

ASSESSING CLASSIFICATION ACCURACY IN MACHINE LEARNING

		Actual Class	
		+	-
Predicted Class	+	True Positive	False Positive
	-	False Negative	True Negative

AN OVERVIEW

Classification Problems

In Machine Learning we will often take on tasks known as **Classification Problems...**

This is where we get an algorithm to learn the patterns in our data - with its output being a **prediction of a class or label**, so one or zero, yes or no, cat or dog, and so on...



VS.



These Classification Problems fall into an area known as **Supervised Learning** where we train the model or algorithm to learn the relationships between inputs & outputs based upon **known** classes or labels.

This means that when training the model, for each data-point, we have both a predicted label *and* an actual label.

Comparing the predictions against the actuals gives us a view of **how good** our model is!

DATA SCIENCE INFINITY

An example...

Let's say we were training a model to predict whether students would pass an exam or not. After training the model we have the following information...

How did our model do?

	Actual Result	Predicted Result
	Pass	Pass
	Pass	Fail
	Fail	Pass
	Fail	Fail
	Pass	Pass
	Fail	Fail

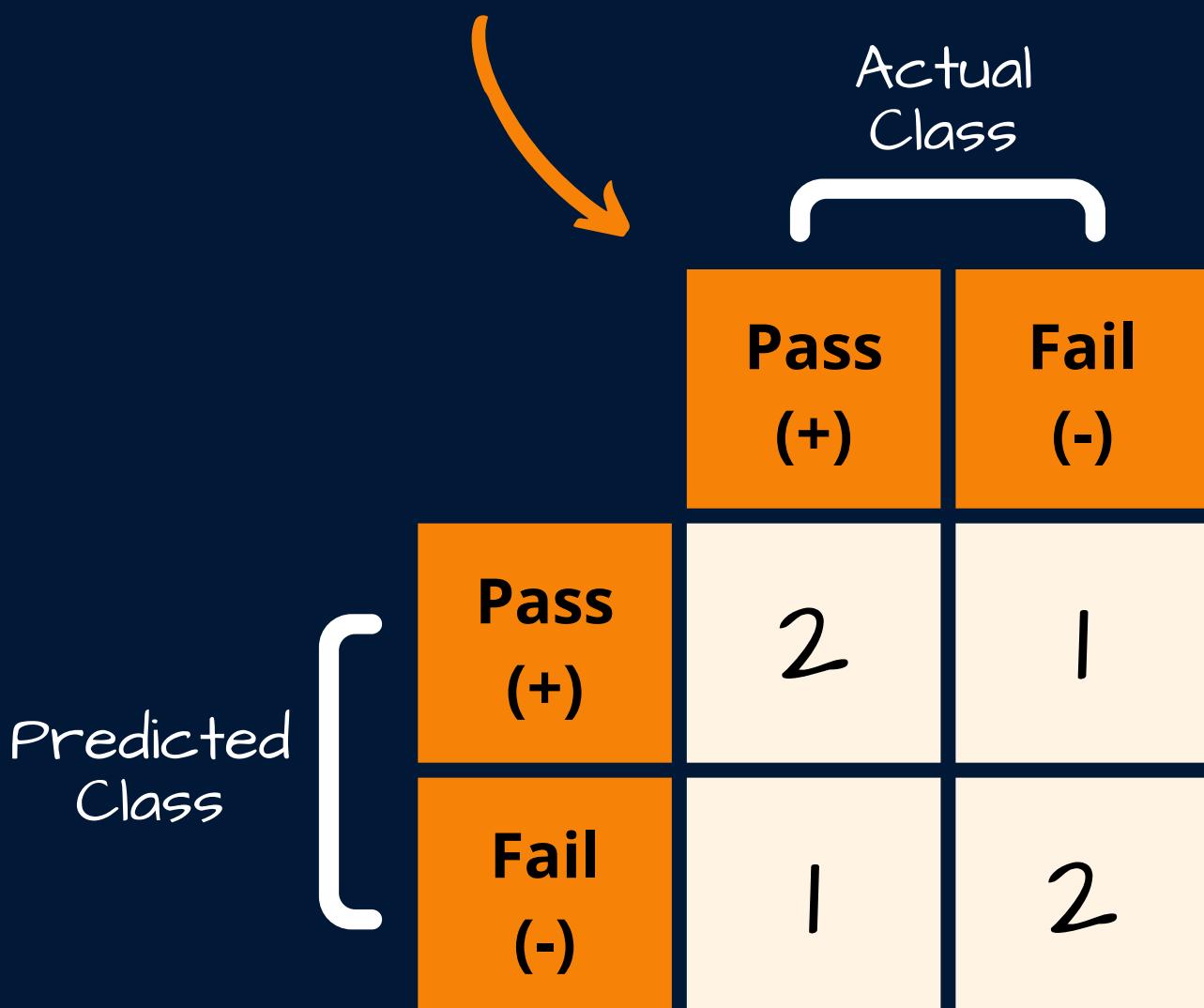
As you can see, sometimes our model predicts the result correctly, and sometimes it gets it wrong. Let's summarise this information down to assess our overall model performance!

DATA SCIENCE INFINITY

	Actual Result	Predicted Result
Person	Pass	Pass
Person	Pass	Fail
Person	Fail	Pass
Person	Fail	Fail
Person	Pass	Pass
Person	Fail	Fail

We can summarise our results into a 2x2 table known as a **Confusion Matrix!**

We simply count up the number of data-points by each class, and then split this by what their prediction was - partitioning everything into 4 segments...



Confusion Matrix...

The 4 segments of the Confusion Matrix have names that you might have heard of...

		Actual Class	
		+	-
Predicted Class	+	True Positive	False Positive
	-	False Negative	True Negative

True Positive

A **True Positive** is where we predict a positive result and are **correct** in our prediction

Example: Our model predicts a patient has a disease, and they actually do.

Actual Class			
Predicted Class	+	True Positive	False Positive
	-	False Negative	True Negative

False Positive

A **False Positive** is where we predict a positive result and are **incorrect** in our prediction. This is often called a **Type 1 Error**

Example: Our model predicts a patient has a disease, but in reality, they do not.

		Actual Class	
		+	-
Predicted Class	+	True Positive	False Positive
	-	False Negative	True Negative

False Negative

A **False Negative** is where we predict a negative result and are **incorrect** in our prediction. This is often called a **Type 2 Error**

Example: Our model predicts a patient does not have a disease, but in reality, they actually do.

		Actual Class	
		+	-
Predicted Class	+	True Positive	False Positive
	-	False Negative	True Negative

True Negative

A **True Negative** is where we predict a negative result and are **correct** in our prediction.

Example: Our model predicts a patient does not have a disease, and that prediction is correct, they do not actually have it.

		Actual Class	
		+	-
Predicted Class	+	True Positive	False Positive
	-	False Negative	True Negative

Classification Accuracy

Classification Accuracy is very intuitive, it simply measures the number of **correct classification predictions** (across both classes) out of **all predictions** made!

Back to our example...

	Actual Result	Predicted Result
	Pass	Pass
♂	Pass	Pass
♀	Pass	Fail
♂	Fail	Pass
♂	Fail	Fail
♀	Pass	Pass
♀	Fail	Fail



		Actual Result	
		Pass (+)	Fail (-)
Predicted Result	Pass (+)	2	1
	Fail (-)	1	2

Our model correctly predicted the results of 4 students, from the total of 6 - giving a **Classification Accuracy** of...

$$4 / 6 = \mathbf{66.6\%}$$



Correct	4
Incorrect	2
Total	6

But be careful...

While **Classification Accuracy** is very intuitive, there are scenarios where it can be misleading. It is important to know when this can happen, and what metrics are more appropriate!

Let us instead consider a scenario where we're building a Classification Model to predict a binary outcome of whether each of 1,000 patients have a **rare** medical disease, or not.

We train the model, and then assess the results using **Classification Accuracy** on our test set of 100 patients. We find that our model has **correctly classified 97 out of the 100** patients, giving a **Classification Accuracy of 97%**.

Classification Accuracy

$$97 / 100 = 97\%$$

Correct	97
Incorrect	3
Total	100

Wow - we are correctly predicting for 97% of patients!

Before we jump for joy - let's take a look at the Confusion Matrix that gave us this result...

DATA SCIENCE INFINITY

While the 98% Classification Accuracy is correct - we notice that this disease we are predicting for, is indeed **very rare!**

We can see from the Confusion Matrix that **only 2 patients out of our set of 100 actually do have the disease!**

Actual Result			
		Disease	No Disease
Predicted Result	Disease	1	2
	No Disease	1	96
		Correct	97
		Incorrect	3
		Total	100

Classification Accuracy = 97%

This is what we call **Imbalanced Data!**

Think about this - if we didn't use any fancy classification model, and we simply predicted that **everyone was disease free** - we'd actually get a higher accuracy rate 98%! Unfortunately, now our model doesn't look quite so amazing!

So what can we do? Well, there are a couple of supplementary metrics we can use over and above Classification Accuracy...

DATA SCIENCE INFINITY

Precision

Definition: Of all observations that were predicted, or classified as positive - what proportion actually were positive

In our example: Of all patients we predicted to have the disease - what proportion actually did

We can easily calculate this from our Confusion Matrix by taking the number of True Positives, and dividing that by the sum of True Positives & False Positives

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

Where TP = 1
FP = 2
Precision = $\frac{1}{(1+2)} = \frac{1}{3} = 33\%$

		Actual Result	
		Disease	No Disease
Predicted Result	Disease	1	2
	No Disease	1	96

TRUE POSITIVE: 1
FALSE POSITIVE: 2
FALSE NEGATIVE: 1
TRUE NEGATIVE: 96

Of all patients we *predicted* to have the disease, only 33% actually did! We would have patients thinking they were ill, when they were not!

This gives us a whole new way to think about **how good** our model really is!

DATA SCIENCE INFINITY

Recall

Definition: Of all positive observations, how many did we predict as positive

In our example: Of all patients who actually had the disease, how many did we correctly predict

We can easily calculate this from our Confusion Matrix by taking the number of True Positives, and dividing that by the sum of True Positives & False Negatives

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

With the data from the confusion matrix:

$$\text{Recall} = \frac{1}{(1 + 1)} = 0.5$$
$$\text{Recall} = \frac{1}{2} = 0.5$$

Or, 50%

		Actual Result	
		Disease	No Disease
Predicted Result	Disease	1	2
	No Disease	1	96

Annotations: TRUE POSITIVE (green), FALSE POSITIVE (red), FALSE NEGATIVE (red), TRUE NEGATIVE (green)

Of all patients who *actually* had the disease, our model only predicted 50% of them!

Again, for a disease prediction model, this potentially isn't great...

Precision vs. Recall

While we've seen that **Classification Accuracy** can be a good indicator of model performance - we should always dig deeper and analyse both **Precision & Recall** before getting too excited about our accuracy measure, especially in cases where the data is biased heavily towards one of the classes....

The tricky thing with these two, Precision & Recall, is that it's hard to optimise both - it's essentially a zero-sum game.

If you try to improve Precision, the Recall falls and vice-versa.



In some scenarios, it will make more sense to try and elevate one of them, in spite of the other.

In the case of the rare disease classification model like in our example, perhaps it's more important to improve Recall as we want to classify as many cases as possible, but then again, we don't want to just classify everyone as having the disease - as that isn't a great solution either...

So - there is one more metric we can use...

F1-Score

The **F1-Score** is the harmonic mean of **Precision & Recall**.

A good F1-Score comes when there is a **balance** between Precision & Recall, **rather than a disparity** between them.

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

In our example we had...

Precision = 0.33

Recall = 0.50

$$\Rightarrow \frac{2 * (0.33 * 0.5)}{(0.33 + 0.5)}$$

$$\Rightarrow \frac{0.33}{0.833}$$

$$\Rightarrow \mathbf{0.396}$$

We now have several metrics that give us a much better understanding of our predictions, and of our model!

Bonus! Precision & Recall Cheat-Sheet!

It can be hard to remember what Precision & Recall each mean, and which way around the go.

Below is a cheat-sheet with a high level overview of what values of each would mean!



Precision	Recall	Meaning
High	High	The model is differentiating between classes well
High	Low	The model is struggling to detect the class, but when it does it is very trustworthy
Low	High	The model is identifying most of the class, but is also incorrectly including a high number of data points from another class
Low	Low	The model is struggling to differentiate between classes



Do you want to **learn more** about this topic - and how to **apply it** in the real world?





Do you want to land an
incredible role in the
exciting, future-proof, and
lucrative field of **Data
Science**?



LEARN THE RIGHT SKILLS



A curriculum based on
input from **hundreds** of
leaders, hiring managers,
and recruiters



BUILD YOUR PORTFOLIO



Create your professionally made portfolio site that includes **10 pre-built** projects



EARN THE CERTIFICATION



Prove your skills with the
**DSI Data Science
Professional Certification**

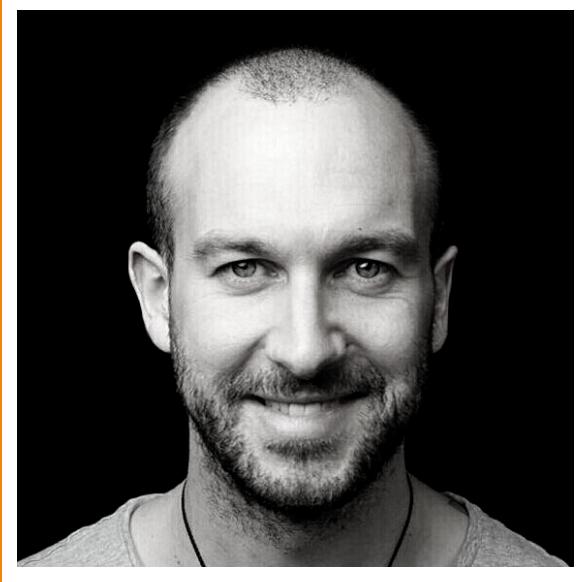


LAND AN AMAZING ROLE



Get guidance & support
based upon **hundreds of**
interviews at top tech
companies





Taught by former Amazon
& Sony PlayStation Data
Scientist **Andrew Jones**

What do DSI
students say?





"I had over 40 interviews without an offer. After **DSI** I quickly got **7 offers** including one at **KPMG** and my amazing new role at **Deloitte**!"

- Ritesh



"The **best program** I've been a part of, hands down"

- Christian



"DSI is **incredible** - everything is taught in such a clear and simple way, even the more complex concepts!"

- Arianna



"**I got it!** Thank you so much for all your advice & help with preparation - it truly gave me the **confidence** to go in and **land the job**!"

- Marta





"I've taken a number of Data Science courses, **and without doubt**, DSI is the best"

- William



"One of the best purchases towards learning I have ever made"

- Scott



"I learned more than on **any other course**, or reading entire books!"

- Erick



"I started a bootcamp last summer through a well respected University, **but I didn't learn half as much from them!**"

- GA





"100% worth it, it is amazing. I have never seen such a good course and I have done plenty of them!"

- Khatuna



"This is a world-class Data Science experience. I would recommend this course to every aspiring or professional Data Scientist"

- David



"Andrew's guidance with my Resume & throughout the interview process helped me land my amazing new role (**and at a much higher salary than I expected!**)"

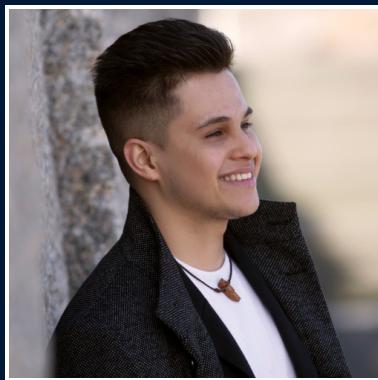
- Barun



"DSI is a **fantastic community & Andrew** is one of the best instructors!"

- Keith





"I'm now at University, and my Data Science related subjects are a **piece of cake** after completing this course!

I'm so glad I enrolled!"

- Jose



"In addition to the great content, Andrew's dedication to the growing DS1 community is **amazing**"

- Sophie



"The course has such high quality content - you get your ROI even from the first module"

- Donabel



"The Statistics 101 section was awesome! I have now started to get confidence in Statistics!"

- Shrikant





"I can't emphasise how good this programme is...well worth the investment!"

- Dejan



"I'd completed my Master's in Business Analytics, but DSI was the first time I felt I had a solid foundation in Data Science to go forward with"

- Scott

...come and join the
hundreds & hundreds of
other students getting the
results they want!



<https://data-science-infinity.teachable.com>