



Факультет вычислительной математики и кибернетики МГУ
Кафедра Математических методов прогнозирования

ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ В УПРАВЛЕНИИ РОЗНИЧНЫМ КРЕДИТНЫМ РИСКОМ

Николай Филипенков

к.ф.-м.н.

Email: n.filipenkov@mail.ru

Спецсеминар «Интеллектуальный анализ данных в бизнесе»

Москва, 1 октября 2010



КРАТКАЯ БИОГРАФИЯ



Николай Филипенков:

09.2001 – поступил на ВМиК

09.2003 – поступил на кафедру ММП

Научный руководитель: член-корреспондент РАН
Рудаков К.В.

11.2005 – поступил на работу в Банк Москвы

07.2006 – окончил ВМиК, поступил в аспирантуру ВЦ РАН

03.2008 – возглавил Отдел анализа скорингового кредитования
Банка Москвы

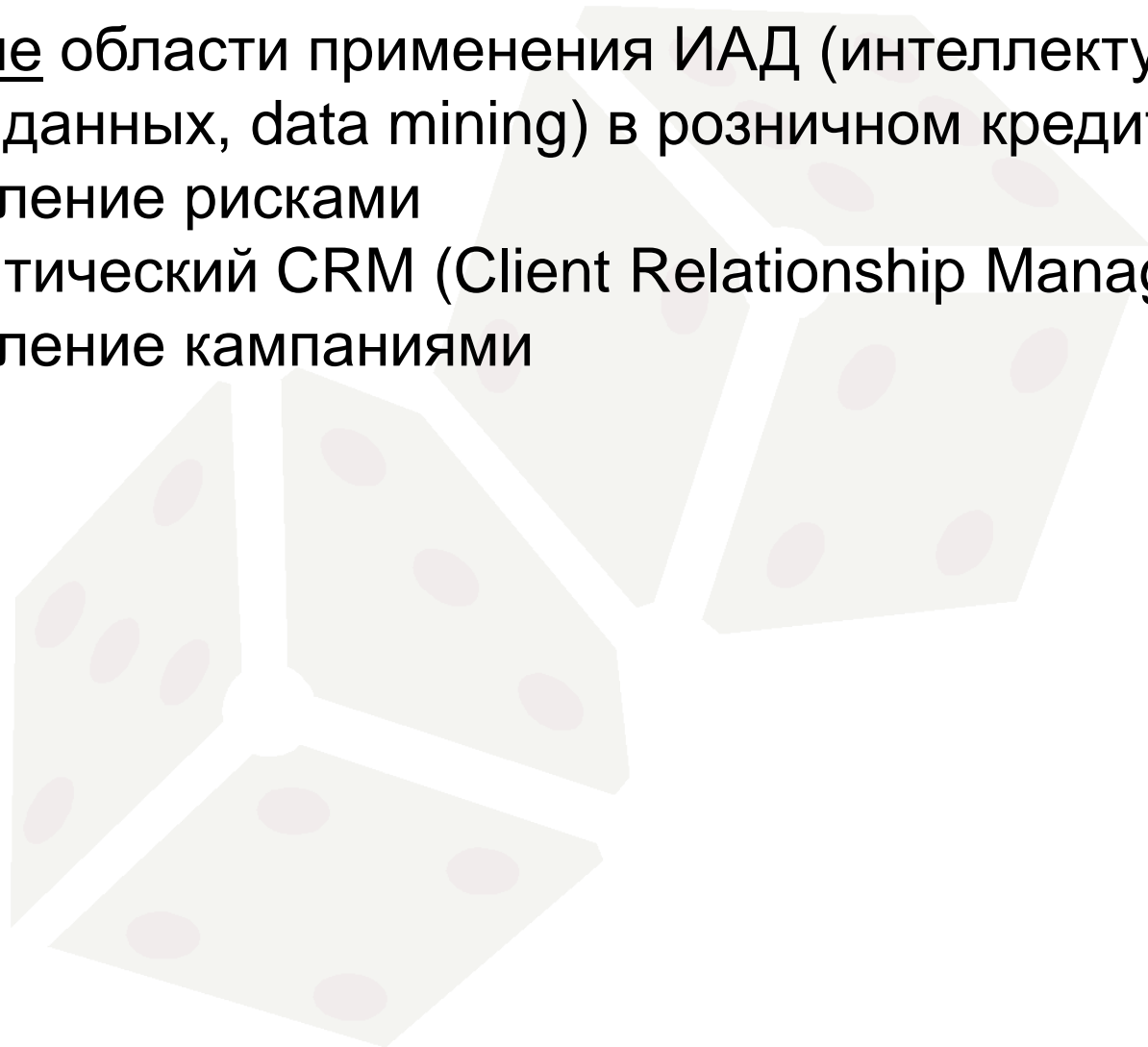
04.2010 – защитил кандидатскую диссертацию
«Об алгоритмах прогнозирования процессов с плавно
меняющимися закономерностями» (к.ф.-м.н.)



ОСНОВНЫЕ ОБЛАСТИ ПРИМЕНЕНИЯ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ

Основные области применения ИАД (интеллектуального анализа данных, data mining) в розничном кредитовании:

- Управление рисками
- Аналитический CRM (Client Relationship Management),
Управление кампаниями





РИСК

Риск – возможность потери части своих активов, недополучения доходов или появления доп. расходов

Виды рисков:

- **РЫНОЧНЫЙ** – возможность потерь в результате колебаний процентных ставок, курсов валют, цен акций и товарных контрактов.
- **КРЕДИТНЫЙ** – возможность потерь в результате неспособности контрагентов (заемщиков) исполнять свои обязательства.
- **ОПЕРАЦИОННЫЙ** – возможность потерь вследствие технических ошибок при проведении операций, умышленных и неумышленных действий персонала, аварийных ситуаций, сбоев аппаратуры, несанкционированного доступа к информационным системам и т. д.
- **Прочие** (Ликвидности, Репутационный, ...)



УПРАВЛЕНИЕ РИСКАМИ

Способы управления рисками:

- Страхование
 - Хеджирование
 - Распределение
 - Резервирование (самострахование)
 - Диверсификация
 - Минимизация (управление активами и пассивами)
 - Избежание (отказ от связанной с риском операции)
- Передача риска 3-му лицу*
- Оставление риска на собственном удержании*

Критерии:

- Вероятность наступления (подверженность риску)
- Чистый ущерб вследствие проявления риска



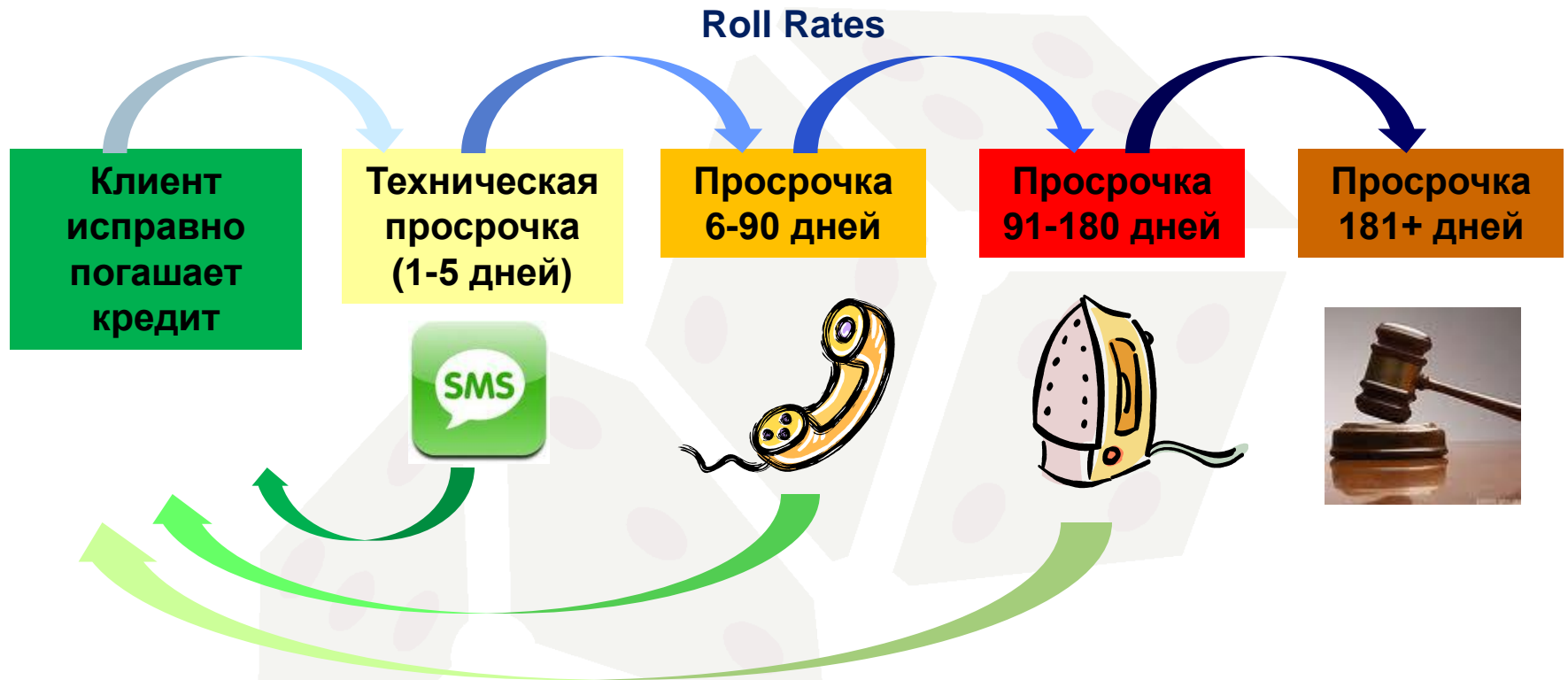
ПРОЦЕДУРА РАССМОТРЕНИЯ АНКЕТЫ-ЗАЯВЛЕНИЯ



Функции	Инструмент проверки
Прогнозирование финансового дефолта (может вернуть?)	Application- скоринговая карта
Выявление мошенничества (хочет вернуть?)	Службы банка, Антимошенническая скоринговая карта



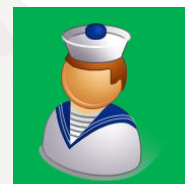
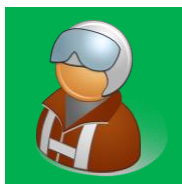
СОПРОВОЖДЕНИЕ КРЕДИТА



Если клиент задолжал несколько платежей и затем погасил часть из них: Вычисление бухгалтерской просрочки происходит по методу LIFO (стек). Задолженность растет пока полностью не погашен долг.



МАРЖА РИСКА



Выдано:

100 000 руб.

100 000 руб.

100 000 руб.

100 000 руб.

100 000 руб.

Возвращено:

100 000 руб.+
x руб.

100 000 руб.+
x руб.

100 000 руб.+
x руб.

100 000 руб.+
x руб.

0 руб.

Процентная ставка включает маржу риска, которая обеспечивает покрытие ожидаемых потерь



МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ В УПРАВЛЕНИИ РИСКАМИ

Задачи риск-менеджмента, использующие математическое моделирование:

- Минимизация потерь (application-, antifraud- скоринг)
- Мониторинг рисков и проактивные действия (behaviour- скоринг)
- Оптимизация сбора задолженности (collection- скоринг)
- Прогноз ожидаемых потерь (моделирование EAD, LGD)
- Стресс-тестирование, прогноз неожиданных потерь



БАЗЕЛЬСКИЙ КОМИТЕТ ПО БАНКОВСКОМУ НАДЗОРУ

The Bank for International Settlements (BIS)

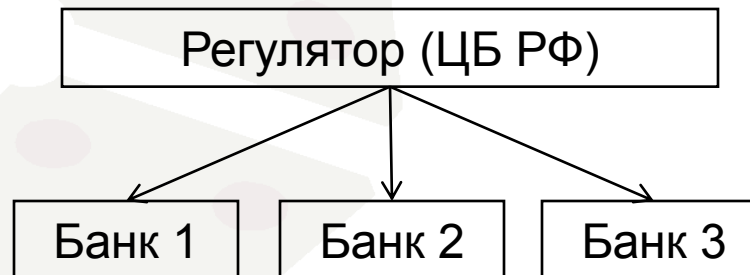
is an international organisation which fosters international monetary and financial cooperation and serves as a bank for central banks.

Established on 17 May 1930, the BIS is the world's oldest international financial organization.

The Basel Committee on Banking Supervision

provides a forum for regular cooperation on banking supervisory matters. Its objective is to enhance understanding of key supervisory issues and improve the quality of banking supervision worldwide.

Basel III: September 2010





КРЕДИТНЫЙ РИСК

ОЖИДАЕМЫЕ ПОТЕРИ (EXPECTED LOSS)

$$EL = EAD * PD * LGD$$

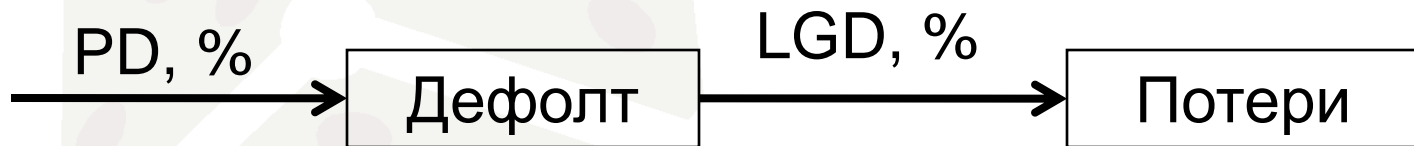
EL (Expected Loss) – ожидаемые потери;

EAD (Exposure at Default) – стоимость, подверженная риску дефолта;

PD (Probability of Default) – вероятность дефолта;

LGD (Loss Given Default) – вероятность потерь в случае дефолта.

Например, событие «дефолт» – попадание в просрочку 90 дней.



Например, EAD (у.е.) = остаток долга – стоимость залога



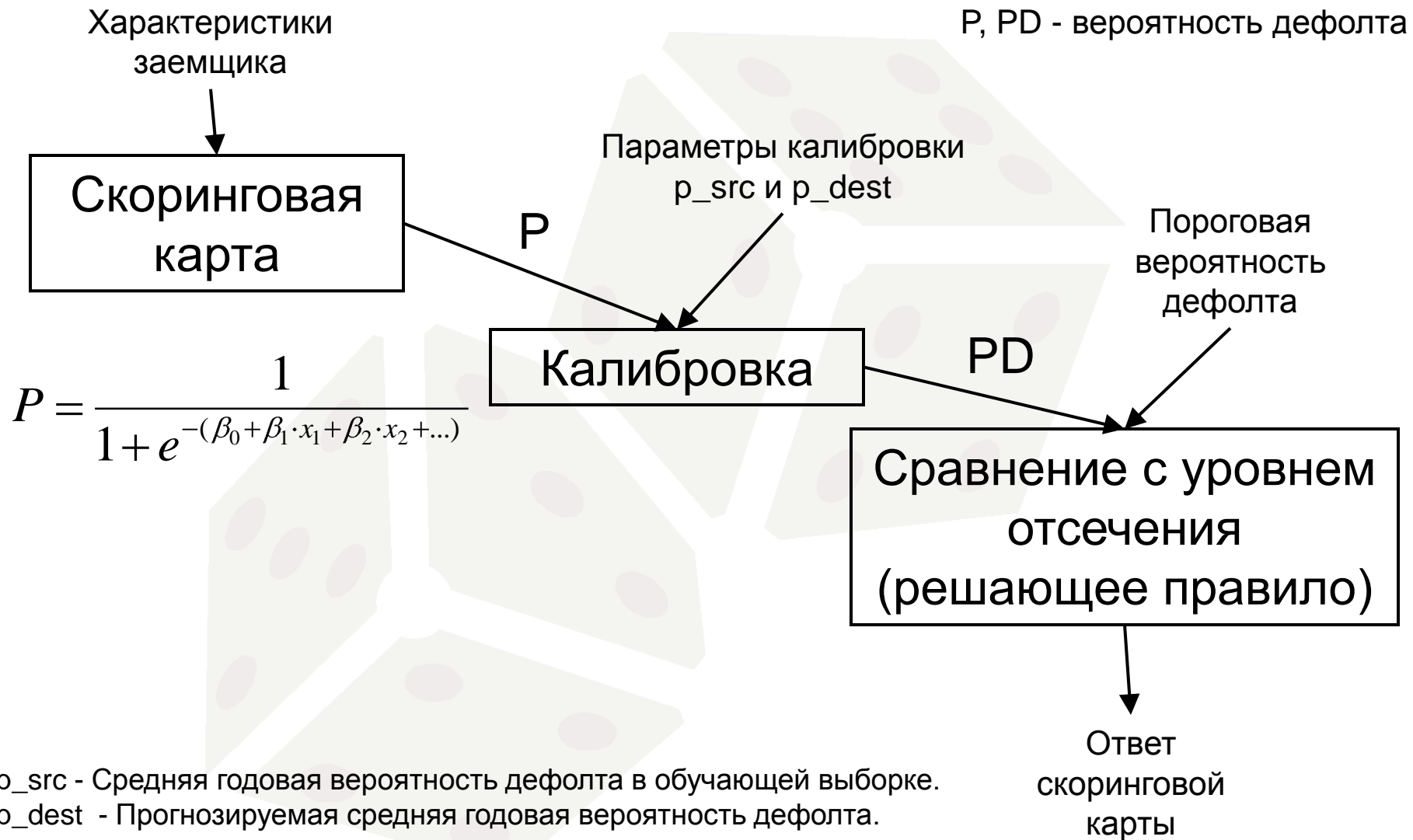


INTERNAL RATINGS BASED (IRB) APPROACH

	PD	LGD	EAD
Foundation approach	Оценка Банка	Оценка регулятора	Оценка регулятора
Advanced approach	Оценка Банка	Оценка Банка	Оценка Банка



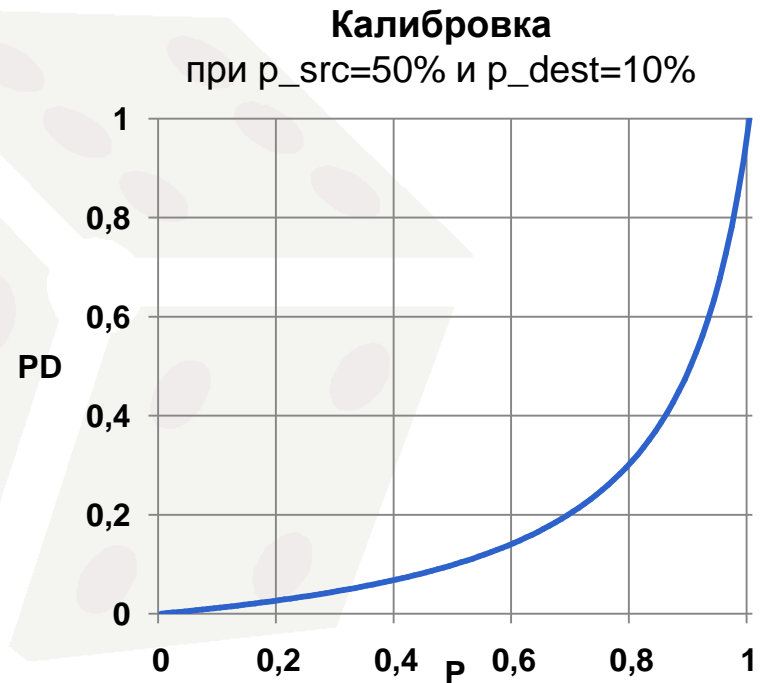
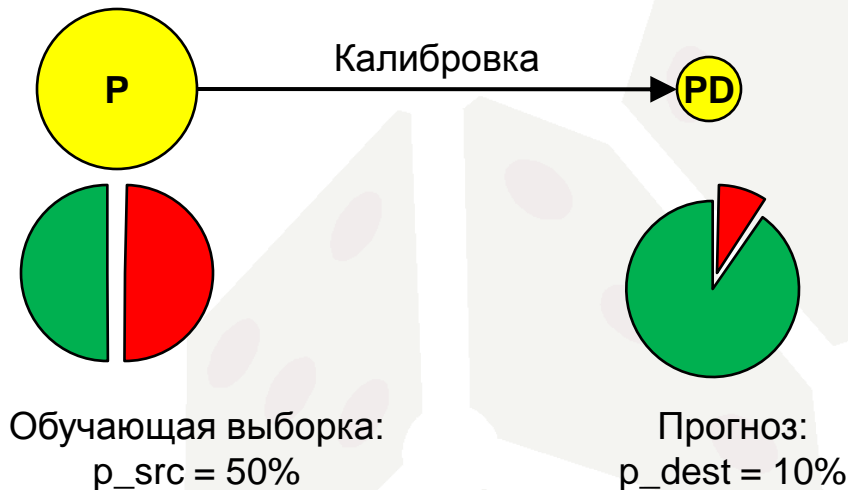
ПРОЦЕСС ПРОВЕРКИ ПО СКОРИНГУ





УРОВНИ ПРИ ПОСТРОЕНИИ МОДЕЛЕЙ

- Уровень 0. Данные
- Уровень 1. Модель
- Уровень 2. Калибровка



Модель только сортирует клиентов.
Калибровка необходима для корректной оценки прогнозируемого параметра (PD, LGD, EAD).



ПРОГНОЗИРОВАНИЕ ВЕРОЯТНОСТИ ДЕФОЛТА (PD), APPLICATION-СКОРИНГ

Задача классификации:

← признаки →

объекты	Заемщики	Возраст	Образование	...	Класс
	Заемщик 1	30	Высшее	...	Хороший
	Заемщик 2	47	Неск. Высших	...	Хороший
	Заемщик 3	25	Среднее	...	Хороший
	Заемщик 4	50	Уч. Степень	...	Хороший
	Заемщик 5	29	Высшее	...	Плохой
	Заемщик 6	37	Среднее	...	Хороший

Новый клиент	34	Высшее	...	???	

Например, плохой – тот, кто хотя бы раз допустил просрочку 90 дней (дефолт).



ЗАДАЧИ НА УРОВНЕ ДАННЫХ

- Определение выборки (sampling)
- Работа с недостающими значениями (missing values)
- Работа с выбросами (outliers)
- Группировка переменных (значений внутри признака)
- Оцифровка переменных (группа → woe)
- Сегментация выборки
- Определение целевой переменной



МОДЕЛИ И АЛГОРИТМЫ

Наиболее распространенные модели анкетного (application-) кредитного скоринга:

- Логистическая регрессия (~99,9% application models, Bart Baesens)
- Деревья решений, решающие списки
- Нейронные сети

This house believes... that Credit Scoring methods and applications are stuck in the 1980s

J Hinder, G Groom, D Hand & N Butler

(Дискуссия на CRC Conference 2009)

<http://www.crc.man.ed.ac.uk/conference/archive/2009.html>

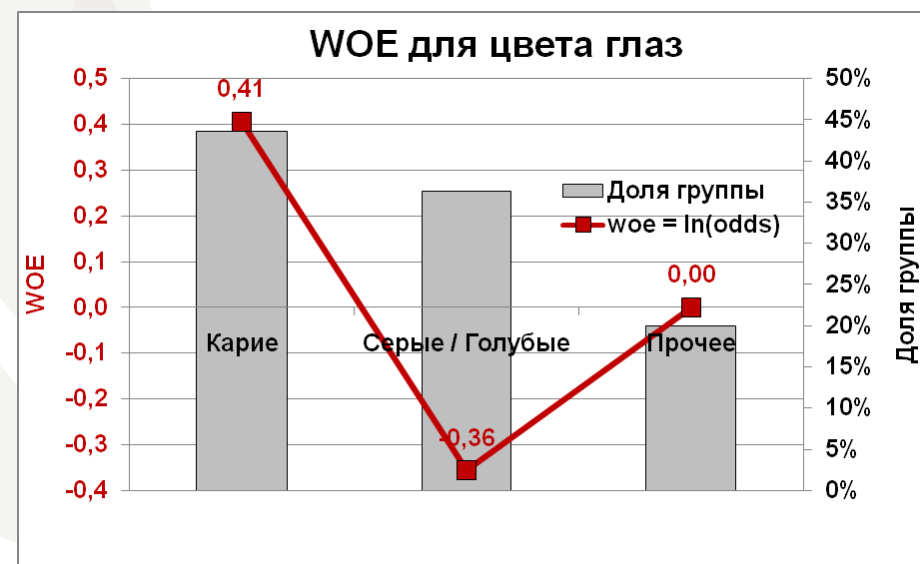
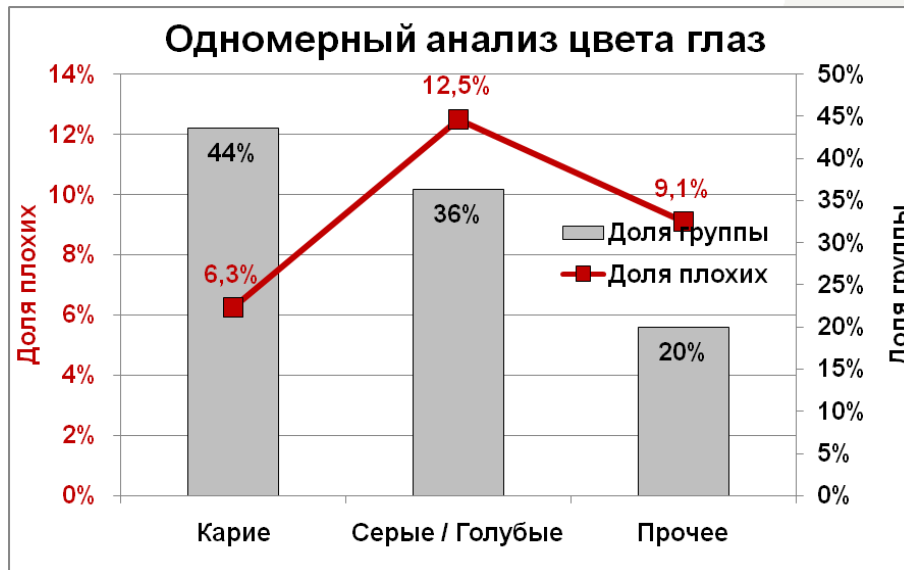


WEIGHT OF EVIDENCE

Weight of evidence $woe_{cat} = \ln(odds) = \ln\left(\frac{p_{cat}^{good}}{p_{cat}^{bad}}\right)$

p_{cat}^{good} - отношение количества хороших в категории cat к числу всех хороших

p_{cat}^{bad} - отношение количества плохих в категории cat к числу всех плохих



Цвет глаз заемщика	Количество плохих	Количество хороших	Количество в группе	Доля группы	Доля плохих	p good	p bad	odds	woe = ln(odds)
Карие	30	450	480	44%	6,3%	45,0%	30,0%	0,67	0,41
Серые / Голубые	50	350	400	36%	12,5%	35,0%	50,0%	1,43	-0,36
Прочее	20	200	220	20%	9,1%	20,0%	20,0%	1,00	0,00
Итого	100	1000	1100	100%	9,1%	100,0%	100,0%	1,00	



ЛОГИСТИЧЕСКАЯ РЕГРЕССИЯ И WOE

Information value (IV) – мера прогностической силы переменной, которая используется для:

- Оценки качества группировки переменной;
- Оценки информативности переменной.

$$IV = \sum_{cat} ((p_{cat}^{good} - p_{cat}^{bad}) \cdot woe_{cat}) = \sum_{cat} \left((p_{cat}^{good} - p_{cat}^{bad}) \cdot \ln \left(\frac{p_{cat}^{good}}{p_{cat}^{bad}} \right) \right)$$

Логистическая регрессия:

$$P = \frac{1}{1 + e^{-(\beta_0 + \beta_1 \cdot woe_{цвет_глаз} + \beta_2 \cdot woe_{цвет_волос} + \dots)}}$$



ДОГАДКА ОБ ОТВЕРГНУТЫХ ЗАЕМЩИКАХ (ПРОБЛЕМА REJECT INFERENCE)

Выборка для построения модели цензурирована: остались клиенты, прошедшие проверку действующей скоринговой картой и службами Банка



Reject Inference: Если бы мы не отказали в кредите, то клиент был бы хорошим или плохим?



ДИСПРОПОРЦИЯ КЛАССОВ

Пусть: 98% хороших и 2% плохих

Тогда: «Все заемщики – хорошие» - неплохая модель

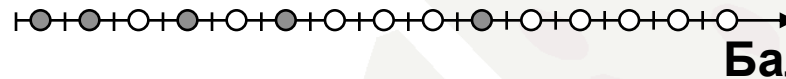
Решения:

- различная стоимость ошибок 1-го и 2-го рода
- увеличение веса («размножение») плохих объектов



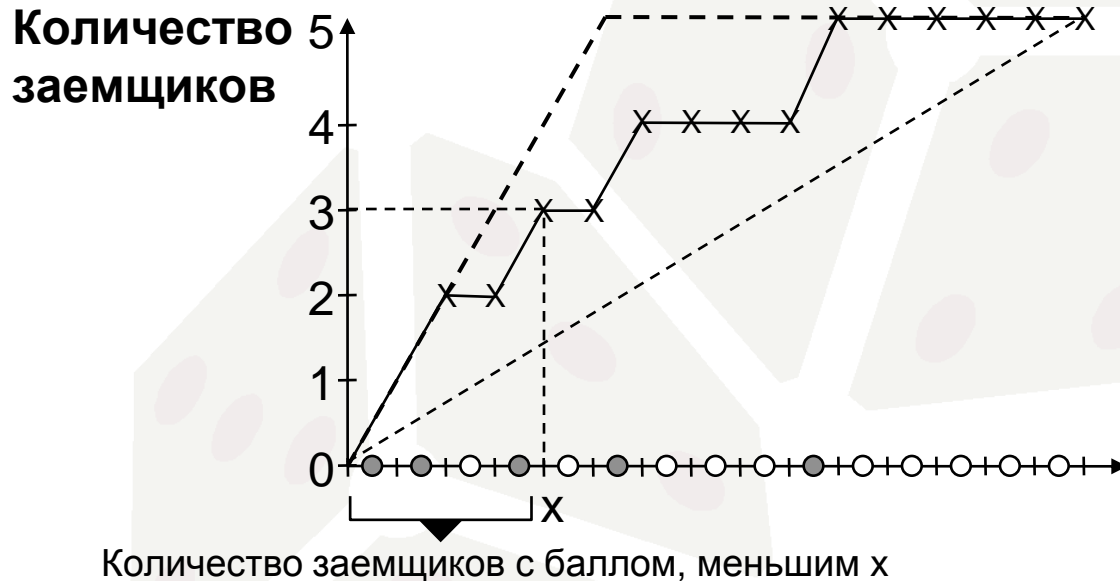
КАЧЕСТВО МОДЕЛЕЙ ПОСТРОЕНИЕ КРИВОЙ GINI

① Упорядочить заемщиков по баллу, отметить дефолты



- - - Идеальная модель
- Реальная модель
- Бесплезная модель

② Значение кривой Gini на заемщике x – это количество дефолтов с баллом, меньшим, чем у заемщика x

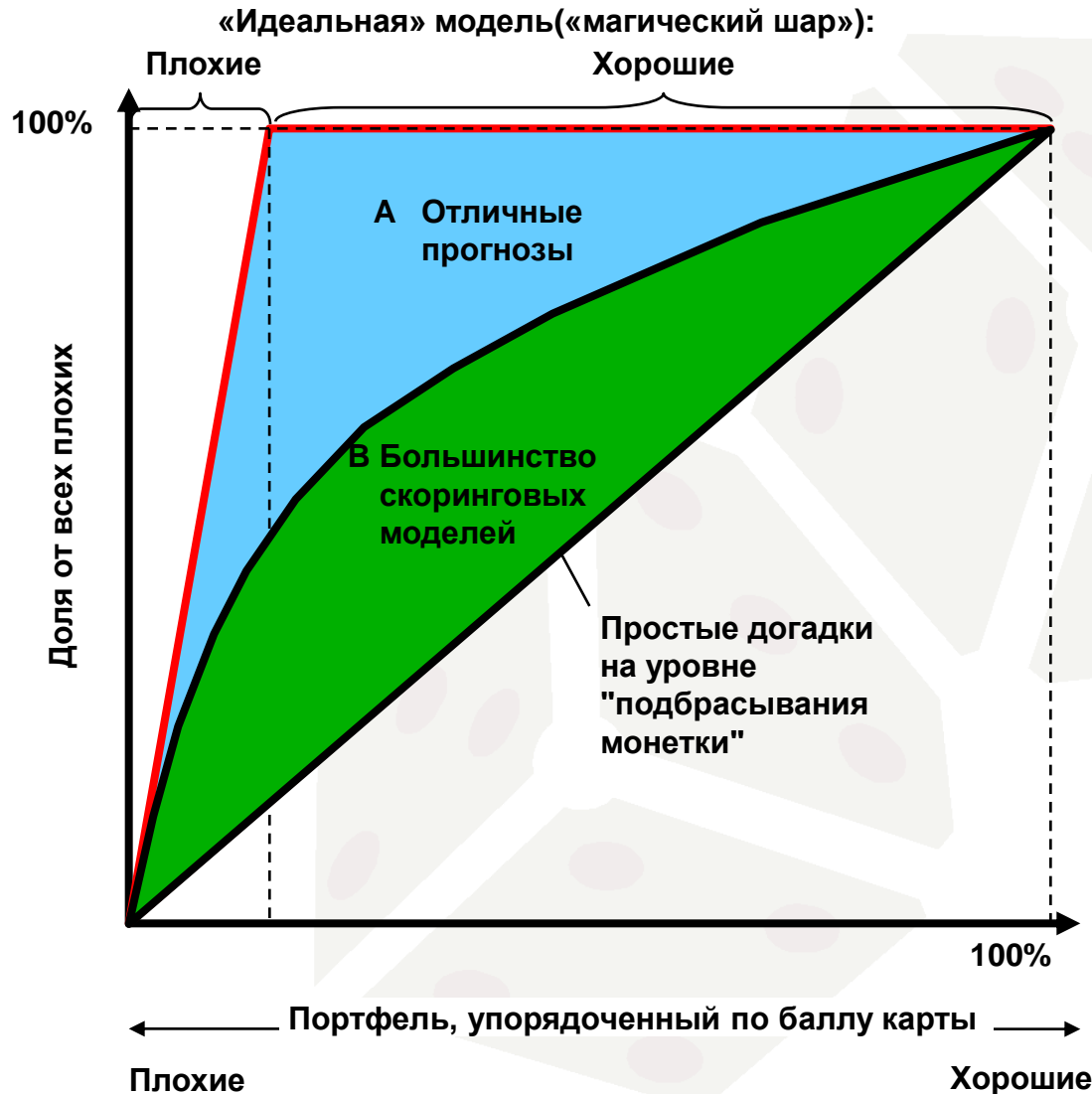


Экстремальные случаи:

- Получаем идеальную модель, если у всех дефолтов балл меньше, чем у худшего из добросовестных плательщиков
- Бесплезная модель соответствует равномерному распределению заемщиков по баллу



КАЧЕСТВО МОДЕЛЕЙ КОЭФФИЦИЕНТ GINI



- Чем больше значение В, сравниваемого с А, тем ближе карта подходит к идеалу отличного прогнозирования дефолта
- Коэффициент Gini = $\frac{B}{A + B}$

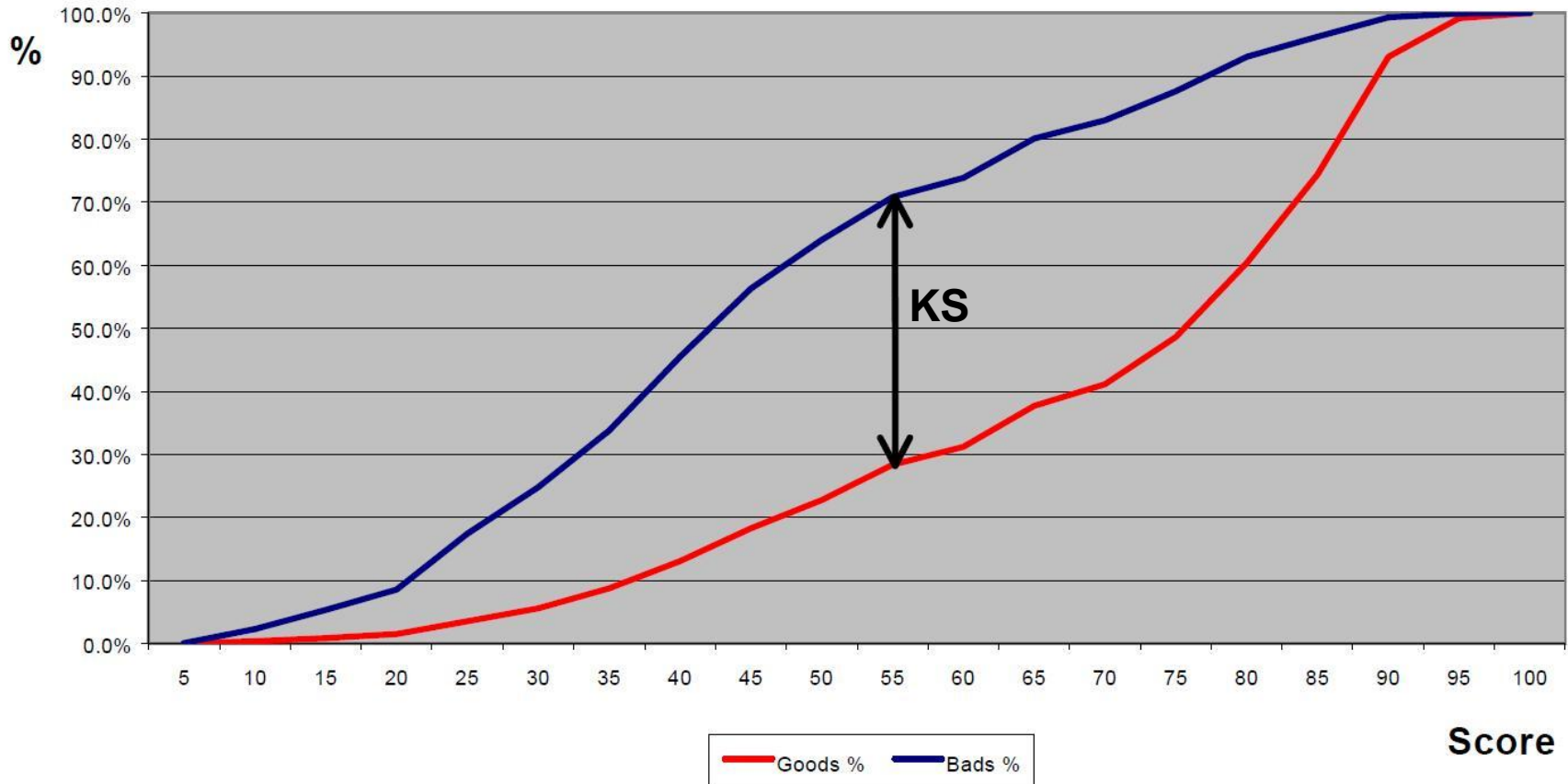
"The Group has found that the AR [=Gini coefficient] and the ROC measure appear to be more meaningful than the other [...] indices* because of their statistical properties."
Basel Committee on Banking Supervision, Working Paper No. 14



КАЧЕСТВО МОДЕЛЕЙ

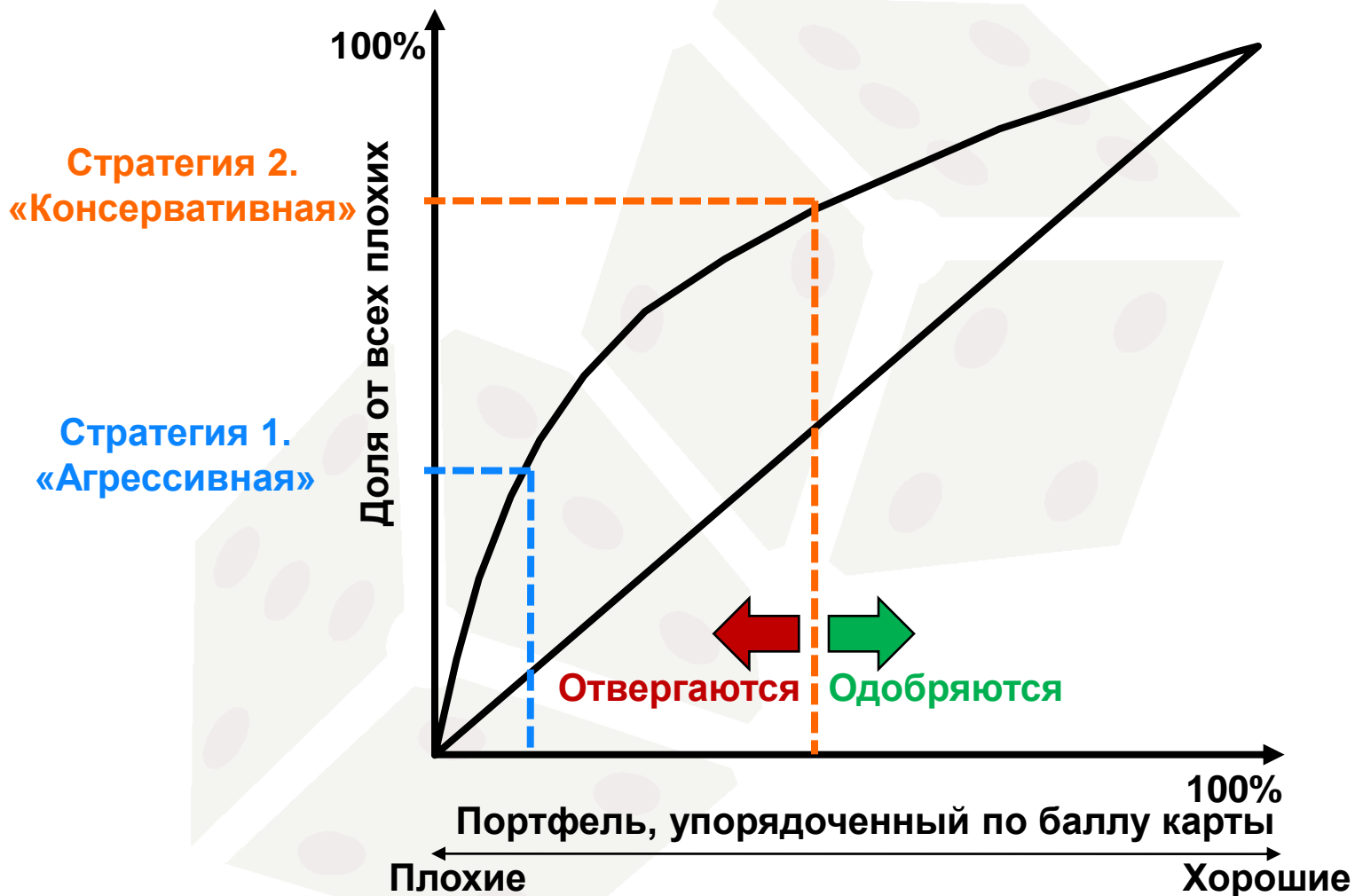
СТАТИСТИКА КОЛМОГороВА-СМИРНОВА

Статистика Колмогорова-Смирнова (KS) – максимальное «расстояние» между функциями распределения по баллу хороших и плохих





СТРАТЕГИИ ПРОДАЖ





ВЫБОР ОПТИМАЛЬНОГО УРОВНЯ ОТСЕЧЕНИЯ (CUTOFF)

Ставка по кредиту =

= Цена фондирования (стоимость привлечения средств) +

+ Операционные расходы (выдача кредита, проверки, сбор задолженности и т.п.) +

+ Маржа банка (какой доход получит банк?) +

+ Маржа риска (зависит от вероятности дефолта)

Пример решающего правила:

Если при данной марже риска (вероятности дефолта) маржа банка $> 0\%$, то выдать кредит.

Тогда **уровень отсечения** – вероятность дефолта, при которой маржа банка = 0%



АНТИМОШЕННИЧЕСКИЙ (ANTIFRAUD-) СКОРИНГ

В качестве целевой переменной (класс) могут использоваться:

- Данные о реальных мошенниках от служб Collection и экономической безопасности;
- Данные о мошенническом поведении (например, клиент ни разу не оплачивал кредит)

Антимошеннические карты обычно используются как сигнал в алгоритме проверки заемщика службами банка

Из выборок для application- карт мошенников обычно исключают, так как они «должны быть нацелены» на прогнозирование финансового дефолта.



ПОВЕДЕНЧЕСКИЙ (BEHAVIOUR-) СКОРИНГ

Примеры поведения:

- Средний / минимальный / максимальный остаток на счете;
- Платежное поведение (просрочки, скоринг БКИ);
- Изменение работы, домашнего адреса.

Поведенческий скоринг может использоваться для:

- Оценки резервов;
- Выявления счетов, которые могут превысить лимит кредитования;
- Управления лимитами кредитных карт;
- Оценки портфеля;
- Коллекторских стратегий;
- Проактивной реструктуризации задолженности;
- Сегментации клиентов и проведения маркетинговых кампаний.

Цель – отранжировать клиентов по вероятности дефолта (такая же как в application- скоринге)



КОЛЛЕКТОРСКИЙ (COLLECTION-) СКОРИНГ

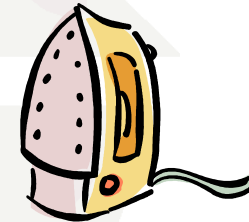
Ранняя
стадия
сбора



Средняя
стадия
сбора



Поздняя
стадия
сбора



Примеры моделей коллекторского скоринга:

- Self-cure (кто вернет долг без дополнительных усилий со стороны банка?)
- Сегментация должников (на кого направить наиболее опытных коллекторов?)
- Форсированный переход на более поздние стадии сбора
- Продажа внешним коллекторским агентствам (кого банку выгодно передать внешним коллекторам?)



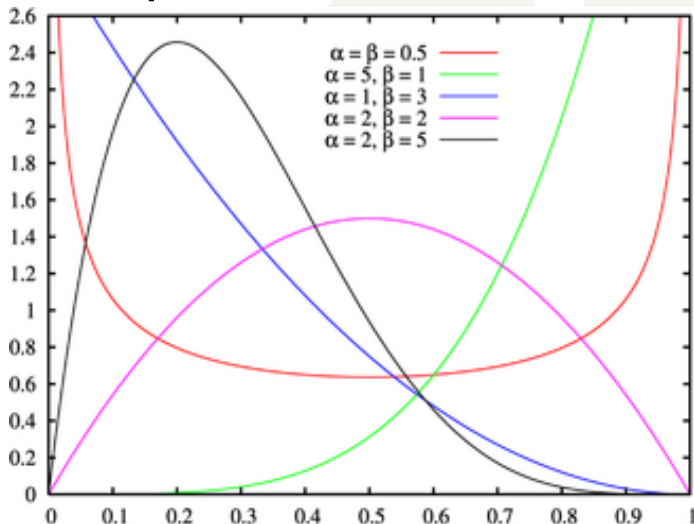
МОДЕЛИРОВАНИЕ LGD

LGD (Loss Given Default) – вероятность потерь в случае дефолта (измеряется в % от EAD; $LGD = 1 - \text{Recovery Rate}$)

Заемщики	Возраст	Образование	...	LGD
Заемщик 1	30	Высшее	...	2%
Заемщик 2	47	Неск. Высших	...	90%

Моделирование во многом аналогично PD, но есть отличия.
Например, непрерывная целевая переменная.

Распределение заемщиков по LGD U-образно, $0 \leq LGD \leq 1$



Наиболее распространенные модели:

- Дерево решений;
- Регрессия с бета-трансформацией;
- 2-стадийные модели.



МОДЕЛИРОВАНИЕ EAD

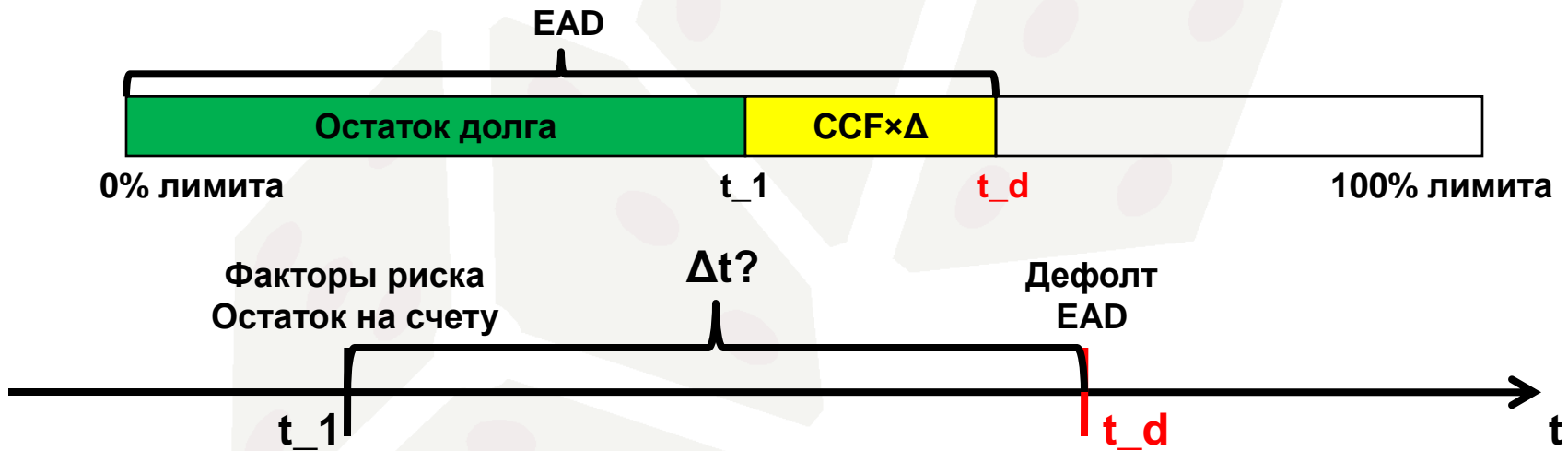
EAD (Exposure at Default) – стоимость, подверженная риску дефолта

Наиболее актуально для таких продуктов, как кредитные карты.

CCF (Credit Conversion Factor) – доля не использованного кредитного лимита

CCF \equiv LEQ (Loan Equivalency Factor) в США

EAD = Остаток долга + CCF (Лимит – Остаток долга)



Моделирование во многом аналогично LGD



КЛАСТЕРИЗАЦИЯ ПРИ РАЗРАБОТКЕ МОДЕЛЕЙ

Город 1: Работодатель ЗАО – это плохо.
Город 2: Работодатель ЗАО – это хорошо.
Кластеры объединяются по близости выборок.





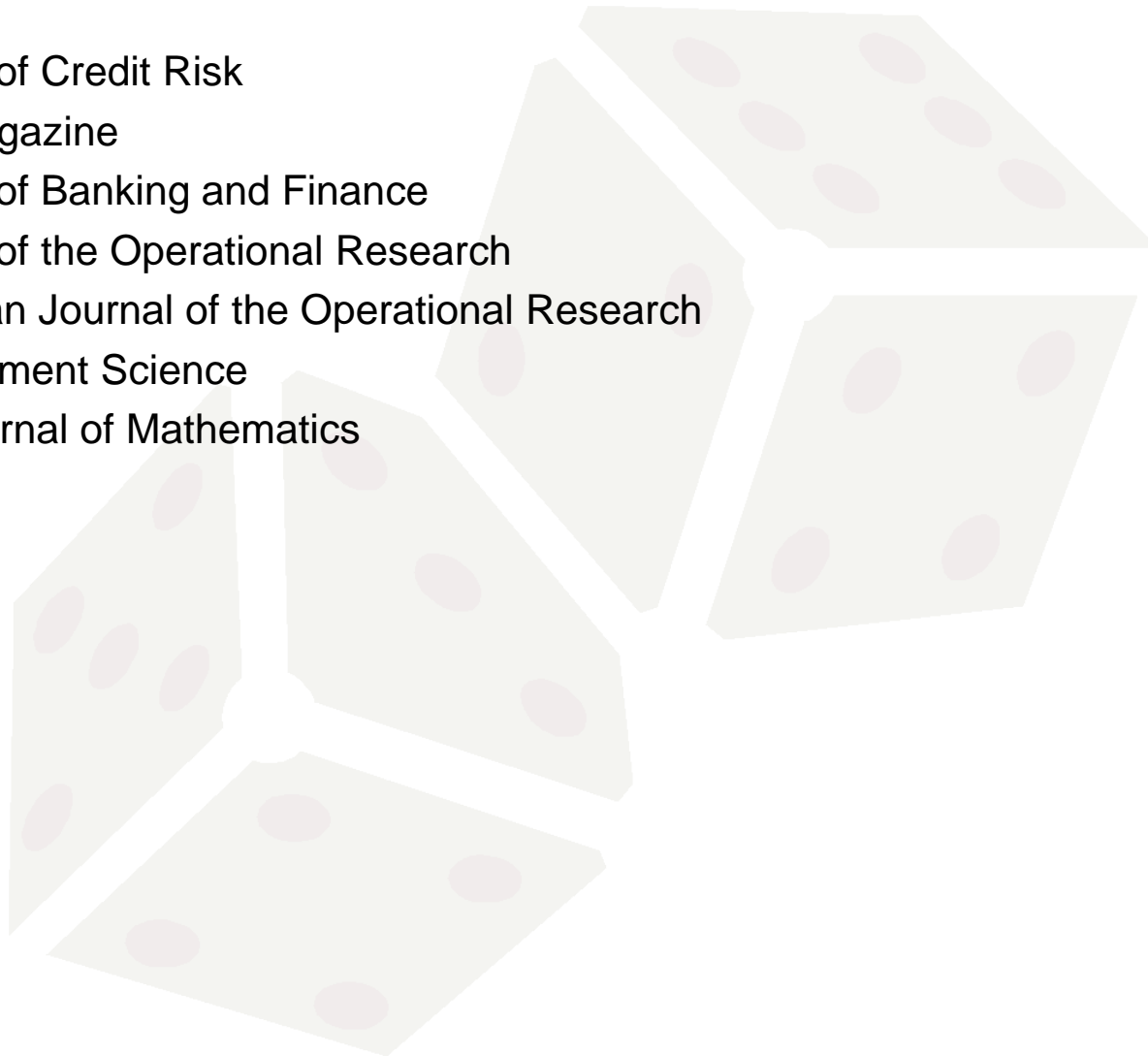
ССЫЛКИ КНИГИ

- Энциклопедия финансового риск-менеджмента / Под ред. к.э.н. А.А. Лобанова и А.В. Чугунова - 4-е изд., испр. и доп. - М.: Альпина Бизнес Букс, 2009.
- Credit Risk Management: Basic Concepts, Van Gestel and Baesens, Oxford University Press, 2008.
- Credit Scoring and Its Applications, Thomas, Edelman and Crook. SIAM Monographs on Mathematical Modeling and Computation, 2002.
- The Credit Scoring Toolkit, Anderson, Oxford University Press, 2007.
- Consumer Credit Models: Pricing, Profit and Portfolios, Thomas, Oxford University Press, 2009
- The Basel Handbook, Ong, 2004.
- Basel II Implementation: A Guide to Developing and Validating a Complicant Internal Risk Rating System, Ozdemir and Miu, McGraw-Hill, 2008.
- The Basel II Risk Parameters: Estimation, Validation and Stress Testing, Engelmann and Rauhmeier, Springer, 2006.
- Recovery Risk: The Next Challenge in Credit Risk Management. Edward Altman, Andrea Resti, Andrea Sironi (editors), 2005.
- Developing Intelligent Systems for Credit Scoring, Bart Baesens, Ph.D. thesis, K.U.Leuven, 2003



ССЫЛКИ ЖУРНАЛЫ

- Journal of Credit Risk
- Risk Magazine
- Journal of Banking and Finance
- Journal of the Operational Research
- European Journal of the Operational Research
- Management Science
- IMA Journal of Mathematics





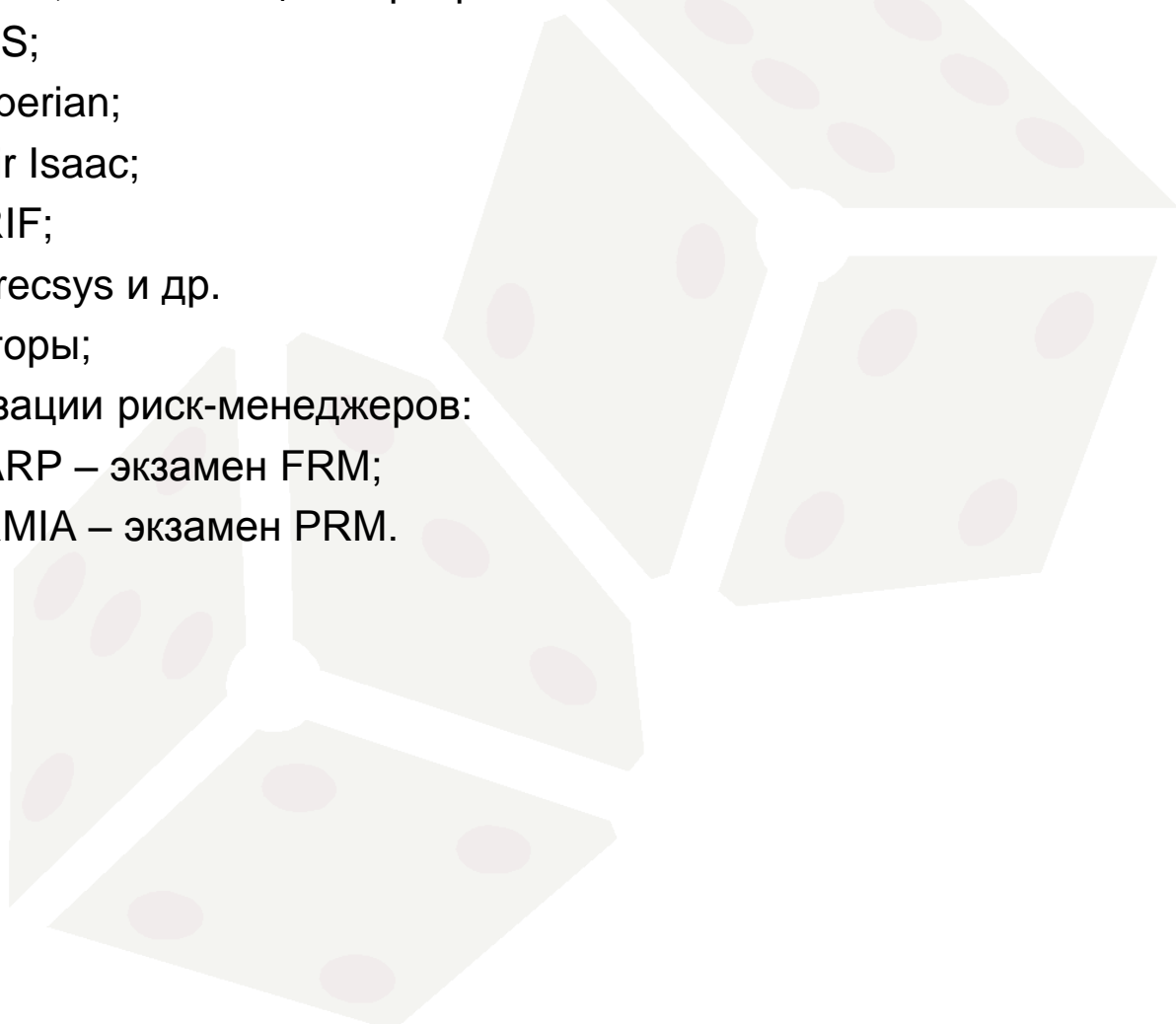
ССЫЛКИ ВЕБ-САЙТЫ

- www.riskofficer.ru
- www.defaultrisk.com
- www.bis.org
- Веб-сайты регуляторов в банковской сфере:
 - www.cbr.ru (ЦБ РФ)
 - www.fsa.gov.uk (United Kingdom)
 - www.hkma.gov.uk (Hong Kong)
 - www.apra.gov.au (Australia)
 - www.mas.gov.sg (Singapore)
- В США:
 - Четыре агентства: OCC, Federal Reserve, FDIC, OTS
 - Proposed Supervisory Guidance for Internal Ratings Based Systems for Credit Risk, Advanced Measurement Approaches for Operational Risk and the Supervisory Review Process (Pillar 2), Federal Register, Vol. 72, No. 39, February 2007.
 - Risk-based capital standards: Advanced Capital Adequacy Framework-Basel II; Final rule, Federal Register, December 2007



ССЫЛКИ WHITE PAPERS

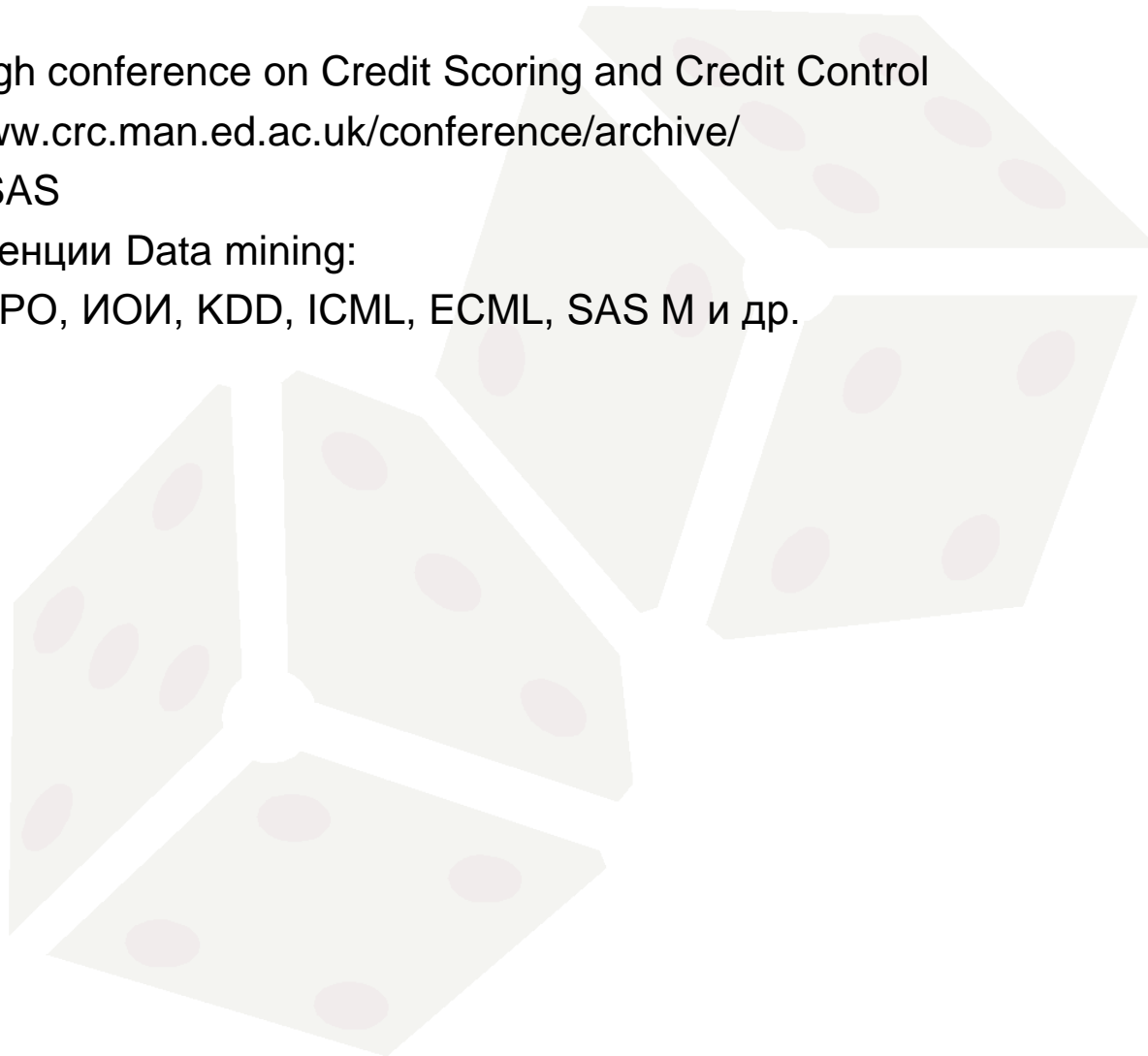
- Компании, занимающиеся разработкой ПО и консалтингом:
 - SAS;
 - Experian;
 - Fair Isaac;
 - CRIF;
 - Forecsys и др.
- Регуляторы;
- Организации риск-менеджеров:
 - GARP – экзамен FRM;
 - PRMIA – экзамен PRM.





ССЫЛКИ КОНФЕРЕНЦИИ И СЕМИНАРЫ

- Edinburgh conference on Credit Scoring and Credit Control
<http://www.crc.man.ed.ac.uk/conference/archive/>
- Курсы SAS
- Конференции Data mining:
 - ММО, ИОИ, KDD, ICML, ECML, SAS M и др.

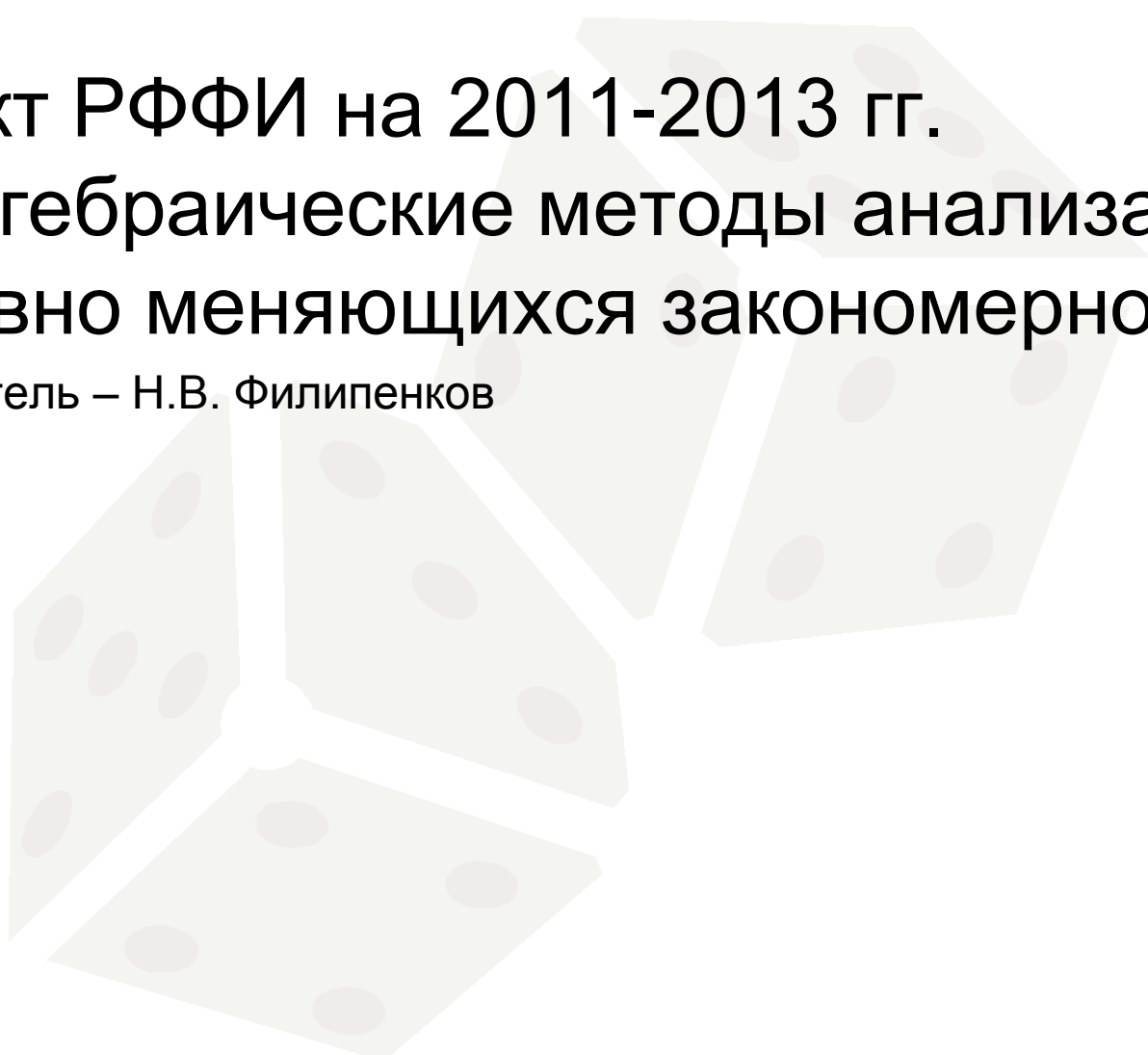




ПРОЕКТ РФФИ

Проект РФФИ на 2011-2013 гг. «Алгебраические методы анализа плавно меняющихся закономерностей»

Руководитель – Н.В. Филипенков





ПЛАВНО МЕНЯЮЩИЕСЯ СКОРИНГОВЫЕ МОДЕЛИ

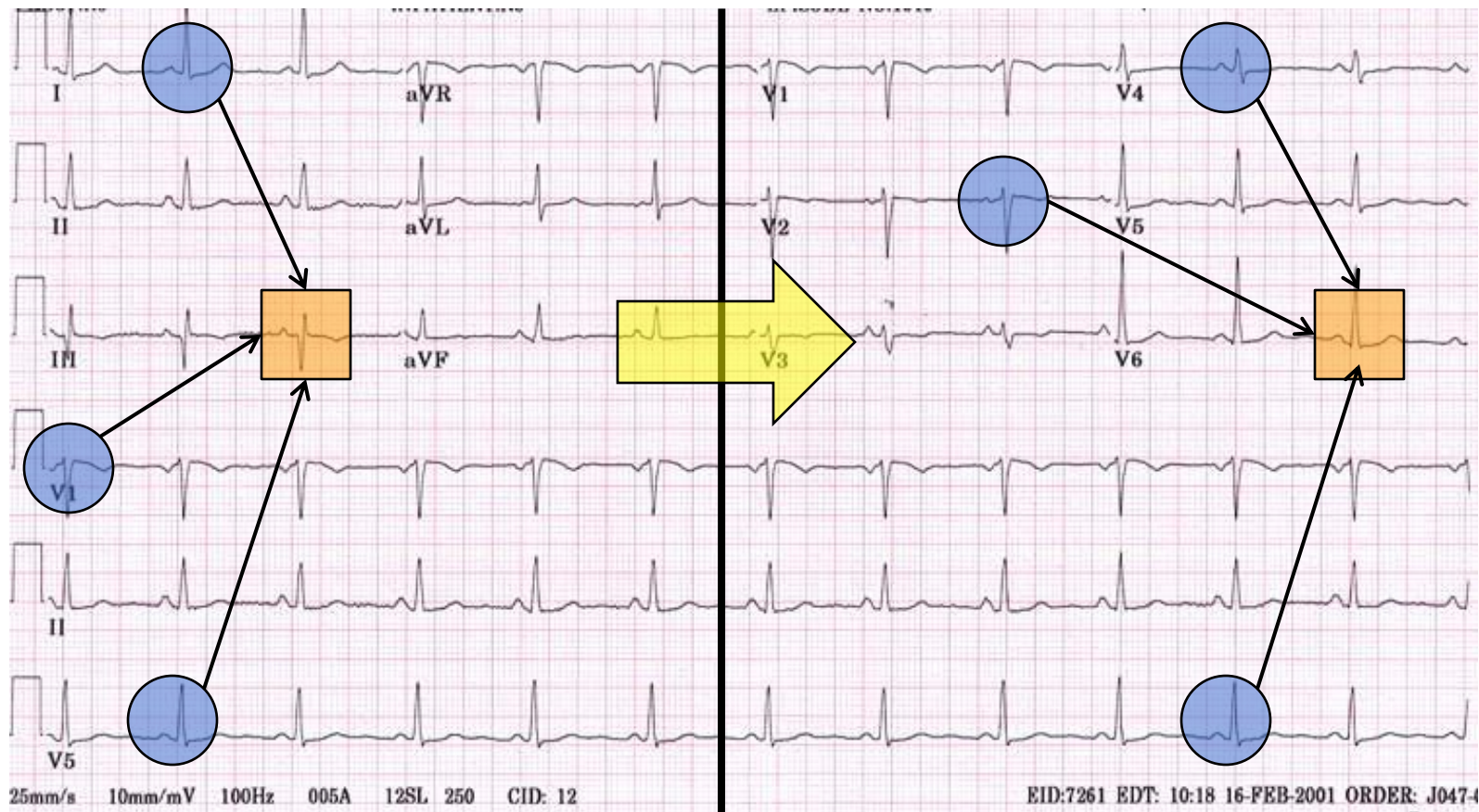
Заемщики	Возраст	Образование	...	Класс
Заемщик 1	30	Высшее	...	Хороший
Заемщик 2	47	Неск. Высших	...	Хороший
Заемщик 3	25	Среднее	...	Хороший
Заемщик 4	50	Уч. Степень	...	Хороший
Заемщик 5	29	Высшее	...	Плохой
Заемщик 6	37	Среднее	...	Хороший
...
Заемщик (n+1)	35	Высшее	...	Хороший
Заемщик (n+2)	49	Неск. Высших	...	Плохой
Заемщик (n+3)	22	Среднее	...	Хороший
Заемщик (n+4)	30	Уч. Степень	...	Хороший
Заемщик (n+5)	29	Высшее	...	Плохой
Заемщик (n+6)	27	Среднее	...	Хороший
...
Заемщик (m+1)	31	Высшее	...	Хороший
Заемщик (m+2)	41	Неск. Высших	...	Хороший
Заемщик (m+3)	26	Среднее	...	Хороший
Заемщик (m+4)	44	Уч. Степень	...	Хороший
Заемщик (m+5)	37	Высшее	...	Плохой
Заемщик (m+6)	33	Среднее	...	Хороший

t





ПЛАВНО МЕНЯЮЩИЕСЯ ЗАКОНОМЕРНОСТИ ВО ВРЕМЕННЫХ РЯДАХ



i -й сегмент временного ряда

$(i+1)$ -й сегмент временного ряда

Закономерность изменилась на близкую с течением времени



ЗАДАЧА: ПОИСК ПЛАВНО МЕНЯЮЩИХСЯ ЗАКОНОМЕРНОСТЕЙ

- 765 временных рядов курсов акций
- 4 537 – длина рядов
почасовой курс с 9:30 13.05.2010 по 15:30 10.12.2004
- 765 csv-файлов со следующей структурой:
YYYYMMDD, hh:mm, Open, High, Low, Close, Volume
- В отдельном файле указаны отрасли для организации рядов в пучки (многомерные ряды)



ПРИСОЕДИНЯЙТЕСЬ К ПРОЕКТУ!

- **Кто?** Студенты и аспиранты кафедры ММП и смежных специальностей (в том числе потенциальные)
- **Как?**
 - n.filipenkov@mail.ru
 - Вконтакте: Группа «Плавно меняющиеся закономерности»
 - Facebook: Data Mining with Nick
- **Материалы по теме:**
 - *Филипенков Н.В.* Об одном методе поиска плавно меняющихся закономерностей в пучках временных рядов // Журнал вычислительной математики и математической физики - 2009. - Т. 49, № 11. С. 2020-2040.
 - *Filipenkov N. V.* An Algorithm for Analysis of Multidimensional Time Series with Smoothly Varying Regularities and Its Application // Pattern Recognition and Image Analysis - 2010. – Vol. 20, № 3 – pp. 251-268.
 - *Филипенков Н.В.* Поиск плавно меняющихся ассоциативных правил // 50-я конф. МФТИ: Тез. докл. - М.: МФТИ, 2007. – С. 117-119.
 - *Филипенков Н.В.* Об эволюционирующих алгоритмах классификации и прогнозирования // ИОИ-2008: Тез. докл. - Симферополь, 2008. – С. 228-230.