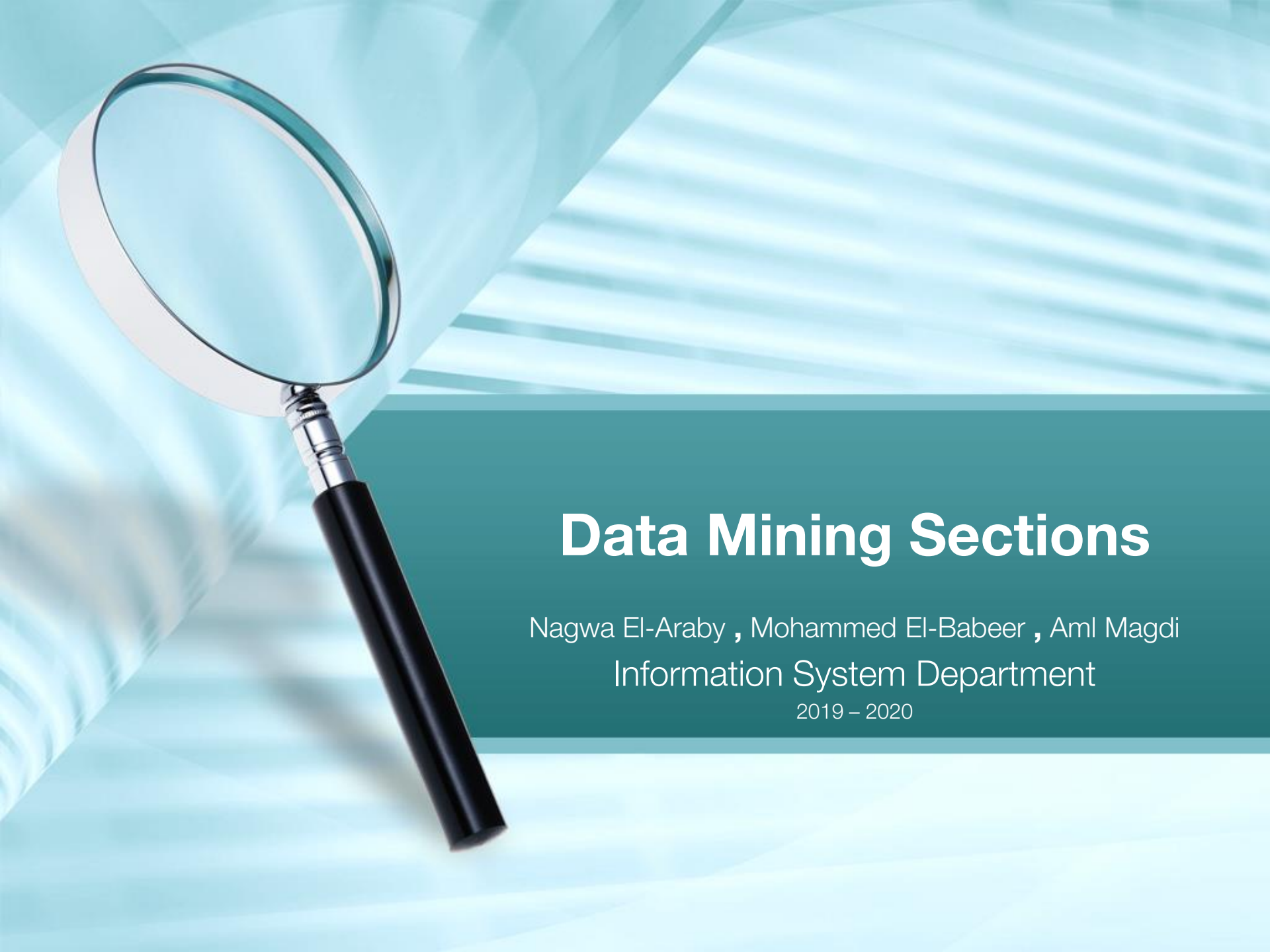


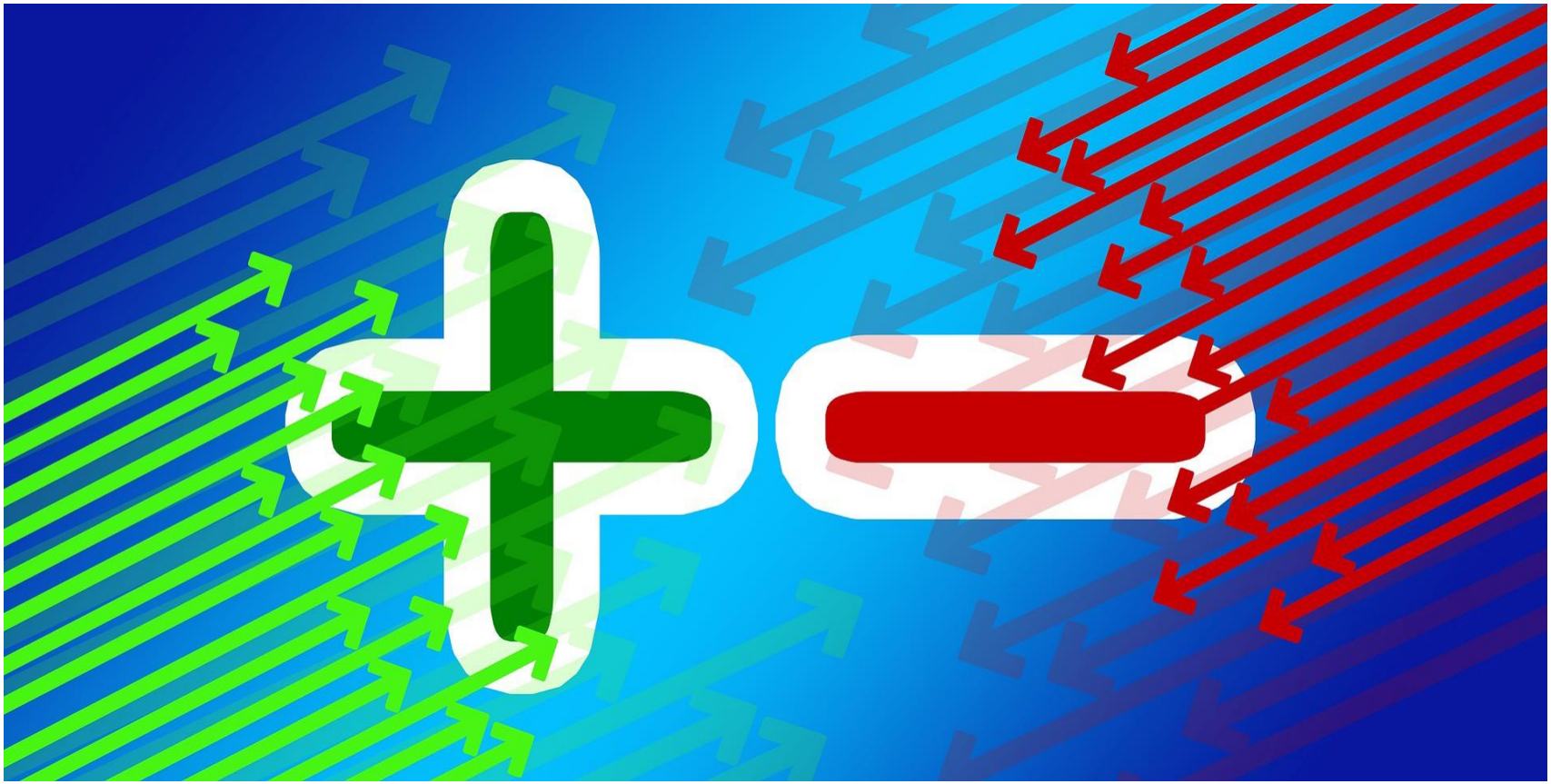
Data Mining Sections

Nagwa El-Araby , Mohammed El-Babeer , Aml Magdi

Information System Department

2019 – 2020

Classification



Classification(cont.)

- It is a Data analysis task.
- The process of finding a model that describes and distinguishes data classes and concepts. Classification is the problem of identifying to which of a set of categories ,a new observation belongs to, on the basis of a training set of data containing observations and whose categories membership is known.

How Classification Works:

- The goal is to create a set of classification rules that answer a question, make a decision, or predict behavior.
- To start, a set of training data is developed that contains a certain set of attributes as well as the likely outcome.
- The job of the classification algorithm is to discover how that set of attributes reaches its conclusion

Two-step process

- **Learning Step (Training Phase):** Construction of classification model, different algorithms are used to build a classifier by making the model learn using the training set available. The model has to be trained for the prediction of accurate results.

- **Classification Step:** Model used to predict class labels and testing the constructed model on test data and hence estimate the accuracy of the classification rules.

Example: Consider a credit-card company trying to determine which prospects should receive a credit card offer. The **company's training data** might include:

Name	Age	Gender	Annual Income	Credit Card Offer
John Doe	25	M	$39,500	No
Jane Doe	56	F	$125,000	Yes

Two-step process(cont.)

- The predictor columns Age, Gender, and Annual Income determine the value of the "predictor attribute" Credit Card Offer. In a training set, the predictor attribute is known. The classification algorithm then tries to determine how the value of the predictor attribute was reached: what relationships exist between the predictors and the decision? It will develop a set of prediction rules, usually an IF/THEN statement.

- Next, the algorithm is given a "prediction set" of data to analyze, but this set lacks the prediction attribute (or decision):

Name	Age	Gender	Annual Income	Credit Card Offer
Jack Frost	42	M	$88,000	
Mary Murray	16	F	$0	

Predictor Data

Real Life Examples

- **Weather Forecasting:** Changing Patterns in weather conditions needs to be observed based on parameters such as **temperature, humidity, wind direction**. This keen observation also requires the use of previous records in order to predict it accurately. Weather predictions use of classification techniques to report whether the day will be **rainy, sunny, or cloudy**.

- **Medical Diagnosis:** Given the symptoms exhibited in a patient and a database of anonymized patient records, predict whether the patient is likely **to have an illness**. A model of this decision problem could be used by a program to provide decision support to medical professionals

- **Spam Detection:** Given email in an inbox, identify those email messages that are spam and those that are not.

- **Speech Understanding**: Given an utterance from a user, identify the specific request made by the user.

Classifiers Of Machine Learning:

- Decision Trees
- Bayesian Classifiers
- Neural Networks
- K-Nearest Neighbour
- Support Vector Machines
- Linear Regression
- Logistic Regression

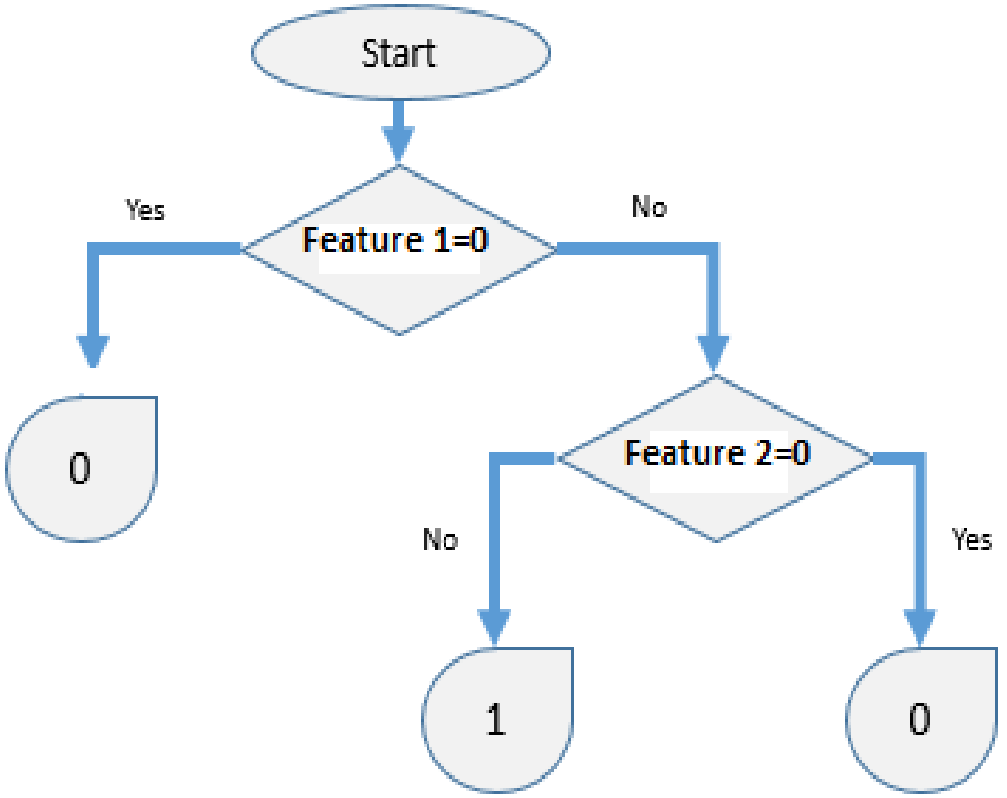
GIST OF DATA MINING :

- Choosing the correct classification method, like decision trees, Bayesian networks, or neural networks.
- Need a sample of data, where all class values are known. Then the data will be divided into two parts, a training set, and a test set.
- If the classifier classifies most cases in the test set correctly, it can be assumed that it works accurately also on the future data else it may be a wrong model chosen.

1-Decision Tree

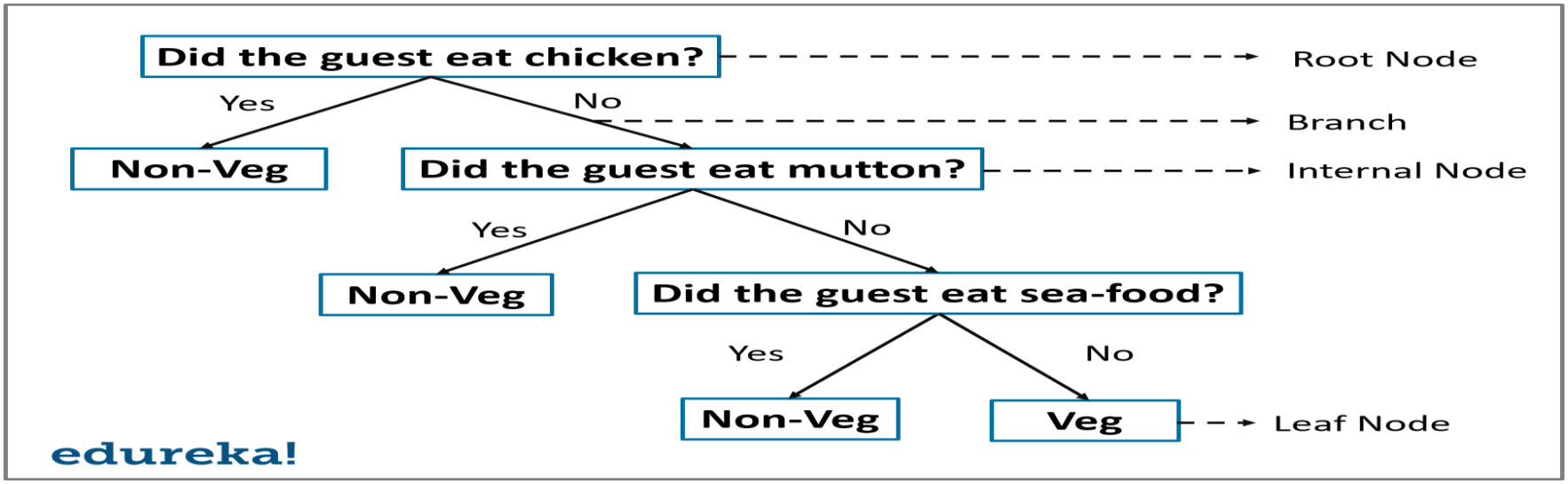
AND truth table

Feature 1	Feature 2	Result
0	0	0
0	1	0
1	0	0
1	1	1

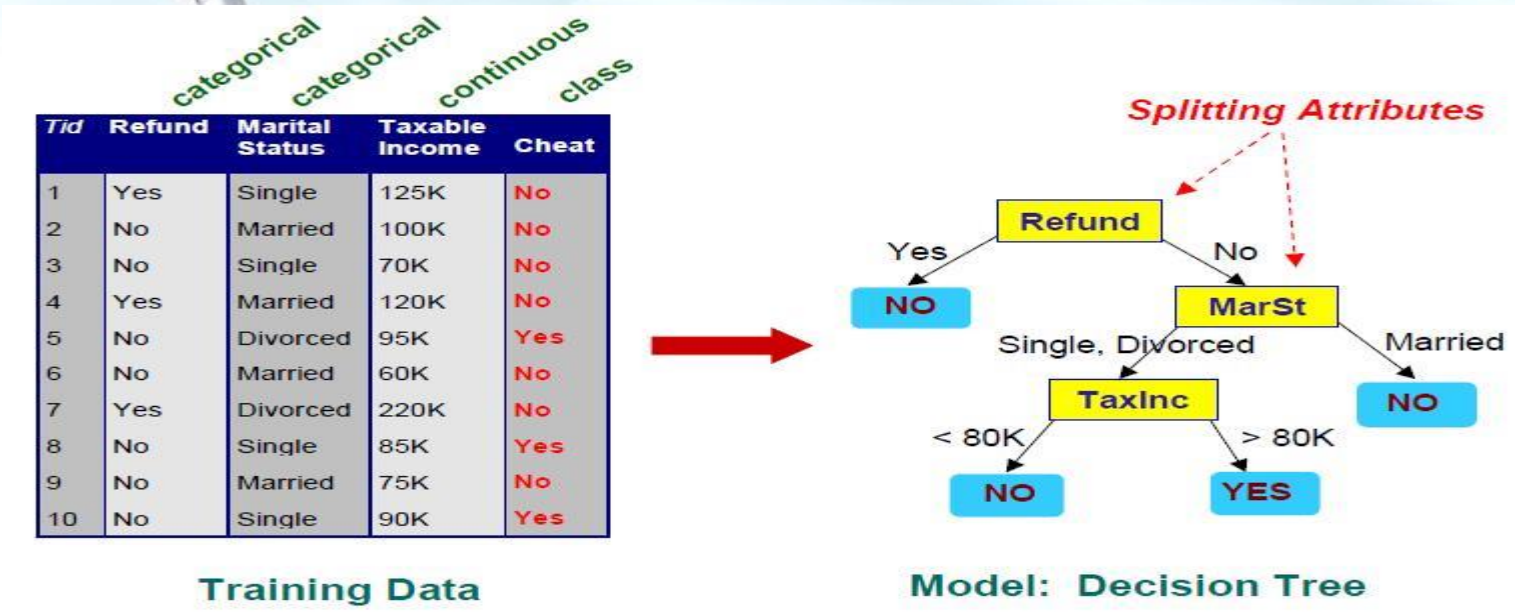


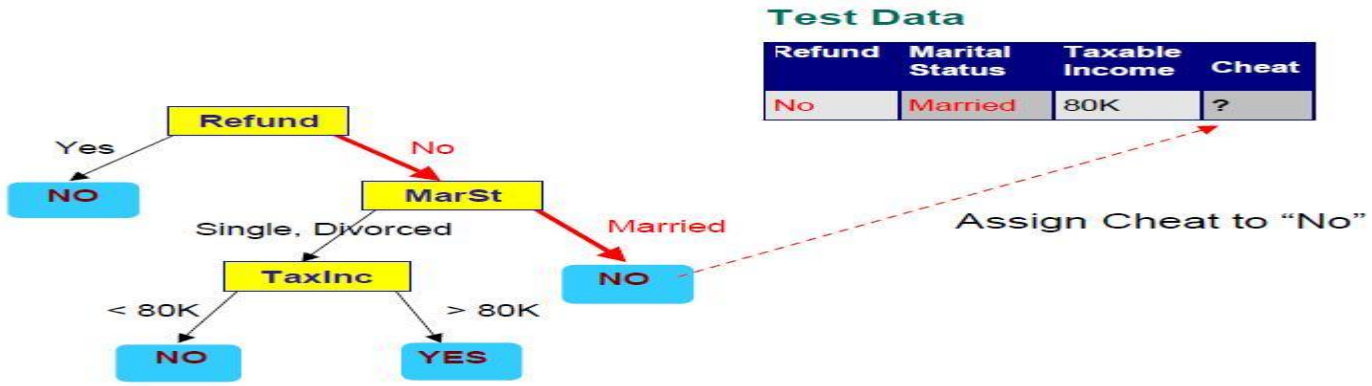
1-Decision Tree(cont.)

- It is used to create data models that will predict class labels or values for the decision-making process. The models are built from the training dataset fed to the system (supervised learning).
- Decision Tree Classifier poses a series of carefully crafted questions about the attributes of the test record. Each time it receive an answer, a follow-up question is asked until a conclusion about the class label of the record is reached.
- Structure Of A Decision Tree:



Decision Tree Example





A Decision Tree has the following structure:

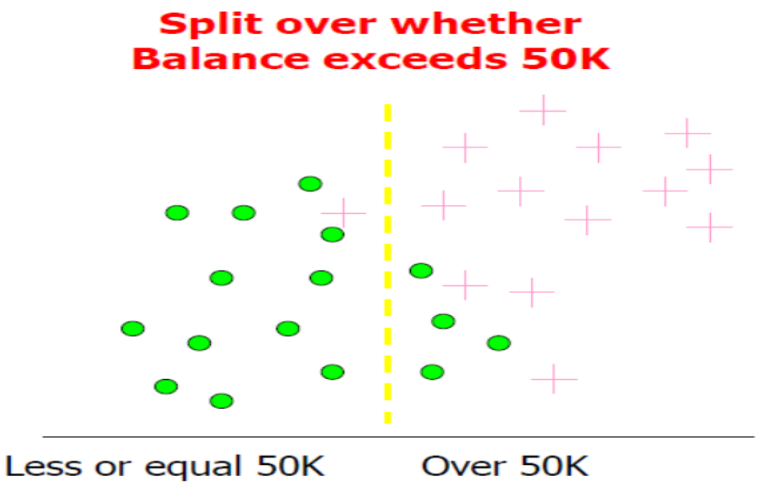
- **Root Node:** The root node is the starting point of a tree. At this point, the first split is performed.
- **Internal Nodes:** Each internal node represents a decision point (predictor variable) that eventually leads to the prediction of the outcome.
- **Leaf/ Terminal Nodes:** Leaf nodes represent the final class of the outcome .
- **Branches:** Branches are connections between nodes, they're represented as arrows. Each branch represents a response such as yes or no.

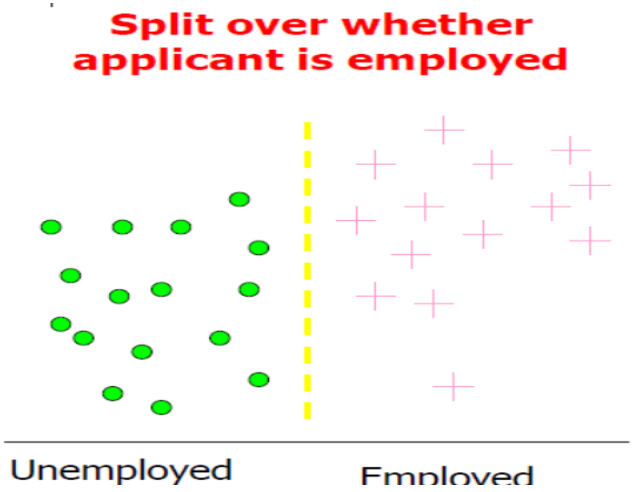
The Decision Tree Algorithm follows the below steps:

1. Select the best attribute using Attribute Selection Measures(ASM) to split the records.
2. Make that attribute a decision node and breaks the dataset into smaller subsets.
3. Starts tree building by repeating this process recursively for each child until one of the condition will match:
    1. All the tuples belong to the same attribute value.
    2. There are no more remaining attributes.
    3. There are no more instances.

Attribute Selection using Information Gain

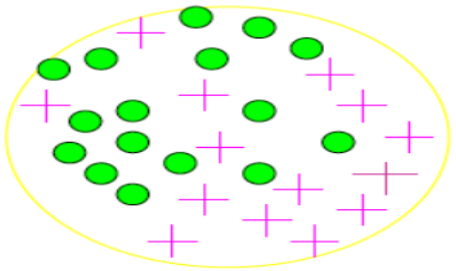
- Idea : Which test is more informative?





- Entropy comes from information theory. The higher the entropy the more the information content.

$$\text{Entropy} = \sum_i -p_i \log_2 p_i$$



p_i is the probability of class i
Compute it as the proportion of class i in the set.

Attribute Selection using Information Gain

- We want to determine which attribute in a given set of training feature vectors is most useful for discriminating between the classes to be learned.
- Information gain tells us how important a given attribute of the feature vectors is. We will use it to decide the ordering of attributes in the nodes of a decision tree.

Easy way to understand Information gain= (overall entropy at parent node) – (sum of weighted entropy at each child node).

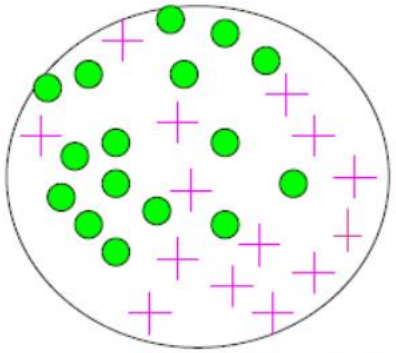
- Attribute with maximum information is best split attribute.

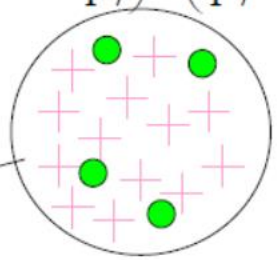
Attribute Selection using Information Gain(cont.)

Information Gain = entropy(parent) – [average entropy(children)]

child entropy $-\left(\frac{13}{17}\cdot\log_2\frac{13}{17}\right)-\left(\frac{4}{17}\cdot\log_2\frac{4}{17}\right)=0.787$

Entire population (30 instances)

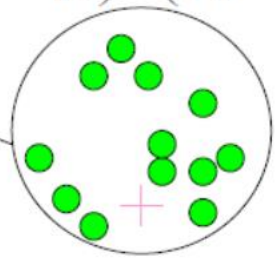




17 instances

child entropy $-\left(\frac{1}{13}\cdot\log_2\frac{1}{13}\right)-\left(\frac{12}{13}\cdot\log_2\frac{12}{13}\right)=0.391$

parent entropy $-\left(\frac{14}{30}\cdot\log_2\frac{14}{30}\right)-\left(\frac{16}{30}\cdot\log_2\frac{16}{30}\right)=0.996$



13 instances

(Weighted) Average Entropy of Children = $\left(\frac{17}{30}\cdot 0.787\right)+\left(\frac{13}{30}\cdot 0.391\right)=0.615$

Information Gain= 0.996 - 0.615 = 0.38 for this split

Decision trees in Python with Scikit-Learn

```
#import needed packages (sklearn,scipy)
!pip install sklearn
!pip install scipy
from sklearn import tree
#construct decision tree
clf = tree.DecisionTreeClassifier()
#[height, hair-length, voice-pitch]
X = [ [180, 15,0],
      [167, 42,1],
      [136, 35,1],
      [174, 15,0],
      [141, 28,1]]

Y = ['man', 'woman', 'woman', 'man', 'woman']
#train the tree by the training data
clf = clf.fit(X, Y)
#predict the class of new data
prediction = clf.predict([[133, 37,1]])
print(prediction)
```


Thank You