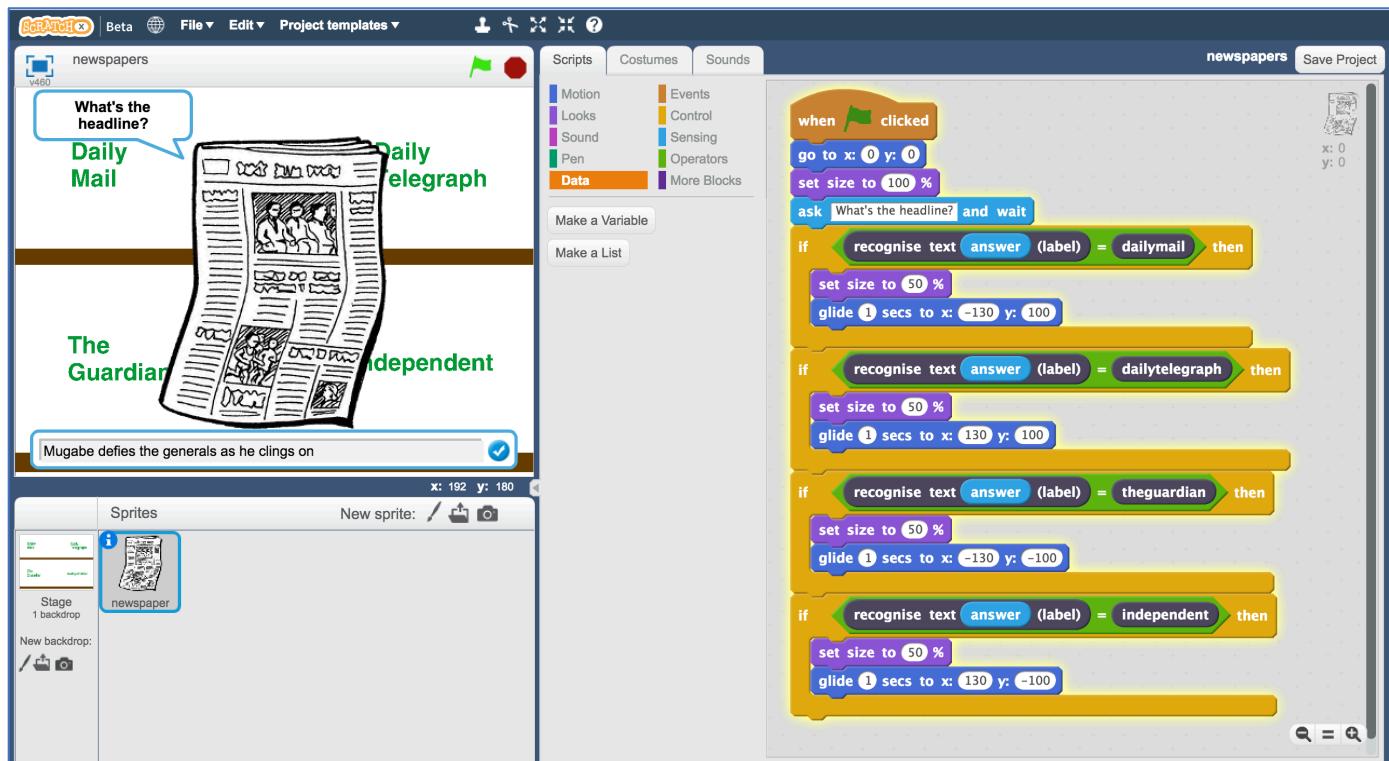


Headlines

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 *

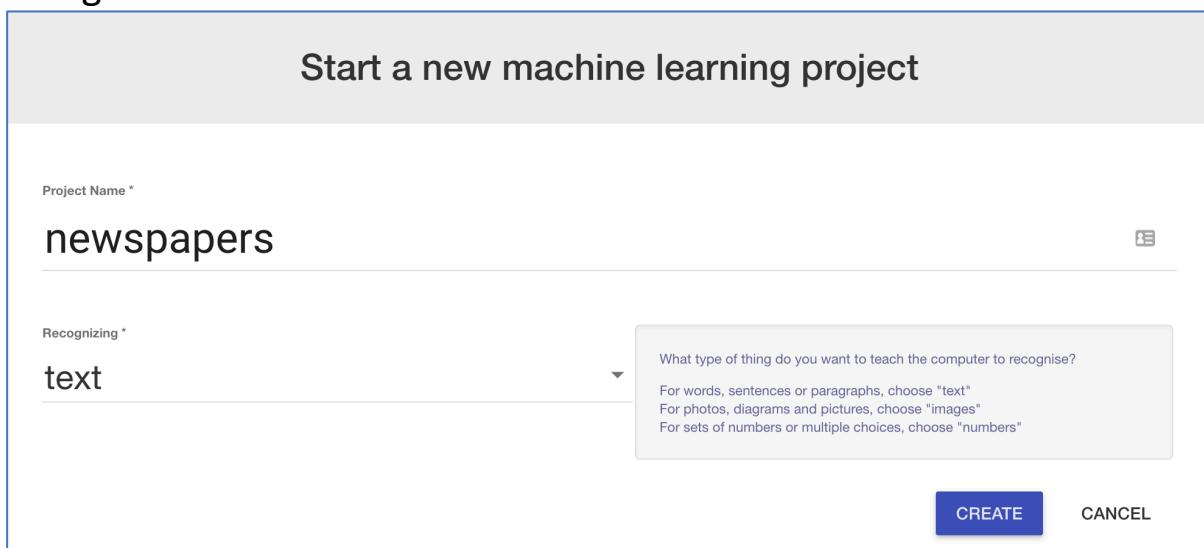
newspapers

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**



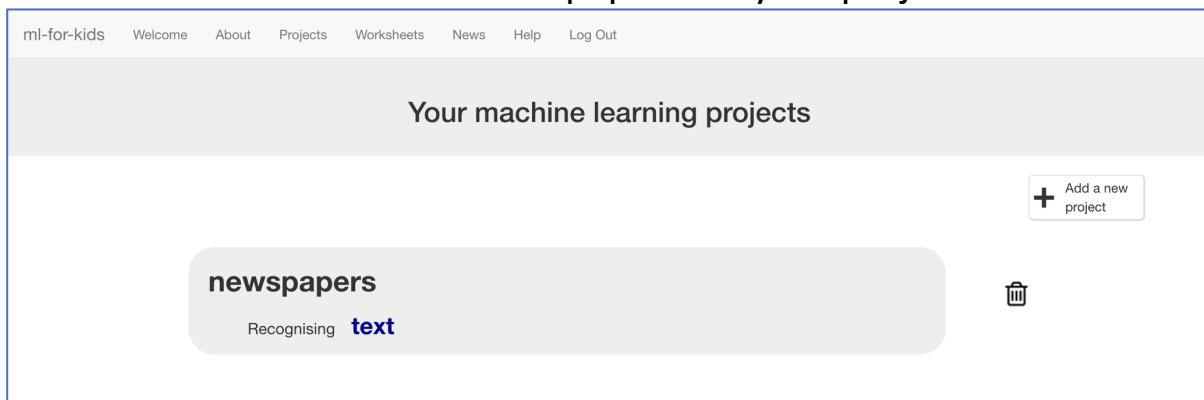
7. You should now see “newspapers” in your projects list. Click on it.

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

The screenshot shows a user interface for a machine learning project. At the top, there is a navigation bar with links: ml-for-kids, Welcome, About, Projects, Worksheets, News, Help, and Log Out. Below the navigation bar, the text "newspapers" is displayed. There are three main buttons: "Train", "Learn & Test", and "Scratch". The "Train" button is highlighted with a blue background and white text. Below each button is a brief description and a blue "Start" button.

Train
Collect examples of what you want the computer to recognise.
Start

Learn & Test
Use the examples to train the computer to recognise text.
Start

Scratch
Use the machine learning model you've trained to make a game in Scratch.
Start

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.

The screenshot shows a list of four labeled buckets for text recognition. The labels are "dailymail", "dailytelegraph", "theguardian", and "independent". Each bucket has a "Add example" button at the bottom. A "Back to project" link is visible at the top left, and a "+ Add new label" button is at the top right.

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

< Back to project

+ Add new label

dailymail

dailytelegraph

theguardian

independent

+ Add example

+ Add example

+ Add example

+ Add example

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 interface for training a machine learning model. On the left, there are four rectangular 'buckets' for classifying text. The top row contains 'dailymail' and 'dailytelegraph'. The bottom row contains 'theguardian' and 'independent'. Each bucket has a small '+ Add example' button at the bottom right. To the right of these buckets is a vertical column of news snippets. The first snippet is from the Daily Express, dated 8/14/2017, with headlines like 'FURY AT PLOT TO GET NEW EU VOTE'. The second snippet is from the Daily Mail, also dated 8/14/2017, with headlines like 'WAR ON HEART DEATHS'. The third snippet is from the Daily Mirror, also dated 8/14/2017, with headlines like 'ANT: MY FIGHT'.

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's 'Add new example' dialog open. The dialog box is centered over the bucket, containing a text input field with the text 'WAR ON HEART DEATHS' and two buttons at the bottom: 'ADD' and 'CANCEL'. The rest of the interface and news snippets are visible in the background.

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 user interface for testing a machine learning model. At the top left, there's a message: "You have trained a machine learning model to recognise when text is dailymail, dailytelegraph or 2 other classes." Below that, it says "You created the model on Tuesday, April 24, 2018 10:42 PM." On the left, under "You have collected:", there's a bulleted list: "• 5 examples of dailymail," "• 5 examples of dailytelegraph," "• 5 examples of theguardian," and "• 5 examples of independent". To the right, there's a note: "try testing the machine learning model below. Enter an example of text below, that you didn't include in the examples you used to train it. It will tell you what it recognises it as, and how confident it is in that." Below this, there's another note: "If the computer seems to have learned to recognise things correctly, then you can go to [Scratch](#) and use what the computer has learned to make a game!" Further down, there's another note: "If the computer is getting too many things wrong, you might want to go back to the [Train](#) page and collect some more examples. Once you've done that, click on the button below to train a new machine learning model and see what difference the extra examples will make!" At the bottom, there's a text input field containing "Cuts 'threaten disaster' for policing terror" and a "Test" button. Below the input field, the text "Recognised as **independent** with 39% confidence" is displayed.

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?

21. Click the “< Back to project link”. Then click “Scratch”.

Your project will add these blocks to the **More Blocks** tab in Scripts.

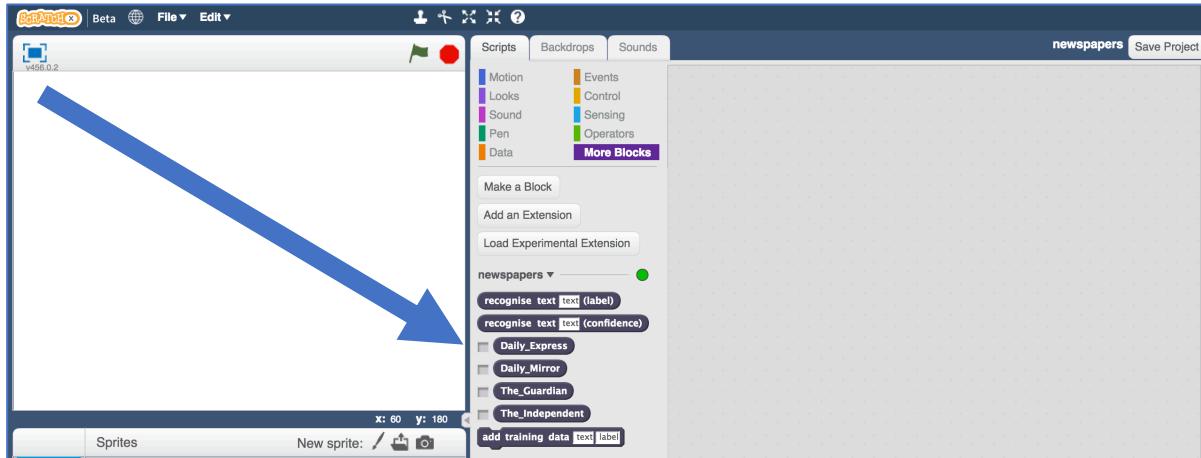
recognise text [text] (label)
Put text in the input for this, and it will return the label that your machine learning model recognises it as.

recognise text [text] (confidence)
This will return how confident your machine learning model is that it recognises the type of text. (As a number from 0 - 100).

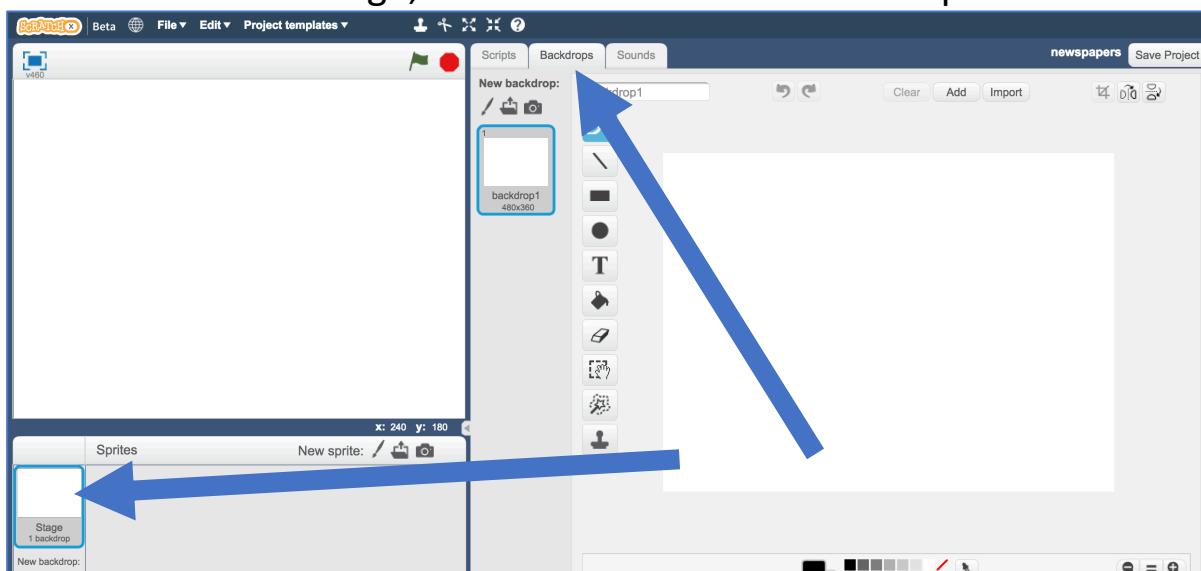
dailymail dailytelegraph theguardian independent
These blocks represent the labels you've created in your project, so you can use their names in your scripts.

It will look something like this - except with the name of your project.

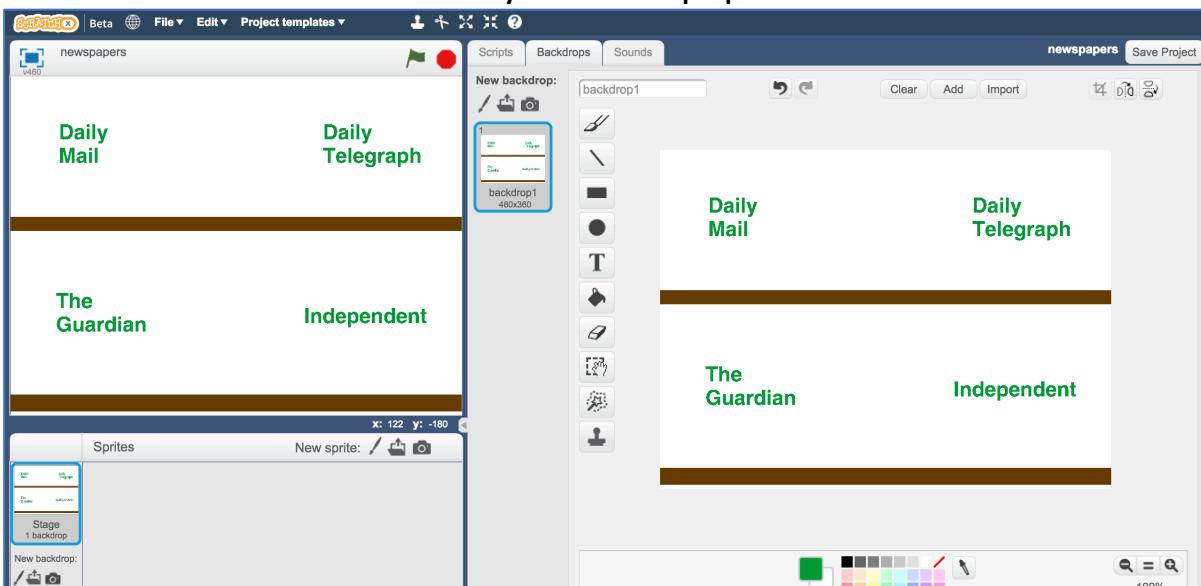
22. Click the “Open in Scratch” button to start making the test. *You should see new blocks in the “More blocks” section from your “newspapers” project.*



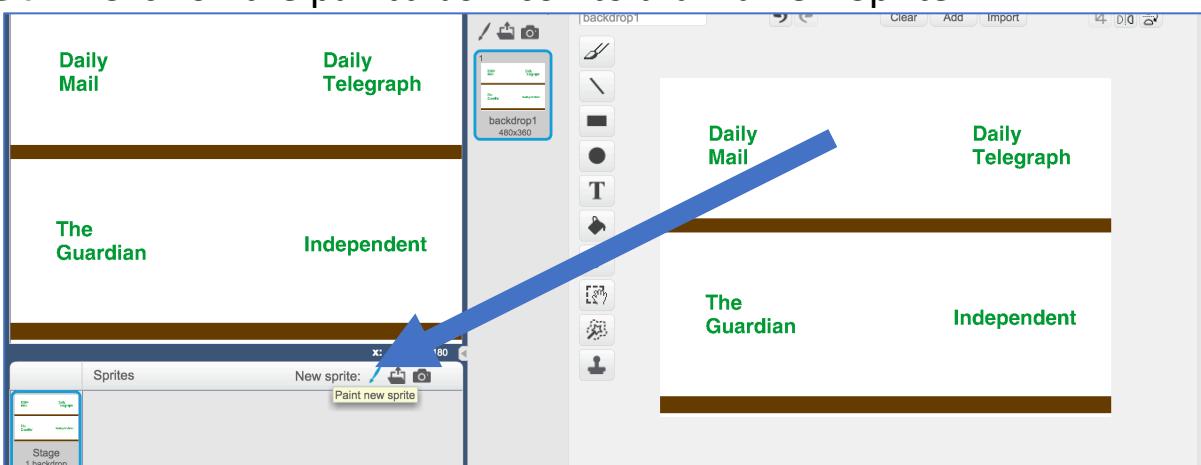
23. Click on the Stage, and then click on the Backdrops tab



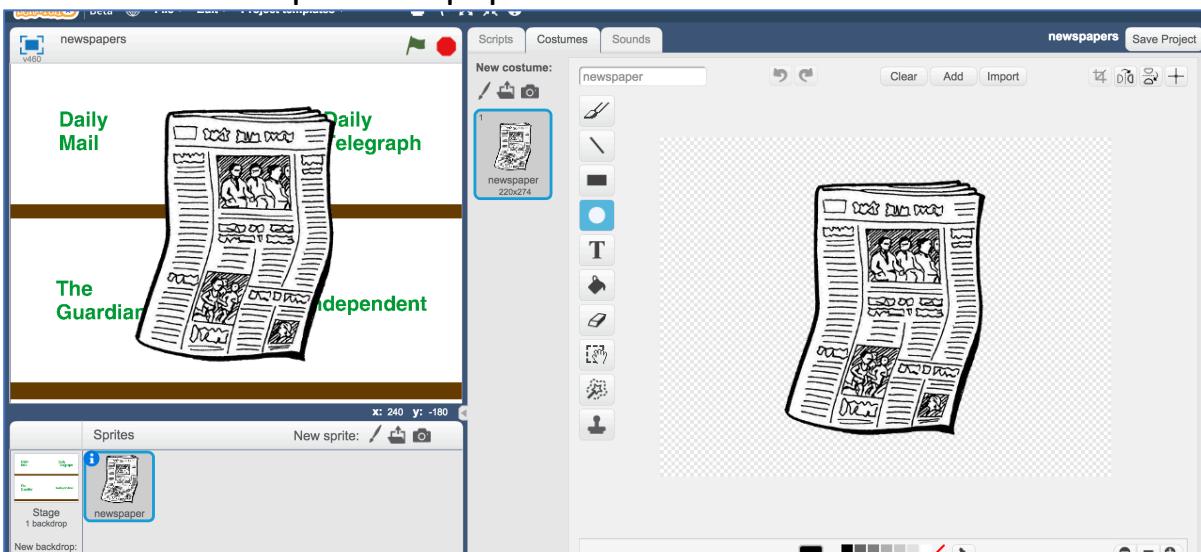
24. Draw some shelves for your newspapers



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



26. Draw a simple newspaper

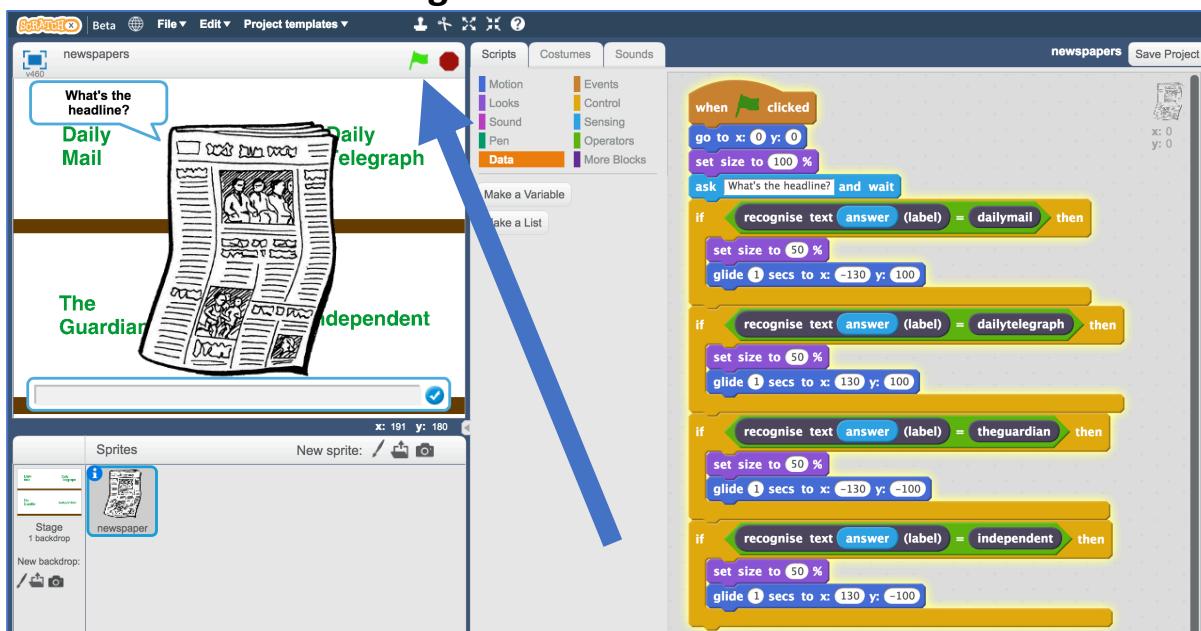


27. Create the following script

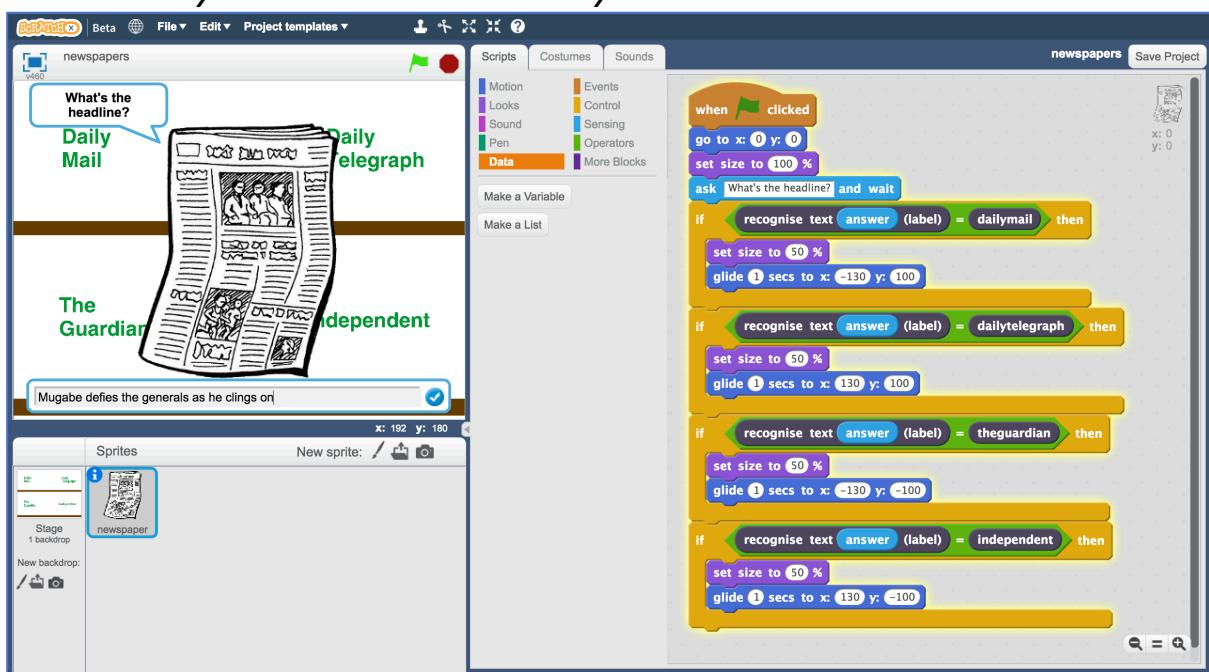
When you type in a headline, this script will predict what newspaper the headline is from and put it on the right shelf.



28. Click the Green Flag to test



29. 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



30. 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.

- 31.** Switch back to the training window
- 32.** Click on the “< Back to project” link.
- 33.** Click the Train button
- 34.** 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, there's a navigation bar with links: ml-for-kids, Welcome, About, Projects, Worksheets, News, Help, and Log Out. Below the navigation, a title reads "Recognising **text** as **dailymail, dailytelegraph or 2 other classes**". There are four main sections, each with a heading and a list of news headlines:

- dailymail** headlines include: BREASTFEEDING LESSONS AT 11, HARRY: NO ROYAL WANTS THR..., SHAME OF INSURANCE GIANTS, MASSACRE OF HOLIDAY FAMILI..., LET'S REIGNITE BRITISH SPIRIT, VOLVO DEATH KNELL FOR PET..., ZOO GIRL MAULED TO DEATH ..., WAR ON HEART DEATHS, THREE LETHAL QUESTIONS. Each headline has a "+ Add example" button below it.
- 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. Each headline has a "+ Add example" button below it.
- 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 ..., Trump fires media chief Scaramu... Each headline has a "+ Add example" button below it.
- 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..., Final reckoning: May on track for ... Each headline has a "+ Add example" button below it.

In the top right corner of the interface, there is a button labeled "+ Add new label".

- 35.** Click the “< Back to project” link. Click the **Learn & Test** button.
- 36.** Click the **Train new machine learning model** button
Wait for the training to complete. This might take a few minutes.
- 37.** 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 “Scratch” button
 - * Click the “Open in Scratch” button
 - * Open the Scratch project you saved before, with “File” -> “Load Project”

38. 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?