WEATHER PREDICTION USING MACHINE LEARNING AND STATISTICAL TECHNIQUES

Abstract

Weather prediction is the use of science and technology for forecasting atmospheric conditions for a specific time and location. People made an attempt to forecast the weather casually for prosperity and correctly since the 19th century. Climate can be predicted by selecting calculable data about current atmospheric state at a specific location and utilizing meteorology to design how atmosphere can vary. Once computed by hand on the basis changes in barometric pressure, present climatic conditions, and sky condition or cloud cover, weather prediction now depends upon computer-based models that take lots of atmospheric factors into consideration.

Keywords: Weather forecasting, Machine learning, linear regression, temperature prediction

Authors

Radhika Sreedharan
Assistant Professor
Department of Computer Science & Engineering
Presidency University, Bangalore, India
radhika.sreedharan@presidencyuniversity.in

Archana Praveen Kumar
Assistant Professor-Senior Scale
Department of Computer Science & Engineering
Manipal Institute of Technology
Manipal, India
archana.kumar@manipal.edu

I. INTRODUCTION

Forecasts on the basis of temperature and rainfall are significant for cultivation, and for businessmen within future exchanges. Temperature predictors are utilized by utility companies for calculating claims on upcoming days. On daily basis, people utilize weather forecasts to find what to wear on a given day. This paper focuses on analysis of the Delhi weather data from 2010-2017.Temperature, dewpoint, humidity are taken as observations from Delhi Weather dataset, and the forecast value is predicted using machine learning algorithms. By taking observations of rain, hailstorm, thunder, atmospheric pressure, temperature, humidity, dew point temperature, etc., the future climatic conditions can be predicted. The predicted value is matched with real-time data from where it is noticed that level component plays a significant role than trend and seasonal component in real-time data, and the predicted value is independent of dataset size.

II. LITERATURE REVIEW

Numerical weather prediction (NWP) has made outstanding enhancement for the past 50 years along with computer technology enlargement, modeling approaches, as well as observations (Molteni et al.,1996; Toth and Kalnay, 1997). In spite of that, NWP model predictions consist of systematized biases because of erroneous model physics, introductory conditions, and outline conditions (Paegle et al., 1997;Mass et al., 2002; Hart et al., 2003; Cheng and Steenburgh,2007; Rudack and Ghirardelli, 2010).

As the result of NWP and observations has different systematized errors, the prediction attainment for different regions, seasons and weather processes differ. Before the release of a weather prediction, a weather examination is compulsory for improving accuracy. For removing systematized fallacy and enhance the output from NWP models, a variance of post- processing processes were enlarged to reproduce climate examination(Wilks and Hamill, 2007; Veenhuis, 2013)—for e.g.,, model output statistics (MOS) (Glahn and Lowry, 1972; Cheng and Steenburgh, 2007; Wu et al., 2007; Glahn et al., 2009; Jacks et.al., 2009; Zhang et al., 2011; Glahn, 2014; Wu et al., 2016), the analog ensemble (Monache et al., 2013; Alessandrini et al.,2015; Junk et al., 2015; Plenković et al., 2016; Sperati et al.,2017), the Kalman filter (Delle Monache et al., 2011; Cassola and Burlando, 2012; Bogoslovskiy et al., 2016; Buehner et al.,2017; Pelosi et al., 2017), anomaly numerical- correction with observations (Peng et al., 2013, 2014), among which MOS is one of the most familiar utilized to yield impartial predictions.

(Glahn et al., 2009). MOS utilizes multiple linear regressions for yielding and enhanced prediction at distinct positions by making use of model forecast variables and previous detections as predictors (Marzban et al., 2006; Cheng and Steenburgh, 2007). MOS is an advantageous means and, t the time of 2002 Winter Olympic Games,MM5-based MOS exceeded the indigenous predictions yielded by MM5 and was equivalent to or more refined than human-generated predictions by the Olympic Forecast Team (Hart et al.,2003). Glahn (2014) utilized MOS with a diminish factor for forecasting temperature and dew-point, and demonstrated how comparable values of the diminish factor influence MOS temperature and dew-point predictions. (Glahn, 2014).Chattopadhyay et al. (2013) stated a nonlinear clustering technique for recognizing the structures of the Madden−Julian Oscillation (Chattopadhyay et al., 2013). Woo and Wong (2017) made use of optical flow techniques to radar-based rainfall forecasting (Woo and Wong, 2017). Climate examination data are matchless, and especially consisted NWP model data as well as observational data. They have distinct structures of data and features, which builds feature engineering a complex activity.

Deepak Ranjan Nayak surveyed about rainfall prediction using Artificial Neural Network which gave good results. Akash Dubey, made a prediction on Pondicherry about the rainfall.

III. METHODOLOGY

A Decision tree technique is used to predict the weather conditions based on the values of dew point temperature, humid, temperature, pressure. Weather conditions can be smoke, fog, haze, Mist, Light rain, Widespread dust etc.

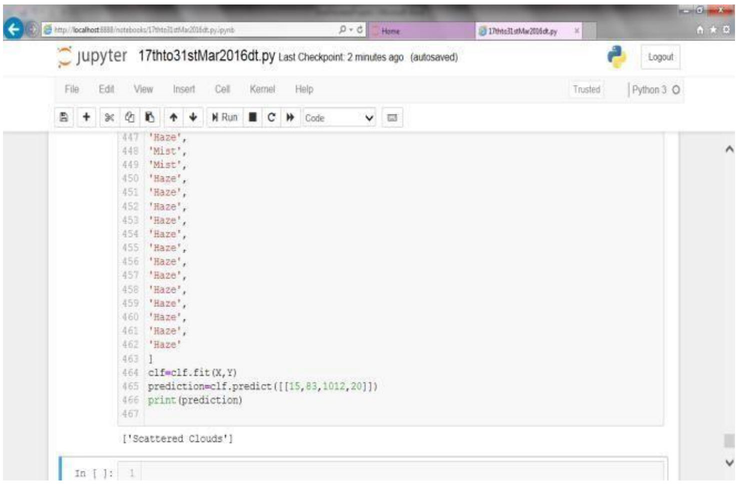

Figure 1: Decision tree to predict weather conditions

A Decision tree Regression algorithm is used to predict temperature based on the values of humidity, dew point temperature, etc.

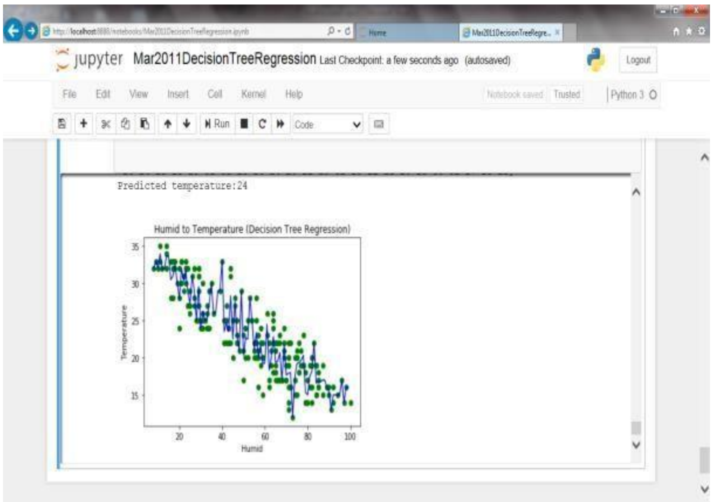

Figure 2: Decision Tree Regression plot between humid and temperature.

A statistical technique like variance is computed for parameters like atmospheric pressure, temperature, humidity. Correlation coefficient can be found between humidity and temperature. By comparing variances of above parameters monthly wise, yearly wise, future weather conditions can be predicted. With the help of Random Forest regression, temperature could be predicted with the help of humidity values.

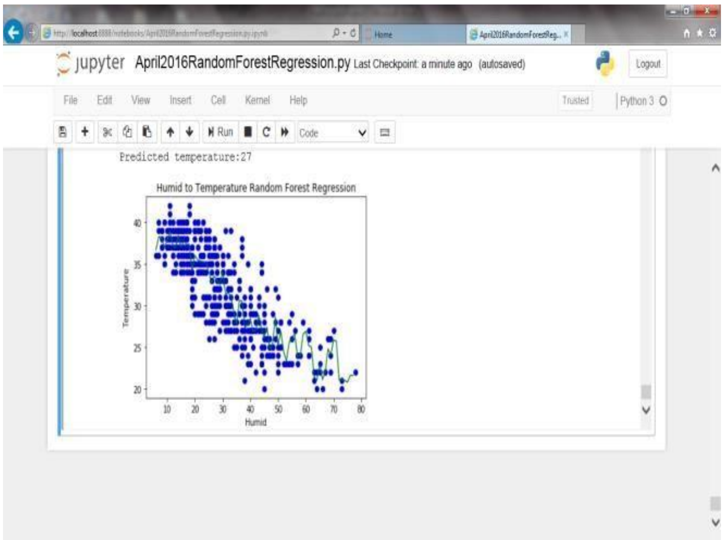

Figure 3: Random Forest regression plot to predict temperature based on humidity values.

K-Means plot is shown below which relates humidity and temperature.

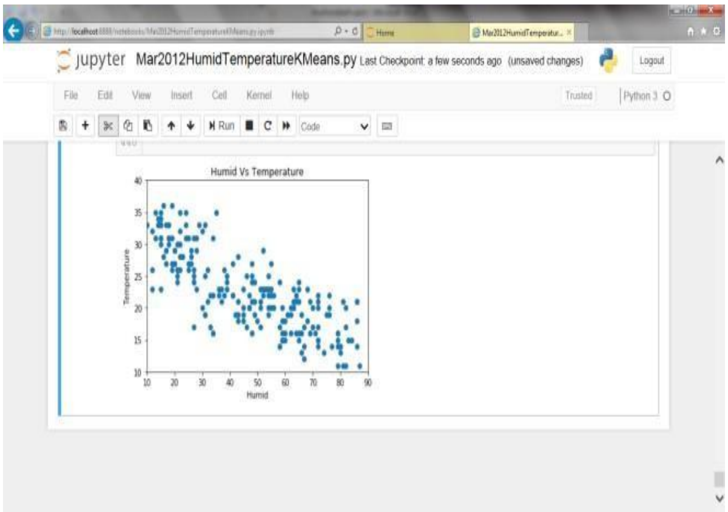

Figure 4: K-Means plot between humid and temperature

IV. RESULTS AND DISCUSSION

Matplotlib: This was used along-with numpy package for scatter plot creation

Numpy: For computing the regression coefficient, Pearson correlation coefficient, numpy package has been used. **sklearn**: For computing KMeans plot, sklearn import tree has been used.

Statistics: For computing variances in temperature, humidity, pressure , dew point temperature, import statistics was used. **Math:** For computing correlation coefficient between humidity and dew point temperature, import math was used. K-Means plots between humid and temperature are created and shown in Fig.4 and Fig.5.

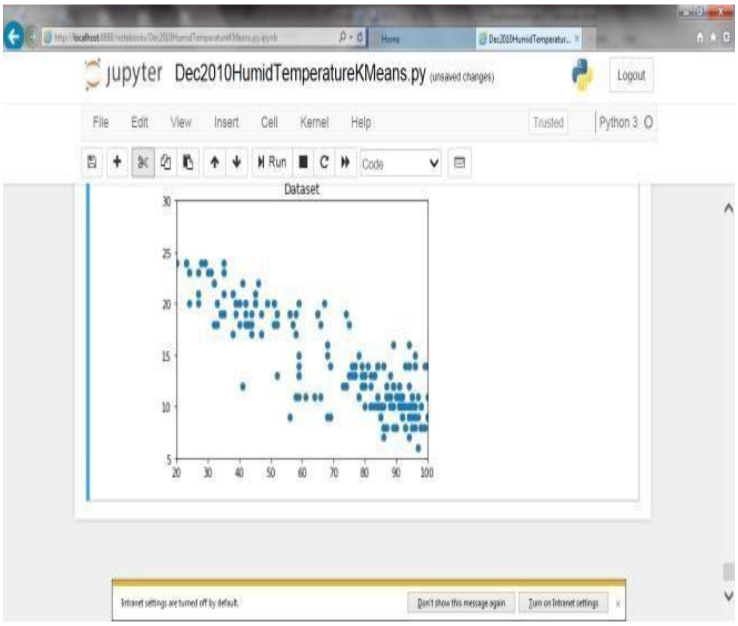

Figure 5: K-Means plot of humid vs temperature

Based on values of humidity, temperature could be predicted based on given data set as shown in Decision tree regression plot in Fig.2

Data Analysis for 2010

In Jan, there was negative correlation between dew point temperature and humidity. If dew point temperature increased, humidity decreased and vice-versa. From February to December, there was positive correlation between same parameters. If humidity increased, dew point also increased and vice-versa.

The variance of humidity increased from Jan to Feb, then, decreased in March, again decreased in April. However, in May, it was increased. Then, again in June and July, there was increase in humidity variance. Then, again it got reduced in August. However, in September, humidity variance was increased drastically, then reduced in October, then, again increased in November and then drastically increased in December. The variance of pressure increased from Jan to Feb, then increased a lot in March. It increased drastically in April. Then in May it increased by almost 4 times. The, it reduced a lot in June. Then, it was increased in July. Then, in August, it increased drastically. In the month of September, variance of pressure reduced by almost onefifth. Then, it was increased in October, then drastically increased in November and drastically increased in December.

The variance of temperature increased from Jan to Feb, and then increased in March. But it decreased in April, then again decreased in May and June. In July and August, it was decreased a lot. Then, in September, it was slightly increased. In October, it was increased a lot, then slightly reduced in November and then slightly increased in December.

Data Analysis for 2011

The K-means plot between humid and temperature for June is shown below:

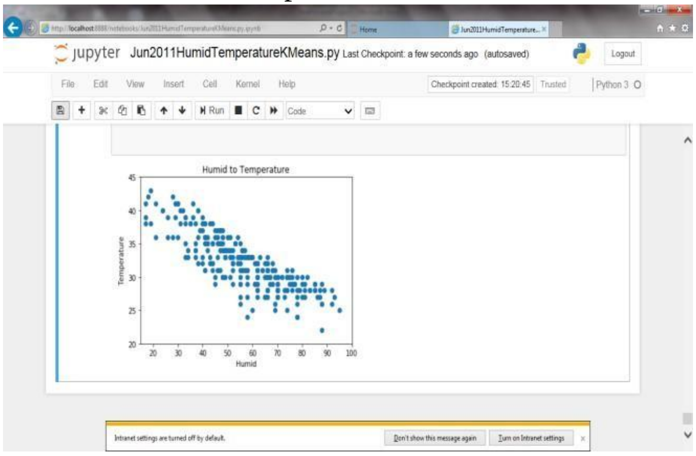

Figure 6: K-Means plot between humid and temperature in June 2010

The k-means plot between dew point temperature and pressure for August is shown below:

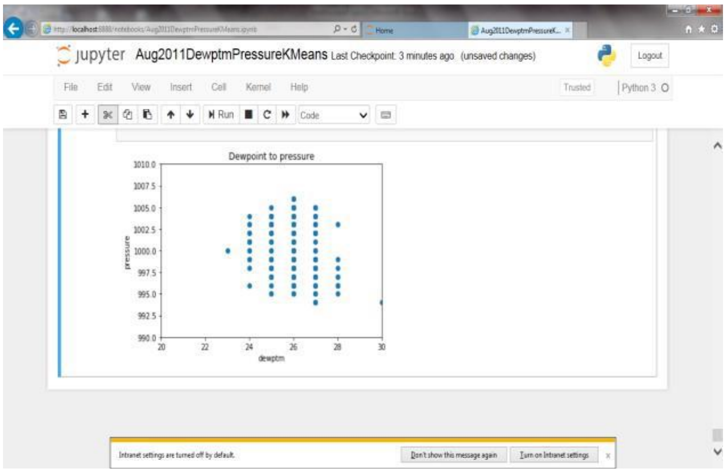

Figure 7: K-Means plot between dew point temperature and pressure in August 2010

Data Analysis for 2012

The K-means plot between dew point temperature and temperature for Feb is shown below:

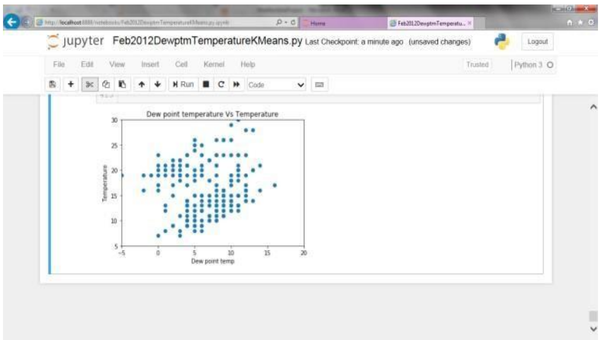

Figure 8: k-means plot between dew point temperature and temperature for Feb 2012

The k-means plot between humid and temperature for March is shown below:

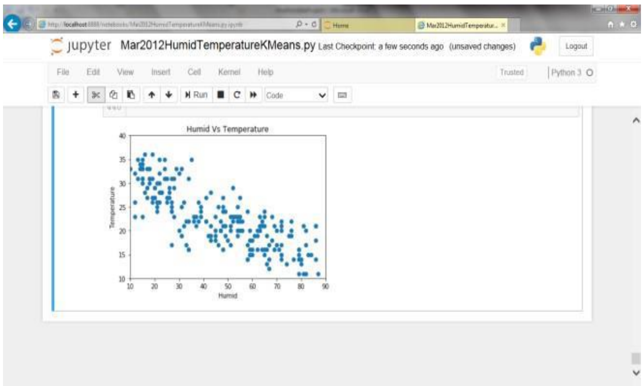

Figure 9: k-means plot between humid and temperature for March 2012

Data Analysis for 2013

Temperature was predicted using humidity values in Feb using Decision Tree Regression algorithm.

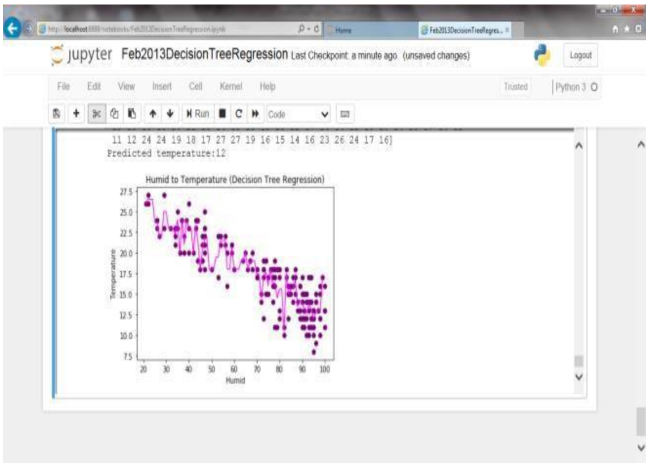

Figure 10: Decision Tree regression plot to predict temperature based on humidity values

Variance of dew point temperature decreased from June to July, then slightly increased in August and then again increased in September. It increased a bit more in October, then reduced in November, then again increased in December.

Variance of humid reduced drastically from June to July, the slightly reduced in August, then again drastically increased in September, then again increased in October and November, then slightly reduced in December.

Data Analysis for 2014

Temperature variance increased from January to February, then again increased in March, April, and then reduced in May, June, July and August.

Data Analysis for 2015

Variance of humid increased drastically from Jan to Feb, then was reduced in March, then reduced drastically in April and May, then increased in June, then again decreased in July, August, September, then again increased in October, November and December.

Temperature variance increased from Jan to Feb, then slightly increased in March, then slightly reduced in April, then slightly increased in May. Then again, it increased in June. In July, temperature variance was drastically reduced; then again, it was reduced in August but slightly increased in September, then drastically increased in October, then reduced slightly in November and then again increased in December.

Data Analysis for 2016

In 2016, variance of humid increased from May to June. Dew point temperature variance reduced from May to June. Variance of pressure increased drastically from May to June. Temperature variance reduced from May to June. For July 2016, temperature could be predicted using humidity values with the help of Support Vector regression.

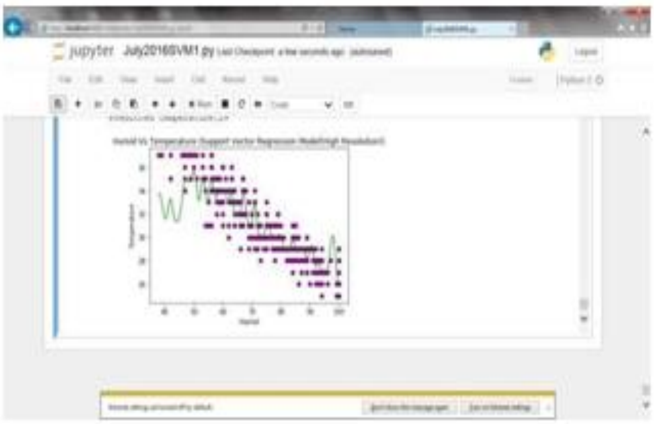

Figure 11: Support Vector regression plot to predict temperature based on humidity value in July 2016

In 2016, Dew point temperature variance increased from September to October, then slightly reduced in November, then slightly reduced in December. Humid variance increased from September to October, and then drastically increased in November, then again increased in December. Pressure variance increased from September to October, then very drastically increased in November, then very drastically decreased in December. Temperature variance increased from September to October, then again increased in November, the decreased in December. Weather condition was predicted for first 16 days of March using decision tree technique.

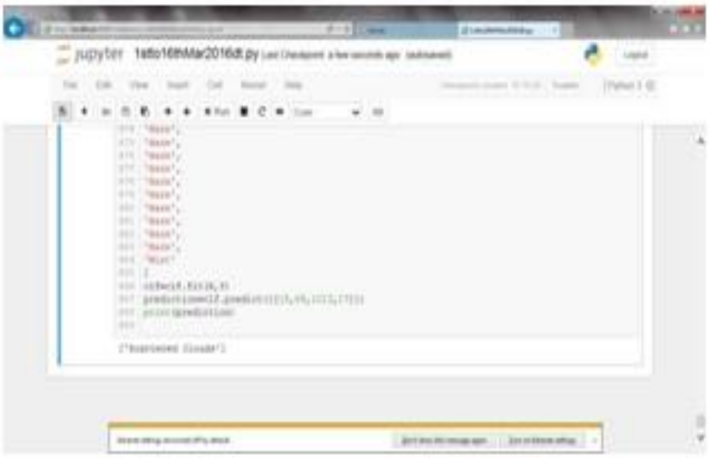

Figure 12: Prediction of weather conditions using Decision tree technique.

Data Analysis for 2017

The temperature in March could be predicted using humidity values using the Support Vector Regression Algorithm

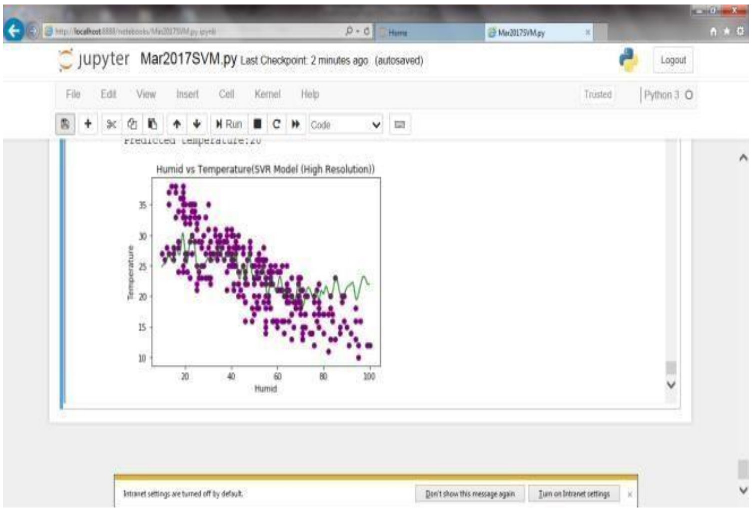

Figure 13: Support vector Regression plot to predict temperature based on humidity values.

V. CONCLUDING REMARKS

Algorithms like Support Vector Regression Algorithm, Random Forest Regression and Decision Tree Regression could predict temperature based on the values of humidity. K-Means plot could be plotted between temperature and humidity and pressure and dew point temperature. With the help of decision tree technique, weather conditions could be predicted based on the values of humidity, temperature, pressure, dew point temperature, etc. In some months, temperature and humidity were directly proportional whereas in certain months of the year, those two quantities were inversely proportional to each other. In future, we need to develop a model which can give results with better accuracy. By comparing climatic conditions of same months in consecutive years, the weather can be predicted for same month in the future. Weather conditions of respective months in future can be predicted with the given data set.

REFERENCES

[1] Li, H.[a,d], Yu, C.[a], Xia, J.[b], Wang, Y.[c], Zhu, J.[b], Zhang, P.-A Model Output Machine Learning Method for Grid Temperature Forecasts in the Beijing Area Advances in Atmospheric Sciences Volume 36, Issue 10, 1 October 2019, Pages 1156-1170

[2] Refonaa,J.[a],Lakshmi,M.[b],Abbas,R.[a],Raziullha,M.[c] –“Rainfall prediction using regression model”- International Journal of Recent Technology and Engineering Volume 8, Issue 2 Special issue 3, July 2019