*Decision Tree*

1. Non-Parametric ML Algo --> iske vajese yeah non Linear Data pe bhi kaam karpata hai
2. It is a White-box model

First understand Black Box Model

* Humne input diya & andar vo processing kuch krega & then output dega

Hum yaha isme yeah idea nahi aayega ki usne input se output kaise nikala hai

E.g Neural Network & Chat GPT

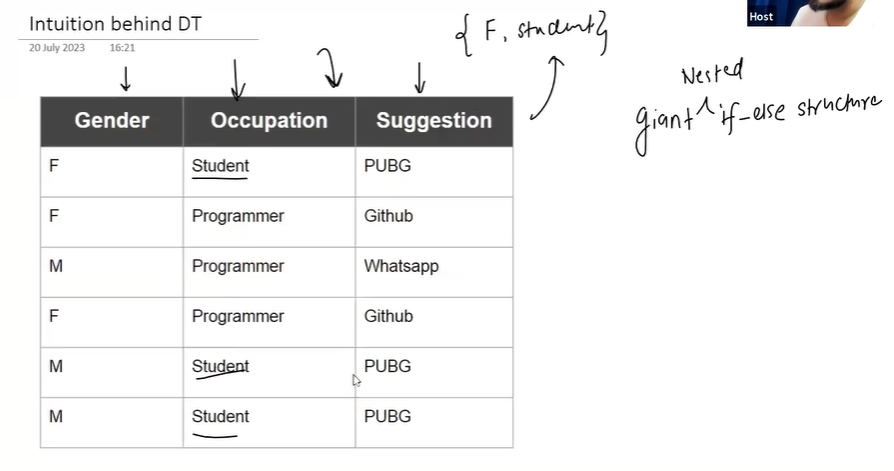
Uska ulta hai yeah Decision Tree Model jo ki White Box Model hai

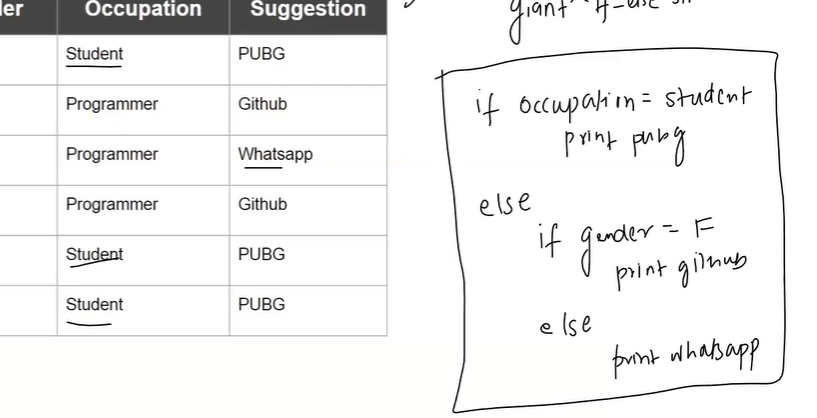
Jisme humko pura smjhega ki Input se Output tak Pochneka decision kaise liya

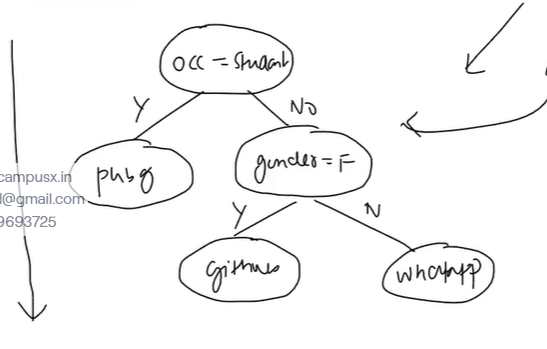
1. Mother of all Tree Based algos

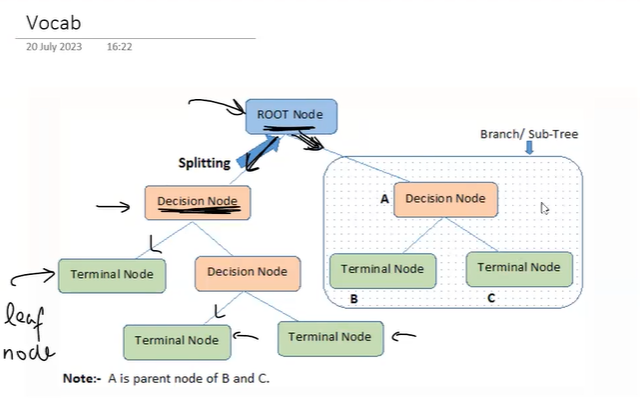
Decision Tree is a building block for Random Forest & XG Boost, Bagging Boosting sbh Decision Tree ke andar aata hai

1. Works with both classification & regression









Decision Tree ke behind bhaut saare algos hote hai

E,g

1. ID3 algo

2. Ch.5 something

3. CART

CART -> Yeahi Sklearn use krta hai behind

CART jo hota hai usme we form Binary Tree

& In ID3 we have more than 2 nodes connected

Impurity function ID3 & Cart dono mai alag rehta hai

We will be focusing more on CART Algorithm as Scikitlearn vahi use krta hai

*CART Algorithm*

CART -> Classification & Regression Trees

It is an algorithm to grow(root to leaf) Decision Tree

**Understanding**

Har node mai 1 Spliting criteria hota hai

Splitting Criteria ko

1 Column or feature chaye & uskesath uski Value

***Impurity Function***

Types of Impurity for classification & Regression ->

Classification -> a.) gini impurity

b.) entropy

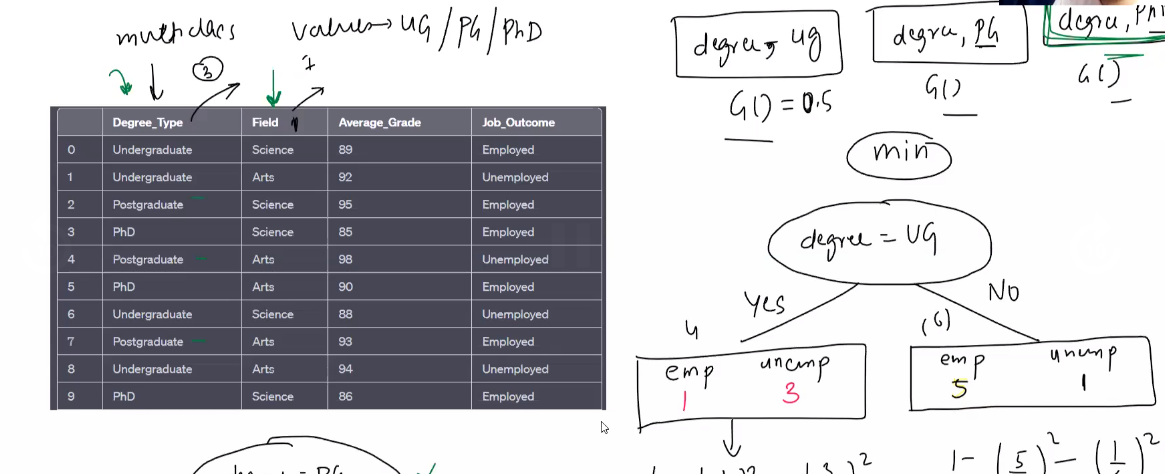
Regression ->

1. MSE
2. MAE

*Classification*

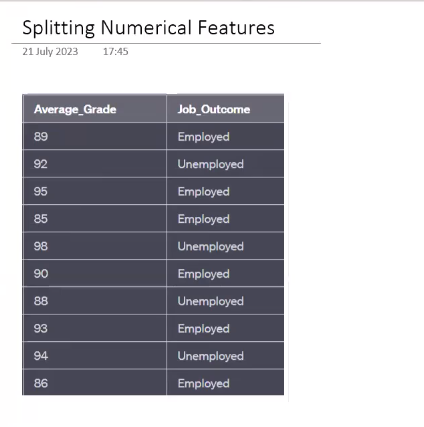
Overall Impurity Value nikalte hai hum saare combinations ka

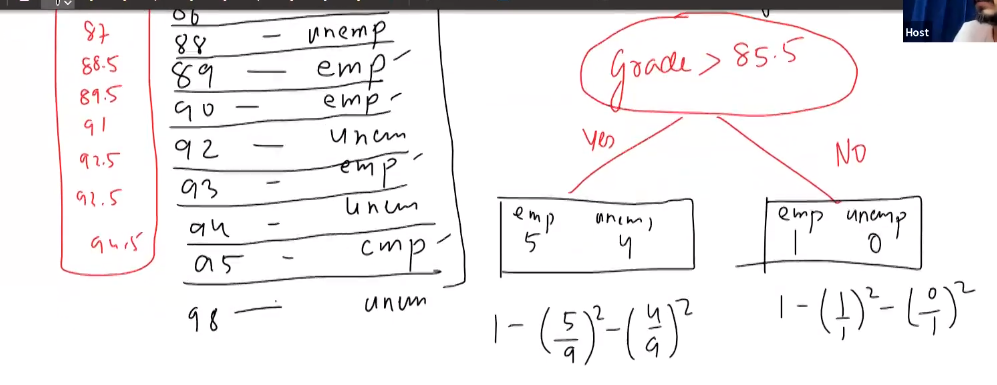
& jis combination ki impurity sbhse kamm hogi vo vala combination hum select krte hai



Agr 2 columns ko gini impurity same aaya then koibhi randomly choice krlo

***Splitting Numerical Features***

******



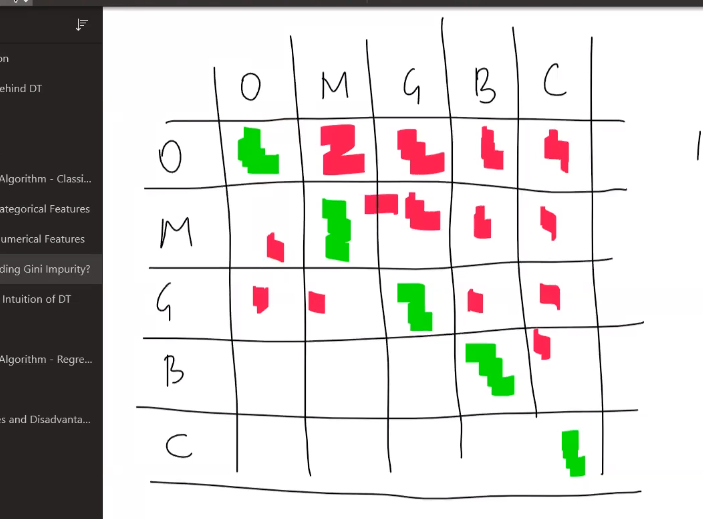
Understanding Gini Impurity

The Gini impurity is a measure of how often a randomly chosen element from the set would be incorrectly labelled if it was randomly labelled according to the distribution of labels in the subset.

Galti kitni krrhe hoo vahi ginni impurity batata hai

Galat honeka probability jo hai vahi Ginni Impurity hai

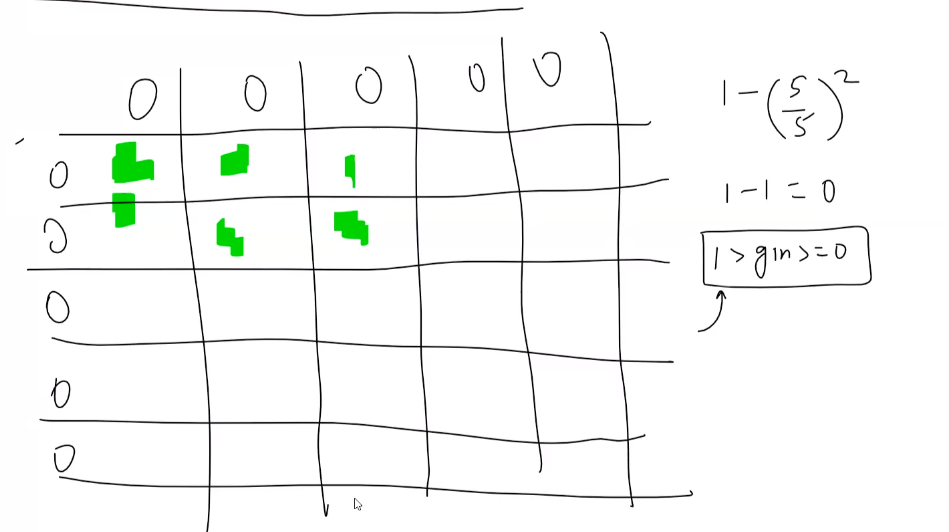
***E.g***



Isme impurity zyaada hogi ki kamm ??

Ofcourse zyaada hogi coz rhightly selected cell bhaut kamm hai

So galat honeki probability high hai yaha

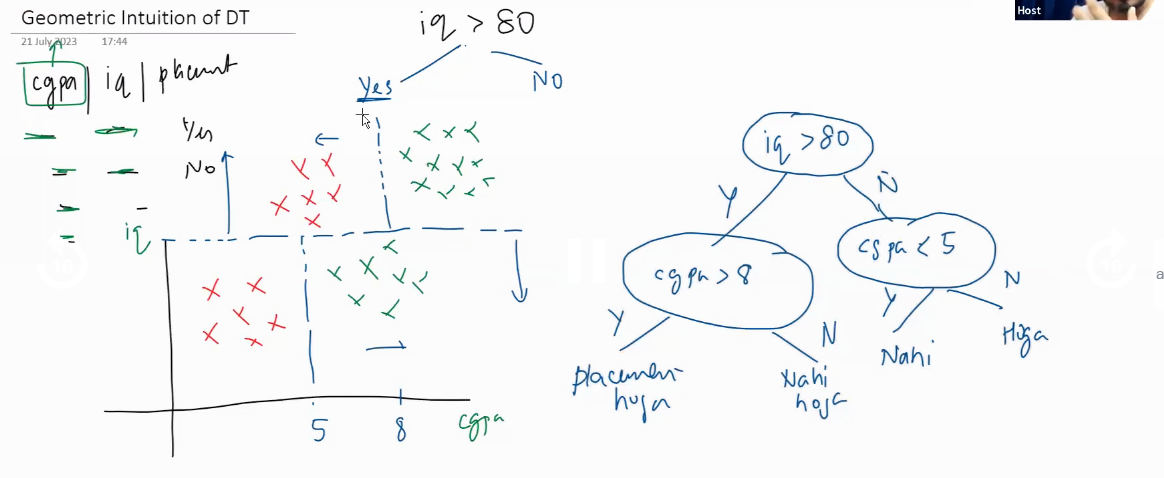


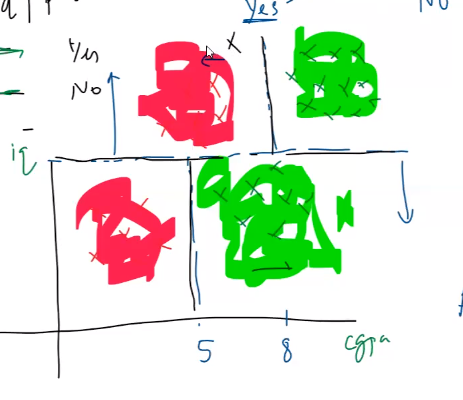
Isme impurity zyaada hogi ki kamm ??

Bhaut kamm ekdum zero % hai impurity

Ginni impurity hum prefer krte hai it must be very close to 0

*Geometric Intuition of Decision Tree*

**

**

*DecisionTree esehie Bss x & y axis ke parallel mai Cut maarte hai*

*& cut marneka yeah goal hai ki*

*Red points alag hojaye & green points alag hojaye*

*Regression ke case mai bss yeahi farak hai ki*

*Impurity function badaljaata hai MSE hojata hai*

*Joki Classification ke case mai Ginni hota hai*