BULLETIN of Cherkasy State Technological University

Journal homepage: https://bulletin-chstu.com.ua/en

Vol. 29 No. 2. 2024

UDC 621.3:004

DOI: 10.62660/bcstu/2.2024.79

Big Data technologies in the process of forecasting electricity generation from solar photovoltaic power plants

Oleksandr Stoliarov*

Postgraduate Student

National University of Food Technologies

01601, 68 Volodymyrska Str., Kyiv, Ukraine

https://orcid.org/0009-0001-7499-4672

Abstract. This research aimed to develop methods for using Big Data technologies to forecast electricity generation from solar photovoltaic power plants, which is crucial for optimising energy production and increasing the efficiency of solar resource utilisation. The study employed a method of analysing the economic feasibility of using energy storage systems and a comparative analysis of electricity buying and selling prices on the market. An experiment involving software tools and algorithms for processing, analysing, and modelling large volumes of data was also conducted. As a result of the research, methodologies were developed that encompass data collection and analysis, information visualisation, selection and training of forecasting models based on available data, as well as monitoring and testing their effectiveness. Graphical diagrams were constructed to illustrate the stages of data processing and analysis, the process of forecasting electricity generation for different time periods, and the process of training a model based on data, monitoring, and testing the model. Additionally, a graph was created to show the typicality and range of values, and a graph to display the change in electricity prices throughout the day. Furthermore, technological tools for using Big Data were described, the cost of electricity was calculated, and the economic attractiveness of using energy storage systems was assessed. As a result of the research, a potential profit indicator from price arbitrage was established, as well as economic parameters for the feasibility of using energy storage management based on an analysis of differences in electricity purchase and sale prices. The results obtained can be useful for energy companies and organisations involved in the production of electricity from solar photovoltaic power plants, allowing them to optimise energy production and increase the efficiency of solar resource utilisation

Keywords: information and analytical tools; energy market; big data processing; energy production optimisation; forecasting models

Article's History: Received: 05.01.2024; Revised: 15.04.2024; Accepted: 27.05.2024.

INTRODUCTION

Existing methods for predicting electricity generation have inherent limitations, as they do not always provide sufficient accuracy or the ability to anticipate changes in environmental conditions. Therefore, there is a need to develop new approaches based on modern Big Data technologies to improve forecasting efficiency and optimise electricity generation from solar photovoltaic power plants. The necessary theoretical knowledge includes concepts related to forecasting electricity generation from solar photovoltaic power plants and the use of Big Data technologies. This encompasses an understanding of the operating principles of solar photovoltaic power plants, their technical characteristics, and the influence of external factors on their performance. Additionally, it is important to understand Big Data technologies and their application in the energy sector, particularly methods for processing large volumes of data for forecasting and optimising electricity generation. This study examined how to implement Big Data technologies in forecasting electricity generation from

Suggested Citation:

Stoliarov, O. (2024). Big Data technologies in the process of forecasting electricity generation from solar photovoltaic power plants. *Bulletin of Cherkasy State Technological University*, 29(2), 79-92. doi: 10.62660/bcstu/2.2024.79.

*Corresponding author

solar photovoltaic power plants, how to use data from photovoltaic power plants to improve energy production efficiency, and which data analysis methods and algorithms can be applied to increase forecast accuracy.

To understand the current state of research in the field of Big Data technologies for electricity generation forecasting, it is necessary to refer to previous studies. In particular, authors U. Pysmenna *et al.* (2021) examined market mechanisms for stimulating the development of energy storage systems in Ukraine. They analysed the current state of such systems and the profitability of investment projects for their installation and operation to participate in various market segments. Researchers V. Stanytsina *et al.* (2023) investigated the possibilities of increasing the energy independence of territorial communities in Ukraine. They considered the criteria and technologies for distributed electricity generation and heat supply, as well as the use of renewable energy sources and local fuels to provide communities with electricity and heat. In their article, M. Kyzym *et al.* (2022) analysed the issues of energy efficiency and energy dependence in Ukraine, considering global trends. They developed a methodology for comparing the resource characteristics of power generation technologies to determine the most efficient Ukrainian solutions. Authors L.C. Kho *et al.* (2023) highlighted the important role of solar photovoltaic energy in distributed energy sources and explained that the amount of solar power generation is steadily increasing worldwide. They also noted that the high dependence of solar power generation on changes in ambient parameters and asset operating conditions is one of the challenges in solar photovoltaics.

Furthermore, S.S. Chandel *et al.* (2023) emphasised that accurate forecasting of electricity generation from industrial solar installations is crucial for ensuring grid stability. Their research presented an updated review of machine learning methods for photovoltaic energy forecasting. The research by K.J. Iheanetu (2022) demonstrated that while electricity production from renewable sources is rapidly increasing, its variable nature complicates forecasting. This poses problems for grid stability and reliability, alongside economic benefits. Additionally, R. Aman *et al.* (2023) highlighted that photovoltaic energy generation is dependent on variable meteorological conditions, and accurate photovoltaic energy estimation is essential for improving the performance of solar installations. A new deep learning model for photovoltaic energy forecasting based on various weather conditions was developed, employing a two-stage deep learning architecture combining long short-term memory (LSTM) and convolutional neural network (CNN). In the research of S. Tajjour & S.S. Chandel (2022), transfer learning was used to train a deep neural network based on solar irradiance data to forecast power output using a small dataset for a 400-kW solar plant in India. On the other hand, deep learning, which is widely used for photovoltaic generation forecasting, requires large amounts of data for proper training, which are initially unavailable for new power plants.

In their research, V.D. Gil-Vera & C. Quintero-López (2023) developed a model for forecasting solar energy production using machine-learning approaches and historical data. The results obtained indicate that the Random Forests method provides the most accurate forecasts, allowing grid operators to plan more effectively. Authors M.M. Hossain *et al.* (2023) highlighted the geographical and environmental characteristics of Bangladesh, where there is a need to find alternative energy sources. Solar power plants, particularly in the Char Fasson region, could be an effective solution as this region has a high potential for solar energy generation.

The aim of this research was to implement new methods for integrating Big Data technologies into the process of forecasting electricity generation from solar photovoltaic power plants. Overall, the study focused on addressing the problem of efficient use of solar resources and optimising energy production through the application of modern Big Data analysis technologies.

MATERIALS AND METHODS

This research addressed challenges related to the implementation of Big Data technologies in forecasting electricity generation from solar photovoltaic power plants, optimising energy production efficiency using data from photovoltaic power plants, and selecting the best data analysis methods and algorithms to improve forecast accuracy. To address these issues, a vast amount of data was collected using Market Operator (n.d.). The data was processed and analysed using tools such as Microsoft Word and Microsoft Excel, including statistical analysis and data visualisation. Trained models for forecasting electricity generation were tested on independent datasets to verify their effectiveness and accuracy.

After processing the data, forecasting models for electricity generation from solar photovoltaic power plants were developed and trained based on the collected data. Input data for training the models included various parameters affecting the electricity generation process. Specifically, data on solar activity levels, weather conditions (temperature, humidity, precipitation, etc.), geographic parameters (latitude, longitude, altitude), and the technical characteristics of the photovoltaic power plant itself. Model training was conducted on input data that had been carefully collected and processed. After training, the models were tested on independent datasets that were not used during training. This allowed for verification of the model's effectiveness and accuracy in forecasting real data that had not been used for their training.

An important aspect was determining the hourly buying and selling price in the "day-ahead" market (DAM) for the period from 14:00 to 15:00 (1):

$$P_{14\text{-}15} = P_{14.00} \cdot 0.5 + P_{15.00} \cdot 0.5, \qquad (1)$$

where $P_{14\text{-}15}$ – the hourly DAM price for the operating period from 14:00 to 15:00 UAH/MWh; $P_{14.00}$ – the hourly DAM price at 14:00 UAH/MWh; $P_{15.00}$ – the hourly DAM price at 15:00 UAH/MWh.

Descriptive and inferential statistical methods were employed to analyse the data. These included analysis of variance, correlation analysis, and regression analysis. Each method was selected based on its suitability for the specific data type and research objectives. The data collection process was standardised and documented to ensure data quality and reliability. To assess the economic attractiveness of using energy storage systems, an analysis of the daily load market was conducted, and electricity costs were calculated based on the difference between electricity purchase and sale prices.

Two approaches were considered for calculations based on actual data from photogalvanic power plants (PGPPs). In the first approach, to assess the additional revenue from arbitrage and the payback period, actual photovoltaic power plant data was used, with the analysis based on hourly buy/sell prices of electricity on the DAM in 2023. To ensure objective assessment, calculations were performed for each month of the seasonal period (February, May, August, and November). Using a graphical method, the day with the minimum and maximum DAM prices for each corresponding month was determined, as well as the volume of electricity generation during the period of minimum prices. The second approach considered a simplified method for calculating additional arbitrage revenue, based on actual solar photovoltaic power plants data for 2023 and the difference between electricity purchase and sale prices determined by the 2023 price caps. Price caps, set by the National Commission, are the maximum and minimum prices for electricity on the Ukrainian electricity market. The application of this method allows for the estimation of the probable volume of electricity sales and its cost during periods of maximum and minimum prices, as well as determining the difference in sales value throughout the year. Formula (1) was used for the calculations.

The research also included an analysis and evaluation of the effectiveness of the developed methods. Additionally, software tools such as Apache Hadoop, Apache Spark, and Apache Kafka were described and analysed. All experiments were conducted using computers with high-performance computing capabilities, as well as specialised software for data processing and analysis.

RESULTS

Development of methods for utilising Big Data technologies to forecast electricity generation from solar photovoltaic power plants. The development of efficient methods for predicting electricity generation from solar photovoltaic power plants is becoming increasingly relevant in the context of the growing role of solar energy in modern energy systems. The use of Big Data technologies in this context opens up wide opportunities for improving forecasting through the analysis of large volumes of data and the application of specialised algorithms.

One method of utilising Big Data for forecasting electricity generation involves the collection and processing of large volumes of data. This procedure includes gathering a variety of information about electricity generation from solar photovoltaic power plants, such as data on solar irradiance, temperature, cloud cover, and other factors. Subsequently, data is processed to remove noise and anomalies, such as data loss or inaccuracies that may occur during collection. Additionally, the collected data can be divided into three main categories: meteorological data, solar photovoltaic power plant status data, and geospatial data. Meteorological data includes both historical and forecasted values of solar radiation, temperature, humidity, wind speed, and cloud cover. Data on the status of solar photovoltaic power plants includes the characteristics of panels and inverters, as well as operational data such as voltage, current, and power. Geospatial data encompasses digital elevation models for determining the tilt angle and azimuth of panels, alongside information on land use and shading. Data storage is also a critical consideration. Large volumes of information, such as data on solar activity, temperature, and other parameters, can be collected from photovoltaic power plants and environmental sensors. Storing this data in specific distributed data storage systems allows for efficient storage and processing of information for further analysis. The flowchart in Figure 1 illustrates the data processing and analysis process. The first block shows data collection and processing. The second block presents the analysis and visualisation of this data. The third block depicts the data storage process.

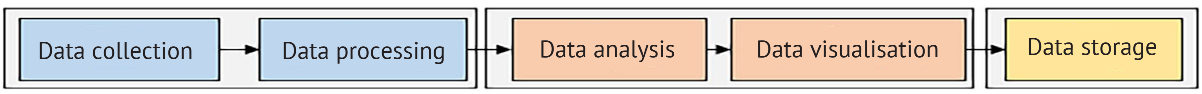

Figure 1. Data collection and processing

Source: compiled by the author

Another method of utilising Big Data involves data analysis and visualisation. After processing information about electricity generation from solar photovoltaic power plants, analysis is conducted to identify dependencies and correlations between various factors such as solar irradiance, temperature, time of day, etc. Various visualisation methods, such as graphs, charts, and heatmaps, are used to better understand the relationships between data. Data analysis includes data cleaning and processing, such as removing errors, missing values, and outliers, as well as normalising and standardising data. It also involves pattern discovery, i.e., using machine learning and data mining techniques to identify relationships between meteorological, operational, and geospatial data and electricity generation. Additionally, an important aspect is modelling, which involves developing models for forecasting electricity generation based on identified patterns. Overall, data analysis involves using machine learning and data mining techniques to identify relationships and patterns in the obtained data, forecasting electricity generation based on historical data and factors such as weather conditions, geographic location, etc. In turn, visualisation involves creating interactive monitoring and forecasting dashboards, visualising electricity generation forecasts, and comparing forecasts with actual values.

The next method involves the selection of a forecasting model (Fig. 2). This could be a statistical model, a machine learning model, or a neural network. Key factors in model selection include forecast accuracy, model complexity, and the ability to be extended to accommodate new data. Forecasting covers various time horizons, including short-term, medium-term, and long-term. Short-term forecasting involves generation in the next few hours or days, medium-term forecasting – in the next few weeks or months, and long-term forecasting – in the next few years, using methods such as energy models and system dynamics. The use of forecasting models involves developing models based on collected data and analysing them using machine learning methods such as linear regression, ensemble methods, and neural networks. These models are used to forecast electricity generation based on various factors. Initially, a forecasting model is selected, which can be statistical or machine learning-based. These models are then used to develop forecasts for short-term, medium-term, and long-term time horizons. The result is electricity generation forecasts that can be used for decision-making in the energy sector.

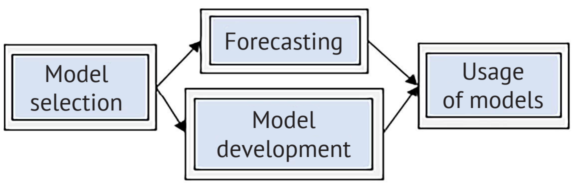

Figure 2. Choosing a forecasting model
Source: compiled by the author

Conversely, the method of training a model based on data involves conducting a detailed analysis and processing of the collected data before the actual training process begins (Fig. 3). After selecting the appropriate model, training starts with the aim of adapting the model to the input data and identifying patterns and relationships between them. However, attention should first be paid to data preparation, including cleaning it of anomalies, removing duplicates, and correcting missing or incorrect values. The training dataset can be divided into two parts: training and testing, for effective model validation. The model is then trained by adapting to the patterns and dependencies in the training data. This involves tuning the model's parameters to maximise forecast accuracy. The model can use various training methods such as gradient descent, backpropagation, etc., depending on the chosen algorithm. After model training is complete, it is evaluated using the test dataset to verify its accuracy and effectiveness. This process allows for an evaluation of how well the model can generalise to new, previously unseen data and draws conclusions about its predictive capabilities. The first block is responsible for the data preparation stage, the second – for model training, and the third – for model evaluation.

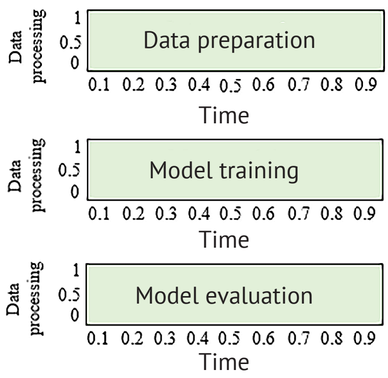

Figure 3. Model training based on data
Source: compiled by the author

The method of monitoring and testing the model involves implementing monitoring systems, testing, and evaluating the model. A crucial aspect of this method is the development of monitoring systems that can collect real-time data on solar activity, weather conditions, and the performance of photovoltaic power plants. The collected data is used for continuous improvement of forecasting models, ensuring their accuracy and adaptability to changing environmental conditions. Attention should also be paid to testing and evaluating the model on a dedicated test dataset. This allows for determining the accuracy of the model's forecasts and identifying its strengths and weaknesses. If the model proves satisfactory during testing, it can be considered suitable for use in forecasting electricity generation from solar photovoltaic power plants. The flowchart in Figure 4 illustrates the model monitoring and testing method. The first block shows the process of continuously monitoring the model's accuracy and adaptability by collecting feedback. The second block highlights the importance of developing monitoring systems to collect real-time data on solar activity and other factors. The third block shows the process of testing and evaluating the model to determine its accuracy, as well as its strengths and weaknesses. Thus, data analysis and visualisation, data-driven model training, and the model monitoring and testing method are key stages in this process. Overall, the use of Big Data methods for forecasting electricity generation from solar photovoltaic power plants allows for improved forecasting accuracy, ensuring that models are adaptable to changing environmental conditions and enabling the efficient use of solar energy in modern energy systems.

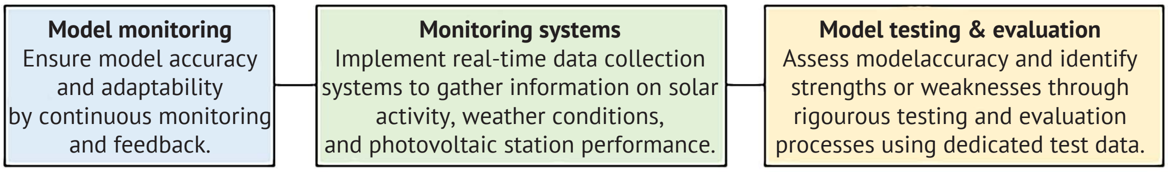

Figure 4. Model monitoring and testing

Source: compiled by the author

The use of Big Data and energy storage to optimise solar power generation and improve the efficiency of energy companies. The question of the effectiveness of the developed methods should also be considered in the context of Ukraine's Energy Strategy until 2035, as the significant impact of these methods on the power balance in the system is highlighted. The introduction of a share of renewable energy sources (RES), in particular photovoltaic power plants, at the level of 25% by 2035, poses the task for the energy system of ensuring the stability and reliability of power supply. The developed forecasting methods prove to be an effective tool for addressing these challenges, contributing to better energy resource management and ensuring the quality of power supply for consumers. In addition, the application of energy storage, particularly based on calculations using the universal energy storage (UES) model, aims to increase the efficiency of energy companies and improve the quality of services in the field of electricity generation. Such an approach to using the developed methods not only contributes to the optimisation of processes in the energy sector but also to achieving strategic goals in the development of renewable energy in Ukraine. The use of Big Data technologies and energy storage in the energy sector has significant potential for optimising solar power generation and increasing the efficiency of energy companies. One of the key aspects in this context is the concept of "arbitrage", which involves the commercially attractive accumulation of electricity during periods of low prices and its subsequent release into the grid during periods of high prices. This strategy allows for smoothing out nighttime troughs as well as morning and evening peaks of the daily load curve of the energy system. To substantiate the feasibility of using energy storage in conjunction with photovoltaic power plants in the conditions of the modern Ukrainian energy market, calculations were carried out based on actual data of the PGPPs operation for two scenarios. The first scenario was based on hourly buy-sell prices of electricity on the DAM in 2023, while the second scenario took into account the difference in buy-sell prices of electricity on the DAM, using price caps for 2023. Such an approach allows for providing convincing data on the prospects of using these methods in solving the current tasks of the energy sector (Fig. 5).

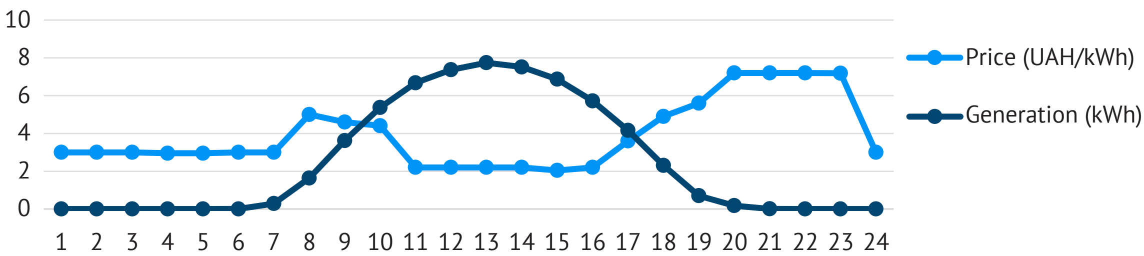

Figure 5. Graph of hourly prices for the buying and selling of electricity on the DAM and the generation of PGPPs for August 2023

Source: compiled by the author based on Market Operator (n.d.)

After this, calculations were conducted to estimate the expected volume of electricity sales and its value during periods of minimum and maximum prices (Table 1). The difference in the value of electricity sales was assessed for each month, as well as in total for the entire year based on the obtained results.

Table 1. Calculation of the difference in the value of electricity sales in 2023

	Sales volume (kWh)	Price (UAH/kWh)	Cost (UAH)	Difference (thousand UAH)	Difference (thousand UAH/ month)
February					
12:00-17:00	1027.33	3.790-3.840	3,905.56	0.039	1.1
19:00-20:00	1027.33	3.840	3,944.95		
May					
12:00-17:00	3099.38	2.650	8,213.36	4.184	129.71
19:00-20:00	3099.38	4.000	12,397.52		
August					
12:00-17:00	2910.94	2.200	6,287.01	14.55	451.05
19:00-21:00	2910.94	7.200	20,958.77		
November					
12:00-17:00	360.24	5.515-5.224	1,959.80	0.13	3.89
19:00-20:00	360.24	5.800	2,089.39		

Source: compiled by the author based on Market Operator (n.d.)

Final calculations were conducted to estimate the expected revenue from arbitrage, taking into account the self-consumption costs of solar photovoltaic power plants, as well as the costs of dispatching and transmitting electricity for 2023. However, these calculations did not consider any negative factors that could impact electricity generation. It is important to highlight the final calculations of the investment attractiveness of employing energy storage systems for collaboration with solar photovoltaic power plants to provide "arbitrage" generation services (Table 2). The assessment of investment attractiveness, considering the stated assumptions and parameters, indicates that the use of energy storage is commercially viable in this case. The projected annual return is approximately 40%, ensuring a return on investment within six years, which aligns with the projected lifespan of the battery energy storage systems. Considering the additional income, the payback period is reduced to 3.8 years. The second scenario employs a simplified method for calculating additional arbitrage revenue, relying on actual data from solar photovoltaic power plants in 2023 and the price difference between buying and selling electricity, determined by the price caps set for 2023. Based on the results obtained, the potential arbitrage revenue for 2023 was calculated, excluding self-consumption costs and other operational expenses (Table 3).

Table 2. Economic indicators of the investment attractiveness of using UES for combined operation with PGPPs based on 2023 data

No.	Indicator	Result
1	Volume of electricity sales (MWh)	1,156.8
2	Revenue from electricity sales per year (thousand UAH)	3,065.29
3	Operational costs per year (thousand UAH)	15.2
4	Costs for transmission system operator (TSO) and distribution system operator (DSO) services per year (thousand UAH)	628.63
5	Actual revenue for the year (thousand UAH)	2,421.46
6	Probable additional revenue for the year (thousand UAH)	1,757.25
7	Probable total revenue for the year (thousand UAH)	4,178.7
8	Average electricity generation from 12:00 to 17:00 (MW)	1,849.5
9	Required installed capacity of UES for RES (kW)	1.0
10	Cost of energy storage installation per kW (USD)	210
11	Exchange rate for the calculation period (UAH)	37.5
12	Estimated budget component of the UES project (CAPEX) (thousand UAH)	15,750
13	Estimated budget component of the UES project (OPEX) (thousand UAH)	123
14	Total costs for UES (thousand UAH)	15,873
15	Payback period of investments (years)	6.6

Continued Table 2.

No.	Indicator	Result
16	Payback period of investments considering probable additional revenue (years)	3.8
17	The projected operational lifespan of battery UES (years)	15

Source: compiled by the author based on Market Operator (n.d.)

Table 3. Potential arbitrage revenue for the year, excluding self-consumption costs, and costs for TSO and DSO services

Period	Potential income (thousand UAH)	Hours of minimum on the DAM	Price caps, min-max (UAH/MW)	Difference per 1 MW by price (UAH)	Difference by 1 MW in fact
December	1,696.20	13-15	2,646.25-4,000.0	max-1,353.75; min-0	54.72
January	1,887.04	12-13	2,646.25-4,000.0	max-1,353.75; min-0	60.87
February	5,889.25	13-14	2,646.25-4,000.0	max-1,353.75; min-0	210.33
March	11,515.95	14-15	2,646.25-4,000.0	max-1,353.75; min-0	371.48
April	27,642.80	14-15	2,646.25-4,000.0	max-1,353.75; min-0	921.43
May	41,644.45	13-14	2,646.25-4,000.0	max-1,353.75; min-0	1,343.37
June	38,286.05	13-14	2,646.25-4,000.0	max-1,353.75; min-0	1,276.20
July	152,020.92	13-14	10 (5,600.0)-7,200.0	max-7,190; min-1,600	4,903.90
August	156,585.05	13-14	10 (5,600.0)-7,200.0	max-7,190; min-1,600	5,051.13
September	154,151.91	13-14	10 (5,600.0)-7,200.0	max-7,190; min-1,600	5,138.40
October	151,152.02	13-14	10 (5,600.0)-7,200.0	max-7190; min-1,600	4,875.87
November	76,690.38	12-13	10 (5,600.0)-7,200.0	max-7,190; min-1,600	3,334.36
Total for the year	819,162				

Source: compiled by the author based on Market Operator (n.d.)

Final calculations were performed to assess the investment attractiveness of implementing a UES system based on its 2023 operational data, disregarding any negative factors that might affect electricity generation (Table 4).

Table 4. Economic parameters of the feasibility of using UES based on the difference in electricity purchase and sale prices on the DAM in 2023

Indicator	Result
Costs for transmission system operator (TSO) and distribution system operator (DSO) services per year (thousand UAH)	628.63
Volume of electricity sales (MWh)	1,156.8
Revenue from electricity sales per year (thousand UAH)	3,065.29
Operational costs per year (thousand UAH)	15.2
Actual revenue for the year (thousand UAH)	2,421.46
Probable additional revenue for the year (thousand UAH)	819,162
Probable total income for the year (thousand UAH)	3,240.62
Required installed capacity of UES for RES (MW)	1.0
Cost of energy storage installation per kW (USD)	210
Exchange rate for the calculation period (UAH)	37.5

Continued Table 3.

Indicator	Result
Estimated budget component of the UES project (CAPEX) (thousand UAH)	15,750
Estimated budget component of the UES project (OPEX) (thousand UAH)	123
Total costs for UES (thousand UAH)	15,873
Payback period based on actual income (years)	6.52
Payback period considering additional income (years)	4.9
The projected operational lifespan of battery UES (years)	15

Source: compiled by the author based on Market Operator (n.d.)

The assessment of investment attractiveness indicators under the considered conditions and parameters confirms that the use of a coordinated UES in this case is economically viable, as the expected additional annual income is approximately 25%. The payback period does not exceed the planned lifespan of battery energy storage systems and is 6.52 years; considering the expected additional income, this period is reduced to 4.9 years.

Thus, the calculated investment attractiveness indicators in both UES usage scenarios are almost identical, indicating its commercial attractiveness in the "arbitrage" generation sector. However, there are two key aspects that can affect the expected profit. Firstly, the highly unstable nature of generation from PGPPs depends on various factors such as weather conditions and equipment technical conditions. Secondly, the regulatory aspects of the Ukrainian electricity market. The implementation of ESS (energy storage systems) requires significant investments and the correct choice and consideration of the expected economic effect is a complex task. Given the various energy storage technologies and their economic maturity, many aspects of their application must be considered. The business model for using UES in the electricity market depends on the intended purpose of the electrical installation and the chosen market segment.

Technological tools and software in Big Data analysis. Within the realm of Big Data, numerous software tools are employed for the collection, processing, and analysis of vast datasets. Apache Hadoop stands out as one of the most popular open-source frameworks for the distributed processing of large data volumes (Fig. 6). It is grounded in the MapReduce model, which enables efficient data processing across a cluster of servers. Hadoop encompasses various components such as the Hadoop Distributed File System (HDFS) for storing data in blocks across multiple servers, MapReduce for data processing and parallel computing, and YARN for resource management and job scheduling within the cluster. A key advantage of Hadoop is its scalability, allowing it to handle massive datasets distributed across dozens or even hundreds of servers.

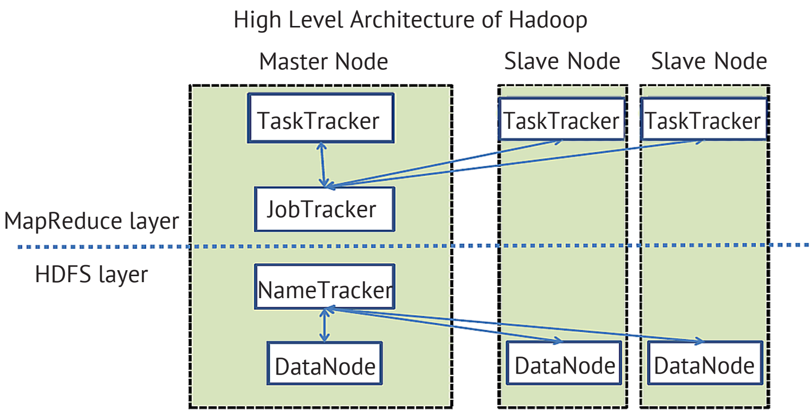

Figure 6. The main components of Apache Hadoop

Source: S. Bappalige (2014)

In turn, Apache Spark is a fast and powerful framework designed for the real-time processing of large datasets (Fig. 7). It provides an application programming interface (API) for distributed data processing, supporting programming languages such as Python, Java, and Scala. Spark leverages in-memory caching and optimised operations, enabling it to perform significantly faster than Hadoop MapReduce. Moreover, Spark incorporates modules for machine learning, deep learning, and graph computations, making it a versatile tool for data analysis. At the core of Spark lies the concept of resilient distributed datasets (RDDs), which serve as its fundamental data structure. Spark offers an API for manipulating RDDs, as well as modules for working with streaming data, machine learning, and structured query language (SQL) queries.

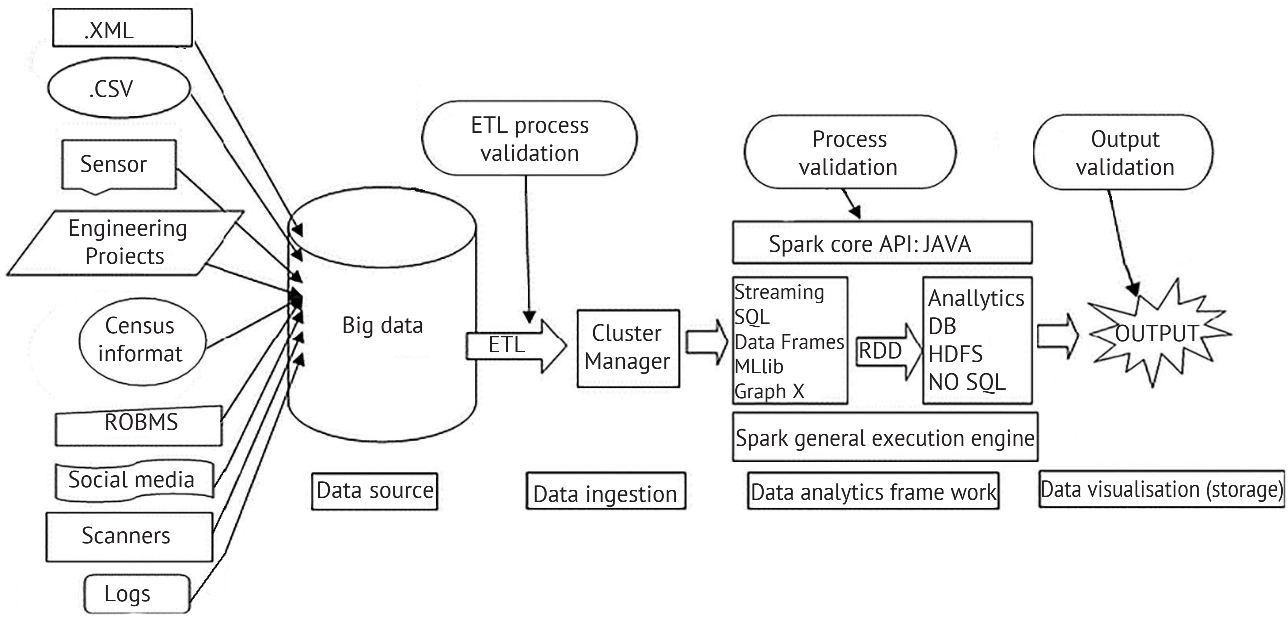

Figure 7. Big Data testing ecosystem with Apache Spark

Source: R.K. Pallamala & P. Rodrigues (2022)

Another notable technology is Apache Kafka, a distributed streaming platform designed for real-time data processing (Fig. 8). It facilitates efficient data exchange between various systems and applications, while also storing data for subsequent analysis. Kafka offers horizontal scalability and guarantees message delivery and order preservation. Apache Kafka is commonly used for real-time data streaming between different systems. It consists of three primary components: producers, which generate data streams; topics, which serve as categories or channels for data streams; and consumers, which process or store the data streams. Kafka ensures high availability and scalability, making it an excellent tool for stream processing.

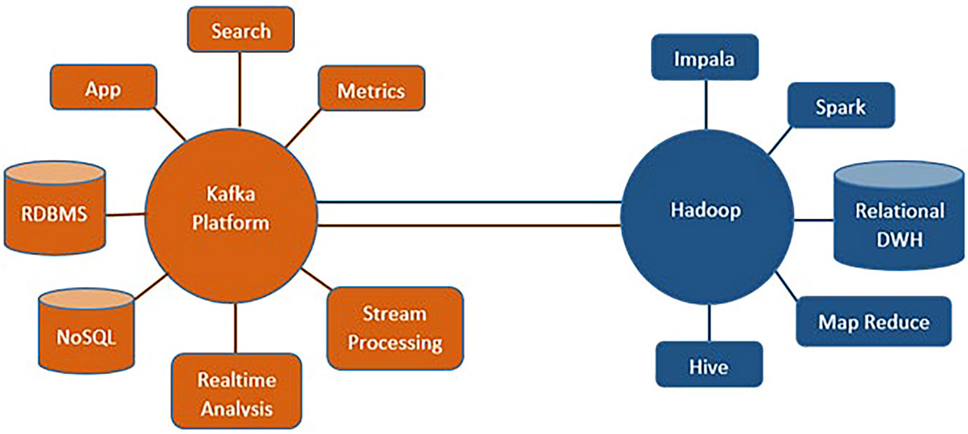

Figure 8. Comparison of Kafka and Hadoop platforms

Source: N. Saxena (2019)

Additionally, during the data processing phase, extract, transform, load (ETL) tools such as Apache NiFi or Talend can be employed. These tools enable the cleaning, transforming, and loading of data from various sources into appropriate systems for further analysis. The ETL process encompasses the stages of data extraction, transformation, and loading from one source to another, ensuring data quality and readiness for analytical use.

DISCUSSION

This study highlights the relevance of using Big Data models for forecasting electricity generation from solar photovoltaic power plants. The study by G. de Freitas Viscondi & S.N. Alves-Souza (2019) corroborates these findings, as it employed a methodology that confirmed the applicability of Big Data models for predicting electricity generation, particularly emphasising the accuracy of machine learning and neural network methods.

Furthermore, the conducted research on Big Data technologies acknowledges machine learning and neural networks as the most precise techniques for addressing this issue. However, this research also expands upon the understanding of this topic by utilising various methods and tools for data analysis.

This research expands upon the study of S.-M. Je *et al*. (2021), attempted to accurately forecast electricity demand through Big Data visualisation and proposed a flexible and balanced Big Data virtualisation for power generation based on a photovoltaic power plant is proposed. In contrast to the researchers' article, this study incorporates not only Big Data visualisation but also other tools for accurately predicting electricity production from solar photovoltaic power plants. This enables a more comprehensive examination of various aspects of this problem and facilitates the discovery of effective solutions to further advance renewable energy.

Similar to the research of M.A. Nazari *et al*. (2023), this research identifies solar energy as a key renewable source with significant potential in the modern energy landscape. However, unlike the researchers' study, this Big Data research focuses on developing and applying various methods and software tools for accurately forecasting electricity generation from solar photovoltaic power plants. This can facilitate the integration of solar energy into production processes.

To address the issue of solar energy intermittency within power grids, Y. Nie *et al*. (2022) identified that image-based forecasting using deep learning is a promising method for predicting short-term fluctuations. However, this study differs from the researchers' approach to solar energy forecasting. The current research focuses on leveraging Big Data to accurately predict electricity generation from solar photovoltaic systems, providing a deeper understanding and more precise forecasts for the integration of solar energy into the grid.

Given the pressing global energy and environmental challenges, the development of renewable energy, particularly photovoltaic electricity generation, is of paramount importance. The research by W. Liu (2023) highlights that the vast potential of solar energy paves the way for the advancement of photovoltaic technologies and industrial growth, with solar energy primarily utilised in photovoltaic power generation systems. The current study on Big Data technologies complements the researchers' article by examining specific methods and software tools for forecasting electricity generation from solar photovoltaic power plants. This study provides more accurate and efficient forecasts to facilitate the integration of solar energy into power grids, leveraging large datasets and diverse analytical approaches.

As the expansion of photovoltaic electricity generation necessitates accurate solar energy forecasting, this research investigates the potential of Big Data and explores specific approaches to predicting electricity production from solar photovoltaic power plants. In turn, the research conducted by Z. Zhang *et al*. (2023) offers an effective model for forecasting electricity generation using deep learning. Nevertheless, this study broadens the arsenal of methods for predicting energy performance and provides additional opportunities for precise and long-term forecasting of electricity generation from solar photovoltaic power plants.

D.K. Dhaked *et al*. (2023) focused on utilising LSTM and back propagation neural network (BPNN) models to forecast solar power generation. While the current study also explores a similar aspect, it concentrates on alternative forecasting methods, employing a variety of approaches and software tools for data analysis and modelling. This expands the repertoire of forecasting methods and offers additional capabilities for accurate and long-term prediction of solar photovoltaic power generation.

In comparing this study with the research conducted by K.E. Sarah *et al*. (2020), it is important to note that the present study addresses the issue of accurately forecasting electricity generation from solar photovoltaic power plants. In contrast, the other study provides a comprehensive overview of the progressive development of solar photovoltaic technologies from first-generation to contemporary configurations, as well as an analysis of solar power generation in various regions. Thus, the conducted research on Big Data technologies broadens the scope by presenting an analysis of diverse software tools and methods for accurate and long-term forecasting of electricity generation from solar photovoltaic power plants.

Accurate forecasting of next-day solar photovoltaic energy is essential for the successful integration of solar power into the electricity grid. J.F. Torres *et al*. (2019) compared the effectiveness of a deep learning-based algorithm with other advanced methods for predicting solar energy and analysed the suitability of all methods for processing large time series data. In turn, this Big Data research expands upon the researchers' study by exploring various methods for modelling and analysing data to forecast electricity generation from solar photovoltaic power plants. It provides additional capabilities for accurate forecasting using diverse approaches, which can contribute to improving the efficiency of energy integration into the electricity grid.

Given the highly variable nature of electricity generation from solar PGPPs due to various factors, accurate forecasting is crucial for reliable grid integration and ensuring energy supply in off-grid systems. M.B. Arias & S. Bae (2021) proposed a forecasting model to estimate solar PGPPs output based on historical data such as solar irradiance, solar energy, and meteorological data, which are processed using Big Data techniques. However, the current research, while also considering electricity generation, explores a broader range of factors. It may also contribute to improving forecasting accuracy and the reliability of integrating solar energy into the grid.

Commonalities between this research and the study of Y. Zhang & Y. Wang (2022) include the exploration of a hybrid CNN and LSTM approach for forecasting solar power generation. M. Sabri & M. El Hassouni (2023) also proposed using LSTM, specifically with an autoencoder, for photovoltaic power generation forecasting. Additionally, R. Vaish & U.D. Dwivedi (2022) compared the LSTM model but also mentioned the Seasonal Autoregressive Integrated Moving Average (SARIMA) model, which is not included in this study.

In contrast to the article of C. Soumya *et al*. (2021), which examines the establishment of decentralised autonomous solar systems in rural areas of countries with limited fossil resources, the present study on Big Data technologies does not focus on addressing the challenges of installing solar systems in specific countries. Meanwhile, J. Gaboitaolelwe *et al*. (2023) provided a comprehensive overview of the latest advancements in existing approaches to forecasting energy generation from solar photovoltaic systems based on machine learning, which are also applied in this study.

While this research specifically analysed solar energy forecasting, Rajnish *et al*. (2024) also considered wind energy, which plays a crucial role in ensuring reliable power supply and grid stability. Conversely, A. Nahar *et al*. (2024) employed a machine learning time series model to forecast real-time power generation in two solar power plants, whereas this study utilises Big Data technologies for forecasting electricity generation from solar photovoltaic power plants.

This study, unlike the article of N. Kumar & M.M. Tripathi (2022), focuses on the utilisation of Big Data technologies for forecasting electricity generation from solar photovoltaic power plants. While the other study analysed the impact of solar power generation on electricity price forecasting, this study, similar to Suwarno & D. Pinayungan (2024), investigates solar power generation as a source of renewable energy.

This research places particular emphasis on developing forecasting models that can overcome the limitations associated with the variability and unpredictability of energy production from solar photovoltaic power plants. It also highlights the importance of accurate power forecasting for ensuring reliable electricity supply and grid stability. Compared to other studies, this research stands out for its unique approach, high complexity, and focus on the real-time efficiency of solar photovoltaic power plant forecasting. This study aimed to improve forecasting tools and methods to ensure the stability and efficiency of using solar energy as a renewable energy source.

CONCLUSIONS

This research developed effective methods for utilising Big Data technologies to forecast electricity generation from solar photovoltaic power plants. It was successfully demonstrated that using real-time data analysis, coupled with advanced forecasting models, enables accurate predictions of electricity generation. To achieve this goal, various algorithms were employed to collect, process, and analyse large volumes of data. Graphical representations were used to visualise the data processing workflow, and forecasting models were monitored and tested. Additionally, the software tools used for Big Data analysis were described and evaluated. Particular attention was paid to the economic aspects of using energy storage systems.

Overall, this study confirmed the economic viability of employing various systems and identified the potential for arbitrage based on electricity price differentials. High accuracy and stability were achieved in forecasting electricity generation from solar photovoltaic power plants. Additionally, the economic benefits of using energy storage systems based on analysing price differences in the electricity market were identified. The results demonstrate the success of the developed methods in forecasting electricity generation from solar photovoltaic power plants. It was established that the analysed forecasting models effectively account for various factors influencing electricity production. Furthermore, the economic viability of using energy storage systems based on the analysis of differences in the buying and selling prices of electricity was confirmed. It was found that such systems can be a crucial component for optimising energy production and increasing the efficiency of solar resource utilisation.

The research was constrained by several factors that influenced its conduct. Notably, there were limitations in the availability of data on solar activity, meteorological parameters, and electricity generation from solar panels, which could impact the accuracy and reliability of the forecasting. The complexity of modelling solar systems and the influence of unpredictable factors should also be considered, as these can lead to a decrease in the accuracy of forecasts. These limitations should be taken into account as the characteristics of solar panels can change over time due to wear and the influence of external factors, which may affect the effectiveness of forecasting in the future.

It is recommended to continuously improve machine learning algorithms using large datasets on solar activity. It is also important to analyse the results to identify opportunities for enhancing the forecasting system. For further enhancement, the utilisation of more detailed data and the implementation of new energy management strategies should be considered, with the aim of optimising production and conserving resources.

ACKNOWLEDGEMENTS

None.

CONFLICT OF INTEREST

None.

REFERENCES

[1] Aman, R., Rizwan, M., & Kumar, A. (2023). Novel and efficient hybrid deep learning approach for solar photovoltaic power forecasting considering meteorological parameters. *Electrical Engineering*. doi: 10.21203/rs.3.rs-3260351/v1.

[2] Arias, M.B., & Bae, S. (2021). Solar photovoltaic power prediction using big data tools. *Sustainability*, 13(24), article number 13685. doi: 10.3390/su132413685.

[3] Bappalige, S. (2014). *An introduction to Apache Hadoop for big data*. Retrieved from https://opensource.com/life/14/8/intro-apache-hadoop-big-data.

[4] Chandel, S.S., Gupta, A., Chandel, R., & Tajjour, S. (2023). Review of deep learning techniques for power generation prediction of industrial solar photovoltaic plants. *Solar Compass*, 8, article number 100061. doi: 10.1016/j.solcom.2023.100061.

[5] de Freitas Viscondi, G., & Alves-Souza, S.N. (2019). A systematic literature review on big data for solar photovoltaic electricity generation forecasting. *Sustainable Energy Technologies and Assessments*, 31, 54-63. doi: 10.1016/j.seta.2018.11.008.

[6] Dhaked, D.K., Dadhich, S., & Birla, D. (2023). Power output forecasting of solar photovoltaic plant using LSTM. *Green Energy and Intelligent Transportation*, 2(5), article number 100113. doi: 10.1016/j.geits.2023.100113.

[7] Gaboitaolelwe, J., Zungeru, A.M., Yahya, A., Lebekwe, C.K., Vinod, D.N., & Salau, A.O. (2023). Machine learning based solar photovoltaic power forecasting: A review and comparison. *IEEE Access*, 11, 40820-40845. doi: 10.1109/ACCESS.2023.3270041.

[8] Gil-Vera, V.D., & Quintero-López, C. (2023). Predictive modeling of photovoltaic solar power generation. *WSEAS Transatcions on Power Systems*, 18, 71-81. doi: 10.37394/232016.2023.18.8.

[9] Hossain, M.M., Pranto, M.A., & Hasan, M.N. (2023). *Design of a 20 MW grid-connected solar photovoltaic power plant for a community in Char Fasson*. Dhaka: Ahsanullah University of Science and Technology. doi: 10.13140/RG.2.2.30335.38560.

[10] Iheanetu, K.J. (2022). Solar photovoltaic power forecasting: A review. *Sustainability*, 14(24), article number 17005. doi: 10.3390/su142417005.

[11] Je, S.-M., Ko, H., & Huh, J.-H. (2021). Accurate demand forecasting: A flexible and balanced electric power production big data virtualization based on photovoltaic power plant. *Energies*, 14(21), article number 6915. doi: 10.3390/en14216915.

[12] Kho, L.C., Ngu, S.S., Annie, J., Sahari, S.K., Kipli, K., & Rulaningtyas, R. (2023). A case study of using long short-term memory (LSTM) algorithm in solar photovoltaic power forecasting. *ASM Science Journal*, 18, 1-8. doi: 10.32802/asmscj.2023.1162.

[13] Kumar, N., & Tripathi, M.M. (2022). Investigation on effect of solar energy generation on electricity price forecasting. *Journal of Intelligent & Fuzzy Systems*, 42(2), 1185-1197. doi: 10.3233/JIFS-189781.

[14] Kyzym, M., Shpilevskyi, V., Zinchenko, V., & Shpilevskyi, O. (2022). Global challenges and prospects for the structural development of Ukraine's electric energy industry. *Business Inform*, 7, 86-98. doi: 10.32983/2222-4459-2022-7-86-98.

[15] Liu, W. (2023). Key technologies for photovoltaic power generation. *Highlights in Science Engineering and Technology*, 43, 74-83. doi: 10.54097/hset.v43i.7407.

[16] Nahar, A., Al Mamun Rudro, R., Al Sohan, F.A., Reya, R.I., & Uddin, M.H. (2024). Predicting photovoltaic power generation by machine learning using time series analysis. *Energy Proceedings*, 43(5), 1-4. doi: 10.46855/energy-proceedings-11024.

[17] Nazari, M.A., El Haj Assad, M., & Salem, M. (2023). Solar energy and power generation. In M. El Haj Assad & M.A. Nazari (Eds.), *Solar collectors and systems* (pp. 15-29). Hauppauge: Nova Science Publishers. doi: 10.52305/EAML0250.

[18] Nie, Y., Li, X., Scott, A., Sun, Y., Venugopal, V., & Brandt, A. (2022). SKIPP'D: A SKy images and photovoltaic power generation dataset for short-term solar forecasting. *Solar Energy*, 255, 171-179. doi: 10.1016/j.solener.2023.03.043.

[19] Official Website of "Market Operator". (n.d.). Retrieved from https://www.oree.com.ua/.

[20] Pallamala, R.K., & Rodrigues, P. (2022). An investigative testing of structured and unstructured data formats in big data application using Apache Spark. *Wireless Personal Communications*, 122(1), 603-620. doi: 10.1007/s11277-021-08915-0.

[21] Pysmenna, U., Sotnyk, I., Kubatko, O., Trypolska, G., & Kurbatova, T. (2021). Market mechanism formation for stimulating the development of energy accumulation systems in Ukraine. *Visnyk of Sumy State University*, 3(3), 31-39. doi: 10.21272/1817-9215.2021.3-4.

[22] Rajnish, Saroha, S., & Kumar Saini, M. (2024). Numerical modeling and performance assessment of machine learning-based solar photovoltaic energy forecasting system. In *Power engineering and intelligent systems* (pp. 391-404). Singapore: Springer. doi: 10.1007/978-981-99-7383-5_29.

[23] Sabri, M., & El Hassouni, M. (2023). Photovoltaic power forecasting with a long short-term memory autoencoder networks. *Soft Computing*, 27(15), 10533-105553. doi: 10.1007/s00500-023-08497-y.
[24] Sarah, K.E., Roland, U., & Okafor, E.N.C. (2020). A review of solar photovoltaic technologies. *International Journal of Engineering Research & Technology*, 9(7), 741-749. doi: 10.17577/IJERTV9IS070244.
[25] Saxena, N. (2019). *Apache Kafka value in Big Data*. Retrieved from https://medium.com/dataseries/big-data-solution-using-apache-kafka-9a9d27cbf51c.
[26] Soumya, C., Deepanraj, B., & Ranjitha, J. (2021). A review on solar photovoltaic systems and its application in electricity generation. *AIP Conference Proceedings*, 2396(1), article number 020011. doi: 10.1063/5.0066291.
[27] Stanytsina, V., Nechaieva, T., Trokhaniak, V., Horskyi, V., & Teslenko, O. (2023). Electricity and heat supply technologies for increasing the energy independence of certain territorial communities. *System Research in Energy*, 4(75), 32-44. doi: 10.15407/srenergy2023.04.032.
[28] Suwarno, & Pinayungan, D. (2024). Solar power forecasting model as a renewable generation source on virtual power plants. *Bulletin of Electrical Engineering and Informatics*, 13(2), 702-712. doi: 10.11591/eei.v13i2.5870.
[29] Tajjour, S., & Chandel, S.S. (2022). Power generation forecasting of a solar photovoltaic power plant by a novel transfer learning technique with small solar radiation and power generation training data sets. *SSRN*. doi: 10.2139/ssrn.4024225.
[30] Torres, J.F., Troncoso, A., Koprinska, I., Wang, Z., & Martínez-Álvarez, F. (2019). Deep learning for big data time series forecasting applied to solar power. In *Proceedings of the international joint conference SOCO'18-CISIS'18-ICEUTE'18* (pp. 123-133). Cham: Springer. doi: 10.1007/978-3-319-94120-2_12.
[31] Vaish, R., & Dwivedi, U.D. (2022). Role of machine learning in forecasting solar and wind power generation. In S.M. Tripathi & A.K. Singh (Eds.), *Energy conversion: Methods, technology and future directions* (pp. 211-241). Hauppauge: Nova Science Publishers. doi: 10.52305/VXCB5652.
[32] Zhang, Y., & Wang, Y. (2022). A hybrid neural network-based intelligent forecasting approach for capacity of photovoltaic electricity generation. *Journal of Circuits, Systems and Computers*, 32(10), article number 2350172. doi: 10.1142/S0218126623501724.
[33] Zhang, Z., Wang, J., Xia, Y., Wei, D., & Niu, Y. (2023). Solar-mixer: An efficient end-to-end model for long-sequence photovoltaic power generation time series forecasting. *IEEE Transactions on Sustainable Energy*, 14(4), 1979-1991. doi: 10.1109/TSTE.2023.3268100.

Технології Big Data в процесі прогнозування електрогенерації від сонячних фотоелектростанцій

Олександр Столяров

Аспірант
Національний університет харчових технологій
01601, вул. Володимирська, 68, м. Київ, Україна
https://orcid.org/0009-0001-7499-4672

Анотація. Мета дослідження полягала в розробці методів використання технологій Big Data для прогнозування електрогенерації від сонячних фотоелектростанцій, що важливо для оптимізації виробництва енергії та підвищення ефективності використання сонячних ресурсів. Було використано метод аналізу економічної вигідності застосування систем управління зберіганням електроенергії та порівняльний аналіз цін на купівлю і продаж електроенергії на ринку, а також експеримент, що включав використання програмних засобів та алгоритмів для обробки, аналізу та моделювання великих обсягів даних. У результаті дослідження було розроблено методики, які охоплюють збір та аналіз даних, візуалізацію інформації, вибір та навчання моделей прогнозування на основі наявних даних, а також проведення моніторингу та тестування їх ефективності. Було побудовано графічні схеми, що вказують на етапи обробки та аналізу даних, процес прогнозування електрогенерації на різні періоди часу, а також процес навчання моделі на основі даних, моніторинг та тестування моделі. Крім того, створено графік, що показує оцінки типовості та розмаху значень, і графік, що відображає зміну цін на електроенергію протягом доби. Додатково було описано технологічні засоби для використання Big Data, розраховано вартість електроенергії та проведено оцінку економічної привабливості застосування систем управління зберіганням електроенергії. У результаті дослідження було встановлено показник імовірного доходу від арбітражу за різницею цін, а також економічні параметри доцільності використання управління зберіганням електроенергії на основі аналізу відмінностей у вартості купівлі та продажу електроенергії. Отримані результати можуть бути корисними для енергетичних компаній та організацій, що працюють у сфері виробництва електроенергії з сонячних фотоелектростанцій, дозволяючи їм оптимізувати виробництво енергії та підвищити ефективність використання сонячних ресурсів

Ключові слова: інформаційно-аналітичні засоби; енергетичний ринок; обробка об'ємних матеріалів; оптимізація енерговиробництва; моделі передбачення