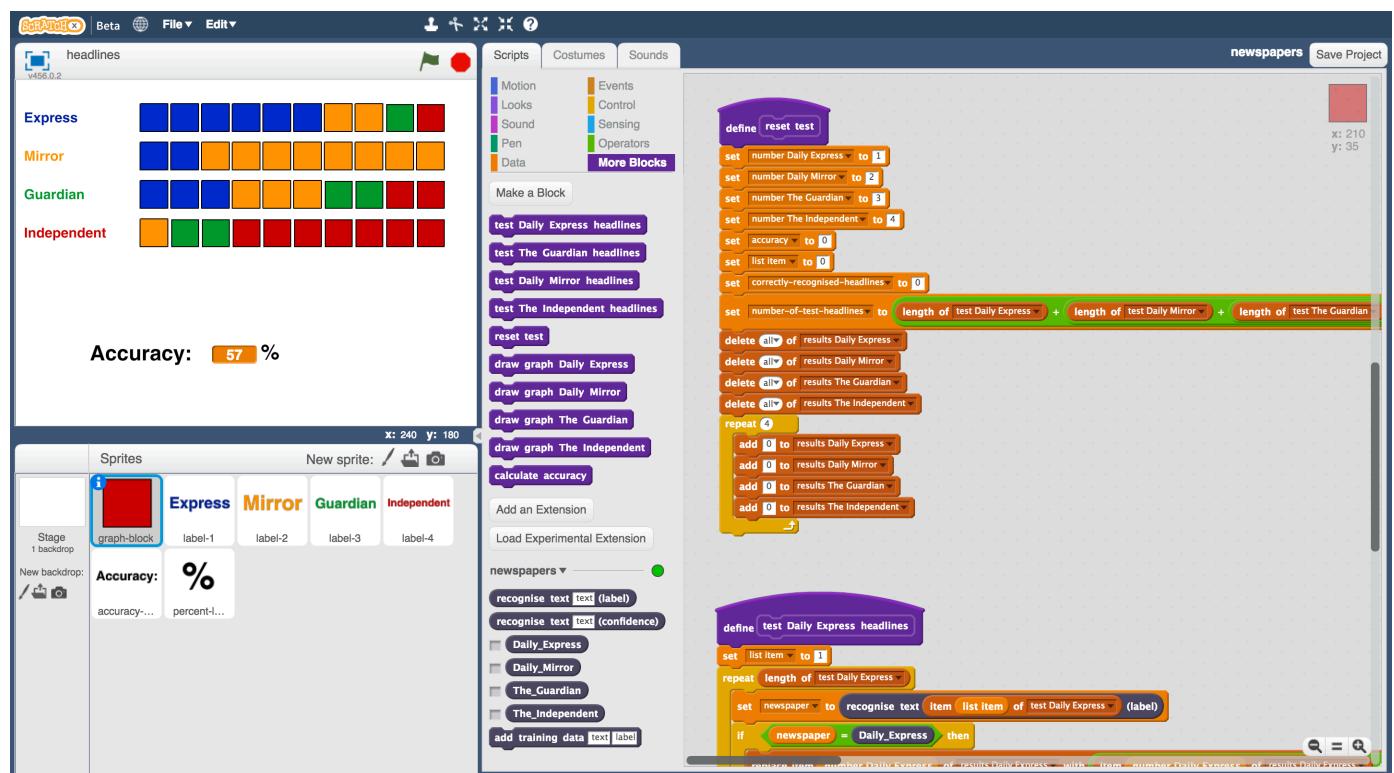


Headlines

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.

You'll also learn about the importance of testing in machine learning projects, and create a simple testing tool in Scratch.



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 *

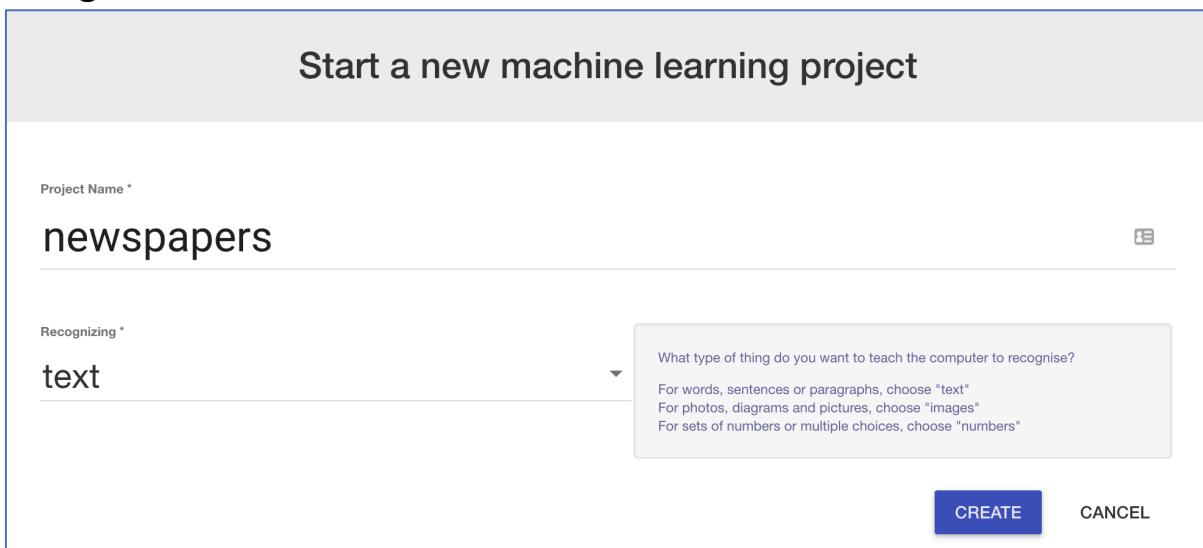
newspapers

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**



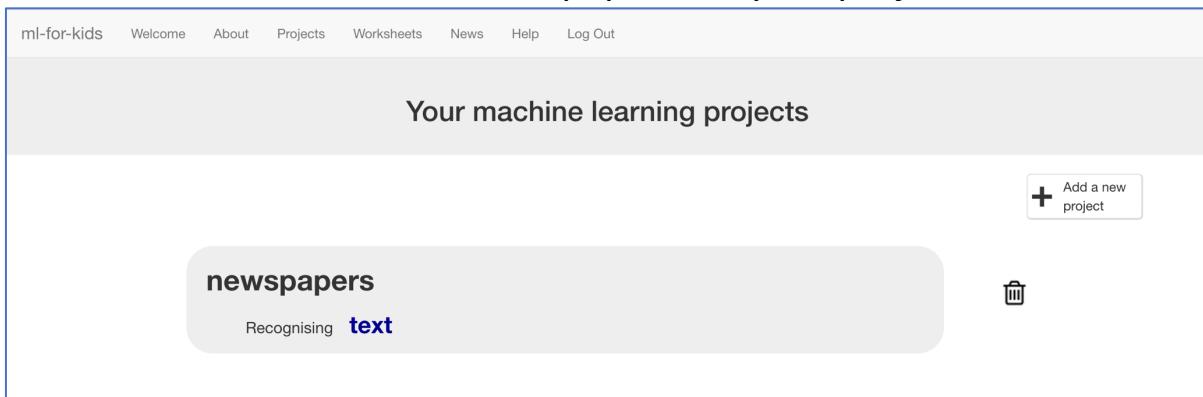
7. You should now see “newspapers” in your projects list. Click on it.

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. A tooltip below it says: "Collect examples of what you want the computer to recognise." Below the "Train" button is a blue "Train" button. The "Learn & Test" button has a tooltip: "Use the examples to train the computer to recognise text." Below it is a blue "Learn & Test" button. The "Scratch" button has a tooltip: "Use the machine learning model you've trained to make a game in Scratch." Below it is a blue "Scratch" 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 Express, Daily Mirror, 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 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 "Recognising text as Daily_Express, Daily_Mirror or 2 other classes" is displayed. There are four labeled boxes representing different newspapers: "Daily_Express", "Daily_Mirror", "The_Guardian", and "The_Independent". Each box has a "Add example" button at the bottom. A "Add new label" button is located in the top right corner of the interface.

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 displays two adjacent browser windows. The left window is titled 'Machine Learning for kids' and contains a classification interface. It features four rectangular 'buckets' for training data, each with a label: 'Daily_Express', 'Daily_Mirror', 'The_Guardian', and 'The_Independent'. Below each bucket is a small button labeled '+ Add example'. The right window is titled 'Newspaper Headlines for Mon' and shows three news snippets from 'Newspaper Headlines for 14 August 2017'. The first snippet is from 'DAILY EXPRESS' with the headline 'FURY AT PLOT TO GET NEW EU VOTE'. The second is from 'Daily Mail' with the headline 'WAR ON HEART DEATHS'. The third is from 'DAILY MIRROR' with the headline 'ANT: MY FIGHT TO SAVE MARRIAGE'. Each snippet includes a small thumbnail image and a brief description below it.

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 'Add new example' dialog box overlaid on the machine learning interface. The dialog box has a blue header bar with the text 'Add new example'. Below this is a white input area with a text field containing the headline 'FURY AT PLOT TO GET NEW EU VOTE'. At the bottom of the input area are two buttons: 'ADD' on the left and 'CANCEL' on the right. The background of the dialog box is semi-transparent, allowing the underlying training interface and news snippets to be visible.

14. Repeat this for all four newspapers on this day

Recognising **text** as **Daily_Express**, **Daily_Mirror** or **2 other classes**

< Back to project

Daily_Express **Daily_Mirror** **The_Guardian** **The_Independent**

FURY AT PLOT TO GET NEW EU V... ANT: MY FIGHT TO SAVE MARRIA... Trump under fire from all sides as ... May's plan to pack Brexit committ...

+ Add example + Add example + Add example + Add example

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



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

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



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

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



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

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 some headlines as text you can copy to save you some typing.

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 **Daily_Express**, **Daily_Mirror** or **2 other classes**

< Back to project

Daily_Express **Daily_Mirror** **The_Guardian** **The_Independent**

FURY AT PLOT TO GET NEW EU VO... WINE IS KEY TO A LONGER LIFE HOUSE PRICES UP BY £10,000 2.37 MILLION EU MIGRANTS WOR... SLAUGHTERED ON THE STREET

ANT: MY FIGHT TO SAVE MARRIAGE SO VULGAR (EVEN FOR YOU, SIR ...) FREEZE FARES BOMB HERO 'STOLE FROM VICTIMS' MASSACRE

Trump under fire from all sides as wo... 'Racism is evil', says Trump, 48 hour... EU chiefs condemn 'fantasy' of UK ... FA under pressure to explain payme... Terror strikes Barcelona

May's plan to pack Brexit committee... Government demands tariff-free trad... Care home crisis: places must doubl... Government crackdown on universiti... Terror in Barcelona: 12 dead in van a...

+ Add example + Add example + Add example + 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've trained a machine learning model to recognise when text is Daily_Express, Daily_Mirror or 2 other classes." Below that, it says "You created the model on Monday, August 21, 2017 8:41 PM." A section titled "You've collected:" lists four categories: "5 examples of Daily_Express," "5 examples of Daily_Mirror," "5 examples of The_Guardian," and "5 examples of The_Independent." To the right, there's a note: "try testing the machine learning model below. Enter an example or 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 is 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, it says: "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 different the extra examples will make!" At the bottom, there's a text input field containing "REVOLT OVER £36BN EU DIVORCE" and a "Test" button. Below the input field, the text "Recognised as Daily_Express with 54% 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?

To answer those questions, you need to test the machine learning model. Next, you'll build a test in Scratch to do this.

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).

Daily_Express Daily_Mirror The_Guardian The_Independent
These blocks represent the labels you've created in your project, so you can use their names in your scripts.

This means you can do something like this:

ask [enter some text here] and wait
if [recognise text [answer] (label)] is [Daily_Express] then
say [I think that was Daily_Express]

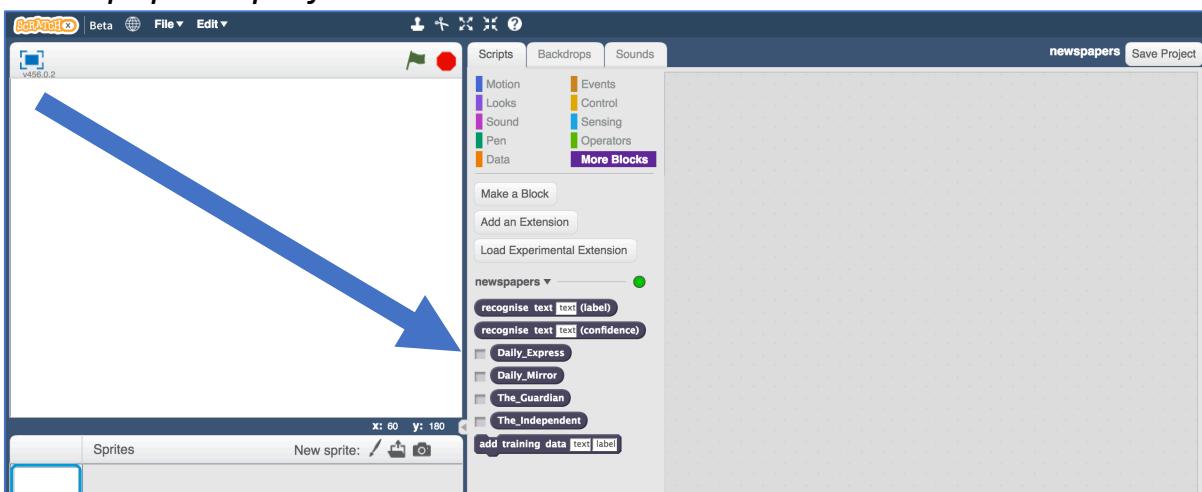
It will look something like this - except with the name of your project.

The coloured circle next to your project name tells you if your machine learning model is okay.

- **green** means your model is trained and ready to go
- **yellow** means your model needs more training
- **red** means your model is not trained at all

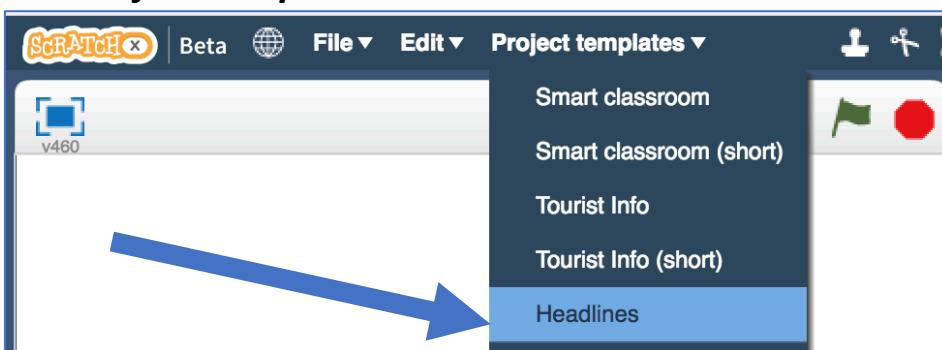
22. Click the “Open in Scratch” button to start making the test.

You should see new blocks in the “More blocks” section from your “newspapers” project.



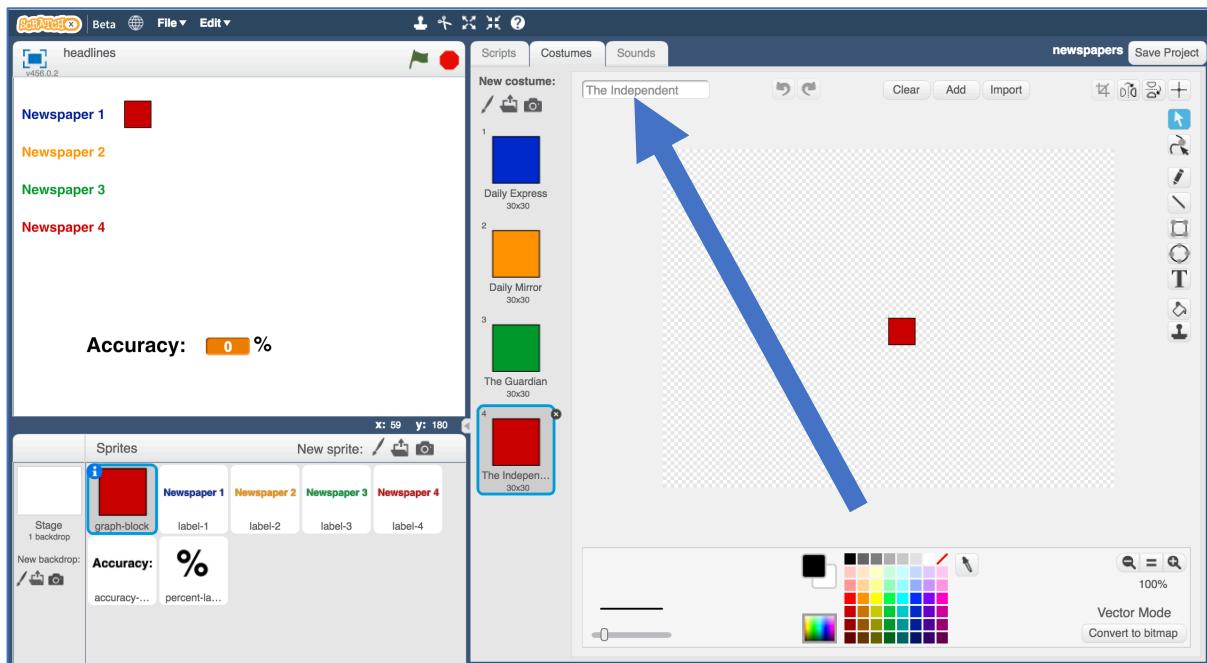
23. Open the Headlines project template

Use Project templates -> Headlines



24. Click on the **Costumes** tab of the graph-block sprite.

Rename the costumes, so instead of being called “newspaper-1”, “newspaper-2”, etc. they have the names of your four newspapers.

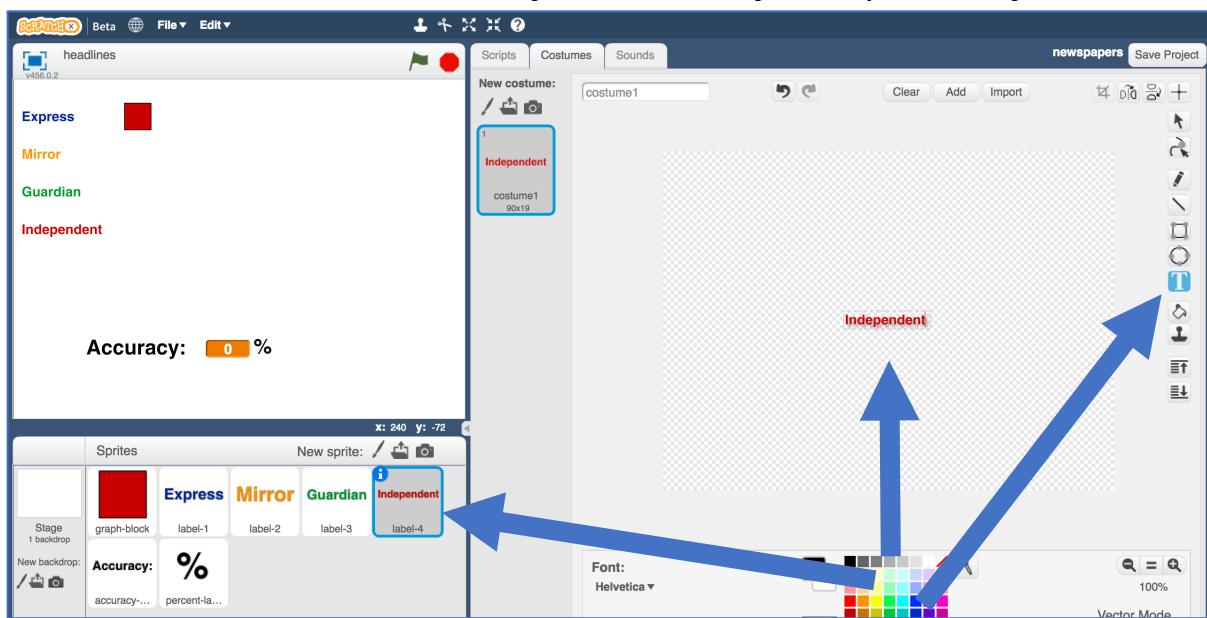


25. Update the text in the label-1, label2, label3, label4 sprites

Make sure that they match the order, and the colours, for the names you put in the graph-block sprite

Click on the costume using the T text tool to change the text.

You can use a shorter version of the names if it helps them fit better.



26. You will need **10** more front-page headlines from each newspaper.

Note: It is very important that none of these are examples you used to train the model.

Ten more examples means another two week's worth of newspapers.

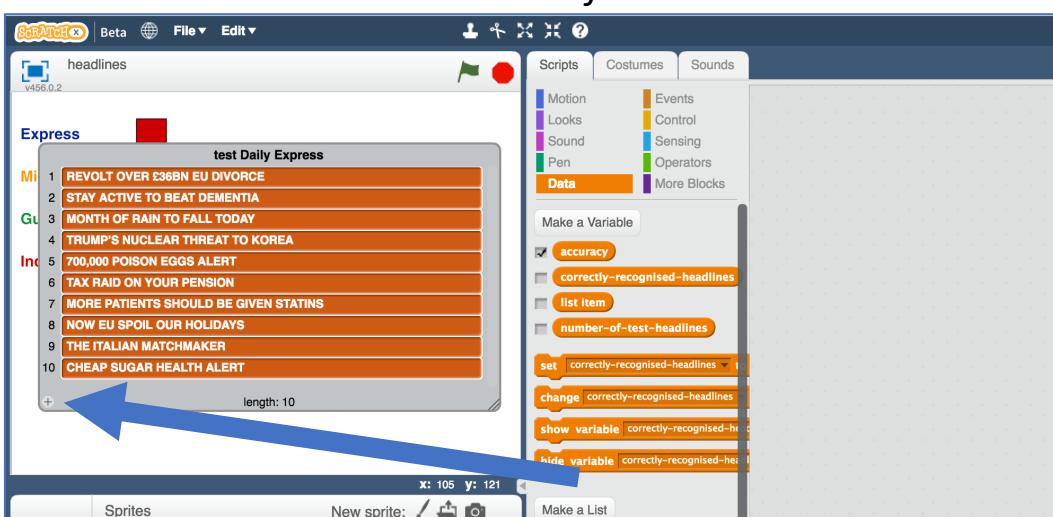
27. Click the **Data** tab, and then click the **Make a List** button

Create a list called “test Daily Express” for all sprites.

If you're using different newspapers, use the name of your first newspaper

28. Enter your 10 test headlines from that newspaper.

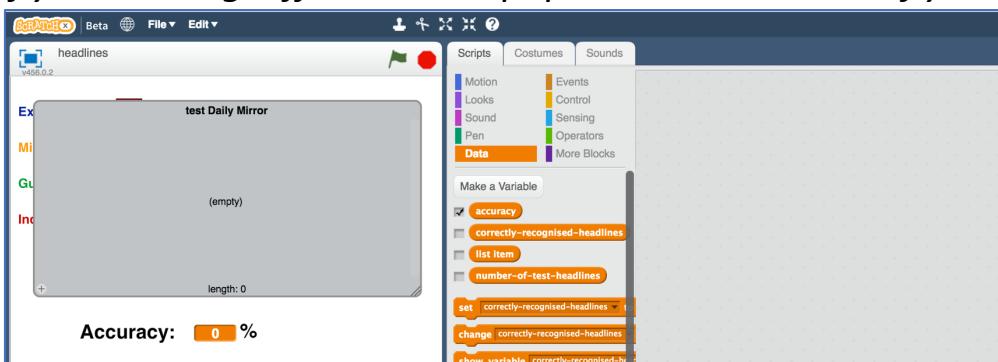
*Resize the list window to make it bigger: that'll make it easier to type into
Click the + button in the bottom left corner to add each new test headline*



29. Untick the “test Daily Express” list so that it is hidden

30. Click the **Make a List** button again, and create a second list for your second newspaper. Call it “test Daily Mirror”

If you're using different newspapers, use the name of your second paper

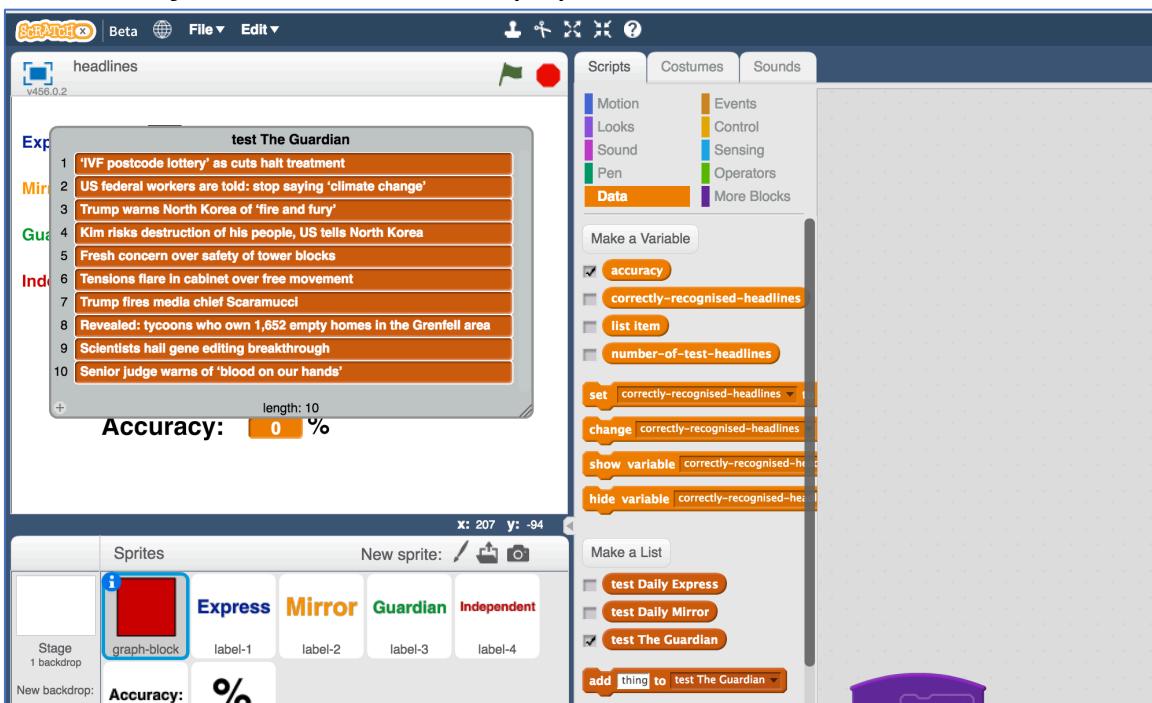


31. Type the 10 test headlines from your 2nd newspaper into this list.

32. Untick the list so that the second list is hidden like the first

33. Repeat for the third newspaper

Create a new list with a name that starts with “test”. Add the ten test headlines from the third newspaper to the list.



34. Hide the third list by unticking it

35. Repeat for the fourth newspaper

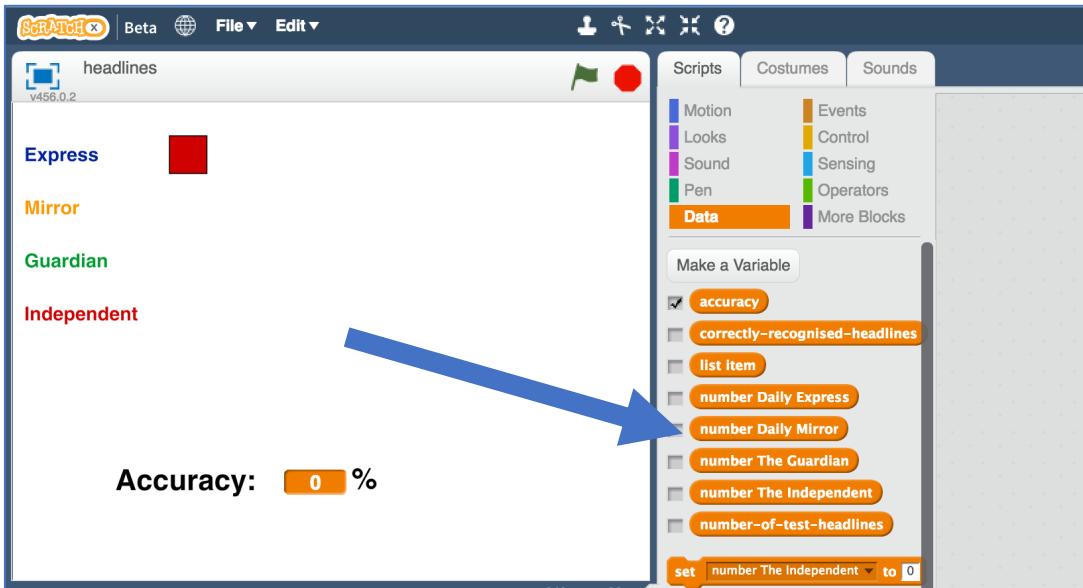
Create a new list with a name that starts with “test”. Add the ten test headlines from the fourth newspaper. Untick the list to hide it.

36. Create four new variables

Create a variable by clicking the **Make a Variable** button on the Data tab.

Create them for all sprites. Untick them so that they are hidden.

Name them “number Daily Express”, “number Daily Mirror”, etc. (using the names of your four newspapers).



37. Create four new lists

Create a list by clicking the **Make a List** button on the Data tab.

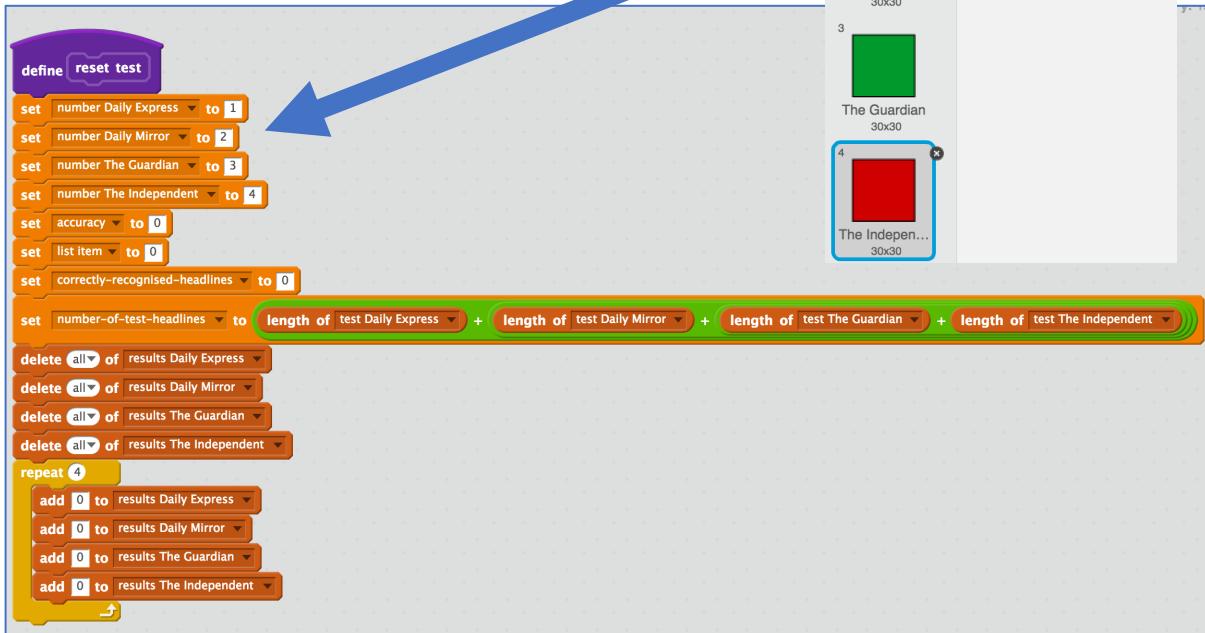
Create them for all sprites. Untick them so that they are hidden.

Name them “results Daily Express”, “results Daily Mirror”, etc. (using the names of your four newspapers).



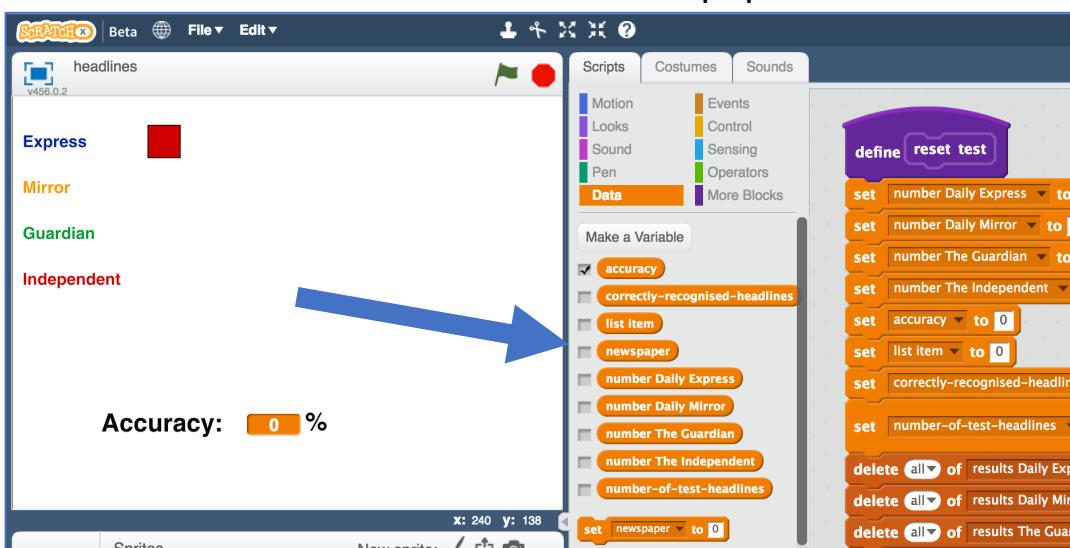
38. Update the script for the **reset test** block to reset the data

The numbers for the first four blocks (1,2,3,4) should match the costume numbers for the newspapers in graph-block. This also means they should be in the same order as the labels on the stage.

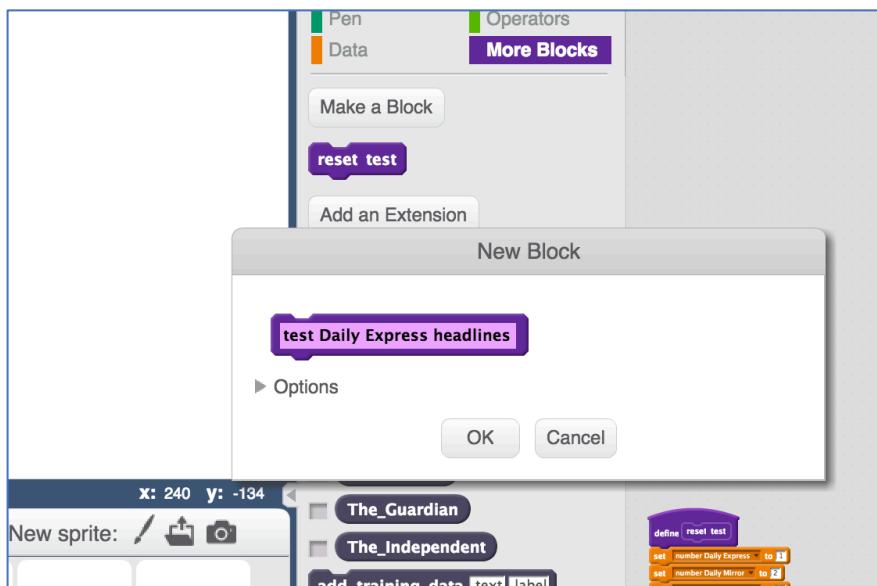


The final '4' in the repeat block is because there are four newspapers.

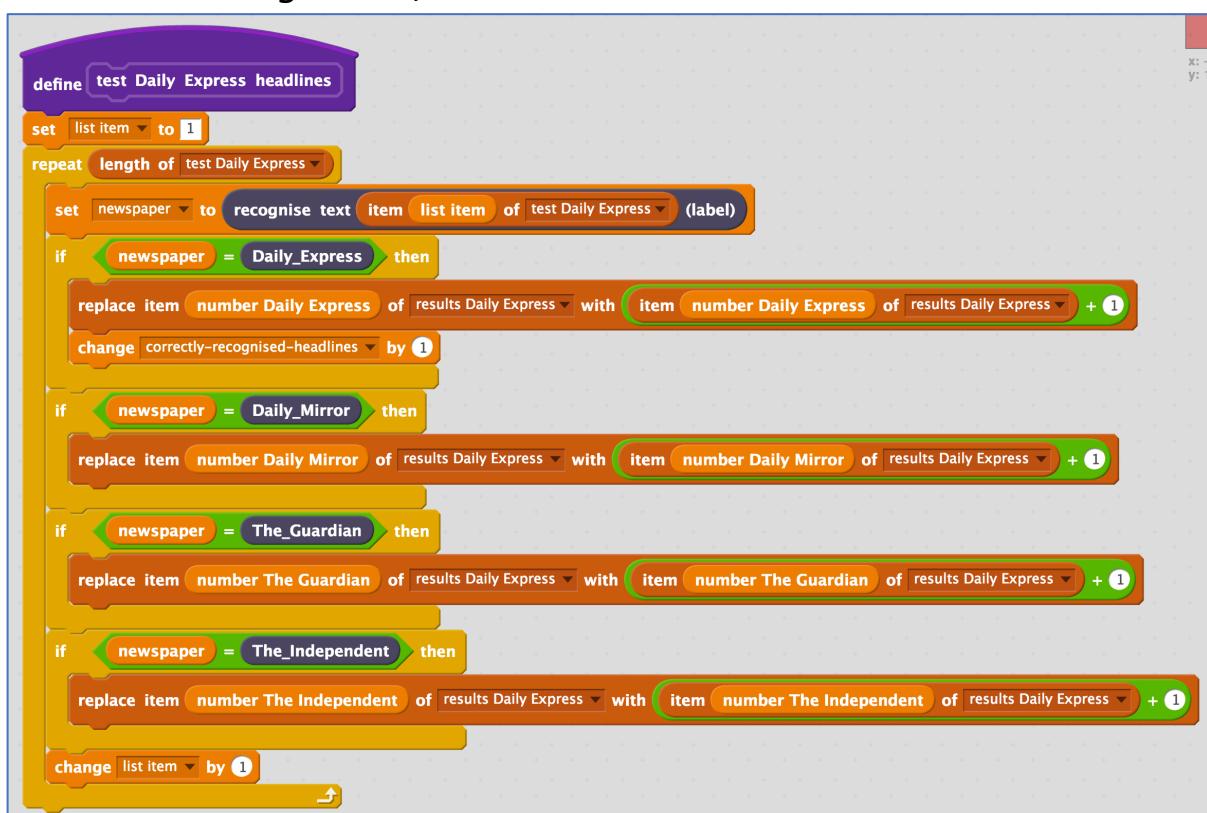
39. Create a new variable called “newspaper”. Untick it to hide it.



- 40.** Create a new block to test your first newspaper
*Click on the graph-block sprite. Click the **Scripts** tab.
 Click the purple **More Blocks** item. Click the **Make a Block** button.
 Name it “test Daily Express headlines”*



- 41.** Enter the following script into the new custom block
It will run your test headlines from the Daily Express through your machine learning model, and count the results it returns.



42. Create a new block to test the headlines from your second newspaper

Click on the graph-block sprite.

*Click on the **Scripts** tab. Click the purple **More Blocks** item.*

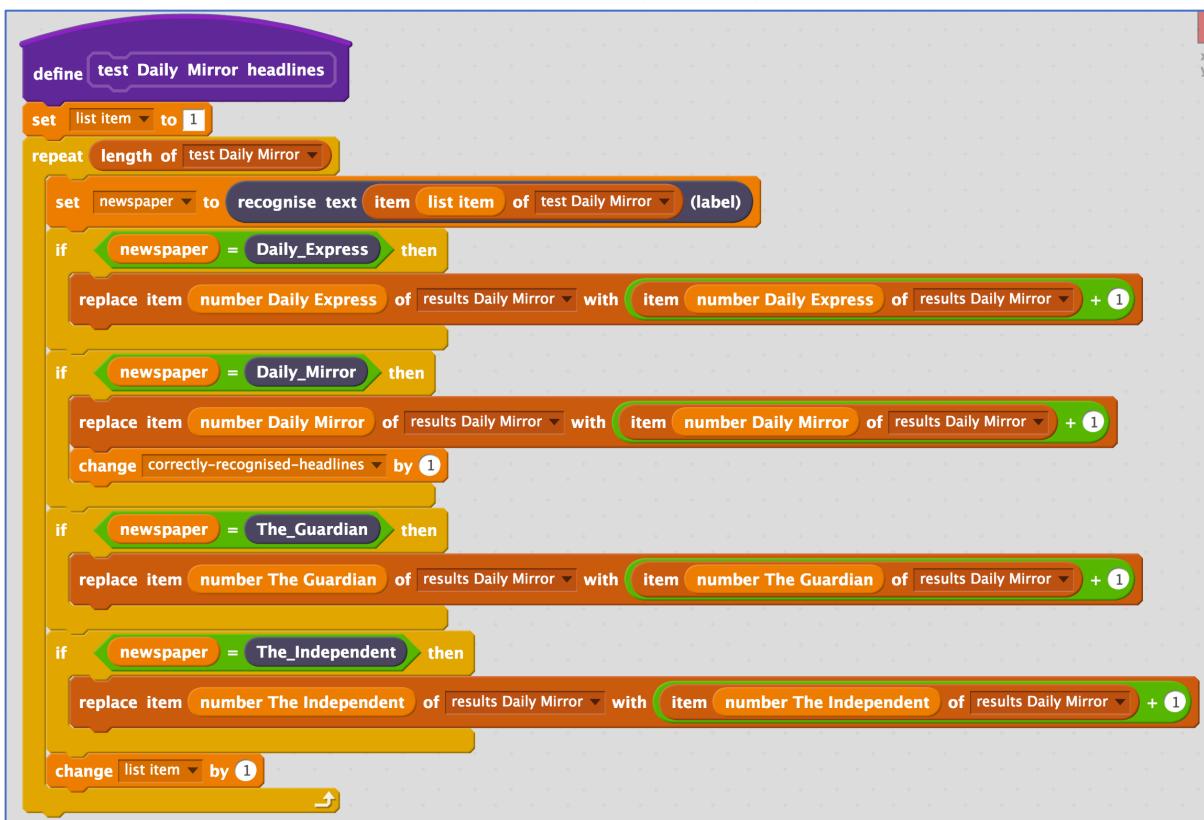
*Click the **Make a Block** button. Name it “test Daily Mirror headlines”*

43. Add the following script to the custom block.

Notice how similar it is to the last one.

You can save yourself some time by duplicating the contents of your last script (right-click on the contents and click on Duplicate).

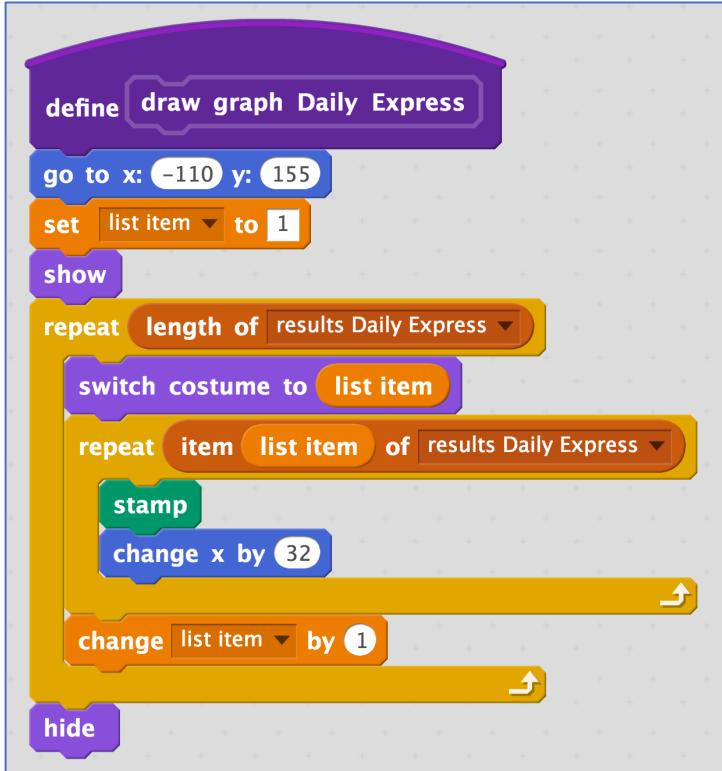
If you do this, make sure you update it to use the Daily Mirror lists and also make sure you move the correctly-recognised-headlines block.



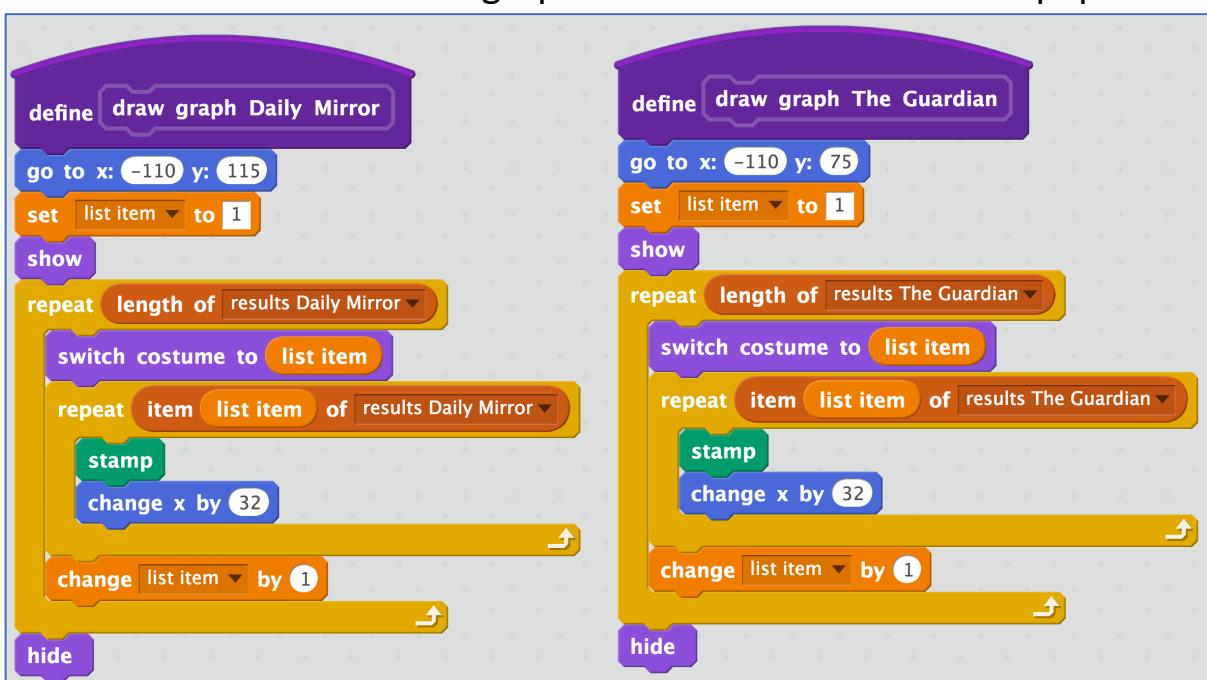
44. Create two more custom blocks to test the third and fourth newspapers.

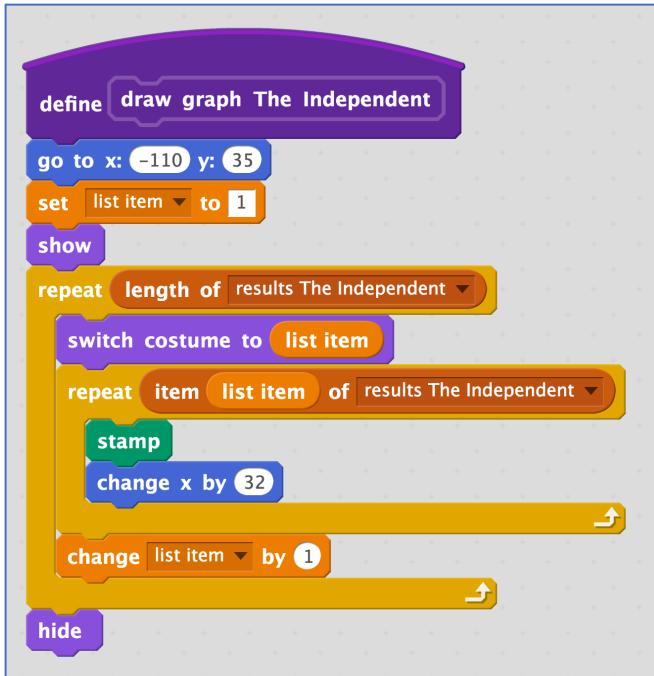


- 45.** Create a new block to draw a graph of the results from the test of the headlines from the first newspaper
Click on the graph-block sprite.
*Click on the **Scripts** tab. Click the purple **More Blocks** item.*
*Click the **Make a Block** button. Name it “draw graph Daily Express”*



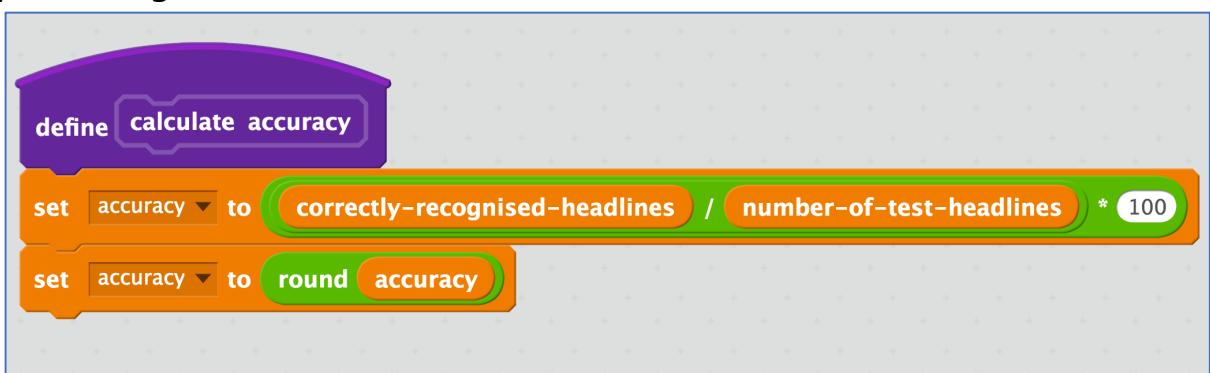
- 46.** Create blocks to draw graphs for the other three newspapers.





47. Create a new block to calculate how accurate your machine learning model is

Create this block in the graph-block sprite along with all the other scripts. This calculates how many headlines are correctly recognised as a percentage.

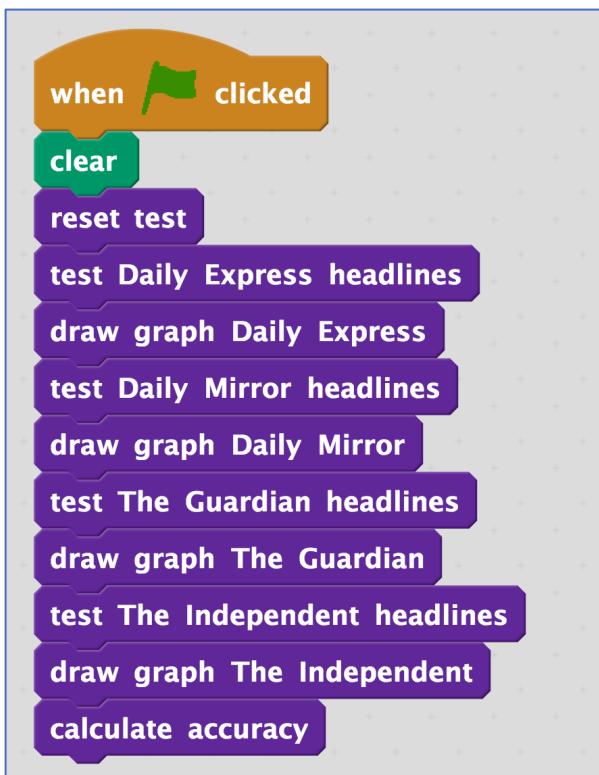


48. Time to put it all together!

Create this script in the graph-block sprite.

It will test your machine learning model using all of your test headlines from all the papers.

It will draw a graph to show you the results, and calculate the overall accuracy to give you a score.

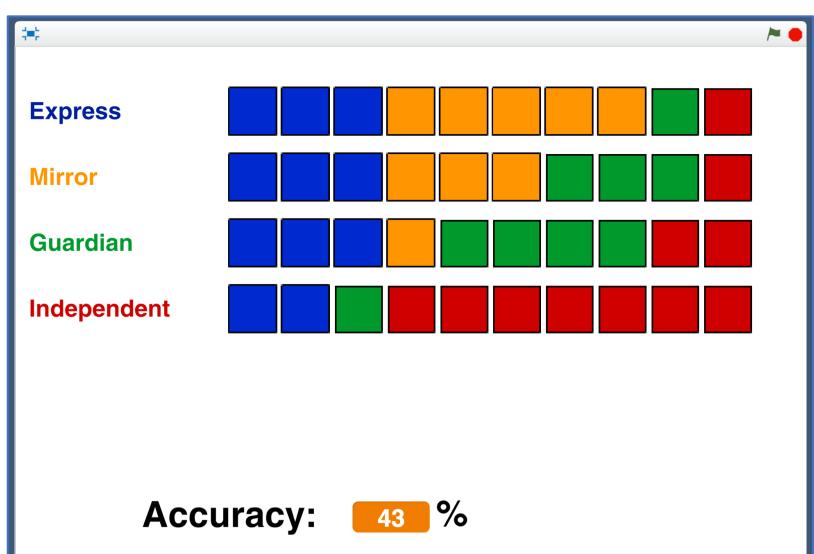


49. Run the test

Click the green flag, and watch it run.

Each row is made up of 10 blocks.

The 10 blocks represent the 10 test headlines from a newspaper.



Independent



I collected 10 headlines from front-pages of *The Independent*.

My machine learning model “thought” 2 of them were from the *Daily Express* (blue).

It “thought” that 1 of them was from *The Guardian* (green).

It correctly recognised the other 7 as being headlines from *The Independent*.

That was pretty good!

Express



I collected 10 headlines from front-pages of the *Daily Express*.

The machine learning model correctly recognised 3 of them.

It thought 5 were from the *Daily Mirror*, 1 was from *The Guardian*, and it thought 1 looked like a headline from *The Independent*.

Not so good. It looks like it's having trouble telling the difference between *Daily Express* and *Daily Mirror* headlines.

Accuracy: 43 %

Overall, my machine learning model correctly recognised 43% of the test headlines.

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 test for your model in Scratch. This test is like an exam for the model, testing how well it does on headlines it hasn't seen before.

It marks how well the machine learning model does. It draws graphs to help you understand what it is good at, and what it's not good at.

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

50. Save your Scratch project

Click on **File -> Save Project**

51. Close the Scratch web browser window

It's important you do this in between giving your machine learning model additional training.

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

53. Click the Learn button

54. Repeat steps 13-15. Collect additional headlines from another couple of weeks of newspapers.

Make sure you don't use any headlines that are in your Scratch test.

If you show it an example that you'll use to test the model, that won't be a fair test.

You should now have fifteen example headlines from each newspaper.

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

Recognising **text** as **Daily_Express, Daily_Mirror or 2 other classes**

< Back to project

Daily_Express

- 2.37 MILLION EU MIGRANTS WORK IN BRITAIN
- 91F BLAST OF HEAT ON WAY
- BREAKTHROUGH ON ALZHEIMER'S
- BRITAIN'S WIDE OPEN TO ILLEGAL MIGRANTS
- DRINK COFFEE TO LIVE LONGER
- EU CAN WHISTLE FOR OUR MONEY
- EU IN CRISIS OVER BOAT MIGRANTS
- EU TOLD: YOU CAN'T FRIGHTEN BRITAIN
- FURY AT PLOT TO GET NEW EU VOTE
- HOUSE PRICES UP BY £10,000
- NEW OPERATIONS CRISIS HITS NHS
- NEW WAY TO HALT DIABETES
- SLAUGHTERED ON THE STREET
- STATINS: NEW HEALTH ALERT
- WINE IS KEY TO A LONGER LIFE

Daily_Mirror

- 'FRAUDSTER WHO PUT 30 TONS OF HORSE ME...
- ANT: MY FIGHT TO SAVE MARRIAGE
- BLUNDER DOCS STERILISED ME AS I GAVE BIRTH
- BOMB HERO 'STOLE FROM VICTIMS'
- BRITISH NAVY... SWEDISH STEEL
- CAM: IT'S SELFISH TO GIVE OUR HEROES PAY R...
- DEMENTIA TIMEBOMB
- FREEZE FARES
- GIVE HEROES A DECENT RISE
- MASSACRE
- MPs TO VOTE ON MAX'S LAW
- OUR CHARLIE IS STILL FIGHTING & SO ARE WE
- SO VULGAR (EVEN FOR YOU, SIR GREEDY)
- TORY MP'S N-WORD SHAME
- TRAGIC REFUGEES CRISIS

The_Guardian

- Delay leaving single market - business chiefs
- EU chiefs condemn 'fantasy' of UK plan to mirror t...
- FA under pressure to explain payment after bullyin...
- Labour threat to defeat May on Brexit bill
- May appeals to Labour for policy ideas
- May under fire as teacher pay rise held at 1%
- PM's EU repeal bill dismissed as 'power grab'
- Revealed: big tobacco's dirty battle for the African ...
- Safety alert on 1,300 'failing' nursing homes
- Scale of public sector salary cuts revealed
- Student debt to hit £57,000 for the poorest
- Terror strikes Barcelona
- Trump under fire from all sides as woman dies at fa...
- UK ditches its 'cake and eat it' Brexit stance
- 'Racism is evil', says Trump, 48 hours after fatal vio...

The_Independent

- Brexit IT 'horror' puts £34bn border tax at risk
- Brexit minister caught in 'tear down EU' rant
- Care home crisis: places must double to meet soar...
- Corbyn's summer tour
- End of austerity? May softens on pay freeze
- Government crackdown on universities handing ou...
- Government demands tariff-free trade to avert bor...
- Grenfell residents face deportation in a year
- May abandons climate change priority at Q20
- May's plan to pack Brexit committee with Tories
- Pay cap 'dead in water' as firefighters land new deal
- Teachers furious as pay is capped at 1 per cent
- Terror in Barcelona: 12 dead in van attack
- The devastation: Mosul finally liberated from Isis
- 'Ripe for abuse': Brexit bill lets ministers scrap righ...

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

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

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

The screenshot shows a web page titled "Machine learning models". 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, there is a link "< Back to project". The main content area is divided into two sections: "What have you done?" and "What's next?".

What have you done?

You've collected examples of text for a computer to use to recognise when text is Daily_Express, Daily_Mirror or 2 other classes.

You've collected:

- 15 examples of Daily_Express,
- 15 examples of Daily_Mirror,
- 15 examples of The_Guardian,
- 15 examples of The_Independent

What's next?

Ready to start the computer's training?

Click the button below to start training a machine learning model using the examples you've collected so far.

(Or go back to the Train page if you want to collect some more examples first.)

57. Click the “< Back to project” link. Click the **Scratch** button.

58. Click the **Open in Scratch** button.

59. Load your test Scratch project

Click on File -> Load Project

60. Test your new model by clicking the Green Flag.

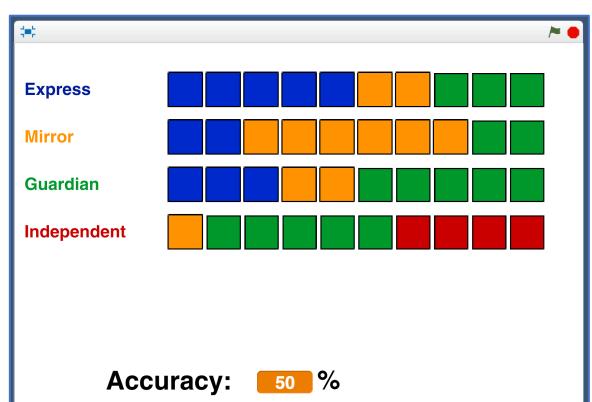
Compare the results with the results you got before.

My results after fifteen examples were a little better overall.

It was a little better at recognising

Express and Guardian headlines, and much better at recognising Daily Mirror headlines.

But it got a lot worse at recognising Independent headlines – getting wrong several that it was recognising correctly before!



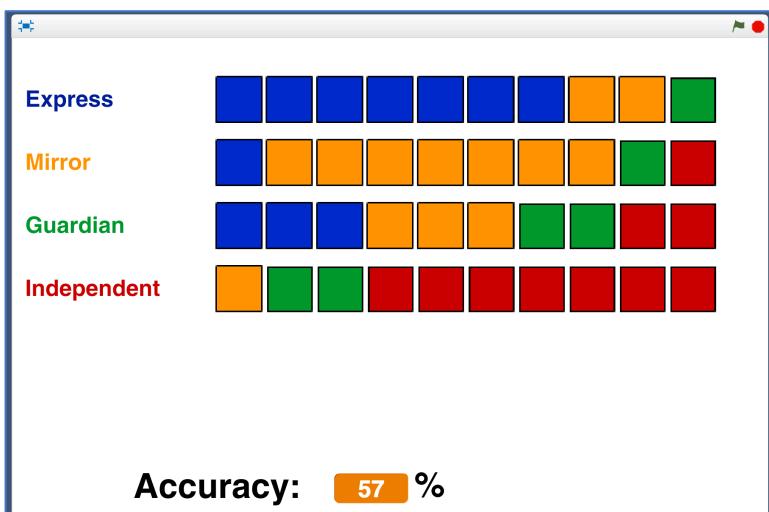
61. Repeat steps 51 – 60 again.

Close Scratch. Go back to the Training tool and collect more examples.

Train a new machine learning model.

Go back into Scratch and run your test to see how your new model is doing.

The screenshot shows a web application interface for training machine learning models. 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 is a title 'Machine learning models'. Underneath the title, there's a link '< Back to project'. The main content area is divided into two sections: 'What have you done?' on the left and 'What's next?' on the right. The 'What have you done?' section contains text about training a model to recognize news sources like Daily_Express, Daily_Mirror, and The_Guardian, along with a list of collected examples. The 'What's next?' section provides instructions for testing the model and offers a 'Train' button. The overall layout is clean and organized, typical of an educational web application.



A little better overall.

It got worse at recognising Guardian headlines, though.

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 repeated this test 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 a bigger test

Try running the test with 20 headlines from each newspaper instead of 10. What difference does that make?

Use the confidence block in Scratch

Think about how you could include the confidence that your machine learning model has in its predictions. Could your test treat predictions with high confidence differently from predictions that have a low confidence?