CS 4320: Machine Learning

Assignment: Decision Tree Classification

Use the [Heart Attack](#) data set at Kaggle. Create a decision tree classification model (such as `sklearn.tree.DecisionTreeClassifier`), to obtain the best `F1` score possible.

It is expected that you will use the Titantic decision tree source code as a starting point for your code development.

Create a report that includes the data exploration plots and analysis, the learning hyper parameters you tried, the `F1` score for these attempts, and the final hyper parameters and `F1` score found.

Include a discussion on the effects of the decision tree hyper parameters.

Report the best trained model's `F1` score on the test data.

Report on the suitability of your model for production on this system.

Required Steps

- Download your data.
- Explore and analyze your data.
- Split the data 80%/20%, for training/testing.
- Write (or modify) a Python program using sklearn to process and fit the training data to decision tree models with various hyper parameters.
- *AFTER* finding your best fit model, measure the model's `F1` score on your test data.
- Create a report with the contents mentioned above.
- Commit and push your code in the git repository.
- Submit the report (as PDF) to Canvas.