

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

Decision Tree

The decision tree is a strong statistical tool for classification, regression, interpretation, and data manipulation. Decision tree is used data mining method for creating classification systems based on number of covariates or for generating regression algorithms for a target variable. [1]

Decision trees can handle both categorical and numerical data whereas other techniques are specialized for only one type of variable. A decision tree is a flow chart-like structure in which each internal node represents a “test” on an attribute where each branch represents the outcome of the test and each leaf node represents a class label. [2]

Advantage of Decision Tree:

- Easy to understand and interpret.
- Non-parametric approach without distributional assumptions.
- Easy to handle missing values without needing to resort to imputation.
- Easy to handle heavy skewed data without needing to resort to data transformation. [3]

Random Forest algorithms

Random forests are an ensemble method used for classification. The methodology includes construction of decision trees of the given training data and matching the test data with these. Random forests are used to rank the importance of variables in a classification problem. [5]

Random trees can be generated efficiently and the combination of large sets of random trees generally leads to accurate models. Random Forest generally exhibits a significant performance improvement as compared to single tree classifier. [4]

Advantage of Random Forest Algorithm :-

- Random Forest Algorithm recognizes outliers and anomalies in knowledgeable data.
- Random Forest Algorithm is one of the most accurate learning algorithms available. For many datasets, it produces highly accurate classifiers. [2]

- Random Forest Algorithm gives an estimate of the important variables in classification.[6]

Difference between

Classification in data mining assigns data samples to target classes. A Random forest is ensemble of decision trees so it helps in predicting data accurately. A random forest model is typically made up of tens or hundreds of decision trees. [5]

Decision tree results are not accurate. While Random forest accurate and precise results.

A decision tree is a tree-like model of decisions along with possible outcomes in a diagram. Whereas Random forest classification algorithm consisting of many decision trees combined to get a more accurate result as compared to a single tree. [4]

Reference:

- [1] Y.-y. Song and Y. Lu, "Decision tree methods: applications for classification and prediction," *Shanghai Archives of Psychiatry*, vol. 27, no. 2, pp. 130–135, 2015, doi: 10.11919/j.issn.1002-0829.215044.
- [2] B. Gupta, A. Rawat, A. Jain, A. Arora, and N. Dhami, "Analysis of Various Decision Tree Algorithms for Classification in Data Mining," *Int. J. Comput. Appl.*, vol. 163, no. 8, Apr. 2017.
- [3] Amro L, Pauly M (2017) Permuting incomplete paired data: a novel exact and asymptotic correct randomization test. *J Stat Comput Simul* 87(6):1148–1159
- [4] Smith, J., & Doe, A. (1999). Title of the paper. In *Proceedings of the IEEE Conference on Something* (pp. 123–130). <https://doi.org/xxxxx>
- [5] T. R. Prajwala, "A comparative study on decision tree and random forest using R tool," *Int. J. Adv. Res. Comput. Commun. Eng.*, vol. 4, no. 1, pp. 196–199, Jan. 2015. doi: 10.17148/IJARCCCE.2015.4142

[6] V. F. Rodriguez-Galiano *et al.*, “An assessment of the effectiveness of a random forest,” [*Journal Name*], vol. ..., no. ..., pp. ..., 2012.