These two models are often used in **classifiers**.

$\equiv$ a machine learning system that makes decisions on input.

the inputs are called characteristics or instance.

the output is called a decision.

We can construct them from Bayesian Networks.

We must first introduce the idea of **entropy**.

$$\text{ENT}(X) = -\sum_{X} Pr(x) log_2(Pr(x))$$

We can show it with the following data table:

![]($(https$: $//paper$ $-$
$attachments.dropbox.com/s_6 6C470E465947B218F12E6833ACF54B222DBA6F153DEEC1F4A1A4D06909A7A0F_1 5909132720$
$drawing + 10.jpg)$

The form we tend to utilize is **CONDITIONAL PROBABILITY**.
If we have ENT(X) and we learn that Y=y, we have

$$E(X|y) = -\sum_{X} Pr(x|y) log_2(Pr(x|y))$$

Alternatively, if we plan to observe Y but do not yet know the value

$$E(X|Y) = -\sum_{Y} Pr(y)\mathrm{ENT}(X|y)$$

It also turns out that information can never increase average entropy, ie

$$\mathrm{ENT}(X|Y) \leq \mathrm{ENT}(X)$$

Note that this specifies average; the entropy of a single value may increase:

![]($(https : //paper - attachments.dropbox.com/s_{6}6C470E465947B218F12E6833ACF54B222DBA6F153DEEC1F4A1A4D06909A7A0F_{1}5909138047$ $drawing + 11.jpg)$

These are used to build classifiers by supervised learning of labeled data.
Our CPT thus effectively functions as our model.

We will now use the notion of a decision tree/random forest to solve a problem.
We will use the following data and corresponding tree:

$![Wehave12labeledvariables](https://paper-attachments.dropbox.com/s_66C470E465947B218F12E6833ACF54B222DBA6F153DEEC1F4A1A4D06909A7A0F_15909141769$
$Shot+2020-05-31+at+1.34.17+AM.png)$

This model is called **interpretable** because it is easy to read, as opposed to a neural network. Classifying a variable is as easy as parsing the tree!

Consider X_{12} — we can just walk; this probability happens to match, but it won't be in general.

$![](https : //paper - attachments.dropbox.com/s_6 6C470E465947B218F12E6833ACF54B222DBA6F153DEEC1F4A1A4D06909A7A0F_1 5909144100$
$Shot + 2020 - 05 - 31 + at + 1.34.28 + AM.png)$

The depth of the decision tree is a sign of its complexity.
Splitting is as easy as making a choice.
Nodes represent attributes.
Leaves represent decisions.

We can equivalently build:

  this has 4 attributes rather than the 10 from above
  this is much shallower, and thus simpler

The algorithm itself is very simple; we just split repeatedly as if tracing the tree.

This assumes a black box for choosing variables, but developing one is not hard
How do we choose which attribute to split on at a given depth?
We can define the algorithm by looking at the following state:

$$

We thus use conditional entropy as a score to determine our next split.

The algorithm is as follows:

![](https://paper-attachments.dropbox.com/s_6C470E465947B218F12E6833ACF54B222DBA6F153DEEC1F4A1A4D06909A7A0F_1 5909865831
drawing+15.jpg)

How do we evaluate an algorithm? We use **cross-validation**.
$\equiv$ split the dataset into 80/20 training/testing data & repeat to find average score.

This can be generalized one more time to a **random forest**.
We build a series of trees and majority vote to determine the output.
We call this type of method an ensemble learning method.

Suppose we have a dataset of 5 values; we may bootstrap data sets by random choice to get:

$$

The count of numbers chosen will be a parameter.
We can test the power using the out of bag examples.

![](https : //paper − attachments.dropbox.com/$s_6$6C470E465947B218F12E6833ACF54B222DBA6F153DEEC1F4A1A4D06909A7A0F$_1$5909874816
drawing + 17.jpg)

A specific subset of these are called naive:

Traditionally, we want AI to be easily explainable.
Consider the following example:

![](https://paper-attachments.dropbox.com/s_6C470E465947B218F12E6833ACF54B222DBA6F153DEEC1F4A1A4D06909A7A0F_15909883463
drawing+20.jpg)

Say we are asked C|{S=1, G=0, F=1, M=1}.
We might say "yes, because F & M!", as S & G are not used.
This is called a **PI-explanation**.
In actuality, we can make a tractable circuit from this data, which is much power powerful.
This, however, is not as interpretable!
In the current day, Random Forests < Bayesian Classifiers < Neural Networks.