Algorithms Used:

1. Decision Tree:

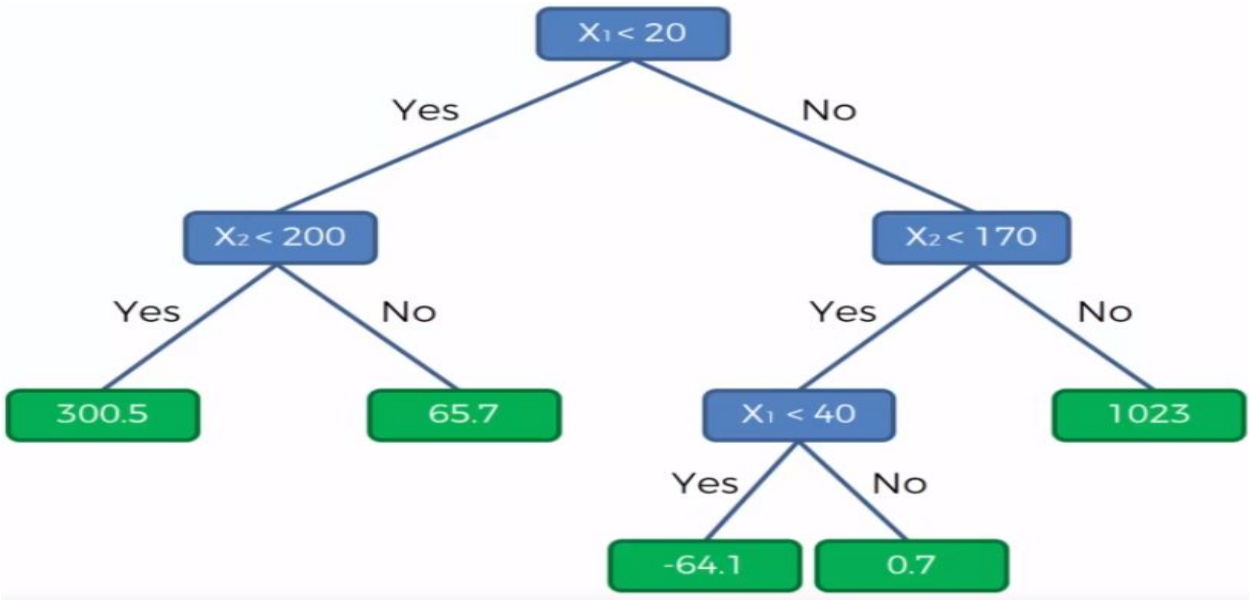
Dependent variable(Y) : Categorical
Independent variable(X): Categorical and Continous
Used for classification and prediction.
Supervised Learning.

Decision Tree is a classification technique used to classify records in a pictorial format.
In Decision Trees, for predicting a class label for a record we start from the **root** of the tree. We compare the values of the root attribute with the record's attribute. On the basis of comparison, we follow the branch corresponding to that value and jump to the next node.

Decision trees use multiple algorithms to decide to split a node into two or more sub-nodes. The creation of sub-nodes increases the homogeneity of resultant sub-nodes. In other words, we can say that the purity of the node increases with respect to the target variable.

Attribute Selection Measures:
Gini Index, Entropy
Gini index: a cost function used to evaluate splits in the dataset. It is calculated by subtracting the sum of the squared probabilities of each class from one.
(Relative Probability of class j at node t: summation over j p(j/t))

$$Gini = 1 - \sum_{i=1}^{C} (p_i)^2$$



Feature selection using Random forest
OOB Out of Bag Score: This metric is the accuracy of examples xi using all the trees in the random forest ensemble for which it was omitted during training.
Use of forests of trees to evaluate the importance of features on an artificial classification task.

Xgboost:
It works on gradient boosting algorithm
Gradient boosting algorithm works on the basic principle gradient descent.
This model is built using tree-based learners(Decision Trees)
Boosting works on the principle of ensemble techniques where errors from earlier models are reduced by the new models
XGradient boosting Algorithm:

Final prediction=Base value(the starting prediction from basic decision tree)+LR*w1+LR*w2+..+LR*wn
Where LR= learning rate=eta
w1=residual predicted value by 1st residual model
wn=residual predicted value by nth residual model
Xgboost is different from other gradient boost is because of its tuning parameters
The main tuning parameters are
1)regularisation parameter(Lambda)

2)threshold that defines auto pruning (Gamma)
3)Learning rate(eta)

We are predicted based on similarity score, gain of branch
similarity score=(sum of residuals)^2/(Lambda+no of residuals)
Gain=(total SS)afternode split-(SS)before node split
If Gain>Gamma then only further splitting occurs else it stops and gives predictions(Wi)
After calculating Wi from the residual model