

MSc-thesis „Risk profiles of school children in sub-Saharan Africa: A machine learning approach”

Project description:

Behavior changes could prevent large components of the burden of disease globally and in all world regions, including sub-Saharan Africa. The major behavioral risks that impede population health and health-related human development are well understood. They include poor diet, consumption of sugar-sweetened beverages, failure to follow good hygiene practices, low physical activity, substance use, excessive alcohol consumption, smoking, and violence.

These behaviors drive the major biological and physical risk factors for ill health including hypertension, diabetes, infections and injuries. However, unlike biological and physical risk factors, which are often causally linked to each other, behavioral risk factors do not necessarily cluster in populations.

It is thus valuable to identify behavioral health risk profiles in different populations. A particularly interesting age group for risk factor analyses are youth, because it is in youth that many risk and protective behaviors are formed and become lifelong habits. Behavioral health risk profiles among youth could be useful both as ‘sensors’ for adult ill health and loss of life and as starting points for tailoring and targeting of public health and behavior change interventions.

However, in most countries and communities the behavioral health risk profiles are not known in youth or other age groups. The small extant literature on this topic is not based on rigorous statistical learning and data mining approaches.

Data:

For our analyses, we will use the Global School-based Student Health Survey (GSHS), which has been funded by the World Health Organization and the US Centers for Disease Control (CDC) and is publicly available online (<https://www.cdc.gov/gshs/>). The GSHS is available in over 130 countries and for several years. The GSHS contains data from students aged 13-17 years on 10 major behavioral health risk factor categories, including tobacco use, alcohol use, drug use, dietary behaviors, hygiene, physical activity, sexual risk behaviors, and violence and injuries.

Job tasks:

- support and conduct analysis of GSHS dataset, including data cleaning – we will mainly focus on dimensionality reduction, data clustering and classification
- support writing of report based on this research – we aim to publish results in a scientific, peer-reviewed journal

Possibilities:

- possibility to:
 - o be involved hands-on in Global Health research

- get to work in a flexible, interdisciplinary and international team
- be a first author or co-author of a scientific, peer-reviewed publication
- receive individual tutoring from Post-Doc
- be involved in a project that will contribute to better understand and shape interventions to improve health in sub-Saharan Africa

Requirements:

- very good command of English
- currently enrolled as student
- good understanding of machine learning methods (classification, clustering) and their application (source GSHS data is ordinal)
- knowledge of Python scientific programming (in particular libraries: numpy, scipy, sklearn, pandas, matplotlib) and R
- basic statistical knowledge

Please send a short motivation letter and CV to Dr. Sandra Barteit: barteit@uni-heidelberg.de