https://www.geeksforgeeks.org/machine-learning/machine-learning-with-python/

https://www.geeksforgeeks.org/machine-learning/ml-implementation-of-knn-classifier-using-sklearn/

LINKS:

https://www.geeksforgeeks.org/machine-learning/how-to-prepare-data-before-deploying-a-machine-learning-model/

https://machinelearningmastery.com/train-test-split-for-evaluating-machine-learning-algorithms/

https://www.geeksforgeeks.org/machine-learning/stratified-sampling-in-machine-learning/

https://www.geeksforgeeks.org/data-science/stratified-random-sampling-an-overview/

https://www.geeksforgeeks.org/machine-learning/cross-validation-machine-learning/

https://scikit-learn.org/0.15/modules/generated/sklearn.cross\_validation.ShuffleSplit.html

https://scikit-learn.org/stable/modules/generated/sklearn.model\_selection.StratifiedShuffleSplit.html

https://scikit-learn.org/stable/modules/generated/sklearn.model\_selection.train\_test\_split.html

https://www.geeksforgeeks.org/machine-learning/how-to-do-train-test-split-using-sklearn-in-python/

https://www.w3schools.com/Python/python\_ml\_train\_test.asp

https://www.geeksforgeeks.org/machine-learning/how-to-split-a-dataset-into-train-and-test-sets-using-python/

https://www.geeksforgeeks.org/machine-learning/cross-validation-machine-learning/

https://www.geeksforgeeks.org/machine-learning/stratified-k-fold-cross-validation/

Example Usage:

Python

from sklearn.model\_selection import ShuffleSplit

import numpy as np

X = np.array([[1, 2], [3, 4], [5, 6], [7, 8], [9, 10], [11, 12]])

y = np.array([0, 1, 0, 1, 0, 1])

*# Create a ShuffleSplit object with 5 splits, 20% test size*

ss = ShuffleSplit(n\_splits=5, test\_size=0.2, random\_state=42)

*# Iterate through the splits*

for train\_index, test\_index in ss.split(X):

X\_train, X\_test = X[train\_index], X[test\_index]

y\_train, y\_test = y[train\_index], y[test\_index]

print(f"Train indices: {train\_index}, Test indices: {test\_index}")

When to Use ShuffleSplit:

* When you need multiple independent train/test splits for evaluating model performance.
* When the dataset is large, and a K-Fold cross-validation might be computationally expensive.
* When you want to control the exact size or proportion of training and testing sets.
* **Stratification:**For classification problems where maintaining class proportions in splits is important, consider using StratifiedShuffleSplit instead, which is a variation of ShuffleSplit that ensures this.

Como Fazer A Divisão Do Dataset Em Treino E Teste?

- Holdout

1) simples: mistura tudo, pega uma parte para Treino e outra para Teste

- K-Fold Cross Validation

Como Fazer O Split?

Desbalanceamento?

Como Fazer Split De Um Dataset-Desbalanceado?

Resposta: Estratificação

Treino -> Teste (Qtas Vezes?)

Kfold -> Validação Cruzada

Random Sampling

- Proporcional Random Sampling --> Estratificação

- Desproporcional Random Sampling

**Estratificação**

https://www.geeksforgeeks.org/machine-learning/how-to-implement-stratified-sampling-with-scikit-learn/

https://datascientyst.com/pandas-random-sampling-stratified-and-weighted/

https://www.reneshbedre.com/blog/stratified-sampling.html

https://www.geeksforgeeks.org/data-science/stratified-random-sampling-an-overview/

https://www.abzu.ai/data-science/stratified-data-splitting-part-1/

**KFOLD**

https://pt.linkedin.com/pulse/valida%C3%A7%C3%A3o-cruzada-estratificada-k-fold-maycon-cypriano-umtdf

https://machinelearninggeek.com/cross-validation-in-scikit-learn/

https://www.kaggle.com/code/satishgunjal/tutorial-k-fold-cross-validation

\*\* https://heartbeat.comet.ml/the-best-way-of-splitting-data-for-machine-learning-54c2f44cf409

| # splitting the dataframe into train and test sets X\_train,X\_test,y\_train,y\_test = train\_test\_split(X,y,test\_size=0.3,random\_state=101) scaler = StandardScaler() scaler.fit(X\_train) X\_train = scaler.transform(X\_train) X\_test = scaler.transform(X\_test) |
| --- |