SMARCO

SMARt COmmunities Skills Development in Europe

Artificial Intelligence

Dr. Assoc. Prof. Sotiris Kotsiantis

HELLENIC OPEN UNIVERSITY

Co-funded by the European Union

smarco

Project 101186291 — SMARCO
Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

SMARCO

SMARt COmmunities Skills Development in Europe

Unit 2 – Regression and Forecasting Methods

Co-funded by the European Union

smarco

Project 101186291 — SMARCO
Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

Unit 2 - Aim and objectives

- This unit introduces trainees to machine learning regression and recommendation methods. Trainees will also become familiar to handle regression problems related to smart cities. Trainees will also become familiar to handle forecasting problems related to smart cities.

Co-funded by the European Union

smarco

Project 101186291 — SMARCO
Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

Unit 2 - Learning outcomes

- Describe the basic machine learning regression techniques.
- Demonstrate a learning algorithm in a smart city related regression problem.
- Demonstrate a forecasting algorithm in a time-series smart city related problem.

Co-funded by the European Union

smarco

Project 101186291 — SMARCO
Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

Table of contents

Co-funded by the European Union

smarco

Project 101186291 — SMARCO
Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

Terms and keywords

- Machine Learning – Supervised machine learning
- Regression algorithm, regressor
- Artificial Neural Network, Linear Regression, Random Forest
- Time-Series Forecasting
- Forecasting algorithms

Co-funded by the European Union

smarco

Project 101186291 — SMARCO
Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

Regression

- The main difference between regression and classification is that regression is used to predict continuous values, while classification is used to predict discrete values.
- For example, regression can be used to predict a person's age, while classification can be used to predict whether a person is male or female.

Co-funded by the European Union

smarco

Project 101186291 — SMARCO
Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

An example of smart city regression problem

- A smart city regression problem could involve predicting the amount of energy that will be used in a particular smart city over a given period of time.
- The problem could involve using data from weather reports, population and economic indicators, historical energy usage data, and other relevant factors to develop a predictive regression model that will provide more accurate estimations of energy consumption.
- This model could be used by smart city planners to better allocate and manage energy resources.

Co-funded by the European Union

smarco

Project 101186291 — SMARCO
Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

Application of a regression algorithm

- The application of a regression algorithm to a dataset begins with understanding the data and the type of regression algorithm that is best suitable for the problem.
- After selecting the algorithm and preprocessing the data, the next step is to train the model using a training dataset.
- After training the model, it is important to analyze the performance of the regression model using various evaluation metrics.
- Finally, the trained model can be used on the test dataset to make predictions.

Co-funded by the European Union

smarco

Project 101186291 — SMARCO
Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

Regression algorithms

- K-Nearest Neighbors
- Linear Regression
- Support Vector Machines
- Artificial Neural Networks
- Model Trees
- Random Forests
- Gradient Boosted Trees

Co-funded by the European Union

smarco

Project 101186291 — SMARCO
Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

Well known evaluation metrics (1)

- Mean absolute error (MAE): It is the average of the absolute errors between the predicted values and the actual values.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

n is the number of observations, y_i is the actual value, and $\hat{y}_i$ is the predicted value.

- Mean square error (MSE): The MSE is the sum of the squared differences between the estimated values and the true values.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

- Root Mean Squared Error (RMSE): RMSE is calculated by taking the square root of the mean square error (MSE).

Co-funded by the European Union

smarco

Project 101186291 — SMARCO
Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

Well known evaluation metrics (2)

- **R-squared (R^2):**

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

- **Root Mean Squared Logarithmic Error (RMSLE):**

$$RMSLE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\log(1 + y_i) - \log(1 + \hat{y}_i))^2}$$

- **Mean Absolute Percentage Error (MAPE):**

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100$$

Co-funded by the European Union

smarco

Project 101186291 — SMARCO
Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

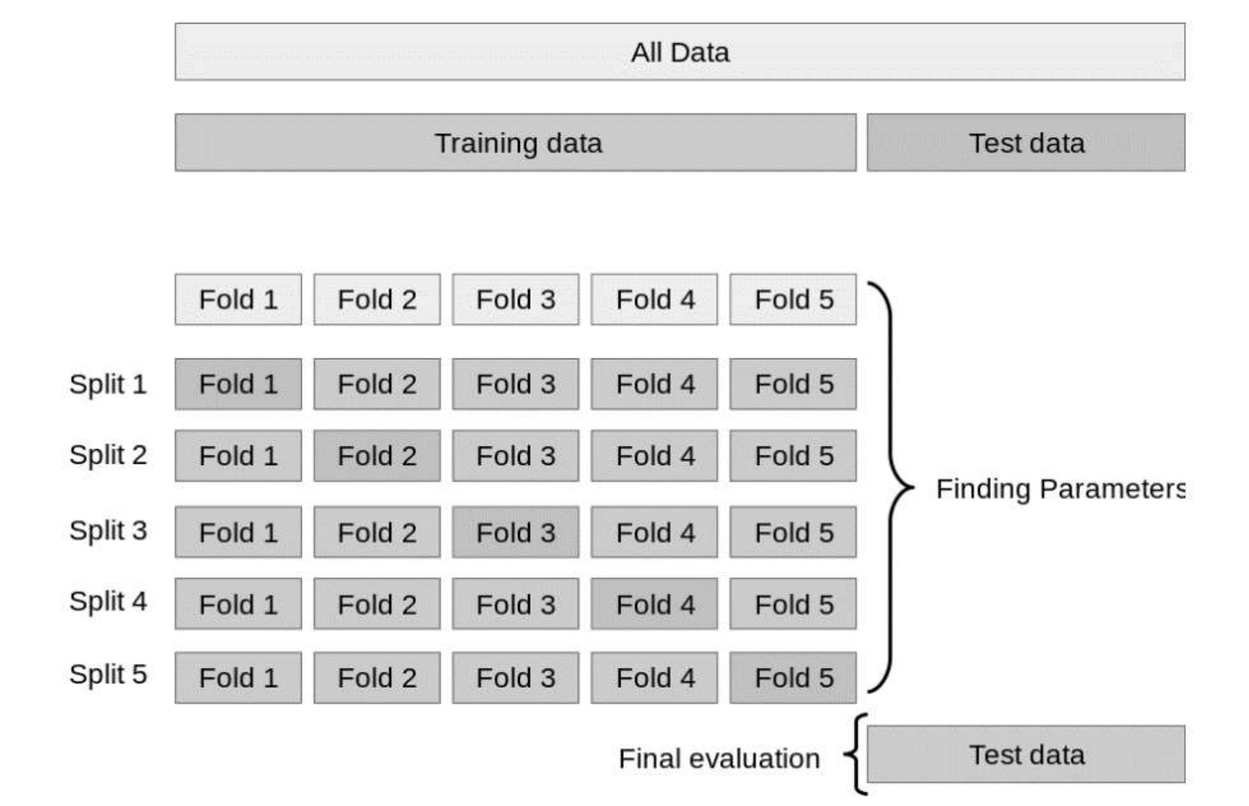
Cross-validation

Cross-validation is a technique used to train and machine learning models on different subsets of data.

It involves splitting the dataset into a number of **folds** and using each fold to train and test the model.

The number of folds is typically set to five, although this can vary depending on the size of the dataset.

Each fold is used to test the **model's performance** a single time, and the overall results are averaged out in order to obtain an overall assessment of the model's generalization ability.



Co-funded by the European Union

smarco

Project 101186291 — SMARCO
Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

Waste management cost prediction - A case study

- A regression algorithm can be used to predict waste management cost in a smart city.
- The model can be used to predict waste management costs by taking into account various inputs such as waste generation rate, transport cost, and landfill capacity etc.
- It is also a powerful tool to analyze and understand how changes in these parameters will affect the waste management costs.
- The regression model can also be used to identify the optimal strategies that can reduce the overall cost of waste management.

Co-funded by the European Union

smarco

Project 101186291 — SMARCO
Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

Run Jupyter Online with Colab

- Google Colab is a cloud-based notebook environment that excels in collaborative work, data analysis, and machine learning tasks.
- Colab comes with many pre-installed Python libraries commonly used in data science and machine learning, such as NumPy, pandas, matplotlib. This saves time and effort in setting up the environment.
- Google Colab provides free access to Graphics Processing Units (GPUs) and Tensor Processing Units (TPUs). This is particularly advantageous for training machine learning models that require significant computational power.
- You can write and execute python code, save and share your analyses, and access powerful computing resources, all for free from your browser.
- To start working with Colab you first need to log in to your google/gmail account, then go to this link https://colab.research.google.com

Co-funded by the European Union

smarco

Project 101186291 — SMARCO
Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

Waste management cost prediction

- Training a learning algorithm in Google Colab

```
!pip install git+https://github.com/pycaret/pycaret.git
import pandas as pd
#the open data can be found at https://www.kaggle.com/datasets/shashwatwork/municipal-waste-management-cost-prediction
!wget --no-check-certificate https://thalis.math.upatras.gr/~sotos/public_data_waste_fee.csv
data = pd.read_csv('public_data_waste_fee.csv')
data
```

- Press Shift+Enter to execute the cell in colab cell

Co-funded by the European Union

smarco

Project 101186291 — SMARCO
Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

Description of the dataset

```
import numpy as np
data = data[data['finance'].notna()]#remove missing values
data=data.drop(columns=['region', 'province','name','gdp','msw_so','msw_un','tc'])#remove not useful independent variables
data
```

	region	province	name	tc	cres	csor	istat	area	pop	alt	...	msw	sor	geo	roads	s_wteregio	s_landfill	gdp	proads	wage	finance
0	Emilia_Romagna	Ferrara	Comacchio	502.250000	129.270004	66.419998	38006	283.750000	22648	1.0	...	33956781	60.07	3.0	285.0	33.105049	15.233530	7.269942	4.354818	9.438692	7.488041
1	Emilia_Romagna	Ferrara	Lagosanto	228.050003	49.520000	44.139999	38011	34.439999	4952	1.0	...	2411867	75.93	3.0	11.0	33.105049	15.233530	7.109984	6.082588	9.510139	7.323284
2	Emilia_Romagna	Ferrara	Goro	268.010010	50.610001	44.599998	38025	26.620001	3895	1.0	...	2159322	78.49	3.0	49.0	33.105049	15.233530	7.267856	4.335555	8.891356	7.485891
3	Emilia_Romagna	Ferrara	Mesola	199.089996	41.110001	40.439999	38014	84.300003	7140	1.0	...	3651915	78.89	3.0	165.0	33.105049	15.233530	7.085936	3.710479	9.433685	7.298514
4	Puglia	Barletta-Andria-Trani	Margherita di Savoia	233.639999	58.270000	25.950001	110005	35.700001	12193	1.0	...	7195880	42.06	1.0	60.0	4.046452	45.411903	7.247444	5.274037	9.125561	7.464867
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
4330	Lombardia	Bergamo	Foppolo	977.419983	469.290009	57.480000	16103	16.139999	202	1508.0	...	188003	38.76	3.0	20.0	38.501492	4.551430	10.538720	2.224623	9.505127	10.854882
4331	Trentino_Alto_Adige	Bolzano	Curon Venosta/Graun im Vinschgau	132.809998	32.380001	71.739998	21027	209.649994	2423	1520.0	...	1466016	70.63	3.0	52.0	18.539640	11.318043	8.413345	3.819020	9.599592	8.665746
4332	Trentino_Alto_Adige	Bolzano	Selva di Val Gardena/Wolkenstein in Gröden	156.429993	62.910000	63.680000	21089	56.240002	2660	1563.0	...	3405016	64.84	3.0	40.0	18.539640	11.318043	8.609627	4.187419	10.174608	8.867916
4333	Trentino_Alto_Adige	Bolzano	Corvara in Badia/Corvara	370.880005	89.139999	260.179993	21026	38.840000	1320	1568.0	...	2527097	75.21	3.0	66.0	18.539640	11.318043	8.750687	3.039358	10.249179	9.013207
4334	Lombardia	Sondrio	Livigno	319.989990	219.429993	67.480003	14037	227.289993	5976	1816.0	...	7683920	44.89	3.0	76.0	38.501492	4.551430	8.463772	4.451475	9.763460	8.717685

3955 rows × 39 columns

Co-funded by the European Union

smarco

Project 101186291 — SMARCO
Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

Columns description

- name: Name of municipality
- tc: Cost per capita eur
- cres: residual cost per capita
- csor: Sorted cost per caputa
- Istat: National code
- area: km2
- pop: population
- alt: altitude m.s.l.
- isle: dummy municipality on isle
- sea: dummy coastal municipality
- pden: population density (people per km2)
- wden: waste per km2
- urb: urbanization index (1 low, 3 high)
- fee: fee scheme
- d_fee: dmmy PAYT
- sample: Reg with PAYT
- organic: organic %
- paper: paper%
- glass: glass %
- wood: wood %
- metal: metal %
- plastic: plastic %
- raee: raee %
- texile: texile %
- other: other %
- msw_so: msw sorted kg
- msw_un: msw unsorted kg
- msw : Municipal solid waste kg
- sor: Share of sorther waste
- geo: 1 South, 2 Center, 3 North
- roads: Km of roads within the municipality
- s_wteregio: Share of sw sent to W2E plants - regional figure
- s_landfill: share of waste to landifil
- gdp: Municipal revenues EUR (p) - log
- proads: People per km of roads (log)
- wage: Taxable income EUR (p) - log
- finance: Municipal revenues EUR (p) - log

Co-funded by the European Union

smarco

Project 101186291 — SMARCO
Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

Estimating the accuracy

- Application of random forest in waste management cost prediction

```
# Import necessary functions from PyCaret's regression module
from pycaret.regression import *
# Set up the regression experiment using PyCaret's setup function
# 'data' is the DataFrame containing the dataset, and 'finance' is the target variable
# 'train_size=0.9' specifies that 90% of the data will be used for training
clf1 = setup(data, target='finance', train_size=0.9)
# Create a Random Forest regression model using PyCaret's create_model function
# 'rf' is the identifier for the Random Forest model
rf = create_model('rf')
```

Fold	MAE	MSE	RMSE	R2	RMSLE	MAPE
0	0.2571	0.1104	0.3322	0.6469	0.0387	0.0343
1	0.2685	0.1358	0.3684	0.6252	0.0416	0.0354
2	0.2667	0.1282	0.3580	0.5984	0.0409	0.0350
3	0.2718	0.1384	0.3720	0.6175	0.0422	0.0355
4	0.2624	0.1214	0.3484	0.6592	0.0394	0.0343
5	0.2721	0.1244	0.3527	0.5650	0.0407	0.0362
6	0.2489	0.1026	0.3203	0.6302	0.0372	0.0332
7	0.2536	0.1108	0.3329	0.6167	0.0383	0.0335
8	0.2532	0.1069	0.3270	0.6740	0.0375	0.0333
9	0.2478	0.1007	0.3174	0.6517	0.0370	0.0331
Mean	0.2602	0.1180	0.3430	0.6285	0.0393	0.0344
Std	0.0088	0.0129	0.0187	0.0302	0.0018	0.0011

Co-funded by the European Union

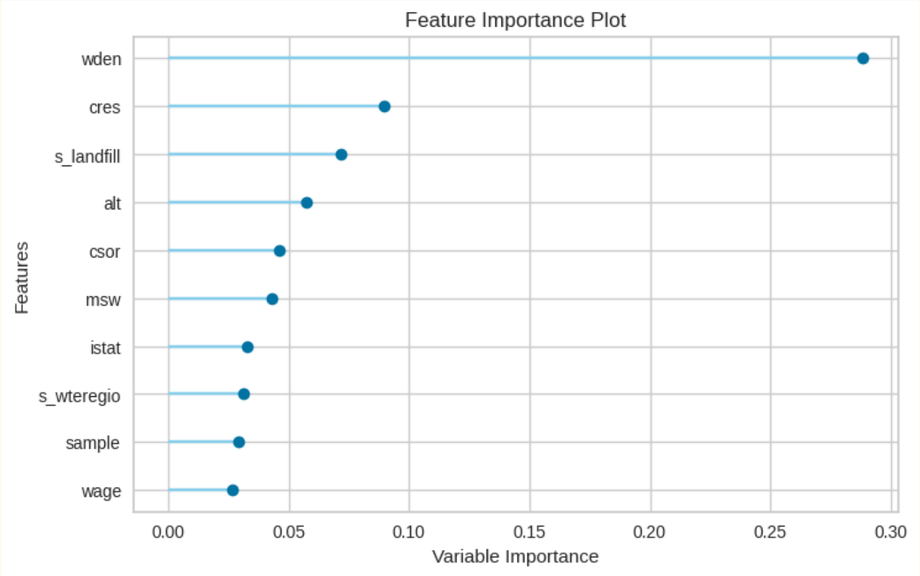
smarco

Project 101186291 — SMARCO
Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

Most important features (dependent variables)

- Some of the features for waste management cost prediction are more important than the others meaning that they have the biggest influence in a prediction model

```
# Plot feature importance for the Random Forest model
# 'plot_model' is a PyCaret function for plotting various aspects of a trained model
# 'plot='feature'' specifies that we want to plot feature importance
plot_model(rf, plot='feature')
```



Co-funded by the European Union

smarco

Project 101186291 — SMARCO
Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

The predictions in the test set

```
predict_model(rf)
```

	Model	MAE	MSE	RMSE	R2	RMSLE	MAPE
0	Random Forest Regressor	0.2501	0.1064	0.3262	0.6868	0.0373	0.0330

	cres	csor	istat	area	pop	alt	isle	sea	pden	wden	...	msw	sor	geo	roads	s_wteregio	s_landfill	proads	wage	finance	prediction_label
336	37.730000	54.439999	37038	43.070000	8674	16.0	0.0	0.0	201.393082	1.047313e+05	...	4510776	74.419998	3.0	103.0	33.105049	15.233530	4.448210	9.785361	7.072884	7.124727
2280	77.849998	6.840000	108024	10.280000	24527	260.0	0.0	0.0	2385.895020	1.008422e+06	...	10366578	80.080002	3.0	97.0	38.501492	4.551430	5.584011	9.688993	7.013426	7.168706
3098	39.700001	53.540001	96004	46.689999	43818	420.0	0.0	0.0	938.487915	4.992527e+05	...	23310108	78.470001	3.0	233.0	24.467649	23.119114	5.253231	9.764866	7.787671	7.425216
1630	14.030000	30.799999	16157	10.790000	5773	157.0	0.0	0.0	535.032410	2.276605e+05	...	2456457	89.239998	3.0	44.0	38.501492	4.551430	4.887790	9.490038	6.796226	6.882728
2909	101.870003	53.560001	53026	174.559998	3596	379.0	0.0	0.0	20.600367	7.722422e+03	...	1348026	64.059998	2.0	54.0	12.075790	31.493038	4.114326	9.233202	7.811202	7.774737
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
2509	20.670000	41.959999	65149	21.030001	2185	295.0	0.0	1.0	103.899185	2.502901e+04	...	526360	67.410004	1.0	12.0	27.876980	3.602713	5.156178	9.038669	8.918715	8.311087
2547	54.970001	39.740002	68026	13.760000	1800	301.0	0.0	0.0	130.813950	2.745727e+04	...	377812	75.519997	1.0	216.0	0.000000	37.241680	2.123026	9.175899	7.144986	7.713515
4222	27.090000	44.740002	91024	112.269997	4062	1000.0	0.0	0.0	36.180637	1.268684e+04	...	1424352	82.949997	1.0	125.0	8.904969	39.524139	3.443352	9.075613	7.467774	7.876756
88	39.549999	50.790001	30120	28.360001	2881	5.0	0.0	1.0	101.586739	4.910141e+04	...	1392516	65.639999	3.0	21.0	17.467550	4.897196	4.908092	9.575512	7.137276	7.331790
3476	50.169998	21.309999	14068	11.150000	624	526.0	0.0	0.0	55.964127	1.612081e+04	...	179747	48.900002	3.0	8.0	38.501492	4.551430	4.367864	9.603958	7.145706	7.419138

Co-funded by the European Union

smarco

Project 101186291 — SMARCO
Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

You can run the full example code

https://colab.research.google.com/drive/1U_a3KuUxTLx3u_TKh-Ee8rZaZXR-CAZO?usp=sharing

Co-funded by the European Union

Project 101186291 — SMARCO
Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

Other applications of regression in a smart city

- Traffic Flow Optimization: Regression algorithms can be used to analyze real-time traffic patterns and adjust lights and signal timings accordingly.
- Energy Demand Forecasting: Regression models can be used to forecast energy demand so that utilities can plan their energy output accordingly.
- Pollution Analysis: Regression algorithms can be used to forecast air and water quality in areas of a smart city in order to identify pollution sources and their implications for public health.

Co-funded by the European Union

smarco

Project 101186291 — SMARCO
Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

Time series

- Smart cities are increasingly relying on data-driven analytics to improve the quality of life for their citizens.
- One such time series problem could be predicting the demand for mobility services like ride-sharing.
- This problem involves forecasting the number of times people will utilize ride-sharing services in the future, such as during peak travel seasons, in different areas of the city, and for different purposes.
- This problem would involve looking at historical travel data, such as the number of people using ride-sharing apps for the past few years, in order to make predictions about future demand.
- Additionally, it would require analyzing the impact of external factors such as special events, weather patterns, socioeconomic phenomena, and other similar variables that could affect travel demand.

Co-funded by the European Union

smarco

Project 101186291 — SMARCO
Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

Time series methods

- The most widely used time series algorithms for smart cities include the Autoregressive Integrated Moving Average (ARIMA) model and support vector machine (SVM).
- ARIMA is a predictive model that uses past data points to forecast future values. It is useful for making predictions about urban dynamics such as traffic, energy supply and demand, air pollution levels, etc.
- SVM is an algorithm used for regression problems, which can be used to predict urban dynamics such as crime, disease spread, transport demands, etc.

Co-funded by the European Union

smarco

Project 101186291 — SMARCO
Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

Forecasting air-quality

- The first step is to gather data on current conditions such as air pollution levels, population density, weather information, traffic patterns, and other relevant factors.
- This data can be used to develop predictive models that will allow cities to forecast air quality levels in the future.
- It is important to consider factors such as the seasonal variation in air quality along with emission sources such as factories and vehicles to ensure accurate forecast.

Co-funded by the European Union

smarco

Project 101186291 — SMARCO
Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

An example

- Simple example for pollution Forecasting (Open data: https://www.kaggle.com/datasets/fedesoriano/air-quality-data-in-india)

```
# Import the pandas library and alias it as 'pd'
import pandas as pd
# Read the CSV file from the provided URL and set the first column as the index
# Parse dates to ensure that date columns are recognized as datetime objects
!wget --no-check-certificate https://thalis.math.upatras.gr/~sotos/air-quality-india.csv
data = pd.read_csv('air-quality-india.csv', index_col=0, parse_dates=True)
# Select the 'PM2.5' column from the dataset
data = data[['PM2.5']]
# Resample the data to have a frequency of one hour ('H')
# This ensures that the time series has a consistent hourly frequency
data = data.asfreq('h')
# Display the resulting DataFrame
data
```

Timestamp	PM2.5
2017-11-07 12:00:00	64.51
2017-11-07 13:00:00	69.95
2017-11-07 14:00:00	92.79
2017-11-07 15:00:00	109.66
2017-11-07 16:00:00	116.50
...	...
2022-06-04 11:00:00	35.89
2022-06-04 12:00:00	33.83
2022-06-04 13:00:00	33.05
2022-06-04 14:00:00	35.29
2022-06-04 15:00:00	40.67

40084 rows × 1 columns

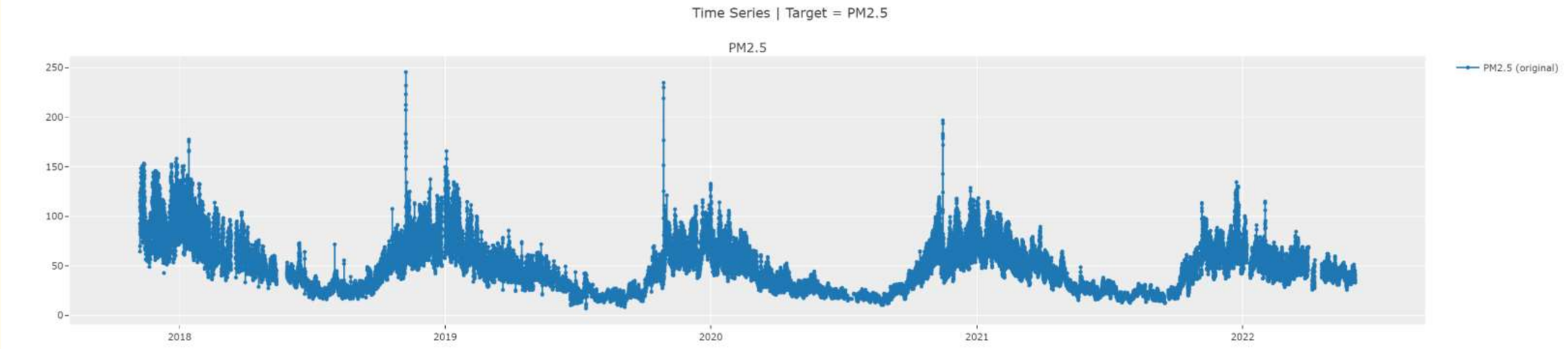
Co-funded by the European Union

smarco

Project 101186291 — SMARCO
Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

Dataset visualization

```
plot_model(plot = 'ts')
```



Co-funded by the European Union

Project 101186291 — SMARCO
Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

Train Arima on the dataset

- Evaluation with 3-cross validation

```
# Install the PyCaret library using pip
!pip install git+https://github.com/pycaret/pycaret.git
!pip install --upgrade sktime
# Import necessary functions from PyCaret's time_series module
from pycaret.time_series import *
# Import the plot_series function from sktime.utils.plotting
from sktime.utils.plotting import plot_series
# Set up the time series experiment using PyCaret's setup function
# 'data' is the time series DataFrame, 'target' is the target variable ('PM2.5')
# 'fh=7' specifies the forecasting horizon (number of future observations to predict)
# 'fold=3' specifies the number of folds for time series cross-validation
# 'numeric_imputation_target='ffill'' specifies forward fill as the numeric imputation strategy
exp_name = setup(data=data, target='PM2.5', fh=7, fold=3, numeric_imputation_target='ffill')
# Create an ARIMA (AutoRegressive Integrated Moving Average) time series model using PyCaret's create_model function
arima = create_model('arima')
```

	cutoff	MASE	RMSSE	MAE	RMSE	MAPE	SMAPE	R2
0	2022-06-03 10:00	0.3766	0.2903	2.1101	2.7709	0.0508	0.0488	0.6682
1	2022-06-03 17:00	0.4050	0.2715	2.2690	2.5913	0.0575	0.0554	0.0512
2	2022-06-04 00:00	0.4259	0.2814	2.3861	2.6853	0.0562	0.0573	0.0343
Mean	NaT	0.4025	0.2810	2.2551	2.6825	0.0548	0.0539	0.2512
SD	NaT	0.0202	0.0077	0.1131	0.0734	0.0029	0.0037	0.2949

$$MASE = \frac{\frac{1}{h}\sum_{t=1}^{h}|y_t - \hat{y}_t|}{\frac{1}{n-1}\sum_{i=2}^{n}|y_i - y_{i-1}|}$$

h is the forecast horizon.

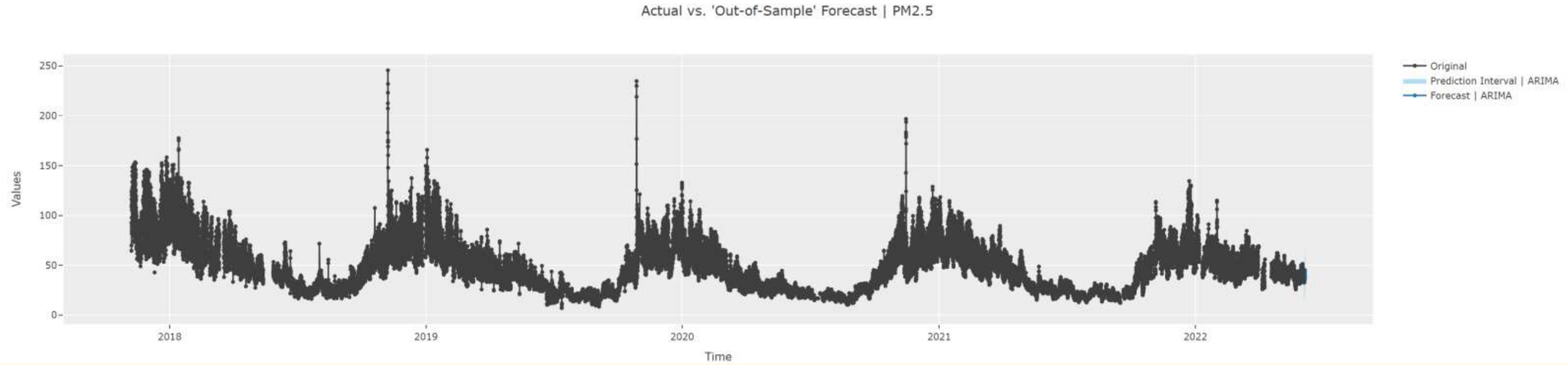
Co-funded by the European Union

smarco

Project 101186291 — SMARCO
Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

Forecasting in the future

```
plot_model(estimator = arima, plot = 'forecast', data_kwargs = {'fh' : 48})
```



Co-funded by the European Union

Project 101186291 — SMARCO
Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

You can run the full example code

https://colab.research.google.com/drive/1C7X9XIkToFLL1Bbr9PPPuHEHnyXoQIMa?usp=sharing

You can save a copy in your drive to execute the code from your PC via colab.

Co-funded by the European Union

smarco

Project 101186291 — SMARCO
Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

Other time series problems in a smart city

- Traffic Congestion
- Parking Availability
- Water Usage
- Energy Consumption
- Crime Activity

Co-funded by the European Union

smarco

Project 101186291 — SMARCO
Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

List of references

- Banga, A., Ahuja, R., & Sharma, S. C. (2021). Performance analysis of regression algorithms and feature selection techniques to predict PM2.5 in smart cities. International Journal of System Assurance Engineering and Management, 14(S3), 732-745. https://doi.org/10.1007/s13198-020-01049-9
- Yu, Z., Zheng, X., Huang, F., Guo, W., Sun, L., & Yu, Z. (2020). A framework based on sparse representation model for time series prediction in smart city. Frontiers of Computer Science, 15(1). https://doi.org/10.1007/s11704-019-8395-7
- Alfarraj, O. (2021). Regression learning assisted efficient energy harvesting method for smart city environment. Sustainable Energy Technologies and Assessments, 44, 101003. https://doi.org/10.1016/j.seta.2021.101003

Co-funded by the European Union

smarco

Project 101186291 — SMARCO
Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

Further reading

- Banga, A., Ahuja, R., & Sharma, S. C. (2022). Stacking Regression Algorithms to Predict PM2.5 in the Smart City Using Internet of Things. Recent Advances in Computer Science and Communications, 15(1). https://doi.org/10.2174/2666255813999200628094351
- Carrera, B., Peyrard, S., & Kim, K. (2021). Meta-regression framework for energy consumption prediction in a smart city: A case study of Songdo in South Korea. Sustainable Cities and Society, 72, 103025. https://doi.org/10.1016/j.scs.2021.103025
- Belhadi, A., Djenouri, Y., Norvag, K., Ramampiaro, H., Masseglia, F., & Lin, J. C.-W. (2020). Space-time series clustering: Algorithms, taxonomy, and case study on urban smart cities. Engineering Applications of Artificial Intelligence, 95, 103857. https://doi.org/10.1016/j.engappai.2020.103857

Co-funded by the European Union

smarco

Project 101186291 — SMARCO
Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

Unit completed - What's next?

- To consolidate your learning and reflect on the key concepts covered, please take a moment to complete this quiz.
- Your feedback and results will help you track your progress and support continuous improvement of the training experience.
- By completing this quiz, you will also become eligible to receive a certificate of successful training completion.
- Click this link to begin the quiz!

Co-funded by the European Union

smarco

Project 101186291 — SMARCO
Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

SMARCO

SMARt COmmunities Skills Development in Europe

www.smarco.eu

info@smarco.eu

We are social! Follow us on:

@smarcoproject

@smarcoproject

@smarcoproject

Co-funded by the European Union

smarco

Project 101186291 — SMARCO
Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.