

ВНЕДРЕНИЕ КРЕДИТНОГО СКОРИНГА: МЕТОДОЛОГИЧЕСКИЕ И ПРАКТИЧЕСКИЕ АСПЕКТЫ

Кредитный скоринг – это метод разделения групп потенциальных клиентов-заемщиков в условиях доступности информации не о параметрах, разделяющих эти группы, а только о некоторых вторичных переменных. Идея разделения групп по вторичным признакам впервые была предложена в 1936 г. Р.А. Фишером, который представил метод определения разновидностей ириса по результатам измерений размеров частей растений, а спустя несколько лет идея использовать аналогичный подход для определения заемщиков с высокой и низкой вероятностью дефолта была высказана Д. Дюрандом.

Типы кредитного скоринга

Назначение кредитного скоринга – автоматизированное принятие решений по выдаче кредитов. В зависимости от типа используемых входных данных о потенциальном заемщике условно выделяют 4 типа скоринга:

1. Application-скоринг – оценка кредитоспособности заемщиков, обратившихся за получением кредита, по указанным в заявке на кредит данным.
2. Collection-скоринг – определение приоритетных направлений работы в отношении заемщиков с кредитным счетом, состояние которого классифицировано как «неудовлетворительное». Данный тип скоринга позволяет вести планомерную работу с просроченной задолженностью до момента ее передачи в коллекторское агентство.
3. Behavioral-скоринг (поведенческий скоринг) – динамическая оценка состояния кредитоспособности существующего клиента-заемщика на основе данных о транзакциях по его счетам, таким как график погашения задолженности, обороты по текущим счетам, наличие новых кредитных заявок. Позволяет определять лимиты кредитования, маркетинговые ходы и меры, которые могут применяться по отношению клиента.
4. Fraud-скоринг – оценка вероятности мошенничества потенциального заемщика.

При построении модели скоринга могут использоваться данные различных видов, поэтому также существует следующая классификация:

1. Априорный скоринг – построение по статистическим данным (макроэкономические показатели, государственная статистика,

результаты переписи и т. п.), которые используются для оценки параметров модели заемщика, применяемой в свою очередь для определения его кредитоспособности.

2. Апостериорный скоринг – построение по историческим данным о клиентах кредитной организации. Такого рода данные обычно представляют собой таблицу, в которой для каждого из клиентов, получившего кредит, указаны данные его заявки, а также результат (погашение долга, досрочное погашение, дефолт).

Если говорить о методах, применяемых в различных типах скоринга, то особого внимания заслуживают методы, применяемые в Application- и Behavioral-скоринге.

Методы Application-скоринга

В настоящее время в application-скоринге используются следующие методы, причем они могут применяться как отдельно друг от друга, так и в различных комбинациях:

- 1) методы статистики (дискриминантный анализ, линейная регрессия, логистическая регрессия, деревья классификации);
- 2) методы исследования операций (линейное программирование, нелинейная оптимизация);
- 3) методы искусственного интеллекта (нейронные сети, экспертные системы, генетические алгоритмы, методы ближайших соседей, байесовские сети, логико-вероятностные методы).

Использование статистических методов сводится к построению правила классификации, основанного на линейной скоринговой функции. При этом используются различные подходы.

Байесовский подход. Суть этого метода состоит в том, что по формуле полной вероятности вычисляется вероятность того, что клиент даст определенные заданные ответы, а затем по теореме Байеса определяется вероятность того, что клиент с данными ответами будет «хорошим» или «плохим». После этого рассчитываются средние потери на одного клиента и определяется условие их минимальности. Далее предполагается нормальность распределений вероятностей, что клиенты дадут заданные ответы, с общей ковариационной матрицей, и строится линейное правило, которое фактически означает, что каждому набору ответов присваивается «счёт», содержащий достаточно информации для различения класса клиента. Нахождение правила классификации сводится к поиску оптимального порога из задачи оптимизации (поиск минимума суммы потерь от классификации «хорошего» клиента как «плохого» и потерь в случае дефолта при ошибочном определении «плохого» клиента как «хорошего»).

Множественная регрессия. В простейшем варианте определяется зависимая переменная Y , принимающая значения 1 в случае «хорошего»

клиента и 0 в случае «плохого», и затем используется метод наименьших квадратов для нахождения весов, обеспечивающих наилучшее предсказание значения Y . При использовании линейной регрессии фактически делается попытка связать линейной функцией вероятность дефолта со значениями ответов на вопросы линейной функцией. Наблюдается очевидное несоответствие, т. к. левая часть функции представляет собой вероятность и должна изменяться от 0 до 1, а правая может принимать любые значения. Для преодоления этой трудности значение вероятности заменяется некоторой функцией от него так, чтобы функция была монотонной и принимала все вещественные значения; например, могут использоваться логистическая регрессия (логарифм шансов дефолта) и пробит-анализ (функция стандартного нормального распределения).

Дискриминантный анализ. При использовании линейных скоринговых функций оценивается вероятность принадлежности клиента тому или иному классу. Далее выбирается класс, которому соответствует больший счёт. В случае двух классов это даёт тот же результат, что и линейная регрессия.

Линейное программирование – ещё один способ получить линейную скоринговую функцию. Здесь производится минимизация по ошибкам поиска «плохих» и «хороших» клиентов. Преимущество метода линейного программирования – возможность включить в программу дополнительные ограничения. Например, если кредитный продукт направлен на людей молодого возраста, можно наложить дополнительное ограничение на соответствующий вес, так что функция будет отдавать предпочтение людям этого возраста.

Генетические алгоритмы – метод оптимизации, не накладывающий стандартных ограничений на целевую функцию (гладкость, выпуклость и т. п.). Генетический алгоритм направлен на генерацию начального множества скоринговых функций с последующим применением к функциям из этого набора операций «скрещивания» и «мутации», а также отбрасыванием наименее пригодных функций.

Деревья классификации (рекурсивные алгоритмы разбиения), в отличие от предыдущих методов, не предназначены для построения линейной скоринговой функции, они последовательно разделяют клиентов на группы по одной из переменных так, чтобы эти группы максимально возможно отличались по величине кредитного риска. Данный процесс продолжается до момента, пока оставшиеся группы не становятся настолько малы, что следующее разбиение не приведёт к статистически значимому различию в уровне риска. Далее каждому «листу» дерева приписывается определённая категория клиентов, ему может соответствовать также не определённая категория, а своя линейная регрессионная модель, что позволяет сна-

чала грубо разделять клиентов на основные группы, а потом для каждой группы использовать регрессионную модель.

Метод ближайших соседей – пример подхода «ленивого обучения», когда обучение сводится к добавлению новых случаев в базу данных. Выбирается метрика на пространстве данных, которая используется для нахождения ближайших исторических анкетных данных к данным нового клиента. Клиент классифицируется в зависимости от того, к какому классу относится большинство его соседей по выбранной метрике. Преимущество метода – простота обучения модели: при поступлении новых данных их добавляют к существующим, а по прошествии времени старые данные удаляют для учета происходящих с течением времени изменений. Недостаток – необходимость полного просмотра базы данных для классификации нового клиента.

Нейронные сети могут рассматриваться в качестве метода нелинейной регрессии. Однако они чаще применяются для скоринга юридических лиц, чем для скоринга частных лиц.

Логико-вероятностный подход основан на методах математической логики и теории вероятностей. Данный подход может рассматриваться как специальный вид нелинейной регрессии, в качестве параметров модели выступают вероятности наступления дефолта, связанного с тем или иным ответом на вопрос анкеты.

Методы Behavioral-скоринга

Для поведенческого скоринга выделяют:

1. Методы, сходные с методами кредитного application-скоринга, – классификация клиентов в зависимости от их предыдущего поведения (главное отличие заключается в том, что в анкету, помимо данных о заёмщике, добавлено описание его поведения).
2. Статистические модели, которые не ставят цель объяснить природу того или иного поведения заёмщиков, а лишь аппроксимируют наблюдаемое поведение. Здесь преобладают модели, основанные на марковских цепях, которые могут быть использованы как для моделирования поведения одного заёмщика, так и для оценки потерь в целом кредитного портфеля.
3. Структурные модели, описывающие динамику кредитоспособности и репутации заёмщика. В данных моделях используются доступные статистические данные для оценки параметров самой модели.

В первом случае в системах поведенческого скоринга изучаются данные о недавнем поведении заёмщика для оценки вероятности его дефолта, при этом применяется методология, сходная с используемой в application-скоринге. Подготавливается выборка данных по уже имею-

щимся заёмщикам, включая информацию о транзакциях по счетам за определенный период времени, который разбивается на период наблюдения и контрольный период, каждый примерно по году. По периоду наблюдения исследуются различные численные показатели поведения заёмщика: обороты по счетам, тренды в платежах и балансе и т. п. Затем заёмщик классифицируется в зависимости от состояния на конец контрольного периода как «плохой» или «хороший». Далее сходными с применяемыми в application-скоринге методами строится модель, классифицирующая заёмщиков на «хороших» и «плохих».

Модели, основанные на марковских цепях (марковские модели), базируются на двух предположениях: набор состояний действительно описывает всевозможные состояния, в которых может находиться заёмщик, и динамика изменения состояний действительно удовлетворяет марковскому свойству. В качестве состояний обычно выбираются уровни задержки платежей (к примеру, количество месяцев задержки). Переходные вероятности определяются по историческим данным, а для построения более точной модели в состояние могут включаться дополнительные данные (состояние счёта, размер долга и т. д.).

Одна из наиболее применяемых *структурных моделей* кредитного риска частных лиц построена с использованием теории оценки опционов и понятия репутации заёмщика. Для этого делается несколько предположений. Первое состоит в том, что существует ненаблюдаемая случайная величина, называемая кредитоспособностью заёмщика, заключающая в себе всю информацию, необходимую для оценки кредитного риска. Второе говорит, что, хотя нет возможности наблюдать эту величину непосредственно, рынок использует доступную информацию о заёмщике, такую как успешный возврат предыдущих кредитов, финансовое состояние заёмщика, другие факторы, для того, чтобы оценить кредитоспособность или построить некий суррогат этой величины. Третье утверждает, что вероятность того, что данному клиенту будет выдан кредит, является строго возрастающей функцией. Четвертое заключается в том, что существуют механизмы передачи информации о дефолте произвольной кредитной организации, т. е. если у заёмщика происходит дефолт, об этом узнают все кредитные организации и у него не будет доступа к кредиту в ближайшее время («потеря репутации»). Пятое предположение: доступ к кредиту представляет ценность для заёмщика, которая, если она выражена в денежном эквиваленте, называется репутацией заёмщика (причем это строго возрастающая функция). Последнее предположение заключается в том, что, если заёмщик не может внести очередной платёж в срок, он возьмёт дополнительный кредит на соответствующую сумму. Сделав указанные предположения, можно использовать для оценки кредитного риска рассуждения, употребляемые в теории оценки опционов Мертона.

Подготовка исходных данных

Для разработки алгоритмов кредитного скоринга необходима обучающая выборка, качество которой (репрезентативность) влияет на точность оценок параметров модели скоринга и эффективность (предиктивную мощность) скорингового алгоритма. Репрезентативность определяется тем, насколько полно в ней присутствуют положительные и отрицательные прецеденты. Один и тот же элемент выборки может быть признан и положительным, и отрицательным для разных постановок задач или вообще не подходить для включения в обучающую выборку. Информация же по отказам не может быть использована в качестве обучающей выборки, т. к. не содержит нужных сведений. Возникает методическая проблема: если в скоринговых расчетах опираться только на фактические данные по выданным кредитам, то предиктивные оценки кредитоспособности новых соискателей будут содержать некоторую систематическую ошибку. Степень этой ошибки можно оценить и снизить, подвергая скорингу данные по отклоненным заявкам и включая их в повторный расчет модели, разделив предварительно данные по отклоненным кредитам на положительные и отрицательные прецеденты, как если бы они были таковыми на самом деле.

Информация на выходе

В зависимости от используемой модели система скоринга может выдавать на выходе следующие данные:

1. Класс клиента. Разделение клиентов на «хороших», которым кредит выдается, и «плохих», которым отказывают в выдаче. Также могут быть определены несколько промежуточных классов, для которых, к примеру, либо кредит выдается, но на других условиях, либо кредитоспособность оценивается вручную, с использованием дополнительной информации.
2. Апостериорное распределение класса клиента. Каждому классу устанавливается вероятность, с которой данный клиент принадлежит этому классу. Выбирается класс с наибольшей вероятностью либо усредняются некоторые показатели по каждому классу. Преимущество данного варианта состоит в том, что в распределении содержится существенно больше информации.
3. Вероятность дефолта. Для принятия решения остаётся только сравнить эту вероятность с пороговым значением допустимой вероятности дефолта. Последняя определяется так, чтобы с учетом этой вероятности и процентной ставки банк в среднем не терпел убытков в случае выдачи кредита.

4. Счёт (score) – количественная оценка кредитоспособности потенциального заёмщика (чем больше счёт, тем она выше). Счёт пропорционален вероятности или шансам успешного возврата кредита, поэтому либо по счёту определяется вероятность дефолта и на этом основании принимается решение, либо, в зависимости от того, в какой промежуток попал счёт, определяется класс клиента и на основании этого – условия выдачи кредита.

Выбор алгоритма

Сравнение алгоритмов скоринга может осуществляться по разным критериям. Один из самых известных и применяемых на практике критериев сравнения состоит в построении графика. Предварительно необходимо упорядочить выборку прецедентов по возрастанию оценок вероятности положительного исхода, полученных с помощью скорингового алгоритма. На оси X откладываются значения долей этой упорядоченной выборки (слева направо по возрастанию вероятности положительного исхода), а на оси Y – доли фактически отрицательных прецедентов в подвыборке, отвечающей координате X. Полученный график иллюстрирует, какой процент Y действительных отрицательных прецедентов содержится в первых X процентах упорядоченной выборки. Это показывает, что чем выше линия графика проходит над осью X, тем более точен алгоритм скоринга. Для сравнения алгоритмов скоринга по предиктивной мощности достаточно сравнить соответствующие им графики. Если график какого-либо метода проходит над графиками остальных методов, то он является самым эффективным. Может оказаться, что абсолютного преобладания ни у одного алгоритма нет. Тогда для консервативной кредитной политики выбирается график, возвышающийся для малых значений оси X, а для более рискованной – преобладающий на продолжении оси.

Внедрение системы кредитного скоринга

Если рассматривать распределение функций при кредитовании с точки зрения банковской инфраструктуры, то выделяют несколько уровней. *Фронт-офис* – непосредственное общение с клиентом и продажа кредитных продуктов; первичная оценка данных, предоставленных заемщиком; сбор и анализ пакета документов заемщика; подготовка договоров. *Мидл-офис* – расчет сумм резервирования; скоринг; проведение кредитной экспертизы; определение перспектив погашения. *Бэк-офис* – ведение кредитного дела; оформление выдачи кредита; открытие счетов и ведение договоров; проведение бухгалтерских операций; оценка качества кредитного портфеля.

Это распределение довольно условно, т. к. функции мидл-офиса на практике часто распределяют между фронт- и бэк-офисом. Рассматривая

место полноценной скоринговой системы с точки зрения программного обеспечения, можно сказать, что она является решением мидл-офиса, расположенным между фронт-офисными и бэк-офисными решениями. После окончательной разработки алгоритма кредитного скоринга необходимо построение фронтальной системы, которая будет использоваться для поддержки принятия решений кредитным офицером в процессе рассмотрения заявок клиентов или для их самодиагностики с помощью web-сервиса, представленного на сайте банка в Интернете. Фронтальная система по информационным признакам клиента определяет его скоринг-балл и сравнивает его с баллом отсека.

Преимущества и недостатки внедрения систем скоринга

При внедрении системы кредитного скоринга в инфраструктуру банка возникает ряд *проблем*.

Во-первых, это отсутствие понимания всей сложности скоринговой системы, т.к. во многих банках до сих пор думают, что анализ данных вполне можно проводить при помощи стандартных средств или простейших разработок собственных IT-отделов.

Во-вторых, это отсутствие необходимых данных. Причиной этого может служить либо то, что банки не успели за время работы собрать достаточное для анализа количество информации, либо просто практически не собирают данные, необходимые для успешной работы. Без общей системы сбора данных очень сложно проследить долгосрочные тенденции, объяснить падение продаж, вырабатывать оптимальные маркетинговые направления и стратегию.

В-третьих, это некорректность, объем и разрозненность данных, связанные с тем, что зачастую данные на разных участках банковской инфраструктуры собираются в совершенно разных форматах. Наиболее оптимальный, но и дорогостоящий вариант – внедрение единого хранилища данных, в котором бы происходил сбор максимально полной информации о клиентах и деятельности банка. В случаях, когда данные собирались уже много лет и разрослись до значительных объемов, это может стать преградой для внедрения системы скоринга, но грамотная интеграция позволяет минимизировать эту проблему.

В-четвертых, серьезной проблемой может стать неполное представление данных в базе в силу непродуманной технологии сбора или из-за ее нарушения. Анализ таких данных может быть небезопасен, т. к. на основе неполных или неверных данных высока вероятность принимать неверные решения.

Если говорить о предиктивной мощности скорингового алгоритма, то с течением времени в силу объективных причин она снижается. Возникает необходимость повторного расчета и замены скоринговых алгоритмов во фронтальных приложениях. Момент перерасчета выбирается на

основе статистических критериев и может происходить с любой периодичностью по мере пополнения базы данных.

Однако, несмотря на наличие столь весомых сложностей, очевидны *преимущества* использования полноценной системы скоринга по сравнению с типовым подходом к оценке потенциальных и существующих заемщиков:

1. Первичная обработка кредитной заявки основывается не на экспертных знаниях кредитного специалиста, а на объективной информации из различных источников.
2. Процесс оценки идентичных заявок: идентичные заявки проходят идентичную процедуру оценки, а не зависят от конкретного кредитного специалиста и субъективных факторов.
3. Процесс внедрения типовой скоринговой методики требует длительного обучения и тренировки каждого кредитного специалиста, наработки опыта и интуиции. Полноценная скоринговая система, в свою очередь, не требует длительного обучения сотрудников, при внедрении необходим контроль со стороны кредитных специалистов высшего звена.
4. Возможность ошибок, злоупотреблений и мошенничества в силу человеческого фактора и злоупотреблений сотрудников в полноценных скоринговых системах практически исключается. Злоупотребления и мошенничества возможны только на уровне высшего звена кредитных специалистов, однако в силу непрерывного контроля их вероятность заметно снижается.
5. Если говорить о гибкости, то при внедрении нового кредитного продукта при типовом скоринге необходима разработка новых инструкций и обучение персонала, а это длительный и мало поддающийся контролю процесс. При внедрении нового кредитного продукта в полноценную систему необходимо создание новых скоринговых моделей (или внесение изменений в уже имеющиеся). Этот процесс полностью контролируем, качество вновь созданных моделей может быть проверено без запуска в работу, а дополнительного обучения персонала практически не требуется.