

Relatório sobre a atividade “4 - Prática: Principais Bibliotecas e Ferramentas Python para Aprendizado de Máquina (I)”

Lucas Gabriel Arenhardt

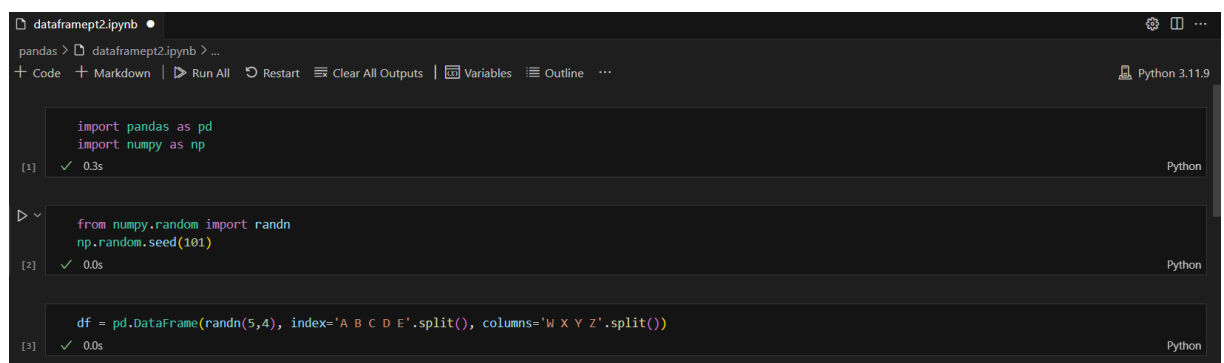
1. Introdução

Nessa atividade, foram demonstradas algumas ferramentas e bibliotecas essenciais no python para lidar com aprendizado de máquina e análise de dados. Primeiramente, foi exibido o Jupyter Notebook, o qual foi utilizado durante todas as aulas para o desenvolvimento dos códigos. Em seguida, foram apresentadas as bibliotecas Numpy e Pandas, respectivamente. Ao final, foram oferecidos exercícios para melhor aprendizado.

2. Conteúdo abordado

2.1. Jupyter Notebook

O Jupyter Notebook é uma ferramenta que permite criar documentos que podem conter, ao mesmo tempo, código ao vivo, texto, equações e visualizações. É amplamente utilizado em campos como análise de dados e aprendizado de máquina. Nesta ferramenta, o código pode ser dividido em células individuais, tornando mais fácil testar e modificar partes específicas do programa. No meu computador, optei por utilizar o Jupyter Notebook juntamente com o Visual Studio Code, pois já é um costume antigo.



```
dataframept2.ipynb
pandas > dataframept2.ipynb > ...
+ Code + Markdown | ▶ Run All | ⌂ Restart | 🗑 Clear All Outputs | 📄 Variables | 📖 Outline | ...
Python 3.11.9

[1] ✓ 0.3s Python
import pandas as pd
import numpy as np

[2] ✓ 0.0s Python
from numpy.random import randn
np.random.seed(101)

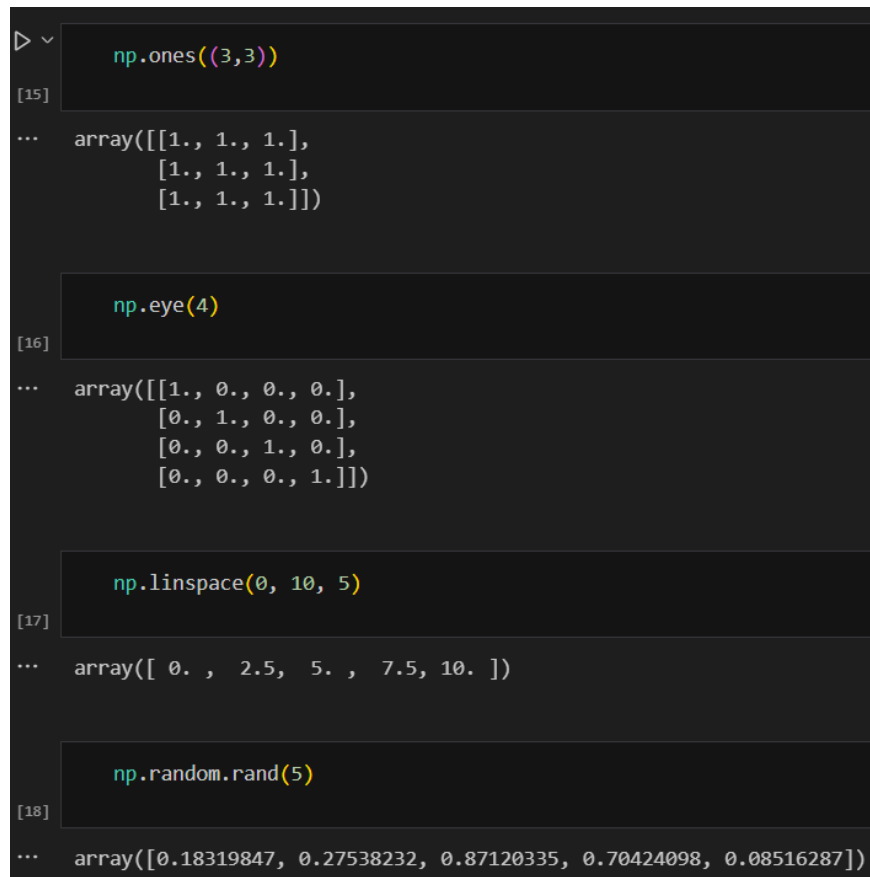
[3] ✓ 0.0s Python
df = pd.DataFrame(randn(5,4), index='A B C D E'.split(), columns='W X Y Z'.split())
```

Imagem 1 - Exemplo de código no Visual Studio Code utilizando o Jupyter Notebook.

2.2. Biblioteca Numpy

O Numpy é, essencialmente, uma biblioteca para operações matemáticas. Ele oferece funções e estruturas de dados eficientes para lidar com arrays multidimensionais e operações matemáticas complexas.

Durante as aulas, foi ensinado como criar Numpy Arrays e realizar operações do tipo indexação, fatiamento, reshape, random, além de outras funções básicas da biblioteca.



```
np.ones((3,3))  
[15]  
... array([[1., 1., 1.],  
          [1., 1., 1.],  
          [1., 1., 1.]])  
  
np.eye(4)  
[16]  
... array([[1., 0., 0., 0.],  
          [0., 1., 0., 0.],  
          [0., 0., 1., 0.],  
          [0., 0., 0., 1.]])  
  
np.linspace(0, 10, 5)  
[17]  
... array([ 0. ,  2.5,  5. ,  7.5, 10. ])  
  
np.random.rand(5)  
[18]  
... array([0.18319847, 0.27538232, 0.87120335, 0.70424098, 0.08516287])
```

Imagem 2 - Exemplos de funções da biblioteca Numpy.

2.3. Biblioteca Pandas

Pandas é uma biblioteca em Python que oferece estruturas para análise e manipulação de conjunto de dados. Suas principais estruturas são as séries e os DataFrames. A série é unidimensional e pode conter mais de um tipo de dado, enquanto o DataFrame é semelhante a uma planilha, organizada em linhas e colunas.

Durante as aulas, foi ensinado como criar séries e realizar operações com as mesmas, como atribuir labels à valores. Seguido disso, foi abordado o conteúdo sobre os DataFrames.

Dentre as operações com os DataFrames, temos: criação e fatiamento de DataFrames, tratamento de dados NaN, junções (bem semelhante às operações em SQL), groupby (semelhante à planilhas), busca no DataFrame, coleta de informações no DataFrame e, por fim, transformação de outros formatos de arquivo (CSV, HTML, XLSX) para um DataFrame Pandas.

3. Conclusão

Após as aulas, foi possível relembrar e adquirir alguns conhecimentos novos acerca de tais bibliotecas e ferramentas tão importantes para a área de aprendizado de máquina e inteligência artificial. Eu já possuía um certo grau de conhecimento sobre tais assuntos, visto que já havia trabalhado com essas ferramentas antes em demais projetos, porém, ainda assim, certas coisas apresentadas nessas aulas eu ainda não conhecia.