Advanced Retail Strategy

Session 3: May 30th, 2019

What do we know about AI/ML?

What are our key questions on the topic?

Jet, a data-powered partner in your life

You see a Jet ad on your phone while on the subway. You'd heard about us and this seems like a good excuse to try us out

You go on the site when you get to work and the same offer is on the homepage. Well, that was easy

You begin to browse products to get a sense of the assortment and your search results and the listed products seem to be all the things you'd be interested in, both in this category and in others. The site asks you some questions as you shop and the recommendations and search get notably better.

You see on the homepage a note about a festival happening in your city next week and there's a great bundle (mesh bag, sunscreen, Kleen Kanteen) of products for it. You've been considering a new phone case for events like this and Jet has a helpful COYA video walkthrough to get you to the right one rather than having to search through hundreds of cases. They didn't even have to ask what phone you have. Huh, cool.

A few weeks later you're back to do your weekly grocery shop. Your cart already has products in it because Jet knows when you need to top up.

It also has recipe recommendations for the week based on the things you already have with offers of bundles to save on the things you still need to plan the week out. We know you're gluten free so there's no worry about any offered products having gluten but when you want to see the details the site has all the nutritional information you need to be confident about what you give to your family.

You make the order and the tracking is proactive and always accurate, letting Jet fit into your already hectic day and take some of the pressure off, rather than being another thing to check off a list that never seems to quite be complete.

Good News.

Computationally this stuff is actually pretty easy.

Process is where most companies fall short.

Reach a little…

But if you can *really* tighten things up, that’s when you unlock the ability to create amazing…

A Day in the NYC Life, 2020

Our Parcel delivery person brings the groceries into your apartment, unpacks them into the fridge and pantry and takes a photo of the fridge and pantry afterwards to send to the customer so they feel good about the result. It’s like magic elves keeping your things in order.

Meanwhile, via product vision files, our systems identify the products in the fridge and pantry and (for products where this is visible) identifies how much is left.

From the quantities identified, re-order reminders and bundles are automatically created and pushed at the right times.

From the brands and products included, user preference profiles are updated, which are leveraged both for the site experience, and for recommendations and standard transactions.

Vans on their delivery routes fill 15% the truck with other products that have not yet been ordered but are likely necessary for customers on that route and ask them on delivery if they will need them, with the ability to add them to orders on the fly, eliminating more of the “fill up” trips that end up taking up the most time out of the week

I Want In! How Do I Learn about AI/ML?

Good question. Here’s a potential path

How Might I Learn Machine Learning / AI?

1. Read the very excellent text from our very own Foster Provost: Data Science for Business to build a conceptual understanding
2. Do some toy examples in **Weka**, a GUI built on Java libraries, to get a feel for what we just read without having to learn to code
3. Start doing more work on data analysis and preparation – this is 80%+ of the job of a data scientist and if it's done poorly, it destroys everything cool you try to do. This is where coding will enter the scene. R and/or Python are the mediums here.
4. Now use some of the standard ML packages in R and/or Python now that you know how to prep data
5. Now try to bring in some real data into those systems and get to work

How Else Can I Engage?

You don’t have to write code to have huge impact in this area. Just being someone who understands what it is (and isn’t!) can get you major benefits:

1) Help your organization move forward its thinking and adopt more forward-looking strategies to attack problems.

2) Help advocate for tech investments *in the right areas* to achieve gains that will offer meaningful advantage to your organization’s business.

3) Defend yourself and your organization from vendor BS that promises you the world in areas where the actual gain is minimal or realistically quite manually and as a result not as great an investment.

Aside: Culture *(Revisiting IT Savvy…)*

The other thing that Google, FB, et al. don't discuss but is equally key is culture. Walmart has a great example of this: Walmart has created tremendous customer success by a policy that is focused on EDLP (Everyday Low Price). This strategy is unlocked by the more internal-focused concept of EDLC (Everyday Low Cost) which makes that possible.

Companies would do well to look at these examples to understand how to materialize value:

1) If you make something a core ideal, it will happen
2) EDLP : EDLC :: Personalization : Data Discipline
   If people treat the creation, aggregation, cleaning, and materialization with seriousness and commitment then they unlock amazing capabilities to leverage data to create the kind of customer experiences that excite.

Signet Bank

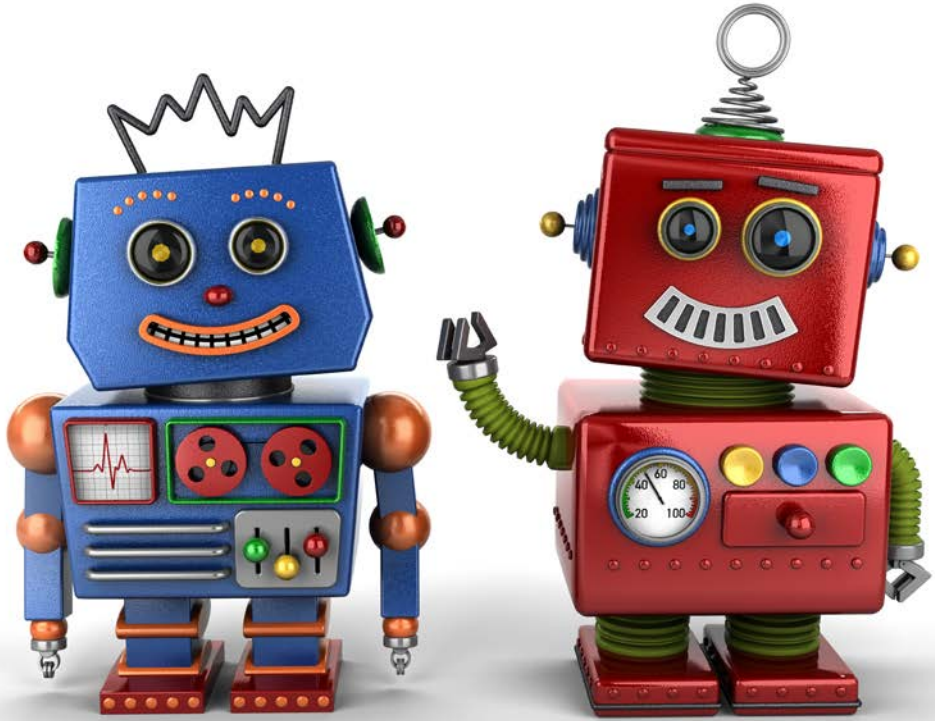
Early Lessons in ML

What did the Signet Bank guys do?

OK I want to know more.

Aside: What is the difference between Data Science and Big Data?

What are the two types of problems we can attack?



Baseline: WTF is AI/ML

AI/ML can be used to do two things:

1) Do things that humans already do well, but do it at fantastic speed and scale to make stuff that is tedious cheaper and faster *(e.g. humans are really good at seeing if email is spam but having humans go through all email is heinous)*.
2) Do things with a level of accuracy that humans can't do, potentially also at amazing scale and speed to unlock new opportunity *(e.g. generate bundles of products that outperform human selections by 80% CVR 😎)*.

What are some types of each?

Type 1:

Type 2:

What are the 9 types of algorithms commonly used in ML?

1. Classification/probability estimation
2. Regression
3. Similarity Matching
4. Clustering
5. Association Rules
6. Profiling
7. Link prediction
8. Data reduction
9. Causal modeling

Supervised v Unsupervised Learning

What’s the difference?

First of All, Some Terms…

Attributes
Features
Independent Variables
Predictors

	Are they cute?	Furry?	Delicious?
Rabbit	Yes	Yes	Yes
Duck	Yes	No	Yes
Hippo	No	No	No

Observations
Rows
Instances
Feature Vectors

Target Variable
Target Attribute
Dependent Variable

Supervised v Unsupervised Learning

Supervised v Unsupervised Learning

Unsupervised

Supervised

Intuitive Understandings of Common Algo Types

- Classification
- Regression
- SVM
- k-Means Clustering
- Neural Nets / Deep Learning

Classification: Enter the Dragon

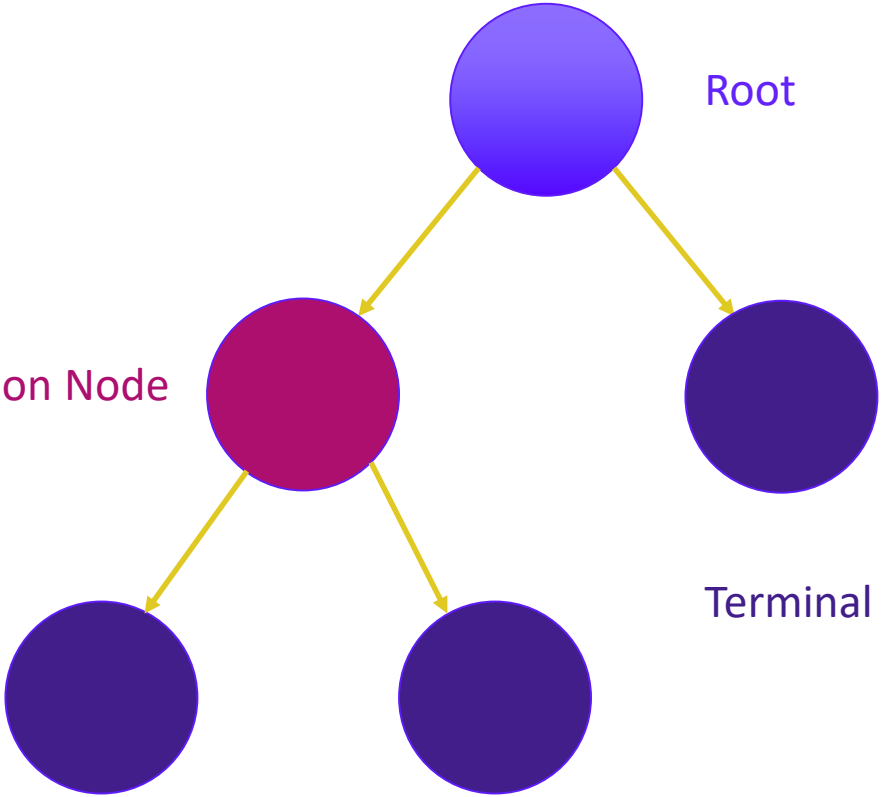
Time to dig in on classification.

We're going to explore two areas:

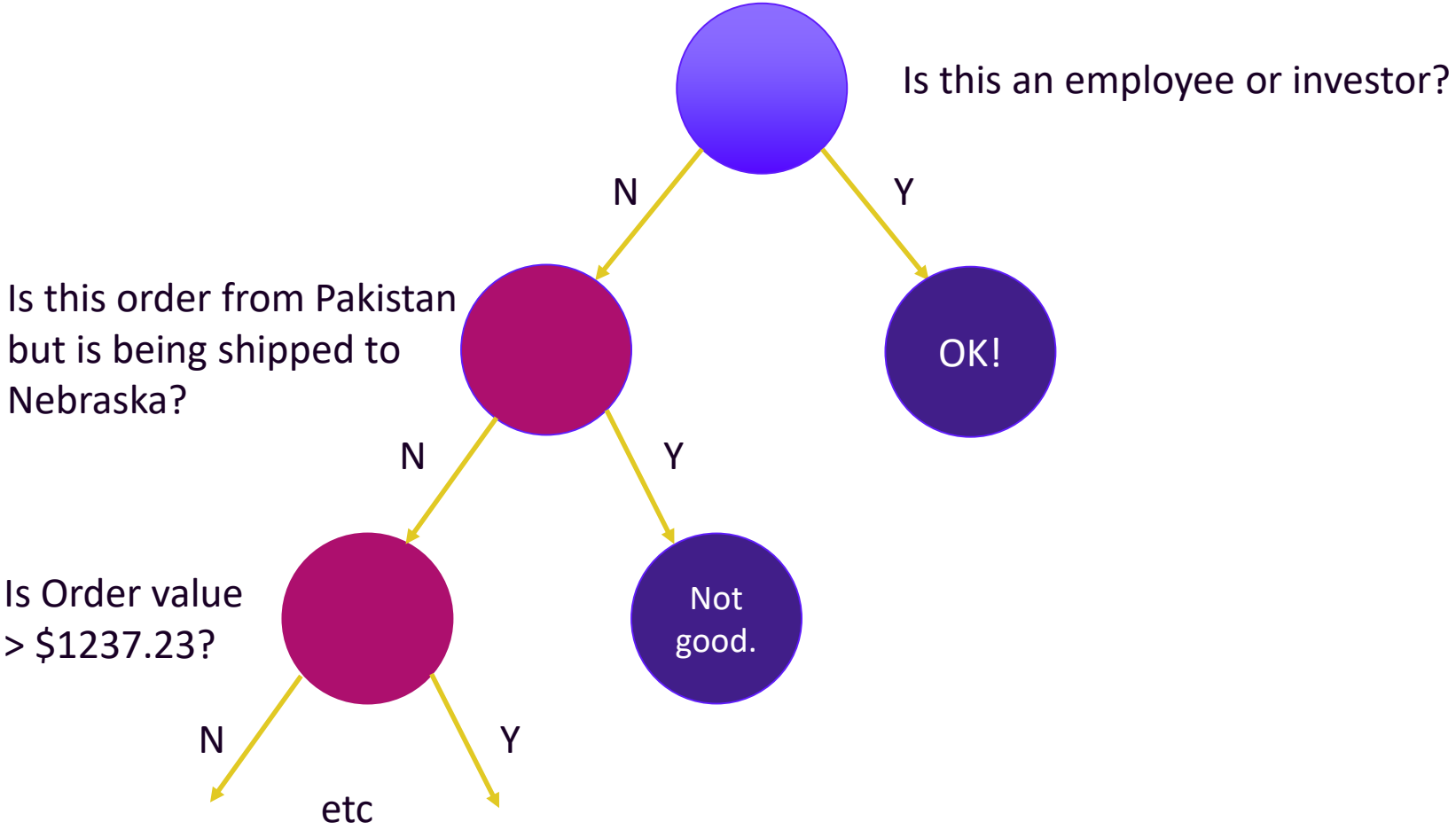
1) How to use a recursive, Information Gain maximization algorithm to create a decision tree
2) Two types of decision tree outputs:
    1) Binary class membership
    2) Probability of class membership

Reading a (Decision) Tree

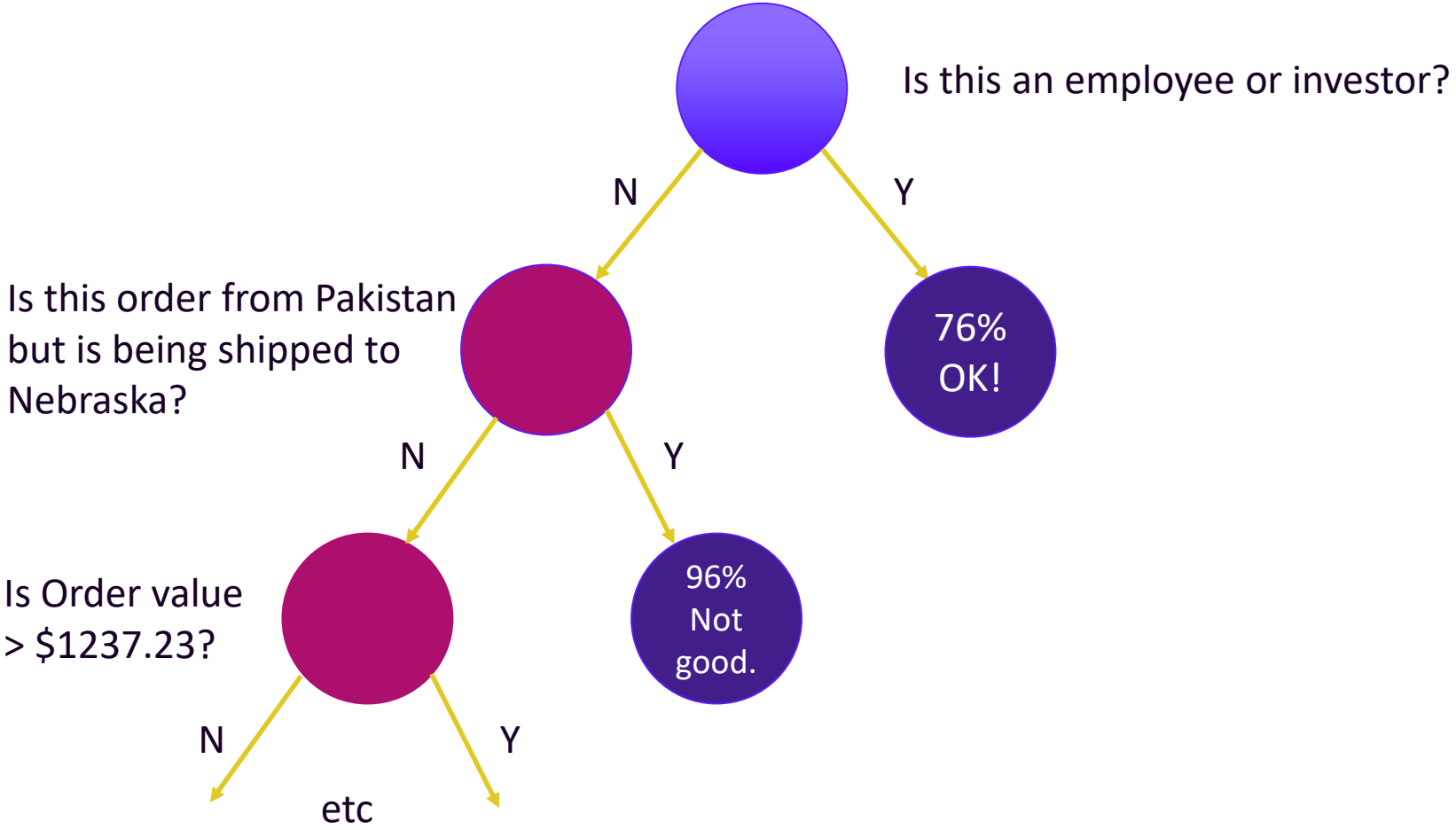
Reading a Tree



Reading a Tree (binary output)



Reading a Tree (probability output)



Arbor Day

How did we get that output?

Claude. M’fing. Shannon.

Claude Shannon is the “Father of Information Theory” based on a landmark paper he published in 1948 (which also provides an absolute limit on the best possible average length of lossless encoding or compression of an information source).

ALSO, as 21-year-old Masters student at MIT he wrote his thesis showing that electrical applications of Boolean algebra could construct any logical numerical relationship. In the last chapter he diagrams several circuits, including a 4-bit full adder. This is the central circuit in all digital computers.

He also helped with codebreaking during WWII. Because of course he did.

Information Theory:
Information Wants to be Free!

Information is a quantity that reduces uncertainty about something.

Information Theory in this context discusses the concept of Entropy, which is essentially how ordered or mixed a system is.

Entropy

The higher the entropy, the more mixed things are.

Low Entropy

Entropy

Equation 3-1. Entropy

$$entropy = - p_1 \log(p_1) - p_2 \log(p_2) - \cdots$$

Each p_i is the probability (the relative percentage) of property i within the set, ranging from $p_i = 1$ when all members of the set have property i, and $p_i = 0$ when no members of the set have property i. The … simply indicates that there may be more than just two properties (and for the technically minded, the logarithm is generally taken as base 2).

Entropy

In terms of actual tactics

The objective of classification algorithms is to reduce entropy as accurately as possible in as few steps as possible

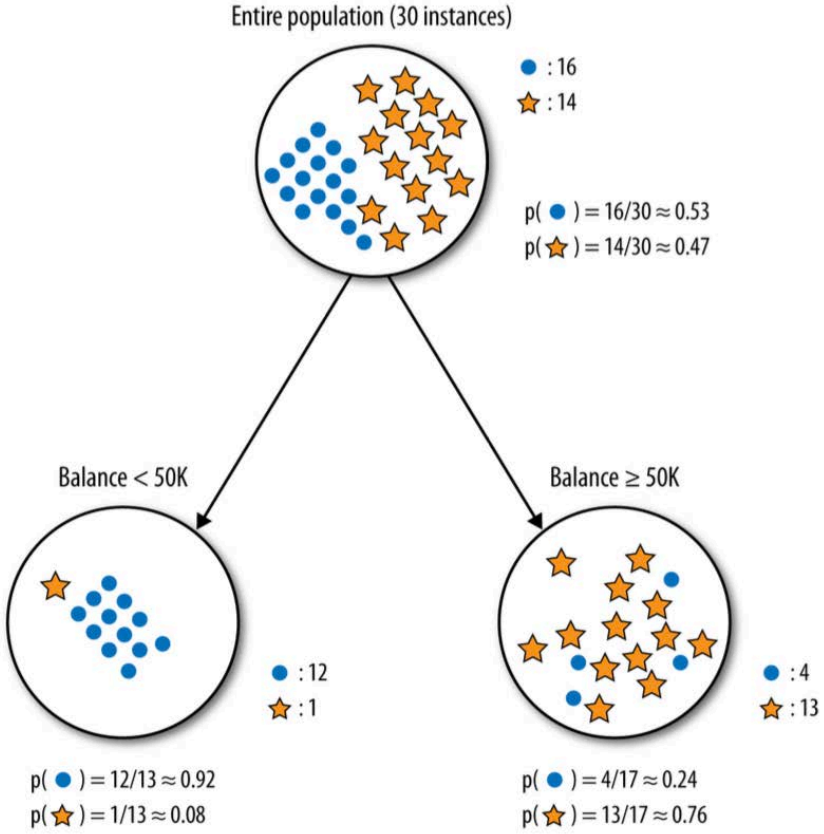
Driver, Roll Up the Partition, Please

We want to measure how much any given split on all values of a single attribute (aka which partitions / decision lines / hyperplanes) decreases entropy.

This is the definition of **INFORMATION GAIN** *(recall, information = a quantity that reduces uncertainty)*

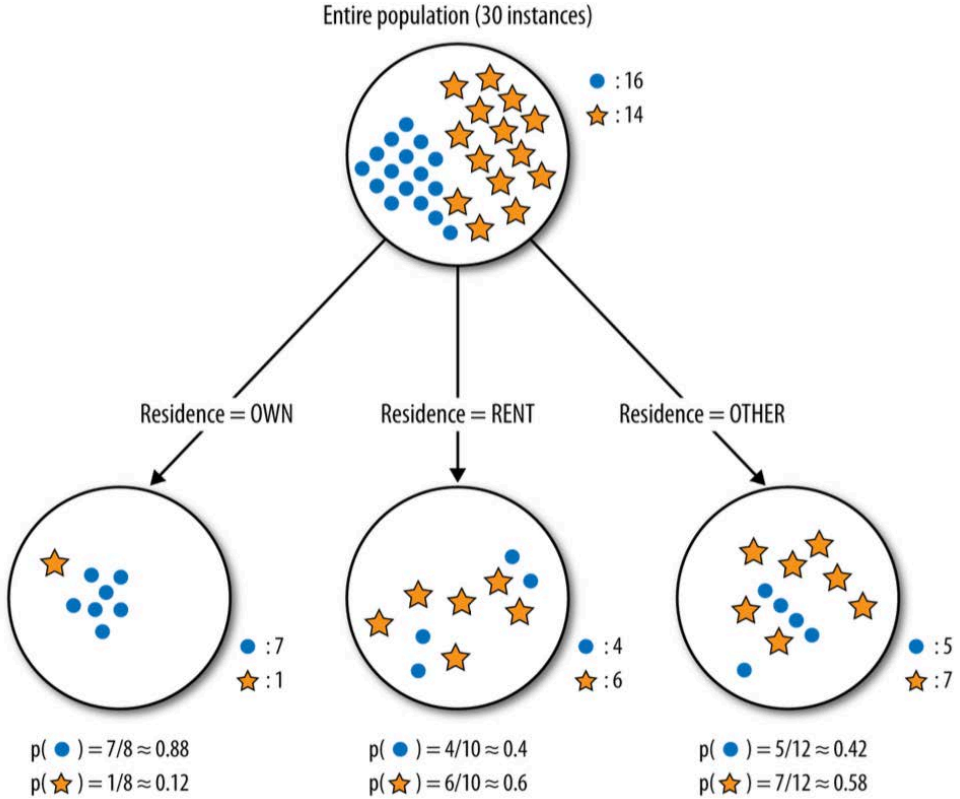
So first, let's look at how the book describes choosing the best feature, measured by Information Gain (p 59-62)

Entropy In Tree Example



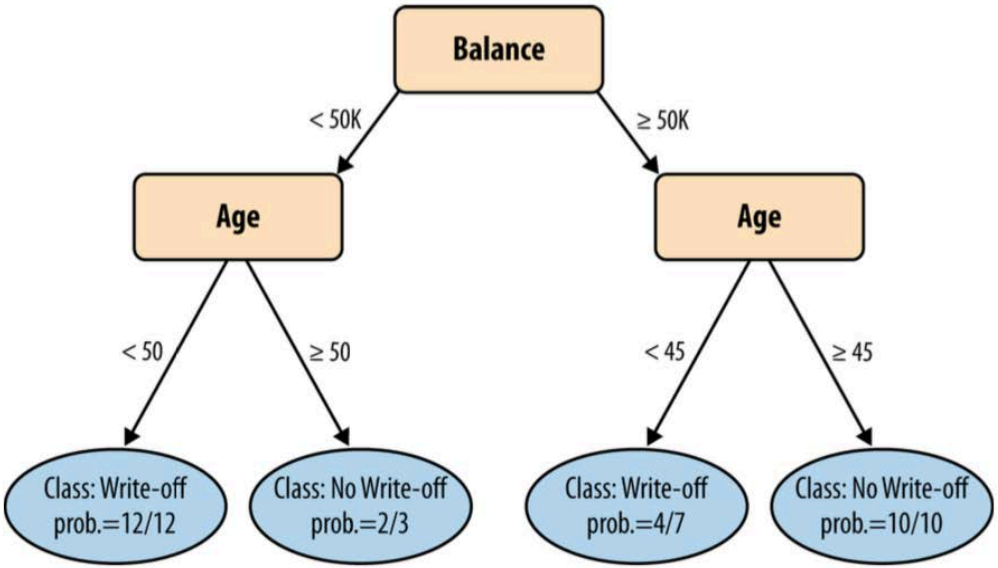
Good Information Gain!

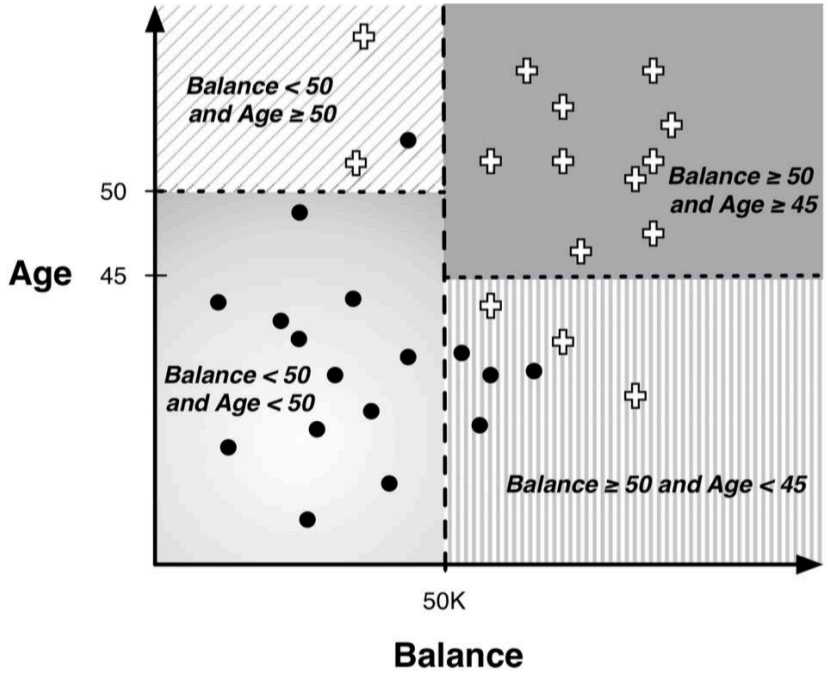
Entropy In Tree Example



Lame Information Gain.

Decision Surface / Partitions





How do we get the Probabilities Instead?

For any given leaf, where

n = instances where something was classified correctly

m = instances where the classification was incorrect

p(c) = n / (n+m)

If you have a very small number of items in a given leaf, might that not over/understate the certainty? **Yes**

Use Laplace Correction:
p(c) = (n+1) / (n+m+2)

How am I going to code this?

Don’t, you fool! It’s already been done by people who are algorithm experts!

Popular algo implementations include…

- CART *(generic form)*
- C4.5 *(award winning form, standard)*
- C5.0 *(commercial upgrade of C4.5)*
- J45 *(open source C4.5)*

Conclusion - The Payoff

Take a dataset

	House Value	Overage Amt	Leftover	College Grad?	Income Level	Call Duration	Churn?
Amos	$520k	98	24	N	$120k	10	**Churn**
Beatrice	$175k	79	7	Y	$85k	12	**Stay**
Clifford	$750k	10	21	Y	$500k	5	**Churn**

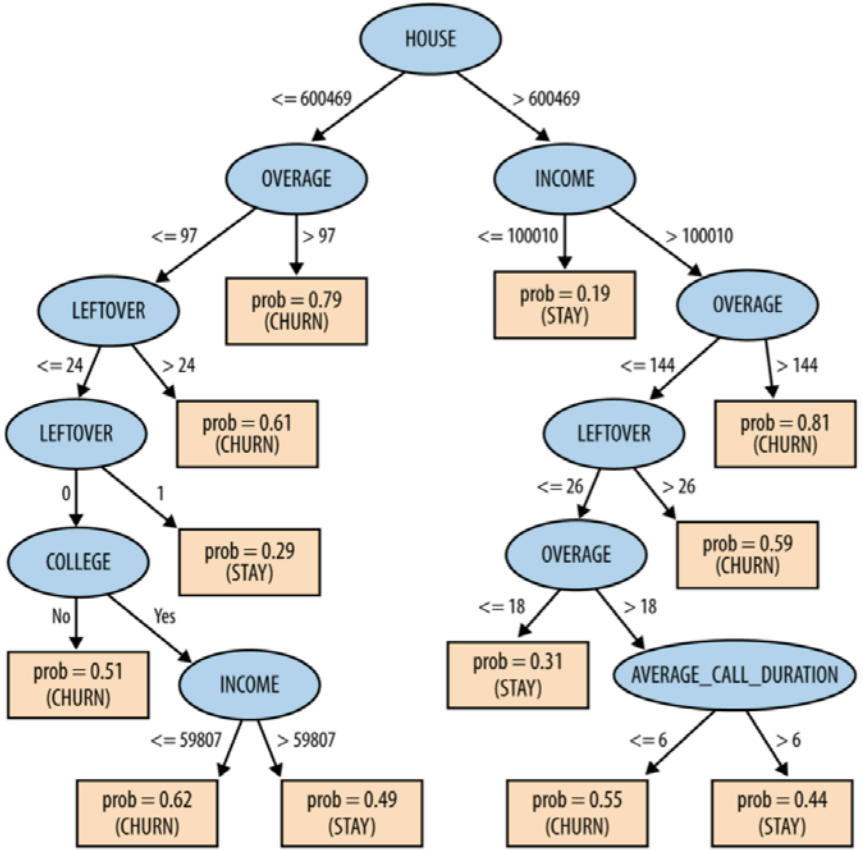
Conclusion - The Payoff

Get it prepared by converting attribute types

	House Value	Overage Amt	Leftover	College Grad	Not College Grad	Income Level	Call Duration	Churn?
Amos	$520k	98	24	0	1	$120k	10	**Churn**
Beatrice	$175k	79	7	1	0	$85k	12	**Stay**
Clifford	$750k	10	21	1	0	$500k	5	**Churn**

Conclusion - The Payoff

Run the algorithm options to create a tree



Conclusion - The Payoff

High five & look at animal photos on Instagram