Data
Science

Information fusion in data analysis

Data Fusion – Project (2)

Alberto Cardoso, Jorge Henriques, Silvia Rodrigues

Departamento de Engenharia Informática
Faculdade de Ciências e Tecnologia
UNIVERSIDADE DE COIMBRA

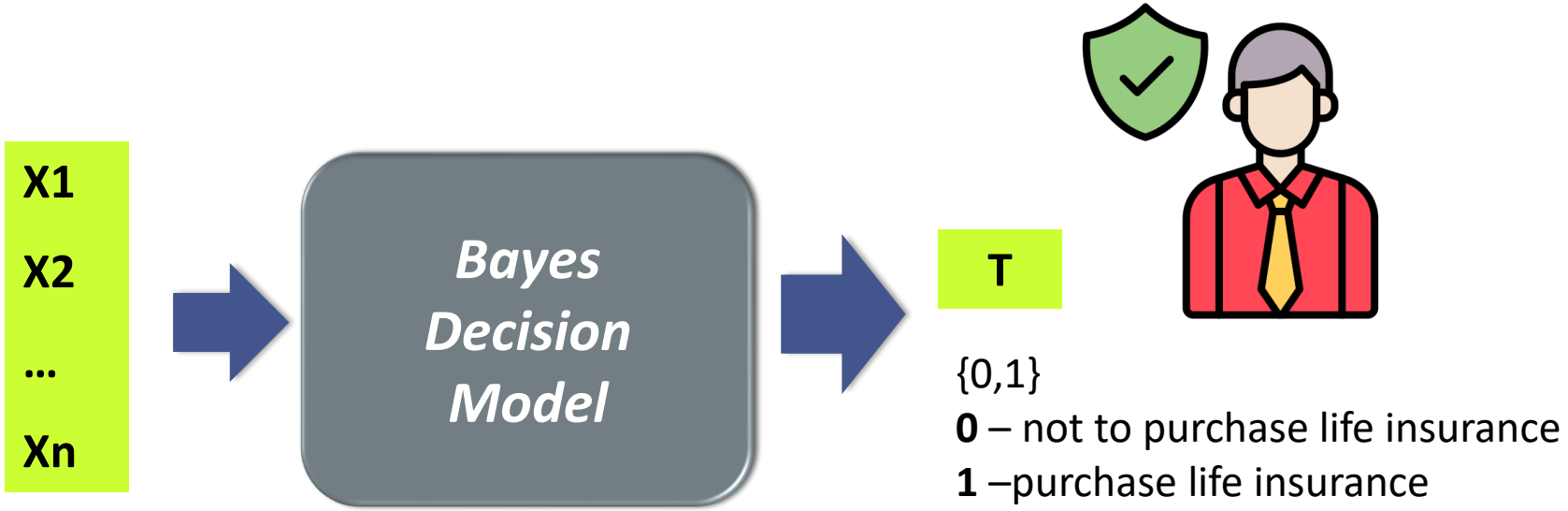
Objective

Problem – Life Insurance

- Based on several variables, implement a Bayesian fusion decision model to help individuals decide whether to purchase life insurance.
- To this aim, a dataset containing examples of individual decisions is provided (assumed to be correct decisions).
  - **lifeInsurance.txt**

Objectivo

Life Insurance



{0,1}
0 – not to purchase life insurance
1 –purchase life insurance

Objectivo

Data set

- **X1** *Gender* {0,1} = { Female, Male}
- **X2** *Age* [34 .. 101]
- **X3** *Marital status* {0,1} = { single, married}
- **X4** *number of children/dependents* {0,1,2,3} = { 0,1,2,>=3}
- **X5** *Physical Status* {0,1,2} = { sedentary, moderately active, active}
- **X6** *chronic diseases* {0,1,2} = { no conditions, moderate, severe}
- **X7** *Monthly Salary* [1370.. 3800]

- *Insurance company suggestion* Based on a score point* (see next page)

- **T** *Decision* Individual decision
  {0,1} = { not purchase, purchase}

* Rule to decide whether to have life insurance or not:

- If the total risk based on age, health, financial situation, and family responsibilities **exceeds 50 points**, it is recommended to purchase life insurance.
- Otherwise, it is not necessary.

- Age:
  - Under 30 years: 5 points
  - Between 30 and 40 years: 10 points
  - Between 40 and 50 years: 15 points
  - Over 50 years: 20 points

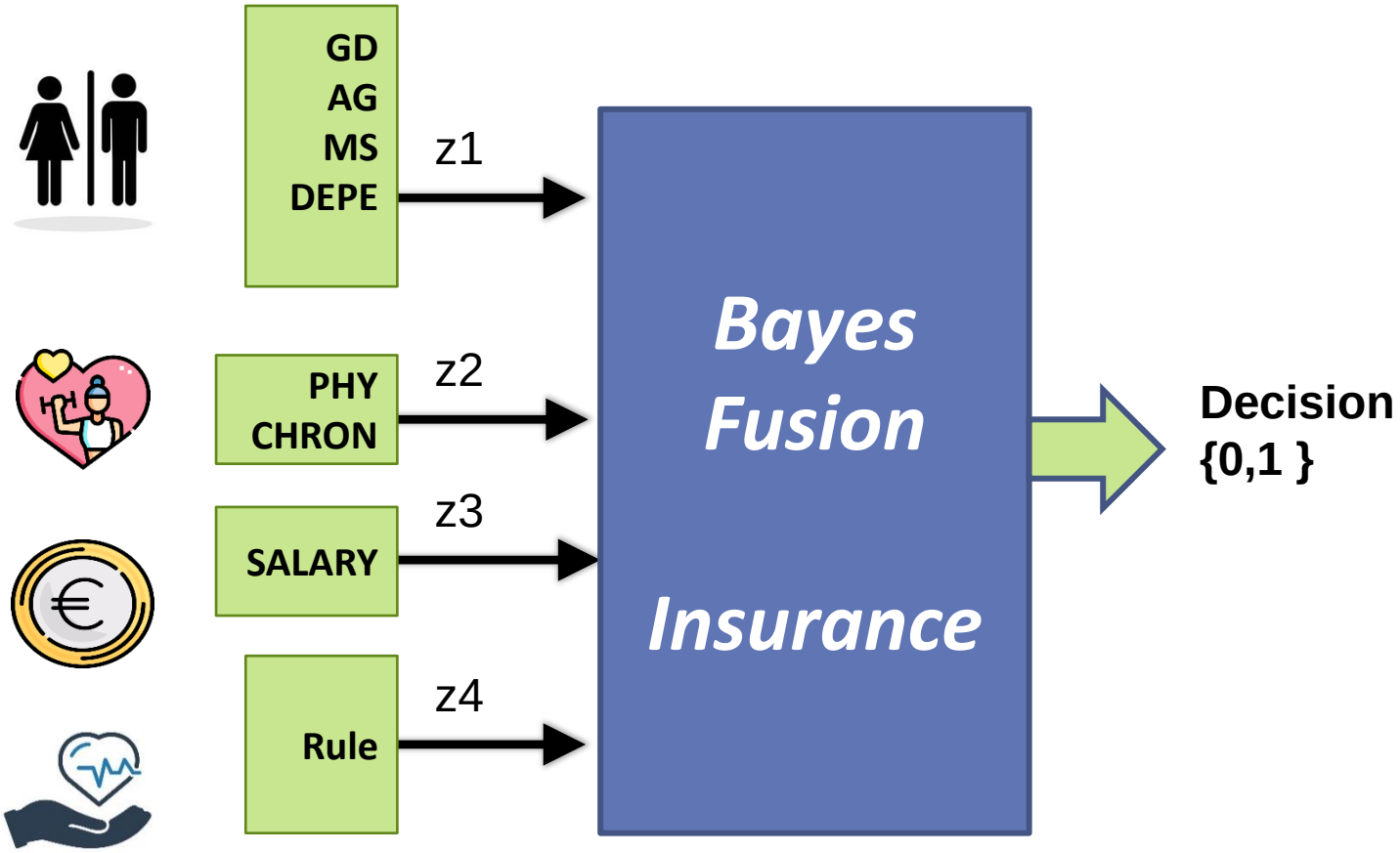
- Health Condition (assessed based on chronic conditions):
  - Healthy: 5 points
  - History of minor diseases (e.g., high cholesterol): 10 points
  - Serious or chronic diseases (e.g., diabetes, hypertension, heart disease): 15 points

- Financial Situation (assessed by income):
  - Monthly income above 3500 €: 5 points
  - Monthly income between 1700€ and 3500 €: 10 points
  - Monthly income below 1700 €: 15 points

- Family Responsibilities (assessed by the number of children/dependents):
  - No dependents: 5 points
  - 1 dependent: 10 points
  - More than 1 dependent: 15 point

Information fusion

- Historical, measurements, clinical knowledge (guidelines)



Questions

- Is the performance of the classifier acceptable ?
- Should all information (inputs/variables) be used ?
- Discrete versus continuous variables ?
- Conditional probabilities : normal distribution ?
- Other distributions ?
- Gaussian mixtures approaches ?
- ….

Elements for evaluation

- IMPORTANT
  - **Bayesian libraries cannot be used!**
  - All code should be provided

- **Report**
  - Maximum 5/6 pages
  - Explain the important decisions

- **Defense**
  - Mandatory, 20 June 2025

- **Deadline for submission**
  - 19 June 2025