



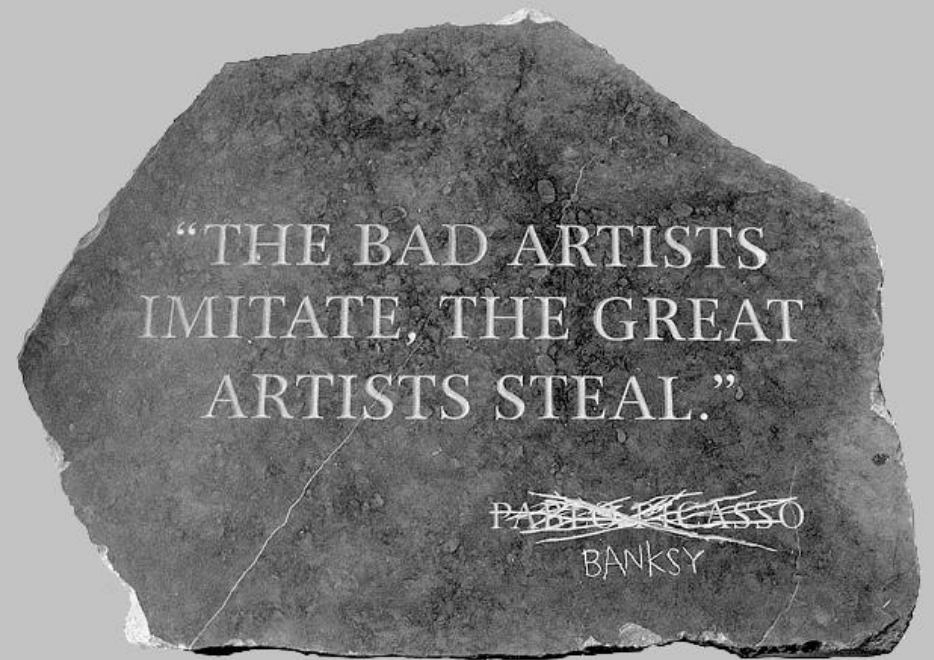
Предиктивная аналитика

**Выявление внутреннего мошенничества
в розничном кредитовании**

Афанасьев Сергей
КБ «Ренессанс Кредит»

Warning System





1. Предикторы на данных FPS

Список правил FPS

1-й набор

Разные заемщики с одним телефоном (не родственники)

2-й набор

У заемщика разные личные данные (адрес, телефон и др.)

3-й набор

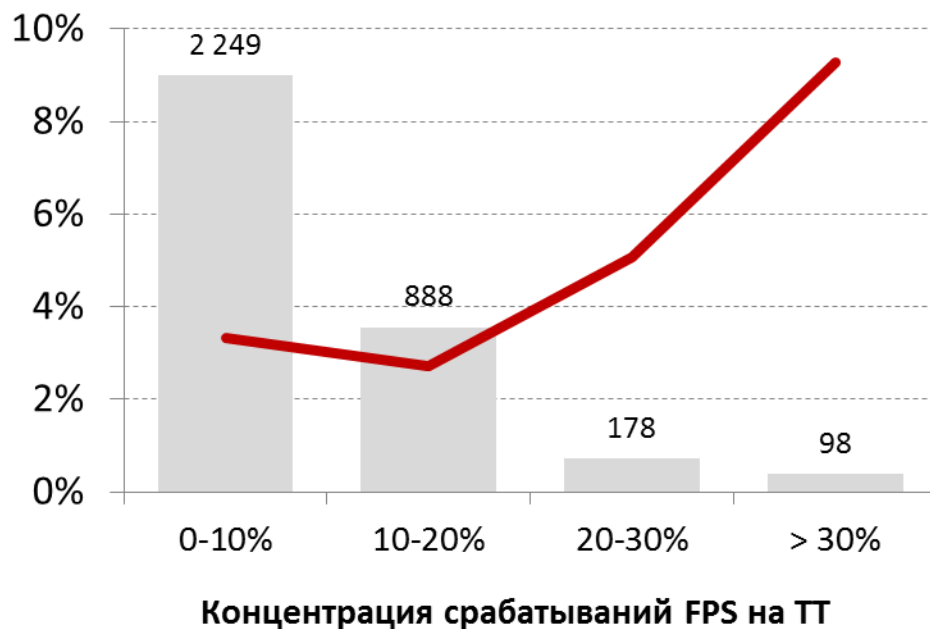
Специальные правила (странная активность, поведение и др.)

FPS_1.11	Один телефон, разные клиенты + доп.условия на ФИО, адреса и др.
FPS_1.12	Один телефон, разные клиенты + доп.условия на ФИО, адреса и др.
FPS_1.13	Использование подставных телефонов
FPS_1.14	Один телефон, разные клиенты + доп.условия на ФИО, адреса и др.
FPS_1.15	Телефон используется разными заемщиками в нескольких банках
FPS_1.16	Телефон встречается у разных заемщиков с разными адресами + доп.условия
FPS_1.17	Телефон встречается у разных заемщиков с разными адресами + доп.условия
FPS_1.18	Телефон используется разными заемщиками в нескольких банках
FPS_2.11	Частая смена мобильного телефона
FPS_2.12	Подозрительное поведение
FPS_2.13	Подозрение на поддельные документы
FPS_2.14	Значительные изменения в данных
FPS_2.15	Значительные изменения в контактных данных
FPS_2.16	Изменение контактных данных заемщика за короткий срок
FPS_2.17	Значительные изменения в контактных данных
FPS_2.18	Изменение контактных данных заемщика за короткий срок
FPS_2.19	После заявки на КЕШ через короткий срок заемщик оформляет заявку на ПОС
FPS_2.20	После отказанной заявки на КЕШ через короткий срок заемщик оформляет заявку на ПОС
FPS_2.21	Высокая кредитная активность заемщика
FPS_2.22	Изменение данных работодателя при посещении других банков
FPS_2.23	Кредитная активность заемщика в высокорисковом сегменте
FPS_2.24	Кредитная активность заемщика в высокорисковом сегменте
FPS_2.25	Высокая кредитная активность заемщика
FPS_2.26	Кредитная активность заемщика в высокорисковом сегменте
FPS_2.27	Высокая кредитная активность заемщика

FPS-концентрации

Высокая концентрация срабатываний правил FPS на ТТ/сотруднике повышает вероятность внутреннего мошенничества

Доля выявленных мошеннических ТТ



Параметры

Минимальное кол-во хитовых заявок на ТТ	Анализируемый период	Концентрация хитовых заявок на ТТ
≥ 3	7 дней	$\geq 5\%$
≥ 5	14 дней	$\geq 10\%$
≥ 10	30 дней	$\geq 15\%$
		$\geq 20\%$
		$\geq 25\%$
		$\geq 30\%$
		$\geq 35\%$
		$\geq 40\%$
		$\geq 45\%$
		$\geq 50\%$
		$\geq 55\%$
		$\geq 60\%$
		$\geq 65\%$

3 x 3 x 13 = **117** комбинаций

Триггеры

Просрочка 30+mob3

Min hit:	>= 3		
Hit-rate	Период (дни)		
	7	14	30
>=5%			
>=10%			
>=15%			
>=20%			
>=25%			
>=30%			
>=35%			
>=40%			
>=45%			
>=50%			
>=55%			
>=60%			
>=65%			

Min hit:	>= 5		
Hit-rate	Период (дни)		
	7	14	30
>=5%			
>=10%			
>=15%			
>=20%			
>=25%			
>=30%			
>=35%			
>=40%			
>=45%			
>=50%			16%
>=55%			18%
>=60%			
>=65%			

Min hit:	>= 10		
Hit-rate	Период (дни)		
	7	14	30
>=5%			
>=10%			
>=15%			
>=20%			
>=25%			
>=30%		15%	
>=35%			24%
>=40%			
>=45%			
>=50%			
>=55%		69%	
>=60%			
>=65%			

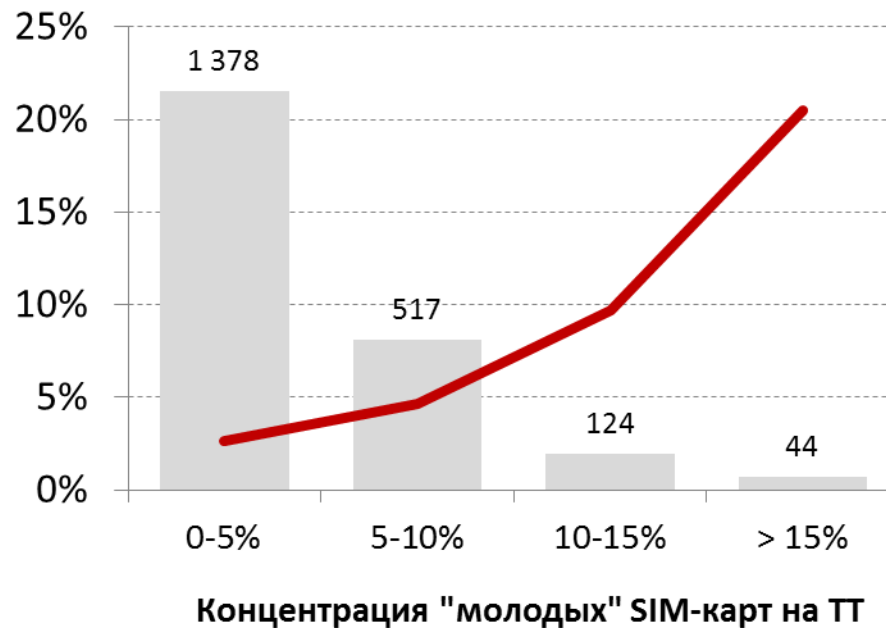


2. Предикторы на данных Мегафона

Молодая SIM-карта

Высокая концентрация заявок с «молодыми» SIM-картами на ТТ/сотруднике повышает вероятность внутреннего мошенничества

Доля выявленных мошеннических ТТ



Параметры

Минимальное кол-во "молодых" SIM-карт на ТТ	"Возраст" SIM-карты	Анализируемый период	Концентрация "молодых" SIM-карт на ТТ
>= 4	0-30 дней	7 дней	>=55%
>= 5	0-90 дней	14 дней	>=60%
>= 6		30 дней	>=65%
>= 7			>=70%
>= 8			>=75%
>= 9			=100%
>= 10			

7 x **2** x **3** x **6** = **252** комбинации

Триггеры

Просрочка 30+mob3

Период (дни):		7						
Возраст сим-карты (дни):		0-30						
	Min hit (кол-во)							
Hit-rate	4	5	6	7	8	9	10	
>=55%								51%
>=60%								
>=65%								
>=70%								
>=75%								
=100%								

Период (дни):		14						
Возраст сим-карты (дни):		0-30						
	Min hit (кол-во)							
Hit-rate	4	5	6	7	8	9	10	
>=55%								
>=60%								
>=65%								
>=70%								
>=75%								
=100%	46%							

Период (дни):		30						
Возраст сим-карты (дни):		0-30						
	Min hit (кол-во)							
Hit-rate	4	5	6	7	8	9	10	
>=55%								44%
>=60%								
>=65%								
>=70%								
>=75%								
=100%								

Период (дни):		7						
Возраст сим-карты (дни):		0-90						
	Min hit (кол-во)							
Hit-rate	4	5	6	7	8	9	10	
>=55%								
>=60%								
>=65%								47%
>=70%								
>=75%								
=100%								

Период (дни):		14						
Возраст сим-карты (дни):		0-90						
	Min hit (кол-во)							
Hit-rate	4	5	6	7	8	9	10	
>=55%								
>=60%								
>=65%								
>=70%								
>=75%								66%
=100%								

Период (дни):		30						
Возраст сим-карты (дни):		0-90						
	Min hit (кол-во)							
Hit-rate	4	5	6	7	8	9	10	
>=55%								
>=60%								
>=65%								
>=70%								
>=75%								59%
=100%								78%

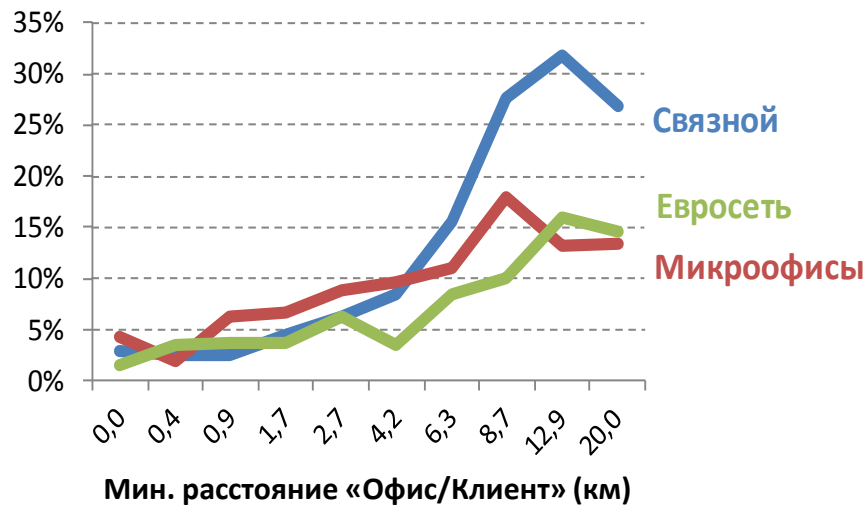
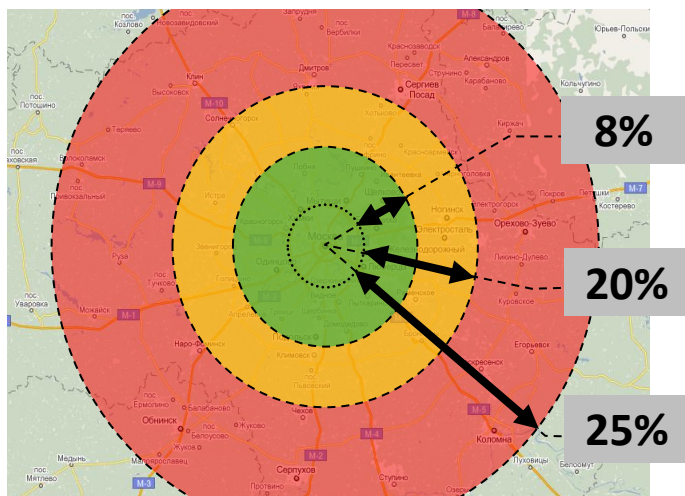


3. Предикторы на данных геолокаций

GPS-зависимости

Чем дальше от офиса продаж живет или работает заемщик, тем выше вероятность того, что он мошенник

FPD-30 в зависимости от расстояния «Офис-Клиент»



Параметры

Расчетное расстояние "ТТ <-> Клиент"	Минимальное кол-во "отдаленных" клиентов	Минимальное расстояние "ТТ <-> Клиент"	Анализируемый период	Концентрация "отдалённых" клиентов
ТТ <-> Факт. адрес	>= 3	0-1 км	7 дней	>=5%
ТТ <-> Рег. адрес	>= 5	1-2 км	14 дней	>=10%
ТТ <-> Работа	>= 10	2-3 км	30 дней	>=15%
		3-5 км	60 дней	>=20%
		5-7 км		>=30%
		7-10 км		>=50%
		10-20 км		>=70%
		>= 20 км		=100%

3 x **8** x **4** x **8** = **768** комбинаций

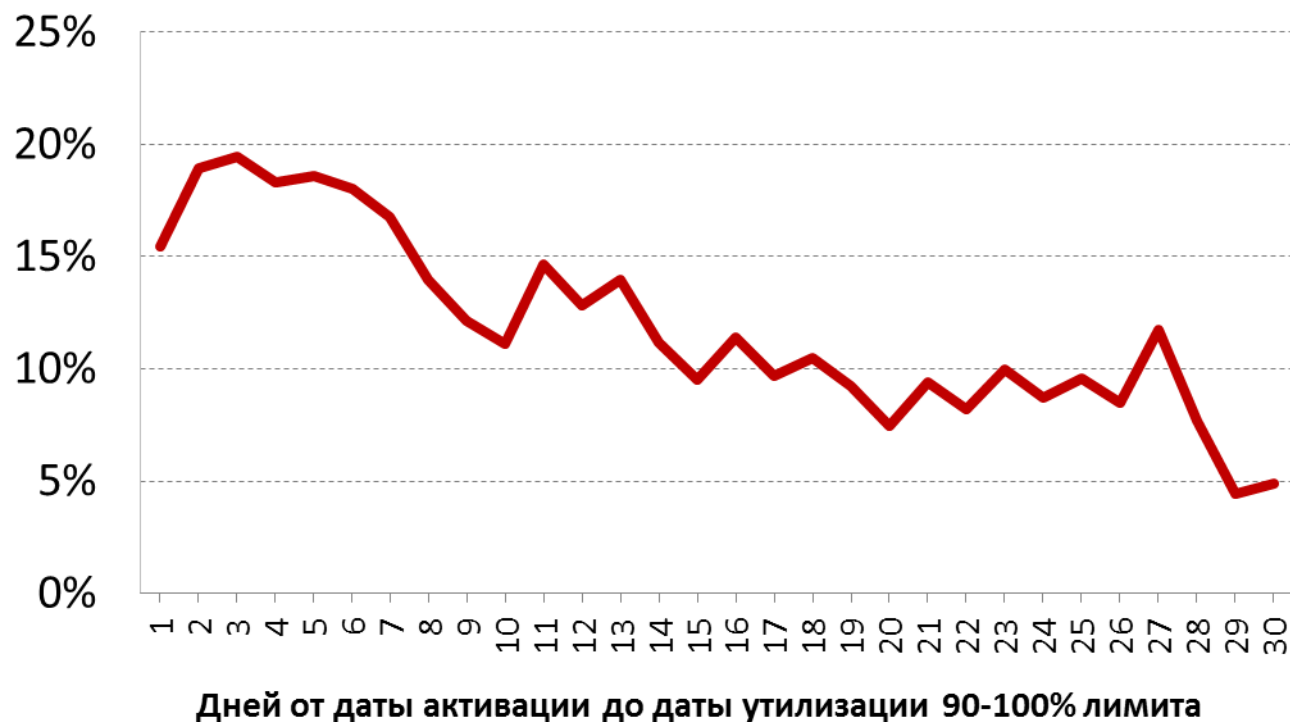


4. Предикторы для кредитных карт

Утилизация лимита

Мошенники обналичивают 90-100% лимита карты в течении первых нескольких дней после активации.

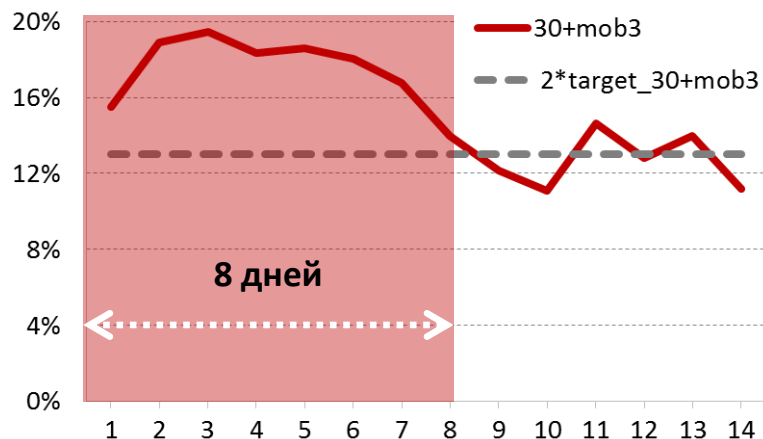
Просрочка 30+mob3



Выбор параметров

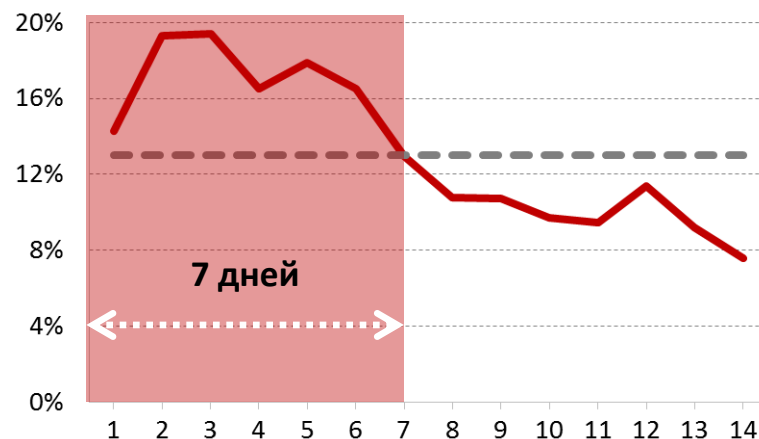
30+mob3

90-100%



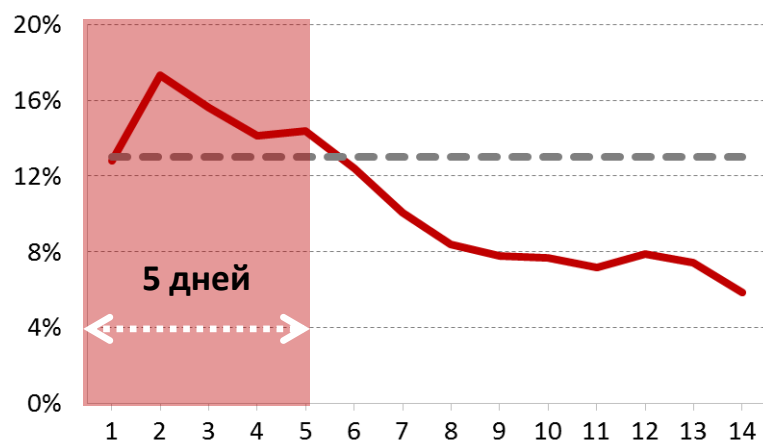
30+mob3

80-90%



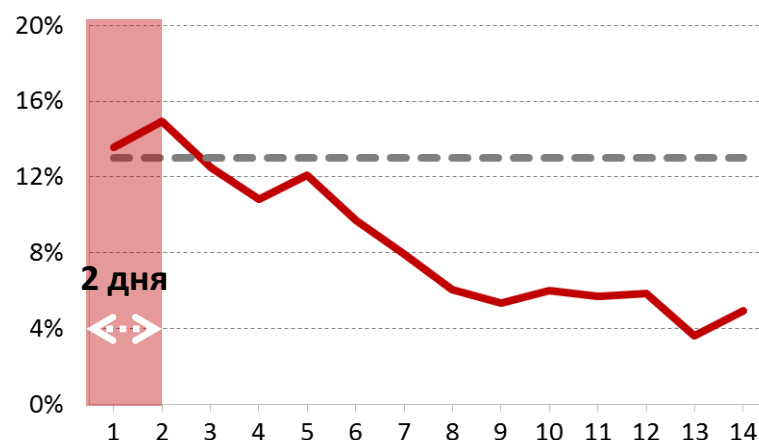
30+mob3

60-80%



30+mob3

40-60%



Параметры

Минимальное кол-во быстрых утилизаций	Доля утилизируемого лимита	Период утилизации	Анализируемый период	Концентрация быстрых утилизаций
≥ 3	20% - 40%	1 день	7 дней	$\geq 5\%$
≥ 5	40% - 60%	2 дня	14 дней	$\geq 10\%$
≥ 10	60% - 80%	3 дня	30 дней	$\geq 15\%$
	80% - 100%	4 дня	60 дней	$\geq 20\%$
		5 дней		$\geq 30\%$
		6 дней		$\geq 50\%$
		7 дней		$\geq 70\%$
		8 дней		$= 100\%$
		9 дней		
		10 дней		

$3 \times 4 \times 10 \times 4 \times 8 = 3\ 840$ комбинаций



5. Б-тесты – аномалии клиентских данных

История

1881г.

Ньюкомб

астроном

Обнаружил, что в логарифмических таблицах единиц больше чем двоек, двоек больше чем троек и т.д.

1938г.

Бенфорд

физик

Проанализировал различные справочники с данными и построил эмпирический закон распределения цифр

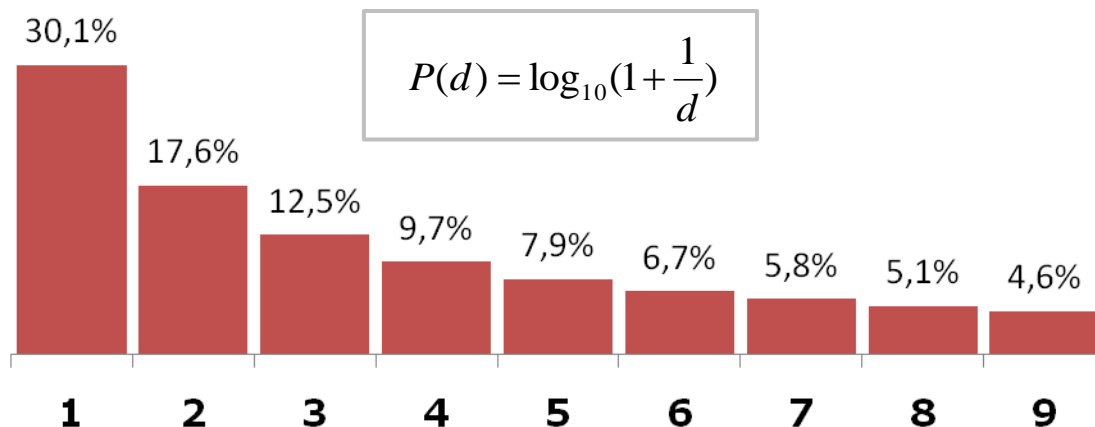
1997г.

Ernst & Yang

компания

Разработали и начали использовать 6 тестов для аудита отчетности компаний и выявления поддельных данных

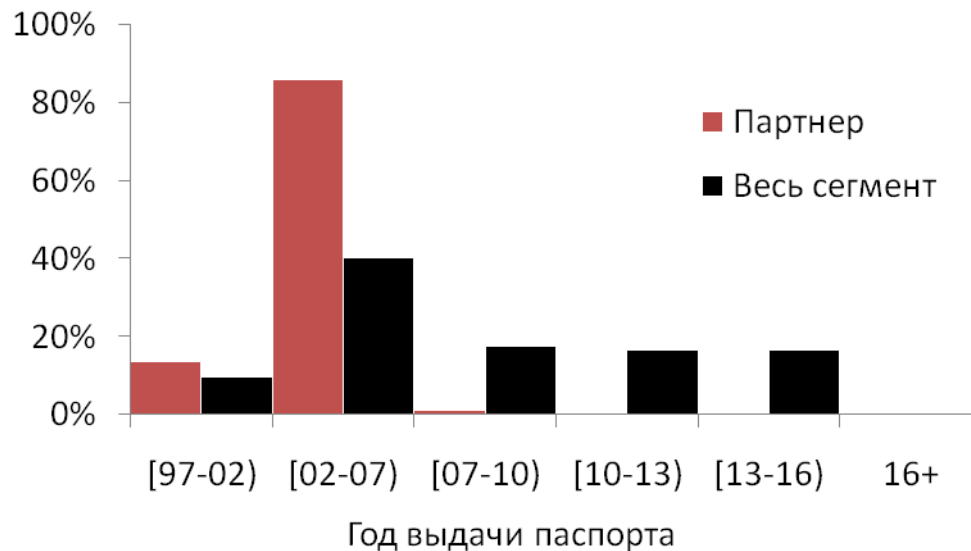
Распределение Бенфорда



Мошенничество

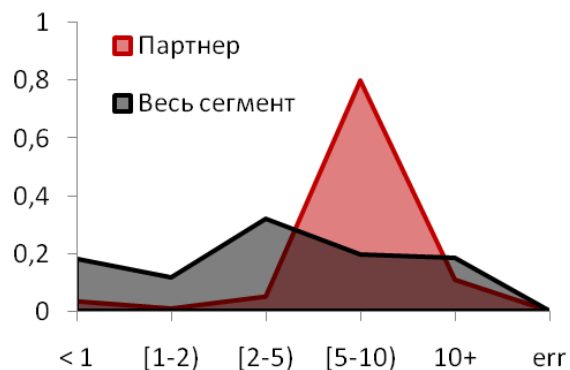
POS-партнер оформил 120 мошеннических кредитов по поддельным документам, используя ПО «Генератор скана паспортов»

Распределение клиентов по году выдачи паспорта

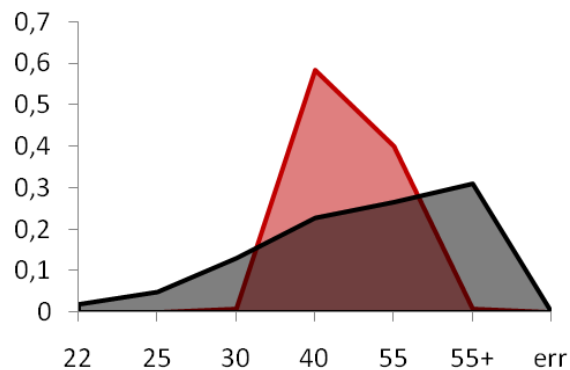


Аномалии

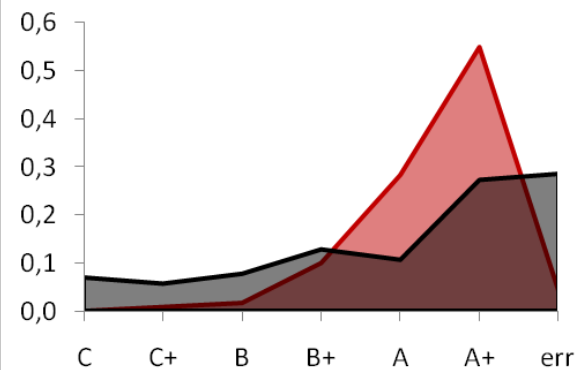
Срок работы



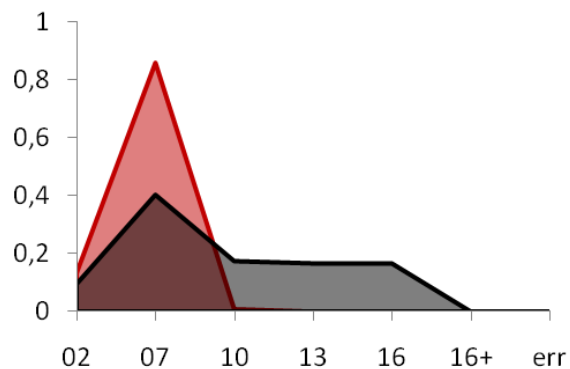
Возраст



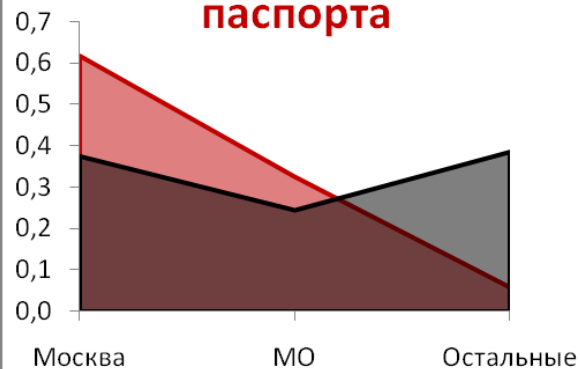
Рейтинг клиента



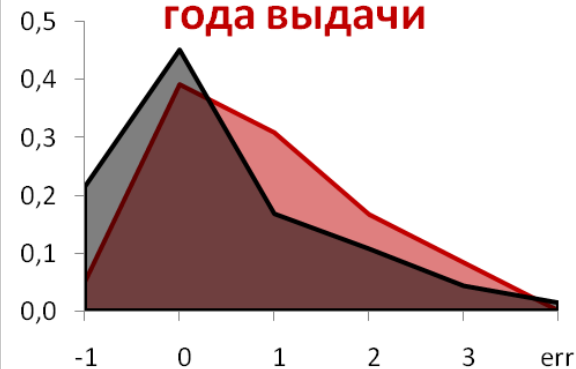
Год выдачи паспорта



Регион выдачи паспорта



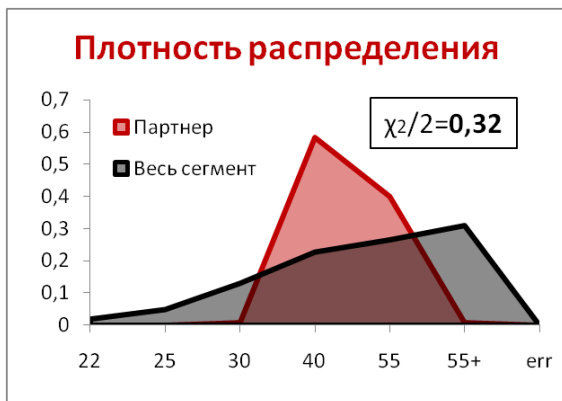
Разность серии и года выдачи



Статистики

Хи-квадрат

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



Минусы:

Значения нормированной статистики Хи-квадрат чаще сосредоточены у нуля и реже у единицы

Колмогорова-Смирнова

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



Минусы:

Для распределений с несколькими локальными максимумами статистика KS может быть заниженной

S-статистика

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



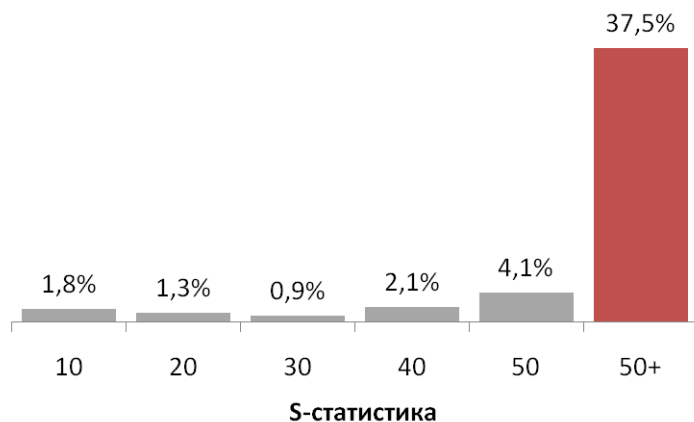
Плюсы:

S-статистика просто интерпретируется — это половина площади под непересекающимися областями

Б-тесты

Триггер

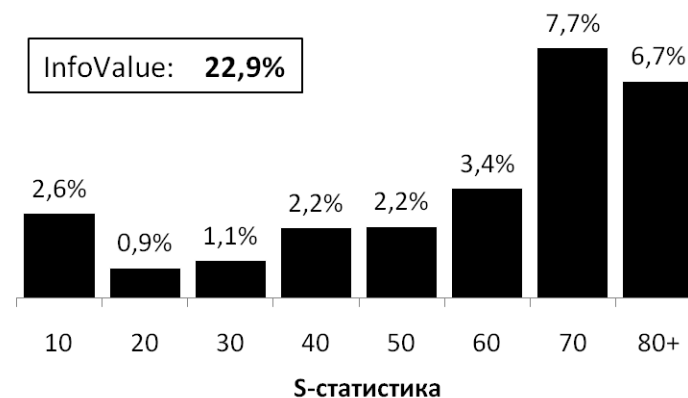
"Год выдачи паспорта по серии"
(доля fraud-точек)



Триггер позволяет выделить высокорисковую группу ТО/сотрудников, которую эффективно отправлять на расследование. По оставшейся выборке пересчитывается InfoValue для использования Б-теста в антифрод-скоркарте.

Предиктор

"4-я с конца цифра дохода"
(доля fraud-точек)



Б-тесты, с помощью которых нельзя выделить высокорисковую группу ТО/сотрудников, используются как предикторы. Предикторы с высоким InfoValue включаются в антифрод-скоркарту.

Параметры

Минимальное кол-во заявок	Кол-во категорий в распределении	Анализируемый период	Порог S-статистики
>= 10	2	7 дней	>=5%
>= 25	3	14 дней	>=10%
>= 20	5	30 дней	>=20%
>= 30	10	60 дней	>=30%
>= 50			>=40%
			>=50%
			>=60%
			>=70%
			>=80%
			>=90%

5 x **4** x **4** x **10** = **800** комбинаций

Триггеры (срез)

Анализируемый период (дни):				60
Минимальное кол-во заявок на ТТ:				>= 10
Кол-во категорий в распределении:				2-6
Описание правила	Тип	S-порог	fraud-group	Info Value
Год по серии паспорта	trigger	>=60%	37,5%	12,2%
Год выдачи паспорта	trigger	>=50%	30,0%	10,6%
5-я цифра дохода (с конца)	predictor	>=50%	3,9%	34,0%
Совпадение адресов	predictor	>=50%	3,5%	28,5%
4-я цифра дохода (с конца)	predictor	>=70%	7,5%	22,9%
Срок работы	predictor	>=50%	5,7%	15,2%
Регион по паспорту	predictor	>=90%	5,9%	13,6%
Возраст	predictor	>=50%	8,2%	12,5%
Регион ТО = Регион по паспорту	predictor	>=60%	3,9%	11,2%
Пол	-	>=30%	3,8%	8,3%
Рейтинг	-	>=40%	2,4%	7,6%
Год по серии минус дата	-	>=30%	2,4%	4,1%
Отрасль	-	-	-	-
Должность	-	-	-	-



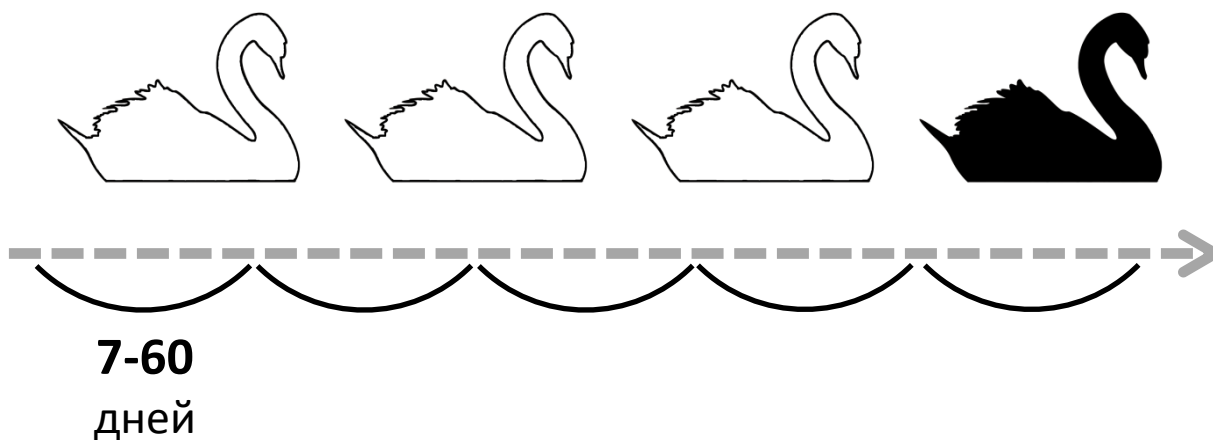
!?

Рекомендации и ограничения

1. Помните о «черных лебедях»

Горизонт анализа: 1-2 года

Достигается за счет покрытия несколькими анализируемыми периодами



2. Следите за устойчивостью

Out-of-time выборка



Вложенность параметров

Период (дни):	14						
Возраст сим-карты (дни):	0-90						
	Min hit (кол-во)						
Hit-rate	4	5	6	7	8	9	10
>=55%							
>=60%							
>=65%							
>=70%							
>=75%							
=100%							

66%

3. Маленькие данные – не ограничение

Случайность



Совпадение



Статистика*



*Проверено Ньютоном

4. Первым делом экспертиза



Спасибо за внимание!

Афанасьев Сергей

Исполнительный директор

Начальник управления расследования мошенничества

КБ «Ренессанс Кредит»

safanasev@rencredit.ru