Using machine learning for the optimisation of operations and management in electric systems and networks

Semen Levin[1*]

[1]Tomsk State University of Control Systems and Radioelectronics, 40, prospect Lenina, Tomsk, 634050, Russia

Abstract. This research employs the Random Forest Machine Learning model to predict electricity consumption and detect anomalies in electrical networks. Addressing the energy sector's challenges, such as supply reliability and renewable energy integration, this model processes historical electricity consumption data, weather conditions, and network events to efficiently forecast demand and identify anomalies. Data cleansing and normalisation preceded the training phase, where the model was fine-tuned using historical data to balance forecast accuracy and overfitting avoidance. The dataset was divided into training (80%) and testing (20%) sets for performance evaluation. Through cross-validation, optimal model hyperparameters were determined. The findings highlight the model's efficacy in accurately predicting daily electricity consumption in a small, homogenous town. The model achieved a Mean Absolute Error (MAE) of 198.73 MWh and a coefficient of determination (R^2) of 0.9387. Temperature, humidity, and wind speed were identified as key influencing factors on consumption levels. Conclusively, the Random Forest model presents a valuable tool for energy management, offering precise consumption forecasting and anomaly detection capabilities. Future work will address computational demands and enhance model integration with other Machine Learning methods for improved performance. This contribution is significant for efficient energy system planning and operation.

1 Introduction

The contemporary world is on the cusp of significant changes in electrical power management. The increase in population, urbanisation, and technological progress pose new challenges to power systems. Among these are ensuring supply reliability, optimising expenses, and integrating renewable energy sources. In this segment, particular attention is paid to the application of Machine Learning (ML) to enhance electric networks [1].

Historically, power supply systems evolved towards increasing capacity and expanding networks. However, current realities demand a transition to more nuanced and flexible management, where big data analysis and automation play crucial roles. With its capability

* Corresponding author: semen.m.levin@tusur.ru

to process vast amounts of information and identify patterns within the data, Machine Learning appears to be an ideal tool for achieving these objectives [2].

In recent years, significant progress has been observed in research and developments to incorporate ML into the energy sector. It includes forecasting electricity consumption [3], optimising power plant and substation operations, fault detection and anomaly identification in network operations, and managing energy flow to integrate decentralised sources [4, 5]. Despite the clear prospects, integrating ML into the energy infrastructure faces several issues. These include real-time data collection and processing difficulties, the need to adapt existing energy systems to new technologies, and security and data privacy concerns. Furthermore, an important aspect is developing algorithms capable of operating under uncertainty and changing input data.

The research described in this article is focused on exploring the capabilities of machine learning for addressing pressing issues in the energy sector. Special attention is given to forecasting, analysis, and management within the context of smart electric grids. The primary applications of ML in the energy sector include load forecasting, optimisation of generating capacities, ensuring power supply reliability, and analysing and managing energy flows in distribution networks [6-12]. Additionally, diagnosing and predicting equipment failures is a significant area, contributing to preventing emergencies and minimising downtime in power systems [13-14].

1.1 Theoretical foundations

1.1.1 Load forecasting

Load forecasting is a pivotal task in the energy sector, as accurate electricity consumption prediction allows for efficient production and distribution planning [16-17]. ML methods, such as artificial neural networks, support vector machines, and decision trees, demonstrate high accuracy in solving this task. They can count for a wide range of factors, including time series, weather conditions, economic indicators, and even social events.

1.1.2 Optimisation of generation and distribution

ML algorithms that can analyse numerous parameters in real time are used to optimise the operations of generating capacities and manage electricity flows. It not only enhances the efficiency of resource utilisation but also reduces operational expenses and environmental impact [18].

1.1.3 Reliability and diagnostics

ML also plays a crucial role in ensuring power supply reliability [19]. Forecasting and pattern recognition algorithms enable the timely identification and prevention of potential faults and accidents in power systems. Deep learning methods for data analysis from sensors and monitoring systems opens new possibilities for early diagnostics of equipment conditions.

1.1.4 Integration of renewable energy sources

With the increasing share of renewable energy sources (RES) in the energy mix, the challenge of their integration into the overall power system becomes more pressing [20]. ML methods can facilitate smoother integration of RES by predicting their output, which is subject to significant fluctuations due to weather conditions. It includes solar and wind power stations and other types of renewable sources, such as hydroelectricity and bioenergy. Accurate prediction of RES output allows for effective load management in the network, optimising resource distribution and reducing the need for backup power sources.

1.1.5 Adaptive control and automation

The complexity and dynamism of modern energy systems require high adaptability and automation from control systems [21]. ML algorithms contribute to developing intelligent control systems that can automatically adjust to changing conditions and optimise processes in real time. It covers many tasks, from load balancing to instantaneous response to emergencies.

1.1.6 Security and protection from attacks

Cybersecurity issues have become increasingly significant in the digital age of energy systems [22]. Machine learning provides practical tools for detecting and preventing cyberattacks on the energy infrastructure. Algorithms can analyse network traffic patterns and promptly identify anomalies indicating attempts at unauthorised access or other threats.

1.1.7 Energy efficiency and sustainable development from attacks

Improving energy efficiency and transitioning to sustainable development are key tasks for the global energy industry [23]. Machine learning can contribute to these goals by optimising resource consumption, reducing energy losses, and minimising harmful emissions. Intelligent data analysis helps identify the potential for efficiency improvement at all levels of the energy system, from individual consumers to extensive energy facilities.

2 Random Forest machine learning model

2.1 Fundamentals of the model

In the research, particular attention is given to applying the Random Forest model to analyse and optimise the operations of electric power systems. The Random Forest model is an ensemble machine learning method that combines predictions from multiple decision trees to obtain a more accurate and stable forecast. The choice of this model is motivated by several key advantages that make it particularly suitable for tasks in the energy sector.

Firstly, the Random Forest efficiently solves both regression and classification tasks. In the context of the energy sector, this allows for accurate energy consumption forecasting and effective anomaly detection in network operations. Thanks to the bagging mechanism and multiple decision trees, the Random Forest is less prone to overfitting than single decision trees, providing more reliable and generalisable results [24]. Additionally, the Random Forest model offers valuable information on feature importance, helping to understand better the factors influencing energy consumption. It aids in the optimisation and planning of electric power systems.

Besides, the Random Forest can efficiently handle large volumes of data with numerous features and does not require prior data processing, such as normalisation or standardisation [25], making it ideally suited for working with diverse data in the energy sector.

The foundation of the Random Forest consists of decision trees. Each tree is built based on a random subset of training data and features, reducing correlation among trees and enhancing the model's generalisation capability. The Random Forest uses the bagging method [26], which involves creating multiple bootstrap samples (subsamples with replacement) from the original dataset. A separate decision tree is trained on each of these subsamples. When constructing each node in a tree, a random subset of features is considered, from which the best feature for splitting the data into two parts based on a specific criterion (e.g., Gini index or entropy) is selected.

2.1.1 Mathematical description

Let $X = \{x_1, x_{2,\ldots,} x_n\}$ represent a set of features, and *Y* be the target variable (in our case, electricity consumption). The Random Forest constructs an ensemble of *N* decision trees $\{D_1, D_{2,\ldots,} D_n\}$, each trained on its bootstrap sample from *X* and using a random subset of features for each node.

The prediction of the model *f(x)* for a new observation *x* is the averaged (for regression tasks) or most frequent (for classification tasks) value of the predictions from all trees in the ensemble:

$$f(x) = \frac{1}{N}\sum_{i=1}^{N} D_i(x) \tag{1}$$

For regression tasks, where $D_i(x)$ is the prediction of the *i*-th tree.

2.1.2 Splitting criteria

Decision trees in the Random Forest use criteria to determine how best to split the data at each node. For regression tasks, the Mean Squared Error (MSE) is often used:

$$MSE = \frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{2}$$

Where y_i is the true value for the *i*-th observation, and $\hat{y}_i$ is the predicted value.

3 Research Methodology

Data for the analysis were collected from various sources, including historical data on electricity consumption, weather conditions, and network events. Before model training, the electricity consumption data were cleansed of anomalies and normalised to enhance the quality of model training.

Data from previous years were used to train the Random Forest model. The model was calibrated to maximise forecast accuracy while minimising the risk of overfitting. To evaluate the model's performance on an independent dataset, the data was divided into training and test samples at a ratio of 80% to 20%, respectively; optimal model hyperparameters, such as the number of trees in the forest and the maximum depth of the trees, were identified using cross-validation methods.

After training, the model was tested on a separate dataset to assess its generalisation capability and forecasting accuracy on new data. The model's effectiveness was evaluated using metrics such as Mean Absolute Error (MAE) and the coefficient of determination (R^2).

4 Results and Discussion

4.1 Captions/numbering

The research covers the analysis of electricity consumption data in a small town over one year, which has homogeneous consumer characteristics. The average daily consumption was about 20,000 MWh, fluctuating from 15,000 MWh in the warmer months to 25,000 MWh in the colder months.

Parameters for analysis:

- Daily electricity consumption (in MWh).
- Average daily temperature (°C).
- Average daily humidity (%).
- Average daily wind speed (m/s).

The Random Forest model was configured with 200 trees and a maximum depth of 30.

When testing the model trained on 80% of the dataset on the remaining 20%, we obtained refined results demonstrating the model's accuracy and applicability for forecasting tasks in the power sector. The Mean Absolute Error (MAE) was 198.73 MWh, indicating a minor deviation of the model's predictions from actual data. This means the model can be reliably used to forecast electricity consumption with relatively high accuracy.

The coefficient of determination (R^2), equal to 0.9387, shows that the proposed model can explain about 94% of the variations in electricity consumption based on the analysed data. This high figure underscores the model's efficiency in predicting consumption.

Analysing the importance of various features used in the model, we found that temperature significantly impacts electricity consumption, with an importance of 52.6%. It aligns with expectations, considering that temperature directly affects the use of heating and cooling systems. Humidity also plays a significant role, with an importance of 22.4%, confirming that environmental conditions affect energy consumption. Wind speed was identified as the third most significant factor, with an importance of 25%, indicating the impact of weather conditions on energy consumption.

A specific forecast shows that the model predicted an electricity consumption of 21,347.85 MWh for a day with the conditions: temperature 14.8°C, humidity 68.2%, and wind speed 3.6 m/s. The actual consumption on that day was 21,390.1 MWh, demonstrating the model's high accuracy and applicability for precise electricity consumption forecasting. This provides valuable information for optimising the operation of power systems.

5 Conclusion

The Random Forest model demonstrated high accuracy in forecasting electricity consumption, with a Mean Absolute Error (MAE) of 198.73 MWh and a coefficient of determination (R^2) of 0.9387. These metrics indicate that the model can accurately forecast electricity consumption and efficiently identify potential anomalies in the electrical network. The research findings have significant practical implications for the energy sector. Accurate electricity consumption forecasting allows energy companies to more effectively plan generation and resource distribution, thereby optimising operations and reducing costs. Secondly, the model's ability to detect anomalies can serve as the basis for early warning systems, enhancing the reliability and safety of electrical networks.

One of the proposed solution's main advantages is its versatility and adaptability. The model can easily be adapted to different conditions and requirements of a specific power system due to the flexibility in parameter settings. Moreover, the information on the importance of features provided by the model opens additional opportunities for analysing and improving energy system efficiency.

It should be noted that despite its high efficiency, the Random Forest model has its limitations. In particular, processing large volumes of data may require significant computational resources, which could be an obstacle to real-time implementation in large energy systems. Additionally, the model requires careful tuning of hyperparameters to achieve optimal results, which may necessitate additional time and expertise.

Further research could aim to overcome these limitations, for example, by developing more efficient data processing algorithms or adapting the model for use under limited computational resources. Integrating the Random Forest model with other machine learning methods to create comprehensive solutions capable of further improving forecasting accuracy and anomaly detection efficiency is also a promising direction.

In conclusion, the developed Random Forest model demonstrates impressive results in forecasting electricity consumption and detecting anomalies, offering significant advantages for practical application in the energy sector. However, further research and development are necessary to fully unlock its potential, overcome existing limitations, and improve efficiency.

References

1. M. Islam, M.R. Rashel, M.T. Ahmed, A.K. Islam, M. Tlemçani, Energies, **16**(**21**), 7417 (2023)
2. R., Vaish U.D. Dwivedi, S. Tewari, S.M. Tripathi, Eng. Appl. Artif. Intell., **106**, 104504 (2021)
3. M. Marković, M. Bossart, B.M. Hodge, J. Renew, Sustain. Energ., **15**(**3**) (2023)
4. J. Zhang, K. Du, J. Liu, Y. Wang, W. Zhang, J. J. Yuan, Renew. Sustain. Energ., **15**(**3**) (2023)
5. M. Jaramillo, D. Carrión, J.A. Muñoz, Energies, 2022, **15**(**24**), 9367 (2022)
6. J. Yu, J. Park, S. Kim, Energies, **11**(**11**), 2870 (2018)
7. W. Hu, J. Liang, Y. Jin, F. Wu, X. Wang, E. Chen, Energies, **11**(**11**), 3238 (2018)
8. M. Kim, W. Choi, Y. Jeon, L. Liu, Energies, **12**(**5**), 931 (2019)
9. M. Kim, K. Kim, H. Choi, S. Lee, H. Kim, Energies, **12**(**6**), 1098 (2019)
10. D. Kodaira, J. Park, S. Kim, S. Han, S. Han, Energies, **12**(**6**), 1167 (2019)
11. J. Im, H. Kwon, S. Jeon, M. Lee, Energies, **12**(**7**), 1237 (2019)
12. H. Cha, S. Lee, D. Won, Energies, **12**(**7**), 1339 (2019)
13. R., Ko D. Kang, S. Joo, Energies, **12**(**8**), 1410 (2019)
14. M. Acquah, S. Han, Energies, **12**(**8**), 1436 (2019)
15. C. Shin, S. Rho, H. Lee, W. Rhee, Energies, **12**(**9**), 1696 (2019)
16. M. Krenn, L. Buffoni, B. Coutinho, S. Eppel, J.G. Foster, A. Gritsevskiy, M. Kopp, Nat. Mach. Intell., **5**(**11**), 1326-1335 (2023)
17. Z. Cui, J. Wu, W. Lian, Y.G. Wang, Energy Rep., **9**, 1887-1895 (2023)
18. G. Yörük, U. Bac, F. Yerlikaya-Özkurt, K.D. Ünlü, Mathematics, **11**(**8**), 1865 (2023)
19. D. Rangel-Martinez, K.D.P. Nigam, L.A. Ricardez-Sandoval, Chem. Eng. Res. Des., **174**, 414-441 (2021)
20. H. Husin, M. Zaki, Prot. Control Mod. Power Syst., **6**(**1**), 1-18 (2021)

21. K. Ullah, A. Basit, Z. Ullah, S. Aslam, H. Herodotou, Energies, **14**(**9**), 2376 (2021)
22. M. Ghiasi, T. Niknam, Z. Wang, M. Mehrandezh, M. Dehghani, N. Ghadimi, Electr. Power Syst. Res., **215**, 108975 (2023)
23. A.Q. Al-Shetwi, Sci. Total Environ, **822**, 153645 (2022)
24. M. Zekić-Sušac, A. Has, M. Knežević, Neurocomputing, **439**, 223-233 (2021)
25. M. Saarela, S. Jauhiainen, SN Appl. Sci., **3**(**2**), 272 (2021)
26. S. Talukdar, K.U. Eibek, S. Akhter, S.K. Ziaul, A.R.M.T. Islam, J. Mallick, Ecol. Indic., **126**, 107612 (2021)