1) a) Large K with noisy data

2) a) They cannot handle missing value and ~~some~~ from ~~for~~ large dataset they overfit

3) d) Improving feature selection

4) b) Features are ignored

5) a) Target variable is categorical

6) c) Sigmoid

7) b) Recall

8) c) Training time

9) a) To reduce bias

10) c) Logistic Regression

11) Decision Tree →

- Decision tree is used in supervised learning.
- It is a heirarical structure.
- It include root node, Parent node, Leaf node

Root node → First node that split the tree
Parent node → node used to split the tree
Leaf node → Last node that predict the output

Now,

[Overfitting in Decision Tree using depth as a parameter]

- Decision tree is a hairical structure some time it become overfit, by the increasing depth of the tree. because during the training time it ~~become~~ Learn the position of the data and give us high accuracy but from the testing time. It can't handel the new data and become overfit.

means increasing the size of data having high - chance of overfiting.

for solving that problem of overfiting we use Bagging and Random forest.

Bagging → Bagging is a technique of preventing model from overfiting by training multiple model from same data set and comparing their results.

Random forest → Random forest is a assembling technique it combine multiple decision trees and give us best result by combining multiple decision tree the size of data (depth of data) is divided into multiple decision

Trees and over model is solve the problem of overfiting

→ Random forest :→

Working :→

- Random forest is assembling technique first it divide the dataset into different decisiont trees

- It use Random feature selection means the data set is divided randomly b/w the differen decision tree.

- Every decision tree perform their operation like find the entropy and calcalate the information gain

  Entropy = $-\Sigma p_i \log_2(p_i)$

  IG = Entropy of parent − weight intropy of child

- At last they combine each the decision trees and give the result by majority voting.

  Majority voting → Count the result of each decision tree and give that result that are mostly given by the model

~~1) In KNN, which situation will mos~~

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

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

a) Accuracy = $\frac{TP+FP}{TP+TN+FP+FN}$

$= \frac{120+(50)}{120+30+50+800}$

$= \frac{170}{1000}$

b) Precision = $\frac{TP}{TP+FP}$

$= \frac{120}{120+50} = \frac{120}{170}$

c) Recall = $\frac{TP}{TP+FN} = \frac{120}{120+800} = \frac{120}{920}$

d) F_1 = $\frac{\text{Precision}}{\text{Recall}}$

$= \frac{12}{17} \times \frac{92}{12} = \frac{92}{17}$

e) Yes

M T W T F S S

Page No.:

Date:

YOUVA STELLAR

13) a) Features are ignored

b) Yes the model overfit because the feature depth of the model = none

c) Biase - high

Variance - high.