Search Medium

Write   Sign up   Sign In

Accuracy, Precision, Recall, F-1 Score, Confusion Matrix, and AUC-ROC

Ritesh Gupta · Follow
5 min read · Apr 5

121   3

In machine learning, there are various evaluation metrics used to measure the performance of a model. Among these metrics, Accuracy, Precision, Recall, F-1 Score, Confusion Matrix, and AUC-ROC are some of the most commonly used ones. In this article, we will explore each of these metrics in detail and provide an example to help understand their significance.

Photo by Pavel Danilyuk

Accuracy:

Accuracy is the most basic evaluation metric, which measures the percentage of correct predictions made by a model. It is calculated by dividing the number of correct predictions by the total number of predictions made by the model.

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{All Samples}}$$

For example, consider a binary classification problem where we have 100 samples in total, out of which 80 are correctly classified by the model. Then the accuracy of the model would be:

Accuracy = (number of correct predictions / total number of predictions) * 100 Accuracy = (80 / 100) * 100 = 80%

However, accuracy can sometimes be misleading if the dataset is imbalanced, i.e., one class has significantly more samples than the other. In such cases, the model may predict the majority class more frequently, resulting in high accuracy, but poor performance on the minority class.

Precision:

Precision measures the proportion of true positive predictions among all the positive predictions made by the model. It is calculated by dividing the number of true positives by the sum of true positives and false positives.

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

Precision = true positives / (true positives + false positives)

For example, consider a spam detection system where the model classifies an email as spam. If the email is genuinely spam, then it is considered as a

true positive. However, if the email is not spam but is classified as spam by the model, then it is considered as a false positive. In this case, the precision of the model would be:

Precision = true positives / (true positives + false positives)

Recall:

Recall, also known as sensitivity, measures the proportion of true positive predictions among all the actual positive samples in the dataset. It is calculated by dividing the number of true positives by the sum of true positives and false negatives.

$$Recall = \frac{True\ Positive(TP)}{True\ Positive(TP) + False\ Negative(FN)}$$

Recall = true positives / (true positives + false negatives)

For example, consider a medical diagnosis system where the model classifies a patient as having a disease. If the patient indeed has the disease, then it is considered a true positive. However, if the patient does not have the disease but is classified as having the disease by the model, then it is considered a false positive. In this case, the recall of the model would be:

Recall = true positives / (true positives + false negatives)

F-1 Score:

F-1 Score is a weighted average of Precision and Recall, where the weights are equal. It is used to balance the trade-off between precision and recall.

$$F1\ Score = 2 \times \frac{recall \times precision}{recall + precision}$$

F-1 Score is calculated as:

F-1 Score = 2 * ((precision * recall) / (precision + recall))

For example, if a model has high precision but low recall, it means that it makes fewer false positives but misses a lot of true positives. In contrast, a model with high recall but low precision makes more false positives but captures more true positives. In such cases, the F-1 Score can help us determine which model is better.

Confusion Matrix:

A confusion matrix is a table that shows the number of true positives, true negatives, false positives, and false negatives predicted by a model. It is used to evaluate the performance of a model on a binary classification problem.

For example, consider a binary classification problem where we have 100 samples, out of which 50 are positive and 50 are negative. A confusion matrix for this problem would look like this:

		Predicted	
		Negative (N) -	Positive (P) +
Actual	Negative -	True Negatives (TN)	False Positives (FP) Type I error
	Positive +	False Negatives (FN) Type II error	True Positives (TP)

In this matrix, the rows represent the actual classes, and the columns represent the predicted classes. The diagonal elements of the matrix (TP and TN) represent the correct predictions, while the off-diagonal elements (FP and FN) represent the incorrect predictions made by the model.

AUC-ROC:

AUC-ROC stands for Area Under the Curve — Receiver Operating Characteristic. It is a metric used to evaluate the performance of a binary classification model based on its ability to distinguish between positive and negative samples.
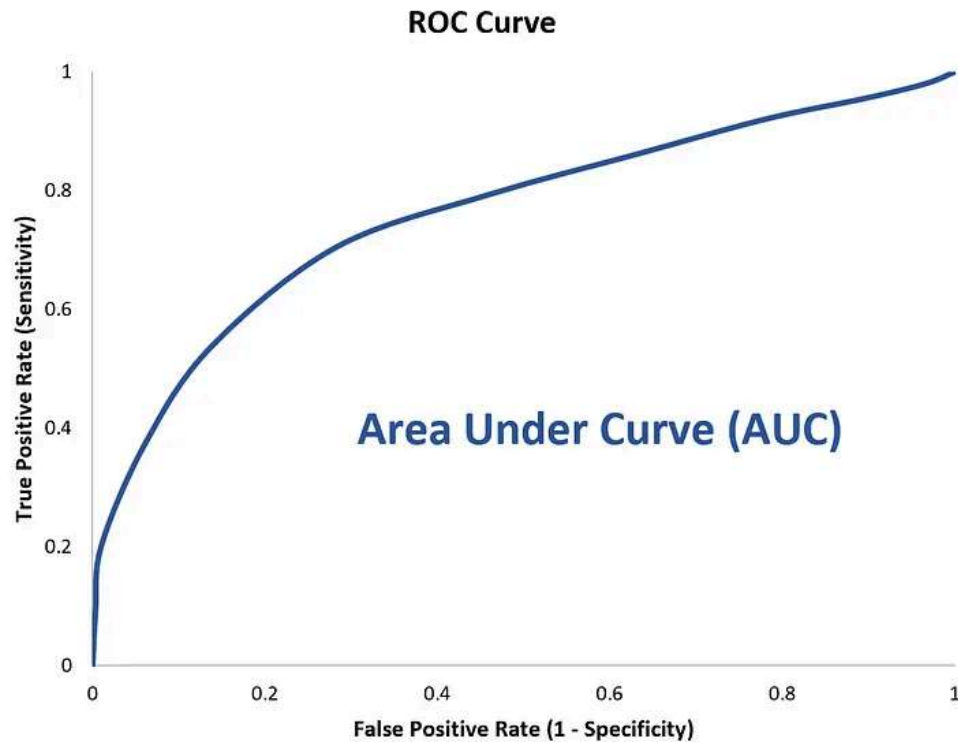
ROC Curve

The ROC curve is a plot of the True Positive Rate (TPR) against the False Positive Rate (FPR) for different classification thresholds. TPR is the same as Recall, while FPR is calculated as the ratio of false positives to the total number of negative samples.

The AUC-ROC score is the area under the ROC curve and ranges between 0 and 1, with 1 representing a perfect classifier and 0.5 representing a random guess.

For example, consider a binary classification problem where we have 100 samples, out of which 50 are positive and 50 are negative. A model with a high AUC-ROC score correctly predicts more positive samples and fewer negative samples than a model with a low AUC-ROC score.

In summary, the evaluation metrics discussed in this article are crucial for assessing the performance of a machine learning model. Accuracy, Precision, Recall, and F-1 Score are commonly used for binary classification problems, while the Confusion Matrix and AUC-ROC are used to gain a more in-depth understanding of the model's performance. Understanding these metrics and their significance is vital for building robust and reliable machine-learning models.

Thanks for Reading!

If you enjoyed this, <u>follow</u> me to never miss another article on data science guides, tricks and tips, life lessons, and more!

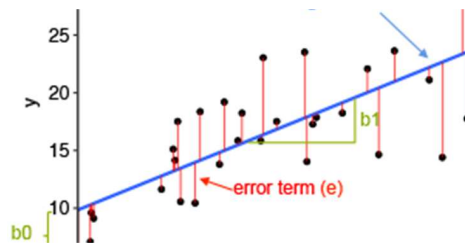Data Science   Machine Learning   Deep Learning   Artificial Intelligence   Python

121   3

Written by Ritesh Gupta

1K Followers

Follow

Data Scientist, I write Article on Machine Learning| Deep Learning| NLP | Open CV |
AI Lover ❤️

More from Ritesh Gupta

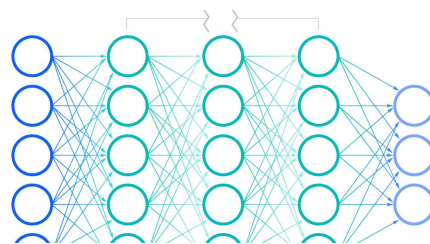Ritesh Gupta

10 Most Common Machine Learning Algorithms Explained...

1. Linear Regression

20 min read · Jan 19

1K   21

Ritesh Gupta

Mastering the Deep Learning Interview: Top 35 Questions and...

What is Deep Learning?

13 min read · Jan 23

417   9
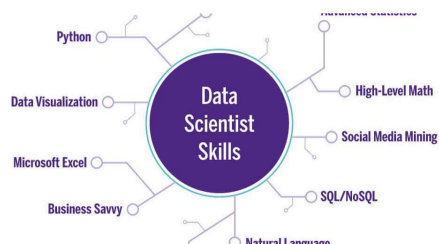
Ritesh Gup... in Artificial Intelligence in Plain Engli...

Ritesh Gupta

Master Data Science with This Comprehensive Cheat Sheet

Comprehensive Cheat Sheet for Data Science: Numpy, Pandas, Python, R, ML, DL,...

6 min read · Jan 30

10 Automated EDA Tools That Will Save You Hours Of Work

Exploratory Data Analysis (EDA) is the process of analyzing and summarizing the...
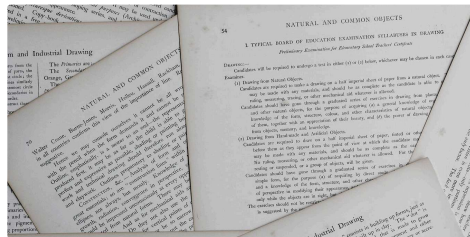
7 min read · Jan 29

(See all from Ritesh Gupta)

Recommended from Medium

Youssef Hosni in Towards AI

How to Read Machine Learning Papers Effectively

The field of machine and deep learning is evolving very fast, and there are new researc...

✦ · 10 min read · Oct 9, 2022

Peter Karas in Artificial Intelligence in Plain English

Logistic Regression in Depth

Logistic regression, activation function, derivation, math
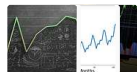
✦ · 7 min read · Jan 31

Lists

Predictive Modeling w/ Python
18 stories · 98 saves

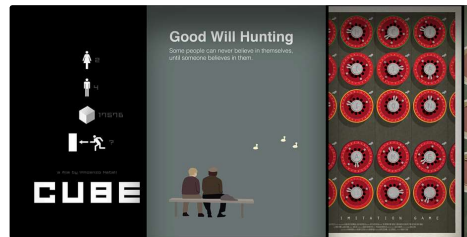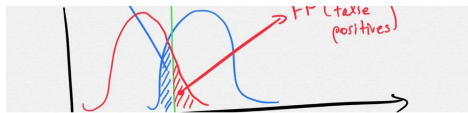### Natural Language Processing
384 stories · 40 saves

Practical Guides to Machine Learning
10 stories · 113 saves

ChatGPT
21 stories · 39 saves

Serafeim Loukas, PhD in Towards AI

How To Estimate FP, FN, TP, TN, TPR, TNR, FPR, FNR & Accuracy...

In this post, I explain how someone can read a confusion matrix and how to extract several...

✨ · 6 min read · Feb 5

👏 65 ○

Ali

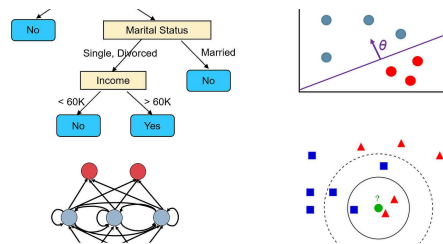### 16 Beautiful Math Movies That Everyone Should Watch at Least...

The existential philosopher Merleau Ponty says, "Film is not thought, it is perceived."

✨ · 16 min read · Jul 5

👏 1.3K ○ 10

Dr. Roi Yehoshua in Towards AI

Which ML Algorithm to Choose?

One of the key decisions you need to make when solving a data science problem is whic...

✨ · 6 min read · Mar 3

👏 234 ○ 3

Peter Karas in Artificial Intelligence in Plain English

L1 (Lasso) and L2 (Ridge) regularizations in logistic...

Logistic regression , Lasso and Rigde regularizations, derivations, math

✨ · 6 min read · Feb 2

👏 60 ○ 2

See more recommendations