

Relatório do projeto de Machine Learning para classificação de flores de Íris

Yasser Schuck Antonio Tuma

Dataset:

O dataset apresenta 150 dados referentes a flores de Íris, cada dado é composto pelos valores de comprimento e largura da sépala, comprimento e largura da pétala, e a espécie a qual pertence a flor de Íris.

Dataset obtido em: <http://archive.ics.uci.edu/ml/datasets/Iris>

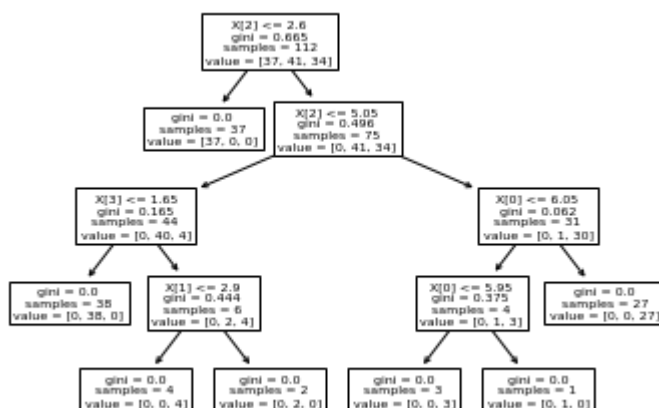
Pré-processamento:

O dataset não apresentava o nome das colunas que representam os atributos da flor, então foram adicionados os nomes “SepalaComprimento”, “SepalaLargura”, “PetalaComprimento”, “PetalaLargura” e “Especie” em suas respectivas colunas.

Em seguida, o dataset foi dividido em amostras para treino (75%, 112) e teste (25%, 38) do modelo de Machine Learning, depois, cada amostra foi dividido em uma amostra com os atributos preditivos (“SepalaComprimento”, “SepalaLargura”, “PetalaComprimento”, “PetalaLargura”) e outra com o atributo de classe (“Especie”).

Treinamento do modelo:

Foi treinado um modelo de machine learning baseado em árvore de decisão, utilizando a amostra de treino.



Métricas de desempenho:

Predições com a amostra de teste:

Acurácia	Recall	F-Score
0.92	0.92	0.92

Matriz de confusão:

```
[[13  0  0]
 [ 0  9  3]
 [ 0  0 13]]
```

Predições com todo o dataset:

Acurácia	Recall	F-Score
0.98	0.98	0.98

Matriz de confusão:

```
[[50  0  0]
 [ 0 50  3]
 [ 0  0 47]]
```