Expert assessment vs. machine learning algorithms: juvenile criminal recidivism in Catalonia

Songül Tolan (JRC), Carlos Castillo (UPF), Marius Miron (JRC), Emilia Gómez (JRC)

Algorithms & Society Workshop, Brussels, 10 December 2018

Joint Research Centre
Universitat Pompeu Fabra

European Commission

Why use ML methods in criminal justice?

- Judge decisions are affected by extraneous factors
  [Danziger et al., 2011; Chen, 2016]
- Algorithms are not affected by cognitive bias
- There can be welfare gains: ML flight risk evaluation can yield substantial reductions in crime rate (with no change in jailing rate) or jailing rates (with no increase in crime rates)
  [Kleinberg et al., 2017]

Why NOT use ML methods in criminal justice?

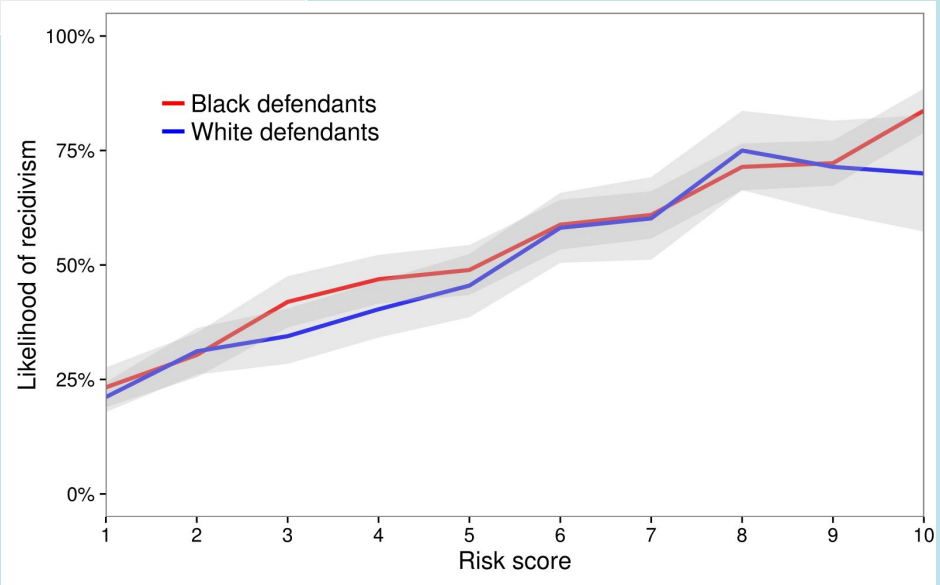
- Machines can inherit human biases through biased data
  - [Barocas and Selbst, 2016]
- In many cases their outputs cannot be explained, so how can we justify?
- "They" can be racist
- There is a need for "fair" ML

Fairness in ML: the case of COMPAS

- ProPublica: COMPAS is unfair! [Angwin et al., 2016]

	White	African American
Predicted higher risk, did not reoffend	23.5%	44.9%
Predicted lower risk, did re-offend	47.7%	28.0%

- NorthPointe: COMPAS is fair!



Corbett-Davies et al., 2017

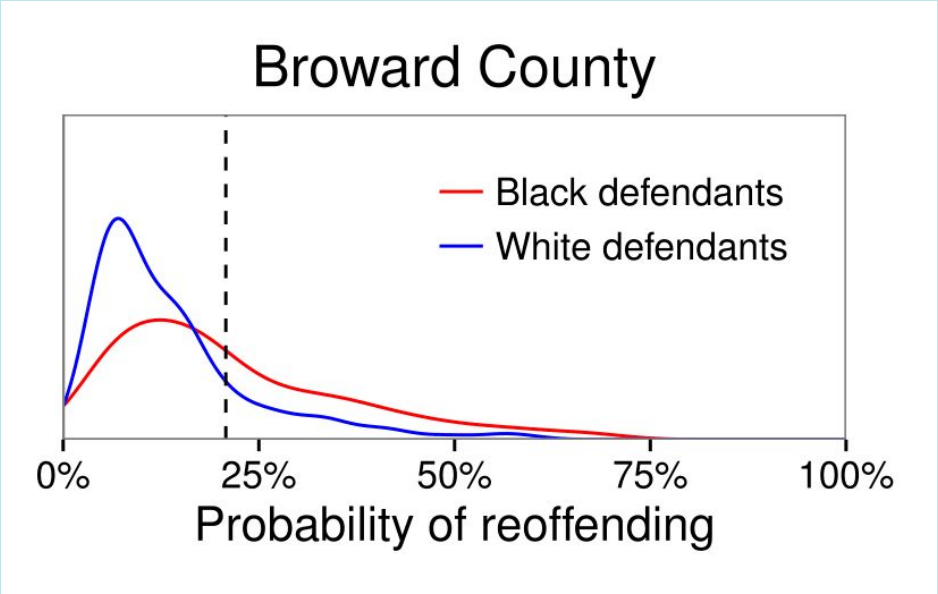
European Commission

Fairness in ML: the case of COMPAS

Impossibility proofs: When base rates differ (in Broward County 51% vs. 39%), you cannot achieve calibration and equal FPR/FNR at the same time
[Kleinberg et al., 2016; Chouldechova, 2017]

Also:

- No single threshold equalizes both FPR and FNR
  - Direct vs. indirect discrimination
- Imposing any fairness criterion has a cost in terms of public safety or defendants incarcerated
- Literature on fairML grows rapidly, but all based on US data



Corbett-Davies et al., 2017

What we do

- Look at European example: SAVRY in Catalonia
- We evaluate SAVRY against ML methods in terms of fairness and predictive performance
- We show some evidence that ML methods of risk assessments introduce unfairness and that their use in criminal justice should be fairness-aware

SAVRY

- Structured Assessment of Violence Risk in Youth (SAVRY)
- Structures Professional Judgement
- Also used to assess the risk of (not only violent) crimes upon release
- Used to inform decisions on interventions
- Sample: Catalonia, 4752 youths aged 12-18, 855 with SAVRY, committed crime between 2002-2010, released in 2010, recidivism by 2015

SAVRY ≠ COMPAS

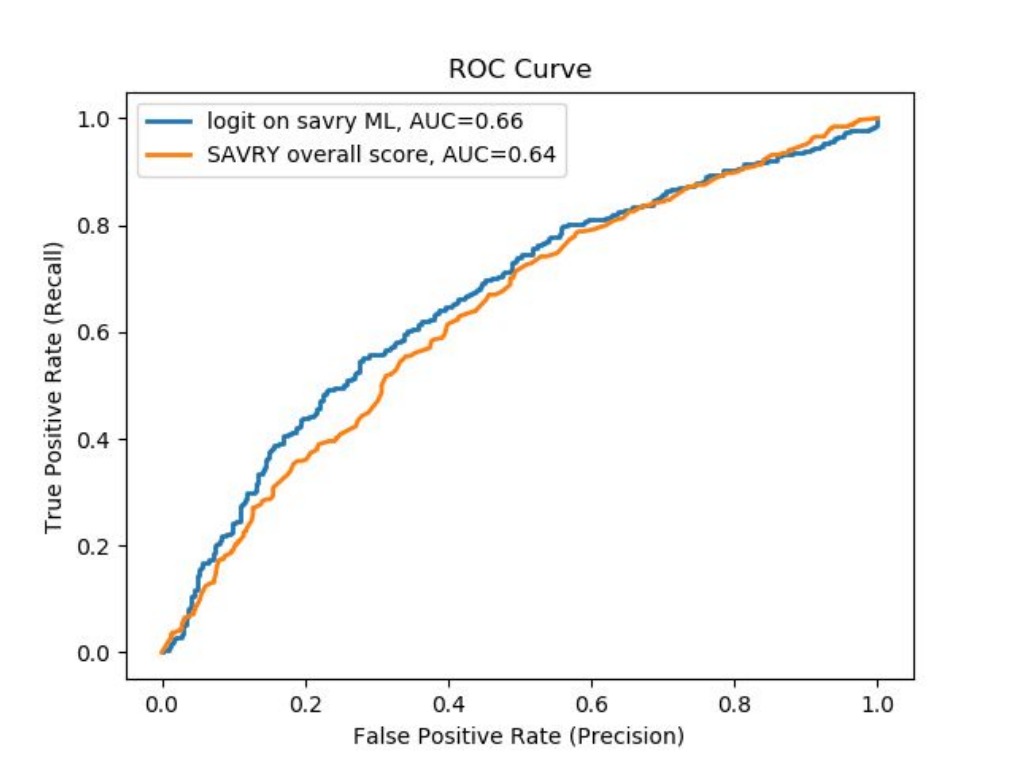
- Detailed and transparent risk assessment
- Based on 6 protective factors
- Based on 24 risk factors: Historical, Social/Contextual, Individual
- We evaluate the sum of 24 risk factors (low, medium, high) against ML methods

Base rates differ

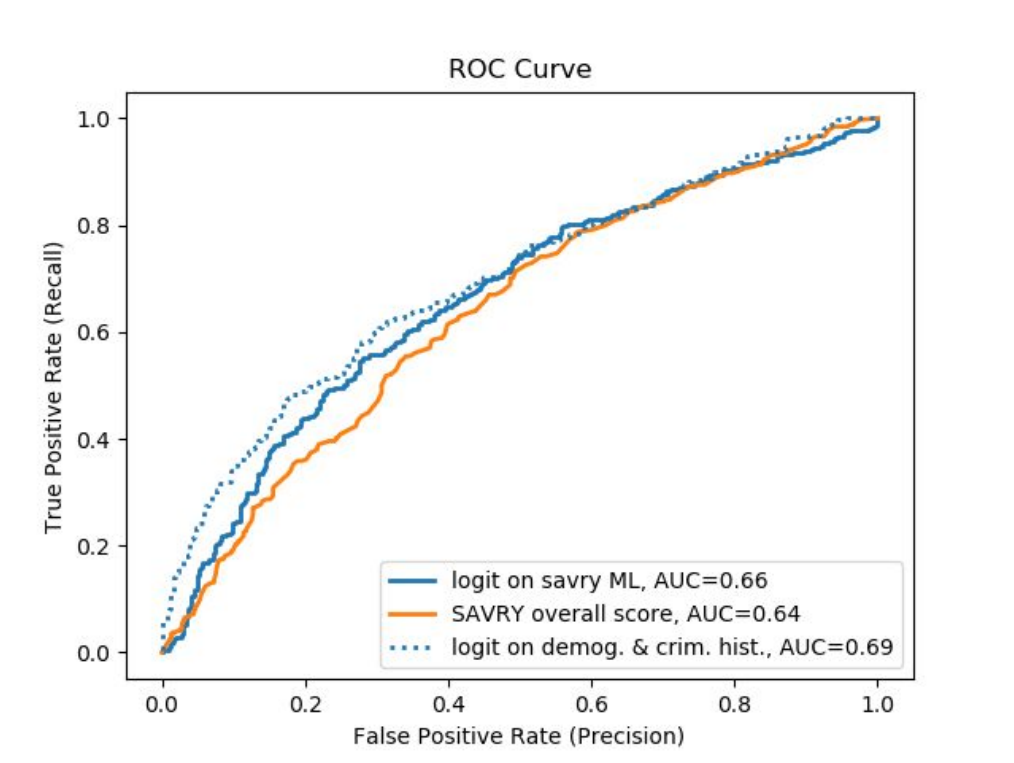
	Recidivated		Not Recidivated		Difference	
	Mean	Std.Dev.	Mean	Std.Dev.	Diff	Std.Dev.
savry	0.20	(0.40)	0.17	(0.38)	-0.03**	(0.01)
female	0.1	(0.31)	0.2	(0.41)	0.1***	(0.0)
foreign	0.44	(0.50)	0.32	(0.47)	-0.12***	(0.01)
national group						
central/south	0.17	(0.38)	0.15	(0.35)	-0.03**	(0.01)
american						
EU	0.05	(0.22)	0.05	(0.22)	0.00	(0.01)
magribian	0.2	(0.40)	0.1	(0.30)	-0.1***	(0.0)
age maincrime	15.59	(1.07)	15.92	(1.07)	0.33***	(0.03)
2 prior	0.21	(0.40)	0.21	(0.41)	0.00	(0.01)
3+ prior	0.10	(0.30)	0.09	(0.28)	-0.01	(0.01)
violent maincrime	0.52	(0.50)	0.54	(0.50)	0.02	(0.02)
action mediat+rep	0.28	(0.45)	0.33	(0.47)	0.05***	(0.01)
action execution	0.29	(0.45)	0.35	(0.48)	0.06***	(0.01)
action duration	118.67	(193.55)	141.05	(216.00)	22.38***	(6.38)
N	1622		3130			

Note: Authors' calculations.

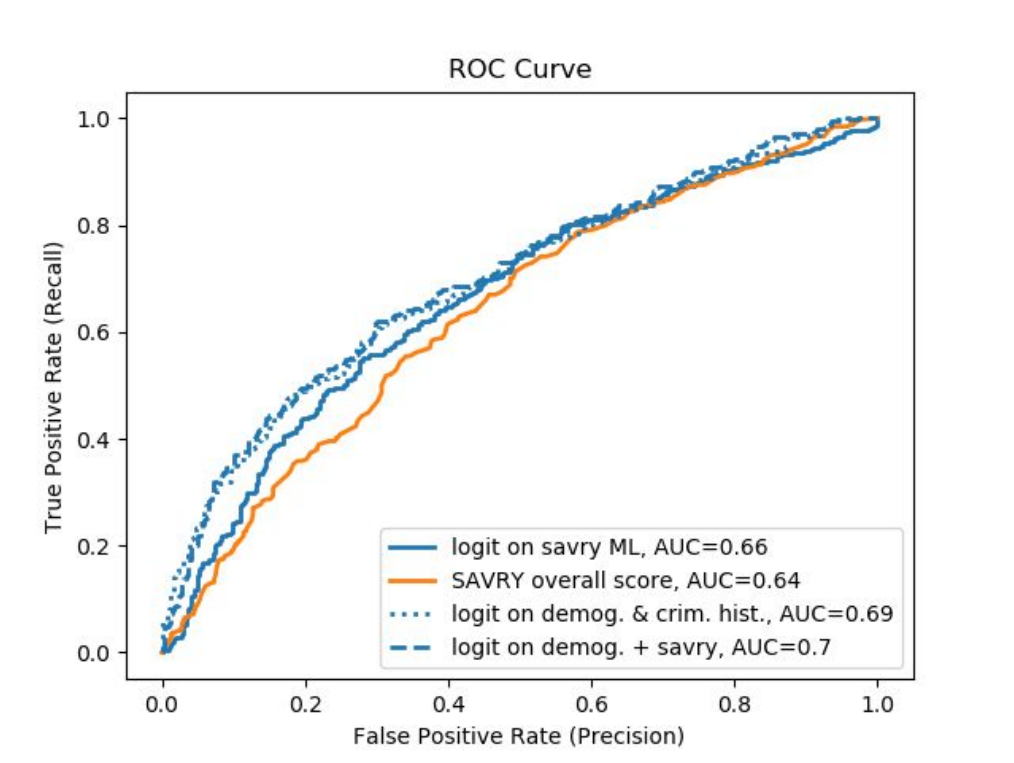
Performance



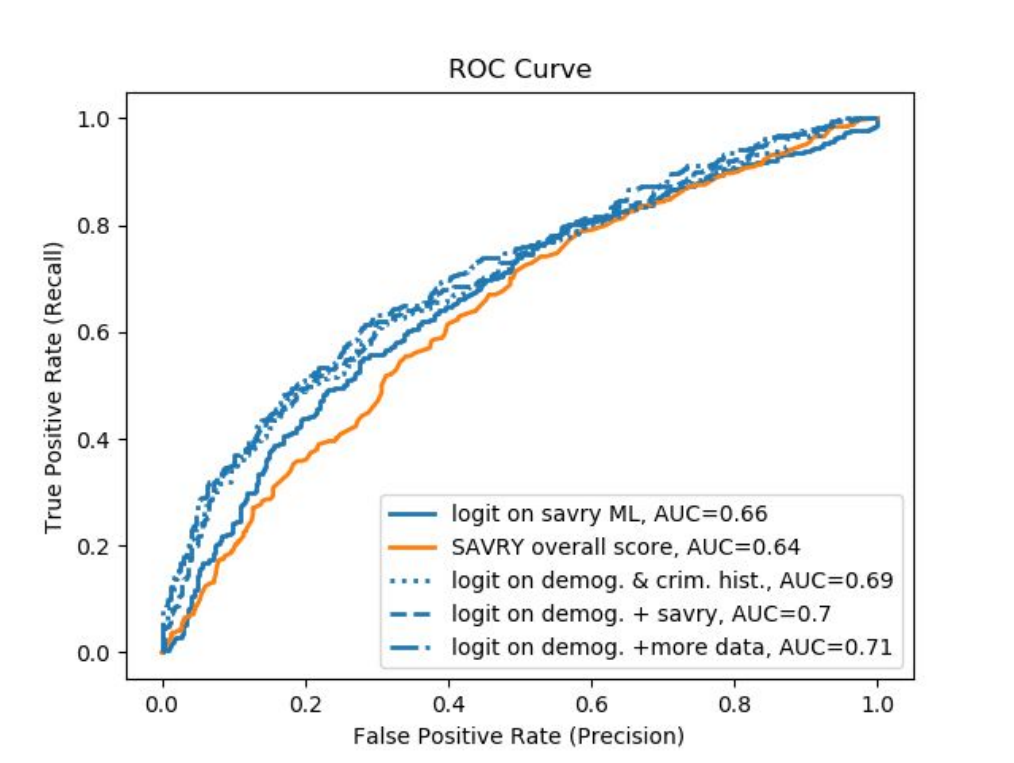
Performance



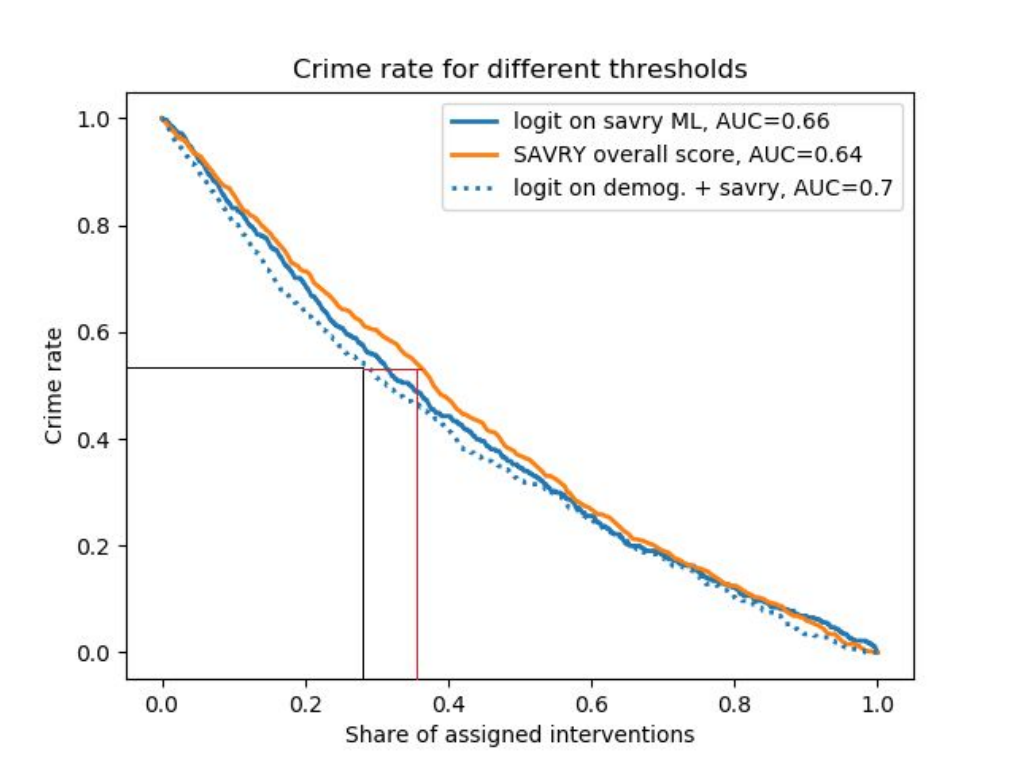
Performance



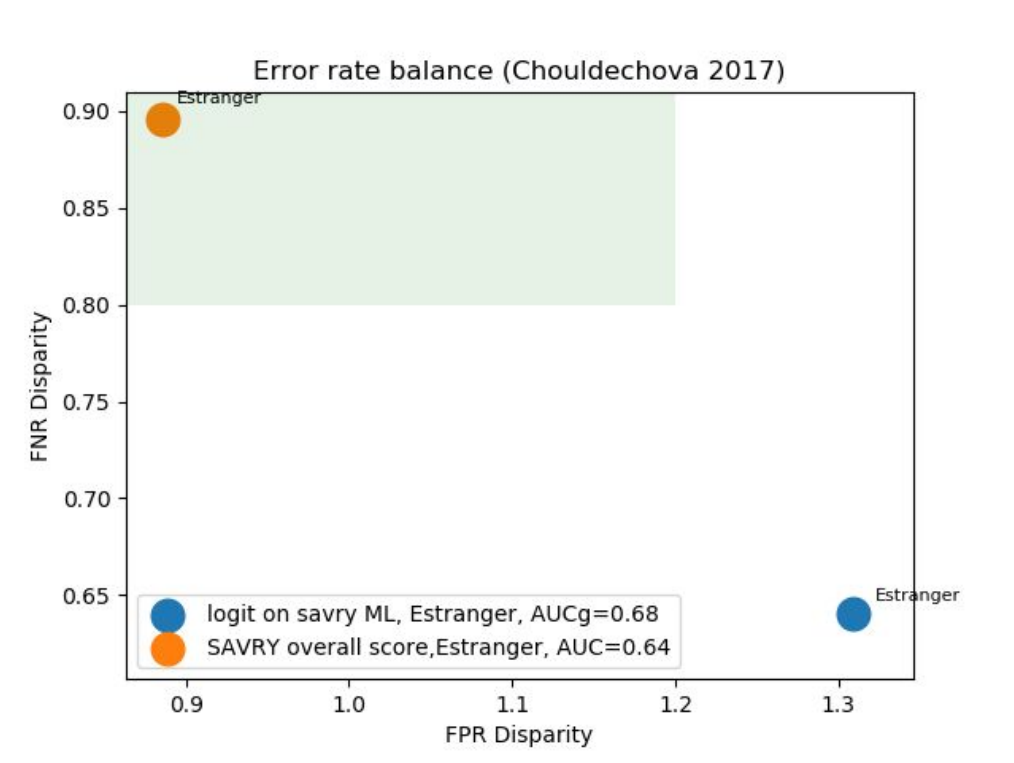
Performance



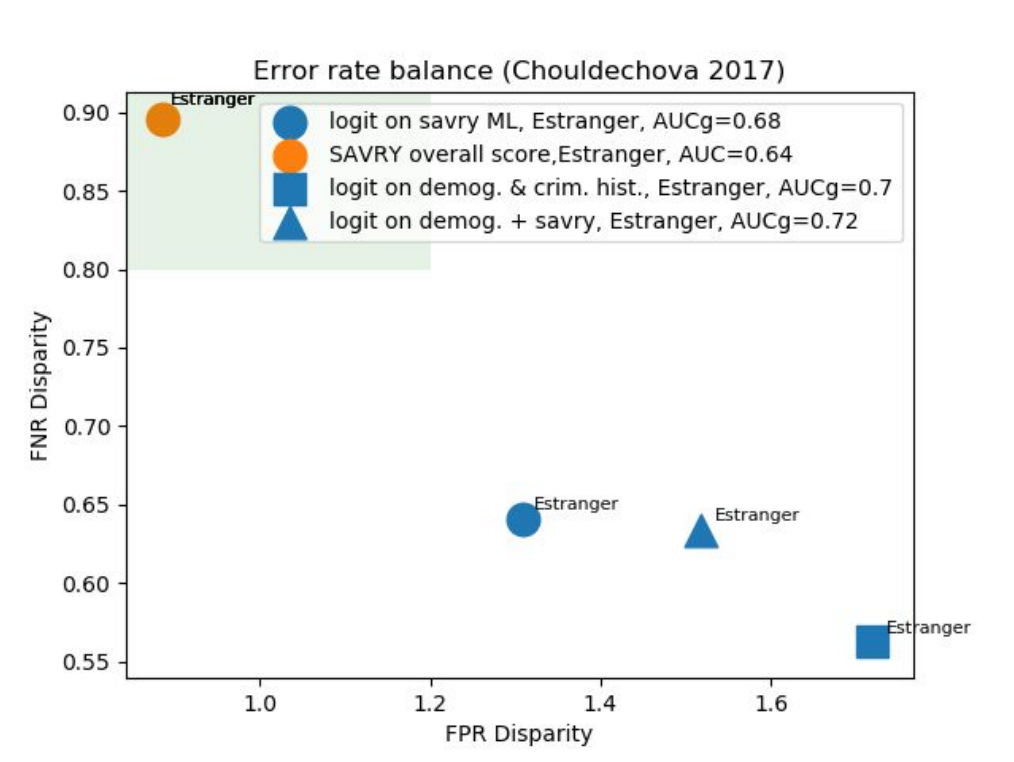
Performance



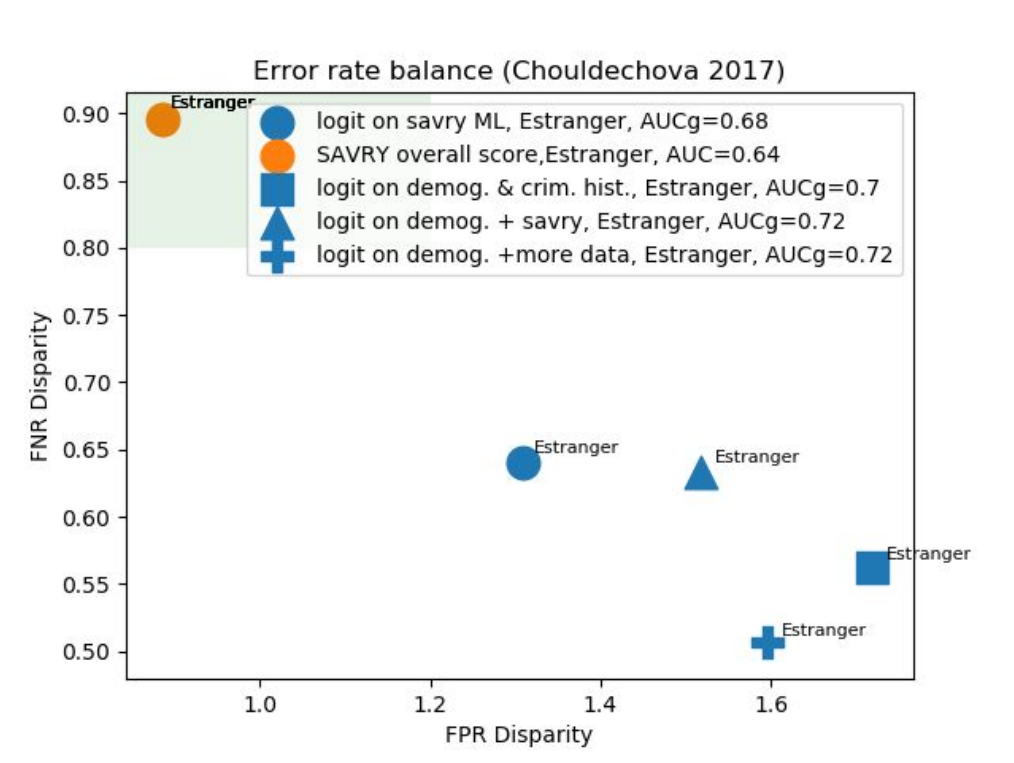
Fairness


Error rate balance (Chouldechova 2017)
Estranger
Estranger
0.90
0.85
0.80
0.75
0.70
0.65
FNR Disparity
0.9
1.0
1.1
1.2
1.3
FPR Disparity
logit on savry ML, Estranger, AUCg=0.68
SAVRY overall score,Estranger, AUC=0.64

Fairness


Error rate balance (Chouldechova 2017)
logit on savry ML, Estranger, AUCg=0.68
SAVRY overall score,Estranger, AUC=0.64
logit on demog. & crim. hist., Estranger, AUCg=0.7
logit on demog. + savry, Estranger, AUCg=0.72
Estranger
Estranger
Estranger
Estranger
FNR Disparity
FPR Disparity
0.90
0.85
0.80
0.75
0.70
0.65
0.60
0.55
1.0
1.2
1.4
1.6

Fairness


Error rate balance (Chouldechova 2017)
logit on savry ML, Estranger, AUCg=0.68
SAVRY overall score,Estranger, AUC=0.64
logit on demog. & crim. hist., Estranger, AUCg=0.7
logit on demog. + savry, Estranger, AUCg=0.72
logit on demog. +more data, Estranger, AUCg=0.72
Estranger
Estranger
Estranger
Estranger
Estranger
FNR Disparity
FPR Disparity
0.90
0.85
0.80
0.75
0.70
0.65
0.60
0.55
0.50
1.0
1.2
1.4
1.6

Summary and Outline

- ML yields a more precise risk assessment
- When base rates differ, ML methods have to be fairness aware
- Use rich information:
  - for a transparent mitigation of unfairness
  - to adjust features that have a substantial effect on increasing unfairness
  - to refocus analysis away from tensions/tradeoffs towards better targeted interventions
- Further Analysis on human-algorithm interaction: RisCanvi

Thank you!

Any questions?

You can find me at songul.tolan@ec.europa.eu

Find HUMAINT at https://ec.europa.eu/jrc/communities/community/humaint
Find Carlos at http://chato.cl/

European Commission