

```
[43]: import pandas as pd
```

```
[44]: data=pd.read_csv(r"C:\Users\VYSHNAVI\Documents\Final!!.csv")
data
```

	PARAMETERS	SO2	CO	NO	NO2	NOX	NH3	O3	AQI	WS	WD	RH	SR	TC
0	01/01/2023	2.02	0.06	8.93	17.84	26.00	8.93	26.07	72	0.33	37.96	80.27	199.80	28.70
1	02/01/2023	1.63	0.12	8.89	17.74	26.00	8.89	25.97	81	0.47	33.84	82.50	189.10	28.11
2	03/01/2023	3.46	0.03	8.91	17.89	26.00	8.91	26.03	82	0.19	54.37	71.00	203.10	28.55
3	04/01/2023	2.79	0.11	8.88	17.77	26.00	8.88	25.95	79	0.54	40.44	78.11	205.20	28.25
4	05/01/2023	1.15	0.22	9.48	17.73	26.00	9.48	23.93	84	0.43	42.53	78.56	202.30	28.55
...
670	12/27/2024	3.53	0.53	1.45	5.34	6.68	1.45	36.20	66	0.80	151.19	83.99	98.19	30.23
671	12/28/2024	3.65	0.48	1.44	5.35	6.72	1.44	36.17	64	0.82	159.29	76.87	95.05	29.99
672	12/29/2024	4.64	0.49	1.42	5.35	6.58	1.42	36.21	60	0.82	148.56	78.66	100.60	30.42
673	12/30/2024	4.18	0.53	1.44	5.37	6.77	1.44	36.20	60	0.70	143.66	86.57	98.00	30.22

```
[49]: data.isnull().sum()
```

```
[49]: PARAMETERS    0  
      SO2          0  
      CO          0  
      NO          0  
      NO2         0  
      NOX         0  
      NH3         0  
      O3          0  
      AQI         0  
      WS          0  
      WD          0  
      RH          0  
      SR          1  
      TC          0  
      dtype: int64
```

```
[50]: data.dropna(inplace=True)
```

```
[51]: data.isnull().sum()
```

```
[51]: PARAMETERS    0  
      SO2          0  
      CO          0  
      NO          0  
      NO2         0
```

dtype: int64

```
[52]: x=data[['SO2', 'CO', 'NO', 'NO2', 'NOX', 'NH3', 'O3','WS',  
            'WD', 'RH', 'SR', 'TC']]
```

```
[53]: y=data['AQI']
```

```
[54]: from sklearn.model_selection import train_test_split  
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.1)
```

```
[55]: x_train.shape
```

```
[55]: (606, 12)
```

```
[56]: from sklearn.neighbors import KNeighborsRegressor  
from sklearn.model_selection import train_test_split  
from sklearn.datasets import make_regression  
import matplotlib.pyplot as plt
```

```
[57]: regressor=KNeighborsRegressor(n_neighbors=3)  
regressor.fit(x_train,y_train)
```

```
[57]: KNeighborsRegressor
```

```
KNeighborsRegressor(n_neighbors=3)
```

```
[58]: y_pred = regressor.predict(x_test)
```

```
[59]: x_test.isnull().sum()
```

```
[59]: SO2      0
      CO      0
      NO      0
      NO2     0
      NOX     0
      NH3     0
      O3      0
      WS      0
      WD      0
      RH      0
      SR      0
      TC      0
      dtype: int64
```

```
[60]: from sklearn.metrics import r2_score
      r2_score(y_test,y_pred)
```

```
[60]: 0.6524257963410344
```

```
[62]: from sklearn.tree import DecisionTreeRegressor
      from sklearn.model_selection import train_test_split
      from sklearn.datasets import make_regression
      import matplotlib.pyplot as plt
```

```
[63]: model=DecisionTreeRegressor()  
      model.fit(x_train,y_train)  
      y_pred=model.predict(x_test)  
  
[64]: from sklearn.metrics import r2_score  
      r2_score(y_test,y_pred)  
  
[64]: 0.7813226547638799  
  
[65]: from sklearn.ensemble import RandomForestRegressor  
      from sklearn.model_selection import train_test_split  
  
[66]: rf_regressor = RandomForestRegressor(n_estimators=100)  
  
[68]: rf_regressor.fit(x_train, y_train)  
      y_pred = rf_regressor.predict(x_test)  
  
[69]: from sklearn.metrics import r2_score  
      r2_score(y_test,y_pred)  
  
[69]: 0.8090077205238984  
  
[70]: from sklearn.linear_model import LinearRegression  
      reg=LinearRegression()  
  
[71]: reg.fit(x_train,y_train)
```

```
[71]: reg.fit(x_train,y_train)
```

```
[71]: ▾ LinearRegression ⓘ ?  
LinearRegression()
```

```
[72]: y_pred=reg.predict(x_test)
```

```
[73]: from sklearn.metrics import r2_score  
r2_score(y_test,y_pred)
```

```
[73]: 0.6847726659391253
```

```
[74]: import pickle  
file=open('projectrf.pkl','wb')  
pickle.dump(rf_regressor,file)  
file.close()
```

```
[ ]:
```