**DECISION TREE REGRESSION**

* **CART -** Classification and Regression Trees
* This is an umbrella term that encompassed two types of Decision Tress

1. Classification Trees
2. Regression Trees

* **Regression Trees**
* We are given two independent variables, x1 and x2, and what we are predicting is a third dependent variable which is y.
* You cannot plot the y variable on a 2D chart/graph, as y is the third dimension, so we usually plot (if needed), the y variable on a 3d plot.
* Once you run the decision/regression tree algorithm, what will happen is your scatter plot, which has the x1 and x2 variable plotted on it, will be split into segments.
* How and where the splits are made is determined by the algorithm.
* It involves looking at the Information Entropy.
* Information Entropy is a mathematical concept and is quite complex.
* So, it basically means, when a split is performed, is the split increasing the amount of information that we have about the points (the plotted points). Is it adding some value to our way that we want to group our points?
* And the algorithm knows when to stop when there is a certain minimum for the information that needs to be added, and once it cannot add any more information to our setup by splitting the leaves (each split is called a leaf).
* So, it stops splitting when there is no information left to add, or when there is less than 5% of the total points in that leaf and that leaf wouldn’t be created.
* The final leaves are called terminal leaves.

**Chart, scatter chart

Description automatically generated** Diagram

Description automatically generated

* So, this is how we create a decision tree. But what happens next, how do we populate the boxes?
* This is where we need to remember about our dependent variable (y), the third dimension. And what we need to check is, how are we going to predict the value of y for a new observation that gets added to the scatter plot or to our dataset?
* Let’s say the new observation has x1 = 30, and x2 = 40. It will fall somewhere in the terminal leaf between greater than 20, less than 170 and less than 40.
* The way the value of new observation is predicted is – you take average of each of your terminal leaf.
* So, you take average of y for all the terminal leaves and that will be assigned to any new point that is assigned to/falls into that terminal leaf.
* Decision Tree Regression is not the best model to use on single feature dataset.

Programing Part –

* You might need to add some data processing like encoding categorical data or handling missing data.
* But you don’t need to apply feature scaling when using Regression Tree nor Random Forest because the predictions from Decision Tree or Random Forest are resulting from successive split of the data, and therefore, they are not some equations like with the previous models.
* For encoding the categorical variable, if the order matters you will apply Label Encoder. If the order does not matter, you will apply ColumnTransformer with OneHotEncoder.
* Decision Tree Regression is not the best model to use on single feature dataset.