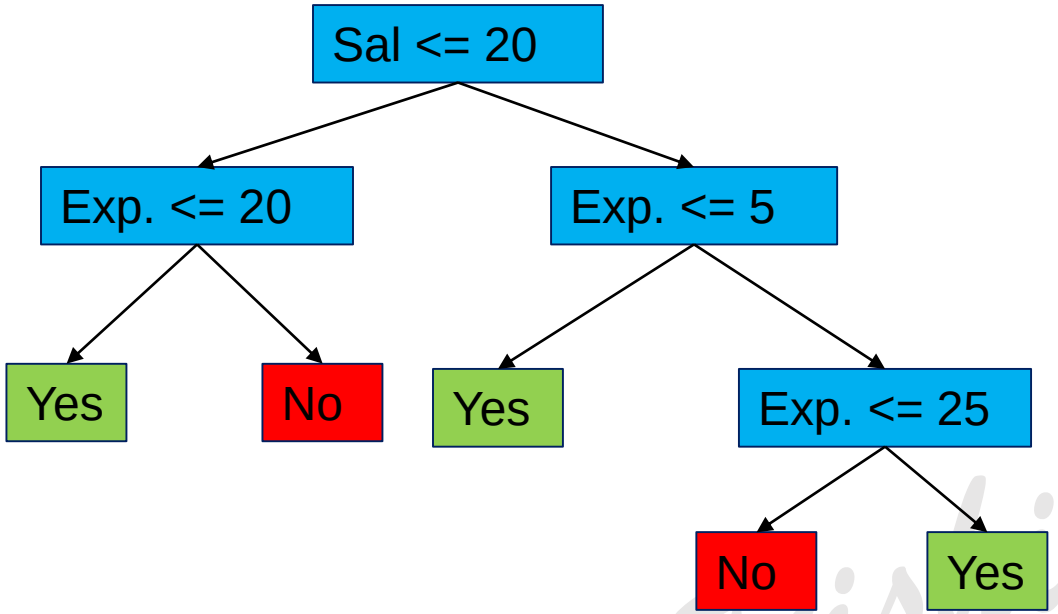
Decision Tree and SVM

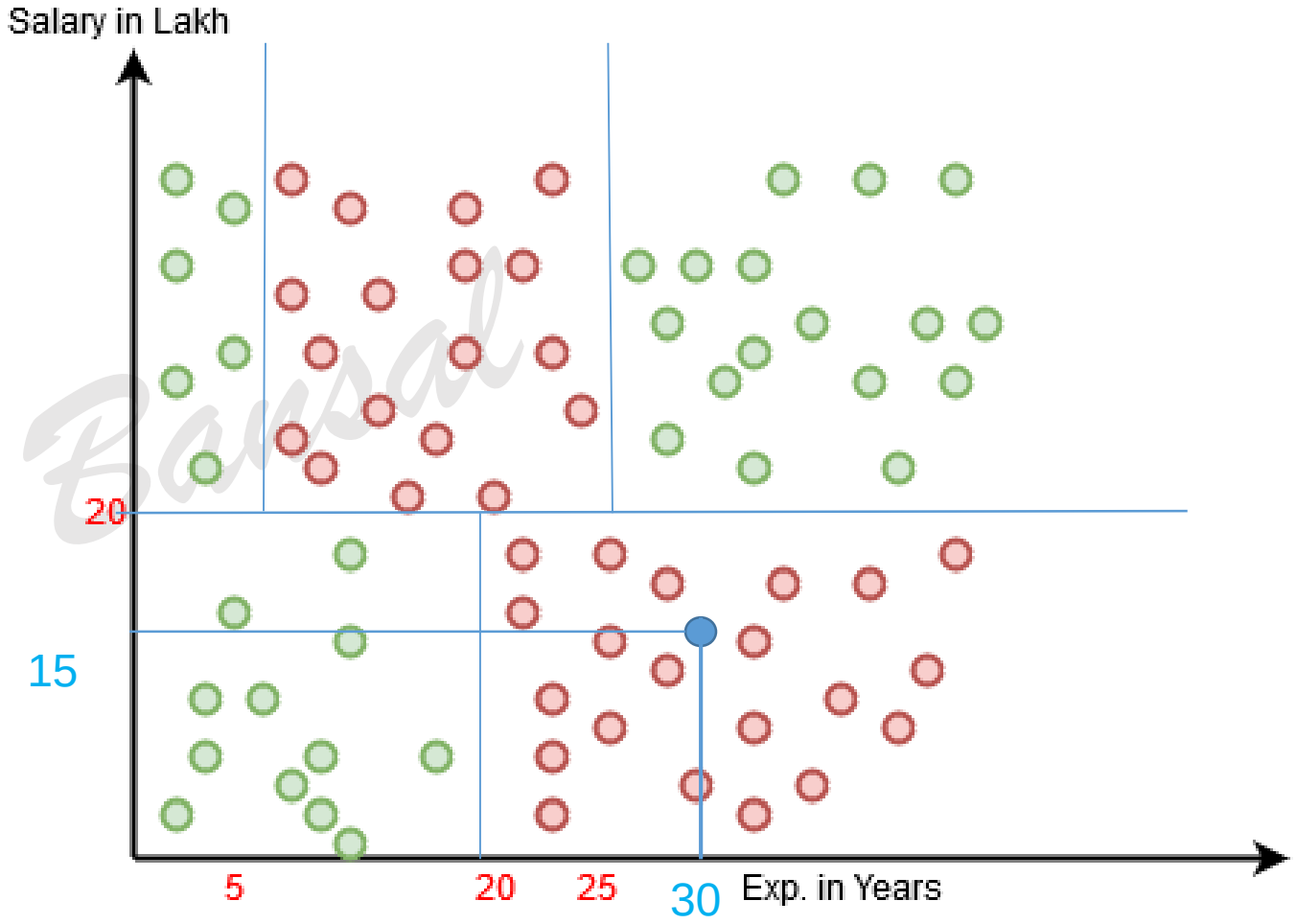
Decision Tree

- Its a tree like data structure to make a model of the data
- uses if-else at every node of the tree
- Supervised Algorithm - For Classification & regression
- DT algorithm generates a decision tree for prediction
- Leafs are the decisions
- DT creation can be based on entropy or gini

Decision Tree Classification

Sal = 15, Exp = 30





Stopping Criteria

Issues without Stopping Criteria:

computationally expensive, difficult to interpret and would probably not work very well with new data

- We decide the maximum depth of the tree (max_depth)
- Purity of the node is more than some pre-specified limit
- Number of cases in the node is less than some pre-specified limit
- Predictor values for all records are identical - in which no rule could be generated to split them

DisAdv:

- tend to degrade the tree's performance

Pruning Trees

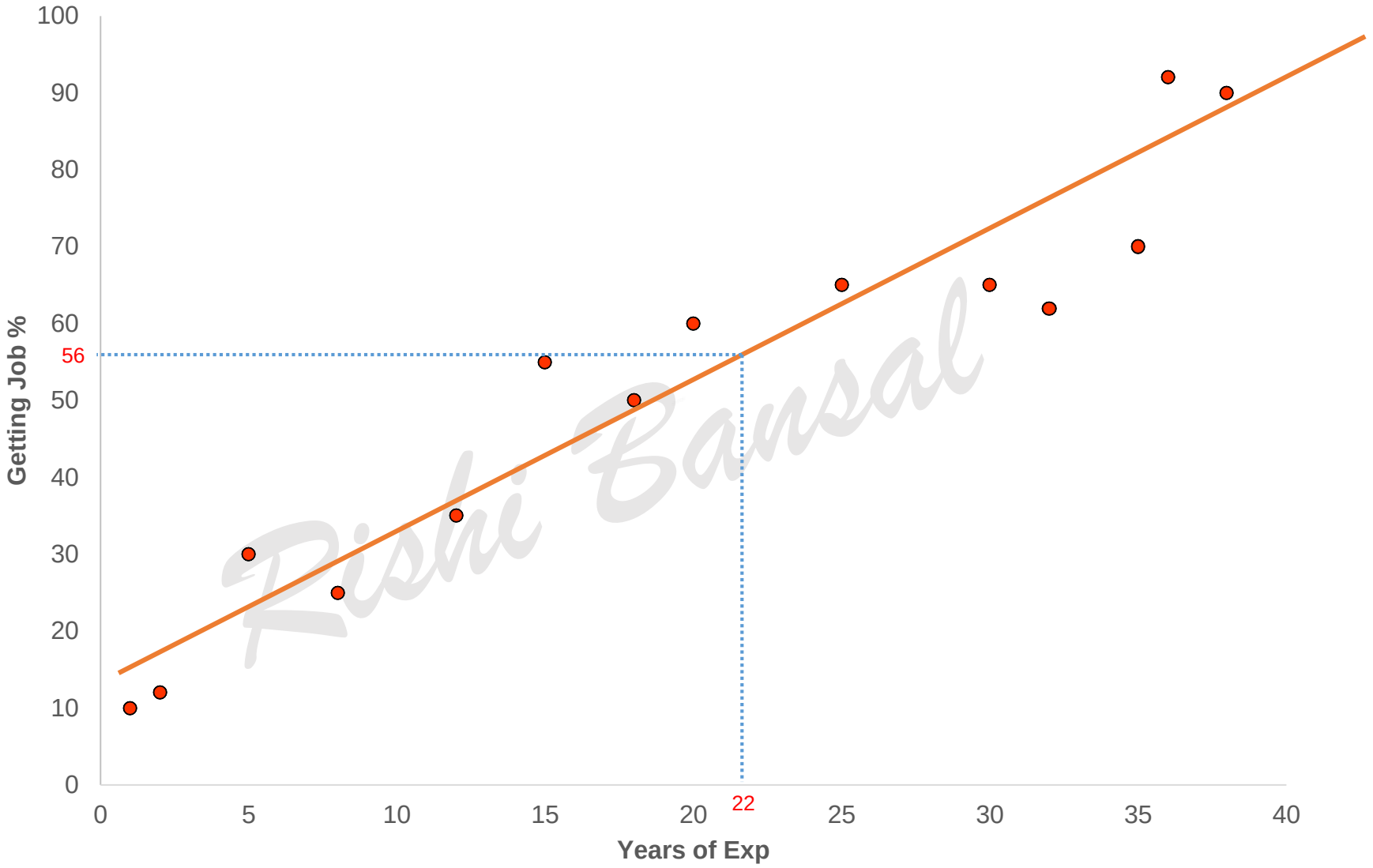
- we allow the tree to overfit on the data and after that we start cutting the branches which are not leading to better accuracy

Gini and Entropy Comparison

- Gini is faster
- Entropy – results are better

Decision Tree Regression

- Continues target is predicted with Tree.
- ID3 is modified for regression by replacing Information Gain with Mean Squared Error.
- Decision Tree tries to partition data into subsets of homogenous contents (minimize mean squared error)


100
90
80
70
60
56
50
40
30
20
10
0
Getting Job %
0
5
10
15
20
22
25
30
35
40
Years of Exp
Rishi Bansal

Decision Tree Regression