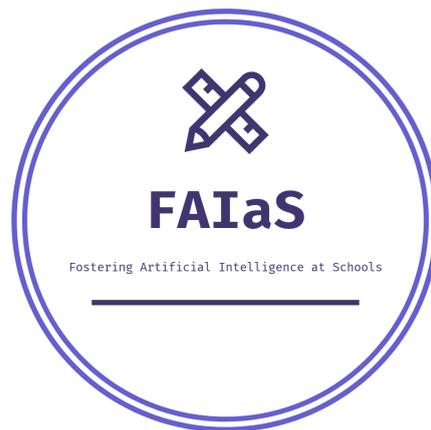




Activity Guide

Activity 1:

Gender Unconscious Bias – Learning ML





Index

| | |
|-----------------------------|-----------|
| Introduction | 4 |
| A step-by-step guide | 5 |
| Step 0 : | 5 |
| Step 1: | 5 |
| Step 2: | 6 |
| Step 3: | 6 |
| Step 4: | 7 |
| Step 5: | 7 |
| Step 6: | 8 |
| Step 7: | 9 |
| Results | 10 |



Glossary

AI(Artificial Intelligence)

CM(Classification model)

Train

Learn

Try



Introduction

Nowadays the first step in the recruitment or selection process is via websites for recruitment or applications, such as LinkedIn. Artificial Intelligence (AI) models are used for selection in these applications. Models of this type should be created to ensure that people of every race, gender, religion, and nationality have the same chances of being hired. When the model is created with unconscious bias, there will be a risk of producing discrimination when hiring someone.

In this activity, we'll show you an example of how we introduced gender bias when developing an artificial intelligence model that will help a startup hire a Computer Science expert. The model will classify candidates into two classes **Hired** or **Not hired** and you can see how the model decides who gets hired and how this bias will affect the results.

The data that is used to train this model are images about women and men in different situations, but we added bias about women, for example we added more pictures of women cooking and caring for children and fewer pictures of them looking like professionals.



A step-by-step guide

Step 0 :

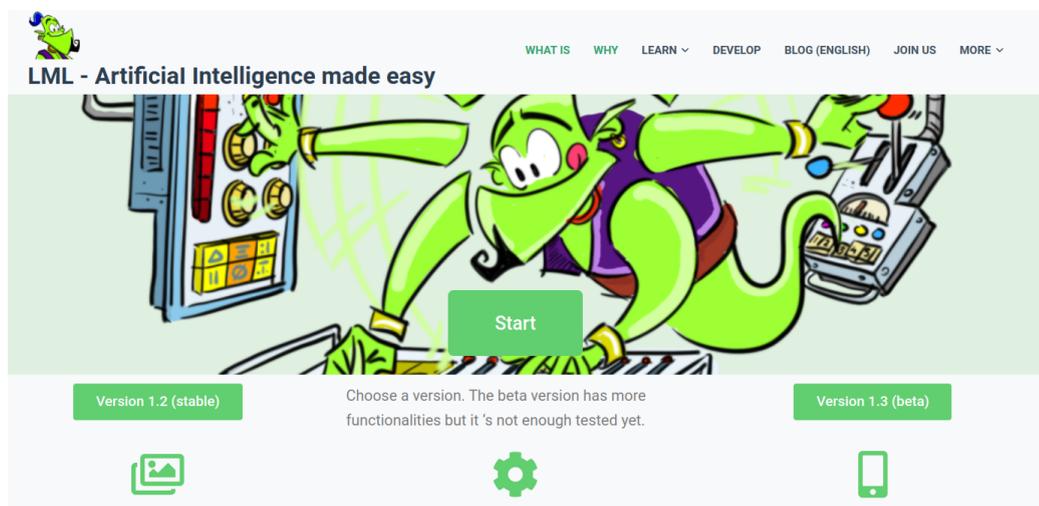
- Click here to visit the following website:
<https://fosteringai.github.io/project/result2/>
- Download this file: *source_activity1.zip* in your Desktop
- Unzip : *source_activity1.zip*

The file *source_activity1* contains two files :

- activity1_gender_ub.json
- test_image.jpeg

Step 1:

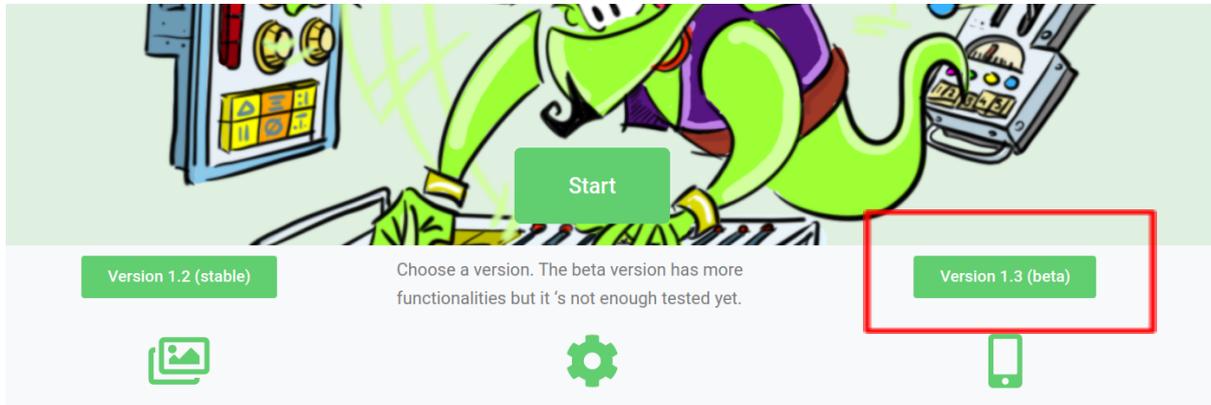
- Click here to visit the following website:
<https://web.learningml.org/en/home-spanish-en-translation/>





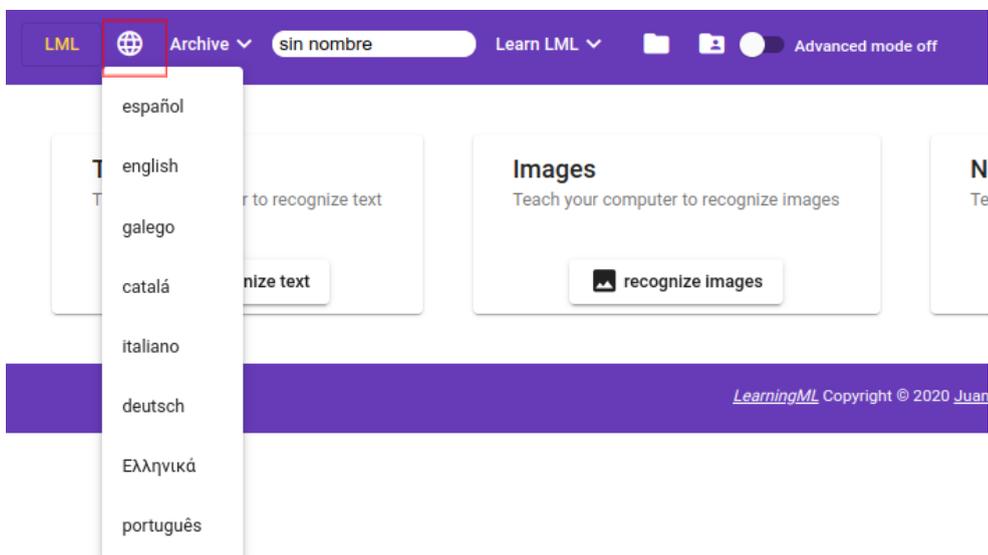
Step 2:

a) Click on : version1.3



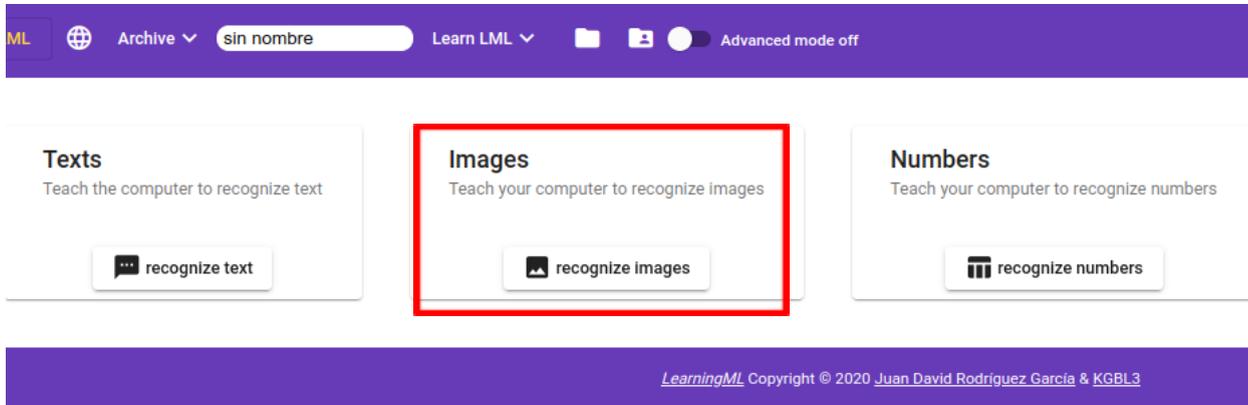
Step 3:

a) Select your language: click on the circle



Step 4:

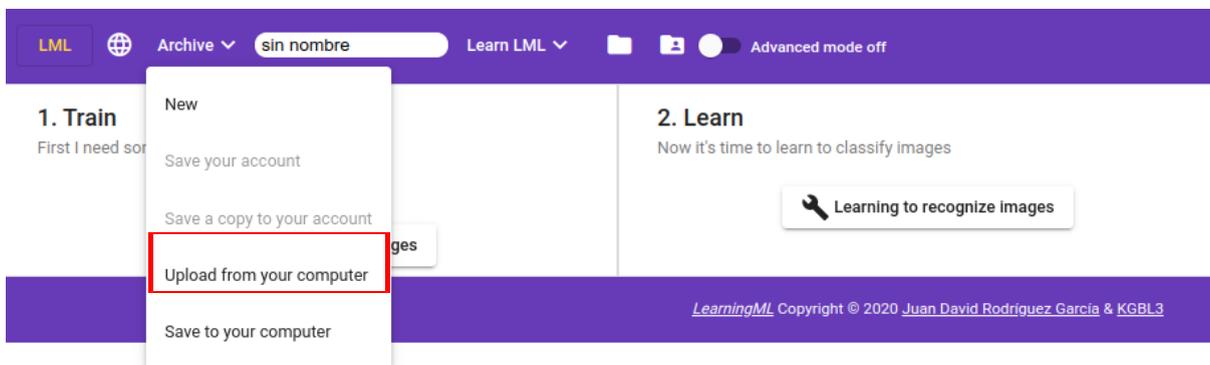
a) Click on : recognizer images



Step 5:

a) Upload file: activity1_gender_ub.json

Click on -> **Archive** -> **Upload from your computer**



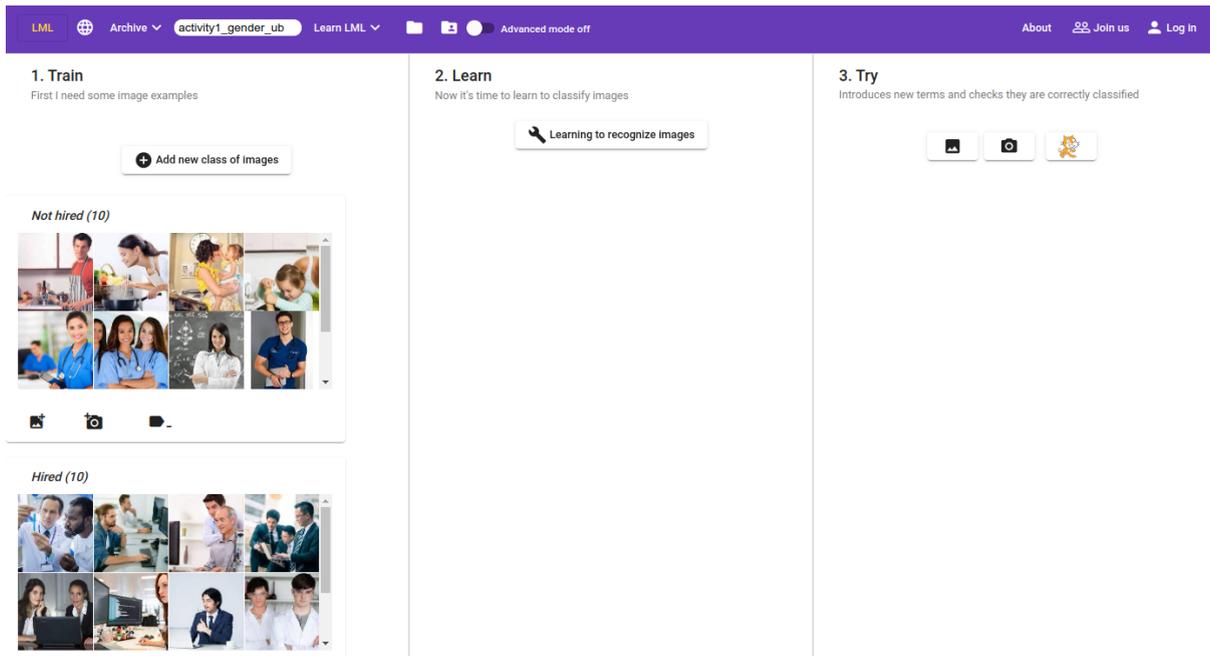
b) Select file **activity1_gender_ub.json** and upload

Desktop -> **source_activity1** -> activity1_gender_ub.json



Note:

As you can see in the following picture the two classes (Hired and Not hired) have been added with their respective images.



Learning MI have 3 stages: Train, Learn and Try

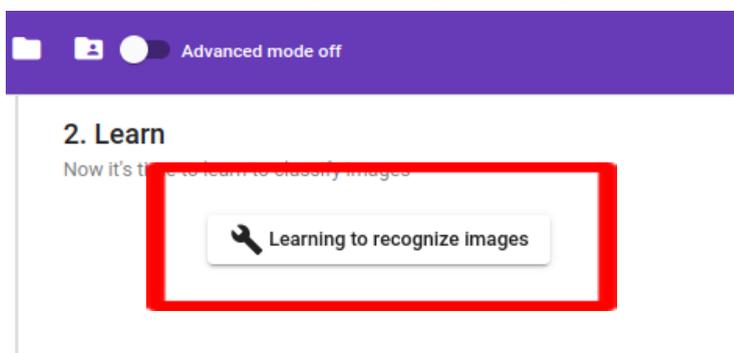
1. Train: This stage consists of adding the data that will be used to train the classification model. The classification model will use these images to learn similarity patterns between the images.

In our case we have 2 classes for classify to a candidate : Hired and Not Hired

In each class we need to add 10 images, but we added data bias in this example. We added more male images than female images in the Hired class and the Not hired class we added more female images. Also, we added more pictures of women cooking and caring for children and fewer pictures of them looking like professionals.

Step 6:

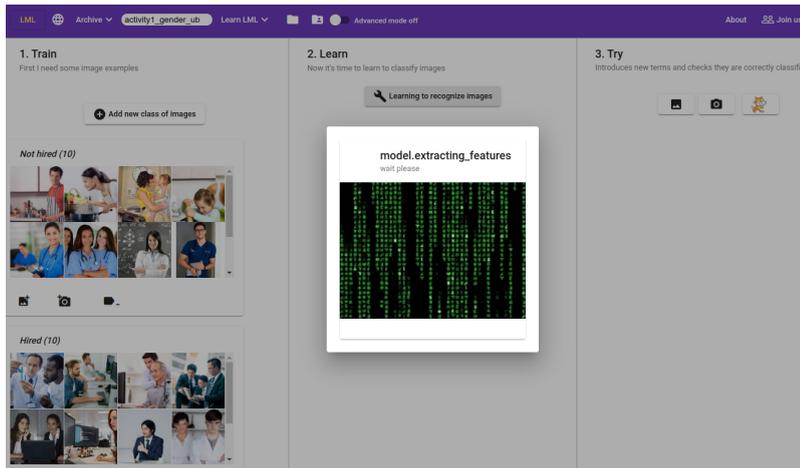
- a) Click on :learning to recognize images



Note:

The second stage in Learning MI is Learn:

2. Learn: In this stage, the classification model learns from the images we added in the previous step. The model learns the similarity patterns of the images and joins the things that are similar.



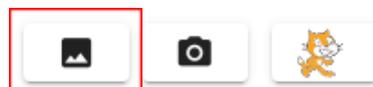
Step 7:

- a) Upload the test image:

Click on : Try -> icon

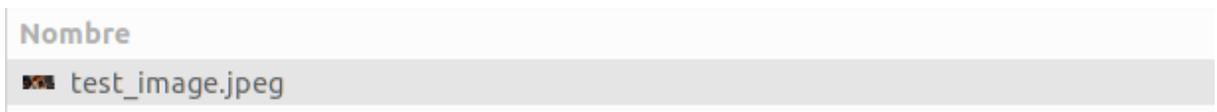
3. Try

Introduces new terms and checks they are correctly classified



- b) Select the test image: test_image.jpeg

Desktop->source_activity1->test_image.jpeg





Note:

The last stage in Learning ML is Try:

3. Try: In this stage the model is going to classify the candidate image as Hired or Not hired. The test image can't not be part of the data train stage.

Results

In this activity, in the *test_image* there are two developer women.

As result, the model classifies a candidate as follows : **Not hired**

3. Try

Introduces new terms and checks they are correctly classified



- **Not hired (82.08 %)**
- **Hired (17.92 %)**

Why these results?

This result was hope, because we introduced a gender bias when adding the images in each class Hired and Not hired. We added more male images in the class Hired and also we added more pictures of women cooking and caring for children and fewer pictures of them looking like professionals so there are more probabilities that the man will be hired. As a result, it is very important that the data used to train the classification model include both women and men in an equal way.