

Weather Forecast through Data Mining

Swati Pandey¹, Shruti Sharma¹, Shubham Kumar¹, Kanchan Bhatt¹, Dr. Rakesh Kumar Arora²

¹Department of Computer Science and Engineering , Krishna Engineering College, Ghaziabad, Uttar Pradesh, India

²Professor, Department of Computer Science and Engineering, Krishna Engineering College, Ghaziabad, Uttar Pradesh, India

ABSTRACT

Article Info

Volume 7, Issue 3

Page Number: 90-95

Publication Issue :

May-June-2021

Article History

Accepted : 02 May 2021

Published : 08 May 2021

Weather Forecasting is the attempt to predict the weather conditions based on parameters such as temperature, wind, humidity and rainfall. These parameters will be considered for experimental analysis to give the desired results. Data used in this project has been collected from various government institution sites. The algorithm used to predict weather includes Neural Networks(NN), Random Forest, Classification and Regression tree (C &RT), Support Vector Machine, K-nearest neighbor. The correlation analysis of the parameters will help in predicting the future values. This web based application we will have its own chat bot where user can directly communicate about their query related to Weather Forecast and can have experience of two-way communication.

Keywords : Weather Forecast, Neural Network, Prediction , Parameters, Analysis, Correlation

I. INTRODUCTION

Weather forecasting is the prediction of what the atmosphere will be like in a particular place by using technology and scientific knowledge to make weather observations. In other words, it's a way of predicting things like cloud cover, rain, snow, wind speed, and temperature before they happen.

Should you wear a raincoat or a sweater today? Should you go to a particular place or not? These are all everyday questions we might only be able to answer with the help of a weather forecast. When dangerous weather conditions like

hurricanes or snowstorms arrive, our choices about what we should or shouldn't do can often mean the difference between life and death.

There is a vast variety of end-users to weather forecasts. Weather warnings are important forecasts because they are used to protect life and property. Forecasts are based on temperature, and therefore to traders within the commodity market. Weather forecasts are made by collecting as much data as possible about the current state of the atmosphere (particularly the temperature, humidity and, wind) and using an understanding of atmospheric processes(through meteorology) to determine how the atmosphere

evolves in the future. Weather forecasting is a vital application in meteorology and has been one of the most scientifically and technologically challenging problems around the world in the last century.

Weather forecasts are made by collecting quantitative data about the current state of the atmosphere at a given place. Once calculated by hand based mainly upon changes in pressure, current weather conditions, and sky condition or cloud cover, weather forecasting now relies on computer-based models that take many atmospheric factors into account. Forecasting weather is one of the hardest and most complicated things scientists do daily. Meteorologists use many tools to tackle the job of weather forecasting. Many starts by looking at images provided by weather satellites. A piece of single satellite information holds tons of information. A meteorologist looking at this image could tell where the mild air is, where the cold and warm storms are, and even identify stormy weather.

Human input is still required to pick the best possible forecast model to base the forecast upon, which involves pattern recognition skills, teleconnections, knowledge of model performance, and knowledge of model basis. The inaccuracy of forecasting is due to the chaotic nature of the atmosphere, the massive computational power required to solve the equations that describe the atmosphere, the error involved in measuring the initial conditions, and an incomplete understanding of atmospheric processes. Hence, forecasts become less accurate as of the difference between the current time and the time for which the forecast is being made increases. The use of ensembles and model consensus helps narrow the error and pick the most likely outcome.

II. LITERATURE SURVEY

On a worldwide scale large numbers of attempts have been made by different researchers to forecast Weather accurately using various techniques but due to nonlinear nature of Weather prediction accuracy obtained by these techniques is still below the satisfactory level.

Sarma , Konwar and Das shows that Artificial Neural Network (ANN) can also combined with different methods . A neural network model for rainfall retrieval over ocean remotely sensed microwave (MW) brightness temperature (BT). They proposed soft computing approach for rainfall prate over ocean using online feature selection , clustering , and hybrid neural network

F. Mekanik and M. A. Imteaz_ found that rainfall is also affected by the key modes of complex climate variables . Few attempts have been made to establish the combined effect of these indices on rainfall in order to develop a better understanding in forecast system.

Artificial neural network have been extensively used in these days various aspect of science and engineering because of its availability to model both linear and nonlinear system without the need to assumption make assumption as are implicit in more traditional statistical approaches . Vamsithar , Rao and Satayapati use the back propagation neural network model for predicting the rainfall based on humidity , dew point and pressure in the country INDIA . Two-Third of the data was used for training and One-Third for testing . The number of training patterns is 250 training and testing patterns are 120 . In the training they obtained 99.79% of accuracy and in Testing 94.28% of accuracy.

These result can predict the rainfall for the future . For rainfall prediction, Artificial Neural Network was applied and the rainfall was predicted in INDIA . According to the result back propagation neural network were acceptably and can be use for predicting the rainfall . So by using this method of prediction we can find the amount of rainfall in the region by using the attributes like humidity , dew point and pressure.

Luk, Ball and Sharma described that due to complexity of the atmospheric processes by which rainfall is generated and the lack of available of data on the necessary temporal and spatial scales, it is not feasible generally to forecast rainfall using a physically based process model.

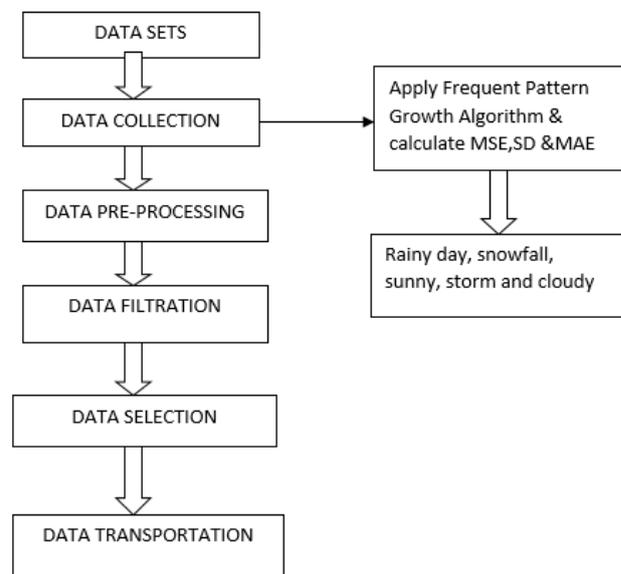
TV Rajinikanth et al. has proposed a methodology for analyzing the metrological data that is very much suite in particular to the Indian weather environments using data mining techniques. The related data is grouped basing on k means algorithm,a decision tree is used for predicting the observations.

III. IMPLEMENTATION

Five data-mining algorithms, neural network (NN), random forest, classification and regression tree (C&RT), support vector machine (SVM), and k-nearest neighbour (k-NN) were used to build the prediction models. NN consists of a group of interconnected neurons, creating it AN adaptive system that may amendment its structure supported external or internal data flowing through the network throughout the learning phase. NNs are sometimes wont to model advanced relationships between input and output variables.. Random forest combines call tree predictors in an exceedingly method that every tree depends on the values of a random

vector sampled independently and with an equivalent distribution for all trees in the forest. It integrates a sacking plan and a random selection of options in constructing a set of call trees. C&RT, popularized by Breiman, may be a statistic technique manufacturing logical if-then rules that are simple to interpret. AN SVM may be a supervised learning technique used for classification and regression analysis. SVM constructs one or a collection of hyper planes in an exceedingly high or infinite dimensional space. The key advantage of SVM is that the use of kernel functions creating SVM appropriate for modelling in advanced nonlinear domains. k-NN is AN instance-based learning method accounting for contributions of the neighbours. It offers sensible performance for a few categories of applications.

The whole project is developed using frequent pattern algorithm in the following data diagram This project is completed in five stages i.e. Data Collection, Data Pre- Processing, data Filtration, Data Selection and Data Transportation



Data Collection: Data used in this project has been collected from various government sites such as: data.gov.in, India stat, India meteorological department.

Data Pre-processing: In this stage, a standardized format for the information model was developed that took care of missing data, finding duplicated data, and removing of dangerous data. Finally, the cleaned information was re-modelled into a format appropriate for data mining.

Data Filtration: At this stage, knowledge relevant to the analysis was selected and retrieved from the dataset.

Data Selection : At this stage , relevant data is retrieved from the dataset for the analysis.

Data Transportation : This stage is also known as data consolidation . In this selected data is transformed in appropriate form for data mining.

The meteorological dataset had eight (8) attributes, their kind and outline is presented in Table 1, whereas associate analysis of the numeric values is bestowed in Table 2. Because of the character of the Cloud Form knowledge wherever all the values are identical and therefore the high percentage of missing values within the sunshine knowledge each were not employed in the analysis.

Table 1 : Attributes of Meteorological Dataset

ATTRIBUTE	Month	Year	Temperature	Humidity	Wind Speed	Sea Level Pressure	Rainfall
TYPE	Numerical	Numerical	Numerical	Numerical	Numerical	Numerical	Numerical
DESCRIPTION	Month Considered	Year Considered	Monthly Minimum Temperature	Monthly Minimum Humidity	Wind Running Km	Pressure in Pascals	Total Monthly Rainfall

This is conjointly called knowledge consolidation. it's the stage in which the chosen knowledge is re-modelled into forms appropriate for data mining. the info file was saved in Comma Separated price (CSV) file format and therefore the datasets were normalized to scale back the impact of scaling on the data. The following table show the analysis of numeric vales:

Table 2. Analysis of numeric data values

No	Variable	Min	Max	Mean	MAE	MSE	SD
1	Month	June (1)	June(30)	-	-	-	-
2	Year	2015	2015	-	-	-	-
3	Temperature	27	32	29.5	0.5957	1.0637	0.6981
4	Humidity	73	82	77.5	0.5957	1.0637	0.6981
5	Wind Speed	6	23	14.5	0.5957	1.0637	0.6981
6	Sea Level Pressure	999	1007	1003	0.5957	1.0637	0.6981
7	Rainfall	Rain	Rain	Rain	Rain	Rain	Rain

Mean Absolute Error (MAE): It is a unremarkably used quantity in time-series analysis that measures however shut the predictions are to the observations.

Mean Square Error: It measures the difference between the predicted model and the value shown by an actual observer. It includes the variance of the predictive model and the variance. The error of an angle is one of the most widely used measures of success in numerical predictions. This value is calculated with the following formula. It is based on the difference in the square mean between each calculated value and the corresponding value.

Standard Deviation: Measures the distinction between values inexplicit by the prediction model

and true observations. It incorporates each the variance of the prediction model and its bias. It indicates what quantity variation exists within the values from the average. it's commonly accustomed live confidence in applied math conclusions.

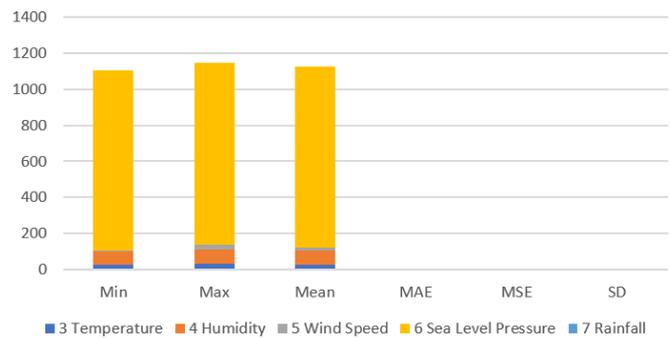
IV. RESULT

On collected data sets we apply FPG Algorithm in correct measurement. And calculate the MSE, SD & MAE with the help of basic equations to evaluate prediction of Rainfall.

Weather forecasting is the attempt to predict the weather conditions at some future time. The algorithm we use here are Neural Networks (NN), Random Forest, Classification and Regression tree (C &RT), the support vector machine, k-nearest neighbor and its technical method to analyze the data and predict the weather. The parameters used to classify weather are humidity, rainfall, temperature. Data used in this project has been collected from data.gov.in and www.indiastat.com.

The project is completed in four stages i.e. Data Collection, Data Pre-Processing, Data Filtration, Data Selection and Data Transportation which has been shown in the project flow diagram.

On collected data sets we apply FPG algorithm in correct measurement. And calculate the MSE, SD and MAE with help of basic equations to evaluate prediction of rainfall. This work is essential to study climatic changes because of variations in weather conditions and they can be studied using these data mining techniques.



V. CONCLUSION

In this word Frequent Pattern Growth algorithm is applied to obtain decision trees and rules for classifying weather parameters such as humidity, rainfall, temperature and wind speed. The data sets we used in this project are collected from data.gov.in, www.indiastat.com for Shimla Himachal Pradesh. On the collected datasets we use FPG algorithm and calculate SD, MSE and MAE. These calculations are more accurate than present model. With FPG algorithm predictions are more accurate than neural network. This work is essential to study climatic changes because of variations in weather conditions like wind speed, rainfall and temperature. And they can be studied using these data mining techniques.

VI. REFERENCES

- [1]. International Journal of Science and Research (IJSR) , By Amruta A. Taksande , P.S. Mohod
- [2]. Roger A. Pielke Sr. , "Mesoscale Meteorological Modelling" , International Geophysics Series 2018.
- [3]. Andrew Kusiak , " Modelling and Prediction of Rainfall Data : A Data Mining Approach" IEEE trans on Geo Science and remote Sensing 2019.
- [4]. Lorenzo Luini and Carlo Capsoni, "A Unified Model for the Prediction of Spatial and Temporal Rainfall Rate Statistics" IEEE trans 2013.

- [5]. State wise Rainfall Statistics in India and monthly Rainfall growth details figures from www.indiaset.com
- [6]. Kumar Abhishek , M.P..Singh , Saswata Ghosh , Abhishek Anand “Weather Forecasting model using Artificial Neural Network” – Procedia Technology 2018
- [7]. Mohamed Akram Zaytar , Chaker El Amrani “ Sequence to Sequence Weather Forecasting With Long Short Term Memory Recurrent Neural Networks” – International Journal of Computer Applications (0975-8887) Volume143 - No.11, June 2016.
- [8]. Liu, J.N.K., Lee, R.S.T. , “Rainfall Forecasting from Multiple Point Source Using Neural Networks”, IEEE Int’l. Conf. Systems 2019.
- [9]. Likas, Nikos Vlassis, Jakob J. Verbeek, "The global k-means clustering algorithm", Pattern Recognition, The Journal of the Pattern Recognition Society, March 2002
- [10]. Fahad Sheikh ,S. Karthick, "Analysis of Data Mining Techniques for Weather Prediction", Indian Journal of Science and Technology , 2016.
- [11]. Rakesh Kumar Arora, Dharmendra Badal, “Evaluating student performance using k means clustering”, International Journal of Computer Science and Technology, 2013.
- [12]. Prashant Biradar, Sarfraz Ansari, Yashavant Paradkar and Savita Lohiya, "Weather Prediction Using Data Mining", International Journal of engineering research and development 2017.
- [13]. Anagha Byakude Aishwarya Dhore, Bhagyashri Sonar and Mansi Waste, “Weather prediction using the data mining Technique”, IEEE 2016.
- [14]. Deepti Gupta and Udayan Ghose, “A Comparative Study of Classification Algorithms for Forecasting Rainfall”, IEEE Publications 2015.
- [15]. Rakesh Kumar Arora, Dharmendra Badal, “Placement Prediction through Data Mining”, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, Issue 7, 2014.
- [16]. Jiawei Han and Micheline Kamber (2006), Data Mining Concepts and Techniques, published by Morgan Kauffman, 2nd ed.
- [17]. India Statistical : <https://www.indiastat.com/>
- [18]. India Meteorological Department : <https://mausam.imd.gov.in>

Cite this article as :

Swati Pandey, Shruti Sharma, Shubham Kumar, Kanchan Bhatt, Dr. Rakesh Kumar Arora, "Weather Forecast through Data Mining", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 7 Issue 3, pp. 90-95, May-June 2021. Available at doi : <https://doi.org/10.32628/CSEIT217318>
Journal URL : <https://ijsrcseit.com/CSEIT217318>