Raspberry Pi Foundation

Lesson 5: Solving problems with ML models

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

Introduction

In this lesson, students will be introduced to the AI project lifecycle and use it to create a machine learning model to solve a problem of their choice.

First, students will order the stages of the AI project lifecycle. They will then be introduced to the idea of needing to take a user-focused approach when working on AI projects. Students will be presented with a choice of projects to select from, then they will be asked to train a machine learning model and test it to determine its accuracy.

Learning objectives

- Describe the stages of the AI project lifecycle
- Use a machine learning tool to import data and train a model
- Test and examine the accuracy of a machine learning model

Key vocabulary

AI project lifecycle, data cleaning, machine learning model, class, label, training, testing, accuracy, confidence score, confidence threshold

Preparation

You will need:

- Slides
- Starter activity ‘The AI project lifecycle’ activity resource
- Project folders:
    - Classifying ocean data
    - Fake news

Starter activity: In this activity, students are asked to order the stages of the AI project lifecycle. The ‘AI project lifecycle’ activity resource includes cards showing the stages and descriptions,

which can be cut out and put on desks, if your classroom allows for this. Alternatively, you can refer to slide 2 in the slide deck, which shows the stages.

Subject knowledge:

Activity 1:

Data cleaning:
In this activity, students are introduced to **data cleaning**. When data is collected, work needs to be done to make sure it is suitable to use for training a machine learning model. If the data is not 'clean', it is likely that the model will produce less accurate results. Cleaning involves taking steps such as removing duplicate or irrelevant data, checking for outliers, and deciding what to do about missing values (such as removing or replacing the data).

Creating classes:
During the planning phase, students are required to think of the classes to use in their project. Students should aim to decide on the classes independently. The classes should be as follows:

- **Classifying ocean data:** subtropical, tropical, and polar
- **Identifying fake news:** Real and Fake

Activity 2:
In this activity, students will use Machine Learning for Kids to train a model. You will need to be familiar with how to set up and train the models using this tool. A support video has been provided, which you can use as a guide before demonstrating the process to the class.

Assessment opportunities

During Activities 1, 2, and 3, the students will be completing a project documentation worksheet. The students will use this worksheet in both this lesson and the next lesson.

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 2–3) 5 mins	**Order the stages of the AI project lifecycle** As the students enter the classroom, display slide 2. Ask the students to read the stages of the AI project lifecycle and put them in the order that they think would be logical. This can be done individually or in

	small groups. Although the students have not been introduced to the AI project lifecycle, they have had experience of a few of the stages in the previous lessons and therefore should be able to attempt to determine the order. Move on to slide 3 to reveal the correct order. Click to build slide, explain to students that they will be exploring the first four life cycle stages in this lesson and the last two stages in the next lesson.
Introduction (Slides 4–7) 5 mins	**User-focused approach** Display slide 4 to show the learning objectives to the students, and explain that in this lesson, they will follow the stages of the AI project lifecycle to create their own machine learning model. Move on to slide 5 to explain a 'user-focused approach'. Then, click on the slide to display the following question: "Can you suggest one way in which you could try to avoid the use of an AI application causing harm?" There are various answers students could give to this question. Accept any answers that show students are thinking about the users who could be harmed. Example answers include the following: • Making sure a wide range of people with different backgrounds, experiences, and opinions are involved in the design process • Making sure that experts in the field are consulted • Identifying the risks and planning how each one can be addressed • Not producing the application if any potential harm to society is identified **United Nations Sustainable Development Goals** Move on to slide 6 to introduce the United Nations Sustainable Development Goals. Explain that the 17 goals adopted by the UN General Assembly in 2015 act as a framework for developing a sustainable and equitable future for humanity. Move on to slide 7 and ask the students to look at the goals in more detail.
Activity 1 (Slides 8–16) 20 mins	**Stage 1: Defining the problem** Use slide 8 to introduce the project that the students will be working on for the remainder of this lesson and the next lesson. Move on to slide 9 to introduce the two possible projects the students can choose to work on.

	Briefly describe each project, and allow the students 5 minutes to open the project folders and read the project briefs and explore the data sets. By the end of the 5 minutes, the students should have chosen a project to work on. Move on to slide 10 and use the video to describe the stages of the AI project lifecycle. Move on to slide 11 to describe the first stage. Part of this stage has already been carried out for the students, and their task is to go to the project documentation worksheet for their project and complete this stage. The worksheet requires them to carry out the following tasks: 1. Identify between one and four UN Sustainable Development Goals that they think the project will support, and justify their answers. 2. Describe why creating a machine learning classification model is suitable for this project (their answers should focus on why a data-driven approach is more suitable than a rule-based approach). Allow no more than 10 minutes for the students to complete this activity. **Stage 2: Preparing the data** Move on to slide 12 to describe the data preparation stage. Describe the process of cleaning data and allow the students 30 seconds to see if they can identify any problems with the data on the slide. Use slides 13 to 15 to highlight the issues that need to be addressed before the data is ready to be used to train the model. Move on to slide 16 and tell the students that they need to decide on the classes for their machine learning model. Instruct the students to return to their worksheet, and allow 2 to 3 minutes for them to write down their classes.
Activity 2 (Slides 17–18) 5 mins	**Stage 3: Training the model** Use slide 17 to describe the students' next task, which is to use Machine Learning for Kids to train their model. Explain that each of their projects are different, and they will need to consider what type of data they will use (text or numbers), as well as what proportion of the data they will set aside for testing. Display slide 18 and demonstrate how to complete the following tasks using Machine Learning for Kids:

	• Copy a project from a template • Select how much training data to use • Train a model Allow the students time to complete this process themselves.
Activity 3 (Slides 19–25) 15 mins	**Stage 4: Testing the model** Use slide 19 to briefly describe how the students will now test their model. Move on to slide 20 to describe how they can measure the accuracy of their model during testing. Use the animation on the slide to talk through a worked example of how to calculate accuracy. Move straight on to slide 21 to remind the students about confidence scores. Explain how a threshold confidence score (a 'confidence threshold') should be decided on to determine whether or not the model has predicted the correct label with an acceptable level of confidence. Click through slides 22 and 23 and allow the students to briefly think about what would be acceptable confidence thresholds for the scenarios on the slides. This is to demonstrate that the thresholds might differ between applications. The model used in an application to help predict the weather might not require an extremely high level of confidence. Depending on who might need to know the conditions, 80% confidence might be acceptable. The model used by a driverless car to identify pedestrians might require a much lower confidence threshold. For example, you could ask the students to consider a human driver driving at night, and whether they would want them to be 100% confident that they could see a pedestrian before slowing down, or whether they would hope they slowed down even if they were uncertain. If there is not enough time, you do not need to hold a class discussion. Move on to slide 24 to explain that the students now need to see accuracy as the proportion of outputs where the label is correct and the confidence threshold is met. Display slide 25. Ask the students to return to their worksheet and complete Stage 4 using Machine Learning for Kids. They should: • Decide on a confidence threshold

	• Download their test data • Test their model at least ten times, noting down whether or not it correctly labelled the data and the confidence score met the threshold • Calculate the accuracy of the model • Reflect on the results **Explorer task:** If any students complete their testing ahead of their classmates, instruct them to try retraining the model to see if they can improve its accuracy, and to try using a different percentage of training/test data.
Plenary (Slides 26–27) 3–5 mins	**Reporting on the accuracy of a model** Use slide 26 to present a scenario in which two students have created models, and highlight the accuracy of those models. Ask the students which model they would use. It might seem as though it is an easy question to answer, as one student is claiming their model is more accurate than the other. Ask the students to think about what information they might also need to help make their decision. Answers could include the following: • What data was used? • Was it the same for both models? • **What was the confidence threshold used to measure the accuracy of their model?** Move on to slide 27, which reveals the confidence thresholds used. You can explain that in this scenario, both students used the same data set, but significantly, the students picked different confidence thresholds, which has had an impact on the reported accuracy of their models. If time allows, ask the students: • Is this fair? • Has it changed their opinion about which model to use? • Should people who make machine learning models show people both the accuracy of the model and the confidence threshold they had decided on? This activity recaps the learning from this lesson, but also acts as a link to the next lesson, in which the students will be asked to evaluate and explain their model.
