

ЦИФРОВЫЕ РЕШЕНИЯ И ТЕХНОЛОГИИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Научно-практический рецензируемый журнал
Издается с 2025 г.

**Издатель: Финансовый университет
при Правительстве Российской Федерации,
Москва, Российская Федерация**

125167, Москва, Ленинградский пр-т, д. 53

Периодичность издания — 4 номера в год

Журнал ориентирован на научное обсуждение актуальных проблем
по научным специальностям:

- 1.2.1. «Искусственный интеллект и машинное обучение»
- 1.2.2. «Математическое моделирование, численные методы и комплексы программ»
- 2.3.6. «Методы и системы защиты информации, информационная безопасность»
- 5.2.2. «Математические, статистические и инструментальные методы в экономике»

Электронная версия журнала на русском и английском языках
находится в открытом доступе на сайте digitari.ru

Журнал публикует материалы на условиях лицензии
Creative Commons Attribution 4.0 International (CC BY 4.0).

DIGITAL SOLUTIONS AND ARTIFICIAL INTELLIGENCE TECHNOLOGIES

Scientific and practical peer-reviewed journal
Published since 2025

ЦИФРОВЫЕ РЕШЕНИЯ И ТЕХНОЛОГИИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Научно-практический
журнал

Том 1, № 3, 2025

Периодичность:

4 раза в год

ISSN 3033-5477 (Print)

Свидетельство о регистрации:

Эл № ФС77-90338

от 01 ноября 2025 г.

Издатель:

Финансовый университет
при Правительстве
Российской Федерации
125167, Российская
Федерация, Москва,
Ленинградский проспект, 53
+7 (499) 553-10-74
dvonegov@fa.ru
Сайт: digitari.ru

Главный редактор
В.Г. Феклин

Заведующий редакцией
Д.В. Онегов

Выпускающий редактор
А.А. Извекова

Переводчик
И.М. Евстратов

Корректор
С.Ф. Михайлова

Верстальщик
Е.А. Смирнова

Подписано в печать:
11.08.2025

Формат 60 x 84 1/8.

Объем 11 п. л.

Отпечатано в ООО «СТ»,
г. Воронеж

Заказ № 2505460

Выход в свет 15.09.2025

© Финансовый университет
при Правительстве
Российской Федерации, Москва

ТЕМА НОМЕРА: ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И МАШИННОЕ ОБУЧЕНИЕ

Кузнецова А.В., Борисова Л.Р., Постовалова Г.А.

Методы машинного обучения для прогнозирования
ожидаемой продолжительности жизни 6

Болтачев Э.Ф., Фархадов М.П., Тюляков А.И.

Современные методы токенизации текстов
в финансовой сфере 19

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ, ЧИСЛЕННЫЕ МЕТОДЫ И КОМПЛЕКСЫ ПРОГРАММ

Кочкаров Р.А., Кочкаров А.А.

Динамические графы и их некоторые приложения. . . 30

Павлов Д.А.

Прогнозирование успешности проектов
на основе анализа структурных характеристик
коммуникационных сетей 37

МЕТОДЫ И СИСТЕМЫ ЗАЩИТЫ ИНФОРМАЦИИ, ИНФОРМАЦИОННАЯ БЕЗОПАСНОСТЬ

Иванов А.В., Царегородцев А.В.

Кибербезопасность цифрового рубля в системе
деструктивных событий цифрового пространства. . . . 44

Рыженко А.А., Козьминых С.И.

Недекларируемые возможности файловой
архитектуры: графические контейнеры 55

Когтева А.Н.

Организационные аспекты применения
искусственного интеллекта в сфере обороны. 62

МАТЕМАТИЧЕСКИЕ, СТАТИСТИЧЕСКИЕ И ИНСТРУМЕНТАЛЬНЫЕ МЕТОДЫ В ЭКОНОМИКЕ

Остапенко Г.А., Рожкова Г.Г., Феклин В.Г., Кочкаров Р.А.

Анализ публикационной активности
и научных коллабораций научно-педагогических
работников Финансового университета. 69

Каган Д.З., Рылов А.А.

Кластерный анализ регионов Российской Федерации
по спросу на транспортные услуги 77

COVER STORY: ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

Kuznetsova A.V., Borisova L.R., Postovalova G.A.

Machine Learning Methods
for Predicting Life Expectancy 6

Boltachev E.F., Farhadov M.P., Tyulyakov A.I.

Modern Tokenization Methods for Text Processing
in the Financial Domain 19

MATHEMATICAL MODELING, NUMERICAL METHODS AND SOFTWARE PACKAGES

Kochkarov R.A., Kochkarov A.A.

Dynamic Graphs and Some of Their Applications 30

Pavlov D.A.

Forecasting the Success of Projects Based
on the Analysis of Structural Characteristics
of Communication Networks 37

METHODS AND SYSTEMS OF INFORMATION PROTECTION, INFORMATION SECURITY

Ivanov A.V., Tsaregorodtsev A.V.

Cybersecurity of the Digital Ruble in the System
of Destructive Events of the Digital Space 44

Ryzhenko A.A., Kozminykh S.I.

Undeclared File Architecture Features:
Graphical Containers 55

Kogteva A.N.

Organizational Aspects of the Use Artificial Intelligence
in the Field of Defense 62

MATHEMATICAL, STATISTICAL AND INSTRUMENTAL METHODS IN ECONOMICS

Ostapenko G.A., Rozhkova G.G., Feklin V.G., Kochkarov R.A.

Thematic Analysis of Publication Activity Among
Academic and Teaching Staff: A Case Study
of the Financial University 69

Kagan D.Z., Rylov A.A.

Cluster Analysis Russian Federation Regions
by Demand for Transport Services 77

DIGITAL SOLUTIONS AND ARTIFICIAL INTELLIGENCE TECHNOLOGIES

*Scientific and practical
journal*

Vol. 1, No. 3, 2025

Publication Frequency:
4 times a year

ISSN 3033-5477 (Print)

Publisher:

Financial University under
the Government of the
Russian Federation
49 Leningradsky Prospekt,
Moscow, 125167, Russian
Federation

+7 (499) 553-10-74

dvonegov@fa.ru

Site: **digitari.ru**

Editor-in-Chief

Vadim G. Feklin

Head of Editorial
Department

Dmitry V. Onegov

Managing Editor

Anna A. Izvekova

Translator

Igor M. Evstratov

Proofreader

Svetlana F. Mikhaylova

Design, make up
Elena A. Smirnova

Signed off to printing: 11.08.2025

Format 60 x 84 1/8.

Size 10 printer sheets.

Printed by ST Ltd, Voronezh

Order № 2505460

Publication date 15.09.2025

© Financial University

under the Government

of the Russian Federation,

Moscow

ГЛАВНЫЙ РЕДАКТОР

ФЕКЛИН В.Г., кандидат физико-математических наук, доцент, декан факультета информационных технологий и анализа больших данных, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

ЗАМЕСТИТЕЛЬ ГЛАВНОГО РЕДАКТОРА

КОЧКАРОВ Р.А., кандидат экономических наук, доцент кафедры искусственного интеллекта, заместитель декана по научной работе факультета информационных технологий и анализа больших данных, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

ЧЛЕНЫ РЕДАКЦИОННОЙ КОЛЛЕГИИ

АНДРИЯНОВ Н.А., кандидат технических наук, доцент, доцент кафедры искусственного интеллекта, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

АФНАСЬЕВ А.А., доктор экономических наук, доцент, ведущий научный сотрудник ЦЭМИ РАН, заведующий кафедрой моделирования и системного анализа, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

БОГДАНОВ Е.А., Ph.D, заведующий кафедрой информационной безопасности, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

ВАСИЛЬЕВА Е.В., доктор экономических наук, доцент, заведующая кафедрой бизнес-информатики, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

ЖУКОВСКАЯ И.Е., доктор экономических наук, доцент, профессор кафедры бизнес-информатики, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

КОРОВИН Д.И., доктор экономических наук, доцент, профессор кафедры моделирования и системного анализа, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

КАШИРИН А.Б., директор Центра продвинутой аналитики АО «Альфа-Банк», заведующий базовой кафедрой Альфа-Банка, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

КОЗЬМИНЫХ С.И., доктор технических наук, доцент, профессор кафедры информационной безопасности, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

КОРОТЕЕВ М.В., кандидат экономических наук, доцент, заведующий кафедрой искусственного интеллекта, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

КОЧКАРОВ А.А., доктор технических наук, доцент, заместитель директора по инновационной работе ФИЦ Биотехнологии РАН, профессор кафедры искусственного интеллекта, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

МАКАРОВ В.Л., доктор физико-математических наук, профессор, академик РАН, научный руководитель ЦЭМИ РАН, научный руководитель факультета информационных технологий и анализа больших данных, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

МЕЛЬНИКОВ Д.А., доктор технических наук, доцент, профессор кафедры информационной безопасности, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

МИХАЙЛОВА С.С., доктор экономических наук, заведующая кафедрой математики и анализа данных, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

НЕИЗВЕСТНЫЙ С.И., доктор технических наук, старший научный сотрудник, профессор кафедры бизнес-информатики, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

ОСТАПЕНКО Г.А., доктор технических наук, профессор, проректор по цифровизации, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

ПЕТРОВСОВ Д.А., кандидат технических наук, доцент, заведующий кафедрой информационных технологий, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

ПРОКОПЧИНА С.В., доктор технических наук, профессор, профессор кафедры искусственного интеллекта, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

РЯБОВ П.Е., доктор физико-математических наук, доцент, профессор кафедры математики и анализа данных, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

СЛАВИН Б.Б., доктор экономических наук, профессор кафедры бизнес-информатики, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

ТИМОШЕНКО А.В., доктор технических наук, профессор, профессор кафедры информационной безопасности, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

ЦАРЕГОРОДЦЕВ А.В., доктор технических наук, профессор, главный научный сотрудник Института цифровых технологий, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

ЩЕТИНИН Е.Ю., доктор физико-математических наук, доцент, профессор кафедры искусственного интеллекта, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

POURMOHAMMADBAGHER L., Dr., Allameh Tabataba'i University (ATU), Iran

BAHRANI M., Dr., Allameh Tabataba'i University (ATU), Iran

PEYMANI M., Dr., Allameh Tabataba'i University (ATU), Iran

SHARMA CH., Prof., GNIOT Institute Of Management Studies, Greater Noida, India

JUNSHENG ZH., Prof., Dalian Neusoft University of Information, China

LIANZHUANG Q., PhD, Dalian Neusoft University of Information, China

KANG L., Prof., Dalian Neusoft University of Information, China

EDITOR-IN-CHIEF

FEKLIN V. G., Cand. Sci. (Phys.-Math.), Associate Professor, Dean of the Faculty of Information Technology and Big Data Analysis, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

DEPUTY EDITOR-IN-CHIEF

KOCHKAROV R.A., Cand. Sci. (Econ.), Associate Professor, Department of Artificial Intelligence, Deputy Dean for Research at the Faculty of Information Technology and Big Data Analysis, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

MEMBERS OF THE EDITORIAL BOARD

ANDRIANOV N.A., Cand. Sci. (Tech.), Associate Professor, Associate Professor of Artificial Intelligence, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

AFANASYEV A.A., Dr. Sci. (Econ.), Associate Professor, Leading Researcher at CEMI RAS, Head of the Department of Modeling and System Analysis, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

BOGDANOV E.A., PhD, Head of the Department of Information Security, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

VASILYEVA E.V., Dr. Sci. (Econ.), Associate Professor, Head of the Department of Business Informatics, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

ZHUKOVSKAYA I.E., Dr. Sci. (Econ.), Associate Professor, Professor, Department of Business Informatics, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

KOROVIN D.I., Dr. Sci. (Econ.), Associate Professor, Professor of the Department of Modeling and System Analysis, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

KASHIRIN A.B., Director of the Center for Advanced Analytics of Alfa-Bank JSC, Head of the Basic Department of Alfa-Bank, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

KOZMINYKH S.I., Dr. Sci. (Tech), Associate Professor, Professor, Department of Information Security, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

KOROTEEV M.V., Cand. Sci. (Econ.), Associate Professor, Head of the Department of Artificial Intelligence, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

KOCHKAROV A.A., Dr. Sci. (Tech), Associate Professor, Deputy Director for Innovation at the Institute of Biotechnology of the Russian Academy of Sciences, Professor, Department of Artificial Intelligence, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

MAKAROV V.L., Dr. Sci. (Phys.-Math.), Professor, Academician of the RAS, Scientific Director of the Central Research Institute of the Russian Academy of Sciences, Scientific Director of the Faculty of Information Technology and Big Data Analysis, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

MELNIKOV D.A., Dr. Sci. (Tech), Associate Professor, Professor, Department of Information Security, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

MIKHAILOVA S.S., Dr. Sci. (Econ.), Head of the Department of Mathematics and Data Analysis, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

NEIZVESTNYJ S.I., Dr. Sci. (Tech), Senior Researcher, Professor, Department of Business Informatics, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

OSTAPENKO G.A., Dr. Sci. (Tech), Professor, Vice-Rector for Digitalization, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

PETROSOV D.A., Cand. Sci. (Tech.), Associate Professor, Head of the Department of Information

Technology, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

PROKOPCHINA S.V., Dr. Sci. (Tech), Professor, Professor, Department of Artificial Intelligence, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

RYABOV P.E., Dr. Sci. (Phys.-Math.), Associate Professor, Professor of the Department of Mathematics and Data Analysis, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

SLAVIN B.B., Dr. Sci. (Econ.), Professor, Department of Business Informatics, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

TIMOSHENKO A.V., Dr. Sci. (Tech), Professor, Professor, Department of Information Security, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

TSAREGORODTSEV A.V., Dr. Sci. (Tech.), Professor, Chief Researcher at the Institute of Digital Technologies, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

SHCHETININ E.Y., Dr. Sci. (Phys.-Math.), Associate Professor, Professor, Department of Artificial Intelligence, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

POURMOHAMMADBAGHER L., Dr., Allameh Tabataba'i University (ATU), Iran

BAHRANI M., Dr., Allameh Tabataba'i University (ATU), Iran

PEYMANI M., Dr., Allameh Tabataba'i University (ATU), Iran

SHARMA CH., Prof., GNIOT Institute Of Management Studies, Greater Noida, India

JUNSHENG ZH., Prof., Dalian Neusoft University of Information, China

LIANZHUANG Q., PhD, Dalian Neusoft University of Information, China

KANG L., Prof., Dalian Neusoft University of Information, China

Manuscripts are submitted vfeklin@fa.ru

Requirements for the design of copyrighted materials:

<https://disk.yandex.ru/i/4TYJzMLvb1gQ-Q>

The Editorial Board assesses the peer-reviewed manuscripts meticulously and executes scientific, literary and technical editing of the author's original in the journal.

DOI: 10.26794/3033-7097-2025-1-3-6-18
УДК 616.98:578.834.1(045)

Методы машинного обучения для прогнозирования ожидаемой продолжительности жизни

А.В. Кузнецова^а, Л.Р. Борисова^б, Г.А. Постовалова^с^а Институт биохимической физики им. Н.М. Эмануэля, Российская академия наук, Москва, Российская Федерация;^{б,с} Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

АННОТАЦИЯ

Прогнозирование ожидаемой продолжительности жизни связано не только с серьезными социальными и финансовыми факторами, но и с состоянием общественного здравоохранения и экономики, а также с состоянием окружающей среды. Использование математических методов позволяет выделить наиболее информативные показатели, влияющие на продолжительность жизни. **Целью** работы является прогнозирование ожидаемой продолжительности жизни по данным Всемирного банка * методами машинного обучения (МО), а также сравнение эффективности прогнозирования ожидаемой продолжительности жизни с использованием различных алгоритмов машинного обучения, включая такие широко распространенные методы, как метод опорных векторов, дерево решений, случайный лес, линейный дискриминант Фишера, нейронные сети, два варианта градиентного бустинга, логистическую регрессию и метод статистически взвешенных синдромов. База данных включила данные для 238 стран. Применены стандартные непараметрические критерии хи-квадрат (χ^2) и критерий Манна-Уиттн (U-тест). Выявлены 11 значимых показателей. Использованы методы машинного обучения (МО) системы анализа данных Data Master Azforus. Результат прогнозирования *методом статистически взвешенных синдромов* (СВС) достиг значения ROC, AUC = 0,986. Представлены одномерные и двумерные диаграммы взаимосвязи исследуемых социально-экономических и медицинских показателей на ожидаемую продолжительность жизни. Из этих диаграмм можно получать прогноз для изменения отдельных показателей с целью улучшения качества и продолжительности жизни. Таким образом, система анализа данных Data Master Azforus (САД ДМА) позволит исследователям создавать рекомендательные системы по прогнозированию ожидаемой продолжительности жизни. Кроме того, проведенное исследование поможет создать более совершенную систему прогнозирования с использованием моделей машинного обучения, которые могут служить руководством для политиков при улучшении прогнозирования ожидаемой продолжительности жизни.

Ключевые слова: показатели; заболевания; регрессия; коэффициент детерминации; машинное обучение; прогностические модели

Для цитирования: Кузнецова А.В., Борисова Л.Р., Постовалова Г.А. Методы машинного обучения для прогнозирования ожидаемой продолжительности жизни. *Цифровые решения и технологии искусственного интеллекта*. 2025;1(3):6-18. DOI: 10.26794/3033-7097-2025-1-3-6-18

ORIGINAL PAPER

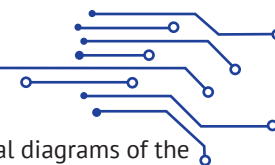
Machine Learning Methods for Predicting Life Expectancy

A.V. Kuznetsova^а, L.R. Borisova^б, G.A. Postovalova^с^а Institute of Biochemical Physics of the Russian Academy of Sciences, Moscow, Russian Federation;^{б,с} Financial University under the Government of the Russian Federation, Moscow, Russian Federation

ABSTRACT

Forecasting of life expectancy is associated not only with serious social and financial factors, but also with the state of public health and economy, as well as with the state of the environment. The use of mathematical methods makes it possible to identify the most informative indicators affecting life expectancy. The aim of the paper is to predict life expectancy from World Bank data using machine learning (ML) methods, and to compare the effectiveness of life expectancy prediction using different machine learning algorithms, including such widely used methods as support vector method, decision tree, random forest, Fisher's linear discriminant, neural networks, two variants of gradient boosting, logistic regression and statistically weighted syndrome method. The database included data for 238 countries. Standard non-parametric chi-square (χ^2) and Mann-Whitney criteria (U-test) were applied. Eleven significant indicators were identified. Machine learning (ML) methods of Data Master Azforus data analysis system was used. The prediction result of the statistically

* URL: <https://data.worldbank.org/indicator/SP.DYN.LE.O0.IN>



weighted syndrome (SWS) method achieved a ROC AUC = 0.986. One-dimensional and two-dimensional diagrams of the relationship between the studied socio-economic and medical indicators on life expectancy are presented. From these charts, predictions can be derived for changes in individual indicators to improve quality and length of life. Thus, the Data Master Azforus data analysis system will enable researchers to create recommendation systems for life expectancy prediction. In addition, the conducted research will help to create a more advanced forecasting system using machine learning models that can serve as a guide for politicians makers in improving life expectancy forecasting.

Keywords: indicators; diseases; regression; coefficient of determination; machine learning; predictive models

For citation: Kuznetsova A.V., Borisova L.R., Postovalova G.A. Machine learning methods for predicting life expectancy. *Digital solutions and artificial intelligence technologies*. 2025;1(3):6-18. DOI: 10.26794/3033-7097-2025-1-3-6-18

ВВЕДЕНИЕ

Увеличение продолжительности жизни и повышение ее качества являются важными показателями развития каждой страны мира. Предшествующие им цели включают проведение правильной социально-экономической политики, направленной на улучшение базовых основ жизни: качества и безопасности питьевой воды, увеличения площади орошаемых земельных угодий, искоренения бедности и социального неравенства, повышения доходов домохозяйств, увеличения доступности квалифицированной медицинской помощи, предотвращение распространения социально-значимых инфекций, повышения уровня жизни в целом.

Как правило, для решения этих задач используется математическое моделирование. Самым распространенным методом является построение регрессионных моделей. В частности, в работе С. П. Куккоева [1] получены коэффициенты корреляции ожидаемой продолжительности жизни (ОПЖ) с основными показателями уровня жизни и поведения, а также математические модели ожидаемой продолжительности жизни при рождении. Определены условия достижения ОПЖ 75, 80 лет в России.

Стоит отметить интересную эконометрическую модель, представленную в работе [2], где авторами описан оригинальный метод построения интегрального показателя жизни в регионах Российской Федерации в зависимости от основных экономических показателей. Коэффициент детерминации данной модели 0,95.

Авторы О.В. Куделина, М.А. Канева описали зависимость между ожидаемой продолжительностью жизни и набором показателей, характеризующих здоровье населения. Они использовали математический аппарат, связанный с построением функций полезности [3].

Регрессионная модель взаимосвязи распространенности онкологических заболеваний и санитарно-гигиенических показателей представлена в работе В. С. Степанова. Коэффициент детермина-

ции данной модели 0,985, ошибка аппроксимации менее 0,02 [4].

SK Khader Basha S. K. и его соавторы отмечают важность прогноза средней продолжительности жизни населения страны, чтобы проанализировать дальнейшие потребности в увеличении или стабилизации темпов роста населения в конкретной стране [5]. Основной целью данной работы стал анализ влияния различных социально-экономических показателей на результат и его вариативность. Показано, что модель «random forest» продемонстрировала наилучшие результаты. Коэффициент детерминации 0,96. Смертность среди взрослых, ВИЧ, школьное образование и индекс массы тела (ИМТ) являются наиболее значимыми факторами, влияющими на ожидаемую продолжительность жизни. Некоторые показатели, такие как ВВП, общие расходы и детская смертность, не оказали существенного влияния на итоговый показатель.

В статье [6] описаны результаты использования композиционного анализа данных для учета согласованности между здоровьем и смертностью. Смертность и состояние здоровья в возрасте 50 лет и старше прогнозируются для женщин Франции, Испании, Швеции и Великобритании. Bergeron Boucher V.-P. и соавторы уверены, что существует четкое обоснование для прогнозирования ожидаемой продолжительности здоровой жизни, так как прогнозы поддерживают социальные, экономические и медицинские решения, а также индивидуальный выбор каждого человека.

В статье [7] приводятся данные, свидетельствующие в пользу учета показателей заболеваемости при прогнозировании уровня жизни. В период 2019–2020 гг. в 27 из 29 стран снизилась ожидаемая продолжительность жизни. Наибольшие потери зафиксированы в 2020 г. в США и Литве: где ожидаемая продолжительность жизни среди мужчин сократилась (2,2 и 1,7 года соответственно); еще в 11 странах зафиксировано сокращение жизни более чем на целый год (среди мужчин) и в 8 (среди женщин). Эта статистика связана

с официально зафиксированными случаями смерти от COVID-19 людей старше 60 лет.

Ожидаемая продолжительность жизни — статистический показатель среднего ожидаемого срока жизни человека, основанный на годе его рождения, текущем возрасте и других демографических факторах, включая пол. В математическом выражении ожидаемая продолжительность жизни — это ожидаемое количество лет, оставшееся человеку в любом возрасте.

Прогноз средней продолжительности жизни очень важен, он позволяет проанализировать дальнейшие потребности в увеличении темпов роста или стабилизации темпов роста в конкретно взятой стране или регионе. Ранее авторы настоящей статьи принимали участие в глобальных исследованиях по изучению основных факторов развития стран [8], в которых доказано, что статистический анализ убедительно подтверждает использование методов машинного обучения при выявлении статистически значимых взаимосвязей между различными показателями, характеризующими развитие стран, даже при наличии пробелов в данных. Были определены статистически значимые 13 из рассмотренных 14 социально-экономических и демографических показателей. Использование методов машинного обучения позволило выявить особенности влияния характеристик общественной среды на результаты институционального дизайна в отношении двух видов форм: заимствованных из иной институциональной среды и опирающихся на национальные исторические традиции, что зафиксировано при эмпирической проверке теории институциональных матриц методами интеллектуального анализа данных [9].

Еще в одной работе [10] описан анализ двух форм социально-экономического уклада стран с точки зрения институциональных матриц (X и Y), применительно к двум общественным сферам (экономика, социология). Исследование проводилось в расширенном понимании взаимодействия стран X и Y с другими странами как источниками или потребителями ресурсов.

Исследование, описываемое в настоящей статье, продолжение предыдущих, но с акцентом на рассмотрение медицинских показателей, не учитываемых ранее. **Цель** настоящего исследования — анализ влияния основных социально-экономических и медицинских показателей на ожидаемую продолжительность жизни по открытым данным. Для этого на первом этапе необходимо выявить наиболее эффективные методы машинного обучения.

МАТЕРИАЛЫ И МЕТОДЫ

База данных. В исследовании использованы открытые данные сайта Всемирного банка для 238 стран. Проанализировано влияние 12 показателей на ожидаемую продолжительность жизни: это индекс Джини; процент населения, использующего безопасную воду; заболеваемость туберкулезом и ВИЧ, смертность от сердечно-сосудистых заболеваний, онкологии, диабета, хронических респираторных заболеваний как мужчин, так и женщин; неонатальная смертность; пожизненный риск материнской смертности; общее потребление алкоголя женщинами; годовой прирост населения; инфляция, измеряемая с помощью индекса потребительских цен, и среднедушевой доход. Страны поделены на две группы: по 119 стран в каждой. В первую группу отнесены страны с низкой ожидаемой продолжительностью жизни, во вторую — с высокой.

Методы анализа. Наряду с традиционными статистическими методами, включая U-тест Манна-Уитни и критерий χ^2 , в работе применяли альтернативные методы интеллектуального анализа данных и методы машинного обучения (МО) [11].

Метод оптимально достоверных разбиений (ОДР) — оригинальная технология анализа данных, основанная на построении оптимальных разбиений пространства признаков.

Использовались также известные основанные на машинном обучении *методы автоматической классификации*: решающих деревьев, статистически взвешенных синдромов, логистической регрессии, варианты градиентного бустинга (XGBoost).

Все они сохраняют эффективность при небольших выборках. Для сравнения приведены также расчеты с использованием нейросетевых технологий, чтобы подчеркнуть их неэффективность на малых выборках.

Метод скользящего контроля (LeaveOneOut) применялся для оценки статистической значимости найденных закономерностей, а *показатели чувствительности, специфичности, общей точности и ROC-анализ* — для оценки значимости. Все перечисленные методы исследования реализованы в системе анализа данных Data Master Azforus, используемой в данной работе [12].

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ И ИХ ОБСУЖДЕНИЕ

На первом этапе исследования сравниваемые группы анализировали с помощью стандартного статистического метода U-тест (табл. 1).

Таблица 1 / Table 1

Сравнение исследуемых групп стран с применением метода U-тест / Comparison of the Studied Groups of Countries using the U-test

Показатель / Indicator	Среднее значение показателя / The average value of the indicator		p-значение / p-value
	группа 1 / group 1	группа 2 / group 2	
Индекс Джини	35,236287	35,090985	0,019354
Использование безопасной воды, %	57,187240	83,824287	0,000000
Заболеваемость туберкулезом (на 100 000 чел.)	171,039604	34,741044	0,000000
Смертность от сердечно-сосудистых заболеваний, рака, сахарного диабета или хронических респираторных заболеваний в возрасте от 30 до 70 лет:			
женщины, %	21,054244	11,802025	0,000000
мужчины, %	25,275356	17,826471	0,000000
Годовой темп прироста населения, %	1,428162	0,732795	0,000000
Коэффициент неонатальной смертности, %	18,661062	5,635258	0,000000
Распространенность ВИЧ, %	2,089426	0,882675	0,001276
Общее потребление алкоголя женщинами на душу населения (чистый алкоголь, л; женщины старше 15 лет)	1,404183	2,684545	0,000000
Пожизненный риск материнской смертности, %	0,808652	0,102283	0,000000
Инфляция, измеряемая с помощью индекса потребительских цен, %	9,669168	8,769882	0,015451
ВВП на душу населения, долл. США	5483,116148	33 745,666421	0,000000

Источник / Source: составлено авторами / Compiled by the authors.

Таблица 1 демонстрирует статистическую значимость признаков (p близко к нулю). Интересно отметить, что такие интегральные социально-экономические показатели, как индекс Джини (характеристика социального неравенства) и инфляция, измеряемая индексом потребительских цен, показали самое большое p -значение при разделении стран на две группы. Этот результат был подтвержден при использовании других стандартных статистических тестов. В табл. 2 приведены результаты разделения стран на две группы с использованием критерия χ^2 .

Страны хорошо разделяются по выбранным показателям. Для всех показателей, за исключением

показателей «Индекс Джини» и «Инфляция», p -значение практически равно нулю, что свидетельствует в пользу статистически значимого разделения совокупности на две группы.

ОЦЕНКА ИНФОРМАТИВНОСТИ ПОКАЗАТЕЛЕЙ

Для изучения закономерностей — связи конкретной группы (1 или 2) с изученными показателями, использовали метод ОДР. Результаты представлены в *табл. 3*.

Из результатов применения метода ОДР, представленных в *табл. 3*, следует, что страны четко

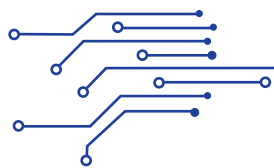


Таблица 2 / Table 2

Сравнение исследуемых групп стран с применением традиционных статистических методов /
Comparison of The Studied Countries Groups Using Traditional Statistical Methods

Показатель / Indicator	Среднее значение показателя / The average value of the indicator		Стандартное отклонение показателя / standard deviation of the indicator		Метод / Method	
	группа 1 / group 1	группа 2 / group 2	группа 1 / group 1	группа 2 / group 2	χ^2	U-test
1. Индекс Джини	35,5	34,99	0,2087	0,4381	0,956	0,000
2. Использования безопасной воды, %	43,18	88,93	1,574	1,161	0,000	0,000
Заболеваемость туберкулезом (на 100 000 чел.)	176,6	26,38	14,55	3,315	0,000	0,000
3. Смертность от сердечно-сосудистых заболеваний, рака, сахарного диабета или хронических респираторных заболеваний в возрасте от 30 до 70 лет:						
женщины, %	21,39	10,75	0,4568	0,2828	0,000	0,000
мужчины, %	25,54	16,98	0,6069	0,5083	0,000	0,000
4. Годовой темп прироста населения, %	1,428	0,7328	0,1448	0,1314	0,001	0,000
5. Коэффициент неонатальной смертности, %	18,77	4,55	0,8024	0,2831	0,000	0,000
6. Распространенность ВИЧ, %	2,267	0,309	0,339	0,01783	0,000	0,000
7. Общее потребление алкоголя женщинами на душу населения (чистый алкоголь, л., прогнозные оценки, женщины старше 15 лет)	1,376	2,846	0,1097	0,1489	0,000	0,000
8. Пожизненный риск материнской смертности, %	0,8147	0,03518	0,08252	0,003735	0,000	0,000
9. Инфляция, измеряемая с помощью индекса потребительских цен, %	9,777	8,674	0,833	1,897	0,908	0,002
10. ВВП на душу населения, долл.	4327	35 185	404,9	3065	0,000	0,000

Источник / Source: составлено авторами / Compiled by the authors.

разделяются по таким основным показателям, как ВВП на душу населения и процент использования населением безопасной воды. И, совсем неожиданно, к этим двум показателям присоединился показатель, характеризующий потребление алкоголя женщинами. Таким образом, умеренное потребление алкоголя вносит в разделение стран на 2 группы такой же вклад, как безопасная вода и душевые доходы.

На рис. 1 продемонстрирован метод ОДР на примере показателя «Смертность от сердечно-сосудистых заболеваний (ССЗ), рака, сахарного диабета или хронических респираторных заболеваний (ХРЗ) в возрасте

от 30 до 70 лет, женщины, %», где группа 1 (с низкой ожидаемой продолжительностью жизни) обозначена красными крестиками, а группа 2 (с высокой ожидаемой продолжительностью жизни) — зелеными кружками. Граница, разделяющая неблагополучные страны от благополучных, проходит на уровне 16,47% (смертность от основных болезней в трудоспособном возрасте). Из рис. 2 следует, что в квадранте I преобладает группа 2, в квадранте II — группа 1. Интерпретация данного результата базируется на том, что для повышения ожидаемой продолжительности жизни необходимо уменьшать коэффициент неонатальной смертности ниже уровня 8,75%.

Преобладание групп / The predominance of groups

Показатель / Indicator	Преобладающая группа в квадранте I / The predominant group in the quadrant I	Граница / Border	Преобладающая группа в квадранте II / The predominant group in the second quadrant II	p-значение / p-Value
1. Годовой темп прироста населения, %	2	1,52	1	0
2. Использование безопасной воды, %	1	66,81	2	0
3. Заболеваемость туберкулезом (на 100 000 чел.)	2	58,5	1	0
4. Смертность от сердечно-сосудистых заболеваний, рака, сахарного диабета или хронических респираторных заболеваний в возрасте от 30 до 70 лет:				
женщины, %	2	16,47	1	0
мужчины, %	2	18,05	1	0
5. ВВП на душу населения, долл. США	1	7830	2	0
6. Коэффициент неонатальной смертности, %	2	8,75	1	0
7. Распространенность ВИЧ, %	2	0,75	1	0
8. Общее потребление алкоголя женщинами на душу населения (чистый алкоголь, л., прогнозные оценки, женщины старше 15 лет)	1	2,48	2	0
9. Пожизненный риск материнской смертности, %	2	0,1357	1	0
10. Инфляция, измеряемая с помощью индекса потребительских цен, %	2	12,59	1	0,03967

Источник / Source: составлено авторами / Compiled by the authors.

На одномерной диаграмме рассеяния *рис. 2* случаи из группы 1 (с низкой ожидаемой продолжительностью жизни) также обозначены красными крестиками, а случаи из группы 2 (с высокой ожидаемой продолжительностью жизни) — зелеными кружками.

Интересно посмотреть, как связь между показателями, представленными на одномерной диаграмме *рис. 2*, изменится в двумерном случае. Двумерные модели обычно более информативные. Результаты для двумерной модели, описывающей связь между коэффициентом неонатальной смертности и смертности трудоспособных женщин от основных заболеваний, представлены на *рис. 3*.

Как следует из *рис. 3*, в первых трех квадрантах преобладает группа 1. В квадранте IV — группа 2 (с высокой ожидаемой продолжительностью жизни). Для IV квадранта обязательно одновременное выполнение двух условий: десятичные логарифмы коэффициента неонатальной смертности и смертности от основных болезней должны быть меньше 0,942 и 1,217% соответственно.

Результаты двумерных моделей были получены для 36 пар показателей. Но наиболее информативные результаты (с *p*-значением, близким к нулю) зафиксированы у 15 пар и представлены в *табл. 4*.

Результаты двумерного исследования свидетельствуют в пользу статистически значимой классификации по выбранным показателям, за исключением

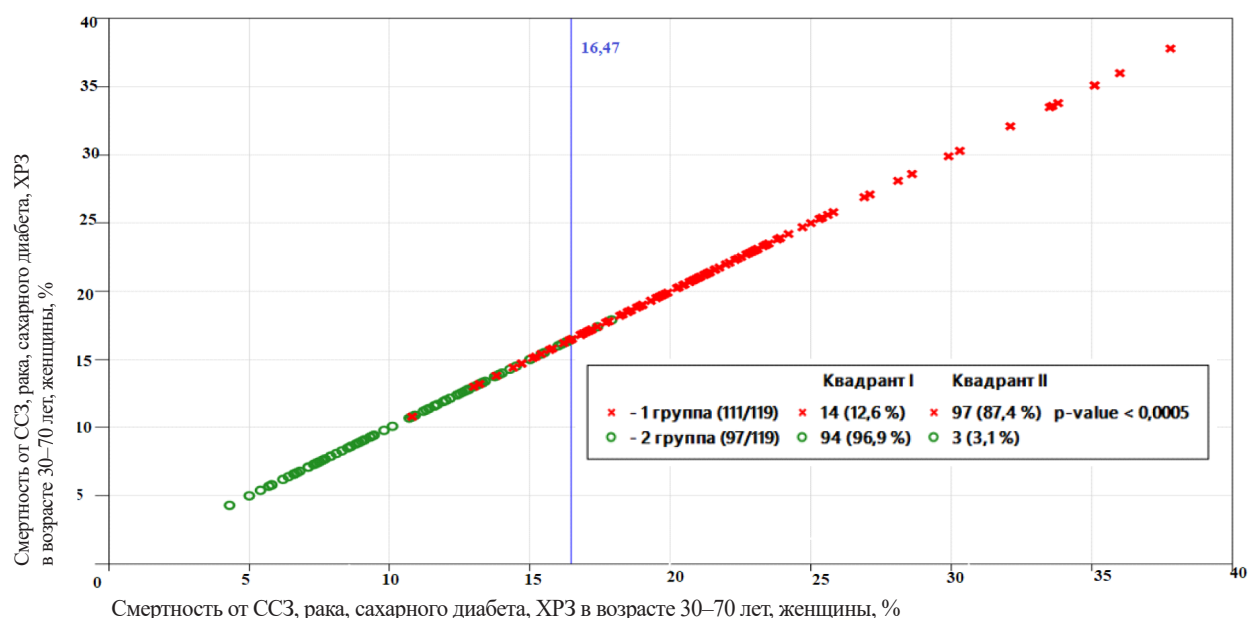


Рис. 1 / Fig. 1. Различие между случаями из группы 1 и группы 2 по показателю «Смертность от ССЗ, рака, сахарного диабета или ХРЗ заболеваний в возрасте от 30 до 70 лет, женщины, %», выявленные методом ОДР, $p < 0,001$ / The Difference Between Cases from Groups 1 and 2 in Terms of “Mortality from CVD, Cancer, Diabetes Mellitus or CRD aged 30 to 70 years, Women, %”, Detected by the SDT Method, $p < 0.001$

Источник / Source: составлено авторами / Compiled by the authors.

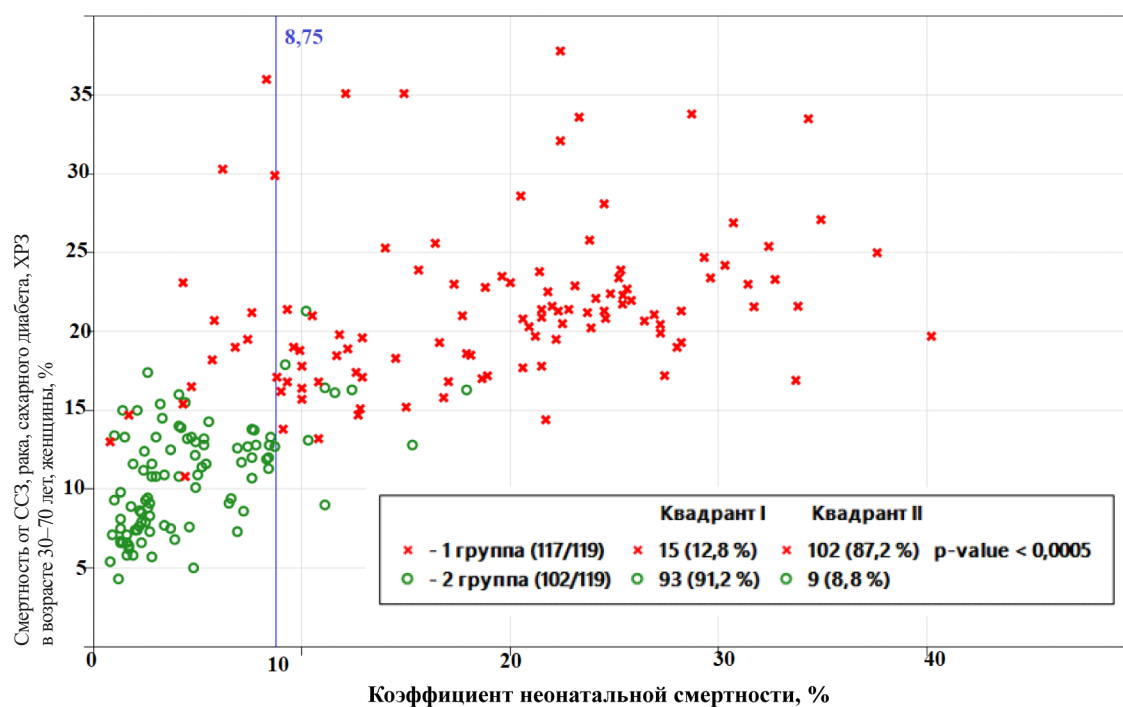


Рис. 2 / Fig. 2. Различия между случаями из группы 1 и 2 по сочетанию коэффициента неонатальной смертности (ось X) и смертностью от ССЗ, рака, сахарного диабета или ХРЗ в возрасте от 30 до 70 лет, женщины, %, ось Y / Differences Between Cases from Groups 1 and 2 in Terms of the Combination of Neonatal Mortality rate (X axis) and Mortality from CVD, Cancer, Diabetes Mellitus or CRD aged 30 to 70 years, Women, %, Y axis

Источник / Source: составлено авторами / Compiled by the authors.

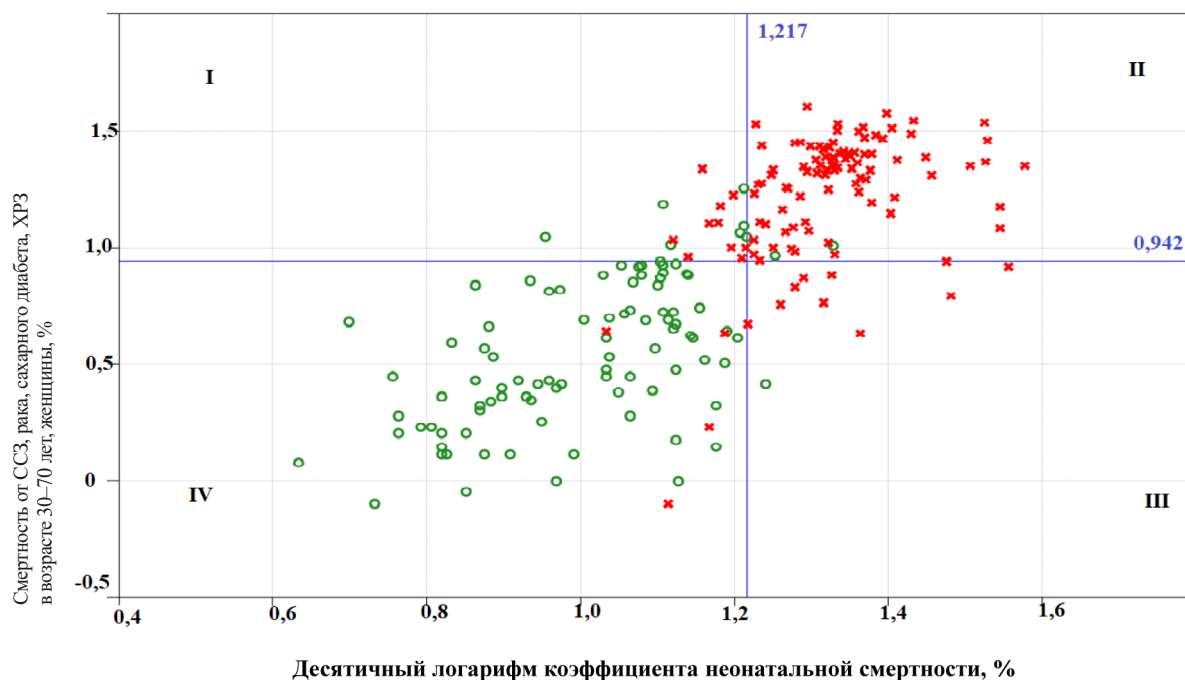


Рис. 3 / Fig. 3. Различия между случаями из группы 1 и 2 по сочетанию коэффициента неонатальной смертности (ось X) и смертностью от ССЗ, рака, сахарного диабета или ХРЗ в возрасте от 30 до 70 лет, женщины, %, ось Y / Mortality rate (X axis) and Mortality from CVD, Cancer, Diabetes Mellitus or CRD aged 30 to 70 years, Women, %, Y axis

Источник / Source: составлено авторами / Compiled by the authors.

Примечание / Note: Представлены десятичные логарифмы показателей по обеим осям / Differences Between Cases from Groups 1 and 2 in Terms of the Combination of Neonatal Decimal Logarithms of Indicators on Both Axes are Presented.

последней пары показателей, а именно, взаимосвязь показателей «Индекс Джини» и «Общее потребление алкоголя женщинами на душу населения». Стоит отметить, что именно для этих показателей было больше всего пропусков в анализируемых данных с сайта Всемирного банка.

РЕЗУЛЬТАТЫ АВТОКЛАССИФИКАЦИИ (AUTO-CLASSIFICATION RESULTS)

Результаты применения методов машинного обучения (МО), участвовавших в классификации, при сравнении двух исследуемых групп показали, что самым эффективным из них оказался оригинальный авторский метод статистически взвешенных синдромов (СВС). Он показал результат распознавания объектов при скользящем контроле: ROC AUC = 0,968. Также хороший результат распознавания получен для методов: логистическая регрессия, адаптивный бустинг и деревья решений. В табл. 5 приведены значения ROC AUC для 9 методов машинного обучения, полученные с помощью метода скользящего контроля (LeaveOneOut) системы анализа данных Data Master Azforus. Нейронная сеть не годится для распознавания (последняя строка

таблицы), если в базе данных ограниченное число объектов (например, 238 в нашем исследовании).

ВЫВОДЫ

Проведено сравнительное исследование 238 стран по 12 показателям по ожидаемой продолжительности жизни (**группа 1, ниже 74 лет, 119 стран**) и 119 стран с ожидаемой продолжительностью жизни больше 74 лет (**группа 2**). Использование стандартных статистических критериев χ^2 и U-теста позволило выделить 11 значимых социально-экономических и медицинских показателей. На методе СВС получен ROC AUC = 0,986. Стоит отметить, что в предыдущей работе [8] качество прогноза ROC AUC = 0,99, возможно, из-за учета 14 различных экономических и демографических показателей. Этот результат свидетельствует в пользу большего учета экономических показателей при прогнозировании ожидаемой продолжительности жизни.

Исследование подтвердило возможность использования технологий машинного обучения для оперативной оценки и прогнозирования ожидаемой продолжительности жизни.

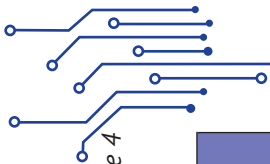


Таблица 4 / Table 4

Двумерные разбиения / Two-dimensional Partitions

№	Показатель / Indicator	χ^2	Граница / Border	<i>p</i> -значение / <i>p</i> -Value	Квадранты I и III / Quadrant I, III		Квадранты II и IV / Quadrant II, IV	
					распределение случаев разбиения	% случаев	распределение случаев разбиения	% случаев
1	Смертность от ССЗ, рака, сахарного диабета или ХРЗ в возрасте от 30 до 70 лет, женщины, %	162,8	16,47	0,0005	10 / 7	9 / 7,3	87 / 2	78,4 / 2,1
	Коэффициент неонатальной смертности, %	–	8,75	0,0005	10 / 1	9 / 1	4 / 86	3,6 / 89,6
2	Смертность от ССЗ, рака, сахарного диабета или ХРЗ в возрасте от 30 до 70 лет, женщины, %	157,6	14,34	0,0005	0 / 71	0 / 74,7	10 / 11	9,6 / 11,6
	ВВП на душу населения, долл. США	–	7830	0,0005	90 / 2	86,5 / 2,1	4 / 11	4 (3,8) / 11 (11,6)
3	Коэффициент неонатальной смертности, %	154,7	8,75	0,0005	6 / 0	5,1 / 0	80 / 1	68,4 / 1
	Пожизненный риск материнской смертности, %	–	0,1623	0,0005	22 / 8	18,8 / 8,1	9 / 90	7,7 / 90,9
4	Коэффициент неонатальной смертности, %	145,1	8,75	0,0005	4 / 80	3,6 / 81,6	10 / 4	9,1 / 4,1
	ВВП на душу населения, долл. США	–	7830	0,0005	86 / 4	78,2 / 4,1	10 / 10	9,1 / 10,2
5	Заболеваемость туберкулезом (на 100 000 чел.)	144,9	26	0,0005	10 / 8	9,8 / 8,9	87 / 3	85,3 / 3,3
	Смертность от ССЗ, рака, сахарного диабета или ХРЗ в возрасте от 30 до 70 лет, женщины, %	–	14,6	0,001	5 / 26	4,9 / 28,9	0 / 53	0 / 58,9
6	Смертность от ССЗ, рака, сахарного диабета или ХРЗ в возрасте от 30 до 70 лет, мужчины, %	144	33,85	0,0005	89 / 9	80,2 / 9,4	8 / 0	7,2 / 0
	Коэффициент неонатальной смертности, %	–	8,75	0,0005	6 / 0	5,4 / 0	8 / 87	7,2 / 90,6
7	Смертность от ССЗ, рака, сахарного диабета или ХРЗ в возрасте от 30 до 70 лет, мужчины, %	141,4	23,24	0,0005	41 / 3	41 (36,9) / 3 (3,1)	47 / 0	42,3 / 0
	Пожизненный риск материнской смертности, %	–	0,1357	0,0005	16 / 13	14,4 / 13,4	7 / 81	6,3 / 83,5
8	Смертность от ССЗ, рака, сахарного диабета или ХРЗ в возрасте от 30 до 70 лет, мужчины, %	137,3	18,05	0,0005	1 / 54	1 / 56,8	9 / 28	8,7 / 29,5
	ВВП на душу населения, долл. США	–	7830	0,0005	88 / 4	84,6 / 4,2	6 / 9	5,8 / 9,5

Продолжение таблицы 4 / Table 4 (continued)

№	Показатель / Indicator	χ^2	Граница / Border	p-значение / p-Value	Квадранты I и III / Quadrant I, III		Квадранты II и IV / Quadrant II, IV	
					распределение случаев разбиения	% случаев	распределение случаев разбиения	% случаев
	Пожизненный риск материнской смертности, %	–	0,1623	0,0005	16 / 7	14,8 / 7,5	14 / 85	13 / 91,4
10	Смертность от, рака, сахарного диабета или ХРЗ в возрасте от 30 до 70 лет, мужчины, %	111,9	17,02	0,0005	4 / 41	3,6 / 42,3	91 / 10	82 / 10,3
	Годовой темп прироста населения, %	–	0,4981	0,0005	14 / 28	12,6 / 28,9	2 / 18	1,8 / 18,6
11	Заболееваемость туберкулезом (на 100 000 чел.)	105,1	29,5	0,0005	11 / 19	10,8 / 21,1	83 / 9	81,4 / 10
	Смертность от ССЗ, рака, сахарного диабета или ХРЗ в возрасте от 30 до 70 лет, мужчины, %	–	18,05	0,0005	6 / 18	5,9 / 20	2 / 44	2 / 48,9
12	Использование безопасной воды, %	100,5	73,6	0,0005	48 / 1	85,7 / 1,4	5 / 9	8,9 / 12,2
	Смертность от ССЗ, рака, сахарного диабета или ХРЗ в возрасте от 30 до 70 лет, женщины, %	–	14,34	0,0005	2 / 54	3,6 / 73	1 / 10	1,8 / 13,5
13	Смертность от ССЗ, рака, сахарного диабета, ХРЗ в возрасте от 30 до 70 лет, мужчины, %	100,1	18,05	0,0005	2 / 42	1,8 / 44,7	27 / 24	24,5 / 25,5
	Общее потребление алкоголя женщинами на душу населения (чистый алкоголь, л., прогнозные оценки, женщины старше 15 лет)	–	1,662	0,0005	75 / 7	68,2 / 7,4	6 / 21	5,5 / 22,3
14	Использование безопасной воды, %	88,32	66,81	0,0005	35 / 4	64,8 / 5	8 / 3	14,8 / 3,8
	Заболееваемость туберкулезом (на 100 000 чел.)	–	56,5	0,0005	3 / 70	5,6 / 87,5	8 / 3	14,8 / 3,8
15	Использование безопасной воды, %	84,93	77,49	0,0005	49 / 6	87,5 / 8,1	4 / 17	7,1 / 23
	Смертность от ССЗ, рака, сахарного диабета или ХРЗ в возрасте от 30 до 70 лет, мужчины, %	–	17,95	0,0005	0 / 41	0 / 55,4	3 / 10	5,4 / 13,5

Источник / Source: составлено авторами / Compiled by the authors.

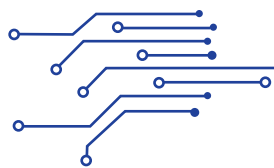


Таблица 5 / Table 5

Результаты классификации методами машинного обучения при сравнении двух групп исследования (относительные единицы) / The Results of Classification by Machine Learning Methods when Comparing two Study Groups (relative units)

Метод / Method	Правильность / Accuracy	Точность / Precision	Чувствительность / Sensitivity	Специфичность / Specificity	F-оценка / F-Score	ROC AUC
Логистическая регрессия (LR)	0,908	0,929	0,882	0,933	0,905	0,968
Статистически взвешенные синдромы (SWS)	0,908	0,937	0,974	0,941	0,904	0,968
Адаптивный бустинг (AB)	0,895	0,970	0,815	0,975	0,886	0,895
Деревья решений (DT)	0,857	0,857	0,857	0,857	0,857	0,857
Метод опорных векторов (SVM)	0,815	0,815	0,815	0,815	0,815	0,827
Градиентный бустинг (XGB)	0,874	0,862	0,891	0,857	0,876	0,943
Метод ближайших соседей (KNN)	0,786	0,793	0,773	0,798	0,783	0,786
Линейный дискриминантный анализ (LDA)	0,895	0,935	0,849	0,941	0,890	0,895
Нейронная сеть	0,559	1,0	0,118	1,0	0,211	0,559

Источник / Source: составлено авторами / Compiled by the authors.

СПИСОК ИСТОЧНИКОВ

1. Куккоев С.П. Моделирование ожидаемой продолжительности жизни в Российской Федерации. *Бюллетень Национального научно-исследовательского института общественного здоровья им. Н.А. Семашко*. 2018;1:29–37. URL: <https://nriph.ru/images/assets/files/archive/2018/2018-1.pdf#page=29>
2. Степанов В.С., Бобков В.Н., Шамаева Е.Ф., Одинцова Е.В. Построение модели, связывающей индикатор уровня жизни населения с комплексом показателей социально-экономической политики в регионах России. *Уровень жизни населения регионов России*. 2022;18(4):450–465. DOI: 10.19181/lsprr.2022.18.4.3
3. Куделина О.В., Канева М.А. Выбор индикатора капитала здоровья в российских регионах. *Экология человека*. 2020;9:18–27. DOI: 10.33396/1728-0869-2020-9-18-27
4. Степанов В.С. Прогноз распространенности онкозаболеваний среди жителей Московского региона на основе модели регрессии. *Моделирование, оптимизация и информационные технологии*. 2024;12(3). DOI: 10.26102/2310-6018/2024.46.3.023
5. SK Khader Basha S.K., et al. Machine learning method for life expectancy prediction. *Mukt Shabd Journal*. 2022;9(6):140–145. URL: https://www.researchgate.net/publication/378874579_MACHINE_LEARNING_METHOD_FOR_LIFE_EXPECTANCY_PREDICTION
6. Bergeron Boucher V.-P., Strozza C., Simonacci Y., Oeppen J. Modeling and Forecasting Healthy Life Expectancy with Compositional Data Analysis Demography. 2025;62(3):787–810. DOI: 10.1215/00703370-11974712
7. Aburto J.M., Schöley J., Kashnitsky I., Zhang L., et al. Quantifying impacts of the COVID-19 pandemic through life-expectancy losses: a population-level study of 29 countries. *Int. J. Epidemiol.* 2022;51(1):63–74. DOI: 10.1101/2021.03.02.21252772
8. Borisova L., Zhukova G., Kuznetsova A., Martin J. Features of machine learning in the study of the main factors of development of countries of the world. *SHS Web of Conferences*. 2021;110(5):02006. DOI: 10.1051/shsconf/202111002006

9. Кирилук И.Л., Волынский А.И., Круглова М.С., Кузнецова А.В., Рубинштейн А.А., Сенько О.В. Эмпирическая проверка теории институциональных матриц методами интеллектуального анализа данных. *Компьютерные исследования и моделирование*. 2015;7(4):923–939. DOI: 10.20537/2076-7633-2015-7-4-923-939
10. Борисова Л.Р., Кузнецова А.В., Алешченко А.В., Кирилук И.Л. Зависимость количественных индикаторов уровня жизни населения от экономического уклада страны. Новые технологии высшей школы. Наука. Техника. Педагогика. Сборник трудов конференции. М.: Московский Политех; 2021:127–132. URL: <https://www.elibrary.ru/ublkvx>
11. Кузнецова А.В., Борисова Л.Р., Демина И.А. Методы машинного обучения для прогнозирования течения заболевания на примере развития тяжелой степени пневмонии. *Цифровые решения и технологии искусственного интеллекта*. 2025;1(1):6–19.
12. Борисова Л.Р., Кузнецова А.В. Использование работающего компьютерного тренажера Data Master Azforus для обучения методам машинного обучения. Цифровая трансформация социальных и экономических систем. Материалы Международной научно-практической конференции; 28 января 2022 г. М.: Московский университет им. С.Ю. Витте; 2022:264–270. URL: <https://www.elibrary.ru/ohcndx>

REFERENCES

1. Kukkoiev S.P. Modeling of life expectancy in the Russian Federation. *Bulletin of the National Scientific Research Institute of Public Health named after N.A. Semashko*. 2018;1:29–37. URL: <https://nrph.ru/images/assets/files/archive/2018/2018-1.pdf#page=29> (In Russ.).
2. Stepanov V.S., Bobkov V.N., Shamaeva E.F., Odintsova E.V. Building a model linking the indicator of the standard of living of the population with a set of indicators of socio-economic policy in the regions of Russia. *The standard of living of the population of the Russian regions*. 2022;18(4):450–465. (In Russ.). DOI: 10.19181/lsprr.2022.18.4.3
3. Kudelina O.V., Kaneva M.A. The choice of health capital indicator in Russian regions. *Human ecology*. 2020;9:18–27. (In Russ.). DOI: 10.33396/1728-0869-2020-9-18-27
4. Stepanov V.S. Forecast of cancer prevalence among residents of the Moscow region based on a regression model. *Modeling, optimization, and information technology*. 2024;12(3). (In Russ.). DOI: 10.26102/2310-6018/2024.46.3.023
5. SK Khader Basha S.K., et al. Machine learning method for life expectancy prediction. *Mukt Shabd Journal*. 2022;9(6):140–145. URL: https://www.researchgate.net/publication/378874579_MACHINE_LEARNING_METHOD_FOR_LIFE_EXPECTANCY_PREDICTION
6. Bergeron Boucher V.-P., Strozza C., Simonacci Y., Oeppen J. Modeling and Forecasting Healthy Life Expectancy with Compositional Data Analysis Demography. 2025;62(3):787–810. DOI: 10.1215/00703370-11974712
7. Aburto J.M., Schöley J., Kashnitsky I., Zhang L., et al. Quantifying impacts of the COVID-19 pandemic through life-expectancy losses: a population-level study of 29 countries. *Int. J. Epidemiol.* 2022;51(1):63–74. DOI: 10.1101/2021.03.02.21252772
8. Borisova L., Zhukova G., Kuznetsova A., Martin J. Features of machine learning in the study of the main factors of development of countries of the world. *SHS Web of Conferences*. 2021;110(5):02006. DOI: 10.1051/shsconf/202111002006
9. Kirilyuk I.L., Volynsky A.I., Kруглова М.С., Kuznetsova A.V., Rubinstein A.A., Senko O.V. Empirical verification of the theory of institutional matrices by data mining methods. *Computer research and modeling*. 2015;7(4):923–939. (In Russ.). DOI: 10.20537/2076-7633-2015-7-4-923-939
10. Borisova L.R., Kuznetsova A.V., Aleshchenko A.V., Kirilyuk I.L. Dependence of quantitative indicators of the standard of living of the population on the economic structure of the country. New technologies of higher education. Science. Technic. Pedagogy. Proceedings of the conference. Moscow: Moskovsky Polytech; 2021:127–132. URL: <https://www.elibrary.ru/ublkvx> (In Russ.).
11. Kuznetsova A.V., Borisova L.R., Demina I.A. Machine learning methods for predicting the course of the disease using the example of the development of severe pneumonia. *Digital solutions and artificial intelligence technologies*. 2025;1(1):6–19. (In Russ.).
12. Borisova L.R., Kuznetsova A.V. Using a working computer simulator Data Master Azforus for teaching machine learning methods. Digital transformation of social and economic systems. Proceedings of the International Scientific and Practical Conference; January 28, 2022; Moscow: Witte Moscow University; 2022:264–270. URL: <https://www.elibrary.ru/ohcndx> (In Russ.).



ИНФОРМАЦИЯ ОБ АВТОРАХ / ABOUT THE AUTORS

Анна Викторовна Кузнецова — кандидат биологических наук, старший научный сотрудник Лаборатории математической биофизики Института биохимической физики им. Н.М. Эмануэля, Российская академия наук, Москва, Российская Федерация

Anna V. Kuznetsova — PhD (Bio.), Senior Researcher Laboratory of Mathematical Biophysics of the Institute of Biochemical Physics of the Russian Academy of Sciences, Moscow, Russian Federation

<http://orcid.org/0000-0003-2705-1935>

azforus@yandex.ru

Людмила Робертовна Борисова — кандидат физико-математических наук, доцент кафедры математики и анализа данных факультета информационных технологий и анализа больших данных, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

Lyudmila R. Borisova — Cand. Sci. (Phys.-Math.) Assoc. Prof., Department of Mathematics and Data Analysis, Faculty of Information Technology and Big Data Analysis, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

<http://orcid.org/0000-0002-5757-0341>

Автор для корреспонденции / Corresponding author:

lrborisova@fa.ru

Галина Александровна Постовалова — кандидат педагогических наук, доцент кафедры математики и анализа данных факультета информационных технологий и анализа больших данных, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

Galina A. Postovalova — Cand. Sci. (Ped.), Assoc. Prof., Department of Mathematics and Data Analysis, Faculty of Information Technology and Big Data Analysis, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

<http://orcid.org/0009-0005-4040-1831>

gapostovalova@fa.ru

Заявленный вклад авторов:

А.В. Кузнецова — расчеты в программном комплексе Data Master Azforus (таблицы 1–5).

Л.Р. Борисова — составление таблицы данных для проведения расчетов, написание текста статьи и введения.

Г.А. Постовалова — редактирование, рисунки 1–3; обсуждение.

Authors' declared contributions:

A. V. Kuznetsova — calculations in the Data Master Azforus software package (tables 1–5).

L. R. Borisova — compilation of a data table for calculations, writing the text of the article and introduction.

G. A. Postovalova — editing, figures 1–3; discussion.

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

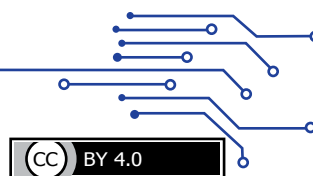
Conflicts of Interest Statement: The authors have no conflicts of interest to declare.

Статья поступила 03.07.2025; принята к публикации 17.07.2025.

Авторы прочитали и одобрили окончательный вариант рукописи.

The article was received 03.07.2025; accepted for publication 17.07.2025.

The authors read and approved the final version of the manuscript.



Современные методы токенизации текстов в финансовой сфере

Э.Ф. Болтачев^а, М.П. Фархадов^б, А.И. Тюляков^с

^{а,с}Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация;

^бИнститут проблем управления им. В.А. Трапезникова РАН, Москва, Российская Федерация

АННОТАЦИЯ

В работе рассматривается токенизация как ключевой этап обработки текстовых данных, особенно в финансовой сфере. Анализируются современные методы токенизации с примерами из последних исследований и их влияние на эффективность NLP-моделей. Исследование показывает, что алгоритмы токенизации по словам (BPE, WordPiece, Unigram) стали стандартом для языковых моделей благодаря гибкости и эффективности сжатия текста.

Обсуждаются ограничения длины входной последовательности в языковых моделях (BPE и WordPiece демонстрируют склонность к избыточному разбиению, Unigram требует сложного обучения, а символьная токенизация создает чрезмерно длинные последовательности) и методы преодоления этих ограничений, включая разбиение текста на части, иерархическую обработку и экстраполяцию предобученных моделей с архитектурой трансформеров для работы с длинными входными данными. Для финансовых данных рекомендуется использование доменно-ориентированных токенизаторов или дообучение на специализированных системах, что подтверждается успешным опытом BloombergGPT. Особое внимание уделяется проблеме обработки длинных текстов. Предложены три подхода к решению: разделение текста на части; иерархическая обработка; экстраполяция моделей-трансформеров. В заключение подчеркивается значимость токенизации для финансовой аналитики, где качество обработки текста напрямую влияет на принятие решений. Развитие методов токенизации продолжается параллельно с совершенствованием NLP-моделей, что делает этот этап обработки текста критически важным компонентом современных аналитических систем.

Ключевые слова: токенизация; длинные последовательности; дообучение; модели-трансформеры; массив информации; финансовая аналитика; минимизация потерь контекста

Для цитирования: Болтачев Э.Ф., Фархадов М.П., Тюляков А.И. Современные методы токенизации текстов в финансовой сфере. *Цифровые решения и технологии искусственного интеллекта*. 2025;1(3):19-29. DOI: 10.26794/3033-7097-2025-1-3-19-29

Modern Tokenization Methods for Text Processing in the Financial Domain

E.F. Boltachev^а, M.P. Farhadov^б, A.I. Tyulyakov^с

^{а,с}Financial University under the Government of the Russian Federation, Moscow, Russian Federation;

^бTrapeznikov Institute of Control Sciences of the Russian Academy of Sciences, Moscow, Russian Federation

ABSTRACT

The paper discusses tokenization as a key step in textual data processing, especially in the financial domain. Current tokenization techniques are analyzed with examples from recent research and their impact on the performance of NLP models. The study shows that word-based tokenization algorithms (BPE, WordPiece, Unigram) have become the standard for language models due to their flexibility and text compression efficiency. We discuss the limitations of input sequence length in language models (BPE and WordPiece show a tendency to over-partition, Unigram requires complex training, and symbolic tokenisation creates excessively long sequences) and methods to overcome these limitations, including text partitioning, hierarchical processing and extrapolation of pre-trained models with Transformer architecture to handle long input data. For financial data, it is recommended to use domain-specific tokenizers or additional training on specialized systems, which is confirmed by the successful experience of BloombergGPT. Special attention is paid to the problem of processing long texts. Three solution approaches are proposed: text partitioning; hierarchical processing; extrapolation of transformer models. In conclusion, the importance of tokenization for financial analytics is emphasized, where the quality of text processing directly affects decision-making. The development of tokenization

methods continues in parallel with the improvement of NLP models, which makes this stage of text processing a critical component of modern analytical systems.

Keywords: tokenization; long sequences; fine-tuning; transformer models; data array; financial analytics; context loss minimization

For citation: Boltachev E.F., Farhadov M.P., Tyulyakov A.I. Modern tokenization methods for text processing in the financial domain. *Digital solutions and artificial intelligence technologies*. 2025;1(3):19-29. DOI: 10.26794/3033-7097-2025-1-3-19-29

ВВЕДЕНИЕ

В последние годы методы обработки естественного языка (**Natural Language Processing, NLP**) стали неотъемлемой частью анализа данных в реальном времени [1] во многих отраслях, в том числе и в финансовой сфере. Современные большие языковые модели, такие как **BERT**, демонстрируют высокую эффективность в разнообразных задачах, включая анализ тональности новостных сообщений и классификацию текстов отчетов.

Эти технологии позволяют финансовым организациям автоматизировать обработку огромных массивов текстовой информации: от новостных сводок до юридических документов. Особое значение имеет способность NLP-систем выявлять рыночные тенденции и прогнозировать движения цен на основе анализа текстовых данных.

В финансовой индустрии NLP применяется для [2]:

- анализа рыночных настроений через обработку новостных потоков;
- выявления подозрительных паттернов в финансовых транзакциях;
- автоматизации комплаенс-процедур при работе с нормативными документами;
- создания структурированных инвестиционных отчетов на основе разнородных источников информации.

Благодаря развитию технологий машинного обучения и появлению мощных языковых моделей, финансовые организации получают инструменты для более глубокого понимания текстовых данных и принятия обоснованных решений в условиях быстро меняющегося рынка.

В настоящее время разработаны специализированные модели для финансовых текстов, например, FinBERT для анализа тональности финансовых новостей, показавшая улучшение качества по сравнению с общими моделями [3]. Однако успешность применения NLP-моделей во многом зависит от корректной токенизации — процесса разбиения текстовых данных на токены (элементарные единицы текста) перед подачей их в модель. Этот этап является фундаментальным для последующего этапа векторизации (пре-

образования токенов в числовые представления) и существенно влияет на качество и эффективность обучения модели.

В финансовых текстах, насыщенных специальной лексикой (например, тикеры акций, финансовые сокращения, числовые значения), выбор метода токенизации приобретает особую важность. Универсального метода, одинаково хорошо подходящего для всех случаев, — не существует: оптимальный способ разбиения текста часто определяется эмпирически и зависит от конкретной задачи и данных. Кроме того, финансовые документы (например, отчеты компаний или новости) могут быть очень объемными, превышающими ограничение на длину последовательности, допустимое в языковых моделях (для BERT — 512 токенов [4]). При превышении предельной длины текста модели вынуждены усекавать или разбивать его, что приводит к потере смыслового контекста. Ограничение длины обусловлено квадратичной сложностью механизма самовнимания трансформеров [5], из-за чего обработка многостраничных документов требует либо несоразмерных вычислительных ресурсов, либо специальных подходов к сегментации входных данных.

Решением вышеописанной проблемы является разработка эффективных методов токенизации и представления длинных текстовых последовательностей, обеспечивающих баланс между полнотой сохраняемой семантической информации и возможностями современных моделей.

В настоящей статье рассмотрены основные методологии токенизации и типы токенизаторов, проблемы токенизации длинных последовательностей и подходы к преодолению этих проблем, включая разделение на фрагменты (чанкинг), иерархическую обработку и экстраполяцию моделей на длинные тексты.

МЕТОДОЛОГИЯ ТОКЕНИЗАЦИИ

Понятие токена и токенизации. Токенизация представляет собой процесс разбиения исходного текста на последовательность токенов — элементарных фрагментов, которые могут быть словами, частями слов (подсловами),



морфемами, отдельными символами или даже объединениями слов в некоторый блок. Однако в большинстве приложений машинного обучения принято дробить слова далее на подсловные единицы. Это позволяет, с одной стороны, ограничить размер словаря модели (многие слова могут иметь общий корень или основу, например, слова **investment** и **investor**), а с другой — дает модели возможность обрабатывать незнакомые слова, раскладывая их на понятные составляющие.

Процесс токенизации обычно включает этапы нормализации текста (приведение регистра, удаление или замена специфических символов) и претокенизации — предварительное разделение простейшим образом (например, по пробелам и пунктуации) на черновые токены. Далее применяется основной алгоритм разбиения на фрагменты согласно выбранной модели. Токены затем преобразуются в числовые векторы (эмбеддинги) и подаются на вход нейросетевой модели.

Выбор методики токенизации критически влияет на эффективность модели как по точности, так и по требуемым ресурсам. Например, слишком грубое фрагментирование (на уровне целых слов) ведет к большому словарю и проблемам с редкими/неизвестными словами, а слишком мелкое (на уровне символов) приводит к чрезмерно длинным последовательностям токенов. В практических задачах обычно стремятся найти компромисс, часто используя подсловную токенизацию — разбиение слов на статистически обоснованные элементы.

Основные типы современных токенизаторов. Токенизация на уровне слов и на уровне символов является простейшим подходом. Токенизаторы на уровне слов разделяют текст по пробелам и знакам пунктуации, формируя токен из каждого слова. Такой подход интуитивно понятен и часто достаточен для простых задач, однако он не справляется с обработкой новых или редко встречающихся слов — любое слово, отсутствующее в словаре, считается неизвестным (out-of-vocabulary, OOV).

Символьная токенизация, напротив, разрезает текст на отдельные символы, полностью устраняя проблему неизвестных слов, — алфавит (набор символов) заведомо известен. Тем не менее последовательность символов для каждого слова получается намного длиннее, чем последовательность слов, что повышает требования к памяти и усложняет обучение модели. Символьные токены плохо отражают семантическое единство

многобуквенных морфем. Поэтому на практике чаще используют методы, находящиеся между этими крайностями, — подсловные токенизаторы.

Подсловные токенизаторы автоматически разрезают слова на статистически значимые фрагменты (субтокены) на основе частотных характеристик корпуса текстов. Наиболее распространенными на сегодня алгоритмами являются Byte Pair Encoding (BPE), WordPiece и Unigram. Эти методы позволяют эффективнее кодировать слова, обеспечивая баланс между размером словаря и длиной токенизированной последовательности.

Алгоритм BPE изначально был разработан для задач сжатия данных, а позже адаптирован для токенизации текста [6]. В NLP он получил популярность после применения в модели GPT от OpenAI и ныне используется в таких моделях, как GPT-2, RoBERTa, BART и многих других. Идея BPE заключается в итеративном слиянии наиболее часто встречающихся пар символов (или более общих токенов) в новом составном токене. На старте каждого символа уникально соответствуют токены, после чего алгоритм подсчитывает частоты всех пар токенов в системе и заменяет самую частотную пару на новый объединенный токен. Этот процесс повторяется, пока размер словаря не достигнет заданного. В итоге часто встречающиеся сочетания букв сливаются в единый токен (например, *in + vest* → *invest*), что позволяет сократить длину последовательности токенов.

Главное преимущество BPE — способность обрабатывать любые новые слова за счет поэтапного разбиения их на известные подчасти: если слово не присутствует целиком в словаре, токенизатор BPE разобьет его на более мелкие субтокены (в пределе — на отдельные символы). Таким образом достигается своего рода адаптивность к лексике, отсутствующей в обучающей выборке. В финансовых текстах, содержащих множество редких названий компаний, аббревиатур и т.д., это свойство крайне важно. BPE прост в реализации и эффективно сокращает длину представления текста, однако его решения основаны исключительно на частоте биграмм и фиксированном словаре — после обучения токенизатора разбиение новых текстов производится жадно по заранее выученному правилу слияний, без учета контекста или вероятностей.

Метод WordPiece во многом похож на BPE. Он также использует идею поэтапного слияния символов и субтокенов, однако обучение словаря происходит несколько иначе. Изначально WordPiece разрабатывался для обработки японского и корейского

Главное преимущество BPE — способность обрабатывать любые новые слова за счет поэтапного разбиения их на известные подчасти: если слово не присутствует целиком в словаре, токенизатор BPE разобьет его на более мелкие субтокены (в пределе — на отдельные символы). Таким образом достигается своего рода адаптивность к лексике, отсутствующей в обучающей выборке. В финансовых текстах, содержащих множество редких названий компаний, аббревиатур и т.д., это свойство крайне важно.

языков без разметки слов и впервые представлен в работах Google по распознаванию речи [4]. Позднее WordPiece был применен в модели BERT, благодаря чему получил широкое распространение. Алгоритм обучения WordPiece также начинается с единичных символов, но при выборе объединяемых пар учитывает не только их частоту, но и максимизацию вероятности данных при данной модели (т.е. основывается на языковой модели, оценивающей вероятность строк). Фактически WordPiece, как и BPE, строит словарь субтокенов, сливая часто встречающиеся последовательности символов, но сохраняет некие оценки вероятностей для токенов.

В реализации результат разбиения часто похож на BPE. Например, токенизатор BERT WordPiece разбивает слово *investment* на токены *invest* и *##ment* (символы *##* обозначают продолжение слова). Существенным отличием WordPiece является сохранение информации о вероятностном языке токенов: в процессе разбиения новых слов алгоритм выбирает сегментацию, максимизирующую вероятность токенов (а не просто жадно берет самый длинный совпадающий токен из словаря). На практике WordPiece также работает достаточно детерминированно и близок к BPE по результатам. Можно считать WordPiece эволюцией BPE с небольшими отличиями в критериях слияния.

Алгоритм Unigram (униграммная токенизация) принципиально отличается от BPE и WordPiece.

Вместо итеративного объединения он, наоборот, начинает обработку с большого списка возможных токенов и последовательно удаляет наименее вероятные, пока размер словаря не достигнет заданного размера. В основе заложена языковая модель «униграмм», предполагающая, что токены независимы и имеют определенные вероятности появления. Обучение токенизатора Unigram похоже на кластеризацию: модель пытается найти набор токенов, оптимально объясняющий (в вероятностном смысле) обучающая система [6].

При разбиении нового слова Unigram может рассматривать несколько сегментаций и выбирать лишь имеющую максимальную вероятность согласно выученным вероятностям токенов. Такой подход позволяет учесть неопределенность: если слово встречается в тексте в новом контексте, то токенизатор может иначе его сегментировать, исходя из вероятностей. Преимущество Unigram — более гибкое, контекстно-зависимое разбиение, которое теоретически может лучше сохранять смысл.

Недавние исследования показали, что алгоритм Unigram позволяет лучше сохранять морфологическую целостность слов по сравнению с BPE [7]. Например, выяснилось, что Unigram чаще выделяет осмысленные корни и аффиксы, ближе соответствуя морфемам языка. В модели BloombergGPT был сделан выбор в пользу Unigram вместо BPE именно на основе результатов, показывающих преимущества Unigram для смешанных систем [8]. В частности, токенизатор BloombergGPT (на базе SentencePiece) обучен по униграммной модели с большим словарем (~ 131k токенов), что позволило лучше охватить финансовую терминологию и снизить общее количество токенов для текстов этого домена.

Однако у Unigram есть и недостатки. Он более сложен в реализации и требует больше данных и вычислений для обучения вероятностной модели токенов. На практике разбиение Unigram может быть менее интерпретируемым, поскольку основывается на глобальной вероятностной модели. В некоторых случаях алгоритм Unigram показывает незначительные преимущества или даже уступает простым методам в метриках качества, особенно если текстовый корпус небольшой. Тем не менее при больших объемах данных и наличии доменно-специфической лексики (как в финансах) преимущества Unigram проявляются в более компактном представлении частых многословных терминов и более осмысленном обращении с редкими словами.

Помимо перечисленных алгоритмов, существуют и другие подходы к токенизации, нацеленные на



Преимущество Unigram – более гибкое, контекстно-зависимое разбиение, которое теоретически может лучше сохранять смысл. На практике разбиение Unigram может быть менее интерпретируемым, поскольку основывается на глобальной вероятностной модели.

специфические языки или задачи. Например, для языков без явных разделителей слов (китайский, японский) используются токенизаторы, основанные на словарях и правилах или на моделях сегментации, которые разбивают текст на слова и фразы осмысленно. В агглютинативных языках (например, тюркские и финно-угорские) эффективна морфологическая токенизация с выделением корней и аффиксов с помощью правил или обучаемых морфанализаторов.

В финансовых текстах может потребоваться отдельная обработка специальных символов (знаков валют, процентов, пунктуации) и сочетаний типа S&P500 или EUR/USD. Например, токенизатор BloombergGPT на этапе претокенизации специально оставляет пробелы в алфавитных последовательностях, позволяя включать многословные финансовые термины как единый токен, а цифры разделяет так, что каждая цифра становится отдельным токеном [8]. Это повышает информационную плотность представления (меньше токенов на многословные названия) и улучшает обработку числовых данных. Такие эвристики и настройки токенизации под конкретный домен нередко позволяют существенно повысить качество моделей на профильных задачах.

Выбор токенизатора определяется балансом между размером словаря, длиной получаемых последовательностей, характером языка и текстов. При дообучении предобученных моделей выбор токенизатора обусловлен требованиями конкретной модели (некоторые модели обучены строго под определенный токенизатор).

КЛАССИФИКАЦИЯ ТОКЕНИЗАТОРОВ

Методы токенизации можно классифицировать по следующим признакам.

По единице разбиения. По словам (whitespace tokenization), по подсловам (BPE, WordPiece, Unigram и др.), по символам. Сюда же можно

отнести N-граммные токенизаторы, разбивающие текст на последовательности из N символов или слов (например, биграммы, триграммы) — иногда используются для сохранения локальных сочетаний, хотя и увеличивают избыточность представления текста.

По способу построения словаря. Правила и словари (статистические правила разбиения, используемые для отдельных языков), обучаемые алгоритмы (статистические — BPE, WordPiece, Unigram; нейросетевые — реже применяются напрямую для токенизации). Обучаемые токенизаторы строят словарь и правила сегментации на основе системных данных. Например, BPE и WordPiece — жадные алгоритмы слияния, Unigram — стохастический алгоритм фильтрации токенов.

По учету контекста при сегментации. Детерминированные (BPE, WordPiece — после обучения разбивают новые тексты предсказуемо, не учитывая семантический контекст предложения), стохастические/контекстуальные (Unigram может выбирать разбиение с учетом вероятностей; также существуют подходы динамической токенизации, где разбиение текста адаптируется под конкретный вход). Детерминированные токенизаторы быстрее и проще, а динамические могут достраивать токены под контекст, увеличивая эффективность использования токенов. Однако полностью контекстно-адаптивные токенизаторы встречаются редко, чаще контекст влияет косвенно через вероятностную модель, как в Unigram.

По размеру словаря. С малым фиксированным словарем (по символам), средним словарем (по подсловам от 30 до 50 тыс. токенов, как в BERT или GPT-2), с большим словарем (специализированные токенизаторы могут иметь более 100 тыс. токенов). Размер словаря влияет на качество: больший словарь уменьшает длину последовательности (так как объединяет часто встречающиеся последовательности в один токен), но увеличивает число параметров модели [8]. Например, в модели BloombergGPT выбрали относительно большой словарь около 131 тыс. токенов, чтобы уменьшить среднюю длину представления документов финансового корпуса, несмотря на рост параметров модели.

По языковой и доменной специфике. Универсальные токенизаторы (обученные на смешанном корпусе, покрывающем множество доменов), специализированные (обученные на узкой предметной области или языке). Пример первого — стандартный токенизатор BERT на основе

WordPiece, обученный на новостных и корпусах книг, пример второго — токенизатор FinBERT, дообученный на финансовых текстах. Специализированные токенизаторы включают больше доменных терминов в качестве цельных токенов, что улучшает понимание модели в узкой области ценой ухудшения на общих текстах.

В финансовой сфере часто целесообразно адаптировать токенизацию: включать сокращения (IPO, EBITDA) и метки как отдельные токены, предотвращая их разбиение на буквы, и обрабатывать числа особым образом (например, разделять число и знак валюты на отдельные токены). Такие настройки можно задать как через правила предварительной токенизации, так и обучив модель токенизации на специализированном корпусе.

Для языков без явных разделителей слов (китайский, японский) используются токенизаторы, основанные на словарях и правилах или на моделях сегментации, которые разбивают текст на слова и фразы осмысленно. В агглютинативных языках (тюркские и финно-угорские) эффективна морфологическая токенизация с выделением корней и аффиксов с помощью правил или обучаемых морфанализаторов.

ПРОБЛЕМЫ ТОКЕНИЗАЦИИ ДЛИННЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ

Финансовые документы зачастую представлены большим набором текстовых данных. С точки зрения токенизации их можно рассматривать как длинные последовательности, поскольку обработка таких документов требует соблюдения порядка следования. Среди основных проблем токенизации длинных последовательностей можно выделить следующие.

Ограничения длины в трансформерах. Большинство современных языковых моделей на основе архитектуры трансформеров имеют фиксированный максимальный размер входной последовательности токенов (контекстное окно). Например, BERT изначально обучен на последовательностях

длиной не более 512 токенов [4], многие модели GPT-2/3 — на 1024 или 2048 токенах. Это значит, что при предсказании или обучении модель просто не сможет учесть токены, выходящие за предел этого окна, — лишние слова отсекаются или разбиваются на отдельные части. Если финансовый документ содержит, скажем, 2000 слов, то напрямую подать его в BERT не удастся. Такая архитектурная особенность обусловлена тем, что вычислительная сложность самовнимания (self-attention) в трансформере растет квадратично с длиной последовательности $O(n^2)$ [5]. Разумным компромиссом между возможностью учесть долгосрочный контекст и приемлемым временем/памятью работы модели является ограничение длины последовательности.

Потенциальные потери при усечении текста.

Если текст превышает максимально допустимую длину, на практике используют усечение (тривиально отбрасывают «хвост» текста) либо разбиение текста на части (чанкинг) для поэтапной обработки. Оба подхода приводят к потере целостного контекста документа. Усечение опасно тем, что важная информация в конце текста будет проигнорирована моделью. Разбиение на части и отдельная обработка фрагментов лишает модель возможности «видеть» взаимосвязи между отдаленными частями текста (например, между началом и концовкой отчета аналитика). В результате качество анализа может снижаться, особенно в задачах, требующих глобального понимания документа (суммаризация, классификация по тональности всего текста и т.п.). Для финансовых документов, содержащих длительные рассуждения или каскадно поданные факты, потеря связности — серьезная проблема.

Избыточное представление коротких текстов. Противоположная ситуация — когда модель рассчитана на большой контекст (например, 2048 токенов), а тексты короткие (100–200 токенов). В этом случае ограничение не мешает обработке, но встает вопрос об эффективности использования длинных токенов. Если токенизация чересчур мелкая, то даже короткий текст может раздуться до сотен токенов, занимая значительную долю окна модели. Например, символьный токенизатор для слова Bloomberg произведет 9 токенов (каждая буква), тогда как подсловный BPE может дать 1 или 2 токена.

Чем меньше токенов, тем эффективнее используется контекстное окно и тем больше информации может модель обработать за раз. Таким образом, токенизация должна стремиться к компактному представлению текста без потери значимой информации. Этого достигают за счет продуманных ал-



горитмов (например, объединяя частые сочетания символов в один токен, как делают BPE/WordPiece). Если же текст крайне короткий, то ограничение на 512 или 1024 токена вообще не играет роли, но стоит учитывать, что модель все равно будет работать с фиксированным входным вектором той же размерности, что и для длинного текста (в BERT это 512).

Доменные особенности длинных финансовых текстов. Финансовые документы — отчеты, прогнозы, новости, аналитические статьи — часто значительно длиннее среднего новостного абзаца. Например, годовой отчет компании может содержать десятки страниц, а лента биржевых новостей — множество обновлений по одной теме. Токенизация таких текстов способна порождать последовательности длиной в тысячи токенов даже при использовании методов токенизации по словам. Кроме того, в финансовых текстах часто встречаются таблицы, списки, структурированные разделы — неочевидно, как их лучше токенизировать (сохранить ли разрывы строк, какие символы считать разделителями и т.п.). Неправильная токенизация может раздуть последовательность или, наоборот, склеить разрозненные части текста, внеся шум. Таким образом, при работе с длинными последовательностями необходимо учитывать не только ограничение модели, но и содержательную структуру документа, чтобы токенизация не нарушила ее семантику.

ПОДХОДЫ К ПРЕОДОЛЕНИЮ ОГРАНИЧЕНИЙ

Для эффективной работы с длинными последовательностями токенов разработано несколько групп методов. Рассмотрим основные из них:

- разделение на части (чанкинг);
- иерархическая обработка;
- экстраполяция трансформеров.

Эти подходы не взаимоисключают, а могут сочетаться друг с другом, обеспечивая более полноценную обработку длинных текстов.

Разделение на части. В контексте обработки текстов под чанкингом понимают разбиение длинного текста на более мелкие фрагменты (чанки), которые затем обрабатываются отдельно или последовательно. В традиционном NLP термин «chunking» чаще относится к разбиению предложения на синтаксические фразы — группы слов, например именные или глагольные группы. Такой фразовый чанкинг является промежуточным этапом между токенизацией отдельных слов и полным синтаксическим разбором предложения. Например,

по заданным грамматическим правилам можно выделять в предложении именные группы (NP), глагольные группы (VP) и т.д. Однако применительно к проблеме длинных последовательностей понятием «чанк» часто обозначают просто фрагмент текста (например, абзац, несколько предложений или фиксированное число токенов), который модель способна обработать целиком. Проще говоря, длинный документ разбивают на части, каждая из которых укладывается в 512 (или иной лимит) токенов, и подают в модель по отдельности. Чанкинг текста — самый прямолинейный способ обойти ограничение длины без модификации модели. Пример — для классификации длинного финансового отчета можно разбить его по абзацам и классифицировать каждый абзац BERT-моделью, а затем как-то агрегировать результаты.

Недавние исследования предлагают улучшенные варианты этого подхода. Так, в 2023 г. был представлен метод ChunkBERT, в котором выходные представления токенов длинного текста делятся на чанки фиксированного размера, после чего поверх каждого чанка применяется сверточный нейронный слой для агрегации информации [4]. Такой подход позволил дообучить BERT для работы с произвольно длинными текстами без сильного роста памяти: модель ChunkBERT сохранила 93,75% памяти по сравнению с обычным увеличением контекста. В целом, чанкинг представляет собой простой и эффективный метод, однако имеющий очевидный недостаток: модель не учитывает взаимосвязей между чанками (если специально не предусмотреть механизм их агрегирования).

В контексте финансов, где вывод в конце документа может зависеть от информации в его начале, простое независимое рассмотрение частей может упустить важные зависимости. Поэтому чанкинг часто дополняется или усложняется другими методами. Тем не менее многие практические системы начинают с чанкинга: например, для резюмирования очень длинных отчетов их сначала делят на разделы, резюмируют каждый раздел отдельно, а затем их агрегируют для подготовки итогового реферата.

Иерархическая обработка. Иерархические методы предполагают построение многоуровневого представления текста: на нижнем уровне модель обрабатывает относительно короткие фрагменты (например, предложения) и получает их эмбединги, на следующем уровне эти эмбединги используются для понимания более крупных единиц (абзац, раздел) и т.д. вплоть до уровня всего документа. Под-

ход был формализован в виде иерархических нейронных сетей, в частности, архитектуры с иерархическим вниманием (Hierarchical Attention Networks, HAN) для классификации документов [9].

В HAN сначала применяется RNN или трансформер для получения векторного представления каждого предложения из слов (эмбеддингов), затем другой уровень сети собирает представления предложений в представление абзаца и глобальный уровень — представление документа из абзацев, при этом на каждом уровне есть механизм внимания, выделяющий важные элементы. Такой подход отражает естественную структуру текста: слова объединяются в осмысленные предложения, предложения — в абзацы, которые формируют документ. Преимущество иерархической схемы — значительное сокращение вычислительной нагрузки, так как каждый уровень оперирует ограниченным числом элементов. Модель фокусируется на наиболее информативных частях на каждом уровне. Важный аспект — возможность использовать разные методы на разных уровнях. В финансовых обзорах, например, можно сначала получить «векторные представления» каждого раздела (аналитика, прогноз, выводы), а затем использовать их для финального вывода модели о тональности или рекомендациях. В математическом плане иерархическая модель может рассматриваться как сжатие последовательности: вместо передачи тысяч токенов напрямую в одну модель, мы передаем, условно, десятки эмбеддингов более высокого уровня. Это своего рода семантическая компрессия текста. Среди недостатков подхода следует выделить возможную потерю детализации: если на нижних уровнях произошло недостаточное извлечение признаков, верхние уровни уже не восстановят утерянную информацию.

Экстраполяция моделей-трансформеров на длинные последовательности. Это подход смещен в сторону улучшения самих моделей для работы с большим числом токенов. Здесь выделяются два направления:

1) разработка новых архитектур трансформеров с более эффективным механизмом внимания (субквадратичной сложностью);

2) методы адаптации уже предобученных моделей к большим контекстам без полного переобучения.

Примеры первого направления — Longformer, BigBird, Reformer, Transformer-XL и множество других вариантов. Longformer вводит разреженное внимание «скользящим окном» и глобальные

токены, достигая линейной сложности по длине [10]. BigBird комбинирует шаблоны разреженного, локального и случайного внимания и демонстрирует, что такая модель теоретически не уступает полной (плотной) по выразительности [11]. Reformer заменяет привычное внимание на локально-чувствительное хеширование, также снижая сложность до $O(n \log(n))$. Transformer-XL использует механизм сегментного рекуррентного внимания: модель запоминает состояния на предыдущих отрезках последовательности и таким образом умеет учитывать контекст длиной более окна, перенося информацию от одного сегмента к другому [12]. Все эти архитектуры требуют полного цикла обучения модели и имеют свои ограничения (например, Longformer и BigBird все еще нуждаются в заданном максимальном окне).

Второе направление — экстраполяция предобученных моделей. Этот подход очень ценен, когда имеется готовая модель и нужно применить ее к более длинному тексту, чем она видела при обучении. Прямое увеличение позиций требует либо добавления обученных позиционных эмбеддингов, либо модификации механизма внимания. В 2023 г. был предложен подход LSG Attention (Local-Sparse-Global), позволяющий преобразовать веса существующего трансформера таким образом, чтобы он мог обрабатывать гораздо более длинные последовательности без дополнительного обучения [5].

LSG-внимание сочетает локальное внимание (каждый токен внимателен только к ближайшим соседям), разреженное и несколько глобальных токенов, которые собирают информацию по всему документу. Благодаря этому комбинированному механизму удастся значительно ускорить обработку длинных последовательностей и при этом адаптировать уже обученные модели к новым длинам контекста. LSG-модификация минимально сказывается на качестве на коротких текстах, но эффективно масштабирует модель до 4096 и более токенов. Подход LSG был продемонстрирован на задачах классификации и резюмирования длинных документов: производительность сопоставима с моделью Longformer, при том, что исходная модель — стандартная модель BERT.

Другие методы экстраполяции включают добавление альтернативных позиционных кодировок. Например, абсолютные позиционные эмбеддинги заменяются на относительные (как в Transformer-XL) или на алгоритмически вычисляемые (как ALiBi), которые позволяют моделям обобщать на длины больше обученных. Эти под-



ходы менее изучены, но заслуживают внимания, поскольку полный цикл обучения модели малоэффективен, а готовые модели ограничены длиной входной последовательности.

ВЫВОДЫ

Токенизация является критически важным этапом при обработке текстов, особенно в специализированных областях, включая финансовую сферу. Правильный выбор и настройка токенизатора позволяют существенно улучшить результаты моделей NLP: повысить точность анализа, обеспечить обобщаемость на новые данные и оптимизировать использование ресурсов. В данной работе проведен обзор современных методов токенизации и показано, что алгоритмы токенизации по словам (BPE, WordPiece, Unigram) на сегодняшний день стали стандартом де-факто для большинства языковых моделей благодаря своей гибкости и эффективному сжатию текста. В то же время подчеркнуто, что ни один метод не лишен недостатков: BPE и WordPiece страдают от жадности разбиения и могут игнорировать морфологию, Unigram сложнее обучать и реже используется, токенизация по словам не справляется с неизвестными словами, а символьная — генерирует слишком длинные входы. Для финансовых данных, содержащих множество специальных терминов, имеет смысл использовать доменно-ориентированные токенизаторы или дообучать словарь на профильном корпусе — это подтверждается опытом (например, модель BloombergGPT достигла лучших результатов, применив

unigram-токенизацию с учетом финансовой терминологии).

Отдельно рассмотрены проблемы длинных текстов: стандартные модели не позволяют обработать целый документ аналитического отчета или новостную ленту целиком из-за ограничений на длину последовательности. Мы описали три ключевых подхода, позволяющих преодолеть это ограничение. Разделение текста по частям предоставляет простой способ разбиения текста на части и анализа их по отдельности, что подходит для многих практических задач, хотя и лишает модель глобального контекста. Иерархическая обработка восстанавливает упущенный контекст, строя многоуровневое представление текста — это более сложный, но и более точный подход, позволяющий моделям учитывать структуру документа и дальние зависимости. Экстраполяция моделей-трансформеров предлагает решение на уровне самой модели, уменьшая вычислительную сложность или вводя механизмы памяти, что делает возможным работу с очень длинными цепочками токенов. Все эти направления активно развиваются: появляются новые алгоритмы токенизации, улучшаются длинноконтекстные модели трансформеров, а также методы предварительного сжатия текста.

Для финансовой сферы, где полнота анализа текста может влиять на принятие важных решений, токенизация — это важный этап работы NLP-систем. Таким образом, токенизация эволюционируя вместе с моделями, остается базовым и крайне важным компонентом процесса обработки текстов в задачах финансовой аналитики.

СПИСОК ИСТОЧНИКОВ

1. Панкратова М.Д., Сковпень Т.Н. Модели NLP с использованием нейронных сетей в анализе тональности новостей. *Аналитические технологии в социальной сфере: теория и практика*. 2023;97–107. URL: <https://www.elibrary.ru/ctabku>
2. Марков А.К., Семеновкин Д.О., Кравец А.Г., Яновский Т.А. Сравнительный анализ применяемых технологий обработки естественного языка для улучшения качества классификации цифровых документов. *International Journal of Information Technologies*. 2024;12(3):66–77. URL: <https://www.elibrary.ru/tubosi>
3. Araci D. FinBERT: Financial Sentiment Analysis with Pre-Trained Language Models. *arXiv preprint arXiv:1908.10063*. 2019;7. DOI: 10.48550/arXiv.1908.10063
4. Jaiswal A., Milios E. Breaking the Token Barrier: Chunking and Convolution for Efficient Long Text Classification with BERT. *arXiv preprint arXiv:2310.20558*. 2023;13. DOI: 10.48550/arXiv.2310.20558
5. Condevaux C., Harispe S. LSG Attention: Extrapolation of Pretrained Transformers to Long Sequences. *Advances in Knowledge Discovery and Data Mining (PAKDD)*. 2023;13935 LNCS:443–454. DOI: 10.48550/arXiv.2210.15497
6. Sennrich R., Haddow B., Birch A. Neural Machine Translation of Rare Words with Subword Units. *Proceedings of ACL*. 2016;1715–1725. DOI: 10.48550/arXiv.1508.07909
7. Bostrom K., Durrett G. Byte Pair Encoding is Suboptimal for Language Model Pretraining. *Proceedings of EMNLP*. 2020;461–466. DOI: 10.18653/v1/2020.findings-emnlp.414



8. Wu S., et al. BloombergGPT: A Large Language Model for Finance. *arXiv preprint arXiv:2303.17564*. 2023;63. DOI: 10.48550/arXiv.2303.17564
9. Yang Z., et al. Hierarchical Attention Networks for Document Classification. *Proceedings of NAACL*. 2016;1480–1489. DOI: 10.18653/v1/N 16–1174
10. Beltagy I., Peters M. E., Cohan A. Longformer: The Long-Document Transformer. *arXiv preprint arXiv:2004.05150*. 2020;12. DOI: 10.48550/arXiv.2004.05150
11. Zaheer M., et al. Big Bird: Transformers for Longer Sequences. *Advances in Neural Information Processing Systems (NeurIPS)*. 2020;33:17283–17297. DOI: 10.48550/arXiv.2007.14062
12. Dai Z., et al. Transformer-XL: Attentive Language Models Beyond a Fixed-Length Context. *Proceedings of ACL*. 2019;2978–2988. DOI: 10.48550/arXiv.1901.02860

REFERENCES

1. Pankratova M. D., Skovpen T. N. NLP Models Using Neural Networks for Sentiment Analysis of News. *Analytical Technologies in the Social Sphere: Theory and Practice*. 2023;97–107. URL: <https://www.elibrary.ru/ctabku>
2. Markov A. K., Semenchkin D. O., Kravets A. G., Yanovsky T. A. Comparative Analysis of Applied Natural Language Processing Technologies for Improving the Quality of Digital Document Classification. *International Journal of Information Technologies*. 2024;12(3):66–77. URL: <https://www.elibrary.ru/tubosi>
3. Araci D. FinBERT: Financial Sentiment Analysis with Pre-Trained Language Models. *arXiv preprint arXiv:1908.10063*. 2019;7. DOI: 10.48550/arXiv.1908.10063
4. Jaiswal A., Milios E. Breaking the Token Barrier: Chunking and Convolution for Efficient Long Text Classification with BERT. *arXiv preprint arXiv:2310.20558*. 2023;13. DOI: 10.48550/arXiv.2310.20558
5. Condevaux C., Harispe S. LSG Attention: Extrapolation of Pretrained Transformers to Long Sequences. *Advances in Knowledge Discovery and Data Mining (PAKDD)*. 2023;13935 LNCS:443–454. DOI: 10.48550/arXiv.2210.15497
6. Sennrich R., Haddow B., Birch A. Neural Machine Translation of Rare Words with Subword Units. *Proceedings of ACL*. 2016;1715–1725. DOI: 10.48550/arXiv.1508.07909
7. Bostrom K., Durrett G. Byte Pair Encoding is Suboptimal for Language Model Pretraining. *Proceedings of EMNLP*. 2020;461–466. DOI: 10.18653/v1/2020.findings-emnlp.414
8. Wu S., et al. BloombergGPT: A Large Language Model for Finance. *arXiv preprint arXiv:2303.17564*. 2023;63. DOI: 10.48550/arXiv.2303.17564
9. Yang Z., et al. Hierarchical Attention Networks for Document Classification. *Proceedings of NAACL*. 2016;1480–1489. DOI: 10.18653/v1/N 16–1174
10. Beltagy I., Peters M. E., Cohan A. Longformer: The Long-Document Transformer. *arXiv preprint arXiv:2004.05150*. 2020;12. DOI: 10.48550/arXiv.2004.05150
11. Zaheer M., et al. Big Bird: Transformers for Longer Sequences. *Advances in Neural Information Processing Systems (NeurIPS)*. 2020;33:17283–17297. DOI: 10.48550/arXiv.2007.14062
12. Dai Z., et al. Transformer-XL: Attentive Language Models Beyond a Fixed-Length Context. *Proceedings of ACL*. 2019;2978–2988. DOI: 10.48550/arXiv.1901.02860

ИНФОРМАЦИЯ ОБ АВТОРАХ / ABOUT THE AUTHORS

Эльдар Филаридович Болтачев — кандидат технических наук, доцент кафедры искусственного интеллекта факультета информационных технологий и анализа больших данных; центр цифровой трансформации и искусственного интеллекта, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

Eldar F. Boltachev — PhD (Tech.), Assoc. Prof., Department of Artificial Intelligence, Faculty of Information Technology and Big Data Analysis; Center for Digital Transformation and Artificial Intelligence, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

<https://orcid.org/0000-0002-6375-0365>

Автор для корреспонденции / Corresponding author:

efboltachev@fa.ru



Маис Паша Оглы Фархадов — доктор технических наук, старший научный сотрудник, заведующий лабораторией «Эргатические системы» Института проблем управления им. В.А. Трапезникова Российской академии наук, Москва, Российская Федерация

Mais P. Farhadov — PhD (Tech.), Senior Researcher, Head of the “Ergatic Systems” Laboratory, Trapeznikov Institute of Control Sciences of the Russian Academy of Sciences, Moscow, Russian Federation

<https://orcid.org/0000-0002-7166-9567>

mais@ipu.ru

Александр Игоревич Тюляков — студент факультета информационных технологий и анализа больших данных, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

Alexander I. Tyulyakov — student of the Faculty of Information Technology and Big Data Analysis, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

<https://orcid.org/0009-0008-0534-0342>

244447@edu.fa.ru

Заявленный вклад авторов:

Э.Ф. Болтачев — введение, подходы к преодолению ограничений.

М.П. Фархадов — методология токенизации, выводы.

А.И. Тюляков — классификация токенизаторов и проблемы токенизации длинных последовательностей.

Authors' declared contributions:

E.F. Boltachev — introduction, approaches to overcoming limitations.

M.P. Farhadov — methodology of tokenization, conclusions.

A.I. Tyulyakov — classification of tokenizers and problems of tokenization of long sequences.

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Conflicts of Interest Statement: The authors have no conflicts of interest to declare.

Статья поступила 18.07.2025; принята к публикации 24.07.2025.

Авторы прочитали и одобрили окончательный вариант рукописи.

The article was received 18.07.2025; accepted for publication 24.07.2025.

The authors read and approved the final version of the manuscript.



Динамические графы и их некоторые приложения

Р.А. Кочкаров^а, А.А. Кочкаров^б^{а,б} Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация;^б ФИЦ Биотехнологии РАН, Москва, Российская Федерация

АННОТАЦИЯ

В статье рассматриваются современные подходы к моделированию сетевых систем и в целом сетей с динамической природой. В работе представлен современный класс динамических графов с описанием их практических применений. В качестве процедуры изменения динамического графа представлены базовые или простые операции, среди которых удаление или добавление вершин и ребер. Выделен специальный подкласс предфрактальных графов со свойствами самоподобия. Для класса динамических графов определено понятие траектории, представленное последовательностью классических графов, меняющихся от одного к другому во времени. Инструментарий динамических графов может стать основой для разработки алгоритмов командно-информационного взаимодействия подвижных абонентов в сетевых системах, в том числе сетевых системах непрерывного пространственного мониторинга. Для описания оптимизационных задач на многовзвешенных графах предложена формальная постановка многокритериальной задачи на предфрактальном графе. Описаны множества допустимых решений, парето-оптимальных и полных решений. Предложены некоторые леммы многокритериальной оптимизации для индивидуальных задач, которые обладают свойством полноты, а также ограничения линейной свертки критериев для поиска парето-оптимальных решений. Исследованы наследственные свойства, проявляющиеся в траекториях динамического графа, — а именно наследственность структурно-функциональных характеристик и, как следствие, наследственность решений при переходе от одного графа к другому в траектории динамического графа. Настоящая работа вносит свой вклад в развитие сетевой науки и теорию динамических сетей, предлагая как подходы, так и частные решения на общих и специальных классах графов.

Ключевые слова: динамический граф; многокритериальная оптимизация; предфрактальный граф; наследственные свойства; многокритериальные задачи; сетевая наука; динамические сети

Для цитирования: Кочкаров Р.А., Кочкаров А.А. Динамические графы и их некоторые приложения. *Цифровые решения и технологии искусственного интеллекта*. 2025;1(3):30-36. DOI: 10.26794/3033-7097-2025-1-3-30-36

ORIGINAL PAPER

Dynamic Graphs and Some of Their Applications

R.A. Kochkarov^а, A.A. Kochkarov^б^{а,б} Financial University under the Government of the Russian Federation, Moscow, Russian Federation;^б Federal Research Center of Biotechnology of the Russian Academy of Sciences, Moscow, Russian Federation

ABSTRACT

The article considers modern approaches to modeling network systems and networks with a dynamic nature in general. The paper presents a modern class of dynamic graphs with a description of their practical implementation. Basic or simple operations, including deletion or addition of vertices and edges, are presented as a procedure for changing a dynamic graph. A special subclass of prefractal graphs with self-similarity properties is identified. For the class of dynamic graphs, the concept of a trajectory is defined, represented by a sequence of classical graphs changing from one to another in timeline. The toolkit of dynamic graphs can become the base for developing algorithms for command-information interaction of mobile subscribers in network systems, including network systems of continuous spatial monitoring. To describe optimization problems on multi-weighted graphs, a formal statement of a multi-criteria problem on a prefractal graph is proposed. Sets of feasible solutions, Pareto-optimal and complete solutions are described. Some lemmas of multicriterial optimization for individual problems that have the property of completeness are proposed, as well as restrictions on the linear convolution of criteria for finding Pareto-optimal solutions. The hereditary properties that manifest themselves in the trajectories of a dynamic graph are investigated, namely, the heredity of structural and functional characteristics and, as a result, the heredity of decisions during the transition from one graph to another in the trajectory of a dynamic graph. This work contributes to the development of network science and the theory of dynamic networks, offering both approaches and particular solutions on general and special classes of graphs.



Keywords: dynamic graph; multicriteria optimization; prefractal graph; hereditary properties; multicriteria problems; network science; dynamic networks

For citation: Kochkarov R.A., Kochkarov A.A. Dynamic graphs and some of their applications. *Digital Solutions and Artificial Intelligence Technologies*. 2025;1(3):30-36. DOI: 10.26794/3033-7097-2025-1-3-30-36

ВВЕДЕНИЕ

При численном моделировании нестационарных течений на адаптивных сетках в лагранжевых координатах возникает необходимость периодически удалять ячейки (если их объем становится слишком мал в ходе расчетов) или добавлять новые. В результате граф функциональной зависимости становится изменяемым во времени $G(t)$. Аналогично нейронным сетям, одним из критериев выбора решения является простота, что позволяет автоматизировать процесс добавления и удаления элементов.

В настоящее время особое внимание уделяется динамике саморазвивающихся систем, в частности социальных сетей. Интерес представляет не только процесс распространения информации по сети с заданной структурой, но и изменение самой структуры графа. Более того, во многих случаях граф обладает уникальными свойствами — например, безмасштабностью, при которой распределение вершин по числу связей подчиняется степенному закону. В связи с этим возникает обратная задача — создание модели, способной воспроизвести динамику роста $G(t)$, наблюдаемую в реальной сети.

ДИНАМИЧЕСКИЙ И ПРЕФРАКТАЛЬНЫЙ ГРАФЫ

Понятие «динамические сети» (Dynamic networks) широко применяется при исследовании сложных сетей с изменяющейся структурой, возникающих в различных областях и имеющих разное происхождение [1, 2]. К динамическим сетям относятся социальные сети, сети связи и коллективного взаимодействия, структуры фондовых рынков, а также системы взаимных обязательств в межбанковской сфере¹ [3–7]. Несмотря на значительные объемы эмпирических данных, накопленных при изучении динамических сетей, пока нельзя говорить о сформировавшейся теории динамического анализа сетей (Dynamic network analysis) или сетевой науки (Network science). Для создания такой прикладной дисциплины необходима

прочная теоретическая база, которой может стать развивающаяся динамическая теория графов, где динамический граф (модель динамической сети) является основным объектом.

Динамический граф как модель динамической сети представляет собой последовательность классических графов, не содержащих параллельных ребер и петель, между которыми происходят переходы, описываемые различными операциями теории графов: удаление или добавление ребра, удаление или добавление вершины, замена вершины «затравкой», приоритетное присоединение вершин и ребер и др.² [8]. Индекс в такой последовательности соответствует своеобразному «топологическому времени», в моментах которого меняется структура графа.

Операции удаления или добавления ребра и вершины принято называть простыми или базовыми. Любая другая операция, которую можно представить как последовательность простых операций, считается сложной. В общем виде динамический граф — это последовательность конечных невзвешенных (не всегда связанных) графов, где переход от одного графа к следующему реализуется при помощи некоторой операции, как простой, так и сложной. Для построения траектории динамического графа может использоваться несколько (конечное число) чередующихся операций. При этом в каждой операции может задаваться механизм выбора элементов графа (ребер, вершин, подграфов), над которыми выполняется соответствующее действие. Последовательность графов, образующая динамический граф, называется *траекторией динамического графа*. Например, фрактальный граф — один из частных случаев динамического графа [9–11].

НЕКОТОРЫЕ ПРИЛОЖЕНИЯ ДИНАМИЧЕСКОЙ ТЕОРИИ ГРАФОВ

Методы динамической теории графов особенно полезны при проектировании командно-информационного взаимодействия между подвижными абонентами в сетевых системах. Под сетевыми

¹ Шерешева М.Ю. Формы сетевого взаимодействия компаний. Курс лекций. М.: Издательский дом Высшей школы экономики; 2010. 339 с. URL: https://www.researchgate.net/publication/342917874_Sereseva_M_U_Formy_setevogo_vzaimodejstvia_kompanij_M_VSE_2010_339_s

² Емеличев В.А., Мельников О.И., Сарванов В.И., Тышкевич Р.И. Лекции по теории графов. М.: УРСС; 2009. 392 с.; Уилсон Р. Введение в теорию графов: Вводные курсы. М.: Мир; 1977. 208 с.



системами понимаются технические комплексы, функционирующие на сетевых структурах. В этом контексте сетевые системы представляют собой преимущественно инженерное понятие, а не строго математическую категорию.

История развития беспроводных сетей демонстрирует расширение сферы применения данного класса телекоммуникационных технологий. В настоящее время беспроводные сети превосходят проводные аналоги по таким параметрам, как безопасность, стоимость, надежность, функциональность и удобство использования. Вместе с тем спектр задач, связанных с новыми приложениями беспроводных технологий и сетей, постоянно расширяется. В частности, выделяются две основные области применения беспроводных сетей — телекоммуникации и мониторинг. В крупных системах беспроводные сети могут одновременно выполнять функции передачи данных и мониторинга [12].

Особое внимание уделяется сетям с подвижными абонентами (датчиками, сенсорами), обеспечение качественной и постоянной связи с которыми является крайне важной задачей. Успешное решение этой задачи позволит повысить связность и скорость передачи информации между мобильными узлами, а также сократить затраты на наземный сегмент сети за счет эффективной маршрутизации и ретрансляции через подвижные узлы. Сложность решения возрастает с увеличением числа абонентов, при этом очевидно, что максимальная эффективность достигается при скоординированных действиях участников сети. За последнее десятилетие в работах зарубежных и отечественных исследователей все чаще появляются разработки, основанные на новой концепции организации совместных действий. Однако большинство существующих алгоритмов сетевого взаимодействия имеют ограниченную область применения и, по сути, представляют собой конкретные инженерные решения.

Развивающаяся динамическая теория графов может стать теоретической основой для создания алгоритмов командно-информационного взаимодействия подвижных абонентов в сетевых системах, а также в сетевых системах непрерывного пространственного мониторинга. Топология сети таких абонентов не может оставаться неизменной — она вынуждена изменяться под воздействием различных факторов, например, при увеличении числа участников. Поскольку эффективность передачи данных зависит от длины цепочки абонентов, при добавлении новых узлов целесообразно не допускать роста диаметра сети.

ПОСТАНОВКА МНОГОКРИТЕРИАЛЬНОЙ ЗАДАЧИ НА МНОВЗВЕШЕННОМ ГРАФЕ

Формулировка многокритериальной задачи на предфрактальном графе представлена следующим образом [13, 14]. Рассматривается предфрактальный граф, сформированный множеством затравок. Каждому ребру графа соответствует набор из действительных чисел, включающий ранг ребра и коэффициенты подобия.

На множестве допустимых решений задается векторная целевая функция (ВЦФ), выраженная через критерии двух типов: метрические и топологические. Метрические критерии являются функциями весовых характеристик ребер, тогда как топологические основаны на топологических свойствах предфрактального графа.

Во множестве допустимых решений необходимо выделить такой элемент, который является экстремальным для векторной целевой функции, т.е. достигает оптимальных значений по всем критериям одновременно.

Количество критериев может варьироваться в зависимости от конкретной задачи. При этом наборы весов считаются несравнимыми и разного типа, что исключает возможность выполнения арифметических операций между ними. Если же предположить, что наборы весов одного типа и арифметические операции над ними допустимы, то их можно заменить одним агрегированным весом, например, средним значением, что сводит задачу к многокритериальной с единым набором весов.

Рассматривая задачу Z_i , будем иметь в виду множество всех ее индивидуальных задач. Обозначим через X_i множество допустимых решений (МДР) всех этих индивидуальных задач.

Многокритериальная задача называется полной, если для любого множества допустимых решений существуют такие параметры ее векторной целевой функции (ВЦФ), при которых выполняется равенство $X^0 = \tilde{X} = X$.

Для предфрактальных графов справедливы следующие леммы из теории многокритериальной оптимизации.

Лемма 1. Добавление новых критериев к ВЦФ отдельной индивидуальной задачи либо не изменяет ее множество парето-оптимальных решений (ПМ) и полное множество альтернатив (ПМА), либо расширяет их за счет новых альтернатив.

Лемма 2. При фиксированном t некоторые индивидуальные задачи Z_i из семейства могут быть полными, тогда как другие — не обладают этим свойством.



Индивидуальные задачи одного семейства имеют одинаковое определение множества допустимых решений, но отличаются размерностью ВЦФ, составом критериев, количеством наборов весов и прочими параметрами.

Для полных задач анализ множеств альтернатив упрощается, что позволяет применять известные теоремы теории графов к предфрактальным графам. Подробные доказательства можно найти в [15].

Для многовзвешенных предфрактальных графов важна следующая лемма.

Лемма 3. Задача нахождения множества парето-оптимальных решений (а также полного множества альтернатив) типовой задачи на многовзвешенном предфрактальном графе неразрешима с помощью алгоритмов линейной свертки.

Тем не менее алгоритмы линейной свертки применимы для поиска нетривиальных множеств альтернатив, в частности лексикографического множества альтернатив (ЛМА).

Лемма 4. Задача нахождения ЛМА типовой целочисленной задачи при условии наличия только весовых критериев на многовзвешенном предфрактальном графе разрешима с использованием алгоритмов линейной свертки.

О НАСЛЕДСТВЕННЫХ СВОЙСТВАХ ДИНАМИЧЕСКИХ ГРАФОВ

В классической теории графов одной из ключевых экстремальных задач является поиск подграфа или остова с определенными свойствами, например, дерева минимального веса. В динамической теории графов основное внимание уделяется выявлению взаимосвязи между решениями таких экстремальных задач на различных «классических» (стационарных) графах, которые входят в состав динамического графа. Если решения на разных графах можно сопоставить по заданным критериям, то говорят о наличии свойства наследственности в классе динамических графов, объединенных общими правилами перехода в последовательностях образующих их графов. Следующим логичным шагом является формализация связи между этим свойством наследственности и операциями перехода в траектории, формирующие динамический граф. При успешном установлении такой связи возникает возможность программируемой самоорганизации, т.е. обеспечения гарантированных наследственных структурных свойств и характеристик динамических графов [16–19].

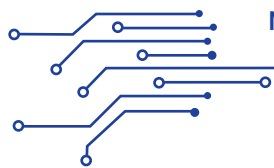
ВЫВОДЫ

В настоящее время особое внимание уделяется изучению динамики саморазвивающихся систем. Интерес вызывает не только процесс распространения информации по сети с фиксированной структурой, но и само изменение этой структуры во времени. Более того, во многих случаях графы обладают уникальными характеристиками, например, свойствами безмасштабных сетей.

Понятие динамических сетей (Dynamic networks) широко применяется при исследовании сложных структур, которые изменяются во времени и имеют различное происхождение. К таким сетям относятся социальные сети, коммуникационные и коллективные системы взаимодействия, структуры фондовых рынков, а также взаимосвязи в межбанковских системах. Несмотря на значительный эмпирический материал, накопленный в области динамических сетей, пока не сформировалась устоявшаяся теория динамического анализа сетей (Dynamic network analysis) или сетевой науки (Network science). Для создания полноценной прикладной дисциплины необходима крепкая теоретическая база, которой может стать зарождающаяся динамическая теория графов, в центре внимания которой находятся динамические графы — модели динамических сетей.

Эта развивающаяся теория графов способна стать фундаментом для разработки алгоритмов командно-информационного взаимодействия подвижных участников сетевых систем. В таких системах топология сети не может оставаться неизменной — она вынуждена изменяться под воздействием различных факторов, например, увеличения числа участников. Поскольку эффективность передачи информации зависит от длины цепочки участников, важно при добавлении новых узлов не допускать роста диаметра сети.

В данной работе предложены основы динамической теории графов и рассмотрены некоторые инженерные приложения. Представлено описание множества альтернативных решений многокритериальных задач на предфрактальных графах. Предложена общая математическая постановка дискретной многокритериальной задачи на многовзвешенном предфрактальном графе. Проведена классификация таких задач и дано определение вычислительной сложности соответствующих алгоритмов. Также описан подход к оценке сложности нахождения множества альтернатив. Одним из ключевых выводов является выявление наследственных свойств динамических графов.



БЛАГОДАРНОСТИ

Статья подготовлена по результатам исследований, выполненных за счет средств Научного фонда Финансового университета при Правительстве Российской Федерации.

ACKNOWLEDGMENTS

The article is based on the results of research carried out at the expense of the Scientific Fund of the Financial University under the Government of the Russian Federation.

СПИСОК ИСТОЧНИКОВ

1. Робертс Ф. С. Дискретные математические модели с приложениями к социальным, биологическим и экологическим задачам. Пер. с англ. М.: Наука; 1986. 494 с.
2. Westaby J. D. Dynamic Network Theory: How Social Networks Influence Goal. American Psychological Association; 2011. 279 p.
3. Губанов Д. А., Новиков Д. А., Чхартишвили А. Г. Социальные сети: модели информационного влияния, управления и противоборства. М.: Физматлит; 2010. 228 с.
4. Кучерявый А. Е., Прокопьев А. В., Кучерявый Е. А. Самоорганизующиеся сети. СПб.: Любавич; 2011. 311 с.
5. Голдсмит А., Медар М., Эффрос М. Самоорганизующиеся беспроводные сети. *В мире науки*. 2012;(6):76–81. URL: <https://www.elibrary.ru/ocsgzt>
6. Визгунов А. Н., Гольденгорин Б. И., Замараев В. А. и др. Применение рыночных графов к анализу фондового рынка России. *Журнал Новой экономической ассоциации*. 2012;3(15):66–81. URL: <https://www.elibrary.ru/pfiuxn>
7. Georg C. P. The Effect of Interbank Network Structure on Contagion and Common Shocks. *Journal of Banking & Finance*. 2013;37(7):2216–2228. DOI: 10.1016/j.jbankfin.2013.02.032
8. Krön B. Growth of self-similar graphs. *J. Graph Theory*. 2004;45(3):224–239. DOI: 10.1002/jgt.10157
9. Кочкаров А. А. Структурная динамика: свойства и количественные характеристики предфрактальных графов. М.: Вега-Инфо; 2012. 120 с.
10. Кочкаров А. А., Малинецкий Г. Г., Кочкаров Р. А. Некоторые аспекты динамической теории графов. *Журнал вычислительной математики и математической физики*. 2015;55(9):1623–1629. DOI: 10.7868/S 0044466915090094
11. Rozenfeld H. D., Gallos L. K., Song Ch., Makse H. A. Fractal and Transfractal Scale-Free Networks. *Mathematics of Complexity and Dynamical Systems*. Robert A. Meyers, ed. New York: Springer; 2012. 1858 p. DOI: 10.1007/978-0-387-30440-3_231
12. Кочкаров А. М. Распознавание фрактальных графов. Алгоритмический подход. Нижний Архыз: РАН CAO; 1998. 170 с.
13. Казанцев А. М., Кочкаров Р. А., Тимошенко А. В., Сычугов А. А. Некоторые подходы к оценке процесса функционирования структурно-динамических систем мониторинга в условиях внешних воздействий. *Моделирование, оптимизация и информационные технологии*. 2021;4(35):14. DOI: 10.26102/2310–6018/2021.35.4.005
14. Кочкаров Р. А. Задачи многокритериальной оптимизации на многовзвешенных предфрактальных графах. М.: Академинновация; 2014. 189 с. URL: <https://www.elibrary.ru/tyiygd>
15. Перепелица В. А. Многокритериальные модели и методы для задач оптимизации на графах. LAP Lambert Academic Publishing; 2013. 336 с. URL: <https://n.eruditor.one/file/1399011/>
16. Павлов Д. А. Многокритериальная задача покрытия предфрактального графа простыми цепями. Дисс. ... канд. физ.-мат. наук. Таганрог: Таганрогский государственный радиотехнический университет; 2004. 110 с. URL: https://new-disser.ru/_avtoreferats/01002738834.pdf
17. Milgram S. The small world problem. *Psychology Today*. 1967;(2):60–67. DOI: 10.1007/978–3–658–21742–6_94
18. Подлазов А. В., Щетинина Д. П. Модель роста социальной сети. *Препринты ИПМ им. М. В. Келдыша РАН*. 2013;(95):1–16. URL: https://keldysh.ru/papers/2013/prep2013_95.pdf
19. Митин Н. А., Подлазов А. В., Щетинина Д. П. Исследование сетевых свойств Живого журнала. *Препринты ИПМ им. М. В. Келдыша РАН*. 2012;(78):1–16. URL: https://keldysh.ru/papers/2012/prep2012_78.pdf

REFERENCES

1. Roberts F. S. Discrete Mathematical Models with Applications to Social, Biological, and Ecological Problems. Transl. from Eng. Moscow: Nauka; 1986. 494 p. (In Russ.).

2. Westaby J.D. Dynamic Network Theory: How Social Networks Influence Goal. American Psychological Association; 2011. 279 p.
3. Gubanov D.A., Novikov D.A., Chkhartishvili A. G. Social networks: Models of information influence, control and confrontation. Moscow: Fizmatlit; 2010. 228 p. (In Russ.).
4. Kucheryavy A.E., Prokopyev A.V., Kucheryavy E.A. Self-organizing networks. St. Petersburg: Lyubavich; 2011. 311 p. (In Russ.).
5. Goldsmith A., Medar M., Effros M. Self-organizing wireless networks. *V mire nauki*. 2012;(6):76–81. URL: <https://www.elibrary.ru/ocsgzt> (In Russ.).
6. Vizgunov A.N., Goldengorin B.I., Zamaraev V.A., et. al. Application of market graphs to the analysis of the Russian stock market. *New Economic Journal*. 2012;3(15):66–81. URL: <https://www.elibrary.ru/pfiuxn> (In Russ.).
7. Georg C.P. The Effect of Interbank Network Structure on Contagion and Common Shocks. *Journal of Banking & Finance*. 2013;37(7):2216–2228. DOI: 10.1016/j.jbankfin.2013.02.032
8. Krön B. Growth of self-similar graphs. *J. Graph Theory*. 2004;45(3):224–239. DOI: 10.1002/jgt.10157
9. Kochkarov A.A. Structural Dynamics: Properties and Quantitative Characteristics of Prefractal Graphs. Moscow: Vega-Info; 2012. 120 p. (In Russ.).
10. Kochkarov A. A., Malinetsky G. G., Kochkarov R. A. Some Aspects of Dynamic Graph Theory. *Computational Mathematics and Mathematical Physics*. 2015;55(9):1623–1629. (In Russ.). DOI: 10.7868/S 0044466915090094
11. Rozenfeld H. D., Gallos L. K., Song Ch., Makse H. A. Fractal and Transfractal Scale-Free Networks. *Mathematics of Complexity and Dynamical Systems*. Robert A. Meyers, ed. New York: Springer; 2012. 1858 p. DOI: 10.1007/978-0-387-30440-3_231
12. Kochkarov A.M. Recognition of fractal graphs. Algorithmic approach. Nizhniy Arkhyz: RAS SAO; 1998. 170 p. (In Russ.).
13. Kazantsev A.M., Kochkarov R.A., Timoshenko A.V., Sychugov A.A. Some approaches to assessing the functioning of structural-dynamic monitoring systems under external influences. *The scientific journal Modeling, Optimization and Information Technology*. 2021;4(35):14. (In Russ.). DOI: 10.26102/2310–6018/2021.35.4.005
14. Kochkarov R. A. Multicriteria Optimization Problems on Multiweighted Prefractal Graphs. Moscow: AkademInnovatsiya; 2014. 189 p. URL: <https://www.elibrary.ru/tyiygd> (In Russ.).
15. Perepelitsa V.A. Multicriteria Models and Methods for Optimization Problems on Graphs. LAP Lambert Academic Publishing; 2013. 336 p. URL: <https://n.eruditor.one/file/1399011/> (In Russ.).
16. Pavlov D.A. Multicriteria Problem of Covering a Prefractal Graph with Simple Chains. Diss. Cand. Sci. (Phys.-Math.) Taganrog: Taganrog State Radio Engineering University; 2004. 110 p. URL: https://new-dissert.ru/_avtoreferats/01002738834.pdf (In Russ.).
17. Milgram S. The Small World Problem. *Psychology Today*. 1967;(2):60–67. DOI: 10.1007/978-3-658-21742-6_94
18. Podlazov A.V., Shchetinina DP Social network growth model. *Preprinty IPM im. M. V. Keldysha RAN*. 2013;(95):1–16. URL: https://keldysh.ru/papers/2013/prep2013_95.pdf (In Russ.).
19. Mitin N.A., Podlazov A.V., Shchetinina D.P. Study of network properties of LiveJournal. *Preprinty IPM im. M. V. Keldysha RAN*. 2012;(78):1–16. URL: https://keldysh.ru/papers/2012/prep2012_78.pdf (In Russ.).

ИНФОРМАЦИЯ ОБ АВТОРАХ / ABOUT THE AUTHORS

Расул Ахматович Кочкаров — кандидат экономических наук, доцент кафедры искусственного интеллекта факультета информационных технологий и анализа больших данных, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

Rasul A. Kochkarov — Cand. Sci. (Econ.), Assoc. Prof. of the Department of Artificial Intelligence, Faculty of Information Technology and Big Data Analysis, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

<https://orcid.org/0000-0003-3186-3901>

Автор для корреспонденции / Corresponding author:

rkochkarov@fa.ru



Азрет Ахматович Кочкаров — доктор технических наук, доцент, заместитель директора по инновационной работе ФИЦ Биотехнологии РАН, Москва, Российская Федерация; профессор кафедры искусственного интеллекта факультета информационных технологий и анализа больших данных, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

Azret A. Kochkarov — Dr. Sci. (Tech.), Assoc. Prof., Federal Research Center of Biotechnology of the Russian Academy of Sciences, Moscow, Russian Federation; Prof. Department of Artificial Intelligence, Faculty of Information Technology and Big Data Analysis, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

<https://orcid.org/0000-0002-3232-5331>

akochkar@fbras.ru

Заявленный вклад авторов:

А.А. Кочкаров — разработка общей концепции статьи, введение и выводы.

Р.А. Кочкаров — приложения динамической теории графов, обзор и список литературы, выводы и леммы.

Authors' declared contributions:

A. A. Kochkarov — development of the general concept of the article, introduction and conclusions.

R. A. Kochkarov — applications of dynamic graph theory, review and list of literature, conclusions and lemmas.

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Conflicts of Interest Statement: The authors have no conflicts of interest to declare.

Статья поступила 04.07.2025; принята к публикации 31.07.2025.

Авторы прочитали и одобрили окончательный вариант рукописи.

The article was received 04.07.2025; accepted for publication 31.07.2025.

The authors read and approved the final version of the manuscript.

Прогнозирование успешности проектов на основе анализа структурных характеристик коммуникационных сетей

Д.А. Павлов

Кубанский государственный аграрный университет им. И.Т. Трубилина, Краснодар, Российская Федерация

АННОТАЦИЯ

В статье представлена методика прогнозирования успешности проектных инициатив на основе анализа структурных характеристик коммуникационных сетей проектных команд. Исследование базируется на данных из систем трекинга задач (Jira, Trello), отражающих формальные взаимодействия участников проекта. **Методология исследования** включает комплекс аналитических инструментов: корреляционный анализ с применением коэффициентов Спирмена и Пирсона, регрессионное моделирование на основе алгоритма Random Forest, а также методы обнаружения аномалий с использованием Isolation Forest. Исследование показало статистически значимые взаимосвязи между ключевыми сетевыми метриками (betweenness centrality, плотность сети, диаметр графа) и показателями эффективности проектов (соблюдение сроков, бюджета, качество результатов). Выявлена статистически значимая отрицательная корреляция между избыточной централизацией и соблюдением сроков ($\rho = -0,72$), а также положительная взаимосвязь между плотностью сети и качеством результатов ($r = 0,68$). **Разработанная модель** на основе Random Forest демонстрирует точность прогнозирования успешности проектов на уровне 84%. Установлено, что избыточная централизация коммуникаций снижает вероятность успешной реализации проекта, тогда как оптимальная плотность коммуникационной сети способствует достижению проектных KPI. **Практическая значимость** исследования заключается в возможности раннего выявления рисков срыва проектов на основе объективных метрик коммуникационной активности. Разработанная методика, апробированная на данных IT-компании, позволяет не только прогнозировать риски, но и формировать рекомендации по оптимизации командных взаимодействий. Результаты исследования представляют интерес для проектных менеджеров, HR-аналитиков и специалистов по data-driven управлению.

Ключевые слова: управление проектами; сетевой анализ; коммуникационные сети; прогнозирование успешности; машинное обучение; трекары задач

Для цитирования: Павлов Д.А. Прогнозирование успешности проектов на основе анализа структурных характеристик коммуникационных сетей. *Цифровые решения и технологии искусственного интеллекта*. 2025;1(3):37-43. DOI: 10.26794/3033-7097-2025-1-3-37-43

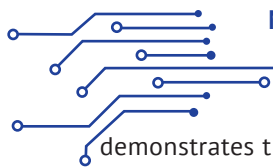
Forecasting the Success of Projects Based on the Analysis of Structural Characteristics of Communication Networks

D.A. Pavlov

I.T. Trubilin Kuban State Agrarian University, Krasnodar, Russian Federation

ABSTRACT

The article presents a methodology for predicting the success of project initiatives based on an analysis of the structural characteristics of project team communication networks. The study is based on data from task tracking systems (Jira, Trello) that reflect formal interactions between project participants. The research methodology includes a set of analytical tools: correlation analysis using Spearman's and Pearson's coefficients, regression modeling based on the Random Forest algorithm, and anomaly detection methods using Isolation Forest. The study revealed statistically significant correlations between key network metrics (betweenness centrality, network density, graph diameter) and project performance indicators (adherence to deadlines, budget, quality of results). A statistically significant negative correlation was found between excessive centralization and adherence to deadlines ($\rho = -0.72$), as well as a positive correlation between network density and quality of results ($r = 0.68$). The developed model based on Random Forest



demonstrates the accuracy of forecasting the success of projects at the level of 84%. It was found that excessive centralization of communications reduces the probability of successful project implementation, while the optimal density of the communication network contributes to the achievement of project KPIs. The practical significance of the study lies in the possibility of early detection of project failure risks based on objective metrics of communication activity. The developed methodology, tested on IT company data, allows not only to predict risks, but also to form recommendations for optimizing team interactions. The results of the study are of interest to project managers, HR analysts, and data-driven management specialists.

Keywords: project management; network analysis; communication networks; success prediction; machine learning; task trackers

For citation: Pavlov D.A. Forecasting the success of projects based on the analysis of structural characteristics of communication networks. *Digital solutions and artificial intelligence technologies*. 2025;1(3):37-43. DOI: 10.26794/3033-7097-2025-1-3-37-43

ВВЕДЕНИЕ

В условиях современной проектной деятельности, характеризующейся высокой динамикой и сложностью, эффективность командных коммуникаций выступает ключевым фактором успеха. Традиционные показатели управления проектами — соблюдение сроков, контроль бюджета и обеспечение качества — часто не позволяют выявить скрытые закономерности взаимодействия между участниками. При этом именно эти паттерны могут заблаговременно сигнализировать о потенциальных рисках и препятствиях в реализации проекта¹. Анализ коммуникационных графов, построенный на данных из систем управления задачами (таких как Jira, Trello, Asana), предоставляет инновационные инструменты для прогнозирования успешности реализации проектов.

Значимость данного исследования определяется растущим трендом на управление на основе данных (data-driven), особенно в сферах с высокой степенью неопределенности: разработке программного обеспечения, научно-исследовательских разработках (R&D) и корпоративных преобразованиях.

Эмпирические исследования демонстрируют, что проекты с недостаточно структурированными коммуникационными процессами чаще подвержены рискам задержек, возникновению конфликтных ситуаций и перерасходу ресурсов [1]. При этом до настоящего момента не разработаны универсальные методики для количественной оценки воздействия сетевых взаимодействий на конечные показатели проектной деятельности.

Цель исследования заключается в разработке комплексного подхода к прогнозированию успешности проектов на основе анализа коммуникационных графов с применением сетевых метрик (центральность, кластеризация, плотность связей) и методов машинного обучения.

В качестве целевых показателей рассматриваются как традиционные KPI (соблюдение сроков, бюджет), так и качественные метрики (удовлетворенность команды, количество инцидентов).

Научная новизна работы состоит в интеграции методов сетевого анализа и прогнозного моделирования в сфере управления проектами. Предложенный подход отличается от существующих исследований, ориентированных преимущественно на ретроспективный анализ, и позволяет:

- выявлять потенциальные проблемы в коммуникационных процессах до их критического проявления;
- оценивать риски проектного провала на основе анализа динамики взаимодействий;
- формировать практические рекомендации по оптимизации командных процессов.

Практическая значимость результатов исследования заключается в возможности их применения в agile-командах, распределенных проектах и корпоративных системах управления для повышения прозрачности и предсказуемости проектных результатов.

ИССЛЕДОВАНИЕ

В рамках исследования рассматривается гипотеза о взаимосвязи структурных характеристик коммуникационной сети проекта и ключевых показателей эффективности (KPI), включающих: бюджет, сроки, качественные метрики.

Предполагается, что отклонения от плановых показателей вызваны следующими неоптимальными коммуникационными паттернами:

- чрезмерной централизацией информационных потоков (высокие значения betweenness centrality отдельных узлов);
- низкой связностью между функциональными подгруппами (повышенный clustering coefficient);
- асимметричным распределением коммуникационной нагрузки (значительный разброс значений degree centrality между участниками) [2, 3].

¹ Поляков Н.А., Мотовилов О.В., Лукашов Н.В. Управление инновационными проектами. Учебник и практикум для вузов. 2-е изд. М.: Юрайт; 2022. 384 с.



Эти отклонения могут количественно фиксироваться через динамику сетевых метрик на ранних этапах проекта, что позволяет прогнозировать риски до их реализации [4–6].

В качестве эмпирической базы использованы данные системы управления проектами Jira² за период 2022–2024 гг.

Анализируемый набор данных включает: историю коммуникаций (назначения задач, упоминания участников, комментарии), временные метки выполнения задач (плановые / фактические сроки), показатели качества (количество критических багов, частота переделок).

Для построения графа коммуникаций в качестве узлов выступали участники проекта (разработчики, тестировщики, менеджеры), а ребрами — взаимодействия (частота и направленность сообщений), где вес соответствует интенсивности взаимодействий.

Для оценки влияния внутрипроектных коммуникаций на эффективность проекта предлагается использовать сетевые метрики [3]:

- 1) размерность;
- 2) плотность;
- 3) диаметр;
- 4) средняя степень;
- 5) меры центральности (центральность по посредничеству, центральность по степени и центральность по близости);
- 6) коэффициент кластеризации;
- 7) максимальная клика;
- 8) K -ядро.

Размер — это главная характеристика сети, определяющая количество участников, обычно называемых узлами или вершинами.

Плотность графа $G(V, E)$ вычисляется как нормированное число ребер:

$$\text{den}(G) = \frac{2|E|}{|V|(|V|-1)},$$

где $|E|$ — число существующих ребер; $|V|$ — количество вершин графа G .

Диаметр сети $d(G) = \max \rho(v, w)$.

Степенью вершины называется количество ребер, присоединенных к вершине. Средняя степень вершин — это среднее значение степеней вершин графа G .

$$\text{deg}(G) = \frac{\sum_{v \in V} \text{deg}(v)}{|V|}.$$

Средняя степень вершин определяет среднее количество связей участников сети.

Меры центральности позволяют оценить важность участников сети.

Степень центральности (*degree centrality*) — простейшая мера центральности для вычисления, основанная на степени узла, позволяющая определить наиболее весомых участников сети:

$$C_d = \frac{1}{(|V|-1)(|V|-2)} \sum_{i=1}^n (\text{deg}_{\max} - \text{deg}_i),$$

где C_d — централизация сети по степени; deg_{\max} — максимальная степень сети; deg_i — степень участника i ; $|V|$ — количество участников сети.

$C_D(i)$ отражает число связей i -го узла.

Центральность по близости (*closeness centrality*) — это способ обнаружения агентов, способных очень эффективно распространять информацию по сети:

$$C_c = \frac{(2|V|-3)}{(|V|-1)(|V|-2)} \sum_{i=1}^n (C_{\max} - C_i),$$

где C_{\max} — максимальная близость участника сети; C_i — близость участника i ; $|V|$ — количество участников сети.

Центральность близости узла измеряет его среднюю удаленность (обратное расстояние) от всех других узлов. Узлы с высокой оценкой близости имеют самые короткие расстояния до всех других узлов.

Центральность по посредничеству (*betweenness centrality*) — это коэффициент, говорящий о том, какое количество раз узел является посредником между всеми кратчайшими путями всех остальных узлов. Узлы с высоким коэффициентом центральности по посредничеству — важные структурные элементы, которые потенциально соединяют ее подструктуры и являются связующими звеньями.

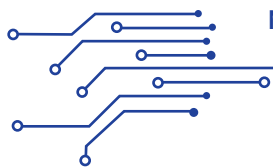
Для вычисления этой метрики используется формула:

$$C_b = \frac{2}{(|V|-1)^2(|V|-2)} \sum_{i=1}^n (B_{\max} - B_i),$$

где C_b — централизация сети по срединности; B_{\max} — максимальный показатель срединности в сети; B_i — срединность участника i ; $|V|$ — количество участников сети.

Коэффициент кластеризации вершины i в графе $G(V, E)$, обозначаемый C_i , показывает, какая часть соседей вершины i связана. Формально коэффициент кластеризации имеет вид:

² URL: <https://www.atlassian.com/ru/software/jira/guides/projects/overview?ysclid=mdiiyunr5p649485372#what-is-a-jira-project>



$$C_i = \frac{2e_i}{k_i(k_i - 1)},$$

где e_i обозначает количество ребер между соседями вершины i , а k_i — это число соседей вершины i .

Когда коэффициент кластеризации высокий — это означает, что сеть чрезвычайно плотно сгруппирована вокруг нескольких узлов; когда низкий — это значит, что связи в сети относительно равномерно распространены среди всех узлов.

Средний коэффициент кластеризации графа $G(V, E)$, обозначаемый \bar{C} , определяется как среднее значение коэффициентов кластеризации всех вершин $v \in V$. Формально средний коэффициент кластеризации определяется выражением:

$$\bar{C} = \frac{1}{|V|} \sum_{i=1}^{|V|} C_i.$$

Коэффициент кластеризации показывает тенденцию участников внутрипроектных коммуникаций формировать более тесные подгруппы, т.е. кластеры. Один из способов изучить подгруппы сети внутрипроектных коммуникаций заключается в исследовании социальной сплоченности [7].

Сетевые метрики представляют собой комплексный показатель, характеризующий свойства коммуникационной сети в целом [8]. Для оценки сплоченности команды проекта используются такие показатели, как метрики клика и k -ядра коммуникационных структур, которые позволяют анализировать степень взаимосвязанности участников проекта.

Рисунок содержит пример представления графов коммуникаций между участниками проекта.

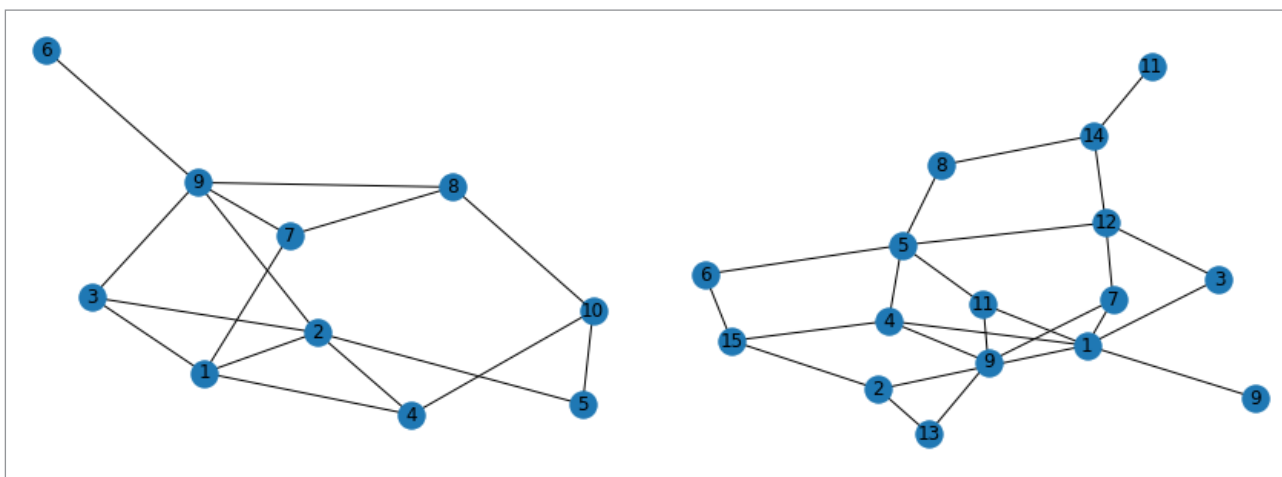


Рис. / Fig. Графы коммуникаций / Communications Network

Источник / Source: составлено автором / Compiled by the author.

НАБОРЫ ДАННЫХ

Экспериментальная проверка проведена на выборке, включающей 27 завершённых IT-проектов длительностью 6–18 месяцев (состав каждой проектной группы 5–14 чел.), рассчитаны сетевые метрики внутренних коммуникаций для выбранных проектов.

Результаты верифицированы через сравнение с экспертной оценкой менеджеров.

МЕТОДЫ АНАЛИЗА КОММУНИКАЦИОННЫХ СЕТЕЙ ПРОЕКТА

Для выявления статистических закономерностей между структурными характеристиками коммуникационных сетей и показателями успешности проектов применен комплекс аналитических методов, включающий корреляционный анализ, регрессионное моделирование и методы обнаружения аномалий.

Корреляционный анализ проводился с использованием коэффициентов Спирмена (ρ) и Пирсона (r). Коэффициент Спирмена применялся для выявления монотонных зависимостей между сетевыми метриками (betweenness centrality, closeness centrality, degree centrality, плотность сети, диаметр графа, модулярность) и показателями успешности проектов (отклонения по срокам, бюджету, качеству результатов). Расчеты показали значимую отрицательную корреляцию между betweenness centrality ключевых участников и соблюдением сроков ($\rho = -0,72$, $p < 0,01$), что свидетельствует о негативном влиянии избыточной централизации коммуникаций [8, 9]. Коэффициент Пирсона выявил сильную положительную линейную зависимость между



плотностью коммуникационной сети и качеством кода ($r = 0,68$, $p < 0,05$).

Регрессионное моделирование включало два подхода: логистическую регрессию для бинарной классификации проектов (успешные/неуспешные) и метод случайных лесов (Random Forest) для мультиклассовой классификации и оценки важности предикторов. Логистическая регрессия продемонстрировала, что увеличение betweenness centrality тимлида на одно стандартное отклонение снижает вероятность успеха проекта на 60% ($OR = 0,4, 95\%$ ДИ $[0,25; 0,64]$). Модель Random Forest с 500 деревьями (точность 84% по данным кросс-валидации) показала наибольшую важность трех предикторов: betweenness centrality владельца продукта (0,32), плотность сети на 3-й месяц проекта (0,25) и closeness centrality технического лидера (0,18).

Метод Isolation Forest применен для раннего обнаружения аномальных коммуникационных паттернов. Алгоритм выявлял проекты с экстремальными значениями сетевых метрик ($score > 0,8$), которые в 87% случаев приводили к значительным отклонениям по срокам или бюджету. Наибольшую прогностическую ценность показали: внезапный рост betweenness centrality одного участника ($score > 0,9$) и резкое снижение плотности сети в критических фазах проекта ($score > 0,85$).

Интеграция различных методов представляет собой комплексную систему оценки и прогнозирования успешности проектов [10, 11]. Полученные результаты подтверждают гипотезу о существовании статистически значимой взаимосвязи между характеристиками коммуникационных сетей и ключевыми показателями эффективности проектов.

ВЫВОДЫ

Проведенное исследование демонстрирует, что анализ структурных характеристик коммуникационных сетей на основе используемых данных позволяет разработать эффективный инструментарий для прогнозирования успешности проектов. Применение комплекса методов (корреляционный анализ, регрессионное моделирование, обнаружение аномалий [12]) выявило статистически значимые взаимосвязи между ключевыми сетевыми метриками (betweenness centrality, плотность сети, диаметр графа) и показателями эффективности проектов (соблюдение сроков, бюджета, качество результатов). Установлено, что избыточная централизация коммуникаций (высокие значения betweenness centrality отдельных участников) достоверно снижает вероятность успешной реализации проекта, в то время как оптимальная плотность коммуникационной сети способствует достижению проектных KPI. Разработанная методика, апробированная на данных IT-компаний, позволяет не только прогнозировать риски на ранних этапах, но и формировать рекомендации по оптимизации командных взаимодействий [13]. Перспективными направлениями дальнейших исследований представляются: учет неформальных коммуникаций, разработка динамических моделей прогнозирования и адаптация методики для различных отраслевых контекстов. Полученные результаты вносят вклад в развитие data-driven подхода к управлению проектами и открывают новые возможности для повышения эффективности командной работы.

СПИСОК ИСТОЧНИКОВ

1. Кон М. Agile: Оценка и планирование проектов. Пер. с англ. М.: Альпина Паблишер; 2022. 418 с. URL: <https://stroki.mts.ru/book/agile-otsenka-i-planirovaniye-proyektov-1759?ysclid=mdhbufcngh64422956>
2. Durland M., Fredericks K. An Introduction to Social Network Analysis. *New Directions for Evaluation*. 2006;2005(107):5–13. DOI: 10.1002/ev.157
3. Otte E., Rousseau R. Social network analysis: A powerful strategy, also for the information sciences. *Journal of Information Science*. 2002;28(6):441–453. DOI: 10.1177/016555150202800601
4. Krackhardt D., Hanson J.R. Informal networks: The company behind the charts. *Harvard Business Review*. 2000;78(4):104–11. URL: https://www.researchgate.net/publication/40961009_Informal_Networks_the_Company_behind_the_Charts
5. Pryke S.D. Towards a social network theory of project governance. *Construction Management and Economics*. 2005;23(9):927–939. DOI: 10.1080/01446190500184196
6. Chinowsky P., Diekmann J. and Galotti V. Social network model of construction. *Journal of Construction Engineering and Management*. 2008;134(10). DOI: 10.1061/(ASCE)0733-9364(2008)134:10(804)
7. Титов С.А. Исследование взаимосвязи структуры внутрипроектных коммуникаций и результативности инновационных проектов с помощью анализа социальных сетей. *Cloud of Science*. 2014;1(4):665–695. URL: <https://www.elibrary.ru/thbiwr>



8. Плохов Д.В. и др. Методика оценки влияния социальных коммуникаций на результативность инновационного проекта. *Cloud of Science*. 2016;3(3). URL: https://www.researchgate.net/publication/309788251_Metodika_ocenki_vliania_socialnyh_kommunikacij_na_rezultativnost_innovacionnogo_proekta
9. Ilin I., Voronova O., Pavlov D. et al. System of project management at a medical hub as an instrument for implementation of open innovation. *Systems*. 2023;11(4):182. DOI: 10.3390/systems11040182
10. Павлов Д.А., Кирий А.В. Методика мониторинга и контроля инновационных проектов методами анализа социальных сетей. *Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета*. 2021;166:169–181. DOI: 10.21515/1990-4665-166-015
11. Павлов Д.А., Кирий А.В. Разработка комплекса проблемно-ориентированных программ для решения проблем сетевого распределения производственных задач. *Труды Кубанского государственного аграрного университета*. 2020;86:26–30. URL: <http://opac.skunb.ru/index.php?url=/notices/index/IdNotice:1396389/Source:default&ysclid=mdhdf669fe544643381#>
12. Рашка С., Мирджалили В. Python и машинное обучение: машинное и глубокое обучение с использованием Python, scikit-learn и TensorFlow 2. 3-е изд. Пер. с англ. СПб.: Дialeктика; 2020. 848 с. URL: <https://www.williamspublishing.com/Books/978-5-907203-57-0.html>
13. Павлов Д.А., Кирий А.В. Прогнозирование успешности выполнения ИТ-проектов. *Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета*. 2024;203:505–514. DOI: 10.21515/1990-4665-203-045

REFERENCES

1. Kohn M. Agile: Project assessment and planning. Transl. from Eng. Moscow: Alpina Publisher; 2022. 418 p. URL: <https://stroki.mts.ru/book/agile-otsenka-i-planirovaniye-proyektov-1759?ysclid=mdhbucngh64422956> (In Russ.).
2. Durland M., Fredericks K. An Introduction to Social Network Analysis. *New Directions for Evaluation*. 2006;2005(107):5–13. DOI: 10.1002/ev.157
3. Otte E., Rousseau R. Social network analysis: A powerful strategy, also for the information sciences. *Journal of Information Science*. 2002;28(6):441–453. DOI: 10.1177/016555150202800601
4. Krackhardt D., Hanson J.R. Informal networks: The company behind the charts. *Harvard Business Review*. 2000;78(4):104–11. URL: https://www.researchgate.net/publication/40961009_Informal_Networks_the_Company_behind_the_Charts
5. Pryke S.D. Towards a social network theory of project governance. *Construction Management and Economics*. 2005;23(9):927–939. DOI: 10.1080/01446190500184196
6. Chinowsky P., Diekmann J. and Galotti V. Social network model of construction. *Journal of Construction Engineering and Management*. 2008;134(10). DOI: 10.1061/(ASCE)0733-9364(2008)134:10(804)
7. Titov S.A. Investigation of the relationship between the structure of intra-project communications and the effectiveness of innovative projects using the analysis of social networks. *Cloud of Science*. 2014;1(4):665–695. URL: <https://www.elibrary.ru/thbiwr> (In Russ.).
8. Plokhov D.V. et al. Methodology for assessing the impact of social communications on the effectiveness of an innovation project. *Cloud of Science*. 2016;3(3). URL: https://www.researchgate.net/publication/309788251_Metodika_ocenki_vliania_socialnyh_kommunikacij_na_rezultativnost_innovacionnogo_proekta (In Russ.).
9. Ilin I., Voronova O., Pavlov D. et al. System of project management at a medical hub as an instrument for implementation of open innovation. *Systems*. 2023;11(4):182. DOI: 10.3390/systems11040182
10. Pavlov D.A., Kiri A.V. Methodology of monitoring and control of innovative projects by methods of social network analysis. *Polythematic online electronic scientific journal of the Kuban State Agrarian University*. 2021;166:169–181. (In Russ.). DOI: 10.21515/1990-4665-166-015
11. Pavlov D.A., Kiri A.V. Development of a set of problem-oriented programs for solving problems of network distribution of production tasks. *Proceedings of the Kuban State Agrarian University*. 2020;86:26–30. URL: <http://opac.skunb.ru/index.php?url=/notices/index/IdNotice:1396389/Source:default&ysclid=mdhdf669fe544643381#> (In Russ.).
12. Rashka S., Mirjalili V. Python and machine learning: machine and deep learning using Python, scikit-learn and TensorFlow 2. 3rd ed. Transl. from Eng. St. Petersburg: Dialectics; 2020. 848 p. URL: <https://www.williamspublishing.com/Books/978-5-907203-57-0.html> (In Russ.).
13. Pavlov D.A., Kiri A.V. Forecasting the success of IT projects. *Polythematic online electronic scientific journal of the Kuban State Agrarian University*. 2024;203:505–514. (In Russ.). DOI: 10.21515/1990-4665-203-045

ИНФОРМАЦИЯ ОБ АВТОРЕ / ABOUT THE AUTHOR

Дмитрий Алексеевич Павлов — кандидат физико-математических наук, научный руководитель лаборатории искусственного интеллекта и анализа данных, доцент кафедры системного анализа и обработки информации, Кубанский государственный аграрный университет им. И. Т. Трубилина, Краснодар, Российская Федерация

Dmitry A. Pavlov — Cand. Sci. (Phys.-Math.), Scientific Director of the Laboratory of Artificial Intelligence and Data Analysis, Associate Professor of the Department of System Analysis and Information Processing, I. T. Trubilin, Kuban State Agrarian University, Krasnodar, Russian Federation

<https://orcid.org/0000-0002-4677-1762>

pavlov.d@edu.kubsau.ru; dp.logic@gmail.com

Конфликт интересов: автор заявляет об отсутствии конфликта интересов.

Conflicts of interest Statement: The author has no conflicts of interest to declare.

Статья поступила 12.07.2025; принята к публикации 22.07.2025.

Автор прочитал и одобрил окончательный вариант рукописи.

The article was received on 12.07.2025; accepted for publication on 22.07.2025.

The author read and approved the final version of the manuscript.



Кибербезопасность цифрового рубля в системе деструктивных событий цифрового пространства

А.В. Иванов^а, А.В. Царегородцев^б

^{а,б} Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация;

^б Российский университет дружбы народов, Москва, Российская Федерация

АННОТАЦИЯ

В статье рассматриваются актуальные проблемы обеспечения кибербезопасности цифрового рубля в условиях растущих угроз цифрового пространства. **Актуальность исследования** определяется следующими факторами: необходимостью разработки новой концепции кибербезопасности цифрового рубля; увеличением числа компьютерных атак на кредитно-финансовую сферу; ростом деструктивного информационного воздействия на финансовые организации; потребностью в выявлении и предупреждении киберугроз; важностью защиты от хакерских атак, вирусов и мошенничества; необходимостью оптимизации процессов и повышения качества финансовых платежей. **Цель исследования** — разработка технического решения по обеспечению кибербезопасности инфраструктуры цифрового рубля. В ходе исследования решены следующие **задачи**: уточнены понятия «инфраструктура доверия» и «кибербезопасность»; проанализированы средства киберпространства в контексте защиты цифрового рубля; разработаны предложения по повышению кибербезопасности инфраструктуры доверия. **Практическая значимость** работы заключается в возможности применения предложенных решений для укрепления системы защиты цифрового рубля. **Ключевые слова**: инфраструктура доверия; система защищенности цифрового рубля; информационная безопасность; кибербезопасность; киберпространство; киберустойчивость; киберпреступление; аутентификация; API; токен

Для цитирования: Иванов А.В., Царегородцев А.В. Кибербезопасность цифрового рубля в системе деструктивных событий цифрового пространства. *Цифровые решения и технологии искусственного интеллекта*. 2025;1(3):44-54. DOI: 10.26794/3033-7097-2025-1-3-44-54

ORIGINAL PAPER

Cybersecurity of the Digital Ruble in the System of Destructive Events of the Digital Space

A.V. Ivanov^а, A.V. Tsaregorodtsev^б

^{а,б} Financial University under the Government of the Russian Federation, Moscow, Russian Federation;

^б Peoples' Friendship University of Russia (RUDN), Moscow, Russian Federation

ABSTRACT

The article discusses the current problems of ensuring the cybersecurity of the digital ruble in the context of growing threats to the digital space. The relevance of the research is determined by the following factors: the need to develop a new cybersecurity concept for the digital ruble; an increase in the number of computer attacks on the credit and financial sector; an increase in the destructive information impact on financial organizations; the need to identify and prevent cyber threats; the importance of protection against hacker attacks, viruses and fraud; the need to optimize processes and improve the quality of financial payments. The purpose of the research is to develop a technical solution to ensure the cybersecurity of the digital ruble infrastructure. During the study, the following tasks were solved: the concepts of "trust infrastructure" and "cybersecurity" were clarified; cyberspace tools were analyzed in the context of protecting the digital ruble; proposals were developed to improve the cybersecurity of the trust infrastructure. The practical significance of the work lies in the possibility of applying the proposed solutions to strengthen the digital ruble protection system. **Keywords**: trust infrastructure; digital ruble security system; information security; cybersecurity; cyberspace; cyber resilience; cybercrime; authentication; API; token

For citation: Ivanov A.V., Tsaregorodtsev A.V. Cybersecurity of the digital ruble in the system of destructive events of the digital space. *Digital Solutions and Artificial Intelligence Technologies*. 2025;1(3):44-54. DOI: 10.26794/3033-7097-2025-1-3-44-54



ВВЕДЕНИЕ

В условиях цифровой трансформации экономики информационная безопасность становится одним из приоритетных направлений развития финансовых систем. Современные методы защиты информации интегрируют как традиционные инструменты информационной безопасности, так и инновационные решения в области кибербезопасности.

Цифровизация финансовых услуг требует создания надежной системы защиты платежных операций. Банком России разработан комплекс нормативных правовых актов, регламентирующих переход к новой форме расчетов — цифровому рублю, который дополняет существующие формы денежного обращения (наличные и безналичные расчеты).

В рамках внедрения платформы цифрового рубля (ПЦР) установлены строгие требования к обеспечению конфиденциальности и целостности платежных транзакций для всех участников системы: администраторов платформы, физических и юридических лиц.

Актуальность исследования кибербезопасности системы цифрового рубля обусловлена следующими факторами:

- Формирование новой концепции кибербезопасности цифрового рубля в условиях растущего числа деструктивных событий в цифровом пространстве и увеличения количества компьютерных атак на кредитно-финансовую сферу.

- Эволюция методов разрушающего информационного воздействия, где объектами атак становятся не только финансовые данные, но и автоматизированные банковские системы и системы дистанционного банковского обслуживания (ДБО).

- Выявление и предупреждение киберугроз как ключевая задача в системе кибербезопасности финансовой инфраструктуры.

Комплексная защита цифрового пространства от деструктивных воздействий, включающая:

- противодействие хакерским атакам;
- защиту от вредоносного программного обеспечения;
- предотвращение мошеннических действий;
- разработку систем мониторинга для оптимизации процессов;
- повышение качества и безопасности финансовых платежей.

Таким образом, исследование направлено на создание эффективной системы защиты цифрового рубля, способной противостоять современным киберугрозам и обеспечивать бесперебойное функционирование платежной системы.

ОСНОВНАЯ ЧАСТЬ

Исследование процессов становления и развития цифрового суверенитета государства позволило провести комплексный анализ цифрового пространства с акцентом на выявление и изучение деструктивной информации. Подобному анализу посвящены работы В.В. Тельбуха, А.В. Десятых, С.С. Андрушкевича, Л.В. Пилипенко [1], вопросы защиты корпоративных информационных систем от комплексных деструктивных воздействий раскрываются в работе Д.К. Левонского [2]. А.Н. Рабчевский исследовал методы и алгоритмы поддержки принятия решений по противодействию деструктивным воздействиям в социальных сетях [3]. Вышеназванные авторы рассматривали принципы регулирования распространения и влияния деструктивной информации на общество, но не предложили возможные способы структуризации и классификации деструктивных событий цифрового пространства, а также проектирование модели разрушений интеллектуального агента безопасности автономии.

В настоящее время в цифровом пространстве сложилось следующее противоречие. С одной стороны, цифровизация способствует ускорению процессов сбора, обработки, передачи информации и принятия по ней соответствующих управленческих решений, с другой — наблюдается рост разрушающих деструктивных событий. Существующее противоречие обусловлено рядом проблем в сфере классификации деструктивных событий цифрового пространства:

- *неточность идентификации.* Классификаторы могут как ошибочно определять деструктивные события, так и пропускать их, что ведет к некорректным управленческим решениям;

- *дефицит данных.* Создание классификаторов требует обширного массива информации о деструктивных событиях, включая их интеграцию, типизацию и нормализацию;

- *оперативность реагирования.* Огромный объем данных в цифровом пространстве не позволяет классификаторам своевременно обнаруживать деструктивные события;

- *динамика угроз.* Постоянное увеличение и изменение характера потенциальных угроз усложняет определение ключевых классификаторов;

- *адаптивность систем.* Классификаторы не всегда способны подстраиваться под новые типы деструктивных событий, что снижает эффективность защиты информации.

Указанные обстоятельства определили актуальность и научную новизну исследования проблем,



связанных с формированием ключевых параметров классификаторов деструктивных событий в цифровом пространстве.

Основания для классификации деструктивных событий цифрового пространства представлены в табл. 1.

В данном контексте ключевыми понятиями выступают информационная безопасность и кибербезопасность.

Информационная безопасность цифровой среды. Информационная безопасность представляет собой комплекс мероприятий (правовых, экономических, финансовых, организационных и управленческих), направленных на защиту информации от несанкционированного доступа, разглашения, модификации и уничтожения.

Кибербезопасность является составной частью информационной безопасности и фокусируется на защите компьютерных систем, сетей, данных и программного обеспечения от киберугроз и киберпреступлений. Ее основные задачи включают:

- обеспечение конфиденциальности, целостности и доступности информационных ресурсов;
- предотвращение несанкционированного доступа к данным;
- защиту от кражи личной информации;
- противодействие кибершпионажу;
- борьбу с распространением вредоносных программ;
- нейтрализацию иных цифровых угроз.

Согласно исследованию А.П. Сидорова, цифровое пространство на технологическом уровне объединяет цифровую инфраструктуру и цифровые ресурсы [4, с. 49]. В научной литературе

цифровое пространство часто рассматривается как цифровая среда [5, с. 7].

В рамках данного исследования под цифровым пространством понимается интегрированная среда, включающая следующие *цифровые* компоненты:

- данные;
- информация;
- информационные системы и сети;
- информационные ресурсы;
- информационная инфраструктура;
- механизмы регулирования; организационные и управленческие инструменты (рис. 1).

В России существуют следующие классификаторы деструктивных событий в цифровом пространстве:

1. Классификатор «Список угроз информационной безопасности». Разработан ФСБ России. Это систематизированный список угроз информационной безопасности с описанием их характеристик, особенностей и требований к защите.

2. Классификатор угроз информационной безопасности (также разработка ФСБ). Определяет различные типы информационных угроз (хакерские атаки, распространение вредоносного программного обеспечения, кибершпионаж и другие).

3. Классификатор атак и инцидентов информационной безопасности (разработка ЦИБ ФСБ России) описывает различные виды атак и инцидентов в цифровом пространстве, включая различные типы атак (фишинг, DOS-атаки, перехват трафика, компрометация систем и др.).

4. Классификатор информационных технологий, применяемых при организации работ по

Таблица 1 / Table 1

Основания для классификации деструктивных событий / Grounds for Classifying Destructive Events

Событие (основание) / Events (Foundation)			
Общие основания	Информационная безопасность	Сфера деятельности	Автоматизированные системы
Характер угрозы	Вид защиты	Социальные	Конфиденциальность
Цель атаки	Уровень защиты	Правовые	Доступность
Методы атак	Область применения	Финансовые	Целостность
Последствия атак	Метод	Психологические	Автоматизируемость
Источник атаки	Уязвимость	Управленческие	Контролируемость

Источник / Source: составлено авторами / Compiled by the authors.



Рис. 1 / Fig. 1. Структурные элементы цифрового пространства / Structural elements of the digital space

Источник / Source: составлено авторами / Compiled by the authors.

обеспечению информационной безопасности (разработка ФСТЭК России), определяющий информационные технологии, используемые при обеспечении информационной безопасности.

5. Национальный классификатор угроз информационной безопасности, содержащий различные типы угроз информационной безопасности (включая методы атак, уязвимости систем, инциденты и технические меры обеспечения безопасности информации).

6. Специальные классификаторы деструктивных событий в области киберугроз.

В общем виде специальные классификаторы деструктивных событий в области киберугроз представлены в табл. 2.

Основой для разработки нормативных правовых актов в сфере защиты цифрового рубля и его инфраструктуры стала Концепция цифрового рубля, разработанная Банком России¹, в основу которой положены результаты общественных опросов и консультаций, сформулированы ключевые положения по обеспечению информационной безопасности цифрового рубля.

Существенный вклад в укрепление защиты цифрового рубля внесло принятие Федерального закона № 340-ФЗ в данной сфере², согласно которому создана специальная платформа

цифрового рубля, оператором которой выступает Банк России³.

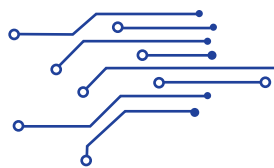
На основании Концепции о цифровом рубле Банком России разработано Положение с требованиями по обеспечению защиты информации для участников платформы цифрового рубля, в числе которых: «защита информации на платформе цифрового рубля; защита информации, необходимой для идентификации, аутентификации и авторизации пользователей платформы цифрового рубля при совершении действий в целях осуществления операций с цифровыми рублями; защита информации о предоставлении, приостановлении, возобновлении или прекращении доступа к платформе цифрового рубля; защита ключевой информации средств криптографической защиты информации (СКЗИ), используемых для обеспечения криптографической защиты операций с цифровыми рублями (криптографические ключи)»⁴. В этом же Положении определены требования в части «организации доступа пользователей к платформе цифрового рубля, а также все операции пользователя с цифровым рублем осуществляются с использованием специализированного программного модуля Банка России, интегрированного с мобильными приложениями кредитных организаций.

¹ Концепция цифрового рубля. Банк России. 2021. URL: https://cbr.ru/Content/Document/File/120075/concept_08042021.pdf

² Федеральный закон от 24.07.2023 № 340-ФЗ «О внесении изменений в отдельные законодательные акты Российской Федерации». URL: <https://base.garant.ru/407424830/?ysclid=mdircnir6589471400>

³ Положение Банка России от 03.08.2023 № 820-П (в ред. от 12.07.2024) «О платформе цифрового рубля». URL: <https://www.garant.ru/products/ipo/prime/doc/407413623/?ysclid=mdirdssmsf128544242>

⁴ Положение Банка России от 07.12.2023 № 833-П «О требованиях к обеспечению защиты информации для участников платформы цифрового рубля». URL: <https://www.garant.ru/products/ipo/prime/doc/408222733/?ysclid=mdirf09jq89319510>

**Специальные классификаторы деструктивных событий в области киберугроз /
Special Classifiers of Bestructive Events in the Field of Cyber Threats**

Тип классификатора / Classifier type	Тип воздействия / Type of impact	Методы мошенничества и последствия / Fraud methods and consequences
1. Классификаторы вирусов и вредоносных программ	Тип вредоносной программы	Метод распространения
	Уровень разрушительного воздействия на систему	Способ обхода защитных механизмов системы
2. Классификаторы кибератак	Тип кибератаки	Метод атаки
	Источник атаки	Последствия атаки для жертв
3. Классификаторы вредоносных фишинговых сайтов	Типы персональной информации, интересной для кражи с сайта	Методы привлечения жертв на вредоносный сайт
	Способы мошенничества на сайте	Возможные последствия посещения вредоносного сайта
4. Классификаторы скам- сайтов	Методы привлечения жертв на скам- сайты	Методы мошенничества на скам-сайте
	Способы защиты от скам-сайта	Последствия для посетителей скам- сайта
5. Классификаторы кибершпионажа	Методы кибершпионажа	Субъекты кибершпионажа
	Уровень организованности и сложности кибершпионажа	Последствия кибершпионажа для жертв
6. Классификаторы деструктивных событий в области кибербезопасности	Способы воздействия и распространения деструктивных событий	Методы обнаружения деструктивных событий
	Предупредительные меры и стратегии защиты от деструктивных событий	Воздействие на безопасность информационных систем
7. Классификаторы кибертерроризма	Критерии кибертеррористических действий	Основные угрозы и уязвимости в киберпространстве
	Механизмы реагирования на кибертеррористические угрозы	Анализ последствий кибертеррористических атак

Источник / Source: составлено авторами / Compiled by the authors.

Научное обоснование формирования инфраструктуры доверия цифрового рубля имеет важное практическое значение для обеспечения его информационной безопасности [6]. Исследователи выделяют несколько ключевых направлений в этой области:

- архитектурное построение инфраструктуры открытых ключей⁵;
- разработка доверенного программного обеспечения как инструмента информационной защиты (И.А. Грачков, А.А. Малюк [7]);

⁵ Концепция цифрового рубля. Банк России. 2021. URL: https://cbr.ru/Content/Document/File/120075/concept_08042021.pdf

- технологическая основа доверия для различных автоматизированных систем, предложенная Д.А. Мельниковым. Автор на основе анализа криптографических средств защиты в инфраструктуре открытых ключей рекомендует «сформировать единую технологическую базу для прикладных информационно-технологических систем» [8].

Отдельные работы посвящены исследованию новых требований по обеспечению информационной безопасности и защищенности цифрового рубля⁶.

⁶ Шашурина И. Что нужно знать о новых стандартах безопасности цифрового рубля? URL: <https://in4security.com/news/tpost/jpmmfiab51-chto-nuzhno-znat-o-novih-standartah-bezo?ysclid=m7c1y7gk5404288694>

Основные элементы инфраструктуры доверия в системе информационной безопасности цифрового рубля представлены на рис. 2.

Для обеспечения информационной безопасности и защиты инфраструктуры доверия в системе цифрового рубля предлагается создать следующие ключевые компоненты:

- узлы Банка России, включая удостоверяющий и выделенный удостоверяющий центры,

а также удостоверяющие центры кредитных организаций. Эти компоненты отвечают за регистрацию и сертификацию ключей клиентов;

- API-платформы цифрового рубля (ПЦР) — программный интерфейс, обеспечивающий подключение кредитных организаций к системе;
- API кредитных организаций (API КО) — интерфейс для взаимодействия между кредитными организациями и их клиентами.

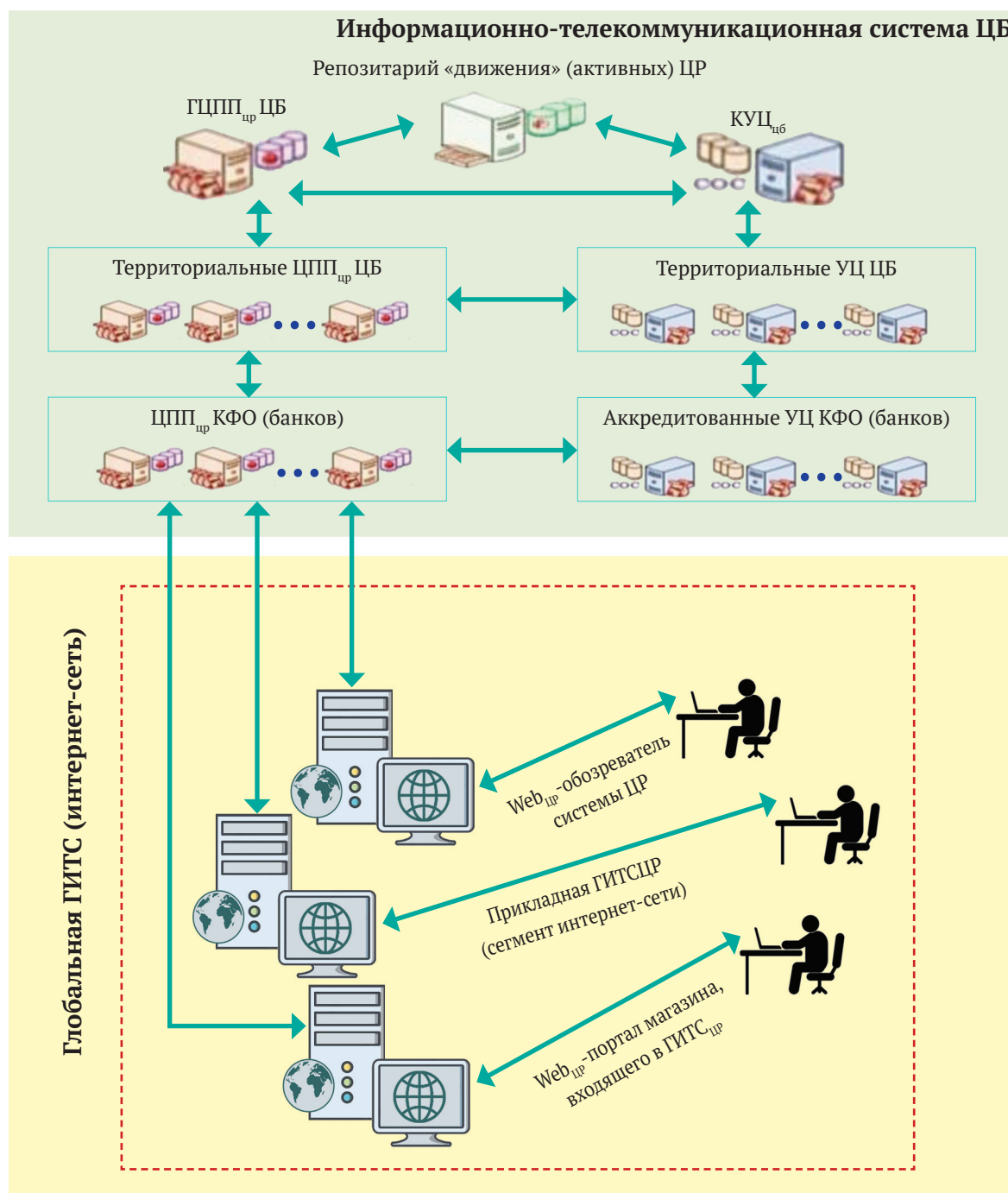


Рис. 2 / Fig. 2. Модель общей иерархической структуры инфраструктуры доверия Банка России / A model of the general hierarchical structure of the Bank of Russia's trust infrastructure

Источник / Source: составлено авторами по Д.А. Мельников [8] / Compiled by the authors from D. Melnikov [8].



Пользовательские устройства — мобильные приложения, которые кредитные организации предоставляют своим клиентам для работы с цифровым рублем⁷.

Модель инфраструктуры доверия цифрового рубля представлена на рис. 3.

Для обеспечения безопасности инфраструктуры доверия в системе цифрового рубля необходимо реализовать следующие организационные и технические меры по:

- защите персональных данных при их обработке в информационных системах с применением средств криптографической защиты⁸;
- соблюдению требований безопасности финансовых (банковских) операций и защиты информации финансовыми организациями⁹.

Проведенный анализ информационной безопасности инфраструктуры цифрового рубля показал, что в банковской сфере активно разрабатываются и внедряются инструменты и программное обеспечение для ее безопасного функционирования. Однако, по нашему мнению, в нормативных правовых актах Банка России вопросам кибербезопасности инфраструктуры цифрового рубля уделяется недостаточное внимание.

Кибербезопасность инфраструктуры — это «комплекс взаимосвязанных организационных и технических мер, направленных на защиту информации, информационных систем, информационно-телекоммуникационных сетей, автоматизированных систем управления технологическим процессом и других ИТ-систем и оборудования, использующих-

ся для обработки и хранения этой информации»¹⁰. В этой связи основной целью кибербезопасности является «обеспечение целостности, доступности и конфиденциальности данных в цифровой среде, реагирование на инциденты, связанные с утечкой данных в цифровом формате»¹¹. В этом случае обеспечивается состояние защищенности организации, его устойчивое функционирование, а также киберустойчивость к компьютерным атакам¹².

В соответствии с недостаточной интерпретацией самого понятия «кибербезопасность» следует провести его теоретическую интерпретацию. Например, Ю.К. Язов понятие «кибербезопасность» рассматривает как «состояние защищенности цифровой информации и устройств ее обработки от киберугроз» [9, с. 4]. С понятием «кибер» связаны и другие производные: «киберугроза» — угроза безопасности цифровой информации; «киберустойчивость» — способность цифровых устройств противостоять киберугрозам; «киберпреступление» — преступление, связанное с нарушением безопасности цифровой информации, путем противоправных действий в киберпространстве; «кибермошенничество» — причинение материального, финансового или иного ущерба путем хищения с использованием личной информации граждан (например, номеров банковских счетов, паспортных данных, кодов, паролей и т.п.) [9].

Общая оценка уровня доверия к субъекту информационного обмена состоит из «оценки ряда внутренних процессов и процедур информационной безопасности субъекта информационного обмена» [10, с. 42].

Киберпространство. Для оценки уровня кибербезопасности инфраструктуры доверия в системе цифрового рубля требуется провести анализ ее киберпространства. С функциональной точки зрения киберпространство включает следующие элементы (средства): каналообразования; распределения ресурсов (маршрутизаторы, коммутаторы и другие устройства); измерения элементов киберпространства и обрабатываемого трафика; разграничения и защиты информационных ресурсов (системы аутентификации и идентификации); управления киберпространством. Особое значение для обеспечения кибербезопасности инфраструктуры

⁷ ГОСТ Р 34.12–2015. Информационная технология. Криптографическая защита информации. Блочные шифры: национальный стандарт Российской Федерации: издание официальное. Федеральное агентство по техническому регулированию и метрологии. М.: Стандартинформ; 2015. URL: https://rosstandarts.ru/file/gost/35/040/gost_r_34.12-2015.pdf?ysclid=md7ivk2ocb394038537

⁸ 15. Приказ ФСБ России от 10.07.2014 № 378 «Об утверждении Состав и содержания организационных и технических мер по обеспечению безопасности персональных данных при их обработке в информационных системах персональных данных с использованием средств криптографической защиты информации, необходимых для выполнения установленных Правительством Российской Федерации требований к защите персональных данных для каждого из уровней защищенности». URL: <https://base.garant.ru/70727118/?ysclid=md7j1brq20107377435>

⁹ ГОСТ Р 57580.1–2017. Национальный стандарт Российской Федерации. «Безопасность финансовых (банковских) операций. Защита информации финансовых организаций. Базовый состав организационных и технических мер». Приказ Росстандарт от 08.08.2017 № 822-ст. URL: <https://base.garant.ru/71742160/?ysclid=md7j57cfar553883364>

¹⁰ Все о кибербезопасности: определение, принцип работы, отличия. URL: <https://ddos-guard.ru/blog/kiberbezopasnost>

¹¹ Что такое кибербезопасность. URL: <https://ddos-guard.ru/blog/kiberbezopasnost>

¹² Кибербезопасность: от понятия до практики. IBS. URL: <https://ibs.ru/media/kiberbezopasnost-ot-ponyatiya-do-praktiki/?ysclid=m9mcgasm1386172313>

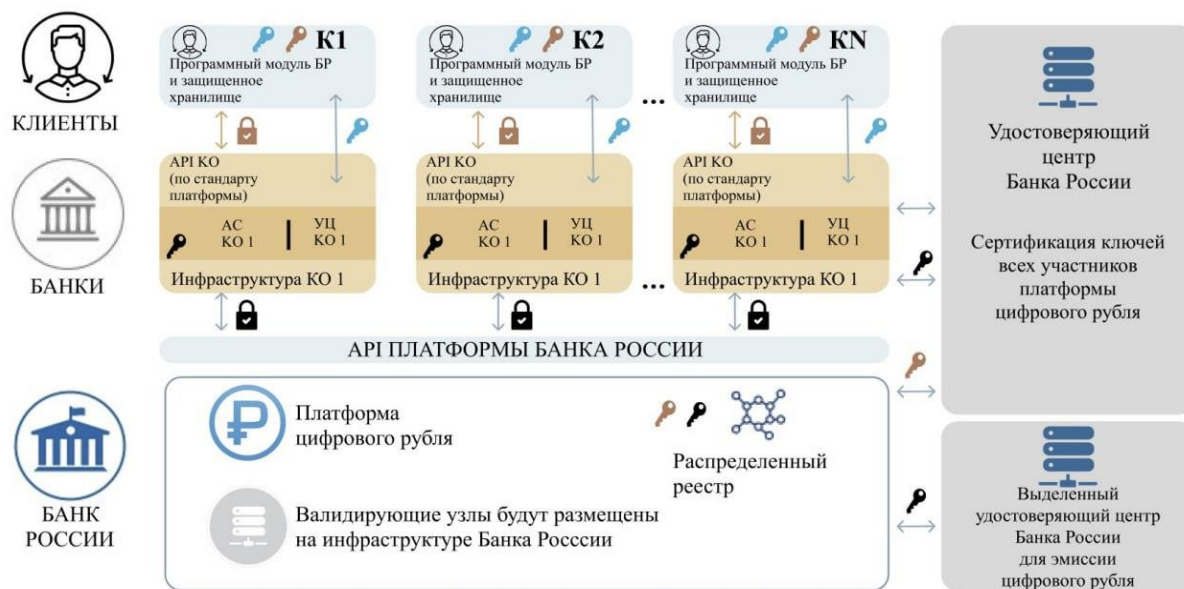


Рис. 3 / Fig. 3. Модель инфраструктуры цифрового рубля / Digital Ruble Infrastructure Model

Источник / Source: Банк России. Концепция цифрового рубля / Bank of Russia. Digital Ruble Concept.

доверия в системе цифрового рубля имеют средства управления киберпространством. К ним можно отнести: «аппаратные средства, которые обеспечивают стабильную работу сети Интернет (дата-центры, сервера, маршрутизаторы); программное обеспечение киберпространства; средства защиты информации (межсетевые экраны, антивирусы)» [11].

Для кибербезопасности инфраструктуры доверия в системе цифрового рубля применяются **криптографические методы защиты**. Например, при совершении онлайн-платежей данные передаются через контролируемые каналы с использованием протоколов TLS, которые позволяют обеспечить надежную и эффективную защиту информации. **Аутентификация и авторизация** в современных системах реализуются через механизм токенов доступа, обеспечивающих управление сессиями и безопасный доступ к ресурсам. Токены позволяют подтверждать личность пользователя без необходимости повторной авторизации при каждом запросе. Существует несколько типов токенов доступа, среди которых наиболее распространены:

1) OAuth-токены для делегированной авторизации. Позволяют одному приложению получать ограниченный доступ к ресурсам пользователя на сервере другого приложения. Эта технология применяется при интеграции сторонних сервисов, которые могут выполнять операции от имени пользователя без доступа к его учетным данным. Процесс начинается с аутентификации пользователя на сервере авторизации, после чего выдает-

ся токен доступа для работы с API. В российской финансовой сфере используется отечественный стандарт ГОСТ OAuth 2.0, разработанный для защиты персональных данных при авторизации;

2) JWT (JSON Web Tokens) — компактные токены, где полезная нагрузка (payload) кодируется в формате JSON. Каждый токен содержит информацию о пользователе, его правах доступа и сроке действия. Подпись токена выполняется с помощью асимметричных или симметричных ключей, что гарантирует его подлинность и целостность.

Управление сессиями с использованием токенов доступа включает несколько ключевых аспектов. Во-первых, необходимо контролировать время жизни токенов. Если не ограничивать срок их действия, то возрастают риски компрометации. В то же время для выпуска нового токена необходимо повторное подтверждение пользователя, что повышает уровень безопасности. Во-вторых, важным аспектом является безопасное хранение токенов на клиентской стороне, так как злоумышленник не должен иметь возможности добраться до токенов, даже если получил доступ к устройству пользователя. Для этого необходимо использовать безопасные каналы связи (HTTPS, TLS 1.3) для передачи токенов, чтобы исключить возможность их перехвата злоумышленниками. В-третьих, если злоумышленник получил токен, необходимо иметь механизм его немедленного аннулирования. Для этого существуют черные списки (blacklists), при каждом запросе сервер проверяет, не входит ли полученный токен в список отозванных. Однако у этого метода есть серьезный



недостаток — каждый новый запрос увеличивает риск компрометации передачи учетных данных.

Сегодня разработаны и применяются альтернативные и более безопасные **технологии API**, широко применяемые в автоматизированных банковских системах (АБС) для управления счетами, кредитами и платежами и системе дистанционного банковского обслуживания (ДБО) для взаимодействия с банковскими сервисами через интернет и мобильные приложения.

API-аутентификация осуществляется через уникальные идентификаторы — API-ключи, выдаваемые клиентам для доступа к ресурса; повышенный уровень защиты обеспечивает многофакторная аутентификация (MFA), требующая от пользователя подтверждения личности несколькими способами (паролем, одноразовым кодом и пр.); OAuth 2.0 реализует модель делегированной авторизации, включающую три основных компонента: хранение защищенных данных (ресурсный сервер), выдачу токенов доступа после успешной аутентификации (сервер авторизации), запросы доступа к данным (клиент).

В целом, безопасность API обеспечивается четырьмя ключевыми компонентами: пользователь, клиент (запрос доступа), сервер авторизации (выдача токена), ресурсный сервер (предоставление доступа при наличии токена). Эти элементы формируют многоуровневую систему защиты, где каждая роль четко определена. Однако эффективность защиты зависит не только от взаимодействия компонентов, но и от используемых инструментов. Шифрование и токенизация остаются основой защиты API, при этом их результативность определяется как качеством алгоритмов, так и грамотностью их реализации.

ВЫВОДЫ

Проведенный анализ системы кибербезопасности инфраструктуры цифрового рубля позволяет сформулировать следующие ключевые выводы.

Комплексный подход к обеспечению кибербезопасности является основополагающим принципом защиты цифрового рубля. Система безопасности охватывает все уровни инфраструктуры, от анализа киберпространства до защиты конечных пользователей.

Многоуровневая система защиты реализована через последовательное применение различных механизмов безопасности: постоянный мониторинг и анализ киберугроз; защищенные каналы взаимодействия; специализированное программное обеспечение; сертифицированные средства криптографической защиты.

Централизованный контроль обеспечивается через использование специализированного программного модуля Банка России, который управляет доступом пользователей и контролирует все операции с цифровыми рублями.

Надежная аутентификация участников системы достигается за счет: двухфакторной аутентификации; использования сертифицированных ключей; защищенных каналов связи.

Защита данных на платформе цифрового рубля гарантируется применением современных СКЗИ, обеспечивающих: целостность данных; достоверность транзакций; надежное подписание операций.

Таким образом, система кибербезопасности цифрового рубля построена на основе современных принципов информационной безопасности и обеспечивает надежную защиту всех компонентов инфраструктуры при проведении операций с цифровыми рублями.

БЛАГОДАРНОСТИ

Статья подготовлена по результатам исследований, выполненных за счет бюджетных средств по государственному заданию Финансовому университету при Правительстве Российской Федерации. Москва, Российская Федерация.

ACKNOWLEDGMENTS

The article is based on the results of research carried out at the expense of budgetary funds under the state assignment of the Financial University under the Government of the Russian Federation. Moscow, Russian Federation.

СПИСОК ИСТОЧНИКОВ

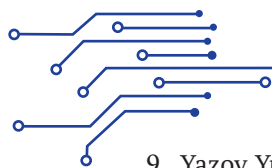
1. Тельбух В. В., Десятых А. В., Андрушкевич С. С., Пилипенко Л. В. Метод выявления деструктивного контента в информационных интернет-ресурсах. *Известия ТулГУ. Технические науки*. 2023;3:423–428. DOI: 10.24412/2071–6168–2023–3–423–429
2. Левоневский Д. К. Методы и модели защиты корпоративных информационных систем от комплексных деструктивных воздействий. Автореф. дисс. ... канд. техн. наук. СПб.: Санкт-Петербургский институт информатики и автоматизации РАН; 2020. 144 с. URL: <https://www.elibrary.ru/swbxfdf>



3. Рабчевский А.Н. Методы и алгоритмы поддержки принятия решений по противодействию деструктивным воздействиям в социальных сетях на основе многофакторного анализа ролей пользователей. Дисс. ... канд. техн. наук. Пермь: Пермский государственный национальный исследовательский университет. 2022. 138 с. URL: https://pstu.ru/files/2/file/adm/dissertacii/rybachevskiy/Dissertaciya_RabchevskiyiAN.pdf
4. Сидорова А.П. Понятие цифрового пространства и его характеристики. Возможности и угрозы использования цифрового пространства. Научный диалог: Молодой ученый. Сб. научных трудов по материалам XXVIII международной научной конференции. СПб.: 2020:48–55. DOI: 10.18411/spc-22-05-2020-11
5. Стельмашонок Е.В., Васильева И.Н., ред. Информационная безопасность цифрового пространства. СПб.: Изд-во СПбГЭУ; 2019. 155 с. URL: https://infosec.spb.ru/wp-content/uploads/2019/02/elibrary_36956934_98598916.pdf?ysclid=md8g277xn6103946799
6. Королев В.И. Архитектурное построение инфраструктуры открытых ключей интегрированного информационного пространства. *Безопасность информационных технологий*. 2015;22(3):59–71. URL: <https://elibrary.ru/vhhxbz>
7. Грачков И.А., Малюк А.А. Проблемы разработки доверенного программного обеспечения, применяемого на объектах критической информационной инфраструктуры (организационные и методические аспекты). *Безопасность информационных технологий*. 2019;26(1):56–63. URL: <https://elibrary.ru/yzellv>
8. Мельников Д.А. и др. Рекомендации по созданию инфраструктуры доверия системы цифрового рубля. *Безопасность информационных технологий*. 2024;31(3):43–63. DOI: 10.26583/bit.2024.3.01
9. Язов Ю.К. Об определении понятия «кибербезопасность» и связанных с ним терминов. *Вопросы кибербезопасности*. 2025;1(65):2–6. URL: <https://cyberrus.info/wp-content/uploads/2025/03/vokib-2025-1-st01-s002-006.pdf> DOI: 10.21681/2311-3456-2025-1-2-6
10. Иванов А.В., Огнев И.А. Проблемы оценки доверия к процессам аудита информационной безопасности. *Вопросы кибербезопасности*. 2024;3(61):40–50. URL: <https://cyberrus.info/wp-content/uploads/2024/05/vokib-2024-3-st05-s040-050.pdf> DOI: 10.21681/2311-3456-2024-3-40-50
11. Стародубцев Ю.И., Закалкин П.В., Иванов С.А. Структурно-функциональная модель киберпространства. *Вопросы кибербезопасности*. 2021;4(44):16–24. URL: https://cyberrus.info/wp-content/uploads/2021/07/16-24-444-21_2-Starodubtsev.pdf DOI: 10.21681/2311-3456-2021-4-16-24

REFERENCES

1. Telbukh V.V., Desyakh A.V., Andrushkevich S.S., Pilipenko L.V. Method of identifying destructive content in Internet information resources. *News of TulsU. Technical sciences*. 2023;3:423–428 (In Russ.). DOI: 10.24412/2071-6168-2023-3-423-429
2. Levonevsky D.K. Methods and models of protection of corporate information systems from complex destructive influences. The abstract Diss. Cand. Sci. (Tech.). St. Petersburg: Saint Petersburg Institute of Informatics and Automation of the Russian Academy of Sciences; 2020. 144 p. URL: <https://www.elibrary.ru/swbxfd> (In Russ.).
3. Rabchevsky A.N. Methods and algorithms for decision support to counteract destructive influences in social networks based on multifactorial analysis of user roles. Diss. Cand. Sci. (Tech.). Perm: Perm State National Research University; 2022. 138 p. URL: https://pstu.ru/files/2/file/adm/dissertacii/rybachevskiy/Dissertaciya_RabchevskiyiAN.pdf (In Russ.).
4. Sidorova A.P. The concept of digital space and its characteristics. Opportunities and threats of using digital space. Scientific dialogue: A young scientist. Collection of scientific papers based on the materials of the XXVIII International Scientific conference. St. Petersburg; 2020:48–55 (In Russ.). DOI: 10.18411/spc-22-05-2020-11
5. Stelmashonok E.V., Vasilyeva I.N., eds. Information security of the digital space. St. Petersburg: SPbGEU Publishing House; 2019. 155 p. URL: https://infosec.spb.ru/wp-content/uploads/2019/02/elibrary_36956934_98598916.pdf?ysclid=md8g277xn6103946799
6. Korolev V.I. Architectural construction of the public key infrastructure of the integrated information space. *Information technology security*. 2015;22(3):59–71. URL: <https://elibrary.ru/vhhxbz> (In Russ.).
7. Grachkov I.A., Malyuk A.A. Problems of developing trusted software used at critical information infrastructure facilities (organizational and methodological aspects). *Information technology security*. 2019;26(1):56–63. URL: <https://elibrary.ru/yzellv> (In Russ.).
8. Melnikov D.A. and others. Recommendations for creating a trust infrastructure for the digital ruble system. *Information technology security*. 2024;31(3):43–63 (In Russ.). DOI: 10.26583/bit.2024.3.01



9. Yazov Yu.K. On the definition of “cybersecurity” and related terms. *Cybersecurity issues*. 2025;1(65):2–6. URL: <https://cyberrus.info/wp-content/uploads/2025/03/vokib-2025-1-st01-s002-006.pdf> (In Russ.). DOI: 10.21681/2311-3456-2025-1-2-6
10. Ivanov A.V., Ognev I.A. Problems of assessing trust in information security audit processes. *Cybersecurity issues*. 2024;3(61):40–50. URL: <https://cyberrus.info/wp-content/uploads/2024/05/vokib-2024-3-st05-s040-050.pdf> (In Russ.). DOI: 10.21681/2311-3456-2024-3-40-50
11. Starodubtsev Yu.I., Zakalkin P.V., Ivanov S.A. Structural and functional model of cyberspace. *Cybersecurity issues*. 2021;4(44):16–24. URL: https://cyberrus.info/wp-content/uploads/2021/07/16-24-444-21_2.-Starodubtsev.pdf (In Russ.). DOI: 10.21681/2311-3456-2021-4-16-24

ИНФОРМАЦИЯ ОБ АВТОРАХ / ABOUT THE AUTHORS

Анатолий Викторович Иванов — доктор социологических наук, профессор кафедры информационной безопасности факультета информационных технологий и анализа больших данных, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация; главный научный сотрудник, Институт цифровых технологий РАН, Москва, Российская Федерация

Anatoly V. Ivanov — Dr. Sci. (Sociol.), Prof. Department of Information Security, Faculty of Information Technology and Big Data Analysis, Financial University under the Government of the Russian Federation, Moscow, Russian Federation; Chief Researcher, Institute of Digital Technologies, Russian Academy of Sciences, Moscow, Russian Federation

<https://orcid.org/0000-0002-0316-1518>

Автор для корреспонденции / Corresponding author:

aivanov@fa.ru

Анатолий Валерьевич Царегородцев — доктор технических наук, профессор, главный научный сотрудник, Институт цифровых технологий РАН, Москва, Российская Федерация; профессор кафедры прикладной информатики и интеллектуальных систем в гуманитарной сфере, Российский университет дружбы народов (РУДН), Москва, Российская Федерация

Anatoly V. Tsaregorodtsev — Dr. Sci. (Tech.), Prof, Chief Researcher, Institute of Digital Technologies of the Russian Academy of Sciences, Moscow, Russian Federation; Professor of the Department of Applied Informatics and Intelligent Systems in the Humanities, Peoples' Friendship University of Russia (RUDN), Moscow, Russian Federation

<https://orcid.org/0000-0002-8447-3352>

anvtsaregorodtsev@fa.ru

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Conflicts of Interest Statement: The authors have no conflicts of interest to declare.

Статья поступила 09.07.2025; принята к публикации 10.08.2025.

Авторы прочитали и одобрили окончательный вариант рукописи.

The article was received 09.07.2025; accepted for publication 10.08.2025.

The authors read and approved the final version of the manuscript.

Недекларируемые возможности файловой архитектуры: графические контейнеры

А.А. Рыженко^а, С.И. Козьминых^б^{а,б} Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация;^б Российский экономический университет им. Г.В. Плеханова, Москва, Российская Федерация

АННОТАЦИЯ

Начало двухтысячных годов показало истинное лицо производителей гигантов цифровой индустрии. Существенные изменения протокольной системы, файловой системы, системы передачи данных и т.д. внесли в общую концепцию информационной безопасности принципиально новое направление — обеспечение конфиденциальности данных изнутри. Данное направление вполне актуально и в настоящее время. Лавинно нарастающее количество недекларируемых функций вызывает объективное опасение наступления различных цифровых рисков во всех без исключения государствах, даже в том, где располагаются основные офисы GAFAM. В данной статье не предполагается рассмотрение обхода запрета с использованием VPN-каналов (официально разрешенных ФСБ России с использованием лицензированных ресурсов популярных провайдеров), но будет представлен алгоритм, описывающий одну из причин, побудивших государство ввести запрет на использование глобальных обменников графической информации. Примеров таких алгоритмов достаточно много, и, как следствие, решение, принятое своевременно РКН, со точки зрения защиты цифрового суверенитета нашего государства обоснованно. В доказательство предлагается ряд статей, посвященных скрытым функциям информационных систем и технологий. Начальный этап будет посвящен файловой системе как для стационарных, так и для мобильных систем. Старт с самых простых возможностей, пропускаемых даже современными системами защиты данных. Статьи несут исключительно образовательные и предупредительные цели для действующих специалистов и не предполагают инструкции к исполнению для нанесения вреда организациям.

Ключевые слова: простая инъекция; файловая система; графический формат; ложный JPEG; безопасность; проникновение; глобальная сеть

Для цитирования: Рыженко А.А., Козьминых С.И. Недекларируемые возможности файловой архитектуры: графические контейнеры. *Цифровые решения и технологии искусственного интеллекта*. 2025;1(3):55-61. DOI: 10.26794/3033-7097-2025-1-3-55-61

ORIGINAL PAPER

Undeclared File Architecture Features: Graphical Containers

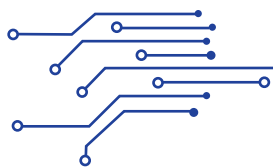
A.A. Ryzhenko^a, S.I. Kozminykh^b^{a,b} Financial University under the Government of the Russian Federation, Moscow, Russian Federation;^b Plekhanov Russian University of Economics, Moscow, Russian Federation

ABSTRACT

The beginning of the two thousand years showed the true face of the digital industry vendors. Significant changes in the protocol system, file system, data transfer system, etc. introduced a fundamentally new direction into the general concept of information security — ensuring data confidentiality from inside. This direction is quite relevant at present time. The avalanche-increasing number of undeclared functions raises objective concerns about the occurrence of various digital risks in all countries, without exception, even in the one where the main offices of GAFAM are located. This article does not intend to consider circumventing the ban using VPN channels (officially authorized by the FSB of Russia using licensed resources from popular providers), but will present an algorithm describing one of the reasons that prompted the government to ban the use of global graphic information exchangers. There are many examples of such algorithms, and as a result, the decision taken by the RCN is justified from the point of view of protecting the digital sovereignty of our state. As proof, a number of articles are proposed on the hidden functions of information systems and technologies. The initial stage will be devoted to the file system for both stationary and mobile systems. Start with the simplest features that are overlooked even by modern data protection systems. The articles are intended Only for educational and preventive purposes for specialists and do not include instructions for harming organizations.

Keywords: injections; file system; graphic format; false JPEG; security; penetration; global network

For citation: Ryzhenko A.A., Kozminykh S.I. Undeclared file architecture features: Graphical containers. *Digital Solutions and Artificial Intelligence Technologies*. 2025;1(3):55-61. DOI: 10.26794/3033-7097-2025-1-3-55-61



ВВЕДЕНИЕ

Классическая файловая архитектура с 2000-х гг. считается морально устаревшей. Большая часть самых популярных форматов файлов перешла из стадии закрытой архитектуры в класс контейнеров открытого типа. Самые популярные формы: полностью открытые и открытый второй (описательный) блок данных используются многими специалистами для передачи данных в скрытой форме через глобальную сеть. Современные средства защиты информации до сих пор не научились обнаруживать даже простые инъекции в файловую инфраструктуру, чем пользуются злоумышленники, заменяя неиспользуемые и избыточные биты изображения-носителя секретными данными (вирусом). Это позволяет обходить меры безопасности, направленные на другие типы данных. В данной статье не предполагается рассмотрение обхода запрета с использованием VPN-каналов (официально разрешенных ФСБ России с использованием лицензированных ресурсов популярных провайдеров), но будет представлен алгоритм заражения JPEG-файла¹ — простая инъекция без знаний программирования [1–6]. Сами по себе форматы JPEG не являются исполняемыми, но вредоносный код может быть вложен в графический файл, чем пользуются злоумышленники для поиска уязвимости в программном обеспечении для открытия и обработки изображений.

Решение Роскомнадзора (РКН) просто отрезать Инстаграм², суть которого — обмен изображениями в разных форматах, от доменной зоны RU, было неожиданным, но своевременным и обоснованным.

АЛГОРИТМ ИНЪЕКЦИИ В ПОПУЛЯРНЫЙ ГРАФИЧЕСКИЙ ФОРМАТ

Вредоносный код скрывается в файле, который выглядит как графический, но в действительности имеет двойное расширение и представляет собой исполняемый код. Рассылка и тиражирование «безобидного» JPEG-файла через соцсети и мессенджеры в считанные часы может поставить под угрозу информационную безопасность не только отдельно взятой организации, но и целого государства.

Структура алгоритма

• **Шаг 1.** Берется произвольная картинка в формате .jpg: .jpg, .jpe, .jpeg и т.д. и переименовывается в jpg.jpg (рис. 1). Несмотря на то что

¹ URL: <https://athomecomputer.co.uk/how-to-hide-a-file-or-folder-inside-an-image/>

² Социальная сеть, цифровая посредническая платформа — деятельность запрещена на территории Российской Федерации.

имеется разница в названиях расширений файлов, после перехода в режим двух блоков разница не существенна. Как файл не будет назван, любой графический редактор сможет прочитать файл и раскодировать изображение. Как следствие, любое изменение в названии расширения на дальнейший алгоритм влиять не будет.

Для последующего шага необходимо создать или использовать готовый файл, который будет внедряться в графический. Можно использовать файл произвольного формата и размера. Но интерес именно в том, чтобы спрятать информацию.

• **Шаг 2.** Создание нового текстового файла в формате.txt (рис. 2).

Здесь необходимо использовать ту самую недекларируемую функцию. Файл для инъекции нужно заархивировать в один из популярных форматов (zip либо rar). Данные архивы хранят внутри собственных алгоритмов сжатия информации блока информационного описания, зашифрованного одним из известных алгоритмов. Именно эта недекларированная возможность позволяет использовать второй блок графического формата как контейнер.

• **Шаг 3.** Создание нового архивного файла в формате.zip для текстового файла в формате.txt (рис. 3).

Следом необходимо сделать инъекцию. Как уже было объявлено ранее, языки программирования не используются. Применяется команда консольного режима COPY с дополнительным ключом /b. Можно перейти в консоль через вызов CMD или в той же папке создать bat-файл, прописав команду внутри файла. Удобство заключается в множественном применении. Итак, создается txt-файл, добавляется команда COPY с ключом и атрибутами. Txt-файл переименовывается в bat-файл на шагах 4 и 5.

• **Шаг 4.** Создание нового текстового файла Code.txt (рис. 4).

• **Шаг 5.** Переименование текстового файла Code.txt в Code.bat. На рис. 5 отражено открытие (запуск) файла.

В результате выполнения команды автоматически создается файл result.jpg. Если запустить/открыть файл в произвольном графическом редакторе, то он откроется без изменений и предупреждений о возможном нарушении целостности.

Далее необходимо попробовать из полученного нового графического файла получить текстовый файл. Здесь необходимо уточнить, что дополнительных специальных программных инструментов не потребуется. Стандартный 7Z-архиватор, встроенный в операционную систему, вполне спра-



Рис. 1 / Fig. 1. Исходный графический файл для инъекции / Source Graphic File for Injection
Источник / Source: открытая зона интернет / Open internet area.

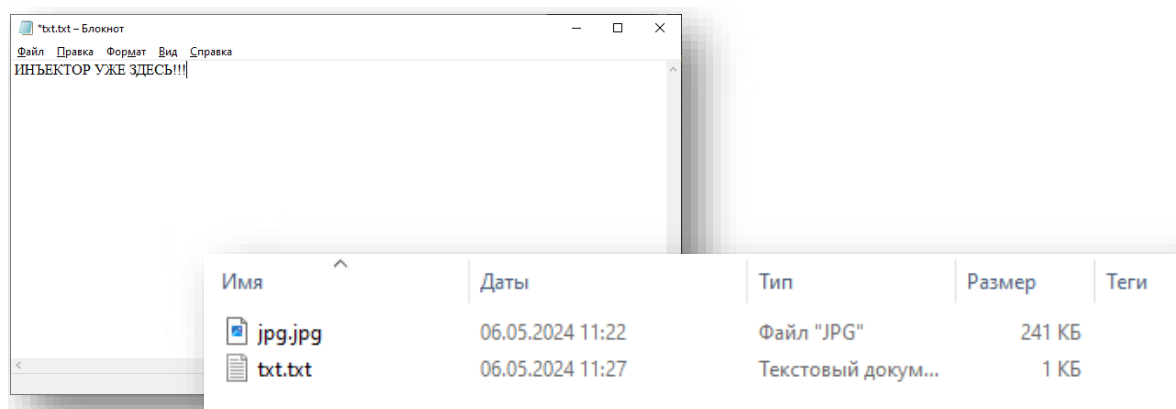


Рис. 2 / Fig. 2. Подготовка файла для инъекции / Preparing the File for Injection
Источник / Source: составлено авторами / Compiled by the authors.

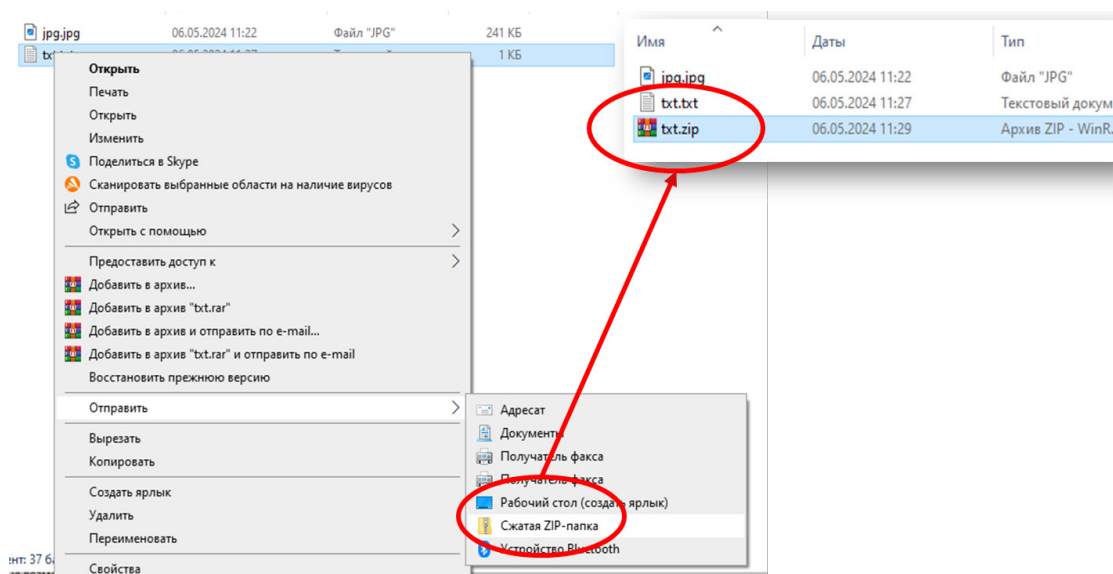


Рис. 3 / Fig. 3. Создание простого архива / Creating a Simple Archive
Источник / Source: составлено авторами / Compiled by the authors.

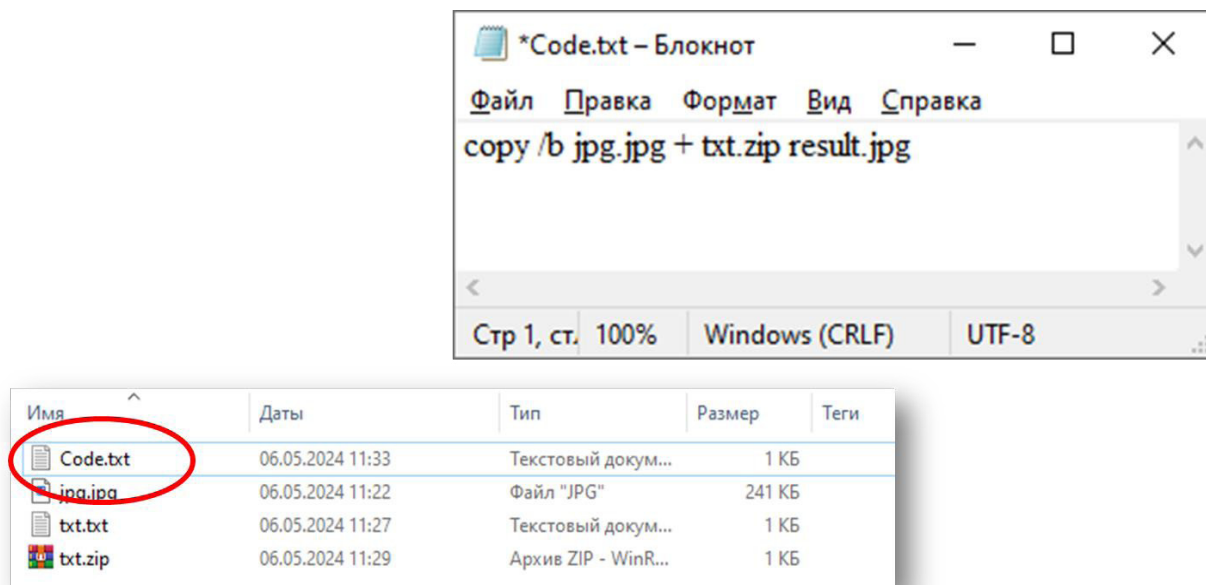


Рис. 4 / Fig. 4. Файл с ключевой командой / The File with the Key Command

Источник / Source: составлено авторами / Compiled by the authors.

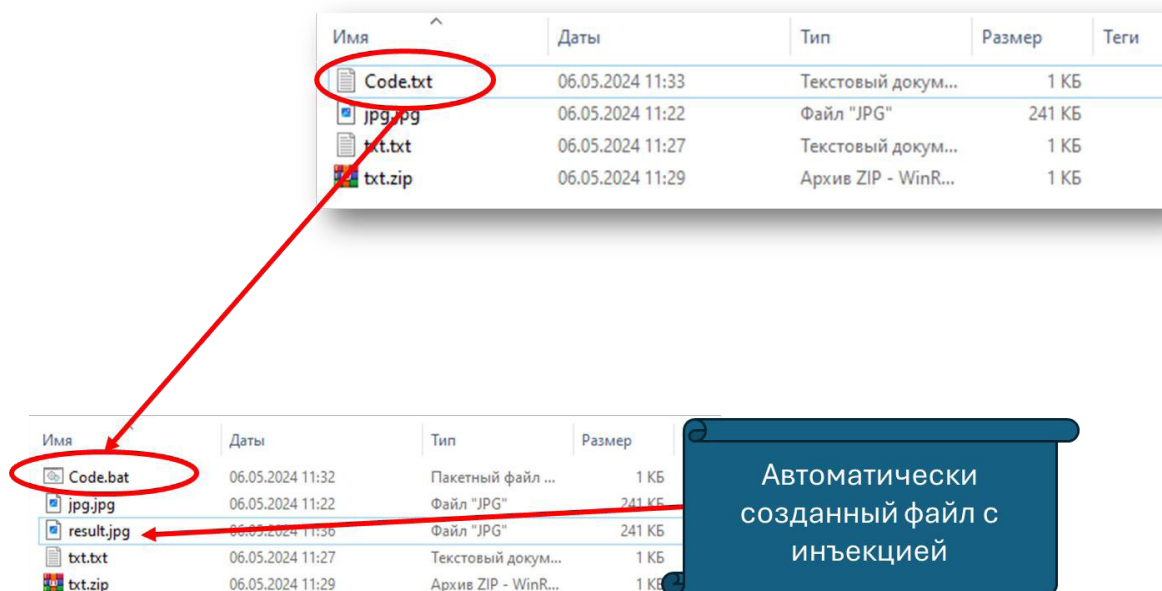


Рис. 5 / Fig. 5. Результат работы команды bat файла / The Result of the Bat File Command

Источник / Source: составлено авторами / Compiled by the authors.

вится с данной задачей. Следующий шаг — переименование файла result.jpg в result.7z. Система не выдаст ошибки и переименует графический файл в файл архива и даже откроет как простой архив — шаги 6–7.

- **Шаг 6.** Проверка контейнера.jpg. Переименовывать файл result.jpg в result.7z или result.rar (если предустановлен) (рис. 6).
- **Шаг 7.** Открыть архив result.7z или result.rar как обычный архив (рис. 7).

Скрытый файл находится внутри графического контейнера. Здесь необходимо проверить целостность графического файла, переименовать обратно архив 7Z в JPG и запустить файл как обычную картинку.

Целостность не нарушена. Проверка пройдена.

- **Шаг 8.** Проверить целостность графического файла. Переименовать обратно в архив result.7z или result.rar в result.jpg. Запустить (рис. 8).



Имя	Даты	Тип	Размер	Теги
Code.bat	06.05.2024 11:32	Пакетный файл ...	1 КБ	
jpg.jpg	06.05.2024 11:22	Файл "JPG"	241 КБ	
result.jpg	06.05.2024 11:36	Файл "JPG"	241 КБ	
txt.txt	06.05.2024 11:27	Текстовый докум...	1 КБ	
txt.zip	06.05.2024 11:29	Архив ZIP - WinR...	1 КБ	

Имя	Даты	Тип	Размер	Теги
Code.bat	06.05.2024 11:32	Пакетный файл ...	1 КБ	
jpg.jpg	06.05.2024 11:22	Файл "JPG"	241 КБ	
result.rar	06.05.2024 11:36	Архив WinRAR	241 КБ	
txt.txt	06.05.2024 11:27	Текстовый докум...	1 КБ	
txt.zip	06.05.2024 11:29	Архив ZIP - WinR...	1 КБ	

Рис. 6 / Fig. 6. Графический формат как архив / Graphic Format as an Archive

Источник / Source: составлено авторами / Compiled by the authors.

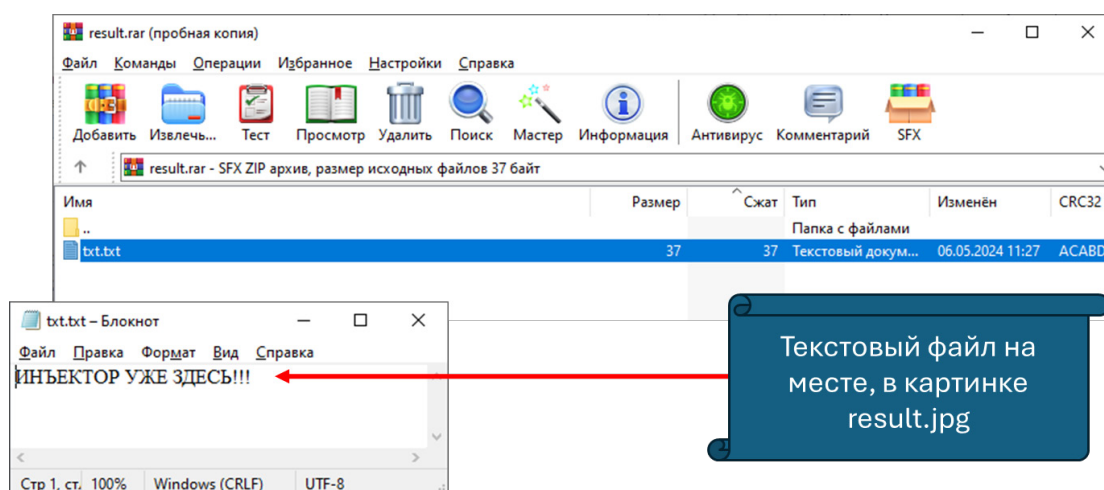


Рис. 7 / Fig. 7. Вскрытый графический формат (второй блок) / The Opened Graphic Format (Second Block)

Источник / Source: составлено авторами / Compiled by the authors.

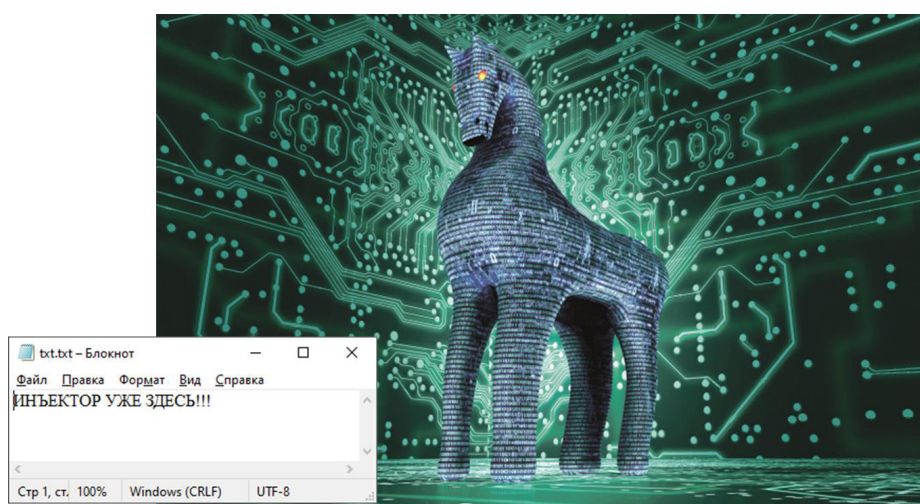


Рис. 8 / Fig. 8. Зараженный графический формат / Infected Graphic Format

Источник / Source: составлено авторами / Compiled by the authors.



Здесь следует обратить внимание на неоспоримый факт, что все манипуляции с целостностью файла не вызывали подозрения ни у одной установленной системы защиты информации.

ВЫВОДЫ

На пошаговом алгоритме наглядно показано, что даже простая инъекция без знаний программирования заставляет задуматься специалистов в области информационной безопасности о степени доверия как к современным системам защиты

информации, так и к самим операционным системам и файловым системам в синтезе. Существует ли панацея или хотя бы методы защиты от простых инъекций? Однозначно пока можно ответить, что нет!

В последующих статьях аналогичным образом планируется представить уязвимость целостности файлов автоматизированного офиса, универсальных форматов файлов, медиафайлов, а также взглянуть на ADS (Advertisement) с иной точки зрения.

СПИСОК ИСТОЧНИКОВ

1. Шумская О. О., Евсютин О. О. Встраивание информации в сжатые JPEG-изображения с минимизацией проявления демаскирующих признаков. GraphiCon 2018: труды 28-й Международной конференции по компьютерной графике и машинному зрению. Томск: Национальный исследовательский Томский политехнический университет; 2018:185–189. URL: <https://elibrary.ru/zamymx>
2. Мельман А. С., Евсютин О. О. Эффективное и безошибочное сокрытие информации в гибридном домене цифровых изображений с использованием метаэвристической оптимизации. *Компьютерные исследования и моделирование*. 2023;15(1):197–210. DOI: 10.20537/2076-7633-2023-15-1-197-210
3. Радаев С. В., Орлов Д. В., Басов О. О. Комбинированный стенографический алгоритм выстраивания конфиденциальной информации в цифровые изображения формата JPEG. *Научные ведомости Белгородского государственного университета. Серия: Экономика. Информатика*. 2017;23(272):185–192. URL: <https://elibrary.ru/yllsbp>
4. Петраков А. В., Малиничев Д. М., Бугаев В. С., Романцов А. П., Федяев Л. С. Реализация методов компьютерной аудиовидеостеганографии для скрытия данных в графических файлах формата jpeg (JFIF). *Электротехнические и информационные комплексы и системы*. 2009;4(5):23–27. URL: <https://elibrary.ru/kwvtml>
5. Шумская О. О., Железны М. Адаптивный алгоритм встраивания информации в сжатые JPEG-изображения на основе операции замены. *Информационно-управляющие системы*. 2018;5:44–56. DOI: 10.31799/1684-8853-2018-5-44-56
6. Секретов М. В. Алгоритм внедрения контейнеров в изображения формата JPEG. Прикладные цифровые технологии и системы XXI века: экономика, менеджмент, управление персоналом, информационная безопасность, право: Материалы региональной научно-практической конференции. Владимир, 17 декабря 2021 г. Владимир; 2022:78–82. URL: <https://elibrary.ru/iilwqn>

REFERENCES

1. Shumskaya O. O., Evsyutin O. O. Embedding information in compressed JPEG images with minimizing the manifestation of unmasking features. GraphiCon 2018: Proceedings of the 28th International Conference on Computer Graphics and Machine Vision. Tomsk: National Research Tomsk Polytechnic University; 2018:185–189. URL: <https://elibrary.ru/zamymx> (In Russ.).
2. Melman A. S., Evsyutin O. O. Effective and error-free information hiding in the hybrid domain of digital images using metaheuristic optimization. *Computer research and modeling*. 2023;15(1):197–210. (In Russ.). DOI: 10.20537/2076-7633-2023-15-1-197-210
3. Radaev S. V., Orlov D. V., Basov O. O. Combined shorthand algorithm for arranging confidential information into digital JPEG images. *Scientific bulletin of Belgorod State University. Series: Economics. Computer science*. 2017;23(272):185–192. URL: <https://elibrary.ru/yllsbp> (In Russ.).
4. Petrakov A. V., Malinichev D. M., Bugaev V. S., Romantsov A. P., Fedyaev L. S. Implementation of computer audio-video steganography methods for hiding data in jpeg (JFIF) image files. *Electrotechnical and information complexes and systems*. 2009;4(5):23–27. URL: <https://elibrary.ru/kwvtml> (In Russ.).
5. Shumskaya O. O., Zhelezny M. Adaptive algorithm for embedding information in compressed JPEG images based on the replacement operation. *Information management systems*. 2018;5:44–56. (In Russ.). DOI: 10.31799/1684-8853-2018-5-44-56

6. Secrets M.V. The algorithm for embedding containers in JPEG images. Applied digital technologies and systems of the XXI century: economics, management, personnel management, information security, law: Proceedings of the regional scientific and practical conference. Vladimir, December 17, 2021. Vladimir; 2022:78–82. URL: <https://elibrary.ru/iilwqn> (In Russ.).

ИНФОРМАЦИЯ ОБ АВТОРАХ / ABOUT THE AUTHORS

Алексей Алексеевич Рыженко — кандидат технических наук, доцент, доцент кафедры информационной безопасности факультета информационных технологий и анализа больших данных, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

Alexey A. Ryzhenko — Cand. Sci. (Tech.), Assoc. Prof., Assoc. Prof. of the Department of Information Security, Faculty of Information Technology and Big Data Analysis, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

<https://orcid.org/0000-0002-7279-9929>

Автор для корреспонденции / Corresponding author:

AARyzhenko@fa.ru

Сергей Игоревич Козьминых — доктор технических наук, доцент, профессор кафедры прикладной информатики и информационной безопасности, Российский экономический университет им. Г.В. Плеханова, Москва, Российская Федерация; профессор кафедры информационной безопасности, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

Sergey I. Kozminykh — Dr. Sci. (Tech.), Associate Professor, Professor, Department of Applied Informatics and Information Security, Plekhanov Russian University of Economics, Moscow, Russian Federation; Professor, Department of Information Security, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

kozminyh.si@rea.ru; SIKozminykh@fa.ru

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Conflicts of Interest Statement: The authors have no conflicts of interest to declare.

Статья поступила 09.07.2025; принята к публикации 10.08.2025.

Авторы прочитали и одобрили окончательный вариант рукописи.

The article was received 09.07.2025; accepted for publication 10.08.2025.

The authors read and approved the final version of the manuscript.

DOI: 10.26794/3033-7097-2025-1-3-62-68
УДК 355.01:004.8 (045)

Организационные аспекты применения искусственного интеллекта в сфере обороны

А.Н. Когтева

Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

АННОТАЦИЯ

В условиях цифровой трансформации ИИ становится ключевым фактором повышения боевой эффективности и обеспечения устойчивого развития Вооруженных Сил Российской Федерации. Статья посвящена анализу организационных аспектов применения искусственного интеллекта в сфере обороны Российской Федерации в условиях необходимости поиска новых путей укрепления национальной безопасности и обороноспособности. **Цель статьи** — проанализировать текущие проблемы внедрения искусственного интеллекта в организационные процессы оборонной деятельности России и предложить практические пути их решения. **Методы.** В исследовании применялись методы системного анализа, сравнительного анализа и структурно-функционального подхода, позволившие изучить особенности интеграции искусственного интеллекта в организационно-управленческие процессы Вооруженных Сил России. Выбор методов исследования обусловлен их взаимодополняющими возможностями и позволяет всесторонне раскрыть процессы интеграции искусственного интеллекта в сферу обороны. **Проблематика.** В работе идентифицированы ключевые проблемы внедрения искусственного интеллекта в сферу обороны: технологические (ограничения искусственного интеллекта, интеграция в существующие системы); ресурсные (финансовая и инфраструктурная недостаточность); кадровые (дефицит специалистов); институциональные (нормативно-правовые и этические вопросы), организационные (недостаток межведомственной координации). **Результаты.** Автором предложены конкретные организационно-управленческие меры, направленные на преодоление этих препятствий и эффективную интеграцию новых технологий. В статье впервые комплексно рассмотрены организационно-управленческие вопросы интеграции искусственного интеллекта в контексте Вооруженных сил Российской Федерации с учетом современных вызовов и национальных особенностей. Выявлены актуальные проблемы внедрения искусственного интеллекта в процессы оборонного управления, такие как технологические, кадровые и нормативные барьеры. **Выводы.** Для эффективного внедрения искусственного интеллекта в сферу обороны необходимо решение выявленных управленческих проблем посредством разработки соответствующей нормативной базы, совершенствования системы подготовки кадров, укрепления межведомственного взаимодействия, адаптации существующих управленческих практик к новым технологическим реалиям. Кроме того, выявлена необходимость внедрения пилотных проектов и поэтапной адаптации управленческих практик при условии постоянного мониторинга и оценки эффективности применяемых мер.

Ключевые слова: искусственный интеллект; оборонная сфера; организационные процессы; вооруженные силы; внедрение технологий; оборонные инновации; оборонное управление

Для цитирования: Когтева А.Н. Организационные аспекты применения искусственного интеллекта в сфере обороны. *Цифровые решения и технологии искусственного интеллекта*. 2025;1(3):62-68. DOI: 10.26794/3033-7097-2025-1-3-62-68

ORIGINAL PAPER

Organizational Aspects of the Use Artificial Intelligence in the Field of Defense

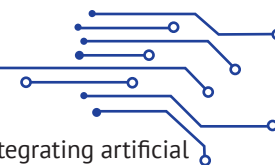
A.N. Kogteva

Financial University under the Government of the Russian Federation, Moscow, Russian Federation

ABSTRACT

In the context of digital transformation, AI is becoming a key factor in increasing combat effectiveness and ensuring the sustainable development of the Armed Forces of the Russian Federation. The article is devoted to the analysis of organizational aspects of the use of artificial intelligence in the field of defense of the Russian Federation, in the context of the need to find new ways to strengthen national security and defense capability. **The purpose of the article** is to analyze the current problems of introducing artificial intelligence into the organizational processes of Russia's defense activities and propose practical ways to solve them. **Methods.** The research used methods of system analysis, comparative analysis, and a structural and functional approach, which made it possible to study the features of the integration of artificial intelligence into the organizational and managerial processes of the Russian Armed Forces. The choice of research methods is based on

© Когтева А.Н., 2025



their complementary capabilities and allows for a comprehensive understanding of the processes of integrating artificial intelligence into the defense sector. **The problem.** The paper identifies key problems in the implementation of artificial intelligence in the defense sector: technological (limitations of artificial intelligence, integration into existing systems); resource-related (financial and infrastructural inadequacies); human resources (shortage of specialists); institutional (regulatory, legal, and ethical issues); organizational (lack of interagency coordination).

Results. The author proposes specific organizational and operational measures aimed at overcoming these obstacles and effectively integrating new technologies. For the first time, the article comprehensively examines the organizational and management issues of integrating artificial intelligence in the context of the Armed Forces of the Russian Federation, taking into account modern challenges and national characteristics. It identifies current problems in the implementation of artificial intelligence in defense management processes, such as technological, personnel, and regulatory barriers.

Conclusions. For the effective implementation of artificial intelligence in the field of defense, it is necessary to solve the identified management problems by developing an appropriate regulatory framework, improving the personnel training system, strengthening interdepartmental cooperation, and adapting existing management practices to new technological realities. In addition, the need for the implementation of pilot projects and step-by-step adaptation of management practices has been identified, subject to constant monitoring and evaluation of the effectiveness of the measures applied.

Keywords: artificial intelligence; defense sector; organizational processes; armed forces; technology implementation; defense innovations; military management

For citation: Kogteva A.N. Organizational aspects of the use artificial intelligence in the field of defense. *Digital Solutions and Artificial Intelligence Technologies*. 2025;1(3):62-68. DOI: 10.26794/3033-7097-2025-1-3-62-68

ВВЕДЕНИЕ

Искусственный интеллект активно интегрируется в различные сферы обороны и военно-промышленного комплекса (ВПК), способствуя повышению эффективности боевых операций и обеспечению устойчивого развития оборонной отрасли. Государства — мировые лидеры разрабатывают стратегии применения искусственного интеллекта в военных целях. В США с 2019 г. функционирует Объединенный центр искусственного интеллекта (Joint AI Center), координирующий проекты в сфере ИИ для военных нужд. Параллельно министерство обороны утвердило комплексную стратегию развития искусственного интеллекта. Китай, в свою очередь, активно реализует концепцию «интеллектуализации» военных операций, рассматривая искусственный интеллект как ключевой фактор достижения оборонно-технического превосходства.

В Российской Федерации данные тенденции также отмечены на высоком уровне и получили признание и поддержку. Национальная стратегия развития искусственного интеллекта до 2030 года¹ закрепляет приоритетное направление по применению ИИ-технологий для обеспечения национальной безопасности и укрепления обороноспособности государства.

В оборонной сфере наблюдается интенсивное развитие технологических инноваций. В 2020 г.

Минобороны России инициировало разработку системы искусственного интеллекта, базирующейся на технологии глубоких нейронных сетей. Бюджет проекта составил 387,8 млн руб. Дальнейшее развитие этого направления получило продолжение в 2021 г., когда в структуре Минобороны России было создано специализированное управление по развитию технологий искусственного интеллекта [1, с. 67]. Особое внимание уделяется применению ИИ в сфере беспилотных летательных аппаратов.

Данные инициативы наглядно демонстрируют стратегическое намерение военного ведомства активно интегрировать передовые технологии искусственного интеллекта в повседневную практику вооруженных сил.

Однако внедрение искусственного интеллекта в оборонную сферу связано с серьезными рисками и вызовами. По мнению исследователей Х.А. Казеяна и Г.Э. Арутюняна, применение ИИ в военных целях может представлять экзистенциальную угрозу для человечества. Среди потенциальных опасностей — неконтролируемая эскалация конфликтов между крупными державами, риск ядерного противостояния и возможность выхода искусственного интеллекта из-под контроля человека [2, с. 36, 37].

Учитывая вышеперечисленные риски и угрозы, требуется особый подход к интеграции ИИ в оборонную сферу. Ключевым аспектом становится заблаговременная разработка нормативных основ и создание инфраструктуры для безопасного применения технологий искусственного интеллекта. Без этого внедрение может привести к непредсказуемым и негативным последствиям. В российской практике особо актуален поиск оптимального ба-

¹ Утверждена Указом Президента РФ от 10.10.2019 № 490 «О развитии искусственного интеллекта в Российской Федерации» (с изм. и доп.). URL: <https://base.garant.ru/72838946/?ysclid=mdpo2rcvv0729111319>



ланса между стремлением использовать потенциал ИИ для укрепления обороноспособности страны и необходимостью минимизации сопутствующих рисков. Все это порождает комплексную научно-практическую задачу: определить необходимый спектр организационных преобразований для эффективного и безопасного внедрения технологий искусственного интеллекта в деятельность вооруженных сил, а также выявить существующие препятствия на этом пути.

СОСТОЯНИЕ И ТЕНДЕНЦИИ ПРИМЕНЕНИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В СФЕРЕ ОБОРОНЫ РОССИИ

В последние пять лет в России формируется институциональная основа для развития технологий искусственного интеллекта в сфере обороны. Принятая в 2019 г. национальная стратегия определила ключевые задачи в этой области до 2030 г., благодаря чему стало возможным создание специальных структур в Минобороны.

Кроме того, в Анапе с 2018 г. функционирует Военный инновационный технополис ЭРА, предназначенный для апробации и внедрения передовых военных технологий. На этой площадке сосредоточены научные роты и лаборатории, где молодые специалисты совместно с предприятиями оборонно-промышленного комплекса занимаются проектами по искусственному интеллекту, робототехнике, информационной безопасности и др. По данным Минобороны России, технополис ЭРА стал одной из основных площадок для проведения испытания технологий искусственного интеллекта и доведения их до практического применения — в том числе непосредственно в образцах вооружений и техники, используемых в ходе современного вооруженного конфликта².

В рамках данной экосистемы налажено сотрудничество с индустриальными партнерами: десятки предприятий, от крупных оборонных концернов до высокотехнологичных стартапов, участвуют в проектах ЭРА. Управление развития технологий искусственного интеллекта Минобороны России осуществляет экспертизу предлагаемых разработок на наличие элементов искусственного интеллекта и ведет специальный реестр инновационных проектов, рекомендованных к включению в го-

сударственный оборонный заказ. Таким образом, создается задел для планомерного внедрения соответствующих технологий в вооружение и военную технику.

Помимо материально-технической и научной базы, уделяется внимание интеграции искусственного интеллекта в системы управления войсками. Руководство Минобороны России подчеркивает приоритет данного направления. В частности, С. К. Шойгу в период пребывания на должности министра обороны указал на необходимость совершенствовать систему управления и связи вооруженных сил за счет активного использования технологий искусственного интеллекта³. Что касается актуальных задач, то в планах оборонного ведомства на ближайшие годы — модернизация контуров управления с внедрением интеллектуальных систем, способных автоматизировать сбор и обработку разведывательной информации, помогать командирам в принятии решений, повышать оперативность и надежность связи.

Уже сегодня элементы искусственного интеллекта применяются в ряде направлений: системах ситуационного анализа, распознавания образов (например, для анализа данных воздушной и космической разведки), кибербезопасности, тренажерах и моделирующих комплексах. Отдельное внимание уделяется беспилотным летательным аппаратам и робототехническим комплексам с элементами автономности, которые доказали свою эффективность [3, с. 102, 103]. Боевые дроны с функциями искусственного интеллекта и роботизированные платформы проходят испытания и частично используются в войсках, что подтверждено опытом боевых действий последнего времени [4, с. 338]. Таким образом, основы для применения искусственного интеллекта в оборонно-промышленном комплексе (ОПК) заложены, и ведется работа по расширению его практического применения.

АКТУАЛЬНЫЕ ПРОБЛЕМЫ ВНЕДРЕНИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В ОРГАНИЗАЦИОННЫЕ ПРОЦЕССЫ В СФЕРЕ ОБОРОНЫ

Несмотря на описанные позитивные сдвиги, на пути внедрения искусственного интеллекта в сфере обороны России существует ряд серьезных проблем организационного характера.

² Реестр технологий ИИ позволяет формировать облик новых образцов вооружения. URL: https://vpk.name/news/902447_reestr_tehnologii_ii_pozvolyaet_formirovat_oblik_novyh_obrazcov_vooruzheniya.html

³ Шойгу С. К. Систему управления и связи ВС РФ улучшат с использованием технологий ИИ. Ведомости. URL: <https://www.vedomosti.ru/politics/news/2023/01/10/958482-shoigu-tehnologii-ii?ysclid=mdpuxye0oh988264903>



1. Технологические и ресурсные ограничения. Современные модели искусственного интеллекта требуют мощной вычислительной инфраструктуры и элементной базы (микроэлектроника, процессоры, датчики). Санкционные и внешнеэкономические ограничения создали дефицит доступа к передовым аппаратным решениям. Российская оборонная промышленность вынуждена полагаться на собственную элементную базу, которая пока уступает мировому уровню. Это учитывается и в Минобороны России: ведомство специально проверяет работоспособность разрабатываемых технологий на отечественной элементной базе, и испытания проходят достаточно успешно⁴.

Тем не менее пока задача импортозамещения высокотехнологичных компонентов остается сложнореализуемой, требуя управленческих решений по координации с гражданскими программами развития микроэлектроники и дополнительного финансирования НИОКР [5, с. 64]. Ограниченность ресурсов также проявляется в конкуренции за перспективные и талантливые кадры и финансирование: искусственный интеллект — дорогостоящее направление, и вооруженным силам нужно обосновывать выделение значительных средств на опытно-экспериментальную деятельность, особенно в условиях повышения нагрузок на оборонный бюджет.

2. Нехватка квалифицированных кадров и компетенций в области искусственного интеллекта. Традиционно основные центры разработки сосредоточены в гражданском секторе (университеты, IT-компании). Вооруженные силы испытывают дефицит специалистов, одновременно разбирающихся и в алгоритмах искусственного интеллекта, и в специфике военного дела [6, с. 139]. Для решения этой проблемы предпринимаются организационные шаги: созданы научные роты, где служат выпускники IT-специальностей, выполняя исследования для армии; налажено взаимодействие с ведущими техническими вузами (МФТИ, МГТУ им. Баумана, Военная академия связи и др.), открыты специализированные кафедры и программы обучения офицеров основам анализа данных и программирования.

Тем не менее кадровый разрыв еще предстоит преодолеть. Необходимы новые подходы к подготовке военных специалистов по искусственному

интеллекту — возможно введение военно-учетных специальностей, связанных с программной инженерией и анализом больших данных, расширение целевого набора в вузы по данным направлениям. Кроме того, существует проблема удержания кадров: молодые специалисты в обозначенной сфере часто предпочитают более высокооплачиваемую работу в коммерческом секторе, поэтому Минобороны России требует разработки стимулирующих мер (социальных гарантий, карьерного роста, интересных задач) для привлечения и удержания талантливых кадров.

3. Институциональные и культурные барьеры внутри вооруженных сил. Внедрение новых технологий требует адаптации устоявшихся структур и процессов управления войсками. Российская военная система характеризуется определенным консерватизмом и инерционностью управленческих практик. Новые концепции ведения боевых действий с применением искусственного интеллекта (например, сетцентрические и когнитивные операции) пока не получили полноценного отражения в официальных доктринальных документах и полевых уставах. Командиры и штабные офицеры среднего звена зачастую недостаточно информированы о возможностях искусственного интеллекта либо не доверяют автоматизированным системам принятия решений, полагаясь на традиционные методы. Имеется также разрыв между разработчиками соответствующих цифровых решений и конечными пользователями сферы обороны, в результате чего созданные алгоритмы могут не полностью учитываться в реальной работе штабов.

Таким образом, назрела необходимость **адаптации организационной структуры, доктрин и процессов** под задачи искусственного интеллекта. Для успешной реализации его потенциала в сфере обороны требуются изменения в концепциях применения войск, структурах сил и подходах к комплектованию, иначе технологии не будут эффективно востребованы.

В современном контексте эта проблема проявляется в опережении прогрессом технических разработок эволюции нормативных процедур и бюрократии. Тем не менее первые шаги к преодолению консерватизма наблюдаются — в ходе учений и реальных операций отмечается более активное использование автоматизированных систем, а руководящий состав получает практический опыт обращения с элементами искусственного интеллекта на поле боя. Вопрос: насколько быстро и полно организационные структуры сумеют усвоить уроки

⁴ Шойгу С.К. Систему управления и связи ВС РФ улучшат с использованием технологий ИИ. Ведомости. URL: <https://www.vedomosti.ru/politics/news/2023/01/10/958482-shoigutehnologii-ii?ysclid=mdpuxye0oh988264903>



и оперативно перестроиться на «цифровые рельсы» во время продолжающегося конфликта?

4. Недостаточность нормативно-правовой базы и этических регламентов. Правовое поле применения искусственного интеллекта в военном деле пока находится в стадии формирования. Отсутствуют официально утвержденные отечественные нормативные правовые акты, детально регламентирующие статус и степень автономности боевых систем с искусственным интеллектом, протоколы принятия решений совместно людьми и машинами, ответственность за ошибки цифровых технологий и т.п. При этом международно-правовые аспекты (например, определение статуса летальных автономных систем оружия — LAWS)⁵ также не до конца урегулированы.

Такая неопределенность порождает риски правового вакуума: внедрение искусственного интеллекта может опережать формирование законодательных норм и способность командования вырабатывать надлежащие правила применения. В результате возможны либо неоправданно рискованные эксперименты, либо, наоборот, чрезмерно осторожное отношение, тормозящее прогресс.

Обозначенная проблема указывает на настоятельную необходимость создания опережающей нормативной базы и инфраструктуры для безопасного использования искусственного интеллекта в военном управлении [2, с. 37]. Другими словами, «правило безопасности» диктует: сначала нормы и меры контроля — потом массовое внедрение. Пока же в России роль подобных регуляторов выполняют разрозненные приказы и методические рекомендации. Требуется на уровне Минобороны России разработать комплекс руководящих документов, посвященных применению технологий искусственного интеллекта: от руководств по испытаниям и сертификации таких систем до этического кодекса его применения в сфере обороны, аналогично введению принципов ответственного искусственного интеллекта для оборонного сектора в отдельных странах⁶.

5. Межведомственная координация и обмен данными. Эффективное использование искусственного интеллекта в большой степени зависит от наличия больших объемов данных и взаимодействия разных источников информации. В силовых структурах России (Минобороны, ФСБ, МВД,

Росгвардия и др.) накоплены значительные массивы данных разведки, наблюдения, оперативной информации. Однако эти данные разобщены по ведомственным хранилищам [7], что затрудняет их интегрированное использование для обучения моделей искусственного интеллекта. Управленческой проблемой является отсутствие единого межведомственного центра или платформы обмена данными и решениями между различными силовыми органами. Каждый из них реализует свои цифровые проекты в относительной изоляции.

Например, системы интеллектуального видеонаблюдения для обеспечения общественной безопасности внедряются МВД и Росгвардией, а военная разведка развивает собственные аналоги — при ограниченном обмене опытом. Отчасти эту задачу решает разрабатываемый упомянутый реестр технологий искусственного интеллекта, в который, по замыслу Минобороны России, будут включаться наиболее перспективные решения, пригодные для Гособоронзаказа и общего пользования силовыми структурами⁷.

Кроме того, на государственном уровне с 2022 г. функционирует Национальный центр развития искусственного интеллекта при Правительстве, куда стекается информация о проектах искусственного интеллекта в различных отраслях, однако военная тематика во многом остается закрытой. Необходимо усилить координацию — например, создать рабочую группу Совета безопасности или межведомственный совет по внедрению искусственного интеллекта в сфере обороны, где представители всех силовых структур могли бы вырабатывать единые подходы, стандарты совместимости систем и совместные проекты. Без этого существует риск дублирования усилий и упущения синергетического эффекта от объединения ресурсов.

ВЫВОДЫ

Организационные аспекты применения искусственного интеллекта в сфере обороны имеют решающее значение для успеха данной технологической трансформации. Проведенный анализ российской практики показал, что в стране созданы первоначальные условия для внедрения искусственного интеллекта в войсках: определены стратегические ориентиры, формируются

⁵ URL: <https://disarmament.unoda.org/the-convention-on-certain-conventional-weapons/background-on-laws-in-the-ccw/>

⁶ URL: <https://www.tadviser.ru/a/719555>

⁷ Шойгу С. К. Систему управления и связи ВС РФ улучшат с использованием технологий ИИ. Ведомости. URL: <https://www.vedomosti.ru/politics/news/2023/01/10/958482-shoigu-tehnologii-ii?ysclid=mdpuxey0oh988264903>

специализированные структуры (технополис ЭРА, управление искусственного интеллекта в Минобороны России), реализуются пилотные проекты.

Вместе с тем выявлен ряд проблем, тормозящих полноценную интеграцию искусственного интеллекта в организационные процессы в сфере обороны. К таким проблемам относятся:

- технологические ограничения (зависимость от импортных компонентов, дефицит вычислительных ресурсов);

- нехватка квалифицированных кадров;
- институциональная инерция и недостаток межведомственной координации;
- пробелы в нормативно-правовом регулировании.

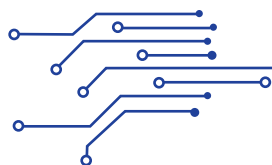
Указанные препятствия носят объективный характер, однако могут быть преодолены за счет целенаправленных управленческих решений. Более подробно вопрос перспектив применения искусственного интеллекта в сфере обороны будет раскрыт в следующей статье цикла.

СПИСОК ИСТОЧНИКОВ

1. Павленко Б. В., Бондаренко В. И., Мартыненко А. А. М. Использование модели YOLO в современных задачах распознавания в реальном времени на примере военной и космической отраслей. *Вестник ДонНУ. Серия Г: Технические науки*. 2023;4:65–75. URL: <https://elibrary.ru/tjkhxi>
2. Казеян Х. А., Арутюнян Г. Э. Проблемы применения искусственного интеллекта в военном управлении. *Управленческое консультирование*. 2023;6:34–45. DOI: 10.22394/1726-1139-2023-6-34-45
3. Иванец В. М., Лукьянчик В. Н., Мельник В. Н. Особенности управления беспилотными летательными аппаратами в составе беспилотной интеллектуальной авиационной системы на основе технологий искусственного интеллекта. *Военная мысль*. 2022;9:100–109. URL: <https://elibrary.ru/heywwn>
4. Лукьянчик В. Н., Константинов С. А., Селезнев А. В. Особенности управления беспилотных летательных аппаратов в составе беспилотной авиационной системы на основе технологий искусственного интеллекта. *Инновационная деятельность в Вооруженных Силах Российской Федерации*. 2021:336–342. URL: <https://elibrary.ru/rumhic>
5. Ленчук Е. Б. Технологическая модернизация как основа антисанкционной политики. *Проблемы прогнозирования*. 2023;4:54–66. DOI: 10.47711/0868-6351-199-54-66
6. Аверьянов А. О., Степуть И. С., Гуртов В. А. Прогноз кадровой потребности для сферы искусственного интеллекта в России. *Проблемы прогнозирования*. 2023;1:129–143. DOI: 10.47711/0868-6351-196-129-143
7. Храпский А. Е. Беспилотные воздушные суда как средство противодействия войскам в ходе выполнения служебно-боевых задач. Опыт борьбы с БВС в зарубежных странах. *Актуальные исследования*. 2023;49–1(179):43–50. URL: <https://elibrary.ru/ihvdbr>

REFERENCE

1. Pavlenko B. V., Bondarenko V. I., Martynenko A. A. M. Using the YOLO model in modern real-time recognition tasks using the example of the military and space industries. *The messenger of Donnu. Series G: Technical Sciences*. 2023;4:65–75. URL: <https://elibrary.ru/tjkhxi> (In Russ.).
2. Kazeyan H. A., Harutyunyan G. E. Problems of using artificial intelligence in military management. *Management consulting*. 2023;6:34–45. (In Russ.). DOI: 10.22394/1726-1139-2023-6-34-45
3. Ivanets V. M., Lukyanchik V. N., Melnik V. N. Features of control of unmanned aerial vehicles as part of an unmanned intelligent aviation system based on artificial intelligence technologies. *Military thought*. 2022;9:100–109. URL: <https://elibrary.ru/heywwn> (In Russ.).
4. Lukyanchik V. N., Konstantinov S. A., Seleznev A. V. Control features of unmanned aerial vehicles as part of an unmanned aircraft system based on artificial intelligence technologies. *Innovative activity in the Armed Forces of the Russian Federation*. 2021:336–342. URL: <https://elibrary.ru/rumhic> (In Russ.).
5. Lenchuk E. B. Technological modernization as the basis of anti-sanctions policy. *Forecasting problems*. 2023;4:54–66. (In Russ.). DOI: 10.47711/0868-6351-199-54-66
6. Averyanov A. O., Stepus I. S., Gurtov V. A. Forecast of personnel requirements for the field of artificial intelligence in Russia. *Forecasting problems*. 2023;1:129–143. (In Russ.). DOI: 10.47711/0868-6351-196-129-143
7. Khrapskiy A. E. Unmanned aerial vehicles as a means of countering troops during the performance of service and combat missions. The experience of fighting MENA in foreign countries. *Current research*. 2023;49–1(179):43–50. URL: <https://elibrary.ru/ihvdbr> (In Russ.).



ИНФОРМАЦИЯ ОБ АВТОРЕ / ABOUT THE AUTHOR

Анна Николаевна Когтева — кандидат экономических наук, доцент кафедры информационной безопасности факультета информационных технологий и анализа больших данных, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

Anna N. Kogteva — Cand. Sci. (Econ.), Assoc. Prof., Department of Information Security, Faculty of Information Technology and Big Data Analysis, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

<https://orcid.org/0000-0001-8109-9900>

ANKogteva@fa.ru

Конфликт интересов: автор заявляет об отсутствии конфликта интересов.

Conflicts of interest Statement: The author has no conflicts of interest to declare.

Статья поступила 10.07.2025; принята к публикации 29.07.2025.

Автор прочитала и одобрила окончательный вариант рукописи.

The article was received on 10.07.2025; accepted for publication on 29.07.2025.

The author read and approved the final version of the manuscript.

Анализ публикационной активности и научных коллабораций научно-педагогических работников Финансового университета

Г.А. Остапенко, Г.Г. Рожкова, В.Г. Феклин, Р.А. Кочкаров

Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

АННОТАЦИЯ

В условиях стремительного роста объема научных публикаций, увеличения междисциплинарных исследований и усиления конкуренции в академической среде особенно актуальными становятся задачи анализа и визуализации научной активности. Современные цифровые инструменты позволяют не только отслеживать публикационную динамику, но и выявлять ключевые исследовательские направления, а также устойчивые коллективы авторов, образующие научные сообщества. Одним из эффективных подходов в данной области является сочетание методов тематического моделирования и сетевого анализа на основе теории графов. Научные организации зачастую сталкиваются с проблемой отсутствия оперативной информации о внутренней структуре исследовательской деятельности: какие тематики наиболее активно развиваются, каковы связи между авторами и коллективами, кто выступает в роли «ядер» научных сообществ. Особенно это актуально для крупных вузов, где работают сотни исследователей, создающих значительное число научных работ. В такой ситуации ручной анализ становится невозможным, и на помощь приходят методы автоматизированной обработки текстов и графовой аналитики. Настоящая статья посвящена анализу публикационной активности авторов Финансового университета. Цель исследования заключается в выявлении тематик научных работ и выделении научных сообществ для понимания развития научно-исследовательской деятельности высших учебных заведений на примере Финансового университета. В исследовании представлен подход к формированию набора данных научных публикаций авторов Финансового университета. Проведены визуализация динамики публикаций и анализ ключевых слов, позволяющие выявить общие тренды. Для решения задачи кластеризации текстов и определения тематик публикаций применена модель BERTopic. Выявление научных сообществ реализовано посредством построения и анализа графа соавторства, что позволяет идентифицировать группы исследователей, активно сотрудничающих в рамках определенных научных направлений.

Ключевые слова: научные сообщества; тематическое моделирование; BERTopic; кластеризация текстов; граф соавторства

Для цитирования: Остапенко Г.А., Рожкова Г.Г., Феклин В.Г., Кочкаров Р.А. Анализ публикационной активности и научных коллабораций научно-педагогических работников Финансового университета. *Цифровые решения и технологии искусственного интеллекта*. 2025;1(3):69-76. DOI: 10.26794/3033-7097-2025-1-3-69-76

ORIGINAL PAPER

Thematic Analysis of Publication Activity Among Academic and Teaching Staff: A Case Study of the Financial University

G.A. Ostapenko, G.G. Rozhkova, V.G. Feklin, R.A. Kochkarov

Financial University under the Government of the Russian Federation, Moscow, Russian Federation

ABSTRACT

In the context of the rapid growth of scientific publications, the increase in interdisciplinary research and increased competition in the academic environment, the tasks of analyzing and visualizing scientific activity are becoming especially actual. Modern digital tools allow not only to track publication dynamics, but also to identify key research areas, as well as stable groups of authors that form scientific communities. One of the effective approaches in this area is a combination of topic modeling methods and network analysis based on graph theory. Scientific organizations often face the problem of lack of operational information about the internal structure of research activities: which topics are most actively developing, what are the connections between authors and teams, who acts as the “cores” of scientific communities. This is especially true for large universities, where hundreds of researchers work, creating a significant number of scientific papers. In such



situation, manual analysis becomes impossible, and automated text processing and graph analytics methods come to the rescue. This article is devoted to the analysis of the publication activity of authors of the Financial University. The purpose of the study is to identify the topics of scientific papers and identify scientific communities to understand the development of research activities of higher education institutions by example of the Financial University. The study presents an approach to forming a data set of scientific publications of authors of the Financial University. Visualization of publication dynamics and keyword analysis were carried out, allowing to identify common trends. The BERTopic model was used to solve the problem of text clustering and determining publication topics. Identification of scientific communities was implemented through the construction and analysis of a co-authorship graph, which allows to identify groups of researchers actively collaborating within certain scientific areas.

Keywords: scientific communities; topic modeling; BERTopic; text clustering; co-authorship graph

For citation: Ostapenko G.A., Rozhkova G.G., Feklin V.G., Kochkarov R.A. Thematic analysis of publication activity among academic and teaching staff: A case study of the financial university. *Digital Solutions and Artificial Intelligence Technologies*. 2025;1(3):69-76. DOI: 10.26794/3033-7097-2025-1-3-69-76

ВВЕДЕНИЕ

Как показывают исследования, применение тематического моделирования (topic modeling), в частности моделей типа LDA или современных BERTopic, позволяет извлекать осмысленные темы из текстов публикаций, включая аннотации и заголовки, даже при значительном тематическом разнообразии [1]. В совокупности с построением графа соавторства и его кластерным анализом такие методы позволяют выявлять научные школы, измерять плотность взаимодействий между исследователями и оценивать структуру академической коллаборации.

В настоящем исследовании предпринята попытка адаптировать указанный подход для анализа научной активности сотрудников Финансового университета за 2020–2024 гг. На основе публикаций научной электронной библиотеки* построены:

- тематическая модель публикаций с использованием BERTopic и различных эмбедингов;
- граф соавторства.

Такой подход позволяет не только выявить наиболее актуальные научные направления, но и оценить внутреннюю структуру взаимодействия исследовательских коллективов.

Последние годы ознаменовались резкими изменениями в глобальной научной повестке, во многом обусловленными пандемией COVID-19. В период 2020–2021 гг. научное сообщество столкнулось с необходимостью быстрой адаптации к новым условиям — дистанционному обучению, удаленной работе, трансформации экономических и социальных процессов. Это неизбежно повлияло на структуру научных исследований: наблюдается рост интереса к тематикам, связанным с цифровой трансформацией, онлайн-образованием, здравоохранением и устойчивым развитием. Публикационная активность от-

ражает эти изменения, формируя новые приоритеты в научной повестке университетов. Предполагается, что тематика публикаций сотрудников Финансового университета с 2020 по 2024 г. претерпела значительные изменения: на первый план вышли исследования в области цифровой экономики, искусственного интеллекта, онлайн-образования и социально-экономической устойчивости. Эти сдвиги не случайны и отражают реакцию научного сообщества на вызовы пандемии, а также государственные приоритеты в области цифровизации. Вместе с тем ожидается, что внутриуниверситетское научное взаимодействие обладает выраженной кластерной структурой, где сообщества формируются вокруг ведущих тем и ключевых исследователей. Отдельное внимание в работе уделяется роли этих исследователей как «узлов связности» — тех, кто объединяет разные тематические или институциональные группы.

АНАЛИЗ ПУБЛИКАЦИЙ

Публикационная активность исследователей Финансового университета в период 2020–2024 гг. демонстрирует заметную динамику. Далее 2025 г. по возможности не принимается к учету, так как год не завершен и не собраны данные. Согласно распределению по годам (рис. 1), 2020 г. отмечен относительно низким числом публикаций, что, вероятно, связано с первоначальным шоком от пандемии COVID-19: переходом на удаленную работу, временной приостановкой научных проектов, а также снижением активности научных коммуникаций в условиях неопределенности. Однако уже с 2021 г. наблюдается устойчивый рост публикационной активности, достигший максимума в 2022 г. — порядка 770 статей. Этот рост может быть обусловлен как накопленным исследовательским материалом, так и необходимостью научного осмысления последствий пандемии, включая цифровую трансформацию, развитие дистанционного образования и вопросы устойчивости социально-экономических систем.

* Научная электронная библиотека eLibrary.ru. URL: <https://www.elibrary.ru>



Рис. 1 / Fig. 1. Распределение количества публикаций по годам (2020–2024 гг.) / Distribution of the Number of Publications by Years (2020–2024)

Источник / Source: составлено авторами / Compiled by the authors.

Однако начиная с 2023 г. фиксируется заметное снижение количества научных публикаций. Одной из возможных причин этого снижения являются внешнеполитические и экономические ограничения, в том числе санкционные барьеры, повлиявшие на международную кооперацию, доступ к зарубежным научным журналам и базам данных, а также участие российских ученых в международных грантах и конференциях.

Примечательно, что значительная часть научных публикаций — более двух третей — выполнена на английском языке (рис. 2). Это свидетельствует о высоком уровне академической культуры и ориентации исследователей Финансового университета на международную научную коммуникацию. Даже в русскоязычных работах ключевые элементы — заголовки, аннотации и ключевые слова — часто представлены также на английском. Это существенно расширяет возможности интеграции в глобальное научное пространство и позволяет эффективно применять англоязычные методы тематического анализа.

Анализ научных публикаций Финансового университета свидетельствует о его устойчивой исследовательской ориентации на приоритетные направления современной экономической повестки. Одним из ключевых векторов является цифровизация экономики: стабильно высокое присутствие терминов «DIGITALIZATION», «DIGITAL ECONOMY», «DIGITAL TECHNOLOGIES» и «DIGITAL TRANSFORMATION».

С 2021 г. наблюдается усиление интереса к тематике пандемии COVID-19. Рост упоминаний

терминов «PANDEMIC», «COVID-19» и «COVID-19 PANDEMIC» отражает научный отклик на глобальный кризис и фокус исследовательского сообщества университета на анализе экономических, социальных и образовательных последствий пандемии.

Отдельное место занимает направление, связанное с искусственным интеллектом и машинным обучением. В период 2020–2024 гг. фиксируется устойчивый рост упоминаний «ARTIFICIAL INTELLIGENCE», «MACHINE LEARNING», «NEURAL

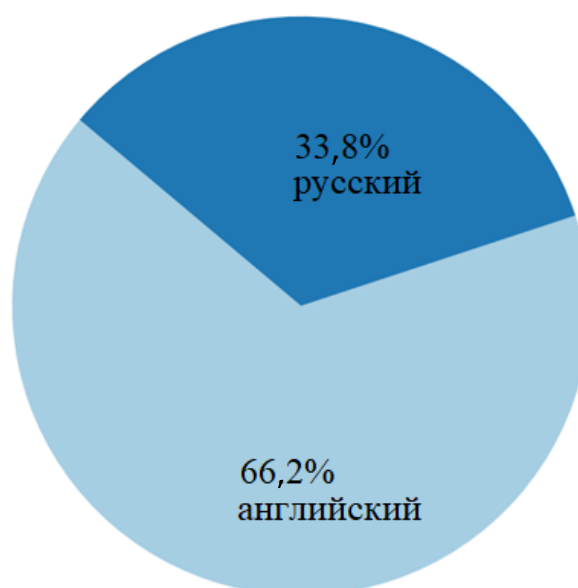


Рис. 2 / Fig. 2. Распределение количества публикаций по языку / Distribution of the Number of Publications by Language

Источник / Source: составлено авторами / Compiled by the authors.



NETWORKS» и «DEEP LEARNING», что свидетельствует об активном освоении современных технологических решений и их применении в экономике, финансах и государственном управлении.

Важной особенностью исследовательской повестки остается акцент на национальный и региональный контексты. Частое использование терминов «RUSSIA», «RUSSIAN EMPIRE», «REGION» и «RUSSIAN FEDERATION» подтверждает направленность части научных работ на изучение историко-экономических процессов и региональной специфики.

Стабильный интерес вызывают темы образования и человеческого капитала, о чем свидетельствует регулярное упоминание таких ключевых слов, как «EDUCATION», «HIGHER EDUCATION», «HUMAN CAPITAL» и «STUDENTS». Традиционные направления экономических исследований — «ECONOMIC GROWTH», «ECONOMIC DEVELOPMENT», «EFFICIENCY», «INVESTMENTS» — сохраняют актуальность и отражают фундаментальные аспекты научного интереса.

Начиная с 2022 г. в структуре исследовательских интересов усиливается экологическая и энергетическая тематика, что выражается в росте частоты употребления терминов «ECOLOGY», «RENEWABLE ENERGY», «ENERGY EFFICIENCY». Это согласуется с глобальными трендами устойчивого развития и «зеленой» экономики (рис. 3).

ОПРЕДЕЛЕНИЕ ТЕМАТИК

Для выявления ключевых тем в каталоге научных публикаций была использована модель тематической кластеризации BERTopic [2]. Модель опирается на современные методы представления текста и кластеризации, позволяя получать интерпретируемые темы, извлекаемые из текстов различной длины и структуры, в том числе научных аннотаций, заголовков и ключевых слов.

Архитектура BERTopic

BERTopic представляет собой модульную модель, в основе которой лежат следующие компоненты:

- Модель эмбединга предложений (Sentence Embeddings): преобразует текст в векторное пространство. В работе используются модели из библиотеки SentenceTransformers.
- UMAP (Uniform Manifold Approximation and Projection): метод понижения размерности, сохраняющий локальную структуру данных. Использовались параметры: `n_neighbors=10`, `n_components=5`, `min_dist=0.0`, `metric='cosine'`.
- HDBSCAN (Hierarchical Density-Based Spatial Clustering): плотностной алгоритм кластеризации, способный выделять кластеры различной формы и отсеивать выбросы. Использованы параметры: `min_cluster_size=10`, `min_samples=5`, `metric='euclidean'`.

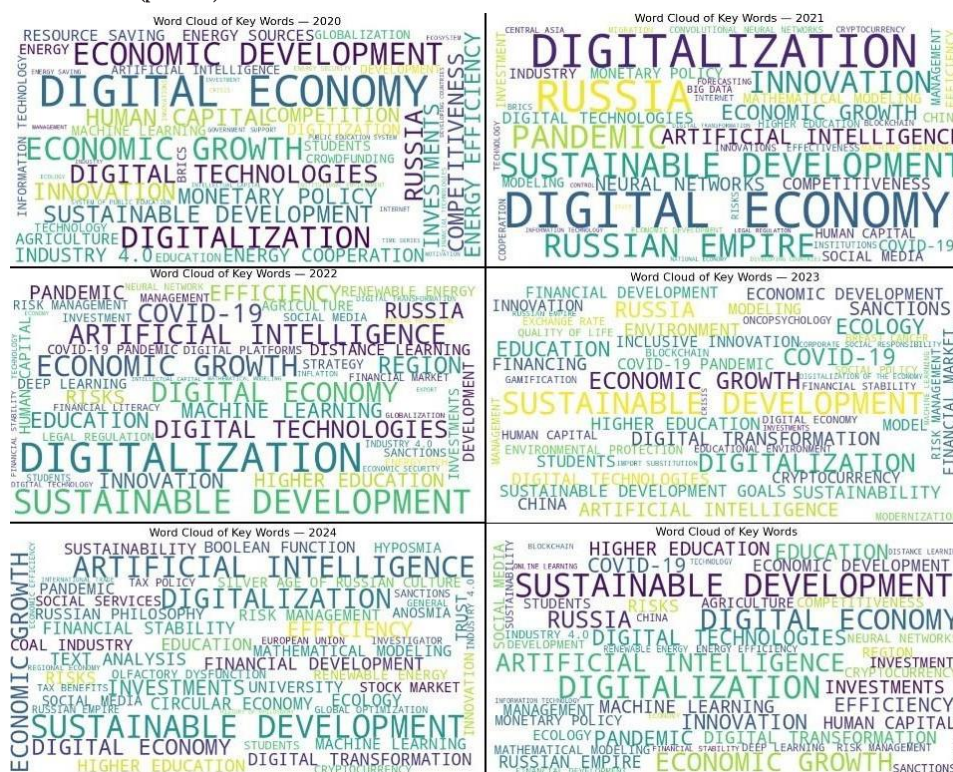


Рис. 3 / Fig. 3. Облака ключевых слов публикаций по годам / Keyword Clouds of Publications by Years

Источник / Source: составлено авторами / Compiled by the authors.



- CountVectorizer: для построения матрицы частот n-грамм (в данной работе использовались uni-, bi- и триграммы), игнорируя стоп-слова.

- Maximal Marginal Relevance (MMR): метод для выбора наиболее разнообразных и репрезентативных терминов в теме (diversity=0.3).

- Модель BERTopic: объединяет вышеуказанные компоненты и возвращает набор тем, сопоставленных с документами, а также визуализации (графики, интерактивные диаграммы).

Для оценки влияния качества эмбедингов на результаты кластеризации проведены эксперименты с пятью предварительно обученными моделями:

- all-mpnet-base-v2 (baseline): мощная модель на базе MPNet, демонстрирует высокое качество на общих задачах семантического сопоставления;

- all-MiniLM-L6-v2: облегченная и быстрая модель MiniLM с хорошим качеством при существенно меньшем размере;

- intfloat/multilingual-e5-large-instruct: мультилингвальная модель, специально адаптированная для поиска и семантического сопоставления в условиях многозначности языков [3];

- mixedbread-ai/mxbai-embed-large-v1: современная универсальная модель для эмбедингов предложений, ориентированная на разнообразие тем и приложений;

- allenai-specter: модель от AllenAI, специально обученная на корпусе научных статей и цитат для задач классификации и поиска по научным публикациям [4].

Для каждой из моделей была построена отдельная BERTopic-модель с одинаковыми параметрами кластеризации и векторизации. Входными данными служили англоязычные тексты, составленные из заголовков, аннотаций и ключевых слов каждой публикации. Для оценки качества кластеризации рассчитывались следующие метрики из пакета sklearn.metrics:

- Silhouette Score измеряет плотность кластеров по расстоянию между элементами, где $a(x_i)$ — среднее расстояние до всех точек того же кластера, $b(x_i)$ — минимальное среднее расстояние до других кластеров:

$$s(x_i) = \frac{b(x_i) - a(x_i)}{\max\{a(x_i), b(x_i)\}}.$$

- Calinski-Harabasz Index. Соотношение межкластерной и внутрикластерной дисперсии, где

n — количество точек, k — количество кластеров, c_i — центроид кластера C_i :

$$CH = \frac{BCSS / (k-1)}{WCSS / (n-k)},$$

$$BCSS = \sum_{i=1}^k n_i \|c_i - c\|^2,$$

$$WCSS = \sum_{i=1}^k \sum_{x \in C_i} \|x - c_i\|^2.$$

- Davies-Bouldin Score измеряет соотношение расстояний внутри кластера к расстояниям между кластерами, показывая, насколько хорошо кластеры разделены и компактны. Более низкие значения DBI означают лучшее качество кластеризации, т.е. кластеры более различимы и тесно группируются.

Также отслеживались временные характеристики: время загрузки модели, обучения, общее время выполнения, число выявленных тем, количество выбросов (outliers). Результаты сохранялись в таблицу и визуализировались.

Благодаря модульной архитектуре, BERTopic позволил сравнить влияние выбора модели эмбедингов на полноту тем, качество кластеров и интерпретируемость ключевых слов. Полученные данные легли в основу выбора оптимального подхода для последующего анализа публикационной активности. Для двух лучших моделей, «all-mpnet-base-v2» и «allenai-specter», был произведен поиск по сетке лучших параметров.

Анализ (рис. 4) распределения долей публикаций по десяти ведущим тематикам в 2020–2024 гг. демонстрирует динамику исследовательских интересов.

Значительно вырос интерес к вопросам налогообложения и фискальной политики (с 6,8 до 18,0%), что, вероятно, связано с возросшим вниманием к макроэкономической стабильности в условиях внешних вызовов.

Тематика цифровизации высшего образования сохраняет значимость: после пикового значения в 2022 г. (17,0%) ее доля остается высокой в 2024 г. (13,5%), что указывает на институционализацию цифровых практик в академической среде.

Публикации по теме политико-экономических институтов и коррупции стабильно занимают 10–13% от общего объема, что отражает устойчивую значимость институциональной проблематики.

Одновременно фиксируется снижение интереса к возобновляемой энергетике и экологии (с 19,9 до 7,5%), а также к цифровой экономике (с 17,0 до 10,5%), вероятно, вследствие насыщения указанных направлений.

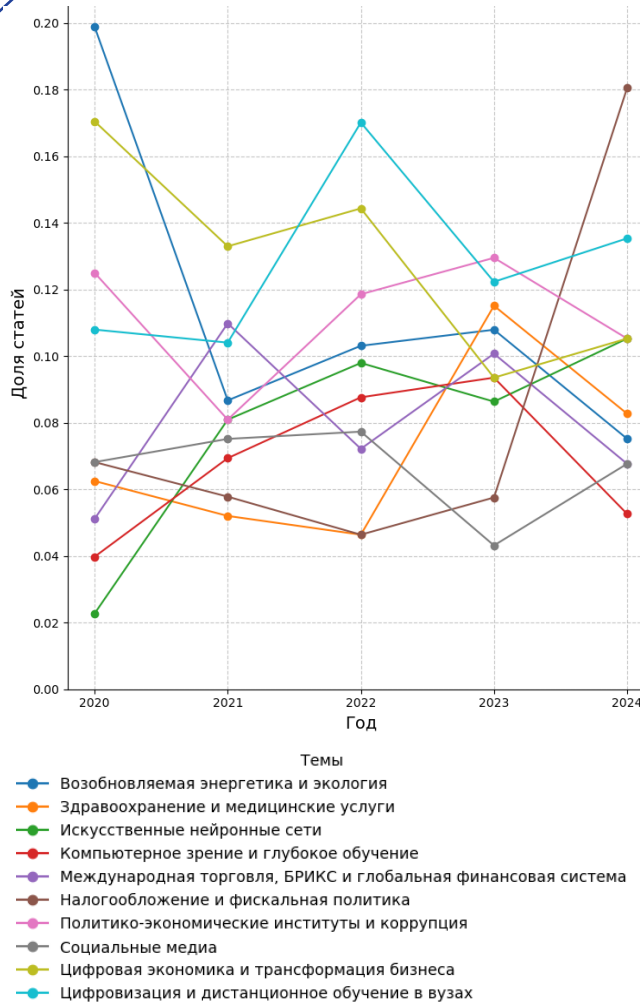


Рис. 4 / Fig. 4. Доля публикаций по темам по годам /
Topic Distribution of Publications per Years

Источник / Source: составлено авторами / Compiled by the authors.

Умеренный стабильный рост доли работ по искусственным нейронным сетям и компьютерному зрению отражает сохраняющееся внимание к технологиям ИИ и их применению.

ВЫЯВЛЕНИЕ НАУЧНЫХ СООБЩЕСТВ

Одним из ключевых направлений сетевого анализа научной активности является идентификация научных сообществ — устойчивых групп исследователей, объединенных совместной публикационной деятельностью. В рамках настоящего исследования был построен граф соавторства на основе англоязычных метаданных статей, где вершины соответствуют авторам, а ребра — фактам соавторства.

После нормализации данных об аффилиации и авторах, а также устранения дублирующих записей был выделен подграф, соответствующий исследователям Финансового университета. В итоговую выборку вошли 579 авторов и 1196 связей. Для

повышения надежности анализа из рассмотрения были исключены узлы с числом публикаций менее двух, что позволило сосредоточиться на активно вовлеченных участниках научного взаимодействия.

Для детекции сообществ использовался алгоритм Louvain, оптимизирующий модулярность и эффективно выявляющий плотные кластеры в больших графах. В результате были обнаружены 67 сообществ с высокой степенью разделения (модулярность $Q = 0,85$), что указывает на выраженную кластерную структуру академических связей в рамках университета.

Визуализация графа соавторства, выполненная с использованием алгоритма Fruchterman-Reingold. Для каждого из них были построены выпуклые оболочки, что позволило наглядно отразить границы взаимодействий между различными научными школами.

Наиболее связными узлами графа стали исследователи с высокой публикационной активностью. В первую десятку по степени центральности вошли: D. Morkovkin, M. Kosov, D. Niyazbekova, S. Korchagin, P. Nikitin, V. Ponkratov, S. Frumina, E.P. Ieshakova, E. Fedorova и E. Fedchenko.

Полученные результаты подтверждают наличие устойчивых научных сообществ внутри Финансового университета и могут быть использованы для последующего анализа структуры академического взаимодействия и развития междисциплинарных связей.

Анализ выявленных сообществ показал наличие как крупных кластеров (численностью от 30 до 55 участников), так и мелких групп (до 5 участников). Крупные сообщества характеризуются умеренной или высокой плотностью связей, наличием устойчивых коллаборационных связей и выраженных центральных акторов, обладающих значительным числом публикаций и высоким значением степени. Такие авторы играют ключевую роль в формировании и поддержании связности внутри сообщества.

Были обнаружены мелкие высокосвязанные кластеры, включающие от 3 до 10 участников, с плотностью, близкой к полной. Эти сообщества отражают локальные коллаборации, зачастую строящиеся вокруг одного ведущего автора и ограниченного круга соавторов. Средняя степень узлов в таких группах составляет от 4 до 6, что указывает на устойчивую совместную публикационную активность.

Структура сети соавторств демонстрирует типичную для научных сообществ организацию, включающую ограниченное число крупных коллаборативных кластеров, значительное количество мелких устойчивых групп, а также периферийные или изолирован-

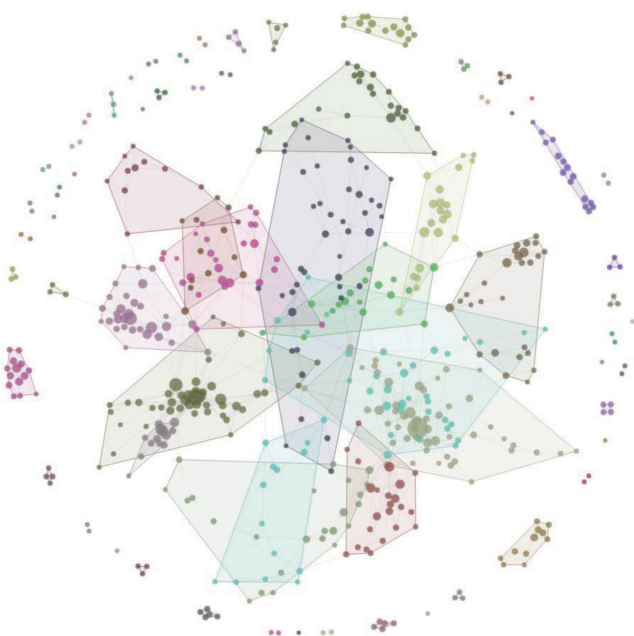


Рис. 5 / Fig. 5. Граф научных сообществ по публикационной активности / Graph of Scientific Communities Based on Publication Activity

Источник / Source: составлено авторами / Compiled by the authors.

ные компоненты. Подобная морфология указывает на наличие как сформировавшихся научных школ, так и эпизодических форм сотрудничества.

Проведенный тематико-сетевой анализ публикационной активности научно-педагогических работников Финансового университета при Правительстве

РФ в 2020–2024 гг. позволил получить представление о динамике научных интересов и структуре внутри-университетского академического взаимодействия.

ВЫВОДЫ

Тематическое моделирование с использованием BERTopic выявило устойчивые исследовательские направления, соответствующие как внутренним академическим приоритетам, так и внешним вызовам. Наибольший рост зафиксирован по темам, связанным с фискальной политикой, цифровизацией образования и применением технологий искусственного интеллекта.

Анализ графа соавторства позволил выявить выраженную кластерную структуру научных взаимодействий. Были обнаружены как крупные сообщества, характеризующиеся устойчивыми связями и высокими значениями степени центральности, так и малые высокосвязанные группы, отражающие локальные формы научного сотрудничества. Высокое значение модулярности подтверждает сегментированность сети и наличие очерченных научных школ. Полученные результаты указывают на наличие сбалансированной структуры академического взаимодействия, сочетающей крупные коллаборативные центры, малые устойчивые группы и периферийные компоненты.

Реализованный подход, основанный на сочетании тематического моделирования и сетевого анализа, продемонстрировал высокую информативность.

СПИСОК ИСТОЧНИКОВ

1. Девицын И.Н., Савин И.В. Инструмент анализа научных сообществ на основе метода моделирования тем и теории графов. *Успехи кибернетики*. 2020;1(4):13–21. DOI: 10.51790/2712-9942-2020-1-4-2
2. Cohan A., Feldman S., Beltagy I., Downey D., Weld D.S. SPECTER: Document-level representation learning using citation-informed transformers. *arXiv preprint*. Cornell University. 2020;arXiv:2004.07180. URL: <https://arxiv.org/abs/2004.07180> DOI: 10.48550/arXiv.2004.07180
3. Wang L., Yang N., Huang X., Yang L., Majumder R., Wei F. Multilingual E 5 Text Embeddings: A Technical Report. *arXiv preprint*. Cornell University. 2024;arXiv:2402.05672. URL: <https://arxiv.org/abs/2402.05672> DOI: 10.48550/arXiv.2402.05672
4. Li X., Li J. AnglE-optimized Text Embeddings. *arXiv preprint*. Cornell University. 2023;arXiv:2309.12871. URL: <https://arxiv.org/abs/2309.12871> DOI: 10.48550/arXiv.2309.12871

REFERENCES

1. Devitsyn I.N., Savin I.V. A tool for the analysis of scientific communities based on topic modeling and graph theory. *Uspekhi kibernetiki*. 2020;1(4):13–21. (In Russ.). DOI: 10.51790/2712-9942-2020-1-4-2
2. Cohan A., Feldman S., Beltagy I., Downey D., Weld D.S. SPECTER: Document-level representation learning using citation-informed transformers. *arXiv preprint*. Cornell University. 2020;arXiv:2004.07180. URL: <https://arxiv.org/abs/2004.07180> DOI: 10.48550/arXiv.2004.07180
3. Wang L., Yang N., Huang X., Yang L., Majumder R., Wei F. Multilingual E 5 Text Embeddings: A Technical Report. *arXiv preprint*. Cornell University. 2024;arXiv:2402.05672. URL: <https://arxiv.org/abs/2402.05672> DOI: 10.48550/arXiv.2402.05672
4. Li X., Li J. AnglE-optimized Text Embeddings. *arXiv preprint*. Cornell University. 2023;arXiv:2309.12871. URL: <https://arxiv.org/abs/2309.12871> DOI: 10.48550/arXiv.2309.12871



ИНФОРМАЦИЯ ОБ АВТОРАХ / ABOUT THE AUTHORS

Григорий Александрович Остапенко — доктор технических наук, профессор, проректор по цифровизации, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

Grigory A. Ostapenko — Dr. Sci. (Tech.), Prof., Vice-Rector for Digitalization, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

<https://orcid.org/0000-0002-1015-5407>

ostg@fa.ru

Галина Геннадьевна Рожкова — студентка бакалавриата, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

Galina G. Rozhkova — undergraduate student, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

<https://orcid.org/0009-0008-9111-9539>

220941@edu.fa.ru

Вадим Геннадьевич Феклин — кандидат физико-математических наук, доцент, декан факультета информационных технологий и анализа больших данных, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

Vadim G. Feklin — Cand. Sci. (Phys.-Math.), Assoc. Prof., Dean of the Faculty of Information Technology and Big Data Analysis, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

<https://orcid.org/0000-0002-1803-6699>

vfeklin@fa.ru

Расул Ахматович Кочкаров — кандидат экономических наук, доцент кафедры искусственного интеллекта факультета информационных технологий и анализа больших данных, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

Rasul A. Kochkarov — Cand. Sci. (Econ.), Assoc. Prof. of the Department of Artificial Intelligence, Faculty of Information Technology and Big Data Analysis, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

<https://orcid.org/0000-0003-3186-3901>

Автор для корреспонденции / Corresponding author:

rkochkarov@fa.ru

Заявленный вклад авторов:

Г.А. Остапенко — разработка общей концепции статьи, введение и выводы.

Г.Г. Рожкова — анализ публикаций, построение графиков, таблиц и диаграмм.

В.Г. Феклин — аннотация, введение, выводы и список литературы.

Р.А. Кочкаров — верификация данных и выводов, обзор и список литературы.

Authors' declared contributions:

G. A. Ostapenko — development of the general concept of the article, introduction, and conclusions.

G. G. Rozhkova — analysis of publications, construction of graphs, tables, and diagrams.

V. G. Feklin — abstract, introduction, conclusions, and list of references.

R. A. Kochkarov — verification of data and conclusions, review, and list of references.

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Conflicts of Interest Statement: The authors have no conflicts of interest to declare.

Статья поступила 09.07.2025; принята к публикации 10.08.2025.

Авторы прочитали и одобрили окончательный вариант рукописи.

The article was received 09.07.2025; accepted for publication 10.08.2025.

The authors read and approved the final version of the manuscript.

Кластерный анализ регионов Российской Федерации по спросу на транспортные услуги

Д.З. Каган, А.А. Рылов

Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

АННОТАЦИЯ

В статье представлено применение кластерного анализа — одного из наиболее распространенных методов машинного обучения — для исследования уровня развития регионального транспорта. Основное внимание уделяется сегментации регионов по существующему спросу на транспортные услуги; определению ключевых компонентов транспортной системы для различных регионов. **Цель исследования** заключается в разделении регионов страны на кластеры, однородные по основным параметрам спроса на транспортные услуги. В каждом кластере объединяются регионы со схожими экономическими, географическими и хозяйственными характеристиками, что определяет схожесть в наиболее востребованных видах транспорта и объектах транспортной инфраструктуры. **Методология исследования** основана на алгоритмах машинного обучения и применении математических метрик к статистическим данным. Исследование включает: отбор значимых факторов; анализ и нормализацию статистических данных; различные методы кластеризации. При нормализации данные приводятся к единой шкале от 0 до 100 баллов с исключением выбросов. В результате кластерного анализа регионы распределяются по четырем основным кластерам. **Техническая реализация** различных вариантов кластерного анализа возможна в табличных редакторах и статистических пакетах. На основе результатов кластеризации проводится интерпретация каждого кластера и выявляются общие характеристики регионов внутри них. **Перспективы исследования** включают его ежегодную актуализацию на основе обновляемых статистических данных. Результаты могут быть использованы для: анализа и развития транспортной системы страны; определения приоритетов в развитии транспортной инфраструктуры регионов; оценки значимости и необходимости реализации региональных транспортных проектов.

Ключевые слова: кластерный анализ; транспортные услуги; метод К-средних; нормализация данных; регионы; транспортная инфраструктура

Для цитирования: Каган Д.З., Рылов А.А. Кластерный анализ регионов Российской Федерации по спросу на транспортные услуги. *Цифровые решения и технологии искусственного интеллекта*. 2025;1(3):77-88. DOI: 10.26794/3033-7097-2025-1-3-77-88

ORIGINAL PAPER

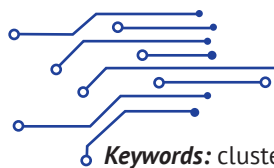
Cluster Analysis Russian Federation Regions by Demand for Transport Services

D.Z. Kagan, A.A. Rylov

Financial University under the Government of the Russian Federation, Moscow, Russian Federation

ABSTRACT

The article presents the application of cluster analysis, one of the most common machine learning methods, to the study of the level of regional transport development, segmentation of regions by the existing demand for transport services, and determination of the most important components of the transport system for certain regions. The purpose of the study is to divide the regions of the country into clusters that are relatively homogeneous in the main aspects of demand for transport services. Each cluster combines regions with similar economic, geographical and economic characteristics, which determines similarity in the most demanded modes of transport and transport infrastructure facilities. The research methodology is based on machine learning algorithms, the application of mathematical metrics to sets of statistical data. The research includes: selection of significant factors; analysis and normalization of statistical data; various clustering methods. In normalization, the data are converted to a single scale from 0 to 100 points with outliers excluded. As a result of cluster analysis, regions are distributed into four main clusters. Technical implementation of different variants of cluster analysis is possible in tabular editors and statistical packages. Based on the clustering results, each cluster is interpreted and common characteristics of the regions within them are identified. The prospects of the study include its annual updating based on updated statistical data. The results can be used to: analyze and develop the country's transport system; identify priorities in the development of regional transport infrastructure; assess the significance and necessity of regional transport projects.



Keywords: cluster analysis; transport services; k-means clustering; data normalization; regions; transport infrastructure

For citation: Kagan D.Z., Rylov A.A. Cluster analysis Russian Federation regions by demand for transport services. *Digital Solutions and Artificial Intelligence Technologies*. 2025;1(3):77-88. DOI: 10.26794/3033-7097-2025-1-3-77-88

ВВЕДЕНИЕ

Транспорт — важнейшая инфраструктурная отрасль экономики страны. Нормальное функционирование и развитие отраслевой инфраструктуры необходимо для полноценной работы других отраслей экономики, обеспечения базовых потребностей жителей. При этом инфраструктура транспортной отрасли должна соответствовать современным требованиям, обеспечивать необходимые пропускные способности, удовлетворять существующий спрос на пассажирские и грузовые перевозки. Транспортная инфраструктура в различных регионах страны демонстрирует значительную неоднородность [1]. Различия наблюдаются и в характере спроса на транспортные услуги [2, 3].

В регионах с высокой плотностью населения и преобладанием городского населения приоритетное значение имеют:

- развитая сеть региональных автомобильных дорог;
- наличие аэропорта с широким спектром направлений полетов.

В северных регионах, характеризующихся низкой плотностью населения и территориями, малопригодными для масштабного дорожного строительства, ключевую роль играет малая авиация. Она обеспечивает как внутрирегиональную связь между населенными пунктами, так и транспортную доступность региона в масштабах всей страны.

Особенности регионального спроса на транспортные услуги играют ключевую роль:

- в определении приоритетов развития транспортной инфраструктуры;
- при отборе приоритетных региональных транспортных проектов;
- в оценке распределения инвестиций в поддержание и развитие транспортной отрасли.

Современные математические методы и методы машинного обучения находят широкое применение в различных областях науки, экономической деятельности [4, 5]. Применение машинного обучения и эконометрических методов при анализе и прогнозировании показателей региональных экономик, а также транспортных показателей уже рассматривается в различных научных работах [6–8]. В данной статье кластерный анализ — один из наиболее эффективных методов машинного обучения — применяется для комплексного исследова-

ния развития транспортной отрасли и оценки состояния транспортной инфраструктуры в различных регионах [9, 10].

Основное применение кластерного анализа в текущем исследовании заключается в моделировании и сегментации существующего спроса на транспортные услуги по регионам страны. На следующем этапе исследования планируется расширение анализа путем распределения регионов по кластерам в соответствии с уровнем развития транспортной системы и обеспеченностью необходимой инфраструктурой.

Цель исследования — провести группировку регионов по характеру спроса на транспортные услуги, выявить наиболее востребованные виды транспорта и объекты транспортной инфраструктуры в каждом конкретном регионе. Такой подход позволяет получить детальное представление о региональных особенностях транспортной системы и определить приоритетные направления ее развития.

В нашем исследовании кластерный анализ регионов преимущественно проводится с учетом факторов, связанных с пассажирскими перевозками. Тем не менее применяемые методы также эффективны для анализа спроса на транспортные услуги в сфере грузовых перевозок. Для адаптации метода оценки под грузоперевозки необходимо скорректировать набор анализируемых показателей и использовать соответствующий массив статистических данных.

В ходе исследования регионы Российской Федерации разделены на 4 кластера по спросу на транспортные услуги. Каждый кластер представляет собой относительно однородную группу регионов с схожими характеристиками транспортного спроса.

Методы кластерного анализа — метод К-средних и иерархическая кластеризация — позволили сформировать группы регионов, однородных по следующим параметрам: характер экономики; особенности расселения населения.

Исходные данные для исследования преимущественно включают статистические показатели по субъектам Российской Федерации, публикуемые Федеральной службой государственной статистики (Росстат) в открытых источниках. Технический инструментарий исследования включает MS Excel, статистический пакет SPSS. Все расчеты и оценки выполнены с использованием указанных программных продуктов.

ЭТАПЫ АНАЛИЗА, ВЫБРАННЫЕ ФАКТОРЫ И ИХ НОРМАЛИЗАЦИЯ

Проводимый анализ разделен на этапы:

- 1) отбор факторов — показателей, по которым будет производиться анализ и сегментация регионов страны;
- 2) сбор статистических данных;
- 3) формализация данных — приведение к единой шкале;
- 4) итерационный процесс разбиения регионов на кластеры — основное содержание собственно кластерного анализа;
- 5) интерпретация полученных результатов.

Для исследования спроса на транспортные услуги были выбраны наиболее важные экономические, демографические и производственные факторы, определяющие характер регионального рынка транспортных услуг. В качестве наиболее важных региональных факторов рассматривались:

X_1 — ВРП (валовый региональный продукт) на душу населения — основной показатель уровня экономического развития региона;

X_2 — объем экспорта на душу населения;

X_3 — среднедушевые денежные доходы населения — главный фактор платежеспособного спроса на услуги транспортной отрасли;

X_4 — доля городского населения в общей численности населения;

X_5 — плотность населения в регионе;

X_6 — доля в произведенной продукции добычи полезных ископаемых;

X_7 — доля в произведенной продукции сельского хозяйства.

Таким образом, учтены факторы, генерирующие основные пассажирские и грузовые потоки. Очевидно, что такие факторы, как плотность населения и доля городского населения, являются ключевыми для спроса на пассажирские перевозки. Основные экономические факторы, приведенные на душу населения, характеризуют платежеспособный спрос на услуги транспорта, а также потенциальную активность населения. Факторы, характеризующие структуру производства в регионах — доля сельского хозяйства и добывающей промышленности — имеют значение для широты географии поездок и их частоты, как внутри, так и за пределами области, а кроме того отражают особенности региональной географии и экономического развития.

Статистические данные по субъектам страны отбираются и рассчитываются на основе данных Росстата за определенный год. Возможны ежегодный пересчет и актуализация распределения регионов на кластеры по мере обновления данных Росстата.

В рамках данного анализа не рассматривались регионы — города федерального значения — Москва, Санкт-Петербург и Севастополь. Для этих регионов, очевидно, не подходит общий алгоритм отбора факторов и распределения на кластеры. Можно считать, что эти три региона образуют отдельный кластер с отличными от других регионов особенностями спроса на транспортные услуги (главную роль для этих регионов играют развитая сеть городского транспорта, наличие современных аэропортов, железнодорожного сообщения с другими регионами).

Для проведения кластерного анализа, в силу несопоставимости абсолютных значений разных признаков, необходимо нормировать все величины так, чтобы диапазон изменений был одинаковым. Также в рамках нормирования должны быть нивелированы «выбросы» — аномально высокие или низкие показатели по какому-либо фактору. Значения всех факторов были приведены к единой шкале — от 0 до 100 баллов. Преобразование значений каждого из факторов X_i проводилось по формуле:

$$X_{ij} = \frac{x_{ij} - x_j^{\min,3}}{x_j^{\max,3} - x_j^{\min,3}},$$

где i — номер субъекта РФ; j — номер признака; X_{ij} — значение j -го фактора X_j для i -го региона; x_{ij} — новое рассчитанное нормированное значение фактора X_j для i -го региона; $x_j^{\max,3}$ — третье максимальное значение j -го признака; $x_j^{\min,3}$ — третье минимальное значение j -го признака.

Значения факторов, третьи по возрастанию и убыванию, берутся, чтобы избавиться от «выбросов», если у одного или двух регионов аномально высокие или низкие показатели по какому-либо фактору. После нормирования для каждого фактора X_j три региона с исходными максимальными значениями будут иметь 100 баллов. Также три региона с исходными минимальными показателями получат значение 0 баллов по этому признаку. Значения для остальных регионов располагаются между 0 и 100 баллами пропорционально исходным значениям.

Спорным остается вопрос: следует ли нормировать показатели, которые уже выражены в процентах? В данном случае мы все-таки нормируем их, чтобы по каждому фактору наблюдался полный диапазон от 0 до 100. Тогда каждый фактор будет иметь равное влияние на итоговое разбиение по кластерам.

В табл. 1 представлены значения отобранных факторов спроса на транспортные услуги по каждому региону, полученные в результате нормирования.

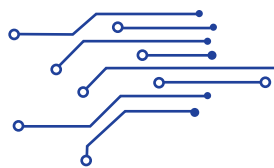
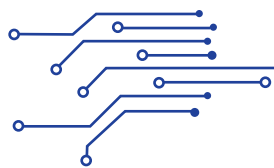


Таблица 1 / Table 1

**Нормированные значения факторов спроса на транспортные услуги /
Normalized Values of Demand Factors for Transport Services**

№	Регион	X_1	X_2	X_3	X_4	X_5	X_6	X_7
1	Алтайский край	6	8	12	28	18	1	95
2	Амурская обл.	18	13	37	56	2	28	55
3	Архангельская обл.	19	57	42	83	3	7	8
4	Астраханская обл.	15	9	16	54	28	52	35
5	Белгородская обл.	26	43	37	56	77	16	96
6	Брянская обл.	7	5	24	63	47	0	88
7	Ненецкий автономный округ	100	0	100	69	0	100	0
8	Ханты-Мансийский автономный округ (Югра)	100	100	83	100	3	100	0
9	Владимирская обл.	10	14	16	84	64	1	28
10	Волгоградская обл.	13	16	10	80	30	7	57
11	Вологодская обл.	22	75	30	69	10	0	18
12	Воронежская обл.	18	15	37	56	60	1	72
13	Еврейская автономная область	11	6	19	59	5	3	42
14	Забайкальский край	8	5	16	58	2	37	24
15	Ивановская обл.	2	4	19	92	64	0	29
16	Иркутская обл.	24	70	14	86	3	47	17
17	Кабардино-Балкарская Республика	0	0	7	18	94	0	100
18	Калининградская обл.	17	38	26	83	88	5	26
19	Калужская обл.	16	21	34	78	46	1	32
20	Камчатский край	35	56	73	83	0	16	14
21	Карачаево-Черкесская Республика	0	1	0	0	44	5	100
22	Кемеровская обл.	15	100	11	100	38	94	19
23	Кировская обл.	6	17	11	79	14	0	40
24	Костромская обл.	9	13	20	68	14	0	39
25	Краснодарский край	18	31	47	24	100	2	59
26	Красноярский край	37	52	32	81	1	30	16
27	Курганская обл.	6	7	9	42	15	2	70

№	Регион	X_1	X_2	X_3	X_4	X_5	X_6	X_7
28	Курская обл.	13	12	25	57	50	18	100
29	Ленинградская обл.	29	82	30	48	28	1	32
30	Липецкая обл.	22	81	34	48	65	1	69
31	Магаданская обл.	62	83	100	100	0	100	5
32	Московская обл.	25=	19	72	93	100	0	8
33	Мурманская обл.	33	100	58	100	6	38	0
34	Нижегородская обл.	16	25	40	87	57	0	19
35	Новгородская обл.	22	47	24	65	14	2	36
36	Новосибирская обл.	18	20	24	86	20	4	25
37	Омская обл.	15	7	24	70	18	1	43
38	Оренбургская обл.	21	33	14	38	21	53	43
39	Орловская обл.	11	7	17	55	41	0	100
40	Пензенская обл.	9	4	13	59	41	0	73
41	Пермский край	22	49	34	77	22	29	11
42	Приморский край	20	35	46	81	15	3	16
43	Псковская обл.	5	5	14	65	15	1	65
44	Республика Адыгея	3	1	19	6	79	3	75
45	Республика Алтай	4	4	0	0	2	4	87
46	Республика Башкортостан	16	43	33	43	38	19	36
47	Республика Бурятия	5	30	23	35	3	12	23
48	Республика Дагестан	3	0	33	0	82	1	58
49	Республика Ингушетия	0	0	0	0	100	3	45
50	Республика Калмыкия	2	0	0	1	4	5	100
51	Республика Карелия	17	33	25	89	4	30	6
52	Республика Коми	41	30	43	83	2	68	5
53	Республика Крым	0	0	1	15	99	6	78
54	Республика Марий Эл	8	9	3	52	39	0	76
55	Республика Мордовия	8	6	0	44	41	0	91



Окончание таблицы 1 / Table 1 (continued)

№	Регион	X_1	X_2	X_3	X_4	X_5	X_6	X_7
56	Республика Саха (Якутия)	56	100	67	52	0	100	8
57	Республика Северная Осетия – Алания	3	2	14	48	100	0	55
58	Республика Татарстан	29	73	47	80	77	29	36
59	Республика Тыва	0	5	0	23	2	38	37
60	Республика Хакасия	15	88	11	61	11	29	26
61	Ростовская обл.	12	40	29	58	56	2	69
62	Рязанская обл.	12	10	22	67	38	0	48
63	Самарская обл.	21	36	29	89	81	23	22
64	Саратовская обл.	9	14	5	77	33	5	73
65	Сахалинская обл.	100	100	100	93	7	87	3
66	Свердловская обл.	23	50	55	100	30	4	11
67	Смоленская обл.	11	30	21	68	25	1	26
68	Ставропольский край	6	9	15	34	57	2	99
69	Тамбовская обл.	16	3	27	39	40	0	99
70	Тверская обл.	10	4	19	77	20	0	28
71	Томская обл.	26	6	21	69	4	39	17
72	Тульская обл.	15	59	30	75	79	1	36
73	Тюменская обл. без автономных округов	41	42	32	53	12	23	22
74	Удмуртская Республика	16	10	19	52	48	39	40
75	Ульяновская обл.	8	10	15	76	45	4	38
76	Хабаровский край	25	35	62	94	1	15	10
77	Челябинская обл.	17	34	18	95	53	7	30
78	Чеченская Республика	0	0	16	0	100	4	38
79	Чувашская Республика	5	3	1	43	91	0	50
80	Чукотский автономный округ	100	59	100	63	0	100	0
81	Ямало-Ненецкий автономный округ	100	96	100	98	0	100	0
82	Ярославская обл.	17	15	32	93	47	0	22

Источник / Source: составлено авторами / Compiled by the authors.

РАСПРЕДЕЛЕНИЕ РЕГИОНОВ ПО КЛАСТЕРАМ, ОСНОВНЫЕ ХАРАКТЕРИСТИКИ

К нормированным данным применены методы кластерного анализа. Кластеризация с разбиением всех рассматриваемых регионов на четыре кластера была проведена двумя методами — K -средних и иерархической кластеризации.

В основе метода K -средних лежит итерационный процесс с минимизацией сумм расстояний между элементами и центроидами кластеров, к которым принадлежат элементы. Центроиды кластеров — средние точки кластеров, вокруг которых собираются рассматриваемые элементы. Элементами в данном случае являются регионы. В начале итерационного процесса произвольным образом определяются координаты (т.е. значения факторов X_1, \dots, X_7) четырех центроидов. В качестве начальных

координат можно выбрать, например, значения факторов по любым четырем регионам. Далее для каждого региона производится расчет четырех расстояний между регионом и всеми центроидами. Для каждого региона определяется центроид с минимальным расстоянием. Соответственно, регион относится к кластеру, определяемому этим центроидом.

В процессе исследования рассчитаны расстояния по наиболее распространенным вариантам метрик — прежде всего использовались евклидова метрика и так называемая метрика «городских кварталов». Стандартная евклидова метрика предусматривает вычисление расстояний между элементами (в данном случае, между регионами и центроидами) по привычной формуле, как корень из суммы квадратов разницы значений координат (в данном случае разницы значений по семи показателям):

$$R(a, b) = \sqrt{\sum_{i=1}^7 (X_i^a - X_i^b)^2} = (X_1^a - X_1^b)^2 + \dots + (X_7^a - X_7^b)^2.$$

В рамках метрики городских кварталов расстояние между регионами определяется как сумма модулей разностей координат. Соответствующая формула имеет следующий вид:

$$R(a, b) = \sum_{i=1}^7 |X_i^a - X_i^b|.$$

После расчета расстояний каждый регион относится к одному из четырех кластеров: к тому, для которого расстояние между регионом и центроидом является минимальным. Далее рассчитываются новые значения координат центроидов, как средние значения показателей по регионам, попавшим в кластер. На следующей итерации происходит перерасчет расстояний между всеми регионами и новыми центроидами; для каждого региона определяется новый кластер по минимальным расстояниям. Алгоритм завершается за конечное число итераций, когда после очередной итерации не происходит никаких изменений в центроидах — соответственно, не изменяется и состав кластеров.

В рамках метода иерархической кластеризации проводилось поэтапное объединение мелких кластеров. Сначала каждый регион определен как отдельный кластер. Далее пары наиболее близких по значениям показателей кластеров объединялись в более крупный кластер. В результате применения различных методов кластеризации получились, соответственно, различные результаты разбиения регионов на кластеры, однако между полученными разбиениями оказалось очень много общего, а различия относительно небольшие. Состав кластеров, полученных разными методами с использовани-

ем различных метрик, практически идентичен. Следовательно, результаты проведенного анализа обладают определенной «устойчивостью» и отражают реальный уровень регионального спроса на транспортные услуги.

Сформированные в результате проведенного анализа четыре группы (кластера) регионов обладают внутренними общими характеристиками (внутренней однородностью); полученное разбиение регионов на группы имеет достаточно четкую интерпретацию. На конечном шаге алгоритма метода K -средних получены следующие координаты центроидов, т.е. значения показателей центральных точек, вокруг которых сгруппированы четыре региональных кластера (табл. 2).

Итоговые координаты центроидов равны средним значениям показателей по всем регионам, вошедшим в соответствующий кластер.

Как видно из табл. 2, кластеру 1 соответствует максимальное по всем кластерам значение показателя X_5 — плотности населения в регионах; также высокое значение показателя X_7 — доли сельского хозяйства в произведенной продукции. Значения остальных показателей для первого кластера сравнительно невелики. Таким образом, в кластер 1 вошли прежде всего регионы с высокой плотностью



населения и важной ролью сельского хозяйства в региональной экономике. Кластеру 2 соответствуют минимальные средние по входящим регионам значения экономических факторов, в том числе среднедушевых доходов населения.

Средние значения регионов из кластера 3 не являются максимальными или минимальными ни по одному фактору. Однако нужно отметить, что в кластере 3 высокое среднее значение показателя X_4 — доля городского населения. Средние значения экономических показателей ниже, чем в кластере 4, но существенно выше кластеров 1 и 2.

Кластеру 1 соответствуют максимальные значения показателей $X_1, X_2, X_3, X_4, X_5, X_6$. Таким образом, для регионов первого кластера характерны максимальные значения ВРП, доходов на душу населения, доли городского населения и доли добычи полезных ископаемых.

Основной результат проведенного кластерного анализа — разбиение регионов страны на четыре кластера представлен в табл. 3.

ВЫВОДЫ

После фиксации результатов исследования крайне важна их интерпретация. Необходимо определить какие особенности соответствуют экономическим и географическим характеристикам регионов каждого кластера и, соответственно, какие транспортные услуги, объекты транспортной инфраструктуры являются наиболее необходимыми.

Сформированные в результате проведенного анализа четыре группы — кластера регионов обладают внутренними общими характеристиками (внутренней однородностью); полученное разбиение

регионов на группы имеет достаточно четкую интерпретацию.

1. Кластер 1 объединяет густонаселенные регионы. В этот кластер входят 13 регионов с очень высокой плотностью населения. Экономические показатели у этих регионов ниже, чем у регионов из кластеров 3 и 4, но выше кластера 2. Относительно высокая доля сельского хозяйства и низкая — добычи полезных ископаемых. Сюда входят прежде всего южные регионы — Краснодарский край, Ростовская обл., большинство республик Кавказа. Отметим, что в этот кластер входят и относительно богатые регионы — уже упомянутые Ростовская обл. и Краснодарский край.

2. Кластер 2 состоит из небогатых регионов с высокой долей сельского хозяйства. Объединяет 16 регионов из различных федеральных округов. У данного кластера самые низкие экономические показатели — ВРП на душу населения, экспорт, среднедушевые доходы. Относительно низкая доля городского населения и самая высокая доля сельского хозяйства.

3. Кластер 3 характеризуется средними значениями всех показателей, кроме доли городского населения, которая в этом кластере близка к максимальной. Данный кластер объединяет немногим более половины регионов — 44 страны (северо-западные, уральские, сибирские и дальневосточные регионы с большой площадью и достаточно развитым промышленным производством или добычей природных ископаемых; центральные регионы, имеющая высокие экономические показатели густонаселенная Московская обл.; регионы с менее благополучной экономикой — Владимирская или

Таблица 2 / Table 2

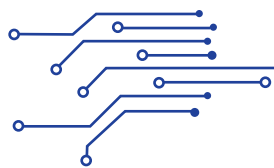
Средние значения показателей регионов по кластерам / Average values of regional indicators by clusters

Фактор	Центроид			
	1	2	3	4
X_1	9	7	19	74
X_2	17	6	32	82
X_3	21	10	29	80
X_4	29	40	74	86
X_5	85	31	29	6
X_6	3	5	15	91
X_7	66	85	27	4

Источник / Source: составлено авторами / Compiled by the authors.

**Распределение регионов страны по кластерам по спросу на транспортные услуги / Regions
Distribution Into Clusters by Demand for Transport Services**

Кластер			
1	2	3	4
Белгородская обл.	Брянская обл.	Владимирская обл.	Ненецкий автономный округ
Воронежская обл.	Курская обл.	Ивановская обл.	Мурманская обл.
Липецкая обл.	Орловская обл.	Калужская обл.	Ханты-Мансийский автономный округ (Югра)
Республика Адыгея	Тамбовская обл.	Костромская обл.	Ямало-Ненецкий автономный округ
Республика Крым	Псковская обл.	Московская обл.	Кемеровская обл.
Краснодарский край	Республика Калмыкия	Рязанская обл.	Республика Саха (Якутия)
Ростовская обл.	Карачаево-Черкесская Республика	Смоленская обл.	Магаданская обл.
Республика Дагестан	Ставропольский край	Тверская обл.	Сахалинская обл.
Республика Ингушетия	Республика Марий Эл	Тульская обл.	Чукотский автономный округ
Кабардино-Балкарская Республика	Республика Мордовия	Ярославская обл.	—
Республика Северная Осетия – Алания	Пензенская обл.	Республика Карелия	—
Чеченская Республика	Саратовская обл.	Республика Коми	—
Чувашская Республика	Курганская обл.	Архангельская обл. без автономного округа	—
—	Республика Алтай	Вологодская обл.	—
—	Республика Тыва	Калининградская обл.	—
—	Алтайский край	Ленинградская обл.	—
—	—	Новгородская обл.	—
—	—	Астраханская обл.	—
—	—	Волгоградская обл.	—
—	—	—	—
—	—	Республика Татарстан	—
—	—	Удмуртская Республика	—
—	—	Пермский край	—
—	—	Кировская обл.	—



Окончание таблицы 3 / Table 3 (continued)

Кластер			
1	2	3	4
—	—	Нижегородская обл.	—
—	—	Оренбургская обл.	—
—	—	Самарская обл.	—
—	—	Ульяновская обл.	—
—	—	Свердловская обл.	—
—	—	Тюменская обл. без автономных округов	—
—	—	Челябинская обл.	—
—	—	Республика Бурятия	—
—	—	Республика Хакасия	—
—	—	Забайкальский край	—
—	—	Красноярский край	—
—	—	Иркутская обл.	—
—	—	Новосибирская обл.	—
—	—	Омская обл.	—
—	—	Томская обл.	—
—	—	Камчатский край	—
—	—	Приморский край	—
—	—	Хабаровский край	—
—	—	Амурская обл.	—
—	—	Еврейская автономная обл.	—

Источник / Source: составлено авторами / Compiled by the authors.

Ивановская обл.). В некоторых вариантах анализа от этого кластера отделяется часть регионов и переходит в кластеры 1 и 3. Отметим, что данный кластер является наименее однородным из 4 полученных. Вероятно, есть целесообразность рассмотреть вариант разделения данного кластера на 2 меньших.

4. Кластер 4 — кластер добывающих регионов с низкой плотностью населения — объединяет 9 регионов, в основном очень богатых и малозаселенных, таких как Югра или Ненецкий автономный округ, главная отрасль которых — добыча природных ископаемых. Средние экономические показатели

этих регионов превосходят все остальные. Для регионов кластера 4 не требуется мощная разветвленная сеть автомобильных или железных дорог в силу малой и неоднородной заселенности их территорий. Для таких регионов наиболее важно наличие сети небольших аэропортов и, возможно, небольших дорог, связывающих удаленные населенные пункты.

Проведенный анализ можно ежегодно актуализировать на основе обновленных годовых данных Росстата. Также данное исследование может быть обогащено данными с добавлением новых факторов и уточнением важных элементов для определения

характеристик регионального спроса на транспортные услуги; возможно увеличение числа кластеров. Интересно и использование различных методов кластеризации, наложение их результатов и выбор итогового состава кластеров с учетом экспертной оценки адекватности каждого полученного варианта распределения.

Результаты настоящего исследования регионального спроса на транспортные услуги (в сочетании с исследованиями существующего уровня транспортной обеспеченности регионов) могут быть использованы при распределении приоритетов в развитии транспортной инфраструктуры регионов и оценке важности реализации транспортных проектов.

СПИСОК ИСТОЧНИКОВ

1. Бардаль А.Б. Спрос на услуги транспортного комплекса региона: экономические факторы грузовых перевозок. *Вестник Тихоокеанского государственного университета*. 2021;63(4):129–138. URL: <https://www.elibrary.ru/iovazb>
2. Капелюк З.А., Попова Я.В. Показатели экономической эффективности транспортных услуг. *Экономика. Бизнес. Банки*. 2022;66(4):8–18. URL: <https://www.elibrary.ru/bxuuet>
3. Иванова Е.А., Соколов Ю.И., Чуверина О.Г. Изучение модели поведения пассажиров для формирования спроса на услуги пассажирского комплекса в городских агломерациях. *Транспорт: наука, техника, управление. Научный информационный сборник*. 2022;(10):47–54. DOI: 10.36535/0236–1914–2022–10–8
4. Афанасьев М.Ю., Гусев А.А. Интегральный индекс структурной сложности региональных экономик. *Экономика и математические методы*. 2025;61(2):57–74. DOI: 10.31857/S 0424738825020054
5. Афанасьев М.Ю., Гусев А.А. Ситуационное моделирование траекторий экономической сложности регионов. *Экономика и математические методы*. 2023;59(4):58–70. DOI: 10.31857/S 042473880028217–7
6. Голубцова П.С., Федотова В.О., Руднев С.Г. Методы кластерного анализа в изучении региональных экономик. *Прикладные экономические исследования*. 2023;(51):99–107. URL: <https://www.elibrary.ru/elhebo> DOI: 10.47576/2949–1908_2023_S 1_99
7. Фроловичев А.И., Ишханян М.В. Факторы роста валового регионального продукта субъектов Российской Федерации. *Транспортное дело России*. 2023;(1):53–57. DOI: 10.52375/20728689_2023_1_53
8. Шаталова О.М., Касаткина Е.В., Лившиц В.Н. Экономическая сложность и вложенность структур региональных экономик. *Экономика и математические методы*. 2021;57(3):67–78. DOI: 10.31857/S 042473880016410–0
9. Малкина М.Ю., Плехова Ю.О., Перова В.И., Сочков А.Л. Исследование влияния отраслевой структуры российских регионов на их экономическое развитие с использованием методов искусственного интеллекта. *Экономический анализ: теория и практика*. 2025;24(2):123–143. DOI: 10.24891/ea.24.2.123
10. Шаталова О.М., Касаткина Е.В., Лившиц В.Н. Кластерный анализ и классификация промышленно ориентированных регионов РФ по экономической специализации. *Экономика и математические методы*. 2022;58(1): 80–91. DOI: 10.31857/S 042473880018971–7

REFERENCES

1. Bardal A.B. Demand for the transport complex services of the region: economic factors of freight traffic. *Bulletin of Pacific National university*. 2021;63(4):129–138. URL: <https://www.elibrary.ru/iovazb> (In Russ.).
2. Kapelyuk Z.A., Popova Ya.V. Indicators of economic efficiency of transport services. *Economy. Business. Banks*. 2022;66(4):8–18. URL: <https://www.elibrary.ru/bxuuet> (In Russ.).
3. Ivanova E.A., Sokolov Yu.I., Chuverina O.G. Study of passenger behaviour to generate demand for passenger complex services in urban agglomerations. *Transport: science, equipment, management. Scientific information collection*. 2022;(10):47–54. (In Russ.). DOI: 10.36535/0236–1914–2022–10–8
4. Afanasiev M. Yu., Gusev A.A. Integral structural complexity index of regional economies. *Economics and mathematical methods*. 2025;61(2):57–74. (In Russ.). DOI: 10.31857/S 0424738825020054
5. Afanasiev M. Yu., Gusev A.A. Situational modeling of trajectories of regions' economic complexity. *Economics and mathematical methods*. 2023;59(4):58–70. (In Russ.). DOI: 10.31857/S 042473880028217–7
6. Golubtsova P.S., Fedotova V.O., Rudnev S.G. Methods of cluster analysis in the study of regional economies. *The applied economic researches journal*. 2023;(51):99–107. URL: <https://www.elibrary.ru/elhebo> (In Russ.). DOI: 10.47576/2949–1908_2023_S 1_99
7. Frolovichev A., Ishkhanyan M. Factors of growth of the gross regional product of the subjects of the russian federation. *Transport business of Russia*. 2023;(1):53–57. (In Russ.). DOI: 10.52375/20728689_2023_1_53
8. Afanasiev M., Kudrov A. Economic complexity and embedding of regional economies' structures. *Economics and mathematical methods*. 2021;57(3):67–78. (In Russ.). DOI: 10.31857/S 042473880016410–0



9. Malkina M. Yu., Plekhova Yu.O., Perova V.I., Sochkov A.L. Studying the influence of the sectoral structure of russian regions on their economic development using artificial intelligence methods. *Economic analysis: theory and practice*. 2025;24(2):123–143. (In Russ.). DOI: 10.24891/ea.24.2.123
10. Shatalova O., Kasatkina E., Livchits V. Cluster analysis and classification of Russia's industrial oriented regions by economic specialization. *Economics and mathematical methods*. 2022;58(1):80–91. (In Russ.). DOI: 10.31857/S 042473880018971–7

ИНФОРМАЦИЯ ОБ АВТОРАХ / ABOUT THE AUTHORS

Дмитрий Зиновьевич Каган — кандидат физико-математических наук, доцент кафедры математики и анализа данных факультета информационных технологий и анализа больших данных, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

Dmitry Z. Kagan — Cand. Sci. (Phys.-Math.), Assoc. Prof., Department of Mathematics and Data Analysis, Faculty of Information Technology and Big Data Analysis, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

<https://orcid.org/0009-0004-9352-3990>

Автор для корреспонденции / Corresponding author:

DZKagan@fa.ru; dmikagan@gmail.com

Александр Аркадьевич Рылов — кандидат физико-математических наук, доцент кафедры математики и анализа данных факультета информационных технологий и анализа больших данных, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

Alexander A. Rylov — Cand. Sci. (Phys.-Math.), Assoc. Prof., Department of Mathematics and Data Analysis, Faculty of Information Technology and Big Data Analysis, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

<https://orcid.org/0000-0001-9474-8519>

ARylov@fa.ru

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Conflicts of Interest Statement: The authors have no conflicts of interest to declare.

Статья поступила 30.06.2025; принята к публикации 10.08.2025.

Авторы прочитали и одобрили окончательный вариант рукописи.

The article was received 30.06.2025; accepted for publication 10.08.2025.

The authors read and approved the final version of the manuscript.