A Hybrid Model Approach for the Prediction of Rainfall

M. Mallika1, M. Nirmala2
1Assistant Professor, Sathyabama Institute of Science and Technology, Chennai, India
2Professor, Sathyabama Institute of Science and Technology, Chennai, India
sindhikas@gmail.com, mnirmalamaths@gmail.com

Abstract. For maintaining the atmospheric balance, precipitation is very much helpful. Rainfall is one of the forms of precipitation. Though excess rainfall causes several damages to the earth in various ways, it is considered very precious as it is one of the essentials for the existence of human beings. As a result, a prediction of rainfall also forms a major part in planning things. This paper proposes a new hybrid model Moving average-kNN for doing the prediction. Error measures Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE) have been used for the validity of the model.

Keywords: Rainfall Prediction, Moving Average, kNN, MAPE, RMSE, MAE

1. Introduction
Forecasting is defined as the process of estimating or evaluating the value of some variable at some future point of time. Forecasting weather parameters for a region plays an important role and in fact it is one of the main functions of National Weather Services. Water is one of the essential things for the survival of human beings. Water is available in different forms of precipitation. Rain is the most important among them. Information about rainfall is very much useful for predicting natural disasters such as droughts and floods and also very helpful for planning and decision making. Research is taken up for study in this area for quite a decade ago by various researchers using various techniques. The most popular being the neural networks which have given quite a desirable result in predicting the rainfall.

2. Literature Survey
Tomoaki Kashiwao et.al. in their paper made a comparison of the performance of prediction using Multilayer Perceptron, Radial Basis Function Network with least squares method, a hybrid algorithm of Back Propagation and Random Optimization method [20]. For carrying out the prediction, Atmospheric Pressure, Precipitation, Temperature, Humidity, Wind Velocity, Vapor Pressure are taken as variables. The results obtained by this proposed method outperformed the prediction done by Japan Meteorological Agency. Mislan et.al. applied Artificial Neural Network with Back Propagation Algorithm consisting of two hidden layers and three different epochs [14]. It was proved that the network with epoch 1000 gave good results for doing prediction. Kumar Abhishek et.al. implements the BackPropagation for predicting the average rainfall over the Udipi district of Karnataka [7]. Pijush Samui et.al. adopts the technique Support Vector Machine (SVM) and the probabilistic basis model Relevance Vector Machine (RVM) for prediction of rainfall in Vellore [17]. For doing the
prediction four weather parameters have been used. It was found that RVM was a more robust model than SVM.

Jyoti Ranjan Mohanty and Manas Ranjan Mohapatra use simple machine learning model SVM Regression [5]. In this paper, Average Mean Square Error (AMSE) for each kernel which includes Linear, Polynomial, RBF, and Sigmoid is evaluated. In the training phase, polynomial kernel produced low MSE whereas for the testing test linear kernel produced low MSE. M. S. Bennet Praba et.al. deployed sensor devices for collecting and recording the weather parameters like temperature, humidity, wind speed, etc., which is then sent into the cloud for storage and the stored data is classified using SVM to predict whether the rain will occur or not which is represented graphically and sent to the user for the information [1]. S. Swain, S. Nandini, and P. Patel developed an Autoregressive Integrated Moving Average (ARIMA) Model for the monthly rainfall prediction of Khordha district, Odisha, India [19]. The rainfall data from 1901 to 1982 was taken as the training set whereas the data from 1981 to 2002 was used as testing set. The best ARIMA model was chosen for the given data and the forecasts thus obtained by the chosen model gave excellent results when compared with the actual monthly rainfall data for the years 1981 to 2002.

Mexoese Nyatuame and Sampson K. Agodzo used Autoregressive Integrated Moving Average (ARIMA) Model for analyzing annual rainfall and maximum temperature over Tordzie watershed in Ghana [12]. Autocorrelation function (ACF) and Partial Autocorrelation Function (PACF) are used to identify the correct ARIMA model based on visual inspection. Two different ARIMA models - one for annual rainfall and the other for temperature were identified as the best. Different error measures were used for the efficiency and the validity of the models. Analysis have shown that the results are adequate. Pinky Saikia Dutta and Hitesh Tabhilder applied the data mining technique for the monthly prediction of rainfall of Assam [18]. The parameters considered for doing the prediction are Maximum and Minimum temperature, Humidity, Wind Speed, and Pressure. The prediction model which is based on multiple linear regression has given results that are of acceptable accuracy.

Himadri Chakrabarti and Sonia Bhattacharya predicted severe thunderstorms by using the k-nearest neighbour (kNN) technique [2]. In this, kNN classifies two events such as “storm days” and “no storm days” with the help of certain weather parameters. As per the technique prediction of “storm days” are more than 91% correct whereas classification of “storm days” and “no storm days” are more than 82% correct. Imran Ahmed et.al. made an empirical approach to predict the rainfall and its relation with other atmospheric parameters using historical data [3]. The methodology of Multiple Linear Regression is applied to the six years data of Coonor, Nilgiris District, Tamilnadu. Appropriate results have been obtained on rainfall prediction based on this methodology. Mingming Huang et.al. in their paper applied k nearest neighbour algorithm for forecasting precipitation [13]. An improved kNN algorithm is proposed in this research article for different choices of ‘k’. This new forecast approach has proved efficient in the prediction of precipitation. M. Mallika and M. Nirmala in their paper made a comparison of the individual models - kNN, Moving average along with the hybrid model kNN-Moving Average for the prediction of Chennai annual rainfall [11]. The study revealed that Moving average outperformed the other models.

In this present research work, another hybrid combination namely Moving Average-kNN is being tried to get a more prediction accuracy. In the case of the hybrid model kNN-Moving Average, the annual rainfall values are smoothened by kNN for which Moving Average is applied whereas in the hybrid model Moving Average-kNN, the annual rainfall values are smoothened by Moving Average for which kNN is applied. Three different error measures are used for the validity of the model.

3. Study Area and Materials
Chennai, formerly known as Madras and one of the metro cities of India has a population of around eighty lakhs. Chennai city experiences a mostly hot climate. More than six months the climate will be hot only. The major rivers flowing through Chennai which had sufficient water in the earlier years proved to be of no use off late. Thus, the city highly relies on the “Nature’s Gift” to replenish its water reservoirs. The only option for the people of Chennai is always the much-awaited annual monsoons.
City receives two thirds of its annual rainfall through North-east monsoon only. South-west monsoon also gives some rain but that is very little which is, in fact, an added bonus for the citizens. As water is a very essential requirement for the residents of Chennai city for day-to-day purposes, prediction of rainfall is a very much needed asset for doing the planning management by the government. For this research work, the Chennai city annual rainfall data for a total of 113 years (1901-2013) has been used. The dataset is obtained from the Indian Institute of Tropical Meteorology (IITM), Pune, India.

4. Methodology

4.1. k-Nearest Neighbour (kNN)

Usually in all techniques, a generalization of the model takes place with the available data which helps in predicting values for the new entry. kNN is one such technique where generalization doesn’t take place. Given data, a prediction is done with the available data instantly. kNN performs its operation based on the nearest neighbours which are established using the distance formula. Though initially it is best suited for classification problems, it now keeps up the role of regression also. Every technique has its pros and cons, kNN also has its own. But this is inconsiderable as long as prediction and
classification go well. The choice of ‘k’ forms the main part of this technique. Proper choice will give the desired results.

4.2 Moving Average

A Time series is a collection of data observed at regular intervals of time. Time Series Analysis deals with the techniques for the analysis of data. One of the simplest and most widely used techniques in time series analysis is the Moving Average. As the name says, the average keeps moving i.e, the average keeps changing with the inclusion of new entry by maintaining the number of observations for calculating the average. The number of terms to be considered for calculating the average is termed as the period of the moving average. Thus, the forecast value using a simple moving average is given by,

\[ F_t = \frac{A_{t-1} + A_{t-2} + A_{t-3} + \ldots + A_{t-n}}{n} \]

where \( F_t \) = Forecast for the period ‘t’, \( n \) = number of terms to be averaged and \( A_{t-1}, A_{t-2}, A_{t-3} \) are the actual values for the periods \( t-1, t-2, t-3 \) and so on.

4.2.1 Model Validity

Different models are considered to the rainfall data for forecasting. To assess the quality of forecasting and to evaluate the consistency of the model, error measures are required. In this work, the error measures Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE) are used. The formula for the error measures are given below:

(i) \[ MAPE = \frac{1}{n} \sum \left| \frac{Actual - Forecast}{Actual} \right| \times 100 \]

(ii) \[ MAE = \frac{1}{n} \sum |Actual - Forecast| \]

(iii) \[ RMSE = \sqrt{\frac{1}{n} \sum (Actual - Forecast)^2} \]

5. Results and Discussions

In the previous work, the choice of ‘k’ in kNN was tried with different values from 2 to 10 and it was finally settled with \( k = 3 \) based on the error measure MAPE. Similarly, for the moving average, different periods ranging from 2 to 15 were considered and the best period was chosen as 14 based on the error measure MAPE. Hence a coupling of this best moving average is done with kNN for its best choice of ‘k’. The model is named Moving Average-kNN in which the annual rainfall data of Chennai are smoothened by Moving average with period 14 and for these smoothened values kNN is applied. The observed and forecast values of this hybrid model are shown in figure 2 and the error measures for this model is shown in table 1.
5

6. Conclusion
Researchers worldwide have been searching for various techniques for making rainfall predictions. In this research article, an attempt of a hybrid model of statistical technique with the data mining technique is done. This combination has outperformed all the other hybrid models discussed in my previous research work. Future work aims at different combinations of hybrid models for still further prediction accuracy.
References

[1] Bennet Praba M S, Antony John Martin, Siddharth Srivastava and Ajay Rana 2018 Weather Monitoring System and Rainfall Prediction Using SVM Algorithm *International Journal of Research in Engineering, Science and Management* 1745-750.

[2] Himadri Chakrabarty and Sonia Bhattacharya 2015 Application of K-Nearest Neighbor Technique to Predict Severe Thunderstorms *International Journal of Computer Applications* 1101-4.

[3] Imran Ahmed, Shruti Menon and Nikitha K B 2013 Rainfall Prediction Using Multiple Regression Technique *International Journal of Applied Engineering Research* 82267-2270.

[4] Jyothis Joseph and Ratheesh T K 2013 Rainfall Prediction using Data Mining Techniques *International Journal of Computer Applications* 8311-15.

[5] Jyoti Ranjan Mohanty and Manas Ranjan Mohapatra 2018 Rainfall Prediction Using Support Vector Machine (SVM) *IJSR Journal of Computer Engineering (IJSR-JCE)* 20, 06-13.

[6] Khalid Alkhatib, Hassan Najadat, Ismail Hmeidi and Mohammed KAli Shatnawi 2013 Stock Price Prediction using K-Nearest Neighbor (kNN) Algorithm *International Journal of Business, Humanities and Technology* 332-44.

[7] Kumar Abhishek, Abhay Kumar, Rajeev Ranjan and Sarthak Kumar 2012 A Rainfall Prediction Model using Artificial Neural Network *IEEE Control and System Graduate Research Colloquium (ICSGRC)* 82-87.

[8] Mallika M Meenakshi Sundaram and Nirmala M 2015 Annual Mean Temperature Prediction of India Using K-Nearest Neighbour Technique *Applied Mathematical Sciences* 9613-616.

[9] Mallika M and Nirmala M 2016 Chennai Annual Rainfall Prediction using k-Nearest Neighbour Technique *International Journal of Pure and Applied Mathematics* 109115-120.

[10] Mallika M and Nirmala M 2018 An Environmental Study on Forecasting Rainfall Using Data Mining Technique and ARIMA Model: An Integrated Approach *Ekoloji* 27, 1133-1141.

[11] Mallika M and Nirmala M 2020 Analysis of Computational and Statistical model For Time Series Prediction *International Journal of Advanced Science and Technology* 29741-745.

[12] Mexoese Nyatuame and Sampson K Agodzo 2018 Stochastic ARIMA model for annual rainfall and maximum temperature forecasting over Tordzie watershed in Ghana *Journal Of Water And Land Development* 37, 127–140.

[13] Mingming Huang, Runsheng Lin, Shuai Huang and Tengfei Xing 2017 A novel approach for precipitation forecast via improved K-nearest neighbor algorithm *Advanced Engineering Informatics* 33, 89-95.

[14] Mislan, Haviluddin, Sigit Hardwinarto and Marlon Aipassa 2015 Rainfall Monthly Prediction Based on Artificial Neural Network: A Case Study in Tenggarong Station, East Kalimantan-Indonesia *Procedia Computer Science* 59, 142-151.

[15] Nirmala M 2015 Computational Models for Forecasting Annual Rainfall in Tamilnadu *Applied Mathematical Sciences* 9617-621.

[16] Nirmala M and Sundaram SM 2011 Rainfall Prediction through Integrated MA-ANN Model *Journal of Computer Applications Research and Development* 127-34.

[17] Pijush Samui, Venkata Ravibabu Mandla, Arun Krishna and Tarun Teja 2011 Prediction of Rainfall Using Support Vector Machine and Relevance Vector Machine *Earth Sciences India* 4188 – 200.
[18] Pinky Saikia Dutta and Hitesh Tahbilder 2014 Prediction Of Rainfall Using Datamining Technique Over Assam Indian Journal of Computer Science and Engineering 585-90.

[19] Swain S Nandi S and Patel P 2018 Development of an ARIMA Model for Monthly Rainfall Forecasting over Khordha District, Odisha, India Recent Findings in Intelligent Computing Techniques 325-331.

[20] Tomoaki Kashiwao, Koichi Nakayama, Shin Ando, Kenji Ikeda, Moonyong Lee and Alireza Bahadori 2017 A neural network-based local rainfall prediction system using meteorological data on the Internet: A case study using data from the Japan Meteorological Agency Applied Soft Computing 56 317-330.