What are Model Performance and its necessity?

Machine learning model performance assessment is just like assessing the scores, how we used to evaluate our scores in school. So apparently, the GOOD score recognizes the fact that the candidate is always good. The same is expected in the machine learning model, and that should achieve the expected results in predictions. Even in the ML world, the model has been trained, tested and used to to Predict.

K-Fold

That k-fold cross validation is a procedure used to estimate the skill of the model on new data. This measure how well a model would perform on the data given, This will help you pcik the best Model to use.

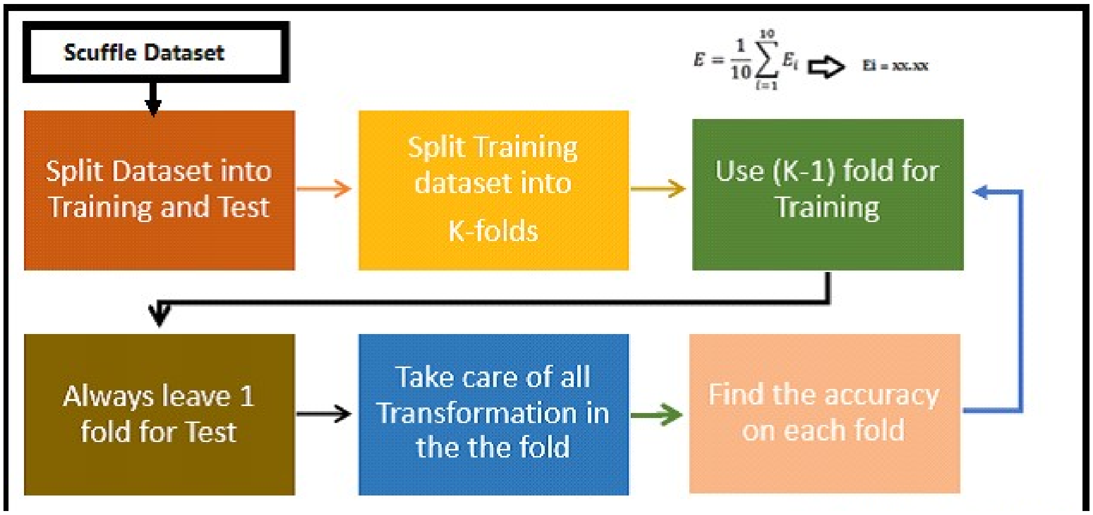
How K-Fold Works

Kfold takes the training data, in this case (X_train, X_train), Divides its into specified number of Folds I.e it be 10 Folds. Each Fold Specific Model. Check the accuracy on Each Fold, Note one Fold is Left for testing.

Take all the accuracy on Each Fold and Get average score.

Repeat this for all models available.

Get the model with Highest Score on Average. Use this model as in Your Training.


Scuffle Dataset
$E = \frac{1}{10}\sum_{i=1}^{10} E_i$
Ei = xx.xx
Split Dataset into Training and Test
Split Training dataset into K-folds
Use (K-1) fold for Training
Always leave 1 fold for Test
Take care of all Transformation in the the fold
Find the accuracy on each fold

Below Code Snippet shows how to perform a Cross Validation on Classification model, below we use a 10 Fold approach

```
# Imports Classification Models
from sklearn.tree import DecisionTreeClassifier
from sklearn.naive_bayes import GaussianNB
from sklearn.neighbors import KNeighborsClassifier
from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import RandomForestClassifier
from sklearn.ensemble import GradientBoostingClassifier
from sklearn.discriminant_analysis import LinearDiscriminantAnalysis
from sklearn.svm import SVC

# Create an Empty List and append all models, Give the Models an alias Name
# i.e we called KNeighborsClassifier() alias KNC etc
models = []
models.append(('DTree', DecisionTreeClassifier()))
models.append(('Gaussian', GaussianNB()))
models.append(('KNC', KNeighborsClassifier()))
models.append(('Random Forest', RandomForestClassifier()))
models.append(('Gradient Boosting', GradientBoostingClassifier()))
models.append(('Linear Disc', LinearDiscriminantAnalysis()))
models.append(('Support Machines', SVC(gamma = 'auto')))
# Import Cross Validation and KFOLD
from sklearn.model_selection import cross_val_score, Kfold
# Create a for loop so that each model is tested in turn
for name, model in models:
    # Here we do a 10 split K-FOLD
    kfold = KFold(n_splits = 10, random_state=42, shuffle=True)

    # We get the results for each Fold
    cv_results = cross_val_score(model, X_train, Y_train, cv = kfold,
    scoring='accuracy')

    # Get the average of all Folds
    print('Model Name:  ', name, '  Results: ', cv_results.mean())
```

Check working example on **Step 4a** in below notebook

https://colab.research.google.com/drive/1tu0UvkIfVszdFlbLzBaO3ux1r9SZEwaI?usp=sharing

More reading

https://en.wikipedia.org/wiki/Cross-validation_(statistics) https://www.statology.org/k-fold-cross-validation-in-python/