Raspberry Pi Foundation

Lesson 3: Bias in, bias out

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

Introduction

In this lesson, students will have an opportunity to create their own machine learning model. The model will classify images of apples and tomatoes, but students will discover that their model is flawed due to the limited data set they will use to train their models. Next, students will explore how bias can appear in sets of data used to train models, which in turn make the models produce biased predictions.

Learning objectives

- Describe the impact of data on the accuracy of a machine learning (ML) model
- Explain the need for both training and test data
- Explain how bias can influence the predictions generated by an ML model

Key vocabulary

Artificial intelligence (AI), machine learning (ML), supervised learning, classification, training data, test data, accuracy, bias, data bias, societal bias

Preparation

You will need:

- Slides
- Activity 1 worksheet and teacher support video
- Activity 3 worksheet

Activity 3 slide preparation:
Slide 23 of the slide deck requires you to add the names of some of **your students** to the subjects that a fictional machine learning model has predicted will be the subjects they will study. This slide is deliberately provocative and designed to prompt debate about the accuracy of a model. It should prompt your students to think about how the model produced the predictions it did (what data was used to train the model).

Subject knowledge:

Starter activity:
In the starter activity, students are asked to name the three types of machine learning (supervised learning, unsupervised learning, and reinforcement learning), therefore it is recommended that you ensure that you are familiar with these terms and how they were addressed in Lesson 2 of this unit.

Introduction:
As part of the introduction, students are asked to think back to Lessons 1 and 2 and describe what is meant by a 'model' in relation to machine learning (slide 6). The answer is displayed on the next slide. Students are also asked to identify that a model is a form of supervised learning.

Activity 1:

Training data and test data:
In this activity, students are introduced to the idea that when training a machine learning model with data, it is important that the data set (introduced as "set of data") is separated into 'training data' and 'test data'. The model is trained using the training data, then the accuracy of the model is tested using the test data that was not used as part of the training process. Once the model has been trained and tested, it is ready to be used with 'unseen data'. This is when an application uses the machine learning model with new data that the model has not seen before. In the example of sentiment analysis from Lesson 2, where a model could be created to classify whether a product review was positive or negative, the model would be trained and tested using reviews that have already been written and labelled as positive or negative. 'Unseen data' would refer to new reviews that the model can be used to label as positive or negative.

Training a model (using Machine Learning for Kids):
Students will use Machine Learning for Kids as a tool to create a machine learning model. Following step-by-step instructions in their worksheets, they will create two classes and label training data before training and testing their model. Before you deliver the lesson, it is recommended that you watch the teacher support video provided, so that you are familiar with the steps involved.

Activity 2:

Bias:
The focus of this activity is the concept of bias and how it appears in machine learning models. The slides are scaffolded to describe data bias and societal bias. Encourage discussion from the students around the examples. The final slide in the activity states that "Societal bias can

lead to data bias". The example on the slide — a facial recognition system that is less accurate in recognising people with certain skin tones — is an example of data bias that can be caused by societal bias: this data bias can occur when not enough data from a range of people with different skin tones is used to train the model, which can happen due to societal bias.

Assessment opportunities

You can use the starter activity and introduction to assess how much the students have remembered of the key concepts from Lessons 1 and 2. These are the different types of machine learning, as well as models and classification.

Activity 2 provides opportunities for discussions with the students where you can assess their understanding of the key differences between data bias and societal bias.

Activity 3 is a written activity where you can assess how the students have applied their understanding of bias to a school timetabling scenario.

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) 2 mins	**The three different types of machine learning** As the students enter the classroom, display slide 2 with the title "What are the three types of machine learning?" As the students are arriving and settling into the class, ask them to discuss their thoughts with the person sitting next to them. Display slide 3 to show the answers.
Introduction (Slides 5–9) 5 mins	**Supermarket AI application** Use slide 5 to introduce the scenario for this activity: a supermarket would like to use the cameras around its stores to recognise the items customers have placed in their baskets. Stress the word "cameras" to the students, as it is important that they notice how the data will be collected. Move on to slide 6 and highlight to the students that they will be required to create a machine learning model. Ask the students, "What is a model?" This is to remind the students of the concepts covered in Lessons 1 and 2.

	Display slide 7 and tell the students that they are going to create their own ML model to recognise apples and tomatoes. Ask the students what type of data they think they will need (note that in this scenario, the data will be collected via cameras). If students answer “data”, ask them specifically what type. The answer is that they will need images of apples and tomatoes — and lots of them. Move on to slide 8. Highlight that image data will be needed to train the model in this scenario. As this is training data, ask the students to identify which of the three types of machine learning they will be using. Move on to slide 9 to reveal that the answer is “supervised learning”.
Activity 1 (Slides 10–13) 20 mins	**Training a model** Use the animation on slide 10 to describe the use of data with a supervised learning model (see the ‘Subject knowledge’ section above for more detail). Move on to slide 11 and ask the students to open their worksheet and follow the instructions to train their model. The students will use Machine Learning for Kids (machinelearningforkids.co.uk) for this activity. They will need to complete the following tasks: • Create a project that recognises images • Create a label (class) for apples and one for tomatoes • Add training data (5 images of apples and 5 images of tomatoes) from the webpage linked on the worksheet (ai-activities.raspberrypi.org/project-files) — encourage the students to think about which images they want to use to train the model • Train the model **Note:** You can watch the teacher support video so that you are familiar with the steps involved before you deliver the lesson. Allow 5 minutes for the students to complete the training process. **Testing the model using the test data** Move on to slide 12. Explain that the students now need to test their model using the test data provided. Allow 3 minutes for testing. Encourage the students to test their model with as many images from the test data provided as possible. The students should find that their model is not as accurate as it needs to be.

	They can determine this by seeing what label the model has applied to the test data, and looking at the confidence score. Move on to slide 13 and ask the students the questions on the slide to encourage them to think about why the model was not accurate enough for the supermarket. This is a 'think, pair, share' activity, so allow approximately 1 minute for the students to think about this independently, before turning to the person next to them to share their thoughts. Once they have had their short discussion, take answers from around the classroom. The answers that the students give should be around the two main issues: 1. Not enough data was used to train the model. 2. The 'right' data was not used: a. There were no images of red apples in the training data. b. The images were not very representative of how the apples and tomatoes might look if they were on a supermarket shelf. For example, the training data included images of apples hanging from trees.
Activity 2 (Slides 14–21) 15 mins	**Bias** The focus of this activity is to introduce students to how machine learning models can make biased predictions. The activity includes opportunities for discussion. Try to keep to time to allow enough time to complete Activity 3 and the plenary activity. Play the video on slide 14 that introduces 'bias'. Move on to slide 15, which starts with a definition of 'bias' in a machine learning context. Read the two examples of machine learning bias on the slide, but at this stage, do not discuss them or how they could have happened. Move on to slide 16 and describe **data bias**. Ask the students to recall how data bias appeared in the model they made for the supermarket. Display slide 17 to show questions to consider to avoid data bias. Move on to slide 18 and describe **societal bias**. Read the example on the slide and ask the students to think about how societal bias could appear in the data. Take suggestions from the class. There are many possible answers to this question, for instance: • Jobs that exist now might not be in the data set • The data set is unlikely to reflect the diversity of the modern workforce in this country

	Move on to slide 19 to emphasise the societal bias that might be in the data set. Ask the students which gender might have been associated with each job role in 1960. Ask whether or not this societal bias is still in existence now. If there is time, you could highlight this point further by demonstrating an image search for the jobs on the slide. What if those images were used as part of the training data? What impact would this have on the predictions made by a machine learning model? Use slide 20 and 21 to assess the students' understanding of the difference between data bias and societal bias. Click through the slides to revisit the examples given at the start of this activity and ask students to vote on whether each one is an example of societal bias or data bias. Use the slide animations to reveal the answers. It might be that some students think that the example on slide 21 — a facial recognition system that is less accurate in recognising people with certain skin tones — is an example of societal bias. However, it is an example of data bias that can be caused by societal bias: the model has likely not been trained with enough data from a range of people with different skin tones, which can happen due to societal bias.
Activity 3 (Slides 22–24) 10 mins	**Student timetable model** Use slide 22 to introduce the next scenario to the students: your school is thinking of investing in timetabling software that uses an AI application. The AI application is used to predict which subjects the students will study. Move on to slide 23, where you will reveal what the model has predicted will be the most appropriate subject choices for some of **your students** (see the 'Preparation' section above). You can point out that this is an entirely fictional outcome, but hopefully this will cause lighthearted discontent among some of your students. Ask the students how the model could have made those predictions. Move on to slide 24 and ask the students to open the worksheet and complete the first task. Allow 5 minutes for the students to evaluate the proposed sets of data and write down which ones they think are suitable to be used in training the model, and which ones they think are unsuitable and why. **Explorer task:** If time allows, students can move on to the explorer task on their worksheet. They should look at the fields of data included in the data sets, mark which fields they would remove to try to reduce the bias, and explain how using all the data could result in a biased outcome.

Plenary (Slides 25–28) 5 mins	**Reducing bias** Use slide 25 to introduce a final scenario to the students: the school also wants to use an AI application to create a weekly menu for the school cafeteria. The aim of this plenary activity is to repack the learning objectives for this lesson. Ask the students what considerations they should make about collecting the data and training the model. Take answers from a number of students, then reveal the answers on slide 26.