

Metrics for Machine learning model

- Evaluation is always good in any field right! In the case of machine learning, it is best the practice. In this post, I will almost cover all the popular as well as common metrics used for machine learning.
- Different performance metrics are used to evaluate different Machine Learning Algorithms. For now, we will be focusing on the ones used for Classification problems.

The Classification Metric Are:

1. Confusion Matrix
2. Accuracy
3. Precision
4. Recall
5. Specificity
6. F1-Score
7. Roc And Auc

1. Confusion Matrix :

- The Confusion matrix is one of the most intuitive and easiest metrics used for finding the correctness and accuracy of the model. It is only used for Classification problem where the output can be of two or more types of classes.

There are 4 terms you should keep in mind:

1. True Positives (TP): It is the case where we predicted Yes and the real output was also yes.
2. True Negatives (TN): It is the case where we predicted No and the real output was also No.
3. False Positives (FP): It is the case where we predicted Yes but it was actually No.
4. False Negatives (FN): It is the case where we predicted No but it was actually Yes.

Lets Take an example to know better this 4 term :

Let's give a label of to our target variable:

1: When a person is having Corona

0: When a person is NOT having Corona.

Now lets apply this corona example in to this 4 term:

1. True Positives (TP) :- Ex: The case where a person is actually having Corona(1) and the model also predict person haing Corona(1) Then it comes under True positive(TP).
2. True Negatives (TN) :- Ex: The case where a person NOT having Corona(0) and the model also prdict person Not Having Corona(0) Then it comes under True Negatives(TN).
3. False Positives (FP) :- Ex: A person NOT having Corona(0) But the model is predict Person haing Corona(1) Then It comes under False Positives (FP).

4. False Negatives (FN) :- Ex: A person having Corona (1) But the model predict Person not Haing Corona (0)
Then it comes under False Negatives(FN).

SO in the above example you see these 4 type of term associate with Confusion Matrix.

	Predicted 0	Predicted 1
Actual 0	TN	FP
Actual 1	FN	TP

- we all want is that the model should give 0 False Positives and 0 False Negatives.
- But that's not the case in real life as any model will NOT be 100% accurate most of the times.
- So we might want to minimise either False Positives or False negatives.

2. Accuracy :

- Accuracy in classification problems is the number of correct predictions made by the model over all kinds predictions made.

	Predicted 0	Predicted 1
Actual 0	TN	FP
Actual 1	FN	TP

$$\text{Accuracy} = \frac{\text{TrueNegatives} + \text{TruePositive}}{\text{TruePositive} + \text{FalsePositive} + \text{TrueNegative} + \text{FalseNegative}}$$

- Accuracy is a good measure when the target variable classes in the data are nearly balanced.
- Accuracy should NEVER be used as a measure when the target variable classes in the data are a majority of one class. (imbalanced dataset)

3. Precision :

	Predicted 0	Predicted 1
Actual 0	TN	FP
Actual 1	FN	TP

$$\text{Precision} = \frac{TP}{TP + FP}$$

- Precision is a measure that tells us what proportion of patients that we diagnosed as having Corona, actually had Corona. The predicted positives (Total People predicted as Corona are TP and FP) and the people actually having a Corona are TP.
- precision gives us information about its performance with respect to false positives(how many did we caught).
- Precision is about being precise. So even if we managed to capture only one Corona case, and we captured it correctly, then we are 100% precise.

4. Recall or Sensitivity:

	Predicted 0	Predicted 1
Actual 0	TN	FP
Actual 1	FN	TP

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

- Recall is a measure that tells us what proportion of patients that actually had Corona was diagnosed by the algorithm as having Corona. The actual positives (People having Corona are TP and FN) and the people diagnosed by the model having a Corona are TP. (Note: FN is included because the Person actually had a Corona even though the model predicted otherwise).
- It is clear that recall gives us information about a classifier's performance with respect to false negatives (how many did we miss).
- Recall is not so much about capturing cases correctly but more about capturing all cases that have "Corona" with the answer as "Corona". So if we simply always say every case as "Corona", we have 100% recall.

5. Specificity :

	Predicted 0	Predicted 1
Actual 0	TN	FP
Actual 1	FN	TP

$$\text{Specificity} = \frac{\text{True Negatives}}{\text{True Negatives} + \text{False Positives}}$$

- Specificity is the exact opposite of Recall.
- Specificity is a measure that tells us what proportion of patients that did NOT have Corona, were predicted by the model as not Corona . The actual negatives (People actually NOT having Corona are FP and TN) and the people diagnosed by us not having Corona are TN. (Note: FP is included because the Person did NOT actually have Corona even though the model predicted otherwie).

All 4 Metric :

1. Accuracy 2. Precision 3. Recall or Sensitivity 4. Specificity

	Predicted 0	Predicted 1
Actual 0	TN	FP
Actual 1	FN	TP

$$\text{Accuracy} = \frac{\text{TrueNegatives} + \text{TruePositive}}{\text{TruePositive} + \text{FalsePositive} + \text{TrueNegative} + \text{FalseNegative}}$$

	Predicted 0	Predicted 1
Actual 0	TN	FP
Actual 1	FN	TP

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

	Predicted 0	Predicted 1
Actual 0	TN	FP
Actual 1	FN	TP

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

	Predicted 0	Predicted 1
Actual 0	TN	FP
Actual 1	FN	TP

$$\text{Specificity} = \frac{\text{True Negatives}}{\text{True Negatives} + \text{False Positives}}$$

f1-score and Roc Auc i am explain in next notebook.

F1-Score

- F1-score is a metric which takes into account both precision and recall as we can't always evaluate both and then take the higher one for our model. It is the harmonic mean of precision and recall. It tells us about the balance that exists between precision and recall.
- F1-score: This is the harmonic mean of Precision and Recall and gives a better measure of the incorrectly classified cases than the Accuracy Metric.

$$F1\text{-score} = \left(\frac{\text{Recall}^{-1} + \text{Precision}^{-1}}{2} \right)^{-1} = 2 * \frac{(\text{Precision} * \text{Recall})}{(\text{Precision} + \text{Recall})}$$

We use the Harmonic Mean since it penalizes the extreme values.

When It used:

- F1-score is used when the False Negatives and False Positives are crucial.
- F1-score is a better metric when there are imbalanced classes.
- In most real-life classification problems, imbalanced class distribution exists and thus F1-score is a better metric to evaluate our model on.

Some advantages of F1-score:

- Very small precision or recall will result in lower overall score. Thus it helps balance the two metrics.
- If you choose your positive class as the one with fewer samples, F1-score can help balance the metric across positive/negative samples.
- As illustrated by the first figure in this article, it combines many of the other metrics into a single one, capturing many aspects at once.

F- β score:

- The F-beta score is the weighted harmonic mean of precision and recall, reaching its optimal value at 1 and its worst value at 0. The beta parameter determines the weight of recall in the combined score.

The general formula for non-negative real β is:

$$F_{\beta} = \frac{(1 + \beta^2) \cdot (\text{precision} \cdot \text{recall})}{(\beta^2 \cdot \text{precision} + \text{recall})}$$

- Here Three Condition arise

1. $\beta=1$ (F1 score) - When FP And FN are equally important.
2. $\beta=0.5$ (F0.5 score) - when importance of FP is greater than FN.

3. $\beta=2$ (F2 score) - When FN impact is high we increase β value 1 to 2.

ROC And AUC Curve

- Before going to ROC and AUC I discuss about main 4 terms

1. TPR (Recall)
2. TNR (Specificity)
3. FPR (1-specificity)
4. FNR

$$TPR = \frac{TP}{Actual\ Positive} = \frac{TP}{TP + FN}$$

$$FNR = \frac{FN}{Actual\ Positive} = \frac{FN}{TP + FN}$$

$$TNR = \frac{TN}{Actual\ Negative} = \frac{TN}{TN + FP}$$

$$FPR = \frac{FP}{Actual\ Negative} = \frac{FP}{TN + FP}$$

- TPR (True Positive Rate): In machine learning, the true positive rate, also referred to as sensitivity or recall, is used to measure the percentage of actual positives which are correctly identified.
- TNR (True Negative Rate) : The Specificity of a test, also referred to as the true negative rate (TNR), is the proportion of samples that test negative using the test in question that are genuinely negative.
- FPR (False Positive Rate) : In statistics, when performing multiple comparisons, a false positive ratio (also known as fall-out or false alarm ratio) is the probability of falsely rejecting the null hypothesis for a particular test. The false positive rate is calculated as the ratio between the number of negative events wrongly categorized as positive (false positives) and the total number of actual negative events (regardless of classification).
- FNR (False Negative Rate) : The rate of occurrence of negative test results in those who have the attribute or disease for which they are being tested.

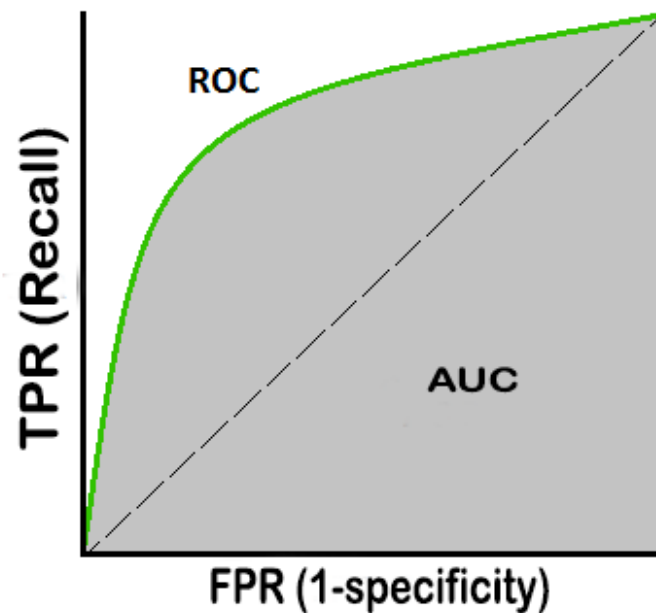
Now I am going to discuss about ROC And AUC

What is ROC curve?

- ROC curve is one of the important evaluating metrics that should be used to check the performance of a classification model. It is also called relative operating characteristic curve, because it is a comparison of two main characteristics (TPR and FPR).
- It is plotted between sensitivity (recall) and False Positive Rate (FPR = 1-specificity).

What is AUC?

- AUC also called as AREA UNDER CURVE. It is used in classification analysis in order to determine which of the used models predicts the classes best. An example of its application are ROC curves.
- AUC ranges in value from 0 to 1. A model whose predictions are 100% wrong has an AUC of 0 and if the predictions are 100% correct has an AUC of 1.



implementation Of ROC And AUC

In [1]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

In [2]:

```
url="https://raw.githubusercontent.com/Suji04/Diabetes-Detection/master/diabetes.csv"
data = pd.read_csv(url)
data.head(5)
```

Out[2]:

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction
0	6	148	72	35	0	33.6	0.62
1	1	85	66	29	0	26.6	0.35
2	8	183	64	0	0	23.3	0.67
3	1	89	66	23	94	28.1	0.16
4	0	137	40	35	168	43.1	2.28

In [3]:

```
X = data.iloc[:,0:-1].values
y = data.iloc[:, -1].values
```

Train-Test Split

In [4]:

```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.20, random_state =
```

Feature Sciling

In [5]:

```
from sklearn.preprocessing import StandardScaler
sc_X = StandardScaler()
X_train = sc_X.fit_transform(X_train)
X_test = sc_X.transform(X_test)
```

Classifier Model

A. Logistic Regrssion

In [6]:

```
from sklearn.linear_model import LogisticRegression
model_logistic = LogisticRegression(C=2)
model_logistic.fit(X_train, y_train)

y_pred_logistic = model_logistic.decision_function(X_test)
```

B. DecisionTreeClassifier

In [7]:

```
from sklearn.tree import DecisionTreeClassifier
dt = DecisionTreeClassifier(max_depth=50)
dt.fit(X_train,y_train)

y_pred_dt = dt.predict_proba(X_test)
```

B. RandomForestClassifier

In [8]:

```
from sklearn.ensemble import RandomForestClassifier
rfr = RandomForestClassifier(n_estimators=100)
rfr.fit(X_train,y_train)

y_pred_rfr = rfr.predict_proba(X_test)
```

Find Roc And Auc Score Of These three model

In [9]:

```
from sklearn.metrics import roc_curve, auc

logistic_fpr, logistic_tpr, threshold = roc_curve(y_test, y_pred_logistic)
auc_logistic = auc(logistic_fpr, logistic_tpr)

dt_fpr, dt_tpr, threshold = roc_curve(y_test, y_pred_dt[:,1:2])
auc_dt = auc(dt_fpr, dt_tpr)

rfr_fpr, rfr_tpr, threshold = roc_curve(y_test, y_pred_rfr[:,1:2])
auc_rfr = auc(rfr_fpr, rfr_tpr)

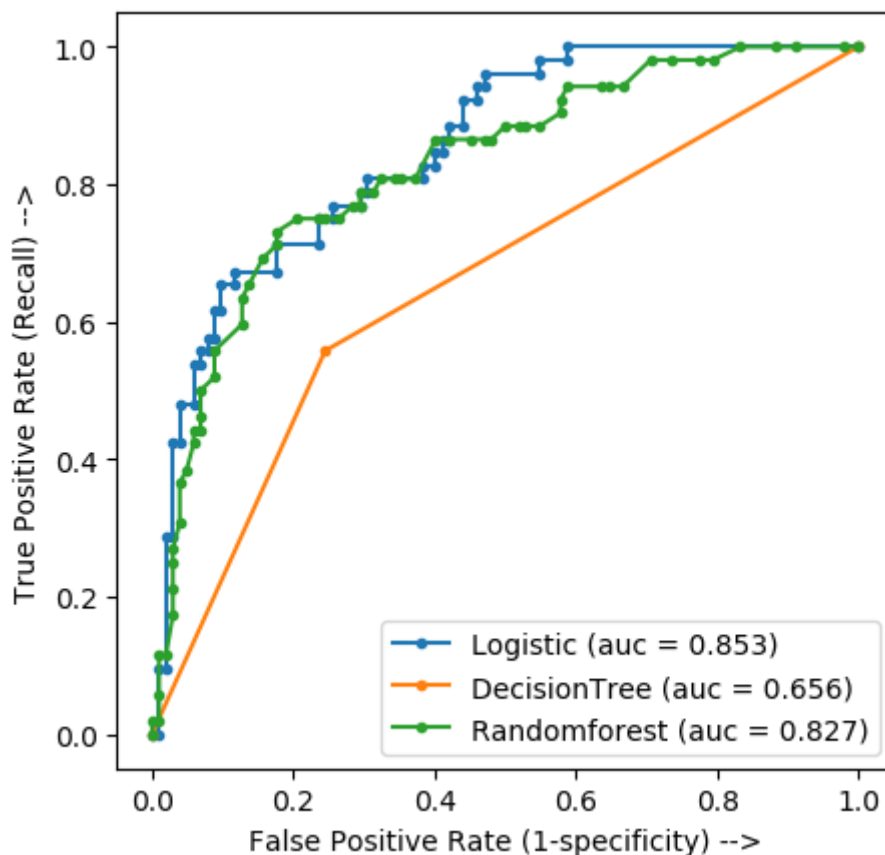
# svm_fpr, svm_tpr, threshold = roc_curve(y_test, y_pred_svm)
# auc_svm = auc(svm_fpr, svm_tpr)

plt.figure(figsize=(5, 5), dpi=100)
# plt.plot(svm_fpr, svm_tpr, linestyle='--', label='SVM (auc = %0.3f)' % auc_svm)
plt.plot(logistic_fpr, logistic_tpr, marker='.', label='Logistic (auc = %0.3f)' % auc_logistic)
plt.plot(dt_fpr, dt_tpr, marker='.', label='DecisionTree (auc = %0.3f)' % auc_dt)
plt.plot(rfr_fpr, rfr_tpr, marker='.', label='Randomforest (auc = %0.3f)' % auc_rfr)

plt.xlabel('False Positive Rate (1-specificity) -->')
plt.ylabel('True Positive Rate (Recall) -->')

plt.legend()

plt.show()
```



- Here you see that 3 model has different Auc score. so here we consider which has higher auc score