

# Course Name: Data Analytics with Cognos

## Project: Air Quality analysis in Tamil Nadu

In this phase Regression method is implemented. Regression is supervised learning technique that fits the data in a given range. Example Regression techniques in Python:

- Random Forest Regressor
- Ada Boost Regressor
- Bagging Regressor
- Linear Regression etc.

From one of these techniques, we chosen the Random Forest regression technique for the project. The python code format of Random forest is

### SAMPLE CODE:

```
# importing Randomforest
from sklearn.ensemble import AdaBoostRegressor
from sklearn.ensemble import RandomForestRegressor

# creating model
ml = RandomForestRegressor()

# separating class label and other attributes
train1 = train.drop(['air_quality_index'], axis=1)
target = train['air_quality_index']

# Fitting the model
ml.fit(train1, target)
'''RandomForestRegressor(bootstrap=True,
ccp_alpha=0.0, criterion='mse',
                        max_depth=None,
max_features='auto', max_leaf_nodes=None,
                        max_samples=None,
min_impurity_decrease=0.0,
                        min_impurity_split=None,
min_samples_leaf=1,
                        min_samples_split=2,
min_weight_fraction_leaf=0.0,
                        n_estimators=100, n_jobs=None,
oob_score=False,
                        random_state=None, verbose=0,
warm_start=False)'''
```

```

# calculating the score and the score
is 97.96360799890066%
m1.score(train1, target) * 100

# predicting the model with other values (testing the
data)
# so AQI is 123.71
m1.predict([[123, 45, 67, 34, 5, 0, 23]])

# Adaboost model
# importing module
# defining model
m2 = AdaBoostRegressor()
# Fitting the model
m2.fit(train1, target)

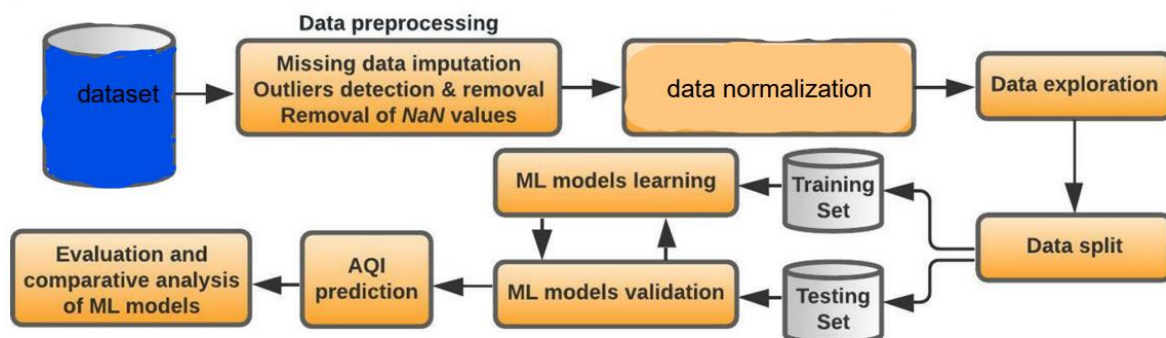
'''AdaBoostRegressor(base_estimator=None,
learning_rate=1.0, loss='linear',
                        n_estimators=50,
random_state=None)'''

# calculating the score and the score
is 96.15377360010211%
m2.score(train1, target)*100

# predicting the model with other values (testing the
data)
# so AQI is 94.42105263
m2.predict([[123, 45, 67, 34, 5, 0, 23]])

```

## MODEL:

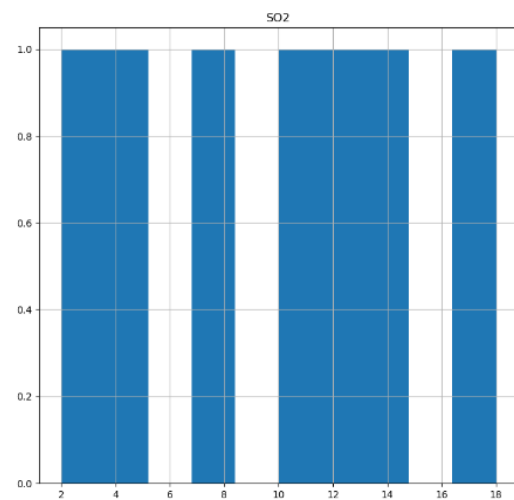
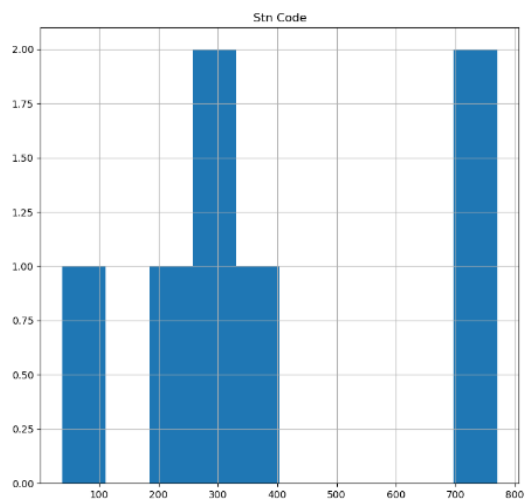


# ANALYSING OF DATA FROM SET

## HISTOGRAM

```
In [16]: import matplotlib.pyplot as plt
import numpy as np
import seaborn as sns
import pandas as pd

data = pd.read_excel("Book1.xlsx")
data.head()
df = pd.DataFrame(data)
data.hist(figsize=(20,30))
```



# BOXPLOT:

```
In [21]: import seaborn as sns
sns.boxplot(df['S02'])
sns.despine()
```

