APES
ENERGY EVOLUTION

**TECHNICAL TRAINING COURSE**

1 day | Instructor-led and hands-on

# Uncertainty Quantification Using Machine Learning

## Practical conformal prediction for regression, classification and production forecasting

A hands-on one-day course for petroleum professionals who build, review or use machine-learning predictions. Learn to add calibrated prediction intervals and prediction sets to realistic engineering workflows, then test whether those uncertainty estimates achieve their intended coverage.

A model can score well on average accuracy and still be unreliable for the next well, fluid sample or production period. This course connects uncertainty methods with engineering decisions through PVT regression, binary well classification and monthly production forecasting.

### WHO SHOULD ATTEND

- Petroleum, reservoir and production engineers
- Geoscientists, petrophysicists and subsurface data specialists
- Technical professionals who build, review or rely on machine-learning forecasts
- Asset-evaluation, reserves, planning and decision-support practitioners
- Data scientists working with sparse, heterogeneous or time-dependent energy data
- Technical leaders reviewing model confidence and prediction risk

### PREREQUISITES

Basic familiarity with regression and classification is helpful. Advanced statistics is not required. Basic Python familiarity is desirable for modifying the notebooks, but guided exercises allow domain professionals to follow the workflow without being expert programmers.

### LEARNING OUTCOMES

- Distinguish point estimates, confidence intervals, prediction intervals, probabilities and prediction sets.
- Explain training, calibration and test data in split conformal prediction.
- Build conformal prediction intervals around a regression model.
- Build prediction sets for binary classification.
- Evaluate empirical coverage, interval width, set size and performance across relevant subgroups.
- Apply chronological splitting and lag design to production time series.
- Choose an uncertainty output suited to the engineering decision.
- Communicate uncertainty visually and in plain language to technical reviewers and decision-makers.

### UNDERSTANDING THE GUARANTEES

Conformal prediction can wrap around many predictive models and, under its stated assumptions, provide finite-sample marginal coverage without a parametric error distribution. Standard validity depends on exchangeability. Time dependence and distribution shift require appropriate methods, chronological validation and diagnostics. Marginal coverage does not guarantee coverage for every subgroup or individual case; stronger distribution-free conditional claims generally need additional restrictions.

# Three petroleum workshops

A connected sequence of regression intervals, classification sets and time-aware forecast intervals, using practical petroleum datasets.

## 01 PVT regression

**Course dataset.** 896 usable records across eight regional sheets. Variables include solution gas-oil ratio (Rs), oil formation-volume factor (Bo), oil relative density, gas relative density and temperature.

**Practical work.** Build a baseline PVT regression workflow using Rs and Bo, then conformalise its predictions. Compare point-error metrics with empirical coverage and interval width. Examine regional heterogeneity, subgroup performance and out-of-domain behaviour.

**Decision lesson.** An interval may be valid on average yet be less useful for a particular subgroup. Test whether a model calibrated on pooled regional data should be used unchanged in a new basin.

## 02 Well classification

**Course dataset.** 203 observations, four documented predictors and a binary class label: 109 class-1 records and 94 class-0 records.

**Practical work.** Train a binary classifier and build prediction sets containing one or both labels. Evaluate class coverage and average set size alongside accuracy. Interpret the documented labels without assigning an unsupported geological meaning.

**Decision lesson.** A one-label set is more decisive; a two-label set makes ambiguity visible and may call for further evidence or review. Class probabilities are not automatically calibrated uncertainty statements, and a forced single-label answer can hide decision risk.

## 03 Volve production forecasting

**Course dataset.** A consolidated course workbook with 112 monthly records from September 2007 to December 2016, including production, injection, operating time, cumulative values and ratio variables.

**Practical work.** Forecast monthly oil production with lag features, rolling statistics and chronological evaluation. Treat cumulative and contemporaneously derived variables as potential leakage sources unless known at forecast time. Explore rolling or adaptive calibration, including EnbPI-style intervals and adaptive conformal inference, with rolling coverage diagnostics.

**Decision lesson.** Random splitting can produce misleadingly optimistic operational forecasts. Check operating interruptions, changing production regimes, rising water cut and forecast degradation. Time-dependent data need specific calibration and diagnostics.

APES
ENERGY EVOLUTION

**TECHNICAL TRAINING COURSE**

1 day | Instructor-led and hands-on

# One-day course programme

Indicative agenda. Classroom or live virtual delivery; timing is confirmed for each course.

| TIME | MODULE AND PRACTICAL OUTPUT |
|---|---|
| 08:30 - 09:00 | **Why point predictions are insufficient**<br>Decision-focused uncertainty checklist |
| 09:00 - 10:00 | **Conformal prediction foundations**<br>Training, calibration, test workflow and coverage interpretation |
| 10:00 - 10:15 | **Break** |
| 10:15 - 11:45 | **PVT regression workshop**<br>Rs and Bo prediction intervals and coverage plot |
| 11:45 - 12:30 | **Interval diagnostics**<br>Coverage, width, subgroup performance and failure modes |
| 12:30 - 13:15 | **Lunch** |
| 13:15 - 14:30 | **Well classification workshop**<br>Binary prediction sets and set-size diagnostics |
| 14:30 - 14:45 | **Break** |
| 14:45 - 16:15 | **Volve time-series workshop**<br>Chronological forecast intervals |
| 16:15 - 16:45 | **From model output to decisions**<br>Reporting template, limitations and model-governance questions |
| 16:45 - 17:00 | **Knowledge check and close**<br>Personal action plan and course feedback |

## DELIVERY AND TOOLS

Instructor-led discussion and hands-on exercises use pre-tested, browser-accessible Python notebooks. Workflows draw on Jupyter, scikit-learn and MAPIE for regression, classification and time-series uncertainty estimation.

Recommended group size is approximately 20-24 participants. A pre-course technical check supports access to the notebooks and delivery environment.

## PARTICIPANT RESOURCES

- Course workbook or PDF handout
- Three guided notebooks
- Cleaned datasets with sources and field definitions retained
- Reusable plotting and reporting templates
- Post-course reading and practice list

**COURSE ENQUIRIES**

**enquiries@apes-energyevolution.com**

www.apes-energyevolution.com