

УДК 004.891

# ОСОБЕННОСТИ ПРИМЕНЕНИЯ МЕТОДОВ DATA MINING В СКОРИНГОВЫХ РЕШЕНИЯХ ДЛЯ КОММЕРЧЕСКИХ БАНКОВ

**Дьяков О.А.,**

магистр кафедры бизнес-информатики,  
Финансовый университет, Москва, Россия  
olegdyakov93@gmail.com

**Аннотация.** В контексте возрастающей потребности в автоматизации и повышении качества процесса оценки кредитоспособности заемщика и его дальнейшего поведения скоринговая модель, построенная с применением технологий Data Mining, является хорошим технологическим подспорьем. Скоринговое решение позволяет сделать процесс оценки кредитоспособности и мониторинга кредитной заявки более прозрачным и контролируемым. Рабочий цикл Data Mining, представляющий собой спиральную модель, хорошо укладывается в границы разработки скоринговой модели и позволяет использовать при ее создании такие методы, как: логистическая регрессия, дискриминантный линейный анализ, нейронные сети и деревья решений. Немаловажным является тот факт, что указанные методы представляют возможность самообучения и развития скоринговой модели на основе статистических и исторических данных. Тем не менее значимость и степень применения тех или иных средств интеллектуального анализа данных определяется руководством организации на основе принятой скоринговой стратегии. В конечном итоге банк получает возможность быстро и эффективно решать широкий спектр задач скоринга в условиях динамической бизнес-среды.

В рамках статьи рассмотрены вопросы значимости технологии кредитного скоринга для финансовой организации, особенности принятия решений в рамках различных видов скоринга, а также особенности и целесообразность применения методов интеллектуального анализа данных в скоринговых моделях.

**Ключевые слова:** кредитный риск; скоринг; банкинг; интеллектуальный анализ данных; методы data mining.

## THE SPECIFIC FEATURES OF USING DATA MINING TECHNIQUES IN SCORING SOLUTIONS FOR COMMERCIAL BANKS

**Dyakov O.A.,**

Master's degree in of Business Informatics,  
Financial university, Moscow, Russia  
olegdyakov93@gmail.com

**Abstract.** Credit scoring technology provides a fast and objective assessment of the risk level for preapproved loans and some recommendations how to work with such loans. Due to the fact of growing need to automate and improve the quality of the evaluation of borrower creditworthiness and to predict his financial behavior, a

---

Научный руководитель: **Громова А.А.**, кандидат экономических наук, доцент кафедры бизнес-информатики.

*scoring model based on Data Mining technology could be an effective solution for such a question. The scoring solution makes a process of credit assessment and monitoring of credit applications more transparent and well controlled at all stages of the loan life cycle: from submitting credit application to the full credit repayment or collection of receivables. Having a spiral model operating cycle of Data Mining technology fits well into the borders of the development of scoring models and the fact of creating them allows to use different methods such as logistic regression, discriminant linear analysis, neural networks and decision trees. It is also important that these methods provide an opportunity for scoring model to learn by itself and develop, being based on statistical and historical data. Nevertheless, the significance of certain data mining methods and their extent of use are based on current scoring strategy and determined by the top-management. Eventually banks will be able to solve a wide range of specified scoring business tasks in a dynamic business environment quickly and efficiently.*

*The scope of the article includes the following issues: the importance of credit scoring technology for financial institutions, the specific features of decision-making for various types of scoring, as well as some features and the expediency of using data mining methods in credit scoring models.*

**Keywords:** credit risk; scoring; banking; data mining; scoring model.

**В** настоящий момент услуга предоставления кредитной организацией займов физическим лицам пользуется большой популярностью среди населения страны. По результатам исследования «Объединенного кредитного бюро» (ОКБ) [1] практически 60% активного населения России имеет непогашенные кредитные обязательства, при этом каждый заемщик в среднем должен банку порядка 210 тыс. руб.

В то же время в банковской сфере на фоне общей сложной экономической ситуации, острой конкурентной борьбы и ужесточившегося надзора со стороны главного финансового регулятора отлаженная работа основного направления, подразумевающего кредитование физических лиц и индивидуальных предпринимателей, особенно важна. Стоит отметить, что фундаментальным звеном в данном сквозном бизнес-процессе является оценка кредитоспособности потенциальных заемщиков. Здесь имеет место проявление кредитного риска, управление которым зачастую определяет эффективность деятельности банка, а также обеспечивает его финансовую устойчивость. Понятие кредитного риска представлено в письме Банка России «О типичных банковских рисках»: кредитный риск — риск возникновения убытков вследствие неисполнения, несвоевременного либо неполного исполнения должником финансовых обязательств перед кредитной организацией в соответствии с условием договора\*.

Принимая во внимание условия Второго Базельского соглашения (Базель II) [2], предписывающего

банкам при определении величины кредитного риска использование либо стандартизированного подхода, либо базового внутреннего рейтинга, основанного на собственных разработках, применение скоринговой системы может оказаться хорошим подспорьем в решении данного вопроса.

Технологическое решение задачи количественной оценки кредитных рисков и разработки методов расчета резервов лежит в области формализованного применения инноваций в области информационных технологий. В данной сфере основной уклон необходимо сделать в сторону интеллектуальных автоматизированных и полуавтоматизированных систем класса Business Intelligence (BI), а также технологий Data Mining, представляющих мощный аппарат современной бизнес-аналитики и исследования данных для обнаружения скрытых закономерностей и построения предсказательных моделей [3]. В этой связи оправдано применение финансовыми организациями скоринговой модели, построенной на основе методов Data Mining — инструмента, включающего в себя математические и программные методы интеллектуальной обработки исторических и статистических данных для принятия кредитных решений, а также являющегося элементом кредитной стратегии розничного банка.

Такая модель позволяет перевести гетерогенную информацию из разряда характеристик заемщика в категорию специфических значений и сформировать некоторый интегральный показатель, указывающий на степень кредитного риска, связанного с этим заемщиком.

Необходимо отметить, что кредитование юридических лиц и организаций в банках чаще всего

\* Письмо Центрального банка Российской Федерации от 23.06.2004 № 70-Т.

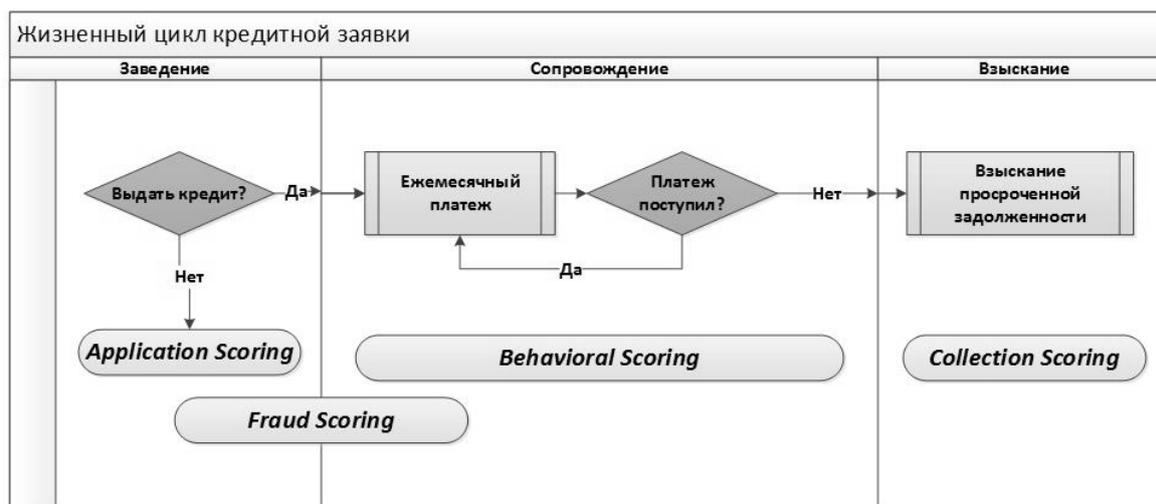


Рис. 1. Этапы скоринга на жизненном цикле заявки

подразумевает индивидуальный подход к каждой отдельной сделке. Условия таких сделок согласуются коллегиально на кредитных комитетах и предусматривают возможность отклонения от стандартных условий кредитования. Таким образом, применение скоринговой модели имеет место главным образом в процессе кредитования физических лиц, особенно в части потребительского кредитования при необеспеченных ссудах. Далее речь пойдет об оценке кредитного риска заемщиков — физических лиц.

Основные задачи практически любой скоринговой модели в банковской сфере включают в себя следующие:

- Увеличение объемов кредитного портфеля за счет снижения доли необоснованных отказов по кредитным сделкам.
- Ускорение процесса оценки кредитоспособности потенциального заемщика при максимальном качестве и точности оценки.
- Снижение общего уровня невозврата заемных средств.
- Использование в качестве инструмента централизованного накопления данных о клиенте.
- Снижение резерва на возможные потери по ссудам (РВПС) по портфельным и непортфельным кредитам.
- Оценка динамики изменений ссудного счета клиента как в разрезе конкретной сделки, так и всего кредитного портфеля в целом.

Скоринговая модель подразумевает осуществление оценки кредитного риска с применением отдельных методов стандартизированного подхода. Такой подход основан на расчете кредитными организациями количественных и качественных оценок

рисков по результатам проведения комплексного анализа деятельности заемщика. Основные элементы такого анализа чаще всего включают в себя информацию о текущем финансовом положении заемщика, качестве обслуживания долга по уже имеющимся ссудам, а также находящуюся в распоряжении кредитной организации информацию о любых рисках заемщика, включая сведения о его внешних обязательствах, полученные из бюро кредитных историй [4].

Каждый из этапов жизненного цикла кредитной заявки и в последующем кредитного договора имеет свои характерные особенности в рамках принятия того или иного решения, а также свои методологии расчета оценок [5]. Так, в зависимости от того, на каком этапе находится кредитная заявка или договор, системой может применяться соответствующий вид скоринга. Если условно разделить жизненный цикл кредитной заявки на этапы «Заведение», «Сопровождение» и «Взыскание», то привязку к скоринговым решениям можно представить, как показано на рис. 1.

На этапе заведения сделки применяется скоринг заявок и обращений по кредитным продуктам (Application Scoring), задачей которого является оказание помощи андеррайтерам и кредитным инспекторам при принятии решения о выдаче кредита на основе имеющейся информации о заявителе. Параллельно с этим на данном этапе жизненного цикла сделки используется скоринг мошенничества (Fraud Scoring) для статистической оценки вероятности мошеннических действий со стороны потенциального заемщика. Учитывая, что Fraud Scoring позволяет риск-аналитикам выявлять зависимости и согласованность действий по множеству кли-

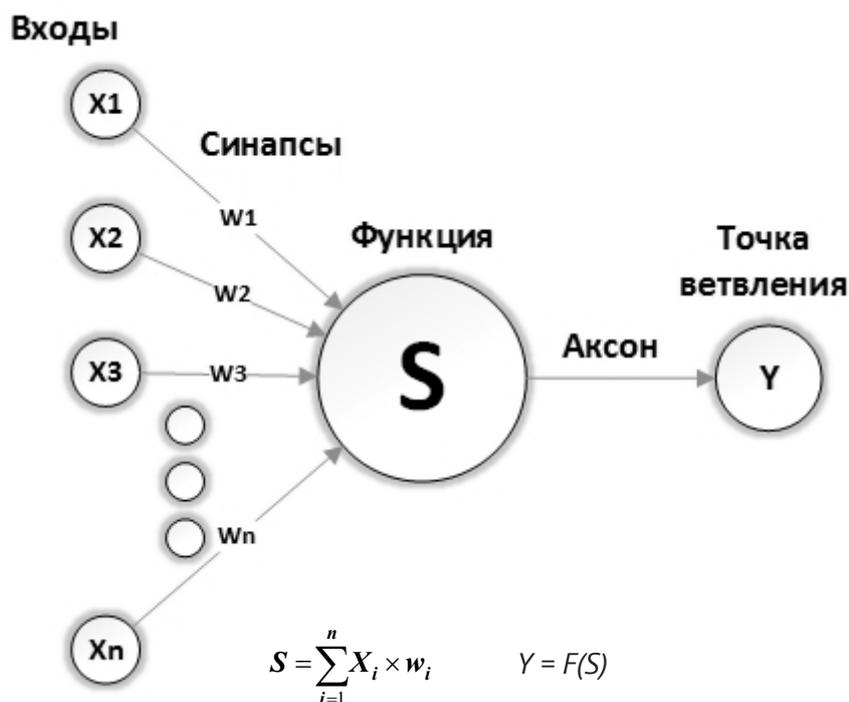


Рис. 2. Искусственный нейрон в нейронной сети

ентов, его применение целесообразно и на этапе сопровождения кредитной сделки. Также на этапе сопровождения задействован поведенческий скоринг (Behavioral Scoring), главной целью которого является анализ всей истории взаимоотношений с действующими клиентами для прогнозирования изменения платежеспособности заемщика, т.е. возможности отвечать по долговым обязательствам на текущий момент времени в целях корректировки установленных кредитных параметров либо формирования индивидуальных предложений. В случае если сделка переходит на этап «Взыскание», для эффективного управления дебиторской задолженностью кредитными организациями часто применяется система скоринга просроченных задолженностей (Collection Scoring).

Скоринговые системы, использующие различные виды скоринга, позволяют кредитной организации снизить издержки и минимизировать операционный риск за счет автоматизации принятия решения, сокращают время обработки заявок на предоставление кредита, дают возможность банкам проводить свою кредитную политику централизованно, обеспечивают дополнительную защиту от мошенничества. При этом необходимо учитывать, что банк имеет возможность самостоятельно определять степень и периодичность использования указанных видов скоринга для каждого конкретного направления кредитования, корректируя свою скоринговую модель.

Возвращаясь к вопросу о технологической реализации скоринговой системы, обеспечить высокую степень автоматизации процессов скоринга с высокой степенью эффективности во всей цепочке работы, начиная со сбора данных и заканчивая получением итоговых баллов и предложениям по стратегии, возможно с помощью методов интеллектуальной обработки данных Data Mining.

Согласно определению компании SAS Institute [6] Data Mining представляет собой процесс выявления скрытых закономерностей, обнаружения в необработанных данных ранее неизвестных, нетривиальных знаний, простых для интерпретации и практически полезных в принятии решений во всех областях человеческой деятельности.

Отличительной особенностью технологии Data Mining является возможность выявлять среди больших массивов данных закономерности, которые не могут быть обнаружены стандартными тривиальными способами обработки информации, но при этом являются практически полезными и объективными. При этом методы Data Mining реализованы на базе различных научных дисциплин, таких как теория вероятности и математическая статистика, теория баз данных, искусственный интеллект, алгоритмизация, визуализация и другие.

В качестве основных методов Data Mining, которые технологически возможно применить при построении скоринговой модели в кредитной ор-

ганизации, можно определить следующие: модель логистической регрессии, дискриминантный линейный анализ, нейронные сети и деревья решений. При этом при выборе того или иного метода необходимо в первую очередь руководствоваться видом скоринга, для которого будет разрабатываться скоринговая модель. Это связано с тем, что каждый из 4 видов скоринга, описанных выше, преследует свою цель при интерпретации результатов, имея при этом различные перечни значимых входных данных и стратегии их обработки.

Нейронные сети представляют собой самообучающиеся статистические модели, построенные по принципу функционирования и организации биологических сетей нервных клеток в живом организме. Каждый нейрон в такой модели — это обрабатывающий элемент, связанный с другими элементами и внешней средой с помощью входных связей, определяемых весовыми коэффициентами (синапсами), и выходных связей (аксонами). Стандартными составными элементами нейрона также являются следующие: входной сумматор связей, агрегирующий силу входных импульсов нейрона, нелинейный преобразователь — основная расчетная функция (ядро нейрона) и точка ветвления, выдающая итоговый результат выполнения функции (рис. 2).

Применительно к скоринговой модели уровень входящих связей можно представить как набор характеристик, используемых для определения платежеспособности клиента. Точка ветвления представляет собой ожидаемый результат — оценку платежеспособности клиента. Входящими и исходящими данными такой модели могут быть как количественные, так и качественные показатели, переведенные в фиктивные переменные. Другими словами, нейронная сеть — это «черный ящик», где разработчик-аналитик с помощью статистических программ задает имеющиеся независимые переменные и ожидаемый результат, а на выходе получает подсчитанные системой коэффициенты модели (скоринговые веса).

Деревья решений — это модель, строящаяся на логической цепочке правил, которые пытаются описать отдельные взаимосвязи между данными относительно ожидаемого результата. Для принятия решения, к какому классу отнести некоторый объект или ситуацию, требуется ответить на вопросы, стоящие в узлах этого дерева, начиная с его корня. При этом структура деревьев решений открыто по-казывает аргументацию правил и поэтому позволяет

легко интерпретировать процесс принятия решения на том или ином узле. Фрагмент дерева решений для скоринга заявок и обращений по кредиту приведен на рис. 3.

Следующий метод интеллектуального анализа данных — множественная линейная регрессия — связывает поведение зависимой переменной (в случае скоринга заявок и обращений — платежеспособность клиента) с линейной функцией определенного набора независимых переменных (характеристиками заемщика). Такая функция позволяет определить наилучшую линейную зависимость путем минимизации суммы стандартных отклонений и имеет следующий вид:

$$y_i = \alpha + \chi_1 \times x_{i1} + \beta_2 \times x_{i2} + \dots + \beta_m \times x_{im},$$

где  $\alpha$  — свободный член уравнения;  $y_i$  — зависимая переменная (платежеспособный/неплатежеспособный клиент);  $x_i$  — независимые переменные (характеристики заемщика);  $\beta$  — параметры модели (скоринговые веса). Полученные в результате построения модели  $\beta$  значения — это параметры, которые определяют характер связи между наблюдаемым значением зависимой переменной «оценка кредитоспособности клиента» и соответствующими скоринговыми характеристиками.

Еще одним методом интеллектуального анализа данных при построении скоринговых моделей является дискриминантный линейный анализ, содержанием которого является выявление решения задач различения (дискриминации) объектов наблюдения по заданным признакам. Применительно к скорингу объектами наблюдения являются собственно заемщики, а признаки — характеристики этих заемщиков (независимые переменные, или предикторы). Процедуры дискриминантного анализа можно условно разделить на две группы. Первая группа процедур предназначена для интерпретации различия между существующими группами, вторая — для проведения классификации новых объектов в тех случаях, когда неизвестно заранее, к какой из существующих групп они относятся.

В ситуации применения метода дискриминантного анализа для скоринга заявок и обращений (Application Scoring) скоринговая модель может полностью дублировать модель множественной регрессии для того же вида скоринга. Это связано с тем, что в этом случае дискриминация применяется только к двум группам клиентов: кредитоспособным

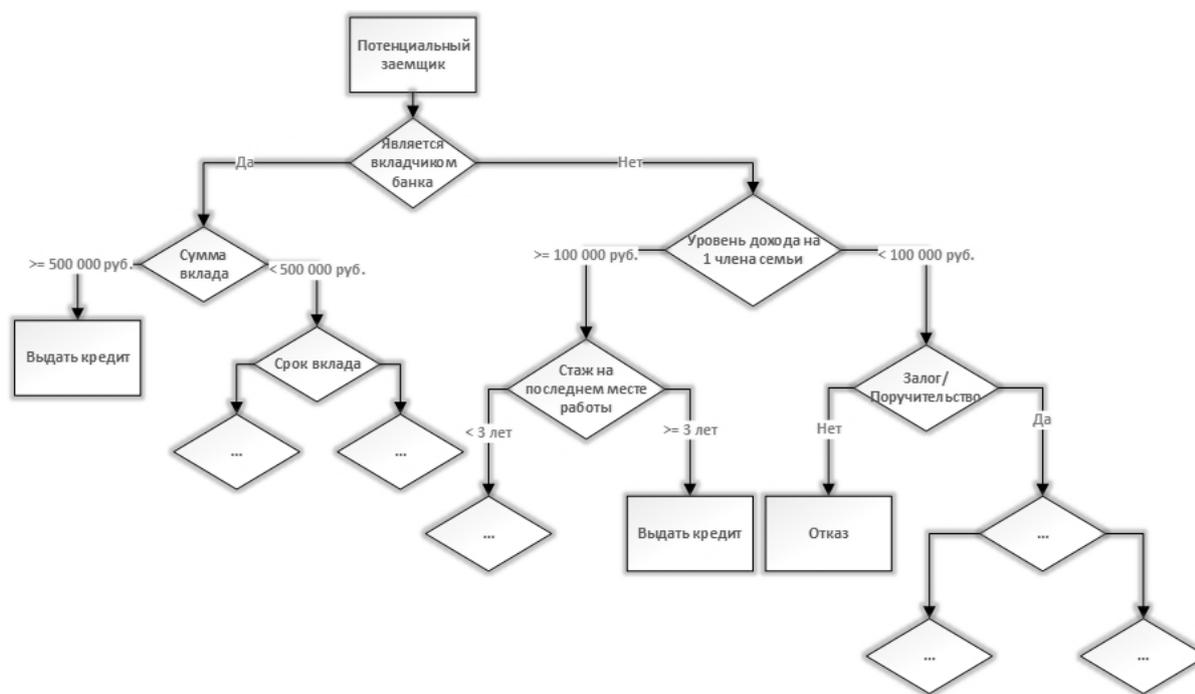


Рис. 3. Фрагмент дерева решений для скоринга заявок по кредиту

и некредитоспособным заемщикам. Таким образом, используя эти группы в качестве зависимых переменных в множественной регрессии, будут получены результаты, аналогичные тем, которые могли бы быть получены с помощью дискриминантного анализа. Другими словами, в случае дискриминации двух групп составляется линейное уравнение следующего типа:

$$Группа(1,2) = a + b_1 \times x_1 + b_2 \times x_2 + \dots + b_m \times x_m,$$

где  $a$  – константа;  $b_1 \dots b_m$  – коэффициенты регрессии.

Интерпретация результатов задачи с двумя группами тесно следует логике использования множественной регрессии: здесь переменные с наибольшими значениями регрессионных коэффициентов приносят наибольший вклад в дискриминацию групп.

Обращаясь к результатам зарубежных исследований в области статистики и анализа данных [7, 8], следует отметить, что алгоритм бинарного дерева решений имеет более высокую степень кластеризации заемщиков по доходному признаку по сравнению с методами логистической регрессии, дискриминантного линейного анализа и нейронных сетей. С другой стороны, выявлено, что метод нейронных сетей имеет наименьший (причем много меньше остальных методов) процент ошибок 2-го рода, подразумевающих, что некредитоспособный

заемщик будет квалифицирован скоринговой системой как кредитоспособный. При этом необходимо учитывать, что ошибки 2-го рода являются наиболее критическими с точки зрения кредитного риска, а ошибки 1-го рода (кредитоспособный заемщик квалифицируется скоринговой системой как некредитоспособный) характеризуют нереализованные банковские продукты и упущенные рыночные возможности.

Тем не менее необходимо учитывать, что степень эффективности применимости того или иного метода или их совокупности варьируются в зависимости от направлений скоринговой стратегии в том или ином банке.

Говоря о целесообразности применения методов Data Mining при построении скоринговой модели, необходимо также отметить возможность тесной увязки этапов жизненного цикла скоринговой системы, рассматриваемой как в качестве самостоятельного программного продукта, и рабочего цикла Data Mining [9] с точки зрения его использования как технологического фундамента для информационной системы. Оба цикла представляют собой набор повторяющихся функциональных шагов, при этом каждый из 11 основных шагов разработки скоринговой модели можно однозначно соотнести с этапами разработки модели Data Mining (рис. 4).

Таким образом, в контексте возрастающей потребности в автоматизации и повышении качества



Рис. 4. Соотнесение рабочего цикла Data Mining с последовательностью разработки скоринговой модели

процесса оценки кредитоспособности заемщика и его дальнейшего поведения скоринговая модель, построенная с применением технологии Data Mining, может стать универсальным и эффективным инструментом, который может быть тиражирован в систему потребительского кредитования многих финансовых организаций.

Стоит отметить, что система скоринга – это динамическая среда, которая требует своевременных доработок и планового сопровождения. Учитывая, что скоринг строится на статистических данных прошлого, система может реагировать на изменения

социально-экономической ситуации с некоторым временным лагом. Однако методы Data Mining при построении скоринговых моделей предполагают использование генерации гипотез о закономерностях, имеющих место в анализируемых данных, и построение моделей, позволяющих количественно оценить уровень корреляции исследуемых факторов, основываясь на имеющейся информации. Такой подход в некоторой степени дает возможность превратить скоринговую модель в самообучающуюся систему и тем самым максимально нивелировать отставание от текущей действительности.

### Список источников

1. Аналитическая Русская служба BBC. У 59% занятого населения России есть непогашенные кредиты [Электронный ресурс] // <http://www.bbc.com/russian/>, 06.04.2016. URL: [http://www.bbc.com/russian/news/2016/04/160406\\_russia\\_bank\\_loans](http://www.bbc.com/russian/news/2016/04/160406_russia_bank_loans) (дата обращения: 19.11.2016).
2. Финансовая газета. Что такое «Базельские соглашения»? [Электронный ресурс] // <http://fingazeta.ru/>, 24.01.2012. URL: <http://fingazeta.ru/how-much/173977/> (дата обращения: 25.09.2016).
3. Дьяков О.А., Солянов К.С. Формирование кредитного конвейера банка на основе систем бизнес-аналитики. Стратегии бизнеса. 2016, (7):2. DOI:10.17747/2311-7184-2016-7-2.
4. Саитова С.Т. Использование скоринговой модели при управлении кредитным риском // Молодой ученый. 2013. № 12. С. 342–344.
5. Коптелов А. Кредитный скоринг: реальные возможности [Электронный ресурс] // <http://www.cnews.ru/>, 04.09.2013. URL: [http://www.cnews.ru/articles/kreditnyy\\_skoring\\_realnye\\_vozmozhnosti](http://www.cnews.ru/articles/kreditnyy_skoring_realnye_vozmozhnosti) (дата обращения: 22.11.2016).
6. SAS Institute. Data Mining. What it is and it matters? [Электронный ресурс] // <http://www.sas.com/>. URL: [http://www.sas.com/en\\_us/insights/analytics/data-mining](http://www.sas.com/en_us/insights/analytics/data-mining) (дата обращения: 17.11.2016).
7. Huseyin Ince, Bora Aktan. A comparison of Data Mining techniques for credit scoring in banking: a managerial perspective // Journal of Business Economics and Management, 2010, Vol. 10, No. 3.
8. Evaristus Didik Madyatmadja, Mediana Aryuni. Comparative study of Data Mining model for credit card application scoring in bank // Journal of Theoretical and Applied Information Technology, 2014, Vol. 59, No. 2.
9. ИНТУИТ, Национальный открытый университет. Лекция 28: Data Mining консалтинг [Электронный ресурс] // <http://www.intuit.ru/>. URL: <http://www.intuit.ru/studies/courses/6/6/lecture/212?page=1> (дата обращения: 16.11.2016).