

Untitled6.ipynb

Untitled6.ipynb\_

Arquivo

Editar

Ver

Inserir

Ambiente de execução

Ferramentas

Ajuda

Todas as alterações foram salvas

ComentárioCompartilhar

[](https://accounts.google.com/SignOutOptions?hl=pt-BR&continue=https://colab.research.google.com/)

Arquivos

..

sample\_data

Arvore\_decisao\_dataset.png

Arvore\_decisao\_dataset\_poda.png

db\_diabetes.csv

Solte arquivos aqui para fazer upload no armazenamento da sessão

Disco

Disponível: 70.74 GB

Código

Texto

RAM

Disco

Editar

[6]

0s

import pandas as pd



[7]

0s

dataset = pd.read\_csv('/content/db\_diabetes.csv', sep=',')



[ ]

dataset.info()



<class 'pandas.core.frame.DataFrame'>

RangeIndex: 768 entries, 0 to 767

Data columns (total 9 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 Pregnancies 768 non-null int64

1 Glucose 768 non-null int64

2 BloodPressure 768 non-null int64

3 SkinThickness 768 non-null int64

4 Insulin 768 non-null int64

5 BMI 768 non-null float64

6 DiabetesPedigreeFunction 768 non-null float64

7 Age 768 non-null int64

8 Outcome 768 non-null int64

dtypes: float64(2), int64(7)

memory usage: 54.1 KB

[ ]

dataset.describe()



[ ]

!pip install dataprep



Collecting dataprep

Downloading dataprep-0.3.0-py3-none-any.whl (1.8 MB)

|████████████████████████████████| 1.8 MB 29.2 MB/s

Requirement already satisfied: tqdm<5.0,>=4.48 in /usr/local/lib/python3.7/dist-packages (from dataprep) (4.62.0)

Requirement already satisfied: bottleneck<2.0,>=1.3 in /usr/local/lib/python3.7/dist-packages (from dataprep) (1.3.2)

Requirement already satisfied: bokeh<3,>=2 in /usr/local/lib/python3.7/dist-packages (from dataprep) (2.3.3)

Requirement already satisfied: ipywidgets<8.0,>=7.5 in /usr/local/lib/python3.7/dist-packages (from dataprep) (7.6.3)

Requirement already satisfied: numpy<2,>=1 in /usr/local/lib/python3.7/dist-packages (from dataprep) (1.19.5)

Collecting nltk<4.0,>=3.5

Downloading nltk-3.6.2-py3-none-any.whl (1.5 MB)

|████████████████████████████████| 1.5 MB 39.4 MB/s

Requirement already satisfied: scipy<2,>=1 in /usr/local/lib/python3.7/dist-packages (from dataprep) (1.4.1)

Collecting jsonpath-ng<2.0,>=1.5

Downloading jsonpath\_ng-1.5.3-py3-none-any.whl (29 kB)

Collecting regex<2021.0.0,>=2020.10.15

Downloading regex-2020.11.13-cp37-cp37m-manylinux2014\_x86\_64.whl (719 kB)

|████████████████████████████████| 719 kB 61.5 MB/s

Collecting metaphone<0.7,>=0.6

Downloading Metaphone-0.6.tar.gz (14 kB)

Collecting pydantic<2.0,>=1.6

Downloading pydantic-1.8.2-cp37-cp37m-manylinux2014\_x86\_64.whl (10.1 MB)

|████████████████████████████████| 10.1 MB 51.6 MB/s

Collecting usaddress<0.6.0,>=0.5.10

Downloading usaddress-0.5.10-py2.py3-none-any.whl (63 kB)

|████████████████████████████████| 63 kB 3.1 MB/s

Collecting aiohttp<4.0,>=3.6

Downloading aiohttp-3.7.4.post0-cp37-cp37m-manylinux2014\_x86\_64.whl (1.3 MB)

|████████████████████████████████| 1.3 MB 45.7 MB/s

Collecting dask[array,dataframe,delayed]<3.0,>=2.25

Downloading dask-2.30.0-py3-none-any.whl (848 kB)

|████████████████████████████████| 848 kB 65.5 MB/s

Requirement already satisfied: pandas<2.0,>=1.1 in /usr/local/lib/python3.7/dist-packages (from dataprep) (1.1.5)

Requirement already satisfied: jinja2<3.0,>=2.11 in /usr/local/lib/python3.7/dist-packages (from dataprep) (2.11.3)

Collecting python-Levenshtein<0.13.0,>=0.12.2

Downloading python-Levenshtein-0.12.2.tar.gz (50 kB)

|████████████████████████████████| 50 kB 7.4 MB/s

Collecting wordcloud<2.0,>=1.8

Downloading wordcloud-1.8.1-cp37-cp37m-manylinux1\_x86\_64.whl (366 kB)

|████████████████████████████████| 366 kB 56.2 MB/s

Collecting multidict<7.0,>=4.5

Downloading multidict-5.1.0-cp37-cp37m-manylinux2014\_x86\_64.whl (142 kB)

|████████████████████████████████| 142 kB 74.5 MB/s

Requirement already satisfied: typing-extensions>=3.6.5 in /usr/local/lib/python3.7/dist-packages (from aiohttp<4.0,>=3.6->dataprep) (3.7.4.3)

Collecting async-timeout<4.0,>=3.0

Downloading async\_timeout-3.0.1-py3-none-any.whl (8.2 kB)

Requirement already satisfied: attrs>=17.3.0 in /usr/local/lib/python3.7/dist-packages (from aiohttp<4.0,>=3.6->dataprep) (21.2.0)

Requirement already satisfied: chardet<5.0,>=2.0 in /usr/local/lib/python3.7/dist-packages (from aiohttp<4.0,>=3.6->dataprep) (3.0.4)

Collecting yarl<2.0,>=1.0

Downloading yarl-1.6.3-cp37-cp37m-manylinux2014\_x86\_64.whl (294 kB)

|████████████████████████████████| 294 kB 77.8 MB/s

Requirement already satisfied: pillow>=7.1.0 in /usr/local/lib/python3.7/dist-packages (from bokeh<3,>=2->dataprep) (7.1.2)

Requirement already satisfied: python-dateutil>=2.1 in /usr/local/lib/python3.7/dist-packages (from bokeh<3,>=2->dataprep) (2.8.2)

Requirement already satisfied: packaging>=16.8 in /usr/local/lib/python3.7/dist-packages (from bokeh<3,>=2->dataprep) (21.0)

Requirement already satisfied: PyYAML>=3.10 in /usr/local/lib/python3.7/dist-packages (from bokeh<3,>=2->dataprep) (3.13)

Requirement already satisfied: tornado>=5.1 in /usr/local/lib/python3.7/dist-packages (from bokeh<3,>=2->dataprep) (5.1.1)

Requirement already satisfied: cloudpickle>=0.2.2 in /usr/local/lib/python3.7/dist-packages (from dask[array,dataframe,delayed]<3.0,>=2.25->dataprep) (1.3.0)

Requirement already satisfied: toolz>=0.8.2 in /usr/local/lib/python3.7/dist-packages (from dask[array,dataframe,delayed]<3.0,>=2.25->dataprep) (0.11.1)

Collecting fsspec>=0.6.0

Downloading fsspec-2021.7.0-py3-none-any.whl (118 kB)

|████████████████████████████████| 118 kB 83.4 MB/s

Collecting partd>=0.3.10

Downloading partd-1.2.0-py3-none-any.whl (19 kB)

Requirement already satisfied: ipython>=4.0.0 in /usr/local/lib/python3.7/dist-packages (from ipywidgets<8.0,>=7.5->dataprep) (5.5.0)

Requirement already satisfied: traitlets>=4.3.1 in /usr/local/lib/python3.7/dist-packages (from ipywidgets<8.0,>=7.5->dataprep) (5.0.5)

Requirement already satisfied: ipykernel>=4.5.1 in /usr/local/lib/python3.7/dist-packages (from ipywidgets<8.0,>=7.5->dataprep) (4.10.1)

Requirement already satisfied: nbformat>=4.2.0 in /usr/local/lib/python3.7/dist-packages (from ipywidgets<8.0,>=7.5->dataprep) (5.1.3)

Requirement already satisfied: jupyterlab-widgets>=1.0.0 in /usr/local/lib/python3.7/dist-packages (from ipywidgets<8.0,>=7.5->dataprep) (1.0.0)

Requirement already satisfied: widgetsnbextension~=3.5.0 in /usr/local/lib/python3.7/dist-packages (from ipywidgets<8.0,>=7.5->dataprep) (3.5.1)

Requirement already satisfied: jupyter-client in /usr/local/lib/python3.7/dist-packages (from ipykernel>=4.5.1->ipywidgets<8.0,>=7.5->dataprep) (5.3.5)

Requirement already satisfied: setuptools>=18.5 in /usr/local/lib/python3.7/dist-packages (from ipython>=4.0.0->ipywidgets<8.0,>=7.5->dataprep) (57.4.0)

Requirement already satisfied: simplegeneric>0.8 in /usr/local/lib/python3.7/dist-packages (from ipython>=4.0.0->ipywidgets<8.0,>=7.5->dataprep) (0.8.1)

Requirement already satisfied: pygments in /usr/local/lib/python3.7/dist-packages (from ipython>=4.0.0->ipywidgets<8.0,>=7.5->dataprep) (2.6.1)

Requirement already satisfied: prompt-toolkit<2.0.0,>=1.0.4 in /usr/local/lib/python3.7/dist-packages (from ipython>=4.0.0->ipywidgets<8.0,>=7.5->dataprep) (1.0.18)

Requirement already satisfied: pexpect in /usr/local/lib/python3.7/dist-packages (from ipython>=4.0.0->ipywidgets<8.0,>=7.5->dataprep) (4.8.0)

Requirement already satisfied: pickleshare in /usr/local/lib/python3.7/dist-packages (from ipython>=4.0.0->ipywidgets<8.0,>=7.5->dataprep) (0.7.5)

Requirement already satisfied: decorator in /usr/local/lib/python3.7/dist-packages (from ipython>=4.0.0->ipywidgets<8.0,>=7.5->dataprep) (4.4.2)

Requirement already satisfied: MarkupSafe>=0.23 in /usr/local/lib/python3.7/dist-packages (from jinja2<3.0,>=2.11->dataprep) (2.0.1)

Collecting ply

Downloading ply-3.11-py2.py3-none-any.whl (49 kB)

|████████████████████████████████| 49 kB 7.9 MB/s

Requirement already satisfied: six in /usr/local/lib/python3.7/dist-packages (from jsonpath-ng<2.0,>=1.5->dataprep) (1.15.0)

Requirement already satisfied: jsonschema!=2.5.0,>=2.4 in /usr/local/lib/python3.7/dist-packages (from nbformat>=4.2.0->ipywidgets<8.0,>=7.5->dataprep) (2.6.0)

Requirement already satisfied: ipython-genutils in /usr/local/lib/python3.7/dist-packages (from nbformat>=4.2.0->ipywidgets<8.0,>=7.5->dataprep) (0.2.0)

Requirement already satisfied: jupyter-core in /usr/local/lib/python3.7/dist-packages (from nbformat>=4.2.0->ipywidgets<8.0,>=7.5->dataprep) (4.7.1)

Requirement already satisfied: click in /usr/local/lib/python3.7/dist-packages (from nltk<4.0,>=3.5->dataprep) (7.1.2)

Requirement already satisfied: joblib in /usr/local/lib/python3.7/dist-packages (from nltk<4.0,>=3.5->dataprep) (1.0.1)

Requirement already satisfied: pyparsing>=2.0.2 in /usr/local/lib/python3.7/dist-packages (from packaging>=16.8->bokeh<3,>=2->dataprep) (2.4.7)

Requirement already satisfied: pytz>=2017.2 in /usr/local/lib/python3.7/dist-packages (from pandas<2.0,>=1.1->dataprep) (2018.9)

Collecting locket

Downloading locket-0.2.1-py2.py3-none-any.whl (4.1 kB)

Requirement already satisfied: wcwidth in /usr/local/lib/python3.7/dist-packages (from prompt-toolkit<2.0.0,>=1.0.4->ipython>=4.0.0->ipywidgets<8.0,>=7.5->dataprep) (0.2.5)

Requirement already satisfied: future>=0.14 in /usr/local/lib/python3.7/dist-packages (from usaddress<0.6.0,>=0.5.10->dataprep) (0.16.0)

Collecting python-crfsuite>=0.7

Downloading python\_crfsuite-0.9.7-cp37-cp37m-manylinux1\_x86\_64.whl (743 kB)

|████████████████████████████████| 743 kB 35.7 MB/s

Collecting probableparsing

Downloading probableparsing-0.0.1-py2.py3-none-any.whl (3.1 kB)

Requirement already satisfied: notebook>=4.4.1 in /usr/local/lib/python3.7/dist-packages (from widgetsnbextension~=3.5.0->ipywidgets<8.0,>=7.5->dataprep) (5.3.1)

Requirement already satisfied: terminado>=0.8.1 in /usr/local/lib/python3.7/dist-packages (from notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets<8.0,>=7.5->dataprep) (0.11.0)

Requirement already satisfied: nbconvert in /usr/local/lib/python3.7/dist-packages (from notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets<8.0,>=7.5->dataprep) (5.6.1)

Requirement already satisfied: Send2Trash in /usr/local/lib/python3.7/dist-packages (from notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets<8.0,>=7.5->dataprep) (1.8.0)

Requirement already satisfied: pyzmq>=13 in /usr/local/lib/python3.7/dist-packages (from jupyter-client->ipykernel>=4.5.1->ipywidgets<8.0,>=7.5->dataprep) (22.2.1)

Requirement already satisfied: ptyprocess in /usr/local/lib/python3.7/dist-packages (from terminado>=0.8.1->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets<8.0,>=7.5->dataprep) (0.7.0)

Requirement already satisfied: matplotlib in /usr/local/lib/python3.7/dist-packages (from wordcloud<2.0,>=1.8->dataprep) (3.2.2)

Requirement already satisfied: idna>=2.0 in /usr/local/lib/python3.7/dist-packages (from yarl<2.0,>=1.0->aiohttp<4.0,>=3.6->dataprep) (2.10)

Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.7/dist-packages (from matplotlib->wordcloud<2.0,>=1.8->dataprep) (0.10.0)

Requirement already satisfied: kiwisolver>=1.0.1 in /usr/local/lib/python3.7/dist-packages (from matplotlib->wordcloud<2.0,>=1.8->dataprep) (1.3.1)

Requirement already satisfied: testpath in /usr/local/lib/python3.7/dist-packages (from nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets<8.0,>=7.5->dataprep) (0.5.0)

Requirement already satisfied: mistune<2,>=0.8.1 in /usr/local/lib/python3.7/dist-packages (from nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets<8.0,>=7.5->dataprep) (0.8.4)

Requirement already satisfied: entrypoints>=0.2.2 in /usr/local/lib/python3.7/dist-packages (from nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets<8.0,>=7.5->dataprep) (0.3)

Requirement already satisfied: defusedxml in /usr/local/lib/python3.7/dist-packages (from nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets<8.0,>=7.5->dataprep) (0.7.1)

Requirement already satisfied: bleach in /usr/local/lib/python3.7/dist-packages (from nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets<8.0,>=7.5->dataprep) (4.0.0)

Requirement already satisfied: pandocfilters>=1.4.1 in /usr/local/lib/python3.7/dist-packages (from nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets<8.0,>=7.5->dataprep) (1.4.3)

Requirement already satisfied: webencodings in /usr/local/lib/python3.7/dist-packages (from bleach->nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets<8.0,>=7.5->dataprep) (0.5.1)

Building wheels for collected packages: metaphone, python-Levenshtein

Building wheel for metaphone (setup.py) ... done

Created wheel for metaphone: filename=Metaphone-0.6-py3-none-any.whl size=13919 sha256=5c5fb65e244fa635e7bbb444dcc24f3665c3087916e489624890c4e847d84f21

Stored in directory: /root/.cache/pip/wheels/1d/a8/cb/6f8902aa5457bd71344e00665c230e9c45255b3f57f2194a0f

Building wheel for python-Levenshtein (setup.py) ... done

Created wheel for python-Levenshtein: filename=python\_Levenshtein-0.12.2-cp37-cp37m-linux\_x86\_64.whl size=149873 sha256=f03a242144e12d70403e6fd93650d5602d12cf87d83b204361db661d9a605d4e

Stored in directory: /root/.cache/pip/wheels/05/5f/ca/7c4367734892581bb5ff896f15027a932c551080b2abd3e00d

Successfully built metaphone python-Levenshtein

Installing collected packages: multidict, locket, yarl, regex, python-crfsuite, probableparsing, ply, partd, fsspec, dask, async-timeout, wordcloud, usaddress, python-Levenshtein, pydantic, nltk, metaphone, jsonpath-ng, aiohttp, dataprep

Attempting uninstall: regex

Found existing installation: regex 2019.12.20

Uninstalling regex-2019.12.20:

Successfully uninstalled regex-2019.12.20

Attempting uninstall: dask

Found existing installation: dask 2.12.0

Uninstalling dask-2.12.0:

Successfully uninstalled dask-2.12.0

Attempting uninstall: wordcloud

Found existing installation: wordcloud 1.5.0

Uninstalling wordcloud-1.5.0:

Successfully uninstalled wordcloud-1.5.0

Attempting uninstall: nltk

Found existing installation: nltk 3.2.5

Uninstalling nltk-3.2.5:

Successfully uninstalled nltk-3.2.5

Successfully installed aiohttp-3.7.4.post0 async-timeout-3.0.1 dask-2.30.0 dataprep-0.3.0 fsspec-2021.7.0 jsonpath-ng-1.5.3 locket-0.2.1 metaphone-0.6 multidict-5.1.0 nltk-3.6.2 partd-1.2.0 ply-3.11 probableparsing-0.0.1 pydantic-1.8.2 python-Levenshtein-0.12.2 python-crfsuite-0.9.7 regex-2020.11.13 usaddress-0.5.10 wordcloud-1.8.1 yarl-1.6.3

[ ]

from dataprep.eda import create\_report

create\_report(dataset)



[ ]

dataset.groupby(['Outcome']).describe()



[26]

0s

from sklearn.model\_selection import StratifiedShuffleSplit



[36]

0s

X = dataset.drop('Outcome', axis=1).values

y = dataset['Outcome'].values



[37]

0s

validador = StratifiedShuffleSplit(n\_splits=1, test\_size=0.1, random\_state=0)

for treino\_id, test\_id in validador.split(X,y):

  X\_train, X\_test = X[treino\_id], X[test\_id]

  y\_train, y\_test = y[treino\_id], y[test\_id]



[38]

0s

from sklearn import tree

classifica\_arvore\_decisao = tree.DecisionTreeClassifier()

arvore\_dataset = classifica\_arvore\_decisao.fit(X\_train, y\_train)

y\_pred = arvore\_dataset.predict(X\_test)



[39]

47s

import matplotlib.pyplot as plt

plt.figure(figsize=(300,200))

tree.plot\_tree(classifica\_arvore\_decisao, filled=True, fontsize=16)

plt.savefig('Arvore\_decisao\_dataset.png')

plt.close()



[40]

0s

from sklearn.metrics import accuracy\_score



[41]

0s

def metricas\_avaliacao(y\_test,y\_pred):

  print('O valor da acurácia é:', accuracy\_score(y\_test,y\_pred))



[42]

0s

metricas\_avaliacao(y\_test,y\_pred)



O valor da acurácia é: 0.7402597402597403

[43]

0s

from sklearn import tree

classifica\_arvore\_decisao = tree.DecisionTreeClassifier(max\_depth=3,random\_state=0)

arvore\_dataset = classifica\_arvore\_decisao.fit(X\_train, y\_train)

y\_pred\_3 = arvore\_dataset.predict(X\_test)



[46]

43s

import matplotlib.pyplot as plt

plt.figure(figsize=(300,200))

tree.plot\_tree(classifica\_arvore\_decisao, filled=True, fontsize=16)

plt.savefig('Arvore\_decisao\_dataset\_poda.png')

plt.close()



[45]

0s

metricas\_avaliacao(y\_test,y\_pred\_3)



O valor da acurácia é: 0.8311688311688312