

Сергей Афанасьев

Ренессанс Кредит

Как переход на Python помогает
улучшить клиентскую аналитику

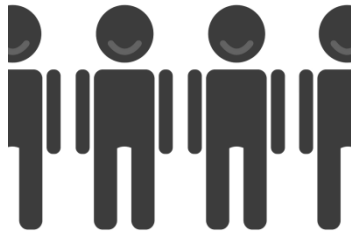


Банк «Ренессанс Кредит»

75 млн
заявок



13 млн
клиентов



9 млн
кредитов



Модели крутятся...

Risk

IL

IL STREET NO HIT	<input type="checkbox"/>
IL STREET HIT	<input type="checkbox"/>
IL REPEAT	<input type="checkbox"/>

GP

GP STREET	<input type="checkbox"/>
GP REPEAT TOPUP	<input type="checkbox"/>
GP REPEAT NON TOPUP	<input type="checkbox"/>

GP X-SELL

GP-XS REPEAT TOPUP	<input type="checkbox"/>
GP-XS REPEAT NON TOPUP	<input type="checkbox"/>

BLENDING

MEGAFON IL STR HIT	<input type="checkbox"/>
MEGAFON IL STR NOHIT	<input type="checkbox"/>
MEGAFON IL REPEAT	<input type="checkbox"/>
MEGAFON GP STREET	<input type="checkbox"/>

OTHERS

INCOME	<input type="checkbox"/>
OVETINDEBTNESS	<input type="checkbox"/>
CC CLIP	<input type="checkbox"/>
INACTIVE	<input type="checkbox"/>
HARD DECLINE	<input type="checkbox"/>

CRM

PTB

PTB OUT CALLS FRESH	<input type="checkbox"/>
PTB OUT CALLS NONFRESH	<input type="checkbox"/>
IL PTB	<input type="checkbox"/>
PTB SMS ROBOT	<input type="checkbox"/>
PTB SMS NON_ROBOT	<input type="checkbox"/>
PTB OUT TRIGGERS	<input type="checkbox"/>
PTB INBOUND CALLS	<input type="checkbox"/>
PTB BY CHANNEL COMM.	<input type="checkbox"/>

ATTRITION

CC ATTRITION	<input type="checkbox"/>
--------------	--------------------------

Collection

ALLOCATION

ALLOCATION "31-60"	<input type="checkbox"/>
ALLOCATION "61-90"	<input type="checkbox"/>
ALLOCATION "91-120"	<input type="checkbox"/>

PRE DELINQUENCY

PRE DELINQUENCY GP/IL	<input type="checkbox"/>
PRE DELINQUENCY CC	<input type="checkbox"/>

FIRST/SECOND WASH

FW	<input type="checkbox"/>
SW	<input type="checkbox"/>

SOFT

ENTRY	<input type="checkbox"/>
SELF CURE	<input type="checkbox"/>
PTP	<input type="checkbox"/>

OTHERS

LEGAL	<input type="checkbox"/>
RECOVERY	<input type="checkbox"/>
AGENCY SCORE	<input type="checkbox"/>

Anti-fraud

WARNING SYSTEM

WS OLD RULES	<input type="checkbox"/>
WS B-TESTS	<input type="checkbox"/>
WS FPS	<input type="checkbox"/>
WS MEGAFON	<input type="checkbox"/>
WS BKI	<input type="checkbox"/>

TRANSACTIONAL FRAUD

OLD RULES	<input type="checkbox"/>
SOCIAL FRAUD P2P	<input type="checkbox"/>
SOCIAL FRAUD C2C	<input type="checkbox"/>
DEPOSIT FRAUD	<input type="checkbox"/>

Finance

IFRS-9 PD MOB_1-5

PD GP STREET	<input type="checkbox"/>
PD GP REPEAT TOPUP	<input type="checkbox"/>
PD GP REPEAT NON TOPUP	<input type="checkbox"/>
PD IL STREET	<input type="checkbox"/>
PD IL REPEAT	<input type="checkbox"/>

IFRS-9 PD MOB_6-12_NODEF

PD GP STREET	<input type="checkbox"/>
PD GP REPEAT TOPUP	<input type="checkbox"/>
PD GP REPEAT NON TOPUP	<input type="checkbox"/>
PD IL STREET	<input type="checkbox"/>
PD IL REPEAT	<input type="checkbox"/>

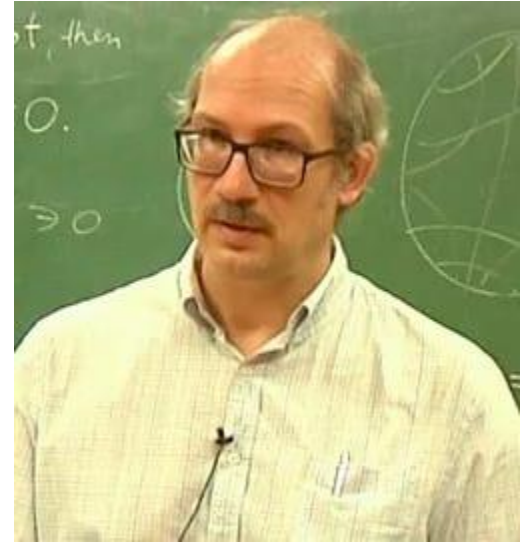
IFRS-9 PD MOB_6-12_DEF

PD GP STREET	<input type="checkbox"/>
PD GP REPEAT TOPUP	<input type="checkbox"/>
PD GP REPEAT NON TOPUP	<input type="checkbox"/>
PD IL STREET	<input type="checkbox"/>
PD IL REPEAT	<input type="checkbox"/>

IFRS-9 OTHERS

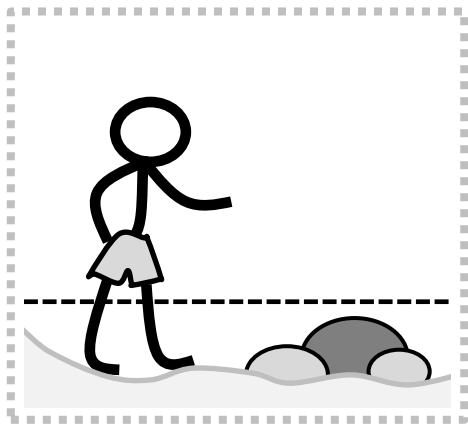
LGD	<input type="checkbox"/>
EAD	<input type="checkbox"/>
MACRO	<input type="checkbox"/>

– Здесь можно наступить на
маленькие подводные грабли

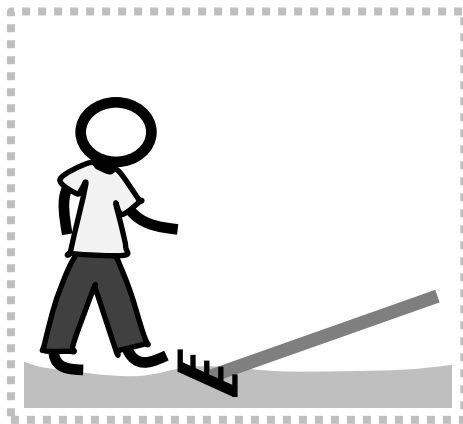


Дынников И.А.

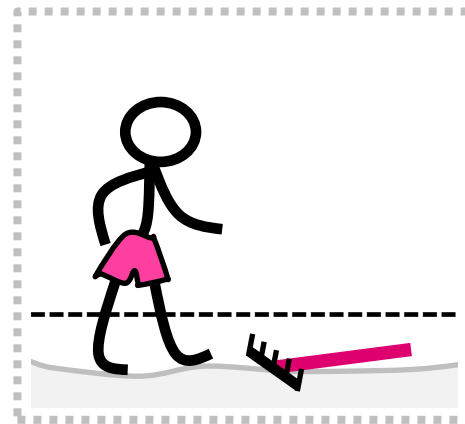
д.ф.-м.н., профессор РАН



Подводные
камни



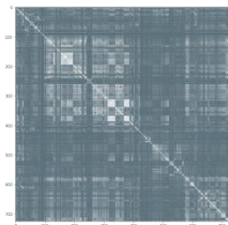
Грабли
обыкновенные



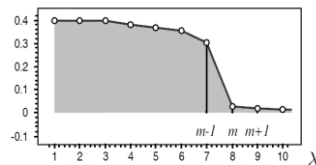
Маленькие
подводные грабли

ОЛДФАГИ

Matrix (Scoring)



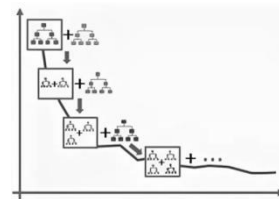
PCA (ML)



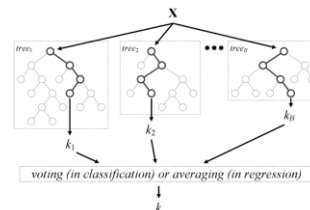
$$E_m = \frac{\|GU^T - F\|^2}{\|F\|^2} = \frac{\lambda_{m+1} + \dots + \lambda_n}{\lambda_1 + \dots + \lambda_n} \leq \varepsilon$$

НЕОФИТЫ

Boosting (Kaggle)



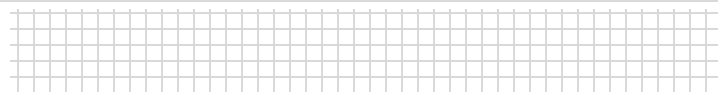
Forest (Science)



В НАУЧНЫХ СТАТЬЯХ

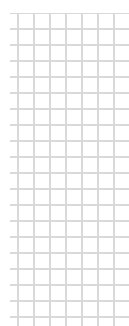
1 000 – 7 000 переменных

100



20 – 100

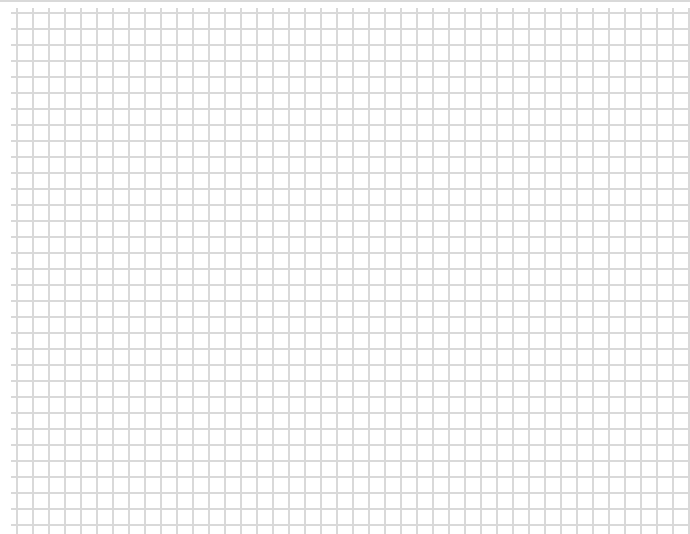
1 000 – 10 000



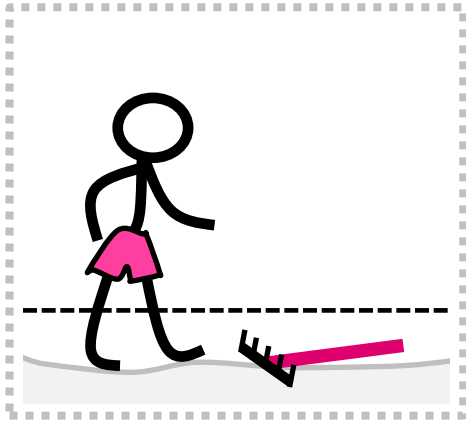
РЕАЛЬНОСТЬ

300 – 1 500 переменных

100 000 – 1 000 000



Проблемы продвинутых методов



Долго и дорого: Two-Forest медленный на больших датасетах и жрет много памяти



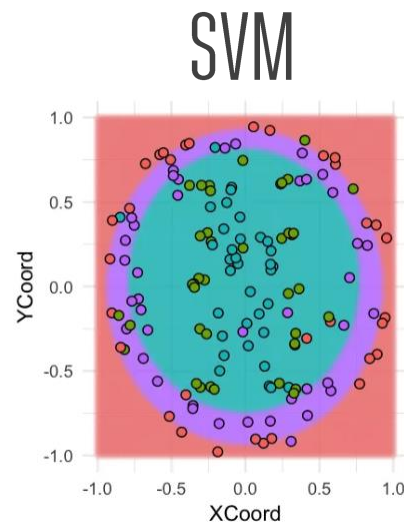
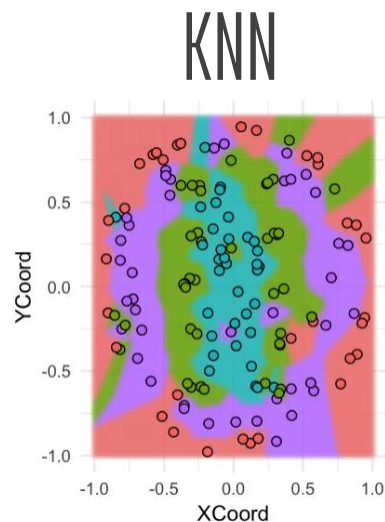
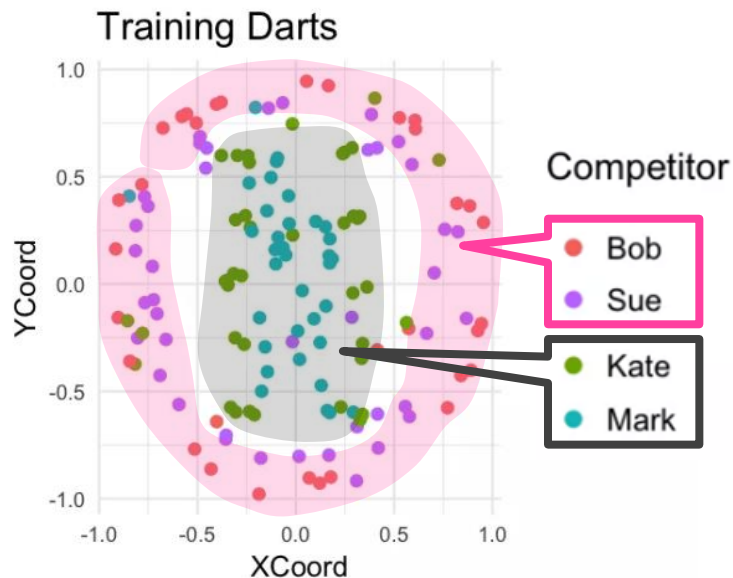
Проклятие размерности: самая сильная фича вылетит с 80% вероятностью

Схема отбора переменных

+2.5% GINI



Как работает stacking?



Ben Gorman. A Kaggler's Guide to Model Stacking in Practice. 12.27.2016
<http://blog.kaggle.com/2016/12/27/a-kagglers-guide-to-model-stacking-in-practice/>

Blending на внутренних данных

APPLICATION

- Пол
- Семейное положение
- Доход
- Образование
- Должность
- И др.

BUREAU

- Исторические просрочки
- Текущие просрочки
- PTI
- DTI
- И др.

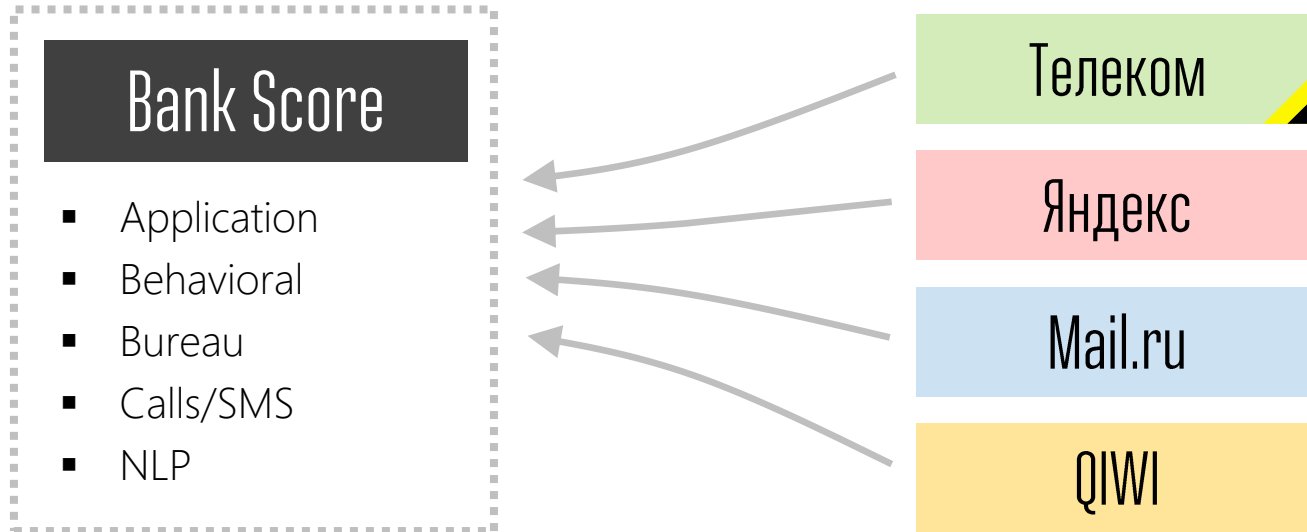
CALLS/SMS

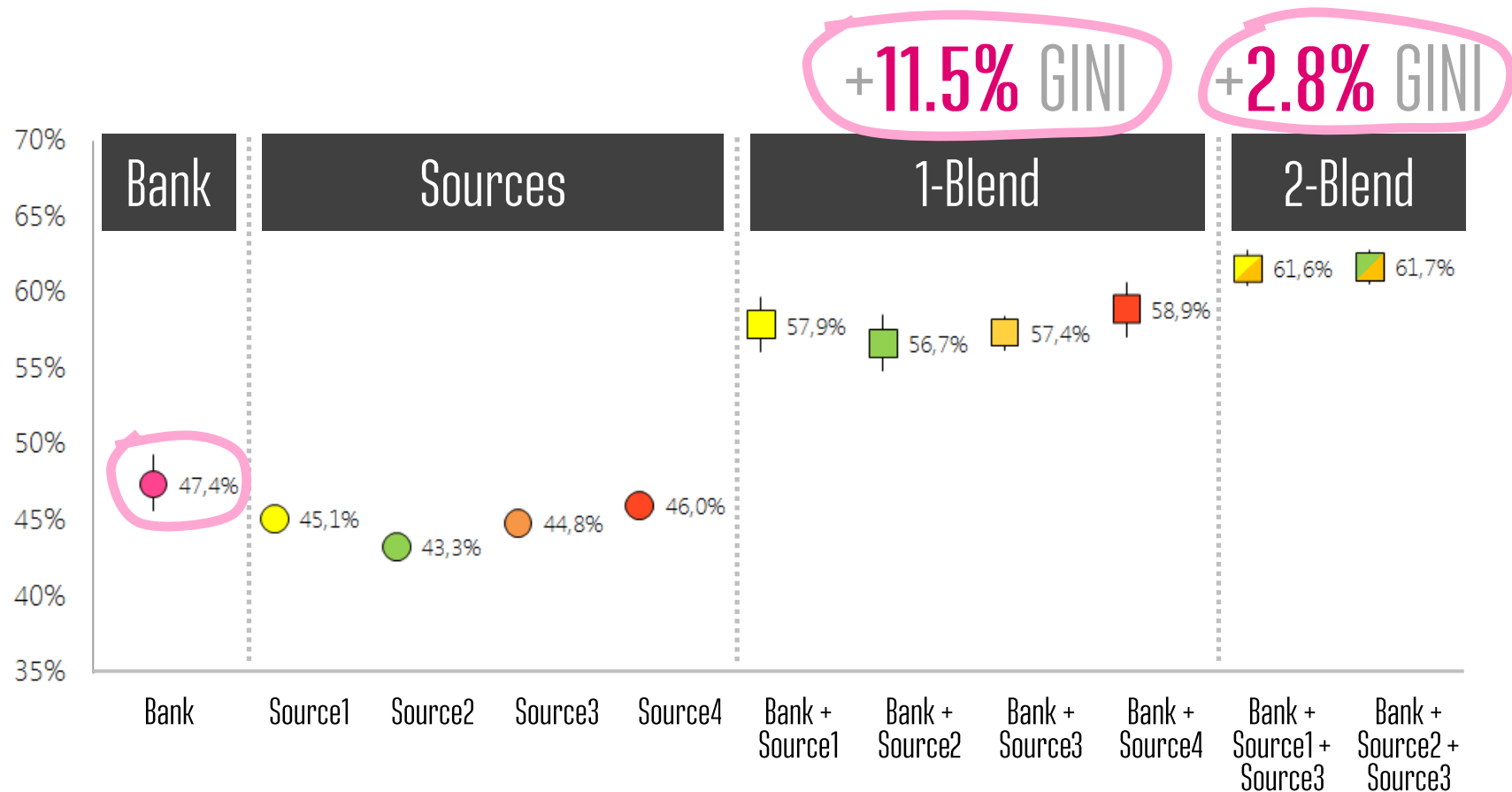
- Кол-во звонков
- Кол-во SMS
- % негативных
- % позитивных
- % недоставленных SMS
- И др.

NLP

- Тексты исходящих звонков X-Sell
- Тексты исходящих звонков Collection
- Тексты входящих звонков
- И др.

Blending с внешними источниками





Бизнес-метрики

		PREDICTED	
OBSERVED	True	True Positive	False Negative
	False	False Positive	True Negative

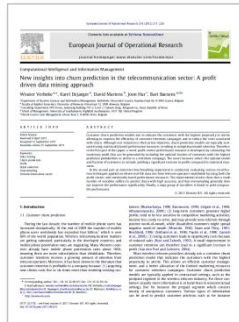
Credit Scoring

$$P = (1 - RR) \left(\frac{t_0}{1 - t_0} * \text{TN} - \text{FN} \right) * s$$

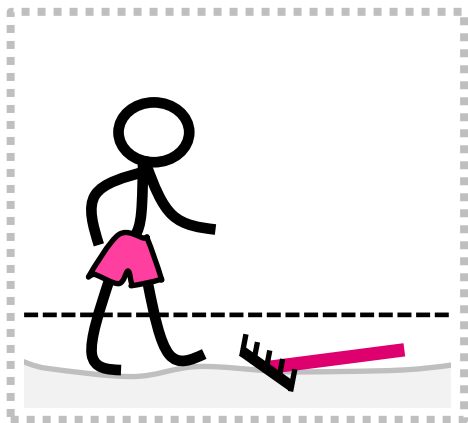
PTB-models

$$P = m * \text{TP} * s - c * (\text{TP} + \text{FP})$$

Идея считать деньги не нова



У осла два уха?



Вся база

X млн клиентов

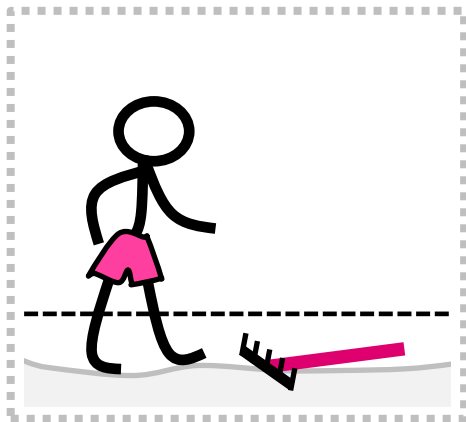
Кампания

$0.1 * X$ млн клиентов

Результат

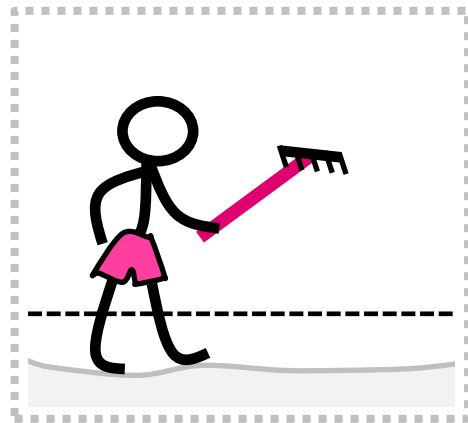
Выгодно звонить всем!

Осторожно ступенька!



	Δ GINI	Δ PROFIT, %
SCORING 1-blend	11.5%	+3.0%
SCORING 2-blend	2,8%	+1.8%
PTB 1-blend	6.0%	$-\infty$

Маленькие
подводные грабли
не смертельны





– По мере общения с людьми из сферы DS приходишь к выводу о том, что в DataScience два типа людей: первые делают кучу ошибок и честно в этом признаются, вторые делают кучу ошибок и строят из себя экспертов. Лучше

сказать «я попробую сделать», «я кое-чему научу». Слишком сложная многогранная область для громких заявлений. Главные причины провала DS проектов – экономия на людях, сильных предметниках, ну не любим мы платить, и боязнь конкуренции, страх перед коллаборацией, «я сейчас расскажу ему про то, как делаю свои фишки с помощью unsupervised алгоритмов, а он потом лекарство от СПИДа изобретет».

Артем Груздев, ИЦ «ГЕВИССТА»

Спасибо за
внимание!

Афанасьев Сергей

Исполнительный директор
Начальник управления статистического анализа

КБ «Ренессанс Кредит»

svafanasev@gmail.com

safanasev@rencredit.ru