



Data Science Academy

[www.datascienceacademy.com.br](http://www.datascienceacademy.com.br)

Engenharia de Dados com Hadoop e Spark

Mini-Projeto 2  
Prevendo a Ocorrência de Doenças  
Cardíacas

## O que é uma doença cardíaca?

Doença cardíaca é um termo geral para designar diversas condições médicas crônicas ou agudas que afetam um ou mais componentes do coração. Entre os pulmões existe uma cavidade conhecida como mediastino. Este é o lugar onde o coração está posicionado – partindo do centro do corpo humano, dois terços para a esquerda. O coração é um órgão muscular do tamanho de um punho, que bombeia o sangue através da rede de artérias e veias chamada sistema cardiovascular.

O coração tem quatro câmaras:

- Átrio direito: recebe o sangue das veias e bombeia para o ventrículo direito
- Ventrículo direito: recebe o sangue do átrio direito e bombeia para os pulmões, onde ele é carregado com oxigênio
- Átrio esquerdo: recebe sangue oxigenado dos pulmões e bombeia para o ventrículo esquerdo
- Ventrículo esquerdo: bombeia o sangue oxigenado para o resto do corpo. As contrações do ventrículo esquerdo criam a nossa pressão arterial.

As artérias coronárias correm ao longo da superfície do coração e fornecem sangue rico em oxigênio ao músculo cardíaco. Uma teia de tecido nervoso também atravessa o coração, conduzindo os sinais neurológicos complexos que regem a contração e relaxamento. Essa teia que envolve o coração é um saco chamado pericárdio.

A doença cardíaca ocorre quando uma dessas estruturas não está funcionando corretamente. Algumas doenças cardíacas comuns: Angina instável e estável, Arritmia cardíaca, Artrose, Aterosclerose (doença cardíaca coronária), Arteriosclerose, Cardiomiopatia, Cardiopatia congênita, Doença arterial periférica, entre outras.

## Diagnóstico e exames

Alguns exames podem ser feitos para diagnosticar ou acompanhar doenças cardíacas. Veja:

- Eletrocardiograma
- Ecocardiograma
- Teste ergométrico
- Cateterismo cardíaco
- Holter 24 horas
- Monitor cardíaco portátil
- Machine Learning

### **O que? Machine Learning?**

Sim, isso mesmo. Cada vez modelos preditivos vem sendo usados para prever a ocorrência de doenças cardíacas, usando algoritmos de classificação. A partir de dados coletados de pacientes, é cada vez mais comum a utilização de Machine Learning para prever a ocorrência de doenças em geral, incluindo doenças cardíacas.

O objetivo deste projeto é servir como um estudo de caso para investigar todos os passos necessários na criação de um modelo preditivo doenças cardíacas. Usaremos diversas das ferramentas do ecossistema Hadoop, que estudamos ao longo do curso.

O dataset pacientes.csv (que você encontra junto com os demais arquivos do projeto) contém as seguintes colunas, com dados coletados de pacientes de um hospital que atende idosos:

Coluna	Descrição
ID	ID único para cada registro
Idade	Idade do paciente
Sexo	Sexo: 0 – Feminino 1 – Masculino
Pressão Sanguínea	Pressão sanguínea medida
Colesterol	Colesterol medido
Açúcar no Sangue	Nível de açúcar no coração: 0 – Nível de açúcar $\leq 120$ mg/dl 1 - Nível de açúcar $> 120$ mg/dl
ECG	Resultados ECG: 0 – Normal 1- Alguma anomalia 2- Anomalia presente
Batimento cardíaco (Valor Máximo)	Batimento cardíaco medido
Doença	Indica se o paciente tem doença cardíaca: 0 – Não 1 – Sim

➔ A variável Doença é a variável target e todas as demais são as variáveis preditoras.

## Dataset (sample)

ID	Idade	Sexo	Pressão Sanguínea	Colesterol	Açúcar no Sangue	ECG	Batimentos Cardíacos	Doença
1001	63	1	145	233	1	2	150	0
1002	67	1	160	286	0	2	108	1
1003	69	1	145	235	1	2	129	0
1004	68	1	120	229	0	1	110	0

Seu trabalho, como Cientista de Dados, é criar um modelo preditivo que utilize as variáveis preditoras e seja capaz de prever (a partir de novos conjuntos de dados) se um paciente pode ou não desenvolver doenças do coração.