# Predictions on heart-failure based on clinical record dataset through Bayesian Network and reasoning

*Master's Degree in Artificial Intelligence, University of Bologna*

*Fundamentals of Artificial Intelligence and Knowledge Representation - Module 3*

**Pietro Tellarini, Giorgia Bravi**

*Email:* pietro.tellarini2@studio.unibo.it,   *Matricola:* 0001145761
*Email:* giorgia.bravi2@studio.unibo.it,   *Matricola:* 0001143630

February 1, 2024

## Abstract

Heart failure is one of the primary causes of death worldwide. This study aims to investigate the existing relationships between symptoms and diseases based on various user characteristics and lifestyle factors. Given the symptoms and characteristics, the developed network in the case study can be used to calculate the probability of the presence of different diseases or the potential occurrence of death. The study is based on the dataset provided by Heumos [2022]. The explored data predict the survival or death of patients with heart failure. The analysis involves various tables related to conditioned probabilities, leading to different conclusions and observations for each of them.

## 1 Introduction

### 1.1 Domain

The dataset contains various elements indicating the characteristics, lifestyle, and symptoms of the patients participating in the study. The collected data align with the study presented by Wu et al. [2023] which identifies 13 categories and their corresponding values for 299 patients. The domain is explained in the code and contains: Anaemia, Creatinine Phosphokinase, Diabetes, Ejection Fraction, Blood Pressure, Platelet, Serum Creatinine, Serum Sodium, Sex, Smoking, Time.

### 1.2 Aim

Bayesian networks are used in a wide range of studies and applications, primarily in the modeling of probabilistic relationships between variables. The analysis and studies conducted are the following:

- Bayesian networks provide a natural framework for Bayesian inference, allowing estimation of the posterior probability distributions of unobserved variables given the observed data, evidences. Therefore the network can be used to predict future events, in this case of study the primary predicted event is the death event and how we can introduce observation on how to reduce it.

- Bayesian networks enable modeling and analysis of probabilistic dependencies and independencies between variables. Permit to identify relevant variables for a given phenomenon.

- Causal relationships between variables can be infered with Bayesian Network, observing the impact of a variable on others.

- Evaluation is another importat point, the fit of Bayesian network models to observed data can be evaluated.

- The sensitivity of model conclusions to variations in the parameters or structures of the Bayesian network can be evaluated.

### 1.3 Method

The Probabilistic Graphical Models in Python defined as pgmpy library has a particular emphasis on Bayesian networks and Markov networks. In this case of study this library implement the network and provides methods for probabilistic inference, performing queries on the Bayesian network. The primary methods used in this case is the MaximumLikelihoodEstimator to estimate the
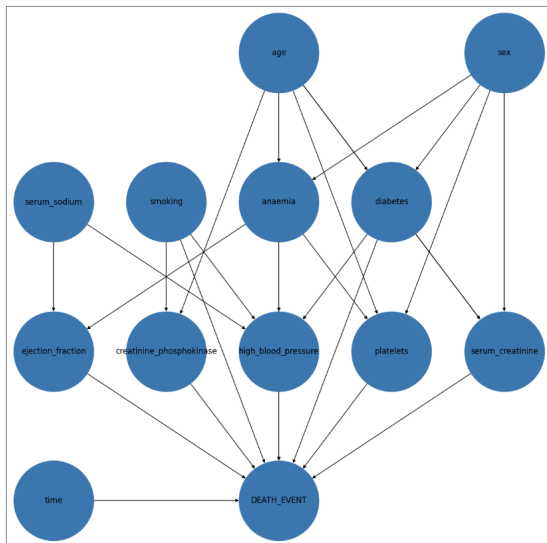
Figure 1: Bayesian Network.

parameters from the provided data and the VariableElimination method for exact inference, that allow to run queries on the Bayesian network.

The pgmpy library primarily works with discrete parameters in the context of Bayesian networks. First we discretized, converted continuous variables into discrete categories, using a special technique, a variable-width binning, that allow to obtain intervals of different but meaningful sizes relative to the assessed domain. The meaningful sizes of the intervals have been assumed by Medscape [2021] and PrelievoADomicilio [2020].

The evaluation of the quality of our Bayesian model was conducted using two distinct methods: the Bayesian Information Criterion (BIC) and the K2 Score.

### 1.4 Results

Based on our analysis, examining the event of death, although conditioned in certain cases by various diseases, the probability conditioned by these factors does not show a substantial difference. Therefore, the study has predominantly focused on identifying how diseases and symptoms interact with each other. In particular, an interesting case study has emerged by analyzing how high blood pressure, often actively influenced by diseases such as anemia and diabetes, can be brought within correct parameters with appropriate sodium levels.

## 2 Model

Our model assumes several binary variables and an equal number of continuous variables. The continuous variables have been discretized into intervals of variable width, as described in the code.

The Conditional Probability Tables (CPTs) are calculated using a Maximum Likelihood Estimator approach and a Bayesian Estimator, so that they could be inserted and adapted to the constructed Bayesian model. The developed code encompasses all the conditional probability tables for the various items.

The Bayesian model was derived from a thorough analysis of the data and their interactions. Information was obtained by reviewing various online research sources, and the validity of the network was assessed by a medical student. The student evaluated the consistency of the relationships between the nodes. The probabilities associated with each node were determined using the data, estimating the Conditional Probability Tables (CPTs) for each node through a Maximum Likelihood Estimator.

## 3 Analysis

### 3.1 Experimental setup

We've carried out different experiments, the most interesting one is how High Blood Pressure is influenced by the level of sodium given Aanemia and Diabetes and how the Ejection Fraction is influenced by the level of sodium given Anaemia.

### 3.2 Results

The observations made about the case study are related to the direct and indirect influences of lifestyle, symptoms, and patient characteristics on different diseases and the occurrence of death.

## 4 Conclusion

Through the analysis of the Bayesian network, we estimated various conditional probabilities related to different items. As previously mentioned, a limitation of the case of study is the homogeneity observed when analyzing the conditional death probabilities based on different factors. The dataset, characterized by a limited number of data points, often does not contribute to obtaining probabilistic values close to real-world scenarios or in specific case studies.

However, it is interesting to observe how symptoms, diseases, and lifestyles influence each other. We have gained insights into manipulating large amounts of data and understanding how they can be interconnected. We believe that the model developed in this document serves as an excellent starting point for delving into the topic of disease diagnosis based on historical data. In a broader context, we consider that this tool can be effectively implemented in a real-world setting.

# 5  Links to external resources

GitHub Repoitory: BayesianNetwork developed by Tellarini and Bravi [2024]

# References

Lukas Heumos. Heart failure clinical records Data Set, 2022. URL https://figshare.com/articles/dataset/Heart_failure_clinical_records_Data_Set/19108337/1. Artwork Size: 14592 Bytes Pages: 14592 Bytes.

Zhuocheng Wu, Yunsi Wu, Xiang He, Lingjin Wan, Li-Chuan Tsai, and Aiguo Chen. Application of Machine Learning for Heart Failure Prediction. In *2023 6th International Conference on Artificial Intelligence and Big Data (ICAIBD)*, pages 276–282, Chengdu, China, May 2023. IEEE. ISBN 978-1-66549-125-9. doi: 10.1109/ICAIBD57115.2023.10206268. URL https://ieeexplore.ieee.org/document/10206268/.

Medscape. Lab Values, Normal Adult: Laboratory Reference Ranges in Healthy Adults. July 2021. URL https://emedicine.medscape.com/article/2172316-overview?form=fpf. Publication: Medscape - eMedicine.

PrelievoADomicilio. Valori del sangue - Valori di riferimento analisi PDF, 2020. URL https://www.prelievoadomicilio.it/valori-riferimento-esame-del-sangue/.

Pietro Tellarini and Giorgia Bravi. PtrTella/BayesianNetwork, January 2024. URL https://github.com/PtrTella/BayesianNetwork. original-date: 2024-01-22T17:24:41Z.