**Capstone\_1**

This program includes the data inspection and preprocessing. All numerical features are standardized. Clean dataset is pickled.

**Capstone\_2**

The program reads the files that include all sample numbers of all incorrect predictions, for each classification model optimized in previous work. Each of the five files follow the following nomenclature: ( [classifier acronym]\_x\_wrong).

Calculates the intersection between the misclassifications of the Logistic Regression and the other four models: Decision Tree, Random Forest, SVC and XGBoost.

**Capstone\_3**

Unpickles the clean dataset, the Logistic Regression and Random Forest models optimized in previous work, the classes in test and the corresponding predictions of the two models.

Accuracy, ROC AUC and confusion matrix of the Logistic Regression and Random Forest models are calculated.

Two ensembling models (voting and stacking) are trained.

The staked model is globally and locally interpreted with SHAP.