Question:

Why scaling (standardization or Normalization) is not required in decision Tree?

Because Decision Tree algorithms don't rely on the scale or distribution of input features. Decision trees make splits in the data based on Individual feature values, and these splits are determined by comparing feature values with a threshold. Since standardization (scaling) only rescales the data without changing the relationships of the features, It doesn't make impact the decision tree's ability to make decisions effectively. It can handle both continuous and categorical features without any pre processing.

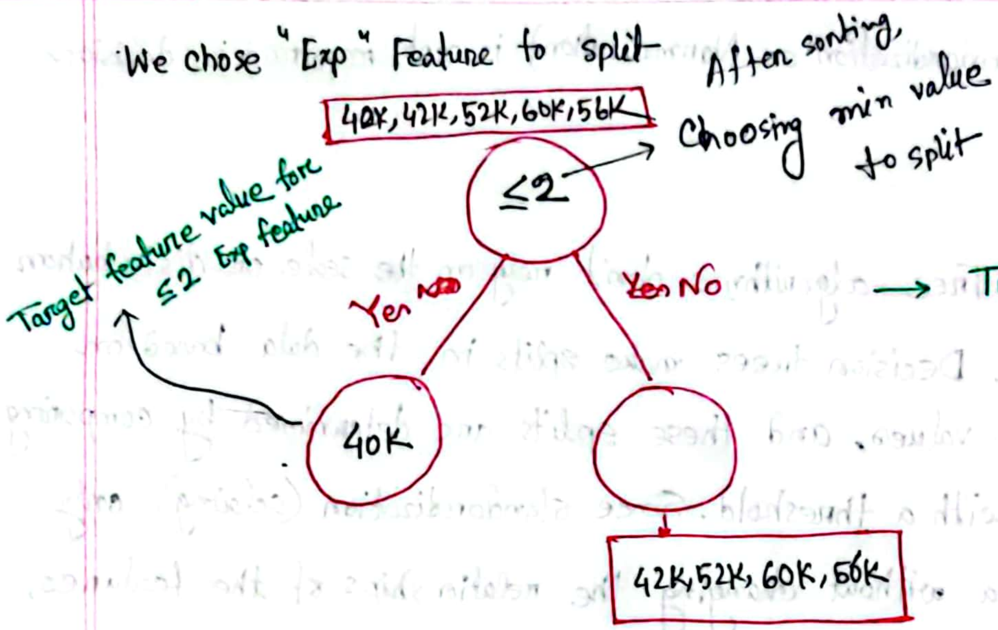
Decision Tree Regressor:

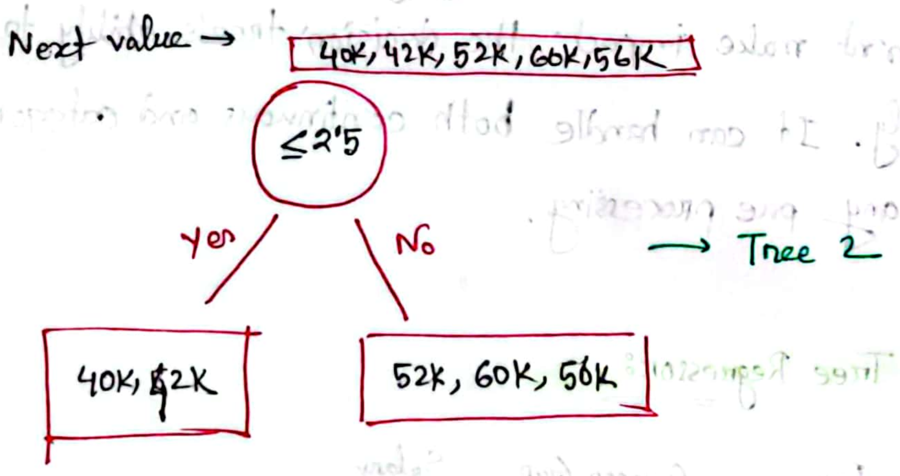
Dataset:

Exp	Career Gap	Salary
2	Yes	40K
2.5	Yes	42K
3	No	52K
4	No	60K
4.5	Yes	56K

Avg (50K)

We chose "Exp" Feature to split





In Decision Tree classifier we use entropy or Gini impurity to check purity but for Regression case we use ~~Variation Reduction~~ Variance Reduction

$$\text{Formula} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2 \quad [\text{Mean squared error}]$$

For Tree 1:

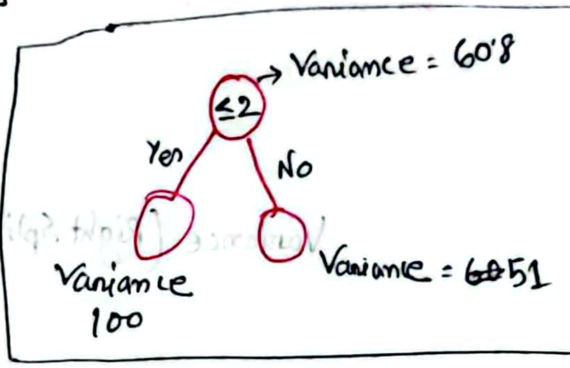
$$\therefore \text{Variance}\,(\text{Root})\;(\leq 2) = \frac{1}{5}\sum_{i=1}^{5}\Big[(40-50)^2 + (42-50)^2 + (52-50)^2 + (60-50)^2 + (56-50)^2\Big]$$

Root : Tree 1 node

n = Number of values in target feature for root node

$$= \frac{1}{5}\,[100+64+4+100+36]$$

$$= 60.8$$



$$\therefore \text{Variance}\,(\text{Left Split}) = \frac{1}{1}\big[(40-50)^2\big]$$

$$= 100$$

$$\therefore \text{Variance}\,(\text{Right Split}) = \frac{1}{4}\Big[(42-50)^2 + (52-50)^2 + (60-50)^2 + (56-50)^2\Big]$$

$$= \frac{1}{4}\,[\cancel{100+}\;64+4+100+36]$$

$$= 51$$

Formula: Variance Reduction = $\text{Var}(\text{Root}) - \sum W_i \, \text{Var}(\text{child})$

$$= 60.8\left[\frac{1}{5}\times 100 + \frac{4}{5}\times 51\right]$$

$$= 0$$

$W_{left} = \frac{1}{5}$

$W_{right} = \frac{4}{5}$

$\text{Var}(\text{left}) = 100$

$\text{Var}(\text{Right}) = 51$

For Tree 2:

$$\text{Variance (Root)} = 60.8$$

$$\text{Variance (Left split)} = \frac{1}{2}\left[(40-50)^2 + (42-50)^2\right]$$

$$= \frac{1}{2}\left[100+64\right]$$

$$= 82$$

$$\text{Variance (Right Split)} = \frac{1}{3}\left[(52-50)^2 + (60-50)^2 + (56-50)^2\right]$$

$$= \cancel{140}\ 46.66$$

$$\text{Variance Reduction} = \text{Var (Root)} - \sum w_i \text{ Var (child)}$$

$$= 60.8 - \left[\frac{2}{5} \times 82 + \frac{3}{5} \times 46.66\right]$$

$$= 0.004$$

~~Varr~~ ~~Variance Reduction (Left Split)~~ Tree

Variance Reduction (Tree 1) < Variance Reduction (Tree 2)

We will take this tree split

Because it has greater variance Reduction value

For example purpose we used two tree with the first two values of "Exp" column. Actually all the values will create it's own tree and we will calculate variance Reduction from each of them and find the greatest variance Reduction value.

"Information gain" will be calculated like before (Decision Tree Classifier)