Prediction of heart disease
Decision tree

Objective

We want to create a model that predicts whether a person has a high risk of heart disease based on certain demographic and diagnostic metrics of the individual.

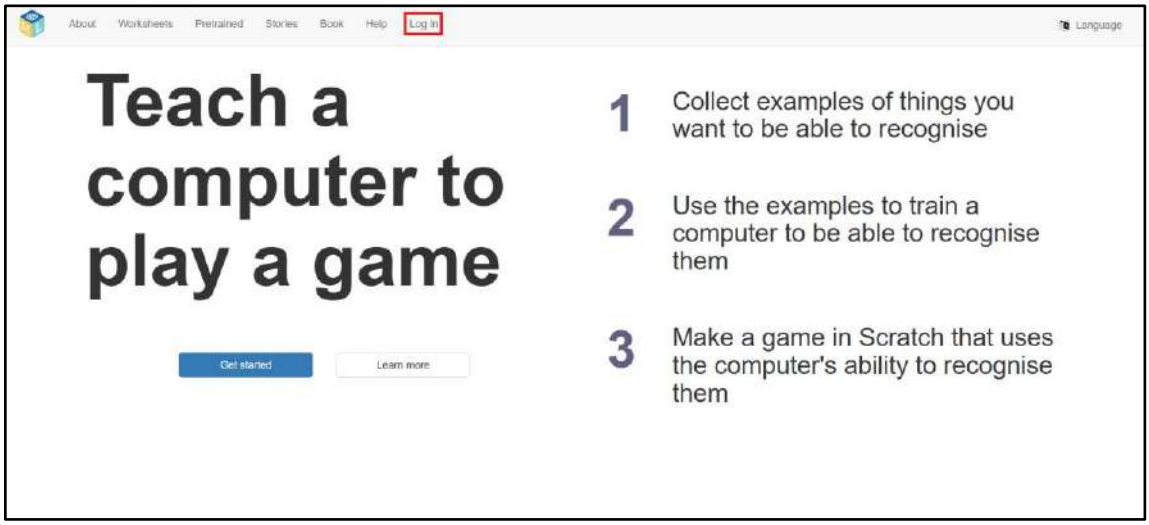
Dataset

- https://huggingface.co/datasets/Shambhavi625/heart_diseases/
- We will keep some of the rows and columns of the original dataset and change the column names to meet the needs of ML4Kids
- The table below shows the columns and their possible values as well as a description of them

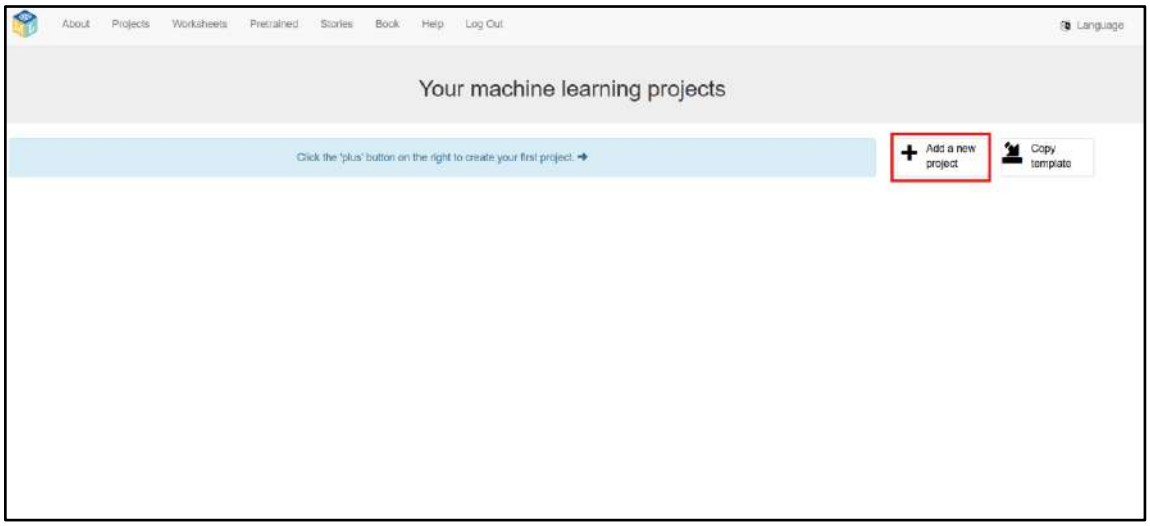
Column	Meaning	Possible values	Price Description
Age	Age of the patient	Numeric (30–80)	Higher age → higher risk of heart disease
Sex	Biological sex	M, F	M = male , F = female
ChestPainTyp	Type of chest pain	TA, ATA, NAP, ASY	TA = Classic heart attack-type pain ATA = Atypical angina - Similar but not school heart pain NAP = Non-angina pain - Chest pain that is NOT caused by the heart ASY = Asymptomatic - No chest pain at all
RestingBP	Blood pressure at rest	Numeric (~100–180)	~120 °C normal 130–139 °C increased ≥140 °C hypertension
Cholesterol	Cholesterol level	Numeric (~100–600)	<200 good 200–239 marginal≥ 240 high (risk of artery blockage)
FastingBS	Fasting blood sugar	0, 1	0 = normal 1 = high (possible diabetes)
RestingECG	Electrocardiogram result at rest	Normal, ST, LVH	Normal = healthy ST = possible ischemia LVH = thick heart muscle
MaxHR	Maximum heart rate	Numeric (~60–200)	Higher = better heart performance Lower = possible heart problem
ExerciseAngi	Chest pain during exercise	Y, N	Y = (Yes) pain when exercising (bad sign) N = (No) no pain
Oldpeak	How much oxygen deficiency does the heart show during the effort	Numeric (0–5+)	0 = normal 1–2 = moderate problem >2 = severe cardiac stress

Create, train, learn, and test

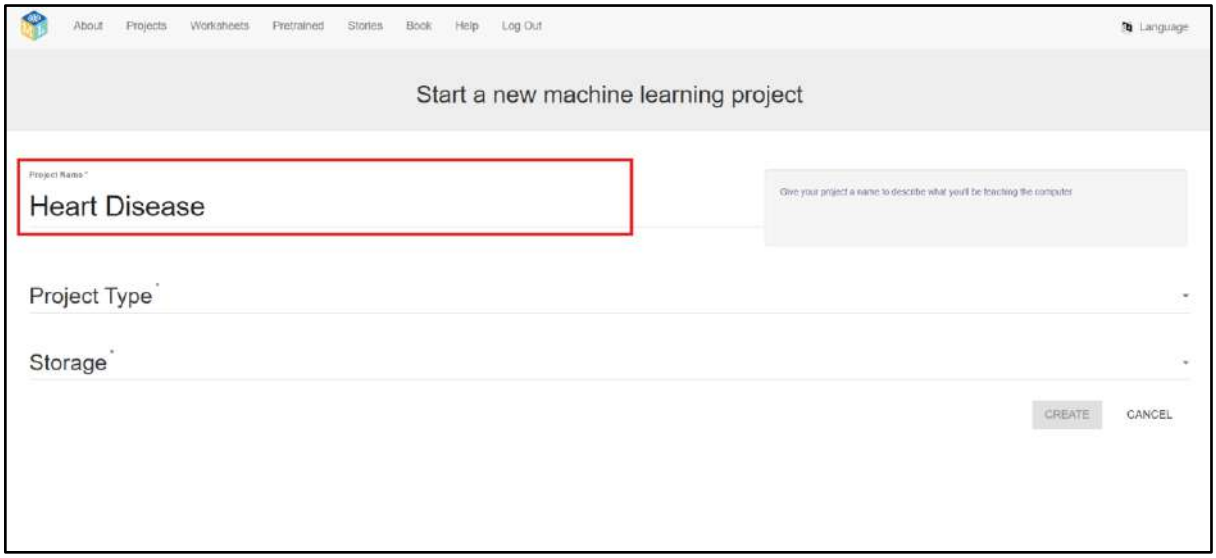
- Open ML4kids by visiting the link:
  - https://machinelearningforkids.co.uk/
- Log in – "**Log in**"



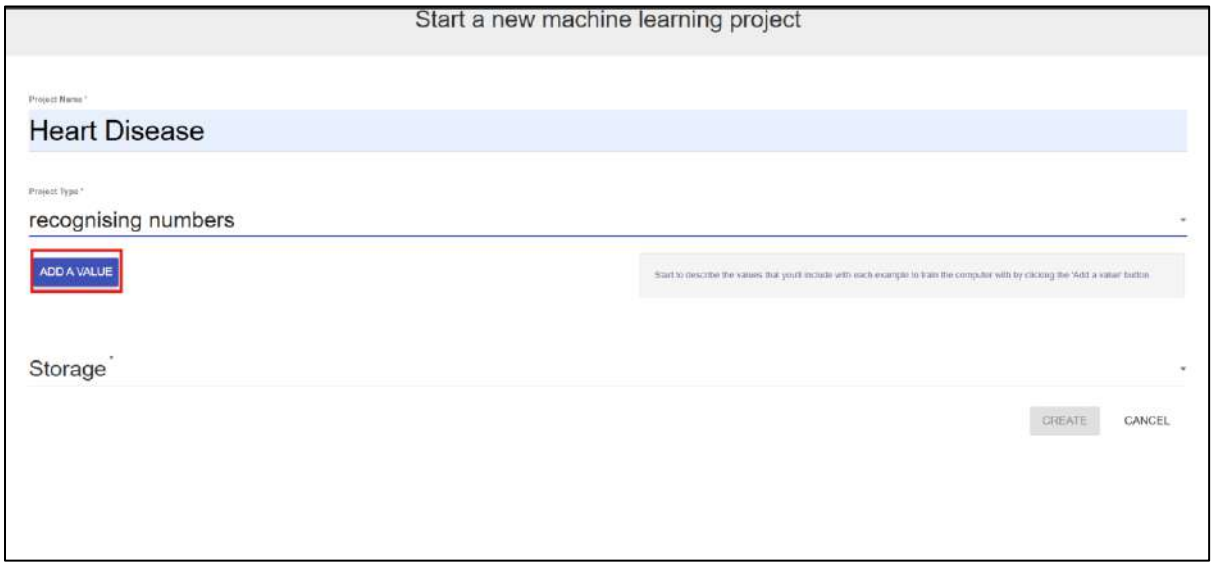
- Click "**Add a new project**"



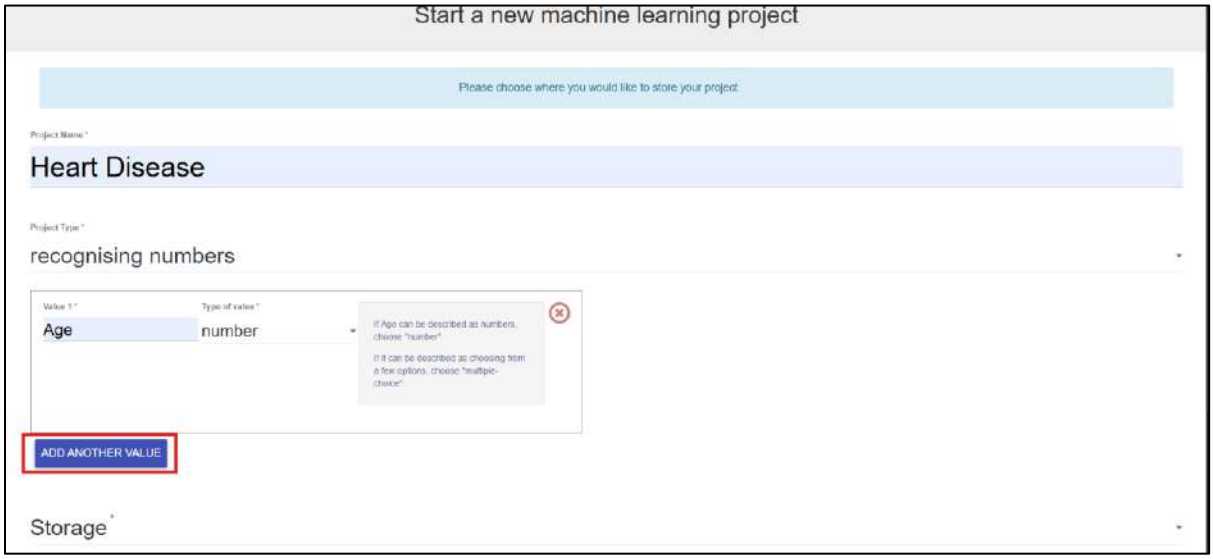
- Add "**Project Name**"



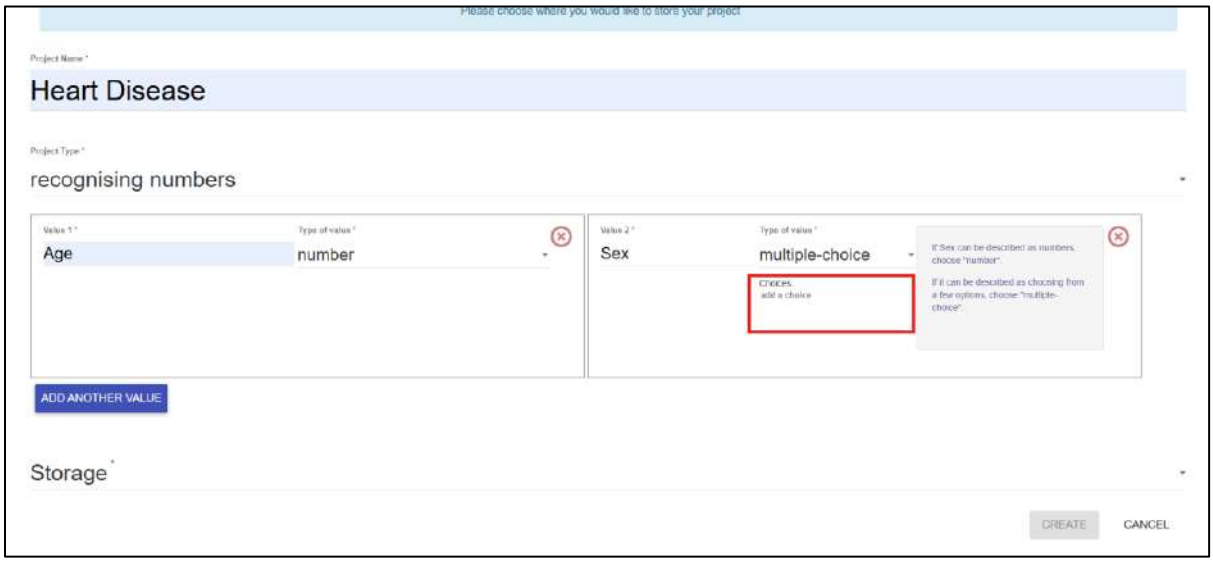
- As "**Project Type**" select:
    - "**Recognising numbers**"
- Click "**ADD A VALUE**"



- For each of the columns in the table above add the name of the value and the "**Type of value**"
- The possible types of value in ML4Kids are "**number**" and "**multiple-choice**", so if the price type is not numeric, you need to choose "**multiple-choice**" for everything else. If you select "**multiple-choice**", you must manually specify the possible values.
- First we add the value "**Age**" which is a numerical value
- We select "**Type of value**" -> **number**
- And then we click on "**ADD ANOTHER VALUE"**

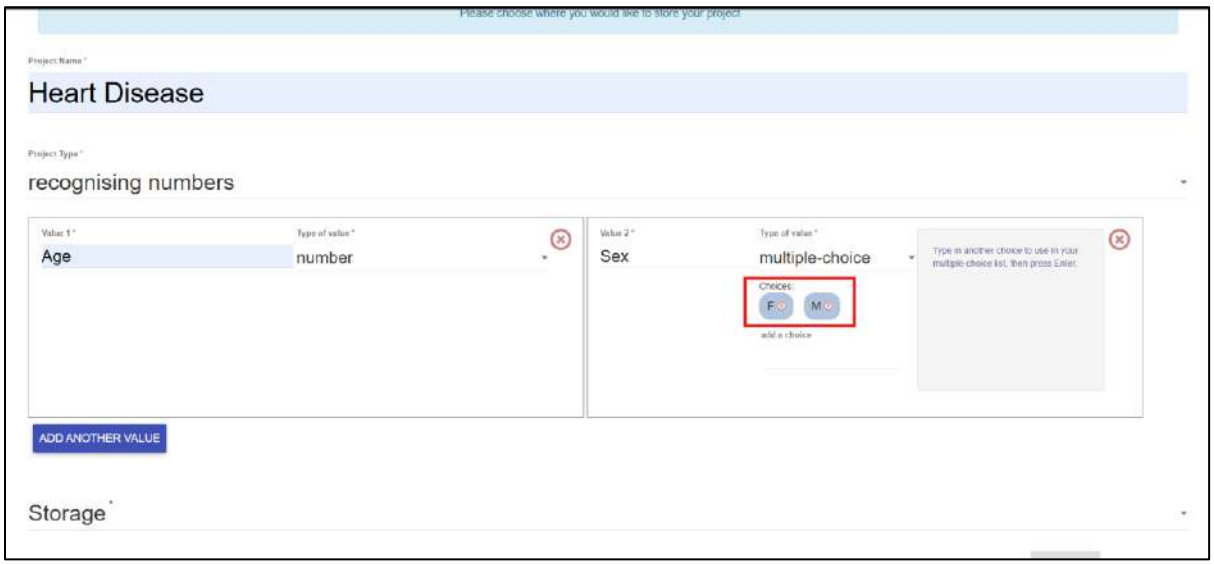


- Now we will add the value "**Sex**" which is not a numerical value, so its formula is "**multiple-choice**"
- A "Choices**"** field appears

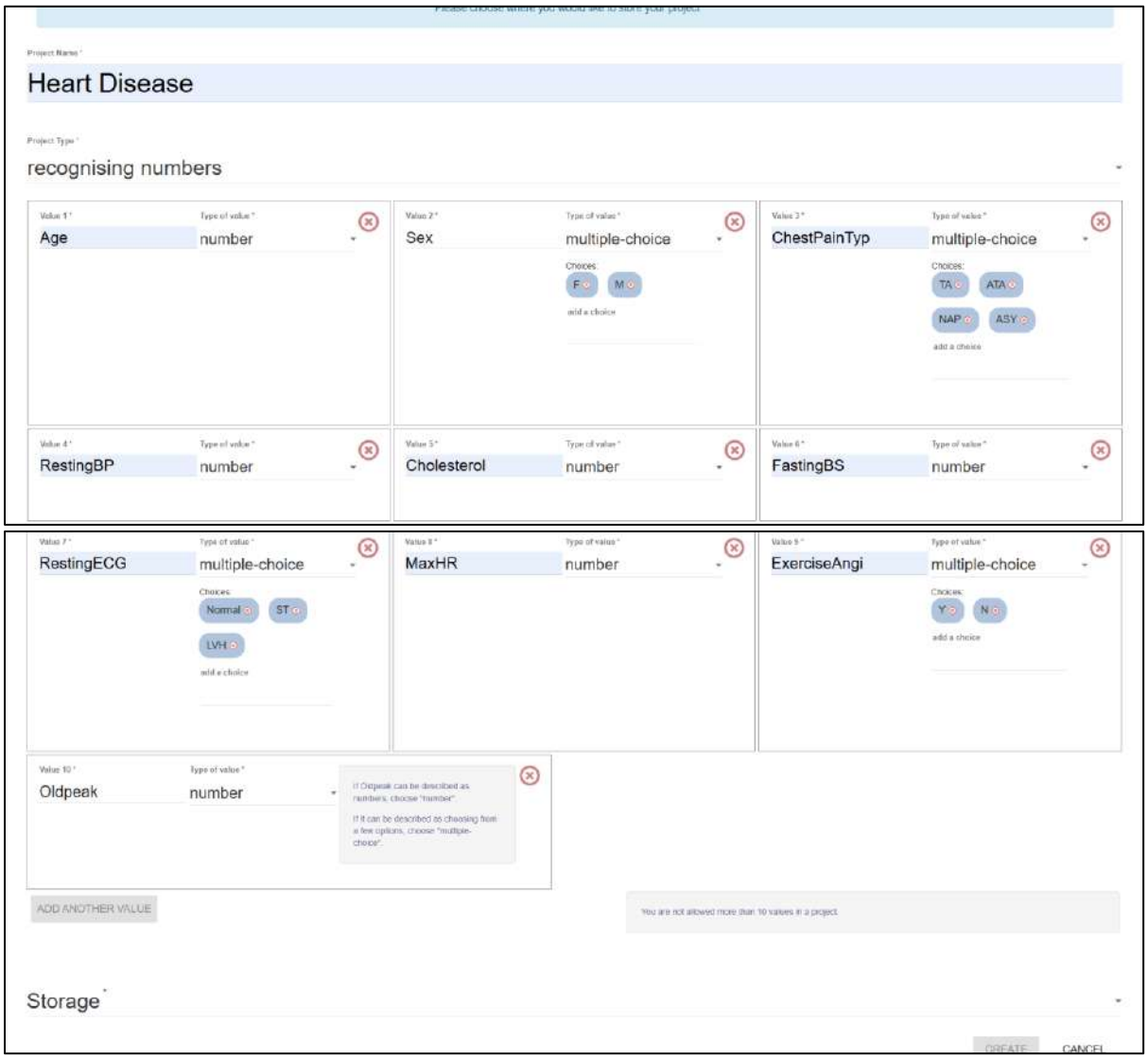


- We will add "**F**" as options
- We click on "**enter**"
- Then write "**M**"
- And click "**enter**"

- After adding the first two values (**Age, Sex**), your screen should look like this:

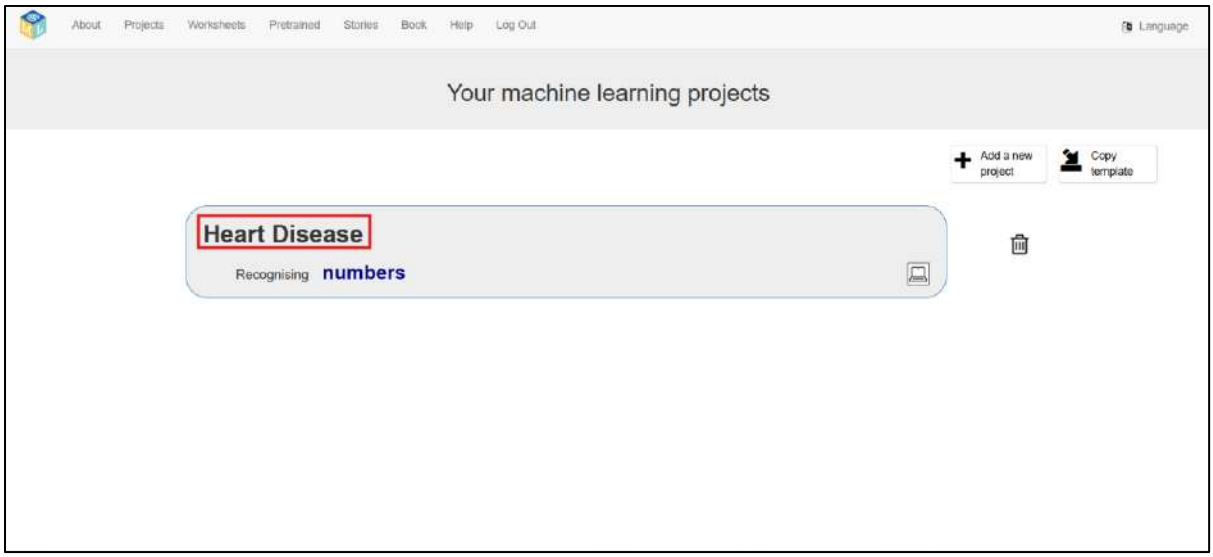


- We do the same for the rest of the column names, after entering all the columns your screen should look like the following:

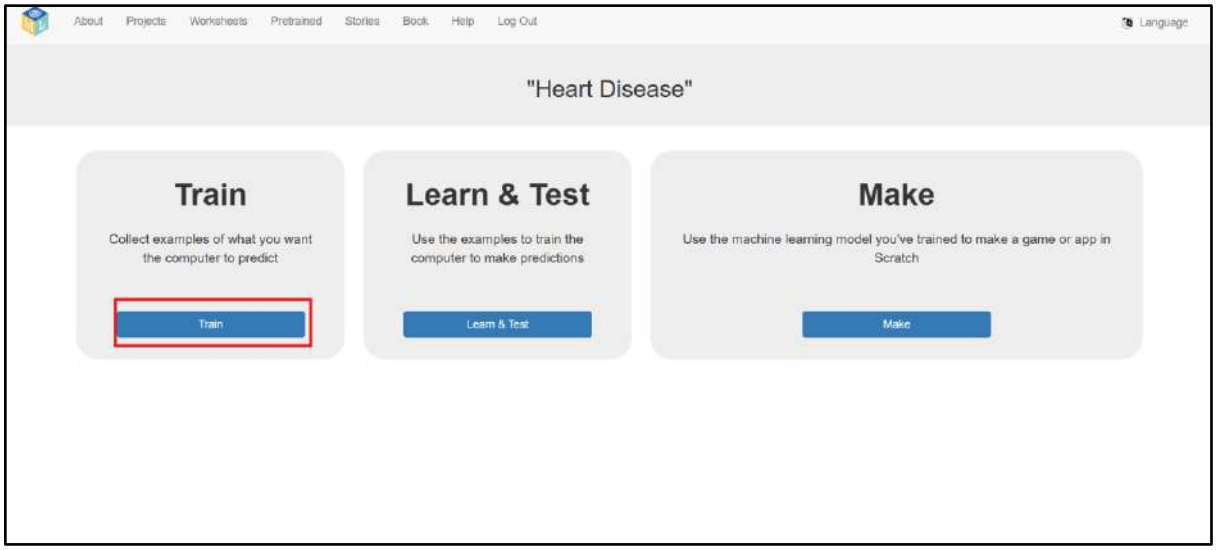


- In the "**Storage**" field, select:
    - "**In your web browser**"
- Click "**CREATE**"

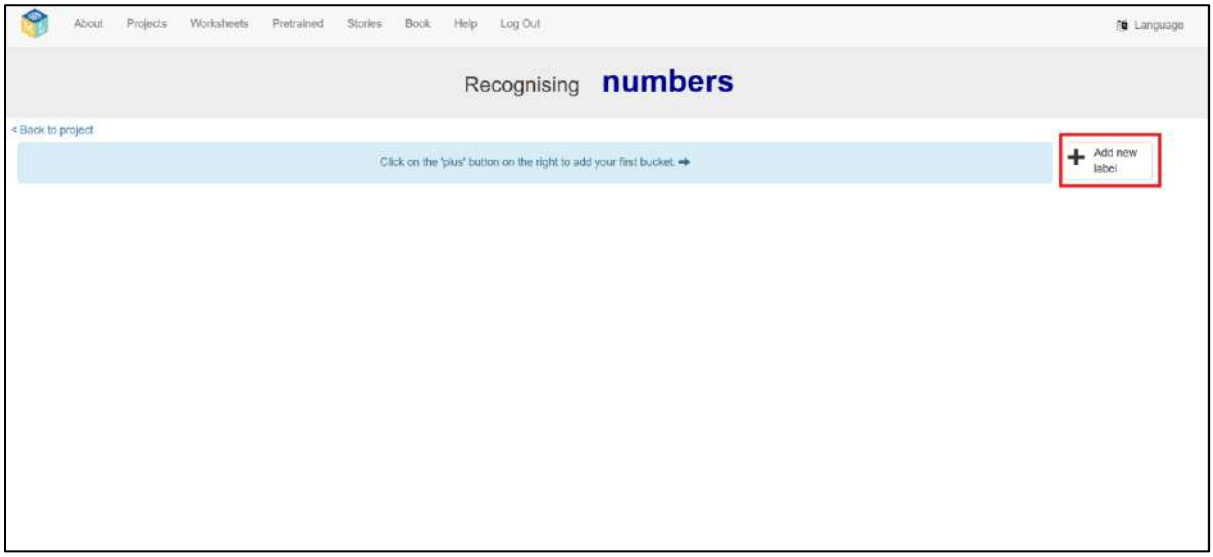
- Click on the project name



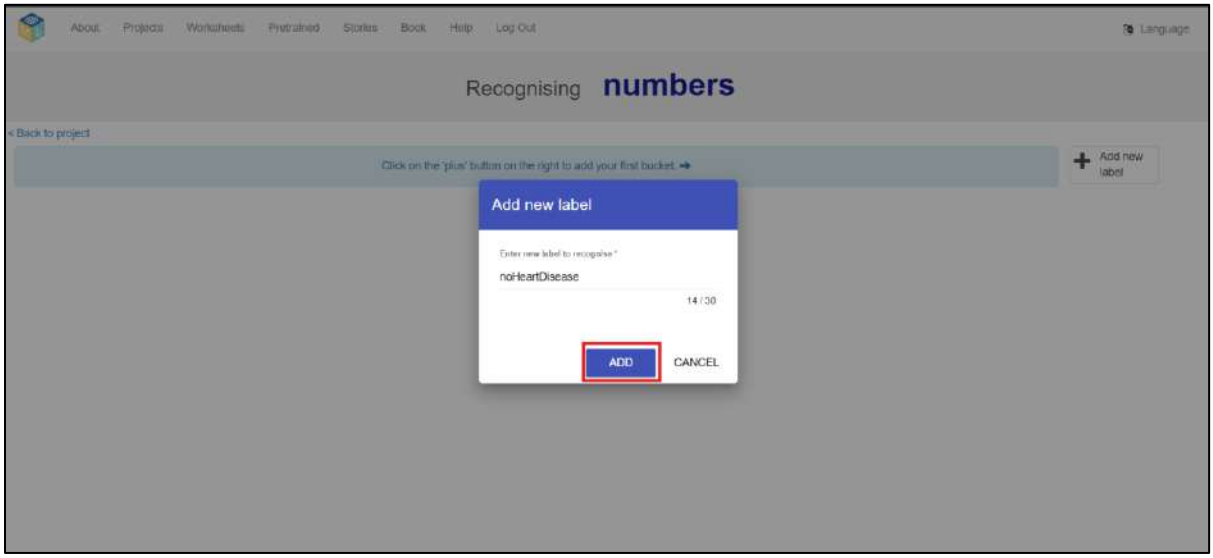
- Click "**Train**"



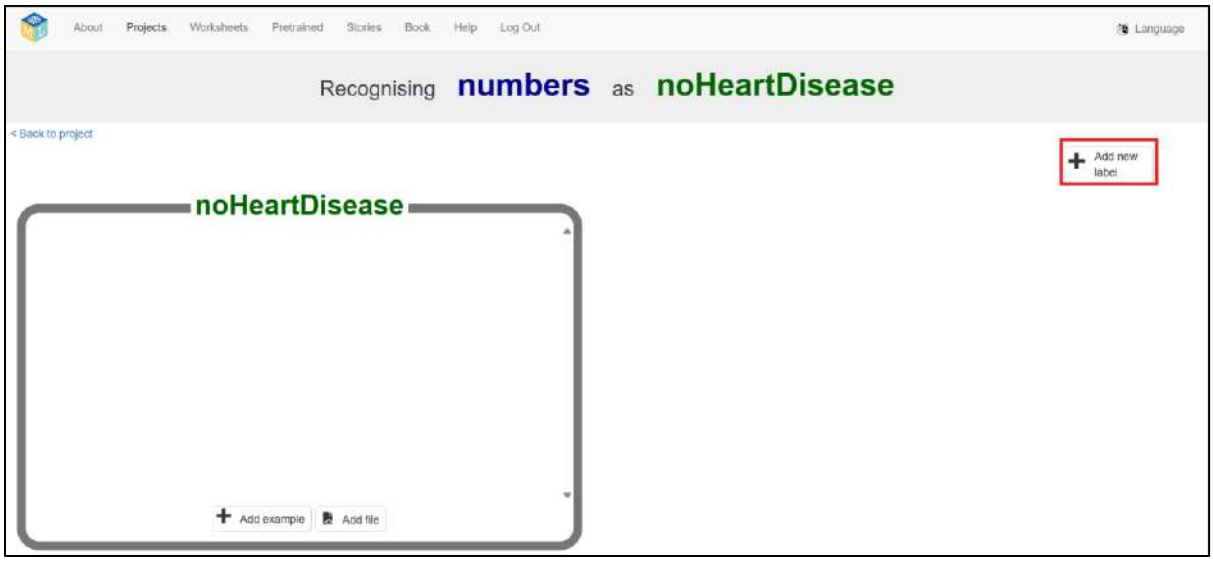
- Click on "**Add new label**"



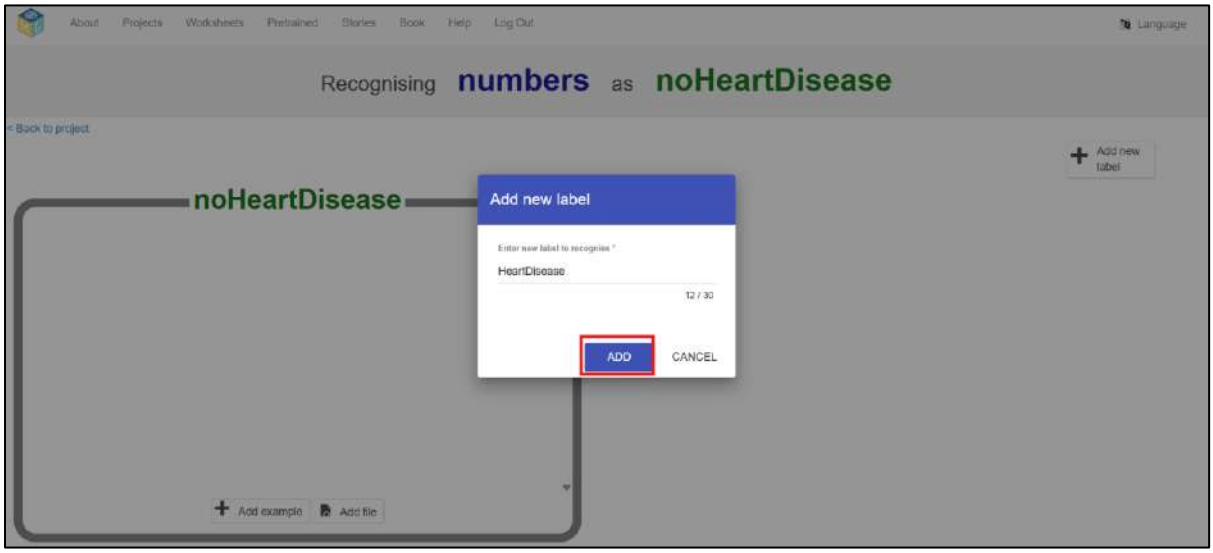
- Name it "**noHeartDisease**"
- Click on "**ADD**"



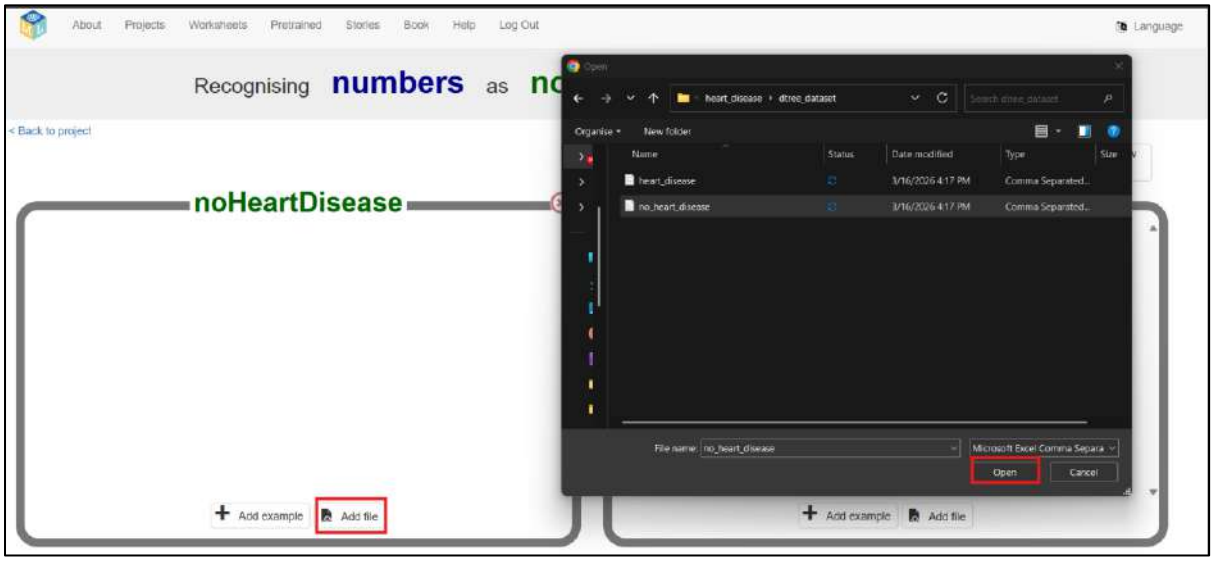
- Click "**Add new label**" again



- And name the second label "**HeartDisease**"
- And click on "**ADD**"

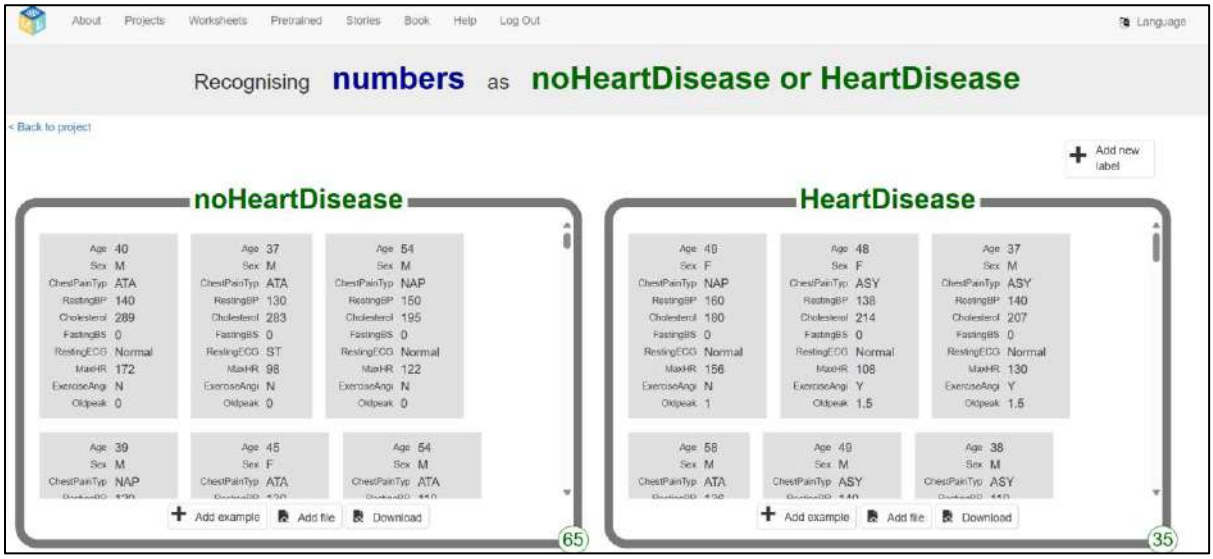


- Download the dataset files placed here: [Training Dataset]
- Inside the noHeartDisease tag click on "**Add file**"
- Select "**no_heart_disease.csv**"
- Click "**Open**"

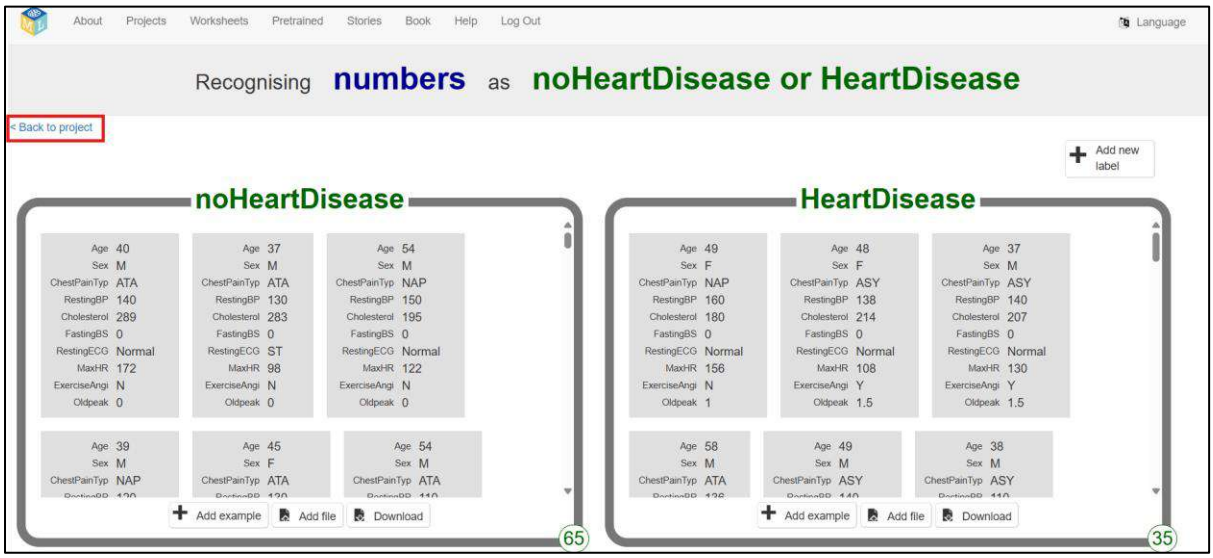


- Do the same for the other tag, but now select the "**heart_disease.csv**" file

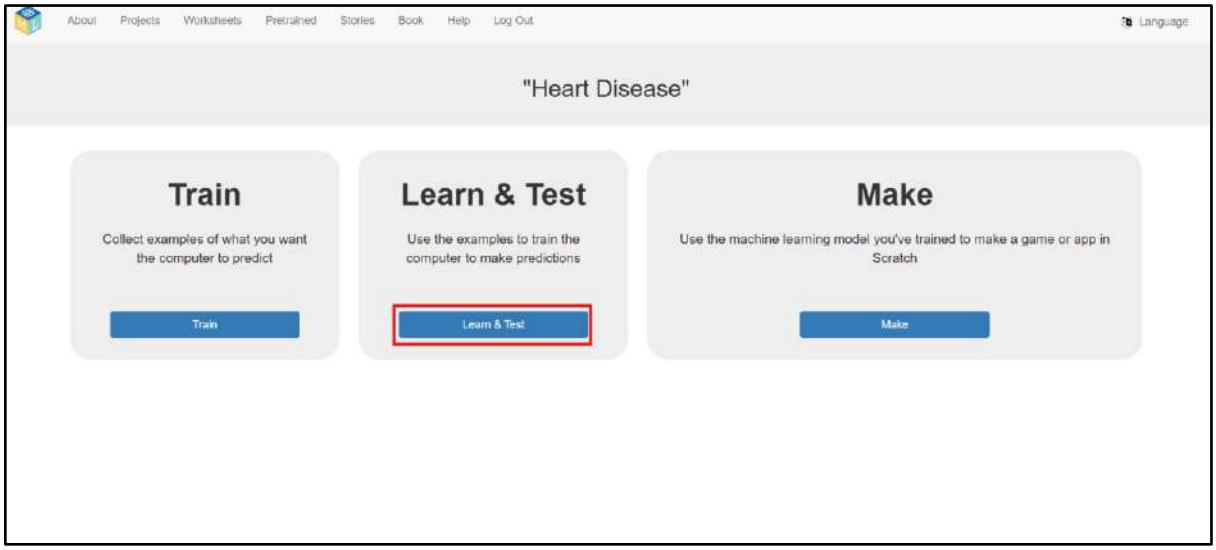
- After this step your screen should look like the following



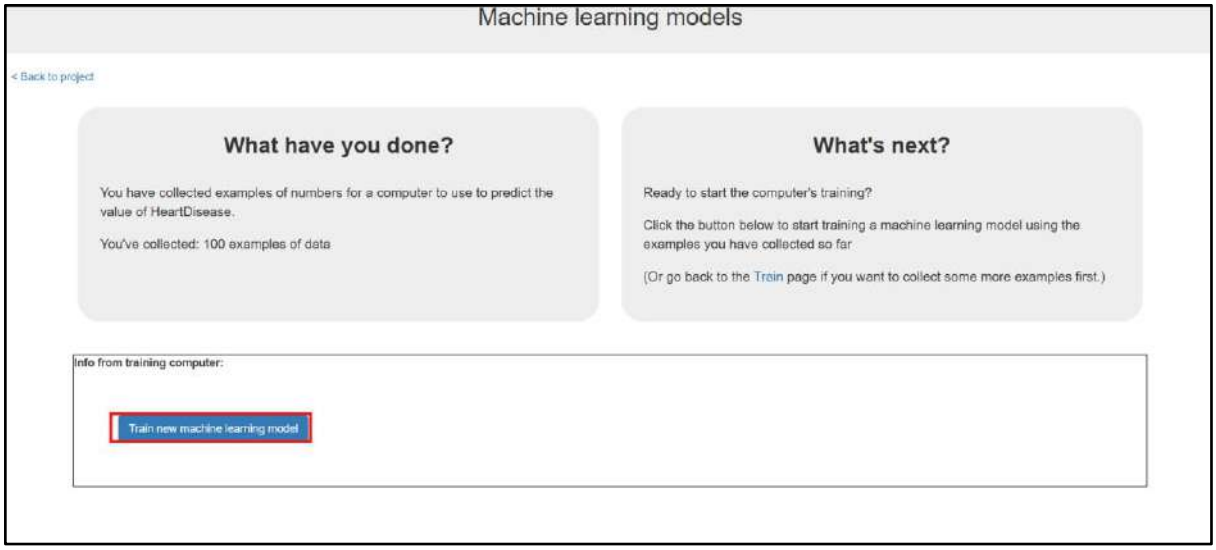
- Click on "**Back to project**"



- Click on "**Learn & Test**"

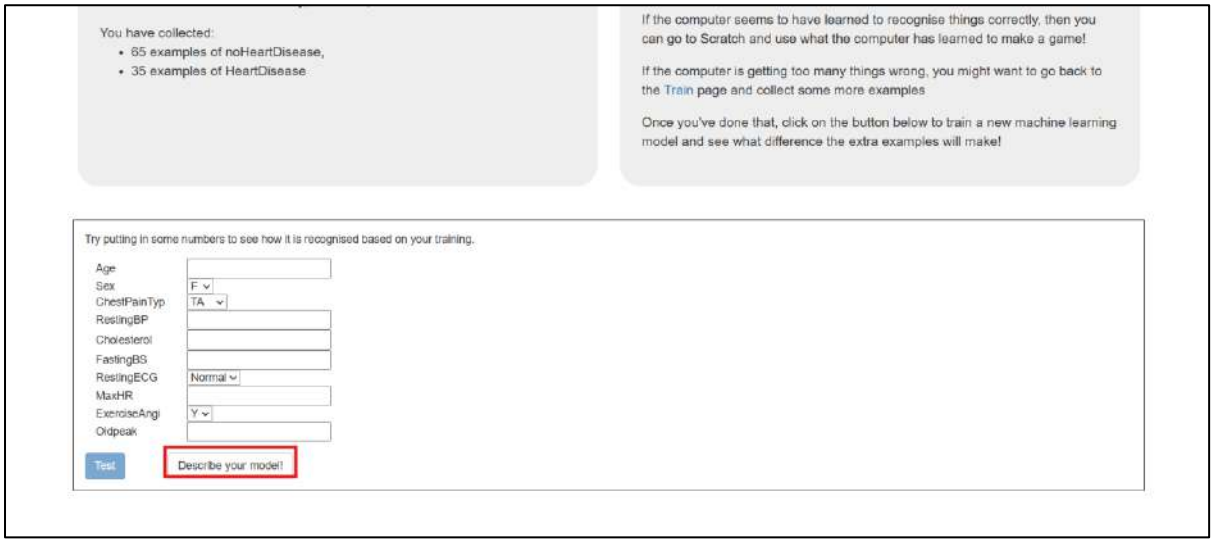


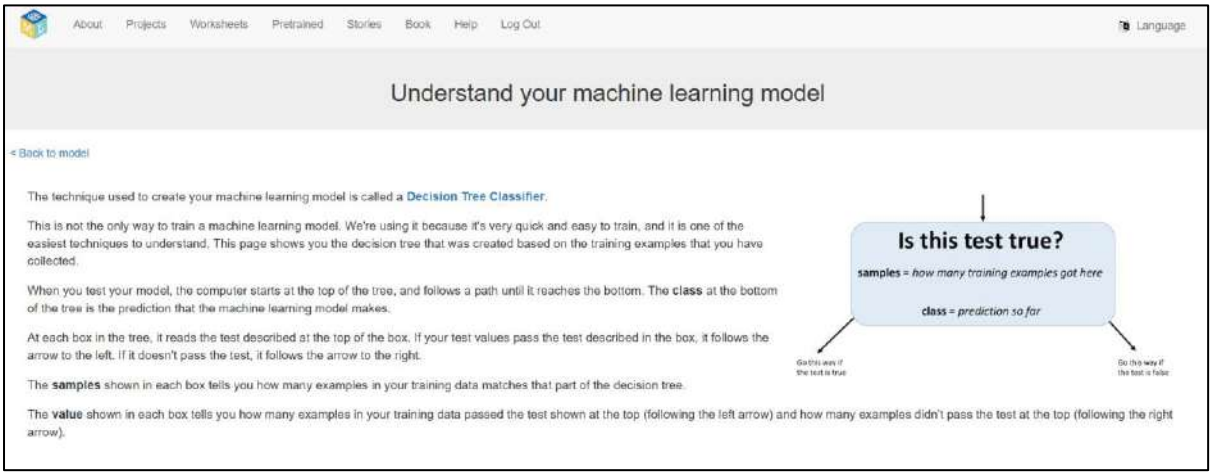
- Click on "**Train new machine learning model**"

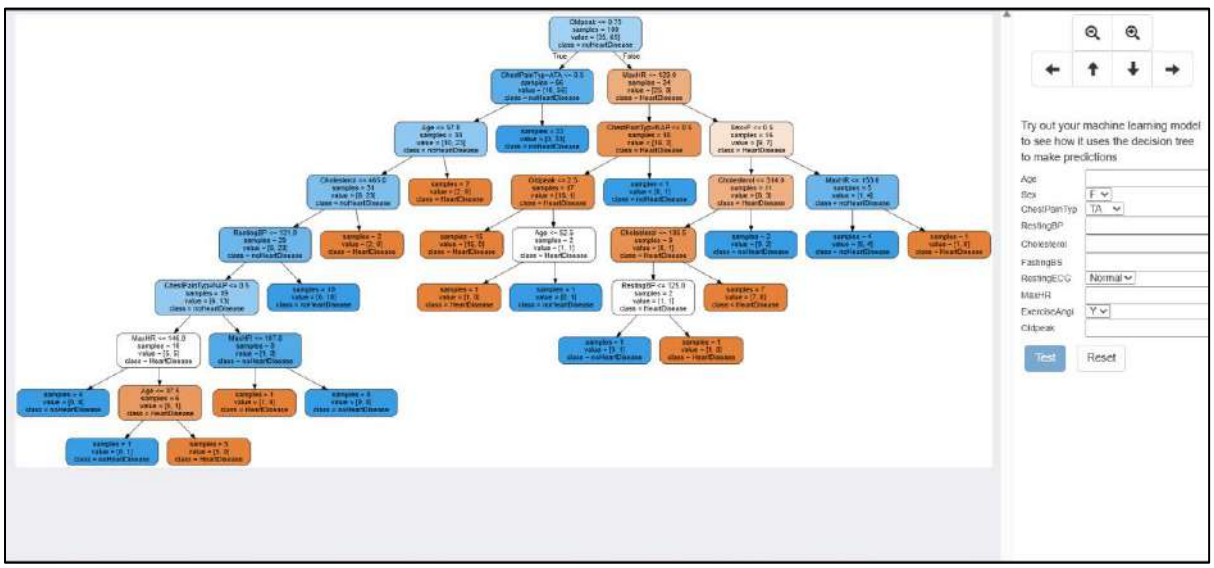


- If your model was successfully trained, it will be available for you to try on

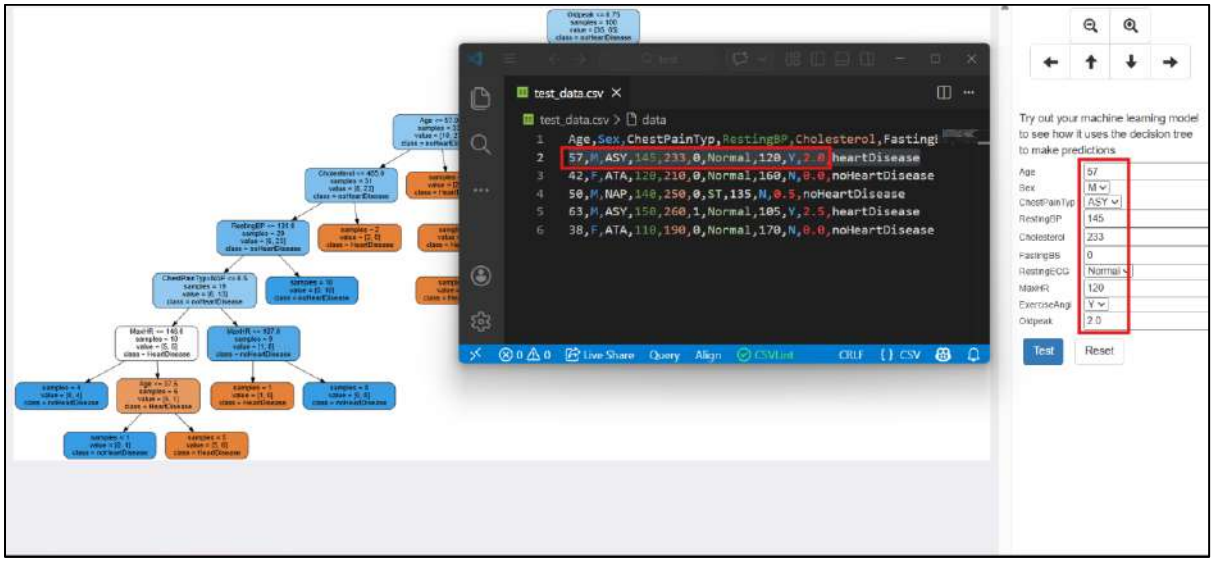
- If you click on "**Describe your model!**You can see the decision tree and an explanation of it



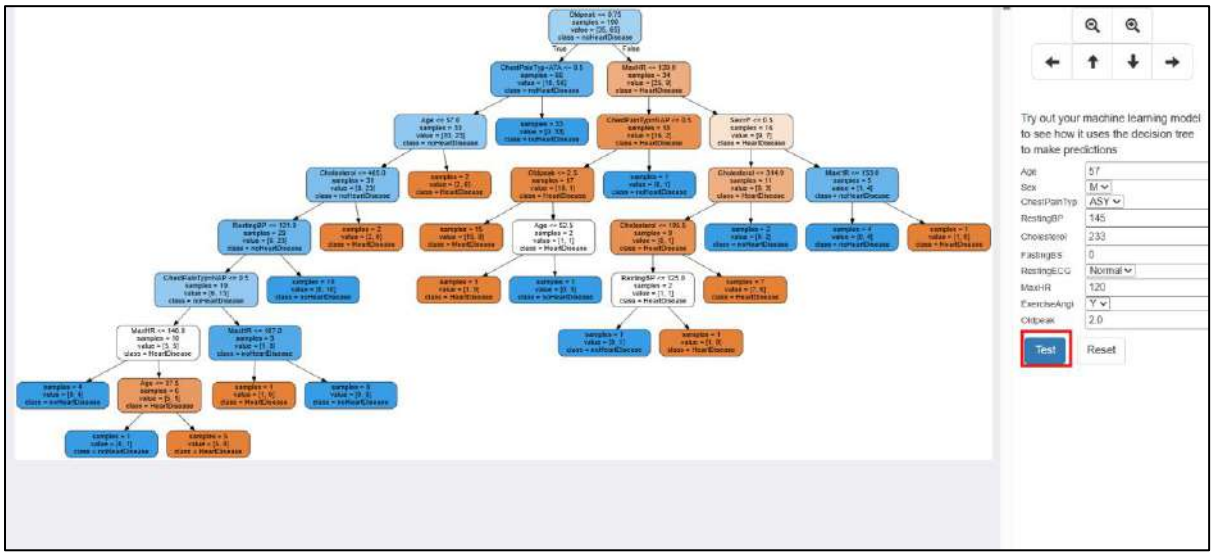




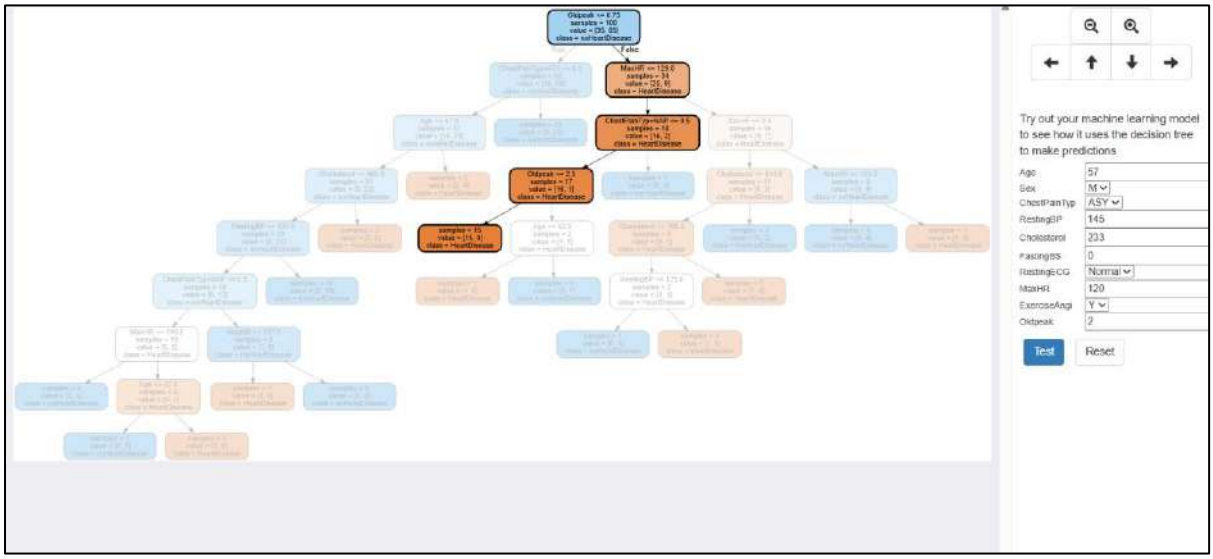
- You can also test your model. Download the file placed here **Test dataset**. This file contains new data that has not been used to train the model.
- Add the values of the first row in the appropriate fields and click test, the value on the label will be the prediction of the model



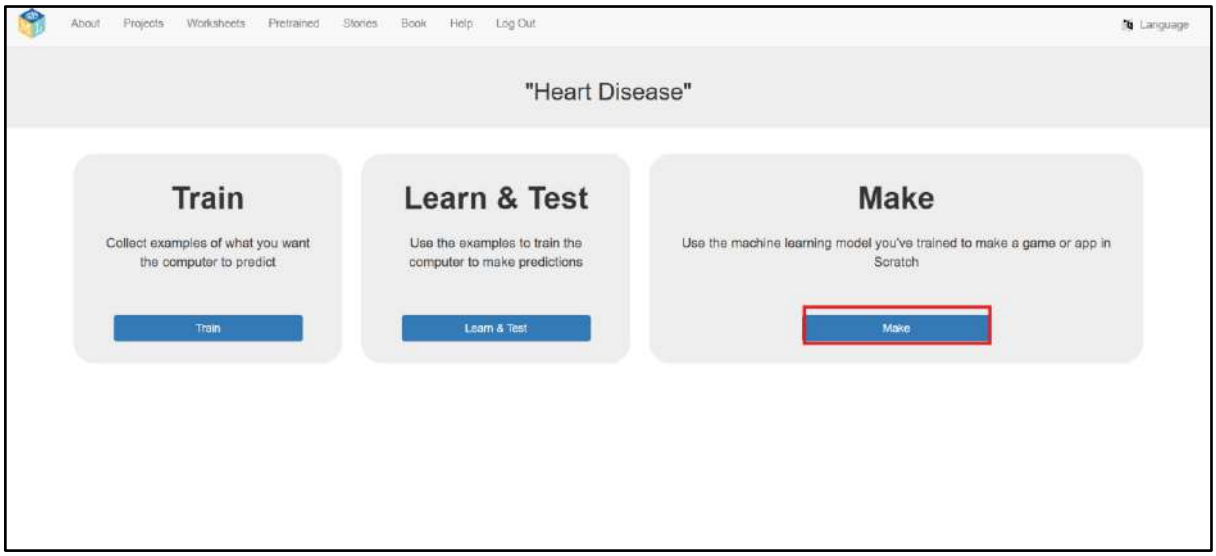
- Click on "**Test**"



- The model correctly predicts "**heartDisease**". As we can see in the **Test Dataset**, the label is also "**heartDisease**" for this data row

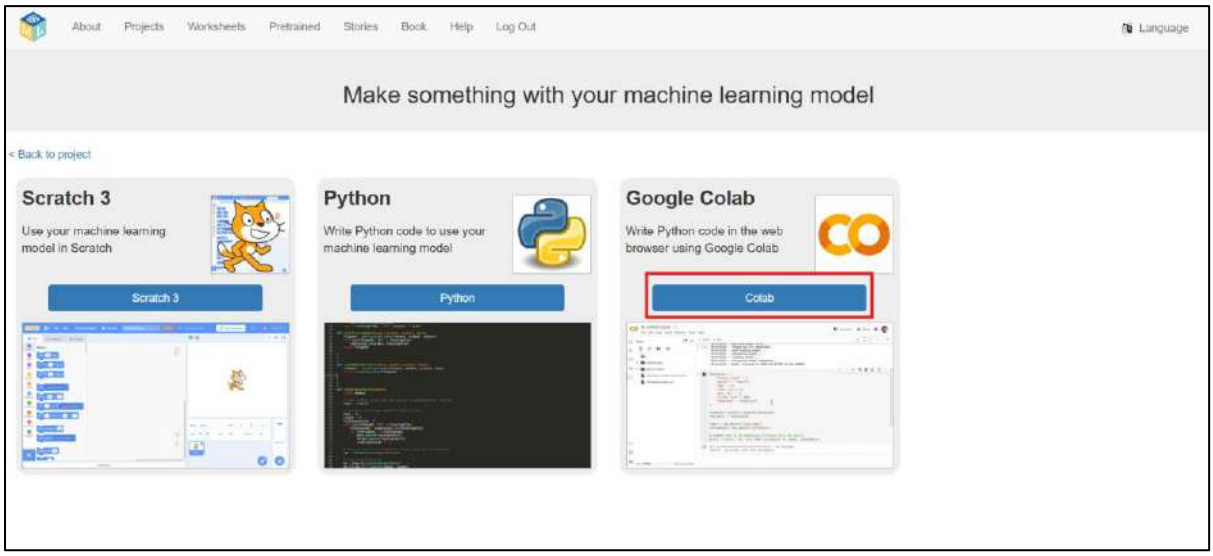


- Click on "**Back to model**"
- Click on "**Back to project**"
- Click "**Make**"



Implementation

- Click on "**Colab**"



- Follow steps 1-7 to enter the code at the end in **colab.google**
- So you can test your model through the Python programming language