

# Relatório Visão Computacional TA4

Eduardo Mathias e Pedro H. Kochinski  
Departamento de Informática  
Universidade Federal do Paraná – UFPR  
Curitiba, Brasil  
ems19@inf.ufpr.br — phks20@inf.ufpr.br

## I. INTRODUÇÃO

Este relatório apresenta uma visão geral das atividades realizadas no quarto projeto de Visão Computacional. A atividade realizada foi fazer experimentos de classificação em cima de um dataset muito conhecido que contém dados de comprimento e largura das sépalas e pétalas de 3 categorias de flores.

## II. DESCRIÇÃO DO PROBLEMA

O dataset que vamos trabalhar em cima, 'Iris Dataset', é um dos datasets mais conhecidos na literatura de reconhecimento de padrões. Os dados desse dataset consistem em 150 amostras de três espécies de flores Íris (50 para cada classe). A primeira coluna do dataset representa o comprimento da sépala, a segunda coluna representa a largura da sépala, a terceira coluna representa o comprimento da pétala, a quarta coluna representa a largura da pétala, e por fim, a última coluna representa a classe da Íris (Iris setosa, Iris versicolor, Iris virginica). As colunas 1-4 são dados em cm e a última coluna é dado em string. Nesse projeto, utilizamos do sci-kit-learn para classificar estas instâncias de acordo com as suas espécies de íris, que serão distinguidas com base nas suas medidas. A imagem das espécies de íris é apresentada de seguida:

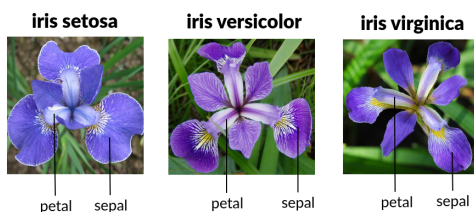


Figura 1: Espécies de Íris presentes no dataset

## III. INFERÊNCIA DE DADOS

Após a leitura do dataset, utilizando a função describe() da biblioteca Pandas temos a seguinte informação sobre os dados: 150 amostras para cada coluna, e,

- Comprimento de Sépala
  - Média = 5.843333
  - Desvio Padrão = 0.828066
  - Menor valor = 4.300000
  - Quantil Inferior = 5.100000

	sepal_length	sepal_width	petal_length	petal_width
count	150.000000	150.000000	150.000000	150.000000
mean	5.843333	3.054000	3.758667	1.198667
std	0.828066	0.433594	1.764420	0.763161
min	4.300000	2.000000	1.000000	0.100000
25%	5.100000	2.800000	1.600000	0.300000
50%	5.800000	3.000000	4.350000	1.300000
75%	6.400000	3.300000	5.100000	1.800000
max	7.900000	4.400000	6.900000	2.500000

Figura 2: Descrição algébrica do Iris Dataset

- Mediana = 5.800000
- Quantil Superior = 6.400000
- Maior valor = 7.900000
- Largura de Sépala
  - Média = 3.054000
  - Desvio Padrão = 0.433594
  - Menor valor = 2.000000
  - Quantil Inferior = 2.800000
  - Mediana = 3.000000
  - Quantil Superior = 3.300000
  - Maior valor = 4.400000
- Comprimento da Pétala
  - Média = 3.758667
  - Desvio Padrão = 1.764420
  - Menor valor = 1.000000
  - Quantil Inferior = 1.600000
  - Mediana = 4.350000
  - Quantil Superior = 5.100000
  - Maior valor = 6.900000
- Largura da Pétala
  - Média = 1.198667
  - Desvio Padrão = 0.763161
  - Menor valor = 0.100000
  - Quantil Inferior = 0.300000
  - Mediana = 1.300000
  - Quantil Superior = 1.800000
  - Maior valor = 2.500000

Abaixo temos os Histogramas das colunas Comprimento de Sépala, Largura de Sépala, Comprimento de Pétala e Largura

de Pétala. Podemos perceber que o comprimento e a largura das sépalas formam uma distribuição normal e o comprimento das pétalas e a largura das pétalas têm duas regiões bem separadas, devido às medidas das diferentes espécies.

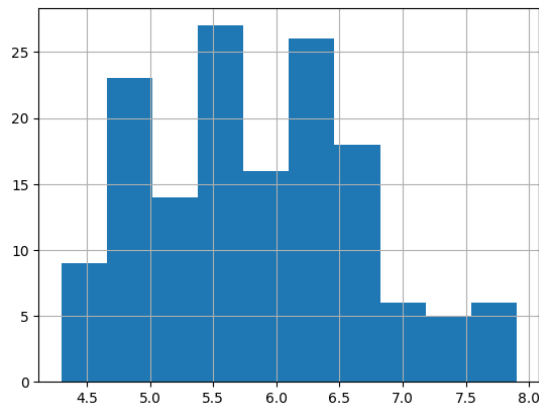


Figura 3: Distribuição dos valores do Comprimento de Sépala em Histograma

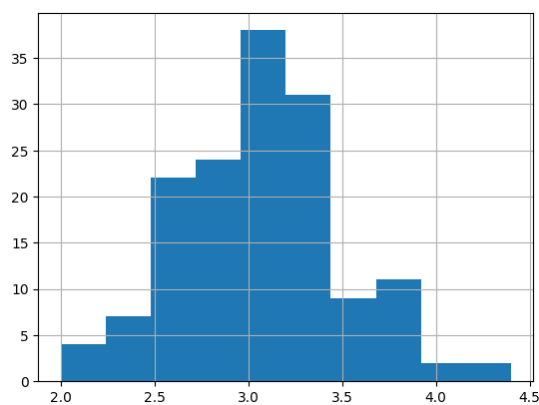


Figura 4: Distribuição dos valores da Largura de Sépala em Histograma

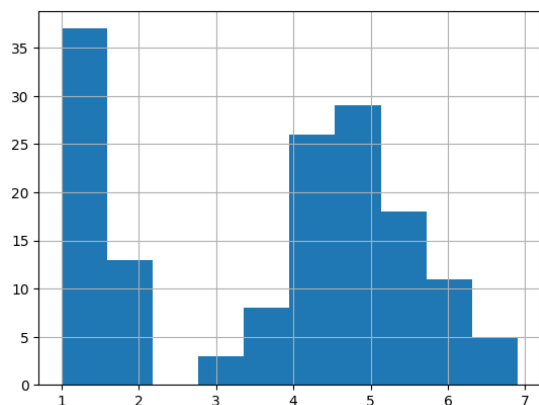


Figura 5: Distribuição dos valores do Comprimento de Pétala em Histograma

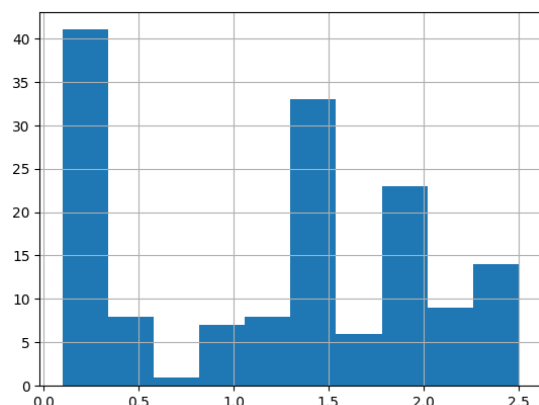


Figura 6: Distribuição dos valores da Largura de Pétala em Histograma

Podemos ver também plotado em um gráfico a relação entre colunas para cada classe. Nesses gráficos, podemos perceber que a classe iris-setosa é facilmente separável das outras 2 classes, e, no gráfico do comprimento e da largura das pétalas, as classes não se sobrepõem.

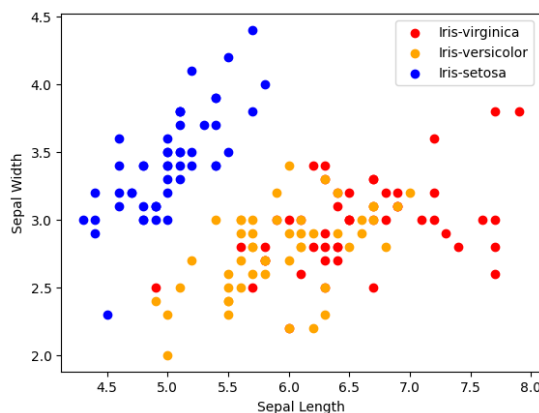


Figura 7: Relação entre Comprimento e Largura de Sépala

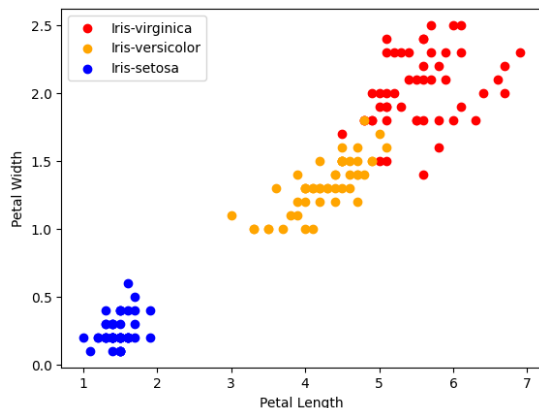


Figura 8: Relação entre Comprimento e Largura de Pétala

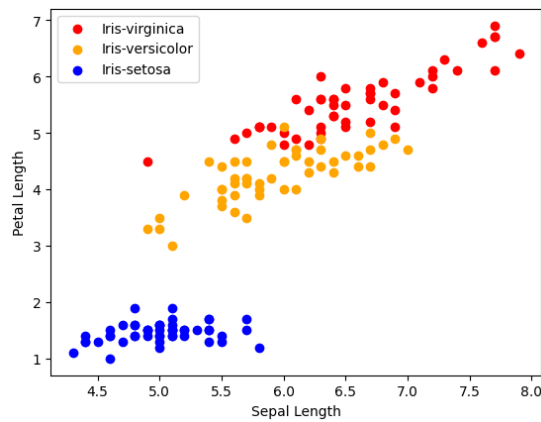


Figura 9: Relação entre Comprimento de Pétala e Sépala

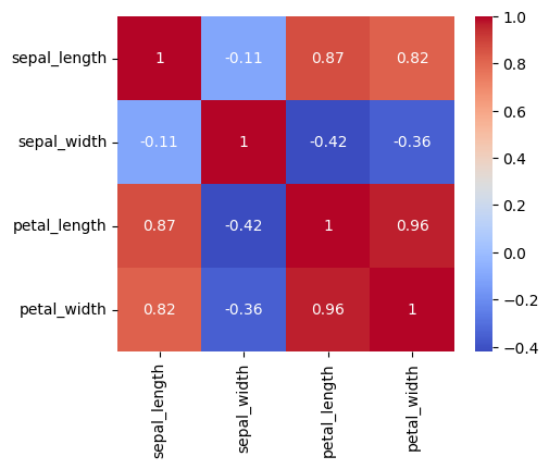


Figura 11: HeatMap de Correlação

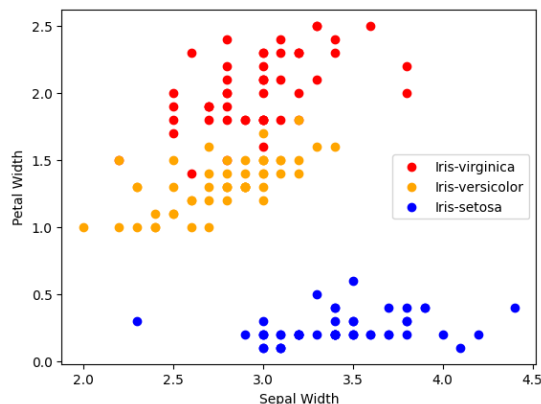


Figura 10: Relação entre Largura de Pétala e Sépala

Por fim, para análise de dados, também podemos plotar uma matriz de Correlação em um HeatMap para melhor visualização. Esse HeatMap nos mostra que o comprimento da pétala e a largura da pétala têm uma correlação positiva elevada de 0,96, se o valor do comprimento da pétala aumenta, a largura da pétala também aumenta, o comprimento da sépala tem uma correlação positiva elevada com o comprimento da pétala e a largura da pétala e a largura da sépala tem uma correlação negativa com o comprimento da pétala e a largura da pétala.

#### IV. CLASSIFICAÇÃO DOS DADOS

Para treinamento e classificação de dados, normalmente lida-se com conjuntos de dados que contêm várias labels numa ou em mais do que uma coluna. Estes labels podem ter a forma de palavras ou de números. Para nosso projeto, realizamos a codificação de labels das classes de string para forma numérica, para facilitar no processo de treinamento.

Com a função `train_test_split()` dividimos a data para treinamento e teste respeitando 80% para treinamento e 20% para teste. Realizando uma regressão linear com 1020 de número máximo iterações, obtivemos uma acurácia 96.66666666666667%. E, realizando um treino com KNN tivemos uma acurácia de 100%.

#### V. CONCLUSÃO

Esse relatório apresentou a classificação e inferência de dados do dataset clássico Iris Dataset. No processo de inferência de dados, conseguimos perceber visualmente que uma classe de Iris era separável das demais e existia diversas correlações pelas quatro colunas de características. Utilizando 80% de dados pra treinamento e 20% para teste, conseguimos uma acurácia de 100% no KNN.