

Redes Bayesianas

Sistema Especialista Probabilístico para Diagnóstico Cardíaco

Diego Carlito Rodrigues de Souza - 221007690

Disciplina: Inteligência Artificial

Universidade de Brasília (UnB)

Novembro de 2025

Resumo

Este relatório apresenta o desenvolvimento de um Sistema Especialista baseado em Redes Bayesianas para a avaliação de risco de doenças cardíacas. Utilizando as variáveis do dataset *Cleveland Heart Disease* como referência para a modelagem do domínio, foi construída manualmente uma topologia causal que reflete o conhecimento médico sobre fatores de risco (idade, colesterol, diabetes) e sintomas. O trabalho detalha a engenharia de conhecimento aplicada na definição das Tabelas de Probabilidade Condicional (CPTs) e analisa a robustez do modelo através de múltiplos cenários de inferência.

1 Introdução

O raciocínio sob incerteza é fundamental em diagnósticos médicos, onde sintomas podem ter múltiplas causas e exames nem sempre são conclusivos. Sistemas inteligentes devem ser capazes de atualizar suas crenças à medida que novas evidências são apresentadas, ponderando informações conflitantes.

Neste projeto, aplicaram-se Redes Bayesianas para modelar essas incertezas. Diferente de abordagens baseadas puramente em dados (Machine Learning), optou-se por uma abordagem de **Sistema Especialista**, onde a estrutura da rede e os parâmetros probabilísticos são definidos a priori baseados na lógica causal do domínio. Isso garante que o modelo seja interpretável e siga princípios fisiológicos conhecidos.

2 O Problema

O diagnóstico de doenças cardíacas envolve a análise de múltiplos fatores de risco que interagem de maneira não-linear. Um sintoma isolado pode ser inconclusivo, mas quando combinado com fatores de risco (como idade ou diabetes), a probabilidade de doença aumenta significativamente.

O desafio abordado é: **Como modelar probabilisticamente essas interações causais para permitir diagnósticos robustos mesmo com informações parciais?** A solução proposta utiliza a biblioteca pgmpy para implementar uma Rede Bayesiana que conecta fatores demográficos, condições fisiológicas e sintomas observáveis.

3 Projeto Desenvolvido

3.1 Ambiente e Bibliotecas

O sistema foi desenvolvido em Python 3.12, utilizando a biblioteca `pgmpy` para a modelagem probabilística e inferência exata. O código foi estruturado para ser compatível com diferentes versões da biblioteca, garantindo portabilidade.

3.2 Dicionário de Dados

As variáveis foram selecionadas do dataset clássico *Cleveland* e adaptadas para estados categóricos para facilitar a modelagem das CPTs manuais. A Tabela 1 descreve o mapeamento.

Tabela 1: Definição das Variáveis do Modelo

Cód.	Significado	Estados	Tipo
age	Idade do Paciente	Jovem / Idoso	Causa Raiz
sex	Sexo Biológico	Feminino / Masculino	Causa Raiz
fbs	Glicemia (Diabetes)	Normal / Diabetes	Causa Raiz
chol	Colesterol Sérico	Normal / Alto	Intermediária
num	Doença Cardíaca	Saudável / Doente	Alvo
cp	Dor no Peito	Assintomático / Angina	Sintoma
thalach	Freq. Cardíaca Máx.	Normal / Anormal	Sintoma

3.3 Topologia da Rede (Grafo Causal)

A estrutura do Grafo Dirigido Acíclico (DAG) foi desenhada seguindo a causalidade médica:

- **Fatores de Risco** \rightarrow **Condição Fisiológica**: Idade e Sexo influenciam a probabilidade de ter Colesterol Alto ($\text{age}, \text{sex} \rightarrow \text{chol}$).
- **Fisiologia** \rightarrow **Doença**: Colesterol alto e Diabetes aumentam diretamente a probabilidade da doença ($\text{chol}, \text{fbs} \rightarrow \text{num}$).
- **Doença** \rightarrow **Sintomas**: A presença da doença causa sintomas observáveis como Angina e anomalias na frequência cardíaca ($\text{num} \rightarrow \text{cp}, \text{thalach}$).

4 Resultados e Cenários de Teste

Para validar a consistência lógica do sistema especialista, foram executados 6 cenários de inferência utilizando o algoritmo de Eliminação de Variáveis. As figuras abaixo apresentam os resultados obtidos diretamente da execução do software.

4.1 Cenário 1: Baixo Risco (Baseline)

Simulou-se um paciente Jovem, do sexo Feminino e sem Diabetes.

```

--- Cenário 1: Paciente de baixo risco (jovem, feminino, sem comorbidades) ---
Evidências: {'age': 'Jovem', 'sex': 'Feminino', 'fbs': 'Normal'}
+-----+-----+
| num          | phi(num) |
+-----+-----+
| num(Saudavel) | 0.8950 |
+-----+-----+
| num(Doente)   | 0.1050 |
+-----+-----+
-> Probabilidade de Doença Cardíaca: 10.50%

```

Figura 1: Inferência para paciente de baixo risco.

Análise: O modelo calculou uma probabilidade baixa (10,5%), coerente com o perfil epidemiológico de menor risco na população modelada.

4.2 Cenário 2: Alto Risco Demográfico

Simulou-se um paciente Idoso, Masculino e Diabético.

```

--- Cenário 2: Paciente de alto risco demográfico (idoso, masculino, diabético) ---
Evidências: {'age': 'Idoso', 'sex': 'Masculino', 'fbs': 'Diabetes'}
+-----+-----+
| num          | phi(num) |
+-----+-----+
| num(Saudavel) | 0.4400 |
+-----+-----+
| num(Doente)   | 0.5600 |
+-----+-----+
-> Probabilidade de Doença Cardíaca: 56.00%

```

Figura 2: Inferência para paciente de alto risco demográfico.

Análise: Apenas com fatores de risco (sem sintomas), a probabilidade subiu para 56%. O modelo capturou corretamente que fatores agravantes aumentam a propensão à doença.

4.3 Cenário 3: Inferência Reversa (Diagnóstica)

Paciente chega com sintomas claros (Angina e Frequência Anormal), sem dados de histórico.

```

--- Cenário 3: Inferência reversa - paciente sintomático ---
Evidências: {'cp': 'Angina', 'thalach': 'Anormal'}
+-----+-----+
| num          | phi(num) |
+-----+-----+
| num(Saudavel) | 0.0485 |
+-----+-----+
| num(Doente)   | 0.9515 |
+-----+-----+
-> Probabilidade de Doença Cardíaca: 95.15%

```

Figura 3: Inferência diagnóstica baseada apenas em sintomas.

Análise: Este cenário demonstra o poder da inferência diagnóstica (do efeito para a causa). A presença forte de sintomas elevou a probabilidade a posteriori para 95,15%, isolando a doença como a explicação mais provável.

4.4 Cenário 4: Evidência Mista

Paciente Jovem (baixo risco a priori) apresentando Angina.

```
--- Cenário 4: Evidência mista - jovem com sintomas preocupantes ---
Evidências: {'age': 'Jovem', 'sex': 'Masculino', 'cp': 'Angina'}
+-----+-----+
| num          | phi(num) |
+-----+-----+
| num(Saudavel) | 0.2839 |
+-----+-----+
| num(Doente)   | 0.7161 |
+-----+-----+
-> Probabilidade de Doença Cardíaca: 71.61%
```

Figura 4: Inferência com evidências conflitantes (Jovem + Angina).

Análise: O modelo ponderou evidências conflitantes. Embora a idade jovem reduza a probabilidade, o sintoma observado tem um peso maior na verossimilhança, resultando em 71,61% de chance de doença.

4.5 Cenário 5: Colesterol Alto Isolado

```
--- Cenário 5: Colesterol alto isolado ---
Evidências: {'chol': 'Alto'}
+-----+-----+
| num          | phi(num) |
+-----+-----+
| num(Saudavel) | 0.3700 |
+-----+-----+
| num(Doente)   | 0.6300 |
+-----+-----+
-> Probabilidade de Doença Cardíaca: 63.00%
```

Figura 5: Inferência com fator de risco isolado.

Análise: A presença isolada de Colesterol Alto elevou o risco para 63%, confirmando sua importância como variável intermediária crítica no modelo.

4.6 Cenário 6: Pior Caso

Todos os fatores de risco e sintomas presentes simultaneamente.

```
--- Cenário 6: Pior caso - todos os fatores de risco presentes ---
Evidências: {'age': 'Idoso', 'sex': 'Masculino', 'fbs': 'Diabetes', 'chol': 'Alto', 'cp': 'Angina', 'thalach': 'Anormal'}
+-----+-----+
| num          | phi(num) |
+-----+-----+
| num(Saudavel) | 0.0052 |
+-----+-----+
| num(Doente)   | 0.9948 |
+-----+-----+
-> Probabilidade de Doença Cardíaca: 99.48%
```

Figura 6: Inferência com todas as evidências positivas.

Análise: Convergência para quase certeza (99,48%), validando a consistência acumulativa do modelo.

5 Dificuldades e Contribuições

5.1 Dificuldades Encontradas

A principal dificuldade na abordagem de Sistema Especialista foi a **calibração das Tabelas de Probabilidade Condicional (CPTs)**. Diferente do aprendizado de máquina onde os parâmetros vêm dos dados, aqui foi necessário estimar valores (ex: "Qual a chance de um homem idoso ter colesterol alto?") que mantivessem a coerência lógica do sistema. Pequenas alterações nas probabilidades a priori causavam grandes impactos na inferência final, exigindo ajustes finos.

5.2 Contribuições

A contribuição central deste trabalho é a modelagem explícita de um domínio médico complexo. O projeto demonstra como Redes Bayesianas podem integrar conhecimento qualitativo (causalidade) com raciocínio quantitativo, permitindo diagnósticos robustos mesmo na ausência de informações completas (como visto no Cenário 3, onde o diagnóstico foi feito apenas pelos sintomas). Além disso, a implementação modular permite fácil expansão para incluir novos sintomas ou exames.