

# **Python e Machine Learning para detecção de câncer**

Jannyfer e Júlia Tamagno

# Câncer de mama

O câncer de mama é uma doença causada pela multiplicação desordenada de células anormais da mama, que forma um tumor com potencial de invadir outros órgãos.

Há vários tipos de câncer de mama. Alguns têm desenvolvimento rápido, enquanto outros crescem lentamente. A maioria dos casos, quando tratados adequadamente e em tempo oportuno, apresentam bom prognóstico e possibilitam melhores resultados estéticos.

O câncer de mama também acomete homens, porém é raro, representando apenas 1% do total de casos da doença.

O Sistema Único de Saúde (SUS) oferece tratamento para o câncer de mama em Unidades Hospitalares especializadas.

Fonte: <https://www.gov.br/inca/pt-br/assuntos/cancer/tipos/mama>



# Informações importantes

1

Um em cada três casos de câncer pode ser curado se for descoberto logo no início. Alguns tipos de câncer, entre eles o de mama, apresentam sinais e sintomas em suas fases iniciais. Detectá-los precocemente traz melhores resultados no tratamento e ajuda a reduzir a mortalidade.

2

Câncer de mama é o tipo mais comum de câncer no Brasil, depois do câncer de pele, e também é o que causa mais mortes por câncer em mulheres. Homens também podem ter câncer de mama, mas isso é raro, ocorrendo em apenas 1% dos casos.

3

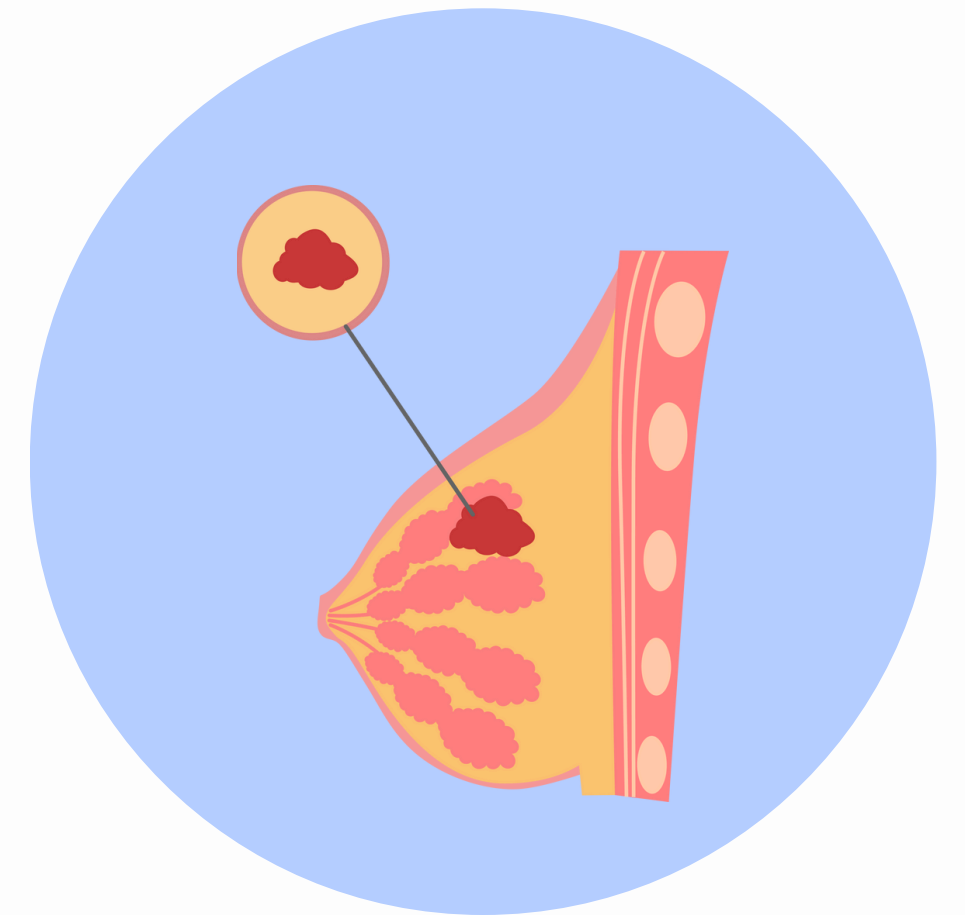
Não há uma causa única para o câncer de mama e o risco de desenvolver a doença aumenta com a idade, sendo maior a partir dos 50 anos. Manter o peso corporal adequado, praticar atividade física e evitar o consumo de álcool ajudam a reduzir o risco. A amamentação também é considerada um fator protetor.

Fonte: Cartilha do câncer de mama do INCA (Instituto Nacional de câncer).

# Como utilizar python para detectar câncer de mama?

Vamos criar um pequeno projeto que utilizará algoritmos de *Machine Learning* implementados em *Python* para analisar um conjunto de dados contendo informações sobre pacientes com massas sólidas nos seios que passaram por um procedimento de biópsia.

A partir da análise destes dados e da aplicação de algoritmos, seremos capazes de criar um projeto que prevê a possibilidade de um paciente possuir câncer.



# Resumo de um projeto

1

Elencar o objetivo do seu projeto e definir o problema que deseja resolver.

Com um escopo definido torna-se mais fácil de criar o roteiro do seu projeto e pesquisar as informações necessárias.

2

Preparar seu dataset

Para projetos pessoais é possível encontrar datasets em diversas fontes (recomendamos o kaggle para iniciantes).

Em projetos profissionais é muito comum que você precise buscar estes dados em bancos e fazer um processo de ETL, talvez até modelar e criar bancos para começar a coletar estes dados.

3

Analisar e tratar os dados

É impossível desenvolver algoritmos e começar a colocar a mão na massa sem conhecer os seus dados. Tome o tempo que for necessário para explorar o dataset e conhecer o seu conteúdo.

# Resumo de um projeto

4

## Seleção de modelos

É muito importante ter uma noção de conceitos de machine learning e conhecer alguns algoritmos que podem ser úteis em contextos parecidos. Para o nosso projeto utilizaremos Random Forest.

5

## Testes

Depois de elencar os modelos é hora de se divertir com o seu dataset.

Treine os modelos e teste para descobrir as métricas dos modelos selecionados. Por fim, decida qual o melhor modelo para utilizar em análises futuras.

6

## Insights e conclusões

Projetos são ótimos pontos de partida para resolver problemas do dia-a-dia utilizando a tecnologia. Analise os dados provenientes do seu projeto e identifique como ele pode ajudar a resolver o seu problema inicial.



```

fromTime = 0
toTime = 150
animLength = toTime - fromTime + 1

# prompt user for directory
filePath = c4d.storage.SaveDialog()
filePath, objName = os.path.split(filePath)
objName = objName + "-"
filePath = filePath + "\\ "
# Check for confirmation
questionDialogText = "Obj Sequence will be saved a\n"
    "" + filePath + objName + "####.obj\n\n"\n
    "From frame " + str(fromTime) + " to " + str(t
proceedBool = c4d.gui.QuestionDialog(questionDialo

if proceedBool == True:

    # loop through animation and export frames
    for x in range(0, animLength):

        # change frame, redraw view
        moveTime = c4d.BaseTime(fromTime, docFps) +
        doc.SetTime(moveTime)
        c4d.EventAdd(c4d.EVENT_FORCEREDRAW)
        c4d.DrawViews(c4d.DRAWFLAGS_FORCEFULLREDRA

        # progress bar
        c4d.StatusSetText("Exporting " + str(x) +
        c4d.StatusSetBar(100.0*x/animLength)

        # add buffer 0001
        bufferedNumber = str(doc.GetTime().GetF
        if len(bufferedNumber) < 4:

```

# Projeto na prática

Vamos desenvolver este projeto utilizando o Google Colaboratory, e caso você deseje replicar os conceitos aprendidos aqui, aconselhamos que utilize o mesmo ambiente.

Todos os materiais utilizados para este workshop serão disponibilizados para você praticar em casa.



# Fontes



01 [Detecção de câncer de mama — Machine Learning | Estephano Bartenski](#)

02 [Breast Cancer Detection Using Deep Learning | AITS Journal](#)

03 [Machine Learning para a Detecção de Câncer de Mama | Carlos Melo](#)

04 [INCA](#)

05 [Random Forest | Didática Tech](#)



**Dúvidas?**



Obrigada!

```
fromTime = 0
toTime = 150
animLength = toTime - fromTime + 1

# prompt user for directory
filePath = c4d.storage.SaveDialog()
filePath, objName = os.path.split(filePath)
objName = objName + "-"
filePath = filePath + "\\ "
# Check for confirmation
questionDialogText = "Obj Sequence will be saved as\n" + filePath + objName + "####.obj\n\n"\n"From frame " + str(fromTime) + " to " + str(toTime)
proceedBool = c4d.gui.QuestionDialog(questionDialogText)

if proceedBool == True:

    # Loop through animation and export frames
    for x in range(0, animLength):

        # change frame, redraw view
        moveTime = c4d.BaseTime(fromTime, docFps) + doc.SetTime(moveTime)
        c4d.EventAdd(c4d.EVENT_FORCEREDRAW)
        c4d.DrawViews(c4d.DRAWFLAGS_FORCEFULLREDRAW)

        # progress bar
        c4d.StatusSetText("Exporting " + str(x) + " of " + str(animLength))
        c4d.StatusSetBar(100.0*x/animLength)

        # add buffer 0001
        bufferedNumber = str(doc.GetTime().GetFrameNumber())
        if len(bufferedNumber) < 4:
```