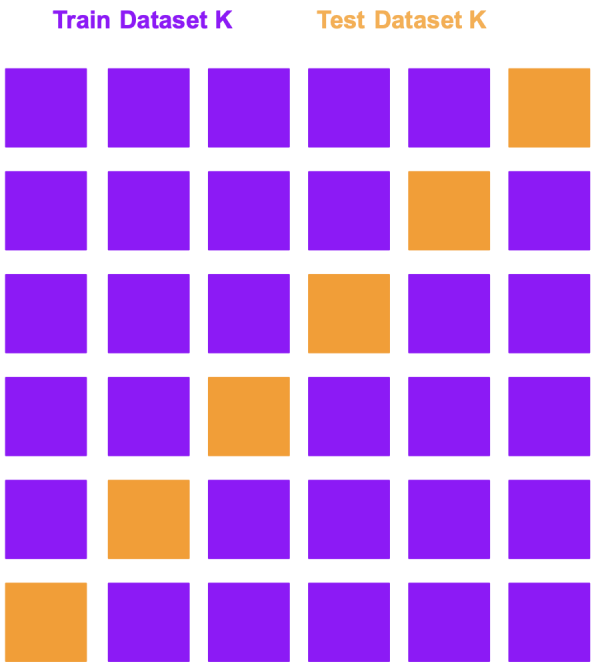
Alberto Andrés Valdés González.
Degree: Mathematical Engineer.
Work position: Data Scientist.
Mail: anvaldes@uc.cl/alberto.valdes.gonzalez.96@gmail.com
Location: Santiago, Chile.

K Folds

K - Folds (Classic)

First of all we have all the data and we split it in Train and Test set.

Then we do many splits of the Train set.



We train the model with the respective Trainset (Train Dataset K) and we measure the performance on the Train dataset K and Test Dataset K. We take decisions based on the average performance on the Test Dataset K (Like hyperparameter tuning).

If we want to train a Neural Network we can do two things:

- Use the Test Dataset K as Val Dataset K (recommended).
- Split the Train Dataset K on two dataset that will be Train Dataset K and Val Dataset K.

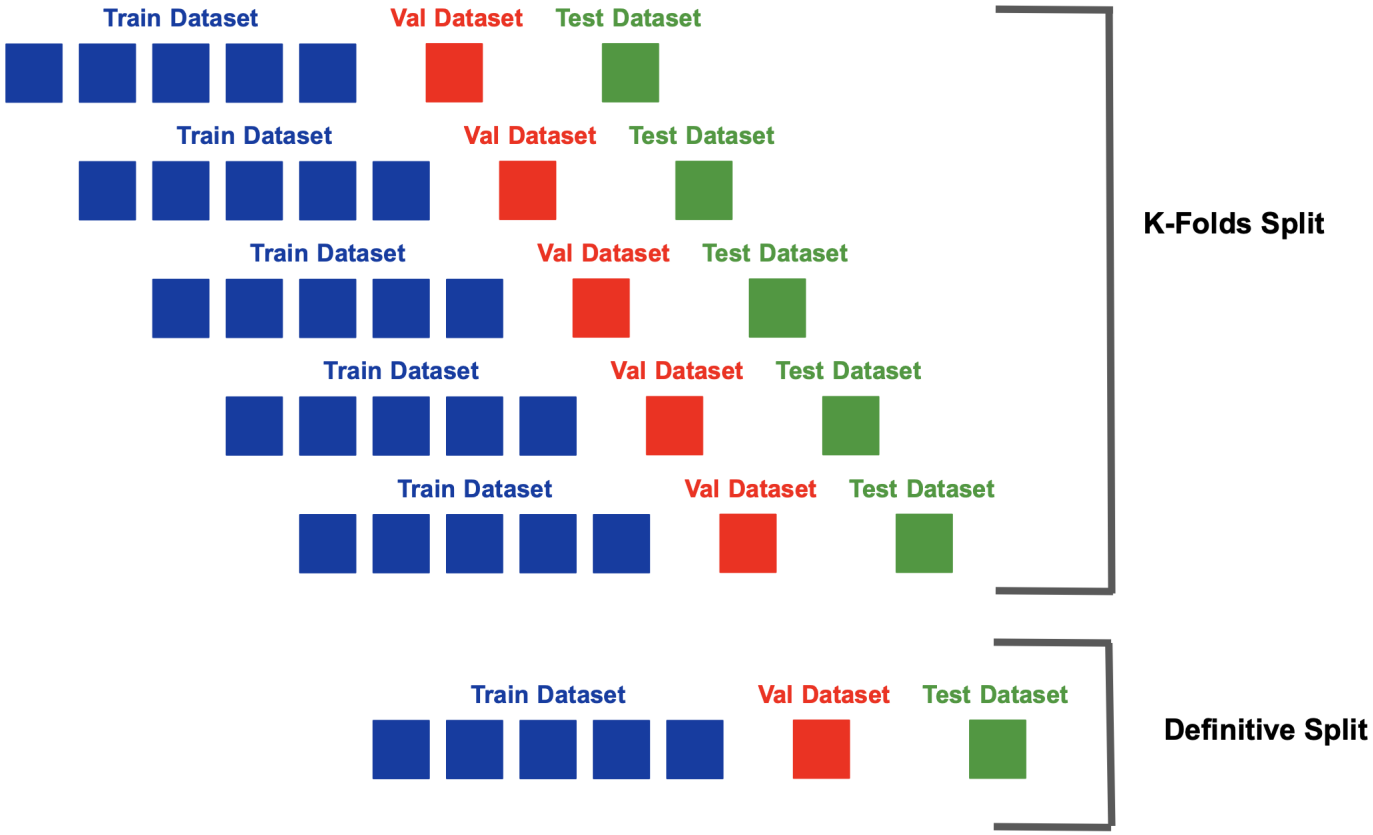
Finally we do a final split to evaluate the performance with the optimal parameters:

As you can see the **size** of the final Train Dataset is very similar to one was used on the K-Folds which is **very important**. Here we simulate the final implementation of the model.

Note in this cases the **calibration** must be done it in the final split.

K - Folds (Time Series)

Here we have temporal dependency and for this reason the application of K-Folds must be different.



We train the model with the respective Trainset and we measure the performance on the Train dataset and Test Dataset. We take decisions based on the average performance on the Test Dataset (Like hyperparameter tuning).

If we want to train a Neural Network we can do it without problems because we have the Train/-Val/Test split.

Finally we take the definitive split which size is very similar to the others was used in the K-Folds which is **very important**. Here we simulate the final implementation of the model.

Note in this case the **calibration** can be done it in all the splits.
