Chapter 1

Introduction

The forecasting of weather conditions and in particular the prediction of precipitation is important for hydro-power operation and flood management. Mechanistic meteorology prediction models based on 3D CFD/Navier Stokes equations (Thibault and Senocak, 2009) isextremely de-manding wrt. computing power. Generating a 14 day weather forecast can easilytake 12 hours even on fast computers. Machine Learning (ML), Big Data, and use of Internet of Things (IoT) are receiving increased interest from the industry. It is well known that large amounts of data coupled with novel ML methods can produce results on par with traditional physics based models.

Due to an interest in weather monitoring in the general public, today a large number of weather stations are connected to the internet, and are thus available as cheap, distributed sensors. Additionally, several organizations that are involved in collection of meteorological data offeronline data servers with accessible Application Programming Interfaces (API) such as the HTTP based GET/REST protocols. In order to simplify experimentation with several sourcesof meteorological data it is of interest to develop a unified API, hence facilitating the extractionof data from different sources. With large quantities of data, both historical and current measurements, it is an attractive solution to use machine learning in order to predict weather conditions based on these relatively simple data sources. Using a large amount of data togetherwith novel machine learning algorithms can then compensate for lack of complex meteorological models and yield usable fore-casts with less computing time.

Simple ML models would base predictions on auto regressive (AR) structures, where, say thecurrent temperature in a location is correlated with several past temperatures in the same location. In a slightly more advanced auto regressive structure, a set of properties, e.g., the tuple (temperature, humidity, and precipitation) could be correlated with several past values of the same tuple. An even more advanced structure is the auto regressive structure with exogenous input (ARX). In such a model, the current (local) set of properties is correlated withboth past values of the same (local) set, but also with other values from the same location or values of the same properties from other locations at *current time*. Finally, in ARMAX structures, exogenous inputs at *different times* (= moving average) are used in the correlation.

1.1 Need for the System

The importance of accurate weather forecasts cannot be over emphasized as the needs for them are always craved for in virtually every aspect of life. These forecasts can be applied in the following areas:

1. **Severe weather alerts and advisories**:

   A major part of modern weather forecasting is the severe weather alerts and advisories, which the national weather services issue in the case that severe or hazardous weather is expected. This is done to protect life and property. Some of the most commonly known severe weather advisories are the severe thunderstorm and tornado warnings, as well as the recent warnings about areas that are prone to flood in some part of Nigeria by the National Meteorological Agency. Other forms of these advisories include winter weather, high wind, flood, tropical cyclone, and fog. Severe weather advisories and alerts are broadcast through the media, including radio, using emergency systems as the Emergency Alert System, which break into regular programming.

2. **Air Traffic** :

   Because the aviation industry is especially sensitive to the weather, accurate weather forecasting is essential considering the fact that a greater number of plane crashes recorded the world over have weather related causes. Just as turbulence and icing are significant in flight hazards, thunderstorms are a major problem for all aircrafts because of severe turbulence due to their updrafts and outflow boundaries, icing due to the heavy precipitation, as well as large hail, strong winds, and lightening, all of which can cause severe damage to an aircrafts in-flight. Volcanic ash is also a significant problem for aviation, as aircrafts can lose engine power with ash clouds. On a day-to-day basis, airliners are routed to take advantage of the jet stream tailwind to improve fuel efficiency. Aircrews are briefed prior to takeoff on the conditions to expect enroute and at their destination. Additionally, airports often change which runway is being used to take advantage of a headwind.

3. **Marine** :

Commercial and recreational use of waterways can be limited significantly by wind direction and speed, wave periodicity and heights, tides, and precipitation. These factors can each influence the safety of marine transit. Consequently, a variety of codes have been established to efficiently transmit detailed marine weather forecasts to vessel pilots through radio, for example the MAFOR (Marine forecast).

4. **Agriculture :**

Farmers rely on weather forecasts to decide what work to do on any particular day. For example, drying hay is only feasible in dry weather. Prolonged periods of dryness can ruin cotton, wheat, and corn crops. While crops can be ruined by drought, their dried remains can be used as a cattle feed substitute in the form of silage. Frosts and freezes play havoc with crops both during the spring and fall. For example, peach tree in full bloom can have their potential peach crop decimated by a spring freeze. Orange groves can suffer significant damage during frosts and freezes, regardless of their timing.

5. **Military** :

applications Similarly to the private sector, military weather forecasters present weather conditions to the war fighters, community. Military weather forecasters provide preflight weather briefs to pilots and provide real time resource protection services for military installations.

6. **Utility companies** :

Electricity and gas companies rely on weather forecasts to anticipate demand, which can be strongly affected by the weather. They use the quantity termed the degree-day to determine how strong of a use there will be for heating (heating degree day) or cooling (cooling degree day). These quantities are based on a daily average temperature of 650 F (180C). Cooler temperatures force heating degree days (one per degree Fahrenheit), while warmer temperatures force cooling degree days. In winter, severe cold weather can cause a surge in demand as people turn up their heating. Similarly, in summer or dry season a surge in demand can be linked with the increased use of air conditioning systems in hot weather. By anticipating a surge in demand, utility companies can produce additional supplies of power or natural gas before the price increases, or in some circumstances, supplies are restricted through the use of brown outs and blackouts.

1.2 Detailed Problem Statement

An important goal of all scientific endeavor is to make accurate predictions. The physicist or chemist who conducts an experiment in the laboratory does so in the hope of discovering certain fundamental principles that can be used to predict the outcome of other experiments based on those principles. In fact, most of the laws of science are merely very accurate predictions concerning the outcome of certain kinds of experiments. But few physical scientists are faced with more complex or challenging prediction problems than the meteorologist

According to Miller and Thompson (1975) and Ayado and Burt (2001) the problem of forecasting then, involves an attempt to observe, analyze and predict the many interrelationships between the solar energy source, the physical feature of the earth, and the properties and motions of the atmosphere. This is the basis on which weather forecasts still go wrong today.

Imperfect data:

The data of today's numerical models still includes a large helping of radiosonde observations. However, the number of radiosonde sites in the World over has actually declined over the past few decades. Developed countries in the world today, spend more money in launching weather satellites than for boring weather balloons. Satellite data are global in average, but researchers in data assimilation are still trying to figure out how this data can be "digested" properly by the models. In addition, important meteorological features still evade detection, especially over the oceans. The model results are only as good as the data in its initial conditions.

Faulty "vision" and "fudges": Today's forecasts also involve an inevitable trade-off between horizontal resolution and the length of the forecast. This is because fine resolution means lots of point at which to make calculations. This requires a lot of computer time. A forecast well into the future also requires millions or billions more calculations. If fine resolution is combined with a long range forecast, the task would choke the fastest supercomputers today. One would not get forecasts for weeks. Future improvement in computing will help speed things up.

1.3 Feasibility of the System

Feasibility study of the system is a very important stage during system design. Feasibility study is a test of a system proposal according to its workability impact on the organization, ability to meet user needs, and effective use of resources. Feasibility study decides whether the system is properly developed or not. There are five types of feasibility as mentioned below:

1. Technical Feasibility
2. Time Schedule feasibility
3. Operational feasibility
4. Implementation feasibility

1. Technical Feasibility

   Technical feasibility corresponds to determination of whether it is technically feasible to develop the

   software. Here those tools are considered, which will be required for developing the project. The tools, which are available, and tools, which will be required, are taken into account.

   Considering all above points and aspects it is observed that the cost incurred in developing this project from a technical perspective would not be too high. Thus, it is feasible for company as well as for me to develop this system.

2. Time Feasibility Time feasibility corresponds to whether sufficient time is available to complete the project. Parameters considered:

   ▪ Schedule of the project.

   ▪ Time by which the project has to be completed.

   ▪ Reporting period Considering all the above factors it was decided that the allotted time that is months was sufficient to complete the project.

3. Operational Feasibility Operational feasibility corresponds to whether users are aware of interface environment and sufficient resources are available or not. Parameters considered:

   ▪ All the relevant necessary resources for implementing and operating this system are already present in office. Bearing in mind the above factor, it was observed that the cost would be incurred in developing this project from an operational standpoint would be low. Thus, it would be operational feasible for the company.

4. Implementation Feasibility Implementation Feasibility is about basic infrastructure required to develop the system. Considering all below points, it is feasible to develop system. Factors considered:
    - All the minimum infrastructure facility required like PC, books, technical manuals are provided. Proper guidance is provided.
    - All necessary data and files are provided.

1.4 Existing Technology

The traditional forecast process employed by most NMHSs involves forecasters producing text-based, sensible, weather-element forecast products (e.g. maximum/minimum temperature, cloud cover) using numerical weather prediction (NWP) output as guidance. The process is typically schedule-driven, product-oriented and labour-intensive. Over the last decade, technological advances and scientific breakthroughs have allowed NMHSs' hydrometeorological forecasts and warnings to become much more specific and accurate.

As computer technology and high-speed dissemination systems evolved (e.g. Internet), National Weather Service (NWS) customers/partners were demanding detailed forecasts in gridded, digital and graphic formats. Traditional NWS text forecast products limit the amount of additional information that can be conveyed to the user community. The concept of digital database forecasting provides the capability to meet customer/partner demands for more accurate, detailed hydrometeorological forecasts. Digital database forecasting also offers one of the most exciting opportunities to integrate PWS forecast dissemination and service delivery, which most effectively serves the user community.

1.6 Organization of the system

Chapter 1 introduction of Weather Forecasting System. Chapter 2 In this described the surveys and presently available system. Chapter 3 described requirement analysis of Weather Forecasting System . Chapter 4: In this chapter describe the system design of Weather Forecasting System. Chapter 5 described implementation of Weather Forecasting System.

Chapter 6 shows the expected result of Weather Forecasting System.

Chapter 2

Literature Survey

2.1 Literature Survey

Weather forecasting is an essential application in meteorology and has been one of the most scientific challenging problems around the world. Weather condition is state of atmosphere at given time and the weather parameters are temperature, humidity and wind speed, The accuracy of the prediction is depends on knowledge of prevailing weather condition over large areas. Weather is the non-linear and dynamic process it varies day-to-day even minute-to-minute. Weather forecasting remains the big challenge of its data intensive and the frenzied nature

Prediction of rainfall the future values is analyzing by Temperature, Rainfall and humidity data is one of the vital parts which can be helpful to the society as well as to the economy. Some of the work in this area as: Ms. Ashwini Mandale, Mrs. Jadhawar B.A et al[1] developed on efficient Data mining techniques it used the algorithms are Artificial Neural Network and Decision tree Algorithms for meteorological to forecast weather. The performance of this algorithm would be compared with the standard performance metrics. It used two approaches that are empirical approach, dynamic approach.The comparison of results carried out by using CART to predict future values of parameters given the Month and Year. Ankita Joshi, Bhagyashri Kamble et al [2] proposed a data mining techniques of decision tree algorithm. The challenging problem is to predict the complicated weather phenomena with limited observations.To predict the weather by numerical means the meteorologists have developed atmospheric models that is approximate by using mathematical equations. They found 82% accuracy in variation of rainfall prediction.

M.Viswambari and Dr.R.Anbu Selvi [3] implemented the data mining techniques is to forecasting rainfall, wind pressure, humidity to forecast the weather data about past historical and future value. Classification is the problem t o identify the set of categories a new observation regards, on the basis of a training data containing the observations whose category membership is known The goal of any supervised learning algorithm is to find a correct output to minimize errors. T V Rajini Kanth, V V SSS Balaram et al [5] implemented the k-clustering technique to grouping the similar data sets to forecast the temperature, rainfall so it would be need higher scientific techniques like machine learning algorithms for effective study and predictions of weather conditions using linear regression. It is found through kmeans cluster analysis.

Pinky saikia dutta and Hitesh tahbilder [6] implemented is done by using multiple linear regressions that presented in the data mining technique in forecasting monthly rainfall of Assam. It was carried out by traditional statistical technique and Multiple Linear Regression. The data include Six years period collected locally from Regional Meteorological Center. They found 63% accuracy in variation of rainfall for our proposed model.

Nikhil Sethi, Dr.Kanwal Garg [7] Rainfall prediction model is implemented with empirical statistical technique. It is used the multiple linear regression (MLR) technique for the early prediction of rainfall. There are two approaches used for predicting rainfall. One is Empirical another one is Dynamical approach. The results prove that there is a close relations between the predicted and actual rainfall amount.

Manisha Kharola and Dinesh Kumar[8] described the back propagation algorithm.ANNs are capable of producing accurate predictions of weather variables for small scale of imperfect datasets. The actual network output is subtracted from the desired outputs in an error signal is produced to predict the future weather with the help of back propagation training algorithm.

Sanjay D. Sawaitul [17] developed a back propagation algorithm for weather forecasting and processing information. They provided the information of coming weather after some period amount of time by changing some parameters of what will be the effect on other parameters are recorded shown on wireless display, to prevent the adverse effect of climate change.

2.2 Presently Available Systems

1. Doppler radar:

Doppler Radar is the meteorologist's window into observing severe storms. With 159 radar towers across the United States, NOAA's National Weather Service has comprehensive coverage of the continental U.S. and partial coverage of Alaska, Hawaii, Puerto Rico and

Guam. Doppler radar detects all types of precipitation, the rotation of thunderstorm clouds, airborne tornado debris, and wind strength and direction.

2. Satellite data

Weather Satellites monitor Earth from space, collecting observational data our scientists analyze. NOAA operates three types of weather satellites. Polar orbiting satellites orbit the Earth close to the surface, taking six or seven detailed images a day. Geostationary satellites stay over the same location on Earth high above the surface taking images of the entire Earth as frequently as every 30 seconds. Deep space satellites face the sun to monitor powerful solar storms and space weather. NOAA also uses data from satellites operated by other agencies and countries.

3. Radiosondes

Radiosondes are our primary source of upper-air data. At least twice per day, radiosondes are tied to weather balloons and are launched in 92 locations across the United States. In its two hour trip, the radiosonde floats to the upper stratosphere where it collects and sends back data every second about air pressure, temperature, relative humidity, wind speed and wind direction. During severe weather, we usually launch weather balloons more frequently to collect additional data about the storm environment.

4. Automated surface-observing systems

ASOS (automated surface observing systems) constantly monitor weather conditions on the Earth's surface. More than 900 stations across the U.S. report data about sky conditions, surface visibility, precipitation, temperature and wind up to 12 times an hour. Nearly 10,000 volunteer NWS Cooperative Observers collect and provide us additional temperature, snowfall and rainfall data. The observational data our ASOS and volunteers collect are essential for improving forecasts and warnings.

5. Supercomputers

Climate Operational Supercomputer System (WCOSS) is the backbone of modern forecasting. With 5.78 petaflop computing capacity it can process quadrillions of calculations per second. Our supercomputers are almost 6 million times more powerful than your average desktop computer. Observational data collected by doppler radar, radiosondes, weather satellites, buoys and other instruments are fed into computerized NWS numerical forecast models. The models use equations, along with new and past weather data, to provide forecast guidance to our meteorologists.

6. AWIPS

AWIPS (NOAA’s Advanced Weather Information Processing System) is a computer processing system that combines data from all the previous tools into a graphical interface that our forecasters use to analyze data and prepare and issue forecasts, watches, warnings. This system uses NOAA supercomputers to process data from doppler radar, radiosondes, weather satellites, ASOS, and other sources using models and forecast guidance products. After meteorologists prepare the forecasts, AWIPS generates weather graphics and hazardous weather watches and warnings. All this helps our meteorologists create more accurate forecasts and faster than ever before.

Chapter 3

Requirement Analysis

Functional Requirements:

- The operator shall be able to input the weather periods to the system to view the desired weather parameters within the particular periods.
- The system shall be able to produce minimum, maximum and the average data of a particular weather parameter when it is requested by an operator.
- The system shall provide the following weather parameters: temperature, pressure, wind speed & direction, rainfall, and humidity.

3.1 Software and Hardware Requirements for Weather Forecasting–

This System is for System Administrators so minimum Requirements will Be

Hardware/ Software	Hardware / Software element	Specification /version
Hardware	Processor	i3
	RAM	2GB
	Hard Disk	250GB
Software	Windows OS	. Jupyter NoteBook. Python 3.
	Python IDE	
	Visual Studio Code	

3.2 System Requirements

The system requirement specification provides a narrative description of each subsystem and a definition of all data that flow between subsystems. The system requirement specification forms the foundation of all engineering work at that follows. These specifications analyze the behavior of the system with the external events and the functionality of system elements.

This phase come into picture while developing software when analysis phase is over. This step takes care that whatever problem has been recognized is correct because the user requirements keeps changing and for creation of successful software it’s necessary that the software deliver goods at every level of work. For this the requirements and problems to be solved must be thoroughly known and they must be worked upon.

Chapter 4

System Design

4.1 Architecture

Main elements in this figure are Dataset, Information in Dataset, Algorithm for prediction, Data Gathering, Data Processing, different charts for showing results. All information of weather and climate are stored in dataset. The data is added to dataset from past climate conditions which contains humidity, rain, winds, temperature, etc and it is imported to program and used for data processing then we apply linear regression algorithm in python and after applying algorithm the predicted climate condition will show charts like histogram and scatterplot.

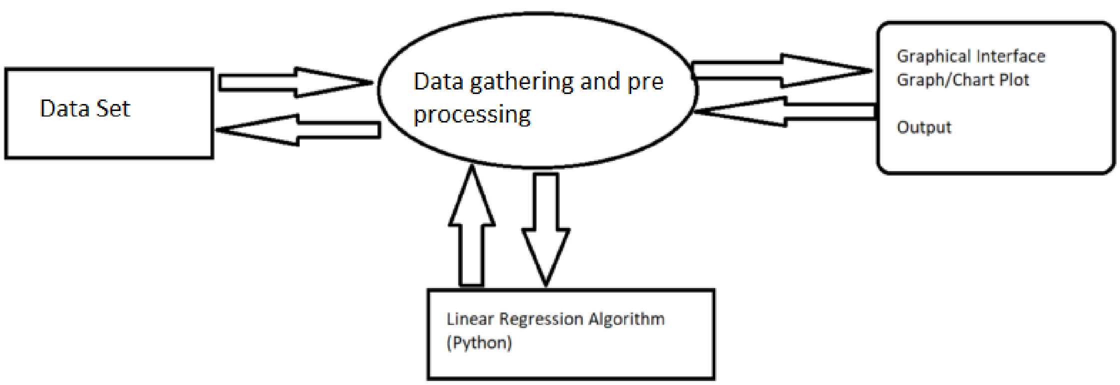


Fig. 4.1 Architecture of Weather Forecasting

4.2 Data Flow Diagrams

The context level data flow diagram (dfd) is describe the whole system. The (o) level dfd describe the all user module who operate the system. The fig 4.2.1 shows Weather Forecasting System shows the administrator can operate the system. Weather Forecasting Application between administrator and weather data. Means Admin can use and manage the weather data for forecasting.

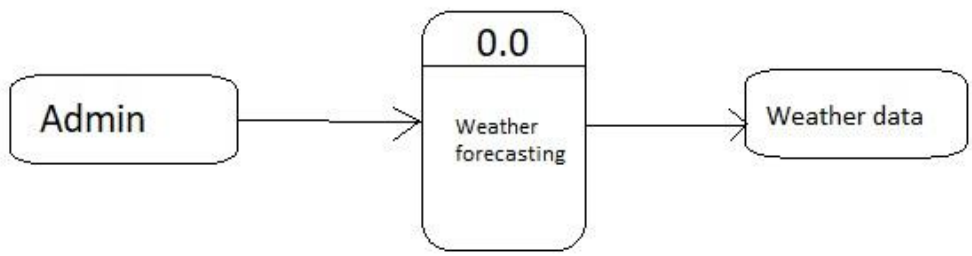


Fig 4.2.1 DFD 0 Level Diagram

A level 1 DFD notates each of the main sub-processes that together form the complete system. In Weather Forecasting the administrator add the dataset which contains weather data and it will store and analyse the data.

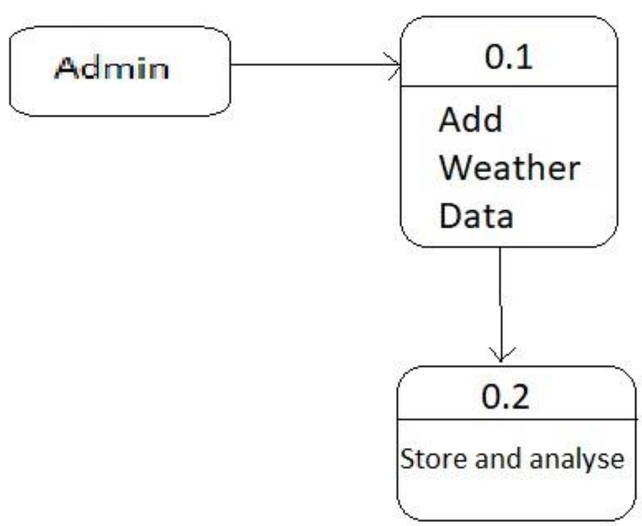


Fig 4.2.2 DFD 1 Level Diagram

Level 2 DFD for Weather forecasting System is also the highest abstraction of the data flow diagram. In Weather forecasting administrator can compare predicted weather data with dataset and shows a prediction with the help of charts like scatterplot and histogram. .

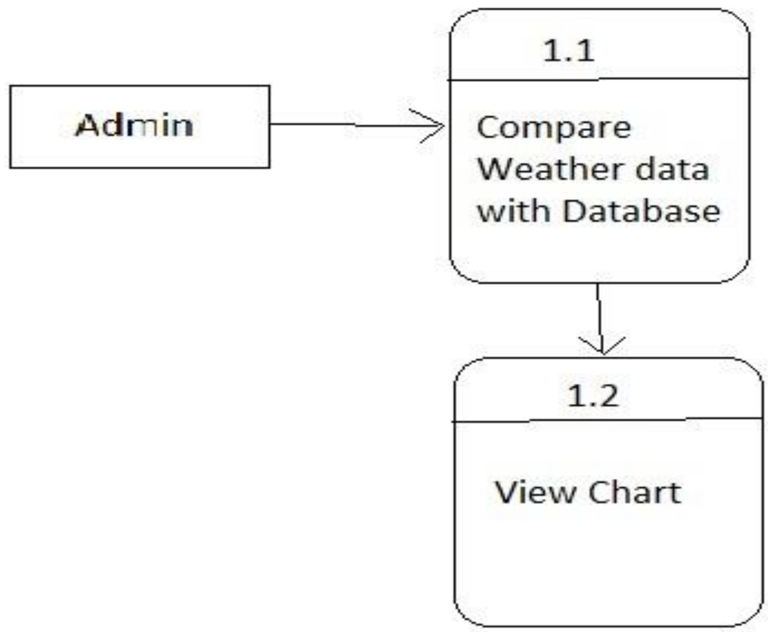


Fig 4.2.3 DFD 2 Level Diagram

4.3 ER Diagram

Entity Relationship Diagram for Weather Forecasting System in an ER model of Weather Forecasting Entities. It is basically a graphical representation of Forecasting. This was complete ER model of Weather Forecasting System project. The box represent entities and oval shape represent attributes. Diamond shape represent the relationship between two entities.

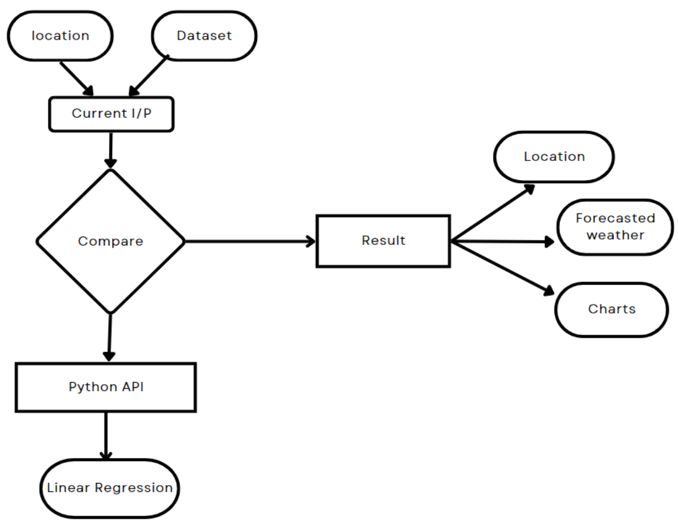


Fig. 4.3 ER – Diagram of Weather Forecasting System

4.4 Sequence Diagram

A sequence diagram is a unified modeling language (UML) diagram that illustrate the sequence of messages between objects in an interaction. A sequences diagram consist of a group of objects that are represented by life lines and the messages that they exchange over time during the interaction. A sequences diagram shows the sequence of messages passed between objects. Sequence diagram can also shows the control between structure between objects.

For example, life line in a sequence diagram for a Weather Forecasting can represent a admin and Weather Prediction The communication between the admin, dataset and prediction algorithm are represented by messages passed between them. The sequences diagram shows the objects of the messages between the objects.

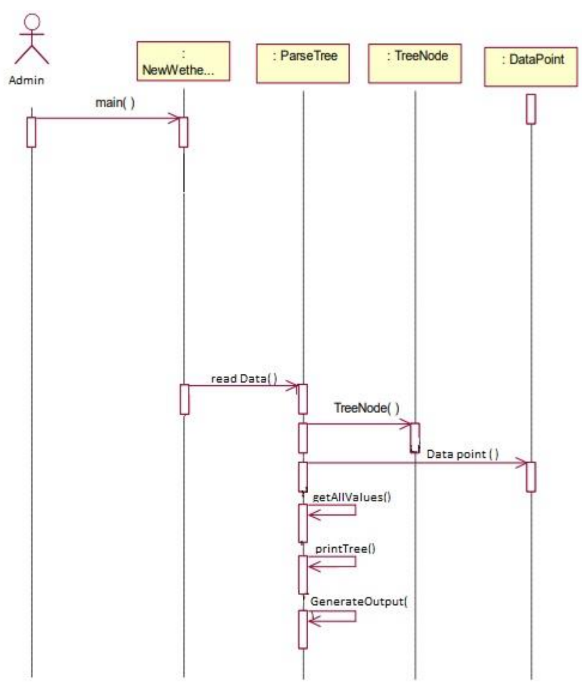


Fig. 4.4 Sequence diagram of Weather Forecasting System

4.5 Use Case Diagram

The use case diagram are usually referred to as behavior diagram used to describe the actions of all user in a system. All user describe in use case are actors and the functionality as action of system. The use case diagram example for the Weather Forecasting has one main illustration. These illustrations describe the system's general and specific processes.

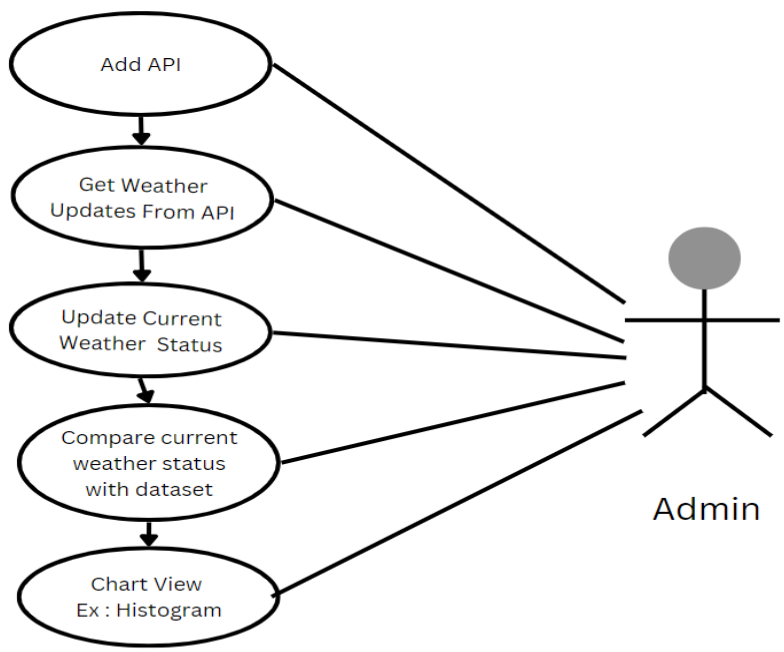


Fig 4.6 Use Case Diagram of Weather Forecasting System

Chapter 5

Implementation

5.1 Modules Description

In this project we have Two modules

1) Data gathering and pre - processing.

2) Applying Algorithm for prediction .

Explanation:

1) In this module we first gather the data(dataset) for our prediction model.Data comes in all forms, most of it being very messy and unstructured. They rarely come ready to use. Datasets, large and small, come with a variety of issues- invalid fields, missing and additional values, and values that are in forms different from the one we require. In order to bring it to workable or structured form, we need to "clean" our data, and make it ready to use. Some common cleaning includes parsing, converting to one-hot, removing unnecessary data, etc.

In our case, our data has some days where some factors weren't recorded. And the rainfall in cm was marked as T if there was trace precipitation. Our algorithm requires numbers, so we can't work with alphabets popping up in our data. so we need to clean the data before applying it on our model.

2)Once the data is cleaned, In this module that cleaned data can be used as an input to our Linear regression model. Linear regression is a linear approach to form a relationship between a dependent variable and many independent explanatory variables. This is done by plotting a line that fits our scatter plot the best, ie, with the least errors. This gives value predictions, ie, how much, by substituting the independent values in the line equation.

We will use Scikit-learn's linear regression model to train our dataset. Once the model is trained, we can give our own inputs for the various columns such as temperature, dew point, pressure, etc. to predict the weather based on these attributes.

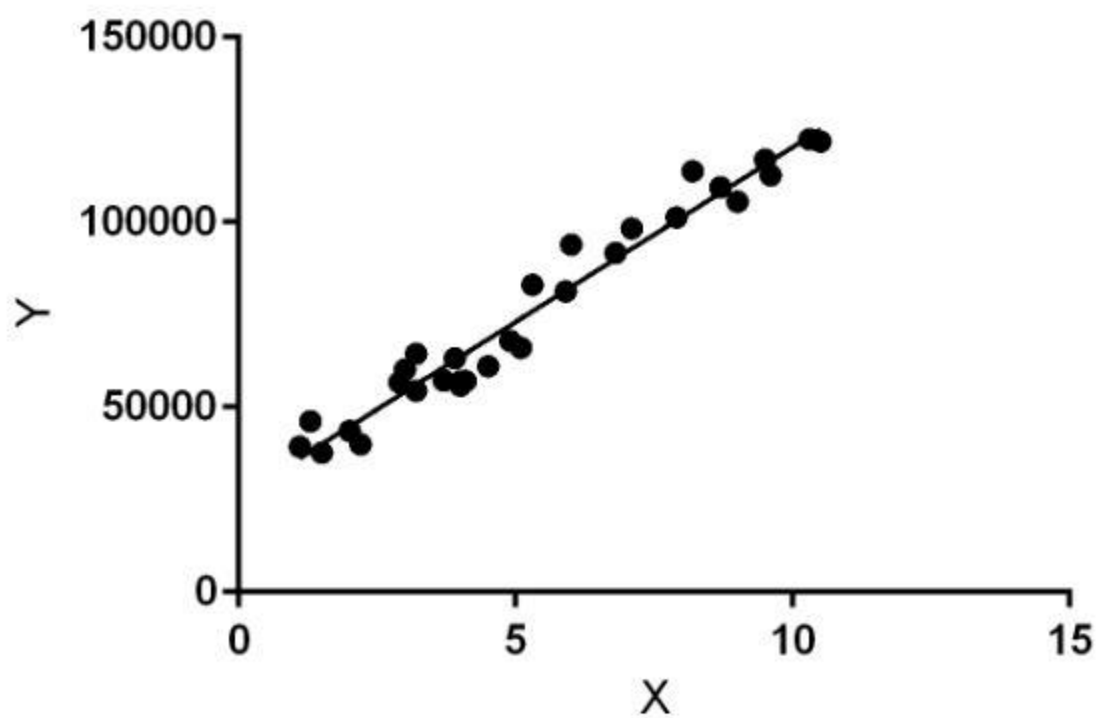
Module Outcomes:

1) By the end of the first module the fully cleaned and useful data is available for the apply the algorithm for the prediction.

1) By the end of the second module the actual prediction will be happen the outcome is the amount of rainfall in inches based upon the users input.

Algorithm:

Linear Regression is a machine learning algorithm based on supervised learning. It performs a **regression task**. Regression models a target prediction value based on independent variables. It is mostly used for finding out the relationship between variables and forecasting. Different regression models differ based on – the kind of relationship between dependent and independent variables, they are considering and the number of independent variables being used.



Linear regression performs the task to predict a dependent variable value (y) based on a given independent variable (x). So, this regression technique finds out a linear relationship between x (input) and y(output). Hence, the name is Linear Regression. In the figure above, X (input) is the work experience and Y (output) is the salary of a person. The regression line is the best fit line for our model.

Hypothesis function for Linear Regression :

y=mx+c

Where

y is the response variable.

x is the predictor variable.

m and **c** are constants which are called the coefficients.

5.2 Dataset

The dataset is a public weather dataset from Austin, Texas available on Kaggle

austin_weather.csv

Columns:

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

DewPointHighF-

High dew point, in degrees Fahrenheit

DewPointAvgF-

Average dew point, in degrees Fahrenheit

DewPointLowF-

Low dew point, in degrees Fahrenheit

HumidityHighPercent-

High humidity, as a percentage

HumidityAvgPercent-

Average humidity, as a percentage

HumidityLowPercent-

Low humidity, as a percentage

SeaLevelPressureHighInches-

High sea level pressure, in inches of mercury

SeaLevelPressureAvgInches-

Average sea level pressure, in inches of mercury

SeaLevelPressureLowInches-

Low sea level pressure, in inches of mercury

VisibilityHighMiles-

High visibility, in miles

VisibilityAvgMiles-

Average visibility, in miles

VisibilityLowMiles-

Low visibility, in miles

WindHighMPH-

High wind speed, in miles per hour

WindAvgMPH-

Average wind speed, in miles per hour

WindGustMPH-

Highest wind speed gust, in miles per hour

PrecipitationSumInches-

Total precipitation, in inches ('T' if trace)

Events-

Adverse weather events (' ' if None)

5.3 Design And Implementation

3.2Source Code

```
# importing libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

# read the data in a pandas dataframe
data = pd.read_csv("C:/Users/TEMP.SANDEEP/Desktop/austin_weather.csv")

#seeing head values
data.head(5)
#seeing shape of the dataset
data.shape

#filling missing NULL values by column means
data.fillna(data.mean())

# drop or delete the unnecessary columns in the data.
data = data.drop(['Events', 'Date', 'SeaLevelPressureHighInches',
'SeaLevelPressureLowInches'], axis = 1)

# some values have 'T' which denotes trace rainfall
# we need to replace all occurrences of T with 0
# so that we can use the data in our model
data = data.replace('T', 0.0)

# the data also contains '-' which indicates no
# or NIL. This means that data is not available
# we need to replace these values as well.
data = data.replace('-', 0.0)
```

```
# dataframe created with
# the above data array
df = pd.DataFrame(data)

# create histogram for numeric data
df.hist()

# show plot
plt.show()
#basic static

# save the data in a csv file
data.to_csv('C:/Users/Windows /austin_final_final.csv')

# importing libraries
import pandas as pd
import numpy as np
import sklearn as sk
from sklearn.linear_model import LinearRegression
import matplotlib.pyplot as plt

# read the cleaned data
data = pd.read_csv("C:/Users/TEMP.SANDEEP/Desktop/austin_final_final.csv")

# the features or the 'x' values of the data
# these columns are used to train the model
# the last column, i.e, precipitation column
# will serve as the label
X = data.drop(['PrecipitationSumInches'], axis = 1)

# the output or the label.
```

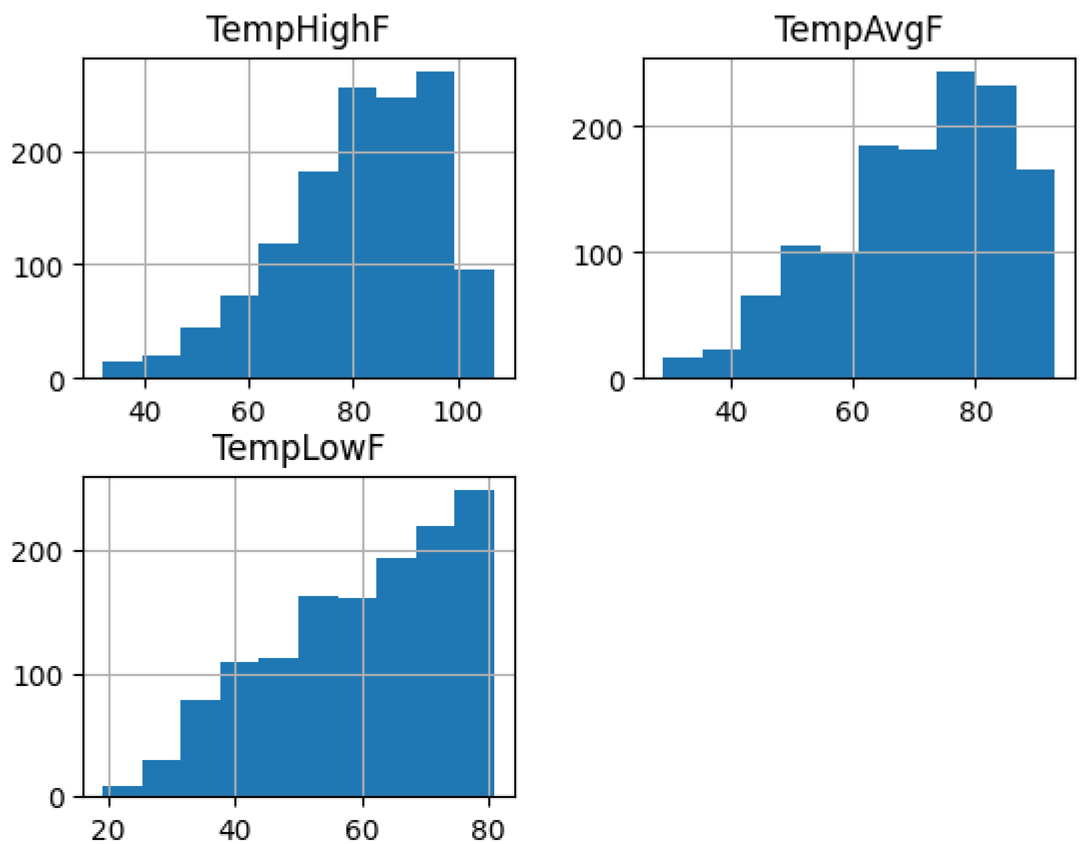
```
Y = data['PrecipitationSumInches']
# reshaping it into a 2-D vector
Y = Y.values.reshape(-1, 1)

# consider a random day in the dataset
# we shall plot a graph and observe this
# day
day_index = 798
days = [i for i in range(Y.size)]

# initialize a linear regression classifier
clf = LinearRegression()
# train the classifier with our
# input data.
clf.fit(X, Y)

# give a sample input to test our model
# this is a 2-D vector that contains values
# for each column in the dataset.
inp = np.array([[74], [60], [45], [67], [49], [43], [33], [45],
                [57], [29.68], [10], [7], [2], [0], [20], [4], [31]])
inp = inp.reshape(1, -1)

# print the output.
print('The precipitation in inches for the input is:', clf.predict(inp))

# plot a graph of the precipitation levels
# versus the total number of days.
# one day, which is in red, is
# tracked here. It has a precipitation
# of approx. 2 inches.
print("the precipitation trend graph: ")
```

```
plt.scatter(days, Y, color = 'g')
plt.scatter(days[day_index], Y[day_index], color ='r')
plt.title("Precipitation level")
plt.xlabel("Days")
plt.ylabel("Precipitation in inches")

plt.show()
x_vis = X.filter(['TempAvgF', 'DewPointAvgF', 'HumidityAvgPercent',
                  'SeaLevelPressureAvgInches', 'VisibilityAvgMiles',
                  'WindAvgMPH'], axis = 1)

# plot a graph with a few features (x values)
# against the precipitation or rainfall to observe
# the trends
print("Precipitation vs selected attributes graph: ")
for i in range(x_vis.columns.size):
    plt.subplot(3, 2, i + 1)
    plt.scatter(days, x_vis[x_vis.columns.values[i][:100]],
                                            color = 'g')
    plt.scatter(days[day_index],
                x_vis[x_vis.columns.values[i]][day_index],
                color ='r')
    plt.title(x_vis.columns.values[i])
plt.show()
```

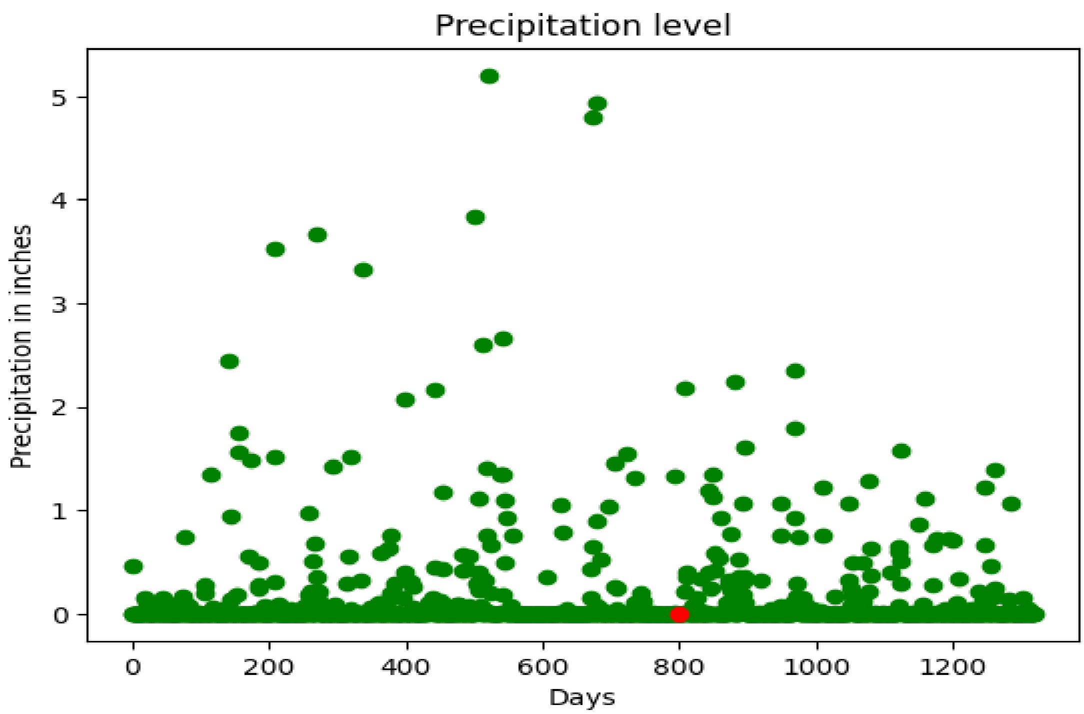
Chapter 6

Result

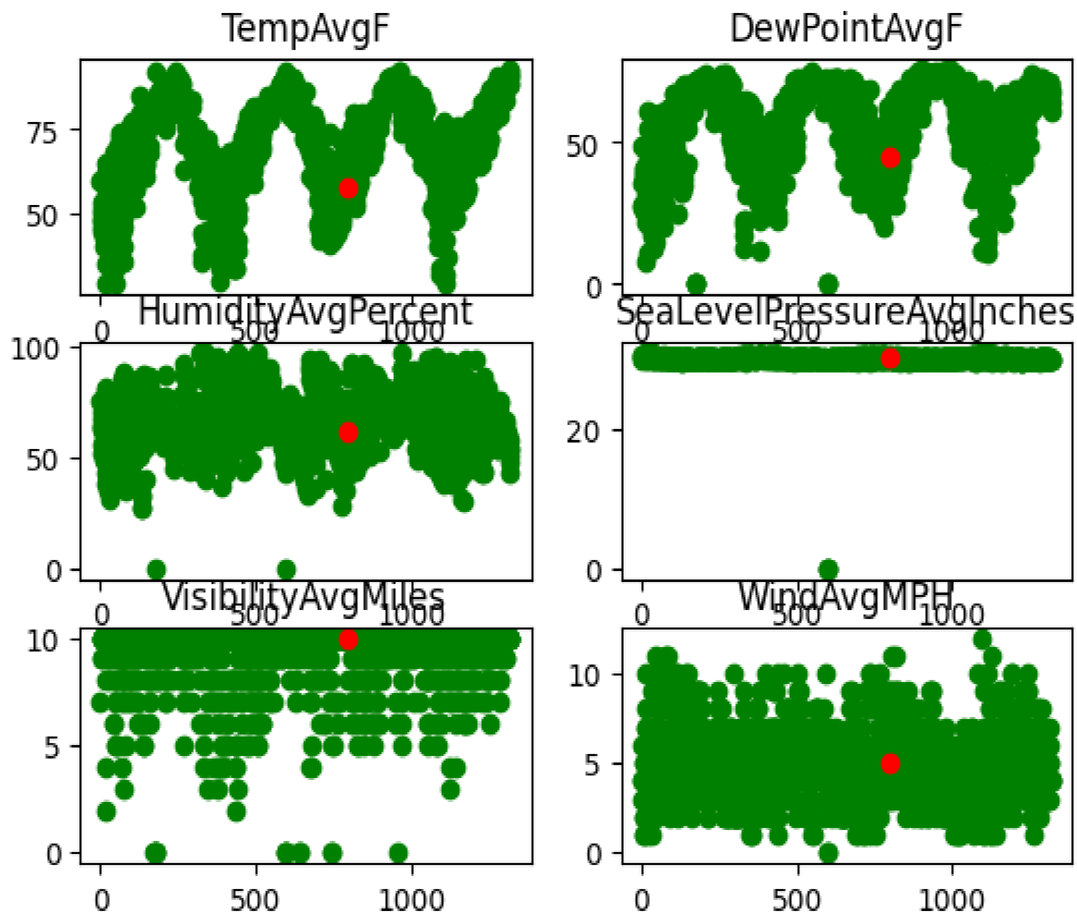
The goal of this work is to predict the temperature using an Existing weather dataset. We will get 4 charts based on our dataset one will show average temperature 2nd will show precipitation level other charts will present humidity, visibility and wind pressure.



This 3 Charts is Showing Temperature based on Different Perspectives like High Temperature it shows number of maximum temperature is between 80 to100 fahrenheit and Low Temperature Chart Shows Number of lowest Temperature recorded is between 20 to 40 Fahrenheit Average Temperature chart shows number of Average Temperature recorded is 60 to 80 Fahrenheit.



This Chart Shows Precitipation of weather. Precipitation is condensation of atmospheric water vapour that falls under gravitational pull from clouds. The Highest Pecitipation is more than 5 inches and it happens between 400 to 600 days and other is lower than 1 or 0 and this data is recorded from 3 years



This last chart shows all the Average Temperature, Humidity, Visibility, Wind,Sea Level, DewPoint This all predictions are used from past 1000 days. This charts is called Scatterplot. The Results are In Fahrenheit measure.

A day (in red) having precipitation of about 2 inches is tracked across multiple parameters (the same day is tracker across multiple features such as temperature, pressure, etc). The x-axis denotes the days and the y-axis denotes the magnitude of the feature such as temperature, pressure, etc. From the graph, it can be observed that rainfall can be expected to be high when the temperature is high and humidity is high.

Chapter 7

Conclusion

We successfully predicted the Temperature, Rainfall, Humidity and etc using the linear regression but here this is not very accurate only some times any way it depends upon the climate changes to season to season. Here we are taking only summer season weather data set it only useful to predict rainfall in summer season. weather forecasts are increasingly accurate and useful, and their benefits extend widely across the economy. While much has been accomplished in improving weather forecasts, there remains much room for improvement. The forecasting community is working closely with multiple stakeholders to ensure that forecasts and warnings meet their specific needs. Simultaneously, they are developing new technologies and observational networks that can enhance forecaster skill and the value of their services to their users. opportunities exist for increasing forecast skill at all time ranges. However, realizing these opportunities will require further research, close international cooperation and coordination, improved observations of the atmosphere, ocean, and land surface, and the incorporation of these observations into numerical models. Also, benefit will be derived from higher spatial resolution of numerical models; increasingly powerful supercomputers; wider use and improvement of model ensembles; the development of data mining and visualization methods that enable forecasters to make better use of model guidance; and collaborative forecast development activities among operational forecasters and researchers.

Chapter 8

References

[1]Ms. Ashwini Mandale, Mrs. Jadhawar B.A, "Weather Forecast Prediction: A Data Mining Application", International Journal of Engineering Research and General Science Volume 3, Issue 2, MarchApril, 2015, ISSN 2091-2730.

[2] Ankita Joshi, Bhagyashri Kamble, Vaibhavi Joshi, Komal Kajale, Nutan Dhange, "Weather Forecasting And Climate Changing Using Data Mining Application", International Journal Of Advanced Research In Computer And Communication Engineering,Vol. 4, Issue 3, March 2015. [3] M. Viswambari, Dr. R. Anbu Selvi, "Data Mining Techniques To Predict Weather: A Survey", ISSN 2348 – 7968, IJISET- International Journal Of Innovative Science, Engineering & Technology, Vol. 1 Issue 4, June 2014.

[4] Divya Chauhan, Jawahar Thakur, "Data Mining Techniques for Weather Prediction: A Review", International Journal On Recent and Innovation Trends In Computing And Communication ISSN: 2321- 8169 Volume: 2 Issue: 8, 2014.

[5] T V Rajini Kanth, V V Sss Balaram And N.Rajasekhar,"Analysis Of Indian Weather Data Sets Using Data Mining Techniques", Pp. 89– 94, 2014. © Cs & It-Cscp 2014.

[6] Pinky Saikia Dutta, Hitesh Tahbilder,"Prediction Of Rainfall Using Data Mining Technique Over Assam", Pinky Saikia Dutta et.al / Indian Journal Of Computer Science And Engineering (IJCSE), ISSN: 0976-5166 Vol. 5 No.2 Apr-May 2014.

[7] Nikhil Sethi, Dr.Kanwal Garg,"Exploiting Data Mining Technique For Rainfall Prediction", Nikhil Sethi Et Al, / (IJCSIT) International Journal Of Computer Science And Information Technologies, Vol. 5 (3), 2014, 3982-3984.

[8] Manisha Kharola and Dinesh Kumar, "Efficient Weather Prediction by Back-Propagation Algorithm", IOSR Journal of Computer Engineering (IOSR-JCE) E-ISSN: 2278-0661, PISSN: 2278-8727volume 16, Issue 3, Ver. IV (May-Jun. 2014), Pp 55-58.

Weblinks:-

1) https://towardsdatascience.com/introduction-to-machine-learning-algorithms-linear-regression-14c4e325882a

2) https://www.kaggle.com/grubenm/austin-weather