Random Forest — Deep Dive for AI Engineers

1. Why Random Forest Exists
Random Forest addresses overfitting in noisy, non-linear real-world data by averaging multiple decision trees, achieving better generalization.

2. Decision Trees as Building Blocks
Decision trees split data using simple rules. They are powerful but unstable when used alone.

3. The Problem with Single Trees
Single trees have high variance and are sensitive to noise and outliers.

4. Core Idea of Random Forest
Random Forest builds many trees using random subsets of data and features, then averages predictions.

5. Randomness in Random Forest
Randomness comes from bootstrapped rows and random feature selection at each split.

6. Bias–Variance Tradeoff
Random Forest achieves low bias and low variance, making it robust for production.

7. Key Hyperparameters
Important parameters include number of trees, depth, minimum samples per leaf, and features per split.

8. Out-of-Bag Evaluation
Unused samples per tree (OOB) provide built-in validation.

9. Feature Importance
Feature importance shows relative impact but should be interpreted cautiously.

10. Limitations
Random Forest cannot extrapolate trends and is best for short-term forecasting.

11. Business Fit
It works extremely well on noisy, tabular ERP and operational data.

12. Forecasting Usage
With lag and rolling features, RF models time-dependent behavior effectively.

13. Interview Summary
Random Forest is an ensemble of decision trees that reduces variance via averaging.

14. Key Takeaway
Random Forest is reliable, stable, and production-proven.