



Heart Disease

Karl Gustav Gailit, Kertu-Carina Kallaste, Markus Kikkatalo

Introduction

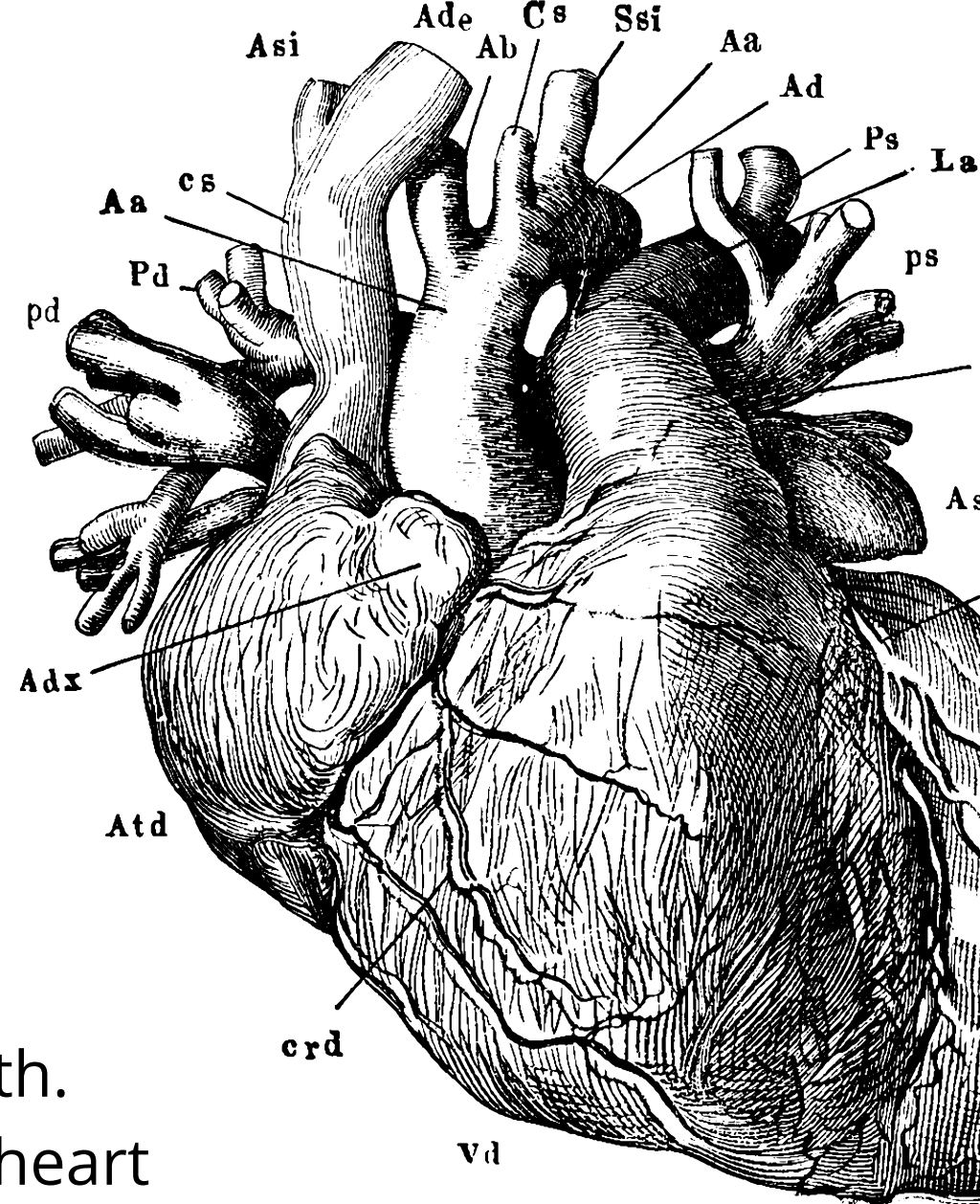
In Estonia, cardiovascular diseases are the leading cause of death. Therefore for our project, we decided to predict the risk of heart disease prior to negative outcomes like myocardial infarctions (heart attacks) taking place.

Data

The data used for this project was collected by the BRFSS in 2015, when 441,456 adults were interviewed. Originally, the dataset contained more than 300 columns, but we selected a subset of 22 features based on extensive research regarding heart disease risk factors and the BRFSS codebook report.

Methods

We trained multiple different models using several algorithms and sampling methods. The classifiers we used were Decision Tree, Random Forest, KNN, Gaussian Naive Bayes and AdaBoost.



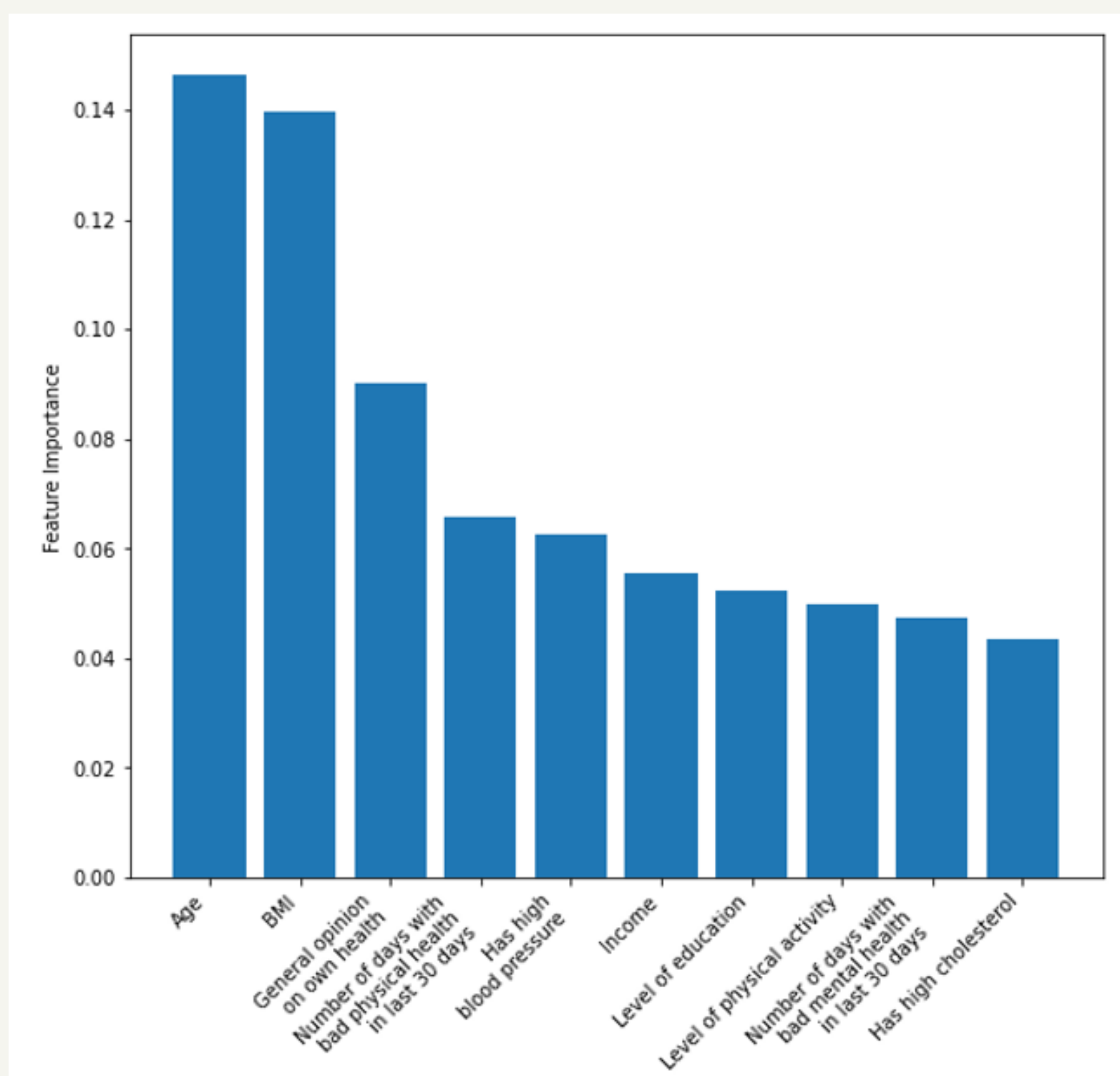
Results

Best model — RFC with oversampled data

F-score — 0.971

Precision — 0.945

Accuracy — 0.999



Features which had the highest correlation with heart disease