

Team GARCH Approach

Track N°3 by Revenue Grid



Andrii Yerko

Nikita Fordui

Oleksii Galganov

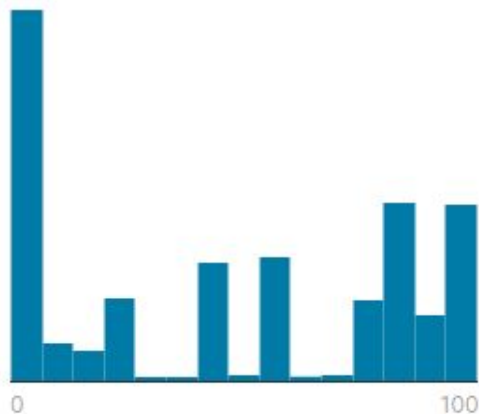
Sofia Shaposhnikova



Task: forecast the probability of successful deal closing

Id	Opportunit...	CreatedById	CreatedDa...	CreatedDa...
00813000020pd6v AAB	0061300001AC2HQ AA1	0051300000BAVA6 AAP	9/23/2014 17:45	2014-09-23 17:45:00
00813000020qCDY AAB	0063000000jwAau AAE	00530000004M1Pm AAK	2/17/2011 4:04	2014-09-26 17:12:00
00813000020qCDh AAN	0063000000myLfb AAE	00530000004M1Pm AAK	8/9/2011 15:48	2014-09-26 17:12:00
00813000020qCE1 AAN	0063000000pxbZx AAI	00530000004M1Pm AAK	1/18/2012 22:44	2014-09-26 17:12:00

Probability



Valid

Mismatched

Missing

Mean

Std. Deviation

Quantiles

34.2k

100%

0

0%

0

0%

43.9

38.3

0

Min

1

25%

30

50%

80

75%

100

Max

			12 18:49	2014-09-26 17:12:00
			012 17:39	2014-09-26 17:12:00
olid	34.2k	100%	12 22:25	2014-09-26 17:12:00
ismatched	0	0%		
issing	0	0%	12 17:36	2014-09-26 17:12:00
ean	43.9			
d. Deviation	38.3		2012	2014-09-26 17:13:00
uantiles	0	Min	2012	2014-09-26 17:13:00
	1	25%		
	30	50%	2012	2014-09-26 17:13:00
	80	75%		
	100	Max	013 15:43	2014-09-26 17:13:00
00813000020qCFp AAN	0063000000EKcmE AAT	00530000000wm2V AAQ	10/18/2007 19:47	2014-09-26 17:13:00



LightGBM

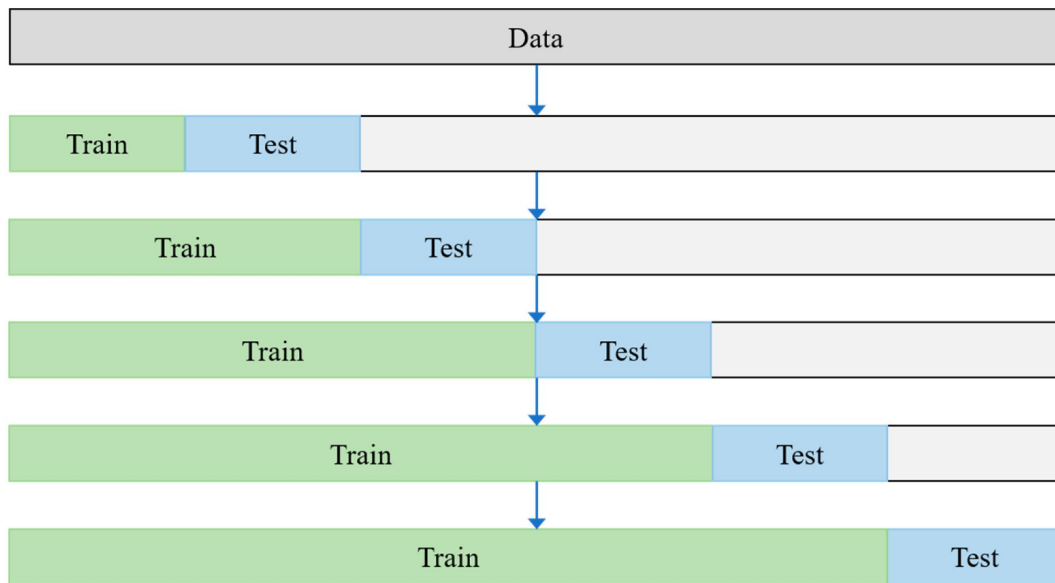


CatBoost



Data distributed over time

Prevent leaks with the correct time-aware prediction scheme



Features

Dates

'CreatedDate',
'CreatedDateForInsert',
'ValidThroughDate',
differences, quarters, years,
sin-cos encoding

Lags

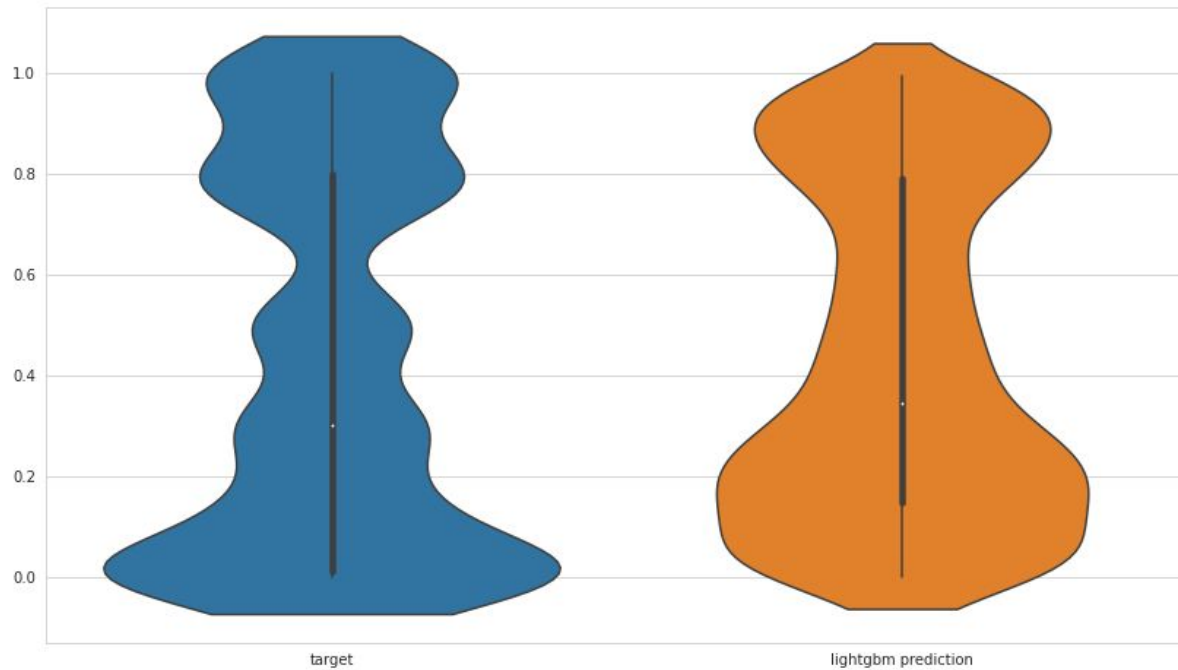
Stats of previous
probabilities grouped by
Opportunity, CreatedBy and
time periods

Categories

'CreatedById', 'AccountId',
'RecordTypeId', 'Type',
'LeadSource', 'CampaignId'
etc.

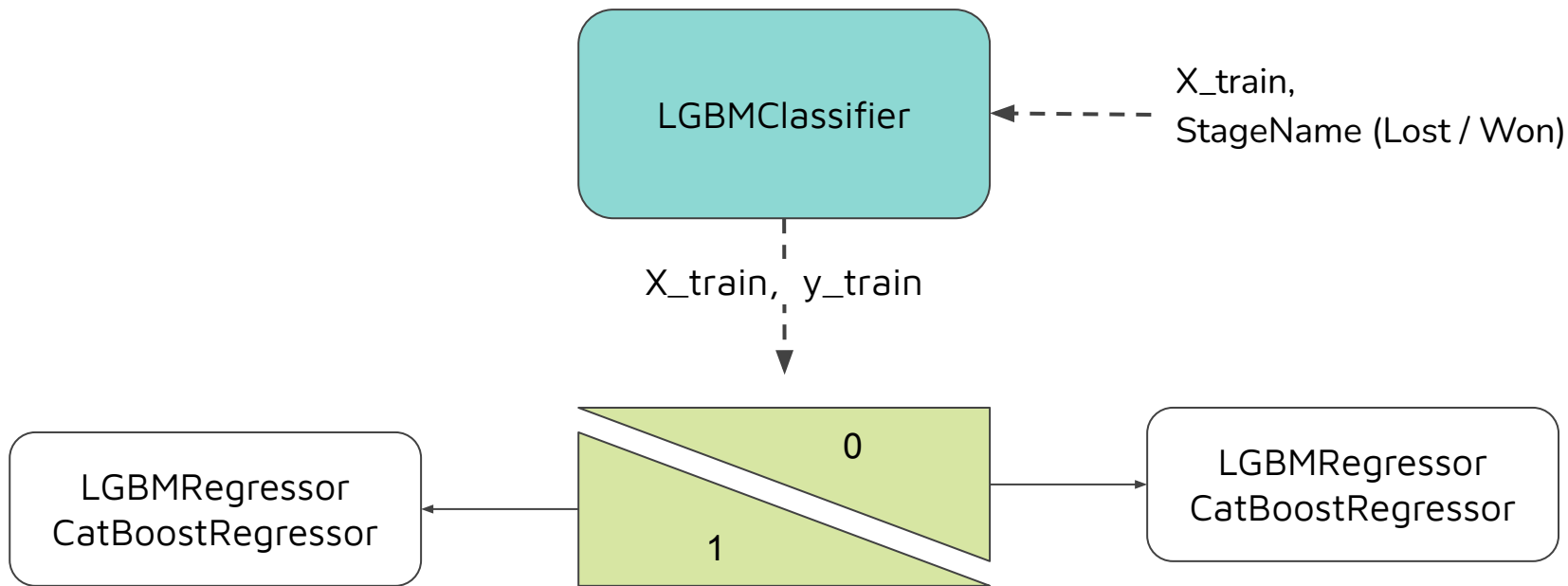


Logloss instead of RMSE



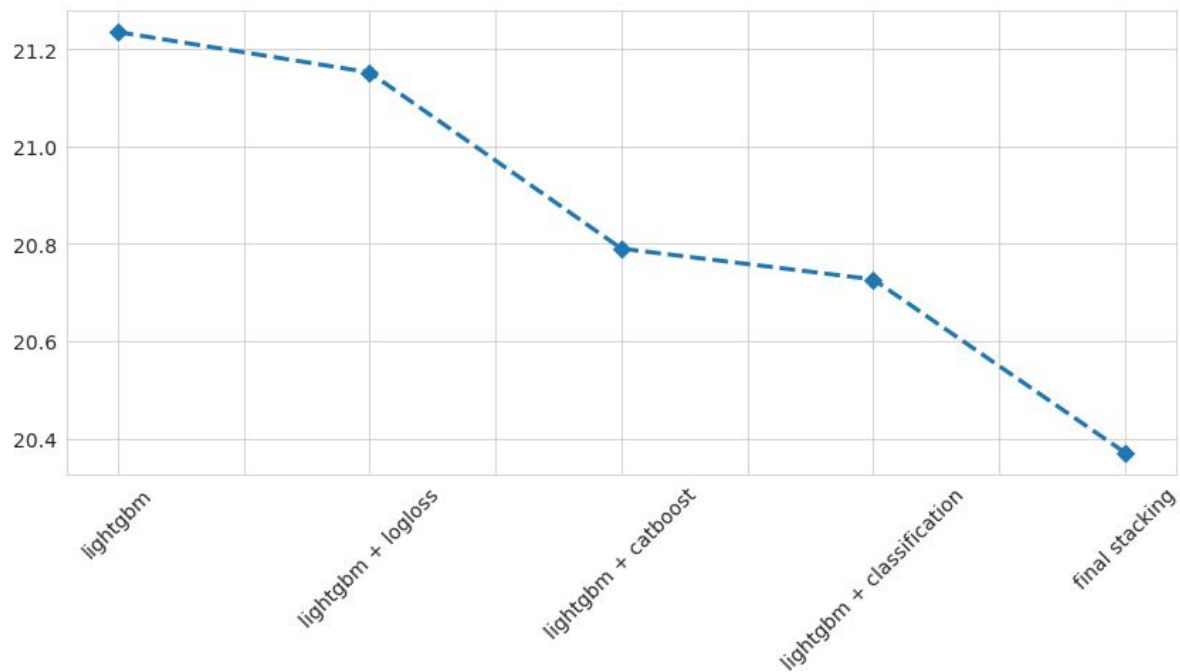


Classifier





Final score: 20.37 RMSE



**Thank you for
attention!**

Any questions?

