

Climate change and Weather Forecast Analysis Model using Data Mining Techniques

Shubham Kumar, Kanchan Bhatt

Professor, Krishna Engineering College, Ghaziabad, Uttar Pradesh, India

Article Information

Received : 02 July 2024
Revised : 06 July 2024
Accepted : 18 July 2024
Published : 22 July 2024

Corresponding Author:

Shubham Kumar

Abstract— 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, KNN 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*

Copyright © 2024: Shubham Kumar, Kanchan Bhatt, This is an open access distribution, and reproduction in any medium, provided Access article distributed under the Creative Commons Attribution License the original work is properly cited License, which permits unrestricted use.

Citation: Shubham Kumar, Kanchan Bhatt, “Climate change and Weather Forecast Analysis Model using Data Mining Techniques”, Journal of Science, Computing and Engineering Research, 7(7), July 2024.

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. RELATED WORKS

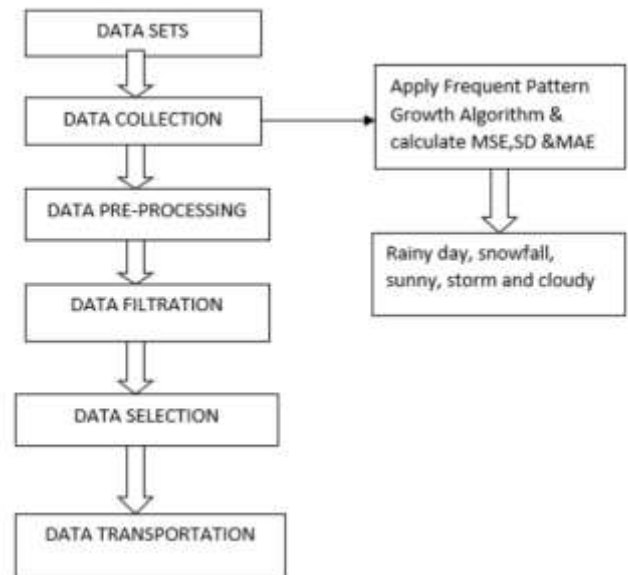
On a worldwide scale large numbers of attempts have been made by different researchers to forecast Weather accurately using various techniques but due to nonlinear nature of Weather prediction accuracy obtained by these techniques is still below the satisfactory level. Sarma , Konwar and Das shows that 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 . TwoThird of the data was used for training and OneThird 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. METHODOLOGY

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 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

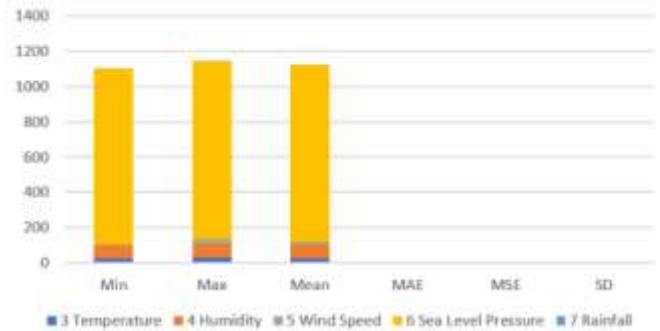
ATTRIBUTE	Month	Year	Temperature	Humidity	Wind Speed	Sea Level Pressure	Rainfall
TYPE	Numerical	Numerical	Numerical	Numerical	Numerical	Numerical	Numerical
DESCRIPTION	Month Considered	Year Considered	Monthly Minimum Temperature	Monthly Minimum Humidity	Wind Run in Km	Pressure in Pascal	Total Monthly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

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

IV. RESULTS AND DISCUSSIONS

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.

REFERENCES

- [1]. P. Nirmala, T. Manimegalai, J. R. Arunkumar, S. Vimala, G. Vinoth Rajkumar, Raja Raju, "A Mechanism for Detecting the Intruder in the Network through a Stacking Dilated CNN Model", *Wireless Communications and Mobile Computing*, vol. 2022, Article ID 1955009, 13 pages, 2022. <https://doi.org/10.1155/2022/1955009>.
- [2]. D. Sathyanarayanan, T. S. Reddy, A. Sathish, P. Geetha, J. R. Arunkumar and S. P. K. Deepak, "American Sign Language Recognition System for Numerical and Alphabets," 2023 International Conference on Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering (RMKMATE), Chennai, India, 2023, pp. 1-6, doi: 10.1109/RMKMATE59243.2023.10369455.
- [3]. J. R. Arunkumar, Tagele berihun Mengist, 2020" Developing Ethiopian Yirgacheffe Coffee Grading Model using a Deep Learning Classifier" *International Journal of Innovative Technology and Exploring Engineering (IJITEE)* ISSN: 2278-3075, Volume-9 Issue-4, February 2020. DOI: 10.35940/ijitee.D1823.029420.
- [4]. Ashwini, S., Arunkumar, J.R., Prabu, R.T. et al. Diagnosis and multi-classification of lung diseases in CXR images using optimized deep convolutional neural network. *Soft*

- Comput (2023). <https://doi.org/10.1007/s00500-023-09480-3>
- [5]. J.R.Arunkumar, Dr.E.Muthukumar," A Novel Method to Improve AODV Protocol for WSN" in Journal of Engineering Sciences" ISSN NO: 0377-9254Volume 3, Issue 1, Jul 2012.
- [6]. R. K, A. Shameem, P. Biswas, B. T. Geetha, J. R. Arunkumar and P. K. Lakineni, "Supply Chain Management Using Blockchain: Opportunities, Challenges, and Future Directions," 2023 Second International Conference on Informatics (ICI), Noida, India, 2023, pp. 1-6, doi: 10.1109/ICI60088.2023.10421633.
- [7]. Arunkumar, J. R. "Study Analysis of Cloud Security Chanllenges and Issues in Cloud Computing Technologies." Journal of Science, Computing and Engineering Research 6.8 (2023): 06-10.
- [8]. J. R. Arunkumar, R. Raman, S. Sivakumar and R. Pavithra, "Wearable Devices for Patient Monitoring System using IoT," 2023 8th International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2023, pp. 381-385, doi: 10.1109/ICCES57224.2023.10192741.
- [9]. S. Sugumar, C. Geetha, S. S, P. C. Bharath Kumar, T. D. Subha and J. R. Arunkumar, "Energy Efficient Routing Algorithm with Mobile Sink Assistance in Wireless Sensor Networks," 2023 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), Chennai, India, 2023, pp. 1-7, doi: 10.1109/ACCAI58221.2023.10201142.
- [10].R. S. Vignesh, V. Chinnammal, Gururaj.D, A. K. Kumar, K. V. Karthikeyan and J. R. Arunkumar, "Secured Data Access and Control Abilities Management over Cloud Environment using Novel Cryptographic Principles," 2023 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), Chennai, India, 2023, pp. 1-8, doi: 10.1109/ACCAI58221.2023.10199616.
- [11].Syamala, M., Anusuya, R., Sonkar, S.K. et al. Big data analytics for dynamic network slicing in 5G and beyond with dynamic user preferences. Opt Quant Electron 56, 61 (2024). <https://doi.org/10.1007/s11082-023-05663-2>
- [12].Krishna Veni, S. R., and R. Anusuya. "Design and Study Analysis Automated Recognition system of Fake Currency Notes." Journal of Science, Computing and Engineering Research 6.6 (2023): 16-20.
- [13].V. RamKumar, S. Shanthi, K. S. Kumar, S. Kanageswari, S. Mahalakshmi and R. Anusuya, "Internet of Things Assisted Remote Health and Safety Monitoring Scheme Using Intelligent Sensors," 2023 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), Chennai, India, 2023, pp. 1-8, doi: 10.1109/ACCAI58221.2023.10199766.
- [14].R. S. Vignesh, R. Sankar, A. Balaji, K. S. Kumar, V. Sharmila Bhargavi and R. Anusuya, "IoT Assisted Drunk and Drive People Identification to Avoid Accidents and Ensure Road Safety Measures," 2023 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), Chennai, India, 2023, pp. 1-7, doi: 10.1109/ACCAI58221.2023.10200809.
- [15].I. Chandra, G. Sowmiya, G. Charulatha, S. D, S. Gomathi and R. Anusuya, "An efficient Intelligent Systems for Low-Power Consumption Zigbee-Based Wearable Device for Voice Data Transmission," 2023 International Conference on Artificial Intelligence and Knowledge Discovery in Concurrent Engineering (ICECONF), Chennai, India, 2023, pp. 1-7, doi: 10.1109/ICECONF57129.2023.10083856.
- [16].G. Karthikeyan, D. T. G, R. Anusuya, K. K. G, J. T and R. T. Prabu, "Real-Time Sidewalk Crack Identification and Classification based on Convolutional Neural Network using Thermal Images," 2022 International Conference on Automation, Computing and Renewable Systems (ICACRS), Pudukkottai, India, 2022, pp. 1266-1274, doi: 10.1109/ICACRS55517.2022.10029202.
- [17].R. Meena, T. Kavitha, A. K. S, D. M. Mathew, R. Anusuya and G. Karthik, "Extracting Behavioral Characteristics of College Students Using Data Mining on Big Data," 2023 International Conference on Artificial Intelligence and Knowledge Discovery in Concurrent Engineering (ICECONF), Chennai, India, 2023, pp. 1-7, doi: 10.1109/ICECONF57129.2023.10084276.
- [18].S. Bharathi, A. Balaji, D. Irene, J. C. Kalaivanan and R. Anusuya, "An Efficient Liver Disease Prediction based on Deep Convolutional Neural Network using Biopsy Images," 2022 3rd International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 2022, pp. 1141-1147, doi: 10.1109/ICOSEC54921.2022.9951870.
- [19].I. Chandra, G. Sowmiya, G. Charulatha, S. D, S. Gomathi and R. Anusuya, "An efficient Intelligent Systems for Low-Power Consumption Zigbee-Based Wearable Device for Voice Data Transmission," 2023 International Conference on Artificial Intelligence and Knowledge Discovery in Concurrent Engineering (ICECONF), Chennai, India, 2023, pp. 1-7, doi: 10.1109/ICECONF57129.2023.10083856.
- [20].Revathi, S., et al. "Developing an Infant Monitoring System using IoT (INMOS)." International Scientific Journal of Contemporary Research in Engineering Science and Management 6.1 (2021): 111-115.
- [21].J.R.Arunkumar, Dr.E.Muthukumar,| A Novel Method to Improve AODV Protocol for WSN| in Journal of Engineering Sciences| ISSN NO: 0377-9254Volume 3, Issue 1, Jul 2012.
- [22].R. S. Vignesh, A. Kumar S, T. M. Amirthalakshmi, P. Delphy, J. R. Arunkumar and S. Kamatchi, "An Efficient and Intelligent Systems for Internet of Things Based Health Observance System for Covid 19 Patients," 2023 International Conference on Artificial Intelligence and Knowledge Discovery in Concurrent Engineering (ICECONF), Chennai, India, 2023, pp. 1-8, doi: 10.1109/ICECONF57129.2023.10084066.
- [23].I. Chandra, K. V. Karthikeyan, R. V, S. K, M. Tamilselvi and J. R. Arunkumar, "A Robust and Efficient Computational Offloading and Task Scheduling Model in Mobile Cloud Computing," 2023 International Conference on Artificial Intelligence and Knowledge Discovery in Concurrent Engineering (ICECONF), Chennai, India, 2023, pp. 1-8, doi: 10.1109/ICECONF57129.2023.10084293.
- [24].R. K, A. Shameem, P. Biswas, B. T. Geetha, J. R. Arunkumar and P. K. Lakineni, "Supply Chain Management Using Blockchain: Opportunities, Challenges, and Future Directions," 2023 Second International Conference on Informatics (ICI), Noida, India, 2023, pp. 1-6, doi: 10.1109/ICI60088.2023.10421633.

- [25].J. R. Arunkumar, and R. Anusuya, "OCHRE: A Methodology for the Deployment of Sensor Networks." American Journal of Computing Research Repository, vol. 3, no. 1 (2015): 5-8.

