IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Weather prediction Using Random Forest Method

Aravind M[1],Thilak S[2],Vigneshwaran B[3] ,Dr.J.B.Jona[4]

[1] *Department of Computer Applications, Coimbatore Institute of Technology,*

India

[2] Department of Computer Applications, Coimbatore Institute of Technology,

[3]*Department of Computer Applications, Coimbatore Institute of Technology,*India,

India

[4]Department of Computer Applications, Coimbatore Institute of Technology,

India

Abstract

The advanced numerical climate models cause an enormous challenge for scientists in weather predictions, particularly for tropical system. This paper is concentrated on presenting the importance of weather prediction victimisation machine learning □ML□ technique. The prediction results supported the random forest milliliter model area unit compared with statistical procedure models and SVM milliliter model. Compared to regression models and SVM milliliter model, the prediction results of random forest milliliter model area unit a lot of correct. foretelling is determination of the correct values of weather parameters and furhermore the longer term atmospheric phenomenon supported these parameters. during this study completely different we tend toather parameters were collected then victimisation Random Forest Model we predict the longer term atmospheric phenomenon victimisation temperature knowledge. once coaching Random Forest model victimisation these parameters the prediction of future weather is completed.

Keywords:Climate models,Random forest method,Prediction of future weather.

1. INTRODUCTION

Weather prediction has been a troublesome task within the meteorologic department since years because of varied reason like the drastically unpredictable behaviour of climate. Even with the newest advancement within the technology, it has been troublesome, notwithstanding the prediction are often created the accuracy in prediction of weather may be a questionable issue. Even in current date, this domain remains as a groundwork topic within which scientists and mathematicians area unit operating to supply a model or rule that may accurately predict the weather. presently one in every of the foremost wide used techniques for weather prediction is mining data represent the weather.Data mining is exploration technique has been wide utilized in major fields wherever the prediction of future is very desired like prediction of costs of stocks over a amount of your time in future. Scientists have currently realised that data processing are often used as a tool for weather prediction in addition. the essential entity needs to mine is that the meteorologic knowledge gathered over an explicit amount of your time. knowledge could be a chunk of knowledge that provides details of the climatical condition such wetness, humidity, etc. the info assortment has been happened in numerous places victimisation the various kinds of devices and gets hold on in some kind that can be rendered to extract the prediction results. Weather prediction is typically done victimisation the info gathered by remote sensing satellites. numerous weather parameters like temperature, rainfall, and cloud conditions area unit projected victimisation image taken by meteorologic satellites to access

future trends The accuracy of the prediction wide depends on information of prevailing atmospheric phenomenon over a large space.

2. LITERATURE STUDY

i. Agrawal et al. (1980) Described the phenomenon of a time-series regression model for predicting rice yields in the Ripple area with weekly data using meteorological parameters.

ii. The author, Kuo and Sun (1993), was used to create an intervention model for

iii. predicting and synthesizing currents for an average of 10 days. It has been studied to cause anomalous phenomena caused by typhoons and various severe weather irregularities. Fresh water. Caused the geographical area of Taiwan.

iv. Chiew et al. (1993) A comparison of six precipitation runoff modeling approaches was performed to identify daily, monthly, and yearly runoffs in eight tolerant catchments. They concluded that the timeseries approach agrees with reasonable estimates of monthly and annual yields within the water capital of river basins.

v. Langu, (1993) is another approach that uses statistical analysis to observe changes in meteorological and runoff patterns and look for significant changes in different parts of meteorological statistics. BoxJenkins (1994),

3. Dataset

First, we want some knowledge.

To use a sensible example, retrieved weather knowledge for port of entry, WA from 2016 victimisation the authority Climate knowledge on-line tool. Generally, concerning eightieth of the time spent in knowledge analysis is cleanup and retrieving knowledge, however this employment are often reduced by finding high-quality knowledge sources. The authority tool is amazingly simple to use and temperature knowledge are often downloaded as clean csv files which might be parsed in languages like Python or R.The input parameters for building the random forest machine learning
model area unit

Following area unit explanations of the columns:

year: 2016 for all knowledge points month: range for month of the year temp_2: easy lay temperature a pair of days previous
temp_1: easy lay temperature one day previous
average: historical average easy lay
temperature actual: easy lay temperature measuring friend: your friend's prediction, a random range between twenty below the common and twenty on top of the common

4. METHODOLOGY

4.1 Random Forest Algorithm

In this study, we use a random forest machine learning algorithm to predict temperature. Random forest algorithm is a supervised learning algorithm in machine learning that uses ensemble learning methods for regression. Ensemble learning is a technique that combines predictions from multiple machine learning algorithms to provide more accurate predictions than a single model. The random forest algorithm is the best algorithm for analyzing large amounts of data.

Due to its high predictive accuracy, this algorithm is the most accessible and provides details about the importance of variables for classification and regression. In contrast to ANNs and SVMs, the training method for random forest algorithms is simple.

The main parameter that needs to be adjusted is the number of trees. Artificial Intelligence and Support Vectors Performs a faster training process compared to machinebased models. This algorithm includes a rulebased approach and does not require data normalization.

4.2 Data preparation

Unfortunately, it's not a place where you can feed raw data into your model and return answers (though people are working on it)! We need to make a few small changes to make the data a machine-understandable term. Use the Pandas Python library for data manipulation. It relies on a structure called a data frame. A data frame is basically an Excel spreadsheet that contains rows and columns. The exact procedure for preparing the data depends on the model used and the data collected, but machine learning applications require some data manipulation..

5. IMPLEMENTATION AND RESULTS:

Plot the OF residue (error) to see if the model tends to overestimate or underestimate. You can also check that the residuals are normally distributed.

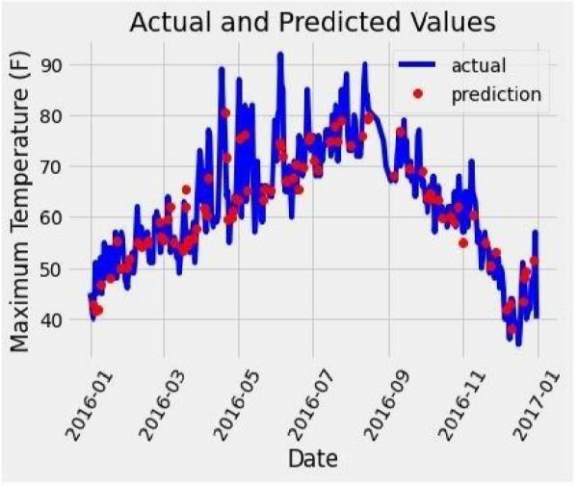

Fig-1(actual and predicted values)

However a final graph is created showing the actual value of , the previous day's temperature,

past averages, and friends' predictions. This is a great way to distinguish between useful variables and useless variables.

```
# Extract the two most important features
important_indices = [feature_list.index('temp_1'), feature_list.index('average')]
train_important = train_features[:, important_indices]
test_important = test_features[:, important_indices]
# Train the random forest
rf_most_important.fit(train_important, train_labels)
# Make predictions and determine the error
predictions = rf_most_important.predict(test_important)
errors = abs(predictions - test_labels)
# Display the performance metrics
print('Mean Absolute Error:', round(np.mean(errors), 2), 'degrees.')
mape = np.mean(100 * (errors / test_labels))
accuracy = 100 - mape
print('Accuracy:', round(accuracy, 2), '%.')
```

```
Mean Absolute Error: 3.92 degrees.
Accuracy: 93.76 %.
```

Fig-2(calculation of accuracy)

To get an idea of the forecast, you can calculate the accuracy using the average average percentage error subtracted from 100%. This model learned and predicted the next day's highest temperature in Seattle, , with an accuracy of 94.

6. CONCLUSION

In this project Random Forest algorithm with dataset comprising of nine weather parameters collected over a period of two years. Though the result of both the algorithms was found to be relatively good as they fall in the category of recommended algorithms for classification and weather prediction problems yet Random Forest proved to be better than the C4.5 Decision Tree.Random Forest is highly recommended over other decision trees when the size of the dataset or the number of parameters in the dataset is high.This project proposes a method to predict the maximum temperature using random forest method to planning weather-based agricultural practices.it can be said that the performance of Random Forest algorithm was is than that of Naïve Bayes in case of dataset dealing with weather. Further improvements can be made to improve the result of the algorithm by applying appropriate filter to the dataset in pre- processing stage.

REFERENCES

[1] Wen-Yeau Chang, "A literature review of wind forecasting methods," *Journal of Power and Energy Engineering*,vol. 2, pp. 161-168, 2014, doi: 10.4236/jpee.2014.24023.

[2] Rahul Sharma and Diksha Singh,"A review of wind power and wind speed forecasting," *Journal of Engineering Research and Application*, vol. 8, pp. 19, 2018.

[3] Saurabh S. Soman, *et al*., "A review of wind power and wind speed forecasting methods with

.....Column Break.....different time horizons," *North American Power Symposium*, pp. 1-8, 26-28 Sep 2010, doi: 10.1109/NAPS.2010.5619586.

[4] R. Meenal, *et al*., "Review on mathematical models for the prediction of global solar radiation," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 15, no. 1, pp. 5459, 2019, doi: 10.11591/ijeecs.v15.i1.pp56-61.

[5] Sani Salisu, *et al*., "A Wavelet Based Solar Radiation Prediction in Nigeria Using Adaptive Neuro-Fuzzy Approach," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 12, no. 3, pp. 907-915,December 2018, doi: 10.11591/ijeecs.v12.i3.pp907-915. [6] MakarandA. *et al*., "Wind speed prediction using statistical regression and neural network," *Journal of Earth System Science*, vol. 117, pp. 457463, 2008, doi: 10.1007/s12040-008-0045-7.

[7] Salisu Muhammad Lawan, *et al*., "Wind speed prediction in non-monitored areas based on topographic radial basis neural network (TRBNN)," *IOP Conference Series: Earth and Environmental Science*, vol. 168, pp. 012012,2018.

[8] Cabral Lima, *et al*., "A model to forecast wind speed through singular spectrum analysis and artifificial neural networks," *International Conference on Artificial Intelligence*, Las Vegas, pp. 235-240, 2017.

[9] R.Meenal, *et al*., "Sensitivity analysis based artificial neural network approach for global solar radiation prediction in India," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 1, pp. 31-38, 2020, doi: 10.11591/ijeecs.v20.i1.pp31-38.

[10] R.Meenal and A.I.Selvakumar, "Review on artificial neural network based solar radiation prediction," *International Conference on Communication and Electronics System*, Coimbatore, pp. 302-305,15-16 Oct 2018, doi: 10.1109 /CESYS. 2017.83212