International Journal of Science and Research Archive
eISSN: 2582-8185
Cross Ref DOI: 10.30574/ijsra
Journal homepage: https://ijsra.net/

(Research Article)

ML in energy sector revolutionizing the energy sector machine learning applications for efficiency, sustainability and predictive analytics

Krishna Gandhi * and Pankaj Verma

Independent Researcher, India

International Journal of Science and Research Archive, 2022, 07(01), 533-541

Publication history: Received on 20 September 2022; revised on 25 October 2022; accepted on 28 October 2022

Article DOI: https://doi.org/10.30574/ijsra.2022.7.1.0226

Abstract

The increasing global energy demand, coupled with the need for sustainability, has necessitated innovative solutions in energy management. This study explores an application of ML techniques to revolutionize the energy sector, emphasizing efficiency, sustainability, and predictive analytics. This study evaluates a performance of proposed ML models in optimizing energy efficiency and predictive analytics for renewable energy applications. Using real-time sensor data encompassing energy consumption, weather conditions, equipment malfunctions, and grid statistics, the dataset was preprocessed and analyzed with proposed models: RF, Neural Networks, GB, SVM, and KNN. These models were assessed using metrics such as accuracy, training time, scalability, interpretability, and energy impact. Among the proposed models, Neural Networks achieved the highest accuracy, 92% and energy impact, 30%, while Random Forest offered a balanced trade-off between accuracy (89%), scalability, and interpretability. The outcomes underscore a potential of the proposed ML models in advancing energy systems, highlighting Neural Networks for optimization and Random Forest for real-time applications. Future work aims to address computational limitations and expand model adaptability for diverse energy scenarios.

Keywords: Renewable Energy (RE); Predictive Analytics, Energy Sector; ML In Energy Sector Revolutionizing; Machine Learning (ML)

1. Introduction

The global society is bearing witness to a monumental transformation in power generation and distribution, use and management. Due to increased globalization of people and the expansion of the Industrial Revolution, demands for energy hit a record high. At the same time, conventional, non-renewable energy sources are becoming more and more unsustainable due to the pressing need to decrease carbon emissions and lessen the effects of climate change. A need for innovative, efficient, and sustainable solutions has never been more critical [1].

The energy sector is changing rapidly, and renewable energy (RE) has come to play a central role in satisfying the need for a more sustainable power system. As the world goes for decarbonization and technological capability increases, the developing world like India has seen the prospect of RE resources in diverting from conventional and carbon-based energy[2]. Nonetheless, the energy sector experiences some barriers in the efficient management and distribution of renewable energy and thus the need to embrace change drivers that can help in efficient management and distribution of renewable energy.

Intelligent grid management, efficiency in energy throughput, and optimization techniques are the buzzwords of the present day energy industry [3]. By offering strong predictive analytics, ML helps in understanding the demand and possible disruptions; make the preemptive action to continuous, smooth energy generation and utilization. The other

* Corresponding author: Krishna Gandhi

major factor has been sustainability in the drive towards digitalization of the energy sector[4]. By introducing AI and ML platforms into energy systems, energy efficiency can be improved, energy losses and emissions minimized alongside optimal resource distribution. However, major technologies that are key to this change are AI and ML[5]. These advanced technologies have been observed to play the role of enhancing efficiency and sustainability as well as the ability to forecast in the energy sub-sector.

1.1. Motivation and Contribution of Study

The motivation behind this study stems from the urgent need to enhance efficiency, sustainability, and predictive capabilities in a renewable energy sector. Innovative strategies to optimize energy management are urgently needed due to the rising worldwide demand for energy and the increasing focus on moving away from fossil fuels and towards cleaner, renewable sources. Through the use of ML techniques, this research seeks to solve these issues by enhancing renewable energy systems' defect detection, forecasting accuracy, and energy efficiency. The following are this study's primary contributions:

- Data collection from diverse sources, including energy-related open repositories, leading publications, and real-time sensor data from renewable energy systems.
- Comprehensive data preprocessing involving handling missing values, noise reduction, and feature scaling using Z-score normalization.
- Implementation of ML models, including RF NN, GB, SVM, and KNN, to evaluate their effectiveness in renewable energy applications.
- Performance evaluation of models using metrics such as accuracy to provide a comprehensive assessment.

1.2. Structure of paper

The structure of this paper is as follows: Section 2 reviews a literature on machine learning in energy. Section 2I details the methodology, including data collection, preprocessing, and models used. Section IV presents results, performance metrics, and analysis. Section V concludes with key findings and future directions for enhancing energy systems with machine learning.

2. Literature Review

This portion of the report examines the body of research on the use of ML methods in the energy industry. Most of a reviewed works focused on leveraging machine learning for energy forecasting, fault detection, and efficiency optimization. Some reviews are:

This study, Goyal, Pandey and Thakur (2020) employ the following eight building attributes to forecast HVAC demands: relative compactness, surface area, wall area, roof area, overall height, orientation, glazing area, and glazing area distribution. MLR, KNN, and SVR are the ML algorithms that have been used in the experiments. The three ensemble algorithms that have also been used are RFs, GBMs, and Extreme GB. Accuracy, R-squared, MSE, RMSE, and MAE are some of the performance indicators that have been used to assess the models. Ensemble methods significantly outperform their Machine Learning counterparts, according to the results [6].

In this study, Lai et al. (2020) examination are shown below. As a first step, this study will examine and evaluate machine learning models used for renewable energy forecasts. As a third step, they calculated the coefficient of determination, MAPE, and different renewable energy sources. Lastly, this poll concluded with a few suggestions for further work [7].

This study, Bokonda, Ouazzani-Touhami and Souissi (2020) surveyed the current state of research in predictive analysis using machine learning techniques and trends. As a result of this procedure, thirty research articles were chosen for this evaluation. Based on the most recent findings in the field, this study aims to help researchers, corporations, or anybody interested in predictive analysis identify the most appropriate ML method(s) for their specific needs[8].

In this study, Serban and Lytras (2020) take advantage of the new trends in the use of AI in the renewable energy industry inside the European Union (EU). Specifically, they looked at (i) how well REs are transformed from GIC to FEC, (2) how this affects the composition of REs according to source (solar, wind, biomass, etc.), (2i) how the RE sector's labor productivity compares to the economy overall and how it correlates with investment levels, and (iv) how AI adoption in RE could impact research into future smart cities [9].

In this study, Mukhopadhyay and Nateghi (2017) provide a non-parametric Bayesian approach to detect the energy demand–climate link. In addition, stakeholders may use their expert knowledge to apply prior probability on key climatic variables using the Bayesian framework. They analyzed the energy consumption patterns of Indiana's homes and businesses to demonstrate the practicality of their suggested approach[10].

The comparative analysis of background study based on their Reference, Study Focus, Methodology, Key Findings, Limitations, and Future Work are provided in Table 1.

Table 1 Summary of Background Study of Machine Learning Applications in the Energy Sector

Reference	Methodology	data	Findings	Limitations/Future study
Goyal, Pandey, and Thakur (2020)	ML algorithms (Multiple LR, KNN, SVM), Ensemble methods (RF, GBM, Extreme Gradient Boosting)	Building energy characteristics (Relative Compactness, Surface Area, etc.)	Ensemble methods outperform machine learning techniques by a significant margin.	Future work on enhancing predictive models for different building types.
Lai et al. (2020)	Literature review and analysis of ML models for renewable-energy predictions	Renewable energy datasets (unspecified)	Reviewed machine learning models, data pre-processing, parameter selection, and performance measurement techniques for renewable energy predictions.	Future opportunities for optimization of prediction models in renewable energy sector.
Bokonda, Ouazzani-Touhami, and Souissi (2020)	Literature review of machine learning methods for predictive analysis	30 research papers from scientific journals (last 5 years)	Identified best machine learning methods for predictive analysis across different fields based on the latest research.	Future studies could include more diverse data sources and application areas.
Serban and Lytras (2020)	Analysis of AI adoption in the renewable energy sector in the EU	Data on renewable energy transformation processes and labor productivity	Evaluated the impact of AI on renewable energy sector transformation, productivity, and investment correlation.	More detailed analysis of AI's impact on smart city development and energy transformation.
Mukhopadhyay and Nateghi (2017)	Bayesian non-parametric method for energy demand-climate nexus	Residential and Commercial Energy Data in Indiana	Proposed probabilistic models to capture energy demand and climate relationship, with applicability to future demand projections.	Future study on incorporating broader climate variables and integrating with existing models.

3. Methodology

The purpose of this research is to examine and contrast how well various ML models improve predictive analytics and energy efficiency. The methodology involves data collection of real-time sensor data from renewable energy systems, covering energy consumption profiles, weather conditions, equipment malfunctions, and grid statistics. After preprocessing the data—handling missing values, eliminating noise, and normalizing using Z-score and Min-Max scaling—the dataset was split into training and testing subsets. Multiple ML models, including RF, Neural Networks, GB, SVM, and KNNs, were trained and tested. Model performance was evaluated based on accuracy, with the results compared to factors such as training time, scalability, interpretability, and energy impact to identify the most suitable model for renewable energy applications. The following steps and phases of the proposed approach are described in the flowchart, which is shown in Figure 1.

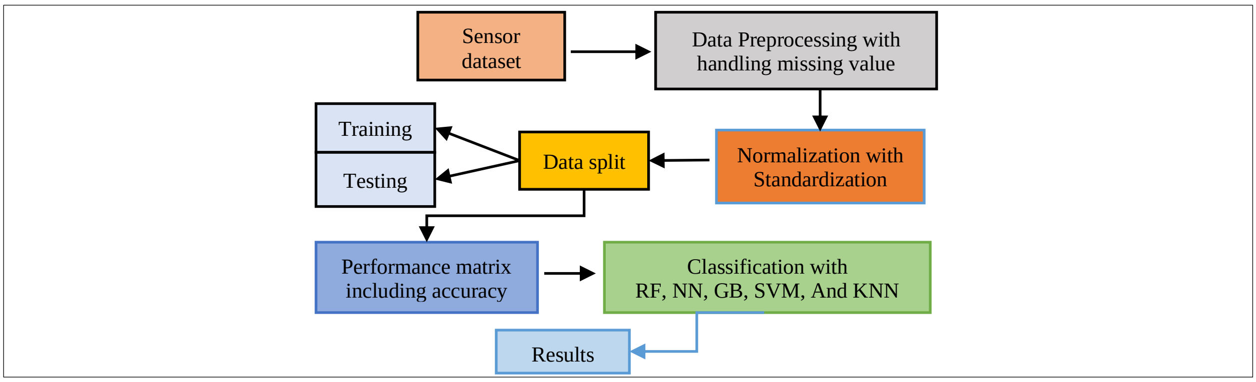

Figure 1 Flowchart for Energy Sector Revolutionizing

The overall steps of the flowchart for ML in the Energy Sector are provided below:

3.1. Data Collection

The dataset is collected from energy-related open repositories, leading case publications, and real-time data from sensors installed in renewable energy (RE) systems. These datasets include energy consumption profiles, weather condition data, equipment malfunction and failure records, and grid operation statistics.

3.2. Data Preprocessing

The concept of "data preprocessing" is used to describe any action taken on unprocessed data before it is used. The act of merely converting unprocessed data into a comprehensible format is known as data preprocessing. Data from the real world is sometimes loud, redundant, inconsistent, and incomplete. Several procedures are used in data preparation to assist in transforming unprocessed data into representations that are logical and processed. The following pre-processing steps are listed below:

Handling Missing Values: Missing values were identified and removed to maintain the integrity of the dataset. Transactions with missing critical features were excluded.

3.3. Normalization with Standardization

They normalized and standardized the data. The Z-score normalization approach was used for standardization (1):

$$z = \frac{x-\mu}{\sigma}$$

the feature values are represented by X, the mean is denoted by μμ, and the standard deviation is denoted by σ.

3.4. Data split

There was a 70% training subset and a 30% testing subset in the dataset. This was executed with the help of scikit-learn's train, test split function.

3.5. Models Selection

The proposed method includes machine-learning algorithms. This study uses RF, NN, GB, SVM, and KNN. Each classifier is described in below:

3.5.1. Random Forest

The Random Forest method is an ensemble learning tool that uses a network of interconnected decision trees to generate predictions [11]. With RF, you get a final forecast by adding together the forecasts of all the decision trees. Each tree makes its own prediction. X represents the input characteristics, *RF* stands for the Random Forest model, and Y represents the target variable[12]. Prediction of *RF* may be expressed as (2), supposing there are N DTs in the forest.

$$RF(X) = mode(Tree_1(X), Tree_2\ (X), \ldots\ldots, Tree_N(X))$$

where $(Tree_1(X)$ stands for the forecast made by the DT with the i-th node. The most common class label across all trees' predictions is returned by mode () in a classification process.

3.5.2. Neural Networks

The mapping, a flotation process model, was implemented in the paper using neural networks trained using the uncertainty back-propagation approach[13].

3.5.3. Extreme gradient boosting (Xgboost)

XGBoost is an ensemble learning technique based on decision trees that builds a robust model by combining weak learners. It involves creating a series of decision trees with the goal of fixing the mistakes of the earlier ones [14]. The outcome is a final model that can handle complicated and big datasets with ease, and it is also very accurate and generalized.

3.5.4. Support Vector Machine

SVMs' capacity to deal with data that is both high-dimensional and non-linearly separable has contributed to their meteoric rise in popularity. Additionally, SVM can deal with imbalanced data and is resilient to outliers. SVMs classify input data using a hyperplane [15]. The greatest margin hyperplane is the one that maximizes the distance between the classes.

3.5.5. k-nearest neighbor

Finding the k training samples that are most similar to the target item in the training set is the suggested use of the K-NN method. Moreover, choose the most prominent category from the k training examples; subsequently, apply this most prominent category to the target object; here, k represents the number of training examples. Predicting how far away the predicted data point is from the known data point is an important step for the K-NN method [16]

3.6. Performance Metrics

A classification model's efficacy may be measured using performance measures. The actual and expected classes make up the two dimensions of the confusion matrix. In contrast to the projected class states shown in the columns, the actual class examples in the rows are represented by the data. The confusion matrix has four columns: TP for true positives, TN for true negatives, FP for false positives, and FN for false negatives. Accuracy, Precision, Recall, F1-Score, and AUC are among the metrics that may be obtained from these numbers.

Accuracy: Accuracy is calculated by dividing the total number of forecasts by the proportion of accurate predictions (True Positive + True Negative). Accuracy was calculated with the following Equation (3).

$$Accuracy = \frac{\mathrm{TN + TP}}{\mathrm{TP + TN + FP + FN}} \qquad (3)$$

Recall: Recall gauges a classifier's sensitivity or completeness. False negatives are reduced when recall is higher and increased when recall is lower. Recall which can be presented by the following Equation (4).

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP{+}FN}} \quad (4)$$

Precision: Precision quantifies a classifier's accuracy. False positive rates decrease with increasing precision and increase with decreasing precision. The following Equation (5).

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP{+}FP}} \qquad (5)$$

F1 Score: F-measure, or the weighted harmonic mean of precision and recall, is a metric that may be created by combining precision and recall. One is the best value, while zero is the worst. This is how the F1-measure is computed: This Equation (6)

$$\mathrm{F1} = \frac{2*(\mathrm{precision}*\mathrm{recall})}{\mathrm{precision}+\mathrm{recall}} \qquad (6)$$

The utilized evaluation metrics are selected due to their effectiveness in evaluating and analyzing the performance.

4. Results And Discussion

The results of the various categorization techniques used in this investigation are examined in this section. Their study utilized ML techniques for optimizing energy efficiency and predictive analytics in energy systems. These methods include RF, NN, GB, SVM, and KNN. The performance of these algorithms was evaluated using datasets collected from energy-related repositories, real-time sensor data, and case studies. Metrics including F1-score, recall, accuracy, and precision were used to compare the models' performance.

Table 2 Comparison of base and proposed model performance of Energy Applications

Performance Metric	Propose Models			Base Models	
	RF	NN	GB	SVM	KNN
Accuracy	89	92	91	85	78
Training Time	12	35	25	10	5
Scalability	High	Medium	High	Low	Low
Interpretability	High	Low	Medium	High	High
Energy Impact Analysis	25%	30%	28%	20%	15%

Table 2 comparison of model performance reveals that Neural Networks (NN) achieved the highest accuracy of 92% and a significant energy impact of 30%, though it required the longest training time of 35 seconds. GB followed with 91% accuracy and 28% energy impact, balancing accuracy and training time (25 seconds). RF demonstrated a strong accuracy of 89% and the shortest training time (12 seconds), making it ideal for real-time applications with high scalability and interpretability. SVM had 85% accuracy, a lower energy impact (20%), and slower scalability, while KNN showed the lowest accuracy 78% and energy impact (15%) but the fastest training time (5 seconds). Overall, NN proved most effective for energy optimization, while RF offered a good balance for real-time use.

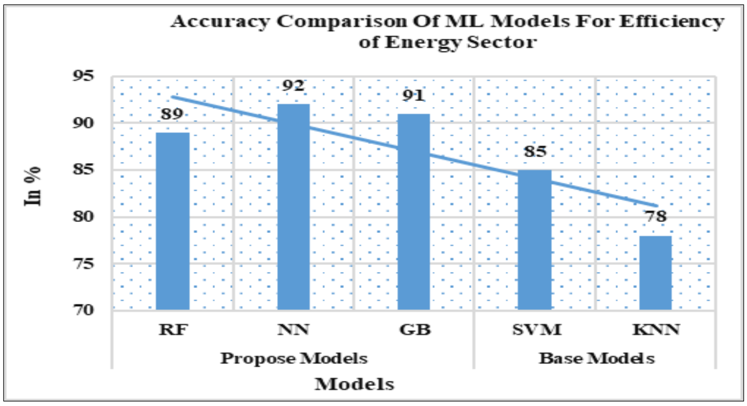

Figure 2 Accuracy of ML Models for Energy Applications

The performance of the proposed models shown in Figure 2, including RF, NN, GB, SVM, and KNN, was evaluated based on accuracy. Among the proposed models, Neural Networks (NN) achieved the highest accuracy at 92%, followed closely by GB at 91% and Random Forest (RF) at 89%. SVM and KNN showed comparatively lower accuracies at 85% and 78%, respectively. This highlights the superior performance of ensemble-based and neural network approaches over traditional machine learning techniques for the given task.

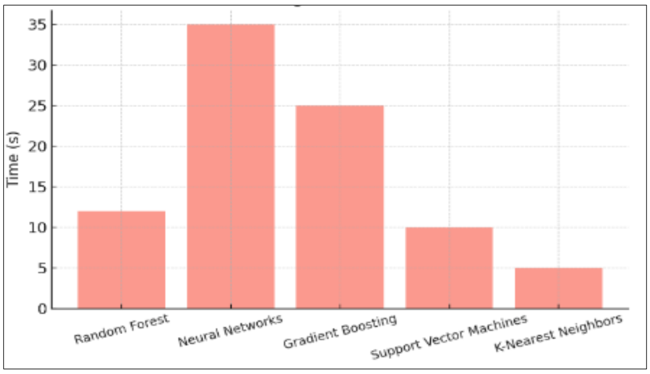

Figure 3 Training Time for ML Models (in Seconds)

Figure 3 compares the training time required for five machine learning models: RF, NN, GB, SVM, and KNN. The training time is measured in seconds. Neural Networks exhibit the longest training time, whereas RF and KNN demonstrate the shortest training times. SVM and GB also require relatively longer training periods.

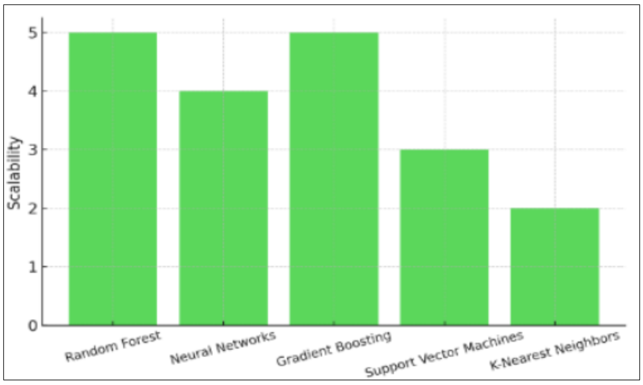

Figure 4 Scalability of ML Models

Figure 4 compares the scalability of five machine learning models: RF, NN, GB, SVM, and KNN. Scalability is a measure of how well a model can handle large datasets and complex problems. According to the graph, RF and GB exhibit the highest scalability, while KNN demonstrates the lowest scalability. Neural Networks and SVM show moderate scalability levels.

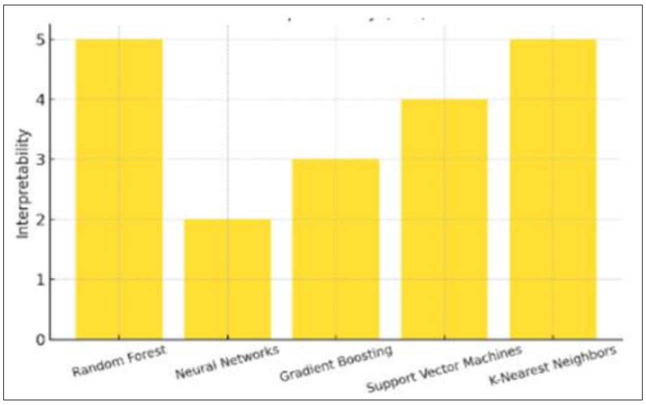

Figure 5 Interpretability of ML Models

Figure 5 compares the interpretability of five machine learning models. According to the graph, RF and KNN are considered to be the most interpretable models. NN and SVM are generally considered to be less interpretable. Gradient Boosting falls somewhere in the middle in terms of interpretability.

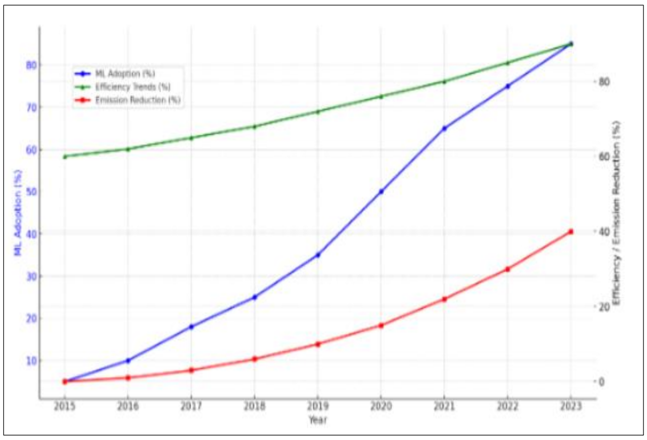

Figure 6 Year-wise trends of ML adoption

Figure 6 tracks the progress of a technology or practice over time. The blue line shows a steady increase in adoption, indicating growing popularity. The green colored line shows an increase in the efficiency measure implying a positive development. The last one, trend line, exhibits a downward trend in emissions proving the environmental effectiveness of the showcased technology. From the above-mentioned trends, implications are that not only is this technology or practice on the rise, but also that it is increasingly becoming more sustainable.

5. Conclusion

The application of utilizing ML for energy efficiency, sustainability, and improved predictive analytics has disrupted the energy industry currently and in the future. With the increasing global energy demand and the changing need to shift to renewable energy sources, the applications of ML are new and insightful for managing energy consumption, predicting energy consumption patterns, maintaining grid stability and controlling carbon emissions. This study demonstrates the efficacy of machine learning models in enhancing energy efficiency and predictive analytics for renewable energy applications, with Neural Networks achieving the highest accuracy 92% and energy impact 30%. This work also has some limitations as follows: NN and GB often present high computational demands, making it unsuitable for a low computational environment. Furthermore, the flexibility of utilized dataset and real-time data application are still problematic for generalizable conditions. Future work is to refine the existing architectures to overcome the computational high cost and include transfer learning in the models, and for the future, there will be a need for developing both single and two-tier hybrid models that are both accurate and interpretable for real-world energy systems.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] R. M. Elavarasan et al., “A Comprehensive Review on Renewable Energy Development, Challenges, and Policies of Leading Indian States with an International Perspective,” IEEE Access, vol. 8, no. April, pp. 74432–74457, 2020, doi: 10.1109/ACCESS.2020.2988011.

[2] C. Bordin, S. Mishra, A. Safari, and F. Eliassen, “Educating the energy informatics specialist: opportunities and challenges in light of research and industrial trends,” SN Appl. Sci., vol. 3, no. 6, 2021, doi: 10.1007/s42452-021-04610-8.

[3] P. L. Donti and J. Z. Kolter, “Machine Learning for Sustainable Energy Systems,” Annu. Rev. Environ. Resour., vol. 46, pp. 719–747, 2021, doi: 10.1146/annurev-environ-020220-061831.

[4] S. Kufeoğlu, G. Liu, A. Karim, and M. Pollitt, “Digitalisation and New Business Models in Energy Sector,” EPRG Work. Pap. . Energy Policy Res. Group, Univ. Cambridge, no. December, pp. 1–39, 2019.

[5] D. Theng and M. Theng, “Machine Learning Algorithms for Predictive Analytics: A Review and New Perspectives,” High Technol. Lett., no. July, pp. 1–10, 2020, doi: 10.37896/HTL26.06/1159.

[6] M. Goyal, M. Pandey, and R. Thakur, “Exploratory Analysis of Machine Learning Techniques to predict Energy Efficiency in Buildings,” in 2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), 2020, pp. 1033–1037. doi: 10.1109/ICRITO48877.2020.9197976.

[7] J. P. Lai, Y. M. Chang, C. H. Chen, and P. F. Pai, “A survey of machine learning models in renewable energy predictions,” Applied Sciences (Switzerland). 2020. doi: 10.3390/app10175975.

[8] P. L. Bokonda, K. Ouazzani-Touhami, and N. Souissi, “Predictive analysis using machine learning: Review of trends and methods,” in 2020 International Symposium on Advanced Electrical and Communication Technologies (ISAECT), 2020, pp. 1–6. doi: 10.1109/ISAECT50560.2020.9523703.

[9] A. C. Serban and M. D. Lytras, “Artificial intelligence for smart renewable energy sector in europe - Smart energy infrastructures for next generation smart cities,” IEEE Access, vol. 8, pp. 77364–77377, 2020, doi: 10.1109/ACCESS.2020.2990123.

[10] S. Mukhopadhyay and R. Nateghi, “Estimating climate — Demand Nexus to support longterm adequacy planning in the energy sector,” in 2017 IEEE Power & Energy Society General Meeting, 2017, pp. 1–5. doi: 10.1109/PESGM.2017.8274648.

[11] M. R. Haque, M. M. Islam, H. Iqbal, M. S. Reza, and M. K. Hasan, “Performance evaluation of random forests and artificial neural networks for the classification of liver disorder,” in 2018 international conference on computer, communication, chemical, material and electronic engineering (IC4ME2), 2018, pp. 1–5.

[12] M. Schonlau and R. Y. Zou, “The random forest algorithm for statistical learning,” Stata J., 2020, doi: 10.1177/1536867X20909688.

[13] D. Jamróz, T. Niedoba, P. Pieta, and A. Surowiak, “The use of neural networks in combination with evolutionary algorithms to optimise the copper flotation enrichment process,” Appl. Sci., vol. 10, no. 9, 2020, doi: 10.3390/app10093119.

[14] N. Mahendran, P. M. Durai Raj Vincent, K. Srinivasan, V. Sharma, and D. K. Jayakody, “Realizing a Stacking Generalization Model to Improve the Prediction Accuracy of Major Depressive Disorder in Adults,” IEEE Access, 2020, doi: 10.1109/ACCESS.2020.2977887.

[15] Tarek Amr, Hands-On Machine Learning with scikit-learn and Scientific Python Toolkits. 2020.

[16] G. F. Fan, Y. H. Guo, J. M. Zheng, and W. C. Hong, “Application of the weighted k-nearest neighbor algorithm for short-term load forecasting,” Energies, vol. 12, no. 5, 2019, doi: 10.3390/en12050916