

Exploratory Analysis of Rain Fall Data in India for Agriculture

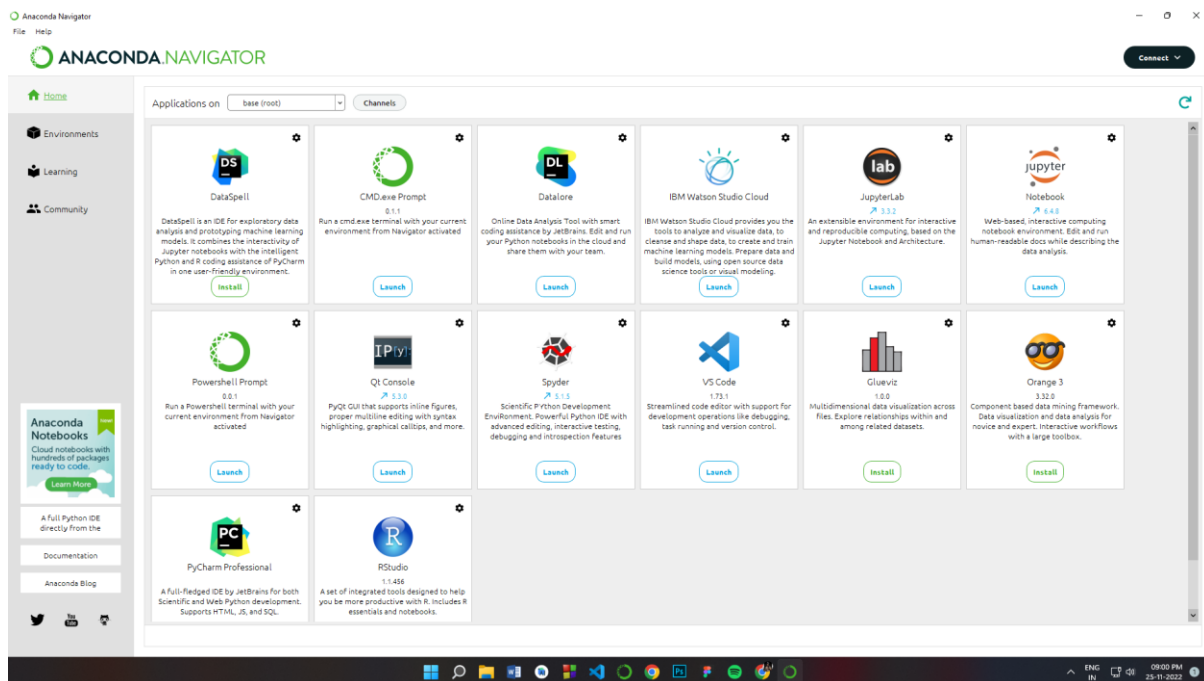
PNT2022TMID30162

Pre – Requisite

Packages

Package	Version
certifi	2022.9.24
charset-normalizer	2.1.1
click	8.1.3
colorama	0.4.6
cycler	0.11.0
Flask	2.2.2
fonttools	4.38.0
ibm-db	3.1.3
idna	3.4
importlib-metadata	5.0.0
itsdangerous	2.1.2
Jinja2	3.1.2
joblib	1.2.0
kiwisolver	1.4.4
MarkupSafe	2.1.1
matplotlib	3.5.3
numpy	1.21.6
packaging	21.3
pandas	1.3.5
pickle5	0.0.12
Pillow	9.3.0
pip	22.3.1
pyarsing	3.0.9
python-dateutil	2.8.2
pytz	2022.6
requests	2.28.1
scikit-learn	1.0.2
scipy	1.7.3
setuptools	40.8.0
six	1.16.0
threadpoolctl	3.1.0
typing_extensions	4.4.0
urllib3	1.26.12
Werkzeug	2.2.2
zipp	3.10.0

Anaconda



Pre Process the Data

1. Import Required Libraries

```
import numpy as np
import pandas as pd
import sklearn
```

2. Read the Dataset

```
data = pd.read_csv(r"H:\IBM Project\Data Set\India Full DS.csv")
```

```
data.head()
```

	Date	Location	MinTemp	MaxTemp	WindDirection	RainToday	RainTomorrow	Unnamed: 7	Unnamed: 8	Unnamed: 9
0	01-01-2022	Andhra Pradesh	16.9	22.9	SW	No	No	NaN	NaN	NaN
1	02-01-2022	Andhra Pradesh	13.9	24.1	SE	No	No	NaN	NaN	NaN
2	03-01-2022	Andhra Pradesh	13.8	29.1	SW	No	No	NaN	NaN	NaN
3	04-01-2022	Andhra Pradesh	15.1	33.3	W	No	No	NaN	NaN	NaN
4	05-01-2022	Andhra Pradesh	14.6	30.5	SW	No	No	NaN	NaN	NaN

3. Check Null Values

```
data.isnull().any()
```

Location	False
MinTemp	False
MaxTemp	False
WindDirection	False
RainToday	False
RainTomorrow	False
dtype:	bool

Model Building

1. Splitting the Data and Train and Test

```
from sklearn.model_selection import train_test_split  
x_train,x_test,y_train,y_test = train_test_split(x,y,test_size=0.3,random_state=1)
```

2. Applying Random Forest

```
from sklearn.ensemble import RandomForestClassifier  
rf = RandomForestClassifier()  
rf.fit(x_train, y_train)
```

```
RandomForestClassifier()
```