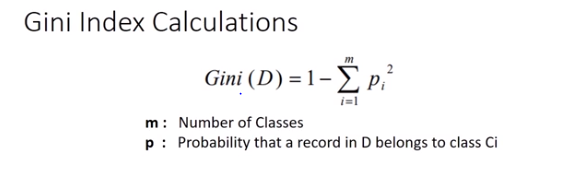
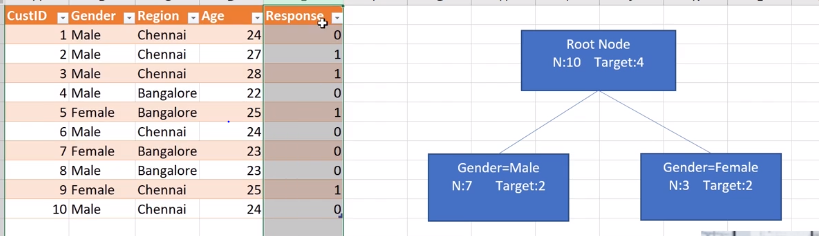
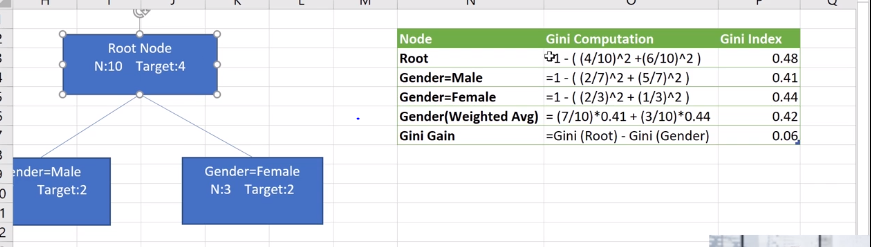
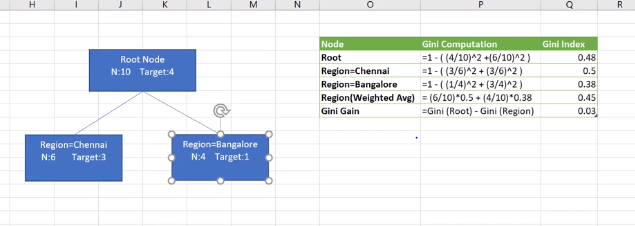
Gini index and Gini gain:



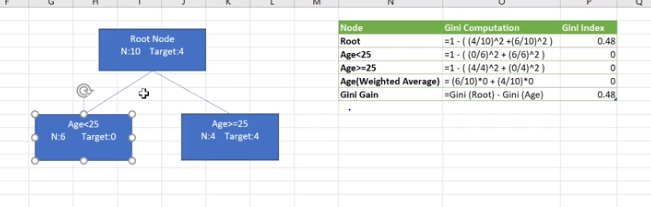
Sample data for calculation



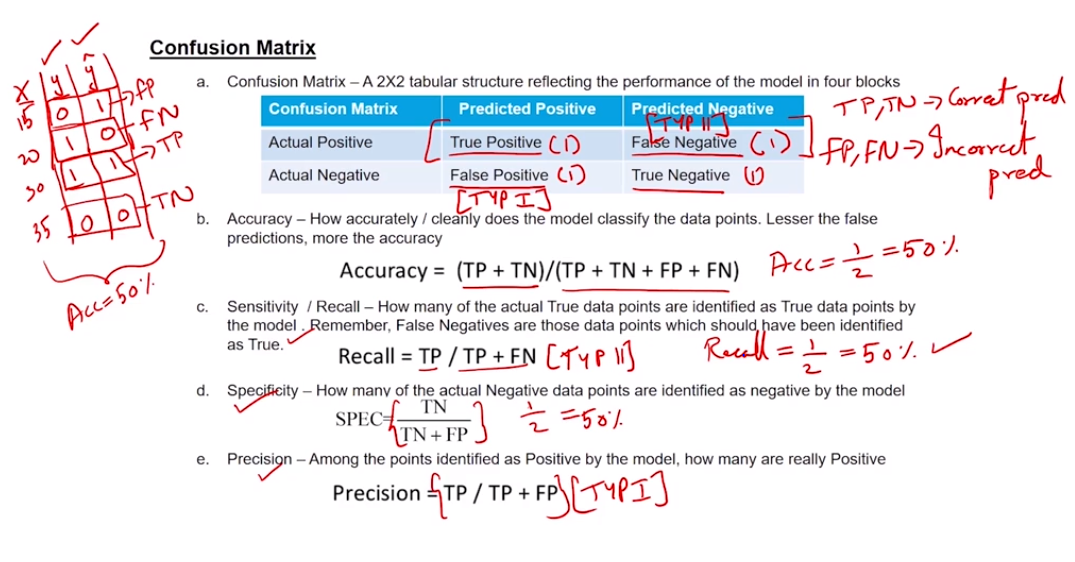


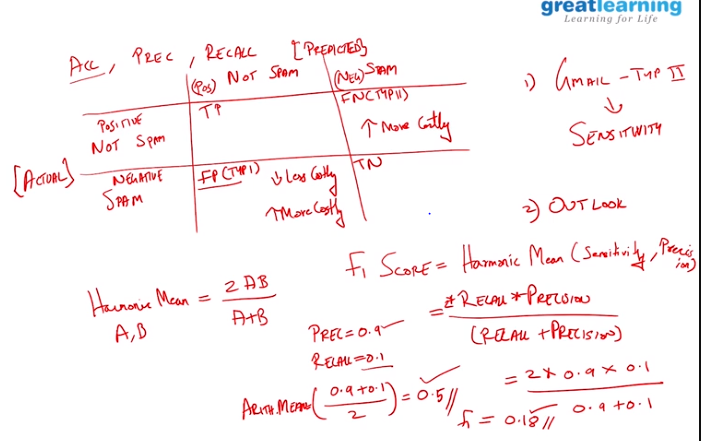


Chennai node above is highly impure and impacts the decision tree accuracy and Gini index for highly impure nodes are always 0.5



When using continuous variable , bucketing them into bins start with median of the range of its values and keep moving left and right until the best combination of bin is achieved towards getting better Gini gain.





Arithmetic mean does not punish the value if either precision or recall is low whereas harmonic mean will while calculation F1 score.