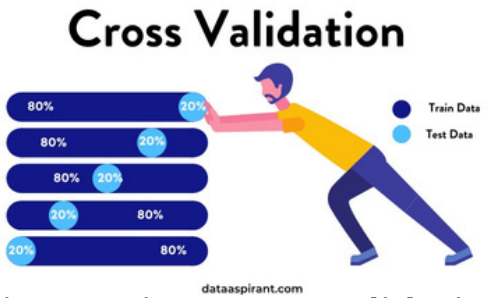


Cross-validation is a technique used in machine learning to check how well a model performs on new data. Instead of testing the model on just one set of data, cross-validation splits the data into multiple parts and tests the model several times. This helps ensure the model is accurate, reliable, and not just memorizing the training data.

Time Series Cross-Validation

Unlike standard cross-validation, Time Series CV ensures that past data is only used to predict the future, preventing data leakage in time-dependent datasets. It is essential for applications like stock market prediction, weather forecasting, and anomaly detection.

Types of Time Series CV:

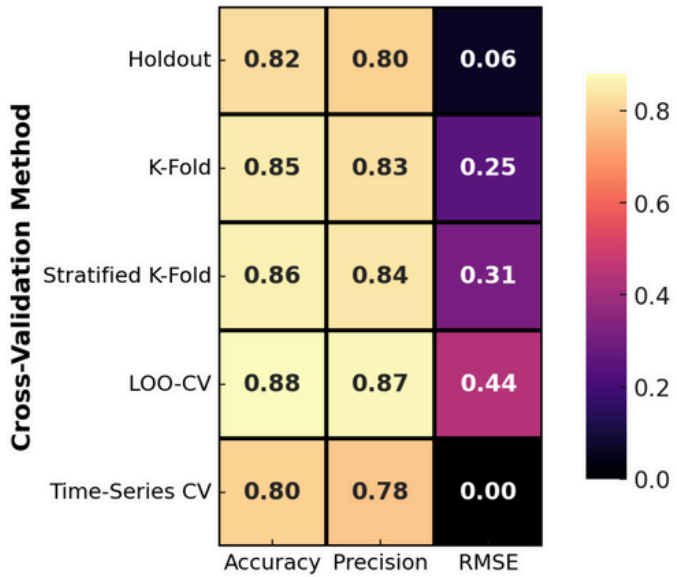
- Rolling Window CV : A fixed-size training set moves forward in time, keeping the test set at the next step.
- Expanding Window CV: The training set grows over time, incorporating more historical data while keeping the test set size constant.

Why Use Time Series CV?

✓ Prevents data leakage by maintaining temporal order.
✓ Mimics real-world scenarios where future data is unknown.
✓ Helps evaluate model stability over different time periods.

	WHAT IS ?	BEST FOR	HANDLED IMBALANCED DATA?	KEY LIMITATIONS
K-FOLD	Splits data into K subsets (folds) and trains K times, each fold used as test data once	Small Dataset	✗	Computationally expensive
STRATIFIED K-FOLD	Similar to K-Fold but maintains the class distribution across folds	Imbalanced classification	✓	Slightly slower
HOLD-OUT CV	Splits data into training and testing sets (e.g., 80-20 split)	Large Datasets	✗	Biased due to one split
LEAVE-ONE-OUT	Uses one sample as test data and the rest as training, repeated for all samples	Very small Datasets	✗	Too slow for large data

Performance Comparison of Cross-Validation Methods




```python
from sklearn.model_selection import  KFold, StratifiedKFold, LeaveOneOut, TimeSeriesSplit
kf = KFold(n_splits=5)
skf = StratifiedKFold(n_splits=5)
loo = LeaveOneOut()
tscv = TimeSeriesSplit(n_splits=5)
```

Famous Real-World Use Cases

- Stock Market Prediction → Time-Series CV ensures no future leakage.
- Medical Diagnosis → Stratified K-Fold balances class distributions.

Poster By : Sahil Khedkar Vedant Gaikwad Omkar Kevate