Evaluating a label clearing AI algorithm

by Massimo Stefan

Supervisor: Prof. Passerini A.
Co-supervisors: Teso S. & Bontempelli A.

INDEX

ML: Sequential learning

Sequential learning is a type of learning in which one part of a task is learnt before the next.

Example: Understanding how a dog looks like receiving one image at the time

Problem: Label noise

1

2

Overall:

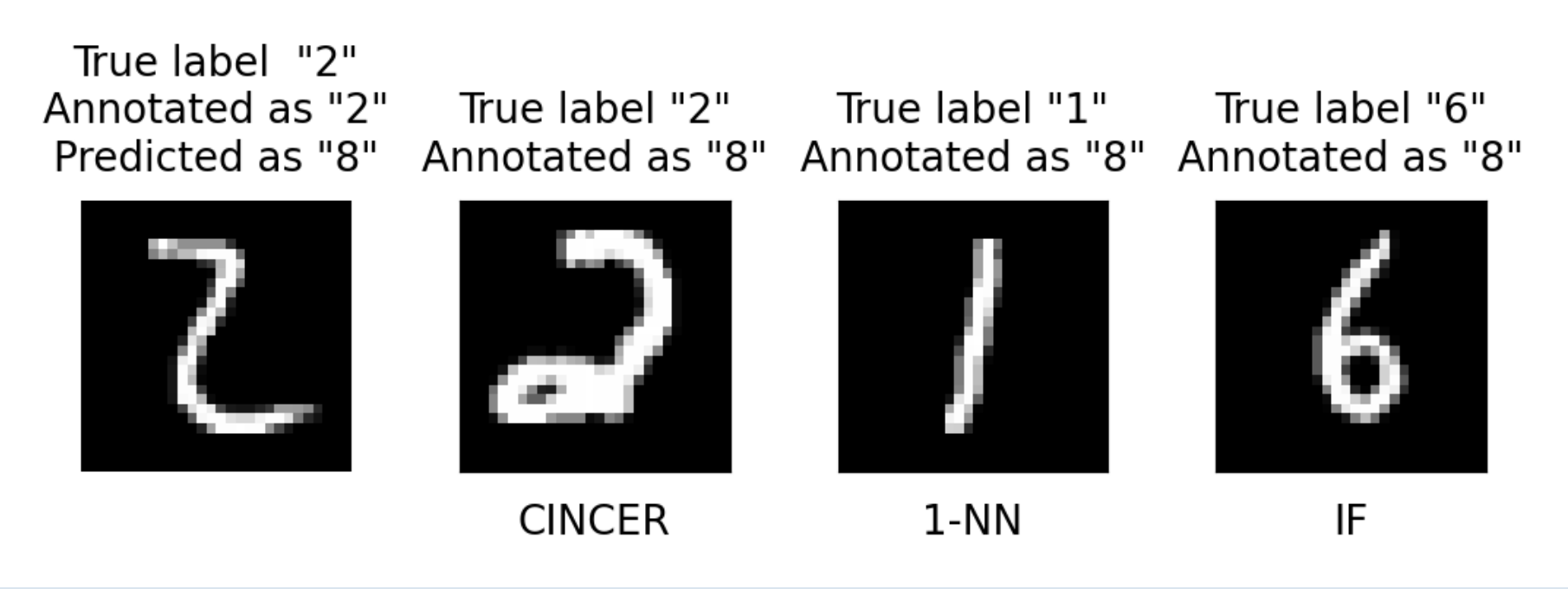
- decrease model performances predictions
- increase the complexity of learning model

Solution? Skeptical learning (SKL)

1. Recognize a suspicious example
2. Identify an example responsible for the mistake (aka counter-example) **explaining why** it has been chosen
3. Correct the example and/or the counter-example

*New features from CINCER

What is a counter-example?



Algorithms description

1-NN (Nearest Neighbor)

- The simplest one
- Uses the Euclidian distance image's pixels
- Counter-examples not necessarily useful to the machine

IF (Influence function)

- Focus on counter-examples that influence the most the machine skepticism
- Too slow for an interactive enviroment

CINCER

- Approximation of IF
- Substitute the Hessian matrix with the Fisher information Matrix (FIM)
- Quicker than IF, suitable in an interactive enviroment

Algorithm 1 Pseudo-code of CINCER. **Inputs**: initial (noisy) training set D_0; threshold τ

fit θ_0 on D_0
for $t = 1, 2, \ldots$ **do**
  receive new example $\tilde{z}_t = (x_t, \tilde{y}_t)$
  if $\mu(\tilde{z}_t, \theta_{t-1}) < \tau$ **then**
    $D_t \leftarrow D_{t-1} \cup \{\tilde{z}_t\}$ ▷ $\tilde{z}_t$ is compatible
  else
    find counterexample z_k using Eq. 2.11 ▷ $\tilde{z}_t$ is suspicious
    present $\tilde{z}_t, z_k$ to the user
    receive possibly cleaned labels y'_t, y'_k
    $D_t \leftarrow (D_{t-1} \backslash \{z_k\}) \cup \{(x_t, y'_t), (x_k, y'_k)\}$
  fit θ_t on D_t

[User] This is a cat

[AI] Ok, it seems a cat to me too

[User] This is a cat

[AI] Not sure... Double-check it and this other example

[User] Yeah, they are both dogs

[AI] All right. Updating...

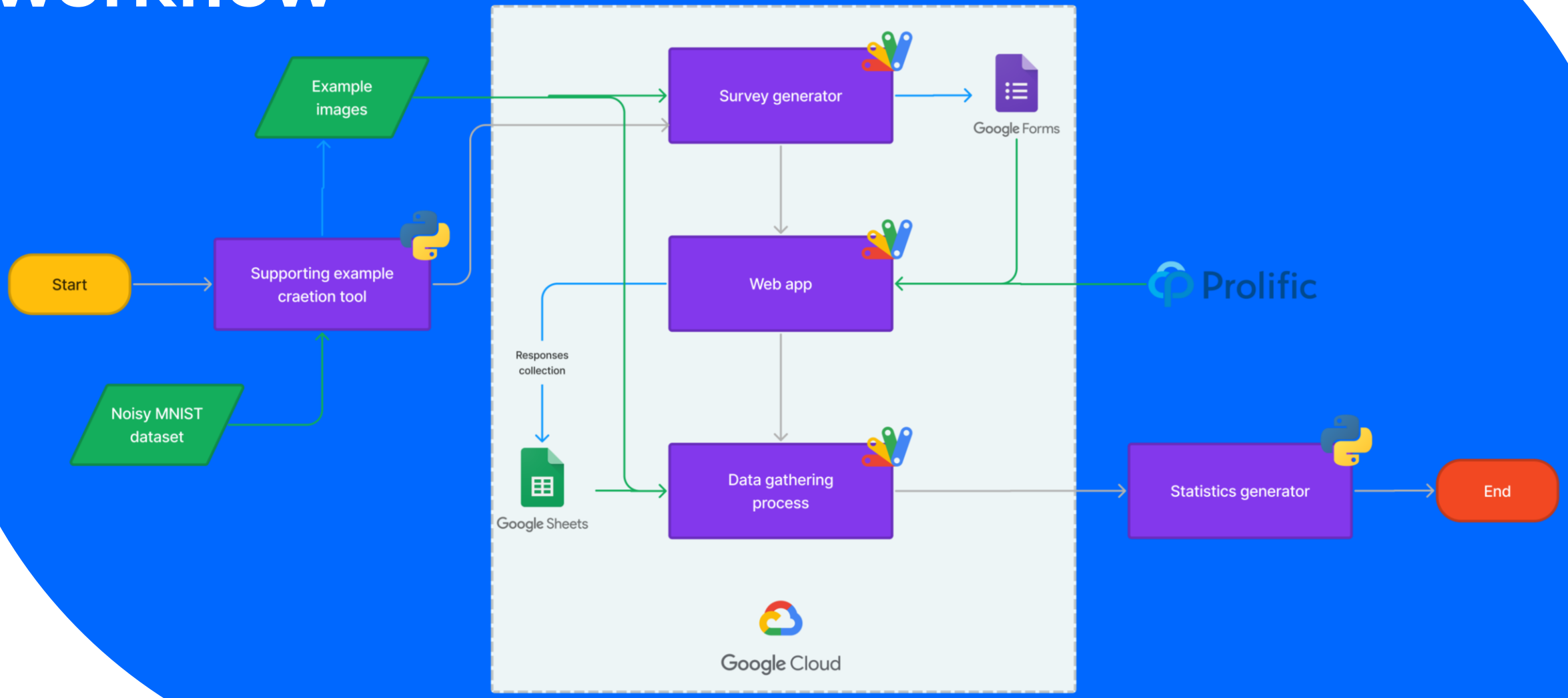
The missing gap

The **evaluation** of CINCER has been only simulated!

My internship fill this gap.

I developed a whole process to prove the effectiveness of CINCER in a real enviroment, comparing the performances of **CINCER**, **IFs**, and **1-NN** with 100 real users.

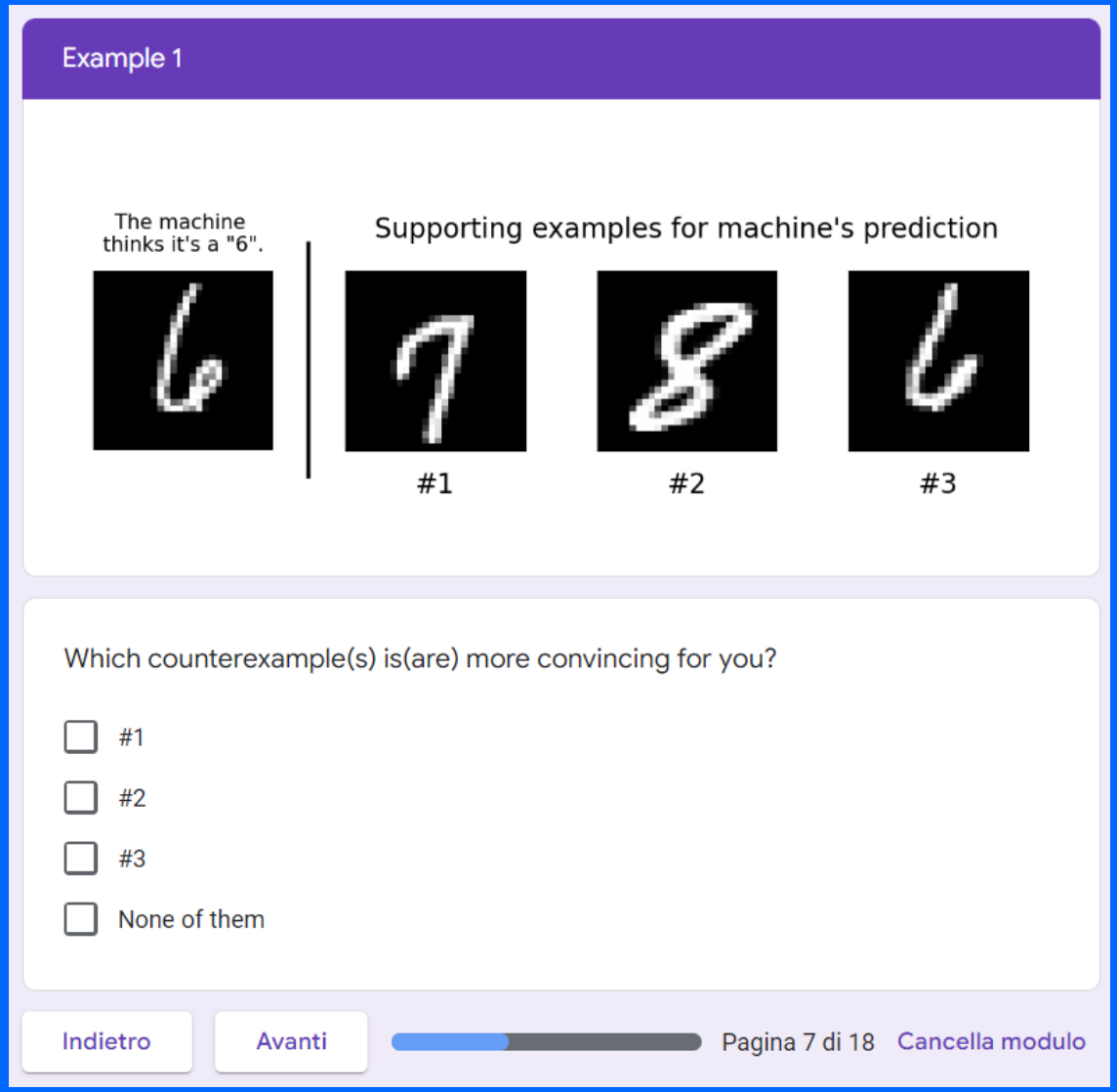
Project workflow



Example question

Structure:

- No algorithm to counterexample reference
- Multiple-choice question
- Sharp separation between example and counter-examples



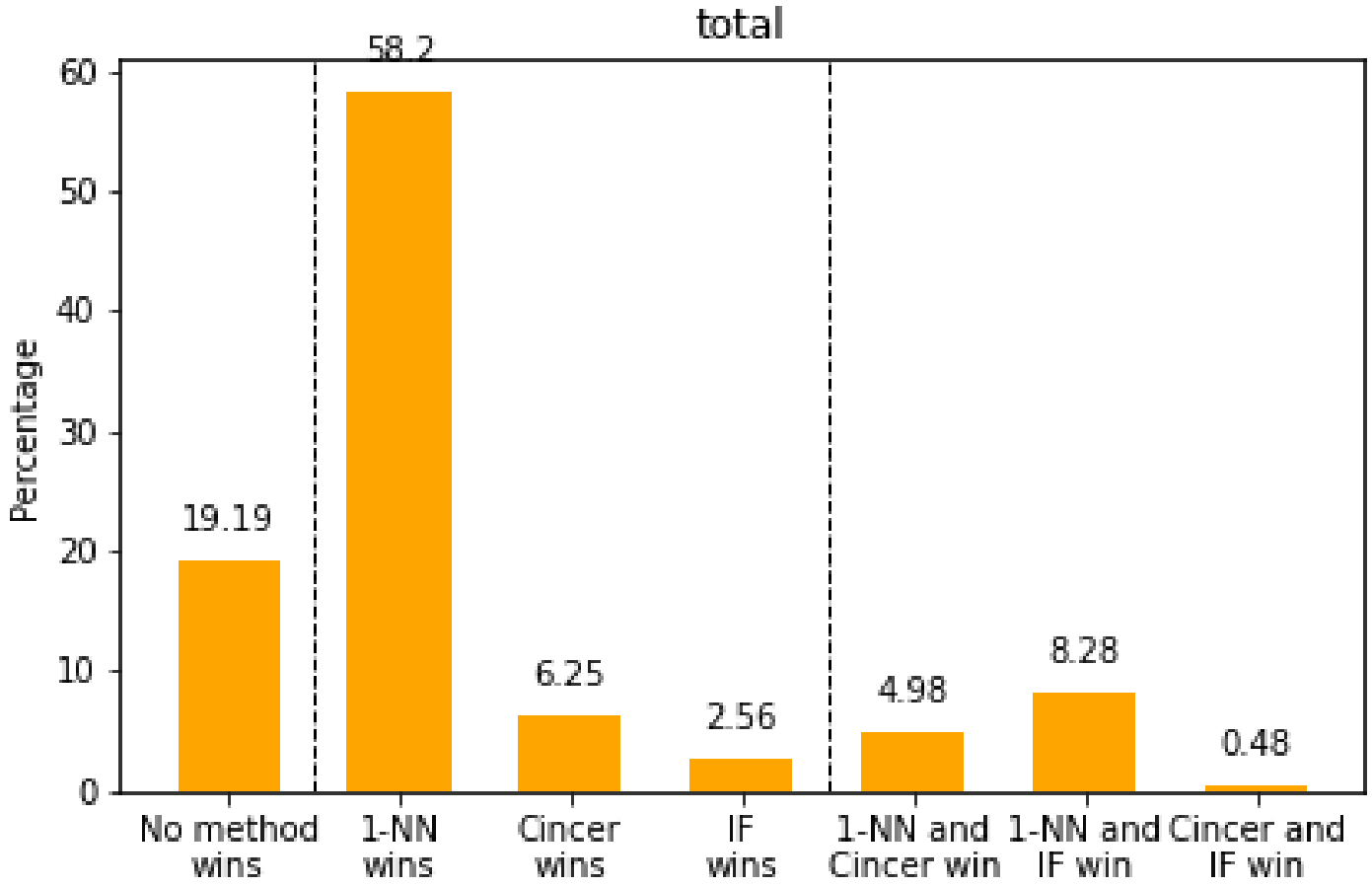
Data gathered

Structure:

- Algorithm to choice link
- Image link
- Anonymized data

	A	B	C	D	E	F
1	12	Please indicate your consent before proceeding:	All clear?	828_I98_188_C68 329__31	315_135_C65_I65 857__21	118_I08_118_C2
2	5/10/2022 21:02	✅ I consent, begin the study	👍 I have understood	#2	None of them	None of them
3	5/16/2022 16:02	✅ I consent, begin the study	👍 I have understood	#2	None of them	#2
4	5/16/2022 16:03	✅ I consent, begin the study	👍 I have understood	#2	#1	#2
5	5/16/2022 16:03	✅ I consent, begin the study	👍 I have understood	#2, #3	#3	None of them
6	5/16/2022 16:03	✅ I consent, begin the study	👍 I have understood	#2	#2	#1
7	5/16/2022 16:03	✅ I consent, begin the study	👍 I have understood	#2	#1	#2
8	5/16/2022 16:03	✅ I consent, begin the study	👍 I have understood	#2	#1	#2
9	5/16/2022 16:04	✅ I consent, begin the study	👍 I have understood	#2	#1	#2
10	5/16/2022 16:04	✅ I consent, begin the study	👍 I have understood	#2	#1	#3
11	5/16/2022 16:04	✅ I consent, begin the study	👍 I have understood	#2	#1	#2
12	5/16/2022 16:04	✅ I consent, begin the study	👍 I have understood	#2	#1	#2
13	5/16/2022 16:07	✅ I consent, begin the study	👍 I have understood	#2	#1	#2
14						

Results: average selection over all the examples



- 1-NN counter-examples are the closest to the human eye
- CINCER counter-examples are more interpretable in respect to IF's
- No instance where the 3 methods are selected together