Project Report on Rainfall Prediction for Mehsana District

Submitted By:
Mehul Katara(1421002)
Riddhi Patel(1421013)
Sahil Desai(1421014)
Saurabh Chauhan(1421015)

Submitted To:
Prof. Sanjay Chaudhary

Contents

List of Figures

1 Introduction

India is basically an agricultural country and the success or failure of the harvest and water scarcity in any year is always considered with the greatest concern.India's agriculture production is highly dependent on its precipitation behavior of the monsoon rainfall. Monsoon is the main source of water. Average rainfall prediction is a prime important factor for crop planning. India receives 80%of rainfall during June to September. These problems are closely linked with the behavior of the monsoon rains in India.An accurate prediction of rainfall is crucial for agriculture based Indian economy. Moreover, it also helps in the pre- vention of flood, the management of water resources, and generating recommendations related to crop for farmers. The variability of rainfall in both time and space makes the rainfall prediction a challenging task. Moreover, the meteorological parameters needed for the rainfall prediction are complex and nonlinear in nature.

Applications of science and technology can be applied to predict the weather of a given location. Attempts for prediction started informally years before and formally since the nineteenth century. Rainfall prediction is made by collecting huge amount of numerical data of the given state of the atmosphere as corpus to show how the atmosphere will be on that place on the basis of current atmospheric parameters. Different types of forecasting methods are employed all over the world.The weather forecasts are divided into the following categories.

- Now casting - in which the details about the current weather and forecasts up to a few hours ahead are given.
- Short range forecasts (1 to 3 days) in which the weather (mainly rainfall) in each successive 24 hr intervals may be predicted upto 3 days. This forecast range is mainly concerned with the weather systems observed in the latest weather charts, although generation of new systems is also considered.
- Medium range forecasts (4 to 10 days) Average weather conditions and the weather on each day may be prescribed with progressively lesser details and accuracy than that for short range forecasts.
- Long range /Extended Range forecasts (more than 10 days to a season). There is no rigid definition for Long Range Forecasting, which may range from a monthly to a seasonal forecast.

A wide range of rainfall forecast methods are employed in weather forecasting at regional and national levels. Fundamentally there are two approaches to predict Rainfall. They are Empirical and Dynamical Methods.

In empirical approach, an analysis of historical data of the rainfall is carried out and its relationship to a various atmospheric and oceanic variables over different parts of the world is determined. The widely used approaches are regression, artificial neural network, fuzzy logic and group method of data handling.

The dynamical approach, predictions are generated by physical models based on system of equations that predict the future Rainfall. The forecasting of weather by computer using equations are known as numerical weather prediction. To predict the weather by numeric means, meteorologist has develop atmospheric models that approximate the change in temperature, pressure etc using mathematical equations.

In our Project ,Rainfall prediction is implemented with the use of empirical statistical technique. Our dataset is from (1996-2014) having attribute such as minimum temperature, maximum temperature ,sea level pressure, wind speed,relative humidity etc.

This model is going to forecasts month-wise and year-wise rainfall(in mm).The resulted rainfall amounts are intended to help farmers in making decision concerning with their crop.Since rainfall is one of the causes of possible calamities like floods and typhoons, predicting the occurrence of rainfall will help us to be prepared for these possible calamities.

The basic procedures involve are firstly identifying an initial model,secondly repeatedly changing the model by removing a predictor variable based on a criteria and then terminating the process when we get a model which fits the data well.

2 The need for Data Science (Model) and Analytics

Analysis of data is a process of inspecting, cleaning, transforming, and modeling data with the goal of discovering useful information, suggesting conclusions, and supporting decision-making. Data analysis has multiple facets and approaches, encompassing diverse techniques under a variety of names, in different business, science, and social science domains.

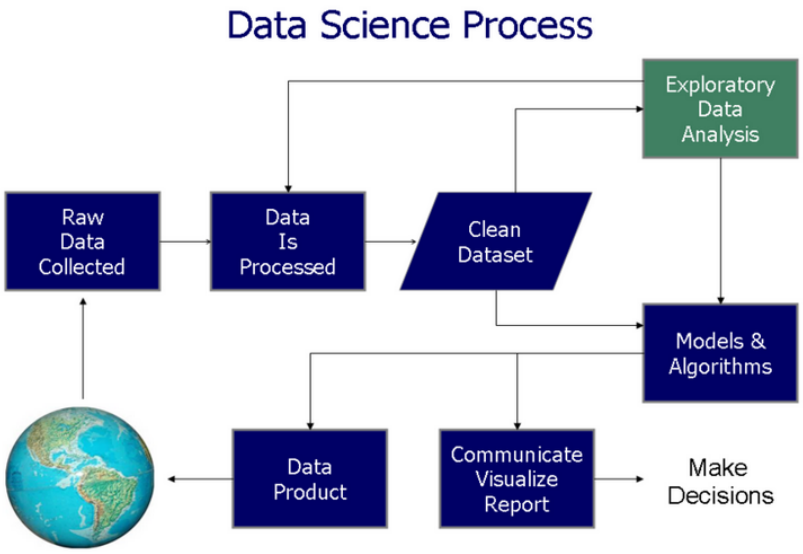

Figure 1: Data Science Process

Data analysis is a process for obtaining raw data and converting it into information useful for decision-making by users.Data is collected from a variety of sources.Data initially obtained must be processed or organized for analysis.Data cleaning is the process of preventing and correcting these errors.Once the data is cleaned, it can be analyzed. Analysts may apply a variety of techniques referred to as exploratory data analysis to begin understanding the messages contained in the data.

Data science is the practice of deriving valuable insights from data. Data science is emerging to meet the challenges of processing very large data sets i.e. "Big Data" consisting of structured, unstructured or semi-structured data that large enterprises produce. A domain at center stage of data science is the explosion of new data generated from smart devices, web, mobile and social media. Data science requires a versatile skill-set. Many practicing data scientists commonly specialize in specific domains such as the fields of marketing, medical, security, fraud and finance.

3 Challenges

Long-term rainfall prediction is a challenging task especially in the modern world where we are facing the major environmental problem of global warming.In general, climate and rainfall are highly non-linear phenomena in nature exhibiting what is known as the "butterfly effect". While some regions of the world are noticing a systematic decrease in annual rainfall, others notice increases in flooding and severe storms. The global nature of this phenomenon is very complicated and requires sophisticated computer modelling and simulation to predict accurately.

The prediction of Indian Summer Monsoon rainfall has remained a challenge over the decades and skill of empirical prediction models have not improved over the years. In addition, it is equally important to predict the extremes of monsoon climate such as droughts and floods. Empirical models systematically fail to predict the extremes. Further, the long range prediction of the seasonal mean monsoon depends on dynamics of its year-to-year variations.

4 Architecture of the proposed system

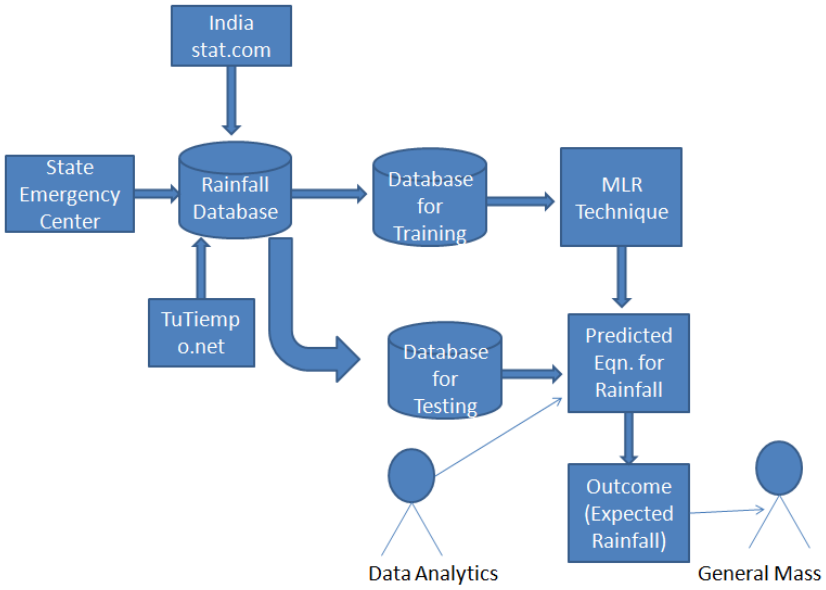

Figure 2: architecture of the rainfall prediction

Data used in the present study is collected from State Emergency Operation Centre,Revenue Department,Gandhinagar , www.Indiastat.com and www.TuTiempo.net.Using collected data, Data is separated into training dataset and testing dataset.Training data set is used to build up a model, while test dataset is to validate the model .So here we develop MLE model using Training data set and test data set is used to validate the same model.

In our project , using MLR model we will predict the rainfall.Once we predict the rainfall we can compare with the actual value of the same so that we can find the efficiency of the model.Here we have implemented model in R language so Goodness of fit in terms of Multiple R Squared, Correlation Coefficient and MSE(Min Square Error).

Finally Predicted value of rainfall are intended to help General Mass, particularly farmers because they are depend upon the rain and so that they have a idea about when to plant, which type of crop to plant and when to cultivate the plant.

5 Methodology/Algorithms

- Time-series Analysis
- K- Nearest Neighbour algorithm
- Decision Tree
- Nave Bayes
- Artificial Neural Network
- Fuzzy logic

In our Project ,Rainfall prediction is implemented with the use of Time-series Analysis with the help of concept of multiple linear regression and moving avgerage.

5.1 Multi Linear Regression

Regression is a statistical empirical technique that utilizes the relation between two or more quantitative variables on observational database so that outcome variable can be predicted from the others. One of the purposes of a regression model is to find out to what extent the outcome (dependent variable) can be predicted by the independent variables. The strength of the prediction is indicated by adjusted R^2 , also known as variance explained or strength of determination. It is a technique widely used in business, the social and behavioural sciences ,climate prediction.Building a Regression model is an iterative process that involves finding effective independent variable to explain the process ,we are trying to model /understand. Then running regression tool to determine which variable are effective predictors then removing variable(s) until we find the best model possible. A multiple linear regressions analysis is work out to predict the values of a dependent variable, Y, given a set of explanatory variables $(X_1, X_2.....X_n)$.

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_n X_n$

where y is the dependent variable,β_0 is intercept and $\beta_1, ..., \beta_n$ are slope ,$X_0, X_1, ...X_n$ are independent variables.

Defining Models in R

To complete a linear regression using R it is first necessary to understand the syntax for defining models.The basic syntax for a regression analysis in R is **lm(Y ∼ model)** ,where Y is the object containing the dependent variable to be predicted and model is the formula for the chosen mathematical model. The command lm() provides the models coefficients but no further statistical information. To obtain a more complete statistical summary of the model, for example, we use the summary() command.The section of output labeled Residuals gives the difference between the experimental and predicted signals. Estimates for the models coefficients are provided along with the their standard deviations (Std Error), and a t-value and probability for a null hypothesis that the coefficients have values of zero. In this case, for example, we see that there is no evidence that the intercept (β_0) is different from zero and strong evidence that the slope (β_1) is significantly different than zero.

5.2 Moving average

In statistics, a moving average (rolling average or running average) is a calculation to analyze data points by creating a series of averages of different subsets of the full data set. It is also called a moving mean (MM) or rolling mean.A moving average is commonly used with time series data to smooth out short-term fluctuations and highlight longer-term trends or cycles.Given a series of numbers and a fixed subset size, the first element of the moving average is obtained by taking the average of the initial fixed subset of the number series. Then the subset is modified by "shifting forward", that is excluding the first number of the series and including the next number following the original subset in the series. This creates a new subset of numbers, which is averaged. This process is repeated over the entire data series.

We have not used inbuilt function (ARIMA) of R for the implementation of the Moving Average.We have develop our own method to implement the same in the R using hard coding with the help of Moving Average theory.We have also examined variations in the implementation of the Moving Average .We used divide by 3 ,divide by 4 and divide by 2 to implement the concept of Moving Average and try to identify the hidden pattern inside that plot.

For example,
Expected Rainfall for June 2012=(Ranifall for June 2009+Ranifall for June 2010+Ranifall for June 2011)/3

Expected Rainfall for July 2012=(Ranifall for July 2009+Ranifall for July 2010+Ranifall for July 2011)/3
and so on.

6 Technology Stack for the project

The model is built using R and requires 4 GB RAM and 2.4 GHz processor.

R is a powerful environment for statistical computing which runs on several platforms.R is available free of charge and is distributed under the terms of the Free Software Foundation's GNU General Public License. You can download the program from the Comprehensive R Archive Network (CRAN).

R is a software language for carrying out complicated (and simple) statistical analyses. It includes routines for data summary and exploration, graphical presentation and data modelling. it's a programmable environment that uses command-line scripting, you can store a series of complex data-analysis steps in R. That lets you re-use your analysis work on similar data more easily than if you were using a point-and-click interface

7 Data sets

- Actually Prediction of Rainfall depends on a no of attributes like,
  - Hr = Hour In Utc (03 Utc = 0830 Ist, 12 Utc = 1730 Ist)
  - Wk = Week No. (1-52)
  - Wslp = Station Level Pressure In Hpa
  - N = Number Of Observations In The Week
  - WMSLP = Mean Sea Level Pressure In Hpa Or Height Of The Station In Geo-Potential Meters (Gpm), If Station Height Is Above 800 Gpm
  - Wdbt = Dry Bulb Temperature In Degrees C
  - Wwbt = Wet Bulb Temperature In Degrees C
  - Wdpt = Dew Point Temperature In Degrees C
  - Wrh = Relative Humidity In
  - Wvpr = Vapour Pressure In Hpa
  - Wsp = Wind Speed In Kmph
  - As = Average Wind Speed In Kmph
  - Wlc = Low Cloud Amount In Oktas
  - Wmc = Medium Cloud Amount In Oktas
  - Whc = High Cloud Amount In Oktas
  - Wtc = Total Cloud Amount In Oktas
  - Wtrf = Total Weekly R/F In Mm
  - Wevp = Evaporation In Mm
  - Wtmp = Water Temperature In Degrees C
- But out of them We are going to consider on following attributes.
  - Min temp.
  - Max temp.
  - Mean temp.
  - Mean Humidity(%)
  - No of Rainy days
  - Sea Level Pressure(hPa)
  - Wind Speed(km/hr)

Rainfall related data is collected from State Emergency Operation Centre,Revenue Department,Gandhinagar. Data related to Min temp.,Max temp.,Mean temp.,Humidity and No of Rainy days are collected from www.indiastat.com(Through DAIICT IP Login) and www.TuTiempo.net.

- Year wise and Month wise dataset from 1996-2013 and 2002-2014 respectively are shown below.

Data Set (Multiple Regression) is shown Below

A	B	C	D	E	F	G	H	I
Year	Mean T	Max T	Min T	Rainfall (mm)	Mean Humidity	No_rainyday	wind speed(km/hr)	Sea level Pressure(hPa)
1996	27.1	34.3	21	432	56.2	61	7.9	1006
1997	26.5	33	20.8	1361	60.9	80	7.6	1008
1998	27.3	34.4	21	941	57.3	65	7.3	1007
1999	26.9	34.3	20.5	371	55.5	50	6.8	1007
2000	27.6	35.1	20.3	289	48.3	38	7	1007
2001	27.1	34.4	20.5	626	50.2	66	7.8	1008
2003	27.2	34.4	21.2	822	48.1	63	8.4	1008
2004	27.3	34.6	21	565	54.1	6	8.6	1008
2006	27.4	34.3	21	1421	57.8	68	7.7	1008
2007	27.5	34.5	21.1	1119	56.3	67	7.9	1007
2008	26.6	33.9	20.7	619	56.9	62	6.8	1007
2009	28	35.3	21.8	459	52.1	42	10.4	1007
2010	27.8	35.1	21.9	783	52.1	87	12.8	1007
2011	28.9	35.3	21.3	863.03	51.4	69	8.2	1007
2012	27.9	35	20.6	578.81	51.8	61	8.4	1007
2013	27	34.1	20.6	1174.59	57.9	80	8.8	1007

Figure 3: Year wise Rainfall Data(1996-2013)

A	B	C	D	E	F	G	H
Monthyear	Rainfall (in M	Mean T	Max T	Min T	Mean Humidity	wind Speed(km/hr)	Sea Level Pressure(hPa)
Jun-02	120.2	32.3	38.7	27.5	54.3	11	1002
Jun-03	63.6	31.6	38.8	27.2	55.1	15	1000
Jun-04	123.8	31.9	38.3	26.6	59.4	14	1002
Jun-05	172.3	30.6	38.4	26	55.2	12	1001
Jun-06	111.4	31.9	38.1	26.9	64.6	10	1002
Jun-07	38.4	32.8	38.9	27.6	60	13	1000
Jun-08	47.4	30.4	37.1	26.6	66	12	1001
Jun-09	1.3	33	40.1	27.8	56.6	16	1003
Jun-10	62.1	33.2	39.8	28.3	49.8	19	1002
Jun-11	29.8	34.1	39.3	28.6	50.3	16	1000
Jun-12	33.93	33.7	40.4	27.6	55.5	12	1002
Jun-13	198.85	31.7	36.9	26.7	66.3	12	1000
Jun-14	0	33.6	40.2	29.2	54.3	17	1002

Figure 4: June Month Rainfall Data(2002-2014)

A	B	C	D	E	F	G	H	I
Monthyear	Rainfall (in MM)	Mean T	Max T	Min T	Mean Humidity	wind_speed(km/hr)	Sea_level_Pressure(hPa)	
Jul-02	11.2	30.2	34.7	26.1	60.8	12	1003	
Jul-03	473.1	28.4	33	25.5	75.2	11	1001	
Jul-04	99.9	29.8	35	26.2	74.5	13	1002	
Jul-05	510.5	28.9	33	26	74.5	12	1002	
Jul-06	352.6	28.7	32.2	25.4	82.1	12	1000	
Jul-07	661.2	29.2	32.7	26.1	77.7	13	1000	
Jul-08	140.6	28.4	33.5	25.5	78.2	8	1001	
Jul-09	284.7	29.6	34.2	25.9	76.9	14	999	
Jul-10	345.3	28.7	33.5	25.7	77.5	15	1001	
Jul-11	189	30.5	34.5	26.2	73.1	10	1001	
Jul-12	131.55	31.4	35.4	26.8	69.5	11	1001	
Jul-13	402.59	28.4	32.1	25.1	85.7	10	1000	
Jul-14	291	312	36	27	70	13	1001	

Figure 5: July Month Rainfall Data(2002-2014)

A	B	C	D	E	F	G	H	I
Monthyear	Rainfall (in MM)	Mean T	Max T	Min T	Mean Humidity	Wind_speed(km/hr)	Sea_level_Presssure(hPa)	
Aug-02	96.5	28.5	32.5	24.9	70.3	9	1004	
Aug-03	286.6	27.7	32	25	76	10	1004	
Aug-04	328.3	27.2	30.3	24.5	87.4	11	1003	
Aug-05	363.6	27.1	31.8	23.8	73.1	10	1003	
Aug-06	803	27.4	30.4	24.7	84.1	11	1002	
Aug-07	335.5	28.1	32.2	25.2	83.1	10	1002	
Aug-08	329.9	28.1	31.4	24.9	79.7	7	1003	
Aug-09	152.2	29.6	34.3	25.3	71.7	14	1004	
Aug-10	281.4	28.5	32.8	25.5	82	11	1003	
Aug-11	520	29.1	32.2	25.4	82.9	9	1001	
Aug-12	156.07	29.4	33.2	25.5	75.5	10	1003	
Aug-13	194.24	28.2	31.9	25.1	84	10	1004	
Aug-14	114	29	34	36	79	9	1004	

Figure 6: August Month Rainfall Data(2002-2014)

Monthyear	Rainfall (in MM)	Mean T	Max T	Min T	Mean Humidity	Wind_speed(km/hr)	Sea_level_Presssure(hPa)
Sep-02	61.3	28.8	34.7	24.1	65.5	8	1008
Sep-03	57.2	27.8	32.6	24.5	70.5	9	1006
Sep-04	37.4	29.6	35.1	25.2	69.9	9	1007
Sep-05	344.5	27.8	32.3	24.3	79.7	6	1004
Sep-06	150	29.7	34	25.4	68.8	7	1006
Sep-07	64.5	28.6	33.3	25.2	79.5	6	1004
Sep-08	73.6	29.3	33.8	25.1	75.7	6	1005
Sep-09	9.6	30.1	35.3	25.1	65.6	10	1006
Sep-10	97.5	28	33.5	24.6	73.7	14	1005
Sep-11	749	29.5	33	24.6	78.4	8	1004
Sep-12	53.08	29.1	33.8	24.6	75.3	6	1005
Sep-13	319.24	29.6	35	25.2	76	8	1005
Sep-14	744	28.6	32.9	25.3	82.9	6.6	1010

Figure 7: September Month Rainfall Data(2002-2014)

- Month wise Sample Data Set (For Moving Average)from 2002-2014 are shown Below.

A	B
Monthyear	Rainfall (in MM)
Jun-03	63.6
Jun-04	123.8
Jun-05	172.3
Jun-06	111.4
Jun-07	38.4
Jun-08	47.4
Jun-09	1.3
Jun-10	62.1
Jun-11	29.8
Jun-12	33.93
Jun-13	198.85

Figure 8: June Month Rainfall(2003-2013)

A	B
Monthyear	Rainfall (in MM)
Jul-03	473.1
Jul-04	99.9
Jul-05	510.5
Jul-06	352.6
Jul-07	661.2
Jul-08	140.6
Jul-09	284.7
Jul-10	345.3
Jul-11	189
Jul-12	131.55
Jul-13	402.59

Figure 9: July Month Rainfall(2003-2013)

A	B
Monthyear	Rainfall (in MM)
Aug-03	286.6
Aug-04	328.3
Aug-05	363.6
Aug-06	803
Aug-07	335.5
Aug-08	329.9
Aug-09	152.2
Aug-10	281.4
Aug-11	520
Aug-12	156.07
Aug-13	194.24

Figure 10: August Month Rainfall(2003-2013)

A	B
Monthyear	Rainfall (in MM)
Sep-03	57.2
Sep-04	37.4
Sep-05	344.5
Sep-06	150
Sep-07	64.5
Sep-08	73.6
Sep-09	9.6
Sep-10	97.5
Sep-11	749
Sep-12	53.08
Sep-13	319.24

Figure 11: September Month Rainfall(2003-2013)

8 Results and Interpretation

8.1 Multiple Regression

Regression Formula:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 \quad (1)$$

where,
Y=Rainfall Prediction
X_1=Mean Temperature(℃)
X_2=Max Temperature (℃)
X_3=Min. Temperature(℃)
X_4=Mean Humidity (٪)
X_5=sea Level pressure(hPa)
X_6=Wind speed (km/h)

Year wise Rainfall Prediction(1996-2013)

	Coefficients
Intercept	-3.796e+05
Mean Temp	1.946e+02
Max Temp	4.057e+01
Min Temp	6.002e+01
Mean Humidity	6.300e+01
No of Rainy days	9.262e+00
Wind speed	-2.620e+01
Mean Sea Level	3.660e+02
Multiple R-squared	0.8765

- Here in below shown (Year Wise Rainfall Prediction) figure, Red line represents Predicted Value of Rainfall where Blue line represents Actual Value of Rainfall.

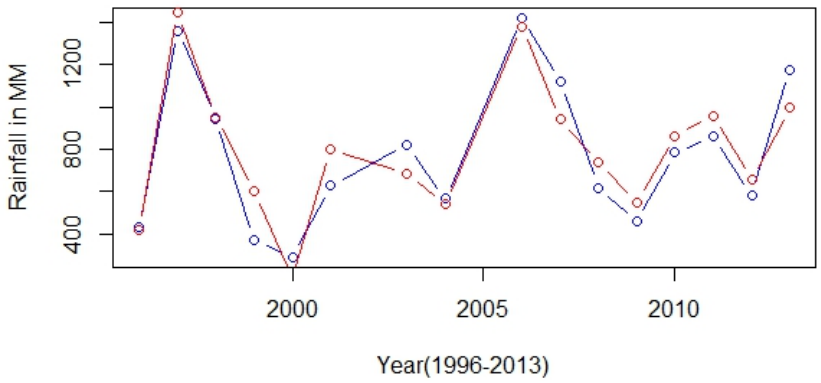

Figure 12: Yearwise Actual Rainfall Vs. Predicted Rainfall(1996-2013)

Month wise Rainfall Prediction(2002-2014)

Coefficients	June	July	August	September
Intercept	-9140.684	18123.1562	149404.322	-75369.28
Mean Temp	32.366	0.5796	9.917	376.12
Max Temp	-62.987	-230.2247	-144.527	-198.37
Min Temp	-54.493	313.3956	23.842	-281.40
Mean Humidity	-7.102	1.1571	-11.508	40.05
Wind speed	-2.196	18.2003	43.488	40.58
Mean Sea Level	12.534	-18.4862	-144.421	74.67
Multiple R-squared	0.7496	0.8539	0.8325	0.8554

- Here in all below shown (Month Wise Rainfall Prediction) figure, Red point represents Predicted Value of Rainfall where Black dash represents Actual Value of Rainfall.

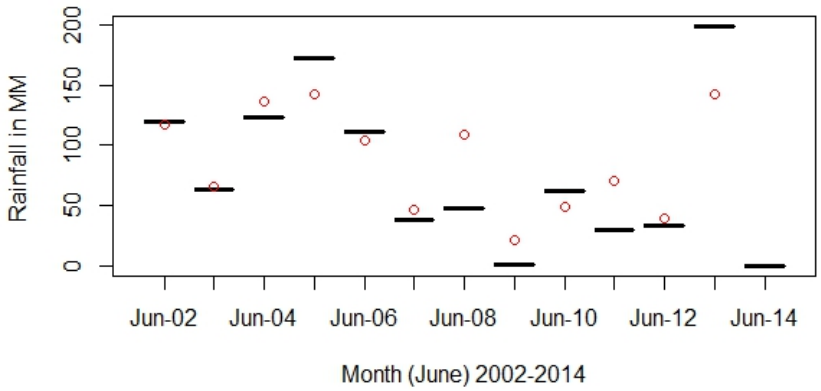

Figure 13: Monthwise Actual Rainfall Vs. Predicted Rainfall(June 2002-2014)

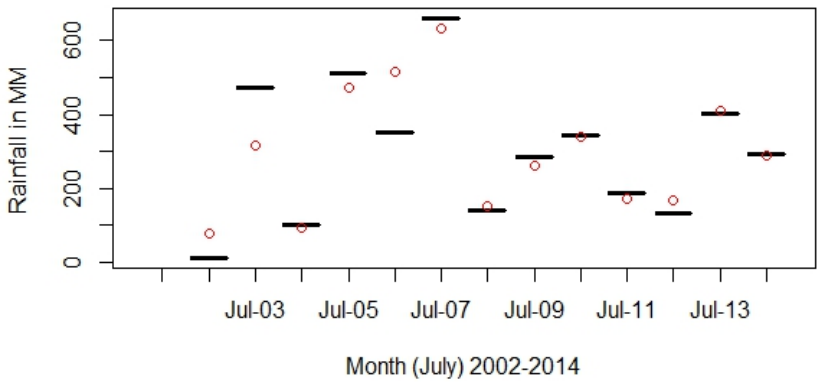

Figure 14: Monthwise Actual Rainfall Vs. Predicted Rainfall(July 2002-2014)

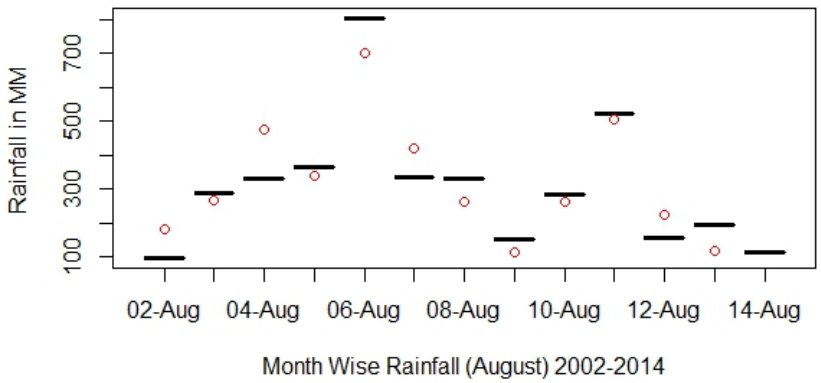

Figure 15: Monthwise Actual Rainfall Vs. Predicted Rainfall(August 2002-2014)

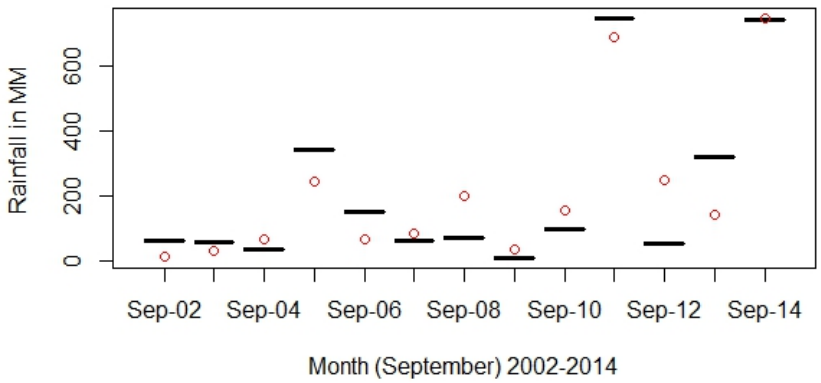

Figure 16: Monthwise Actual Rainfall Vs. Predicted Rainfall(September 2002-2014)

8.2 Moving Average

- Here in all below shown (Moving avg.) figure, Red line represents Predicted Value of Rainfall where Blue line represents Actual Value of Rainfall.

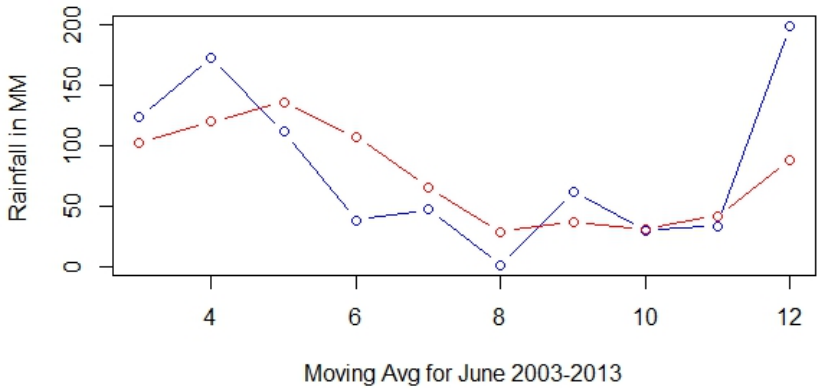

Figure 17: Moving average for June 2003-2013

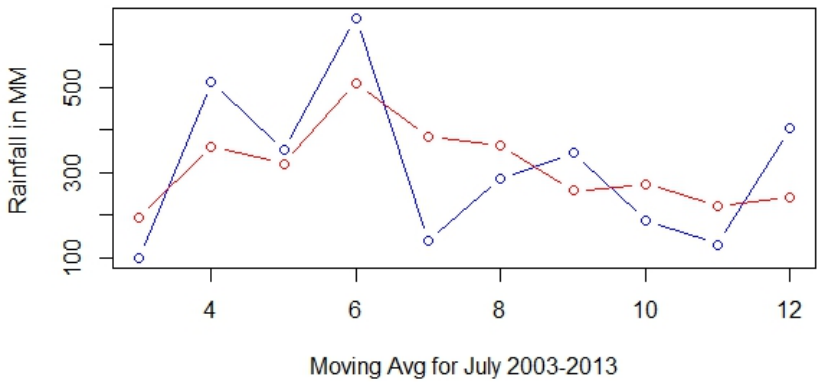

Figure 18: Moving average for July 2003-2013

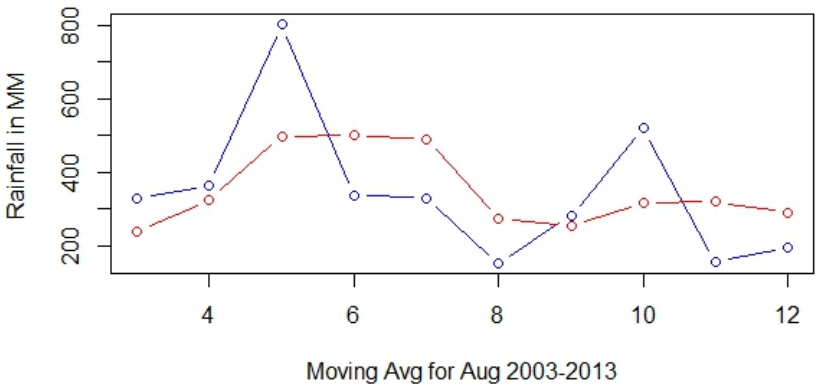

Figure 19: Moving average for August 2003-2013

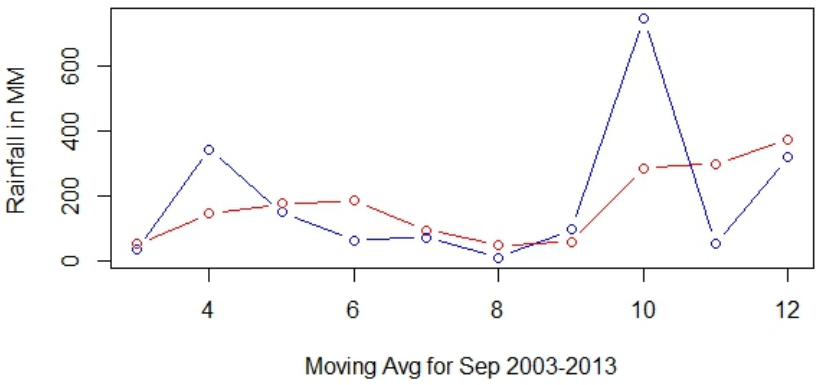

Figure 20: Moving average for September 2003-2013

8.3 Interpretation

Year	Actual Rainfall (in MM)	Excepted Rainfall (1)	Excepted Rainfall(2)
1996	432	513.689	420.233
1997	1361	1542.58	1450.666
1998	941	809.327	951.278
1999	371	560.318	600.141
2000	289	205.787	186.829
2001	626	657.092	797.21
2003	822	616.451	682.878
2004	565	855.26	543.274
2006	1421	1187.668	1381.487
2007	1119	848.089	940.061
2008	619	666.153	736.881
2009	459	606.454	550.181
2010	783	736.065	863.059
2011	863.03	964.519	958.925
2012	578.81	645.511	656.004
2013	1174.59	939.006	994.149

Here,Excepted Rainfall(1) values is estimated with using the MLR technique with the help of parameter Mean temp.,Max temp.,Min temp. ,Mean Humidity, wind Speed, and Sea Level Pressure and Multiple R-squared is 0.7524 for our model.

Excepted Rainfall(2) values is estimated with using the MLR technique with the help of parameter Mean temp.,Max temp.,Min temp. ,Mean Humidity,No of Rainy Days, wind Speed, and Sea Level Pressure and Multiple R-squared is 0.8765 for our model.

In Excepted Rainfall(2) Some predictor like No of Rainy Days is included due to constraints on data collection which could give more accurate result than the previous model.

For June Month (2002-2013)

Year	Actual Rainfall (in MM)	Excepted Rainfall (in MM)
2002	120.2	117.8568
2003	63.6	65.7162
2004	123.8	136.3407
2005	172.3	142.3484
2006	111.4	104.4438
2007	38.4	46.0517
2008	47.4	108.3609
2009	117.8568	21.2027
2010	62.1	48.4971
2011	29.8	70.7411
2012	33.93	39.9236
2013	198.85	142.9202

For the June Month (2002-2013) model consider mean temperature, maximum temperature, minimum temperature, wind speed, Mean sea level as Predictors.We found Multiple R-squared 0.7496 for our model.

For July Month (2002-2013)

Year	Actual Rainfall (in MM)	Excepted Rainfall (in MM)
2002	11.2	78.58487
2003	473.1	316.32056
2004	99.9	93.16395
2005	510.5	472.21229
2006	352.6	514.00513
2007	661.2	631.66856
2008	140.6	150.07861
2009	284.7	259.64505
2010	345.3	339.52374
2011	189	170.94738
2012	131.55	166.33889
2013	402.59	410.60000

For the July Month (2002-2013) model consider mean temperature, maximum temperature, minimum temperature, wind speed, Mean sea level as Predictors.We found Multiple R-squared 0.8539 for our model.

For August Month (2002-2013)

Year	Actual Rainfall (in MM)	Excepted Rainfall (in MM)
2002	96.5	167.1904
2003	286.6	211.7969
2004	328.3	497.3311
2005	363.6	383.9359
2006	803	672.0276
2007	335.5	398.7619
2008	329.9	271.4731
2009	152.2	128.8161
2010	281.4	234.8909
2011	520	516.6819
2012	156.07	217.3194
2013	194.24	141.5283

For the August Month (2002-2013) model consider mean temperature, maximum temperature, minimum temperature, wind speed, Mean sea level as Predictors.We found Multiple R-squared 0.8525 for our model.

For September Month (2002-2013)

Year	Actual Rainfall (in MM)	Excepted Rainfall (in MM)
2002	61.3	13.072
2003	57.2	32.459
2004	37.4	67.210
2005	344.5	245.630
2006	150	66.864
2007	64.5	86.886
2008	73.6	201.605
2009	9.6	37.431
2010	97.5	157.400
2011	749	690.850
2012	53.08	251.061
2013	319.24	141.432

For the September Month (2002-2013) model consider mean temperature, maximum temperature, minimum temperature, wind speed, Mean sea level as Predictors.We found Multiple R-squared 0.8554 for our model.

9 Limitation

- In our project, We have used only structured data which is easy to handle and we have predict the value of rainfall only for Mehsana district so we are sticky on particular Mehsana District.
- In our Project we are not going to consider some attributes like Pressure of station , longitude or latitude.Bocz we are not able to collect data of the same.So our accuracy might not be good.
- In MLR Technique , Predictors are Min temp. , Max temp. , Mean temp. , Mean Humidity(%) , No of Rainy days , Sea Level Pressure(hPa) , Wind Speed(km/hr) and we have only last 12 years Month-wise data and Last 17 years Year-wise data.
- In this project We can predict the Long Term(Year-wise and Month-wise) Rainfall with a good efficiency approx more than 60 % but we are not able to predict the real time rainfall(Real time Analysis) like what will be possibility of Rainfall within 2-3 days and amount of Rainfall (in mm).
- We have implemented the concept of Moving Average, in that we have not consider any attributes (that is used to predict the rainfall).So the concept of moving average is only used to identify trends or Long term averages actually which is not create a great impact.

10 Conclusion And Future Direction

The model consider mean temperature,maximum temperature, minimum temperature,mean humidity, wind speed, Mean sea level as Predictors.The model can predict month-wise and year-wise rainfall.

- The values of rainfall are calculated using the data collected over 17 Years.
- In this project we are dealing with structured data but in future we may try to work with unstructured data.
- In this project , we are sticky only on Mehsana district but in future we should try to Predict the Rainfall for whole State(Considering all District and Taluka) and should come up with the better visualization and interpretation so it can be easily understand by the end user.
- From technology point of view , We may apply different machine learning technique and try to come up with the best technique suitable for the forecasting.

References

[1] N. Sen, Forecast models for Indian South-West Monsoons Season Rainfall, in Current Science, vol. 84, No. 10, May 2003, pp.1290-1291.

[2] Gadgil, S et al (2002) On Forecasting the Indian summer Monsoon: the Intriguing Season of 2002.

[3] M. Kannan (2010) International Journal of Engineering and Technology, Vol.2 (6),2010, 397-401.

[4] Rajeevan, M (2001) Prediction of Indian summer monsoon: Status, problems and prospects, Current Science, 81, 1451-1457.

[5] http://www.ijcse.com/docs/INDJCSE14-05-02-081.pdf

[6] http://www.iucaa.ernet.in/ nspp/ieee.pdf