# MACHINE LEARNING ASSIGNMENT

## Classification


Submitted by:

Name: Khuram Shahzad

Roll No: p218742

Subject Machine Learning (MS DS)

Submitted to:

Dr. Muhammad Taimoor Khan


APRIL 24, 2022


FAST NATIONAL UNIVERSITY OF COMPUTER AND EMERGING SCIENCES – PESHAWAR CAMPUS
Jamrud Road 160 Industrial Estate Road, Peshawar - PK

# 1 TABLE OF CONTENTS





**Prepared by: Khuram Shahzad (p21-8742)**

# 2 Assignment Question:

*Classify the following dataset to distinguish between skin color and nonskin color. You may try different models but report at least the two top scoring classifiers. Use the following settings for the experiments.*

1. *Stratified 5-Fold cross-validation (the dataset is imbalanced)*
2. *Report precision, recall, and F1-score (with micro as averaging scheme)*

*Data Set: https://archive.ics.uci.edu/ml/datasets/Skin+Segmentation*

*Submit a code document and a word document with setup, results and discussion.*

# 3 About Data (discussion):

1. As we know data is imbalanced so we used stratified 5-fold validation.
2. Our data label is 1 and 2 but we changed them into 0 and 1 respectively for smooth classification and evaluation.

# 4 Steps(discussion):

It's to be noted that we have defined a function for each step and we have made the generalized code.

## 4.1 Importing Libraries

In this step we just imported the all libraries which is being used in assignment

## 4.2 Importing the Data

Here we just imported / read our text file.

## 4.3 Exploratory Data Analysis (EDA)

Here we just perform some Exploratory data analysis to find the co-relation between the data.

## 4.4 Data Preprocessing

In Data Preprocessing we just convert our data into NumPy Array and then we have just changed our data label is 1 and 2 to 0 and 1 respectively for smooth classification and evaluation.

## 4.5 Training and Testing

Here we have defined our training and testing generalized function for classifier.

## 4.6 EVALUATION

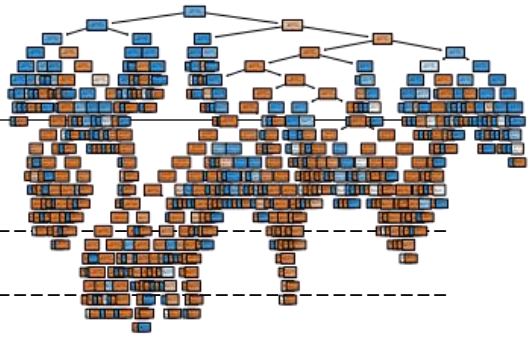
Here we have a function for evaluation report.

## 4.7 DRIVER CLASS

Finally, here the driver class from where we will drive the whole code.



**Prepared by: Khuram Shahzad (p21-8742)**



# 5 DECISION TREE CLASSIFIER:

## 5.1 OUTPUT

---

**Classifier Name: DecisionTreeClassifier**

---

**1 of KFold 5**
>Train: 1=40686, 2=155358, Test: 1=10172, 2=38840
ROC AUC score: 0.9986740068757636
**2 of KFold 5**
>Train: 1=40686, 2=155359, Test: 1=10172, 2=38839
ROC AUC score: 0.9988577450877426
**3 of KFold 5**
>Train: 1=40686, 2=155359, Test: 1=10172, 2=38839
ROC AUC score: 0.9992767287386224
**4 of KFold 5**
>Train: 1=40687, 2=155358, Test: 1=10171, 2=38840
ROC AUC score: 0.9989641889396774
**5 of KFold 5**
>Train: 1=40687, 2=155358, Test: 1=10171, 2=38840
ROC AUC score: 0.9992228591396087

---

- **List of possible accuracy are :** 0.99925 , 0.99929 , 0.99941
- **Maximum Accuracy That can be obtained from this model is:** 99.94286996797 %
- **Minimum Accuracy:** 99.91634531023648 %
- **Overall(Mean) Accuracy:** 99.92532322806302 %
- **Standard Deviation is:** 0.00011044005334967038

---

- **Cv:** [0.9986740068757636, 0.9988577450877426, 0.9992767287386224, 0.9989641889396774, 0.9992228591396087]
- **Mean cv Score :** 0.9989991057562829

---

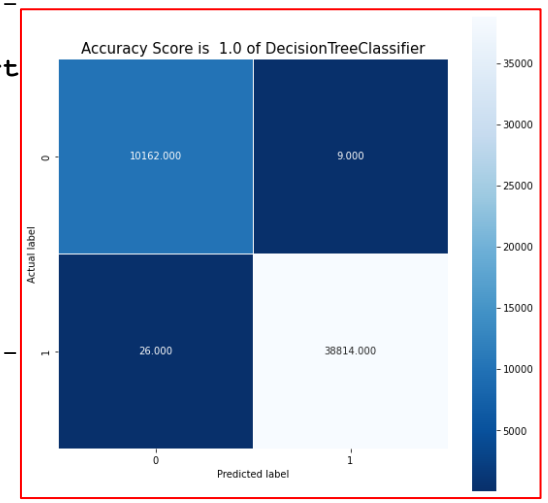
**Confusion Matrix on tested 5th k fold data**

---

**Classification report:**

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 1.00 | 1.00 | 1.00 | 10171 |
| 1 | 1.00 | 1.00 | 1.00 | 38840 |
| accuracy | | | 1.00 | 49011 |
| macro avg | 1.00 | 1.00 | 1.00 | 49011 |
| weighted avg | 1.00 | 1.00 | 1.00 | 49011 |

---






**Prepared by: Khuram Shahzad (p21-8742)**

# 6 K NEIGHBORS CLASSIFIER:

## 6.1 OUTPUT

```
--------------------------------------------------------------------------------
Classifier Name:  KNeighborsClassifier
--------------------------------------------------------------------------------
1 of KFold 5
           >Train: 1=40686, 2=155358, Test: 1=10172, 2=38840
           ROC AUC score: 0.9996079431714774
2 of KFold 5
           >Train: 1=40686, 2=155359, Test: 1=10172, 2=38839
           ROC AUC score: 0.9996313440998963
3 of KFold 5
           >Train: 1=40686, 2=155359, Test: 1=10172, 2=38839
           ROC AUC score: 0.9997167795257345
4 of KFold 5
           >Train: 1=40687, 2=155358, Test: 1=10171, 2=38840
           ROC AUC score: 0.9995541014359904
5 of KFold 5
           >Train: 1=40687, 2=155358, Test: 1=10171, 2=38840
           ROC AUC score: 0.9996547541165535
--------------------------------------------------------------------------------
```

- **List of possible accuracy are :** 0.99955 , 0.99953 , 0.99955
- **Maximum Accuracy That can be obtained from this model is:** 99.95511303354%
- **Minimum Accuracy:** 99.94082960967945 %
- **Overall(Mean) Accuracy:** 99.95103158428502 %
- **Standard Deviation** is: 5.948772589477762e-05

```
--------------------------------------------------------------------------------
Cv:  [0.9996079431714774, 0.9996313440998963, 0.9997167795257345, 0.9995541014359904, 0.9996547541165535]
Mean cv Score : 0.9996329844699303
--------------------------------------------------------------------------------
```

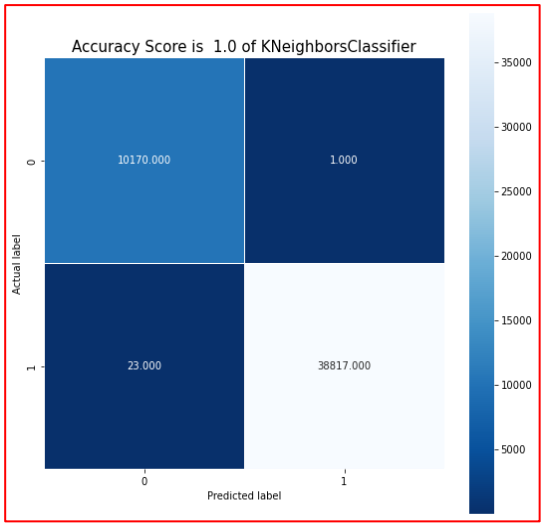
**Confusion Matrix on tested 5th k fold data**

```
--------------------------------------------------------
```

**Classification report:**

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 1.00 | 1.00 | 1.00 | 10171 |
| 1 | 1.00 | 1.00 | 1.00 | 38840 |
| accuracy | | | 1.00 | 49011 |
| macro avg | 1.00 | 1.00 | 1.00 | 49011 |
| weighted avg | 1.00 | 1.00 | 1.00 | 49011 |

```
--------------------------------------------------------
```






**Prepared by: Khuram Shahzad (p21-8742)**

# 7 AdaBoostClassifier:

## 7.1 Output

```
--------------------------------------------------------------------------------
Classifier Name:  AdaBoostClassifier
--------------------------------------------------------------------------------
1 of KFold 5
           >Train: 1=40686, 2=155358, Test: 1=10172, 2=38840
           ROC AUC score: 0.9235979362989536
2 of KFold 5
           >Train: 1=40686, 2=155359, Test: 1=10172, 2=38839
           ROC AUC score: 0.9227771870418569
3 of KFold 5
           >Train: 1=40686, 2=155359, Test: 1=10172, 2=38839
           ROC AUC score: 0.9245092964566701
4 of KFold 5
           >Train: 1=40687, 2=155358, Test: 1=10171, 2=38840
           ROC AUC score: 0.9282899038187469
5 of KFold 5
           >Train: 1=40687, 2=155358, Test: 1=10171, 2=38840
           ROC AUC score: 0.9231340410089429
--------------------------------------------------------------------------------
```

- **List of possible accuracy are :** 0.95332 ,  0.95368  ,  0.9547
- **Maximum Accuracy That can be obtained from this model is:** 95.53569606822958 %
- **Minimum Accuracy:** 95.30513558180816 %
- **Overall(Mean) Accuracy:** 95.40227568604807 %
- **Standard Deviation is:** 0.0009746834098173316

```
--------------------------------------------------------------------------------
```

- **Cv:** [0.9235979362989536, 0.9227771870418569, 0.9245092964566701, 0.9282899038187469, 0.9231340410089429]
- **Mean cv Score :** 0.9244616729250341

```
--------------------------------------------------------------------------------
```

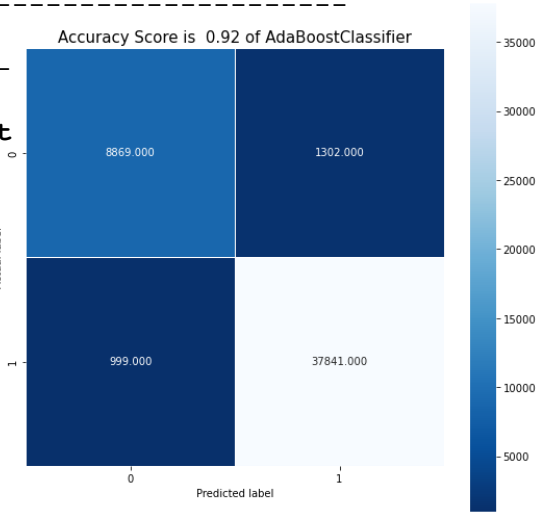
**Confusion Matrix on tested 5th fold data**

```
--------------------------------------------------------
```

**Classification report:**

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.90 | 0.87 | 0.89 | 10171 |
| 1 | 0.97 | 0.97 | 0.97 | 38840 |
| accuracy | | | 0.95 | 49011 |
| macro avg | 0.93 | 0.92 | 0.93 | 49011 |
| weighted avg | 0.95 | 0.95 | 0.95 | 49011 |

```
--------------------------------------------------------
```






**Prepared by: Khuram Shahzad (p21-8742)**