

УДК 004-855-5

EDN
[JJYVPU](#)

Использование методов машинного обучения без учителя для обнаружения аномалий в энергопотреблении зданий

Д.А. Луговой*

Ярославский государственный технический университет, пр. Московский, 88, Ярославль, 150023, Россия

*E-mail: lugovoidmal@yandex.ru

Аннотация. В условиях стремительного роста энергопотребления и нарастающей угрозы энергетического кризиса эффективное управление энергоресурсами становится критически важной задачей. Одним из ключевых направлений в этой сфере является обнаружение аномалий в потреблении энергии, что позволяет своевременно выявлять случаи неэффективного использования ресурсов, технические неисправности оборудования, возможные утечки энергии, а также потенциальные кибератаки, направленные на энергосистемы. Если такие отклонения не выявлять оперативно, возрастает риск значительных финансовых потерь, увеличения эксплуатационных затрат, возникновения аварийных ситуаций, а также снижения общей надежности и устойчивости энергетической инфраструктуры. Для решения данной проблемы активно применяются современные методы машинного обучения, которые позволяют анализировать большие объемы данных, выявлять скрытые закономерности и прогнозировать возможные риски. Такие методы находят широкое применение в интеллектуальных системах мониторинга энергопотребления, позволяя оперативно обнаруживать неисправности, повышать энергоэффективность и предотвращать критические сбои в работе энергетических объектов. В данной работе рассмотрены передовые подходы к анализу данных энергопотребления с использованием методов машинного обучения, а также их практическое применение для выявления аномалий, повышения надежности энергосистем и оптимизации управления ресурсами.

Ключевые слова: информационные технологии, датасет, аномалии, энергопотребление, Python.

Using machine learning techniques without teacher to detect anomalies in the energy consumption of buildings

D.A. Lugovoi*

Yaroslavl State Technical University, 88 Moskovsky pr., Yaroslavl, 150023, Russia

*E-mail: lugovoidmal@yandex.ru

Abstract. In the context of rapidly growing energy consumption and the growing threat of an energy crisis, effective management of energy resources is becoming a critical task. One of the key areas in this area is the detection of anomalies in energy consumption, which allows timely identification of inefficient resource utilization, technical equipment malfunctions, possible energy leaks, and potential cyberattacks targeting energy systems. If such deviations are not promptly detected, the risk of significant financial losses, increased operating costs, emergencies, and a decrease in the overall reliability and sustainability of the energy infrastructure increases. To solve this problem, modern machine learning methods are actively used, which allow analyzing large amounts of data, identifying hidden patterns and predicting possible risks. Such methods are widely used in intelligent energy monitoring systems, enabling rapid fault detection, improving energy efficiency and preventing critical failures in energy facilities. This paper discusses advanced approaches to analyzing energy consumption data using machine learning techniques and their practical applications for anomaly detection, improving power system reliability, and optimizing resource management.

Keywords: information technology, dataset, anomalies, energy consumption, Python.

1. Введение

В современных условиях стремительного роста энергопотребления и возрастания вероятности глобального энергетического кризиса эффективное управление энергоресурсами приобретает особую значимость. Одним из ключевых направлений в данной сфере является своевременное выявление аномалий, сигнализирующих о нерациональном использовании ресурсов, технических неисправностях или попытках несанкционированного подключения к энергосетям.

Здания потребляют свыше 40 % всей вырабатываемой в мире энергии и являются источником более 30 % глобальных выбросов CO₂. Современные счетчики и интеллектуальные датчики, установленные в жилых и коммерческих объектах, ежедневно собирают и обрабатывают огромные массивы данных. При эффективном анализе эта информация может быть использована для выявления отклонений в энергопотреблении, что позволит конечным пользователям, поставщикам энергии и коммунальным предприятиям оптимизировать использование ресурсов.

2. Постановка задачи (Цель исследования)

Неисправность электрических устройств в зданиях может приводить к их нестабильной работе, что выражается в повышенном энергопотреблении, увеличении времени функционирования по сравнению с нормативными показателями и, в некоторых случаях, может привести к полному выходу оборудования из строя [1].

Для решения этой проблемы можно применять технологии искусственного интеллекта (ИИ), которые обеспечивают новые подходы к управлению энергопотреблением и повышению энергоэффективности зданий. Благодаря ИИ становится возможным автоматизировать обработку данных, что значительно ускоряет и улучшает процесс выявления отклонений в энергопотреблении.

3. Методы и материалы исследования

Аномалии в данных – это редкие или неожиданно возникающие отклонения, которые могут указывать на потенциальные проблемы, такие как поломки оборудования, ошибки в измерениях, незаконное вмешательство в систему или кибератаки. Методы машинного обучения позволяют автоматизировать процесс выявления таких отклонений, обеспечивая более высокую точность и надёжность по сравнению с традиционными методами анализа.

Машинное обучение представляет собой процесс, при котором компьютерные системы самостоятельно распознают закономерности в данных без явного программирования. Этот метод можно сравнить с обучением человека, который постепенно учится классифицировать объекты и находить связи между событиями. Современные алгоритмы машинного обучения позволяют автоматизированным системам решать сложные задачи, которые ранее требовали участия специалистов, тем самым расширяя возможности интеллектуального управления энергоресурсами [2].

Машинное обучение представляет собой совокупность методов и алгоритмов, которые позволяют компьютерам анализировать данные, выявлять закономерности и принимать обоснованные решения. В зависимости от характера задачи применяются различные подходы, одним из которых является обучение без учителя.

Обучение без учителя – это метод, при котором модель самостоятельно выявляет скрытые структуры в данных без заранее заданных правильных ответов. Такой метод широко применяется в задачах кластеризации, поиска аномалий и сокращения размерности данных [3]. Основная задача этого метода — выявление скрытых аномалий или нестандартных моделей потребления без предварительной информации о них. Алгоритм анализирует данные и ищет отклонения, которые ранее не были известны, позволяя обнаруживать неожиданные закономерности.

Кластерный анализ представляет собой метод разделения набора объектов на группы (кластеры) таким образом, чтобы внутри каждой группы элементы имели максимальное сходство по определённым характеристикам, а между группами различия были наиболее выраженными [4]. Кластеризация представляет собой метод машинного обучения, позволяющий разделять данные о потреблении электроэнергии на группы, чтобы выявлять нормальные и аномальные случаи без необходимости предварительной разметки.

Также известный как одноклассовое обучение (One-Class Learning, OCL), данный метод рассматривает исходные данные как принадлежащие одной из двух категорий: нормального или аномального потребления. При этом отрицательная группа (аномальные случаи) может быть частично известна, плохо определена или вовсе отсутствовать, что усложняет задачу классификации и требует разработки специальных алгоритмов.

В машинном обучении одним из эффективных способов предварительной обработки данных является снижение размерности, позволяющее исключить незначимые параметры и уменьшить избыточность информации. Это снижает вычислительные затраты и улучшает качество классификации данных об энергопотреблении. Для этой цели применяются различные методы, такие как анализ главных компонент (PCA), линейный дискриминантный анализ (LDA), квадратичный дискриминантный анализ (QDA) и множественный дискриминантный анализ (MDA).

Методы обучения без учителя (unsupervised learning) обладают значительными преимуществами, особенно в случаях, когда разметка данных недоступна или требует слишком больших затрат. Они позволяют моделям самостоятельно выявлять структуры в данных и адаптироваться к новым условиям без предварительного вмешательства человека.

Датасет представляет собой упорядоченный набор данных, организованный в табличной форме, где каждому объекту присваиваются определённые характеристики, связи или параметры. Такие структуры используются для построения аналитических моделей, тестирования гипотез и обучения нейросетей на основе исторической информации [5].

Выбранный для решения описанной проблемы набор данных включает в себя почасовые показания счетчиков за один год из 1636 нежилых зданий, собранных с 16 различных площадок по всему миру. Кроме этого, в нем содержатся такие метаданные о здании, как `square_feet`, `year_built` и `floor_count` для описания структуры здания (задается идентификатором `building_id`). Кроме того, к ним прилагаются различные погодные параметры, которые помогают лучше моделировать энергопотребление зданий.

Этот набор данных содержит измерения, полученные с помощью четырех различных типов счетчиков энергии (электричество, охлажденная вода, пар и горячая вода). Для задачи обнаружения аномалий были использованы данные почасовых показаний 1413 электросчетчиков из 16 различных типов зданий, таких как офисные, которые отслеживались в течение одного года. В данной работе делается акцент на счетчиках электроэнергии (1413).

Данный набор данных может быть разделена на раздел с точечными аномалиями и раздел с последовательными или коллективными аномалиями:

- 1) точечная аномалия;

2) последовательная или коллективная аномалия.

4. Полученные результаты

Для обработки датасета была использована специально разработанная программа на языке Python. Она производит подготовку данных из датасета для последующего анализа. Программа работает следующим образом. Изначально выделяются данные по энергопотреблению для одного здания (1319). Затем заполняются пропуски. После программа оставляет только одиночные аномалии. Данные разделяются на обучающие и тестовые множества. Полученные множества стандартизируются. В конце работы программы выводятся готовые графики (рисунки 1, 2, 3, 4).

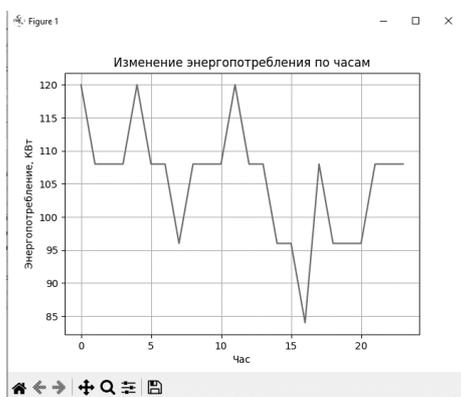


Рисунок 1. График № 1.

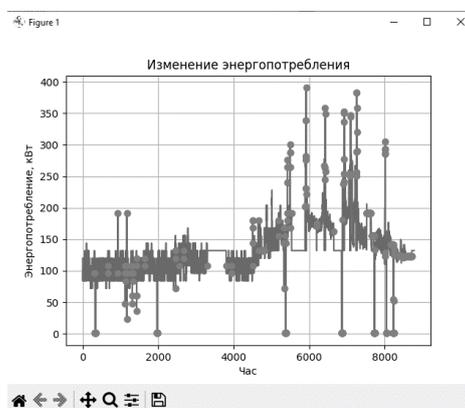


Рисунок 2. График № 2.

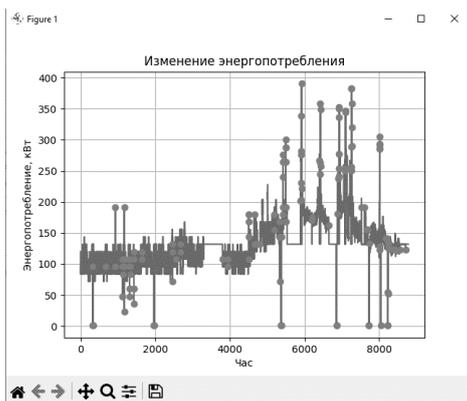


Рисунок 3. График № 3.

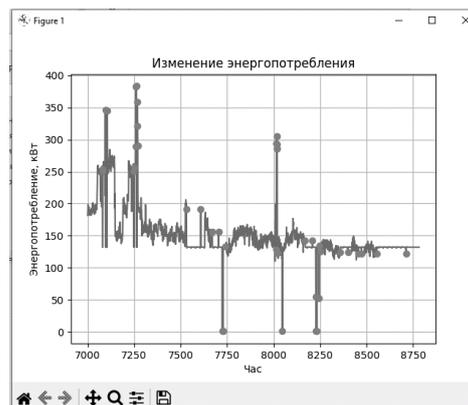


Рисунок 4. График № 4.

5. Выводы

В рамках работы был проведен всесторонний обзор современных методов обучения без учителя для анализа аномалий. Изучены и рассмотрены основные алгоритмы, применяемые для выявления аномалий в данных. Эти методы позволяют

эффективно обнаруживать выбросы в данных без необходимости наличия заранее размеченных данных, что особенно актуально для анализа энергопотребления, где аномалии могут возникать по различным причинам.

Для практической части исследования был найден и использован подходящий датасет, содержащий данные об энергопотреблении различных зданий. А с помощью специально разработанной программы была произведена обработка выбранного датасета.

Таким образом, проведенное исследование показало, что методы обучения без учителя могут успешно применяться для анализа аномалий в энергопотреблении, предоставляя полезные решения и способствуя более эффективному управлению энергоресурсами.

Список литературы

1. Himeur, Y. Artificial intelligence based anomaly detection of energy consumption in buildings: A review, current trends and new perspectives / Y. Himeur, Kh. Ghanem, A. Alsalemi, F. Bensaali, A. Amira // Applied Energy. – 2021. – V. 287. – P. 1-9. – URL: <https://www.sciencedirect.com/science/article/pii/S0306261921001409> (дата обращения: 30.01.2025).
2. Машинное обучение: просто о сложном. – cloud.ru: сайт. – 2021. – URL: <https://cloud.ru/ru/blog/machine-learning-about> (дата обращения: 30.01.2025).
3. Обучение без учителя (Unsupervised learning). – Loginom Wiki: сайт. – URL: <https://wiki.loginom.ru/articles/unsupervised-learning.html> (дата обращения: 30.01.2025).
4. Кластеризация. – Университет ИТМО: сайт. – 2022. – URL: <https://neerc.ifmo.ru/wiki/index.php?title=%D0%9A%D0%BB%D0%B0%D1%81%D1%82%D0%B5%D1%80%D0%B8%D0%B7%D0%B0%D1%86%D0%B8%D1%8F> (дата обращения: 30.01.2025).
5. Датасет: почему аналитику данных не обойтись без этого инструмента. – Россия – страна возможностей: сайт. – 2022. – URL: <https://rsv.ru/blog/dataset-pochemu-analitiku-dannyh-ne-obojtis-bez-etogo-instrumenta/> (дата обращения: 30.01.2025).