

Meteorological Data Analysis for Arid Region of Karnataka

Prajwala Talanki¹, Dassapa Ramesh², Venugopal³

¹Department of Computer Science, CMR institute of technology (VTU), Bangalore, Karnataka, India

²Department of Computer Applications, Sri Siddhartha Institute of Technology, Tumkur, Karnataka, India

³Department of Computer Science, Sri Siddhartha Institute of Technology, Tumkur, Karnataka, India

Email address:

prajwalatr@gmail.com (P. Talanki), rameshd_ssit@yahoo.com (D. Ramesh), venussit@gmail.com (Venugopal)

To cite this article:

Prajwala Talanki, Dassapa Ramesh, Venugopal. Meteorological Data Analysis for Arid Region of Karnataka. *International Journal of Data Science and Analysis*. Vol. 6, No. 2, 2020, pp. 69-71. doi: 10.11648/j.ijdsa.20200602.12

Received: May 18, 2020; **Accepted:** May 29, 2020; **Published:** June 15, 2020

Abstract: Meteorological data analysis is one of the time series prediction applications. Analysis of meteorological data give insights to the weather forecast and makes country more prepared for the worst situation like drought and flood. Northern part of Karnataka is usually a drought region. The paper provides insights into application of random forest and decision tree for a region of Karnataka called Raichur. The results of accuracy precision and recall are tabulated for Raichur region. There are 10 input features of climate considered in prediction of rainfall for a region. An accuracy of 96% is obtained after applying random forest to the meteorological data collected from IMD (Indian Meteorological Department). Raichur is an arid region of Karnataka which receives less rainfall. There were 13 input features considered for prediction of rainfall. The data was collected from Indian Meteorological Department (IMD) for a span of 17 years from January 1999 to December 2016 for prediction of rainfall. The decision tree classifier was applied to get an accuracy of 88%. The classification report shows a precision and recall of 0.90 and 0.97. Random forest an ensemble classifier was run through the dataset for an accuracy of 96%. The precision and recall of 1.00 and 0.99 was achieved. For both the algorithms a total of 11159 tuples were considered. There are total 11158 samples. The total training observations are 7810. The total testing samples are 3348. The decision rules are documented. Random forest algorithm shows a relative importance of parameters for Raichur rainfall prediction. A highest importance on rainfall prediction is Wet Bulb Temperature (WBT) and least important factor is Wind direction (FFF).

Keywords: Random Forest Regressor, Gini Index, Classification Report, Decision Tree

1. Introduction

Meteorological data analysis a field growing rapidly in atmospheric science. Large scale meteorological data is not supporting traditional approach. It requires a comprehensive approach to handle the data. One of the key area of climate research is extreme weather conditions like drought and flood. This paper focuses on one such issue which is the drought conditions north Karnataka regions like Raichur, Gulbarga, Bellary, Bagalkot. The paper focuses on prediction of rainfall for the Raichur region of Karnataka. Raichur is one of the drought stuck areas of Karnataka. The prediction shows an accuracy of 96% with 70% training sample and 30% test data set. There are total 10 parameters considered. These 10 features are related to humidity, pressure, temperature used to predict the rainfall of Raichur region. The data is collected from Indian Meteorological department

(IMD) from 1999 to 2016. There are total 11158 samples. The total training observations are 7810. The total testing samples are 3348. A precision and recall is 0.99 and 0.97 respectively.

The Decision Tree algorithm [1] is applied for the Raichur data for prediction of rainfall. Decision tree is a classification learning technique which uses best split to categorize the data into target classes. Random forest is ensemble of trees which uses random sampling of input. It considers the output of different decision tree. Random forest combines output of different decision tree and then classifies the class target value.

2. Methodology

Supervised learning techniques are a model prepared through a training process. The input data called training data

set is run on a model say a classification model to predict the output. If the predicted output is correct it is classified for the correct class label. This process continues until model achieves desired level of accuracy. One such supervised learning technique is classification. Classification has set of input features along with correct output feature. The algorithm learns by comparing the class label and predicted output.

Decision tree is a classification algorithm. It can be used to solve regression problems also. The algorithm goes as follows. It initially considers entire dataset to be root node. The tree is split based on maximum information gain or minimum impurity. This process is repeated until all nodes are pure, which is the terminal condition. [3]

Random forest [5] is a classification algorithm which is an ensemble technique.

Random forest works as follows. It creates many decision trees. The output of each decision tree is given weightage. The output is calculated of different decision tree and weights given to it. The random forest uses random sampling. The advantage of random forest classifier is it won't overfit the data and applicable to categorical data.

Raichur is a region of north Karnataka which has defiant rainfall. Rainfall usually occurs in month of September. The driest months of Raichur are January and February. The data is considered from period of 1999 to 2016. The input features considered are SLP Station Level Pressure (in hpa), MSLP (Mean Sea-Level Pressure) (in hpa), DBT (Dry Bulb Temperature) (in deg. C), WBT (Wet Bulb Temperature)(in deg. C), Dpt (Dew Point Temperature)(in deg. C), RH (Relative Humidity) (In%), VP (Vapour Pressure)(in hpa), DD (Wind Direction) FFF (Wind Speed) (in Kmph), AW (Average Wind Speed) (in kmph). The class label is rainfall in mm. The data is considered for span of 17 years and monthly data is recorded. The missing data values were calculated by taking mean of the previous values.

Decision tree algorithm was used for prediction of rainfall for Raichur region. The data is collected from Indian Meteorological department (IMD) from 1999 to 2016. There are total 11158 samples. The total training observations are 7810. The total testing samples are 3348. A precision and recall is 0.99 and 0.97 respectively. The data was initially considered to be root node. The Gini index is used to find best split of the node. Gini index is defined using the following formula:

$$\text{Gini index (node)} = 1 - \sum_j p_j^2$$

Where P_j is probability that the sample belongs to class j . In this case class label is rainfall.

The best split attribute at first level is vapour pressure. The level 2 of tree is split based on attribute temperature. At level 3 the splitting attribute is Station level Pressure. From the above results we can conclude that temperature, Humidity and station level pressure are the features used for predicting the rainfall in the region of Raichur.

Random forest is applied to Raichur data. The data is collected from Indian Meteorological department (IMD) from 1999 to 2016. There are total 11158 samples. The total

training observations are 7810. The total testing samples are 3348. A precision and recall is 0.99 and 0.98 respectively. The measure for splitting each node is minimum Mean Squared Error (MSE). It is used as loss function for least squared regression. After running random forest algorithm it is found that the features affecting weather of Raichur are RH (Relative Humidity), Vapour pressure, MSLP (Mean Sea-Level Pressure) and DBT (Dry Bulb Temperature). The tree was constructed for depth of level 3.

3. Results and Discussion

The Raichur data was collected for a span of 17 years. There are 10 input features considered for the target label rainfall. The input features considered are SLP Station Level Pressure (in hpa), MSLP (Mean Sea-Level Pressure) (in hpa), DBT (Dry Bulb Temperature) (in deg. C), WBT (Wet Bulb Temperature)(in deg. C), Dpt (Dew Point Temperature)(in deg. C), RH (Relative Humidity) (In%), VP (Vapour Pressure)(in hpa), DD (Wind Direction) FFF (Wind Speed) (in Kmph), AW (Average Wind Speed) (in kmph). The algorithms run are decision tree and random forest Regressor.

The results of accuracy, precision, recall and F1 score are tabulated as follows.

Table 1. Results of algorithm

	Decision tree	Random forest
Accuracy	95%	96%
Precision	0.98	0.99
Recall	0.97	0.98
F1 score	0.98	0.98

The decision tree for Raichur region is as shown in figure 1.

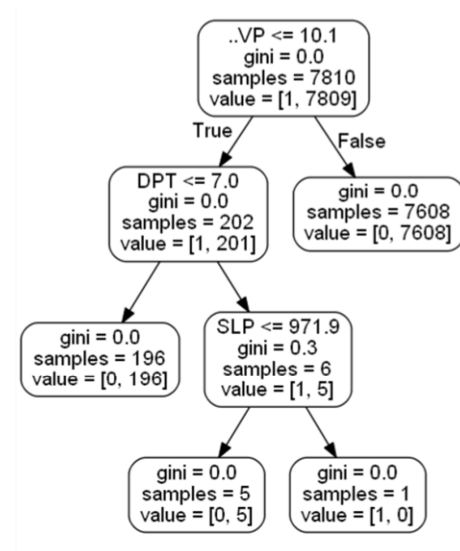


Figure 1. Decision tree for Raichur region.

From the above results it is clear that the factors affecting rainfall of Raichur region are Vapour Pressure, temperature and sea level pressure. The least affecting factors are wind direction and wind speed. The Random forest Regressor

provides a high accuracy of 96%.

4. Conclusion

Meteorological data analysis is a field growing rapidly in atmospheric science. Large scale meteorological data is not supporting traditional approach. It requires a comprehensive approach to handle the data. Decision tree algorithm and Random forest is applied to Raichur data. The data is collected from Indian Meteorological department (IMD) from 1999 to 2016. There are total 11158 samples. The rainfall prediction for drought region of Karnataka is having accuracy of 96%. From the above results it is clear that the factors affecting rainfall of Raichur region are Vapour Pressure, temperature and sea level pressure. The least affecting factors are wind direction and wind speed.

Acknowledgements

I would like to express my gratitude to Indian Meteorological Department (IMD) for providing the data for research purpose

References

- [1] "Deep Learning based architecture for rainfall estimation integrating heterogeneous data sources", Gianluigi Folino; Massimo Guarascio; Francesco Chiaravalloti; Salvatore Gabriele et. al, 2019 International Joint Conference on Neural Networks (IJCNN).
- [2] "Daily Rainfall Data Construction and Application to Weather Prediction" Choujun Zhan; Fujian Wu; Zhengdong Wu; Chi K. Tse et. al, 2019 IEEE International Symposium on Circuits and Systems (ISCAS).
- [3] Rainfall prediction based on 100 year meteorological data", sandeep kumar et. al, 2018 International Conference on Computing and Communication Technologies for Smart Nation (IC3TSN)", feb 2018.
- [4] "Analyze the Rainfall of land slide on Apache Spark" Chouyann-lee et. al IEEE conference on advanced computer Intelligence, march 2018.
- [5] "A Deep Neural Network Approach for Crop Selection and Yield Prediction in Bangladesh", Tanhim Islam et. al, 2018 IEEE Region 10 Humanitarian Technology Conference (R10-HTC).
- [6] "Rainfall Prediction: Accuracy Enhancement Using Machine Learning and Forecasting Techniques", Urmay shah et. al, 2018 Fifth International Conference on Parallel, Distributed and Grid Computing (PDGC).
- [7] "A Method of Rainfall Runoff Forecasting Based on Deep Convolution Neural Networks", Xiaoli Li; Zhenlong Du; Guomei Song,, 2018 Sixth International Conference on Advanced Cloud and Big Data (CBD).
- [8] "Deep learning multilayer perceptron (MLP) for flood prediction model using wireless sensor network based hydrology time series data mining", Indrastanti R. Widiyari et. al, 2017 International Conference on Innovative and Creative Information Technology (ICITech).
- [9] "Rainfall prediction of a maritime state (Kerala), India using SLFN and ELM techniques", Yajnaseni Dash, et. al.; 2017 International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICT) IEEE transactions, April 2018.
- [10] "Early Prediction System Using Neural Network in Kelantan River, Malaysia" Mohd Azrol Syafiee Anuar* et. al, IEEE conference, 2017.
- [11] G. B. Huang, M. B. Li, L. Chen, C. K. Siew, "Incremental extreme learning machine with fully complex hidden nodes," *Neurocomputing*, vol. 71 (x), pp. 576-583, 2008.