

СОВЕРШЕНСТВОВАНИЕ ИСПОЛЬЗОВАНИЯ ТЕХНОЛОГИИ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ, КАК СПОСОБ СНИЖЕНИЯ КРЕДИТНОГО РИСКА

*Шварцкопф Н.В., старший преподаватель,
Ачинский филиал ФГБОУ ВО Красноярский ГАУ*

Аннотация: технология интеллектуального анализа данных или Data Mining позволяет существенно снизить риски при кредитовании физических лиц. Банки заинтересованы в снижении процентных ставок по кредитам. Это позволяет привлечь дополнительных клиентов, увеличить прибыль, что рассматривается в качестве основной цели коммерческой деятельности.

Риски вероятных убытков, связанных с невыполнением заемщиком условий кредитного договора, могут быть уменьшены за счет применения вероятностного подхода к оценке имущественного состояния потенциального клиента.

Ключевые слова: банковская система, банковские риски, база данных, скоринг, модель Дюрана

Один из популярных методов оценки рисков в данной сфере – это скоринг. Особенность данного подхода состоит в придании каждому фактору, отражающему риск кредитования, количественной оценки. В итоге составляется кредитный портрет заемщика, что позволяет сделать предварительные выводы относительно его возможности в будущем выполнять договорные обязательства. Следует отметить, что специфика скоринга состоит в неодинаковом удельном весе каждого фактора. Так, отдельные параметры оценки имеют различный диапазон порогов (значений). Существуют второстепенные и основные вопросы, на которые необходимо найти ответ при формировании кредитного портрета физического лица.

Международный опыт применения скоринга позволил выделить несколько эффективных форм его реализации на практике. В России популярностью пользуется модель Дюрана. Данный подход основан на выделении группы

факторов, каждый из которых в отдельности позволяет сделать вывод о допустимой степени риска кредитования.

Однако существует один системный недостаток в модели Дюрана – сложная адаптация в новых условиях рассматриваемого подхода и зависимость количественных оценок с реальным положением дел в банковской сфере. Примером может служить ситуация на рынке труда в США. Так, при систематической смене должностей конкретным человеком банки сделают вывод о его компетентности, что отразится на предлагаемых кредитных условиях. Иначе ситуация будет решена в РФ: такой человек в нашей стране будет считаться асоциальным, т.е. неспособным ужиться в коллективе, что в последствии скажется на периодичности внесения ежемесячных платежей [5].

Исходя из этого, модель Дюрана сложно приживается в новых рыночных условиях. Это также актуально для одних и тех же стран, которые переживают различные этапы

становления экономики. Иначе обстоят дела с адаптацией скоринговой модели. Специалист банковского учреждения должны разрабатывать индивидуальные стратегии оценки кредитоспособности потенциального клиента. Это предполагает высокую квалификацию работника и повышенный оклад.

Реализации на практике скоринговой модели предполагает учет специалистами текущей ситуации на рынке труда. В результате создается уникальная модель с индивидуальными весовыми коэффициентами, каждый из которых имеет пороговое значение. Преодоление этих маркеров позволяет сделать положительное заключение о кредитоспособности физического лица.

Важно учитывать, что полученный вывод – это субъективное мнение банковского служащего. Полученные результаты в минимальной степени подкреплены статистическими данными, что не позволяет использовать такие модели для отражения текущего финансового состояния конкретного заемщика.

В качестве альтернативного метода можно рассматривать Data Mining, который предоставляет ряд инструментов (деревьев решений), с помощью которых происходит решение задач классификации выбранных алгоритмов.

Важно отметить, что для успешного применения Data Mining на практике необходимо уделить особое внимание исходным данным. Для реализации методики важно выполнить ряд действий:

1) Выдвинуть гипотезу. Это предположение, от степени конкретизации которого зависит использование конкретных методов исследуемой

задачи. Особая роль здесь отведена экспертами, которые полагаются на личный опыт и знания.

2) Систематизировать имеющиеся данные. Формализация и подготовка имеющихся в распоряжении фактов, создание стройной модели доказательств и выводов – все это позволяет проводить глубокий анализ финансового положения заемщика [2].

3) Выбор оптимальной модели. Здесь также происходит её тестирование, т.е. подбор комбинаций актуальных механизмов анализа исходных данных. В рамках данных мероприятий предполагается оценка последствий совершения очередного шага, а также возвращение на предыдущий этап при невозможности объективной оценки возможных результатов.

4) Реализация модели на практике и её совершенствование. Следует отметить, что предлагаемые заемщикам анкеты – это результат сбора и систематизации огромного объема информации, связанной с созданием стройного механизма оценки финансового состояния физического лица.

Основная проблема состоит в необходимости создания такой модели оценки классификации заемщиков, при которой банковские работники, обладая достоверной классификацией рисков, могут использовать элементы методики на практике. Это позволит снизить процентную ставку по кредитам при одновременном повышении спроса населения на услуги банковских организаций.

Вышеописанная методика основывается на гипотезе, согласно которой существует ряд исчерпывающих факторов определения финансового состояния человека. Количественная оценка этих значений позволяет определить

суммарный риск, в рамках которого делается вывод об отнесении физического лица к категории добросовестных или недобросовестных заемщиков.

Автоматический анализ данных может быть реализован путем применения методики дерева

решений (Рис. 1). В результате формируется модель, где все элементы подчинены заданной иерархической структуре, а каждый нижестоящий узел определяет направление развития логики исследования.

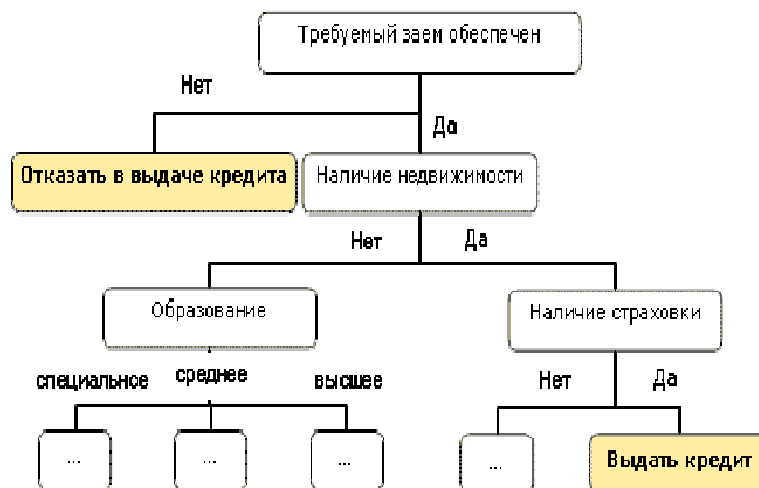


Рис. 1. Дерево решений

Данный метод основан на учете следующих категорий:

1) Происходит анализ эмпирических и статистических данных за прошлые периоды. Итогом такой работы является построение дерева решений. Важно учитывать, что для создания такой модели необходимо заранее знать класс и специфику каждой ситуации, лежащей в основе классификации рисков. Это касается получения исчерпывающей информации о том, были ли в прошлом ситуации с неуплатой кредита, какого средняя сумма займа, определяется факт наличия просрочек в ежемесячных платежах. Исходные данные сначала попадают в верхний узел дерева решений, а затем по нисходящей распределяются по другим узлам, что не исключает возможности деления их на второстепенные подгруппы. В основе критерия дифференциации лежит характеристика значения исходного фактора. Чтобы определить масштаб разбиения на группы,

используются мера неопределенности или энтропия. Для этого выбирается то поле, которое позволяет устранить наибольшее количество возникающих неопределенностей [1].

2) Создается определенный класс модели. На практике это выражается в волеизъявлении: одобрить заявку на кредит или отказать в удовлетворении запроса.

Важно учитывать специфическую особенность дерева решений – каждый элемент модели может быть изменен при существенной корректировке ситуации в банковском секторе. Адаптация модели к новым условиям происходит без существенных временных и ресурсных затрат.

Визуализация метода будет осуществляться с использованием специализированного программного обеспечения Tree Analyzer, которое представлено в качестве ключевого элемента пакета Deductor.

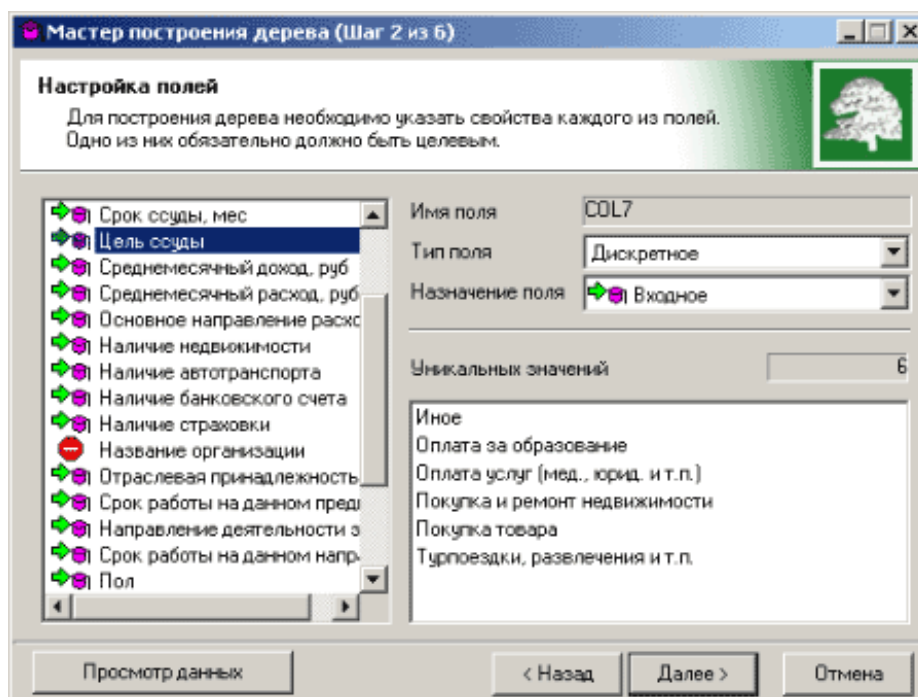


Рис. 2. Настройка определяющих и целевых факторов

Целевые факторы определяются по методологии хранилища, т.е. исходная информация заложена в наборе процессов и фактов, каждый из которых реализован по схеме «звезда». Центральное место здесь занимают факты, а измерению имеют форму лучей. Процесс предполагает оценку возможности выдачи кредита. Самой ценной информацией для банка здесь является определение статуса кредита. Если по итогам анализа будет сделан вывод об

отсутствии просрочек по выплатам, то в целом, такое физическое лицо может рассматриваться в качестве добросовестного заемщика.

При создании модели оценки кредитного статуса потенциального клиента эксперт изучает аналитическую отчетность. Исходя из того, что в хранилище она представлена в обобщенном виде (многомерная форма), то оптимальный вариант – исследование аналитических данных в табличном варианте.

Возраст	Давать кредит		
	Да	Нет	Итого
19	4 360 000,00	1 624 000,00	5 984 000,00
20	4 647 000,00	1 622 500,00	6 269 500,00
21	4 893 500,00	1 291 500,00	6 185 000,00
22	5 369 500,00	1 573 500,00	6 943 000,00
23	4 740 000,00	1 933 000,00	6 673 000,00
24	4 919 000,00	1 491 500,00	6 410 500,00
65	332 500,00	351 000,00	683 500,00
66	195 500,00	728 500,00	924 000,00
67	148 000,00	299 500,00	447 500,00
68	175 500,00	552 500,00	728 000,00
69	236 500,00	278 500,00	515 000,00
70	97 500,00	333 000,00	430 500,00
Итого	38 098 000,00	119 066 000,00	157 164 000,00

Рис. 3. Суммы, которые получили заемщики (или требовали потенциальные заемщики)

Рисунок 3. демонстрирует особенность распределения кредитов с учетом возраста потенциальных клиентов. Здесь также представлены сведения о кредитной истории физических лиц.

Детальный анализ представлен в кросс-диаграмме (Рис. 4.).

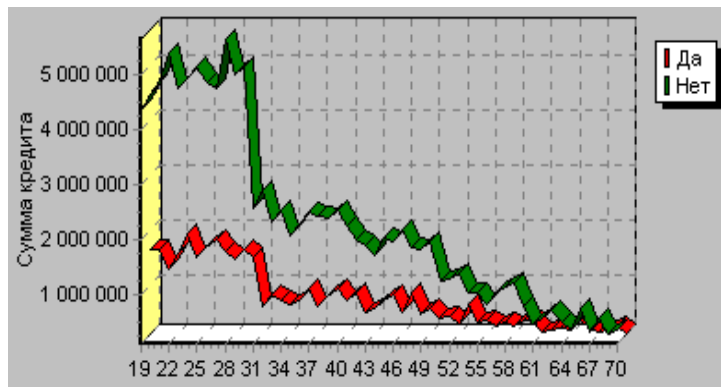


Рис. 4. Представление данных в виде кросс-диаграммы

Анализ представленных результатов позволяет сделать вывод о том, что основной круг заемщиков сводится к лицам, возраст которых до 30 лет. Физические лица в возрасте от 30 до 50 реже прибегают к услугам банкам в рамках процедуры кредитования. Для привлечения указанной группы заемщиков необходима корректировка кредитных условий. Это может быть снижение размера ежемесячного платежа. Особую роль играет и процесс оценки причин отказов от предоставления займов. Выявленные

тенденции позволяют адекватно оценить перспективы сотрудничества с новыми заемщиками.

Программа Deductor обладает расширенным функционалом в плане агрегации фактов с учетом выбранного измерения (селектор). Всегда можно оставить только те объекты, которые соответствуют заявленным условиям. Это позволяет объединения суммы кредитов с учетом возрастных особенностей физических лиц. Наглядно информация представлена на рис. 5 и 6.

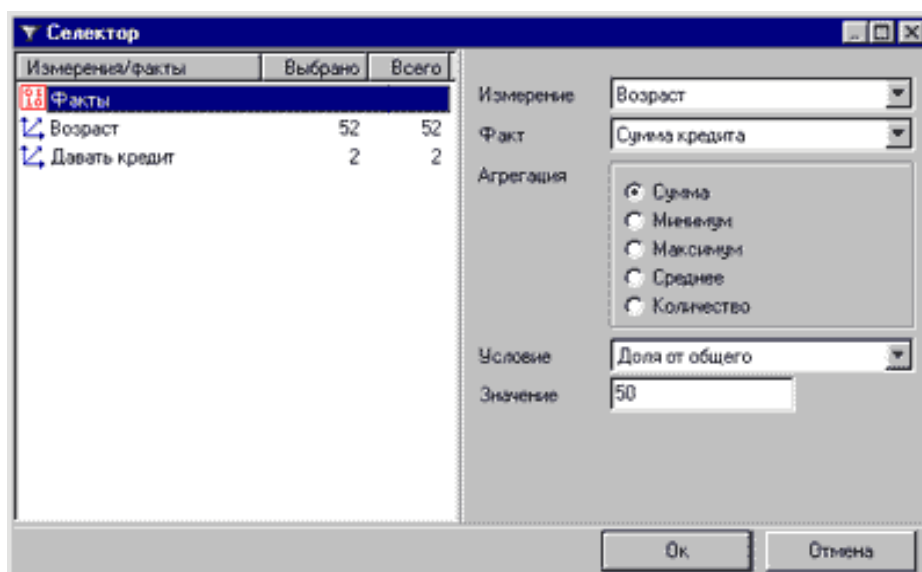


Рис. 5. Селектор

В качестве примера эффективного использования данных, представленных в аналитических отчетах, следует отметить популяризацию системы кредитования среди молодежи. С учетом специфики оказания кредитных услуг конкретный банк может самостоятельно выбирать и анализировать информацию о заемщиках, которая размещается в хранилищах. Это наглядно представлено на рис. 6.

Давать кредит			
Возраст	Да	Нет	Итого
19	1 624 000,00	4 360 000,00	5 984 000,00
20	1 622 500,00	4 647 000,00	6 269 500,00
21	1 291 500,00	4 893 500,00	6 185 000,00
22	1 573 500,00	5 369 500,00	6 943 000,00
23	1 933 000,00	4 740 000,00	6 673 000,00
24	1 491 500,00	4 919 000,00	6 410 500,00
25	1 676 500,00	5 121 500,00	6 798 000,00
26	1 811 500,00	4 855 500,00	6 667 000,00
27	1 597 500,00	4 753 000,00	6 350 500,00
28	1 446 000,00	5 633 000,00	7 079 000,00
29	1 642 000,00	4 993 500,00	6 635 500,00
30	1 474 500,00	5 118 500,00	6 593 000,00
Итого	19 184 000,00	59 404 000,00	78 588 000,00

Рис. 6. Заемщики, взявшие 50% всех кредитов

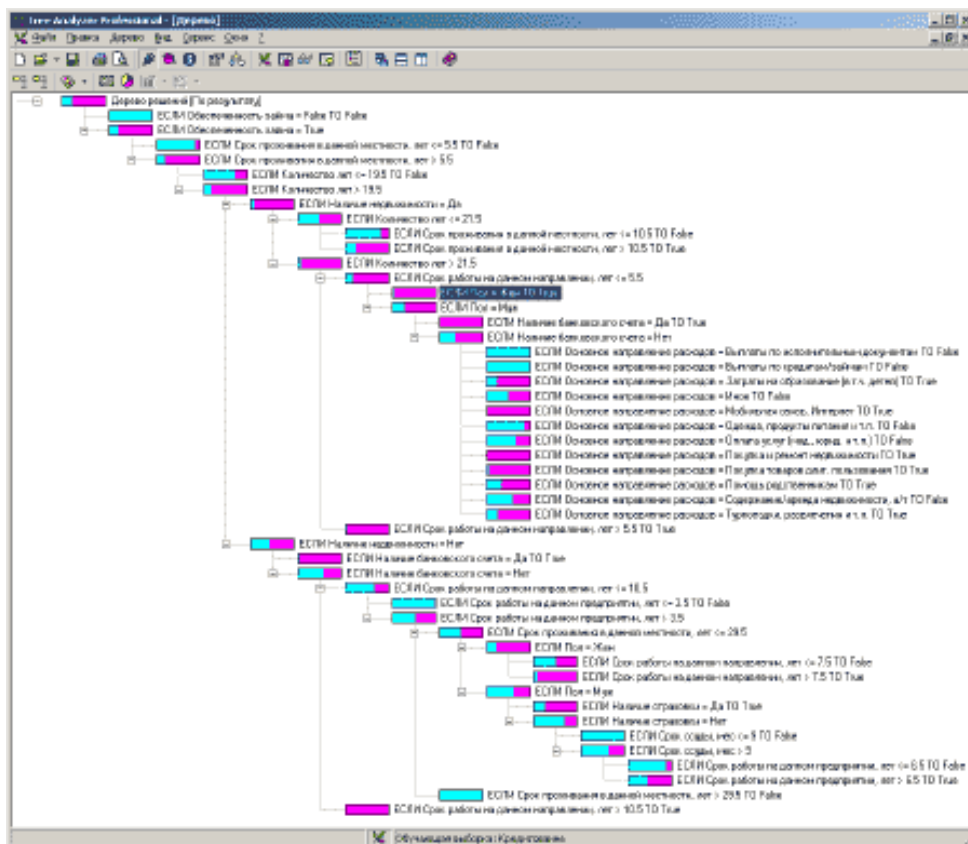


Рис. 7. Дерево решений – модель определения кредитоспособности физических лиц

Представленное дерево решений предоставляет возможность экспертам анализировать выбранные факторы влияния на финансовую устойчивость заемщика. Это становится возможным в связи с использованием на каждом иерархическом уровне критерия устранения неопределенности. Это дублируется при разделении каждого уровня на дочерние узлы. Исходя из этого, наиболее значимые для анализа факторы размещаются в корнях дерева.

Модель дерева решений строится на основе правил, которые записывают заемщика в конкретную кредитную группу с учетом индивидуальных требований. Они могут иметь форму:

1) Определения степени обеспеченности займа. Происходит построение следующей цепочки: наличие обеспеченности кредита, срок проживания в конкретном регионе – более 5.5 лет,

возраст – более 19 лет, наличие собственного недвижимого имущества, личный счет в банке = одобрение заявки на кредит.

2) Учета личных предпочтений заемщика. Допускается построение следующей цепочки умозаключений: заём обеспечен, срок проживания в регионе составляет более 5 лет, есть недвижимость, возраст – более 21 года, основная статья расходов заемщика – одежда и продукты питания = отказ в одобрении заявки.

Если воспользоваться диалоговым окном в Deductor «Эксперимент», то будет предложен тест. В зависимости от выбранных вариантов ответов система самостоятельно примет решение о том, выдавать ли кредит или стоит отказать в одобрении заявки.

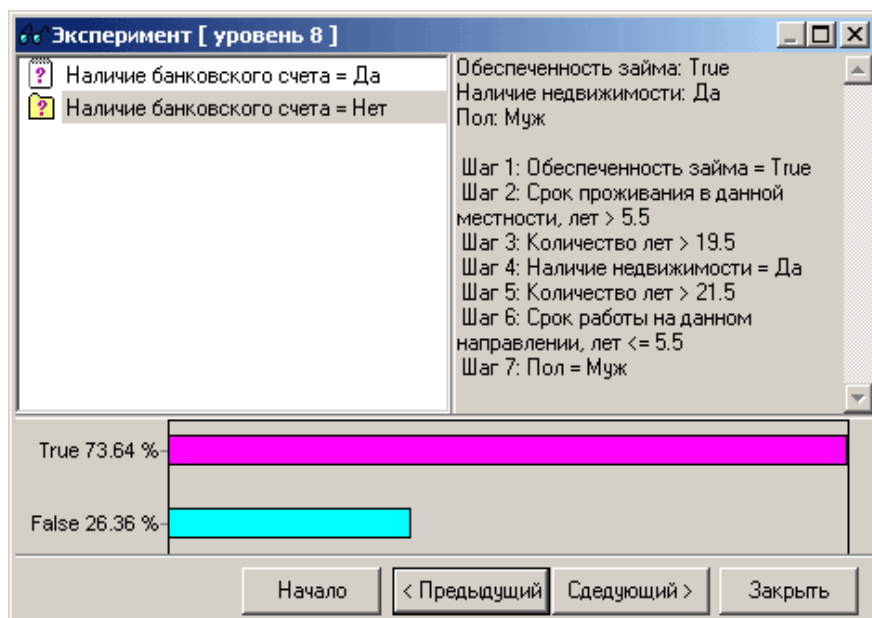


Рис. 8. Окно «Эксперимент»

Такой подход одновременно позволяет устранить известные недостатки скоринговой системы:

1) Минимизируется стоимость адаптации модели к новым условиям применения, т.к. дерево решений представляет собой самоадаптируемый

комплекс, где роль человека сводится к минимуму.

2) Эффективность полученного результата достаточно велика, что обусловлено анализом наиболее значимых факторов конечного ответа.

Ключевые преимущества модели дерева решений:

1) возможность её интеграции со сторонними программами и ресурсами.

2) независимый сбор информации в хранилище для последующего анализа рисков.

3) разнообразие инструментов определения факторов финансового состояния заемщика.

Технология интеллектуального анализа данных или Data Mining позволяет существенно снизить

риски при кредитовании физических лиц. Банки заинтересованы в снижении процентных ставок по кредитам. Это позволяет привлечь дополнительных клиентов, увеличить прибыль, что рассматривается в качестве основной цели коммерческой деятельности.

Риски вероятных убытков, связанных с невыполнением заемщиком условий кредитного договора, могут быть уменьшены за счет применения вероятностного подхода к оценке имущественного состояния потенциального клиента.

Литература

1. Анализ банковских рисков. Система оценки корпоративного управления и управления финансовым риском: пер.с англ.; вступ. сл. д.э.н. К.Р. Тагирбекова. М.: Издательство "Весь Мир", 2014. 304с.

2. Белоглазова Г.Н., Кроливецкая Л.П. Банковское дело. Организация деятельности коммерческого банка. М.: Высшее образование, 2014.

3. Дробозина Л.А., Можайсков О.В. Финансовая и денежно-кредитная система Англии. М.: Финансы, 2014.

4. Ефимова Л.Г. Банковское право. М.: «БЕК», 2014.

5. Жарковская Е.П., Банковское дело, Омега-Л, Москва, 2013.

Жарковская Е.П., Финансовый анализ деятельности коммерческого банка, Омега-Л, Москва, 2015.

References

1. Analiz bankovskih riskov. Sistema ocenki korporativnogo upravleniya i upravleniya finansovym riskom: per.s angl.; vstup. sl. d.eh.n. K.R. Tagirbekova. M.: Izdatel'stvo "Ves' Mir", 2014. 304s.

2. Beloglazova G.N., Kroliveckaya L.P. Bankovskoe delo. Organizaciya deyatel'nosti kommercheskogo banka. M.: Vysshee obrazovanie, 2014.

3. Drobozina L.A., Mozhajskov O.V. Finansovaya i denezhno-kreditnaya sistema Anglii. M.: Finansy, 2014.

4. Efimova L.G. Bankovskoe pravo. M.: «BEK», 2014.

5. Zharkovskaya E.P., Bankovskoe delo, Omega-L, Moskva, 2013.

Zharkovskaya E.P., Finansovyy analiz deyatel'nosti kommercheskogo banka, Omega-L, Moskva, 2015.

IMPROVING THE USE OF DATA MINING TECHNOLOGY AS A WAY OF REDUCING CREDIT RISK

*Schwarzkopf N.V., Senior Lecturer,
Krasnoyarsk State Agrarian University in Achinsk*

Abstract: data mining technology or Data Mining can significantly reduce risks when lending to individuals. Banks are interested in lowering interest rates on loans. This allows to attract additional customers, increase profits, which is considered as the main goal of commercial activity.

The risks of probable losses associated with the borrower's failure to comply with the terms of the loan agreement can be reduced by applying a probabilistic approach to assessing the property status of a potential client.

Keywords: banking system, banking risks, database, scoring, Duran model