CROSS VALIDATION

CROSS VALIDATION

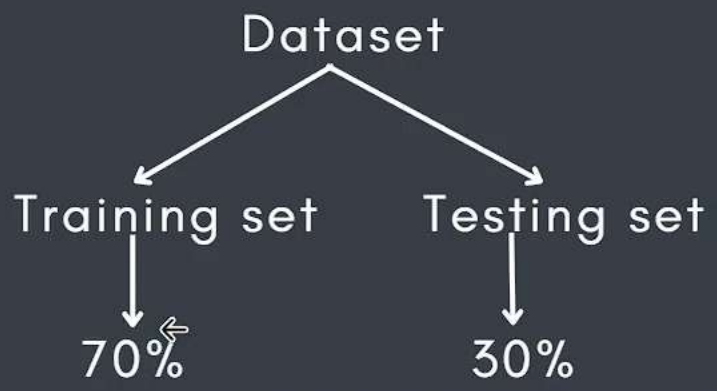
Cross Validation (CV) is a way to check how well a machine learning model will work on new, unseen data by splitting the dataset into several smaller parts, training on some parts, and testing on the remaining part — and then repeating this process multiple times.

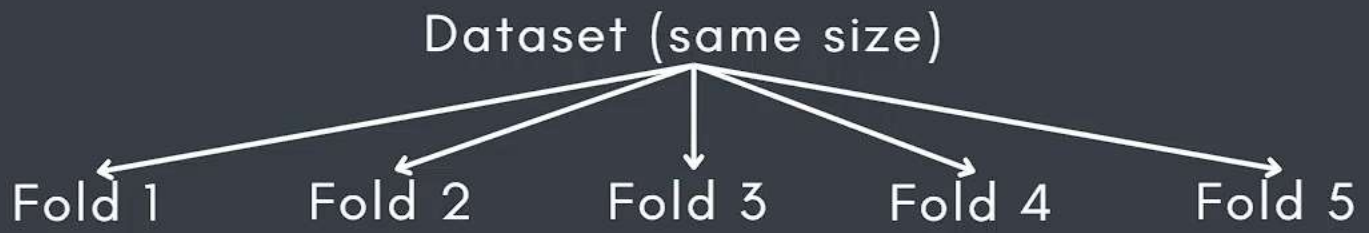
It helps to:

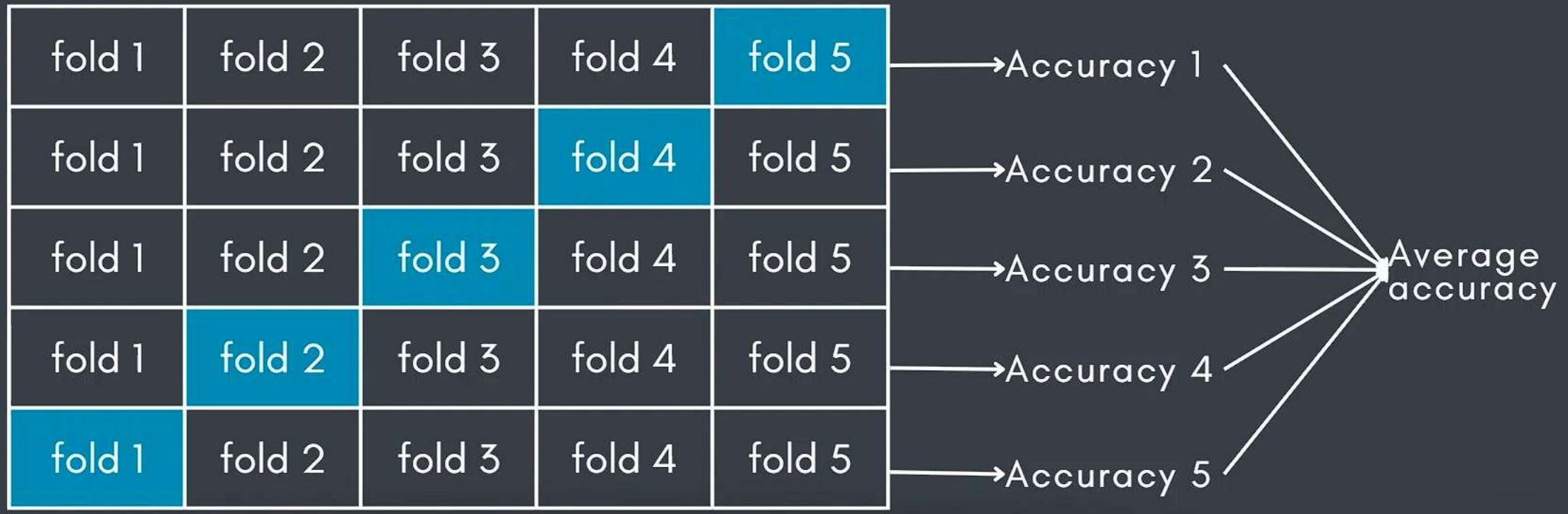
- Avoid overfitting (model memorising data instead of learning patterns)
- Get a more reliable accuracy score
- Make better decisions when choosing algorithms or tuning parameters

Think of it like testing a student on different question papers to see if they've really learned the subject, not just memorised one set of answers.

WHY CROSS VALIDATION



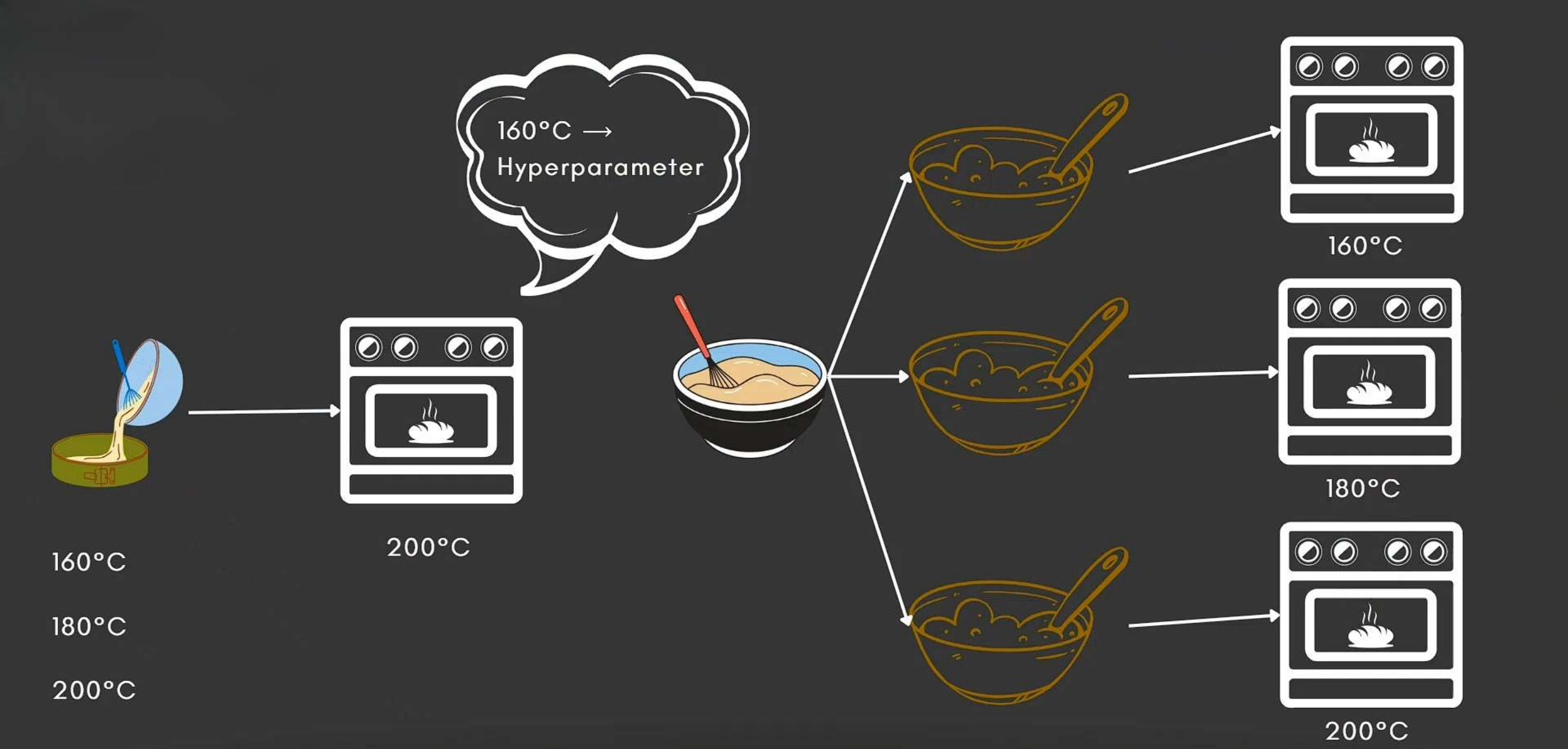




WHY CROSS VALIDATION

1 Cross-validation is an essential technique in machine learning because it helps to prevent overfitting or underfitting of a model. Overfitting occurs when the model is too complex and fits the training data too closely, resulting in poor performance on new data. On the other hand, underfitting occurs when the model is too simple and does not capture the underlying patterns in the data, resulting in poor performance on both the training and test data.

2 Cross-validation also helps to determine the optimal model hyperparameters. Hyperparameters are the settings that control the behavior of the model. For example, in a decision tree algorithm, the maximum depth of the tree is a hyperparameter that determines the level of complexity of the model. By using cross-validation to evaluate the performance of the model at different hyperparameter values, we can select the optimal hyperparameters that maximize the model's performance.


160°C →
Hyperparameter
200°C
160°C
180°C
200°C
160°C
180°C
200°C

MCQS

1. Which of the following best describes the purpose of cross-validation?
   A) To shuffle the dataset randomly before training
   B) To evaluate model performance on different subsets of the data
   C) To reduce the size of the dataset
   D) To increase the number of features in the dataset

2. In k-fold cross-validation, if k = 5, how many times will the model be trained and tested?
   A) 1 time
   B) 3 times
   C) 5 times
   D) 10 times