Raspberry Pi Foundation

Lesson 4: Decision trees

Enjoyed teaching these lessons? Found a mistake? Share your feedback at rpf.io/experienceai-feedback.

Introduction

In this lesson, students will take their first in-depth look at a type of model: decision trees. The activities build on students' learning from Lessons 1–3 about classification, training and test data, and the data-driven nature of models. The aim of this lesson is for students to gain an understanding of the processes used to create machine learning models.

First, students will learn about the structure of a decision tree, introducing them to the key terminology and parts of a decision tree. Then, they will see how a decision tree is used to process data and predict a label.

Next, students will see how a decision tree is made using training data. You will demonstrate the process for them using the slide deck, then they will apply the process independently with new training data. Students will also have an opportunity to see what the term 'data-driven' really means, as the two decision trees that they create with separate training data will be different.

Finally, students will explore the reasons machine learning is useful when creating decision trees, with regard to both scale (data sets used to create ML models are very large) and adaptability (being data-driven). They will use Machine Learning for Kids to create a decision tree using a larger set of data, and will use the decision tree that they create.

Learning objectives

- Describe how decision trees are used to build a classification ML model
- Describe how training data changes an ML model
- Explain why ML is used to create decision trees

Key vocabulary

Decision tree, feature, node, root node, decision node, leaf node, classification, explainability

Preparation

You will need:

- Slides
- Activity 2 Part 1 worksheet and Part 2 worksheet
- Activity 3 worksheet

Subject knowledge:

Activity 1: In this activity, you will guide the students through the key terminology of a decision tree, namely:

1. The root: the top node or box in a decision tree
2. A decision node: any node or box in a decision tree that contains a condition
3. A leaf: a node or box without a condition that usually represents a single class

Activity 2: This activity involves you demonstrating the method used to create a decision tree, choosing conditions to split the training data. In reality, decision trees are made through trial and error, so it is not important that your splits are completely correct, nor that the students' are.

It is recommended that you complete the activity yourself ahead of the lesson so you know what the splits will look like.

Activity 3: In this activity, the students will watch a video to guide them to create a star classifier using Machine Learning for Kids. There is also a step-by-step guide on the worksheet. You may need to assist any students who encounter difficulties, so it is recommended you watch the video and follow along in advance of the lesson so you know what all the steps are.

Assessment opportunities

The starter activity is an opportunity for you to check what students have learned about classification so far, and to challenge any alternative conceptions they may have before starting the lesson.

In Activity 3, the students will complete a self-guided activity using Machine Learning for Kids. This will give you an opportunity to have discussions with students about the major concepts covered so far in the unit, such as training a machine learning model, the split between training and test data, and the key terminology of classification (for example, 'class' and 'label').

Outline plan

Please note that the slide deck labels the activities in the top right-hand corner to help you navigate the lesson.

**Timings are rough guides*

Starter activity (Slides 2–4) 5 mins	**What is classification?** As the students enter the classroom, show the question "What is classification?" on slide 2. This question should prompt the students to recall their prior learning from Lessons 1–3. They will be using decision trees to classify in this lesson, so it is important they remember what classification is. Ask them to discuss their answers with the person next to them. You can also encourage them to think a bit deeper and name some of the examples they have already seen, and the classes they used (the model classifying tomatoes and apples from Lesson 3 is a good example). When you are ready to begin the lesson, show slide 3 and run through the objectives with the class. Show slide 4 to remind the students of the key terminology of classification.
Introduction (Slides 5–6) 5 mins	**What does a model look like?** Display slide 5, and explain that machine learning models can take many forms and the students have not really seen what a model looks like yet. Outline the considerations for choosing a type of model for a project shown on the slide: • What type of data will you be working with? • What problem are you trying to solve? • Is there a need to explain the results? The slide introduces a new term — 'explainability'. Tell the students that it is a way of measuring how understandable a model's predictions are. Some types of model will be more explainable than others. Move on to slide 6 and play the video. In the video, some DeepMind staff members explain some different types of model and where they might be used, such as neural networks and transformers. The video ends with a discussion of decision trees, including where they are useful and how explainable they are.
Activity 1 (Slides 7–13)	**Decision trees**

15 mins	Display slide 7, which shows an example of a decision tree. Explain that decision trees are a form of model made with supervised learning, and can be used for classification. Decision trees are made up of 'nodes', and the top node is called the 'root'. Show slide 8, where the components of a decision tree are further broken down to help the students understand the flow of a decision tree. Explain that all nodes are either a 'decision node' or a 'leaf'. Each decision node contains a 'condition', and the result of that condition applied to the data tells you which route to follow from the decision node. Move on to slide 9. Explain that each route through a decision tree ends at a leaf. Leaf nodes usually represent a single class. When data is processed through a decision tree, the leaf you end on provides the predicted label for that data. Next, display slide 10 to show a real decision tree to the students. Explain that this decision tree can be used to classify animals as mammals, fish, reptiles, etc. Explain that the data used to create a decision tree is made up of **features**. In this case, the data used to create the decision tree is made up of Boolean (True/False) and numeric data. Move on to slide 11, which is the start of an example of how to use a decision tree. Ask the students to work in pairs to use the data and the decision tree to work out the label that would be predicted for the animal in this example. Give the students 3 minutes to do this, then collect some answers and ask the students to describe the path through the decision tree. Display slide 12 to highlight the correct path through the decision tree, as well as the relevant data in the table. Finally, show slide 13, which provides some important information about decision trees. Explain that they will only work with numeric or categorical data (like True/False), and not with complex data like images, audio, video, or long-form text. Remind the students about the explainability of decision trees. Now that they have used one, they should be able to see that it is straightforward to follow a decision tree and explain the conditions used to arrive at a prediction.
Activity 2 (Slides 14–33)	**How decision trees are made** In this activity, the students will see how the conditions in a decision tree are

15 mins	created using training data. Show slide 14 and explain that you and the students are going to create a decision tree to classify stars found by the new James Webb Space Telescope. This telescope takes long-range images of distant solar systems and nebulae containing thousands of stars. Highlight the fact on the slide about the amount of data the telescope collects every day, to help the students understand the scale of the problem and why a machine learning model might help. Move on to slide 15 and tell the students that stars have many different stages of their lives. Identifying what type of star we have found can help scientists understand our own sun by studying similar stars. The classes for this problem are stages of stars' life cycles: Red Dwarf, White Dwarf, Brown Dwarf, Main Sequence, Supergiant, and Hypergiant. Explain that these will be the classes and labels for the decision tree. Show slide 16 and explain the data. Remind the students that they cannot use the images directly (decision trees will not work with images), so instead they will use numeric and categorical data extracted from the images. The features of the stars will be temperature, radius, brightness, and colour. One important note is that both radius and brightness in this case are **relative** to our sun. This means that stars that are larger or brighter than our sun will have positive values for these features and stars that are smaller or less bright will have negative values. Next, display slide 17 and explain the process of creating a decision tree. As the tree is created, each decision node will 'split' the training data. Therefore, as you move down the tree, the number of examples being considered will get smaller. Show slide 18 and remind the students that the splits are made using conditions. Explain that they will need to decide which feature of the stars to use to split the data at each decision node. Display slide 19. This slide introduces 'star cards', which the students will use in this activity as the training data. Each card contains data about a single star. Move on to slide 20, and explain that for the first split, you are going to use temperature. Show slide 21 and explain that the split has been made, using the midpoint

	between the two temperature values. Display slide 22 to show a new star card, for the students to use to test the split. Ask them if the decision tree correctly groups together the data in the same class. It does not. Display slide 23 and ask the students what they would do to fix it. Explain that they can adjust the split, then display slide 24 to reveal the result. Explain to the students that during the process of training a decision tree, there will be thousands of little adjustments like this. Show slide 25 and ask the students to open the Part 1 worksheet, in which they will find six 'star cards' that will be their training data. The worksheet also includes the decision tree, with the nodes already determined filled in. Explain that the students can drag the cards around to sort them by their features and split the data. Direct the students to try making the last split. Tell them that the only rule is that they cannot use the same feature twice, so they cannot split the data by temperature. Give them around 2 minutes to complete the task. Once the students have finished, use slide 26 to show them the completed decision tree. Theirs might look slightly different if they chose a different feature, but that is fine. Use the next four slides (27–30) to guide the class through a test of the decision tree. There are two example sets of test data that they should run through the decision tree by tracing it to see if it classifies them correctly. The second star will not be classified correctly. Display slide 31 and tell the students that they are going to make another decision tree, as there are errors in the one that they have just made. Explain that the training data will have an impact on the decision tree that is created, which is what it means for a model to be **data-driven** rather than rule-based. Move on to slide 32 and instruct the students to open the Part 2 worksheet, which includes some different training data. Tell the students that they have 4 minutes to create a new decision tree based on the new training data. Their only clue is that the best feature for the first split this time is brightness (which is already filled in on the worksheet). Some students may not finish within 4 minutes, but it is not important that they do. Instead, the aim of this part of the activity is for the students to apply what they have just seen demonstrated on the slides. Display slide 33 to show the two decision trees side by side. This is to

	underline how different data can result in a different decision tree.
Activity 3 (Slides 34–39) 15 mins	**Using machine learning to create a decision tree** Show slide 34 and ask the students why machine learning might be helpful when creating a decision tree. Display slide 35 and explain some of the reasons ML is useful, with regard to scale and adaptability. Data sets are usually very large (much larger than the data set containing twelve stars used in the previous activity). ML also helps when adjusting the conditions based on the training data. Move on to slide 36 and tell the students that they are going to create a decision tree for the whole data set of stars using Machine Learning for Kids. Play the video on slide 37, which demonstrates how to complete the activity and shows the students the decision trees they will produce. Display slide 38 and ask the students to open the worksheet, which will guide them through the process of creating the star classifier decision tree using Machine Learning for Kids. The process is as follows: 1. Open machinelearningforkids.co.uk. 2. Click on **Get started**, then click on **Try it now** or log in to their account. 3. Go to **Projects** and click on **Copy template**. 4. Find the 'Star types' template and click on **Import**. 5. Tick the box that says "Do you want to use some of the data for testing?" 6. Choose a percentage (the worksheet recommends 10–30%). 7. Click on **IMPORT**. 8. Select the newly created 'Star types' project. 9. Click on **Train**. 10. Examine the data on the screen. 11. Click on **Back to project**, then click on **Learn & Test**. 12. Click on **Train new machine learning model**. Next, display slide 39 and ask the students to test their decision tree using the test data they set aside. The process is as follows: 1. Click on **Download test data** and open the spreadsheet that is downloaded. 2. Go back to Machine Learning for Kids and click on **Describe your model!**

	3. Choose three sets of star data from the spreadsheet. 4. Enter one set of test data into the form provided and click on **Test**. 5. Repeat step 4 for two more stars. When the students have done this, ask them to look at the decision tree that the person next to them has made, to compare them and see again how different data results in a different decision tree.
Plenary (Slides 40–41) 5 mins	**Decision trees in medicine** To conclude the lesson, the students are going to reflect on what they have learned about decision trees and apply it to a particular context — medicine. Show slide 40 and ask the students to discuss with the person next to them why decision trees might be used in a medical context. You can remind them of the explainability of decision trees to help guide their discussion. Show slide 41 to summarise that decision trees would work well in medicine due to both their explainability and the type of data that doctors collect.