1. ID3 Algorithm:-

*. A well-known decision tree approach for machine learning is the Iterative Dichomiser 3 (ID3) algorithm.

*. By choosing the best characteristic at each node to paritition the data depending on information gain, it recursively constructs a tree.

*. The goal is to make the final subsets as homogeneous as possible.

*. By choosing features that offer the greatest reduction in entropy or uncertainty ID3 iteratively grows the tree.

*. The procedure keeps going until a halting requirement is satisfied, like a minimum subset size of a maximum tree depth.

How ID3 works:-

The ID3 algorithm is specifically designed for building decision trees from a given dataset. Its primary objective is to construct a tree that best explains the relationship between attributes in the data and in the data and their corresponding class labels.

1. selecting the Best attribute:-

*. ID3 employs the concept and information gain to determine the attribute that best separates data. Entropy measures the impurity or randomness in the dataset.

2. creating Tree Nodes:-

*. The chosen attribute is used to split the dataset into subsets based on its distinct values.

*. For each subset, ID3 recurses to find the next best attribute to further partition the data, forming branches and new nodes accordingly.

3. Stopping criteria:-

*. The recursion continues until one of the stopping criteria is met, such as when all instances in a branch belong to the same class or when all attributes have been used for splitting

4. Handling Missing values:-

*. ID3 can handle missing values by employing various strategies like attribute mean/mode substitution or using majority class values.

5. Tree pruning:-

*. Pruning is a technique to prevent overfitting. while not directly included in ID3, post-processing techniques or variations like C4.5 incorporate pruning to improve the tree's generalization

Mathematical concepts of ID3 Algorithms:-

* Now let's examine the formula linked to the main theoretical ideas in the ID3 algorithm.

1. Entropy:-

A measure of disorder or uncertainty in a set of data is called entropy. Entropy is tool in ID3 to measure of dataset's disorder or impurity. by dividing data into as homogenous subsets as feasible, the objective is to minimize entropy

$$\text{Entropy}(S) = \sum_{i=1}^{c} -p_i \log_2(p_i)$$

2. Information Gain:-

A measure of how well a certain quality reduces uncertainty is called information Gain

ID3 splits the data at each stage, choosing the property that maximizes information Gain

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{v \in \text{values}(A)} \frac{|S_v|}{|S|} \times \text{Entropy}(S|v)$$

3. Gain Ratio:-

Gain ratio is an improvement on Information Gain that considers the inherent worth of characteristics that have a wide range of possible values.

It deals with bias of Information Gain in favour of characteristics with more pronounced values

$$\text{SplitInformation}(S, A) = -\frac{|S_{yes}|}{|S|}\log_2\frac{|S_{yes}|}{|S|} - \frac{|S_{no}|}{|S|}\log_2\frac{|S_{no}|}{|S|}$$

$$\text{Gain Ratio}(S, A) = \frac{\text{Gain}(S, A)}{\text{SplitInformation}(S, A)}$$

2. Regression Trees:-

*. A regression tree is a type of decision tree that is used to predict continuous target variables.

*. It works by partitioning the data into smaller and smaller subsets based on certain criteria, and then predicting the average value of the target variable within each subset.

CART Algorithm:-

*. Classification and Regression Trees (CART) is a decision tree algorithm that is used for both classification and regression tasks.

Tree structure:-

CART builds as tree-like structure consisting of nodes and branches possible outcomes of those decisions.

Splitting criteria:-

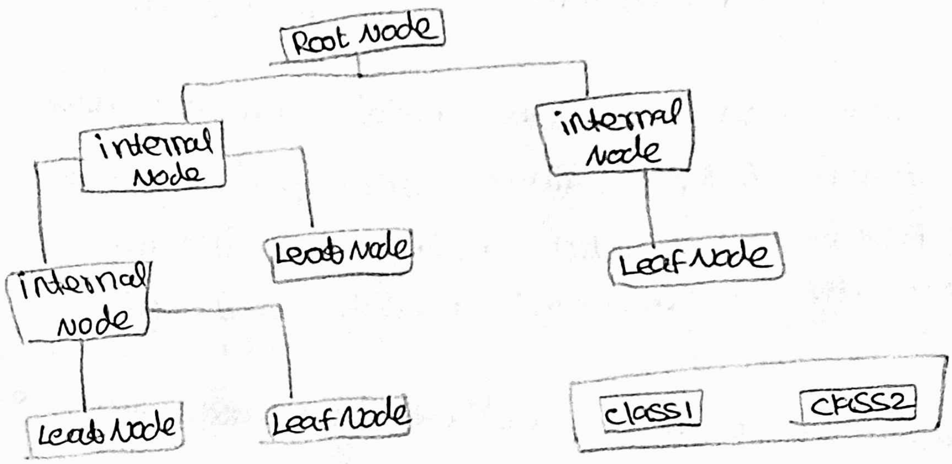
CART uses a greedy approach to split the data at each node. It evaluates all possible splits and selects the one that best reduces the impurity of the resulting subsets.

Pruning:-

To prevent overfitting of the data, pruning is a technique used to remove the nodes that contribute little to the model accuracy. Cost complexity pruning and information gain pruning are two popular pruning techniques.

How does CART algorithm works:-

* The best - split point of each input is obtained
* Based on the best - split points of each input in step 1, the new "best" split point is identified
* split the chosen input according to the best split point
* continue splitting until a stopping rule is satisfied or not further desirable is available.



CART for Regression:-

* A Regression tree is an algorithm where the target variable is continuous and the is used to predict its value
* Regression trees are used when the response variable is continuous. CART for regression is a decision tree learning method that creates a tree-like structure to predict continuous target variables.

3. Performance Measures:-

1. For every classification model, a confusion matrix is used to check the performance of any given set of the data.

Confusion Matrix:-

A confusion matrix is a summary of correct and incorrect predictions and helps visualize the outcomes.

confusion matrix something looks like this

	Actual 0	Actual 1
Predicted 0	True Negative (TN)	False Negative (FN)
Predicted 1	False Positive (FP)	True Positive (TP)

where,

True Positive (TP) : Predicted positive and it's true

True Negative (TN) : predicted negative and it's true

False Positive (FP) : Predicted Positive and it's false

False Negative (FN) : predicted negative and it's false.

Accuracy:-

Accuracy is one of the most commonly used evaluation metrics in classification problems.

Accuracy = Number of correct predictions / Total Number of predictions

Mathematically it is defined as.

Precision = True positive (TP) / True Positive (TP) + False Positive (FP)

Recall:-

The recall is also known as sensitivity or true positive rate. It is the ratio of the number of true positive predictions to total number of actual positive measures the ability in the instances dataset

Recall = True Positive (TP) / True Positive (TP) + False Negative (FN)

F1-score:-

* F1-score is the harmonic mean of precision and recall. It provide a single metric that balances the trade-off between precision and recall.

F1 score = 2x[(precision x Recall) / (precision + Recall)]

The F1-score ranges between 0 & 1

1: indicates perfect precision and recall

0: neither precision nor recall

AUC-ROC curve:-

* AUC-ROC stands for the Area under the Receiver operating characteristic curve. ROC curve is a graphical representation classification model performances at different thresholds.

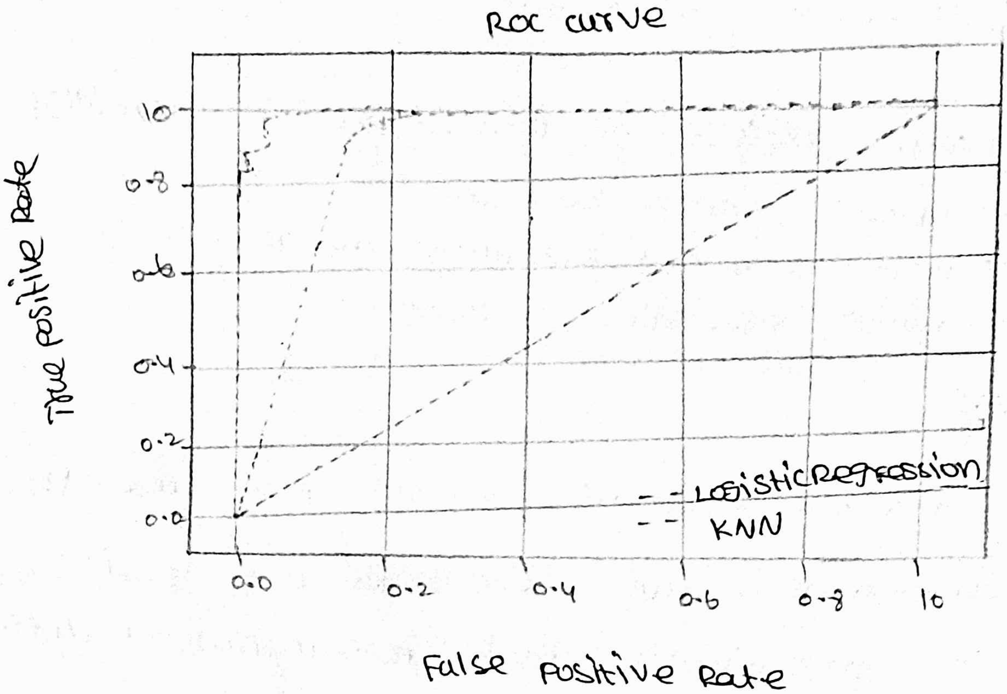
* It is created by plotting the True positive Rate (TPR) against the false positive Rate (FPR).

The formula of TPR & FPR:

True positive rate (TPR/ sensitivity/Recall) = True positive / True positive + false negative

False positive Rate (FPR) = False positive / False positive + True Negative

A typical AUC-ROC curve looks like:-



Impurity	Task	Formula	Description
Gini impurity	classification	$\sum_{i=1}^{C} f_i(1-f_i)$	f_i is the frequency of lable i at a node and C is the number of unique lables.
Entropy	classification	$\sum_{i=1}^{C} -f_i \log(f_i)$	f_i is the frequency of label i at a node and C is the number of unique lables.
variance / Mean square Error (MSE)	Regression	$\frac{1}{N}\sum_{i=1}^{N}(Y_i-\mu)^2$	Y_i label for an instance, N is the number of instances and μ is the mean given by $\frac{1}{N}\sum_{i=1}^{N} Y_i$
variance/Mean Absolute Error (MAE) (scikit-learn only)	Regression	$\frac{1}{N}\sum_{i=1}^{N}\lvert Y_i-\mu)$	Y_i label for an instance, N is the number of instanes and μ is the mean given by $\frac{1}{N}\sum_{i=1}^{N} Y_i$