



UNIVERSIDADE FEDERAL DE
UBERLÂNDIA
INTELIGÊNCIA ARTIFICIAL

Rafael Henrique Guimarães
12021BCC002
Gabriel Peres da Silva
12021BCC015

**Trabalho de Inteligência Artificial –
Classificação de Base de Dados ÍRIS
utilizando Árvores de Decisão e KNN**

Uberlândia

2023

Introdução

Este trabalho visa aplicar as técnicas de aprendizagem de máquina Árvores de Decisão e KNN na classificação da base de dados Iris, que contém informações sobre três espécies de flores.



Este dataset é composto por quatro características:

Comprimento da Sépala, Largura da Sépala, Comprimento da Pétala, Largura da Pétala,
como podemos observar:

```
[[5.1 3.5 1.4 0.2]
[4.9 3. 1.4 0.2]
[4.7 3.2 1.3 0.2]
[4.6 3.1 1.5 0.2]
[5. 3.6 1.4 0.2]
[5.4 3.9 1.7 0.4]
[4.6 3.4 1.4 0.3]
[5. 3.4 1.5 0.2]
[4.4 2.9 1.4 0.2]
[4.9 3.1 1.5 0.1]
[5.4 3.7 1.5 0.2]
[4.8 3.4 1.6 0.2]
[4.8 3. 1.4 0.1]]
```

Dividimos a base em três experimentos distintos(A, B e C) e avaliamos a eficácia dos métodos de classificação:

- Primeiro Experimento: Treinamento com (A+B), Teste com (C)
- Segundo Experimento: Treinamento com (A+C), Teste com (B)
- Terceiro Experimento: Treinamento com (C+B), Teste com (A)

Calcularemos métricas como acurácia, sensibilidade, especificidade e precisão para cada experimento, utilizando as técnicas Árvores de Decisão e KNN. Além disso, apresentaremos a estrutura da árvore de decisão para cada experimento. Ao final, forneceremos as médias das métricas para uma análise abrangente da performance dos métodos de classificação na **base Iris**.

Experimentos Realizados

Para treinarmos o nosso modelo, utilizamos Árvores de Decisão e KNN, logo obtemos esses resultados:

```
Métricas do teste 1:
Acurácia: [0.92, 0.98]
Sensitividade: [0.92, 0.98]
Especificidade: [0.92, 0.9811111111111112]
Precisão: [0.9595959595959597, 0.98989898989899]

Métricas do teste 2:
Acurácia: [0.96, 0.96]
Sensitividade: [0.96, 0.96]
Especificidade: [0.96, 0.9644444444444444]
Precisão: [0.9800950683303625, 0.9803921568627452]

Métricas do teste 3:
Acurácia: [0.92, 1.0]
Sensitividade: [0.92, 1.0]
Especificidade: [0.9352380952380952, 1.0]
Precisão: [0.9595959595959597, 1.0]

Médias das Métricas:
Acurácia: [0.93333333 0.98      ]
Sensitividade: [0.93333333 0.98      ]
Especificidade: [0.966429  0.99009705]
Precisão: [0.9384127  0.98185185]
```

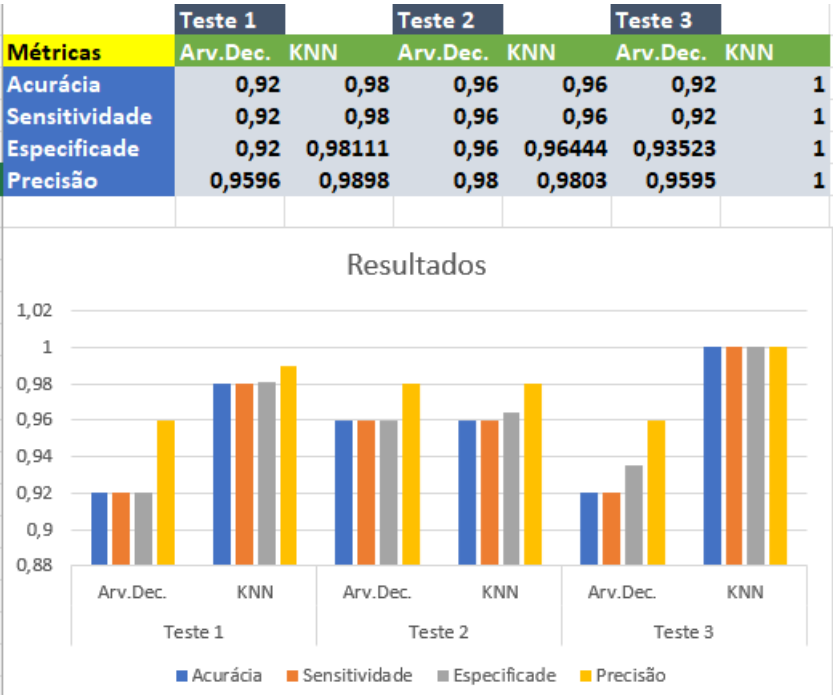
Análise dos Dados

Acurácia: Na classificação da base de dados Iris, a acurácia medirá a proporção de predições corretas, ou seja, a precisão global do modelo ao classificar as três espécies de flores. Cada experimento (Treinamento A+B e Teste C, Treinamento A+C e Teste B, Treinamento C+B e Teste A) resultará em uma acurácia específica para cada método de classificação (Árvores de Decisão e KNN).

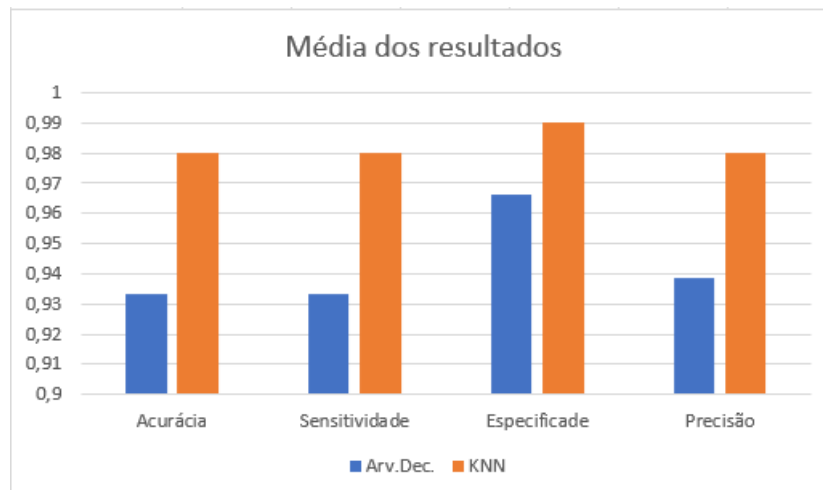
Sensitividade (Recall): A sensibilidade, no contexto da Iris, avaliará a capacidade do modelo em identificar corretamente cada espécie de flor. Para cada experimento, a sensibilidade será calculada para cada classe (setosa, versicolor e virginica), fornecendo informações sobre a capacidade de capturar todos os exemplos positivos de cada espécie.

Especificidade: Como a especificidade é mais relevante em problemas de classificação binária, para o caso da Iris, onde há três classes, podemos adaptar a métrica para cada classe. Podemos calcular a especificidade para cada espécie, avaliando a capacidade do modelo em identificar corretamente os exemplos negativos (ou seja, as outras duas espécies) para cada experimento.

Precisão: A precisão, ao lidar com a base de dados Iris, indicará a qualidade das predições positivas para cada espécie de flor. Para cada experimento, a precisão será calculada para cada classe, fornecendo insights sobre a confiabilidade das predições positivas para cada método de classificação.



	Média	
Métricas	Arv.Dec.	KNN
Acurácia	0,93333	0,98
Sensitividade	0,93333	0,98
Especificade	0,96643	0,99
Precisão	0,93841	0,98



- Os métodos de classificação, em média, apresentaram boas taxas de acurácia e sensibilidade, indicando uma capacidade satisfatória de classificação das espécies de flores da Iris.
- A especificidade, mostra a habilidade do modelo em identificar corretamente os exemplos negativos para cada espécie.
- A precisão destaca a qualidade das predições positivas, sendo relevante para avaliar a confiabilidade das classificações positivas para cada classe.

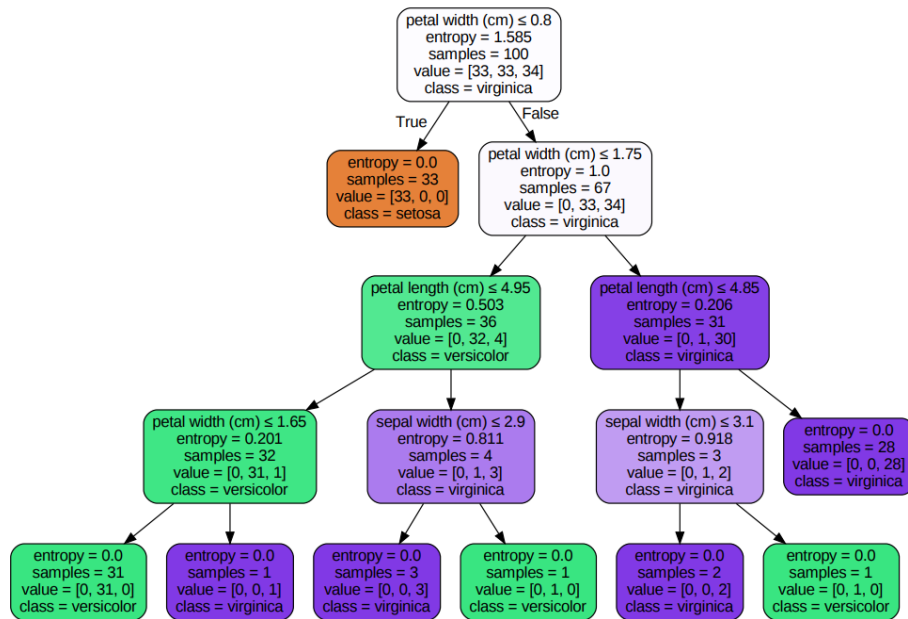
Primeiro Experimento

Treinamento (A+B) e Teste (C)



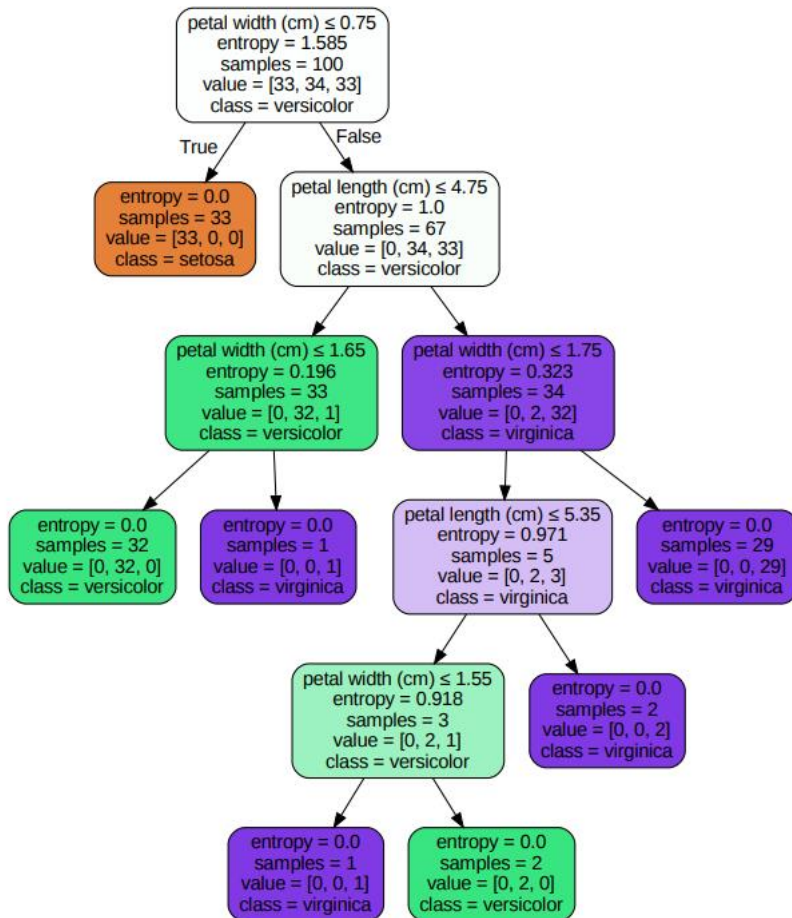
Segundo Experimento

Treinamento (A+C) e Teste (B)



Terceiro Experimento

Treinamento (C+B) e Teste (A)



Linguagem Utilizada: Python

Biblioteca: Scikit- Learn



Conclusão

Durante o trabalho foi explorado a aplicação de duas técnicas de aprendizado de máquina, Árvores de Decisão e KNN, na classificação da base de dados Iris. Esta base, desenvolvida por Ronald Fisher, contém medidas morfológicas de três espécies de flores Iris.

Os resultados dos experimentos, utilizando a estratégia de divisão Treinamento/Teste com três conjuntos distintos (A, B, C), revelaram desempenhos consistentes dos métodos de classificação. As métricas de avaliação, como acurácia, sensibilidade, especificidade e precisão, foram calculadas para cada experimento e suas médias apresentam uma visão geral do desempenho dos modelos.

Destaca-se a habilidade dos modelos em classificar corretamente as espécies de Iris, evidenciada pelas altas taxas de acurácia e sensibilidade. A análise da especificidade adaptada para o contexto de múltiplas classes forneceu insights sobre a capacidade dos modelos em identificar corretamente exemplos negativos. A precisão, por sua vez, indicou a confiabilidade das previsões positivas para cada classe.

Por fim, os resultados positivos sugerem que tanto Árvores de Decisão quanto KNN são métodos eficazes na classificação da base Iris, proporcionando um alicerce sólido para a utilização dessas técnicas em problemas semelhantes de classificação de padrões florais.