

OPEN TALKS DELL'INNOVAZIONE 2021

**AI for Good: idee, progetti e
sfide per lo sviluppo sostenibile**

—
15-06-2021 // h. 11:00

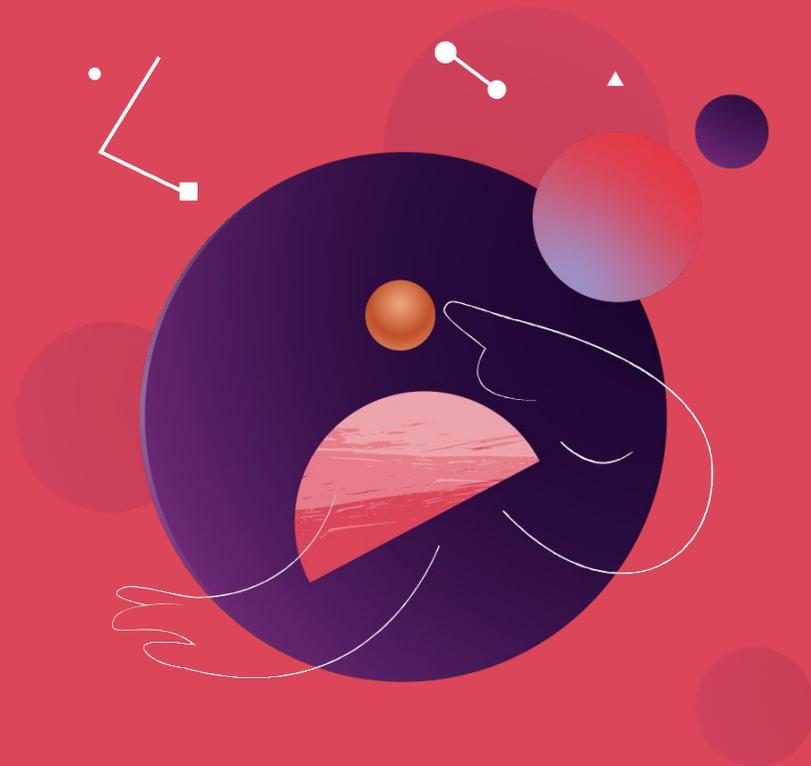


AI&ML: FROM DATA TO KNOWLEDGE

Paolo Palmerini, Gnucoop

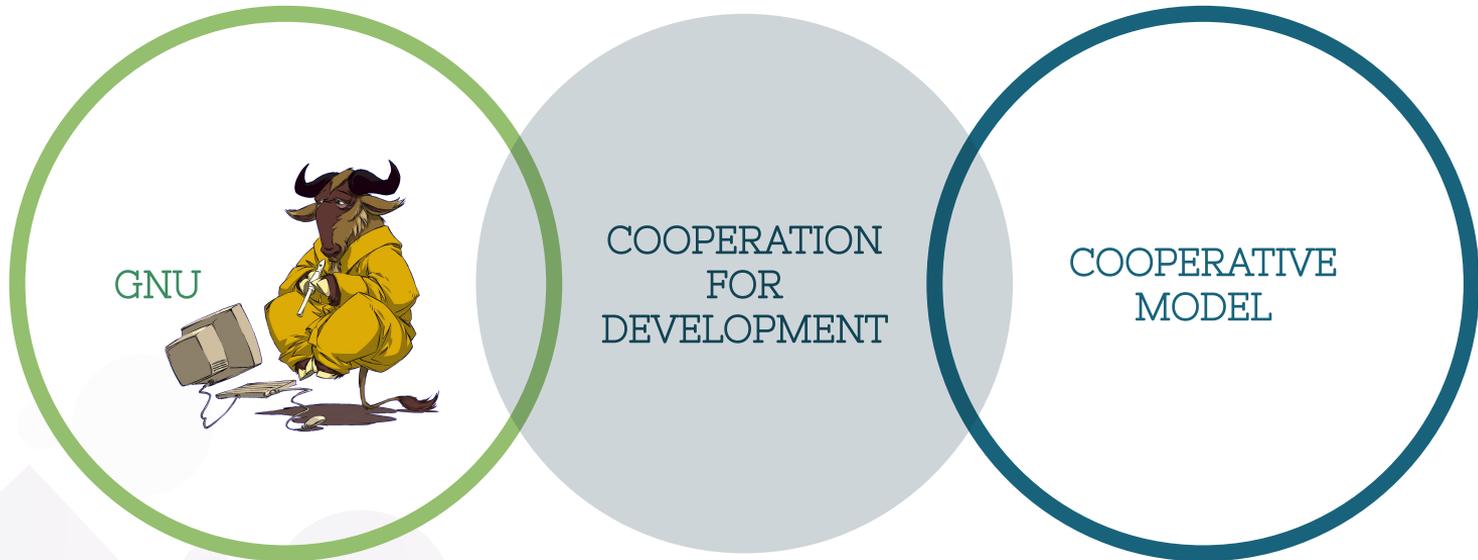
—

18-05-2021 // h. 11:00



GNUCOOP, IT FOR NON PROFIT

<https://www.gnucoop.com>





The project

Burkina Faso, Central West Region.

40% of the children are working (gold mining) and school dropout rate is very high

Objective: Increase enrollment and reduce dropout ⇒DATA!

IT as a driver for effectiveness

Primary schools in rural Burkina Faso, are often overcrowded (more than 50 child per class). Several classes are combined in the same room.

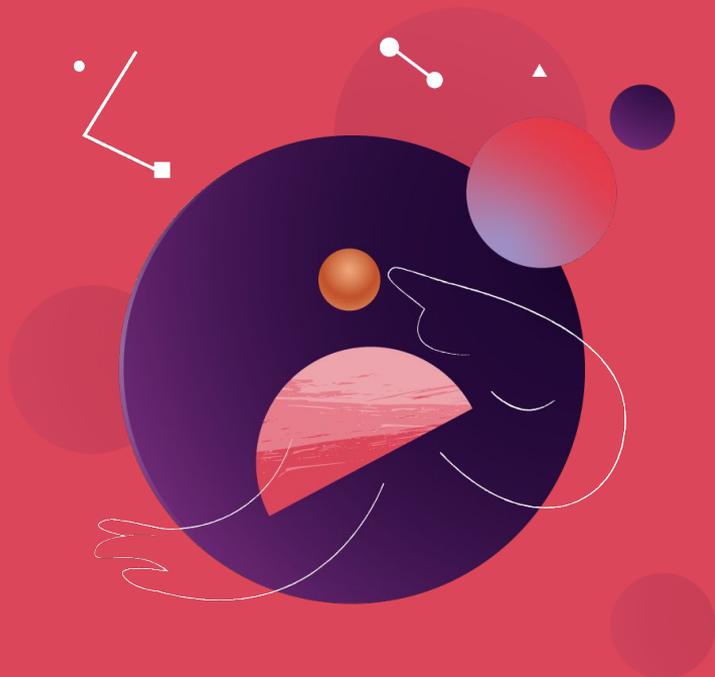
School attendance is collected manually. There is no structured and analytical sharing system.





The challenge

- 1) Easily count the number of children in a class
- 2) Correctly identify the gender (M/F)
- 3) Use affordable technology (cheap smartphone without internet connection)

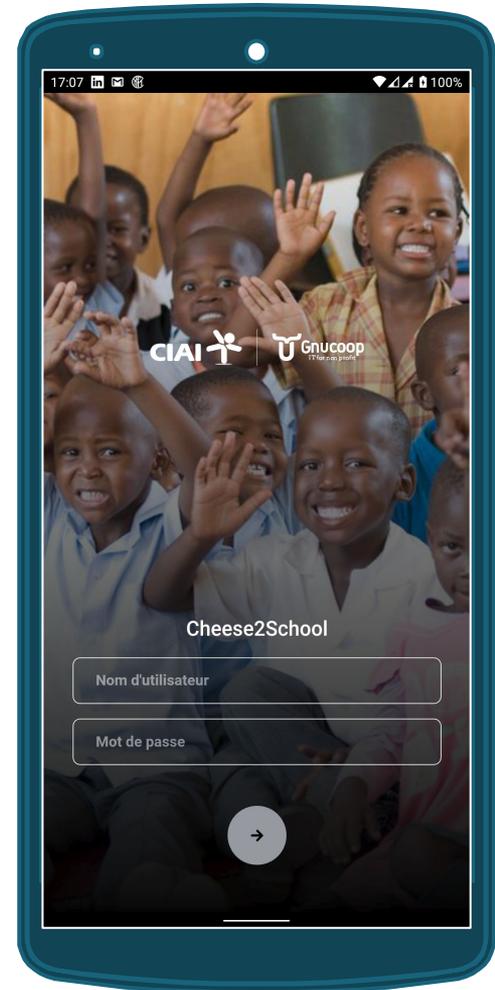


01

THE APP

Cheese2School

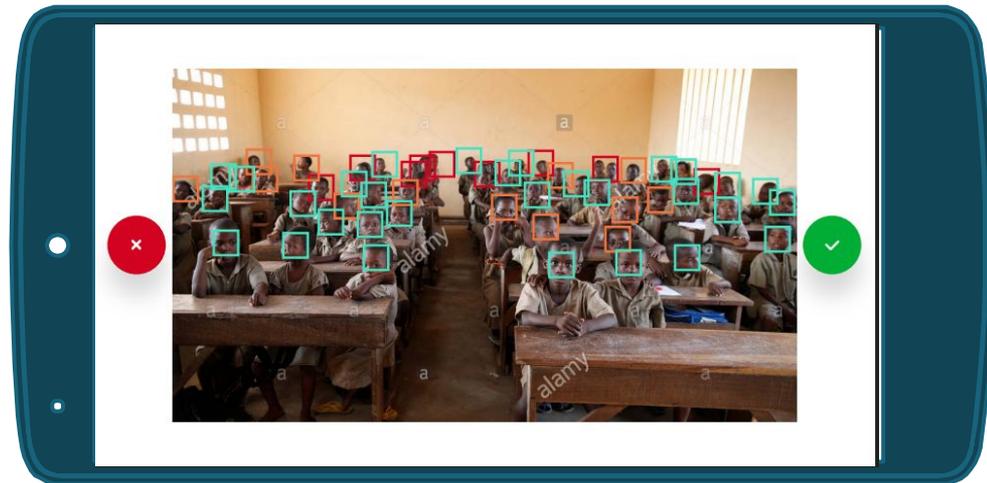
It allows the teacher/operator to collect and register school attendance through a facial recognition system.



Cheese2School1

Categorizing per gender: each of the faces is reworked using a specific artificial neural network to recognize the male or female gender

The data collected in schools will be aggregated and centralized inside a database to which educational institutions and the project committee will have access



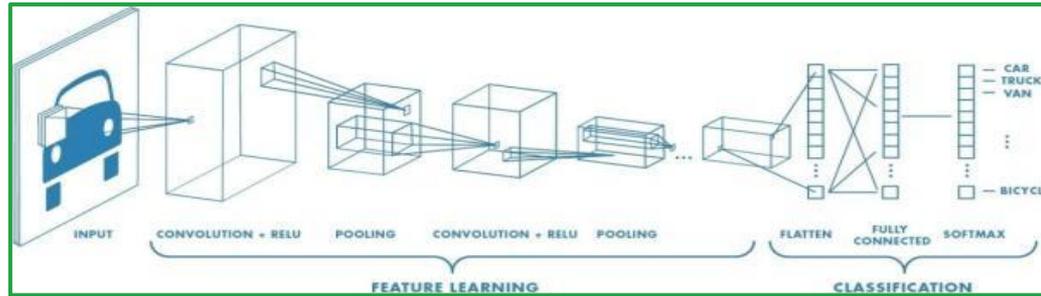
Learning

In **Deep Learning**, the classification of an image is nothing but the application of a multilayer filter to it. Each layer is information.



The strategy

Use artificial neural networks, designed on the rich open-source tools for Deep Learning.



Considering the context with constraints on both **devices power and poor connectivity**, we adopted the model **MobileNet**, proposed by Google, and framework Open Source **TensorFlow**, released by Google Brain

Neural network training

00

Step Zero

The model MobileNet was already trained by the database ImageNet, composed of more than 14 millions of images!

01

First Step

2.000 photographs taken by 2 digital cameras. One for females and one for males. Certain classification from the beginning.

02

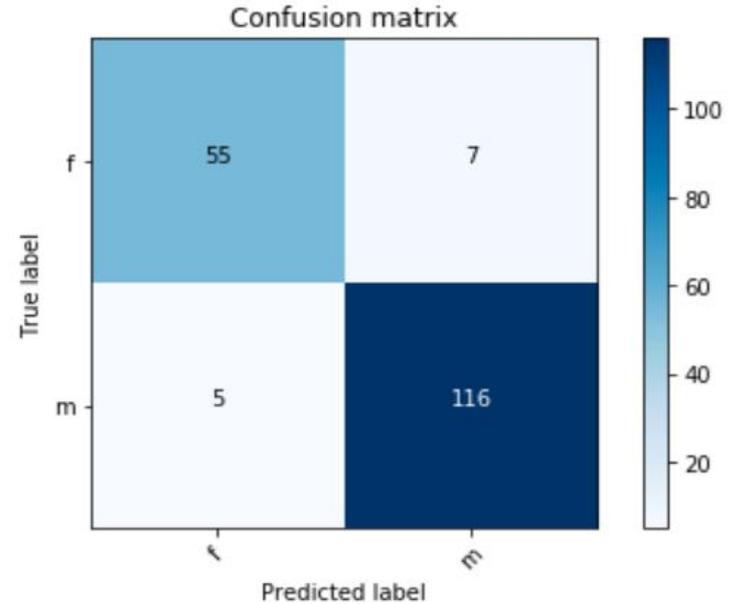
Second Step

We trained the neural network on the last 20 layers of the model using the images collected.

Testing the model

Test on a set of 183 new images

Accuracy of the results of first test 94%





02

Prediction model

A new challenge

- After the project started, we realised that we had more data than we thought
- Not “only” attendance data, but also data from a socio-economic survey used for beneficiaries selection
- We asked ourselves: analysing all this data, is it possible to “predict” school dropout?

The data we had

Beneficiaries:

1800 households

Data:

Age, study, occupation,
income, properties

Household members:

Age, sex, birth registration,
school attendance

1791 rows (beneficiaries) and 233 columns (attributes)

Data cleaning

Empty columns

Sparse rows

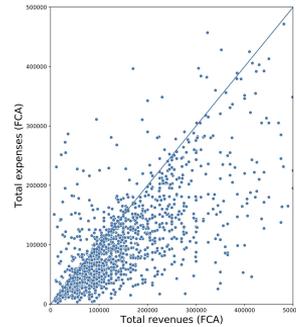
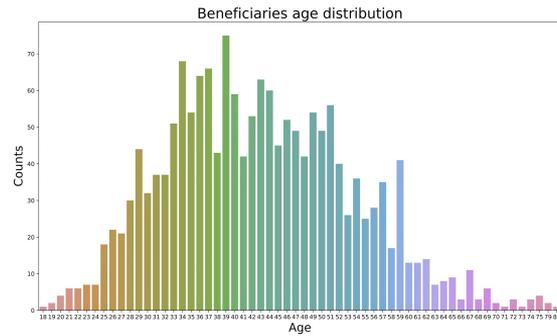
Merging of features

Noise filtering

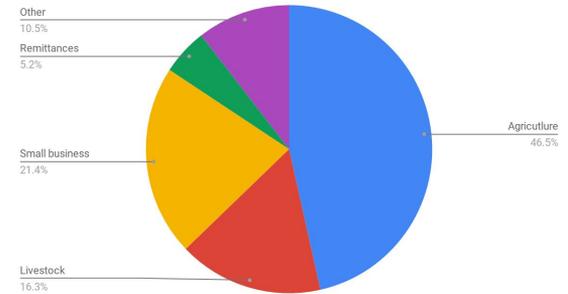


1665 rows (beneficiaries) and 45 columns (attributes)

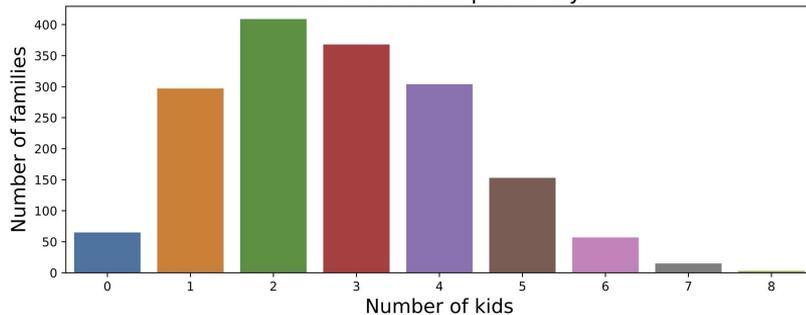
Descriptive phase



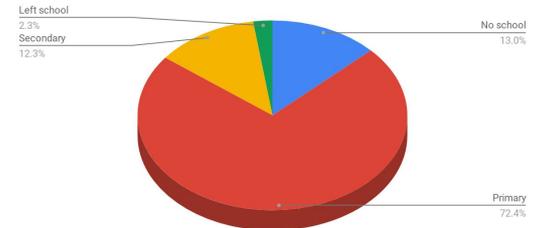
Total revenues per year



Number of kids per family



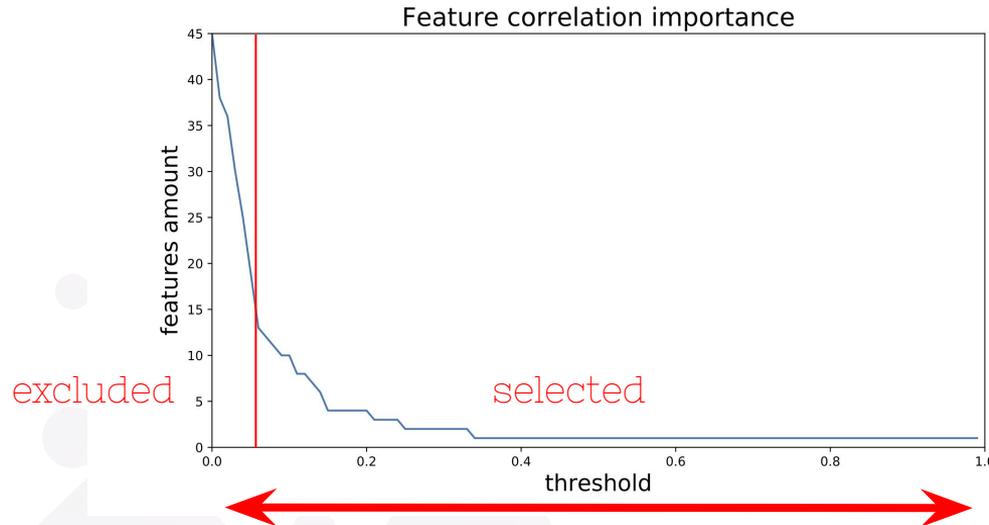
Kids with schooling age



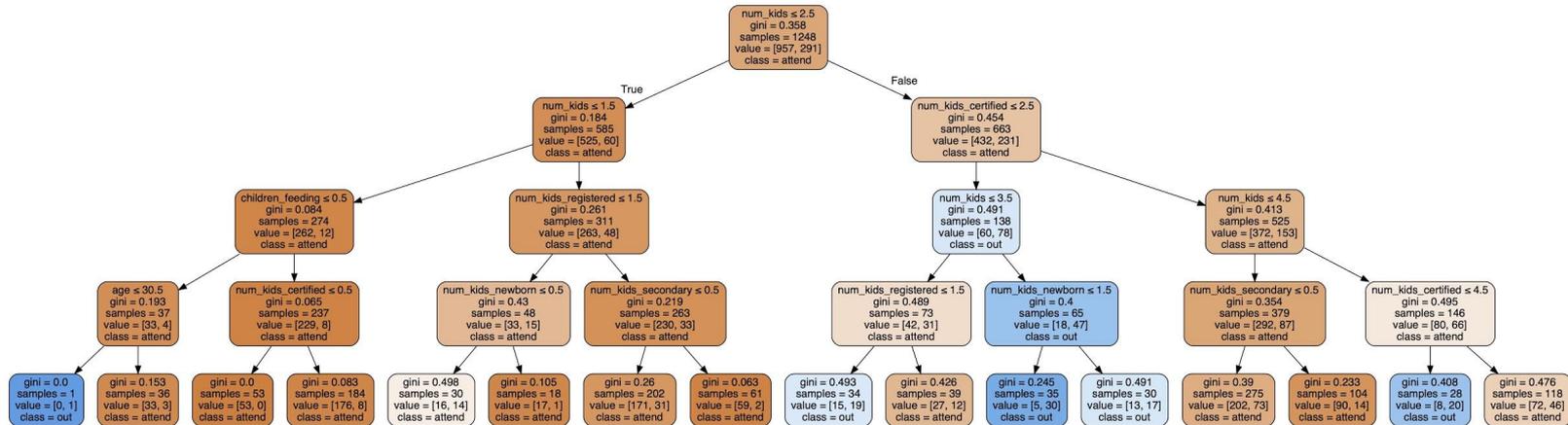
Predictive phase

Goal: find the probability that a family has out-of-school children

FEATURE SELECTION: From 45 to 15 features!



The tree-based classifier

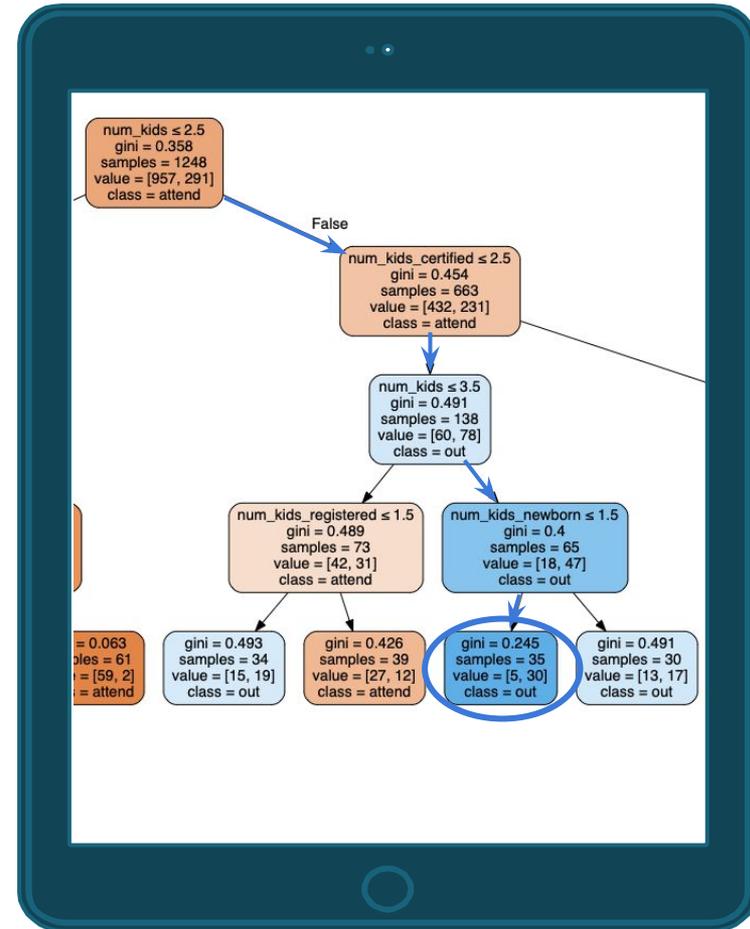


The tree gives us the probability that a child is out-of-school

The rules

If not all children in a family are registered at birth and there are infants, there is 86% chance that some of the school age children of that family will drop-out of school

We can use this information to prevent drop-out



Conclusions and lessons learned

What we have learnt so far

- 1) Technology does provide a solution to our problem
- 2) We always have more data than we think

What we still need to learn

- 1) Will this solution be actually used?
- 2) What is the actual added value of the solution? If it's only saving time...

“In God we trust, all others must bring data”

– Deming