М., Фримане В., Геймар К.А., Хонг Т., Кей М.Дж., Киллинджер С., Кляйсл Дж., Лорет П., Лоренц Э., Чжан Дж. Проверка детерминированных солнечных прогнозов. Солнечная энергия, 210, 20–37. 2020. <https://doi.org/10.1016/j.solener.2020.04.019>
- [11] Меллит А., Калогиру С.А. Методы искусственного интеллекта для фотоэлектрических приложений: обзор. Прогресс в области энергетики и науки о горении, 34 (5), 574–632. 2008. <https://doi.org/10.1016/j.pecs.2008.01.001>.
- [12] Ли З., Чжан Ю., Чжан Г. Двухэтапное стохастическое программирование для вторичного распределения рафинированной нефти с неопределенным спросом и ограниченными запасами. 119487–119500. IEEE. 2020 г. <https://doi.org/10.1109/ACCESS.2020.3004849>.
- [13] Лю Ю. Анализ и прогноз цены акций на основе алгоритма LSTM. // 2021 Международная конференция IEEE по компьютерным наукам, электронной информационной инженерии и технологиям интеллектуального управления (CEI), 76–79. 2021 г. <https://doi.org/10.1109/CEI52496.2021.9574519>.
- [14] Буктиф С., Физат А., Уни А., Серхани Г-жа А. Глубокое обучение Multi-Sequence LSTM-RNN и метаэвристика для прогнозирования электрической нагрузки // Энергии, 13(2), 391. 2020. <https://doi.org/10.3390/en13020391>.