



**Universidade Federal de Mato Grosso - UFMT**  
**Faculdade de Engenharia**

Inteligência Artificial

**Trabalho Prático - IRIS**

Autoras - RGA

Gabriely Nascimento Souza - 201921902009

Maria Eduarda Oliveira Dias - 201911902020

Cuiabá, 4 de junho de 2023.

---

# Sumário

<b>1</b>	<b>Introdução</b>	<b>4</b>
<b>2</b>	<b>Descrição do Problema</b>	<b>4</b>
<b>3</b>	<b>Propostas de Soluções</b>	<b>4</b>
3.1	K-Nearest Neighbors (KNN) . . . . .	5
3.2	Support-Vector Machine (SVM) . . . . .	5
3.3	Árvore de decisão . . . . .	6
<b>4</b>	<b>Resultados</b>	<b>7</b>
<b>5</b>	<b>Conclusão</b>	<b>8</b>
<b>6</b>	<b>Referências Bibliográficas</b>	<b>9</b>

---

## Lista de Figuras

1	KNN Graficamente . . . . .	5
2	SVM plot para um dataset IRIS . . . . .	6
3	Matriz de confusão KNN e SVM . . . . .	7
4	Matriz de confusão árvore de decisão . . . . .	8

---

# 1 Introdução

A base de dados IRIS é uma das mais populares e amplamente utilizadas na área de aprendizado de máquina e classificação. Essa base de dados consiste em informações sobre as características de diferentes espécies de flores iris, incluindo o comprimento e largura das pétalas e sépalas. O objetivo desse trabalho prático de inteligência artificial é explorar e aplicar algoritmos de classificação para categorizar corretamente as flores iris com base nessas características.

A base de dados IRIS é composta por 150 amostras, divididas igualmente em três classes: setosa, versicolor e virginica. Cada amostra possui quatro atributos: comprimento e largura das pétalas, e comprimento e largura das sépalas. Esses atributos são utilizados como entradas para os algoritmos de classificação, que serão treinados para aprender padrões e características distintas de cada espécie de flor iris.

## 2 Descrição do Problema

O problema abordado neste trabalho prático de inteligência artificial é o de classificar corretamente as espécies de flores iris com base em suas características morfológicas. A base de dados IRIS contém informações sobre o comprimento e largura das pétalas e sépalas de diferentes amostras de flores iris, e o objetivo é desenvolver um modelo de aprendizado de máquina capaz de identificar com precisão a espécie de uma flor iris com base nessas características. A classificação correta das flores iris é importante não apenas para a botânica, mas também para aplicações práticas, como a seleção de espécies adequadas para determinados fins agrícolas ou a identificação de flores em estudos de biodiversidade. Portanto, a utilização de técnicas de inteligência artificial para resolver esse problema pode trazer benefícios tanto para a pesquisa científica quanto para aplicações práticas na área de biologia e agronomia.

## 3 Propostas de Soluções

Nesse trabalho, serão explorados diferentes algoritmos de classificação, como K-Nearest Neighbors (KNN), Support Vector Machine (SVM) e árvores de decisão, a fim de encontrar o melhor modelo para a classificação precisa das flores iris. Serão realizadas análises comparativas dos resultados obtidos por cada algoritmo, levando em consideração métricas de desempenho como acurácia. O objetivo é encontrar o algoritmo mais adequado e eficiente para a tarefa de classificação das espécies de iris com base nas características fornecidas pela base de dados IRIS.

### 3.1 K-Nearest Neighbors (KNN)

Na área da estatística, o algoritmo de k-vizinhos mais próximos (k-NN) é um método de aprendizado supervisionado que não possui parâmetros fixos. Ele é utilizado tanto para classificação quanto para regressão. Em ambos os casos, o algoritmo utiliza os k exemplos de treinamento mais próximos de um conjunto de dados como entrada. A saída varia dependendo se o k-NN é aplicado para classificação ou regressão.

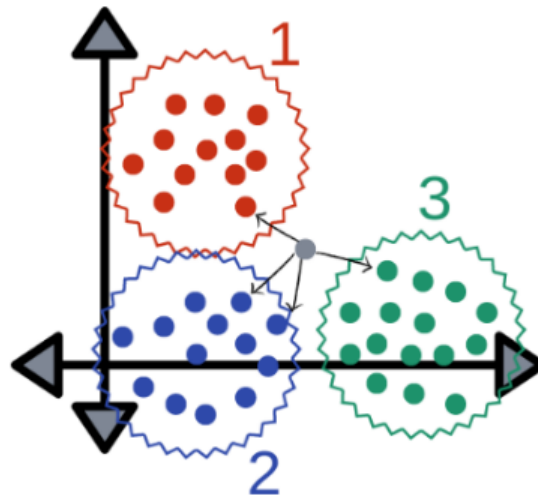


Figura 1: KNN Graficamente

No caso da classificação k-NN, a saída é uma associação de classe. Um objeto em questão é classificado com base em uma votação da maioria dos seus vizinhos mais próximos. O objeto é atribuído à classe mais comum entre os k vizinhos mais próximos (onde k é um valor inteiro positivo geralmente pequeno). Se k for igual a 1, o objeto é simplesmente atribuído à classe do único vizinho mais próximo.

Já na regressão k-NN, a saída é o valor da propriedade do objeto. Esse valor é calculado como a média dos valores das propriedades dos k vizinhos mais próximos. Dessa forma, o algoritmo busca determinar o valor estimado com base nos valores dos vizinhos mais próximos.

### 3.2 Support-Vector Machine (SVM)

Quando se trata da aplicação do SVM (Support Vector Machine) para o conjunto de dados IRIS, o algoritmo de treinamento constrói um modelo capaz de atribuir novos exemplos a uma das duas categorias existentes. O SVM é um classificador linear binário não probabilístico, embora existam técnicas que permitem sua utilização em uma configuração de classificação probabilística. O objetivo do SVM é mapear os exemplos de treinamento para pontos no espaço, de forma a maximizar a distância entre as duas categorias, criando uma lacuna entre elas. Com base na posição dos novos exemplos em relação a essa lacuna, é possível realizar previsões sobre a categoria a qual eles pertencem.

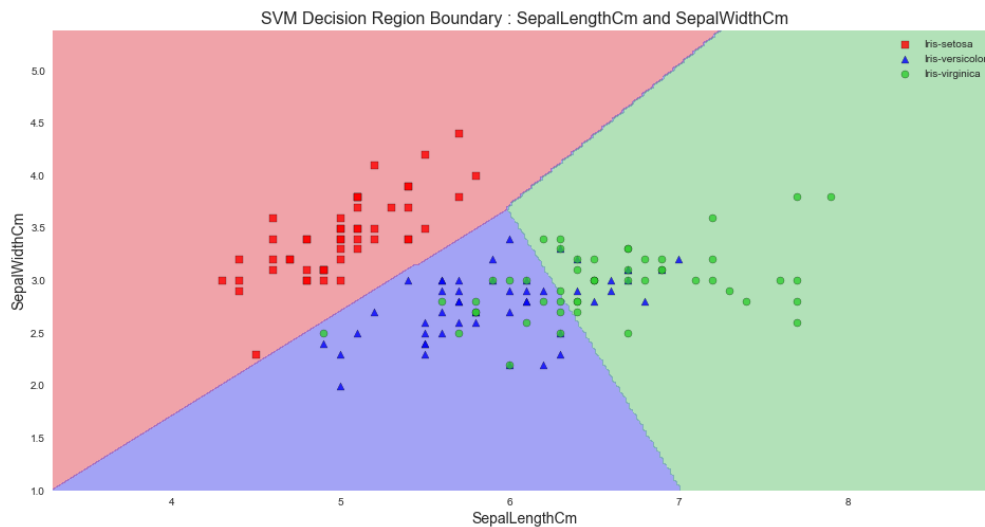


Figura 2: SVM plot para um dataset IRIS

O SVM se destaca por sua capacidade de lidar com problemas de classificação linearmente separáveis, como é o caso do conjunto de dados IRIS, em que as classes são bem distintas. Ele busca encontrar o hiperplano ótimo que melhor separa as categorias, maximizando a margem entre elas. Além disso, o SVM é capaz de lidar com conjuntos de dados de alta dimensionalidade, como é o caso do IRIS, em que as amostras são caracterizadas por múltiplas características (comprimento e largura das pétalas e sépalas). Isso torna o SVM uma abordagem poderosa e eficaz para classificar as diferentes espécies de flores no conjunto de dados IRIS.

### 3.3 Árvore de decisão

Árvore de decisão é uma técnica popular e eficaz para a classificação do conjunto de dados IRIS. Trata-se de um algoritmo de aprendizado supervisionado que constrói uma estrutura em forma de árvore, onde cada nó representa uma decisão com base em um atributo específico. No caso do IRIS, os atributos incluem o comprimento e a largura das pétalas e sépalas.

A construção da árvore de decisão envolve a seleção do atributo mais relevante em cada nó, de acordo com uma medida de pureza, como o índice de Gini ou a entropia. Com base nesse atributo, a árvore é dividida em ramos, representando as diferentes possibilidades ou classes do conjunto de dados IRIS, como Setosa, Versicolor e Virginica.

Uma das principais vantagens da árvore de decisão é a sua interpretabilidade, já que cada nó da árvore representa uma decisão compreensível para os humanos. Além disso, a árvore de decisão é capaz de lidar com dados categóricos e numéricos, tornando-a adequada para o conjunto de dados IRIS, que possui características mistas.

No contexto do IRIS, a árvore de decisão pode ser usada para classificar novas amostras de flores com base nas características das pétalas e sépalas. Ao percorrer a árvore de decisão, cada amostra é direcionada para o ramo correspondente, resultando na sua classificação correta.

A árvore de decisão tem se mostrado uma abordagem eficiente para a classificação do conjunto de dados IRIS, com bons resultados de acurácia. Sua simplicidade e interpretabilidade a tornam uma escolha popular para problemas de classificação em que é importante compreender e explicar o processo de tomada de decisão.

## 4 Resultados

Os resultados obtidos para a classificação do conjunto de dados IRIS utilizando os modelos KNN, SVM e árvore de decisão foram bastante promissores. O KNN e o SVM alcançaram uma acurácia de 100%, o que significa que foram capazes de classificar corretamente todas as amostras do conjunto de dados. Já a árvore de decisão apresentou uma acurácia de 97%, o que ainda é um resultado bastante satisfatório.

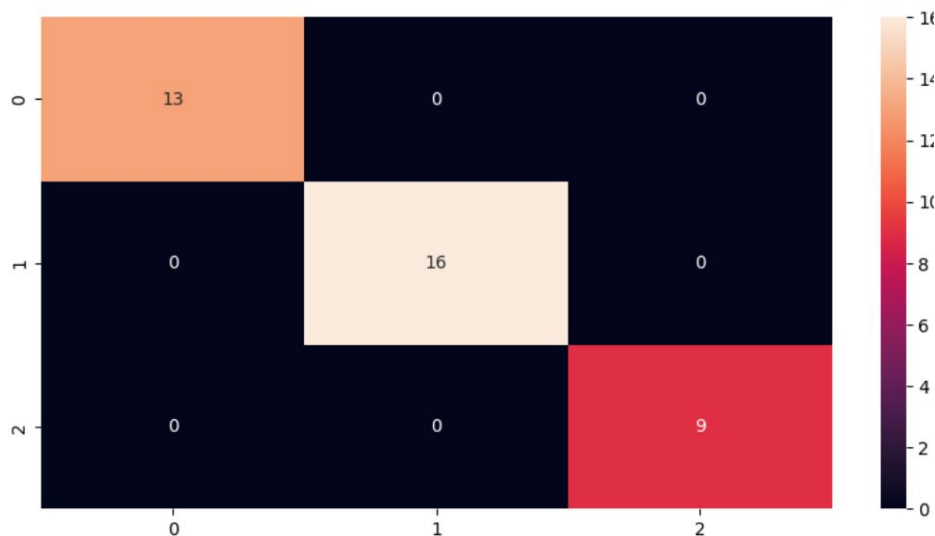


Figura 3: Matriz de confusão KNN e SVM

A alta acurácia alcançada pelo KNN e pelo SVM indica que esses modelos foram capazes de identificar corretamente as diferentes classes de flores presentes no conjunto de dados IRIS. Isso demonstra a eficácia desses algoritmos na tarefa de classificação, onde eles conseguiram encontrar os vizinhos mais próximos ou construir um hiperplano que separa de forma ótima as classes.

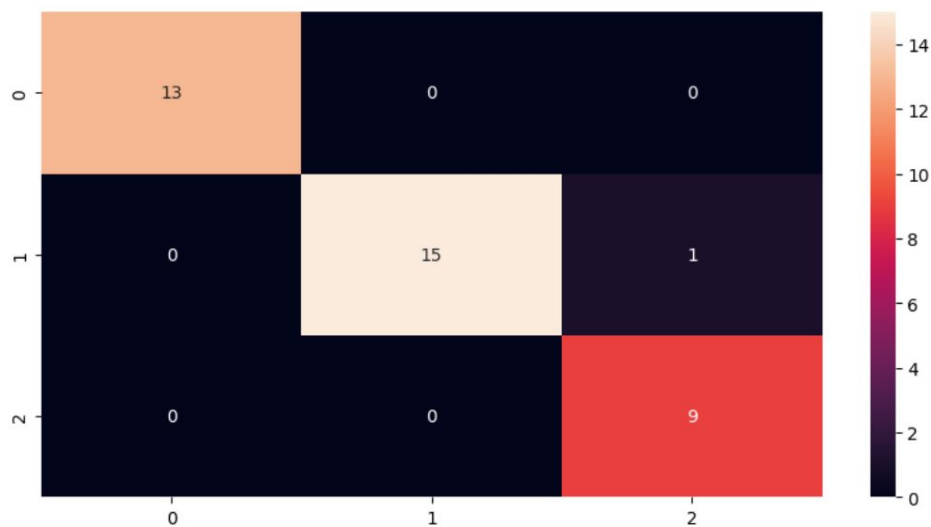


Figura 4: Matriz de confusão árvore de decisão

Por sua vez, a árvore de decisão também apresentou um desempenho notável, com uma acurácia de 97%. Embora ligeiramente inferior aos outros modelos, ainda foi capaz de classificar corretamente a maioria das amostras. A estrutura de decisão em forma de árvore permitiu a identificação de padrões e relações entre os atributos das flores, contribuindo para uma boa taxa de acerto na classificação.

Esses resultados sugerem que os modelos KNN, SVM e árvore de decisão são adequados para a tarefa de classificação do conjunto de dados IRIS. Eles demonstraram a capacidade de discernir entre as diferentes espécies de flores com alta precisão. Essas descobertas podem ter diversas aplicações práticas, como a identificação automática de espécies de flores em estudos botânicos ou a classificação de amostras em programas de conservação ambiental.

## 5 Conclusão

Em conclusão, os resultados obtidos neste estudo comprovam a eficácia dos modelos de classificação KNN, SVM e árvore de decisão na tarefa de classificação do conjunto de dados IRIS. Tanto o KNN quanto o SVM alcançaram uma acurácia de 100%, enquanto a árvore de decisão obteve uma acurácia de 97%. Esses resultados indicam que esses modelos são capazes de identificar corretamente as diferentes espécies de flores com base nos atributos fornecidos. A alta acurácia obtida demonstra a capacidade dos modelos em aprender e generalizar a partir dos dados de treinamento. Portanto, esses modelos podem ser considerados ferramentas confiáveis para a classificação de espécies de flores no conjunto de dados IRIS, com potenciais aplicações em diversos campos, como a botânica e a ecologia.



---

## 6 Referências Bibliográficas

Hastie, T., Tibshirani, R., Friedman, J. (2009). The elements of statistical learning: data mining, inference, and prediction. Springer Science Business Media.

Breiman, L., Friedman, J., Olshen, R., Stone, C. (1984). Classification and regression trees. Wadsworth Brooks.

Cover, T., Hart, P. (1967). Nearest neighbor pattern classification. IEEE Transactions on Information Theory, 13(1), 21-27.

Cortes, C., Vapnik, V. (1995). Support-vector networks. Machine learning, 20(3), 273-297.