Real Time Air Quality Monitoring &Weather Forecasting System

Project By

S. ASRITHA

G. VAISHNAVI

K. SAI RAKESH

G. RAVI TEJA

TABLE OF CONTENTS

CHAPTERS

ABSTRACT

Weather forecasting is the application of science and technology to predict the state of the atmosphere for a given location.Ancient weather forecasting methods usually relied on observed patterns of events, also termed pattern recognition. For example, it might be observed that if the sunset was particularly red, the following day often brought fair weather.However, not all of these predictions prove reliable.

Here this system will predict weather based on parameters such as temperature, humidity and wind. User will enter current temperature; humidity and wind, System will take this parameter and will predict weather(rainfall in inches) from previous data in database(dataset). The role of the admin is to add previous weather data in database, so that system will calculate weather(estimated rainfall in inches) based on these data. Weather forecasting system takes parameters such as temperature, humidity, and wind and will forecast weather based on previous record therefore this prediction will prove reliable. This system can be used in Air Traffic, Marine, Agriculture, Forestry, Military, and Navy etc.

1.INTRODUCTION

- **Data Warehousing**

Data Warehouse is electronic storage of a large amount of information by a business which is designed for query and analysis instead of transaction processing. It is a process of transforming data into information and making it available to users for analysis.

- **Data Mining**

Data mining is looking for hidden, valid, and potentially useful patterns in huge data sets. Data Mining is all about discovering unsuspected/ previously unknown relationships amongst the data.It is a multi-disciplinary skill that uses machine learning, statistics, AI and database technology.

1.1. Introduction

Rainfall Prediction is the application of science and technology to predict the amount of rainfall over a region. It is important to exactly determine the rainfall for effective use of water resources, crop productivity and pre-planning of water structures.

In this project, we used Linear Regression to predict the amount of rainfall. Linear Regression tells us how many inches of rainfall we can expect.

1.2 Problem Definition

It is important to exactly determine the rainfall for effective use of water resources, crop productivity and pre-planning of water structures.

1.3 Scope

It tells us how many inches of rainfall we can expect.

1.4 Purpose

There are several reasons why weather forecasts are important. They would certainly be missed if they were not there. It is a product of science that impacts the lives of many people. The following is a list of various reasons why weather forecasts are important:

1. Helps people prepare for how to dress (i.e. warm weather, cold weather, windy weather, rainy weather)
2. Helps businesses and people plan for power production and how much power to use (i.e. power companies, where to set thermostat)
3. Helps people prepare if they need to take extra gear to prepare for the weather (i.e. umbrella, rain coat, sun screen)
4. Helps people plan outdoor activities (i.e. to see if rain/storms/cold weather will impact outdoor event)
5. Helps curious people to know what sort of weather can be expected (i.e. a snow on the way, severe storms)
6. Helps businesses plan for transportation hazards that can result from the weather (i.e. fog, snow, ice, storms, clouds as it relates to driving and flying for example)
7. Helps people with health related issues to plan the day (i.e. allergies, asthma, heat stress)
8. Helps businesses and people plan for severe weather and other weather hazards (lightning, hail, tornadoes, hurricanes, ice storms)
9. Helps farmers and gardeners plan for crop irrigation and protection (irrigation scheduling, freeze protection)

1.5 Problem and 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 Proposed System

User will enter current temperature; humidity and wind, System will take this parameter and will predict weather from previous data in database. The role of the admin is to add previous weather data in database, so that system will calculate weather based on these data. Weather forecasting system takes parameters such as temperature, humidity, and wind and will forecast weather based on previous record therefore this prediction will prove reliable.

2.REQUIREMENTS

2.1. Platform Requirements

Hardware/Software	Hardware / Software element	Specification /version
Hardware	Processor	i3
	RAM	2GB
	Hard Disk	250GB
Software	OS	Windows,Linux.
	Python IDE Microsoft Azure	JupyterNoteBook. Python 3.

2.2. 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.

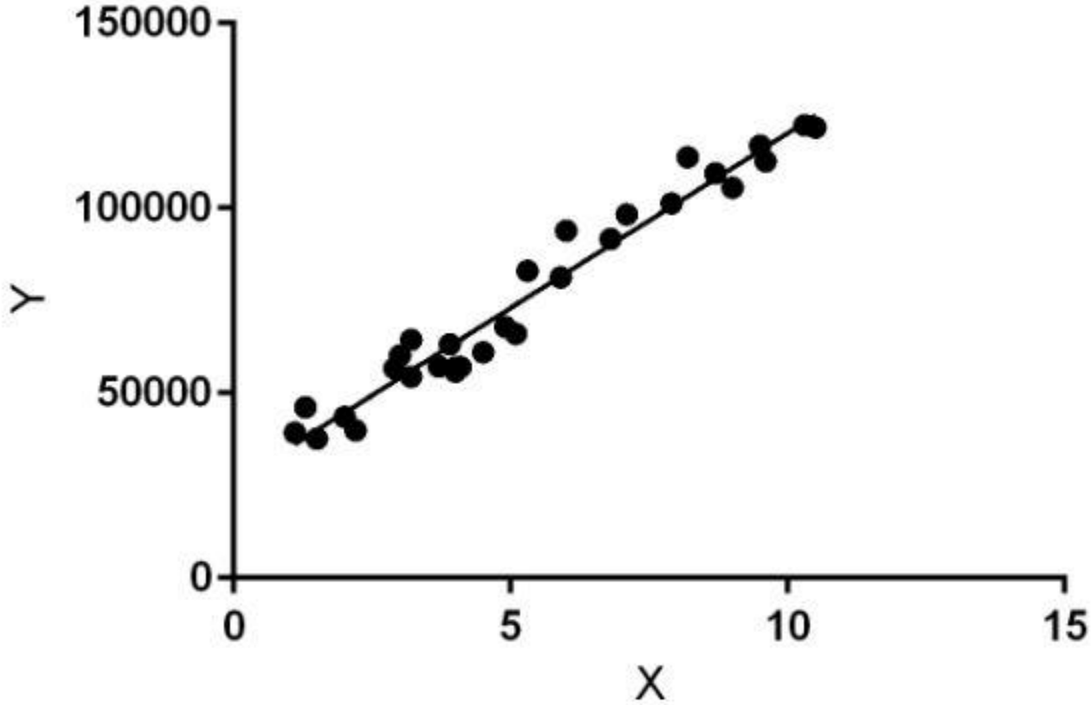
Module Outcomes:

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 consideringand 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.

2.3. Data Set

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)

3.DESIGN AND IMPLEMENTATION

3.1 Algorithms:

Linear Regression:

Module-1 : Data gathering and pre - processing.

Module-2: Applying Algorithm for prediction .

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/TEMP.SANDEEP/Desktop/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')
```

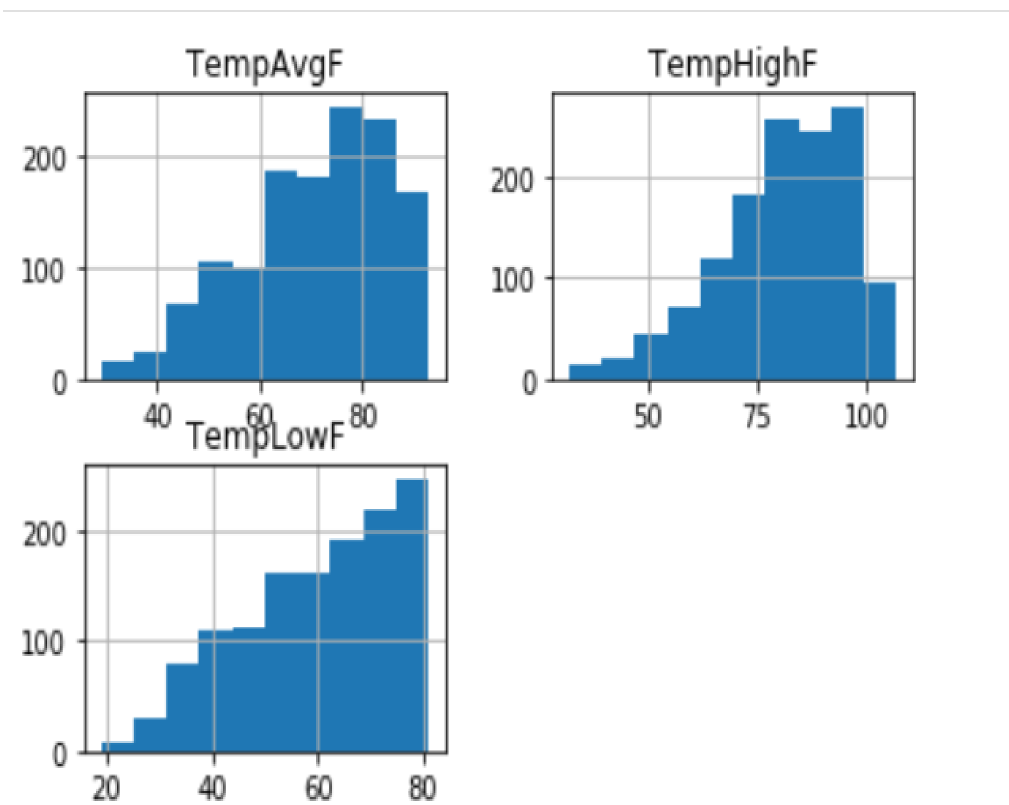
```
plt.scatter(days[day_index],
x_vis[x_vis.columns.values[i]][day_index],
color ='r')
plt.title(x_vis.columns.values[i])
plt.show()
```

OUTPUT:

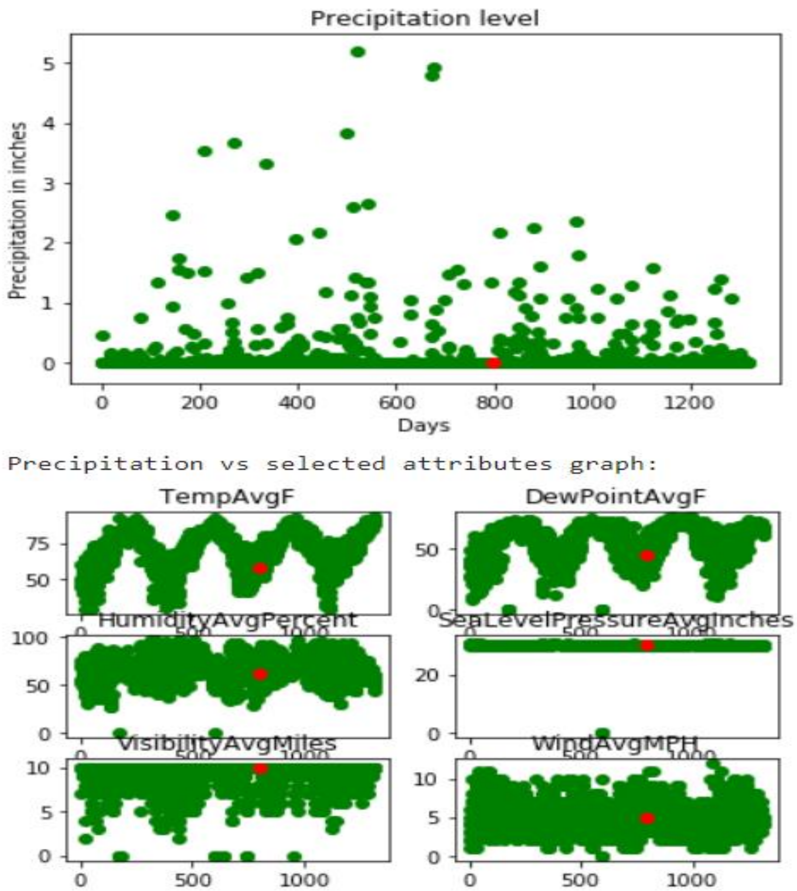
The precipitation in inches for the input is: [[1.33868402]]

Graphs:

1) Histogram for Temp

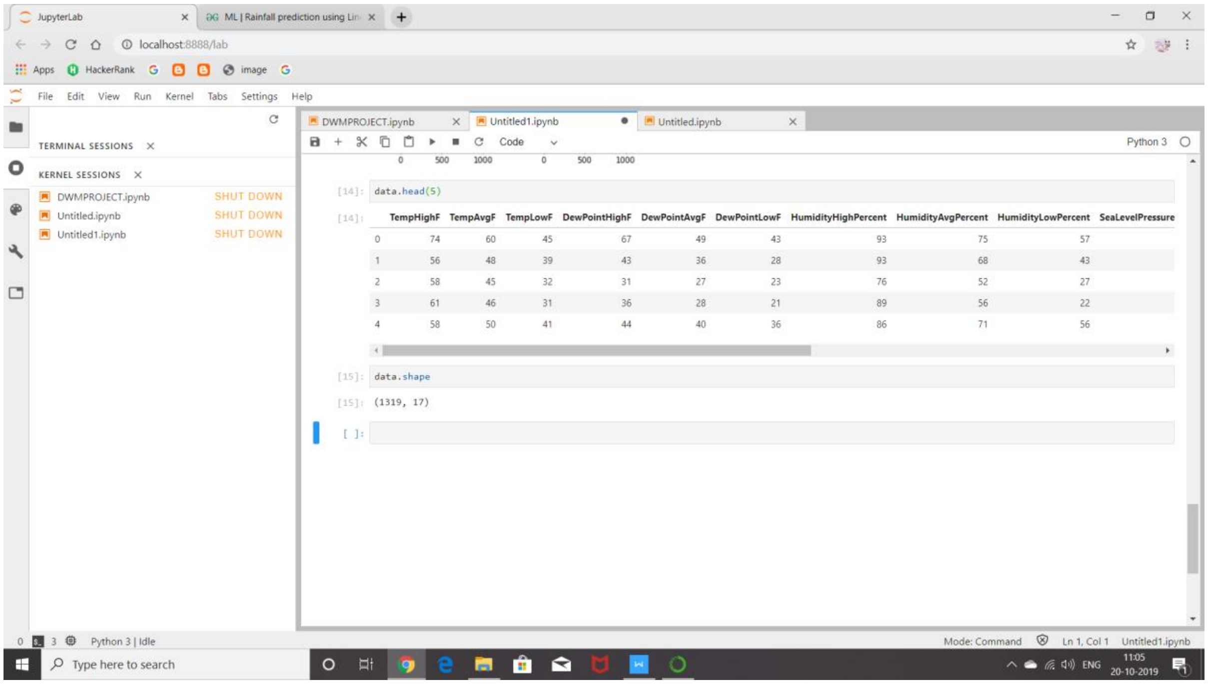


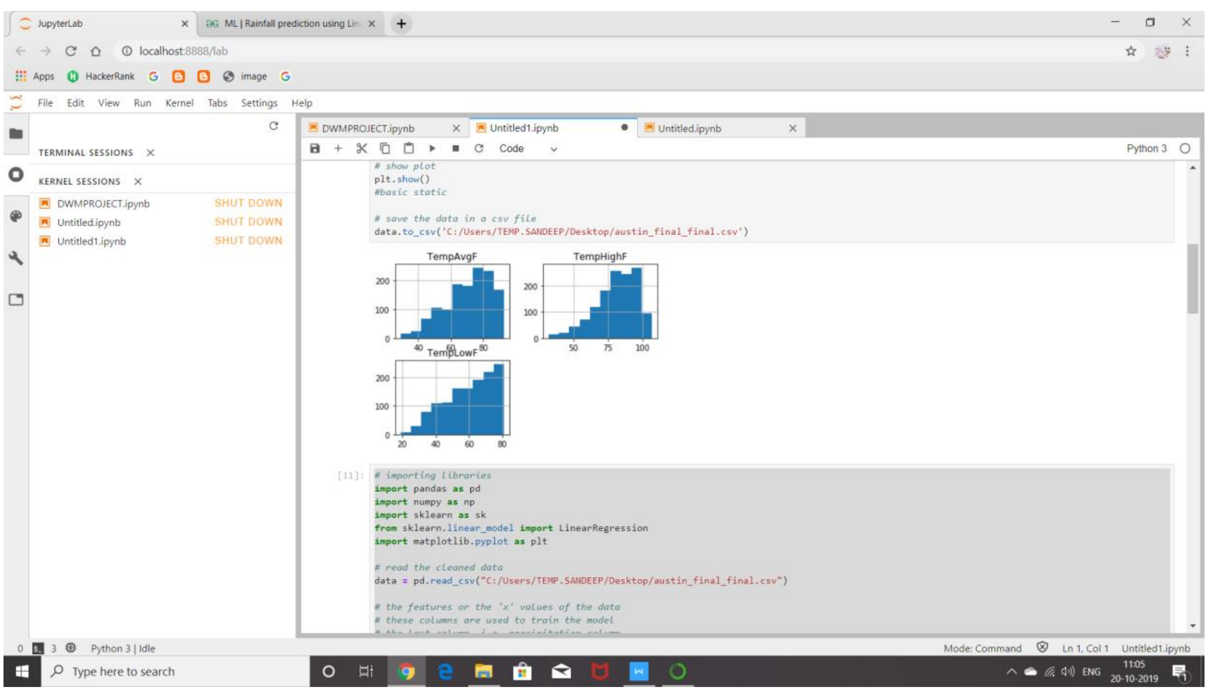
2) The precipitation trend graph:

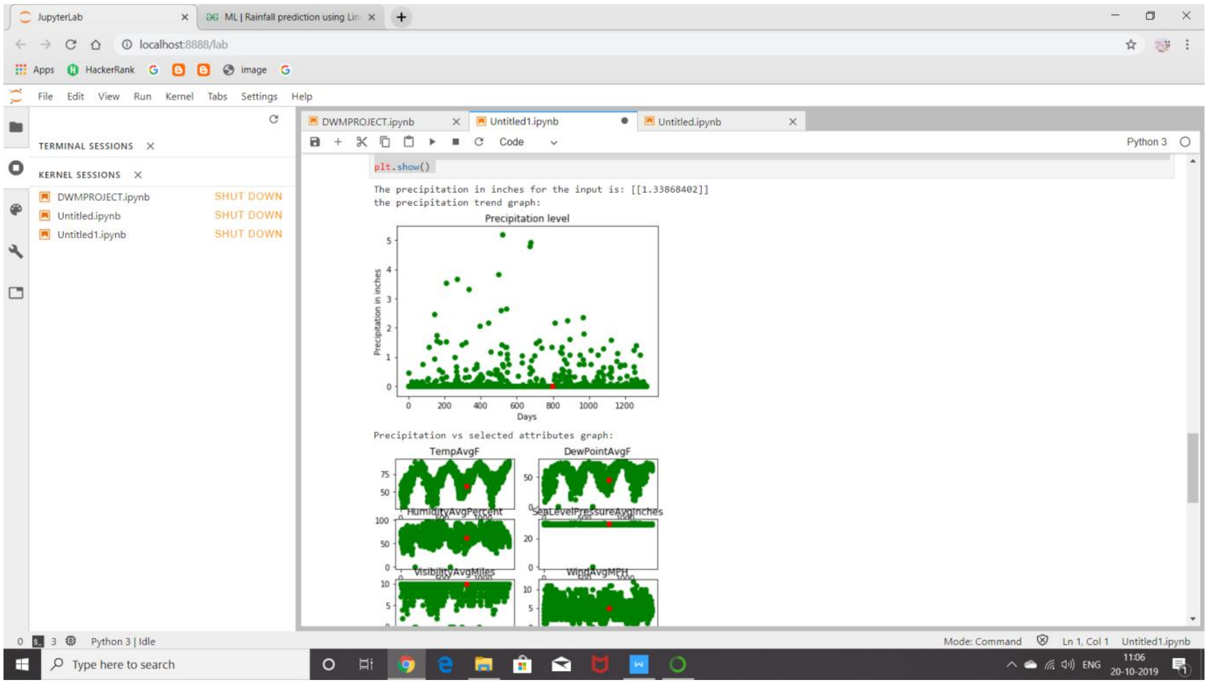


A day (in red) having precipitation of about 2 inches is tracked across multiple parameters (the same day is tracker across multiple features such as temperature, pressure, etc). The x-axis denotes the days and the y-axis denotes the magnitude of the feature such as temperature, pressure, etc. From the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

4. SCREENSHOTS







5.CONCLUSION

We successfully predicted the rainfall 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.

Weblinks:-

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

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