



How (not) to talk about your ML project

Martyna Urbanek-Trzeciak

FANDOM

WiMLDS meeting, 10th December 2018



How to talk better in ML project

Martyna Urbanek-Trzeciak

FANDOM

WiMLDS meeting, 10th December 2018

Who am I?

- Data Analyst in FANDOM
 - working mainly with Ad Engineering
- Moved from biomolecular science to data science
 - PhD in biochemistry
 - MSc in Computer Science

Our team

- Not a data science team!
- Product Manager (likes ML & AI)
- Full-stack Engineers
- QA Engineer
- Data Analyst (and now two!)

What is the message here?



Communication...

communication...

a bit of domain knowledge...

and...

communication

Table of Content

- Story of communication in ML projects
 - Episode 1: First model evaluation
 - Episode 2: How to compare multiple models
- Learning points

Example project

If user will not go to the next page?

Can we make her stay?

Binary classification problem

Episode 1 - We have a model!

- Is it good?
- How do we evaluate it?
 - **Precision/recall/f1 score**
 - **Confusion matrix (False positives vs False negatives)**
 - Benchmark

Let's talk...

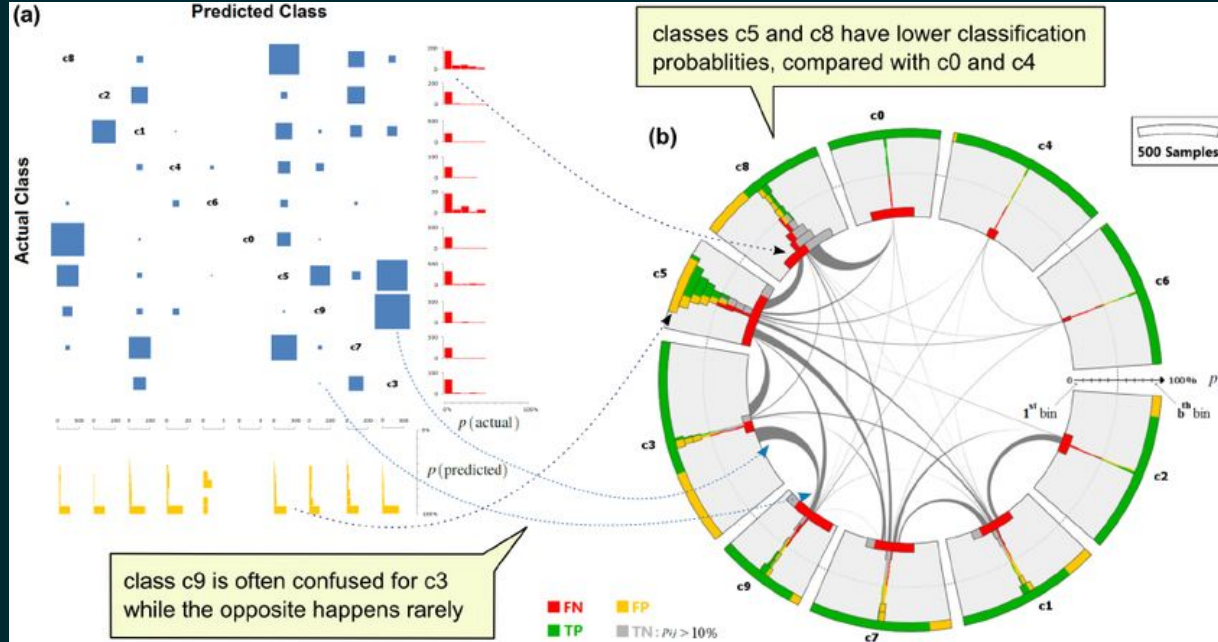
	precision	recall	f1-score	support
0	0.92702	0.79973	0.85869	25541
1	0.50470	0.76422	0.60792	6820
micro avg	0.79225	0.79225	0.79225	32361
macro avg	0.71586	0.78198	0.73330	32361
weighted avg	0.83802	0.79225	0.80584	32361

[[20426 1608]
[5115 5212]]
[1689 4]]

10

[illegible]

Let's talk...



Let's talk...

	precision	recall	f1-score	support
0	0.92702	0.79973	0.85869	25541
1	0.50470	0.76422	0.60792	6820
micro avg	0.79225	0.79225	0.79225	32361
macro avg	0.71586	0.78198	0.73330	32361
weighted avg	0.83802	0.79225	0.80584	32361

```
[[20426 1608]
 [ 5115 5212]]
[ 1689 4]]
```

Confusion matrix

n=165	Predicted: NO	Predicted: YES	
Actual: NO	TN = 50	FP = 10	60
Actual: YES	FN = 5	TP = 100	105
	55	110	

<https://jovianlin.io/confusion-matrix/>

Confusion matrix

n=165	Predicted: NO	Predicted: YES	
Actual: NO	TN = 50	FP = 10	60
Actual: YES	FN = 5	TP = 100	105
	55	110	

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

<https://jovianlin.io/confusion-matrix/>

<https://towardsdatascience.com/taking-the-confusion-out-of-confusion-matrices-c1ce054b3d3e>

Confusion matrix

n=165	Predicted: NO	Predicted: YES	
Actual: NO	TN = 50	FP = 10	60
Actual: YES			

	Class 1 Predicted	Class 2 Predicted
Class 1 Actual	TP	FN
Class 2 Actual	FP	TN

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

<https://jovianlin.io/confusion-matrix/>

<https://towardsdatascience.com/taking-the-confusion-out-of-confusion-matrices-c1ce054b3d3e>

<https://www.geeksforgeeks.org/confusion-matrix-machine-learning/>

Confusion matrix

n=165	Predicted: NO	Predicted: YES	
Actual: NO	TN = 50	FP = 10	60
Actual: YES			

	Class 1 Predicted	Class 2 Predicted
Class 1 Actual	TP	FN
Class 2 Actual	FP	TN

Actual Values	
	Positive (1) Negative (0)
Predicted Values	
Positive (1)	TP FP
Negative (0)	

	Not Pregnant	Pregnant
Positive Test Result	False Positive	True Positive
Negative Test Result	True Negative	False Negative

<https://jovianlin.io/confusion-matrix/>

<https://towardsdatascience.com/taking-the-confusion-out-of-confusion-matrices-c1ce054b3d3e>

<https://www.geeksforgeeks.org/confusion-matrix-machine-learning/>

<https://whatis.techtarget.com/definition/confusion-matrix>

Confusion matrix

n=165		Predicted: NO
Actual: NO		TN = 50
Actual: YES	

	Class 1 Actual
	Class 2 Actual



	Pregnant
	True Positive
	False Negative

<https://jovianlin.io/confusion-matrix/>

<https://towardsdatascience.com/taking-the-confusion-out-of-confusion-matrices-c1ce054b3d3e>

<https://www.geeksforgeeks.org/confusion-matrix-machine-learning/>

<https://whatis.techtarget.com/definition/confusion-matrix>

Episode 1 - We have a model!

- Is it good?
- How do we evaluate it?
 - Precision/recall/f1 score
 - Confusion matrix (False positives vs False negatives)
 - **Benchmark**

Benchmark (stratified coin flip)

	In reality: didn't go	In reality: did go
Prediction: will not go	9%	21%
Prediction: will go to N+1	21%	49%

Project: If user will not go to the next page?

Benchmark evaluation

	In reality: didn't go	In reality: did go
Prediction: will not go	Main focus area 9%	Big error, so wrong 21%
Prediction: will go to N+1	Lost opportunity 21%	Whatever, carry on, user

Project: If user will not go to the next page?

Results

	In reality: didn't go	In reality: did go
Prediction: will not go	16.2%	19.0%
Prediction: will go to N+1	9.5%	55.3%

Project: If user will not go to the next page?

Results vs benchmark

	In reality: didn't go	In reality: did go
Prediction: will not go	+80% (good)	-10% (good)
Prediction: will go to N+1	-55% (good)	Whatever, carry on, user

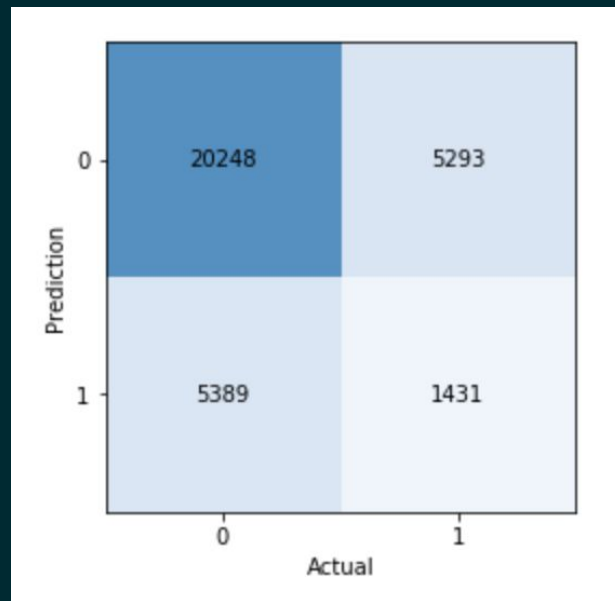
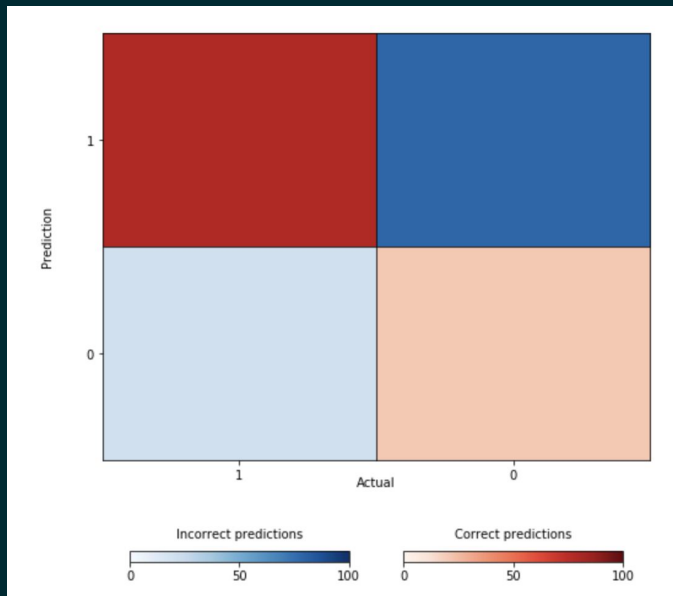
Project: If user will not go to the next page?

Results evaluation

	In reality: didn't go	In reality: did go
Prediction: will not go	Main focus area 16%, up from 9%	Big error, so wrong 19%, down from 21%
Prediction: will go to N+1	Lost opportunity 9.5%, down from 21%	Whatever, carry on, user

Project: If user will not go to the next page?

Plot-like confusion matrix

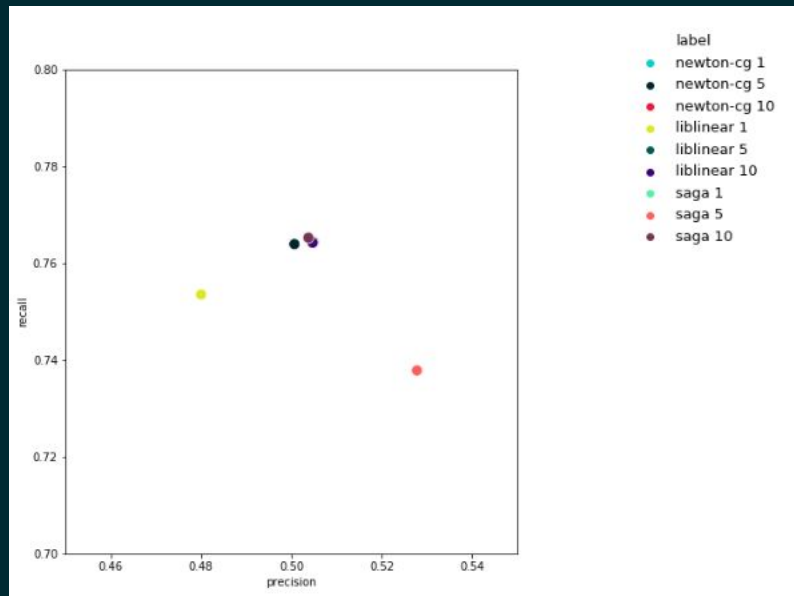
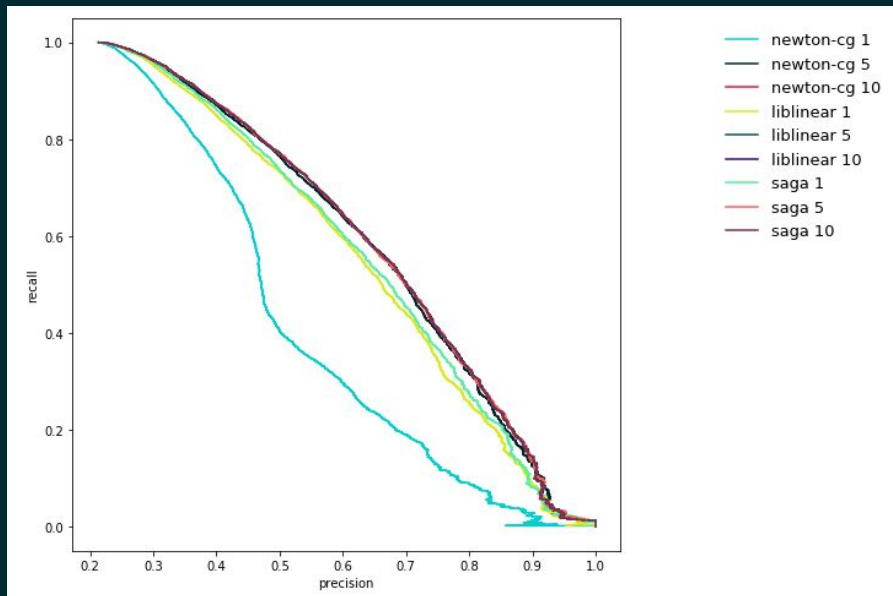


Notebook...

Episode 2 - many models!

- How to compare them?
- How to choose the best one?
- Confusion matrix for 100 models?

Plotting multiple models



Notebook...

Episode N

And... return to the beginning :)

More features!

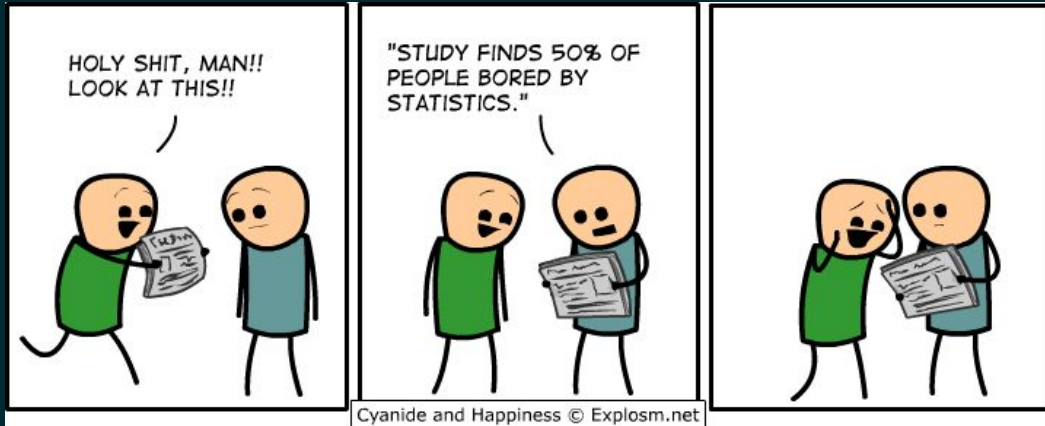


Main learning points

- **Talk :)**
- Make **simple** plots and **simple** tables with **precise** description where you can
- Confront your model evaluation methods with business needs
 - Set your benchmark **together**
 - Set your key metric(s) **together**

Questions? Comments?

murbanek@fandom.com



Appendix: Resources

- Jupyter notebook with analysis (github)
 -
- Presentation (github)
 -