Calories burnt Prediction data set

Kaggle link : **https://meet.google.com/xoh-hwyr-yrf**

- ✅ Total Number of Samples : 15000

- ✅ Number of Features = 7

- ✅ Features Names :

  User_ID ,Gender ,Age ,Height ,Weight Duration ,Heart_Rate ,Body_Temp

- ✅ Number of Samples in Train Set = 12000

- ✅ Number of Samples in Test Set = 3000

- ✅ Number of Samples in Validation Set = 2400 , Number of Folds = 5

Used Libraries:

```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.model_selection import cross_val_score
from sklearn.model_selection import KFold
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_absolute_error , mean_squared_error , r2_score
from sklearn.neighbors import KNeighborsRegressor
```

Preprocessing Phase:

I Used Function Info from pandas library to now information About my Dataset

1-The Number of Missing Data (no missing data)

2-The Data Type of each column (there is one categorical column)

```python
#Converting the text data to numerical values
gender = {
    'male':1,
    'female':0
}
df['Gender']=df['Gender'].map(gender)
df.head()
```

|	User_ID	Gender	Age	Height	Weight	Duration	Heart_Rate	Body_Temp	Calories
0	14733363	1	68	190.0	94.0	29.0	105.0	40.8	231.0
1	14861698	0	20	166.0	60.0	14.0	94.0	40.3	66.0
2	11179863	1	69	179.0	79.0	5.0	88.0	38.7	26.0
3	16180408	0	34	179.0	71.0	13.0	100.0	40.5	71.0
4	17771927	0	27	154.0	58.0	10.0	81.0	39.8	35.0

Correlation between Columns

- Create a correlation plot using seaborn
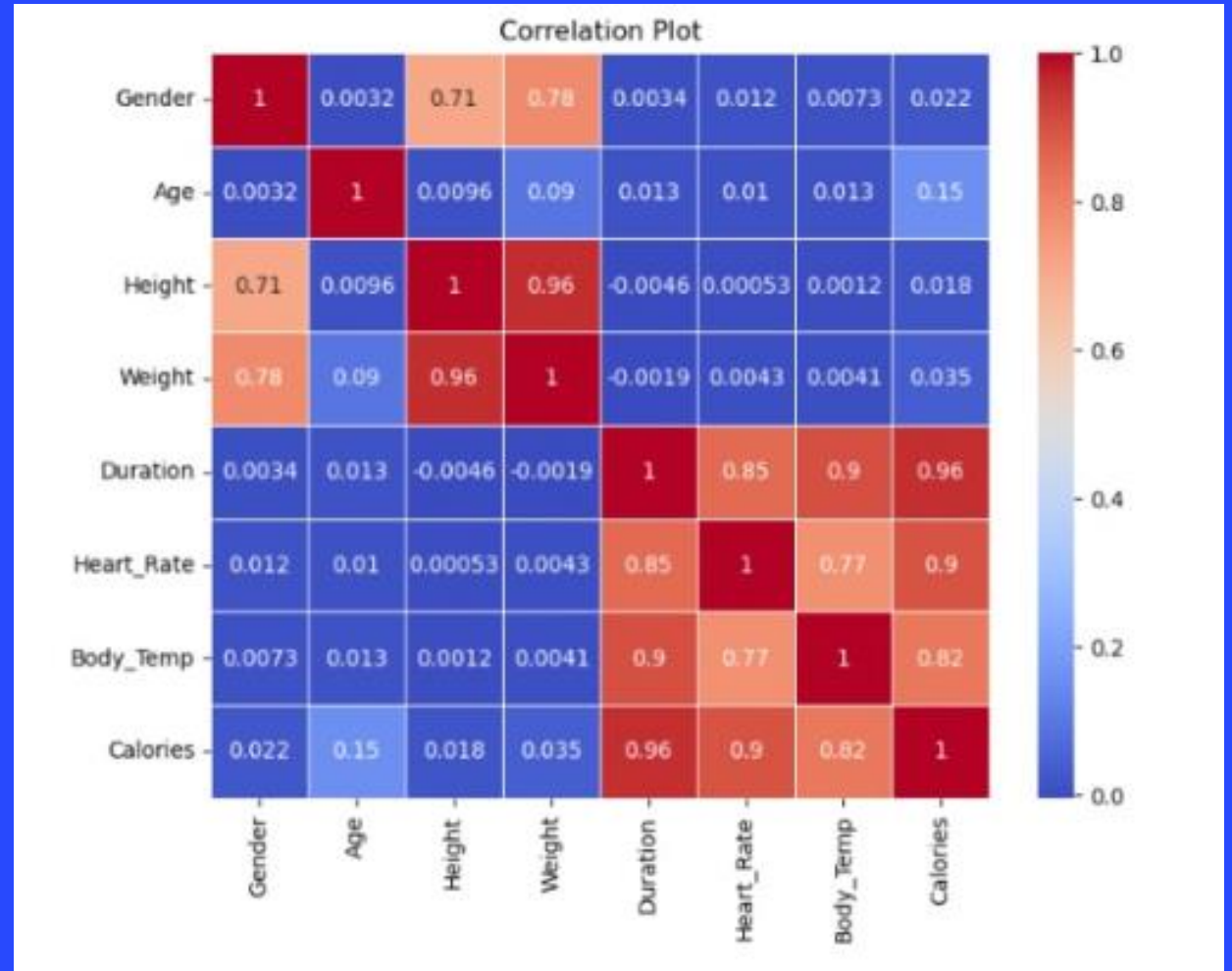
Regression Details:

Using Linear Regression from sklearn to Predict Calories Burnt During Exercise
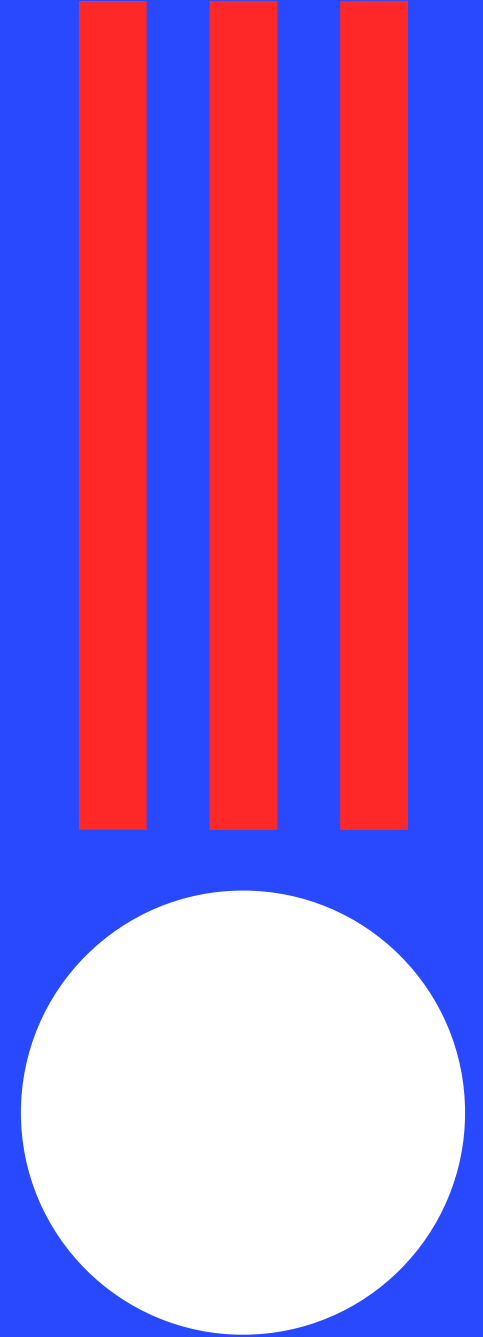
Regression Model Evaluation:

Cross-validation scores:

 [0.96718642 , 0.9656019 , 0.96779367
0.96628163 , 0.96641819]

Mean CV Score: 0.9666563627372389
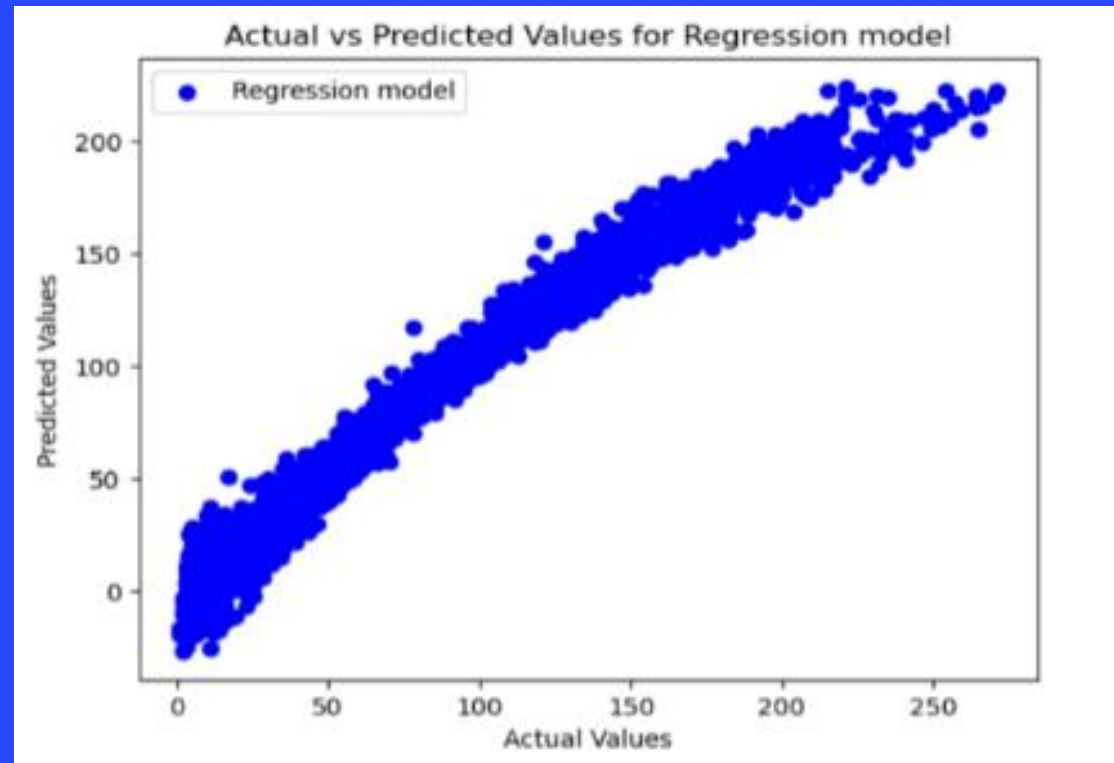
R-squared Score:
0.9683635333933688

Mean Absolute Error:
8.09067963631315l
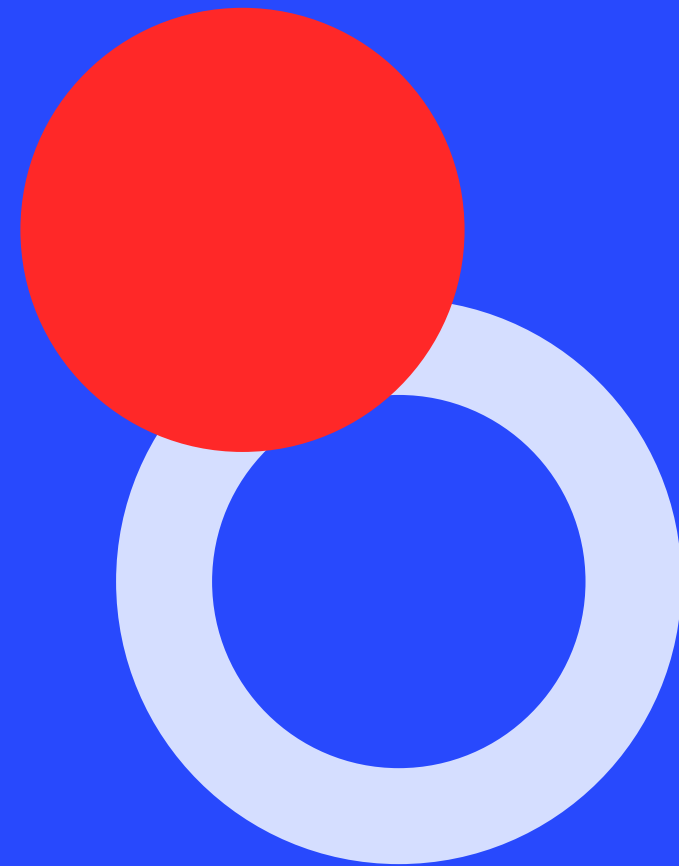
Mean Squared Error:
118.79074609385702

Predicted Values vs Actual Values:

Actual vs Predicted Values for Regression model

Knn Details:

Using KNeighborsRegressor from sklearn to Predict Calories Burnt During Exercise

The number of Neighbors = 5

Knn Model Evaluation:

KNN Cross-validation scores

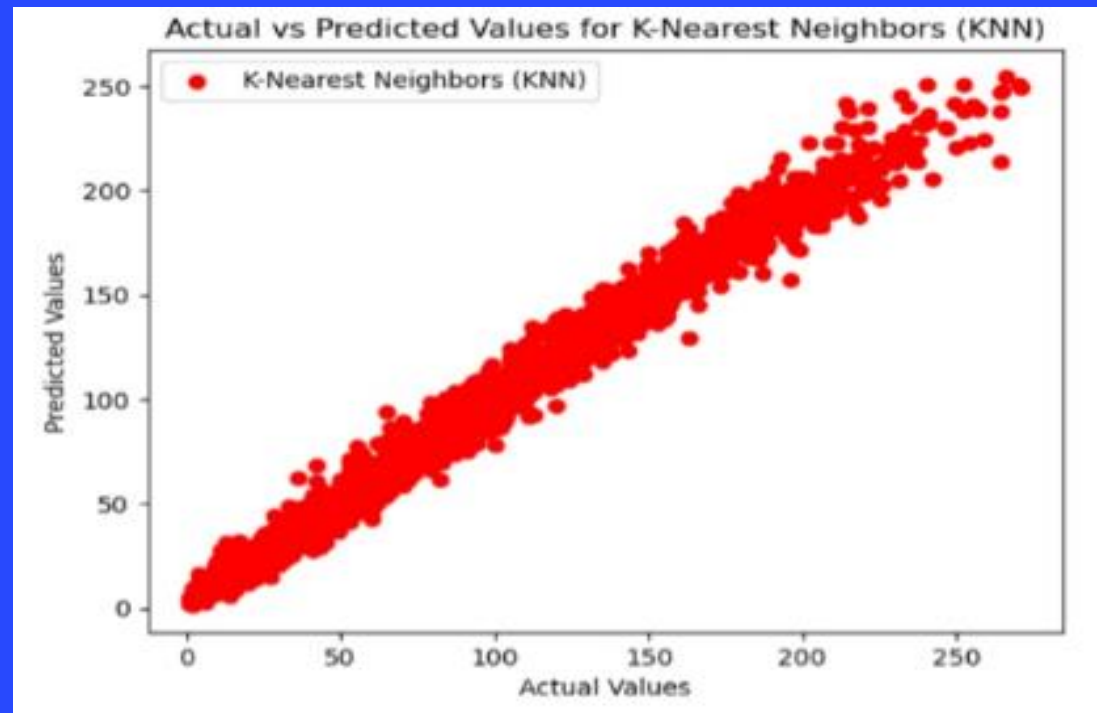: [0.98652628 0.98503894 0.98675647 0.98475885 0.98491032]

KNN Mean CV Score: 0.9855981714151658

R2 Score for knn model : 0.9877563495985314

KNN Mean Absolute Error: 4.995533333333333

KNN Mean Squared Error: 47.24380000000001

Predicted Values vs Actual Values:

Actual vs Predicted Values for K-Nearest Neighbors (KNN)

Stanford Dogs Data set

Dataset link :
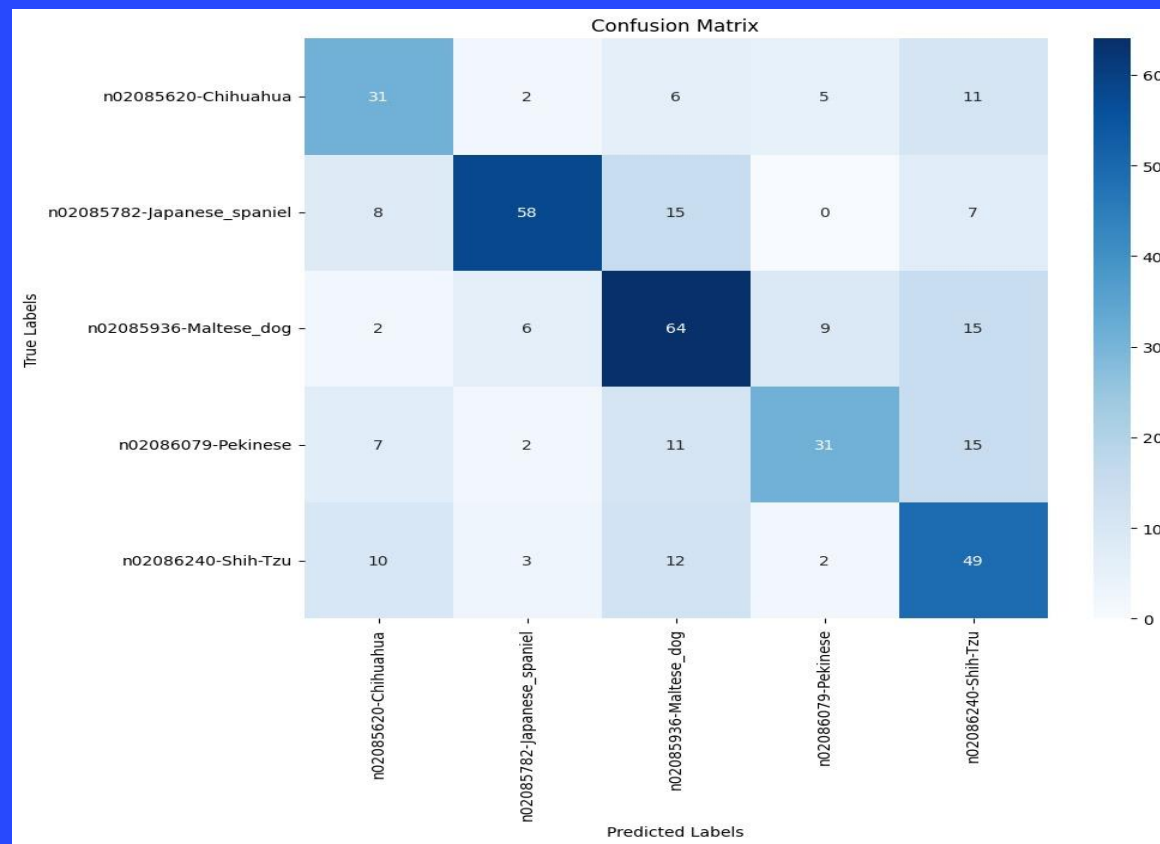http://vision.stanford.edu/aditya86/ImageNetDogs/main.html

- Total Number of Samples : 952

- Number of Classes : 5

- Size of Image is 128 * 128

- Number of Samples in Training/Validating Set = 761

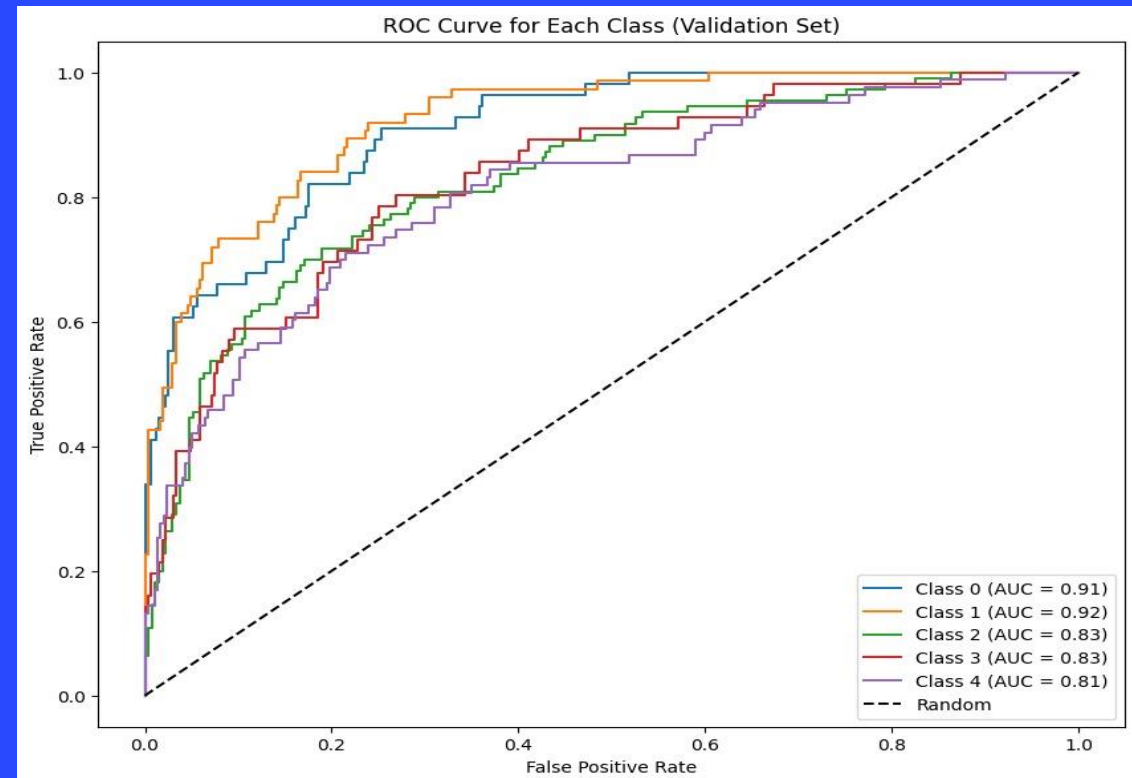- Number of Samples in Testinf Set = 191

- Accuracy : 61.15%

Learning Curve for Logistic Regression:
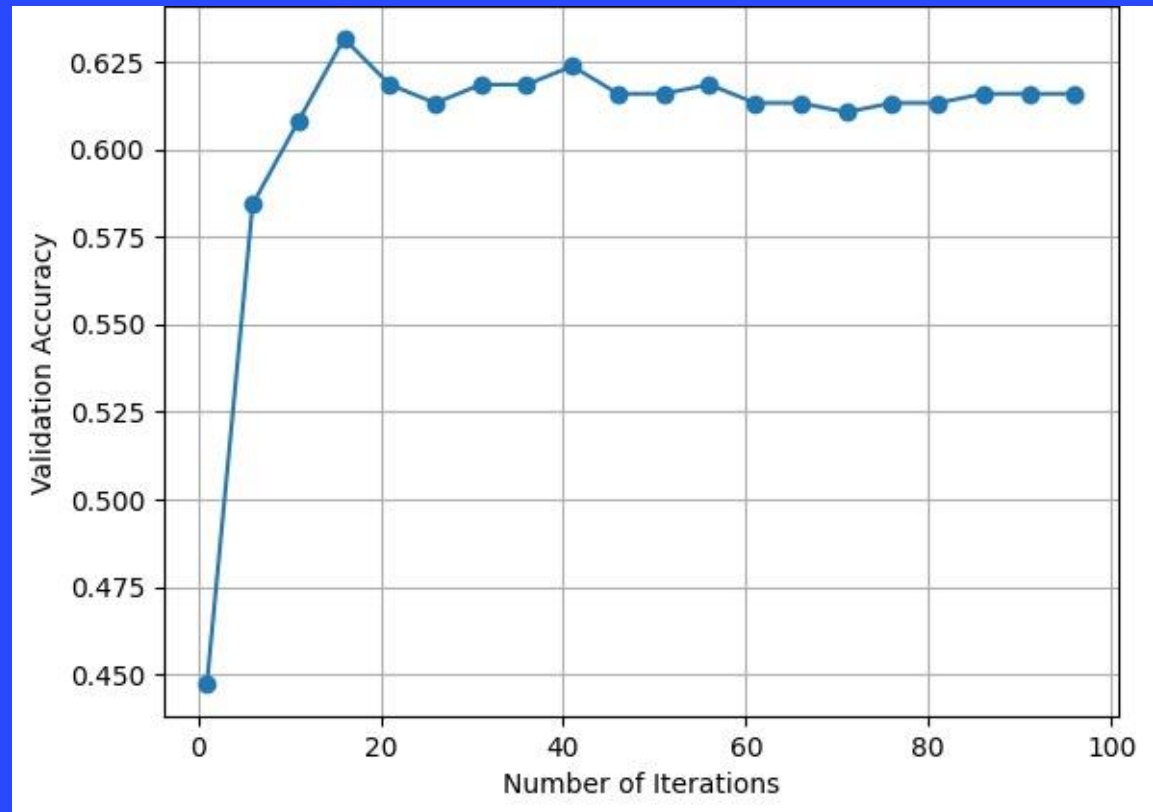
Confusion Matrix:

ROC Curve for each Class (validation test)

ROC Curve for Each Class (Validation Set)

Class 0 (AUC = 0.91)
Class 1 (AUC = 0.92)
Class 2 (AUC = 0.83)
Class 3 (AUC = 0.83)
Class 4 (AUC = 0.81)
Random

Thank you

Omar Mohamed 20210625

Amr Mohamed 20210645

Farida Ahmed 20210674

Farah Khaled 20210673

Fatma Mohamed 20210667

Nouran Yasser 20211015