# Classification Using Tree Based Models

## BUILDING DECISION TREES

**Swetha Kolalapudi**
CO-FOUNDER, LOONYCORN

www.loonycorn.com

# Overview

Recognize classification problems

Understand how decision trees are used to solve classification problems

Understand how machine learning can be used to build decision trees

# Gender Detection

**Given the first name of a user**

**or**

# Weather Forecasting

**Given a time of day**

or

or

# Quant Trading

**Given a trading day**

**or**

# Fraud Detection

**Given a transaction**

**or**

# Classification Problems

## Classifying something into predefined set of categories

**Gender Detection**
- Male, Female

**Weather Forecasting**
- Cloudy, Sunny, Rainy

**Quant Trading**
- Up day, Down day

**Fraud Detection**
- Fraud, Not fraud

Classification Problems

We are given a **problem instance**

**A name**

**A time of day**

**A trading day**

**A transaction**

# Classification Problems

We need to assign a **label** to the problem instance

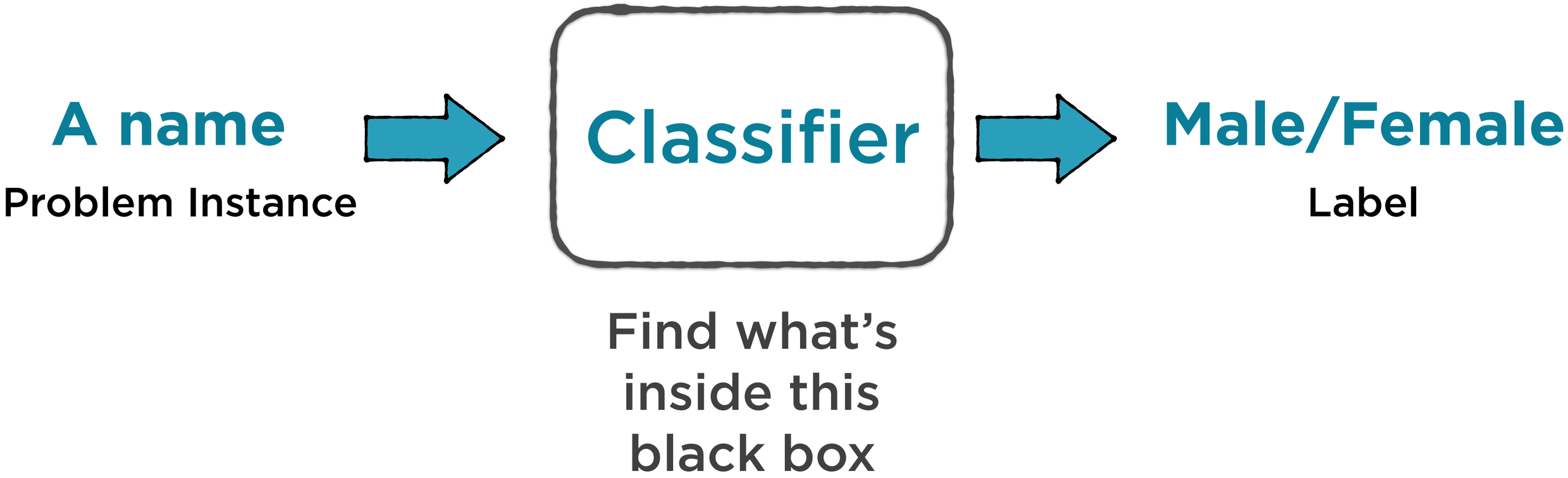
**Male** or **female**?

**Cloudy** or **rainy** or **sunny**?

**Up-day** or **down-day**?

**Fraud** or **Not fraud**?

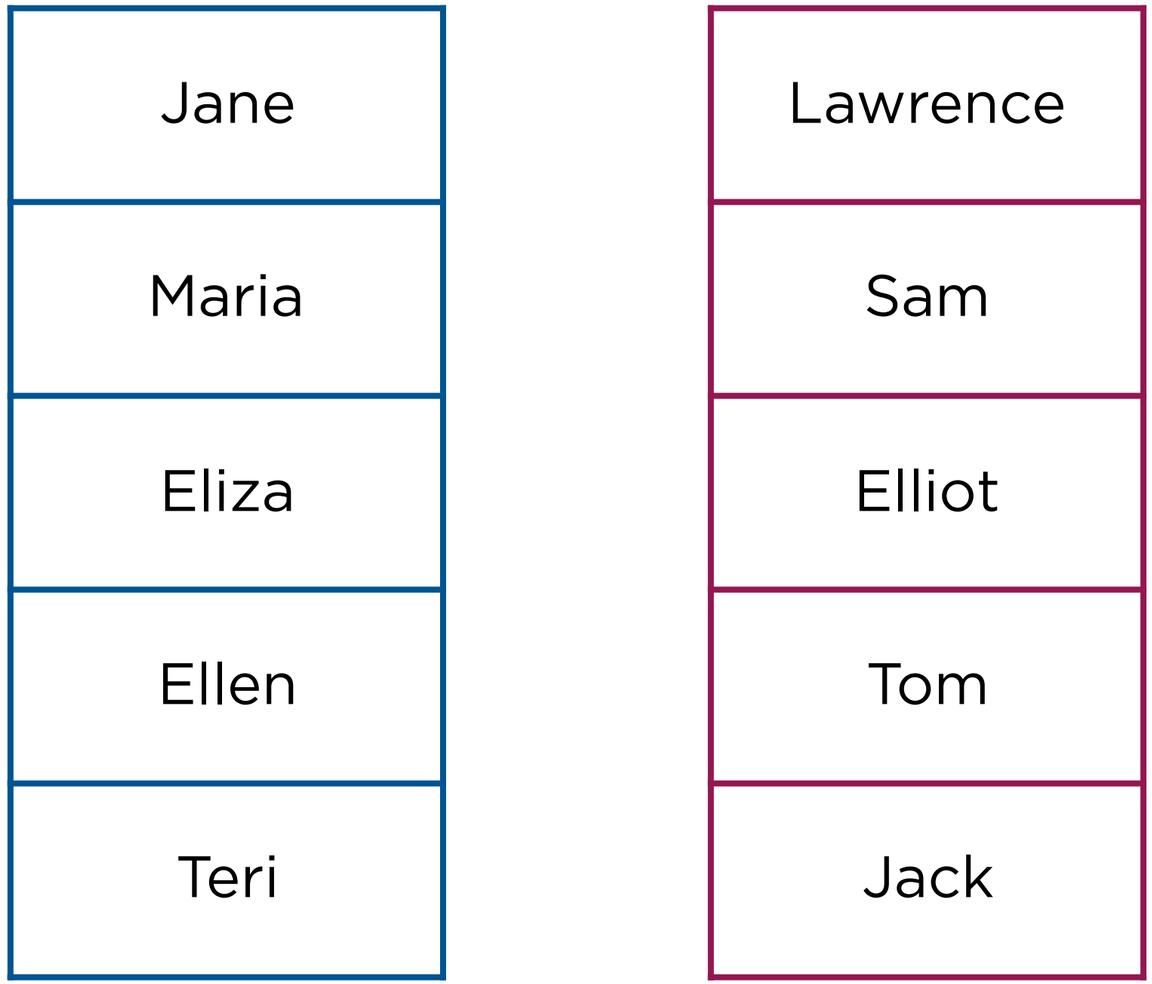
# Solving Classification Problems



One way to solve
classification problems

# Define a set of rules

**Find the set of rules that can classify these names correctly**



**Simply do a visual inspection**

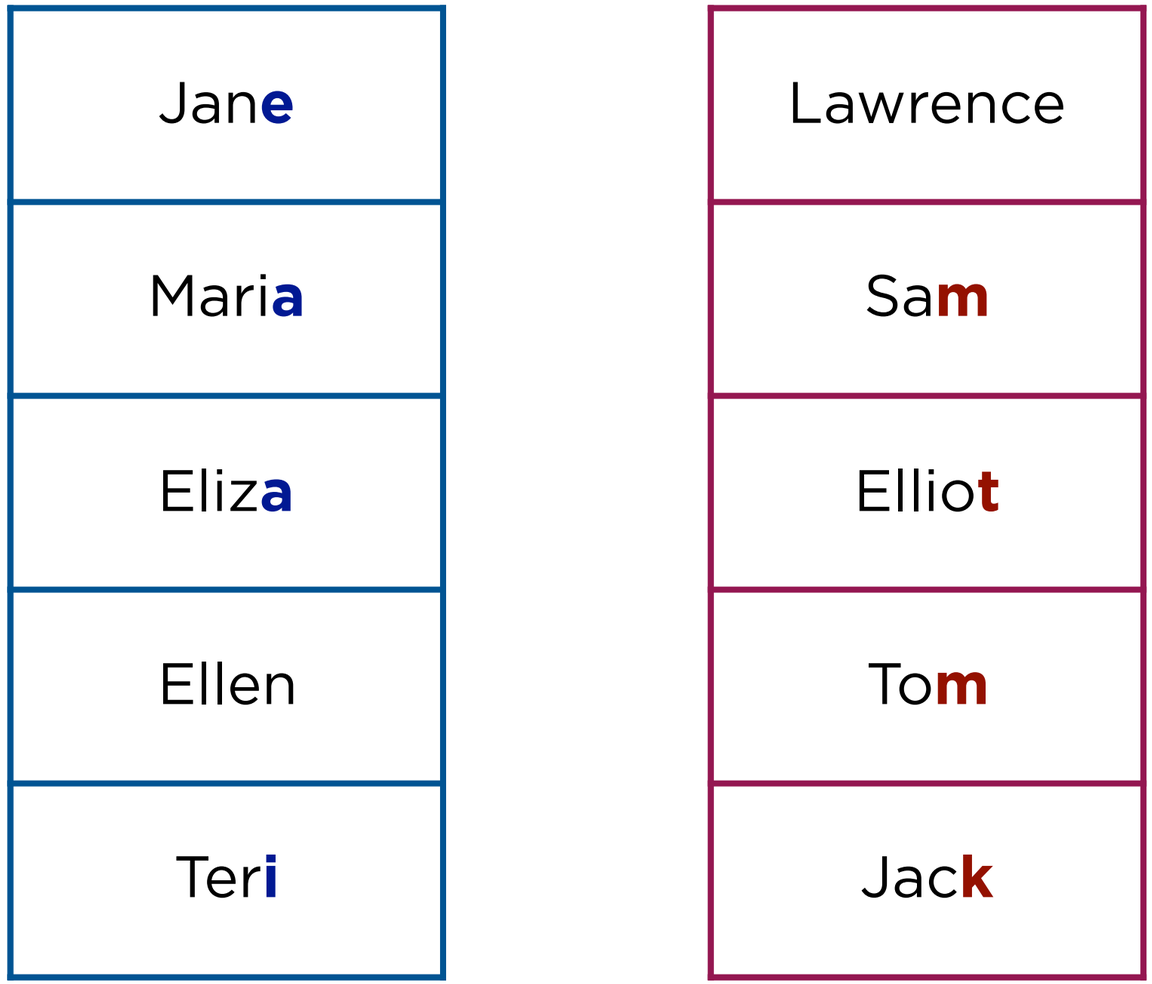
| |
|---|
| Jane |
| Maria |
| Eliza |
| Ellen |
| Teri |

| |
|---|
| Lawrence |
| Sam |
| Elliot |
| Tom |
| Jack |

In most cases

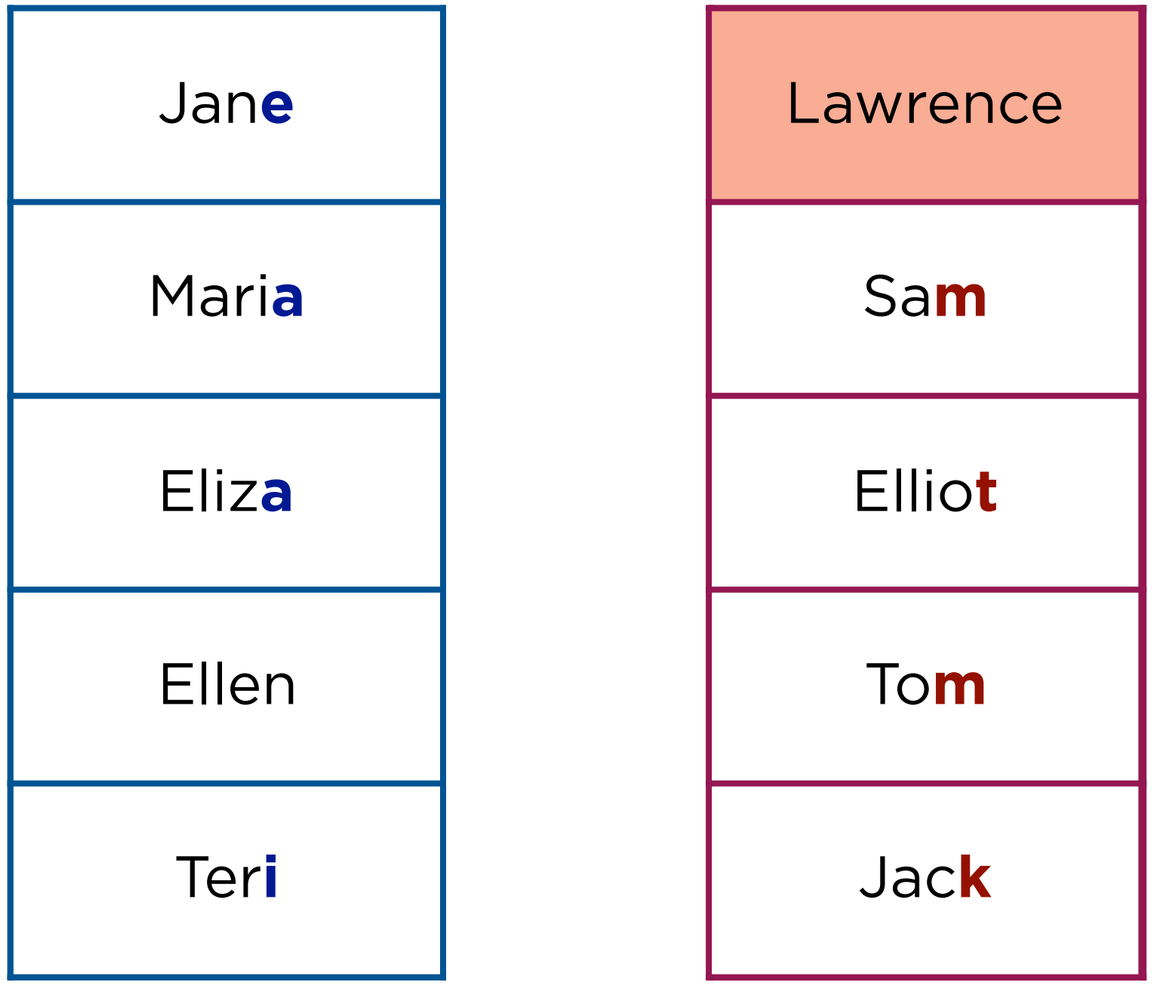
**Female** first names
end in **vowels**

**Male** first names
end in **consonants**



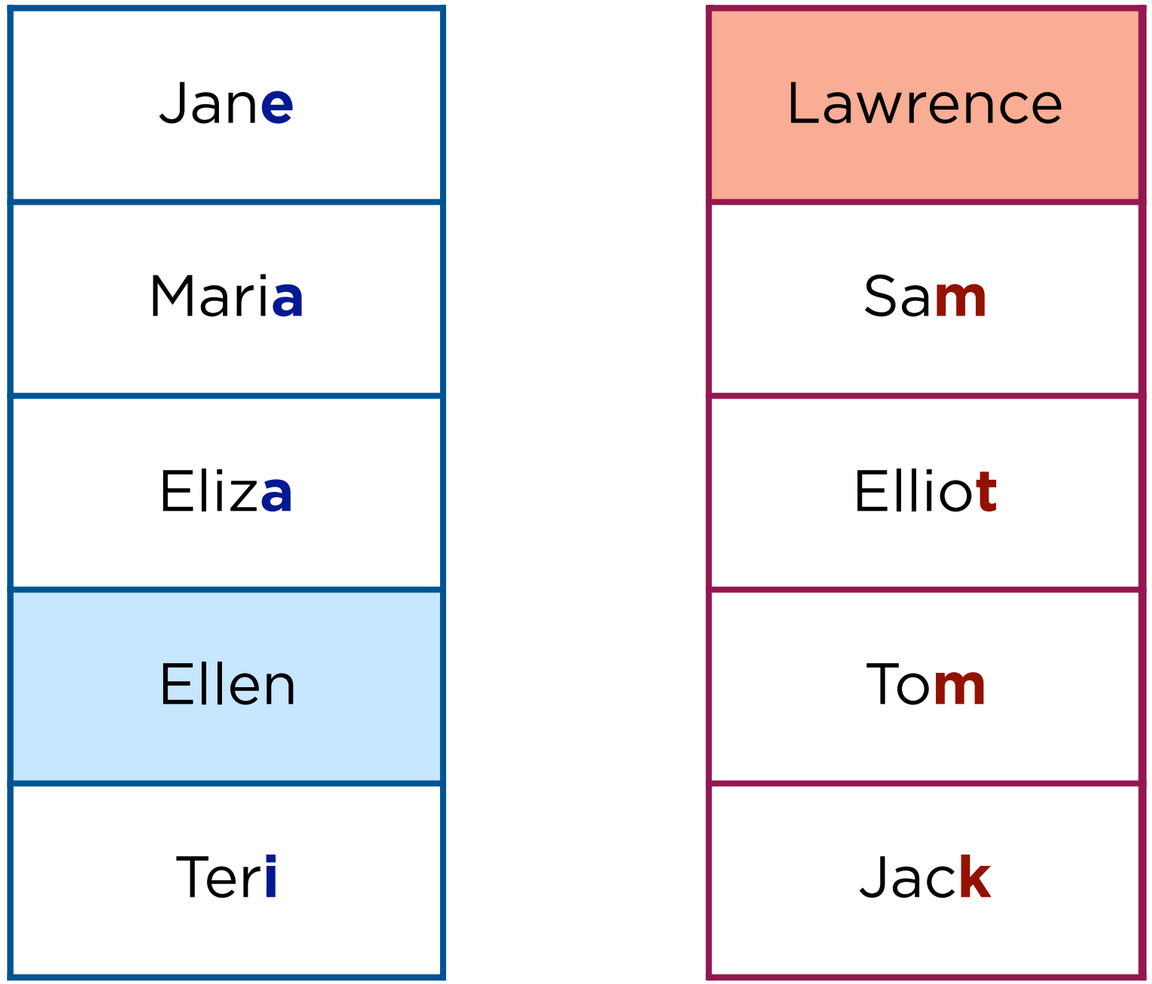
**Except..**

**Vowel-ending names which begin with L are male names**

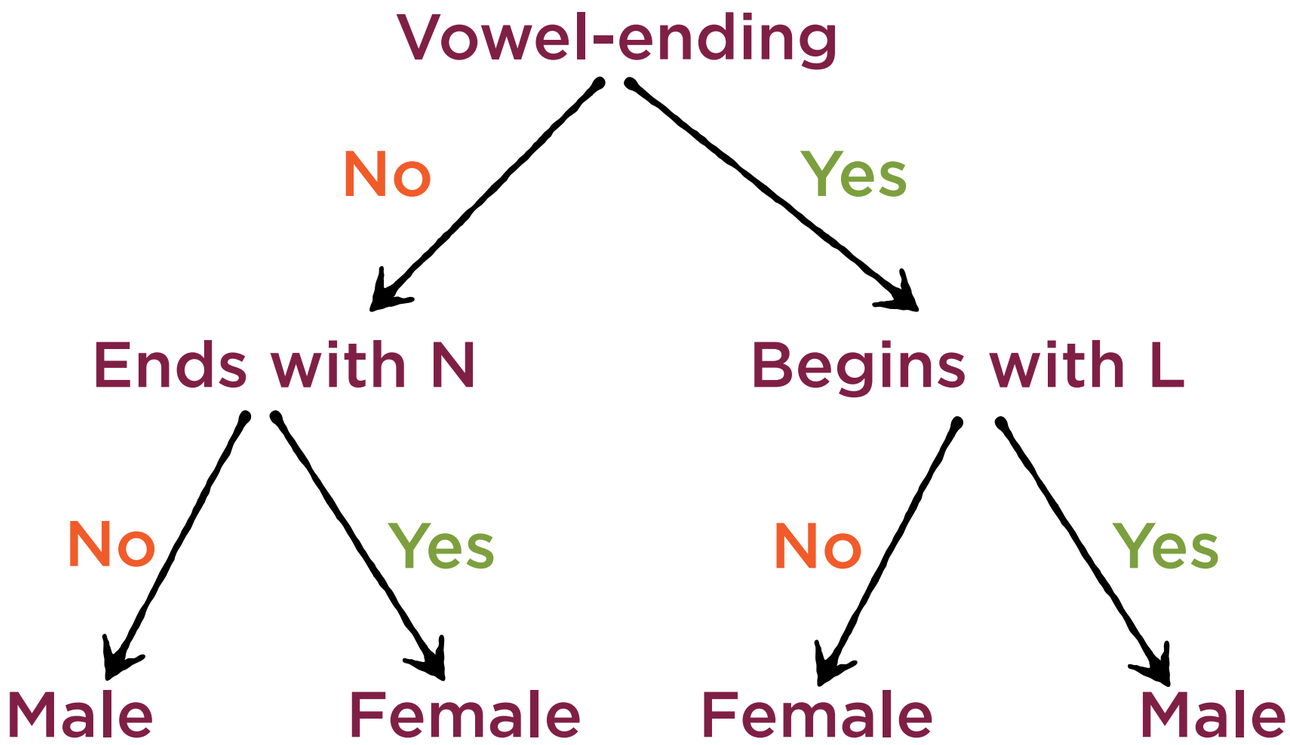


**Except..**

**Consonant-ending names which end with N are female names**



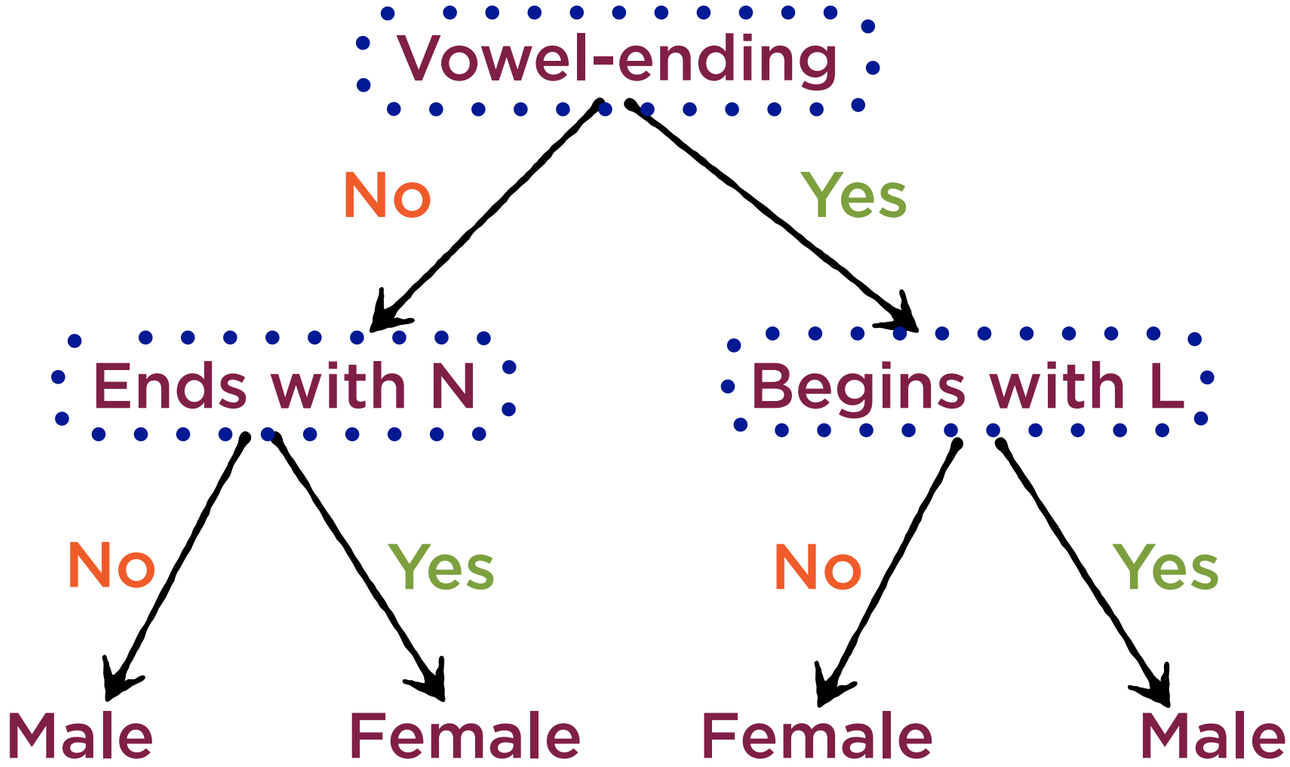
Visualize these rules using a tree representation



The tree represents a series of choices

i.e. decisions

# Decision Tree



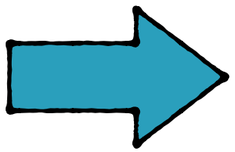
# Decision Tree

A set of rules
used to classify
something

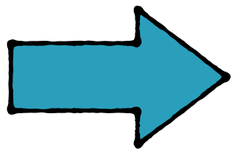
# Building a Decision Tree

**Training Data**

| | |
|---|---|
| Jane | Lawrence |
| Maria | Sam |
| Eliza | Elliot |
| Ellen | Tom |
| Teri | Jack |



**Machine Learning Algorithm**



**Decision Tree**

# Tree Based Models

**Machine learning algorithms which build decision trees from training data**

# Tree Based Models

**Decision Tree Learning Algorithms**

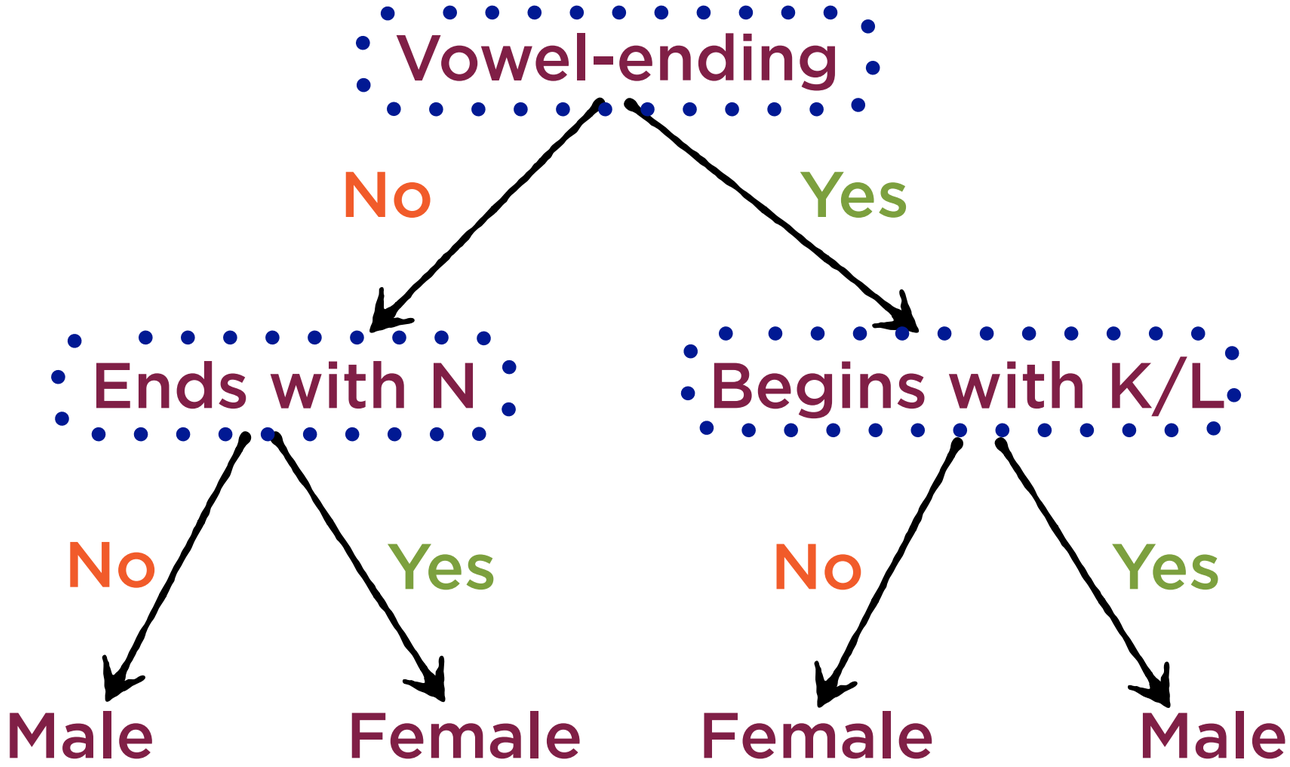
Build a single decision tree

**Ensemble Learning Algorithms**

Build multiple decision trees and combine their results

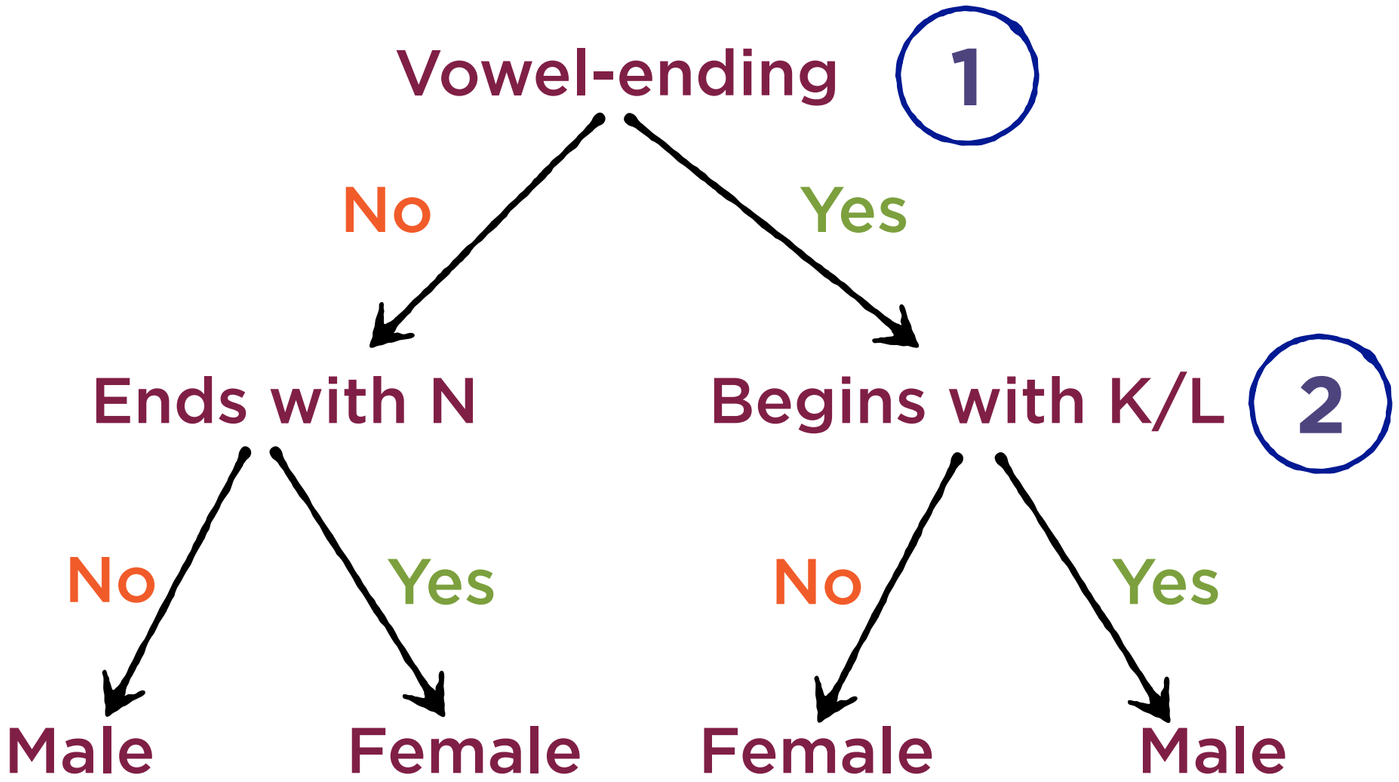
Each decision depends on the value of some attribute of the problem instance

# Decision Tree



The order in which we look at the attributes is important

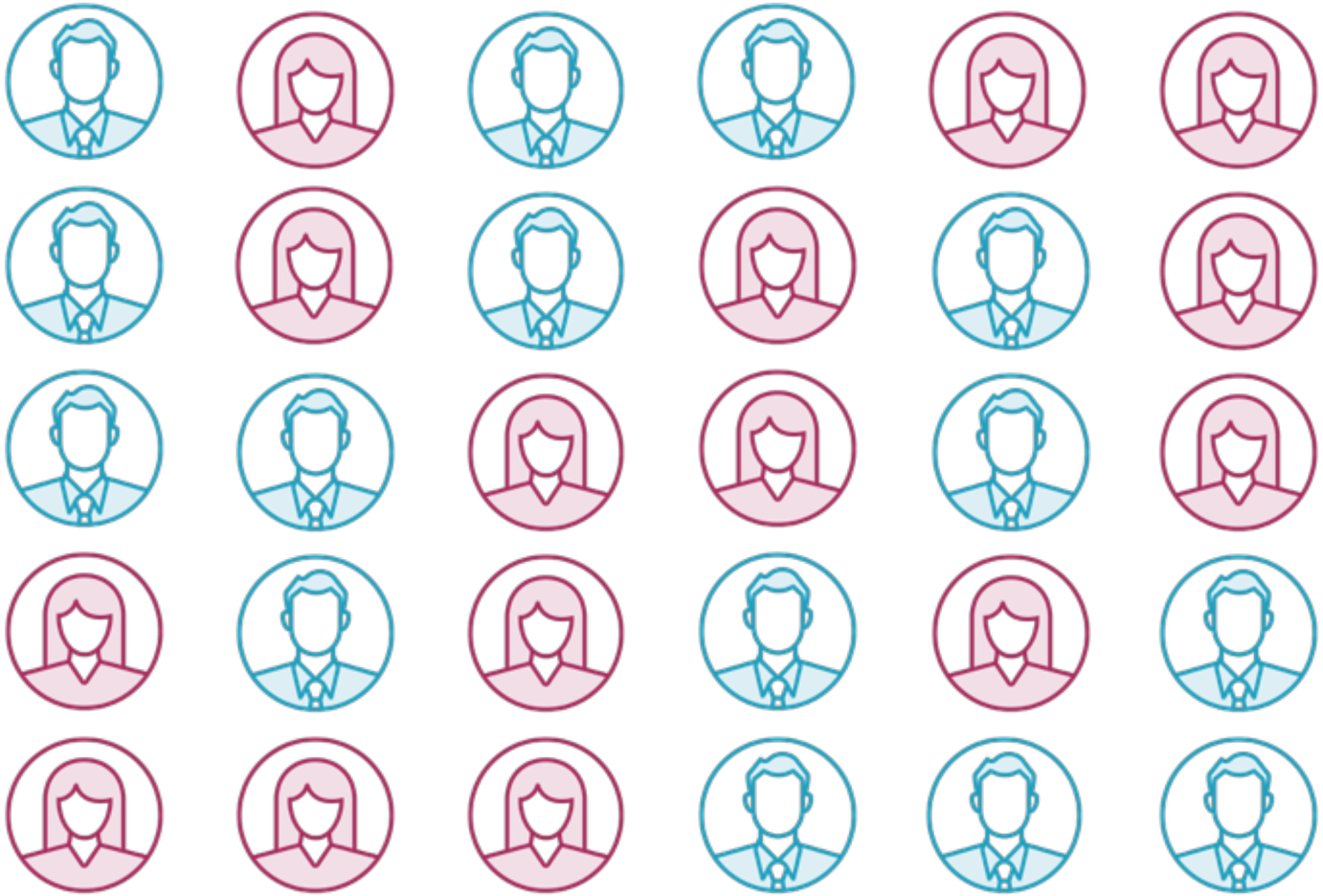
# Decision Tree



# How Decision Tree Learning Works

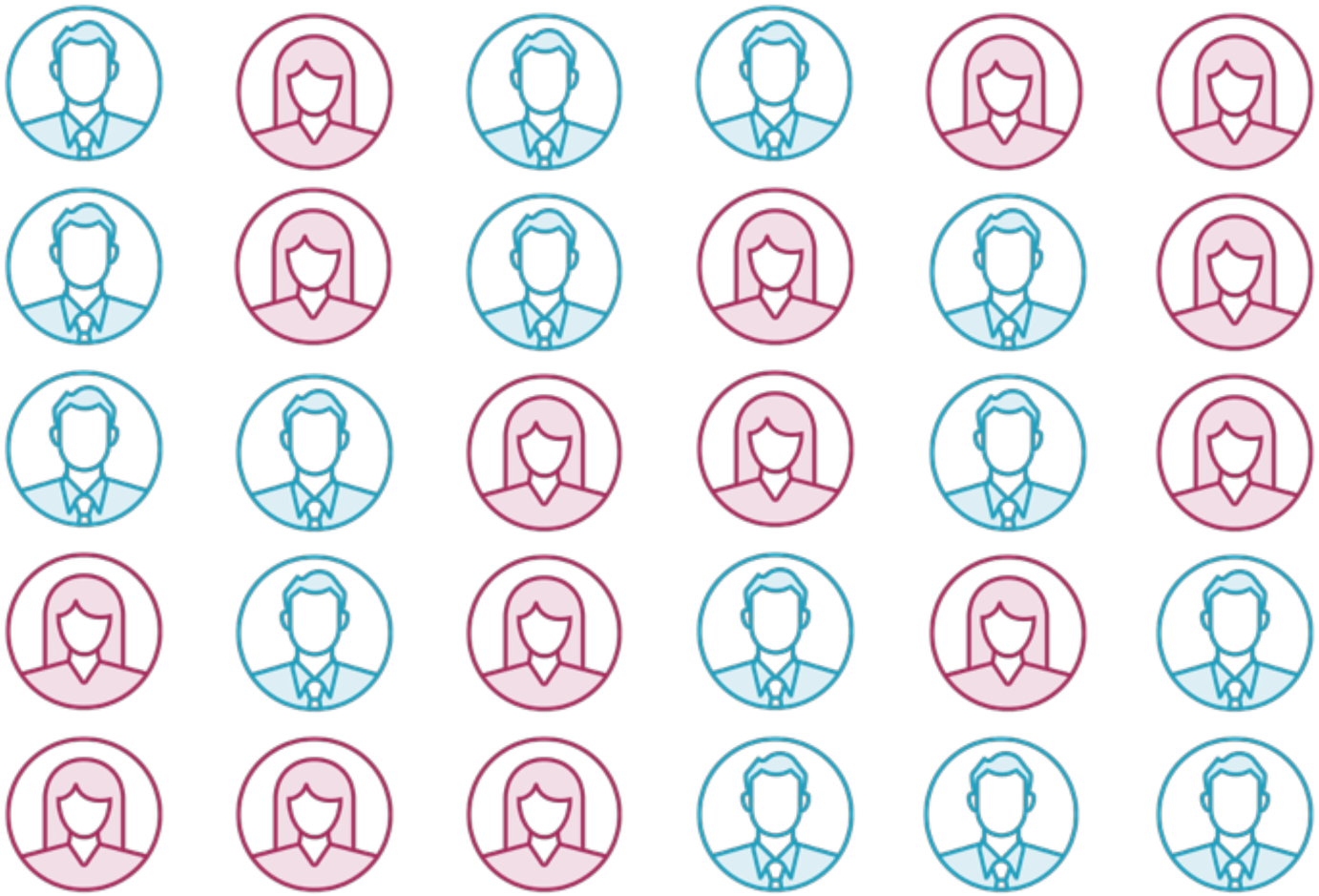
**Choose an attribute/ feature that divides the training data into homogenous subsets**

# How Decision Tree Learning Works



**Training Data**

# How Decision Tree Learning Works

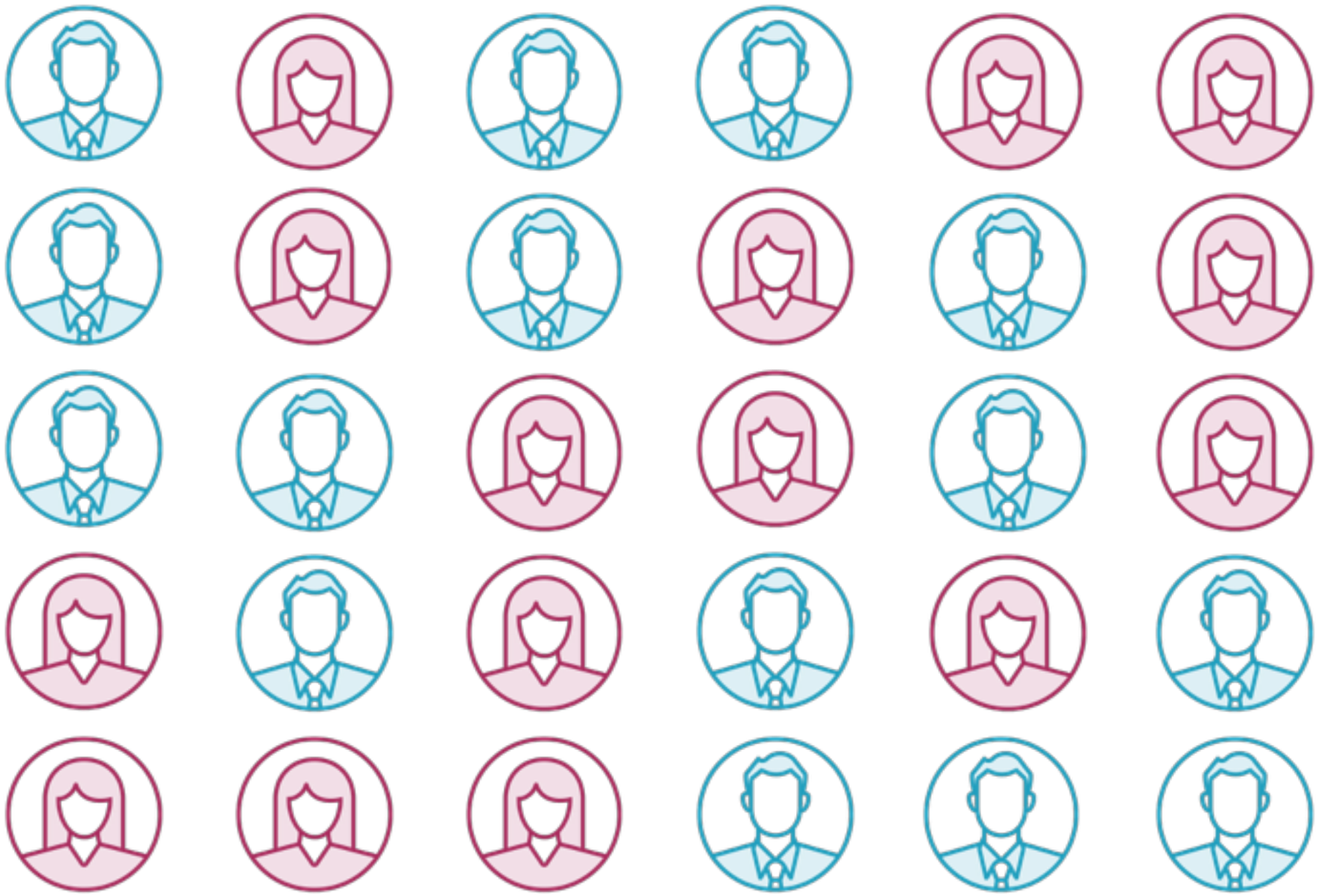


**Attributes are**

**Vowel-ending?**

**Begins with K?**

# How Decision Tree Learning Works



Attributes are

Vowel-ending?

**Begins with K?**

# How Decision Tree Learning Works

Yes

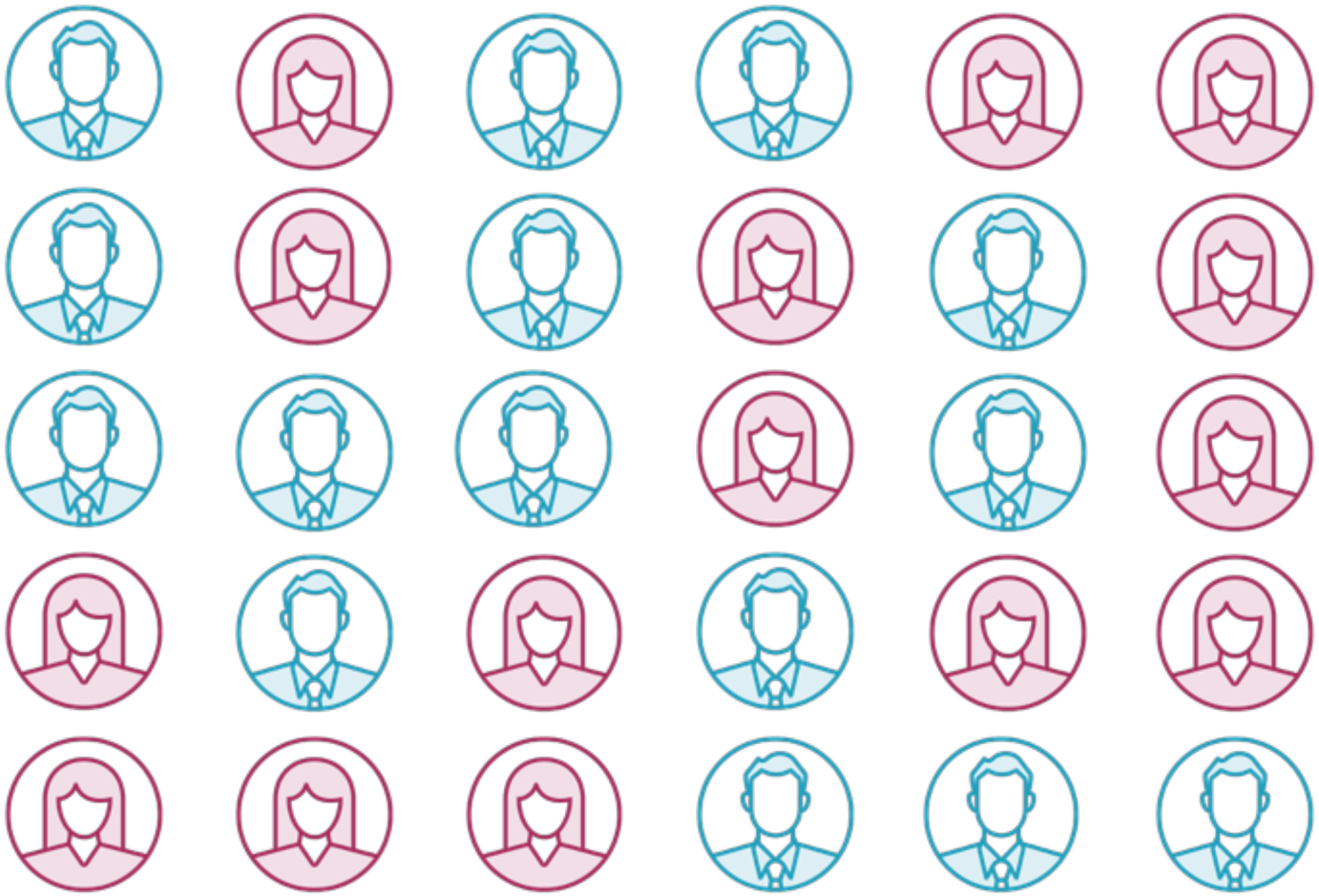
No

Attributes are

Vowel-ending?

**Begins with K?**

# How Decision Tree Learning Works



Attributes are

**Vowel-ending?**

Begins with K?

# How Decision Tree Learning Works

Yes

No

Attributes are

**Vowel-ending?**

Begins with K?

# How Decision Tree Learning Works

Vowel-ending?

**Leads to more homogenous subsets**

Begins with K?

**Leads to non-homogenous subsets**

# How Decision Tree Learning Works

Vowel-ending?

**Gives us more information**

Begins with K?

**Gives comparatively less information**

# How Decision Tree Learning Works

Vowel-ending?

**First attribute in the decision tree**

Begins with K?

**Next attribute in the decision tree**

# How Decision Tree Learning Works

Vowel-ending?

Begins with K?

Ends with N?

**If there are more attributes, repeat this process within each subset using the remaining attributes**

# How Decision Tree Learning Works

**Information Gain**

**Gini Impurity**

**Ways to measure homogeneity of the subsets formed**

# Summary

Recognize classification problems

Understand how decision trees are used to solve classification problems

Understand how machine learning can be used to build decision trees