

ИСПОЛЬЗОВАНИЕ ИНСТРУМЕНТАРИЯ DATAMINING В УПРАВЛЕНИИ КРЕДИТНЫМИ РИСКАМИ

Эффективная система банковского кредитования – необходимое условие функционирования промышленности, которая, в свою очередь, создает реальные условия для перехода страны на новый этап развития. В Украине, по данным НБУ[1], по состоянию на конец апреля 2013 г. нефинансовым корпорациям были предоставлены кредиты на сумму 617 млрд грн (в том числе 161,561 млрд грн кредитных ресурсов размещено в промышленности), тогда как домохозяйства получили кредитов на 190 млрд грн, то есть приток кредитных ресурсов в производство почти в 3,3 раза больше, чем сумма тех кредитов, которые используются преимущественно на потребление, что свидетельствует о более высоких темпах развития производства. Но развитию эффективного кредитования мешает высокая стоимость кредитов, что во многом определяется высоким уровнем кредитного риска.

Кредитная деятельность во многом определяет результативность функционирования банка, поскольку значительная часть дохода банков формируется в результате кредитных операций. При этом кредитование всегда связано с риском. Невозврат кредитов может привести банк к банкротству, а это, в свою очередь, может повлечь за собой банкротство связанных с ним предприятий, банков и частных лиц. Поэтому проблема результативного управления кредитным риском является необходимой частью стратегии и тактики выживания и развития любого коммерческого банка.

Кредитный риск можно определить как вероятность убытков вследствие несоблюдения заемщиками первоначальных условий договоров по исполнению ими принятых на себя денежных обязательств. Исходя из этого

кредитный риск – это максимально ожидаемый убыток, который может произойти с заданной вероятностью в течение определённого периода времени в результате уменьшения стоимости кредитного портфеля в связи с частичной или полной неплатёжеспособностью заёмщиков к моменту погашения кредита [2, с. 24]. Под управлением кредитными рисками в работе [3, с. 9] подразумевается система взаимосвязанных и взаимозависимых методов сознательного, целенаправленного воздействия, направленных на недопущение вероятностного отклонения действительности от ожидаемых результатов (наступление рискованного события) или извлечение дополнительной выгоды (дохода, прибыли) в сравнении с ожидаемым результатом в условиях преодоления неопределённости в движении кредитов.

Один из основных методов управления банковским кредитным риском – это предупреждение, то есть ликвидация предпосылок возникновения кредитного риска в будущем. Немаловажную роль в этом играет оценка кредитоспособности заемщика и установление его кредитного рейтинга. То есть для повышения эффективности кредитного процесса следует уделять серьезное внимание разработке методов анализа кредитоспособности и платежеспособности заемщиков, позволяющих давать обоснованные рекомендации о предоставлении кредитов, повышая, таким образом, их качество.

Проблеме управления кредитным риском посвящены многочисленные исследования зарубежных и отечественных ученых и специалистов: Е.А. Барановой, Р.А. Давыдова, П.П. Ковалева, Г.Г. Коробовой, Л.В. Лепетикова, Г.Г. Меликьяна, В.А. Путиловского, С.В. Пыхтина, Л.М. Резвановой, И.Н. Ры-

ковой, Н.А. Савинской, А.А. Слуцкого, О.А. Солдатовой, М.И. Сухова, Н.В. Фисенко и других.

Возможности оценивания кредитоспособности заемщиков и проблемы использования скоринга для анализа кредитоспособности подробно описаны в работах Ю.Г. Максугова, А.С. Манделя, М.Ф. Наумова, А.А. Строева, и А.В. Бекарева, М.С. Помазанова, В.В.Ковалева, П.П. Ковалева, Н.Б. Паклина и других ученых. Эти работы преимущественно посвящены проблеме выяснения кредитоспособности юридических лиц на основе проведения комплексного анализа финансового состояния предприятий – потенциальных заемщиков, для чего имеется множество отечественных и зарубежных методик, основанных на использовании качественных и количественных инструментов.

Оценить кредитоспособность физических лиц намного сложнее, поскольку для этого, зачастую, недостаточно информации. Этот недостаток можно восполнить, используя технологию DataMining, которая (по определению одного из основателей Г. Пятецкого-Шапиро) представляет собой процесс «обнаружения в сырых данных ранее неизвестных, нетривиальных, практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности» [4].

В банковской сфере вопрос применения технологии DataMining становится все более актуальным. Это обусловлено, прежде всего, накоплением банками больших объемов информации, увеличением количества случаев невозврата кредитов, ужесточением конкурентной борьбы. В отчете Ассоциации американских банкиров (АВА) отмечается, что 45 из 100 крупнейших банков США уже внедрили у себя системы интеллектуального анализа данных, и еще около 50 банков запустили пилотные проекты или планируют это сделать в ближайшее время [5]. С помощью инструментов DataMining банк может получить «профили» добросовестных и неблагонадежных заемщиков. Кроме того, возможно классифицировать заемщика по группам риска, то есть не только решить вопрос о воз-

можности кредитования, но и установить лимит кредита, проценты по нему и срок возврата.

Целью статьи является использование современных математических методов и ИТ-технологий, в частности, методов Data-Mining, для оценки кредитоспособности физических лиц – потенциальных заемщиков.

Из медицинской практики хорошо известно, что легче предупредить болезнь, чем ее вылечить. Аналогично рассуждая, следует основные усилия направить на предупреждение возникновения кредитного риска.

Банковский кредитный риск можно выявить через количество проблемных (просроченных и сомнительных) кредитов. Действительно, доля проблемных кредитов в общем объеме выданных кредитов является одним из индикаторов финансовой безопасности банковского сектора Украины и характеризует кредитные риски [6, с.156].

Удельный вес просроченных и сомнительных кредитов в общем объеме кредитования в банковской системе Украины достаточно велик (рис. 1), причем он заметно вырос, начиная с 2008 г.

По мнению одного из руководителей Банка Москвы [10], в подавляющем большинстве проблемные кредиты возникают из-за реализации коммерческого, предпринимательского риска. Наиболее типичными причинами возникновения таких кредитов является сокращение рынка сбыта и фатальные проблемы с неплатежами дебиторов. Цепочки неплатежей выстраиваются по всей экономике. Средние и мелкие предприятия, имеющие наименьшие запасы прочности, часто сталкиваются с проблемой невозврата взятых кредитов.

На конец 2012 г. объем рынка потребительского кредитования в Украине достиг 40-45 млрд грн (не считая ипотеки, кредиты под залог недвижимости и автокредитование). Эксперты прогнозируют его рост как минимум на ближайшие 2-3 года. Это связано, в первую очередь, с тем, что проникновение услуги потребительского кредитования в Украине по сравнению с другими странами до сих пор находится на очень низком уровне [11].



Рассчитано по данным работ [7, с. 19; 8, с. 39; 9, с. 8]

Рис. 1. Соотношение величины кредитования физических лиц и проблемных кредитов в Украине

Из рис. 1 хорошо видно, что характер изменения доли проблемных кредитов с небольшим лагом следует за количеством кредитных средств, предоставленных заемщикам – физическим лицам. В этом нет ничего удивительного, поскольку и в потребительском кредитовании налицо все предпосылки к росту проблемной задолженности.

Таким образом, очевидно, ощутимый вклад в образование проблемной задолженности (в величину кредитного риска) вносят кредиты, выданные физическим лицам.

Единственный способ полностью исключить появление проблемных кредитов состоит в отказе от осуществления кредитных операций. Однако, поскольку кредитование является приоритетной сферой размещения банковских активов, которая приносит наибольший доход, такой способ не может найти применения. Поэтому целесообразно сосредоточить усилия на профилактической работе, которая должна быть направлена на прогнозирование возможных осложнений по возвращению кредитов и устранение причин их возникновения. Причем превентивные мероприятия необходимо применять не только по отношению к текущим кредитам, но еще на стадии рассмотрения заявки от заемщика на выдачу кредита.

Комплексная оценка кредитоспособности заемщиков коммерческого банка является

сложным процессом, состоящим из различных этапов, дополняющих друг друга. Это проверка благонадежности, основанная на изучении кредитной истории, скоринговая оценка, а также оценка финансового положения клиента (на основе финансовых показателей платёжеспособности).

Под оценкой кредитоспособности заемщика чаще всего банком подразумевается анализ возможности и целесообразности предоставления заемщику денежных средств, определение вероятности их возврата своевременно и в полном объеме [12]. Кредитная история представляет собой систематизированную информацию о том, какие кредиты брал заемщик и насколько дисциплинированно он соблюдал кредитные обязательства. Для формирования, хранения и использования кредитных историй существуют специализированные бюро, деятельность которых строго регламентирована законодательством, известны они под названием "бюро кредитных историй".

Первые кредитные бюро появились в середине прошлого века в Нью-Йорке, в России соответствующее учреждение было создано в 2000 г.

Верховная Рада в 2005 г. приняла Закон Украины "Об организации формирования и обращения кредитных историй" [13], который определяет правовые и организаци-

онные основы формирования и ведения кредитных историй. Согласно ст. 7 закона кредитная история содержит данные, идентифицирующие личность заемщика: ФИО, дату рождения, паспортные данные, место жительства, сведения о текущей трудовой деятельности, семейное положение заемщика и число граждан, находящихся на его иждивении. Кроме того, история кредитов содержит информацию о самом денежном обязательстве (кредите), сумме обязательства по заключенной кредитной сделке, виде валюты обязательства, сроке и порядке исполнения кредитного договора, а также сведения о размере погашенной суммы и окончательной сумме обязательства по кредитному договору. Согласно закону сбор, хранение, использование информации, которая составляет кредитную историю, осуществляет бюро кредитных историй.

Важная особенность функционирования бюро кредитных историй – использование так называемых "меток" в кредитных историях, или критериев ранжирования нарушений по погашению своих финансовых обязательств. Существуют три основных критерия ("метки"): невозврат ссуды в прошлом (или "черная метка"); средняя просрочка платежей ("серая метка"); "нормальная", приемлемая банками просрочка платежа не более 5 дней. Наличие либо отсутствие подобных "меток" в кредитной истории заемщика может стать решающим фактором при принятии решения о выдаче ему ссуды или кредита. Предполагается, что если нарушений в выплатах по ссудам и кредитам у заемщика нет, ему будут предоставляться льготные условия как по получению кредита, так и по его погашению.

Информация, содержащаяся в кредитной истории, используется при построении скоринговых систем. Основная идея скоринга состоит в том, чтобы, используя кредитные истории заемщиков прошлых периодов, оценить риск того, что потенциальные заемщики, обратившиеся в банк в настоящее время, не вернут полученные ссуды. Оценка осуществляется с помощью математических моделей, строящихся на такой кредитной истории.

Для построения модели оценки кредитоспособности заемщика – физического лица в условиях скоринга сначала осуществляется отбор клиентов кредитной организации, ко-

торые уже так или иначе себя зарекомендовали. Скоринговая система оценки потенциальных заемщиков, как правило, предполагает наличие трех разделов: информация по кредиту, сведения о клиенте, финансовое положение клиента. Такая выборка может включать от нескольких тысяч до сотен тысяч наблюдений в зависимости от накопленной статистики и объема кредитного портфеля. Следовательно, в распоряжении ЛПР (лица, принимающего решение) имеется большой объем разнообразной информации о клиентах, разобраться в которой можно только с помощью современного математического аппарата.

Наиболее перспективным направлением в построении скоринговых систем для оценки кредитоспособности заемщиков, по нашему мнению, является технология Data Mining. Название Data Mining (с англ. data – данные, сведения; mining – добыча руды в шахте) по-разному переводится на русский язык: добыча данных, глубинный анализ данных, интеллектуальный анализ данных, извлечение данных и т. д. По образному высказыванию Ю.А. Денисова, «датамайнинг – рутинный процесс поиска необходимых сведений в полноводных источниках информации, которые обрушиваются на человека в немислимых количествах» [14].

В основе Data Mining лежат различные методы классификации, прогнозирования и моделирования, которые могут базироваться на деревьях решений, искусственных нейронных сетях, генетических алгоритмах, эволюционном программировании, ассоциативной памяти, нечеткой логике. Кроме того, к Data Mining иногда причисляют такие статистические методы, как дескриптивный, корреляционный, регрессионный, факторный, дисперсионный, компонентный, дискриминантный анализ, а также анализ временных рядов. Тем не менее такие статистические методы анализа требуют, чтобы исследователь изначально имел какие-то представления об изучаемых данных, в то время как при использовании Data Mining цель состоит в том, чтобы обнаружить данные, которые ранее не были известны.

В настоящее время в скоринге для расчета рейтинга заемщиков и управления кредитными рисками получила распространение логит-регрессия. Поэтому, несмотря на свое

"происхождение" из статистики, логит-регрессию и ROC-анализ почти всегда можно увидеть в наборе DataMining алгоритмов [15].

Логит (логистическая) регрессионная модель предназначена для решения задач предсказания значения непрерывной зависимой переменной, при условии, что эта переменная может принимать значения на интервале от 0 до 1, причем значения зависимой переменной не могут быть меньше (или равными) 0, или больше (или равными) 1, независимо от значений факторов. Построение логит-регрессии актуально при оценке вероятности наступления того или иного события (например, выданный кредит оказался проблемным) в зависимости от значений независимых переменных (факторов, предикторов).

Использование логит-регрессии возможно и для решения задач двоичного выбора (или задач с бинарным откликом). Такие задачи появляются, когда в роли зависимой переменной выступает факт наступления данного события. Зависимая переменная принимает значение 1, если данное событие произошло, и 0 – в противном случае. Например, зависимая переменная принимает значение 0, если заемщик не в состоянии расплатиться по кредиту, и значение 1, если выполнены кредитные обязательства.

Формула логит-модели (логистической регрессии) имеет вид

$$p_i = F(z_i) = \frac{e^z}{1 + e^z},$$

$$\text{или } p_i = F(z_i) = \frac{1}{1 + e^{-z}},$$

где p_i – вероятность наступления события;

e – основание натурального логарифма

$$z = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n + \varepsilon;$$

x_i – значения независимых переменных;

β_i – коэффициенты, оценка которых является задачей бинарной логистической регрессии, весовые коэффициенты.

Если p получит значение, меньшее 0,5, то можно предположить, что событие не наступит; в противном случае предполагается наступление события.

Для оценивания параметров β модели бинарного выбора, а именно такой является рассматриваемая модель логит-регрессии, используется метод максимального правдоподобия – это метод оценивания неизвестно-

го параметра путём максимизации функции правдоподобия, который основан на предположении о том, что вся информация о статистической выборке содержится в функции правдоподобия.

Интерпретация коэффициентов логит-модели отличается от модели линейной регрессии. В модели бинарного выбора коэффициенты показывают, насколько изменится вероятность получения значения $Z=1$ при изменении величины независимой переменной на единицу и при неизменных значениях других переменных. Отрицательный знак при коэффициенте регрессии говорит об уменьшении вероятности при увеличении соответствующих переменных, положительный – об увеличении.

Оценка адекватности построенной модели основана на анализе тестовых характеристик и статистической проверке гипотез:

для оценки статистической надежности оценок параметров применяется Р-статистика;

для анализа уравнения в целом проверка нулевой гипотезы о значимости коэффициентов проводится с помощью тестов Вальда (W), множителей Лагранжа (LR), отношения правдоподобия (LM). Во всех этих тестах нулевая гипотеза формулируется следующим образом:

$$H_0: Q\beta = r,$$

где Q – известная матрица ограничений;

β – вектор тестируемых параметров;

r – вектор констант.

Суть нулевой гипотезы состоит в том, что коэффициенты при всех включенных в модель переменных одновременно равны нулю. Если нулевая гипотеза отклоняется, то, значит, в модели присутствуют факторы, оказывающие статистически значимое влияние на эндогенную переменную.

Кроме того, эти тесты объединяет и то обстоятельство, что критические статистики для всех них имеют распределение χ^2 , и проверка нулевой гипотезы проводится следующим образом: вычисленное значение статистики сравнивается с табличным значением распределения χ^2 для заданного уровня значимости (α) и при определенном числе степеней свободы (ν). Если вычисленное значение превышает критическое, то нулевая гипотеза отклоняется в пользу альтернативной (т.е. не все коэффициенты одновременно равны нулю).

Статистики для рассматриваемых тестов определяются следующим образом:

1) Вальда

$$W = (Q\hat{\beta} - r)'(Q(X'\Omega^{-1}X)^{-1}Q')^{-1} \cdot$$

$$\cdot (Q\hat{\beta} - r) \sim \chi^2(v),$$

где Ω – известная матрица ковариаций ошибок;

2) множителей Лагранжа (LM)

$$LM = \tilde{l}'Q(X'\Omega^{-1}X)^{-1}Q'\tilde{l} \sim \chi^2(v),$$

где $l = -2(\ln L(\tilde{\beta}) - \ln L(\hat{\beta}))$ – логарифмическая функция правдоподобия рассматриваемой модели;

\tilde{l} – ограниченная логарифмическая функция правдоподобия, т. е. логарифмическая функция правдоподобия для модели, в которой величина всех параметров, кроме свободного члена, равна нулю;

3) отношения правдоподобия

$$LR = -2(\ln L(\tilde{\beta}) - \ln L(\hat{\beta})) = 2(\tilde{l} - l),$$

где $\hat{\beta}$ и $\tilde{\beta}$ – соответственно оценки параметров для регрессии без ограничения и с ограничением.

Таким образом, в качестве критической статистики теста берется разность максимумов логарифмических функций правдоподобия.

Для исследования прогностической способности модели можно использовать классификационную таблицу, которая представляет собой характеристику правильных и ошибочных классификаций используемой выборки объектов. Построение таблицы основывается на использовании порогового значения C и на вычислении ожидаемых значений зависимой переменной. В результате, чем больше полученных правильных классификаций и чем меньше значения оценок вероятности ошибок, тем выше прогностическая ценность построенной модели.

По материалам о кредитовании физических лиц одного из банков выполнено построение и исследование логистической модели диагностики кредитоспособности потенциального клиента, в которой в качестве зависимой переменной используется бинарная переменная, отражающая статус клиента. Зависимая переменная будет принимать нулевое значение, если кредит оказался проблемным, и значение равно 1, в противном случае. Значение в интервале от 0 до 1 будет

свидетельствовать о вероятности невозврата кредита или несвоевременного возврата.

В число независимых переменных были введены данные о кредитной истории заемщика, возрасте, половой принадлежности, сведения о семейном положении, наличии и числе иждивенцев, наличии в собственности заемщика движимого и недвижимого имущества, об уровне дохода и сроке проживания в данном регионе и работы на последнем месте. Причем две переменные – характеристика половой принадлежности и сведения о семейном положении, являются качественными и поэтому будут представлены следующим образом:

$$x_2 = \begin{cases} 1, & \text{если это женщина} \\ 0, & \text{если это мужчина,} \end{cases}$$

$$x_3 = \begin{cases} 1, & \text{если состоит в браке} \\ 0, & \text{если не состоит в браке.} \end{cases}$$

Бинарная логистическая регрессия в таком случае рассчитывает вероятность наступления события в зависимости от значений независимых переменных.

Построение модели было выполнено с использованием пакета STATISTICA в подмодуле Логит-регрессия модуля Нелинейная оценка.

Параметры полученной модели приведены на рис.2.

На основе информации, приведенной на рис. 2, можно сделать первоначальный вывод о качестве построенной модели. Так, l – логарифмическая функция правдоподобия – принимает значение, равное 166,922, и \tilde{l} – ограниченная логарифмическая функция правдоподобия – значение 443,5017. Таким образом,

$$LR = 2(\tilde{l} - l) = 2(443,5017 - 166,922) = 553,159.$$

Кроме того, здесь содержится p -уровень гипотезы. Поскольку этот уровень меньше 5% ($p = 0,000000$), то модель значима. Значение статистики $\chi^2 = 276,5797$ для разницы между текущей моделью и моделью, содержащей лишь свободный член, высоко значимо. Можно вычислить индекс отношения правдоподобия (коэффициент детерминации Макфаддена), который принимает значение

$$R_{MF}^2 = 1 - \frac{1}{1 + \frac{LR}{N}} = \frac{LR}{LR + N} = 0,633.$$

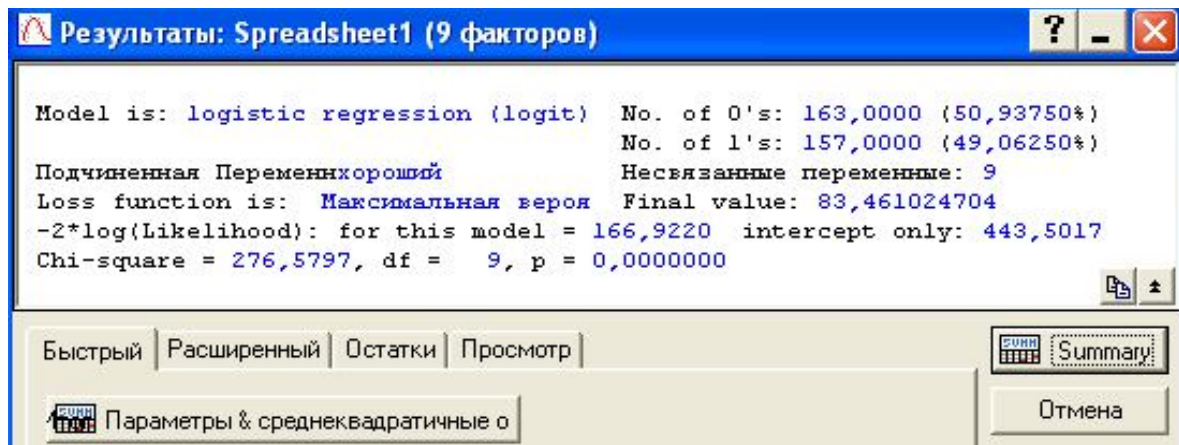


Рис. 2. Результаты оценивания логистической регрессии

Этот коэффициент аналогичен коэффициенту детерминации в линейной регрессии. То есть можно сказать, что модель на 63,3% характеризует изменчивость исследуемого показателя.

Поэтому можно сделать вывод, что рассматриваемые факторы в совокупности определяют успешность оценивания кредитоспособности заемщиков.

На рис. 3 представлен результат оценивания коэффициентов регрессии.

	Const.B0	возраст	пол	в браке	на иждив.	доход	работа	проживание	собственность	плата по кр.
Оценить	-3,32301	-0,02012	-0,60443	0,153801	-1,23615	7,689026E-01	0,0738	0,01554	0,02544	-1,16815
Стандартная ош	1,40294	0,04286	0,43187	0,415651	0,26852	9,917734E-02	0,0546	0,01865	0,01002	0,17271
t(310)	-2,36861	-0,46954	-1,39956	0,370023	-4,60359	7,752805E+00	1,3508	0,83333	2,53898	-6,76359
p-level	0,01847	0,63902	0,16265	0,711618	0,00001	1,300834E-13	0,1777	0,40530	0,01161	0,00000
-95%CL	-6,08349	-0,10446	-1,45419	-0,664054	-1,76450	5,737567E-01	-0,0337	-0,02115	0,00572	-1,50799
+95%CL	-0,56252	0,06421	0,24534	0,971654	-0,70780	9,640484E-01	0,1813	0,05223	0,04515	-0,82832
Wald's Chi-square	5,61030	0,22046	1,95877	0,136917	21,19306	6,010599E+01	1,8248	0,69444	6,44644	45,74615
p-level	0,01786	0,63869	0,16165	0,711368	0,00000	9,254028E-15	0,1768	0,40466	0,01112	0,00000
Odds ratio (unit ch)	0,03604	0,98008	0,54639	1,166258	0,29050	2,157398E+00	1,0766	1,01566	1,02576	0,31094
-95%CL	0,00228	0,90081	0,23359	0,514760	0,17127	1,774922E+00	0,9669	0,97907	1,00574	0,22136
+95%CL	0,56977	1,06632	1,27805	2,642313	0,49273	2,622291E+00	1,1988	1,05362	1,04619	0,43678
Odds ratio (range)		0,47492	0,54639	1,166258	0,00712	3,580828E+11	8,5028	2,14122	9,37983	0,00005
-95%CL		0,02096	0,23359	0,514760	0,00086	4,184141E+08	0,3763	0,35474	1,65489	0,00000
+95%CL		10,75915	1,27805	2,642313	0,05894		192,1208	12,92446	53,16434	0,00087

Рис. 3. Результаты оценивания коэффициентов логистической регрессии

Полученная модель имеет вид
 $Z = -3,323 - 0,02x_1 - 0,604x_2 + 0,154x_3 - 1,236x_4 + 0,769x_5 + 0,074x_6 + 0,016x_7 + 0,025x_8 - 1,168x_9$,
 тогда

$$p_i = F(z_i) = \frac{1}{1 + e^{-(-3,323 - 0,02x_1 - 0,604x_2 + 0,154x_3 - 1,236x_4 + 0,769x_5 + 0,074x_6 + 0,016x_7 + 0,025x_8 - 1,168x_9)}}$$

Таким образом, видим, что возраст, количество лиц, находящихся на иждивении, размер ежемесячной платы по кредиту оказывают на вероятность того, что кредит окажется не проблемным ($Z = 1$), обратное влия-

ние, то есть чем больше величина этих переменных, тем меньше указанная вероятность. Такие факторы, как наличие в собственности заемщика движимого и недвижимого имущества, уровень дохода, время проживания в

данном регионе и работы на последнем месте влияют прямо, т.е. с их ростом вероятность того, что кредит будет возвращен в полном объеме и в установленные сроки, растет. Учитывая, что среди независимых переменных были фиктивные, введенные для отображения половой принадлежности заемщика и сведений о семейном положении, то модель для прогноза кредитоспособности может быть представлена с учетом соответствующих обстоятельств. Так, если конкретный заемщик – женщина, которая состоит в браке, то в модели переменные x_2 и x_3 следует приравнять единице.

Из таблицы видим, что не все оценки параметров модели можно считать статистически значимыми. С вероятностью 0,95 статистически надежными являются только коэффициенты при переменных x_4 (количество иждивенцев), x_5 (доход), x_8 (недвижимость в собственности) и x_9 (ежемесячная плата по кредиту), а также величина свободного члена (Const.ВО). Уровни значимости для них (p -уровень) меньше 0,05. Остальные коэффициенты имеют низкие значения t -статистик и величину p -значений, превышающую 0,05. Поэтому можно сделать вывод, что

возраст, пол, семейное положение, а также длительность проживания в данной местности и трудовой стаж существенного влияния на кредитоспособность не оказывают. Это отчасти можно объяснить высокой степенью связи между возрастом и стажем: коэффициент корреляции между этими показателями оказался равным 0,717.

Для того чтобы исключить влияние мультиколлинеарности были построены еще 3 модели: в одну из них были включены все независимые переменные, кроме возраста; в другую – кроме длительности работы на последнем месте; в третью – только те переменные, которые имеют статистически надежные коэффициенты. Следует отметить, что все модели в целом оказались адекватными.

Третья модель (модифицированная) имеет следующий вид:

$$p_i = F(z_i) = \frac{1}{1 + e^{-(-3,0 - 1,0x_4 + 0,769x_5 + 0,025x_8 - 1,152x_9)}}$$

Все коэффициенты данной модели являются статистически надежными с высокой степенью доверительной вероятности.

Выводы о прогностической ценности моделей можно сделать, изучая классификационную таблицу.

Таблица

Матрица ошибок классификации

Спрогнозировано моделью	Фактически	
	«хороший»	«плохой»
«Хороший»	<p>TP</p> <p>Истинно положительные («хороший» классифицирован как «хороший»)</p>	<p>FP</p> <p>Ложно положительные («плохой» классифицирован как «хороший»)</p> <p>Ошибка II рода</p>
«Плохой»	<p>FN</p> <p>Ложно отрицательные («хороший» классифицирован как «плохой»)</p> <p>Ошибка I рода</p>	<p>TN</p> <p>Истинно отрицательные («плохой» классифицирован как «плохой»)</p>

Показателями, отражающими эффективность модели бинарной классификации, являются чувствительность и специфичность.

Чувствительность определяется как отношение числа истинно положительных наблюдений к числу фактически положительных и характеризует долю истинно положительных наблюдений относительно количества наблюдений, классифицированных моделью как положительные

$$S_e = \frac{TP}{TP + FN}$$

где TP – число истинно положительных наблюдений; FN – число ложно отрицательных наблюдений.

Чувствительность тем меньше, чем больше ложно отрицательных наблюдений, т.е. положительных наблюдений, ошибочно распознанных, как отрицательные. Верхний предел чувствительности равен 1, когда ложно отрицательные наблюдения отсутствуют,

т.е. не допущено ни одной ошибки. Модель, обладающая высокой чувствительностью, обеспечивает большую вероятность правильного распознавания для положительных примеров.

Специфичность – это показатель, отражающий точность работы бинарной классификационной модели. Он определяется как отношение истинно отрицательных наблюдений к числу фактически отрицательных

$$S_p = \frac{TN}{TN + FR}$$

Если число ложно положительных наблюдений велико, т.е. модель допустила большое количество ошибок, распознав положительные наблюдения как отрицательные, то специфичность стремится к 0; если же

ложно отрицательных наблюдений нет, то она равна 1. Модель, обладающая высокой специфичностью, обеспечивает большую вероятность правильного распознавания для отрицательных наблюдений.

На рис. 4 представлены классификационные таблицы для модели с полным набором независимых переменных (рис. 4а) и с переменными, отличающимися статистической надежностью (рис. 4б).

Выполненные расчеты показывают, что чувствительность как для одной, так и для другой модели высока: 90,798 и 91,4118% соответственно; то же касается и специфичности, величина которой составляет для рассматриваемых моделей 91,083 и 91,720% соответственно.

Классификация случаев (Spreadsh Odds ratio: 100,78)				Классификация случаев (Spreadsh Odds ratio: 117,89)			
Измеренный	Предсказ	Предсказ	Процент	Измеренный	Предсказ	Предсказ	Процент
0,000000	0,000000	1,000000	Исправле	0,000000	0,000000	1,000000	Исправле
0,000000	148	15	90,79755	0,000000	149	14	91,41104
1,000000	14	143	91,08280	1,000000	13	144	91,71974

а

б

Рис. 4. Классификационная таблица

В таблицах, представленных на рис. 4, выведено отношение шансов (oddsratio) для двух групп объектов. Шанс – это отношение вероятности того, что события произойдет к вероятности того, что событие не произойдет.

Поскольку отношение шансов > 100 (причем для модифицированной модели эта величина больше – 117,89), то шанс для первой группы больше шанса для второй группы.

Графическое представление проверки прогностических свойств модели выполняется на основе построения ROC-кривой (ReceiverOperatorCharacteristic), которая показывает зависимость количества верно классифицированных положительных примеров (истинно положительных) от количества неверно классифицированных отрицательных примеров (ложно отрицательных).

На графике ROC-кривой по оси ординат откладывается чувствительность (истинно положительные примеры), по оси абсцисс – либо специфичность (истинно отрица-

тельные примеры), либо 1 минус специфичность (ложно положительные примеры).

Чем кривая ближе к верхнему левому углу, тем выше предсказательная способность модели. Наоборот, чем ближе она расположена к диагональной прямой, тем менее эффективна модель.

Таким образом, изучение кредитоспособности клиента является одним из наиболее важных методов снижения кредитного риска и успешной реализации кредитной политики банка, поскольку позволяет избежать необоснованного риска еще на этапе рассмотрения заявки на предоставление кредита. С целью формирования кредитных историй в странах создаются и функционируют кредитные бюро. Скоринг представляет собой математическую модель, с помощью которой на основе кредитных историй других клиентов банк пытается определить, насколько велика вероятность того, что конкретный потенциальный заемщик вернет кредит в срок.

Для построения скоринговой модели определения кредитного рейтинга заемщика целесообразно использовать логистическую регрессию – инструмент DataMining – для решения задачи регрессии и классификации. Логит-модель позволяет классифицировать очередного заемщика по степени принадлежности к одному или другому классу – добросовестный или неблагонадежный – с определенной вероятностью, что позволит банку принять правильное решение относительно предоставления кредита и тем самым снизить кредитный риск.

Дальнейшие исследования предполагается направить на получение «профилей» недобросовестных и благонадежных заемщиков с помощью таких инструментов DataMining, как деревья решений, представляющие собой систему, разделяющую клиентов на группы, внутри которых уровень риска одинаков и максимально отличается от уровня риска других групп, а также использовать кластерный анализ при построении так называемых обучающих выборок для дискриминантного анализа, с помощью которого можно классифицировать заемщика по группам риска.

Литература

1. Статистичний бюлетень (електронне видання) (2013). Національний банк України [Електронний ресурс]. – Режим доступу: http://www.bank.gov.ua/control/uk/publish/category?cat_id=57897.
2. Помазанов М. Количественный анализ кредитного риска / М. Помазанов // Банковские технологии. – 2004. – № 2. – С. 22-28.
3. Ковалев П.П. Пути повышения результативности кредитного риск-менеджмента в коммерческом банке: автореф. дис. ... на соиск. уч. степени канд. экон. наук: спец. 08.00.10 – Финансы, денежное обращение и кредит. – М.: Государственный НИИ системного анализа Счетной палаты РФ. – 2006. – 24 с.
4. DataMining в системе управления знаниями. – [Интернет-издание]. – Режим доступа. – <http://www.smart-edu.com/upravlenie-znaniyami/izvlechenie-znaniy-data-mining-v-sisteme-upravleniya-znaniyami.html>.
5. DATA MINING в банках: перспектива или реальность? <http://www.inftech.web-servis.ru/it/database/datamining/ar5.html>
6. Фінансова безпека підприємств і банківських установ: моногр. /за заг. ред. д-ра екон. наук, проф. А.О. Єпифанова, [А.О. Єпифанов, О.Л. Пластун, В.С. Домбровський та ін]. – Суми: ДВНЗ «УАБС НБУ», 2009. – 295 с.
7. Основные показатели деятельности банков Украины на 01.01.2009 року // Весн. нац. банка Украины. – 2009. – № 2. – С. 19.
8. Основные показатели деятельности банков Украины на 01.01.2012 року// Весн. нац. банка Украины. – 2012. – № 2. – С. 39.
9. Ежемесячный дайджест «Обзор банковского сектора Украины» за январь 2013 г., с. 8. – [Интернет-издание]. – Режим доступа. – <http://www.kreditprombank.com/upload/content/510/January2013!.pdf>.
10. Необходима профилактика долгов. [Интернет-издание]. – Режим доступа. – <http://eco-nomicus.ru/banki/89-neobhodima-profilaktika-dolgov.html>.
11. Рынок потребительского кредитования продолжит бурный рост [Интернет-издание]. – Режим доступа. – <http://minfin.com.ua/2013/03/30/738661>.
12. Ефимов А.М. Современные методы оценки кредитоспособности физических лиц // "Банковский ритейл". – 2010. – № 2. – [Интернет-издание]. – Режим доступа. – <http://www.lawmix.ru/bux/10172>.
13. Закон Украины «Об организации формирования и обращения кредитных историй» от 23 июня 2005 года №2704-IV. - [Интернет-издание]. – Режим доступа. – <http://zakon4.rada.gov.ua/laws/show/2704-15>.
14. Денисов Ю.А. Учебник по датамайнингу. - [Интернет-издание]. – Режим доступа. – http://www.simple-clerk.narod.ru/DATAMINING/index_vol01.htm.
15. Паклин Н.Б. Применение логистической регрессии в медицине и скоринге. - [Интернет-издание]. – Режим доступа. – http://www.basegroup.ru/files/image/library/practice/logis_medic_scoring/loans.txt.

Представлена в редакцию 17.05.2013 г.