**Answers**

I built 2 models to predict if a patient has survived 2 years (Regression.ipynb & Classification.ipynb) , a new attribute “survived-2” is created where it is 1 if “survival” is greater or equal to 24, and it is 0 when “survival” less than 24.

1. To predict the “survival” length then see if it is 24 months+ or not

This model is a regression model (Random Forest Regression) where the value to predict is the “survival”. Once getting these predictions, I check if the value of “survived-2”, then compare the value of the newly generated “survived-2” with the actual “survived-2” to get the accuracy.

I choose this model as it supposes to find the small differences in survival periods, by that I mean a patient could has lived 23 month and get 0 for “survived-2” and another patient could has lived 24 months and get 1. Although the survival period is almost the same, yet the classification is different.

This model was supposed to pick these small differences and give better results

1. To predict “survived-2”

This model is a classification model (KNN), to predict “survived-2” is 1 for patients who lived more than 24 months and 0 for the ones who did not survive 24 months.

Comparing the results of both models I got these results

Random Forest Regression Accuracy = 77.5% 🡸🡸🡸🡸🡸🡸🡸🡸🡸 This one is better

KNN Accuracy = 74.6%

Files for predicting “survived-2” for “echocardiogram.test” file attached

I would add features such as patient gender, and patient weight.