import pandas as pd

import numpy as np

‘’’Description

The Iris dataset is a classic dataset in machine learning, consisting of 150 samples of iris flowers.

It contains four features: sepal length, sepal width, petal length, and petal width, all measured in centimeters.

These features are used to classify the flowers into three species: setosa, versicolor, and virginica.

The dataset is widely used for testing classification algorithms.

Source

<https://www.kaggle.com/datasets/uciml/iris>

‘’’

df = pd.read\_csv('IRIS.csv')

df.columns

df['species'].unique()

df['species'].value\_counts()

df.isnull().sum()

df['sepal\_length'] = df['sepal\_length'].fillna(df['sepal\_length'].mean())

print(f"Mean of 'sepal\_length': {df['sepal\_length'].mean()}")

df['sepal\_width'] = df['sepal\_width'].fillna(df['sepal\_width'].mean())

print(f"Mean of 'sepal\_width': {df['sepal\_width'].mean()}")

df['petal\_length'] = df['petal\_length'].fillna(df['petal\_length'].mean())

print(f"Mean of 'petal\_length': {df['petal\_length'].mean()}")

df['petal\_width'] = df['petal\_width'].fillna(df['petal\_width'].mean())

print(f"Mean of 'petal\_width': {df['petal\_width'].mean()}")

df.isnull().sum()

df.describe(include='all')

df.dtypes

df.shape

# Summarize the variable types based on data types

print("\nSummary of Variables:")

for col in df.columns:

if df[col].dtype == 'object':

print(f"{col}: Character (String)")

elif df[col].dtype == 'int64':

print(f"{col}: Integer")

elif df[col].dtype == 'float64':

print(f"{col}: Numeric")

elif df[col].dtype == 'bool':

print(f"{col}: Logical (Boolean)")

else:

print(f"{col}: Unknown")

df1 = pd.read\_csv('IRIS.csv')

from sklearn.preprocessing import LabelEncoder

le = LabelEncoder()

df1['species'] = le.fit\_transform(df1['species'])

df1.sample(5)