The task of this work consists in the analysis of a dataset containing informations about cardiovascular disease [1]. It is made of 70 000 records of patients data composed by 11 features and a target indicating the presence or absence of the disease. After having done some preliminary data manipulation, the main goal of the project is to build the best classification predictive model using Bayesian Networks. In order to find the most probable belief-network structure, the K2 algorithm [2] is implemented and used. The results in terms of accuracy of the predictions are compared to other classification models such as decision trees.

[1] <https://www.kaggle.com/sulianova/cardiovascular-disease-dataset>

[2] G. F. Cooper and E. Herskovits, A Bayesian Method for the Induction of Probabilistic Networks from Data, Machine Learning 9, (1992) 309