e-ISSN 2320 –7876 www.ijfans.org
Vol.9, Iss.2, Feb 2020
Research Paper **UGC CARE Listed (Group -I) Journal**

AIR TEMPERATURE PREDICTION USING RANDOM FOREST AND PARTICLE SWARM OPTIMIZATION ALGORITHM

[1]YALLA S J V DURGA BHAVANI DEVIKARANI, [2]VENKATA RAJU ATHILI, [3]SIRISHA BALLA, [4]K.AJITA

[1,2,3]Assistant Professor, Department of Computer Science and Engineering, BVC College of Engineering, Rajahmundry, Andhra Pradesh, India

[4]B.Tech Scholar, Department of Computer Science and Engineering, BVC College of Engineering, Rajahmundry, Andhra Pradesh, India

ABSTRACT: The complex numerical climate models pose a big challenge for scientists in weather predictions. Air temperature is an essential climatic component particularly in water resources management and other agro-hydrological/meteorological activities planning. Machine Learning (ML) approaches have been popular in the field of prediction. On the other hand temperature measurement tolerance rises due to topographic and structural variation when this data is generated with respect to temporal measurements at reference stations. This paper presents Air Temperature Prediction using Random Forest and Particle Swarm Optimization algorithm. The prediction capability of combination of two models is evaluated using different time lags input combinations with help of root mean square error (RMSE), the mean absolute error (MAE). The obtained results indicated that the combined model (RF+PSO) is more accurate in temperature forecasting than individual models.

KEYWORDS: Air temperature prediction, Random Forest and Particle Swarm Optimization, RMSE, MAE.

I. INTRODUCTION

Air temperature represents a crucial meteorological variable that affects several meteorological and hydrologic processes at different spatial and temporal scale [1]. In case of water resources management or land evaluation, temperature is usually applied as an input variable to derive other parameters such as degree of soil degradation, vegetation growth and evapotranspiration resulting from moister content in soil and plants. Temperature prediction remembers skill for different disciplines. The prediction of climatic boundaries is fundamental for different applications [2]. Some of them incorporate environment observing, dry season location, serious climate prediction, horticulture, and in energy industry, aeronautics industry, correspondence, contamination dispersal and so forth Exact prediction of climate boundaries is a troublesome errand because of the unique idea of climate. Accurate prediction of the temperature have an essential contribution for the planning of the most suitable agriculture site for crop planting, especially in those regions characterized by high intra- and inter-annual weather conditions variability [3].

In recent years, increasing awareness towards Global warming has drawn attention not only of the scientist but also decision-makers and other related stakeholders [4]. According to different climate models, during this century is excepted a continues increase of the earth surface temperature at the global scale, which in turn may result in a wide range of consequences on ecosystem and humans as well. In these circumstances, in addition to technology advancement, there are essential needs of developing and applying novel models that allow for accurate estimation and which help to address the variability of the air temperature more accurately.

There are three broad categories of solar and wind forecasting models namely physical model, statistical and computational models. Artificial intelligence, machine learning and deep learning methods, such as CNN and LSTM, are widely used in many applications like load forecasting, prediction of solar radiation, prediction of heating energy consumption and image classification and so on. The deep learning neural networks produce good reasonable weather forecasts in spite of having no clear knowledge of atmospheric physics. Incorporation of artificial intelligence; machine learning and deep learning methods in a climate model will give a faster and more accurate model. Machine Learning approach that is being experimented in a wide variety of climate change studies or prediction problems. Regression incorporates numerous strategies to display few variables when the attention is on the connection between a needy variable and at least one free variable. Most usually, regression is used for assumption for the reliant variable given the free variables On the whole cases, the assessment target is a component of the free variables called the regression work [5].

Temperature forecasting means predicting the temperature conditions (Degree of temperature) of a particular given area or location. More importantly, accurate temperature prediction is very important to pursue day-to-day activities. Living and non-living things are somehow dependent on temperature predictions. Even after decades of weather forecasting, one of the major obstacles that temperature forecasting faces is the arbitrary & ill-suited expectations from the nature. Machine learning is the ability of computer to learn without being explicitly programmed. It allows machines to find hidden patterns and insights. In supervised learning, we build a model based on labeled training data. The model is then used for mapping new examples. So, based on the observed temperature patterns from the past, a model can be built and used to predict the temperature. In this Random Forest Particle Swarm Optimization Algorithm is used to train the machine with the previous collected data and for the new input the trained machine will predict the temperature. With this research, weather can be forecasted with greater accuracy, which will be helpful in daily activities.

II. LITERATURE SURVEY

H. S. P. H. S., K. Y. Bae and D. K. Sung et. al. [6], propose a solar power prediction model based on various satellite images and a support vector machine (SVM) learning scheme. The motion vectors of clouds are forecasted by utilizing satellite images of atmospheric motion vectors (AMVs). We analyze 4 years' historical satellite images and utilize them to configure a large number of input and output data sets for the SVM learning. We compare the performance of the proposed SVM-based model, the conventional time-series model, and an artificial neural network (ANN) model in terms of prediction accuracy.

Yan Ren, Xiaomin Zhou, Yanjun Lu, Li Fu, Rui Fang, et. al. [7] presents a measuring method based on Support Vector Machine(SVM), which is used to solve the high temperature measuring problem. As we all known, it is difficult to measure directly in complex industrial environment. Thus, the normal support vector machine(NOR-SVM) is improved, and then a new regression algorithm is proposed. Simulation results demonstrate that the improved algorithm has good nonlinear modeling, generalization ability and predictive ability. What's more, this model needs less Support Vectors(SVs), so it learns more faster.

Kisi and Shiri [8] applied the Adaptive Neuro-Fuzzy Inference System (ANFIS) and Artificial Neural Networks (ANN) algorithm to predict the long-term monthly temperature in an arid and semiarid region, and they found that the ANN model performed better than ANFIS.

Wu Lv, Zhizhong Mao, Mingxing Jia, et. al. [9] presents accurate prediction of molten steel temperature which is important for optimal control of Ladle furnace (LF) process. Under this conception, a novel LF temperature prediction model is constructed based on extreme learning machine (ELM), which is a new learning algorithm for single hidden layer feedforward neural networks (SLFNs). ELM is chose due to its good generalization performance and extremely fast learning speed. Furthermore, online sequential learning is adopted on the sequentially arriving data to correct the ELM based prediction model. We introduce a forgetting factor in this learning scheme for the sake of successfully accommodate to the variation in the production process. The simulation results show that the proposed predictor has a good accuracy and fast sequential learning speed, which ensure its ability for practical application.

Ustaoglu, B.; Cigizoglu, H.K.; Karaca, M. et. al. [10], authors assessed two intelligent algorithms, respectively; Feed-Forward Back Propagation (FFBP) and GRNN to predict several parameters of air temperature, which then were compared to Multiple Linear Regression (MLR) model. The finding of that study showed that all models provide satisfactory results in terms of several performance criteria.

III. AIR TEMPERATURE PREDICTION

The input parameters for building the random forest machine learning model are maximum temperature (Tmax), minimum temperature (Tmin), surface pressure in kPa, percentage relative humidity (RH), months and latitude and longitude. The measured data for this work namely the monthly mean maximum temperature, minimum temperature, daily GSR in MJ/m2 /day and wind speed in m/s for different locations of India were collected from India meteorological department, Pune. The selected influencing input parameters are month, latitude and maximum temperature. For 15 Indian locations hourly global solar radiation data is collected from IMD, Pune. Daily mean and monthly mean global solar radiation data is computed from hourly data set. Monthly mean data available for few Indian locations is utlised to train the solar radiation models to predict the solar radiation for other remote solar potential Indian locations where the solar data is unavailable.

The entire dataset is divided into two sections such as training set (installation sub-data set -70%) and testing set (validation sub-data set -30%). The scikit-learn python machine learning library 'sklearn.ensemble' provides an implementation of machine learning. This library holds the built in function for the random forest regressor and Particle Swarm Optimization (PSO) model. After importing the model the training data must be fitted into the model for the machine to learn about the data and to build the model.

It is applied for both classification and regression applications. For the analysis of large datasets, random forest algorithm is the most suitable algorithm. Due to high-prediction accuracy, this algorithm is most approachable and presents particulars on the importance of variables for classification and regression. One more advantage of the RF algorithm is that all those data points which are not used up in the training while bootstrapping are used for cross-validating the model, thus for cross-validation, more data is not needed, these data points are called out-of-bag points. It is empirically proven that this does not decrease the overall performance of the model.

PSO is a meta heuristic optimization computation technique which optimizes the equation iteratively. Initially, the particles, probable solutions, are randomly "flown"

e-ISSN 2320 –7876 www.ijfans.org
Vol.9, Iss.2, Feb 2020
  UGC CARE Listed (Group -I) Journal

into the sample space and are given random velocities. The fitness is evaluated and the particles are given velocity in the direction of the best fitness particle along with the random velocity. The acceleration given is also weighed randomly and with time-optimized solution is obtained.

RF does not need cross-validation to estimate the parameters because it has a built-in estimation of accuracy. The optimal parameter search is done by analyzing the model performance by varying each parameter individually. This decreases the error considerably, but is certainly neither optimal nor satisfactory. Therefore, to improve the algorithm the parameter selection is done by using PSO technique.

This model was implemented using collecting data that we used in this paper and the results were then compared with that of the hybrid RF-PSO model developed in this study. The performance of the machine learning model is evaluated using the performance metrics Accuracy, RMSE and MSE.

IV. RESULT ANALYSIS

The performance of the proposed hybrid RF-PSO model is evaluated in this section. We compared it with Decision Tree, and Multi Layer Perceptron Neural Network. Each prediction scheme is evaluated using three statistical terms, Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and accuracy score are used in this study.

$$RMSE = \sqrt{\frac{\sum_{I=M+1}^{N} \left(y_i - f(x_i)\right)^2}{N-M}} \ldots\ldots(1)$$

$$MAE = \frac{1}{N-M} \sum_{I=M+1}^{N} |(y_i - f(x_i))| \ldots\ldots(2)$$

Comparative performance analysis of the described Air Temperature Prediction using Random Forest and Particle Swarm Optimization algorithm and other models as Decision Tree, and Multi Layer Perceptron Neural Network is represented in below Table 1. Graphical representation of performance parameters for these models is shown in below Fig. 1. The minimum MAE and RMSE are found to be of RF-PSO scheme.

Table 1: PREDICTIVE ERRORS OF DIFFERENT MODELS

Model	MAE	RMSE	Accuracy
Decision Tree	37.5	74.2	89
Multi Layer Perceptron	51.6	77.8	90
RF-PSO model	22.3	48.1	97

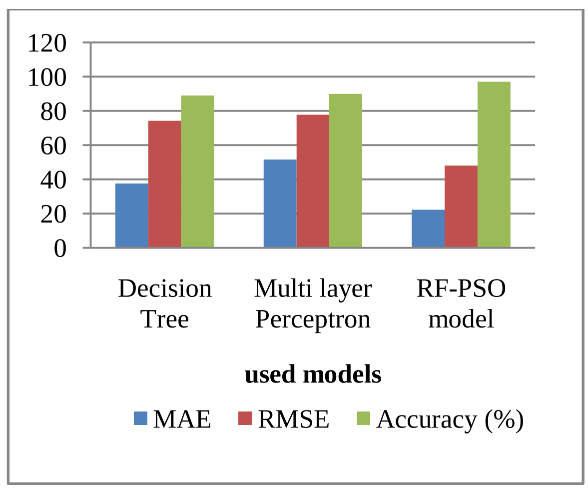

Fig. 1: COMPARATIVE PERFORMANCE ANALYSIS

The predicted air maximum and minimum temperature for collected data is represented in below Fig. 2 and Fig. 3 respectively.

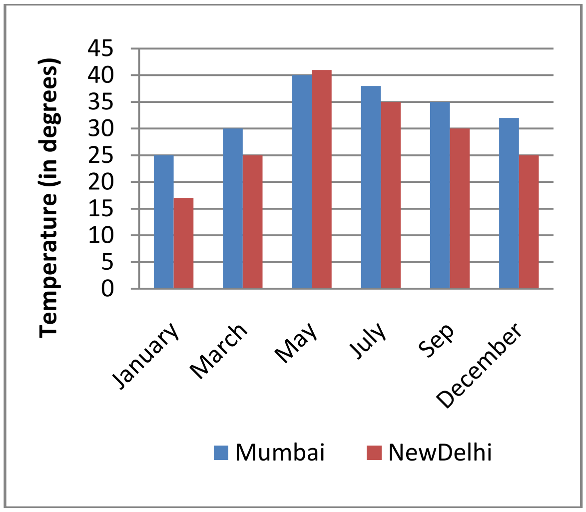

Fig. 2: PREDICTED MAXIMUM TEMPERATURE

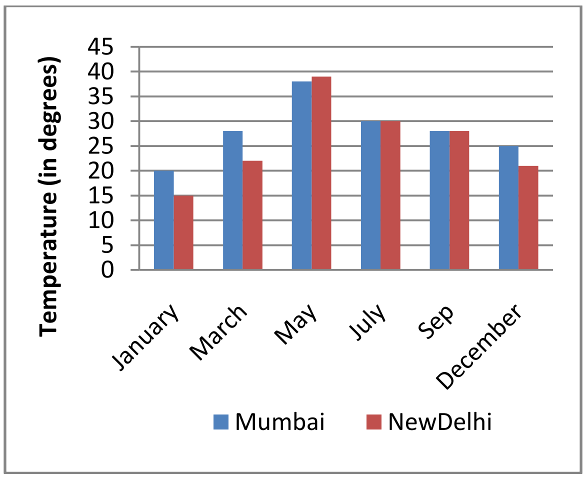

Fig. 3: PREDICTED MINIMUM TEMPERATURE

From comparative results it is clear that described model is predicted air temperature levels accurately than other models. Based on these results, the model with the best accuracy for temperature prediction was RF-PSO hybrid model.

V. CONCLUSION

In this paper, Air Temperature Prediction using Random Forest and Particle Swarm Optimization algorithm is described. Air temperature is an essential climatic component particularly in water resources management and other agro-hydrological/meteorological activities planning. measured data for this work namely the monthly mean maximum temperature, minimum temperature, daily GSR in MJ/m2 /day and wind speed in m/s for different locations of India were collected from India meteorological department, Pune. The entire dataset is divided into two sections such as training set (installation sub-data set -70%) and testing set (validation sub-data set -30%). The performance of the machine learning model is evaluated using the performance metrics Accuracy, RMSE and MSE. Based on these results, the model with the best accuracy for temperature prediction was RF-PSO hybrid model.

VI. REFERENCES

[1] Min Ye Thu, Wunna Htun, Yan Lin Aung, Pyone Ei Ei Shwe, Nay Min Tun, "Smart AIR Quality Monitoring System with LoRaWAN", 2018 IEEE International Conference on Internet of Things and Intelligence System (IOTAIS), Year: 2018

[2] Gui-Mei Cui, Zhao-Guo Jiang, Pi-Liang Liu, Zhi-Hui Chen, Lin Shi, "Prediction of Blast Furnace Temperature Based on Multi-information Fusion of Image and Data", 2018 Chinese Automation Congress (CAC), Year: 2018

[3] Seung Ho Kim, Jong Mun Jeong, Min Tae Hwang, Chang Soon Kang, "Development of an IoT-based atmospheric environment monitoring system", 2017 International Conference on Information and Communication Technology Convergence (ICTC), Year: 2017

[4] V. Natarajan, "Global warming impact and performance enhancement of propane/isobutane mixtures as alternative in air conditioning systems", 2017 Third International Conference on Science Technology Engineering & Management (ICONSTEM), Year: 2017

[5] Yuya Tarutani, Kazuyuki Hashimoto, Go Hasegawa, Yutaka Nakamura, Takumi Tamura, Kazuhiro Matsudax, Morito Matsuoka, "Reducing Power Consumption

in Data Center by Predicting Temperature Distribution and Air Conditioner Efficiency with Machine learning", 2016 IEEE International Conference on Cloud Engineering (IC2E), Year: 2016

[6] H. S. P. H. S., K. Y. Bae and D. K. Sung, "Solar power prediction based on satellite images and support vector machine," IEEE Transactions on Sustainable Energy, vol. 7, no. 3, p. 1255–1263, 2016.

[7] Yan Ren, Xiaomin Zhou, Yanjun Lu, Li Fu, Rui Fang, "Approach to radiation temperature measuring and its application via support vector machine", The 27th Chinese Control and Decision Conference (2015 CCDC), Year: 2015

[8] Kisi, O.; Shiri, J. Prediction of long-term monthly air temperature using geographical inputs. International Journal of Climatology 2014, 34, 179-186, doi:10.1002/joc.3676.

[9] Wu Lv, Zhizhong Mao, Mingxing Jia, "ELM based LF temperature prediction model and its online sequential learning", 2012 24th Chinese Control and Decision Conference (CCDC), Year: 2012

[10] Ustaoglu, B.; Cigizoglu, H.K.; Karaca, M. "Forecast of daily mean, maximum and minimum temperature time series by three artificial neural network methods", Meteorological Applications 2008, 15, 431-445, doi:10.1002/met.83