Weather Parameters Forecasting and Modeling Comparison between ANN-MLP and Random Forest

1. Introduction:

As per the literature reviewed the deep learning algorithms and artificial intelligence techniques have been used in several diverse areas widely in recent years. In our research case study, it will compare the forecasting and modeling between Random Forest **(RF)** and Artificial Neural Networking Multilayer Perceptron **(ANN-MLP)**. Random forest machine learning is one of the most popular of these algorithms. Random forests **(RF)** are a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest by Breiman 2001[1]. A research case study for weather forecast exhibits that Artificial Neural Networking Multilayer Perceptron **(ANN-MLP)** model is more significantly forecasted the data, which is reliable and near to the current values as compare to the Auto Regressive Integrated Moving Average **(ARIMA)** and Exponential Smoothing Algorithm **(ETS)** models by Dr Bushra, Zubair, Zara 2019[7].

2. Literature Review:

Several literatures on the Random Forest **(RF)** algorithm are being reviewed. Initially, methods of selection of the random subset and split the training data set by Breiman 2001[1], and the impact of success prediction of the weather forecasting **RF** is numerously used in research papers, such as Remote detection and diagnosis of thunderstorm turbulence by Williams 2008[2], Automated approach for classifying storm type from weather radar reflectivity using decision trees by Gagne, 2009 [3], **RF** is also applied to diagnose aviation turbulence by Williams, J., 2014 [4], Using the **RF** algorithm to decrease the false alarm ratio **(FAR)**, while minimally impacting the probability of detection **(POD)** by Clark, 2015[5], The **RF** shows the remarkable ability to detect the mesoscale convective systems **(MCS-I)** events by Ahijevych 2016 [6], Khan, Muhammed & Shamshad, Bushra & Omar, Zara. (2019). Modeling and Forecasting Weather Parameters using **ANN-MLP**, **ARIMA** and **ETS** model: A case study for Lahore, Pakistan [7], Cafer Mert Yeşilkanat, Spatio-temporal estimation of the daily cases of COVID-19 in worldwide using random forest machine learning algorithm, Chaos, Solitons & Fractals, Volume 140, 2020 [8].

3. Research Object:

The object of the research is to perform the assessment between **ANN-MLP** and **RF** models for forecasting the essential weather parameters of Madrid, Spain.

4. Data Resource:

In this research we are using the monthly average five main weather parameters of Madrid, Spain. The data is measured for eighteen-year period (from 1997 to 2015), obtained from the website https://www.kaggle.com/juliansimon/weather_madrid_lemd_1997_2015.csv.

1. Temperature.
2. Dew Point.
3. Humidity.
4. Sea level Pressure Pa.
5. Wind Speed Km/h.

5. Methodologies:

5.1 Random Forest (RF) Method and Steps

The random forest is a machine learning algorithm with many decision trees. It is a combination of **Bagging** and **Random Subspaces** methods. This method has proved its success in both regression and classification problems in recent years and is one of the best machine learning algorithms used in many different fields, Because of its ability to randomly obtain training data from subsets and form trees with random algorithms, the RF algorithm is better than other machine learning algorithms. Furthermore, the random forest algorithm retains the overfitting amount while testing is done by boot-strap sampling on randomly chosen various sub-datasets. Initially step is to divide the data set randomly into two parts training data set and testing data set. Later, many decision trees are randomly created with (boot strap samples) samples from the data set. Finally, **RF** estimate the average of all results from all trees by Cafer 2020 [8].

5.2 Automatic Artificial Neural Network Multilayers Perceptron (ANN-MLP) Method

The first is to decide the design of the networks. Secondly, to find the best properties for the association loads (determination of the training algorithm). In the process of defining the size of the network, an insufficient number of hidden nodes create difficulties in learning data by Zealand et al., 1999 [9]. Initially we divide the data set into three sets training set, validation set and testing set. Secondly, define the model structure input layer and parameters setting hidden nodes, furthermore train MLP by the hidden nodes and then train the model and validation an optimal or suitable parameter set in the validation set. Finally, simulate the prediction model by using the testing set and calculate the performance level based on the value to verify the trained model by D. C. Lo, Chih [10].

5.3 Forecast Performance

Three different forecast performance indicators are used to evaluate the model performance, such as Mean Error **(ME)**, Mean Absolute Error **(MAE)**, and Root Mean Squared Error **(RMSE).**

Results:

In this section, presenting the major findings of **(ANN-MLP)** and **(RF)** models for forecasting the key weather parameters, Also the graphically representation of both models and forecasting performance indicator table.

Conclusion:

We can forecast the five main parameters of weather with two different models and the comparison of the forecasting accuracy can be done by the values of the lowest **ME**, **MAE** and **RMSE** of the model.

References:

[1] Breiman, L., 2001: Random forests. Mach. Learn., 45, 5–32, https://doi.org/10.1023/A:1010933404324.

[2] Williams, R. Sharman, J. Craig, and G. Blackburn, 2008: Remote detection and diagnosis of thunderstorm turbulence. Proc. SPIE, 7088, 708804, https://doi.org/10.1117/12.79557.

[3] Gagne, D. J., A. McGovern, and J. Brotzge, 2009: Classification of convective areas using decision trees. J. Atmos. Oceanic Technol., 26, 1341–1353, https://doi.org/10.1175/2008JTECHA1205.1.

[4] Williams, J., 2014: Using random forests to diagnose aviation turbulence.Mach. Learn., 95, 51–70, https://doi.org/10.1007/s10994-013-5346-7.

[5] Clark, A. J., A. MacKenzie, A. McGovern, V. Lakshmanan, and R. Brown, 2015: An automated, multiparameter dryline identification algorithm. Wea. Forecasting, 30, 1781–1794, https://doi.org/10.1175/WAF-D-15-0070.1.

[6] Ahijevych, D., J. O. Pinto, J. K. Williams, and M. Steiner, 2016: Probabilistic forecasts of mesoscale convective system initiation using the random forest data mining technique. Wea. Forecasting, 31, 581–599, https://doi.org/10.1175/WAF-D-15-0113.1.

[7] Khan, Muhammed & Shamshad, Bushra & Omar, Zara. (2019). Modeling and Forecasting Weather Parameters using ANN-MLP, ARIMA and ETS model: A case study for Lahore, Pakistan. https://www.researchgate.net/publication/334163943_Modeling_and_Forecasting_Weather_Parameters_using_ANN-MLP_ARIMA_and_ETS_model_A_case_study_for_Lahore_Pakistan.

[8] Cafer Mert Yeşilkanat, Spatio-temporal estimation of the daily cases of COVID-19 in worldwide using random forest machine learning algorithm, Chaos, Solitons & Fractals, Volume 140, 2020, 110210, ISSN 0960-0779, https://doi.org/10.1016/j.chaos.2020.110210.

[9] Zealand, C. M., Burn, D. H., & Simonovic, S. P. (1999). Short term streamflow forecasting using artificial neural networks. Journal of hydrology, 214(1-4), 32-48.

[10] Lo, D. & Wei, Chih-Chiang & Tsai, En-Ping. (2015). Parameter Automatic Calibration Approach for Neural-Network-Based Cyclonic Precipitation Forecast Models. Water. 7. 3963-3977. 10.3390/w7073963. https://www.researchgate.net/publication/281953313_Parameter_Automatic_Calibration_Approach_for_Neural-Network-Based_Cyclonic_Precipitation_Forecast_Models.