



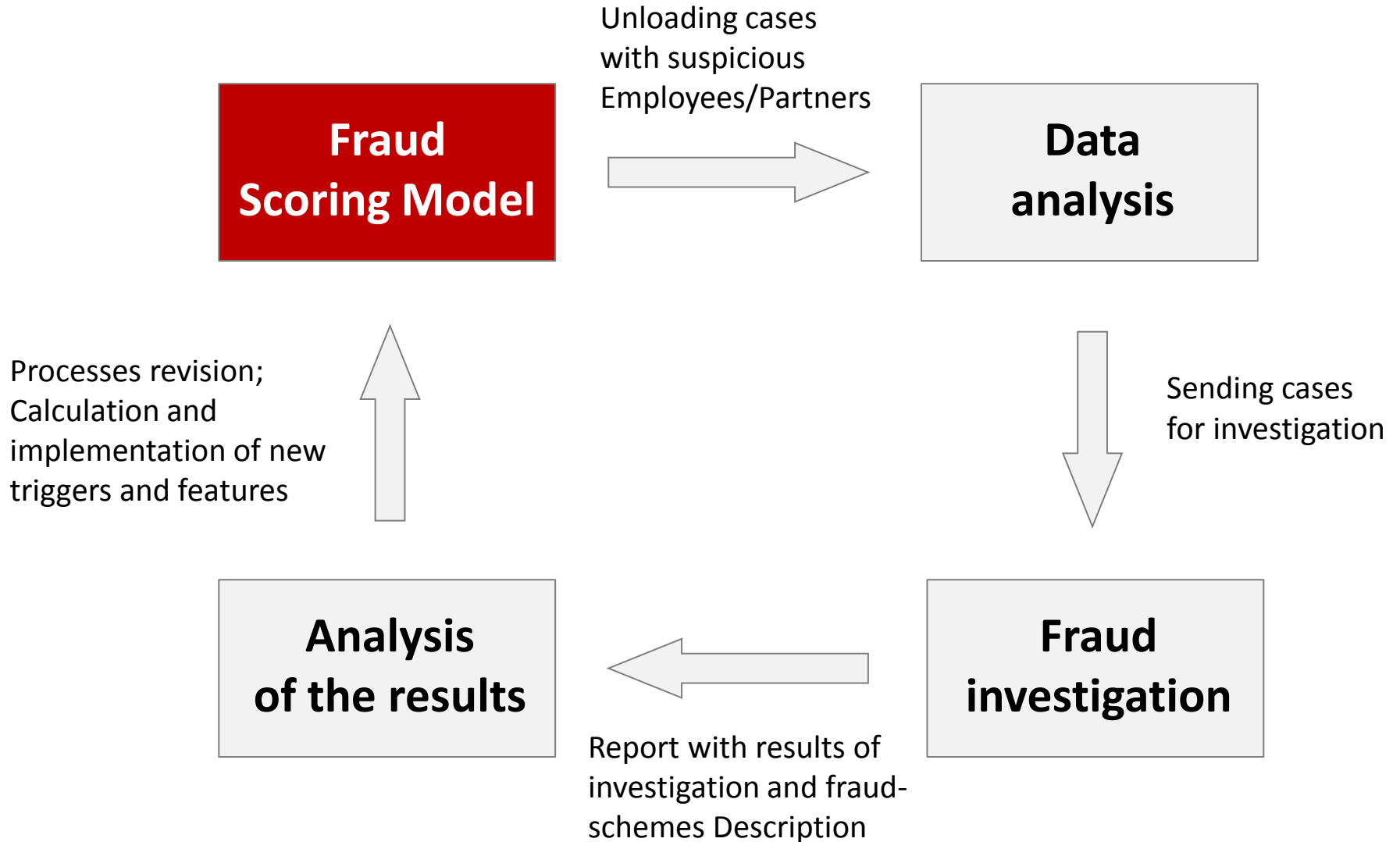
Predictive Anti-Fraud Analytics

Detection of Internal and External Fraud in Retail Banking

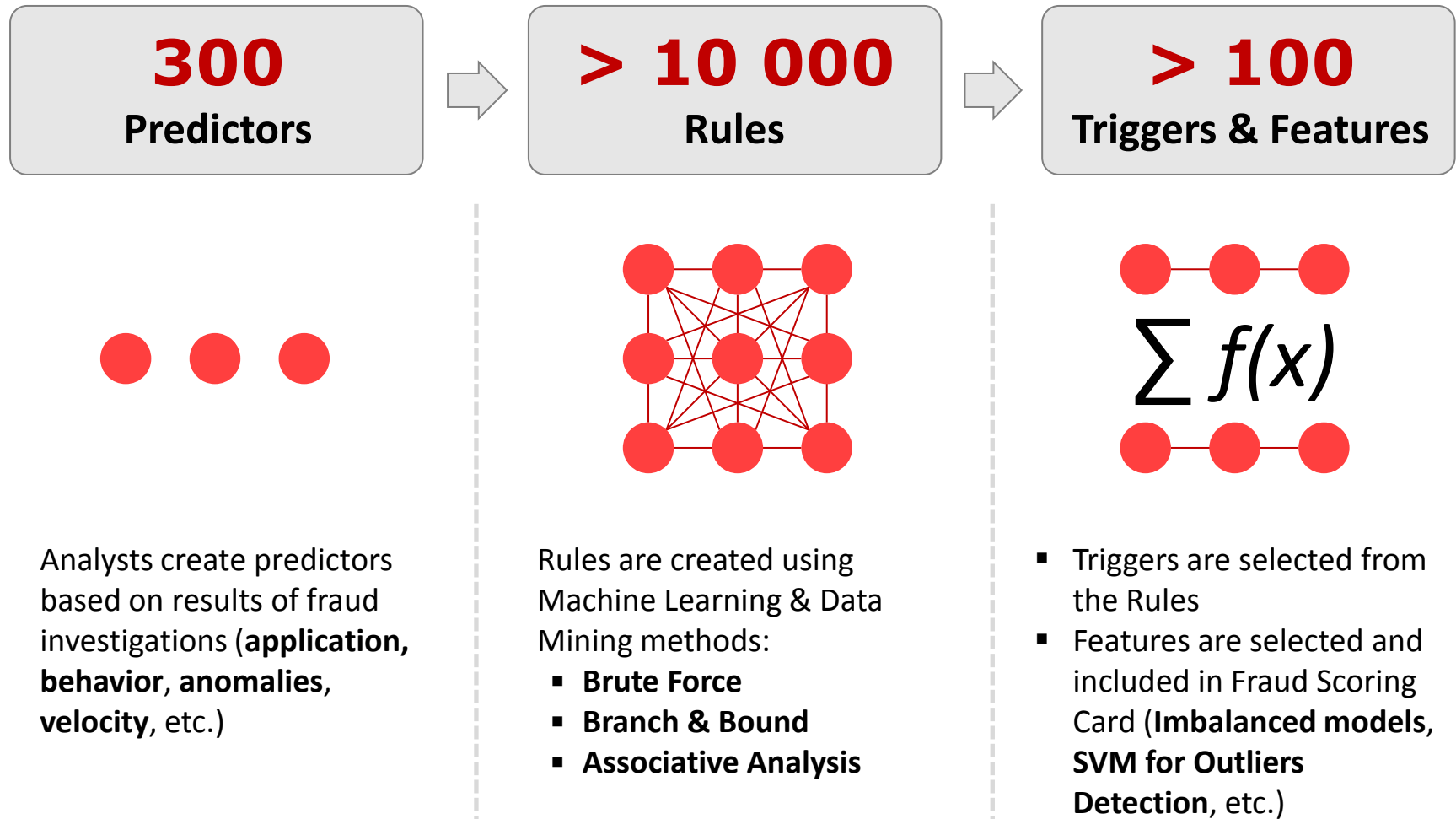
Sergey Afanasiev

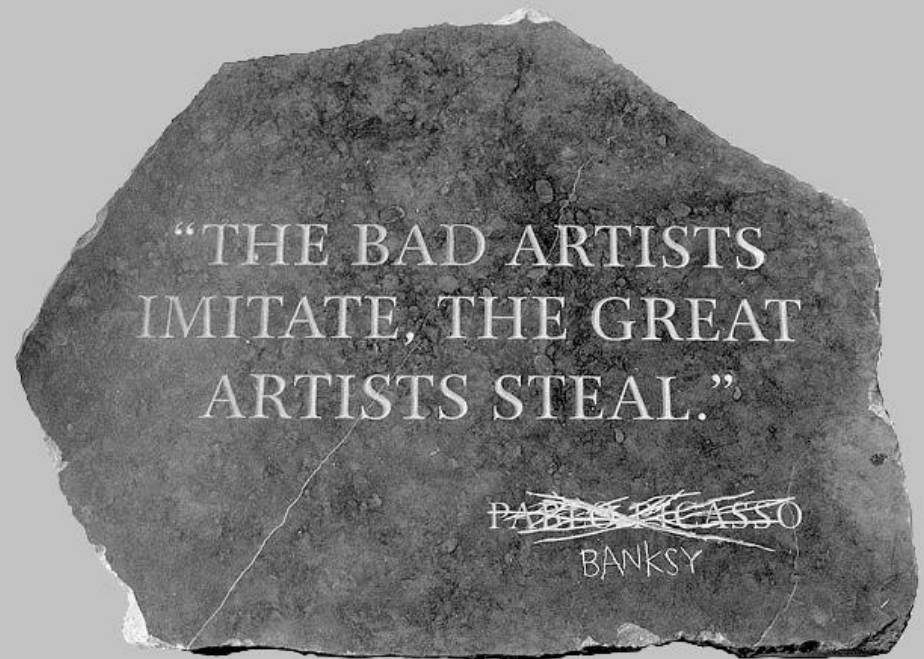
Head of Anti-Fraud, Bank Renaissance Credit
Russia, Moscow, 2017

Warning System



Fraud Scoring Model





1. Predictors based on Rules from External Solutions

Anti-fraud systems: FPS, Hunter etc.

FPS-Rules

1st set

Different applicants
with the same phone
number (not relatives)

2nd set

The same applicant has
different personal data
(address, phone, etc.)

3rd set

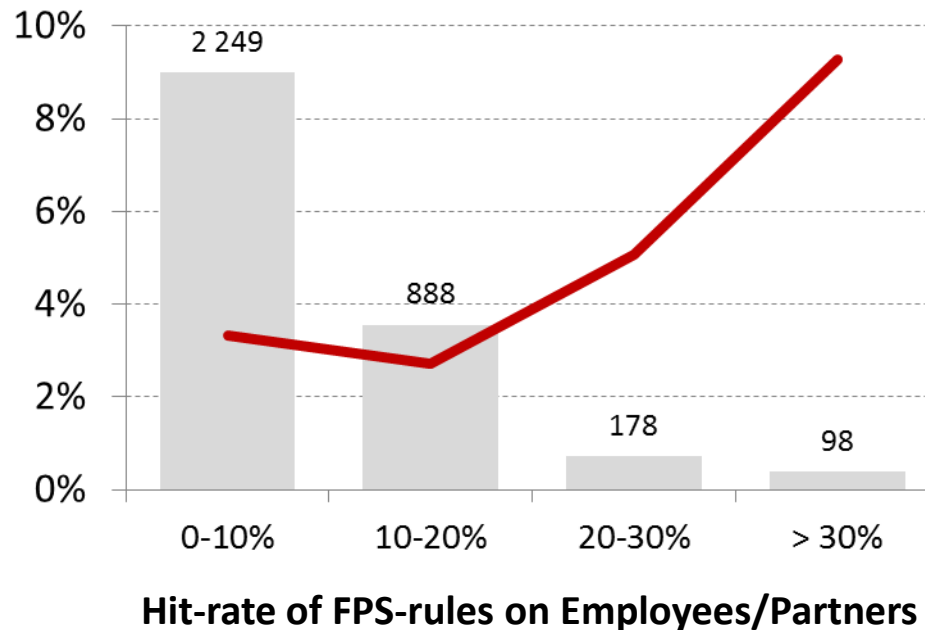
Special rules
(suspicious activity,
behavior, etc.)

FPS_1.11	One phone, different applicants + special criteria on name, address, etc.
FPS_1.12	One phone, different applicants + special criteria on name, address, etc.
FPS_1.13	Usage of fake phones
FPS_1.14	One phone, different applicants + special criteria on name, address, etc.
FPS_1.15	Different applicants used one phone in several banks
FPS_1.16	One phone is found in different applicants with different addresses + special criteria
FPS_1.17	One phone is found in different applicants with different addresses + special criteria
FPS_1.18	Different applicants used one phone in several banks
FPS_2.11	Frequent change of mobile phone
FPS_2.12	Suspicious behavior
FPS_2.13	Possible fake documents
FPS_2.14	A lot of changes in data
FPS_2.15	A lot of changes in contact data
FPS_2.16	Modify contact data of applicant for a short period
FPS_2.17	A lot of changes in contact data
FPS_2.18	Modify contact data of applicant for a short period
FPS_2.19	After Cash-application client makes a request for POS-credit
FPS_2.20	After rejection Cash-application client makes a request for POS-credit
FPS_2.21	High credit activity of applicant
FPS_2.22	Different data of employer in different banks
FPS_2.23	High-risk segment
FPS_2.24	High-risk segment
FPS_2.25	High credit activity of applicant
FPS_2.26	High-risk segment
FPS_2.27	High credit activity of applicant

FPS-Predictor for Internal Fraud

The high rate of FPS-hits on Employees/Partners increases the probability of internal fraud by these Employees/Partners

The percentage of fraud detected



Parameters

Min number of "FPS-hits" for one Employee	Period	Concentration of "FPS-hits" on Employee
≥ 3	7 days	$\geq 5\%$
≥ 5	14 days	$\geq 10\%$
≥ 10	30 days	$\geq 15\%$
		$\geq 20\%$
		$\geq 25\%$
		$\geq 30\%$
		$\geq 35\%$
		$\geq 40\%$
		$\geq 45\%$
		$\geq 50\%$
		$\geq 55\%$
		$\geq 60\%$
		$\geq 65\%$

$3 \times 3 \times 13 = 117$ Rules

Triggers

Brute Force method:

Triggers are selected in several iterations until target variable becomes lower than a target value

Target variable: **30+mob3** \geq **15%**

Min hit: ≥ 3			
Hit-rate	Period (days)		
	7	14	30
$\geq 30\%$			
$\geq 35\%$			
$\geq 40\%$			
$\geq 45\%$			
$\geq 50\%$			
$\geq 55\%$			

Min hit: ≥ 5			
Hit-rate	Period (days)		
	7	14	30
$\geq 30\%$			
$\geq 35\%$			
$\geq 40\%$			
$\geq 45\%$			
$\geq 50\%$			
$\geq 55\%$			

Min hit: ≥ 10			
Hit-rate	Period (days)		
	7	14	30
$\geq 30\%$	15%		
$\geq 35\%$		24%	
$\geq 40\%$			
$\geq 45\%$			
$\geq 50\%$			
$\geq 55\%$			

4th iteration

5th iteration

2nd iteration

3rd iteration

1st iteration

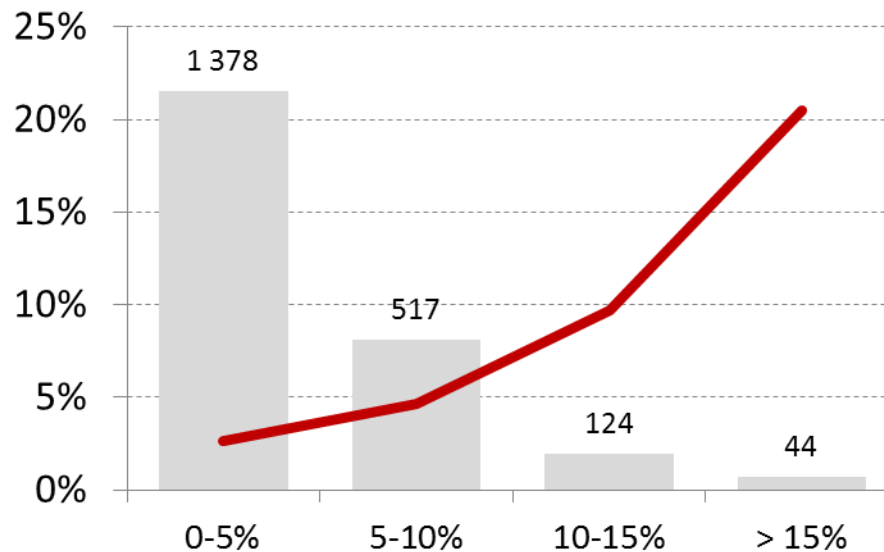


2. Predictors based on Data from Mobile Operators

Young SIM

The high rate of “Young SIM” applications of Employees/Partners increases the probability of internal fraud

The percentage of fraud detection



Hit-rate of “Young SIM” of Employees/Partners

Parameters

Min number of "Young SIM" for one Employee	Age of SIM	Period	Concentration of "Young SIM" on Employee
>= 4	0-30 days	7 days	>=55%
>= 5	0-90 days	14 days	>=60%
>= 6		30 days	>=65%
>= 7			>=70%
>= 8			>=75%
>= 9			=100%
>= 10			

7 x 2 x 3 x 6 = 252 Rules

Triggers

Target variable: **30+mob3** \geq 40%

Period (days) 7							
Age of SIM (days) 0-30							
Hit-rate	Min hit						
	4	5	6	7	8	9	10
$\geq 55\%$							51%
$\geq 60\%$							
$\geq 65\%$							
$\geq 70\%$							
$\geq 75\%$							
$\geq 100\%$							

Period (days) 14							
Age of SIM (days) 0-30							
Hit-rate	Min hit						
	4	5	6	7	8	9	10
$\geq 55\%$							
$\geq 60\%$							
$\geq 65\%$							
$\geq 70\%$							
$\geq 75\%$							
$\geq 100\%$						46%	

Period (days) 30							
Age of SIM (days) 0-30							
Hit-rate	Min hit						
	4	5	6	7	8	9	10
$\geq 55\%$							44%
$\geq 60\%$							
$\geq 65\%$							
$\geq 70\%$							
$\geq 75\%$							
$\geq 100\%$							

Period (days) 7							
Age of SIM (days) 0-90							
Hit-rate	Min hit (кол-во)						
	4	5	6	7	8	9	10
$\geq 55\%$							
$\geq 60\%$							
$\geq 65\%$						47%	
$\geq 70\%$							
$\geq 75\%$							
$\geq 100\%$							

Period (days) 14							
Age of SIM (days) 0-90							
Hit-rate	Min hit						
	4	5	6	7	8	9	10
$\geq 55\%$							
$\geq 60\%$							
$\geq 65\%$							
$\geq 70\%$							
$\geq 75\%$						66%	
$\geq 100\%$							

Period (days) 30							
Age of SIM (days) 0-90							
Hit-rate	Min hit						
	4	5	6	7	8	9	10
$\geq 55\%$							
$\geq 60\%$							
$\geq 65\%$							
$\geq 70\%$							
$\geq 75\%$							59%
$\geq 100\%$						78%	

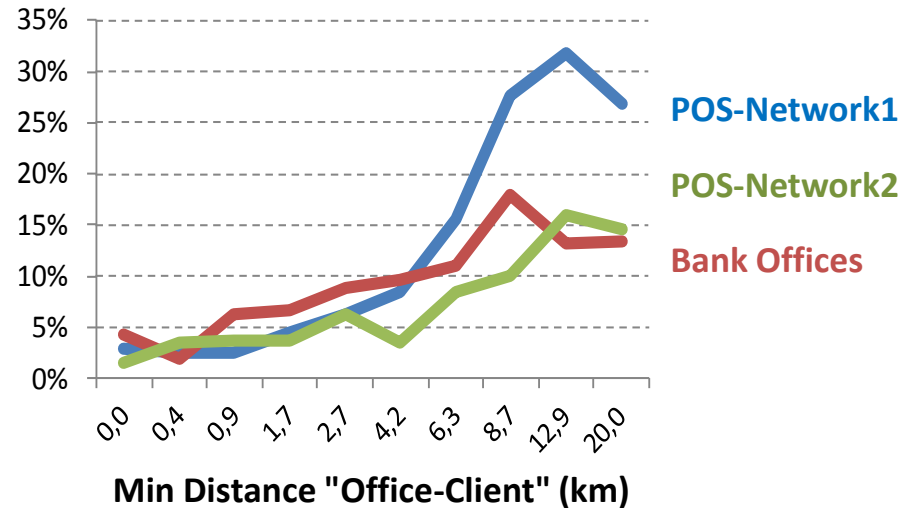
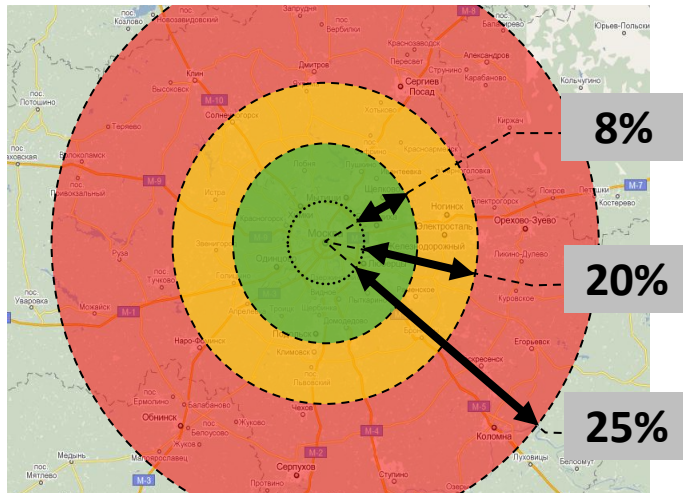


3. Predictors based on Geolocation Data

GPS-Predictors

If the applicant lives or works far from Bank office, there is a high probability that he is a fraudster

FPD-30 vs. Distance "Office-Client"



Parameters

Distance "Office <-> Client"	Min number of "Distant clients" for one Employee	Min distance "Office-Client"	Analysis period	Concentration of "Distant clients" on Employee
Office <-> Reg. address	>= 3	0-1 km	7 days	>=5%
Office <-> Actual address	>= 5	1-2 km	14 days	>=10%
Office <-> Work address	>= 10	2-3 km	30 days	>=15%
		3-5 km	60 days	>=20%
		5-7 km		>=30%
		7-10 km		>=50%
		10-20 km		>=70%
		>= 20 km		=100%

3 x 8 x 4 x 8 = 768 Rules

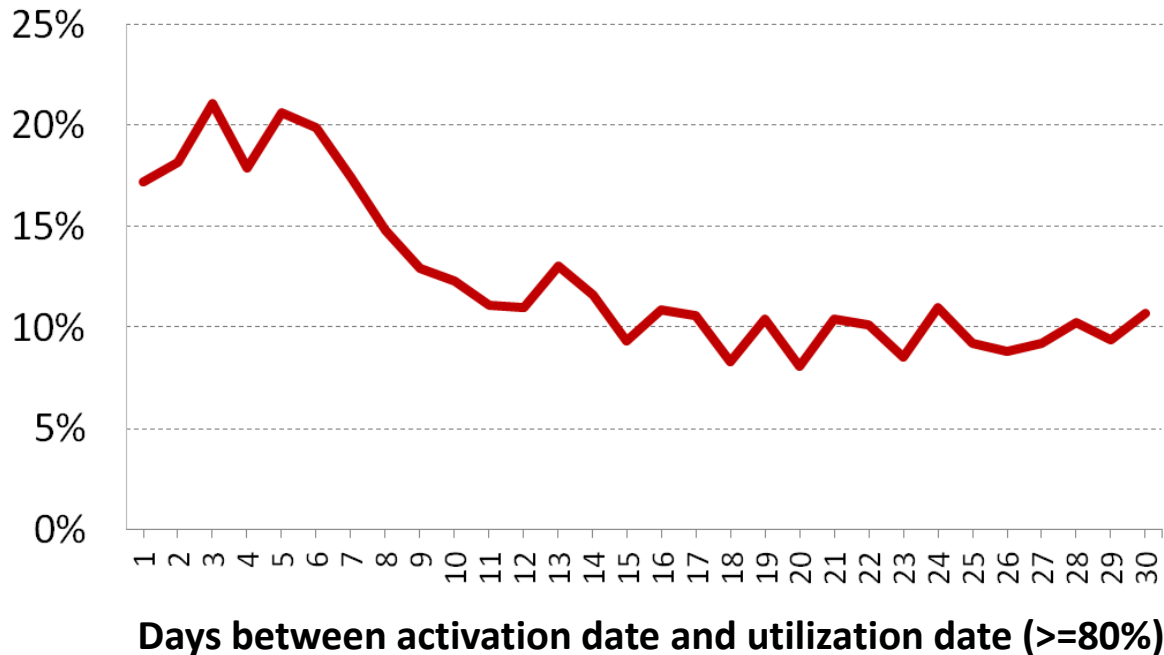


4. Special Predictors for Credit Cards

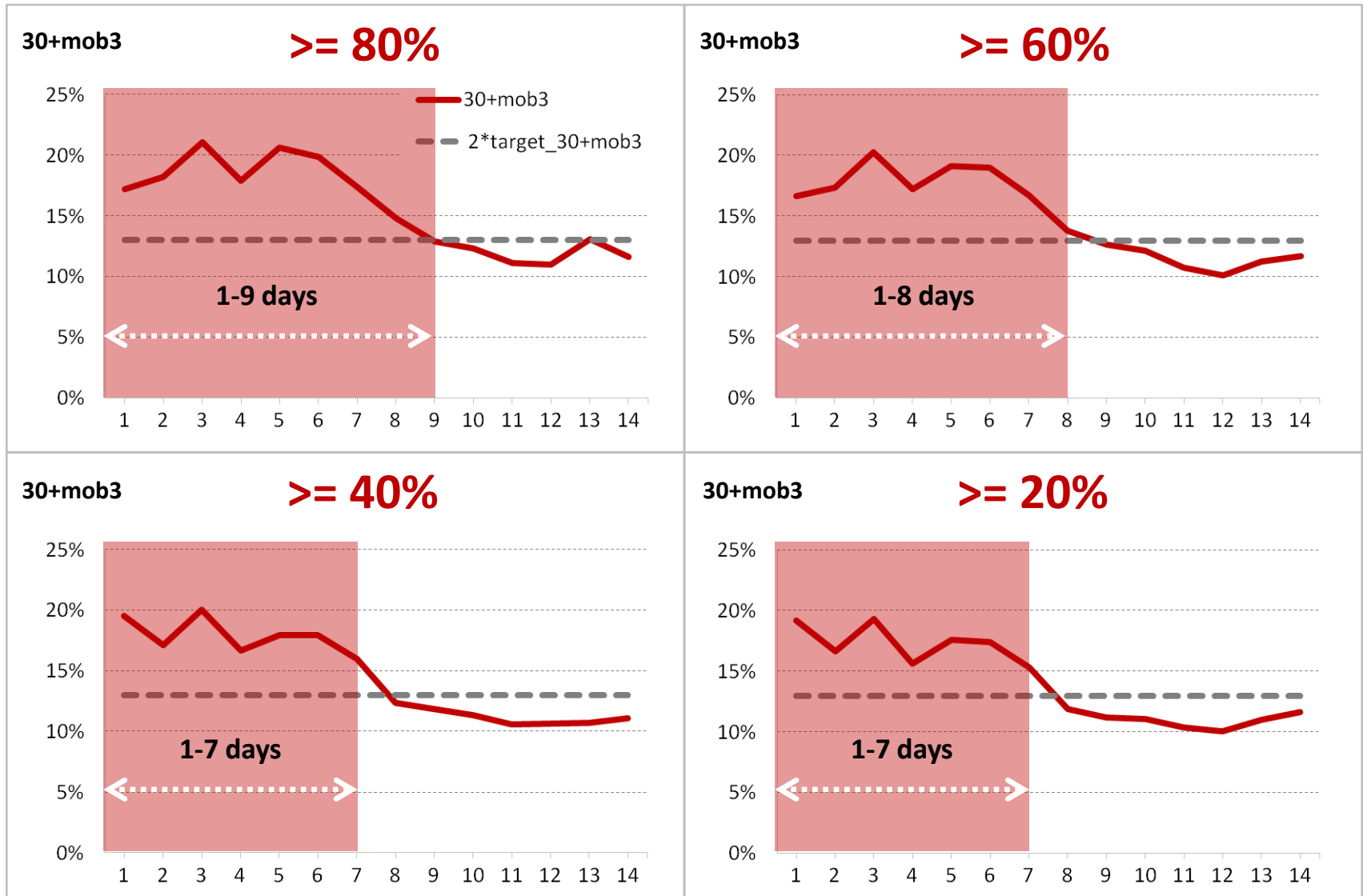
Utilization of Credit Card Limit

Fraudsters spend 80-100% of card limit within few days after activating the credit card

30+mob3 vs. Period of limit utilization



Parameters Selection



Parameters

Min number of "Quick Utilization" for one Employee	% of limit utilized	Period of utilization	Period	Concentration of "Quick Utilization" on Employee
>= 3	>= 20%	1 day	7 days	>=5%
>= 5	>= 40%	2 days	14 days	>=10%
>= 10	>= 60%	3 days	30 days	>=15%
	>= 80%	4 days	60 days	>=20%
		5 days		>=30%
		6 days		>=50%
		7 days		>=70%
		8 days		=100%
		9 days		
		10 days		

3 x 4 x 10 x 4 x 8 = 3 840 Rules

Triggers

Target variable: **30+mob3** $\geq 20\%$

Min number of "Quick Utilization" >=3									>=5									>=10									>=3									>=5									>=10								
Hit-rate >= 70% % utilized limit >= 20%									>= 70% % utilized limit >= 20%									>= 70% % utilized limit >= 20%									>= 70% % utilized limit >= 40%									>= 70% % utilized limit >= 40%									>= 70% % utilized limit >= 40%								
Period	Period of utilization								Period of utilization	Period of utilization								Period of utilization	Period of utilization								Period of utilization	Period of utilization								Period of utilization	Period of utilization																
	1	2	3	4	5	6	7	8		1	2	3	4	5	6	7	8		1	2	3	4	5	6	7	8		1	2	3	4	5	6	7	8		1	2	3	4	5	6	7	8									
7																																																					
14																																																					
30																																																					
60																																																					
Min number of "Quick Utilization" >=3									>=5									>=10									>=3									>=5									>=10								
Hit-rate = 100% % utilized limit >= 20%									= 100% % utilized limit >= 20%									= 100% % utilized limit >= 20%									= 100% % utilized limit >= 40%									= 100% % utilized limit >= 40%									= 100% % utilized limit >= 40%								
Period	Period of utilization								Period of utilization	Period of utilization								Period of utilization	Period of utilization								Period of utilization	Period of utilization								Period of utilization	Period of utilization																
	1	2	3	4	5	6	7	8		1	2	3	4	5	6	7	8		1	2	3	4	5	6	7	8		1	2	3	4	5	6	7	8		1	2	3	4	5	6	7	8									
7																																																					
14																																																					
30																																																					
60																																																					
Min number of "Quick Utilization" >=3									>=5									>=10									>=3									>=5									>=10								
Hit-rate >= 50% % utilized limit >= 60%									>= 50% % utilized limit >= 60%									>= 50% % utilized limit >= 60%									>= 50% % utilized limit >= 80%									>= 50% % utilized limit >= 80%									>= 50% % utilized limit >= 80%								
Period	Period of utilization								Period of utilization	Period of utilization								Period of utilization	Period of utilization								Period of utilization	Period of utilization								Period of utilization	Period of utilization																
	1	2	3	4	5	6	7	8		1	2	3	4	5	6	7	8		1	2	3	4	5	6	7	8		1	2	3	4	5	6	7	8		1	2	3	4	5	6	7	8									
7																																																					
14																																																					
30																																																					
60																																																					
Min number of "Quick Utilization" >=3									>=5									>=10									>=3									>=5									>=10								
Hit-rate >= 70% % utilized limit >= 60%									>= 70% % utilized limit >= 60%									>= 70% % utilized limit >= 60%									>= 70% % utilized limit >= 80%									>= 70% % utilized limit >= 80%									>= 70% % utilized limit >= 80%								
Period	Period of utilization								Period of utilization	Period of utilization								Period of utilization	Period of utilization								Period of utilization	Period of utilization								Period of utilization	Period of utilization																
	1	2	3	4	5	6	7	8		1	2	3	4	5	6	7	8		1	2	3	4	5	6	7	8		1	2	3	4	5	6	7	8		1	2	3	4	5	6	7	8									
7																																																					
14																																																					
30																																																					
60																																																					
Min number of "Quick Utilization" >=3									>=5									>=10									>=3									>=5									>=10								
Hit-rate = 100% % utilized limit >= 60%									= 100% % utilized limit >= 60%									= 100% % utilized limit >= 60%									= 100% % utilized limit >= 80%									= 100% % utilized limit >= 80%									= 100% % utilized limit >= 80%								
Period	Period of utilization								Period of utilization	Period of utilization								Period of utilization	Period of utilization								Period of utilization	Period of utilization								Period of utilization	Period of utilization																
	1	2	3	4	5	6	7	8		1	2	3	4	5	6	7	8		1	2	3	4	5	6	7	8		1	2	3	4	5	6	7	8		1	2	3	4	5	6	7	8									
7																																																					
14																																																					
30																																																					
60																																																					
Min number of "Quick Utilization" >=3									>=5									>=10									>=3									>=5									>=10								
Hit-rate = 100% % utilized limit >= 60%									= 100% % utilized limit >= 60%									= 100% % utilized limit >= 60%									= 100% % utilized limit >= 80%									= 100% % utilized limit >= 80%									= 100% % utilized limit >= 80%								
Period	Period of utilization								Period of utilization	Period of utilization								Period of utilization	Period of utilization								Period of utilization	Period of utilization								Period of utilization	Period of utilization																
	1	2	3	4	5	6	7	8		1	2	3	4	5	6	7	8		1	2	3	4	5	6	7	8		1	2	3	4	5	6	7	8		1	2	3	4	5	6	7	8									
7																																																					
14																																																					
30																																																					
60																																																					
Min number of "Quick Utilization" >=3									>=5									>=10									>=3									>=5									>=10								
Hit-rate = 100% % utilized limit >= 60%									= 100% % utilized limit >= 60%									= 100% % utilized limit >= 60%									= 100% % utilized limit >= 80%									= 100% % utilized limit >= 80%									= 100% % utilized limit >= 80%								
Period	Period of utilization								Period of utilization	Period of utilization								Period of utilization	Period of utilization								Period of utilization	Period of utilization								Period of utilization	Period of utilization																
	1	2	3	4	5	6	7	8		1	2	3	4	5	6	7	8		1	2	3	4	5	6	7	8		1	2	3	4	5	6	7	8		1	2	3	4	5	6	7	8									
7																																																					
14																																																					
30																																																					
60																																																					
Min number of "Quick Utilization" >=3									>=5									>=10									>=3									>=5									>=10								
Hit-rate = 100% % utilized limit >= 60%									= 100% % utilized limit >= 60%									= 100% % utilized limit >= 60%									= 100% % utilized limit >= 80%									= 100% % utilized limit >= 80%									= 100% % utilized limit >= 80%								
Period	Period of utilization								Period of utilization	Period of utilization								Period of utilization	Period of utilization								Period of utilization	Period of utilization								Period of utilization	Period of utilization																
	1	2	3	4	5	6	7	8		1	2	3	4	5	6	7	8		1	2	3	4	5	6	7	8		1	2	3	4	5	6	7	8		1	2	3	4	5	6	7	8									
7																																																					
14																																																					
30																																																					
60																																																					
Min number of "Quick Utilization" >=3									>=5									>=10									>=3									>=5									>=10								
Hit-rate = 100% % utilized limit >= 60%									= 100% % utilized limit >= 60%									= 100% % utilized limit >= 60%									= 100% % utilized limit >= 80%									= 100% % utilized limit >= 80%									= 100% % utilized limit >= 80%								
Period	Period of utilization								Period of utilization	Period of utilization								Period of utilization	Period of utilization								Period of utilization	Period of utilization								Period of utilization	Period of utilization																
	1	2	3	4	5	6	7	8		1	2	3	4	5	6	7	8		1	2	3	4	5	6	7	8		1	2	3	4	5	6	7	8		1	2	3	4	5	6	7	8									
7																																																					
14																																																					
30																																																					
60																																																					
Min number of "Quick Utilization" >=3									>=5									>=10									>=3									>=5									>=10								
Hit-rate = 100% % utilized limit >= 60%									= 100% % utilized limit >= 60%									= 100% % utilized limit >= 60%									= 100% % utilized limit >= 80%									= 100% % utilized limit >= 80%									= 100% % utilized limit >= 80%								
Period	Period of utilization								Period of utilization	Period of utilization								Period of utilization	Period of utilization								Period of utilization	Period of utilization								Period of utilization	Period of utilization																
	1	2	3	4	5	6	7	8		1	2	3	4	5	6	7	8		1	2	3	4	5	6	7	8		1	2	3	4	5	6	7	8		1	2	3	4	5	6	7	8									
7																																																					
14																																																					
30																																																					
60																																																					
Min number of "Quick Utilization" >=3									>=5									>=10									>=3									>=5									>=10								
Hit-rate = 100% % utilized limit >= 60%									= 100% % utilized limit >= 60%									= 100% % utilized limit >= 60%									= 100% % utilized limit >= 80%									= 100% % utilized limit >= 80%									= 100% % utilized limit >= 80%								
Period	Period of utilization								Period of utilization	Period of utilization								Period of utilization	Period of utilization								Period of utilization	Period of utilization								Period of utilization	Period of utilization																
	1	2	3	4	5	6	7	8		1	2	3	4	5	6	7	8		1	2	3	4	5	6	7	8		1	2	3	4	5	6	7	8		1	2	3	4	5	6	7	8									
7																																																					
14																																																					
30																																																					
60																																																					
Min number of "Quick Utilization" >=3									>=5									>=10									>=3									>=5									>=10								
Hit-rate = 100% % utilized limit >= 60%									= 100% % utilized limit >= 60%									= 100% % utilized limit >= 60%									= 100% % utilized limit >= 80%									= 100% % utilized limit >= 80%									= 100% % utilized limit >= 80%								
Period	Period of utilization								Period of utilization	Period of utilization								Period of utilization	Period of utilization								Period of utilization	Period of utilization								Period of utilization	Period of utilization																
	1	2	3	4	5	6	7	8		1	2	3	4	5	6	7	8		1	2	3	4	5	6	7	8		1	2	3	4	5	6	7	8		1	2	3	4	5	6	7	8									
7																																																					
14																																																					
30																																																					
60																																																					
Min number of "Quick Utilization" >=3									>=5									>=10									>=3									>=5									>=10								
Hit-rate = 100% % utilized limit >= 60%									= 100% % utilized limit >= 60%									= 100% % utilized limit >= 60%									= 100% % utilized limit >= 80%									= 100% % utilized limit >= 80%									= 100% % utilized limit >= 80%								
Period	Period of utilization								Period of utilization	Period of utilization								Period of utilization	Period of utilization								Period of utilization	Period of utilization								Period of utilization	Period of utilization																
	1	2	3	4	5	6	7	8		1	2	3	4	5	6	7	8																																				



5. B-Tests: Anomalies in Customer Data

History

1881

Simon Newcomb

Astronomer

Simon found that the logarithmic reference book contain the digits "1" more than digits "2", the digits "2" more than digits "3," etc.

1938г.

Frank Benford

Physicist

Frank analyzed different reference book with data and he calculated empirical law of distribution of first digits

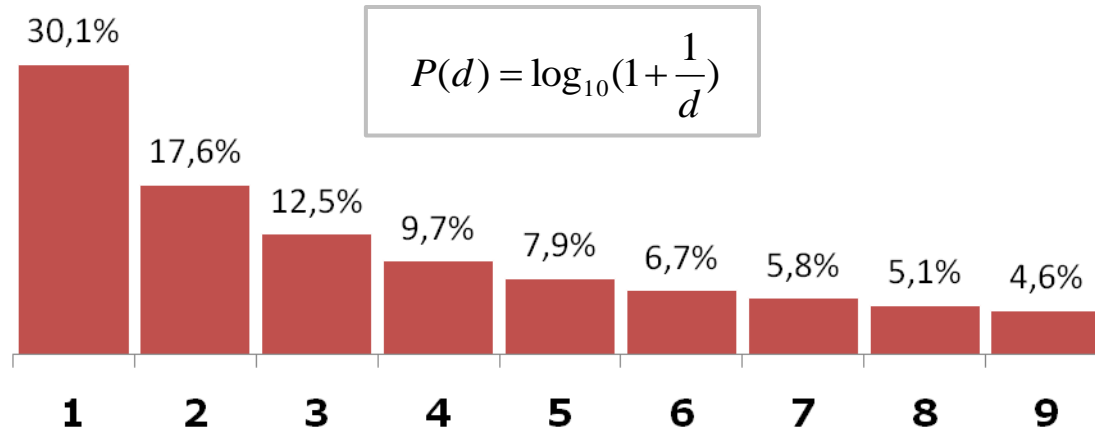
1993г.

Mark J. Nigrini

Accounting

Mark developed tests for financial audit and revealed the embezzlement of \$2 million from Treasury of the state of Arizona

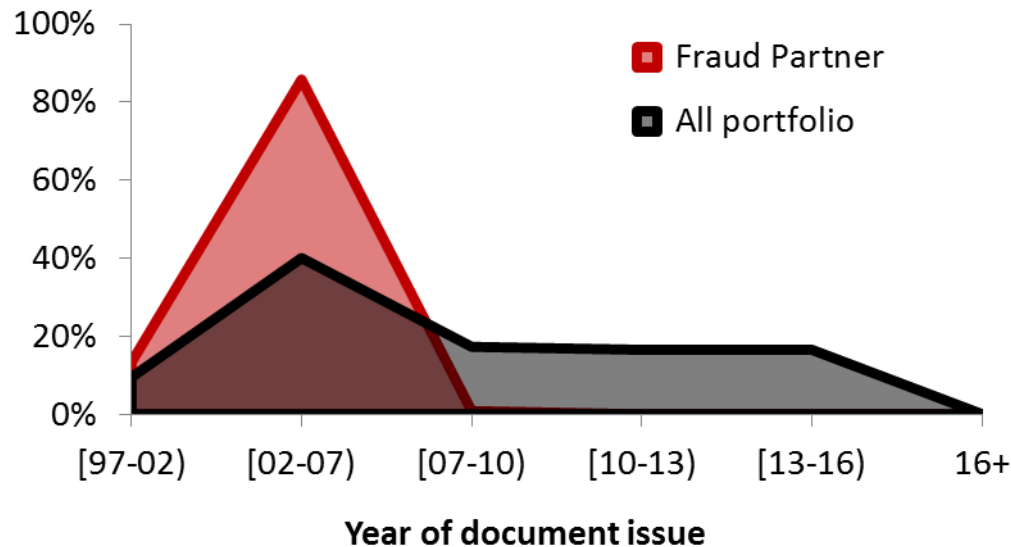
Benford's law



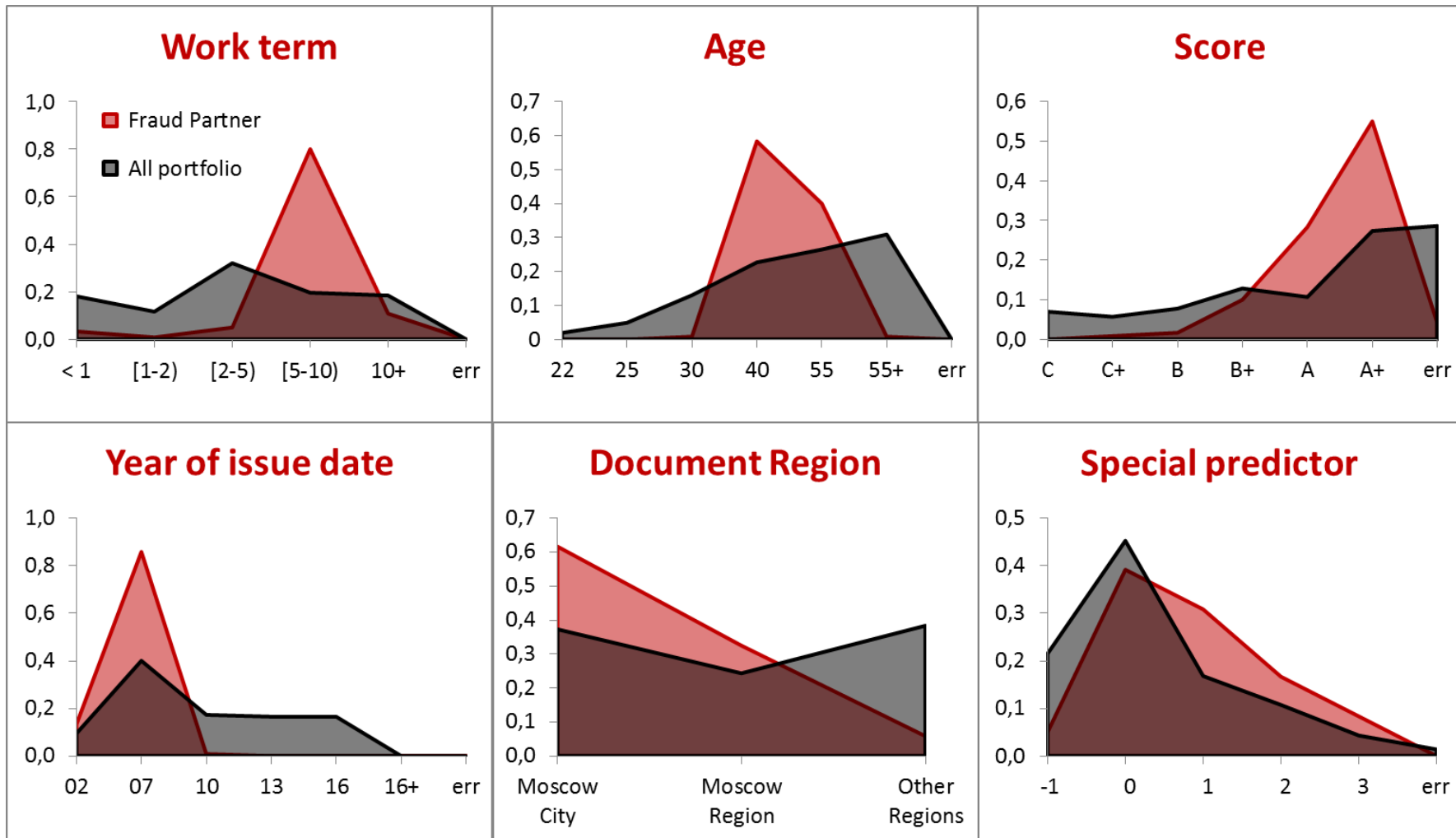
Fraud Case

POS-partner applied for 120 fraudulent loans by using fake documents created by “Documents scan-copy generator”

Distribution of applications



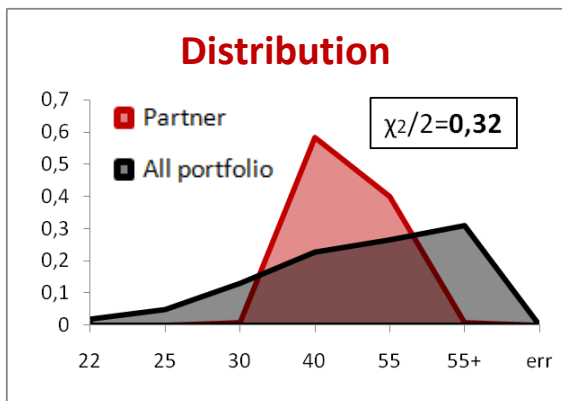
Anomalies



Measures

Chi-squared

$$\chi^2 = \sum_{i=1}^n \frac{(a_i - b_i)^2}{a_i + b_i}$$

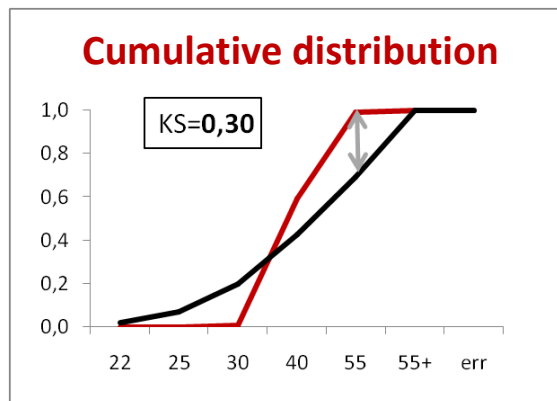


Cons:

Chi-square values are non-linear distributed: they are often concentrated at zero values

Kolmogorov–Smirnov

$$KS = \max_i |F(a_i) - F(b_i)|$$

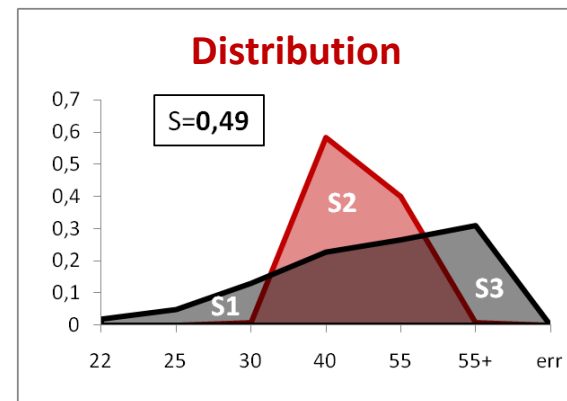


Cons:

K-S may be understated for distributions with multiple local maxima (see example above)

S-metric

$$S = \sum_{i=1}^n \frac{|a_i - b_i|}{2}$$



Pros:

S-metric is distributed linearly and can be simply interpreted: It equals to the half of area of non-overlapping segments

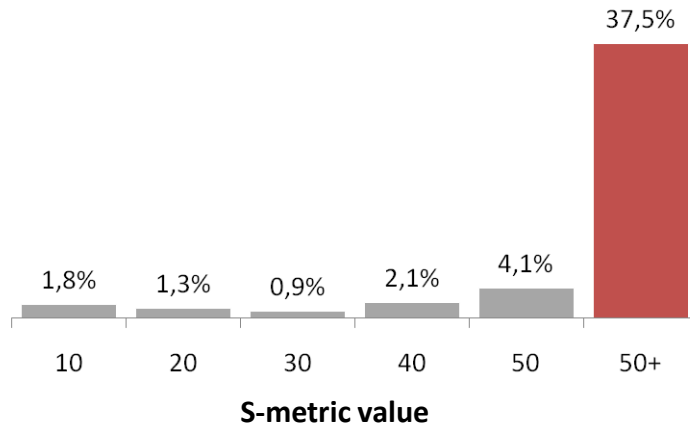
B-Tests

Description	Fraud case values		
	$\chi^2/2$	KS	S
1. Industry	70%	48%	80%
2. 5th figure income (from the end)	43%	55%	61%
5. Term work	39%	53%	60%
6. Working position	40%	54%	59%
4. B-test for home addresses	37%	56%	56%
7. B-test for passport series	31%	50%	50%
8. Passport issue date	33%	50%	50%
3. Age	32%	30%	49%
9. Score	25%	24%	45%
10. Special B-Test for region	16%	34%	34%
11. Region passport	16%	33%	33%
12. Sex	11%	32%	32%
13. Special B-Test for passport series	9%	23%	24%
14. 4th figure income (from the end)	7%	20%	22%

Type of B-Tests

Triggers

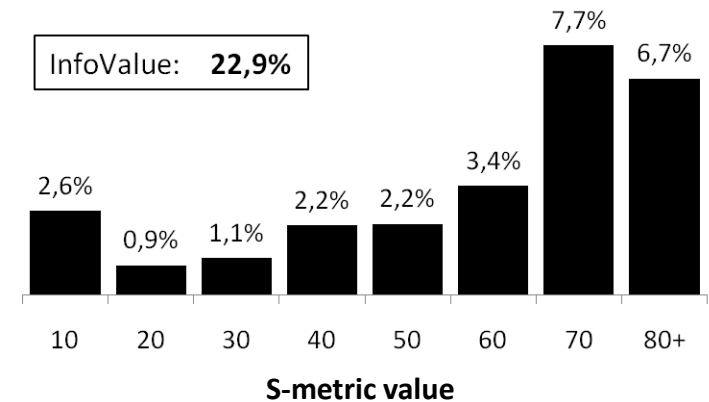
Year of issue
(% fraud)



Triggers allow to select a high-risk group of Employees and Partners , which can be sent immediately for investigation.

Features

4th figure income (from the end)
(% fraud)



The rest of B-tests are used as features. Features with high Information Value are included into Fraud Scoring Card.

Parameters

Min number of applications for one Employee	Number of categories in distribution	Period	S-metric cut-off
>= 10	3	7 days	>=5%
>= 15	6	14 days	>=10%
>= 20	10	30 days	>=20%
>= 30		60 days	>=30%
>= 50			>=40%
			>=50%
			>=60%
			>=70%
			>=80%
			>=90%

14 B-Tests

$$14 \times (5 \times 3 \times 4 \times 10) = 8\,400 \text{ Rules}$$

Trigger “Year of Issue”

Target variable: **30+mob3** \geq **20%**

Period	30	Number of categories					3	Period	30	Number of categories					6	Period	30	Number of categories					10
Min applic.	S-metric cut-off, %						Min applic.	S-metric cut-off, %						Min applic.	S-metric cut-off, %								
	40+	50+	60+	70+	80+	90+		40+	50+	60+	70+	80+	90+		40+	50+	60+	70+	80+	90+			
10							10							10	60%								
15							15							15									
20							20							20	100%								
30	31%						30							30									
50							50							50									
Period	60	Number of categories					3	Period	60	Number of categories					6	Period	60	Number of categories					10
Min applic.	S-metric cut-off, %						Min applic.	S-metric cut-off, %						Min applic.	S-metric cut-off, %								
	40+	50+	60+	70+	80+	90+		40+	50+	60+	70+	80+	90+		40+	50+	60+	70+	80+	90+			
10							10							10									
15							15							15									
20							20							20									
30							30							30									
50							50	20%						50									

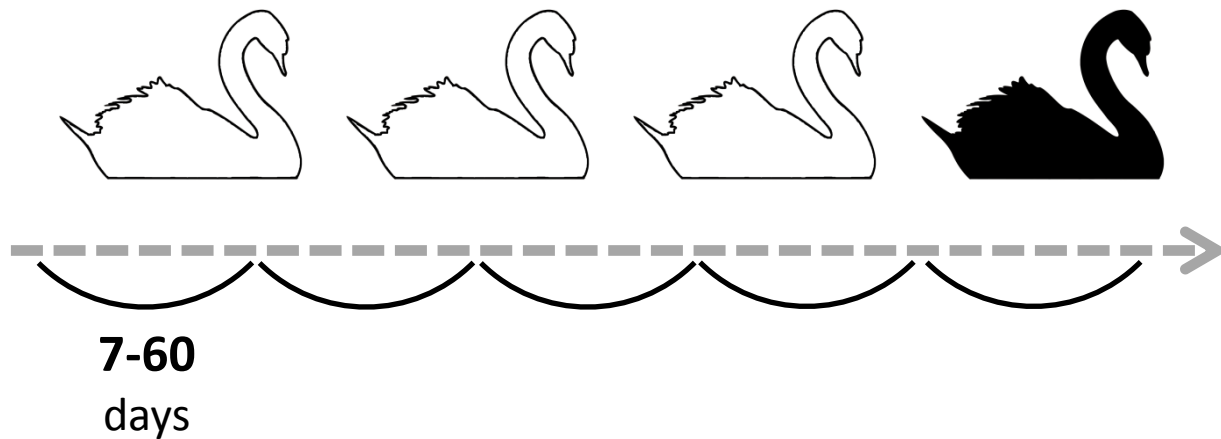


Recommendations and Warnings

1. Remember about “Black Swans”

Analysis horizon: 1-2 years

Multiple overlapping “periods”



2. Mind the Triggers Stability

Out-of-Time Sample



Hierarchy of Parameters

Period (days)	14						
Age of SIM (days)	0-90						
	Min hit						
Hit-rate	4	5	6	7	8	9	10
>=55%							
>=60%							
>=65%							
>=70%							
>=75%							
=100%							

66%

3. Small Data is not a Problem

Accident



Coincidence



Pattern*



* Proved by Newton

4. Expertise is a Must



Thank you for your attention!

If you have any questions
please send me an e-mail:

svafanasev@gmail.com