Using ONE of the following sources, complete the questions for only that source.

Credit approval: <https://archive.ics.uci.edu/ml/datasets/Statlog+%28Australian+Credit+Approval%29>

Cardiac Arrhythmia: <https://archive.ics.uci.edu/ml/datasets/Arrhythmia>

Abalone age: <https://archive.ics.uci.edu/ml/datasets/Abalone> - this one is a bit harder since its not binary like the others, but if you really want to master these concepts, you should pick this one. Use RMSE as a performance metric if you do this as regression. You should target a value of under 2.

Note: at least one of your models should have the most relevant performance metric above .90 . All performance metrics should be above .75 . You will partially be graded on model performance.

1. Preprocess your dataset. Indicate which steps worked and which didn’t. Include your thoughts on why certain steps worked and certain steps didn’t.
2. Create a decision tree model tuned to the best of your abilities. Explain how you tuned it.
3. Create a random forest model tuned to the best of your abilities. Explain how you tuned it.
4. Create an xgboost model tuned to the best of your abilities. Explain how you tuned it.
5. Which model performed best? What is your performance metric? Why?