

## Test AD & MC

- ① Small K with Noisy data
- ② Small changes in data lead to different trees.
- ③ By Reducing variance
- ④ All features are considered at each split
- ⑤ Target variable is categorical
- ⑥ Sigmoid
- ⑦ Accuracy
- ⑧ Overfitting Training time
- ⑨ Because distance calculation depend on scale
- ⑩ Logistic Regression

Ans) when tree needs to grow deep. then the Depth as parameter is used in overfitting in decision tree. A tree learn all the training data and memorize it. They can learn noisy data as well. They have many nodes and branches because tree grow deep.

They won't memorize new data because they already read and memorize training data.

A3.2 Bagging Address this problem by creating multiple trees on some training data

Random forest Address this problem by Adding layer as on Bagging Randomness.

A3 Random forest working in detail:-

- Bootstrap sampling:- Bootstrap sampling is also known as 'Bagging'. They create multiple trees of training data to reduce overfitting on decision tree.
- Random feature selection:- for each tree which is created by bootstrap sampling a random subset of feature is select.
- Majority Voting:- In Majority voting ~~they~~ this resume find result or prediction of data.

## ~~Fraud~~ Fraud Detection Model

	Detected Fraud	Detected Not Fraud
Actual Fraud	120	30
Actual Not Fraud	50	800

- Calculate Accuracy

- 11 11 Precision

- 11 11 Recall

- 1111 F1 Score

- Is this Model Acceptable for Fraud Detection?

$$TP = 120$$

$$FN = 30$$

$$FP = 50$$

$$TN = \frac{800}{1000}$$

① Accuracy :-  $A = \frac{TP + TN}{\text{Total}} = \frac{120 + 800}{1000} = \frac{920}{1000} = 0.92$

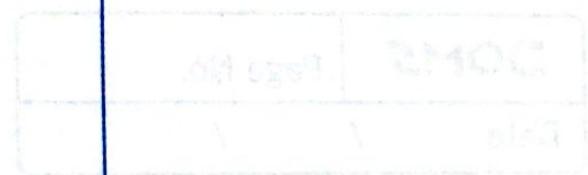
0.92%

② Precision :-  $P = \frac{TP}{TP + FP} = \frac{120}{120 + 50} = \frac{120}{170} = 0.70$

③ Recall =  $\frac{TP}{TP + FN} = \frac{120}{120 + 30} = \frac{120}{150} = 0.8$

④ F1 Score =  $\frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$

$$F1 = \frac{2 \times 0.70 \times 0.8}{0.70 + 0.8} = \frac{1.12}{1.5} = 0.74$$



DOMS	Page No.
Date	/ /

- Yes model is acceptable for fraud detection.