🌲 Feature Importance using Decision Trees & Random Forests

Feature importance tells us **which features (columns) in the dataset contribute the most** to making predictions. Both **Decision Trees** and **Random Forests** can compute this automatically.

⌛ Time-Stamped Concept Map

- **What is Feature Importance?**
  - It measures how much a feature helps reduce uncertainty (impurity) in the dataset.
  - In trees, every split tries to reduce impurity (Entropy, Gini, or Variance for regression).
  - A feature is **important** if splits on it **consistently reduce impurity a lot**.

Feature Importance Documentation

- `feature_importances_` in sklearn =

$$\text{Importance}(f) = \frac{\sum(\text{Impurity Decrease at node using } f)}{\text{Total Impurity Decrease across all features}}$$

Calculating Importance using Decision Trees

- For each split:
  1. Calculate **parent impurity** (Entropy/Gini/Variance).
  2. Calculate **weighted child impurity**.
  3. **Impurity Decrease** = Parent − Weighted Child.
- Feature importance = **sum of impurity decrease** over all nodes where that feature is used.

Example:

- If splitting on Age reduces Gini impurity by 0.15, and Salary reduces it by 0.05, then **Age is 3x more important**.

Calculating Importance using Random Forest

- Random Forest builds many trees.
- Each tree calculates feature importance as above.
- Final importance = **average across all trees**.

- More robust than a single Decision Tree because it reduces bias toward noisy features.

🎯 Key Insights

- **Decision Trees**: Feature importance = impurity reduction per feature.
- **Random Forests**: Average importance across many trees (more stable, less variance).
- **Higher score** → feature is more influential.
- **Caution**: Importance can be biased toward features with more categories or higher variance.

⚡ Quick Example (Intuition)

Dataset: Predict whether someone buys a product.

- Features: Age, Income, City.
- Tree splits mostly on Income, sometimes on Age, almost never on City.

Feature Importance might look like:

- Income → **0.70**
- Age → **0.25**
- City → **0.05**

So, **Income** is the strongest predictor.

✓ Now, whenever you revisit this, just think:

- Trees measure **how much impurity each feature reduces**.
- Random Forest **averages this across many trees**.
