LAB 1 - Iris Dataset Classification

By Bilel RAHMOUNI

Dependencies

In [ ]:

```
pip install ydata-profiling
```

```
Collecting ydata-profiling
  Downloading ydata_profiling-4.9.0-py2.py3-none-any.whl.metadata (20 kB)
Requirement already satisfied: scipy<1.14,>=1.4.1 in /usr/local/lib/python3.10/dist-packa
ges (from ydata-profiling) (1.13.1)
Requirement already satisfied: pandas!=1.4.0,<3,>1.1 in /usr/local/lib/python3.10/dist-pa
ckages (from ydata-profiling) (2.1.4)
Requirement already satisfied: matplotlib<3.10,>=3.5 in /usr/local/lib/python3.10/dist-pa
ckages (from ydata-profiling) (3.7.1)
Requirement already satisfied: pydantic>=2 in /usr/local/lib/python3.10/dist-packages (fr
om ydata-profiling) (2.8.2)
Requirement already satisfied: PyYAML<6.1,>=5.0.0 in /usr/local/lib/python3.10/dist-packa
ges (from ydata-profiling) (6.0.2)
Requirement already satisfied: jinja2<3.2,>=2.11.1 in /usr/local/lib/python3.10/dist-pack
ages (from ydata-profiling) (3.1.4)
Collecting visions<0.7.7,>=0.7.5 (from visions[type_image_path]<0.7.7,>=0.7.5->ydata-prof
iling)
  Downloading visions-0.7.6-py3-none-any.whl.metadata (11 kB)
Requirement already satisfied: numpy<2,>=1.16.0 in /usr/local/lib/python3.10/dist-package
s (from ydata-profiling) (1.26.4)
Collecting htmlmin==0.1.12 (from ydata-profiling)
  Downloading htmlmin-0.1.12.tar.gz (19 kB)
  Preparing metadata (setup.py) ... done
Collecting phik<0.13,>=0.11.1 (from ydata-profiling)
  Downloading phik-0.12.4-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.meta
data (5.6 kB)
Requirement already satisfied: requests<3,>=2.24.0 in /usr/local/lib/python3.10/dist-pack
ages (from ydata-profiling) (2.32.3)
Requirement already satisfied: tqdm<5,>=4.48.2 in /usr/local/lib/python3.10/dist-packages
(from ydata-profiling) (4.66.5)
Requirement already satisfied: seaborn<0.14,>=0.10.1 in /usr/local/lib/python3.10/dist-pa
ckages (from ydata-profiling) (0.13.1)
Collecting multimethod<2,>=1.4 (from ydata-profiling)
  Downloading multimethod-1.12-py3-none-any.whl.metadata (9.6 kB)
Requirement already satisfied: statsmodels<1,>=0.13.2 in /usr/local/lib/python3.10/dist-p
ackages (from ydata-profiling) (0.14.2)
Requirement already satisfied: typeguard<5,>=3 in /usr/local/lib/python3.10/dist-packages
(from ydata-profiling) (4.3.0)
Collecting imagehash==4.3.1 (from ydata-profiling)
  Downloading ImageHash-4.3.1-py2.py3-none-any.whl.metadata (8.0 kB)
Requirement already satisfied: wordcloud>=1.9.1 in /usr/local/lib/python3.10/dist-package
s (from ydata-profiling) (1.9.3)
Collecting dacite>=1.8 (from ydata-profiling)
  Downloading dacite-1.8.1-py3-none-any.whl.metadata (15 kB)
Requirement already satisfied: numba<1,>=0.56.0 in /usr/local/lib/python3.10/dist-package
s (from ydata-profiling) (0.60.0)
Collecting PyWavelets (from imagehash==4.3.1->ydata-profiling)
  Downloading pywavelets-1.7.0-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl
.metadata (9.0 kB)
Requirement already satisfied: pillow in /usr/local/lib/python3.10/dist-packages (from im
agehash==4.3.1->ydata-profiling) (9.4.0)
Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.10/dist-packages
(from jinja2<3.2,>=2.11.1->ydata-profiling) (2.1.5)
Requirement already satisfied: contourpy>=1.0.1 in /usr/local/lib/python3.10/dist-package
s (from matplotlib<3.10,>=3.5->ydata-profiling) (1.3.0)
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.10/dist-packages (f
```

```
from matplotlib<3.10,>=3.5->ydata-profiling) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in /usr/local/lib/python3.10/dist-packag
es (from matplotlib<3.10,>=3.5->ydata-profiling) (4.53.1)
Requirement already satisfied: kiwisolver>=1.0.1 in /usr/local/lib/python3.10/dist-packag
es (from matplotlib<3.10,>=3.5->ydata-profiling) (1.4.5)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.10/dist-packages
(from matplotlib<3.10,>=3.5->ydata-profiling) (24.1)
Requirement already satisfied: pyparsing>=2.3.1 in /usr/local/lib/python3.10/dist-package
s (from matplotlib<3.10,>=3.5->ydata-profiling) (3.1.4)
Requirement already satisfied: python-dateutil>=2.7 in /usr/local/lib/python3.10/dist-pac
kages (from matplotlib<3.10,>=3.5->ydata-profiling) (2.8.2)
Requirement already satisfied: llvmlite<0.44,>=0.43.0dev0 in /usr/local/lib/python3.10/di
st-packages (from numba<1,>=0.56.0->ydata-profiling) (0.43.0)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/dist-packages (f
rom pandas!=1.4.0,<3,>1.1->ydata-profiling) (2024.1)
Requirement already satisfied: tzdata>=2022.1 in /usr/local/lib/python3.10/dist-packages
(from pandas!=1.4.0,<3,>1.1->ydata-profiling) (2024.1)
Requirement already satisfied: joblib>=0.14.1 in /usr/local/lib/python3.10/dist-packages
(from phik<0.13,>=0.11.1->ydata-profiling) (1.4.2)
Requirement already satisfied: annotated-types>=0.4.0 in /usr/local/lib/python3.10/dist-p
ackages (from pydantic>=2->ydata-profiling) (0.7.0)
Requirement already satisfied: pydantic-core==2.20.1 in /usr/local/lib/python3.10/dist-pa
ckages (from pydantic>=2->ydata-profiling) (2.20.1)
Requirement already satisfied: typing-extensions>=4.6.1 in /usr/local/lib/python3.10/dist
-packages (from pydantic>=2->ydata-profiling) (4.12.2)
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.10/dist
-packages (from requests<3,>=2.24.0->ydata-profiling) (3.3.2)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (f
rom requests<3,>=2.24.0->ydata-profiling) (3.8)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.10/dist-packa
ges (from requests<3,>=2.24.0->ydata-profiling) (2.0.7)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.10/dist-packa
ges (from requests<3,>=2.24.0->ydata-profiling) (2024.8.30)
Requirement already satisfied: patsy>=0.5.6 in /usr/local/lib/python3.10/dist-packages (f
rom statsmodels<1,>=0.13.2->ydata-profiling) (0.5.6)
Requirement already satisfied: attrs>=19.3.0 in /usr/local/lib/python3.10/dist-packages (
from visions<0.7.7,>=0.7.5->visions[type_image_path]<0.7.7,>=0.7.5->ydata-profiling) (24.
2.0)
Requirement already satisfied: networkx>=2.4 in /usr/local/lib/python3.10/dist-packages (
from visions<0.7.7,>=0.7.5->visions[type_image_path]<0.7.7,>=0.7.5->ydata-profiling) (3.3
)
Requirement already satisfied: six in /usr/local/lib/python3.10/dist-packages (from patsy
>=0.5.6->statsmodels<1,>=0.13.2->ydata-profiling) (1.16.0)
Downloading ydata_profiling-4.9.0-py2.py3-none-any.whl (356 kB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 356.2/356.2 kB 11.4 MB/s eta 0:00:00
Downloading ImageHash-4.3.1-py2.py3-none-any.whl (296 kB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 296.5/296.5 kB 18.6 MB/s eta 0:00:00
Downloading dacite-1.8.1-py3-none-any.whl (14 kB)
Downloading multimethod-1.12-py3-none-any.whl (10 kB)
Downloading phik-0.12.4-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (686 k
B)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 686.1/686.1 kB 32.8 MB/s eta 0:00:00
Downloading visions-0.7.6-py3-none-any.whl (104 kB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 104.8/104.8 kB 7.8 MB/s eta 0:00:00
Downloading pywavelets-1.7.0-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (
4.5 MB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 4.5/4.5 MB 55.4 MB/s eta 0:00:00
Building wheels for collected packages: htmlmin
  Building wheel for htmlmin (setup.py) ... done
  Created wheel for htmlmin: filename=htmlmin-0.1.12-py3-none-any.whl size=27081 sha256=9
e65ef35e77e24fd3e404960a2e50ab8452dd37ba1bfcde302867f57e3c26913
  Stored in directory: /root/.cache/pip/wheels/dd/91/29/a79cecb328d01739e64017b6fb9a1ab9d
8cb1853098ec5966d
Successfully built htmlmin
Installing collected packages: htmlmin, PyWavelets, multimethod, dacite, imagehash, visio
ns, phik, ydata-profiling
Successfully installed PyWavelets-1.7.0 dacite-1.8.1 htmlmin-0.1.12 imagehash-4.3.1 multi
method-1.12 phik-0.12.4 visions-0.7.6 ydata-profiling-4.9.0
```

Part 2 : Data Presentation

In [ ]:

```
from sklearn.datasets import load_iris
import pandas as pd
```

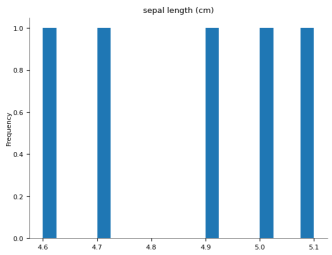
In [ ]:

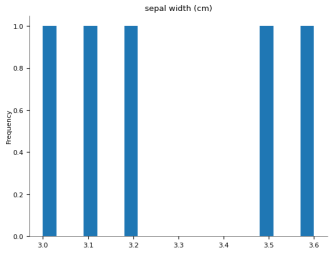
```
# Load the Iris dataset
iris = load_iris()
# Convert to DataFrame
iris_df = pd.DataFrame(data=iris.data, columns=iris.feature_names)
iris_df['species'] = pd.Categorical.from_codes(iris.target, iris.target_names)
# Display the first few rows of the dataset
iris_df.head()
```

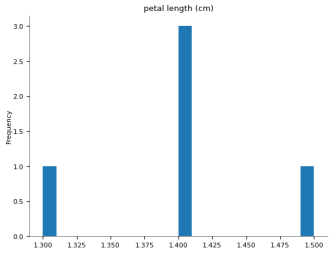
Out[ ]:

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)	species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa

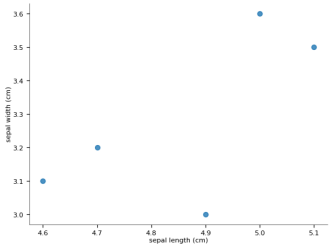
Distributions

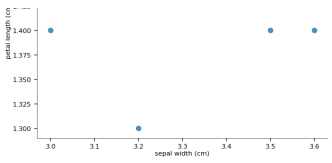




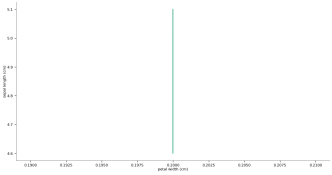


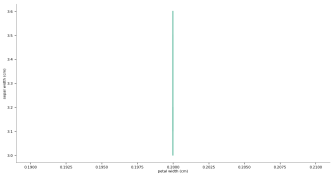
2-d distributions

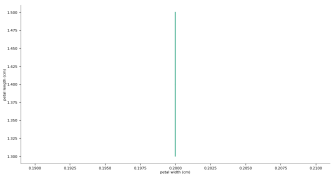


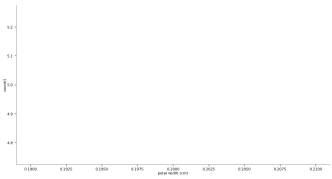


Time series

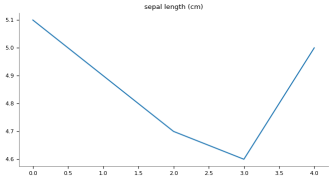


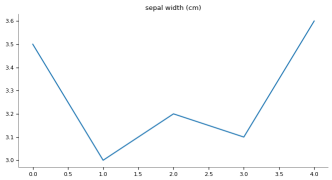


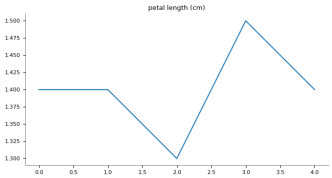




Values







Part 3: Data Exploration with ydata_profiling

In [13]:

```
from ydata_profiling import ProfileReport
# Generate the profiling report
profile = ProfileReport(iris_df, title="Iris Dataset Profiling Report")
profile.to_notebook_iframe()
```

Part 4: Classification Task

Split the Data into Training and Testing Sets:

In [ ]:

```
from sklearn.model_selection import train_test_split
X = iris_df.drop(columns=['species'])
y = iris_df['species']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42
)
```

Train a Classification Model:

In [ ]:

```
from sklearn.linear_model import LogisticRegression
# Initialize the Logistic Regression model
model = LogisticRegression(max_iter=200)
# Train the model on the training data
model.fit(X_train, y_train)
```

Out[ ]:

```
LogisticRegression(max_iter=200)
```

In [ ]:

```
from sklearn.svm import SVC
# Initialize the model
model = SVC(kernel='linear')
# Train the model
model.fit(X_train, y_train)
```

Out[ ]:

```
SVC(kernel='linear')
```

Evaluate the Model:

In [ ]:

```
from sklearn.metrics import classification_report, accuracy_score
# Make predictions
y_pred = model.predict(X_test)
# Print evaluation metrics
print("Accuracy:", accuracy_score(y_test, y_pred))
print("Classification Report:", classification_report(y_test, y_pred))
```

```
Accuracy: 1.0
Classification Report:               precision    recall  f1-score   support

      setosa       1.00      1.00      1.00        19
  versicolor       1.00      1.00      1.00        13
   virginica       1.00      1.00      1.00        13

    accuracy                           1.00        45
   macro avg       1.00      1.00      1.00        45
weighted avg       1.00      1.00      1.00        45
```

We can see that the tests will always be correct since the precision is 1.00