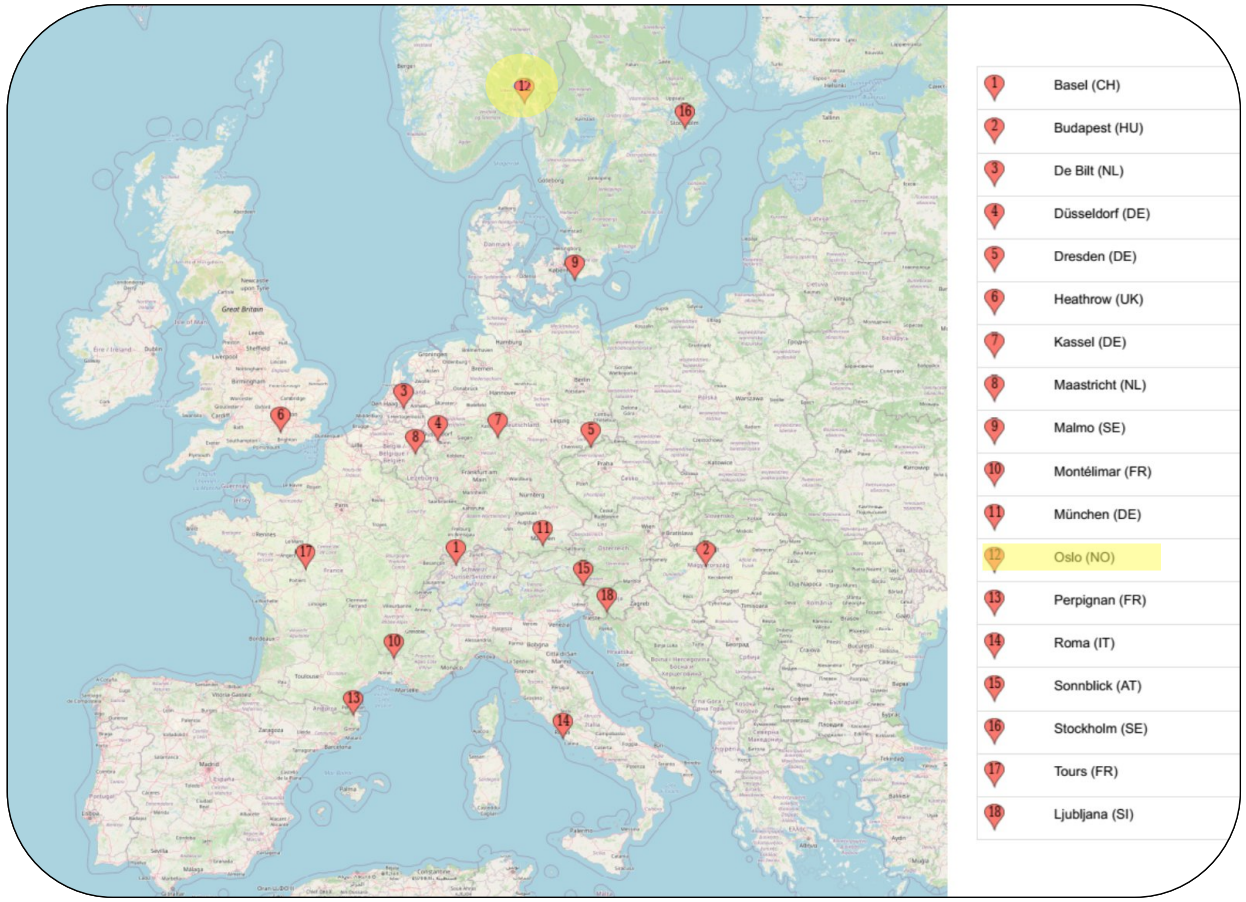
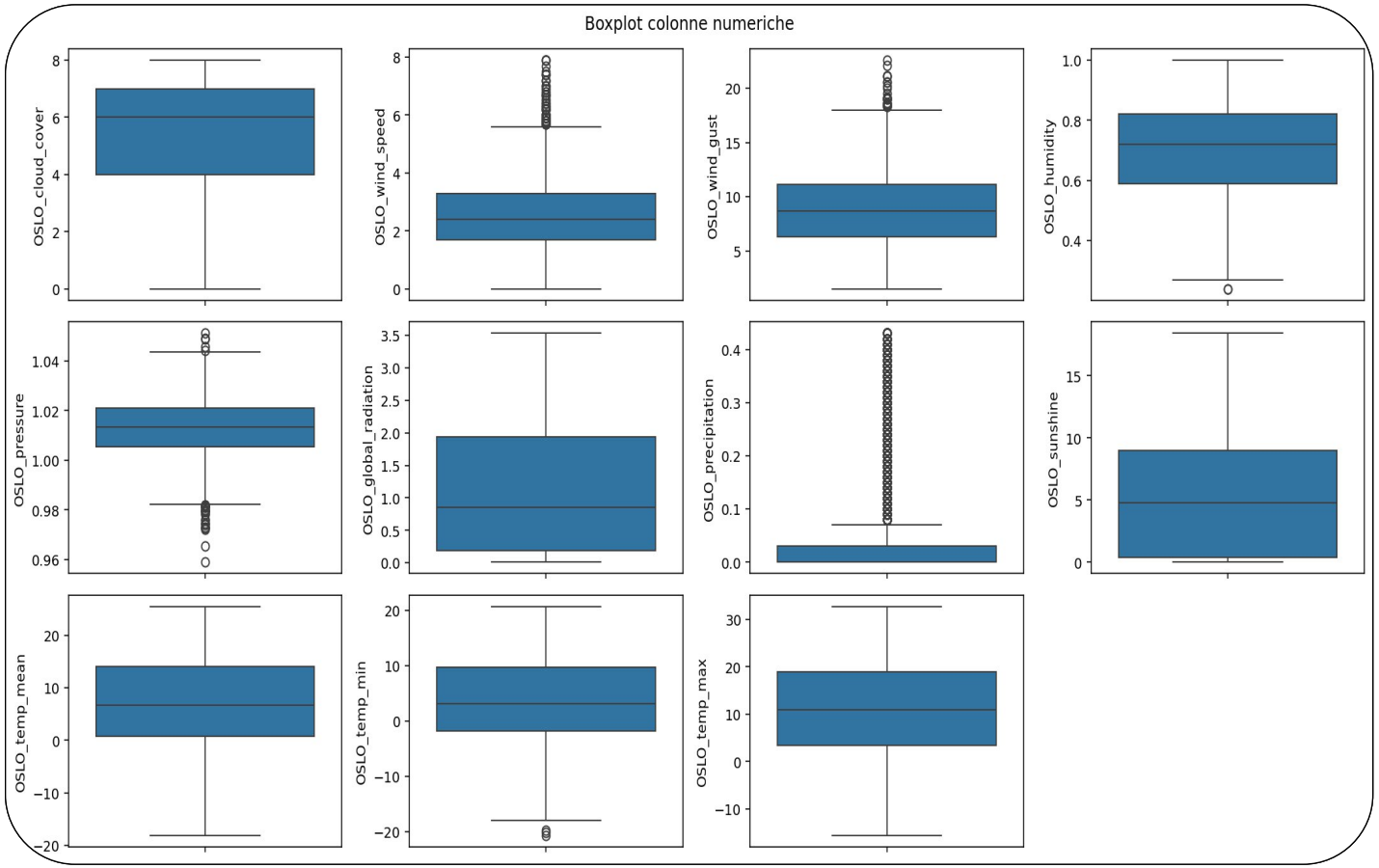


Nome feature	Descrizione	Unità di misura
_cloud_cover	Nuvolosità	okta
_wind_speed	Velocità vento	1 m/s
_wind_gust	Velocità raffiche di vento	1 m/s
_humidity	Umidità	1 %
_pressure	Pressione	1000 hPa
_global_radiation	Irraggiamento	W/m^2
_precipitation	Precipitazioni	10 mm
_sunshine	Ore di luce	0.1 h
_temp_mean	Temperatura media	ºC
_temp_min	Temperatura minima	ºC
_temp_max	Temperatura massima	ºC

Weather dataset

Filippo Bucciarelli
Dataset reference on [Github]

- 3654 registrazioni giornaliere
- 18 città, 17 con classificazione
- 165 parametri meteorologici registrati (massimo 11 per città)



Pre-Processing

- null
- NaN
- duplicates

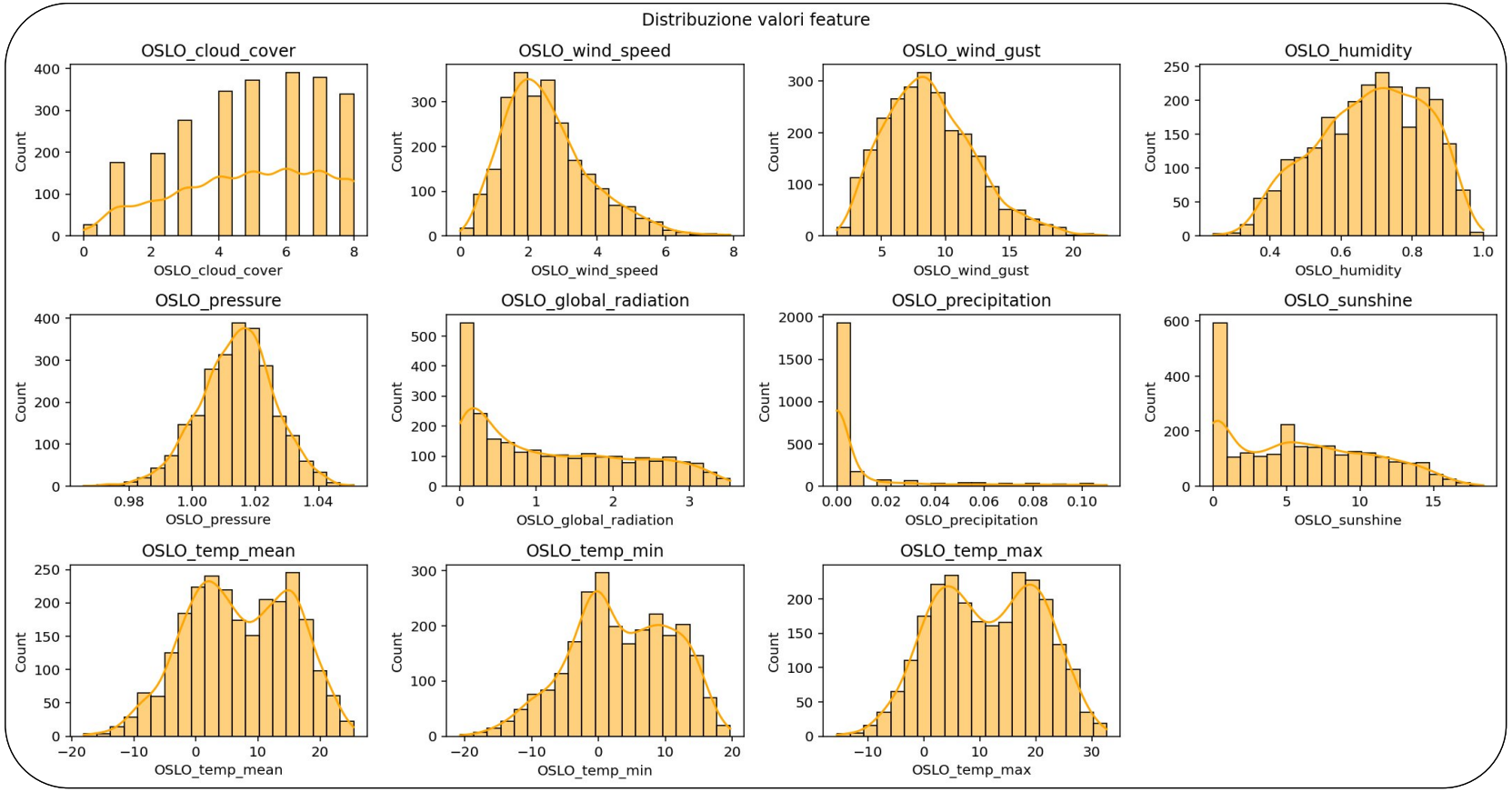
Record rimossi

- `df[df[f"{citta}_sunshine"] < 20]`
- `df[~((df[f"{citta}_temp_min"] < 0) & (df["MONTH"].isin([6, 7, 8])))]`

Valori errati

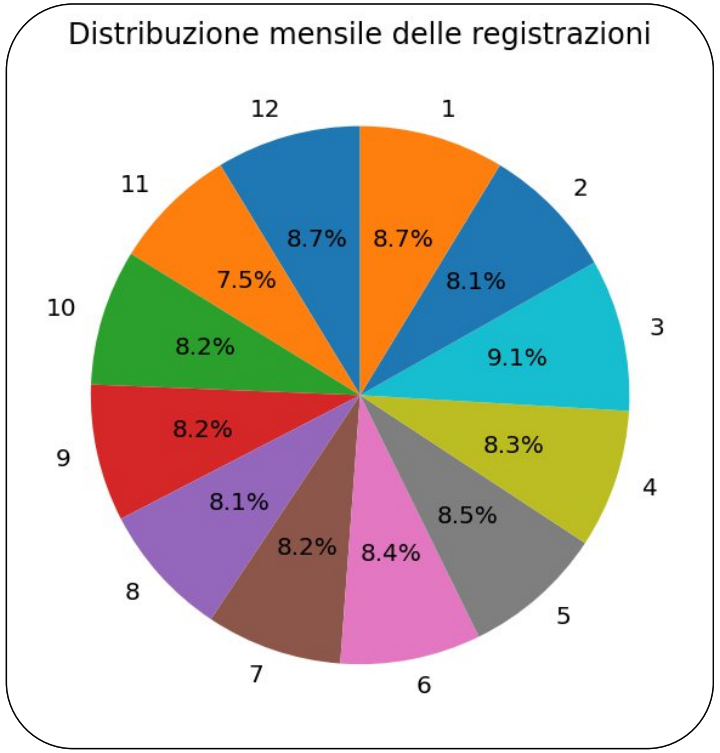
- wind_speed
- wind_gust
- humidity
- pressure
- precipitation

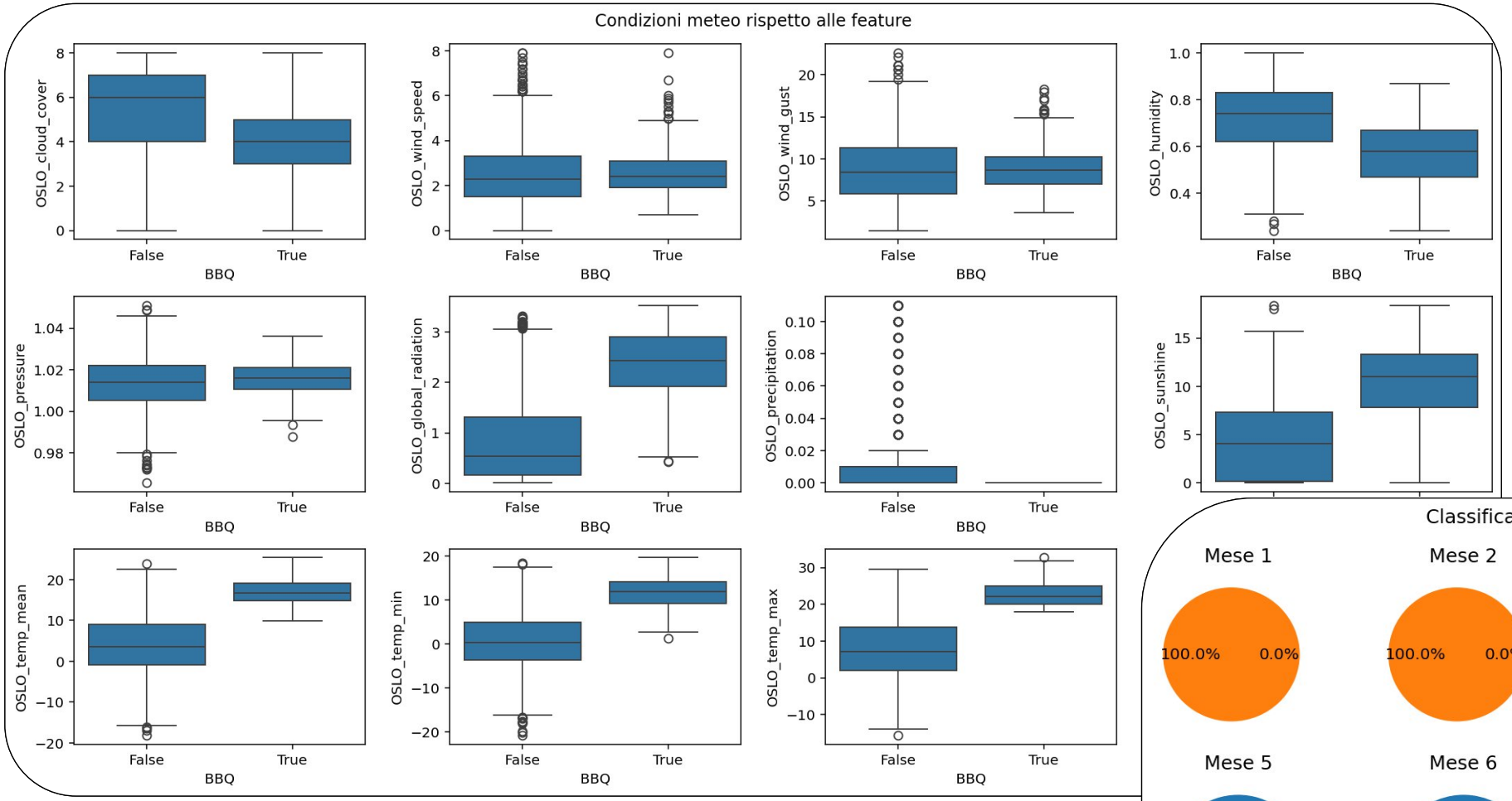
Rimozione outliers sospetti

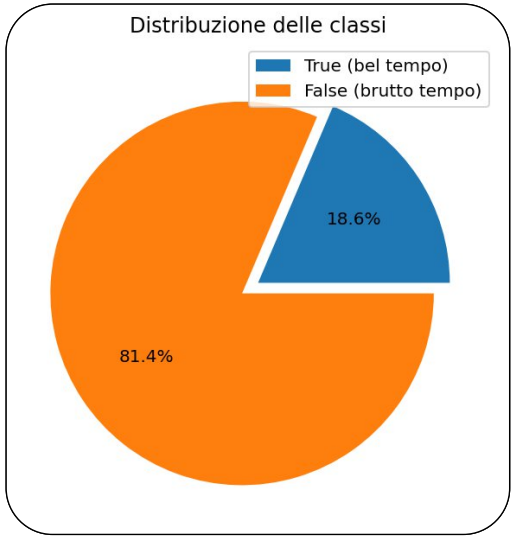


EDA

Distribuzione valori

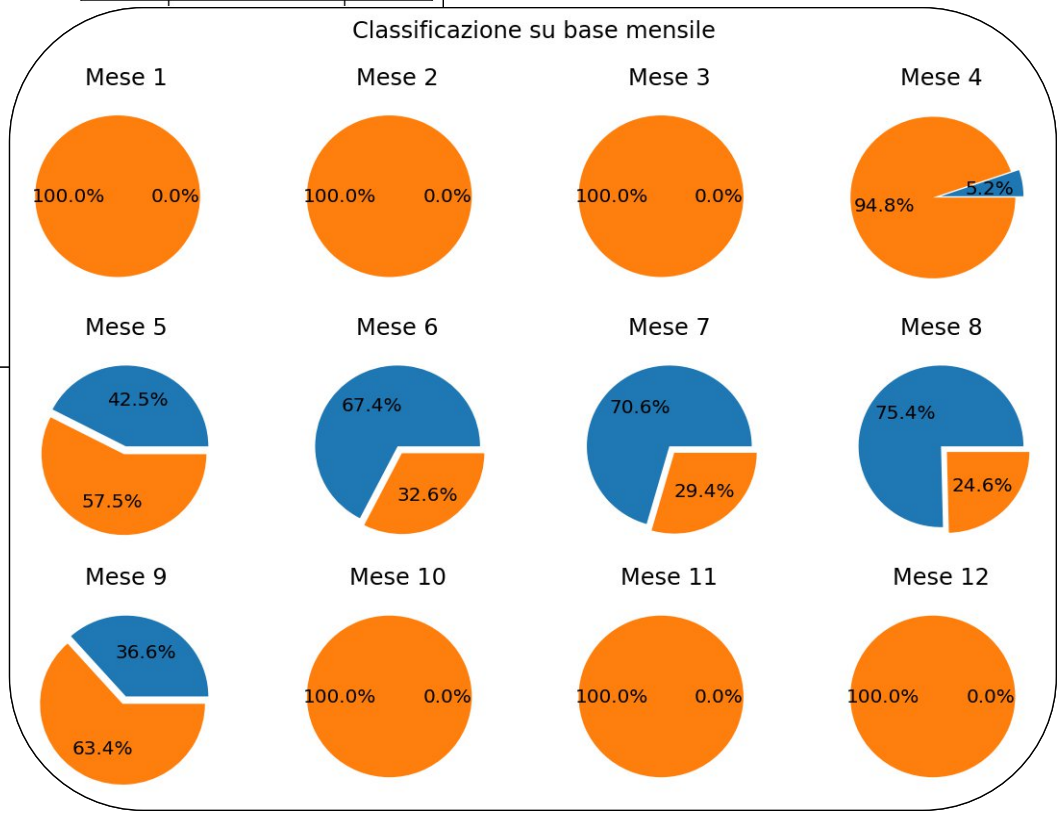


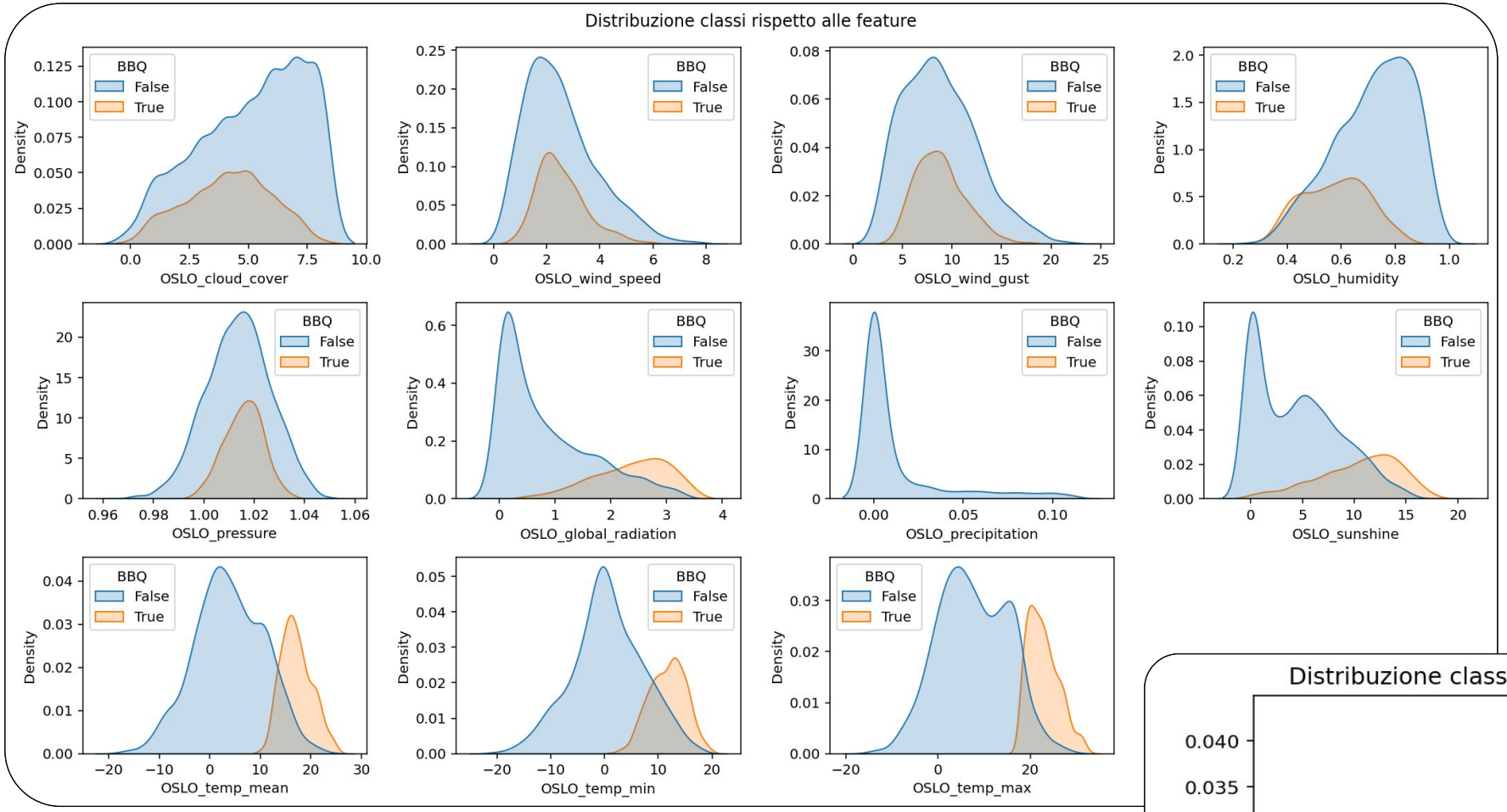


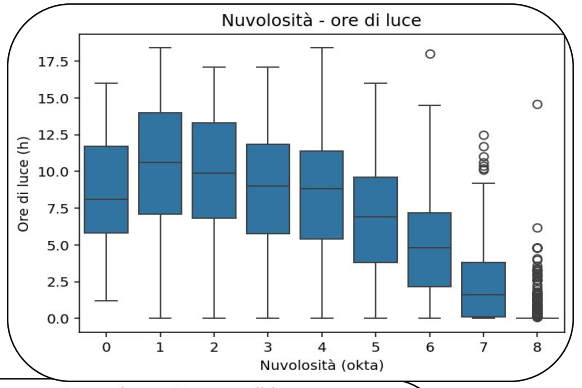


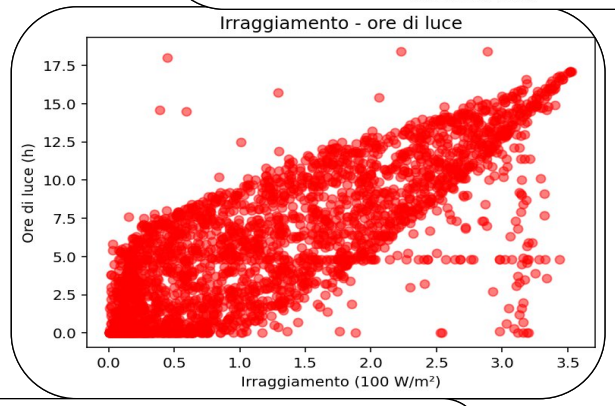
EDA

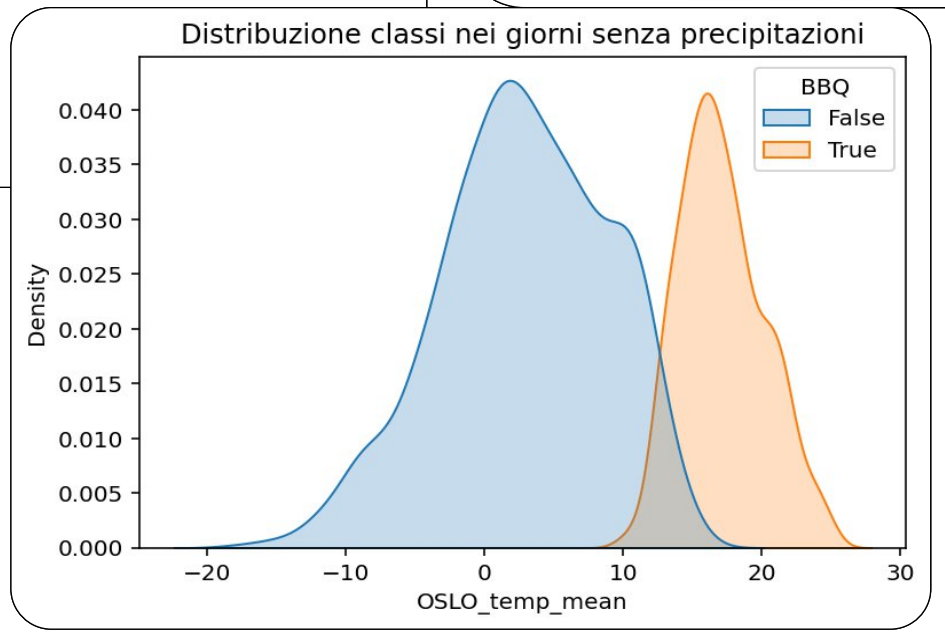
Distribuzione classi











EDA

Analisi bivariata e multivariata

Classificazione

- SVC
- Logistic Regression
- SVM poly
- SVM rbf

Modelli allenati

C: [0.1, 1, 10, 100]
degree: [2, 3, 4]
gamma: [scale, auto, 1]

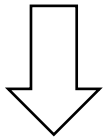
Valori iperparametri SVM

solver: [saga, liblinear]
C: [0.1, 1, 10, 100]

Valori iperparametri Logistic Regression

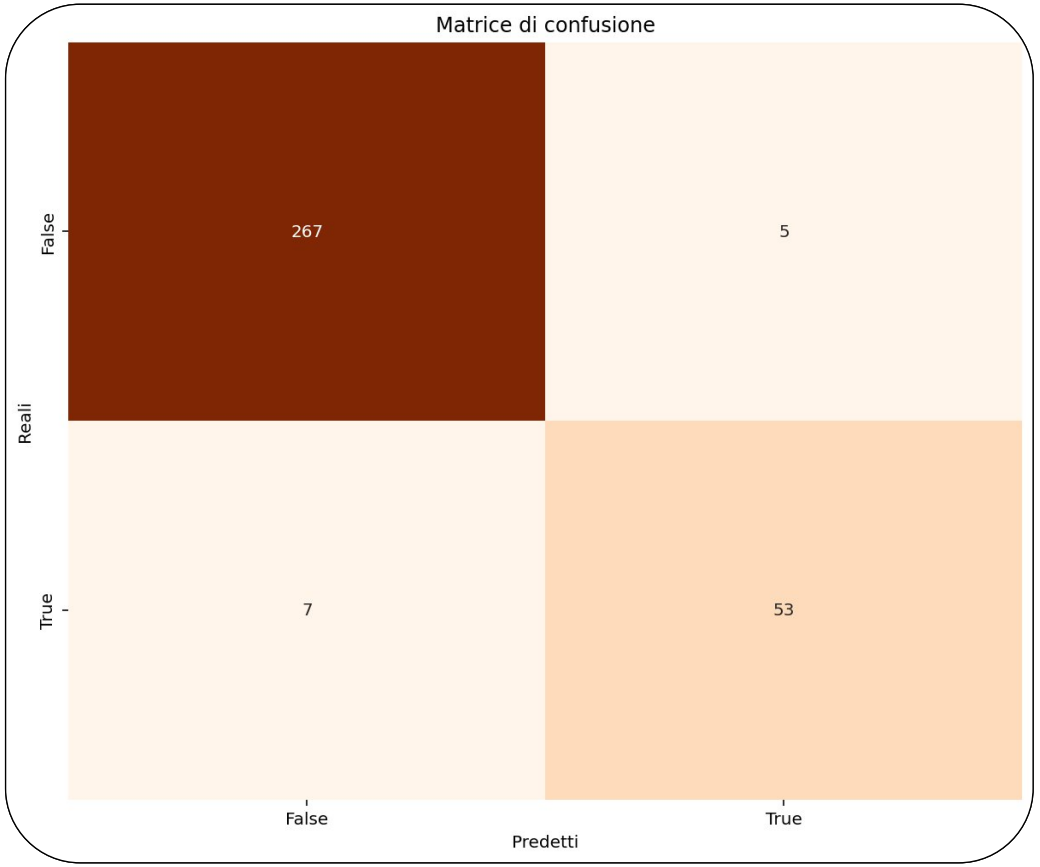
SVC(C=**100**, degree=**2**, kernel='**linear**')

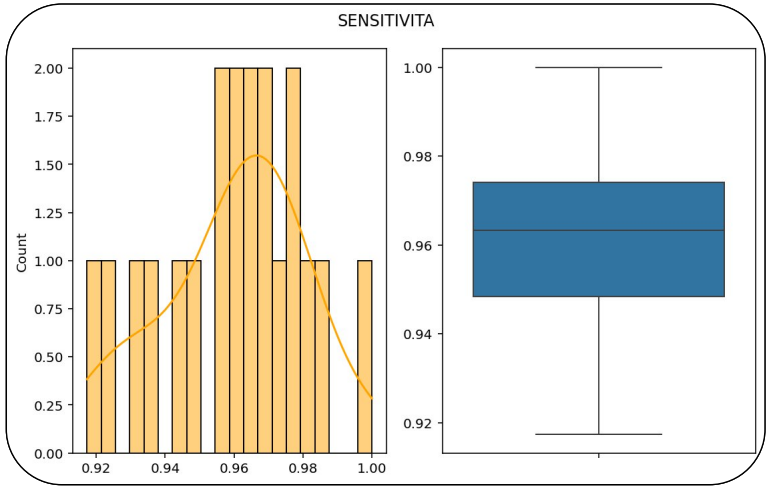
Modello scelto



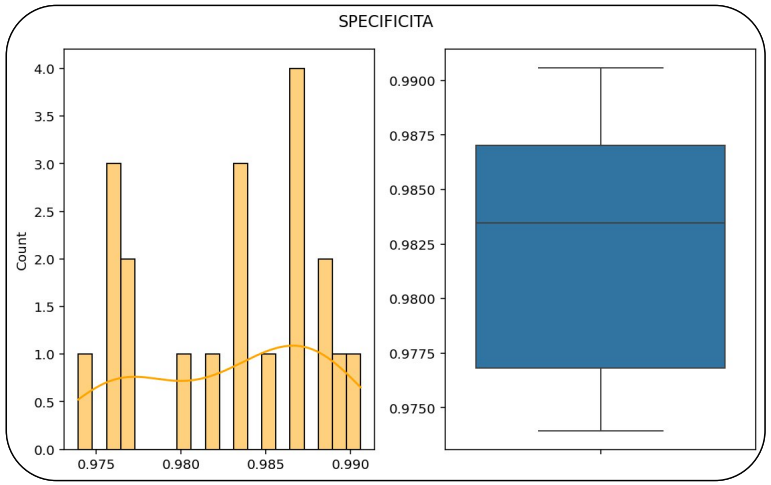
Accuratezza

Testing set: 0.9639
Validation set: 0.9759
Training set: 0.958500



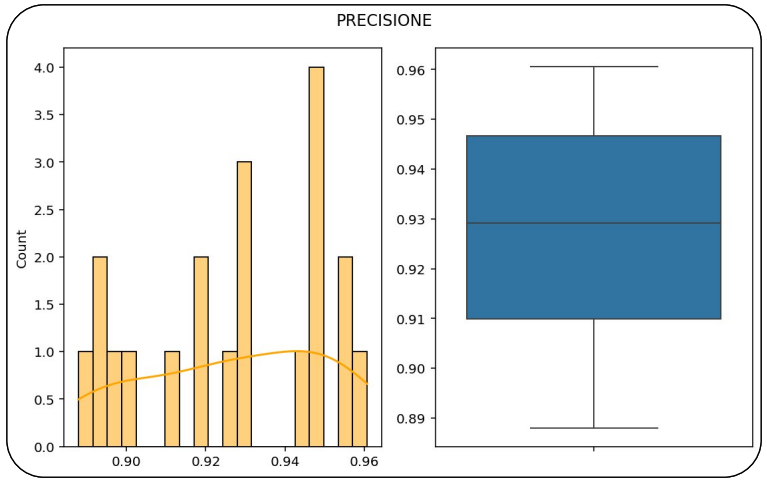


Media: 0.9829
Intervallo di confidenza: [0.9804; 0.9854]
Mediana: 0.9835
Dev. standard: 0.0052
IQR: 0.0102



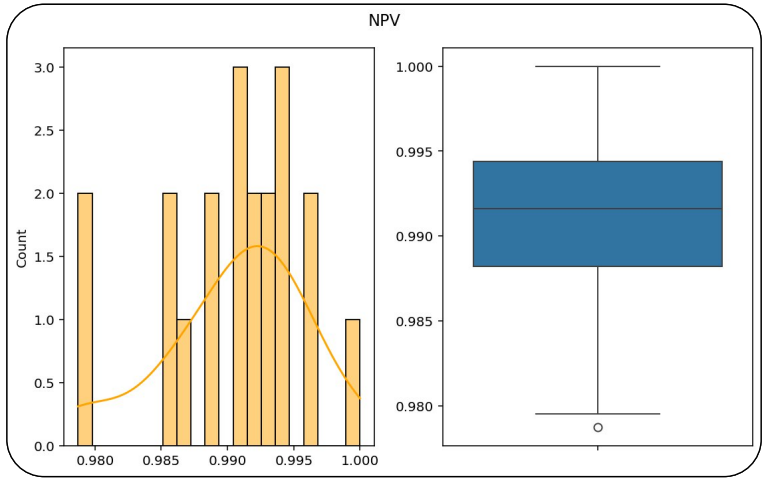
Media: 0.9598
Intervallo di confidenza: [0.9500; 0.9696]
Mediana: 0.9632
Dev. standard: 0.0204
IQR: 0.0256

Classificazione

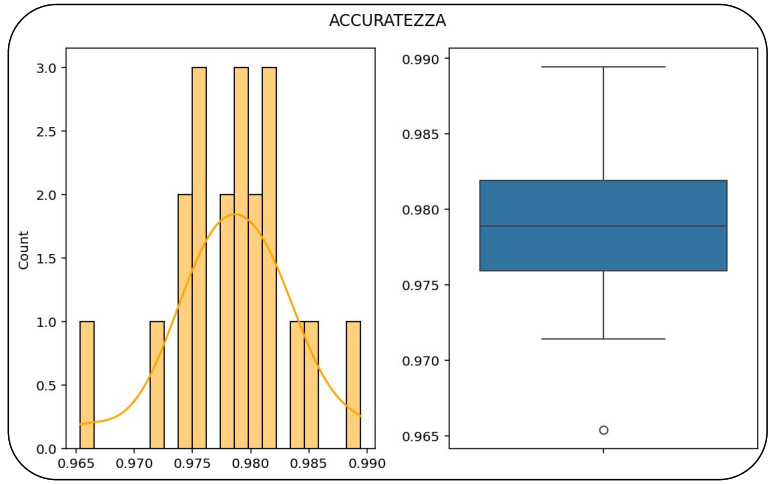


Media: 0.9273
Intervallo di confidenza: [0.9164; 0.9382]
Mediana: 0.9292
Dev. standard: 0.0227
IQR: 0.0368

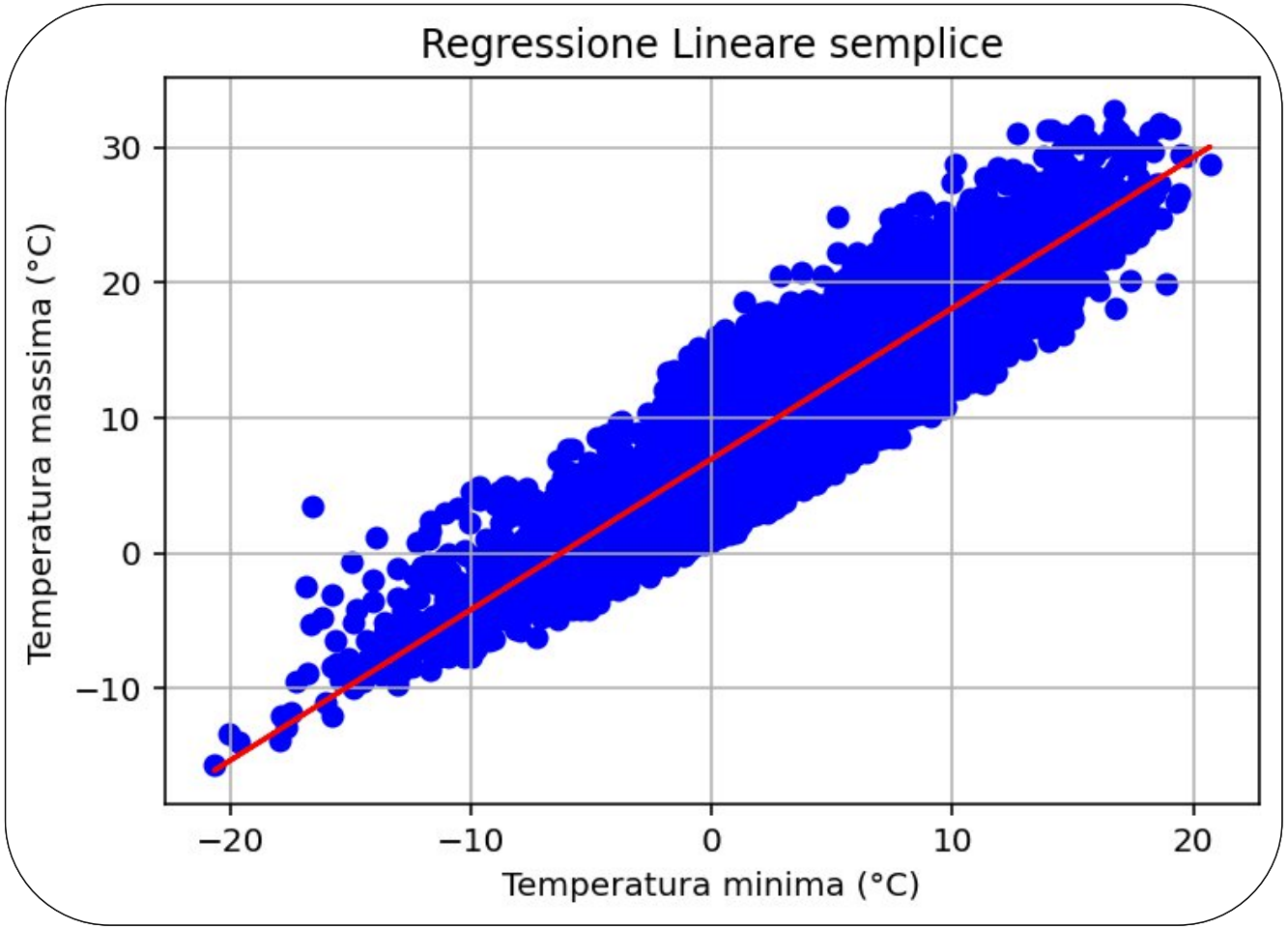
Media: 0.9785
Intervallo di confidenza: [0.9760; 0.9809]
Mediana: 0.9789
Dev. standard: 0.0050
IQR: 0.0060



Media: 0.9906
Intervallo di confidenza: [0.9880; 0.9931]
Mediana: 0.9916
Dev. standard: 0.0053
IQR: 0.0062



Classificazione



Regressione lineare

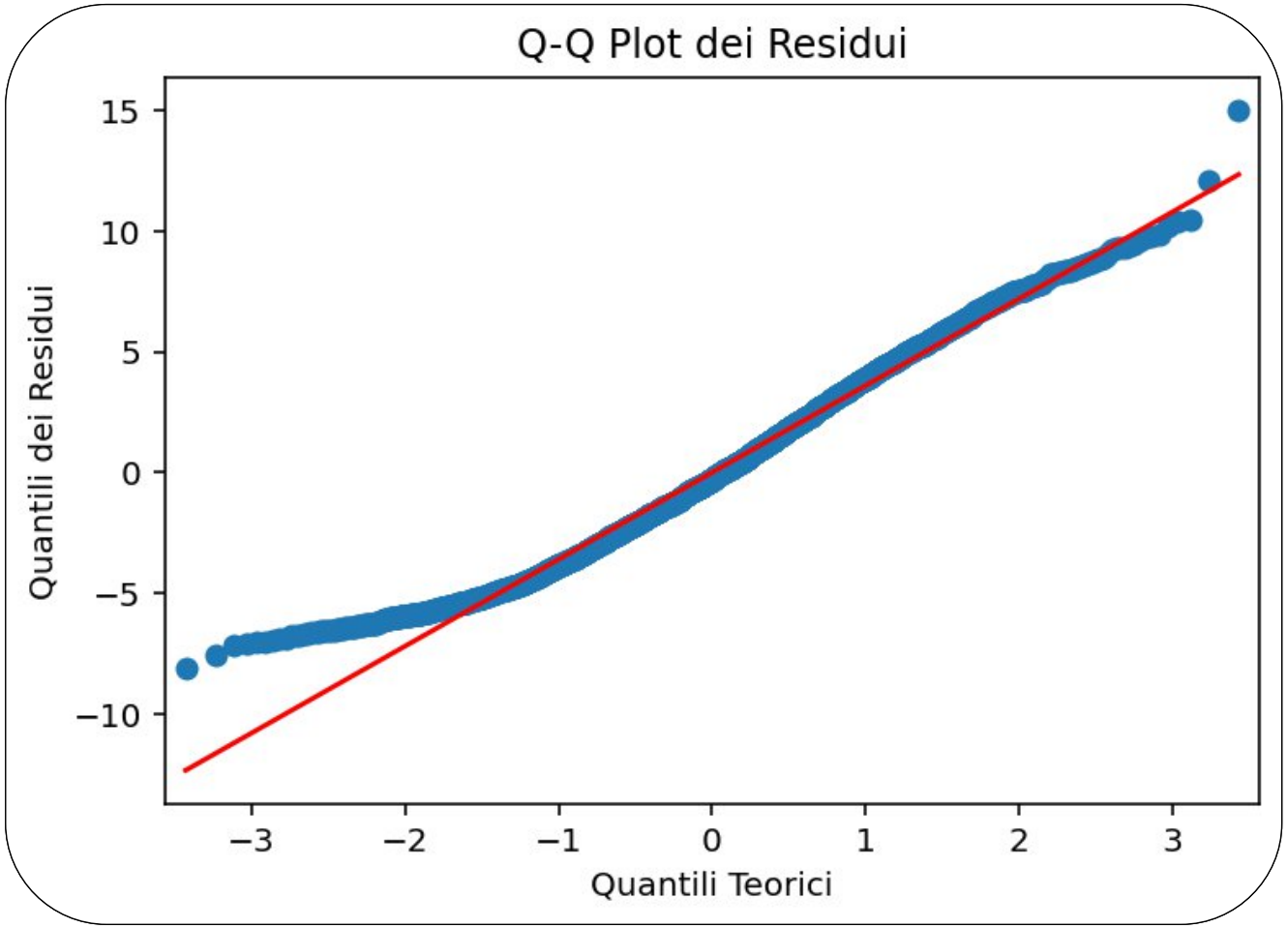
$$Y = \mathbf{1.11}X + \mathbf{6.99}$$

Equazione retta di regressione

r^2: 0.8450

MSE: 12.9536

Metriche di valutazione



Regressione lineare

Analisi di normalità dei residui

p-value: 4.2677e-19

Test di Shapiro-Wilk

media: 1.1142e-16

Media dei residui

$$\varepsilon_i = y_i - (\beta_1 x_i + \beta_0)$$