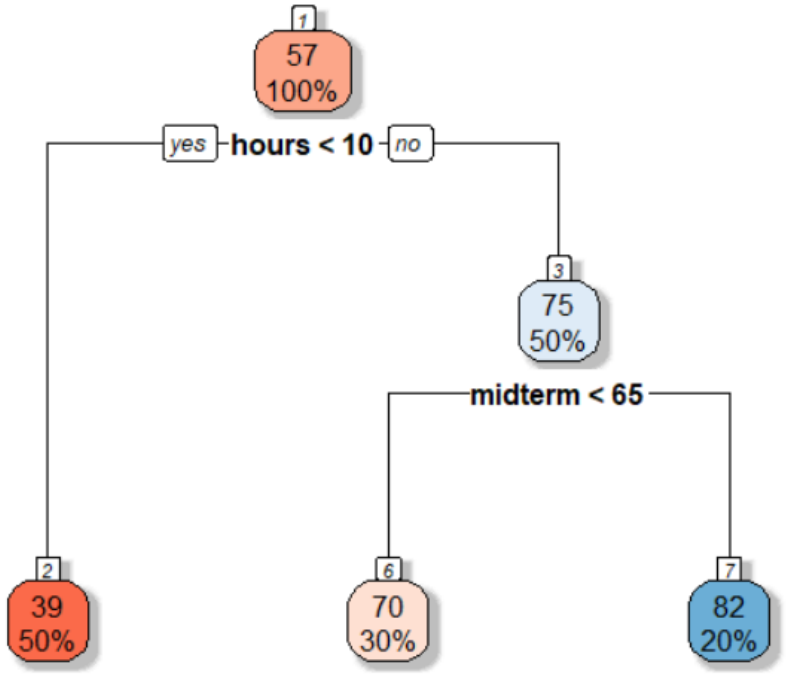
Start-Tech Academy

# Steps



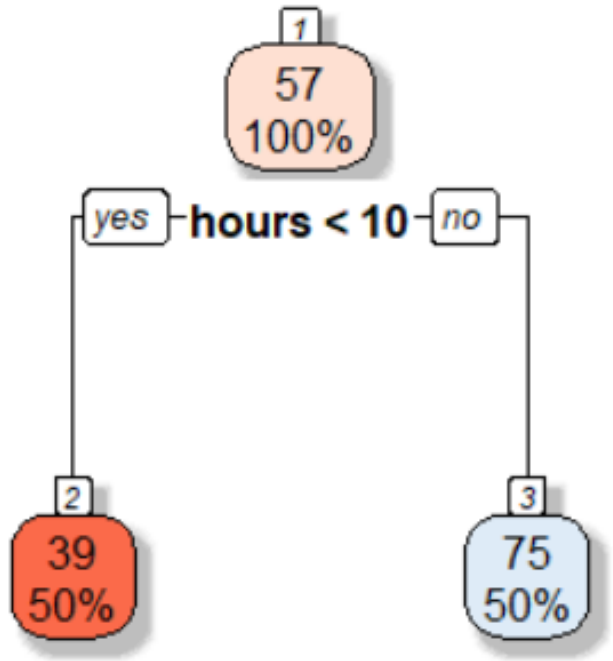

1. We divide the predictor space—that is, the set of possible values for X1,X2, . . .,Xp—into J distinct and non-overlapping regions, R1,R2, . . . , RJ .

2. For every observation that falls into the region Rj, we make the same prediction, which is simply the mean of the response values for the training observations in Rj .

Goal is to minimize RSS

$$\sum_{j=1}^{J} \sum_{i \in R_j} (y_i - \hat{y}_{R_j})^2$$

# Building tree



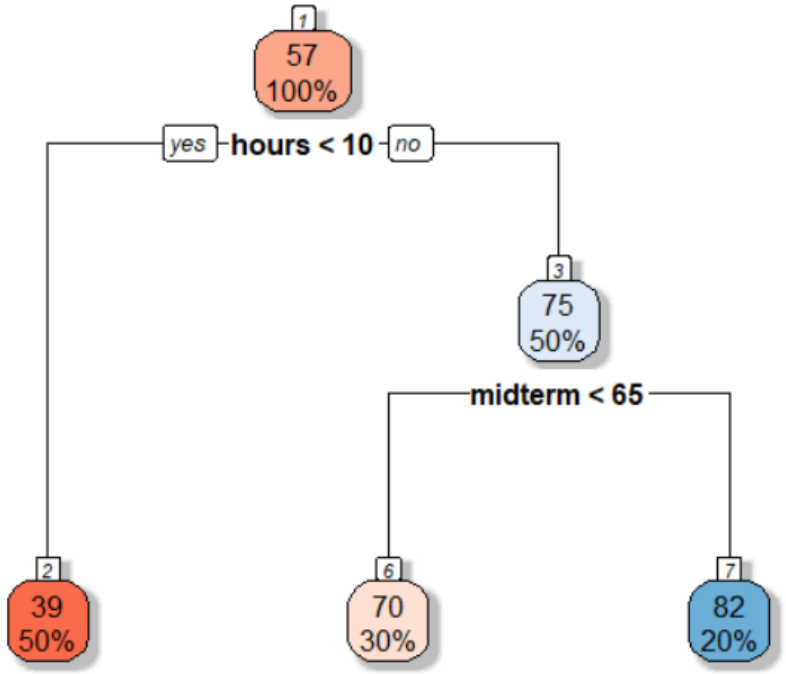

| | score | hours | midterm |
|---|---|---|---|
| 1 | 35 | 6 | 42 |
| 2 | 38 | 5 | 65 |
| 3 | 40 | 7 | 35 |
| 4 | 45 | 6 | 75 |
| 5 | 35 | 8 | 60 |
| 6 | 65 | 11 | 50 |
| 7 | 70 | 12 | 45 |
| 8 | 75 | 18 | 40 |
| 9 | 80 | 14 | 80 |
| 10 | 85 | 12 | 82 |

Mean score 39

Mean score 75

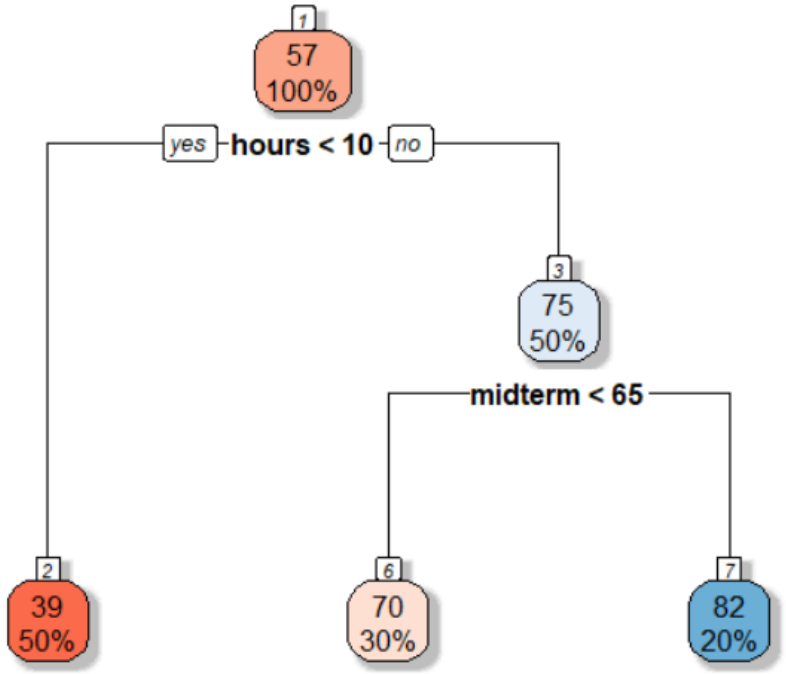
$$\sum_{j=1}^{J} \sum_{i \in R_j} (y_i - \hat{y}_{R_j})^2$$

# Approach




- Top-down, greedy approach that is known as recursive binary splitting.
- Top-down because it begins at the top of the tree and then successively splits the predictor space
- Each split is indicated via two new branches further down on the tree.
- It is greedy because at each step of the tree-building process, the best split is made at that particular step, rather than looking ahead and picking a split that will lead to a better tree in some future step.

# Steps




1. Considers all predictors and all possible cut point values
2. Calculates RSS for each possibility
3. Selects the one with least RSS
4. Continues till stopping criteria is reached