Implementation of Big Data Analytics for Simulating, Predicting & Optimizing the Solar Energy Production

Jabar Yousif

International Applied Computing & Applications publisher

October 30, 2023

Abstract

This paper investigated and reviewed the current big data methods and tools in solar energy production. It discusses the comprehensive two-stage design and evaluation for examining the optimal structure for renewable energy systems. In the design stage, technical and economic aspects are discussed based on a robust analysis of all input/output variables for determining the highest performance. Next, assess and evaluate the effectiveness of each method under different circumstances conditions. Then convert each qualitative indicator into a quantitative measure using extensive data analysis methods to determine the overall performance of the various qualitative variables. The paper also provides an in-depth analysis of the mathematical techniques used in measuring the efficiency of the renewable energy production system and discussing future axes of work in the field of specific energy.
Implementation of Big Data Analytics for Simulating, Predicting & Optimizing the Solar Energy Production

Jabar H. Yousif1,*
1 Faculty of Computing & Information Technology, University of Sohar, Oman
* Corresponding author: Jabar H. Yousif1, jyousif@su.edu.om

Abstract

The notable developments in renewable energy facilities and resources help reduce the cost of production and increase production capacity. Therefore, developers in renewable energy evaluate the overall performance of the various equipment, methods, and structure and then determine the optimal variables for the design of energy production systems. Variables include equipment characteristics and quality, geographical location, and climatic variables such as solar irradiance, temperature, humidity, dust, etc. This paper investigated and reviewed the current big data methods and tools in solar energy production. It discusses the comprehensive two-stage design and evaluation for examining the optimal structure for renewable energy systems. In the design stage, technical and economic aspects are discussed based on a robust analysis of all input/output variables for determining the highest performance. Next, assess and evaluate the effectiveness of each method under different circumstances conditions. Then convert each qualitative indicator into a quantitative measure using extensive data analysis methods to determine the overall performance of the various qualitative variables. The paper also provides an in-depth analysis of the mathematical techniques used in measuring the efficiency of the renewable energy production system and discussing future axes of work in the field of specific energy.

Keywords: Big Data; machine learning; solar energy; ANN; model optimization;
1. Introduction

Student Big Data has become a significant interest research area for both academic and business organizations. It describes the use of large, hard-to-manage volumes of data. It comprises the analysis of the three V’s (Volume, Velocity, and Variety) for insights that lead to better decisions and strategic business moves (Yousif & Saini, 2020). On the other hand, Solar energy is an inherent energy source that uses light and heat to apply photovoltaic and thermal energy systems. Big data techniques help promote solar/thermal energy generation for efficient and reliable usage of renewable sources. Energy plays a vital role in the economy and environment (Singh et al., 2021). Therefore, the evaluation of energy production, capacity factor, and cost of power is a potential research focus. Figure 1 shows the growth in renewable electricity generation globally.

![Figure 1. growth in renewable electricity generation globally (IRENA, 2020)](image)

In It shows that the renewable electricity production in 2018 reaches 376 TWh, higher than in 2017, increasing 6.1%. The generation growth of Solar and wind grow strongly by 28% and 11% to become the third-largest source of renewable electricity generation in 2018 (IRENA, 2020).

The transformation of renewable energy technology into the global economy will boost technological innovations to enhance power production and reduce prices. This requires the application of new algorithms and methods to increase the performance of solar energy production (khan I, 2021). Also, to find more efficient systems to store and
transfer energy generated from the grid-connected system to sources of consumption quickly and cheaply. Applying big data analysis based on machine learning and artificial intelligence will efficiently help analyze problems and find practical solutions (He et al., 20202). This work aims to present modern technologies and explore practical solutions that help increase solar and thermal energy production performance to guide new researchers.

2. Challenges in Big Data analysis

Specialists working in Big Data analysis specify many challenges in exploring large data sets when extracting information to build relationships between different data types (Ren et al., 2021). Difficulties in Big Data analysis include data capturing, warehouse, exploration, sharing, accounting, management, and visualization. Big Data management aims to ensure the reliability of the data and provide simple ways to access and control the stored data securely. One of the most crucial challenges in dealing with Big Data is collecting data from different sources and integrating them into a standard format (Devaraj et al., 2021). Also, finding the efficient model for storing the generated power to reduce storage size. Another significant challenge is the efficient management of Big Data to accelerate the extraction of information and facilitate the transfer of information at a low cost. The demand for renewable energy is increasing; therefore, it is requisite to recognize the factors that prevent its scalability. Today renewable energy achieves several benefits, but industries will fail without efficient forecasting and intelligent scheduling of the resources (Fan Z, 2021). Hence, examining appropriate data analytics and machine learning techniques can help simulate and predict weather conditions efficiently and improve productivity. Incorporating these technologies will eventually lead to the development of the renewable energy sector, which is an important factor in the economy of every country.

3. Big Data Techniques for Renewable Energy

This work aims to investigate and review the current big data methods and tools in solar energy production. It should focus on the comprehensive two-stage design and evaluation for examining the optimal variables and structure for renewable energy systems (Akbari-Dibavar et al., 2021). In the design stage, technical and economic aspects are discussed based on a robust analysis of all input/output variables for determining the highest performance. Next, assess and evaluate the effectiveness of each method under different circumstances conditions for optimizing the renewable energy systems.
The photovoltaic/thermal (PVT) is a hybrid system that combines solar thermal and electrical power simultaneously (Gül & Akyüz, 2021). The performance of PV systems is influenced significantly by various weather conditions, such as (Solar Irradiance (SR), Temperature (TM), Relative Humidity (RH), Wind Speed (WS), Wind Direction (WD), Sunshine Hours (SH), Evaporation (EVA), Pressure (PRE), Cloud (CL), Cell Temperature (TC), Fluid Temperature (FT), etc.

Several fields of research could be investigated that includes:

- Power Data production and forecasting.
- Resource management.
- Storage management.
- Predicting and optimization of future figures.

Big Data are mostly based on machine learning techniques can include Artificial Neural Network (ANN), Support Vector Machine (SVM), Linear/Logistic regression (LR) (Yousif & Fekihal, 2012).

These approaches help for examining and enhancing renewable energy production could include the following techniques:

- **Supervised Learning**, for example (Multilayer perceptron (MLP), Generalized Regression Neural Network (GRNN), Learning Vector Quantization (LVQ)).
- **Unsupervised Learning**, for example (Recurrent Neural Network (RNN), Self-Organizing Feature Map (SOFM), Principal Component Analysis (PCA)).
- **Semi-supervised Learning** (semi-supervised Support Vector Machines (S3VMs)).
- Reinforcement Learning is based on Markov decision process (MDP)
- **Deep Learning** such as convolutional neural networks (CNN), Long Short-Term Memory Networks (LSTM), and Generative Adversarial Networks (GANs).

Table 1 presents and explores several studies that deploys Artificial Neural Network (ANN) for estimating, simulating, predicting, and optimizing the solar energy production.
Table 1. Review of studies that deploy Artificial Neural Network (ANN) for solar energy production.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Location</th>
<th>ANN Method</th>
<th>Estimation/ Simulation / Prediction / Optimization</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Jabar &amp; Hussein, 2021)</td>
<td>Oman</td>
<td>MLP, SOFM, SVM</td>
<td>Prediction &amp; Optimization</td>
</tr>
<tr>
<td>(Zayed et al., 2021)</td>
<td>USA, Arizona</td>
<td>RVFL, PSO, SO, WOA</td>
<td>Prediction &amp; Optimization</td>
</tr>
<tr>
<td>(Al-Waeli, et al., 2020)</td>
<td>Malaysia</td>
<td>MLP, SOFM</td>
<td>Estimation &amp; Prediction</td>
</tr>
<tr>
<td>(Li, Y et al., 2020)</td>
<td>Hong Kong</td>
<td>FFMLP- LM</td>
<td>Simulation &amp; Prediction</td>
</tr>
<tr>
<td>(Al-Waeli, et al., 2019a)</td>
<td>Malaysia</td>
<td>MLP, SOFM</td>
<td>Simulation &amp; Prediction</td>
</tr>
<tr>
<td>(Kazem et al., 2019a)</td>
<td>Oman</td>
<td>TLRN, FRNN</td>
<td>Simulation &amp; Prediction</td>
</tr>
<tr>
<td>(Alnaqi et al., 2019)</td>
<td>Malaysia</td>
<td>MLP</td>
<td>Prediction &amp; Optimization</td>
</tr>
<tr>
<td>(Li, Y et al., 2020)</td>
<td>Hong Kong</td>
<td>MLP, SOFM</td>
<td>Prediction &amp; Optimization</td>
</tr>
<tr>
<td>(Al-Waeli, et al., 2019b)</td>
<td>Oman</td>
<td>PMM</td>
<td>Prediction</td>
</tr>
<tr>
<td>(Yousif et al., 2019a)</td>
<td>Australia / Queensld</td>
<td>MLP, SVR, GPML, GP</td>
<td>Prediction</td>
</tr>
<tr>
<td>(Kazem &amp; Yousif, 2017)</td>
<td>Oman</td>
<td>Comparative study</td>
<td>Simulation</td>
</tr>
<tr>
<td>(Kazem et al., 2017)</td>
<td>Iran</td>
<td>ML</td>
<td>Simulation &amp; Prediction</td>
</tr>
<tr>
<td>(Yousif et al., 2017)</td>
<td>Oman</td>
<td>ML</td>
<td>Prediction &amp; Optimization</td>
</tr>
<tr>
<td>(Bassam et al., 2017)</td>
<td>Mexico</td>
<td>MLP, SOFM, SVM, PMM</td>
<td>Prediction &amp; Optimization</td>
</tr>
<tr>
<td>(Kalani et al., 2017)</td>
<td>Malaysia</td>
<td>MLP, PBF-NN, ANFIS</td>
<td>Estimation &amp; Prediction</td>
</tr>
<tr>
<td>(Yousif et al., 2016)</td>
<td>Oman</td>
<td>ML</td>
<td>Prediction</td>
</tr>
<tr>
<td>(Shamisi et al., 2014)</td>
<td>UAE Al-Ain</td>
<td>MLP / RBF-NN</td>
<td>Estimation</td>
</tr>
<tr>
<td>(Premalatha &amp; Arasu, 2012)</td>
<td>USA Maries County</td>
<td>MLP / RBF-NN</td>
<td>Prediction</td>
</tr>
<tr>
<td>(Alzahrani et al., 2014)</td>
<td>UAE Al-Ain</td>
<td>MLP / RBF-NN</td>
<td>Prediction</td>
</tr>
<tr>
<td>(Alzahrani et al., 2014)</td>
<td>UAE Al-Ain</td>
<td>MLP / RBF-NN</td>
<td>Prediction</td>
</tr>
<tr>
<td>(Assi et al., 2013)</td>
<td>Algeri/ Bechar</td>
<td>MLP -LM</td>
<td>Estimation</td>
</tr>
<tr>
<td>(Hasni et al., 2012)</td>
<td>India / Tamil Nadu</td>
<td>FFMLP</td>
<td>Estimation</td>
</tr>
<tr>
<td>(Qin et al., 2011)</td>
<td>China/ Tibetan Plateau</td>
<td>MLP-LM</td>
<td>Estimation</td>
</tr>
</tbody>
</table>

Random Vector Functional Link (RVFL); Particle Swarm Optimization (PSO); Spherical Search Optimization (SSO); Whale Optimization Algorithm (WOA); Levenberg-Marquardt algorithm (LM); Polynomial Mathematical Model (PMM); Gaussian Process Machine Learning (GPML); Genetic Programming (GP); Single layer Feed-forward network (SLFN); Feedforward Multilayer Perceptron (FFMLP); Recurrent Neural Network (RNN); Full Recurrent Neural Network (FRNN); Nonlinear Autoregressive Network (NARX); Support Vector Machine (SVM); Multilayer Perceptron (MLP); Self-Organizing Feature Map (SOFM); Time Lagged Recurrent Network (TLRN); Random Forest (RF); Extra Trees (ET); Decision Trees (DT); Polynomial basis function (PBF-NN); Adaptive Neuro Fuzzy Inference System (ANFIS); Generalized Feed-Forward Neural Networks (GFF-NN); Generalized Regression Neural Network (GRNN); Radial Basis Function Neural Network (RBF-NN); Feed-Forward Neural Networks (FF-NN);
4. Conclusion

Several essential technologies are helping to create an efficient, productive, and sustainable future globally. Solar energy and big data are two primary methods of these technologies. They will change how energy is purchased, marketed, distributed, and used responsibly. Advanced technologies such as big data and analytics have an enormous effect on every aspect of energy production and today's modern world. The aggregate of big data and solar energy has also supported the creation of other brand-new energy distribution models. Solar panels are affordable enough for some companies and customers to buy outright. In other cases, having access to the correct data can help identify issues where additional agreements are more appropriate for the parties involved. In these ways and more, big data and solar power make an ideal match that will continue paying significant interest. Solar panels let us tap into an almost inexhaustible source of energy. Big data analysis helps us manage and enhance power production and its cost-effectively.

The following are several recommendations:

a) One of the most significant problems of managing electric power generation is the combination of different types of renewable energy (wind, solar cells, thermal units, etc.) and knowing the exact amount of production during different year periods. The increase and decrease in production will generate a storage defect that may exceed the current storage capacity.

   o The solution is to find efficient mathematical models that can predict wind and solar energy production, which helps reduce uncertainty about the changing renewable energy. Predictive models will help network operators commit extreme events more efficiently, reducing system balancing costs.

b) Most of the energy production network systems differ depending on the type of tools used, location, and various weather conditions such as (Solar Irradiance (SR), Temperature (TM), Relative Humidity (RH), Wind Speed (WS), Wind Direction (WD), Sunshine Hours (SH), Evaporation (EVA), Pressure (PRE), Cloud (CL), Cell Temperature (TC), Fluid Temperature (FT), etc.

   o So, there is a significant need to find optimal solutions to address the integration between these different variables, which is a major branch of research and development.

Acknowledgment

The research leading to these results has received no Research Project Grant Funding.
References


or photovoltaic electricity production in [44].


