**Artificial Intelligence & Machine Learning**

***Decision Tree Activity***

Alexander Hrabski & Zena Lapp

We have designed this to have as much flexibility as possible so that you may change this lesson plan as you see fit (e.g. more or less detail and guidance).

**Activity summary**

Big data, artificial intelligence, and machine learning are often buzzwords in the news, but what do they actually mean? In this activity, students will be introduced to these topics and will implement a machine learning method themselves!

**Materials**

* Digital slide deck (if you want to give an intro presentation) & way to present it.
* Printout of the **DisneyCharacterGroups.pdf** and **TestData.pdf** *(don’t hand this one out until the end!)* for each individual/group. These are in the **Activity Printouts** folder.
* Large sheet(s) of paper for each individual/group.
* Pencil/pen for each individual/group.

**Suggested learning objectives**

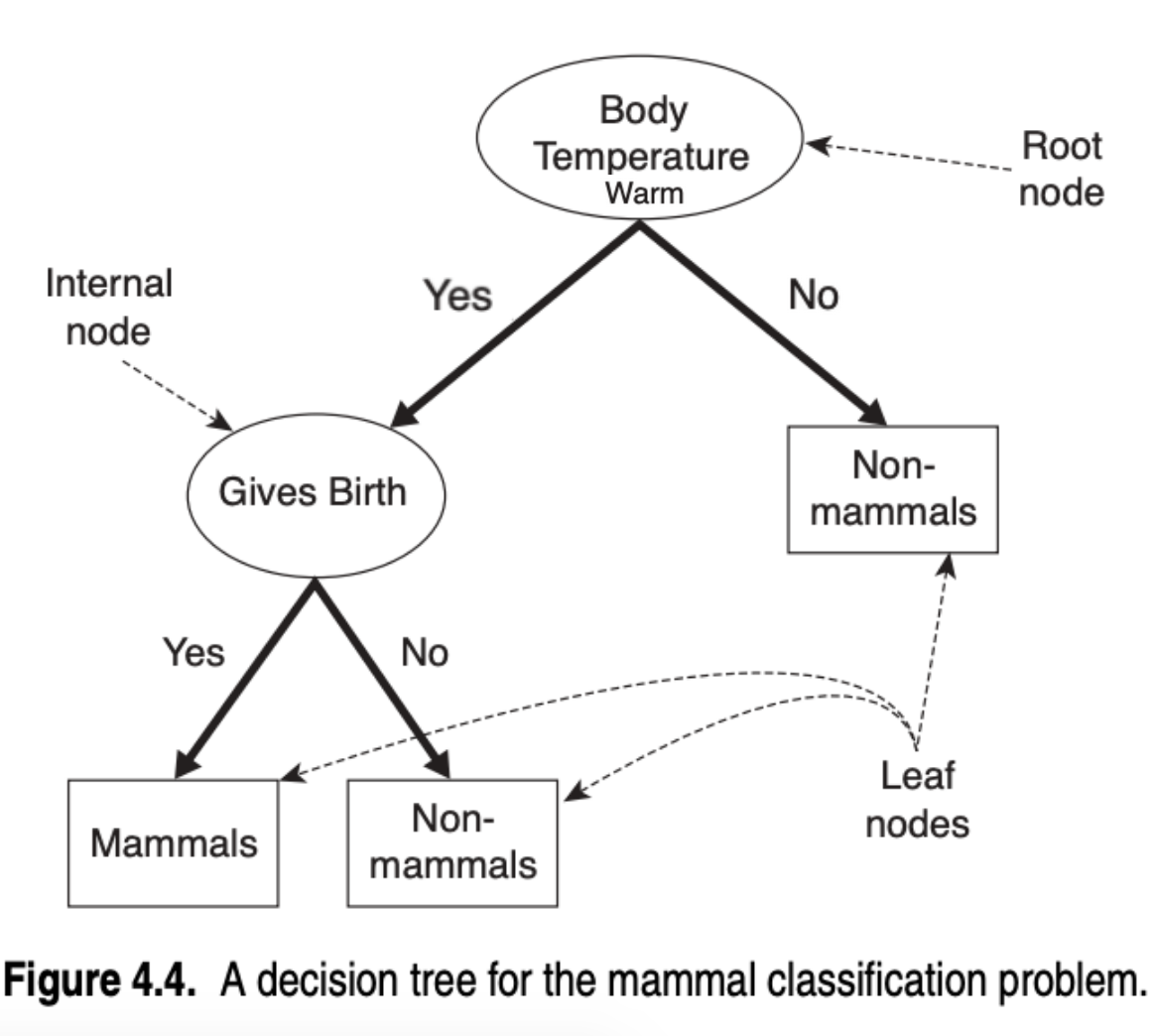
* Describe what data science is and why it is useful.
* Conceptually explain what artificial intelligence and machine learning are.
* Understand the concept of a decision tree and how you use it for classification.
* Generate a decision tree.

**Background information**

Computer science permeates our everyday lives - we use it every time we search for something on the internet, or use an app on our phone, and even when we check out at the grocery store. A lot of computer science focuses on analyzing large sets of data (i.e “big data”) -- this process is called *Data Science*. Often, we use *Artificial Intelligence (AI)* to analyze large datasets. AI describes machines (or computers) that mimic cognitive functions that people associate with the human mind, such as learning and problem solving. One subset of AI is *Machine Learning (ML)*, which is when the computer uses a set of instructions to find patterns in the data to classify items into different categories, or to predict a certain outcome. Some examples are classifying different types of vehicles (e.g. cars, trucks, boats, trains, airplanes) into categories, classifying patients as at risk for getting an infection, and checking for credit card fraud (fraud or not).

The set of instructions that the computer uses to perform machine learning is called an *algorithm*. Algorithms are the basis of all computer programs because they tell the computer exactly what it needs to do. There are all sorts of machine learning algorithms that are specialized for different types of data or tasks. A lot of machine learning algorithms are made to sort things into different categories (often called classes) based on differences between them.

One machine learning algorithm is called a *decision tree*. This algorithm splits the data into classes using a set of *rules* based on certain *features* in the data. These rules can be thought of as yes or no questions. The next question that the algorithm asks depends on the answer to the previous question. Below is an example of a decision tree to classify animals as mammals or non-mammals. In this example, the features are body temperature (is the body temperature warm?), and gives birth (does the animal give birth?).



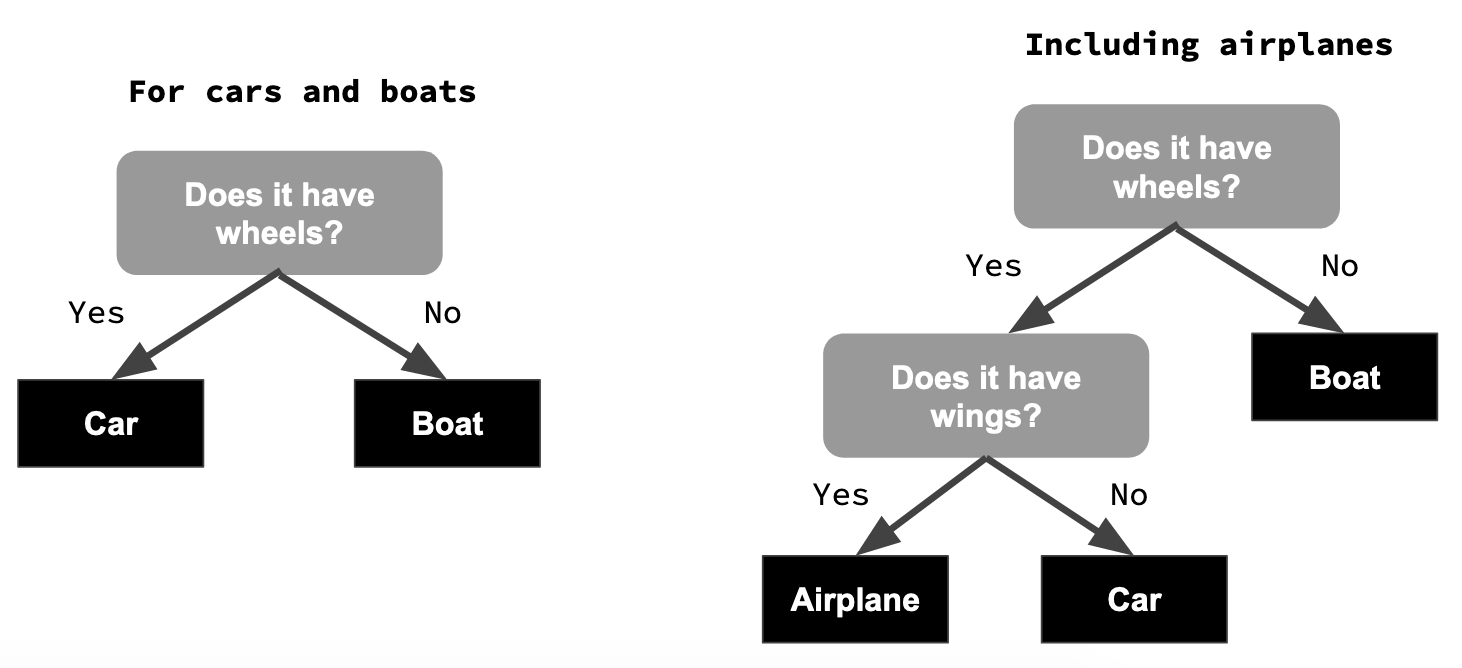
In machine learning, you first use *training data* to make a set of rules that works well to classify that data into different categories. For instance, the training data for the above decision tree could be snake, human, cat, bird, and fish. This set of rules is called a *model*. You can try out this model on new *test data* that wasn’t in your training data to see how well it performs on something it’s never seen before. Sometimes, the test data works - for instance, if you tested this model out on ants, dogs, and elephants it would work. But biology is complicated, and some mammals are not really warm- or cold-blooded ([ref](http://www.bbc.com/earth/story/20160609-time-to-bust-a-myth-not-all-mammals-are-warm-blooded); e.g. the Arctic ground squirrel). In this case, the model would not classify the animal correctly. Testing your model on new data provides an idea of the *generalizability* of your model to classify - i.e. how well it works on data it’s never seen before. For the mammal example, it usually works pretty well, but there are some cases when it doesn’t work. This shows that a model is only as good as the data you use to train it. Depending on your goal, your model does not necessarily need to be perfect; in fact, it’s almost impossible to make a perfect model for complicated data.

**Preparation steps**

1. Print the provided **DisneyCharacterGroups.pdf** and **TestData.pdf** *(don’t hand this one out until the end!)* for each individual/group. These are in the **Activity Printouts** folder.
2. Cut out the individual characters from **DisneyCharacterGroups.pdf** (one set for each individual/group).

**Activity protocol**

1. Go over background info (data science and/or AI & ML presentations)
2. Have the class do a simple decision tree (as an entire class or in groups):
   1. *Question for class:* How can we classify cars and boats using a decision tree? What would the first rule/question be?
      1. *Possible answer:* Does it have wheels?
         1. Yes: Car
         2. No: Boat
   2. *Question for class:* With the current decision tree, what would an airplane be classified as?
      1. *Answer:* Car
   3. *Question for class:* What rule can we add to classify airplanes correctly?
      1. *Answer:* After checking if it has wheels: Does it have wings?
         1. Yes: Airplane
         2. No: Car

Example answer visual: ****

1. As individuals or groups:
   1. *Part 1:* Build a Decision Tree to separate the animal Disney characters into their subcategories: Fish, Birds, Reptiles, Arthropods, Marsupials, and Mammals
   2. *Part 2:* Build a Decision Tree to separate the Disney characters into three categories: Human, Animal, and Other
   3. **Note:** For Parts 1 and 2 you will have to “*extract*” your own features for the characters to build your Decision Tree rule set. In other words, you have to choose features to use to ask your yes/no questions.
   4. *Part 3:* Come together as a class and share your Decision Trees with each other.
   5. *Part 4 (don’t show this to the students until after they finish parts 1 and 2):* Check your Decision Tree from Part 2 on the test data (this data is in **TestData.pdf**, and also in the presentation).
      1. Do this by going through your decision tree rules (asking the questions) for each of the Disney characters in the test data.
      2. Were they classified correctly? If not, why not? How could you fix your decision tree to classify them correctly?
      3. **Lesson:** a model is only as good as the data you use to *train* it. If you don’t train the model with diverse data, it won’t work on diverse data.
2. Note: In the Artificial Intelligence presentation, there is an additional activity related to deep learning. I kept this in the slides in case you’re interested in using it, but it requires computers so it is not a part of this activity.

**Additional resources**

*Data Science*

* Overview of what data science is and what data scientists do
  + <https://datascience.berkeley.edu/about/what-is-data-science/>
* Difference between data scientist, analyst, and engineer
  + <https://www.innoarchitech.com/blog/what-is-data-science-does-data-scientist-do>

*Artificial intelligence*

* Overview of AI
  + <https://www.youtube.com/watch?v=2ePf9rue1Ao>
* Applications of AI/AI & Society (Royal Society)
  + <https://www.youtube.com/watch?v=nASDYRkbQIY>
* Interactive websites:
  + Train AI with Image: <https://teachablemachine.withgoogle.com/train/image>
  + Ocean AI Activity: <https://studio.code.org/s/oceans/stage/1/puzzle/3>
  + Quick Draw: <https://quickdraw.withgoogle.com/>
  + Recognizing Handwritten Digits: <http://scs.ryerson.ca/~aharley/vis/conv/flat.html>
  + Tensor flow: <https://playground.tensorflow.org>
* Ethical concerns of AI
  + <https://www.weforum.org/agenda/2016/10/top-10-ethical-issues-in-artificial-intelligence/>
  + <https://www.forbes.com/sites/cognitiveworld/2020/12/29/ethical-concerns-of-ai/#1b072c0523a8>
  + <https://towardsdatascience.com/ethics-of-ai-a-comprehensive-primer-1bfd039124b0>
  + <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>

*Machine learning*

* Overview of machine learning
  + <https://www.youtube.com/watch?v=ukzFI9rgwfU>
* Different types of machine learning algorithms
  + <https://towardsdatascience.com/types-of-machine-learning-algorithms-you-should-know-953a08248861>
  + <https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms/>
  + <https://machinelearningmastery.com/a-tour-of-machine-learning-algorithms/>
  + Analogies to war/armies (not sure this is school-appropriate, but it helps with the intuition behind different ML algorithms): <https://www.analyticsvidhya.com/blog/2015/12/10-machine-learning-algorithms-explained-army-soldier/>
* Long list of types and applications of ML
  + <https://en.wikipedia.org/wiki/Outline_of_machine_learning>