Notes from model:

Observe descriptive stats:

There’s about 105 participants with missing data for age, ie, with age listed as 0.

[[ Aside:

I may consider dropping these rows.

When these rows are removed the minimum changes to 14.

In any case the current range extends from around 0/14 to about 83.

Alternatively, I can replace the 0’s in age with the mean or median age. That’ll be a simpler solution.

To be enacted once my model is responding better.

{{Btw, this can be enacted even now. We can just lift the script and place it up top at a later stage}}

Best template to use for this is Lillian pierson’s titanic dataset data cleaning step

]]

Once again, the model performance is 72% accuracy, only because it predicts only 0’s.

This precision-recall graph is again quite a nice one.

May want to draw on it for future analyses.

The classification report, something we’ve seen before, again is a rehash, and a useful tool, clearly.

Here’s a different rendering of the confusion matrix from the template previous to this course.

Model evaluation