Confusion Matrix “Un-confused”

Breaking down the confusion matrix

Kurtis Pykes Follow
Feb 16 · 5 min read ★

The goal of applied machine learning in industry is to drive business value. Therefore being able to evaluate your machine learning algorithms performance is extremely important for deriving insights into your model.

In this post, I aim to dive into the confusion matrix in a way that is accessible for those actively using them and for those who are just curious. With that being said, this article will be structured as follows:

1. What is a Confusion Matrix
2. How to interpret the Confusion Matrix

What is the Confusion Matrix?

The confusion matrix is a useful tool used for classification tasks in machine learning with the primary objective of visualizing the performance of a machine learning model.

In a binary classification setting where the negative class is 0 and the positive class is 1, the confusion matrix is constructed with a 2x2 grid table where the rows are the actual outputs of the data, and the columns are the predicted values from the model.

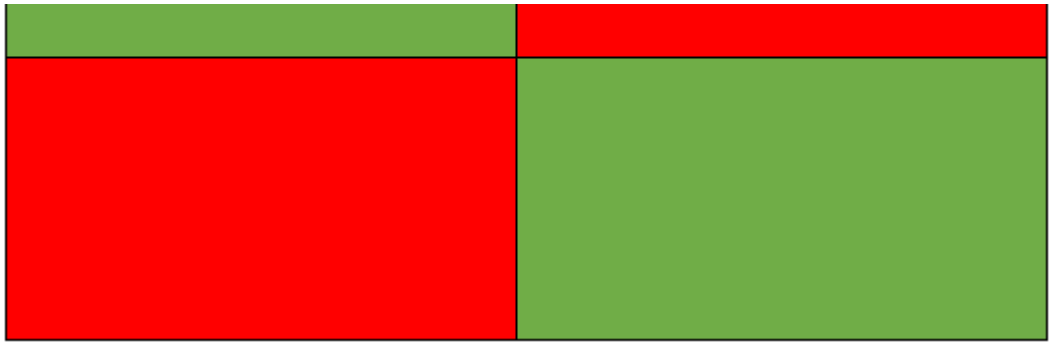

Figure 1: A visual representation of the confusion matrix (without annotations)

I will explain why certain grids are one color whilst the others have another, for now, this is just a visual representation of what the confusion matrix is and what it would look like for a binary classification task.

Note: Many sources structure the confusion matrix differently, for instance, the rows may be the predicted values and columns the actual values. If you are using a framework (i.e. Sci-kit learn) it is important to read the documentation for the confusion matrix to see what format the confusion matrix is in.

> *No matter what patch of the grid that you land on in* figure 1*, it's adjacent grid would be highlighted with another color. This takes into account the different variations of a confusion matrix that you may see if you simply rotate the grid.*

Figure 2: Variations of confusion matrix

Good! Now that you have become acquainted with the format of the confusion matrix for binary classification task, it is now important to know how to fill in a confusion matrix.

Let's use the classical cats and dogs example. If our classification task has been trained to distinguish between a cat and dog, the confusion matrix will create a summary of results for the prediction model for further analysis.

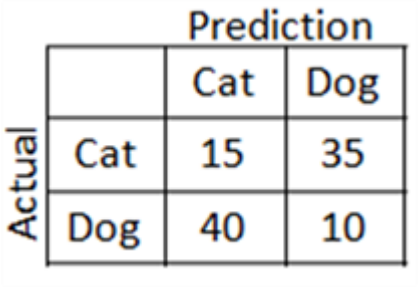

Figure 3: Confusion matrix results taken from Badgerati

Of the 50 cats in *figure 3,* the model predicted that 15 were actually cats and 45 were dogs. In respect to the dogs, the model predicted that 40 were cats and 10 were dogs. All the correct predictions are allocated in the diagonal in the table (from top to left) — yep, you guessed it! That is why the boxes were colored as they were in *figure 1*.

You've now learnt how to plot your own confusion matrix. As Data Scientist we want to know why the model is not doing a good job so that we have better insight into our data that we can use to engineer new features for the next modelling phase.

. . .

How to interpret a Confusion Matrix

I will start by first describing some terminology, if you are able to work out where they would go then you are halfway there:

True Positives: The model predicted positive and the label was actually positive.

True Negatives: The model predicted negative and the label was actually negative.

False Positives: The model predicted positive and the label was actually negative — I like to think of this as ***falsely classified as positive***.

False Negatives: The model predicted negative and the label was actually positive— I like to think of this as ***falsely classified as negative***.

Prediction

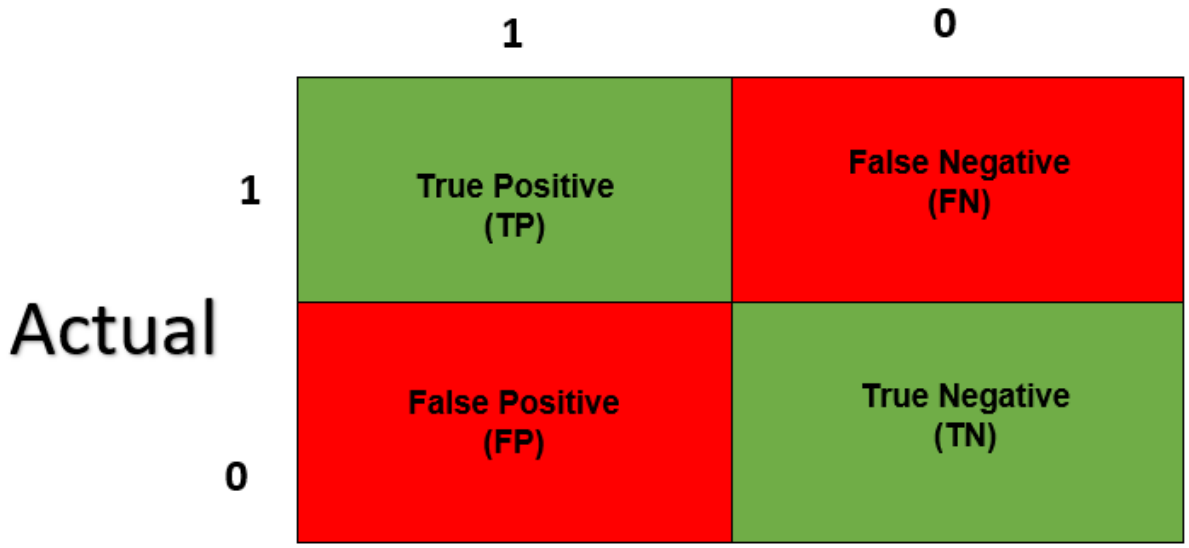

Figure 4: Annotated Confusion Matrix

If you were able to categorize the labels into these columns then that is fantastic, you now have a comprehensive grasp of the confusion matrix. Let's dig further!

Our confusion matrix has two red patches on our grid, which infer to us the type of errors that our model is making. Type I error is when our model falsely classify something as positive when it is actually a negative (False positive) — so it can be thought of as a false alarm — for example predicting that someone has asthma when they actually do not. Type II error is when we falsely classify something as negative when it is actually positive (False Negative) — In our asthma case, this will be saying that somebody does not have asthma when in actual fact they do.

I know the wordings can make this hard to grasp so if you feel like it has not sat well yet, return to figure 4 for a visual concept.

You have now learnt that a confusion matrix is a tool used to evaluate the performance of a predictive model in classification tasks. In addition, you also know the terminology associated with each output of our predictive model in accordance with its actual label.

In conclusion, there are further metrics that can be derived from our confusion matrix such as recall, precision, true positive rate and more. Having a firm understanding of the terminologies from the confusion matrix is a solid foundation to progress on to

understanding other metrics that can be derived from the confusion matrix. If you would like to push on, below are valuable sources that dive deeper into those aspects.

. . .

Understanding Confusion Matrix

When we get the data, after data cleaning, pre-processing and wrangling, the first step we do is to feed it to an...

towardsdatascience.com

https://en.wikipedia.org/wiki/Confusion_matrix