

## 1. Importance of entropy and information gain?

In simple term, Entropy measures the impurity or uncertainty present in the data.

How much information a particular variable or feature contributes to the model. Information gain basically tells the importance of a particular variable or feature toward the target variable or final result.

## 2. How decision tree can be used for regression?

A regression tree is basically a decision tree that is used for the task of regression which can be used to predict continuous valued outputs instead of discrete outputs

A measure that tells us how much our predictions deviate from the original target and that's the entry-point of mean square error.

$$\text{MSE} = 1/n (y - \hat{y})^2$$

where  $\hat{y}$  is actual prediction

### Step 1. Sorting of Data Values

step 2. we calculated the average for the first 2 numbers of sorted X and split the dataset based on that and calculated the predictions.

Step3. we need to choose the point at which we are going to split the dataset. and that point is the point, which resulted in the lowest mean squared error on splitting at it

5. List down advantages and dis-advantages of decision tree?

A decision tree does not require normalization of data. A decision tree does not require scaling of data as well. Missing values in the data also do NOT affect the process of building a decision tree to any considerable extent. A Decision tree model is very intuitive and easy to explain to technical teams as well as stakeholders.