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 1 – Machine Learning Classification 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.

Module Aim and objectives

- This module targets the ability to identify, select and apply AI solutions in a smart city context. It includes being able to (i) understand fundamental concepts, practical applications, and ethical considerations of AI in smart cities, (ii) distinguish different types of Machine Learning (ML), (iii) describe basic ML approaches, techniques and algorithms, and (iv) provide real-world examples of the use of AI in 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.

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.

Unit 1 - Aim and objectives

- This unit introduces trainees to artificial intelligence and machine learning classification methods. Trainees will also become familiar to handle classification 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 1 - Learning outcomes

- Explain the concepts of artificial intelligence and machine learning
- List machine learning categories
- Describe the basic machine learning classification techniques.
- Demonstrate a learning algorithm in a smart city related classification 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

- Artificial Intelligence
- Machine Learning – Supervised machine learning
- Classification algorithms, classifiers
- Artificial Neural Network, Decision Tree, Random Forest

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.

Artificial Intelligence (AI) usage

- Governments can use AI to analyze open data, allowing for more informed public policy decisions. For example, analysis of open data on traffic, health, or education can reveal trends and inform policy to address societal needs.
- AI can assist in predicting housing demands, identifying crime hotspots, and even enhancing public health monitoring through open epidemiological data.
- Open data allows AI applications to be transparent in terms of the data sources they rely on. This transparency builds trust in AI system.

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.

Concepts and terminology

- Artificial intelligence (AI) and machine learning (ML) are related terms that deal with how computers can be programmed to make decisions and to think like humans.
- AI is a broader term that refers to systems that can perform tasks and exhibit behavior that appears to be intelligent. This generally involves using data and algorithms to enable a computer to think and act in a way that is akin to a human — this can include using natural language processing and decision-making.

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.

Machine Learning

- ML is a subset of AI and is concerned with how a computer system can learn from data. A computer can be trained to understand data and identify patterns in that data. It can use this understanding to make decisions and predictions and take action accordingly.
- Machine learning categories:
  - Supervised Machine Learning
  - Unsupervised Machine Learning
  - Reinforcement Learning

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.

Supervised Machine Learning

- Supervised learning is a type of machine learning algorithm that uses a set of labeled training data to make predictions.
- The algorithm uses a mapping function to make predictions based on new data that falls within the range of the labeled data.
- Examples of supervised learning include classification and regression.

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.

Classification

- In a classification problem, the independent variables are the features used to make predictions while the dependent variable is the label of the class or the decision to be made.
- The independent variables are used to calculate the classification boundary, while the dependent variable determines the category that the data point belongs to.

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 classification problem

- A smart city classification problem could involve using available data to classify a city's inhabitants into different groups according to their income, population size, or access to resources.
- The goal of this problem would be to gain insights into how to create more equitable and sustainable cities by understanding the factors that affect a city's citizens.
- Data used in this classification problem could include census data, school enrollment records, economic indicators, and surveys about housing, transportation, health, and safety.

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 classification algorithm

- The application of a classification algorithm to a dataset begins with understanding the data and the type of classification 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 classification 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.

Classification algorithms

- K-Nearest Neighbors
- Logistic Regression
- Support Vector Machines
- Artificial Neural Networks
- Decision Trees
- Naive Bayes
- Random Forests
- Gradient Boosted Trees

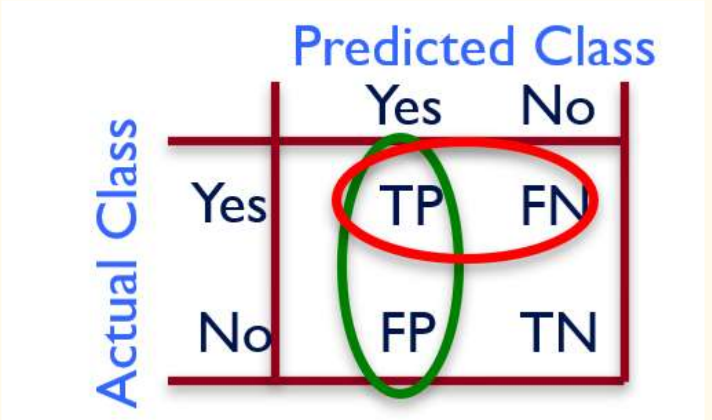
Co-funded by the European Union

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.

Supervised Machine Learning

- A confusion matrix is a table used to evaluate the performance of a classification algorithm. It is a table of confusion that helps to understand where the classifier is making the correct predictions and where it is going wrong. It shows the actual class against the predicted class.
- Each cell in the table shows the number of instances that fall into a given combination of actual and predicted classes. For example, the upper-left cell in the confusion matrix represents the number of instances classified as the first class and were actually in the first class.



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)

- **Accuracy**: It is the number of correct predictions divided by the total number of predictions. The higher number indicates better accuracy.
- **Precision**: It is the fraction of true positives which were correctly identified as such.
- **Recall**: It is the fraction of all the relevant cases of a class that were identified correctly.
- **F Score**: It is the harmonic mean of precision and recall.

$$\text{accuracy} = \frac{TP + TN}{P + N}$$

$$\text{precision} = \frac{TP}{TP + FP}$$

$$\text{recall} = \frac{TP}{TP + FN}$$

$$F_\beta = (1 + \beta^2) \cdot \frac{\text{precision} \cdot \text{recall}}{\beta^2 \cdot \text{precision} + \text{recall}}$$

$$F_{\beta=1} = \frac{2PR}{P + R}$$

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)

MCC is often preferred when dealing with imbalanced datasets due to its focus on all four confusion matrix elements.

$$\text{MCC} = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP) \cdot (TP + FN) \cdot (TN + FP) \cdot (TN + FN)}}$$

Cohen's kappa has similar interpretation as MCC, with +1 indicating perfect agreement and 0 indicating agreement due to chance.

$$\kappa = \frac{2 \cdot (TP \cdot TN - FP \cdot FN)}{(TP + FP) \cdot (FP + TN) + (TP + FN) \cdot (FN + TN)}$$

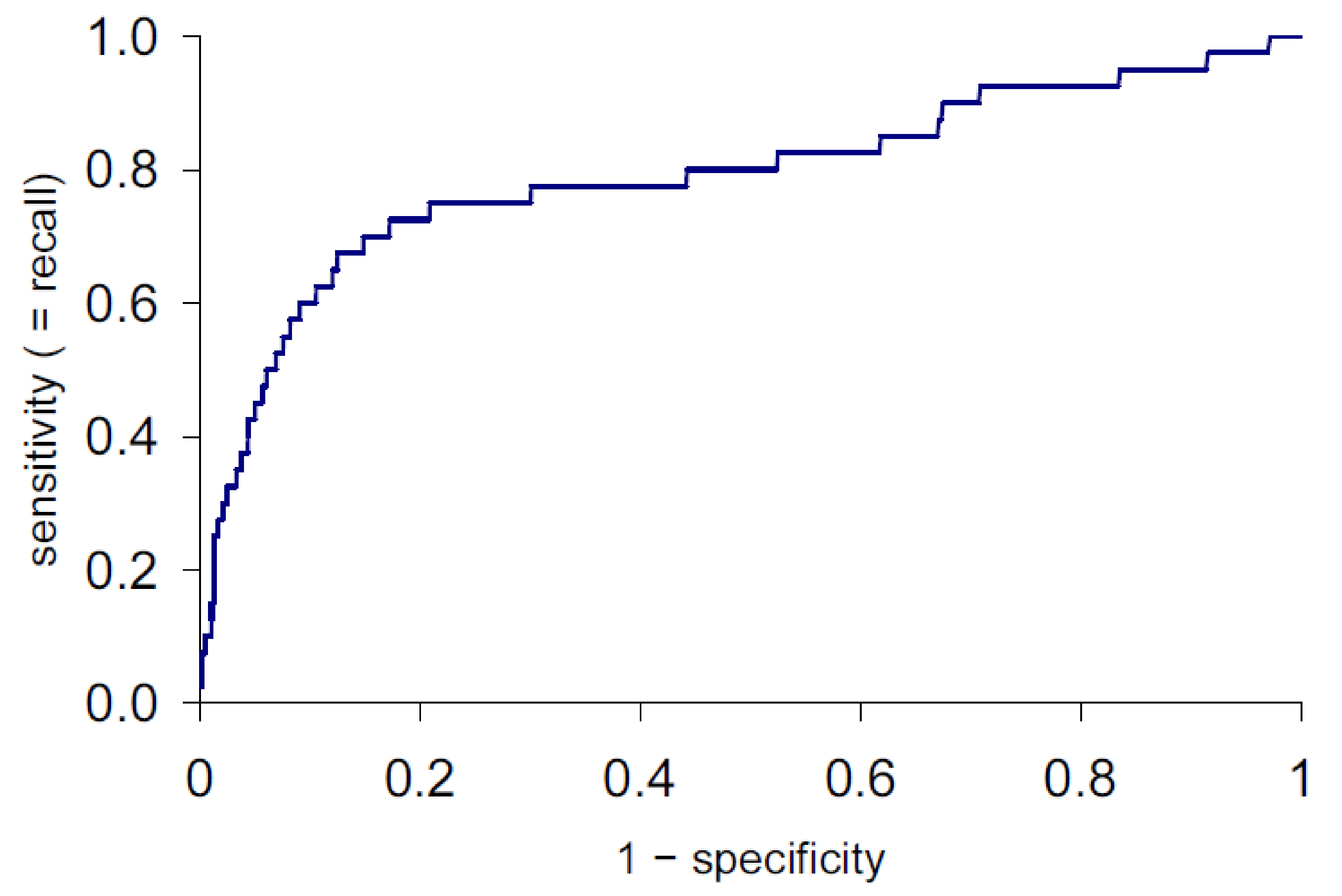
Co-funded by the European Union

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 (3)

AUC ROC Curve: Area Under Curve (AUC) Receiver Operating Characteristic (ROC) Curve is a measure of the accuracy of a predictive model. It measures the probability that a model will correctly identify an instance from two classes.



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.

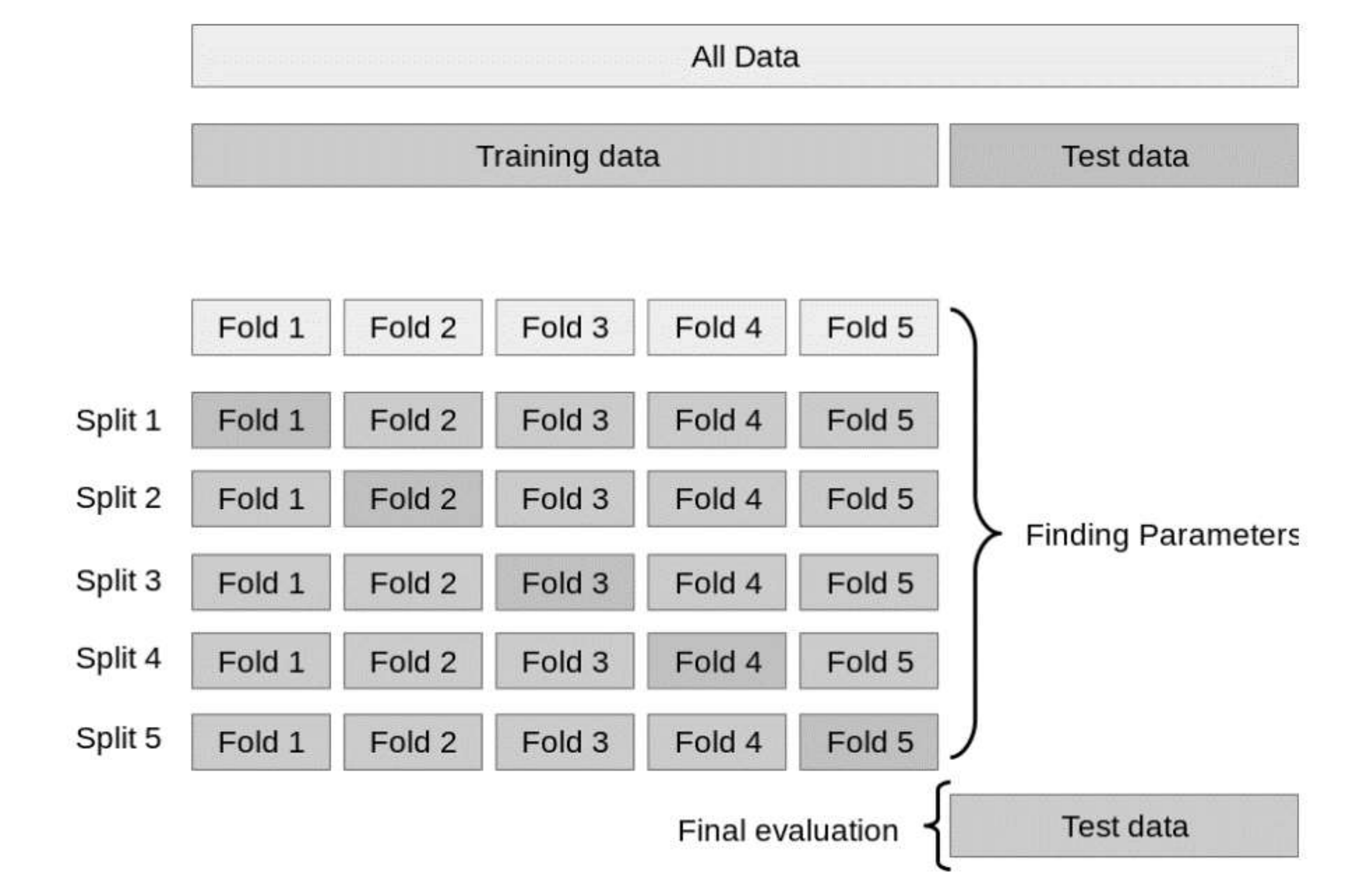
Cross-validation

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.

Water potability - A case study

- A classifier can be used to determine the potability of water in a smart city.
- The classifier would take in information about the source of the water, such as if it is from a tap, a stream, or a well, and use this information to classify the water as potable or non-potable.
- Additionally, the classifier can consider other factors, such as the source of the water, chemical composition of the water, presence of any contaminants, and the temperature of the water, to determine its potability.
- The classifier can then be used to alert authorities and citizens of potential water contamination problems so that action can be taken to address them.

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.

CSV, basic sharing

- A basic approach to share data is the comma separated value (CSV) format
  - it is a text format, accessible to all apps
  - each line (even if blank) is a row
  - in each row, each value is separated from the others by a comma (even if it is blank)
  - cannot capture complex things like formula

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.

Estimation of water potability

- Training a learning algorithm in Google Colab

```
# run this cell to install pycaret in Google Colab
!pip install git+https://github.com/pycaret/pycaret.git
import pandas as pd
#the open dataset can be found at https://www.kaggle.com/datasets/adityakadiwal/water-potability
import ssl
!wget --no-check-certificate https://thalis.math.upatras.gr/~sotos/water_potability.csv
data = pd.read_csv('water_potability.csv')
data
```

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

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

```
data.describe()
```

	ph	Hardness	Solids	Chloramines	Sulfate	Conductivity	Organic_carbon	Trihalomethanes	Turbidity	Potability
count	2785.000000	3276.000000	3276.000000	3276.000000	2495.000000	3276.000000	3276.000000	3114.000000	3276.000000	3276.000000
mean	7.080795	196.369496	22014.092526	7.122277	333.775777	426.205111	14.284970	66.396293	3.966786	0.390110
std	1.594320	32.879761	8768.570828	1.583085	41.416840	80.824064	3.308162	16.175008	0.780382	0.487849
min	0.000000	47.432000	320.942611	0.352000	129.000000	181.483754	2.200000	0.738000	1.450000	0.000000
25%	6.093092	176.850538	15666.690297	6.127421	307.699498	365.734414	12.065801	55.844536	3.439711	0.000000
50%	7.036752	196.967627	20927.833607	7.130299	333.073546	421.884968	14.218338	66.622485	3.955028	0.000000
75%	8.062066	216.667456	27332.762127	8.114887	359.950170	481.792304	16.557652	77.337473	4.500320	1.000000
max	14.000000	323.124000	61227.196008	13.127000	481.030642	753.342620	28.300000	124.000000	6.739000	1.000000

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

- **pH**: The pH level of the water.
- **Hardness**: Water hardness, a measure of mineral content.
- **Solids**: Total dissolved solids in the water.
- **Chloramines**: Chloramines concentration in the water.
- **Sulfate**: Sulfate concentration in the water.
- **Conductivity**: Electrical conductivity of the water.
- **Organic_carbon**: Organic carbon content in the water.
- **Trihalomethanes**: Trihalomethanes concentration in the water.
- **Turbidity**: Turbidity level, a measure of water clarity.
- **Potability**: Target variable; indicates water potability with values 1 (potable) and 0 (not potable).

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.

Statistic measures

- **Count:** The number of non-null values in each column.
- **Mean:** The average or arithmetic mean of the values in each column. It is calculated by summing up all values and dividing by the count.
- **Std (Standard Deviation):** A measure of the amount of variation or dispersion in a set of values. It indicates how much the values in a column deviate from the mean.
- **Min (Minimum):** The smallest value in each column.
- **25% (25th Percentile or Q1):** The value below which 25% of the data falls.
- **50% (50th Percentile or Median or Q2):** The middle value of the dataset when it is sorted in ascending order. It represents the point below which 50% of the data falls.
- **75% (75th Percentile or Q3):** The value below which 75% of the data falls.
- **Max (Maximum):** The largest value in each column.

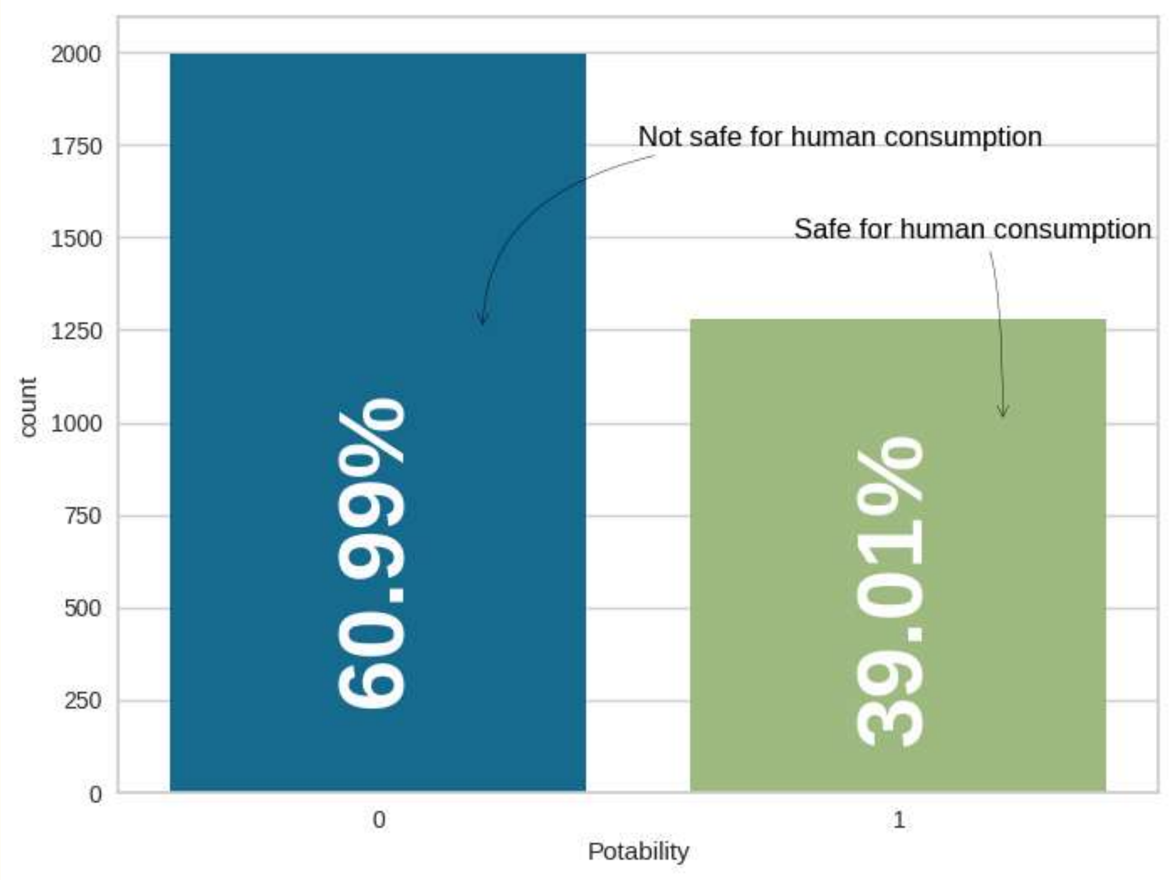
Co-funded by the European Union

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 variable that we want to estimate

```python
# Import necessary libraries
import matplotlib.pyplot as plt  # Matplotlib for basic plotting
import seaborn as sns            # Seaborn for statistical data visualization
import numpy as np                # NumPy for numerical operations
# Create a figure with a specified size
fig = plt.figure(figsize=(8, 6))
# Create a count plot using Seaborn
ax = sns.countplot(data=data, x='Potability')
# Annotate each bar in the count plot with the percentage
for i in ax.patches:
    ax.text(
        x=i.get_x() + i.get_width() / 2,       # x-coordinate for the text
        y=i.get_height() / 7,                  # y-coordinate for the text
        s=f"{np.round(i.get_height() / len(data) * 100, 2)}%",  # Text with percentage
        ha='center',                           # Horizontal alignment
        size=40,                               # Text size
        weight='bold',                         # Text weight (bold)
        rotation=90,                           # Text rotation angle
        color='white'                          # Text color
    )
# Annotate additional information about the plot using plt.annotate
# Annotate "Not safe for human consumption" with an arrow
plt.annotate(
    text="Not safe for human consumption",       # Annotation text
    xytext=(0.5, 1750),                          # Text position
    xy=(0.2, 1250),                              # Arrow position
    arrowprops=dict(arrowstyle="->", color='black', connectionstyle="angle3,angleA=0,angleB=90"),  # Arrow properties
    color='black'                               # Text color
)
# Annotate "Safe for human consumption" with an arrow
plt.annotate(
    text="Safe for human consumption",           # Annotation text
    xytext=(0.8, 1500),                          # Text position
    xy=(1.2, 1000),                              # Arrow position
    arrowprops=dict(arrowstyle="->", color='black', connectionstyle="angle3,angleA=0,angleB=90"),  # Arrow properties
    color='black'                               # Text color
)
# Show the plot
plt.show()
```



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 prediction of water potability

```
from pycaret.classification import *
from pycaret.classification import ClassificationExperiment
exp1 = ClassificationExperiment()
exp1 = setup(data, target = 'Potability',train_size=0.9) #90% of the dataset will be used for training the learning algorithm
rf = create_model('rf') #the accuracy of the learning algorithm estimated by 10-cross validation in the training set
```

Fold	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
0	0.6576	0.6489	0.2696	0.6458	0.3804	0.1957	0.2314
1	0.6678	0.6966	0.3043	0.6604	0.4167	0.2264	0.2596
2	0.6746	0.6757	0.3565	0.6508	0.4607	0.2551	0.2788
3	0.6678	0.6617	0.3043	0.6604	0.4167	0.2264	0.2596
4	0.6881	0.7074	0.3739	0.6825	0.4831	0.2862	0.3128
5	0.6610	0.6810	0.3478	0.6154	0.4444	0.2267	0.2459
6	0.6780	0.7126	0.2957	0.7083	0.4172	0.2435	0.2879
7	0.6780	0.6534	0.3130	0.6923	0.4311	0.2488	0.2869
8	0.7041	0.7285	0.4261	0.7000	0.5297	0.3320	0.3538
9	0.6259	0.6267	0.2435	0.5490	0.3373	0.1277	0.1482
Mean	0.6703	0.6793	0.3235	0.6565	0.4317	0.2368	0.2665
Std	0.0196	0.0305	0.0507	0.0447	0.0505	0.0509	0.0515

Co-funded by the European Union

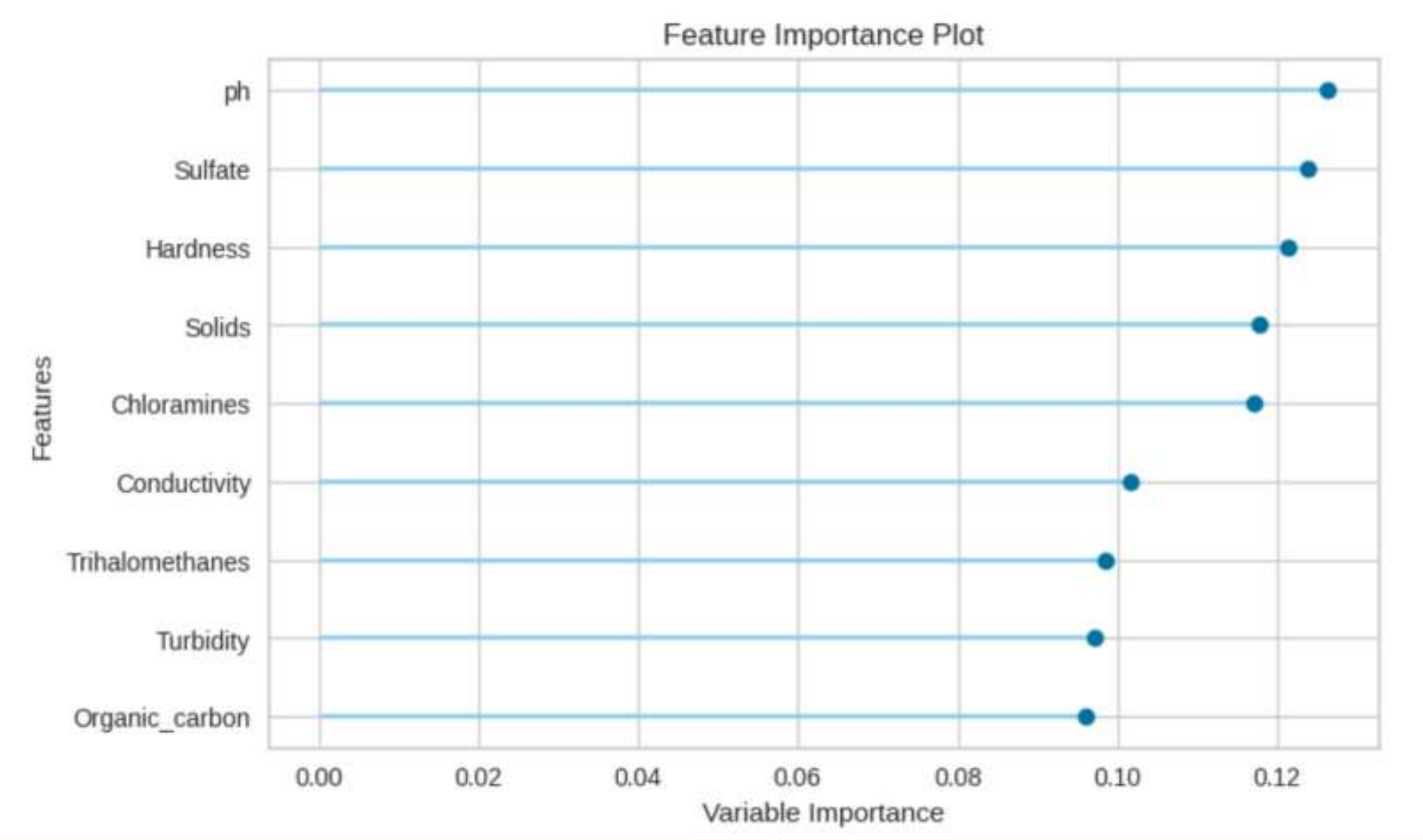
smarco

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 (the input measurements e.g., pH, turbidity, dissolved solids, etc.) for prediction of water potability are more important than the others meaning that they have the biggest influence in a prediction model

```
plot_model(rf, plot = 'feature') #the most important features (variables) for the decision
```



One of the advantages of Random Forest is its ability to provide feature importance scores, which help in understanding the contribution of each feature to the model's 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.

The predictions in the test set

```
predict_model(rf) # the predictions in the test set
```

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
0	Random Forest Classifier	0.6860	0.7480	0.3438	0.6984	0.4607	0.2738	0.3080

	ph	Hardness	Solids	Chloramines	Sulfate	Conductivity	Organic_carbon	Trihalomethanes	Turbidity	Potability	prediction_label	prediction_score
2941	4.275160	229.443115	26098.638672	6.525203	281.896820	508.792084	16.133204	66.654190	3.732856	0	0	0.85
1009	NaN	194.806091	32981.238281	9.554786	NaN	286.794434	10.415889	71.905144	3.231680	0	0	0.57
803	7.804369	216.673874	14160.686523	6.060143	281.355988	468.212524	15.147547	62.860626	4.075557	1	0	0.63
2182	5.915807	195.744080	12431.802734	6.661616	380.725342	367.854034	21.300648	90.394897	4.513201	0	1	0.51
545	NaN	179.149292	4784.967773	5.702187	378.398224	517.496155	14.387223	68.687027	2.380084	0	0	0.59
...	...	...	...	...	...	...	...	...	...	...	...	...
2781	8.323982	262.218597	21292.828125	7.076944	350.325745	538.792297	17.684654	72.343811	4.153783	1	0	0.62
2002	7.035037	220.501892	19746.494141	7.059383	371.912628	536.183899	15.303433	67.901756	4.557029	1	1	0.62
621	NaN	193.400269	30630.759766	6.873568	342.193237	430.395599	17.446817	49.883652	2.913049	0	0	0.69
2994	4.497585	193.936508	27768.382812	6.906348	330.430695	556.032898	22.618412	81.265816	4.965739	0	0	0.73
1269	3.846814	190.992874	26895.257812	5.629536	NaN	660.254944	18.125202	78.153572	3.584202	0	1	0.53

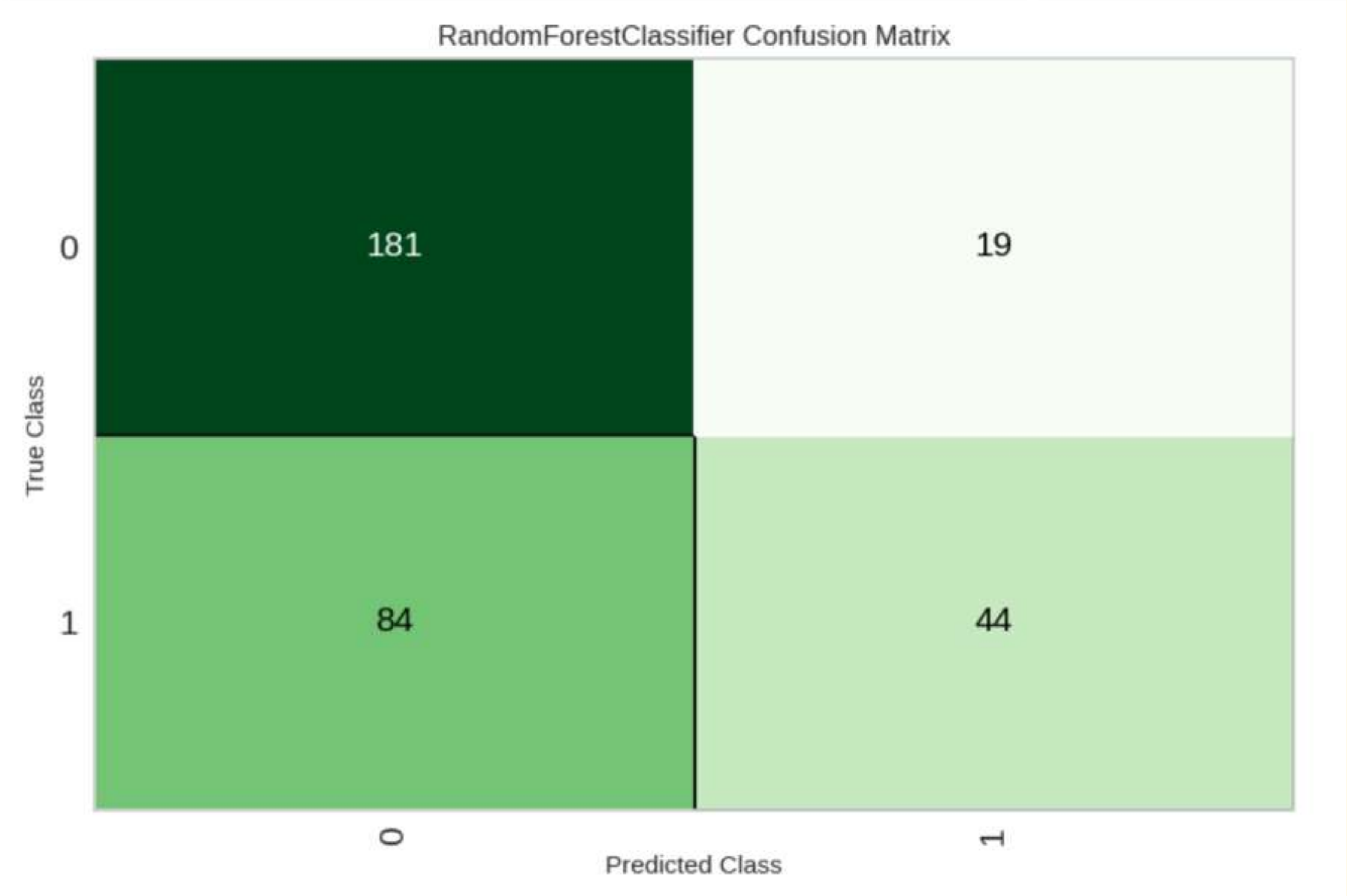
Co-funded by the European Union

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 confusion matrix for the test set

```
plot_model(rf, plot = 'confusion_matrix') # the confusion matrix for the test set
```



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/1RepTGNjf2D6tF2TGz2hP0h3QRRVdQtAL?usp=sharing

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 applications of classification in a smart city

- Classifiers could also be used to detect anomalies in data collected from connected devices such as sensors and IoT devices, helping to ensure that all infrastructure and services remain safe and secure.
- Classification algorithms can also be used to segment different parts of the city into different regions based on socio-demographic factors, income level, or crime rate. This can help city planners and decision makers better understand and plan for the needs of a given neighborhood or section of the city.

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

- Ullah, Z., Al-Turjman, F., Mostarda, L., & Gagliardi, R. (2020). Applications of Artificial Intelligence and Machine learning in smart cities. Computer Communications, 154, 313-323. https://doi.org/10.1016/j.comcom.2020.02.069
- Allam, Z., & Dhunny, Z. A. (2019). On big data, artificial intelligence and smart cities. Cities, 89, 80-91. https://doi.org/10.1016/j.cities.2019.01.032
- Yigitcanlar, T., & Cugurullo, F. (2020). The Sustainability of Artificial Intelligence: An Urbanistic Viewpoint from the Lens of Smart and Sustainable Cities. Sustainability, 12(20), 8548. https://doi.org/10.3390/su12208548

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

- Singh, S., Sharma, P. K., Yoon, B., Shojafar, M., Cho, G. H., & Ra, I.-H. (2020). Convergence of blockchain and artificial intelligence in IoT network for the sustainable smart city. Sustainable Cities and Society, 63, 102364. https://doi.org/10.1016/j.scs.2020.102364
- Yigitcanlar, T., Mehmood, R., & Corchado, J. M. (2021). Green Artificial Intelligence: Towards an Efficient, Sustainable and Equitable Technology for Smart Cities and Futures. Sustainability, 13(16), 8952. https://doi.org/10.3390/su13168952
- Rojek, I., & Studzinski, J. (2019). Detection and Localization of Water Leaks in Water Nets Supported by an ICT System with Artificial Intelligence Methods as a Way Forward for Smart Cities. Sustainability, 11(2), 518. https://doi.org/10.3390/su11020518
- Chui, K. T., Lytras, M. D., & Visvizi, A. (2018). Energy Sustainability in Smart Cities: Artificial Intelligence, Smart Monitoring, and Optimization of Energy Consumption. Energies, 11(11), 2869. https://doi.org/10.3390/en11112869

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.