INTERNET OF THINGS

ECE – 3502

PROJECT REPORT

Weather Forecasting using Machine Learning

Under the Guidance of

Prof. Manimaran P

School of Electronics and Communication Engineering

Vellore Institute of Technology (VIT)

Vellore, TamilNadu - 632014

Team Members

Ayushman Mookherjee – 18BEC0888

Vaibhav Jain – 18BEC0898

Sabthesh JK Sriniwas – 18BEC0877

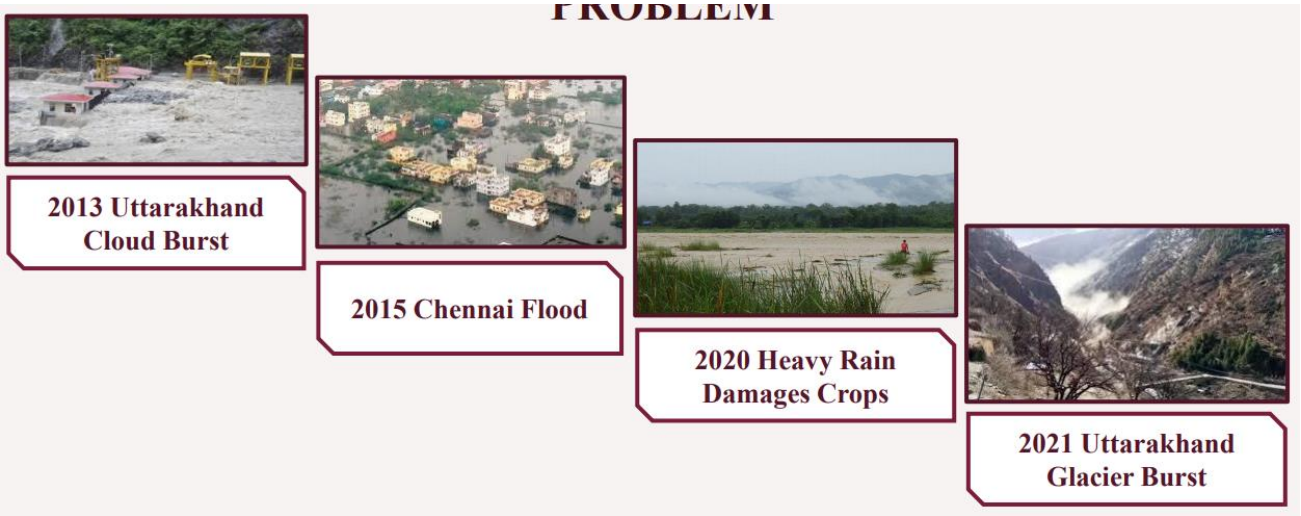
Muskan Sarkar – 18BEC0892

Table of Content

Abstract-

The activities of many primary sectors depend on the temperature for production, e.g. farming. The climate is changing at a drastic rate nowadays, which makes the old temperature prediction methods less effective and more hectic. To overcome these difficulties, the improved and reliable weather prediction methods are required. These predictions affect a nation's economy and the lives of people. To develop a temperature prediction system that can be used in remote areas is the main motivation of this work. A low-cost and portable solution for temperature prediction is devised.

Problem-



Reasons-

➢ India Meteorological Department (IMD) has forecast a below normal rainfall during the summer monsoon. In a country where 60 per cent of farms are rain-fed, weak rains could lower agricultural output, resulting in a spike in inflation. This poses a major challenge for the country's newly elected government of Narendra Modi.

➢ Its four key Union ministries—agriculture, food, fertilizers and water resources—and state governments are frantically preparing to deal with any possible failure of monsoon.

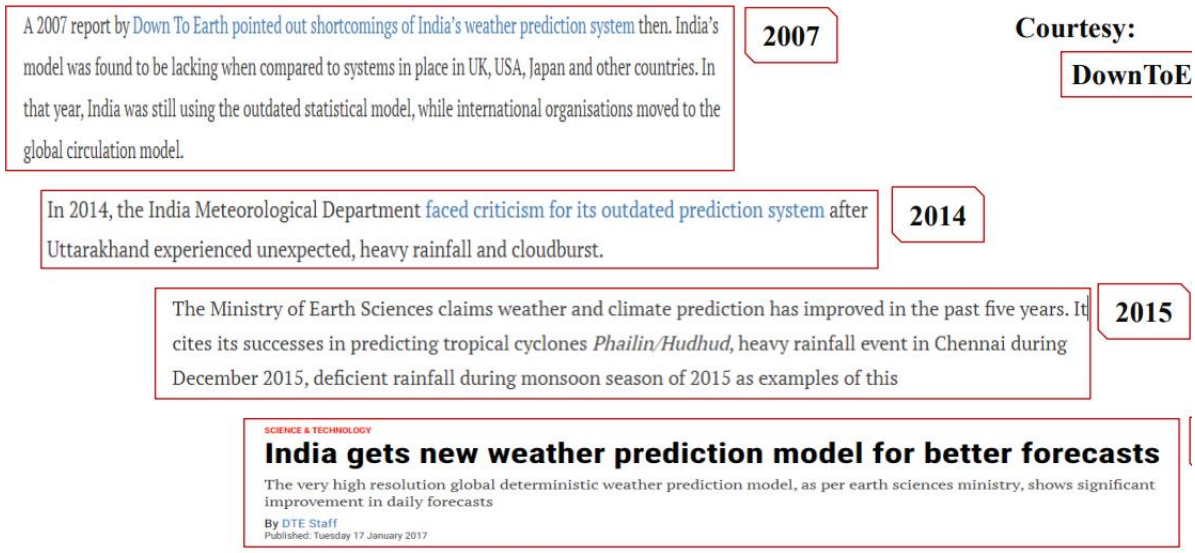
➢ Contingency plans have been chalked out in 500 districts, where authorities are preparing nurseries for short-duration crops that can withstand low rainfall, issuing location-specific advisories and monitoring the situation closely

➢ According to a study published in Nature Communications in October 2013, El Niño affects 22 per cent to 24 per cent of harvested areas worldwide. But the prevailing fears mostly stem from the lack of understanding of the weather phenomenon.

➢ The problem is scientists do not really understand El Niño and cannot predict when the weather phenomenon, which lasts from eight months to over

a year, will strike and with what vigor. Such uncertainty leaves both farmers and governments undecided about their future course of action.

➢ For instance, the El Niño of 1997-98 has been the strongest since record keeping of the phenomenon began in 1950. Indian scientists and the authorities were dreading a bad monsoon that year. Though the eastern Pacific Ocean remained warm till February 1998, it did not affect Indian monsoon.



Forecast Models-

➢ There are two main types of forecast models: global ones, covering the entire planet, and local ones, covering specific areas, such as continents, countries, mountain ranges and so on.

➢ Both global and local models also vary in their resolutions, which is the distance between two grid points.

➢ Bigger resolutions of 50 to 10 km in size are usually deployed in relatively flat terrains, while mountain ranges require the nodes to be a lot closer to each other, usually 5, 2 or 1 km.

Global Models

- **ECMWF** (European Centre for Medium-Range Weather Forecasts)
- **GFS** (The Global Forecast System)
- **ICON** (Global German Standard)
- **ICON** (Global German Standard)
- **UM** (United Kingdom Met Office)
- **CFS** (Climate Forecast System)

Regional Models

- **ECMWF**
- **WRF** (Weather Research and Forecasting)
- **ALADIN** (Aire Limitée Adaptation dynamique Développement Inter National)
- **NAM** (North American Mesoscale)
- **HRRR** (High Resolution Rapid Refresh)
- **GEM** (The Global Environmental Multiscale Model)

Algorithm-

➢ The data which is received from the satellite is in RAW format which doesn't provide any kind of information.

➢ The process of converting RAW data into the cleaned data is known as Knowledge Discovery Process (Data Mining). Various method of data mining is used,

• Rule-based Methods

• Neural Networks • Naive Bayes

• Bayesian Belief Network

• Support Vector Machine

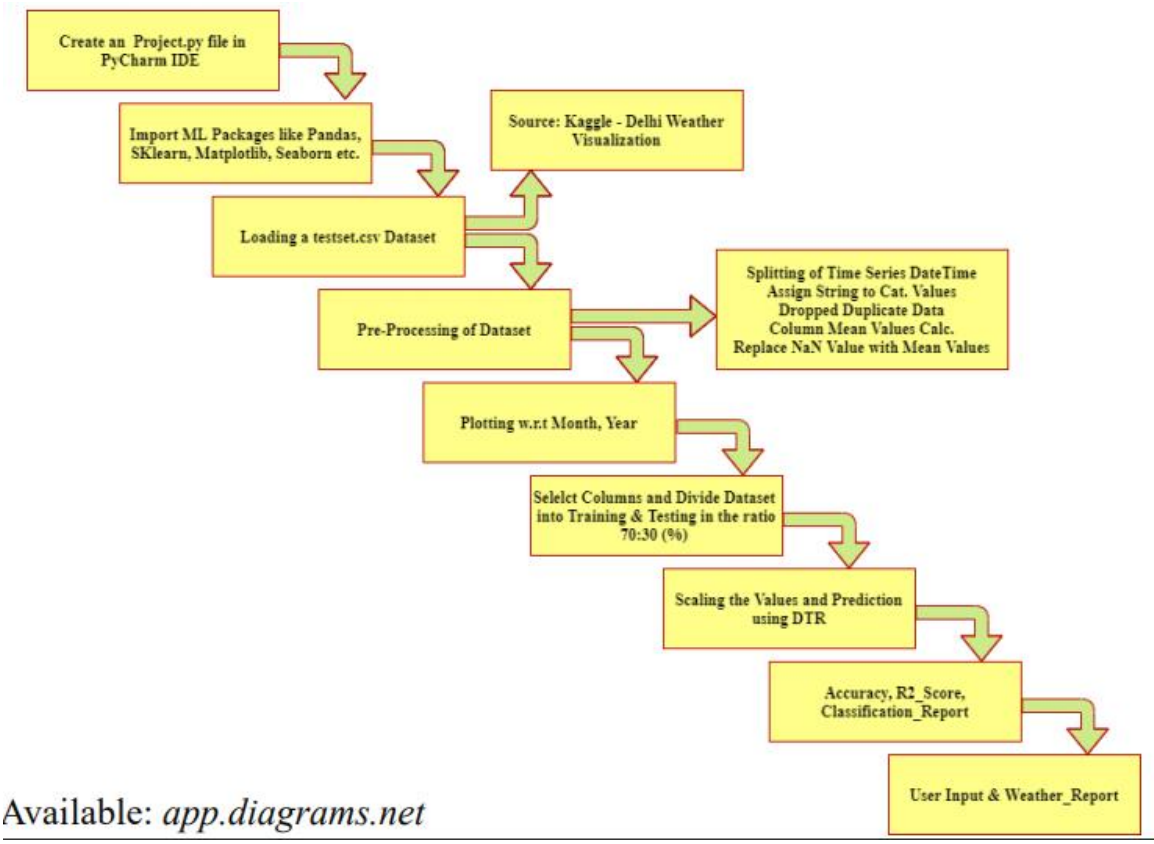
Proposed Tools-

➢ Source: Kaggle datasets

➢ Packages: NumPy, Pandas, Plotly, Scikit, Matplotlib

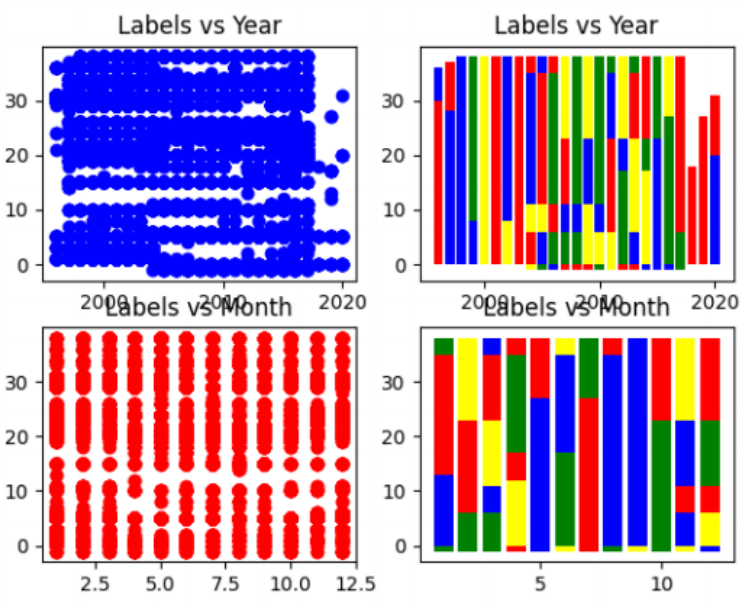
➢ IDE: Visual Studio / Google Colab / PyCharm

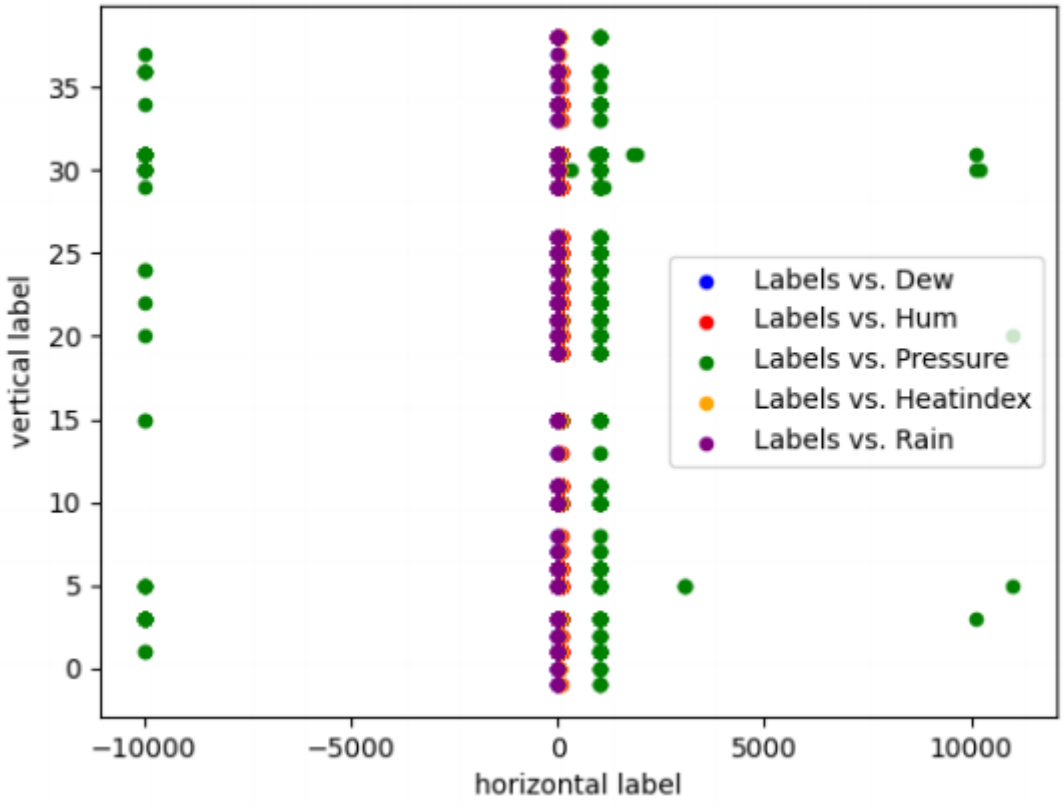
Block Diagram-

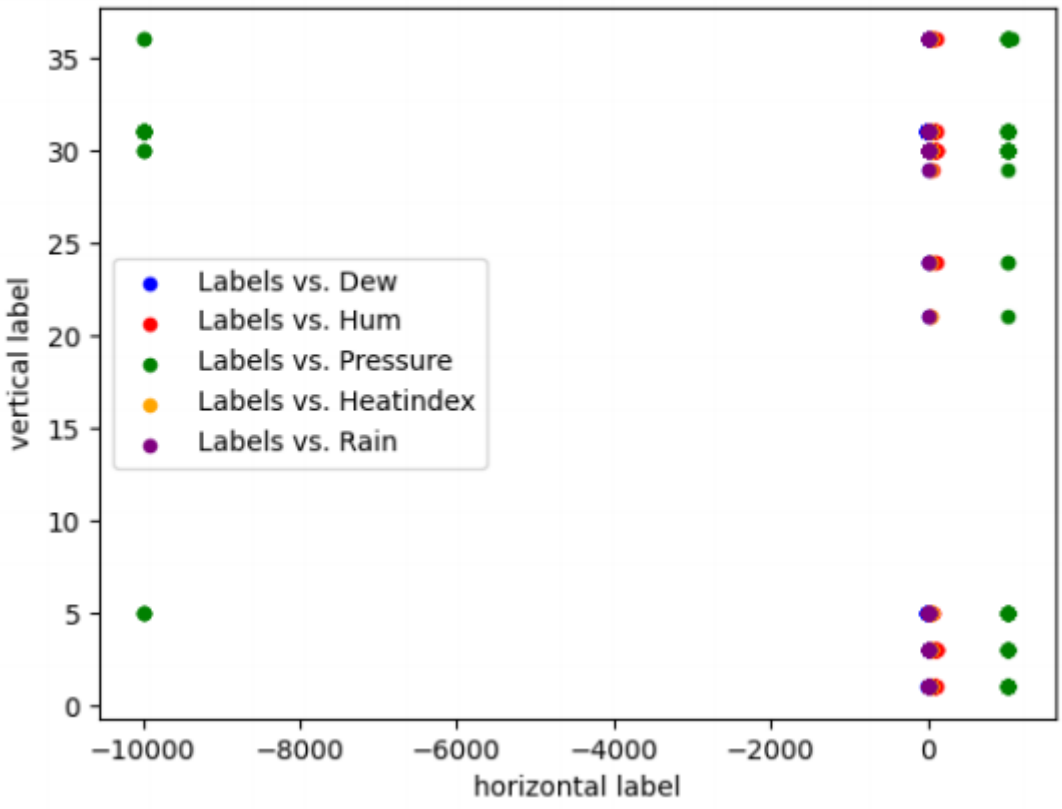


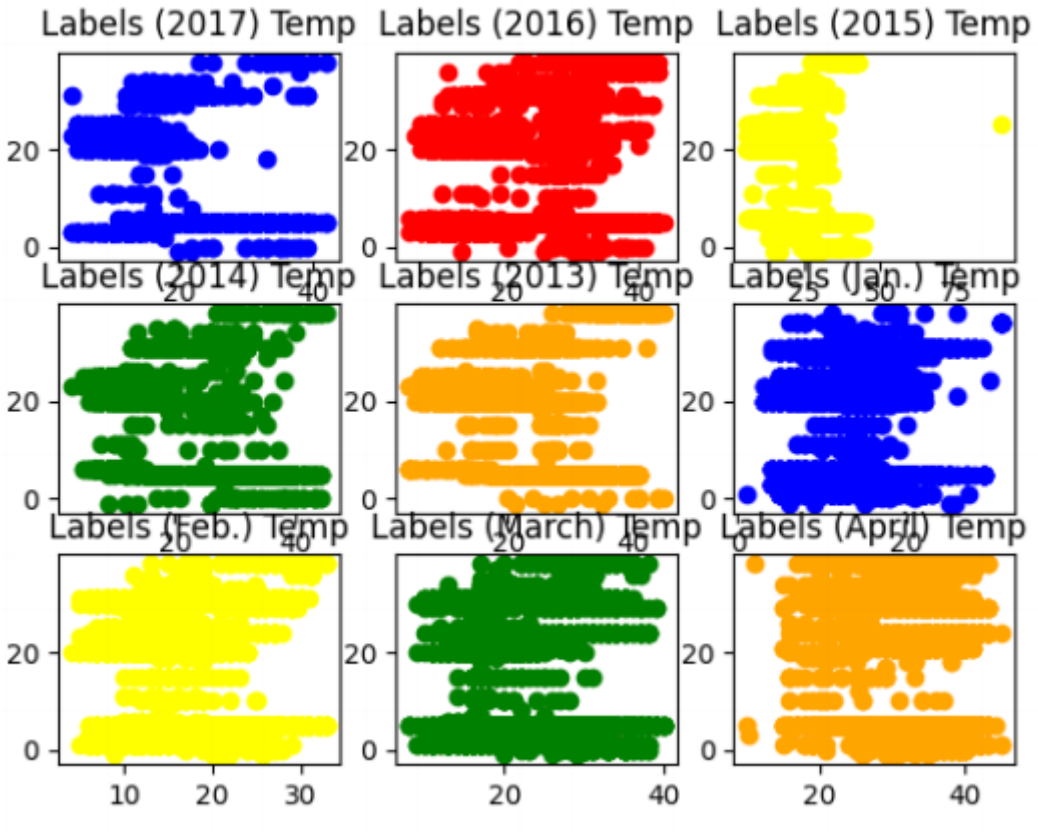
Available: *app.diagrams.net*

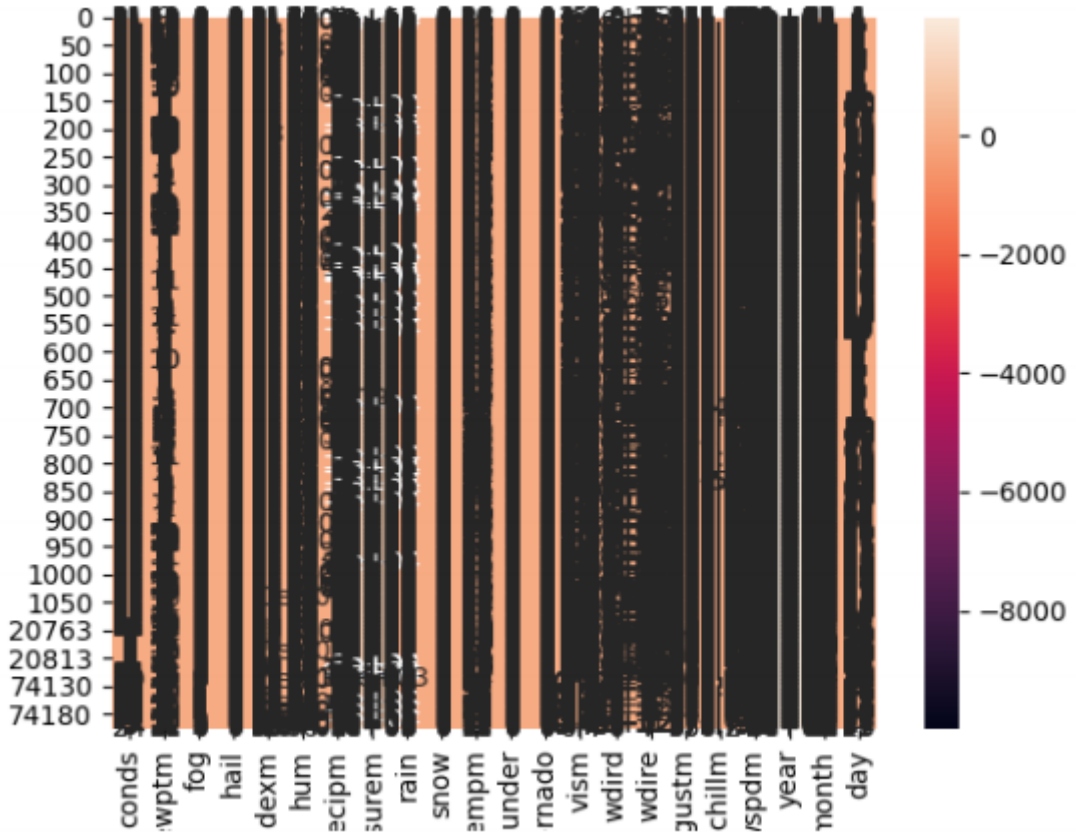
Outputs and Data Visualizations:

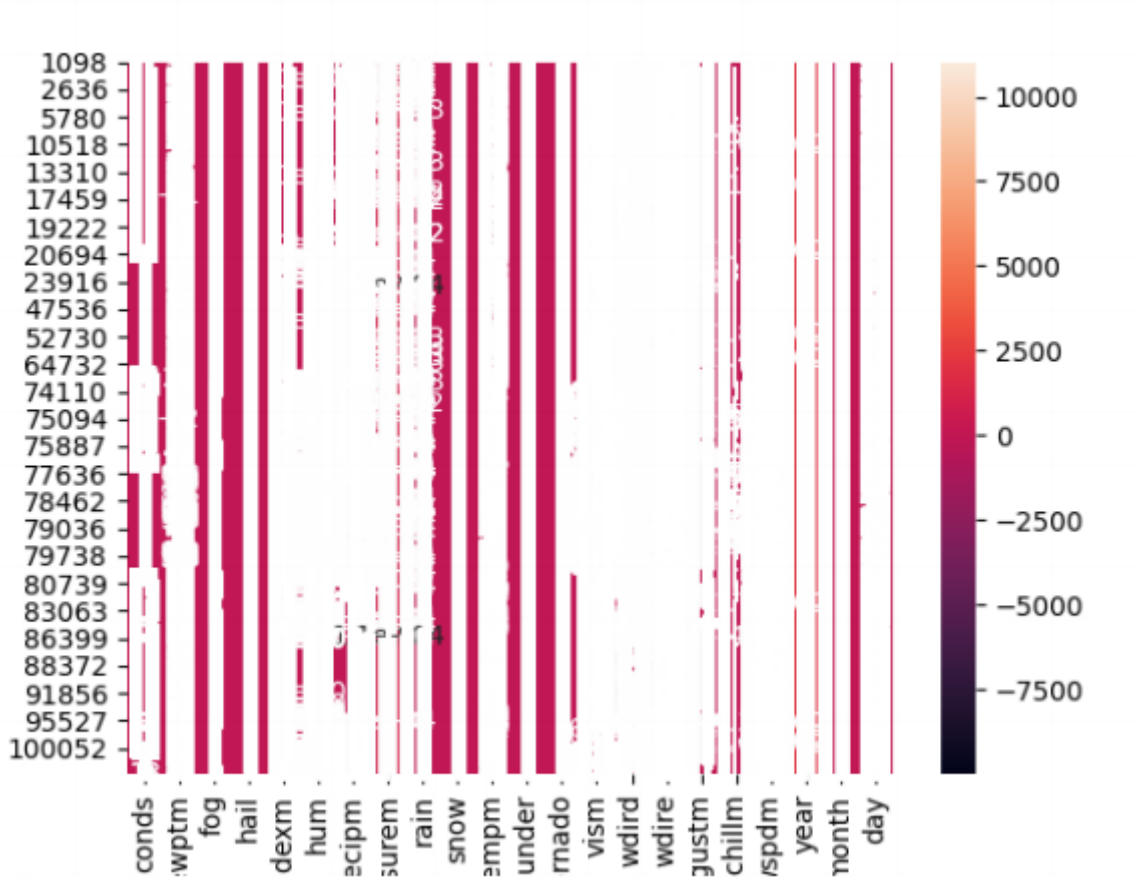



Labels vs. Dew
Labels vs. Hum
Labels vs. Pressure
Labels vs. Heatindex
Labels vs. Rain
vertical label
horizontal label
35
30
25
20
15
10
5
0
−10000
−5000
0
5000
10000


Labels vs. Dew
Labels vs. Hum
Labels vs. Pressure
Labels vs. Heatindex
Labels vs. Rain
vertical label
horizontal label
35
30
25
20
15
10
5
0
−10000
−8000
−6000
−4000
−2000
0


Labels (2017) Temp
Labels (2016) Temp
Labels (2015) Temp
Labels (2014) Temp
Labels (2013) Temp
Labels (Jan.) Temp
Labels (Feb.) Temp
Labels (March) Temp
Labels (April) Temp


0
50
100
150
200
250
300
350
400
450
500
550
600
650
700
750
800
850
900
950
1000
1050
20763
20813
74130
74180
conds
wptm
fog
hail
dexm
hum
ecipm
surem
rain
snow
empm
under
rnado
vism
wdird
wdire
gustm
chillm
spdm
year
month
day
0
−2000
−4000
−6000
−8000



Before Pre Processing

Starting of Programme

First Ten Data

	datetime_utc	conds	dewptm	fog	hail	heatindexm	hum	precipm	pressurem	...	tempm	thunder	tornado	vism	wdird	wdire	wgustm	windchillm	wspdm
0	19961101-11:00	Smoke	9.0	0	0	NaN	27.0	0.0	1010.0	...	30.0	0	0	5.0	280.0	West	NaN	NaN	7.4
1	19961101-12:00	Smoke	10.0	0	0	NaN	32.0	0.0	-9999.0	...	28.0	0	0	NaN	0.0	North	NaN	NaN	NaN
2	19961101-13:00	Smoke	11.0	0	0	NaN	44.0	5.0	-9999.0	...	24.0	0	0	NaN	0.0	North	NaN	NaN	NaN
3	19961101-14:00	Smoke	10.0	0	0	NaN	41.0	0.0	1010.0	...	24.0	0	0	2.0	0.0	North	NaN	NaN	NaN
4	19961101-16:00	Smoke	11.0	0	0	NaN	47.0	0.0	1011.0	...	23.0	0	0	1.2	0.0	North	NaN	NaN	0.0
5	19961101-17:00	Smoke	12.0	0	0	NaN	56.0	0.0	1011.0	...	21.0	0	0	NaN	0.0	North	NaN	NaN	NaN
6	19961101-18:00	Smoke	13.0	0	0	NaN	60.0	0.3	1010.0	...	21.0	0	0	0.8	0.0	North	NaN	NaN	0.0
7	19961101-19:00	Smoke	13.0	0	0	NaN	60.0	0.8	-9999.0	...	21.0	0	0	NaN	0.0	North	NaN	NaN	NaN
8	19961101-20:00	Smoke	13.0	0	0	NaN	68.0	2.0	-9999.0	...	19.0	0	0	NaN	0.0	North	NaN	NaN	NaN
9	19961101-21:00	Smoke	13.0	0	0	NaN	68.0	0.3	1010.0	...	19.0	0	0	NaN	0.0	North	NaN	NaN	NaN

[10 rows x 20 columns]

Last Ten Data

	datetime_utc	conds	dewptm	fog	hail	heatindexm	hum	precipm	pressurem	...	tempm	thunder	tornado	vism	wdird	wdire	wgustm	windchillm	wspdm
101024	20200318-18:00	Haze	14.0	0	0	NaN	65.0	NaN	1010.0	...	19.0	0	0	2.0	NaN	NaN	NaN	NaN	0.0
101025	20200318-21:00	Mist	14.0	0	0	NaN	78.0	NaN	1010.0	...	17.0	0	0	2.0	NaN	NaN	NaN	NaN	0.0
101026	20200319-00:00	Mist	14.0	0	0	NaN	78.0	NaN	1010.0	...	17.0	0	0	2.0	NaN	NaN	NaN	NaN	0.0
101027	20200319-03:00	Haze	14.0	0	0	NaN	66.0	NaN	1012.0	...	20.0	0	0	2.0	230.0	SW	NaN	NaN	3.7
101028	20200319-03:30	Haze	14.0	0	0	NaN	64.0	NaN	1013.0	...	21.0	0	0	3.0	250.0	WSW	NaN	NaN	7.4
101029	20200319-04:30	Haze	14.0	0	0	NaN	53.0	NaN	1013.0	...	24.0	0	0	3.0	250.0	WSW	NaN	NaN	14.8
101030	20200319-05:00	Haze	14.0	0	0	NaN	50.0	NaN	1013.0	...	25.0	0	0	3.0	250.0	WSW	NaN	NaN	14.8
101031	20200319-05:30	Haze	14.0	0	0	NaN	50.0	NaN	1013.0	...	25.0	0	0	3.0	250.0	WSW	NaN	NaN	14.8
101032	20200319-06:00	Haze	15.0	0	0	NaN	39.0	NaN	1012.0	...	27.0	0	0	4.0	290.0	WNW	NaN	NaN	18.5
101033	20200319-06:30	Haze	15.0	0	0	27.3	48.0	NaN	1013.0	...	27.0	0	0	4.0	320.0	NW	NaN	NaN	14.8

[10 rows x 20 columns]

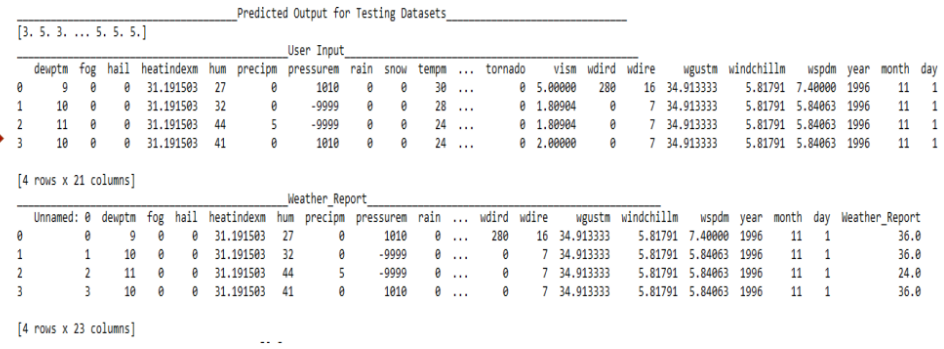
After Pre Processing

Good Dataset

	datetime_utc	conds	dewptm	fog	hail	heatindexm	hum	precipm	pressurem	...	wdird	wdire	wgustm	windchillm	wspdm	year	month	day	time
0	1996-11-01 11:00:00	31	9.0	0	0	31.191503	27.0	0.0	1010.0	...	280.0	16	34.913333	5.81791	7.400000	1996	11	1	11:00:00
1	1996-11-01 12:00:00	31	10.0	0	0	31.191503	32.0	0.0	-9999.0	...	0.0	7	34.913333	5.81791	5.840895	1996	11	1	12:00:00
2	1996-11-01 13:00:00	31	11.0	0	0	31.191503	44.0	5.0	-9999.0	...	0.0	7	34.913333	5.81791	5.840895	1996	11	1	13:00:00
3	1996-11-01 14:00:00	31	10.0	0	0	31.191503	41.0	0.0	1010.0	...	0.0	7	34.913333	5.81791	5.840895	1996	11	1	14:00:00
4	1996-11-01 16:00:00	31	11.0	0	0	31.191503	47.0	0.0	1011.0	...	0.0	7	34.913333	5.81791	0.000000	1996	11	1	16:00:00
5	1996-11-01 17:00:00	31	12.0	0	0	31.191503	56.0	0.0	1011.0	...	0.0	7	34.913333	5.81791	5.840895	1996	11	1	17:00:00
6	1996-11-01 18:00:00	31	13.0	0	0	31.191503	60.0	0.3	1010.0	...	0.0	7	34.913333	5.81791	0.000000	1996	11	1	18:00:00
7	1996-11-01 19:00:00	31	13.0	0	0	31.191503	60.0	0.8	-9999.0	...	0.0	7	34.913333	5.81791	5.840895	1996	11	1	19:00:00
8	1996-11-01 20:00:00	31	13.0	0	0	31.191503	68.0	2.0	-9999.0	...	0.0	7	34.913333	5.81791	5.840895	1996	11	1	20:00:00
9	1996-11-01 21:00:00	31	13.0	0	0	31.191503	68.0	0.3	1010.0	...	0.0	7	34.913333	5.81791	5.840895	1996	11	1	21:00:00
10	1996-11-01 22:00:00	31	13.0	0	0	31.191503	68.0	0.0	1009.0	...	0.0	7	34.913333	5.81791	5.840895	1996	11	1	22:00:00
11	1996-11-01 23:00:00	31	12.0	0	0	31.191503	64.0	0.0	1009.0	...	0.0	7	34.913333	5.81791	5.840895	1996	11	1	23:00:00
12	1996-11-02 00:00:00	31	11.0	0	0	31.191503	60.0	0.0	1010.0	...	0.0	7	34.913333	5.81791	5.840895	1996	11	2	00:00:00
13	1996-11-02 01:00:00	31	11.0	0	0	31.191503	60.0	3.0	1010.0	...	0.0	7	34.913333	5.81791	5.840895	1996	11	2	01:00:00
14	1996-11-02 02:00:00	31	10.0	0	0	31.191503	52.0	0.0	1011.0	...	200.0	10	34.913333	5.81791	9.300000	1996	11	2	02:00:00

[15 rows x 24 columns]

Weather_Report



Predicted Output for Testing Datasets

```
[3. 5. 3. ... 5. 5. 5.]
```

User Input

	dewptm	fog	hail	heatindexm	hum	precipm	pressurem	rain	snow	tempm	...	tornado	vism	wdird	wdire	wgustm	windchillm	wspdm	year	month	day
0	9	0	0	31.191503	27	0	1010	0	0	30	...	0	5.00000	280	16	34.913333	5.81791	7.40000	1996	11	1
1	10	0	0	31.191503	32	0	-9999	0	0	28	...	0	1.80904	0	7	34.913333	5.81791	5.84063	1996	11	1
2	11	0	0	31.191503	44	5	-9999	0	0	24	...	0	1.80904	0	7	34.913333	5.81791	5.84063	1996	11	1
3	10	0	0	31.191503	41	0	1010	0	0	24	...	0	2.00000	0	7	34.913333	5.81791	5.84063	1996	11	1

[4 rows x 21 columns]

Weather_Report

	Unnamed: 0	dewptm	fog	hail	heatindexm	hum	precipm	pressurem	rain	...	wdird	wdire	wgustm	windchillm	wspdm	year	month	day	Weather_Report
0	0	9	0	0	31.191503	27	0	1010	0	...	280	16	34.913333	5.81791	7.40000	1996	11	1	36.0
1	1	10	0	0	31.191503	32	0	-9999	0	...	0	7	34.913333	5.81791	5.84063	1996	11	1	36.0
2	2	11	0	0	31.191503	44	5	-9999	0	...	0	7	34.913333	5.81791	5.84063	1996	11	1	24.0
3	3	10	0	0	31.191503	41	0	1010	0	...	0	7	34.913333	5.81791	5.84063	1996	11	1	36.0

[4 rows x 23 columns]

Overall Report

R2_Score

```
0.9936791773747968
```

OVERALL CLASSIFICATION REPORT

	precision	recall	f1-score	support
-1.0	0.63	0.68	0.65	25
0.0	0.88	0.93	0.91	121
1.0	0.99	1.00	1.00	875
2.0	1.00	1.00	1.00	30
3.0	1.00	1.00	1.00	835
4.0	0.00	0.00	0.00	0
5.0	1.00	1.00	1.00	14297
6.0	0.75	0.77	0.76	122
7.0	0.86	1.00	0.92	6
8.0	0.67	0.80	0.73	5
9.0	0.00	0.00	0.00	1
10.0	0.99	1.00	1.00	133
11.0	0.70	0.47	0.56	15
12.0	0.00	0.00	0.00	1
13.0	0.00	0.00	0.00	1
15.0	0.99	0.99	0.99	393
16.0	1.00	0.67	0.80	3
17.0	0.00	0.00	0.00	7
18.0	0.94	1.00	0.97	15
19.0	0.98	0.95	0.97	65
20.0	1.00	1.00	1.00	2827
21.0	1.00	1.00	1.00	476
22.0	0.98	0.98	0.98	92
23.0	0.89	0.90	0.90	311
24.0	1.00	1.00	1.00	612
25.0	1.00	1.00	1.00	288
26.0	0.99	0.99	0.99	110
27.0	1.00	1.00	1.00	1
28.0	0.00	0.00	0.00	0
29.0	1.00	1.00	1.00	696
30.0	1.00	0.99	1.00	577
31.0	1.00	1.00	1.00	6231
32.0	1.00	1.00	1.00	1
33.0	1.00	1.00	1.00	64
34.0	1.00	1.00	1.00	138
35.0	1.00	1.00	1.00	5
36.0	1.00	0.98	0.99	121
37.0	0.00	0.00	0.00	2
38.0	1.00	1.00	1.00	822
accuracy			1.00	30324
macro avg	0.78	0.77	0.77	30324
weighted avg	1.00	1.00	1.00	30324

ACCURACY

```
0.9959108297058435
```

Results and References-

Successfully we have achieved our target so that it can predict the Weather_Report of an given Area with an Accuracy of 99.5%. But Area can be extend if we better technique to access that data. Low Cost and Portable Solution for this Problem.

1. DownToEarth [Online]. Available: *India gets new weather prediction model for better forecasts (downtoearth.org.in)*
2. Woma. T. Y & Iseh. A. J, "Weather Forecasting Models, Methods and Applications" in *International Journal of Engineering Research & Technology (IJERT),* 2013, pp. 1945-1948.
3. Wai Yan Nyein & Zaw Zaw Htike, "Forecasting of Monthly Temperature Variations using Random Forests " in *ARPN Journal of Engineering and Applied Sciences,* 2015, pp. 10109-10112.
4. Nitin Singh, Saurabh Chaturvedi & Shamim Akhter, "Weather Forecasting Using Machine Learning Algorithm" in *IEEE 2019 International Conference on Signal Processing and Communication (ICSC),* 2019, pp. 171-172.
5. A H M Jakaria, Md Mosharaf Hossain & Mohammad Ashiqur Rahman, "Smart Weather Forecasting Using Machine Learning: A Case Study in Tennessee" in *ACM Mid-Southeast conference,* 2020, pp. 1945-1948.