



INSTITUTO POLITÉCNICO NACIONAL

ESCOM

INTELIGENCIA ARTIFICIAL

LAB 8: DATASET

MARTÍNEZ CHÁVEZ JORGE ALEXIS

6CV3

31 OCTUBRE 2024

Introducción

Desarrollo

Primero para la versión de Python, debemos de instalar una librería llamada pandas la cual no ayudara a manipular los datos, para ello simplemente debemos de abrir un terminal e insertar el siguiente comando.

```
C:\Users\jorge>pip install pandas
Collecting pandas
  Downloading pandas-2.2.3-cp310-cp310-win_amd64.whl.metadata (19 kB)
Requirement already satisfied: numpy>=1.22.4 in c:\users\jorge\appdata\local\programs\python\python310\lib\site-packages (from pandas) (1.26.0)
Requirement already satisfied: python-dateutil>=2.8.2 in c:\users\jorge\appdata\local\programs\python\python310\lib\site-packages (from pandas) (2.8.2)
Collecting pytz>=2020.1 (from pandas)
  Downloading pytz-2024.2-py2.py3-none-any.whl.metadata (22 kB)
Collecting tzdata>=2022.7 (from pandas)
  Downloading tzdata-2024.2-py2.py3-none-any.whl.metadata (1.4 kB)
Requirement already satisfied: six>=1.5 in c:\users\jorge\appdata\local\programs\python\python310\lib\site-packages (from python-dateutil>=2.8.2->pandas) (1.16.0)
Downloading pandas-2.2.3-cp310-cp310-win_amd64.whl (11.6 MB)
  11.6/11.6 MB 16.0 MB/s eta 0:00:00
Downloading pytz-2024.2-py2.py3-none-any.whl (508 kB)
  508.0/508.0 kB 31.1 MB/s eta 0:00:00
Downloading tzdata-2024.2-py2.py3-none-any.whl (346 kB)
  346.6/346.6 kB 21.0 MB/s eta 0:00:00
Installing collected packages: pytz, tzdata, pandas
Successfully installed pandas-2.2.3 pytz-2024.2 tzdata-2024.2

[notice] A new release of pip is available: 24.1 -> 24.3.1
[notice] To update, run: python.exe -m pip install --upgrade pip
```

Dentro de nuestro código de Python importamos las herramientas para poder utilizarlas únicamente con las letras “pd”

```
LABORATORIO 8_DATASETS > dataset.py > ...
import pandas as pd
```

Posteriormente asignamos la ruta en donde se encuentra el archivo que deseamos leer, en este caso usamos la ruta completa.

```
# Ruta al archivo bezdekIris.data
archivo = "C:/Users/jorge/OneDrive/Documentos/JORGE/ESCOM/9_SEMESTRE/IA/LABORATORIO/8_DATASETS/iris/b
```

Dentro del sitio web nos indica que los valores son los que tiene cada una de las columnas, entonces estos los asignamos dentro de un arreglo para que sirvan de título de nuestro dataset.

Variable Name	Role	Type	Description	Units	Missing Values
sepal length	Feature	Continuous		cm	no
sepal width	Feature	Continuous		cm	no
petal length	Feature	Continuous		cm	no
petal width	Feature	Continuous		cm	no
class	Target	Categorical	class of iris plant: Iris Setosa, Iris Versicolour, or Iris Virginica		no

Rows per page 10 0 to 5 of 5

```
# Nombres de las columnas según el dataset Iris
columnas = ["longitud sepalo en cm", "ancho sepalo en cm", "longitud petalo en cm", "ancho petalo en cm", "clase"]
```

Y ahora cargamos el archivo, para ello ocuamos panda, en donde le decimos que deber leer el archivo que se encuentra en la ruta especificada, y le asignamos el nombre de las columnas que hemos asignado.

```
# Carga el archivo en un DataFrame
df = pd.read_csv(archivo, header=None, names=columnas)
```

Por ultimo simplemente imprimos una parte de lo que se ha leído y hacemos una prueba de ejecución.

```
# Imprime 5 filas del DataFrame
print(df.head())
```

```

ive/Documentos/JORGE/ESCOM/9_SEMESTRE/IA/LABORATORIO/8_DATASETS/dataset.py
logitud sepalo en cm ancho sepalo en cm longitud petalo en cm ancho petalo en cm clase
0      5.1         3.5         1.4         0.2  Iris-setosa
1      4.9         3.0         1.4         0.2  Iris-setosa
2      4.7         3.2         1.3         0.2  Iris-setosa
3      4.6         3.1         1.5         0.2  Iris-setosa
4      5.0         3.6         1.4         0.2  Iris-setosa

```

Ahora para la versión de java debemos de usar buffer y un lector de archivos, para ello importamos las herramientas.

```
import java.io.BufferedReader;
import java.io.FileReader;
import java.io.IOException;
```

Ahora debemos de señalar la ruta completa en donde se encuentra nuestro archivo.

```
// Ruta al archivo bezdekIris.data
String filePath = "C:/Users/jorge/OneDrive/Documentos/JORGE/ESCOM/9_SEMESTRE/IA/LABORATORIO/8
```

De la misma forma asignamos el nombre de las columnas correspondientes.

```
// Nombres de las columnas según el dataset Iris
String[] columnas = {"longitud sepalo en cm", "ancho sepalo en cm", "longitud petalo en cm", "
```

Y ahora creamos una matriz en donde debemos de almacenar cada uno de los leído, para ello creamos un buffer en donde le asignaremos el archivo leído y generaremos un ciclo en donde debemos de ir línea por línea. Ahora dentro de cada una de las líneas debemos de ir separando las columnas, mismas que están separadas por “,” y vamos almacenando dentro del arreglo.

```
while ((line = br.readLine()) != null && row < 150) {
    // Dividir la línea por comas y almacenar en la matriz
    String[] values = line.split(",");
    System.arraycopy(values, 0, data[row], 0, values.length);
    row++;
}
```

Por último de la misma forma imprimimos los resultados de lo leído, para ello de nuevo debemos de hacer un ciclo que recorra la matriz que hemos generado.

```
// Imprimir los nombres de las columnas
for (String columna : columnas) {
    System.out.print(columna + "\t");
}
System.out.println();

// Imprimir las primeras 5 filas para verificar
for (int i = 0; i < 5; i++) {
    for (int j = 0; j < 5; j++) {
        System.out.print(data[i][j] + "\t");
    }
    System.out.println();
}
}
```

```
PS C:\Users\jorge\OneDrive\Documentos\JORGE\ESCOM\9_SEMESTRE\IA\LABORATORIO\8_DATASETS> java Dataset
logitud sepalo en cm    ancho sepalo en cm    longitud petalo en cm    ancho petalo en cm    clase
5.1    3.5    1.4    0.2    Iris-setosa
4.9    3.0    1.4    0.2    Iris-setosa
4.7    3.2    1.3    0.2    Iris-setosa
4.6    3.1    1.5    0.2    Iris-setosa
5.0    3.6    1.4    0.2    Iris-setosa
PS C:\Users\jorge\OneDrive\Documentos\JORGE\ESCOM\9_SEMESTRE\IA\LABORATORIO\8_DATASETS>
```

Enlace

A continuación en enlace en donde se encuentran los código correspondientes a esta practica:https://github.com/Jorge300403/IA_6CV3_MartinezChavez/tree/main/LAB_8_DATASET