MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE, GWALIOR

(A Govt. Aided UGC Autonomous & NAAC Accredited Institute Affiliated to RGPV, Bhopal)

Skill Based Mini Project Report

on

Rain Prediction Model using ML

Submitted By:

Utkarsh Mishra

0901CS201130

Faculty Mentor:

Dr. Ranjeet Kumar Singh, Assistant Professor, CSE

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE
GWALIOR - 474005 (MP) est. 1957

JAN-JUNE 2022

MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE, GWALIOR

(A Govt. Aided UGC Autonomous & NAAC Accredited Institute Affiliated to RGPV, Bhopal)

CERTIFICATE

This is certified that **Utkarsh Mishra** (0901CS201130) has submitted the project report titled **Rain Prediction Model using ML** under the mentorship of **Dr. Ranjeet Kumar Singh**, in partial fulfilment of the requirement for the award of degree of Bachelor of Technology in Computer Science and Engineering from Madhav Institute of Technology and Science, Gwalior.

Dr. Ranjeet Kumar Singh
Faculty Mentor
Assistant Professor
Computer Science and Engineering

MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE, GWALIOR

(A Govt. Aided UGC Autonomous & NAAC Accredited Institute Affiliated to RGPV, Bhopal)

DECLARATION

I hereby declare that the work being presented in this project report, for the partial fulfilment of requirement for the award of the degree of Bachelor of Technology in Computer Science and Engineering at Madhav Institute of Technology & Science, Gwalior is an authenticated and original record of my work under the mentorship of **Dr. Ranjeet Kumar Singh**, **Assistant Professor**, Computer Science & Engineering.

I declare that I have not submitted the matter embodied in this report for the award of any degree or diploma anywhere else.

Utkarsh Mishra
0901CS201130
II Year,
Computer Science and Engineering

MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE, GWALIOR

(A Govt. Aided UGC Autonomous & NAAC Accredited Institute Affiliated to RGPV, Bhopal)

ACKNOWLEDGEMENT

The full semester project has proved to be pivotal to my career. I am thankful to my institute, **Madhav Institute of Technology and Science** to allow me to continue my disciplinary/interdisciplinary project as a curriculum requirement, under the provisions of the Flexible Curriculum Scheme (based on the AICTE Model Curriculum 2018), approved by the Academic Council of the institute. I extend my gratitude to the Director of the institute, **Dr. R. K. Pandit** and Dean Academics, **Dr. Manjaree Pandit** for this.

I would sincerely like to thank my department, **Department of Computer Science and Engineering, for allowing** me to explore this project. I humbly thank **Dr. Manish Dixit**, Professor and Head, Department of Computer Science and Engineering, for his continued support during the course of this engagement, which eased the process and formalities involved.

I am sincerely thankful to my faculty mentors. I am grateful to the guidance of **Dr. Ranjeet Kumar Singh**, Assistant Professor, Computer Science & Engineering, for his continued support and guidance throughout the project. I am also very thankful to the faculty and staff of the department.

Utkarsh Mishra
0901CS201130
II Year,
Computer Science and Engineering

ABSTRACT

Using the Prediction Model, the objective is to predict the possibility of rain in a particular area. It would help in obtaining the data of future rain patterns and we can prepare for less or excess rainfall. It would help in predicting if there is any possibility of flood or drought in the coming future. More the accuracy of the model, more it would be trustworthy as the result would be more genuine.

In this **Rain Prediction Model**, we are predicting the rainfall different parts of India by the help of the data that is collected by studying the rainfall in those parts over the past years. It would also help in the future if government wants to provide water to areas where there is scarcity of rainfall by building infrastructures. The prediction model can be used in many more possible ways as water is one of the most important necessity for humankind.

TABLE OF CONTENTS

Chapter 1: INTRODUCTION

The motive of the Rain Prediction Model is to predict whether there would be rain in an area at a particular time of the year or not which would be very beneficial as it would help in preparing if there would be shortage of water in a particular area at a particular time of the year. It would also predict if there is any chance of excess of rainfall if there is rainfall in most of the timeframe across the year. The dataset created from the rainfall records of the previous years in the areas across the year is used in this model to predict the future prospects and the pattern of rainfall. With time the model would be able to predict the result with more accuracy as new entries would be edited and the model would grow more and more. The result would not always be guaranteed true but according to the accuracy it would very much be close to the actual result.

The motivation behind the project was to be able to prepare according to the rainfall so that there would be no problem regarding the deficiency or excess of water which could create a hazard and people could know beforehand about the situation of rainfall in their area throughout the year.

The drawbacks of this model could be its inefficiency to predict the correct result every time as a model can never be one hundred percent accurate and there would be always a chance that the predicted result would be different from the actual result and the preparations done according to the predicted result would be a waste.

Chapter 2: TOOLS

2.1 HARDWARE REQUIREMENTS:

- Processor: Minimum 1 GHz; Recommended 2GHz or more.
- Ethernet connection (LAN) OR a wireless adapter (Wi-Fi).
- Hard Drive: Minimum 32 GB; Recommended 64 GB or more.
- Memory (RAM): Minimum 2 GB; Recommended 4 GB or above.

2.2 SOFTWARE REQUIREMENTS:

- Any Web Browser (eg: Chrome).
- Operating system: Windows or MacOs or Linux.
- Language: Python.
- Jupyter Notebook.

Chapter 3: METHODOLOGY

Importing additional library

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score
```

Load the Dataset

```
rf= pd.read_csv("district wise rainfall normal.csv")
rf.describe()
```

	JAN	FEB	MAR	APR	MAY	JUN	JU
count	641.000000	641.000000	641.000000	641.000000	641.000000	641.000000	641.00000
mean	18.355070	20.984399	30.034789	45.543214	81.535101	196.007332	326.03369
std	21.082806	27.729596	45.451082	71.556279	111.960390	196.556284	221.36464
min	0.000000	0.000000	0.000000	0.000000	0.900000	3.800000	11.60000
25%	6.900000	7.000000	7.000000	5.000000	12.100000	68.800000	206.40000
50%	13.300000	12.300000	12.700000	15.100000	33.900000	131.900000	293.70000
75%	19.200000	24.100000	33.200000	48.300000	91.900000	226.600000	374.80000
max	144.500000	229.600000	367.900000	554.400000	733.700000	1476.200000	1820.90000

Checking the null values

```
rf.isnull().sum()
```

```
STATE_UT_NAME    0
DISTRICT         0
JAN              0
FEB              0
```

```
MAR             0
APR             0
MAY             0
JUN             0
JUL             0
AUG             0
SEP             0
OCT             0
NOV             0
DEC             0
ANNUAL          0
Jan-Feb         0
Mar-May         0
Jun-Sep         0
Oct-Dec         0
target          0
dtype: int64
```

Describing the Dataset

```
rf.describe()
```

```
rf.describe()
```

	JAN	FEB	MAR	APR	MAY	JUN	JU
count	641.000000	641.000000	641.000000	641.000000	641.000000	641.000000	641.00000
mean	18.355070	20.984399	30.034789	45.543214	81.535101	196.007332	326.03369
std	21.082806	27.729596	45.451082	71.556279	111.960390	196.556284	221.36464
min	0.000000	0.000000	0.000000	0.000000	0.900000	3.800000	11.60000
25%	6.900000	7.000000	7.000000	5.000000	12.100000	68.800000	206.40000
50%	13.300000	12.300000	12.700000	15.100000	33.900000	131.900000	293.70000
75%	19.200000	24.100000	33.200000	48.300000	91.900000	226.600000	374.80000
max	144.500000	229.600000	367.900000	554.400000	733.700000	1476.200000	1820.90000

Counting the number of True and False values in Traget Attribute.

```
rf['target'].value_counts()
```

```
0    429
1    212
```

```
Name: target, dtype: int64
```

Creating new Dataset to avoid the attributes which have strings.

```
nf = rf.drop('STATE_UT_NAME',axis=1)
nf = nf.drop('DISTRICT',axis=1)
```

0 → NO RAIN

1 → RAIN

Splitting the Features and Target

```
x = nf.drop('target',axis=1)
y = nf['target']
print(x)
```

```
       JAN   FEB    MAR    APR    MAY    JUN     JUL    AUG    SEP    OCT
0    107.3  57.9   65.2  117.0  358.5  295.5   285.0  271.9  354.8  326.0
1     43.7  26.0   18.6   90.5  374.4  457.2   421.3  423.1  455.6  301.2
2     32.7  15.9    8.6   53.4  343.6  503.3   465.4  460.9  454.8  276.1
3     42.2  80.8  176.4  358.5  306.4  447.0   660.1  427.8  313.6  167.1
4     33.3  79.5  105.9  216.5  323.0  738.3   990.9  711.2  568.0  206.9
..     ...   ...    ...    ...    ...    ...     ...    ...    ...    ...
636   13.4  22.1   43.6  150.4  232.6  651.6   788.9  527.3  308.4  343.2
637    2.3   1.0    8.4   46.9  217.6  999.6  1108.5  636.3  263.1  234.9
638   19.8  45.2   73.9  184.9  294.7  556.9   539.9  352.7  266.2  359.4
639    4.8   8.3   17.5   83.3  174.6  698.1  1110.4  592.9  230.7  213.1
640   20.8  14.7   11.8   48.9  171.7  330.2   287.7  217.5  163.1  157.1

       NOV    DEC  ANNUAL  Jan-Feb  Mar-May  Jun-Sep  Oct-Dec
0    315.2  250.9  2805.2    165.2    540.7   1207.2    892.1
1    275.8  128.3  3015.7     69.7    483.5   1757.2    705.3
2    198.6  100.0  2913.3     48.6    405.6   1884.4    574.7
3     34.1   29.8  3043.8    123.0    841.3   1848.5    231.0
4     29.5   31.7  4034.7    112.8    645.4   3008.4    268.1
..     ...    ...     ...      ...      ...      ...      ...
636  172.9   48.1  3302.5     35.5    426.6   2276.2    564.2
637   84.6   18.4  3621.6      3.3    272.9   3007.5    337.9
638  213.5   51.3  2958.4     65.0    553.5   1715.7    624.2
639   93.6   25.8  3253.1     13.1    275.4   2632.1    332.5
640  117.7   58.8  1600.0     35.5    232.4    998.5    333.6

[641 rows x 17 columns]
```

```
print(y)
```

```
0      1
1      1
2      1
3      1
4      1
      ..
636    1
637    1
638    1
639    1
640    1
Name: target, Length: 641, dtype: int64
```

Splitting the data into training data & test data

```
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.2, stratify = y,
random print(x.shape, x_train.shape, x_test.shape)
```

```
(641, 17) (512, 17) (129, 17)
```

Model Training | Logistic Regression

```
model = LogisticRegression();
```

Training the logistic regression model with Training data

```
 model.fit(x_train, y_train)
```

Model Evaluation | Accuracy Score

```
x_train_prediction = model.predict(x_train)
training_data_accuracy = accuracy_score(x_train_prediction, y_train)
print("Accuracy on training data : ", training_data_accuracy)
```

```
Accuracy on training data : 0.85546875
```

```
x_test_prediction = model.predict(x_test)
test_data_accuracy = accuracy_score(x_test_prediction, y_test)
print("Accuracy on test data : ", test_data_accuracy)
```

```
Accuracy on test data : 0.813953488372093
```

Predictive System

```
input_data = (28,15,10,2,3,29,28,21,30,31,30,20,110,16,53,120,89)
```

```
# change the input data to a numpy array
input_data_as_numpy_array = np.asarray(input_data)
```

```
# reshape the numpy array as we are predicting for only one instance
input_data_reshaped = input_data_as_numpy_array.reshape(1,-1)
```

```
prediction = model.predict(input_data_reshaped)
print(prediction)
   if(prediction[0] == 0):
  print("It won't rain !!")
else:
  print("It will rain !!")
```

```
[0]
It won't rain !!
```

Chapter 4: DISTRICT WISE RAINFALL DATASET

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	STATE_UT	DISTRICT	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL	Jan-Feb	Mar-May	Jun-Sep	Oct-Dec	target	
2	ANDAMAI	NICOBAR	107.3	57.9	65.2	117	358.5	295.5	285	271.9	354.8	326	315.2	250.9	2805.2	165.2	540.7	1207.2	892.1	1	
3	ANDAMAI	SOUTH AN	43.7	26	18.6	90.5	374.4	457.2	421.3	423.1	455.6	301.2	275.8	128.3	3015.7	69.7	483.5	1757.2	705.3	1	
4	ANDAMAI	N & M AN	32.7	15.9	8.6	53.4	343.6	503.3	465.4	460.9	454.8	276.1	198.6	100	2913.3	48.6	405.6	1884.4	574.7	1	
5	ARUNACH	LOHIT	42.2	80.8	176.4	358.5	306.4	447	660.1	427.8	313.6	167.1	34.1	29.8	3043.8	123	841.3	1848.5	231	1	
6	ARUNACH	EAST SIAN	33.3	79.5	105.9	216.5	323	738.3	990.9	711.2	568	206.9	29.5	31.7	4034.7	112.8	645.4	3008.4	268.1	1	
7	ARUNACH	SUBANSIR	28	48.3	85.3	101.5	140.5	228.4	217.4	182.8	159.8	75.9	20.9	11.6	1300.4	76.3	327.3	788.4	108.4	0	
8	ARUNACH	TIRAP	42.2	72.7	141	316.9	328.7	614.7	851.9	500.6	418.3	218.7	42.9	22.9	3571.5	114.9	786.6	2385.5	284.5	1	
9	ARUNACH	ANJAW (L	42.2	80.8	176.4	358.5	306.4	447	660.1	427.8	313.6	167.1	34.1	29.8	3043.8	123	841.3	1848.5	231	1	
10	ARUNACH	LOWER DI	83.7	153.9	303.5	383.6	268	374.2	272	160.5	266.7	167.2	64	56	2553.3	237.6	955.1	1073.4	287.2	1	
11	ARUNACH	CHANGLA	70.3	170.9	367.9	554.4	334.2	526.2	460.8	291.5	353.6	275	64.9	74.2	3543.9	241.2	1256.5	1632.1	414.1	1	
12	ARUNACH	PAPUM PA	33.5	67.8	106.1	226.9	453	640.5	609.5	503.4	492.3	214.7	19.2	11.3	3378.2	101.3	786	2245.7	245.2	1	
13	ARUNACH	LOW SUBA	97.5	109.3	92.4	204.3	266.2	284.1	248.9	270.5	192.7	78.5	49.5	27.2	1921.1	206.8	562.9	996.2	155.2	1	
14	ARUNACH	UPPER SIA	74.3	176.7	362.6	397.5	408.7	801.9	653	417.9	686	264.9	86.9	71.7	4402.1	251	1168.8	2558.8	423.5	1	
15	ARUNACH	WEST SIAI	26	66.7	76.8	229.2	239.5	416.6	592.4	312.4	291.1	126.8	33.7	29.5	2440.7	92.7	545.5	1612.5	190	1	
16	ARUNACH	DIBANG V	83.7	153.9	303.5	383.6	268	374.2	272	160.5	266.7	167.2	64	56	2553.3	237.6	955.1	1073.4	287.2	1	
17	ARUNACH	WEST KAI	35.2	43.5	58.9	134.3	341.1	665.3	749.9	579.1	490.9	233.9	40.3	27	3399.4	78.7	534.3	2485.2	301.2	1	
18	ARUNACH	EAST KAM	49	74.4	96.5	156.9	208	345.7	368.5	256.2	275.9	138.2	34.4	27.2	2030.9	123.4	461.4	1246.3	199.8	1	
19	ARUNACH	TAWANG(	35.2	43.5	58.9	134.3	341.1	665.3	749.9	579.1	490.9	233.9	40.3	27	3399.4	78.7	534.3	2485.2	301.2	1	
20	ARUNACH	KURUNG K	82.7	70	128.2	245.7	271.4	292.7	404	276.3	283.5	92.3	32.3	42.4	2221.5	152.7	645.3	1256.5	167	1	
21	ASSAM	CACHAR	13.3	50.2	168.3	262.5	386.4	532.1	526.2	470.8	360.8	182.4	34.8	11.4	2999.2	63.5	817.2	1889.9	228.6	1	
22	ASSAM	DARRANG	13.1	21.4	53.5	168.8	320	419.7	345.8	272.1	221.5	95.4	17.2	9.3	1957.8	34.5	542.3	1259.1	121.9	1	
23	ASSAM	GOALPARA	12.7	20.4	51.1	196.6	399.8	567.8	502.8	334.6	304.9	157.7	21.7	5.2	2575.3	33.1	647.5	1710.1	184.6	1	
24	ASSAM	KAMRUP	12	20.8	58.6	151.7	293.4	365.5	345.1	248.7	188.4	106.6	15.1	7.5	1813.4	32.8	503.7	1147.7	129.2	1	
25	ASSAM	LAKHIMPU	27.7	48.6	76.7	165.5	331.9	528.3	605.2	467.6	424.1	140.3	23	20.4	2859.3	76.3	574.1	2025.2	183.7	1	
26	ASSAM	NORTH CA	16.7	47.5	158.9	207.9	308	328.1	270.3	201.3	189.1	196.4	42.1	11.2	1977.5	64.2	674.8	988.8	249.7	1	
27	ASSAM	NAGAON	12	22.5	48.1	128.9	171.3	285.9	326.3	294.1	218.6	120	21.6	10.8	1660.1	34.5	348.3	1124.9	152.4	1	

Chapter 5: RESULT

By the help of the prediction model now we are able to predict whether there would be rainfall in a particular area across the year or not which would be beneficial to prepare for any situation. We used the dataset that is created by the help of the rainfall stats from the previous years. We used Machine Learning libraries and Python to achieve the result. We divided the dataset into training and testing sets. The training set is used to train the prediction model and the testing set was used afterwards to check the accuracy of the model and then modifications were done to increase the accuracy of the prediction system. We used logistic regression to predict the rainfall as it satisfies the dataset and the distribution of the data.

Predictive System

```
input_data = (28,15,10,2,3,29,28,21,30,31,30,20,110,16,53,120,89)
```

```
# change the input data to a numpy array
input_data_as_numpy_array = np.asarray(input_data)
```

```
# reshape the numpy array as we are predicting for only one instance
input_data_reshaped = input_data_as_numpy_array.reshape(1,-1)
```

```
prediction = model.predict(input_data_reshaped)
print(prediction)
   if(prediction[0] == 0):
  print("It won't rain !!")
else:
  print("It will rain !!")
```

```
[0]
It won't rain !!
```

Out of the liner and logistic regression, we used logistic as it was best suited according to the dataset and it was doing the work that was needed to be done. The model was trained and tested and then accuracy of the model was checked which came out to be approximate eighty five percent which is decent if not brilliant. It has the potential to predict the rainfall with a decent accuracy with eighty five percent accuracy.

Chapter 6: CONCLUSION

The conclusion that we got after working on the prediction model was that Python and Machine Learning are very important and beneficial in today's world as it is being used in most of the technologies that are around us. They played an important role in this prediction model as well to calculate the result that whether there would be rainfall at a particular time of the year or not. We used various python libraries in this project to achieve our goal and it would be very difficult or may be impossible to develop this project in any other programming language.

At last we conclude that python programming is an important aspect of the programming world as industries like Machine Learning, Artificial Intelligence, Natural Language Processing are fully based on it and would be inefficient without the python language. These industries plays an important role in our life and in the project as well it played an important role to enable it towards its execution.

REFERENCES

1. https://www.python.org/
2. https://en.wikipedia.org/wiki/Python_(programming_language)
3. https://www.programiz.com/python-programming/online-compiler/
4. https://www.tutorialspoint.com/python/index.htm
5. https://www.w3schools.in/python/gui-programming