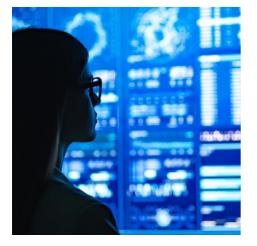# An introduction to data science and machine learning

This resource was developed by teachers within the Royal Society Schools Network

iStock image: credit Artem Peretiatko

Curriculum key words

Data
Machine learning

Curriculum links

Computing
- use logical reasoning to predict the behaviour of simple programs
- recognise common uses of information technology beyond school

Science: I can gather and record data including information from other sources to help in answering questions.

Equipment needed

– sheet of paper to record data around a person or book character
– optional: internet access to explore an infographic

Resources to use

– Example of Google map
– Example of a bar chart for CBeebies

KS1

Introduction

The aims of the National Curriculum for computing state that:

"A high-quality computing education equips students to use computational thinking and creativity to understand and change the world. Computing has deep links with mathematics, science, and design and technology, and provides insights into both natural and artificial systems. The core of computing is computer science, in which students are taught the principles of information and computation, how digital systems work, and how to put this knowledge to use through programming."

In today's world it is impossible to ignore the presence of data, data science and machine learning. As more and more data is collected about us and is used by data scientists, it is vital that we educate our students on how data is collected, how it is used and how it impacts our daily lives both now and in the future.

The purpose of this lesson is to introduce students to the concept of data, data science and machine learning. It could take place over several lessons and could link with data handling in maths.

Learning objective:

To understand what data is and how it can be used to solve problems.

Success criteria (SC)

- SC1: I can give some examples of data

- SC2: I can interpret a bar chart or pictogram

- SC3: I can compare computer abilities and learning with human intelligence and learning.

An introduction to data science and machine learning

Starter activity: What is data?

(Approximately 10 – 20 mins) [SC1]

Ask the students to write down or tell an adult as many facts as they can either about a character from a book that the class is reading or about someone they all know well. Add these around a central image of the character.

Explain that each of these facts is a piece of data about the person.

Explain that data is collected about all sorts of things in lots of different ways and all the time.

Location icon

Give some examples about how data is collected e.g. browsing a website providing data about what websites we like to visit, use of a loyalty card collects data about what we buy, mobile phones and GPS data giving data about location – show them the location icon and ask if they recognise this.

Open Google Maps, select a local area to you where you know there is traffic and find a busy journey. Show the students how the route changes colour to show where there is really busy traffic.
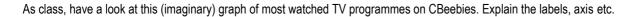

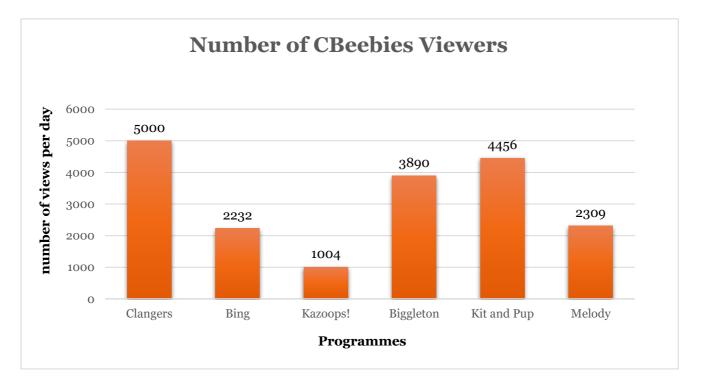Example Google map screenshot: showing where busy traffic is located in a part of London.

Ask the students how Google Maps knows when a route is busy (Google uses anonymized data from mobile phones, calculates the speed users are traveling along a length of road and then generates traffic data accordingly).

This shows us how computer scientists use data to help us find out about something.

Activity A: let's be data scientists

(Approximately 10 – 15 mins) [SC2]

As class, have a look at this (imaginary) graph of most watched TV programmes on CBeebies. Explain the labels, axis etc.

Number of CBeebies Viewers

A bar chart titled "Number of CBeebies Viewers" with y-axis "number of views per day" (0 to 6000) and x-axis "Programmes". The bars show: Clangers 5000, Bing 2232, Kazoops! 1004, Biggleton 3890, Kit and Pup 4456, Melody 2309.

Tell the class to imagine that they are a team of producers who have to decide on three programmes to air at the time most students watch TV (between 4pm and 7pm).

Ask the students which programmes would they choose to make sure that students will be able to watch the shows they enjoy and why?

The students should come up with an answer to convince the rest of the class/group about their choice of programme by explaining how they have used the graph to come to a conclusion.

Activity B: do you think a computer is clever?

(Approximately 5 – 10 mins) [SC3]

Ask the students if they think computers are clever.

- Why do they think this?

- What things can a human do that a computer can't do?

A computer follows instructions called algorithms. Computers can now follow algorithms which help them to learn! It is called machine learning and computers are said to have artificial intelligence.

Activity C: Royal Society 'What is machine learning?' infographic

(Approximately 20 - 30 mins, or as long as you have time for)

Find out how computers learn using The Royal Society's What is machine learning infographic (https://royalsociety.org/topics-policy/projects/machine-learning/what-is-machine-learning-infographic/). The app takes a few seconds to load, this may be longer on slower internet connections.

Notes for teachers

Much of the infographic may be too complex for KS1 students but some of the ideas could be summarized for them.

There is a photo challenge activity which explains how Google trained a network of computers to learn the difference between cats and dogs by feeding it with millions of YouTube clips. The system started to identify features of cats and dogs and grouped them into the same thing. It then used those groups of images to identify other images of cats and dogs. By the end of three days it was able to identify cat and dog images with high levels of accuracy even when the photos were unclear.

Photo challenge activity (optional)

Show the students random images of cats and dogs which they have to identify as such as quickly as they are able. Students can use the activity on the infographic (half way down the page). At the end, the teacher can show on the infographic how quickly they can manage it. It is then possible to click on a play arrow which shows an animation of a computer performing the same task – in less than a millisecond!

The students can repeat the task numerous times trying to beat their score.

Explain to the students: The computer has learnt about the different features of a dog and cat and can use what it has learnt, just like us, to determine whether each image shows a dog or a cat but can do this much more quickly than we can. BUT it doesn't really know what a cat or a dog is in the same way as we do. This shows us that computers can perform specific tasks but aren't 'clever' like human beings. Instead they have 'artificial intelligence' by learning about something using lots of bits of data. This is called machine learning.

Activity D: Royal Society 'Machine learning in the world around you' infographic

(Approximately 20 - 30 mins, or as long as you have time for)

Exploring machine learning and artificial intelligence in the real world via the Machine learning and the world around you infographic (https://royalsociety.org/topics-policy/projects/machine-learning/machine-learning-in-the-world-around-you-infographic/).

This infographic describes different ways in which machine learning is, or will be used in the future, in our everyday lives. Many of the examples will be too difficult for a KS1 student to understand but ask the students if they have any ideas for how machine learning could be

used in each of the areas. Summarise at a level suitable for the students some of the ways in which machine learning is used (shopping – making suggestions for clothes you might like, banking – voice recognition, shopping – detecting purchases and taking money from an account without having to queue at a check out). The Farm activity should be suitable for KS1 students.

Farm activity (optional)

When you open up the Farm activity it explains that: 'programs – such as LettuceBot - can use machine learning to photograph 5,000 young plants a minute and decide which are crops and which are weeds. Robotic agricultural machinery then removes the weeds'.

The students are then challenged to compete with the program. There are six images of similar looking lettuces and weeds for the students to learn which is which. They are then shown an example and have to decide whether or not to hoe it (they may need hoe explaining to them first) to get rid of the weed or leave it (if it is a lettuce).

At the end it tells you how many images you have processed in 20 seconds and how many Lettuce Bot would have processed (many more!).

Discuss with the students what would be the advantages of this and what would be the disadvantages.

Some thoughts:

- Advantages, quicker, more weeds removed, cheaper in the long run so perhaps food would become cheaper

- Disadvantages: people will lose their jobs; the machinery will be very expensive to begin with, over farming??