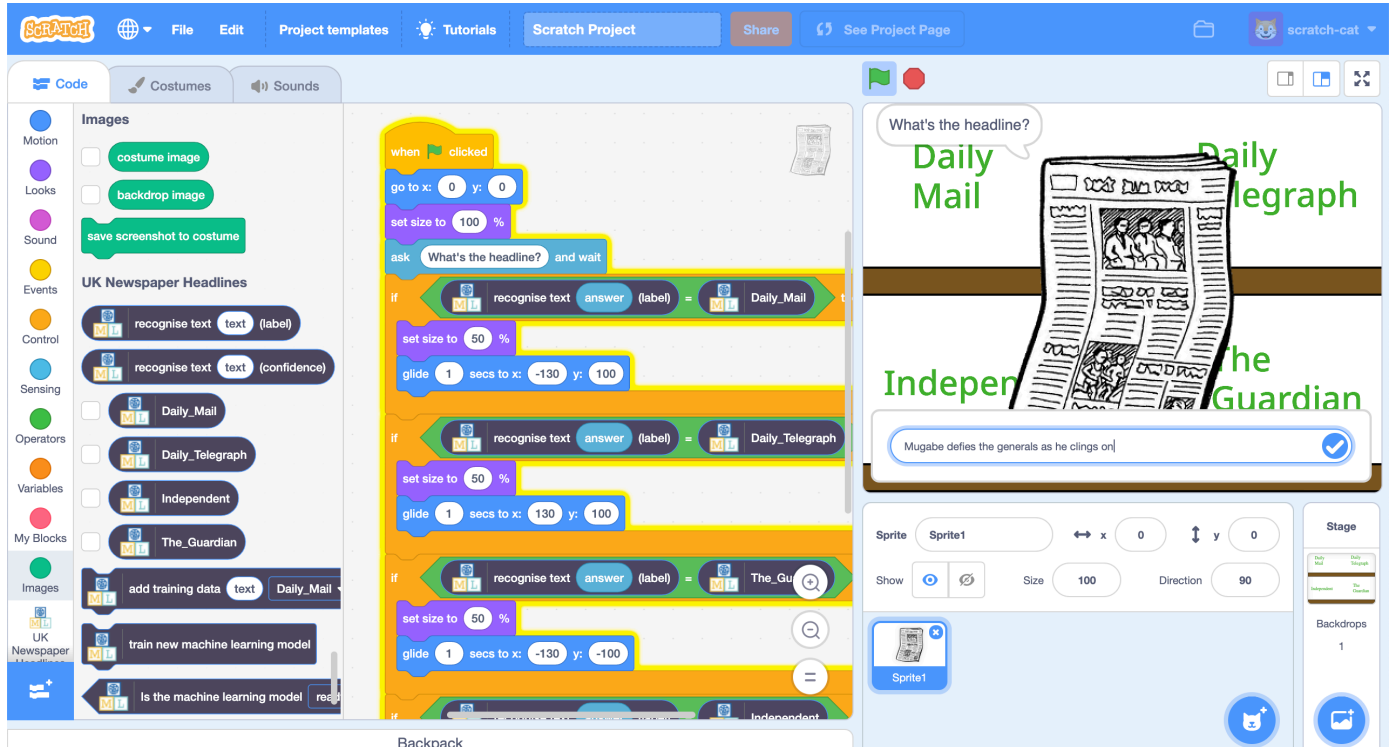




Newspaper shelves

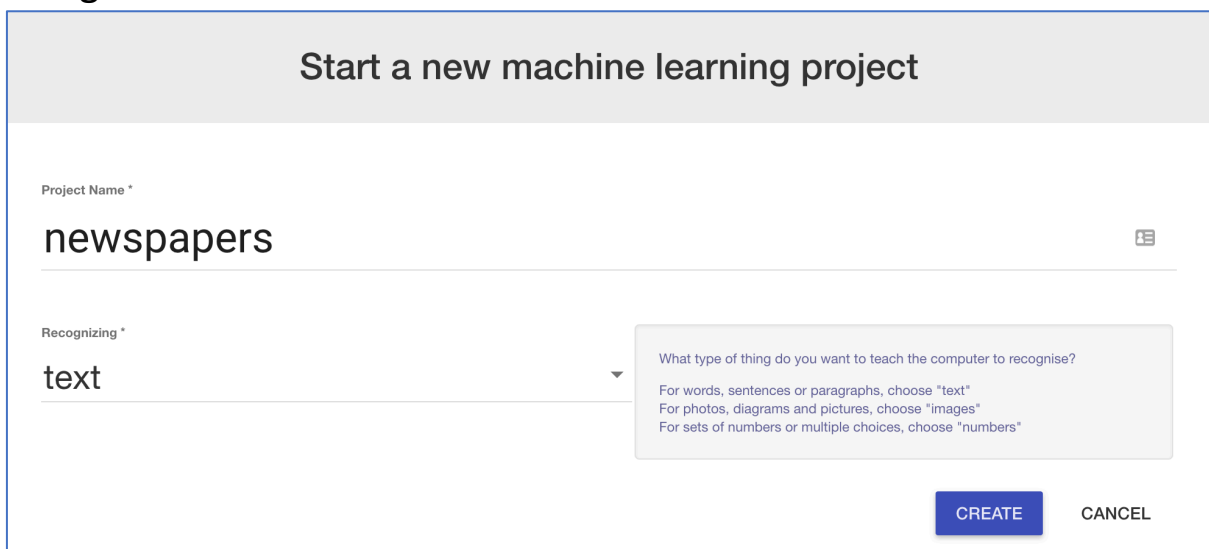
In this project, you will train a computer to recognise headlines from different newspapers.

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.



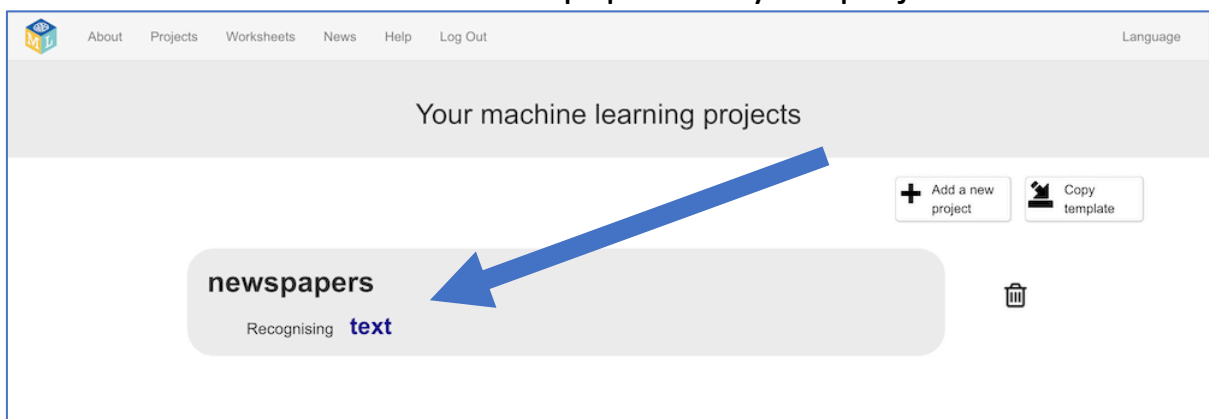
This project worksheet is licensed under a Creative Commons Attribution Non-Commercial Share-Alike License
<http://creativecommons.org/licenses/by-nc-sa/4.0/>

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**”

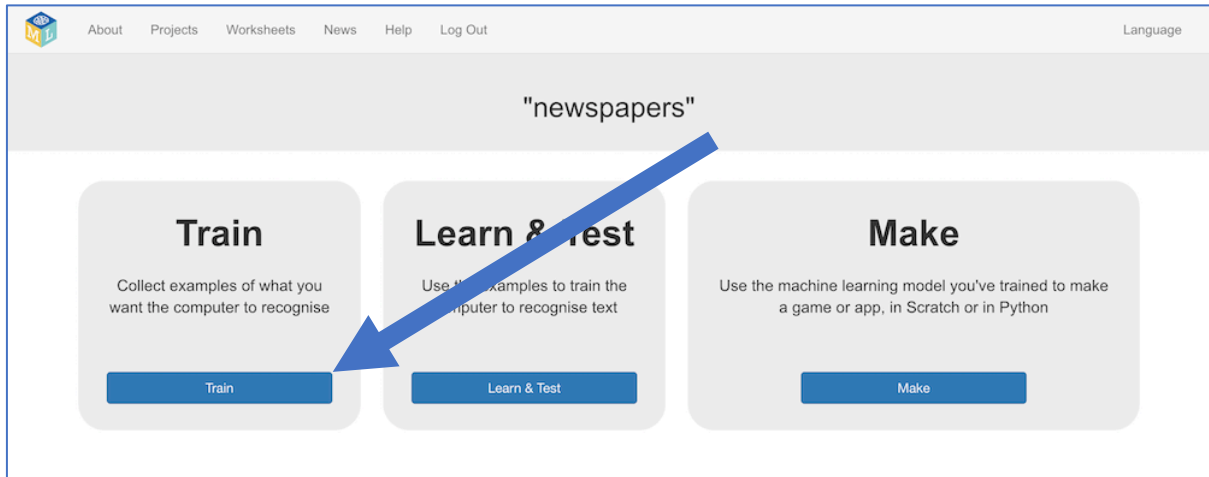


The screenshot shows a form titled "Start a new machine learning project". It has two main input fields: "Project Name *" with the text "newspapers" and "Recognizing *" with the text "text". A dropdown menu is open next to the "Recognizing" field, showing a list of options: "What type of thing do you want to teach the computer to recognise?", "For words, sentences or paragraphs, choose 'text'", "For photos, diagrams and pictures, choose 'images'", and "For sets of numbers or multiple choices, choose 'numbers'". At the bottom right, there are two buttons: "CREATE" and "CANCEL".

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



8. Click the “Train” button



9. Choose four national newspapers.

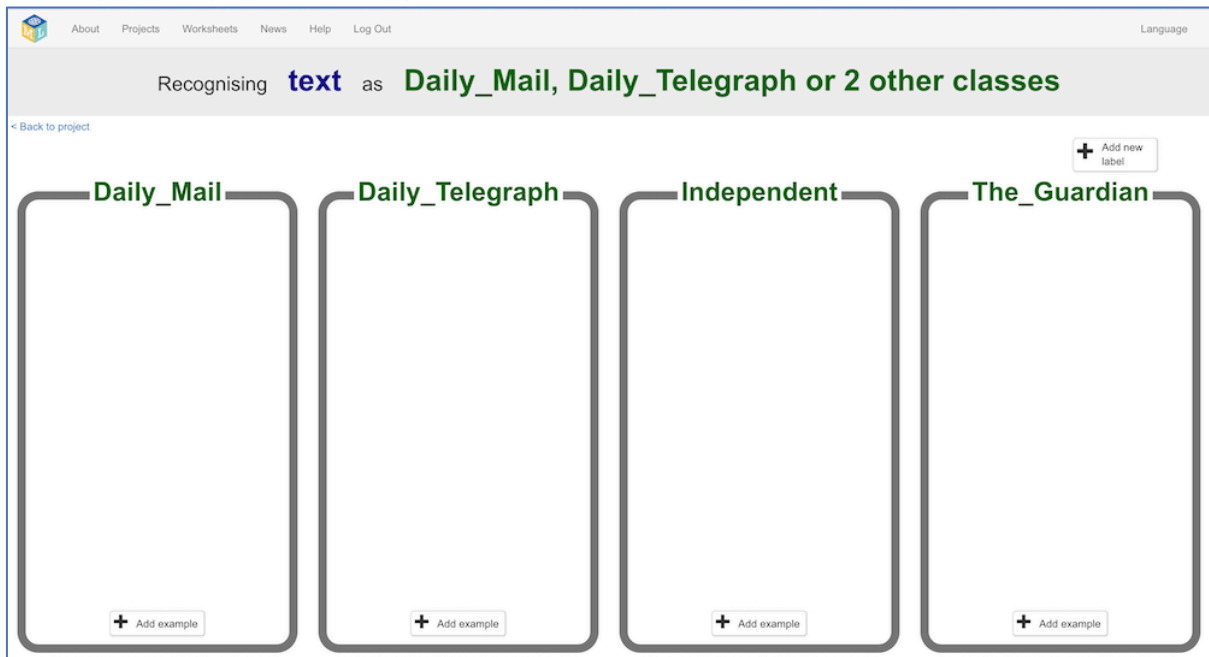
To save having to buy actual physical newspapers, you can use an online collection of newspaper front pages.

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.

Recognising **text** as **Daily_Mail**, **Daily_Telegraph** or 2 other classes

< Back to project

+ Add new label

Daily_Mail + Add example

Daily_Telegraph + Add example

Independent + Add example

The_Guardian + Add example

Newspaper Headlines for 14 August 2017

[Click Here to Switch to Front Pages View](#)

8/14/2017: FRONT PAGE: FURY AT PLOT TO GET NEW EU VOTE
FEARS were raised last night that opponents of Brexit will push for a second referendum after Labour veteran David Miliband provoked outrage by insisting there should be a referendum on the terms of Britain's final exit deal with the...

- Visit [Daily Express Web Site](#)
- Read the [Digital Edition of Daily Express at PressDisplay](#)
- View [Daily Express Twitter Feed](#)

8/14/2017: Front Page: WAR ON HEART DEATHS
MILLIONS are to be offered checks at GP surgeries and pharmacies in an effort to reduce heart disease and early deaths. The strategy aims to raise dramatically the 'killer' conditions including high blood pressure, high cholesterol and obesity.

- Visit [Daily Mail Web Site](#)
- Read the [Digital Edition of Daily Mail at PressDisplay](#)
- View [Daily Mail Twitter Feed](#)

Visit [Daily Mirror Web Site](#)

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.

Recognising **text** as **Daily_Mail**, **Daily_Telegraph** or 2 other classes

< Back to project

+ Add new label

Daily_Mail + Add example

Daily_Telegraph + Add example

Independent + Add example

The_Guardian + Add example

Add example

Enter an example of 'Daily_Mail'

WAR ON HEART DEATHS 19 / 1000

ADD **CANCEL**

8/14/2017: FRONT PAGE: FURY AT PLOT TO GET NEW EU VOTE

- View [Daily Express Twitter Feed](#)

8/14/2017: Front Page: WAR ON HEART DEATHS

- Visit [Daily Mail Web Site](#)
- Read the [Digital Edition of Daily Mail at PressDisplay](#)
- View [Daily Mail Twitter Feed](#)

Visit [Daily Mirror Web Site](#)

- View [Daily Mirror Twitter Feed](#)

8/14/2017: News: Rally girl is named

THE woman run over and killed at an anti-fascist rally in London was named as 21-year-old Hannah...

14. Repeat this for all four newspapers on this day

Recognising **text** as **Daily_Mail, Daily_Telegraph or 2 other classes**

< Back to project

+ Add new label

Daily_Mail

WAR ON HEART DEATHS

+ Add example 1

Daily_Telegraph

Rees-Mogg: Cut stamp duty now

+ Add example 1

Independent

May's plan to pack Brexit commit...

+ Add example 1

The_Guardian

Trump under fire from all sides a...

+ Add example 1

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

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

- Visit [The Guardian Web Site](#)
- Read the [Digital Edition of The Guardian at PressDisplay](#)
- View [The Guardian Twitter Feed](#)

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

Heather Heyer, 32, was killed when a car driven 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 medal

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.

About Projects Worksheets News Help Log Out Language

Recognising **text** as **Daily_Mail, Daily_Telegraph or 2 other classes**

< Back to project

+ Add new label

Daily_Mail

BREASTFEEDING LESSONS AT ...

HARRY: NO ROYAL WANTS TH...

SHAME OF INSURANCE GIANTS

MASSACRE OF HOLIDAY FAMI...

WAR ON HEART DEATHS

+ Add example 5

Daily_Telegraph

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...

Rees-Mogg: Cut stamp duty now

+ Add example 5

Independent

May abandons climate change pri...

Government demands tariff-free t...

NHS faces £1bn Brexit bill for tre...

Male fertility crisis is 'wake-up call...

May's plan to pack Brexit committ...

+ Add example 5

The_Guardian

'Racism is evil', says Trump, 48 h...

May intensifies personal attack a...

Ban from 2040 on diesel and petr...

Backbenchers tell May: sack the ...

Trump under fire from all sides as...

+ Add example 5

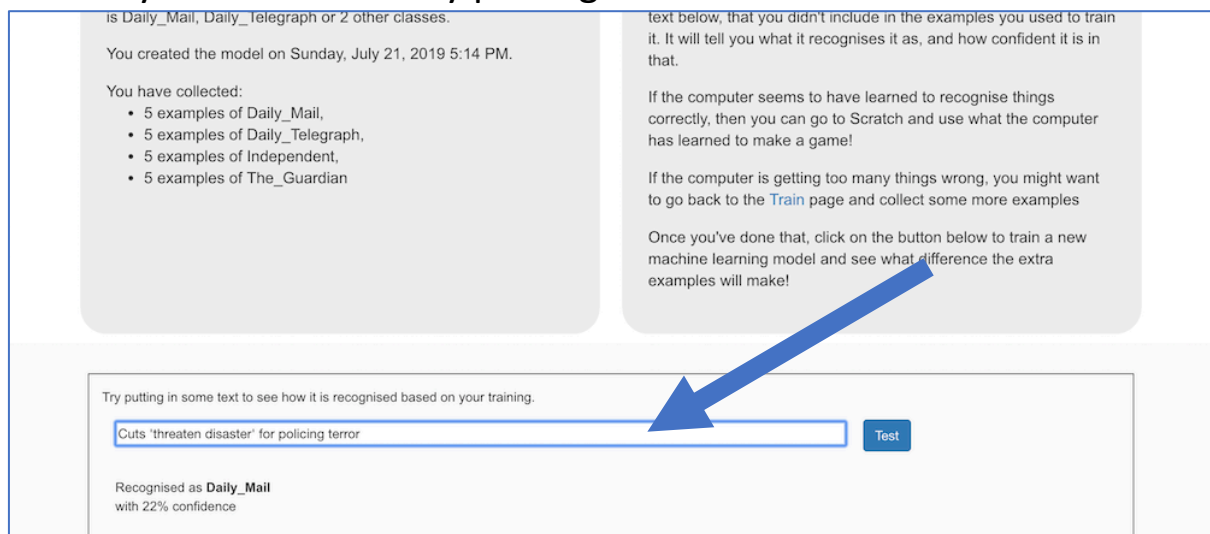
17. Click the “< Back to project” link. Then click “Learn & Test”

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



21. Click the “< Back to project” link

What have you done so far?

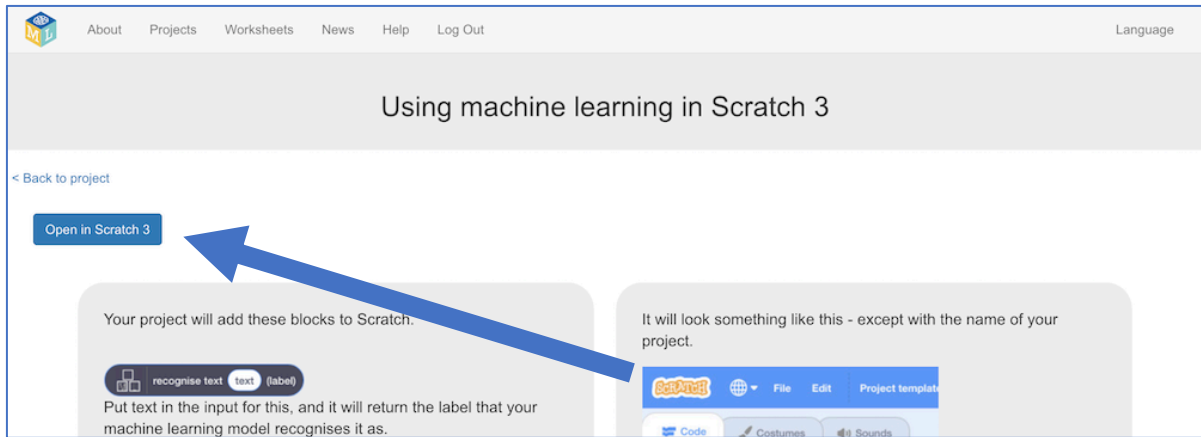
You’ve started to train a machine learning model to recognise patterns in the way that different newspapers use words, phrasing and punctuation in their front-page headlines.

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

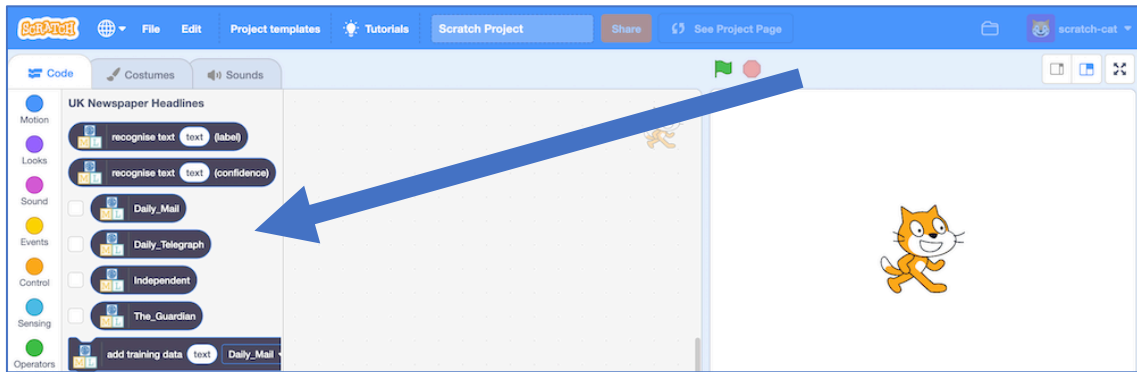
What is it good at? What is it bad at? Which newspapers does it get confused between?

22. Click the “**Make**” button, then click the “**Scratch 3**” button

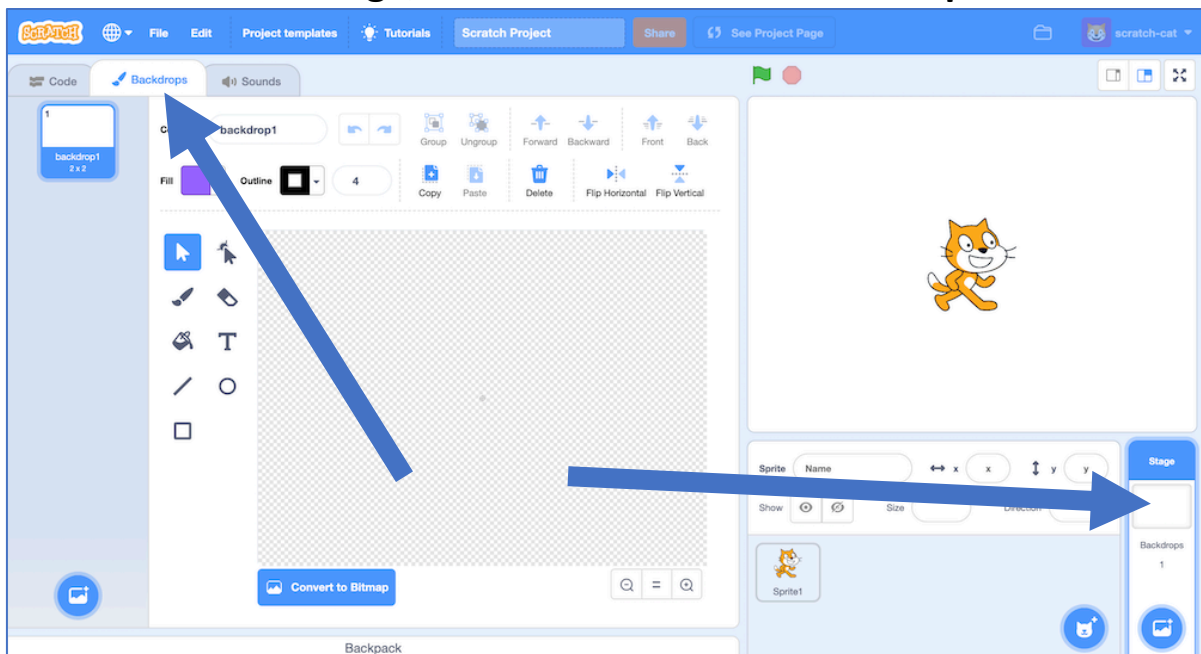
23. Click the “**Open in Scratch 3**” button



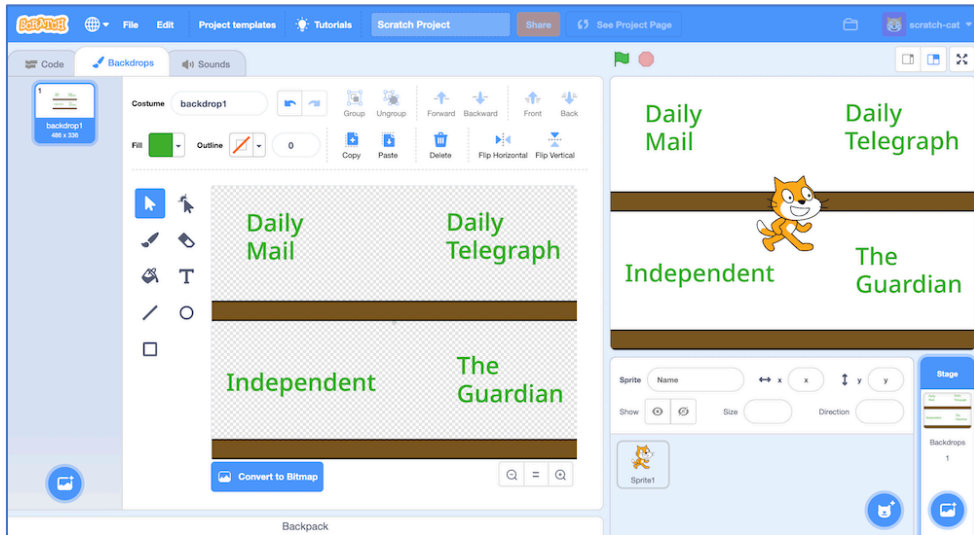
24. You should see blocks (in the bottom section) from your project.



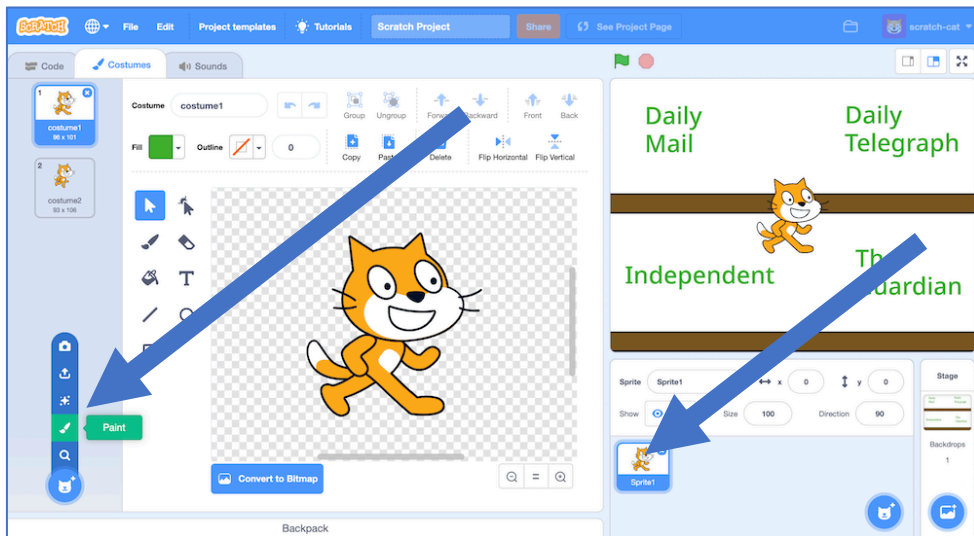
25. Click on the **Stage**, and then click on the **Backdrops** tab



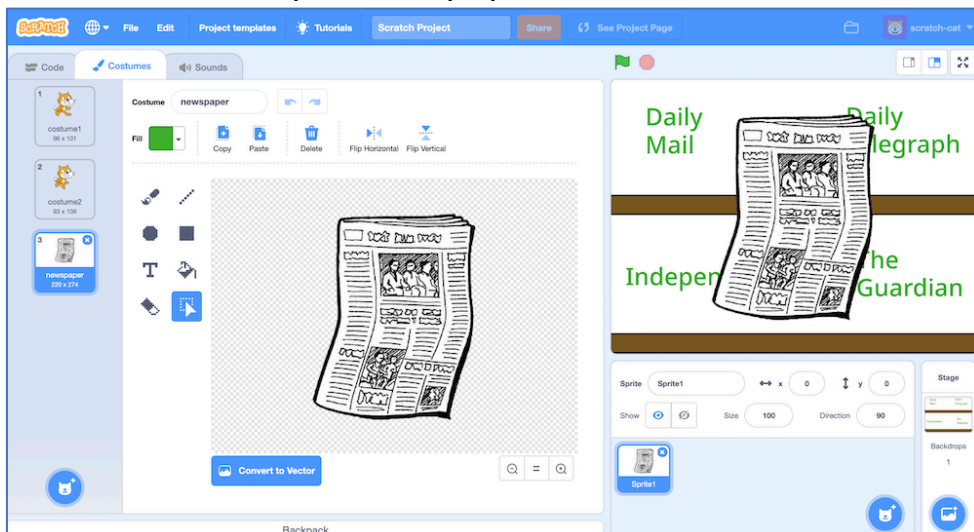
26. Draw some shelves for your newspapers



27. Click on the cat sprite and then click on **Paint** to draw a new Sprite

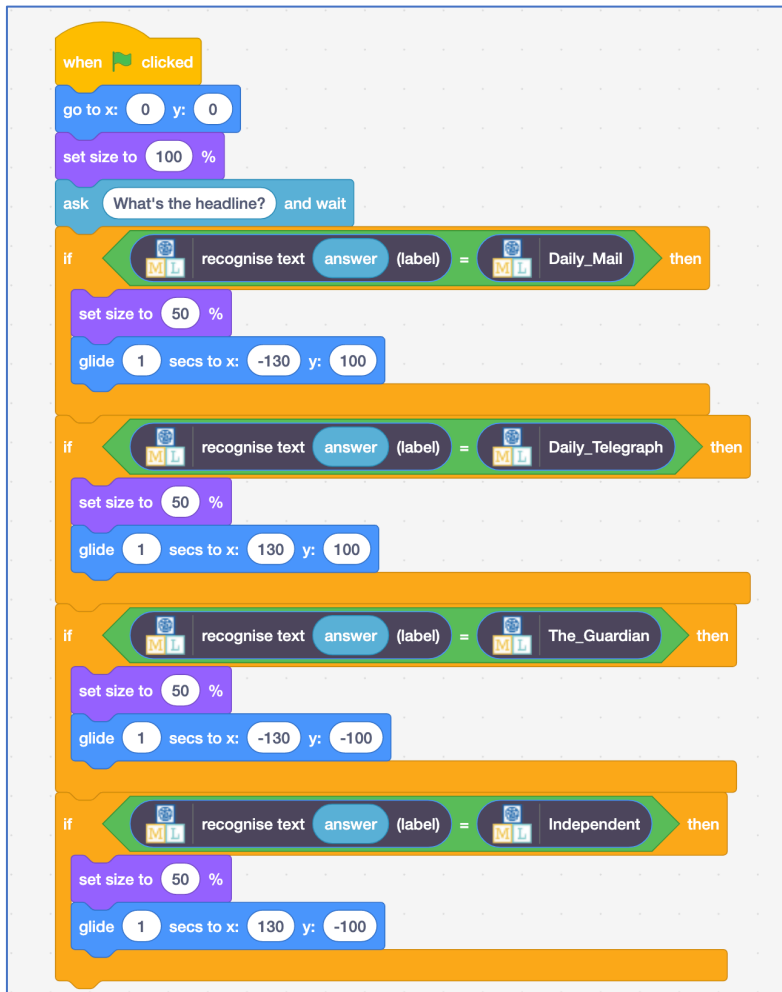


28. Draw a simple newspaper



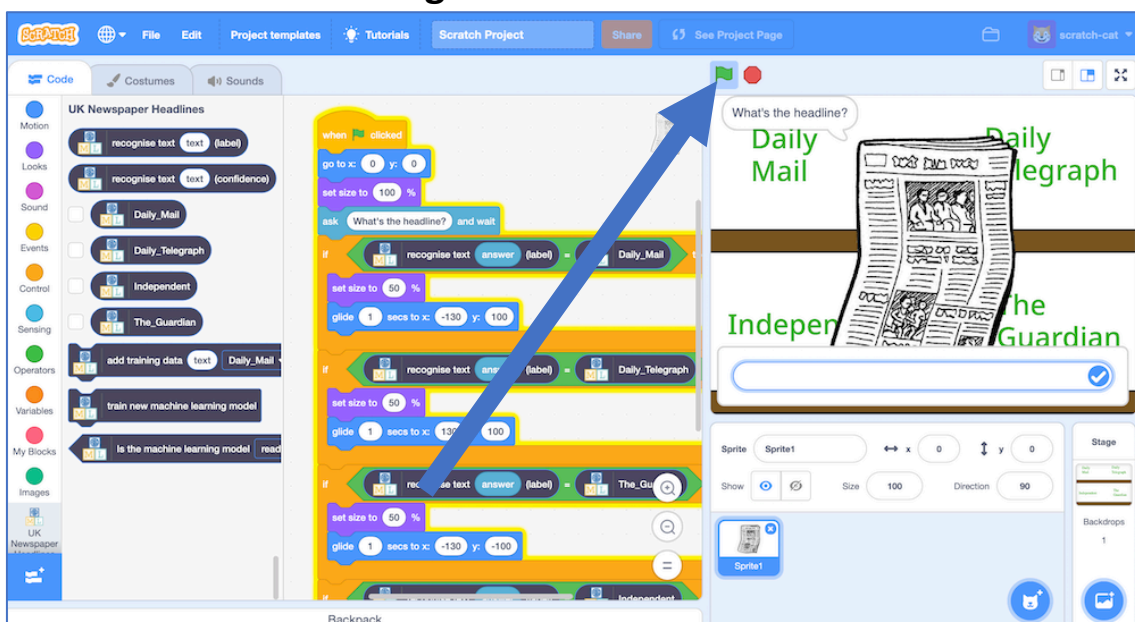
29. Create the following script

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

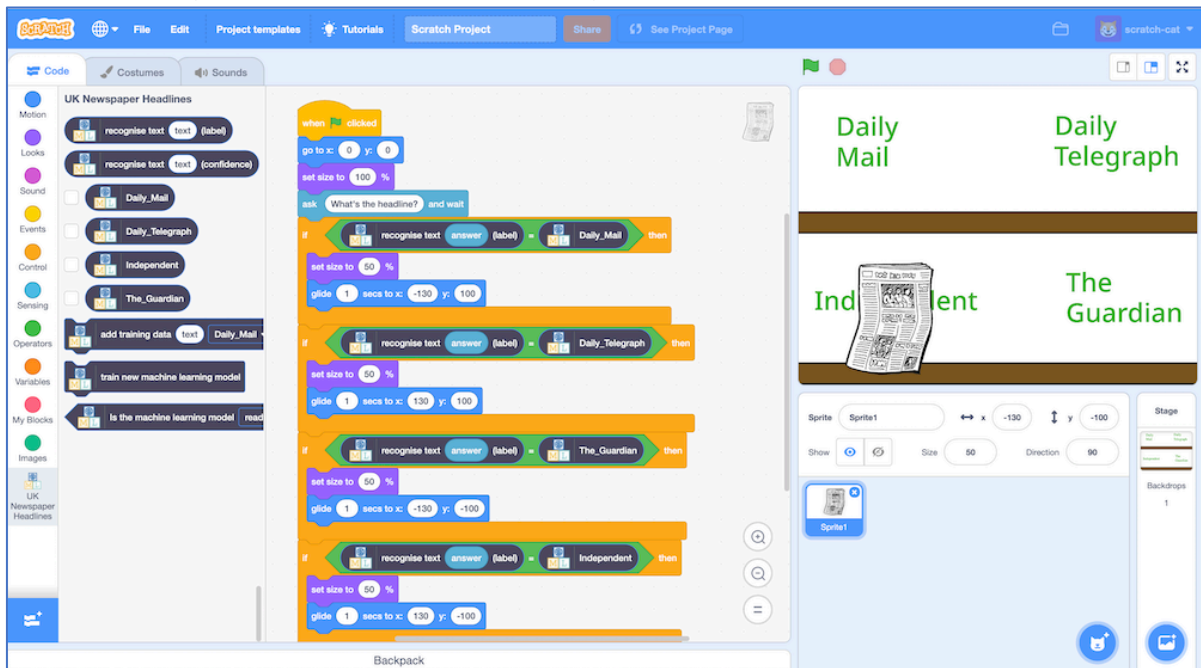


```
when clicked
  go to x: 0 y: 0
  set size to 100 %
  ask "What's the headline?" and wait
  if recognise text answer (label) = Daily_Mail then
    set size to 50 %
    glide 1 secs to x: -130 y: 100
  if recognise text answer (label) = Daily_Telegraph then
    set size to 50 %
    glide 1 secs to x: 130 y: 100
  if recognise text answer (label) = The_Guardian then
    set size to 50 %
    glide 1 secs to x: -130 y: -100
  if recognise text answer (label) = Independent then
    set size to 50 %
    glide 1 secs to x: 130 y: -100
```

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 to your computer

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 interface for training a machine learning model. The title is "Recognising text as Daily_Mail, Daily_Telegraph or 2 other classes". There are navigation links for "About", "Projects", "Worksheets", "News", "Help", and "Log Out". A "Language" dropdown is in the top right. A "Back to project" link is on the left. A "Add new label" button is in the top right. The interface is divided into four columns, each representing a newspaper category:

- Daily_Mail** (71 examples): BREASTFEEDING LESSONS AT..., HARRY: NO ROYAL WANTS TH..., SHAME OF INSURANCE GIANTS, MASSACRE OF HOLIDAY FAMI..., LET'S REIGNITE BRITISH SPIRIT, VOLVO DEATH KNELL FOR PE..., ZOO GIRL MAULED TO DEATH ..., WAR ON HEART DEATHS, THREE LETHAL QUESTIONS, IAROLIR IN MFI TDOWN. + Add example
- Daily_Telegraph** (69 examples): 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', Stalins 'needlessly doled out to m..., Corbyn ducks terror challenge, Quarter of care homes are unsafe, Terror strikes the heart of Barcelona, Rees-Mogg: 'Cut stamp duty now. + Add example
- Independent** (69 examples): May abandons climate change pri..., Government demands tariff-free t..., NHS faces £1bn Brexit bill for tre..., Male fertility crisis is 'wake-up call...', 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 ..., Trump fires press chief - 10 days. + Add example
- The_Guardian** (69 examples): '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 Br..., US federal workers are told: stop ..., Macron wins French presidency -..., Corbyn and May make last pitch f..., Trump fires media chief Scaramucci, Corbyn tells Paxman: 'I'm no diet. + Add example

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 3**” button
- * Click the “**Open in Scratch 3**” button
- * Open the project you saved before, with **File -> Load from your computer**

40. Repeat steps 30 – 31 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?