# MOTIVATION

What problem are we trying to solve?

According to the World Health Organization heart diseases, also known as Cardiovascular diseases (CVDs) represent the number one cause of death globally. With an estimated 17.9 million cases in 2016 they make up a staggering 31% of all global deaths. As deaths of CVDs can be prevented through appropriate treatment, we are trying to find a way to ensure, that high-risk patients can be identified.

Why do you choose this topic?

We chose this subject because we believe that a lot of the deaths would have been easily preventable, if patients would have received the necessary care. Using a statistical approach, we hope to ensure that patients with a high risk of CVDs can get identified, even before the first symptoms show. This is especially important, since the first warning of a cardiovascular disease is often already a heart attacks or a stroke.

What will be the potential impact of your project?

Our project aims to ensure, that people with a high risk of cardiovascular disease can get diagnosed and subsequentially receive the necessary treatment. As three quarters of the world’s deaths from CVDs occur in low- and middle-income countries we are also aiming to bridge a global equality gap and reduce the burden on the countries’ economies.

Current Practices

Currently there are two types of interventions to reduce the likelihood of cardiovascular diseases: population-wide and individual.  
Population wide interventions include measures, such as taxation on fatty and salty foods, tobacco and alcohol, building walking and cycle paths to encourage physical activity and providing healthy school meals to children. As these are general measures this paper will not cover them more in depth.  
Individual level practices … (see who print-out)

There have also been previous efforts to use a data-driven approach such as <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4981580/> (look up how to quote correctly), who have used a Naïve Base Classifier with an accuracy of up to 87.98% and a true positive recognition rate between 82.06% and 100% depending on the risk level of the patient.

What do you want to improve?

Our aim is to contribute to the research in the field of cardiovascular diseases and to contribute to the World Health Organisations “Global action plan for the prevention and control of NCDs 2013-2020”.

# METHODS

Data Source

For our project we are using the Heart Disease dataset from the UCI Machine Learning Repository. It consists of 303 individuals, with 165 cases of cardiovascular diseases and 136 members of the control group.

Variables of Interest

The dataset originally contained a total of 76 variables, many of which were deemed statistically irrelevant by previous researchers. As previous research has only focused on a smaller, complete subset of data, we planned on following the precedents and chose the same data-subset. It’s compromised of the following 14 variables:  
  
ADD VARIABLES, DESCRIPTIONS

Analytical Procedure

Our main objective was to find the model with the highest accuracy....

RESULTS

Sources

<https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)> (WHO on Heart disease (statistics, etc.)

<https://www.escardio.org/Journals/E-Journal-of-Cardiology-Practice/Volume-15/prevention-of-cardiovascular-disease-recent-achievements-and-remaining-challeng> (not used yet)

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4981580/> (predicting CVDs with a Bayesian classifier)