


# Deductor Credit Scorecard Modeler 2.2: возможности решения



## Назначение

Решение Deductor Credit Scorecard Modeler (DCSM) предназначено для построения, анализа и мониторинга скоринговых карт портфелей однородных ссуд с целью предупреждения и снижения рисков, связанных с оценкой кредитоспособности заемщика.

Основная сфера применения решения – в банковских организациях.

## Решаемые задачи

- Определение зрелости счетов кредитного портфеля.
- Построение скоринговой карты.
- Сравнение и контроль качества скоринговых карт.
- Контроль актуальности внедренной скоринговой карты.

## Состав решения

Решение из состоит из:

- Витрины данных (ВД) – специализированная предметно-ориентированная база, предназначенная для хранения данных, необходимых для работы модулей. В решение включены скрипты для создания структуры ВД для соответствующей СУБД (FireBird, MS SQL Server, Oracle), пустая ВД на СУБД FireBird, ВД на СУБД FireBird с демонстрационными данными.
- Модулей – группы сценариев решения, реализующий основной функционал. Каждый сценарий запускается средствами аналитической платформы Deductor.
- Документации – руководство по работе и требования к данным.

## Этапы разработки скоринговой карты

Решение состоит из последовательно запускаемых модулей. В состав модулей может входить несколько методов реализации расчетов, что позволяет настроить наиболее подходящий вариант работы решения. На рисунке ниже показан реализованный в DCSM цикл разработки скоринговой карты (пунктиром обозначены необязательные блоки).



*Цикл разработки скоринговой карты*

## Функциональные возможности модулей

Сценарии модулей реализуют не только минимально необходимый, но и расширенный функционал построения, анализа, мониторинга скоринговых карт, загрузки всех необходимых данных.

Значения ключевых параметров вынесены в переменные и могут быть легко изменены аналитиком на отличные от выставленных по умолчанию.

Средствами Deductor Studio любой сценарий решения DCSM может быть изменен аналитиком.

Цикл разработки скоринговой карты разделен на логические блоки, реализованные модулями.

### **Модуль загрузки данных в хранилище**

Влияние данных и их качества на точность получаемой скоринговой карты трудно переоценить. Поэтому в фундаментальную основу DCSM заложено специализированная предметно-ориентированная витрина данных (ВД). ETL-сценарии этого модуля отвечают за сбор, обработку и загрузку данных в ВД из разнообразных источников, включая НБКИ, операционные и заявочные системы банка.

Целью данного модуля является сокращение времени и объема работ, связанных со сбором, трансформацией и загрузкой необходимых данных в ВД DCSM.

Модуль содержит примеры реализации процесса ETL и предоставляет возможности:

- Проведения аудита входных данных на отсутствие ошибок и противоречий.
- Загрузки широкого круга всесторонней информации о счетах и заемщике, необходимой для построения скоринговой карты с широким профилем риска, в том числе:
  - Любая информация о финансовом положении заемщика (доход, индекс Платеж/Доход, данные из БКИ и т.д.), которой располагал кредитор на момент принятия решения о выдаче кредита.
  - Социально-демографические данные о заемщике.
  - Параметры кредитования.
  - Данные о состоянии счета (закрытый, открытый, списанный долг, досрочно погашенный, проданный третьей стороне).
  - Данные о просрочках (по периодам в зависимости от длительности просрочки).
- Предобработки сырых данных перед загрузкой: фильтрация записей с ошибками, объединение редких значений.

### **Модуль анализа жизненных циклов счетов**

Загруженные в ВД данные нужно подготовить к процессу построения и тестирования скоринговых карт. Самый первый этап – это формирование репрезентативных рабочих и тестовой выборки. В этих выборках счета должны принадлежать к определенному классу: хороший или плохой. Определение этого класса является нетривиальной решаемой задачей на данном этапе и требует проведение анализа имеющихся данных специализированными методами.

Целью данного модуля является получение репрезентативных подвыборок данных (сэмплов) для дальнейшего построения и тестирования скоринговых карт.

Сценарии данного модуля позволяют:

- Определить «показательный период» (performance window) с помощью винтажного анализа (по каждому критерию определения наступления дефолта).
- Определить корректность выбора критериев определения дефолта счета методом миграции просрочек.
- Определить статус для каждого счета (плохой или хороший) согласно выбранному критерию наступления дефолта.
- Провести сэмплирование выборок для дальнейшего построения и проверки скоринговой карты, учитывая счета:
  - Не имеющие ни одного платежа (hard-bads).
  - С редким набором атрибутов характеристик.
- Выявить ошибки в данных, включая:
  - Проверки правильности ведения счета.
  - Проверки заполнения просрочки по методам её начисления LIFO/FIFO.

### **Модуль двумерного анализа**

Данный этап подготовки данных связан с их трансформацией. Основная цель этого этапа – выявить или получить важные для построения логистической регрессии характеристики и атрибуты, а также удалить всё лишнее, чтобы модель получалась не только достаточно точной, но и интерпретируемой.

Целью данного модуля является получение оптимизированной для обучения модели структуры данных.

Сценарии данного модуля позволяют:

- Осуществить автоматическое квантование значений характеристик с возможностью интерактивного подбора границ интервалов.
- Снизить число уникальных значений внутри атрибутов методом оптимального квантования (Fine&Coarse Classing).
- Рассчитать предсказательную силу характеристик и атрибутов.
- Провести общий анализ индексов IV (Information Value) и WoE (Weight of Evidence).

### **Модуль моделирования и оценки качества**

В основе скоринговой карты DCSM лежит модель логистической регрессии. Главные решаемые задачи данного модуля связаны с обучением модели и оценкой её качества.

Целью данного модуля является построение модели, пригодной для использования скоринговой картой.

Сценарии данного модуля позволяют:

- Построить модель на базе логистической регрессии с поправкой на априорные вероятности.
- Получить балльную скоринговую карту в стандарте FICO.
- Провести оценки и сравнение скоринговых карт, включая:
  - Расчёт индекса Gini.
  - Расчёт статистики Колмогорова-Смирнова для анализа получившихся скоринговых карт.
  - Построение CAP-кривых.

### **Модуль оценки невыданных кредитов**

Выборка одобренных кредитных заявок является смещенной относительно всей генеральной совокупности заявок и не учитывает поведения потенциальных заемщиков по отклоненным заявкам.

Целью данного модуля является полученные скоринговой карты с профилем рисков, максимально отражающим кредитоспособность потенциальных заемщиков на всем множестве имеющихся заявок.

Сценарии данного модуля позволяют произвести перекалибровку скоринговой карты с учетом отклоненных кредитных заявок одним из методов Reject Inference:

- Hard Cut Off.
- Random Parceling
- Fuzzy Augmentation.

### **Модуль мониторинга**

Под действием множества внешних и внутренних экономических факторов выборка поступающих кредитных заявок подвергается постоянным изменениям. Как следствие, риски всё чаще реализуются в тех условиях, которые имеющиеся скоринговые карты «считают» безопасными для выдачи кредита.

Целью данного модуля является предупреждение и минимизация риска использования внедренной и ставшей со временем неактуальной скоринговой карты.

В сценариях данного модуля реализованы тесты оценки актуальности скоринговой карты, тесты на стабильность и сдвиги в популяции, включая расчет индекса стабильности популяции (Population Stability Index, PSI).

### **Модуль расчет порогового балла**

Принятие решение о кредитоспособности заемщика основано на сравнении полученного скорингового балла с некоторым пороговым значением.

Целью данного модуля является получение оптимального порогового балла, разделяющего плохие заявки от хороших, для новой скоринговой карты.

Сценарии данного модуля позволяют:

- Рассчитать уровень отказов, исходя из статистики в портфелях.
- Определить пороговый балл с предварительным осуществлением процедуры ретроскоринга по одной из следующих стратегий:
  - С сохранением доли отказов на прежнем уровне.
  - С сохранением доли плохих на прежнем уровне.
  - С сохранением требуемого соотношения Хороший/Плохой.
- Произвести сравнение вышеуказанных стратегий.

### **Отчеты модулей**

В модуле анализа жизненных циклов счетов доступны следующие отчеты:

- Винтажная матрица (по каждому критерию определения дефолта).
- Зависимость доли дефолтных счетов в поколении от периода (по каждому критерию определения дефолта).
- Зависимость доли дефолтных счетов в периоде от поколения (по каждому критерию определения дефолта).
- Кросс-таблица с матрицей миграции.
- Распределение целевой переменной по выборкам после сэмплинга (целевая переменная – флаг принадлежности счета к классу хороших).
- Кросс-таблица и диаграмма, показывающие долю хороших счетов из кредитного портфеля, попавших в выборку для разработки скоринговой карты.
- Кросс-таблица и диаграмма, показывающие итоговое распределение статусов счетов после сэмплинга.
- Список найденных зрелых счетов hard-bads.
- Список найденных незрелых счетов hard-bads.
- Результаты сэмплинга счетов с редким в имеющей выборке сочетанием атрибутов (распределение таких счетов по подвыборкам).
- Сводный отчет по ошибкам в данных, выявленных в результате аудита.
- Количество счетов с ошибками и без ошибок.
- Распределение количества уникальных счетов с ошибками по каждому периоду и портфелю.
- Детализированные отчеты по ошибкам в счетах.
- Правильность заполнения просрочек по FIFO.

В модуле двумерного анализа доступны следующие отчеты:

- Статистика по результатам оптимального квантования непрерывных и дискретных характеристик.
- Индексы WoE (кросс-таблица и диаграмма с детализацией по атрибутам).
- Информационные индексы для каждой характеристики.

В модуле моделирования и оценки качества доступны следующие отчеты:

- Анализ регрессионной модели (статистические показатели для лог-регрессии и её коэффициентов).
- Таблица скоринговой карты (OLAP-куб с распределением количества хороших, плохих счетов, суммарного балла, уровня дефолта по атрибутам и характеристикам).
- Индекс Gini.
- Статистика Колмогорова-Смирнова.
- Распределения скоринговых баллов плохих и хороших счетов.
- Рисковые сегменты (гистограмма с двумя распределениями шансов: для плохих и для хороших счетов).
- Рейтинги счетов (таблица с оцененными скоринговым баллом и вероятностью дефолта для каждого счета).
- Анализ рассогласованности (OLAP-куб, в котором оценивается изменение степени влияния атрибутов характеристики на целевую переменную в построенной модели в сравнении с оценкой их влияния на целевую переменную по индексу WoE).
- AR-BR кривая.
- CAP-кривые.
- ROC-кривые.

В модуле оценки невыданных кредитов предусмотрены следующие отчеты:

- Итоговые рейтинги для отказов (таблица с оцененными скоринговым баллом и вероятностью дефолта для каждого счета из отказов).
- Интервальная статистика (таблица распределения количества и долей хороших и плохих счетов, количества отказов и необходимого количества плохих счетов по интервалам скорингового балла).
- Конечная выборка с отказами (таблица счетов с характеристиками и атрибутами и дополнительной информацией).

В модуле «Мониторинг» доступны следующие отчеты:

- Анализ характеристик на сдвиги в атрибутах.
- Диаграмма среднего балла в популяции.
- Диаграмма индекса стабильности популяции (Population Stability Index, PSI).
- Детальный отчет по PSI.
- Диаграмма динамики PSI (за месяц, квартал, год).
- Итоговые скоринговые отчеты за
  - Актуальный период (заданный вручную аналитиком).
  - Последний месяц.
  - Последний квартал.

В модуле «расчет порогового балла» предусмотрены следующие отчеты:

- Таблица с уровнями (процентом) отказов по портфелям.
- Диаграмма AR-BR.
- Изменения статусов счетов по различным совокупностям: KGB, AGB, совокупность счетов с выделением отдельной категории swap-in.

## Программные требования

Для работы с DCSM необходимо:

- Установленное ПО Deductor Studio Professional/Enterprise версии 5.3.
- Витрина данных на базе одной из трех СУБД: Firebird, MS SQL Server, Oracle и настроенное подключение к нему.
- Сценарии наполнения витрины данных (ВД). Конфигурация комплектуется шаблонами загрузки данных в ВД из DDF-файлов Deductor Studio (прим.: DDF – это специальный формат хранения данных, используемый аналитической платформой Deductor).
- Наличие специальной опции в лицензионном ключе Deductor, позволяющей открывать сценарии конфигурации.

Для ознакомления с демо-стендом веб-сервиса – Deductor Analytic Server и Deductor Integration Server.