Validation & Cross-Validation H2O Parameters

Choose one of two approaches:

1. Validation
   - **`validation_frame`**: id of the validation data frame.
2. Cross-Validation
   - **`nfolds`**: number of folds for k-fold cross-validation
   - **`fold_column`**: column with cross-validation fold index assignment per observation.
   - **`fold_assignment`**: cross-validation fold assignment scheme, if fold column is not specified.
   - **`keep_cross_validation_fold_assignment`**: keep the cross-validation fold assignment.
   - **`keep_cross_validation_predictions`**: keep the predictions of the cross-validation models.

Data Leakage

What Is It

- Leakage is allowing your model to use information that will not be available in a production setting.
- Example: using the Dow Jones daily gain/loss as part of a model that predicts stock performance.

What Happens

- Model is overfit.
- Predictions will be inconsistent with those scored during model training (even with a validation set).
- Insights derived from the model will be incorrect.

What to Do

- Understand the nature of your problem and data.
- Scrutinize model feedback, such as relative influence or linear coefficient.

H_2O.ai

Validation & Cross-Validation H2O Parameters

Choose one of two approaches:

1. Validation
   - **`validation_frame`**: id of the validation data frame.
2. Cross-Validation
   - **`nfolds`**: number of folds for k-fold cross-validation
   - **`fold_column`**: column with cross-validation fold index assignment per observation.
   - **`fold_assignment`**: cross-validation fold assignment scheme, if fold column is not specified.
   - **`keep_cross_validation_fold_assignment`**: keep the cross-validation fold assignment.
   - **`keep_cross_validation_predictions`**: keep the predictions of the cross-validation models.