Journal of Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.7750393

RAINFALL PREDICTION USING HYPER PARAMETER TUNING

Dr.V.Ramesh Babu[1] , Dr.A.Vinoth kumar[1] ,Dr.D.Usha[1]
Potru Manishanvitha[2] ,Nunna Lahari[2] ,Papolu Shiney[2]
[1]Professor(s),Dr.M.G.R Educational and Research Institute
[2]Final Year B.Tech CSE,Dr.MGR Educational and Research Institute
[1]rameshbabu.cse@drmgrdu.ac.in, dravinothkumar@gmail.com, usha.cse@drmgrdu.ac.in
[2]manishanvitha16@gmail.com, nunnalahari@gmail.com, shineypapolu1426@gmail.com.

Abstract: Predicting rainfall is one of the more difficult tasks because there are so many variables to take into account. A brand-new technique for predicting rainfall has been suggested. The suggested approach employed hyper-tuned settings using machine learning techniques based on regression. Three different machine learning algorithms, specifically regression models, trained the dataset. For the purpose of predicting rainfall, the proposed model included machine learning, Ridge regression, Logistic regression, and Lasso regression. Online data collection for experimental purposes spans the years 1901 to 2015. The experimental findings point to a respectable advancement over established techniques for rainfall forecasting. By achieving lower MAE error values, the proposed performed better than the compared models.
Keywords: rainfall predictions, HyperTuning, Machine Learning, MAE,CNN.

I. INTRODUCTION

Since rainfall has a significant impact on our environment and nature, rainfall prediction is regarded as one of the most crucial weather forecasting-related studies. Rainfall has a significant impact on a number of natural phenomena, including flood, drought, and weather indicators like relative humidity. Therefore, reliable and precise rainfall prediction technologies are of relevance. Quantitative rainfall forecasts have been successfully made on numerous occasions. The nature of meteorological data makes rain prediction difficult. Since there are numerous large special sectors of data involved, forecasting is complicated. The scientific model for precipitation is nonlinear and depends on a variety of meteorological factors. Running numerical climate prediction frameworks that replicate the conditions of meteorological phenomena using a mesoscale or local forecast. Neural networks are increasingly being used in research of nonlinear systems since they are more appropriate for them. Rainfall prediction is important requirement to help water source managers in countries like India. This prediction is useful for government to know prior on droughts or heavy rainfall. The prediction of rainfall with high accuracy is inevitable in countries like India, as our economy majorly sdepends on agriculture, which in turn needs rainfall. Among all weather conditions, rainfall is considered as major one, as human life is much dependent on rainfall. The motivation is that neural network techniques are imperative for the rainfall prediction is that it analyses better on the trend from the given volume of historical data. The existing work is based on statistical or mathematical models, which calculated accurately, but there are not used on predictions. These days, how to improve the precision of prediction is a major and troublesome in the study of machine learning techniques. In big data era, how to utilize the extensive measure of climate information gathered to improve the accuracy of the conventional climate figure rate has been a major challenge test of rainfall prediction. Our goal is to create a fantastic rainfall prediction model that makes extensive use of rainfall data to reveal hidden information and ultimately increase the accuracy of the rainfall prediction. There are numerous published works on predicting precipitation and the weather, among which statistical models were thoroughly studied and applied. Regression analysis is the basis for statistical methods. However, it occasionally fails to deliver positive outcomes.

II.LITERATURE SURVEY

[1]. Koizumi K. Weather Forecast., 14, 109–118, 1999. An objective method of forecasting precipitation coverage with a neural network is presented. This method uses as predictors all available data at local weather stations including both numerical model results and weather data obtained later than the model initial time, which sometimes contradict each other and hence have to be handled subjectively by well-experienced forecasters. Since the method gives an objective and also realistic forecast of areal precipitation coverage, its skill scores are better than those of the persistence forecast (after 3 h), the linear regression forecasts, and numerical model precipitation prediction. Koizumi employed an ANN model using radar, satellite and weather-station data together with numerical products generated by the Japan Meteorological Agency (JMA) Asian Spectral Model for 1-year training data. Koizumi found that the ANN skills were better than persistence forecast, the linear regression forecasts, and numerical model precipitation prediction. As the ANN used only 1 year data for training, the results were limited.

[2] Quantitative Rainfall Prediction in Thailand First International Conference on Hydrology and Water Resources on Asia Pacific Region (APHW), Kyoto, Japan. Manusthiparom C., Oki T., and Kanae, S. , 2003. The rainfall is one of the significant data set of water resource management. With the monthly historical rainfall data in the period of 1941-1999 form 245 rainfall monitor stations in Thailand around Chao Phraya River, the rainfall prediction with an artificial intelligent technique is possible. Artificial neural networks is one the most widely supervised techniques of data mining. It can be applied on predictive mining tasks to make a prediction. The main contribution of this paper is to utilize a neural network model for monthly rainfall prediction. The training and testing patterns are prepared as a time-series data of the past ten months. The numbers of training and testing patterns are 372 and 96, respectively. In the training step, the neural network gives 99.6 % of accuracy and 96.9 % of accuracy in the testing step. The results show that it is possible to predict annual rainfall one year ahead with acceptably accuracy. Manusthiparom et al. investigated the correlations between El Nin˜o Southern Oscillation indices, namely, Southern Oscillation Index (SOI), and sea surface temperature (SST), with monthly rainfall in Chiang Mai, Thailand, and found that the correlations were significant. The study suggested that it might be better to adopt various related climatic variables such as wind speed, cloudiness, surface temperature and air pressure as the additional predictors.

[3] Long-range monsoon rainfall prediction of 2005 for the districts and sub-division Kerala with artificial neural network Guhathakurta, P, Current Science 90:773-779, 2006 The advantages of artificial neural network technique for explaining the nonlinear behavior between the inputs and output is explored to forecast the monsoon rainfall of 36 meteorological sub-divisions of India. The model uses the past years of monsoon rainfall data only to forecast the monsoon rainfall of coming year. Monthly rainfall time series data for each of the 36 meteorological sub-divisions constructed by Guhathakurta and Rajeevan (2007) is used for the present study. The model captures well the input-output nonlinear relations and predicted the seasonal rainfall quite accurately during the independent period. All India monsoon rainfall forecasts were generated by using area weighted rainfall forecasts of all the sub-divisions. For the first time the idea of up-scaling is introduced in monsoon rainfall prediction using neural network technique and it is shown that up scaling helps to capture the variability of the all India rainfall better. This helps to predict the extreme years like 2002, 2004 better than the neural network model developed based on single time series of all India rainfall. However, derivation of smaller scale (sub-divisions) forecast model may be more useful than the all India forecast. Present contribution deviates from the study of Guhathakurta in the sense that instead of choosing a particular state, the authors implement Back propagation ANN to forecast the average summer-monsoon rainfall over the whole country. The applicability of the model is limited to monthly rainfall data.

[4] Modeling inter-annual variation of a local rainfall data using a fuzzy logic technique Halide, H. and Ridd P., Proceedings of International Forum on Climate Prediction, 2002, James Cook Universiy, Australia, pp: 166-170, 2002. The present study investigates the ability of fuzzy rules/logic in modeling rainfall for South Western Nigeria. The developed Fuzzy Logic model is made up of two functional components; the knowledge base and the fuzzy reasoning or decision making unit. Two operations were performed on the Fuzzy Logic model; the fuzzification operation and defuzzification operation. The model predicted outputs were compared with the actual rainfall data. Simulation results reveal that predicted results are in good agreement with measured data. Prediction Error, Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and the Prediction Accuracy were calculated, and on the basis of the results obtained, it can be suggested that fuzzy methodology is efficiently capable of handling scattered data. The developed fuzzy rule-based model shows flexibility and ability in modeling an ill-defined relationship between input and output variables. Halide and Ridd used fuzzy logic to rainfall prediction. The fuzzy logic technique is used to model and predict local rainfall data. The RMSE between data and model output is found to be 319.0 mm which is smaller than that by using either the local rain or the Niño 3.4 alone of 349.2 and 1557.3 mm, respectively.

III. EXISTING SYSTEM

In order to forecast rainfall in India, numerous ANN-based architectures-based rainfall prediction models have been presented. In Bangkok, Thailand, rainfall has been predicted using an artificial neural network (ANN) based on four years of hourly data. The wet bulb temperature, air pressure, relative humidity, and cloudiness served as the basis for the prediction model. Wet bulb temperature, according to the scientists, may be a deciding element in rainfall forecasting. The learning algorithm looks for the best possible collection of weights for the ANN's neural connections. Since an error function is often minimised, the training step can be viewed as an optimization issue.

Disadvantage of existing system:

- Less feature compatibility.
- Low accuracy.
- The procedure is extremely slow.
- It has been revealed that the standard algorithms may be unable to approximate the exact pattern of the data if it is reasonably complex. .

IV. PROPOSED SYSTEM

Proposed Rainfall prediction based on rainfall by month-wise data. The data has been collected over the period from 1901 to 2015. There are three techniques applied for prediction Logistics regression, Ridge regression and Lasso regression. The machine learning model used Hyper parameter tuning. Grid Search CV tries all possible combination of the parameters taken from the data and evaluates the model for each combination through cross validation method

Advantages of Proposed System:

- Accurate Classification.
- Less Complexity
- High performance.
- Easy identification.

V.SYSTEM ARCHITECTURE

Dataset Collection: The selection of high-quality data for analysis is a step in the data collection process. Here, we implemented machine learning using a dataset of phishing websites that was acquired from uci.edu. A data analyst's task is to identify methods and resources for gathering accurate and complete data, interpreting that data, and doing outcomes analysis using statistical methods.

Data visualization: It's simpler to comprehend and analyse a lot of data when it's displayed graphically. A data analyst may be required by some firms to know how to make slides, diagrams, charts, and templates. Our method uses a histogram plot and feature extraction that are visualized.

Model Training: Following preprocessing of the gathered data and division into train and test sets, a data scientist can move on with a model training. The algorithm is "fed" training data during this process. An algorithm will evaluate the data and produce a model that can locate a target value (attribute) in fresh data, giving you the result you're looking for with predictive analysis. Model development is the goal of model training. The objective of this stage is to create the most basic model that can quickly and adequately formulate a target value. By fine-tuning the models, a data scientist can accomplish this objective. To get the best performance out of an algorithm, model parameters are optimised in this way.

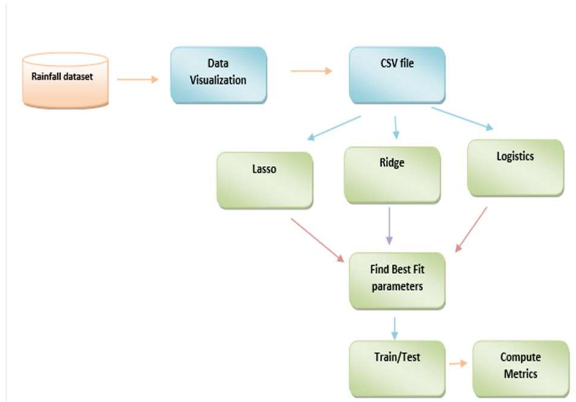

Fig 1.1 System Architecture

VI. IMPLEMENTATION

Module 1: Data preprocessing

Preprocessing is used to convert unprocessed data into a machine learning-friendly format. By utilising structured and clean data, a data scientist can utilise an applied machine learning model to produce conclusions that are more accurate. The process includes data formatting, cleansing, and sampling.

Module 2: Dataset splitting

Three subsets are training, test and validation sets should be created from a dataset used for machine learning. Exercise equipment to train a model and provide the best parameters that it should learn from data, a data scientist utilises a training set. a test set. To assess the trained model's generalizability and evaluate its performance, a test set is required. The latter refers to a model's capacity, following training over training data, to recognise patterns in fresh, previously unobserved data. To prevent model overfitting, which causes the lack of generalizability as discussed above, it is essential to employ separate subsets for training and testing.

Module 3: Training and Testing

A data scientist can start building a model after pre-processing the acquired data and separating it into train and test sets. This procedure comprises "

feeding" training data to the algorithm. Predictive analysis uses an algorithm to evaluate data and produce a model that can locate a target value (attribute) in fresh data. To create a model is the goal of model training.

Convolutional Neural Network (CNN): The monthly rainfall at a chosen location is predicted using a convolutional neural network. CNN is an intricate preceptor. In general, CNN does better at obtaining more precise monthly and yearly averages. The generated results are encouraging and can be widely used to predictions and other scenarios.. .

VII. RESULT AND CONCLUSION

Rainfall plays a crucial influence in determining the weather as well as the likelihood of natural disasters like floods and droughts. The benefit of knowing the weather conditions in advance and taking precautions accordingly is available to agricultural sectors. Additionally, this directly contributes to the growth of the national economy. In the current work, an unique machine learning-based prediction has been developed in response to the need for an accurate and reliable model to predict rainfall. In order to select the optimal parameter to suit the model, the machine learning algorithm employed GridSearchCV for hyper-tuning parameters. The data was gathered between 1901 and 2015. The suggested model has a lower error rate and is effective at predicting whether it will rain. Results from experiments demonstrated that the Ridge and Lasso algorithm performed well.

VIII. REFERENCES

[1] Nayak, D. R., Mahapatra, A., & Mishra, P. (2013). A survey on rainfall prediction using artificial neural network. International Journal of Computer Applications, 72(16)

[2] Dubey, A. D. (2015). Artificial neural network models for rainfall prediction in Pondicherry. International Journal of Computer Applications, 120(3).

[3] Chatterjee, S., Sarkar, S., Hore, S., Dey, N., Ashour, A. S., &Balas, V. E. (2016). Particle swarm optimization trained neural network for structural failure prediction of multistoried RC buildings. Neural Computing and Applications, 1-12.

[4] Chatterjee, S., Ghosh, S., Dawn, S., Hore, S., &Dey,N. (2016). Forest Type Classification: A Hybrid NN-GAModel Based Approach. Information Systems Design and Intelligent Applications (pp. 227-236). Springer India.

[5] Bengio, Y. Lecun, Yann. (1997). Convolutional Networks for Images, Speech, and Time-Series

[6] Reddy, P.C and Babu, A.S(2017).Survey on weather prediction using big data Analystics.2017 Second Intb Conf. Electr. Comput Commu. Technol,.pp.1-6

[7]N.Sharma, M.Zakallua and H.Tiwari, and D.Kumar"Runoff and sediment yield modeling using ANN and Support vector machines: A case study from Nepal Watershed," Model.Earth Syst.Environ., vol.1,no.3,p.23,2015

[8] S.M.Sumi,M.F.Zaman, and H.Hirose, " A rainfall Forecasting method using machine learning models and its application to the Fukuoka citycase,"Int.J.Appl.Math.Comput.Sci.,vol.22,no.4,pp. 841-854, Dec.2012.

[9] S. Lee, J.-C. Kim, H.-S. Jung, M. J. Lee, and S. Lee,''Spatial prediction of flood susceptibility using random-forest and boosted-tree models in Seoul metropolitan city, Korea,'' Geomatics, Natural Hazards Risk, vol. 8, no. 2, pp. 1185–1203, Dec. 2017.

[10] S. Das, R. Chakraborty, and A. Maitra, ''A random forest algorithm for nowcasting of intense precipitation events,'' Adv. Space Res., vol. 60, no. 6, pp. 1271–1282, Sep. 2017.