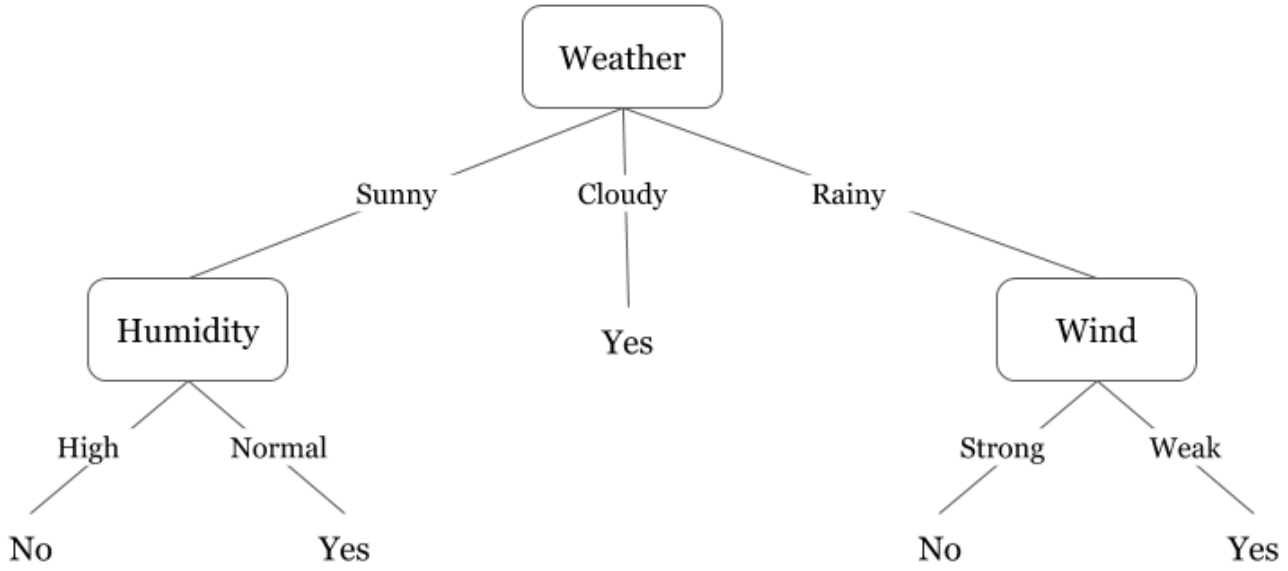
Decision Tree

Objective

- Concept of Decision Tree
- Use of Decision Tree to classify data
- Basic algorithm to build Decision Tree
  - Some illustrations
- Concept of Entropy
  - Basic concept of entropy in information theory
  - Mathematical formulation of entropy
  - Calculation of entropy of a training set
- Decision Tree induction algorithms
  - ID3

Introduction

- Basic idea behind building a decision tree is to map all the possible decision paths in the form of a tree.
- Efficient machine learning algorithm.
- Need to create new tree once seen whole new data
- Data driven programing the conditions.



Reference : Weather Prediction

Case Study: Drug Prediction

- A medical researcher compiling data for a study.
- During course of treatment, each patient responded to one of two medications; we'll call them Drug A and Drug B.
- Job is to build a model to find out which drug might be appropriate for a patient with same illness.
- Feature sets: Age, Gender, Blood Pressure, and Cholesterol
- Target: Drug that each patient responded to.

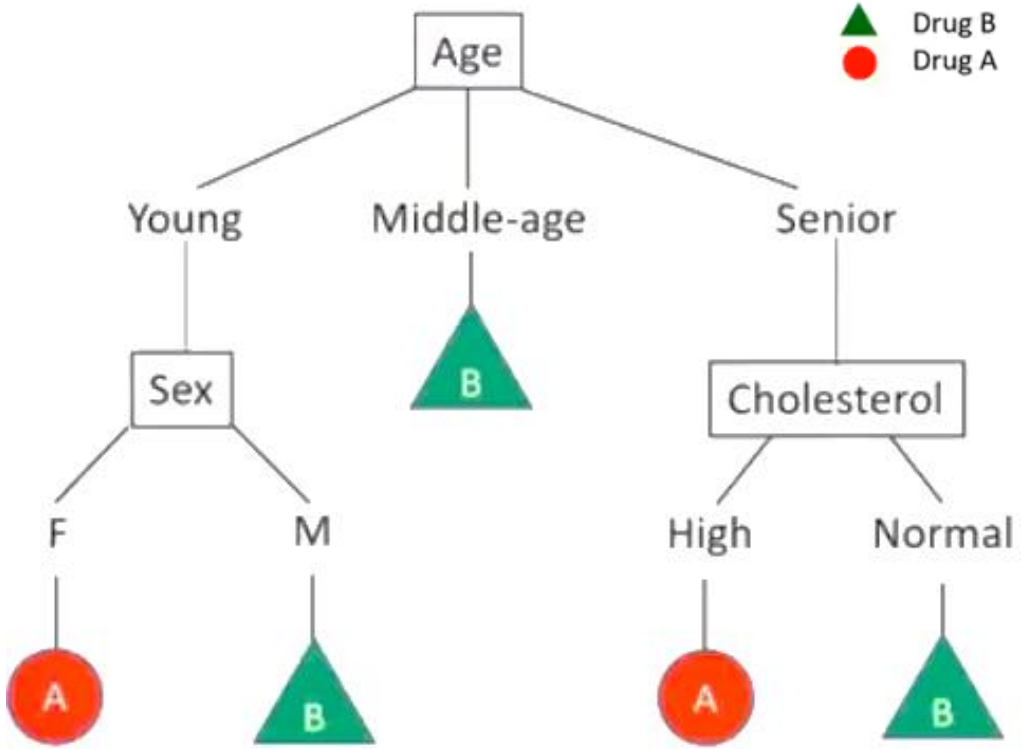
Features Target

Patient ID	Age	Sex	BP	Cholesterol	Drug
p1	Young	F	High	Normal	Drug A
p2	Young	F	High	High	Drug A
p3	Middle-age	F	Hiigh	Normal	Drug B
p4	Senior	F	Normal	Normal	Drug B
p5	Senior	M	Low	Normal	Drug B
p6	Senior	M	Low	High	Drug A
p7	Middle-age	M	Low	High	Drug B
p8	Young	F	Normal	Normal	Drug A
p9	Young	M	Low	Normal	Drug B
p10	Senior	M	Normal	Normal	Drug B
p11	Young	M	Normal	High	Drug B
p12	Middle-age	F	Normal	High	Drug B
p13	Middle-age	M	High	Normal	Drug B
p14	Senior	F	Normal	High	Drug A
p15	Middle-age	F	Low	Normal	?

Reference : Patient Drug Data

Decision Tree

- It is a sample of binary classifiers.
- **Internal node**: Test
- **Branch**: Test result
- **Leaf Node**: Patient Class (classification)



Reference : Decision Tree

How to Build Decision Tree ?

- Choose an attribute from dataset.
- Calculate the significance of the attribute (effective attribute) in the splitting of the data.
- Split the data based on the value of the best attribute.
- Then, go to each branch and repeat above steps for the rest of the attributes.
- Once your tree is build, you can use it to predict the class of unknown cases (ex. Proper Drug for new patient), based on his/her characteristics.

How to select the effective attribute ?

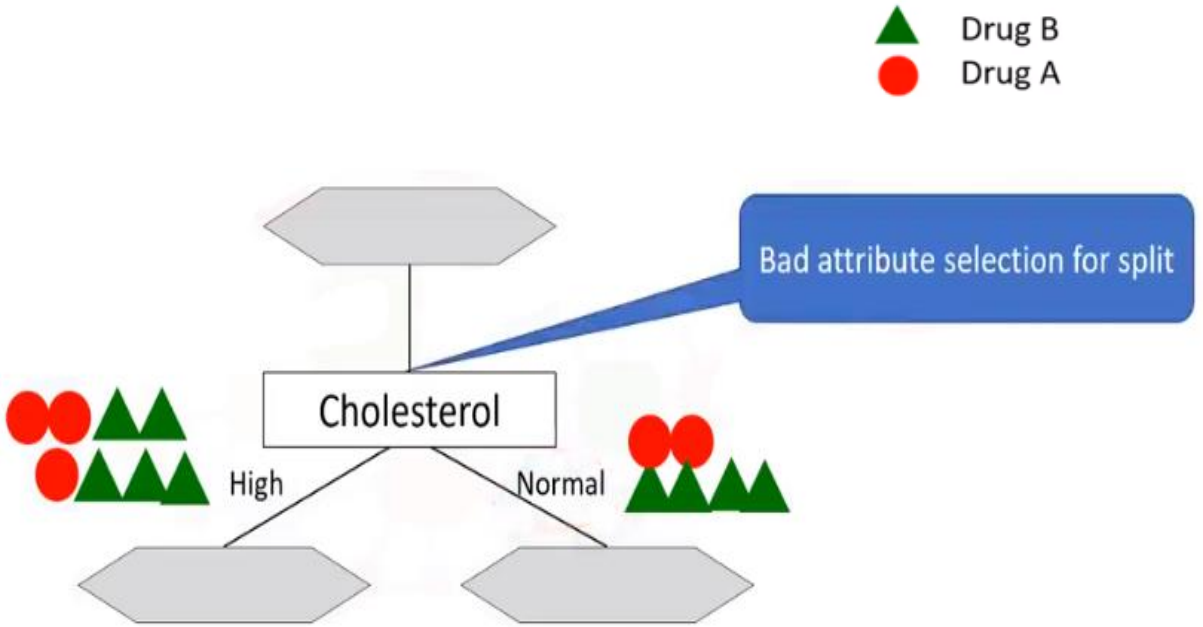
- The algorithm chooses the most predictive feature to split the data on.
- Root node can be any valid feature.
- Root node divides to create branches.

Patient ID	Age	Sex	BP	Cholesterol	Drug
p1	Young	F	High	Normal	Drug A
p2	Young	F	High	High	Drug A
p3	Middle-age	F	Hiigh	Normal	Drug B
p4	Senior	F	Normal	Normal	Drug B
p5	Senior	M	Low	Normal	Drug B
p6	Senior	M	Low	High	Drug A
p7	Middle-age	M	Low	High	Drug B
p8	Young	F	Normal	Normal	Drug A
p9	Young	M	Low	Normal	Drug B
p10	Senior	M	Normal	Normal	Drug B
p11	Young	M	Normal	High	Drug B
p12	Middle-age	F	Normal	High	Drug B
p13	Middle-age	M	High	Normal	Drug B
p14	Senior	F	Normal	High	Drug A
p15	Middle-age	F	Low	Normal	?

Reference : Patient Drug Data

Effective Attribute Quest!

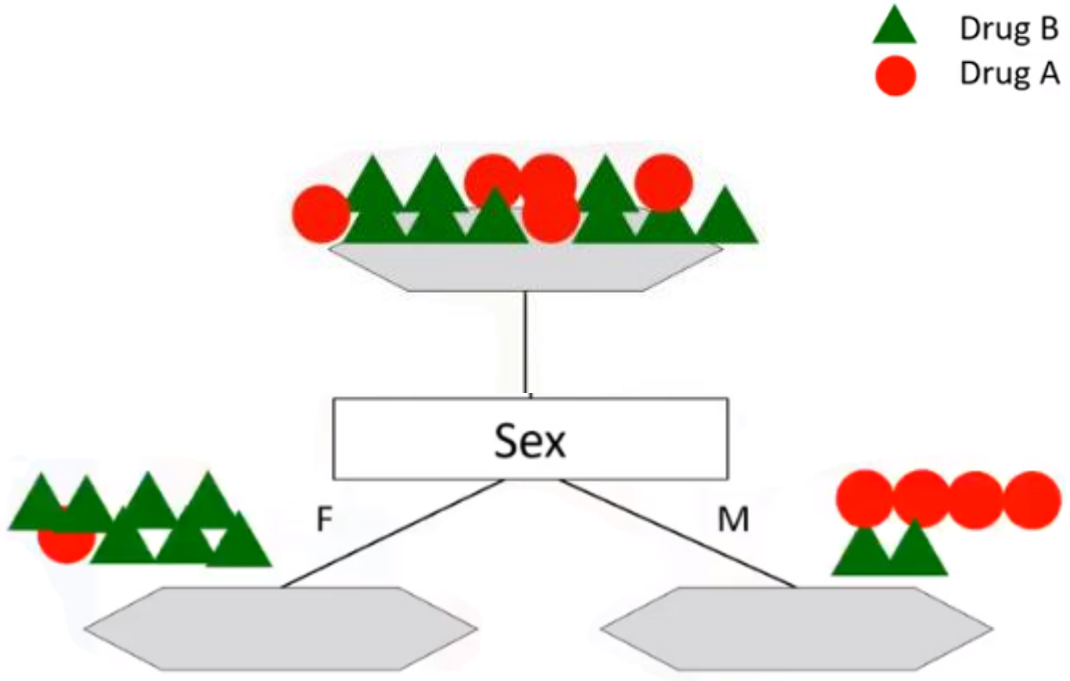
- Let’s begin our quest by picking “Cholesterol” as the first attribute to split data.
- It is a sample of bad attribute selection for splitting data.
- Bad in terms of impurity cleanliness.
- Improper class balance attained.



Reference : How Decision Tree Works

Effective Attribute Quest!

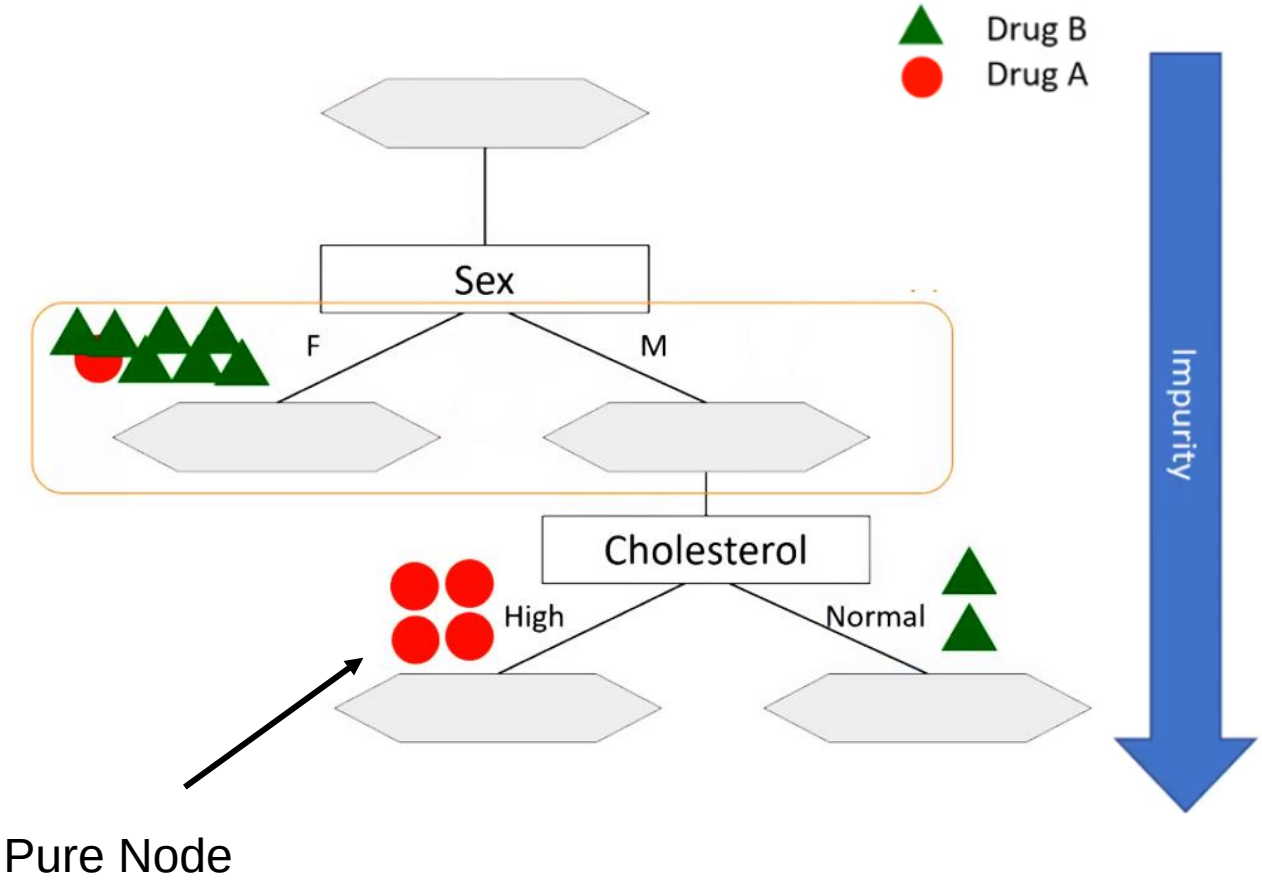
- Let's try with "sex" attribute of patients.
- However, it is still a better choice in comparison with the "Cholesterol" attribute, because the result in the nodes are more pure, i.e. nodes which are either mostly Drug A or Drug B.
- Thus, it's more predictive than the other attributes.



Reference : How Decision Tree Works

Effective Attribute Quest!

- Predictiveness is based on decrease in “impurity” of nodes.
- So, the “Sex” feature is a good candidate in the following case, because it almost found the pure patients.
- We test “Cholesterol” again here.
- As you can see, it results in even more pure leaves.
- So, we can easily make a decision here.



Reference : How Decision Tree Works

Intuition of Node Impurity

- Method uses recursive partitioning to minimize the “impurity” at each step.
- ”Impurity” of nodes is calculated by “Entropy” of data in the node.
- So, what is “Entropy”?
- Entropy is the amount of information disorder, or the amount of randomness in the data.
- In decision trees, we're looking for trees that have the smallest entropy in their nodes.
- Lower the Entropy, distribution is less uniform and nodes are more pure!

Entropy

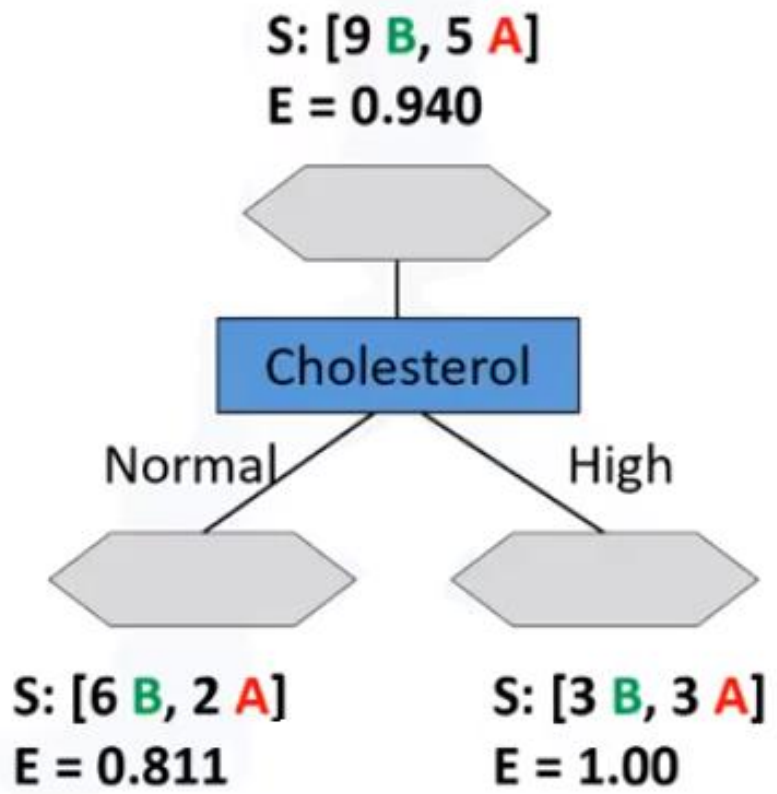
- To calculate entropy, formulae is:

$$Entropy = -p(A)\ \log(p(A)) - p(B)\ \log(p(B))$$ **base 2*

- *p* is for the proportion or ratio of a category, such as Drug *A* or *B*.
- Let's calculate the entropy of the dataset in our case, before splitting it.
- We have 9 occurrences of Drug *B* and 5 of Drug *A*.
- Entropy = 0.528 + (-0.410) = 0.940 (approx.).

Entropy at each node (for each Attribute)

- Consider all the attributes and calculate the “Entropy” after the split, and then chose the best attribute.
- Calculate node entropy for Cholesterol feature.
- Separate test is conducted for each valid feature.



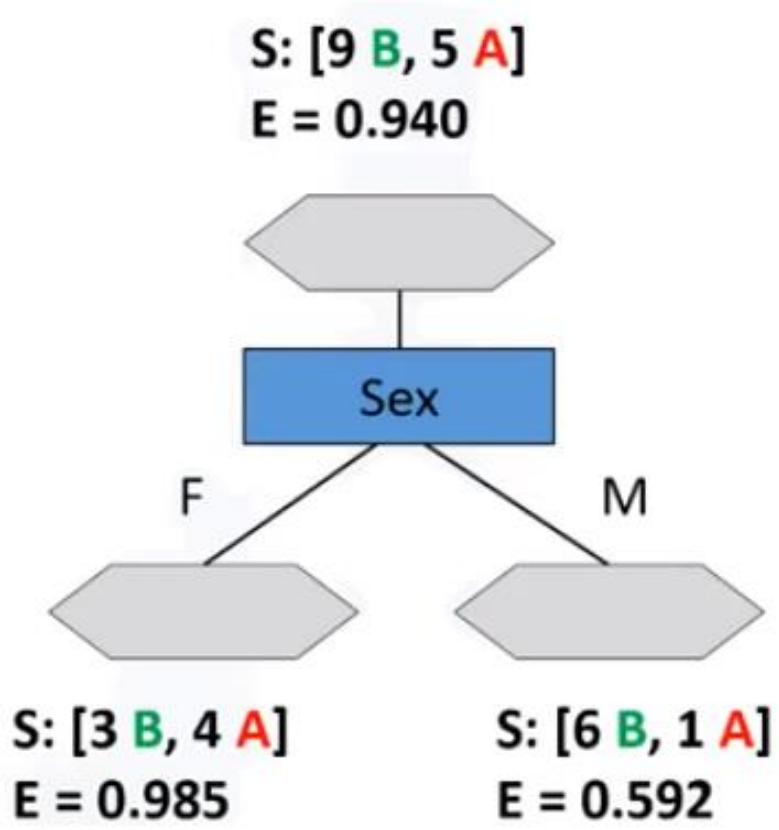
Reference : How Decision Tree Works

Entropy at each node (for each Attribute)

- Calculate node entropy for Sex feature
- -(9/14*log(9/14) + 5/14*log(5/14))

  =0.940
- Entropy of branch **F**
- -(3/7*log(3/7) + (4/7*log(4/7))

  =0.985



Reference : How Decision Tree Works

Information Gain (ID3)

- Before taking our splitting decision, lets understand Information Gain!
- Information gain is the information that can increase the level of certainty after splitting.
- As entropy, or the amount of randomness, decreases, the information gain, or amount of certainty, increases, and vice-versa.
- So, constructing a decision tree is all about finding attributes that return the highest information gain.
- Information Gain = Total Entropy – Sum of Conditional Entropies

Comparison of Attributes

- Information Gain (Sex)

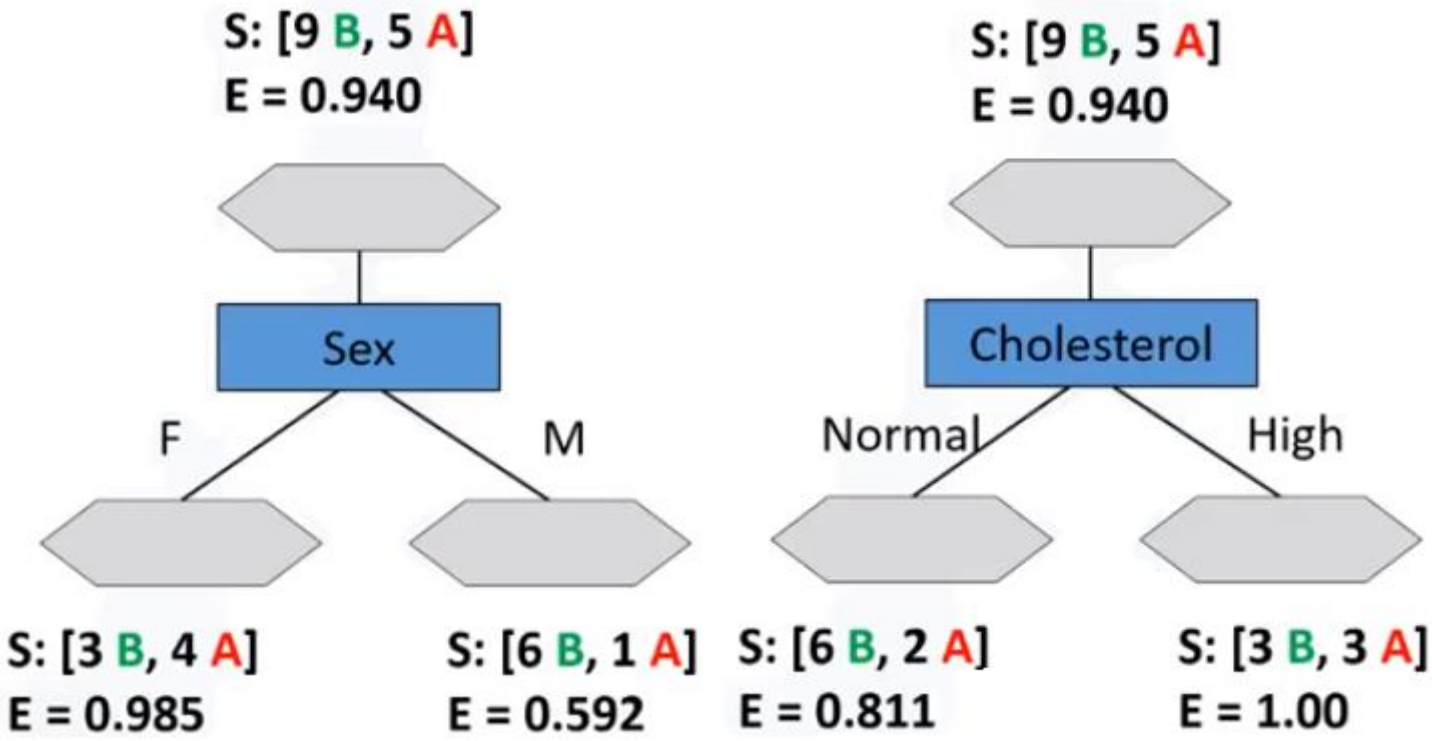
  0.940-(7/14*0.985)-(7/14*0.592)

  = **0.151**

- Information Gain (Cholestrol)

  0.940-(8/14*0.811)-(6/14*1)

  = **0.048**



Reference : How Decision Tree Works

Question ??

- Between the Cholesterol and Sex attributes, which one is a better choice?
- Which one is better as the first attribute to divide the dataset into 2 branches?
- Which attribute results in more pure nodes for our drugs?
- Answer: **“Sex”** attribute

Repeat!

- So, we select the “Sex” attribute as the first splitter.
- Now, what is the next attribute after branching by the “Sex” attribute?
- We should repeat the process for each branch, and test each of the other attributes to continue to reach the most pure leaves.
- This is the way that you build a decision tree!

Hands On