4- Supervised Classification Models - Random Forest

أكاديمية طويق
Tuwaiq Academy

By: eng. Esraa Madhi

https://mlu-explain.github.io/random-forest/

Random Forest •

The Random Forest algorithm offers several advantages over a single Decision Tree, making it a more robust and accurate choice for many machine learning tasks. Here are some of the key benefits:

1. Improved Accuracy

Random Forests improve prediction accuracy by combining the predictions from multiple decision trees (each built on a random subset of the data and features). This ensemble approach generally leads to better performance and more reliable predictions compared to a single decision tree.

2. Overfitting Reduction

A common problem with decision trees, especially those that are deep, is overfitting—they can become too tailored to the training data, capturing noise along with the signal. Random Forest mitigates this by averaging multiple trees, which by their construction (each tree built on a random subset of data) are less likely to overfit.

3. Handling Unbalanced Data

Random Forest can handle unbalanced data sets by balancing itself to some extent. In contrast, a single decision tree might develop a bias toward the classes that are predominant.

4. Robustness to Noise

The ensemble approach of Random Forest makes it more robust to noise in the dataset. Individual decision trees might be sensitive to noise, but averaging across multiple trees reduces the effect of noisy data points.

5. Feature Importance

Random Forests naturally provide insights into feature importance—how much each feature contributes to the prediction accuracy. This is derived during the construction of the trees and is valuable for understanding the data and the model.