

Prevent leakage in industrial time series

Why randomly splitting sensor rows can exaggerate real-world performance.

In short

- Adjacent readings often share the same conditions, so windows from one failure event must not land in both training and test sets.
- Scaling, imputation and feature selection must be fitted on training data only, because including the test mean in a scaler leaks future information.
- Choose the model, window and alert threshold on validation data, and keep the final test out of those decisions.

01 / Rows are not independent

Adjacent readings often share the same conditions. Neighboring windows from one failure event can land in both training and test sets, exposing the event before evaluation. Many rows may still represent only a few independent failures. Report the number of rows, events and assets separately.

02 / Consider both time and assets

For next-month prediction, train on the past and test on a later period. For deployment to new machines, hold out entire machines. Leave a gap when window lengths or prediction horizons overlap the boundary. A single split may not establish both new-asset and new-season generalization.

03 / Preprocessing learns too

Fit scaling, imputation and feature selection on training data only. Including the test mean in a scaler leaks future information. Refit the preprocessing pipeline inside each validation fold. Check whether a maintenance outcome field was populated after the moment at which predictions would have been made.

04 / Protect the final test

Choose the model, window and alert threshold using validation data. Keep the final test out of those decisions. If the result is poor, development can continue, but repeatedly using that test reduces its independence. Confirm the next release on a newly reserved period.

Try it

Draw your validation workflow in Flow Maker, from chronological split to final holdout.

References and online guide

[Open the online guide](#)

scikit-learn: common pitfalls

scikit-learn: cross-validation

Proposed pilot steps do not claim field accuracy. Synthetic examples are not measurements from a production site.