

Data Taxonomy

Static Datasets

- Cross-sectional tables (e.g., reference data, security master)
- Snapshots / embeddings / cross-sectional features
- Use cases: schema/constraint checks, cross-source reconciliation, outlier rows

Time Series

- Prices, quotes, trades, volumes, derived indicators
- Seasonality/trend, calendar effects, microstructure noise
- Use cases: point anomalies, level shifts, drifts, regime changes

Algorithm Taxonomy

Baselines (Column-wise)

- Range/null/regex checks; rule validation
- Simple statistics: z-scores, modified z, etc.
- Control charts for sequential data

Unsupervised (No Labels)

- Learn the “normal” latent structure and flag outliers
- Patterns leveraged: density/distance, low-rank/projections, reconstruction error, robust distributional fit
- Time-series view: forecast residuals and decomposition residuals

Supervised (With Labels)

- Train a model on *historical labeled issues*
- Emphasis on class imbalance, cost-sensitive objectives, human-in-the-loop review
- Best for recurring, well-defined error types