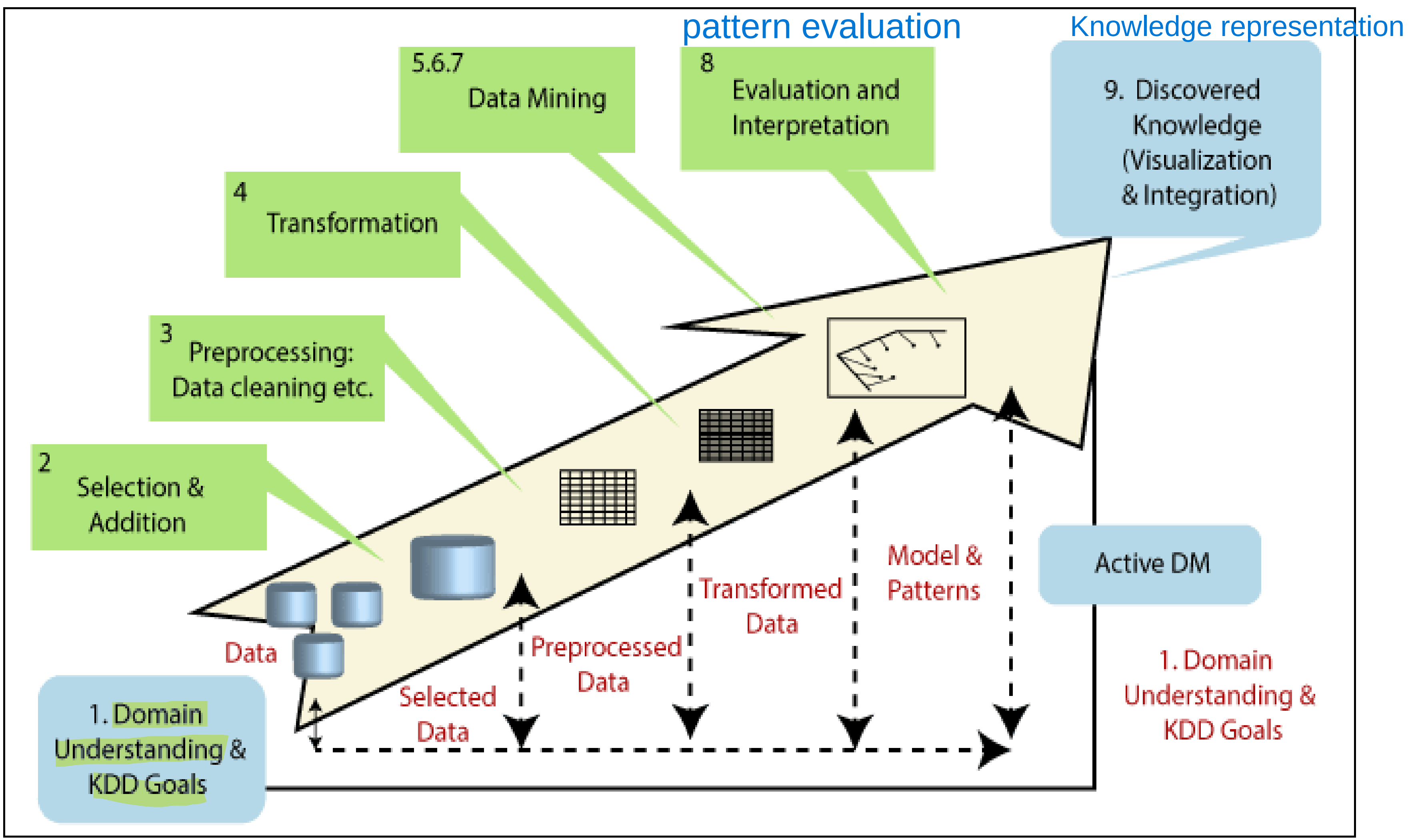
Introduction to Classification

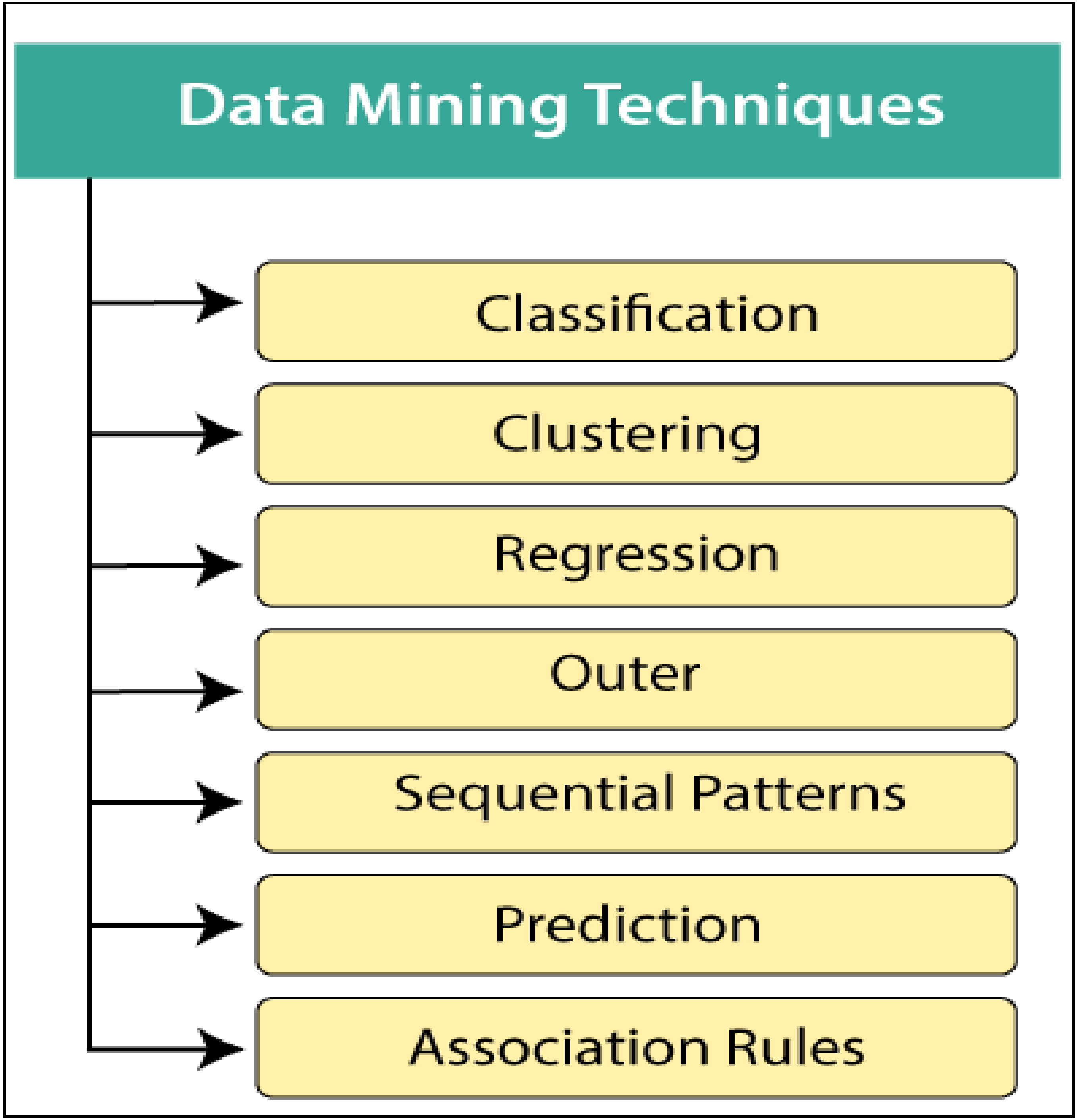
Tree based Classification

ID3 algorithm

Steps involved in KDD

Knowledge Data Discovery




Data Mining Techniques
Classification
Clustering
Regression
Outer
Sequential Patterns
Prediction
Association Rules

Classification: Definition

- Given a collection of records (*training set*)
  - Each record contains a set of *attributes*,
  - one of the attributes is the *class*.
- Find a *model* for class attribute as a function of the values of other attributes.

- Goal: previously unseen records should be assigned a class as accurately as possible.
  - A *test set* is used to determine the accuracy of the model.
  - Usually, the given data set is divided into training and test sets, with training set used to build the model and test set used to validate it. and determine the accuracy of the model

Classification: Definition

- Given a collection of records (training set)
  - Each record is by characterized by a tuple (***x***,*y*), where ***x*** is the attribute set and *y* is the class label
    - ***x***: attribute, predictor, independent variable, input
    - *y*: class, response, dependent variable, output

- Task:
  - **Learn a model that** maps each attribute set ***x*** into one of the predefined class labels *y*

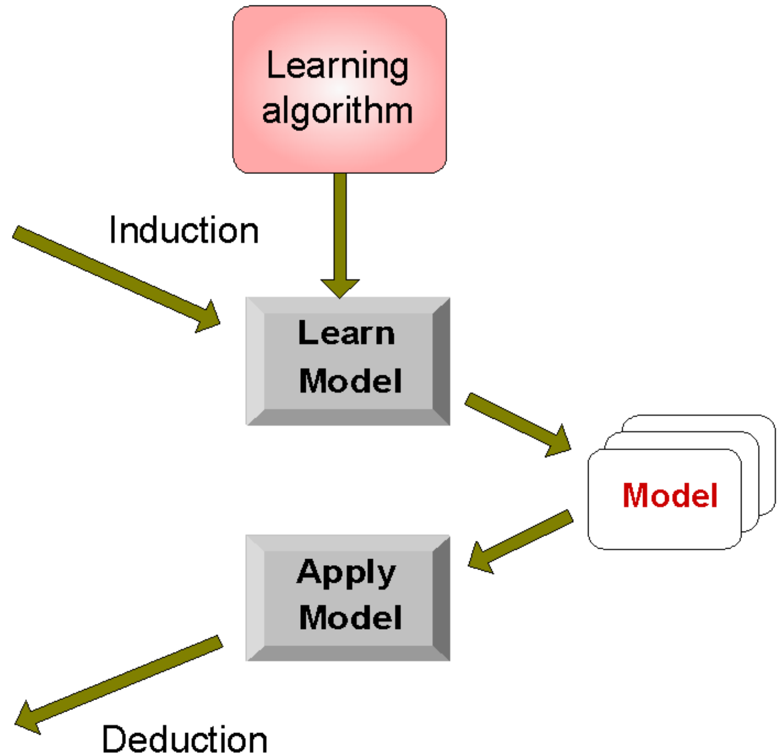
Prediction: Classification vs. Regression

- Classification
  - predicts categorical class labels (discrete or nominal)
  - classifies data (constructs a model) based on the training set and the values (class labels) in a classifying attribute and uses it in classifying new data
- Regression
  - models continuous-valued functions, i.e., predicts unknown or missing values

Illustrating Classification Task

Tid	Attrib1	Attrib2	Attrib3	Class
1	Yes	Large	125K	No
2	No	Medium	100K	No
3	No	Small	70K	No
4	Yes	Medium	120K	No
5	No	Large	95K	Yes
6	No	Medium	60K	No
7	Yes	Large	220K	No
8	No	Small	85K	Yes
9	No	Medium	75K	No
10	No	Small	90K	Yes

Training Set



Tid	Attrib1	Attrib2	Attrib3	Class
11	No	Small	55K	?
12	Yes	Medium	80K	?
13	Yes	Large	110K	?
14	No	Small	95K	?
15	No	Large	67K	?

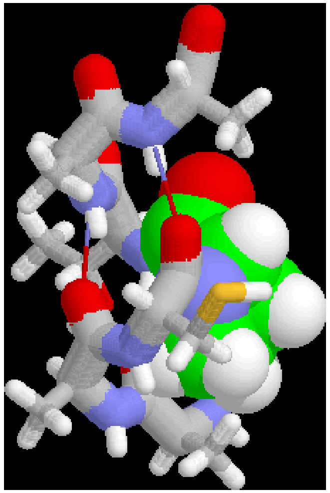
Test Set

Examples of Classification Task

- Classifying tumor cells as benign or malignant

- Classifying credit card transactions as legitimate or fraudulent

- Classifying secondary structures of protein as alpha-helix, beta-sheet, or random coil



- Categorizing news stories as finance, weather, entertainment, sports, etc

Learning Algorithm

- Probabilistic Functions (Bayesian Classifier)
- Functions to partitioning Vector Space
  - Non-Linear: KNN, Neural Networks, ...
  - Linear: Support Vector Machines, Perceptron, ...
- Boolean Functions (Decision Trees)

Classification Techniques

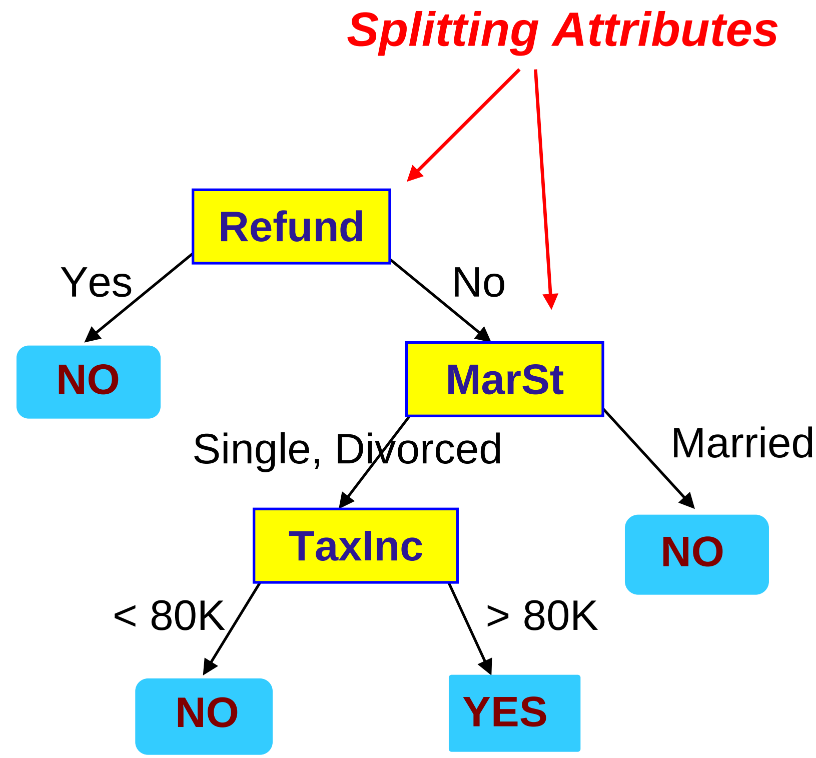
- Base Classifiers
  - Decision Tree based Methods
  - Rule-based Methods
  - Nearest-neighbor
  - Neural Networks
  - Deep Learning
  - Naïve Bayes and Bayesian Belief Networks
  - Support Vector Machines
- Ensemble Classifiers
  - Boosting, Bagging, Random Forests

Example of a Decision Tree

categorical categorical continuous class

Tid	Refund	Marital Status	Taxable Income	Cheat
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

Training Data

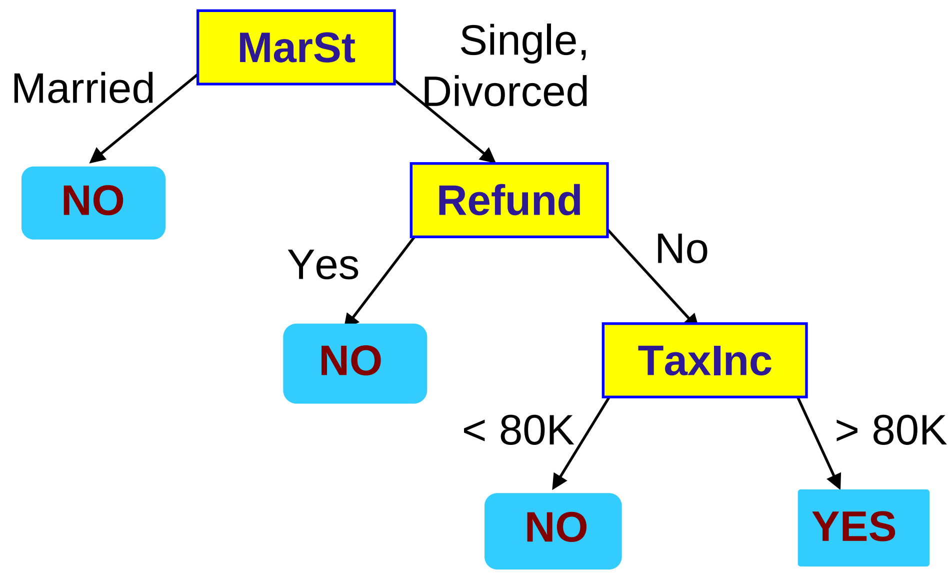


Model: Decision Tree

Another Example of Decision Tree

categorical categorical continuous class

Tid	Refund	Marital Status	Taxable Income	Cheat
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes



There could be more than one tree that fits the same data!

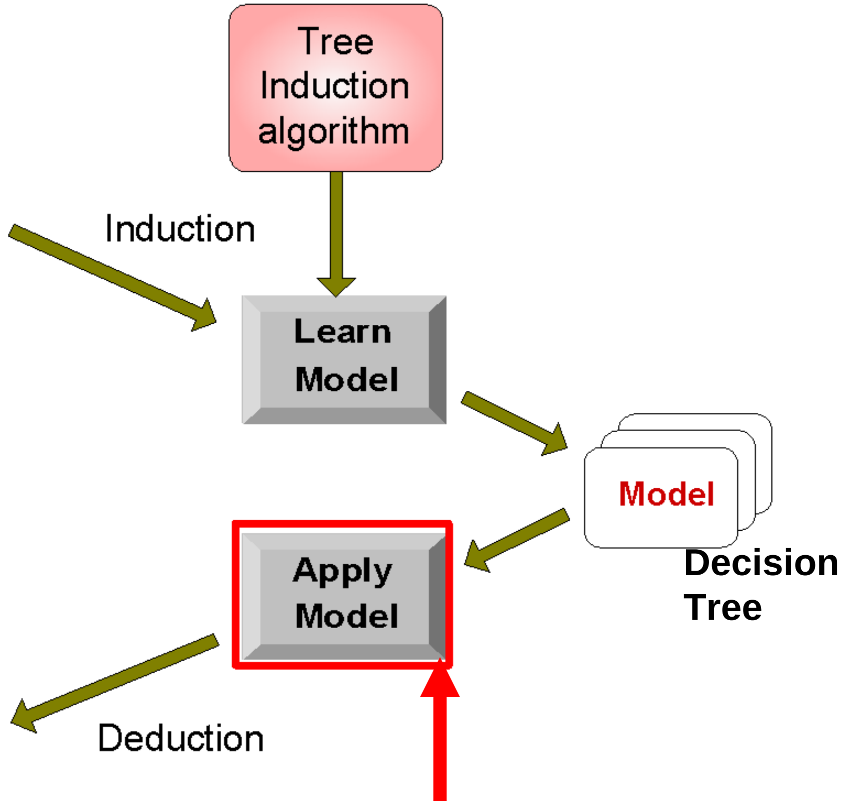
Decision Tree Classification Task

Tid	Attrib1	Attrib2	Attrib3	Class
1	Yes	Large	125K	No
2	No	Medium	100K	No
3	No	Small	70K	No
4	Yes	Medium	120K	No
5	No	Large	95K	Yes
6	No	Medium	60K	No
7	Yes	Large	220K	No
8	No	Small	85K	Yes
9	No	Medium	75K	No
10	No	Small	90K	Yes

Training Set

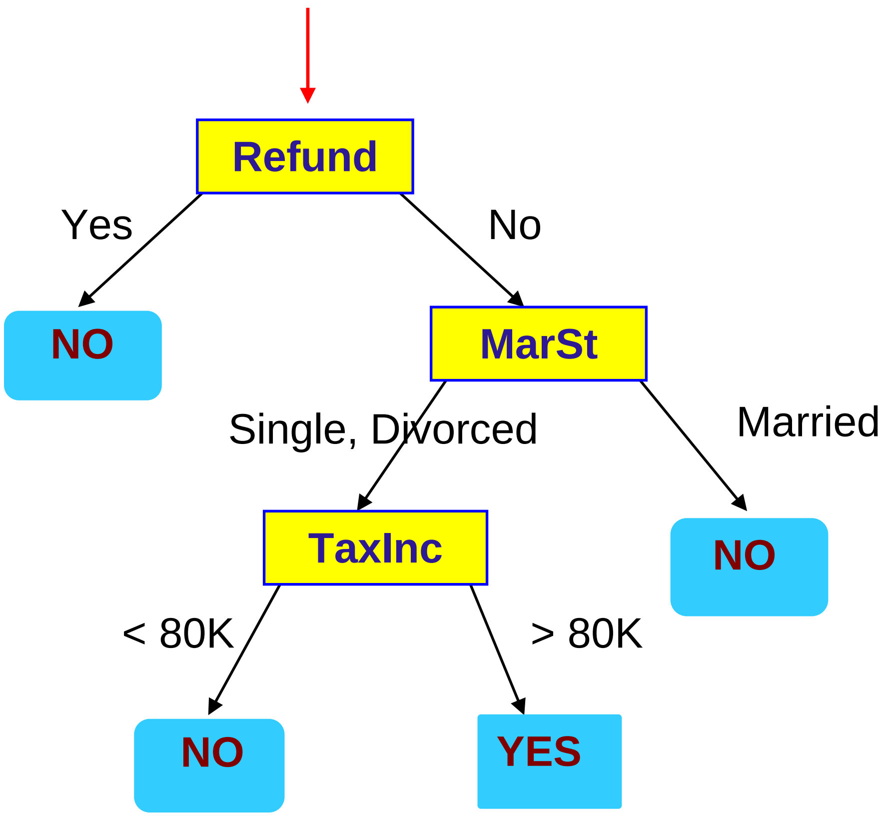
Tid	Attrib1	Attrib2	Attrib3	Class
11	No	Small	55K	?
12	Yes	Medium	80K	?
13	Yes	Large	110K	?
14	No	Small	95K	?
15	No	Large	67K	?

Test Set



Apply Model to Test Data

Start from the root of tree.



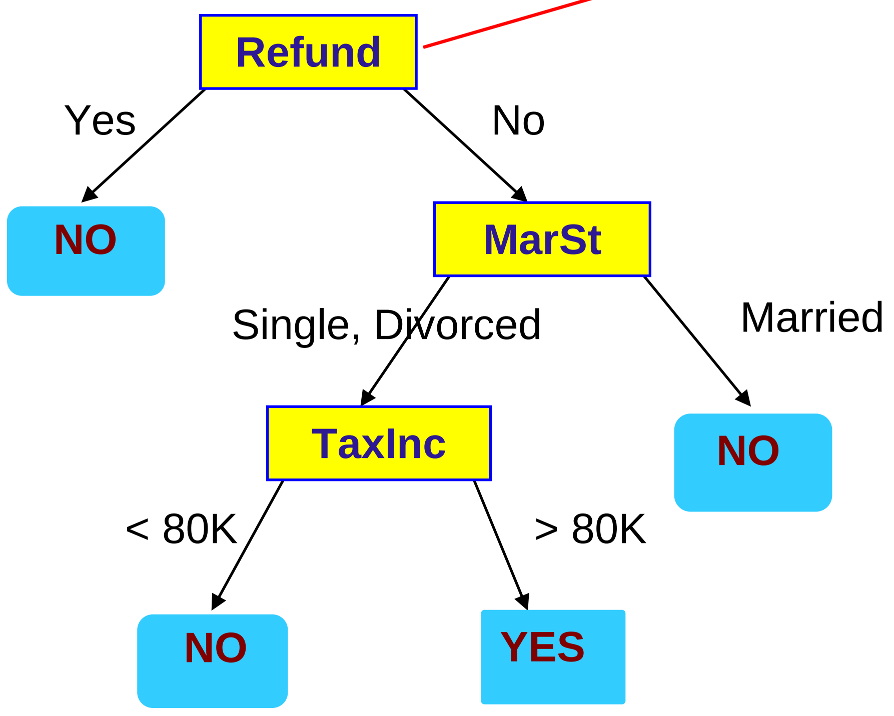
Test Data

Refund	Marital Status	Taxable Income	Cheat
No	Married	80K	?

Apply Model to Test Data

Test Data

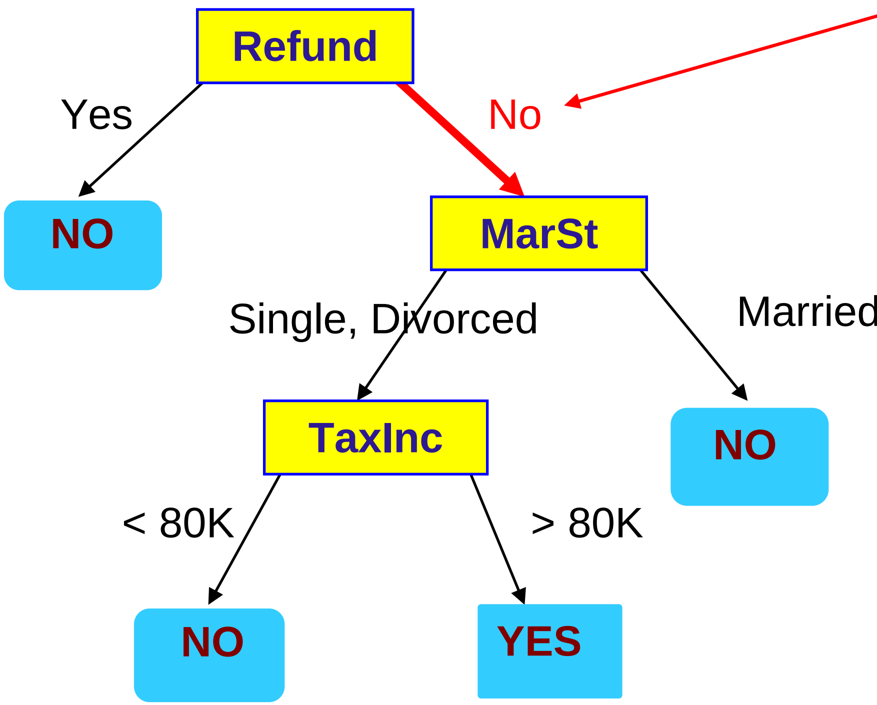
Refund	Marital Status	Taxable Income	Cheat
No	Married	80K	?



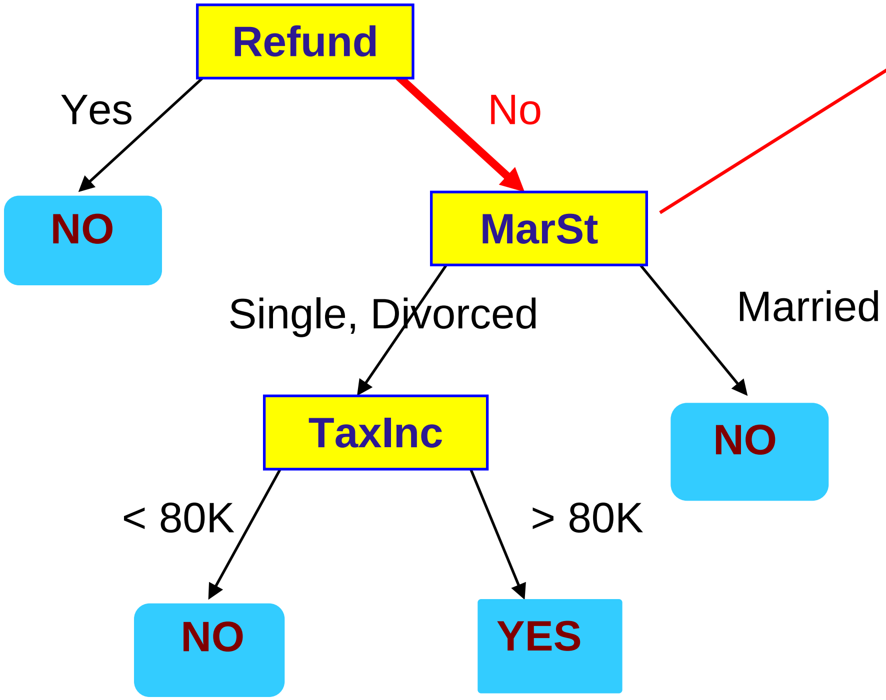
Apply Model to Test Data

Test Data

Refund	Marital Status	Taxable Income	Cheat
No	Married	80K	?



Apply Model to Test Data



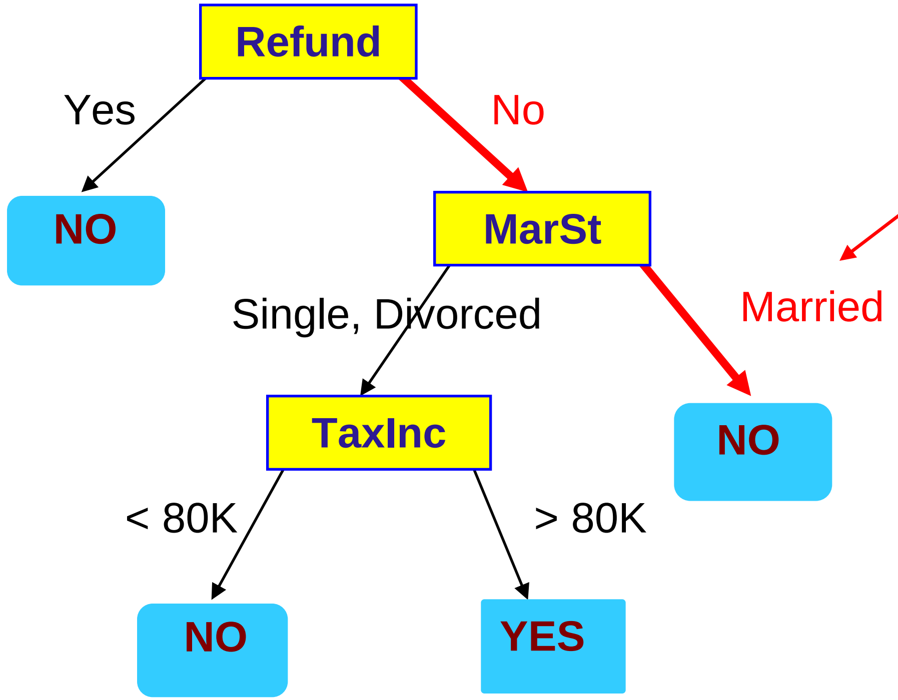
Test Data

Refund	Marital Status	Taxable Income	Cheat
No	Married	80K	?

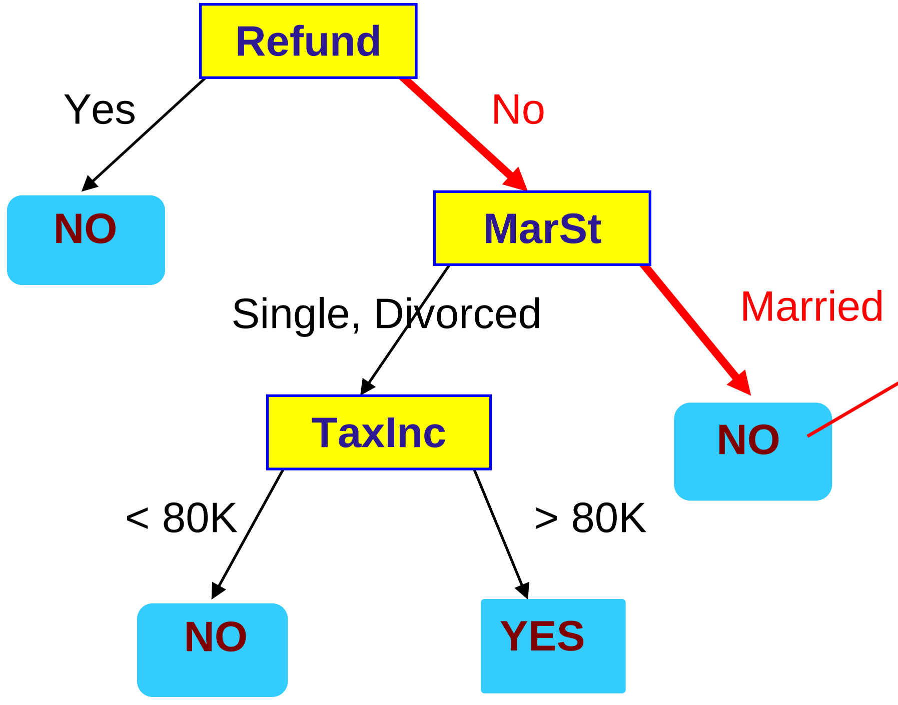
Apply Model to Test Data

Test Data

Refund	Marital Status	Taxable Income	Cheat
No	Married	80K	?



Apply Model to Test Data



Test Data

Refund	Marital Status	Taxable Income	Cheat
No	Married	80K	?

Assign Cheat to “No”

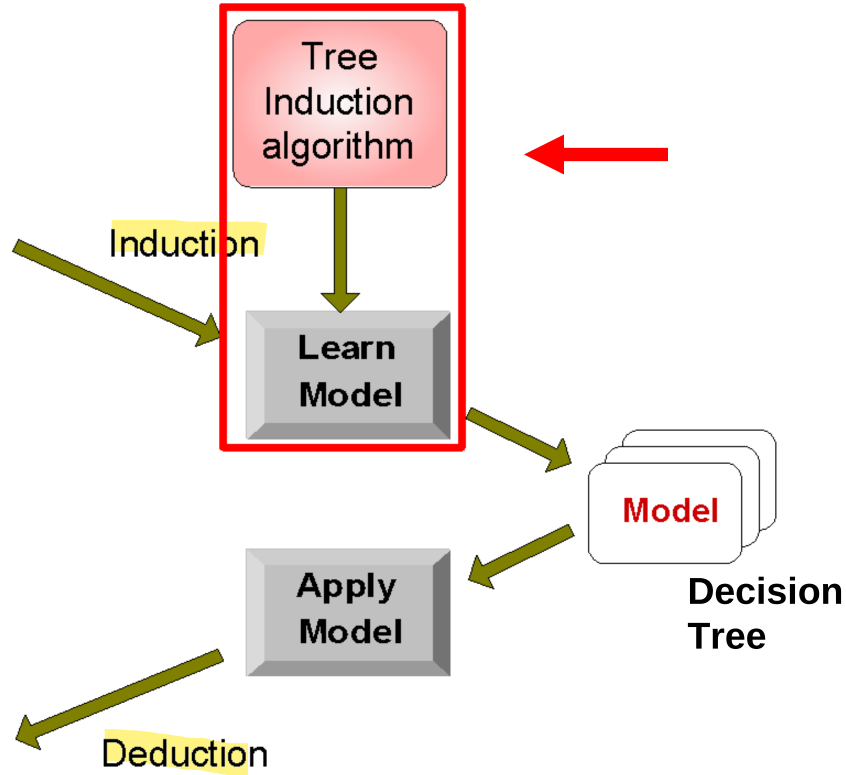
Decision Tree Classification Task

Tid	Attrib1	Attrib2	Attrib3	Class
1	Yes	Large	125K	No
2	No	Medium	100K	No
3	No	Small	70K	No
4	Yes	Medium	120K	No
5	No	Large	95K	Yes
6	No	Medium	60K	No
7	Yes	Large	220K	No
8	No	Small	85K	Yes
9	No	Medium	75K	No
10	No	Small	90K	Yes

Training Set

Tid	Attrib1	Attrib2	Attrib3	Class
11	No	Small	55K	?
12	Yes	Medium	80K	?
13	Yes	Large	110K	?
14	No	Small	95K	?
15	No	Large	67K	?

Test Set



Classification-Decision Tree

Decision Tree

Age	Competition	Type	Profit
old	Yes	S/w	Down
old	No	S/w	Down
old	No	H/w	Down
mid	Yes	S/w	Down
mid	Yes	H/w	Down
mid	No	H/w	Up
mid	No	S/w	Up
new	Yes	S/w	Up
new	No	H/w	Up
new	No	S/w	Up

$$I.G = -\frac{P}{P+N}\log_2\left(\frac{P}{P+N}\right) - \frac{N}{P+N}\log_2\left(\frac{N}{P+N}\right)$$

$$E(A) = \sum_{i=1}^{v} \frac{P_i + N_i}{P+N} I(P_i N_i)$$

$$Gain = I.G - E(A)$$

$$\log_2 x = \frac{\log_{10} x}{\log_{10} 2}$$

Age :-

	Down	Up
old	3	0
mid	2	2
new	0	3

$$I(old) = -\left[\frac{3}{3}\log_2\left(\frac{3}{3}\right) + \frac{0}{3}\log_2\left(\frac{0}{3}\right)\right] = 0 \quad \times 3/10 = 0$$

$$I(mid) = -\left[\frac{2}{4}\log_2\left(\frac{2}{4}\right) + \frac{2}{4}\log_2\left(\frac{2}{4}\right)\right] = 1 \quad \times 4/10 = 0.4$$

$$I(new) = -\left[\frac{0}{3}\log_2\left(\frac{0}{3}\right) + \frac{3}{3}\log_2\left(\frac{3}{3}\right)\right] = 0 \quad \times 3/10 = 0$$

$$E(Age) = 0.4$$

$$I.G = -\left[\frac{5}{10}\log_2\left(\frac{5}{10}\right) + \frac{5}{10}\log_2\left(\frac{5}{10}\right)\right]$$

$$= -\left[0.5 \times \log_2 2^{-1} + 0.5 \log_2 2^{-1}\right]$$

$$= -\left[0.5 \times (-1 \log_2 2) + 0.5 \times (-1 \log_2 2)\right]$$

$$= -\left[-0.5 - 0.5\right] = -\left[-1\right]$$

$$I.G = 1$$

$$Gain = 1 - 0.4$$
$$= 0.6$$

Decision Tree

Age	Competition	Type	Profit
old	Yes	S/W	Down
old	NO	S/W	Down
old	NO	H/W	Down
mid	Yes	S/W	Down
mid	Yes	H/W	Down
mid	NO	H/W	Up
mid	NO	S/W	Up
new	Yes	S/W	Up
new	NO	H/W	Up
new	NO	S/W	Up

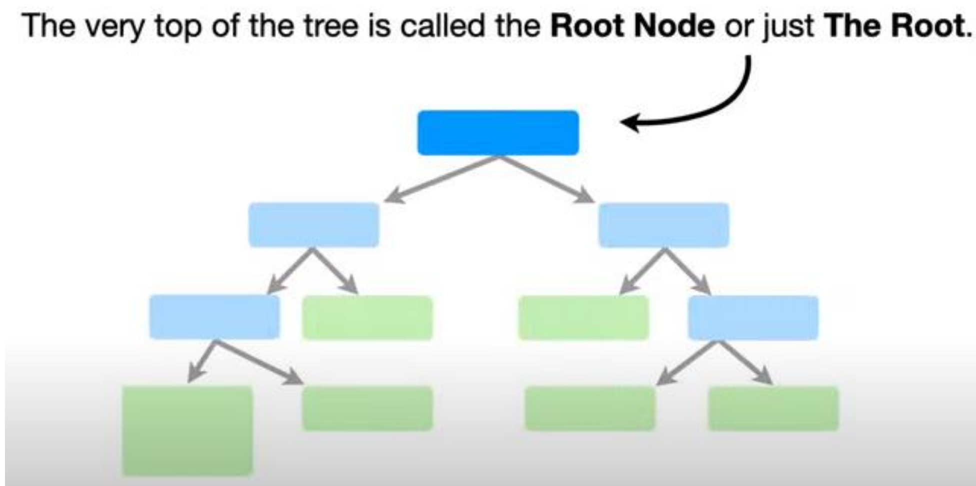
Gain(Age) → 0.60

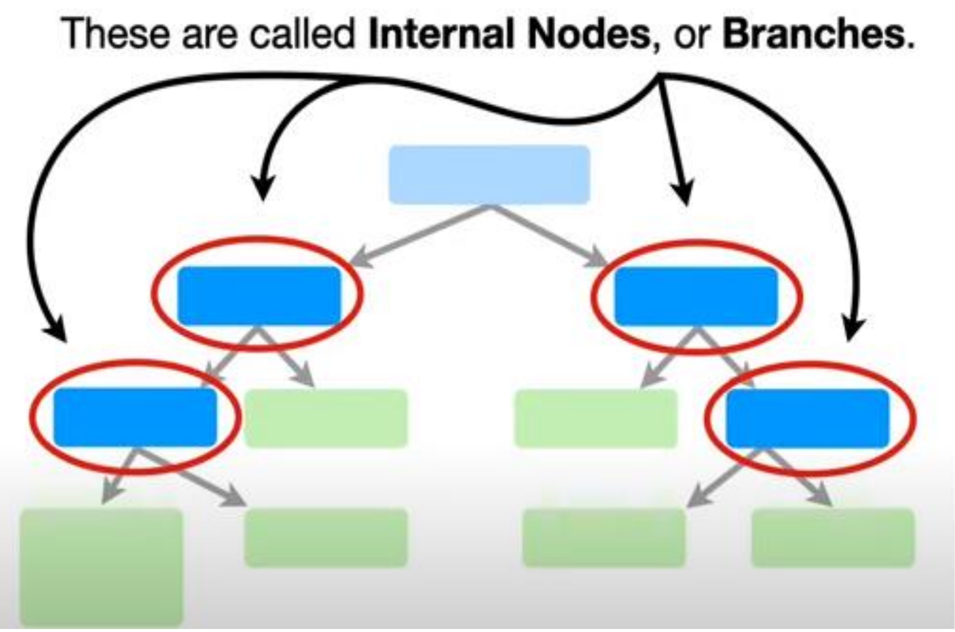
Gain(Competition) → 0.124

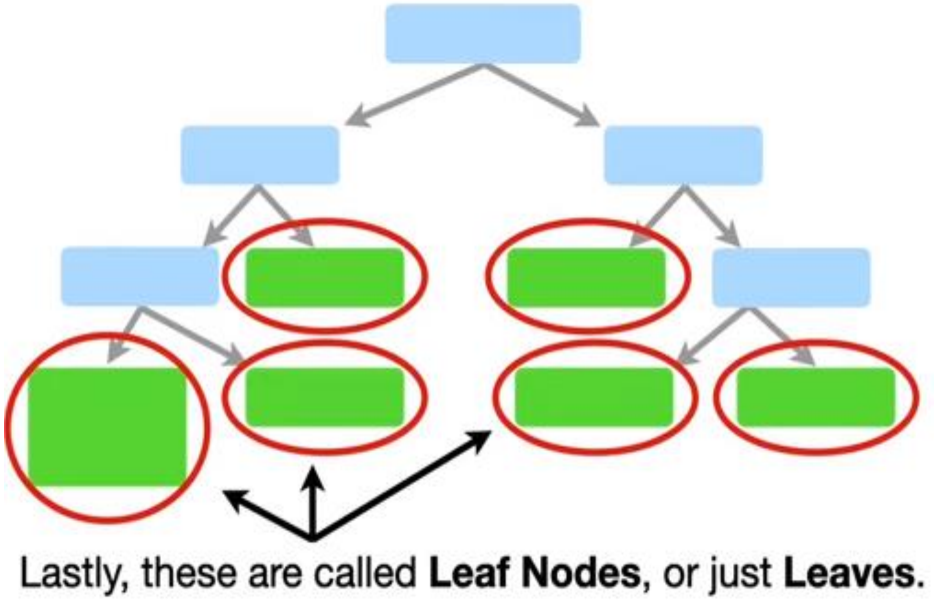
Gain(Type) → 0

I.G = 1

Decision Tree



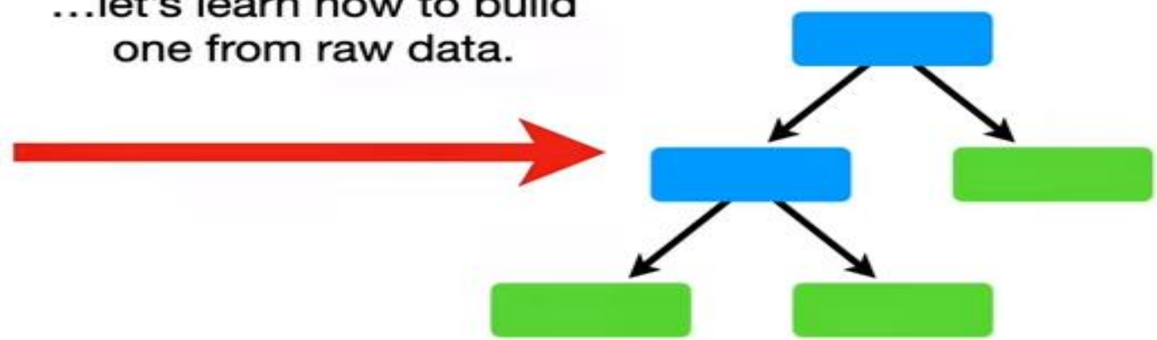




Decision Tree

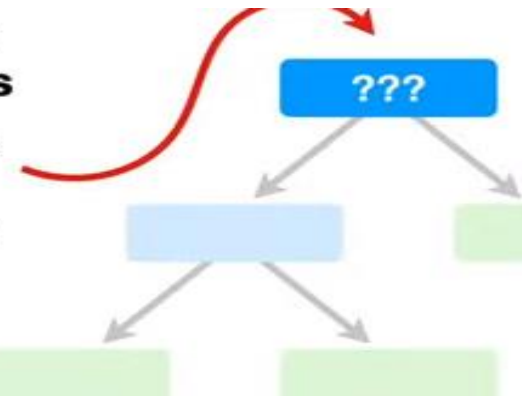
Loves Popcorn	Loves Soda	Age	Loves Cool As Ice
Yes	Yes	7	No
Yes	No	12	No
No	Yes	18	Yes
No	Yes	35	Yes
Yes	Yes	38	Yes
Yes	No	50	No
No	No	83	No

…let's learn how to build one from raw data.



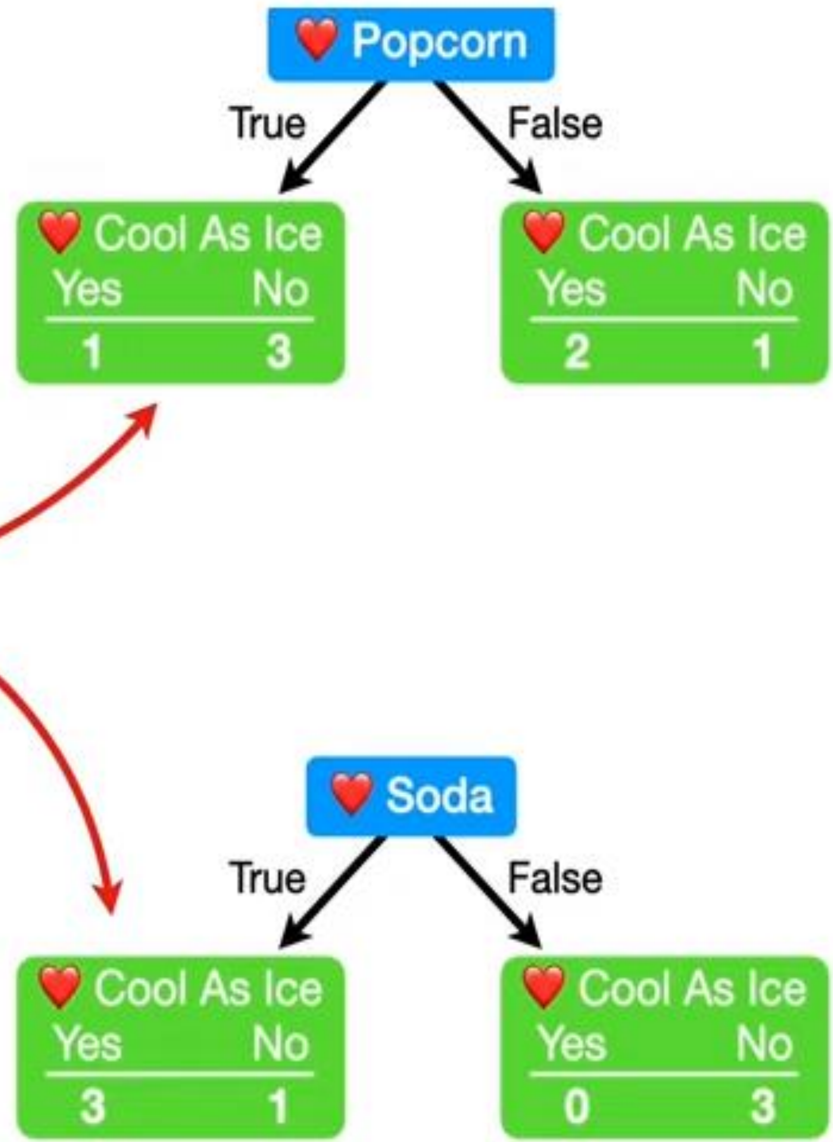
Loves Popcorn	Loves Soda	Age	Loves Cool As Ice
Yes	Yes	7	No
Yes	No	12	No
No	Yes	18	Yes
No	Yes	35	Yes
Yes	Yes	38	Yes
Yes	No	50	No
No	No	83	No

The first thing we do is decide is whether **Loves Popcorn, Loves Soda,** or **Age** should be the question we ask at the very top of the tree.

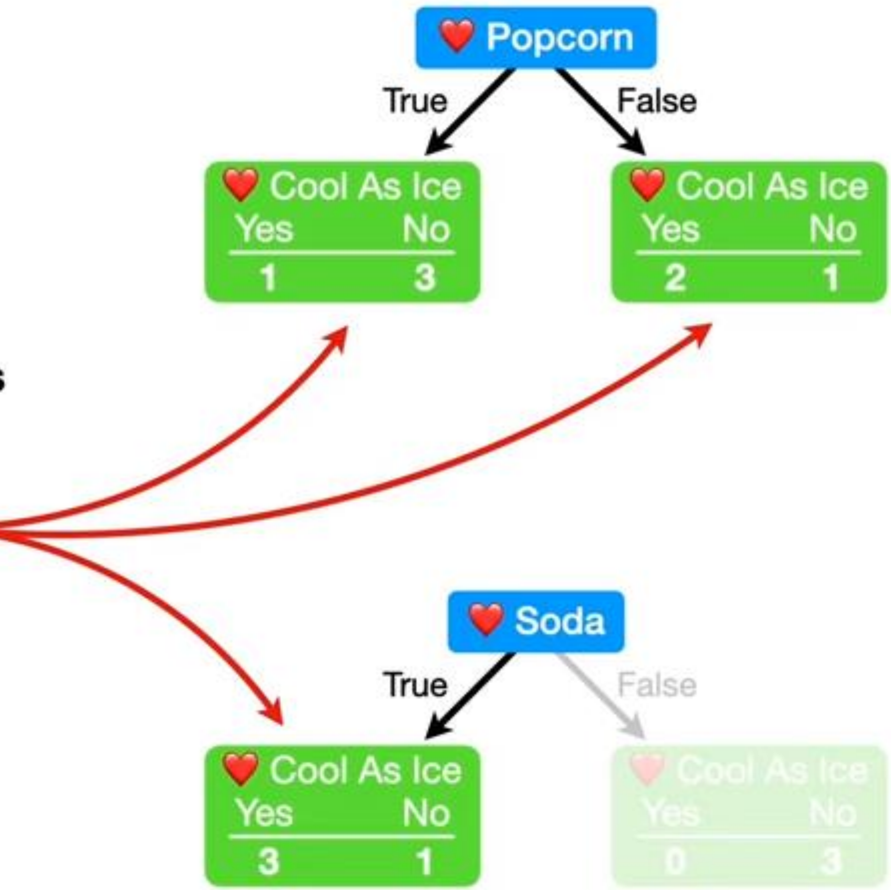


Decision Tree

Looking at the two little trees, we see that neither one does a perfect job predicting who *will* and who *will not* **Love Cool As Ice.**



Decision Tree



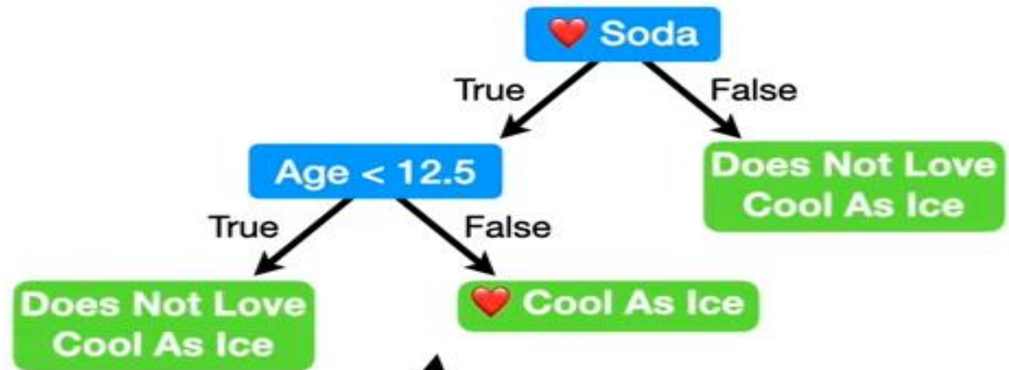
Because these three **Leaves** all contain a mixture of people who *do* and *do not* **Love Cool As Ice**…

…they are called **Impure**.

One of the most popular methods is called **Gini Impurity**, but there are also fancy sounding methods like **Entropy** and **Information Gain**.

Decision Tree

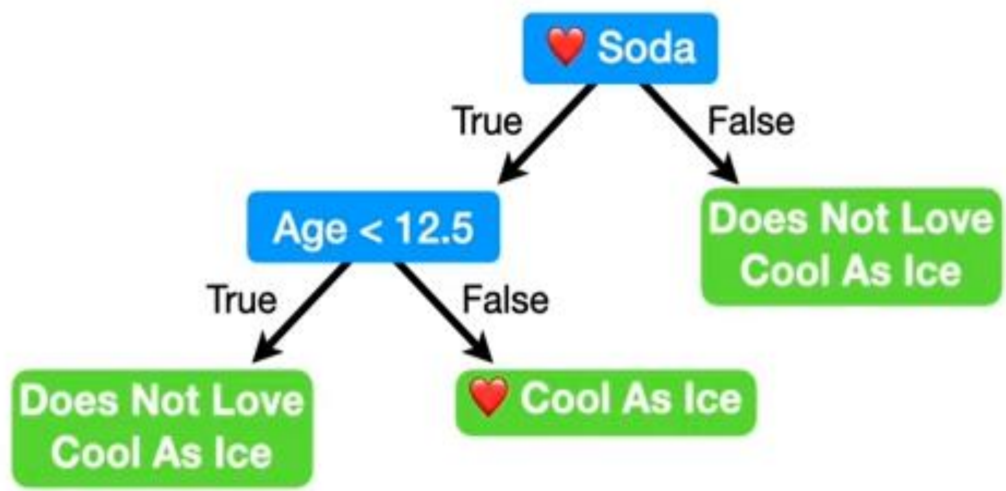
Loves Popcorn	Loves Soda	Age	Loves Cool As Ice
Yes	Yes	7	No
Yes	No	12	No
No	Yes	18	Yes
No	Yes	35	Yes
Yes	Yes	38	Yes
Yes	No	50	No
No	No	83	No



Hooray!!!
We finished building a **Tree** from this data.

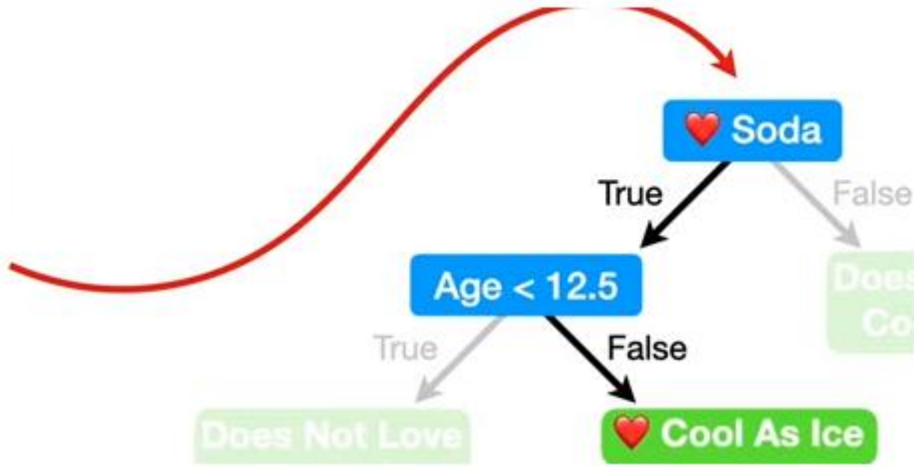
Loves Popcorn	Loves Soda	Age	Loves Cool As Ice
Yes	Yes	15	???

Now, if someone new comes along...



Decision Tree

Loves Popcorn	Loves Soda	Age	Loves Cool As Ice
Yes	Yes	15	YES!!!



Decision Tree

Instance	a1	a2	a3	Classification
1	True	Hot	High	No
2	True	Hot	High	No
3	False	Hot	High	Yes
4	False	Cool	Normal	Yes
5	False	Cool	Normal	Yes
6	True	Cool	High	No
7	True	Hot	High	No
8	True	Hot	Normal	Yes
9	False	Cool	Normal	Yes
10	False	Cool	High	Yes

Lazy Learners

- Learning from neighbors
- Simply stores training data and wait until it gets a test tuple
- i.e. works only when it gets a new example
- Less training time
- More Prediction time
- Example: KNN algorithm

K Nearest Neighbour

Pseudo code for K Nearest Neighbour (classification):

- Load the training data.
- Prepare data by scaling, missing value treatment, and dimensionality reduction as required.
- Find the optimal value for K:
- Predict a class value for new data:

  Calculate distance(X, Xi) from i=1,2,3,….,n.

  where X= new data point, Xi= training data, distance as per your chosen distance metric.

  Sort these distances in increasing order with corresponding train data.

  From this sorted list, select the top 'K' rows.

  Find the most frequent class from these chosen 'K' rows. This will be your predicted class.

K Nearest Neighbour

query ⇒ x = (Maths = 6, CS = 8) , K=3

	maths	CS	Result
1)	4	3	Fail
2)	6	7	Pass
3)	7	8	Pass
4)	5	5	Fail
5)	8	8	Pass

Euclidean distance :-

$$d = \sqrt{|x_{o1} - x_{A1}|^2 + |x_{o2} - x_{A2}|^2}$$

I $\sqrt{(6-4)^2 + (8-3)^2} = \sqrt{29} = 5.38$

II $\sqrt{(6-6)^2 + (8-7)^2} = 1$

III $\sqrt{(6-7)^2 + (8-8)^2} = 1$

IV $\sqrt{(6-5)^2 + (8-5)^2} = \sqrt{10} = 3.16$

V $\sqrt{(6-8)^2 + (8-8)^2} = 2$

I $\sqrt{\underset{4}{(6-4)^2} + \underset{25}{(8-3)^2}} = \sqrt{29} = 5.38$

II $\sqrt{\underset{0}{(6-6)^2} + \underset{1}{(8-7)^2}} = 1$

III $\sqrt{\underset{1}{(6-7)^2} + (8-8)^2} = 1$

IV $\sqrt{(6-5)^2 + (8-5)^2} = \sqrt{10} = 3.16$

V $\sqrt{(6-8)^2 + (8-8)^2} = 2$

P P P = 3P

= 0F

3 > 0