

Mammographic Mass Data Set

Jefferson Pires de Sousa





Sobre o dataset

- Mammographic Mass Data Set
- Discriminação de massas mamográficas em benignas e malignas com base em atributos de BI-RADS e a idade do paciente.
- Número de linhas: 916 (80% para treino e 20% para teste)
- Número de colunas: 6



Sobre o dataset

- **Objetivo:** Descobrir se uma massa mamográfica é benigna ou maligna com base em atributos de BI-RADS e a idade do paciente.
- Este conjunto de dados pode ser usado para prever a gravidade (benigna ou maligna) de uma lesão de massa mamográfica, com base em atributos BI-RADS e da idade do paciente.



Sobre o dataset

- **Atributos:**

Avaliação de BI-RADS: é o nome de um sistema padronizado, utilizado para uniformizar os relatos de radiologia quando se analisam as imagem de mamografia.

Idade: idade do paciente em anos

Forma: massa forma

Margem: margem de massa

Densidade: densidade de massa

Gravidade: benigna = 0 ou maligna = 1



Análise dos dados

- **Colunas:**

1. Avaliação de BI-RADS: 1 a 5 (ordinal)
2. Idade: idade do paciente em anos (inteiro)
3. Forma: massa forma: rodada = 1 oval = 2 lobular = 3 irregular = 4 (nominal)
4. Margem: margem de massa: circunscrito = 1 microlobulado = 2 obscurecido = 3 mal definido = 4 espiculados = 5 (nominal)
5. Densidade: densidade de massa alta = 1 iso = 2 baixa = 3 contem gordura = 4 (ordinal)
6. Gravidade: benigna = 0 ou maligna = 1 (binominal)



Análise dos dados

- Campos Nulos:
 1. BI-RADS assessment: 2
 2. - Age: 5
 3. - Shape: 31
 4. - Margin: 48
 5. - Density: 76
 6. - Severity: 0



Análise dos dados

Variáveis categóricas: BI-RADS, Age, Shape, Margin, Density

Campos novos criados: Nenhum

Campos descartados: BI-RADS



Análise dos dados

- Técnica de análise : Matriz de Confusão

```
from sklearn.metrics import confusion_matrix  
confusion_matrix(teste_marcacoes, resultado)
```

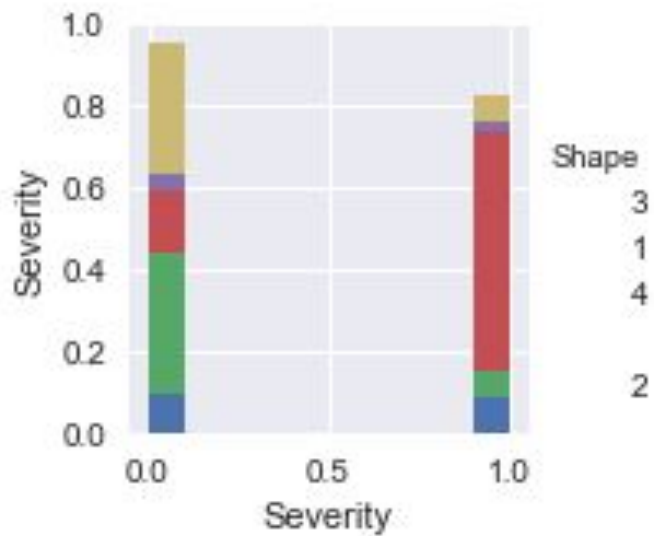
Acerto MultinomialNB: 80.21

Acerto MultinomialNB com K-fold:80.21%

```
array([[80, 20],  
       [18, 74]], dtype=int64)
```

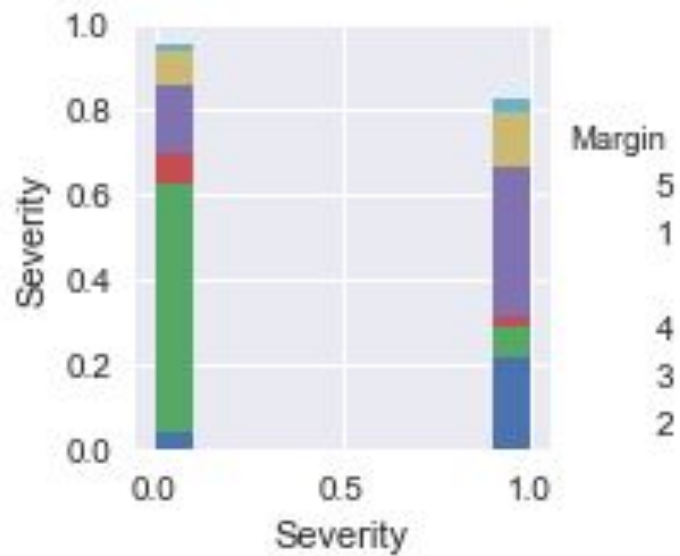



Severidade x Forma



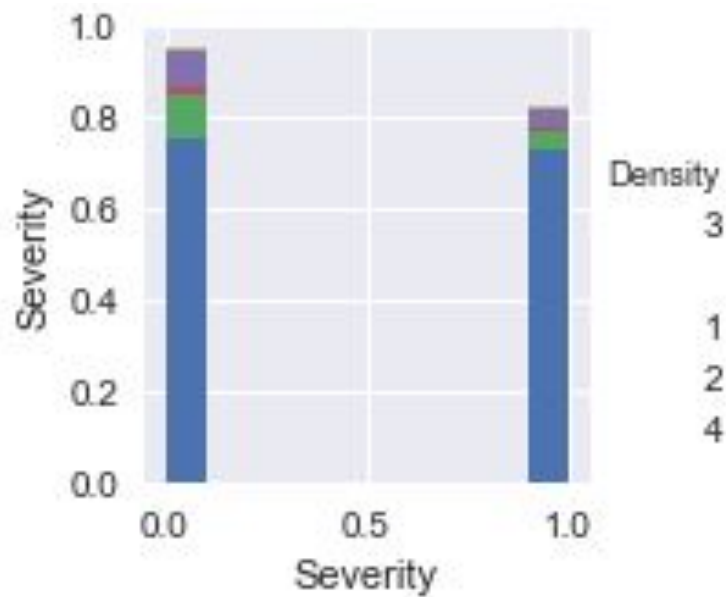


Severidade x Margem

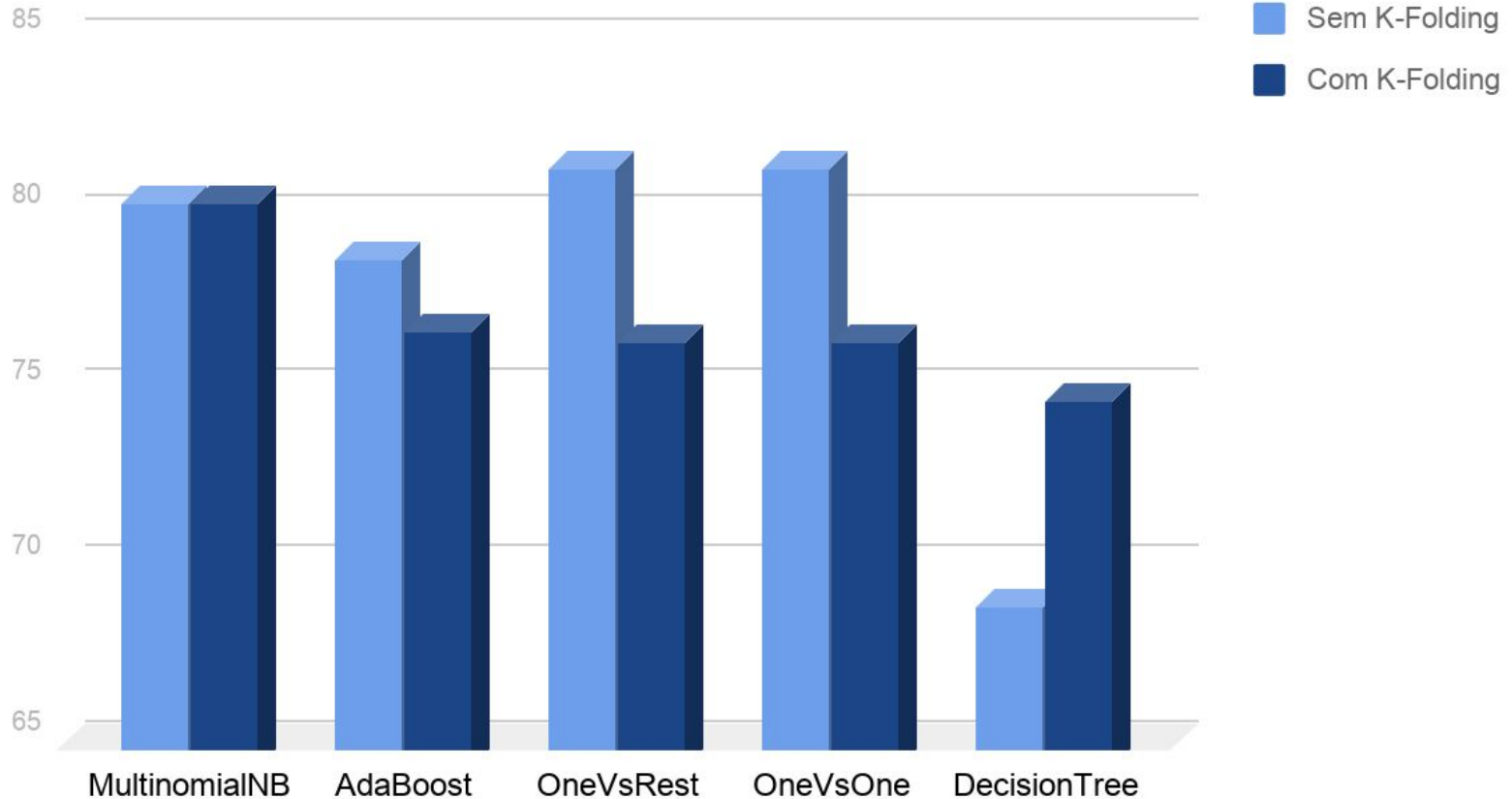




Severidade x Densidade



Execução dos algoritmos





Referências

<https://archive.ics.uci.edu/ml/datasets/Mammographic+Mass>

<https://www.ncbi.nlm.nih.gov/pubmed/18072480>

Mammographic Mass Data Set

Jefferson Pires de Sousa

