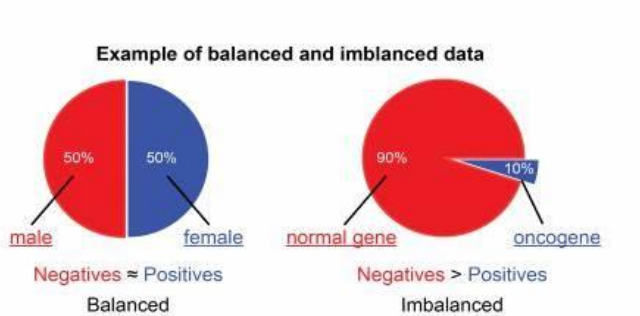
Imbalanced Dataset



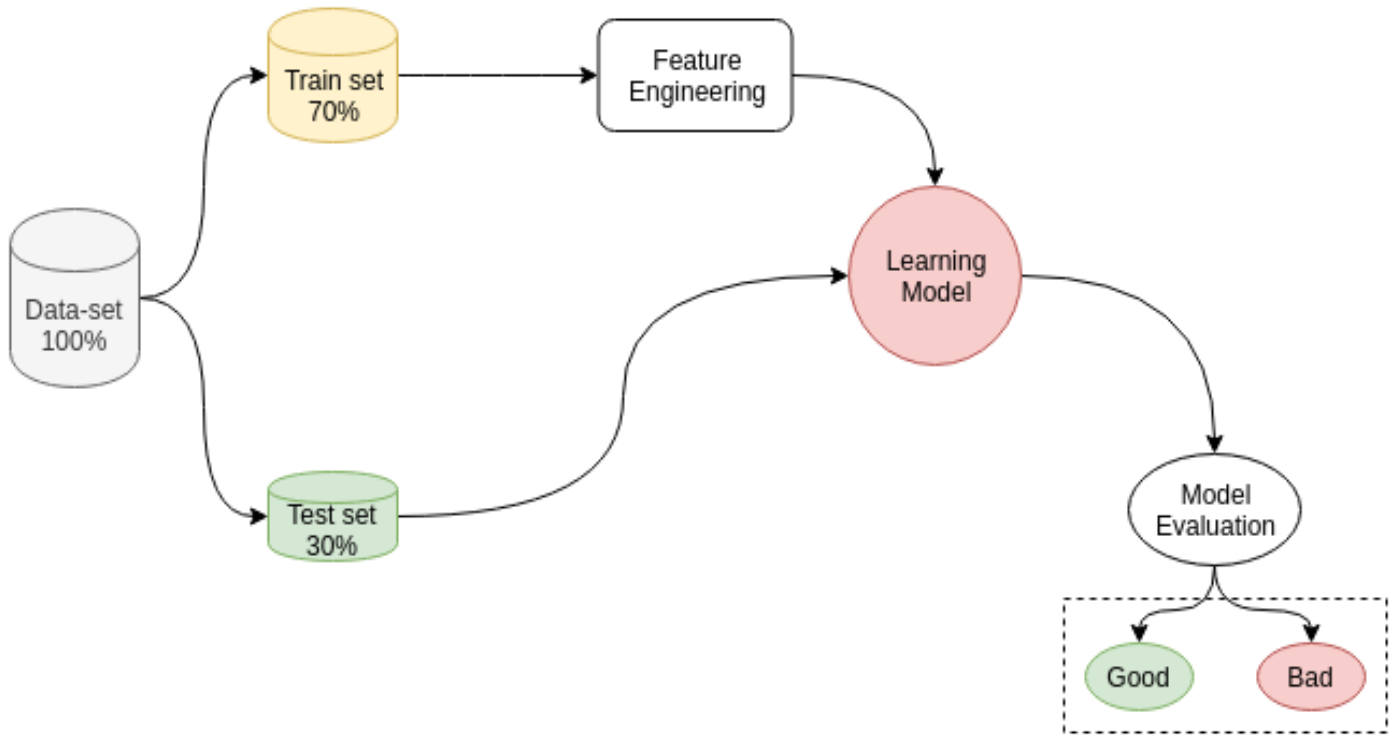
Cross Validation

Cross-validation is a resampling procedure used to evaluate machine learning models on a limited data sample. It is a technique used to protect against overfitting in a predictive model, particularly in a case where the amount of data may be limited. In cross-validation, you make a fixed number of folds (or partitions) of the data, run the analysis on each fold, and then average the overall error estimate.

Cross Validation

- Hold Out Cross Validation
- K-Fold Cross Validation
- Leave One-Out Cross Validation (LOOCV)
- Stratified K Fold Cross Validation

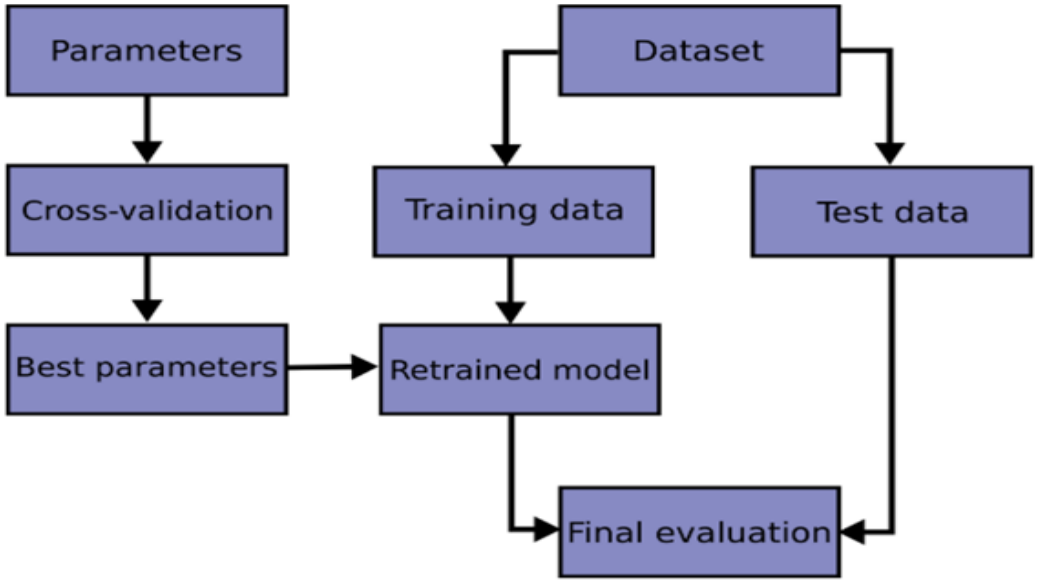
Machine Learning Model



Cross Validation

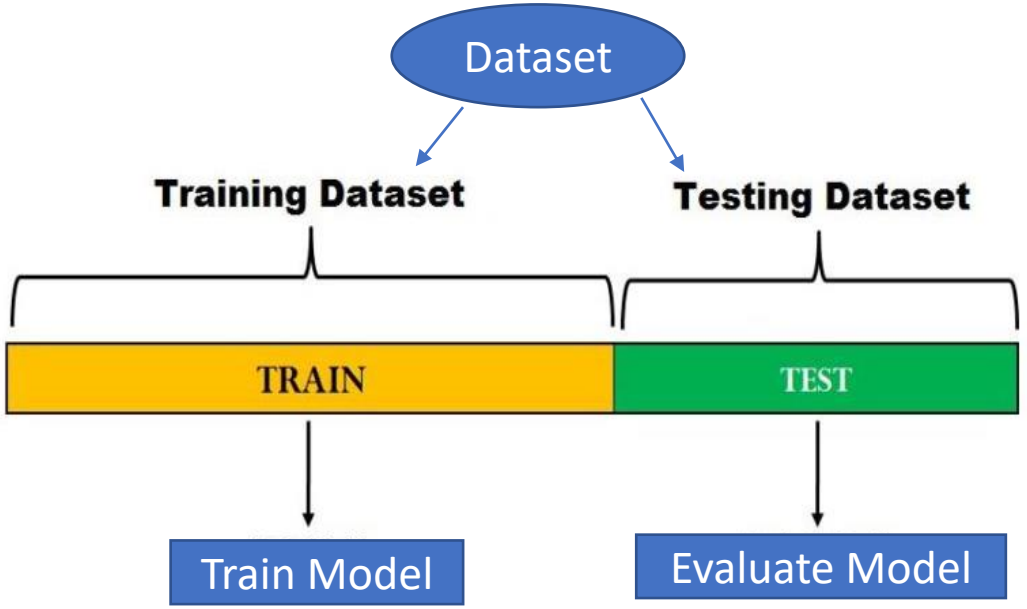
Cross-validation *is a resampling procedure used to evaluate machine learning models on a limited data sample.*

K = 5



Cross Validation

Hold Out Cross Validation



Cross Validation

Hold Out Cross Validation

```
from sklearn.model_selection import train_test_split

xtrain, xtest, ytrain, ytest = train_test_split(x,y, train_size=0.7, random_state=1)
```

Cross Validation

K-Folds Cross Validation

The general procedure is as follows:

1. Shuffle the dataset randomly.
2. Split the dataset into k groups
3. For each unique group:
    1. Take the group as a hold out or test data set
    2. Take the remaining groups as a training data set
    3. Fit a model on the training set and evaluate it on the test set
    4. Retain the evaluation score and discard the model
4. Summarize the skill of the model using the sample of model evaluation scores

Cross Validation

K-Folds Cross Validation

The general procedure is as follows:

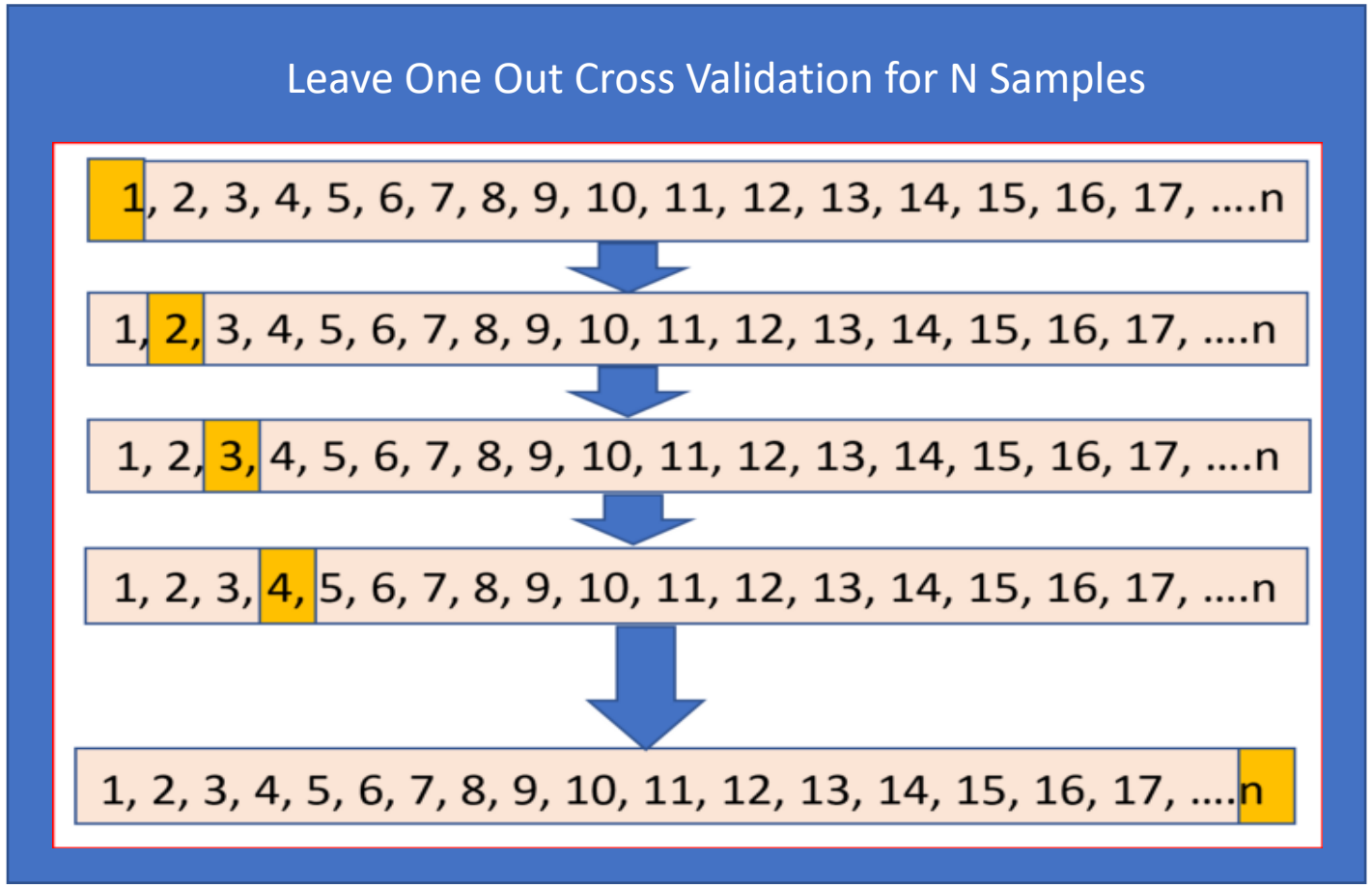
1. Shuffle the dataset randomly.
2. Split the dataset into k groups
3. For each unique group:
    1. Take the group as a hold out or test data set
    2. Take the remaining groups as a training data set
    3. Fit a model on the training set and evaluate it on the test set
    4. Retain the evaluation score and discard the model
4. Summarize the skill of the model using the sample of model evaluation scores

Cross Validation

K-Folds Cross Validation

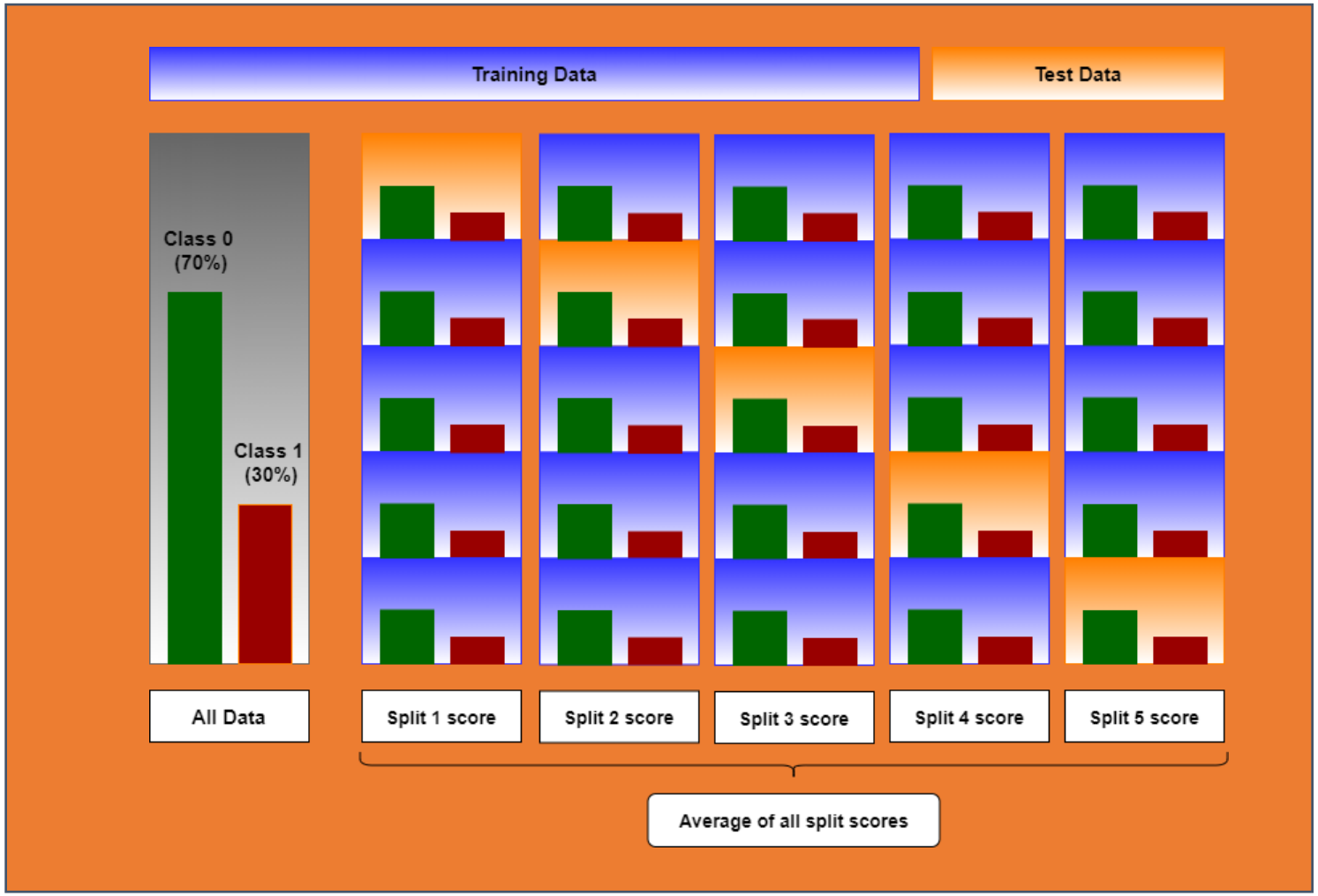
Cross Validation

Leave One-Out Cross Validation (LOOCV)



Cross Validation

Stratified K Fold Cross Validation



Cross Validation

Let’s Do it with PYTHON