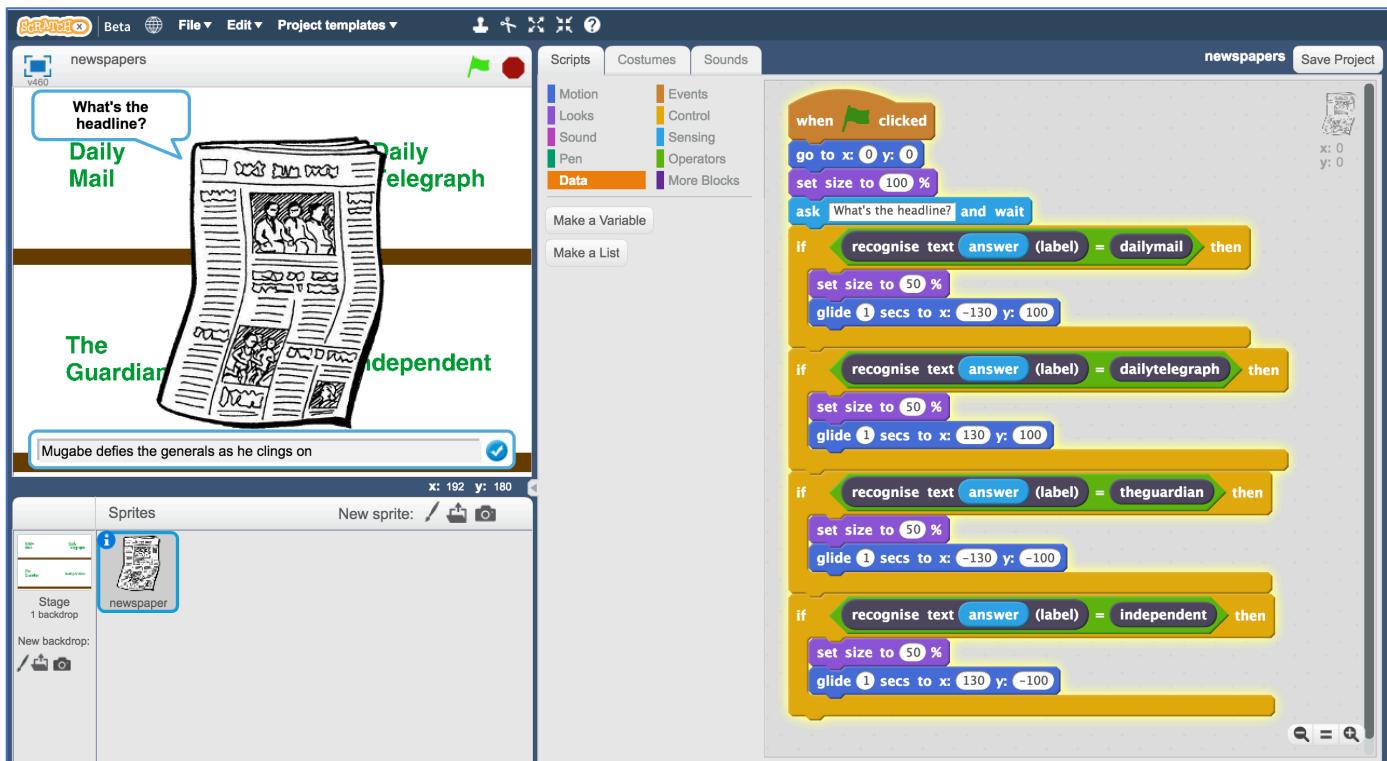


Headlines

In this project, you will train a computer to recognise headlines from different newspapers.

You will collect examples of the front-page headlines from different newspapers. You'll use those to train a machine learning model that can learn from the use of language to be able to recognise which newspaper a headline is from.



This project worksheet is licensed under a Creative Commons Attribution Non-Commercial Share-Alike License
<http://creativecommons.org/licenses/by-nc-sa/4.0/>

1. Go to <https://machinelearningforkids.co.uk/> in a web browser
2. Click on “**Get started**”
3. Click on “**Log In**” and type in your username and password
If you don't have a username, ask your teacher or group leader.
4. Click on “**Projects**” on the top menu bar
5. Click the “**+ Add a new project**” button.
6. Name your project “newspapers” and set it to learn how to recognise “text”. Click “**Create**”

Start a new machine learning project

Project Name *

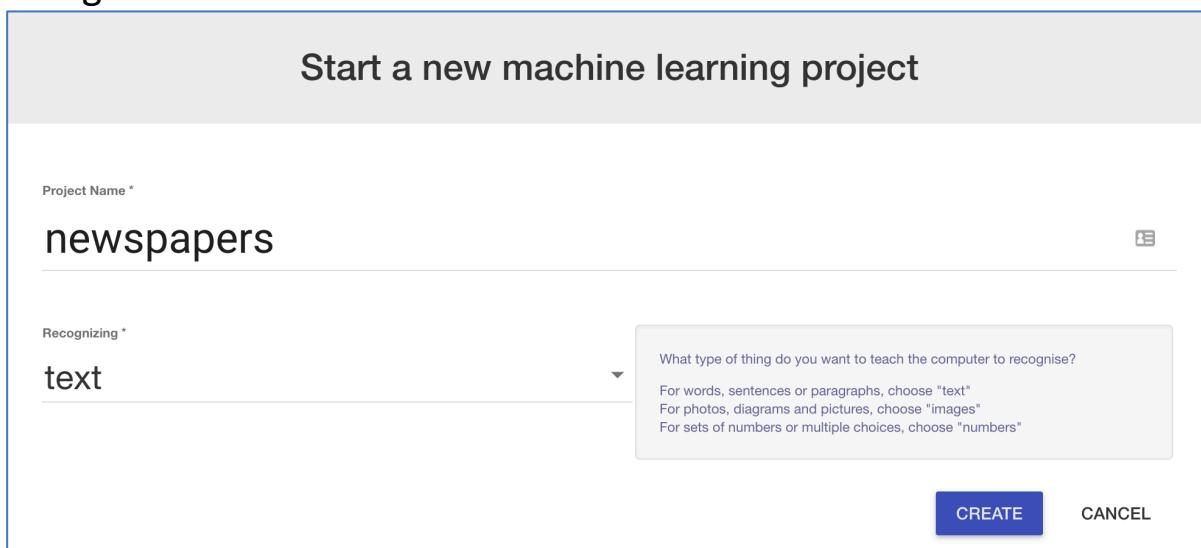
newspapers

Recognizing *

text

What type of thing do you want to teach the computer to recognise?
For words, sentences or paragraphs, choose "text"
For photos, diagrams and pictures, choose "images"
For sets of numbers or multiple choices, choose "numbers"

CREATE **CANCEL**



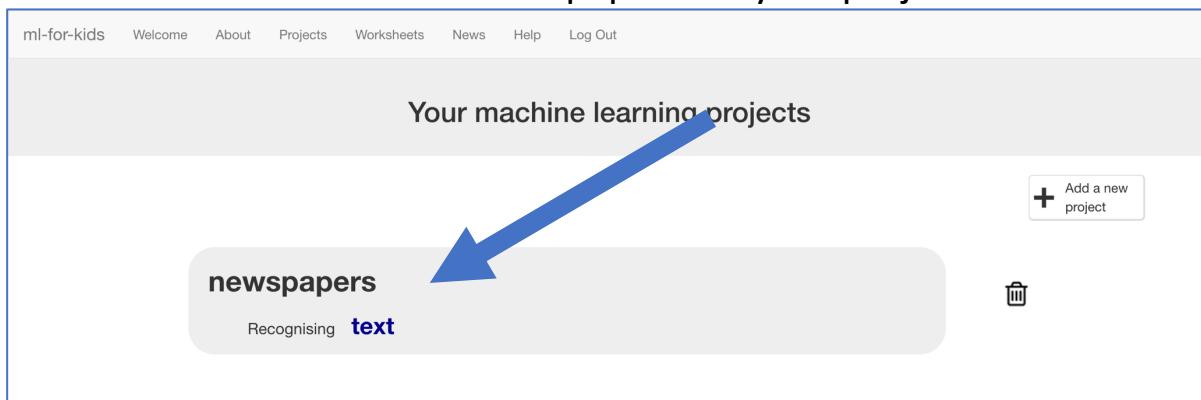
7. You should now see “newspapers” in your projects list. Click on it.

ml-for-kids Welcome About Projects Worksheets News Help Log Out

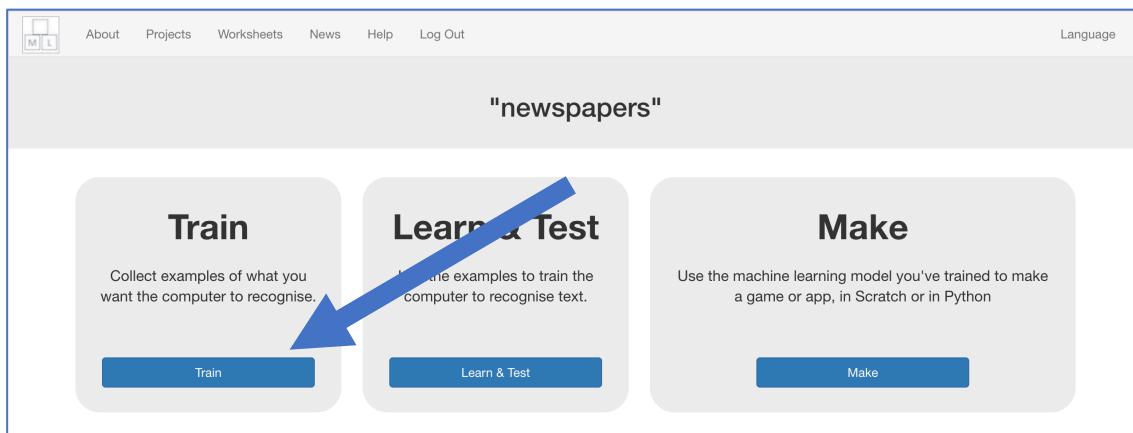
Your machine learning projects

newspapers  Recognising **text**

+ Add a new project 



8. Click the “Train” button



9. Choose four national newspapers.

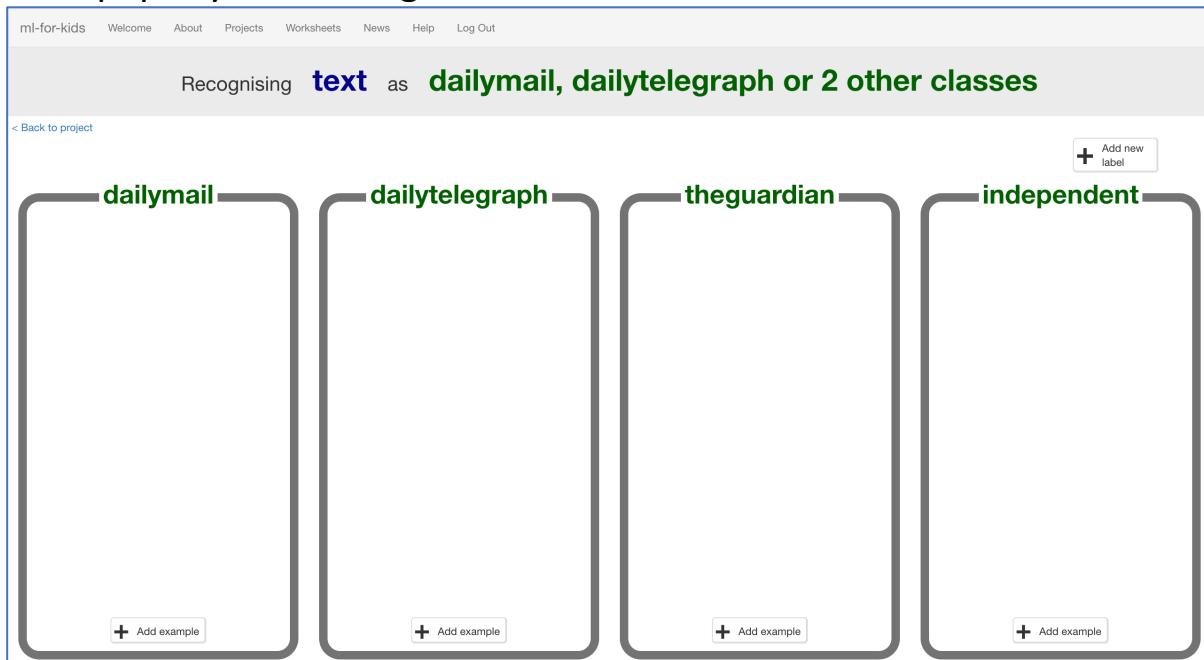
To save having to buy actual physical newspapers, you can use an online collection of newspaper front pages.

For the rest of this worksheet, I'll be using:

Daily Mail, Daily Telegraph, The Guardian, The Independent and getting the headlines from:

<https://www.thepaperboy.com/uk/front-pages.cfm>

10. Use the “+ Add new label” button to create a bucket for each newspaper you’re using.



11. In another web browser window, find newspaper front pages.
Arrange the windows so your training buckets are next to the book site.

The screenshot shows a web application for training a machine learning model. On the left, there are four rectangular "buckets" for collecting examples, each labeled with a newspaper name: "dailymail", "dailytelegraph", "theguardian", and "independent". Each bucket has a small "+ Add example" button at the bottom right. On the right side of the screen, there are three separate sections, each displaying a different newspaper's front page from August 14, 2017. Each section includes a link to "Click Here to Switch to Front Pages View" and a list of links to the newspaper's website, digital edition, and Twitter feed.

12. Agree on some rules for the examples of headlines that you'll collect, so that the training is consistent.

For the rest of this worksheet, I'll be using:

- * Newspapers from weekdays only (Mon-Fri) not weekend editions
- * Only the biggest main headline from each newspaper

13. Click the “+ Add example” button in one of the buckets and copy the newspaper headline into the “Add new example” dialog.

This screenshot shows the same web application as the previous one, but with the "dailymail" bucket highlighted and a modal dialog box overlaid on it. The dialog box is titled "Add new example" and contains a text input field with the headline "WAR ON HEART DEATHS". Below the input field are two buttons: "ADD" and "CANCEL". The rest of the interface, including the other buckets and the newspaper front pages on the right, remains the same.

14. Repeat this for all four newspapers on this day

Recognising **text** as **dailymail, dailytelegraph or 2 other classes**

< Back to project

dailymail WAR ON HEART DEATHS + Add example

dailytelegraph Rees-Mogg: Cut stamp duty now + Add example

theguardian Trump under fire from all sides as... + Add example

independent May's plan to pack Brexit commit... + Add example

The Daily Telegraph Total Football podcast

8/14/2017: FRONT PAGE: Trump under fire from all sides as woman dies rally

Politicians from all sides have rounded on Donald Trump explicitly to condemn white supremacy groups or domestic terrorism after a woman was killed when a car antiracism protesters at the weekend, as the US...

Visit The Daily Telegraph Web Site
Read the Digital Edition of The Daily Telegraph at PressDisplay
View The Daily Telegraph Twitter Feed

the guardian

Trump under fire from all sides as... + Add example

independent May's plan to pack Brexit commit... + Add example

THE SCOTSMAN

8/14/2017: Front Page: Charlottesville's innocent victim

Heather Heyer, 32, was killed when a car driven by a Nazi drove at peaceful demonstrators

Visit The Independent Web Site
Read the Digital Edition of The Independent at PressDisplay
View The Independent Twitter Feed

8/14/2017: Front Page: Scottish runners lead the way as Team GB lands

15. Repeat this for five days of newspaper headlines.

If you're using front pages from thepaperboy.com you can use the calendar at the bottom of the page to switch to a different day.
Remember the rules: Weekdays only. Biggest headline you see.
Tip: Use the 'Front Page Headline View' to get headlines as text.

16. You should now have examples from one week of newspapers Keep the newspaper site window open – you'll need it again later.

ml-for-kids Welcome About Projects Worksheets News Help Log Out

Recognising **text** as **dailymail, dailytelegraph or 2 other classes**

< Back to project

dailymail BREASTFEEDING LESSONS AT 11 + Add example

HARRY: NO ROYAL WANTS THR... + Add example

SHAME OF INSURANCE GIANTS + Add example

MASSACRE OF HOLIDAY FAMIL... + Add example

LET'S REIGNITE BRITISH SPIRIT + Add example

dailytelegraph BBC's gender pay gap revealed + Add example

Draw up your will in a text message + Add example

Set young free online 'to save the... + Add example

France's new hope puts cloud ov... + Add example

'Your country needs you' + Add example

theguardian 'Racism is evil', says Trump, 48 h... + Add example

May intensifies personal attack a... + Add example

Ban from 2040 on diesel and petr... + Add example

Backbenchers tell May: sack the ... + Add example

Labour threat to defeat May on B... + Add example

independent May abandons climate change pr... + Add example

Government demands tariff-free t... + Add example

NHS faces £1bn Brexit bill for tre... + Add example

Male fertility crisis is 'wake-up ca... + Add example

Brexit IT 'horror' puts £34bn bord... + Add example

17. Click the “< Back to project” link. Then click “Learn & Train”

18. Click “Train new machine learning model”.

As long as you've collected enough examples, the computer should start to learn how to recognise headlines from the examples you've given to it.

19. Wait for the training to complete. This might take a few minutes.

20. Try the model out by putting a headline into the **Test** box

The screenshot shows a web-based machine learning interface. At the top, there's a message about the model's creation date and the types of news sources it has learned from. Below this, a text input field contains the headline "Cuts 'threaten disaster' for policing terror". A blue arrow points from the left towards this input field. To the right of the input field, there's a "Test" button. Below the input field, the text "Recognised as independent with 39% confidence" is displayed. The entire interface is contained within a light gray box with a blue border.

21. Click the “< Back to project” link

What have you done so far?

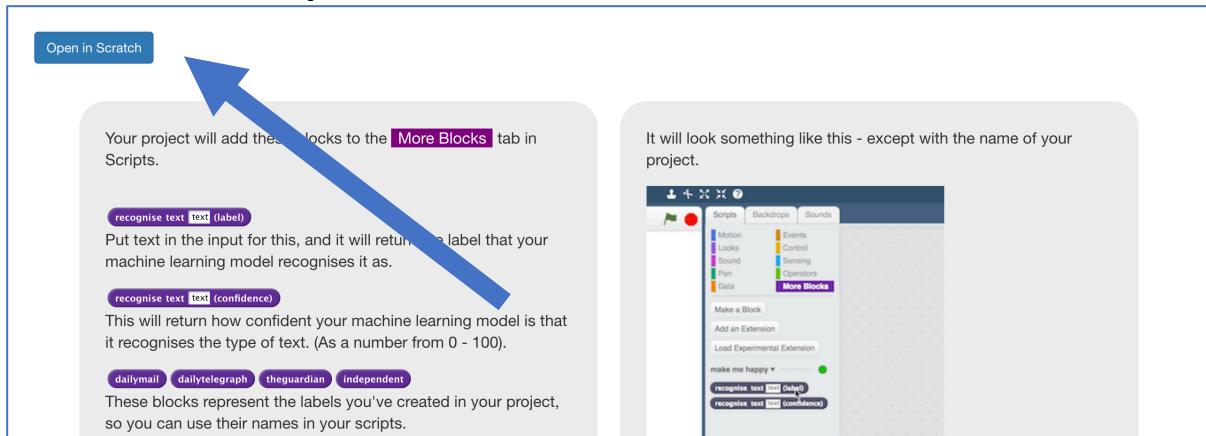
You've started to train a machine learning model to recognise patterns in the way that different newspapers use words, phrasing and punctuation in their front-page headlines.

But is it working well? How much has it learned so far? Does it need more examples to be able to recognise headlines correctly?

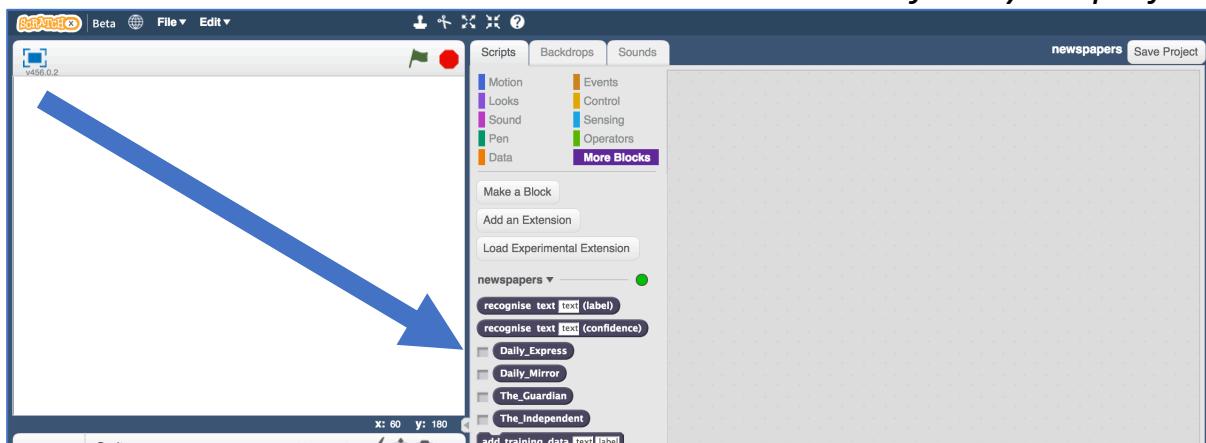
What is it good at? What is it bad at? Which newspapers does it get confused between?

22. Click the “Make” button, then click the “Scratch” button

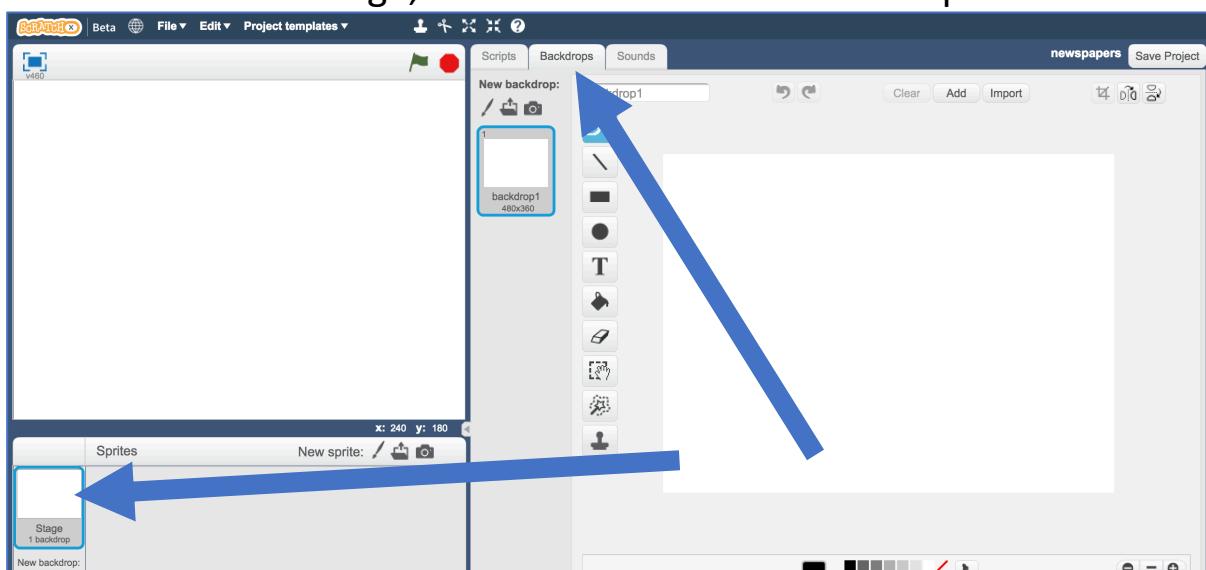
23. Click the “Open in Scratch” button



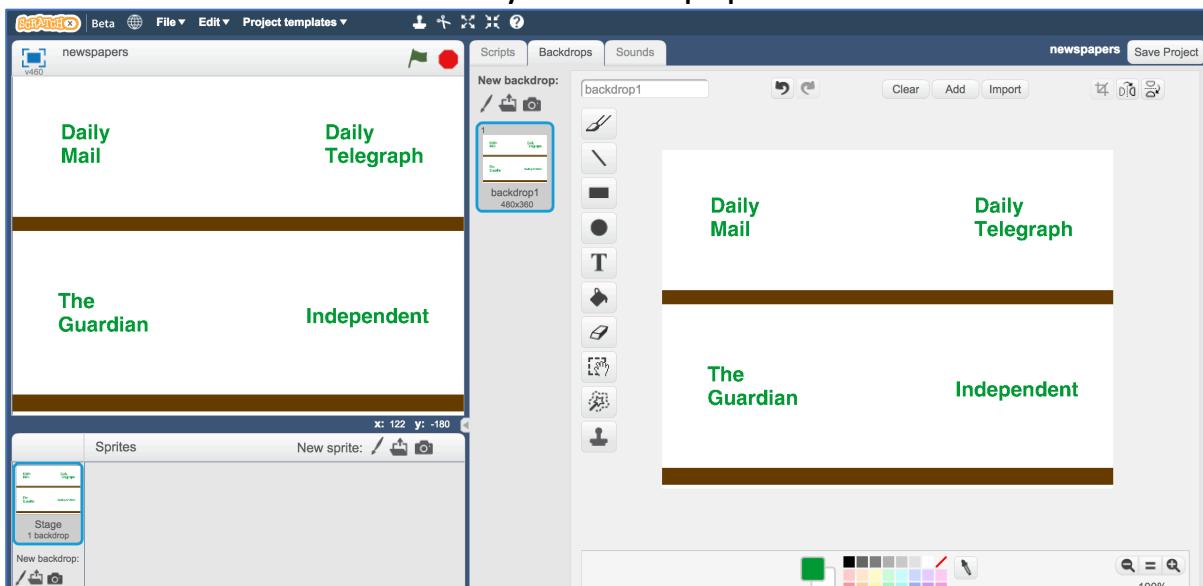
24. You should see blocks in “More blocks” section from your project.



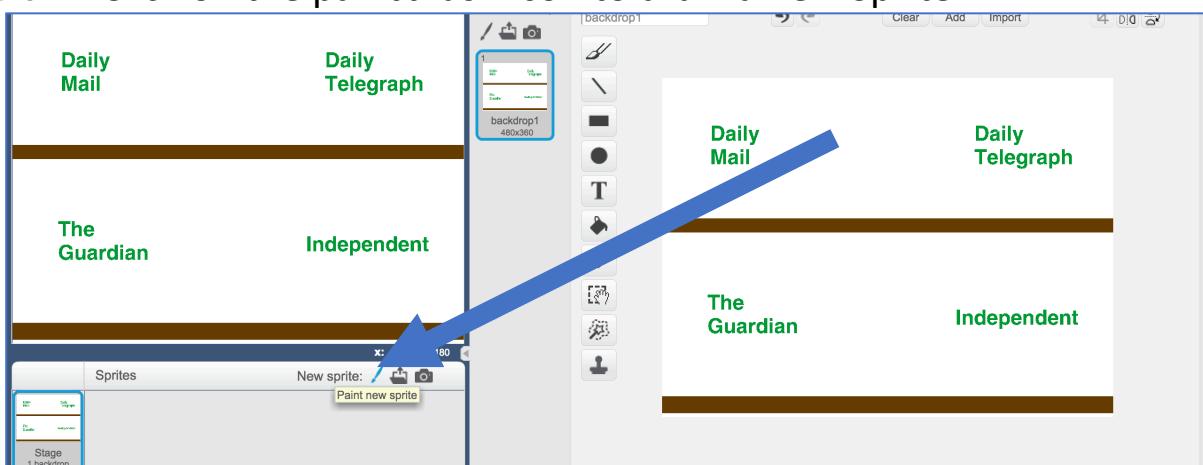
25. Click on the Stage, and then click on the Backdrops tab



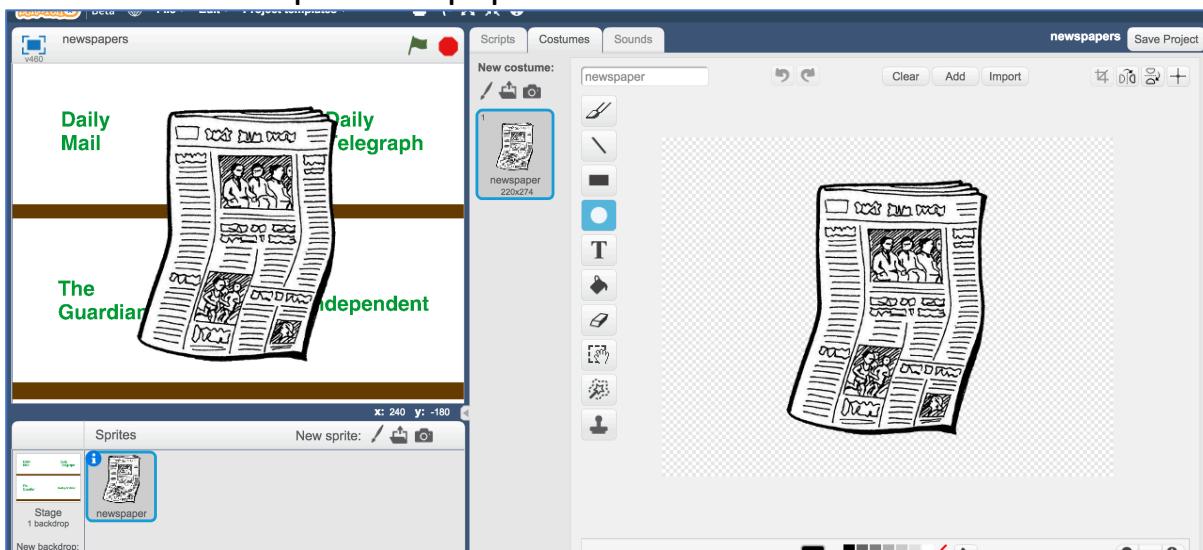
26. Draw some shelves for your newspapers



27. Click on the paintbrush icon to draw a new Sprite



28. Draw a simple newspaper

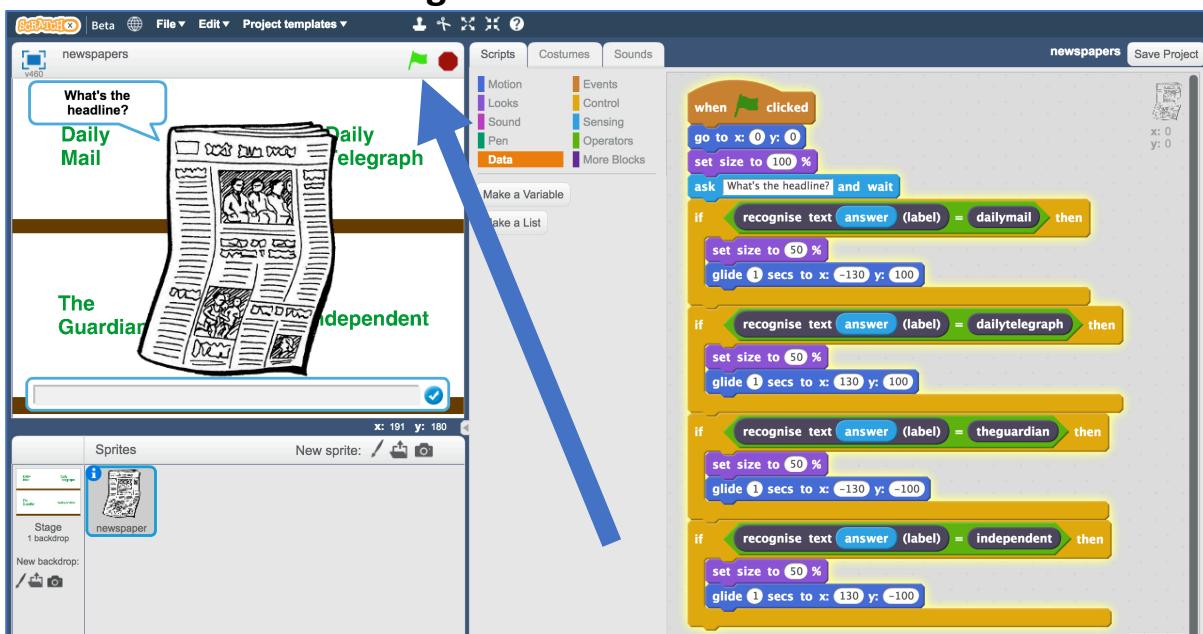


29. Create the following script

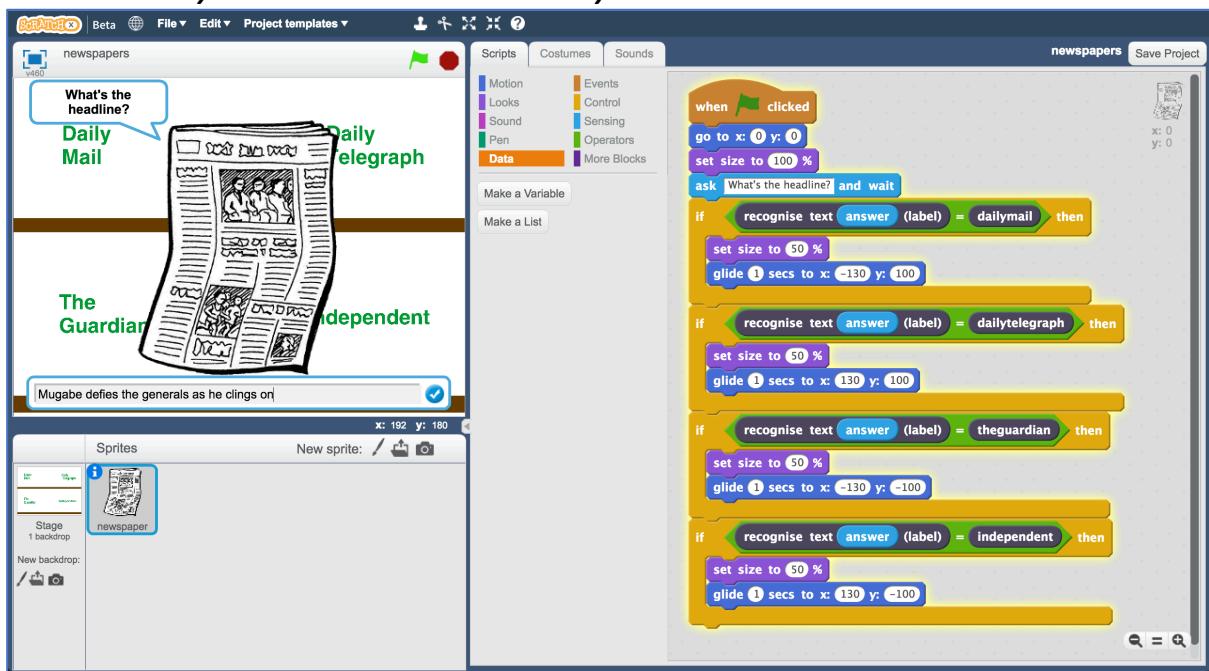
When you type in a headline, this script will predict what newspaper the headline is from and put it on the right shelf.



30. Click the Green Flag to test



31. Try typing in different headlines and see how the computer behaves
Make sure you use headlines that you didn't use to train the model



32. Save your Scratch project
*Click on **File** -> **Save Project***

What have you done so far?

You've started to train a machine learning model to recognise newspaper front-page headlines. You've given it a very small number of examples to learn from – only a single week's worth of newspapers.

You've created a project in Scratch to test your model on headlines it hasn't seen before. You're probably seeing that it gets a lot wrong.

The next step is to try and improve the results by giving our machine learning model some more training. With more examples of headlines, it will hopefully do better.

33. Switch back to the training window

34. Click on the “< Back to project” link.

35. Click the Train button

36. Repeat steps 13-15.

Collect additional headlines from another couple of weeks of newspapers.

You should now have fifteen example headlines from each newspaper.

The screenshot shows a web-based machine learning training interface. At the top, it says "ml-for-kids" and includes links for Welcome, About, Projects, Worksheets, News, Help, and Log Out. Below this, the main title is "Recognising **text** as **dailymail, dailytelegraph or 2 other classes**". A "Back to project" link is visible. On the right, there's a button labeled "+ Add new label". The interface is divided into four sections, each representing a news outlet:

- dailymail**: Headlines include "BREASTFEEDING LESSONS AT 11", "HARRY: NO ROYAL WANTS THR...", "SHAME OF INSURANCE GIANTS", "MASSACRE OF HOLIDAY FAMIL...", "LET'S REIGNITE BRITISH SPIRIT", "VOLVO DEATH KNELL FOR PET...", "ZOO GIRL MAULED TO DEATH ...", "WAR ON HEART DEATHS", and "THREE LETHAL QUESTIONS".
- dailytelegraph**: Headlines include "BBC's gender pay gap revealed", "Draw up your will in a text message", "Set young free online 'to save the...'", "France's new hope puts cloud ov...", "Your country needs you!", "Statins 'needlessly doled out to ...'", "Corbyn ducks terror challenge", "Quarter of care homes are unsafe", and "Terror strikes the heart of Barcelona".
- theguardian**: Headlines include "'Racism is evil', says Trump, 48 h...", "May intensifies personal attack a...", "Ban from 2040 on diesel and petr...", "Backbenchers tell May: sack the ...", "Labour threat to defeat May on B...", "US federal workers are told: stop...", "Macron wins French presidency ...", "Corbyn and May make last pitch ...", and "Trump fires media chief Scaramu...".
- independent**: Headlines include "May abandons climate change pr...", "Government demands tariff-free t...", "NHS faces £1bn Brexit bill for tre...", "Male fertility crisis is 'wake-up ca...', "Brexit IT 'horror' puts £34bn bord...", "May given go-ahead to fire cabin...", "France stems tide of the populist ...", "Government accused of failing di...", and "Final reckoning: May on track for ...".

Each section has a "+ Add example" button at the bottom.

37. Click the “< Back to project” link. Click the **Learn & Test** button.

38. Click the **Train new machine learning model** button

Wait for the training to complete. This might take a few minutes.

39. Switch back to the Scratch window.

If you accidentally closed it, you can get back to it by doing this:

- * Click the “**< Back to project**” link
- * Click the “**Make**” button
- * Click the “**Scratch**” button
- * Click the “**Open in Scratch**” button
- * Open the Scratch project you saved before, with “**File**” -> “**Load Project**”

40. Repeat steps 28 – 29 again.

See how your new model is doing.

Does your model predict which newspaper a headline comes from more often than before?

What have you done?

You've trained a machine learning model to recognise the use of language in the front page headlines of four national newspapers.

You've tested the model to assess the performance of the machine learning model.

You checked this at various stages of your model's training to see how the performance improved as you continued to train it.

Ideas and Extensions

Now that you've finished, why not give one of these ideas a try?

Or come up with one of your own?

Look for other patterns in newspapers

Instead of comparing tabloid papers with broadsheet newspapers, try some other project ideas.

What about comparing local newspapers with national newspapers? Or weekend editions with week day papers?

Try different numbers of newspapers

What happens if you add additional newspapers? Is it still as good? What happens if you remove some, and only train it to tell headlines from two newspapers? Does it do better then?

Use newspaper stories instead of headlines

What about the text of newspaper stories instead of headlines? Can you train a machine learning model to recognise the style of writing in the story?