

Understanding the Difference between Decision Tree and Random Forest Algorithms

Machine learning is a branch of artificial intelligence that focuses on developing models capable of learning patterns from data to make predictions or decisions. Supervised learning techniques such as Decision Trees and Random Forests are widely used for this purpose [1].

A Decision Tree is a tree-structured model that recursively splits data based on feature values to classify instances, making it simple, interpretable, and effective for small datasets; however, it is prone to overfitting and performance degradation when applied to large datasets [2].

Random Forest extends Decision Trees by constructing an ensemble of multiple unpruned trees using bagging and random feature selection, where final predictions are obtained through majority voting. This approach significantly improves classification accuracy, robustness, and generalization while reducing overfitting [2].

While Decision Trees are computationally efficient and easy to interpret, their instability limits performance on complex datasets. In contrast, Random Forest provides superior results for large and high-dimensional datasets, albeit at the cost of increased computational complexity. In practice, Decision Trees are preferred for smaller datasets and applications requiring transparency, whereas Random Forest is more suitable for large-scale classification problems where high accuracy and reliability are essential [1], [2].

References

[1] T. M. Mitchell, Machine Learning, New York, NY, USA: McGraw-Hill, 1997.

2] J. Ali, R. Khan, N. Ahmad, and I. Maqsood, "Random Forests and Decision Trees," International Journal of Computer Science Issues, vol. 9, no. 5, no. 3, pp. 272-278, Sep. 2012.

