

Project Title: Weather Prediction using Machine Learning

Problem Description:

The goal of this project is to develop a machine learning model that can predict weather conditions in advance. By leveraging historical weather data, the aim is to create a model that can forecast future weather patterns with high precision.

Data Exploration:

1. Data Sources:

The dataset utilized in this arrangement has been gathered from Kaggle which is “Historical Weather Data for Indian Cities” from which we have chosen the data for “Kanpur City”.

	date_time	maxtempC	mintempC	totalSnow_cm	sunHour	uvIndex	uvIndex.1	moon_illumination	moonrise	moonset	...	WindChillC	WindGustKmph	cloudcover	hu
0	2009-01-01 00:00:00	24	8	0.0	8.7	4	1	31	10:15 AM	10:03 PM	...	11	17	1	
1	2009-01-01 01:00:00	24	8	0.0	8.7	4	1	31	10:15 AM	10:03 PM	...	11	17	1	
2	2009-01-01 02:00:00	24	8	0.0	8.7	4	1	31	10:15 AM	10:03 PM	...	10	18	1	
3	2009-01-01 03:00:00	24	8	0.0	8.7	4	1	31	10:15 AM	10:03 PM	...	9	18	1	
4	2009-01-01 04:00:00	24	8	0.0	8.7	4	1	31	10:15 AM	10:03 PM	...	11	15	1	

Data Types of the data:

```

date_time      object
maxtempC       int64
mintempC       int64
totalSnow_cm   float64
sunHour        float64
uvIndex        int64
uvIndex.1      int64
moon_illumination  int64
moonrise      object
moonset      object
sunrise      object
sunset      object
DewPointC     int64
FeelsLikeC    int64
HeatIndexC    int64
WindChillC    int64
WindGustKmph  int64
cloudcover    int64
humidity      int64
precipMM      float64
pressure      int64
tempC         int64
visibility     int64
winddirDegree int64
windspeedKmph int64
dtype: object

```

Null values:

```

date_time      False
maxtempC       False
mintempC       False
totalSnow_cm   False
sunHour        False
uvIndex        False
uvIndex.1      False
moon_illumination False
moonrise      False
moonset      False
sunrise      False
sunset      False
DewPointC     False
FeelsLikeC    False
HeatIndexC    False
WindChillC    False
WindGustKmph  False
cloudcover    False
humidity      False
precipMM      False
pressure      False
tempC         False
visibility     False
winddirDegree False

```

Describe() method:

	maxtempC	mintempC	totalSnow_cm	sunHour	uvIndex	uvIndex.1	moon_illumination	DewPointC	FeelsLikeC	HeatIndexC	WindChillC
count	96432.000000	96432.000000	96432.0	96432.000000	96432.000000	96432.000000	96432.000000	96432.000000	96432.000000	96432.000000	96432.000000
mean	32.788701	21.469886	0.0	11.234769	6.776506	4.656711	46.094077	9.496474	28.293896	28.39490	27.667880
std	6.602928	7.482104	0.0	1.999857	1.500904	3.320682	31.249725	8.676747	8.543340	8.40839	7.940635
min	15.000000	3.000000	0.0	3.900000	3.000000	1.000000	0.000000	-21.000000	5.000000	5.00000	5.000000
25%	28.000000	15.000000	0.0	8.700000	6.000000	1.000000	18.000000	3.000000	22.000000	23.00000	22.000000
50%	33.000000	23.000000	0.0	11.600000	7.000000	5.000000	46.000000	8.000000	29.000000	29.00000	28.000000
75%	37.000000	27.000000	0.0	13.300000	8.000000	8.000000	73.000000	18.000000	35.000000	35.00000	33.000000
max	48.000000	36.000000	0.0	13.900000	10.000000	11.000000	100.000000	27.000000	61.000000	61.00000	52.000000

Machine learning approach:

Machine Learning Algorithm can be broadly classified into three types: Supervised Learning Algorithms Unsupervised Learning Algorithms Reinforcement Learning algorithm

There are various machine learning algorithms for the prediction like Linear regression, Decision Tree, Random Forest, Support Vector Machines, K-Means Clustering, Logistic regression.

Supervised learning is the types of machine learning in which machines are trained using well "labelled" training data, and on basis of that data, machines predict the output. The labelled data means some input data is already tagged with the correct output. Supervised learning can be used for regression tasks.

In this project, we are concentrating on the temperature prediction of Kanpur city with the help of various machine learning algorithms and various regressions. By applying various regressions on the historical weather dataset of Kanpur city we are predicting the temperature like first we are applying Multiple Linear regression, then Decision Tree regression, and after that, we are applying Random Forest Regression.

Information about Data:

There are 96432 rows in the data. The dataset utilized in this arrangement has been gathered from Kaggle which is "Historical Weather Data for Indian Cities" from which we have chosen the data for "Kanpur City". The dataset was created by keeping in mind the necessity of such historical weather data in the community. The dataset was used with the help of the worldweatheronline.com API and the wwo_hist package. The datasets contain hourly weather data from 01-01-2009 to 01-01-2020. The data of city is for more than 10 years. This data can be used to visualize the change in data due to global warming or can be used to predict the weather for upcoming days, weeks, months, seasons, etc. The main target of this dataset can be used to predict the weather for the next day or week with huge amounts of data provided in the dataset. Furthermore, this data can also be used to make visualization which would help to understand the impact of global warming over the various aspects of the weather like precipitation, humidity, temperature, etc.