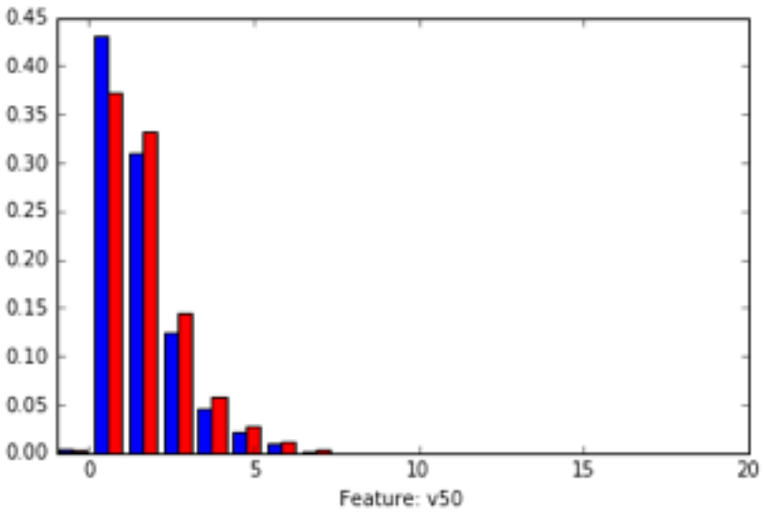
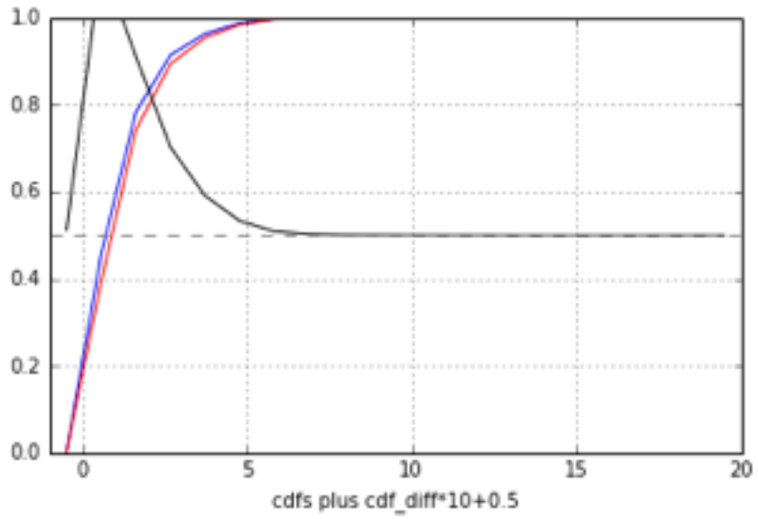
BNP Cardif claims management

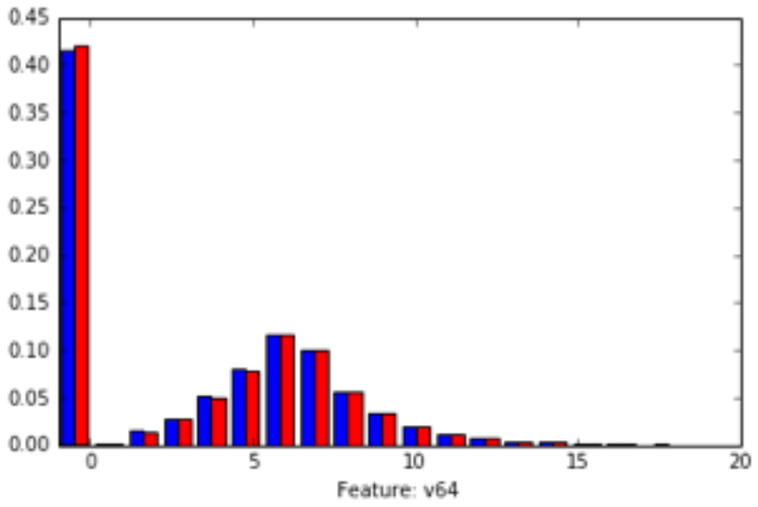
1 Fast. Claims for which approval could be accelerated leading to faster payments
2 Audit. Claims for which additional information is required before approval

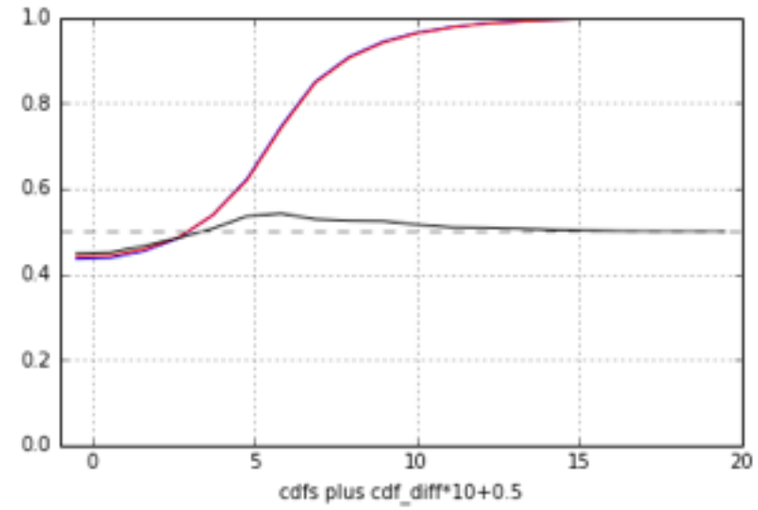
2 class classification

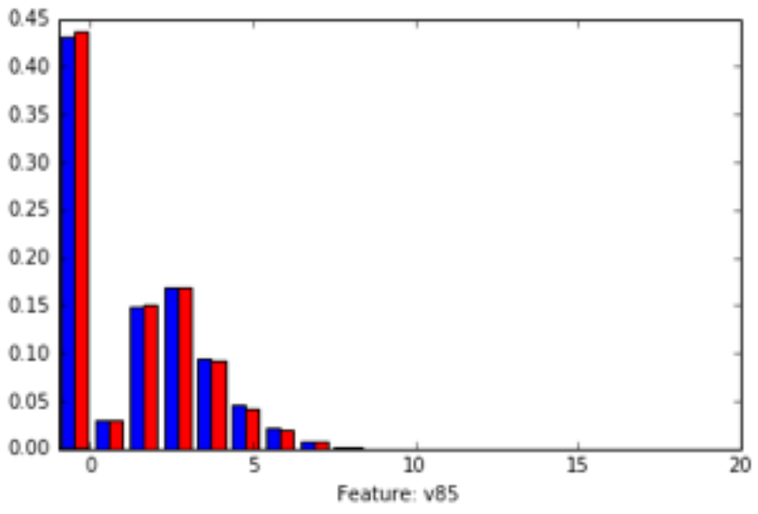
$$logloss = -\frac{1}{N}\sum_{i=1}^{N}\left(y_i \log(p_i) + (1 - y_i)\log(1 - p_i)\right)$$

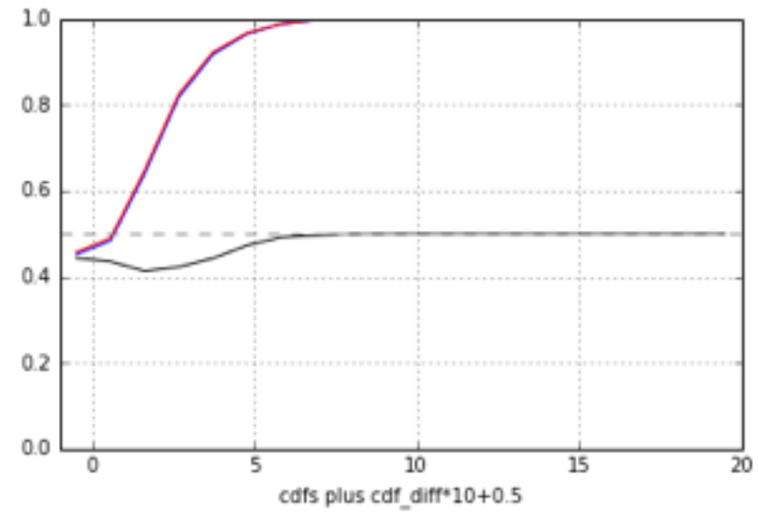

0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0
5
10
15
20
Feature: v50

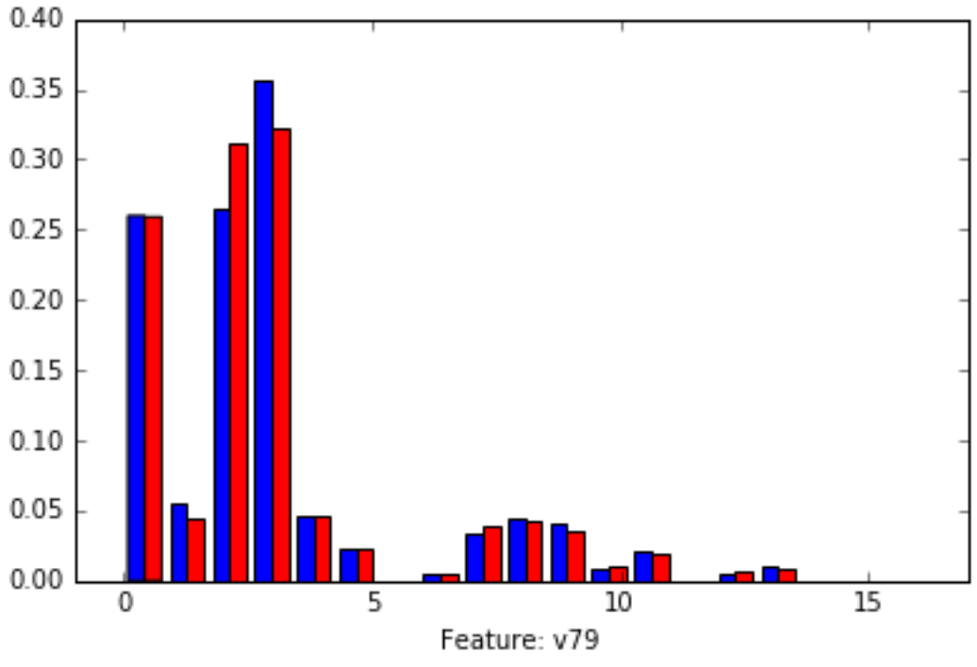

1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15
20
cdfs plus cdf_diff*10+0.5

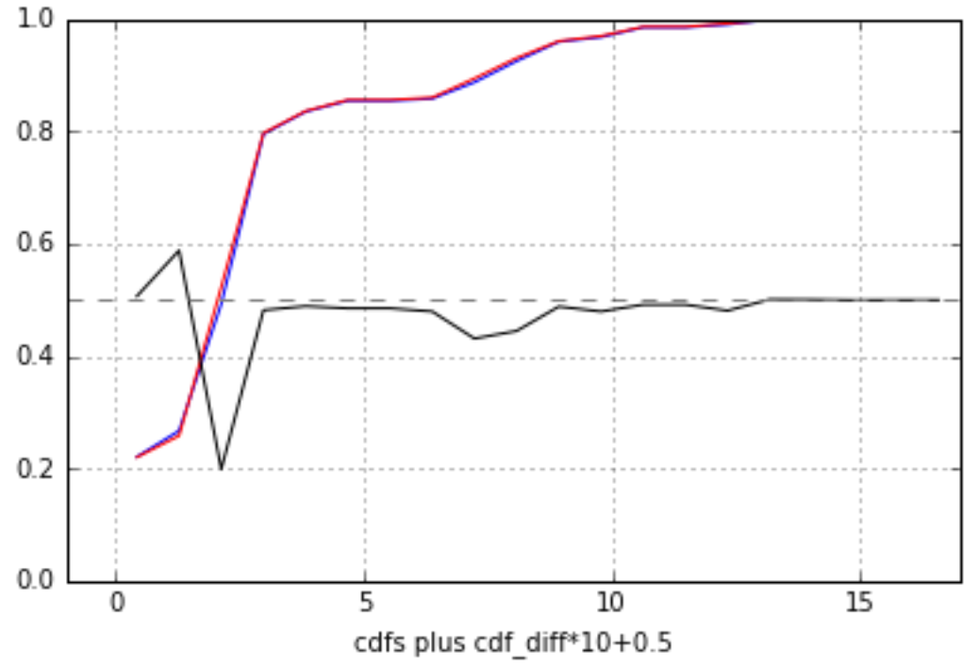

0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0
5
10
15
20
Feature: v64

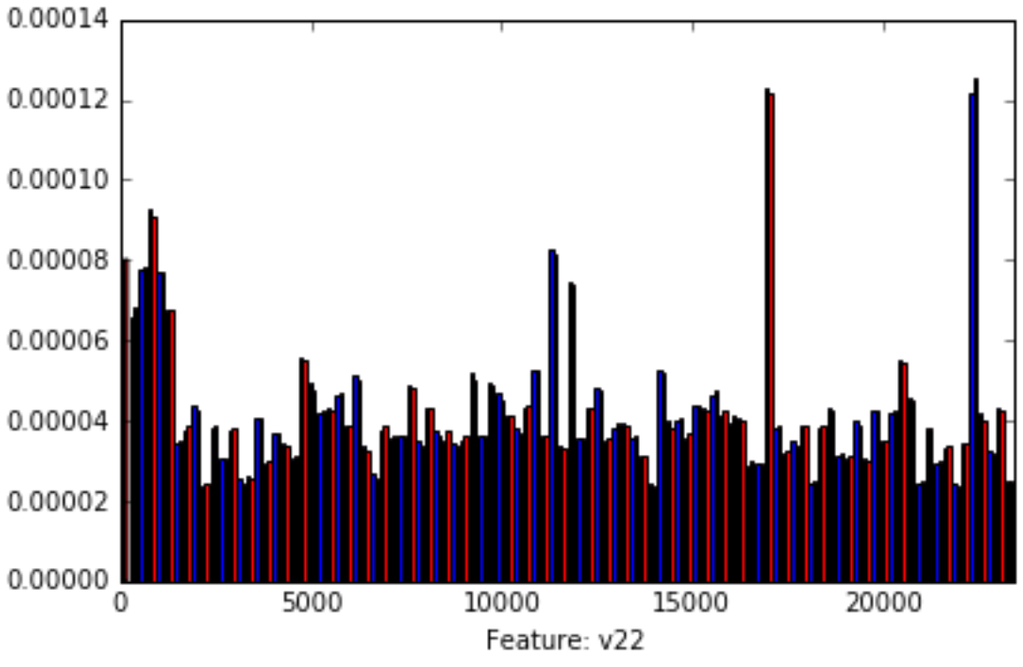

1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15
20
cdfs plus cdf_diff*10+0.5

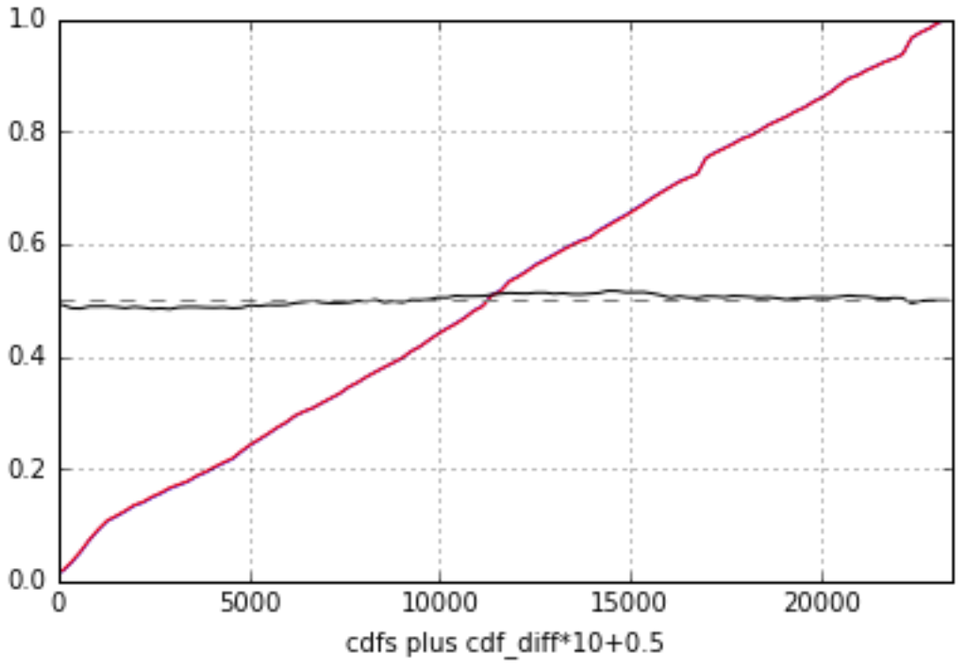

0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0
5
10
15
20
Feature: v85

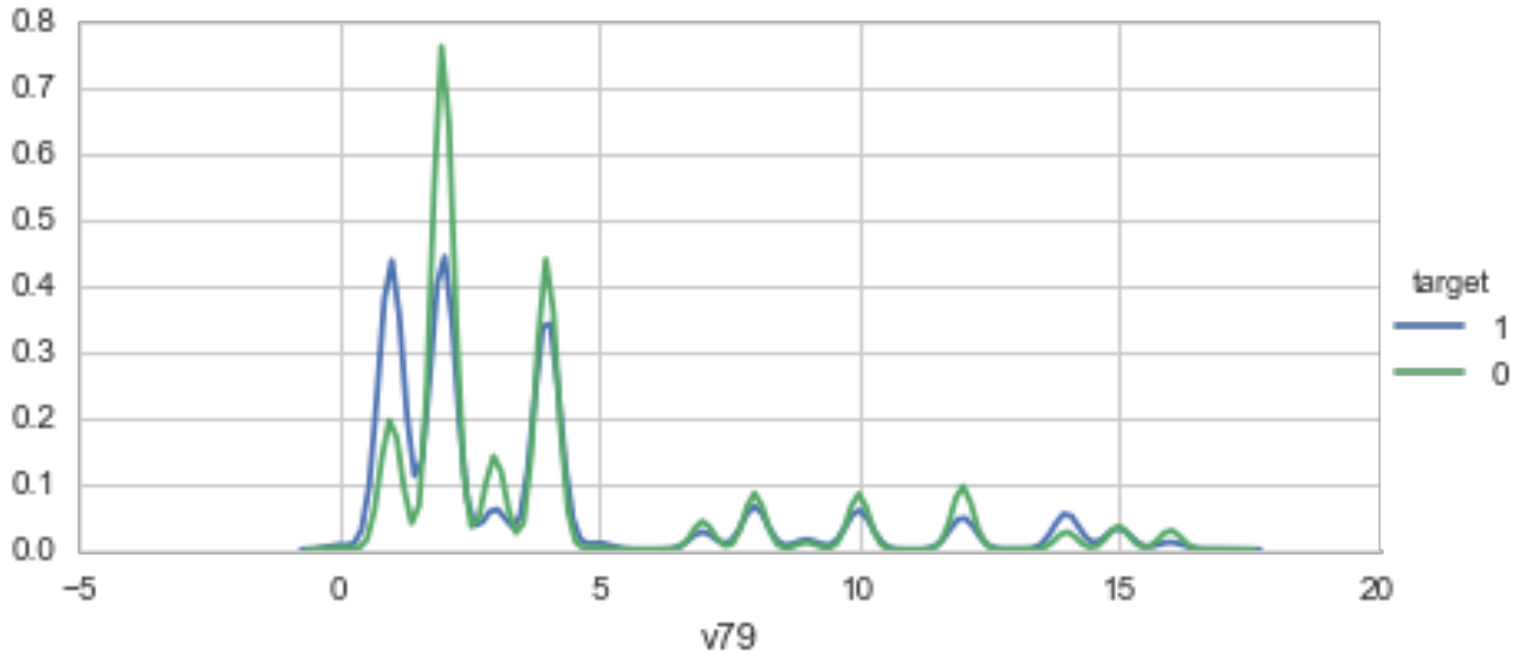

1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15
20
cdfs plus cdf_diff*10+0.5


0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0
5
10
15
Feature: v79


1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15
cdfs plus cdf_diff*10+0.5


0.00014
0.00012
0.00010
0.00008
0.00006
0.00004
0.00002
0.00000
0
5000
10000
15000
20000
Feature: v22


1.0
0.8
0.6
0.4
0.2
0.0
0
5000
10000
15000
20000
cdfs plus cdf_diff*10+0.5


0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
−5
0
5
10
15
20
v79
target
1
0

Plan

- Feature engineering
  %percentage
  interactions
  genetic programming

- LogReg + Ridge/Lasso/Elastic

- Text feature interpretation

- LibFM
  http://libfm.org

- RGF
  http://stat.rutgers.edu/home/tzhang/software/rgf/

- NeuraI Network
  Keras / Lasagne / Caffe / Tensorflow

- SVM

- ET / RF / XGBoost hyperopt

- Stacking / BIending / Voting

Links

http://opendatascience.ru
http://facebook.com/playittodeath