European Journal of Engineering and Technologies

Vol. 1 No. 1 2025

Review Open Access

The Promoting Role of Data Analysis Technology in Sustainable Energy

Bin Li [1,*]

[1] Columbia Climate School, Columbia University, New York, 10027, NY, United States
[*] Correspondence: Bin Li, Columbia Climate School, Columbia University, New York, 10027, NY, United States

Received: 20 April 2025
Revised: 23 April 2025
Accepted: 29 May 2025
Published: 03 June 2025

Abstract: Promoting the development of renewable energy is a key strategy to address global energy supply shortages and climate change issues. With the rapid advancement of data analysis technology, it has injected new vitality into the renewable energy industry. This article discusses the specific applications of core data analysis technologies such as artificial intelligence and machine learning, data mining and pattern recognition, predictive analysis and optimization algorithms in the field of energy. By utilizing data analysis technology, it is possible to monitor the energy network in real time and promptly issue fault alerts. At the same time, it can optimize the operation of the smart grid, increase the output efficiency of green energy, and provide scientific data support for the formulation of energy allocation strategies. Research has shown that data analysis technology provides solid support for the popularization and efficient management of sustainable energy, promotes the green transformation of energy structure, and provides feasible solutions for achieving global sustainable energy development.

Keywords: sustainable energy; data analysis; artificial intelligence; smart grid; clean energy

1. Introduction

Faced with the increasing global energy consumption and the severe challenges brought by climate change, countries have turned their attention to the exploration and application of renewable energy. However, the volatility and instability of renewable energy remain stumbling blocks in its promotion process. In this context, the in-depth application of data analysis technology in the energy industry has brought innovative solutions for the effective management of renewable energy. The advancement of artificial intelligence and machine learning, as well as the application of data mining and pattern recognition technologies, have improved the operational efficiency of energy systems, reduced operation and maintenance costs, and laid a solid foundation for promoting sustainable energy development strategies. The article analyzes the current application status of data analysis technology in the sustainable energy industry and looks forward to its future development direction, exploring the enormous potential of this technology in enhancing energy efficiency, improving energy allocation, and reducing environmental burden.

2. Overview of Sustainable Energy

Renewable energy refers to the types of energy converted from continuously renewable resources in nature, such as sunlight, wind, water flow, biomass energy, and geothermal energy. This type of energy is known for its virtually inexhaustible supply and low environmental burden, with minimal interference to the natural environment in production and application processes. Compared with conventional fossil fuels, the development and application of renewable energy can help reduce greenhouse gas emissions, slow down global climate change, and alleviate air pollution. In addition, promoting the development of renewable energy is also a key path to ensuring energy security, which helps to reduce dependence on overseas fossil fuels, improve the reliability and self-sufficiency of energy supply. With the continuous innovation of technology and the large-scale expansion of industries, the input cost of renewable energy continues to decline, and its economic value and market competitiveness are constantly increasing [1].

3. Data Analysis Techniques

3.1. Artificial Intelligence and Machine Learning

Among the numerous technologies in data analysis, artificial intelligence and machine learning are at the core, and their application in the field of sustainable energy is gradually expanding, as shown in Figure 1. Artificial intelligence technology can simulate certain aspects of human cognitive processes and model and optimize the complexity of energy systems. Machine learning technology focuses on mining patterns from data and building models for prediction and decision-making. The integration of these two technologies can efficiently process large datasets in energy systems and extract key information that is crucial for improving energy production, distribution, and utilization efficiency.

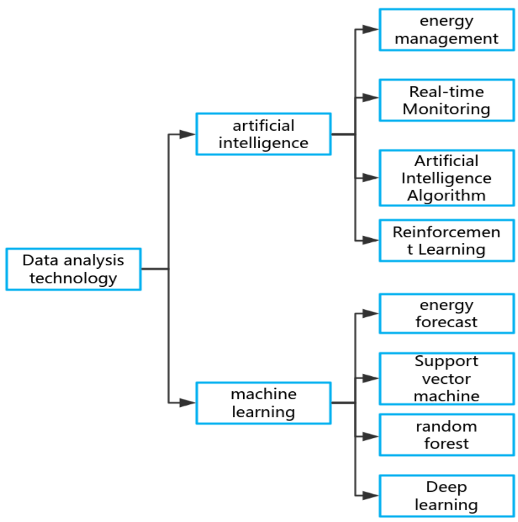

Figure 1. Data Analysis Techniques.

In terms of energy forecasting, machine learning algorithms can predict energy demand and renewable energy production through historical data and environmental variables. Algorithms such as Support Vector Machine (SVM) and Random Forest perform well in short-term load forecasting, playing a key role in balancing energy supply and demand contradictions [2]. Meanwhile, Long Short Term Memory (LSTM) networks in the field of deep learning have demonstrated extremely high efficiency in processing time-series data and are often used to predict the output of wind and solar photovoltaic power generation. Artificial intelligence technology in the field of energy management can intelligently adjust the allocation of energy through real-time analysis of energy network data, achieving efficiency improvement. In smart grid applications, artificial intelligence programs can optimize the transmission route of electrical energy and reduce losses during the transmission process. AI that integrates reinforcement learning technology can automatically adjust energy allocation strategies, which helps reduce operating expenses and environmental pollution. Taking wind power generation prediction as an example, the future wind power generation, denoted as Pt, is predicted using relevant environmental variables including wind speed v, air density ρ, and the swept area A of the wind turbine. Machine learning models can be built based on the following physical formulas:

$$Pt = \frac{1}{2} \cdot \rho \cdot A \cdot v_t^3 \cdot \eta \tag{1}$$

In formula (1), η is the efficiency of the fan, usually an empirical value. In practical applications, the model will be trained based on historical data and optimize prediction results by learning temporal patterns of v, ρ, and other environmental variables. The machine learning algorithm constructed with formulas helps energy managers achieve precise planning of wind farm operations, reduce energy losses or system overload caused by prediction deviations, and promote the optimized use of green energy.

3.2. Data Mining and Pattern Recognition

In the field of data analysis, data mining and pattern recognition are core technical means, focusing on extracting key information and typical feature models from complex energy information. Data mining utilizes statistics, intelligent algorithms, and database management techniques to process and analyze large-scale datasets. Pattern recognition focuses on exploring the inherent regularity and unique attributes of data. The integration of these two technologies has numerous application scenarios in the energy industry, including but not limited to energy demand forecasting, anomaly detection, and system performance optimization. In the analysis of electricity consumption, data mining techniques can be used to integrate past electricity usage records, meteorological conditions, and user daily behavior information to reveal user electricity consumption patterns and provide data support for the rational allocation of electricity resources [3]. By using clustering analysis and other algorithms, it is possible to estimate changes in energy demand in different regions and assist decision-makers in planning corresponding adjustment plans in advance. In the daily operation of smart grids, line damage or equipment degradation may cause abnormal conditions. Using pattern recognition methods can quickly identify abnormal patterns from numerous grid monitoring information, issue timely alerts, and prevent serious accidents from occurring. The method of integrating deep learning and pattern recognition is also applicable for detecting fault patterns in wind and solar energy equipment, improving maintenance plans, and extending the service life of equipment. The following is a statistical report on the electricity consumption of residents in a certain area (Table 1).

Table 1. Electricity Consumption Data Statistics Based on Data Mining.

User ID	Monthly electricity consumption (kWh)	Proportion of electricity consumption during peak hours (%)	Average daily electricity consumption fluctuation (kWh)	Classification tags
001	320	45	15	Peak users
002	280	25	8	Uniform user
003	450	60	20	Peak users
004	200	30	5	Uniform user
005	500	70	25	Peak users

The above data can be analyzed through data mining techniques. The above data can be analyzed through data mining techniques. Using the K-Means clustering algorithm, users can be divided into peak users and uniform users to optimize power grid operation strategy. The formula is as follows:

$$J = \sum_{i=1}^{k} \sum_{j \in C_i} \left\| x_j - \mu_i \right\|^2 \quad (2)$$

In formula (2), J is the clustering objective function, C_i is the i-th cluster, μ_i is the centroid of the i-th cluster, and x_j is the j-th sample. Based on clustering data analysis, power grid managers can implement differentiated time of use electricity pricing strategies for users during peak hours, incentivizing users to adjust their electricity consumption during peak periods, in order to balance the burden on the power grid and improve the overall efficiency of energy utilization.

3.3. Predictive Analysis and Optimization Techniques

Predictive analysis and optimization techniques, as the core part of data analysis, are committed to relying on past data and current situations to make advance judgments on future development trends and formulate optimal action strategies [4]. They integrate statistical principles, machine learning algorithms, and optimization techniques for operational research, providing precise decision assistance to complex energy networks. These technologies are widely used in the field of sustainable energy, covering energy consumption prediction, green energy output estimation, optimization configuration of power networks, and management of energy storage systems. By constructing a time series data model, predictive analysis can effectively predict future changes in energy supply and demand. By using regression analysis, deep learning networks, or time-series forecasting methods, it is possible to make forward-looking predictions on the fluctuations of urban power grid loads, which provides a scientific basis for the layout and management of the power system. Based on these predicted data, various optimization strategies have been implemented, relying on linear programming, dynamic programming, and metaheuristic algorithms (genetic algorithm and particle swarm optimization algorithm), which have been widely used in energy scheduling and configuration. In the management process of intelligent power grid, these optimization strategies help managers achieve load balance, reduce energy consumption costs, and minimize environmental pollution to the greatest extent possible. Taking power grid load forecasting as an example, it is necessary to predict future electricity demand y_t based on historical load data. Time series prediction models such as ARIMA (Autoregressive Integral Moving Average Model) are commonly used for such problems, and their mathematical formula is:

$$y_t = c + \varphi_1 y_t - 1 + \varphi_2 y_t - 2 + \ldots + \varphi p y t - p + \varepsilon t - \theta_1 \varepsilon t - 1 - \ldots - \theta q \varepsilon t - q \quad (3)$$

In formula (3), y_t is the predicted value of time t, c is a constant term, φ_i is the autoregressive coefficient, θ_j is the moving average coefficient, and $\epsilon\ t$ is the error term. By combining load forecasting and optimized scheduling, managers can scientifically allocate energy sources, such as prioritizing the use of low-carbon energy to reduce environmental impact. By combining predictive analysis and optimization techniques, the reliability and economy of the energy system have been improved, providing solid technical support for the efficient utilization of sustainable energy.

4. The Role of Data Analysis in Sustainable Energy

4.1. Real Time Monitoring and Fault Warning

The main purpose of real-time monitoring and fault warning system is to timely detect potential risks and faults in energy equipment or systems, reduce maintenance costs and energy waste, and ensure stable and reliable system operation. In the new energy industry, such as wind and solar power generation, monitoring systems rely on sensors to collect key data, including wind speed, light level, machine vibration, and temperature indicators. This information is transmitted in real time to the central monitoring system, and then processed by data analysis models to identify abnormal signal patterns [5]. The fault warning system utilizes pattern recognition and prediction algorithms to pre evaluate the danger of equipment operation. Anomaly detection methods based on machine learning algorithms, such as support vector machines (SVM) and deep neural networks, can automatically detect abnormal signals in energy networks and issue timely alarms. These models continuously improve the accuracy and response speed of early warning by analyzing historical data, ensuring the smooth operation of the energy system. Taking the fault prediction of wind turbines as an example, the key parameters monitored include speed (R), vibration amplitude (V), and temperature (T), and the real-time collected data is input into the anomaly detection model. A commonly used anomaly scoring formula is the Mahalanobis distance based detection method:

$$DM = \sqrt{(x-\mu)^T \sum -1\,(x-\mu)} \tag{5}$$

In formula (5), x is a real-time data vector, such as $[R, V, T]$. μ is the mean vector of historical data, $\sum$ is the covariance matrix of historical data, and DM is the Mahalanobis distance, representing the degree of difference between the current data point and the normal state. When $DM > \tau$ (preset threshold), the system determines that the state is abnormal and issues a warning. For example, in real-time monitoring, if $DM = 12$ at a certain moment exceeds the set threshold of 10, the warning system will immediately notify the operation and maintenance personnel to check the equipment to prevent potential faults from causing greater losses.

4.2. Optimization and Operation Management of Smart Grid

By integrating advanced data collection, transmission, and analysis technologies, smart grids can dynamically manage and optimize the flow of energy and information [6]. Its purpose is to improve the energy utilization efficiency of the power grid, reduce losses during transmission and distribution, and ensure that users enjoy a more reliable power supply. In the operation process of smart grid, data analysis technology is key and widely applied in various aspects such as load forecasting, supply and demand matching, and scheduling of energy storage systems. By utilizing past electricity consumption records and real-time electricity consumption data, data analysis techniques can effectively predict future electricity demand, make resource adjustments in advance, and prevent electricity shortages or surpluses. The following is an analysis of electricity demand in a certain region over a period of time:

Analyzing the data in Table 2, the system accurately calculated the energy difference for each time period and adopted optimization algorithms to adjust the operation of the energy storage system in real time. With the optimization and regulation methods of smart grid, the supply and demand of energy have been instantly balanced, which has built a solid defense line for the full utilization of sustainable energy.

Table 2. Energy Supply and Demand Data during a Certain Period of Time.

Point of time	Wind power generation (MW)	Solar power generation capacity (MW)	Energy storage system available capacity (MW)	User demand (MW)	Energy shortfall/surplus (MW)
08:00	15	20	10	40	+5
12:00	10	50	15	70	+5
18:00	5	10	30	55	-10
22:00	20	5	20	35	+10

4.3. Improving the Production Efficiency of Clean Energy

Green energy sources such as wind and photovoltaic energy are often affected by surrounding environmental factors, which leads to unpredictable fluctuations in their production capacity. With the help of data analysis technology, it is possible to accurately optimize the generation process of these renewable energy sources, reduce energy losses caused by external interference, and improve equipment utilization and electricity generation efficiency. In the power generation process of wind and solar energy, data analysis relies on monitoring real-time environmental information (including wind speed, light intensity, temperature, etc.), establishing predictive models, and adjusting the operating parameters of power generation equipment. Photovoltaic systems adjust the angle and operating conditions of solar panels through data analysis to reduce energy loss. Taking the photovoltaic power generation system as an example, the output power P of the photovoltaic panel is related to the light intensity I, ambient temperature T, and conversion efficiency η, and its formula is:

$$P=I \cdot A \cdot \eta(T) \tag{6}$$

In formula (6), P represents output power (watts), I is light intensity (watts/square meter), A is photovoltaic panel area (square meters), and $\eta(T)$ is the conversion efficiency that varies with temperature. A nonlinear regression model of $\eta(T)$ can be constructed through data analysis to predict the efficiency of photovoltaic panels at different temperatures. A certain photovoltaic power station has detected that the light intensity I = 800 W/m², A = 10 m², and T = 35°C. According to analysis, η (35°C) = 0.18, then the output power is calculated as follows: $P=800 \cdot 10 \cdot 0.18=1440W$. Through real-time monitoring and analysis, the output efficiency of renewable energy can significantly increase, making a solid contribution to the environmental transformation of the energy structure [7].

4.4. Optimizing Energy Dispatch Strategies

Optimizing energy scheduling strategies is the core approach to promoting efficient integration of renewable energy and conventional energy, and achieving sustainable development [8]. Energy dispatch refers to the scientific allocation of various energy sources in different temporal and spatial contexts, with the aim of improving energy efficiency, reducing negative environmental impacts, and lowering operational economic costs. Given the volatility and unpredictability of renewable energy, data analysis techniques are particularly important in the field of energy allocation. By relying on data analysis technology, the energy system can track changes in energy supply and demand, environmental conditions, and equipment operation in real time, and flexibly optimize allocation strategies based on this. For example, a certain energy dispatch system needs to meet the user demand D_t at different time periods, while optimizing the allocation of wind energy W_t, solar energy S_t, and energy storage system E_t, with the goal of minimizing costs. The optimization objective function and constraint formula are:

$$\text{Minimize } C = \sum_{t=1}^{T}(c_W W_t + c_S S_t + c_E E_t) \tag{7}$$

The constraint condition in formula (7) is that the total supply satisfies the demand $W_t+S_t+E_t \geq D_t, \forall_t$. The production limits for wind and solar energy are $0 \leq W_t \leq W_{max}$, $0 \leq S_t \leq S_{max}$, and the storage capacity limit is $0 \leq E_t \leq E_{max}$. At a certain time period, the user demand is D_t = 100MW, and the unit costs of wind energy, solar energy, and energy storage are 0.05, 0.06, and 0.08 yuan/kWh, respectively. Data analysis shows that the optimal allocation plan is W_t = 40MW, S_t = 30MW, and E_t = 30MW. The total cost calculation is $C=(0.05 \cdot 40) + (0.06 \cdot 30) + (0.08 \cdot 30)=6.3yuan$. This optimized scheduling scheme reduces costs while meeting demand, and maximizes the use of clean energy, promoting the sustainable development of the energy system.

5. Conclusion

With the rapid advancement of data analysis technology, it has shown new development space in promoting the efficient application and management of renewable energy. Data analysis technology is important for real-time system monitoring, early warning of faults, upgrading the intelligence of the power grid, improving the efficiency of green energy output, and optimizing energy allocation strategies. The data-driven energy sector is gradually moving away from traditional models and towards a transformation in intelligence and environmental protection, providing a solid technological foundation for the grand blueprint of global carbon neutrality. With the acceleration of technological innovation and the deepening of application, data analysis will play a more critical role in promoting the development of renewable energy, laying a key support for creating a more environmentally friendly, intelligent, and sustainable energy new era.

References

1. A. M. Dahunsi and B. A. K. Foli, “Assessment of past and future potential of ocean wave power in the Gulf of Guinea,” *Int. J. Sustain. Eng.*, vol. 16, no. 1, pp. 302–323, 2023, doi: 10.1080/19397038.2023.2269204.
2. K. Suriyan, P. Adhikary, K. Karpoora Sundari, M. C. Madhu, C. S. Madhusudhana, and B. N. Badari Narayana, "A novel reconfigurable hybrid DC-AC home technique with renewable energy resources and converters," *Int. J. Sustainable Eng.*, vol. 16, no. 1, pp. 285–301, 2023, doi: 10.1080/19397038.2023.2205872.
3. P. Sharma, “Analyzing the role of renewables in energy security by deploying renewable energy security index,” *J. Sustain. Dev. Energy Water Environ. Syst.*, vol. 11, no. 4, pp. 1–21, 2023, doi: 10.13044/j.sdewes.d11.0463.
4. P. Rentschler, C. Klahn, and R. Dittmeyer, “The need for dynamic process simulation: A review of offshore power-to-X systems,” *Chem. Ing. Tech.*, vol. 96, no. 1–2, pp. 114–125, 2024, doi: 10.1002/cite.202300156.
5. S. Kunwar, N. Pandey, P. Bhatnagar, G. Chadha, N. Rawat, N. C. Joshi, et al., “A concise review on wastewater treatment through microbial fuel cell: sustainable and holistic approach,” *Environ. Sci. Pollut. Res.*, vol. 31, no. 5, pp. 6723–6737, 2024, doi: 10.1007/s11356-023-31696-x.
6. M. I. Nazir, A. Ahmad, and I. Hussain, “Water cycle algorithm based parametric tuning of non-negative LMMN control of grid tied renewable energy systems,” *IETE J. Res.*, vol. 69, no. 12, pp. 9428–9444, 2023, doi: 10.1080/03772063.2022.2089748.
7. S. H. Ashrafi Niaki, Z. Chen, B. Bak-Jensen, K. Sharifabadi, Z. Liu, and S. Hu, “DC protection coordination for multi-terminal HB-MMC-based HVDC grids,” *IET Renew. Power Gener.*, vol. 18, no. 2, pp. 187–199, 2024, doi: 10.1049/rpg2.12903.
8. T. Papi Naidu, G. Balasubramanian, and B. Venkateswararao, “Optimal power flow control optimisation problem incorporating conventional and renewable generation sources: a review,” *Int. J. Ambient Energy*, vol. 44, no. 1, pp. 1119–1150, 2023, doi: 10.1080/01430750.2022.2163287.