Time Series Cross-Validation: Walk-Forward Method
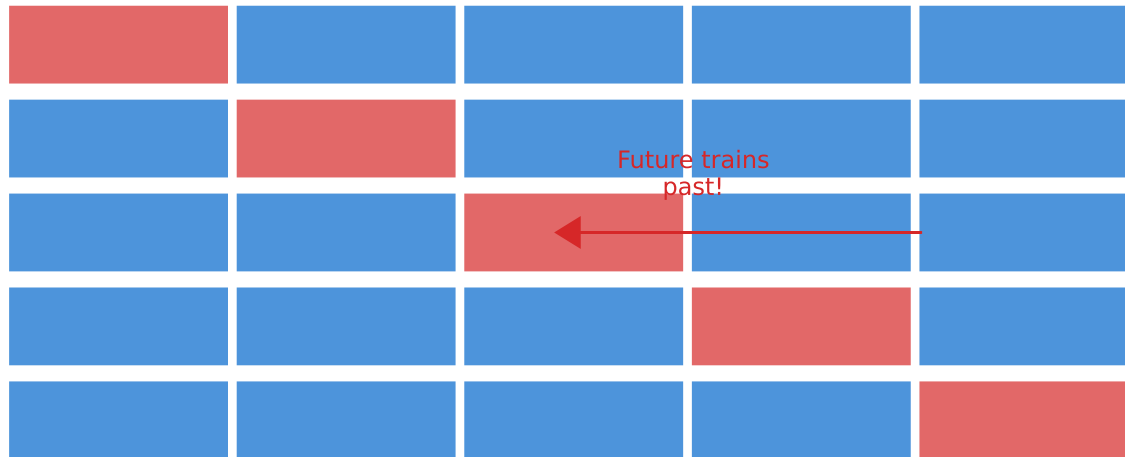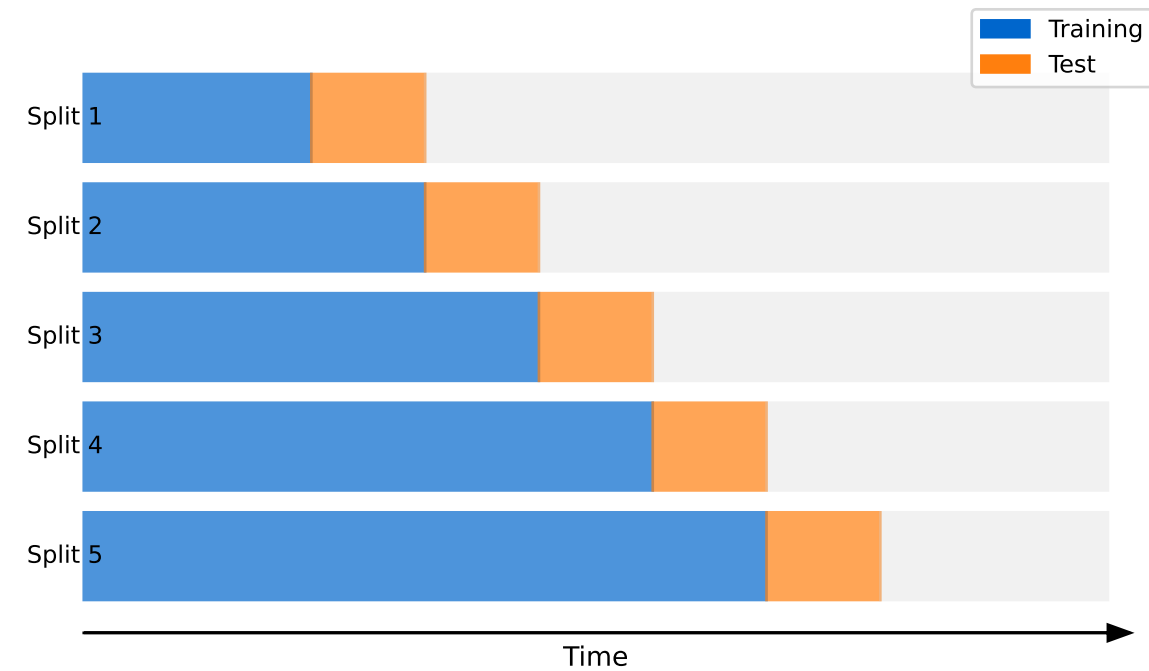
Problem: Standard CV Leaks Future Information
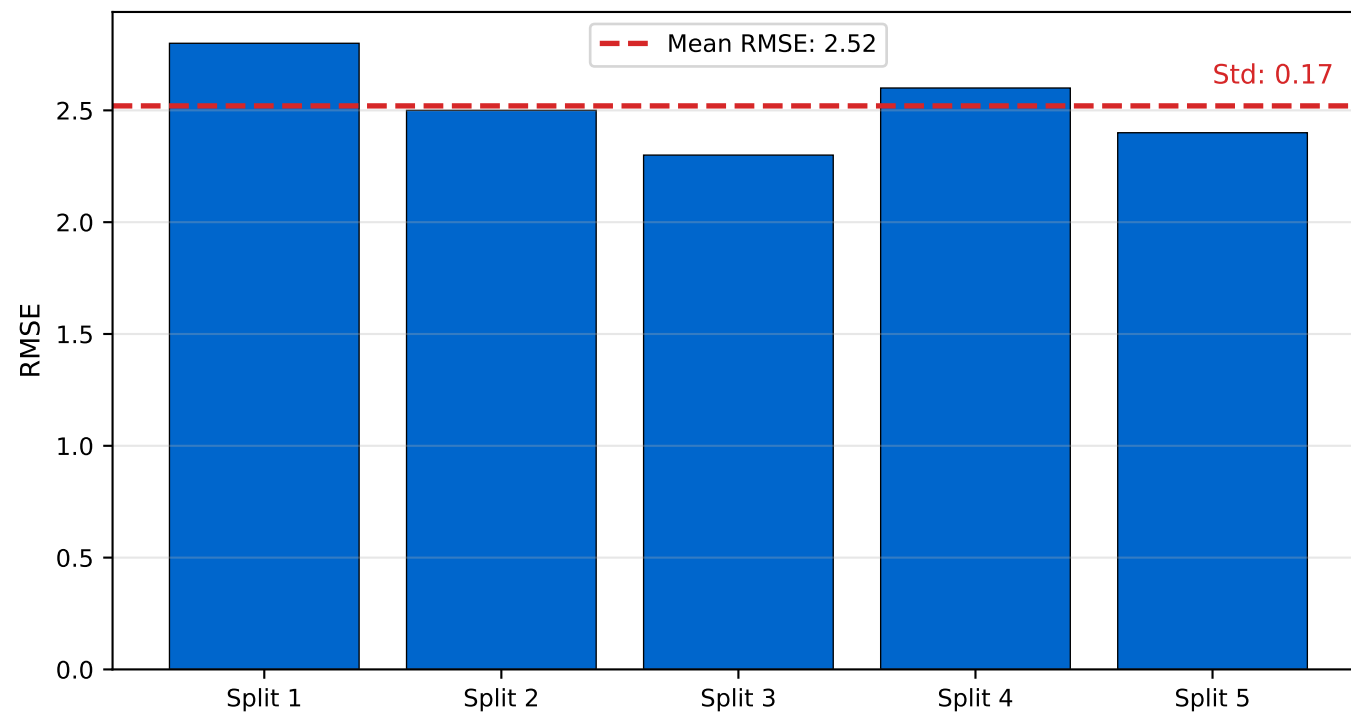

Future trains past!

Standard CV: Future data trains model predicting past!

Solution: Walk-Forward CV (Expanding Window)

Legend: Training, Test

Split 1
Split 2
Split 3
Split 4
Split 5

Time

Walk-Forward CV Results

Mean RMSE: 2.52

Std: 0.17

RMSE

Split 1, Split 2, Split 3, Split 4, Split 5

sklearn Implementation

```
TIME SERIES CROSS-VALIDATION IN SKLEARN

from sklearn.model_selection import TimeSeriesSplit
from sklearn.linear_model import Ridge
from sklearn.metrics import mean_squared_error
import numpy as np

# Create time series splitter
tscv = TimeSeriesSplit(n_splits=5)

# Walk-forward validation
scores = []
for train_idx, test_idx in tscv.split(X):
    X_train, X_test = X[train_idx], X[test_idx]
    y_train, y_test = y[train_idx], y[test_idx]

    model = Ridge(alpha=1.0)
    model.fit(X_train, y_train)
    y_pred = model.predict(X_test)

    rmse = np.sqrt(mean_squared_error(y_test, y_pred))
    scores.append(rmse)

print(f"Mean RMSE: {np.mean(scores):.3f}")
print(f"Std RMSE:  {np.std(scores):.3f}")

# Alternative: use cross_val_score
from sklearn.model_selection import cross_val_score

scores = cross_val_score(
    model, X, y, cv=tscv,
    scoring='neg_root_mean_squared_error'
)
```