

Скоринговая модель кредитных историй

Виктор ПЛЁНКИН



Начальник Управления «Кредитный регистр» Национального банка

Мантас ТАРТЕНАС



Руководитель отдела эконометрического моделирования Creditinfo Lietuva (Литва)

Большинство кредитных бюро начинают свою деятельность с представления кредиторам не-обработанных данных в форме кредитных отчетов. Кредитные отчеты позволяют обоснованно подойти к принятию решения о выдаче кредита, но требуют времени и навыков для анализа и интерпретации данных. Кроме того, их сложно интегрировать в автоматизированные системы

оценки кредитных рисков. Поэтому в целях более эффективного использования данных предлагаются такие аналитические продукты, как кредитный скоринг.

Кредитный скоринг представляет собой систему присвоения баллов кредитополучателю на основании его способности и потенциала погасить долг. Баллы рассчитываются на основании имеющейся кредитной информации о прошлых периодах с использованием статистической модели или математического алгоритма.

Основным преимуществом кредитного скоринга является способность установить количественно измеримую степень риска, что способствует:

- снижению издержек за счет автоматизации принятия решения о выдаче кредита;
- сокращению времени обработки заявлений и принятия решения о выдаче или отказе в кредите;
- снижению влияния человеческого фактора при принятии кредитного решения.

Готовность кредитного бюро к предложению аналитических продуктов определяется «зрелостью» баз данных. Необходимо время, чтобы данные, содержащиеся в базах данных, обладали способностью прогнозировать будущие результаты.

В Республике Беларусь система сбора и распространения информации об исполнении кредитных обязательств функционирует с 2007 г., передача данных является обязательной для всех банков. С 1 января 2009 г. собирается информация по всем кредитным сделкам без ограничения по сумме договора. С 21 августа 2009 г. собирается информация по всем договорам обеспечения. Обязательность передачи данных способствовала обеспечению полноты и достоверности сведений. Более половины взрослого населения страны (~ 60%) имеют кредитную историю. Вероятность

нахождения кредитной истории (hit rate) составляет более 80%. Таким образом, количественные и качественные характеристики Кредитного регистра Национального банка Республики Беларусь (далее – Кредитный регистр) позволяют разрабатывать более сложные продукты, такие как кредитный скоринг. Немаловажным является и то, что база данных Кредитного регистра построена на четких, однозначных идентификаторах.

Если сравнивать ситуацию в Республике Беларусь с ситуацией в других странах Таможенного союза, то можно отметить следующее. По количественным показателям кредитной информации (охват взрослого населения и вероятность нахождения кредитной истории) наша страна опережает партнеров по Таможенному союзу, но по количеству и качеству предоставляемых услуг отстает. В Российской Федерации и Республике Казахстан созданы системы кредитных бюро с участием ведущих мировых операторов кредитных историй. Используя опыт и ноу-хау своих зарубежных партнеров, кредитные бюро России и Казахстана предоставляют услуги, основанные на аналитической обработке данных. В Беларуси до сих пор разработка услуг велась собственными силами, что, к сожалению, обусловило отсутствие доступа к мировым технологиям и ноу-хау в этой области.

В отчете группы экспертов Всемирного банка по Программе оценки финансового сектора от 10 июля 2014 г. отмечается эффективность нынешней системы кредитной отчетности и даются рекомендации по ее дальнейшему развитию за счет расширения охвата кредитной информации и создания частных кредитных бюро, преимуществом которых являлось бы предоставление дополнительных услуг, таких как скоринг. Начиная с 2014 г. в рамках исследования «Ведение бизне-

са» учитывается предоставление скоринговых оценок кредитными бюро (кредитными регистрами) при расчете индекса глубины кредитной информации.

В соответствии с протоколом заседания Правления Национального банка Республики Беларусь от 31 января 2014 г. № 5/1/П Управлению «Кредитный регистр» поручено обеспечить предоставление скоринговой оценки кредитной истории (кредитного скоринга).

Учитывая международный опыт и собственные возможности, необходимо признать, что оптимальным для внедрения кредитного скоринга на текущем этапе является привлечение международной компании, оказывающей подобные услуги.

В связи с этим был проведен открытый конкурс на разработку скоринговой модели. Основным требованием к исполнителям было наличие опыта построения скоринговых моделей по данным кредитных бюро. Участвовали региональные представительства всех международных операторов кредитных историй. Наиболее выгодным предложением признано предложение литовской компании Creditinfo Lietuva. Кроме того, литовская компания гарантировала высокое качество скоринговой модели (коэффициент Gini¹ > 60%).

Много времени заняла подготовка данных. Как оказалось, высокое требование к качеству модели – это в первую очередь высокое требование к качеству данных. Потраченные силы стоили полученного результата, и сейчас Национальный банк представляет принципиально новый продукт на основе кредитных историй – кредитный скоринг.

Компанией Creditinfo Lietuva разработана универсальная скоринговая модель (generic scorecard) оценки кредитоспособности физических лиц. В качестве «плохого» было определено следующее событие: значимая по величине просроченная задолженность (три базовых величины) сроком более 90 дней в течение 12 месяцев со дня подписания договора.

«Хорошими» признаются следующие события:

- максимальное количество дней просроченной задолженности ≤ 30 дней в течение 12 месяцев со дня подписания договора;
- количество дней просроченной задолженности от 30 до 90, сумма просроченной задолженности < 3 базовых величин;
- количество дней просроченной задолженности ≥ 90, сумма просроченной задолженности < 1 базовой величины.

Все остальные события признаются «неопределенными».

Для построения модели был выбран интервал с 01.01.2013 по 30.09.2013. Соотношение «плохих» клиентов (bad rate) ко всем клиентам в группе в этот период было стабильным и репрезентативным, что видно на *рисунке 1*.

При разработке модели были исключены из анализа «неопределенные» события. Их оказалось ~ 2% от общего количества контрактов. Вся выборка была случайным образом разделена на две части. Для построения скоринговой модели использовалось 80% данных, для тестирования – оставшиеся 20%.

В первоначальной выборке по каждому договору присутствовало более 1200 характеристик. После проведения первичного анализа осталось 375 характеристик, демонстрирующих ста-

бильную и/или логическую тенденцию к bad rate. После группирования взаимосвязанных характеристик и выбора наиболее сильных из каждого кластера количество характеристик уменьшилось до 62.

После анализа и отсева «лишних» характеристик построена статистическая модель на основе множественной логарифмической регрессии, которая включает 25 переменных.

Переменные модели можно объединить в несколько блоков:

- демографическая информация (возраст субъекта кредитной истории, место жительства и т. д.);
 - информация о запросах (количество запросов, пользователь, время с момента последнего запроса и т. д.);
 - информация о договорах (количество договоров, типы договоров, время с момента последнего договора и т. д.);
 - информация о просроченной задолженности (количество дней просроченной задолженности, время с момента последней просрочки и т. д.);
 - исполнение обязательств (типы договоров, относительная непоплаченная сумма договоров, продолжительность оставшихся обязательств и т. д.).
- Каждый блок имеет разный вес в модели (*рисунок 2*).

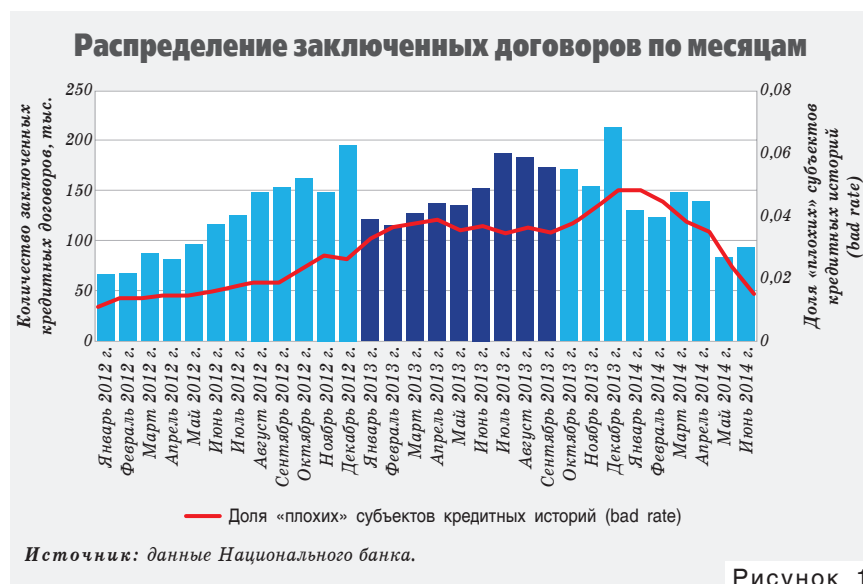
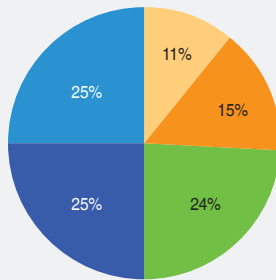


Рисунок 1

¹ Коэффициент Джини (Gini coefficient) – мера оценки качества скоринговой модели. Позволяет судить о дискриминирующей способности модели, позволяющей отличать «хороших» и «плохих» кредитополучателей. Используется для сравнения качества различных скоринговых моделей.

Вес переменных в модели



- Демографическая информация
- Информация о запросах
- Информация о договорах
- Информация о просрочке
- Исполнение обязательств

Источник: данные Национального банка.

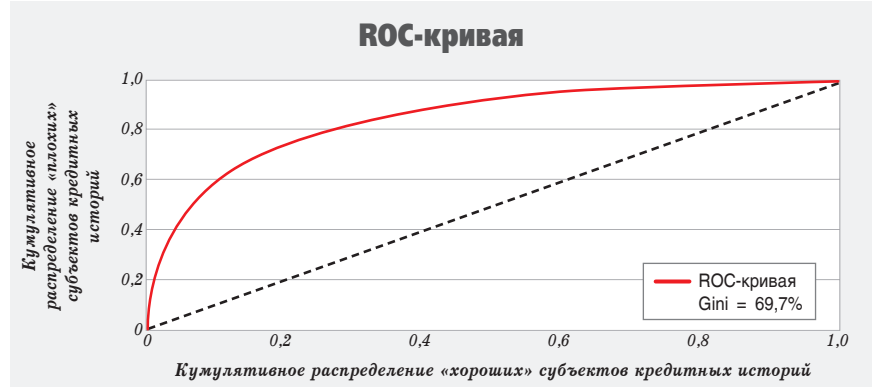
Рисунок 2

Коэффициент Gini для экспериментальной выборки составляет 70,1%, для тестовой выборки – 69,7%. Значения Gini, полученные на экспериментальной выборке и тестовой, близки. Это является показателем того, что параметры модели рассчитаны правильно и модель является эффективной для разных ситуаций.

Индекс Gini можно рассчитать, сравнивая кумулятивные распределения GOOD и BAD из графика ROC (от англ. Receiver Operational Characteristic) (рисунок 3). Каждая точка кривой соответствует определенному баллу и показывает разницу между кумулятивными распределениями GOOD и BAD. Чем эта кривая более выпуклая, тем Gini больше. Если бы кривая совпала с пунктирной линией, Gini модели был бы 0%.

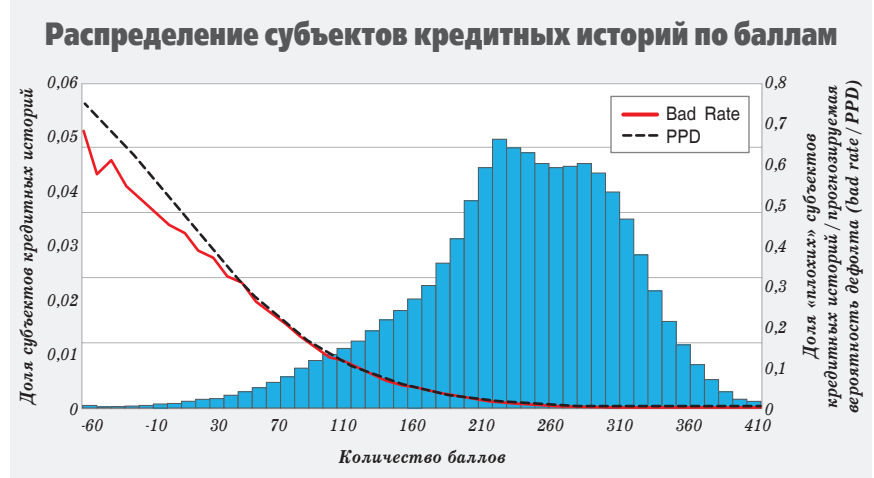
Анализ распределения населения по баллам был проведен с использованием интервала в 10 пунктов, 99,5% всех договоров попали в интервал [0;400], представляющий собой диапазон возможных баллов, которые будут предоставлены пользователям (рисунок 4).

Для лучшей демонстрации того, насколько хорошо модель отделяет «хороших» от «плохих», на одном графике построены кривые GOOD и BAD. Разница между кривыми определяет эффективность модели (чем больше разница, тем выше эффективность). Статистический тест Колмогорова – Смирнова



Источник: данные Национального банка.

Рисунок 3



Источник: данные Национального банка.

Рисунок 4



Источник: данные Национального банка.

Рисунок 5

рассчитывает максимальную разницу, существующую между этими двумя кривыми кумулятивных распределений, и может использоваться в качестве индикатора эффективности модели. Результат

проверки модели по тесту Колмогорова – Смирнова составляет 0,55, он достижим при количестве баллов 215 (рисунок 5). Это означает, что для данного значения модель дает наилучший результат.

Это может означать «одобрить», если количество баллов > 215, и «отклонить», если количество баллов ≤ 215. На практике кредиторы для определения граничного значения принимают во внимание много других характеристик (например, рентабельность и т. п.).

Еще один способ продемонстрировать различающую способность модели – на одном графике построить кривые маргинального распределения вероятностей GOOD и BAD (рисунок 6). Степень взаимного наложения маргинальных распределений GOOD и BAD определяет коэффициент Gini. Если взаимного наложения не происходит, то коэффициент Gini = 100%, если произойдет полное наложение распределений, то коэффициент Gini = 0%.

Рисунки 1–4 часто используются для визуализации эффективности модели. На практике график ROC и график кумулятивных распределений (вместе с анализом рентабельности) часто используются для определения точки отсечения, когда запрос клиента на получение кредита будет автоматически приниматься, отклоняться или же нуждаться в оценке эксперта.

Кроме скоринговых баллов, будут предоставляться классы рейтинга по таблице. Класс рейтинга меняется через каждые 25 баллов. Это означает, что log-odds² увеличивается постоянной величиной. Log-odds соотношений между классами рейтинга C1 и C2 является таким же, как и между классами рейтинга C2 и C3. До определения классов от А до Е определяется класс F по условию, указанному в таблице.

Классы легче сравнивать, они более понятны как для пользователей кредитных историй, так и для субъектов. Кроме того, они полезны для мониторинга, нет смысла отслеживать незначительные изменения в баллах. Распределение субъектов кредитных историй по классам рейтинга представлено на рисунке 7. Около половины популяции попадает в В класс, а 15% в А класс.

Оценку кредитной истории конкретного лица в баллах можно трансформировать в PPD (от англ. Predicted Probability of Default,

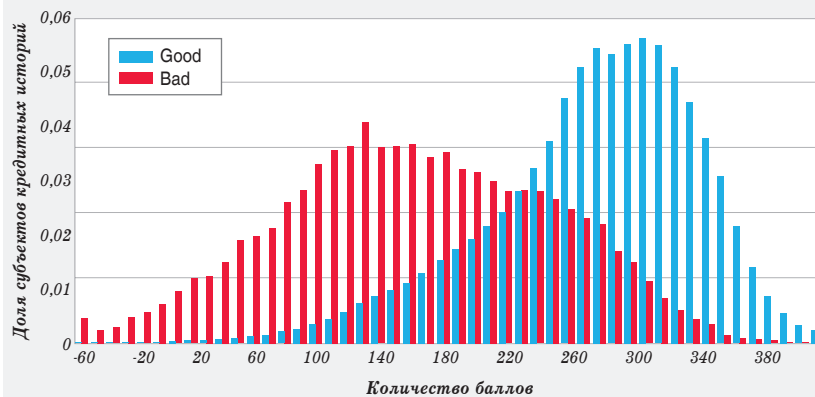
Таблица

Классы рейтингов

Class	>Score	≤Score	PPD lo	PPD Hi
A1	375	-	0,00%	0,25%
A2	350	375	0,25%	0,39%
A3	325	350	0,39%	0,60%
B1	300	325	0,60%	0,92%
B2	275	300	0,92%	1,41%
B3	250	275	1,41%	2,16%
C1	225	250	2,16%	3,30%
C2	200	225	3,30%	4,99%
C3	175	200	4,99%	7,50%
D1	150	175	7,50%	11,11%
D2	125	150	11,11%	16,16%
D3	100	125	16,16%	22,92%
E1	75	100	22,92%	31,44%
E2	50	75	31,44%	41,42%
E3	-	50	41,42%	100%
F	Текущая просроченная задолженность > 130 000 BYR и > 90 дней			

Источник: данные Национального банка.

Распределение «хороших» и «плохих» субъектов кредитных историй по баллам скоринга



Источник: данные Национального банка.

Рисунок 6

прогнозная вероятность дефолта) по формуле:

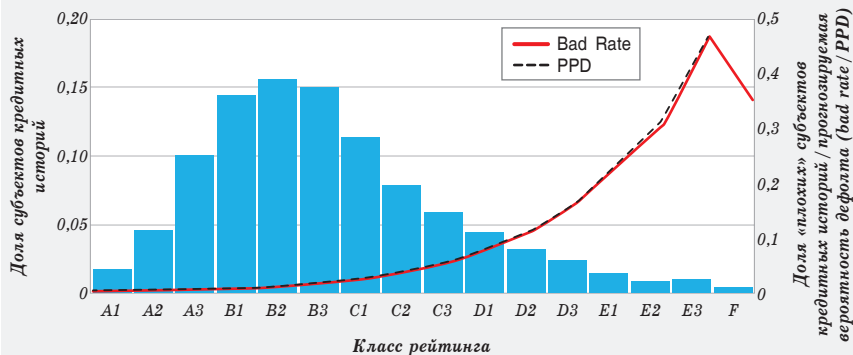
$$PPD(score) = \frac{100}{1 + \exp\left(\frac{score - 30}{40} \ln(2)\right)},$$

где score – итоговая оценка в баллах.

Скоринговая модель тестировалась на данных 2014 г. и первых месяцев 2015 г. Индекс стабильности популяции (PSI) имел максимальное значение 0,12

² Log-odds – логарифм отношения вероятностей наступления и ненаступления события, логарифм отношения шансов.

Распределение субъектов кредитных историй по классам рейтинга



Источник: данные Национального банка.

Рисунок 7

в декабре 2014 г., а все остальные месяцы тестируемого периода не превышал 0,1. Это позволяет утверждать, что модель ведет себя стабильно вне периода разработки.

В Республике Беларусь некоторые банки применяют собственные

скоринговые модели, и это дает им конкурентное преимущество. Предложение кредитного скоринга на основе данных Кредитного регистра может несколько нивелировать технологический разрыв между банками. В то же время

кредитный скоринг Кредитного регистра будет полезен всем банкам независимо от наличия собственных скоринговых моделей. Скоринговая оценка, рассчитанная на основе объединенной информации, полученной от многих кредиторов, будет включать в себя характеристики, которые недоступны для отдельного кредитора.

Согласно мировой практике, банки сочетают скоринги кредитных бюро со своими внутренними скорингами и правилами принятия решений (матричный подход). Матричный подход (когда банк использует два скоринга: один, разработанный за счет внутренних ресурсов на основе данных из своей клиентской базы, и другой, общий скоринг кредитного бюро, разработанный по широкому спектру информации национальных кредитных бюро) использует лучшее из двух подходов, и могут быть реализованы стратегии, учитывающие оба скоринга.