

Title: Understanding the Difference between Decision Tree and Random Forest algorithms.

Decision Tree:

Decision tree is a classical machine learning algorithm, which is widely used because of its intuitive and clear model. It has been widely used in classification and other fields of data mining, since it was proposed in the 1960s. DT algorithms are playing an increasingly important role with the development of machine learning, while the traditional DT is facing unprecedented challenges with the innovation of branch rules and the selection of variable [1].

The learning procedure of decision trees involves a series of steps where the data is split into homogenous subsets. There is a root node, which is the starting point of the tree, represents the entire dataset. The algorithm identifies the feature and the threshold that leads to the best split based on a specific criterion. The process continues recursively, with each subset of the data being further split at each child node. This continues until a stopping criterion is reached, typically when the nodes are pure (i.e., all data points in a node belong to the same class) or when a predefined depth of the tree is reached. The nodes where the tree ends, called leaf node or terminal node, represent the outcomes or class labels. The decision to split at each node is made using mathematical formulations such as information gain, Gini impurity, or variance reduction [2].

Random Forest:

The RF model consists of an ensemble classifier that produces many decision trees through the use of a randomly selected subset of samples and training variables. The RF model has assumed importance within the scientific community thanks to its performance. The accuracy of its classifications and prediction has allowed the use of RF in several research domains, which have benefited from it. The RF is a user friendly, intuitive, and very fast approach. It less training time than Decision Tree and Support Vector Machine. RF becomes functional in cases of large datasets because it helps avoid the problem of overfitting by managing the noises present in datasets [3].

Difference:

Random forest models can provide good prediction models due to their efficacy and sensitivity and specificity. Whereas, decision tree can be visualized as a single tree compare to Random Forest where there are multiple trees.[4].

REFERENCES:

- [1] S. R. Jiao, J. Song, and B. Liu, "A Review of Decision Tree Classification Algorithms for Continuous Variables," *Journal of Physics: Conference Series*, vol. 1651, no. 1, p. 012083, 2020, doi: 10.1088/1742-6596/1651/1/012083.
- [2] I. D. Mienye and N. Jere, "A Survey of Decision Trees: Concepts, Algorithms, and Applications," in *IEEE Access*, vol. 12, pp. 86716-86727, 2024, doi: 10.1109/ACCESS.2024.3416838.
- [3] M. Aria, A. Gnasso, L. D'Aniello, *et al.*, "Twenty Years of Random Forest: Preliminary results of a systematic literature review," in *IES 2022 Innovation & Society 5.0: Statistical and Economic Methodologies for Quality Assessment*, PKE srl, 2022, pp. 225–230.
- [4] H. Esmaily, M. Tayefi, H. Doosti, M. Ghayour-Mobarhan, H. Nezami, and A. R. Amirabadizadeh, "A Comparison between Decision Tree and Random Forest in Determining the Risk Factors Associated with Type 2 Diabetes," *J. Res. Health Sci.*, vol. 18, no. 2, p. e00412, 2018.