Atividade_02

October 3, 2023

Classificação - Mineração de Dados

Nome: Davi Augusto Neves Leite

Data de Entrega: 10/10/2023

---

1   Materiais

Os principais recursos para a execução desta atividade podem ser vistos a seguir.

1. **Software**

- Sistemas Operacionais: Windows 11 para *desktop*;
- Ambiente de Desenvolvimento Integrado: Microsoft Visual Studio Code;
- Linguagem de Programação: Python 3.11.5 64-bit.

2. **Hardware**

- Notebook pessoal Lenovo Ideapad 330-15IKB com: processador Intel Core i7-8550U, HDD WD Blue WD10SPZX de 1TB, SSD Crucial BX500 de 1TB, 12 GB DDR4 de Memória RAM e placa de vídeo NVIDIA GeForce MX150 (2 GB GDDR5 de memória).

---

2   Instalação das Bibliotecas Principais

Nota: ao decorrer deste Notebook, outras bibliotecas podem ser utilizadas em quaisquer respectiva seção/conjunto de dados, dependendo da necessidade. Abaixo, há a instalação das principais que são comuns e utilizadas em todas ou quase todas seções/conjunto de dados.

```
[ ]: %pip install numpy
     %pip install pandas
     %pip install seaborn
     %pip install matplotlib
     %pip install scikit-learn
```

```
Requirement already satisfied: numpy in c:\python311\lib\site-packages (1.23.5)
Note: you may need to restart the kernel to use updated packages.
```

```
DEPRECATION: Loading egg at c:\python311\lib\site-
packages\vboxapi-1.0-py3.11.egg is deprecated. pip 23.3 will enforce this
behaviour change. A possible replacement is to use pip for package
installation..

Requirement already satisfied: pandas in c:\python311\lib\site-packages (2.1.0)
Requirement already satisfied: numpy>=1.23.2 in c:\python311\lib\site-packages
(from pandas) (1.23.5)
Requirement already satisfied: python-dateutil>=2.8.2 in c:\python311\lib\site-
packages (from pandas) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in c:\python311\lib\site-packages
(from pandas) (2022.7.1)
Requirement already satisfied: tzdata>=2022.1 in c:\python311\lib\site-packages
(from pandas) (2023.3)
Requirement already satisfied: six>=1.5 in c:\python311\lib\site-packages (from
python-dateutil>=2.8.2->pandas) (1.16.0)
Note: you may need to restart the kernel to use updated packages.

DEPRECATION: Loading egg at c:\python311\lib\site-
packages\vboxapi-1.0-py3.11.egg is deprecated. pip 23.3 will enforce this
behaviour change. A possible replacement is to use pip for package
installation..

Requirement already satisfied: seaborn in c:\python311\lib\site-packages
(0.12.2)
Requirement already satisfied: numpy!=1.24.0,>=1.17 in c:\python311\lib\site-
packages (from seaborn) (1.23.5)
Requirement already satisfied: pandas>=0.25 in c:\python311\lib\site-packages
(from seaborn) (2.1.0)
Requirement already satisfied: matplotlib!=3.6.1,>=3.1 in c:\python311\lib\site-
packages (from seaborn) (3.7.1)
Requirement already satisfied: contourpy>=1.0.1 in c:\python311\lib\site-
packages (from matplotlib!=3.6.1,>=3.1->seaborn) (1.0.7)
Requirement already satisfied: cycler>=0.10 in c:\python311\lib\site-packages
(from matplotlib!=3.6.1,>=3.1->seaborn) (0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in c:\python311\lib\site-
packages (from matplotlib!=3.6.1,>=3.1->seaborn) (4.39.0)
Requirement already satisfied: kiwisolver>=1.0.1 in c:\python311\lib\site-
packages (from matplotlib!=3.6.1,>=3.1->seaborn) (1.4.4)
Requirement already satisfied: packaging>=20.0 in c:\python311\lib\site-packages
(from matplotlib!=3.6.1,>=3.1->seaborn) (23.0)
Requirement already satisfied: pillow>=6.2.0 in c:\python311\lib\site-packages
(from matplotlib!=3.6.1,>=3.1->seaborn) (9.5.0)
Requirement already satisfied: pyparsing>=2.3.1 in c:\python311\lib\site-
packages (from matplotlib!=3.6.1,>=3.1->seaborn) (3.0.9)
Requirement already satisfied: python-dateutil>=2.7 in c:\python311\lib\site-
packages (from matplotlib!=3.6.1,>=3.1->seaborn) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in c:\python311\lib\site-packages
(from pandas>=0.25->seaborn) (2022.7.1)
```

```
Requirement already satisfied: tzdata>=2022.1 in c:\python311\lib\site-packages
(from pandas>=0.25->seaborn) (2023.3)
Requirement already satisfied: six>=1.5 in c:\python311\lib\site-packages (from
python-dateutil>=2.7->matplotlib!=3.6.1,>=3.1->seaborn) (1.16.0)
Note: you may need to restart the kernel to use updated packages.

DEPRECATION: Loading egg at c:\python311\lib\site-
packages\vboxapi-1.0-py3.11.egg is deprecated. pip 23.3 will enforce this
behaviour change. A possible replacement is to use pip for package
installation..

Requirement already satisfied: matplotlib in c:\python311\lib\site-packages
(3.7.1)
Requirement already satisfied: contourpy>=1.0.1 in c:\python311\lib\site-
packages (from matplotlib) (1.0.7)
Requirement already satisfied: cycler>=0.10 in c:\python311\lib\site-packages
(from matplotlib) (0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in c:\python311\lib\site-
packages (from matplotlib) (4.39.0)
Requirement already satisfied: kiwisolver>=1.0.1 in c:\python311\lib\site-
packages (from matplotlib) (1.4.4)
Requirement already satisfied: numpy>=1.20 in c:\python311\lib\site-packages
(from matplotlib) (1.23.5)
Requirement already satisfied: packaging>=20.0 in c:\python311\lib\site-packages
(from matplotlib) (23.0)
Requirement already satisfied: pillow>=6.2.0 in c:\python311\lib\site-packages
(from matplotlib) (9.5.0)
Requirement already satisfied: pyparsing>=2.3.1 in c:\python311\lib\site-
packages (from matplotlib) (3.0.9)
Requirement already satisfied: python-dateutil>=2.7 in c:\python311\lib\site-
packages (from matplotlib) (2.8.2)
Requirement already satisfied: six>=1.5 in c:\python311\lib\site-packages (from
python-dateutil>=2.7->matplotlib) (1.16.0)
Note: you may need to restart the kernel to use updated packages.

DEPRECATION: Loading egg at c:\python311\lib\site-
packages\vboxapi-1.0-py3.11.egg is deprecated. pip 23.3 will enforce this
behaviour change. A possible replacement is to use pip for package
installation..

Requirement already satisfied: scikit-learn in c:\python311\lib\site-packages
(1.2.2)
Requirement already satisfied: numpy>=1.17.3 in c:\python311\lib\site-packages
(from scikit-learn) (1.23.5)
Requirement already satisfied: scipy>=1.3.2 in c:\python311\lib\site-packages
(from scikit-learn) (1.10.1)
Requirement already satisfied: joblib>=1.1.1 in c:\python311\lib\site-packages
(from scikit-learn) (1.2.0)
Requirement already satisfied: threadpoolctl>=2.0.0 in c:\python311\lib\site-
packages (from scikit-learn) (3.1.0)
```

```
Note: you may need to restart the kernel to use updated packages.

DEPRECATION: Loading egg at c:\python311\lib\site-
packages\vboxapi-1.0-py3.11.egg is deprecated. pip 23.3 will enforce this
behaviour change. A possible replacement is to use pip for package
installation..
```

---

3 Importação das Bibliotecas Principais

Nota: ao decorrer deste Notebook, outras bibliotecas podem ser utilizadas em quaisquer respectiva seção/conjunto de dados, dependendo da necessidade. Abaixo, há a importação das principais que são comuns e utilizadas em todas ou quase todas seções/conjunto de dados.

```python
[ ]: import numpy as np        # Manipulação de listas
     import pandas as pd        # Manipulação de tabelas
     import seaborn as sbn      # Geração de gráficos estatísticos
     import matplotlib.pyplot as plt  # Geração de gráficos de listas
     import sklearn as skl      # Biblioteca para pré-processamento
     from copy import copy as cp  # Possibilitar copiar os objetos

     # Ignorar os avisos não importantes durante a execução deste notebook
     import warnings
     warnings.filterwarnings('ignore')
```

---

4 Conjunto de Dados: *Fashion MNIST*

Descrição do Dataset: este conjunto é composto por **70 mil imagens** a respeito de **10 peças de roupas distintas**. Cada imagem possui a **resolução de 28x28 (784 pixels) em escala de cinza e 256 níveis de cinza**. Desse total, exitem **7000 imagens** para cada uma das peças de roupas, ou classes.

Cada classe pode ser vista a seguir.

1. Camiseta
2. Calça
3. Pulôver
4. Vestido
5. Casaco
6. Sandália
7. Camisa
8. Tênis
9. Bolsa
10. Bota de Tornozelo

Este conjunto de dados pode ser acessado por meio de: Fashion MNIST (última data de acesso: 02 de out. de 2023).

Para fins práticos, apenas as classes *camiseta*, *vestido*, *casaco* e *camisa* serão utilizadas para esta atividade, tendo em vista que, por se tratar de uma base de dados de imagens, o processamento necessário para classificação é elevado.

4.1 Importação da Base de Dados

Para importar as imagens, utiliza-se o módulo **scikit-learn** capaz de carregar diversas bases de dados através do portal **OpenML**, o qual essa base de dados está disponível online.

```
[ ]: # Importação da função necessária para importação de base de dados OpenML
from sklearn.datasets import fetch_openml

# Carrega as imagens e suas classes separadamente
images, targets = fetch_openml(
    'Fashion-MNIST', return_X_y=True, as_frame=True, parser='auto')

# Conversão das imagens para NumPy
images = images.to_numpy()
targets = targets.to_numpy().astype(int)
```

4.2 Pré-Processamento dos Dados