

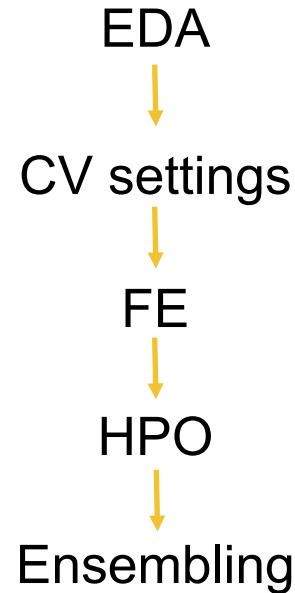
kaggle tips & tricks

kaggleDays meetup
11th June, Los Angeles



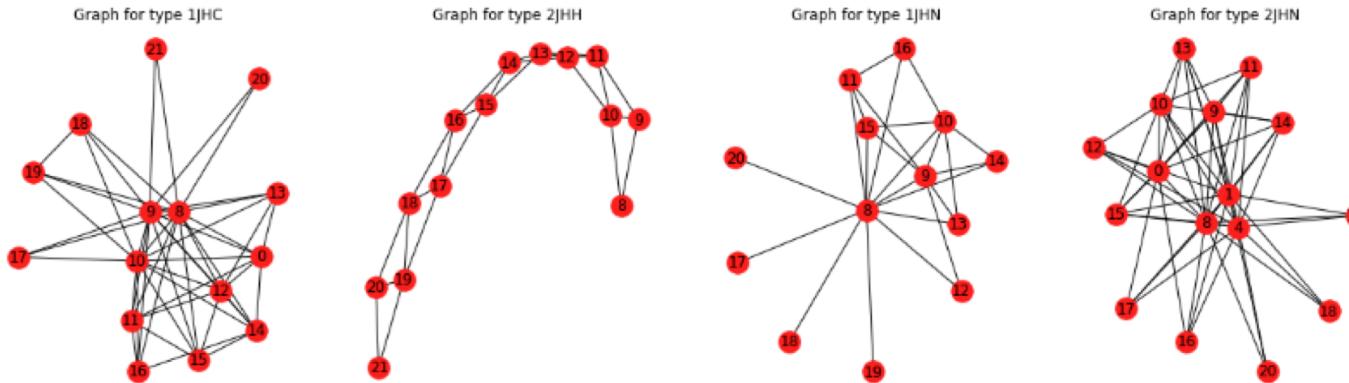
Michał Bugaj, meykds

How to start a competition?



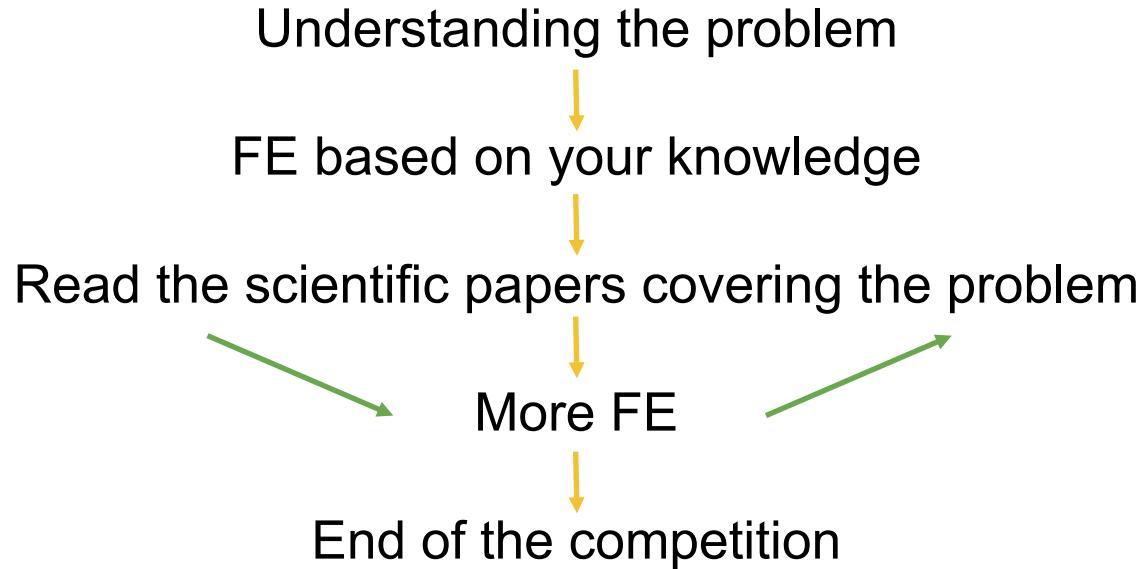
Exploratory Data Analysis

Well prepared EDA is crucial during the competition



EDA - Molecular Properties - made by Andrew Lukyanenko

How to do proper FE:



Unique ideas:

- what if I were ...
- understanding the business problem

1	Kraków, Lublin i Zhabinka		0.81724	329	9mo
2	ikiri_DS		0.81241	477	9mo
3	circlecircle		0.81124	231	9mo

Output (is default?)

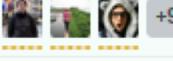
Bidirectional LSTM

1D Convolution layer (96 x 2,
128 filters)

Input „user“ image
(before normalization
- for interpretability)

Data Category	Data Field	Months before loan application								
		-96	-95	...	-5	-4	-3	-2	-1	
Internal	Account_inflows	0	0	...	4000	4000	4000	5000	4000	
Internal	Account_outflows	0	0	...	3000	3000	3000	3000	3000	
Internal	Number of installments due	0	0	...	1	1	1	1	1	
Internal	Number of installments 90DPD	0	0	...	0	0	1	1	1	
...	
External	Applied_for_loan_in_other_banks[flag]	0	0	...	0	0	1	0	0	
External	Got_rejected_by_other_banks[flag]	0	0	...	0	0	1	0	0	
...	

Why interpretability of the models matter?

1	▲ 10	Home Aloan		+3	0.80570	499	9mo
2	—	ikiri_DS		+9	0.80561	477	9mo
3	▲ 1	alijs & Evgeny			0.80511	143	9mo
4	▲ 6	Quad Machine			0.80474	178	9mo
5	▼ 4	Kraków, Lublin i Zhabinka			0.80449	329	9mo

Details ...



Toxic Comment Classification Challenge

Identify and classify toxic online comments

Featured · a year ago · 🗃 arguments, text data



88/4551
Top 2%

What have we done but we haven't used at the end of the competition

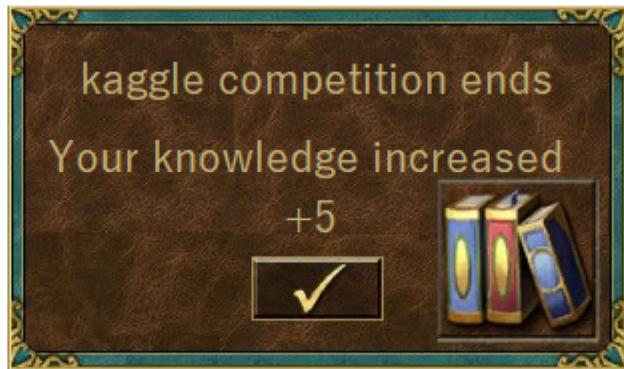
CV: train vs validation sets

minimize the gap between train score and validation score



Private LB vs Public LB difference will be smaller

kaggle competitions Pros & Pros



Analytics Prediction Competition

Data Science for Good: City of Los Angeles

Help the City of Los Angeles to structure and analyze its job descriptions

City of Los Angeles · 11 days to go

Overview Data Kernels Discussion Rules

Implementation

Knowledge