

Recap – Random Forest

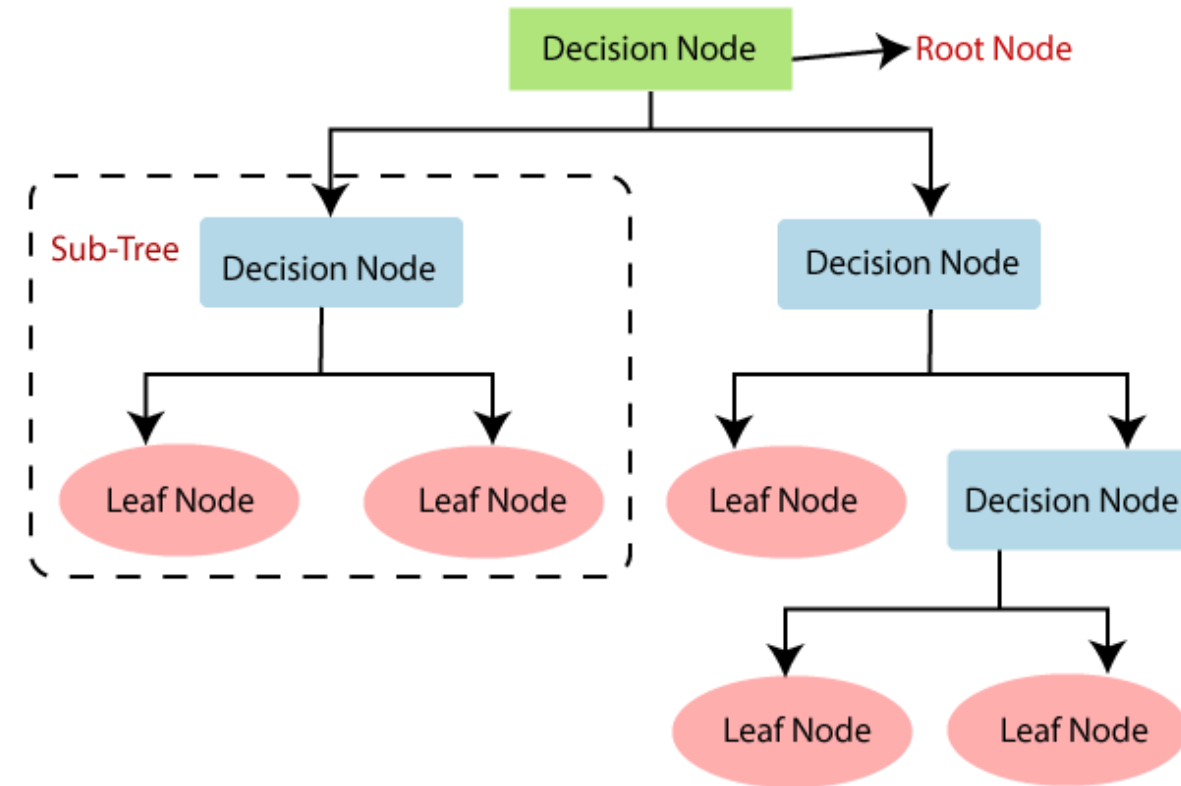
Rina BUOY



AMERICAN UNIVERSITY
OF PHNOM PENH
STUDY LOCALLY. LIVE GLOBALLY.

Decision Tree

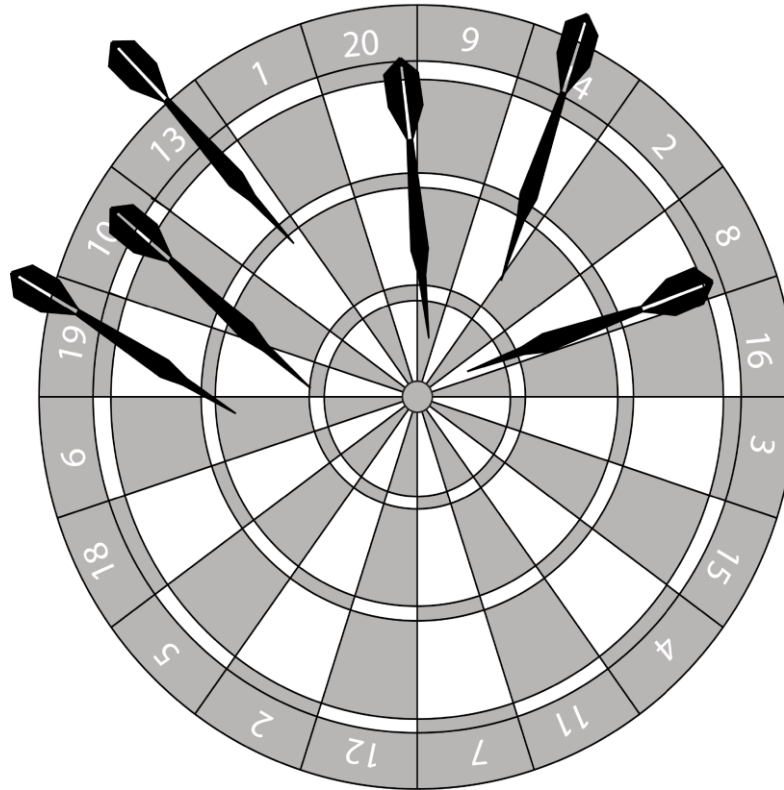
Outlook	Temperature	Humidity	Windy	Play Tennis?
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No



A LG model will require multiple parameters, while a DT requires multiple branches.

Deep Tree => Accuracy => High Variance

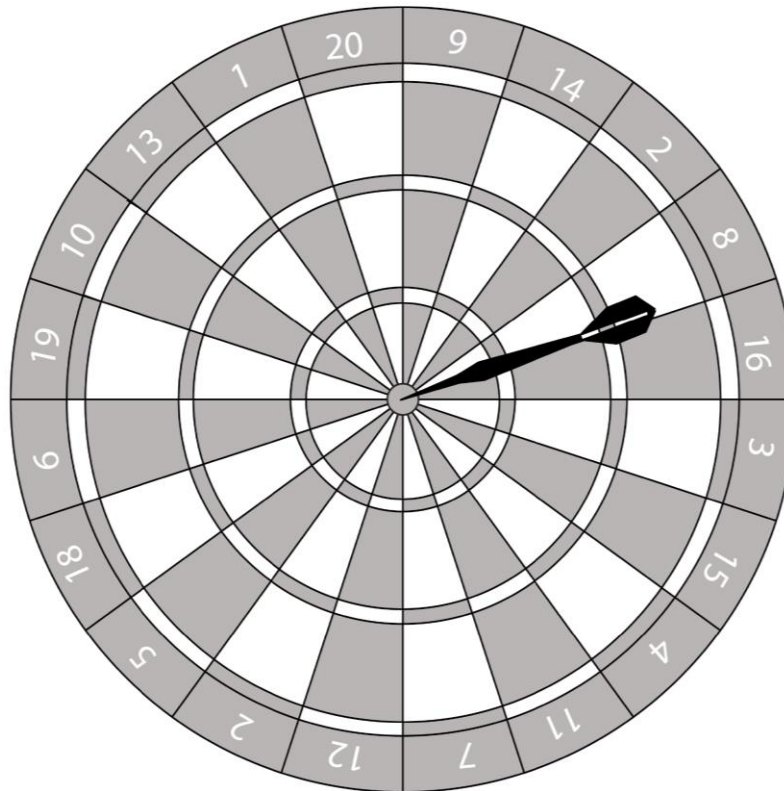
Decision Trees



Solution : Many Shallow Trees

- A shallow tree => high bias and almost no variance
- Many shallow trees => low bias, low variance (compared with a deep tree)

Random Forests



Bootstrapping

N = 3

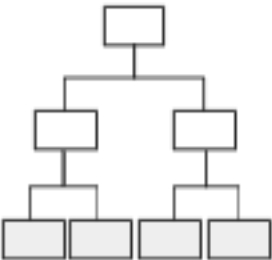
	debt	assets	price	status
0	500	2500	1250	OK
1	250	4500	1500	OK
2	500	2500	1250	OK
3	1000	4000	4500	default

Original Dataset

N = 1
sample(features, 2)

Sample 1

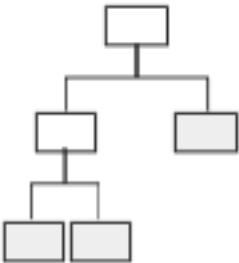
	debt	assets	status
0	500	2500	OK
1	250	4500	OK
2	500	2500	OK
3	1000	4000	default



N = 2
sample(features, 2)

Sample 2

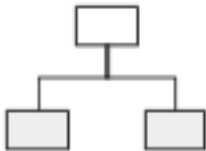
	debt	price	status
0	500	1250	OK
1	250	1500	OK
2	500	1250	OK
3	1000	4500	default



N = 3
sample(features, 2)

Sample 3

	assets	price	status
0	2500	1250	OK
1	4500	1500	OK
2	2500	1250	OK
3	4000	4500	default



Important Hyper Parameters

Number of decision trees

- Specifies the number of independent decision trees in your ensemble.
- Higher value usually result in better / more stable predictions, since errors average out.

Maximum depth of trees

- Specifies the maximum depth a tree in the ensemble can have.
- Rule of thumb: Deeper trees give better accuracy but increase the risk of overfitting.

Minimum leaf size

- Determines the smallest size of a leaf node in the ensemble.
- Too many leaves can cause overfitting and poor model generalization.