



Published in DataDrivenInvestor



Namratesh Shrivastav

Follow

Jan 18, 2020 · 3 min read · Listen



## Confusion Matrix(TPR,FPR,FNR,TNR), Precision, Recall, F1-Score

Confusion matrix is used to measure the performance of the classification model. Checking our model performance by accuracy sometimes it's misleading when we have imbalanced data. You can read more about accuracy [here](#).

So what is a Confusion matrix?

It is performance metrics to measure classification models where output is binary or multiclass. It has a table of 4 different combinations.



124

3

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

There are two things to noticed in the above image

- Predicted values- Values that are predicted by the model.
- Actual Value- Values that are actually in a dataset.

### Machine Learning in Finance | Data Driven Investor

Before we cover some Machine Learning finance applications, let's first understand what Machine Learning is. Machine...

[www.datadriveninvestor.com](http://www.datadriveninvestor.com)

Here, we are taking binary classification for understanding the model. Positive points belong to a positive class and Negative points to negative class. So it can be understood by these 4 points.

1. **True Positive(TP):** Values that are actually *positive* and predicted *positive*.
2. **False Positive(FP):** Values that are actually *negative* but predicted to *positive*.
3. **False Negative(FN):** Values that are actually *positive* but predicted to *negative*.
4. **True Negative (TN):** Values that are actually *negative* and predicted to *negative*.

Rate is a measure factor in a confusion matrix. It has also 4 type TPR, FPR, TNR, FNR

True Positive Rate(TPR): True Positive/positive

False Positive Rate(FPR): False Positive /Negative

False Negative Rate(FNR): False Negative/Positive

True Negative Rate(TNR): True Negative/Negative

For better performance, *TPR*, *TNR* should be high *and FNR*, *FPR* should be low.

Suppose we have 100 n points and our model's confusion matrix look like this.

		Actual Values	
		0	1
Predicted Values	0	850 TN	6 FN
	1	50 FP	94 TP

Now,

$$\text{TPR} = \text{TP}/P = 94/100 = 94\%$$

$$\text{TNR} = \text{TN}/N = 850/900 = 94.4\%$$

$$\text{FPR} = \text{FP}/\text{N} = 50/900 = 5.5\%$$

$$\text{FNR} = \text{FN}/\text{p} = 6/100 = 6\%$$

Here, TPR, TNR is high and FPR, FNR is low. So our model is not in underfit or overfit.

### **Precision**

It is used in information retrieval, pattern recognition. *Precision is all the points that are declared to be positive but what percentage of them are actually positive.*

$$\text{Precision} = \text{True Positive} / \text{Predicted Positive}$$

### **Recall**

It is all the points that are actually positive but what percentage declared positive.

$$\text{Recall} = \text{True Positive} / \text{Actual Positive}$$

### **F1-Score**

It is used to measure test accuracy. It is a weighted average of the precision and recall. When F1 score is 1 it's best and on 0 it's worst.

$$F1 = 2 * (\text{precision} * \text{recall}) / (\text{precision} + \text{recall})$$

- Precision and Recall should always be high.

#### References:

**sklearn.metrics.f1\_score - scikit-learn 0.22.1 documentation**

scikit-learn: machine learning in Python

scikit-learn.org

[https://en.wikipedia.org/wiki/Precision\\_and\\_recall](https://en.wikipedia.org/wiki/Precision_and_recall)

#### **Confusion matrix**

In the field of machine learning and specifically the problem of statistical classification, a confusion matrix, also...

en.wikipedia.org



# *Data Driven Investor*

## Gain Access to Expert Views

Give me access!



I agree to leave [medium.datadriveninvestor.com](https://medium.com/datadriveninvestor) and submit this information, which will be collected and used according to [Upscribe's privacy policy](#).

6K signups

Accuracy

Machine Learning

Performance

Mathematics

Model

[About](#) [Help](#) [Terms](#) [Privacy](#)