



## Методологические аспекты разработки скоринговых моделей кредитного риска в банковском секторе

Мария Владимировна Чернова<sup>1</sup>, Элиза Юрьевна Захариadis<sup>2</sup>

<sup>1,2</sup>Российская академия народного хозяйства и государственной службы при Президенте Российской Федерации Москва, Россия

<sup>1</sup>chernova-mv@ranepa.ru, <https://orcid.org/0009-0001-7904-8773>

<sup>2</sup>zakhariadis-ey@ranepa.ru, <https://orcid.org/0009-0005-6588-7540>

### Аннотация

**Введение.** Современный банковский сектор функционирует в условиях высокой конкуренции, роста объемов кредитования и увеличения требований к управлению рисками. В этих условиях особую значимость приобретает проблема оценки кредитоспособности заемщиков и минимизация риска невозврата кредита. Развитие цифровых технологий и увеличение объемов данных обуславливает необходимость совершенствования аналитических инструментов, которые используются в кредитном скоринге. Традиционная модель логистической регрессии, обладающая высокой интерпретируемостью, конкурирует с «черными ящиками» машинного обучения, обеспечивающими более высокую точность прогнозирования. Это формирует актуальную научную задачу выбора оптимальной методологии оценки кредитного риска.

**Целью** исследования является анализ теоретико-методологических основ построения скоринговых моделей на базе логистической регрессии, а также определение факторов, влияющих на качество оценки кредитного риска и эффективность применения данных моделей в банковском секторе.

**Материалы и методы.** Методологическую основу исследования составляют статистический, эконометрический и кибернетический подходы теорий кредитного скоринга, риск-менеджмента и машинного обучения. В работе использованы методы сравнительного анализа, обобщения и систематизации научных подходов к построению скоринговых моделей. В качестве основного инструмента анализа рассматривается логистическая регрессия для прогнозирования вероятности невозврата кредита, а также учтены современные подходы, основанные на алгоритмах машинного обучения.

**Результаты.** Прежде всего определены ключевые этапы разработки скоринговых моделей, включая подготовку данных, отбор признаков и оценку параметров. Установлено, что логистическая регрессия сохраняет значительную практическую ценность благодаря своей интерпретируемости и соответствию регуляторным требованиям. Выявлены основные ограничения логит-моделей, такие как чувствительность к качеству данных и ограниченная способность учитывать сложные нелинейные зависимости.

**Выводы.** Скоринговые модели являются ключевым инструментом управления кредитным риском в банковской деятельности. Логистическая регрессия, несмотря на развитие более сложных алгоритмов, остается востребованной благодаря прозрачности и надежности результатов. Повышение эффективности скоринговых моделей требует комплексного учета качества данных, корректного выбора признаков и адаптации моделей к изменяющимся условиям внешней среды. Ядром логистической регрессии является метод максимального правдоподобия. В рамках эконометрики функцию максимального правдоподобия максимизируют, в машинном обучении минимизируют обратную функцию потерь. Перспективным направлением является комбинированное применение логистической регрессии с более точными методами.

**Ключевые слова:** банковский скоринг, кредитный скоринг, скоринговая модель, логистическая регрессия, оценка кредитоспособности, сигмоида, логит-модель, риск-менеджмент, машинное обучение, метод максимального правдоподобия, градиентный метод спуска

**Для цитирования:** Чернова М. В., Захариадис Э. Ю. Методологические аспекты разработки скоринговых моделей кредитного риска в банковском секторе // Государственное и муниципальное управление. Ученые записки. 2026. № 2. С. 96–110. EDN RFUFNW

Original article

## Methodological aspects of the development of credit risk scoring models in the banking sector

Maria V. Chernova<sup>1</sup>, Eliza Yu. Zakhariadis<sup>2</sup>

<sup>1,2</sup>Russian Presidential Academy of National Economy and Public Administration, Moscow, Russia

<sup>1</sup>chernova-mv@ranepa.ru, <https://orcid.org/0009-0001-7904-8773>

<sup>2</sup>zakhariadis-ey@ranepa.ru, <https://orcid.org/0009-0005-6588-7540>

### Abstract

**Introduction.** The modern banking sector operates in a highly competitive environment, with increasing lending volumes and risk management requirements. In this context, assessing borrowers' creditworthiness and minimizing the risk of loan default become crucial. The development of digital technologies and the increasing volume of data necessitates the improvement of analytical tools used in credit scoring. The traditional logistic regression model, with its high level of interpretability, competes with the "black boxes" of machine learning, which offer higher prediction accuracy. This creates an urgent scientific challenge of choosing the optimal methodology for assessing credit risk.

**Purpose.** The purpose of the study is to analyze the theoretical and methodological foundations of building scoring models based on logistic regression, as well as to identify the factors that affect the quality of credit risk assessment and the effectiveness of using these models in the banking sector.

**Materials and methods.** The study is based on the works of authors in the field of credit scoring, risk management, and machine learning. The study uses methods of comparative analysis, generalization, and systematization of scientific approaches to building scoring models. Logistic regression is considered as the main analysis tool for predicting the probability of loan default. Modern approaches based on machine learning algorithms are also taken into account.

**Results.** The study identified key stages in the development of scoring models, including data preparation, feature selection, and parameter estimation. It was found that logistic regression retains significant practical value due to its interpretability and compliance with regulatory requirements. The study also identified key limitations of logistic models, such as their sensitivity to data quality and limited ability to account for complex nonlinear relationships.

**Conclusions.** Scoring models are a key tool for managing credit risk in banking. Despite the development of more complex algorithms, logistic regression remains popular due to its transparency and reliable results. Improving the efficiency of scoring models requires a comprehensive approach to data quality, proper feature selection, and model adaptation to changing environmental conditions. The core of logistic regression is the maximum likelihood method. In econometrics, the maximum likelihood function is maximized, while in machine learning, the inverse loss function is minimized. A promising area is the combined use of logistic regression with more accurate methods.

**Keywords:** banking scoring, credit scoring, scoring model, logistic regression, creditworthiness assessment, sigmoid, logit model, risk management, machine learning, maximum likelihood method, gradient descent method

**For citation:** Chernova M.V., Zakhariadis E.Yu. Methodological aspects of the development and validation of credit risk scoring models in the banking sector. *State and Municipal Management. Scholar Notes*. 2026;(2):96–110. (In Russ.). EDN RFUFNW

## Введение

В современных условиях банковский сектор использует актуальные механизмы, позволяющие управлять кредитными рисками. В этой связи, на фоне усиления конкурентной борьбы и увеличения кредитных операций особое внимание уделяется вопросам кредитоспособности клиентов и управлению риском невозврата [1, с. 312]. Стремительное увеличение объемов данных, усложнение структуры клиентских профилей и ужесточение требований к системам риск-менеджмента финансовые институты вынуждены переосмысливать используемый аналитический инструмент. Прежде всего, на первый план выходит задача выбора между классическими эконометрическими моделями, которые обладают высокой степенью интерпретируемости и предсказательной способностью, и современными алгоритмами, основанными на нейросети, которые обеспечивают более высокую точность для сложноструктурированных данных.

Выбор правильной методологии приобретает принципиальное значение при решении прикладных задач кредитного скоринга, определение мошенничества, прогнозирование клиентского оттока и комплексной оценки платежеспособности заемщиков [2, с. 413]. Более того, определяющую роль в контексте управления кредитным риском приобретает этап количественной оценки, на котором формируется баланс между целями максимизации доходности и обеспечения финансовой устойчивости кредитной организации. Корректность данной оценки играет решающую роль, так как ошибки могут привести к разнонаправленным экономическим последствиям. Завышенная оценка риска потенциального заемщика может спровоцировать необоснованный отказ в предоставлении кредита экономически состоятельным клиентам, что обуславливает ограничение роста кредитного портфеля и сокращает прибыль финансовой организации. В свою очередь, недооценка кредитного риска, связанная с неполным учетом факторов, которые определяют платежеспособность заемщика, повышает вероятность просроченной задолженности и негативно сказывается на качестве кредитного портфеля [3, с. 59].

Таким образом, банковский скоринг является ключевым инструментом управления кредитным риском, способствующий принятию решения о предоставлении кредитных продуктов заемщику. Банковский скоринг – это механизм оценки кредитоспособности клиентов с помощью эконометрических, статистических методов и машинного анализа. В широком смысле слова, скоринг представляет собой моделирование клиентской базы по классификациям. Функционирование скоринговых моделей основывается на гипотезе о поведенческой однородности экономических агентов с схожими социально-экономическими характеристиками. Это позволяет строить статистические модели с высоким уровнем универсальности. Наряду с этим, методологической проблематикой при разработке скоринговых моделей выступает обоснованный выбор наборов признаков и правильное определение их весовых коэффициентов, которые оценивают вклад каждого фактора в результативный признак оценки кредитного риска [4, с. 17].

**Целью** данного исследования является комплексный анализ теоретических подходов к построению скоринговых моделей кредитного риска, основанных на применении логистической регрессии и позволяющих выявить, а также систематизировать факторы, оказывающих влияние на качество оценки кредитного риска, и идентифицировать условия достижения эффективности использования данных моделей в банковской деятельности.

## Материалы и методы

В условиях возрастающей значимости скоринговых моделей в системе управления кредитным риском особую актуальность приобретает анализ степени научной проработанности данной проблематики. Несмотря на значительное число исследований отдельных проблемам построения скоринговых моделей, в научной литературе четко тестируется явный недостаток комплексных работ, систематизирующих различные подходы к применению логистической регрессии в процессе оценки кредитного риска. В частности, ограничено число публикаций, посвященных одновременному изучению вопросов прогностической точности моделей, интерпретируемости, устойчивости к изменениям выборки и соответствию регуляторным требованиям.

Данная статья представляет собой аналитический обзор, в рамках которого рассматриваются вопросы теоретико-методологических основ построения скоринговых моделей на базе логистической регрессии, ключевые этапы разработки моделей кредитного скоринга, в том числе отбор признаков, кодирование переменных и оценка параметров, а также ограничения логит-моделей, мероприятия, способствующие увеличению качества скоринга и практические аспекты внедрения в банковскую деятельность.

### Теоретические основы скоринговых моделей кредитного риска

Формирование банковского скоринга как механизма оценки кредитоспособности клиентов относится к середине XX века. На начальном этапе использовались логистическая регрессия и дискриминантный анализ, которые основывались на предположении о наличии линейной зависимости между исследуемыми факторами и вероятностью наступления дефолта [1, с. 313].

По мере развития банковской практики и расширения спектра решаемых задач подходы к оценке кредитного риска претерпели существенную дифференциацию. Это привело к формированию различных видов скоринговых моделей, различающихся по целям применения, используемым данным и методам построения моделей (табл. 1).

**Таблица 1 – Виды скоринговых моделей<sup>1</sup>**

Table 1 – Types of scoring models

Вид модели	Описание
Анкетный скоринг	Сбор информации о клиенте посредством анкетирования для принятия решения об одобрении или отказе в кредитовании
Противомошеннический скоринг	Сбор информации о клиенте посредством анкетирования для принятия решения об одобрении или отказе в кредитовании
Коллекторный скоринг	Используется при работе с клиентами для взыскания задолженности по платежам. Оценивается вероятность возврата просроченного кредита, на основе которой составляется план дальнейшей работы коллекторского отдела.
Поведенческий скоринг	Применяется для прогнозирования потенциального снижения кредитоспособности клиента, основанного на статических данных об операциях по кредитным картам. Данный метод позволяет контролировать и изменять предлагаемые лимиты займа.

Как и любой другой процесс, банковский скоринг обладает достоинствами и недостатками. К числу достоинств следует отнести:

- 1) скорость и стандартизация принятия решений, то есть снижается продолжительность рассмотрения заявки;
- 2) объективность и прозрачность – полностью исключен человеческий фактор при принятии решения;
- 3) увеличение качества портфеля благодаря точным способам прогнозирования вероятности дефолта;
- 4) снижение операционных затрат и повышение производительности сотрудников банка;
- 5) возможность риск-ориентированного ценообразования – взаимосвязь процентной ставки от риска клиента.

Наряду с достоинствами банковского скоринга, существуют определенные критерии, выполнение которых способствует поддержанию и увеличению точности построенной модели.

- Необходимость регулярной проверки адекватности модели, обусловленная ее зависимостью от исторических данных. Скоринговые модели строятся на основе информации о клиентах прошлых периодов, поэтому для качественного определения, какие заемщики склонны возвращать кредит, а какие – нет, необходимо постоянное обновление модели. [4, с. 18].

- Высокое качество и полнота исходных данных. Наличие пропусков и искажений способствует некорректной оценке кредитоспособности клиентов, приводящей как к завышению, так и к занижению риска. В то же время, недостаток информации о клиентах снижает предсказательную способность модели.

- Учет влияния внешних макроэкономических факторов (табл. 2).

<sup>1</sup> Составлено авторами по: [4].

Таблица 2 – Факторы, влияющие на точность скоринговых моделей<sup>1</sup>

Table 2 – Factors influencing the accuracy of scoring models

Факторы	Содержание
Процентная ставка	Повышение ставки приводит к удорожанию кредитных ресурсов, влекущему за собой рост долговой нагрузки клиентов, что в свою очередь, потенциально увеличивает вероятности дефолта. Поэтому, если в модели не была учтена динамика ставок, то это может привести к недооцениванию кредитного риска, особенно в условиях ужесточения денежно-кредитной политики.
Инфляция	Ускорение инфляционных процессов способствует снижению реальной покупательной способности доходов населения и увеличению расходов домохозяйств, что может отразиться на погашении задолженности.
Доходность населения	Увеличение безработицы, нестабильная занятость, снижение реальных доходов увеличивает финансовую уязвимость. В условиях экономического спада ранее надежные клиенты могут переходить в более рискованные категории, что требует от скоринговых моделей периода адаптации к изменяющимся условиям.

Процедура одобрения выдачи кредита (рис. 1) начинается с приема кредитным менеджером заявки клиента, содержащей информацию о цели, сумме, сроке кредита и возможном гарантийном обеспечении. На данном этапе работы кредитного консультанта используется упрощенная скоринговая модель (скоринг «первого уровня»), решающая задачу предварительного отбора заявителей, которые имеют очевидные признаки низкой платежеспособности: возраст, доход, занятость, кредитная нагрузка. Благодаря скорингу «первого уровня» сокращается нагрузка на последующие подразделения и ускоряется процесс рассмотрения заявок.

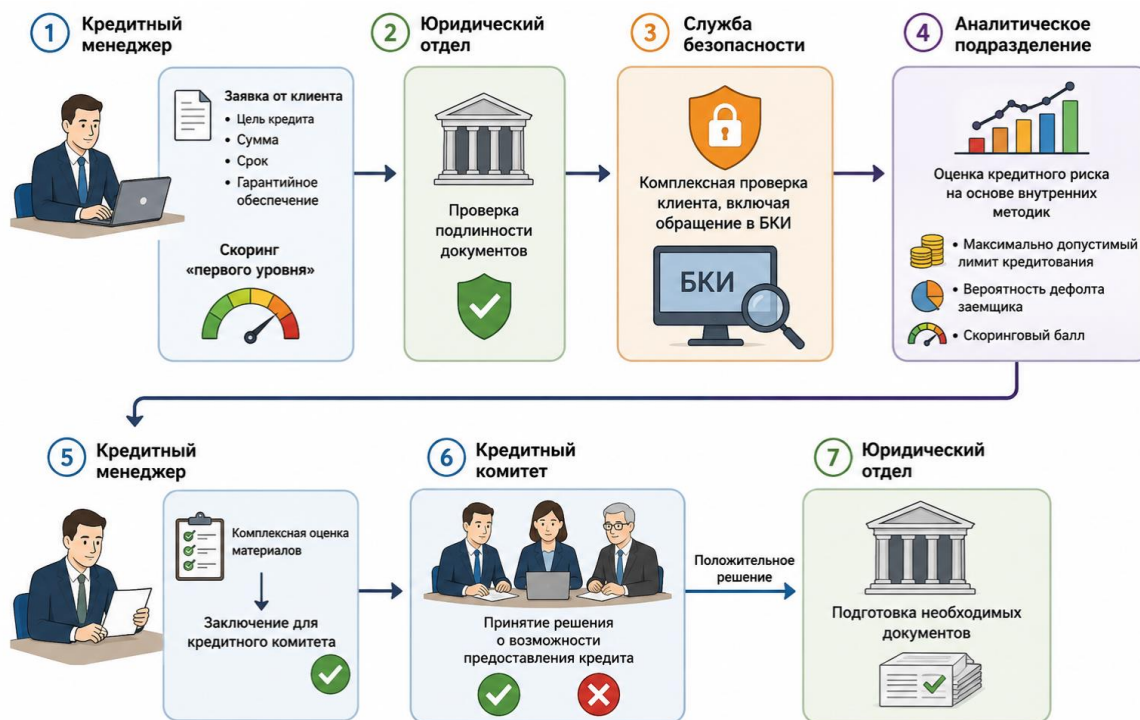


Рис. 1 Схема процедуры оценки и принятия решения по кредитной заявке<sup>2</sup>

Fig. 1 Procedure for credit application assessment and credit decision-making

<sup>1</sup> Составлено авторами по: [4].

<sup>2</sup> Составлено авторами по [4]. При разработке схемы была использована генеративная модель нейросети Dall-E 3, версия – dall-e-3 (OpenAI), апрель 2026, <https://openai.com/ru-RU/index/dall-e-3/>.

Далее сформированный пакет документов направляется в юридический отдел для проверки подлинности. После этого предоставленные документы передаются в службу безопасности, которая проводит комплексную проверку клиента, включающую обращение в бюро кредитных историй. По итогам проверки заявка и данные кредитной истории направляются в аналитическое подразделение, которое на основе внутренних методик оценивания кредитного риска рассчитывает максимально допустимый лимит кредитования, вероятность дефолта заемщика и скоринговый балл. Именно на данном этапе скоринговая модель используется в полном объеме, основываясь на данных анкеты, кредитной истории и финансовых показателей.

На следующем этапе кредитный менеджер осуществляет комплексную оценку полученных материалов и составляет заключение для кредитного комитета, который принимает окончательное решение о возможности предоставления кредита и передает информацию в юридический отдел для подготовки необходимых документов в случае положительного решения [4, с. 19].

После описанной многоэтапной процедуры рассмотрения заявки и принятия решения целесообразно отметить, что данный процесс направлен на создание комплексной оценки платежеспособности заемщика. В практической деятельности результаты проверок требуют приведения их к единому количественному и категориальному показателю. Поэтому возникает необходимость использования рейтинговых и скоринговых инструментов, благодаря которым стандартизируется процедура принятия решений.

Ключевым элементом данной системы оценки выступает кредитный рейтинг, который является сводным показателем определения кредитоспособности и степени надежности потенциального заемщика. Присвоение рейтинга осуществляется на основе факторов, характеризующих финансовое состояние клиента, его платежную дисциплину и уровень кредитного риска. В результате заемщики распределяются по определенным классам, что способствует дифференциации условий предоставления кредита и контролю рисков. Рейтинговая система включает в себя базовые компоненты кредитного риска, используемые в современных программах риск-менеджмента: вероятность дефолта, уровень потерь при дефолте и величину кредитного займа.

Применение кредитных рейтингов регламентируется рекомендациями Базельского комитета по банковскому надзору, в соответствии с которым различают внешние и внутренние рейтинговые системы. В российской практике к числу аккредитованных рейтинговых агентств, формирующих внешние рейтинги, относят АКРА, Эксперт РА, НКР и НРА [5, с. 53]. Внутренние рейтинги разрабатываются непосредственно кредитными организациями.

Скоринговые инструменты в отличие от рейтинговых подходов представляют собой количественные модели, которые направлены на автоматизированную оценку кредитоспособности заемщика, используя статистические данные прошлых периодов, применение которых снижает субъективный фактор принятия решений.

Результаты скоринговых моделей также используются:

- при расчете вероятности дефолта (PD) для определения резервов на возможные потери по ссудам;
- при оценке риск-взвешенных активов (RWA) в соответствии с требованиями Базельского соглашения и Инструкции Банка России №199-И;
- при построении внутренних рейтинговых систем (IRB).

Обобщая представленные подходы, можно отметить, что процесс разработки скоринговой модели носит последовательный характер и реализуется через совокупность взаимосвязанных процедур. Типичный процесс построения модели включает этапы, представленные на рис. 2.

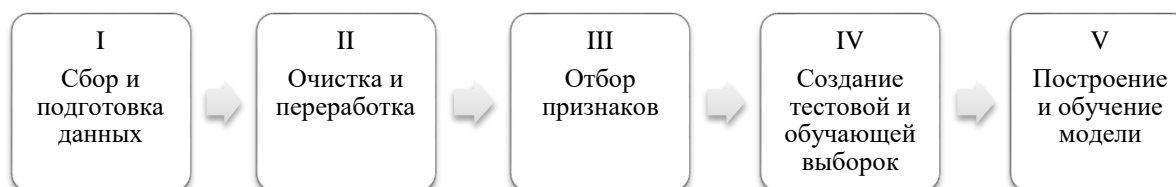


Рис. 2. Этапы построения скоринговой модели<sup>1</sup>  
Fig. 2. Stages of building a scoring model

<sup>1</sup> Составлено авторами.

На первоначальном этапе осуществляется формирование исходной выборки, которая служит эмпирической базой для последующего моделирования. В выборку включаются данные по кредитам, содержащие как характеристики заемщика, так и сведения о параметрах кредитного продукта. Ключевым элементом является определение целевой переменной (таргета) с меткой «возврат» / «невозврат».

Следующий этап направлен на повышение качества исходной информации: обработка пропущенных переменных, выявление и корректировка выбросов, а также дополнительно может проводиться процедура группировки значений признаков (биннинг).

На третьем этапе исключаются из модели неинформативные переменные, не оказывающие статистически значимого воздействия на результирующий признак, а также мультиколлинеарные переменные, способные привести к проблемам со статистической интерпретацией коэффициентов.

Четвертый этап предполагает разделение данных на обучающую и тестовую выборки, для обеспечения объективной оценки качества модели. Обучающая выборка применяется для оценки параметров модели в то время, как тестовая (валидационная) используется для проверки ее прогностической способности на ранее неиспользованных данных.

На заключительном этапе осуществляется выбор и реализация конкретного метода моделирования. Существует несколько подходов к построению скоринговых моделей: простая логистическая регрессия и более сложные методы машинного обучения (деревья решений, случайный лес, градиентный бустинг). Эти подходы имеют принципиальные отличия с точки зрения методологических особенностей, а также преимуществ и недостатков каждого из них (табл. 3).

**Таблица 3 – Классификация и сравнение моделей кредитного скоринга<sup>1</sup>**

Table 3 – Classification and comparison of credit scoring models

Метод	Определение	Достоинства	Недостатки
Логистическая регрессия	Оценка вероятности наступления события по значениям предикторов.	Высокая интерпретируемость. Низкая склонность к переобучению	Отсутствует возможность описывать нелинейные связи.
Деревья решений	Каждый узел проверяет один признак, ветки – возможные ответы, конечные листья дают прогноз.	Способна учитывать нелинейных формы зависимостей. Нет необходимости масштабирования. Способна выявить ключевые факторы.	Средняя устойчивость к переобучению. Чувствительность к незначительным корректировкам данных.
Случайный лес	Ансамбль решающих деревьев, усредняется результат отдельных деревьев.	Устойчивость к переобучению. Высокая точность прогнозов.	Низкое качество интерпретации. Высокие вычислительные требования.
Градиентный бустинг	Ансамбль слабых моделей, каждая обучается и исправляет ошибки предыдущих.	Высокая точность прогнозов. Возможность работы с разными типами данных. Эффективное выявление сложных зависимостей.	Сложность интерпретации. Чувствительность к переобучению. Высокая вычислительная нагрузка.

Машинное обучение (*англ. Machine Learning, ML*) – это класс методов искусственного интеллекта (рис. 3), в которых специалисты обучают алгоритмы на данных решать поставленную задачу без описания четкой последовательности действий.

Главная идея ML-метода – обучить модель сама выявлять скрытые закономерности и находить решение, то есть аналогично человеку думать и действовать на основе предшествующего опыта. При этом при реализации метода программирование в явном виде не используется. Одной из востребованных задач, решаемых ML-методом, является классификация объектов (рис. 4), то есть распределение объектов совокупности по определенным классам, опираясь на значения ряда известных признаков.

<sup>1</sup> Составлено авторами по: [6].

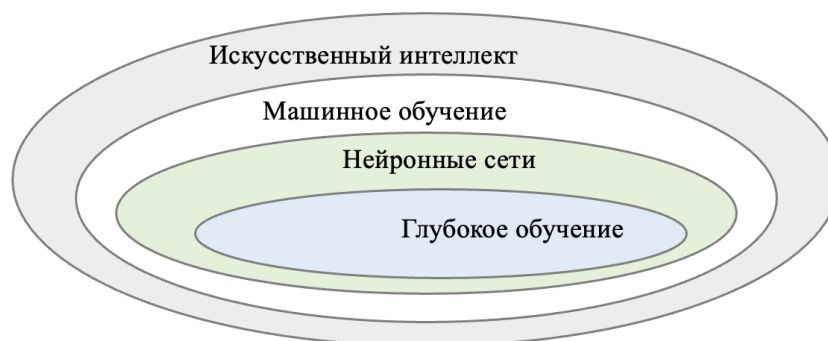
Рис. 3. ML-метод в структуре искусственного интеллекта<sup>1</sup>

Fig. 3. ML method in the structure of artificial intelligence

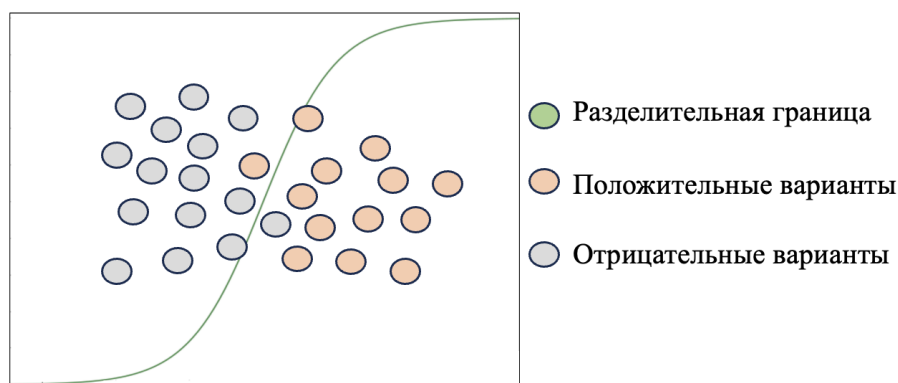
Рис. 4. Бинарная классификация объектов выборки<sup>2</sup>

Fig. 4. Binary classification of sample objects

Алгоритмы машинного обучения принято разделять на интерпретируемые (объяснимые) и плохо интерпретируемые, то есть «черные ящики». Среди интерпретируемых алгоритмов выделяют логистическую регрессию.

Логистическая регрессия (англ. *logit regression*) – это статистическая модель для прогнозирования вероятности появления события или принадлежности объекта определенному классу по значениям определенного набора признаков. Метод является достаточно «старым» инструментом в сравнении с более сложными методами, реализуемыми в машинном обучении. Однако одним из его неоспоримых преимуществ, обеспечивающим неугасающий интерес специалистов в этой области, является высокая интерпретируемость. Более сложные «черные ящики» таким качеством не обладают и в ситуациях, когда требование к интерпретируемости находится на критически высоком уровне, альтернативы логистической регрессии просто нет. Коэффициенты логистической регрессии напрямую показывают вклад каждого отдельного признака в вероятность появления ключевого события.

Математическим ядром модели логистической регрессии является метод максимального правдоподобия (англ. *maximum likelihood function*). По выборке  $x_1, x_2, \dots, x_n$ , имеющейся в распоряжении, исследователь пытается выяснить, как устроена генеральная совокупность. Для этого делается предположение, что выборка хорошо описывается некоторой функцией  $f(x|\theta)$ , в которой  $\theta$  – это параметр (или вектор параметров), который потребуется оценить.

Правдоподобие – это мера возможности или вероятность при каждом фиксированном значении параметра  $\theta$  получить наблюдаемую выборку  $x_1, x_2, \dots, x_n$ . В качестве оценки неизвестного параметра  $\theta$  выбирается такое значение, при котором правдоподобие достигает максимума [9; 10].

<sup>1</sup> Составлено авторами по: [7].

<sup>2</sup> Составлено авторами по: [8].

Суть метода сводится к решению задачи максимизации функции правдоподобия, которая представляет собой плотность вероятности совместного появления результатов выборки  $x_1, x_2, \dots, x_n$ :

$$L(x_1, x_2, \dots, x_n; \theta) = f(x_1|\theta) \cdot f(x_2|\theta) \cdot \dots \cdot f(x_n|\theta) \rightarrow \max(\theta). \quad (1)$$

Для упрощения решения задачи максимизации функция правдоподобия предварительно логарифмируется. В результате дальнейшее дифференцирование упрощается, так как по свойству логарифмов произведение функций преобразуется в сумму:

$$\ln L(x_1, x_2, \dots, x_n; \theta) = \ln f(x_1|\theta) + \ln f(x_2|\theta) + \dots + \ln f(x_n|\theta) \rightarrow \max(\theta). \quad (2)$$

Для определения оценки параметра  $\theta$  требуется решить уравнение (систему уравнений) правдоподобия, которое будет получено взятием производной и приравняв ее к нулю:

$$\frac{d \ln L}{d \theta} = 0. \quad (3)$$

Когда в распоряжение представителя аналитического подразделения банка поступает информация по ряду потенциальных заемщиков (данные анкет, кредитные истории, финансовые показатели и т. д.) формируется задача принятия решения на основании этих данных о выдаче кредита или об отказе в выдаче. Ответ на этот вопрос может быть описан бинарной переменной:

$$y = \begin{cases} 1, & \text{одобрить выдачу кредита,} \\ 0, & \text{отказать в выдаче кредита.} \end{cases} \quad (4)$$

По определению математического ожидания бинарной переменной  $y$  ее среднее равно вероятности положительного исхода:

$$M(y) = 1 \cdot P(y = 1) + 0 \cdot P(y = 0) = P(y = 1). \quad (5)$$

Одной из функций, которые используют для моделирования вероятности, является сигмоидальная функция (рис. 5).

Особенность сигмоиды состоит в возможности преобразовывать любые числа из интервала  $(-\infty; +\infty)$  в значения из промежутка  $[0; 1]$ :

$$\varphi(z) = \frac{1}{1 + e^{-z}}, \quad (6)$$

$$z = w_0 + w_1 x_1 + w_2 x_2 + \dots + w_n x_n, \quad (7)$$

где:

$z$  – линейная комбинация факторов,

$w_0, w_1, w_2, \dots, w_n$  – свободный член и коэффициенты регрессии,

$x_1, x_2, \dots, x_n$  – факторы, определяющие результат принятия решения.

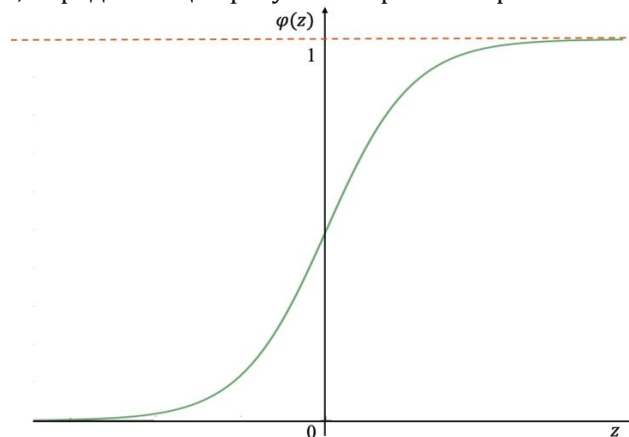


Рис. 5. График сигмоидальной функции

Fig. 5. Graph of a sigmoid function<sup>1</sup>

Через сигмоидальную функцию можно установить связь между вероятностью положительного исхода (вероятностью одобрения кредита) и линейной комбинацией характеристик заемщиков [11]:

<sup>1</sup> Построено авторами (Gretl).

$$P(y = 1) = \frac{1}{1 + e^{-z}}. \quad (8)$$

Отношение шансов – это отношение вероятности положительного исхода (одобрения) к вероятности отрицательного (отказа). Учитывая (8), отношение шансов может быть преобразовано:

$$\frac{P}{1 - P} = \frac{\frac{1}{1 + e^{-z}}}{1 - \frac{1}{1 + e^{-z}}} = \frac{1}{e^{-z}} = e^z, \quad (9)$$

$$\frac{P}{1 - P} = e^{w_0 + w_1 x_1 + \dots + w_n x_n} = e^{w_0} \cdot e^{w_1 x_1} \cdot \dots \cdot e^{w_n x_n}. \quad (10)$$

В основе внутреннего механизма логистической регрессии лежит функция логит или функция логарифма шансов:

$$\text{logit}(P) = \ln \frac{P}{1 - P}, \quad (11)$$

$$\text{logit}(P) = \ln(e^{w_0} \cdot e^{w_1 x_1} \cdot \dots \cdot e^{w_n x_n}), \quad (12)$$

$$\text{logit}(P) = w_0 + w_1 x_1 + \dots + w_n x_n, \quad (13)$$

$$\text{logit}(P) = z. \quad (12)$$

Фактически функция логарифма шансов является обратной сигмоидальной функции, так как выполняет преобразование ограниченного промежутка значений вероятностей  $[0; 1]$  в бесконечный промежуток  $(-\infty; +\infty)$ . На определенном этапе внутри компьютера как бы временно отменяется действие сигмоиды только потому, что так проще реализуются численные методы. Благодаря этому компьютер «думает», что это обычная регрессия, только строит ее не по  $y$ , а по  $\text{logit}(y)$ . Пользователь напрямую с функцией  $\text{logit}(y)$  не встречается, он оперирует функцией  $y = \text{sigmoid}(x)$ .

В случае логистической регрессии функция максимального правдоподобия имеет специальный вид:

$$L(w) = \prod_{i=1}^n (P(y_i = 1)^{y_i} \cdot (1 - P(y_i = 1))^{(1-y_i)}). \quad (14)$$

Выражение (13) преобразуется с учетом сигмоидальной функции:

$$L(w) = \prod_{i=1}^n \left( \varphi(z^{(i)})^{y^{(i)}} \cdot (1 - \varphi(z^{(i)})^{1-y^{(i)}}) \right). \quad (15)$$

После преобразования выполняется логарифмирование:

$$\log L(w) = \sum_{i=1}^n \left( y^{(i)} \log(\varphi(z^{(i)})) + (1 - y^{(i)}) \log(1 - \varphi(z^{(i)})) \right). \quad (16)$$

В рамках модели логистической регрессии обычно заменяют поиск максимума функции правдоподобия вычислением минимума отрицательного логарифма правдоподобия, то есть логистическую потерю или *Log Loss*:

$$J(w) = - \sum_{i=1}^n \left( y^{(i)} \log(\varphi(z^{(i)})) + (1 - y^{(i)}) \log(1 - \varphi(z^{(i)})) \right). \quad (17)$$

В последнем выражении при  $y^{(i)} = 0$  в ноль обращается первое слагаемое, а при  $y^{(i)} = 1$  – второе:

$$J(w) = \begin{cases} -\log(\varphi(z^{(i)})), & \text{если } y^{(i)} = 1, \\ -\log(1 - \varphi(z^{(i)})), & \text{если } y^{(i)} = 0. \end{cases} \quad (18)$$

Таким образом, *Log Loss* является измерителем ошибки или погрешности в методе логистической регрессии. Для определения параметров (весов модели), минимизирующих функцию *Log Loss*, принято использовать метод градиентного спуска (англ. *Gradient Descent*), точнее его усовершенствованные версии (стохастический градиентный спуск, мини-батч, добавление импульса и т.д.). При этом фундаментом всех модификаций является классический алгоритм [9; 10].

Общая идея метода градиентного спуска для поиска минимума функции заключается последовательном итерационном движении в направлении обратном градиенту этой функции (рис. 6). Прежде чем переходить к итерациям, выполняется подготовка.

На первом этапе требуется выбрать функцию, которая будет измерять ошибку на каждой итерации, то есть величину разницы между фактическими и прогнозируемыми значениями. В качестве такого индикатора обычно берут среднеквадратическую ошибку (англ. *Mean Squared Error, MSE*).

На втором этапе необходимо задать начальные значения параметрам модели, то есть выбрать вектор весов  $W_0 = (w_0, w_1, \dots, w_n)$ . В общем случае все параметры модели можно установить на нулевом уровне, однако в машинном обучении это может привести к симметрии при обучении. На этом этапе для параметров модели следует установить случайные значения.

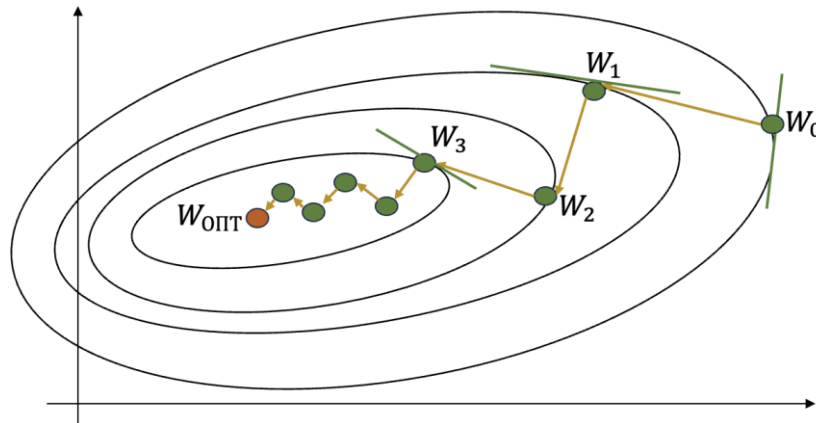


Рис. 6. Общая схема метода скорейшего градиентного спуска<sup>1</sup>  
 Fig. 6. General scheme of the steepest gradient descent method

Третий этап подготовки заключается в вычислении вектора-градиента функции *Log Loss*, то есть вектора частных производных по весам модели логистической регрессии:

$$\nabla J = (J'_{w_1}; J'_{w_2}; \dots; J'_{w_n}). \quad (19)$$

После трех подготовительных этапов переходят непосредственно к итерациям. Построение последовательных приближений, то есть обновление весов модели логистической регрессии происходит по формуле:

$$W_{k+1} = W_k - \lambda \cdot \nabla J(w), \quad (20)$$

где:

$W_i$  и  $W_{i+1}$  – исходный и обновленный векторы весов модели соответственно;

$\lambda$  – положительное число, «шаг» каждой итерации (англ. *learning rate*);

$\nabla J$  – градиент функции *Log Loss*.

Градиентные методы отличаются способом выбора «шага». Например, метод скорейшего спуска предполагает вычисление  $\lambda$  из условия ортогональности градиентов в двух последовательных точках  $W_k$  и  $W_{k+1}$ :

$$\nabla J_{k+1} \cdot \nabla J_k = 0. \quad (21)$$

В результате такого выбора «шага» для любого  $k$  луч, идущий от точки  $W_k$  и  $W_{k+1}$  перпендикулярен линии уровня функции *Log Loss*, проходящей через точку  $W_k$  и касается линии уровня, проходящей через точку  $W_{k+1}$ , так как перпендикулярен к следующему лучу (рис. 6).

Выбор в качестве направления движения вектора-градиента и специальный подбор «шага» обеспечивают сходимость последовательности  $W_0, W_1, \dots, W_k, \dots$  к оптимальной точке  $W_{\text{ОПТ}}$ . В редких исключениях это может произойти за конечное число итераций, но чаще этот процесс бесконечен и на определенном шаге определенный набор весов  $W_k$  берется в качестве приближенного значения оптимального решения. Момент, когда можно считать, что функция *Log Loss* достигла минимума,

<sup>1</sup> Составлено авторами по: [10].

определяется условием, выбранным на первом этапе в качестве индикатора завершения итераций (MSE). Условие проверяется на каждой итерации.

На практике чаще применяют улучшенные версии метода градиентного спуска, так как базовый метод может допустить ошибку и выбрать в качестве минимума функции не глобальный, а локальный. Так же существуют методы, альтернативные градиентному спуску. Например, метод Ньютона, который оперирует со вторыми производными функции потерь. В результате вычислительная сложность возрастает значительно, хотя в некоторых случаях метод может помочь быстрее достигнуть оптимальную точку. Кроме того, градиентный метод обладает еще рядом недостатков (переобучение, взрыв градиентов, затухающий градиент), но разработчики уже научились их преодолевать.

Таким образом, процесс реализации логистической можно изобразить в виде схемы (рис. 7). Этапом, предшествующим логистической регрессии, является формулирование задачи или вопроса, на который требуется получить ответ в формате «да» или «нет», и выполняется сбор и предобработка данных. Результатом этого этапа является сформированный набор данных со значениями признаков  $x_1, x_2, \dots, x_n$ .

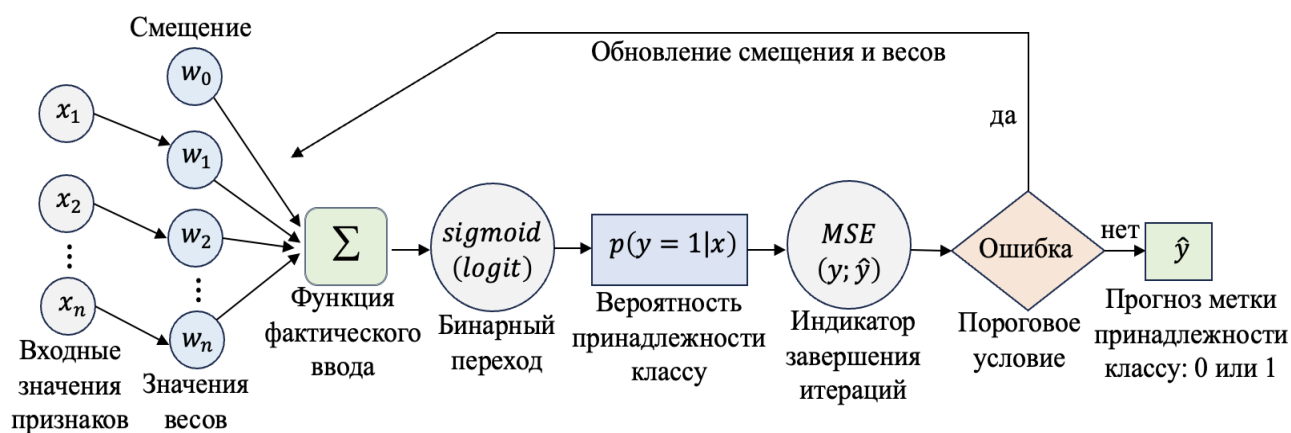


Рис. 7. Схема реализации логистической регрессии<sup>1</sup>

Fig. 7. Implementation scheme for logistic regression

Первым этапом логистической регрессии является выбор случайных значений смещения  $w_0$  и весов  $w_1, w_2, \dots, w_n$ . Далее формируется и подается на вход функция фактического ввода, как сумма произведений значений признаков и соответствующих весов с добавлением смещения.

На втором этапе функция фактического ввода пропускается через сигмоидальную функцию  $y = \text{sigmoid}(x; w)$ , в результате чего числовые входы преобразуются в вероятности принадлежности классу ( $y = 1|x$ ). На этом этапе функция потерь *Log Loss* минимизируется одним из усовершенствованных градиентных методов.

Третий этап состоит в сравнении фактических и прогнозируемых меток принадлежности классу (0 или 1). На этом этапе вычисляется ошибка (обычно среднеквадратическая) и ее значение сверяется с пороговым, после чего принимается решение о продолжении или завершении итераций. В случае, если значение ошибки превышает пороговое значение, алгоритм возвращается к первому этапу для корректировки весов. Так продолжается много раз, пока не будет достигнуто пороговое условие завершения итераций.

В результате применения логистической регрессии помимо непосредственно предсказанной вероятности метки принадлежности классу, коэффициенты регрессии, позволяют оценить влияние каждого отдельного признака на эту вероятность.

В процессе реализации методов машинного обучения, в том числе более сложных, отдельное внимание уделяется показателям, которые используются для оценки качества и сравнения разных моделей (табл. 4).

<sup>1</sup> Составлено авторами по: [12].

Таблица 4 – Показатели оценки качества логистической регрессии<sup>1</sup>

Table 4 – Logistic Regression Quality Assessment Indicators

Показатель	Определение	Интерпретация
Accuracy	Отношение числа правильных предсказаний к общему числу всех предсказаний.	Оценивает долю верных предсказаний среди всех предсказаний.
Precision (точность)	Доля действительно положительных среди всех положительных предсказаний.	Показывает степень доверия положительному предсказанию.
Recall (полнота)	Доля обнаруженных положительных объектов среди всех положительных объектов.	Характеризует чувствительность модели.
F1-score	Среднее гармоническое Precision и Recall.	Сбалансированность между точностью и полнотой.
ROC-AUC	Площадь под ROC-кривой.	Демонстрирует способность модели ранжировать объекты при разных порогах.
Log Loss	Логарифмическая функция потерь (отрицательный логарифм правдоподобия).	Измеряет расхождение между предсказанными вероятностями и истинными метками.

### Заключение

1. Повышение эффективности скоринговых моделей требует комплексного учета качества данных, корректного выбора признаков и адаптации моделей к изменяющимся условиям внешней среды. Полнота информации о заемщиках, отсутствие значительного объема пропусков, корректность обработки выработки оказывают непосредственное влияние на точность прогнозирования кредитного риска и стабильность модели.

2. Выявлено, что скоринговые модели требуют регулярной адаптации к изменениям внешней среды, включая макроэкономические колебания изменения поведения заемщиков, трансформацию банковских продуктов и регуляторных требований. Отсутствие своевременной корректировки и переоценки параметров модели может приводить к снижению точности прогнозов и росту кредитных рисков.

3. Несмотря на развитие более сложных алгоритмов, логистическая регрессия как одна из моделей кредитного скоринга не теряет свою популярность, благодаря относительной простоте и особенно высокой интерпретируемости результатов. При этом она может проигрывать в точности более сложным алгоритмам. Решение о выборе между логистической регрессией и другими более сложными моделями классификации должно приниматься на основе специфики конкретно решаемой задачи, отталкиваясь от того, насколько высоки и критичны требования к интерпретируемости. По сравнению с более сложными «черными ящиками» логистическая регрессия в определенных сферах может быть более востребованной, даже несмотря на некоторую потерю точности. Также перспективным направлением является комбинированное применение логистической регрессии с более точными методами.

4. Математическим ядром логистической регрессии является метод максимального правдоподобия. В случае реализации логистической регрессии средствами эконометрики функцию правдоподобия максимизируют, а в машинном обучении ее разворачивают в противоположную сторону и называют функцией потерь, которую соответственно минимизируют. Одним из приоритетных численных методов, средствами которого выполняется оценка модели, является метод градиентного спуска и его модификации. Преимущества метода в простоте, относительно невысокой вычислительной сложности и хорошей сходимости. Имеются и недостатки, но большую часть из них разработчики научились обходить. Также в основе реализации логистической регрессии лежат сигмоидальная функция и функция логит. Первую видит разработчик, вторую использует компьютер для градиентного метода. Построение логистической регрессии проходит под обязательным контролем значений показателей, отвечающих за качество.

<sup>1</sup> Составлено авторами по: [1].

## Список источников

1. Комиссаров А. Р. Анализ эффективности методов машинного обучения в предсказании кредитного риска // *Наукосфера*. 2024. № 6-2. С. 312–316. <https://doi.org/10.5281/zenodo.12621417>. EDN: EHJLKI
2. Силенко А. Н., Туйгунов А. И., Теслюк В. С. Сравнение эффективности традиционных и нейросетевых подходов в банковской практике // *Journal of Monetary Economics and Management*. 2025. № 9. С. 412–419. <https://doi.org/10.26118/2782-4586.2025.75.72.029>. EDN: KOOXXR
3. Гибадуллина А. А. Банковский скоринг на основе современных цифровых технологий / *Цифровая экономика глазами студентов. Материалы Международной научной конференции*. Казань, 2023. С. 59–61. EDN: THIXOL
4. Ванина М. Ф., Ерохин А. Г. Применение методов машинного обучения в банковском скоринге // *DSPA: вопросы применения цифровой обработки сигналов*. 2025. Том 15. № 1. С. 17–35. EDN: PCWMBW
5. Холда Е. Д., Идрисов Ш. А. Об оценке кредитных рисков в современной банковской системе РФ // *Журнал экономических исследований*. 2025. Том 11. № 2. С. 49–60. EDN: QBBODG
6. Кантаев Н. К. Перспективы использования методов машинного обучения для оценки кредитных рисков // *Вестник Евразийской науки*. 2025. Том 17. № S2. EDN: KOGIEQ
7. Дьяконова Л. П., Константинов А. Ф., Китова О. В., Китов В. А. Применение методов машинного обучения и искусственного интеллекта в сфере кредитования и обнаружения мошенничества // *Уголь*. 2025. № 10 (1198). С. 63–69. <https://doi.org/10.18796/0041-5790-2025-10-63-69>. EDN: FLXTEK
8. Дядюнов Д. А. Машинное обучение для риск-менеджмента в банке: возможности и вызовы // *Вестник науки*. 2025. Том 1. № 1(82). С. 265–273. EDN: OJWRUL
9. Кремер Н. Ш. Теория вероятностей и математическая статистика : учебник и практикум для вузов. – 5-е изд., перераб. и доп. М.: Издательство Юрайт, 2026. 538 с. ISBN 978-5-534-10004-4
10. Исследование операций в экономике : учебник для вузов / под редакцией Н. Ш. Кремера. – 4-е изд., перераб. и доп. М.: Издательство Юрайт, 2026. 414 с. ISBN 978-5-534-12800-0
11. Михайлова С. С., Гринева Н. В., Концевая Н. В. Эконометрика: Учебник для бакалавриата. М.: Прометей, 2024. 512 с. ISBN: 978-5-00172-626-5. EDN: RHNTBU
12. Рашка С., Мирджалили В. Python и машинное обучение. Машинное и глубокое обучение с использованием Python, scikit-learn и TensorFlow 2. Перевод с английского и редакция Ю. Н. Артеменко. – 3-е изд. М.: Диалектика; СПб.: Диалектика, 2020. 848 с. ISBN 978-5-907203-57-0

## References

1. Komissarov A.R. Analysis of the effectiveness of machine learning methods in predicting credit risk. *Naukosfera*. 2024;(6-2):312–316. (In Russ.). <https://doi.org/10.5281/zenodo.12621417>. EDN: EHJLKI
2. Silenko A.N., Tuigunov A.I., Teslyuk V.S. Comparison of the effectiveness of traditional and neural network approaches in banking practice. *Journal of Monetary Economics and Management*. 2025;(9):412–419. (In Russ.). <https://doi.org/10.26118/2782-4586.2025.75.72.029>. EDN: KOOXXR
3. Gibadullina A.A. Bank scoring based on modern digital technologies. In: *Digital economy through the eyes of students*. Materials of the International Scientific Conference. Kazan; 2023. P. 59–61. (In Russ.). EDN: THIXOL
4. Vanina M.F., Erokhin A.G. Application of machine learning methods in banking scoring. *DSPA: issues of digital signal processing application*. 2025;15(1):17–35. (In Russ.). EDN: PCWMBW
5. Holda E.D., Idrisov S.A. On the assessment of credit risks in the modern banking system of the Russian Federation. *Journal of Economic Studies*. 2025;11(2):49–60. (In Russ.). EDN: QBBODG
6. Kantaev N.K. Prospects of using machine learning methods to assess credit risks. *The Eurasian Scientific Journal*. 2025;17(S2). (In Russ.). EDN: KOGIEQ
7. Dyakonova L.P., Konstantinov A.F., Kitova O. , Kitov V.A. Application of machine learning and artificial intelligence methods in the field of lending and fraud detection. *UGOL' (Russian Coal Journal)*. 2025;10(1198):63–69. (In Russ.). <https://doi.org/10.18796/0041-5790-2025-10-63-69>. EDN: FLXTEK
8. Dyadyunov D.A. Machine learning for risk management in a bank: opportunities and challenges. *Vestnik nauki*. 2025;1(1(82)):265–273. (In Russ.). EDN: OJWRUL

9. Kremer N.S. *Probability theory and mathematical statistics* : textbook and practical course for universities, 5th ed., revised. Moscow: Yurait Publishing House; 2026. 538 p. (In Russ.). ISBN 978-5-534-10004-4

10. *Operations Research in Economics* : textbook for universities. N. S. Kremer (ed.). – 4th ed., revised and additional. Moscow: Yurait Publishing House, 2026. 414 p. (In Russ.). ISBN 978-5-534-12800-0

11. Mikhailova S.S., Grineva N.V., Kontsevaya N.V. *Econometrics*: Textbook for undergraduate studies. Moscow: Prometey; 2024. 512 p. (In Russ.). ISBN: 978-5-00172-626-5. EDN: RHNTBU

12. Rashka S., Mirjalili V. *Python and machine learning. Machine learning and deep learning using Python, scikit-learn and TensorFlow 2*. Translated from English and edited by Yu.N. Artemenko. – 3rd ed. Moscow: Dialektika; St. Petersburg: Dialektika; 2020. 848 p. (In Russ.). ISBN 978-5-907203-57-0

### ***Информация об авторах***

М. В. Чернова – кандидат физико-математических наук, доцент, доцент кафедры финансовых рынков, технологий и регулирования, Российская академия народного хозяйства и государственной службы при Президенте Российской Федерации.

Э. Ю. Захариадис – кандидат экономических наук, доцент кафедры финансовых рынков, технологий и регулирования, Российская академия народного хозяйства и государственной службы при Президенте Российской Федерации.

### ***Information about the authors***

M.V. Chernova – Cand. Sci. (Physics and Mathematics), Associate Professor at the Department of Financial Markets, Technologies and Regulation, Russian Presidential Academy of National Economy and Public Administration.

E.Yu. Zachariadis – Cand. Sci. (Econ.), Associate Professor at the Department of Financial Markets, Technologies and Regulation, Russian Presidential Academy of National Economy and Public Administration.

***Вклад авторов:*** все авторы сделали эквивалентный вклад в подготовку публикации.

Авторы заявляют об отсутствии конфликта интересов.

***Contribution of the authors:*** the authors contributed equally to this article. The authors declare no conflicts of interests.

Статья поступила в редакцию 22.04.2026; одобрена после рецензирования 25.05.2026; принята к публикации 28.05.2026.

The article was submitted 22.04.2026; approved after reviewing 25.05.2026; accepted for publication 28.05.2026.