Heart Disease analysis with classification

Authors: Lola Sõukand, Robin Mürk

Data: https://www.kaggle.com/datasets/alexteboul/heart-disease-health-indicators-dataset/discussion

Aim is to build a binary classifier that predicts a person's risk of a heart attack and to analyze individual attributes to determine what lowers and what increases the risk of heart attack using regression analysis.

Goal 1: train a model to predict whether a patient is at high risk or at low risk of a heart attack.
Goal 2: find factors, that lower the risk of heart attack
Goal 3: find factors, that increase the risk of heart attack.

Tasks

1. Prodect planning- Robin, Lola (4 h):
   Planning the project, setting up the goals and a roadmap of the project.
2. Data describing and first visualitsion – Lola (8 h):
   First data visualization, first comparisons and descriptions. Making first plots: historgrams, heatmaps etc using Seaborn,Pyplot etc.
3. Data cleaning - Lola (1 h)

   Going over the data and filtering or replacing any NaN or missing values. If there are huge, singular, outliers are found, they will be removed.

4. Training ja test splittid -Robin (2 h)

   Creating a main test and training split from the data. After that deviding the training data into training and validation data. For each model in the ensamble the train/validation split will be different.

5. Models (Goal 1) – Robin (20 h)

   Creating a classifier for predicting heart disease risk with an ensamble of model, using SVC, KNN, RandomForestClassifier. Doing hyperparameter tuning on all of them, and training each one with a different training/valitation

set. Effectivness measures will be AUC, accuracy, and percicion. Must check to not overfit the data to the models.

6. Visualistaion and finding correlations (Goal 2, 3) – Lola (12 h)

   Visualizing the results of the model. Finding attributes with positive moderate ($r>0.3$) and negative moderate ($r<-0.3$) correlation with heart disease. Visualizing the findings.

7. Summary -Robin, Lola (1-2 h)
   Creating a summary of the overall findings, summerizing the effectivness of the model and explaining the found correlations.
8. Making poster – Robin, Lola (2 h)
   Creaing a final poster for the presentation with neccesary info and illustrations.