# Headline testing

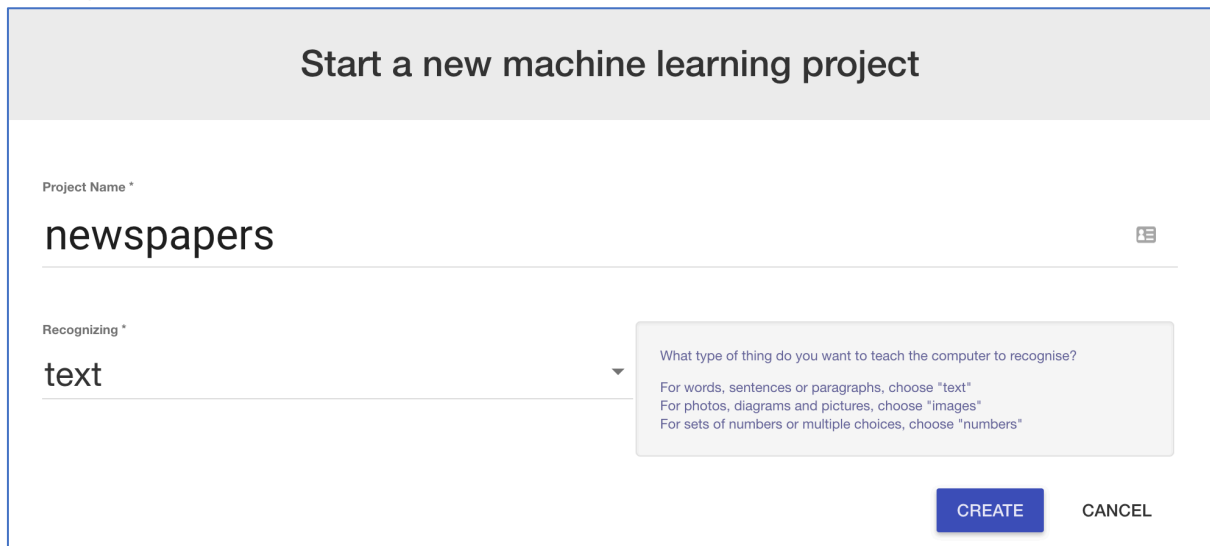In this project, you will train a computer to recognise headlines from different newspapers.

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.
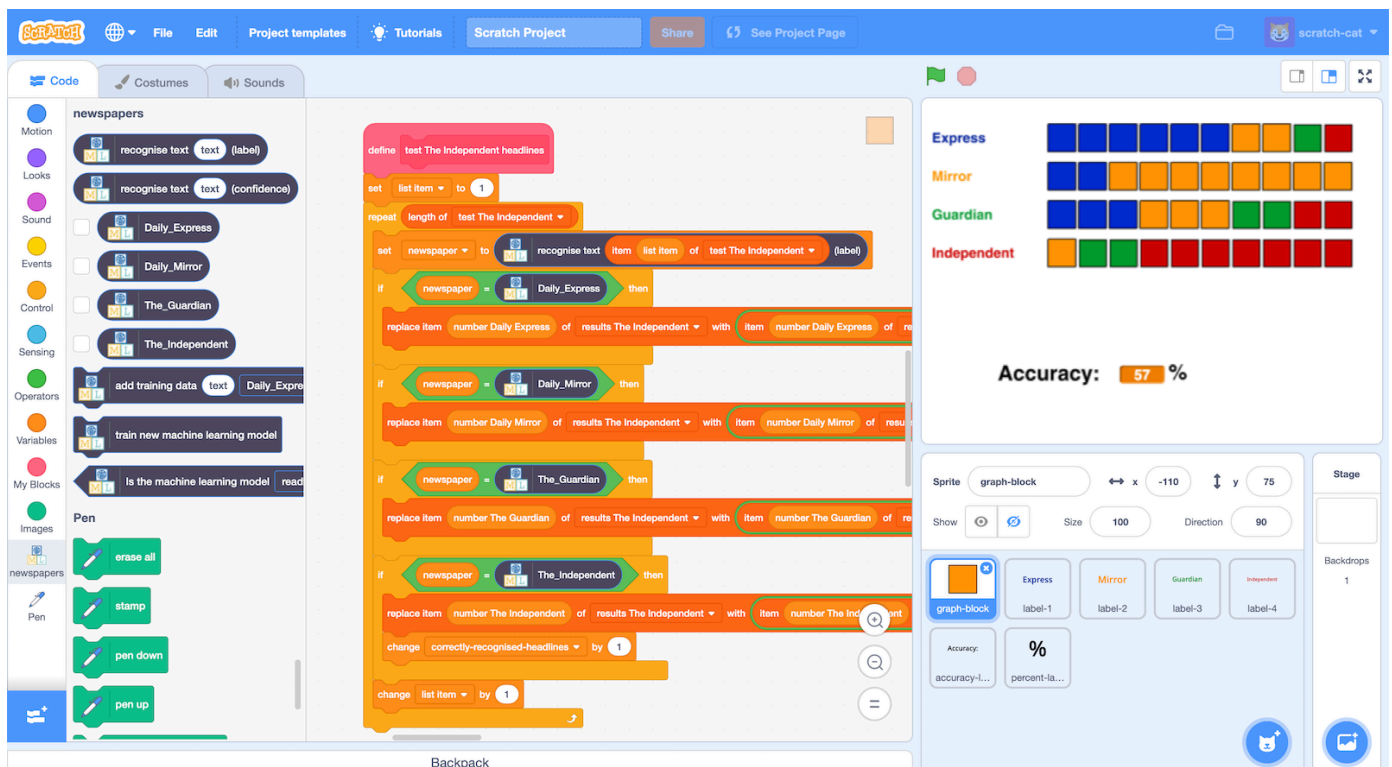
1. Go to https://machinelearningforkids.co.uk/ in a web browser

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.

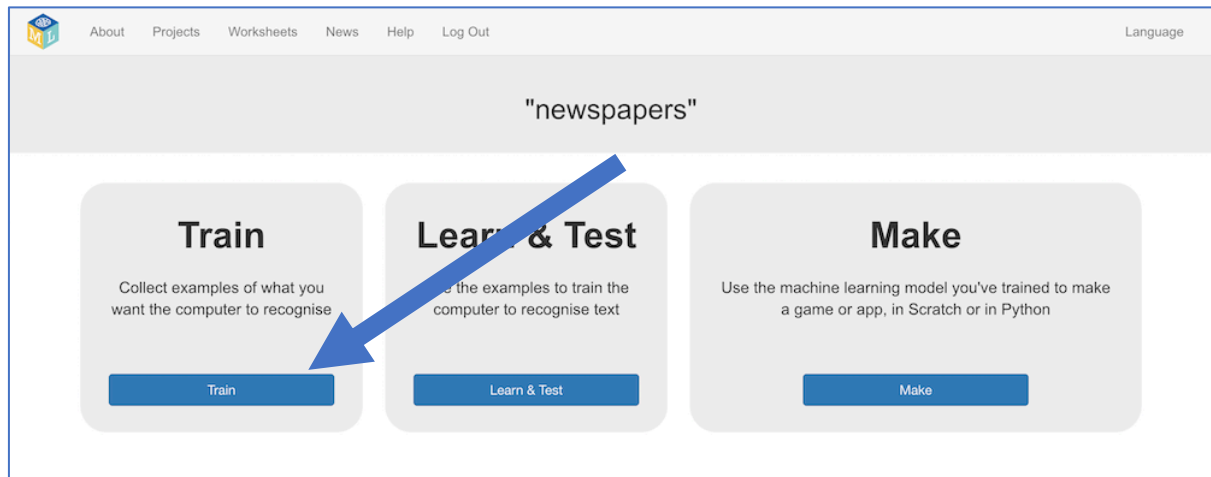**6.** Name your project "newspapers" and set it to learn how to recognise "**text**". Click "**Create**"

7. You should now see "newspapers" in your projects list. Click on it.
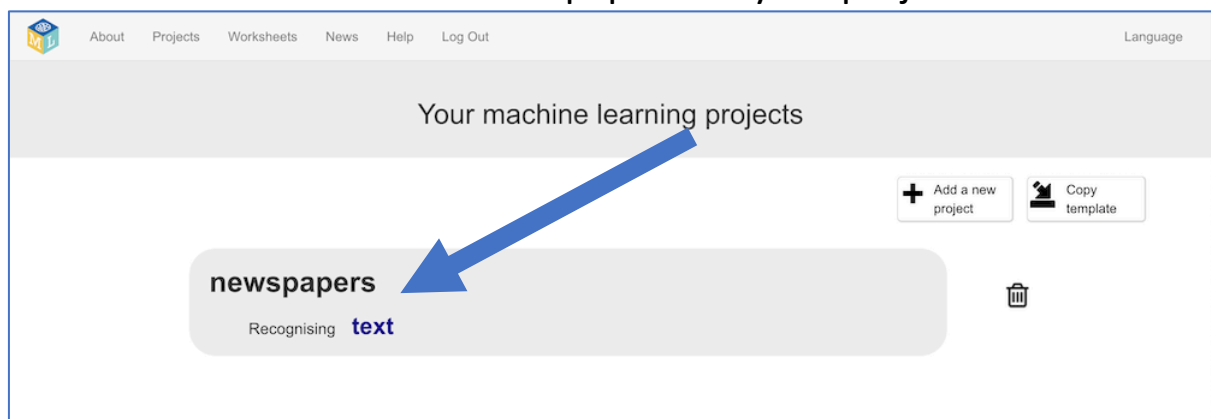
8. Click the "**Train**" button

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 Express, Daily Mirror, The Guardian, The Independent*
and getting the headlines from:
> [*https://www.thepaperboy.com/uk/front-pages.cfm*](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.
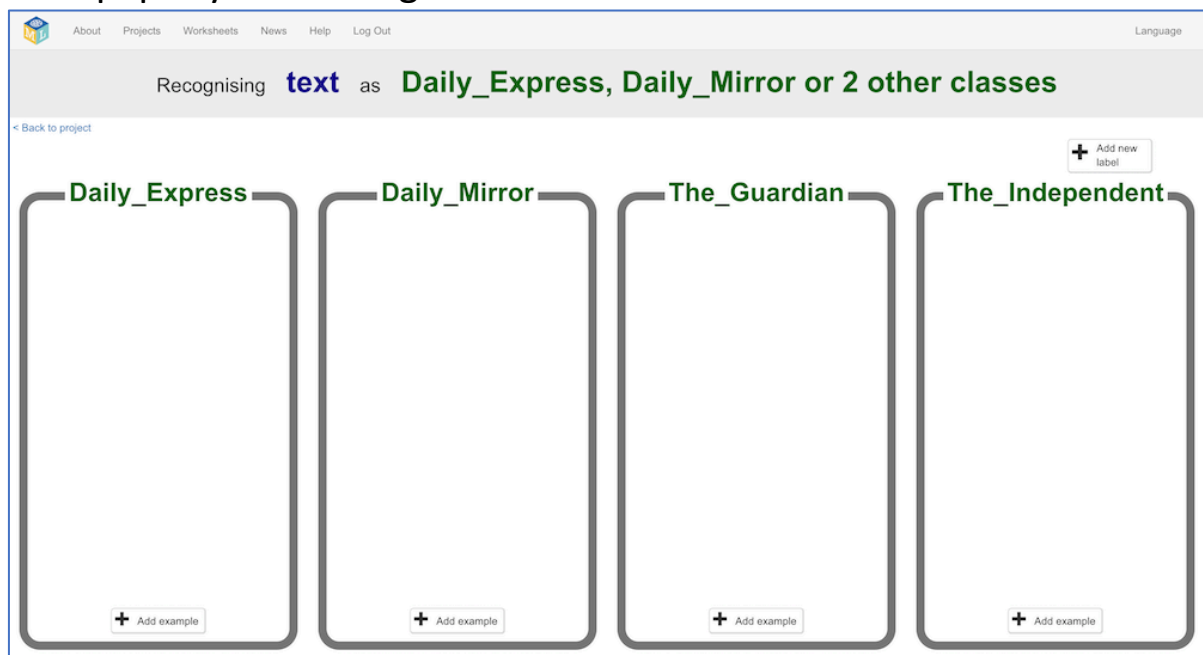
11. In another web browser window, find newspaper front pages.
Arrange the windows so your training buckets are next to the book site.
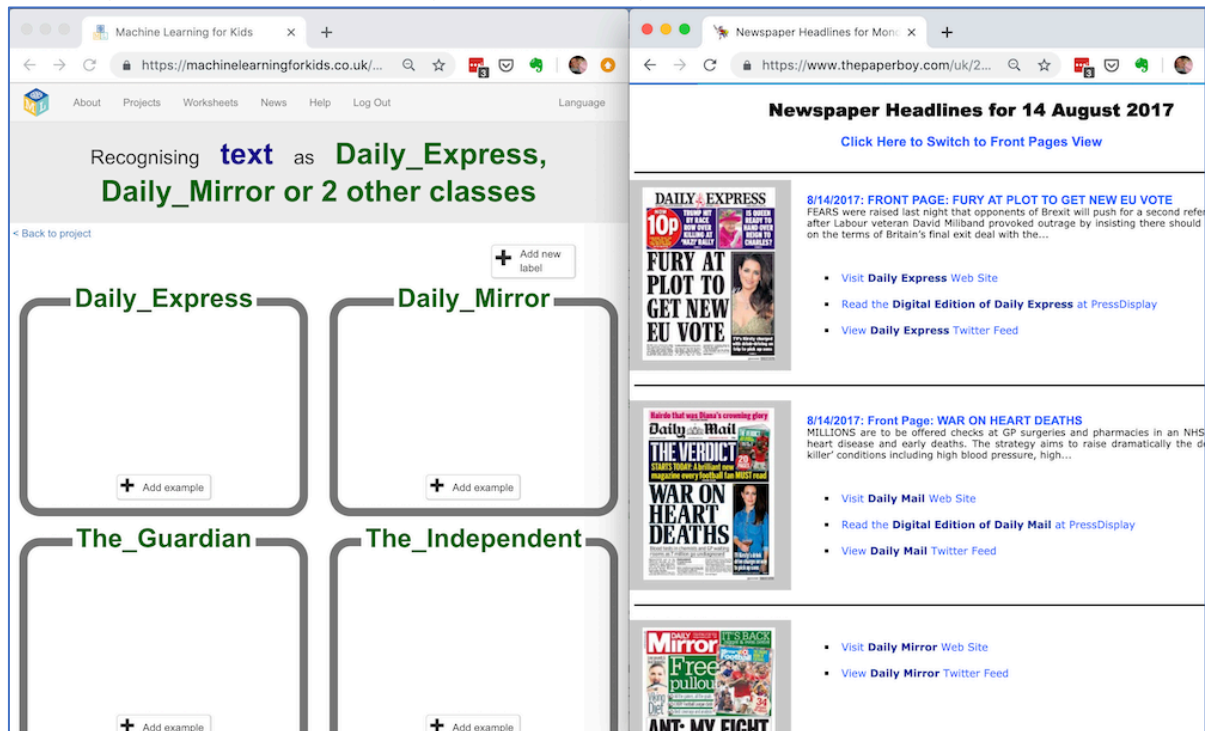
12. Agree on some rules for the examples of headlines that you'll collect, so that the training is consistent.
For the rest of this worksheet, I'll be using:
* Newspapers from weekdays only (Mon-Fri) not weekend editions
* Only the biggest main headline from each newspaper

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

14.   Repeat this for all four newspapers on this day

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.

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

You've trained a machine learning model to recognise when text is Daily_Express, Daily_Mirror or 2 other classes.

You created the model on Monday, August 21, 2017 8:41 PM.

You've collected:
- 5 examples of Daily_Express,
- 5 examples of Daily_Mirror,
- 5 examples of The_Guardian,
- 5 examples of The_Independent

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.

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!

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!

Try putting in some text to see how it is recognised based on the training.

| REVOLT OVER £36BN EU DIVORCE | | Test |

Recognised as **Daily_Express**
with 54% confidence

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.

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

23. Click the "**Open in Scratch 3**" button
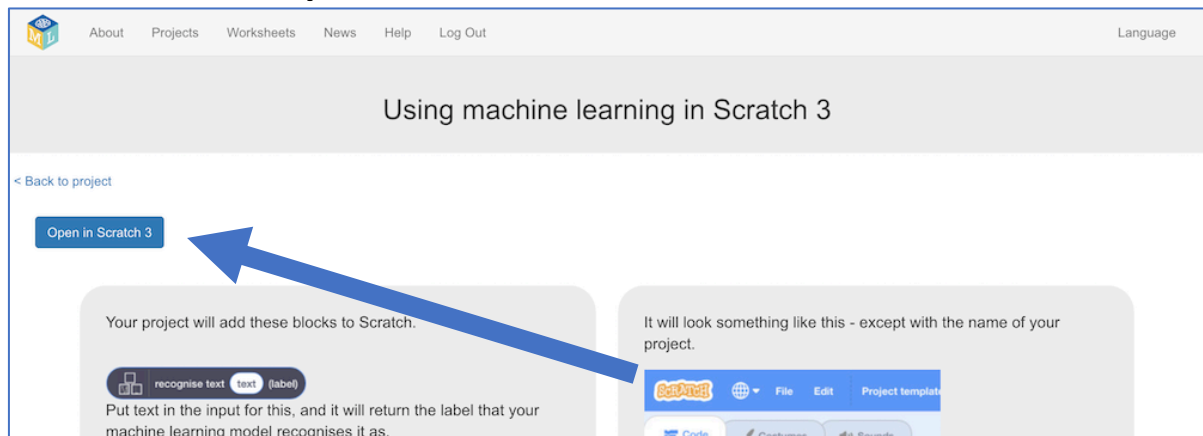
24. *You should see blocks from your project at the bottom of the list.*
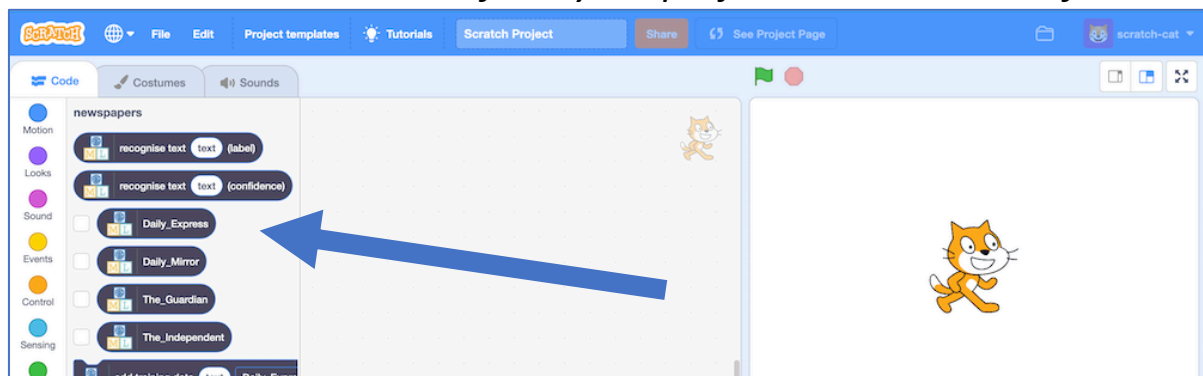
25. Open the Headlines project template
*Click on **Project templates** and then click on **Headlines***
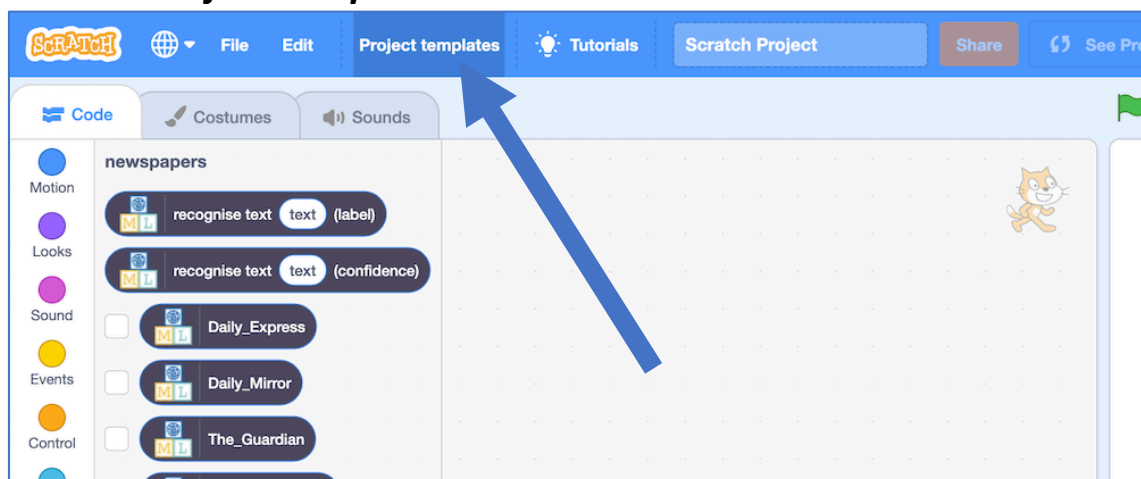
26. Click on the **Costumes** tab of the graph-block sprite.
Rename the costumes, so instead of being called "newspaper-1",
"newspaper-2", etc. they have the names of your four newspapers.
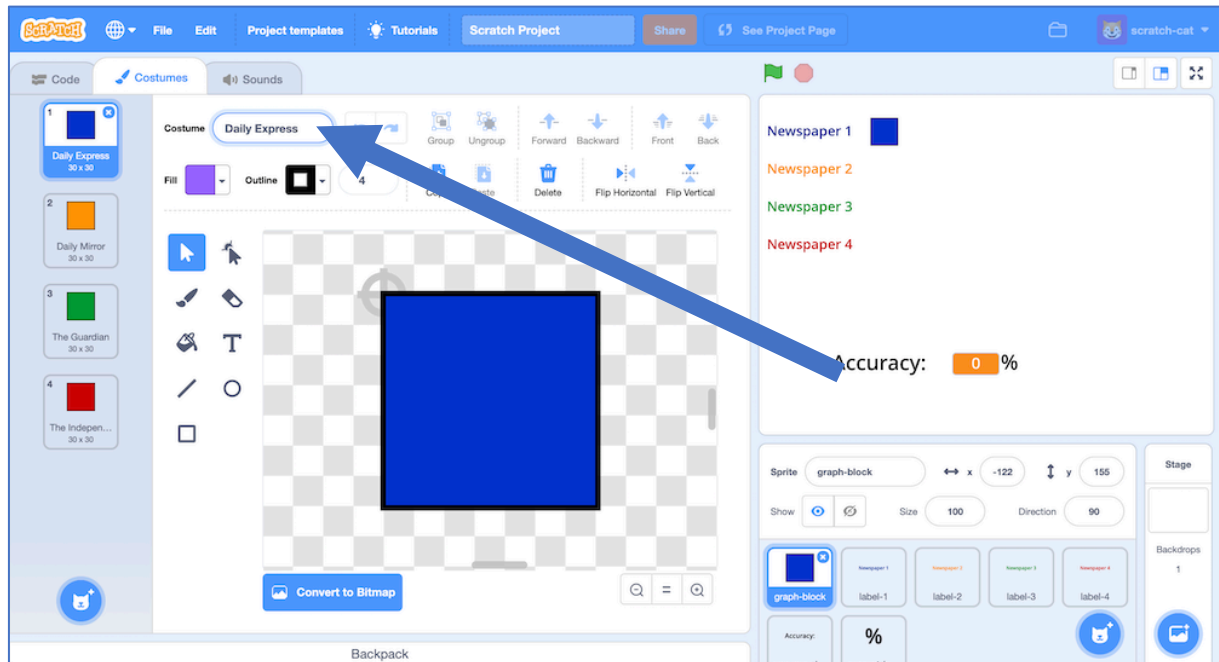
27. Update the text in the label-1, label2, label-3, label4 sprites
Make sure that they match the order, and the colours, for the names you put in the graph-block sprite
Click on the costume using the T text tool to change the text.
You can use a shorter version of the names if it helps them fit better.
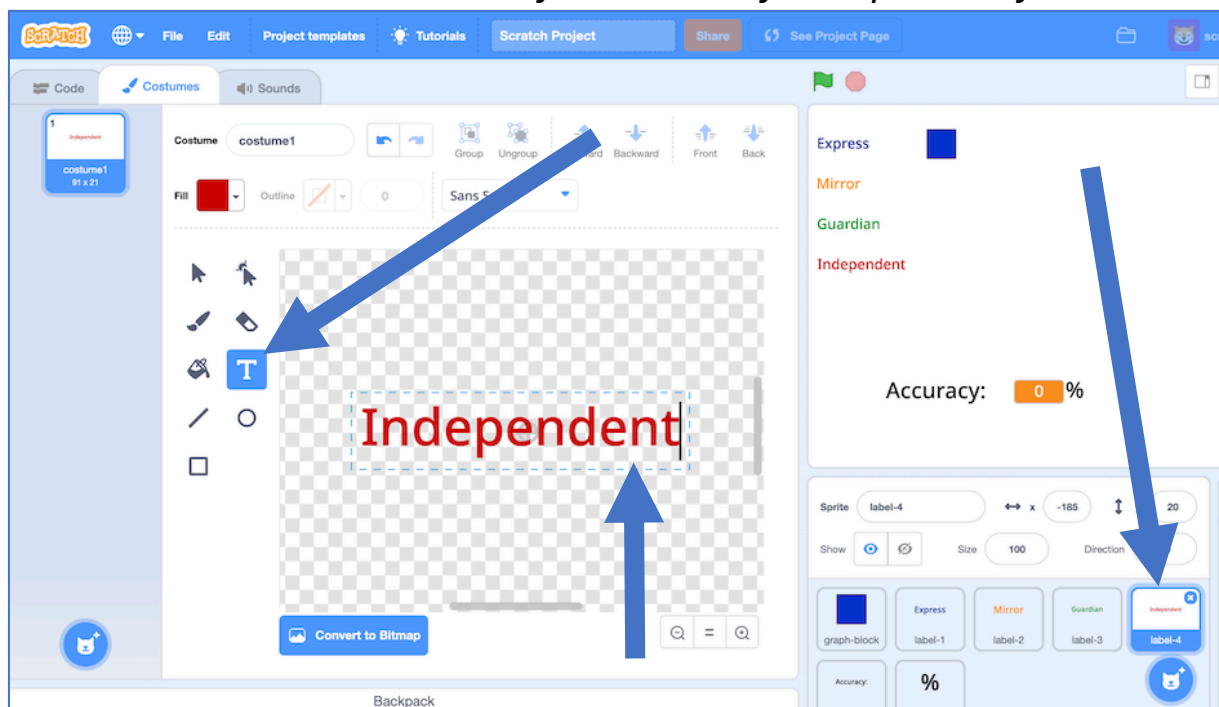
28.   Click on the **Code** tab

29.   You will need **10** more front-page headlines from each newspaper.
Note: It is very important that none are examples you used for training.
Ten more examples means another two weeks of newspapers.

30.   Click the **Make a List** button in the **Variables** section
Create a list called "test Daily Express" for all sprites.
If you're using different newspapers, use the name of your first newspaper

31.   Enter your 10 test headlines from that newspaper.
Resize the list using the = button to make it easier to type into
Click the + button in the bottom left corner to add each new test headline
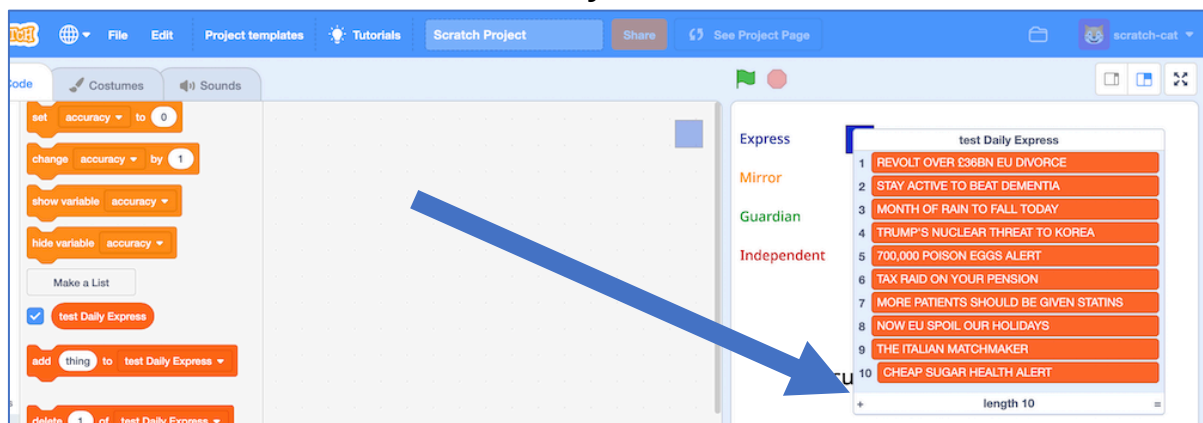
32.   Untick the "test Daily Express" list so that it is hidden

33.   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

Last updated: 21 July 2019
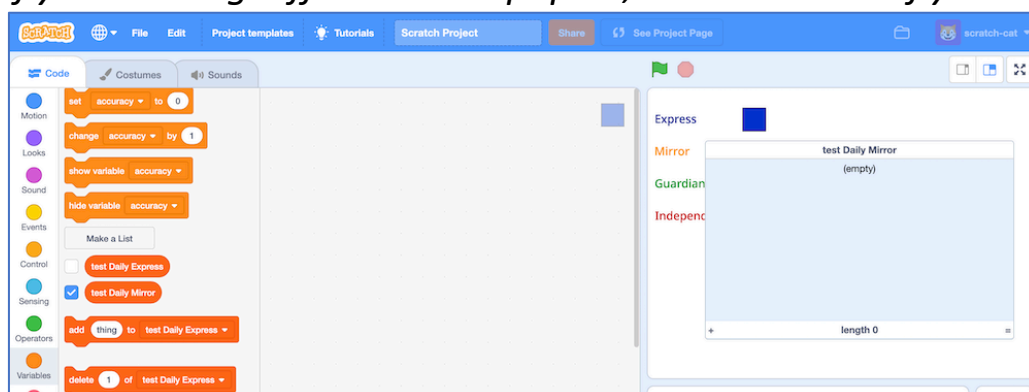
34. Type the 10 test headlines from your 2nd newspaper into this list.

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

36. 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.
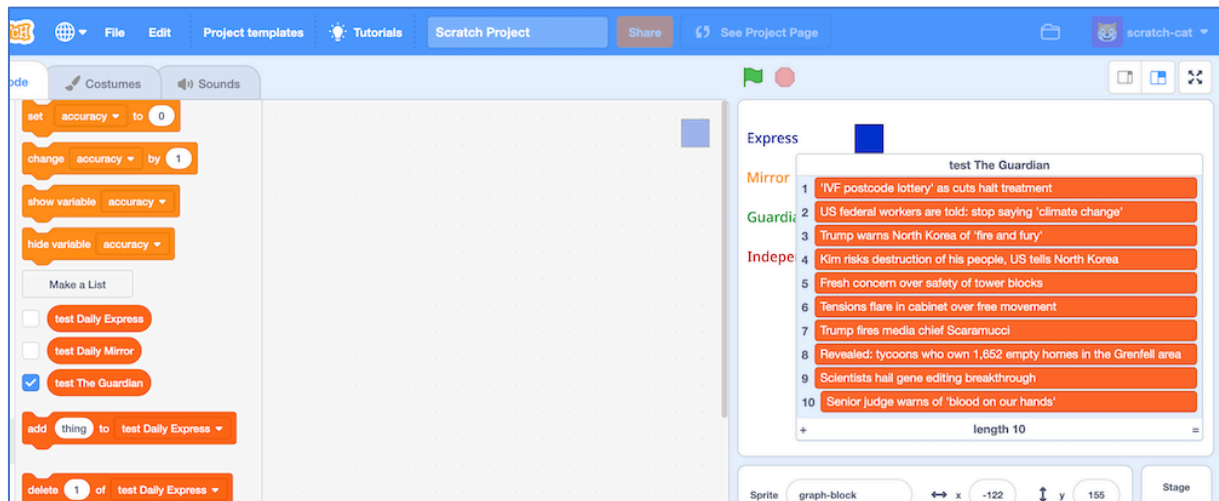
37. Hide the third list by unticking it

38. 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.

Last updated: 21 July 2019
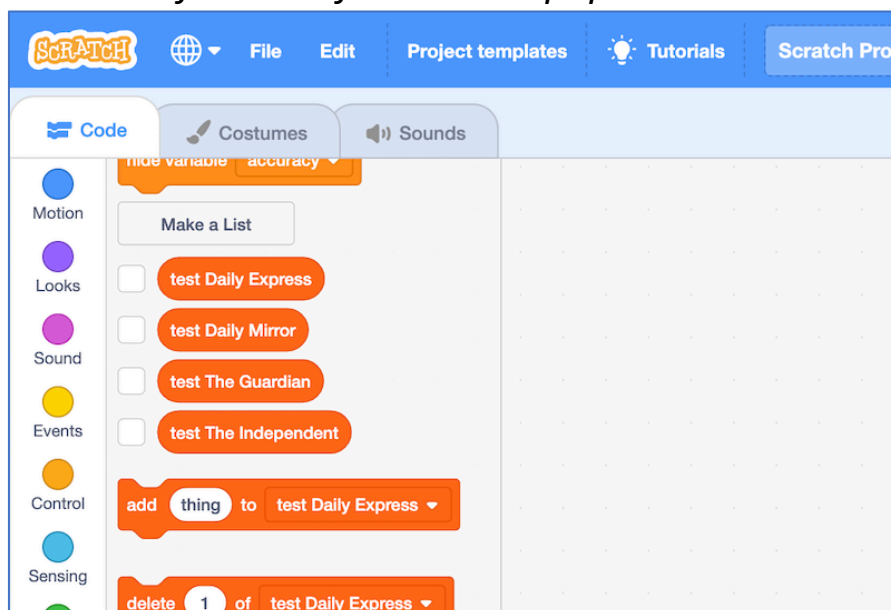
39. Create four new variables

*Create a variable by clicking the **Make a Variable** button.*
*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).
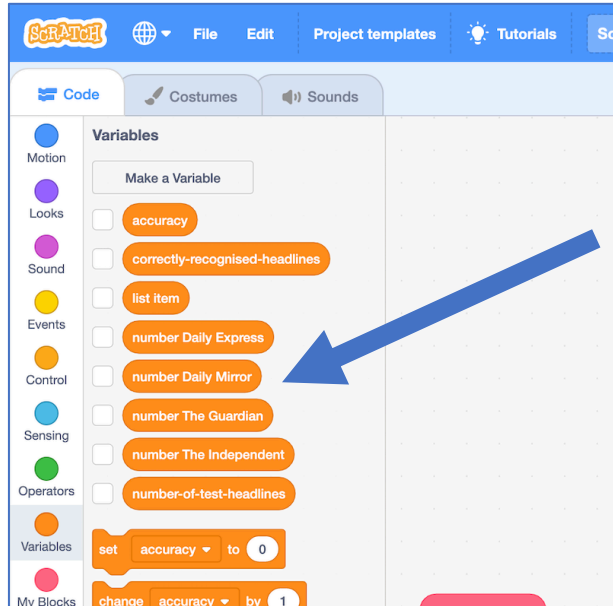
40. Create four new lists

*Create a list by clicking the **Make a List** button.*
*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).
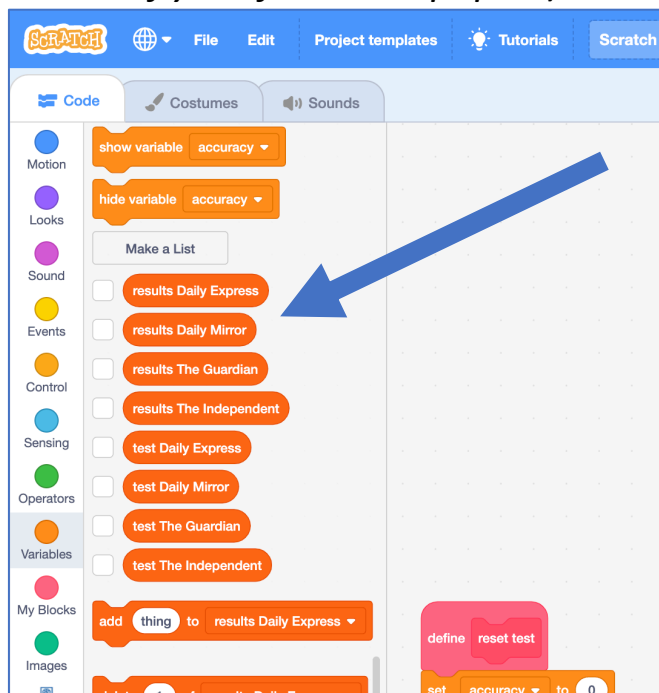
41.   Update the script for the **reset test** block to reset the data

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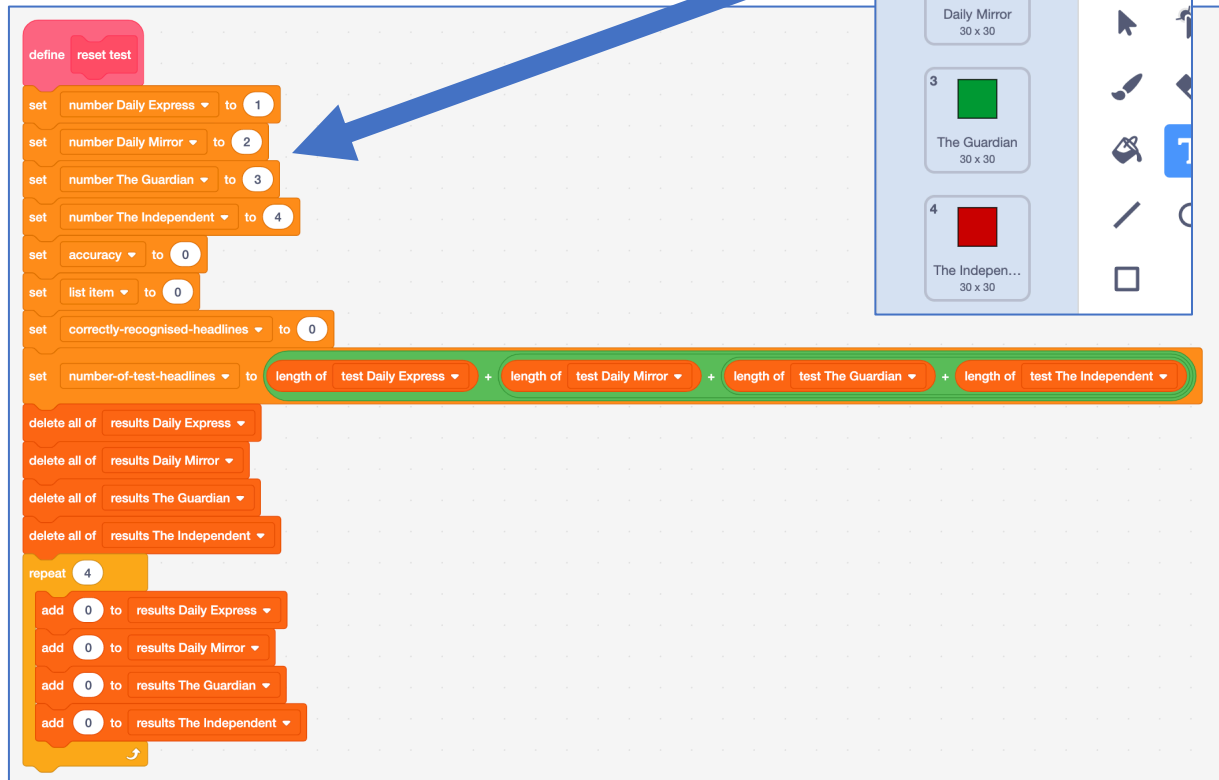
The final '4' in the repeat block is because there are four newspapers.

42.   Create a new variable called "newspaper". Untick it to hide it.
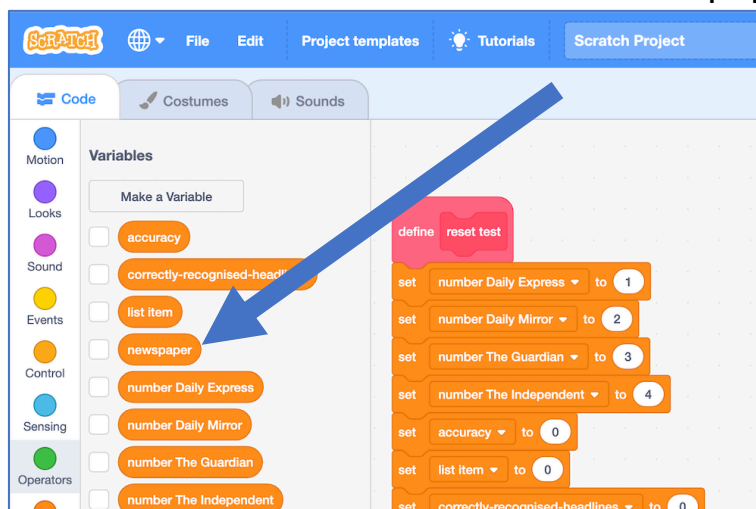
43. Create a new code block to test your first newspaper
*Click on the graph-block sprite. Click the **Code** tab.*
*Click the **Make a Block** button.*
Name it "test Daily Express headlines" (replace the name with your newspaper)
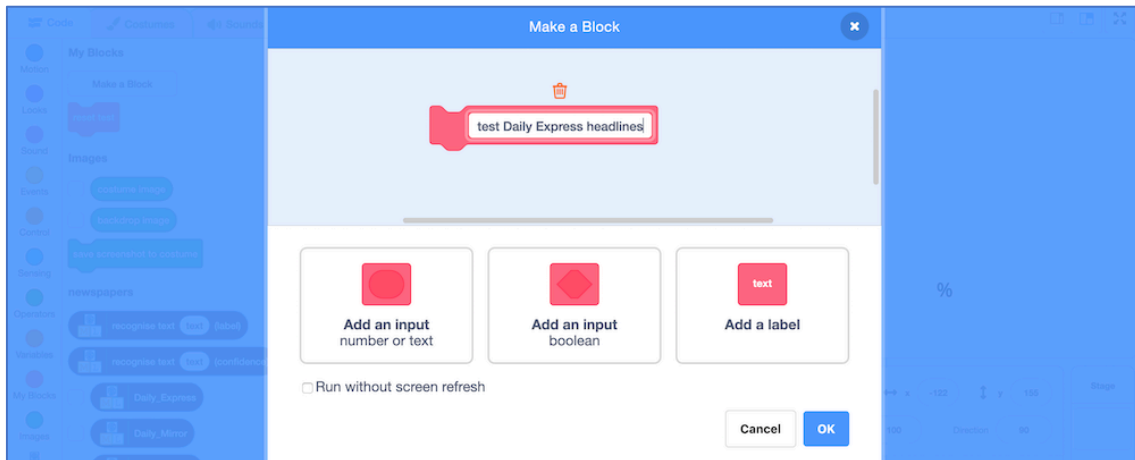
44. Enter the following script into the new custom block
It will run your test headlines from the Daily Express through your machine learning model, and count the results it returns.
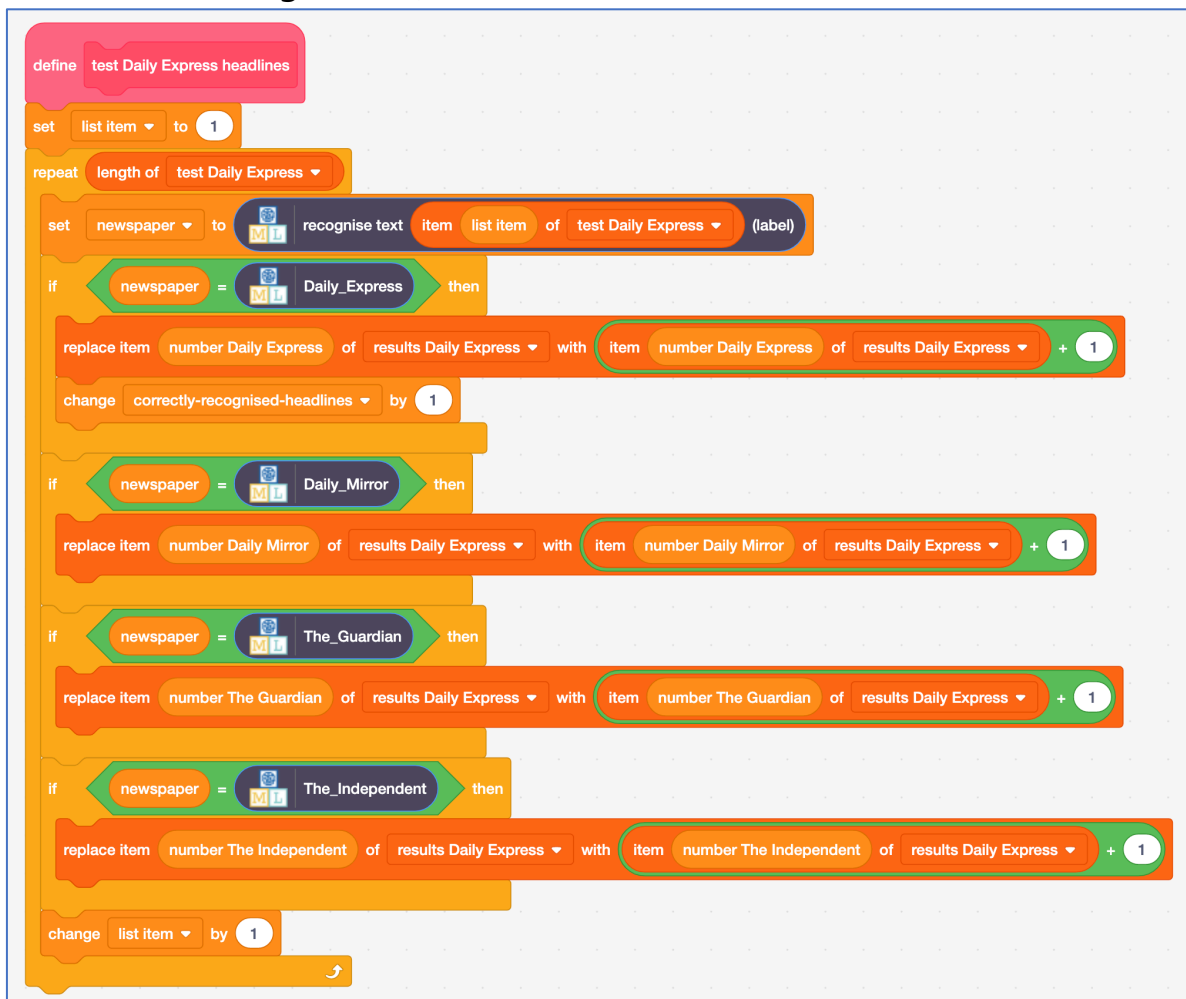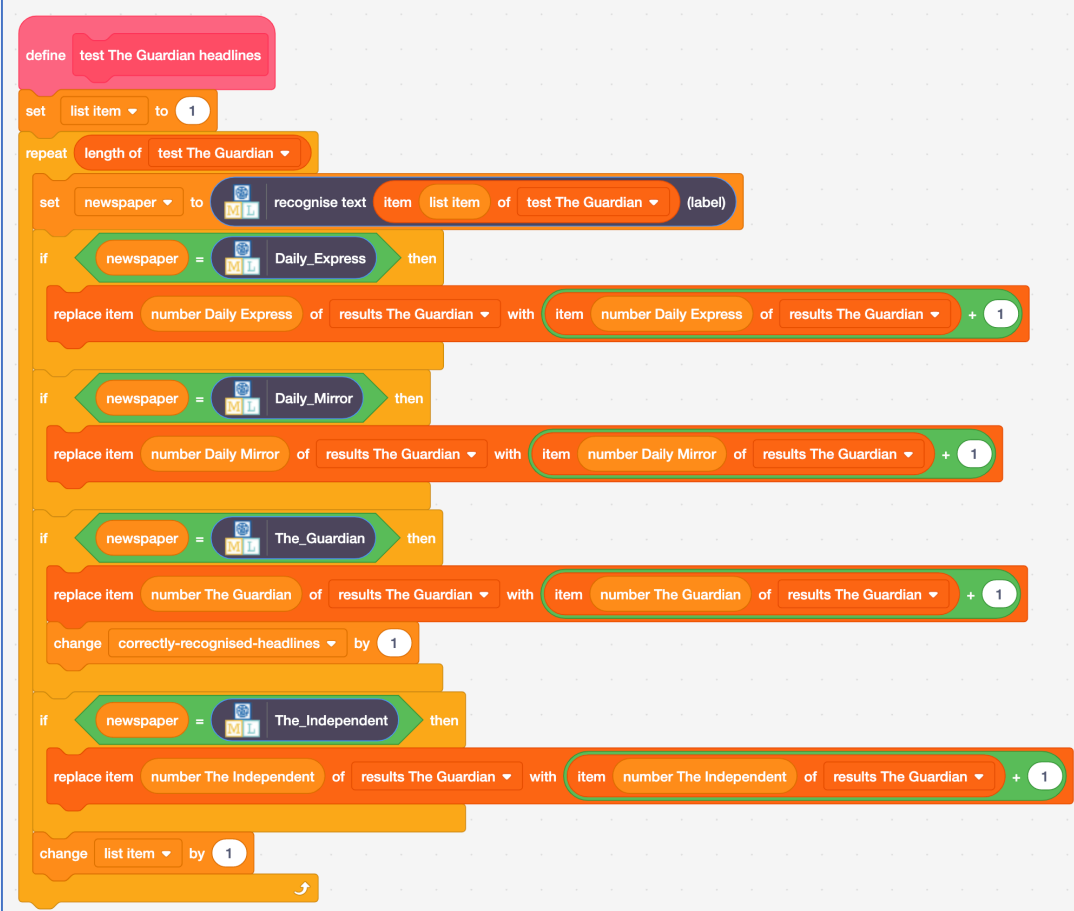
45. Create a new block to test headlines from your second newspaper

Click on the graph-block sprite.

*Click on the **Code** tab. Click the **Make a Block** button.*

Name it "test Daily Mirror headlines" (replace the name with your newspaper)

46. 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.
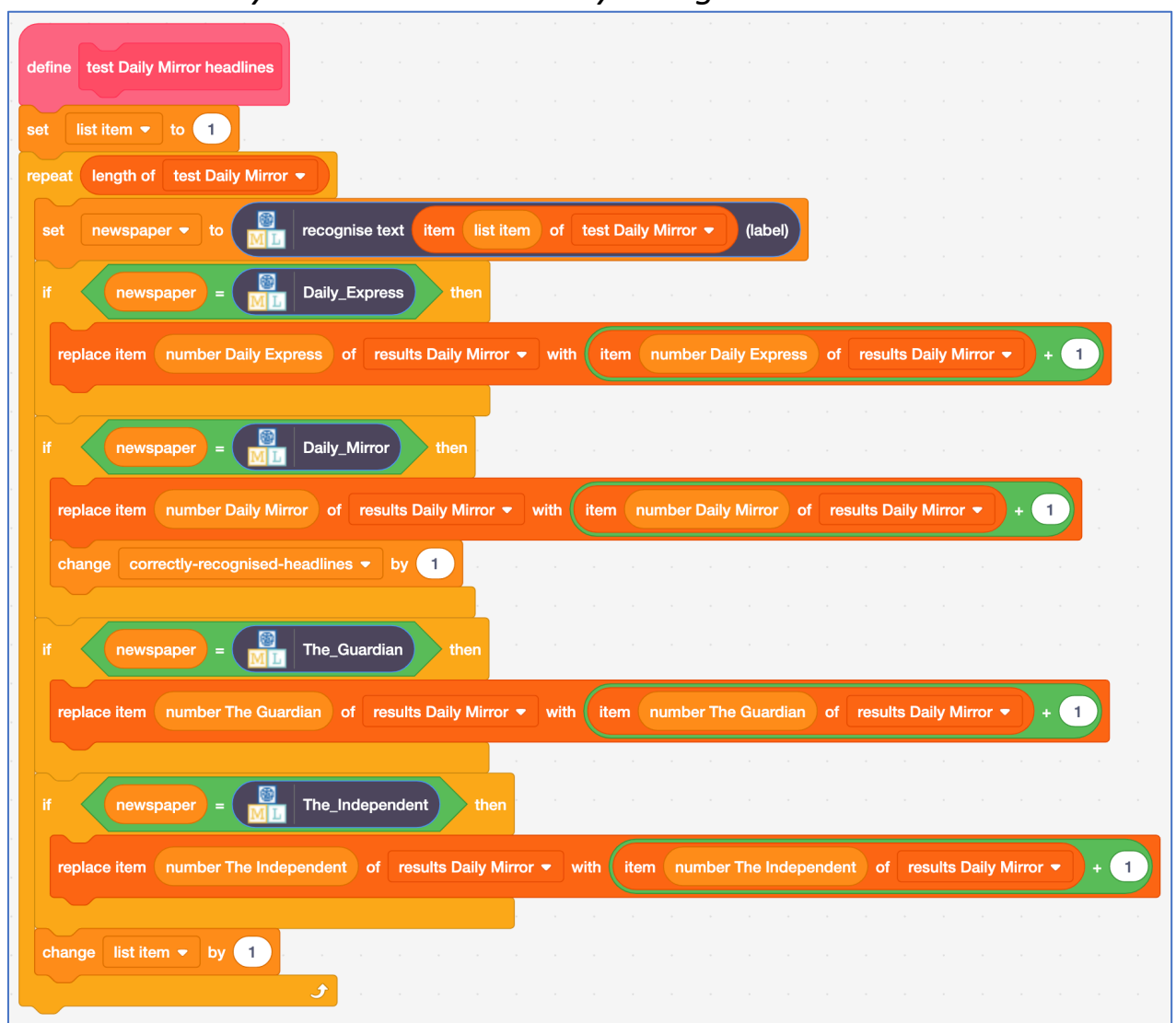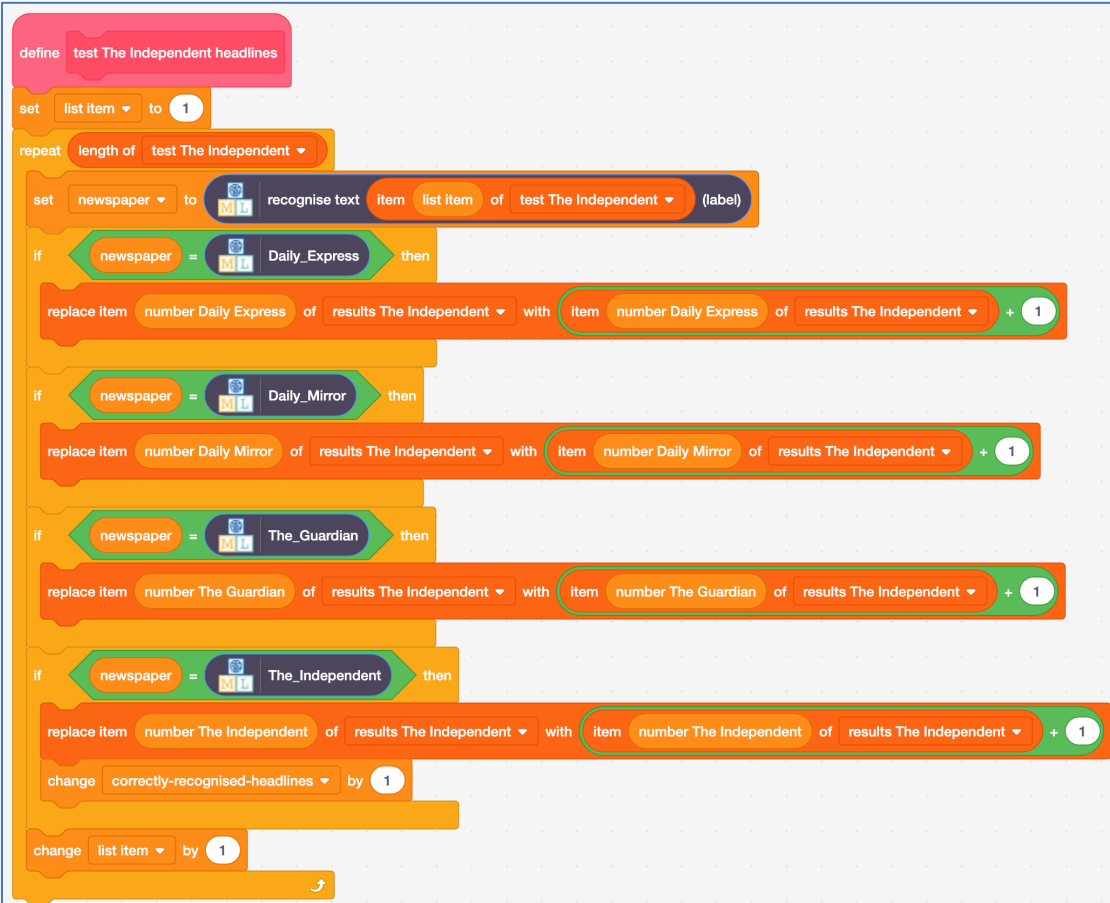
47.   Create two custom blocks to test the third & fourth newspapers

define test The Guardian headlines

set list item ▾ to 1

repeat length of test The Guardian ▾

　　set newspaper ▾ to 🖥 recognise text item list item of test The Guardian ▾ (label)

　　if newspaper = 🖥 Daily_Express then

　　　　replace item number Daily Express of results The Guardian ▾ with item number Daily Express of results The Guardian ▾ + 1

　　if newspaper = 🖥 Daily_Mirror then

　　　　replace item number Daily Mirror of results The Guardian ▾ with item number Daily Mirror of results The Guardian ▾ + 1

　　if newspaper = 🖥 The_Guardian then

　　　　replace item number The Guardian of results The Guardian ▾ with item number The Guardian of results The Guardian ▾ + 1

　　　　change correctly-recognised-headlines ▾ by 1

　　if newspaper = 🖥 The_Independent then

　　　　replace item number The Independent of results The Guardian ▾ with item number The Independent of results The Guardian ▾ + 1

　　change list item ▾ by 1

---

define test The Independent headlines

set list item ▾ to 1

repeat length of test The Independent ▾

　　set newspaper ▾ to 🖥 recognise text item list item of test The Independent ▾ (label)

　　if newspaper = 🖥 Daily_Express then

　　　　replace item number Daily Express of results The Independent ▾ with item number Daily Express of results The Independent ▾ + 1

　　if newspaper = 🖥 Daily_Mirror then

　　　　replace item number Daily Mirror of results The Independent ▾ with item number Daily Mirror of results The Independent ▾ + 1

　　if newspaper = 🖥 The_Guardian then

　　　　replace item number The Guardian of results The Independent ▾ with item number The Guardian of results The Independent ▾ + 1

　　if newspaper = 🖥 The_Independent then

　　　　replace item number The Independent of results The Independent ▾ with item number The Independent of results The Independent ▾ + 1

　　　　change correctly-recognised-headlines ▾ by 1

　　change list item ▾ by 1

48. Click on the extensions icon

49. Click on the **Pen** extension
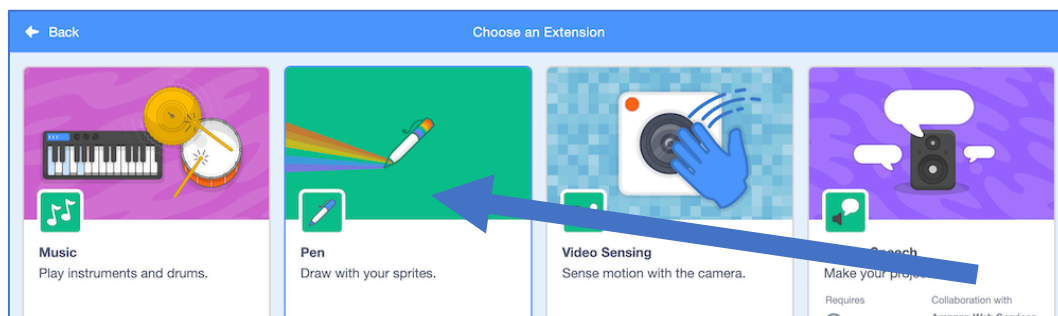
50. Create a new block to draw a graph of the results from the test of the headlines from the first newspaper
*Click on the **Code** tab. Click the **Make a Block** button.*
Name it "draw graph Daily Express" (replace the name with your paper)

Last updated: 21 July 2019
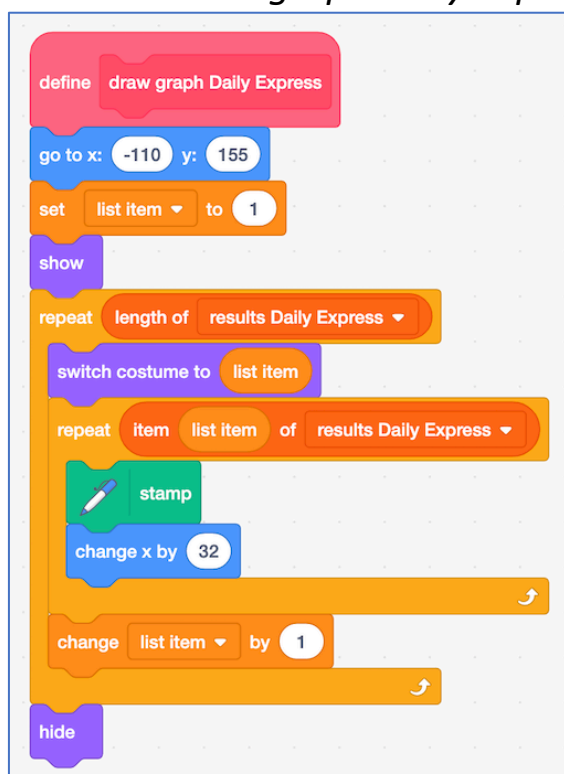
51. Create blocks to draw graphs for the other three newspapers.

```
define draw graph Daily Mirror

go to x: -110 y: 115
set list item ▼ to 1
show
repeat length of results Daily Mirror ▼
    switch costume to list item
    repeat item list item of results Daily Mirror ▼
        stamp
        change x by 32
    change list item ▼ by 1
hide
```

```
define draw graph The Guardian

go to x: -110 y: 75
set list item ▼ to 1
show
repeat length of results The Guardian ▼
    switch costume to list item
    repeat item list item of results The Guardian ▼
        stamp
        change x by 32
    change list item ▼ by 1
hide
```

```
define draw graph The Independent

go to x: -110 y: 35
set list item ▼ to 1
show
repeat length of results The Independent ▼
    switch costume to list item
    repeat item list item of results The Independent ▼
        stamp
        change x by 32
    change list item ▼ by 1
hide
```

52. 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.

```
define calculate accuracy
set accuracy ▼ to ( correctly-recognised-headlines / number-of-test-headlines )
set accuracy ▼ to ( accuracy * 100 )
set accuracy ▼ to ( round accuracy )
```
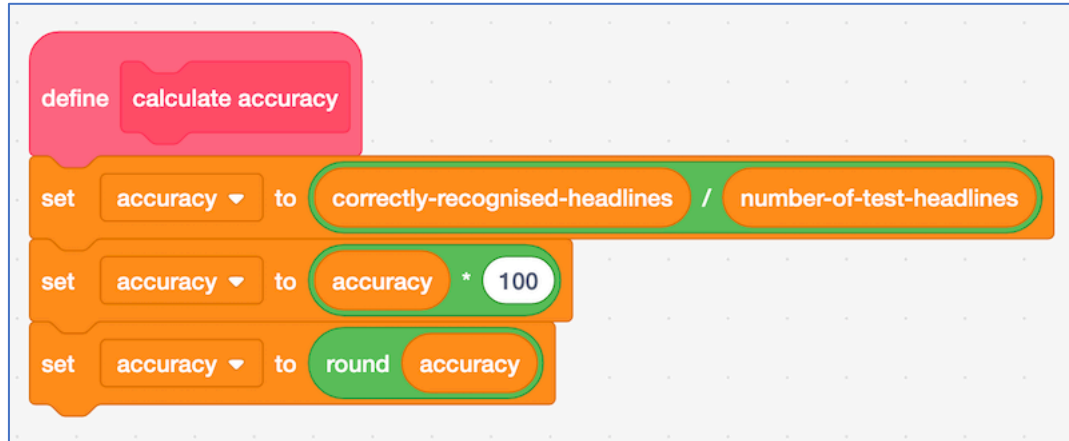
53. Time to put it all together!

Create this script in the graph-block sprite.
It will test your ML model using all your test headlines from all papers.
It will draw a graph to show you the results and calculate the overall accuracy to give you a score.

```
when 🏳 clicked
erase all
reset test
test Daily Express headlines
draw graph Daily Express
test Daily Mirror headlines
draw graph Daily Mirror
test The Guardian headlines
draw graph The Guardian
test The Independent headlines
draw graph The Independent
calculate accuracy
```
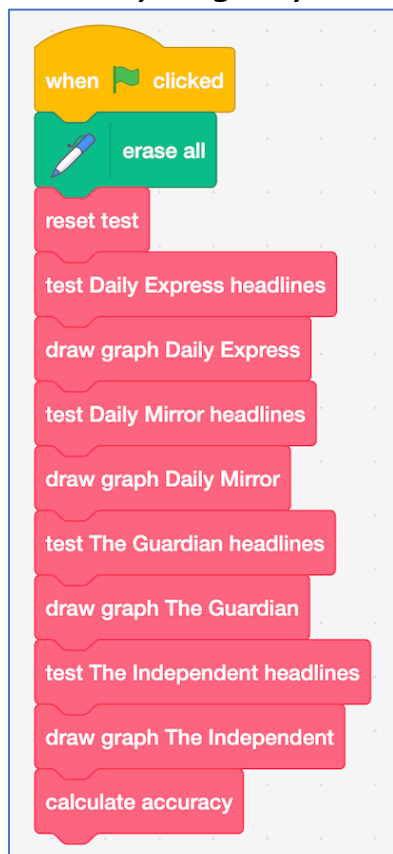
54. Run the test
Click the green flag, and watch it run.

Each row is made up of 10 blocks.

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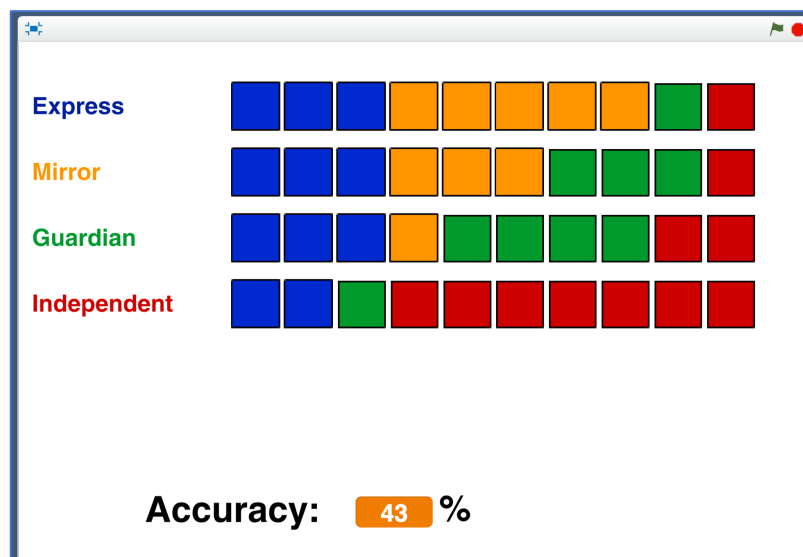
Independent

I collected 10 headlines from front-pages of The Independent.
My machine learning model "thought" 2 of them 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.

55. Save your Scratch project
*Click on **File** -> **Save to your computer***

56. Go back to the training tool window

57. Click on the **"< Back to project**" link.

58. Click the **Train** button

59. 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.

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

< Back to project

Add new label

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

Add example

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

Add example

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...

Add example

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 G20
- 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...

Add example

60. Click the "**< Back to project**" link. Click the "**Learn & Test**" button.

61. Click the "**Train new machine learning model**" button
Wait for the training to complete. This might take a few minutes.

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

Machine learning models

< Back to project

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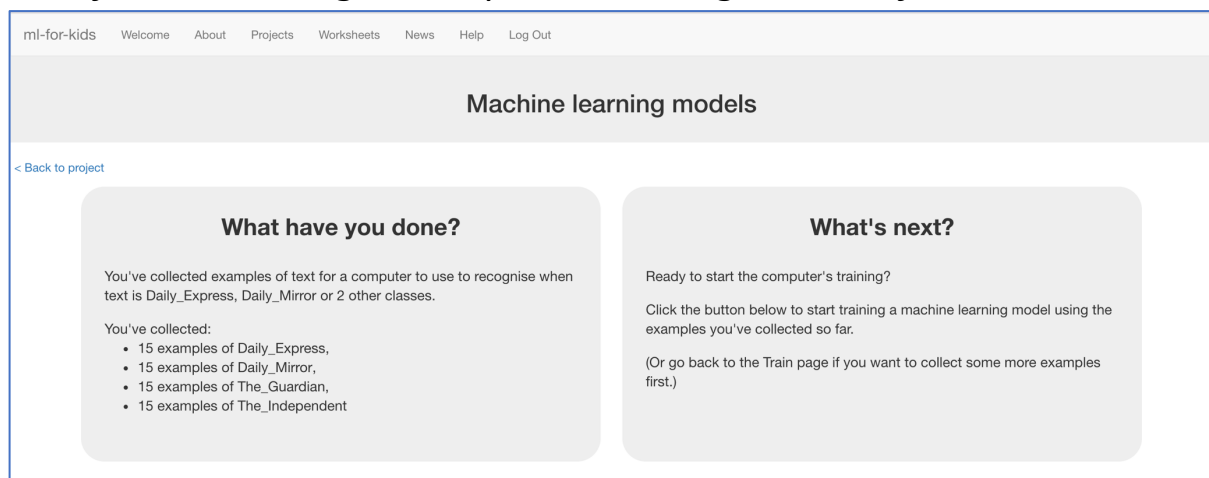- 15 examples of Daily_Mirror,
- 15 examples of The_Guardian,
- 15 examples of The_Independent

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.)

62. 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 Scratch project you saved before ("**File**" -> "**Load from your computer**")*

**63.**   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!
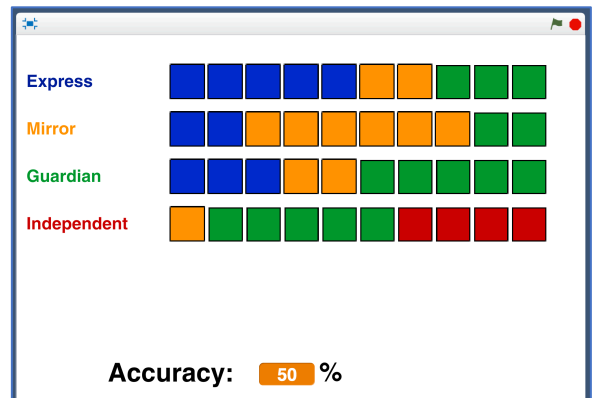
64.   Repeat steps 56 – 63 again.
Go back to the Training tool and collect more examples.
Train a new machine learning model.
Go back to Scratch and run your test to see how your new model is doing.

Machine learning models

< Back to project

What have you done?

You've trained a machine learning model to recognise when text is Daily_Express, Daily_Mirror or 2 other classes.

You created the model on Tuesday, August 22, 2017 2:38 AM.

You've collected:
- 25 examples of Daily_Express,
- 25 examples of Daily_Mirror,
- 25 examples of The_Guardian,
- 25 examples of The_Independent

What's next?

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.

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!

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!
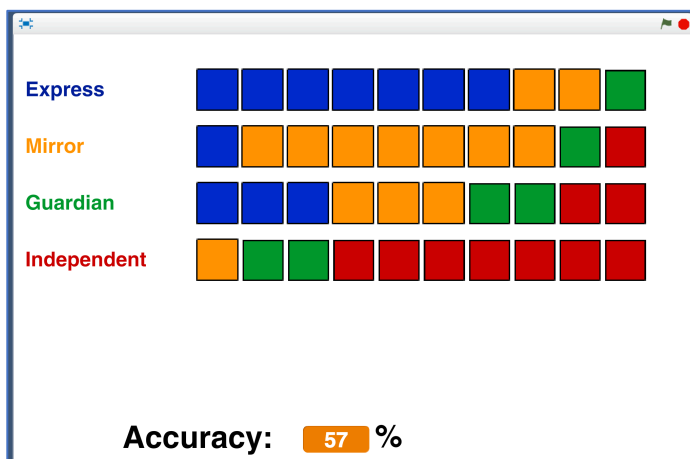
A little better overall.

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 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?