

① Accuracy, Precision, Recall & F1 Score: Interpretation of Performance Measures

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

September 9, 2016 / 28 likes / 8 comments / Renuka Joshi

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

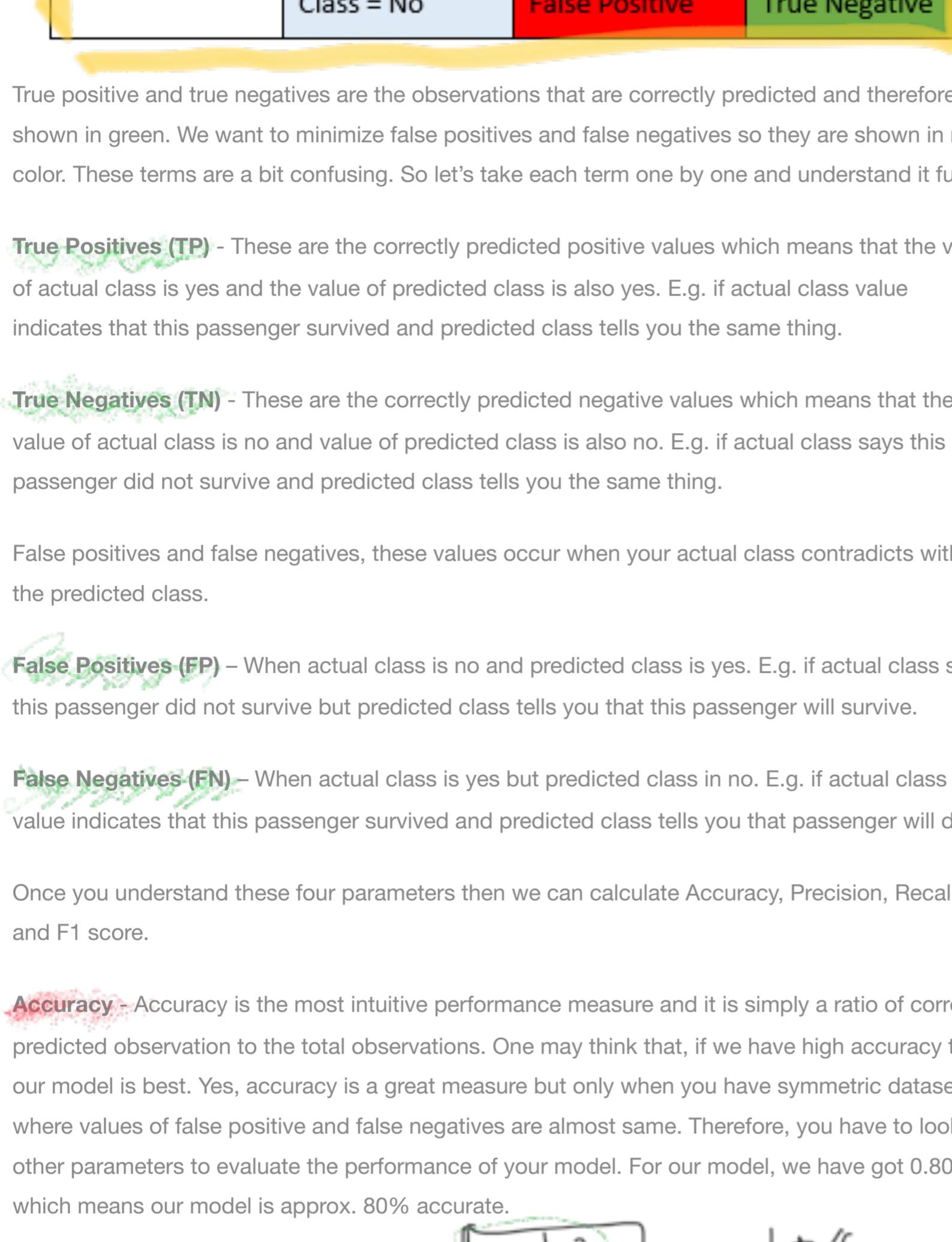


Fig. Evaluation results for classification model

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.

Confusion Matrix

		Predicted class	
Actual Class			
	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 is 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.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} = \frac{1+4}{1+2+3+4}$$

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.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} = \frac{1}{1+3}$$

Recall - 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.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{1}{1+2}$$

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.

$$\text{F1 Score} = 2 * (\text{Recall} * \text{Precision}) / (\text{Recall} + \text{Precision})$$

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)

Users feedback (8)

Faruk Ahmad says:

July 10, 2017 at 6:34 am

This post was very helpful. I have moved to the internet to understand the concept of precision and recall but all that I have got was tough for me to understand. But your explanation was very helpful. Thanks.

REPLY

Joseph says:

September 22, 2017 at 6:32 am

Crisp and clear! Thanks!

REPLY

Andres Luna says:

October 21, 2017 at 9:34 pm

Thank you so much for this post! You helped me a lot!

REPLY

Wayne says:

November 5, 2017 at 8:37 am

Great article! It help me a lot in understanding these.

REPLY

Arjun says:

November 15, 2017 at 7:58 am

Helpful post!

Simple and thought provoking explanations.

REPLY

Anonym says:

November 23, 2017 at 11:46 am

Thanks for this great article.

REPLY

Marcelo says:

January 23, 2018 at 4:26 am

Very clear and very helpful, Congratulations !!!

REPLY

Alan says:

September 26, 2018 at 8:50 am

Thanks for this article.

REPLY

Arjun says:

November 15, 2017 at 7:58 am

Helpful & Regards,

Arjun

REPLY

Anonym says:

November 23, 2017 at 11:46 am

Thanks for this great article.

REPLY

Marcelo says:

January 23, 2018 at 4:26 am

Very clear and very helpful, Congratulations !!!

REPLY

Alan says:

September 26, 2018 at 8:50 am

Thanks for this article.

REPLY

FIND YOUR INTERESTS

- Blog Homepage (109)
- Business Consulting (13)
- Business Intelligence (3)
- Client Management (3)
- Creative Services (6)
- Data (27)
- Events (6)
- Exsilio (16)
- Exsilio Culture (17)
- GIS (1)
- HTML5 (3)
- Marketing (17)
- News (8)
- Project Management (16)
- Search (9)
- SEM (7)
- Social Media (14)
- Software Development (15)
- SQL (6)

© Exsilio Solutions 2013
All rights reserved.



RELATED POSTS

Common Data Warehouse Problems and How to Fix Them **How to use Change Data Capture** **Ready Hacker One: A Hack-athon hosted by Exsilio Solutions**

October 16, 2018

July 16, 2018

July 16, 2018

SSIS Safari Adventure: How to Hack an XPath through the Occasional ETL Jungle

June 5, 2018

May 15, 2018

July 16, 2018

Employee Spotlight – Hyo Lee

June 5, 2018

May 15, 2018

July 16, 2018

Any thoughts?

* YOUR COMMENT

* Name

* E-mail

Website

* Captcha

REPLY

TAGS

#accuracy #azure ML #Azure ML Videos #confusion metrics #data #Data Science

#exsilio #exsilio solutions #F1Score #performance measures #precision #recall

#Two-class boosted Decision tree algorithm