

Data Mining Techniques for Weather Forecasting and Climate Change Prediction System: A Review

¹Laxmi Kumari, ²Dr.Deepika Punj,

¹Master of Technology Student, ²Assistant Professor

^{1,2}Department of Computer Engineering,

^{1,2}J.C. Bose University of Science & Technology YMCA, Faridabad, Haryana, India

Abstract : Data Mining is the process of cultivating and analyzing enormous sets of data and extracting the meaningful data from them. In simple words, data mining is the process of mining knowledge from the data. The two main functions of data mining include descriptive function and classification and prediction. Descriptive function deals with the general properties of data in database. Classification and prediction includes discovering a model that narrates the data classes and then using this model to predict the class of object whose label is unknown respectively.

This paper presents a survey of available literatures of different data mining techniques employed by various researchers to predict the weather condition. The work of different researchers has been reviewed and represented in tabular form. The decision tree and K-mean clustering is found out to be best suited for this application.

IndexTerms - Weather Prediction, Data Mining, Clustering, Classification.

I. INTRODUCTION

Weather Prediction is a really very enthralling concept and electrifies the weather experts to track close/almost accurate prediction. Weather prediction is challenging task as it depends upon various factors such as temperature, humidity, wind speed, etc which keeps on changing. Weather Prediction is important for numerous applications such as climate monitoring, aviation industry, drought detection, communication, severe weather prediction, agriculture and production, planning in energy industry, disaster prevention, pollution dispersal, etc. Weather prediction even affect for our day to day planning as we decide our action according to the predicted weather like what we will wear, is there is any need to carry umbrella or rain coat, etc.

II. LITERATURE REVIEW

Different researchers advocated their different models and methods on weather forecasting for early prediction of weather conditions.

S.S.Badhiye et al[1]proposed the use of KNN(K-Nearest Neighbor) classification algorithm along with clustering technique for hidden pattern recognition from large weather related dataset and convey this retrieved information into usable knowledge for climate change prediction. For a particular time interval humidity and temperature is procured. For prediction and analysis of parameters in distant areas a data logger system embedding software within is used. Prediction accuracy obtained is high for humidity and temperature.

Neha Khandelwal and Ruchi Davey [2] presented a system that predict the rainfall of a year by considering four different climatic parameters, which are pressure, temperature, sea level and humidity and as result of that calculating drought chances in Rajasthan using the dataset. Data mining techniques are used to find certain factors. Then correlation is established in the factors by applying correlation analysis on dataset. The factors having positive correlations are picked and utilized for regression analysis for predicting rainfall, in regression analysis, multiple linear regressions (MLR) is used. Then statistical analysis is applied on the data for predicting drought possibility. Only single parameter rainfall is taken into consideration for inspecting drought condition in contrast with the other climatic factors may affect the condition to an ample range. Consequently it is not very much accurate.

M. A. Kalyankar and S. J. Alaspurkar [3] implemented data mining techniques to obtain weather related data. The techniques are used to find the hidden patterns inside the large dataset in order to transfer the retrieved information into utilizable knowledge for the classification and the prediction of weather condition to build a system that can learn dynamically to meet the nature of expeditiously changeable weather conditions and unforeseen events, Dynamic data mining methods are needed. . Data mining process is implemented to extract knowledge from weather dataset of Gaza city. This knowledge can be utilized to acquire useful predictions and assist the decision making process.

Subana Shanmuganathan and Philip Sallis [4] used the data mining methods to look for the patterns in the ad- hoc weather conditions, such as month of the year, wind direction, time of the day, speed, and severity using a data set related to single location. The historical weather data, between 2008 and 2012 is taken into consideration, collected using telemetry devices installed in a vineyard in the north of New Zealand.The decision tree algorithms used are C5, Quest, CRT and CHAID. SOM is used for the clustering purpose. The wind gust is predicted using Multilayered supervised ANN. Data mining methods and statistical techniques are executed using SPSS. It provides a good tool for analyzing adhoc dataset.

Pinky Saikia Dutta and Hitesh Tahbilder [5] used data mining technique for predicting monthly Rainfall of Assam. Multiple Linear Regression- Traditional statistical technique is used. The dataset consist of data collected over Six years period between 2007 and 2012 locally from Regional Meteorological Center, Guwahati, Assam, India. Parameters taken for the model are maximum temperature, minimum temperature, wind speed, mean sea level pressure and rainfall. Acceptable accuracy is provided by prediction model implemented using multiple linear regression. Some parameters like wind direction is not included in the dataset.

Kavita Pabreja [6] examined the use of data mining techniques to derive the sub-grid scale weather systems from Numerical Weather Prediction model output products which cannot be achieved through normal Model Output Statistics technique. The early prediction of cloudburst formation is done by applying data mining technique, clustering on divergence and relative humidity. Two days data of real life is taken into consideration and K mean clustering is applied on the datasets. One disadvantage is found that it cannot be used for long term predictions.

Gaurav J. Sawale and Sunil R. Gupta [7] predicted weather for future using an artificial neural network method in a given location. Initial modeling is done using Back Propagation Neural Network. Then results of BPN model are fed to Hopfield Networks. The attributes considered include temperature, humidity and wind speed. As the prediction errors are less and learning process is faster, the proposed system can be used as good alternative to traditional approaches for weather prediction.

III. COMPARATIVE ANALYSIS

Table 3.1 Comparison Table

Author	Applications	Techniques	Advantages	Disadvantages
S Badhiye et al[1]	Temperature and humidity prediction	Clustering techniques	Satisfactory for multi-modal classes	Can't predict remote area data
Neha Khandelwal, Ruchi Davey[2]	Prediction of Drought	Regression	Correlation and statistical analysis is also performed	Verification not finished
Meghali A. Kalyankar, S. J. Alaspurkar [3] Meteorological data analysis	Temperature, rain, wind speed, humidity	Clustering, K-mean clustering	Have Good prediction accuracy	Dynamic data mining methods are required.
Phillip Sallis, S Shanmuganathan [4] Wind gust prediction	Humidity, temperature, Dew point, wind speed, wind direction	Decision tree, ANN C5.0, CRT, QUEST, CHAID, SOM	Good for analyzing ad hoc dataset	Do not handle continuous data. Data is recorded at irregular intervals
PS Dutta, H Tahbilder [5] Rainfall prediction	Rainfall, Min and max temperature, humidity, wind direction	Regression MLR (multiple linear regression)	Acceptable accuracy	Attribute elimination required for better accuracy
K Pabreja [6] Cloud burst prediction.	Temperature, humidity	Clustering K- mean clustering	Supplement with Numerical Weather Prediction models.	Poor performance for long term predictions
GJ Sawale, Dr.S.Gupta [7]	Temperature, humidity, wind speed	ANN BPN, Hopfield networks	Have better prediction accuracy	Requirement of Attribute normalization

Table 3.1 displayed the comparison among the different techniques and methodologies used by various researchers for weather forecasting and prediction.

IV. CONCLUSION

This paper provides a survey of available literatures of various data mining techniques used for predicting weather. The review narrates the study of predicting several weather phenomena such as temperature, thunderstorms, rainfall, cloudburst using various data mining methods like decision trees, artificial neural networks, clustering and regression algorithms. Out of different data mining techniques used for the application, Decision Tree and K-Mean clustering is found to be best suited for the application and can be used as good alternative to the traditional metrological approaches .

V. ACKNOWLEDGMENT

I would like to thank my guide Dr. Deepika Punj, Computer Engineering Department, J.C. Bose university of Science and Technology, YMCA ,Faridabad for their continuous guidance. Also, I would like to thank my family and friends for their support in my research work.

REFERENCES

- [1] Badhiye S. S., Dr. Chatur P. N., Wakode B. V., January 2012, "Temperature and Humidity Data Analysis for Future Value Prediction using Clustering Technique: An Approach", International Journal of Emerging Technology and Advanced Engineering, 2250-2459, Volume 2, Issue 1.
- [2] Neha Khandelwal, Ruchi Davey, September -October 2012, "Climatic Assessment Of Rajasthan's Region For Drought With Concern Of Data Mining Techniques", International Journal Of Engineering Research and Applications (IJERA), Vol. 2, Issue 5, 1695-1697.
- [3] Meghali A. Kalyankar, S. J. Alaspurkar, February – 2013, "Data Mining Technique to Analyse the Metrological Data", International Journal of Advanced Research in Computer Science and Software Engineering 3(2), 114-118.
- [4] Subana Shanmuganathan and Philip Sallis, 2014, "Data Mining Methods to Generate Severe Wind Gust Models", 5, 60-80.
- [5] Pinky Saikia Dutta, Hitesh Tahbilder, Apr-May 2014 "Prediction Of Rainfall Using Data mining Technique Over Assam", Indian Journal of Computer Science and Engineering (IJCSSE), Vol. 5 No.2.
- [6] Kavita Pabreja, 2012, "Clustering technique to interpret Numerical Weather Prediction output products for forecast of Cloudburst", International Journal of Computer Science and Information Technologies (IJCSIT), Vol. 3 (1) , 2996 - 2999.
- [7] Gaurav J. Sawale, Dr. Sunil R. Gupta, April 2013, "Use of Artificial Neural Network in Data Mining For Weather Forecasting", International Journal Of Computer Science And Applications Vol. 6, No.2.