Datastat Research
Building Capacity in Training and Research

# DATASTAT TRAINING IN

OFFICIAL COURSE PROSPECTUS

Center for Advanced Research & Professional Development

Nairobi Training Center • Global Virtual Academy • Corporate On-Site

DATA SCIENCE

# Training Course on Ensemble Methods for Predictive Modeling

Comprehensive Course Syllabus, Practical Modules & Upcoming Schedules

## PROGRAM SPECIFICATIONS & HIGHLIGHTS

DURATION
**10 Days Practical Hands-on**

DELIVERY MODES
**In-Person Class & Live Virtual**

CERTIFICATION
**Accredited Certificate of Completion**

INTAKES / DATES
**Multiple Cohort Schedules Included**

PREPARED FOR:

**Valued Participant**

Email: training@datastatresearch.org | Issued: November 22, 2026

**Academic & Professional Training Year: 2026**

**DATASTAT TRAINING INSTITUTE**

P.O. Box 102802-00101, Nairobi, Kenya

Training Centers: Nairobi • Mombasa • Kigali • Dubai • Live Virtual Academy

Tel: +254 724 527 104 | +254 734 969 612

Email: training@datastatresearch.org

www.datastatresearch.org

# Training Course on Ensemble Methods for Predictive Modeling

**Data Science • 10 Days Intensive • Accredited Certification**

## Course Overview

Training Course on Ensemble Methods for Predictive Modeling is designed to equip participants with the theoretical foundations and practical expertise to implement and fine-tune advanced ensemble methods

### Key Skills & Software Tools Covered

Ensemble Learning • Random Forest • Gradient Boosting • Python • Scikit-learn • ML Models

## Upcoming Training Schedules

Confirmed cohort intakes and venues. Both physical in-person classroom and live virtual online seats are available.

| TRAINING DATES | LOCATION / VENUE | FORMAT / DURATION | INVESTMENT (FEE) |
|---|---|---|---|
| **5 Jan 2026 – 16 Jan 2026** | Nairobi | Classroom / 10d | **$3,000 USD / KES 220,000** |
| **5 Jan 2026 – 16 Jan 2026** | Kigali | Classroom / 10d | **$6,000 USD / KES 110,000** |
| **5 Jan 2026 – 16 Jan 2026** | Dubai | Classroom / 10d | **$10,000 USD / KES 110,000** |
| **5 Jan 2026 – 16 Jan 2026** | Cape Town | Classroom / 10d | **$9,000 USD / KES 110,000** |
| **5 Jan 2026 – 16 Jan 2026** | Addis Ababa | Classroom / 10d | **$6,000 USD / KES 110,000** |
| **5 Jan 2026 – 16 Jan 2026** | Singapore | Classroom / 10d | **$10,000 USD / KES 110,000** |
| **5 Jan 2026 – 16 Jan 2026** | Istanbul | Classroom / 10d | **$10,000 USD / KES 110,000** |
| **5 Jan 2026 – 16 Jan 2026** | Bangkok | Classroom / 10d | **$10,000 USD / KES 110,000** |
| **5 Jan 2026 – 16 Jan 2026** | Accra | Classroom / 10d | **$8,000 USD / KES 110,000** |
| **19 Jan 2026 – 30 Jan 2026** | Nairobi | Classroom / 10d | **$3,000 USD / KES 220,000** |
| **19 Jan 2026 – 30 Jan 2026** | Kigali | Classroom / 10d | **$6,000 USD / KES 110,000** |
| **19 Jan 2026 – 30 Jan 2026** | Dubai | Classroom / 10d | **$10,000 USD / KES 110,000** |

* Corporate discounts and customized cohort scheduling are available for teams of 3 or more participants.

## Course Curriculum & Modules

**Training Course on Ensemble Methods for Predictive Modeling**

**Introduction**

**In the rapidly evolving landscape of data science and machine learning, the pursuit of highly accurate and robust predictive models is paramount. Traditional single models often grapple with the inherent bias-variance trade-off, leading to either underfitting or overfitting and ultimately limiting their real-world applicability. This course delves into ensemble methods, a powerful paradigm that synergistically combines multiple individual models to overcome these limitations, yielding superior prediction performance and model generalization. By leveraging the collective intelligence of diverse learners, ensemble techniques like bagging, boosting, and stacking have become indispensable tools for data scientists and machine learning engineers seeking to build cutting-edge solutions across various industries.**

**Training Course on Ensemble Methods for Predictive Modeling is designed to equip participants with the theoretical foundations and practical expertise to implement and fine-tune advanced ensemble methods. We will explore the nuances of popular algorithms such as XGBoost and LightGBM, renowned for their efficiency and predictive power in competitive machine learning environments. Through hands-on exercises, real-world case studies, and best practices, attendees will master the art of building high-performing predictive systems, enhancing their ability to tackle complex challenges in fraud detection, risk assessment, medical diagnosis, and beyond. Prepare to transform your approach to predictive modeling and unlock new levels of accuracy and insights.**

**Course Duration**

**10 days**

## Course Objectives

- Master the foundational concepts of ensemble learning and its significance in advanced predictive analytics.
- Differentiate between bagging, boosting, and stacking techniques, understanding their underlying principles and applications.
- Implement Bootstrap Aggregating (Bagging) from scratch and apply it effectively using libraries like Scikit-learn.
- Deeply understand and leverage Random Forests for robust feature importance and variance reduction.
- Grasp the core mechanics of Gradient Boosting Machines (GBM) and its iterative error correction.
- Become proficient in using XGBoost for highly efficient and scalable gradient boosting, including hyperparameter tuning.
- Explore LightGBM as an alternative high-performance boosting framework for large datasets.
- Construct stacked generalization (stacking) models, combining diverse base learners with a meta-learner for optimal performance.
- Evaluate and compare the performance of various ensemble models using appropriate evaluation metrics (e.g., AUC, F1-score, RMSE).
- Apply cross-validation strategies to ensure robust model evaluation and prevent overfitting.
- Interpret the results of ensemble models and understand their feature importance using techniques like SHAP and LIME.
- Address common challenges in ensemble modeling, including computational complexity and model interpretability.
- Develop strategies for selecting the most appropriate ensemble method for different machine learning problems and datasets.

Organizational Benefits

- Organizations will benefit from highly accurate and reliable predictive models, leading to better decision-making across various business functions.
- Improved predictive capabilities in areas like fraud detection, risk assessment, and anomaly detection will significantly mitigate financial and operational risks.
- More precise forecasting and resource prediction enable efficient allocation of budgets, personnel, and inventory.
- Companies adopting advanced ensemble methods will gain a significant edge in their respective markets through superior data-driven insights.
- Accurate prediction of customer behavior, churn, and preferences can lead to personalized services and enhanced customer satisfaction.
- Equipping teams with these advanced techniques fosters a culture of innovation and empowers them to solve complex, real-world problems.
- Reliable predictive models provide a solid foundation for strategic planning and proactive decision-making.

Target Audience

- Data Scientists.
- Machine Learning Engineers
- Analytics Professionals.
- AI Developers.
- ResearchersStatisticians.
- Software Engineers (with ML interest)
- Business Analysts (with strong technical aptitude)

Course Outline

## Module 1: Introduction to Ensemble Learning

- What is Ensemble Learning? Why combine models?
- Benefits: Bias-Variance Trade-off, Robustness, Generalization.
- Types of Ensemble Methods: Bagging, Boosting, Stacking (Overview).
- Weak vs. Strong Learners.
- Case Study: Early challenges of single-model limitations in customer churn prediction, highlighting the need for ensemble approaches.

## Module 2: Foundations of Predictive Modeling

- Review of Supervised Learning: Regression and Classification.
- Key Concepts: Training, Validation, Test Sets, Overfitting, Underfitting.
- Performance Metrics: Accuracy, Precision, Recall, F1-Score, AUC-ROC, MSE, MAE.
- Data Preprocessing: Feature Scaling, Encoding, Missing Value Imputation.
- Case Study: Analyzing a financial dataset for loan default prediction, demonstrating the impact of data quality on model performance.

## Module 3: Bagging (Bootstrap Aggregating)

- Introduction to Bootstrapping and Aggregation.
- Bagging Algorithm Steps and Working Principle.

- Random Subsampling with Replacement.
- Advantages and Disadvantages of Bagging.
- Case Study: Bagging applied to image classification, demonstrating how multiple decision trees can reduce variance and improve accuracy on a complex image dataset.

### Module 4: Random Forests

- Building on Bagging: Decorrelating Decision Trees.
- Feature Randomness in Random Forests.
- Out-of-Bag (OOB) Error Estimation.
- Feature Importance from Random Forests.
- Case Study: Predicting housing prices using a Random Forest Regressor, emphasizing its ability to handle high-dimensional data and provide feature insights.

### Module 5: Introduction to Boosting

- Sequential Learning: Correcting Previous Model Errors.
- Concept of Weak Learners and Adaptive Learning.
- Boosting vs. Bagging: Key Differences.
- Additive Models and Gradient Descent.
- Case Study: Adopting boosting for a medical diagnosis problem where misclassifications have high costs, illustrating the focus on difficult examples.

### Module 6: AdaBoost (Adaptive Boosting)

- Understanding AdaBoost Algorithm: Weighting Samples and Models.
- Iterative Learning and Error Minimization.
- Strengths and Limitations of AdaBoost.
- Implementation with Scikit-learn.
- Case Study: Applying AdaBoost to a spam email detection dataset, showing how it iteratively improves classification on mislabeled emails.

### Module 7: Gradient Boosting Machines (GBM)

- The Essence of Gradient Boosting: Gradient Descent for Residuals.
- Loss Functions in GBM.
- Shrinkage and Subsampling for Regularization.
- GBM Hyperparameters and Tuning.
- Case Study: Predicting customer lifetime value (CLTV) using GBM, showcasing its effectiveness in regression tasks with complex relationships.

### Module 8: XGBoost (Extreme Gradient Boosting) - Part 1

- Introduction to XGBoost: Optimized GBM Implementation.
- Key Features: Regularization (L1 & L2), Parallel Processing.
- Handling Missing Values and Sparse Data.
- Tree Construction and Splitting in XGBoost.
- Case Study: Competing in a Kaggle-style tabular data prediction challenge (e.g., predicting sales), demonstrating XGBoost's speed and accuracy.

### Module 9: XGBoost (Extreme Gradient Boosting) - Part 2

- Advanced XGBoost Hyperparameter Tuning Strategies.
- Early Stopping and Cross-Validation within XGBoost.
- Feature Importance Interpretation with XGBoost.
- Building Classification and Regression Models.
- Case Study: Fraud detection in banking transactions, leveraging XGBoost for high recall on rare fraudulent activities.

### Module 10: LightGBM (Light Gradient Boosting Machine)

- Introduction to LightGBM: Leaf-wise Tree Growth.
- Comparison with XGBoost: Speed and Memory Efficiency.
- Categorical Feature Handling in LightGBM.
- LightGBM Hyperparameters and Optimization.
- Case Study: Predicting server failures in a large IT infrastructure, highlighting LightGBM's efficiency with massive datasets.

### Module 11: Stacking (Stacked Generalization)

- Concept of Stacking: Combining Diverse Base Models.
- Level-0 Models (Base Learners) and Level-1 Model (Meta-Learner).
- Training Stacking Models with Cross-Validation.
- Choosing Base Models and Meta-Learner.
- Case Study: Developing a robust recommendation system by stacking various collaborative filtering and content-based models to predict user preferences.

### Module 12: Advanced Stacking Techniques

- Blending vs. Stacking.
- Multi-Layer Stacking Architectures.
- Ensemble Diversity Strategies in Stacking.
- Practical Considerations for Stacking Complexity.
- Case Study: Improving credit risk scoring by stacking logistic regression, decision trees, and neural networks, leading to a more reliable risk assessment.

### Module 13: Model Evaluation and Comparison

- Advanced Metrics for Imbalanced Datasets (e.g., Precision-Recall Curve).
- Statistical Significance Testing for Model Comparison.
- Visualizing Model Performance: Confusion Matrices, ROC Curves.
- Techniques for Model Interpretability (SHAP, LIME).
- Case Study: Evaluating and comparing different ensemble models for predicting disease outbreaks, focusing on robust performance metrics and interpretability for public health officials.

### Module 14: Regularization and Overfitting in Ensembles

- Understanding Regularization in Ensemble Methods.
- Techniques to Prevent Overfitting (e.g., Subsampling, Shrinkage).
- The Role of Hyperparameter Tuning in Regularization.
- Bias-Variance Decomposition Revisited for Ensembles.
- Case Study: Optimizing a stock market prediction model to avoid overfitting to historical data, ensuring generalization to unseen market conditions.

### Module 15: Deployment and Best Practices for Ensemble Models

- Model Serialization and Deployment Strategies.
- Monitoring Ensemble Model Performance in Production.
- Ethical Considerations in Ensemble Modeling (Fairness, Transparency).
- Workflow for Building and Maintaining Ensemble Systems.
- Case Study: Deploying an ensemble model for real-time traffic prediction in a smart city initiative, addressing latency and continuous model updates.

Training Methodology

This training course employs a blended learning approach, combining theoretical instruction with extensive hands-on practice.

- Interactive Lectures: Engaging presentations explaining complex concepts with clear examples.
- Live Coding Sessions: Demonstrations of implementation using Python and popular machine learning libraries (Scikit-learn, XGBoost, LightGBM).
- Hands-on Labs & Exercises: Practical sessions where participants apply learned concepts to real datasets.
- Case Study Analysis: In-depth examination of real-world business problems and how ensemble methods provide solutions.
- Group Discussions: Fostering collaborative learning and problem-solving.
- Q&A Sessions: Dedicated time for addressing participant queries and clarifying concepts.
- Project-Based Learning: A final project where participants build and evaluate their own ensemble model.
- Cloud-Based Lab Environment: Access to pre-configured environments (e.g., Jupyter Notebooks, Google Colab) to facilitate seamless coding and experimentation.

Register as a group from 3 participants for a Discount

Send us an email: info@datastatresearch.org or call +254724527104

Certification

Upon successful completion of this training, participants will be issued with a globally- recognized certificate.

Tailor-Made Course

We also offer tailor-made courses based on your needs.

Key Notes

a. The participant must be conversant with English.

b. Upon completion of training the participant will be issued with an Authorized Training Certificate

c. Course duration is flexible and the contents can be modified to fit any number of days.

d. The course fee includes facilitation training materials, 2 coffee breaks, buffet lunch and A Certificate upon successful completion of Training.

e. One-year post-training support Consultation and Coaching provided after the course.

<span style="font-size: 12.0pt; line-height: 150%; font

## Certification & Learning Methodology

- Hands-on practical delivery with real-world datasets, case studies, and lab exercises.
- Post-training mentorship, access to class resources, and customized follow-up support.
- Participants who successfully complete the course receive an accredited Datastat Certificate of Completion.

## How to Enroll & Reserve Your Seat

1. Online Registration: Visit https://datastatresearch.org/courses/training-course-on-ensemble-methods-for-predictive-modeling to register online.
2. Email Booking: Send participant details to training@datastatresearch.org for official invoice & reservation.
3. Support & Assistance: Inquire via phone +254 724 527 104 | +254 734 969 612.