



РАЗРАБОТКА НЕЙРОСЕТЕВОЙ МОДЕЛИ КРЕДИТНОГО СКОРИНГА ФИЗИЧЕСКИХ ЛИЦ

Япарова Наталья Михайловна

доктор технических наук, доцент,
заведующий кафедрой
«Математическое обеспечение
информационных технологий»
Южно-Уральского государственного университета,
Челябинск, Россия
E-mail: iaparovnm@susu.ru

Щеголев Александр Владимирович

магистрант кафедры
«Математическое обеспечение
информационных технологий»
Южно-Уральского государственного университета,
Челябинск, Россия
E-mail: njznew@gmail.com

Предмет исследования: методы и подходы к построению нейросетевых моделей, включающих обработку и анализ информации о данных заемщика и предназначенных для принятия решений в области кредитования.

Цель исследования: создание нейросетевой модели системы кредитного скоринга, способной точно и надежно прогнозировать кредитоспособность заемщиков на основе их социально-экономических характеристик.

Методы и объекты исследования: в качестве объекта исследования использовались данные о заемщиках, включающие различные социально-экономические признаки. Методы исследования включают предварительную обработку данных, отбор значимых признаков с использованием метода SHAP, построение полносвязной нейронной сети с тремя скрытыми слоями и функцией активации ReLU. Для выходного слоя используется функция активации Sigmoid. Оптимизация модели проводилась с помощью Adam.

Основные результаты исследования: разработанная нейросетевая модель прошла тестирование на тестовой выборке и продемонстрировала достаточную точность предсказаний. Валидация показала минимальное значение функции потерь в 0.2145 при оптимальном количестве эпох. Исследование подтвердило эффективность нейросетевых моделей в задачах кредитного скоринга.

Ключевые слова: нейронные сети, кредитный скоринг, анализ данных, оценка метода, принятие решений.

DEVELOPMENT OF A NEURAL NETWORK MODEL OF CREDIT SCORING FOR PHYSICAL PERSON

Natalia M. Iaparova

Doctor of Technical Sciences, Associate Professor,
Head of the Department of Computational
Mathematics and High-Performance Computing
South Ural State University,
Chelyabinsk, Russia
E-mail: iaparovnm@susu.ru

Alexander V. Shchegolev

master's student of
Computational Mathematics
and High-Performance Computing
South Ural State University,
Chelyabinsk, Russia
E-mail: njznew@gmail.com

Subject of research: methods and approaches to building neural network models that include processing and analyzing information about borrower's data and are intended for decision-making in the field of lending.

Purpose of research: to create a neural network model of a credit scoring system capable of accurately and reliably predicting the creditworthiness of borrowers based on their socio-economic characteristics.

Methods and objects of research: the object of research is data on borrowers, including various socio-economic characteristics. The research methods included preliminary data processing, selection of significant features using the SHAP method, and the construction of a fully connected neural network with three hidden layers and a ReLU activation function. The Sigmoid activation function is used for the output layer. The model was optimized using Adam.

Main results of research: the developed neural network model was tested on a test sample and demonstrated sufficient accuracy of predictions. The validation showed the minimum value of the loss function at 0.2145 with the optimal number of epochs. The study confirmed the effectiveness of neural network models in credit scoring tasks.

Keywords: neural networks, credit scoring, data analysis, method evaluation, decision making.

ВВЕДЕНИЕ

Современная экономика сталкивается с проблемой роста закредитованности населения и предприятий. Это создает дополнительные риски в инвестиционных вложениях, связанные с проблемой частичного или полного невозвращения выданных заемных средств, что в свою очередь приводит к снижению эффективности использования инструмента кредитования. Системы кредитного скоринга подразумевают оценку кредитоспособности заемщика и являются ключевым инструментом решения проблемы оценки уровня его финансовой ответственности, способствуют совершенствованию механизма управления

финансовыми рисками и повышению эффективности принятия решений в сфере кредитования. Системы скоринга оценивают вероятность того, что заемщик вернет кредитные средства в установленные сроки. Это помогает снизить риски невозврата средств и принимать обоснованные решения о выдаче кредита. Высокий уровень кредитоспособности может указывать на надежность заемщика, тогда как низкий рейтинг может быть сигналом о возможных финансовых проблемах [1–3]. Наиболее известные системы кредитного скоринга за рубежом включают FICO, VantageScore и Schufa, а в России популярностью пользуются решения, предлагаемые



такими бюро кредитных историй, как ОКБ и НБКИ [4, 5].

Разработка систем кредитного скоринга привлекает внимание многих исследователей, включая Кадиева А. Д. [6], Богданова А. Л. [7], которые изучали применение нейронных сетей в задачах скоринга, а также Полищука Ф. С. [8], разрабатывавшего рейтинговую систему для оценки кредитных рисков физических лиц. Высокий интерес к этой области обусловлен необходимостью повышения точности и эффективности анализа кредитоспособности, что достигается благодаря использованию инструмента нейросетевого моделирования.

Наиболее часто учитываемыми факторами при кредитном скоринге являются демографические данные, сведения о занятости, данные о кредитной истории и предыдущих отношениях с кредитором, особенности предоставляемой услуги, а также информация о финансовом состоянии клиента [9]. Таким образом, системы, учитывающие разнообразные источники данных, включая альтернативные платежные истории, социальные связи и трендовые данные, предоставляют более полную картину финансового поведения заемщика. Также это может быть важным фактором для оценки кредитоспособности тех, у кого ограниченная кредитная история.

Потребительские кредиты играют важную роль в повышении качества жизни, предоставляя доступ к ресурсам для реализации различных жизненных целей. Однако с увеличением объемов кредитования возрастает необходимость автоматизации процессов принятия решений, что требует использования современных технологий. В условиях больших объемов данных и сложных взаимосвязей между различными факторами традиционные подходы становятся менее эффективными. В этой связи использование методов искусственного интеллекта позволяет решать проблемы, возникающие в различных сферах деятельности, на качественно новом уровне [10–14].

В статье рассмотрены подходы к разработке нейросетевой модели для оценки кредитного скоринга физических лиц. Описан процесс анализа данных заемщиков, их предварительной обработки и отбора значимых признаков. Предложенная модель показала достаточную точность и устойчивость, что подтверждает её пригодность для практического применения в кредитных организациях.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Обработка исходных данных

Для оценки эффективности построенных моделей выборку необходимо разбить на обучающий и тестовый наборы. Обучающая выборка будет использоваться для построения модели нейронной сети. Тестовый набор будет использоваться для прогнозирования кредитоспособности новых заемщиков. Размер обучающего набора будет составлять 70 % от исходной выборки, а тестовый 30 % соответственно.

Исходные данные были взяты из открытых источников и содержат реальную информацию о 1000 заемщиков. Зависимая переменная представляет собой показатель кредитоспособности клиентов и является бинарной: принимает значение 1 в случае кредитоспособности и 0 в противном случае. Независимые переменные включают 20 факторов, описывающих каждого заемщика: текущий баланс счета, продолжительность кредита, кредитная история клиента, цель кредита, сумма кредита, накопления, стаж работы, рассрочка платежа, семейное положение, созаемщики или поручители, количество лет проживания по текущему месту, данные об имуществе, возраст, другие имеющиеся кредиты, условия проживания, количество активных кредитов, вид деятельности, количество иждивенцев, наличие телефона, гражданство.

Для более детального анализа данных заемщиков было принято решение рассмотреть каждый фактор по отдельности. Такой подход позволит выявить ключевые закономерности и связи между различными переменными и характеристиками клиентов, что в свою очередь может помочь в прогнозировании их кредитоспособности и управлении рисками. Рассмотрение каждого фактора отдельно также даст возможность глубже понять влияние различных переменных на целевую переменную – класс заемщика (хороший или плохой).

На этапе отбора признаков была произведена оценка их важности для модели с помощью алгоритма SHAP (SHapley Additive exPlanations). SHAP – это популярный метод, который предоставляет аналитические сведения о вкладе каждой входной функции в заданный прогноз. Это позволило сразу выявить менее значимые факторы, которые могли бы увеличить шум и усложнить процесс обучения. В результате были выделены 15 наиболее значимых признаков, оказывающих влияние на результат прогнозирования. Результат работы алгоритма SHAP можно увидеть на рисунке 1.

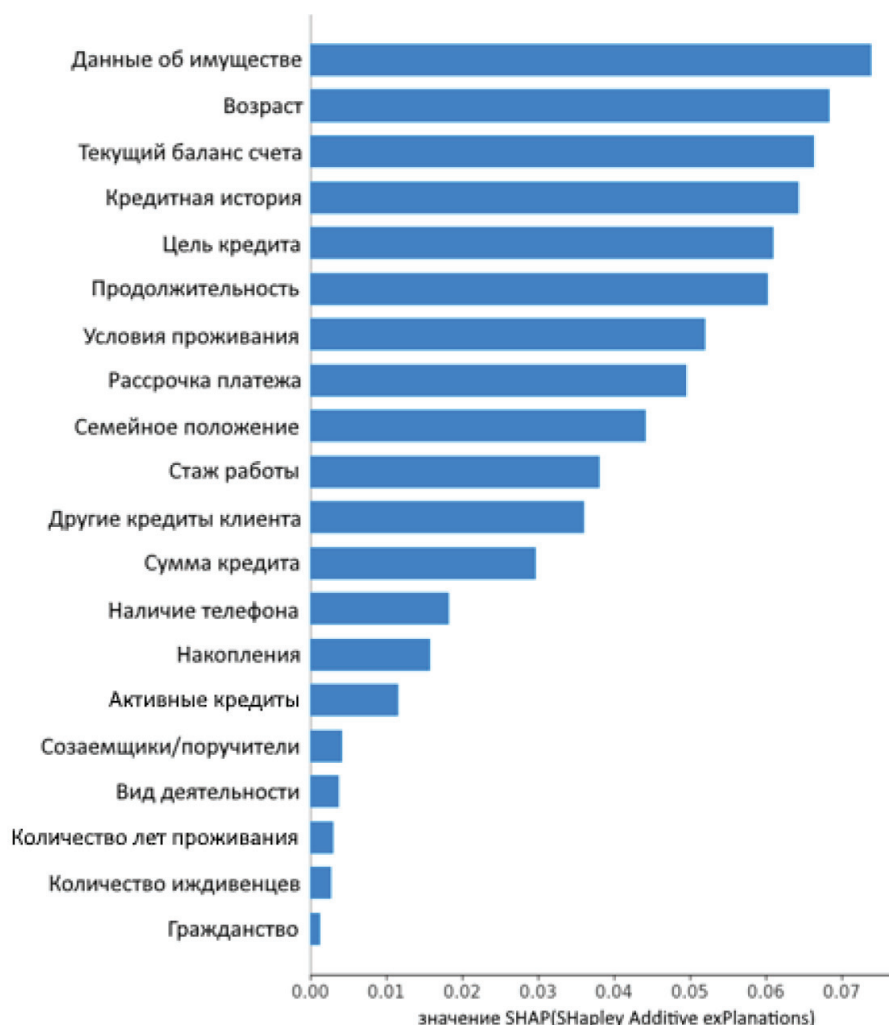


Рисунок 1. Значимость признаков

Архитектура нейросетевой модели

Предлагаемая нейронная сеть представлена следующими уравнениями:

$$z^{(l)} = \sum_{i=1}^{n_l} W_i^{(l)} a_i^{(l-1)} + b_i^{(l)}, \quad a^{(l)} = g^{(l)}(z^{(l)}),$$

где:

$W_i^{(l)} \in R^{n_l \times n_{l-1}}$ – веса слоя l ,

$b_i^{(l)} \in R^{n_l}$ – вектор смещения слоя l ,

$a^{(l)}$ – выходы слоя l ,

$a^{(l-1)}$ – выходы предыдущего слоя l ,

$g^{(l)}$ – функция активации, используемая в слое l ,

n_l – число нейронов в слое l ,

$z^{(l)}$ – выход предыдущего слоя l ,

$z^{(l-1)}$ – выход предыдущего слоя до функции активации,

l – количество слоев.

Модель реализована в виде полносвязной нейронной сети, предназначенной для прогнозирования кредитоспособности клиентов. В скрытых слоях используются функции

активации ReLU, которые помогают справляться с нелинейностями в данных. В выходном слое применяется функция активации Sigmoid, сжимающая выходное значение в диапазон от 0 до 1, что позволяет интерпретировать результат как вероятность принадлежности к одному из двух классов, необходимую для бинарной классификации.

Итоговая нейронная сеть состоит из 3 скрытых слоев с функцией активации ReLU. В выходном слое используется Sigmoid для бинарной классификации. Для предотвращения переобучения на каждом скрытом слое был использован Dropout, который случайно обнуляет 30 % нейронов во время обучения. В качестве оптимизатора был использован Adam с параметром скорости обучения 0.005, обеспечивая более быстрое и стабильное обучение. В качестве функции потерь была использована бинарная кросс-энтропия. Схему архитектуры полученной сети можно увидеть на рисунке 2. В таблице 1 приведены гиперпараметры, используемые для нейронной сети.

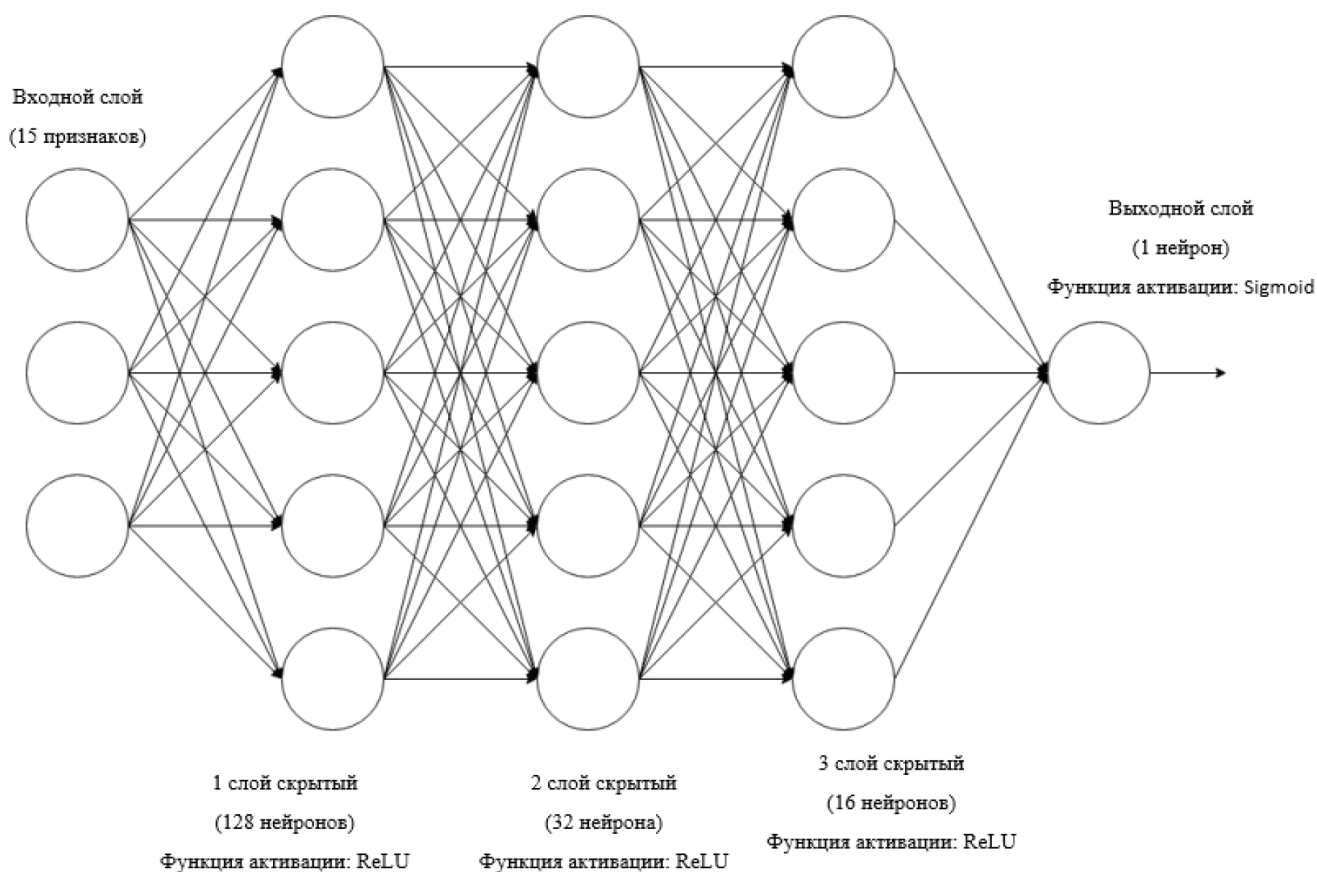


Рисунок 2. Архитектура нейронной сети

Таблица 1. Гиперпараметры нейронной сети

Оптимизатор	Adam (скорость обучения: 0.005)
Количество скрытых слоев	3
Количество нейронов	Входной слой: 15
1 скрытый слой: 128	
2 скрытый слой: 32	
3 скрытый слой: 16	
Выходной слой: 1	
Функция активации (скрытые слои)	ReLU
Функция активации (выходной слой)	Sigmoid
Dropout	0.3
Оптимизатор	Adam (скорость обучения: 0.005)
Функция потерь	Бинарная кросс-энтропия
Количество эпох	102

Вычислительные эксперименты

Цель вычислительных экспериментов заключалась в нахождении оптимального количества эпох для нейросетевой модели с

приведенным ранее набором гиперпараметров для минимизации функции потерь. График потерь с функциями потерь с изменением количества эпох представлен на рисунке 3.

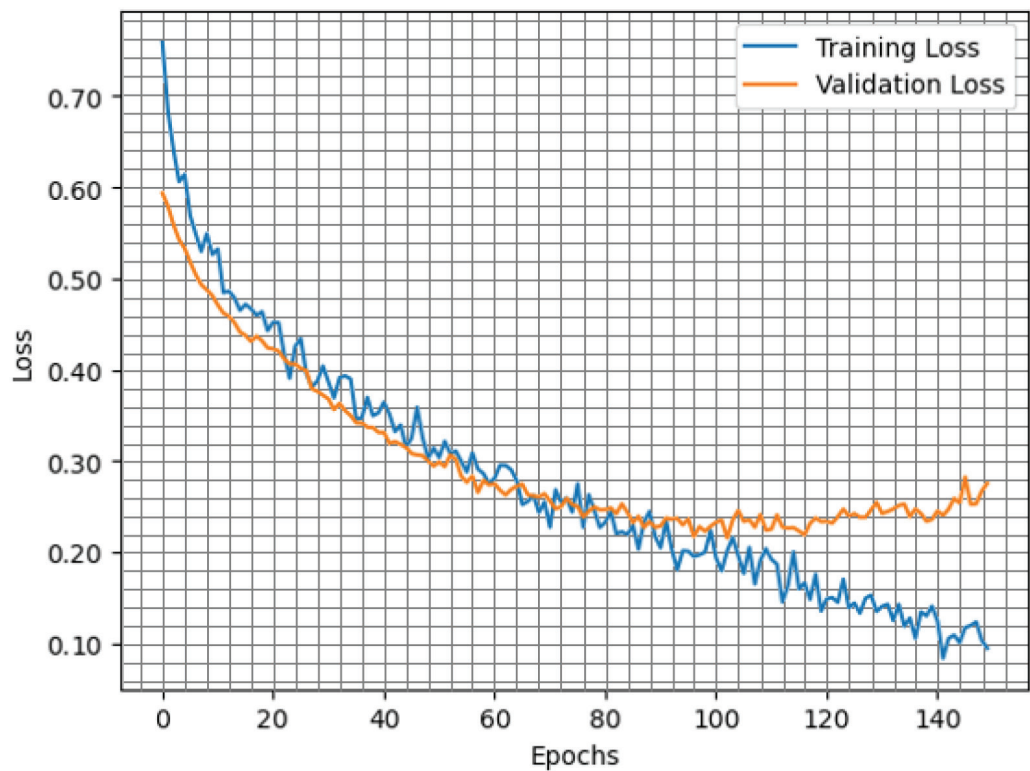


Рисунок 3. График функций потерь

Синяя линия соответствует графику функции потерь, полученной при обработке обучающей выборки. Значения функции потерь свидетельствуют об устойчивом снижении объема потерь с увеличением количества эпох, что свидетельствует об успешном обучении модели. Оранжевая линия отображает значения функции потерь на тестовой выборке. На начальных этапах она тоже уменьшается, но начиная с определенного момента наблюдается расхождение с потерями на

обучающей выборке. Однако с определенного количества эпох значения функции потерь существенно не меняются. Это свидетельствует о начале переобучения модели. В таблице 2 приведены значения функции потерь относительно количества эпох обучения на тестовой выборке. Результаты эксперимента свидетельствуют о том, что оптимальное количество эпох составило 102 со значением функции потерь 0.2145, что говорит о достаточной точности построенной модели.

Таблица 2. Значения функции потерь относительно количества эпох обучения на тестовой выборке

Количество эпох	Значения функции потерь
97	0.2300
98	0.2268
99	0.2316
100	0.2348
101	0.2364
102	0.2145
103	0.2347
104	0.2464
105	0.2367
106	0.2383
107	0.2306

ЗАКЛЮЧЕНИЕ И ВЫВОДЫ

В работе был предложен подход к разработке системы кредитного скоринга физических лиц. Для решения поставленной задачи использовались методы предварительной обработки данных, включающие нормализацию, кодирование категориальных признаков и отбор значимых факторов с применением алгоритма SHAP. На основе обработанных данных была построена полносвязная нейронная сеть, состоящая из трёх скрытых слоёв с функцией активации ReLU и выходным слоем с функцией активации Sigmoid. Результаты проведённого тестирования подтвердили достаточную точность модели на тестовой выборке. Оптимальное количество эпох, определённое в ходе экспериментов, составило 102, при этом минимальное значение функции потерь достигло 0.2145. Полученные результаты подтверждают целесообразность использования нейросетевых подходов для прогнозирования кредитоспособности заемщиков, что может значительно повысить качество принятия решений в кредитных организациях.

СПИСОК ЛИТЕРАТУР

1. Глинкина, Е. В. Кредитный скоринг как инструмент эффективной оценки кредитоспособности / Е. В. Глинкина // Финансы и кредит. – 2011. – № 16 (448). – С. 43–47.
2. Митрофанова, К. Б. Понятие кредитного риска и факторы, на него влияющие / К. Б. Митрофанова // Молодой ученый. – 2015. – № 2. – С. 284–288.
3. Абдуллаев, Н. А. Перспективы внедрения современных технологий искусственного интеллекта в скоринговые системы / Н. А. Абдуллаев // Экономика и финансы (Узбекистан). – 2023. – № 1 (161). – С. 39–49.
4. Волкова, А. С. Скоринговая система оценки кредитоспособности / А. С. Волкова // Проблемы экономики и менеджмента. – 2016. – № 5 (57). – С. 136–138.
5. Кочеткова, В. В. Обзор методов кредитного скоринга / В. В. Кочеткова, К. Д. Ефремова // Juvenis scientia. – 2017. – № 6. – С. 22–25.
6. Кадиев, А. Д. Нейросетевые методы решения задачи кредитного скоринга / А. Д. Кадиев, А. В. Чибисова // Математическое моделирование и численные методы. – 2022. – № 4 (36). – С. 81–92.
7. Богданов, А. Л. Применение нейронных сетей в решении задачи кредитного скоринга / А. Л. Богданов, И. С. Дуля // Вестник Томского государственного университета. Экономика. – 2018. – № 44. – С. 173–183.
8. Полищук, Ф. С. Кредитный скоринг: разработка рейтинговой системы оценки риска кредитования физических лиц / Ф. С. Полищук, А. Ю. Романов // Новые информационные технологии в автоматизированных системах. – 2016. – № 19. – С. 280–282.
9. Мельников, А. А. Разработка автоматизированной системы кредитного скоринга / А. А. Мельников, Д. С. Стельмаш, С. Н. Ефимов. – Текст : непосредственный // Актуальные проблемы авиации и космонавтики. – 2010. – № 6. – С. 233–234.
10. Применение методов машинного обучения для прогнозирования производительности доменной печи / Т. А. Барбасова, Е. В. Бауман, П. А. Самолетова, С. А. Черепанова // Известия ТулГУ. Технические науки. – 2023. – № 9. – С. 48–51.
11. Применение нейронной сети для определения режимов работы доменной печи / Т. А. Барбасова, Е. В. Бауман, П. А. Самолетова, С. А. Черепанова // Математическое и программное обеспечение систем в промышленной и социальной сферах. – 2021. – Т. 9, № 2. – С. 17–20.
12. Свидетельство о регистрации программы для ЭВМ RU 2022681794. Программа анализа русскоязычных текстов с выделением некоторых синтаксических зависимостей : Заявка № 2022681036 от 10.11.2022 / Н. М. Япарова, Ю. В. Перцев.
13. Свидетельство о регистрации программы для ЭВМ RU 2022681794, 16.11.2022. Ранжирование депозитных вкладов в разных валютах методом EDAS : Заявка № 2022681036 от 10.11.2022 / К. Н. Кудрявцев, П. К. Симаков.
14. Япаров, Д. Д. Разработка нейросетевой модели для управления светофором / Д. Д. Япаров, П. А. Бурьянов // Успехи кибернетики. – 2024. – Т. 5, № 4. – С. 122–127.