

Machine Learning

Regression, trees and gradient boosting on real tabular data. You build models that predict honestly, measure them properly, and learn to spot the leakage that makes bad results look good.

Foundational

4 phases

21 sessions

32 contact hours

Syllabus

4 phases, 21 sessions. Each phase ends in something you have built.

1

Before You Fit Anything

Build the numeracy this course actually uses, and beat guessing without a model.

5 sessions

\$72 alone

1.1

What a Prediction Problem Is

Rows, features, a target, and framing a question someone actually has.

[FREE PREVIEW](#)

1.2

Tables in Python

Load a CSV, select rows and columns, check types, and find what is missing.

1.3

Just Enough Statistics

Mean, spread, base rate, and probability as a number you can already predict with.

1.4

Just Enough Linear Algebra

A row is a vector, a weighted sum is a dot product, and that is most of a model.

1.5

Predicting Without Machine Learning

Write a rule by hand, measure its error, and set the number every model must beat.

By the end: **A data profile and a hand-written rule with a measured error, written before any model exists.**

2

The Supervised Loop

Train a model, measure it honestly, and know when the measurement is a lie.

5 sessions

\$72 alone

2.1 Fit, Predict, Score

The three calls every scikit-learn estimator shares, run on a real dataset.

2.2 Held-Out Data and the Dummy

Train/test splits, DummyClassifier, and why 92% accuracy is often worthless.

2.3 Linear and Logistic Regression

A weighted sum, a loss, downhill, and reading what the coefficients claim.

2.4 Overfitting and Underfitting

Capacity, learning curves, regularization, and recognizing memorization on sight.

2.5 Cross-Validation and the Right Metric

Confusion matrices, precision, recall, ROC, and choosing the metric before you train.

By the end: **A cross-validated model with a written defense of its split and its metric.**

3

The Toolkit That Wins

Learn the models that actually win on tables, and compare them without cheating.

6 sessions

\$72 alone

3.1 Decision Trees

Splits you can read out loud, and watching a tree memorize its training set.

3.2 Forests and Gradient Boosting

Bagging, boosting, HistGradientBoosting, LightGBM, and why they still own large tables.

3.3 Pipelines and Leakage

ColumnTransformer, encoding, scaling, and fitting on training data only.

3.4 Feature Engineering

Dates, counts, ratios, and TF-IDF text features that beat a fancier model.

3.5 Tuning Without Fooling Yourself

Randomized search, nested validation, and spending your sealed test set once.

3.6 Tabular Foundation Models

TabPFN-3 against a tuned booster on your data, plus its size, cost, and license limits.

By the end: **A leakage-free pipeline comparing four model families under one harness on a sealed test set.**

4

Messy Data, Real Users

Handle the data you actually get, then put a model in front of people and defend it.

5 sessions

\$72 alone

4.1 When There Is No Label

k-means, PCA, anomaly detection, and what a cluster does not prove.

4.2 Imbalance, Missingness and Time

Rare classes, missing values, and splits that respect time and groups.

4.3 Explaining One Prediction

Permutation importance, partial dependence, SHAP, and an answer a non-engineer accepts.

4.4 Serving Without Skew

An API that loads the exact fitted pipeline you validated, not a retyped copy.

4.5 Drift, Fairness and the Refusal

Retraining triggers, error broken out by subgroup, and when the answer is do not ship.

By the end: **A deployed model API with a model card, a subgroup error table, and a monitoring plan.**

Tools you will use

Python

NumPy

pandas

scikit-learn

LightGBM

TabPFN-3

SHAP

FastAPI

What you will build

1 The Rule You Have to Beat

A hand-written baseline, measured before any model

2 Model Bake-Off

Logistic, forest, booster, TabPFN-3, one harness

3 Deployed Risk Scorer

Served pipeline with a model card and drift alerts

Registration

Phase 1: Before You Fit Anything	\$72
Phase 2: The Supervised Loop	\$72
Phase 3: The Toolkit That Wins	\$72
Phase 4: Messy Data, Real Users	\$72
Whole track, all 4 phases	\$240

Buying every phase separately costs \$288, so the whole track saves \$48.

No payment is taken online. Registering creates an order and payment instructions are sent with it. Access opens once payment is confirmed.