

Siddhardhan

Model Evaluation in Machine Learning



Work Flow of a ML Project



Data



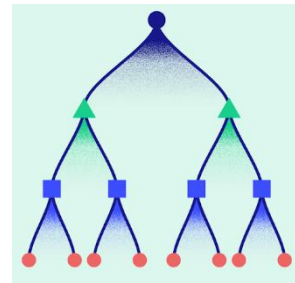
Data pre processing



Data Analysis



Train Test split



XGBoost Regressor



Evaluation

Types of Supervised Learning

Supervised Learning

```
graph TD; SL[Supervised Learning] --> C[Classification]; SL --> R[Regression];
```

Classification

Classification is about predicting a class or discrete values

Eg: Male or Female; True or False

Evaluation metric for

Classification: **Accuracy score**

Regression

Regression is about predicting a quantity or continuous values

Eg: Salary; age; Price.

Evaluation metric for

Regression: **Mean Absolute Error**

Accuracy Score

In Classification, **Accuracy Score** is the ratio of **number of correct predictions** to the **total number of input data points**.



$$\text{Accuracy Score} = \frac{\text{Number of correct predictions}}{\text{Total Number of data points}} \times 100 \%$$

Number of correct predictions = 128

Accuracy Score = 85.3 %

Total Number of data points = 150

```
from sklearn.metrics import accuracy_score
```

Mean Squared Error

Mean Squared Error measures the average of the squares of the errors, that is, the average squared difference between the estimated values and the actual value.



$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

Actual Value ($Y_i = 140 \text{ mg/dL}$)

Predicted Value ($\hat{Y}_i = 160 \text{ mg/dL}$)

```
from sklearn.metrics import mean_squared_error
```