Rainfall Prediction System

DeepKumar Patel
AU1841058

Devarsh Suthar
AU1841068

Dhruv Panchal
AU1841081

Vishvas Patel
AU1841083

Abstract—**India is primarily an agricultural country, with crop productivity and rainfall playing a large role in its economy. Rainfall prediction is required and mandatory for all farmers in order to analyze crop productivity. Predicting the condition of the atmosphere using science and technology is known as rainfall prediction. Hence, rainfall prediction becomes essential as heavy rainfall can lead to many disasters. Hardware components often fail to predict rainfall accurately. As a result, using machine learning techniques, farmers will benefit greatly from accurately predicted rainfall results. For these purposes, we have compared and evaluated various regression models in this paper and analyzed their regression scores.**

Keywords—***Rainfall | Prediction | Machine Learning | Regression | Accuracy***

I. Introduction

Rainfall prediction is a major problem for the meteorological department as it is closely associated with the economy and human life. Accuracy of rainfall forecasting has great importance for countries like India whose economy is highly dependent on agriculture. There are hardware devices for predicting rainfall by using the weather conditions like temperature, humidity, etc. But these traditional methods cannot work efficiently. However, by using machine learning techniques we can produce accurate results by analyzing historical rainfall data. We extracted data of Ahmedabad city from the past 11 years (i.e, 2009-2020) from an API [2]. The extracted data includes hourly forecasts for parameters including precipitation(MM), temperature, wind speed, pressure, visibility, and more. There were many redundant columns in the rainfall data. These unnecessary columns were discarded. We performed feature selection for our model using the correlation method. For that, we found highly correlated features among all the features. If two of the features are highly correlated (i.e, if the correlation is > 90%), we discarded one of the features. The heat map for correlation between features is shown in the article. Further, we defined labels and features for our dataset. After splitting our data into training and testing categories, we applied different regression models on our dataset such as linear, ridge, lasso and logistic regression. The regression score for each regression method was tested and analyzed. The optimal regression model can be found by comparing the regression score of all the models that have been used. To compare outcomes, we used linear, ridge, lasso and logistic regression. The graphs for the same are depicted below in the results section of this article.

II. Literature Overview

Maulana, Rositha .et al [1] has described the rainfall prediction using monthly data from 1901-2009 using regression methods like Multiple linear regression, Support vector machine, Lasso regressions and obtained 99% of accuracy.

Nikhil Oswal [2] has described the rainfall prediction using date wise data using various machine learning techniques and models like Logistic regression, Decision tree, k-mean, Random forest and obtained 84% of accuracy.

III. Implementation

A. Platform

We have used python as a platform to perform regression analysis and comparison.

B. Data Preparation

- We have generated an API on a world weather online platform to generate Ahmedabad Rainfall data. We wrote a python script and gained day wise data from 2009-2020.
- After converting the data into CSV files we removed some columns(features) which are irrelevant (i. e SnowFall_cm : this column shows snowfall information for the day in centimeters, This is irrelevant because snowfall events never happen in ahmedabad.)
- Then we separated dependent(label) and independent variables(Features).
- Now, we started our coding. We perform some data preprocessing tasks to optimise output. We performed auto correlation to our dataset. This task will give us the highly correlated columns. We remove these highly correlated columns, because high correlation affects the prediction value.

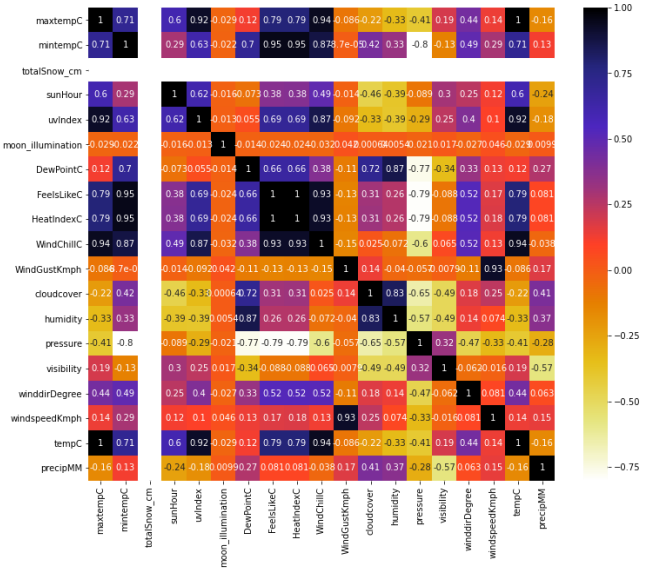
- After removing correlated features we performed normalization. This converts a wide range of data in between 0 and 1 so machines can learn optimally.

C. Regression Model

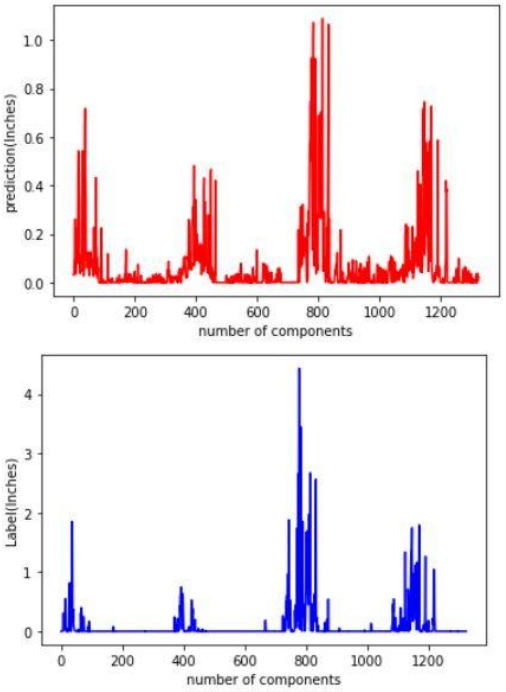
- Then we splitted data into training and testing data to perform linear regression. [Figure 3]
- We further applied different regression models on our dataset such as linear, ridge, lasso and logistic regression.
- The regression score for each regression method was tested and analyzed.
- To compare outcomes, we used linear, ridge, and lasso regression. However, since rain precipitation has several spikes (for example, in the monsoon, some days there is no rain at all and the next day it is heavy rain). [Figure 2]
- To achieve an accuracy of nearly 90%, we used logistic regression and compact out classification in binary form (i.e. rain vs. no rain). [Figure 3]

IV. Results

Feature Correlation Heatmap:

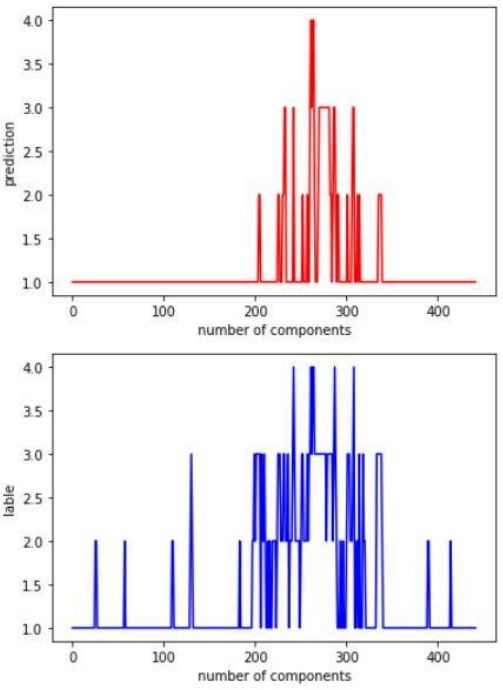


Regression analysis (Predicted rainfall vs Label):



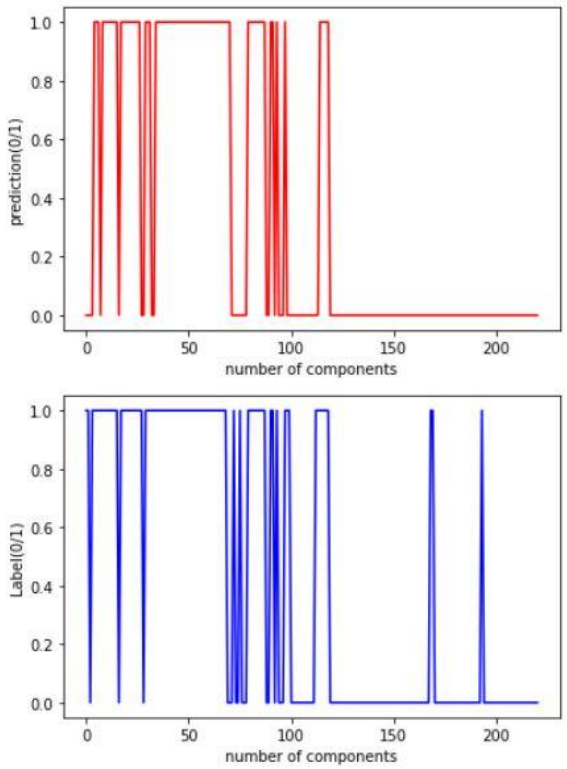
[Figure 1]

- **Figure 1 displays a linear regression prediction graph (red). A label value is represented by the blue graph.**



[Figure 2]

- **Figure 2 is logistic regression with 4 combinations: no rain, drizzle, moderate rain, and heavy rain. The red graph represents predictions and blue graph represents the label values.**



[Figure 3]

- **Figure 3 is logistic regression with 2 combinations: no rain and rain. The red graph represents predictions and blue graph represents the label values. (more than 90% accuracy)**

V. Conclusion

We have analysed the regression score of various regression models. Further after selecting the optimal regression model, the accuracy of this regression model can be increased. The rainfall prediction is required and mandatory for all farmers in order to analyze crop productivity. The aim of rainfall prediction is to provide information to farmers and citizens so that they can reduce weather related losses. As can be seen from the above data, the logistic regression has given better accuracy of more than 90% so far.

References

[1] http://www.ijstr.org/final-print/jan2020/Prediction-Of-Rainfall-Using-Machine-Learning-Techniques.pdf
[2] https://towardsdatascience.com/obtain-historical-weather-forecast-data-in-csv-format-using-python-5a6c090fc828
[3] https://arxiv.org/pdf/1910.13827v1.pdf