

CS 4320: Machine Learning

Assignment: Support Vector Classification

Use the [heart_failure_clinical_records_dataset](#) data set at Kaggle. You'll need to identify which features are categorical and which are numerical. Use hyper parameter search with cross validation to create a decision tree classification model and a support vector classification model to obtain the best $F1$ scores possible.

It is expected that you will use the Titanic hyper parameter search with cross validation decision tree source code as a starting point for your code development.

Create a report that includes the data exploration plots and analysis. The report will also include for each type of model (decision tree and svc) which hyper parameters were used in the search, the range or set of values used for each hyper parameter, the hyper parameters selected, the number of cross validation sets, the $F1$ cross-validation score obtained, the training $F1$ score of the model when trained on all training data, and finally, the $F1$ score of the model on the testing data.

Include a comparison of the cross validation and full training $F1$ scores between the two models, and which model you would select, based only on those scores. Finally, discuss whether the $F1$ scores on the testing data support your decision or not.

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 the training data and fit classification models with a hyper parameter search and cross validation.
- *AFTER* finding your best fit models, measure each 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.