

naithub

March 13, 2025

```
[ ]: import pandas as pd
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error
from sklearn.datasets import load_diabetes
# Load the dataset

# pandas: A powerful data manipulation library used to read, analyze, and
↳manage datasets.
# matplotlib.pyplot: A plotting library used to create visualizations like
↳scatter plots, line plots, etc.
# train_test_split: A function from sklearn that splits your dataset into
↳training and testing sets. This is crucial for evaluating model performance.
# LinearRegression: A simple linear regression model from sklearn, used to
↳predict a continuous target variable based on one or more features.
# mean_squared_error: A metric used to evaluate the accuracy of a regression
↳model by calculating the average squared difference between actual and
↳predicted values.
# load_diabetes: A built-in dataset in sklearn, commonly used for regression
↳tasks.
```

```
[14]: # Regression is a type of supervised learning used to predict
# continuous numerical values. The goal is to find
# a relationship between independent variables (features) and a dependent
↳variable (target).

# examples Example: Predicting house prices, predicting temperature, or
↳estimating sales.

# 1. load_boston() (Deprecated) his dataset contains information about housing
↳prices in Boston.

# from sklearn.datasets import load_boston
# # data = load_boston()
# load_diabetes()
```

```
# A dataset measuring disease progression in diabetes patients.
# Features: 10 features such as age, BMI, blood pressure, etc.
# Target: A continuous measure of disease progression.
# load_california_housing()
# print(data['DESCR']) # Displays dataset description
```

```
[15]: from sklearn.datasets import fetch_openml
data = fetch_openml(name='abalone', version=1, as_frame=True)
print(data['frame'].head())
```

	Sex	Length	Diameter	Height	Whole_weight	Shucked_weight	Viscera_weight	\
0	M	0.455	0.365	0.095	0.5140	0.2245	0.1010	
1	M	0.350	0.265	0.090	0.2255	0.0995	0.0485	
2	F	0.530	0.420	0.135	0.6770	0.2565	0.1415	
3	M	0.440	0.365	0.125	0.5160	0.2155	0.1140	
4	I	0.330	0.255	0.080	0.2050	0.0895	0.0395	

	Shell_weight	Class_number_of_rings
0	0.150	15
1	0.070	7
2	0.210	9
3	0.155	10
4	0.055	7

```
[5]: data = pd.read_csv('weather_data.csv')

display(data.head())

# Display the first 5 rows
df.head()
```

	Outdoor Drybulb Temperature [C]	Outdoor Relative Humidity [%]	\
0	17.81	68.12	
1	16.14	74.75	
2	16.10	75.88	
3	16.10	78.25	
4	16.16	77.75	

	Diffuse Solar Radiation [W/m2]	Direct Solar Radiation [W/m2]	\
0	0.0	0.0	
1	0.0	0.0	
2	0.0	0.0	
3	0.0	0.0	
4	0.0	0.0	

	6h Prediction Outdoor Drybulb Temperature [C]	\
0	16.19	

1	17.20
2	18.76
3	21.43
4	23.37

	12h Prediction Outdoor Drybulb Temperature [C] \
0	25.29
1	25.96
2	25.90
3	26.08
4	25.84

	24h Prediction Outdoor Drybulb Temperature [C] \
0	18.31
1	17.25
2	15.07
3	15.21
4	12.91

	6h Prediction Outdoor Relative Humidity [%] \
0	76.47
1	72.65
2	67.54
3	56.74
4	50.20

	12h Prediction Outdoor Relative Humidity [%] \
0	44.10
1	41.52
2	41.96
3	42.34
4	42.92

	24h Prediction Outdoor Relative Humidity [%] \
0	67.32
1	72.46
2	82.54
3	80.86
4	93.52

	6h Prediction Diffuse Solar Radiation [W/m2] \
0	0.00
1	41.67
2	101.65
3	151.99
4	187.61

	12h Prediction Diffuse Solar Radiation [W/m2] \
--	---

0	171.45
1	132.00
2	118.76
3	69.99
4	34.75

	24h Prediction Diffuse Solar Radiation [W/m2] \
0	0.0
1	0.0
2	0.0
3	0.0
4	0.0

	6h Prediction Direct Solar Radiation [W/m2] \
0	0.00
1	65.46
2	329.89
3	436.51
4	486.97

	12h Prediction Direct Solar Radiation [W/m2] \
0	577.41
1	708.69
2	595.70
3	582.30
4	329.68

	24h Prediction Direct Solar Radiation [W/m2]
0	0.0
1	0.0
2	0.0
3	0.0
4	0.0

[5]:	Outdoor Drybulb Temperature [C]	Outdoor Relative Humidity [%] \
0	17.81	68.12
1	16.14	74.75
2	16.10	75.88
3	16.10	78.25
4	16.16	77.75

	Diffuse Solar Radiation [W/m2]	Direct Solar Radiation [W/m2] \
0	0.0	0.0
1	0.0	0.0
2	0.0	0.0
3	0.0	0.0
4	0.0	0.0

6h Prediction Outdoor Drybulb Temperature [C] \	
0	16.19
1	17.20
2	18.76
3	21.43
4	23.37

12h Prediction Outdoor Drybulb Temperature [C] \	
0	25.29
1	25.96
2	25.90
3	26.08
4	25.84

24h Prediction Outdoor Drybulb Temperature [C] \	
0	18.31
1	17.25
2	15.07
3	15.21
4	12.91

6h Prediction Outdoor Relative Humidity [%] \	
0	76.47
1	72.65
2	67.54
3	56.74
4	50.20

12h Prediction Outdoor Relative Humidity [%] \	
0	44.10
1	41.52
2	41.96
3	42.34
4	42.92

24h Prediction Outdoor Relative Humidity [%] \	
0	67.32
1	72.46
2	82.54
3	80.86
4	93.52

6h Prediction Diffuse Solar Radiation [W/m2] \	
0	0.00
1	41.67
2	101.65

3	151.99
4	187.61

	12h Prediction Diffuse Solar Radiation [W/m2] \
0	171.45
1	132.00
2	118.76
3	69.99
4	34.75

	24h Prediction Diffuse Solar Radiation [W/m2] \
0	0.0
1	0.0
2	0.0
3	0.0
4	0.0

	6h Prediction Direct Solar Radiation [W/m2] \
0	0.00
1	65.46
2	329.89
3	436.51
4	486.97

	12h Prediction Direct Solar Radiation [W/m2] \
0	577.41
1	708.69
2	595.70
3	582.30
4	329.68

	24h Prediction Direct Solar Radiation [W/m2]
0	0.0
1	0.0
2	0.0
3	0.0
4	0.0

Key Differences and Use Cases Dataset Type Use Case load_boston() Housing Prices Predict house prices load_diabetes() Medical Data Predict diabetes progression load_linnerud() Health & Fitness Predict physiological response load_california_housing() Housing Data Predict California house prices make_regression() Synthetic Data Custom regression testing data fetch_openml() Real-World Data Flexible for multiple projects

[]: