

Перевод книги осуществлен
Исследовательским Центром

Gewissta

Наим Сиддики

**Скоринговые Карты
Для Оценки Кредитного
Риска**

Разработка и Внедрение
Интеллектуальных Методов
Кредитного Скоринга

СОДЕРЖАНИЕ

ГЛАВА 3. Процесс Разработки Скоринговой Карты, Этап 1: Подготовка Данных и Планирование	2
ГЛАВА 4. Процесс Разработки Скоринговой Карты, Этап 2: Обзор Данных и Параметров Проекта	7
ГЛАВА 5. Процесс Разработки Скоринговой Карты, Этап 3: Создание Базы Данных для Разработки Модели	29
ГЛАВА 6. Процесс Разработки Скоринговой Карты, Этап 4: Разработка Скоринговой Карты	39
ГЛАВА 7. Процесс Разработки Скоринговой Карты, Этап 5: Управленческие Отчеты по Скоринговой Карте	83
ГЛАВА 8. Процесс Разработки Скоринговой Карты, Этап 6: Внедрение Скоринговой Карты	86
ГЛАВА 9. Процесс Разработки Скоринговой Карты, Этап 7: Мероприятия после Внедрения	105

Процесс Разработки Скоринговой Карты, Этап 1: Подготовка Данных и Планирование

Существует ошибочное мнение, что разработка скоринговой карты начинается со сбора данных. Это неверно. Интеллектуальная скоринговая карта требует правильного планирования, которое проводится перед началом анализа. Данный процесс включает обоснование причины и цели проекта, определение главных лиц, участвующих в разработке и внедрении скоринговой карты, и распределение задач между этими лицами. При этом каждый сотрудник, принимающий участие в разработке скоринговой карты, отвечает за свое направление работы.

Подготовка бизнес-плана

Определение организационных целей и роли скоринговой карты

Первый этап разработки скоринговой карты – это определение организационных целей и их расстановка по приоритетности. Он позволяет сфокусировать внимание на той или иной задаче и помогает выявить приоритеты по спорным вопросам, возникающим в ходе разработки скоринговой карты (например, что является более приоритетным – повышение доходов или снижение убытков). Данный этап гарантирует проекту практическую осуществимость и стабильность, устраняя «сюрпризы», когда, например, банк неожиданно принимает решение об уменьшении балла отсечения с целью выдачи большего количества кредитов или об увеличении того же балла с целью сокращения потерь.

Примеры организационных целей:

- Сокращение безнадежных долгов/случаев банкротства/выплат по претензиям/случаев мошенничества
- Увеличение количества выдаваемых кредитов или увеличение доли рынка в сегменте обеспеченных кредитов, когда низкий уровень задолженности указывает на возможность расширения
- Увеличение прибыльности
- Увеличение операционной эффективности (например, улучшить процесс обработки и оценки кредитных заявок)
- Уменьшение расходов или увеличение скорости оборота капитала за счет автоматизации оценки с использованием скоринговых карт
- Улучшение предсказательной способности модели (в сравнении с пользовательскими моделями или моделью кредитного бюро)

Организационные цели влияют на способ валидации и выбор «наилучшей» скоринговой карты в тех случаях, когда разработано несколько скоринговых карт. Обычно для большинства организаций характерно свое сочетание целей и задач, включая вышеупомянутый список.

Дополнительный вопрос, который нужно прояснить на данном этапе – какая роль отводится скоринговой карте в процессе принятия решений? Она будет единственным инструментом оценки или станет инструментом поддержки принятия решения? Особенно это важно для организаций, которые ранее не использовали скоринг, поскольку внедрение скоринговых карт, возможно, повлияет на организационную культуру и последовательность операций. Скоринговую карту можно использовать в процессе принятия решений, варьируя ее роль в зависимости от продукта, организационной культуры, структуры организации, нормативных требований. Например, компания-эмитент кредитных карт может использовать скоринговые карты риска в качестве главного инструмента оценки, за исключением небольшого числа заявителей, например, 3% от общего числа (на основе скоринговых баллов или правил кредитной политики), направленных кредитному аналитику на дополнительное рассмотрение. В данном случае большинство решений будет принято автоматически и только при помощи скоринговой карты, без вмешательства человека, за исключением ключевых данных.

С другой стороны, ипотечный банк или страховая компания могут использовать скоринг в качестве одного из способов оценить кредитоспособность заявителей (или риск выплат по претензиям) наряду с подробным рассмотрением вручную и субъективными соображениями.

Осмысление данных вопросов поможет в разработке эффективных скоринговых карт и подходящих стратегий. В данном ключе скоринговые карты рассматриваются как неотъемлемая часть непрерывного и логически последовательного процесса принятия решений внутри организации. Желательно, чтобы характеристики скоринговой карты не дублировались. Например, если скоринговая карта является единственным инструментом решения, то она должна основываться на максимально возможном количестве независимых переменных (в отличие от скоринговой карты, построенной по нескольким характеристикам, с ограниченным количеством типов информации). Образно говоря, скоринговая карта должна имитировать хорошего, опытного специалиста, который анализирует кредитную заявку (т.е. проводит полный анализ кредитоспособности заявителя). Поэтому скоринговая карта должна содержать максимально полный профиль риска.

Если скоринговая карта является инструментом поддержки решения, то ее характеристики должны дополнять прочие критерии, использованные для оценки риска. Например, если используются правила кредитной политики, то из скоринговой карты желательно исключить характеристики, на основе которых были сформулированы данные правила. Это условие должно обеспечить логическую последовательность и эффективное использование имеющейся информации. Кроме того, если происходит пересмотр решения, определение объема пересмотренных заявок и характеристик, использованных для пересмотра, поможет выявить смещенные данные.

Сравнение преимуществ и недостатков собственной и внешней разработки, определение типа скоринговой карты

Бизнес-план определяет, какие скоринговые карты будут использоваться в проекте, внутренние (построенные в процессе проекта) или внешние (готовые), и обосновывает этот выбор. Готовые карты используются не только тогда, когда в

банке нет достаточной кредитной истории. Подобное решение может быть также вызвано отсутствием определенных ресурсов, необходимых для проекта, или отсутствием опыта построения скоринговой карты для специфического продукта, сжатыми сроками по времени или стоимостью внутренней разработки.

В тех случаях, когда данных недостаточно, когда использование данных затруднительно или качество данных является сомнительным, могут потребоваться дженериковые карты, разработанные внешними поставщиками или кредитными бюро. Кроме того дженериковые карты необходимы, когда организация проводит маркетинговую кампанию, нацеленную на новый сегмент, канал продаж или предлагает новый продукт, по которому не накоплено исторических данных, но существуют статистические данные по отрасли; когда объем оборотов по продукту не покрывает стоимости разработки пользовательской скоринговой карты; или когда график запуска продукта не предусматривает времени, достаточного для разработки пользовательской скоринговой карты.

В некоторых случаях статистически невозможно разработать пользовательские или дженериковые скоринговые карты. Обычно это обусловлено очень низкими объемами продаж, которые не оправдывают затраты, связанные с любой скоринговой картой. Встречаются также продукты, для которых не существует дженериковой модели или она неприемлема. В данных обстоятельствах необходимо разработать модель на основе экспертных оценок. Такие модели называют «экспертными системами» или «моделями на основе опыта».

Разработка такой модели включает в себя селективный отбор характеристик, обладающих (предположительно) хорошей прогностической силой, и назначение баллов каждому атрибуту, аналогично тому, как это происходит со статистическими моделями. Однако внедрение подобной модели производится на основе коллективного опыта и интуиции, а получающаяся модель отражает проводимую банком политику. Несмотря на отсутствие статистической основы модель экспертных оценок может стать более логичным и объективным инструментом принятия решений, чем оценка риска отдельными специалистами.

Разработка модели экспертных оценок должна происходить при участии отдела маркетинга, оценки и принятия решения по кредитным заявкам (кредитного отдела), отдела управления рисками и других соответствующих структур.

Подготовка плана проекта

План проекта должен содержать ясно обозначенный объем работ и временные рамки. Также он должен решить вопросы, связанные с документацией по проекту и стратегией внедрения. План проекта включает все предполагаемые осложнения и риски, гарантирует последовательность процесса разработки и внедрения скоринговой карты. Он обеспечит плавный переход от команды разработчиков к лицам, ответственным за тестирование, разработку стратегии и внедрение. Правильное планирование на данном этапе предотвратит сценарии, когда скоринговая карта разработана группой сотрудников, но не внедрена, потому что отделу IT не сообщили об этом заранее или потому что отчеты по новым скоринговым картам, необходимые для внедрения, не могут подготовить вовремя. Все заинтересованные стороны должны знать, что необходимо выполнить, кем и когда.

Определение рисков проекта

Успешная разработка скоринговой карты зависит от различных связанных между собой операций, при этом каждая последующая операция может начинаться только после того, как удовлетворительно завершена предыдущая. Как и решение, операция полностью зависит от исходных данных. В результате возникают некоторые риски, связанные с разработкой скоринговой карты:

- Непригодность данных или недостаточный объем данных
- Низкое качество данных («грязные» или недостоверные данные)
- Запаздывание/трудности в получении данных
- Не обладающие предсказательной силой или слабые данные
- Характеристики скоринговой карты или выводы, которые невозможно обработать с помощью действующей системы
- Изменение организационных направлений/приоритетов
- Возможные препятствия в ходе внедрения скоринговой карты
- Другие правовые и оперативные вопросы

На данном этапе следует определить риски проекта, серьезные препятствия и другие причины, которые могут повлиять на качество данных, и, когда это необходимо, следует составить запасной план.

Определение участников проекта

План проекта также определяет участников и собирает их в команду проекта, состоящую из специалистов различных отделов. Список предполагаемых участников приведен в Главе 2.

Данный список должен определить роли и обязанности участников, руководителей проекта и сотрудников, заключения которых требуются для успешного завершения того или иного этапа разработки скоринговой карты. Еще один список необходимо составить для лиц, которым нужно владеть информацией о предварительных результатах, сроках и предполагаемых стратегиях. Это обычно отделы, которые напрямую не участвуют в разработке скоринговой карты, но могут влиять на нее, внося изменения в кредитную стратегию.

После утверждения бизнес-плана, определяющего в общих чертах наблюдение для скоринговой карты, руководители проекта и сотрудники отделов, ответственных за разработку и внедрение скоринговой карты, утверждают полный план проекта, объем работ и документацию по проекту.

Целью такого бизнес-планирования не является создание дополнительных бюрократических препятствий для разработчиков скоринговых карт. Идея проста – в больших организациях, где различные отделы совместно выполняют работу по разработке, внедрению и управлению скоринговыми картами, для эффективности работы и прогнозирования результатов необходима некоторая согласованность действий.

Встреча по итогам работы скоринговой карты – это не лучшее время выяснить, что скоринговую карту, на которую вы потратили два месяца разработки, нельзя внедрить. Такое бизнес-планирование не должно быть формальным. Всем вопросам необходимо уделить достаточно внимания.

Почему формат скоринговой карты?

Эта книга описывает разработку обычной скоринговой карты (как показано на Рисунке 1.1). Несмотря на то, что прогнозные модели могут разрабатываться и в других форматах, например, в SAS-коде или С-коде, самым распространенным является формат скоринговой карты. Приведем несколько причин, по которым формат скоринговой карты предпочтительнее:

- Данный формат легко интерпретируем, привлекателен для широкого круга риск-менеджеров и аналитиков, которые не располагают глубокими познаниями в области статистики и добычи данных.
- Причины отказов, низких или высоких баллов можно объяснить пользователям, аудиторам, надзорным органам, топ-менеджерами и другому персоналу в понятной деловой форме.
- Процесс разработки данных скоринговых карт перестает быть «черным ящиком» и понятен широкой аудитории. Поэтому он соответствует требованию надзорных органов о прозрачности метода.
- Очень удобно проводить диагностику и мониторинг скоринговой карты, используя стандартные отчеты. Структура скоринговой карты позволяет аналитикам выполнять эти операции без глубоких познаний в области статистики и программирования. Скоринговая карта становится эффективным инструментом управления риском.

Процесс Разработки Скоринговой Карты, Этап 2: Обзор Данных и Параметров Проекта

Вероятно, по времени это наиболее длительный этап, требующий привлечения большого количества ресурсов. На этом этапе определяется, насколько осуществим проект по разработке скоринговой карты с практической точки зрения, а также его основные параметры. Параметры подразумевают определение исключений, целевой переменной, «окна выборки» и «окна созревания».

Доступность и качество данных

На данной стадии первоочередное внимание уделяется вопросу доступности данных с точки зрения их качества и количества. Для разработки скоринговой карты требуются достоверные и чистые данные, а также минимально допустимое количество «хороших» и «плохих» клиентов. Этот процесс можно сделать более удобным и эффективным, если данные размещать в специальных хранилищах или витринах данных.

Количество данных варьирует, но в целом оно должно соответствовать требованиям статистической значимости и случайности. На данной стадии точное количество не играет решающей роли, поскольку оно зависит от определения «плохого» заемщика, которое будет дано на следующем шаге. Однако по правилам для разработки аппликационной скоринговой карты требуется примерно 2000 «плохих» и 2000 «хороших» клиентов, которые могут быть случайно выбраны для каждой скоринговой карты из набора одобренных заявителей, получивших кредит в течение определенного интервала времени. Для поведенческих скоринговых карт это будет уже другой набор данных, отражающий «платежное» поведение клиента. Для коллекторских скоринговых карт используются записи со статусом просрочки платежа. Для аппликационных скоринговых карт, в которых применяется анализ отклоненных заявок, дополнительно потребуются данные об отказах (2000 записей). Отчеты компании об убытках/просрочках по платежам/выплатах по претензиям или другим показателям, а также данные об объеме кредитных заявок должны подтвердить первоначальное определение целевой переменной. Как правило, труднее найти достаточное количество «плохих» записей, чем «хороших».

Также команде проекта нужно будет определить, не искажены ли внутренние данные, предназначенные для разработки скоринговой карты, или же они являются недостоверными по другим причинам. Непроверенные демографические данные и другие характеристики анкеты, такие, как доход, могут быть в большей мере искажены, однако такие характеристики, как данные кредитного бюро, данные по недвижимости, финансовые коэффициенты и прочие показатели являются более устойчивыми и могут использоваться. Если

установлено, что данные, взятые из анкет, являются недостоверными, то скоринговые карты могут разрабатываться только на основе данных кредитного бюро.

Как только определено, что есть достаточное количество внутренних данных хорошего качества, нужно оценить внешние данные, определить количество данных и охарактеризовать их. Организация может разработать скоринговые карты на основе только внутренних данных или совместить эти данные с информацией из внешних источников, например, с данными кредитного бюро, данными центральных хранилищ страховых выплат, геодемографическими данными провайдеров. Некоторые организации получают и сохраняют подобные сведения по каждому заявителю в формате электронных баз данных. В тех случаях, когда данные недоступны в электронном виде или сохраняются на бумажных носителях, организация может либо вбить их в базу данных, либо купить ретроспективные данные у внешнего поставщика. Временной интервал для ретроспективных данных задается в соответствии с определениями «окна созревания» и «окна выборки».

На заключительной стадии, когда можно подтвердить как качество, так и количество внутренних и внешних данных, разрешено начинать сбор исходных данных для определения параметров проекта.

Сбор данных для определения параметров проекта

Чтобы определить проектные параметры для разработки скоринговой карты, сведения о клиентах сперва должны быть собраны в базу данных. Проектные параметры в первую очередь содержат определения «хороших» и «плохих» клиентов, определение «окна созревания» и «окна выборки», а также определение исключений для проектирования обучающей выборки и для самого процесса разработки скоринговой карты.

Следующие характеристики обычно собираются по кредитным заявкам за период от двух до пяти лет или отбираются из достаточно большой выборки:

- Клиент/идентификационный номер
- Дата обращения/получения кредита
- История долгов/выплат по претензиям
- Индикатор заявки заемщика - Одобрить/Отказать
- Продукт/канал продаж и другие идентификаторы сегментов
- Текущий статус клиента (например, нет операций по счету, счет закрыт, потеря кредитной карты, кража средств по кредитной карте, мошенничество)

Для разработки поведенческой скоринговой карты клиенты выбираются в определенный момент времени и их поведение анализируется обычно в течение 6 или 12 месяцев.

В зависимости от бизнес-целей карты в нее могут добавляться всевозможные другие данные, в том числе демографические: возраст, регион проживания и существующие зависимости; время проведения кампаний по привлечению клиентов; средние значения различных индикаторов на основе данных кредитного бюро; другие критерии, которые помогут в построении полного профиля клиентской базы вашей организации.

Определение параметров проекта

Последующий анализ проводится не только для определения параметров проекта. Это еще способ понять бизнес с помощью данных.

Исключения

Определенные типы клиентов необходимо исключить из обучающей выборки. В целом для разработки скоринговой карты используют сведения по тем клиентам, которых вам приходится оценивать в ходе анализа типичного ежедневного потока клиентов. Клиенты с нетипичным поведением – например, мошенники – оцениваются с помощью нескольких критериев, независимых от скоринговой оценки, и не должны попасть в обучающую выборку. Они могут включать определенных клиентов (например, персонал самого банка, VIP-клиенты, зарубежные клиенты, предварительно одобренные, клиенты с потерянными/украденными картами, умершие, несовершеннолетние, «отказники» по кредитам) в «окне созревания». Заметим, что некоторые разработчики включают клиентов, отказавшихся от кредита, в категорию «серая зона», поскольку эти клиенты прошли скоринг и были одобрены. Поэтому они попали в категорию «скоринговая оценка типичного ежедневного потока заявок». С точки зрения логики этот подход является наилучшим. Если есть географические зоны или рынки, где компания прекратила свою деятельность, то данные по этим зонам или рынка должны быть исключены, чтобы обучающие данные отражали ожидаемый статус клиента. Например, компания, выдающая автокредиты, раньше финансировала покупку развлекательных транспортных средств, таких, как снегоходы, судна и вездеходы. Однако год спустя она решила сосредоточиться на финансировании личного автотранспорта и прекратило финансировать все другие активы. В целях разработки скоринговой карты обучающие данные для этой компании должны включать только заявки на автокредит.

Другой способ взглянуть на исключения – это рассмотреть только определенный сегмент (однородную аудиторию, которая принимается за типичную). Например, если вы хотите разработать скоринговую карту, применимую к жителям крупных городов, вы можете не включать в обучающую выборку тех, кто живет в сельской местности. Аналогично, любой клиент или заявитель, который не проходит процедуру скоринга или ведет себя нетипично, не должен быть включен в обучающую выборку.

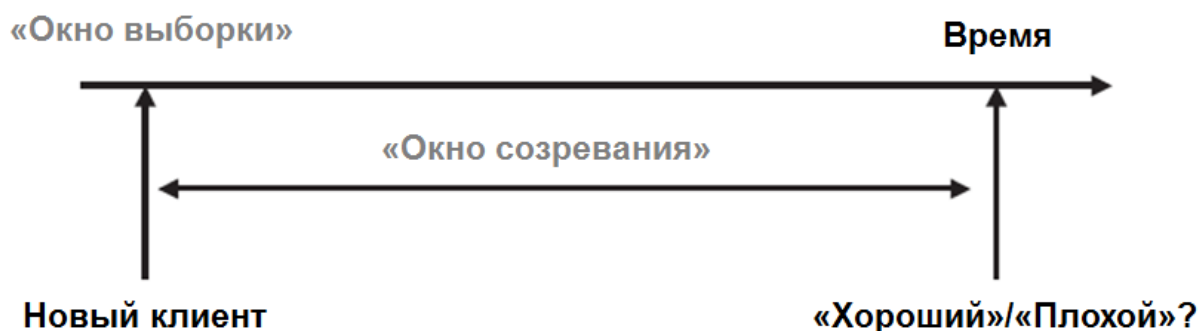
«Окно созревания» и «окно выборки», определение «плохого» клиента

Скоринговые карты разрабатываются, исходя из предположения «прошлое отражает будущее». На основе этого предположения анализируется поведение прошлых клиентов, чтобы спрогнозировать поведение будущих клиентов.

Для того, чтобы корректно выполнить этот анализ, нужно собрать необходимые данные за определенный промежуток времени, а затем осуществить их мониторинг в течение другого определенного отрезка времени и оценить, были ли они «хорошими» или «плохими». Собранные данные (переменные) наряду с соответствующей классификацией (целевая переменная: «хороший»/«плохой») составляют основу для разработки скоринговой карты.

На рисунке 4.1 сделано предположение, что новый клиент получил кредит в определенный период времени (например, 1 февраля 2003 г.). В некоторый момент времени в будущем (например, через 90 дней) вы должны будете определить, был ли этот клиент «хорошим» или «плохим» (чтобы классифицировать поведение).

Рисунок 4.1 Определение «окна созревания» и «окна выборки»



«Окно созревания» представляет собой тот промежуток времени, когда клиент, собственно говоря, имел возможность себя проявить (целевая переменная: 90+). «Окно выборки» представляет собой тот промежуток времени, когда те или иные клиенты отбираются для анализа (попадают в выборку). Рекомендуется также проанализировать, какой период «созревания» является идеальным для того или иного продукта (региона, типа клиента и т.п.). В некоторых случаях, таких как мошенничество и банкротство, временной период уже известен или предопределен. Но тем не менее вышеописанный анализ полезно выполнить для того, чтобы определить идеальное «окно созревания».

Самый простой способ определить «окно созревания» и «окно выборки» состоит в том, чтобы проанализировать портфель на предмет просроченной задолженности и построить графики с различными сценариями «плохих» случаев в течение определенного времени: просрочка более 30/60/90 дней. Хороший источник для подобных данных - ежемесячный или ежеквартальный отчет по когортному или винтажному анализу, имеющийся в любом отделе кредитных рисков.

Пример винтажного анализа для просроченной задолженности свыше 90 дней и «окна созревания» 9 месяцев приведен на Рисунке 4.2. Данные, выделенные жирным шрифтом, показывают текущий статус просрочки платежа на определенный отчетный период времени.

Рисунок 4.2 Пример когортного/винтажного анализа

<i>Bad = 90 days</i>	<i>Time on Books</i>								
<i>Open Date</i>	<i>1 Mth</i>	<i>2 Mth</i>	<i>3 Mth</i>	<i>4 Mth</i>	<i>5 Mth</i>	<i>6 Mth</i>	<i>7 Mth</i>	<i>8 Mth</i>	<i>9 Mth</i>
Jan-03	0.00%	0.44%	0.87%	1.40%	2.40%	2.80%	3.20%	3.60%	4.10%
Feb-03	0.00%	0.37%	0.88%	1.70%	2.30%	2.70%	3.30%	3.50%	
Mar-03	0.00%	0.42%	0.92%	1.86%	2.80%	3.00%	3.60%		
Apr-03	0.00%	0.65%	1.20%	1.90%	2.85%	3.05%			
May-03	0.00%	0.10%	0.80%	1.20%	2.20%				
Jun-03	0.00%	0.14%	0.79%	1.50%					
Jul-03	0.00%	0.23%	0.88%						
Aug-03	0.00%	0.16%							
Sep-03	0.00%								

Таблица имеет достаточно простую интерпретацию. Так, на первой строчке 2.4% заемщиков, получивших кредит в январе 2003 г., «выпали в просрочку» более 90 дней через 5 месяцев.

Рисунок 4.3 представляет собой график «плохих» случаев за 14 месяцев.

Рисунок 4.3 Рост «плохих» случаев



Этот рисунок показывает пример типичного портфеля кредитных карт — приведены данные о «плохих» клиентах за 14 месяцев. Он иллюстрирует быстрый рост «плохих» случаев в первые несколько месяцев, а затем стабилизацию, начиная с 12-го месяца.

Поэтому выборку формируют по тому промежутку времени, когда процент «плохих» клиентов считается постоянным или когда когорта считается «созревшей» (т.е. когда процент «плохих» клиентов начинает выравниваться). В предыдущем примере наилучшим «окном выборки» будет промежуток времени между 12-м и 14-м месяцами (например, с января по март 2003 г.), а «окно созревания» составит в среднем 13 месяцев.

Отбор выборок из «созревшей» когорты осуществляется так, чтобы минимизировать вероятность ошибочной классификации (т.е. всем клиентам предоставляют достаточное время, в течение которого они могут стать «плохими») и гарантировать, что полученный в результате сценарий «плохих» случаев из «несозревшей» выборки не занижает итоговый ожидаемый процент «плохих» клиентов. Например, если выборка была сформирована семь месяцев назад, то около 4.5% наблюдений выборки будут классифицированы как «плохие» случаи (т.е. процент просрочки составит 4.5%). В то время как «созревшая» выборка по данному портфелю должна включать 6% наблюдений, отнесенных к «плохим» (т.е. процент просрочки составляет 6%). Поэтому некоторые «плохие» клиенты, отобранные за семь месяцев, будут ошибочно классифицированы как «хорошие», если выборка была взята за указанный период времени.

Временной горизонт «созревания» зависит от продукта и определения «плохого» клиента. Счета по кредитным картам становятся «зрелыми» после 18-24 месяцев, в то время как счета сроком 3-4 года являются типичными для разработки скоринговой карты по ипотечным кредитам. В какой-то степени это выглядит как самосбывающееся пророчество, так как портфели по кредитным карточкам имеют по своей природе более высокий риск, чем ипотечные кредиты, и поэтому достигают того же самого уровня просрочки намного быстрее. В случае отсутствия денег клиенты с большей вероятностью прекращают платить по кредитной карте, чем по обеспеченному кредиту. Аналогично процент просрочки на меньший срок

(30 дней) значительно превышает уровень просрочки 60 или 90 дней. Скоринговые карты, разрабатываемые для выплат по претензиям, мошенничества, банкротства и других показателей, вероятно, получают уникальные постоянные профили, которые можно будет определить, выполнив схожий анализ. Поведенческие скоринговые карты предусматривают, как правило, «окно созревания» из расчета 6-12 месяцев. Коллекторские модели строятся, как правило, на данных одного месяца, но все чаще компании строят такие карты для более коротких временных интервалов – до двух недель, чтобы облегчить разработку более подходящих способов взыскания долгов. Когда речь идет о разработке прогнозных моделей в соответствии с определенными требованиями надзорных органов – например, в соответствии с требованиями соглашения Базель II – «окно созревания» может определяться регулятором.

Разработка скоринговых карт по просрочкам требует проведения многократного анализа для различных статусов просрочек. Он выполняется, потому что различные статусы просрочек дают различное количество наблюдений в выборке. Разные факторы, такие, как «окно выборки» и определение «хороший»/«плохой» в ряде случаев необходимо скорректировать, чтобы получить достаточно большую выборку (смотрите следующую главу). При разработке скоринговых карт банкротства или списания безнадежного долга вполне достаточно однократного анализа, поскольку используется лишь один возможный статус «плохого» клиента. По возможности данный анализ должен проводиться, используя понятие «хотя бы раз плохого» клиента (т.е. клиент считается «плохим» в случае, если он хотя бы раз в период наблюдения выпадал в просрочку). Если это не представляется возможным, из-за трудностей с данными, достаточно «текущего» статуса «плохого» клиента, когда статус просрочки клиентов определяется на конец самого последнего месяца наблюдения. На рисунке 4.4 приведен пример, показывающий 24-месячную историю просрочек по одному клиенту. В графе «Месяц» указан текущий месяц, а в графе «Просрочка» указано количество месяцев просрочки.

Рисунок 4.4 24-месячная история просрочек по клиенту

Month	1	2	3	4	5	6	7	8	9	10	11	12
Delq	0	0	1	1	0	0	0	1	2	3	0	0
Month	13	14	15	16	17	18	19	20	21	22	23	24
Delq	0	0	1	2	0	0	0	1	0	1	0	0

Используя определение «хотя бы раз плохого» клиента, статус просрочки по данному клиенту можно отнести к 3 месяцам просрочки. Однако, используя статус «текущего плохого» клиента, данного заемщика можно назвать не имеющим просрочки (т.е. ноль месяцев просрочки).

Сезонные эффекты

Здесь необходимо рассмотреть изменение объема кредитных заявок и процента одобрения в зависимости от времени, а также эффект сезонности. Необходимо убедиться в том, что обучающая выборка (сформированная по «окну выборки») не содержит данные, взятые из «нетипичных» периодов. Таким образом, выборка, используемая для разработки скоринговой карты, соответствует типичным периодам деловой активности, адекватно представляя типичную популяцию клиентов «с улицы». Цель – подтвердить предположение, что «будущее похоже на прошлое» и таким образом, обучающая выборка репрезентативна для будущих заявителей (т.е. «типичных» клиентов). Фактически это помогает получить точные

прогнозы процента одобренных заявок/процента «плохих» клиентов, и что более важно, позволяет разрабатывать скоринговые карты, которые будут стабильными и выдержат проверку временем. В реальности такие процедуры в значительной степени выполняются для поиска нетипичного поведения, поскольку определить стандарт «типичного» клиента затруднительно.

Есть несколько способов устранить эффекты нетипичных периодов, когда популяция заявителей не репрезентативна для типичной популяции заявителей «с улицы». В любом случае необходимо сперва определить причины нетипичных периодов. Лучше всего это сделать путем сравнения характеристик среднестатистического клиента с характеристиками клиентов, отобранных по «окну выборки». Дополнительные причины изменений профиля могут быть получены на основе информации по маркетинговым кампаниям, проводившимся в период «окна выборки», или других факторов, которые могут повлиять на профиль заявителей. Например, организация прогнозирует, что заявителями по кредитным картам будут преимущественно зрелые люди, однако обнаруживает, что заявителями, отобранными по «окну выборки» длительностью один месяц, являются главным образом молодые мужчины. Анализ маркетинговых кампаний показывает, что организация в течение этого месяца активно привлекала клиентов на одной из выставок во время автошоу (обычно автомобили привлекают внимание молодых мужчин). Вооружившись данной информацией, компания может увеличить «окно выборки» до трех и более месяцев, чтобы уменьшить эффект, характерный для данного месяца.

Другой техникой «нормализации» данных является выявление источника нетипичного поведения. В предыдущем примере, если компания уверена в том, что в будущем молодые мужчины не являются ее целевой аудиторией и кредитоспособность этих молодых мужчин будет искажать качество ожидаемого кредитного портфеля, она может исключить молодых мужчин из обучающей выборки. В результате обучающая выборка (и статистические показатели кредитного портфеля) будут соответствовать типичному ежедневному потоку заявителей. Сезонные эффекты можно также устранить, используя большое количество «окон выборки», но при этом с одинаковым «окном созревания». Например, можно сформировать три выборки по январю, февралю и марту 2002 года, с «окном созревания» 24 месяца каждая. Поэтому статус «плохого» клиента для каждой выборки будет определяться кредитоспособностью клиента на январь, февраль и март 2004 года соответственно. Этот способ противоположен подходу, когда дата окончания периода наблюдения остается одной и той же для всех выборок (например, определение кредитоспособности на февраль 2004 года для всех трех когорт в приведенном примере – в результате получим когорты с различными «окнами созревания» по одной и той же выборке).

В тех случаях, когда речь идет о неустойчивых выборках или расширение «окна выборки» невозможно, но при этом причины нетипичного профиля заявителей известны и характерны для одного определенного месяца, можно сформировать выборку, исключив нетипичные наблюдения. Однако для этого потребуются подробная информация о распределении характеристик типичного потока заявителей. Рекомендуется, чтобы в выборке исключенных наблюдений до ее отбраковки был проведен анализ трендов.

Определение «плохого» клиента

На данной стадии заемщики делятся на три главные группы: «плохие», «хорошие» и «серая зона». Для банкротства, безнадежных долгов или мошенничества определение «плохого клиента» является однозначным. Относительно просроченной задолженности возможны разные интерпретации в зависимости от

уровня просрочки платежа. Как упоминалось, анализ различных статусов «плохого» клиента (как показано на Рисунке 4.3) будет давать различный процент «плохих» клиентов. Наиболее подходящий статус «плохого» клиента выбирается для этих наблюдений с помощью некоторых факторов, перечисленных ниже. Определение «плохого» клиента основывается на нескольких положениях:

- Определение должно соответствовать организационным целям. Если цель состоит в том, чтобы увеличить прибыльность, то определение должно производить разбиение в точке, когда клиент перестает приносить прибыль. Здесь возникают свои трудности: заемщик, который регулярно выпадает в просрочку, опаздывая с платежами, но не более чем на два или три месяца, может быть прибыльным (выплачивая штрафы за просрочку платежа). Для клиентов страховых компаний можно определить сумму выплат по претензиям. Если целью является выявление просрочки, то определение будет еще проще (например, «когда-либо имел просрочку» 60 или 90 дней).
- Определение должно соответствовать продукту или цели, для которой строится скоринговая карта, например банкротство, мошенничество, выплаты по претензиям (свыше 1000 \$) и сбор долгов по кредитам (у которых возврат меньше, чем 50% за последние три месяца).
- Чем «строже» определение, например, «списание» или «120 дней просрочки», тем более точно производится дифференцирование клиентов, но в некоторых случаях это может привести к резкому уменьшению размеров выборки.
- Более «свободное» определение (например, 30 дней просрочки) приводит к большему количеству записей, но уже, возможно, недостаточно хорошо делит клиентов на «хороших» и «плохих», и, таким образом, скоринговая карта получается слабой.
- Определение должно быть легко интерпретируемым и удобным для контроля (например, 90 дней просрочки, банкротство, подтвержденное мошенничество, выплата по претензии свыше 1000 \$). Такие определения, как «три просрочки по 30 дней» или «две просрочки по 60 дней» или «одна просрочка 90 дней», возможно, более точно описывают «плохого» заемщика, но они намного сложнее для контроля, к тому же, возможно, не подойдут любой компании. Выбор более простого определения приводит к упрощению принятия решения и более простой интерпретации «плохих» клиентов (например, 4% «плохих» клиентов означает, что 4% попали в просрочку более 90 дней в течение срока их существования).
- Часто компании выбирают определение «плохого» клиента с бухгалтерской точки зрения списания долгов.
- В некоторых случаях может оказаться разумным сформулировать единые определения «плохих» клиентов для различных сегментов и скоринговых карт, применяемых в одном и той же компании. Это делается для упрощения принятия решения, особенно в ситуации, когда используется большое количество скоринговых карт. Вместе с масштабированием баллов это снижает затраты на обучение и программирование при повторной разработке скоринговых карт.
- Возможно также существование регулирующих или других внешних требований, определяющих понятие просрочки, которое не совпадает с «рабочим» определением самой организации. Так, регулирующие требования к отчетности Базеля II приводят свое определение «плохого» клиента, которое непосредственно связано с размером экономических потерь, установленным уровнем просроченной задолженности или

определенным уровнем ожидаемых потерь. Согласно Базелю II дефолтом считается наличие просрочки длительностью 90 дней.

В некоторых случаях определение «плохого» клиента задают в зависимости от дефолта по причине недостатка исторических данных (например, у организации есть только исторические данные за 12 месяцев) так, чтобы статус «плохого» клиента, который сформировался в ходе анализа (см. Рисунок 4.3), соответствовал клиентам, имевшим хотя бы одну просрочку длительностью 30 дней. Некоторые организации не сохраняют данные о ежемесячных платежах, и единственный способ дать определение «плохого» клиента – это текущий статус просрочки (сравниваемый с наличием хотя бы одной просрочки у клиента).

Подтверждение статуса «плохого» клиента

После того как банк определится с исходным понятием «плохой», выполнив анализ, описанный в предыдущих разделах, ему необходимо провести дальнейший анализ для того, чтобы удостовериться, что выявленные заемщики действительно «плохие». В большей степени это актуально для ситуации, когда определение «плохого» клиента не является окончательным (например, на основе определенного количества дней просрочки или суммы возмещения). Для этого возможно применение экспертизы или анализа, а также комбинации того и другого в зависимости от имеющихся в распоряжении ресурсов и доступных данных. Важно отметить, что описанные методы могут существенно помочь при проведении анализа, но не являются единственно верными.

Метод «Консенсус» Экспертный метод или метод «консенсус» заключается в том, что в результате совместного обсуждения определения «плохой заемщик» кредитным отделом, отделом маркетинга и отделом кредитных операций принимается формулировка, кого именно считать «плохим», принимая во внимание опыт и различные точки зрения сотрудников подразделений, а также результаты анализа, описанного в предыдущих разделах.

Аналитические методы Существуют два аналитических метода для того, чтобы подтвердить предполагаемое определение «плохого» клиента:

1. Анализ уровней перехода (Roll Rate)
2. Сравнительный анализ текущего уровня задолженности по отношению к наихудшему

В дополнение к этим двум методам рекомендуется также использовать анализ рентабельности, который поможет убедиться, что определенные банком «плохие» клиенты действительно нерентабельны или имеют отрицательную чистую приведенную стоимость (NPV). Стоит также отметить, что на практике это гораздо легче проверить на отдельных продуктах (например, пластиковые карты), чем на банковском уровне в целом.

Анализ уровней перехода Данный метод заключается в сравнении наихудшего уровня просрочки за указанные «предыдущие x» месяцы с «последующими x» месяцами. Затем вычисляется процент заемщиков, которые, сохраняют наихудший уровень просрочки, «улучшаются» или переходят в следующие категории просрочек.

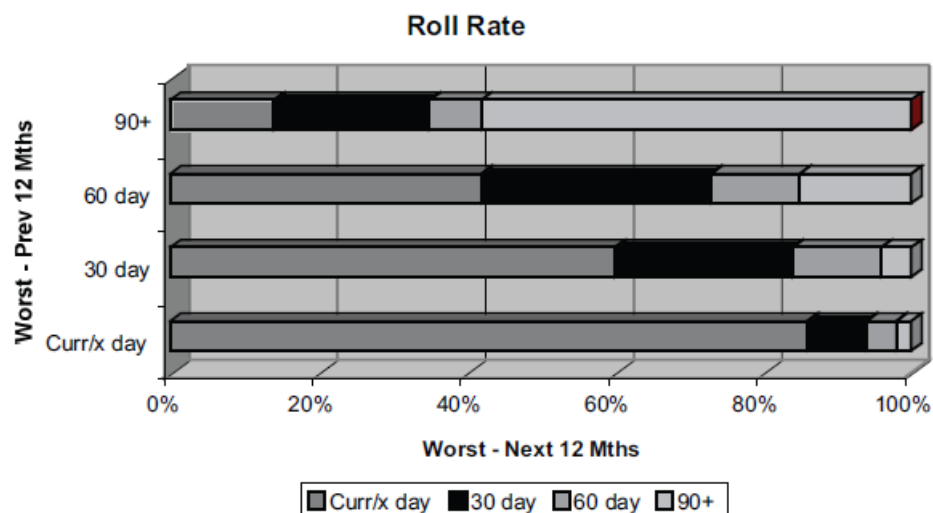
Например, Рисунок 4.5 показывает статус просрочки возобновляемой кредитной линии за 24-месячный период, разбитый на два равных 12-месячных периода: «предыдущий» и «следующий».

Рисунок 4.5 История просрочек клиента

“Previous” 12 Months												
Month	1	2	3	4	5	6	7	8	9	10	11	12
Delq	0	0	1	1	0	0	0	1	2	3	0	0
“Next” 12 Months												
Month	13	14	15	16	17	18	19	20	21	22	23	24
Delq	0	0	1	2	0	0	0	1	0	1	0	0

Исходя из этих данных, наихудшая просрочка платежа за «предыдущий» 12-месячный период наблюдалась в октябре (десятый месяц) – просрочка в три месяца. Кроме того, просрочку в два месяца мы наблюдали на 16-м месяце уже в «следующем» 12-месячном периоде. Данная информация собрана по всем клиентам и графически представлена на Рисунке 4.6.

Рисунок 4.6 Графический анализ уровней перехода (Roll Rate)



Основная цель здесь состоит в том, чтобы распознать ту «точку невозврата», когда уровень просрочки заемщика становится безнадежным. На практике мы сталкиваемся с тем, что большинство клиентов, достигших 90 дней просрочки, уже не восстанавливаются, а становятся только хуже (на «последующем» периоде), подтверждая, таким образом, определение «плохого» клиента. В представленном примере только 18% клиентов, которые хотя бы раз имели просрочку в 30 дней за предыдущие 12 месяцев, попали в просрочку 60 и 90 дней. Однако почти 70% просрочек, которые достигают 90-дневного срока, остаются в наихудшей категории просрочки. В этом случае определение «просрочка хотя бы раз более 90 дней» для «плохого» клиента имеет больше смысла, поскольку она действительно изолирует «безнадежных» должников. Наоборот, определение «плохого» как достигшего просрочки 30 дней, не будет адекватным, поскольку большинство таких просрочек восстанавливается. Неубедительные доказательства, полученные в ходе такого анализа, возможно, укажут статус потенциального «серого» клиента.

Стоит отметить, что Базельское соглашение определяет «дефолт» как период, после которого банк признает должника неспособным погасить обязательство в

полной мере и дает определение «плохого» клиента как достигшего 90 дней просрочки (некоторые регуляторы могут изменить это определение до 180 дней для определенных продуктов).

Сравнительный анализ текущего уровня задолженности по отношению к наихудшему Этот метод в принципе похож на предыдущий, но более простой в выполнении. Мы сравниваем наихудшую (возникшую хотя бы раз) просрочку клиента с его текущим статусом просрочки. Как и в предыдущем методе, основная цель здесь состоит в том, чтобы найти «точку невозврата». Пример приведен на Рисунке 4.7.

Рисунок 4.7 Сравнение наихудшего уровня задолженности с текущим

		Worst Delinquency					Write-off
	Current	Current	30 days	60 days	90 days	120 days	
Current	Current	100%	84%	53%	16%	7%	
Delinquency	30 days		12%	28%	10%	8%	
	60 days		4%	11%	14%	10%	
	90 days			8%	44%	13%	
	120 days				16%	62%	
	write-off						100%

Рисунок показывает, что из всех клиентов, которые хотя бы раз имели 30-дневную просрочку, большинство (84%) не имеют просрочек в настоящее время. И наоборот, 60% клиентов, достигших 90-дневной просрочки, остаются в ней или становятся хуже. Это снова подтверждает, что 90- или 120-дневное определение «плохого» заемщика более корректное, поскольку за это время может быть собрано уже достаточный процент «плохих» клиентов.

Следует отметить, что описанные методы для формирования и подтверждения корректного определения «плохого» клиента применимы как в аппликационных, так и в поведенческих скоринговых картах. Несмотря на то, что поведенческие скоринговые карты обычно разрабатываются с использованием 6-месячного временного окна, тем не менее полезно выполнить этот анализ для того, чтобы определить действительно «плохого» клиента.

Определение «хорошего» клиента и «серой зоны»

После того, как «плохие» клиенты окончательно определены, можно применить тот же самый анализ для определения «хороших». Напомним еще раз, что определение должно соответствовать организационным целям и другим критериям, оговоренным ранее. Определение «хороших» клиентов является менее аналитическим и, как правило, очевидно. Некоторые характеристики «хорошего» клиента:

- Ни одной просрочки или просрочка с ролл-рейтом менее 10%
- Прибыльный клиент (или с позитивным NPV)
- Ни одной выплаты по претензии
- Никогда не был банкротом
- Отсутствие мошенничества
- Уровень восстановления (recovery rate) при сборе задолженностей, например 50%

Стоит также отметить, что в то время как «хорошие» клиенты должны сохранять свой статус на протяжении всего «окна созревания», «плохие» клиенты проявляются, лишь достигая указанной стадии просрочки в любое время в пределах «окна созревания» (согласно определению «наличие просрочки хотя бы один раз»).

«Неопределенные» клиенты или «серая зона» - это те, которые в результате не попадают в категорию ни «хороших», ни «плохих». Эти клиенты не имеют достаточной кредитной истории для их классификации или имеют некоторую небольшую просрочку, поэтому их нельзя отнести к «хорошим», но и к «плохим» они также пока не попали. «Серую зону» можно описать следующим образом:

- Клиенты, попавшие в просрочку 30 или 60 дней, но не уходящие в более глубокую просрочку (то есть неокончательно «плохие»)
- Неактивные или добровольно отказавшиеся клиенты, например, когда заявление о предоставлении кредита было одобрено, но клиент не воспользовался своим правом получения кредита, а также клиенты, которые были одобрены, но не «созрели» для какой-либо классификации
- Клиенты с небольшой суммой просрочки (например, клиенты, у которых самый высокий баланс по кредитной карте меньше 20\$)
- Клиенты страховой компании с размерами выплат по претензиям меньше установленной суммы
- Клиенты с NPV = 0

Отметим, что некоторые разработчики скоринговых карт относят всех добровольно отказавшихся и «неучтенных» клиентов к отклоненным, чтобы исключить их из анализа. Однако заявители, самостоятельно отказавшиеся от кредита, являются предполагаемыми клиентами и, возможно, отказались от кредита из-за проблем с обслуживанием клиентов. Если данные по ним используются повторно, то данные клиенты окажутся повторно оцененными и, возможно, снова одобренными. Поэтому этих клиентов следует учитывать в процессе разработки скоринговой карты в качестве «серой зоны».

«Серая зона» используется только в тех случаях, когда «плохой» клиент может быть определен по-разному, и не требуется, когда определение является четким (например, банкрот). Принято, что «серая зона» не должна составлять более 10-15% кредитного портфеля.

Включение «серой зоны» в обучающий набор данных увеличивает ошибочную классификацию клиентов. Например, можно ошибочно отнести неопределенного заемщика к «хорошим» и тем самым недооценить процент «плохих» клиентов.

В тех случаях, когда «серая зона» является очень большой (например, когда имеется большое количество неактивных клиентов), анализ должен быть направлен на выяснение причин данной ситуации, например, существуют другие кредитные карты с более высокими лимитами или более низкими процентными ставками, есть другие кредитные карты с более лучшими программами лояльности или также отсутствует активность по всем другим картам. После выяснения причин, в силу которых клиенты не пользуются кредитным продуктом, могут быть предприняты соответствующие действия для того, чтобы исправить ситуацию (например, увеличение лимита или снижение процентных ставок для «лучших» клиентов, начисление бонусных баллов для постоянных клиентов, предложение скидок по банковским картам, выпущенных под брендами розничных сетей, или отмена кредитных карточек для неактивных клиентов).

При построении скоринговой карты используются только клиенты, определенные как «хорошие» и «плохие» (и отклоненные клиенты для аппликационной

скоринговой карты). «Серую зону» включают для того, чтобы прогноз точно отражал популяцию клиентов «с улицы», поскольку они получили скоринговую оценку и по ним было принято решение.

После окончательного определения "плохого" заемщика нужно зафиксировать их количество в кредитном портфеле, которое будет использоваться в ходе оверсемплинга.

Данный раздел ограничивается клиентами, кредитоспособность которых известна. В случае применения аппликационных скоринговых с использованием анализа отклоненных заявок в обучающую выборку необходимо включить дополнительную категорию кредитоспособности, «отклоненных» клиентов (т.е. тех клиентов, которым отказали в выдаче кредита или предоставлении услуг). Это позволяет спроектированной обучающей выборке точно отражать популяцию заявителей в целом, а не только популяцию одобренных. Анализ отклоненных заявок будет освещен позже в Главе 6.

Сегментация

Иногда использование нескольких скоринговых карт для одного портфеля обеспечивает лучшее дифференцирование риска, чем использование одной скоринговой карты. Это происходит каждый раз, когда мы имеем дело с «разношерстной» аудиторией, состоящей из различных групп, и одна и та же скоринговая карта не может работать достаточно эффективно для всех, поскольку различные характеристики требуются для прогноза риска по различным группам. Процесс идентификации этих групп называют сегментацией. Существуют два основных способа сегментации клиентов:

1. Деление на группы производится на основе опыта и отраслевых знаний, а затем проверяется аналитическим путем
2. Сегментация производится статистическими методами типа кластерного анализа или деревьев решений

В любом случае полученные сегменты должны быть достаточно большими в целях получения достоверной выборки для разработки отдельной скоринговой карты. Сегменты, которые указывают на различную кредитоспособность, но при этом недостаточно большие для разработки отдельной скоринговой карты, должны рассматриваться дифференцированно с использованием различных порогов отсека и кредитных стратегий.

Стоит также отметить, что в ходе разработки скоринговой карты риска, «характерный» сегмент выявляется не просто на основе заданных характеристик (например, демографических), а на основе его кредитоспособности. Основная цель состоит в том, чтобы определить сегменты, основанные на реализации риска, а не только на профиле риска.

Выявление различного поведения, однако, не является достаточной причиной для проведения сегментации. Различия среди заемщиков необходимо перевести в измеряемые бизнес-результаты (например, выделение сегмента с более низкими потерями или более высоким процентом одобренных заявок). Пример такого измерения дан в разделе «Сравнительный анализ улучшений» данной главы.

Сегментацию, независимо от того, на основе чего она построена – опыта или статистических методов, нужно проводить с учетом будущих задач. В большинстве случаев анализ и опыт базируется на прошлых данных, однако скоринговые карты необходимо разрабатывать для будущих клиентов. Единственный способ добиться этого – это скорректировать сегментацию, например, с учетом предполагаемого целевого рынка. Обычно сегментацию

проводят, чтобы выявить оптимальный набор сегментов, которые максимально увеличат эффективность модели. Предложенный подход заключается в поиске сегментов с оптимальной кредитоспособностью (например, целевых рынков). Он подчеркивает важность максимальной эффективности с точки зрения бизнеса и гарантирует, что процесс разработки скоринговой карты увеличит ценность бизнеса.

Базельское соглашение также уделяет соответствующее внимание сегментации, определяя сегменты как "однородные пулы риска". Оно не указывает какого-то определенного подхода к сегментации, разрешая банкам по своему усмотрению выделять собственные уникальные сегменты.

Сегментация на основе опыта (Эвристическая сегментация)

Эвристическая сегментация использует бизнес-знания, операционные аспекты и практический опыт банковской индустрии. Источниками для возникновения идей могут служить:

- Отделы маркетинга и управления риском, выявляющие различные профили заявителей в заданном сегменте
- Кредитный портфель, оценка которого произведена на основе одной и той же скоринговой карты и одного и того же балла отсечения, но с сегментами, показывающими значительное различие в поведении (например, более высокий уровень «плохих» клиентов)
- Улучшение (новая разработка) существующего продукта
- Умышленное выделение определенной группы (например, «золотые клиенты»)
- Новые маркетинговые стратегии

Обычно сегментация базируется на следующих факторах:

- **Демографические характеристики.** Региональный фактор (региональный/ федеральный уровень, городская/сельская местность, почтовый индекс, соседство), возраст, стиль потребления, срок постановки в кредитном бюро, срок пребывания в банке и т.п.
- **Тип продукта.** Золотая/платиновая карта, срок ипотеки, тип страховки, обеспеченный/необеспеченный кредит, размер кредита и т.п.
- **Источники бизнеса (каналы).** Фронт-офис, филиал, торговая точка, интернет, дилеры, брокеры и т.п.
- **Доступность данных.** Наличие у банка кредитной истории по заемщику, данные кредитного бюро и т.п.
- **Тип заявителя.** Существующий/новый клиент, первый кредит/продление ипотеки, профессиональные группы (например, инженеры, доктора и т.д.)
- **Клиенты по другим продуктам банка.** Заемщик по ипотеке, обратившийся на получение кредитной карточки в том же самом банке

После того, как сформулированы идеи насчет возможной сегментации клиентов, необходимо провести дальнейший анализ. Он обусловлен двумя причинами. Первая причина заключается в том, что эти идеи необходимо подтвердить, используя как минимум несколько эмпирических доказательств. Вторая причина состоит в том, что анализ поможет лучше определить сегменты по таким критериям, как кредитная история заемщика, почтовый индекс, молодой/старый возраст с помощью подходящих точек разбиения (например, точки разбиения помогут ответить на вопрос «Что представляет собой хорошая кредитная история?», «Кого считать «молодым» заявителем?»).

Еще один простой способ подтвердить идеи по поводу сегментации и определить необходимость в ней – это анализ поведения одной и той же характеристики для различных заранее определенных сегментов. Если одна и та же характеристика в каждом сегменте дает различные прогнозы, то это, возможно, указывает на необходимость разработки сегментированных скоринговых карт. Однако если характеристика предсказывает один и тот же уровень риска независимо от сегмента, то разработка дополнительной скоринговой карты не требуется, поскольку никакого дифференцирования не наблюдается. Рисунок 4.8 показывает наблюдаемый процент «плохих» клиентов для характеристик «Статус проживания» и «Количество кредитных линий», сегментированных по возрасту: старше 30 лет и моложе 30 лет. При этом наблюдаемый процент «плохих» клиентов для несегментированных атрибутов указан в крайнем столбце справа.

Рисунок 4.8 Проценты «плохих» клиентов по характеристикам, сегментированным на основе возраста

Bad Rate			
	<i>Age > 30</i>	<i>Age < 30</i>	<i>Unseg</i>
Res Status			
Rent	2.1%	4.8%	2.9%
Own	1.3%	1.8%	1.4%
Parents	3.8%	2.0%	3.2%
Trades			
0	5.0%	2.0%	4.0%
1-3	2.0%	3.4%	2.5%
4+	1.4%	5.8%	2.3%

В примере процент «плохих» различен для разных групп клиентов. Одна и та же информация (то есть атрибуты «Съемщики» и «Те, кто живет с родителями») предсказывает процент «плохих» по-разному для клиентов более старшего и более молодого возраста. Это говорит о том, что сегментация по возрасту является хорошей идеей. Заметим, что если бы применялась только одна скоринговая карта (т.е. сегментация не использовалась бы), то все клиенты получили бы одинаковые баллы, а так они получают разные баллы в соответствии с их уровнем риска.

То же самое верно для клиентов с количеством кредитных линий «0» и «4+» - одна и та же информация предсказывает процент «плохих» по-разному для клиентов старше и моложе 30 лет. Оба примера взяты из американской практики. Клиенты старше 30 лет, живущие со своими родителями, являются более высокорискованными. То же самое можно сказать о клиентах старше 30 лет, у которых нет открытых кредитных линий. Это, возможно, типично для клиентов более молодого возраста, однако в США обычно принято, что к 30 годам заявитель уже накопил достаточную историю платежей по нескольким кредитным продуктам. Поэтому отсутствие кредитов в этом возрасте свидетельствует о высоком риске.

Другой способ подтверждения идей по поводу сегментации и определения уникальных сегментов состоит в том, чтобы рассмотреть процент «плохих» в различных отобранных группах. Этот метод подразумевает анализ процента «плохих» клиентов для различных атрибутов в отобранных характеристиках и выявление соответствующих сегментов, сильно различающихся с точки зрения кредитоспособности.

Пример подобного анализа приведен на Рисунке 4.9. Данный рисунок показывает типичную сегментацию по возрасту, каналы привлечения и типу клиента. Что касается возраста, отчетливо видно, что категория клиентов «моложе 30 лет» значительно хуже по кредитоспособности, чем остальные две категории. В данном случае сегментация по возрасту имеет смысл. Вместо широких категорий, приведенных в примере, можно использовать более точные группировки для возраста и очевидно, что число сегментов должно быть более двух.

Рисунок 4.9 Проценты «плохих» клиентов по заранее определенным сегментам

<i>Attributes</i>	<i>Bad Rates</i>
Age	
Over 40 yrs	1.8%
30-40 yrs	2.5%
Under 30	6.9%
Source of Business	
Internet	20%
Branch	3%
Broker	8%
Phone	14%
Applicant Type	
First-time buyer	5%
Renewal mortgage	1%

Относительно каналов привлечения, все четыре атрибута показывают различный процент «плохих» клиентов и могут быть отнесены к уникальным сегментам. Однако канал привлечения клиентов «По интернету» начал использоваться совсем недавно и составляет очень маленькую долю кредитного портфеля. В данном случае имеет смысл объединить сегменты «По интернету» и «По телефону» и заново разработать скоринговую карту после того, как будет накоплено достаточное количество клиентов данного типа. Несмотря на то, что данный анализ приведен здесь в качестве способа сегментации, его можно также проводить ежеквартально с целью выявления потенциальных проблемных вопросов, связанных с просрочкой платежа.

Оба вышеописанных метода достаточно легко внедрить. Они могут в некоторой степени измерить приемлемую точность сегментации, осуществленную с помощью опыта и интуиции. Также они помогут выбрать точки разбиения, например, по возрасту для характеристик, выбранных в качестве основы сегментации (как это показано на Рисунке 4.8).

Статистическая сегментация

Кластеризация Кластеризация (объединение в кластеры) - довольно широко используемая техника идентификации групп, схожих по определенным признакам (переменным). Кластеризация баз данных заключается в размещении объектов в различные группы (кластеры) в зависимости от самих данных. Объекты внутри кластера имеют тенденцию быть схожими по определенному признаку, а объекты из различных групп отличаются. Двумя наиболее известными методами, применяемыми в кластерном анализе, является метод k-средних и самоорганизующиеся карты.

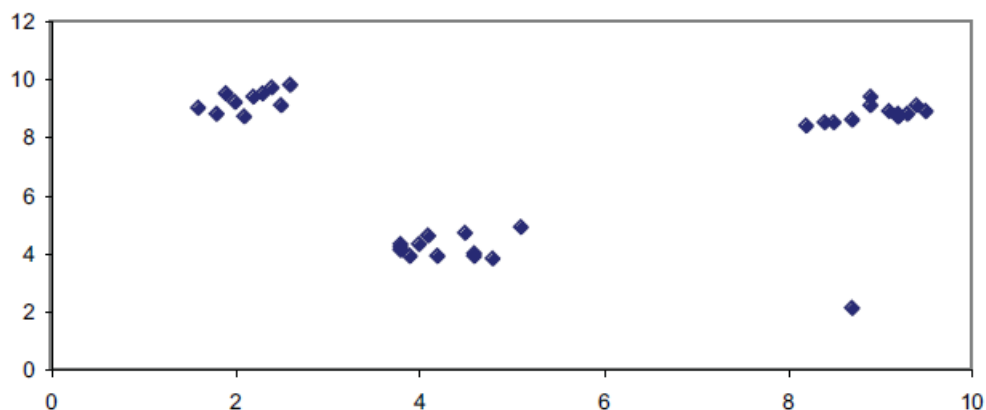
Кластеризация может быть произведена на основе Евклидовых расстояний, вычисляемых по одной или нескольким количественным переменным.

Наблюдения распределяются по кластерам таким образом, чтобы каждое наблюдение принадлежало только одному кластеру.

Самоорганизующиеся карты повторяют способ, с помощью которого различные человеческие сенсорные сигналы отображаются на соответствующие области коры головного мозга так, чтобы зависимости между входными стимулами в пространстве соответствовали зависимостям между нейронами (т.е. кластерами) в пространстве.

На рисунке 4.10 приведена кластеризация на основе двух переменных. Он показывает точки (наблюдения), образующие три отдельных кластера или группы. В нижнем правом углу можно увидеть выброс.

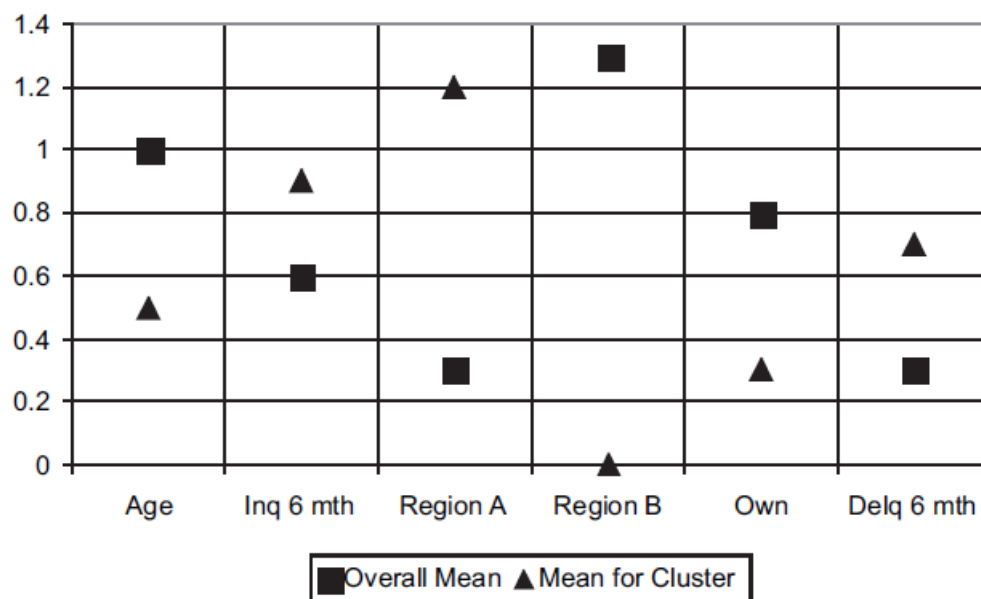
Рисунок 4.10 Кластеризация



Результаты, полученные по одному кластеру с использованием данной техники, приведены на рисунке 4.11, где ось y – это нормализованное среднее. Данный кластер имеет следующие характеристики:

- Возраст ниже среднего
- Количество запросов за последние шесть месяцев выше среднего
- Тенденция проживания в регионе А
- Нет клиентов, постоянно проживающих в регионе В
- Меньшая вероятность наличия собственного жилья
- Большая вероятность просрочек за последние шесть месяцев

Рисунок 4.11 Характеристики кластера



С помощью аналогичных графиков можно проанализировать остальные кластеры по определенным характеристикам. Дальнейший анализ распределения характеристик внутри каждого кластера выполняют для того, чтобы получить набор критериев, позволяющих выделить уникальные группы. Для примера, приведенного на рисунке 4.11, можно сформировать следующие кластеры:

- Молодые владельцы жилья, проживающие в регионе А
- Молодые съемщики жилья, проживающие в регионе А
- Владельцы жилья более старшего возраста, проживающие в регионе А
- Съемщики жилья более старшего возраста, проживающие в регионе А

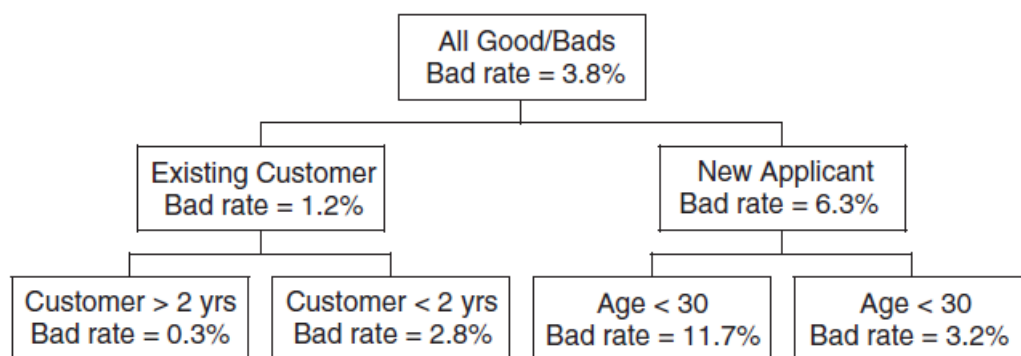
Также кластеризация может быть на основе других характеристик, дающих наилучшую дифференциацию между кластерами.

Следует отметить, что кластеризация определяет группы, которые схожи по своим характеристикам, а не кредитоспособность. Таким образом, кластеры могут показаться отличающимися друг от друга, однако все они имеют схожую кредитоспособность. Поэтому необходимо провести дальнейшее исследование кластеров, используя, например, анализ численности «плохих» клиентов. Это гарантирует, что полученная сегментация выделяет группы с различными профилями кредитоспособности.

Деревья решений. Еще один способ выполнения статистической сегментации - деревья решений. Деревья решений выделяют сегменты на основе характеристик кредитоспособности (то есть разделяют клиентов на «хороших» и «плохих»). Они понятны и их легко интерпретировать. Помимо выявления характеристик для сегментации деревья решений также определяют оптимальные точки разбиения для каждой характеристики. Поэтому они представляют очень эффективный и удобный метод сегментации. Пример, приведенный на Рисунке 4.12, показывает сегментацию на основе двухуровневого дерева решений.

В результате мы получаем четыре возможных сегмента для данного портфеля, основанных на критериях: существующий/новый клиент, длительность кредита и возраст.

Рисунок 4.12 Сегментация на основе деревьев решений



Сравнительный анализ улучшений

И эвристические и статистические методы позволяют сформулировать идеи по поводу возможной сегментации и обосновать причины разбиения на группы. Однако они не могут количественно оценить преимуществ сегментации. Существуют довольно простые доступные способы для того, чтобы оценить, насколько эффективна сегментация.

Первый шаг состоит в измерении предсказательной силы модели благодаря сегментации, которая выражается статистическими показателями, такими, как критерий Колмогорова-Смирнова, метод с-статистики и т.д. Рисунок 4.13 показывает пример такого анализа с использованием с-статистики (подробно о с-статистике говорится в Главе 6).

Рисунок 4.13 Сравнительное улучшение за счет сегментации

<i>Segment</i>	<i>Total c-stat</i>	<i>Seg c-stat</i>	<i>Improvement</i>
Age < 30	0.65	0.69	6.15%
Age > 30	0.68	0.71	4.41%
Tenure < 2	0.67	0.72	7.46%
Tenure > 2	0.66	0.75	13.64%
Gold Card	0.68	0.69	1.47%
Platinum Card	0.67	0.68	1.49%
Unsegmented	0.66	—	—

На рисунке приведены различные способы расчета с-статистики для несегментированной скоринговой карты и шести сегментированных скоринговых карт. «Общая с-статистика» - с-статистика для выделенной группы на основе несегментированной скоринговой карты. «Сегментированная с-статистика» - с-статистика для выделенной группы на основе скоринговой карты, построенной специально по данному сегменту. В любом случае сегментированные скоринговые карты обладают лучшей предсказательной силой, чем общая скоринговая карта. Сегментация платиновых/золотых карт не приведет к существенному улучшению. Вероятно, потому что кредитор не проводил значимых различий между клиентами золотых и платиновых карт. Наилучшими сегментами, похоже, будут сегменты на основе срока кредита, поскольку это наиболее надежная информация о поведении клиентов из имеющейся. Данный анализ проводится на основе одноуровневой сегментации. Аналогичный анализ можно провести и для более сложной сегментации. Затем пользователю необходимо решить, какой уровень улучшения является достаточно значимым для разработки и внедрения.

Решая данный вопрос, лучше всего использовать не статистические, а бизнес-показатели. Для бизнеса не важно максимальное значение с-статистики или КС. Бизнес интересуют индикаторы кредитоспособности: процент одобренных заявок, размер прибыли, размер убытков и т.д. Поэтому было бы полезно преобразовать улучшенную предсказательную способность в ожидаемую эффективность кредитного портфеля, как показано на Рисунке 4.14.

На рисунке приводятся два общеизвестных показателя кредитоспособности, а именно процент одобренных заявок и ожидаемый процент «плохих» клиентов по каждой сегментированной скоринговой карте. Также на рисунке приведена информация о приблизительном размере сегмента. С помощью приведенного на рисунке примера можно решить, достаточен ли размер сегмента и улучшение кредитоспособности, чтобы внедрить дополнительную скоринговую карту. Сравнение процента одобренных заявок и процента «плохих» клиентов должно основываться на некоем ожидаемом значении кредитоспособности. Например, если планируемый процент одобренных заявок составляет 70%, то и проценты «плохих» клиентов должны быть рассчитаны для данного значения. В приведенном примере сегментация по сроку кредита дает улучшение с т.з. уменьшения процента «плохих» клиентов: для заемщиков со сроком кредита менее 2 лет – с 5.7% до 4.2%, для заемщиков со сроком кредита более 2 лет – с 2.9% до 2.1%. Используя данную информацию, можно выдвинуть некоторые идеи по поводу сокращения убытков.

Рисунок 4.14 Оценка преимуществ сегментации

<i>Segment</i>	<i>Size</i>	<i>After Segmentation</i>		<i>Before Segmentation</i>	
		<i>Approve</i>	<i>Bad</i>	<i>Approve</i>	<i>Bad</i>
Total	100%	70%	3.5%	70%	4.1%
Age < 30	65%	70%	4.1%	70%	6.3%
Age > 30	35%	70%	2.7%	70%	3.6%
Tenure < 2	12%	70%	4.2%	70%	5.7%
Tenure > 2	88%	70%	2.1%	70%	2.9%
Gold Card	23%	70%	3.9%	70%	4.3%
Platinum Card	77%	70%	3.1%	70%	3.8%

Отбор сегментов

Однако почему же на практике специалисты не спешат внедрять все разработанные ими скоринговые карты, которые обладают лучшей предсказательной способностью? Причины могут быть следующие:

- **Стоимость разработки.** Это включает работы (внутренние и внешние) для разработки скоринговой карты с полной документацией.
- **Стоимость внедрения.** Дополнительные расходы связаны с внедрением скоринговой карты, особенно если при этом используются нестандартные характеристики или необходимо написать сложный программный код.
- **Обработка.** Чем больше скоринговых карт применяется, тем больше дополнительные затраты, связанные с их обработкой.
- **Развитие стратегии и мониторинг.** Каждая скоринговая карта включает в себя набор связанных стратегий или правил, а также отчетности по ее работе. Управление и мониторинг большого количества скоринговых карт требуют дополнительных ресурсов.

Для больших кредитных портфелей эти затраты и усилия незначительны (вышеуказанные затраты окупаются довольно быстро). Однако при меньших

портфелях и в небольших финансовых организациях такой анализ может быть достаточно трудоемким, что, возможно, не окупит дополнительных усилий, сложности работ и понесенных затрат.

Методология

Существуют различные математические методы, применяемые для построения скоринговых карт, предсказывающих риск – например, логистическая регрессия, нейронные сети, деревья решений и т.д. Предпочтение в выборе используемого метода может зависеть от следующих факторов:

- Качество имеющихся данных. Деревья решений будут более предпочтительны в ситуации, когда есть достаточный объем пропущенных данных или зависимость между характеристиками (предикторами) и целевой переменной является нелинейной.
- Тип целевой переменной – бинарная («хороший»/«плохой») или непрерывная (прибыль/потери).
- Объем выборки.
- Платформа внедрения (т.е. позволяет ли существующая система рассмотрения кредитных заявок внедрить предложенный тип скоринговой карты). Например, модель нейронной сети может быть идеальной, но неприменимой на практике, если ее невозможно внедрить на базе существующей системы.
- Интерпретируемость результатов, например, удобство применения скоринговых карт с использованием баллов, полученных на основе регрессии.
- Соответствие методологии требованиям прозрачности и ясности, обычно выдвигаемым местными органами надзора.
- Возможность мониторинга и диагностики работы скоринговой карты.

Здесь также можно обсудить вопросы масштабирования и структуры скоринговой карты (например, возможные скоринговые диапазоны, количество баллов, которое удваивает шансы, при том, что скоринговая оценка сама по себе представляет ожидаемый процент «плохих» и т.д.). Метод и предложенный формат скоринговой карты должен быть обсужден со специалистами отдела рисков и IT отдела, чтобы тщательно рассмотреть вопросы, связанные с использованием данных и теоретическими аспектами методов. Таким образом, результаты скоринговой карты будут правильно интерпретированы и карту после разработки можно будет внедрить.

Экспертиза плана внедрения

Дополнительная информация, полученная на данной стадии, возможно, потребует внести изменения в исходный план внедрения и временные рамки проекта. Особенно, если необходимое число скоринговых карт после проведения сегментации больше, чем ожидалось ранее, то данная методология становится трудоемкой и требует внесения изменений в платформу внедрения, ужесточаются требования к данным, а проекту нужно больше времени. Для получения реалистичных прогнозов необходимо провести экспертизу тестирования и плана внедрения. Это важно для тех компаний, в которых обязанности по разработке, тестированию, внедрению и мониторингу скоринговых карт возложены на различные отделы. Руководитель проекта должен убедиться в том, что изменения обоснованы и их влияние на исходный план проекта оценено количественно, чтобы стадии проекта последовательно сменяли друг друга.

Теперь, когда выдвинуты требования к данным и подготовлена документация к плану проекта, можно начинать проектирование базы данных.

Процесс Разработки Скоринговой Карты, Этап 3: Создание Базы Данных для Разработки Модели

Теперь можно начинать создание базы данных, следуя параметрам, заданным в Главе 2. База данных будет включать набор характеристик (или предикторов) и целевую переменную, который затем используется для разработки скоринговой карты.

Формулирование требований к обучающей выборке

После того, как определены параметры, разбиение на сегменты и методология проекта, необходимо сформулировать требования к созданию базы данных (на основе обучающей выборки), используемой на этапе построения модели. Основываясь на результатах предыдущих этапов, сформулированы и задокументированы следующие требования:

- Количество требуемых скоринговых карт и формулирование требований по каждому сегменту, включая инструкции по кодировке, как идентифицировать различные сегменты
- Определения «плохого», «хорошего» и «серой зоны»
- Процент «плохих» клиентов в кредитном портфеле и процент одобрения для каждого сегмента
- «Окно созревания» и «окно выборки»
- Определение исключительных случаев

Кроме того, на данном этапе сформулированы и добавлены следующие требования:

- Размер выборки, необходимый для каждого сегмента и класса кредитоспособности (включая «отклоненных» для аппликационных скоринговых карт)
- Полученный из внутренних и внешних источников подробный перечень характеристик, необходимый для каждого сегмента обучающей выборки
- Расчетные характеристики

Выбор характеристик

Выбор характеристик, которые будут включены в обучающую выборку, - это важная часть процесса разработки. Данный этап, в ходе которого проводится осторожный отбор характеристик, подтверждает необходимость делового подхода, применяемого на каждом этапе разработки скоринговой карты. Альтернативный вариант – импорт моментальных «снимков» витрин данных или других хранилищ данных в базу данных, применяемую для разработки

скоринговой карты – неэффективен и в отличие от выбора характеристик не улучшает понимание внутренних данных. Характеристики отбираются предварительно для улучшения процесса разработки. Их нужно выбирать на основе различных факторов, включая:

- **Ожидаемая предсказательная сила.** Данную информацию получают на основе совместного опыта (специалистов по взысканию просроченной задолженности и риск-аналитиков), предыдущего анализа и проектов по разработке скоринговой карты. Здесь в значительной степени может помочь интервью с кредитными аналитиками и специалистами по взысканию долгов.
- **Надежность и устойчивость.** Некоторые неподтвержденные данные могут быть подделаны или подвержены подтасовке (например, доход), особенно в тех случаях, когда ввод данных осуществляется персоналом, заинтересованным активнее продавать продукт, например, служащие филиалов банков или кредитные брокеры. В ряде случаев достаточно накладно проверять такие данные (например, низкоприбыльные кредитные продукты) и поэтому подобного рода данные не следует включать.
- **Легкость сбора данных.** Также нужно исключить пункты анкеты заявителя, которые заполняются по желанию (и поэтому пропущены заявителями), или рассматривать для скоринговых карт только, когда эти пункты носят обязательный характер.
- **Интерпретируемость.** Некоторые характеристики, такие, как профессия и тип отрасли, подвержены субъективной интерпретации. Различные специалисты могут отнести одного и то же человека к различной профессии или типу отрасли, а по мере роста необычных профессий становится трудно распределять людей по традиционным профессиям. Данный факт служит причиной того, что в большинстве организаций категория «другое» часто составляет 75% от всех наблюдений по характеристике «профессия». Даже в тех случаях, когда подобного рода характеристики показали, что могут использоваться для прогнозов, проблемы будущей интерпретации могут поставить под сомнение их устойчивость. Исключение может быть сделано, когда субъективная интерпретация подтверждается практикой кредитного риска. Пример вышесказанного – такое поле, как «качество управления», которое может использоваться в скоринговых картах для малого и среднего бизнеса (SME). Несмотря на то, что оно является субъективным, оценка ставится (или должна ставиться) опытными кредитными аналитиками и позволяет использовать данное поле с большей степенью уверенности.
- **Вмешательство человека.** Это относится к характеристикам, на которые значительно повлияло вмешательство человека (например, индикаторы банкротства следует исключать в тех случаях, когда банкроты были отклонены правилами кредитной политики). Несмотря на то, что анализ отклоненных заявок может в какой-то степени исправить данную ситуацию, правила кредитной политики и характеристики скоринговой карты должны быть независимыми друг от друга.

- **Правовые проблемы, связанные с использованием определенных типов информации.** Если есть характеристики, которые собраны в течение исторического периода (например, семейный статус, пол и т.д.), но по правовым и политическим соображениям сомнительны, лучше их исключить из разработки скоринговой карты.
- **Создание коэффициентов на основе деловых соображений.** Пользователям необходимо избегать метода «ковровых бомбардировок». Он использует всевозможные переменные в наборе данных, деля их на все остальные и получая перечень коэффициентов, которые хоть и могут обладать предсказательной силой, но не поддаются объяснению. Любой созданный коэффициент необходимо обосновать. Например, отношение количества запросов в кредитное бюро за последние 3 месяца к количеству запросов в кредитное бюро за последние 12 месяцев дает представление о краткосрочной потребности в кредите в сравнении с долгосрочной. Аналогично кратко- и долгосрочные сравнения других индикаторов, таких, как покупки, платежи, утилизация, баланс, обязательства/доход и др. доказали, что могут быть хорошими индикаторами риска – благодаря тому, что хорошо согласуются с деловой практикой.
- **Возможность сбора данных в будущем.** Убедиться в том, что любая переменная, рассматриваемая в ходе разработки скоринговой карты, будет собрана в будущем.
- **Изменения конкурентной среды.** Характеристики, которые в данный момент можно не рассматривать в качестве обладающих предсказательной силой, могут стать таковыми в будущем в силу таких трендов, как увеличение кредитных линий или появление новых продуктов.

Теперь известно, что скоринговая карта будет разработана на основе данных за последние два-три года, и ее планируется использовать приблизительно в течение следующих двух лет. Поэтому на данный момент нужно принять во внимание прошлые и ожидаемые в будущем тренды. Один из способов сделать это – это обратиться в кредитное бюро по поводу того, как изменилась информация за последние два-три года – изменились ли значительно за последние несколько лет такие индикаторы, как баланс, количество кредитных линий, утилизация и являются ли данные изменения восходящим или нисходящим трендом. Несмотря на то, что это не повлияет на обучающие данные, можно регулировать ожидаемые значения и разрабатывать соответствующие стратегии. Например, рост конкуренции между банками увеличит среднее количество запросов. Скоринговые карты, разработанные по историческим данным, будут рассматривать количество запросов больше четырех за 12 месяцев в качестве индикатора высокого риска на основе прошлой информации о кредитоспособности. Однако новые данные могут указать на то, что заявители, имеющие четыре запроса, теперь имеют нормальную, среднерисковую кредитоспособность. Поэтому есть вариант, либо искусственно изменить распределение скоринговых баллов для запросов, либо скорректировать порог отсека так, чтобы принять во внимание изменения тренда. По крайней мере, это дает понимание данной проблемы, которую таким образом можно скорее учесть, чем проигнорировать.

Объем вопросов, затронутых в рамках данной задачи, еще раз подчеркивает необходимость сотрудничества и консультаций разработчиков скоринговых карт с остальными участниками команды.

Формирование выборки

Формирование выборки в ходе разработки скоринговой карты обусловлено двумя задачами, а именно, разбиением общей выборки на обучающий и тестовый наборы данных и принятие решения о доле «хороших», «плохих» и отклоненных для включения в итоговую выборку.

Обучающая выборка/Тестовая выборка

Есть различные методы разбиения выборки на обучающий (выборка, на основе которой скоринговая карта разрабатывается) и тестовый наборы данных (выборка, на основе которой скоринговая карта проверяется). Обычно 70-80% выборки используется для разработки скоринговой карты, оставшиеся 20-30% исключаются из разработки и затем используется для тестирования или валидации скоринговой карты. Когда размеры выборки являются небольшими, скоринговая карта может быть разработана на 100% выборки и проверена на нескольких случайных выборках, содержащих от 50% до 80% наблюдений в каждой.

«Хорошие»/ «Плохие»/Отклоненные

Как упоминалось ранее, обычно около 2000 «хороших», «плохих» и «отклоненных» в каждой категории достаточно для разработки скоринговой карты. Метод называется оверсемплинг и широко применяется в отрасли. Методы корректировок оверсемплинга (освещается ниже в данной главе) применяются позже, чтобы получить реалистичный прогноз. Дополнительное преимущество достаточно большой выборки заключается в том, что она снижает эффект мультиколлинеарности и дает статистически значимые результаты логистической регрессии.

Можно также использовать пропорциональное формирование выборки, пока есть достаточное количество «хороших» и «плохих» для обеспечения статистической достоверности. Использование пропорционального формирования выборки, например, для кредитного портфеля с процент «плохих» 4% потребует обучающей выборки, содержащей 4% «плохих» (например, 4000 «плохих» и 96000 «хороших»). Нет необходимости корректировать данный набор данных в соответствии с априорными вероятностями, поскольку выборка уже отражает фактические вероятности.

Есть различные статистические техники, позволяющие определить оптимальные размеры выборки на основе количества рассматриваемых характеристик и их изменчивости.

Один из методов – это анализ мощности (например, используя PROC POWER/GLMPower/PSS в SAS). Как вариант, можно использовать один из двух методов расчета:

Предположим, вы хотите на 95% быть уверенным, что среднее значение переменной в фактической популяции находится в заданных пределах оценки среднего, вычисленной по выборке – например, вы хотите на 95% быть уверенным, что среднее расстояние до места работы в фактической популяции находится в пределах 10 миль от выборочной оценки.

Дано: Z – z-статистика для уровня значимости (например, 1.96 для уровня значимости 0.05 или уровня надежности 95%)
 σ – стандартное отклонение популяции
 d – половина ширины доверительного интервала (точность оценки)

Значит: $n = (Z \sigma / d)^2$

В предыдущем примере, если вы оцениваете σ , равным 15, значит $n = (1.96 \cdot 15 / 10)^2 \approx 9$. То есть вам необходима выборка 9 человек. Если вы оцениваете σ , равным 30, значит $n = (1.96 \cdot 30 / 10)^2 \approx 35$.

С помощью второго метода, зная долю признака в популяции, задав z-статистику, определяемую в зависимости от уровня значимости, и ширину доверительного интервала (точность оценки), вы можете найти минимальный объем выборки.

Дано: Z – z-статистика для уровня значимости (например, 1.96 для уровня значимости 0.05 или уровня надежности 95%)
 p – доля признака в популяции
 d – половина ширины доверительного интервала для доли признака (точность оценки доли признака)

Значит: $n = [Z^2 p(1-p)] / d^2$

Например, предположив наихудший сценарий 50/50 (задающий максимально возможный размер выборки) и точность оценки 5% или 0.05,
 $n = [1.96^2 \cdot 0.5(1-0.5)] / 0.05^2 \approx 385$

Однако нужно помнить, что эти методы дают минимальные размеры выборки. Пользователи должны придерживаться такого метода формирования выборки, который считают наиболее обоснованным до тех пор, пока размеры выборки позволяют убедиться в удовлетворительных статистических и практических результатах. Когда разрабатываются скоринговые карты, использующие сгруппированные характеристики (о чем детально рассказано в этой книге), конечная цель любого применяемого метода заключается в том, чтобы каждая группа имела достаточное количество «хороших» и «плохих» для получения содержательных результатов анализа. Это подразумевает вычисление WOE для отдельных групп, а также построение WOE трендов по атрибутам.

Сбор данных и создание обучающей выборки

Основываясь на требованиях к обучающей выборке, данные по заявителям (или клиентам для поведенческих скоринговых карт) и характеристикам их кредитоспособности собираются из различных источников (например, внутренние базы данных, данные кредитное бюро, запросы к хранилищам, базы недвижимости и т.д.). Если некоторые требуемые данные хранятся в неэлектронном формате, они должны быть подготовлены для перевода в формат компьютерных данных. Конечный результат данной процедуры – набор данных, который содержит выбранные описательные переменные (характеристики) и целевую переменную (индикатор «хорошего»/«плохого» поведения).

Случайность и репрезентативность

Отбор заявителей и клиентов должен быть случайным, отражать тот сегмент, для которого должна быть разработана скоринговая карта (например, должна быть

репрезентативна для заявителей, которые в будущем получают скоринговые баллы), и не должен включать определенные типы клиентов. Набор данных, смещенный по определенному региону или возрастной группе, не может быть эффективен для анализа общей популяции заявителей, поскольку может недооценивать другие сегменты или создавать видимость, что они являются менее рискованными.

Несегментированный набор данных

Для проектов, в рамках которых разрабатывается более одной скоринговой карты, необходимо сформировать отдельный набор данных для каждого сегмента, а также набор данных, который является несегментированным.

Несегментированный набор данных используется для анализа измерения любого дополнительного «прироста информации» или преимущества при использовании сегментированных скоринговых карт, таких, какие показаны на рисунке 4.14.

Неожиданные изменения данных

В ходе сбора данных пользователь должен быть осведомлен о наличии исторических изменений, произошедших в базах данных, особенно в промежутки времени, приходящийся на «окно выборки» и после него. Большинство аналитиков, которые имели дело с финансовыми данными, осведомлены о ситуациях, когда форматы или процедуры сбора данных изменились – например код списания долга изменился с «W» на «C»; жилищный статус заносился до декабря в поле 32, а теперь в поле 22; «другие кредитные карты», начиная с прошлого марта, больше не заносится; классификация профессий изменилась четыре месяца назад и т.д. Информация о подобных изменениях гарантирует, что программный код или запросы, написанные для построения выборки, обрабатывают данные согласно заданным требованиям, и обеспечивают «отсутствие сюрпризов».

В большинстве организаций такие неожиданные изменения данных не задокументированы. Это позволяет специалистам по data mining оставлять эти данные на произвол судьбы, надеясь на свою память или память тех, кто уже проработал в организации в течение какого-то периода времени. Ясно, что подобные ситуации не способствуют стабильности и эффективности. В идеале, каждая база данных в организации должна иметь *Журнал изменения данных*, который фиксирует все изменения в базе данных, начиная с самого начала. Изменения, приведенные в предыдущем параграфе, не будут больше неожиданными, поскольку специалист, работающий с базами, имеет сведения о модифицированных данных.

Поправка на априорные вероятности

Оверсемплинг – это стандартная практика для большинства процедур прогнозного моделирования и применяется в ситуациях, когда доля «хороших» и «плохих» в обучающей выборке отличается от доли «хороших» и «плохих» в фактической популяции. Оверсемплинг также известен как смещенное, основанное на выборе, стратифицированное, зависимое от отклика формирование выборки.

В подобных ситуациях обучающую выборку нужно будет скорректировать с учетом априорных вероятностей. Этот метод, также известный как «разложение на множители», корректирует частоты наблюдений в обучающей выборке таким образом, что процент «плохих» (или процент одобренных заявок) выборки отражает процент «плохих» в фактической популяции. Он полезен как в процессе разработки, так и для составления управленческих отчетов по скоринговым

картам, таких, как График выигрышей. Корректировка оверсемплинга дает более реалистичные прогнозы и понимание кредитоспособности клиентов по группировкам атрибутов – таким образом, предоставляя ценное бизнес-преимущество, а также выгоды с т.з. статистического анализа. Корректировка оверсемплинга также необходима для проведения валидации модели и получения таких показателей эффективности, как статистика Колмогорова-Смирнова, статистика, коэффициент Джини и т.д.

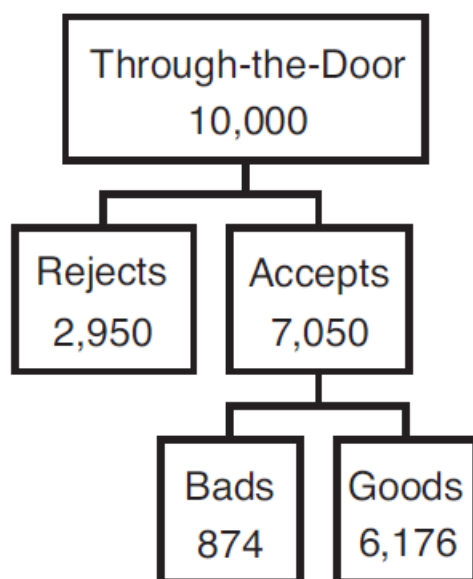
Поэтому корректировка полезна, когда необходимо получить прогнозы. Она не является обязательной, когда цель моделирования – исследование взаимосвязей между предикторами и целевой переменной или когда требуется лишь ранжирование скоринговых баллов (т.е. когда более высокий балл должен обозначать более низкий риск, но необязательно определенный процент «плохих»). Можно с уверенностью предположить, что данная корректировка потребуется для большинства разрабатываемых скоринговых карт. Кредитный скоринг используется для получения практических решений, проведения специальных вычислений, назначения порогов отсека и поэтому требует, чтобы зависимость между скоринговым баллом и процентом «плохих» была известна.

Например, если обучающая выборка содержит по 2000 «хороших», «плохих» и «отклоненных», процент «плохих» и процент одобрения в выборке будет 50% и 67%. Предположим, что фактический процент одобрения равен 70.5%, а процент «плохих» в фактической популяции равен 12.4%. На рисунке 5.1 показана выборка, разложенная применительно к популяции «с улицы» в 10000 наблюдений.

Несмотря на то, что корректировка оверсемплинга может быть выполнена до или после подгонки модели, когда разрабатываются скоринговые карты на основе сгруппированных атрибутов, полезно проводить ее перед процедурой группировки. Это обусловлено тем, что взаимосвязи между предикторами и целевой переменной могут быть спрогнозированы лучше – реалистичное распределение процента «плохих» и процента одобрения по атрибутам может дать информацию о том, искусственно ли повлияли на кредитоспособность правила кредитной политики или оценки, сделанные вручную (например, если известный негативный атрибут имеет низкий процент одобрения и низкий процент «плохих», или если известный положительный атрибут имеет высокий процент «плохих»). Он также проводит проверку разумности созданных группировок, чтобы убедиться – группы достаточно сильно отличаются друг от друга по кредитоспособности.

Существует два главных метода корректировки оверсемплинга: метод смещения и метод выборочных весов.

Рисунок 5.1 **Корректировка уравновешенной выборки для моделирования популяции «с улицы» в 10000 наблюдений**



Метод смещения

Стандартная логистическая функция, использованная в регрессии, выглядит:

$$\text{Logit} (p_i) = \beta_0 + \beta_1 x_1 + . . . + \beta_k x_k$$

где p_i , β_k и x_k являются апостериорной вероятностью, регрессионными коэффициентами и переменными соответственно.

Для оверсемплинга логиты смещаются с помощью поправки, $\ln (\rho_1 \pi_0 / \rho_0 \pi_1)$ и тогда новая логистическая функция (псевдомодель) становится:

$$\text{Logit}(p^*_i) = \ln \left(\frac{\rho_i \pi_0}{\rho_0 \pi_1} \right) + \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$$

где ρ_1 и ρ_0 – доли классов в выборке, π_1 и π_0 – доли классов в популяции.

Вышеприведенный метод является корректировкой, применяемой перед тем, как модель подогнана. Скорректированная апостериорная вероятность может быть вычислена аналогичным образом после подгонки модели, используя:

$$p^{\wedge}_i = \frac{(p^{\wedge}*_i \rho_0 \pi_1)}{[(1 - p^{\wedge}*_i) \rho_1 \pi_0 + p^{\wedge}*_i \rho_0 \pi_1]}$$

где $p^{\wedge}*_i$ - нескорректированная оценка апостериорной вероятности.

Обе эти корректировки могут быть выполнены в SAS Enterprise Miner с использованием вектора априорных вероятностей или, как вариант, в SAS PROC LOGISTIC с использованием настройки Смещение, как показано на рисунке 5.2. Заметим, что в обоих случаях «поправка» – это $\ln (\rho_1 \pi_0 / \rho_0 \pi_1)$.

Метод выборочных весов

При корректировке оверсемплинга с использованием выборочных весов каждое наблюдение умножается на установленный вес, чтобы выборка отражала фактическую популяцию. Весами данного наблюдения являются p_1/p_0 и p_0/p_1 для классов 1 и 0 соответственно.

Как вариант, вес каждого «плохого» может быть установлен, равным 1, а вес каждого «хорошего» устанавливается как соотношение p («хороший»)/ p («плохой»), где p («хороший») и p («плохой») – вероятности «хороших» и «плохих» в фактической популяции. Например, если процент «плохих» – т.е. p («плохой») – в кредитном портфеле равен 4%, а обучающая выборка содержит 2000 «хороших» и 2000 «плохих», то скорректированная выборка будет включать 2000 «плохих» и 48000 «хороших», т.е. $(0.96/0.04) \times 2000$.

Данный метод можно применить, используя настройку «sampwt» в программном коде SAS или переменную частоты в SAS Enterprise Miner. Пример, использующий программирование на языке SAS, показан на рисунке 5.3.

Следует отметить, что итоговые оценки параметров регрессии при использовании двух вышеперечисленных методов корректировки будут другими.

В общем, когда логистическая модель задана верно, то лучше использовать метод смещения. Когда логистическая модель является приближением некоей нелинейной модели, лучше использовать метод весов. Разрабатывая прогнозные модели на основе несгруппированных переменных, специалисты-статистики предпочитают использовать метод смещения. Однако в случае сгруппированных переменных, скоринговых карт на основе баллов, метод весов лучше, т.к. вместо корректировки лишь спрогнозированных вероятностей он корректирует оценки параметров, использованных для получения баллов. Нормализация, которая достигается при использовании выборочных весов, также уменьшает искажения p -value и стандартных ошибок.

Рисунок 5.2 Программный код SAS для корректировки оверсемплинга с помощью метода смещения

<u>Premodel adjustment</u>	<u>Postmodel adjustment</u>
Data develop;	Proc logistic data=develop ...;
Set develop;	Run;
Off=(offset calc);	Proc score ... out=scored...;
Run;	Run;
Proc logistic data=develop ...;	Data scored; set scored;
Model ins=...../ offset=off;	Off = (offset calc);
Run;	$p = 1 / (1 + \exp(-(ins - off)))$;
Proc score;	Run;
$p = 1 / (1 + \exp(-(ins - off)))$;	Proc print data=scored ..;
Proc print;	Var p
Var p;	Run;
Run;	

**Рисунок 5.3 Программный код SAS для корректировки
оверсемплинга с помощью метода выборочных весов**

```
Data develop;  
Set develop;  
sampwt=(  $\pi_d / \rho_o$  ) * (ins=0) + (  $\pi_d / \rho_i$  ) * (ins=1);  
Run;
```

```
Proc logistic data=develop ...;  
Weight=sampwt;  
Model ins=.....;
```

```
Run;
```

Процесс Разработки Скоринговой Карты, Этап 4: Разработка Скоринговой Карты

После того, как создание базы данных завершено, разработчик скоринговой карты должен подготовить базу данных с набором характеристик (предикторов) и целевой переменной. Существует различные методы, которые можно использовать для разработки скоринговой карты на основе полученной информации. Все они включают выявление и квантификацию взаимосвязей между характеристиками и «хорошим»/«плохим» поведением (целевой переменной).

Данная глава посвящена исключительно разработке модели с использованием сгруппированных атрибутов и применению метода логистической регрессии. Мы также осуществим анализ отклоненных заявок (reject inference), и скоринговая карта получит отмасштабированные баллы. Результатом станет скоринговая карта, представленная на рисунке 1.1. Этот метод уравнивает два ключевых требования к успешной разработке скоринговой карты: необходимость убедительной статистической основы и бизнес-ориентированность.

Рисунок 6.1 иллюстрирует последовательность построения скоринговой карты, которая вытекает из применения данной методологии. Заметим, что она актуальна для аппликационной скоринговой карты. Поведенческие скоринговые карты используют те же самые процедуры, за исключением анализа отклоненных заявок. Поэтому для поведенческих скоринговых карт нужно провести анализ исходных характеристик, а затем сразу приступить к финальному этапу разработки скоринговой карты.

Рисунок 6.1 Этапы разработки скоринговой карты



Исследование данных

Исследование выборочных данных является хорошей практикой перед началом моделирования. Такие простые статистики, как распределения значений, среднее/медиана, доля пропущенных значений и ранжирование значений каждой характеристики может дать более глубокое понимание бизнеса, анализ этих статистик – хороший способ проверить целостность данных. Часто эти задачи очень хорошо решает визуальное представление данных. Распределения выборочных данных необходимо сравнить с распределениями кредитного портфеля в целом, чтобы убедиться в том, что выборка репрезентативна для кредитного портфеля. Данные должны быть проверены на возможность интерпретации (например, убедиться, что «0» означает ноль, а не пропущенные значения) и подтвердить, что зафиксированы специальные значения, такие, как 99 или 999. Кроме того, этот шаг позволяет убедиться, что данные собраны согласно заданным требованиям, все свойства данных изучены, включая странности поведения данных.

Пропущенные значения и выбросы

Большинство финансовых данных содержит пропущенные значения или значения, которые не имеют смысла для отдельных характеристик. Они могут быть либо полями, которые не заполнены оператором, неправильно введены, пропущены заявителем, либо простыми выбросами, обозначающими крайние случаи.

Несмотря на то, что некоторые статистические техники, такие, как деревья решений, могут работать с пропущенными значениями, логистическая регрессия требует полные наборы данных без пропущенных значений (анализ полных наблюдений). Есть четыре основных способа обработки пропущенных значений:

1. Исключить все данные с пропущенными значениями – это анализ полных наблюдений, и, вероятно, результатом будет очень маленький набор данных для работы.
2. Исключить из модели характеристики или наблюдения, которые имеют значительное число пропущенных значений (более 50%), особенно, если ожидается, что пропуски продолжат появляться в будущем.

3. Включить характеристики с пропущенными значениями в скоринговую карту. «Пропущенные» могут быть обработаны как отдельные атрибуты, сгруппированы и использованы в регрессионном анализе в качестве входной характеристики. В скоринговой карте можно назначить веса данному атрибуту. В некоторых случаях этот вес может быть близок к «нейтральному» или среднему значению, но в тех случаях, когда вес близок к значению другого атрибута, это может пролить свет на истинную причину пропусков.
4. Импутировать пропущенные значения с помощью статистических техник.

Несмотря на то, что второй способ может оказаться более подходящим, способ 3 дает много преимуществ. Способы 1, 2 и 4 предполагают, что пропуски не несут содержательный смысл – дополнительная информация не может быть получена из анализа пропущенных данных. Это необязательно верно – пропущенные данные могут быть частью тренда, связанного с другими характеристиками, являться индикатором «плохого» заявителя. Пропуски не являются обычно случайными. Например, те, кто недавно начал работать, с большей вероятностью оставят пустым поле «Стаж работы» в анкете. Если характеристики или записи с пропущенными значениями исключаются, ни одно из этих предположений не может приниматься в расчет. Поэтому рекомендуется включать пропущенные данные в анализ и присваивать им баллы в скоринговой карте. Данный метод допускает, что пропущенные данные содержат какие-то значения IV и таким образом достигается бизнес-преимущество за счет включения таких данных в анализ. Кроме того, присвоение весов пропущенным значениям облегчит скоринг заявителей, которые в будущем будут оставлять поля в анкете пустыми. В любом случае, пропущенные данные должны быть сперва проанализированы и, если обнаружено, что они являются случайными или не влияют на определение статуса кредитоспособности, могут быть исключены или импутированы.

Некоторое программное обеспечение data mining, такое, как SAS Enterprise Miner, имеет алгоритмы импутации пропущенных данных. Подобного рода алгоритмы включают импутацию на основе деревьев решений и замену средним или медианой. Для разработки скоринговой карты рекомендуются методы импутации, которые учитывают значения других характеристик и записей. Присвоение пропущенным значениям «самых часто встречающихся» или «средних» значений может привести к появлению пиков в данных, и дифференциация между данными с присвоенными средними значениями и данными с реальными средними значениями будет невозможна – таким образом, бизнес-информация может потеряться. Также полезно присвоить специальные значения пропущенным данным (например, 99 или 999, лежащие вне диапазона обычных значений) и исключить их из анализа.

Выбросы – это значения, которые выпадают из нормального диапазона значений для определенной характеристики. Например, распределение возраста охватывает все население в диапазоне от 18 до 55, включая несколько наблюдений – 99, 112 и 134. Несмотря на то, что эти наблюдения могут быть реальными, они, вероятнее всего, являются результатом ошибочного ввода данных со стороны персонала. Эти наблюдения негативно влияют на результаты регрессионного анализа и обычно исключаются. В некоторых случаях, когда таких наблюдений немного и они не повлияют неблагоприятно на результаты, им могут быть присвоены средние значения. Однако во всех случаях выбросы должны быть сперва исследованы, поскольку они могут указывать на проблемы, такие, как случаи мошенничества.

Корреляция

Анализ исходных характеристик, описанный в следующем разделе, лишь рассматривает отдельные характеристики - ни корреляция, ни мультиколлинеарность, ни частные взаимосвязи в нем не рассматриваются. Однако корреляция существует и ее нужно контролировать. Кроме того, есть несколько способов обнаружить корреляцию. Один из таких методов – это процедура PROC VARCLUS в SAS. Данная процедура SAS использует разновидность анализа принципиальных компонент, чтобы обнаружить группы характеристик, которые коррелируют между собой.

Процедура PROC VARCLUS лучше использования простого корреляционного анализа, она учитывает коллинеарность, а также корреляцию и поэтому является более оптимальным методом отбора характеристик для разработки скоринговой карты.

Мультиколлинеарность (MC) не представляет значительной угрозы, когда разработка прогнозных моделей осуществляется на основе больших массивов данных. Эффекты мультиколлинеарности (MC), уменьшающие статистическую эффективность модели, могут быть преодолены за счет использования достаточно большой выборки, такой, при которой отдельные эффекты каждой входной переменной могут быть надежно оценены. В этом случае оценки параметров, полученные с помощью регрессии методом наименьших квадратов, будут устойчивыми.

Поиск корреляции можно осуществить до или после анализа исходных характеристик, но перед этапом регрессионного анализа. И корреляция, и группировка дают ценную информацию о данных, имеющихся в распоряжении, и значат больше, чем обычные статистические процедуры. Несмотря на то, что сокращение числа характеристик, которые можно сгруппировать (проверив сначала корреляцию), экономит время, оно лишает возможности увидеть природу взаимосвязей между многочисленными характеристиками и кредитоспособности. Поэтому наилучший метод – это, вероятно, исключение некоторых характеристик в сочетании с отбором характеристик из каждого коррелированного «кластера», осуществляемом на основе бизнес-опыта и интуиции. Он используется для того, чтобы уравновесить эффективность модели и возможность лучшего понимания данных.

Анализ исходных характеристик

Анализ исходных характеристик решает две задачи. Первый задача заключается в том, чтобы оценить силу каждой характеристики заявителя отдельно в качестве предиктора кредитоспособности. Данная процедура известна еще как одномерный анализ и выполняется для того, чтобы обнаружить слабые или противоречащие логике характеристики.

Самые сильные характеристики подвергают группировке. Она применяется по отношению к атрибутам как непрерывных, так и дискретных характеристик, что продиктовано очевидной причиной. Группировка требуется для построения скоринговой карты в формате, приведенном на рисунке 1.1.

Скоринговые карты могут быть разработаны и разрабатываются с помощью непрерывных (несгруппированных) характеристик. Однако группировка дает некоторые преимущества:

- Она облегчает обработку выбросов по интервальным переменным и редких классов.
- Группировка упрощает понимание взаимосвязей и поэтому позволяет получить больше информации о кредитном портфеле. График,

показывающий взаимосвязь между атрибутами характеристики и кредитоспособностью, является более мощным инструментом, чем обычная статистика переменной. Помимо силы взаимосвязи он позволяет пользователям объяснить природу этой взаимосвязи.

- Нелинейные зависимости могут быть построены с помощью линейных моделей.
- Она позволяет осуществлять беспрецедентный контроль над процессом разработки – с помощью назначения групп.
- Процесс группировки характеристик позволяет пользователю более детально понять поведение предикторов риска и получить больше знаний о портфеле, что в дальнейшем может помочь в разработке более оптимальных стратегий по управлению кредитным портфелем.

После того, как самые сильные характеристики сгруппированы и проранжированы, отбор переменных выполнен. По итогам анализа исходных характеристик разработчик скоринговой карты получит набор сильных, сгруппированных характеристик для использования в регрессионном анализе.

Сила характеристик измеряется с использованием четырех главных критериев:

- Предсказательная сила атрибута. Для этого используется Weight Of Evidence (WOE).
- Ранжирование и тренд значений WOE по сгруппированным атрибутам внутри характеристики.
- Предсказательная сила характеристики. Для этого используется Information Value (IV).
- Текущие и деловые соображения (например, исходя из коэффициента обслуживания долга, установленного в соответствии с ограничениями кредитной политики).

Некоторые аналитики перед группировкой характеристик используют другие алгоритмы отбора переменных (например, ранжирование предсказательной силы с использованием хи-квадрат или R-квадрат). Это дает им представление о силе характеристик с использованием средних значений независимых переменных и предупреждает их о наблюдениях, когда значение Information Value является высоким/низким по сравнению с другими метриками.

Анализ исходных характеристик может быть проведен в диалоговом режиме, вовлеченность в анализ сотрудников, от бизнес-пользователей до операторов, должна поощряться. В частности, они могут дать более глубокое понимание неожиданных или нелогичных форм поведения заемщикам и улучшить группировку переменных.

Первый этап анализа исходных характеристик заключается в том, чтобы провести первоначальную группировку переменных и проранжировать их по IV или другим показателям предсказательной силы. Это делается с помощью техники биннинга. В SAS Credit Scoring для этого может использоваться Interactive Grouping Node (Узел Интерактивной Группировки).

Если используются другие программы, наилучший способ биннинга – разбить интервальные переменные на 50 или на другое равное количество групп и вычислить WOE и IV по сгруппированным атрибутам и характеристикам.

Затем можно использовать любое программное обеспечение для тонкой настройки группировки более сильных характеристик, принципы которой изложены в следующем разделе. Аналогично для категориальных характеристик могут быть вычислены WOE каждого атрибута и IV каждой характеристики. Возможна длительная тонкая настройка группировки для характеристик,

превосходящих минимально приемлемое значение. Часто для группировки переменных используются деревья решений. Однако большинство пользователей используют их для того, чтобы сделать исходные предположения, кроме того, используют альтернативное программное обеспечение для интерактивной тонкой настройки группировки.

Статистики

Рисунок 6.2 показывают типичную таблицу для анализа сгруппированных характеристик. В примере приводится характеристика «Возраст» после группировки. «Общ Распр», «Распр Хор» и «Распр Плох» показывает по столбцу общее распределение, распределение «хороших» и распределение «плохих» соответственно. Например, 17.5% от общего числа заявителей, 18.81% «хороших» и 5.21% «плохих» попадает в возрастную группу 35–44.

Рисунок 6.2 Анализ сгруппированных переменных

<i>Age</i>	<i>Count</i>	<i>Tot Distr</i>	<i>Goods</i>	<i>Distr Good</i>	<i>Bads</i>	<i>Distr Bad</i>	<i>Bad Rate</i>	<i>WOE</i>
Missing	1,000	2.50%	860	2.38%	140	3.65%	14.00%	-42.719
18–22	4,000	10.00%	3,040	8.41%	960	25.00%	24.00%	-108.980
23–26	6,000	15.00%	4,920	13.61%	1,080	28.13%	18.00%	-72.613
27–29	9,000	22.50%	8,100	22.40%	900	23.44%	10.00%	-4.526
30–35	10,000	25.00%	9,500	26.27%	500	13.02%	5.00%	70.196
35–44	7,000	17.50%	6,800	18.81%	200	5.21%	2.86%	128.388
44+	3,000	7.50%	2,940	8.13%	60	1.56%	2.00%	164.934
Total	40,000	100%	36,160	100%	3,840	100%	9.60%	

Information Value = 0.668

На рисунке 6.2 нужно обратить внимание на несколько моментов:

- «Missing» («Пропущенные значения») сгруппированы отдельно. WOE этой группы означает, что большая часть пропущенных данных относится к возрастной группе между 23 и 29.
- Для проведения анализа применено правило «минимум 5% в каждой группе».
- Нет групп с количеством «хороших» и «плохих», равным 0.
- Процент «плохих» и WOE должны в достаточной мере отличаться по группам (группировка выполняется так, чтобы максимизировать разницу между «хорошими» и «плохими»). Одна из целей данной процедуры – выявить и отделить атрибуты, которые хорошо дифференцируют заявителей. Несмотря на то, что абсолютное значение WOE важно, разница между WOE групп играет ключевую роль. Чем больше разница между последующими группами, тем выше предсказательная способность данной характеристики.
- WOE непропущенных значений подчиняется логическому распределению, восходя от отрицательных значений к положительным, без смены зависимости на обратную.

Как упоминалось ранее, WOE измеряет силу каждого атрибута или сгруппированных атрибутов, как способность отделять «хороших» от «плохих» клиентов. Этот показатель измеряет разность между долей «хороших» и долей «плохих» в каждом атрибуте (т.е. шансы, что клиент с данным атрибутом является

«хорошим» или «плохим»). WOE основывается на вычислении логарифма шансов:

$$(\text{Распр Хор} / \text{Распр Плох})$$

который измеряет шансы быть «хорошим» клиентом (например, для атрибута 23–26, $13.61/28.13 = 0.48$). Клиент в возрастной группе 23–26 имеет шансы быть «хорошим» 0.48:1.

$$[\ln (\text{Распр Хор} / \text{Распр Плох})] \times 100$$

Более удобный способ вычисления WOE используется в таблице, приведенной на рисунке 6.2:

Например, WOE атрибута 23–26 равен:

$$\ln\left(\frac{0.1361}{0.2813}\right) \times 100 = -72.613$$

Умножение на 100 делает числа более удобными для работы. Отрицательные числа обозначают, что отдельно взятый атрибут выделяет большую пропорцию «плохих», чем «хороших».

Information Value, или общая сила характеристики, измеряется по формуле:

$$\sum (\text{Распр Хор} - \text{Распр Плох}) \times \ln (\text{Распр Хор} / \text{Распр Плох})$$

Заметим, что «Распр Хор» и «Распр Плох» используются в этой формуле в десятичном формате, например, 0.136 и 0.28.

При интерпретации значений IV руководствуются правилом:

- Меньше 0.02 – характеристика не обладает предсказательной способностью
- от 0.02 до 0.1 – слабая предсказательная способность
- от 0.1 до 0.3 – средняя предсказательная способность
- 0.3 и выше – высокая предсказательная способность

Характеристики со значением IV выше 0.5 следует проверить – они либо исключаются из процесса моделирования, либо используются в рамках специальных методов, которые будут описаны в разделе «Подготовка скоринговой карты».

IV является общераспространенной метрикой, но различные специалисты применяют разные правила, определяющие слабость или силу характеристики.

Когда скоринговая карта разрабатывается с использованием несгруппированных характеристик, для оценки предсказательной силы применяются статистики хи-квадрат и R-квадрат. Для оценки характеристик оба метода используют критерий качества подгонки. Техника R-квадрат использует метод пошагового отбора, который отклоняет характеристики, которые не удовлетворяют заданному порогу пошагового увеличения R-квадрат (обычно 0.005). Техника хи-квадрат использует схожий принцип, используя минимальное значение порога 0.5. Пороговые значения могут быть увеличены, если в модели слишком много характеристик.

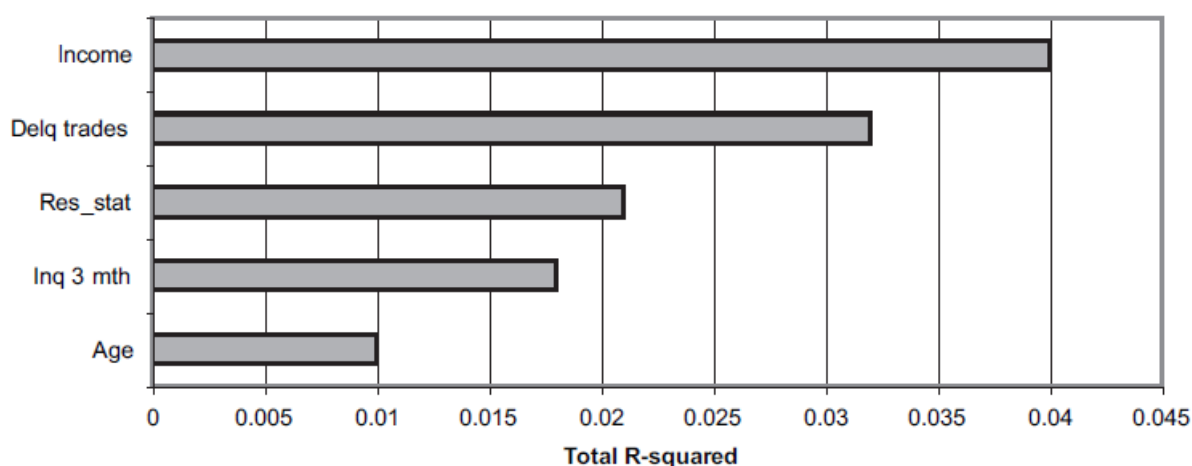
Поскольку техника использует сгруппированные переменные, цель здесь отобрать характеристики для регрессии (следующего этапа моделирования).

Кроме того, важно отметить, что одномерный анализ, вне зависимости от того, используется группировка или нет, не объясняет частичные взаимосвязи и взаимодействия между входными характеристиками. Частичная взаимосвязь возникает, когда эффект одной характеристики меняется в присутствии другой. Многомерные методы, которые исследуют объединенные поднаборы переменных, могут быть более предпочтительными в данном случае. В любом случае, цель данной процедуры – сформировать набор сильных переменных, чтобы подать на вход регрессии (или другого метода, при необходимости).

Некоторое программное обеспечение, использующееся для моделирования, предлагает группировать характеристики с помощью методов R-квадрат и хи-квадрат и тестировать взаимодействия категориальных входных переменных. Примеры 2-факторных взаимодействий, которые могут тестироваться – доход*статус проживания, возраст*доход и т.д. Данная методология выходит за пределы анализа исходных характеристик и может дать более весомые результаты, рассматривая взаимодействия между характеристиками.

Результаты, полученные с помощью анализа R-квадрат, показаны на рисунке 6.3, где приведено пошаговое увеличение значения R-квадрат по мере добавления в модель характеристик, начиная с возраста и заканчивая доходом.

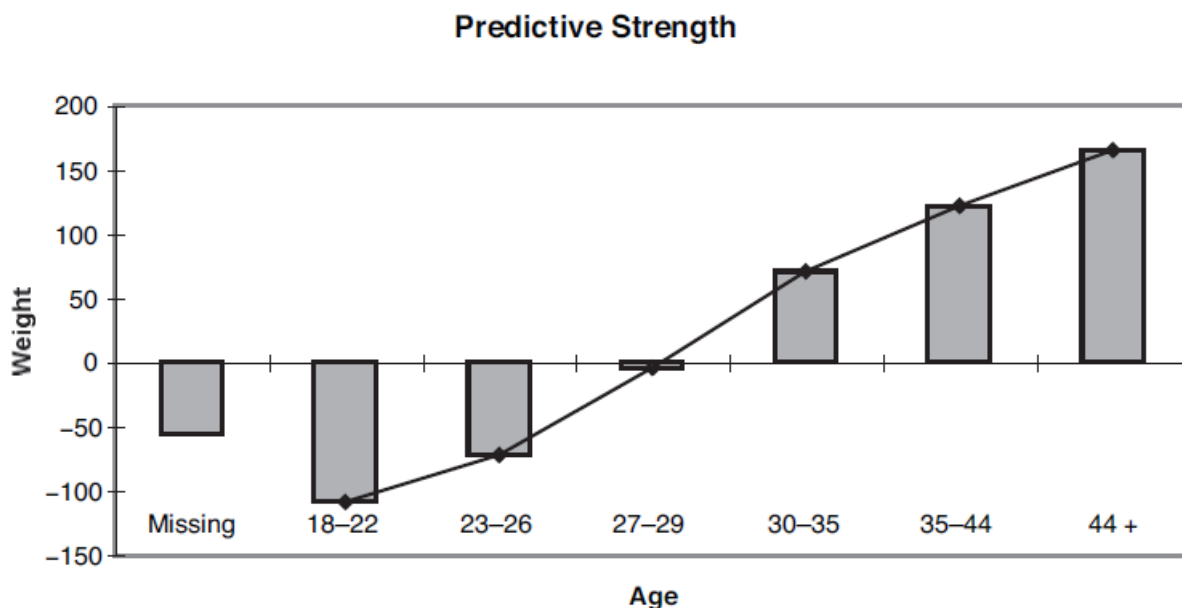
Рисунок 6.3 Характеристики модели



Логичный тренд

Однако, предсказательная сила, измеряемая в терминах WOE и IV, не является единственным фактором отбора характеристик для дальнейшего анализа или определения в качестве сильного предиктора. В скоринговых картах, построенных по сгруппированным характеристикам, значения силы атрибутов должны также следовать в логическом порядке и иметь смысл. Например, на рисунке 6.4 приведено распределение WOE по атрибутам характеристики «Возраст», взятое из таблицы на рисунке 6.2.

Рисунок 6.4 Логичный тренд WOE для возраста



Как видно, группировки этой характеристики, за исключением «Missing» («Пропущенных значений») линейно взаимосвязаны с WOE; то есть, они показывают линейную и логическую взаимосвязь между атрибутами возраста и пропорцией «плохих». Это подтверждает практические наблюдения в области кредитования и страхования, что более молодые люди имеют более высокий риск, чем старшее поколение. Установление таких логических (необязательно линейных) взаимосвязей с помощью группировки является целью данной процедуры. Процесс построения логичного тренда – это процесс проб и ошибок, в ходе которого соблюдают равновесие между построением логических трендов и сохранением приемлемого значения IV.

Экспериментирование с различным числом группировок устраняет смены зависимостей (когда тренд меняет направление на противоположное) и другие нелогические взаимосвязи. Общие тренды могут быть рассмотрены с помощью поиска взаимосвязей между WOE и необработанными (несгруппированными) атрибутами – группировка только сглаживает кривую. Однако в некоторых случаях смены зависимостей могут отражать реальное поведение или данные, их устранение может снизить общую силу характеристики. Они должны быть сперва тщательно изучены, чтобы понять, есть ли надежное объяснение такого поведения. В целом группировка снижает «переобучение», по причине которого моделируются скорее странности поведения данных, чем общий прогнозный тренд. Когда возникают обоснованные нелинейные взаимосвязи, они должны использоваться, при условии, что могут быть объяснены опытом или трендами в экономике. Кроме того, необходимо подтвердить, что общий тренд или профиль получен с помощью модели, а не вызван случайными колебаниями данных. В этом плане лучшим средством проверки будет бизнес-опыт. Например, в Северной Америке утилизация по револьверным кредитным линиям согласно WOE имеет бананообразную кривую. Клиенты с очень низкой утилизацией являются высокорисковыми, затем риск снижается до некоторой отметки, но в итоге риск начинает увеличиваться по мере того, как увеличивается утилизация. Прочие достоверные взаимосвязи могут принимать U-образную форму, эти зависимости оставляют до тех пор, пока их можно объяснить.

Номинальные переменные группируются так, чтобы собрать вместе атрибуты со схожими WOE и, как и в случае с непрерывными переменными, максимизировать различия между группами.

Очевидно, что данная процедура может быть неправильно выполнена, когда осуществляется человеком, не знающим бизнеса. Это снова подчеркивает необходимость выполнения процедуры совместно с другими членами команды, участвующей в разработке скоринговой карты

На рисунке 6.5. приведен пример нелогичного тренда. В этом наборе данных характеристика является слабой и не показывает логической взаимосвязи между возрастом и соотношением «хороших» и «плохих».

Рисунок 6.5 Нелогичный тренд WOE для возраста

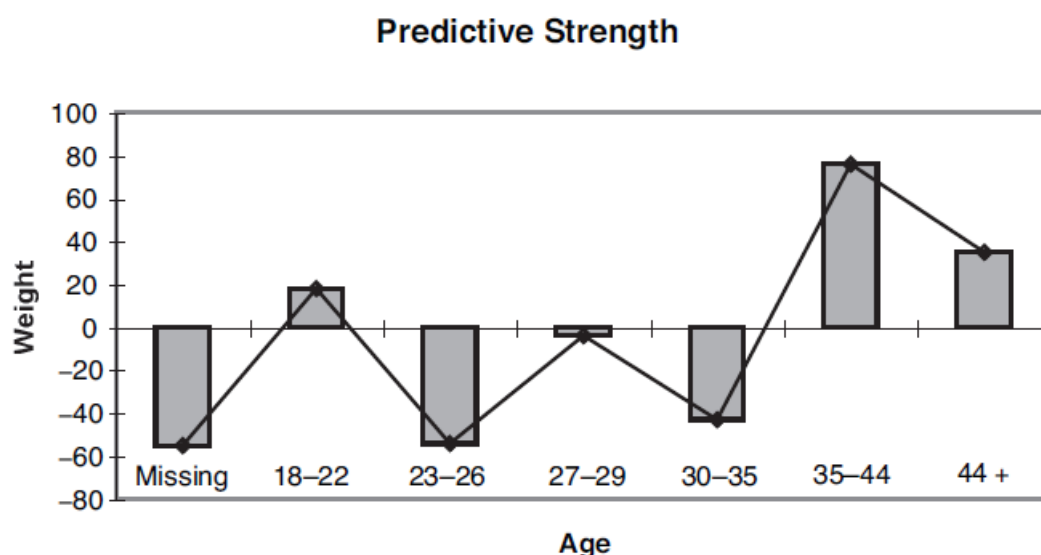
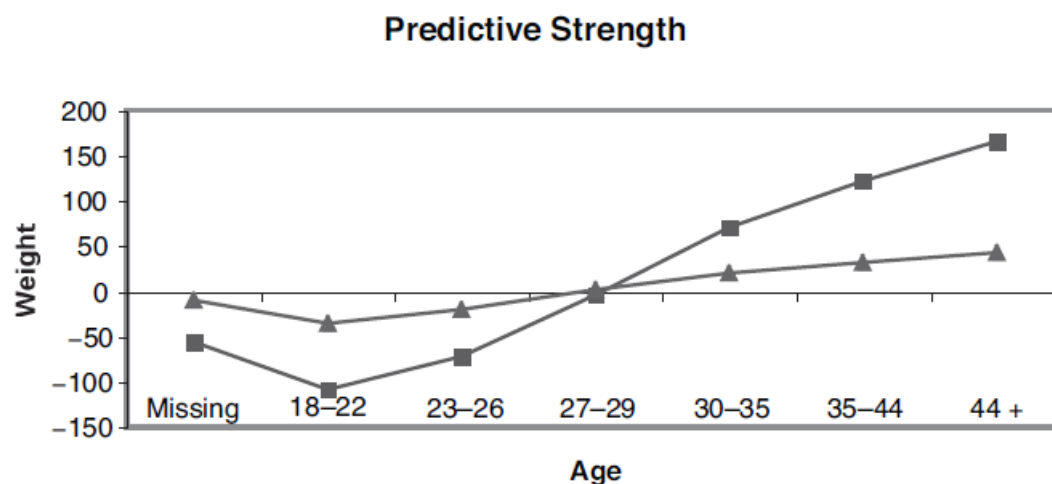


Рисунок 6.6. показывает две зависимости WOE от атрибутов характеристики. Однако более крутая линия (маркеры – квадраты) соответствует более сильной предсказательной взаимосвязи между возрастом и соотношением классов. Этот факт будет отражен в значениях IV.

Рисунок 6.6 Логически обоснованный тренд и предсказательная сила



Анализ исходных характеристик включает создание логических взаимосвязей, опирающихся на бизнес-опыт, с помощью группировки атрибутов, превосходящих минимальное значение IV. Альтернативный, чисто статистический подход включает установление взаимосвязей, которые лишь максимизируют IV или другие показатели, вне зависимости, сгруппированы атрибуты или нет. Бизнес-ориентированный подход лучше по нескольким причинам:

- Логические взаимосвязи гарантируют, что финальные веса после проведения регрессионного анализа будут иметь смысл. Они также гарантируют, что когда атрибуты распределяются по баллам, чтобы построить скоринговую карту, эти баллы несут логический смысл (например, чем старше заявитель по возрасту, тем больше баллов он получает).
- Логические взаимосвязи лучше принимаются конечными пользователями. Когда скоринговая карта подтверждает жизненный опыт, она пользуется более высоким уровнем доверия при автоматическом принятии решения.
- Логические взаимосвязи подтверждают бизнес-опыт, таким образом они делают шаг вперед по сравнению с чисто статистическим оцениванием. Это позволяет применить бизнес-опыт, чтобы улучшить прогнозное моделирование, сделать его релевантным для использования в бизнесе.
- Важнее всего, что обобщение взаимосвязей посредством их логической группировки снижает переобучение. Вы больше не моделируете каждую индивидуальную особенность данных, присваивая несгруппированным атрибутам неограниченное число весов. Теперь вы ранжируете риск и моделируете тренды так, что скоринговая карта может применяться к текущему потоку клиентов с некоторым запасом эластичности (модель способна учитывать некоторые изменения в популяции) и будет оставаться стабильной в течение долгого периода времени. Справедливое беспокойство здесь должно вызывать переобобщение, в силу чего модель, по-видимому, будет работать, даже когда изменения в популяции станут критическими. Решение этой проблемы заключается в том, чтобы построить расширенный профиль риска, а не скоринговую карту, основанную на узком наборе критериев. Долгий период использования отличает разработку скоринговой карты от маркетинговых моделей, которые часто строятся под конкретные кампании и затем отвергаются. Поэтому нельзя позволить, чтобы модель была нестабильной.

Деловые/текущие соображения

В качестве подходов, используемых для группировки атрибутов, рассматриваются статистический анализ и бизнес-логика. Третий подход – это значимость для бизнеса или текущей деятельности.

Для нечисловых дискретных переменных – например, номинальных, таких, как код жизненного стиля – группировки создаются на основе схожих весов, чтобы построить логичный тренд (т.е. атрибуты со схожими весами группируются вместе). Группировки должны быть изучены с учетом различных факторов.

В некоторых случаях имеет смысл назначать группы с помощью кредитной политики – ставить отсечки. Например, если правила кредитной политики требуют, чтобы кредиты с коэффициентом обслуживания долга более 42% отправлялись на дополнительное рассмотрение, можно ввести порог отсечения 42%, т.е. разбить кредиты на две группы – кредиты с коэффициентом обслуживания долга менее 42% и кредиты с коэффициентом обслуживания долга более 42%. Такие группировки могут проверить общепринятое мнение и

предыдущую кредитную политику – например, чтобы посмотреть, есть ли смысл устанавливать правило «42%» в данной точке или было бы лучше установить более высокий коэффициент обслуживания долга, чтобы максимизировать дискриминирующую способность.

Предварительная скоринговая карта

Анализ исходных характеристик определяет набор сильных характеристик, которые нужно рассмотреть для включения в итоговую модель. Он преобразует их в сгруппированные переменные. На этапе предварительной скоринговой карты для того, чтобы выбрать набор характеристик, обладающий наибольшей предсказательной силой, могут быть использованы различные техники моделирования. Некоторые из техник, используемых в бизнесе, - это логистическая регрессия, деревья решений и нейронные сети. Вообще итоговые скоринговые карты, создаваемые на этом этапе, должны содержать от 8 до 15 характеристик. Это делается, чтобы гарантировать стабильно работающую скоринговую карту, предсказательная способность которой будет высокой, даже если профиль одной или двух характеристик должен был измениться. Скоринговые карты с небольшим количеством характеристик в целом не способны выдержать проверку временем, поскольку они чувствительны к незначительным изменениям в профиле заявителя.

Вне зависимости от используемой техники моделирования, необходимо разработать скоринговую карту с оптимальным сочетанием характеристик, принимая во внимание такие проблемы, как:

- Корреляция между характеристиками
- Итоговая предсказательная сила скоринговой карты
- Интерпретируемость модели
- Реализуемость
- Прозрачность методологии для регулирующих органов

Понятие профиля риска

Скоринговые карты могут разрабатываться для различных целей – максимизации качества статистических показателей, эффективности (с использованием наименьшего числа переменных) и т.д. В условиях бизнеса скоринговые карты разрабатываются, чтобы помочь специалисту смоделировать процесс принятия решения. Опытный кредитный аналитик никогда не будет принимать свое решение, исходя только из 4-5 правил анкеты или истории клиента. Вероятнее всего, он проанализирует сразу несколько ключевых показателей для формирования *профиля риска* клиента. Так какой смысл разрабатывать скоринговые карты всего с 4-5 переменными или характеристиками?

Цель процесса разработки скоринговой карты - построить наиболее полный профиль риска для клиента. Такой широкий подход делает скоринговую карту не только более эффективными, но и менее восприимчивыми к изменениям в одной отдельной области. Такой профиль риска должен включать в себя характеристики, отражающие столько независимых типов информации, сколько возможно. Например, скоринговая карта для кредитной карты должна включать в себя: демографическую информацию о клиенте (возраст, регион и стаж работы), информацию кредитного бюро (запросы на открытие новых счетов, обслуживание кредита, информация из открытых источников, включающая данные о судебных постановлениях, банкротствах, штрафах, просроченных залоговых операциях, арестах имущества за просроченную задолженность по какому-либо налогу,

просрочках по выплатам алиментов), показатели обслуживания долга (валовой коэффициент обслуживания долга), другую важную информацию о текущих клиентах.

Профиль риска также помогает сделать последующий мониторинг скоринговой карты более релевантным. Большинство риск-аналитиков предпочитает использовать ежемесячные отчеты типа "стабильность системы" или "стабильность популяции", чтобы убедиться в достоверности результатов скоринговой карты по текущим заявителям или клиентам. Эти отчеты эффективно оценивают изменение популяции, *исходя лишь из характеристик, используемых в скоринговой карте*. Общий профиль риска может более реалистично отразить текущие изменения в популяции, чем при использовании ограниченного количества переменных из скоринговой карты.

Создание скоринговой карты на основе профиля риска в теории практически ничем не отличается от других процедур прогнозного моделирования. Разница состоит лишь в методе получения конечного набора характеристик. Большая часть техник, упоминавшихся в предыдущих главах, может и должна быть умело использована для решения вопросов, обсуждавшихся в предыдущих параграфах, поскольку запуск алгоритмов моделирования без вмешательства кредитного аналитика вряд ли приведет к построению профиля риска. Оставшаяся часть данного раздела будет посвящена методам, использующим логистическую регрессию для построения скоринговых карт.

Логистическая регрессия

Логистическая регрессия – широко распространенная техника, применяемая для разработки скоринговых карт в большинстве программных приложений, где прогнозируемые переменные являются категориальными. В случае непрерывных прогнозируемых переменных используется линейная регрессия. Оставшаяся часть данного раздела будет посвящена использованию множественной логистической регрессии для прогноза бинарной переменной («хорошие»/«плохие»).

Логистическая регрессия, как и большинство других методов прогнозного моделирования, использует набор характеристик (предикторов) для определения вероятности наступления заданного события. Уравнение логит-преобразования вероятности наступления события выглядит следующим образом:

$$\text{Logit} (p_i) = \beta_0 + \beta_1 x_1 + . . . + \beta_k x_k$$

где

p = итоговая вероятность наступления "события," обусловленная входными переменными

x = входные переменные (предикторы)

β_0 = константа (начальный уровень линии регрессии)

β_k = параметры

Логит-преобразование - это логарифм шансов, т.е. логарифм отношения вероятности наступления события к вероятности его ненаступления:

$$\ln \left(\frac{p_i}{1 - p_i} \right)$$

Оно используется для линеаризации итоговой вероятности, ограничивая вероятность от 0 до 1. Для оценки параметров $\beta_1 \dots \beta_k$ используется метод максимального правдоподобия. Эти параметры оценивают меру изменения логарифма шанса при изменении входной переменной на одну единицу (в согласовании с другими входными переменными). Фактически коэффициенты показывают наклон линии регрессии между целевой переменной и соответствующей входной переменной $x_1 \dots x_k$. Параметры зависят от единицы измерения входной переменной (например, выражаются в процентном отношении), и их необходимо стандартизировать для облегчения анализа. Стандартизация может быть выполнена различными методами, включая и метод стандартизированных оценок. Другой способ стандартизации состоит в общей отмене единиц измерения входных данных, и выполнении регрессионного анализа не по входным данным, а по WOE для каждой группировки, созданной на прошлом шаге.

Регрессия подразумевает наличие целевой переменной и входных переменных. Эти входные переменные могут иметь различные типы измерения. Наиболее общий способ - это использовать необработанную входную информацию для числовых данных и создавать дамми-переменные для категориальных переменных. Далее в анализе используется метод стандартизированных оценок для нейтрализации эффекта, вызванного различными единицами измерения входных переменных. Однако этот метод не релевантен, когда необходимо разработать скоринговую карту по сгруппированным переменным.

В случае скоринговых карт по сгруппированным переменным, входные переменные могут быть представлены в виде средних значений для числовых переменных, например средний возраст каждой группы, или взвешенное среднее, или дамми-переменные для групп в категории. Использование дамми-переменных для категориальных переменных имеет существенный недостаток – оно предполагает, что разница между группами внутри категориальной переменной одна и та же. Более предпочтительный метод для сгруппированных переменных – использование WOE каждой группы в качестве входных данных. Данный подход не только решает проблемы различий единиц измерения входных переменных, но и принимает в расчет точный тренд и шкалу зависимостей между группами. Он помогает при разработке скоринговой карты тем, что сохраняет каждую характеристику неизменной. Кроме того, если группировка выполнена верно, можно быть уверенным, что распределение баллов по каждой группе в ходе масштабирования скоринговой карты является логически обоснованным и отражают разницу во взаимосвязях между группами. Гораздо чаще используются следующие три пошаговые техники логистической регрессии:

- **Метод пошагового включения.** Сначала выбирает наилучшую характеристику модели, основываясь на индивидуальной предсказательной силе каждой характеристики, затем пошагово добавляет остальные характеристики в эту модель до тех пор, пока не останется переменных с p -value меньше заданного уровня значимости (например, 0.5) или хи-квадрат выше заданного уровня. Этот метод эффективен, но может не работать, если имеется очень большое количество переменных или существует высокая корреляция.
- **Метод пошагового исключения.** Противоположный методу пошагового включения, этот метод работает сразу со всеми характеристиками модели,

и последовательно исключает переменные с наименьшим уровнем значимости. Процесс идет до тех пор, пока все оставшиеся переменные не будут иметь *p-value* ниже уровня значимости (например, 0.1). Этот метод предоставляет переменным с низким уровнем значимости больше шансов войти в модель, чем метод пошагового включения и метод пошагового отбора.

- **Метод пошагового отбора.** Комбинация двух предыдущих техник. Использует и добавление и удаление переменных динамически из скоринговой карты на каждом этапе до тех пор, пока не будет достигнута наилучшая комбинация переменных. Пользователь может задать минимальные *p-value*, при которых переменная добавляется в создаваемую модель, или остается в модели.

Проектирование скоринговой карты

Хотя и возможно построить скоринговую карту, введя все характеристики в регрессионную модель и сгенерировав статистически оптимальный результат, этот метод не может принести хорошие результаты. Разработчик скоринговой карты обычно опирается на некоторые статистические показатели, такие как *p-value*, хи-квадрат, R-квадрат и некоторые другие для определения качества построенной модели. Однако есть некоторые бизнес-задачи, решение которых необходимо учесть при разработке скоринговой карты.

Первая задача состоит в определении наилучшего набора характеристик, и построении полного профиля рисков. Методика построения профиля рисков была описана выше. В идеале, этот профиль должен быть построен с использованием как можно большего числа независимых типов информации, например демографических, временных, финансовых, запросов в кредитное бюро, платежеспособности и т.д. Процесс разработки должен учитывать проблемы корреляции и коллинеарности и другие факторы, влияющие на надежность модели.

Разработанная скоринговая карта должна соответствовать структуре принятия решений в организации. Если модель является единственным решающим фактором, необходимость построения полного профиля рисков становится еще значимее. Если модель используется как инструмент принятия решений, то переменные, включаемые в карту, должны согласовываться с остальными показателями и не противоречить им. Например, включение таких характеристик, как банкротство, TDSR (общий коэффициент обслуживания долга), просрочки по предыдущим кредитам и других, обычно уже включенных в правила кредитной политики, должно быть сведено к минимуму.

Пример, приведенный на рисунке 6.7, показывает характеристики скоринговой карты просрочек клиента. Данная скоринговая карта представляет собой полный профиль риска. Заметьте, что она включает различные типы информации, как из внутренних, так и из внешних источников. Включение отношения запросов за последние 3 месяца и за последние 12 месяцев сделано для того, чтобы можно было определить кратко- и долгосрочную потребность в кредитовании. Эти два показателя могли быть включены независимо друг от друга. «Банкротство» и «Информация из открытых источников» не были включены в скоринговую карту, так как они используются в правилах кредитной политики, чтобы автоматически отсеивать соответствующих заявителей. «Общий стаж работы» используется вместо «Стаж на последнем месте работы», поскольку общий период занятости

является лучшим индикатором риска, чем время, проведенное на последнем месте работы, особенно в городах с высокой мобильностью рабочей силы. Отношение числа запросов на открытие новых счетов за последние 12 месяцев к числу кредитных линий, открытых в течение того же периода времени, измеряет долю успешных попыток заявителя получить кредит. Один кредитный аналитик, у которого я брал интервью, называет это «средним уровнем успеха» заявителя.

Рисунок 6.7 Пример скоринговой карты на основе профиля риска

- **Возраст**
- **Статус проживания**
- **Почтовый индекс**
- **Общий стаж работы**
- **Срок проживания по последнему адресу**
- **Запросы на открытие новых счетов за 3 месяца / Запросы на открытие новых счетов за 12 месяцев**
- **Запросы на открытие новых счетов за 12 месяцев / Кредитные линии, открытые за 12 месяцев**
- **Кредитные линии с платежом, просроченным более чем на 90 дней / Общее количество кредитных линий**
- **Кредитные линии, открытые за последние 3 месяца / Кредитные линии, открытые за 12 месяцев**
- **Утилизация (сумма балансов, деленная на сумму кредитных линий)**
- **Количество счетов в банке**
- **Задолженность перед банком**
- **Общий коэффициент обслуживания долга**

Обычно такая скоринговая карта не является результатом автоматического регрессионного алгоритма. Как же получается подобная скоринговая карта? Спроектируем ее.

Разработчик скоринговой карты имеет на вооружении несколько методов, с помощью которых он может построить итоговую модель. Эти методы предполагают отбор обязательных характеристик, в силу чего они, считаясь необходимыми или включенными по принципу «должны быть», принудительно вводятся на начальном этапе в скоринговую карту. Проводится настройка модели, чтобы максимально увеличить шансы включения определенных характеристик в итоговую модель.

Один из способов добиться результата - это учесть значимые характеристики для включения в модель, шаг за шагом, причем переменные, рассматриваемые на каждом шаге, задаются индивидуально. Этот способ не отличается от пошаговой регрессии. Пример приведен на рисунке 6.8.

Используя данную технику, регрессионный алгоритм сначала выбирает характеристики, применяя либо метод пошагового включения, либо метод пошагового исключения, либо метод пошагового отбора. Характеристики, удовлетворяющие критерию отбора (например, *p-value* параметра оценивается по заданному уровню статистической значимости), добавляются в скоринговую карту первыми (или удаляются первыми, если используется метод пошагового исключения). В приведенном примере характеристики «возраст», «срок проживания по последнему адресу», «срок обслуживания в банке» и «общий стаж работы» будут вычислены при первой итерации, принимая в расчет корреляцию. Предположим, что характеристика «возраст» обладает наибольшей предсказательной силой – таким образом, «возраст» будет добавлен в модель.

На второй итерации того же уровня, алгоритм рассмотрит три оставшиеся характеристики, принимая во внимание уже отобранную переменную "возраст". Если окажется, что либо одна, либо все оставшиеся характеристики в

достаточной мере увеличивают предсказательную силу скоринговой карты, они тоже будут добавлены в модель. Регрессионный анализ остановится, когда не будет ни одной характеристики, которую можно было бы добавить, либо исключить из модели.

Все характеристики, включенные в модель на шаге 1, будут участвовать в ней и на шаге 2. Регрессионный алгоритм на этом шаге, который должен рассмотреть такие характеристики, как «район», «почтовый индекс» и «область», начнет с характеристик, отобранных на шаге 1 и уже включенных в модель. Кроме того, такие показатели, как *p-value* и уровень значимости будут использованы для включения характеристик в модель на данном шаге.

Подобный анализ будет проведен для каждого последующего уровня до тех пор, пока не будет получена итоговая скоринговая карта. Характеристики, включенные в модель на более ранних шагах, будут включены в модель и на последующих шагах.

Статистические показатели, такие как хи-квадрат или стандартизированные оценки, могут быть использованы для измерения силы прогнозной модели на каждой итерации.

Опытный пользователь может настроить процесс анализа для увеличения шансов получения тех или иных итоговых характеристик скоринговой карты. Сравнительно более слабые и «предпочтительные» характеристики могут быть помещены на более ранние шаги, чтобы максимально увеличить их шансы войти в модель или чтобы максимально увеличить влияние проверенных переменных, введя их первыми, а затем позволить другим переменным увеличить предсказательную силу модели.

Более сильные характеристики помещаются в конец и могут не войти в скоринговую карту, поскольку их влияние на прогноз может быть уже смоделировано по одному или нескольким другим критериям. Использование нескольких более слабых критериев для моделирования поведения одного сильного критерия применяется для стабилизации, причем без потери силы прогноза (например, пять характеристик, добавляющих 200 баллов каждая в скоринговой карте, предпочтительнее, чем две характеристики, добавляющие по 500 баллов каждая). Модель будет эффективна, основываясь на более широком наборе характеристик, что соответствует идее создания профиля риска.

Схожие критерии («возраст», «срок проживания по последнему адресу», «общий стаж работы» или «запросы за 3 месяца», «запросы за 6 месяцев», «запросы за 12 месяцев») размещаются на одном и том же шаге для того, чтобы в дальнейшем могла быть учтена корреляция между этими характеристиками. Наилучшая среди коррелированных характеристик войдет в скоринговую карту. Связанные между собой коэффициенты должны быть также помещены в один и тот же шаг в качестве информации о числителе и знаменателе. Кроме того, рассмотрение отдельно различных типов независимой информации на каждом шаге максимально увеличивает шансы того, что по крайней мере одна переменная от каждого типа информации войдет в итоговую скоринговую карту.

Рисунок 6.8 Определение характеристик на каждом шаге регрессии

Шаг 1 Возраст, срок проживания по последнему адресу, срок обслуживания в банке, общий стаж работы

Шаг 2 Район, почтовый индекс, регион

Шаг 3 Текущий клиент (Да/Нет)

Шаг 4 Запросы за 3 месяца; запросы за 6 месяцев; запросы за 9 месяцев; запросы за 3 месяца/запросы за 12 месяцев

Шаг 5 Количество кредитных линий с просрочками; кредитные линии с платежом, просроченным более чем на 90 дней/общее количество кредитных линий; общее количество кредитных линий

Шаг 6 Утилизация, информация из открытых источников

Шаг 7 Банкротство

Регрессионный анализ повторяется, используя в ходе итеративного процесса различные комбинации характеристик на разных шагах и различные уровни значимости, чтобы получить скоринговую карту с наибольшей предсказательной способностью. Характеристики могут быть перемещены на более ранние или более поздние шаги, чтобы получить различные комбинации характеристик для скоринговых карт. Эти карты будут оценены позже, с использованием бизнес-критериев, комбинации характеристик и статистических показателей прогнозной силы модели.

На практике этого можно достичь, используя настройки последовательности характеристик в пошаговой регрессии. Есть два часто используемых метода:

1. Простая регрессия
2. Множественная регрессия

Простая регрессия Выполняется однократный запуск регрессии, характеристики располагаются упорядоченно, основываясь на типе информации и предсказательной силе. На рисунке 6.9 приведен пример.

Более слабые типы информации располагаются вверху, т.е. в начале (основываясь на среднем значении IV), а более сильные – внизу, т.е. в конце. Внутри каждого типа информации характеристики могут быть отсортированы от самой слабой к самой сильной. Ранжирование каждой характеристики также может выполняться с помощью IV. Пример на рисунке 6.9 показывает характеристики, проранжированные от самой слабой до самой сильной, основываясь на общем IV. Внутри каждой характеристики, таких как «Срок» и «Запросы», проводится дополнительное ранжирование на основе IV в каждом типе информации. Это задает последовательность, в соответствии с которой регрессионный алгоритм рассмотрит каждую характеристику. Это отправная точка и заданная последовательность должна быть скорректирована в последующих запусках регрессионного алгоритма, пока не будут достигнуты желаемые результаты. Другой способ проранжировать характеристики для простой регрессии состоит в том, чтобы расположить характеристики в зависимости от значений IV, от самого низкого к самому высокому, независимо от типа информации.

Рисунок 6.9 Входные характеристики для простой регрессии

<div style="display: flex; flex-direction: column; align-items: center;"> <div style="margin-bottom: 20px;">Weaker</div> <div style="margin-bottom: 20px;">↓</div> <div style="margin-bottom: 20px;">Stronger</div> </div> <div style="display: flex; flex-direction: column; align-items: center;"> <div style="margin-bottom: 20px;">Weaker</div> <div style="margin-bottom: 20px;">↓</div> <div style="margin-bottom: 20px;">Stronger</div> </div> <div style="display: flex; flex-direction: column; align-items: center;"> <div style="margin-bottom: 20px;">Weaker</div> <div style="margin-bottom: 20px;">↓</div> <div style="margin-bottom: 20px;">Stronger</div> </div>	Info Type		
		Characteristic	IV
		Time 1	0.02
		Time 2	0.04
		Time 3	0.06
		Demographics 1	0.09
		Demographics 2	0.12
		Demographics 3	0.2
		Inquiries 1	0.15
		Inquiries 2	0.18
		Inquiries 3	0.19
		Inquiries 4	0.26
		Financial 1	0.25
		Financial 2	0.34

Множественная регрессия При использовании данного метода алгоритм регрессионного анализа многократно повторяется, рассматривая каждый тип информации отдельно на каждом шаге.

- Более слабые типы информации рассматриваются в первую очередь на начальных шагах регрессионного анализа.
- Внутри каждого шага характеристики располагаются в порядке от самой слабой к самой сильной.
- Характеристики, включенные в скоринговую карту на более ранних шагах анализа, включаются во все последующие шаги.

В SAS упорядоченная регрессия, такая, как показано на рисунке 6.9, может быть построена с помощью процедуры PROC LOGISTIC с использованием настройки “SEQUENTIAL=”. Настройка “INCLUDE=” может быть использована для принудительного включения характеристик, чтобы оставить в скоринговой карте, тогда как настройка “START=” запускает пошаговую регрессию с заданными заранее x переменными (но не в любом заданном порядке), однако эти переменные могут быть удалены на последующих шагах.

Выполнение регрессионного анализа с заданными последовательностями, как было описано выше, может быть выполнено в SAS Enterprise Miner с использованием настройки Model Ordering (Упорядочение модели) в узле регрессии.

Кроме того, как и группировка, данный метод для разработки скоринговых карт может быть неправильно использован из-за своей гибкости. Понимание статистических показателей, а также используемых данных снизит шансы неправильного применения. Данный метод должен быть апробирован с использованием нескольких различных комбинаций для того, чтобы понять динамику изменения данных, перед разработкой итоговой скоринговой карты.

Данный процесс сочетает статистический подход (например, регрессионный анализ) с бизнес-подходом к разработке скоринговой карты, которая эффективна и устойчива, включает характеристики из различных источников и представляет различные типы информации, все вместе формирующие профиль риска (демографические данные, запросы, кредитные линии и т.д.). Заметим, что регрессионный анализ выполняется с использованием наиболее эффективного набора переменных, выбранных с помощью анализа исходных характеристик, и все слабые критерии исключены. Все тесты на значимость участвуют в выборе итоговой структуры скоринговой карты, но это не единственный критерий для включения. Построенная скоринговая карта обладает диагностируемой статистической силой и влиянием. Что еще более важно, скоринговая карта – это полезный бизнес-инструмент, который может использоваться для разработки стратегий, скорректированных с учетом риска, риск-менеджерами и другими специалистами, ответственными за принятие решений. Другие преимущества перечислены в разделе «Понятие профиля риска».

После того, как список характеристик для включения в скоринговую карту получен, они могут быть проанализированы с помощью регрессии в сгруппированном виде, для получения итоговых параметров регрессии. В случае сегментации подобные процессы сопутствуют разработке каждой скоринговой карты – для каждого сегмента в отдельности. Традиционно для каждого сегмента строится несколько скоринговых карт с использованием различных комбинаций характеристик. Скоринговые карты оцениваются в соответствии со стратегическими целями, чтобы сделать итоговый выбор. Скоринговая карта с более низкой предсказательной силой может обладать большей эффективностью с точки зрения стратегической цели (например, больше прибыли), чем скоринговая карта с более высокой предсказательной силой и поэтому лучше оценить несколько скоринговых карт данным способом, чем полагаться только на статистические показатели. Заметим, что выбор критериев эффективности скоринговых карт и их валидация будет рассмотрен в последующих разделах.

Итогом данного этапа являются несколько различных скоринговых карт, включающих список характеристик и их параметры регрессии.

Анализ отклоненных заявок

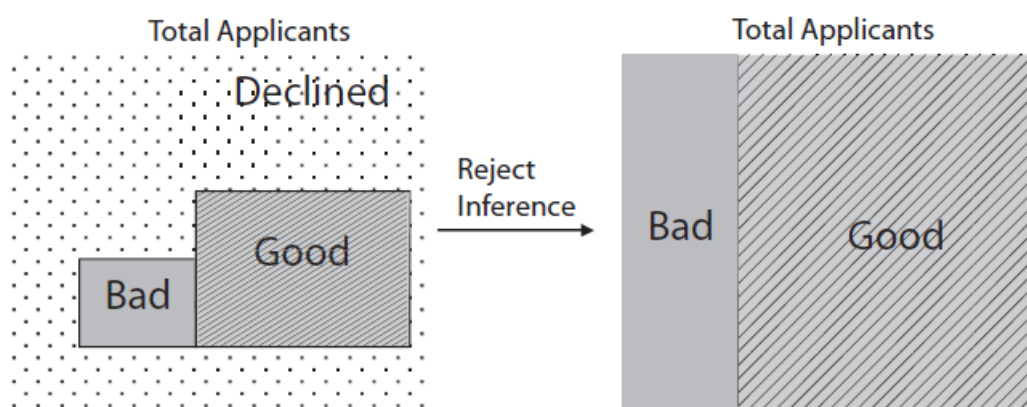
Все оценки, связанные с разработкой модели до этого момента, учитывали клиентов с известным результатом погашения кредита. Их обычно относят к выборке «Фактические «Хорошие»/«Плохие»». Аппликационные скоринговые карты разрабатываются так, чтобы спрогнозировать поведение всех заявителей, но применение модели, построенной лишь по одобренным заявителям, может привести к неверным результатам («смещение выборки»). Это особенно верно, когда предыдущие решения «одобрить/отклонить» принимались систематически и не были случайными, т.е. численность одобренных представляет собой смещенную выборку и нерепрезентативна для отклоненных. Описываемый метод необходим, чтобы объяснить наблюдения, где поведение заявителя неизвестно. Отметим, что если были разработаны поведенческие скоринговые карты, этот этап не нужен.

Анализ отклоненных заявок – это метод, с помощью которого исследуется исполнение кредитных обязательств ранее отклоненными заявителями, чтобы присвоить им класс кредитоспособности. Точно так же, как есть какие-то «плохие», которые получают одобрение, будут и какие-то «хорошие», которые

будут отклонены. Этот метод делает процесс разработки скоринговой карты релевантным, воссоздавая исполнение кредитных обязательств при проценте одобрения 100% (получаем отношение шансов получить одобрение и получить отказ).

Рисунок 6.10 показывает, как это происходит. В левой части рисунка – картинка до анализа отклоненных заявок, с фактическими «хорошими», «плохими» и отклоненными заявителями. Для разработки скоринговой карты, применимой ко всем заявителям, необходима правая часть рисунка, где представлены все заявители, классифицированные на «хороших» или «плохих».

Рисунок 6.10 Анализ отклоненных заявок



Причины для проведения анализа отклоненных заявок

Первая причина проведения анализа отклоненных заявок – это релевантность результатов скоринговой карты, игнорирование отклоненных является причиной того, что скоринговая карта не применима к общей популяции заявителей. Ранее упоминалась проблема смещения выборки. Анализ отклоненных заявок также включает в себя влияние предыдущих решений, принятых в ходе разработки скоринговой карты. Это особенно верно для отделов андеррайтинга, где заявителей анализируют вручную. Например, рассмотрим сценарий, в котором 1000 из 10000 заявителей имеют серьезную просрочку. Кредитные аналитики отклоняют 940 и одобряют 60 заявителей. Последующее исполнение кредитных обязательств показывает, что большая часть из 60 одобренных заявителей аккуратно гасят кредит и классифицированы как «хорошие». Несправедливость отбора вряд ли стала неожиданной новостью. Теперь если скоринговая карта разрабатывается с использованием фактических «хороших» и «плохих», напрашивается вывод, что те, у кого есть серьезные просрочки, являются «хорошими». Анализ отклоненных заявок может нейтрализовать искажающий эффект «наиболее понравившихся» заявителей и даже правила кредитной политики, включив вероятность одобрения «наиболее понравившихся» заявителей, классифицированных на «хороших» и «плохих».

С точки зрения процесса принятия решений анализ отклоненных заявок дает более точный и реалистичный прогноз исполнения кредитных обязательств всеми заявителями. Например, рассмотрим банк, который, используя свою скоринговую карту, традиционно одобрял всех заявителей, которые набирали 200 баллов и выше. Теперь банк считает, что он слишком консервативен, и хочет одобрять заявителей, набравших от 170 до 200 баллов. Если банк в прошлом никогда не одобрял этих заявителей, как он узнает возрастающий уровень риска при

снижении порога отсечения? Анализ отклоненных заявок поможет принять решение, позволяя оценить процент «плохих» по баллам отклоненных. Он дает возможность улучшить работу модели в будущем, создавая «набор для обмена данными». «Набор для обмена данными» – это обмен фактическими «плохими» с предполагаемыми «хорошими», как показано на рисунке 6.11. Предполагаемые «хорошие» - это те заявители, которые ранее были отклонены, но после проведения анализа отклоненных заявок определены как потенциальные «хорошие». Они являются заявителями, которые будут одобрены в будущем. В сочетании с отклонением фактических «плохих» (меняем фактических «плохих» на предполагаемых «хороших») это даст кредитной организации возможность одобрить то же самое количество заявителей, но при этом сделает работу модели эффективнее за счет более лучшего выбора. Факторы деловой среды, которые позволяют сделать процесс принятия решений более оптимальным и обоснованным, подчеркивают тот факт, что анализ отклоненных заявок – это больше, чем просто процедура для соблюдения статистических принципов, она имеет значимость для бизнеса.

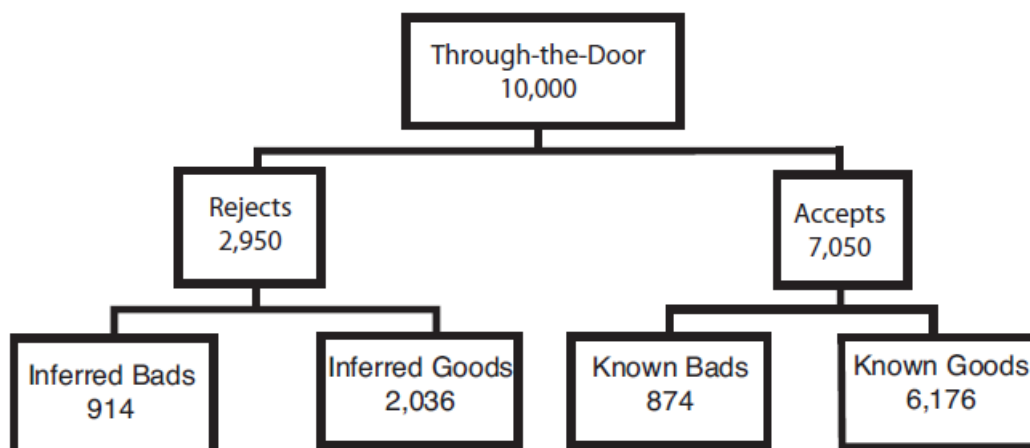
Рисунок 6.11 Набор для обмена данными

		Old Scorecard	
		Approve	Decline
New Scorecard	Approve	Known G	Inf G
	Decline	Known B	Inf B

Важно осознавать, что анализ отклоненных заявок прогнозирует неизвестное и будет всегда нести степень неопределенности. Уровень неопределенности можно снизить за счет использования лучших техник и здравого смысла. Пользователи должны понимать, что анализ отклоненных заявок может улучшить процесс принятия решений, а может и не улучшить, но в любом случае не будет на 100% точным.

Популяция, включающая как фактических, так и предполагаемых «хороших» и «плохих» («Все «Хорошие»/«Плохие») используется для получения итоговой скоринговой карты. После анализа отклоненных заявок выборка разбивается на подвыборки, как показано на рисунке 6.12. Основываясь на нашем предположении о текущем проценте одобрения 70.5% и проценте «плохих» в текущей популяции 12.4%, предполагаемая выборка показывает общий процент «плохих» 17.9% при проценте одобрения 100%. Заметим, что процент «плохих» для предполагаемой популяции составляет около 31% (914/2950). Это проверка релевантности, чтобы убедиться в том, что анализ отклоненных заявок был проведен корректно. Если предполагаемая популяция имеет более низкий процент «плохих», чем фактическая популяция, это может означать, что отклоненные фактически лучше по качеству, чем одобренные компанией.

Рисунок 6.12 Структура выборки после анализа отклоненных заявок



Когда следует применять анализ отклоненных заявок Влияние и важность анализа отклоненных заявок в ходе разработки скоринговой карты зависят от уровня одобрения заявителей и уровня достоверности критериев предоставления кредита. Очень высокий уровень достоверности в сочетании с высоким процентом одобрения (выдвигает предположение о том, что «все отклоненные=«плохие»»), и очень низкий уровень достоверности (предполагает почти случайную оценку) снижает необходимость анализа отклоненных заявок. В отношении заявителей с плохой кредитной историей или отсутствием кредитной истории даже средний процент одобрения позволяет сделать предположение «все отклоненные=«плохие»», если уровень достоверности оценок высок.

Анализ отклоненных заявок также менее важен в условиях высокого процента одобрения и соответственно высокого процента «плохих», поскольку популяция одобренных очень близка к популяции всех заявителей и может без ущерба использоваться для замещения популяции заявителей, пришедших «с улицы». То же самое верно, когда решения были приняты либо случайно, либо с использованием неточных инструментов. В ситуации с высоким процентом одобрения и очень низким процентом «плохих» – довольно трудной для вынесения оценки – можно без риска предположить, что все отклоненные являются «плохими». Здесь анализ отклоненных заявок не сможет провести значимую разницу между «хорошими» и «плохими».

В условиях низкого или среднего процента одобрения и низкого процента «плохих» анализ отклоненных заявок помогает найти возможности для увеличения доли рынка в соответствии со стратегиями, скорректированных с учетом риска. В данной ситуации, если кредитор понимает, что есть достаточно кредитоспособные заявители, которых он на данный момент отклоняет, упуская выгоду, анализ отклоненных заявок поможет ему найти подобных заявителей. Анализ отклоненных заявок будет иметь важное значение в тех случаях, когда есть уверенность в оценках, но при этом предполагается, что процент «плохих» можно снизить за счет более лучшего отбора.

Техники, применяемые в анализе отклоненных заявок

Для выполнения анализа отклоненных заявок существует различные техники. Обсудим ниже некоторые из них.

Рассматривать всех отклоненных как «плохих» Данный метод в целом не может быть признан удовлетворительным, потому что, опираясь на информацию кредитных бюро, мы знаем, что значительная часть отклоненных заявителей может оказаться «хорошими». Единственной ситуацией, где это было бы приемлемо – это когда процент одобрения очень высок, например, составляет 97%, и есть высокая степень уверенности в достоверности оценок. В этом случае с некоторой долей уверенности может быть сделано предположение о том, что все отклоненные – «плохие».

Рассматривать отклоненных в той же самой пропорции «хороших и плохих» в качестве учтенных среди одобренных Для обеспечения надежности этого метода нужно предположить, что нет совершенно никакой согласованности в текущем разбиении заявителей. Это подразумевает, что решения до настоящего времени принимались случайно.

Полностью игнорировать отклоненных Строим скоринговую модель только по одобренным заявителям. Это подразумевает 2-этапный процесс: (1) сперва выбираем одобренных и (2) оцениваем всех одобренных заявителей и отклоняем тех, кто оказался ниже заранее установленного порога. Данный подход означает систематический пересмотр текущей модели, что считается неэффективным, непопулярным среди разработчиков или среди тех, кто все еще уверен в эффективности текущей модели.

Одобрять всех заявителей Это единственный способ оценить фактическое (а не предполагаемое) исполнение обязательств отклоненными заявителями. Он подразумевает одобрение всех заявителей «с улицы» в течение определенного периода времени (например, три месяца). Позволяет собрать выборку из фактической популяции всех заявителей и оценить их кредитоспособность. Нет необходимости говорить, что одобренные заявители должны быть репрезентативны для всех диапазонов скорингового балла, чтобы не допустить переоценки или недооценки процента «плохих» по отклоненным заявителям. Хотя данный метод, возможно, самый научный и простой, сама мысль об одобрении заявителей с очень высокими рисками может отпугивать (по этой причине метод также известен как «покупка данных»). Однако нет необходимости одобрять любого заявителя в течение длительного периода времени. Учтите тот факт, что как минимум для разработки скоринговой карты потребуется около 2000 «плохих». Если 500 из них лежат ниже установленного порога отсека, а процент «плохих» заявителей ниже порога составляет 25%, то 2000 заявителей, находящихся ниже порога, нужно одобрить.

Настройки для получения данных следующие:

- Одобрять всех заявителей в течение определенного времени в количестве, достаточном для генерирования выборки в 500 заявителей.
- Одобрять всех заявителей выше порога, но при этом случайно отбираем заявителей ниже порога
- Одобрять всех заявителей, получивших 10-20 баллов ниже отсечки, а остальных отбираем случайно, чтобы получить лучшую выборку заявителей, находящихся в зоне принятия решений (где, вероятнее всего, должны быть установлены пороги отсека)

В условиях большого объема данных рекомендуется, чтобы данные по заявителям собирались в течение нескольких месяцев с целью минимизировать

сезонные колебания. Дальнейшая стратегия уменьшения убытков заключается в том, чтобы давать заявителям, находящимся ниже отсечки, меньшие суммы кредитов. Заслуживает внимания тот факт, что в определенных случаях могут возникнуть законные препятствия для реализации этого метода. Одобрение одних заявителей и отклонение других со схожими характеристиками или случайное одобрение заявителей может создать проблемы.

Метод, построенный на сходстве собственных данных и данных кредитного бюро Этот метод включает собственные данные по кредитоспособности заявителей, отклоненных по одному продукту, но одобренных по схожему продукту одним и тем же банком. Метод, близкий к нему, использует данные бюро о кредитоспособности заявителей, отклоненных одним банком, но одобренных по схожему продукту другим.

Например, банк может получить список заявителей, которые были отклонены по кредиту, но получили одобрение по кредитной карте в том же банке, или, похожая ситуация, когда банк-эмитент кредитных карт может получить список заявителей, которые были им отклонены, но получили одобрение в других банках-эмитентах. Обслуживание просрочки этими клиентами в другом банке-эмитенте или по схожему продукту в одном и том же банке может отслеживаться с помощью отчетов кредитных бюро и ежемесячным отчетам эффективности. Обслуживание обязательств по другим продуктам или в других банках рассматривается в качестве предположения о том, как отклоненные заявители погашали бы кредит, если бы их изначально одобрили.

Этот метод дает нам приближенное представление о фактическом исполнении кредитных обязательств, но у него есть несколько недостатков. Во-первых, органы надзора могут не позволить банку получать данные кредитного бюро об отклоненных заявителях (существует временной лимит). Однако это возможно, если отклоненный заявитель является текущим клиентом мультипродуктового банка, информация о котором отправляется в кредитное бюро и регулярно приобретается банком. С другой стороны, когда использование данных кредитного бюро строго регулируется, это может быть неосуществимо. Выбранные заявители должны получить схожий кредит в течение схожего периода времени (т.е. вскоре после того, как были отклонены). Кроме того, определение «плохой», полученное с помощью анализа фактических «хороших» и «плохих», должно применяться в отношении этих клиентов с использованием разных источников данных – что нелегко реализовать. Заявители, отклоненные одним банком или по одному продукту, вероятно, должны быть отклонены и другими банками, таким образом, снижается потенциальный объем выборки.

Приращение исторически оцененных данных (Soft Cutoff) Этот метод строится по одобренным заявителям, располагает их по скоринговым баллам и с помощью модели, построенной на одобренных, подсчитывает отклоненных в каждом скоринговом балле. Идея приращения состоит в том, что заявители, отклоненные в данном диапазоне скоринговых баллов, будут погашать кредит так же, что и те, кто был одобрен в этом диапазоне. В каждом диапазоне подсчитывается количество одобренных и отклоненных. Множитель приращения вычисляется так: $(A + R)/A$, где A – количество одобренных в диапазоне, а R – количество отклоненных в диапазоне.

Рисунок 6.13 Soft Cutoff

Диапазон	N Плохие	N Хорошие	Принятые	Отклоненные	Множитель приращения	Отклоненные-Плохие	Отклоненные-Хорошие
0–50	590	870	1460	1154	1,79	1056	1557
51–100	430	1115	1545	3258	3,11	1337	3468
101–150	255	1179	1434	1569	2,09	533	2464
151–200	224	2158	2382	2977	2,25	504	4856
201–250	125	2175	2300	895	1,39	174	3023
251–300	100	2890	2990	2594	1,87	187	5404
301–350	89	3401	3490	1257	1,36	121	4625
351–400	72	3891	3963	1107	1,28	92	4980
400+	59	4500	4559	987	1,22	72	5490
Overall	1944	22179	24123	15798	1,65	3208	36595

Например, есть 590 плохих и 870 хороших среди отклоненных в диапазоне 0-50. Множитель приращения в этом диапазоне равен $(1460+1154)/1460=1.79$. 1056 отклоненных будут определены как «плохие» ($590 \cdot 1.79$) и 1557 как «хорошие» ($870 \cdot 1.79$).

Простое приращение (Hard Cutoff). Данный метод состоит из следующих этапов:

1. Строим модель, используя фактических «хороших» и «плохих».
2. С помощью данной модели выставаем скоринговые баллы отклоненным, устанавливаем по ним ожидаемый процент «плохих» или вероятность «плохого» исхода.
3. Задаем ожидаемый процент «плохих», выше которого заявитель объявляется «плохим», а все заявители ниже этого уровня считаются «хорошими». Хорошая и надежная точка выбора - ожидаемый предельно допустимый процент «плохих» наихудшего заявителя, которого вы хотите одобрить.
4. Добавляем предполагаемых «хороших» и «плохих» к фактическим «хорошим» и «плохим», строим модель заново с учетом этого добавления.

Метод прост, но имеет несколько недостатков. Классификация отклоненных заявителей на «хороших» и «плохих» может быть произвольной, даже если можно использовать итерации с различными порогами отсека и простые правила для проверки разумности модели (например, процент «плохих» отклоненных должен быть в 2-4 раза выше, чем процент «плохих» одобренных). Скоринговая карта, дискриминирующая фактических «хороших» и «плохих», должна быть высокого качества, поскольку только она используется для назначения класса кредитоспособности. Этот метод также не берет в расчет вероятность отклоненного заявителя быть одобренным, поэтому отклоненные заявители объединяются с фактическими в пропорции 1:1. Следующий метод пытается устранить этот отдельный недостаток.

Приращение 2⁶ (Augmentation 2⁶) Этот метод присваивает веса фактическим «хорошим» и «плохим», оценивая вероятность одобрения (т.е. вероятность заявителя быть включенным в фактическую популяцию). Он включает два этапа:

1. Строим модель одобренных/отклоненных, чтобы получить вероятность одобрения или отклонения для каждого наблюдения.

- Используя только фактических «хороших» и «плохих» (т.е. одобренных), строим модель «хороших» и «плохих» на основе популяции, скорректированной с помощью установленных весов для одобренных и отклоненных. Это делается так, чтобы новые веса наблюдений были обратно пропорциональны вероятности одобрения. Таким образом, наблюдения взвешиваются, чтобы более точно представлять общую популяцию заявителей.

Метод признает необходимость установить вероятность одобрения. Он лучше, чем простое приращение. Схожая техника, которая использует отклоненных, освещается ниже под названием «Нечеткое приращение».

Парцелинг (Parceling) Этот метод похож на простое приращение, но вместо классификации всех отклоненных на «хороших» и «плохих» в заданном диапазоне, он распределяет их пропорционально ожидаемому проценту «плохих» в данном диапазоне. На рисунке 6.14 приведен пример.

Рисунок 6.14 Парцелинг

Диапазон	N Плохие	N Хорошие	% Плохие	% Хорошие	Отклоненные	Отклоненные-Плохие	Отклоненные-Хорошие
0–169	290	971	23.0%	77.0%	1646	379	1267
170–179	530	2414	18.0%	82.0%	1732	312	1420
180–189	365	2242	14.0%	86.0%	3719	521	3198
190–199	131	1179	10.0%	90.0%	7334	733	6601
200–209	211	2427	8.0%	92.0%	1176	94	1082
210–219	213	4047	5.0%	95.0%	3518	176	3342
220–229	122	2928	4.0%	96.0%	7211	288	6923
230–239	139	6811	2.0%	98.0%	3871	77	3794
240–249	88	10912	0.8%	99.2%	4773	38	4735
250+	94	18706	0.5%	99.5%	8982	45	8937

Первые четыре столбца в таблице – это распределение фактических «хороших» и «плохих» по баллам, полученным с помощью скоринговой карты. Столбец «Отклоненные» представляет собой распределение отклоненных заявителей по баллам, вычисленным с помощью скоринговой карты на основе фактических «хороших» и «плохих». Два последних столбца представляют случайное распределение отклоненных заявителей, оцененных моделью, на «хороших» и «плохих».

Например, есть 7334 отклоненных заявителя в диапазоне 190-199, ожидаемый процент «плохих» равен 10%, 733 отклоненных заявителя будут определены как «плохие» ($7334 \cdot 0.10$) и 6601 как «хорошие» ($7334 \cdot 0.90$). Назначение классов внутри каждого диапазона является случайным.

Тем не менее с точки зрения бизнеса разумно предположить, что пропорция «хороших» и «плохих» среди отклоненных заявителей не может быть такой же, как и среди одобренных (отклоненные должны быть хуже по исполнению кредитных обязательств). Консервативный подход относит большую долю отклоненных к «плохим». Здесь могут использоваться итеративный подход и эмпирические правила. Как минимум, процент «плохих» отклоненных должен быть в 2-4 раза больше, чем процент «плохих» одобренных.

Этот метод также достаточно быстрый и простой для внедрения. Как и в случае с простым приращением, скоринговая карта, дискриминирующая фактических «хороших» и «плохих», должна быть высокого качества, поскольку только она

используется для назначения класса кредитоспособности. Кроме того, распределение нуждается в корректировке (например, на основе консервативного подхода, упомянутого выше) так, чтобы процент «плохих» отклоненных заявителей не был недооценен.

Нечеткое приращение (Fuzzy Augmentation) Этот метод довольно схож с обычным приращением, но вместо назначения классов «хороший» или «плохой», он причисляет каждого отклоненного к частично «хорошему» и частично «плохому» классу. Процедура включает в себя классификацию и приращение, поэтапно:

1. Классификация.
 - Выставляем баллы отклоненным заявителям с помощью модели, построенной по фактическим «хорошим» и «плохим».
 - Вычисляем вероятность «хорошего» исхода и вероятность «плохого» исхода для каждого отклоненного, опираясь на ожидаемый процент «плохих».
 - Назначаем каждого отклоненного частично «хорошим» и частично «плохим» (т.е. создаем два наблюдения с весами для каждого отклоненного, чтобы в сумме эти две записи давали вес 1).
 - Взвешиваем отклоненных «хороших», используя вероятность «хорошего» исхода, и отклоненных «плохих», используя вероятность «плохого» исхода. Как и в случае с парцелингом, каждому отклоненному может быть присвоена большая вероятность «плохого» исхода в качестве консервативной меры.
2. Приращение.
 - Объединяем отклоненных с одобренными, внося поправку на процент одобрения, вероятность одобрения.
 - Например, частота «хорошего» среди отклоненных = вероятность «хорошего» исхода \times вес, где «вес» – вероятность отклоненного заявителя быть включенным в дополненный набор данных.

Дополнительное взвешивание на этапе приращения выполняется в силу того, что объединение одобренных и отклоненных по принципу «один-к-одному» подразумевает равные шансы попасть в набор данных.

Этот метод включает не просто вероятность «плохого» исхода для отклоненного заявителя, но прежде всего вероятность быть одобренным. Этот метод лучше, т.к. он позволяет учесть степень важности отклоненных заявителей в итоговой выборке. Кроме того, применение частичной классификации лучше произвольной.

Итеративная переклассификация⁷ Этот метод сначала строит модель фактических «хороших» и «плохих», присваивает классы, используя вероятность «плохого» исхода (как в простом приращении), объединяет отклоненных с одобренными, повторяя этот процесс до тех пор, пока не будет достигнут некоторый порог сходимости. Включает следующие этапы:

1. Строим модель фактических «хороших» и «плохих».
2. Выставляем скоринговые баллы отклоненным и назначаем класс, опираясь на минимально ожидаемый процент «плохих» или выбранную вероятность «плохого» исхода.
3. Объединяем отклоненных и одобренных, а затем перестраиваем скоринговую карту.
4. Заново выставляем баллы отклоненным и переназначаем класс, затем объединяем и перестраиваем скоринговую карту.

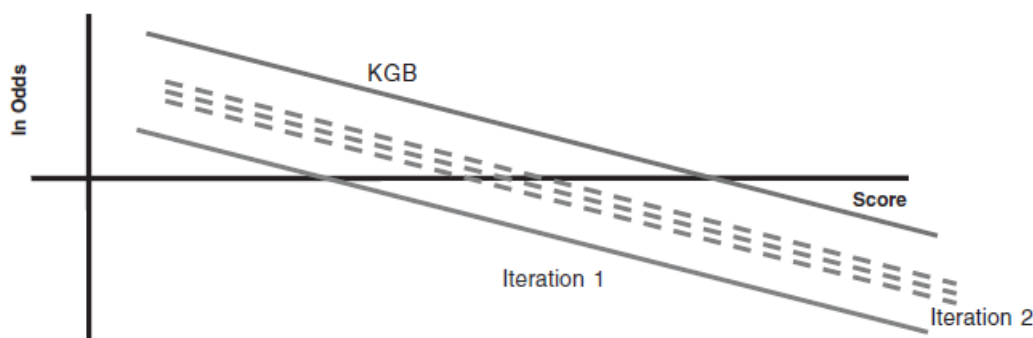
5. Повторяем этот процесс до тех пор, пока оценки параметров (или вероятность «плохого» исхода) не сойдутся.

Отметим, что один из способов модифицировать этот метод – применение классификации частично «хороших» и частично «плохих» вместо классификации на основе произвольно выбранной вероятности «плохого» исхода.

Сходимость может измеряться с помощью оценок параметров или вероятности «плохого» исхода для каждого диапазона скоринговых баллов или каждого прохода алгоритма, или с помощью графика зависимости логарифма шансов от скорингового балла, как показано на рисунке 6.15. Пунктирные линии соответствуют итерациям.

Каждая итерация должна быть ниже линии, соответствующей набору фактических «хороших» и «плохих», подтверждая тот факт, что объединенная популяция имеет более высокий процент «плохих», чем отдельно одобренные. Если линия объединенного набора данных выше линии фактического набора, это может означать, что отклоненные заявители выше по качеству исполнения обязательств, чем одобренные.

Рисунок 6.15 Анализ отклоненных заявок с использованием итеративной переклассификации



Метод ближайшего соседа (Кластеризация) Данная техника использует кластеризацию для определения «хороших» и «плохих» в выборке отклоненных и не зависит от ранее построенных моделей. Техника достаточно проста:

1. Создаем два набора кластеров – один для фактических «хороших», второй – для фактических «плохих».
2. Оцениваем отклоненных по обоим кластерам.
3. Сравниваем евклидовы расстояния, чтобы присвоить наиболее вероятный класс кредитоспособности (т.е. если отклоненный заявитель ближе к кластеру «хороших», чем к кластеру «плохих» то, вероятно, он является «хорошим»).
4. Объединяем одобренных и отклоненных, чтобы создать набор, учитывающий предполагаемых заявителей, и перестраиваем модель.

В качестве дополнительных мер по улучшению метода может быть введена поправка на вероятность получить одобрение, когда создается набор, включающий предполагаемые данные, и выполнены частичные классификации. Данный метод сравнительно прост. Его недостаток в том, что используемые метрики являются условными – по сравнению с другими параметрами, обсуждаемыми выше.

Рассуждения, основанные на памяти (Memory-Based Reasoning) Рассуждения, основанные на памяти, или рассуждения, основанные на прецедентах, предполагают 2-этапный процесс назначения классов. Сначала метод определяет схожие наблюдения, например, «хороших» и «плохих» в выборке. Затем, обучившись на первом этапе, он назначает класс новому наблюдению. Этот процесс имитирует ход мыслей людей, которые проводят анализ отклоненных заявок. Например, кредитный аналитик рассматривает большое количество фактических «хороших» и «плохих» клиентов, отмечая характеристики тех и других. Затем он оценивает отклоненных и классифицирует их по схожим характеристикам на фактических «хороших» и «плохих».

Метод включает следующие этапы:

1. Выполнить кластеризацию, используя фактических «хороших», «плохих» и отклоненных. SAS Enterprise Miner, например, применяет метод k ближайших соседей, чтобы классифицировать наблюдения. Для этого используется Memory Based Reasoning Node.
2. Затем измерить расстояние между каждым наблюдением – фактическим заявителем («хорошим» или «плохим»), и наблюдением – отклоненным заявителем. K фактических наблюдений («хороших» или «плохих»), которые имеют наименьшее евклидово расстояние до наблюдения – отклоненного заявителя, являются для него k ближайшими соседями.
3. Целевые классы («хорошие»/«плохие») k ближайших соседей используются для определения вероятности «хорошего» или «плохого» исхода для каждого отклоненного.
4. Создаем объединенную выборку и перестраиваем модель.

Например, если значение $k=10$, а из 10 соседей, ближайших к отклоненному заявителю, 7 являются «хорошими» и 3 «плохими», то итоговая вероятность «плохого» исхода для данного отклоненного заявителя равна 30%. Можно использовать либо частичную классификацию каждого отклоненного заявителя, либо полную классификацию на основе некоторой вероятности «плохого» исхода.

Проверка результатов

После того, как анализ отклоненных заявок завершен, может быть выполнен простой анализ для проверки результатов. Он включает:

- Сравнение процента «плохих»/шансов предполагаемых и фактических заявителей и применение практических правил, обсуждавшихся выше. Анализ отклоненных заявок может проводиться с использованием различных параметров до тех пор, пока он не будет удовлетворять этим правилам. Правила следует применять, опираясь на процент одобрения и степень уверенности в надежности прежней системы выдачи кредитов. Например, если прежняя система была хорошей, процент одобрения был высоким, то процент «плохих» отклоненных должен быть по крайней мере в 3-4 раза выше, чем процент «плохих» одобренных. Средний процент одобрения допускает, что процент «плохих» отклоненных может быть лишь в 2 раза выше, чем процент «плохих» одобренных.
- Сравнение WOE или процент «плохих» для сгруппированных атрибутов в наборе данных до и после проведения reject inference. Атрибуты с низким процентом одобрения и высокими WOE должны показать изменения WOE, согласующиеся с бизнес-опытом или объяснимые другим способом.

Различные техники анализа отклоненных заявок, их параметры могут тестироваться с помощью «ненастоящих» отклоненных. Это подразумевает случайное разбиение выборки одобренных на одобренных и отклоненных, например 70/30. Модель, построенная на 70% «одобренных», затем может использоваться для оставшихся 30%. Поскольку фактическая кредитоспособность 30% «отклоненных» уже известна, можно использовать ошибочную классификацию, чтобы оценить эффективность применяемой техники.

После того, как анализ отклоненных заявок завершен, создается объединенный набор данных (из одобренных и отклоненных) и используется на следующем этапе разработки скоринговой карты. Теперь проблема смещения выборки решена, итоговая скоринговая карта отражает популяцию заявителей, пришедших «с улицы».

Финальный этап разработки скоринговой карты

Итоговые скоринговые карты разрабатываются по данным, полученным после анализа отклоненных заявок, с использованием того же самого анализа исходных характеристик и тех же самых статистических алгоритмов (например, регрессия). Итог - окончательный набор характеристик для скоринговой карты. Отметим, что на данном этапе вы не ограничены характеристиками предварительной скоринговой карты. После проведения анализа отклоненных заявок некоторые характеристики могут оказаться более слабыми, некоторые более сильными, таким образом, процесс выбора переменных необходимо повторить.

На данном этапе допустим, что анализ исходных характеристик и логистическая регрессия для построения итоговой скоринговой карты проводится по популяции всех заявителей, классифицированных на «хороших» и «плохих». У нас есть теперь набор характеристик наряду с результатами логистической регрессии – константой, оценками параметров и статистическими критериями качества модели.

Задачи, которые теперь нужно решить, - это масштабирование скоринговых баллов, достоверность распределения баллов, ошибочная классификация и предсказательная сила скоринговой карты.

Масштабирование

Скоринговые карты могут разрабатываться в различных форматах. В некоторых случаях использование скоринговой карты в режиме онлайн или реального времени часто зависит от платформы, инструкций надзорных органов, касающихся обоснования причин отказа в кредите, легкости в использовании и других факторов, упомянутых в разделе «Формат скоринговой карты» главы 3. Скоринговая карта должна быть представлена в особом формате (смотрите рисунок 1.1). Для этого нужно применить масштабирование. Оно включает ранжирование и формат скоринговых баллов, устанавливает масштаб изменения шансов в зависимости от скорингового балла. Скоринговые баллы могут быть представлены в виде десятичных или дискретных чисел:

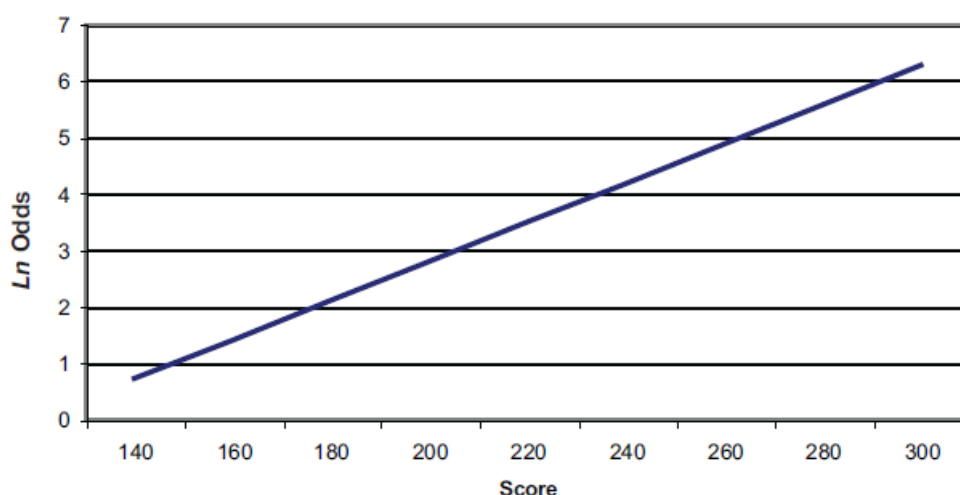
- Где скоринговый балл – это отношение шансов «хороших» к «плохим» или вероятность «плохого» исхода
- Где есть некая заданная числовая шкала с минимумом и максимумом (например, 0–1000, 150–350) с установленным отношением шансов «хороших» к «плохим» в заданной точке (например, отношение шансов 5:1 в точке 500 баллов) и масштаб изменения шансов стать «хорошим» (например, каждые 50 баллов удваивают шансы стать «хорошим»)

Масштабирование не влияет на предсказательную силу скоринговой карты. Оно продиктовано следующими соображениями:

- Удобство внедрения скоринговой карты.
- Легкость использования для обслуживающего персонала (например, с дискретными числами проще работать).
- Преемственность с существующими скоринговыми картами или другими скоринговыми картами в банке. Это позволяет избежать обучения персонала по использованию скоринговой карты и интерпретации баллов.

Существуют различные техники масштабирования. Наиболее распространенной является скоринговая карта с дискретными числами, отмасштабированных логарифмически, где каждые 20 баллов удваивают шансы стать «хорошим» заявителем. Пример такого масштабирования приведен на рисунке 6.16.

Рисунок 6.16 Отмасштабированные скоринговые баллы



Расчет масштабирования В целом, зависимость между отношением шансов и скоринговым баллом можно представить в виде линейного преобразования:

$$\text{Балл} = \text{Смещение} + \text{Множитель} * \ln(\text{шансы})$$

Когда скоринговая карта разрабатывается, используя заданное отношение шансов в какой-либо точке и установленное количество баллов, которое удваивает шансы («points to double the odds» или pdo), множитель и константа могут быть легко вычислены с помощью следующих уравнений:

$$\text{Балл} = \text{Смещение} + \text{Множитель} * \ln(\text{шансы})$$

$$\text{Балл} + \text{pdo} = \text{Смещение} + \text{Множитель} * \ln(2 * \text{шансы})$$

Решив уравнения выше, получаем:

$$\text{pdo} = \text{Множитель} * \ln(2), \text{ отсюда}$$

$$\text{Множитель} = \text{pdo} / \ln(2);$$

$$\text{Смещение} = \text{Балл} - \{\text{Множитель} * \ln(\text{шансы})\}$$

Например, если скоринговая карта масштабировалась так, что в точке 600 баллов отношение шансов составляет 50:1 и каждые 20 баллов удваивают шансы (т.е. pdo=20), то множитель и константа будут:

$$\begin{aligned}\text{Множитель} &= 20 / \ln(2) = 28.8539 \\ \text{Смещение} &= 600 - \{28.8539 \ln(50)\} = 487.123\end{aligned}$$

Теперь можно вычислить балл, соответствующий заданному отношению шансов (или атрибуту):

$$\text{Балл} = 487.123 + 28.8539 \ln(\text{шансы})$$

Та же самая формула может использоваться для масштабирования скоринговой карты, в которой определенное количество баллов утраивает или учетверяет шансы. Однако количество баллов, которое удваивает шансы, является самым распространенным в кредитном скоринге.

Данная формула может использоваться для подсчета баллов для любого наблюдения, используя любую модель, выдающую вероятность «плохого» исхода или шансы. Она позволит включить модели, использующие другие техники (помимо регрессии), как это было описано в данной книге.

Поскольку скоринговая карта разрабатывалась с использованием WOE в качестве входных характеристик, вышеописанную зависимость можно представить в виде:

$$\text{Балл} = \ln(\text{шансы}) * \text{Множитель} + \text{Смещение} =$$

$$= \sum_{j,i=1}^{k,n} \left(-(woe_i * \beta_i + \frac{a}{n}) * \text{Множитель} + \frac{Offset}{n} \right)$$

где:

WOE – это WOE каждого сгруппированного атрибута

β – регрессионный коэффициент каждой характеристики

a – константа

n – количество характеристик

k – количество групп (атрибутов) в каждой характеристике

Данная формула позволяет вычислить баллы, присвоенные каждому сгруппированному атрибуту по каждой характеристике в скоринговой карте. Суммирование всех баллов по каждому атрибуту даст итоговый балл. Здесь стоит отметить, что на баллы, назначаемые с помощью этого метода, влияют тренд WOE и разница между WOE сгруппированных атрибутов. Это подчеркивает важность как логической обоснованности тренда, так и максимизации разницы между WOE групп.

Коды причин отказа (Adverse Codes) В некоторых юрисдикциях, особенно в США, кредитные организации обязаны раскрывать заявителям причины, по которым отказали им в выдаче кредита. Это осуществляется с помощью кодов причин отказа. Используя методику распределения баллов, описанную выше, коды причин отказа можно получить прежде всего с помощью «нейтральных баллов». Нейтральным будет балл, у которого WOE равно 0. После того, как

вычислены множитель и смещение, в уравнение выше можно подставить WOE=0 и получить нейтральный балл. Уравнение для нейтрального балла имеет вид:

$$-\left(\frac{a}{n}\right) * \text{Множитель} + \frac{\text{Offset}}{n}$$

Характеристика, у которой скоринговый балл ниже нейтрального балла, является потенциальной причиной отказа, т.к. вероятность «плохого» исхода составит более 50% (заметим, что при WOE=0 вероятность «хорошего» или «плохого» исхода равна 50%). На рисунке 6.17 приведен пример того, как заявителю выставляют баллы на основе скоринговой карты. Также приведен нейтральный балл для данной скоринговой карты.

Рисунок 6.17 Анализ причин отказа по нейтральному баллу

Скоринговая карта	
Возраст	56
Период проживания по месту жительства	43
Почтовый индекс	22
Запросы за последние 3 месяца	20
% кредитных линий с просрочками	43
Возраст первой кредитной линии	68
Коэффициент обслуживания долга	42
Утилизация	25
Наихудший рейтинг	30
Нейтральный балл	31

Исходя из расчета баллов по данному заявителю, тремя главными причинами отказа будут «Запросы за последние 3 месяца», «Почтовый индекс» и «Утилизация». Они имеют самый низкий скоринговый балл, лежащий ниже нейтрального балла.

Некоторые организации вычисляют «нейтральный балл» на основе средневзвешенного балла каждого атрибута. На рисунке 6.18 приведен пример расчета средневзвешенного значения для характеристики «Срок проживания».

Средневзвешенное среднее вычисляется по следующей формуле:

$$\sum_{i=1}^n (\text{Распределение}_i * \text{Балл}_i)$$

Рисунок 6.18 Нейтральный балл, вычисленный с помощью средневзвешенного значения

Срок проживания	Распределение	Балл	Распределение * Балл
0–6	18%	12	2.16
7–18	32%	25	8
19–36	26%	28	7.28
37–	24%	40	9.6
Взвешенное среднее			27.04

Причина отказа формируется по атрибутам, у которых баллы оказались ниже вычисленного средневзвешенного значения. Ранжирование осуществляется с помощью расчета процентной разницы между баллом заявителя и средневзвешенным значением, по всем характеристикам, по которым заявитель получил балл ниже средневзвешенного значения – выбираются три наибольших отклонения станут тремя главными причинами отказа.

Распределение баллов

После того, как скоринговая карта построена, нужно проверить распределение баллов по каждому атрибуту и общую предсказательную силу скоринговой карты. Распределение баллов должно быть логичным, соответствуя трендам, выявленным в ходе анализе исходных характеристик. На рисунке 6.19 приведен пример распределений баллов по «возрасту», взятых из двух отдельно разработанных скоринговых карт.

Скоринговая карта 1 имеет логичное распределение баллов: с увеличением возраста заявители получают больше баллов. Это соответствует распределению весов атрибутов и бизнес-опыту. Однако, скоринговая карта 2 в атрибуте «27-29» показывает смену зависимости. Это могло быть вызвано корреляцией или странностями данных. Также это происходит, если две группы имеют веса, недостаточно далеко расположенные друг от друга. Отметим, что если следовать методу, изложенному в этой книге, применив логическую группировку и регрессию с использованием WOE, данная смена зависимости не произойдет. Эта тенденция проявляется, когда в качестве входных переменных регрессии используются сырые данные и затем назначаются баллы с помощью других методов. Поскольку это единственная смена зависимости в данной характеристике, а распределение остальных баллов логично, то обычно выполняется субъективная корректировка распределения баллов. В зависимости от важности смены зависимости и порядка распределения баллов по остальным атрибутам, может потребоваться перегруппировка характеристик и пересмотр метода пошаговой регрессии.

Последняя процедура является итеративным процессом, продолжающимся до тех пор, пока не будет получена статистически значимая и удобная в применении скоринговая карта.

Рисунок 6.19 Логичное распределение баллов

Возраст	Вес	Скоринговая карта 1	Скоринговая карта 2
Missing	-55.50	16	16
18-22	-108.41	12	12
23-26	-72.04	18	18
27-29	-3.95	26	14
30-35	70.77	35	38
35-44	122.04	43	44
44+	165.51	51	52

Выбор скоринговой карты

Большинство разработчиков скоринговых карт должны получить по крайней мере 2-3 различные скоринговые карты. Разработка нескольких различных скоринговых карт облегчает выбор, обусловленный объемом контроля и гибкостью метода разработки, описанного в этой книге. Выбор итоговой скоринговой карты среди прочих подразумевает два вопроса: «Какая скоринговая карта является лучшей?», «Насколько хорошей является скоринговая карта?» Ответы можно получить, используя статистические показатели в сочетании с бизнес-опытом.

Ошибочная классификация

Скоринговые карты разрабатываются для прогноза вероятности «хорошего» или «плохого» исхода. Еще более важно то, что они используются для дифференциации между «хорошими» и «плохими». Ошибочная классификация – это хороший способ определить, осуществляет ли скоринговая карта правильную дифференциацию. Многие организации обычно выбирают в качестве порога

отсечения минимально приемлемый процент «плохих» (опираясь на скоринговый балл). Заявители, получившие балл ниже порога отсечения получают отказ в кредите или рассматриваются как потенциальные мошенники. В результате всегда есть вероятность того, что фактический «хороший» может быть классифицирован как «плохой» и получить отказ в кредите и, наоборот. То же самое верно для поведенческих скоринговых карт, в которых порог отсечения используется для того, чтобы определить положительную или отрицательную динамику по определенным клиентам. Итоговую скоринговую карту нужно выбирать так, чтобы минимизировать уровень ошибочной классификации.

Есть несколько показателей, используемых для оценки уровня ошибочной классификации и сравнения различных скоринговых карт. Эти показатели сравнивают в заданном пороге отсечения количество фактических «хороших» и «плохих» с количеством спрогнозированных «хороших» и «плохих». «Хорошими» и «плохими» являются наблюдения, лежащие выше и ниже выбранного порога соответственно.

Эти показатели основываются на матрице ошибок (или ошибочной классификации), приведенной на рисунке 6.20.

Лучшей будет та скоринговая карта, у которой максимально возможное количество «истинных» исходов и минимально возможное количество «ложных» исходов.

Рисунок 6.20 Матрица ошибок

	Спрогнозировано	
Фактически	Положительные исходы	Отрицательные исходы
Положительные исходы	ТР истинно положительные («хороший» классифицирован как «хороший»)	FN ложно отрицательные («хороший» классифицирован как «плохой») ОШИБКА I РОДА
Отрицательные исходы	FP ложно положительные («плохой» классифицирован как «хороший») ОШИБКА II РОДА	TN истинно отрицательные («плохой» классифицирован как «плохой»)

Отрицательными исходами будут «Плохие», а положительными – «Хорошие».

Положительные исходы состоят из истинно положительных (ТР) и ложно положительных (FP).

Отрицательные исходы состоят из ложно отрицательных (FN) и истинно отрицательных (TN).

ТР (True Positives) или истинно положительные исходы – это верно классифицированные положительные исходы. «Хороший» прогнозируется как «хороший».

TN (True Negatives) или истинно отрицательные исходы – это верно классифицированные отрицательные исходы. «Плохой» прогнозируется как «плохой».

FN (False Negatives) или ложно отрицательные исходы — это положительные исходы, неверно классифицированные как отрицательные (ошибка I рода). «Хороший» ошибочно прогнозируется как «плохой». Это так называемый «ложный пропуск», когда интересующее нас событие ошибочно не обнаруживается.

FP (False Positives) или ложно положительные исходы — это отрицательные исходы, неверно классифицированные как положительные (ошибка II рода). «Плохой» ошибочно прогнозируется как «хороший». Это «ложная тревога», когда при отсутствии события ошибочно выносится решение о его присутствии.

Существует четыре главных показателя для оценки ошибочной классификации:

- **Специфичность (Specificity)** — доля истинно отрицательных исходов от общего числа фактических отрицательных исходов. Кратко, это доля отрицательных исходов у «плохих» клиентов. Речь идет о способности модели правильно определять тех клиентов, у которых дефолт *есть*. Модель с высокой специфичностью характеризуется ужесточенным отбором клиентов — максимальным предотвращением пропуска «плохих», является консервативной. Ее задача — минимизировать кредитный риск, связанный с выдачей кредита «плохому» клиенту (уменьшить вероятность совершения ошибки II рода).
- **Чувствительность (Sensitivity)** — доля истинно положительных исходов от общего числа фактических положительных исходов. Проще говоря, это доля положительных исходов у «хороших» клиентов. Речь идет о способности модели правильно определять, у кого *нет* дефолта. Модель с высокой чувствительностью характеризуется менее тщательным отбором клиентов — максимальным предотвращением пропуска «хороших», является рискованной. Ее задача — минимизировать упущенную выгоду, связанную с отказом «хорошему» клиенту в выдаче кредита (уменьшить вероятность совершения ошибки I рода).
- **Точность (Accuracy)** — доля истинно положительных и истинно отрицательных исходов от общего количества исходов.
- **Уровень ошибки (Error rate)** — доля ложно положительных и ложно отрицательных исходов от общего количества исходов.

Основываясь на этих показателях, компания может, например, максимизировать отказ «плохим». В этом случае, когда скоринговые карты строятся для снижения потерь, следует выбрать скоринговую карту, которая максимально увеличивает специфичность. В ситуации, когда компания желает заполучить *большую* долю рынка, пренебрегая одобрением некоторых «плохих», можно минимизировать отказ «хорошим». Для этого выбирают скоринговую карту, которая максимально увеличивает чувствительность. Здесь показатели специфичности и чувствительности используются в разрезе бизнес-задач, поставленных перед скоринговой картой.

Когда для сравнения было разработано несколько моделей, то эти показатели необходимо получить для каждой модели, основываясь на аналогичных порогах отсека (например, при итоговом проценте одобрения 70% или проценте «плохих» 5%).

Отметим, когда скоринговая карта была разработана без настройки оверсемплинга, цифры в матрице ошибок нужно скорректировать так, чтобы они отражали настоящие частоты. Это выполняется путем умножения частот в ячейках на выборочные веса π_1 и π_2 , как показано на рисунке 6.21.

Рисунок 6.21 Матрица ошибок для нескорректированных выборок

Фактически	Спрогнозировано	
	Положительные исходы	Отрицательные исходы
Положительные исходы	$n^*(\text{Факт. } P_s / \text{Текущ. } P_s) * p_1$	$n^*(1 - \text{Чувствительность}) * p_1$
Отрицательные исходы	$n^*(1 - \text{Специфичность}) * p_0$	$n^*(\text{Специфичность}) * p_0$

Сила скоринговой карты

Сила скоринговой карты – это статистический показатель, измеряющий предсказательную способность скоринговой карты. Большая часть показателей нужна для сравнения скоринговых карт. Эти метрики обычно используются в сочетании с показателями ошибочной классификации и стратегическими задачами, чтобы выбрать итоговую наиболее предпочтительную скоринговую карту.

В некоторых случаях ранее использовавшаяся или текущая скоринговые карты сравниваются с новыми разработанными. Нужно проявлять осторожность в этих сравнениях, т.к. изменения в данных, профилях заявителей и маркетинговых стратегиях могут сделать эти сравнения нерелевантными или в лучшем случае слабыми. Например, если текущая скоринговая карта стала нестабильной и непригодной для использования, необоснованно будет сравнивать, например, KS данной скоринговой карты с KS старой скоринговой карты на тот момент, когда она была разработана. Скоринговые карты следует разрабатывать в меру возможностей, предоставляемых данными. Некоторые компании обращаются к прошлому, используют как новую, так и существующую скоринговую карту, а затем изучают эффективность обеих скоринговых карт. Кроме того, если текущая скоринговая карта не является стабильной, то процедура сравнения совершенно не учитывает нынешние обстоятельства. Нужно провести бенчмаркинг текущей модели, лучшим способом была бы разработка модели с использованием нескольких методов на одних и тех же данных. Например, можно построить скоринговую карту с использованием логистической регрессии по сгруппированным переменным, логистической регрессии по сырым данным, нейронных сетей и дерева решений. Таким образом, будет лучше обозначено преимущество или слабость выбранного метода и модели.

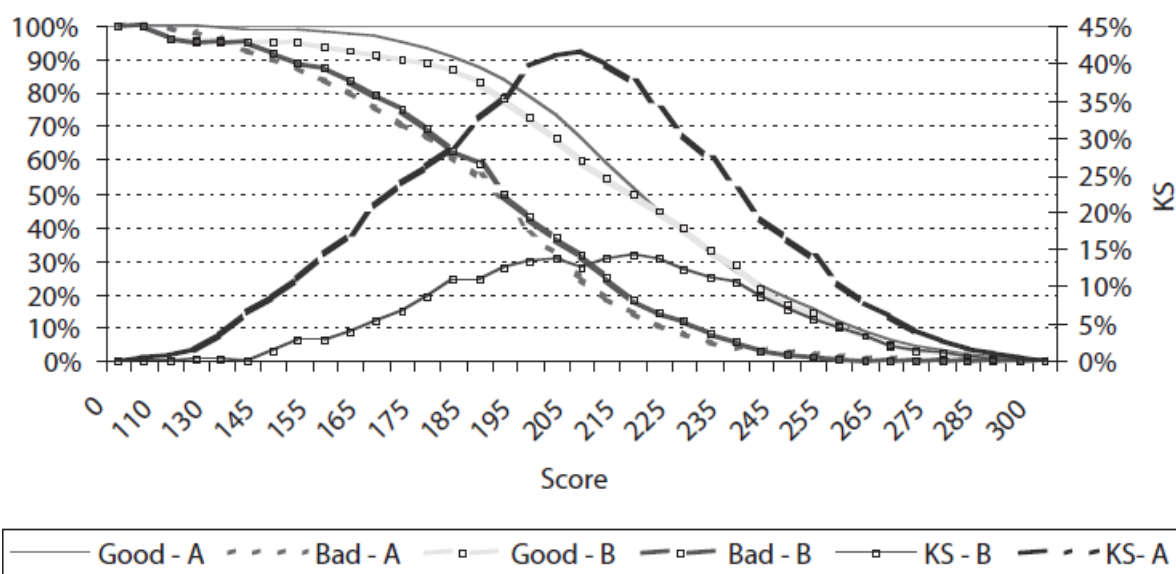
Методы, использующиеся для сравнения предсказательной способности скоринговой карты, включают следующие показатели:

- **Информационный критерий Акаике (Akaike's Information Criterion, AIC)**
Штрафует за добавление излишнего количества параметров модели. Предпочтительны наименьшие значения AIC.
- **Байесовский критерий Шварца (Schwarz's Bayesian Criterion, SBC).** SBC также штрафует за добавление излишнего количества параметров модели. Предпочтительны наименьшие значения AIC. В случае скоринговых карт «профиля риска», построенных с использованием методов, описанных в данной книге, SBC и AIC не могут являться наилучшими критериями для

оценки предсказательной силы, поскольку приоритетом в данной ситуации является скорее скоринговая карта, построенная на широком наборе характеристик, чем скоринговая карта с минимальным числом критериев.

- **Статистика Колмогорова-Смирнова (КС).** Измеряет максимальную разницу между кумулятивными распределениями «хороших» и «плохих». Слабость этой метрики заключается в том, что разница измеряется только в одной точке (которая может и не совпадать с ожидаемым порогом отсеечения), а не по всему диапазону скоринговых баллов. Точка достижения максимального значения КС обычно находится ближе к середине диапазона. Поэтому если предполагаемый порог скоринговой карты находится в нижней (начальной) или верхней (конечной) части диапазона, то данная метрика не может быть хорошим критерием для сравнения карт. В таких случаях лучше сравнивать отклонение в предполагаемом пороге отсеечения.

Рисунок 6.22 Статистика Колмогорова-Смирнова



Теоретически КС может принимать значения от 0 до 100, однако на практике она обычно оказывается в диапазоне от 25 до 75. Обычно руководствуются правилом:

- меньше 20 - скоринговая карта непригодна к применению
- 20-40 – среднее качество скоринговой карты
- 41-50 – хорошее качество скоринговой карты
- 51-60 – очень хорошее качество скоринговой карты
- 61-75 – отличное качество скоринговой карты
- больше 75 - вероятно, были допущены ошибки в разработке скоринговой карты.

На рисунке 6.22 показан расчет КС для двух скоринговых карт, где максимальное значение метрики KS достигается в диапазоне 205-215. Скоринговая карта «А» имеет большую предсказательную силу, чем скоринговая карта «В», поскольку

«А» имеет максимальную разницу около 41% по сравнению с «В», у которой максимальная разница 15%.

- **с-статистика.** Один из самых мощных непараметрических двухвыборочных тестов, данная метрика эквивалентна площади под ROC-кривой, коэффициенту Джини и тесту Уилкоксона-Манна-Уитни. Она измеряет эффективность классификатора по всему диапазону скоринговых баллов и является наилучшей метрикой для оценки общей предсказательной силы скоринговой карты. С-статистика измеряет площадь под кривой *Чувствительность в сопоставлении с (1 - Специфичность)* для всего диапазона скоринговых баллов. На рисунке 6.23 приведен пример сравнения скоринговых карт с использованием с-статистики. На нем видно, что скоринговая карта «А» имеет большую предсказательную силу, поскольку имеет большую площадь под ROC-кривой, чем скоринговая карта «В».

Линия «случайного» классификатора обозначает с-статистику, равную 0.5. По причине того, что скоринговая карта должна быть лучше, чем «случайный» классификатор, с-статистика должна быть выше 0.5. В большинстве случаев, когда используются данные высокого качества, с-статистика 0.7 и выше может рассматриваться как вполне пригодная.

В ряде программных продуктов, предназначенных для разработки скоринговых карт, таких, как SAS Credit Scoring, данная статистика вычисляется автоматически. Программный код, приведенный на рисунке 6.24, - это пример кода SAS для вычисления с-статистики.

Рисунок 6.23 ROC-кривая

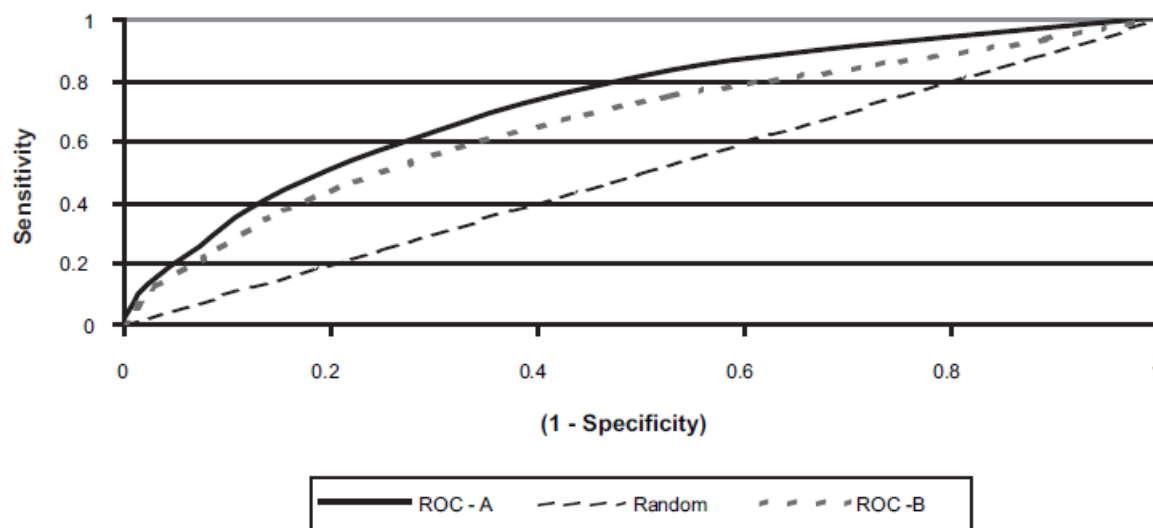


Рисунок 6.24 Код SAS для расчета с-статистики

```
proc rank data=&_VALID out=rscored;
    var PREDICTIONVARIABLE;
run;

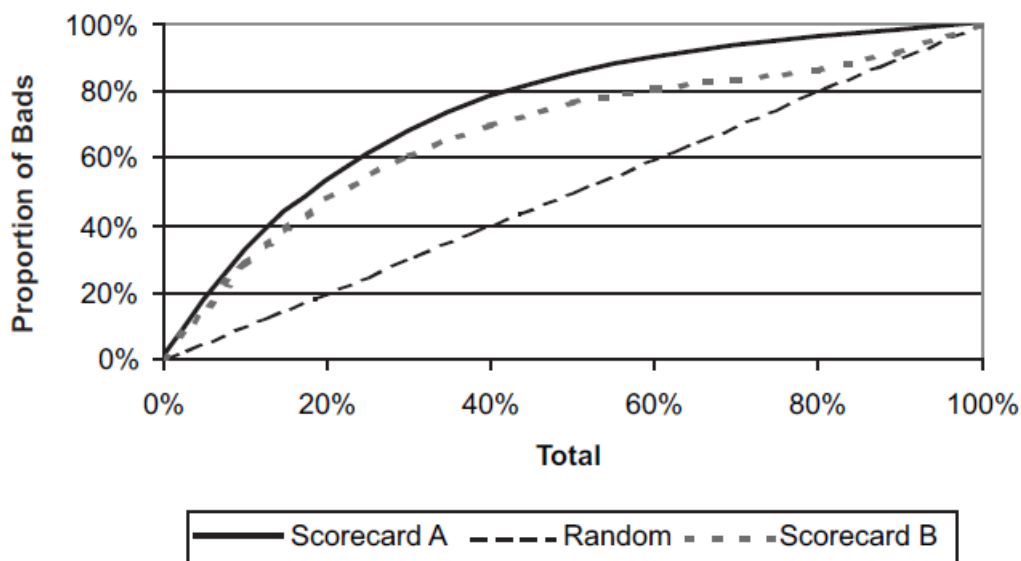
proc sql;
    select sum(TARGET=1) as n1,
           (sum(PREDICTIONVARIABLE*(TARGETVARIABLE=1))-0.5*(calculated
n1)*(calculated n1+1))
           /(((calculated n1)*(count(TARGETVARIABLE)-(calculated n1))))
    as c
    from rscored;
quit;
```

- **Кривая Лоренца.** Метрикой, аналогичной ROC-кривой и используемой в бизнесе для сравнения моделей, является график распределения «плохих» исходов и всех исходов, разбитых по децилям, по всему диапазону скоринговых баллов. Ее называют кривой Лоренца и она измеряет, насколько хорошо скоринговая карта отличает «плохих» от «хороших» в выбранном дециле. Пример кривой Лоренца приведен на рисунке 6.25.

На рисунке 6.25 показано, что скоринговая карта на 60% данных выявляет около 90% всех «плохих», тогда как скоринговая карта «В» выявляет только около 80%. Поэтому скоринговая карта «А» демонстрирует большую эффективность. Отметим, что отношение площади между кривой Лоренца для скоринговой карты и диагональной линией в 45% к общей площади треугольника под диагональной линией в 45% эквивалентно индексу Джини.

Здесь важно сравнить эффективность скоринговой карты по децилям, подразумевая, что если ожидаемый процент одобрения составляет около 60%, то эффективность должна сравниваться по 60% процентилю. Сравнение эффективности в первых 10% данных нерелевантно в тех случаях, когда необходимо добиться наибольшей эффективности, в данном случае в отметке 60%. Однако в ходе работы со скоринговыми картами банкротства или отклика на предложение, сравнение по наименьшим процентиям имеет смысл и должно быть выполнено – поскольку в данной ситуации цель заключается в том, чтобы отделить наихудших/наилучших клиентов от остальных для проведения маркетинговых акций.

Рисунок 6.25 Кривая Лоренца



Как и в случае с другими решениями, которые были приняты в ходе разработки скоринговой карты, решение о выборе итоговой скоринговой карты должно быть вынесено с учетом целей, для которых разрабатывалась карта. Это может потребовать дальнейшего анализа, чтобы определить, какая скоринговая карта, например, максимизирует прибыль или минимизирует риски (для скоринговых карт мошенничества) в заданном пороге отсечения. При выборе итоговой скоринговой карты нужно сперва отобрать ту, которая помогает банку реализовать его стратегические цели, а затем убедиться в том, что она имеет приемлемую предсказательную силу.

Есть много и других метрик для оценки предсказательной силы скоринговой карты, включая:

График выигрышей (Gains Chart). Отношение кумулятивного числа спрогнозированных положительных исходов к общему числу спрогнозированных положительных исходов в выборке.

Lift-кривая. Отношение чувствительности к общему числу спрогнозированных положительных исходов в выборке.

$$Lift = \frac{\text{Число спрогнозированных положительных исходов}}{\% \text{ положительных исходов в выборке}}$$

Стоимости ошибочной классификации (Misclassification Costs). Ложно положительные и ложно отрицательные исходы рассматриваются как убыточные случаи. Оптимальное решение минимизирует общую ожидаемую стоимость.

Правило Байеса. Минимизирует ожидаемую стоимость (т.е. стоимость ошибочной классификации). Правило Байеса и стоимости ошибочной классификации трудно реализовать на практике из-за проблем в получении точного количества убытков.

Отношение стоимостей. Отношение стоимостей – это отношение ошибочной классификации «хороших» как «плохих» (ложно отрицательные исходы) к ошибочной классификации «плохих» как «хороших» (ложно положительные исходы).

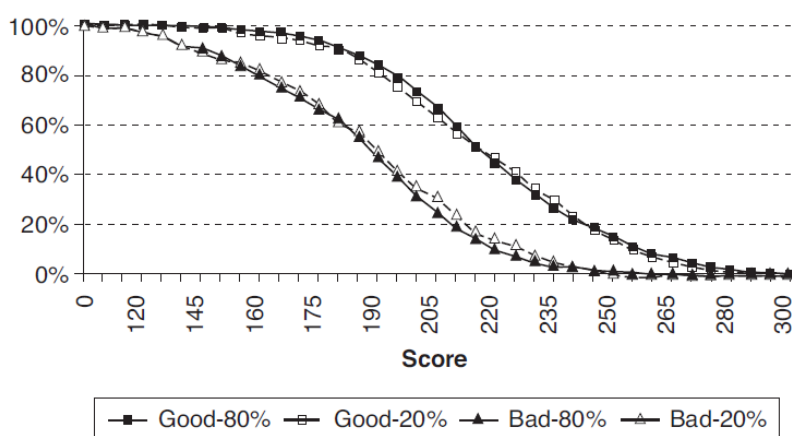
D Сомерса, гамма Гудмана и Крушкала, тау-а Кенделла. Основываются на числе конкордантных и дискордантных пар. Эти меры связаны с с-статистикой.

Валидация скоринговой карты

После того, как итоговая скоринговая карта выбрана, результаты моделирования необходимо проверить. Валидация выполняется для того, чтобы убедиться: модель применима к популяции и модель не переобучена. Как упоминалось ранее, рекомендуется разбить данные на обучающую и тестовую выборки (70-80% и 20-30%). На обучающей выборке строить скоринговую карту, а на тестовой проверять ее. Если скоринговая карта разрабатывается на небольшой выборке, необходимо строить ее на 100% наблюдений, а затем проверять на нескольких случайных выборках, содержащих от 50% до 80% наблюдений.

Первый метод валидации сравнивает распределения «хороших» и «плохих», оцененных по скоринговому баллу, в обучающей и тестовой выборках. На рисунке 6.26 приведены результаты валидации.

Рисунок 6.26 График валидации



Скоринговая карта считается прошедшей проверку, если два набора данных не имеют значимых различий. Обычно визуальной проверки двух кривых достаточно. Однако можно использовать и другие критерии качества, такие, как метод наименьших квадратов или IV.

Второй метод валидации сравнивает различные метрики скоринговой карты в обучающей и тестовой выборках. Пример приведен на рисунке 6.27.

Скоринговая карта считается прошедшей проверку, если нет значимых различий между метриками для обучающей и тестовой выборок.

Следующий метод валидации сравнивает показатель дивергенции для обучающей и тестовой выборок.

Дивергенция – это разница между средними значениями «хороших» и «плохих» в данном скоринговом балле, деленными на их дисперсию. Она вычисляется по следующей формуле:

$$\text{Дивергенция} = (\text{средн}_G - \text{средн}_B)^2 / [0.5 (\text{дисп}_G + \text{дисп}_B)]$$

где средн_G , средн_B , дисп_G и дисп_B являются средними значениями и дисперсиями «хороших» и «плохих», оцененных по скоринговому баллу.

Цель этого метода – построить модель так, чтобы в скоринговом балле распределение «хороших» находилось как можно дальше от распределения «плохих». Затем разница (расстояние) между распределением «хороших» и «плохих» по скоринговым баллам сравнивается в обучающей и тестовой выборках.

Валидацию можно также проводить, сравнивая отношение «хороших» к «плохим» по диапазону скоринговых баллов в обучающей и тестовой выборках. Если проценты «плохих» по скоринговым баллам в обеих популяциях не имеют значимых различий, то скоринговая карта считается прошедшей проверку. Значимые различия, обнаруженные с помощью любого из вышеперечисленных методов, требуют дальнейшего анализа. Обычно характеристики с большими диапазонами скоринговых баллов представляют проблемы, поскольку небольшое смещение распределения данных может привести к значительному смещению баллов. То же самое верно и для скоринговых карт, разработанных на основе небольших и неслучайных выборок. Такие характеристики, включенные в модель, являются причиной переобучения.

Рисунок 6.27 Валидация путем сравнения показателей эффективности модели

<i>Fit Statistic</i>	<i>Label</i>	<i>Development</i>	<i>Validation</i>
AIC	Akaike's Information Criterion	6214.0279153	.
ASE	Average Squared Error	0.0301553132	0.0309774947
AVERR	Average Error Function	0.1312675287	0.1355474611
DFE	Degrees of Freedom for Error	23609	.
DFM	Model Degrees of Freedom	7	.
DFT	Total Degrees of Freedom	23616	.
DIV	Divisor for ASE	47232	45768
ERR	Error Function	6200.0279153	6203.7361993
FPE	Final Prediction Error	0.0301731951	.
MAX	Maximum Absolute Error	0.9962871546	0.9959395534
MSE	Mean Square Error	0.0301642541	0.0309774947
NOBS	Sum of Frequencies	23616	22884
NW	Number of Estimate Weights	7	.
RASE	Root Average Sum of Squares	0.1736528525	0.1760042464
RFPE	Root Final Prediction Error	0.1737043324	.
RMSE	Root Mean Squared Error	0.1736785944	0.1760042464
SBC	Schwarz's Bayesian Criterion	6270.5156734	.
SSE	Sum of Squared Errors	1424.295752	1417.777979
SUMW	Sum of Case Weights Times Freq	47232	45768
MISC	Misclassification Rate	0.0320121951	0.0325117986
PROF	Total Profit for GB	3430000	2730000
APROF	Average Profit for GB	145.24051491	119.29732564

Процесс Разработки Скоринговой Карты, Этап 5: Управленческие Отчеты по Скоринговой Карте

После того, как итоговая скоринговая карта выбрана, необходимо подготовить ряд управленческих отчетов по скоринговой карте. Эти отчеты являются инструментами управления, используемыми для принятия таких решений, как выбор порога отсеечения для скоринговой карты, определение критериев одобрения клиента и управленческих стратегий, мониторинг эффективности будущей скоринговой карты. Их нужно разрабатывать и получать для того, чтобы помочь бизнес-пользователю ответить на вопросы «Где мне следует поставить порог отсеечения, чтобы решить поставленные передо мной задачи?», «Каким образом это повлияет на мой кредитный портфель?». Поэтому оптимально получить входную информацию от разработчиков скоринговых карт. На основе их отчетов бизнес-пользователь найдет информацию, полезную для принятия решений и использовать как руководство для разработки последующих отчетов.

Обычно эти отчеты включают распределения скоринговых баллов и характеристик скоринговой карты, графики и таблицы ожидаемого процента «плохих»/процента одобрения и результаты скоринговой карты по ключевым группам. Они строятся на обучающем наборе, использованном для разработки скоринговой карты, включая «серую зону» и отклоненных заявителей.

В дополнение к этим отчетам следует подготовить документацию по скоринговой карте, детально описывающую каждый этап разработки (обоснование проекта, определение «хороших»/«плохих»/«серой зоны», исключения, сегментация, сбор данных, выборка данных, анализ исходных характеристик, разработка модели, анализ отклоненных заявок, показатели эффективности скоринговой карты и валидация), и результаты. Она послужит справочным материалом в ходе разработки последующих скоринговых карт, подготовки требований к аудиту и соблюдению правовых норм, поиска и устранения ошибок.

Таблица выигрышей (Gains Table)

Таблица выигрышей включает в себя общее распределение, распределение «хороших» и «плохих» по баллам или диапазонам баллов. На рисунке 7.1. приведен фрагмент таблицы выигрышей с использованием индивидуальных скоринговых баллов. Таблица выигрышей разрабатывается по всей выборке, а также по отобраным подвыборкам. Ключевой информацией в этой таблице является:

- Ожидаемый процент «плохих» для каждого балла или диапазона баллов (процент «плохих» в данном диапазоне или предельно допустимый процент «плохих»)
- Ожидаемый процент «плохих» для всех заявителей, находящихся выше определенного балла (кумулятивный процент «плохих»)
- Ожидаемый процент одобрения по каждому баллу

Данная информация, опираясь на ожидаемый процент «плохих» или процент одобрения, в сочетании с финансовыми и текущими соображениями используется для определения порога отсечения: в каком пороге отсечения следует одобрить новых заявителей? Цель построения таблицы выигрышей для групп заключается в том, чтобы выявить любые аномальные результаты по сегментам, важным для бизнеса. Например, обычно скоринговые карты присваивают более молодым клиентам меньше баллов. Группами могут быть географические районы, отрасли, возрастные группы, текущие/новые клиенты, сегменты, выбранные для будущих маркетинговых кампаний и т.д.

Рисунок 7.1 Фрагмент таблицы выигрышей

<i>Score</i>	<i>Count</i>	<i>Cumulative Count</i>	<i>Goods</i>	<i>Bads</i>	<i>Cumulative Goods</i>	<i>Cumulative Bads</i>	<i>Interval Bad Rate</i>	<i>Cumulative Bad Rate</i>	<i>Approval Rate</i>
210	345	6,965	311	34	6,538	427	9.86%	6.13%	69.50%
211	500	6,620	462	38	6,227	393	7.60%	5.94%	66.20%
212	450	6,120	418	32	5,765	355	7.11%	5.80%	61.20%
213	345	5,670	323	22	5,347	323	6.38%	5.70%	56.70%

Распределение популяции по баллам также используется для оценки стабильности скоринговой карты, итоговый отчет по баллам применяется для мониторинга скоринговой карты (который будет освещен в Главе 9).

Отчеты «Характеристики»

Данные отчеты содержат распределения каждой характеристики, включенной в скоринговую карту, а также процент одобрения и процент «плохих» для каждого атрибута по скоринговому баллу. На рисунке 7.2 приведен пример отчета «Характеристики».

Верхняя часть рисунка – это основа отчетов «Характеристики» (смотрите Главу 9), создаваемых в ходе регулярного мониторинга скоринговой карты. Нижняя часть используется для получения результатов разбиения по заданному порогу для атрибутов, а также для сегментов, важных для банка. На рисунке 7.2. показано, что если порог отсечения установлен в скоринговом балле 212, только 58% заявителей в возрасте от 18 до 22 лет получают одобрение. Сравните это с процентом одобрения 92%, установленным для заявителей старше 44 лет. Данная политика может быть приемлема для некоторых организаций, которые не нацелены на более молодых заявителей. Аналогичные отчеты, содержащие информацию об ожидаемом проценте «плохих» для каждого сегмента по каждому баллу отсечения, предупредят бизнес-пользователя о наличии высокорисковых групп, которые получают одобрение. В случае подобного сценария нужно попытаться разработать сегментированную скоринговую карту для более молодых заявителей, чтобы максимально увеличить эффективность модели в данном сегменте, или использовать ту же самую скоринговую карту с различными порогами отсечения для различных сегментов.

Рисунок 7.2 Отчет «Характеристики»

Age	Distr	Points	Bad Rate				
Missing	8%	16	16%				
18-22	9%	12	24%				
23-26	15%	18	18%				
27-29	26%	26	10%				
30-35	10%	35	5%				
35-44	20%	43	3%				
44 +	12%	51	2%				

Score	Missing	18-22	23-26	27-29	30-35	35-44	44 +
210	74%	70%	78%	80%	83%	91%	97%
211	67%	64%	71%	77%	80%	88%	95%
212	61%	58%	66%	70%	76%	84%	92%
213	56%	50%	61%	67%	72%	80%	87%

Также настоятельно рекомендуется подготовить вариант отчета, в котором распределение значений и уровень одобрения по сегментам строится для характеристик, не включенных в скоринговую карту. Эти характеристики следует выбрать так, чтобы они представляли важнейшие группы или целевые рынки. По ним более детально оценивают результаты разбиения по заданному порогу для каждого сегмента в отдельности. Такие отчеты можно использовать для разработки стратегии по каждому сегменту, включая назначение различных порогов отсека для разных сегментов.

Подготовка данных отчетов обычно означает завершение процесса разработки скоринговой карты. Следующие главы будут посвящены тому, как внедрить эти скоринговые карты и использовать их для принятия решений.

Процесс Разработки Скоринговой Карты, Этап 6: Внедрение Скоринговой Карты

Данный раздел посвящен анализу скоринговой карты после ее разработки и затронет три главных темы:

1. Рассмотрение результатов анализа и деловых аспектов в ходе внедрения скоринговых карт.
2. Понимание того, как используются скоринговые карты и управленческие отчеты.
3. Понимание того, как разработана стратегия.

Валидация перед внедрением

Начатая после разработки скоринговой карты подготовка к внедрению включает тестирование точности скоринга и фронт-энд валидацию. Процедура фронт-энд валидации схожа с валидацией, осуществляемой в ходе разработки скоринговой карты, но при этом решает другие задачи. В то время как цель валидации в ходе разработки скоринговой карты заключалась в том, чтобы подтвердить стабильность скоринговой карты, сравнивая распределения в обучающей и тестовой выборках, цель валидации перед внедрением – убедиться в том, что разработанная скоринговая карта точно отражает текущую популяцию заявителей. В некоторых случаях, когда обучающей выборке 2-3 года, могут произойти значительные изменения профиля заявителя и нужно выявить их. Результаты данной валидации также используются в ходе анализа порогов отсека.

Перед выполнением валидации, все новые внешние и внутренние характеристики скоринговой карты нужно перевести в программный код, создать внешний интерфейс, систему оценки и принятия решений.

После того, как программирование завершено, можно выполнить тестирование предсказательной точности скоринговой карты. Отчеты по валидации могут быть подготовлены как составная часть тестирования, которая позволит повысить эффективность процедуры. В идеале, тестирование следует выполнять в тех же самых условиях, в которых скоринговая карта будет внедряться. Если это невозможно, то программное обеспечение, например, SAS, можно настроить так, что оно смоделирует скоринг и распределения значений в популяции. Однако это не даст нам точного представления о том, как текущая система будет интерпретировать различные скоринговые характеристики, особенно расчетные, и может привести к неточным прогнозам. Поэтому важно, чтобы условия проведения тестирования были максимально приближены к условиям работы текущей системы.

После того, как точность скоринга определена, можно подготовить отчеты по фронт-энд валидации с помощью скоринга текущих заявителей.

Одним из важнейших показателей эффективности скоринговой карты является не только ее предсказательная точность, но и отсутствие значительных изменений в профиле заявителя на протяжении ее построения. Если построение скоринговой

карты длится достаточно долго, в кредитном портфеле могут произойти значимые изменения. Поскольку такие изменения не отражаются в исходных данных, которые были использованы для разработки скоринговой карты (обучающая выборка), то необходимо убедиться в том, что качество скоринговой карты не изменилось. А если изменилось, то необходимо учесть такие отклонения. Для решения этих задач проводится фронт-энд валидация.

В ходе фронт-энд валидации используется новая скоринговая карта и сравнивается распределение текущих заявителей (фактический входящий поток) с распределением заявителей в обучающей выборке. Обычно выполняется скоринг заявителей, недавно подавших заявки на кредит. Однако хорошей практикой является последовательный анализ текущих заявителей по нескольким последним периодам времени (последний месяц, последние три месяца, последние шесть месяцев и т.д.), чтобы обнаружить любые появляющиеся тренды или подтвердить предположение о том, что изменения, выявленные за отдельно взятый месяц, не являются долгосрочным трендом.

Обычно для этого подготавливают отчеты «Стабильность системы» и «Анализ характеристик». Данные отчеты могут разрабатываться как для аппликационных, так и для поведенческих скоринговых карт.

Отчет «Стабильность системы»

На рисунке 8.1 приведен пример отчета «Стабильность системы». Отметим, что отчет «Стабильность системы» иногда относят к отчету «Стабильность популяции» или «Стабильность скоринговой карты».

Рисунок 8.1 Отчет «Стабильность системы»

<i>Score Range</i>	<i>Actual %</i>	<i>Expected %</i>	<i>(A-E)</i>	<i>A/E</i>	<i>ln(A/E)</i>	<i>Index</i>
0-169	7%	8%	-1%	0.8750	-0.1335	0.0013
170-179	8%	10%	-2%	0.8000	-0.2231	0.0045
180-189	7%	9%	-2%	0.7778	-0.2513	0.0050
190-199	9%	13%	-4%	0.6923	-0.3677	0.0147
200-209	11%	11%	0%	1.0000	0.0000	0.0000
210-219	11%	10%	1%	1.1000	0.0953	0.0010
220-229	10%	9%	1%	1.1111	0.1054	0.0011
230-239	12%	10%	2%	1.2000	0.1823	0.0036
240-249	11%	11%	0%	1.0000	0.0000	0.0000
250+	14%	9%	5%	1.5556	0.4418	0.0221
Index =						0.05327824

Столбцы «Текущие %» и «Ожидаемые %» показывают распределение наблюдений в текущей и обучающей выборках для каждого диапазона скоринговых баллов. Индекс оценивает изменение популяции заявителей, произошедшее за период разработки скоринговой карты, в выбранном диапазоне.

Индекс вычисляется следующим образом:

$$\sum (\% \text{ Текущие} - \% \text{ Ожидаемые}) \times \ln (\% \text{ Текущие} / \% \text{ Ожидаемые})$$

по всем диапазонам скоринговых баллов.

Индекс меньше 0.10 показывает отсутствие значимого изменения популяции, индекс 0.10–0.25 говорит о незначительном изменении, которое необходимо исследовать, индекс выше 0.25 свидетельствует о значительном смещении популяции заявителей.

Для оценки смещения могут также использоваться другие методы, такие, как хи-квадрат с заданным уровнем значимости. Метод, приведенный на рисунке 8.1, широко используется в бизнесе.

Дополнительный способ выявить причины смещений – построить график распределения текущих и ожидаемых заявителей по скоринговым баллам. Это может дать дополнительную информацию (является ли смещение баллов нисходящим, восходящим или эксцессом).

Индекс, упомянутый выше, не является универсальной метрикой. Перед тем, как принять решение о том, что смещение значимо, необходимо учесть такие факторы, как тренды (изменение носит временный или долговременный характер?), величина смещения, причины смещений.

Изменения скоринговых баллов могут быть вызваны несколькими причинами:

- Объективные изменения в профиле заявителя (например, возраст).
- Маркетинговые кампании, конкурентная борьба за рыночные ниши. Например, если заявители, отобранные за последний месяц, были значительно моложе или были из определенного региона, это могло быть вызвано маркетинговой активностью. Изменение продукта, такие как, дополнительные условия по программам лояльности, изменения в структуре оплаты кредита – «0% на срок 6 месяцев» и прочие «бонусы», или переключение на нетрадиционные каналы продаж могут также привлечь новых заявителей. Внешняя конкуренция, особенно из-за новых участников на рынке, нацеленная на узкие демографические группы, также может повлиять на состав ваших заявителей. В поведенческом скоринге более агрессивные акции или кредитные стратегии, внедрение программ лояльности, пересмотр ставок, кросс-продажи могут изменить профиль текущего клиента.
- Ошибки ввода данных.
- Ошибки в сборе данных, данные представляют собой неслучайную или неправильно сформированную выборку или включены наблюдения, исключенные из обучающей выборки.

Важно отметить, что для строгого отчета «Стабильность системы» выборку текущих заявителей нужно сформировать по тем же самым критериям отбора, что использовались для проектирования обучающей выборки. Однако компании для определения порога отсечения обычно проводят повторный анализ, который включает всех заявителей. Он дает более реалистичную оценку ожидаемого процента одобрения и результатов разбиения при выбранном пороге по ключевым сегментам.

Отчет «Стабильность системы» лишь показывает, произошло ли смещение и дает представление о величине этого смещения. Бизнесу для принятия решений важнее найти причину смещения, т.к. это позволит внести необходимые коррективы.

Дальнейший анализ причин может быть осуществлен с помощью:

- Анализ смещений характеристик скоринговой карты
- Анализ смещений характеристик, не вошедших в скоринговую карту. Предположительно они влияют на «качество» заявителя и могут объяснить скрытые причины смещений характеристик в скоринговой карте (более подробно об этом в следующем разделе)
- Сбор информации о последних маркетинговых кампаниях и предложениях на рынке

Отчет «Анализ характеристик»

Отчет «Анализ характеристик» содержит информацию о смещениях в распределении характеристик (вошедших и не вошедших в скоринговую карту), позволяет оценить влияние этих смещений на баллы. На рисунке 8.2 приведен анализ характеристик по «Возрасту».

Столбцы «Текущие %» и «Ожидаемые %» снова показывают распределение наблюдений в текущей и обучающей выборках для каждого диапазона скоринговых баллов. Индекс вычисляется просто:

$$\Sigma(\% \text{ Текущие} - \% \text{ Ожидаемые}) \times (\text{Баллы})$$

по всем диапазонам скоринговых баллов.

Рисунок 8.2 Отчет «Анализ характеристик»

<i>Age</i>	<i>Expected %</i>	<i>Actual %</i>	<i>Points</i>	<i>Index</i>
18-24	12%	21%	10	0.9
25-29	19%	25%	15	0.9
30-37	32%	28%	25	-1
38-45	12%	6%	28	-1.68
46+	25%	20%	35	-1.75
				-2.63

На рисунке 8.2 приведен анализ характеристики «Возраст». Исходя из таблицы, можно сделать вывод о смещении в пользу более молодых заявителей. Текущие заявители по характеристике «возраст» получают на 2.63 балла меньше, чем ожидаемые. Аналогичным способом выполняется анализ по остальным характеристикам скоринговой карты, чтобы подробно выяснить причины смещения баллов.

Сравнение текущих и ожидаемых клиентов выполняется и по другим характеристикам, включая:

- Характеристики, не вошедшие в скоринговую карту, но предположительно влияющие на качество заявителя. Они включают надежные характеристики, которые не являлись таковыми, будучи включенными в скоринговую карту. Например, если смещения характеристик скоринговой карты указывают на ухудшение кредитоспособности заявителей, то анализ характеристик, не включенных в скоринговую карту, поможет подтвердить данный факт. В основе анализа предположение, что оценка кредитоспособности заявителя опиралась на различные источники информации и поэтому более устойчива.
- Характеристики, схожие с характеристиками скоринговой карты. Например, если «возраст» включен в скоринговую карту, мониторинг характеристик, зависящих от времени, таких, как стаж работы, стаж проживания, возраст первого кредитного счета – могут в дальнейшем объяснить причины изменения профиля заявителя. Если «запросы за последние 6 месяцев» включены в скоринговую карту, мониторинг запросов за последние 3 месяца и за последние 12 месяцев помогут в дальнейшем объяснить изменение количества запросов. Смещение значений связанных характеристик должно быть однонаправленным (например, более старшие по возрасту заявители имеют больший стаж работы и больший возраст кредитной истории).
- Когда в скоринговой карте используются коэффициенты, необходимо исследовать распределение знаменателя и числителя, чтобы объяснить

изменение самого коэффициента. Например, если утилизация (баланс, деленный на кредитный лимит) уменьшилась, это может быть обусловлено либо снижением балансов, либо увеличением кредитных лимитов.

Кроме того, необходимо выполнить данный анализ, сравнив обучающие данные с данными за последний месяц или за последние 6 месяцев, чтобы обнаружить какие-либо тренды и подтвердить, что смещения распределений не являются временным явлением.

Смещение характеристик (как включенных, так и не включенных в скоринговую карту) должно согласовываться и объясняться логикой и деловыми соображениями. Например, если анализ показывает, что заявители «помолодели», но при этом у них большой «стаж работы», то это указывает, что при вводе данных могли быть допущены ошибки. Такие нелогичные смещения должны быть тщательно исследованы и объяснены.

Что делать, если скоринговая карта не прошла валидацию?

Если анализ показывает, что популяция изменилась значительно, то в распоряжении пользователя есть несколько средств. Повторная разработка скоринговой карты не может быть оптимальным решением, поскольку выборка для повторного моделирования, скорее всего, будет за тот же период, что и первоначальная скоринговая карта, не прошедшая валидацию. В таких случаях пользователи могут скорректировать популяцию ожидаемых заявителей с учетом новых распределений и в дальнейшем настроить порог отсека и кредитные стратегии. Это обозначает повторную подготовку отчетов эффективности скоринговой карты, как было описано в предыдущем разделе, с учетом нового распределения значений, но с сохранением той же самой зависимости отношения шансов от скорингового балла. По итогам качественного и количественного анализа ожидаемый процент «плохих» также может быть пересмотрен. Например, если выдвинуто предположение, что популяция ниже по своему качеству, чем ожидалось, можно предпринять консервативные меры, увеличив ожидаемый процент «плохих» и снизив суммы по кредитным линиям до определенного порога.

Качественный анализ, проводимый на данном этапе, исследует причины смещений скоринговых баллов. Например, если анализ выявляет более молодых заявителей с более высоким уровнем просрочки и более низкими показателями стабильности, то ожидаемая кредитоспособность очевидна. В некоторых случаях причины смещений менее очевидны. Например, наблюдения, у которых «утилизация кредитной линии» стала ниже, несут меньший риск. Это может быть обусловлено снижением балансов (которое на самом деле показывает более низкий риск) или увеличением доступных кредитных лимитов из-за конкурентного давления банков (которое не может быть индикатором более низкого риска). Объяснение причин поможет дать обоснованную оценку ожидаемой кредитоспособности.

Если причины смещений баллов надежно объяснены, можно попытаться внести изменения в скоринговые баллы. Например, в тех случаях, когда общее число кредитных линий возрастает, искусственно снижая утилизацию, заявителям с более низкой утилизацией можно назначить меньше баллов, чтобы учесть более высокий риск по сравнению с тем, который предполагался в скоринговой карте. Этот метод не из тех, которые можно применять, не вникая в причины смещения баллов. Он является лишь наиболее подходящим решением проблемы. Другой подход заключается в том, чтобы выделить причину смещения и разработать для нее отдельную скоринговую карту. Например, банк был нацелен на заявителей с устоявшейся кредитной историей, однако в последнее время стал привлекать

заявителей – новых должников. Кредитный портфель может быть разбит на два сегмента: первый – для клиентов с устоявшейся кредитной историей и второй – для новых должников, разрабатываются две отдельные скоринговые карты. Отметим, что если сегмент клиентов, имевших долги в недавнем прошлом, является совершенно новым для банка, банку необходимо использовать дженериковую скоринговую карту.

Также существуют статистические методы, которые могут максимально «приблизить» распределения в обучающей выборке к самым последним распределениям в текущей выборке. Однако эти методы лишь корректируют распределения характеристик и не могут дать информацию о том, как от этого изменится эффективность исходной выборки.

Как только установлено, что скоринговая карта точно отражает текущую популяцию заявителей, валидация считается завершенной. Теперь можно начинать разработку стратегии.

Разработка стратегии

Общие рассуждения

Скоринговые карты разрабатываются для решения определенных бизнес-задач. После того, как скоринговая карта составлена, пользователю нужно решить, как она будет применяться для достижения этих задач. Это включает проведение анализа и принятие решений, скорректированных с учетом риска, по таким вопросам, как минимальный балл одобрения (порог отсечения), первоначальная сумма кредитной линии, автоматическое увеличение кредитной линии, определение правил кредитной политики, внедрение стратегий «Чемпион – Претендент». Обычно для этих целей используют отчеты по фронт-энд валидации в сочетании с отчетами эффективности скоринговой карты.

Есть несколько важных моментов, которые нужно помнить при разработке стратегии:

- **Результаты по ключевым сегментам.** Любая внедряемая стратегия, любое принятое решение должно быть проанализировано с т.з. получения результатов по ключевым сегментам (региональные, демографические, каналы распространения). Ключевыми называют сегменты, наиболее привлекательные для банка, или важные целевые рынки. Это снижает вероятность неожиданных ситуаций и позволит банку использовать ту же самую скоринговую карту, но при этом с различными порогами отсечения для определенных сегментов. Данный анализ еще более важен, когда для этих сегментов не разработаны отдельные скоринговые карты и профиль заявителя значимо отличается по сегментам (клиент, пришедший в филиал банка, или клиент, пришедший на интернет-сайт).
- **Анализ «Что-если».** Когда компании используют программное обеспечение для автоматизированного принятия решений, необходимо выполнить анализ «что-если» для стратегий «Чемпион-Претендент». Цель – получить первоначальное представление о влиянии новой предполагаемой стратегии на существующий бизнес, будет ли она работать лучше, чем существующий «Чемпион», оценить количество клиентов, учитываемых в каждом из заранее определенных сценариев. Кроме того,

целью является принятие наиболее обоснованного решения, насколько это возможно.

- **Правила кредитной политики.** Внедрение новых стратегий дает организации возможность пересмотреть и улучшить правила кредитной политики, особенно при вводе в эксплуатацию новых скоринговых карт. Особое внимание нужно обратить на согласованность решений, основанных на скоринге и правилах кредитной политики. Важно, чтобы эти решения не противоречили друг другу, а их эффективность могла быть тщательно исследована и независимо оценена.

Скоринговая стратегия

В условиях, где для какого-то сегмента используется одна скоринговая карта, скоринговая стратегия довольно простая. Каждый заявитель или клиент получает баллы по данной скоринговой карте и оценивается в зависимости от установленного порога отсечения. Однако когда используются сложные скоринговые карты, применяются различные методы. Сложные скоринговые карты применяются, когда, например, заявитель получает баллы в зависимости от уровня просрочки, ухода, банкротства, прибыльности, а также когда скоринговая модель кредитного бюро используется как дополнение к внутренней скоринговой модели.

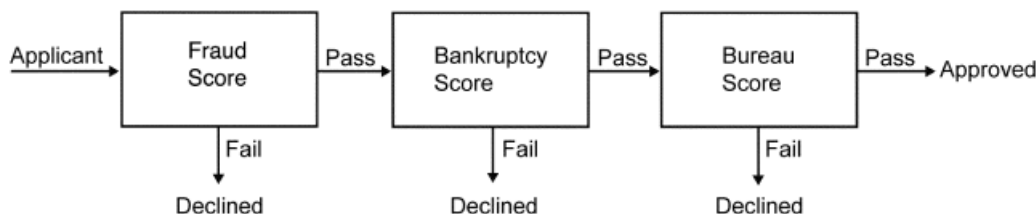
Существует три главных подхода к внедрению решений на основе сложных скоринговых карт:

1. Последовательный
2. Матричный
3. Матрично-последовательный

Последовательный При использовании этого метода заявитель последовательно получает баллы по каждой скоринговой карте с отдельным порогом отсечения. См. рисунок 8.3.

На рисунке 8.3 приведен пример последовательного подхода. Для оценки каждого заявителя используются три скоринговые карты. Отметим, что кроме решений «пропустить» и «отказать» может использоваться решение «отправить на дополнительное рассмотрение». Однако эту стратегию лучше всего внедрять когда используются абсолютные «барьеры» (например, заявитель, чтобы получить одобрение, должен пройти по скоринговому баллу банкротства или минимальному скоринговому баллу кредитного бюро) и нет «серых» зон. Когда есть «серые зоны» (например, оценки доходности), матричный подход предпочтительнее.

Рисунок 8.3 Последовательный подход в скоринге



Матричный В матричном подходе сложные скоринговые карты используются одновременно с использованием матрицы решений, основанной на комбинации порогов отсечения для различных скоринговых карт. Пример на рисунке 8.4 показывает матрицу ожидаемого риска и оттока (или ухода), где выявлены «серые» зоны.

Этот подход наиболее часто используется когда на основе различных, желательно независимых типов информации, нужно сделать сбалансированный выбор. Хороший балл по одной скоринговой карте может уравниваться плохим баллом по другой; например, вы хотите одобрить заявителя с низким риском просрочки и высокой вероятностью ухода или вы хотите одобрить заявителя, который, вероятно, будет в просрочке, но при этом у него низкая вероятность дефолта?

Пример, приведенный на рисунке 8.4, показывает:

- Заявители с высоким баллом просрочки (обозначающим низкий риск) и высоким баллом оттока (обозначающим низкую вероятность оттока) одобряются
- Заявители с низкими баллами просрочки и оттока отклоняются
- Заявители, находящиеся в «серой» зоне, подлежат дальнейшему анализу
- Заявители с низким баллом просрочки и высоким баллом оттока (т.е. у них высокий риск и низкая вероятность оттока) отклоняются сразу

Рисунок 8.4 Матричный подход в скоринге

		Churn Score				
		0–549	550–619	620–649	650–699	700 +
In- House Delq Score	0–189					Decline
	190–209	Decline	Refer			
	210–229					
	230–249			Approve		
	250 +					

Баланс зависит от внутренних приоритетов и целей. Наиболее распространенными примерами являются сравнение внутренней скоринговой карты со скоринговой картой кредитного бюро (обслуживание кредита в нашем банке сравнивается с обслуживанием кредитов в других банках), сравнение просрочки с прибыльностью, сравнение просрочки с оттоком/уходом и сравнение просрочки с вероятностью банкротства/списания (чтобы выделить заявителей, которые находятся в просрочке, но и в итоге расплатятся по кредиту). Важный аспект заключается в том, чтобы оба параметра должны быть независимы друг от друга и выдавать конкурирующую информацию.

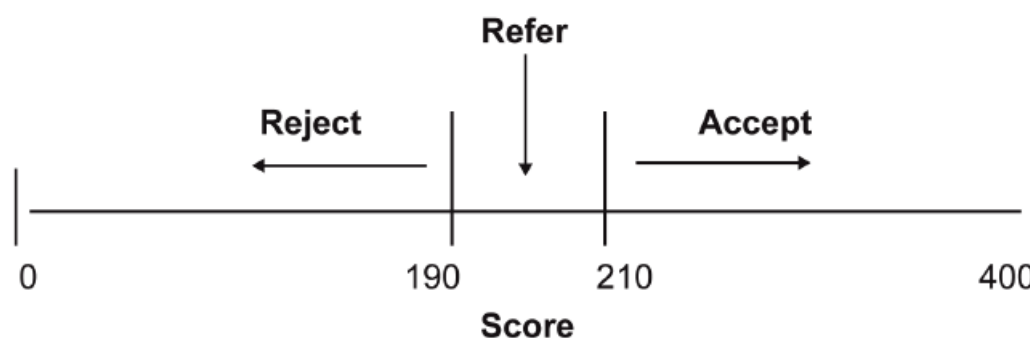
На рисунке 8.4 на основе баллов скоринговой карты риска и скоринговой карты доходности выделены три отдельные зоны. Аналогичным способом, когда это возможно, для более сложного процесса принятия решений можно построить трехмерные матрицы.

Матрично-последовательный В некоторых сложных скоринговых картах используется сочетание ранее упомянутых методов. Заявители предварительно оцениваются с помощью последовательного подхода, а затем применяется матричный подход. Например, заявители сперва оцениваются по модели банкротства и, пройдя порог отсеечения, сводятся в матрицу, составленную из скоринговых баллов просрочки/прибыли/оттока. Этот подход проще, чем многомерный матричный и более гибкий, чем последовательный. Его лучше всего использовать когда применяется более трех скоринговых карт и можно уравновесить несколько конкурирующих задач. Матрично-последовательный подход можно также использовать в сочетании с правилами кредитной политики для предварительной оценки заявителей.

Назначение порогов отсеечения

Большинство организаций, использующих скоринговые карты, устанавливают минимальную скоринговую оценку, при которой они могут одобрить заявителей (или оценить их в плане последующего обслуживания кредитного счета с помощью поведенческих карт). Эта минимальная скоринговая оценка называется «порогом отсеечения» и может представлять собой пороговое значение риска, прибыли или другого показателя в зависимости от задач банка, решаемых с помощью скоринговой карты. На рисунке 8.5 приведен простой пример назначения порога отсеечения для отбора новых заявителей.

Рисунок 8.5 Назначение порога отсеечения



В данном случае заявители, получившие больше 210 баллов по скоринговой карте, автоматически одобряются; заявители, получившие меньше 190 баллов, отклоняются; те, кто получил от 190 до 210 баллов, тщательно оцениваются вручную. Когда речь идет о работе с клиентами, для варьирования коллекторских стратегий (от мягких форм воздействия к более жестким) или определения более высоких кредитных лимитов может быть несколько порогов отсеечения. Для заявок, по которым труднее вынести решение, могут быть разработаны и более сложные стратегии, например:

- Назначение различных уровней «Одобрения», основанное на соответствующей проверке или дополнительной информации, необходимой для получения итогового одобрения (например, находящееся на рассмотрение подтверждение доходов или ожидающая решения оценка недвижимости по ипотечным кредитам). В некоторых случаях порог отсеечения устанавливают так, что выше него подтверждение дохода не

требуется. Это снижает требования для низкорисковых заявителей и низкоприбыльных кредитов. Кроме того, для высокорисковых заявителей могут потребоваться более серьезные методы подтверждения доходов, (например, копия справки о зарплате с работы), тогда как низкорисковые заявители могут быть просто опрошены по факсу.

- Назначение «жесткого нижнего порога отсечения» в точке, ниже которой пересмотры решений скоринговой карты (overrides) не допускаются. Например, банк может назначить итоговый порог отсечения 200 баллов, необходимых для одобрения, и может установить «жесткий нижний порог» 180 баллов. Это обозначает, что кредитные аналитики могут пересмотреть решение по отклоненным заявителям, набравшим от 180 до 200 баллов, если на это есть веские причины. «Жесткий верхний порог» обычно не используется. Банки могут всегда пересмотреть заявителей, опираясь на правила кредитной политики и т.п., вне зависимости от того, насколько высокие баллы они получили.

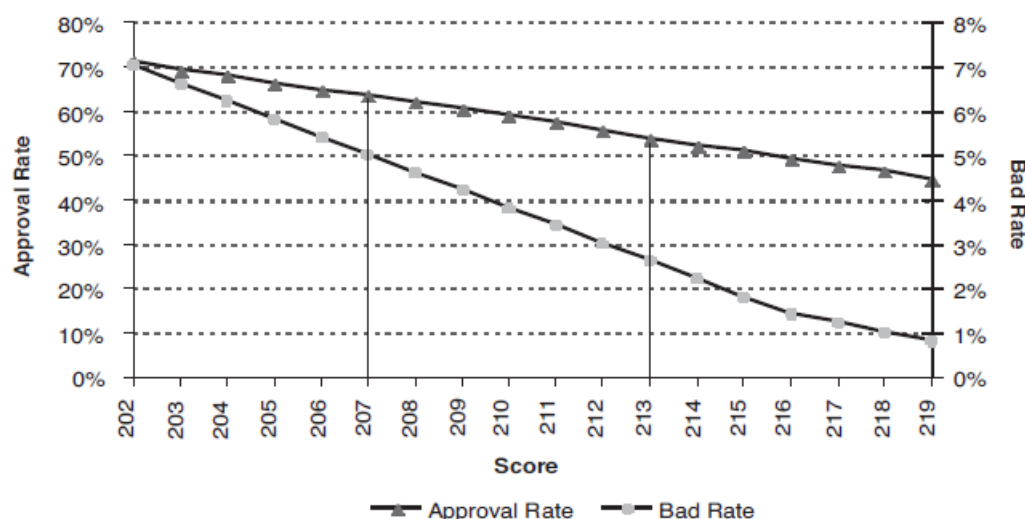
Анализ порогов отсечения В аппликационных скоринговых картах отправной точкой для выбора подходящего порога отсечения является анализ соотношения между ожидаемым процентом одобрения и процентом «плохих» в скоринговом балле. Есть лишь два главных критерия (в данном случае это процент одобрения и процент «плохих»). Схожий анализ можно выполнить, чтобы оценить влияние порога отсечения на такие параметры «компромисса», как прибыльность и доходность, процент «плохих» и доходность и другие релевантные показатели. Хороший способ получить сбалансированный компромисс между процентом «плохих» и процентом одобрения – выявить два ключевых пункта в скоринговом диапазоне:

1. Балл отсечения, который соответствует текущему проценту одобрения (и дает новый ожидаемый процент «плохих»); т.е. отвечаем на вопрос «Каким будет процент «плохих», если я оставлю тот же процент одобрения?»
2. И наоборот, балл отсечения, который соответствует текущему проценту «плохих» (и дает новый ожидаемый процент одобрения); т.е. отвечаем на вопрос «Каким будет процент одобрения, если я оставлю тот же процент «плохих»?»

Вообще каждое новое поколение разработанных скоринговых карт должно быть лучше, чем предыдущее, предшествующий анализ должен дать ответ, где организация получает более низкий процент «плохих» при том же самом проценте одобрения или более высокий процент одобрения при том, что процент «плохих» остается постоянным.

На рисунке 8.6 приведен «график компромиссного решения» («tradeoff chart»). Он показывает процент «плохих» и процент одобрения по выбранному диапазону скорингового балла, используя данные из таблицы выигрышей, построенной ранее. Отметим, что предшествующий анализ нужно выполнить, используя текущее распределение заявителей (для процента одобрения) и ожидаемый процент «плохих» в обучающей выборке. Рисунок показывает пример компании, у которого текущий процент «плохих» около 5% и процент одобрения 55%.

Рисунок 8.6 График компромиссного решения



Если цель разработки скоринговой карты – увеличить долю рынка (или если потери не важны), компания может сохранить текущий процент «плохих», выбрав порог отсечения в 207, и добиться нового, более высокого процента одобрения около 62%. Это обозначает, что организация ведет более агрессивную стратегию, одобряет большее количество заявителей, но при этом подверженность риску остается неизменной за счет более лучшего отбора заявителей для одобрения. Отметим, что это обусловлено наличием «набора для обмена данными», о чем было рассказано в разделе «Анализ отклоненных заявок» главы 6.

И наоборот, если цель разработки скоринговой карты – уменьшить потери, компания может выбрать порог отсечения в 213, сохранив процент одобрения, но получив более низкий ожидаемый процент «плохих» около 2.5%. Это будет обозначать, что компания снижает подверженность риску, но при этом отбирает одинаковое количество новых клиентов.

Когда есть заранее поставленные цели, может применяться одна из двух вышеописанных настроек. Обычно если нет такой цели, компании выбирают порог в интервале между двумя основными точками отсечения (207 и 213 на рисунке 8.6), где одновременно наблюдается увеличение процента одобрения и уменьшение ожидаемого процента «плохих».

Выбор порога не ограничен интервалом между этими двумя точками (от 207 до 213 на рисунке 8.6). Некоторые предпочитают устанавливать более низкий порог отсечения, чтобы получить более высокий процент «плохих», чем текущий, но при этом ожидая более высокую прибыльность. Например, управляющий портфелем ипотечных кредитов, имея потери ниже 10 базовых пунктов, может принять решение о снижении порога отсечения и увеличении потерь еще на несколько базовых пункта, ожидая охватить большую долю рынка и получить более высокую прибыльность.

Графики компромиссного решения, аналогичные графику на рисунке 8.6, могут быть построены с использованием других критериев, в зависимости от целей, поставленных перед скоринговой картой, например, графики компромиссного решения типа «риск/прибыль», «риск/отток».

Оценка влияния Как только предварительный порог отсечения выбран, нужно выяснить, как это повлияет на ключевые группы и сегменты. Кроме того, нужно убедиться, что все возможные последствия учтены до того, как будет

разрабатываться итоговая стратегия. Отчеты, приведенные на рисунках 7.1 и 7.2, можно использовать для анализа ожидаемого процента одобрения и процента «плохих» по критериям скоринговой карты, а также важнейшим сегментам. Разумный способ выявить такие сегменты – это обратиться в отдел маркетинга, чтобы получить информацию о планируемых и текущих целевых рынках. Нужно ознакомиться с результатами проведенных ранее исследований по выявлению сегментов, чтобы найти сегменты, которые считались совершенно другими по составу, но не дали достаточное количество информации для разработки скоринговой карты.

Для поведенческого скоринга сегменты – это группы клиентов, которые, вероятно, должны быть целевыми аудиториями для таких акций, как продление кредита, обслуживание кредитной линии.

Данные отчеты следует подготовить на основе текущей выборки заявителей, чтобы получить более точное представление об ожидаемом поведении одобренных заявителей.

Особенно рекомендуется применение этих отчетов когда используются характеристики, содержащие признаки неблагоприятного поведения (например, банкротство). Этот метод должен выровнять ожидания банка и предотвратить вопрос «Каким образом клиент-банкрот получил одобрение?». Пользователям необходимо понимать: есть основанная на пороге отсеечения статистическая вероятность того, что определенная доля клиентов-банкротов получит одобрение. Эта доля может быть очень маленькой в зависимости от порога отсеечения и способности скоринговой карты отделять клиентов-банкротов от остальных клиентов в нижней (начальной) части скорингового диапазона.

Также нужно понимать, что если порог отсеечения выбран с учетом общего профиля риска, клиенты-банкроты должны получить достаточно высокие баллы по другим характеристикам, чтобы достичь порога отсеечения. Пример такого анализа приведен на рисунке 8.7.

Рисунок показывает, что в скоринговом диапазоне, рассматриваемом для установления порога отсеечения, могут быть одобрены от 8% до 22% клиентов, имевших банкротство в прошлом. Организация может скорректировать порог отсеечения, основываясь на приемлемом для него уровне. Однако единственный путь не допустить прохождения банкротов в числе одобренных клиентов – применить жесткие правила кредитной политики. Если у организации есть такие правила, то характеристики, содержащие информацию о признаках банкротства, не должны использоваться в скоринговых картах, поскольку применение правил повлечет за собой смещенные результаты и ухудшит эффективность скоринговой карты. Это нужно понимать на первоначальном этапе, когда отбираются характеристики для включения в скоринговую карту (анализ исходных характеристик).

Рисунок 8.7 Уровень одобрения по скоринговым баллам – наибольшей просрочке по данным кредитного бюро

<i>Score</i>	<i>Nvr Delq</i>	<i>30 days</i>	<i>60 days</i>	<i>90 days</i>	<i>120 days</i>	<i>Bankrupt</i>
210	97%	70%	56%	50%	41%	22%
211	95%	62%	48%	41%	36%	18%
212	92%	58%	41%	38%	31%	12%
213	87%	50%	37%	31%	25%	8%

В том случае, если для отдельных сегментов обнаружена недостаточная эффективность скоринговой карты, у разработчика есть несколько альтернатив, включая:

- Использование одной и той же скоринговой карты, но с различными порогами отсечения для разных сегментов. Например, предположим, в стране есть пять регионов, с различным процентом «плохих» в соответствии с предполагаемым общим порогом отсечения. Если общий планируемый процент «плохих» равен, например, 2.5%, можно выбрать для каждого сегмента приемлемый порог отсечения, который даст в результате прогнозируемый процент «плохих» 2.5% для каждого сегмента. Таким образом, лучшие сегменты получают более высокий процент одобрения, в то время как более рискованные сегменты более низкий. Анализ может быть аналогичным способом выполнен, используя процент одобрения. Тот же самый принцип можно применить к скоринговым оценкам поведения заявителя и стратегиям определения лимита кредитной линии или пересмотра ставок.
- Повторная разработка скоринговой карты с различными характеристиками.
- Введение правил кредитной политики, чтобы не допустить одобрения определенных сегментов или задать для некоторых сегментов более строгие требования.
- Анализ дальнейшего сегментирования по переоцененным и недооцененным сегментам. Значительное отклонение от «средней» кредитоспособности указывает на необходимость сегментации, особенно если данные сегменты являются очень важными.

Стратегия коммуникаций между отделами в ходе разработки скоринговой карты

В предыдущем разделе обсуждалась важность понимания того, как стратегия влияет на получаемые результаты. Хорошей практикой является учет мнений различных заинтересованных сторон, принимавших участие в разработке скоринговой карты (упоминалось в разделе «Участники разработки скоринговой карты» главы 2). Они могут помочь, предупредив о произошедших изменениях в скоринговой карте и подготовив ее соответствующим образом.

Отдел маркетинга Это ценный источник информации, позволяющий выявить важные сегменты и дать более точный прогноз влияния на маркетинговые кампании. Отдел маркетинга может сообщить отделу риск-менеджмента о текущих маркетинговых кампаниях, других кампаниях по удержанию и привлечению клиентов, на которые может негативно повлиять введение новых скоринговых карт и порогов отсечения. Маркетинговые кампании могут объяснить изменения в профиле заявителя (выявленные в ходе валидации, предшествующей внедрению скоринговой карты).

Отдел IT. Предполагаемые изменения могут потребовать участие отдела IT и задержать внедрение скоринговой карты, особенно если скоринговые карты изменились, разработаны новые скоринговые карты или применены стратегии сложных порогов отсечения (при этом требуется соответствующее увеличение количества отчетов). Кроме того, если масштабирование скоринговой карты изменилось, отделу IT может потребоваться написать новый программный код для правил кредитной политики, стратегий, основанных на скоринговых баллах, и интерфейса, используемого для принятия решений по кредиту.

Кредитный отдел/Отдел автоматизации. Пороги отсечения и правила кредитной политики, увеличивающие число клиентов, отправленных на

дополнительное рассмотрение, могут привести к возрастанию объемов работ, лежащих на кредитных аналитиков. Аналогичным образом, большее количество транзакций по кредитным картам, нуждающихся в ручной оценке, потребует большего количества специалистов, чтобы обеспечить обслуживание клиентов на прежнем уровне. В тех случаях, когда наращивание производительности не представляется возможным, порог отсечения должен устанавливаться с учетом данного ограничения.

Отдел по взысканию просроченной задолженности. В некоторых случаях банк может взять на себя дополнительные риски, что увеличит объем работы, поступающий в отдел взыскания просроченной задолженности. Риски можно достаточно точно спрогнозировать, чтобы специалисты отдела по взысканию лучше подготовились к обработке большего объема работ в будущем. Например, расчетный риск может быть принят когда процент одобрения значимо возрос (наряду с ожидаемым процентом «плохих»), но в сочетании с ожидаемой большей прибыльностью. В этом случае отдел взыскания может столкнуться со значительным увеличением объема работы, которое, возможно, потребует найма дополнительного персонала. Кроме того, учет большего объема потерь (но в сочетании с большей прибылью, обусловленной более высокими доходами) стал бы хорошей идеей для отдела стратегического планирования, но не для специалистов, занимающихся возмещением потерь.

Отдел финансов. Информация, необходимая для расчета прибыльности новых стратегий, поступает обычно из отдела финансов.

Отдел корпоративных рисков. Более крупные суммы кредитных линий требуют распределения большего капитала; определение ставки по кредиту в зависимости от уровня кредитоспособности (Risk Based Pricing) может потребовать пересмотр системы хеджирования. В некоторых случаях слишком агрессивные стратегии могут противоречить стандартам управления рисками.

Отдел по работе с клиентами. Необходимо урегулировать негативные последствия для различных сегментов, вызванные введением новой скоринговой карты. Например, текущая маркетинговая кампания нацелена на определенную группу клиентов, которым по телефону предлагают воспользоваться продуктами банка. Если данный сегмент в ходе разработки скоринговой карты будет признан высокорискованным, то это негативно отразится на данной кампании. В этом случае, чтобы минимизировать этот негативный эффект, компания может составить для менеджеров по работе с клиентами письменные инструкции и предложить бонусные баллы клиентам, получившим отказ в кредите. Аналогично, если правила кредитной политики или изменения порога отсечения ведут к увеличению количества заявителей или клиентов, проходящих через специалистов кредитных отделов и отделов автоматизации, то это может вызвать недопустимо долгий период рассмотрения заявок.

Отдел обучения Возможно, персоналу нужно пройти переподготовку, чтобы понять масштабирование новой скоринговой карты, пороги отсечения и стратегии. Затраты на такое обучение можно снизить за счет введения единого стандарта масштабирования для банковских продуктов и сегментов. Скоринговые баллы в новой карте будут обозначать то же самое, что и в старой. Например, скоринговый балл 200 одинаково трактуется для всех скоринговых карт. Другой способ уменьшить затраты на обучение (и больше подходит для обеспечения

секретности скоринговых баллов) – преобразовать скоринговые баллы в отметки типа А, В, С, D и т.д. Разработка новой карты потребует лишь изменение назначения отметок вместо переобучения. Уполномоченный персонал увидит отметки, а не баллы, правила кредитной политики, связанные с каждой отметкой, не нужно будет менять.

Юридический отдел. Анализ влияния порогов отсечения на отбор клиентов в определенных географических зонах может уменьшить риск «географической дискриминации» и снизить риск кредитора быть обвиненным в дискриминации.

Мероприятия, скорректированные с учетом риска

На данном этапе у нас есть порог отсечения и поэтому мы знаем уровень риска у заявителей, находящихся под отсечкой и над отсечкой (одобренные). Данную информацию используют для разработки стратегий или мероприятий, скорректированных с учетом риска, чтобы максимально решить бизнес-задачи. Эти мероприятия будут различаться друг от друга в зависимости от предложенных продуктов. Стратегии можно разработать для заявителей или текущих клиентов – задача поиска решения, скорректированного на риск, остается той же самой. Например, в зависимости от скорингового балла и других критериев можно разработать различные стратегии, такие, как:

- Скорректированное на риск определение ставок по кредитам и другим продуктам, страховые премии для новых клиентов и кредитов с пересмотренными ставками по причине продления кредита
- Предложение обновлений продукта (например, золотая карта, платиновая карта) для лучших клиентов или привлечение клиентов более низкой процентной ставкой по карте в зависимости от их уровня риска.
- Определение размера первоначального взноса/залога для таких продуктов, как автокредит и ипотека, или определение условий и сроков продления кредита
- Кросс-продажи дополнительных продуктов для лучших клиентов на основе предварительных одобрений. Отметим, что в кросс-продажах оценки риска и готовности к покупке следует всегда использовать в сочетании с обслуживанием долга. Это гарантирует, что кросс-продажи ориентированы на клиентов, которые не только примут предложение и имеют низкий риск дефолта, но и смогут расплатиться по дополнительно взятому кредиту.
- Предоставление лучшим клиентам большей суммы кредитной линии или лимита по кредитной карте, как для заявителей, так и для текущих клиентов (смотрите пример на рисунке 8.8).
- Определение общего размера риска на одного клиента, приемлемого для банка (на основе профиля риска), в целях управления рисками концентрации.
- Использование более мягких методов взыскания долгов для низкорисковых клиентов (звонки и письма) и более жестких для высокорисковых клиентов (передача в коллекторское агентство).

- Варьирование условий и сроков платежей для клиентов (для лучших клиентов – более мягкие условия и сроки, в то время как для высокорисковых клиентов срок платежа наступает раньше, чем для «хороших»).
- Предоставление низкорисковым клиентам или клиентам на ранних сроках просрочки возможности совершать покупки на сумму, превышающую лимит по кредитной карте.
- Анализ мошеннических действий или требование тщательной оценки стоимости недвижимости для высокорисковых ипотечных кредитов.

На рисунке 8.8. приведена стратегия по определению размера лимита.

Рисунок 8.8 Определение суммы кредитной линии по карте в зависимости от скорингового балла и обслуживания долга

Score	Debt Service Ratio				
	0-10%	11-15%	16-24%	25-35%	36% +
230-234	\$3,500	\$3,500	\$2,500	\$2,500	\$2,000
235-239	\$4,000	\$3,500	\$3,000	\$2,500	\$2,000
240-244	\$4,500	\$4,000	\$3,500	\$3,000	\$2,500
245-249	\$5,000	\$4,500	\$4,000	\$3,500	\$3,000
250 +	\$7,500	\$5,000	\$4,500	\$4,000	\$3,500

Первое, что нужно отметить, данная стратегия основывается на двух независимых и релевантных показателях. Скоринговые баллы (слева) измеряют вероятность дефолта, а коэффициент обслуживания долга измеряет долю дохода клиента, затрачиваемого на обслуживание долгов. Более низкий коэффициент обслуживания долга лучше, т.к. указывает на то, что клиент имеет достаточный доход, чтобы расплатиться по кредиту. Согласно предложенной стратегии клиенты с более высокими скоринговыми баллами и низкими коэффициентами обслуживания долга получают большие кредитные лимиты. После того, как показатели для определения размера кредитной линии установлены, у пользователя появляются различные варианты управления кредитной линией:

- Определить порядок предоставления кредитной линии на основе обоих показателей и затем выбрать компромиссное решение, чтобы предоставить лучшим клиентам больший размер кредитной линии, а высокорисковым клиентам – меньший размер кредитной линии. Например, наилучшие клиенты могут получить увеличение на 30%, наихудшие – уменьшение на 30%. После того, как ячейки в нижнем левом углу таблицы (наилучшие клиенты) и ячейки в верхнем правом углу таблицы (наихудшие клиенты) заполнены, можно заполнить оставшиеся ячейки.
- Определить максимальный и минимальный размер кредитной линии, предоставляемый клиентам, и назначить их наилучшим и наихудшим клиентам в таблице. Заполнить оставшиеся ячейки, постепенно увеличивая или уменьшая размер кредитной линии.
- На основе ожидаемых потерь (*Expected Loss, EL*) для выбранной когорты, ожидаемой вероятности дефолта (*Probability of Default, PD*), доли потерь в случае дефолта (*Loss Given Default, LGD*) и величины средств под риском

(*Exposure at Default, EAD*) можно назначить оптимальный максимальный размер риска (кредитный лимит) для каждой ячейки таблицы. Необходимо выдвинуть некоторые предположения касательно распределения ожидаемых убытков. Для удобства предположим, что потери на одного клиента в ячейке составляют 500\$. Если вероятность дефолта равна 5%, а доля потерь в случае дефолта составляет 97% от лимита, то максимальный кредитный лимит для этой ячейки можно вычислить по формуле:

$$\begin{aligned}EL &= EAD \times PD \times LGD \\500 &= EAD \times 0.05 \times 0.97 \\EAD &= 10309\$ \end{aligned}$$

Изложение этих трех вариантов не ставит своей целью рекомендовать их в качестве окончательных способов управления кредитной линией. Во-первых, выбор параметров, используемых в принятии решений, играет ключевую роль. Многие организации применяют только оценку риска, чтобы определить сумму кредитной линии или кредита. Но это лишь одна сторона медали. Как показано на рисунке 8.8, баланс между вероятностью выплаты и способностью оплачивать кредит имеет большее значение. Во-вторых, по каждому решению нужно оценить варианты, начиная от самого простого и заканчивая самым сложным. Иногда самые простые варианты, такие, как первые два, приведенные выше, являются наилучшими. Они являются простыми, надежными и основаны на двух понятных метриках.

Правила кредитной политики

Правила кредитной политики представляет собой принятый банком перечень принципов, которого необходимо придерживаться в ходе принятия решений по кредиту. Они включают в себя юридические и связанные с минимизацией риска правила, такие как, минимальные требования (стоп-факторы) и формулировки риска:

- Требование совершеннолетнего возраста (например, если возраст <18, то отклонить)
- Занятость (например, если не работает, самозанятый, работает меньше года, то отклонить или отправить на дополнительное рассмотрение)
- Банкротство (например, недавнее банкротство или банкротство менее 2 лет назад, то отправить на дополнительное рассмотрение или отклонить)
- Лимит на количество просрочек (например, если более трех просрочек по данным бюро, то отклонить)
- Тип заявителя (например, если VIP, то отправить на дополнительное рассмотрение)
- Предыдущие внутренние сведения о заявителе (если есть кредит, списанный как безнадежный долг, то отправить на дополнительное рассмотрение).
- Минимальный размер кредитной линии для случаев, где есть пересмотр решения скоринговой карты, или определенных типов клиентов (например, студент или клиент, у которого был дефолт в прошлом)
- Запрет на увеличение кредитной линии, если не прошло шести месяцев

Правила кредитной политики являются неотъемлемой частью риск-менеджмента, предупреждающей банк от неосторожных решений. Важно, что они должны

основываться на независимых и надежных показателях. Нежелательно их разрабатывать на основе характеристик скоринговой карты, т.к. это снизит ценность скоринга. Например, если информация о банкротстве или просрочках по предыдущим кредитам включается в скоринговую карту, то нежелательно использовать данные критерии для определения правил кредитной политики. Когда использование таких правил критично, лучше построить скоринговые карты с применением других показателей.

Правила кредитной политики часто носят субъективный характер и редко тестируются опытным путем. Некоторые правила существуют лишь потому, что их приняли много лет назад и после этого никто не потрудился заново пересмотреть их. Когда это возможно, правила должны время от времени (например, ежегодно) пересматриваться, тестироваться и получать четкое обоснование. Иногда разработка скоринговой карты на этапе анализа исходных характеристик может подтвердить эффективность некоторых правил. Однако заметим, поскольку некоторые правила используются в качестве общего отказа без обсуждения причин, необходимо минимизировать эффект «наиболее понравившихся» заявителей. Для этого нужно использовать данные, полученные после проведения анализа отклоненных заявок.

Правила кредитной политики часто используются для пересмотров решений скоринговой карты (принятие противоположных решений, вопреки порогу отсека) в определенном скоринговом диапазоне. Как и стратегии, они должны разрабатываться при участии операционного отдела, юридического отдела, кредитного отдела, чтобы учесть все возможные последствия.

Пересмотры решений скоринговой карты (Overrides)

Пересмотром называют ручную или автоматизированную процедуру, которая меняет решение скоринговой карты на противоположное, заданное на основе порогов отсека (т.е. данная процедура опровергает решение, выданное скоринговой картой). Она применяется для скоринга новых заявителей. Есть два типа пересмотров:

1. **«Жесткий нижний порог».** Заявители, находящиеся под отсечкой, одобряются
2. **«Жесткий верхний порог».** Заявители, находящиеся над отсечкой, отклоняются.

Как и правила кредитной политики, пересмотры – необходимая часть риск-менеджмента, предупреждающая от неосторожных решений и требующая правильного применения. Общее правило для пересмотра решения в том, что он должен приниматься на основе значимой доступной информации, не зависящей от скоринговой карты. Поскольку скоринговые карты обычно разрабатываются с использованием эмпирических методов и представляют собой анализ тысяч наблюдений на протяжении многих лет, будет правильным применять пересмотры решений только в том случае, если вы знаете что-то, чего не знает скоринговая карта. Приведем подобные ситуации:

- Правила кредитной политики, обсуждавшиеся выше.
- Локальное знание. Например, специалист в филиале банка может знать заявителя и его ближайшее окружение, имеет возможность использовать такие характеристики, как семейный доход, трудовой стаж на последнем месте работы и т.п. для принятия более правильного решения. Это может быть как положительная, так и негативная информация.

- Допустимое смягчение требований. Например, заявитель не работал, имел пропущенные платежи, но теперь у него хорошо оплачиваемая работа, или студент пропустил платежи на незначительную сумму во время каникул.
- Необходимые проверки и справки. Например, заявитель не способен предоставить документы, подтверждающие доход, для получения ипотеки/автокредита.
- Другие исключительные обстоятельства, когда уровень риска, спрогнозированный с помощью одной лишь скоринговой карты, недостаточен для принятия решения.

Доля заявок, по которым пересмотрено решение скоринговой карты (или процент пересмотров), зависит от продукта и суммы кредита. Обычно для мелких/крупных продуктов, где используется автоматическое принятие решений (например, кредитные карты), устанавливают очень низкий процент пересмотров. Однако, для крупных продуктов, где решения принимаются вручную и требуют гораздо более тщательной проверки (кредиты предприятиям и ипотечные кредиты), устанавливают более высокий процент пересмотров. Процент пересмотров будет также выше в условиях, когда скоринговая карта считается скорее дополнительным инструментом поддержки (которое рассматривается наравне со многими другими инструментами), чем единственным источником принятия решений. Кроме того, когда скоринговые карты разрабатываются с небольшим количеством характеристик (вместо расширенной скоринговой карты «профиля риска»), процент пересмотров чаще всего высокий. Это обычно происходит из-за того, что скоринговая карта фиксирует несколько характеристик и поэтому гораздо больше информации оказывается за пределами скоринговой карты. В свою очередь скоринговые карты, построенные на узком наборе характеристик, вызывают у специалистов меньше доверия, заставляя в большей степени предугадывать решения скоринговой карты.

В любом случае процент одобрения заявителей под отсечкой нужно держать на каком-то минимальном уровне – и когда это допустимо, необходимо отслеживать работу модели, желательно выявив «причину пересмотра решения». Кроме того, каждое испытание должно быть проведено так, чтобы причиной не могла стать какая-либо «противоречивая» характеристика. Иначе это даст основание утверждать, что пересмотр решения носит субъективный характер, а не опирается на более веские основания. Необходимо гарантировать, что пересмотры решений основываются на опыте (а точнее, его отсутствии), а «внутреннее чутье» проверено с помощью показателей эффективности и применяется обдуманно.

Представители регулирующих органов требуют, чтобы доля пересмотренных решений в скоринге составляла не более 3-5%. Важно провести мониторинг конкретных причин пересмотров решений и мониторинг трендов. В ипотечном кредитовании более высокий процент пересмотров для заявителей, входящих в защищаемый от дискриминации сегмент (например, пол, раса, религиозная принадлежность, сексуальная ориентация и др.), признаются как ущемляющие права.

Процесс Разработки Скоринговой Карты, Этап 7: Мероприятия после Внедрения

Мероприятия после внедрения скоринговой карты включают два основных участка работ:

1. Подготовка отчетов
2. Экспертиза

Отчеты по мониторингу скоринговой карты и кредитного портфеля

Настоящий раздел посвящен подготовке некоторых стандартных отчетов, используемым риск-менеджерами для мониторинга эффективности скоринговой карты и кредитного портфеля. Большая часть управленческих отчетов по скоринговой карте и кредитному портфелю разрабатывается, используя такие показатели эффективности кредитного портфеля и скоринговой карты, как процент одобрения, процент «плохих», процент пересмотров и различные индексы. Однако есть более важные задачи и причины, в силу которых необходимо подготовить данные отчеты. Данные задачи и отчеты, используемые для их решения, изложены ниже.

Управленческие отчеты по скоринговой карте и заявителям:

- Подтвердить предположение о том, что «будущее похоже на прошлое». Скоринговые карты разработаны по определенному профилю заявителя или клиента (представлены распределением скоринговых баллов или значений характеристик). Данное предположение нужно проверить на текущих заявителях.
 - Отчет «Стабильность системы» (также называемый «Стабильность популяции» и «Стабильность скоринговой карты»)
 - Анализ характеристик скоринговой карты
 - Анализ характеристик, не вошедших в скоринговую карту
- Провести мониторинг и выявить причины изменений, произошедших в профиле заявителей и одобренных (или клиентов в поведенческом скоринге). Недостаточно просто знать, что произошло изменение, поскольку это не позволит предпринять конкретные действия. Необходимо выявить источник (причину) изменения.
 - Анализ характеристик, вошедших и не вошедших в скоринговую карту
 - Анализ действий конкурентов и анализ маркетинговых кампаний
 - Анализ по регионам и другим сегментам
- Изучить профиль риска текущих клиентов и заявителей.
 - Отчет «Стабильность системы»

- Анализ характеристик, вошедших и не вошедших в скоринговую карту
- Отчет «Распределение скоринговых баллов по одобренным заявителям/клиентам»
- Вывести показатели по принятию/пересмотру решений скоринговой карты.
 - Итоговый скоринговый отчет
 - Отчет «Пересмотры решений скоринговой карты»

Управленческие отчеты по кредитному портфелю:

- Провести мониторинг кредитоспособности клиентов.
 - Отчет «Просрочки»
 - Винтажный анализ
 - Отчет «Миграция просрочек»
 - Отчет «Уровни перехода» (Roll Rate)
- Провести мониторинг и выявить источники просрочек и прибыли. Как уже говорилось ранее, знание причин убытков помогает нам предпринять решения, скорректированные с учетом риска.
 - Отчет «Просрочки», по регионам и другим сегментам
 - Анализ маркетинговых кампаний и действий конкурентов
- Оценить уровень будущих убытков.
 - Винтажный анализ и отчет «Уровни перехода» (Roll Rate)
- Оценить спрогнозированный процент «плохих» и установить его приемлемый уровень. Сравнение фактической кредитоспособности с прогнозируемой позволит скорректировать прогнозы будущих потерь.
 - Винтажный анализ
 - Отчет «Просрочки»
 - Отчет «Миграция просрочек»

Все выше сказанное применяется как в отчетах по аппликационным скоринговым картам, так и в отчетах по поведенческим скоринговым картам.

Не нужно говорить, что тщательная подготовка отчетов и исследование структуры данных должны быть осуществлены до внедрения скоринговой карты, чтобы можно было изучить первые результаты, полученные с помощью новой скоринговой карты или стратегии. На начальных этапах внедрения рекомендуют подготавливать еженедельные отчеты для того, чтобы любые отклонения от ожидаемой работы модели могли быть быстро выявлены и исправлены. Когда установлено, что скоринговые карты или стратегии работают так, как и ожидалось, вполне достаточно регулярных ежемесячных или квартальных отчетов.

Несмотря на то, что кредитование – динамический процесс, он меняется довольно медленными темпами. В отличие от рыночного риска, где флуктуации факторов риска могут происходить ежеминутно, индикаторы и факторы розничного кредитного риска имеют тенденцию меняться в течение более длительного периода времени. Поэтому важно выявлять тренды, а не случайные колебания данных, особенно в тех ситуациях, когда есть необходимость изменить решение. В разделе «Валидация перед внедрением» главы 8 было рекомендовано подготавливать отчет «Стабильность системы» по заявителям за последние три и

шесть месяцев. Причиной было выявление долговременных трендов перед тем, как решить, является ли скоринговая карта все еще надежной или нет. Нет особой разницы в том, как мы рассуждали тогда и сейчас – пытаюсь решить, является ли скоринговая карта все еще надежной, определить необходимость изменения порога отсечения для отдельного сегмента, ввести более жесткий контроль за просрочками и выдачей кредита для определенной когорты или сегмента, в любом случае вы должны убедиться в том, что результаты указывают скорее на долгосрочный непрерывный тренд, чем на разовое событие.

В конечном счете, как было подчеркнуто в главе, посвященной разработке скоринговой карты, нужно объяснить причины изменений, произошедших в профиле заявителя и его кредитоспособности. Недостаточно просто взглянуть на процент одобрения, процент «плохих» или индексы стабильности – чтобы предпринять грамотные/скорректированные с учетом риска решения, вы должны объяснить, почему эти показатели стали такими.

Управленческие отчеты по скоринговой карте

Эти отчеты часто называют «фронт-энд» отчетами, потому что они используются для мониторинга входящего потока заявителей.

Отчет «Стабильность системы/популяции/скоринговой карты» Об этом отчете рассказывалось ранее (в Главе 8), когда освещалась валидация перед внедрением скоринговой карты. Во время валидации, предшествующей внедрению скоринговой карты, распределения скоринговых баллов в самой последней группе заявителей сравнивались с распределениями скоринговых баллов в обучающей выборке, чтобы выявить какие-либо изменения в профиле заявителя. Здесь задача и метод те же самые. В этом отчете также сравнивается распределение скоринговых баллов в текущей выборке («фактические») с распределением скоринговых баллов в обучающей выборке («ожидаемые»). Это делается для того, что обнаружить смещения профиля заявителя, показанные с помощью распределения по скоринговым баллам. Отчет можно легко модифицировать для использования в поведенческом скоринге, сравнивая распределения существующих клиентов с распределениями по баллам в обучающей выборке. Это сравнение включает два аспекта:

1. Оно проверяет предположение «прошлое отражает будущее» (т.е. прогнозы процента «плохих» основаны на профиле будущего заявителя, который похож на профиль заявителя, использованный для построения модели). Хотя и неокончательно, но доказательство схожести обеспечивает приемлемый уровень прогноза.
2. Оно дает представление о качестве заявителей/клиентов (например, если происходит смещение баллов вниз, то это может указывать на ухудшение качества заявителей или существующих клиентов).

Как правило, отчеты «Стабильность системы» подготавливаются ежемесячно, однако для устойчивых кредитных портфелей достаточно ежеквартальных отчетов. Отметим, что для строгого отчета «Стабильность системы» популяция заявителей должна быть сформирована по тем же самым критериям, что и обучающая выборка. Однако компании для определения порога отсечения обычно проводят повторный анализ, который включает всех заявителей. Он дает более реалистичную оценку ожидаемого процента одобрения и результатов разбиения при выбранном пороге по ключевым сегментам.

Пример отчета «Стабильность системы» показан на рисунке 9.1. Отметим, что отчет «Стабильность системы» иногда также называют отчет «Стабильность популяции» или «Стабильность скоринговой карты».

Столбцы «Текущие %» и «Ожидаемые %» показывают распределение наблюдений в текущей и обучающей выборках для каждого заданного диапазона баллов. Заметим, что на рисунке баллы сгруппированы так, что каждый диапазон баллов включает 10% от общего числа «ожидаемых». Это сделано с целью эффективности, так, чтобы можно было легко выявить любые смещения вверх или вниз (установив 10%-ный порог для каждой группы).

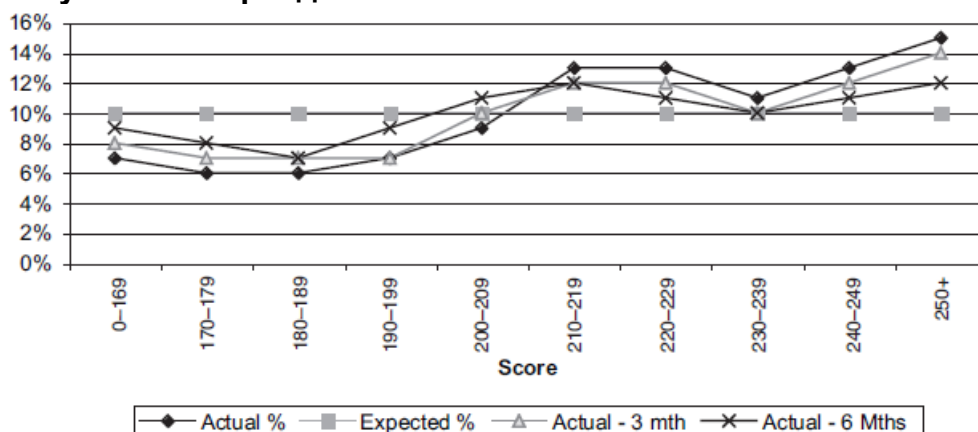
Есть два вида анализа, которые можно выполнить с помощью рисунка 9.1. Первый вид анализа подразумевает, что причину смещения популяции можно выявить, изучив график распределения текущих и ожидаемых заявителей/клиентов по скоринговым баллам. Это может дать дополнительную информацию (является ли смещение баллов нисходящим, восходящим или эксцессом). Кроме того, отображение на том же самом графике распределений по историческому периоду вместе с распределением за последний месяц или квартал (например, за последние три месяца, за последние шесть месяцев и т.д.) поможет выявить долгосрочные тренды. Это поможет выяснить, например, происходит ли медленное ухудшение качества клиента, склонно ли распределение к циклическим изменениям. Изменение, которое представляет собой долгосрочный тренд, может рассматриваться как устойчивое событие и поэтому зависит от изменений кредитной политики и других решений. Данная ситуация является лучшим поводом для принятия решений, чем ответные меры на ежемесячные изменения, которые не являются стабильными.

Рисунок 9.1 Отчет «Стабильность системы»

<i>Score Range</i>	<i>Actual %</i>	<i>Expected %</i>	<i>(A-E)</i>	<i>A/E</i>	<i>ln(A/E)</i>	<i>Index</i>
0-169	7%	10%	-3%	0.7000	-0.3567	0.0107
170-179	6%	10%	-4%	0.6000	-0.5108	0.0204
180-189	6%	10%	-4%	0.6000	-0.5108	0.0204
190-199	7%	10%	-3%	0.7000	-0.3567	0.0107
200-209	9%	10%	-1%	0.9000	-0.1054	0.0011
210-219	13%	10%	3%	1.3000	0.2624	0.0079
220-229	13%	10%	3%	1.3000	0.2624	0.0079
230-239	11%	10%	1%	1.1000	0.0953	0.0010
240-249	13%	10%	3%	1.3000	0.2624	0.0079
250+	15%	10%	5%	1.5000	0.4055	0.0203
Index						0.1081

На рисунке 9.2. на одном и том же графике показано сравнение распределений ожидаемых заявителей, текущего потока самых последних заявителей, текущего потока заявителей за последние три месяца, текущего потока заявителей за последние шесть месяцев. Количество текущих и ожидаемых заявителей взято из рисунка 9.1. На рисунке 9.2. четко показано, что по сравнению с ожидаемыми текущие заявители последовательно получают более высокие баллы, а наиболее высокие баллы – у текущих заявителей за последние шесть месяцев.

Рисунок 9.2 Тренд «Стабильность системы»



Второй вид анализа исходит из того, что величину смещения популяции можно измерить математически. Он включает только измерение различия между обоими распределениями. Для этого используется индекс стабильности популяции. Он измеряет величину смещения популяции, сравнивая распределения текущих и ожидаемых заявителей. Индекс вычисляется следующим образом:

$$\Sigma(\% \text{ Текущие} - \% \text{ Ожидаемые}) \times \ln(\% \text{ Текущие} / \% \text{ Ожидаемые})$$

по всем диапазонам скоринговых баллов.

Вычисление происходит точно так же, как и вычисление IV, рассмотренное ранее в Главе 6. Индекс измеряет отклонение между распределением в текущей выборке и распределением в обучающей выборке. В целом индекс можно интерпретировать следующим образом:

- Индекс меньше 0.10 показывает отсутствие значимого изменения популяции.
- Индекс 0.10–0.25 говорит о незначительном изменении, которые необходимо исследовать.
- Индекс выше 0.25 свидетельствует о значительном смещении популяции заявителей.

Для оценки смещения могут также использоваться другие методы, такие, как хи-квадрат с заданным уровнем значимости. Метод, приведенный здесь, широко используется кредитными аналитиками.

Данный индекс не является универсальной метрикой. Он лишь дает общее представление. Фактически он только говорит нам, произошло ли смещение популяции или нет, и дает представление о величине смещения. Прежде чем решить, что смещение популяции является значимым, необходимо принять во внимание такие факторы, как тренды (изменение носит временный или долгосрочный характер?), величину смещения и причины изменения. Из всего вышеприведенного, что можно использовать в отчете для принятия решения, самая главная задача – найти причины смещений баллов.

Изменения скоринговых баллов могут быть вызваны несколькими причинами:

- Объективные изменения в профиле заявителя (например, возраст).
- Динамика рынка, которая включает маркетинговые кампании, конкурентную борьбу за рыночные ниши и разработку продукта. Например, если заявители, отобранные за последний месяц, были значительно моложе или

были из определенного региона, это могло быть вызвано маркетинговой активностью. Изменения продукта, такие как, дополнительные условия по программам лояльности, изменения в структуре оплаты кредита – «0% на срок 6 месяцев» и прочие «бонусы», или переключение на нетрадиционные каналы продаж могут также привлечь новых заявителей. Внешняя конкуренция за привлечение узких демографических групп также может повлиять на состав ваших заявителей. Это могут быть мероприятия, которые позволяют выделить целевые группы покупателей с помощью изменений условий продукта, программ лояльности и более высоких процентных ставок. Например, банк, у которого не было программ лояльности по своим кредитным картам, обнаружил, что качество его заявителей стало хуже со временем. Анализ показал, что большая часть его заявителей были отклонены другими банками, предлагавшими программы лояльности по своим кредитным картам (т.е., те, кто смог, получили кредитные карты, предлагавшие дополнительные бонусы, в то время как те, кто не смог, ушли в этот конкретный банк). В поведенческом скоринге более агрессивные акции или кредитные стратегии, внедрение программ лояльности, пересмотр ставок, кросс-продажи могут изменить скоринговую оценку профиля существующего клиента.

- Ошибки ввода данных.
- Ошибки в сборе данных, данные представляют собой неслучайную или неправильно сформированную выборку или включены наблюдения, исключенные из обучающей выборки.

Есть несколько способов провести дальнейший анализ причин смещений. Они включают:

- Анализ характеристик скоринговой карты. (Он может показаться очевидным, но выполняется, потому что скоринговые баллы основаны на характеристиках скоринговой карты. Кроме того, он подчеркивает важность построения скоринговой карты на основе расширенного «профиля риска», чтобы выявленные смещения отражали реальность, а не были просто следствием флуктуаций одной или нескольких характеристик в скоринговой карте, построенной на узком наборе характеристик.)
- Анализ смещений характеристик, не вошедших в скоринговую карту. Он включает три главные категории характеристик:
 - Сильные характеристики, которые не вошли в скоринговую карту. (Поскольку уже есть ранжирование риска, то их распределения помогут достоверно ответить на вопрос, становится популяция лучше или хуже.)
 - Характеристики, связанные с характеристиками, включенными в скоринговую карту (Они должны изменяться в одном направлении вместе с характеристиками, включенными в скоринговую карту, таким образом подтверждая смещение характеристик скоринговой карты)
 - Числители и знаменатели коэффициентов в скоринговой карте.
- Сбор информации об изменениях нормативных документов, последних маркетинговых кампаниях, проводимых внутри банка и конкурентами, и связанных с ними изменениях продукта.

Отчет «Анализ характеристик скоринговой карты» Отчет «Анализ характеристик скоринговой карты» сравнивает текущее распределение каждой характеристики скоринговой карты с ожидаемым и оценивает влияние смещений

распределения на скоринговые баллы. Этот отчет помогает в дальнейшем провести анализ причин смещений баллов, обычно его подготавливают ежеквартально или всякий раз, когда отчет «Стабильность системы» выявляет значительное смещение популяции. Как и в случае с отчетом «Стабильность системы», снова рекомендуется сравнить ожидаемые распределения с текущим распределением, а также с текущими распределениями за различные периоды времени, чтобы выявить тренды.

На рисунке 9.3 приведен пример анализа характеристики «Возраст», включенной в скоринговую карту.

Столбцы «Текущие %» и «Ожидаемые %» снова показывают распределение наблюдений в текущей и обучающей выборках. Есть два вида анализа, которые можно выполнить по этим данным. В первом виде анализа, как показано на рисунке 9.2, строится график, на котором откладываются ожидаемое, текущее, а также текущее, взятое за различные периоды времени, распределение каждой характеристики. Он даст визуальное представление о причине смещений.

Во втором виде анализа мы можем вычислить величину этого смещения математически. В этой ситуации мы оценим влияние этого смещения в баллах. Это влияние или индекс характеристики скоринговой карты вычисляется просто

Индекс вычисляется просто:

$$\sum (\% \text{ Текущие} - \% \text{ Ожидаемые}) \times (\text{Баллы})$$

по всем диапазонам скоринговых баллов.

Заметим, что включение столбца «% Одобрённых по критерию» в данный анализ может дать представление о том, является ли скоринговая карта дискриминирующей в отношении выделенного сегмента.

Рисунок 9.3 показывает смещение в пользу более молодых заявителей. Текущие заявители по характеристике «возраст» получают на 2.63 балла меньше, чем ожидаемые.

Рисунок 9.3 Анализ характеристик скоринговой карты

<i>Age</i>	<i>Expected %</i>	<i>Actual %</i>	<i>Points</i>	<i>Index</i>
18-24	12%	21%	10	0.9
25-29	19%	25%	15	0.9
30-37	32%	28%	25	-1
38-45	12%	6%	28	-1.68
46+	25%	20%	35	-1.75
Index				-2.63

Аналогичные отчеты подготавливают по всем характеристикам скоринговой карты и обычно размещают для анализа на одной странице. Пример такого отчета (фрагмент аппликационной скоринговой карты) приведен на рисунке 9.4.

Анализируя характеристики, можно сделать вывод:

- Заявители становятся моложе (более рискованные).
- У них меньше срок проживания по постоянному адресу (более рискованные). Другие характеристики, не вошедшие в скоринговую карту и зависящие от времени, должны меняться в одном и том же направлении.
- Они еще не меняли регион проживания.

- У них значительно большее число запросов в кредитное бюро за последние шесть месяцев (более рискованные). Анализ запросов за последние три месяца и за последние двенадцать месяцев подтвердит, являются ли эти запросы краткосрочным явлением или носят постоянный характер.
- Число просрочек по данным бюро стало больше (более рискованные). Другие характеристики, полученные от бюро и зависящие от просрочки, должны подтвердить это.
- Утилизация револьверного кредита по данным бюро стала меньше (менее рискованные).

Рисунок 9.4 Полный отчет по всем характеристикам скоринговой карты

<i>Age</i>	<i>Expected</i>	<i>Actual</i>	<i>Points</i>	<i>Index</i>	<i># Delq</i>	<i>Expected</i>	<i>Actual</i>	<i>Points</i>	<i>Index</i>
18-24	12%	21%	10	0.9	0	80%	65%	45	-6.75
25-29	19%	25%	15	0.9	1-2	12%	21%	20	1.8
30-37	32%	28%	25	-1	3-5	5%	8%	12	0.36
38-45	12%	6%	28	-1.68	6+	3%	6%	5	0.15
46+	25%	20%	35	-1.75					-4.44
				-2.63	<i>Utilization at Bureau</i>				
<i>Time at Res</i>					0	12%	8%	15	-0.6
0-6	18%	29%	12	1.32	1-9	10%	19%	40	3.6
7-18	32%	32%	25	0	10-25	14%	20%	30	1.8
19-36	26%	22%	28	-1.12	26-50	22%	25%	25	0.75
37+	24%	17%	40	-2.8	50-69	11%	6%	20	-1
				-2.6	70-85	13%	9%	15	-0.6
<i>Region</i>					86-99	14%	8%	10	-0.6
Major Urban	55%	58%	20	0.6	100+	4%	5%	5	0.05
Minor Urban	26%	24%	25	-0.5					3.4
Rural	19%	18%	15	-0.15					
				-0.05					
<i>Inq 6 mth</i>									
0	63%	34%	40	-11.6					
1-3	19%	31%	30	3.6					
4-5	10%	16%	15	0.9					
6+	8%	19%	10	1.1					
				-6					

Несмотря на то, что все характеристики указывают на группу высокорискованных заявителей, уровень утилизации свидетельствует об обратном. Тенденции, похожие на вышеописанные, нелогичны и должны быть тщательно изучены. В этом случае отдельно анализируют суммы балансов и суммы кредитных линий (т.е. числитель и знаменатель для расчета утилизации). Несмотря на то, что средние суммы балансов показали небольшое увеличение за наблюдаемый период, кредитные линии увеличились на более значительную сумму. Это произошло в условиях высокой конкуренции, когда банки увеличили клиентам суммы кредитных линий, чтобы выглядеть более конкурентноспособными - очевидно не принимая в расчет увеличившуюся рискованность этих клиентов. Поэтому заявители получили более высокий скоринговый балл по утилизации, но это не свидетельствует о более низком риске.

Вышесказанное подчеркивает, необходимо дополнять анализ скоринговых баллов бизнес-логикой, не ограничиваясь просто вычислением смещений баллов.

В таких ситуациях, как эта, когда риск-менеджер опасается, что заявители получают более высокие скоринговые баллы, чем нужно, можно рассмотреть несколько настроек:

- Уменьшить скоринговые баллы по утилизации для высокорискованных заявителей.
- Увеличить порог отсечения.
- Скорректировать ожидаемый процент потерь/процент «плохих», чтобы учесть более высокий фактический риск, чем тот риск, который показывают скоринговые баллы.

Анализ качественно улучшен за счет того, что от простого вычисления баллов переходит к ключевому индикатору риска (KRI). Он заключается в сравнении распределений остальных характеристик, не вошедших в скоринговую карту, в обучающей и текущей выборках. Эти характеристики могут включать:

- Характеристики, не вошедшие в скоринговую карту, но предположительно влияющие на качество заявителя. Они включают надежные характеристики, которые не являлись таковыми, будучи включенными в скоринговую карту. Например, если смещения характеристик скоринговой карты указывают на ухудшение качества заявителей, то анализ характеристик, не включенных в скоринговую карту, поможет подтвердить данный факт. В основе анализа предположение, что оценка качества заявителя опиралась на различные источники информации и поэтому более устойчива.
- Характеристики, схожие с характеристиками скоринговой карты. Например, если «возраст» включен в скоринговую карту, мониторинг характеристик, зависящих от времени, таких, как стаж работы, стаж проживания, возраст первого кредитного счета – могут дополнительно объяснить причины изменения профиля заявителя. Если «запросы за последние 6 месяцев» включены в скоринговую карту, мониторинг запросов за последние 3 и за последние 12 месяцев поможет в дальнейшем объяснить изменение количества запросов. На рисунке 9.4 распределение по характеристике «запросы за последние 12 месяцев» не показало значительного изменения. Это значит, что увеличение запросов может быть вызвано возросшей конкуренцией между банками за последние шесть месяцев или свидетельствует о том, что такие факторы, как изменение условий продукта или маркетинг в течение последних шести месяцев, могли привлечь клиентов, наиболее сильно нуждающихся в кредите. Смещение значений связанных характеристик должно быть однонаправленным (например, более старшие по возрасту заявители имеют больший стаж работы и больший возраст кредитной истории). Примеры таких характеристик – кредитные линии, социально-демографические характеристики, запросы и финансовые коэффициенты.
- Если в скоринговой карте используются коэффициенты, то необходимо исследовать распределение знаменателя и числителя, чтобы объяснить изменение самого коэффициента. Например, если утилизация (баланс, деленный на кредитный лимит) уменьшилась, это может быть обусловлено либо снижением балансов, либо увеличением кредитных лимитов (как показано на рисунке 9.4).

Кроме того, данный анализ следует выполнить на основе сравнения обучающей выборки и выборки текущих заявителей, взятых за период от одного до шести месяцев, чтобы обнаружить какие-либо тренды, и проверить, не являются ли смещения распределений временным явлением.

Смещение характеристик (как включенных, так и не включенных в скоринговую карту) должно согласовываться и объясняться логикой и деловыми

соображениями. Например, если анализ показывает, что заявители «помолодели», но при этом у них большой «стаж работы», то это указывает, что при вводе данных могли быть допущены ошибки. Такие нелогичные смещения должны быть тщательно исследованы и объяснены.

Итоговый скоринговый отчет Итоговый скоринговый отчет применяется для аппликационных скоринговых карт. Его составляют для получения рабочих показателей, таких, как процент одобрения и процент пересмотров. Его также можно использовать, чтобы определить избыточную долю пересмотренных решений, а также оценить качество заявителей и одобренных клиентов. На рисунке 9.5 приведен обычный итоговый скоринговый отчет, использующий скоринговую карту с порогом отсечения 200 баллов.

Рисунок показывает процент одобрения 64.2%, процент пересмотров для нижнего и верхнего порогов составляет 4.84% и 9.09% соответственно. Отметим, что некоторые кредитные организации вычисляют процент пересмотров на основе общего количества заявителей в качестве знаменателя, тогда как другие используют в качестве знаменателя количество заявителей ниже или выше порога отсечения (для нижнего и верхнего порогов пересмотра решений).

Для таких продуктов, как ипотека и кредиты, создается дополнительный столбец «Непринятые», чтобы показать заявителей, которые были одобрены кредитной организацией, но не приняли предложения. В ходе разработки скоринговой карты их рассматривают как «серую зону».

Отчет на рисунке 9.5 обычно подготавливают для каждой скоринговой карты, но большинство пользователей разрабатывают его для различных групп и сегментов. Это делается для того, чтобы убедиться, важные сегменты адекватно представлены и процент одобрения по скоринговым баллам является постоянным для различных сегментов. Невыполнение данного условия может обозначать, что либо скоринговая карта не прогнозирует так, как планировалось, либо имеет место чрезмерное количество пересмотров решений. В первом случае может потребоваться альтернативная сегментация, тогда как во втором потребуются дальнейший анализ и настройка требований.

Рисунок 9.5 Итоговый скоринговый отчет

<i>Score Range</i>	<i>Applicants</i>	<i>Approved</i>	<i>% Approved</i>	<i>Lowside</i>	<i>Highside</i>
0-169	700	—	0%	0	
170-179	800	16	2%	16	
180-189	700	35	5%	35	
190-199	900	99	11%	99	
200-209	1,100	924	84%		176
210-219	1,100	968	88%		132
220-229	1,000	900	90%		100
230-239	1,200	1,092	91%		108
240-249	1,100	1,045	95%		55
250+	1,400	1,344	96%		56
	10,000	6,423	64.2%	150	627
Above Cutoff	6,900	6,273		4.84%	9.09%
Below Cutoff	3,100	150			

На рисунке 9.6. приведен мониторинг качества зарегистрированных клиентов по месяцам.

На рисунке приведены распределения зарегистрированных клиентов по скоринговым баллам за три периода времени. Они сравниваются с ожидаемым

распределением, чтобы определить, стало ли лучше или хуже качество клиентов по сравнению с тем, что ожидалось. На рисунке 9.6 показана постоянная величина, используемая для отклонения заявок, когда доля одобренных в диапазоне баллов 200-209 (выше порога отсечения) увеличивается, в то время как доля тех, кто находился в начальных диапазонах, уменьшается. Кроме того, процент пересмотренных решений по нижнему порогу отсечения не только постепенно увеличивается, но и растет с уменьшением скорингового балла.

Рисунок 9.6 Качество клиента

<i>Score</i>	<i>Expected</i>	<i>Q1 03</i>	<i>Q2 03</i>	<i>Q3 03</i>	<i>Q4 03</i>
0-169	0%	0%	1%	1%	
170-179	0%	1%	3%	4%	
180-189	0%	2%	3%	5%	
190-199	0%	2%	4%	5%	
200-209	20%	26%	28%	26%	
210-219	15%	19%	20%	20%	
220-229	20%	22%	22%	20%	
230-239	16%	15%	12%	11%	
240-249	18%	4%	3%	5%	
250+	11%	9%	4%	3%	
Total	100%	100%	100%	100%	

Наряду с тем, что отчет «Стабильность системы» и «Итоговый скоринговый отчет» показывают качество заявителей, отчет «Качество клиента» дает представление о качестве одобренных клиентов. Этот отчет помогает получить более оптимальный ожидаемый процент «плохих» по каждой новой группе клиентов вместо того, чтобы полагаться на простую зависимость процент одобрения/процент «плохих», взятую из таблицы выигрышей. Данная зависимость – говорящая вам, что если ваш процент одобрения составляет, например, 70%, то вы можете ожидать общий процент «плохих» 3% - основана на предположении, что распределение клиентов выше порога отсечения остается стабильным. Как видно на рисунке 9.6, как только это предположение перестает быть достоверным, нужно заново построить прогнозы процента «плохих». В некоторых случаях процент одобрения может остаться тем же самым для данного портфеля, но если состав портфеля меняется, прогнозы, взятые из таблицы выигрышей, больше не являются достоверными.

Отчет «Пересмотры решений скоринговой карты» Этот отчет проводит мониторинг пересмотров решений по нижнему и верхнему порогу отсечения (одобрение клиентов под отсечкой, отклонение клиентов над отсечкой), исследуя причины пересмотров. В ситуации, когда решения принимались как вручную, так и автоматически (т.е. с помощью программного обеспечения), необходимо провести мониторинг обоих типов решений.

Как уже обсуждалось в разделе «Пересмотры решений скоринговой карты» Главы 8, несмотря на то, что некоторые пересмотренные решения обоснованы и должны быть приняты, их избыточное и неконтролируемое число приводит к увеличению потерь. Поэтому этот отчет является средством контроля, которое предупреждает руководство о ситуациях, когда количество пересмотров возрастает или когда пересмотры решений не обоснованы. Также его можно использовать для определения качества пересмотренных решений. По мере возможности все пересмотры решений должны основываться на законных и проверяемых причинах. Поэтому данный анализ можно выполнить, чтобы определить, какие причины пересмотров допустимы, а от каких причин следует отказаться. Один из

способов сделать это – подготовить отчет по эффективности на основе причин пересмотров.

На рисунке 9.7 показана выдержка из отчета «Пересмотры решений по скоринговой карте».

Рисунок 9.7 показывает пересмотренные заявки по причине и типу решения. Заметим, что трактовки автоматических и принятых вручную решений будут различными для банков в зависимости от того, как устроен в них процесс одобрения. Некоторые организации требуют, чтобы кредитный аналитик повторно рассмотрел и подтвердил автоматические решения по всем заявителям, в то время как другие автоматически одобряют или отклоняют большую часть заявителей, пересматривая вручную лишь небольшое количество. Поэтому этот отчет, показанный на рисунке 9.7, можно модифицировать, включив пункты «Предварительное решение» и «Итоговое решение» вместо автоматических и принятых вручную решений.

Рисунок 9.7 Отчет «Пересмотры решений скоринговой карты»

<i>Override Reason</i>	<i>Number</i>	<i>System 'D'</i>	<i>System 'A'</i>	<i>Manual 'D'</i>	<i>Manual 'A'</i>
Lowside					
Local Knowledge	34	34	0		34
Justifiable Delq	96	96	0		96
VIP	12	12	0		12
VP Override	8	8	0		8
	150	150	0		150
Highside					
Bankruptcy	125	120	0		5
Local Knowledge	102	0	102	102	
Derogatory	200	0	200	185	15
Policy 1	55	55	0		
Policy 2	73	73	0		
Policy 3	92	92	0		
	647	340	302	287	20

Также отметим, что этот отчет не включает столбцы «Другие» или «Не определены» в качестве причин пересмотров решений. Разумеется, они должны быть сведены к минимуму. Хорошей практикой перед проектированием системы обработки кредитных заявок и подготовкой отчетов является проведение опроса кредитных аналитиков для составления полного перечня причин, которым они руководствовались, пересматривая решение скоринговой карты по тому или иному заявителю. Потом эти причины можно разделить на группы. Данная практика минимизирует шансы пересмотров решений, основанных на причинах, которые не могут быть включены в систему.

Пересмотры решений по нижнему порогу (одобренные заявители, находящиеся под отсечкой) приведены в столбцах «Автоматически «Отклоненные» и «Вручную «Одобренные». Это обозначает, что автоматическая система отклонила всех из них (скоринговые баллы ниже порога отсечения), но в итоге они все были одобрены вручную по определенной причине.

Количество пересмотров решений по верхнему порогу (отклоненные заявители, находящихся над отсечкой) варьирует:

- Строка «Банкротство» показывает, что все 125 заявителей были отклонены автоматической системой, свидетельствуя о том, что наличие банкротства – это автоматическое правило отклонения. Однако, пять заявителей в итоге были одобрены кредитным аналитиком, несмотря на то, что у них в прошлом было банкротство.

- Кредитные правила 1, 2 и 3 также являются автоматическими правилами отклонения.
- Строка «Локальное знание» показывает всех 102 заявителей, одобренных автоматической системой и затем последовательно отклоненных вручную.
- Строка «Негативные сведения» (например, обслуживание просрочки по данным банка и кредитного бюро) показывает интересную ситуацию. Все 200 заявителей изначально были автоматически одобрены системой, означая, что отдельная негативная характеристика кредитоспособности не является автоматическим правилом отклонения. Позднее, несмотря на то, что 185 заявителей из 200 в силу негативных сведений были отклонены вручную, кредитный аналитик решил одобрить 15 заявителей.

Поэтому для мониторинга пересмотренных решений лучше использовать отчет, показанный на рисунке 9.7, чем простое распределение по причинам пересмотров, которое иногда может скрывать определенные виды решений, такие, как одобрения вручную для строки «Банкротство» (в ходе пересмотров решений по верхнему порогу). Если отчет подготовлен, исходя из логики «если скоринговый балл > порог отсечения, то решение «одобрить» (чтобы определить одобренных клиентов), он не выявит пять банкротов в качестве пересмотренных решений и включит их как обычных одобренных. Можно поспорить с тем, что 15 заявителей с негативными характеристиками, которые были вручную рассмотрены и одобрены, также являются пересмотренными решениями. В таких случаях их нужно рассмотреть отдельно, чтобы оценить эффективность одобрения клиентов с негативными сведениями.

В условиях, когда процент пересмотров является высоким, нужно подготовить дополнительный отчет, составив в общих чертах профиль риска для пересмотренных заявок. Профиль риска можно получить различными способами, например:

- Сравнить характеристики, вошедшие и не вошедшие в скоринговую карту по пересмотренным и непересмотренным заявкам, особенно такие неблагоприятные показатели обслуживания кредита, как «Наибольшая просрочка».
- Оценить распределение пересмотренных заявок по скоринговым баллам за определенный период времени.

Это позволит получить качественную оценку профиля риска для пересмотренных заявок.

Отчеты эффективности кредитного портфеля

Отчеты эффективности кредитного портфеля часто называют «бэк-энд» отчетами. Они включают анализ обслуживания просрочек клиентами.

Отчет «Просрочки» (или «Обслуживание») Отчет «Просрочки» (или «Обслуживание») используется для определения эффективности кредитного портфеля. Он обычно состоит из таблиц, показывающих распределение процента «плохих» по скоринговым баллам для различных категорий «плохих» клиентов. Кроме того, этот отчет подготавливается по различным сегментам (например, регион, канал продаж, демографические характеристики и др.) для выявления особых зон с высоким и низким уровнем просрочки. Он также составляется по клиентам на основе дат открытия счетов, чтобы выявить отдельные когорты, представляющие более высокий риск.

Пример отчета «Просрочки» для портфеля кредитных карт показан на рисунке 9.8.

Этот отчет может быть составлен как в виде частот (показано на рисунке 9.8), так и в денежном выражении. Он применяется как для поведенческого, так и аппликационного скоринга. Рисунок 9.8 показывает:

- Обслуживание просрочек согласно скоринговой карте с порогом отсечения 200.
- Отдельный столбец для действующих клиентов (те, кто уже утилизировали свой доступный кредит). Определение «действующий» различно для кредитных организаций, но обычно основано на давности использования кредита.
- Процент «плохих» по каждому уровню просрочки, с использованием в качестве знаменателя общего количества действующих клиентов. Некоторые кредитные организации используют общее количество открытых счетов для определения портфеля револьверных кредитов. Для таких продуктов, как ипотека и займы, в качестве знаменателя должно использоваться общее количество открытых счетов или общее количество активно оплачиваемых счетов. Если скоринговая карта была разработана, используя такое сложное определение «плохого» клиента, как «1 x 90 дней или 2 x 60 дней или 3 x 30 дней», то необходимо добавить столбец с этим определением, чтобы можно было оценить текущую эффективность скоринговой карты. Отметим, что если разрабатывались поведенческие карты, то этот отчет может дать представление о приблизительном количестве «плохих» клиентов, которые могут быть полезны для разработки скоринговой карты.
- Тренд увеличения процента «плохих» по мере уменьшения скорингового балла, как и должно быть. Он показывает, что скоринговая карта ранжирует риск.

«Плохой» клиент может быть определен либо как «хотя бы раз плохой» клиент, либо как «текущий плохой» клиент. Некоторые кредитные организации составляют отчеты на основе этих двух определений.

В некоторых случаях кредитоспособность клиентов, находящихся непосредственно под отсечкой (нижний порог пересмотра решений) лучше кредитоспособности клиентов прямо над отсечкой. Это главным образом происходит из-за отбора «наиболее понравившихся» клиентов (клиенты под отсечкой были вручную выбраны после внимательного рассмотрения в качестве «лучших», в то время как клиенты над отсечкой были автоматически одобрены).

Данный отчет основывается на обслуживании просрочек – схожие отчеты могут быть подготовлены по оттоку, прибыли, доходу, сумме возмещения и т.д. в зависимости от задач, для решения которых была разработана скоринговая карта. Другая задача мониторинга эффективности – это оценка точности скоринговой карты. Для обеспечения надежного сравнения оценка кредитоспособности текущих заявителей должна основываться на тех же самых критериях, что и оценка кредитоспособности на обучающей выборке для скоринговой карты (с тем же самым определением «плохого» клиента, сегментацией, «окном созревания», исключительными случаями). Пример такого сравнения показан на рисунке 9.9.

Этот рисунок показывает текущий процент «плохих» по портфелю («Плохие %»), сравниваемый с ожидаемым процентом «плохих» в обучающей выборке («Ож Плохие %»). Есть несколько факторов, которые стоит отметить здесь:

- Поскольку этот отчет может быть составлен только по завершении «окна созревания», он имеет ограниченную оперативную ценность. В большинстве случаев для мониторинга кредитоспособности отдельных когорт в сравнении с ожидаемой кредитоспособностью используются

винтажный анализ и таблицы, как показано на рисунке 4.3. Если когорта имеет лучшую кредитоспособность, можно сделать вывод, что процент «плохих» на момент вызревания кредита будет меньше, чем ожидаемый, и наоборот. Эта информация полезна для управления клиентами и прогнозирования.

- Текущая кредитоспособность почти всегда отличается от ожидаемой в силу стратегий управления клиентами. Для риск-менеджера важно в любой момент времени точно спрогнозировать ожидаемую кредитоспособность.
- Текущий мониторинг процента «плохих» на основе дат открытия счетов и сравнение с ожидаемой кредитоспособностью используются для того, чтобы определить, на самом ли деле скоринговая карта работает. В большинстве случаев скоринговые карты ранжируют риск, но спрогнозированный процент «плохих» отличаются от ожидаемого. Когда скоринговая карта не ранжирует риск, ее, возможно, надо заменить. В тех случаях, когда наблюдается различие между ожидаемой и спрогнозированной кредитоспособностью, могут понадобиться такие меры, как изменение порогов отсека (для некоторых сегментов, если это допустимо), изменение правил кредитной политики или перевзвешивание баллов.

Винтажный (или Когортный) Анализ Винтажный (или когортный) анализ подразумевает получение процента «плохих» по различным когортам (счетам, открытым в течении определенного периода времени) на основе дат открытия счетов.

Как и отчет «Просрочки», данный отчет разрабатывается для различных категорий «плохого» клиента и по различным сегментам и группам. Отчет используется, чтобы:

- Выявить высокорисковые когорты (т.е. если счета, открытые в определенном месяце или квартале, имеют более высокий риск, чем остальные).
- Провести мониторинг роста процента «плохих» на протяжении времени – заметим, что Рисунок 4.3, который показывает рост процента «плохих» на протяжении времени, был получен, используя информацию из таблицы с результатами когортного анализа на Рисунке 4.2. Эта информация используется для сравнения кредитоспособности клиентов в новых когортах с долгосрочной кредитоспособностью, чтобы контролировать ожидаемое распределение и получать более точные прогнозы.

Пример винтажного анализа показан на рисунке 9.10. Он показывает качество кредитных счетов, открытых с января 2003 по март 2004, измеренных по одинаковым периодам времени (в данном случае, по кварталам). Отметим, что схожие отчеты можно разработать и для других показателей, таких, как отток, прибыль, банкротство, сумма возмещения и т.д. в зависимости от бизнес-задач, а также целей, поставленных перед скоринговой картой.

Данный отчет может быть получен для различных категорий «плохих» клиентов (просрочка 90+, просрочка 60+ и т.д.). Рисунок 9.10 показывает, что счета, открытые в марте и апреле 2003, несут больший риск, чем счета, открытые в других месяцах. Здесь можно обратиться к отчету «Стабильность системы» и другим отчетам за март и апрель 2003, чтобы выявить причины такого различия в кредитоспособности. Это различие может быть обусловлено такими факторами, как плохой отбор, изменения порога отсека, маркетинговая кампания,

нацеленная на высокорисковые группы, системные ошибки, в силу которых по невнимательности были одобрены высокорисковые клиенты, или чрезмерное количество пересмотров решений. Короче говоря, для поиска этих причин нужно использовать практически все отчеты, упомянутые в данной главе, а также информацию на основе накопленной бизнес-практики.

После того, как эти причины выяснены, можно определить, носят ли они разовый или повторяющийся характер. В последнем случае можно предпринять меры, которые позволят избежать одобрения высокорисковых клиентов.

Рисунок 9.8 Отчет «Обслуживание просрочек»

<i>Score</i>	<i>Accounts</i>	<i>Active</i>	<i>%</i>	<i>Current</i>	<i>%</i>	<i>1–29 days</i>	<i>%</i>	<i>30–59</i>	<i>%</i>	<i>60–89</i>	<i>%</i>	<i>90+</i>	<i>%</i>	<i>Writeoff</i>	<i>%</i>	<i>Bankrupt</i>	<i>%</i>
0–169	200	198	99%	127	64%	22	10.9%	15	7.50%	12	6.10%	9	4.50%	6	3.00%	8	4.00%
170–179	348	300	86%	207	69%	36	12.0%	20	6.80%	9	3.00%	9	3.10%	9	3.10%	9	3.00%
180–189	435	367	84%	264	72%	48	13.0%	15	4.00%	12	3.20%	10	2.80%	8	2.20%	10	2.80%
190–199	466	387	83%	286	74%	43	11.0%	21	5.50%	11	2.80%	8	1.94%	10	2.56%	9	2.20%
200–209	2,456	1,876	76%	1,482	79%	225	12.0%	43	2.30%	47	2.48%	18	0.96%	39	2.10%	22	1.16%
210–219	4,563	3,600	79%	2,952	82%	342	9.5%	93	2.58%	83	2.30%	23	0.65%	67	1.87%	40	1.10%
220–229	5,678	4,325	76%	3,676	85%	389	9.0%	93	2.16%	67	1.54%	14	0.32%	51	1.18%	35	0.80%
230–239	7,658	4,598	60%	4,046	88%	359	7.8%	87	1.90%	41	0.90%	18	0.40%	28	0.60%	18	0.40%
240–249	5,786	3,546	61%	3,333	94%	142	4.0%	35	1.00%	18	0.50%	—	0.00%	7	0.20%	11	0.30%
250+	4,987	2,176	44%	2,089	96%	44	2.0%	17	0.80%	9	0.40%	9	0.40%	4	0.20%	4	0.20%
Total	32,577	21,373	66%	18,463	86%	1,648	7.71%	441	2.06%	307	1.44%	118	0.55%	230	1.08%	165	0.77%

Рисунок 9.9 Отчет «Точность скоринговой карты»

<i>Score</i>	<i>Accounts</i>	<i>Active</i>	<i>%</i>	<i>Bad</i>	<i>%</i>	<i>Exp Bad %</i>
0–169	200	198	99%	35	18%	23.0%
170–179	348	300	86%	37	12%	18.0%
180–189	435	367	84%	40	11%	14.0%
190–199	466	387	83%	37	10%	10.0%
200–209	2,456	1,876	76%	126	7%	8.0%
210–219	4,563	3,600	79%	213	6%	5.0%
220–229	5,678	4,325	76%	166	4%	4.0%
230–239	7,658	4,598	60%	106	2%	2.0%
240–249	5,786	3,546	61%	35	1%	0.8%
250+	4,987	2,176	44%	26	1%	0.5%
Total	32,577	21,373	66%	821	4%	

Рисунок 9.10 Винтажный анализ

<i>Open Date</i>	<i>1 Qtr</i>	<i>2 Qtr</i>	<i>3 Qtr</i>	<i>4 Qtr</i>	<i>5 Qtr</i>
Jan-03	0.00%	0.44%	0.87%	1.40%	2.40%
Feb-03	0.00%	0.37%	0.88%	1.70%	2.30%
Mar-03	0.00%	0.42%	0.92%	1.86%	2.80%
Apr-03	0.00%	0.65%	1.20%	1.90%	
May-03	0.00%	0.10%	0.80%	1.20%	
Jun-03	0.00%	0.14%	0.79%	1.50%	
Jul-03	0.00%	0.23%	0.88%		
Aug-03	0.00%	0.16%	0.73%		
Sep-03	0.00%	0.13%	0.64%		
Oct-03	0.20%	0.54%			
Nov-03	0.00%	0.46%			
Dec-03	0.00%	0.38%			
Jan-04	0.00%				
Feb-04	0.00%				
Mar-04	0.00%				

Помимо всего прочего, как только обнаружено, что отдельные когорты несут больший риск, чем остальные, для управления ими могут быть приняты решения, скорректированные с учетом риска. Эти решения включают:

- Увеличение процентной ставки за продление кредита
- Приостановление автоматического увеличения кредитной линии для отмеченных высокорисковых когорт, кредитные лимиты увеличивают только по запросу и после рассмотрения
- Приостановление карточного кредитования для клиентов на период просрочки или по причине превышения предоставленного кредитного лимита
- Назначение более строгих мер взыскания долгов для этих когорт

Использование информации отчетов для принятия вышеперечисленных решений играет ключевую роль. Оно делает отчет средством принятия решения, а не просто процедурой подготовки необходимых бумаг и статистических показателей. Анализ, упомянутый ранее (для выяснения причин различия в кредитоспособности разных когорт), легче выполнить, если задокументированы определенные сведения. Они включают:

- Изменения скоринговых карт, порогов отсека, правил кредитной политики, продуктов и требований органов надзора
- Информация о крупных маркетинговых кампаниях

Их нужно собрать по каждому кредитному портфелю и задокументировать для упрощения диагностики и исправлений ошибок в будущем. Подобного рода документация является общепринятой для кредитного скоринга и ее часто относят к Журналу хронологии кредитного портфеля. Отметим, что он отличается от Журнала изменения данных, который включает все внутренние изменения в базах данных.

Отчет «Миграция просрочек» Ежемесячно большинство кредитных организаций проводит мониторинг перемещения клиентов из одной категории просрочки в другую. Пример отчета «Миграция просрочек» приведен на рисунке 9.11.

Отчет, показанный на рисунке 9.11, показывает миграцию клиентов из одного класса просрочки в другой, от «предыдущего месяца» к «текущему месяцу». Заметим, что «%» в столбце «Предыдущий месяц» показывает распределение по столбцу, «%» в столбце «Текущий месяц» показывает распределение по строке. Например, из всех клиентов, которые в прошлом месяце находились в категории «Просрочка 30-59 дней», 40% являются заплатившими в срок, 10% попадают в категорию «Просрочка 1-29 дней», 14.5% в категорию «Просрочка 30-59 дней» и т.д. Тот же самый отчет можно подготовить не по частотам, а по денежным суммам в каждой категории просрочки.

Данные отчеты могут помочь в составлении прогнозов. Можно модифицировать их, чтобы построить долгосрочные матрицы переходов (roll-rate). В некоторых случаях, когда проектирование модели 90-дневной просрочки или списания долга невозможно, могут разрабатываться модели, прогнозирующие меньшую просрочку, например, 60 дней. Затем можно объединить прогноз модели с информацией, полученной с помощью матрицы переходов, чтобы предсказать 90-дневную просрочку или списание долга. Отчет похож на анализ матрицы переходов, обсуждавшийся в Главе 4, и содержит аналогичные данные о «точке невозврата» в обслуживании просрочек. Приведенный выше отчет показывает, что 80% из тех, кто в прошлом месяце находились в категории «Просрочка 1-29 дней», погасили просрочку и стали клиентами, заплатившими в срок, в то время как лишь 3.8% из тех, кто в прошлом месяце были в категории «Просрочка 90-119 дней», погасили просрочку.

Отчет «Уровни перехода» (Roll Rate) Модифицированный вариант рисунка 9.11 применяется для мониторинга количества клиентов или неуплаченных сумм в каждой категории просрочки за определенный временной период. Пример такого отчета приведен на рисунке 9.12. В целях общего ознакомления он показывает только две категории просрочки. В реальности этот отчет подготавливают для всех категорий просрочки.

Данный отчет помогает вам понять формирование просроченной задолженности на протяжении времени, выразив его в виде количества клиентов, имеющих просрочки, или сумм просрочек. Относительный рост сумм просрочек дает представление об увеличивающихся балансах и позволяет ответить, возрастает ли доля потерь в случае дефолта (LGD) или нет.

Рисунок 9.11 Отчет «Миграция просрочек»

This Month																		
Previous Month	Previous Month		Current		1–29 days		30–59 days		60–89 days		90–119 days		120–179 days		180+		Bankrupt	
	#	%	#	%	#	%	#	%	#	%	#	%	#	%	#	%	#	%
Current	54,782	72%	52,591	96.0%	2,082	3.8%											210	0.2%
1–29 days	12,640	17%	10,112	80.0%	632	5.0%	1,871	14.8%									85	0.2%
30–59 days	3,254	4%	1,302	40.0%	325	10.0%	472	14.5%	1,139	35.0%							59	0.5%
60–89 days	2,271	3%	182	8.0%	204	9.0%	227	10.0%	227	10.0%	1,413	62.2%					39	0.8%
90–119 days	1,449	2%	55	3.8%	65	4.5%	80	5.5%	87	6.0%	72	5.0%	1,065	73.1%			42	2.1%
120–179	887	1.2%	16	1.8%	12	1.4%	19	2.1%	20	2.2%	27	3.0%	64	4.0%	550	78.0%	150	7.5%
180+	632	0.8%	1	0.2%	8	1.2%	11	1.8%	16	2.5%	13	2.0%	3	3.0%	371	60.3%	210	29.0%
Total	75,915		64,258	85%	3,329	4%	2,679	3.5%	1,488	2.0%	1,524	2.0%	1,132	1.5%	921	1.2%	585	0.8%

Рисунок 9.12 Отчет «Уровни переходов»

<i>Month</i>	<i>Total Receivable</i>		<i>Current</i>				<i>1–29 days</i>			
	<i>Accounts</i>	<i>Dollars</i>	<i>Accounts</i>	<i>%</i>	<i>Dollars</i>	<i>%</i>	<i>Accounts</i>	<i>%</i>	<i>Dollars</i>	<i>%</i>
May-04	80,895	\$256,987	71,188	88.0%	\$230,260	89.6%	\$6,472	8.0%	\$35,978	14.0%
Jun-04	81,229	\$277,125	71,075	87.5%	\$245,533	88.6%	\$6,986	8.6%	\$36,026	13.0%
Jul-04	86,985	\$289,541	75,851	87.2%	\$251,901	87.0%	\$6,872	7.9%	\$41,115	14.2%
Aug-04	89,524	\$298,654	77,796	86.9%	\$261,322	87.5%	\$7,162	8.0%	\$41,513	13.9%
Sep-04	92,458	\$311,897	80,069	86.6%	\$270,103	86.6%	\$7,027	7.6%	\$39,923	12.8%
Oct-04	97,114	\$318,694	84,004	86.5%	\$276,626	86.8%	\$7,478	7.7%	\$40,155	12.6%
Nov-04	99,365	\$322,145	85,851	86.4%	\$283,488	88.0%	\$7,651	7.7%	\$39,302	12.2%

Экспертиза

После того, как скоринговые карты построены и внедрены, экспертиза является хорошим способом выявить недоработки или недостатки, допущенные в процессе разработки и внедрения скоринговой карты и, наоборот, определить эффективные области применения. Ее назначение – сделать процесс разработки скоринговой карты более эффективным. Экспертизу следует выполнять вместе с лицами, принимающими участие в разработке скоринговой карты, как было описано в Главе 2. Некоторые ключевые вопросы, которые должны быть выяснены:

- Очистка данных требовалась? Если да, логика очистки данных должна быть сохранена.
Были ли интервью со специалистами кредитного отдела, отдела взыскания успешными с т.з. определения прогнозирующих переменных и изменений характеристик на протяжении времени? Эти интервью следует повторить в ходе разработки будущих проектов. В ситуации, когда анализ, проведенный в ходе разработки карты, подтвердил, что информация, полученная от интервьюируемых, возможно, была некорректной, ее необходимо передать обратно, например, специалистам кредитного отдела, чтобы они могли скорректировать свои предположения. Например, специалисты, возможно, подумали, что молодые люди, живущие в мегаполисах, были высокорисковыми, потому что у них высокая арендная плата, они часто меняют работу и место жительства. Однако анализ показал, что они – низкорисковые клиенты, потому что это рассматривается как «нормальное» поведение для мегаполисов. В большинстве случаев для специалистов, ответственных за оценку рисков и управление кредитным портфелем, могут быть представлены результаты интерактивной группировки, поскольку данная информация интуитивно понятна, наглядна и легка для интерпретации. Она способствует оптимизации системы риск-менеджмента за счет лучшего понимания факторов риска, влияющих на ваш кредитный портфель.
- Были ли особые факторы, которые пришлось учитывать при формировании кредитного портфеля, такие, как сезонность или периоды аномальной активности? Если да, то подобного рода информацию следует задокументировать для разработки скоринговой карты и управления кредитным портфелем в будущем.
- Были ли какие-либо непредвиденные события, когда проект был завершен? Можно предпринять какие-то действия в будущем, чтобы избежать подобных случаев? Эти события включают ожидание данных, невозможность интерпретации результатов или данных, недоработки других отделов при выполнении своих обязанностей из-за плохого планирования, не позволяющие внедрить скоринговую карту, невозможность согласовать порог отсечения или другую стратегию кредитования и т.д.
- Применялись ли «манипуляции данными» или преобразования данных, чтобы упростить разработку скоринговой карты?
- Были ли какие-либо элементы данных, которые вызвали проблемы?
- Наблюдались ли какие-либо неожиданности и непредвиденные последствия после того, как была разработана и внедрена стратегия кредитования? Это было вызвано проблемами коммуникации между отделами или ошибкой прогноза? Если да, снова необходимо выполнить анализ типа «что-если».
- Были ли подготовлены отчеты перед разработкой скоринговой карты?

Большая часть вопросов – и ответы на них – освещены в данной книге. Несмотря на это часто совершаются ошибки. Однако важно учиться на них и понимать, что каждая последующая разработка скоринговой карты будет более эффективной, лучше организованной и – прежде всего – более интеллектуальной.