Weather Forecasting using R

Sachin Kumar Srivastava[1] Robin Singh[2] Rishabh Kumar[3]
[1,2,3]B.Tech Student
[1,2,3]Department of Computer Science & Engineering
[1,2,3]IMS Engineering College (AKTU), India

Abstract— Forecasting[1] is about predicting or finding the future values as accurately as possible, when all the required values are available, including past years data and understanding the events that might affect the forecasts. Weather forecasting is a process of recording of weather like temperature, rainfall, humidity, wind speed etc. By R packages ,models build on the parameters of the weather by which we predict the values of the future weather, considering that the one parameter is dependent on other parameters, by which it changes when the value of other parameter changes. In this paper, different models which were used in the past for weather forecasting are discussed. This paper is divided in three parts , first part is about discussing the related work of the same domain, second part is about operation is done in process of weather forecasting and the third part is about the techniques we used which is regression.

Key words: R-Packages, Regression, Weather, Forecast, Prediction Techniques

I. INTRODUCTION

Weather forecasting is an important application of science for planning our day to day activities. Weather predictions are used in agriculture, air traffic, marine, forestry, military applications and utility companies. Previous year's weather data is used for the weather forecasting. Using Machine Learning Techniques, the weather forecasting process can be done easily. The weather datasets is gathered from the local and web resources. The framework like R, Python, MATLAB and the Machine Learning Techniques are used for future precipitation forecasting based on historical data. Statistcal methods are used for the prediction of future forecasting and possible climate change. The technique we used to build the model on the parameter of weather is regression with the help of R-packages in the Rstudio IDE.

II. RELATED WORK

In the previous years, there are number of models were designed to predict the future values of weather depending on the weather dataset.

ANKITA JOSHI, BHAGYASHRI KAMBLE [2] they proposed technique of decision tree algorithm. The problem is to predict the values of complicated weather phenomena with less data. To predict weather, they developed atmospheric models that are approximate by using mathematical equations. The accuracy of rainfall prediction is found to be 82%.

G VASMI KRISHNA [3], by using ARIMA model, calculate the value of regression and auto regression. Using correlation analysis, high correlation is considered to be most likely change in weather and it assumed to produced destructive effects. The ARIMA model is used to predict the future values of weather of next 15 days.

Nikhil Sethi, Dr.Kanwal Garg [4], the prediction of rainfall is done by statistical technique. The multiple linear regression technique is applied for the early prediction of rainfall. The results imply that there is a very minimal difference between values of the predicted and actual rainfall amount.

Manisha Kharola and Dinesh Kumar [5] implemented the back propagation algorithm. This algorithm produces accurate value of weather variable on small data set of imperfect dataset. The output in an error signal which produced to predict the future variable of weather with the help of it.

III. MULTIPLE LOGISTIC REGRESSION

Multiple logistic regression is used to predict the probability of the occurrence of an event using more than one explanatory variable.

Multiple logistic regression applies when there is a single dichotomous outcome and more than one independent variable.

The output of the logistic regression analysis is either 1 or 0, where output 1 implies success of the outcome of interest. If we define p is the probability which is equal to 1, Multiple Logistic Regression [6] model can be written as follows:

$$\hat{p} = \frac{\exp(b_0 + b_1X_1 + b_2X_2 + ... + b_pX_p)}{1 + \exp(b_0 + b_1X_1 + b_2X_2 + ... + b_pX_p)}$$

$\hat{p}$ is the probability which is success of an event (1); X1 to Xp are independent variables; and b0 to bp are the coefficients of the variable by regression.

IV. STEP BY STEP DESCRIPTION OF OPERATION

A. Data Collection

The weather data is collection of parameters like humidity, rainfall, temperature etc. which is recorded by different sensor for over past several years. These datasets can be found on the websites like Kaggle.com etc. in csv file. Then this data is send to for pre-processing step.

B. Pre-Processing

In this step, the removal of unwanted data or noise from the datasets took place. And the process of categorical and quantitative variable analysis is also done.

C. Data Transfer

Then the resulted is send for the statistical process.

D. Regression Technique

Then the data is divided into learning data and testing data to find the accuracy with k-cross validation. Then applied the

regression technique (multiple logistic regression) to build the model.

E. Prediction of Future Weather

According to the model we give the date as input for prediction of values of weather of that day and the outcome is the parameter values of weather on the given date.

V. RESULTS & CONCLUSION

In this paper the weather dataset with the parameter such as humidity, minimum and maximum temperature, and rainfall etc. of Canberra, Australia for 365 days (year) is considered. The prediction of the future values of weather is carried out for a specific date (or next no. of day) by the model produced by the multiple logistic regression applying on the input data. The model is used to produce the result for a specific data (only temperature) , the comparison between predicted results and real data is shown in figure 1 & 2.

Date	Location	MinTemp	MaxTemp	Rainfall	Evaporati	Sunshine	WindGust	WindGust	WindDir9	WindDir3	WindSpee	WindSpee	Humidity9	Humidity3	Pressure9	Pressure3	Cloud9am	Cloud3pm	Temp9am
11/1/2007	Canberra	8	24.3	0	3.4	6.3	NW	30	SW	NW	6	20	68	29	1019.7	1015	7	7	14.4
11/2/2007	Canberra	14	26.9	3.6	4.4	9.7	ENE	39	E	W	4	17	80	36	1012.4	1008.4	5	3	17.5
11/3/2007	Canberra	13.7	23.4	3.6	5.8	3.3	NW	85	N	NNE	6	6	82	69	1009.5	1007.2	8	7	15.4
11/4/2007	Canberra	13.3	15.5	39.8	7.2	9.1	NW	54	WNW	W	30	24	62	56	1005.5	1007	2	7	13.5
11/5/2007	Canberra	7.6	16.1	2.8	5.6	10.6	SSE	50	SSE	ESE	20	28	68	49	1018.3	1018.5	7	7	11.1
11/6/2007	Canberra	6.2	16.9	0	5.8	8.2	SE	44	SE	E	20	24	70	57	1023.8	1021.7	7	5	10.9
11/7/2007	Canberra	6.1	18.2	0.2	4.2	8.4	SE	43	SE	ESE	19	26	63	47	1024.6	1022.2	4	6	12.4
11/8/2007	Canberra	8.3	17	0	5.6	4.6	E	41	SE	E	11	24	65	57	1026.2	1024.2	6	7	12.1
11/9/2007	Canberra	8.8	19.5	0	4	4.1	S	48	E	ENE	19	17	70	48	1026.1	1022.7	7	7	14.1

Fig. 1: Sample Input of Weather Data

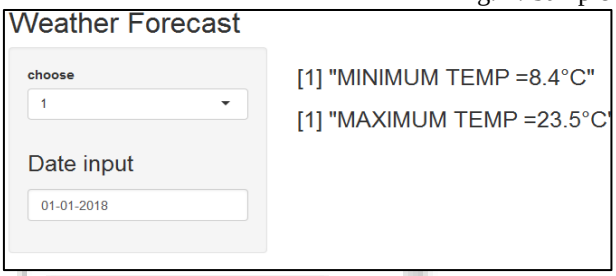

Fig. 2: Prediction of Values of Minimum & Maximum Temperature by MLR Model

REFERENCES

[1] Rob j Hyndman & Goerge "Forecasting: Principles and Practice" BOOK.
[2] Joshi A, Kamble B, Joshi B, Kajale K nd Dhange N, "Weather Forecasting And Climate Changing Using Data Mining Application" (2015) IJARCEE.
[3] G Vasmi Krishna, "An Integrated Approach for Weather Forecasing Based On Data Mining and Forecast Analysis" VOL-120 NO-11, JUNE 2015.
[4] N Sethi and DR. Kanwal Garg, "Exploiting Data Mining Technique for Rainfall Prediction", IJSCIT VOL-5(3), 2014
[5] Manisha Kharola & Dinesh Kumar, ISRO Journal of Computer Engineering (2014) – "Efficient Weather Prediction by Back-Propagation algorithm."
[6] Multiple Logistic Regression Analysis by Boston University School of Public Health (mobile page) January 17 2013.

