

Анкета участника

Информация	Описание
ФИО студента	Якупов Роберт Марсович
Направление/специальность	080801 Прикладная информатика
Вуз	ГОУ ВПО «Ивановский государственный энергетический университет им. В.И. Ленина»
Вуз-партнер	Да
Город	Иваново
Кафедра	Информационные технологии
ФИО зав. кафедрой	Белов А.А.
Тема ВКР	Модернизация системы информационного обеспечения кредитования физических лиц коммерческого банка
Версия Deductor	5.0
Дата защиты	16.06.2010
Оценка	Отлично
Руководитель ВКР	Канд. техн. наук, доцент Баллод Борис Анатольевич
Представлено	<ul style="list-style-type: none">▪ Анкета участника▪ Аннотация▪ Пояснительная записка▪ Сканированные титульные листы

Аннотация

Модернизация системы информационного обеспечения кредитования физических лиц коммерческого банка

В настоящее время в российской экономике наблюдается стабилизация, постепенное увеличение жизненного уровня населения. Складывающаяся ситуация явилась одной из основных причин развития рынка кредитования физических лиц: выдачи потребительских кредитов, автокредитования, ипотечного кредитования, образовательного кредитования, кредитования при помощи пластиковых карт. При этом нужно принимать во внимание, что кредитование всегда связано с риском. Таким образом, можно выделить ряд проблем, связанных с кредитованием физических лиц.

Уже в 2002 году, по данным Центробанка, населению был выдан 141 млрд. рублей. В 2003-м - 298 млрд., а в 2004-м - 616,5 млрд. А в январе 2010-го, по некоторым оценкам, объем рынка потребительского кредитования в России превысил 1 трлн. руб. На этом фоне клиента начинают уважать, за него борются и, единожды получив добросовестного плательщика, стараются его не упустить. Если же заставлять его доказывать свою платежеспособность и ожидать "вердикта" несколько недель, а то и месяцев, клиент может обидеться и уйти в другой банк. Так что быстрота реагирования на обращение клиента становится прямым конкурентным фактором. Здесь необходимо отметить рекламные заявления некоторых коммерческих банков уже в прошлом году, по поводу сокращения срока принятия решения до одного дня.

В то же время оперативность не должна отражаться на качестве принимаемых решений. Если выдавать кредиты всем без разбору, то лидирующее положение на рынке обернется финансовой катастрофой, подобной той, которая постигла в конце 2003 года Южную Корею. Объем просроченных потребительских кредитов к тому времени составлял 57,47 млрд. долларов, или 13,5% от общего числа выданных кредитов. Причиной столь безрадостной картины послужило отсутствие у банков отработанных систем управления рисками на фоне жесткой борьбы за рынок.

Т.о., деятельность в данной области связана с большим количеством рисков, касающихся в основном длительности периода кредитования. Поэтому стоимость кредитной услуги очень велика. Для привлечения клиентов необходимо, в первую очередь, снижение процентной ставки за счет исключения из нее риска неплатежа. Для этого банки должны отсеять "плохих" заемщиков и предупредить случаи невозврата и, соответственно, дополнительные расходы с этим связанные.

На фоне правовой неопределенности вопросов, связанных с кредитованием, перед банками возник целый ряд проблем снижения рисков и устранения случаев мошенничества.

Кредитование как процесс можно рассматривать с точки зрения прогнозирования, т.е. как говорил Сэмюэл Батлер «Всякая коммерция – попытка предвидеть будущее».

Любой процесс прогнозирования, как правило, строится в следующей последовательности:

1. Формулировка проблемы.
2. Сбор информации и выбор метода прогнозирования.
3. Применение метода и оценка полученного прогноза.
4. Использование прогноза для принятия решения.
5. Анализ "прогноз-факт". (совпали или ожидания с результатом)

Скоринговые модели

Задача скоринга - не только выявить потенциально неплатежеспособного клиента, но и пресечь деятельность мошенников, изначально не планирующих возвращать кредит. Если в первом случае шанс получения от клиента правдивых данных достаточно велик, то злоумышленники

пытаются вводить специально составленный набор данных, для того чтобы найти и использовать "слабые места" в системе скоринга.

Среди первых скоринговых моделей можно выделить модель Дюрана, которая была предложена еще в 1941 году Дэвидом Дюраном. К тому же времени можно отнести появление термина "кредитный скоринг", означающего сложную математическую систему определения степени платежеспособности заемщика, основанную на анализе нескольких характеристик клиента. Модель Дюрана широко применяется в финансовых организациях по сей день. В данной модели фигурируют группы факторов для определения степени кредитного риска и указаны коэффициенты для различных факторов, характеризующих кредитоспособность клиента:

- 1 Пол: женский (0,4 балла), мужской (0).
- 2 Возраст: 20 лет и меньше (0), 21 год (0,1), 22 года (0,2), 23 года и выше (0,3).
- 3 И т.д.

По мере развития ИТ банки стали заказывать разработчикам и скоринг-системы. В результате к настоящему времени даже сформировался рынок таких ИТ-решений, хотя большинство банков предпочитают уникальные разработки.

Итак, проведем обзор современных скоринговых решений:

- статистические методы, основанные на дискриминантном анализе (линейная регрессия, логистическая регрессия);
- различные варианты линейного программирования;
- дерево классификации или рекурсионно-партиционный алгоритм (РПА);
- нейронные сети;
- генетический алгоритм;
- метод ближайших соседей.

Традиционными и наиболее распространенными являются регрессионные методы, прежде всего линейная многофакторная регрессия :

$$p = w_0 + w_1x_1 + w_2x_2 + \dots + w_nx_n ,$$

где p -- вероятность дефолта, w -- весовые коэффициенты, x -- характеристики клиента.

Недостаток данной модели заключается в том, что в левой части уравнения находится вероятность, которая принимает значения от 0 до 1, а переменные в правой части могут принимать любые значения от $-\Gamma$ до $+\Gamma$.

Логистическая регрессия позволяет преодолеть этот недостаток:

$$\log(p/(1-p)) = w_0 + w_1x_1 + w_2x_2 + \dots + w_nx_n.$$

Для применения логистической регрессии необходимы гораздо более сложные расчеты для получения весовых коэффициентов и, следовательно, более мощная компьютерная база и усовершенствованное компьютерное обеспечение. Но при современном уровне развития компьютерной техники это не является проблемой, и в настоящее время логистическая регрессия является лидером скоринговых систем.

Преимущество логистической регрессии еще и в том, что она может подразделять клиентов как на две группы (0 -- плохой, 1 -- хороший), так и на несколько групп (1, 2, 3, 4 группы риска).

Все регрессионные методы чувствительны к корреляции между характеристиками, поэтому в модели не должно быть сильно коррелированных независимых переменных.

Линейное программирование также приводит к линейной скоринговой модели. Провести абсолютно точную классификацию на плохих и хороших клиентов невозможно, но желательно

свести ошибку к минимуму. Задачу можно сформулировать как поиск весовых коэффициентов, для которых ошибка и будет минимальной.

Дерево классификации и нейронные сети представляют собой системы, которые разделяют клиентов на группы, внутри которых уровень риска одинаков и максимально отличается от уровня риска других групп. Нейронные сети используются главным образом при определении кредитоспособности юридических лиц, где анализируются выборки меньшего размера, чем в потребительском кредите. Но наиболее успешной областью их применения стало выявление мошенничества с кредитными карточками благодаря их способности выявлять нестандартные ситуации

Генетический алгоритм основан на аналогии с биологическим процессом естественного отбора. В сфере кредитования это выглядит следующим образом: имеется набор классификационных моделей, которые подвергаются «мутации», «скрещиваются», и в результате отбирается «сильнейший», т. е. модель, дающая наиболее точную классификацию.

При использовании метода ближайших соседей выбирается единица измерения для определения расстояния между клиентами. Все клиенты в выборке получают определенное пространственное положение. Каждый новый клиент классифицируется исходя из того, каких клиентов - плохих или хороших - больше вокруг него.

Относительно недавно получили развитие финансовые коэффициенты оценки платежеспособности клиентов коммерческого банка (используются при кредитовании индивидуальных предпринимателей).

Выбор финансовых коэффициентов определяется особенностями клиентуры банка, возможными причинами финансовых затруднений, кредитной политикой банка. Можно выделить пять групп коэффициентов:

- I — ликвидности;
- II — эффективности, или оборачиваемости;
- III — финансового левериджа;
- IV — прибыльности;
- V — обслуживания долга.

Метод «Деревьев решений»

Наибольшую популярность, а также достоверность получаемой информации в процессе работы доказали 2 наиболее конкурирующих подхода в определении платежеспособности клиентов: «Деревья решений» и нейросетевые методы. Результаты сравнения представлены в таблице:

Из приведенных выше данных, несложно сделать вывод о схожести результатов, но все же некоторое преимущество можно отдать методу «Деревья решений», в особенности в вопросах потребительского кредитования.

Скоринговые системы используют статистические методы работы, а значит, нуждаются в массивах исходных данных о ранее выданных кредитах (выборке) для выявления зависимостей. Например, для выявления степени влияния семейного положения на возврат кредита. Существуют два способа поиска исходных данных. Можно воспользоваться уже собранной кем-то другим базой данных или использовать методы, требующие для корректного применения меньших по объему массивов опытных данных. Первый способ свойствен в первую очередь при разработке банками собственных решений. Вторым пользуются вендорами "серийных" систем скоринга. Однако ни один из этих вариантов нельзя назвать идеальным.

Очевидно, что данные о клиентах, например, американского банка мало подойдут для адекватной оценки российских заемщиков. Совершенно другие доходы, уровень жизни, менталитет делают невозможным использование такой БД даже с сильной коррекцией

результатов. Классический подход к скорингу предусматривает принятие положительного решения о выдаче кредита в том случае, если выходной показатель превышает некий критический порог. А выходной показатель вычисляется как сумма численных характеристик параметров (возраст, количество иждивенцев, доход, наличие кредитной карты и т. д.), помноженных на соответствующий "вес" (значимость показателя в общей оценке). Да и само слово "scoring" можно перевести как "подсчет очков". Обучение системы сводится к подгонке "весов". Так вот "веса" одного и того же параметра для США и РФ будут существенно различаться.

Результаты прогнозирования качества клиента при нейросетевом подходе

Класс клиента	Фактические		Прогноз	Правильный прогноз	
	0	1			
Прогнозируемые	0	25195	1061	26256	55,1%
	1	20549	3195	23744	75,1%
Итого факт		45744	4256	50000	65,1%

Результаты прогнозирования кредитного качества клиента при методе «Деревья решений»

Класс клиента	Фактические		Прогноз	Правильный прогноз	
	0	1			
Прогнозируемые	0	32070	1789	33859	70,1%
	1	13674	2467	2926	58%
Итого факт		45744	4256	50000	69,1%

Итак, задача заключается в построении модели оценки (классификации) потенциальных заемщиков. Решение задачи также должно обладать большой достоверностью классификации, возможностью адаптации к любым условиям, простотой использования модели.

Пользуясь приведенной выше методикой, была предложена гипотеза о том, какие факторы влияют на кредитоспособность человека. По мнению экспертов по этим факторам можно учесть суммарный риск. Тем самым должно достигаться и отнесение потенциального заемщика к способным вернуть кредит или не способным.

Одним из вариантов решения вышеставленной задачи является применение алгоритмов, решающих задачи классификации. Задача классификации – это задача отнесения какого-либо объекта (потенциальный заемщик) к одному из заранее известных классов (Давать/Не давать кредит). Такого рода задачи с большим успехом решаются одним из методов Data Mining – при помощи деревьев решений. Деревья решений – один из методов автоматического анализа данных. Получаемая модель – это способ представления правил в иерархической, последовательной структуре, где каждому объекту соответствует единственный узел, дающий решение.

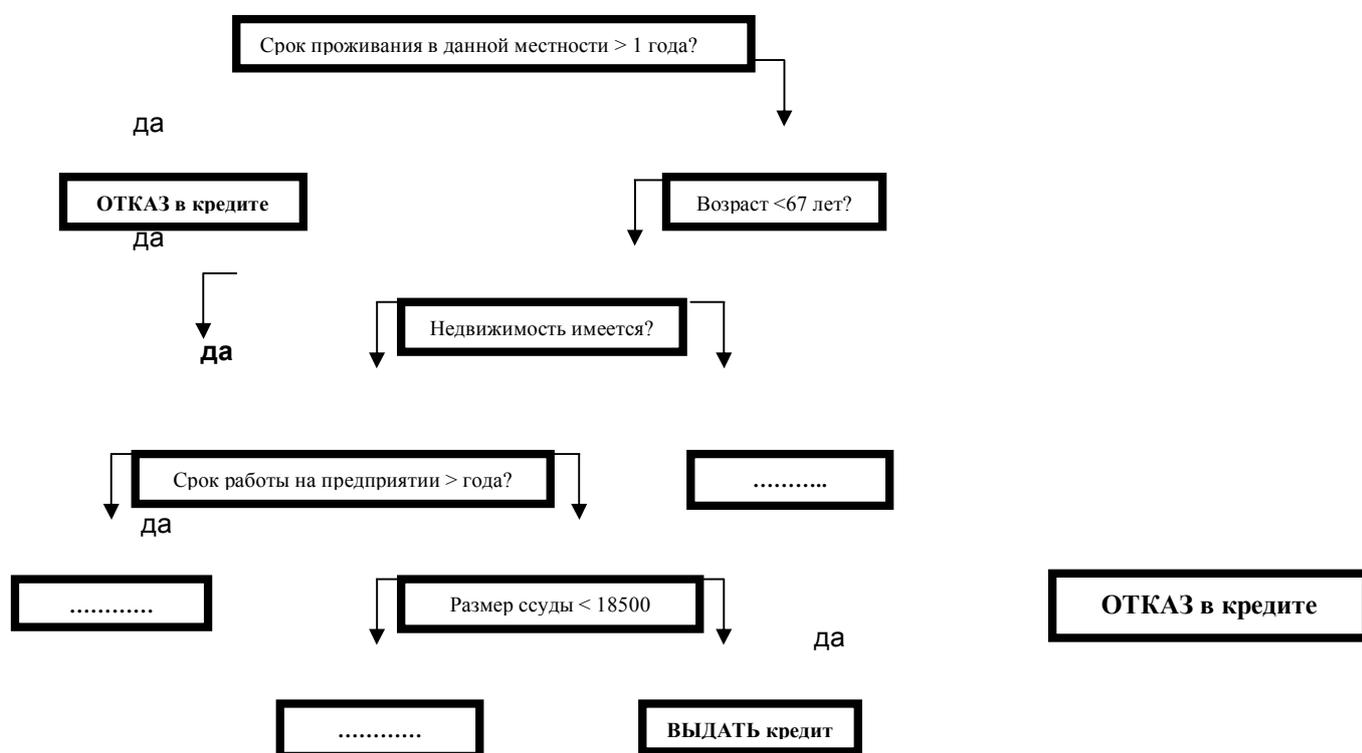


Рис. 1 Пример «Дерева решений»

1. На основе данных за прошлые периоды строится дерево. При этом класс каждой из ситуаций, на основе которых строится дерево, заранее известен. В нашем случае должно быть известно, была ли возвращена основная сумма долга и проценты и не было ли просрочек в платежах. При построении дерева все известные ситуации обучающей выборки сначала попадают в верхний узел, а потом распределяются по узлам, которые в свою очередь также могут быть разбиты на дочерние узлы. Критерий разбиения – это различные значения какого-либо входного фактора. Для определения поля, по которому будет происходить разбиение, используется показатель, называемый энтропия – мера неопределенности. Выбирается то поле, при разбиении по которому устраняется больше неопределенности. Неопределенность тем выше, чем больше примесей (объектов, относящихся к различным классам) находятся в одном узле. Энтропия равна нулю, если в узле будут находиться объекты, относящиеся к одному классу.

2. Полученную модель используют при определении класса (Давать/Не давать кредит) вновь возникших ситуаций (поступила заявка на получение кредита).

3. При существенном изменении текущей ситуации на рынке, дерево можно перестроить, т.е. адаптировать к существующей обстановке.

То есть, грубо говоря, на выходе системы получается не бинарное решение: "отдаст/не отдаст" - "1/0", а "0,53" означает, что 53 шанса из 100, что кредит будет возвращен.

Работа с аналитической платформой Deductor

Методика работы с платформой Deductor выглядит следующим образом:

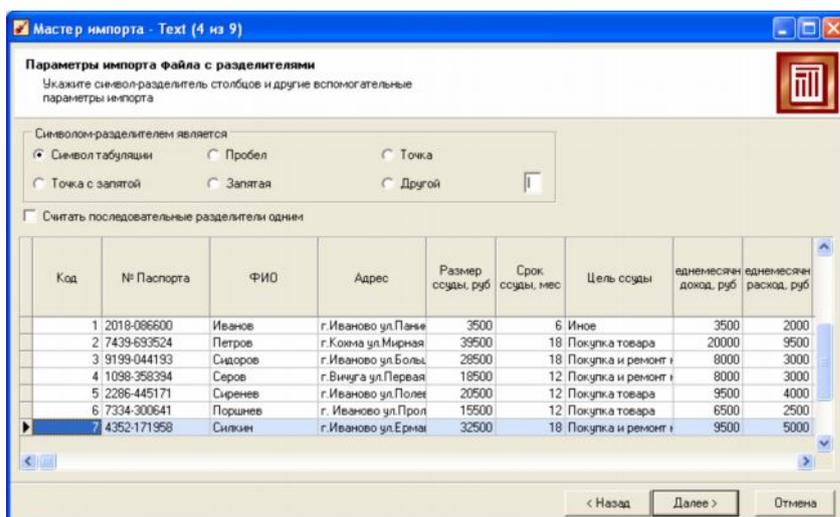


Рис. 2 Импорт данных из текстового файла

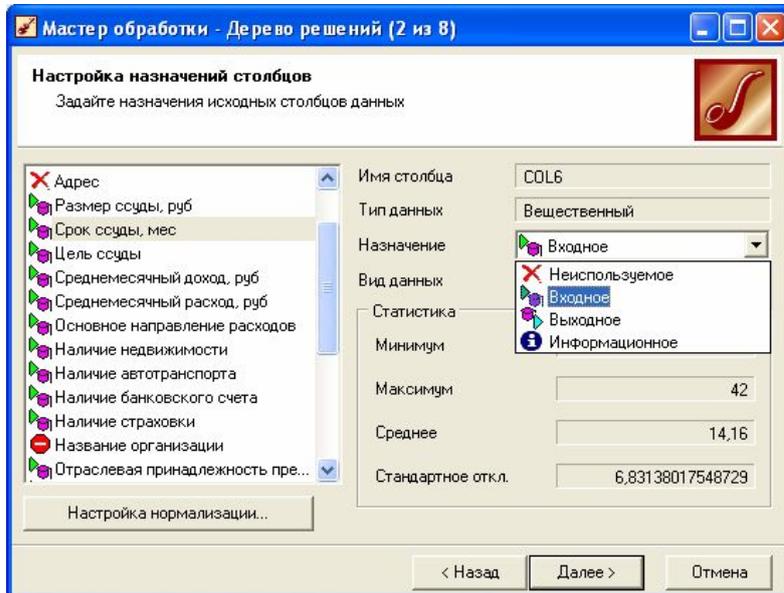


Рис. 3 Запуск мастера обработки данных

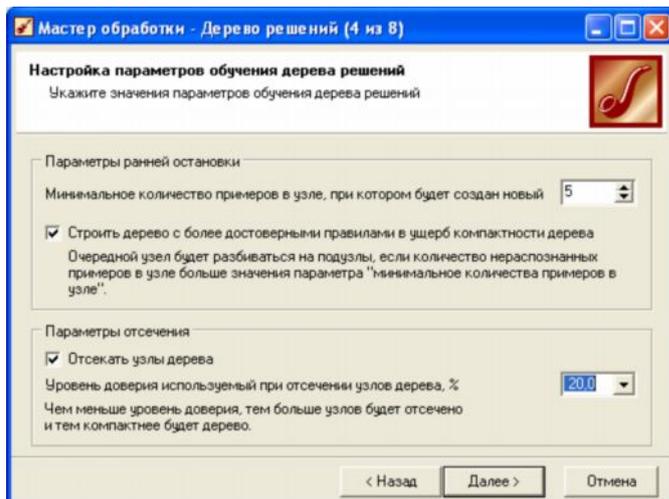


Рис. 4 Настройка «Дерева Решений»

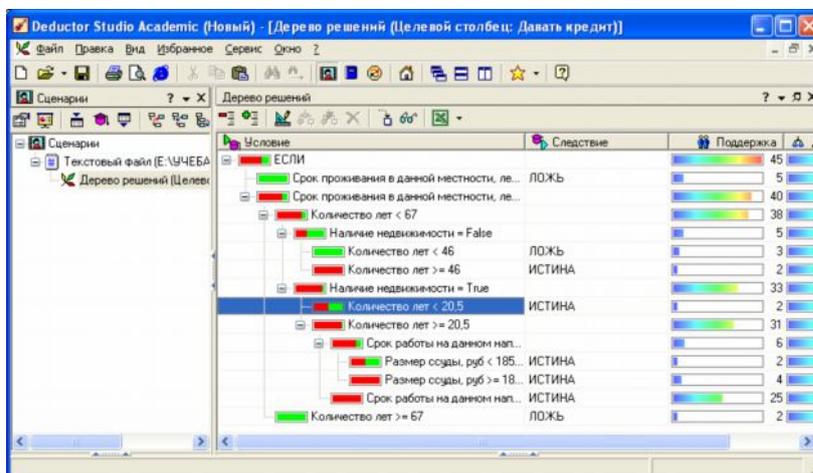


Рис. 5 Пример «Дерева Решений»

Необходимо отметить, что ни одна приобретаемая скоринговая система, как правило, не пригодна для практического использования без предварительной "настройки". Суть такой настройки состоит в том, чтобы на имеющихся у банка данных (обучающая выборка) по закрытым кредитам (с известным результатом погашения) провести настройку скоринговой системы, включающую, в частности, отбор наиболее значимых (из числа имеющихся) характеристик потенциального заемщика, для решения задач скоринга. Как показывает практика, такой набор характеристик существенно отличается не только для разных стран Западной Европы, но и для разных регионов одной страны (например, Москвы и небольших городов с численностью населения до 100 тыс. человек). Так, в ряде регионов для небольших городов одной из важнейших характеристик заемщика нередко оказывается место работы и срок работы на каком-либо градообразующем предприятии. Для крупнейших городов страны данный фактор может и не быть определяющим. Это означает, что многофилиальные банки, осуществляющие кредитование в различных регионах страны, будут вынуждены проверять настройку скоринговых систем для каждого филиала или групп филиалов. Иначе говоря, в многофилиальных банках может иметь место ситуация, когда в разных филиалах функционируют разные версии скоринговой системы. Более того, и постоянная модификация (обновление) скоринговой системы должна проводиться дифференцированно в разрезе филиалов и групп филиалов.

Таким образом, еще одним важным преимуществом метода «Деревьев решений» в кредитовании является и простота подстройки дерева под тот или иной регион (город, район и

т.п.), под изменение того или иного фактора (например наличие автомобиля сейчас отнюдь не может считаться показателем платежеспособности) т.к. оно написано на естественном языке, что не предполагает дополнительных затрат на модернизацию.

Использование системы позволяет:

- жестко регламентировать процесс принятия решения по заявке;
- сократить время ожидания ответа до 1 мин;
- повысить объективность принимаемого решения;
- управлять качеством кредитного портфеля;
- оперативно корректировать кредитные правила и этапы прохождения заявки.

Еще одним преимуществом данного метода является вариация возможностей взаимодействия Deductor с инфраструктурой Банка:

Независимое приложение

Аналитическая платформа Deductor может быть использована в качестве отдельного desktop-приложения, функционирующего непосредственно на рабочем месте специалиста, и не предусматривает никакой интеграции с инфраструктурой банка. Разработанные модели экспортируются в виде файла открытой структуры с расширением .ded, которые затем интегрируются в систему принятия решений банка для последующей работы. Разработку моделей можно начинать сразу же после инсталляции приложения на рабочем месте специалиста.

Базовая интеграция в инфраструктуру банка

Подразумевается привязка механизма экспорта (внедрения) разрабатываемых Deductor скоринговых моделей непосредственно в систему принятия решений банка.

Комплексное интегрированное решение

Комплексное интегрированное решение подразумевает работу Deductor внутри скоринговой инфраструктуры — т.е. системы кредитного скоринга.