UPPSALA
UNIVERSITET

Data Mining

Project Proposal

Group 9

Teng-Sung Yu, Seppe Maes

Research Problem

In this project, we would like to focus on the problem of worldwide mortality. For example, investigate the risk factors contributing to mortality, such as age and gender. Analyze the geographical distribution of mortality rates to identify regions or areas with significantly higher mortality rates. This can help allocate resources more effectively for public health interventions. We plan to use data from World Heath Organizations (WHO) database and to extract some useful, previously unknown information patterns by using data mining algorithms. With the reference paper ”Data mining of WHO data warehouse with PASW modeler”, we plan to do both clustering analysis and classification analysis on data from WHO mortality database.

Data

Data from the Global Health Observatory Data Repository: https://www.who.int/data/gho
We plan to use the data from WHO mortality database.

Reference Article

S. Masih, S. Dubey, D. Pathak and N. Rahatekar, ”Data mining of WHO data warehouse with PASW modeler,” 2011 3rd International Conference on Electronics Computer Technology, Kanyakumari, India, 2011, pp. 52-56, doi: 10.1109/ICECTECH.2011.5942049.

Data Mining Algorithm

1. Clustering analysis: We plan to implement the K-means clustering algorithm. By the grouping result, we can analyse the similarity within the same group, and the difference between different groups. For example, by applying clustering algorithms to group countries or regions with similar mortality trends. This can reveal patterns that might not be apparent when analyzing data at a global level.

2. Classification analysis: We would like to use the CART algorithm to get variable importance of the data and analyse in which age range do people have a higher death rate.