MACHINE LEARNING

Algorithms

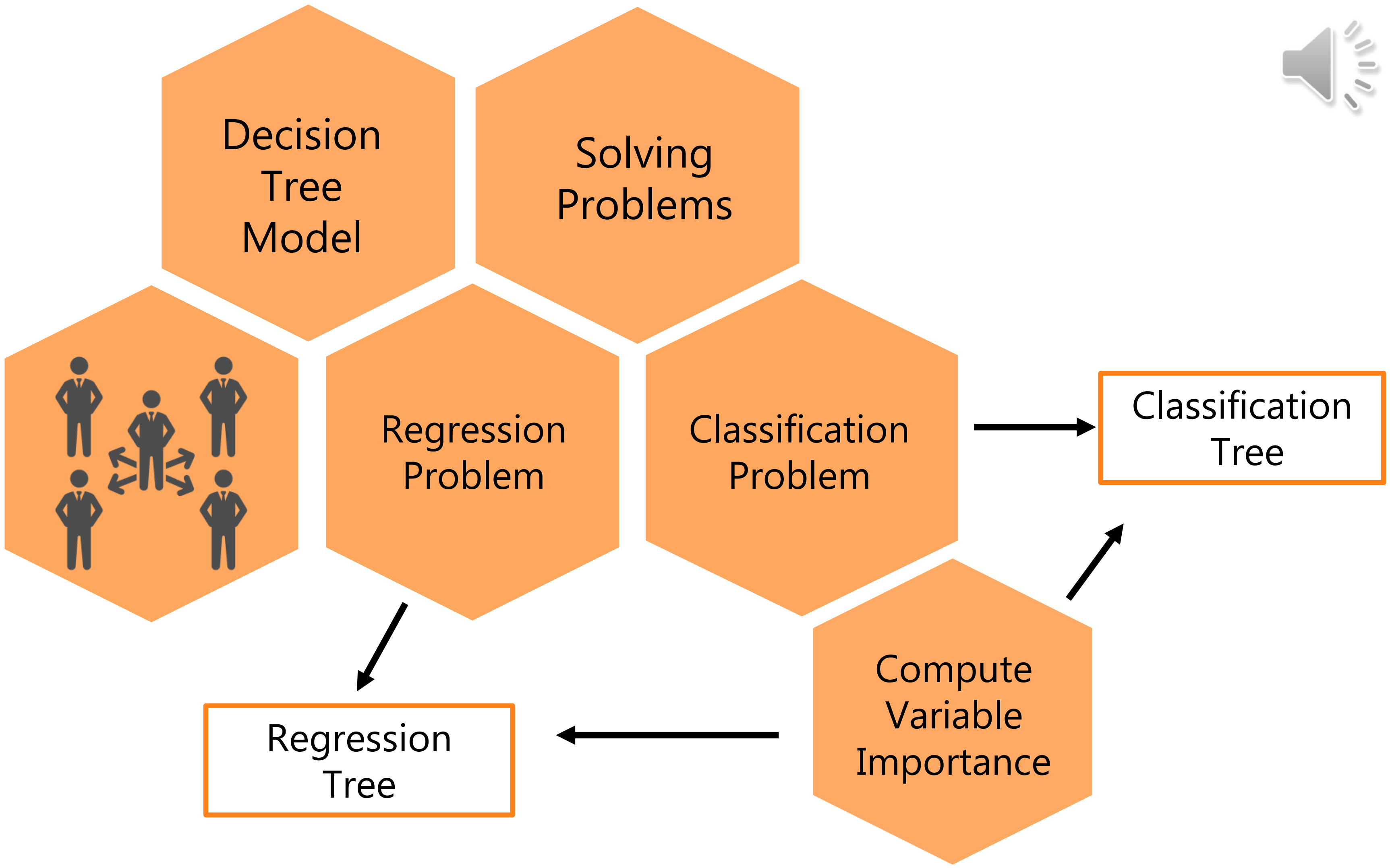
Class

Tree Based Models

Topic

Introduction to Classification Trees

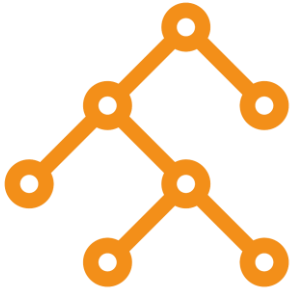
Agenda



© Jigsaw Academy Education Pvt Ltd

Decision Tree: Overview

Solve both regression and classification problems



Decision Tree works is based on a branch of computer science known as **Information Theory**

The classic use case of decision trees is analysis of segments in business data

© Jigsaw Academy Education Pvt Ltd

Decision Tree

Existing Data of a Bank

Customer	Age	Gender	Marital Status	# cr. Cards	Profitability
1	36	M	M	1	P
2	32	M	S	3	U
3	38	M	M	2	P
4	40	M	S	1	U
5	44	M	M	0	P
6	56	F	M	0	P
7	58	F	S	1	U
8	30	F	S	2	P
9	28	F	M	1	U
10	26	F	M	0	U

Profitable

Unprofitable

To build a predictive model classifying customers logistic, Regression Classifier can be used

© Jigsaw Academy Education Pvt Ltd

Decision Tree

Existing Data of a Bank

Customer	Age	Gender	Marital Status	# cr. Cards	Profitability
1	36	M	M	1	P
2	32	M	S	3	U
3	38	M	M	2	P
4	40	M	S	1	U
5	44	M	M	0	P
6	56	F	M	0	P
7	58	F	S	1	U
8	30	F	S	2	P
9	28	F	M	1	U
10	26	F	M	0	U

Total Population = 10
Profitable = 5
Unprofitable = 5
Profitability rate = 50%

> 35 **Age** **<= 35**

Total Population = 6
Profitable = 4
Unprofitable = 2
Profitability rate = 66%

Total Population = 4
Profitable = 1
Unprofitable = 3
Profitability rate = 25%

© Jigsaw Academy Education Pvt Ltd

Decision Tree

Total Population = 10
Profitable = 5
Unprofitable = 5
Profitability rate = 50%

> 35 **Age** **<= 35**

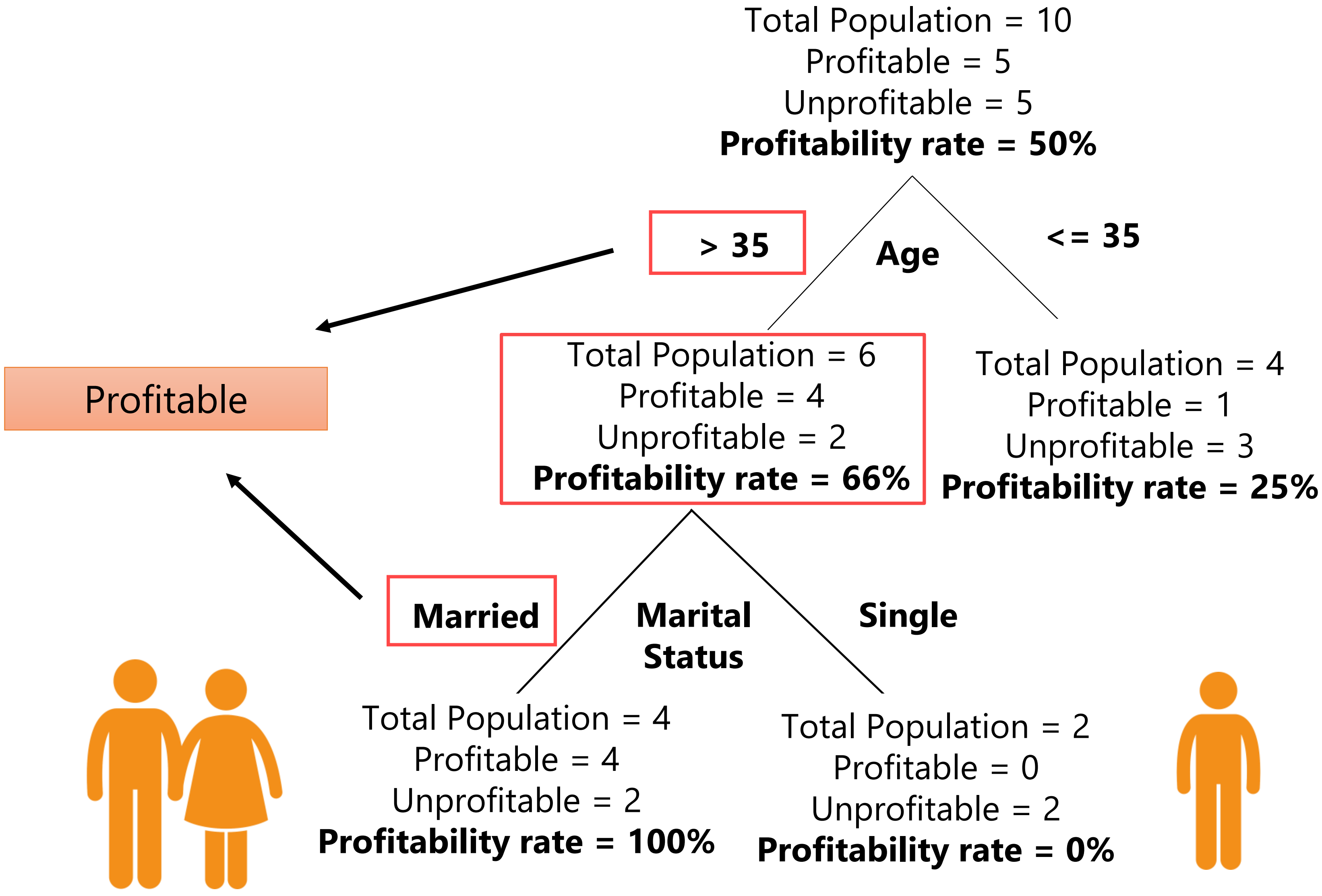
Total Population = 6
Profitable = 4
Unprofitable = 2
Profitability rate = 66%

Total Population = 4
Profitable = 1
Unprofitable = 3
Profitability rate = 25%

The segment of data which is >35 has a higher chance of seeing a profitable customer

© Jigsaw Academy Education Pvt Ltd

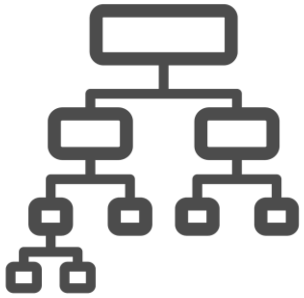
Decision Tree



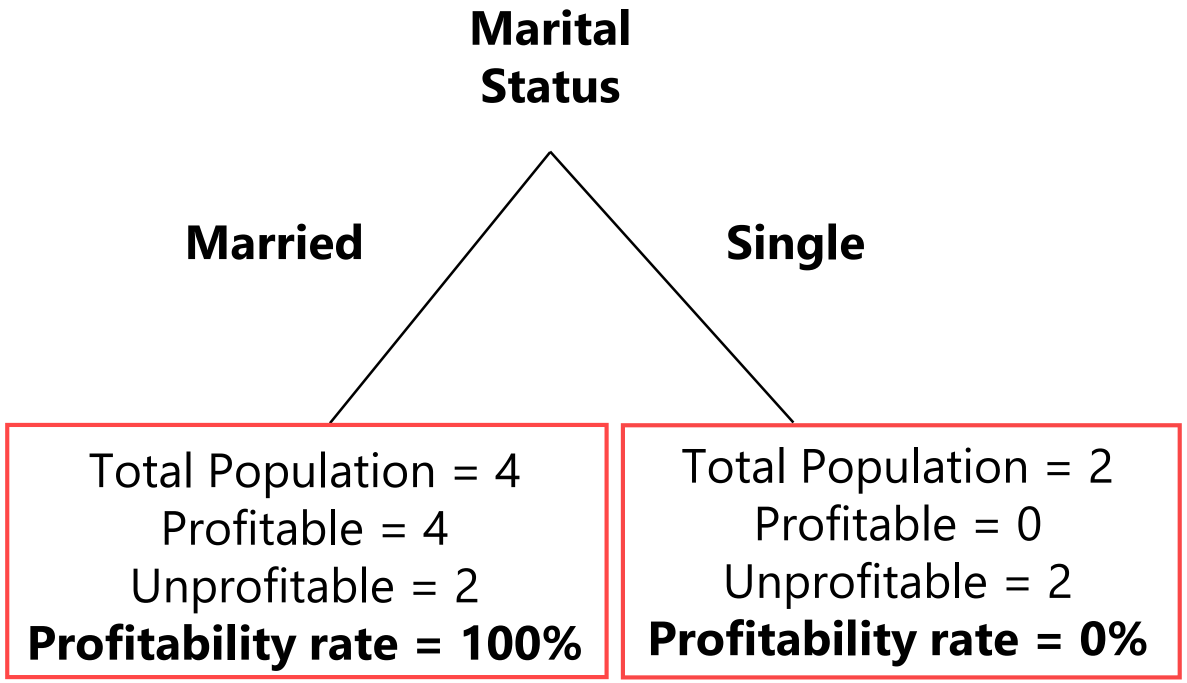
© Jigsaw Academy Education Pvt Ltd

Decision Tree

Decision tree classifier - Recursively sub-setting data can reveal interesting patterns



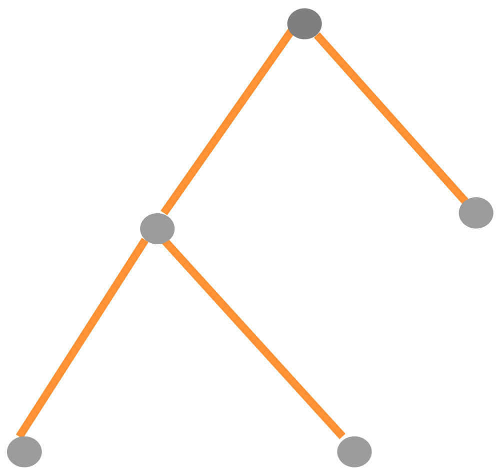
Data needs to be split in such a way so that the subsets of data end up being dominated by one class of the target variable



© Jigsaw Academy Education Pvt Ltd

Decision Tree

Decision Tree splits into 2 parts at each node



Binary Tree

Most implementations of a decision trees produce binary splits

© Jigsaw Academy Education Pvt Ltd

Decision Tree: Algorithm

How to decide which variable should be used to create splits?

Understand the intuition behind creating splits

The intuition will be formalized by introducing purity metrics

© Jigsaw Academy Education Pvt Ltd

Decision Tree: Algorithm

Previous Example

Total Population = 10
Profitable = 5
Unprofitable = 5
Profitability rate = 50%

Age

> 35

Total Population = 6
Profitable = 4
Unprofitable = 2
Profitability rate = 66%

<= 35

Total Population = 4
Profitable = 1
Unprofitable = 3
Profitability rate = 25%

Total Population = 10
Profitable = 5
Unprofitable = 5
Profitability rate = 50%

Gender

Male

Total Population = 5
Profitable = 3
Unprofitable = 2
Profitability rate = 60%

Female

Total Population = 5
Profitable = 2
Unprofitable = 3
Profitability rate = 40%

Both splits can be compared to understand which split is better

© Jigsaw Academy Education Pvt Ltd

Decision Tree: Algorithm

Total Population = 10
Profitable = 5
Unprofitable = 5
Profitability rate = 50%

Age

> 35

Total Population = 6
Profitable = 4
Unprofitable = 2
Profitability rate = 66%

<= 35

Total Population = 4
Profitable = 1
Unprofitable = 3
Profitability rate = 25%

Total Population = 10
Profitable = 5
Unprofitable = 5
Profitability rate = 50%

Gender

Male

Total Population = 5
Profitable = 3
Unprofitable = 2
Profitability rate = 60%

Female

Total Population = 5
Profitable = 2
Unprofitable = 3
Profitability rate = 40%

Which variable produces better splits?

Age or Gender?

© Jigsaw Academy Education Pvt Ltd

Decision Tree: Algorithm

Total Population = 10
Profitable = 5
Unprofitable = 5
Profitability rate = 50%

Age

> 35

Total Population = 6
Profitable = 4
Unprofitable = 2
Profitability rate = 66%

<= 35

Total Population = 4
Profitable = 1
Unprofitable = 3
Profitability rate = 25%

Total Population = 10
Profitable = 5
Unprofitable = 5
Profitability rate = 50%

Gender

Male

Total Population = 5
Profitable = 3
Unprofitable = 2
Profitability rate = 60%

Female

Total Population = 5
Profitable = 2
Unprofitable = 3
Profitability rate = 40%

Good split in context of classification problem

Split produced by variable age are better than the splits produced by variable gender

Greater the **class imbalance**, better the split

© Jigsaw Academy Education Pvt Ltd

Decision Tree: Algorithm

Class imbalance can be measured by computing Gini or Entropy

$$Gini = 1 - \sum p_i^2$$

$$Entropy = -\sum p_i log_2 p_i$$

© Jigsaw Academy Education Pvt Ltd

Decision Tree: Algorithm

$$Gini = 1 - \sum p_i^2$$

Total Population = 10
Profitable = 5
Unprofitable = 5
Profitability rate = 50%

Age

> 35

Total Population = 6
Profitable = 4
Unprofitable = 2
Profitability rate = 66%

$$1 - [\left(\frac{4}{6}\right)^2 + \left(\frac{2}{6}\right)^2]$$

0.44

<= 35

Total Population = 4
Profitable = 1
Unprofitable = 3
Profitability rate = 25%

$$1 - [\left(\frac{1}{4}\right)^2 + \left(\frac{3}{4}\right)^2]$$

$\boxed{0.375}$

Total Population = 10
Profitable = 5
Unprofitable = 5
Profitability rate = 50%

Gender

Male

Total Population = 5
Profitable = 3
Unprofitable = 2
Profitability rate = 60%

$$1 - [\left(\frac{3}{5}\right)^2 + \left(\frac{2}{5}\right)^2]$$

0.48

Female

Total Population = 5
Profitable = 2
Unprofitable = 3
Profitability rate = 40%

$$1 - [\left(\frac{2}{5}\right)^2 + \left(\frac{3}{5}\right)^2]$$

0.48

© Jigsaw Academy Education Pvt Ltd

Decision Tree: Algorithm

$$Gini = 1 - \sum p_i^2$$

Total Population = 10
Profitable = 5
Unprofitable = 5
Profitability rate = 50%

> 35 **Age** **<= 35**

> 35	**<= 35**
Total Population = 6 Profitable = 4 Unprofitable = 2 **Profitability rate = 66%**	Total Population = 4 Profitable = 1 Unprofitable = 3 **Profitability rate = 25%**

$$\left(\frac{6}{10}\right) * 0.44 \quad + \quad \left(\frac{4}{10}\right) * 0.375$$

$\boxed{0.41}$

Total Population = 10
Profitable = 5
Unprofitable = 5
Profitability rate = 50%

Male **Gender** **Female**

Male	**Female**
Total Population = 5 Profitable = 3 Unprofitable = 2 **Profitability rate = 60%**	Total Population = 5 Profitable = 2 Unprofitable = 3 **Profitability rate = 40%**

$$\left(\frac{5}{10}\right) * 0.48 \quad + \quad \left(\frac{5}{10}\right) * 0.48$$

0.48

© Jigsaw Academy Education Pvt Ltd

Decision Tree: Algorithm

$$Entropy = -\sum p_i \log_2 p_i$$

Split on Age

Total Population = 10
Profitable = 5
Unprofitable = 5
Profitability rate = 50%

Age

> 35

Total Population = 6
Profitable = 4
Unprofitable = 2
Profitability rate = 66%

$$-[\left(\frac{4}{6}\right) * \log_2\left(\frac{4}{6}\right) + \left(\frac{2}{6}\right) * \log_2\left(\frac{2}{6}\right)]$$

0.91

<= 35

Total Population = 4
Profitable = 1
Unprofitable = 3
Profitability rate = 25%

$$-[\left(\frac{1}{4}\right) * \log_2\left(\frac{1}{4}\right) + \left(\frac{3}{4}\right) * \log_2\left(\frac{3}{4}\right)]$$

$\boxed{0.81}$

Split on Gender

Total Population = 10
Profitable = 5
Unprofitable = 5
Profitability rate = 50%

Gender

Male

Total Population = 5
Profitable = 3
Unprofitable = 2
Profitability rate = 60%

$$-[\left(\frac{3}{5}\right) * \log_2\left(\frac{3}{5}\right) + \left(\frac{2}{5}\right) * \log_2\left(\frac{2}{5}\right)]$$

0.97

Female

Total Population = 5
Profitable = 2
Unprofitable = 3
Profitability rate = 40%

$$-[\left(\frac{2}{5}\right) * \log_2\left(\frac{2}{5}\right) + \left(\frac{3}{5}\right) * \log_2\left(\frac{3}{5}\right)]$$

0.97

© Jigsaw Academy Education Pvt Ltd

Decision Tree: Algorithm

$$Entropy = -\sum p_i log_2 p_i$$

Total Population = 10
Profitable = 5
Unprofitable = 5
Profitability rate = 50%

Age

> 35

Total Population = 6
Profitable = 4
Unprofitable = 2
Profitability rate = 66%

<= 35

Total Population = 4
Profitable = 1
Unprofitable = 3
Profitability rate = 25%

$$\left(\frac{6}{10}\right) * 0.91 + \left(\frac{4}{10}\right) * 0.81$$

$\boxed{0.87}$

Total Population = 10
Profitable = 5
Unprofitable = 5
Profitability rate = 50%

Gender

Male

Total Population = 5
Profitable = 3
Unprofitable = 2
Profitability rate = 60%

Female

Total Population = 5
Profitable = 2
Unprofitable = 3
Profitability rate = 40%

$$\left(\frac{5}{10}\right) * 0.97 + \left(\frac{5}{10}\right) * 0.97$$

0.97

© Jigsaw Academy Education Pvt Ltd

Decision Tree: Algorithm Overview

For each split the purity metric is computed

Choose the lowest variable which results in lowest value of purity metric

Continue doing these till some **stopping criteria** is met

© Jigsaw Academy Education Pvt Ltd

Decision Tree: Algorithm Overview

Stopping Criteria

Depth of tree

Specifying the levels of the tree

Improvement in purity metric

Specifying the minimum change in purity metric from one split to another

Value in terminal node

Specifying the number of value in the terminal node

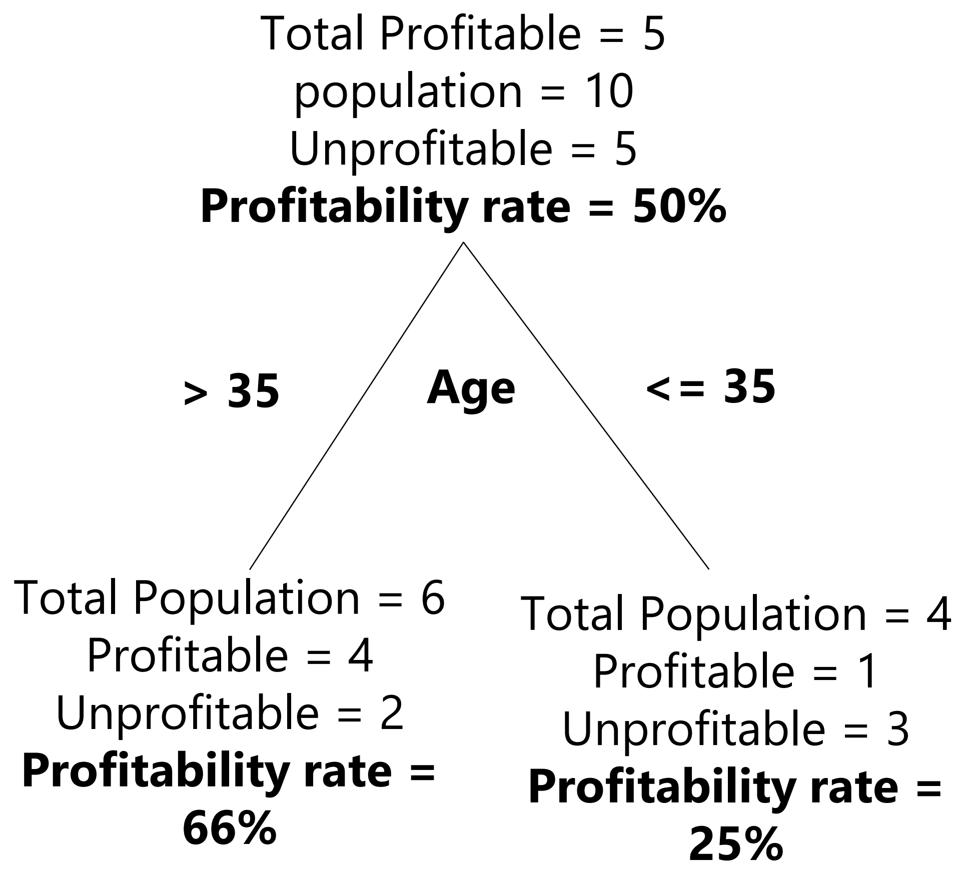
© Jigsaw Academy Education Pvt Ltd

Decision Tree: Prediction

Use decision tree classifier as prediction

Available data – 20 year old person

Prediction – 25% Chance of him being profitable



© Jigsaw Academy Education Pvt Ltd

Decision Tree: Performance Metrics

Decision tree classifier output probabilities

ROC curves

Confusion metrics

Performance of the decision tree classifier

Area under ROC curves

For multiclass problems, accuracy is used as a performance measure

© Jigsaw Academy Education Pvt Ltd

Decision Tree: Parameters and Hyperparameters

Parameters of
a decision tree

Data

Purity metric- Gini or Entropy?

Depth of the tree

These parameters are estimated using cross validation

At the model level of decision tree rules are decided for predicting probabilities or classes

© Jigsaw Academy Education Pvt Ltd

Recap

- Decision Tree Overview
- Decision Tree Algorithms – Gini and Entropy
- Decision Tree Performance Metrics
- Decision Tree Parameter and Hyperparameter

© Jigsaw Academy Education Pvt Ltd

MACHINE LEARNING

Algorithms

Class

Tree Based Models

Topic

Introduction to Regression Tree

Decision Tree: Regression

Decision Tree can be used to do regression tasks

When the target variable is continuous decision tree regressor can be used

Prediction → Mean value of the target variable

© Jigsaw Academy Education Pvt Ltd

Decision Tree: Regression

Example

Country	Rim	Tires	Type	Price
Japan	R14	195/60	Small	11.95
Japan	R15	205/60	Medium	24.76
Germany	R15	205/60	Medium	26.9
Germany	R14	175/60	Compact	18.9
Germany	R14	195/60	Compact	24.65
Germany	R15	225/60	Medium	33.2
USA	R14	185/75	Medium	13.15
USA	R14	205/75	Large	20.225
USA	R14	205/75	Large	16.145
USA	R15	205/70	Medium	23.04

Build a decision tree model to predict price

Price is a continuous variable

Regression tree

Recursively subset the data

© Jigsaw Academy Education Pvt Ltd

Decision Tree: Regression

Example

Country	Rim	Tires	Type	Price
Japan	R14	195/60	Small	11.95
Japan	R15	205/60	Medium	24.76
Germany	R15	205/60	Medium	26.9
Germany	R14	175/60	Compact	18.9
Germany	R14	195/60	Compact	24.65
Germany	R15	225/60	Medium	33.2
USA	R14	185/75	Medium	13.15
USA	R14	205/75	Large	20.225
USA	R14	205/75	Large	16.145
USA	R15	205/70	Medium	23.04

Total Population =10
Average price = 21.9

R14

Yes

Total Population = 6
Average Price= 17.50

Price
11.95
18.9
24.65
13.15
20.22
16.14

No

Total Population = 4
Average Price = 26.97

Price
24.76
26.90
33.20
23.04

© Jigsaw Academy Education Pvt Ltd

Decision Tree: Regression

Obtain the average value of continuous target variable

Total Population – 10
Average price – 21.9

Yes | **R14** | No

Total Population = 6
Average Price= 17.50

Total Population = 4
Average Price = 26.97

Yes | **Tire – 225/60** | No

Total Pop = 1
Average Price= 24.9

Total Pop = 3
Average Price= 33.2

A regression tree can also be summarised as a series of rules

© Jigsaw Academy Education Pvt Ltd

Purity Metrics

How does a regression tree algorithm pick up which variable to split on?

Predictions need to be accurate

The prediction is the average value of target variable in decision node

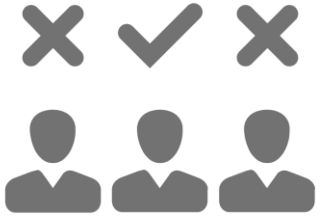
Higher the accuracy of prediction, the better the split is

Mean Squared Error (MSE) or Residual Sum of Square (RSS) as a proxy of accuracy in each node

© Jigsaw Academy Education Pvt Ltd

Purity Metrics

Country	Rim	Tires	Type	Price
Japan	R14	195/60	Small	11.95
Japan	R15	205/60	Medium	24.76
Germany	R15	205/60	Medium	26.9
Germany	R14	175/60	Compact	18.9
Germany	R14	195/60	Compact	24.65
Germany	R15	225/60	Medium	33.2
USA	R14	185/75	Medium	13.15
USA	R14	205/75	Large	20.225
USA	R14	205/75	Large	16.145
USA	R15	205/70	Medium	23.04



MSE or RSS helps in deciding which variable to choose for a split

© Jigsaw Academy Education Pvt Ltd

Purity Metrics

Country	Rim	Tires	Type	Price
Japan	R14	195/60	Small	11.95
Japan	R15	205/60	Medium	24.76
Germany	R15	205/60	Medium	26.9
Germany	R14	175/60	Compact	18.9
Germany	R14	195/60	Compact	24.65
Germany	R15	225/60	Medium	33.2
USA	R14	185/75	Medium	13.15
USA	R14	205/75	Large	20.225
USA	R14	205/75	Large	16.145
USA	R15	205/70	Medium	23.04

Total Population =10
Average price = 21.9

R14

Yes

Total Population = 6
Average Price= 17.50

Price
11.95
18.9
24.65
13.15
20.22
16.14

No

Total Population = 4
Average Price = 26.97

Price
24.76
26.90
33.20
23.04

© Jigsaw Academy Education Pvt Ltd

Purity Metrics

Country	Rim	Tires	Type	Price
Japan	R14	195/60	Small	11.95
Japan	R15	205/60	Medium	24.76
Germany	R15	205/60	Medium	26.9
Germany	R14	175/60	Compact	18.9
Germany	R14	195/60	Compact	24.65
Germany	R15	225/60	Medium	33.2
USA	R14	185/75	Medium	13.15
USA	R14	205/75	Large	20.225
USA	R14	205/75	Large	16.145
USA	R15	205/70	Medium	23.04

Total Population =10

Average price = 21.9

Country - Germany

Yes

Total Population = 4

Average Price= 25.91

Price
26.90
18.90
26.45
33.20

No

Total Population = 6

Average Price = 18.21

Price
11.95
24.76
13.15
20.22
16.14
23.04

© Jigsaw Academy Education Pvt Ltd

Purity Metric

Total Population =10
Average price = 21.9

Yes | **R14** | No

Total Population = 6
Average Price= 17.50

Total Population = 4
Average Price = 26.97

Total Population =10
Average price = 21.9

Yes | **Country - Germany** | No

Total Population = 4
Average Price= 25.91

Total Population = 6
Average Price = 18.21

Rim or Country?

Which variable helps in creating a more accurate prediction?

© Jigsaw Academy Education Pvt Ltd

Purity Metric

Total Population =10
Average price = 21.9

Yes | **R14** | No

Total Population = 6
Average Price= 17.50

Total Population = 4
Average Price = 26.97

Total Population =10
Average price = 21.9

Yes | **Country - Germany** | No

Total Population = 4
Average Price= 25.91

Total Population = 6
Average Price = 18.21

Use Mean Squared Error (MSE) or Residual Sum of Square (RSS)

$$MSE = \frac{1}{n}\sum(y_i - \mu)^2$$

MSE is just the average of RSS

Nothing but variance in the values of target in variable in a node

© Jigsaw Academy Education Pvt Ltd

Purity Metric

Total Population =10
Average price = 21.9

R14

Yes

Total Population = 6
Average Price= 17.50

Price	Pred.
11.95	17.50
18.9	17.50
24.65	17.50
13.15	17.50
20.22	17.50
16.14	17.50

No

Total Population = 4
Average Price = 26.97

Price	Pred.
24.76	26.97
26.90	26..97
33.20	26.97
23.04	26.97

Total Population =10
Average price = 21.9

Country - Germany

Yes

Total Population = 4
Average Price= 25.91

Price	Pred.
26.90	25.91
18.90	25.91
26.45	25.91
33.20	25.91

No

Total Population = 6
Average Price = 18.21

Price	Pred.
11.95	18.21
24.76	18.21
13.15	18.21
20.22	18.21
16.14	18.21
23.04	18.21

MSE tries to find out how accurate a prediction is in each node

© Jigsaw Academy Education Pvt Ltd

Purity Metric

$$MSE = \frac{1}{n}\sum(y_i - \mu)^2$$

R14

Total Population =10
Average price = 21.9

Yes

Total Population = 6
Average Price= 17.50

Price	Pred.
11.95	17.50
18.9	17.50
24.65	17.50
13.15	17.50
20.22	17.50
16.14	17.50

$$\frac{1}{6}(11.95 - 17.50)^2 + (18.90 - 17.50)^2 + .. + (16.14 - 17.50)^2$$

No

Total Population = 4
Average Price = 26.97

Price	Pred.
24.76	26.97
26.90	26..97
33.20	26.97
23.04	26.97

$$\frac{1}{4}(24.76 - 26.97)^2 + (26.90 - 26.97)^2 + .. + (23.04 - 26.97)^2$$

Country - Germany

Total Population =10
Average price = 21.9

Yes

Total Population = 4
Average Price= 25.91

Price	Pred.
26.90	25.91
18.90	25.91
26.45	25.91
33.20	25.91

No

Total Population = 6
Average Price = 18.21

Price	Pred.
11.95	18.21
24.76	18.21
13.15	18.21
20.22	18.21
16.14	18.21
23.04	18.21

© Jigsaw Academy Education Pvt Ltd

Purity Metric

$$MSE = \frac{1}{n}\sum(y_i - \mu)^2$$

Total Population =10
Average price = 21.9

R14

Yes

Total Population = 6
Average Price= 17.50

MSE – 18.67

No

Total Population = 4
Average Price = 26.97

MSE – 14.78

$$\frac{6}{10} * 18.67 + \frac{4}{10} * 14.78 \quad \boxed{= 17.114}$$

Total Population =10
Average price = 21.9

Country - Germany

Yes

Total Population = 4
Average Price= 25.91

MSE – 26.21

No

Total Population = 6
Average Price = 18.21

MSE – 23.22

$$\frac{4}{10} * 26.21 + \frac{6}{10} * 23.22 \quad = 24.416$$

Rim is better than country at producing more accurate predictions

© Jigsaw Academy Education Pvt Ltd

Hyperparameters

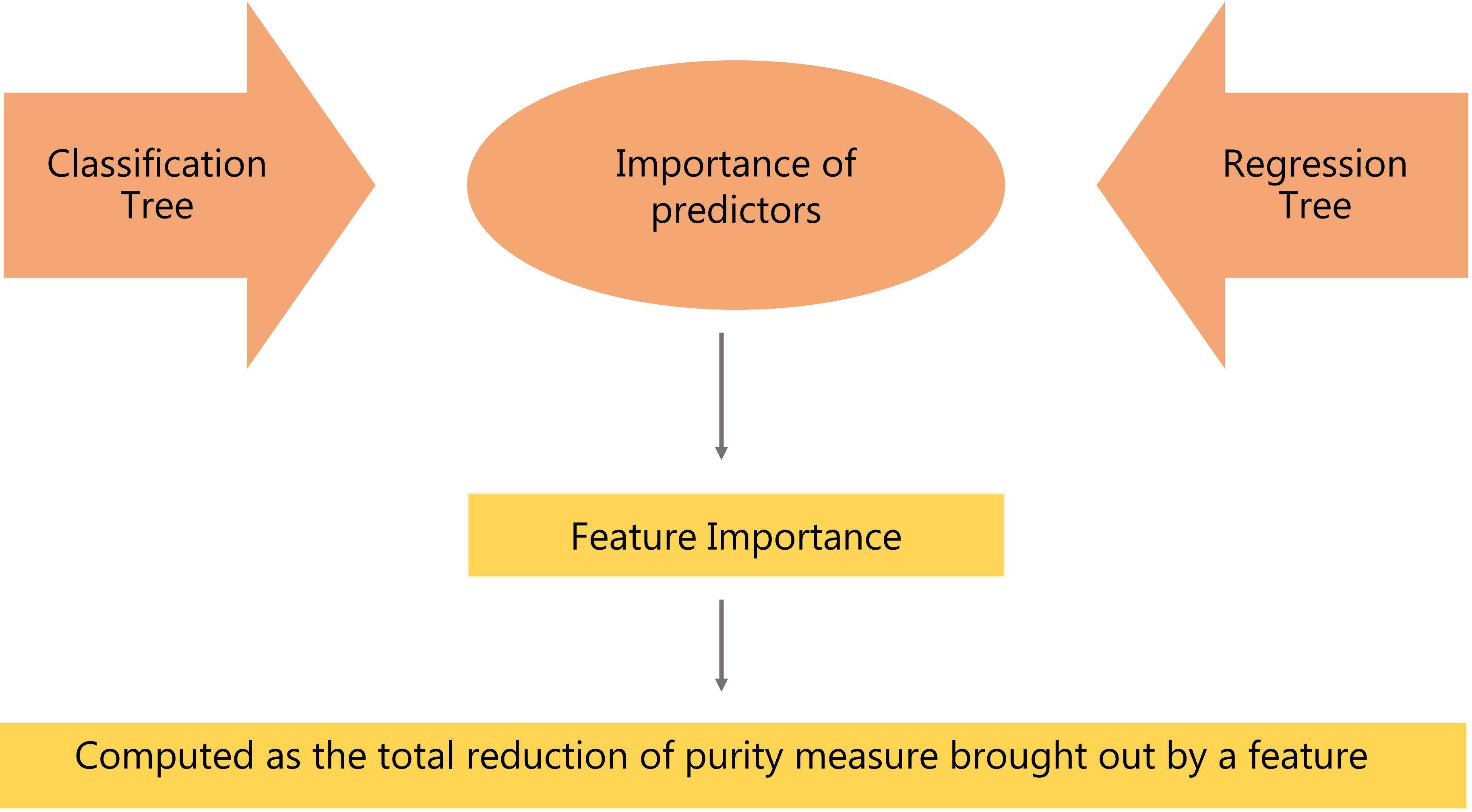
Regression Tree

- Depth of tree
- Number of observations in terminal node

Grid search procedure to compute the appropriate values of these hyperparameters

© Jigsaw Academy Education Pvt Ltd

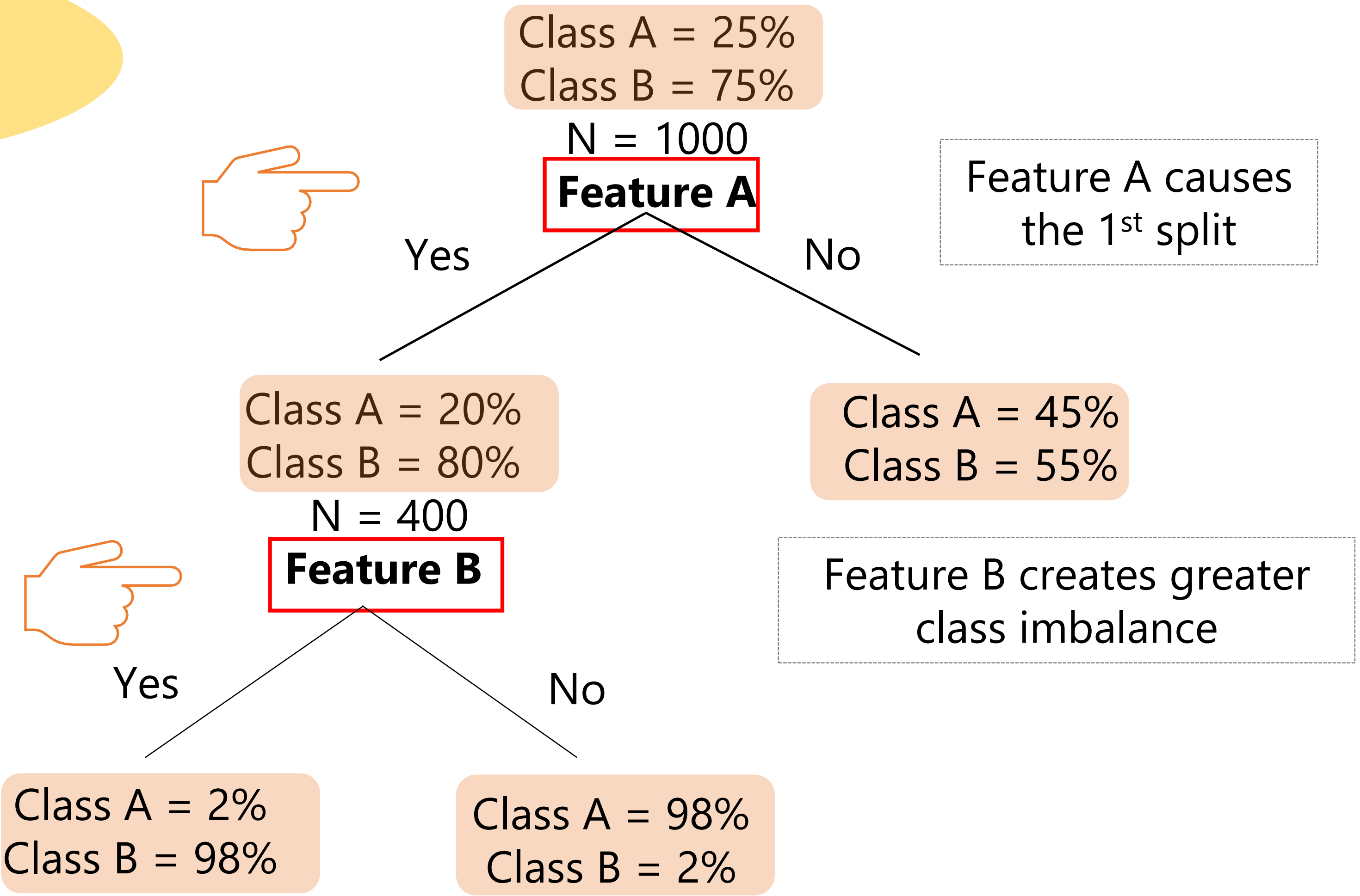
Feature Importance



© Jigsaw Academy Education Pvt Ltd

Feature Importance

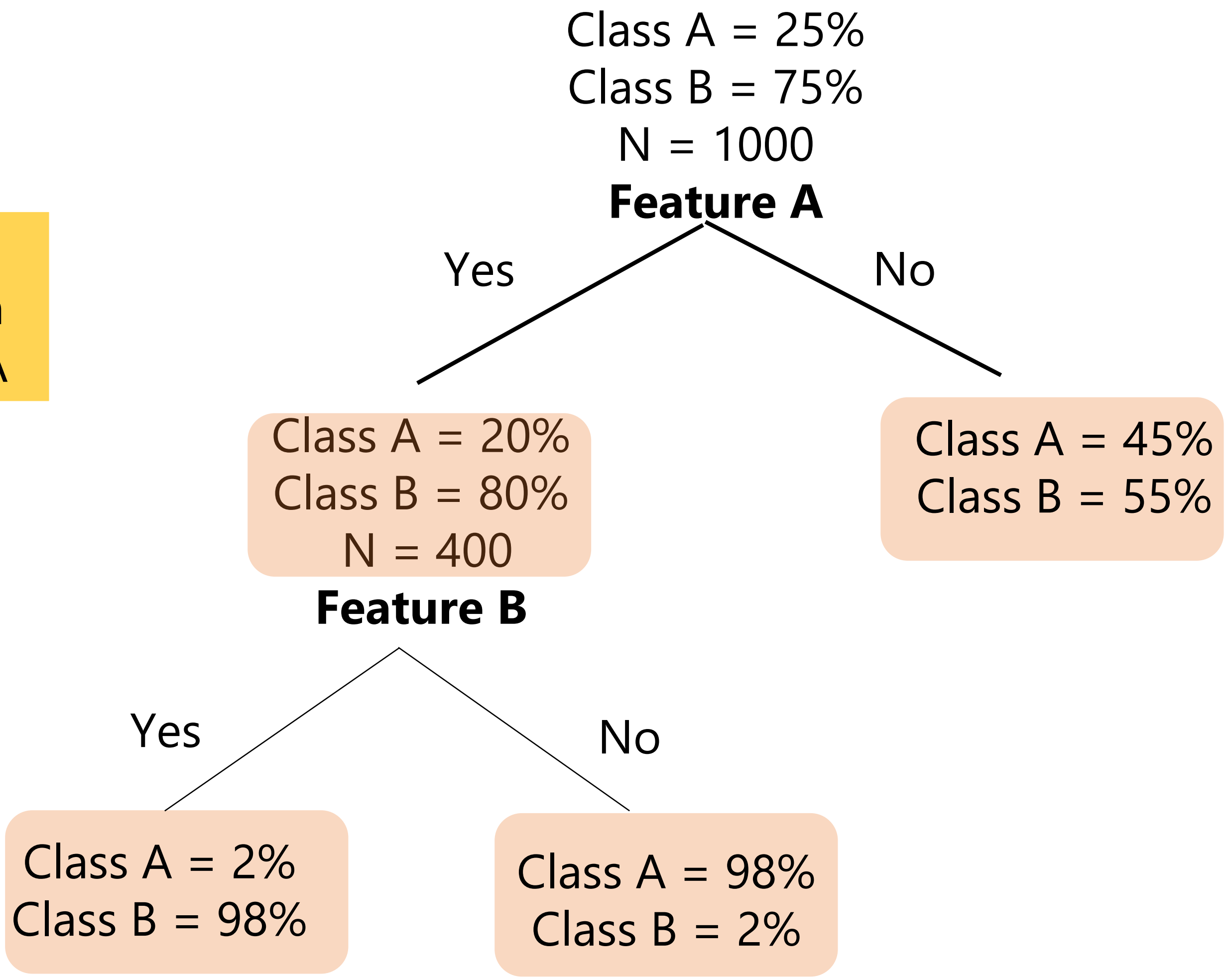
Which feature is more important?



© Jigsaw Academy Education Pvt Ltd

Feature Importance

Proportion of classes are **more disproportionate** in Feature B than in Feature A



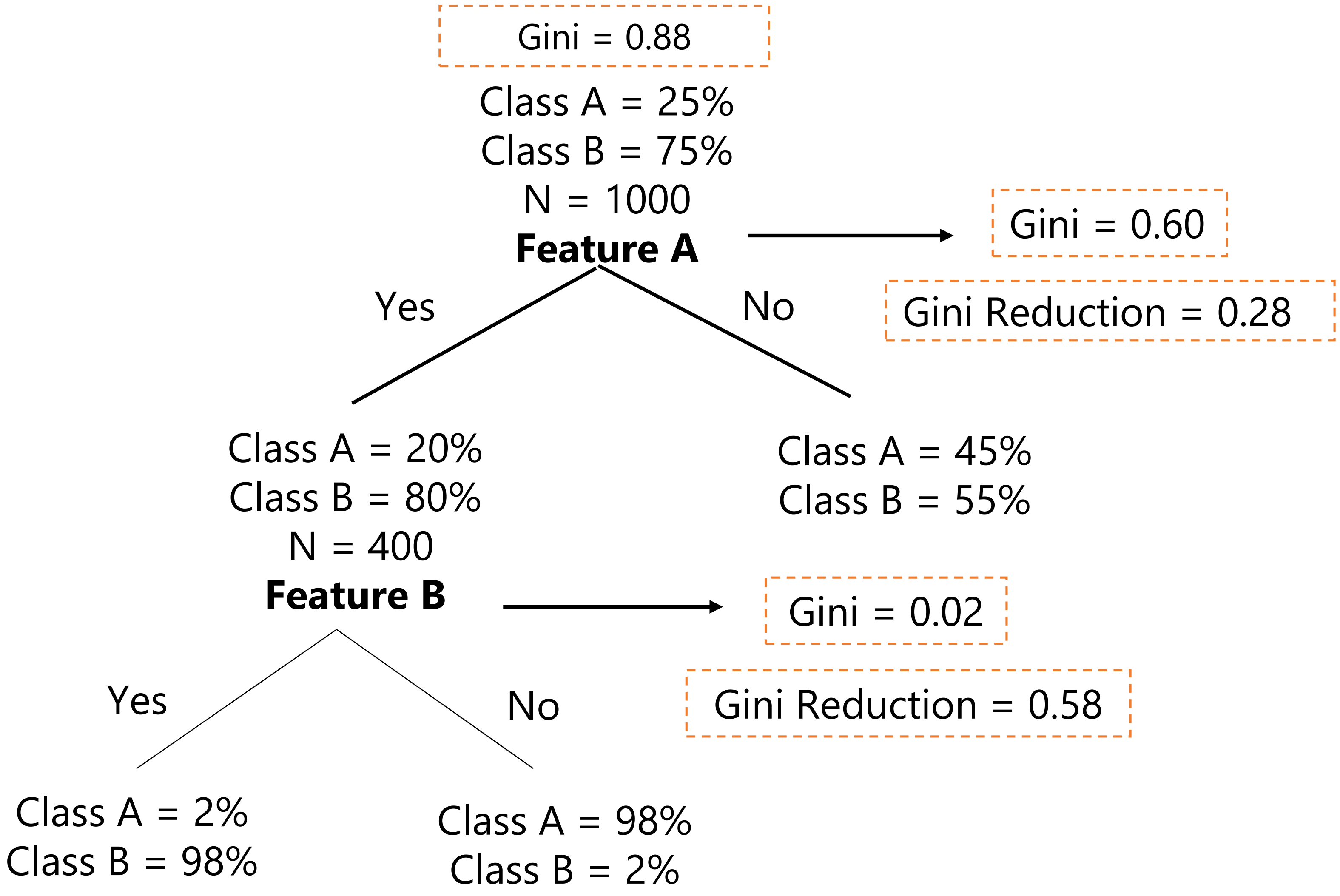
© Jigsaw Academy Education Pvt Ltd

Feature Importance

In Variable Importance both the sequence of the split and the purity of a node should be considered

Feature A precedes Feature B

Feature B creates greater node purity



© Jigsaw Academy Education Pvt Ltd

Feature Importance

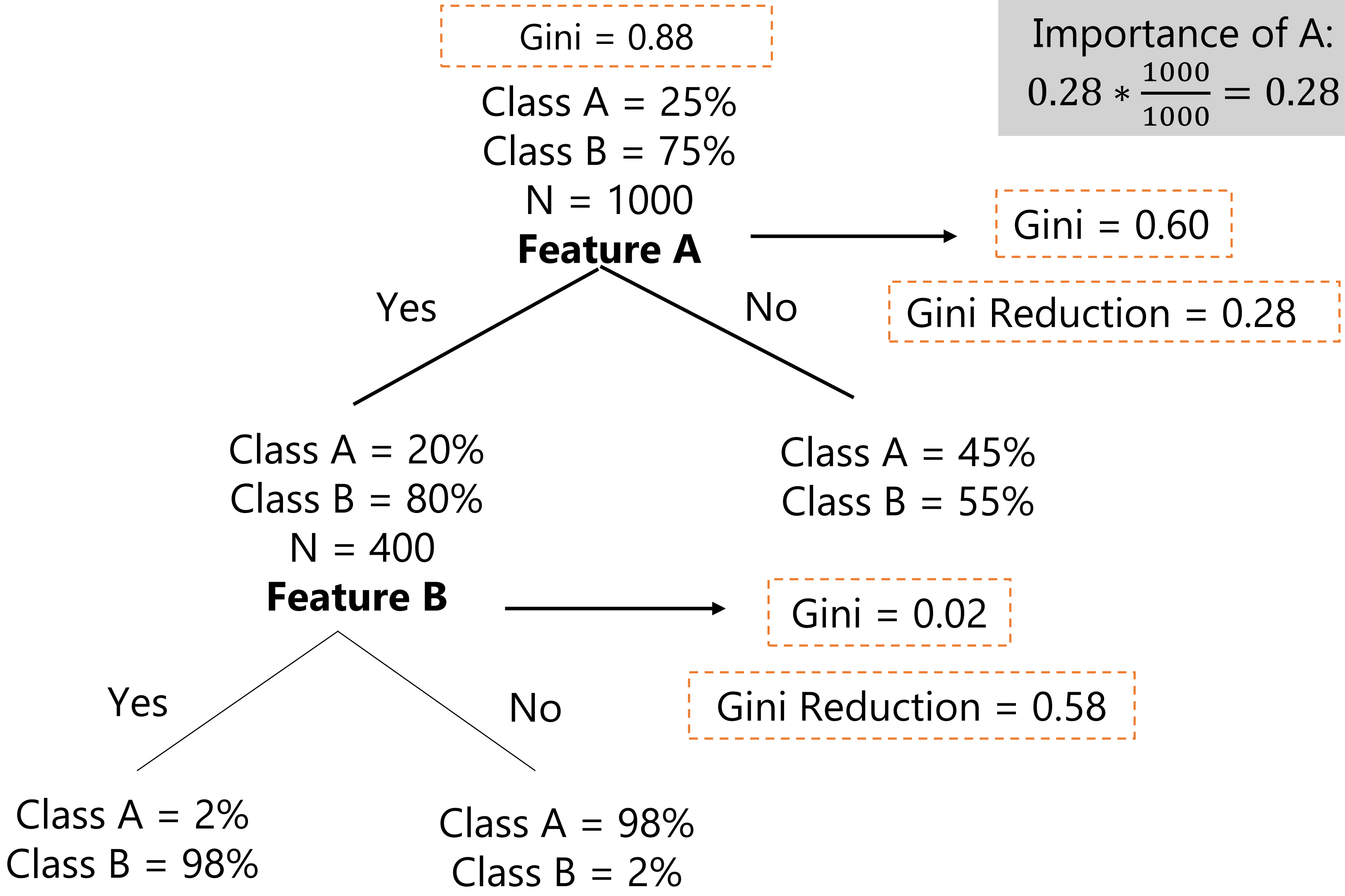
Importance of A: Decrease in Gini * Proportion of data

Decrease in Gini
Ability of a variable to create class imbalance compared to preceding split

Proportion of data
Sequence in which variable causes the split

More observations will pass through the node caused by an early split

Importance of A:
$0.28 * \frac{1000}{1000} = 0.28$

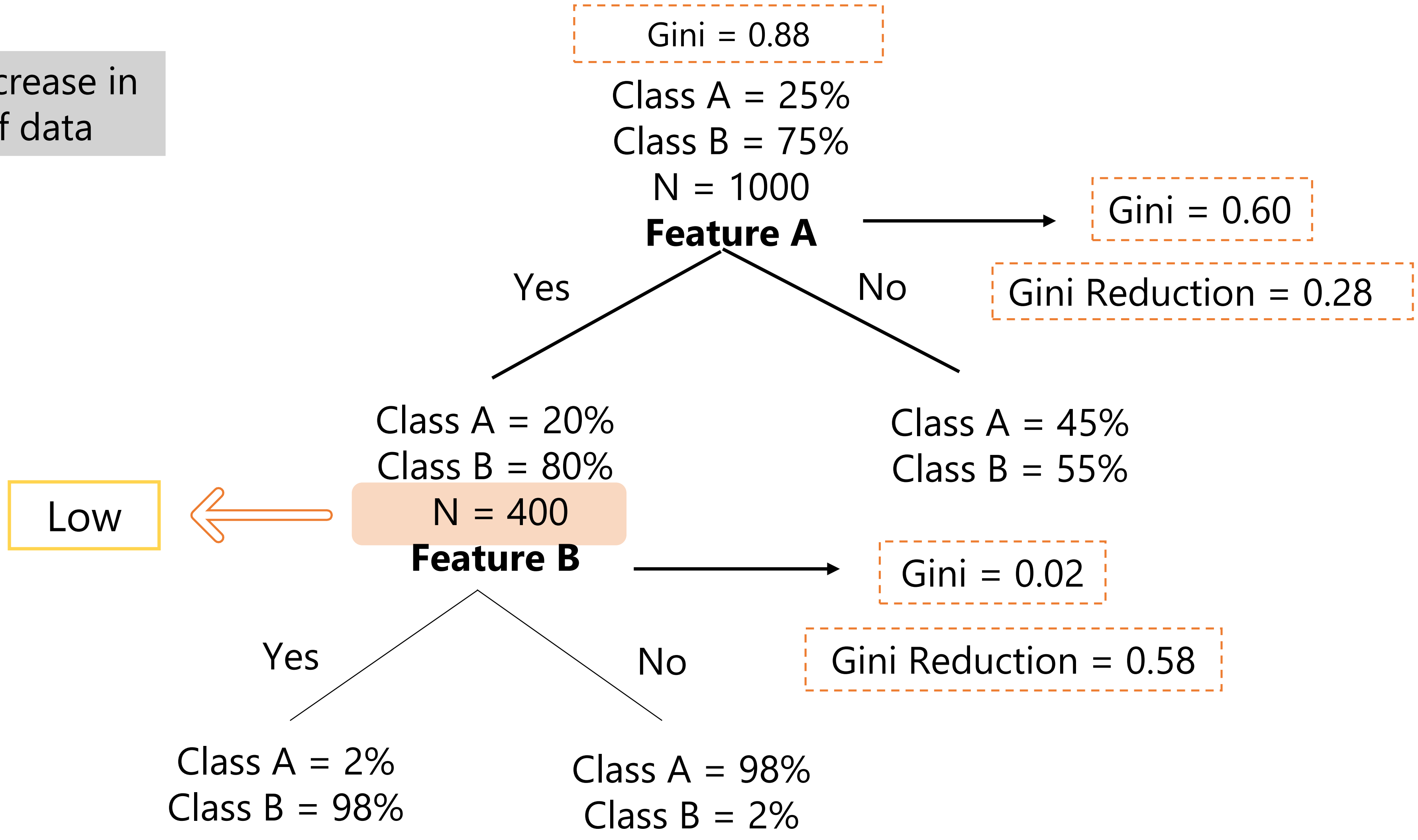


© Jigsaw Academy Education Pvt Ltd

Feature Importance

Importance of B: Decrease in Gini*Proportion of data

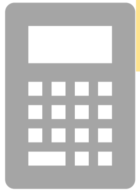
Importance of B:
$$0.58 * \frac{400}{1000} = 0.23$$



© Jigsaw Academy Education Pvt Ltd

Feature Importance

Weigh the decrease in Mean Squared Error and Residual Sum Square appropriately



Feature Importance

© Jigsaw Academy Education Pvt Ltd

Recap

1. Decision tree – Regression
2. Purity Metric
3. Hyperparameters
4. Feature Importance

© Jigsaw Academy Education Pvt Ltd

MACHINE LEARNING

Algorithms