

MTMC-512 Programming Lab IV(Machine Learning)

Lab Assignment 1: Python ML Environment Setup

1 Task 1: Environment Setup

1.1 Installing Required Libraries

```
1 import numpy as np
2 import pandas as pd
3 import seaborn as sns
4 import matplotlib.pyplot as plt
5 %matplotlib inline
6 import warnings
7 warnings.filterwarnings('ignore')
```

1.2 Verifying Installation Versions

```
1 print(pd.__version__) # Output: 1.3.4
2 print(np.__version__) # Output: 1.20.3
3 print(sns.__version__) # Output: 0.11.2
```

2 Task 2: Load and Explore the Dataset

2.1 Loading the Iris Dataset

```
1 df = sns.load_dataset('iris')
2 df.head()
```

Output:

	sepal_length	sepal_width	petal_length	petal_width	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

2.2 Display Basic Dataset Information

2.2.1 Number of Rows and Columns:

```
1 df.shape # Output: (150, 5)
```

2.2.2 Data Types of Features:

```
1 df.dtypes
```

Output:

```
sepal_length    float64
sepal_width     float64
petal_length    float64
petal_width     float64
species         object
```

2.2.3 Summary Statistics:

```
1 df.describe()
```

3 Task 3: Data Visualization

3.1 Histogram for Feature Distributions

```
1 plt.figure(figsize=(10,6))
2 sns.histplot(df, kde=True)
3 plt.title('Histogram for Iris Dataset')
4 plt.savefig("histogram_iris.png")
5 plt.show()
```

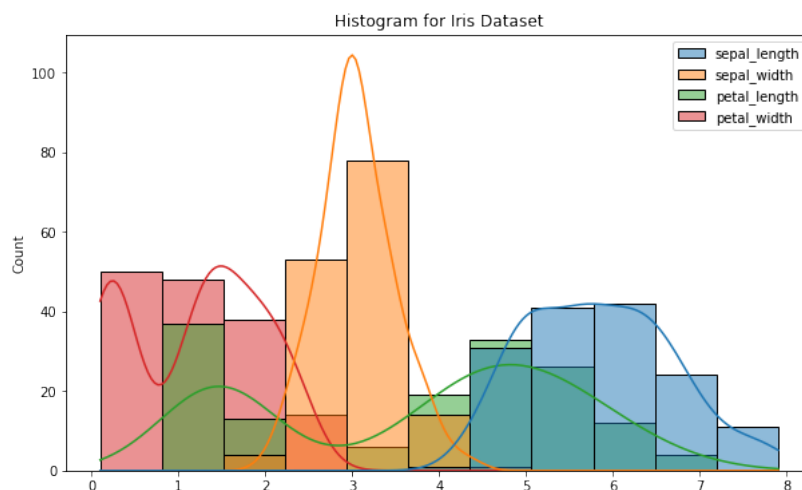


Figure 1: Histogram for Iris Dataset

3.2 Box Plots to Detect Outliers

```
1 plt.figure(figsize=(10,6))
2 sns.boxplot(data=df)
3 plt.title('Boxplot for Iris Dataset')
4 plt.savefig("boxplot_iris.png")
5 plt.show()
```

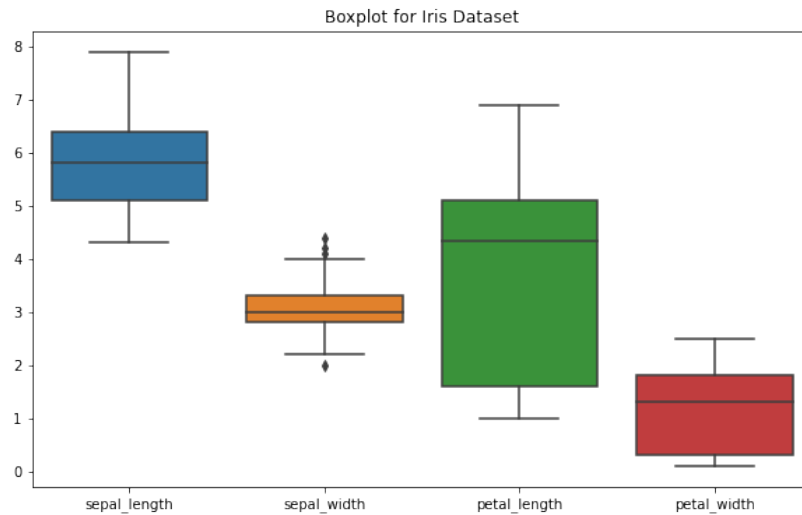


Figure 2: Boxplot for Iris Dataset

3.3 Pair Plots to Explore Feature Relationships

```
1 sns.pairplot(df, hue='species')
2 plt.savefig("pairplot_iris.png")
3 plt.show()
```

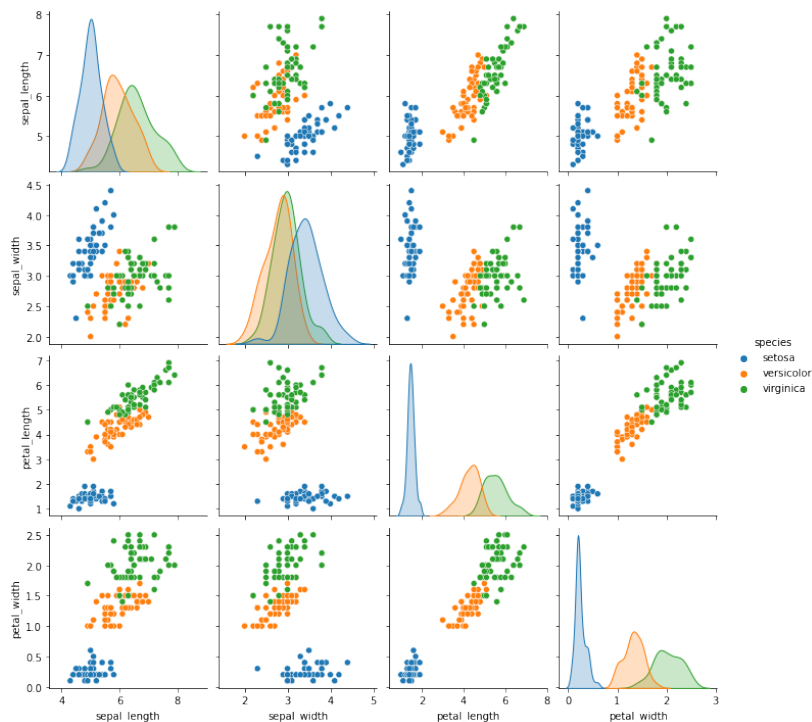


Figure 3: Pairplot for Iris Dataset

3.4 Correlation Heatmap

```
1 plt.figure(figsize=(20,12))
2 sns.heatmap(df.corr(), annot=True)
3 plt.savefig("heatmap_iris.png")
4 plt.show()
```

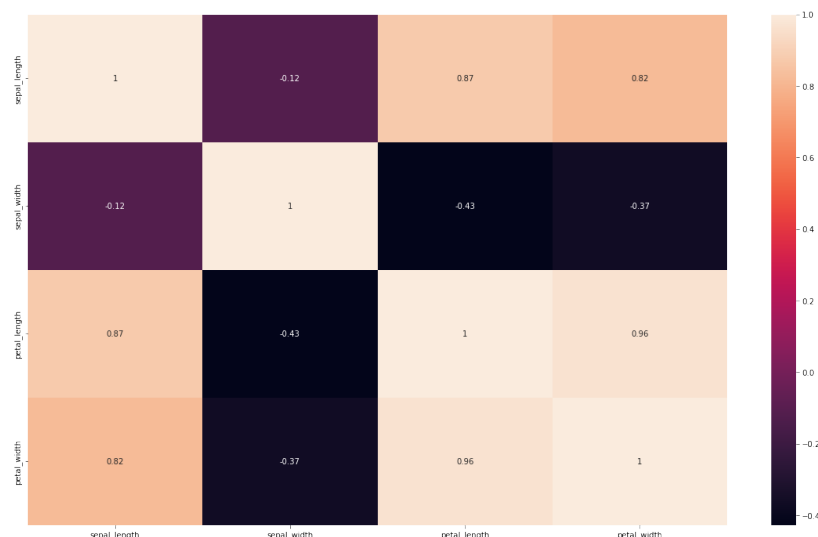


Figure 4: Correlation Heatmap for Iris Dataset

4 Task 4: Basic Data Preprocessing

4.1 Handling Missing Values

```
1 df.isnull().sum() # Output: No missing values
```

4.2 Feature Normalization

```
1 from sklearn.preprocessing import StandardScaler
2 scaler = StandardScaler()
3 scaled_features = scaler.fit_transform(df.iloc[:, :-1])
```

4.3 Dataset Splitting (80-20)

```
1 from sklearn.model_selection import train_test_split
2 X = df.iloc[:, :-1].values
3 y = df.iloc[:, -1].values
4 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.20,
    random_state=42)
```

5 Conclusion

This assignment demonstrated the setup of a Python environment for machine learning, data loading, exploration, visualization, and preprocessing using the Iris dataset. The results provide a comprehensive understanding of the data and initial steps in a typical ML workflow.