EasyChair Preprint

№ 3565

Weather Forecasting using Linear Regression In Machine Learning

Om Narayan Jaiswal

June 7, 2020

Weather Forecasting using Linear Regression In Machine Learning

1. Abstract

Prediction requires accurate classification of data .In order to predict the uncertain things, we need to analyse various factors which involved either directly or indirectly. Weather is one of the most influential environmental constraints in every phase of our lives on the earth. So as to make everyday tasks we are very much rely on weather and need to know weather condition on before hands. This could be achieved by predicting the weather condition such as humidity, rainfall, temperature, thunder, fog, etc. This helps us in protecting ourselves from abnormal conditions and avoids unnecessary delays. The main objective of this paper is to design an effective weather prediction model by the use of multivariate regression or multiple linear regressions and support vector machine (SVM). As of now, there are various debates going on around the world either scientifically or non-scientifically regarding the change of Earth's climate in fore coming decades/centuries and what impact it will cause on all the living creatures. Scientific models which predict future climates offer the best plan or aspiration for providing the information which will allow the world's policy maker to take preventive measures and make better decisions for the future of the Earth and for the future lives. This paper explores about weather forecast in effective way.

Keywords: - Linear Regression, Coefficient correlation, Data preprocessing

2. Introduction

Weather forecasting is basically the prediction of the future weather and for the specified geographical location. Weather conditions are changing very rapidly around the world and it affects all the major areas. Weather forecasts become very essential in today’s world. Today

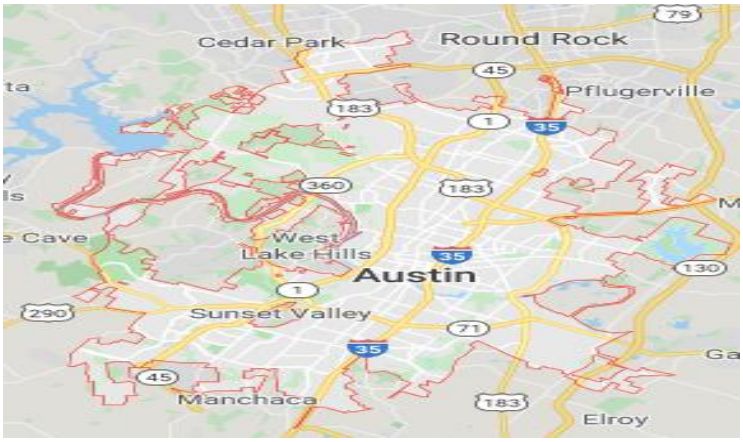
we are heavily depend on weather forecasts whether it is from industrial to agriculture, from travelling to daily commuting, anything where weather plays a role. For the easy and seamless mobility it becomes very important that we predict the weather correctly and ensure that it has no error.

The Scientists are still in working process of overcoming the limitations of computer models to improvise the accuracy rate of prediction through recent technologies of adding intelligence to machine. To add intelligence for system as human we have given a study platform called Artificial Neural networks, Machine learning, rule based techniques where there exist ample impetus to study the weather occurrence and prediction.

Here we implemented machine learning algorithm is Scikit's learn linear regression model. A large historical datasets of Austin was obtained and used to train this algorithm. The input was the weather conditions of Austin of many years and thus the output obtained is weather condition of upcoming days.

3. Datasets and Implementation

Datasets: - Once the data was collected, we split that data into training set and test set. However target variable is always the same. The training set contains 10 years of data starting from 2013-02-21 to 2017-07-31. In contrast, test set is the next day after 2017-07-31. Thus trained model predict weather condition on the inputting of training data. Different weather conditions and factor were chosen for this model. Daily weather data for the Austin KATT station from 2013-02-21 to 2017-07-31. In this, dataset taken into consideration is from kaggle.com of Austin weather. Some glimpse of data is as follows: -



	Date The date of the collection (YYYY-MM-DD)	TempHighF High temperature, in degrees Fahrenheit	TempAvgF Average temperature, in degrees Fahrenheit	TempLowF Low temperature, in degrees Fahrenheit
1	2013-12-21	74	60	45
2	2013-12-22	56	48	39
3	2013-12-23	58	45	32
4	2013-12-24	61	46	31

Implementation: - Most important thing in model is the implementation of that model. So in this model various steps were followed to get the desired output. (A) Loading of historical data. Important thing is that to define a dataset using which one can analyze and perform some action. (B) Data Cleaning – in this we drop unwanted data, remove miscellaneous data and place a zero in that place where data entry is null. Also allocate index to every row which later on makes traversing easier. (C) Training Model – After data cleansing we proceed we updated data, thus by the use of that data we train our model for our target. (D)Prediction and result – This is the last step where one get result in form of prediction for which model is created. Thus on providing the feature or input or dependent variable, one can get the target or independent variable outcome.

4. Related Work

The impact of "weather conditions" on transportation systems could be a general term that will pose some confusion. Researchers have used different classification schemes for weather conditions, because these conditions differ considerably both in type and magnitude. Some weather conditions are harsh in nature and thus may activate a different response by the drivers.

Such extreme conditions are outside the immediate focus of this study. Other inclement weather gives a less compressed timeframe to the choice makers, and allow drivers to retain an appropriate amount of control on their vehicles; this control could also be but under "normal everyday" situation thanks to physical factors like visibility, physical discomfort (cold or hot temperatures) and reduced pavement friction with the tires when there's precipitation or icy condition prevail.

Most existing studies don't describe all "weather conditions" within the sort of measurable objective parameters, making it difficult to elucidate or quantify the effect of such conditions on the mobility systems and their users.

But there is some other dimension which we need to consider such as:

- ✓ Severity
- ✓ Duration
- ✓ Area
- ✓ Demand

5. Proposed Model

The proposed model consists of collected historical weather data that includes various important factors responsible for the weather change that includes the temperature, both maximum and minimum temperature, the moisture and humidity in the atmosphere, precipitation, dew point, sea level, visibility, etc.

4.1 Data preprocessing:-In this proposed model the collected dataset is segregated into the parts which in use and which aren't of any use to the machine learning model. After that the dataset goes through the data preprocessing part wherein the data is passed on to a process where the missing and the error values in the dataset are replaced by the mean values or the most occurring value in that filed. Other way is not to consider those values and replacing those empty values with EAN and the carrying out the other tasks.

After the data preprocessing is completed there comes the part wherein cleaned dataset is segregated into two parts namely training set and test set. The training set is used to train machine learning model to teach and compute results. Testing set is used to find results and comparing actual value with calculated value. Using the error value as the benchmark to teach machine learning model further.

The training phase contain fold cross validation wherein the dataset is divided in to k sets k times and then the dataset is divided into test and training sets such that in a set training sets are selected randomly and then in sets the model is trained. The kth set is then used as the test set for the trained machine learning model. This technique not only does help us to reduce the condition of under fitting and condition of over fitting as well.

4.2. Linear Regression: - linear regression is a linear approach to modeling the relationship between a scalar response and one or more explanatory variables. The case of

one explanatory variable is called simple linear regression. For more than one explanatory variable, the process is called multiple linear regression.

When a **correlation coefficient** shows that data is likely to be able to predict future outcomes and a scatter plot of the data appears to form a straight line, you can use simple linear regression to find a predictive function.

```
# the features or the 'x' values of the data
# these columns are used to train the model
# precipitation column
X = data.drop(['PrecipitationSumInches'], axis = 1)
# temperature
X1 = data.drop(['TempAvgF'], axis = 1)
# Humidity Percent
X2 = data.drop(['HumidityAvgPercent'], axis = 1)
```

```
Main purpose of this paper is to show prediction of weather by the
means of machine learning techniques.
```

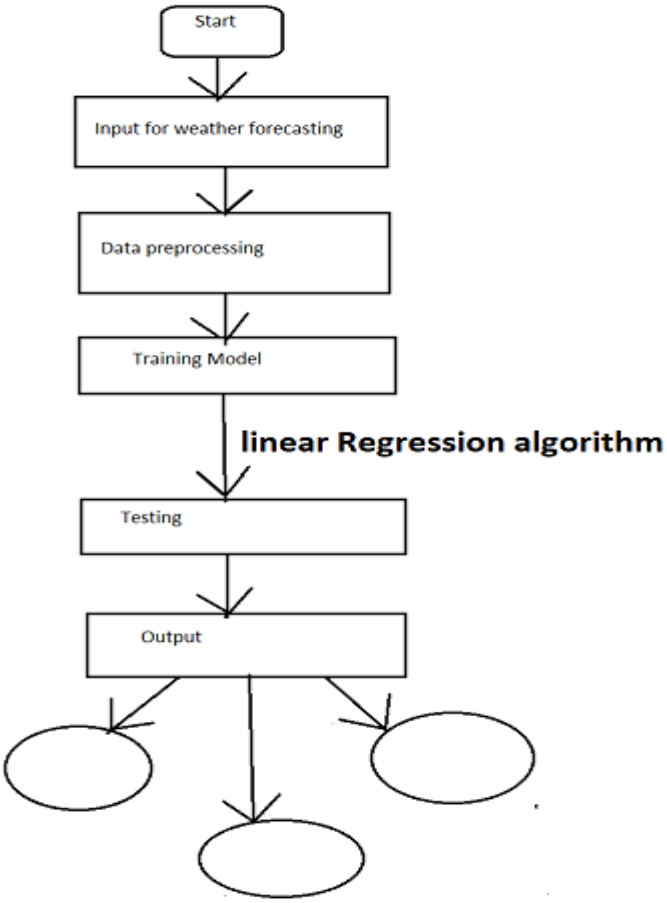


Figure shows flow chart of Model

6. Result

The goal of this work is to predict various weather condition using Linear Regression thus by the use of this model one can predict the temperature, humidity, precipitation, etc. This model can become more accurate by the training of large datasets .So it can easily be said how these parameters helps in prediction of weather.

Here in provided chart in precipitation level dark green and in temperature light green shows the plotting of the data onto the map whereas in precipitation level blue and in temperature red shows the predicted value for the input day on the basis of given data.

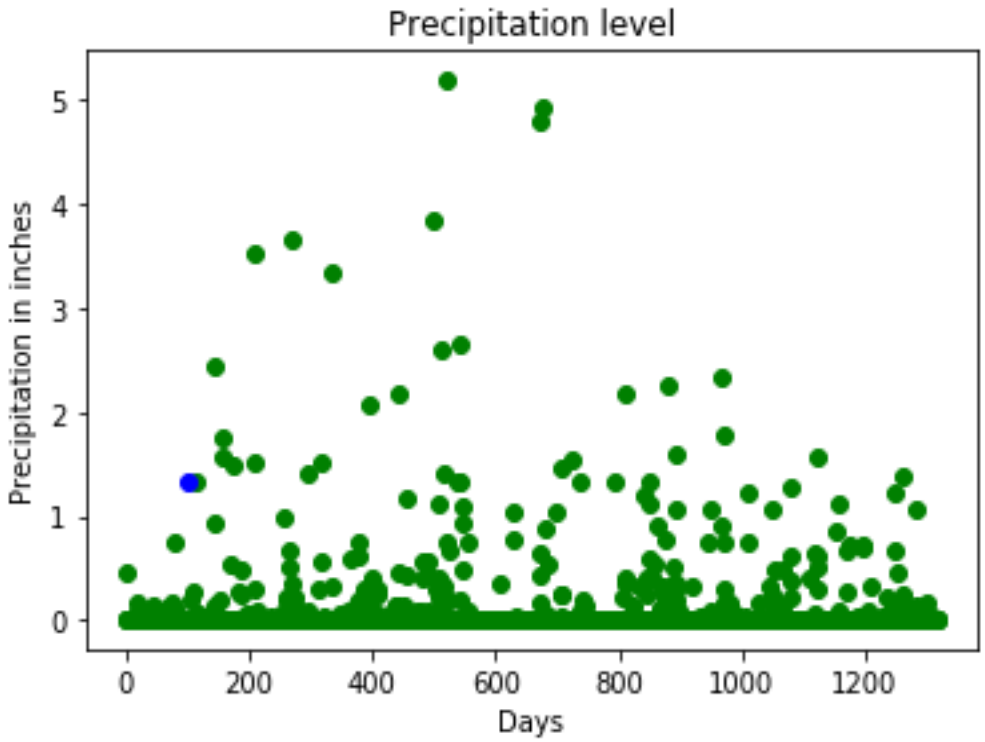


Fig 1: - shows predicted result of precipitation level

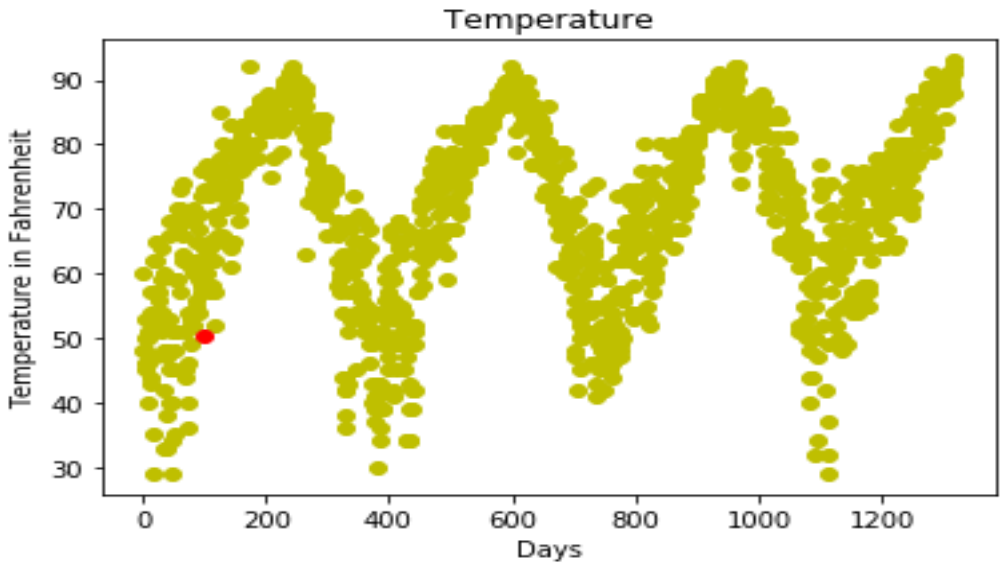


Fig 2: - shows the predicted result of temperature

7. Conclusion

Therefore in this we provided how the machine learning techniques can be trained and used for the weather forecasting. In this Machine learning models are much accurate than human prediction and physical models prepared by human. For weather forecasting we used dataset for the Austin KATT station from 2013-02-21 to 2017-07-31. Accuracy obtained here was measured on the basis of coefficient correlation. We also utilize the historical data to predict the weather conditions which is much faster than the traditional models. The new pattern is combining deterministic and machine learning or statistical components, can provide fast and accurate calculations of these processes as well and help in predicting value of independent variable accurately. In future work, going to do research and make a model on how the neighboring weather can affect the weather of our area

8. References

1. Dixian Zhu , Changjie Cai , Changjie Cai , Changjie Cai –A Machine Learning Approach for Air Quality

Prediction: Model Regularization and Optimization| MDPI December 24, 2013

2. A H M jakaria, Mosharaf Hossain, Mohammad Ashiqur Rahman –Smart weather forecasting using Machine Learning techniques. |November.2018
3. International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-7 Issue-6C, April 2019
4. International Journal of Pure and Applied Mathematics Volume 118 No. 22 2018, 1945-1949 ISSN: 1314-3395 (on-line version) url: http://acadpubl.eu/hub