Grenze International Journal of Engineering and Technology, January Issue

Comparative-Analysis of Machine Learning Algorithms for Renewable Energy Forecasting

Anuj Tanwar[1], Anik Kumar[2], Anand Saroj[3] and Prof. Narendra Kumar[4]
[1-4]Delhi Technological University/ Department of Electrical Engineering, New Delhi, India
Email: kumartanwar123@gmail.com, {anikk3460, anandsaroj550}@gmail.com, dnk_1963@yahoo.com

***Abstract*— Recent emphasis on climate change and successive COP meetings have led global attention on renewable energy particularly solar and wind as a pivotal solution especially with many countries committing to complete clean energy by 2070. As we see increase in synchronization of renewable energy sources (RES) into the electricity grid, accurate forecasting and predictive modeling play pivotal roles in optimizing the utilization of sustainable energy sources. In this paper, we present the methodologies, results, conclusions and future scope and deploy Deep learning models instead of techniques, offering enhanced accuracy and computational efficiency utilizing datasets from authenticated sources, the study incorporates parameters like time, dew point, humidity, temperature, wind-speed, cloud cover etc. Furthermore, this paper explores different deep learning techniques like Long-Short term memory, Recurrent Neural networks, Support Vector Machines and asses their algorithmic efficacy using performance metrics like Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Square Error (MSE).**

***Index Terms*— Renewable energy, Machine Learning, Forecasting, Grid Management, Deep Learning and Sustainability.**

I. INTRODUCTION

As a result of the increasing demand for sustainable and pure energy, renewable energy research and development have garnered substantial attention in recent years [1,2]. Renewable energy is vital in the effort to reduce greenhouse gas emissions and slow down climate change, and the world is going in that direction. Furthermore, the utilization of renewable energy sources (RES) has several benefits such as decreased reliance on foreign energy sources, generation of employment opportunities, and the possibility of cost savings [1, 6]. The broad use of renewable energy is, however, severely hampered by the inherent unpredictability and fluctuation of RES [7, 8]. For instance, the weather has a significant impact on the production of wind energy and can be unpredictable and challenging to anticipate with accuracy [9]. Similar to this, variables like cloud cover and variations in sunlight throughout the year affect the production of solar energy. It is difficult to effectively integrate renewable energy sources (RES) into the electrical grid due to their significant variability and unpredictability.

One way to tackle this problem is to create precise forecasting models for the production of renewable energy. The negative effects of renewable energy generation's unpredictability and uncertainty on the electricity system can be reduced with the aid of precise forecasting models.

Grenze ID: 01.GIJET.11.1.401

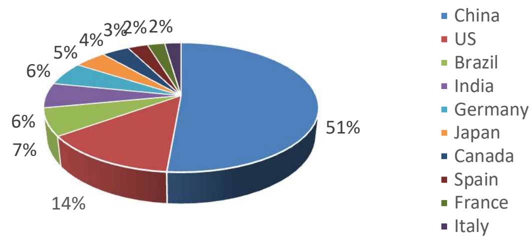

Fig.1 Top 10 countries cumulative Renewable Energy (GW)

TABLE I. ML ALGORITHMS AND THEIR PERFORMANCE

S.No.	Performance (Mean Absolute Error)		
	Algorithms	Wind Energy	Solar Energy
1	Convolutional Neural Networks + Long Short-Term Memory	1979.94	1911.23
2	Long Short-Term Memory	2140.62	11498.99
3	Recurrent Neural Networks	1850.53	552.235
4	Support Vector Machines+ Radial Basis Function kernel	1748.84	496.98
5	Random Forest Regression	1748.42	496.14

Statistical models like the autoregressive integrated moving average (ARIMA) method have certain limitations. They often struggle to manage complex, nonlinear relationships and the high-dimensional characteristics typical of renewable energy data.

In this paper, we delve into advanced algorithms such as Long Short-Term Memory (LSTM), Recurrent Neural Networks (RNN), Convolutional Neural Networks (CNN), and Random Forest Regression (RFR) for forecasting energy data. We then conduct a comparative analysis of these methods by evaluating their regression metrics and comparing their R-squared (R^2) values to determine their performance.

II. ALGORITHMS

A. Support Vector Machines

Support vector machine (SVM) is a supervised learning model that analyses data and identifies patterns, mainly used in classification and regression analysis. The main goal of SVM is to find the best hyperplane that best separates different classes of data points in hyperspace. The ideal hyperplane is the one that allows separation of classes. They are important in defining the position and orientation of the hyperplane. SVM aims to complement these flowers to provide maximum separation between classes. Distribution. Kernels include linear kernels, polynomial kernels, radial basis functions (RBF), and sigmoid kernels**.**

B. Random Forest Regression

Random Forest (RF) is a type of supervised machine learning algorithm which is based on ensemble learning. Ensemble learning combines multiple algorithms, either of the same or different types, to create a stronger model. RF specifically uses a combination of multiple Decision Trees (DT), forming a "forest," hence the name Random Forest. In this, each tree is built using a random subset of features and samples. The two key parameters needed for RF are the number of trees in the forest and the minimum number of observations required per node. Unlike some other algorithms, RF does not rely on cross-validation to estimate its parameters. Instead, it uses an out-of-bag (OOB) estimation method, which assesses performance using samples not included in the training set.

C. Neural Networks

Neural networks are computational models inspired by the human brain, designed to recognize patterns and learn from data. They consist of layers of interconnected nodes, or neurons, which process input data and generate output. The basic types include feedforward neural networks, where connections do not form cycles, and recurrent neural networks (RNNs), which handle sequential data through feedback connections. Variants like convolutional neural networks (CNNs) excel in image processing by detecting spatial hierarchies, while long short-term memory (LSTM) networks and gated recurrent units (GRUs) address the vanishing gradient problem in RNNs, improving their ability to learn long-term dependencies. Neural networks are foundational in deep learning, enabling advancements in image and speech recognition, natural language processing, and autonomous systems. They learn through backpropagation.

D. Long Short-Term Memory

Long Short-Term Memory (LSTM) is a type of Recurrent Neural Network that can learn long-term dependent information. They were introduced by Hochreiter & Schmid Huber (1997). They not only have the memory function of historical information but also overcome the long-term dependence of the model and can selectively forget the invalid information and update the effective information, thus solving the problem of gradient dispersion to some extent. LSTM is composed of the input layer, an output layer and several recursive hiding layers between them. LSTM network can learn to recognize an important input, store it in the long-term state, learn to preserve it for as long as it is needed, and learn to extract it whenever it is needed. LSTM networks are widely used for Natural Language Processing (NLP), Time Series Detection, and Anomaly Detection due to their capability to capture temporal dependencies.[10][11]

III. Performance Indicators

The efficacy of trained machine learning models is determined by the analysis of the discrepancy between predicted and observed labels using a variety of metrics. This is accomplished through the use of holdout data. The following indicators are used to reduce the prediction performance of the proposed model: R-Squared (R^2), mean absolute error (MAE), root mean squared error (RMSE), and mean absolute scaled error (MASE).

A. Root Mean Square Error

Root Mean Square Error (RMSE) is a standard way to measure the error of a model in predicting quantitative data.

$$\text{RMSE} = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} (\hat{y} - y)^2}$$

B. Mean Absolute Error

Mean absolute error gives the mean of the absolute difference between algorithm’s prediction and required value. The MAE is determined by using the following formula:

$$\text{MAE} = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} |y - \hat{y}|}$$

C. Mean Absolute Scaled Error

The MASE, or Mean Absolute Scaled Error, is a statistical metric that is used to assess one's accuracy.
It quantifies the mean absolute error of the data in comparison to the mean absolute error of a naive forecast, which is primarily derived from a straightforward method such as the historical average.

$$\text{MASE} = \frac{1}{n} \cdot \sum_{i=1}^{n} \left| \frac{y - \hat{y}}{MAE} \right|$$

D. R-Squared Error

R-Squared is a measure that shows how well a regression model fits data. Since $\boldsymbol{X}$-square has an ideal value of 1, the closer $\boldsymbol{U}$-square is to 1 the better the model fits the data.

$$R^2 = 1 - \frac{\sum (y - \hat{y})^2}{\sum (y - y')^2}$$

IV. Methodology

In this section different ML classification algorithm’s result and the dataset on which they are applied are presented with graphs and theory and numbers. For further knowledge on the classification model, an interested reader can go through [1].

A. Dataset description

The dataset used is taken from Open Power System Data [8], a free-of-charge platform with data regarding renewable generation capacity of Germany, in time series format and weather data that persisted during that time

duration. It represent the generation of renewable energy between 1st January 2016 to 1st January 2017 and another dataset that has the climatic information for the same.
Time series has wind and solar production capacity in hourly interval.
Here the wind data is trimmed down as in accordance to the cut-in and cut-off speed of as an average 3metres/second and 25metres/second respectively of the turbine.
Weather data with wind speed, radiation, temperature and other measurements
The Weather data contains the following parameters:

- Wind
  - v1: speed [m/s] At height h1 (two meters above the height of displacement)
  - v2: speed [m/s] @ height h2 (10 meters above the height of displacement)
  - v_50m: speed [m/s] @ 50 meters above sea level
  - h1: height above ground [m] (h1 = displacement height +2m)
  - h2: altitude above the surface [m] (h2 = height of displacement + 10 m)
  - z0: length for roughness [m]
- solar parameters:
  - SWTDN: total horizontal radiation at the top of the atmosphere [W/m2].
  - SWGDN: total ground horizontal radiation [W/m²]
- temperature data
  - T: Temperature [K] at a height of two meters above the displacement (refer to h1)
- air data
  - Rho: air density [kg/m³] @ surface
  - p: air pressure [Pa] @ surface

B. Algorithms performance for Solar and Wind Energy

In this sub-section we have taken the algorithms namely Random Forest Regression, Long- Short Term Memory, Recurrent Neural Network, Convolutional Neural Networks and Support Vector Machines, and have trained them across the dataset.
Having done that, we find out their Regression metrics (particularly Root-mean Squared error, Mean Squared Error and Mean Absolute error) to find their efficiency and compare them each with each other for Solar energy as well as Wind energy Fig.3 & Fig.4 respectively.
Here, the graph and the values show that Random Forest algorithm is better suited for solar energy prediction as well as for wind energy forecasting when compared to the rest.

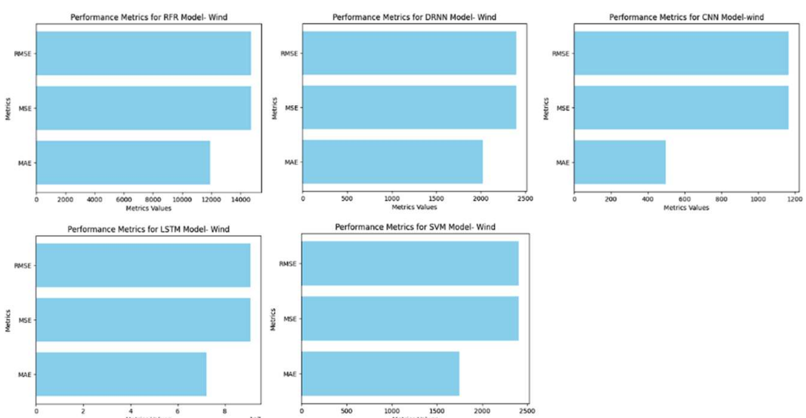

Fig. 3 Performance analysis of algorithms for Wind Energy

V. RESULT

THE MEAN Absolute Error (wind= 1748.421168, solar= 518.724614) and Root Mean Squared Error (wind= 2433.214386, solar= 1201.170098) values show that the model's average errors are low while the Mean Squared Error (wind=49.3276, solar= 34.657901) shows that the model is effective in reducing larger errors. The Root Mean Squared Logarithmic Error (wind= 3.898484, solar= 3.545526) shows the model's ability to handle wind

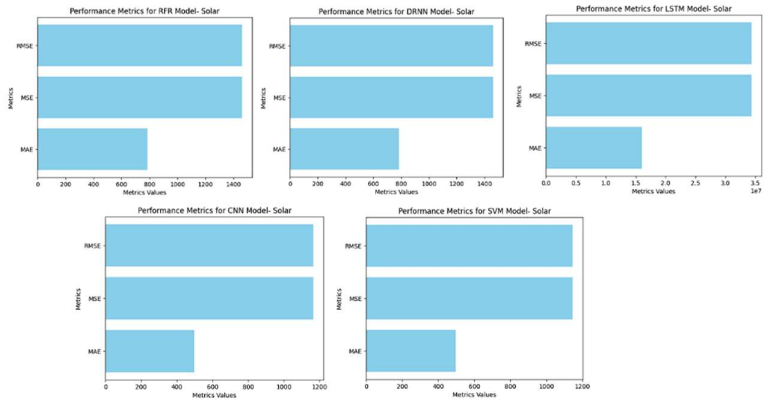

Fig. 4 Performance analysis of algorithms for solar energy

and solar metrics. The most crucial point is that the high R-Squared (wind= 0.895874, solar= 0.913412) value indicates that the model fits the data very well, accounting for the substantial amount of variability in the wind and solar metrics.

This shows that the performance of the Random Forest Regressor model is more accurate for forecasting the wind and solar metrics.

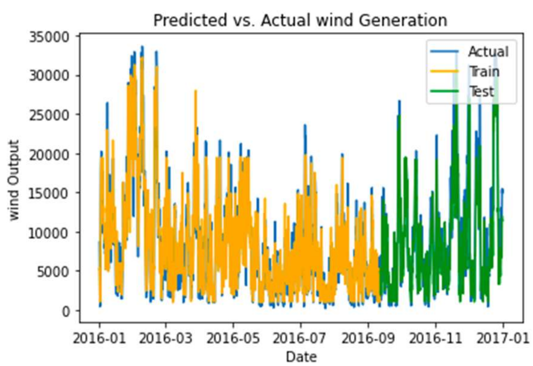

Fig. 5 Random Forest on wind energy

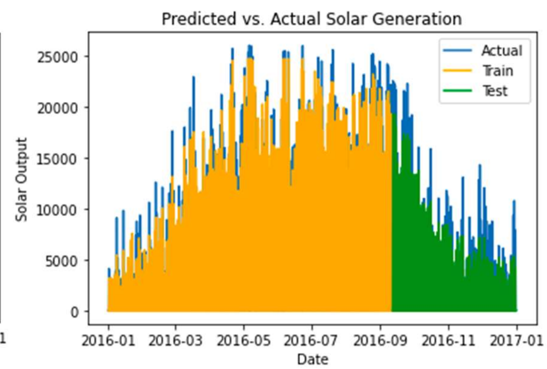

Fig. 6 Random Forest on solar energy

VI. CONCLUSIONS

In this paper, we analyse various machine learning (ML) algorithms and their performance on geophysical (weather and load of renewable energy) datasets to classify the best suitable algorithm for the forecasting of renewable energy (majorly solar and wind). We briefly discussed different ML algorithms and present their performance metrics, including Root Mean Squared Error, mean squared error and Mean Absolute error. Our performance analysis shows that the random forest algorithm achieves the best R^2 score of 0.92 and 0.90 for solar energy and wind energy respectively. Future research can improve ML algorithms and their consistency and hence, energy storage and their direct integration in grid, aiding in the efficient utilization of natural resources for a sustainable future.

REFERENCES

[1] Artificial Intelligence for Renewable Energy Systems by (Ajay Kumar Vyas, Balamurugan, Kamal Kant Hiran and Harsh S. Dhiman).
[2] Forecasting Renewable Energy Generation with Machine learning and Deep Learning: Current Advances and Future Prospects-by (Natei Ermias Benti *, Mesfin Diro Chaka and Addisu Gezahegn Semie *)
[3] Statista.com.
[4] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature.
[5] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.

[6] Optimization of decentralized renewable energy system by weather forecasting and deep machine learning techniques- 2016 IEEE Innovative Smart Grid Technologies - Asia (ISGT-Asia) Melbourne, Australia, Nov 28 - Dec 1, 2016.
[7] Revie paper on Renewable Energy Forecasting and Predictive Modelling by Author Hubert K (2015)
[8] Comparative Analysis of Different Machine Learning Algorithms in Classification of Suitability of Renewable Energy Resource by Aamir Shahab and M.P. Singh (2019)
[9] Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd Edition Queue by Aurélien Géron.
[10] Predicting density and velocity ahead of the bit with zero-offset VSP using deep learning by Vladimir Kazei, Aleksei Titov, Weichang Li.
[11] A Novel Deep Learning Approach for Wind Power Forecasting Based on WD-LSTM Model by Bingchun Liu 1,Shijie Zhao 1,Xiaogang Yu et.