

DATA SCIENCE

ADVANCED

5 DAYS

Predictive Modeling in Data Science

Build a strong foundation in predictive modeling by understanding core concepts, algorithms, and real world applications. Learn how to preprocess data, engineer features, evaluate models, and build accurate predictive solutions using Python or R for data driven decision making.

Why Predictive Modeling?

Improved Decision-Making

Predictive models enable data-driven decision-making by analysing historical data, identifying patterns, and forecasting future outcomes. You can leverage them to anticipate risks, identify opportunities, and understand customer behavior.

Risk Management and Fraud Detection

Predictive models can help you identify potential risks and detect fraudulent activities. Fraud detection models analyse transaction patterns and anomalies to identify fraudulent activities, reducing financial losses.

Enhanced Operational Efficiency

Predictive modeling can optimise your operations by identifying bottlenecks, streamlining processes, and reducing costs.

What you'll be able to do

- ✓ Gain a clear understanding of predictive modeling, its applications, and its role in data science.
- ✓ Gain proficiency in data preprocessing techniques
- ✓ Develop knowledge of various predictive modeling algorithms
- ✓ Learn how to select relevant features and engineer new ones to enhance their models' predictive power.
- ✓ Understand model evaluation and selection.
- ✓ Gain hands-on experience in building predictive models using popular programming languages (e.g., Python or R) and relevant libraries and frameworks.

Who this is for

- Data Scientists
- Analysts
- Data Engineers
- Machine Learning Engineers

Curriculum

01 Predictive Modeling Foundations and Workflow

- Map business outcomes to predictive modeling problems
- Walk the end-to-end modeling workflow from data to decision
- Define success metrics before building models
- Identify common failure modes in production predictions

02 Data Preparation and EDA

- Clean and transform data for modeling readiness
- Explore distributions, correlations, and leakage risks
- Visualize patterns that inform feature and model choice
- **Hands-on:** Produce an EDA brief for a modeling dataset

03 Feature Selection and Engineering

- Select relevant features and create predictive signals
- Encode categorical and temporal features safely
- Avoid target leakage in engineered features
- Track feature definitions for reuse in scoring

04 Linear and Logistic Regression

- Build linear models for continuous outcomes
- Apply logistic regression for binary classification
- Interpret coefficients and odds in business language
- Diagnose underfitting and overfitting early

05 Tree-Based Models (Decision Trees, Random Forests, Gradient Boosting)

- Train decision trees for transparent baselines
- Improve performance with random forests
- Apply gradient boosting for strong tabular results
- Compare tree-family trade-offs for accuracy vs explainability

06 SVM and Neural Networks

- Use SVMs for classification and regression tasks
- Introduce neural networks for non-linear patterns
- Match model complexity to data volume and signal
- Set realistic expectations for training cost and tuning

07 Time Series Forecasting

- Structure time-dependent data for forecasting
- Apply core forecasting approaches for business horizons
- Validate forecasts with time-aware splits
- Communicate uncertainty and forecast limits

08 Evaluation, Cross-Validation, and Hyperparameter Tuning

- Choose metrics for ranking, classification, and regression
- Use cross-validation for trustworthy estimates
- Tune hyperparameters without leaking test data
- Build a model comparison scorecard

09 Interpretability and Imbalanced Data

- Explain predictions with practical interpretability methods
- Handle class imbalance with sampling and thresholding
- Align decision thresholds to business cost
- Review fairness and bias signals in model outputs

10 Ensemble Methods and Deep Learning Intro

- Combine models to improve robustness
- Understand when ensembles help vs hurt latency
- Introduce deep learning use cases for predictive work
- Decide when classical models remain the better default

11 Deployment, Monitoring, and Ethics

- Deploy models for batch or real-time scoring
- Monitor drift, performance decay, and data quality
- Plan retraining and model maintenance loops
- Apply ethics, fairness, and transparency checkpoints

12 Case Studies and Advanced Topics

- Review industry predictive modeling case studies
- Practice scoping a prediction use case end to end
- Survey advanced directions (Bayesian methods, RL overview)
- Present recommendations with risks and next steps