L04: Random Forests
Ensemble Learning for Robust Predictions

Methods and Algorithms – MSc Data Science

By the end of this lecture, you will be able to:

1. Explain how decision trees partition feature space
2. Implement Random Forests using bagging and feature randomization
3. Interpret feature importance and out-of-bag error
4. Apply ensemble methods to fraud detection problems

Finance Application: Fraud detection with interpretable feature importance

From single models to ensemble methods that combine many weak learners

The Business Problem

Fraud Detection Challenge
- Need high accuracy: fraudulent transactions cost millions
- Need interpretability: explain why transaction flagged
- Complex patterns: fraud evolves and adapts

Why Random Forests?
- Combines many trees for robust predictions
- Built-in feature importance ranking
- Handles non-linear relationships naturally

Ensemble methods: "wisdom of crowds" for machine learning

01_decision_tree/chart.pdf

02_feature_importance/chart.pdf

03_bootstrap/chart.pdf

04_oob_error/chart.pdf

05_ensemble_voting/chart.pdf

06_bias_variance/chart.pdf

07_decision_flowchart/chart.pdf