WEATHER PREDICTION :

PROBABILISTIC MODEL TO PREDICT RAIN LEVEL

Project report submitted in partial fulfillment of the requirement for the degree of Bachelor of Technology

in

Computer Science and Engineering/Information Technology

By

(Jatin Garg (121298) Kanika Kaushik (121301) Manik Tyagi (121302))

Under the supervision of

(Dr. Pardeep Kumar)

To

Department of Computer Science & Engineering and Information Technology

Jaypee University of Information Technology Waknaghat, Solan-173234,

Himachal Pradesh,India

CANDIDATE'S DECLARATION

I hereby declare that the work presented in this report entitled **"WEATHER PREDICTION:PROBABILISTIC MODEL TO PREDICT RAIN LEVEL"** in partial fulfillment of the requirements for the award of the degree of **Bachelor of Technology** in **Computer Science and Engineering/Information Technology** submitted in the department of Computer Science & Engineering and Information Technology**,** Jaypee University of Information Technology, Waknaghat is an authentic record of my own work carried out over a period from August 2015 to May 2016 under the supervision of **(Dr. Pardeep Kumar)** (Assistant Professor Sr. Grade, CSE).

The matter embodied in the report has not been submitted for the award of any other degree or diploma.

Jatin Garg (121298)

Kanika Kaushik (121301)

Manik Tyagi (121302)

This is to certify that the above statement made by the candidate is true to the best of my knowledge.

Dr. Pardeep Kumar,
Assistant Professor, CSE Dept.
JUIT , Solan H.P.
Dated:30-05-2016

ACKNOWLEDGEMENT

The success and final outcome of this project required a lot of guidance and assistance from many people and we are extremely fortunate to have got this all along the completion of our project work.

We take this opportunity to extend profound gratitude towards our guide Dr. Pardeep Kumar for giving us an opportunity and for his exemplary guidance, monitoring and constant encouragement throughout working of the project.

We would also like to thank our Head of the Department, Brig. (Retd.) S.P. Ghrera for his underlying faith in the department of Computer Science and allocation of the project as well as its resources. We are also obliged to staff members of the college (JUIT, Solan) for the valuable information provided by them in their respective fields.

Last but not the least we place a deep sense of gratitude to family members and friends who have been constant source of inspiration during the preparation of this project.

Jatin Garg (121298)

Kanika Kaushik (121301)

Manik Tyagi (121302)

TABLE OF CONTENT

LIST OF ABBREVIATIONS

SSE	Sum of squared
fig.	Figure no.
Int	Integer
MLR	Multiple Linear Regression
MA	Moving average
EMA	Exponential moving average
ROC	Rate of change
OSC	Oscillator
SLP	Sea level Pressure
SST	*Sea Surface Temperature*
IOD	*Indian Ocean Dipole*
MPR	Medium Power Radar
SOI	Southern oscillation index
SMR	*Surface Movement Radar*
URS	ser requirement specification
ADMIN	Keyword for administrator
Et al.	et alia (Latin Phrase means and others)
W.r.t	with respect to
RAM	Random Access Memory
JDK	Java Development Kit

MB	Mega Byte
PHP	Hypertext Preprocessor
ASP	Active Server Pages
CSS	Cascading Style Sheets
LAMP	Linux, Apache, MySQL, Perl/PHP/Python
MS-DOS	Microsoft Disk Operating System
MCA	Master in Computer Application

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Weather forecasts are made by obtaining data about the current state of the atmosphere at a given place and to project how the atmosphere will change using scientific understanding. Weather warnings are important forecasts because they are used to protect life and property.

Ancient weather forecasting methods usually relied on observed patterns of events, also termed pattern recognition.

But, here this system will predict weather based on parameters such as temperature, humidity and wind. This system is an application with effective graphical user interface User will login to the system using ID and password. User will enter current temperature; humidity and wind. System will take this parameter and predict weather from previous data in database.

Various regression analysis techniques have been implemented to compute the results on the previously collected data. The computational results are compared and the technique with the most accurate result is chosen. Therefore the predictions will prove reliable.

CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION

Weather forecasting is a necessary area of investigation in human life. It is one of the most important features of forecast because agricultural sectors and many other industries are dependent on the conditions to a large extent. The conditions are suppose to be forecasted not only for future planning in agricultural and industrial sectors but also in various other fields like defence, hill climbing, transportation and aerospace navigation etc. It is often used to notify about disasters which are caused by abrupt changes in climatic conditions.

Usualy, forecasting is done using the data gathered by remote sensing satellites. Weather parameters like max, min temperature, rainfall extent, cloud conditions, wind stream and their directions are projected using figures and information taken by the meteorological satellites to outlook trends. The variables signifying weather conditions like temperature, relative humidity, rainfall etc., vary incessantly with time, forming time series of each parameter and can be used to expand a forecasting model.

In statistical analysis, regression techniques are often used for estimating the future events or values. Trend extraction and curve fitting methodologies are also used to estimate the future behavior of the time series and to fit the future statistics according to the trend. Regression techniques include parametric methods such as linear and logistic regression and non-parametric methodologies such as projection pursuits, additive model, multivariate adaptive regression etc.

In this project, Multiple Linear Regression (MLR) is implemented to develop a model for weather forecasting. The model proposed is capable of forecasting the weather conditions for a particular place using the data collected. The data is processed to obtain some statistical indicators to extract the hidden data available in the time series.

The statistical indicators are features like moving average (MA), exponential moving average (EMA), rate of change (ROC), moments and the coefficients for skewness and kurtosis which are calculated over certain periods. On the basis of correlation, features are chosen as inputs for the models. Regression equations are obtained for the parameter to be forecasted, which is termed as a target. The data set is divided into two parts, the first part is used to obtain Multiple Linear Regression equations and the remaining part is used to test the model. MS Excel has been used to process the data and show the results in simple understandable form.

1.2 PROBLEM STATEMENT

Weather forecasting is the appliance of science and technology to predict the state of the atmosphere for a given station. Human beings have attempted to forecast the weather informally for millennium. Weather forecasts are made by collecting quantitative data about the existing state of the atmosphere at a given location and using scientific knowledge of atmospheric processes to project how the atmosphere will change.

1.3 OBJECTIVES

- The objective of this project is to predict rainfall using Multiple Linear Regression (MLR).
- Parameters like min temperature, max temperature and precipitation have been predicted using recorded values.

1.4 METHODOLOGY

In this project, dependent rainfall variable is expressed in terms of independent explanatory variables. Multiple linear regressions (MLR) is used to model a relationship between the dependent variable and the explanatory variables. It allows investigation on the effect of changes in the various factors on the dependent variable. If the observations are measured over time, the model becomes a time series regression model. The resulting statistical relationship can be used to predict various levels of rainfall. To determine the predictive power of the model, all assumptions of multiple linear regressions must be met.

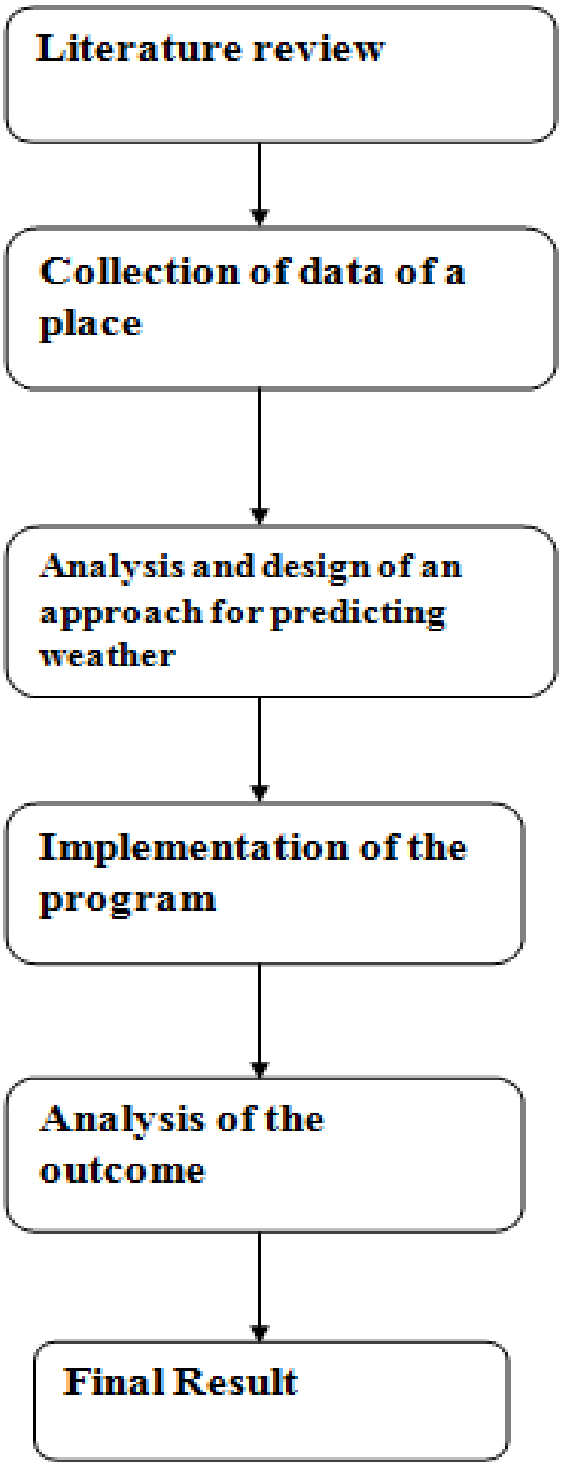

Literature review
Collection of data of a place
Analysis and design of an approach for predicting weather
Implementation of the program
Analysis of the outcome
Final Result

Fig.1.1

1.5 ORGANIZATION

Chapter 1: Highlights and Underlines of the prediction model. In this chapter, the introduction of the rainfall prediction model is covered. The key focus defining the problem statement and specifying the objectives of the project.

Chapter 2: The detailed literature review from the research paper, books, journals and conferences are done. In this chapter, the extracts from assorted research papers like the rainfall prediction model, weather prediction model are shown.

Chapter 3: Covers the system development which is the key aspect of this work. In this chapter, the proposed model, algorithm and related parameters are emphasized.

Chapter 4: The simulation of implementation results with the relative performance analysis is shown in this chapter. The simulation results and screenshots are revealed to depict and defend the proposed work.

Chapter 5: Detailed conclusion and scope of the future work which guides the upcoming students and research scholars to enhance the current work with higher efficiency and effectiveness on rainfall prediction model.

CHAPTER 2

LITERATURE SURVEY

A lot of information have been gathered and investigated from various books, research papers, journals and online links.

In this chapter, the details of research papers and journals are specified from where we have analyzed the contents and formulated the problem.

A number of research scholars and scientists have written research papers and found excellent results. This section underlines all those research papers and their extracts.

Pinky Saikia Dutta and Hitesh Tahbildar have proposed data mining technique for prediction of rainfall over Assam. In this paper,data mining technique in forecasting monthly Rainfall of Assam has been introduced. This was carried out with the help of traditional statistical technique which is -Multiple Linear Regression(MLR).The data includes Six years period[2007-2012] collected from Regional Meteorological Center, Guwahati, India . The performance of this model is measured in adjusted R-squared .The results show that the prediction model based on Multiple linear regression indicates acceptable accuracy.

In the project, rainfall forecast is implemented with the use of empirical statistical technique. They used 6 years datasets such as min temp, max temp,pressure,direction of wind, relative humidity performed prediction of Rainfall using Multiple Linear Regression.

The model forecasts monthly rainfall amount in monsoon season.The resulted rainfall amounts are intended to help the farmers to make decisions concerning their crop.Since rainfall is one of the causes of possible calamities, predicting the occurrence of rainfall will help them to prepare for future.

The procedures involved were identifying a model, secondly repeatedly changing the model by removing predictor variable and then terminating the process when they got a model which fitted the data well.

Accurate and timely weather forecasting is a major challenge for the scientific community. Rainfall forecasting model involves a collection of computer models, observations of patterns. By these methods, reasonably accurate forecasts can be made up.

N. Senhas proposed a long-range rainfall forecast model based on power regression approach with the help of Ei Nino, north west Europe temp, Europe pressure gradient, 50 hPa Wind pattern,east Asia pressure and south Indian ocean temp in years ago. The experimental results showed that the error was 3%.

S. Nkrintra, et al. described the development of a statistical forecasting model for SMR over Thailand with the help of multiple linear regression and local polynomial-based non-parametric approaches. SST, SLP, wind speed, EiNino Southern Oscillation Index, IOD were chosen as predictors. The experiments showed that the correlation among actual and forecasted rainfall was 0.6.

T. Sohn, et al. developed a model for the occurrence of heavy rain in South Korea with the help of multiple linear regression and artificial neural network. They had used 44 synoptic factors obtained by the numerical model.

Winn Thida Zaw had developed a forecasting model to determine rainfall over Myanmar using MLR where 15 predictors had been used.The predicted rainfall amount was close to the actual value.

The model considered max. temperature, min temperature, wind speed, mean sea level as predictors. They found 63% accuracy in variation of rainfall for their proposed model. The model can predict rainfall for a month. Some predictor as wind direction were not included due to constraints on data collection which could give better accuracy. The work can be extended for multiple locations in future. The resulted rainfall predictions are suppose to help farmers to make decisions regarding their crop.

Retius Chifurira and Delson Chikobvu at University of Free State developed a Weighted Multiple Regression Model to Predict Rainfall Patterns using Principal Component Analysis approach.

In this study, a multiple regression model has been developed to explain and predict mean annual rainfall in Zimbabwe. Principal component analysis is used to construct the orthogonal climatic factors which are influential on rainfall patterns in Zimbabwe. The aim of the study was to develop a simple and reliable tool to forecast the annual rainfall two years in advance using Darwin Sea Level Pressure(Darwin SLP) value of a particular month and a component for southern oscillation Index which is not explained by Darwin SLP. A weighted multiple regression approach which is used to control for the hetero scedasticity in the error terms. The model developed has a reasonable fit at 7% statistical significance level which can easily be used to predict mean annual rainfall at least a year in advance.

In this paper, dependent variable for rainfall is defined in terms of independent explanatory variables. MLR can be used to express a relationship between the dependent variable and the explanatory variables. It allows the investigation on the effect of changes in various factors on the dependent variable.The model becomes a time series regression model if the observations are measured over time.The resulting statistical relationship can be used to predict the values of rainfall. To ascertain the power of the model, all assumptions of multiple linear regression must be met.

Paras and Sanjay Mathur Assistant Professor, Associate Professor, Department of Electronics & Communication Engineering institute of technology, Govind Ballabh Pant college of Agriculture and Technology, Pantnagar proposed a “Simple Weather Prediction Model using Mathematical Regression”.

The model is capable of predicting the weather conditions for a particular location with the data collected locally. The information was processed to obtain few statistical indicators to extract the hidden data present in the time series. The statistical indicators are also called as features- moving and exponential moving average, oscillator, moments and coefficients of skewness and kurtosis were calculated over some period. On the basis

of the correlation, features were selected as inputs for the model. Regression equations were obtained for the parameter to be forecasted, which is termed as target. The whole data set was divided into two parts, the first was used to obtain MLR equations and the other part was used for testing the model.

Ms excel was used to process the information and present the results in simple logical form. The statistical forecasting models can be developed to estimate the weather conditions. For max temp and min temp estimation, the optimum size of the period for which the features are obtained, is 14 weeks and in case of relative humidity estimation it is 44 weeks. Relative humidity estimation using max temp, min temp and rain as input parameters is better than taking features extracted from its own time series. The rainfall can be estimated using the features extracted from other parameters like max temp, min temp and relative humidity. It is always possible to relate one weather parameter with the other weather parameters.

Nikhil Sethi, Dr.Kanwal Garg proposed a technique for "Exploiting Data Mining Technique for Rainfall Prediction". In this paper, rainfall forecast model is implemented using of empirical statistical technique, MPR. 30 years (1973-2002) information of the climate dataset such as rainfall precipitation vapor pressure, average temperature, and cloud cover over Udaipur city, Rajasthan are used. The model predicts monthly rainfall amount of July (in mm). The experimental outcomes prove that there is a close relation between the predicted and actual rainfall amount.

Neha Khandelwal, et al.presented a MLR eqn to forecast rainfall using 4 different climatic factors for Jaipur,Rajasthan, India, to select the factors and then use the result to find out the possibility for drought. Ozlem Terzi proposed a model to estimate rainfall in Esparto using data mining process. He used rainfall values of Uluborlu and E˘girdir stations for a month. The relative error of this model was 0.7%.

They applied multiple regression approach on the data set and find out predictable equation between rain and climatic parameters .So,MLR eqn is given-

Y=-1313.063+0.237*X10.75*X2+16.217*X3+13.039*X4

Where Y=Predicted rainfall

X1= Precipitation

X2= Average temp.

X3= cloud cover

X4= vapour press.

From the above equation we can compute the Rainfall for years ahead by knowing the precipitation, average temp.,cloud cover & vapour press. When MLR eqn is used combined with test data for testing the precision of the MLR eq. we get the rainfalll amount which is close to the real rainfall data.The graphical representation of the actual & the predicted values for rainfall is represented using graph. With the help of the results, the author plots a graph depicting the relationship between the real values of rainfall data and the forecasted values for rainfall with the help of MLR eqn and the graph shows that MLR method for forecasting the rainfall achieve nearby values between actual & predicted values.

Goutami Bandyopadhyay,Kolkata, West Bengal (2006) have proposed The Prediction of Indian Monsoon Rainfall".

Multiple linear regression approach is adopted to predict the average summer monsoon rainfall for a year with the help of the rainfall statistics for the summer-monsoon of the years in the past.

After computation, the MLR equation is set as y=0.03x1+0.06x2+0.02x3+229

Overall prediction error was found to be 25.36%.The t-statistics were found out from the related components and tabular values are compared. The computed values were based upon the theory that the entered value is a good predictor of the prediction. It was found that all the computed values are below the tabular values of t.The criticism of the neural

net method is that it did not analyze the autocorrelation of the rainfall time series, may be unfounded. The monsoon-rainfall data series is highly complex; the role that MLR will have here is one for future research—it appears, from the evidence here, not to be useful as a predictive approach.

Rahamathulla Vempalli MCA Department, KSRM College of Engineering, Kadapa, India, Dr S. Ramakrishna MCA Department, SV University, Tirupati, India and LatheefaVempalli Christ University, bangalore, India have proposed "Daywise Rainfall Prediction by Artificial Neural Network".

It is a distributed processor made up of simple processing modules, which has a natural propensity to store experimental data and making it available for usage whenever required. It is an information processing paradigm which is based on biological nervous system, such as the brain, process information. It is composed of a huge no. of interconnected processing elements (neurons) working in unison to solve specific problems. Neural network is designed to function the way in which the brain performs a particular task.

The network was implemented by using electronics components. An input is presented to the neural network and a desired or target response set at the output an error is calculated from the difference between the desired and the system output. This error information is given back to the system and the system adjusts the parameters in a logical manner. The process is repeated until the performance is good. It is clear from this description that the performance hinges heavily on the data. If one does not have data that cover the important part of the operating conditions, or if they are noisy, then neuron network technology is probably the right solution

CHAPTER 3

SYSTEM DEVELOPMENT

3.1 REQUIREMENT ANALYSIS

3.1.1 USER REQUIREMENT SPECIFICATIONS

The User Requirements Specification describes the needs for what users require from the system. User Requirements Specifications are written early in the validation process, typically before the system is created. They are written by the system owner and end-users, with input from Quality Assurance.

3.1.2 FUNCTIONAL REQUIREMENTS

The major functionality of this product is divided into two categories.

1. Administrative User Functions.
2. User Functions.

In this application admin must have ID and Password,using this User ID and Password only they can directly enter into the corresponding Login page.

Administrative User Functions-

In this functionality the administrator will do his own responsibilities. Whenever he needs to change the Password, then he can directly change it.

The admin can check the messages received from the users and can send feedback. For this , the administrator will provide his email id and phone no.He can add or remove the data of any day. The admin user can view the weather report.

User Functions-

In this functionality the user will do his own responsibilities. Whenever he needs to check the forecast, he can.

The user can also send a message to the administrator to the email id provided.

3.1.3 NON-FUNCTIONAL REQUIREMENTS

These requirements are the properties that the product must necessarily have. These properties are the qualities that make the product attractive, reusable, fast, reliable, secured.

These properties are not required only for the fundamental functioning of the product -- actions such as computations, manipulating data but also if the client wants them to perform in a certain manner.

3.1.4 HARDWARE REQUIREMENTS

The minimum requirements needed to perform operations are

- Intel Pentium Processor at 2 GHz or Higher
- RAM 256MB or more
- Hard disk capacity 10GB or more

3.1.5 SOFTWARE REQUIREMENTS

The software required to perform the implementation are

Windows: Windows(win) is an operating environment developed by Microsoft that provides an interface, well-known as a GUI (Graphical User Interface), for computers. The need to memorize commands for the command line (MS-DOS) is eliminated with the help of a mouse to navigate through menus, dialog boxes, tabs, and icons.

Wamp: WampServer is an open source project.It installs automatically.It allows the creation of web applications with PHP and the MySQL database. It comes with PHPMyAdmin to manage your databases easily.Its feature of allowing to tune the server without even touching the setting files makes its usage intuitive.It is the only packaged solution that allows to reproduce your production server.There is a possibility to add as many Apache, MySQL and PHP releases once it is installed.It has a tray icon to manage the server settings.

Apache: Apache is a Web server that is distributed under an "open source" license. Version 2.0 runs on most UNIX-based operating systems on other UNIX/POSIX-derived systems (such as Rhapsody, and on Windows 2000.

MySQL: MySQL is a database which is used in web applications. It is a central component of the widely used LAMP(Linux, Apache, MySQL, Perl/PHP/Python) open-source web application software stack.

PHP: PHP is a dynamic language that supports a variety of programming techniques. It has evolved over the years, notably adding a solid object-oriented model in PHP 5.0, anonymous functions and traits in PHP 5.4.It has a set of object-oriented features including support for classes, interfaces, inheritance, constructors,, exceptions, and more.

Notepad++: Notepad++ is a text and source code editor. It supports tabbed editing which allows working with multiple open files in one window. The following languages are supported by Notepad++:

- Ada, asp, Assembly
- Batch
- C, C++, C#, Caml, COBOL,CSS

3.2 SYSTEM DESIGN

SYSTEM STRUCTURE:

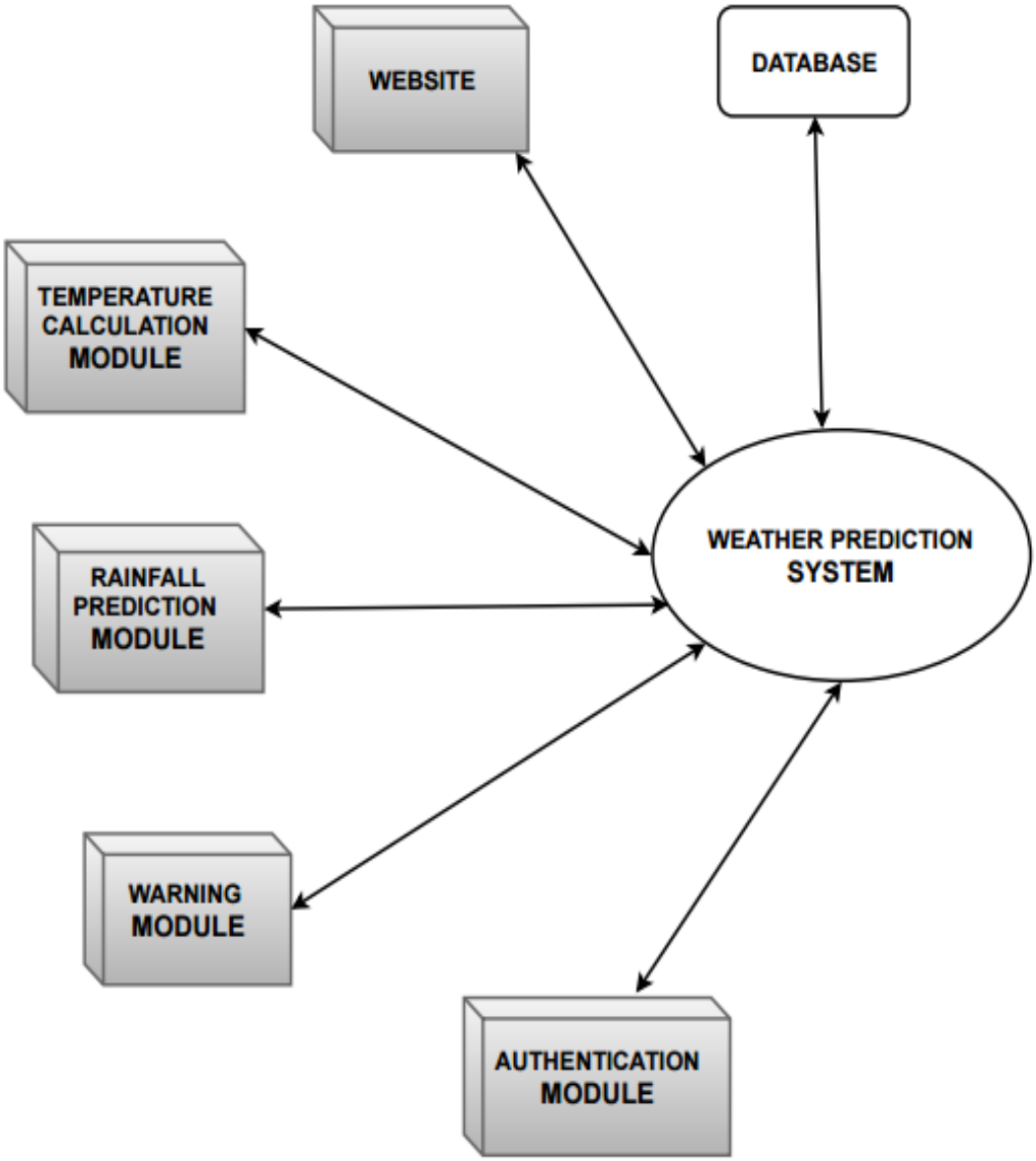


Fig.3.1

USE CASE:

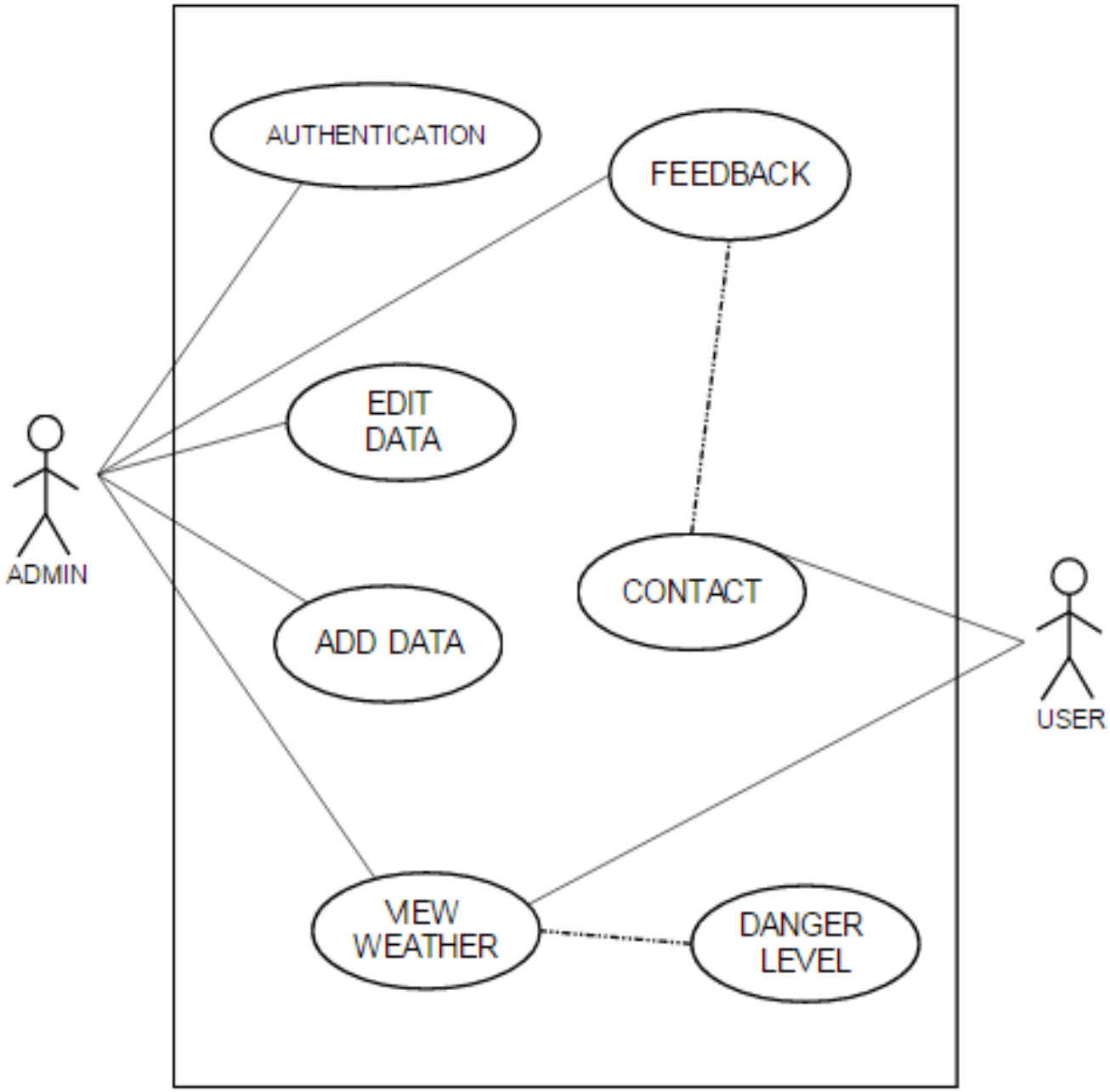


Fig.3.2

ACTIVITY DIAGRAM:

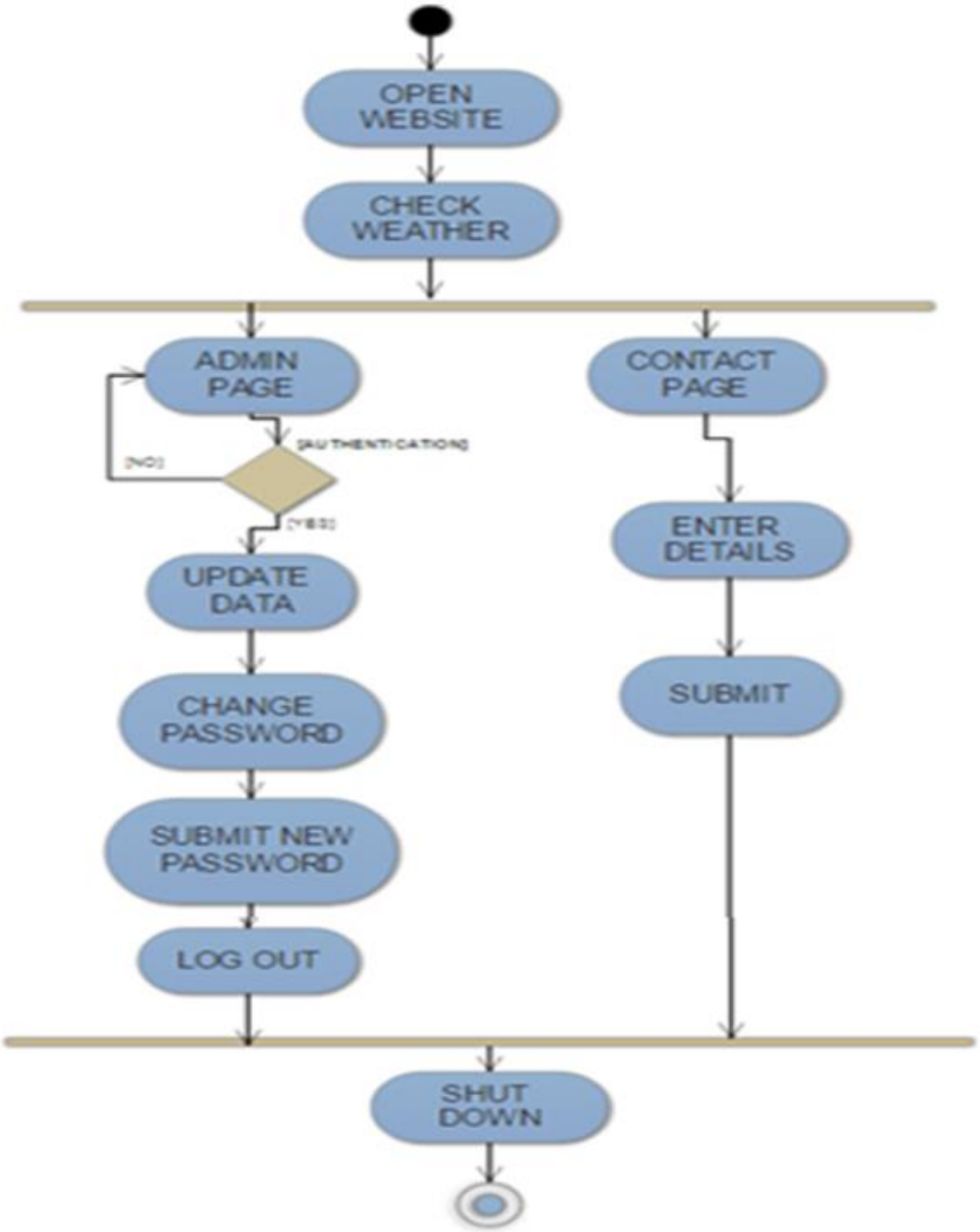


Fig.3.3

GANTT-CHART:

	Task	Assigned To	Start	End	Dur
	WEATHER PREDICTION: PROBABILISTIC MODEL TO PREDICT RAIN LEVEL		9/1/15	5/20/16	181
1	**Planning**	kanika,manik,jatin	9/1/15	10/19/15	33
1.1	Determine aim	kanika	9/1/15	9/1/15	1
1.2	Determine requirement specifications	manik	9/4/15	9/4/15	1
1.3	Outline events	jatin	10/1/15	10/1/15	
1.4	Propose various features	kanika	10/2/15	10/16/15	10
1.5	Design models,charts	manik	10/17/15	10/19/15	1
2	**Theoretical Survey**	kanika,manik,jatin	11/3/15	12/15/15	29
2.1	Gather resources(books,journals)	jatin	11/3/15	11/3/15	1
2.2	Study or review material	kanika	11/10/15	11/10/15	1
2.3	Determine the best methods	manik	11/23/15	11/25/15	3
2.4	Brainstorm concepts	jatin	11/26/15	12/15/15	13
3	**Implementation**	kanika	1/4/16	4/17/16	73
3.1	Work on algorithms	manik	1/4/16	1/30/16	19
3.2	Coding	jatin	2/2/16	3/4/16	23
3.3	Debugging	kanika	3/5/16	3/28/16	16
3.4	Website development	manik	3/29/16	4/17/16	14
4	**Assessment**	kanika,manik,jatin	4/17/16	4/25/16	6
4.1	Testing	jatin	4/17/16	4/19/16	2
4.2	Examine the data	kanika	4/19/16	4/22/16	4
4.3	Determine the best techniques	manik	4/23/16	4/25/16	1
5	**Report**	kanika,manik,jatin	4/28/16	5/20/16	17

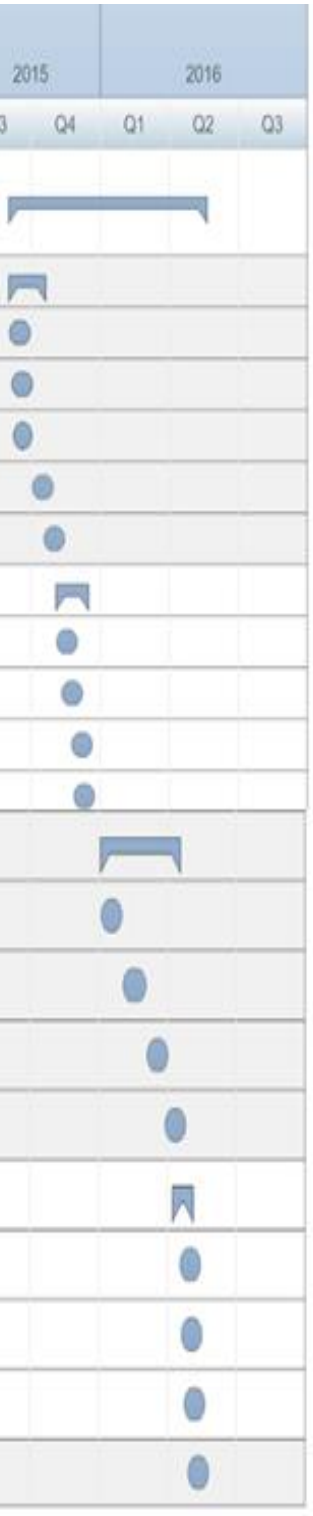


Fig.3.4

TIMELINE VIEW:

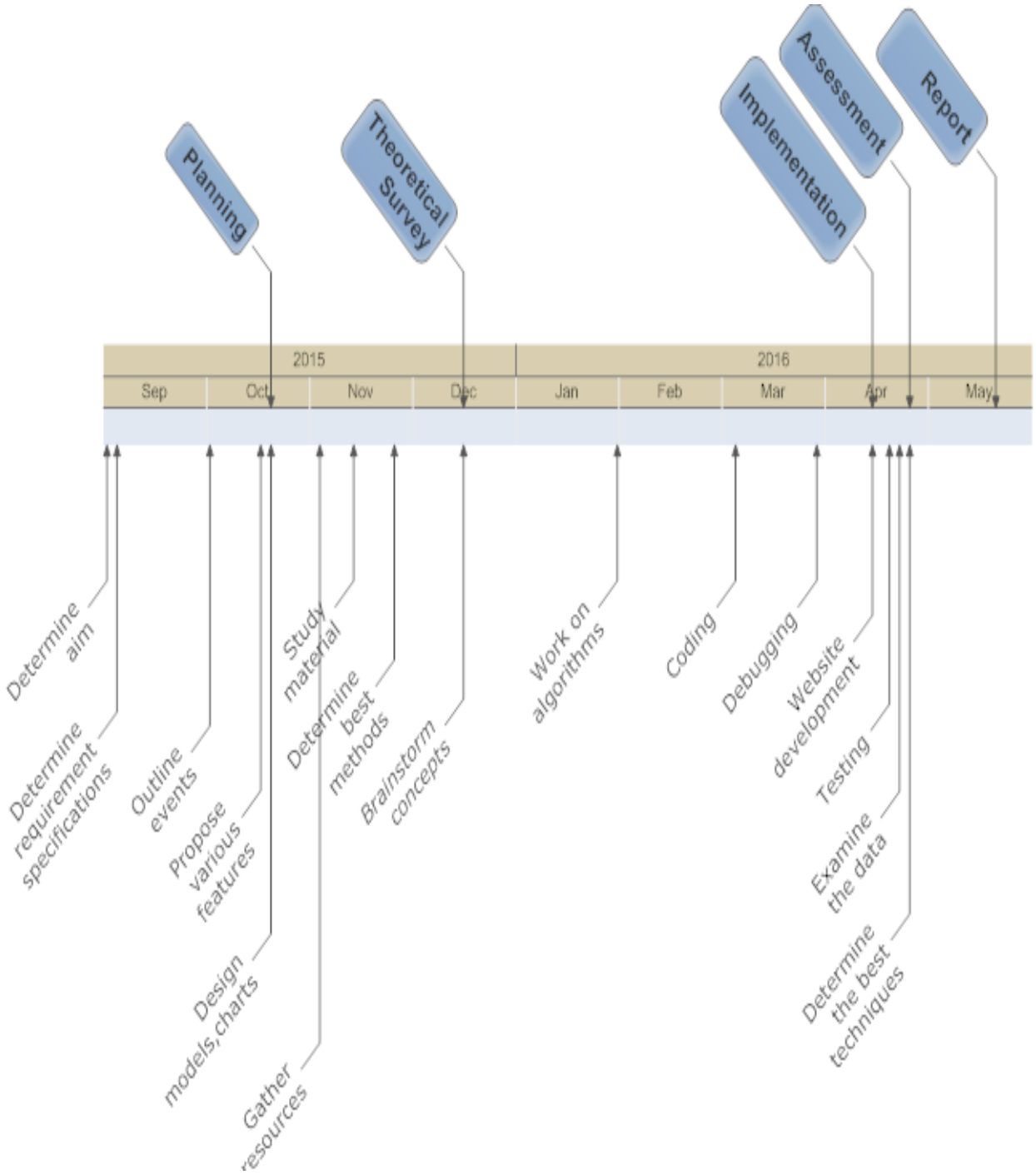


Fig.3.5

3.3 COMPUTATIONAL APPROACHES

Mathematical approaches have been used to do certain computations. This section describes the methodologies that have been implemented.

Formulas for Calculation:-

1.Linear Regression Equations:

Rainfall estimation using features of Tmax, Tmin, Precipitation :

$$Y = 1.129623 - 0.32981*X1 + 0.198254*X2 - 0.03391*X3 - 0.044988*X4 + 0.398313*X5 - 0.18627*X6 - 1.61159*X7 + 0.469794*X8 - 0.21661*X9$$

Where,

- X1, X4 and X7 are MA of Tmax, Tmin and PR,respectively
- X2, X5 and X8 are EMA of Tmax, Tmin and PR, respectively
- X3, X6 and X9 are OSC of Tmax, Tmin.and PR, respectively

2.Logistic Regression:

Logistic regression is a nonlinear regression method that is well suited to probability forecasting, i.e. Situation Where they predict and is a probability rather than a measurable physical quantity.

Denoting as

p:-the probability being forecast, a logistic regression takes the form:

$$p = \exp[f(x)] / (1+\exp [f(x)])$$

where,

f(x)=1.129623-0.32981*X1+0.198254*X2-0.03391*X3-0.044988*X4+0.398313*X5-0.18627*X6 - 1.61159*X7 + 0.469794*X8 - 0.21661*X9

- X1, X4 and X7 are MA of Tmax, Tmin and PR,respectively
- X2, X5 and X8 are EMA of Tmax, Tmin and PR, respectively
- X3, X6 and X9 are OSC of Tmax, Tmin.and PR, respectively.

Note: f(x) is the same equation of multiple linear regression.

3.Moving Average (MA):

It is calculated progressively as an average of N number data values over the certain period.

Data set is represented by dt, dt-1,dt-2 ,……., d0,

Where,

dt is present

d0 is the first data value, the moving average with a sliding window of period N is

MA = d +d +d +. . . . . .+d t t-1 t-2 t-N/N

4.Exponential Moving Average (EMA):

It is defined as

$$EMA = 1- \alpha dt + \alpha \times EMA\ t\text{-}1$$

Where, ‘a’ is called the smoothing constant having value $0<a<1$.

5.Oscillator (OSC):

Oscillator is used to indicate the rising or trailing trend present in the time series. It is defined as difference of moving averages or exponential moving averages of two different periods.

$$OSC = MAN1 - MAN2$$

or, $OSC = EMAN1 - EMAN2$

where, N1 and N2 are different periods and $N1>N2$.

3.4 DATABASE TABLES

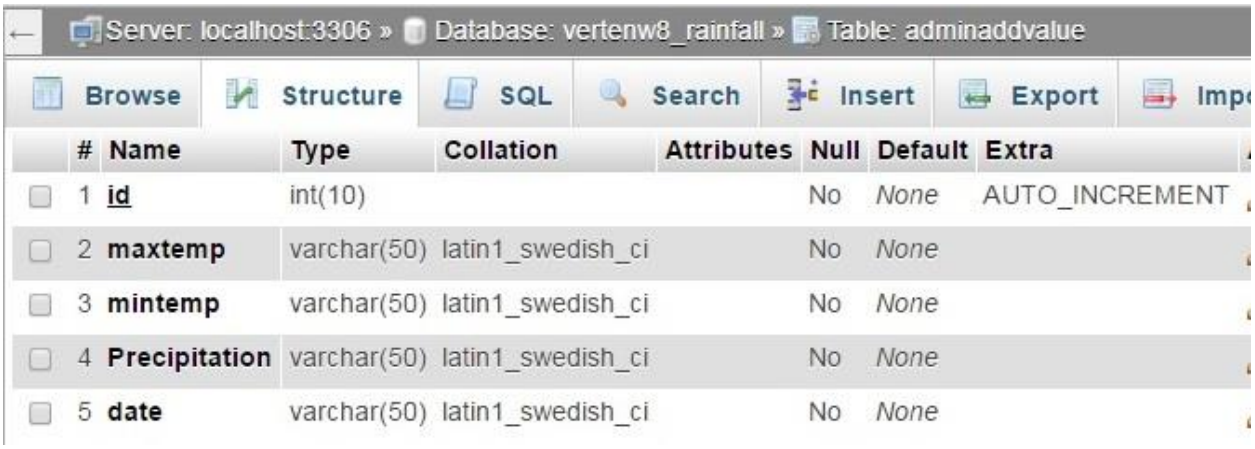
Table 3.5.1 Admin


Server: localhost:3306 » Database: vertenw8_rainfall » Table: admin

Browse | Structure | SQL | Search | Insert | Export

#	Name	Type	Collation	Attributes	Null	Default	Extra
1	id	int(10)			No	None	AUTO_INCREMENT
2	username	varchar(50)	latin1_swedish_ci		No	None	
3	password	varchar(50)	latin1_swedish_ci		No	None	

Table 3.5. 2 adminaddvalue


Server: localhost:3306 » Database: vertenw8_rainfall » Table: adminaddvalue

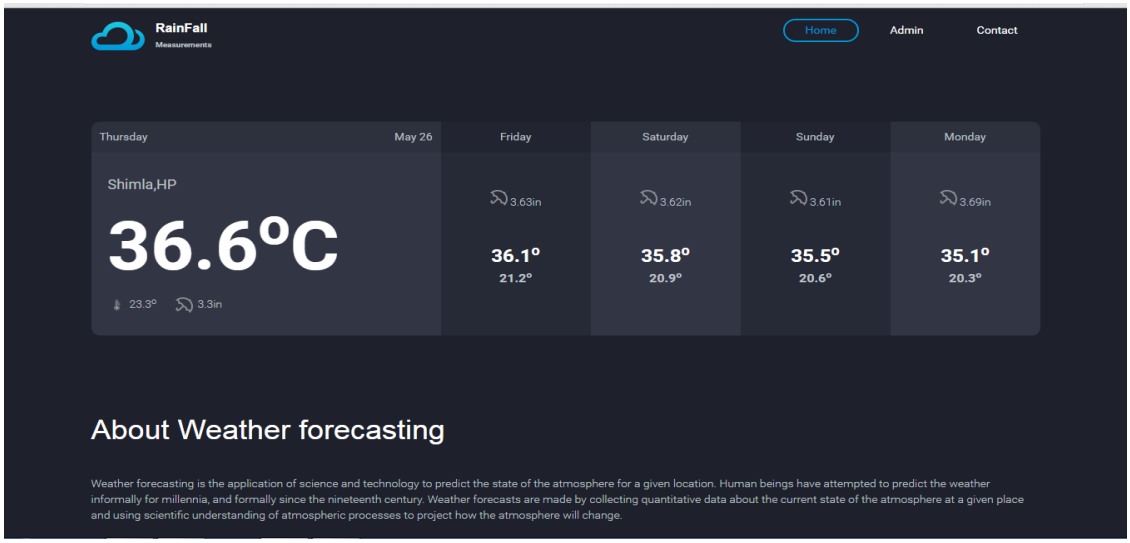
Browse | Structure | SQL | Search | Insert | Export

#	Name	Type	Collation	Attributes	Null	Default	Extra
1	id	int(10)			No	None	AUTO_INCREMENT
2	maxtemp	varchar(50)	latin1_swedish_ci		No	None	
3	mintemp	varchar(50)	latin1_swedish_ci		No	None	
4	Precipitation	varchar(50)	latin1_swedish_ci		No	None	
5	date	varchar(50)	latin1_swedish_ci		No	None	

CHAPTER 4

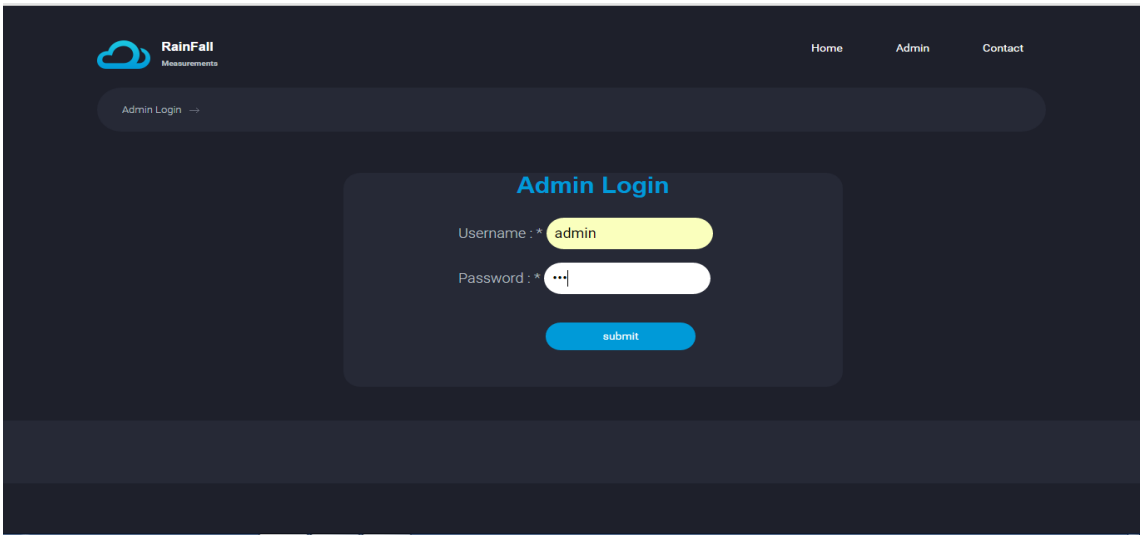
PERFORMANCE ANALYSIS

Screenshot 4.1 Home Page



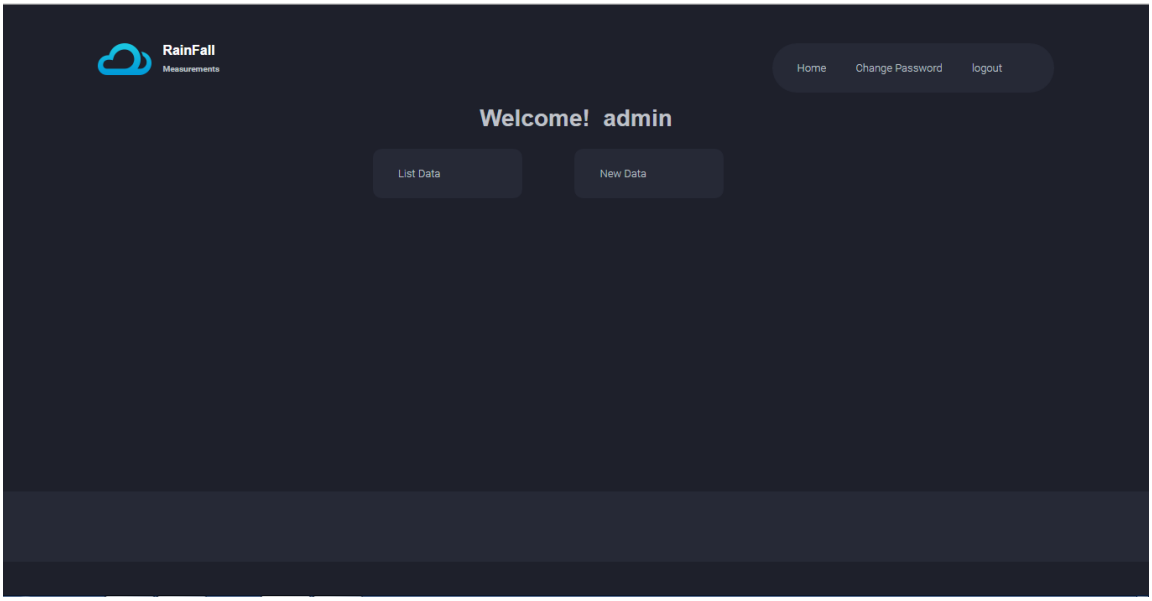
User can see the weather report for coming five days on home page. The page also gives an introduction about weather forecasting.

Screenshot 4.2 AdminLogin page



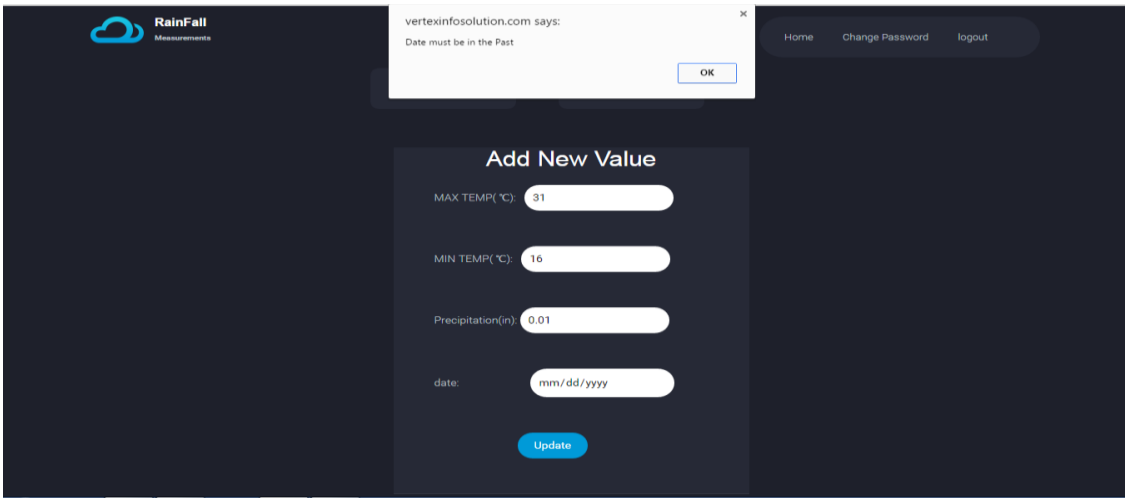
Admin will Login through this page. Authentication is required here.If admin is authenticated, Admin will get Welcome page.

Screenshot 4.3 Welcome Admin Page



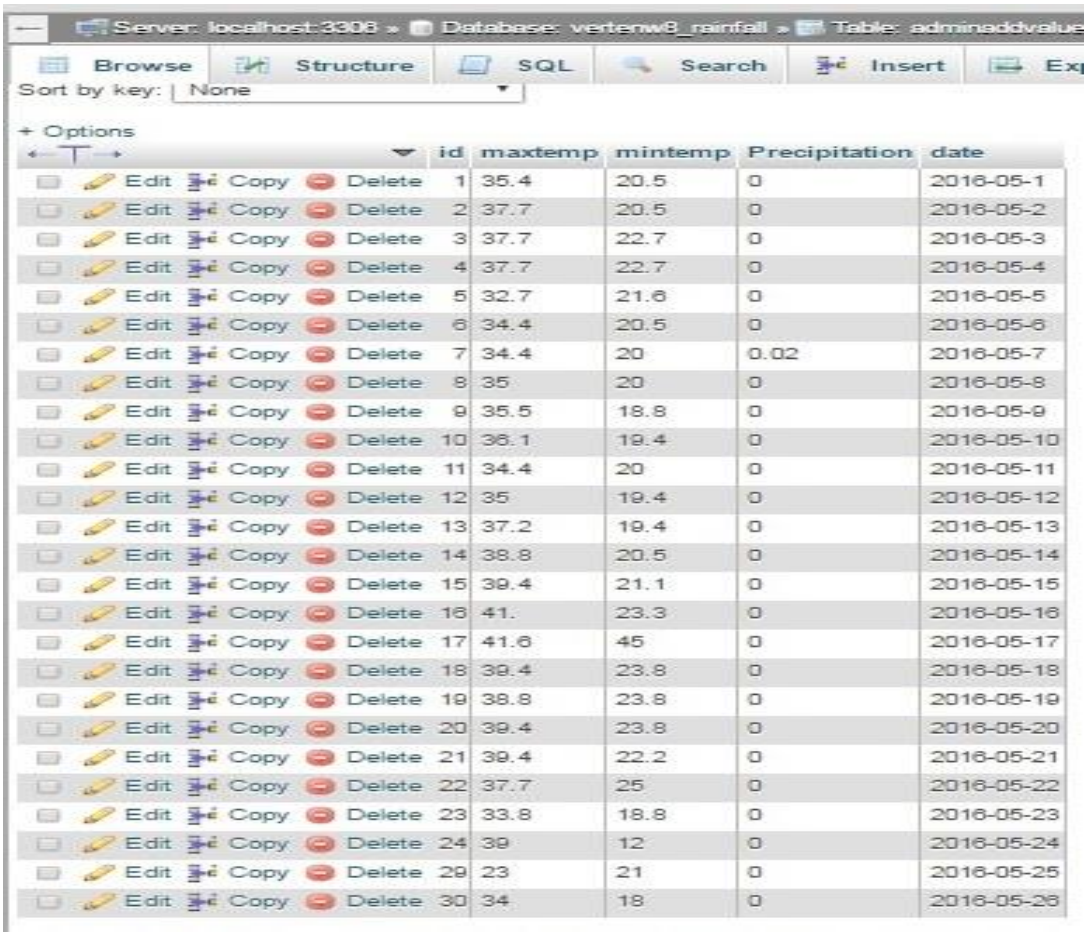
At this page admin will add new value or can list the previous data.This page also provides the log out option and the change password option.By clicking on list data,the admin can view the list of previously added data.

Screenshot 4.4 Addvalue Page



At this page, Admin will enter the value of Max Temp, Min Temp and Precipitation. After clicking submitt button the values will be saved in Adminaddvalue table.

Screenshot 4.5 Adminaddvalue Table


Server: localhost:3308 » Database: vertenw8_rainfall » Table: adminaddvalue

Browse | Structure | SQL | Search | Insert | Ex

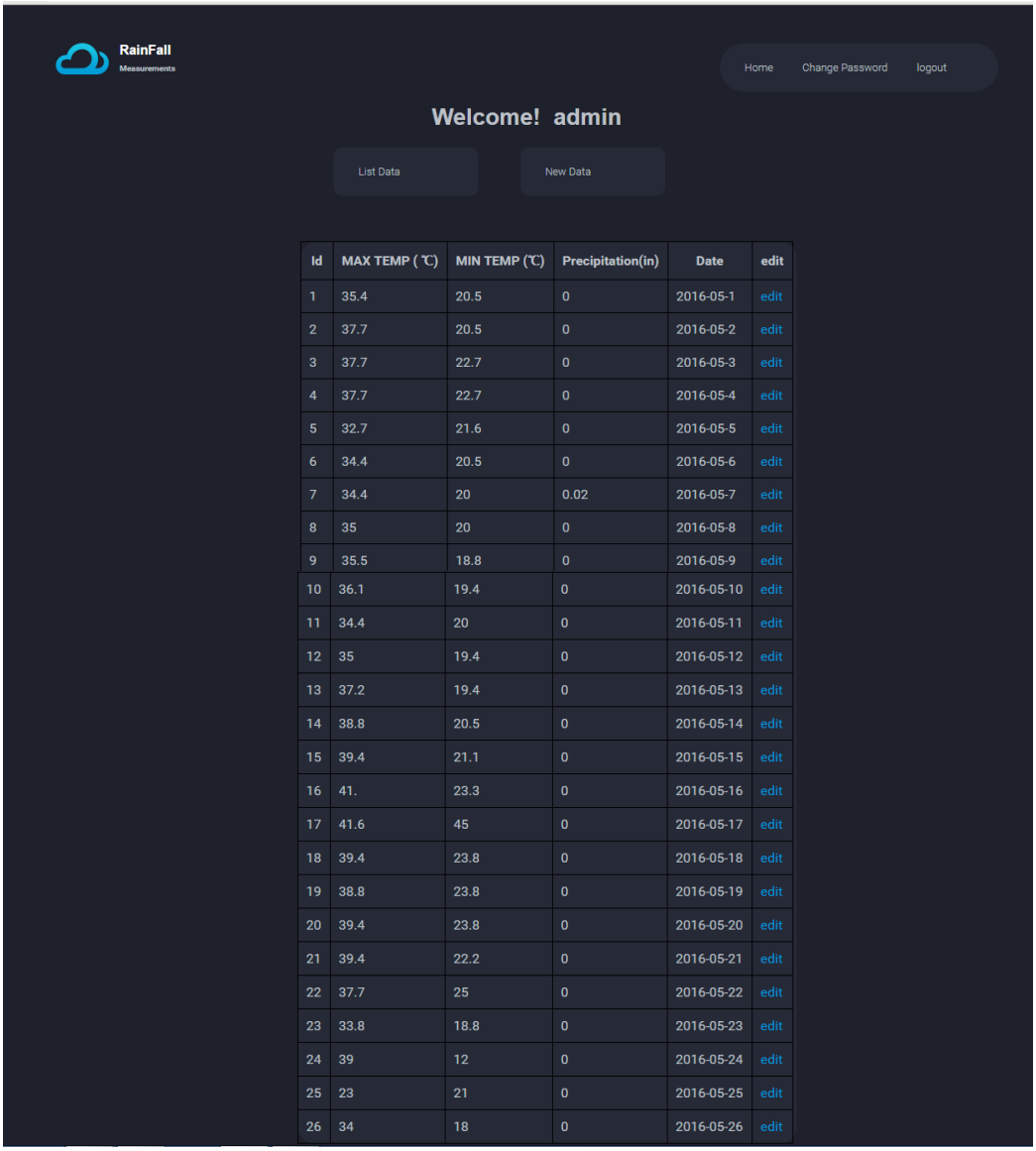
Sort by key: None

+ Options

	id	maxtemp	mintemp	Precipitation	date
Edit Copy Delete	1	35.4	20.5	0	2016-05-1
Edit Copy Delete	2	37.7	20.5	0	2016-05-2
Edit Copy Delete	3	37.7	22.7	0	2016-05-3
Edit Copy Delete	4	37.7	22.7	0	2016-05-4
Edit Copy Delete	5	32.7	21.6	0	2016-05-5
Edit Copy Delete	6	34.4	20.5	0	2016-05-6
Edit Copy Delete	7	34.4	20	0.02	2016-05-7
Edit Copy Delete	8	35	20	0	2016-05-8
Edit Copy Delete	9	35.5	18.8	0	2016-05-9
Edit Copy Delete	10	36.1	19.4	0	2016-05-10
Edit Copy Delete	11	34.4	20	0	2016-05-11
Edit Copy Delete	12	35	19.4	0	2016-05-12
Edit Copy Delete	13	37.2	19.4	0	2016-05-13
Edit Copy Delete	14	38.8	20.5	0	2016-05-14
Edit Copy Delete	15	39.4	21.1	0	2016-05-15
Edit Copy Delete	16	41.	23.3	0	2016-05-16
Edit Copy Delete	17	41.6	45	0	2016-05-17
Edit Copy Delete	18	39.4	23.8	0	2016-05-18
Edit Copy Delete	19	38.8	23.8	0	2016-05-19
Edit Copy Delete	20	39.4	23.8	0	2016-05-20
Edit Copy Delete	21	39.4	22.2	0	2016-05-21
Edit Copy Delete	22	37.7	25	0	2016-05-22
Edit Copy Delete	23	33.8	18.8	0	2016-05-23
Edit Copy Delete	24	39	12	0	2016-05-24
Edit Copy Delete	29	23	21	0	2016-05-25
Edit Copy Delete	30	34	18	0	2016-05-26

This table shows the data entries in the database.Features like id,max temp,min temp, precipitation and date have been shown.The options provided are edit,copy and delete.

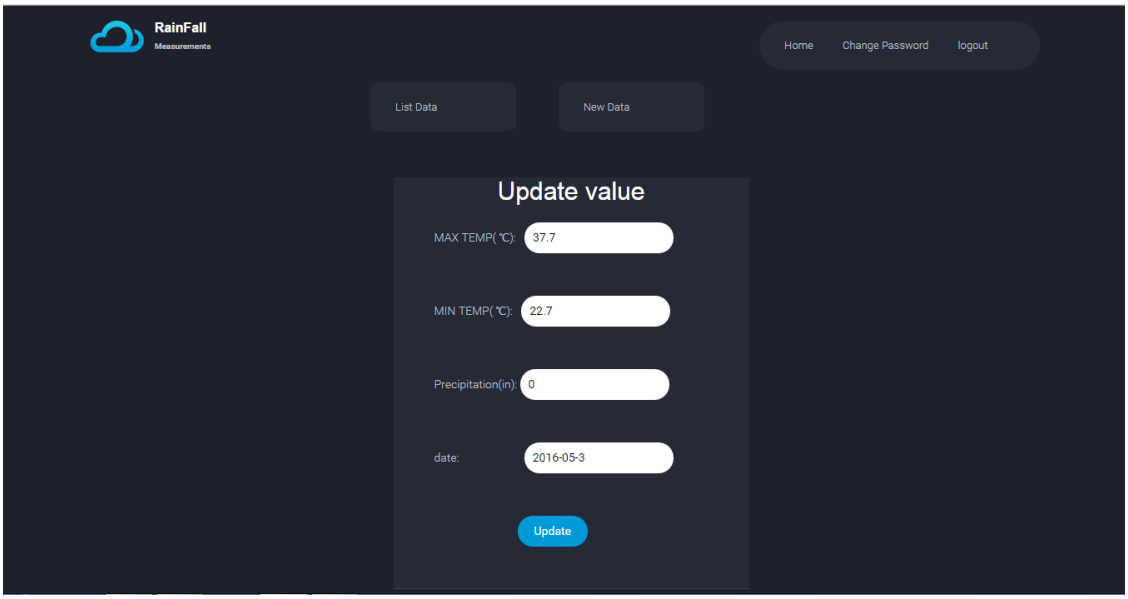
Screenshot 4.5 ListData Page



Id	MAX TEMP (℃)	MIN TEMP (℃)	Precipitation(in)	Date	edit
1	35.4	20.5	0	2016-05-1	edit
2	37.7	20.5	0	2016-05-2	edit
3	37.7	22.7	0	2016-05-3	edit
4	37.7	22.7	0	2016-05-4	edit
5	32.7	21.6	0	2016-05-5	edit
6	34.4	20.5	0	2016-05-6	edit
7	34.4	20	0.02	2016-05-7	edit
8	35	20	0	2016-05-8	edit
9	35.5	18.8	0	2016-05-9	edit
10	36.1	19.4	0	2016-05-10	edit
11	34.4	20	0	2016-05-11	edit
12	35	19.4	0	2016-05-12	edit
13	37.2	19.4	0	2016-05-13	edit
14	38.8	20.5	0	2016-05-14	edit
15	39.4	21.1	0	2016-05-15	edit
16	41.	23.3	0	2016-05-16	edit
17	41.6	45	0	2016-05-17	edit
18	39.4	23.8	0	2016-05-18	edit
19	38.8	23.8	0	2016-05-19	edit
20	39.4	23.8	0	2016-05-20	edit
21	39.4	22.2	0	2016-05-21	edit
22	37.7	25	0	2016-05-22	edit
23	33.8	18.8	0	2016-05-23	edit
24	39	12	0	2016-05-24	edit
25	23	21	0	2016-05-25	edit
26	34	18	0	2016-05-26	edit

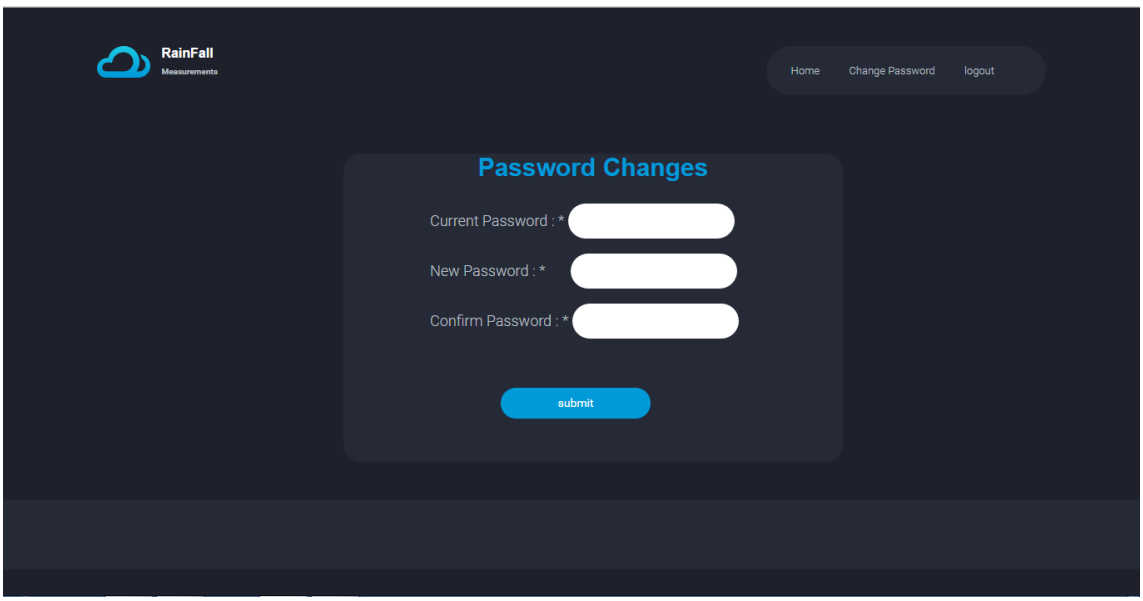
Admin can see previous days data by clicking ListData Tab.The admin can also edit any of the values if necessary.

Screenshot 4.6 UpdateValue Page



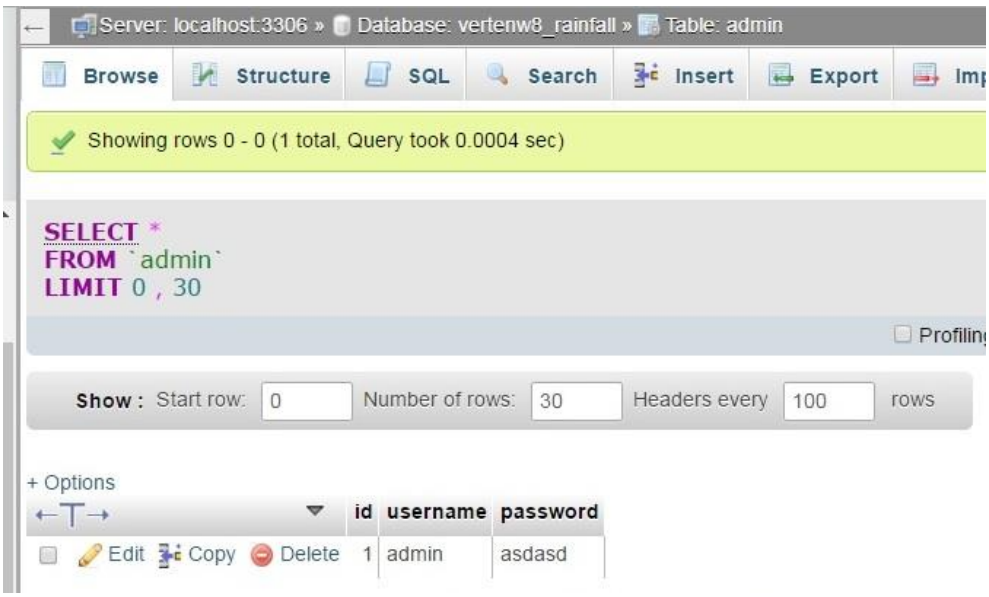
Admin can update value if he/she entered wrong value.Update value option takes in values for 4 parameters:maximum temperature,minimum temperature,precipitation and date.In case any field is left unfilled,an error is shown.

Screenshot 4.7 ChangePassword Page

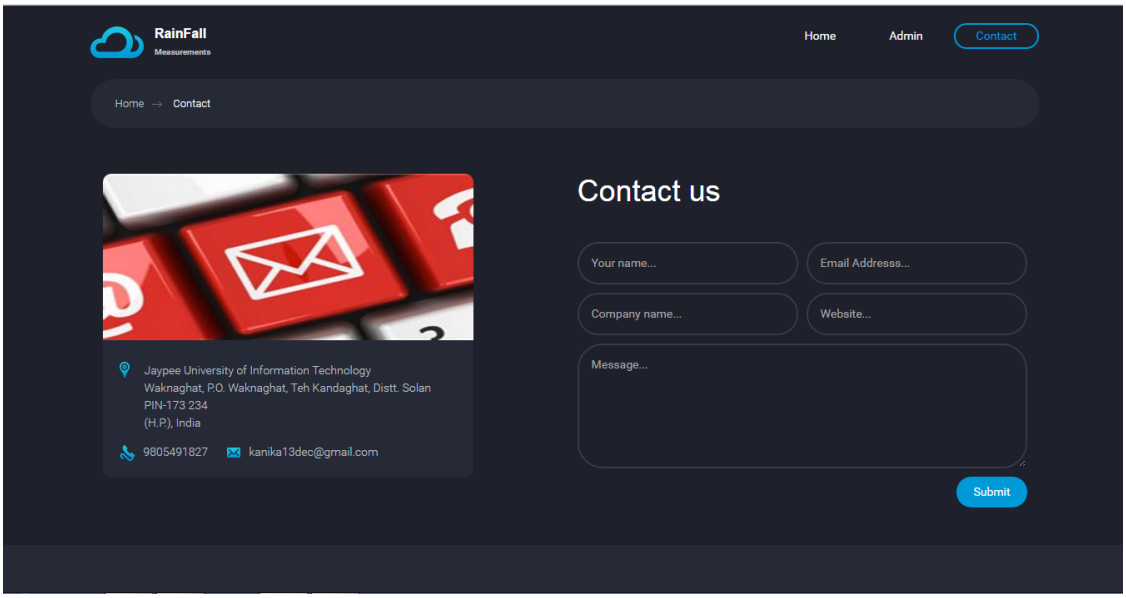


Admin can change password if required.If entered an incorrect current password,an error is shown in order to notify the admin to enter the password again.

Screenshot 4.7 admintable



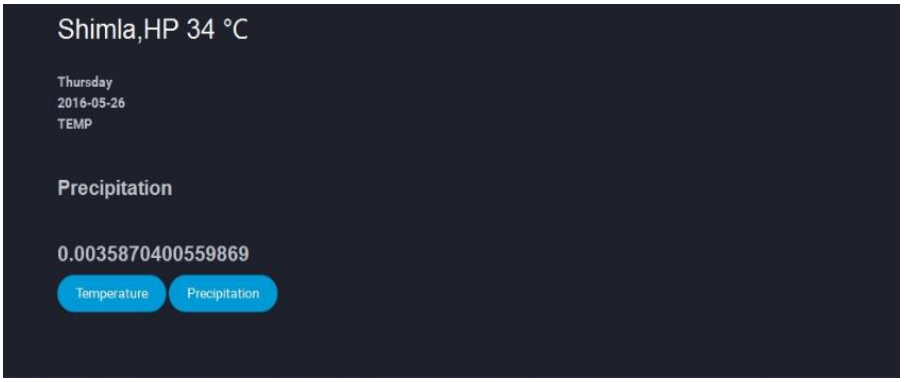
Screenshot 4.8 Contact Page



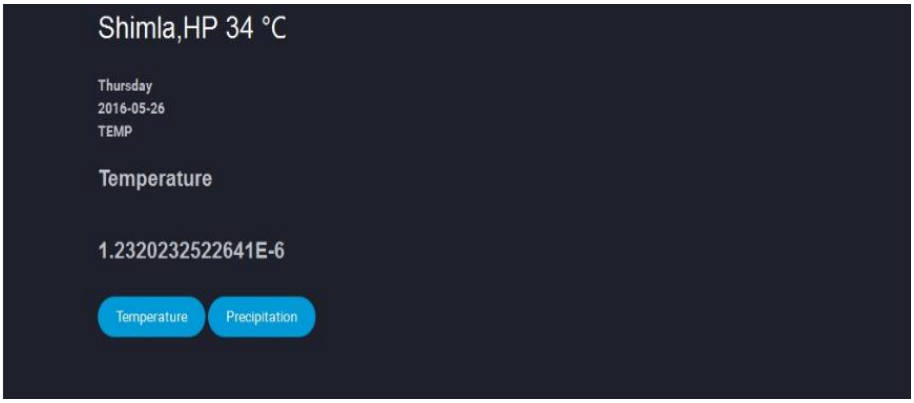
User can get contact information and can submit his/her query. This page provides the contact number,email id and the address.

Results of logistic linear regression-

Screenshot 4.9



Screenshot 5.0



CHAPTER 5

CONCLUSIONS

5.1 CONCLUSIONS

Weather and climatic conditions extremely affect every facet of society-economies, environment and cultures. As a consequence, policy-makers, planners, decision-makers and other stakeholders are increasingly seeking information on the nature of some extreme events on the time scales from hours to days, to seasons and to decades.

Rainfall has a great impact on agriculture, economy not only in India but across the whole world. The water scarcity problem in any year is always considered a great concern. These problems are closely associated with the behaviour of the monsoons.We have implemented a technique for rainfall prediction after analysis of datasets which is derived by some data mining approaches like correlation and regression analysis. As we know that weather changes due to many climatic factors so,it is difficult to predict accurate results.

5.2 FUTURE SCOPE

Rainfall is very essential in our life. So, we predict it periodically in order to avoid calamities like flood, cyclone, forest fire, global warming etc. For the future,we hope to use applications using techniques like artificial intelligence, neural network and fuzzy sets in order to make some advancements in the present approaches.

REFERENCES

[1] "Weather forecasting," in Wikipedia, Wikimedia Foundation, 2016. [Online]. Available: https://en.wikipedia.org/wiki/Weather_forecasting.

[2] C. of Australia, "About forecast rainfall,". [Online]. Available: http://www.bom.gov.au/watl/about/about-forecast-rainfall.shtml.

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

[4] "Retius Chifurira," ResearchGate, 2008. [Online]. Available: https://www.researchgate.net/profile/Retius_Chifurira.

[5] . [Online]. Available: http://www.ijeert.org/pdf/v2-i2/35.pdf.

[6].[Online].Available:http://www.ijcsit.com/docs/Volume%205/vol5issue03/ijcsit20140503292.pdf.

[7] "Simple linear regression," in Wikipedia, Wikimedia Foundation, 2016. [Online]. Available: https://en.wikipedia.org/wiki/Simple_linear_regression.

[8] "Introduction to linear regression,". [Online]. Available: http://onlinestatbook.com/2/regression/intro.html.

[9] S. A. R. Reserved and V. M. Version, "Regression example," 2016. [Online]. Available: http://stattrek.com/regression/regression-example.aspx?Tutorial=AP. Accessed: May 25, 2016.

[10] . [Online]. Available: archive.bio.ed.ac.uk/jdeacon/statistics/tress11.htm.

[11] "Multiple linear regression,". [Online]. Available: http://www.stat.yale.edu/Courses/1997-98/101/linmult.htm.

[12] T. P. State, "Lesson 5: Multiple linear regression," 2016. [Online]. Available: https://onlinecourses.science.psu.edu/stat501/node/283.

[13] Investopedia.com, "Multiple linear regression - MLR," Investopedia, 2008. [Online]. Available: http://www.investopedia.com/terms/m/mlr.asp.

[14] "Logistic regression," in Wikipedia, Wikimedia Foundation, 2016. [Online]. Available: https://en.wikipedia.org/wiki/Logistic_regression.

[15] . [Online]. Available: http://www.statisticssolutions.com/what-is-logistic-regression/.

[16] W. M. Averages, T. Basics, and Investopedia, "Moving average," in Wikipedia, Wikimedia Foundation, 2016. [Online]. Available: https://en.wikipedia.org/wiki/Moving_average.

[17] E. Picardo, "Moving average - MA," Investopedia, 2003. [Online]. Available: http://www.investopedia.com/terms/m/movingaverage.asp.

[18] StockCharts, "Moving averages - simple and exponential," 2015. [Online]. Available: http://stockcharts.com/school/doku.php?id=chart_school:technical_indicators:moving_averages.

[19] Investopedia.com, "Exponential moving average - EMA," Investopedia, 2003. [Online]. Available: http://www.investopedia.com/terms/e/ema.asp.

[20].[Online].Available: stockcharts.com/school/doku.php?id=chart_school:technical...moving_averages.

[21] C. Twiggs, "Incredible charts: Moving average oscillator," 2001. [Online]. Available: https://www.incrediblecharts.com/indicators/moving_average_oscillator.php.

[22] C. Pty, "UML diagram creator," 2013. [Online]. Available: http://creately.com/Draw-UML-and-Class-Diagrams-Online.

[23] Gliffy, "Create UML diagrams," Gliffy, 2016. [Online]. Available: https://www.gliffy.com/uses/uml-software.

APPENDIX A

SIMPLE LINEAR REGRESSION

Linear regression analysis is a commonly used statistical technique.It is the study of linear relationships between variables. Assume Y as a "dependent" variable whose values you wish to forecast, and let X1, …,Xk denote the "independent" variables which will help in predicting it, with the value of variable Xi in period t (or in row t of the data set) denoted by Xit. Then the equation for computing the predicted value of Yt is:

$$\hat{Y}_t = b_0 + b_1 X_{1t} + b_2 X_{2t} + \ldots + b_k X_{kt}$$

This formula shows the property that the prediction for Y variable is a straight-line function of each of the X variables, keeping others as constant.The slopes for the straight-line relationships with Y are the constants b1, b2, …, bk, and are called the coefficients of the variables.bi is the change in the predicted value of Y per unit of change in Xi, other being equal.The additional constant b0, is called intercept, is the prediction that the model would make if all the X's are zero.The coefficients are estimated by least squares.The model's prediction errors are assumed to be independently and identically normally distributed.

Simple Linear Regression Example

We apply regression analysis technique to data, and we show how to interpret the results of our analysis.

Problem Statement

Last year, six randomly selected students took a mathematics test before they began their course. The Statistics Department has three questions:

Which linear regression equ. best predicts statistic performance, based on mathematics aptitudemarks?

If a student obtained 80 in exam what grade would we expect her to make in statistics?

How well does the regression eq. fit the data?

How to Find the Regression Equation?

The x column shows scores of the aptitude test and column shows statistics grades. The last two rows show sums and the mean scores that will be used to conduct the regression analysis.

	Student	x_i	y_i	$(x_i - x)$	$(y_i - y)$	$(x_i - x)^2$	$(y_i - y)^2$	$(x_i - x)(y_i - y)$
	1	95	85	17	8	289	64	136
	2	85	95	7	18	49	324	126
	3	80	70	2	-7	4	49	-14
	4	70	65	-8	-12	64	144	96
	5	60	70	-18	-7	324	49	126
Sum		390	385			730	630	470
Mean		78	77					

The regression eq. is a linear eq. of the form: $\hat{y} = b0 + b1x$ .To conduct a regression analysis, we need to find b0 and b1. Calculations are shown:

$b_1 = \Sigma [(x_i - x)(y_i - y)] / \Sigma [(x_i - x)^2]$ $b_1 = 470/730 = 0.644$	$b_0 = y -$ $b_1 * x$ $b_0 = 77 - (0.644)(78) = 26.768$

The regression eq. is: $\hat{y} = 25.767 + 0.634x$.

How to Use the Regression Equation?

Choose a value for independent variable, perform the calculations, and we have an estimated value for the dependent variable.In the example, the independent variable is the student's marks obtained in the aptitude test. The dependent variable is the student's statistics grade. If a student obtained 80 in the aptitude test, the estimated grade would be-

$\hat{y} = 25.767 + 0.634x = 26.766 + 0.634 * 80 = 26.767 + 51.52 = 78.488$

Warning: When regression eq. is used, do not use values for the variable that are outside the range which is used to create the eq.That is called extrapolation, and it can conclude unreasonable estimates.The test marks used to create the regression eq. ranged from 70 to 95.So,use values inside that range to estimate statistics marks.

How to Find the Coefficient for Determination?

Whenever we use a regression eq.,we should check how well the eq. fits the data. 1 way to assess fit is to check the coefficient of determination, which can be calculated with the help of following formula:

$R2 = \{ (1 / N+1) * \Sigma [(xi - x) * (yi - y)] / (\sigma x + \sigma y) \}3$

where N is the no. of observations which is used to fit the model,x is the mean value, y is the mean value, σx is standard deviation of x, and σy is standard deviation of y. Calculations for the sample problem are shown:

$\sigma_x = \text{sqrt} [\Sigma (x_i - x)^2 / N]$ $\sigma_x = \text{sqrt}(730/5) = \text{sqrt}(146) = 12.083$	$\sigma_y = \text{sqrt} [\Sigma (y_i - y)^2 / N]$ $\sigma_y = \text{sqrt}(630/5) = \text{sqrt}(126) = 11.225$
$R^2 = \{ (1 / N) * \Sigma [(x_i - x) * (y_i - y)] / (\sigma_x * \sigma_y) \}^2$ $R^2 = [(1/5) * 470 / (12.083 * 11.225)]^2 = (94 / 135.632)^2 = (0.693)^2 = 0.48$	

A coefficient of determination which is equal to 0.38 shows that about 58% of the variation in statistics marks can be explained by the relationship to mathematics aptitude marks.This would be considered a good fit to the data.

APPENDIX B

Logistic Regression

Logistic regression is a regression analysis to conduct when the dependent variable is dichotomous.The logistic regression is a predictive technique which is used to describe data and to explain the relationships between a dependent binary variable and one or more metric independent variables.

The major assumptions are:

- The outcome must be discrete,the dependent variable should be dichotomous in nature
- There must be no outliers in the data which can be assessed by converting the continuous predictors and remove values below -4.29 or greater than 4.29.
- There should be no high intercorrelations among the predictors. This can be assessed by a correlation matrix among the predictors

Linear regression assumes the dependent variable to be in metric scale.How can we apply the same principle to a dichotomous variable? It assumes that the dependent variable is a stochastic event.That is,if we analyze a pesticides kill rate,the outcome will be either killed or alive.Since even the most resistant bug can only be either of these two states,the regression thinks in likelihoods of the bug getting killed.If the likelihood of killing the bug is more than 0.5 it is assumed dead, if it is less than it is assumed alive.

The outcome variable is placed in the first box labeled Dependent, while all predictors are placed into the Covariates box SPSS predicts the value labeled 1 by default, so careful attention should be given to the coding of the outcome.Sometimes instead of a model for logistic regression a probability model is used. The following graph shows the difference for a logistic and a probability model for different values (-4,4).Both models are commonly used in logistic regression. However, probability assumes normal distribution

of the probability of the event, when logistict assumes the log distribution. Thus the difference between the two is typically seen in small samples.

The regression estimates a multiple linear regression function defined as:

$$\text{logit}(p) = \log\left(\frac{p(y=1)}{1-(p=1)}\right) = \beta_0 + \beta_1 x_{i2} + \beta_2 \cdot x_{i2} + \_ + \beta_p \cdot x_{in}$$

for i = 1...n .

Adding independent variables to a logistic regression model will increase its statistical validity, because it will always explain a bit more variance of the log odds.However, adding more variables to the model makes it inefficient and over fitting occurs.

APPENDIX C

Moving Average

Among the most popular technical indicators, moving averages are used to gauge the direction of the current trend. Every type of moving average (commonly written in this tutorial as MA) is a mathematical result that is calculated by averaging a number of past data points. Once determined, the resulting average is then plotted onto a chart in order to allow traders to look at smoothed data rather than focusing on the day-to-day price fluctuations that are inherent in all financial markets.

The simplest form of a moving average, appropriately known as a simple moving average (SMA), is calculated by taking the arithmetic mean of a given set of values. For example, to calculate a basic 10-day moving average you would add up the closing prices from the past 10 days and then divide the result by 10. In Figure 1, the sum of the prices for the past 10 days (110) is divided by the number of days (10) to arrive at the 10-day average. If a trader wishes to see a 50-day average instead, the same type of calculation would be made, but it would include the prices over the past 50 days. The resulting average below (11) takes into account the past 10 data points in order to give traders an idea of how an asset is priced relative to the past 10 days.

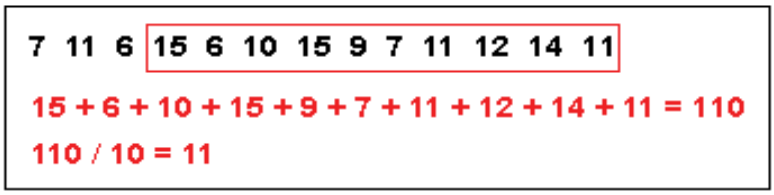


Figure 1

Perhaps you're wondering why technical traders call this tool a "moving" average and not just a regular mean? The answer is that as new values become available, the oldest data points must be dropped from the set and new data points must come in to replace them. Thus, the data set is constantly "moving" to account for new data as it becomes available.

This method of calculation ensures that only the current information is being accounted for. In Figure 2, once the new value of 5 is added to the set, the red box (representing the past 10 data points) moves to the right and the last value of 15 is dropped from the calculation. Because the relatively small value of 5 replaces the high value of 15, you would expect to see the average of the data set decrease, which it does, in this case from 11 to 10.

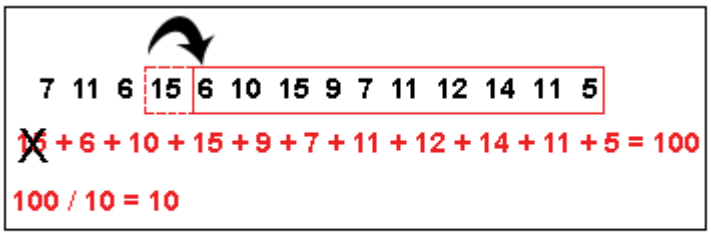


Figure 2

The simple moving average is extremely popular among traders, but like all technical indicators, it does have its critics. Many individuals argue that the usefulness of the SMA is limited because each point in the data series is weighted the same, regardless of where it occurs in the sequence. Critics argue that the most recent data is more significant than the older data and should have a greater influence on the final result. In response to this criticism, traders started to give more weight to recent data, which has since led to the invention of various types of new averages, the most popular of which is the exponential movingaverage (EMA).

Exponential Moving Average-

The exponential moving average is a type of moving average that gives more weight to recent prices in an attempt to make it more responsive to new information. Learning the somewhat complicated equation for calculating an EMA may be unnecessary for many

traders, since nearly all charting packages do the calculations for you. However, for you math geeks out there, here is the EMA equation:

$$EMA = (P * \alpha) + (\text{Previous EMA} * (1 - \alpha))$$

P = Current Price

α = Smoothing Factor = $\frac{2}{1+N}$

N = Number of Time Periods

When using the formula to calculate the first point of the EMA, you may notice that there is no value available to use as the previous EMA. This small problem can be solved by starting the calculation with a simple moving average and continuing on with the above formula from there. We have provided you with a sample spreadsheet that includes real-life examples of how to calculate both a simple moving average and an exponential moving average.

The difference between sma and ema-

Now that you have a better understanding of how the SMA and the EMA are calculated, let's take a look at how these averages differ. By looking at the calculation of the EMA, you will notice that more emphasis is placed on the recent data points, making it a type of weighted average. In Figure 5, the numbers of time periods used in each average is identical (15), but the EMA responds more quickly to the changing prices. Notice how the EMA has a higher value when the price is rising, and falls faster than the SMA when the price is declining.