

- [SEM](#)
- [SOCIAL MEDIA](#)
- [SOFTWARE DEVELOPMENT](#)
- [VIRTUALIZATION](#)

• Search for:

## • FIND YOUR INTERESTS

- [Blog Homepage](#) (118)
- [Business Consulting](#) (13)
- [Business Intelligence](#) (5)
- [Client Management](#) (3)
- [Creative Services](#) (9)
- [Data](#) (30)
- [Events](#) (6)
- [Exsilio](#) (18)
- [Exsilio Culture](#) (17)
- [GIS](#) (1)
- [HTML5](#) (3)
- [Marketing](#) (19)
- [News](#) (8)
- [Project Management](#) (16)
- [Search](#) (9)
- [SEM](#) (7)
- [Social Media](#) (14)
- [Software Development](#) (18)
- [SQL](#) (6)
- [Uncategorized](#) (1)

---

© [Exsilio Solutions](#) 2013

All rights reserved.

# What's on our mind?

Exsilio Solutions is proud to introduce our new blog! Not only is this forum completely integrated with all our social networks, but it is also 100% responsive. Now, you can take Exsilio with you on your phone, tablet, and desktop - redefine what you thought possible!

# Accuracy, Precision, Recall & F1 Score: Interpretation of

# Performance Measures

How to evaluate the performance of a model in Azure ML and understanding “Confusion Metrics”

Post September 9, 2016 / [111 likes](#) / [24 comments](#) / Exsilio Solutions

This blog demonstrates how to evaluate the performance of a model via Accuracy, Precision, Recall & F1 Score metrics in Azure ML and provides a brief explanation of the “Confusion Metrics”. In this experiment, I have used Two-class Boosted Decision Tree Algorithm and my goal is to predict the survival of the passengers on the Titanic.

Once you have built your model, the most important question that arises is how good is your model? So, evaluating your model is the most important task in the data science project which delineates how good your predictions are.

The following figure shows the results of the model that I built for the project I worked on during my internship program at Exsilio Consulting this summer.

Two-Class Boosted Decision Tree > Evaluate Model > Evaluation results

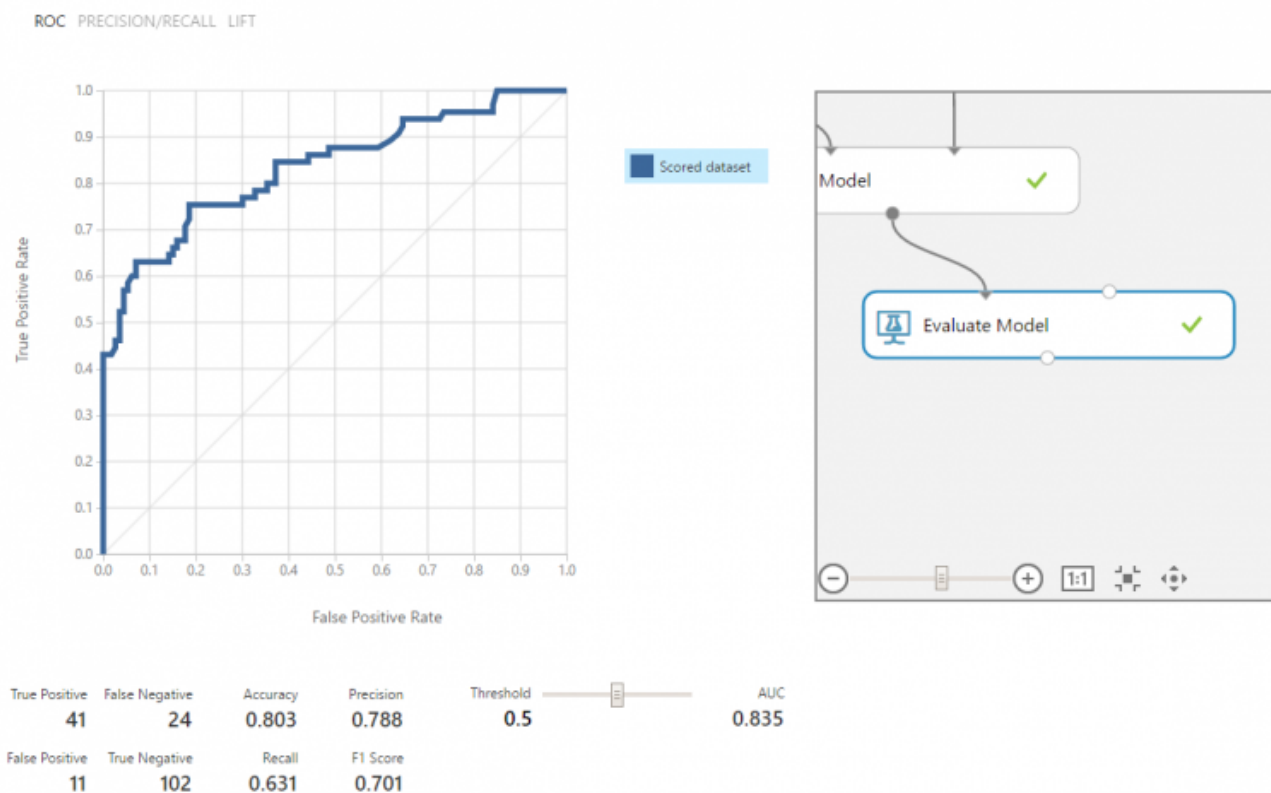


Fig. Evaluation results for classification model

Let's dig deep into all the parameters shown in the figure above.

The first thing you will see here is ROC curve and we can determine whether our ROC curve is good or not by looking at AUC (Area Under the Curve) and other parameters which are also called as Confusion Metrics. A

confusion matrix is a table that is often used to describe the performance of a classification model on a set of test data for which the true values are known. All the measures except AUC can be calculated by using left most four parameters. So, let's talk about those four parameters first.

	Predicted class		
Actual Class		Class = Yes	Class = No
	Class = Yes	True Positive	False Negative
	Class = No	False Positive	True Negative

True positive and true negatives are the observations that are correctly predicted and therefore shown in green. We want to minimize false positives and false negatives so they are shown in red color. These terms are a bit confusing. So let's take each term one by one and understand it fully.

**True Positives (TP)** - These are the correctly predicted positive values which means that the value of actual class is yes and the value of predicted class is also yes. E.g. if actual class value indicates that this passenger survived and predicted class tells you the same thing.

**True Negatives (TN)** - These are the correctly predicted negative values which means that the value of actual class is no and value of predicted class is also no. E.g. if actual class says this passenger did not survive and predicted class tells you the same thing.

False positives and false negatives, these values occur when your actual class contradicts with the predicted class.

**False Positives (FP)** – When actual class is no and predicted class is yes. E.g. if actual class says this passenger did not survive but predicted class tells you that this passenger will survive.

**False Negatives (FN)** – When actual class is yes but predicted class in no. E.g. if actual class value indicates that this passenger survived and predicted class tells you that passenger will die.

Once you understand these four parameters then we can calculate Accuracy, Precision, Recall and F1 score.

**Accuracy** - Accuracy is the most intuitive performance measure and it is simply a ratio of correctly predicted observation to the total observations. One may think that, if we have high accuracy then our model is best. Yes, accuracy is a great measure but only when you have symmetric datasets where values of false positive and false negatives are almost same. Therefore, you have to look at other parameters to evaluate the performance of your model. For our model, we have got 0.803 which means our model is approx. 80% accurate.

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN}$$

**Precision** - Precision is the ratio of correctly predicted positive observations to the total predicted positive observations. The question that this metric answer is of all passengers that labeled as survived, how many actually

survived? High precision relates to the low false positive rate. We have got 0.788 precision which is pretty good.

Precision =  $TP / (TP + FP)$

**Recall** (Sensitivity) - Recall is the ratio of correctly predicted positive observations to the all observations in actual class - yes. The question recall answers is: Of all the passengers that truly survived, how many did we label? We have got recall of 0.631 which is good for this model as it's above 0.5.

Recall =  $TP / (TP + FN)$

**F1 score** - F1 Score is the weighted average of Precision and Recall. Therefore, this score takes both false positives and false negatives into account. Intuitively it is not as easy to understand as accuracy, but F1 is usually more useful than accuracy, especially if you have an uneven class distribution. Accuracy works best if false positives and false negatives have similar cost. If the cost of false positives and false negatives are very different, it's better to look at both Precision and Recall. In our case, F1 score is 0.701.

F1 Score =  $2 * (Recall * Precision) / (Recall + Precision)$

So, whenever you build a model, this article should help you to figure out what these parameters mean and how good your model has performed.

I hope you found this blog useful. Please leave comments or send me an email if you think I missed any important details or if you have any other questions or feedback about this topic.

**\*\*Please Note** that the above results and analysis of numbers is based on the Titanic model. Your numbers and results may vary upon which model you work on and your specific business use case.

Related videos: <https://www.youtube.com/channel/UC9jOb7yEfGwxjjdpWMjmKJA>

By: Renuka Joshi (Intern at Exsilio)



#### RELATED POSTS

- [Common Data Warehouse Problems and How to Fix Them](#)

October 16, 2018

Learn about common issues in a data warehouse and the approaches you can use to resolve them....

- [How to use Change Data Capture](#)

July 16, 2018

Here's a step-by-step tutorial on using the Change Data Capture function....

- [Ready Hacker One: A Hack-a-thon hosted by Exsilio Solutions](#)

July 16, 2018

The recap of our 1st international hack-a-thon event! See all the highlights of the event from each office