Assignment - 7

1. (b) - Small K with noisy data

2. (b) - Small changes in data lead to different trees.

3. (c) - Reducing variance

4. (c) - All features are considered at each split.

5. (a) - Target variable is categorical

6. (c) - Sigmoid

7. (c) - Accuracy

8. (d) - overfitting

9. (c) - Because distance calulation depends on scale

10. (c) - Logistic Regression.

11. overfitting in Decision Trees using depth as a parameter.

Decision tree splits the data again and again to make decisions. if the tree depth small, the model is simple.

if the tree depth is large the tree is noise and small details

how Bagging & Random forest address this problem differently

Bagging address this problem

Data is sampled randomly with replacement.
Each tree is trained on a slightly different dataset.
final result is decided by voting.

Random forest

It is also uses different data samples. And Also uses random features at each split

Random Forest working in detail including

- Bootstrap Sampling.
- Random feature selection.
- Majority voting.

Random forest is a esemble learning algorithm that builds many decision tree and Combines their result to make a final prediction.

① Bootstrap Sampling :- Its from the original dataset, multiple new dataset are created. Data is selected randomly with replacement. Each new dataset is called a bootstrap sample.

② Random feature Selection:- when a tree is spliting a Node, it does not use all feature. thus, they selects a random subset of feature.

③ Majority voting:- In the majority voting it is the final result or prediction. The class with the maximum votes becomes the final output

⑫ A fraud detection model produced the following results:

	Predicted Fraud	Predicted Not Fraud
Actual fraud	120	30
Actual Not fraud	50	800

a) Calculate Accuracy
b) Calculate Precision
c) Calculate Recall
d) Calculate F1 score
e) IS this model acceptable for fraud detection? Justify your Answer.

	Predicted Fraud	Predicted Not Fra
Actual fraud	120	30
Actual Not fraud	50	800

TP = 120
FN = 30
FP = 50
TN = 800
Total = 1000

a) Accuracy

$$A = \frac{TP + TN}{Total} = \frac{120 + 800}{1000} = \frac{920}{1000} = 0.92$$

$= 92\%$

Precision :-

(b) $P = \frac{TP}{TP + FP} = \frac{120}{120 + 50} = \frac{120}{170} = 0.706$

(70.6%)

Page No.

Date

(c) Recall :- $\frac{TP}{TP + FN} = \frac{120}{120+30} = \frac{120}{150} = 0.8$

$= (80\%)$

(d) F1 Score :- $\frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$

$F1 = \frac{2 \times 0.706 \times 0.8}{0.706 + 0.8} = 0.75$

(e) It's Acceptable for fraud detection.

Because :- Accuracy is high (92%)

Recall is 80% → 20% fraud case are missed.