

MACHINE LEARNING TECHNIQUES OF WEATHER FORECASTING: A REVIEW

Abstract:

Weather is a particular state of the atmosphere that describes the degrees to which it is hot or cold, wet or dry, calm or stormy, clear or cloudy. On earth, most weather phenomena occurs in the lowest layer of the planet's atmosphere i.e. the troposphere. Weather forecasting tools are used in the field of science and technology to forecast atmospheric conditions for a certain place and period. It is a very challenging task for the researchers of this field in this modern era. In this review paper we have tried to estimate the accuracy level of different machine learning methods of weather forecasting.

Key word: random forest , decision tree , support vector Machine , Naïve Bays , logistic regression , K-nearest neighbors , weather data

Introduction:

Now a days weather and climate of this planet is changing very drastically and continuously as a result of the climate change. Earth is suffering very badly due to this continuous climate change. It is very crucial and essential task to predict the exact current weather conditions. Presently machine learning techniques are widely used to forecast the weather. These techniques depend on complex physical models and uses of the dataset of some previous years having the weather parameters, i.e. temperature, dew, humidity, pressure, visibility, wind speed and wind direction also. The terms of forecasting weather in Machine learning methods may be 'No rain', 'Fog', 'Rain', 'Thunderstorms', 'Fog Rain', 'Tropical Cyclones', 'Heat waves', 'Cold snaps', 'Heavy downpours', 'Floods' etc. In this research process the predicted outcomes are continuous numeric values temperature.

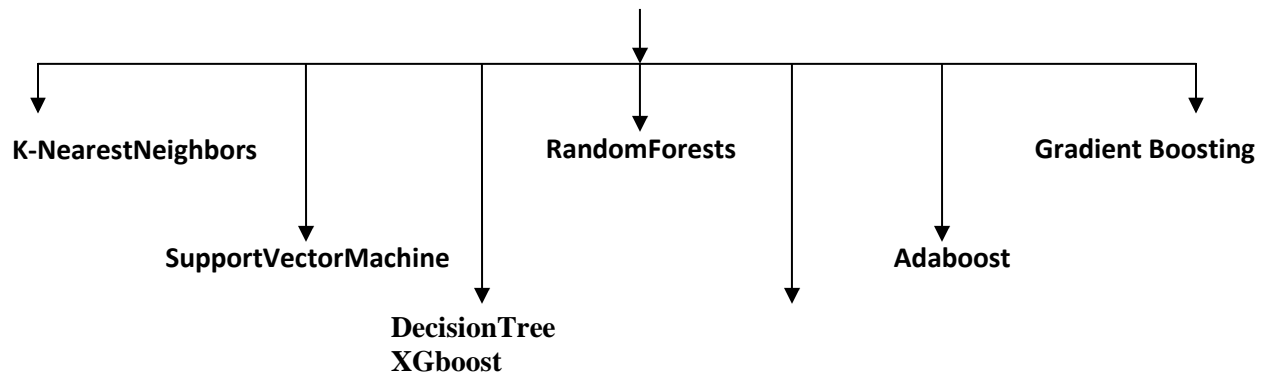
The different Machine Learning Methods which is used in this technique are random forest, decision tree, support vector Machine, KNN, Adaboost, Xgboost, Gradient Boosting, Naïve Bays and logistic regression etc. The Models are evaluated in the basis of their accuracy in prediction and verification process. The used techniques of Machine learning method for weather forecasting is analysed in respect of their accuracy level to get an overall idea and its applications.

Methodology and Techniques:

Here we aim to use Machine Learning techniques to predict the weather day by day and hourly. We use the techniques namely Chi square test and Naïve Bayes statistics applied on the dataset to examine the useful information from that dataset. First of all here we collect the data by user information. Then we process the data by transformation and combination and prepare a database. By Data Mining Techniques the data is stored. At last by making Decision Tree, here we process the dataset and give the dataset for weather prediction.

The techniques that are used in this machine learning process depicted below in the chart and described briefly to understand the process

Techniques Used



(i) K-nearestneighbors(KNN)

KNN method may be used for both regression and classification. Mainly it is utilized for classification. The right class of testing data is being predicted by KNN of a given set of data of various classes by calculating the distance between both the testing data and all training points. After that K point is chosen which are most similar to the test.

(i) SupportVectorMachine

This technique is used for the predictive data processing. Also for the purpose of classification the support vector machine is utilized.

(ii) DecisionTree

Among the effective techniques of machine learning, decision tree is the most effective and best used techniques. An item or scenario with a set of attributes is used as an input and provides a 'yes/no' decision in the decision tree method.

(iii) RandomForests

A collection of decision trees is called as Random forests strategy which generates more tree variety.

(iv) XGboost

This method is also known as "Extreme Gradient Boosting." To solve a number of data science problems more precisely XGboost uses parallel tree boosting technique.

(v) Adaboost

Adaboost a statistical categorization method which is used for the improvement of the results and is combined with a number of learning algorithms.

(vi) Gradient Boosting

Among the other boosting techniques, Gradient Boosting is the prominent technique which corrects the former mistakes.

Analysis of Results:

The different models have been used here to predict the weather data of 21 years (1996-2017). For a particular given region, the data of the accuracy of the different models is summarized in Table 1 and its percentage of accuracy is shown in fig. 1 with graphical representation. Based upon the F1 score of the models used, the performance of all the models is summarized in Table 2 and represented in the fig. 2

Table 1: Dataset used in the experiment

SL	Temp.	Dew	Humidity	Pressure	Visibility	Wind	Event
01	28	24	76	1002	05	11	NoRain
02	29	26	85	1003	05	08	NoRain
03	32	26	78	1004	05	11	NoRain
04	31	26	81	1003	04	13	NoRain
05	31	26	86	1001	04	10	Rain, thunderstorm

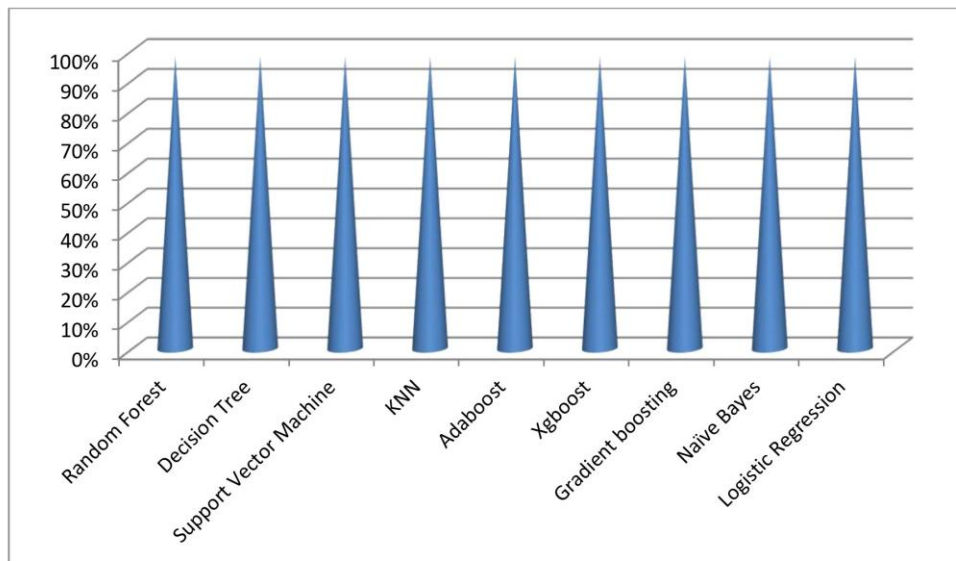


Fig 1: Graphical comparison of the accuracy of different models

Table2: F1Scoreofdifferentmodels

SI No	Modelname	F1score
1	Randomforest	0.81
2	DecisionTree	0.75
3	SupportVectorMachine	0.74
4	KNN	0.80
5	Adaboost	0.74
6	Xgboost	0.80
7	Gradientboosting	0.83
8	NaïveBayes	0.79
9	Logisticregression	0.80

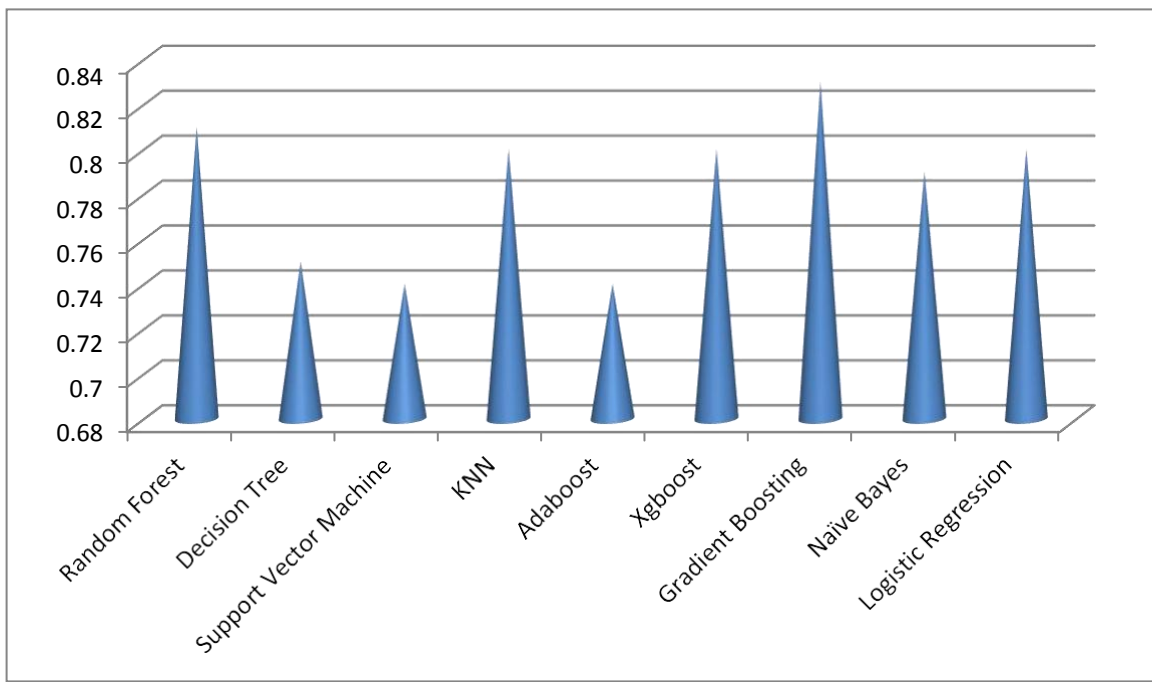


Fig. 2 The graphical comparison of F1 score of different models.

A maximum accuracy upto 81.67% observed in these models which needs dataset from predefined resources. Highest accuracy level is observed in Gradient Boosting method. It is evident to use low cost internet of things (IoT) devices in near future such as temperature and humidity sensors to collect the weather data from different parts of a place or an area and to be fed in the machine learning methods.

Conclusions:

Machine learning technology can easily be run on almost all the computers including ML devices. Traditional physical models are more resource oriented than machine learning technology. It can provide intelligent models and are much more simpler in comparison to machine learning technique. ML technique can predict weather parameters more precisely and accurately than the traditional models.

References:

1. Singh N, Chaturvedi S, Akhter S (2019) Weather forecasting using machine learning algorithm. In: International conference on signal processing and communication (ICSC), pp 171-174. <https://doi.org/10.1109/ICSC45622.2019.8938211>
2. Khajure S, Mohod SW (2016) Future weather forecasting using soft computing techniques. *Proc Comput Sci* 78:402-407. ISSN 1877-0509. [h.aps://doi.org/10.1016/j.procs.2016.02.081](https://doi.org/10.1016/j.procs.2016.02.081)
3. Bhardwaj R, Duhoon V (2018) Weather forecasting using soft computing techniques. In: International conference on computing, power and communication technologies (GUCON), pp 1111-1115. [h.aps://doi.org/10.1109/GUCON.2018.8675088](https://doi.org/10.1109/GUCON.2018.8675088)
4. Haghbin M, Sharafati A, Motta D (2021) Applications of soft computing models for predicting sea surface temperature: a comprehensive review and assessment. *Prog Earth Planet Sci* 8(4). <https://doi.org/10.1186/S40645-020-00400-9>
5. Vathsala H, Koolagudi SG (2021) Neuro-fuzzy model for quantified rainfall prediction using data mining and soft computing approaches. *IETE Res*. <https://doi.org/10.1080/03772063.2021.1912648>
6. Balogh B, Saint-Martin D, Ribes A (2021) A toy model to investigate stability of AT-based dynamical systems. *Geophys Res Lett* 48(8). <https://doi.org/10.1029/2020GL092133>
7. Jayasingh SK, Gountia D, Samal N, Chinara PK (2021) A novel approach for data classification using neural network. *IETE J Res*. <https://doi.org/10.1080/03772063.2021.1986152>
8. Litta AJ, Idicula SM, Mohanty UC (2013) Artificial neural network model in prediction of meteorological parameters during premonsoon thunderstorms. *Int J Atmos Sci* 2013:14. Article ID 525383. <https://doi.org/10.1155/2013/525383>
9. Schultz MG, Betancourt C, Gong B, Kleinert F, Langguth M, Leufen LH, Mozaffari A, Stadler S (2021) Can deep learning beat numerical weather prediction. *Phil Trans R Soc A* 379:20200097. <https://doi.org/10.1098/rsta.2020.0097>
10. Sharma A, Agarwal S (2012) Temperature prediction using wavelet neural network. *Res J Inf Technol* 4:22-30. <https://doi.org/10.3923/rjit.2012.22.30>. <https://scialert.net/abstract/?doi=rjit.2012.22.30>
11. Lin S, Wang G, Zhang S, Li J (2006) Time series prediction based on support vector regression. *Inf Technol J* 5:353-357. <https://doi.org/10.3923/itj.2006.353.357>. <https://scialert.net/abstract/?doi=itj.2006.353.357>
12. Askari M, Askari H (2011) Time series grey system prediction-based models: gold price forecasting. *Trends Appl Sci Res* 6:1287-1292. <https://doi.org/10.3923/tasr.2011.1287.1292>. <https://scialert.net/abstract/?doi=tasr.2011.1287.1292>
13. Lee J, Lee J (2016) Constructing efficient regional hazardous weather prediction model through big data analysis. *IJFIS* 16:1-12. [h.aps://doi.org/10.5391/15TS.2016.16.1.1](https://doi.org/10.5391/15TS.2016.16.1.1)
14. Jayasingh SK, Mantri JK, Pradhan S (2021) Weather prediction using hybrid soft computing models. In: Udgata SK, Sethi S, Srirama SN (eds) *Intelligent systems. Lecture notes in networks and systems*, vol 185. Springer, Singapore. [h.aps://doi.org/10.1007/978-981-33-6081-5_4](https://doi.org/10.1007/978-981-33-6081-5_4)
15. Sofian IM, Affandi AK, Iskandar I, Apriani Y (2018) Monthly rainfall prediction based on artificial neural networks with backpropagation and radial basis function. *Int J Adv Intell Inf* 4(2):154-166. ISSN 2548-3161. <https://doi.org/10.26555/ijain.v4i2.208>