🌲 OOB Score | Out-of-Bag Evaluation in Random Forest

Random Forest uses **Bootstrap Sampling** → training each tree on a random subset of the dataset **with replacement**. This means:

- Some samples are **included** in training for a tree.
- Some samples are **left out** → these are called **Out-of-Bag (OOB) samples**.

---

1. What is OOB Evaluation?

- Each tree in the forest sees only ~63% of the dataset (because of sampling with replacement).
- The remaining ~37% of data is **unused for that tree's training** → OOB samples.
- These OOB samples act like a **built-in validation set**.

---

2. How OOB Score Works

- For every training sample:
  1. Look at predictions from only the trees where that sample was **not included** (OOB for those trees).
  2. Aggregate those predictions (majority vote for classification, average for regression).
- Compare aggregated prediction with true label → this gives an **error estimate**.

---

3. Why OOB is Powerful

- **No need for a separate validation set** → you save data.
- Gives a performance estimate **during training**.
- Reduces computation compared to cross-validation.
- Works well for large datasets where cross-validation would be costly.

---

4. OOB Score vs. Test Set

- OOB score is like **internal cross-validation**.
- Still, you need a **final test set** to report unbiased performance.
- Think of OOB as a **quick estimate** during training, not a replacement for test evaluation.

🎯 Quick Intuition Recap

- Random Forest trains on bootstrapped subsets.
- Left-out samples (OOB) act like **free validation points**.
- OOB score ≈ **cross-validation** but cheaper.
- Use OOB during training, test set for final evaluation.
