

Disciplina: Inteligência Computacional.
Docente: Anusio Menezes.

Equipe:

- Leonardo Sena;
- Pedro Gustavo Santana;
- Vítor Pontes.

Aplicação de MLP

Problema 1

Tentar identificar quando o câncer de mama pode ser reincidente em pacientes.

Base de dados 1

Título : Breast Cancer data

Dados obtidos das universidades: University Medical Centre (Londres), Institute of Oncology Ljubljana (Eslovênia), Yugoslavia.

Agradecimento a M. Zwitter e M. Soklic por disponibilizar os dados.

Base de dados 1

Consiste em 286 instâncias que possuem 9 atributos de caracterização e 1 de classificação.

Características:

- Idade.
- Menopausa.
- Tamanho de tumor.
- Invasão linfonodal.
- Nódulo capsulado.
- Grau histológico do tumor
 - 1 - Predomínio de células cancerígenas.
 - 2 - Consiste em características usuais de células cancerígenas.
 - 3 - Células altamente afetadas.
- Mama (lado).
- Mama (quadrante).
- Histórico de radioterapia.

Classificação: Recorrente ou não recorrente.
Pessoas que voltaram a ter o câncer de mama.

Base de dados 1

Ajustes para utilização:

- Idade:
 - 30 - 39 -> 3
 - 40 - 49 -> 4
 - 50 - 59 -> 5
 - 60 - 69 -> 6
 - 70 - 79 -> 7
- Menopausa:
 - premeno -> 0
 - lt40 (antes) -> 1
 - ge40 (depois) -> 2
- Tamanho de tumo:
 - 0-4 -> 1
 - 5-9 -> 2
 - 10-14 -> 3
 - 15-19 -> 4
 - 20-24 -> 5
 - 25-29 -> 6
 - 30-34 -> 7
 - 35-39 -> 8
 - 40-44 -> 9
 - 45-49 -> 10
 - 50-54 -> 11
 - 55-59 -> 12

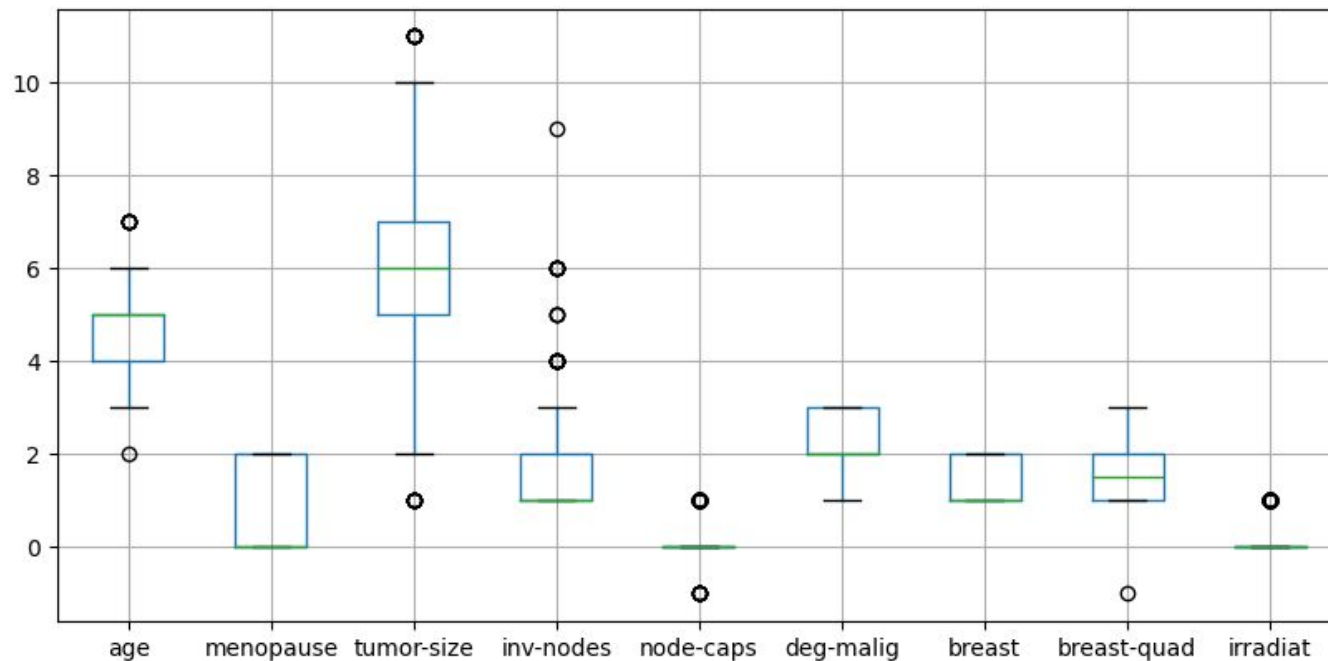
Base de dados 1

Ajustes para utilização:

- Invasão linfonodal:
 - 0-2 -> 1
 - 3-5 -> 2
 - 6-8 -> 3
 - 9-11 -> 4
 - 12-14 -> 5
 - 15-17 -> 6
 - 18-20 -> 7
 - 21-23 -> 8
 - 24-26 -> 9
 - 27-29 -> 10
 - 30-32 -> 11
 - 33-35 -> 12
 - 36-39 -> 13
- Nó capsulado:
 - Não -> 0
 - Sim -> 1
- Mama (lado):
 - Esquerdo -> 1
 - Direito -> 2
- Mama (quadrante):
 - Esquerdo baixo -> 1
 - Esquerdo alto -> 1.5
 - Direito baixo -> 2
 - Direito alto -> 2.5
- Histórico de radioterapia:
 - Não -> 0
 - Sim -> 1

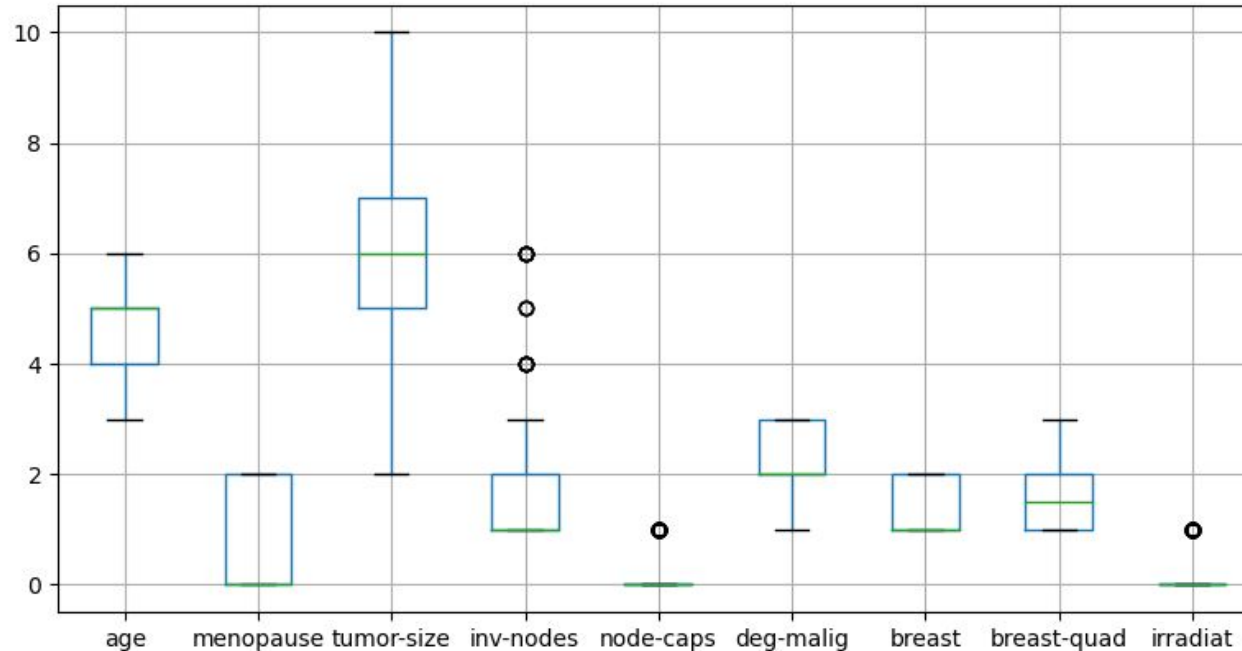
Base de dados 1

Original:

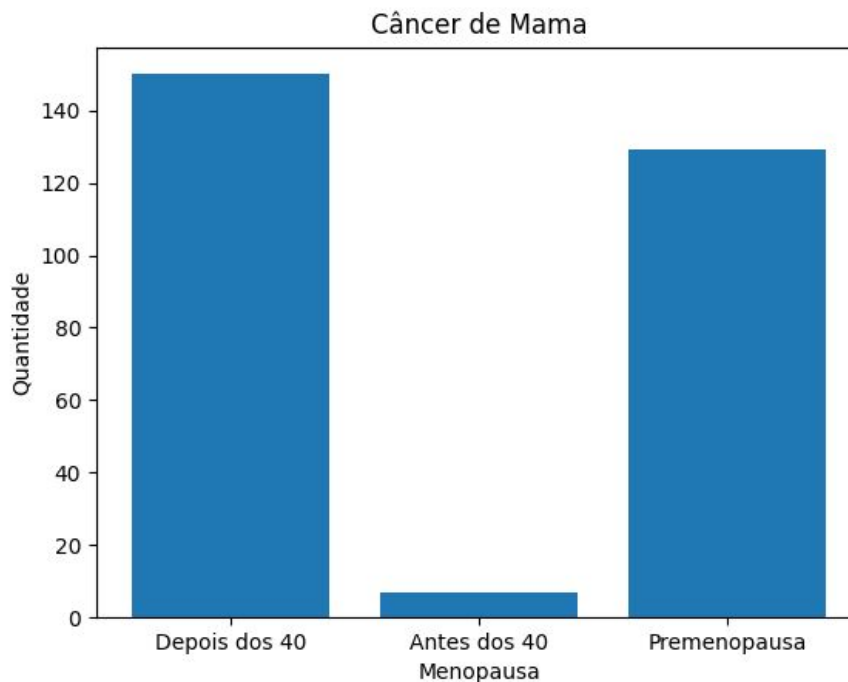
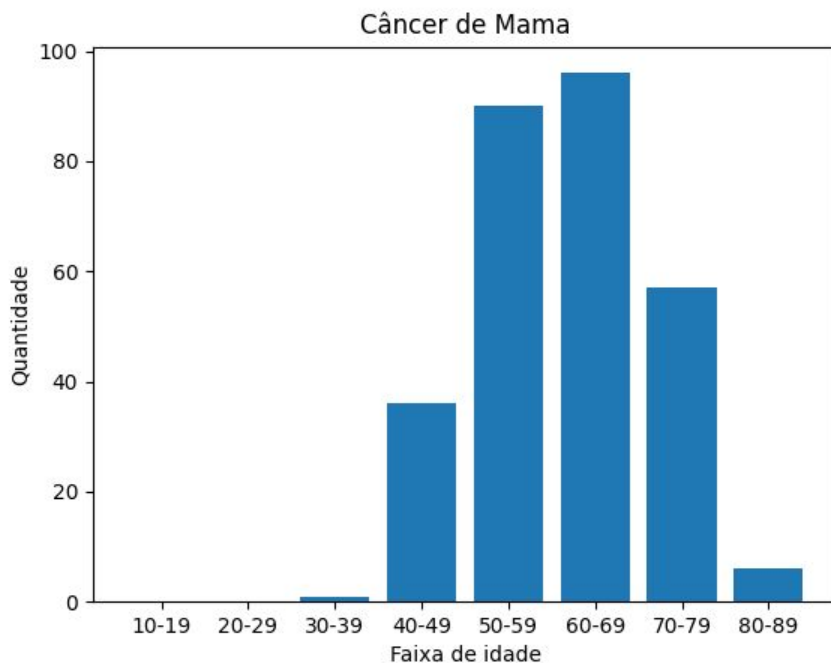


Base de dados 1

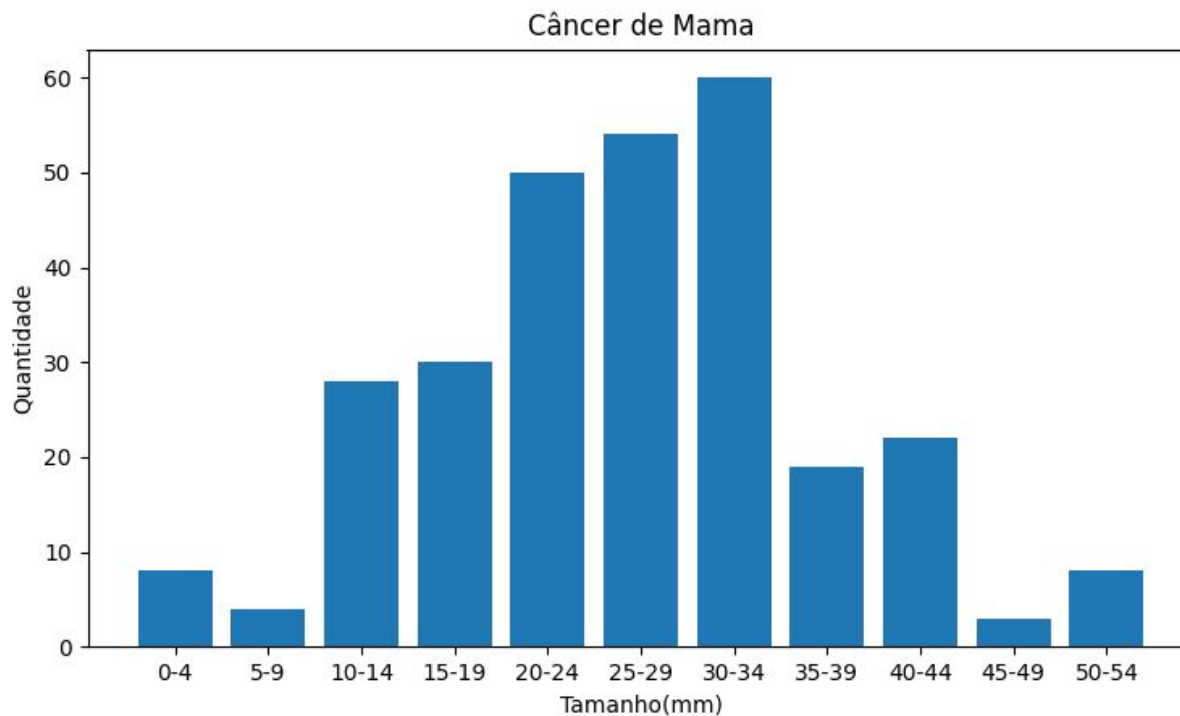
Após alterações:



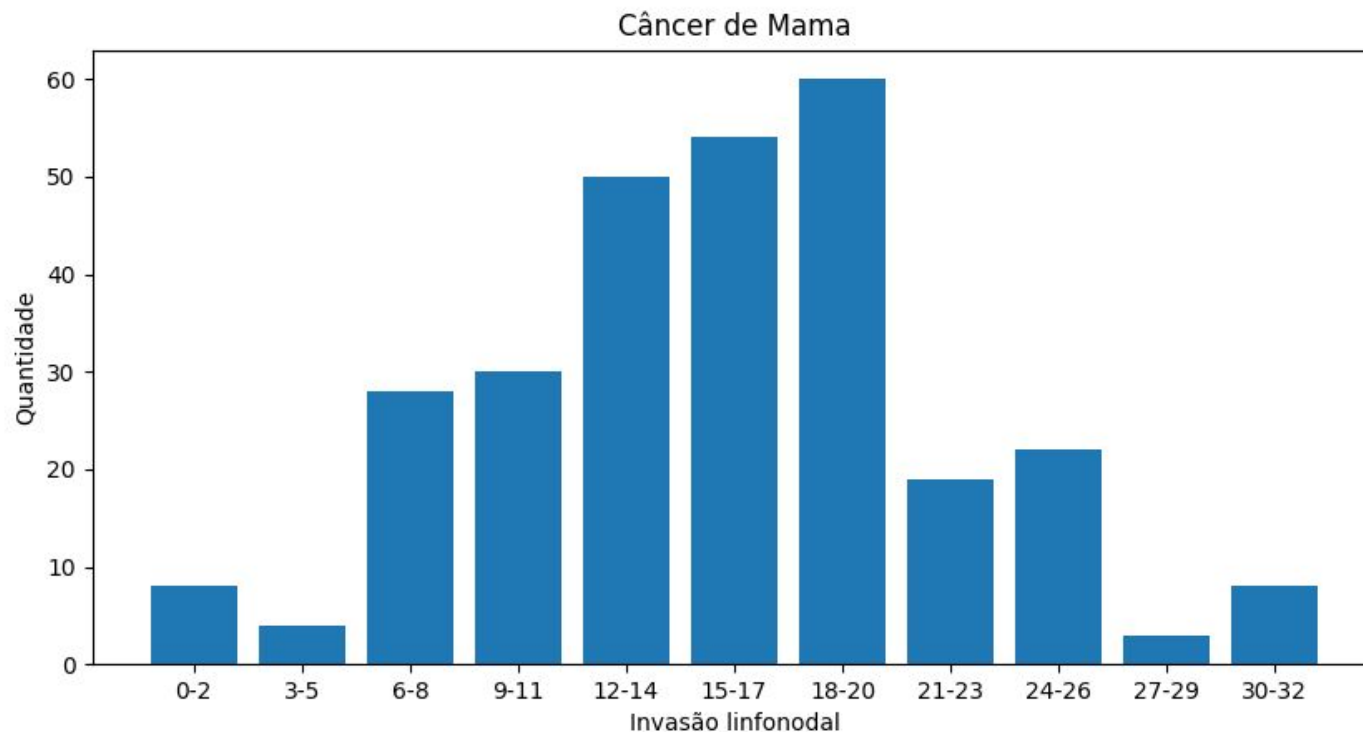
Base de dados 1



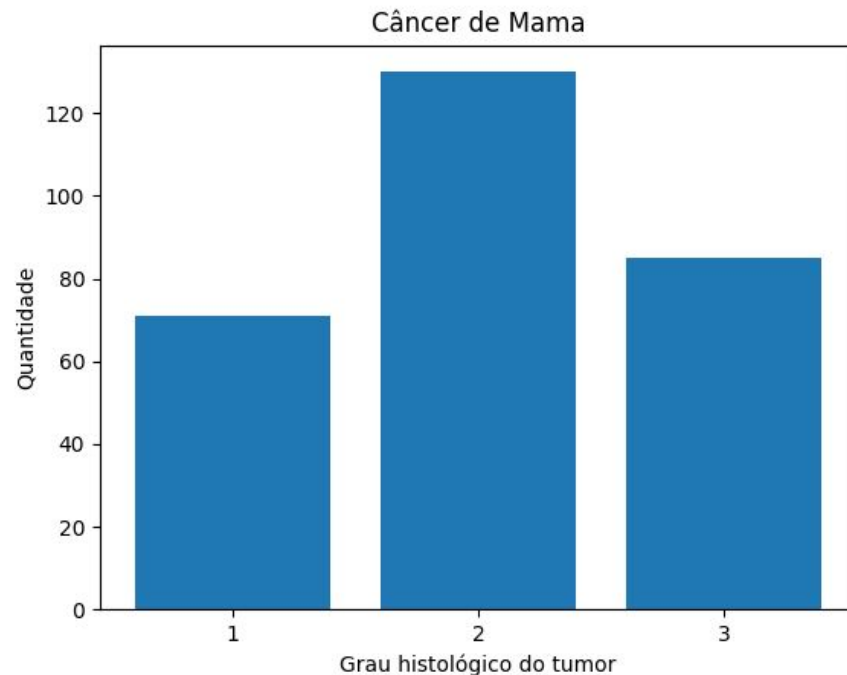
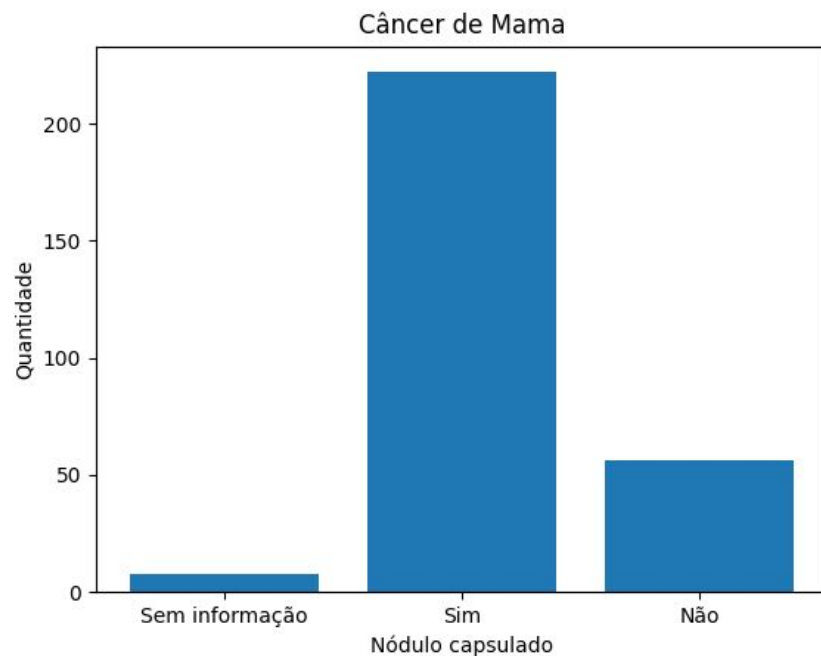
Base de dados 1



Base de dados 1

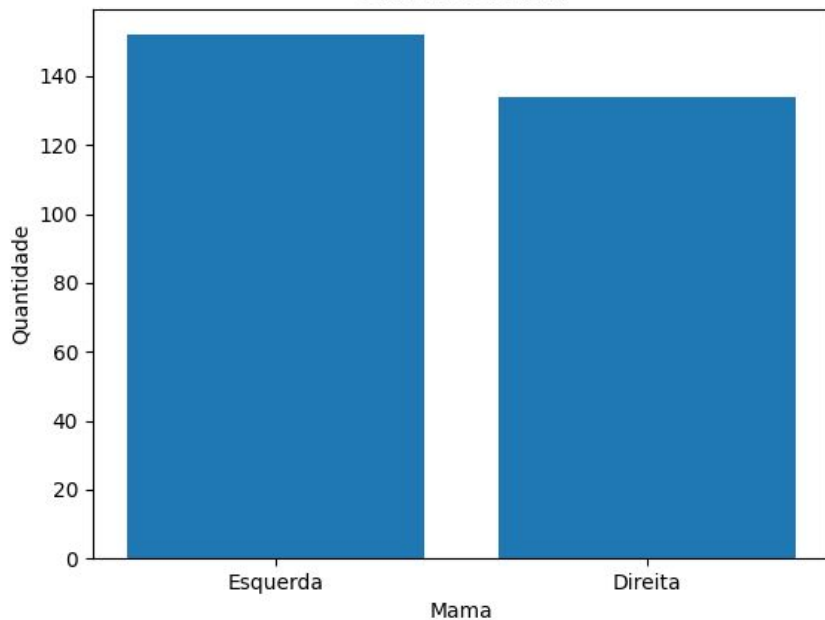


Base de dados 1

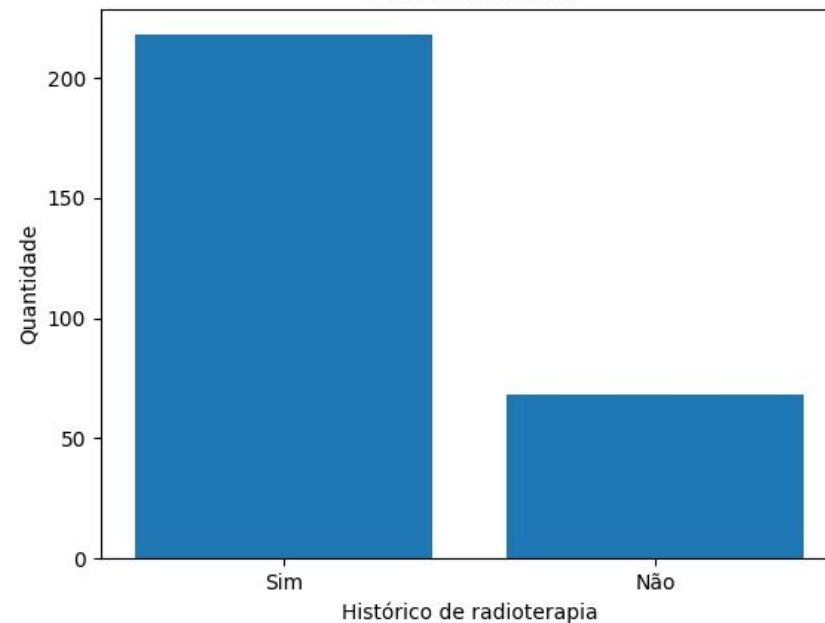


Base de dados 1

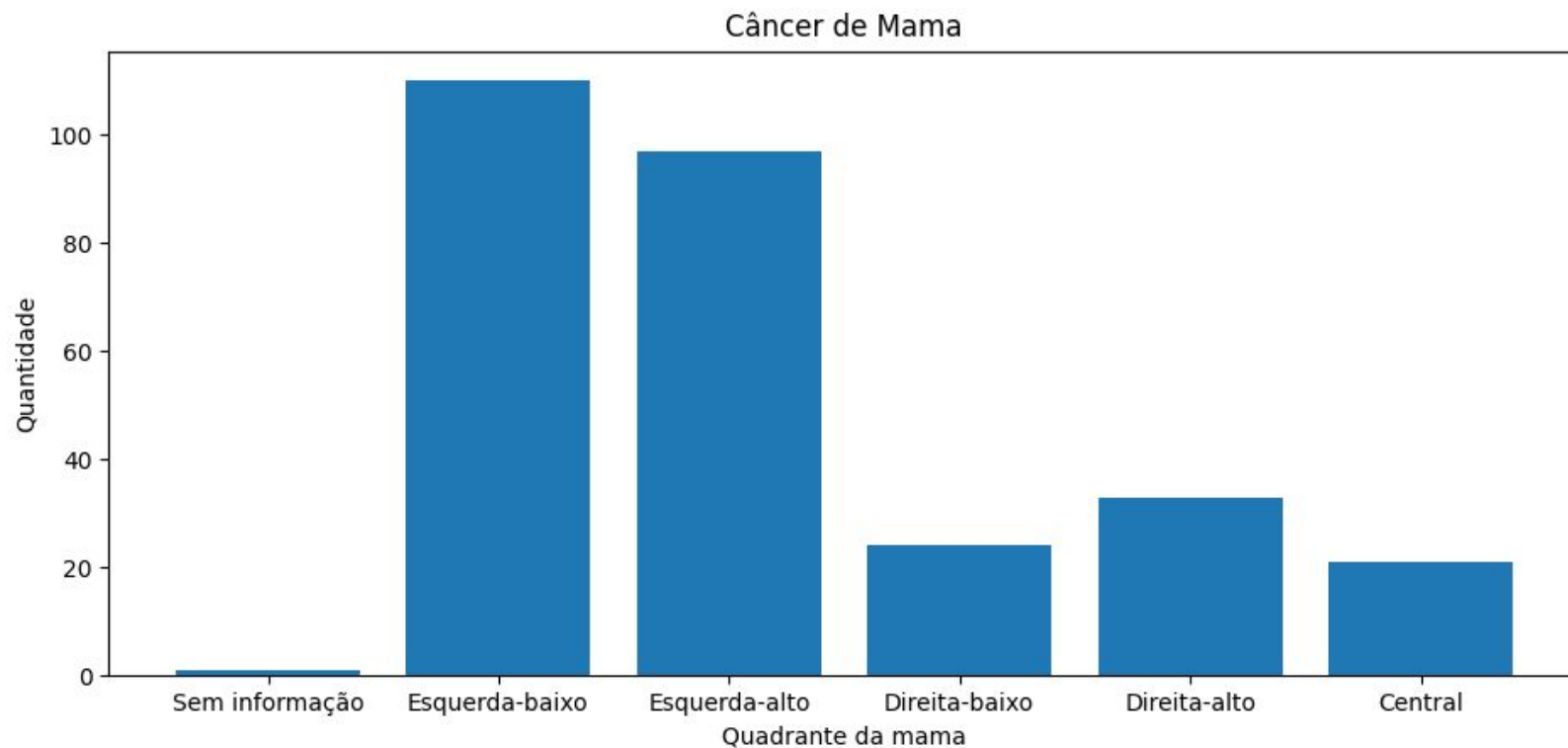
Câncer de Mama



Câncer de Mama



Base de dados 1



Experimento 1

Preparação do modelo:

- Perceptron de multicamada (MPL).
- 4 camada escondidas (12, 8, 4 e 1).
- Ativação: 'softmax' (3 primeiras) e 'relu' (última).
- Otimizador: 'adamax'.
- Perda: 'mean_squared_logarithmic_error'

Experimento 1

A base foi dividida em 3 partes:

1. Dados de treinamento (os primeiros 85).
2. Dados de avaliação (os 85 subsequentes).
3. Dados de predição (os últimos 85).

Dois treinamentos foram utilizados, ambas com *learn rate* de 0.12:

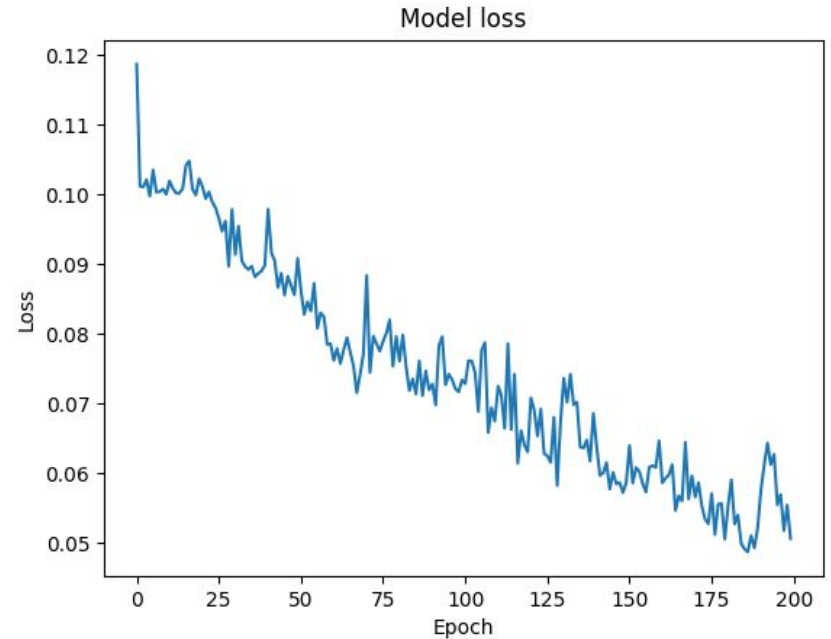
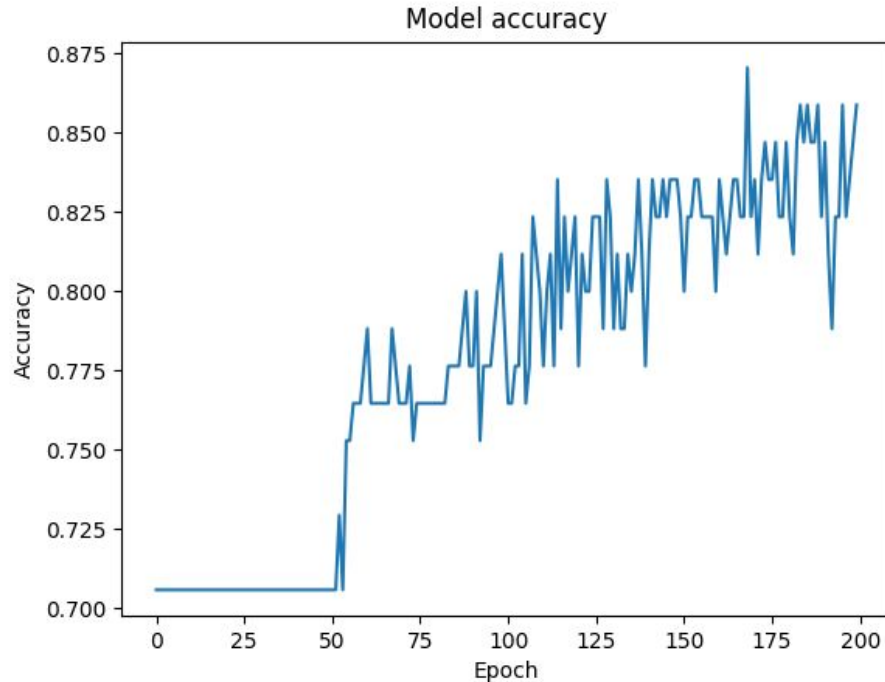
1. 200 épocas e *batch size* de 16.
2. 125 épocas e *batch size* de 64.

Experimento 1

Resultados:

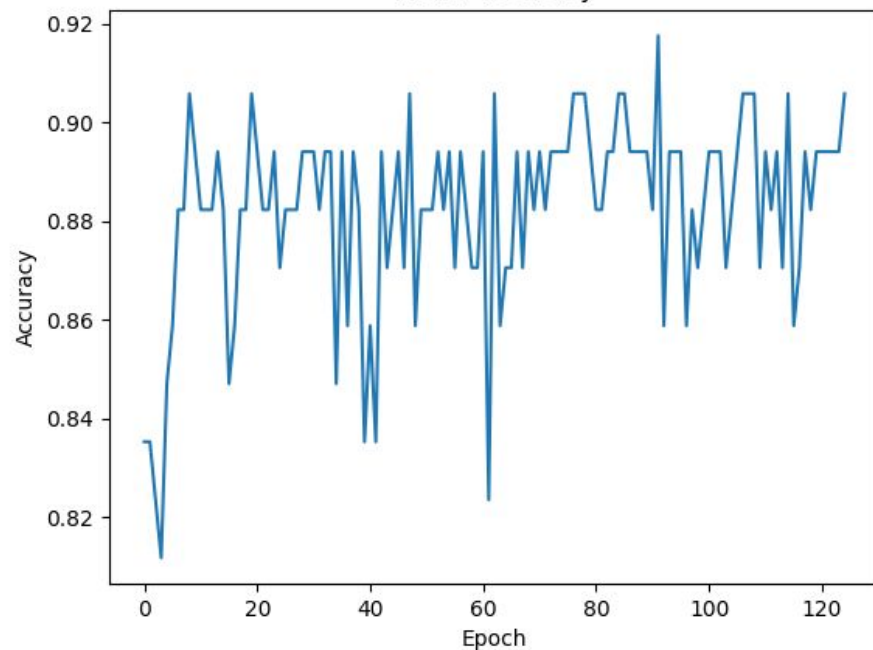
- Treinamento:
 - 1 - Taxa de acerto médio de 77,81% com um erro médio de 0,07.
 - 2 - Taxa de acerto médio de 88,27% e um erro de 0,05.
- Avaliação:
 - Base de avaliação: 74,42% de acerto e erro de 0.11
 - Base de predição: 39% de acerto.

Experimento 1

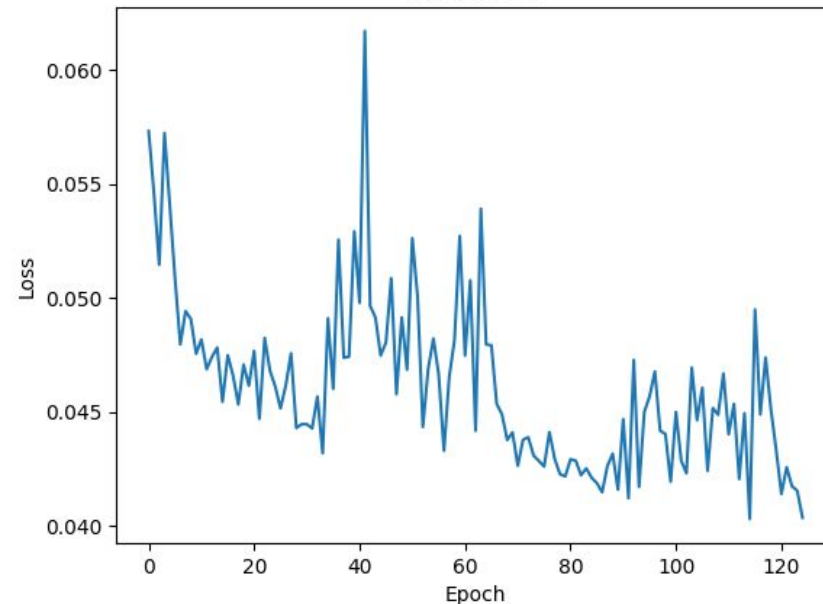


Experimento 1

Model accuracy



Model loss



Experimento 1

Conclusão:

Apesar dos bons resultados apresentados no treinamento, quando submetida a outro conjunto de dados, a rede não apresentou resultados satisfatórios. Sendo assim, não podendo ser aplicada para tentar indicar se o paciente irá ter reincidência do câncer de mama.

Base de dados 2

Título : Tic Tac Toe Endgame

Dados obtidos a partir das possibilidades de jogada do jogo da velha.

Agradecimento a David W. Aha por disponibilizar os dados.

Base de dados 2

Consiste em 958 instâncias que possuem 9 atributos de caracterização e 1 de classificação.

Características: { x, o, b }

- Esquerdo superior
- Meio superior
- Direito superior
- Esquerdo meio
- Meio meio
- Direito meio
- Esquerdo inferior
- Meio inferior
- Direito inferior

Classificação: Positivo ou Negativo

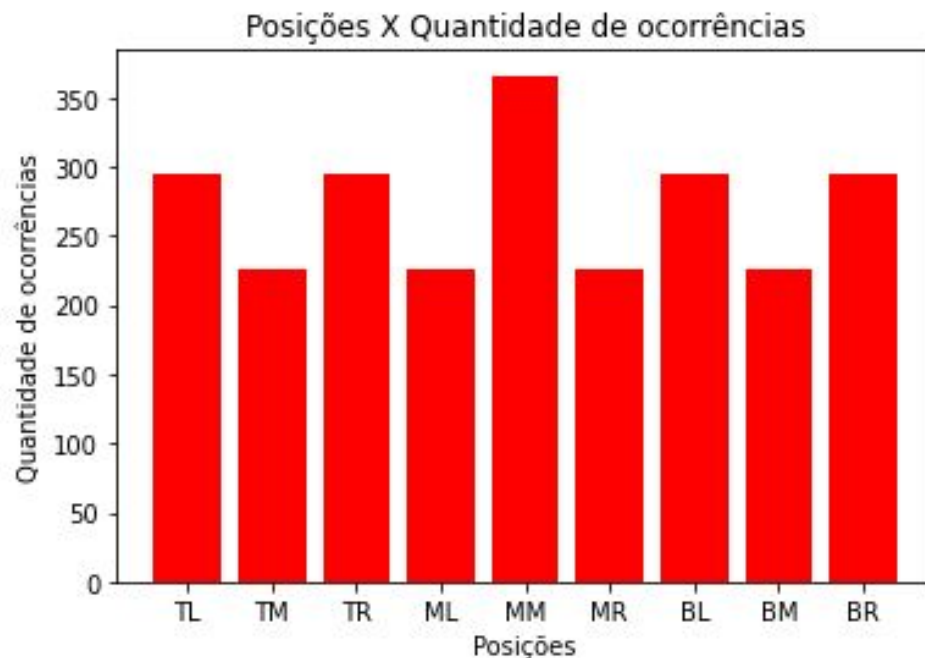
Base de dados 2

Ajustes para utilização:

- Valores de caracterização:
 - X -> 1
 - O -> 0
 - B -> 2

- Valores de Classificação:
 - Positive - > 1
 - Negative -> 0

Base de dados 2



Positivo 626 0.6534446764091858
Negativo 332 0.3465553235908142

Total de dados: 958
Total ocorrências: 2446

Ocorrência por posição:

295 225 295
225 366 225
295 225 295

Porcentagem da ocorrência:
0.1206 0.0919 0.1206
0.0919 0.1496 0.0919
0.1206 0.0919 0.1206

Experimento 2

Foi utilizado perceptron de multicamada (MLP), com 3 camadas escondidas. Além disso, função de ativação sigmoide, com 'RMSprop' e lr de 0.12 como otimizador da rede. Para o treinamento, 250 épocas e binary_crossentropy como parâmetro de perda.

Experimento 2

A base foi dividida em 3 partes:

1. Dados de treinamento.
2. Dados de predição.
3. Dados para validação

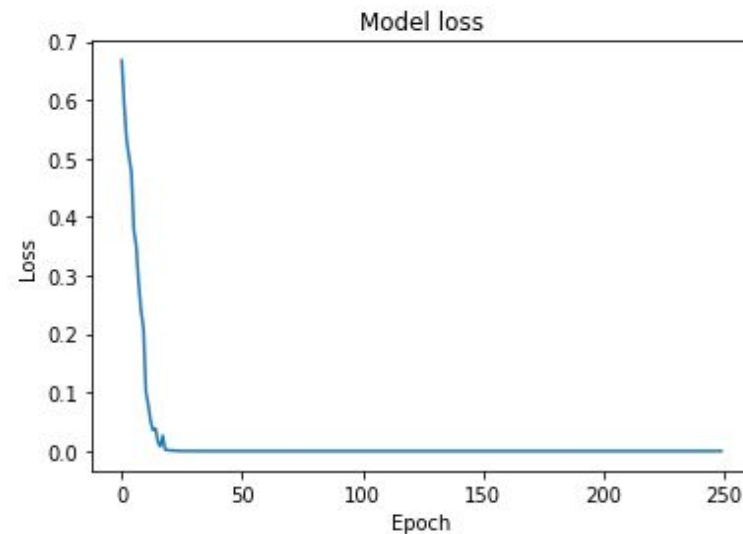
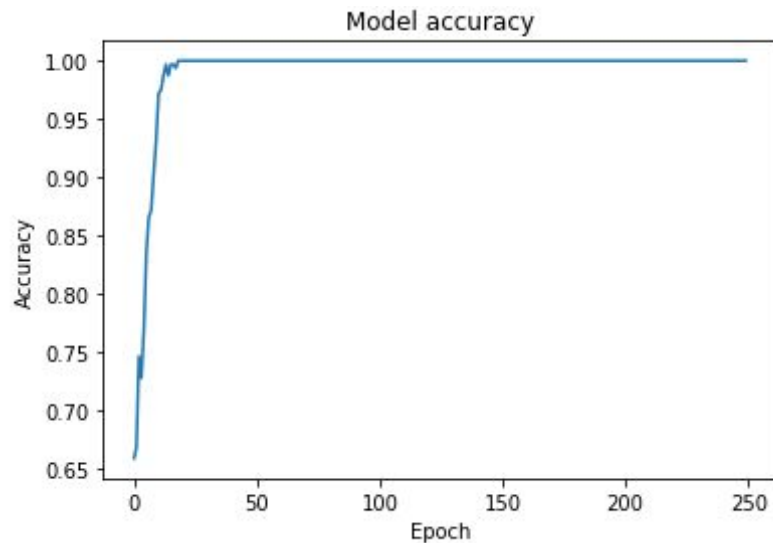
O treinamento foi definido com 250 épocas e uma taxa de aprendizado de 0,12.

Experimento 2

Resultados:

- A rede no treinamento obteve uma taxa de acerto médio de 99,15% e um erro médio de 0,02.
- Avaliação:
 - Base de treinamento: 69,91% de acerto e erro de 5,82

Experimento 2



Experimento 2

Conclusão:

O treinamento apresentou bons resultados. Entretanto, quando submetida a outro conjunto de dados, a rede obteve queda significativa na sua capacidade assertiva. Sendo assim, a mesma possui uma relativa capacidade de identificar se é possível o jogador “X” ganhar o jogo, mas não se deve confiar 100% nela.

Referências

<https://archive.ics.uci.edu/ml/datasets/Breast+Cancer>

<http://archive.ics.uci.edu/ml/datasets/Tic-Tac-Toe+Endgame>

<http://d.researchbib.com/f/5nnJcwp21wYzAioF9xo2AmY3OupTIlpl9XLJ51LKW5ZwNkAP9JZ0xkZwNkAQNIYaOxMt.pdf>

<https://keras.io>