JUPITER PRIME
Consulting • Training • Advisory
جوبيتر برايم للاستشارات والتدريب

COURSE BROCHURE

Predictive Model Validation and Data Leakage Control Training Course

3 – 7 May 2027
Dar es Salaam

Predictive Model Validation and Data Leakage Control Training Course

Ref: 1493_55303 **Date:** 3 – 7 May 2027 **Location:** Dar es Salaam **Fees:** 7900 GBP

Predictive Model Validation and Data Leakage Control is a 1-week training course for analysts and business professionals who build regression and classification models and must prove they will hold up on new cases, and it produces a Model Validation Dossier. As more decisions rest on scores built quickly in notebooks and automated tools, an unchecked leak or overfitted model can pass every test and still fail in use. Jupiter Prime Training delivers this course on Predictive Model Validation and Data Leakage Control.

Why Predictive Model Validation and Data Leakage Control Matters Now

Automated modelling tools let analysts train dozens of candidate models in an afternoon, so one may look impressive purely by luck. Business sponsors, meanwhile, want a single accuracy figure and rarely ask how it was measured. Organisations that separate training, tuning and test data properly, and that hunt for leakage before launch, get models whose live performance matches the pilot; those that skip this step discover the gap only after customers or budgets are affected. The course draws on ISO/IEC 42001, the AI management system standard, and the NIST AI Risk Management Framework NIST AI RMF.

Who Gains Most from Predictive Model Validation and Data Leakage Control

- Analysts who build churn, demand or risk models in spreadsheets, notebooks or automated tools
- Professionals who choose between candidate models before a business launch
- Reviewers who sign off model results for finance, operations or marketing
- Data specialists who prepare features and join tables for modelling
- Report owners who present predictions and their uncertainty to sponsors

Analytics groups that share a model review process gain most when several members attend together.

Predictive Model Validation and Data Leakage Control Capabilities

Participants practise separating genuine predictive signal from artefacts of the way data was split, joined or tuned.

- Design train, validation and test splits that respect time and customer grouping — model reviewers
- Test feature sets for target leakage before any model is fitted — data specialists
- Calculate regression and classification metrics suited to the business cost — sponsors
- Weigh simpler against complex models using learning curves and regularisation — analytics leads
- Apply cross-validation and nested tuning without contaminating the test set — analysts
- Communicate expected live accuracy with honest error ranges — decision owners

Key Predictive Model Validation and Data Leakage Control Terms

- Data leakage — information in training data that would not be available at prediction time, inflating measured accuracy.
- Overfitting — a model learning noise in training data so it performs worse on new cases.
- Cross-validation — rotating which slice of data is held out, to estimate performance more reliably.
- Holdout test set — data kept untouched until the final check of a chosen model.
- Regularisation — a penalty on model complexity that reduces overfitting.
- Calibration — agreement between predicted probabilities and how often outcomes actually occur.

Predictive Model Validation and Data Leakage Control Day by Day

Day 1: Why Promising Models Fail in Use

- Pilot Versus Live Performance Review — Diagnosing Common Failure Stories
- Prediction Moment Definition — Fixing What Is Known When Scoring
- Problem Framing Canvas — Choosing Regression or Classification Targets
- ISO/IEC 42001 Verification Expectations — Documenting Model Testing Evidence
- NIST AI RMF Measure Function — Linking Metrics to Trustworthiness

Day 2: Splitting Data and Hunting for Leakage

- Time-Based and Grouped Splits — Preventing Customer Overlap Between Sets
- Target Leakage Audit — Removing Post-Outcome Fields From Features
- Join and Aggregation Leakage Checks — Catching Future Data in Summaries
- Preprocessing Inside the Pipeline — Fitting Scalers and Encoders on Training Only
- Duplicate and Near-Duplicate Detection — Protecting Test Set Independence

Day 3: Building and Tuning Regression and Classification Models

- Baseline Model Benchmark — Setting the Bar Any Model Must Beat
- Feature Engineering Experiments — Testing Ratios, Lags and Interactions
- Regularised Regression and Tree Ensembles — Comparing Bias and Variance
- Nested Cross-Validation — Tuning Hyperparameters Without Optimistic Scores
- Learning Curve Reading — Deciding Between More Data and Simpler Models

Day 4: Metrics, Model Selection and Communication

- Error Metric Selection — Matching Mean Absolute Error or Recall to Cost
- Calibration and Threshold Checks — Aligning Probabilities With Observed Rates
- Subgroup Performance Comparison — Spotting Segments Where the Model Struggles
- Model Selection Scorecard — Ranking Candidates on Accuracy, Stability and Simplicity
- Prediction Explanation Summary — Presenting Drivers and Uncertainty to Sponsors

Day 5: Practising the Model Validation Dossier

- Consumer Lending Default Case — Applying Grouped Splits to Repeat Borrowers
- Telecom Churn Case — Applying Leakage Audits to Account Histories
- Peer Validation Review — Challenging Each Other's Metrics and Splits
- Monitoring Trigger Plan — Setting Retest Points After Launch
- Model Validation Dossier — Completing Evidence for Sign-Off

Applying Predictive Model Validation and Data Leakage Control After the Course

- Run a target leakage audit on one model currently in use and report any fields to remove.
- Rebuild the test split of a live model with time-based or grouped rules and compare the new score.

Predictive Model Validation and Data Leakage Control Questions FAQ

What is predictive model validation and data leakage control and why does it matter now?

It is the discipline of proving a model will perform on unseen cases. It matters now because automated tools produce many candidate models quickly, and leaked or overfitted ones look convincing until they fail in use.

Which roles gain most from predictive model validation and data leakage control?

Analysts who build models, reviewers who sign them off and data specialists who prepare features gain most.

How does predictive model validation and data leakage control differ from a broad machine learning course?

It concentrates on testing and selecting regression and classification models rather than surveying deep learning, language or vision methods, so each day goes deeper into splits, metrics and leakage.

What do you bring back from predictive model validation and data leakage control training?

You bring back a validation file for a real model, showing its splits, leakage checks, chosen metrics and expected live accuracy.

Predictive Model Validation and Data Leakage Control: The Way Forward

The Model Validation Dossier first informs whether a candidate model is approved for launch or sent back for rework. Analysts, reviewers and business sponsors use it as the shared record of evidence. After its first use, revisit the monitoring triggers and subgroup scores against live results, to see which checks need tightening.