

Data set

The data set consist of features that could be associated to heart disease Like age , chest pain type and other features you can check data set and its documentation on

<https://www.kaggle.com/fedesoriano/heart-failure-prediction>

Summary of findings

I was able to build a model to predict heart disease with accuracy 86.5% , Fixing data outliers , Find most effective features on heart disease and proportions of features unique values and heart disease

Key insights

Focusing at first on balancing of data set and scattering plot of specific features then introducing proportions

Used references

<https://www.kaggle.com/andreshg/xgboost-optuna-hyperparameter-tunning>