Comparative Analysis of Machine Learning Models for Daytime Power Generation Prediction

Nima Khodadadi[1,*], **Benyamin Abdollahzadeh**[2]

[1]Department of Civil and Architectural Engineering, University of Miami, Coral Gables, FL, USA
[2]Department of Mathematics, Faculty of Science, University of Hradec Králové, 500 03, Hradec Králové, Czech Republic

Emails: nima.khodadadi@miami.edu; benyamin.abdolahzade@gmail.com

Abstract

This paper proposes to evaluate how different machine learning techniques can be used to predict daytime power generation based on the "Daily Power Generation Data" data set. As a result of six models, which contain Random Forest Regressor, Decision Tree Regressor, Nearest Neighbors, Linear Regression, MLP Regressor, and SVR, a clear understanding has been accomplished by assessing the performance using multiple metrics. First, the Random Forest Regressor turned out to be the best in terms of the Mean Squared Error (MSE) of 3.57E-06, which was the lowest among the three ML models. The introduction of the paper highlights the role of precise planning of the power market and the consecutive sections describing the topic mathematically. The table below, with a total list of performance issues, explains why the Random Forest Regressor is the superior full-proof model using the lowest MSE, highest explained variance, and great resistance to outlying samples. The paper thus gave various useful approval criteria that we can largely choose the best model out of them because the Random Forest Regressor was able to get the highest performance metrics.

Keywords: Power Generation; Daily Power Generation; Machine Learning; Random Forest Regressor

1. Introduction

The escalating energy demand and environmental concerns are driving a significant transformation in the global energy landscape, necessitating a shift towards innovative and sustainable power generation solutions [1]. Power generation optimization, amidst technological advancements and resource constraints, is paramount in this endeavor [2]. Fossil fuels have long been the backbone of the global energy system, but their limited availability, environmental impacts, and geopolitical complexities underscore the urgency of transitioning to cleaner energy sources. This research focuses on optimizing power generation to balance rising energy needs with environmental preservation [3]. The guiding principle of this paper is to provide comprehensive insights into power generation phenomena and advocate for sustainable practices [4]. By informing policy decisions, facilitating technological advancements, and engaging diverse stakeholders, including scientists, engineers, policymakers, and environmentalists, we aim to contribute to transforming the global energy landscape and achieving a sustainable future. The study's research objectives serve as road markers, directing attention toward innovative and sustainable solutions to power-generating issues. We delve into three key research questions: Firstly, we explore the principal challenges and limitations associated with current power generation methods, aiming to identify barriers to sustainability and effectiveness. Secondly, we investigate cutting-edge technologies that could enhance electricity generation efficiency, driving our technological research forward. Finally, we examine strategies to minimize the environmental impact of power generation and promote sustainability, guided by our commitment to environmental stewardship and the preservation of ecological equilibrium [5,6]. As we embark on this intellectual voyage toward innovation, our collective endeavor is to seek creative, sustainable, and inclusive approaches to energy generation. This research paper will focus on current power-generating methods, explore technological

DOI: https://doi.org/10.54216/JAIM.080201
Received: March 24, 2024 Revised: May 25, 2024 Accepted: November 01 2024

advancements, and advocate for environmentally friendly practices [7]. We invite the reader to join us on this journey, which promises to be both enlightening and impactful, as we contribute to the development of innovative power generation sectors and work towards sustaining the Earth's oasis state for future generations.

2. Literature Review

Power generation is vital for modern societies, supporting industries, infrastructure, and daily life. With rising energy demand worldwide, ensuring efficient, sustainable, and reliable power generation is crucial. This review covers traditional and renewable energy sources, like solar and wind power, and investigates smart grid technologies to improve system efficiency and resilience..

Within the sphere of solar energy, achieving optimal solar radiation levels is paramount for the effective design and operation of solar energy systems. Research endeavors such as [8] delve into innovative modeling techniques, like the hybrid PSO-ELM model, which significantly enhances the accuracy of daily solar radiation estimation, particularly in regions where comprehensive onsite data is lacking. Moreover, [9] focuses on advancing power prediction methodologies for photovoltaic plants through the application of sophisticated machine learning algorithms. Concurrently, [10] introduces a novel hybrid model, the SDA-GA-ELM, tailored for precise hourly estimation of PV power production, showcasing the relentless pursuit of accuracy in solar energy forecasting.

Meanwhile, the transition to smart grids presents a host of complex challenges, including demand-side management, cybersecurity, and the optimization of grid infrastructure. The evaluation conducted in [11] provides invaluable insights into the effectiveness of machine learning algorithms within smart grid technologies, highlighting the need for further research to optimize data processing and enhance network management capabilities. Additionally, [12] introduces LSTM-PC, an advanced model aimed at accurately forecasting PV plant energy production, addressing inherent challenges such as weather-dependent predictability and fluctuating energy demand.

Furthermore, [13] undertakes the crucial task of predicting greenhouse gas emissions from Turkey's electricity sector, employing state-of-the-art deep learning techniques alongside traditional methodologies. This research underscores the urgent need to curb emissions and transition to cleaner energy sources, given the alarming rise in fossil fuel consumption and its detrimental environmental impact.

Navigating the complexities of power system operation requires innovative solutions to effectively manage uncertainty. [14] presents a pioneering method for generating statistically valid scenarios from probabilistic projections, facilitating improved decision-making in power system management. Meanwhile, [15] proposes a data-driven framework for accurate monthly renewable energy forecasts, leveraging advanced techniques like STL and LSTM to enhance prediction performance and ensure grid stability.

Lastly, ensuring grid observability remains paramount for efficient electricity industry operation, as discussed in [16]. By employing state estimation approaches and machine learning techniques, this study showcases the potential for optimizing grid operation under varying conditions, paving the way for real-time status prediction and improved system resilience in the face of evolving energy demands and environmental challenges.

In summary, this literature review underscores the imperative of transitioning to clean and sustainable technologies within the power generation sector. Collaboration among stakeholders, along with active interdisciplinary research, is essential for navigating the evolving energy landscape and fostering the development of a resilient, resourceful, and environmentally friendly power generation system that meets the needs of society while minimizing environmental impact.

3. Dataset

3.1 Dataset Description

The "Daily Power Generation Data" dataset [17], is notable for its comprehensive information on power generation activities, offering insights sorted by geographical stations. Covering a significant period from September 1, 2017, to January 19, 2023, this dataset provides a detailed account of electricity production history, enabling a deep understanding of power generation dynamics over time. However, it's essential to highlight significant data gaps on specific dates, including October 2, 2017, November 19, 2017, November 26, 2017, April 3, 2018, and April 4, 2018, as well as the prolonged gap from March 19, 2020, to May 31, 2020. These gaps obscure critical insights into power generation during these periods, presenting challenges for comprehensive analysis and decision-making processes.

Furthermore, the dataset is available in two distinct file formats, necessitating precise processing and analysis to ensure consistency and accuracy across both formats. Overcoming these challenges is paramount for reliable data preparation, particularly for predictive analysis and informed decision-making in energy management. Exploring strategic approaches to address data gaps and associate power generation processes with level estimations will be essential to enhance the dataset's trustworthiness and usefulness for future analysis. By confronting these

DOI: https://doi.org/10.54216/JAIM.080201
Received: March 24, 2024 Revised: May 25, 2024 Accepted: November 01 2024

challenges head-on, researchers can unlock valuable insights and facilitate informed decision-making in the realm of power generation dynamics.

3.2 Dataset Pre-processing Steps

Before commencing the analysis of daily power generation data, a meticulous preparatory phase was undertaken to ensure the dataset's integrity and consistency. The subsequent processes delineate the preprocessing [18] methods employed:

- Handling Missing Data: The identification and resolution of dates with missing data involved a thorough review to understand the underlying reasons. Imputation techniques were then applied to fill these gaps, offering researchers flexibility in choosing methods tailored to the data's nature and analytical objectives. Techniques such as mean imputation, interpolation, or predictive modeling were considered based on the context of the missing data and the research objectives.
- File Format Harmonization: Given the disparity in file formats across the dataset, a harmonization method was implemented to ensure accurate representation of the dataset. This involved converting and aligning the storage format to establish a well-ordered structure conducive to future analyses. Additionally, efforts were made to reconcile any discrepancies in data encoding or formatting conventions between the two files to ensure seamless integration and consistency.
- Standardization of Units: A systematic standardization procedure was executed on the dataset's units of measurement to enhance clarity and facilitate comparisons across variables. Specifically, a methodical transformation from mega units (MU) to megawatts (MW) was conducted, promoting a cohesive and uniform analysis.

This conversion facilitated a more intuitive understanding of power generation metrics and eliminated potential confusion arising from disparate unit representations. Ensuring dataset integrity was crucial during preparation, with subsequent verifications conducted to address any inconsistencies. Through cross-variable consistency tests, researchers scrutinized relationships between variables, rectifying any discrepancies detected to ensure coherence and accuracy.

3.3 Descriptive statistics

Descriptive statistics were instrumental in comprehensively understanding the traits and patterns within the "Daily Power Generation Data," offering a quantitative overview of key characteristics and scrutinizing the dataset's major tendencies and variation. Spanning from September 1, 2017, to January 19, 2023, the dataset captures a rich tapestry of temporal dynamics in daily power generation. Across diverse power stations, each contributing uniquely to the overall landscape, computed descriptive statistics shed light on individual performance metrics. The dynamic heat map in Figure 1 illustrates the daily power generating statistics. By utilizing color gradients, it reveals the variations in power levels over time and across different stations, providing valuable information on geographical and temporal trends.

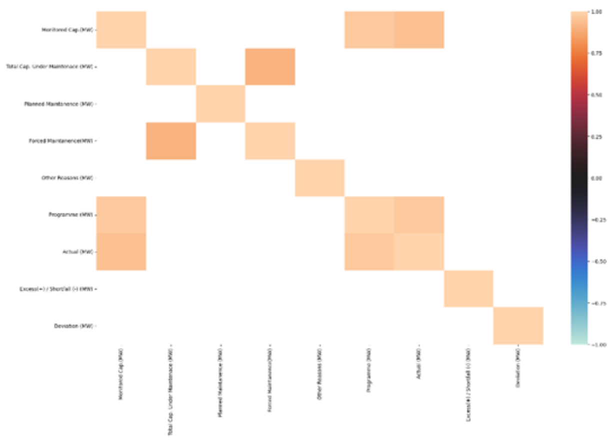

Figure 1. Heatmap of the Dataset

DOI: https://doi.org/10.54216/JAIM.080201
Received: March 24, 2024 Revised: May 25, 2024 Accepted: November 01 2024

Figure 2's histogram vividly displays variations between projected and actual power generation, using bars to represent real data.

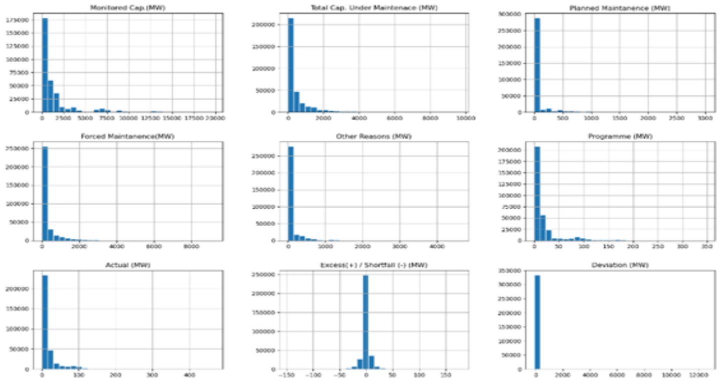

Figure 2. Histograms of the Dataset

4. Results

This area is devoted to the outcomes of applying several machine learning cases to the "Daily Power Output Data" dataset. The model was used in the back-end system to forecast power generation from different power stations, using features such as Expected Power Generation, Actual Power Generation, Deviation, and a few others. The objective of this study is to test the models and select the one that provides better power generation prediction than the other.

The following machine learning models were evaluated:

1. RandomForestRegressor
2. DecisionTreeRegressor
3. NearestNeighbors
4. LinearRegression
5. MLPRegressor
6. SVMRegressor

For each model, we developed a set of indicators to offer a comprehensive insight into the model's accuracy in forecasting power production. The metrics that we will consider are The Mean Squared Error (MSE), the Root Mean Squared Error (RMSE), the Mean Absolute Error (MAE), the Explained Variance Score (EVS), the Max Error, the Median Absolute Error (MedAE), the Mean Absolute Percentage Error (MAPE), the R-squared (R2), the Modified Tchebyche.

The commingling of the forthcoming segments will hone in on the exact results concerning each design, thereby supplying a wholesome grasp of their pros and cons in capturing the specialized details of the electricity-producing dataset. This analysis will provide the required basis for making more informed decisions about the most suitable model choice for forecasting power generation.

4.1 Machine Learning Models

In this sub-unit, we talk in detail about the specific machine learning models or the ML models that are utilized in analyzing the "Daily Power Generation Data" dataset. Participants are provided with a set of the best-suited tools for predictive analytics that have different design features and mathematical foundations, leading to a wide range of predictive analytics models. The choice of these models is the very purpose of representing the highly interrelated nature of the data, which contributes to the overall picture of the power generation processes [19].

Random Forest Regressor:

Explanation: The Random Forest Regressor is an ensemble learning method that uses a collection of decision trees formed during training. This method works by combining the trees' forecasts. The forest delivers a robust and accurate model. Every tree is trained on a different slice of the data, which lowers the overall diversity of datasets, thus reducing the associated risk of overfitting.

DOI: https://doi.org/10.54216/JAIM.080201
Received: March 24, 2024 Revised: May 25, 2024 Accepted: November 01 2024

Mathematical Foundation:

Let h(x, θ_i) represent the prediction of the i-th tree in the forest. The final prediction of the Random Forest is given by averaging these individual predictions:

$$H(x) = \frac{1}{N}\sum_{i=1}^{N} h(x, \theta_i)$$

Decision Tree Regressor:

Explanation: A Decision Tree Regressor evenly splits its dataset into subsets, the most important feature at each node. A simple averaging of the relative target variables at the leaves of the decision tree achieves the final output. A decision tree is natural and complex enough to represent relationships of the data that are not linear.

Mathematical Foundation: Let R_m represent the region represented by the m-th leaf node, and $y^{-}{}_{m}$be the mean of the target variable in R_m. The prediction is given by:

$$H(x) = \sum_{m=1}^{M} y^{-}{}_{m} I(x \in R_m)$$

Nearest Neighbors:

Explanation: The k-nearest Neighbor algorithm predicts a desired variable based on the average of the k-nearest neighbors in the feature space. It hypothesizes that the same series of inputs results in the same output values, so it will be weighted.

Mathematical Foundation: For a given point x, the prediction is the average of the k nearest neighbors' target values:

$$H(x) = \frac{1}{k}\sum_{i=1}^{k} y_i$$

Linear Regression:

Explanation: Linear regression assumes a straight-line association between the target variable and the independent variables through linear function. However, the linear relationship between the independent variables and the outcome variable is also assumed.

Mathematical Foundation: Given features X, weights θ, and bias b, the prediction is given by:

$$H(X) = X \cdot \theta + b$$

MLP Regressor (Multi-Layer Perceptron):

Explanation: MLP Regressor is a class of multilayered neural networks based on the structure of two-layer networks with input, intermediary, and output layers. The network has the power to unveil the hidden links in the data by propagating the activating neuron signals throughout it.

Mathematical Foundation: For a given input X, the output of the network is calculated through the activation of neurons in each layer:

$$H(X) = \sigma(W^{(2)} \cdot \sigma(W^{(1)} \cdot X + b^{(1)}) + b^{(2)})$$

SVR (Support Vector Regressor):

Explanation: SVR utilizes support vectors to determine a nonlinear hyperplane with the fewest total deviations from the given data for the best projections in an artificially created space. It provides a useful function of recording non-respective relations between data.

Mathematical Foundation: Given input features X, target values y, a set of support vectors, and a kernel function K, the prediction is given by:

$$H(X) = \sum_{i=1}^{n} \alpha_i y_i K(X, X_i) + b$$

where $\alpha_i$are support vector weights, y_i are target values, $X_i$are support vectors, and b is a bias term.

DOI: https://doi.org/10.54216/JAIM.080201
Received: March 24, 2024 Revised: May 25, 2024 Accepted: November 01 2024

Delving into the mathematical details governing these models' models lends a good insight into how they work, making the subsequent interpretation of the results more efficient.

4.2 Performance Metrics

The performance metrics employed for assessing machine learning models before using the "Daily Power Generation Data" dataset have been designed for a particular reason. The metrics are used to assess model predictive performance and define different aspects of performance.

Mean Squared Error (MSE): Explanation: MSE calculates the squared deviation of actual values from predicted values and their average. It scores larger deviations harder than small errors, which, in turn, provides a measure of overall model accuracy.

Interpretation: The formation of the MSE lower values indicates an optimal performance of the model better that builds upon small prediction errors.

Root Mean Squared Error (RMSE): Explanation: RMSE is an algebraic formula of MSE, which makes it more comprehensible as RMSE tells errors in the same unit as the target variable.

Interpretation: Like MSE, smaller RMSE values primarily point to higher accuracy of the model.

Mean Absolute Error (MAE): Explanation: MAE presents the average of the deviation between the real value and the predicted value. It is a very rigorous test of the accuracy of the model's prediction capability.

Interpretation: Units of lower MAE are believed to have higher accuracy since they more suitably ignore outliers.

Explained Variance Score (EVS): Explanation: EVS indicates the model accuracy by the ratio of residual variance among the explained variance of the target variable. A score of 1 means the perfect prediction, and a margin below indicates less accurate predictions.

Interpretation: EVS observations with a higher value display better concordance, while a 1 score stands for the perfect fit, reflecting the actual estimation.

Max Error: Explanation: The Max Error expresses the largest possible difference in absolute values of the predicted and actual values. It determines the largest error between the actual observation(s) and the forecast(s) produced by machine learning.

Interpretation: Smaller Max Error relevance's mean less severe prediction disparate error indices.

Median Absolute Error (MedAE): Explanation: MedAE stands for the Median absolute difference between the expected and resulted outcomes. It introduces the means to assess prediction erroneousness.

Interpretation: With regard to AraE data, smaller values indicate more accuracy, especially in datasets with outliers.

MAPE (Mean Absolute Percentage Error): Explanation: MAPE gives the common average percentage error that occurred between the actual and the predicted values, which enables us to estimate the size of errors.

Interpretation: Larger MAPEs indicate worse accuracy and the advantage of indicating errors as a percentage of the errors.

R-squared (R2): Explanation: The R2 explains the model in terms of the percentage of variance in the target variables, which the model explains. A ratio of 1 means that the model perfectly fits the data, while 0 means that there is no connection modeled between the predictors and the dependent variable.

Interpretation: Stronger R2 values reflect a better model fit when 1 is ideal, with perfect variance explained.

Modified Tchebycheff Distance (MTD): Explanation: MTD compares the similarity between two probability distributions, giving the measure of how close and productive the models are that are aiming at predicting the true distribution of the target variable.

Interpretation: Lower MTD (mean target deviation) represents better tuning between forecasts and the real phenomenon.

Relative root mean squared: Explanation: RMMSE restructures RMSE to take the mean of the observed values, facilitating the comparison of a range of datasets. Interpretation: Lower RRMSE values are proposed that model performance is much higher while the dataset is large enough.

Willmott Index: Explanation: The Willmott Index employs preferred risk, negativity, and correlation to deduce whether the model functioned properly. It rises from -1 to 1, with 1 standing for excellent harmony.

DOI: https://doi.org/10.54216/JAIM.080201
Received: March 24, 2024 Revised: May 25, 2024 Accepted: November 01 2024

Interpretation: Values closer to 1 mean high accuracy of the model if the weighted average comprising both associations and bias is not high enough.

Mean Bias Error (MBE): Explanation: MBE figures out the average difference between the observed and estimated values, which is useful for identifying systematic errors. Interpretation: MBE values close to 0 inform us that a machine learning model is highly unbiased and its predictions are free of any biases.

Standard Deviation (SD): Explanation: SD indicates the level of the split between the predicted values by indicating the amount of variation or dispersion in the values set. It is able to give us reliable information and a route of consistency. Interpretation: Shorter SDs imply smaller spreads around the mean.

Knowing what the performance metric measures and how to interpret it will unite all the ML models' contributions to predicting power generation based on the given dataset.

4.3 Regression Results

Here, we go further and peek into the results of our machine-learning model experiments using a variety of measures. The main goal is to conduct a thorough examination of the models' abilities to accurately predict power generation issues within a given set of "Daily Power Generation Data." Following the information conveyed in the above result, it is evident that a series of different metrics are being used to highlight different aspects of predictive performance individually. The Table 1 below showcases the performance metrics for each evaluated machine-learning model.

Table 1: Regression Models Results

Models	mse	rmse	mae	evs	max_error	MedAE	MAPE	R2	mtd	RRMSE	Willmott	MBE	SD
RandomForestRegressor	3.57E-06	0.00189	0.000714	0.780449	0.14408	0.000348	4.3E+09	0.780449	3.57E-06	0.008007	0.000346	-9.9E-07	0.0189
DecisionTreeRegressor	6.16E-06	0.002482	0.001056	0.621495	0.182195	0.000552	2.28E+10	0.621422	6.16E-06	0.010515	0.989277	3.45E-05	0.002481
NearestNeighbors	8.55E-06	0.002924	0.000408	0.475002	0.176012	2.82E-05	1.16E+11	0.474369	8.55E-06	0.01239	0.960413	0.000102	0.002922
LinearRegression	1.33E-05	0.003651	0.001833	0.181297	0.226556	0.000782	1.63E+12	0.18041	1.33E-05	0.015471	0.030117	-0.0012	0.003649
MLPRegressor	1.46E-05	0.003822	0.002061	0.110707	0.2254	0.000895	1.68E+12	0.101939	1.46E-05	0.016195	0.089541	-0.0038	0.003803
SVR	0.008185	0.090469	0.090336	-0.53908	0.133703	0.091572	2.5E+13	-502.15	0.008185	0.383326	0.02338	0.090331	0.005004

Discussion of Key Findings

1. Random Forest Regressor: Laboratory experimental results revealed the lowest MSE, RMSE, and MAE, consequently showing highly precise predictions.

High EVS and R2 values indicate great modeling ability for the data.

Demonstrates a less maximum absolute error and medAE, suggesting good removal of outliers.

2. Decision Tree Regressor: It allows for the ability to compete in order to have performance metrics across measures.

High Willmott Index score and R2 values that are high suggest a good fit for the case data.

DOI: https://doi.org/10.54216/JAIM.080201
Received: March 24, 2024 Revised: May 25, 2024 Accepted: November 01 2024

Only 0.4 and 2.15 units of MAE and Max Error are higher than that of Random Forest, respectively.

3. Nearest Neighbors: Less MAE and MedAE are evident, suggesting high accuracy and precision when predicting real energy generation values.

Diagnostic R2 and EVS are significantly less than those of Random Forest and Decision Tree; hence, they reflect a smaller portion of the response variability.

4. Linear Regression: A will not miserably (done reasonably well) in case of moderate values under most of the metrics.

Ideally, the R-squared and adjusted R-squared values (R2 and EVS) of tree-based models should be higher than those of tree-based models when predicting non-linear relationships.

5. MLP Regressor: One of the things that this video does best is eloquently illustrate complex relationships with competitive performance metrics.

The Max Error in this model has also been slightly higher than in Random Forest and Decision Tree models.

6. SVR: The models fell ahead of every other model: Negative R2 and EVS imply the lack of a suitable representation of the underlying pattern within the data.

The Random Forest Regressor [20] achieves the best performance results among the considered models according to the complete assessment for the given dataset, which suggests it to be the most probable model for the power production prediction. Its joint of low error metrics high explained variance, and robustness on outliers contribute to the fact that it is the top-notch method for tackling such a problem.

5. Conclusion

In summary, this research aims to achieve accurate forecasts of power generation using a data-driven approach accomplished by in-depth analysis of the performance of the multiple machine learning models in the context of the "Daily Power Generation Data" dataset. Exploration covers six models: Random Forest Regressor, Decision Tree Regressor, Nearest Neighbors, Linear Regression, MLP Regressor, and SVR Evaluated by Performance Metrics at all Fronts. It is, therefore, noteworthy that Random Forest Regressor was adjudged as most excellent in terms of the lowest Mean Squared Error (MSE) by 3.57E-06, which is a good indication of its productiveness. The thorough evaluation of Metrics like RMSE, MAE, EVS, and Max Error proves the great resistance of Random Forest Regression, which illustrates outliers and performs the external variations in power data analysis. The results gained from this study have the highest significance for those who are involved in power company management; they can get useful information on what method will be the best for the successful implementation of predictive analytics. The Random Forest Regressor was the only model among others that had dominantly demonstrated competence, the other models having enough but less than the RF Regressor. This statistically explains that the RF Regressor is the best among them all for accurate and reliable power generation predictions. Nevertheless, the development of an ultimate framework to meet all the target aspects and limitations is a challenging process.

Funding: “This research received no external funding”

Conflicts of Interest: “The authors declare no conflict of interest.”

References

[1] Almetwally, E. M., & Meraou, M. A. (2022). Application of environmental data with new extension of Nadarajah-Haghighi distribution. Computational Journal of Mathematical and Statistical Sciences, 1(1), 26–41. https://doi.org/10.21608/cjmss.2022.271186

[2] Muhammed, H. Z., & Almetwally, E. M. (2024). Bayesian and non-Bayesian estimation for the shape parameters of new versions of bivariate inverse Weibull distribution based on progressive type II censoring. Computational Journal of Mathematical and Statistical Sciences, 3(1), 85–111. https://doi.org/10.21608/cjmss.2023.250678.1028

[3] Towfek, S. K. (2023). Navigating the storm: Cutting-edge risk mitigation and analysis for volatile markets. Journal of Artificial Intelligence and Metaheuristics, 4(2), 36–44. https://doi.org/10.54216/JAIM.040204

[4] Towfek, S. K. (2023). A semantic approach for extracting the medical association rules. Journal of Artificial Intelligence and Metaheuristics, 5(1), 46–52. https://doi.org/10.54216/JAIM.050105

DOI: https://doi.org/10.54216/JAIM.080201
Received: March 24, 2024 Revised: May 25, 2024 Accepted: November 01 2024

[5] Alotaibi, R., AL-Dayian, G. R., Almetwally, E. M., & Rezk, H. (2024). Bayesian and non-Bayesian two-sample prediction for the Fréchet distribution under progressive type II censoring. AIP Advances, 14(1), 015137. https://doi.org/10.1063/5.0174390

[6] El-Kenawy, E.-S. M., Khodadadi, N., Mirjalili, S., Abdelhamid, A. A., & Eid, M. M. et al. (2024). Greylag goose optimization: Nature-inspired optimization algorithm. Expert Systems with Applications, 238, 122147. https://doi.org/10.1016/j.eswa.2023.122147

[7] El-Kenawy, E.-S., Ibrahim, A., Mirjalili, S., Zhang, Y.-D., & Elnazer, S. et al. (2022). Optimized ensemble algorithm for predicting metamaterial antenna parameters. Computers, Materials & Continua, 71(3), 4989–5003. https://doi.org/10.32604/cmc.2022.023884

[8] Khafaga, D. S., Ibrahim, A., El-Kenawy, E.-S. M., Abdelhamid, A. A., & Karim, F. K. et al. (2022). An Al-Biruni earth radius optimization-based deep convolutional neural network for classifying monkeypox disease. Diagnostics, 12(11), Article 11. https://doi.org/10.3390/diagnostics12112892

[9] Djaafari, A., Ibrahim, A., Bailek, N., Bouchouicha, K., & Hassan, M. A. et al. (2022). Hourly predictions of direct normal irradiation using an innovative hybrid LSTM model for concentrating solar power projects in hyper-arid regions. Energy Reports, 8, 15548–15562. https://doi.org/10.1016/j.egyr.2022.10.402

[10] AlEisa, H., El-Kenawy, E.-S., Alhussan, A., Saber, M., & Abdelhamid, A. et al. (2022). Transfer learning for chest X-rays diagnosis using dipper throated algorithm. Computers, Materials & Continua, 73(2), 2371–2387. https://doi.org/10.32604/cmc.2022.030447

[11] El-Kenawy, E.-S. M., Khodadadi, N., Mirjalili, S., Makarovskikh, T., & Abotaleb, M. et al. (2022). Metaheuristic optimization for improving weed detection in wheat images captured by drones. Mathematics, 10(23), Article 23. https://doi.org/10.3390/math10234421

[12] Abdelhamid, A. A., Towfek, S. K., Khodadadi, N., Alhussan, A. A., & Khafaga, D. S. et al. (2023). Waterwheel plant algorithm: A novel metaheuristic optimization method. Processes, 11(5), Article 5. https://doi.org/10.3390/pr11051502

[13] Khafaga, D., El-Kenawy, E.-S., Karim, F., Alshetewi, S., & Ibrahim, A. et al. (2022). Optimized weighted ensemble using dipper throated optimization algorithm in metamaterial antenna. Computers, Materials & Continua, 73(3), 5771–5788. https://doi.org/10.32604/cmc.2022.032229

[14] Bhavsar, S., Pitchumani, R., & Ortega-Vazquez, M. A. (2021). Machine learning enabled reduced-order scenario generation for stochastic analysis of solar power forecasts. Applied Energy, 293, 116964. https://doi.org/10.1016/j.apenergy.2021.116964

[15] Ding, S., Zhang, H., Tao, Z., & Li, R. (2022). Integrating data decomposition and machine learning methods: An empirical proposition and analysis for renewable energy generation forecasting. Expert Systems with Applications, 204, 117635. https://doi.org/10.1016/j.eswa.2022.117635

[16] Mukherjee, D., Chakraborty, S., & Ghosh, S. (2022). Power system state forecasting using machine learning techniques. Electrical Engineering, 104(1), 283–305. https://doi.org/10.1007/s00202-021-01328-z

[17] Daily power generation data. (2024). Kaggle. Retrieved March 13, 2024, from https://www.kaggle.com/datasets/arvindnagaonkar/power-generation-data

[18] Alam, S., & Yao, N. (2019). The impact of preprocessing steps on the accuracy of machine learning algorithms in sentiment analysis. Computational and Mathematical Organization Theory, 25(3), 319–335. https://doi.org/10.1007/s10588-018-9266-8

[19] Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning. Electronic Markets, 31(3), 685–695. https://doi.org/10.1007/s12525-021-00475-2

[20] Xue, L., Liu, Y., Xiong, Y., Liu, Y., & Cui, X. et al. (2021). A data-driven shale gas production forecasting method based on the multi-objective random forest regression. Journal of Petroleum Science and Engineering, 196, 107801. https://doi.org/10.1016/j.petrol.2020.107801

DOI: https://doi.org/10.54216/JAIM.080201
Received: March 24, 2024 Revised: May 25, 2024 Accepted: November 01 2024