1. Original one provided by Prof:A table with numbers and letters

   Description automatically generated
2. (Drop missing values) All variables after dropping high correlation variables:

A screenshot of a computer

Description automatically generated

1. (Drop missing values) Top 12 variables selected from feature importance after dropping high correlation variables:

A screenshot of a computer

Description automatically generated

1. (Impute missing values) Top 18 variables selected from feature importance after dropping high correlation variables A screenshot of a computer

   Description automatically generated

Highest AUROC/AUPRC : Dataset with dropping missing values + all variables + GB + (Upsample 30%)

1. Tried with 2 datasets with dropped/imputed missing values
2. Feature selection:
3. (['outcome\_inhospital\_mortality'] == 0) & (train['ed\_death'] == 0)
4. Correlation check
5. Extracted the **hour** of “intime\_ed”&”outtime\_ed” (e.g. extracted “16” from 2166-09-20 16:52:00)

? Whether need to bin the hour

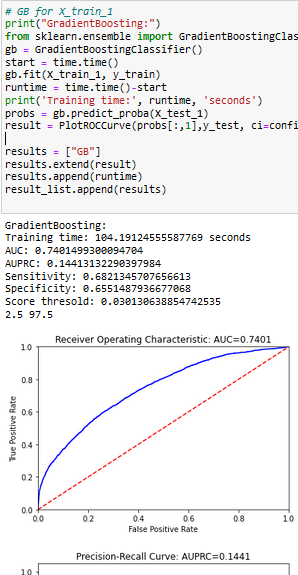
1. Decision tree: top important features
2. Modeling + XGBoost with all/top important features
3. Objective variable imbalance tackling:

Upsampled minor class (=1) to take 30% of whole data

Downsample

Min/max scale standardiza

Hyperparameter : default

A screen shot of a graph

Description automatically generatedA screen shot of a graph

Description automatically generated

1. Distribution of 18 features
2. Feature importance
3. Auprc & auroc