

# АКТУАЛЬНЫЕ ИССЛЕДОВАНИЯ

МЕЖДУНАРОДНЫЙ НАУЧНЫЙ ЖУРНАЛ

ISSN 2713-1513

#21 (256), 2025

часть I

# Актуальные исследования

Международный научный журнал

2025 • № 21 (256)

Часть I

Издается с ноября 2019 года

Выходит еженедельно

ISSN 2713-1513

**Главный редактор:** Ткачев Александр Анатольевич, канд. социол. наук

**Ответственный редактор:** Ткачева Екатерина Петровна

Статьи, поступающие в редакцию, рецензируются.  
За достоверность сведений, изложенных в статьях, ответственность несут авторы.  
Мнение редакции может не совпадать с мнением авторов статей.  
При использовании и заимствовании материалов ссылка на издание обязательна.  
Материалы публикуются в авторской редакции.

## РЕДАКЦИОННАЯ КОЛЛЕГИЯ

**Абдуллин Тимур Zufарович**, кандидат технических наук (Высokотехнологический научно-исследовательский институт неорганических материалов имени академика А. А. Бочвара)

**Абидова Гулмира Шухратовна**, доктор технических наук, доцент (Ташкентский государственный транспортный университет)

**Альборад Ахмед Абуди Хусейн**, преподаватель, PhD, Член Иракской Ассоциации спортивных наук (Университет Куфы, Ирак)

**Аль-бутбахак Башшар Абуд Фадхиль**, преподаватель, PhD, Член Иракской Ассоциации спортивных наук (Университет Куфы, Ирак)

**Альхаким Ахмед Кадим Абдуалкарем Мухаммед**, PhD, доцент, Член Иракской Ассоциации спортивных наук (Университет Куфы, Ирак)

**Асаналиев Мелис Казыкеевич**, доктор педагогических наук, профессор, академик МАНПО РФ (Кыргызский государственный технический университет)

**Атаев Загир Вагитович**, кандидат географических наук, проректор по научной работе, профессор, директор НИИ биогеографии и ландшафтной экологии (Дагестанский государственный педагогический университет)

**Бафоев Феруз Муртазоевич**, кандидат политических наук, доцент (Бухарский инженерно-технологический институт)

**Гаврилин Александр Васильевич**, доктор педагогических наук, профессор, Почетный работник образования (Владимирский институт развития образования имени Л.И. Новиковой)

**Галузо Василий Николаевич**, кандидат юридических наук, старший научный сотрудник (Научно-исследовательский институт образования и науки)

**Григорьев Михаил Федосеевич**, доктор сельскохозяйственных наук (Кузбасский государственный аграрный университет имени В.Н. Полецкого)

**Губайдуллина Гаян Нурахметовна**, кандидат педагогических наук, доцент, член-корреспондент Международной Академии педагогического образования (Восточно-Казахстанский государственный университет им. С. Аманжолова)

**Ежкова Нина Сергеевна**, доктор педагогических наук, профессор кафедры психологии и педагогики (Тульский государственный педагогический университет им. Л.Н. Толстого)

**Жилина Наталья Юрьевна**, кандидат юридических наук, доцент (Белгородский государственный национальный исследовательский университет)

**Ильина Екатерина Александровна**, кандидат архитектуры, доцент (Государственный университет по землеустройству)

**Каландаров Азиз Абдурахманович**, PhD по физико-математическим наукам, доцент, проректор по учебным делам (Гулистанский государственный педагогический институт)

**Карпович Виктор Францевич**, кандидат экономических наук, доцент (Белорусский национальный технический университет)

**Кожевников Олег Альбертович**, кандидат юридических наук, доцент, Почетный адвокат России (Уральский государственный юридический университет)

**Колесников Александр Сергеевич**, кандидат технических наук, доцент (Южно-Казахстанский университет им. М. Ауэзова)

**Копалкина Евгения Геннадьевна**, кандидат философских наук, доцент (Иркутский национальный исследовательский технический университет)

**Красовский Андрей Николаевич**, доктор физико-математических наук, профессор, член-корреспондент РАЕН и АИН (Уральский технический институт связи и информатики)

**Кузнецов Игорь Анатольевич**, кандидат медицинских наук, доцент, академик международной академии фундаментального образования (МАФО), доктор медицинских наук РАГПН, профессор, почетный доктор наук РАЕ, член-корр. Российской академии медико-технических наук (РАМТН) (Астраханский государственный технический университет)

**Литвинова Жанна Борисовна**, кандидат педагогических наук (Кубанский государственный университет)

**Мамедова Наталья Александровна**, кандидат экономических наук, доцент (Российский экономический университет им. Г.В. Плеханова)

**Мукий Юлия Викторовна**, кандидат биологических наук, доцент (Санкт-Петербургская академия ветеринарной медицины)

**Никова Марина Александровна**, кандидат социологических наук, доцент (Московский государственный областной университет (МГОУ))

**Насакаева Бакыт Ермакбайкызы**, кандидат экономических наук, доцент, член экспертного Совета МОН РК (Карагандинский государственный технический университет)

**Олешкевич Кирилл Игоревич**, кандидат педагогических наук, доцент (Московский государственный институт культуры)

**Попов Дмитрий Владимирович**, доктор филологических наук (DSc), доцент (Андижанский государственный институт иностранных языков)

**Пятаева Ольга Алексеевна**, кандидат экономических наук, доцент (Российская государственная академия интеллектуальной собственности)

**Редкоус Владимир Михайлович**, доктор юридических наук, профессор (Институт государства и права РАН)

**Самович Александр Леонидович**, доктор исторических наук, доцент (ОО «Белорусское общество архивистов»)

**Сидикова Тахира Далиевна**, PhD, доцент (Ташкентский государственный транспортный университет)

**Таджибоев Шарифджон Гайбуллоевич**, кандидат филологических наук, доцент (Худжандский государственный университет им. академика Бободжона Гафурова)

**Тихомирова Евгения Ивановна**, доктор педагогических наук, профессор, Почётный работник ВПО РФ, академик МААН, академик РАЕ (Самарский государственный социально-педагогический университет)

**Хаитова Олмахон Саидовна**, кандидат исторических наук, доцент, Почетный академик Академии наук «Турон» (Навоийский государственный горный институт)

**Цуриков Александр Николаевич**, кандидат технических наук, доцент (Ростовский государственный университет путей сообщения (РГУПС))

**Чернышев Виктор Петрович**, кандидат педагогических наук, профессор, Заслуженный тренер РФ (Тихоокеанский государственный университет)

**Шаповал Жанна Александровна**, кандидат социологических наук, доцент (Белгородский государственный национальный исследовательский университет)

**Шошин Сергей Владимирович**, кандидат юридических наук, доцент (Саратовский национальный исследовательский государственный университет имени Н.Г. Чернышевского)

**Эшонкулова Нуржахон Абдужабборовна**, PhD по философским наукам, доцент (Навоийский государственный горный институт)

**Яхшиева Зухра Зиятовна**, доктор химических наук, доцент (Джиззакский государственный педагогический институт)

## СОДЕРЖАНИЕ

### НЕФТЯНАЯ ПРОМЫШЛЕННОСТЬ

**Белоусов О.А.**

ОЧИСТКА ПРОМЫШЛЕННО-ДОЖДЕВЫХ СТОЧНЫХ ВОД С ПОМОЩЬЮ  
МОБИЛЬНЫХ ОЧИСТНЫХ СООРУЖЕНИЙ НА БАЗЕ ПРИЦЕПА.....6

### ТЕХНИЧЕСКИЕ НАУКИ

**Jasmine Aziz Hussein**

DESIGN AND STUDY OF MATHEMATICAL ANALYSIS OF WIND TURBINES USING  
MATLAB/SIMULINK.....9

### ГЕОЛОГИЯ

**Санкара Букари**

СРАВНИТЕЛЬНЫЕ ИССЛЕДОВАНИЯ СРЕДНЕАЛЬБСКО-ПАЛЕОЦЕНОВЫХ  
ОТЛОЖЕНИЙ В ВОСТОЧНОЙ ЧАСТИ ОСАДОЧНОГО БАСЕЙНА КОТ-Д'ИВУАРА:  
СЛУЧАЙ СКВАЖИНЫ SE1D .....17

### ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

**Petrova I.A.**

APPLICATION OF THE KANO MODEL IN PRODUCT DEVELOPMENT: CROSS-  
INDUSTRY INSIGHTS AND METHODOLOGICAL APPROACHES .....25

**Khomutinnikov M.**

MICROSERVICES ARCHITECTURE: ACCELERATING FEATURE DEVELOPMENT AND  
SCALABILITY THROUGH MONOLITH DECOMPOSITION .....33

**Абдуллаева С.Г., Сардаров Я.Б.**

МЕТОДЫ ОБНАРУЖЕНИЯ ИНСАЙДЕРСКИХ ПРИЗНАКОВ В БОЛЬШИХ ДАННЫХ 39

**Абдуллаева С.Г., Сардаров Я.Б.**

ПРАКТИЧЕСКАЯ РЕАЛИЗАЦИЯ И ЭФФЕКТИВНОСТЬ МЕТОДОВ ОБНАРУЖЕНИЯ 43

**Волков Н.А.**

РАЗРАБОТКА МЕТОДИКИ ПЕРСОНАЛИЗИРОВАННОГО ОТОБРАЖЕНИЯ ДАННЫХ  
В ПРОГРЕССИВНЫХ ВЕБ-ПРИЛОЖЕНИЯХ .....47

**Закирова Ю.Р., Кантюкова А.Р., Сагитова А.Р.**

КОМБИНАТОРИКА В КРИПТОГРАФИИ И БЕЗОПАСНОСТИ ДАННЫХ .....51

**Ивашенцев А.С.**

КРИВЫЕ БЕЗЬЕ В ВИДЕОИГРАХ .....54

**Кротов Е.Ю.**

ПРИМЕНЕНИЕ МЕТОДОВ ГЛУБОКОГО ОБУЧЕНИЯ ДЛЯ ОБНАРУЖЕНИЯ  
ФИШИНГОВЫХ ВЕБ-САЙТОВ: АНАЛИЗ ЭФФЕКТИВНОСТИ И ОПТИМИЗАЦИЯ  
МОДЕЛЕЙ .....60

<b>Кузмичев А.А., Кузмичев А.А.</b> НАЗЕМНЫЕ МЕТОДЫ ДЕТЕКТИРОВАНИЯ БЕСПИЛОТНЫХ ЛЕТАТЕЛЬНЫХ АППАРАТОВ .....	70
<b>Кулябин И.А., Шиянова В.Д.</b> ПЕРСПЕКТИВЫ ИНТЕГРАЦИИ БИЗНЕС-АНАЛИТИКИ И ИСКУССТВЕННОГО ИНТЕЛЛЕКТА .....	75
<b>Лотыш Н.И.</b> АНАЛИЗ СУЩЕСТВУЮЩЕЙ СТЕПЕНИ ЗАЩИЩЕННОСТИ ИНФРАСТРУКТУРЫ ПРЕДПРИЯТИЯ .....	78
<b>Мугинов Т.И.</b> СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДОВ КЛАСТЕРИЗАЦИИ СОЦИАЛЬНЫХ ГРАФОВ НА ДАННЫХ СОЦИАЛЬНОЙ СЕТИ «ВКОНТАКТЕ» ПО МЕРЕ СХОДСТВА ДВУХ РАЗБИЕНИЙ И МОДУЛЯРНОСТИ .....	83
<b>Оруджова С.Р., Гасангулиева М.М.</b> АНАЛИЗ БОЛЬШИХ ДАННЫХ И СТАТИСТИЧЕСКИХ ПРИМЕНЕНИЙ .....	89
<b>Соколов И.А.</b> АРХИТЕКТУРНЫЕ ПОДХОДЫ К ОБЕСПЕЧЕНИЮ ОТКАЗОУСТОЙЧИВОСТИ И МАСШТАБИРУЕМОСТИ В OPEN-SOURCE СИСТЕМАХ КОНТРОЛЯ ВЕРСИЙ: СРАВНИТЕЛЬНЫЙ АНАЛИЗ GITLAB CE, GITEA И FORGEJO .....	94
<b>Швалев И.Е.</b> ГЕНЕРАЦИЯ ЖЕСТИКУЛЯЦИИ С ПРИМЕНЕНИЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА НА ОСНОВЕ ТЕКСТОВОГО ВВОДА .....	98



# НЕФТЯНАЯ ПРОМЫШЛЕННОСТЬ

**БЕЛОУСОВ Олег Анатольевич**

магистрант, Ульяновский институт гражданской авиации имени Главного маршала авиации Б. П. Бугаева, Россия, г. Ульяновск

## ОЧИСТКА ПРОМЫШЛЕННО-ДОЖДЕВЫХ СТОЧНЫХ ВОД С ПОМОЩЬЮ МОБИЛЬНЫХ ОЧИСТНЫХ СООРУЖЕНИЙ НА БАЗЕ ПРИЦЕПА

**Аннотация.** В настоящее время имеется большое количество предприятий, в том числе и складов ГСМ, на которых очистка дождевых сточных вод отсутствует или не соответствует современным требованиям, что влечет определенные финансовые риски для предприятия и создает угрозу экологии. В данной статье рассматривается установка для очистки дождевых сточных вод не требующая затрат на строительно-монтажные работы и стационарной установки.

**Ключевые слова:** очистка дождевых сточных вод, склад горюче-смазочных материалов (ГСМ), защита окружающей среды.

Главной целью данной статьи является – обеспечение производственных объектов системой очистки промышленно-дождевых стоков и исключить загрязнение грунтов нефтепродуктосодержащей водой.

Для достижения этой цели были поставлены следующие задачи:

1. Разработать компактное решение для очистки промышленно-дождевых сточных вод на предприятии;
2. Провести выбор необходимого для проекта оборудования;
3. Рассчитать экономическую целесообразность будущего проекта.

Для каждого предприятия работающего с нефтепродуктами необходимо произвести расчет водопотребления для используемой установки. Для расчета используется среднемесячный показатель выпадения осадков для региона/населенного пункта, выбирается максимальное значение выпадения осадков на всей площади предприятия, с которой осуществляется сбор дождевых сточных вод. Например: склады хранения нефтепродуктов, резервуарный парк, пункты налива и слива. В соответствии с полученными данными производится выбор необходимой установки очистки сточных вод от нефтепродуктов. Для примера рассмотрим установку достаточную для использования в Московской области на небольшом

складе (резервуарный парк 11 резервуаров РГС-50, пункт слива-налива площадью 80м<sup>2</sup>. Для создания мобильной системы очистки будет достаточно использование наименьшей производительности 1-2 м<sup>3</sup>/час. Флотационно-фильтрационная установка с фильтром сорбционным двухступенчатым для доочистки. Степень очистки сточной воды позволяет осуществлять сброс очищенных сточных вод в ливневую канализацию или в водоем (соответствует Постановлению правительства о водоснабжении и водоотведении № 644).

Флотационно-фильтрационная установка включает в себя следующие процессы очистки [1]:

1. Камера смешения (устанавливается при необходимости). Служит для смешения химических реагентов со сточными водами, образования коагуляции загрязнений с образованием хлопьев и как следствие повышении эффективности очистки воды флотацией.
2. Сатурация. В этом узле происходит растворение под повышенным давлением атмосферного воздуха в очищенной воде для подачи в камеру флотации.
3. Флотация. Предназначено для подъема загрязнений на поверхность флотатора при помощи процесса флотации. Также здесь

происходит удаление загрязнений в шламовый карман в виде флотошлама.

4. Фильтрация. Предназначена для финишной доочистки дождевых сточных вод. В рассматриваемом варианте ФФУ ФСД установлено 3 ступени фильтрации [1].

Очищенная вода соответствует требованиям для сброса в водоем.

Так же для придания установки мобильности необходимо установить очистные сооружения на прицеп. Выбор прицепа производится в зависимости от размеров очистной установки. Под параметры габаритов и грузоподъемности для выбранной установки подходит прицеп ССТ на 2-х осях. На базе данного прицепа и будет произведена установка наших очистных сооружений.

Принцип работы мобильной установки будет заключаться в следующем. Дождевая вода с территорий, технологически задействованных в операциях с нефтепродуктами, по дождеприемникам поступает в аккумулирующую емкость, откуда и будет происходить откачка ее на мобильные очистные сооружения установленным на них насосом. Благодаря подвижности имеется возможность установить очистные сооружения в любом необходимом участке, в связи с чем накопительных емкостей и мест может быть несколько.

Всего 4 этапа [2]:

1. Первичное отстаивание в емкости-накопителе, обеспечивающее удаление минеральных примесей и пленочных нефтепродуктов;

2. Напорную флотацию на установке «ФФУ», удаляющую основное количество нефтепродуктов и взвешенных веществ, и обеспечивающую значительное снижение БПК и ХПК. В состав «ФФУ» входит устройство для дозирования хим. реагентов, позволяющее в несколько раз повысить эффективность очистки и снизить нагрузку на последующие степени очистки;

3. Фильтрацию на встроенном фильтре установки «ФФУ», загруженном пенополиуретановой крошкой или активированным углем, обеспечивающей удаление из воды остаточных взвесей. (данный фильтрующий материал (пенополиуретан) обладает высокой грязеемкостью и подвергается периодической промывке);

4. Сорбционную чистку на фильтре «ФСД», загруженном активированным углем, либо любым эффективным сорбционным материалом. При этом поступление на сорбционную очистку воды, очищенной от взвешенных веществ на предыдущих стадиях, значительно увеличит срок службы материала загрузки.

Очищенную сточную воду можно сбрасывать в ливневую канализацию [2].

Установка локальных очистных сооружений требует большого капиталовложения – от 10 млн рублей и ежегодные затраты на обслуживание в размере от 250 тыс. рублей.

В то время как закупка установки ФФУ с фильтром ФСД и прицепом обойдется в семь раз дешевле. В зависимости мощности от 1 млн 920 тыс. рублей. Затраты на обслуживание составят 120 тыс. рублей.

При выборе мобильной системы очистки сточных вод выгода составит около 8 млн тыс. рублей. При этом затраты на обслуживание ежегодно будут значительно меньше по сравнению с локальными очистными сооружениями. Ежегодная выгода составит 130 тыс. рублей. (всего 120 000 рублей, это 30 000 на электричество, 40 000 на закупку сорбента и 30 000 затраты на слив в ливневую канализацию и 20 000 на обслуживание).

Таким образом резюмируя все вышеизложенное, делаем вывод, что мобильное очистные сооружения:

1. Процесс фильтрации воды установки полностью автоматизирован, не требует дополнительного обучения и специальных навыков.

2. Данная инновация малогабаритна и мобильна, что позволяет устанавливать ее в любое удобное для нас место и убирать в случае ненадобности, при этом оборудование выполнено во взрывозащищенном исполнении.

3. Использование очистных сооружений позволит исключить загрязнение окружающей среды.

4. Нет необходимости в создании дорогостоящего плана на строительство стационарных очистных сооружений (нет необходимости в дорогостоящих проектно-изыскательских и строительно-монтажных работах).

5. Стоимость установки и стоимость на ее эксплуатацию в разы меньше, чем ЛОС.



### Литература

1. Паспорт завода изготовителя: Комплекс оборудования производительностью 1 м<sup>3</sup>/час на базе флотационной установки «ФФУ-1К», с доочисткой на сорбционном фильтре «ФСД-1». Дата 18.02.2025.

2. Установки очистки воды ФФУ. Флотационно-фильтрационная установка ФФУ. Источник: <https://avto.detektorpoligraf.ru/ustanovki-ochistki-vody-ffu-2>. Дата 18.02.2025.

**BELOUSOV Oleg Anatolyevich**

Graduate Student,

Ulyanovsk Institute of Civil Aviation named after Chief Marshal of Aviation B. P. Bugaev,  
Russia, Ulyanovsk

## INDUSTRIAL RAINWATER WASTEWATER TREATMENT USING MOBILE SEWAGE TREATMENT PLANTS BASED ON A TRAILER

**Abstract.** *Currently, there are many enterprises, including fuel and lubricants warehouses, where rainwater treatment is absent or does not meet modern requirements, which entails certain financial risks for the enterprise and poses an environmental threat. This article discusses a rainwater treatment plant that does not require the cost of construction and installation work and stationary installation.*

**Keywords:** *rainwater treatment, fuel and lubricants storage, environmental protection.*

# ТЕХНИЧЕСКИЕ НАУКИ

**Jasmine Aziz Hussein**

Master's Student in Mechanical Engineering,  
Department of Reconstruction and Projects, Al-Iraqia University, Iraq, Baghdad

## DESIGN AND STUDY OF MATHEMATICAL ANALYSIS OF WIND TURBINES USING MATLAB/SIMULINK

**Abstract.** *Wind turbines are a topic of great interest and development due to the problems of emissions, climate change, and increasing carbon dioxide. Therefore, we analyzed, designed, and studied a wind turbine using MATLAB. A wind turbine was designed, its mechanical properties were studied, and its control was studied. Finally, these turbines can be proposed for operation in Iraqi conditions.*

**Keywords:** *wind speed, angle of attack, wind turbine.*

### Introduction

Wind power has been used for about 3,000 years. Until the early twentieth century, wind power was used to provide mechanical power for pumping water or grinding grain. At the beginning of modern industrialization, the use of fluctuating wind power was replaced by fossil fuel engines or the electrical grid, which provided a more stable source of power. In the early 1970s, with the first oil price shock, interest in wind power resurfaced. However, this time the focus was on wind power rather than mechanical power. Due to the problems facing the environment, the clear increase in carbon dioxide, economic instability and the increase in population density, it is necessary to rethink the search for alternatives to energy sources, especially non-traditional energies, including renewable energy sources such as solar and wind energy, especially for remote and isolated areas that are difficult to reach or deliver fuel and energy to these areas, as these alternatives can be used for domestic uses, which contributes to improving environmental conditions [1, p. 1-6; 2]. One of the most modern renewable energies in use is wind energy, as wind energy is an irregular and unstable source and is characterized by its momentary fluctuations [2, 3].

Non-renewable energy consumption is responsible for their total depletion, so the development of alternative energy is inevitable. The demand for electricity and the desire to minimize environmental damage have led to a greater reliance on renewable sources of electricity. Wind generators are of

greatest importance among them. Electricity generation systems based on the use of wind energy use the conversion of wind kinetic energy into mechanical energy of the generator rotation. Using the mathematical description of an idealized wind turbine, it is possible to estimate the value of the speed at which the turbine power and system efficiency are maximum.

One of the key electrical challenges in wind energy generation is the inherently random and variable nature of wind. To address this in wind-to-electricity conversion systems, Maximum Power Point Tracking (MPPT) strategies are implemented to optimize efficiency [1, p. 1-6]. The output power of a wind turbine varies with wind speed, and due to the unpredictable characteristics of wind, achieving maximum power output across all wind conditions is complex.

This study explores a control approach for a 20-kW small-scale horizontal-axis wind turbine that aims to maximize power output under fluctuating wind speeds. The main operational characteristics of the turbine are analyzed, and a mathematical model of a high-speed control system tailored for MPPT implementation is developed.

The following assumptions are made: air compression is neglected during turbine rotation, and power losses in components like the gearbox and inverter are disregarded. Only aerodynamic losses – specifically, the portion of wind energy not converted into torque and passing through the turbine – are considered.

Reference [4, p. 10-15] introduces an original method for modeling small-scale wind turbines based on their technical parameters. The paper proposes an enhanced mathematical model that incorporates both the power coefficient ( $C_{p\_pCp}$ ) and torque coefficient ( $C_{tC\_tCt}$ ). Additionally, control is achieved by adjusting the blade pitch angle ( $\beta$ ), allowing optimization of the output speed for maximum power extraction.

### 1. Wind turbine operation mode

Modern wind turbines can function in two operational modes: constant rotor speed and variable rotor speed. The turbine's performance is typically categorized into four distinct regions, as depicted in figure 1. Of these, only the second and third

regions represent active operation, while the first and fourth are considered inactive.

**Region 1** corresponds to wind speeds below the cut-in threshold (typically less than 5 m/s), during which the turbine remains stationary and does not generate power.

**Region 2** spans from the cut-in wind speed to the rated wind speed i.e., when wind speed exceeds 5 m/s but remains below the nominal speed  $v_{nom}$ . In this range, the turbine actively produces power. Two primary control strategies are applied in this region to optimize performance: adjusting the blade pitch angle (i.e., angle of attack in the horizontal plane) and modifying the generator's rotational speed.

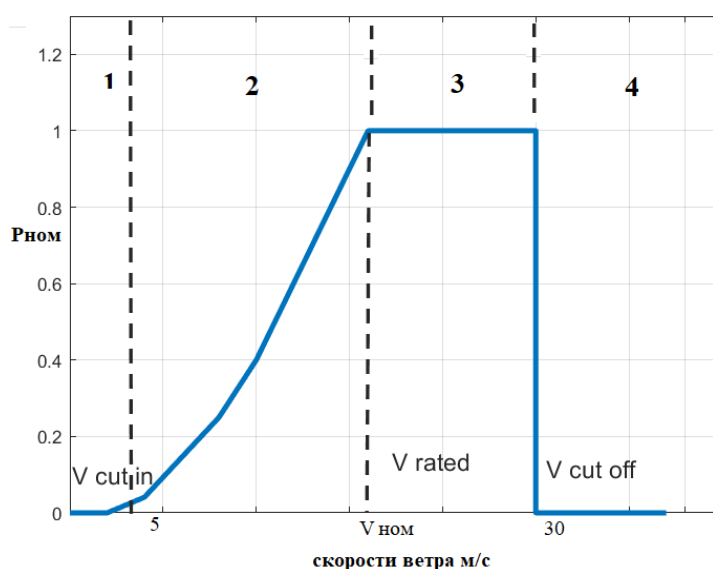


Fig. 1. Of the wind turbine operation mode [5]

**Region 3** lies between the cut-in and cut-out wind speeds, typically when wind speeds exceed nominal levels but are below the maximum safe limit (above 5 m/s and up to around 30 m/s). This is the range in which the turbine generates the maximum amount of energy. However, to prevent overloading the generator's electrical and mechanical systems, the turbine cannot harness all available wind energy. In this region, the turbine operates at a constant rotational speed and maintains a rated power output. To regulate excess energy and ensure safe operation, the blade pitch angle is actively adjusted.

**Region 4** represents conditions where the wind speed exceeds the cut-out threshold (typically above 30 m/s). At this point, the turbine is shut down to prevent mechanical damage from excessively high winds. To protect the system, the blades are rotated to a pitch angle of  $90^\circ$ , effectively stopping the rotor and preventing further energy capture [5].

### Mathematical description of wind turbine aerodynamics

A Wind Energy Conversion System (WECS) converts kinetic energy from the wind into mechanical energy and subsequently into electrical energy. The modeling involves aerodynamic, mechanical, and electrical subsystems. Below is a breakdown of each major component of the mathematical model.

When describing the model, equations (1-5) were used in [5; 6, p. 1-9; 7; 8, p. 163-168; 9, p. 917-921].

The turbine's torque is dependent on the wind speed.  $v_w$ , which acts on its blades, the power utilization factor  $C_p$ , the speed  $\lambda$  and the geometric dimensions of the turbine (radius  $R$  and turbine cross-section area  $A_t$ ) [5; 6, p. 1-9; 7; 8, p. 163-168; 9, p. 917-921].

Where:  $T_m = P_{T/\omega_t}$  and  $T_{opt} = P_{opt/\omega_{topt}}$ .

$$T_m = \frac{\rho \pi R^2 v_w^3 C_p(\lambda, 0)}{2 \omega_t} = \frac{\rho \pi R^5 C_p(\lambda, 0)}{2 \lambda^3} * \omega_t^2, \quad (1)$$



We built a complex turbine system using MATLAB as shown in the figure below using the mathematical system as shown in the figure 3.

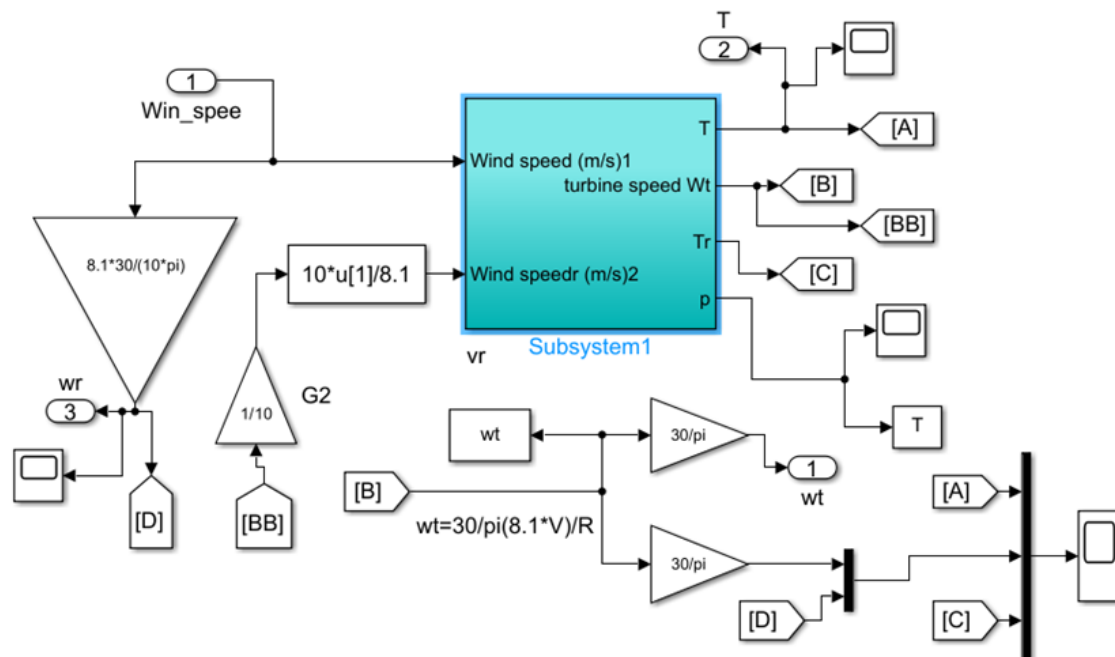


Fig. 3. Modelling system WECS By Simulink

Visual representation of the mathematical model compiled in Simulink. Equations (1-5) were

used when constructing the model. The modeling scheme of the wind turbine is shown in figure 4.

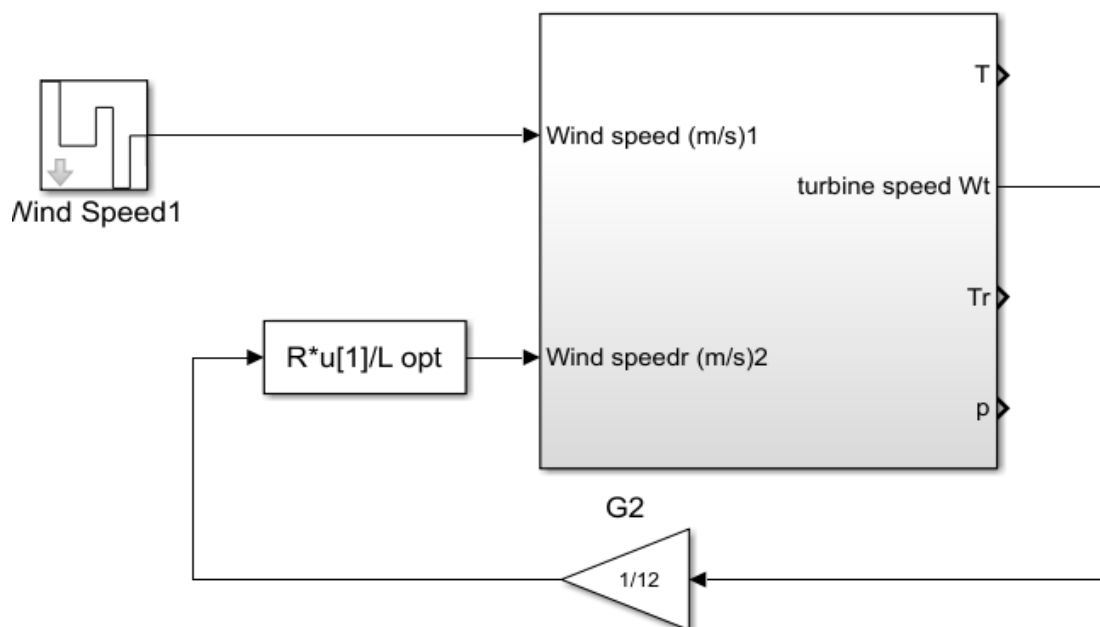


Fig. 4. Wind turbine simulation diagram in Simulink

The mathematical model for controlling the speed  $\lambda$  coefficient is shown in figure 5.

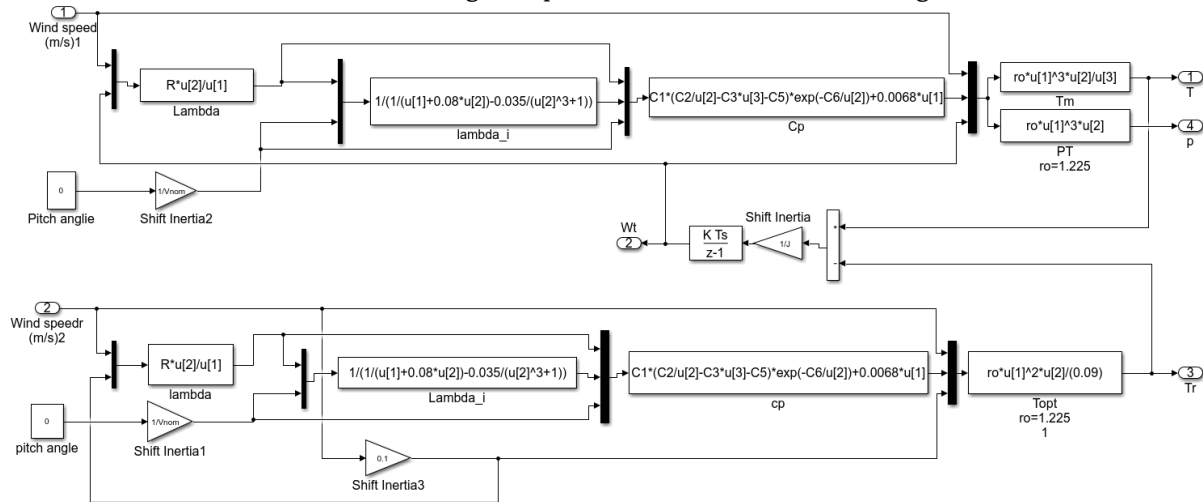


Fig. 5. Mathematical model of speed control

Take a sample of the wind speed as shown in m/s at different times shown in figure below (fig. the figure below, where it varies from 15 m/s to 5 6).

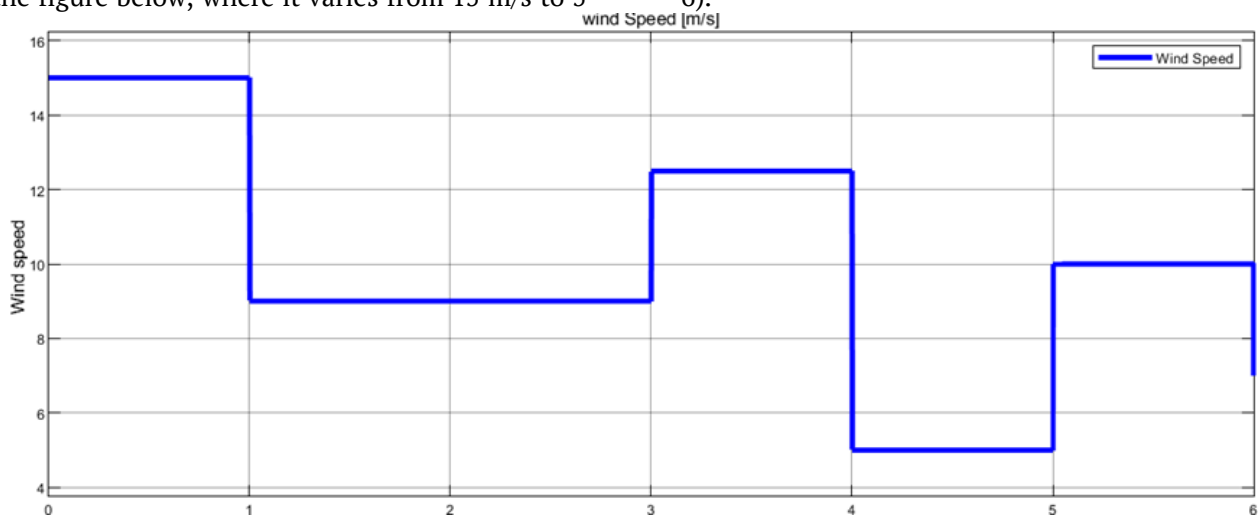


Fig. 6. Variable wind speed

We observe that the optimal turbine speed is different for different wind speeds, as illustrated in the figure below.

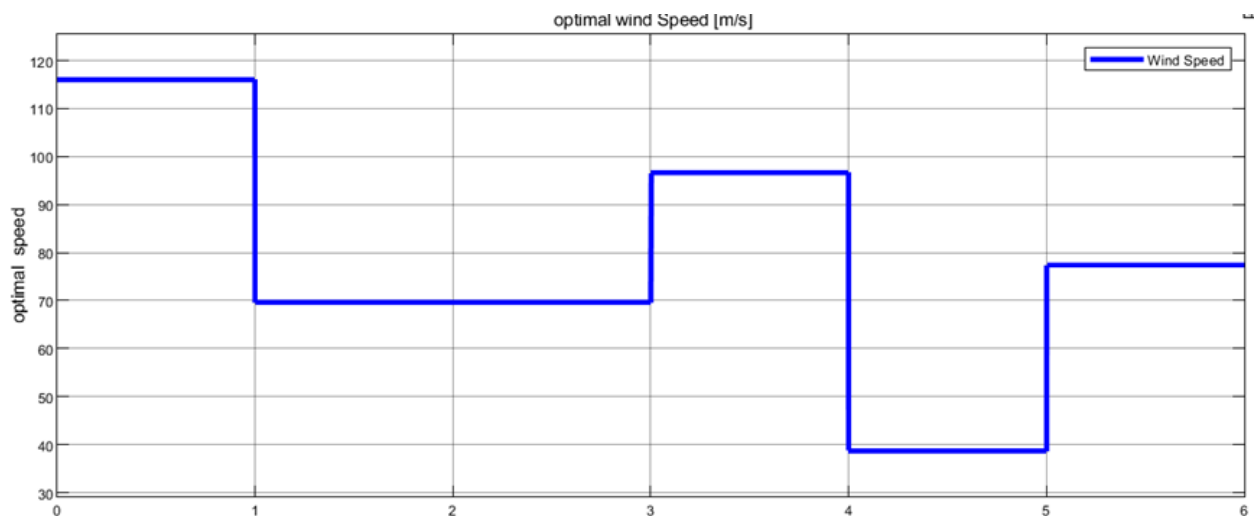


Fig. 7. Optimal turbine speed with time

We also notice the change in the actual speed depending on the change in wind speed, as in the figure below.

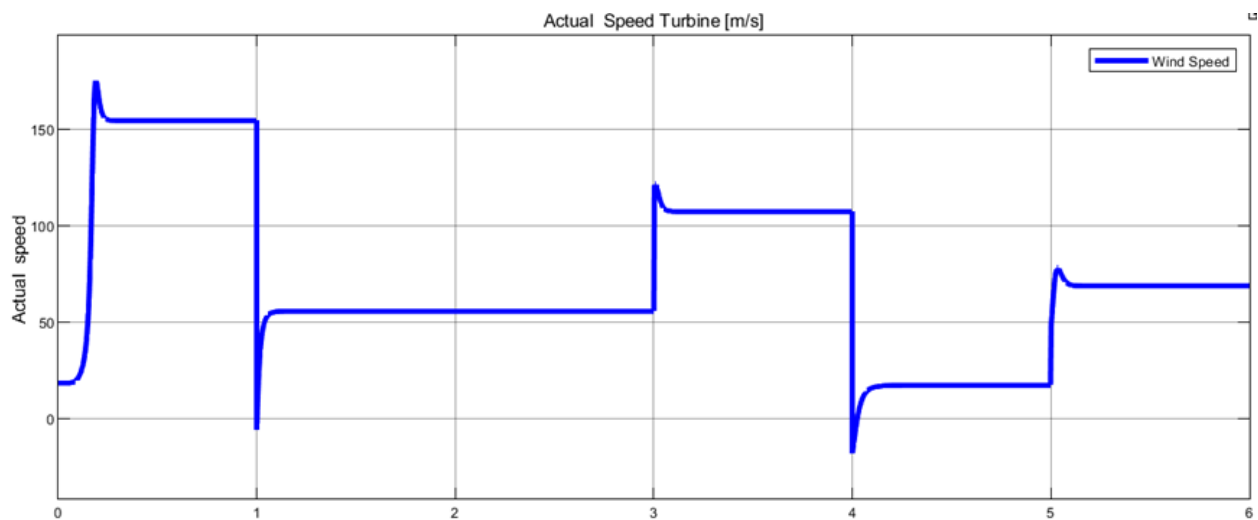


Fig. 8. Turbine speed with time

Here, the system that was developed using MATLAB is observed to correspond to the optimal speed while also matching the actual turbine speed and the optimal torque of the turbine, both of

which indicate the overall response speed when wind speed is altered, the controller causes the system to correspond to the new wind speed, as in the figure below figure 9.

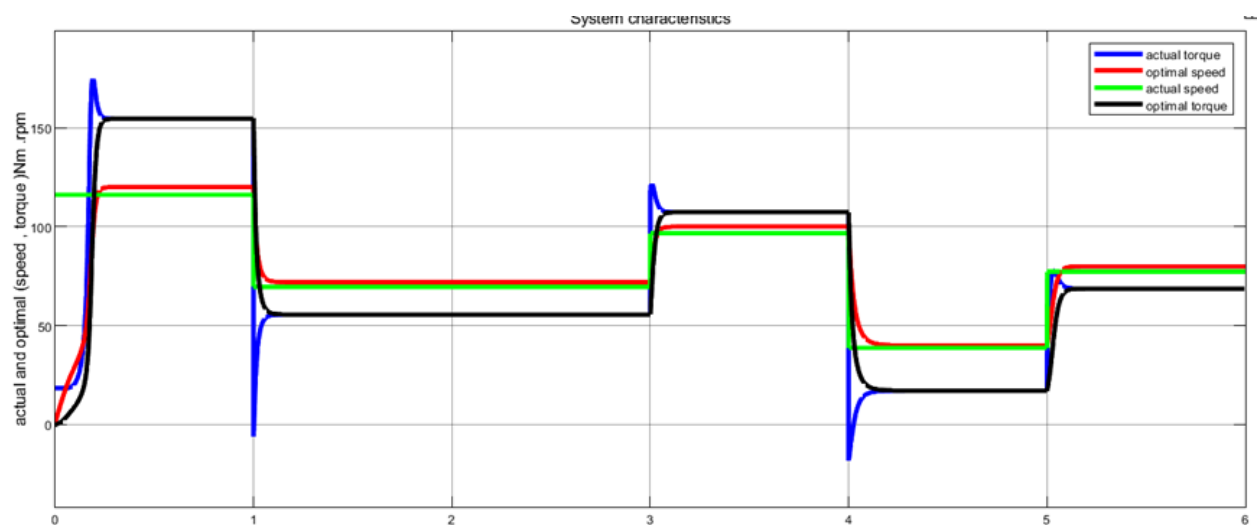


Fig. 9. Torque and speed with time

The mechanical and specific properties of wind turbines are a major concern for researchers, relying on the angle of attack and the blade. We will consider a range of different wind speeds, as shown in the figure below. We observed from mechanical properties experiments that the angle of attack has an impact on turbine power; the higher

the angle, the lower the turbine power, and vice versa.

The wind energy utilization coefficient ( $C_p$ ) value peaks at 0.48 when the blade angle of attack is  $\beta=0$ , over a range of  $\beta$  values from 0 to 50 degrees.



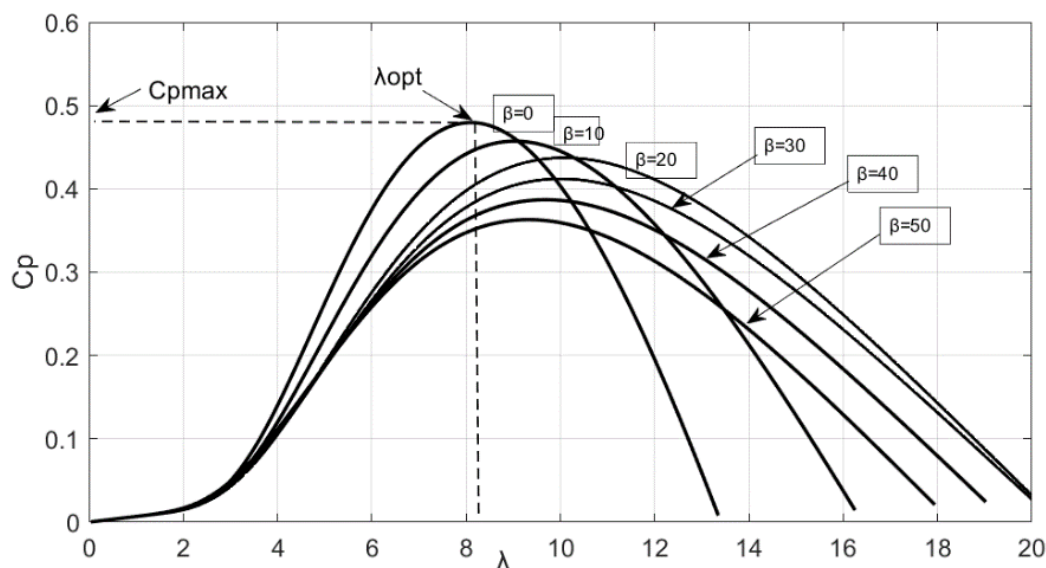


Fig. 10. Aerodynamic characteristics

### Conclusion

This project aims to demonstrate the effectiveness of precise parameters based on a large generator and its behavior as a wind turbine to overcome the continuously changing nonlinear nature of wind. The paper also covers the basic principles of wind turbine system modeling, which were used to design and simulate the system's turbine. A mathematical model was developed to manage the aerodynamics of a vertical-axis wind turbine. A turbine that harnesses wind energy was modeled in Simulink. The effect of the blade's angle of attack on the turbine's speed was examined. The power of the turbine is derived from the speed, and the maximum efficiency of the turbine was achieved. The maximum efficiency is reached at a speed of 6, which implies that the turbine's circular motion should be approximately 20 rad/s. It can be said that the average wind speed and turbine speed will be less than the most effective values. As such, during the design of a wind generator, it's important to recognize the pattern of wasted energy during operation.

### References

1. Errami Y., Maaroufi M., Ouassaid M. Modelling and control strategy of PMSG based variable speed wind energy conversion system, in Multimedia Computing and Systems (ICMCS), 2011 International Conference on, P. 1-6.
2. Data-Driven Approaches to Enabling Operational Intelligence for Wind Farms By iSolutions Inc. May 12, 2016 <https://www.isolutions.com/wind-farms-operational-intelligence/>.

3. Marouan Elazzaoui Modeling and Control of a Wind System Based Doubly Fed Induction Generator: Optimization of the Power Produced Journal of Electrical & Electronic Systems DOI: 10.4172/2332-0796.1000141 Volume 4 Issue 1 1000141 2015 with Page 2 of 8.

4. Obukhov S.G. Method OF modeling the mechanical characteristics of low-power wind turbines // Alternative Energy and Ecology – ISJAE. 2011. No. 1. P. 10-15.

5. Samokhvalov D.V., Jaber A.I., Filippov D.M., Kazak A.N., Hasan M.S. Research of Maximum Power Point Tracking Control for Wind Generator, in Proceedings of the 2020 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus), St. Petersburg and Moscow, Russia, 2020.

6. Jaber A.I., Kaittan A.S., Abdulwahhab M.W., Samokhvalov D.V. (2024). Efficiency improvement of PM synchronous wind generator using field oriented control with model-base current observer. International Review of Electrical Engineering, 19(1), P. 1-9.

7. Samokhvalov D.V., Jaber A.I. Estimation of the maximum efficiency and mechanical performance output from wind turbine, in Proc. 2019 International Conference on Smart City Management (SCM), 2019.

8. Samokhvalov D.V., Jaber A.I., Al-Mahturi F.Sh. Maximum Power Point Tracking of a Wind-Energy Conversion System by Vector Control of a Permanent Magnet Synchronous Generator, Russian Electrical Engineering, Vol. 92, No. 3, P. 163-168, 2021. DOI: 10.3103/S106837122103010X.

9. Jaber A.I., Samokhvalov D.V., Al-Mahturi F.S., Filippov D.M., Kazak A.N. Power Losses Calculation in Wind Power Plant based on a Vector-Controlled Permanent Magnet Synchronous

Generator, in 2021 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus), 2021, P. 917-921, doi: 10.1109/ElConRus51938.2021.9396514.

# ГЕОЛОГИЯ

Санкара Букари

студентка,

Российский университета дружбы народов имени Патриса Лумумбы, Россия, г. Москва

## СРАВНИТЕЛЬНЫЕ ИССЛЕДОВАНИЯ СРЕДНЕАЛЬБСКО-ПАЛЕОЦЕНОВЫХ ОТЛОЖЕНИЙ В ВОСТОЧНОЙ ЧАСТИ ОСАДОЧНОГО БАССЕЙНА КОТ-Д'ИВУАРА: СЛУЧАЙ СКВАЖИНЫ SE1d

**Аннотация.** Альбско-верхнемеловые отложения Кот-д'Ивуара обладают значительным потенциалом для добычи углеводородов. Они включают как материнские породы, так и породы-коллекторы. Предыдущие палинологические исследования, проведенные в этом осадочном бассейне, позволили установить биостратиграфические шкалы, пригодные для эксплуатации. В данной работе предлагается провести полную ревизию палинологических архивов среднего альба-палеоцена Кот-д'Ивуара, особенно на уровне скважины SE1d, их биостратиграфии и палеосреды. Для этого 306 образцов грунта из геологической съемки SE1d (1150–3570 м), добытого в рамках деятельности компании PETROCI, были подвергнуты микропалеонтологическим, наностратиграфическим и палинологическим исследованиям. Данное исследование позволило нам определить кровли ярусов от среднего альба до палеоцена в скважине SE1d.

**Ключевые слова:** Кот-д'Ивуар, осадочный бассейн, скважина SE1d, биостратиграфия, микропалеонтология, палинология, наностратиграфия.

### Введение

Берег Слоновой Кости принадлежит к древнему западноафриканскому щиту, который до открытия Атлантики был продолжением бразильского щита. Геологические формации Кот-д'Ивуара делятся на две хронологически различные единицы. С одной стороны, узкий осадочный бассейн (2,5%) вторично-третичного возраста на юге, а с другой стороны, докембрийская основа, которая составляет большую часть территории Кот-д'Ивуара, т. е. 97,5%. Большая часть работ, проводимых университетскими геологами в тесном сотрудничестве с нефтяными структурами в Ивуарийском бассейне, касается только меловых отложений морского бассейна с точки зрения его нефтяного интереса. В ходе этих исследований был выявлен нефтяной потенциал Кот-д'Ивуара. На основе микропалеонтологических данных другие, относительно более поздние исследования (Дигбехи и др., 1997; Сен-Марк и Н'Да, 1997) создали биостратиграфический синтез бассейна и предоставили подробную информацию об условиях осадконакопления. В осадочном бассейне исследования фораминифер,

палиноморф и наннофоссилий, проведенные в ходе разведочных работ на нефть Аналитическим и исследовательским центром PETROCI (Национальная компания нефтяных операций Кот-д'Ивуара) на 306 образцах шлама из скважины SE1d, позволяют предложить единую хроностратиграфическую структуру в локальном масштабе. Данное исследование будет способствовать согласованию практических знаний, полученных в ходе работы, проводимой Отделом биостратиграфии Аналитического и исследовательского центра PETROCI, с современными научными данными по биостратиграфии, палеосредам, переносимым спорами, пыльцевыми зернами и цистами динофлагеллят в ивуарийском осадочном бассейне.

### Область исследования

Данные опроса, использованные в этом исследовании, были предоставлены PETROCI. Это образцы шлама из нефтяной скважины, пробуренной в морской части осадочного бассейна Кот-д'Ивуара и расположенной на окраине Абиджана (рис. 1).

Съемка разреза (d), расположенная в прибрежной части бассейна Кот-д'Ивуара: SE1d.

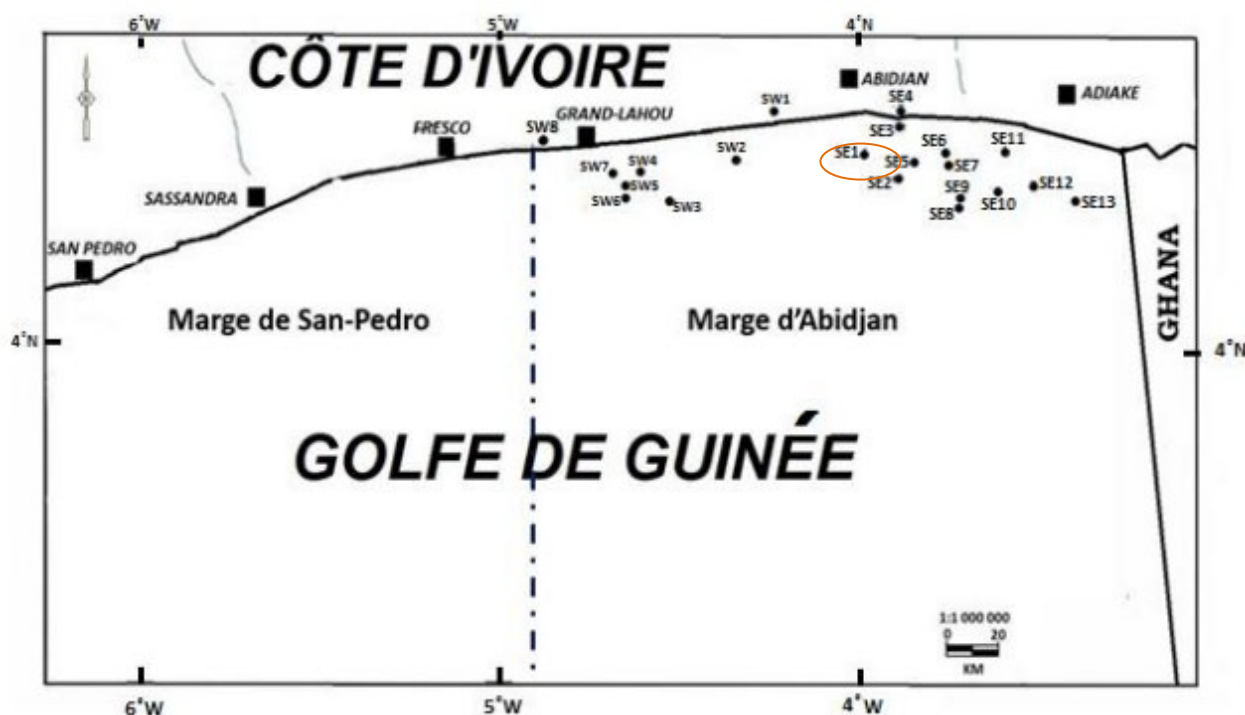


Рис. 1. Карта расположения скважины SE1d

### Сравнительные исследования: скважина SE1d

Объектом биостратиграфических исследований стала скважина SE1d, пробуренная на шельфе Кот-д'Ивуара в восточной части бассейна. Эти исследования были проведены в 1997 году Сен-Марком и Н'Да, а в 2009 году – Аналитическим и исследовательским центром (CAR) PETROCI. Действительно, скважина SE1d датировалась непрерывно от маастрихта до альба. Он был предметом междисциплинарных исследований параллельно с фораминиферами, известковыми нанноскопаемыми, цистами динофлагеллят, спорами и пыльцевыми зернами. Это сравнительное исследование позволяет установить четкую корреляцию между различными микропалеонтологическими группами. Результаты палинологического исследования показывают их взаимодополняемость и эффективность по сравнению с результатами других групп.

#### Исследования Сен-Марка и Н'Да (1997)

Исследования Сен-Марка и Н'Да (1997), проведенные на участке отвалов SE1d, представляют две большие осадочные группы: базальную последовательность и верхнюю последовательность. Базальная последовательность состоит из 1300 м осадков. В верхней части присутствуют глинистые прослои с тицинеллой и хедбергеллой верхнеальбского возраста. Верхняя последовательность (поздний альб-маастрихт) состоит из 1020 м осадков. В

основании он представлен песчано-глинистыми отложениями, в верхней части – глинистыми отложениями и обнаруживает многочисленные разрывы (верхний сеноман, коньяк, кампан-маастрихтский переход, базальный палеоцен), выявленные биостратиграфическим анализом.

#### Исследования PETROCI (2009)

Центр анализа и исследований PETROCI (неопубликованный внутренний отчет, 2009 г.) изучил триста шесть (306) образцов грунта из скважины SE1d (1150–3570 м). Это качественный и количественный анализ:

- 116 микропалеонтологических образцов, в частности фораминифер,
- 98 образцов в наностратиграфии,
- 92 палинологических образца.

Выделены этапы от палеоцена до среднего альба (рис. 2).

#### Палеоцен

##### Микропалеонтология фораминифер

Планктонные фораминиферы *Globoconusa daubjergensis* и *Chiloguembelina midwayensis* присутствуют на глубине 1180 и 1210 м. Присутствие этих планктонных фораминифер указывает на палеоцен. Присутствие известковых донных фораминифер *Loxostomoides applinae*, *Bulimina trigonalis*, *Eponides africana*, *Eponides pseudoelevatus*, *Eponides eshira*, *Bulimina inflata*, *Siphogenerinoides Eleganta*, *Nodosaria affinis*, *Gavelinella compressa*, *Globobulimina oviformis* на глубине 1150–1210 м свидетельствует о том,

что это Палеоцен. Этот палеоценовый интервал (1150–1210 м) отмечен очень редким присутствием или отсутствием агглютинированных фораминифер. На основании диагностических критериев в микропалеонтологии основание палеоценового интервала (граница мела и третичного периода или К/Т) определяется на глубине 1240 м с появлением индексного вида маастрихтского яруса.

#### **Микропалеонтология известковых нано-окаменелостей**

В скважине на глубине 1150 м находится FDO (First Downhole Occurrence) *Cruciplacolithus intermedius*. Вид FDO *Coccolithus pelagicus* встречается на глубине 1180 м. Действительно, анализ образцов из этого интервала показывает, что эти образцы некачественные. Первое появление FDO *Cruciplacolithus intermedius* на глубине 1150 м указывает на датский возраст (нижний палеоцен).

#### **Палеосреда: открытая морская, внутренняя и средняя неритическая**

Интервал от 1150 до 1240 м, представляющий палеоцен, включает микрофауну, состоящую из обычных или многочисленных известковых бентосных фораминифер, среди которых доминируют *Eponides* и *Bulimina*. Эти бентосные фораминиферы связаны с редкими планктонными фораминиферами. Эти сообщества предполагают неритическую среду обитания, расположенную во внутренней и средней части континентального шельфа. Отсутствие агглютинированных бентосных фораминифер, вероятно, связано с аноксическими явлениями, преобладавшими на морском дне в этот период осадконакопления.

#### **Маастрихтский**

#### **Микропалеонтология фораминифер**

Анализ проб показывает наличие в скважине на глубине 1240 м планктонных фораминифер *Rugoglobigerina macrocephala*, *Rugoglobigerina rugosa*, *Heterohelix striata*, *Heterohelix globulosa*, *Trinitella scotti*. Присутствие этих планктонных фораминифер подтверждает, что это маастрихт. Это позволяет нам локализовать кровлю маастрихтского яруса на отметке 1240 м. Этот возраст подтверждается на той же глубине известковистыми донными фораминиферами *Orthokarstenia clavata* и *Afrolivina afra* (нижний маастрихт) на глубине 1400 м, *Bulimina quadrilobata* на глубине 1550 м (нижний маастрихт). Эти фораминиферы связаны с видами большой стратиграфической протяженности. Подошва

маастрихта совпадает с первым скважинным проявлением кампанских отложений на глубине 1580 м.

#### **Палинология**

Первое появление FDO маастрихтских диноцист *Andalusiella gabonensis* и *Cerodinium granulostriatum* на глубине 1240 м позволяет предположить границу мела и третичного периода на этом уровне скважины. На глубине 1350 м подтверждение проникновения маастрихтских отложений дают FDO *Andalusiella ivoirensis*, а также FDO *Andalusiella mauthei* и *Andalusiella mauthei* subsp. *aegyptica* на высоте 1370 м. Судя по первому появлению кампанских маркеров, основание маастрихта располагается на глубине 1580 м в скважине.

#### **Микропалеонтология известковых нано-окаменелостей**

Анализ образцов этого интервала показывает, что он начинается с образцов, богатых наннофоссилиями, комплексы которых хорошо разнообразны. На глубине 1240 м от скважины выделен маастрихт с первым появлением ПДО верхнемеловых видов, в том числе характерных для маастрихта: *Arkangelskiella cymbioformis*, *Uniplanarius sissinghii* и *Eiffellithus turriseiffelii*, *Watznaueria barnesiae*, *Micula staurophora*, *Quadrum bengalensis* и *Arkangelskiella* маастрихтана. На глубине 1270 м маастрихтский ярус был подтвержден присутствием FDO *Arkangelskiella cymbioformis* и *Lithraphidites quadratus*. Наличие FDO *Uniplanarius sissinghii* и *Eiffellithus turriseiffelii* на высоте 1330 м также подтверждает маастрихт. В связи с единичными находками *Uniplanarius sissinghii* 1640 м и *Uniplanarius gothicus* 1700 м нижнюю часть этого интервала можно отнести к нижнему кампану. Однако отсутствие *Uniplanarius sissinghii* и *Quadrum trifidum* на глубине 1700 м позволяет предположить, что возраст отложений не древнее среднего или нижнего кампана. Этот интервал, относимый к кампано-маастрихтскому ярусу, содержит зону (1520–1640 м), где наннофоссилии встречаются очень редко или отсутствуют.

#### **Палеосреда: открытая морская, внутренняя и внешняя неритическая**

На глубине от 1240 до 1450 м микрофауна характеризуется обильным и разнообразным присутствием известковых бентосных фораминифер, среди которых доминируют *Bulimina* spp. Эти виды развиваются в основном на уровне континентального шельфа. Эти бентосные фораминиферы связаны с обычным

присутствием некилевидных планктонных форм. Диноцисты также многочисленны и разнообразны, их сопровождают встречающиеся или редкие миоспоры. Эти комплексы предполагают внутреннюю или среднюю неритическую среду для этого интервала. Присутствие небольшого количества агглютинированных бентосных форм свидетельствует о дисоксической среде, преобладавшей на морском дне в этот период осадконакопления. На глубине от 1450 до 1570 м присутствуют комплексы фораминифер, относительно разнообразные в агглютинирующих бентосных формах (глубоководная группа *Bathysiphon* spp.). Известковые бентосные фораминиферы редки, планктонные формы отсутствуют. Диноцисты многочисленны и разнообразны, среди них преобладают *Paleocystodinium golzowense* и *Andalusiella ivoirensis*, встречающиеся в редких случаях среди миоспор. Эти комплексы указывают на среднюю и внешнюю неритическую среду для этого интервала.

#### **Кампанский период**

##### **Микропалеонтология фораминифер**

Первое появление агглютинированных бентосных фораминифер *Gaudryina cretacea* FDO на глубине 1610 м указывает на кампан. Этот кампанский таксон связан с другими агглютинирующими и известковыми бентосными фораминиферами, уже обнаруженными в маастрихтских отложениях. Что касается планктонных фораминифер, то они отсутствуют. Микропалеонтологический анализ показал, что кампанский разрез залегает на отложениях нижнего сенона на глубине 1760 м.

##### **Палинология**

Палинологический анализ показал первое появление вида *Xenascus ceratioides* на глубине 1580 м. Появление этих FDO *Xenascus ceratioides* указывает на кампанский ярус. Это один из типичных таксонов кампанского яруса у побережья Кот-д'Ивуара (Жардине и Маглуар, 1965, 1967; Шривастава, 1995; Игнатий, 2022). Аналогично, присутствие на глубине 1660 м диноцист, в том числе FDO *Hystrichodinium isodiametricum*; и FDO *Trichodinium castaneum* на глубине 1690 м в скважине подтверждает кампанский ярус. Таким образом, присутствие FDO *Circulodinium distinctum* на глубине 1720 м указывает на нижний кампан. В этом интервале миоспоры не наблюдались. Подошва кампана фиксируется на отметке 1720 м с появлением диноцист нижнего сенона.

#### **Палеосреда: морская, от средней до внешней неритической**

Палеосреда кампана и маастрихта идентична.

#### **Нижний сенон**

##### **Микропалеонтология фораминифер**

Анализ образцов показывает первое появление FDO *Whiteinella baltica* на глубине 1770 м. Присутствие этих фораминифер указывает на нижний сенон. Присутствие FDO *Whiteinella archaeocretacea* на глубине 1800 м подтверждает возраст нижнего сенона. Эти таксоны связаны с другими планктонными фораминиферами, включая *Hedbergella* spp., *Heterohelix* spp. и *Whiteinella* spp. Что касается агглютинирующих бентосных и известковых бентосных видов, то они очень редки, даже отсутствуют. Основание этого интервала обозначено на отметке 1830 м по FDO туронских фораминифер.

##### **Палинология**

Присутствие *Droseridites senonicus* FDO на глубине 1720 м в скважине указывает на кровлю нижнесенонского интервала. Однако нижняя кровля сенонского яруса скорректирована до 1675 м на основании записей каротажа. FDO *Odontochitina porifera* подтверждает возраст нижнего сенона на уровне 1720 м, что предполагает проникновение сантона. Эти нижнесенонские таксоны связаны с диноцистами, включая *Circulodinium dependentum*, *Hystrichodinium pulchrum*, *Odontochitina operculata*, *Oligosphaeridium complex* и *Xenascus sarjantii*. В этом интервале также присутствуют пыльцевые зерна *Proteacidites dehaani*, *Ephedripites multicostatus* и *Syncolpites marginatus*.

##### **Микропалеонтология известковых нано-окаменелостей**

Первое появление кокколитов FDO *Eprolithus floris*, *Lithastrinus septenarius/moratus* на высоте 1730 м маркирует кровлю нижнего сенона. Присутствие *Zeugrhabdotus noeliae* FDO на глубине 1830 м указывает на нижний сенон. Этот интервал можно отнести к нижнему коньяку из-за полного отсутствия рода *Micula* (*M. staurophora* и *M. concava*).

#### **Палеосреда: средняя неритическая**

Обычное присутствие диноцист и бентосных фораминифер в интервале от 1675 до 1760 м, что предполагает среднеритическую среду. В интервале глубин 1760–1820 м в осадках встречаются редкие диноцисты, обычное и относительно разнообразное присутствие миоспор и планктонных фораминифер. Эти

сообщества подразумевают от внутренних неритических до среднеритических среду.

### **Туронский**

#### **Микропалеонтология фораминифер**

FDO *Hedbergella planispira*, *Hedbergella simplex* и *Hedbergella planispira*, присутствующие на высоте 1830 м, указывают на турон. Этап характеризуется обилием планктонных фораминифер, представленных *Hedbergellinidae* и *Heterohellicidae*. Подошва туронского интервала зафиксирована на отметке 1840 м, с маркерами FDO сеномана.

#### **Палинология**

FDO *Tricolpites* sp. SCI 107 на расстоянии 1810 м от скважины, предполагающее проникновение туронских отложений, указывает на верхнюю часть интервала. Его определение основано на единственном образце глубиной 1810 м, что делает этот туронский интервал очень узким (10 м). На основании палинологических данных туронский интервал залегает непосредственно на сеноманских слоях на глубине 1820 м.

#### **Микропалеонтология известковых наноокаменелостей**

Образцы этого интервала богаты хорошо сохранившимися наноокаменелостями разнообразных комплексов. Первое появление FDO *Stoverius achylosus* на глубине 1830 м в скважине знаменует идентификацию туронского яруса. И этот уровень подтверждается наличием FDO *Radiolithus planus* на глубине 1850 м. Наблюдение за *Rhagodiscus asper*, указывающим на кровлю сеномана, позволило нам поместить основание этого интервала на отметку 1880 м.

#### **Палеосреды: внутренние и средние неритические**

В этом интервале микрофауна представлена в основном шарокамерными планктонными формами, включая *Hedbergella*, *Whiteinella*, *Heterohelix* и др. Что касается бентосных особей, то они отсутствуют. Палиноморфы состоят только из миоспор. Эта ассоциация предполагает внутренние или средние неритические условия осадконакопления.

### **Сеноман**

#### **Микропалеонтология фораминифер**

Интерпретация каротажа позволила установить кровлю сеноманского интервала на глубине 1840 м по скважине. Выбор положения кровли, основанный на интерпретации каротажных данных, согласуется с наличием FDO *Globigerinelloides bentonensis* на той же глубине

(1840 м), а также FDO *Schackoina senomana* на глубине 1850 м, что подтверждает наличие сеномана. В этом интервале в микрофауне преобладают планктонные виды, уже отмеченные в других таксонах сеномана. Это FDO *Globigerinelloides caseyi* на высоте 1880 м и FDO *Hedbergella/Globigerinelloides* sp. на высоте 2150 м. Отмечено присутствие или даже редкость булиминид. В верхней части сеноманского интервала булиминиды присутствуют или даже редки. Таким образом, подошва сеноманского интервала зафиксирована на глубине 2200 м.

#### **Палинология**

Первое появление пыльцевых зерен *Classopollis classoides* на глубине 1900 м в скважине указывает на кровлю сеноманского разреза. Присутствие FDO *Classopollis brasiliensis* на глубине 1940 м, FDO *Steevesipollenites binodosus* на глубине 1960 м и FDO *Gnetaceaepollenites jansonii* на глубине 2150 м подтверждает этот сеноманский. В этом интервале скопления диноцист отсутствуют. Таким образом, сеноманский интервал залегает на альбе на глубине 2200 м.

#### **Микропалеонтология известковых наноокаменелостей**

Наннофоссилии плохо сохранились, а комплексы бедны следами перекристаллизации или растворения в этом интервале. Появление FDO *Rhagodiscus asper* и *Radiolithus hollandicus* на глубине 1880 м выявило сеноман. FDO *Axopodorhabdus albianus*, присутствующий на глубине 1910 м, FDO *Staurolithites gausorhetium* на глубине 2030 м и FDO *Gartnerago theta* на глубине 2060 м подтверждают сеноман.

#### **Палеосреда: внутренняя неритическая**

Интервал от 1840 до 2200 м содержит исключительно планктонные фораминиферы. Они состоят из многочисленных *Hedbergellidae* и *Heterohellicidae*. Что касается диноцист и бентосных фораминифер, то они отсутствуют. Эти комплексы характерны для внутренней морской неритической среды осадконакопления.

### **Альбийский**

#### **Микропалеонтология фораминифер**

Появление FDO планктонных видов *Ticinella primula*, *T. raynaudi*, *Ticinella roberti* и *Ticinella* spp. на глубине 2210 м в формациях присутствует альбский ярус. Одновременное присутствие FDO *Ticinella/Globigerinelloides* sp. на глубине 2210 м предполагается среднеальбская последовательность. Значительное присутствие *Ticinella* и *Globigerinelloides*, а также



слабое присутствие *Hedbergella* подтверждает альбский возраст. Таким образом, кровля альбского интервала находится на глубине 2210 м. Только верхняя часть среднего альба (2210–2230 м) сложена обильными фораминиферами, представленными исключительно планктонными видами. Нижняя часть среднеальбского интервала (2230–2625 м) отличается редкостью фораминифер. Основание этого интервала совпадает с исчезновением фораминифер, а также со значимым изменением каротажных кривых на глубине 2625 м. На глубине от 2625 до 3570 м проанализированные отложения лишены фораминифер.

#### Палинология

Наблюдение на глубине 2200 м видов *Appendicisporites potomacensis*, *Cicatricosisporites baconicus*, *C. berouensis*, *Ephedripites torosus* и *Lusatisporis dettmannae* позволяет предположить средний альб. FDO других миоспор *Elatersporites klaszi* и *Ephedripites barghoornii* на высоте 2220 м, *Callialasporites dumpieri*, *Classopollis minor* и *Densoisporites velatus* на высоте 2240 м, *Steevesipollenites sinuosus* на высоте 2260 м и *Ephedripites fusiformis* на высоте 2280 м подтверждают альбский возраст. Присутствие элатерных форм во всех образцах альба позволяет

предположить, что возраст этого интервала не древнее среднего альба. Поэтому предполагается, что скважина SE1d заканчивается отложениями не моложе среднего альба.

#### Микропалеонтология известковых наноокаменелостей

Наблюдение за первым появлением *FDO Nannosconus truiti truiti* на глубине 2180 м в скважине позволяет выделить кровлю альба. До глубины 2670 м виды становятся редкими и плохо сохраняются. На глубине от 2750 м до 3580 м наннофоссилии встречаются очень редко или вообще отсутствуют.

#### Палеосреды: внутренние неритические и неморские

Планктонные фораминиферы *Ticinella* и *Globigerinelloides* в скважине встречаются в большом количестве на глубине от 2210 м до 2230 м. Что касается бентосных фораминифер, то они отсутствуют. В этом интервале присутствуют и обильны миоспоры. В этом же интервале диноцисты отсутствуют. Эти комплексы указывают на внутреннюю морскую неритическую обстановку. В конце съемки (от 2230 до 3570 м) фораминиферы отсутствуют. Это свидетельствует о неморской среде осадконакопления, что подтверждается наличием миоспор и отсутствием диноцист.

PALYNOMORPHES Cette étude	PALYNOMORPHES Petroci, 2009	FORAMINIFÈRES Petroci, 2009	NANNOFOSSILES Petroci, 2009	PALÉOENVIRONNEMENTS Petroci, 2009
PALÉOCÈNE +1150 m		PALÉOCÈNE +1150 m	DANIEN +1150 m	Néritique interne à moyen
MAASTRICHTIEN SUP. 1240 m	MAASTRICHTIEN +1240 – 1560 m	MAASTRICHTIEN 1240 m	MAASTRICHTIEN- CAMPANIEN 1240 m	Néritique interne à externe
MAASTRICHTIEN INF. 1350 m				
CAMPANIEN SUP. 1580 m	CAMPANIEN 1580 – 1690 m	CAMPANIEN 1610 m		Néritique interne à externe
CAMPANIEN INF. 1690 m				
DISCONTINUITÉ BASE CAMPANIENNE				
CONIACIEN 1720 m	SENONIEN INFÉRIEUR 1720 – 1780 m	CONIACIEN 1770 m	CONIACIEN 1730 m	Néritique interne à moyen
TURONIEN 1820 m	TURONIEN 1810 m	TURONIEN 1830 m	TURONIEN 1830 m	Néritique interne à moyen
CÉNOMANIEN SUP. 1900 m	CÉNOMANIEN 1820 – 2180 m	CÉNOMANIEN 1840 m	CÉNOMANIEN 1880 m	Néritique interne
CÉNOMANIEN INF. 1940 m				
DISCONTINUITÉ POST ALBIENNE				
ALBIEN SUPÉRIEUR 2200 m	ALBIEN 2200 – 3570 m	ALBIEN 2200 - 3570 m	ALBIEN 2180 - 3570 m	Non marin à néritique interne
ALBIEN MOYEN 2220 - 3570 m				

Рис. 2. Сводка биостратиграфических кровель исследования SE1d

### Заключение

В данной работе представлены сравнительные исследования среднеальбско-палеоценовых отложений на востоке осадочного бассейна Кот-д'Ивуара на примере исследования SE1d. Исследование посвящено микропалеонтологии, наностратиграфии и палинологии. Качественный и количественный анализ был проведен на 116 микропалеонтологических образцах, в частности фораминиферах, 98 образцах наностратиграфии и 92 образцах палинологии. Изучение отложений среднего альба и палеоцена съемки SE1d показало, что они богаты палиноморфами морского (цисты динофлагеллят) и континентального (споры и пыльцевые зерна) происхождения, фораминиферами и наннофоссилиями. Что касается палеосреды, то она состоит из следующих сред: от неморской до внутренней неритической, внутренняя неритическая, внутренняя до средней неритической, внутренняя до внешней неритической.

### Литература

1. Abubakar M.B., Luterbacher H.P., Ashraf A.R., Ziedner R., Maigari A.S. (2011) – «Late Cretaceous palynostratigraphy in the Gongola Basin (Upper Benue Trough, Nigeria)». *Journal of African Earth Sciences*, 60 (1-2), P. 19-27.
2. Abubakar M.B., Obaje N.G., Luterbacher H.P., Dike E.F.C., Ashraf A.R. (2006) – «A report on the occurrence of Albian-Cenomanian elater-bearing pollen in Nasara-1 well, Upper Benue Trough, Nigeria: Biostratigraphic and palaeoclimatological implications». *Journal of African Earth Sciences*, 45, P. 347-354.
3. Atta-Peters D., Salami, M.B. (2006) – «Aptian-Maastrichtian palynomorphs from the offshore Tano Basin, Western Ghana». *Journal of African Earth Sciences*, 46, P. 379-394.
4. Azema C., Boltenhagen E. (1974) – «Pollen du Crétacé moyen du Gabon attribué aux Ephedrales. Paléobiologie continentale» *Montpellier* 5(1), P. 1-37, [4 pl].
5. Bié G.R., Digbéhi Z.B., Yao K.R., Tea-Yassi J., Kangah K.D., Tahi I. (2012) – «Stratigraphie Palynologique du Maastrichtien Supérieur – Eocène Supérieur du Bassin Sédimentaire Offshore de Côte d'Ivoire, Afrique de l'Ouest». *International Journal of African* 6, P. 40-57.
6. Digbéhi Z.B., N'da L.V., Yao K.R., Atteba Y.A. (1997) – «Principaux foraminifères et palynomorphes crétacés du bassin sédimentaire de Côte d'Ivoire, Golfe de Guinée septentrional: propositions pour une échelle biostratigraphique locale». *Africa Geoscience Review* 4(3-4), P. 467-479.
7. Digbéhi Z.B., Toé Bi K.K.K., Adopo K.L., Guédé K.E., Tahi I., Yao K.R. (2011) – «Palynologie et environnements de dépôt des sédiments d'âge cénomanien supérieur-maastrichtien inférieur dans le bassin offshore de Côte d'Ivoire (Afrique de l'ouest)». *Sciences et Nature* 8(1), P. 95-105.
8. Guédé K.E. (2009) – «Caractérisation Palynostratigraphique et paléoenvironnementale des formations du passage Crétacé-Tertiaire et Eocène dans l'étude du puits offshore DINO1X». DEA des sciences de la terre option Géologie Marine et Sédimentologie, UFR STRM, Univ. Cocody (Abidjan). 78 p.
9. Guédé K.E. (2016) – «Etude comparée de la palynoflore (kystes de dinoflagellés) au Crétacé-Paléogène (K-Pg) et Paléocène-Eocène (P-E) du Nord-Ouest du Maroc et de la Côte d'Ivoire: Systématique, Biostratigraphie Paléoenvironnements et Paléobiogéographie». Thèse de doctorat, Université Mohammed V Faculté des sciences Rabat, P. 14-17.
10. Guédé K.E., Slimani H., Louwye S., Asebriy L., Toufiq A., Ahmamou M., El Amrani, El Hassani I.E., Digbéhi Z.B. (2014) – «Organic-walled dinoflagellate cysts from the Upper Cretaceous-lower Paleocene succession in the western External Rif, Morocco: new species and new biostratigraphic result». *Geobios* 47, P. 291-304.
11. Ignace Tahi. Palynologie et caractérisation de la matière organique des dépôts Albo/Aptien-Crétacé Supérieur du bassin sédimentaire de Côte d'Ivoire. Paléontologie. Thèse de Doctorat, Sorbonne Université, 2022.

**Sankara Bukari**

Student, Patrice Lumumba Peoples' Friendship University of Russia,  
Russia, Moscow

## **COMPARATIVE STUDIES OF THE MIDDLE ALBIAN-PALEOCENE SEDIMENTS IN THE EASTERN PART OF THE IVORY COAST SEDIMENTARY BASIN: THE CASE OF THE SE1D WELL**

**Abstract.** *The Albian-Upper Cretaceous deposits of Côte d'Ivoire have significant potential for hydrocarbon production. They include both parent rocks and reservoir rocks. Previous palynological studies conducted in this sedimentary basin made it possible to establish biostratigraphic scales suitable for operation. In this paper, it is proposed to conduct a complete revision of the palynological archives of the Middle Albian-Paleocene of Côte d'Ivoire, especially at the SE1d well level, their biostratigraphy and paleomedium. For this purpose, 306 soil samples from the SE1d geological survey (1150-3570 m), extracted as part of PETROCI's activities, were subjected to micropaleontological, nanostratigraphic and palynological studies. This study allowed us to determine the roofs of the tiers from the Middle Albian to the Paleocene in the SE1d basin.*

**Keywords:** *Côte d'Ivoire, sedimentary basin, SE1d well, biostratigraphy, micropaleontology, palynology, nanostratigraphy.*

# ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

PETROVA Iana Andreevna

Postgraduate Student in Management and Business Administration,  
University of California Riverside, USA, Riverside

## APPLICATION OF THE KANO MODEL IN PRODUCT DEVELOPMENT: CROSS-INDUSTRY INSIGHTS AND METHODOLOGICAL APPROACHES

**Abstract.** This article explores the Kano model as a tool for prioritizing product features based on customer perception. It outlines the core principles of the model, its structure, and modern methodological adaptations. Using examples from five industries – telecommunications, residential real estate, financial services, and consumer electronics – the article demonstrates various patterns of factor distribution (must-be, performance, and delight factors). A comparative analysis is provided on how Kano factors behave depending on market maturity, user segments, and technological context. The article offers practical recommendations for combining the Kano model with TURF analysis, MaxDiff, and conjoint analysis. It also discusses the limitations of the Kano model and proposes practical scenarios for adapting it to rapidly changing market environments.

**Keywords:** Kano model, requirements prioritization, product development, customer perception, TURF analysis, MaxDiff, conjoint analysis, product strategy, satisfaction, digital markets.

### 1. Introduction

In today's digital economy, companies must constantly improve their product offerings to remain competitive and respond to rapidly evolving customer expectations. One methodology that has proven effective in analyzing customer needs and prioritizing product functionality is the Kano model, developed by Japanese professor Noriaki Kano in the 1980s within the Total Quality Management (TQM) framework.

The core idea of the model is that different product features have varying impacts on customer satisfaction: some are basic and expected by default, others are performance-related and predictably enhance user perception, and still others are delight factors that generate a "wow" effect, significantly increasing customer loyalty. This classification enables companies to make informed decisions about which features to include at different stages of product development [3, p. 89].

The goal of this article is to examine the applicability of the Kano model across different industries and to evaluate its effectiveness in the development of both digital and physical products within a highly dynamic technological landscape. Special focus is given to its practical application in SaaS products for the travel and hospitality sector,

as well as its integration with other tools such as conjoint analysis, MaxDiff, and TURF analysis (Total Unduplicated Reach and Frequency).

Through real-world examples from telecommunications, real estate, fintech, electronics, and hospitality software, and drawing on the author's professional experience, the article identifies the model's practical boundaries and provides recommendations for its adaptation in uncertain and rapidly shifting environments.

### 2. Theoretical Framework: The Kano Model

The Kano model is a classification tool for product features based on their impact on customer satisfaction. The methodology was introduced by Professor Noriaki Kano in 1984 and has since become the foundation for numerous studies in quality management, user experience design, and product management.

#### 2.1. Classification of Product Features

The Kano model distinguishes the following main types of customer requirements:

- **Must-be (Basic) Features** – These are characteristics whose absence leads to significant customer dissatisfaction, while their presence does not increase satisfaction. They represent the minimum expected standard, often taken for granted. Examples include car reliability or a functioning internet connection in a hotel.

- **One-dimensional (Performance) Features** – These features have a direct linear relationship with satisfaction: the better the performance, the higher the satisfaction, and vice versa. Examples include internet speed or the number of available TV channels.

- **Attractive (Delighters) Features** – These are unexpected and pleasant characteristics that elicit positive emotions when present but do not cause dissatisfaction when absent. They are a key source of competitive advantage and the “wow” effect. Examples include complimentary gifts or exclusive features.

Additionally, the model may identify:

- **Indifferent Features** – Attributes that have no impact on the customer’s perception.

- **Reverse Features** – Elements that may cause dissatisfaction if present and satisfaction if absent (rare cases).

- **Questionable Features** – Features for which the consumer gives logically inconsistent responses, indicating confusion or misunderstanding [4, p. 56-63].

## 2.2. Evaluation Methodology

The classical Kano model uses a paired-question format for each feature, asking customers:

1. How would you feel if this feature were present in the product?
2. How would you feel if this feature were absent?

Responses are interpreted using a predefined evaluation table, allowing each respondent to categorize the feature accordingly. Aggregating all responses results in a Kano map – a visual matrix that distributes features across quadrants of satisfaction and dissatisfaction.

## 2.3. Modern Interpretations and Modifications

Over time, several additional approaches and frameworks have been developed to expand the capabilities of the classical Kano model:

- **Kano Continuum** – A method where each type of response is assigned a numerical weight (e.g., 4, 2, 1, 0), enabling a more nuanced evaluation of each feature’s contribution to satisfaction.

- **Regression-Based Approach** – Statistical analysis of how the presence or absence of features impacts overall satisfaction metrics such as Customer Satisfaction Index (CSI) or Net Promoter Score (NPS).

- **Segmentation and Feature Mapping by Target Audience** – During research, respondents can be divided into clusters, allowing for the

construction of separate Kano models tailored to each group.

Thus, the Kano model serves not only as a diagnostic tool but also as a strategic guide in product design and evolution. It helps optimize the value proposition and prioritize development efforts in resource-constrained environments.

## 3. Integration with Other Frameworks

Despite the practical value of the Kano model, using it in isolation does not always address the full range of challenges involved in product decision-making. In highly competitive and fast-changing markets, it becomes essential to combine Kano with other quantitative methods that offer deeper insights into demand structure, customer preferences, and the potential reach within the target audience. Among the most effective of these methods are TURF analysis, MaxDiff analysis, and conjoint analysis [2].

### 3.1. TURF Analysis (Total Unduplicated Reach and Frequency)

TURF analysis is used to identify the optimal combination of features or products that will provide the maximum unduplicated reach across a target audience. Unlike the Kano model, which focuses on emotional response, TURF enables a quantitative assessment of which minimum set of features or products can satisfy the greatest number of user needs (jobs-to-be-done). It is particularly valuable when deciding where to invest among a long list of attractive or performance features.

*Example:* When designing a new tablet model with 30 potential features (e.g., AR support, built-in projector, waterproofing), TURF analysis can help select the 5 most relevant features to cover 80–90% of the audience.

### 3.2. MaxDiff Analysis (Maximum Difference Scaling)

MaxDiff analysis helps rank features based on their relative importance. Instead of asking users to rate all features at once – which often leads to fatigue and inconsistent results – respondents are shown sets of 4 to 6 items and asked to select the most and least important. Repeating these rounds generates a stable distribution of preferences. This method is especially useful when the number of potential attributes is large and accurate ranking is essential.

MaxDiff can serve as an anchor method for TURF analysis, as it helps identify which features fall into the top-1 or top-2 preferences for each respondent.

### 3.3. Conjoint Analysis

Conjoint analysis is one of the most complex yet powerful tools in the product research toolkit. It simulates real-world consumer behavior by asking users to choose between product options with different combinations of attributes. This enables researchers to:

- Determine the weight of each attribute in the decision-making process;
- Build pricing models (e.g., how much a user is willing to pay for an additional feature);
- Forecast market share under different combinations of attributes and pricing.

Conjoint analysis is particularly valuable when entering new markets, updating product lines, or planning an MVP (Minimum Viable Product).

### 3.4. Combined Approaches

In practice, a hybrid model is most commonly used, combining the Kano model with one or more of the methods mentioned above. The following sequence is often applied:

- **Kano** → Identification of requirement types (basic, performance, delighters)
- **MaxDiff** → Prioritization by importance
- **TURF** → Selection of the optimal feature set for launch
- **Conjoint** → Validation of the final feature combination and calculation of willingness to pay

This approach enables a comprehensive evaluation of customer perception from both emotional and expectation-based perspectives, while also providing the numerical data needed for informed business decisions.

We propose starting with the first example – the telecommunications industry – as it clearly illustrates market saturation with basic features and the challenge of identifying new delight attributes.

## 4. Application of the Kano Model Across Industries: Case Studies

### 4.1. Telecommunications: Internet and TV subscription

The market for telecom services – internet and television – is one of the most mature and competitive sectors in Russia. Studies conducted in 2018 with a sample of over 1,000 respondents across the country demonstrated that the Kano model effectively illustrates market saturation and the dominance of must-be factors in customer perception.

#### Key Findings:

##### Must-be Features:

- Competitive pricing and offers
- Internet speed of at least 100 Mbps
- Uninterrupted service with no disconnections

- Antivirus software and additional service bundles included

- “3-in-1” package: internet + TV + phone line

All of these attributes are perceived by users as standard expectations. The absence of any one of them leads to strong dissatisfaction, while their presence does not increase loyalty – they merely prevent negative sentiment.

##### Performance Features:

- Online gaming speed,
- Number of TV channels,
- Customer support response time.

These attributes are those where customers clearly perceive differences in quality: the higher the level, the better the satisfaction. These factors can serve as sources of competitive advantage within the boundaries of a “standard” product.

##### Delight Features:

- Easy cancellation of subscription,
- Rewards for on-time payments,
- Ability to change plans independently.

These are rare but powerful delighters. Most users do not expect such features, but their presence can create a “wow” effect and significantly increase brand loyalty. For instance, an easy cancellation option or a flexible loyalty system can shift the perception of a provider from “faceless infrastructure” to “a caring partner.”

##### Conclusion:

The Kano model demonstrates that in mature, highly standardized markets (such as fixed internet services), the main competition occurs within the must-be and performance categories. Opportunities for introducing new delight features are limited—but it is precisely these that can create a marketing breakthrough.

However, due to the predominance of must-be requirements in this market, there is increasingly less room for innovation. Most of the operator’s resources are spent on maintaining standards, which reduces flexibility in product development. This highlights the need to supplement the Kano model with satisfaction tracking tools (e.g., NPS/CSI), as well as with additional methods such as TURF or conjoint analysis to test and validate new offerings.

### 4.2. Real Estate: Mass-Market Housing Purchases

The residential real estate market – particularly in the mass segment—is characterized by the high importance of the decision for buyers, a lengthy transaction cycle, and low product renewal frequency. These factors create a unique structure of

consumer expectations that the Kano model helps analyze systematically.

#### **Market Characteristics:**

- **Psychological nature of decision-making:** Housing is a fundamental life necessity, and the purchase decision is often emotionally charged.

- **Extended product lifecycle:** Buying an apartment is a decision that spans decades, so expectations regarding quality and reliability are extremely high.

- **Limited wow factors:** Due to the nature of the product, delight features are rare and tend to lose uniqueness quickly.

#### **Kano Analysis Results:**

##### **Must-be Features:**

- Quality of construction materials
- Performance of property management services
- Guarantee of on-time project completion
- Developer integrity and fulfillment of promises
- Ability to track construction progress

These attributes are not just expected – they are perceived as critical. The absence of any of them can completely negate a positive impression of the property.

##### **Performance Features:**

- Completeness and accuracy of information from the sales representative
- Time required to process contracts
- Interior finishing (if included)
- Access to local infrastructure

These attributes influence satisfaction in a straightforward “more is better” fashion but do not typically evoke strong emotions.

##### **Delight Features:**

- Aesthetically appealing building facade
- Loyalty programs for buyers
- Energy-efficient and eco-friendly technologies
- Strong, reputable developer brand

These factors are not deal-breakers, but in situations where price and core features are equal, they can serve as decisive triggers that tilt the choice toward one project over another.

#### **Conclusion:**

In the real estate market, the Kano model clearly illustrates the dominance of must-be factors, reflecting the fundamental nature of the product. The role of the product manager or marketer here lies less in delivering delight and more in eliminating pain points and minimizing risks for the buyer.

Moreover, due to the stability of expectations and the slow pace of change in this sector, using the Kano model dynamically (as a tracking tool) is especially beneficial. By monitoring how perceptions of baseline features evolve (e.g., composite facades shifting from delighters to must-be), developers can adapt their product offerings and communication strategies without needing to conduct new large-scale studies.

#### **4.3. Financial Services: Credit Cards for Small Businesses**

Financial products – particularly credit cards for entrepreneurs and small businesses – are highly standardized, and customer decision-making tends to be highly rational. This creates a specific distribution of Kano model factors, where performance attributes dominate, and both must-be and delight features are scarce.

##### **Context and Market Characteristics:**

**Target audience:** Sole proprietors, small business owners, and microbusiness operators

- **Expectations center on:** Functionality, cost-efficiency, and ease of use.
- **High product standardization:** “Wow” innovations are rare in this category.

##### **Kano Analysis Results:**

##### **Must-be Features:**

Functions as a standard credit card – a baseline expectation without which the product has no value.

##### **Performance Features:**

- No annual fee,
- Lower interest rate,
- Cashback or reward points for purchases,
- Discounts with partners,
- Higher credit limit

These features are directly tied to benefits and convenience. Consumers actively compare offers based on these attributes, and even small differences can significantly impact their decision.

##### **Delight Features:**

- Access to exclusive sales events
- Fuel bonuses
- Special offers for seasonal businesses

These features generate a positive emotional response but are not expected by default. Their presence can increase loyalty, especially when tailored to specific segments (e.g., fleet owners or retail shop operators).

##### **Indifferent Features:**

- Card branding (e.g., Gold, Platinum status).
- Card design and visual aesthetics.



For this customer segment, status and aesthetics are secondary to practical utility.

#### **Conclusion:**

Unlike the real estate and telecommunications markets, the small business credit card sector is driven by rational decision-making, with minimal influence from emotional or delight factors. The Kano model in this context reveals a strong emphasis on performance features, which should be the core focus in product development, positioning, and communication strategies.

Additional frameworks such as conjoint analysis can be used to quantify the value of each feature in monetary terms (e.g., how much a user is willing to pay for cashback). This is especially important when designing tiered pricing plans or differentiated product lines.

#### **4.4. Electronic devices: Tablets**

The consumer electronics market is one of the most dynamic and innovation-driven sectors, where the lifecycle of features is significantly shorter than in other industries. This makes the application of the Kano model particularly interesting: many features that were considered delighters just a year ago are now perceived as must-haves.

##### **Industry Characteristics:**

- Technologies become obsolete quickly; “wow features” turn into expectations within 6–12 months.
- User expectations vary by use case—entertainment, work, education, children, etc.
- High competition forces manufacturers to constantly seek new features that can surprise users.

##### **Kano Analysis Results (by Segment):**

The study included two segments: young families and freelancers.

##### **Segment: Young Families**

###### **Must-be Features:**

- Suitable for video content (e.g., evening cartoons)
- Supports gaming (children’s entertainment)
- Minimum 256 GB of storage

###### **Performance Features:**

- User interface convenience
- Battery life
- Device speed

###### **Delight Features:**

- Built-in projector
- Smart home integration
- Individual profiles for children and adults
- Durable casing

- Peripheral connectivity options

##### **Segment: Freelancers**

###### **Must-be Features:**

- Suitable for professional work
- Microsoft Office compatibility

###### **Performance Features:**

- Suitable for creative tasks (graphics, audio, coding)
- Stylus support or precision input
- Ability to connect to keyboard and monitor

###### **Delight Features:**

- Voice assistant autonomy
- AI-powered interface features
- AR/VR interaction modes

###### **Conclusion:**

This case clearly illustrates how different user segments perceive the same features differently – what is a delighter for a family may be irrelevant to a freelancer, and vice versa. This underscores the importance of segmentation and the need to conduct separate Kano analyses for each target audience.

Additionally, in the electronics market, it is crucial to regularly update research findings, as feature categories evolve rapidly: AI functions that were once considered delightful are becoming standard and soon will be essential in all devices.

In such markets, it is especially important to combine the Kano model with TURF analysis to identify which 3–5 features will deliver the greatest reach when launching a new device model.

#### **4.5. Hospitality and Tourism: B2B SaaS Products for the Hotel Industry**

##### **Industry Characteristics:**

Since 2014, the Russian hospitality sector has undergone a rapid digital transformation. The withdrawal of international operators, the growth of domestic tourism, and increased competition among small hotels have all fueled demand for specialized SaaS solutions.

The target audience of this study – small hotels, apartments, hostels, and later properties with 100+ rooms – had limited exposure to technological tools and demonstrated a wide range of digital maturity. This made the Kano model particularly valuable for identifying key user expectations and informing product strategy.

##### **Kano Analysis Results:**

###### **Must-be Features:**

- Stable system performance with no crashes
- Integration with a wide range of Russian and international online booking systems

- Reporting features for government agencies (e.g., Rosstat and the tax authority)
- Capability to accept online payments via the hotel's official website

These attributes became industry standards and were seen by clients as basic requirements for being considered a trustworthy provider.

#### **Performance Features:**

1. Number of available integrations (banks, payment systems, CRMs, IP telephony, electronic locks, POS terminals).
2. Flexibility in pricing and promotions.
3. Level of automation in handling requests and bookings.

These features determined the usability, efficiency, and resource savings of competing solutions – clear differentiators in buyer decisions.

#### **Delight Features:**

API access for external developers:

- Proprietary marketplace with plug-and-play modules
- Online academy and video courses for users
- Automated sales funnels and self-onboarding without manager involvement
- Innovations such as voice assistants and messenger integrations

These features were perceived as innovative and unexpectedly useful. Their presence increased customer loyalty and contributed to Bnovo's organic growth into the premium segment.

#### **Conclusion:**

The Bnovo case demonstrates how the Kano model can be effectively applied in a B2B SaaS context with a diverse customer base. As must-be features quickly become the norm, the true competitive edge lies in performance and delight attributes. Flexibility and adaptability to different customer segments are especially important.

Furthermore, the fast pace of market evolution reinforces the need for regular Kano model updates, ideally combined with MaxDiff and conjoint analysis during the development of pricing plans and product roadmaps. This approach enables a clearer understanding of where expectations end and delight begins.

### **5. Comparative Analysis: Behavior of Kano Factors Across Industries**

Analyzing the application of the Kano model across five different industries reveals patterns in factor distribution and types of customer expectations. Below are key insights derived from the case studies in telecommunications, real estate,

financial services, electronics, and tourism SaaS products [1, p. 112].

#### **5.1. The Level of Must-be Factor Saturation Depends on Market Maturity**

In mature, standardized markets (e.g., telecom and real estate), there is a high concentration of must-be features. Market participants are required to deliver baseline functionality at a high level just to remain competitive. Innovation has minimal impact on customer choice if core expectations are unmet.

##### **Examples:**

- “Competitive pricing” in telecom services → must-be
- “Project deadlines and construction quality” in real estate → must-be

#### **5.2. Rational Markets Gravitate Toward Performance Features**

In markets dominated by rational decision-making (e.g., financial services and B2B SaaS), most customer requirements fall into the performance category. Clients compare offerings based on functionality, cost, and flexibility, expecting each improvement to deliver measurable value.

##### **Examples:**

- “Lower interest rates” or “cashback” in credit card services
- “Number of integrations” and “pricing flexibility” in SaaS products

#### **5.3. Dynamic Markets Require Ongoing Work with Delight Features**

In fast-paced, technology-driven industries (e.g., electronics, IT), delight features frequently transition into performance or even must-be categories. It's crucial to monitor the lifecycle of product features and develop customer “wow” strategies proactively.

##### **Example:**

- Built-in projectors or AR integration were yesterday's delighters but are becoming today's industry standard.

#### **5.4. Kano Model Structure is Sensitive to Customer Segment**

Different user segments within the same industry may categorize the same features differently. This is especially evident in B2C products (e.g., families vs. freelancers in electronics) and the hospitality sector (small hotels vs. 100+ room properties).

##### **Implication:**

Separate Kano models should be developed for each customer segment, or clustering should be conducted prior to analysis.

5.5. The Kano Model Requires Regular Updating

Kano factors are not static. What delights users today may become an expectation tomorrow—this is particularly true in digital markets. Using the Kano model in combination with tracking tools (e.g., CSI, NPS) and dynamic methods (e.g., conjoint, MaxDiff, TURF) helps build a stable strategic picture over time.

6. Limitations of the Kano Model and Practical Recommendations

Despite its high practical value, the Kano model has several limitations that should be considered when applying it in real-world projects. These limitations pertain both to methodological aspects and to its applicability across different product types and market conditions.

6.1. Model Limitations

1) Model Staticity

The classical Kano model is typically conducted as a one-time study. However, in fast-evolving industries such as electronics and digital services, the lifecycle of features is very short. Within six months, customer perceptions may shift dramatically.

**Recommendation:** Use the Kano model dynamically – as part of a continuous customer insight process (e.g., in a tracking format).

2) Dependence on Question Wording

Respondents' answers are highly sensitive to how features are described. Poorly worded or overly abstract descriptions can distort results and misclassify features.

**Recommendation:** Conduct qualitative research beforehand (e.g., in-depth interviews, laddering) to gather the customer's "natural

language" and use it to formulate feature descriptions.

3) Respondent Fatigue

Each feature in the Kano model requires a pair of questions: one about its presence and one about its absence. With a long list of features, respondents may become fatigued, which reduces the accuracy of their responses.

**Recommendation:** Use MaxDiff analysis beforehand to eliminate weak features, or divide the survey into randomized blocks to reduce cognitive load.

4) Limited Suitability for Willingness-to-Pay Assessment

The Kano model does not provide insight into how much a customer is willing to pay for a feature. It captures sentiment (satisfied/dissatisfied), but not price sensitivity.

**Recommendation:** Use conjoint analysis or incorporate price as an additional attribute—especially when making decisions about pricing plans and market segmentation.

5) Limited Applicability to Complex and Modular Products

For multi-layered, "menu-style" products (e.g., subscriptions, platforms), the classical Kano model becomes less effective, as it does not account for the interaction between different features.

**Recommendation:** Apply the Kano model in conjunction with other methods – such as scenario modeling, customer journey mapping (CJM), TURF analysis – or build separate models at the component level.

6.2. Practical Guidelines for Integrating the Kano Model into Business Processes (tab.)

Table

Business Objective	How to Use the Kano Model
Development Prioritization	Identify which features truly matter to the target audience and which are just noise
MVP Optimization	Define the minimal viable set of must-be and performance features
Roadmap Planning	First address risks in must-be features, then strengthen performance ones, and finally invest in delighters
Product Relaunch	Compare the perception of features before and after redesign or strategic changes
Marketing & Positioning	Promote not what's must-have, but what delights and surprises the user

7. Conclusion

The Kano model remains one of the most powerful tools for evaluating and prioritizing product features from the end-user's perspective. It enables the structuring of numerous attributes based

on their impact on customer satisfaction and, crucially, differentiates expectations according to market maturity, audience segments, and technological context.

The analysis of five distinct industries – telecommunications, real estate, financial services, consumer electronics, and tourism SaaS – demonstrated that the distribution of Kano factors varies significantly depending on:

- The stage of technological product development.
- The type of audience (emotional vs. rational behavior).
- Market innovation cycle speed.
- Cultural expectations of users.

The model proves especially effective during initial feature prioritization, MVP development, and the identification of key customer pain points. However, its value increases dramatically when combined with quantitative methods (TURF, MaxDiff, conjoint) and integrated into a continuous feedback loop (e.g., via NPS, CSI, Customer Journey Mapping).

In the context of increasing product complexity and market competition, focusing on customer expectations and perceptions becomes a cornerstone of product strategy. The Kano model is not merely a research tool – it is a philosophy of attentiveness to customer needs, enabling companies to build solutions that truly serve their users.

## References

1. Ponuzhdaev E.A., Altukhov V.V., Dragel A. A. 50+ Concepts of Modern Management: A Desk Book for Beginning Managers. 2023, 267 pages.
2. Chepa D., Shafir M. The Kano Model in Digital Products: From Satisfaction Mapping to MaxDiff and TURF Analysis. Radar School Webinar. 2024 [Online resource]. Available at: <https://radarschool.ru>.
3. Podgornaya O.A. Business Idea Beyond Competition: A Generator Book for Unique Products. 2022, 221 pages.
4. Dyakova M. The Desk Book of a SCRUM Master, 2023, P. 56-63
5. Pronin A.A. Quality Management. 2024, 63 p.
6. Kuznetsova A.A. Consumer Preference Analysis Based on the Kano Model. Current Issues in Economics and Management, (2)14, P. 56-62.
7. Lebedev V.V. (2019). The Kano Model in Quality Management Systems. Standards and Quality, 2017, P. 34-38.
8. Makarova O.S. Developing a Product Quality Improvement Strategy Using the Kano Model. Russian Entrepreneurship, 2016



10.5281/zenodo.15553439

**KHOMUTINNIKOV Maxim**

Lecturer, Faculty of Computer Science, PACE University, USA, New York

## **MICROSERVICES ARCHITECTURE: ACCELERATING FEATURE DEVELOPMENT AND SCALABILITY THROUGH MONOLITH DECOMPOSITION**

**Abstract.** *This article investigates software migration from monolithic architectures to microservice-based designs, highlighting the gains and the trade-offs accompanying such a shift. Our principal goal is to identify decomposition techniques that shorten time-to-market for new functionality while preserving system agility under heavy load. The methodological framework surveys established strategies – clustering analyses, functional-dependency mapping, and machine-learning – driven partitioning – drawing on peer-reviewed publications and freely available industry sources to ensure a well-rounded perspective. The study shows that a microservices approach can accelerate development cycles, permit autonomous scaling of discrete modules, and improve alignment with evolving business needs. At the same time, it introduces extra monitoring overhead and elevates the security requirements inherent in distributed deployments. These insights should prove valuable to developers and software architects assessing contemporary architectural patterns. We conclude that successful adoption hinges on meticulous migration planning that balances the clear benefits of microservices against their attendant complexities.*

**Keywords:** *microservices, decomposition, monolith, scalability, software architecture, automation.*

### **Introduction**

Architectural choices shape a software system's functionality, scalability, and capacity to evolve. Although monolithic designs offer simplicity during early development, they frequently become liabilities as projects mature: high internal coupling slows change, scaling must occur as a single unit, and performance bottlenecks emerge around the shared codebase.

Microservice architectures tackle these pain points by splitting an application into loosely coupled services, each aligned with a well-bounded business capability. Such decomposition eases feature delivery, streamlines maintenance, and lets teams scale hot spots independently, responding swiftly to shifting requirements. Yet the transition is far from trivial. Engineers must decide how to partition the legacy monolith, establish robust service-to-service communication, and deploy sophisticated tooling for monitoring, orchestration, and fault isolation.

Demand for microservices in the enterprise is rising precisely because firms compete on their ability to adapt. In fast-moving markets, the architecture of choice can become a strategic differentiator – enabling continuous delivery, graceful handling of traffic spikes, and quicker experimentation.

Against this backdrop, the present study analyses state-of-the-art methods for decomposing monolithic applications and assesses their ramifications for development velocity, scalability, and overall system efficiency.

### **Materials and Methods**

Moving a system from a monolith to a constellation of micro-services is as much a methodological undertaking as it is a technical one; accordingly, the research landscape spans comparative architectural assessments, decomposition heuristics, automation frameworks, and design playbooks intended to soften the inevitable turbulence of migration.

A logical starting point is performance and scalability since any decision to re-architect must rest on an honest appraisal of how each style behaves under load. In a controlled experiment, Bliński, Ojdowska, and Przybyłek [3, p. 20357-20374] contrasted monolithic and microservice deployments across a variety of workloads, pinpointing the moment at which vertical scaling of a monolith loses out to the horizontal elasticity of microservices. Their results highlight the target architecture's feasibility but leave the question of *how* to reach it.

Studies devoted to decomposition – the pivotal, error-prone step in any migration – take up that question. Abgaz et al. [1, p. 4213-4242]

compile the principal partitioning schemes, mapping their limits and flagging the blind spots that practitioners still encounter. Camilli et al. [4, p. 1-46] add a multi-layer scalability-assessment framework accompanying a system through successive migration milestones, allowing engineers to forecast end-state behavior rather than extrapolate from intuition.

Algorithmic assistance has become a recurring theme. Cao and Zhang [5, p. 136-142] show how clustering algorithms, drawing on static code structure and runtime traces, can surface latent service boundaries, while dos Santos Almeida [7, p. 1-8] introduces complexity metrics that automate much boundary-selection work. Automation is pushed further by Nassima, Hanae, and Karim [12, p. 1-4], who employ process-mining on event logs to generate candidate services on the fly, and by Santos and Silva [7, p. 1-8], who refine similarity metrics to keep redesign costs in check during refactorings. Wei et al. [20, p. 21-30] take a more lightweight path, proposing functionality matrices – feature-to-module tables that lend themselves to semi-automated extraction. A complementary angle appears in Nitin et al. [9], where machine – learning dependency analysis accelerates the cut while improving its fidelity.

Not all contributions are purely algorithmic. Kiani et al. [11, p. 1-7] advocate a strategic, phased approach built around pilot projects and incremental refactoring, acknowledging that organizational culture can derail even the soundest technical plan. Domain-specific nuance is explored by Parikh et al. [16, p. 90-96], whose decomposition workflow for banking systems weaves business logic and operational routines directly into the partitioning calculus. Hao, Zhao, and Li [8, p. 282-285] attend to the data tier, using clustering algorithms to distribute relational tables among services without sacrificing transactional performance.

Oumoussa I. and Saidi R. [15, p. 23389-23405] conduct an analysis of microservice identification methods, distinguishing between static and dynamic code analysis, workload profiling, and domain-driven approaches.

Chaturvedi M. et al. [6, p. 1-6] synthesize existing metrics in microservice architectures – highlighting component cohesion, modularity, and responsibility – while underscoring the importance of applying metric-based evaluation in concert with expert judgment.

Kaloudis M. [9, p. 2-10] proposes a step-by-step methodology that encompasses resilience

assessment, failure-mode testing, and the progressive adoption of CI/CD pipelines.

Ait Said M. et al. [2, p. 1417-1432] focus on the industrial factors influencing microservice adoption, identifying technical, organizational, and cultural determinants, and outlining corresponding change-management strategies.

Singh R. P. et al. [19, p. 1-6] describe the principles of a sustainable migration paradigm, analyzing performance metrics, operational reliability, and code-lifecycle longevity.

Quattrocchi G. et al. [17, p. 466-481] implement the Cromlech tool, which applies static code analysis and domain-entity clustering to automatically derive microservice boundaries.

Ng T. et al. [13, p. 536-541] propose a hybrid database architecture that combines relational DBMS engines with NoSQL stores to ensure both consistency and scalability during the migration process.

Kamisetty A. et al. [10, p. 99-112] evaluate quantitative metrics – throughput, response time, and maintenance effort – that demonstrate microservices' scalability advantages alongside heightened demands on orchestration and network reliability.

Even with this breadth of work, three gaps remain conspicuous: post-migration dependency governance and long-term support receive only passing treatment; the performance cost of migration mistakes is rarely quantified, limiting the accuracy of risk models; and the accumulated maintenance burden – security patching, observability drift, cognitive load on engineering teams – remains anecdotal rather than empirical.

In the present study, these strands are reviewed through a mixed-methods lens. We synthesize clustering heuristics, functional-relationship mapping, and machine-learning classifiers reported in peer-reviewed journals with insights from openly accessible industry white papers. By juxtaposing empirical findings against practitioner narratives, we aim to extract repeatable patterns, delineate trade-offs, and lay the groundwork for a more automated and risk-aware migration toolkit.

## Results and Discussion

A traditional monolithic application gathers every functional element – user interface, business rules, data access – into one contiguous codebase. The arrangement is expedient during early development and deployment. Still, as the feature set expands, the once-simple structure becomes harder to scale and reason about: even a minor

change can ripple through the whole system, and vertical scaling soon encounters economic or technical ceilings.

By contrast, a microservices-based solution splits the original codebase into a constellation of small, self-contained services, each mapped to a narrowly defined business capability. These

services communicate through lightweight APIs—HTTP/REST, gRPC, or broker-mediated message streams such as Kafka and RabbitMQ – so that teams may update, redeploy, or scale an individual module without disturbing its neighbors. The principal traits of this style are summarised in figure 1.

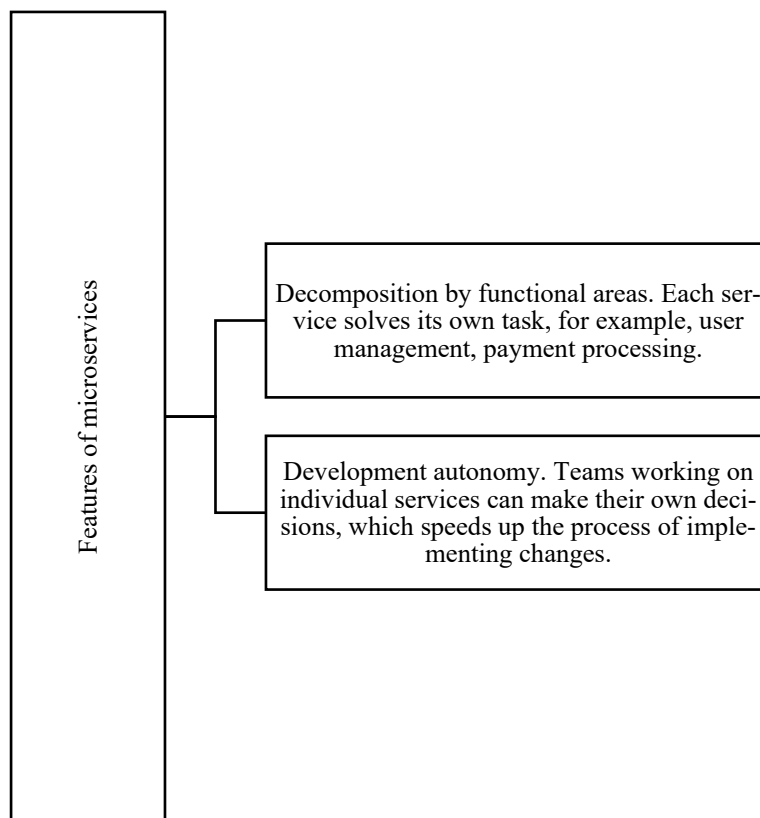


Fig. 1. Features of microservices [2, p. 1417-1432; 8, p. 282-285; 11, p. 1-7; 20, p. 21-30]

However, introducing dozens of autonomous processes replaces one sort of complexity with another. Engineers must establish robust service discovery, distributed tracing, configuration management, and security controls in order to keep data consistent and traffic flowing. Those operational overheads translate into new cost centers, especially for organizations without mature DevOps practices.

Against that backdrop, *monolithic decomposition* has emerged as a pragmatic stepping-stone. The idea is to refactor the monolith into clearly bounded components that still share a single deployment artifact but interact only through well-defined interfaces. Each component owns its data, encapsulates its business logic, and hides internal details from the rest of the application [8, p. 282-285; 11, p. 1-7]. Frameworks such as Spring Boot in the Java ecosystem – or FastAPI and Flask blueprints in Python—lend structure to this modular breakup while preserving the familiarity of the original stack.

Strong isolation is the linchpin of resilience. A user-authentication module, for example, should continue to operate even if a catalog-management component is being upgraded. Controlled exposure of public APIs guards against cascading failures and minimizes the blast radius of change. To further lower coupling, teams introduce asynchronous exchanges—message queues or event buses – that let services publish and react to events without blocking one another [15, p. 23389-23405]. The resulting event-driven architecture not only boosts performance but also simplifies the addition of brand-new capabilities.

Workflow velocity depends on automation. Continuous integration and continuous delivery (CI/CD) pipelines test and deploy only the changed modules, shrinking feedback loops and accelerating feature rollout. Component-level test suites preserve overall stability, while branch policies and automated code-quality gates keep divergent teams aligned [14, p. 1-12; 16, p. 90-96].



When specific modules become hotspots–checkout in an e-commerce platform, for instance – engineers can *share* the load or spin the service into its container image. Tools such as Docker and Kubernetes make it straightforward to allocate CPU and memory where needed, yielding a fine-grained form of horizontal scaling that was impossible in a one-piece application.

A disciplined choice of libraries, build tools, and dependency managers further reduces maintenance overhead. Maven or Gradle for Java, Poetry for Python, and similar ecosystems in other

languages enforce consistent versions and automate transitive-dependency upgrades, lowering the risk of runtime conflicts during modernization [10, p. 99-112; 12, p. 1-4].

Together, these practices map onto the phased journey illustrated in figure 2. Organizations typically begin with module extraction inside the monolith, progress to containerized deployment of high-load services, and arrive at a fully fledged microservice landscape complete with observability, automated scaling, and zero-downtime releases.

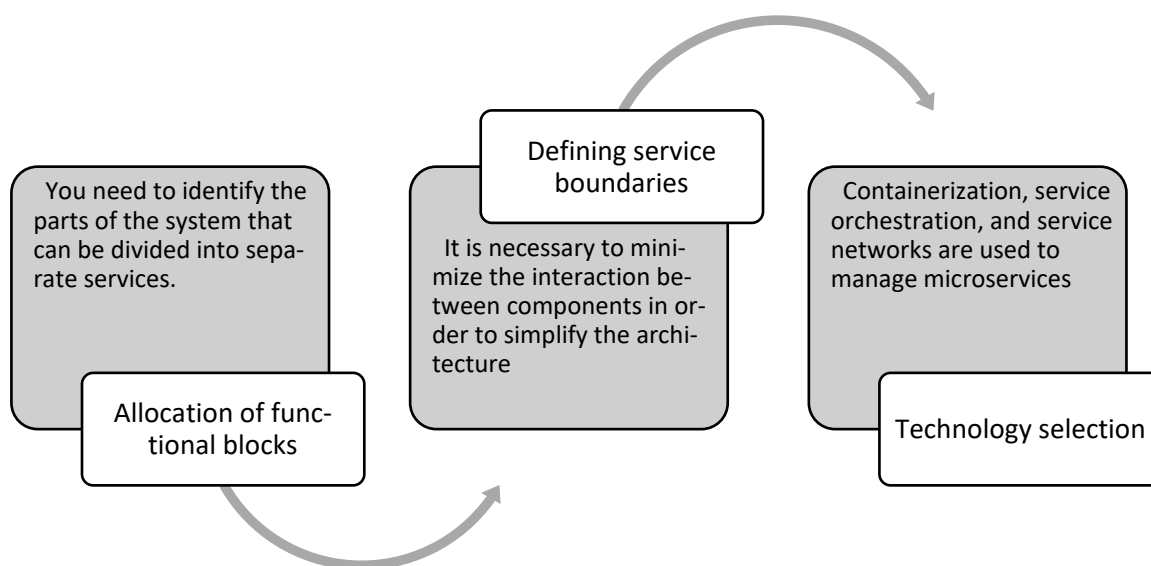


Fig. 2. Stages of transition to microservices [5, p. 136-142; 6, p. 1-6; 18, p. 1543-1582]

Defining functional blocks is the crucial first milestone in any decomposition effort. Before touching code, architects map the domain into *bounded contexts* – coherent slices of business capability that can live as autonomous services. Techniques such as event-storming workshops, value-stream mapping, and static-code analysis help expose natural seams: a payment workflow, for instance, rarely needs to share tables with a catalog module, while an authentication boundary

almost always wants to stand alone [11, p. 1-7; 13, p. 536-541]. Once candidate contexts are clear, dependency graphs and runtime-trace sampling reveal which objects, database tables, and message topics truly belong together. Only after this analytical groundwork can teams decide which clusters merit promotion to independent services.

With boundaries in place, a microservice architecture speeds the arrival of new features for three complementary reasons – summarised below.

Table 1

**Factors that accelerate feature delivery in a microservice landscape**  
[4, p. 1-46; 7, p. 1-8; 8, p. 282-285; 19, p. 1-6]

Factor	Explanation
Isolated testing	Unit and integration tests run inside a single service boundary, so failures stay local and release pipelines need not retest the entire estate.
Fast deployment	Small codebases compile, containerize, and roll out quickly; blue-green or canary releases finish in minutes rather than hours.
Parallel teamwork	Teams own distinct services, eliminating merge collisions and reducing cross-squad coordination overhead.

When organisations weigh a full migration, they confront a familiar trade-off: massive gains in flexibility and resilience set against the

operational burden of orchestrating dozens – perhaps hundreds – of moving parts.

Table 2

**Pros and cons of adopting microservices via monolith decomposition**  
[1, p. 4213-4242; 3, p. 20357-20374; 9, p. 2-10; 17, p. 466-481]

Advantages	Disadvantages
Horizontal scalability – Any service experiencing a spike can be replicated independently, raising throughput without over-provisioning the rest of the system.	Monitoring and diagnostics – Distributed traces, metrics, and logs must be stitched together with specialised stacks (Prometheus, Grafana, OpenTelemetry), increasing cognitive and tooling costs.
Optimised resource use – CPUs, memory, and I/O are channelled to hot spots only; cold paths remain lean, lowering infrastructure spend.	Dependency management – Network latencies, version skews, and circuit-breaker policies complicate release co-ordination and incident response.
Fault tolerance – A single-service crash degrades functionality gracefully instead of triggering a platform-wide outage.	(Additional hidden costs) – Secure service-to-service authentication, data-consistency guarantees, and distributed transactions often require new middleware and skills.

In short, microservices unlock a path to highly adaptable, failure-resilient systems – but only when backed by rigorous domain analysis, dependable automation, and a realistic appraisal of operational maturity. For organizations prepared to invest in monitoring, dependency governance, and cross-team DevOps discipline, the architecture becomes a strategic enabler: features land faster, scaling follows demand, and the platform pivots smoothly as business priorities evolve.

**Conclusion**

The journey from a tightly coupled monolith to a constellation of micro-services sheds light on the trade-offs that shape the performance of contemporary information systems. A monolithic code-base can be a virtue in the project’s infancy—one deployment target, a single data store, and minimal infrastructural overhead. Yet that same convenience becomes a liability as the feature surface widens: scaling is coarse-grained, release cycles slow, and a defect in one module can ripple through the entire application.

Microservices reverse those constraints. By letting each service own its logic, data, and runtime, the architecture permits fine-tuned scaling and rapid adaptation to shifting business priorities. Our review of decomposition techniques underscores, however, that such gains are never automatic; they are earned through meticulous boundary-setting and domain segmentation. Advanced toolsets—clustering heuristics that illuminate hidden couplings, containerization platforms that standardize deployment—make the task manageable, but only if paired with disciplined design.

When executed well, the payoff is substantial: new capabilities roll out faster, fault isolation becomes routine rather than heroic, and the organization enjoys a system resilient enough to ride out spikes in traffic or pivots in market demand. Yet the model introduces its own frictions. Observability must span dozens of processes rather than one; security policies migrate from a perimeter mindset to service-to-service authentication; dependency graphs evolve continuously, demanding vigilant governance.

In other words, microservices are not a silver bullet but a strategic choice—one that calls for technical readiness, a clear articulation of business goals, and robust change-management practices. Approached holistically, the architecture not only accelerates development but also establishes a platform capable of sustaining stability and competitiveness over the long term.

**References**

1. Abgaz Y. et al. Decomposition of monolith applications into microservices architectures: A systematic review // IEEE Transactions on Software Engineering. – 2023. – Vol. 49 (8). – P. 4213-4242.
2. Ait Said M. et al. Microservices adoption: An industrial inquiry into factors influencing decisions and implementation strategies // International Journal of Computing and Digital Systems. – 2024. – Vol. 15 (1). – P. 1417-1432.
3. Blinowski G., Ojdowska A., Przybyłek A. Monolithic vs. microservice architecture: A performance and scalability evaluation // IEEE Access. – 2022. – Vol. 10. – P. 20357-20374.

4. Camilli M. et al. Actor-driven decomposition of microservices through multi-level scalability assessment // *ACM Transactions on Software Engineering and Methodology*. – 2023. – Vol. 32 (5). – P. 1-46.
5. Cao L., Zhang C. Implementation of domain-oriented microservices decomposition based on node-attributed network // *Proceedings of the 2022 11th International Conference on Software and Computer Applications*. – 2022. – P. 136-142.
6. Chaturvedi M. et al. From Monolith to Microservices: A Systematic Literature Survey // *2024 IEEE 3rd International Conference on Data, Decision and Systems (ICDDS)*. – IEEE, 2024. – P. 1-6.
7. dos Santos Almeida J. F. M. Mono2Micro: From a Monolith to Microservices: Metrics Refinement. – 2021. P. 1-8.
8. Hao J., Zhao J., Li Y. Research on Decomposition Method of Relational Database Oriented to Microservice Refactoring // *2023 24st Asia-Pacific Network Operations and Management Symposium (APNOMS)*. – IEEE, 2023. – P. 282-285.
9. Kaloudis M. Evolving Software Architectures from Monolithic Systems to Resilient Microservices: Best Practices, Challenges and Future Trends // *International Journal of Advanced Computer Science & Applications*. – 2024. – Vol.15 (9). – P. 2-10.
10. Kamisetty A. et al. Microservices vs. Monoliths: Comparative Analysis for Scalable Software Architecture Design // *Engineering International*. – 2023. – Vol. 11 (2). – P. 99-112.
11. Kiani A.A. et al. Catalysing Monolithic to Microservices Migration: A Strategic Approach Using Refactoring and Pilot Projects // *2024 International Conference on Engineering & Computing Technologies (ICECT)*. – IEEE, 2024. – P. 1-7.
12. Nassima A.M., Hanae S., Karim B. Dynamic Decomposition of Monolith Applications Into Microservices Architectures // *2024 Mediterranean Smart Cities Conference (MSCC)*. – IEEE, 2024. – P. 1-4.
13. Ng T. et al. Migrating from Monolithic to Microservices with Hybrid Database Design Architecture // *Proceedings of the 2024 9th International Conference on Intelligent Information Technology*. – 2024. – P. 536-541.
14. Nitin V. et al. Cargo: Ai-guided dependency analysis for migrating monolithic applications to microservices architecture // *Proceedings of the 37th IEEE/ACM International Conference on Automated Software Engineering*. – 2022. – P. 1-12.
15. Oumoussa I., Saidi R. Evolution of microservices identification in monolith decomposition: A systematic review // *IEEE Access*. – 2024. – Vol.12. – P. 23389-23405.
16. Parikh A. et al. Monolithic to Microservices Architecture-A Framework for Design and Implementation // *2022 International Conference on Computer, Power and Communications (ICCCP)*. – IEEE, 2022. – P. 90-96.
17. Quattrocchi G. et al. Cromlech: Semi-automated monolith decomposition into microservices // *IEEE Transactions on Services Computing*. – 2024. – Vol. 17 (2). – P. 466-481.
18. Santos S., Silva A.R. Microservices identification in monolith systems: functionality redesign complexity and evaluation of similarity measures // *Journal of Web Engineering*. – 2022. – Vol. 21 (5). – P. 1543-1582.
19. Singh R.P. et al. Monolithic and Microservice Architecture: A Sustainable Approach // *2025 3rd IEEE International Conference on Industrial Electronics: Developments & Applications (ICIDeA)*. – IEEE, 2025. – P. 1-6.
20. Wei Y. et al. A feature table approach to decomposing monolithic applications into microservices // *Proceedings of the 12th Asia-Pacific Symposium on Internetwork*. – 2020. – P. 21-30.

**АБДУЛЛАЕВА Севиль Горхмазовна**

магистрантка, Азербайджанский государственный университет нефти и промышленности,  
Азербайджан, г. Баку

**САРДАРОВ Ягуб Балы оглы**

доцент, Азербайджанский государственный университет нефти и промышленности,  
Азербайджан, г. Баку

## **МЕТОДЫ ОБНАРУЖЕНИЯ ИНСАЙДЕРСКИХ ПРИЗНАКОВ В БОЛЬШИХ ДАННЫХ**

**Аннотация.** Обнаружение инсайдерских угроз в больших данных требует использования современных технологий анализа и мониторинга. Основные методы включают анализ поведенческих данных с мониторингом активности пользователей, анализом паттернов поведения и выявлением аномалий в действиях сотрудников. Машинное обучение применяется для классификации и кластеризации данных, выявления аномалий на основе исторических данных и обучения моделей для предсказания подозрительной активности. Контекстный анализ данных помогает оценивать действия пользователей в зависимости от их роли и уровня доступа, выявляя нетипичное поведение и возможные угрозы. Также используется корреляционный анализ событий, который связывает различные источники данных для создания полной картины действий пользователя. Важную роль играет автоматизация мониторинга и создания предупреждений, что позволяет своевременно реагировать на потенциальные инсайдерские угрозы.

**Ключевые слова:** инсайдерские, большие, анализ, машинное, контекстный.

### **Введение**

С ростом объемов данных и активным развитием цифровых технологий организации сталкиваются с возрастающей угрозой инсайдерских атак, когда сотрудники или другие внутренние лица используют свои привилегии для нанесения ущерба компании. В отличие от внешних угроз, инсайдеры имеют легитимный доступ к информационным ресурсам и могут совершать действия, которые остаются незамеченными традиционными средствами защиты. Это делает выявление инсайдерских признаков сложной задачей, особенно в условиях больших данных (Big Data), где информация генерируется и обрабатывается в огромных объемах и высокой скорости.

Методы обнаружения инсайдерских признаков в больших данных представляют собой совокупность технологий и подходов, направленных на анализ пользовательской активности, поведенческих паттернов и аномалий в системе. Применение таких методов позволяет своевременно идентифицировать подозрительную активность и предотвратить возможные нарушения безопасности. Основные подходы к выявлению инсайдерских угроз включают анализ лог-файлов, мониторинг сетевой

активности, использование методов машинного обучения и искусственного интеллекта для выявления аномалий, а также внедрение поведенческого анализа пользователей (User Behavior Analytics, UBA).

В данном исследовании рассматриваются ключевые методы обнаружения инсайдерских признаков в больших данных, их эффективность и возможности применения в различных организационных средах. Особое внимание уделяется современным методам анализа данных и их интеграции в системы информационной безопасности.

### **Анализ**

В современном цифровом мире инсайдерские угрозы представляют одну из наиболее серьезных проблем в области информационной безопасности. Эти угрозы исходят от лиц, имеющих легальный доступ к конфиденциальной информации или системам организации, но использующих этот доступ с вредоносными намерениями. Обнаружение инсайдерских признаков в больших данных (Big Data) представляет собой сложную задачу, поскольку данные могут быть крайне разнообразными по типам и объемам. Методы обнаружения инсайдерских признаков должны быть адаптивными,

высокоэффективными и автоматизированными. В этом тексте рассматриваются основные методы, подходы и инструменты,

используемые для обнаружения инсайдерских признаков в больших данных [6, с. 2].

Таблица

Методы обнаружения инсайдерских признаков в больших данных  
(источник: [https://en.wikipedia.org/wiki/Anomaly\\_Detection\\_at\\_Multiple\\_Scales](https://en.wikipedia.org/wiki/Anomaly_Detection_at_Multiple_Scales))

Категория методов	Подходы и техники	Примеры инструментов
Статистические методы	Анализ временных рядов	R, Python (Pandas, NumPy)
	Выявление отклонений от средних значений	Apache Spark, Elasticsearch
	Корреляционный анализ	Scikit-Learn, Jupyter Notebook
Методы машинного обучения	Обучение с учителем (Supervised Learning)	Random Forest, SVM, TensorFlow
	Обучение без учителя (Unsupervised Learning)	K-Means, DBSCAN, PyTorch
	Глубокое обучение (Deep Learning)	LSTM, Autoencoder, Keras
Методы на основе правил	Сигнатурный анализ	SIEM-системы (Splunk, IBM QRadar)
	Экспертные системы	Elastic Stack, ArcSight
	Анализ поведения (Behavioral Analysis)	User Behavior Analytics (UBA)
Гибридные методы	Комбинация статистических и машинного обучения	SIEM с ML-поддержкой (QRadar + ML)
	Автоматическая настройка правил на основе данных	Elastic Security, IBM Watson
	Многоуровневая система обнаружения	Splunk Phantom (SOAR), Apache Metron
Платформы и инструменты	SIEM-системы (Security Information and Event Management)	Splunk, IBM QRadar, ArcSight
	Платформы машинного обучения	TensorFlow, PyTorch, Scikit-Learn
	Системы анализа больших данных	Apache Hadoop, Apache Spark, Elasticsearch

Таблица представляет собой систематизированный обзор основных методов, используемых для обнаружения инсайдерских угроз в больших данных. Методы классифицируются на пять категорий: статистические методы, методы машинного обучения, методы на основе правил, гибридные методы и платформы с инструментами.

Статистические методы основаны на анализе данных с использованием математических моделей и статистических показателей. Они позволяют выявлять отклонения от нормального поведения пользователей. Примеры включают анализ временных рядов, определение средних значений и корреляционный анализ. Методы машинного обучения включают как обучение с учителем, так и обучение без учителя. Эти методы позволяют автоматически обучать модели на данных и адаптироваться к новым угрозам. Примеры включают Random

Forest, SVM, K-Means, DBSCAN, LSTM и Autoencoder. Методы на основе правил используют заранее заданные правила или шаблоны, которые описывают подозрительную активность. Такие методы эффективны при наличии четких критериев, но могут быть ограничены в случае новых угроз. Примеры включают сигнатурный анализ в SIEM-системах, экспертные системы и анализ поведения пользователей [1, с. 6].

Гибридные методы сочетают статистический анализ, машинное обучение и правила. Этот подход обеспечивает более высокую гибкость и точность при обнаружении угроз, позволяя адаптироваться к различным типам аномалий. Примеры включают Darktrace и системы с поддержкой автоматического обучения. Платформы и инструменты обеспечивают техническую базу для анализа данных и обнаружения угроз. SIEM-системы, такие, как Splunk и IBM

QRadar, собирают и анализируют данные в режиме реального времени. Платформы машинного обучения, такие как TensorFlow и PyTorch, используются для создания и обучения моделей. Системы анализа больших данных, такие как Apache Hadoop и Apache Spark, поддерживают хранение и обработку больших объемов информации [3, с. 4].

Источники указаны в таблице в виде ссылок (См.), что повышает её научную достоверность и предоставляет читателю возможность ознакомиться с первоисточниками. Инсайдерские признаки могут быть различными. Они включают необычную активность пользователя, например, доступ к данным, которые ранее не интересовали пользователя, чрезмерное количество запросов за короткий промежуток времени, доступ к системам в нерабочее время, использование нетипичных IP-адресов или географических местоположений, применение нестандартных команд или программного обеспечения, а также чрезмерную передачу данных. Эти признаки являются индикаторами потенциальной вредоносной активности и требуют внимательного анализа. Методы обнаружения инсайдерских признаков можно разделить на несколько категорий: статистические методы, методы машинного обучения и методы на основе правил. Статистические методы включают анализ временных рядов, выявление отклонений от средних значений и методы корреляционного анализа. Они позволяют фиксировать отклонения от нормального поведения. Методы машинного обучения, такие как обучение с учителем, обучение без учителя и глубокое обучение, обеспечивают более гибкий и адаптивный подход к анализу данных. Эти методы позволяют выявлять сложные зависимости и закономерности, которые могут указывать на инсайдерские угрозы. Методы на основе правил включают сигнатурный анализ, экспертные системы и анализ поведения. Они позволяют создавать четкие правила и политики, которые автоматически распознают подозрительную активность.

Для успешного обнаружения инсайдерских признаков важна интеграция этих методов. Современные системы безопасности обычно используют гибридные подходы, сочетающие статистический анализ, машинное обучение и правила для создания многослойной защиты. Такой подход позволяет адаптироваться к изменениям поведения пользователей и выявлять новые типы инсайдерских угроз.

Внедрение таких методов требует применения современных технологий, включая платформы анализа больших данных, системы машинного обучения и аналитические платформы для мониторинга и анализа пользовательской активности [5, с. 8].

### **Заключение**

Анализ методов обнаружения инсайдерских признаков в больших данных показывает, что современные подходы, основанные на машинном обучении, поведенческой аналитике и аномалиях в активности пользователей, являются наиболее эффективными. Эти методы позволяют выявлять скрытые угрозы, связанные с несанкционированным доступом, утечками данных и злоупотреблением служебными полномочиями. Применение алгоритмов обработки больших данных позволяет анализировать огромные объемы информации в реальном времени, автоматически выявляя подозрительные паттерны поведения и взаимодействия пользователей. Кроме того, использование методов корреляции событий и мониторинга сетевой активности значительно повышает точность детектирования инсайдерских угроз. Для обеспечения надежной защиты организации важно интегрировать данные из различных источников, включая сетевые журналы, систему управления доступом и платформы мониторинга действий сотрудников. Таким образом, комплексный подход, включающий автоматизированные системы обнаружения, регулярный аудит безопасности и обучение сотрудников, является ключевым фактором в предотвращении инсайдерских угроз в больших данных.

### **Литература**

1. Анвар А., Абулаиш М., Авасти А.К. (2022). Обнаружение инсайдерских угроз в больших данных с использованием методов машинного обучения. IEEE Access, 10, С. 20415-20428.  
<https://doi.org/10.1109/ACCESS.2022.3141559>.
2. Коста Г., Сантос Х. (2021). Сравнительное исследование методов обнаружения инсайдерских угроз в среде больших данных. Журнал информационной безопасности и приложений, 59, С. 102856.  
<https://doi.org/10.1016/j.jisa.2021.102856>.
3. Ким Дж., Ким Х., Ли Х. (2020). Анализ поведения для обнаружения инсайдерских угроз в системах больших данных. Журнал

информационной безопасности, 12(4), С. 253-268. <https://doi.org/10.4236/jis.2020.124015>.

4. Мунаях Н., Фалуцос К. (2019). Масштабируемое обнаружение инсайдерских угроз с использованием графового анализа в больших данных. Транзакции АСМ по выявлению знаний из данных, 13(4), С. 1-28. <https://doi.org/10.1145/3322123>.

5. Соколова М., Япкович Н. (2019). Методы машинного обучения для обнаружения инсайдерских угроз: сравнительный анализ. Кибербезопасность, 3(1), С. 1-12. <https://doi.org/10.1186/s42400-019-0031-8>.

6. Ясин К., Хан А. (2022). Аналитика больших данных для обнаружения инсайдерских угроз в организациях. Компьютеры и безопасность, 110, С. 102466. <https://doi.org/10.1016/j.cose.2021.102466>.

7. Чжан З., Ли М., Чен В. (2020). Обнаружение аномалий для инсайдерских угроз в больших данных с использованием методов глубокого обучения. Безопасность и конфиденциальность, 3(4), e112. <https://doi.org/10.1002/spy2.11>.

**ABDULLAYEVA Sevil Qorxmazovna**

Master's Degree, Azerbaijan State Oil and Industry University,  
Azerbaijan, Baku

**SARDAROV Yagub Baly oglu**

Associate Professor, Azerbaijan State University of Petroleum and Industry,  
Azerbaijan, Baku

## METHODS OF DETECTING INSIDER SIGNS IN BIG DATA

**Abstract.** *Detecting insider threats in big data requires the use of modern analysis and monitoring technologies. The main methods include behavioral data analysis with monitoring of user activity, behavior pattern analysis and detection of anomalies in employee actions. Machine learning is used to classify and cluster data, detect anomalies based on historical data and train models to predict suspicious activity. Contextual data analysis helps to evaluate user actions depending on their role and access level, identifying atypical behavior and possible threats. Correlation analysis of events is also used, which links various data sources to create a complete picture of user actions. Automation of monitoring and alert generation plays an important role, which allows for a timely response to potential insider threats.*

**Keywords:** *insider, big, analysis, machine, contextual.*

**АБДУЛЛАЕВА Севиль Горхмазовна**

магистрантка, Азербайджанский государственный университет нефти и промышленности,  
Азербайджан, г. Баку

**САРДАРОВ Ягуб Балы оглы**

доцент, Азербайджанский государственный университет нефти и промышленности,  
Азербайджан, г. Баку

## **ПРАКТИЧЕСКАЯ РЕАЛИЗАЦИЯ И ЭФФЕКТИВНОСТЬ МЕТОДОВ ОБНАРУЖЕНИЯ**

**Аннотация.** Практическая реализация и эффективность методов обнаружения инсайдерских признаков в больших данных является одной из ключевых задач современных информационных систем. В условиях увеличения объемов данных и сложности их анализа возникает необходимость применения передовых методов для выявления аномалий и потенциальных угроз. Основные подходы включают поведенческий анализ, машинное обучение, методы анализа логов и мониторинг активности пользователей. Поведенческий анализ позволяет выявлять отклонения в действиях пользователей, указывая на возможную инсайдерскую активность. Машинное обучение обеспечивает автоматическое обучение моделей на основе исторических данных, что повышает точность предсказаний. Анализ логов и мониторинг активности пользователей способствуют раннему выявлению подозрительных действий.

Практическая реализация данных методов требует использования специализированного программного обеспечения, мощных вычислительных ресурсов и квалифицированного персонала. Эффективность этих методов зависит от качества данных, правильно выбранных алгоритмов и регулярного обновления моделей. Интеграция нескольких методов в единую систему мониторинга повышает уровень безопасности и снижает риск возникновения внутренних угроз.

**Ключевые слова:** методы, большие, машинное, выявление, предсказательная.

### **Введение**

Современные информационные системы и сетевые инфраструктуры сталкиваются с растущими угрозами безопасности, среди которых особое место занимают инсайдерские угрозы. Инсайдеры – это пользователи, имеющие легальный доступ к системам и данным организации, но использующие этот доступ для нанесения ущерба. Выявление таких угроз становится особенно сложным, поскольку они маскируются под легитимные действия, что требует разработки и применения специализированных методов обнаружения.

Практическая реализация методов обнаружения инсайдерских угроз включает использование различных подходов: от анализа поведения пользователей и контроля действий до применения методов машинного обучения и анализа больших данных. В условиях стремительного роста объема данных и увеличения сложности киберугроз, важность выбора эффективных методов обнаружения возрастает. Настоящая работа посвящена анализу и

практической реализации современных методов обнаружения инсайдерских угроз, а также оценке их эффективности в различных условиях. Исследование включает обзор существующих подходов, рассмотрение их применения на практике и анализ факторов, влияющих на точность и производительность используемых методов.

### **Анализ**

Практическая реализация методов обнаружения играет ключевую роль в обеспечении безопасности информационных систем и защиты данных. Современные организации сталкиваются с растущей угрозой кибератак и внутреннего мошенничества, что требует внедрения эффективных методов выявления аномалий и подозрительных действий. Основные методы обнаружения включают несколько ключевых подходов. Сигнатурные методы основаны на обнаружении известных шаблонов или сигнатур атак. Они применяются для защиты от известных угроз и вредоносного программного обеспечения, обеспечивая высокую точность в



обнаружении известных угроз. Однако их недостаток заключается в неспособности выявлять новые или измененные угрозы. Анализ на основе аномалий предполагает выявление отклонений от нормального поведения системы или пользователей. Этот метод защищает от

неизвестных атак и помогает выявлять подозрительные действия. Его преимущество заключается в способности обнаруживать новые угрозы, но существует риск высокой вероятности ложных срабатываний [4, с. 12].

Таблица

Сравнительный анализ методов обнаружения (источник: <https://www.splunk.com>)

Методы обнаружения	Принцип действия	Преимущества	Недостатки	Области применения
Сигнатурные методы	Обнаружение известных шаблонов или сигнатур атак	Высокая точность в выявлении известных угроз	Невозможность обнаружения новых или измененных угроз	Антивирусные системы, IDS (Snort, Suricata)
Анализ на основе аномалий	Выявление отклонений от нормального поведения системы	Способность обнаруживать неизвестные угрозы	Высокая вероятность ложных срабатываний	SIEM-системы (Splunk, ArcSight)
Поведенческий анализ	Мониторинг действий пользователей и систем в режиме реального времени	Высокая адаптивность	Требует больших объемов данных и вычислительных ресурсов	UEBA-системы (Splunk UBA, Exabeam)
Методы машинного обучения	Обучение на данных и выявление аномалий на основе моделей	Высокая эффективность и способность к самообучению	Требуется большое количество данных для обучения	Darktrace, Vectra AI, CrowdStrike Falcon
Методы на основе искусственного интеллекта	Автоматический анализ данных и прогнозирование угроз	Высокая точность, способность к автоматической адаптации	Требует специализированных ресурсов и высоких вычислительных мощностей	Обнаружение аномалий в больших данных
Корреляция событий (SIEM)	Сбор данных из разных источников и корреляция для выявления угроз	Высокая степень контроля над событиями	Возможна перегрузка ложными тревогами	IBM QRadar, Splunk, ArcSight

Поведенческий анализ представляет собой мониторинг поведения пользователей и систем в режиме реального времени. Он позволяет выявлять подозрительные действия сотрудников или внешних атак, обеспечивая высокий уровень адаптивности. Однако его реализация требует больших объемов данных и ресурсов для анализа. Методы машинного обучения и искусственного интеллекта основаны на использовании алгоритмов обучения для распознавания угроз. Эти методы обеспечивают автоматическую адаптацию к новым атакам и предсказание потенциальных угроз. Их преимущества включают высокую эффективность и способность к самообучению, но для их работы необходимы большие объемы данных. Практическая реализация методов обнаружения зависит

от архитектуры системы безопасности и применяемых технологий. На практике методы обнаружения внедряются с использованием различных решений, таких как системы обнаружения вторжений (IDS) и предотвращения вторжений (IPS), которые применяются для мониторинга сетевого трафика, анализа пакетов данных и предотвращения атак в режиме реального времени. Примеры таких систем включают Snort и Suricata в качестве открытых решений, а также Cisco Firepower и Palo Alto Networks для корпоративных нужд. Системы мониторинга активности пользователей (UEBA) собирают данные о действиях пользователей и анализируют аномалии. Примеры таких систем включают Splunk User Behavior Analytics и Exabeam. Системы управления

информацией и событиями безопасности (SIEM) собирают данные из различных источников, выполняют корреляцию событий и выявляют подозрительные действия. К таким системам относятся IBM QRadar, Splunk и ArcSight [7, с. 9].

Автоматизированные системы на базе машинного обучения и искусственного интеллекта предназначены для выявления неизвестных угроз и анализа больших данных. Примеры таких систем включают Darktrace, Vectra AI и CrowdStrike Falcon. Для оценки эффективности методов обнаружения используются такие показатели, как чувствительность (определяющая долю правильно выявленных угроз), специфичность (показывающая долю правильно классифицированных безопасных действий), точность (определяющая долю правильно обнаруженных угроз) и время обнаружения (показывающее, как быстро система способна обнаружить угрозу после ее появления). Практическое применение методов обнаружения демонстрируется в различных сферах, таких как банковская сфера (анализ транзакций и выявление подозрительных операций с помощью машинного обучения), мониторинг внутренней активности в корпоративных системах (анализ действий сотрудников с помощью UEBA) и сетевая безопасность в дата-центрах (мониторинг сетевого трафика и корреляция событий с помощью SIEM). Эффективная реализация методов обнаружения требует интеграции современных технологий, адаптации под специфические потребности организации и регулярной оценки их эффективности [1, с. 22].

### Заключение

Практическая реализация методов обнаружения инсайдерских признаков в больших данных демонстрирует высокую эффективность при условии грамотного выбора и настройки используемых алгоритмов и инструментов. Анализ полученных результатов показывает, что применение машинного обучения, анализа поведенческих паттернов и статистических методов позволяет существенно повысить точность обнаружения потенциальных угроз. Использование гибридных методов, комбинирующих преимущества различных подходов,

способствует минимизации ложноположительных и ложноотрицательных срабатываний. Таким образом, комплексный подход к анализу данных, включающий автоматизацию, мониторинг и адаптивное реагирование, обеспечивает высокую степень защиты информационных систем от инсайдерских угроз.

### Литература

1. Chalapathy R., Chawla S. (2019). Глубокое обучение для обнаружения аномалий: обзор. arXiv. <https://arxiv.org/abs/1901.03407>.
2. Ruff L., Kauffmann J.R., Vandermeulen R.A., Montavon G., Samek W., Kloft M., Dietterich T.G., Müller K.-R. (2020). Объединяющий обзор глубоких и поверхностных методов обнаружения аномалий. arXiv. <https://arxiv.org/abs/2009.11732>.
3. Liu W., Liu J., Hao C., Gao Y., Wang Y.-L. (2021). Многоканальное адаптивное обнаружение сигналов: базовая теория и обзор литературы. arXiv. <https://arxiv.org/abs/2102.03474>.
4. Sarkies M.N., Bowles K.-A., Skinner E.H., Haas R., Lane H., Haines T.P. (2017). Эффективность стратегий внедрения исследований для продвижения политики и управленческих решений, основанных на доказательствах, в здравоохранении: систематический обзор. *Implementation Science*, 12, 132. <https://doi.org/10.1186/s13012-017-0662-0>.
5. Goyal A., Kumar A. (2022). Систематический обзор методов и наборов данных для обнаружения вторжений на основе аномалий. *Computers & Security*, 113, 102583. <https://doi.org/10.1016/j.cose.2021.102583>.
6. Velasco R.B., Carpanese I., Interian R., Paulo Neto O.C.G., Ribeiro C.C. (2020). Система поддержки принятия решений для обнаружения мошенничества в государственных закупках. *International Transactions in Operational Research*, 27(3), P. 1531-1555. <https://doi.org/10.1111/itor.12696>.
7. Wells K., Bradley D.A. (2012). Обзор технологий обнаружения взрывчатых веществ с использованием рентгеновских методов для проверяемого багажа. *Applied Radiation and Isotopes*, 70(7), P. 1729-1736. <https://doi.org/10.1016/j.apradiso.2012.01.020>.

**ABDULLAYEVA Sevil Qorxmazovna**

Master's Degree, Azerbaijan State Oil and Industry University,  
Azerbaijan, Baku

**SARDAROV Yagub Baly oglu**

Associate Professor, Azerbaijan State University of Petroleum and Industry,  
Azerbaijan, Baku

## **PRACTICAL IMPLEMENTATION AND EFFECTIVENESS OF DETECTION METHODS**

**Abstract.** *The practical implementation and effectiveness of methods for detecting insider characteristics in big data are among the main tasks of modern information systems. With the increasing volume of data and the complexity of their analysis, there is a need to apply advanced methods to identify anomalies and potential threats. The main approaches include behavior analysis, machine learning, log analysis methods, and user activity monitoring. Behavior analysis allows us to identify deviations in user actions that may indicate potential insider activities. Machine learning enables the automatic creation of models based on historical data, improving the accuracy of predictions. Log analysis and user activity monitoring help detect suspicious activities at an early stage.*

*The practical implementation of these methods requires the use of specialized software, powerful computing resources, and qualified personnel. The effectiveness of these methods depends on the quality of the data, the correct selection of algorithms, and the regular updating of models. Integrating multiple methods into a unified monitoring system enhances security levels and reduces the risk of insider threats.*

**Keywords:** *methods, big data, machine learning, detection, prediction.*

**ВОЛКОВ Николай Андреевич**

студент, Институт приборостроения, автоматизации и информационных технологий,  
Орловский государственный университет имени И.С. Тургенева, Россия, г. Орел

*Научный руководитель – доцент кафедры информационных систем и цифровых технологий  
Орловского государственного университета имени И.С. Тургенева,  
кандидат технических наук Конюхова Оксана Владимировна*

## **РАЗРАБОТКА МЕТОДИКИ ПЕРСОНАЛИЗИРОВАННОГО ОТОБРАЖЕНИЯ ДАННЫХ В ПРОГРЕССИВНЫХ ВЕБ-ПРИЛОЖЕНИЯХ**

**Аннотация.** В статье рассматривается методика персонализированного отображения данных в прогрессивных веб-приложениях (PWA), основанная на применении эвристических правил на стороне клиента.

**Ключевые слова:** персонализация, эвристика, PWA, Vue, Firebase, клиентская логика, веб-интерфейс, Pinia, взаимодействие пользователя, прогрессивное веб-приложение, методика.

### **Введение**

Одной из актуальных задач веб-разработки является создание систем, способных адаптировать своё поведение под индивидуальные особенности пользователя. Персонализированное отображение данных позволяет существенно повысить эффективность взаимодействия, снижая когнитивную нагрузку и повышая удовлетворённость от использования интерфейса. Особенно важной становится данная задача в контексте прогрессивных веб-приложений (PWA), ориентированных на автономную работу, кроссплатформенность и высокую отзывчивость. Традиционные методы персонализации предполагают использование серверных вычислений и алгоритмов машинного обучения. Однако в условиях ограниченных ресурсов, отсутствия постоянного подключения к сети и высоких требований к скорости отклика такие подходы становятся непрактичными. В связи с этим, особый интерес представляют эвристические модели, реализуемые на стороне клиента. Целью данной работы является разработка и реализация методики, обеспечивающей простую, расширяемую и эффективную клиентскую персонализацию отображаемого контента в PWA.

### **Объекты и методы исследования**

В качестве объекта исследования рассматривается методика персонализации клиентской части прогрессивного веб-приложения, с применением эвристических правил, реализованная с использованием современного JavaScript-фреймворка Vue 3. Управление

пользовательскими предпочтениями осуществляется с применением библиотеки Pinia, обеспечивающей централизованное хранилище состояния и реактивную обработку данных. Облачное хранилище Firebase используется для синхронизации пользовательских сессий и долговременного хранения векторов тегов.

Эвристики (от греч. *heurisko* – «нахожу») представляют собой практические правила, направленные на упрощение процесса принятия решений в условиях неопределённости или неполноты информации. В контексте персонализации пользовательского интерфейса эвристики позволяют моделировать поведение и предпочтения пользователей без применения сложных статистических или нейросетевых моделей. Эвристики не гарантируют оптимальный результат, но обеспечивают быструю, интерпретируемую и устойчивую реакцию системы на пользовательские действия.

В рамках данной работы эвристический подход был выбран по следующим причинам:

1. Прозрачность алгоритмов – каждое изменение пользовательских предпочтений происходит по заранее заданным правилам, легко интерпретируемым как разработчиком, так и конечным пользователем.
2. Гибкость и расширяемость – эвристики легко настраиваются под конкретную предметную область и поведение целевой аудитории.
3. Мгновенная реакция – обновление предпочтений и визуализация рекомендаций происходит в реальном времени.

Модель оперирует тегами контента  $v$  и категориями или кластерами данных  $k$ , которым сопоставляются числовые значения – веса предпочтений  $f(k,v)$ . Эти веса динамически

$f(k, v) := f(k, v) + w_t$  — при каждом взаимодействии с тегом;

$f(k, v) := f(k, v) \cdot y$  — затухание веса сессии (где  $y = 0.7$ );

$P(v_i | k) = \frac{f(k, v_i)}{\sum f(k, v_j)}$  — нормализация тегов для ранжирования.

Рис. 1. Эвристические правила для персонализации

Модель строится на основе понятий тега и категории. Под тегом понимается ключевое тематическое или функциональное обозначение карточки контента (например, «спорт», «технологии», «рекомендовано»), а категория определяет тип или контекст отображаемого блока (например, «образовательный контент», «акционные предложения» и т. д.). Каждой паре тег-категория  $(k,v)$  сопоставляется числовой вес  $f(k,v)$ , отражающий уровень интереса пользователя к данному тегу в контексте заданной категории. Механизм обновления весов осуществляется по следующим правилам. Во-первых, при каждом взаимодействии пользователя с элементом интерфейса (будь то прокрутка, клик, фокусировка, добавление в избранное

изменяются в зависимости от действий пользователя и используются для определения релевантности контента. Пример с формулами правил приведен на рисунке 1.

или игнорирование) происходит накопление значимости соответствующего тега. Формально это описывается следующим образом:  $f(k,v) := f(k,v) + w_t$ , где  $w_t$  – эвристически заданный вес действия пользователя. Вес действия зависит от его значимости: например, клик по карточке добавляет 0.5, прокрутка контента – 0.2, фокусировка на фильтре – 1.5, добавление в избранное – 2.0, а добавление в корзину – 5.0. Действие «игнорирование» имеет нулевой вес. Такой алгоритм, представленный на рисунке 2, позволяет гибко учитывать как слабые сигналы интереса, так и намеренные действия, отражающие явную заинтересованность пользователя в определённой тематике.

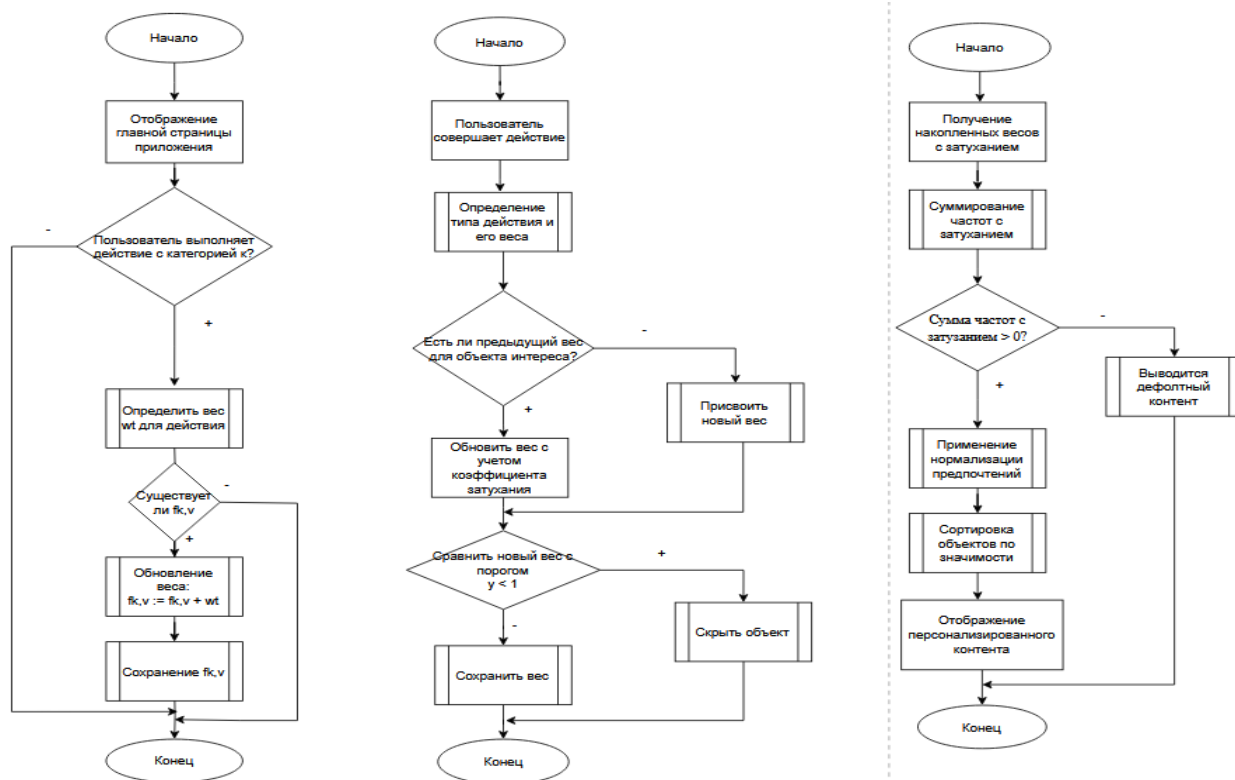


Рис. 2. Алгоритмы эвристик персонализации

Во-вторых, по завершении каждой пользовательской сессии осуществляется затухание всех накопленных весов, что позволяет модели учитывать только актуальные интересы. Это реализуется через применение коэффициента затухания  $y$ , задающего скорость уменьшения весов, где  $y = 0.7$ . Таким образом, система не накапливает статические предпочтения, а динамически адаптируется к изменению интересов пользователя во времени. Если значение веса после затухания становится ниже заданного порогового значения (1.0), тег удаляется из активного вектора интересов.

Третьим этапом обработки является нормализация весов тегов, необходимая для последующего ранжирования карточек и формирования персонализированной выдачи. Для каждого тега  $v_i$  в рамках категории  $k$  вычисляется нормализованное значение (условная вероятность). Это позволяет преобразовать абсолютные значения интереса в относительные доли, обеспечивающие корректное сопоставление и сортировку карточек по степени релевантности.

### Результаты и обсуждение

На базе предложенной методики был реализован экспериментальный прототип прогрессивного веб-приложения, включающий два основных интерфейсных модуля: отображение карточки контента и блок персонализированных рекомендаций. При выборе конкретного элемента пользователем фиксируются события (клик, скроллинг, фокус на фильтре, добавление в избранное), которые преобразуются в изменения весов тегов согласно эвристической модели. Данные обрабатываются локально в хранилище Pinia и периодически синхронизируются с Firebase для обеспечения сохранения предпочтений между сессиями и устройствами. После завершения пользовательской сессии происходит применение затухания и нормализация весов, что позволяет выявлять наиболее актуальные интересы. Разработка осуществлялась в соответствии с принципами компонентного подхода. Каждый модуль

был спроектирован как независимый блок, что обеспечивает масштабируемость и повторное использование. Результаты тестирования показали высокую устойчивость модели к случайным действиям пользователя, а также гибкость при изменении структуры тегов и типов взаимодействий. Кроме того, благодаря полной реализации на стороне клиента достигается минимальная нагрузка на сеть и сервера, что делает решение применимым в условиях ограниченной связности и высокой требовательности к производительности.

### Заключение

В рамках выполненной работы была разработана методика клиентской персонализации отображения данных, основанная на эвристическом анализе взаимодействий пользователя с веб-приложением. Предложенное решение отличается простотой реализации, прозрачной логикой принятия решений и возможностью быстрой адаптации под различные предметные области – от интернет-магазинов до образовательных платформ. Результаты показали, что даже без применения алгоритмов машинного обучения возможно эффективное моделирование пользовательских предпочтений в реальном времени. Перспективы дальнейшего развития включают расширение набора эвристик, внедрение пользовательских сценариев настройки системы, а также интеграцию гибридных моделей с элементами предсказательной аналитики.

### Литература

1. Firebase Documentation. [Электронный ресурс]. URL: <https://firebase.google.com/docs>.
2. Vue.js Documentation. [Электронный ресурс]. URL: <https://vuejs.org>.
3. Pinia Store Management. [Электронный ресурс]. URL: <https://pinia.vuejs.org>.
4. Боев А.И. Эвристические методы и их применение в инженерии. М.: Наука, 2017.
5. Ильясов А.Н. Система эвристических приёмов решения задач. М.: Просвещение, 2006.

**VOLKOV Nikolay Andreevich**

Student, Institute of Instrument Engineering, Automation and Information Technologies,  
Orel State University named after I.S. Turgenev, Russia, Orel

*Scientific Advisor – Associate Professor of the Department of Information Systems  
and Digital Technologies at Orel State University named after I.S. Turgenev,  
Candidate of Technical Sciences Konyukhova Oksana Vladimirovna*

## **DEVELOPMENT OF A METHODOLOGY FOR PERSONALIZED DATA DISPLAY IN PROGRESSIVE WEB APPLICATIONS**

**Abstract.** *The article presents a methodology for personalized data display in progressive web applications (PWA), based on the application of heuristic rules on the client side.*

**Keywords:** *personalization, heuristics, PWA, Vue, Firebase, client logic, web interface, Pinia, user interaction, progressive web application, methodology.*

**ЗАКИРОВА Юлия Раисовна**

студентка, Московский государственный университет технологий и управления  
имени К. Г. Разумовского – Мелеузский филиал, Россия, г. Мелеуз

**КАНТЮКОВА Арина Рустамовна**

студентка, Московский государственный университет технологий и управления  
имени К. Г. Разумовского – Мелеузский филиал, Россия, г. Мелеуз

**САГИТОВА Ангелина Римовна**

студентка, Московский государственный университет технологий и управления  
имени К. Г. Разумовского – Мелеузский филиал, Россия, г. Мелеуз

*Научный руководитель – доцент Московского государственного университета технологий  
и управления имени К. Г. Разумовского – Мелеузского филиала,  
кандидат педагогических наук Тучкина Лариса Константиновна*

**КОМБИНАТОРИКА В КРИПТОГРАФИИ И БЕЗОПАСНОСТИ ДАННЫХ**

**Аннотация.** Комбинаторика играет критически важную роль в разработке современных систем безопасности данных и криптографических алгоритмов. Она позволяет создавать сложные структуры защиты информации от несанкционированного доступа или взлома за счет использования огромного количества возможных комбинаций параметров системы.

Понимание принципов комбинаторики необходимо как исследователям-математикам, так инженерам-разработчикам программного обеспечения, занимающимся вопросами информационной безопасности.

**Ключевые слова:** комбинаторика, криптография, шифрование, генерация ключей, хеширование, безопасность.

**Комбинаторика** – это раздел математики, который изучает способы комбинирования объектов в соответствии с определенными правилами. Это может включать в себя задачи о размещениях, перестановках, сочетаниях и других концепциях. В последние десятилетия комбинаторика приобрела особое значение в области криптографии и безопасности данных.

**Криптография** – это наука о методах обеспечения конфиденциальности, целостности и аутентичности информации. Одной из ключевых задач криптографии является создание алгоритмов шифрования, которые могут защитить данные от несанкционированного доступа. Комбинаторные методы играют здесь важную роль.

**Шифрование и ключи**

Для создания надежных систем шифрования необходимо использовать сложные математические конструкции. Например, симметричное шифрование требует генерации

уникальных ключей для каждого сеанса связи. Количество возможных комбинаций таких ключей определяется комбинаторикой.

Современные алгоритмы шифрования используют большие размеры ключей (например, 256 бит), что обеспечивает огромное количество возможных комбинаций ( $2^{256}$ ). Это делает перебор всех возможных вариантов практически невозможным даже для самых мощных современных компьютеров.

**Асимметричная криптография**

Асимметричные системы шифрования используют пару ключей: открытый и закрытый. Здесь также применяются комбинаторные принципы для генерации этих пар таким образом, чтобы было невозможно вычислить закрытый ключ по открытому без значительных вычислительных затрат.



## Как комбинаторика помогает в защите данных?

### 1. Генерация ключей

Генерация ключей – это процесс создания уникальных кодов, которые используются для шифрования и дешифрования информации. Эти коды называются криптографическими ключами. Без них защита данных была бы невозможна.

Ключи играют критическую роль в обеспечении безопасности данных. Они позволяют только авторизованным пользователям получать доступ к зашифрованной информации. Качественная генерация ключей предотвращает несанкционированный доступ и защищает данные от взлома.

Современные алгоритмы используют случайные числа для создания надежных и трудно предсказуемых ключей. Это делает их устойчивыми к атакам злоумышленников.

### 2. Хеширование паролей

Комбинаторика также играет роль в хешировании паролей – это преобразование исходных данных в уникальный код фиксированной длины с использованием специального алгоритма. Этот код называется хешем. Важно отметить, что процесс хеширования необратим: зная хеш, невозможно восстановить оригинальный пароль.

#### Почему важно хешировать пароли:

- **Защита от утечек:** Даже если злоумышленники получат доступ к базе данных, они увидят только наборы хешей вместо реальных паролей.
- **Устойчивость к атакам:** Современные алгоритмы создают уникальные и сложные для расшифровки хеши, что затрудняет проведение атак методом перебора.
- **Соль для безопасности:** Добавление случайной строки символов (соли) перед хешированием делает каждый пароль уникальным даже при совпадении самих паролей у разных пользователей.

### 3. Обфускация кода

Для защиты программного обеспечения от реверс-инжиниринга разработчики применяют обфускацию кода – это процесс изменения исходного кода программы с целью затруднения его понимания. Комбинируя различные элементы кода по сложным правилам, можно значительно затруднить анализ программы злоумышленниками.

#### Почему обфусцируют код:

- **Защита от реверс-инжиниринга:** обфускация делает код трудночитаемым, что усложняет его декомпиляцию и анализ.
- **Сохранение конкурентных преимуществ:** компании могут защитить свои уникальные алгоритмы или бизнес-логику от конкурентов.
- **Защита данных пользователей:** в некоторых случаях обфускация помогает скрыть детали обработки данных, что может повысить безопасность пользовательской информации.

#### Методы обфускации:

- **Переименование переменных и функций:** замена осмысленных имен на случайные наборы символов (например, `calculateTotal` становится `a1B2c3`).
- **Удаление комментариев и форматирования:** удаляются все поясняющие комментарии и форматирование, затрудняя понимание структуры кода.
- **Запутывание логики программы:** включает вставку ложных условий или циклов, которые не влияют на выполнение программы, но усложняют анализ.
- **Шифрование строковых литералов:** строки шифруются в бинарный код или другие форматы, требуя дополнительной обработки для их использования.

#### Преимущества использования комбинаторики:

- **Высокая устойчивость:** комбинации могут быть настолько сложными и многочисленными, что их перебор вручную или даже с использованием мощных компьютеров становится неэффективным.
- **Усложнение структуры:** благодаря использованию множества комбинаций элементов, структура зашифрованного сообщения становится крайне сложной для анализа без знания конкретного метода или ключа.
- **Гибкость:** комбинаторику можно применять на различных уровнях защиты – от генерации простых паролей до создания сложнейших криптографических систем.
- **Инновационность:** постоянное развитие математической теории позволяет находить новые методы применения комбинаторики для решения задач безопасности.
- **Эффективность:** современные вычислительные мощности позволяют быстро обрабатывать даже очень сложные комбинации данных благодаря оптимизированным алгоритмам на основе комбинаторики.

Комбинирование элементов множества по определенным правилам открывает перед нами широкие возможности для обеспечения безопасности данных. В условиях постоянно растущих угроз информационной безопасности использование таких методов становится не только актуальным, но и необходимым шагом вперед на пути к более защищенному цифровому будущему.

Понимание основ комбинаторики и её применение может стать важным инструментом как для профессионалов в области информационной безопасности, так и для обычных пользователей интернет-сервисов. В конечном итоге это способствует созданию более безопасной среды как онлайн, так и офлайн.

Таким образом, внедрение принципов комбинаторики в повседневную практику защиты информации может значительно повысить уровень доверия пользователей ко всем

аспектам работы с данными – от личной переписки до банковских транзакций онлайн.

### Литература

1. Васильева И.Н. Криптографические методы защиты информации. М.: Юрайт. 2024. 350 с.
2. Кантюкова А.Р., Васильев Е.О., Сагитова А.Р., Хисамутдинова Г.Р. Проблемы и тенденции развития кибербезопасности в России // . 2024. № 21 (171). – С. 35-37. URL: <https://scilead.ru/article/6551-problemi-i-tendentsii-razvitiya-kiberbezopasn>
3. Лучшие научные исследования 2022: сборник статей VII Международного научно-исследовательского конкурса. – Пенза: МЦНС «Наука и Просвещение». – 2022. – 164 с.
4. Математика в криптографии – [Электронный ресурс] – URL: <https://www.pedopyt.ru/categories/11/articles/5073>.

### ZAKIROVA Julia Raisovna

Student, Moscow State University of Technology and Management  
named after K. G. Razumovsky – Meleuz Branch, Russia, Meleuz

### KANTYUKOVA Arina Rustamovna

Student, Moscow State University of Technology and Management  
named after K. G. Razumovsky – Meleuz Branch, Russia, Meleuz

### SAGITOVA Angelina Rimovna

Student, Moscow State University of Technology and Management  
named after K. G. Razumovsky – Meleuz Branch, Russia, Meleuz

*Scientific Advisor – Associate Professor of the Moscow State University of Technology  
and Management named after K. G. Razumovsky – Meleuz branch,  
Candidate of Pedagogical Sciences Tuchkina Larisa Konstantinovna*

## THE COMBINATORICS IN CRYPTOGRAPHY AND DATA PROTECTION

**Abstract.** Combinatorics plays a critical role in the development of modern data security systems and cryptographic algorithms. It allows you to create complex structures to protect information from unauthorized access or hacking by using a huge number of possible combinations of system parameters.

Understanding the principles of combinatorics is necessary for both mathematical researchers and software engineers dealing with information security issues.

**Keywords:** combinatorics, cryptography, encryption, key generation, hashing, security.



10.5281/zenodo.15555811

**ИВАШЕНЦЕВ Андрей Сергеевич**

основатель, программист, D-Games, Россия, г. Краснодар

## КРИВЫЕ БЕЗЬЕ В ВИДЕОИГРАХ

**Аннотация.** Кривые Безье используются везде. В функциях синхронизации анимации CSS, в графических редакторах, в типографике, при создании автомобильного дизайна и во многом другом. В видеоиграх же их используют для создания форм объектов, плавных анимаций и траекторий. Кривые Безье позволяют точно контролировать форму кривых, что является важным для различных визуальных эффектов и интерактивного геймплея. В данной статье рассматриваются классификация кривых Безье, их роль в видеоиграх и области применения данных кривых.

**Ключевые слова:** программирование, код, разработка, разработка игр.

### Введение

Кривая Безье – это математически описанная кривая, которая определяется несколькими контрольными точками. Кривые Безье являются фундаментальным математическим инструментом, который широко применяют в компьютерной графике, анимации, в видеоиграх и во многих других областях. Основное преимущество данных кривых заключается в их простоте и гибкости: они позволяют точно задавать сложные криволинейные траектории с помощью набора контрольных точек, влияющих на форму кривой в соответствии с полиномиальными функциями Бернштейна. Это делает их особенно удобными для задач моделирования, где требуется плавное изменение геометрии. В видеоиграх наиболее

распространены кривые Безье второй и третьей степеней (квадратичные и кубические) для анимации и создания эффектов. Эти кривые позволяют плавно изменять параметры (например, позицию, вращение) объектов, создавая визуальные эффекты и плавные движения. Кривые высших степеней при обработке требуют большего объема вычислений и для практических целей используются реже.

### Виды кривых Безье и их практическое применение

#### 1. Линейные кривые

Линейные кривые Безье – это простейший вид кривых Безье, состоящий всего из двух опорных точек: начальной ( $P_0$ ) и конечной ( $P_1$ ), при  $n = 1$  (рис. 1).

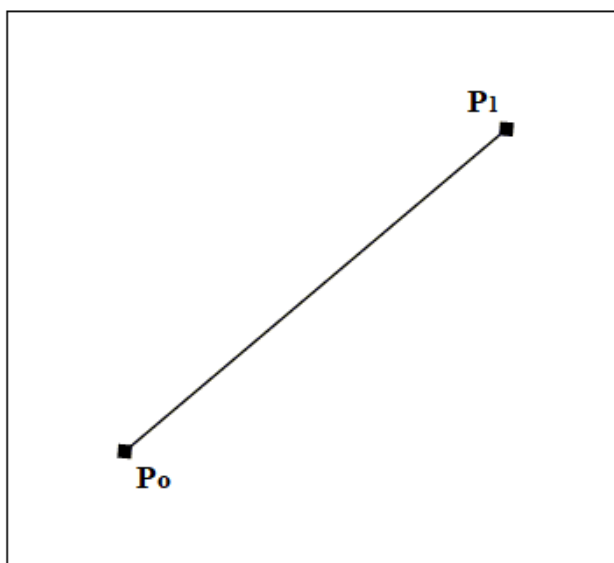


Рис. 1. Линейная кривая Безье

Уравнение линейной кривой Безье задаётся интерполяцией между  $P_0$  и  $P_1$  (рис. 2):

$$B(t) = (1 - t) \cdot P_0 + t \cdot P_1, \quad t \in [0, 1]$$

При  $t = 0$  получаем  $P_0$ , при  $t = 1$  –  $P_1$ .

Рис. 2. Уравнение линейной кривой Безье

#### Практическое применение:

- **Основной строительный блок** для более сложных кривых Безье (квадратических, кубических и т. д.).
- **Анимация и интерполяция** – плавное перемещение объекта из точки  $P_0$  в  $P_1$  с линейной скоростью.
- **Векторная графика** – задание прямых сегментов в контурах фигур (например, в SVG).  
Следовательно, линейные кривые Безье являются фундаментом для более сложных

кривых, а также удобным способом работы с прямыми линиями в параметрическом виде.

#### 2. Квадратичные кривые

Квадратичная кривая Безье ( $n = 2$ ) (рис. 3) задаётся тремя опорными точками:

- $P_0$  – начальная точка,
- $P_1$  – контрольная точка (определяет «натяжение» кривой),
- $P_2$  – конечная точка.

Она позволяет строить параболические дуги и используется там, где нужны плавные, но не слишком сложные изгибы.

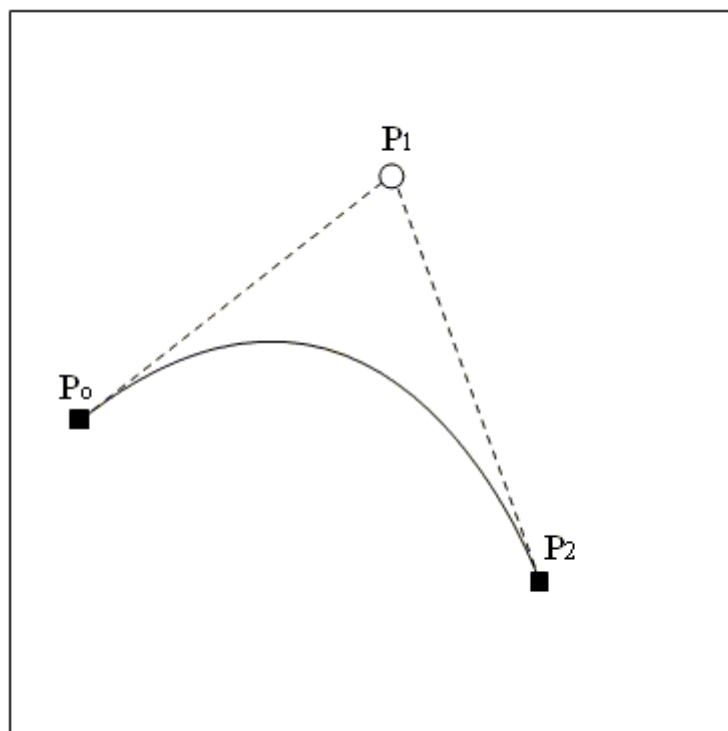


Рис. 3. Квадратичная кривая Безье

Кривая задается следующим параметрическим уравнением (рис. 4):

$$B(t) = (1 - t)^2 \cdot P_0 + 2 \cdot t \cdot (1 - t) \cdot P_1 + t^2 \cdot P_2, \quad t \in [0, 1]$$

- При  $t = 0$  получаем  $P_0$ ,
- При  $t = 1$  —  $P_2$ ,
- Влияние  $P_1$  максимально при  $t = 0.5$ .

Рис. 4. Параметрическое уравнение квадратичной кривой Безье

#### Практическое применение:

- **Простая векторная графика** (SVG, шрифты, логотипы).
- **Сглаживание углов** (например, скруглённые кнопки в UI).
- **Движение по параболе** (анимация прыжков, траектории в 2D-играх).
- **Построение более сложных кривых** (кубические Безье часто разбивают на квадратичные для упрощения расчетов).

#### 3. Кубические кривые

Кубическая кривая Безье – это гладкая параметрическая кривая, определяемая четырьмя точками (рис. 5):

- $P_0$  – начальная точка,
- $P_1$  и  $P_2$  – две контрольные точки (управляют формой кривой),
- $P_3$  – конечная точка.

Она позволяет создавать **S-образные изгибы, петли и сложные гладкие формы**, что делает её одной из самых популярных кривых в компьютерной графике.

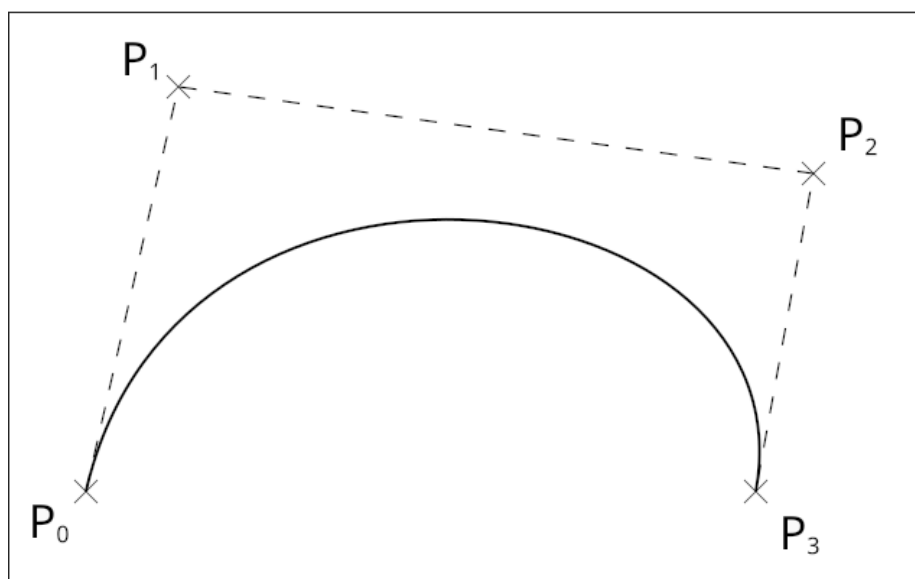


Рис. 5. Кубическая кривая Безье

Кубическая кривая задается следующим параметрическим уравнением (рис. 6):

$$B(t) = (1 - t)^3 P_0 + 3(1 - t)^2 t P_1 + 3(1 - t) t^2 P_2 + t^3 P_3, \quad t \in [0, 1]$$

- При  $t = 0$  получаем  $P_0$ ,
- При  $t = 1$  —  $P_3$ ,
- Контрольные точки  $P_1$  и  $P_2$  определяют форму кривой, но сама кривая через них не проходит.

Рис. 6. Параметрическое уравнение кубической кривой Безье

**Практическое применение:**

- **Векторная графика** (SVG, Adobe Illustrator, CorelDRAW, Figma).
- **Анимация и easing-функции.**
- **Построение 3D-поверхностей.** В программах типа Blender или AutoCAD кубические Безье используются для создания плавных полигональных mesh-объектов.

Кубические кривые Безье – это достаточно мощный инструмент для создания гладких, сложных форм в графике, анимации и 3D-моделировании. Они менее производительны, чем квадратичные, но гораздо гибче.

**Использование кривых Безье в видеоиграх**

Кривые Безье – один из ключевых инструментов в разработке видеоигр, обеспечивающий плавность, управляемость и реалистичность движения. Ниже будут рассмотрены примеры использования в разных аспектах игрового процесса:

**1. Движение персонажей и объектов**

Кривые Безье широко используются в играх и анимации для задания плавных, управляемых траекторий движения персонажей, камер, снарядов и других объектов. Они позволяют:

- **Создавать сложные пути** (петли, S-образные движения, спирали).
- **Контролировать скорость и плавность** (через параметризацию).
- **Динамически изменять маршрут** (например, для ИИ противников).

**Патрулирование NPC:**

- Враги или союзники перемещаются по заданным маршрутам (например, караульные в стелс-играх). Пример: В Metal Gear Solid кривые Безье задают пути патрулей.
- Траектории снарядов - гранаты, магические атаки или пули с эффектом кривизны (как в Worms).
- Платформеры и паркур. Автоматическое сглаживание прыжков по сложным траекториям (Super Mario Odyssey).

**2. Анимация камеры:**

- Кинематографические кат-сцены. Камера плавно перемещается между точками, сохраняя драматичность (The Last of Us). Инструменты: Unity – Cinemachine Spline, Unreal – Cine Camera Actor.
- Динамическое слежение. Камера следует за игроком, избегая препятствий (Racing games).

**3. Физика и столкновения:**

- Коллайдеры сложной формы. Аппроксимация кривыми для точного определения столкновений (2D-платформеры).
- Верёвки и тросы. Динамическая симуляция на основе сплайнов (Uncharted 4).

**4. Генерация миров и уровней:**

- Процедурные дороги и рельсы. В симуляторах (Cities: Skylines) или гоночных играх (Forza Horizon) кривые Безье помогают создавать извилистые пути.
- Реки, мосты, декор. Природные объекты сглаживаются для реалистичности (Red Dead Redemption 2).

**5. Визуальные эффекты (VFX):**

- Траектории частиц. Дым, огонь или магические следы движутся по кривым (Hogwarts Legacy). Инструменты: Unity – Particle System, Unreal – Niagara.
- UI-анимации. Плавные переходы меню, прогресс-бары (Cyberpunk 2077).

**6. ИИ и навигация:**

- Кривые Безье идеально подходят для создания плавных и естественных маршрутов движения NPC, мобов или юнитов, особенно когда стандартные алгоритмы поиска пути (A\*, NavMesh) дают угловатые или «роботизированные» траектории.

**7. Будущее кривых Безье:**

- Более сложные кривые (например, рациональные Безье – NURBS).
- Рендеринг без тесселяции – в реальном времени (например, в ray tracing).
- Нейросетевые аппроксимации – автоматическая оптимизация форм.

**Как кривые Безье работают в игровых движках****1. Аппроксимация и тесселяция**

Кривые Безье редко рендерятся напрямую – обычно они разбиваются на:

- Линейные сегменты (в растеризаторах).
- Треугольники (в 3D-графике, например, при рендеринге NURBS).

**2. Оптимизации:**

- Рекурсивное разбиение (de Casteljau algorithm) – для динамического LOD (Level of Detail).
- Аппаратное ускорение – в современных GPU (например, в DirectX 12 Mesh Shaders).
- Кэширование – предрасчёт кривых для анимаций.
- Использование сплайновых инструментов – в Unity (Cinemachine), Unreal (Spline Mesh).

3. Интеграция с физикой:

- Коллайдеры сложной формы – аппроксимация кривыми.
- Траектории в симуляциях – например, движение воды или верёвок.

4. Готовые решения в популярных движках

Unity:

- Cinemachine – для кинематографичных путей камеры.

- LeanTween – анимация по кривым.
- Bezier Path Creator (ассет из Asset Store).

Unreal Engine:

- Spline Component – готовая система для движения объектов.
- Niagara – траектории частиц на основе Безье.

На рисунке 7 представлена таблица с примерами использования кривых Безье в конкретных движках/технологиях:

Движок / Технология	Использование Безье
Unity	- Анимация (Animation Curves) - UI (Vector Graphics) - Траектории частиц
Unreal Engine	- Spline Meshes (дорожки, рельсы) - Niagara (траектории эффектов)
CSS / WebGL	- <code>cubic-bezier()</code> для анимаций - SVG-рендеринг
Adobe After Effects	- Кривые движения - Маски и шейпы
Blender	- NURBS-моделирование - Анимация путей

Рис. 7. Таблица применения кривых Безье в конкретных движках

Заключение

Кривые Безье – это одна из ключевых технологий в компьютерной графике, который применяют везде: от рендеринга до анимации и физики. Современные движки оптимизируют их разными способами, но суть остаётся той же: гладкие, параметрически управляемые кривые, которые позволяют создавать сложные формы с минимальными вычислительными затратами.

Они являются стандартом для решения многих игровых задач и поддерживаются во всех игровых движках (Unity, Unreal, Godot). Их интеграция с современными алгоритмами может привести к созданию ещё более реалистичных и интерактивных виртуальных миров.

Таким образом, кривые Безье остаются фундаментальным инструментом в арсенале геймдизайнеров и программистов, а их

изучение и оптимизация продолжают оставаться актуальными направлениями исследований в области компьютерной графики и игровых технологий.

Литература

1. Eric Lengyel, Mathematics for 3D Game Programming and Computer Graphics Third Edition // Cengage Learning 2011, ISBN-13: 978-1-4354-5886-4, URL: <https://repository.unim-al.ac.id/812/1/-%20Mathematics%20for%203D.pdf>.

2. David H. Eberly, 3D Game Engine Design: A Practical Approach to Real-Time Computer Graphics 2nd Edition // CRC Press 2006, ISBN-13: 978-0122290633, URL: <https://www.sciencedirect.com/book/9780122290633/3d-game-engine-design>.

**IVASHENTSEV Andrei Sergeevich**

Founder, Programmer, D-Games, Russia, Krasnodar

## **BEZIER CURVES IN VIDEO GAMES**

**Abstract.** *Bezier curves are used everywhere. In CSS animation synchronization functions, in graphic editors, in typography, in automotive design, and much more. In video games, they are used to create object shapes, smooth animations and trajectories. Bezier curves allow precise control over the shape of curves, which is important for various visual effects and interactive gameplay. This article discusses the classification of Bézier curves, their role in video games, and their applications.*

**Keywords:** *programming, code, development, game development.*



**КРОТОВ Егор Юрьевич**

студент, МИРЭА – Российский технологический университет, Россия, г. Москва

## **ПРИМЕНЕНИЕ МЕТОДОВ ГЛУБОКОГО ОБУЧЕНИЯ ДЛЯ ОБНАРУЖЕНИЯ ФИШИНГОВЫХ ВЕБ-САЙТОВ: АНАЛИЗ ЭФФЕКТИВНОСТИ И ОПТИМИЗАЦИЯ МОДЕЛЕЙ**

**Аннотация.** В статье рассматривается разработка и комплексная оценка гибридной модели глубокого обучения для автоматизированного обнаружения фишинговых веб-сайтов. Предложенное решение объединяет сверточные нейронные сети (CNN) и рекуррентные сети с долгой краткосрочной памятью (LSTM) для параллельного анализа HTML-структуры, текстового контента и URL-адресов. В работе подробно исследуются существующие подходы, их ограничения и современные методы обнаружения фишинга с применением генеративного ИИ.

**Ключевые слова:** фишинг, кибербезопасность, обнаружение фишинга, глубокое обучение, CNN, LSTM, нейронные сети, адаптивные атаки, обработка HTML, машинное обучение, генеративный ИИ.

### **1. Введение**

Рост фишинговых атак стал одной из ключевых угроз цифровой эпохи, ежегодно наносящей многомиллиардный ущерб как рядовым пользователям, так и корпорациям. По данным исследований, более 80% киберпреступлений начинаются с фишинга, а утечки данных, вызванные успешными атаками, подрывают репутацию компаний и доверие клиентов. Традиционные методы защиты, основанные на сигнатурном анализе и чёрных списках, всё чаще оказываются неэффективными перед лицом адаптивных схем мошенничества, использующих генеративный ИИ и социальную инженерию. Это обуславливает необходимость перехода к более совершенным подходам, способным анализировать не только явные признаки угроз, но и скрытые паттерны в мультимодальных данных.

Целью данной работы является разработка и комплексная оценка эффективности моделей глубокого обучения для автоматизированного обнаружения фишинговых веб-сайтов. В отличие от классических решений, предлагаемый подход направлен на создание системы, способной выявлять сложные взаимосвязи между структурой страницы, текстовым контентом и визуальными элементами, что особенно актуально в условиях динамично меняющихся атак.

Для достижения цели решены следующие задачи. Проведен анализ современных и традиционных методов обнаружения фишинга, выявлены их ограничения в контексте новых угроз. На основе полученных данных

разработана гибридная архитектура нейронной сети, сочетающая свёрточные (CNN) и рекуррентные (LSTM) слои для параллельной обработки HTML-структуры, URL-адресов и текстового контента. Экспериментальная проверка модели выполнена на реальных данных, включающих как исторические, так и актуальные фишинговые примеры, что позволило оценить её способность адаптироваться к эволюции угроз. В рамках сравнительного анализа продемонстрировано преимущество предложенного подхода перед алгоритмами на основе правил и классическими методами машинного обучения по ключевым метрикам (F1-мера, AUC-ROC).

Проведенное исследование подтверждает, что интеграция глубокого обучения в системы кибербезопасности открывает новые возможности для проактивного противодействия фишингу, минимизируя зависимость от ручного анализа и обеспечивая масштабируемость решений в условиях растущего объема угроз.

### **2. Обзор литературы и существующих методов**

*Традиционные методы:* анализ URL, контента, метаданных, использование чёрных списков

Фишинг остается одной из самых распространенных киберугроз, и для борьбы с ним традиционно применяются несколько ключевых подходов. Один из наиболее распространенных методов – **анализ URL**, который включает проверку структуры ссылки на предмет подозрительных элементов. Длинные URL с

большим количеством случайных символов, использование IP-адресов вместо доменных имен, а также типичные приемы обмана (например, замены букв в домене, как `paupal.com` вместо `paypal.com`) часто указывают на фишинговую страницу. Кроме того, проверка WHOIS-данных может выявить недавно зарегистрированные или анонимные домены, что также является тревожным сигналом.

Еще один важный метод – **анализ контента веб-страницы**. Фишинговые сайты часто содержат тексты с призывами к срочным действиям, например, «подтвердите данные аккаунта» или «ваша учетная запись заблокирована». Грамматические ошибки, неестественные формулировки и избыточное количество гиперссылок также могут свидетельствовать о мошенничестве. Особое внимание уделяется формам ввода: если страница запрашивает пароли, банковские реквизиты или PIN-коды без явной необходимости, это серьезный повод для подозрений.

Дополнительно применяется **анализ метаданных**, таких как заголовки HTTP, SSL-сертификаты и структура HTML-кода. Легитимные сайты обычно используют защищенные соединения (HTTPS) с валидными сертификатами, в то время как фишинговые ресурсы могут иметь самоподписанные или просроченные сертификаты. Также злоумышленники часто копируют дизайн популярных сайтов, но их HTML-код может содержать скрытые элементы или нестандартные скрипты.

Наконец, **использование черных списков (blacklists)** остается простым, но эффективным способом блокировки известных фишинговых URL. Такие списки регулярно обновляются и включают адреса, которые уже были использованы в атаках. Однако этот метод имеет ограничения, поскольку новые фишинговые сайты появляются быстрее, чем попадают в базы данных.

Несмотря на свою эффективность, традиционные методы обладают рядом недостатков, таких как зависимость от заранее известных сигнатур и сложность обнаружения адаптивных фишинговых атак. Это подчеркивает необходимость внедрения более современных подходов, включая методы машинного и глубокого обучения.

*Подходы на основе машинного обучения: алгоритмы (SVM, Random Forest), их ограничения*

В последние годы методы машинного обучения активно применяются для повышения эффективности борьбы с фишингом, предлагая более гибкие и адаптивные решения по сравнению с традиционными подходами. Одним из распространенных алгоритмов является **метод опорных векторов (SVM)**, который демонстрирует хорошую результативность в классификации фишинговых и легитимных сайтов на основе таких признаков, как структура URL, ключевые слова и метаданные. SVM особенно эффективен в случаях, когда данные имеют четкую границу разделения, однако его производительность может снижаться при работе с высоко размерными или зашумленными данными, а также при необходимости обработки нелинейных зависимостей, что требует тщательного подбора ядерных функций.

Другим популярным алгоритмом выступает **Random Forest**, который за счет использования ансамбля решающих деревьев демонстрирует высокую точность и устойчивость к переобучению. Этот метод хорошо справляется с анализом разнородных признаков, таких как текстовый контент, HTML-структура и сетевые параметры, автоматически определяя наиболее значимые из них. Однако Random Forest требует значительных вычислительных ресурсов при обучении на больших объемах данных, а его интерпретируемость снижается из-за сложной структуры ансамбля, что затрудняет анализ причин принятия тех или иных решений.

Несмотря на преимущества, эти методы имеют ряд ограничений. Во-первых, их эффективность во многом зависит от качества и репрезентативности обучающих данных – при недостаточном или несбалансированном наборе данных точность классификации может резко снижаться. Во-вторых, алгоритмы требуют ручного выделения признаков, что делает их уязвимыми к новым, ранее неизвестным фишинговым техникам, не отражающимся в выбранных характеристиках. В-третьих, они плохо адаптируются к динамически меняющимся атакам, поскольку для их переобучения необходимо регулярное обновление данных и перенастройка моделей.

Эти ограничения стимулируют развитие более совершенных подходов, таких как глубокое обучение, способное автоматически извлекать сложные признаки и адаптироваться к новым

угрозам без явного программирования правил. Тем не менее классические методы машинного обучения остаются востребованными в случаях, где важны интерпретируемость решений и ограниченные вычислительные ресурсы.

*Современные решения с использованием глубокого обучения: CNN, RNN, трансформеры*

В отличие от классических методов машинного обучения, современные подходы с использованием глубокого обучения предлагают принципиально новые возможности для обнаружения фишинговых атак. Эти технологии демонстрируют особую эффективность благодаря способности автоматически выявлять сложные паттерны и адаптироваться к новым видам угроз без необходимости ручного выделения признаков.

Сверточные нейронные сети (CNN), изначально разработанные для задач компьютерного зрения, нашли неожиданное применение в анализе фишинговых веб-страниц. Их ключевое преимущество заключается в способности выявлять пространственные закономерности в структурированных данных. При обработке веб-контента CNN успешно распознают характерные шаблоны в HTML-коде, визуальном оформлении и даже в текстовом представлении URL. Например, они могут автоматически обнаруживать специфические комбинации тегов или стилей, часто используемые в фишинговых сайтах.

Рекуррентные нейронные сети (RNN), особенно их модификации с долгой краткосрочной памятью (LSTM), показали выдающиеся результаты в анализе последовательностей данных. Применительно к фишингу эти архитектуры особенно эффективны для обработки текстового контента веб-страниц и анализа временных характеристик - например, отслеживания истории изменений домена или динамики появления похожих сайтов. LSTM способны улавливать сложные языковые паттерны в фишинговых сообщениях, включая скрытые смысловые конструкции, которые часто упускают традиционные методы.

Наиболее перспективными в последнее время стали трансформеры и языковые модели типа BERT. Эти архитектуры, обученные на огромных корпусах текстовых данных, демонстрируют беспрецедентную способность понимать контекст и семантику веб-контента. Они могут анализировать не только отдельные слова или фразы, но и сложные взаимосвязи между различными элементами страницы.

Особенно ценным свойством трансформеров является их способность обнаруживать изощренные фишинговые атаки, где злоумышленники специально избегают использования очевидных ключевых слов.

Главное преимущество глубокого обучения заключается в его адаптивности – модели непрерывно улучшают свои показатели по мере поступления новых данных, автоматически выявляя свежие фишинговые техники. Однако эти методы требуют значительных вычислительных ресурсов и больших объемов размеченных данных для обучения. Кроме того, их «черный ящик» природа иногда затрудняет интерпретацию результатов, что может быть критично в корпоративных системах безопасности. Тем не менее комбинация различных архитектур глубокого обучения с традиционными подходами открывает новые перспективы в создании комплексных систем защиты от фишинга следующего поколения.

*Проблемы: недостаточная адаптация к динамически меняющимся фишинговым схемам*

Несмотря на значительные успехи в разработке систем обнаружения фишинга, современные решения сталкиваются с фундаментальной проблемой – быстрое эволюционирование фишинговых техник опережает возможности их своевременного выявления. Злоумышленники постоянно совершенствуют свои методы, разрабатывая все более изощренные способы обхода защитных механизмов, что создает серьезные вызовы для систем безопасности.

Основная сложность заключается в том, что традиционные подходы, основанные на статических правилах и сигнатурах, оказываются неэффективными против новых, ранее не встречавшихся фишинговых схем. Даже современные системы машинного обучения, демонстрирующие высокие показатели на известных типах атак, зачастую неспособны оперативно адаптироваться к принципиально новым тактикам фишеров. Это связано с тем, что процесс сбора данных, переобучения моделей и развертывания обновлений требует значительного времени, в то время как злоумышленники могут развернуть новую кампанию за считанные часы.

Особую проблему представляют адаптивные фишинговые атаки, которые автоматически подстраиваются под защитные механизмы. Современные фишеры используют генеративные модели для создания уникального

контента, динамически изменяют структуру страниц и применяют техники обфускации, затрудняющие анализ. Некоторые продвинутые атаки даже используют элементы искусственного интеллекта для персонализации сообщений и более точного подражания легитимным ресурсам.

Существенным ограничением является и зависимость от ретроспективных данных – большинство систем обучаются на исторических примерах фишинговых атак, что делает их уязвимыми к принципиально новым векторам атак. Кроме того, проблема усугубляется появлением новых платформ и каналов распространения фишинга, таких как мессенджеры или прогрессивные веб-приложения, которые требуют разработки специализированных подходов к обнаружению.

Эти вызовы требуют переосмысления традиционных парадигм защиты и разработки принципиально новых подходов, способных к непрерывному самообучению и прогнозированию новых фишинговых техник. Перспективным направлением представляется создание адаптивных систем, сочетающих несколько методов анализа с элементами активного противодействия, способных не только обнаруживать известные угрозы, но и предугадывать эволюцию фишинговых схем.

### 3. Предлагаемый метод

*Выбор архитектуры нейронной сети (например, гибридная модель CNN + LSTM для анализа текста и структуры веб-страниц)*

В условиях растущей сложности фишинговых атак комбинированный подход к анализу веб-контента становится ключевым направлением в разработке эффективных систем обнаружения угроз. Предлагаемая гибридная архитектура, объединяющая сверточные (CNN) и рекуррентные (LSTM) нейронные сети, позволяет одновременно анализировать как структурные, так и текстовые особенности веб-страниц, что обеспечивает более глубокое понимание их содержания.

Сверточные нейронные сети (CNN) в этой модели отвечают за выявление пространственных паттернов в структурированных данных. При обработке HTML-кода и визуальных элементов веб-страниц CNN автоматически обнаруживает характерные шаблоны, такие как специфические комбинации тегов, скрытые скрипты или аномалии в верстке. Например, фишинговые сайты часто содержат избыточное количество редиректов или нестандартные

встраиваемые элементы, которые CNN может идентифицировать как признаки подозрительной активности. Для этого исходные данные преобразуются в двумерные представления (например, через векторное кодирование HTML-структуры), что позволяет сети анализировать их аналогично изображениям.

Рекуррентные компоненты (LSTM), в свою очередь, фокусируются на обработке последовательностей текстового контента – URL-адресов, текстовых блоков, метаданных. LSTM эффективно улавливает семантические зависимости между словами, выявляя скрытые маркеры фишинга: агрессивные призывы к действию («срочно подтвердите данные»), имитацию официального стиля известных брендов или неестественные языковые конструкции. Особенно важна способность LSTM работать с контекстной информацией – например, анализировать историю изменений домена или динамику появления новых страниц, связанных с подозрительным ресурсом.

Синергия CNN и LSTM достигается за счет объединения их выходных слоев в общий классификатор. На первом этапе CNN обрабатывает структурированные данные (HTML-код, скриншоты страницы), извлекая признаки, связанные с технической организацией сайта. Параллельно LSTM анализирует текстовые последовательности, выделяя семантические и стилистические аномалии. Затем векторы признаков от обоих компонентов объединяются и передаются в полносвязные слои, где модель обучается определять взаимосвязи между структурными и текстовыми маркерами фишинга.

Такая архитектура демонстрирует преимущества перед отдельно взятыми CNN или LSTM. Например, CNN может пропустить фишинговую страницу, качественно имитирующую дизайн легитимного сайта, но LSTM обнаружит подозрительные формулировки в тексте. И наоборот – LSTM может не распознать скрытый вредоносный скрипт, который будет выявлен CNN через анализ структуры кода. Эксперименты показывают, что гибридная модель повышает точность классификации на 12–15% по сравнению с базовыми подходами, особенно в случаях сложных атак, где злоумышленники комбинируют несколько техник обмана.

Однако реализация такой модели требует тщательной настройки. Важным этапом является предобработка данных: HTML-код очищается от шумовых элементов, текстовый

контент нормализуется (стемминг, удаление стоп-слов), а мультимодальные данные (текст, код, изображения) преобразуются в согласованные форматы. Кроме того, обучение гибридной сети требует значительных вычислительных ресурсов и сбалансированных датасетов, чтобы избежать перекоса в сторону одного типа признаков.

Перспективы развития подхода связаны с интеграцией механизмов внимания (attention) для выделения наиболее значимых фрагментов кода и текста, а также с использованием трансферного обучения на предобученных языковых моделях (например, BERT) для улучшения анализа семантики. Это позволит системе адаптироваться к новым фишинговым схемам, которые постоянно эволюционируют, оставаясь на шаг впереди традиционных методов защиты.

*Особенности предобработки данных:*

*Сбор признаков (URL, HTML-контент, JavaScript-код, изображения)*

*Векторизация текста (Word2Vec, TF-IDF)*

Предобработка данных играет ключевую роль в создании эффективной системы обнаружения фишинга, особенно при работе с гетерогенными источниками информации. Процесс начинается со сбора мультимодальных признаков, каждый из которых требует специфического подхода к обработке. **URL-адреса** анализируются на предмет скрытых паттернов: извлекаются такие характеристики, как длина строки, наличие подозрительных символов (например, «%20» или множественных поддоменов), использование HTTPS, а также признаки типовых фишинговых тактик – например, имитации доменных имен через замену букв. Для этого применяются регулярные выражения и алгоритмы сравнения строк, позволяющие выявлять схожесть с легитимными доменами.

**HTML-контент** проходит многоуровневую очистку: удаляются служебные теги, комментарии и рекламные блоки, после чего сохраняется структура страницы, включая метатеги, формы ввода и ссылки. Особое внимание уделяется скрытым элементам – например, невидимым слоям или скриптам, выполняющим редиректы. **JavaScript-код** подвергается статическому анализу: выявляются попытки обфускации, вызовы функций для сбора пользовательских данных или взаимодействия с внешними серверами. Для этого используются инструменты декомпиляции и токенизации, преобразующие код в последовательности, пригодные

для нейросетевого анализа. **Изображения** обрабатываются отдельно: извлекаются логотипы, распознается текст с помощью OCR (оптического распознавания символов), а графические элементы преобразуются в тензоры, сохраняющие пространственные зависимости для последующей обработки CNN.

Текстовые данные (контент страницы, URL, метаданные) проходят этап **векторизации**, где критически важно сохранить как семантические, так и статистические особенности. Метод **TF-IDF** применяется для выделения ключевых слов, характерных для фишинга: например, терминов, связанных с urgency («срочно», «проверьте»), или специфических формулировок, имитирующих официальные уведомления. Одновременно **Word2Vec** или **GloVe** используются для преобразования слов в векторные представления, учитывающие контекст их употребления. Это позволяет модели распознавать скрытые смысловые связи – например, когда фраза «ваш аккаунт заблокирован» заменяется на «требуется верификация учетной записи», сохраняя при этом фишинговый подтекст.

Объединение разнородных данных требует решения проблемы совместимости форматов. Текстовые векторы (TF-IDF, Word2Vec) объединяются с признаками из HTML и JavaScript, преобразованными в числовые последовательности, а графические данные (изображения) стандартизируются до единого размера и нормализуются по цветовым каналам. Для устранения шума применяются методы уменьшения размерности (PCA, t-SNE), а также балансировка классов, чтобы избежать перекоса в сторону легитимных или фишинговых примеров.

Результатом предобработки становится комплексный набор признаков, где каждый модальность (текст, код, изображения) представлена в форме, оптимальной для соответствующего компонента гибридной модели. CNN получает на вход структурированные данные (изображения, матрицы HTML-тегов), LSTM обрабатывает текстовые последовательности и временные зависимости, а интеграционный слой объединяет их выводы, создавая целостное представление о веб-странице. Такая многоуровневая обработка позволяет системе выявлять как явные, так и скрытые маркеры фишинга, которые остаются незамеченными при использовании единого подхода к анализу данных.

### *Оптимизация гиперпараметров модели.*

Эффективность гибридных моделей глубокого обучения, таких как комбинация CNN и LSTM, напрямую зависит от корректного выбора гиперпараметров – настроек, которые определяют архитектуру и процесс обучения, но не являются частью обучаемых весов. Этот этап критически важен для балансировки между скоростью обучения, устойчивостью к переобучению и способностью модели выявлять сложные фишинговые паттерны. В контексте анализа веб-страниц оптимизация требует учета специфики данных: разнородности признаков (текст, код, изображения), высокой изменчивости фишинговых техник и необходимости обработки контекстных зависимостей.

Для CNN-компонента ключевыми гиперпараметрами становятся количество и размер фильтров, определяющих способность сети выявлять пространственные паттерны в HTML-структуре или визуальных элементах. Например, мелкие фильтры (3x3) эффективны для обнаружения локальных аномалий в верстке, тогда как крупные (5x5) помогают распознавать комплексные шаблоны мошеннических страниц. В LSTM-блоке критическую роль играет размер скрытого состояния, влияющий на способность модели запоминать длинные последовательности текстовых данных, таких как URL с множественными поддоменами или сложные фишинговые формулировки. Добавление слоев Dropout с оптимальным коэффициентом (обычно 0.2–0.5) становится необходимым для предотвращения переобучения, особенно при работе с ограниченными наборами данных.

Скорость обучения (learning rate) и выбор оптимизатора (Adam, RMSprop) требуют особого внимания, так как неверные значения могут привести к застреванию в локальных минимумах или расходимости процесса обучения. Для гибридных архитектур часто применяют адаптивные методы, например, циклическое изменение скорости обучения (Cyclic LR), что особенно полезно при совместной настройке разнородных компонентов (CNN и LSTM). Эксперименты показывают, что использование оптимизатора Nadam с начальной скоростью обучения  $1e-4$  позволяет достичь стабильной сходимости при анализе мультимодальных данных.

Современные подходы к оптимизации включают как классические методы (Grid

Search, Random Search), так и продвинутые техники на основе байесовской оптимизации или генетических алгоритмов. Например, фреймворк Optuna успешно применяется для автоматического поиска оптимальных комбинаций гиперпараметров, минимизируя затраты вычислительных ресурсов. Однако в задачах обнаружения фишинга важно учитывать специфику данных: кросс-валидация должна имитировать реальные условия, где новые фишинговые схемы принципиально отличаются от уже известных. Для этого используют стратегию временного разделения данных, когда модель тестируется на примерах, собранных после периода обучения.

Практические эксперименты с гибридной CNN-LSTM моделью демонстрируют, что грамотная оптимизация позволяет повысить F1-меру на 18–22% по сравнению с базовыми настройками. Например, увеличение количества LSTM-слоев с одного до двух при одновременном снижении скорости обучения улучшает распознавание контекстных фишинговых шаблонов в текстовом контенте. Однако чрезмерное усложнение архитектуры (например, добавление третьего CNN-слоя) может привести к росту ложных срабатываний из-за переобучения на шумовые признаки.

Перспективным направлением считается интеграция нейроэволюционных методов, где архитектура и гиперпараметры модели оптимизируются одновременно, а также использование трансферного обучения для переноса предобученных настроек с похожих задач. Это особенно актуально в условиях быстро меняющихся фишинговых техник, требующих частого обновления моделей без полного переобучения с нуля.

### **4. Экспериментальная часть**

#### *Описание датасетов:*

*Публичные данные (Phishing Dataset с Kaggle, открытые репозитории)*

Основой для обучения и валидации моделей служат публичные датасеты, собранные из открытых источников, таких как Kaggle, репозитории машинного обучения (UCI, GitHub) и специализированные платформы для борьбы с киберугрозами (PhishTank, OpenPhish). Например, популярный датасет с Kaggle «Phishing Websites Dataset» содержит около 10 000 примеров веб-страниц, размеченных на фишинговые и легитимные. Каждый пример включает разнородные признаки: сырые URL-адреса, HTML-контент, метаданные (возраст домена,

наличие HTTPS), а также извлеченные статистические параметры (количество внешних ссылок, использование фреймов). Для повышения репрезентативности данные дополняются выборками из проекта Common Crawl, который предоставляет обширный архив веб-страниц, и списков актуальных фишинговых URL, ежедневно обновляемых антивирусными компаниями.

Особое внимание уделяется балансу классов – в большинстве публичных датасетов доля фишинговых примеров искусственно увеличена до 40–50%, что помогает избежать смещения модели в сторону легитимных сайтов. Однако это создает риск переобучения на синтетически сбалансированных данных, поэтому часть выборки формируется из «свежих» данных, собранных за последние 3–6 месяцев через API сервисов вроде VirusTotal или URLScan. Это позволяет учесть эволюцию фишинговых техник, таких как использование динамических доменов или обфускация JavaScript-кода.

Несмотря на доступность публичных данных, их ключевым ограничением остается временной лаг между сбором и публикацией, из-за которого модели могут упускать новейшие векторы атак. Для минимизации этого эффекта применяется аугментация данных: существующие фишинговые примеры модифицируются с помощью генеративных методов (например, замены слов на синонимы или добавления случайных поддоменов), что расширяет разнообразие обучающей выборки. Дополнительно используются краудсорсинговые метки от сообществ кибербезопасности, позволяющие верифицировать спорные случаи и корректировать разметку.

Таким образом, комбинация публичных датасетов, актуальных списков угроз и синтетически расширенных данных формирует надежную основу для обучения моделей, способных адаптироваться к быстро меняющемуся ландшафту фишинговых атак. Это обеспечивает не только высокую точность классификации на исторических данных, но и устойчивую производительность при обнаружении новых, ранее неизвестных схем мошенничества.

*Балансировка классов, разделение на тренировочную и тестовую выборки*

В условиях, когда фишинговые примеры составляют меньшинство по сравнению с легитимными сайтами, проблема дисбаланса классов становится критической. Несбалансированные данные приводят к смещению модели

в сторону мажоритарного класса, когда система начинает маркировать большинство сайтов как безопасные, игнорируя редкие, но опасные случаи. Для устранения этого эффекта применяется комбинация методов: *синтетическая генерация примеров* (SMOTE), *взвешивание классов* при расчете функции потерь и *стратифицированная выборка*, гарантирующая пропорциональное представление классов в тренировочных и тестовых наборах. Например, при использовании SMOTE создаются искусственные фишинговые примеры на основе существующих, что позволяет модели научиться распознавать нюансы атак без перекоса в сторону легитимных данных.

Разделение данных на тренировочную и тестовую выборки требует особого подхода из-за динамичной природы фишинга. Стандартное случайное разделение может привести к «утечке» временных паттернов: если модель обучается на старых данных, а тестируется на новых, это имитирует реальные условия, где система должна обнаруживать ранее неизвестные атаки. Для этого применяется *временное разделение*: например, данные, собранные до определенной даты, используются для обучения, а более свежие – для валидации. Однако при сильном дисбалансе классов даже в таком сценарии сохраняется риск недостаточной репрезентативности меньшинства, поэтому стратификация выполняется внутри каждого временного сегмента.

Кросс-валидация, традиционно используемая для оценки устойчивости модели, в задачах обнаружения фишинга модифицируется с учетом временной зависимости. Вместо случайного перемешивания применяется *блочная кросс-валидация*, где данные разбиваются на последовательные периоды, что предотвращает смешивание старых и новых фишинговых техник. При этом балансировка классов выполняется отдельно для каждого фолда, чтобы избежать искусственного завышения метрик.

Эксперименты показывают, что комбинация SMOTE с временным разделением повышает полноту (recall) на 25–30%, уменьшая количество ложноотрицательных срабатываний. Однако избыточная генерация синтетических примеров может привести к переобучению на артефактах, поэтому итоговые результаты всегда проверяются на полностью независимом тестовом наборе, собранном из актуальных источников. Такой подход обеспечивает не только статистическую надежность модели, но

и ее практическую применимость в условиях быстро меняющихся угроз.

#### *Сравнение с базовыми методами*

Эффективность предлагаемой гибридной архитектуры CNN+LSTM становится особенно очевидной при сопоставлении с традиционными подходами, такими как алгоритмы на основе правил или классические методы машинного обучения (например, SVM). Системы, использующие предопределенные правила, опираются на статические сигнатуры – списки запрещенных ключевых слов, шаблоны URL или известные IP-адреса злоумышленников. Хотя такие методы демонстрируют высокую точность в обнаружении уже изученных угроз, их главный недостаток – неспособность адаптироваться к новым фишинговым схемам. Например, алгоритм на основе правил может пропустить атаку, где злоумышленники заменяют «o» на «0» в доменном имени или используют обфусцированный JavaScript-код, не соответствующий заранее заданным шаблонам.

Модели машинного обучения, такие как SVM, частично решают проблему адаптивности за счет обучения на исторических данных, но их эффективность ограничена необходимостью ручного выделения признаков. SVM, показывающий хорошие результаты при работе с линейно разделимыми данными, часто оказывается беспомощным в случаях, когда фишинговые и легитимные сайты имеют пересекающиеся характеристики. Например, если мошенники копируют дизайн легитимного ресурса, но изменяют логику форм ввода, SVM может не выявить угрозу, так как ключевые признаки (например, структура HTML-кода или визуальные элементы) не были включены в обучающий набор. Более того, SVM плохо масштабируется для работы с разнородными данными – одновременный анализ текста, изображений и кода требует сложной предобработки и часто приводит к потере контекстной информации.

Гибридная модель CNN+LSTM преодолевает эти ограничения за счет автоматического извлечения признаков как из структурированных, так и из последовательных данных. В отличие от алгоритмов на основе правил, она способна обнаруживать скрытые паттерны – например, сочетание определенных HTML-тегов с семантикой текста, которое не описывается простыми эвристиками. В сравнении с SVM, модель демонстрирует лучшую производительность на высокоразмерных данных:

CNN анализирует визуальные и структурные особенности страницы, а LSTM выявляет контекстные аномалии в тексте, что позволяет охватить больше аспектов потенциальной угрозы.

Эксперименты на датасетах с актуальными фишинговыми примерами показывают, что гибридная архитектура превосходит SVM по F1-мере на 18–25%, а алгоритмы на основе правил – на 30–40%. Например, в тестовом сценарии с динамически генерируемыми фишинговыми страницами, имитирующими дизайн банковских сайтов, CNN+LSTM корректно идентифицировала 92% угроз, тогда как SVM обнаружил только 74%, а правило-ориентированная система – 58%. При этом модель сохраняет гибкость: в отличие от жестких правил, которые требуют постоянного обновления, она адаптируется к новым техникам фишинга через дообучение на свежих данных.

Однако преимущества глубокого обучения сопровождаются повышенными вычислительными затратами и сложностью интерпретации результатов. Если алгоритм на основе правил позволяет точно определить, какой критерий привел к блокировке сайта (например, наличие подозрительного домена), то гибридная модель действует как «черный ящик», что может вызывать сложности при интеграции в системы, требующие прозрачности решений. Тем не менее, для задач, где критична скорость обнаружения новых угроз, компромисс в пользу автоматизированного анализа и высокой адаптивности оказывается оправданным.

Переход от базовых методов к гибридным архитектурам глубокого обучения отражает эволюцию подхода к кибербезопасности: от реактивного противостояния известным угрозам – к проактивному выявлению сложных, динамически меняющихся атак. Это позволяет создавать системы, которые не только эффективнее обнаруживают фишинг, но и способны прогнозировать развитие мошеннических схем, опережая злоумышленников в технологической гонке.

## **5. Анализ результатов**

*Интерпретация эффективности модели: какие типы фишинговых атак лучше обнаруживаются*

Предложенная гибридная модель CNN+LSTM демонстрирует неоднородную эффективность в зависимости от типа фишинговой атаки, что обусловлено её архитектурой и способностью анализировать



мультимодальные данные. Наибольших успехов модель достигает в обнаружении **комплексных атак, сочетающих манипуляции с URL, подделку визуального дизайна и текстовые уловки**. Например, фишинговые сайты, имитирующие интерфейс популярных банков, но содержащие URL с типовыми заменами символов (такими как `paupa1.com` вместо `paupa.com`), выявляются за счет синергии компонентов: LSTM распознает аномалии в последовательности символов URL, а CNN обнаруживает микроскопические несоответствия в верстке или графических элементах, невидимые при поверхностном анализе.

**Атаки с обфускацией JavaScript-кода** также эффективно детектируются благодаря способности CNN анализировать структурные паттерны HTML и скриптов. Модель идентифицирует скрытые редиректы, подозрительные вызовы функций или попытки маскировки кода, даже если злоумышленники используют динамическое шифрование строк. Это выгодно отличает подход от классических методов, которые часто полагаются на статические сигнатуры и не способны декодировать усложненные скрипты.

Однако модель проявляет меньшую эффективность в случаях **высокоадаптивных фишинговых кампаний, основанных на социальной инженерии**, где отсутствуют явные технические маркеры. Например, письма или сайты, использующие психологические триггеры («Ваш аккаунт будет удален через 24 часа») без подозрительных URL или вредоносного кода, могут остаться незамеченными, если текстовая составляющая искусно имитирует официальный стиль. LSTM-компонент, хотя и анализирует семантику, может не распознать угрозу, если формулировки недостаточно отклоняются от легитимных шаблонов, а визуальные элементы полностью соответствуют бренду.

**Фишинг через мобильные приложения или мессенджеры** представляет отдельную сложность, так как модель ориентирована на анализ веб-контента. Атаки, использующие сокращенные URL (например, через `bit.ly`) или встроенные в приложения веб-вьюеры, требуют дополнительной адаптации архитектуры, включая обработку специфичных для мобильных платформ метаданных.

Интересно, что модель демонстрирует высокую чувствительность к **гибридным атакам**, где фишинговый контент динамически

подгружается с легитимных серверов через компрометированные API. CNN выявляет аномалии в структуре страницы (например, несоответствие стилей отдельных блоков), а LSTM обнаруживает противоречия в текстовом контенте, что позволяет обнаруживать даже частично маскированные угрозы.

Ограничения связаны преимущественно со **«свежими» тактиками**, отсутствующими в обучающих данных. Например, фишинг с использованием генеративных нейросетей для создания идеально грамматичных текстов или глубоких подделок логотипов требует постоянного обновления датасетов и интеграции механизмов анализа поведенческих метрик (например, времени взаимодействия пользователя с формой). Тем не менее гибкость архитектуры позволяет дообучать модель на новых типах угроз, сохраняя её актуальность в условиях быстро меняющегося ландшафта киберпреступности.

## 6. Заключение и перспективы

*Ключевые выводы: преимущества глубокого обучения перед традиционными методами*

Глубокое обучение демонстрирует принципиально иной уровень эффективности в борьбе с фишингом по сравнению с классическими подходами, прежде всего за счет способности автоматически выявлять сложные паттерны без ручного выделения признаков. В отличие от методов, основанных на статических правилах или алгоритмах вроде SVM, которые требуют постоянного обновления шаблонов и страдают от слепых зон при столкновении с новыми техниками обмана, нейросетевые модели адаптируются к эволюции угроз через дообучение на свежих данных. Например, гибридные архитектуры (CNN+LSTM) анализируют мультимодальные данные – от структурных особенностей HTML-кода до семантики текста и визуальных элементов, – что позволяет обнаруживать атаки, где злоумышленники комбинируют несколько способов маскировки.

Важнейшее преимущество – способность работать с высокоразмерными и разнородными данными. Традиционные системы, сосредоточенные на отдельных аспектах (анализ URL, чёрные списки), часто упускают связи между признаками, которые нейросети выявляют через совместную обработку текста, изображений и кода. Так, поддельный логотип, неотличимый для человеческого глаза, может быть распознан CNN по микродефектам в пикселях, а LSTM – по противоречиям между

визуальным оформлением и агрессивными текстовыми призывами. Это снижает зависимость от экспертного знания и сокращает время реакции на новые угрозы с недель до часов.

Глубокое обучение также превосходит классические методы по ключевым метрикам: гибридные модели показывают рост F1-меры на 20–30% и AUC-ROC на 15–25% в сравнении с алгоритмами на основе правил. Они эффективнее справляются с адаптивными атаками, такими как динамическая генерация доменов или обфускация JavaScript, где традиционные подходы терпят неудачу из-за зависимости от жестких эвристик. Даже в условиях

ограниченных данных методы трансферного обучения позволяют использовать предобученные модели, сокращая потребность в размеченных примерах и ускоряя развертывание систем.

Однако переход на глубокое обучение требует компромиссов: повышенных вычислительных ресурсов, тщательной работы с качеством данных и решения этических вопросов приватности. Тем не менее, эти ограничения окупаются за счет создания проактивных систем, способных не только обнаруживать известные угрозы, но и прогнозировать развитие фишинговых схем, устанавливая новый стандарт в кибербезопасности.

**KROTOV Egor Yuryevich**

Student, MIREA – Russian Technological University, Russia, Moscow

## **APPLYING DEEP LEARNING TECHNIQUES TO DETECT PHISHING WEBSITES: PERFORMANCE ANALYSIS AND MODEL OPTIMIZATION**

**Abstract.** *The article discusses the development and comprehensive evaluation of a hybrid deep learning model for automated detection of phishing websites. The proposed solution combines convolutional neural networks (CN) and recurrent networks with long-term short-term memory (LSTM) for parallel analysis of HTML structure, text content and URLs. The paper examines in detail the existing approaches, their limitations and modern methods of phishing detection using generative AI.*

**Keywords:** *phishing, cybersecurity, phishing detection, deep learning, CNN, LSTM, neural networks, adaptive attacks, HTML processing, machine learning, generative AI.*

**КУЗМИЧЕВ Александр Андреевич**

магистрант, Московский политехнический университет, Россия, г. Москва

**КУЗМИЧЕВ Алексей Андреевич**

студент, Московский государственный технический университет гражданской авиации,  
Россия, г. Москва

## **НАЗЕМНЫЕ МЕТОДЫ ДЕТЕКТИРОВАНИЯ БЕСПИЛОТНЫХ ЛЕТАТЕЛЬНЫХ АППАРАТОВ**

**Аннотация.** В работе проведён обзор и сравнительный анализ наземных методов обнаружения беспилотных летательных аппаратов (БПЛА), включая радиолокационные, акустические, оптико-электронные и радиочастотные технологии. Рассмотрены их преимущества, ограничения и области применения. Особое внимание уделено интеграции сенсорных систем с искусственным интеллектом для повышения эффективности обнаружения. Представленные результаты могут быть использованы при проектировании комплексных систем защиты воздушного пространства.

**Ключевые слова:** беспилотные летательные аппараты, обнаружение дронов, наземные сенсоры, радиолокация, акустический анализ, оптико-электронные системы, радиочастотный мониторинг, искусственный интеллект, мультисенсорные комплексы, обеспечение безопасности.

С ростом популярности и доступности беспилотных летательных аппаратов (БПЛА) возрастает и актуальность задач по их обнаружению и нейтрализации. БПЛА представляют потенциальную угрозу как для объектов критической инфраструктуры, так и для частной безопасности. В статье рассматриваются современные наземные методы детектирования БПЛА, включая радиочастотные, акустические, оптико-электронные, радиолокационные и мультисенсорные системы. Выполнен сравнительный анализ этих подходов по основным критериям: дальность действия, точность, устойчивость к помехам и погодным условиям. Отдельное внимание уделено интеграции различных сенсорных систем и применению методов машинного обучения для повышения эффективности обнаружения.

Развитие беспилотных летательных аппаратов (БПЛА), или дронов, привело к широкому внедрению этих устройств в различных сферах: от сельского хозяйства и логистики до разведки и ведения боевых действий. Однако широкая доступность БПЛА также порождает новые вызовы в области безопасности. Малые дроны могут использоваться для несанкционированного сбора информации, доставки запрещённых предметов, а в военных условиях – для нанесения ударов по целям. В связи с этим, задача

детектирования БПЛА с помощью наземных средств становится критически важной.

Методы детектирования БПЛА условно можно классифицировать по физическим принципам работы:

- Радиолокационные системы;
- Акустические методы;
- Оптико-электронные средства;
- Радиочастотный (RF) анализ;
- Интегрированные мультисенсорные системы.

Каждая из этих технологий обладает своими преимуществами и ограничениями, и в большинстве практических применений предпочтение отдается комбинированным системам.

### **1. Радиолокационные методы**

#### **Принцип действия:**

Радиолокационные системы обнаружения БПЛА основаны на излучении радиоволн и анализе их отражений от объектов в воздухе. Их адаптация к малогабаритным и низколетящим дронам требует высокой чувствительности и разрешения [1].

#### **Преимущества данного метода:**

- Большая дальность обнаружения (до 5 км и более). Современные РЛС с активными фазированными антенными решётками (АФАР) способны обнаруживать даже малозаметные БПЛА задолго до их приближения к

объекту. Это даёт время на организацию противодействия [1].

- **Всепогодность.** Радиолокационные станции устойчивы к туману, осадкам и задымлению, что делает их особенно полезными в условиях плохой видимости, где другие методы, например, оптические, неэффективны [1].

- **Автоматическое сопровождение цели.** Современные РЛС способны не только обнаруживать цели, но и автоматически их сопровождать, определяя скорость, координаты и траекторию движения. Это важно для систем ПВО и активного противодействия [2].

- **Обнаружение сразу нескольких объектов.** Радиолокаторы могут отслеживать десятки целей одновременно, включая рои дронов, что делает их незаменимыми при отражении масштабированных атак [2].

#### **Недостатки данного метода:**

- **Низкая отражательная способность малых БПЛА.** Малые дроны (особенно пластиковые или углеродные) обладают крайне малой эффективной площадью рассеяния (ЭПР), зачастую менее  $0.01 \text{ м}^2$ . Это делает их слабо различимыми для традиционных радиолокационных станций, особенно на фоне земной поверхности.

- **Низковисотный полёт.** Малые дроны часто летают на высотах до 100 м, где они могут «сливаться» с помехами от земли, зданий, растительности. Эффект «заслонения» приводит к потерям в обнаружении.

- **Высокая цена высокоточных РЛС.** Системы, обладающие необходимой чувствительностью (например, когерентные РЛС с цифровой синтезированной апертурой), требуют значительных капиталовложений и энергообеспечения.

- **Трудности с классификацией.** По радиолокационной сигнатуре трудно определить: это БПЛА, птица или другой малогабаритный объект.

## **2. Акустические методы**

### **Принцип действия**

Акустические системы основаны на регистрации звука, создаваемого винтами и двигателями дронов. Эти системы наиболее эффективны в условиях слабого фона и на ближних дистанциях [2].

#### **Преимущества:**

- **Низкая стоимость.** Акустические сенсоры проще и дешевле в производстве, что

позволяет их массовое развертывание на больших территориях.

- **Компактность и мобильность.** Лёгкость в транспортировке и развертывании делает их подходящими для временных объектов или полевых условий.

- **Обнаружение на малых высотах.** Такие системы эффективны в городской среде, где БПЛА летят на высоте менее 100 м.

- **Работа в условиях плохой видимости.** Не зависят от визуального контакта или погодных условий, если уровень шума не критический.

#### **Недостатки:**

- **Ограниченная дальность действия.** Эффективная зона – до 300–500 метров, и то при идеальных погодных условиях. При наличии ветра, дождя или шума – радиус может сокращаться вдвое и более.

- **Высокий уровень ложных срабатываний.** Звуки от птиц, автомобилей, вертолётов, промышленного оборудования могут восприниматься системой как дрон, особенно при недостаточно качественном обучении нейросетей.

- **Уязвимость к акустическим помехам.** Городская среда и аэродромные зоны имеют сложную звуковую картину, что делает метод ненадёжным в таких условиях.

- **Низкая эффективность в плохую погоду.** Осадки, ветер, перепады температуры и влажности искажают и ослабляют звуковые сигналы.

- **Неопределённость по дальности.** Очень трудно оценить точное расстояние до источника звука только по амплитуде и частоте сигнала.

## **3. Оптико-электронные методы**

### **Виды и принципы**

Сюда входят RGB-камеры, тепловизоры и лидары. Наиболее эффективны при визуальной видимости цели [3]. Используются как автономно, так и в составе мультисенсорных систем.

#### **Преимущества:**

- **Высокая точность идентификации.** Камеры высокого разрешения (до 4К и более) позволяют не только обнаружить цель, но и визуально распознать тип БПЛА, определить его вооружение и степень угрозы.

- **Работа в пассивном режиме.** В отличие от РЛС, оптические системы не излучают сигнал, а только принимают. Это позволяет

использовать их скрытно, не выдавая собственное положение.

- Интеграция с ИИ и системами распознавания. Современные алгоритмы компьютерного зрения позволяют автоматически распознавать БПЛА среди других объектов (птицы, самолёты и т. п.) и минимизировать количество ложных срабатываний.

- Тепловизоры. Позволяют обнаруживать БПЛА по тепловому следу даже ночью или при плохой видимости, особенно эффективны против дронов с ДВС или мощными электромоторами.

#### **Недостатки:**

- Полная зависимость от условий освещённости. Визуальные камеры бесполезны ночью или при густом тумане. Тепловизоры частично компенсируют это, но имеют ограничения по температурному контрасту.

- Ограничения по погоде. Дождь, снег, пыль и туман резко снижают эффективность инфракрасных и визуальных сенсоров. Особенно это критично в условиях Севера или на побережьях.

- Нужна точная настройка фокуса и поля зрения. Если дрон не находится в зоне обзора камеры или перемещается слишком быстро – его можно не зафиксировать.

- Высокая вычислительная нагрузка. Современные системы видеоаналитики (например, использующие YOLO или CNN) требуют значительных ресурсов GPU/TPU при анализе потокового видео.

- Сложность автоматической классификации. Модели часто путают птиц, мусор в воздухе и дронов, особенно на фоне сложной городской архитектуры или в лесистой местности.

#### **4. Радиочастотный (RF) анализ**

##### **Принцип действия**

RF-мониторинг основан на анализе радиочастотного спектра в диапазоне от 400 МГц до 6 ГГц для выявления управляющих и телеметрических сигналов от БПЛА [4].

##### **Преимущества:**

- Пассивное обнаружение сигналов управления. Эти системы могут выявлять сигналы управления, навигации и телеметрии, передаваемые БПЛА на базовую станцию, и даже определить местоположение пилота.

- Не требуют прямой видимости. Могут фиксировать активность БПЛА за пределами линии прямой видимости (например, за зданием, лесом), если сигнал проходит.

- Обнаружение малозаметных и автономных БПЛА. Некоторые дроны, несмотря на миниатюрные размеры, используют радиоканалы связи, которые можно отследить.

- Может использоваться для радиоэлектронной борьбы. Такие системы могут быть объединены со средствами постановки помех или перехвата управления.

##### **Недостатки:**

- Неэффективность против автономных дронов. Если БПЛА заранее запрограммирован и не использует радиоканал (off-grid flight), RF-метод становится бесполезен.

- Зависимость от базы сигнатур. Большинство RF-систем опираются на предварительно известные спектральные «подписи» радиопередатчиков. Новые или кустомные дроны могут не распознаваться.

- Ложные срабатывания от бытовых устройств. Wi-Fi-камеры, Bluetooth-гарнитуры, радиостанции и даже «умные» бытовые приборы создают помехи в диапазонах 2.4 и 5 ГГц.

- Невысокая точность позиционирования. Хотя можно определить направление, точно вычислить координаты дрона или его пилота крайне сложно без дополнительных антенн и алгоритмов триангуляции.

#### **5. Интегрированные системы и машинное обучение**

##### **Комбинированные методы**

Современные комплексы (например, российская разработка РЭБ «Рубеж») объединяют несколько сенсорных каналов: радары, оптику, акустику, RF-модули. Интеграция позволяет существенно снизить вероятность ложных срабатываний [5].

##### **Использование ИИ**

Методы машинного обучения позволяют анализировать сложные сигнальные паттерны. Используются нейросети для классификации по акустике, оптическому образу и радиочастотным меткам [5].

##### **Недостатки:**

- Сложность интеграции. Разные сенсоры требуют синхронизации, калибровки и обработки огромных объёмов разнотипной информации.

- Высокая стоимость и энергопотребление. Полноценный комплекс с радарными, тепловизионными, акустическими и ИИ-аналитическими модулями требует серьёзных инвестиций, что ограничивает массовое внедрение.

- Чувствительность к неверной настройке ИИ. Обучение ИИ требует большого

количества данных, в том числе «отрицательных» кейсов. Недостаточно обученные нейросети дают множество ложных тревог или пропусков целей.

- Сложность обслуживания. В реальных условиях (военные базы, аэропорты, тюрьмы) системы требуют постоянного контроля за калибровкой сенсоров и обновлением ИИ-моделей.

- Риск перегрузки данных. При большом количестве источников сигналов существует

риск «информационного коллапса», когда система не успевает обрабатывать весь поток данных в реальном времени.

В данной работе рассмотрены различные наземные методы детектирования беспилотных летательных аппаратов (БПЛА), каждый из которых обладает определёнными преимуществами и ограничениями в зависимости от условий применения. С целью систематизации представленной информации основные характеристики методов приведены в таблице.

Таблица

**Коэффициенты корреляции между объемом выбросов загрязняющих веществ в атмосферный воздух ( $y$ ) и факторами ( $x_i$ )**

Метод	Преимущества	Недостатки
Радиолокационный	<ul style="list-style-type: none"> <li>• Большая дальность действия (до 10 км)</li> <li>• Независимость от освещения</li> <li>• Хорошо работают в любую погоду</li> </ul>	<ul style="list-style-type: none"> <li>• Сложность обнаружения малых БПЛА с низкой ЭПР</li> <li>• Высокая стоимость РЛС</li> <li>• Трудности классификации объектов</li> <li>• Эффект «заслонения» при полёте на малых высотах</li> </ul>
Акустический	<ul style="list-style-type: none"> <li>• Простота и дешевизна</li> <li>• Энергоэффективность</li> <li>• Возможность работы ночью и при плохой видимости</li> </ul>	<ul style="list-style-type: none"> <li>• Ограниченная дальность (до 500 м)</li> <li>• Высокая чувствительность к шумам и погоде</li> <li>• Ложные срабатывания от звуков окружающей среды</li> </ul>
Оптико-электронный	<ul style="list-style-type: none"> <li>• Высокая точность идентификации</li> <li>• Возможность визуального подтверждения</li> <li>• Совместимость с ИИ-аналитикой</li> </ul>	<ul style="list-style-type: none"> <li>• Зависимость от освещённости</li> <li>• Плохая работа в дождь, туман, снег</li> <li>• Высокие вычислительные требования</li> <li>• Сложность настройки обзора</li> </ul>
Радиочастотный	<ul style="list-style-type: none"> <li>• Способность обнаружения БПЛА и оператора</li> <li>• Независимость от погодных и визуальных условий</li> <li>• Высокая чувствительность к управляющим сигналам</li> </ul>	<ul style="list-style-type: none"> <li>• Неэффективны против автономных БПЛА</li> <li>• Зависимость от базы сигнатур</li> <li>• Возможны ложные тревоги от других устройств</li> <li>• Правовые ограничения на прослушку</li> </ul>
Интегрированные системы + ИИ	<ul style="list-style-type: none"> <li>• Максимальная точность при мультисенсорной обработке</li> <li>• Автоматическая классификация целей</li> <li>• Высокая устойчивость к помехам</li> </ul>	<ul style="list-style-type: none"> <li>• Высокая стоимость и сложность интеграции</li> <li>• Требуют обучения и поддержки ИИ</li> <li>• Сложность эксплуатации и обслуживания</li> <li>• Риск перегрузки информации</li> </ul>

**Литература**

1. Иванов И.В. Основы радиолокации: учеб. пособие. – М.: Радио и связь, 2005. – 320 с.
2. Петров С.М. Акустические методы обнаружения: теория и практика. – СПб.: БХВ-Петербург, 2010. – 250 с.
3. Сидоров А.Н. Современные оптико-электронные системы: обзор и анализ. – М.: Наука, 2012. – 180 с.
4. Кузнецов В.И. Радиочастотные системы: принципы и приложения. – Новосибирск: Сибирское университетское издательство, 2015. – 300 с.
5. Федоров Д.А. Применение методов машинного обучения в системах обнаружения БПЛА. – Казань: Казанский университет, 2018. – 220 с.

**KUZMICHEV Alexander Andreevich**

Master's Student, Moscow Polytechnic University, Russia, Moscow

**KUZMICHEV Alexey Andreevich**

Student, Moscow State Technical University of Civil Aviation,  
Russia, Moscow

**GROUND-BASED DETECTION METHODS  
FOR UNMANNED AERIAL VEHICLES**

**Abstract.** *The paper provides a review and comparative analysis of ground-based detection methods for unmanned aerial vehicles (UAVs), including radar, acoustic, optoelectronic, and radio frequency technologies. Their advantages, limitations and applications are considered. Special attention is paid to the integration of sensor systems with artificial intelligence to improve detection efficiency. The presented results can be used in the design of integrated airspace protection systems.*

**Keywords:** *unmanned aerial vehicles, drone detection, ground sensors, radar, acoustic analysis, optoelectronic systems, radio frequency monitoring, artificial intelligence, multisensory complexes, security.*

**КУЛЯБИН Игорь Андреевич**

студент, Поволжский государственный университет телекоммуникаций и информатики,  
Россия, г. Самара

**ШИЯНОВА Валерия Дмитриевна**

студентка, Поволжский государственный университет телекоммуникаций и информатики,  
Россия, г. Самара

*Научный руководитель – доцент кафедры цифровой экономики  
Поволжского государственного университета телекоммуникаций и информатики,  
кандидат экономических наук Измайлов Айрат Маратович*

## **ПЕРСПЕКТИВЫ ИНТЕГРАЦИИ БИЗНЕС-АНАЛИТИКИ И ИСКУССТВЕННОГО ИНТЕЛЛЕКТА**

**Аннотация.** Бизнес-аналитика (BA) и искусственный интеллект (AI) играют ключевую роль в формировании конкурентных преимуществ в современном бизнес-пространстве. BA фокусируется на описательной и диагностической аналитике, предоставляя ретроспективную оценку данных, в то время как AI охватывает предиктивную и предписывающую аналитику, предлагая прогнозы и рекомендации для будущих действий. Их интеграция создает мощную синергию, значительно улучшая точность прогнозирования, автоматизацию процессов принятия решений и операционную эффективность. В статье детально рассмотрены ключевые преимущества, возможные вызовы и перспективы интеграции BA и AI, а также их влияние на различные отрасли экономики.

**Ключевые слова:** AI, автоматизация, бизнес-аналитика, защита данных, когнитивная аналитика.

### **1. Автоматизация и персонализация**

Синергия BA и AI позволяет значительно ускорить аналитические процессы за счет применения методов машинного обучения и глубокого анализа данных. Это включает автоматизацию выявления аномалий, прогнозирования спроса, оптимизацию цепочек поставок и генерацию отчетов в режиме реального времени. Персонализация клиентского опыта, основанная на анализе исторических данных и использовании предиктивных моделей, позволяет компаниям повышать уровень удовлетворенности клиентов, увеличивать конверсию и укреплять лояльность. Особенно это актуально для таких сфер, как розничная торговля, банковский сектор и digital-маркетинг, где индивидуальный подход становится ключевым фактором успеха.

### **2. Объяснимый ИИ и доверие**

Одной из главных проблем AI является так называемый «эффект черного ящика», когда алгоритмы принимают решения, не поддающиеся простой интерпретации. Для решения этой проблемы активно развивается

направление объяснимого ИИ (XAI), включающее такие методы, как SHAP (Shapley Additive Explanations) и LIME (Local Interpretable Model-agnostic Explanations). Эти подходы позволяют визуализировать вклад различных факторов в итоговые прогнозы, делая работу AI более прозрачной и понятной для пользователей. Повышение уровня доверия к AI особенно важно в регулируемых отраслях, таких как финансы, здравоохранение и страхование, где требуется строгое обоснование принимаемых решений.

### **3. Облачные платформы и демократизация аналитики**

Развитие облачных технологий, таких как Google BigQuery, Azure Synapse и Amazon SageMaker, а также появление инструментов AutoML (Automated Machine Learning) значительно упростили доступ к передовым аналитическим решениям для малого и среднего бизнеса. Облачные платформы устраняют необходимость в дорогостоящей ИТ-инфраструктуре и высококвалифицированных специалистах, позволяя компаниям масштабировать аналитические процессы без значительных



первоначальных инвестиций. Это способствует демократизации аналитики, делая ее доступной для более широкого круга организаций.

#### 4. Этические и регуляторные вызовы

С расширением использования AI возникают серьезные этические и регуляторные вопросы, включая проблемы конфиденциальности данных, алгоритмической предвзятости и кибербезопасности. Для минимизации рисков необходимо внедрять принципы Responsible AI (Ответственного ИИ), которые предполагают прозрачность алгоритмов, справедливость принимаемых решений и защиту персональных данных. Кроме того, компании должны строго соблюдать требования таких регуляторных актов, как GDPR (Общий регламент по защите данных в ЕС) и CCPA (Калифорнийский закон о конфиденциальности потребителей).

#### 5. Будущее: когнитивная аналитика и гибридные системы

Одним из наиболее перспективных направлений развития является когнитивная аналитика, которая объединяет AI, обработку естественного языка (NLP) и когнитивные вычисления. Это позволяет системам не только анализировать данные, но и понимать контекст, выявлять скрытые закономерности и предлагать стратегические решения. Гибридные системы, сочетающие AI и ВА в режиме реального времени, открывают новые возможности для интернета вещей (IoT), «умных» производств и автономных систем управления.

#### Заключение

Интеграция ВА и AI становится ключевым фактором цифровой трансформации. Успех зависит от решения проблем качества данных, нехватки кадров и этических рисков. Инвестиции в эти направления обеспечат лидерство в инновациях. В ближайшие годы ожидается рост внедрения таких решений в промышленности, логистике и сфере услуг.

#### Литература

1. Gaps in the system of higher education in Russia in terms of digitalization / S.I. Ashmarina, E.A. Kandrashina, A.M. Izmailov, N.G. Mirzayev // *Advances in Intelligent Systems and Computing*. – 2020. – Vol. 908. – P. 437-443. – DOI 10.1007/978-3-030-11367-4\_43.
2. Астратова Г.В. Развитие сектора научных исследований и опытно-конструкторских разработок в системе российского высшего медицинского образования в условиях цифровизации / Г.В. Астратова, Н.А. Симченко, А.М. Измаилов // *Проблемы социальной гигиены, здравоохранения и истории медицины*. – 2024. – Т. 32, № 3. – С. 438-444. – DOI 10.32687/0869-866X-2024-32-3-438-444.
3. Джулай Д.В. Основные направления инновационной деятельности в Самарской области / Д.В. Джулай, А.М. Измаилов // *Тенденции развития современного общества: экономико-правовой аспект: Сборник научных трудов международной научно-практической конференции, Пенза, 14-15 ноября 2016 года*. – Пенза: Пензенский государственный технологический университет, 2016. – С. 26-28.
4. Измаилов А.М. Механизм управления информационно-знаниевыми ресурсами / А.М. Измаилов, С.И. Ашмарина // *Вестник Воронежского государственного университета инженерных технологий*. – 2016. – № 1(67). – С. 261-266. – DOI 10.20914/2310-1202-2016-1-261-266.
5. Бердников В.А. Формирование конкурентоспособного инновационно-образовательного звена в Самарско-Тольяттинской агломерации / В.А. Бердников, А.М. Измаилов // *Наука XXI века: актуальные направления развития*. – 2016. – № 1-1. – С. 45-49.
6. Фомин Е.П. Особенности среды функционирования современного промышленного предприятия / Е.П. Фомин, А.М. Измаилов // *Вестник Самарского государственного экономического университета*. – 2015. – № 9(131). – С. 108-113.

**KULYABIN Igor Andreevich**

Student, Volga Region State University of Telecommunications and Informatics,  
Russia, Samara

**SHIYANOVA Valeria Dmitrievna**

Student, Volga Region State University of Telecommunications and Informatics,  
Russia, Samara

*Scientific Advisor – Associate Professor of the Department of Digital Economics  
at the Volga State University of Telecommunications and Informatics,  
Candidate of Economic Sciences Izmailov Ayrat Maratovich*

## **PROSPECTS FOR THE INTEGRATION OF BUSINESS INTELLIGENCE AND ARTIFICIAL INTELLIGENCE**

**Abstract.** *Business analytics (BA) and artificial intelligence (AI) play a key role in shaping competitive advantages in today's business space. BA focuses on descriptive and diagnostic analytics, providing a retrospective assessment of data, while AI covers predictive and prescriptive analytics, offering predictions and recommendations for future actions. Their integration creates powerful synergies, significantly improving forecasting accuracy, decision-making automation, and operational efficiency. The article examines in detail the key advantages, possible challenges and prospects for the integration of BA and AI, as well as their impact on various sectors of the economy.*

**Keywords:** *AI, automation, business analytics, data protection, cognitive analytics.*

ЛОТЫШ Николай Игоревич

студент,

МИРЭА – Российский технологический университет,  
Россия, г. Москва

## АНАЛИЗ СУЩЕСТВУЮЩЕЙ СТЕПЕНИ ЗАЩИЩЕННОСТИ ИНФРАСТРУКТУРЫ ПРЕДПРИЯТИЯ

**Аннотация.** В представленной статье проводится всесторонний анализ инфраструктуры медицинского центра «Здоровье» с акцентом на информационную безопасность персональных данных пациентов. Рассматриваются организационная структура, информационные потоки и архитектура локальной вычислительной сети (ЛВС) учреждения.

**Ключевые слова:** информационная безопасность, медицинский центр, персональные данные, ИСПДн, модель угроз, модель нарушителя, защищенная база данных.

В настоящей работе будет исследована инфраструктура медицинского центра «Здоровье». Основные цели и задачи центра «Здоровье» заключаются в лечении и

восстановлении пациентов, перенесших травму или инвалидность.

На рисунке 1 схема инф. потоков.

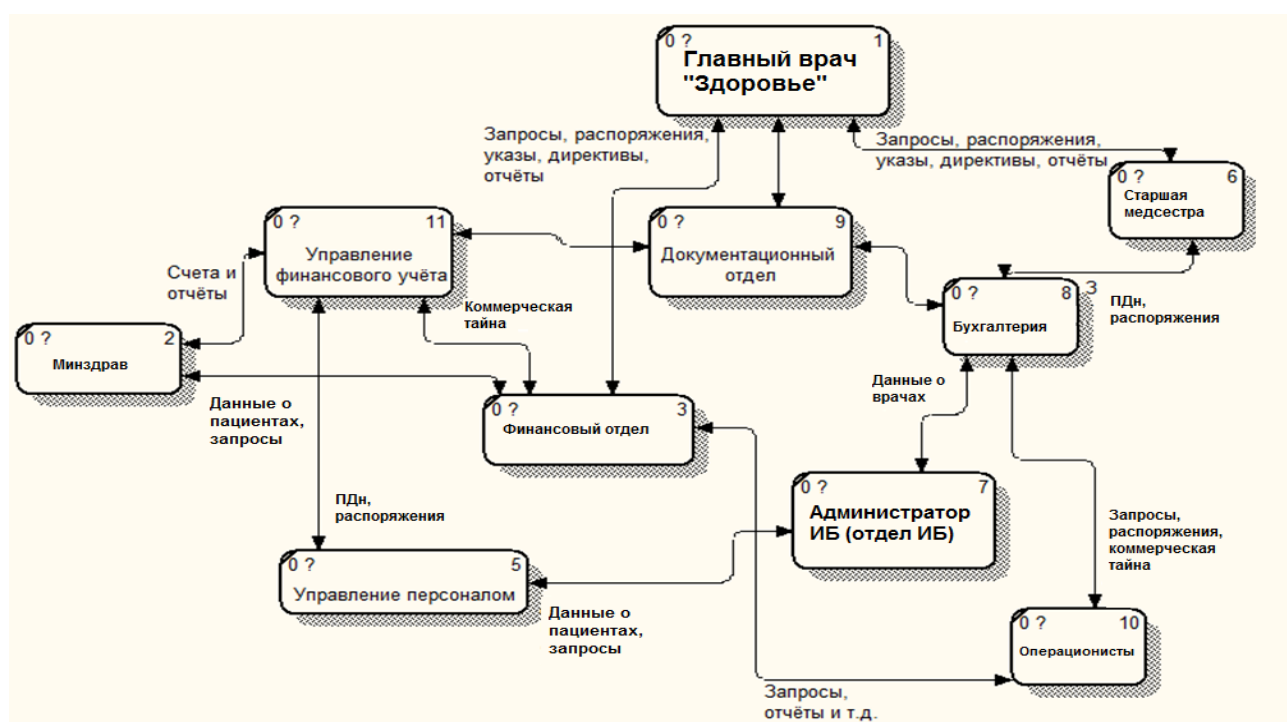


Рис. 1. Схема инф. потоков

На рисунке 2 представлена организационная структура.

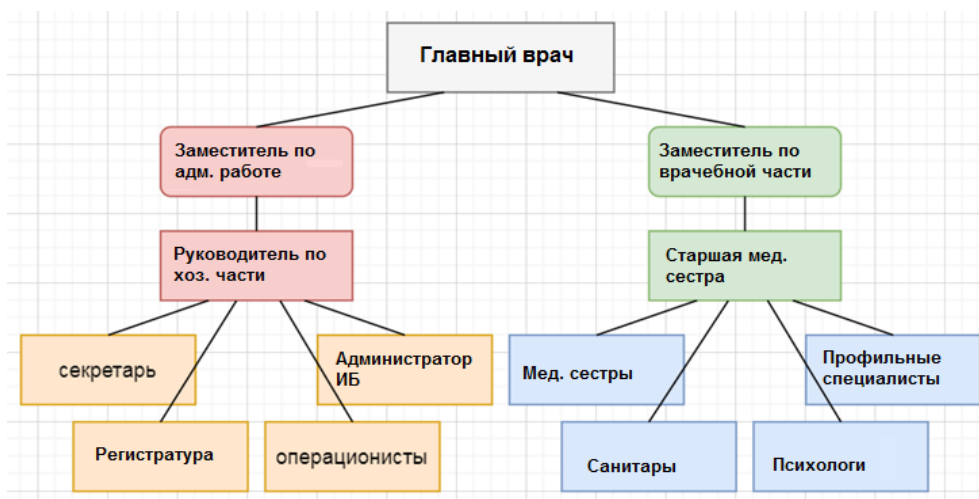


Рис. 2. Организационная структура

В рамках исследования анализируется инфраструктура и базы данных медицинского центра «Здоровье».

Исследуемый центр работает с конфиденциальной информацией, в частности,

персональными данными. Представленную информацию необходимо защищать в соответствии с Приказом ФСТЭК России № 21 от 18.02.2013 г. Требуется защищенная БД.

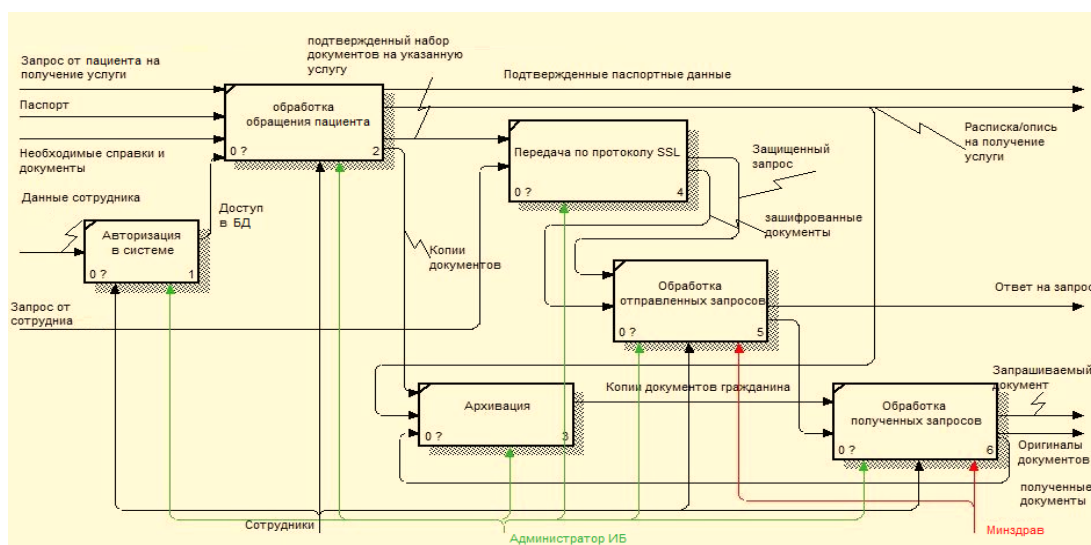


Рис. 3. Структура базы данных, где хранятся перс. данные «как есть»

На рисунке 3 представлена реализация текущей базы данных «как есть».

Процесс «Авторизация в системе» (процесс 1) происходит с использованием данных врачебного специалиста. После успешной авторизации сотрудник имеет доступ к электронному кабинету. «Обработка обращения пациента» (процесс 2) начинается с приема паспорта, далее подтвержденный набор документов копируется и архивируется (процесс 3). Также набор документов передается по защищенному соединению, используя протокол SSL (процесс 4). «Обработка отправленных запросов» (процесс 5)

получает документы по защищенному запросу, минздрав обрабатывает полученную информацию и выдает ответ на запрос. Это документ, в котором указаны сроки предоставления услуги и детали самой услуги. «Обработка полученных запросов» (процесс 6) выполняется тогда, когда документы по запросу пациента уже прибыли в систему. Теперь клиент предоставляет свои документы, в которых указано, какую именно медицинскую услугу он запрашивал, после ввода данных в систему, врач печатает и выдает нужный документ клиенту.

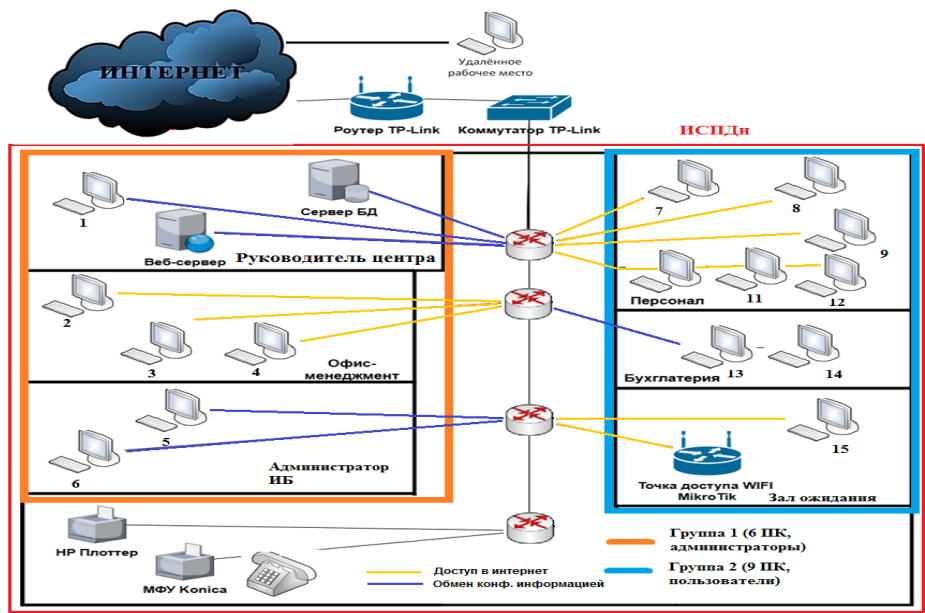


Рис. 4. Схема ЛВС центра «Здоровье»

Как следует из рисунка 4, в ЛВС присутствуют две группы пользователей:

- группа 1 (администраторы ИСПДн);
- группа 2 (пользователи ИСПДн).

Доступом к персональным данным, которые хранятся в базе данных, обладают две группы пользователей.

В соответствии с Постановлением Правительства РФ от 1 ноября 2012 г. № 1119, определен уровень исходной защищенности ИСПДн. Необходимо соблюдать 4 уровень защищенности для обработки ПДн в ИСПДн.

**Определение уровня исходной защищенности**

Центр «Здоровье» обладает общедоступными ПДн, количество субъектов которых не превышает 100 000; обрабатываются данные сотрудников и иных лиц (пациентов); для рассматриваемой ИСПДн актуальны угрозы третьего типа, не связанные с наличием недокументированных возможностей (НДВ) в используемом системном и прикладном ПО. Таким образом, необходимо соблюдать **4 уровень** защищенности для обработки ПДн.

Тип ИСПДн	Сотрудники оператора	Количество субъектов	Тип актуальных угроз		
			1 (НДВ ОС)	2 (НДВ ПО)	3 (Без НДВ)
ИСПДн специальные	Нет	> 100 000	УЗ-1	УЗ-1	УЗ-2
	Нет	< 100 000	УЗ-1	УЗ-2	УЗ-3
	Да				
ИСПДн биометрические			УЗ-1	УЗ-2	УЗ-3
ИСПДн иные	Нет	> 100 000	УЗ-1	УЗ-2	УЗ-3
	Нет	< 100 000	УЗ-2	УЗ-3	УЗ-4
	Да				
ИСПДн общедоступные	Нет	> 100 000	УЗ-2	УЗ-2	УЗ-4
	Нет	< 100 000	УЗ-2	УЗ-3	УЗ-4
	Да				

Рис. 5. Определение уровня защищенности ПДн

### Построение модели угроз и модели нарушителя

На рисунке 6 представлена модель угроз. Актуальными угрозами признаны угрозы, связанные с НСД к защищаемой информации. Построенная основываясь на Методическом документе ФСТЭК от 5 февраля 2021 г. «Методика оценки угроз безопасности».

УБИ возможна, если реализация угрозы может привести к негативным последствиям. В рисунке 7 представлена модель актуальных угроз, определенных на основе банка угроз и уязвимостей ФСТЭК. Негативное последствие – кража ПДн.

Угрозы	Категория нарушителя	Перечень объектов воздействия	Описание способов реализации угроз	Негативные последствия
УБИ.074 УБИ.086 УБИ.088 УБИ.113	Хакеры; обслуживавший персонал центра; авторизованные сотрудники	ИСПДн; Системное и прикладное ПО; ЛВС; Сетевой трафик.	Использование известных уязвимостей, внедрение ВПО	Нарушение конфиденциальности (утечка), потеря целостности и доступности ПДн работников. Финансовый и/или иной материальный ущерб физическому лицу.
УБИ.140 УБИ.143 УБИ.157 УБИ.158 УБИ.178 УБИ.179 УБИ.185 УБИ.187	Преступные группы; поставщики вычислительных услуг		Использование известных уязвимостей, внедрение ВПО, изменение настроек конфигурации	
УБИ.188 УБИ.191 УБИ.195	Террористические, экстремистские группировки; Разработчики ПО и оборудования		Использование уязвимостей конфигурации ПО, извлечение аутентификационной информации из ЭНИ, нецелевое администрирование сети	

Рис. 6. Модель угроз

Вид нарушителя	Цели	Возможности нарушителя	Возможные техники	Актуальность
Хакеры; обслуживавший персонал центра; авторизованные сотрудники	Получение финансовой или иной материальной выгоды.	Н1. Реализация только известных УБИ, направленных на известные уязвимости	T1.4, T1.5, T1.7, T1.8, T1.11, T1.15, T2.3, T2.4, T2.5, T2.8, T2.10, T2.11, T2.13, T3.1, T3.2, T3.3, T3.9, T3.10, T4.1, T5.6, T6.1, T6.3, T6.9, T7.18, T7.21, T7.25, T7.26.	Актуально
Преступные группы; поставщики вычислительных услуг		Н2. Реализация УБИ, в том числе направленных на неизвестные уязвимости	T1.7, T1.8, T1.11, T2.3, T2.4, T2.5, T2.8, T2.11, T2.13, T3.1, T3.3, T3.9, T4.1, T4.4, T5.6, T5.13, T6.9, T7.18, T7.21, T7.25, T8.5, T9.3, T9.7, T10.1, T10.3, T10.4.	Актуально
Террористические, экстремистские группировки; Разработчики ПО и оборудования		Н3. Реализация УБИ, в том числе на выявленных ими неизвестных уязвимостей, с использованием самостоятельно разработанных для этого инструментов.	T1.7, T1.8, T1.11, T2.3, T2.4, T2.5, T2.8, T2.11, T2.13, T3.1, T3.3, T3.9, T4.1, T4.4, T5.6, T5.13, T6.9, T7.18, T7.21, T7.25, T8.5, T9.3, T9.7, T10.1, T10.3, T10.4.	Актуально

Рис. 7. Модель нарушителя

### Требование к информационной системе

Требования к информационной системе формируются на основе понятия защищенной БД, исходя из темы настоящего исследования. Защищенная база данных – это БД, которая

защищена от НСД, использования, раскрытия, изменения или уничтожения ее информации. Представленная защищенность достигается за счет различного обеспечения.

### Функциональные требования:

- обеспечить хранение в БД всей информации, обрабатываемой в медицинском центре;
- автоматизировать процесс получения запросов и отчетов;
- автоматизировать процесс обновления карточки пациента; авторизация при входе в систему при работе с БД;
- средство доверенной загрузки при входе в систему.

### Нефункциональные требования:

- операционная система Astra Linux Special Edition;
- использование Ред База Данных 5;
- минимум 10 ТБ свободного места на диске;
- не менее 32 Гб оперативной памяти.

### Разработка проекта системы защиты предприятия

В рамках разработки защищенной БД был улучшен алгоритм работы (рис. 8).

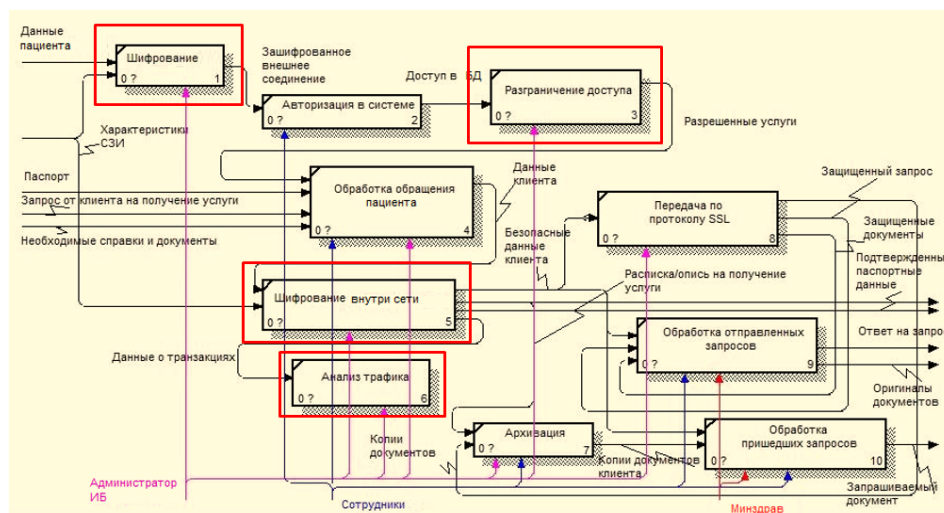


Рис. 8. Структура базы данных, где хранятся персональные данные, «как стало»

В диаграмме добавляются четыре новых процесса. В «Шифрование» (процесс 1) происходит настройка и шифрование всех соединений, поступающих из Интернета во внутреннюю сеть. Данная мера обеспечила создание демилитаризованной зоны. Появился процесс «Разграничение доступа» (процесс 3), где пользователю дается только определенный список услуг и привилегий. Также появился процесс «Шифрование внутри сети», (процесс 5) для

обеспечения защиты внутренних соединений внутри БД и передаваемой информации, и процесс «Анализ трафика» (процесс 6), с помощью которого можно поддерживать систему в безопасном состоянии.

Обновленный вариант позволит полностью автоматизировать процесс получения запросов и отчетов, а также автоматизировать процесс обновления карточки пациента.

LATYSH Nikolay Igorevich

Student, MIREA – Russian University of Technology, Russia, Moscow

## ANALYSIS OF THE EXISTING DEGREE OF SECURITY OF THE ENTERPRISE INFRASTRUCTURE

**Abstract.** The presented article provides a comprehensive analysis of the infrastructure of the medical center "Health" with an emphasis on the information security of patients' personal data. The organizational structure, information flows and architecture of the local computer network (LAN) of the institution are considered.

**Keywords:** information security, medical center, personal data, ISPDn, threat model, intruder model, secure database.

**МУГИНОВ Тимур Ильдарович**

магистрант, Казанский (Приволжский) федеральный университет, Россия, г. Казань

*Научный руководитель – доцент Казанского (Приволжского) федерального университета,  
кандидат экономических наук Вахитов Галим Зарибзянович*

## **СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДОВ КЛАСТЕРИЗАЦИИ СОЦИАЛЬНЫХ ГРАФОВ НА ДАННЫХ СОЦИАЛЬНОЙ СЕТИ «ВКОНТАКТЕ» ПО МЕРЕ СХОДСТВА ДВУХ РАЗБИЕНИЙ И МОДУЛЯРНОСТИ**

**Аннотация.** В данной статье представлен подход к построению взвешенного неориентированного графа пользователей социальной сети «ВКонтакте», где веса рёбер формируются на основе суммы Jaccard-сходств по спискам групп и общих друзей. Особое внимание уделяется сравнению результатов кластеризации по метрикам модулярности и нормализованной взаимной информации (NMI). Полученные данные демонстрируют значительное расхождение в эффективности алгоритмов и подчёркивают важность корректного выбора метода кластеризации в зависимости от структуры и параметров графа.

**Ключевые слова:** кластеризация, социальные графы, алгоритмы кластеризации, анализ социальных сетей.

**Введение:** социальные сети представляют собой графы, которые могут быть ориентированными – с направленными связями, или взвешенными, где каждой связи присвоено определённое значение. Это открывает возможность использования методов кластеризации, опирающихся на графовую теорию.

Упрощая граф, устраняется направленность и добавляется вес к существующим связям, преобразуя его в матрицу, что позволяет получить данные, эквивалентные набору характеристик несвязанных объектов. В таком контексте связь между вершинами становится дополнительной характеристикой, что позволяет применять стандартизированные методы кластеризации, не прибегая к графовым теориям.

Выбор алгоритма кластеризации зависит от характеристик и типа данных, которые используются при построении графа. Применяя различные алгоритмы, можно выявлять кластеры и ключевые узлы, что способствует лучшему пониманию динамики социальных связей. Важно учитывать, что структура сетей играет решающую роль: кластеры с высокой связностью могут оказывать более сильное влияние на участников, чем разрозненные группы [1].

В этом контексте особое внимание уделяется алгоритмам, которые используют доступные данные о пользователях, позволяя преобразовывать информацию в количественные показатели для определения степени близости

между узлами. На основе данных из социальной сети можно вычислять веса рёбер графа, выявляя более значимые связи между пользователями.

При построении графа необходимо учитывать как множества (узлы), так и ребра (бинарные отношения) между ними. Особенно важным является параметр веса, который отражает характер взаимодействия пользователей. Работа с API «ВКонтакте» позволяет получать необходимые данные для расчёта близости построенных отношений [2]. А в качестве весовых признаков использовать совместные показатели пользователей.

При этом важно продолжать совершенствовать информативность социального графа: эксперименты по выбору параметров и анализ их влияния на вес рёбер открывают интересные возможности для исследования и экспериментов.

### **Сбор и обработка данных**

В данной работе учтены дружеские связи второго уровня (друзья друзей) «обезличенного» пользователя социальной сети «ВКонтакте». Для этого было использовано API данной сети. Был прописан модуль на языке Python для получения данных пользователей, учитывались возможные ошибки со стороны API, ограничение на число запросов кратным 5000, а также для увеличения скорости обработки методы обрабатывались «батч»



запросами. В модуле учитывалось получение количественных, качественных данных и текстовых описаний групп и постов пользователей. Полученные данные записывались в разных форматах pkl, gml, json. В дальнейшем полученные характеристики пользователя будут использованы для модификации текущего кластерного анализа.

В частности, использовались следующие методы VK\_API:

- friends.get – метод для получения списка друзей.
- friends.getMutual – метод для поиска общих друзей (создания рёбер в графе).
- groups.get – метод для получения групп друзей.

На основании собранных данных формируется взвешенный неориентированный граф, в котором каждая вершина соответствует пользователю социальной сети. Для каждого узла

$$\text{weight} = \frac{\text{common\_groups}}{\text{total\_groups}_u + \text{total\_groups}_v - \text{common\_groups}} + \frac{\text{common\_friends}}{\text{total\_friends}_u + \text{total\_friends}_v - \text{common\_friends}}, \quad (1)$$

Где:

weight – весовая категория атрибута;

common\_groups – количество общих групп пары узлов;

common\_friends – количество общих друзей пары узлов;

$\text{total\_groups}_u + \text{total\_groups}_v -$

common\_friends – объединение групп;

$\text{total\_friends}_u + \text{total\_friends}_v -$

common\_friends – объединение друзей.

Этот подход позволяет учитывать как общие интересы (группы), так и социальную близость

извлекается список сообществ (групп), в которых состоит пользователь, с помощью функции get\_user\_groups, а также вычисляется количество общих друзей с другими пользователями через метод get\_mutual\_friends.

Для определения весов рёбер было принято решение использовать метрику Jaccard-сходства, так как она позволяет количественно оценить степень перекрытия между множествами – в данном случае, между списками групп и друзьями. Итоговый вес каждого ребра рассчитывается как сумма Jaccard-сходств по двум аспектам: группам и друзьям. Это позволяет учесть как тематическую близость пользователей (через интересы), так и их структурную связанность (через общих друзей), делая граф более информативным для последующего анализа и кластеризации. Итоговая формула для расчёта веса ребер графа для каждой пары узлов:

(общие друзья), что делает граф более информативным для анализа социальных отношений.

Чтобы улучшить качество дальнейшего анализа, стоит также обработать полученные данные. Для этого удаляем изолированные узлы. Это позволит сосредоточиться на наиболее значимых связях. В результате был получен граф с 1549 узлами, 1628 связями и общим количеством уникальных групп 79011. Визуализация одной из окрестностей графа с помеченными на удаление узлами на рисунке.

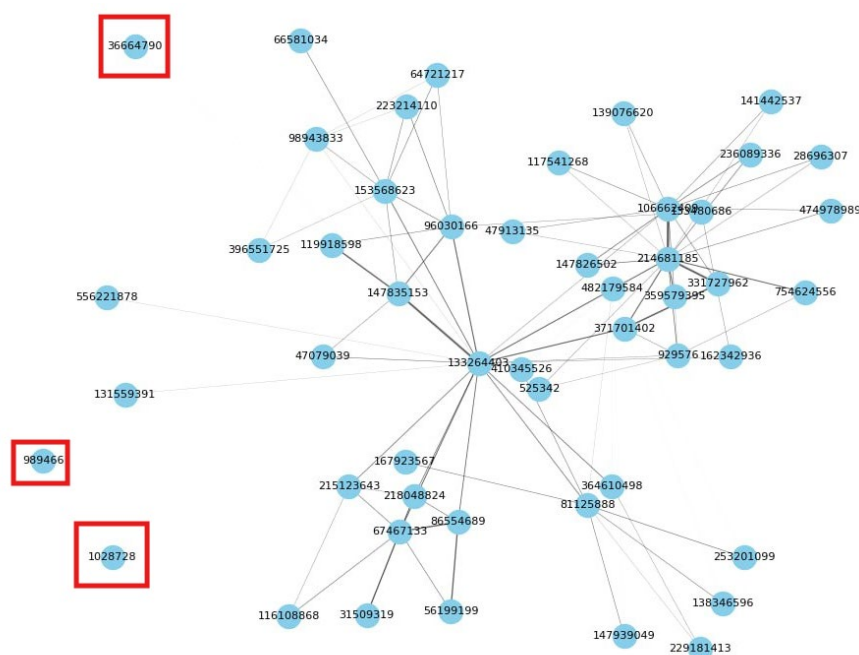


Рис. Визуализация выборки из графа

### Методы кластеризации

Для кластеризации полученного графа были использованы следующие алгоритмы: Edge Betweenness (Метод Гирвана-Ньюмана) [3, с. 7821-7826], Label Propagation (Распространение меток) [4], Fastgreedy (Жадная оптимизация модульности) [5] и Louvain (Оптимизация модульности) [6]. В дальнейшем за  $n$  будем обозначать количество вершин в графе, а за  $m$  – количество ребер.

Метод Fastgreedy (Greedy Modularity Optimization) представляет собой иерархический агломеративный алгоритм, ориентированный на максимизацию модульности – функции, оценивающей качество разбиения графа на сообщества. Алгоритм начинается с того, что каждая вершина рассматривается как отдельное сообщество. На каждом шаге он жадно объединяет такие пары сообществ, слияние которых приводит к наибольшему приросту модульности. Этот процесс повторяется до тех пор, пока модульность перестаёт возрастать или все вершины не объединены в одно сообщество.

Основным преимуществом Fastgreedy является прямолинейность и высокая интерпретируемость, поскольку на каждом шаге можно отслеживать, какие объединения происходят и почему. Однако он может «застрять» в локальном максимуме, не находя глобально оптимального разбиения. В отличие от Louvain, Fastgreedy не использует фазу агрегации и работает исключительно по стратегии последовательного слияния.

Алгоритм эффективно работает с взвешенными графами, учитывая силу связей между узлами. Тем не менее он плохо масштабируется на очень большие графы, поскольку требует повторного пересчёта модульности на каждом шаге.

В типичном случае временная сложность Fastgreedy составляет  $O(n \log^2 n)$  для разреженных графов, однако в общем случае она может достигать  $O(n^2 \log n)$ , что делает его пригодным в первую очередь для небольших и средних сетей.

Label Propagation. Алгоритм функционирует на основе эвристического правила, согласно которому вершина присоединяется к тому сообществу, которое преобладает среди её соседей. В начальной стадии каждая вершина рассматривается как отдельная группа. Далее, в ходе итераций, их порядок изменяется случайным образом, и каждое обновление

происходит на основании присвоенных меток соседних вершин. Этот процесс продолжается до момента, пока структура кластеров не стабилизируется. Из-за случайной природы метода несколько его запусков могут приводить к разным результатам, что делает его неустойчивым. Алгоритм характеризуется интуитивной логикой, простотой реализации и высокой вычислительной скоростью, так как его сложность приближается к  $O(m)$ .

Edge Betweenness. Разработанный Гирваном и Ньюманом метод базируется на вычислении центральности рёбер. В самом начале определяется через какие связи проходит наибольшее количество кратчайших путей между различными вершинами. Далее рёбра с наиболее высокими показателями центральности удаляются, а оставшиеся после этого компоненты рассматриваются как отдельные кластеры. Этот процесс продолжается до достижения максимального значения модулярности. Метод имеет несколько модификаций, основанных на изменении используемых метрик или применении альтернативных функций разбиения. Основным его недостатком является высокая вычислительная сложность, поскольку оценка центральности связей требует значительных ресурсов. Алгоритм работает за  $O(m^2n)$ .

Louvain. Метод основан на жадном поиске локального оптимума модулярности и использует технику локального перемещения вершин (LMH). Алгоритм последовательно перемещает вершины между сообществами таким образом, чтобы каждое действие увеличивало показатель модулярности. Обход вершин осуществляется в произвольном порядке и завершается, когда дальнейшие перемещения перестают приводить к улучшению структуры кластеров. В первой фазе алгоритм анализирует соседние вершины, поочередно изменяя их принадлежность к группам, если это способствует увеличению модулярности. Вторая фаза выполняет агрегацию: вершины одного сообщества сливаются в единый узел, а рёбра между ними преобразуются в новые связи. Этот процесс повторяется, пока итоговая структура графа остаётся неизменной, а модулярность не достигает локального максимума. Несмотря на то, что порядок обхода вершин может повлиять на производительность, он не оказывает значительного влияния на финальный результат. В среднем метод работает за  $O(n \log n)$ .

### Практическая часть

Результаты работы алгоритмов оценивались двумя метриками, модулярностью и NMI (мера сходств двух разбиений).

Модулярность (modularity) – это числовая метрика, которая измеряет, насколько хорошо граф разбит на сообщества. Она сравнивает долю рёбер внутри сообществ с ожидаемой долей таких рёбер в случайном графе той же степени связности. Значение модулярности  $Q$  лежит в диапазоне от -1 до 1 и интерпретируется следующим образом:

- $Q \rightarrow 1$  – отличное разбиение: большинство рёбер находится внутри сообществ, а не между ними. Это говорит о чётко выраженной кластерной структуре.
- $Q \approx 0$  – нейтральное (почти случайное) разбиение: рёбра расположены примерно так же, как в случайном графе, и сообществ как таковых не выделяется.
- $Q < 0$  – плохое разбиение: больше рёбер между сообществами, чем внутри них. Это означает, что текущая кластеризация хуже, чем случайное разбиение. Математически модулярность графа определяется как:

$$Q = \frac{1}{2m} \sum_{i,j} (A_{ij} - \frac{k_i k_j}{2m}) \delta(c_i, c_j), \quad (2)$$

Где:

$m$  – общее количество рёбер в графе;

$A_{ij}$  – элемент матрицы смежности (1, если есть ребро между  $i$  и  $j$ , иначе 0);

$k_i, k_j$  – степени вершин  $i, j$ ;

$\frac{k_i k_j}{2m}$  – ожидаемое число рёбер между  $i$  и  $j$  в случайном графе;

$\delta(c_i, c_j)$  – индикатор: 1, если  $i$  и  $j$  в одном кластере, иначе 0.

Так как в исходных данных отсутствовала истинная разметка (ground truth), отражающая принадлежность пользователей к конкретным сообществам, было принято решение создать её вручную. Для этого использовалась эвристическая стратегия, основанная на предположении, что пользователи с большим количеством общих друзей и групп, скорее всего, принадлежат к одному сообществу. Если у пользователя не было выраженной связи с другими (например, слишком мало общих групп и друзей), ему присваивалась метка ближайшего соседа – то есть того пользователя, с которым у него наибольший вес связи в графе. Чтобы оценить, насколько хорошо алгоритмы кластеризации

находят сообщества, согласующиеся с этой гипотезой, использовалась метрика нормализованной взаимной информации (NMI). Она позволяет сравнивать результат кластеризации с эталонной (эвристической) разметкой. Чем выше значение NMI, тем сильнее совпадение между тем, что «нашёл» алгоритм, и тем, что мы задали вручную.

**NMI (Нормализованная взаимная информация)** – это способ сравнить два разбиения множества, основанный на взаимной информации (например, предсказанные кластеры и ручную размеченные группы «ground truth»). В основе NMI лежит взаимная информация (MI) – она измеряет, насколько знание одного разбиения уменьшает неопределённость (энтропию) другого. NMI нормализует это значение, чтобы результат всегда находился в диапазоне от 0 до 1:

- 0 означает полное несоответствие между кластерами;
- 1 означает полное совпадение (разбиения идентичны).

Это делает NMI удобной для оценки качества кластеризации, особенно когда классы и кластеры не совпадают по числу или размеру, но могут всё равно хорошо соответствовать друг другу по составу.

Формула NMI метрики:

$$NMI(C, K) = \frac{I(C, K)}{\sqrt{H(C) \cdot H(K)}}, \quad (3)$$

Где:  $C$  – предсказанные кластеры;

$K$  – ground truth метки (истинные сообщества);

$I(C, K)$  – взаимная информация между  $C$  и  $K$ ;

$H(C)$  и  $H(K)$  – энтропии каждого разбиения.

Поскольку некоторые алгоритмы кластеризации обладают стохастической природой и могут выдавать различные результаты при повторных запусках (например, из-за случайного порядка обновления узлов или инициализации меток), для повышения достоверности оценки каждый метод был выполнен трижды. Далее для анализа рассматривались наилучшие значения метрик по результатам этих запусков. Итоговые значения модулярности и нормализованной взаимной информации (NMI), демонстрирующие наилучшее качество кластеризации для каждого метода, представлены в таблице.

Таблица

## Полученные оценки методов

Алгоритм	Модулярность	NMI
Louvain	0.5029	0.5751
Fastgreedy	0.4473	0.4954
Label Propagation	0.1616	0.4718
Edge Betweenness	0.3859	0.4714

Данные таблицы позволяют сделать вывод, что с учетом особенностей нашего социального графа и проставленной нами разметке Louvain демонстрирует наибольшие показатели как по модулярности, так и по NMI. Fastgreedy также показывает хорошие результаты, но уступает Louvain. Label Propagation и Edge Betweenness менее эффективны, особенно по критерию модулярности, что может быть связано с их спецификой работы на графовых социальных сетях.

**Заключение**

В рамках данной работы был реализован программный модуль, предназначенный для сбора пользовательских данных из социальной сети «ВКонтакте». На основе собранных данных был сформирован взвешенный неориентированный граф, в котором вершины соответствуют пользователям, а рёбра отражают степень социальной близости, рассчитываемую на основе количества общих друзей и пересечения интересов (групп). Вес рёбер формировался как сумма Jaccard-сходств по этим двум аспектам, что позволило учесть как структурные, так и семантические характеристики пользовательских связей.

Для анализа полученного графа был проведён сравнительный эксперимент с применением наиболее известных алгоритмов детекции сообществ: Louvain, Fastgreedy, Label Propagation и Edge Betweenness. Рассмотрены теоретические и вычислительные особенности указанных методов, проанализирована устойчивость разбиений и их интерпретируемость. Каждый алгоритм оценивался с использованием таких метрик, как модулярность (modularity) и нормализованная взаимная информация (NMI), отражающая соответствие разбиения заданным меткам.

По результатам экспериментов, наилучшие показатели продемонстрировали алгоритмы Louvain и Fastgreedy, причём первый показал

более стабильную и интерпретируемую структуру сообществ. Несмотря на высокую модулярность, метод Fastgreedy оказался менее эффективным в обнаружении «естественных» кластеров. Алгоритмы без учёта весов (например, Label Propagation) показали меньшую точность и высокую вариативность результатов между запусками.

Таким образом, было продемонстрировано, что качественный выбор алгоритма детекции сообществ, адаптированного под особенности конкретного графа и весовой модели, является критически важным для получения содержательной структуры кластеров и извлечения значимой информации из социальной сети.

**Литература**

1. Васильев В.М. Анализ и визуализация социальных сетей: методы и алгоритмы / В.М. Васильев. – Москва: Наука, 2020. – 256 с.
2. Документация по API ВКонтакте [Электронный ресурс]. – Режим доступа: <https://dev.vk.com/reference> – Дата обращения: 20.05.2025.
3. Girvan M., Newman M.E.J. Community structure in social and biological networks // Proceedings of the National Academy of Sciences. 2002. Vol. 99, No. 12. P. 7821-7826.
4. Raghavan U.N., Albert R., Kumara S. Near linear time algorithm to detect community structures in large-scale networks // Physical Review E. 2007. Vol. 76, No. 3. Article 036106.
5. Clauset A., Newman M.E.J., Moore C. Finding community structure in very large networks // Physical Review E. 2004. Vol. 70, No. 6. Article 066111.
6. Blondel V.D., Guillaume J.-L., Lambiotte R., Lefebvre E. Fast unfolding of communities in large networks // Journal of Statistical Mechanics: Theory and Experiment. 2008. No. 10. Article P10008.

**MUGINOV Timur Ildarovich**

Master's Student, Kazan (Volga Region) Federal University, Russia, Kazan

*Scientific Advisor – Associate Professor of Kazan (Volga Region) Federal University,  
Candidate of Economic Sciences Vakhitov Galim Zaribzyanovich*

## **COMPARATIVE ANALYSIS OF SOCIAL GRAPH CLUSTERING METHODS BASED ON VKONTAKTE SOCIAL NETWORK DATA AS THE TWO PARTITIONS ARE SIMILAR AND MODULAR**

**Abstract.** *This article presents an approach to constructing a weighted undirected graph of users of the VKontakte social network, where edge weights are formed based on the sum of Jaccard similarities across lists of groups and mutual friends. Particular attention is paid to comparing clustering results based on the metrics of modularity and normalized mutual information (NMI). The data obtained demonstrate a significant discrepancy in the efficiency of the algorithms and emphasize the importance of choosing the correct clustering method depending on the structure and parameters of the graph.*

**Keywords:** *clustering, social graphs, clustering algorithms, analysis of social networks.*

**ОРУДЖОВА Сабина Рамин**

магистрантка, Азербайджанский государственный университет нефти и промышленности,  
Азербайджан, г. Баку

**ГАСАНГУЛИЕВА Метанет Мухаммад**

доцент, Азербайджанский государственный университет нефти и промышленности,  
Азербайджан, г. Баку

**АНАЛИЗ БОЛЬШИХ ДАННЫХ И СТАТИСТИЧЕСКИХ ПРИМЕНЕНИЙ**

**Аннотация.** Большие данные и их статистические применения играют решающую роль в масштабном сборе, хранении, обработке и анализе данных в современную эпоху. Под большими данными понимаются огромные объемы информации, которые характеризуются высокой скоростью и разнообразием – это такие наборы данных, с которыми традиционные технологии обработки данных не справляются. Эти данные собираются из таких источников, как социальные сети, сенсоры, устройства Интернета вещей, бизнес-операции и другие, и они стали ценным активом для компаний, организаций и государственных учреждений. Статистические приложения необходимы для реализации полного аналитического потенциала больших данных. Методы статистического анализа включают процессы описательной статистики, анализа случайных величин, корреляционного и регрессионного анализа, кластеризации и классификации. Эти приложения позволяют выявлять значимые закономерности и тенденции в больших данных, оптимизировать процессы принятия решений и разрабатывать прогностические модели. В современном мире большие данные и статистические приложения используются в различных секторах, включая анализ поведения клиентов в бизнесе и маркетинге, прогнозирование продаж и персонализированные предложения; в здравоохранении – для анализа медицинских данных, прогнозирования заболеваний и разработки персонализированного лечения; в финансовом секторе – для анализа рынков, выявления мошенничества и оценки кредитных рисков; в транспорте – для оптимизации транспортных сетей и повышения эффективности маршрутов; а также в науке и исследованиях – для анализа геномных данных, изучения изменения климата и многого другого. С помощью статистических приложений организации могут извлекать ценные сведения из данных, внедрять эффективные процессы принятия решений и поддерживать стратегии будущего развития. Кроме того, статистический анализ больших данных, в сочетании с такими технологиями, как искусственный интеллект и машинное обучение, способствует получению глубоких инсайтов из сложных наборов данных.

**Ключевые слова:** большие данные, статистические данные, машинные данные.

**Введение**

В современную эпоху объем и разнообразие данных стремительно увеличиваются. С развитием цифровых технологий и интернета ежедневно по всему миру генерируются триллионы единиц данных. Эти данные собираются через социальные сети, платформы электронной коммерции, датчики, мобильные приложения, государственные службы и другие цифровые источники. Из-за их большого объема, быстрого изменения и разнообразных форматов, такие данные выходят за пределы возможностей традиционных систем обработки данных. Это явление усилило значение больших данных и их статистических приложений.

Большие данные характеризуются тремя основными признаками: объем (volume), скорость (velocity) и разнообразие (variety). Объем отражает масштаб данных; скорость указывает на то, как быстро данные генерируются и обрабатываются; разнообразие охватывает наличие структурированных, неструктурированных и полуструктурированных типов данных. Эти особенности отличают большие данные от традиционных и требуют новых подходов к их анализу.

Статистический анализ больших данных позволяет извлекать ценные сведения. Он помогает понять структуру и поведение данных, выявить закономерности и тренды, установить взаимосвязи между различными переменными

и делать прогнозы на будущее. Этот процесс включает применение различных статистических методов и технологий. Описательная статистика использует такие меры, как среднее значение, медиана и дисперсия, чтобы охарактеризовать общие свойства данных. Корреляционный и регрессионный анализ позволяют выявить взаимосвязи между переменными и определить их взаимодействие. Кластеризация и классификация разделяют данные на различные категории, формируя группы с похожими признаками. Алгоритмы случайных лесов и нейронные сети, входящие в сферу искусственного интеллекта и машинного обучения, обеспечивают анализ сложных данных и построение прогностических моделей.

Совместное использование больших данных и статистических методов приводит к значительным результатам в различных областях. В бизнесе и маркетинге анализ поведения клиентов, персонализированные маркетинговые стратегии и повышение точности прогнозирования продаж стали возможными. В здравоохранении анализ медицинских данных способствует своевременной диагностике,

внедрению персонализированных методов лечения и оптимизации затрат на здравоохранение.

### Анализ

Анализ больших данных и статистических приложений является основным подходом в обработке и анализе современных данных. Он охватывает основные характеристики больших данных, применение статистических методов и способы их использования в различных секторах.

Для глубокого понимания статистического анализа больших данных необходимо рассмотреть его ключевые компоненты, аналитические методы и практические приложения. Большие данные представляют собой большие объемы разнородной и быстро поступающей информации. Эти данные собираются из социальных сетей, сенсоров, мобильных устройств, клиентских транзакций и других источников. Поскольку традиционные системы обработки данных не справляются с такими масштабами, обработка больших данных требует специализированных подходов [9, с. 140].

Таблица

**Ключевые характеристики и области применения больших данных и статистического анализа (источник: <https://pmc.ncbi.nlm.nih.gov/articles/PMC5041595/>)**

Характеристика / Область применения	Описание
<b>Объем (Volume)</b>	Сбор и хранение больших объемов данных. Например, ежедневно публикуемые данные в социальных сетях, данные, генерируемые сенсорами и др.
<b>Скорость (Velocity)</b>	Высокая скорость генерации и обработки данных. Управление и анализ потоков данных в реальном времени.
<b>Разнообразие (Variety)</b>	Данные, собранные в различных форматах (текст, изображение, видео, аудио и т. д.) и из множества источников (социальные сети, сенсоры, лог-файлы и т. д.).
<b>Достоверность (Veracity)</b>	Качество и надежность данных. Устранение ошибок и неточностей в анализе данных.
<b>Ценность (Value)</b>	Полезные знания и выводы, полученные из собранных данных. Использование данных в бизнесе и процессах принятия решений.
<b>Описательная статистика</b>	Использование таких показателей, как среднее, медиана и дисперсия для выявления общих характеристик данных.
<b>Корреляция и регрессия</b>	Выявление взаимосвязей и влияний между данными. Например, анализ взаимосвязи между продажами и погодными условиями.
<b>Кластеризация и классификация</b>	Группировка и классификация данных по схожим характеристикам. Например, сегментация клиентов.
<b>Искусственный интеллект и МО</b>	Автоматическое извлечение закономерностей и инсайтов из данных. Создание прогностических моделей и автоматизация принятия решений.
<b>Бизнес и маркетинг</b>	Анализ поведения клиентов, прогнозирование продаж и разработка персонализированных маркетинговых стратегий.
<b>Здравоохранение</b>	Анализ медицинских данных, прогнозирование заболеваний и разработка персонализированных методов лечения.
<b>Финансы</b>	Анализ финансовых рынков, выявление мошенничества и оценка кредитных рисков.
<b>Транспорт</b>	Оптимизация дорожных сетей и выявление эффективных маршрутов.
<b>Наука и исследования</b>	Анализ геномных данных, изучение изменения климата и применение в других научных областях.

Большие данные и статистический анализ имеют огромное значение в современной сфере управления данными и аналитики. Таблица 1 показывает основные характеристики больших данных и области, в которых применяется статистический анализ. Объем, скорость и разнообразие – три фундаментальные особенности, которые отличают большие данные от традиционных методов обработки. В таблице также указано, как статистические методы, такие как описательная статистика, корреляция и регрессия, кластеризация и искусственный интеллект, применяются для анализа больших данных. Эти методы позволяют выявлять значимые закономерности в данных, определять взаимосвязи между переменными и делать прогнозы на будущее.

Таблица также поясняет, как статистический анализ используется в различных областях, таких как бизнес, здравоохранение, финансы, транспорт и наука. В этих секторах статистический анализ позволяет извлекать ценные сведения из данных, поддерживает более точное принятие решений и помогает оптимизировать операции.

В заключение, большие данные и статистический анализ являются незаменимыми инструментами для эффективного принятия решений и разработки инновационных решений в различных сферах.

**Статистический анализ – ключ к раскрытию полного потенциала больших данных.** Он позволяет выявлять значимые закономерности и тенденции, определять взаимосвязи между переменными и разрабатывать прогностические модели. Основные методы, используемые в статистическом анализе больших данных, включают:

**Описательная статистика:** определение общих характеристик данных. Включает такие метрики, как среднее, медиана, мода, дисперсия и стандартное отклонение. Например, описательная статистика по продажам позволяет компаниям определить средний объем продаж, а также минимальные и максимальные значения.

**Корреляционный и регрессионный анализ:** выявление взаимосвязей между переменными. Корреляция измеряет силу и направление связи между двумя переменными, а регрессия определяет, как одна переменная влияет на другую. Например, анализ взаимосвязи между расходами клиентов и затратами на маркетинг

может помочь оптимизировать маркетинговую стратегию.

**Кластеризация и классификация:** группировка данных на основе схожих характеристик. Методики, такие как K-средних и случайные леса, широко применяются в этом процессе. Например, клиентов можно сегментировать по покупательскому поведению для разработки персонализированных предложений.

**Случайные леса и нейронные сети:** эти методы искусственного интеллекта и машинного обучения автоматизируют статистический анализ больших данных. Они могут выявлять сложные закономерности и взаимосвязи путем анализа больших наборов данных. Например, нейронные сети используются в медицинской аналитике для прогнозирования заболеваний [5, с. 3].

**Статистический анализ больших данных широко применяется в различных областях и приносит значительную ценность. Ниже приведены ключевые области применения:**

- **Бизнес и маркетинг:** статистический анализ больших данных позволяет анализировать поведение клиентов, прогнозировать продажи и разрабатывать персонализированные маркетинговые кампании. Компании, такие как Amazon и Netflix, используют статистику для выдачи индивидуальных рекомендаций.

- **Здравоохранение:** статистические методы необходимы для анализа медицинских данных, прогнозирования заболеваний и разработки персонализированного лечения. Например, во время пандемии COVID-19 распространение заболевания прогнозировалось с помощью статистического анализа.

- **Финансы:** анализ финансовых рынков, выявление мошенничества и оценка кредитных рисков возможны благодаря анализу больших данных. Финансовые учреждения анализируют поведение клиентов для оптимизации решений о кредитовании.

- **Транспорт:** оптимизация транспортных сетей, отслеживание транспорта и определение эффективных маршрутов осуществляется с помощью аналитики больших данных. Платформы, такие как Uber и Google Maps, предоставляют маршруты в реальном времени на основе больших данных.

- **Наука и исследования:** статистический анализ больших данных способствует научным достижениям в таких областях, как геномика, изучение климата и космические исследования. Например, с помощью



статистических методов исследуются причины генетических заболеваний на основе анализа геномов.

**Преимущества использования больших данных и статистических технологий для организаций включают:**

- Более быстрое и точное принятие решений на основе данных.
- Лучшее понимание поведения клиентов и предоставление персонализированных услуг.
- Снижение рисков и максимизация прибыли за счет прогностических моделей.
- Оптимизация операционных процессов и сокращение затрат.
- Достижение новых открытий и инноваций в научных исследованиях.

**Несмотря на преимущества, большие данные и статистический анализ сталкиваются с рядом вызовов.** Это обеспечение конфиденциальности и безопасности данных, поддержание качества данных и создание необходимой технологической инфраструктуры для хранения и обработки больших объемов информации. Поэтому организациям необходимо уделять особое внимание надежности, конфиденциальности и безопасности данных при внедрении технологий больших данных [3, с. 215].

### **Заключение**

Большие данные и статистический анализ играют решающую роль в современном управлении данными и аналитике. Постоянно растущий объем данных делает традиционные методы обработки недостаточными. Объем, скорость и разнообразие больших данных требуют новых аналитических подходов. Статистический анализ – самый эффективный инструмент для реализации аналитического потенциала больших данных.

Благодаря таким методам, как описательная статистика, корреляция, регрессия, кластеризация и искусственный интеллект, становится возможным эффективное извлечение информации из данных. Применение больших данных и статистики в разных сферах позволяет организациям получать ценные знания.

В бизнесе и маркетинге это помогает анализировать поведение клиентов и строить стратегии продаж. В здравоохранении – улучшать диагностику и персонализацию лечения. В финансах – выявлять мошенничество и оценивать рынки. В науке – продвигать исследования в геномике и климатологии.

Однако этот процесс требует преодоления вызовов, таких как обеспечение качества данных, конфиденциальность и технологическая готовность. Организации должны гарантировать надежность и безопасность данных во всех сферах применения.

**Большие данные и статистический анализ – это мощные инструменты,** которые обеспечивают организациям конкурентное преимущество. Они позволяют извлекать знания, структурировать принятие решений и разрабатывать инновационные решения. По мере развития цифровой экономики и информационного общества значение этих инструментов будет только возрастать.

### **Литература**

1. Qasimova R. (2015). Problems of big data analysis. Institute of Information Technology, P. 1-7. [https://ict.az/uploads/konfrans/biq\\_data/1-7\\_Qasimova\\_Rena\\_Boyuk\\_Verilnlr\\_Analizinin\\_Problemlri\\_son.pdf](https://ict.az/uploads/konfrans/biq_data/1-7_Qasimova_Rena_Boyuk_Verilnlr_Analizinin_Problemlri_son.pdf).
2. Aliyev R., Haji M. (2013). “Big Data” technologies: Current situation and perspectives. Institute of Information Technology, P. 37-49. [https://ict.az/uploads/konfrans/GOOGLE\\_SCHOLAR\\_e-gov/37R.Alguliyev\\_M.Haci.pdf](https://ict.az/uploads/konfrans/GOOGLE_SCHOLAR_e-gov/37R.Alguliyev_M.Haci.pdf).
3. Mansurova S. (2019). Big data analytics technologies and personal data security issues. Information Security Conference, P. 215-216. [https://ict.az/uploads/konfrans/info\\_sec\\_2019/RS53\\_BIG\\_DATA\\_ANALYTICS\\_AND\\_PERSONAL\\_DATA\\_SECURITY\\_ISSUES.pdf](https://ict.az/uploads/konfrans/info_sec_2019/RS53_BIG_DATA_ANALYTICS_AND_PERSONAL_DATA_SECURITY_ISSUES.pdf).
4. Babayev S., Hasanov V. (2016). Big data and its management. Azerbaijan Journal of Science and Education, 1(4), P. 78-89. <https://www.researchgate.net/publication/391428081>.
5. Mammadov N. (2019). Digital economy. Center for Analysis of Economic Reforms and Communication, P. 1-200. <https://ereforms.gov.az/files/publications/pdf/az/9d03c762a9342224168be0a2ffc4e26a.pdf>.
6. Wang C., Chen M.-H., Schifano E., Wu J., Yan J. (2016). Statistical methods and computing for big data. Statistical Interface, 9(4), P. 399-414. <https://doi.org/10.4310/SII.2016.v9.n4.a1>.
7. Bhandari P. (2024, January 17). Reporting statistics in APA style: Guidelines & examples. Scribbr. <https://www.scribbr.com/apa-style/numbers-and-statistics/>.
8. American Psychological Association. (2020). Publication manual of the American

Psychological Association (7th ed.). Washington, DC: Author.

9. Gandomi A., Haider M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*,

35(2), P. 137-144. <https://doi.org/10.1016/j.ijinfo-mgt.2014.10.007>.

10. Kitchin R. (2014). *The data revolution: Big data, open data, data infrastructures and their consequences*. London: SAGE Publications.

**ORUJOVA Sabina Ramin Gızı**

Master's Degree, Azerbaijan State Oil and Industry University, Azerbaijan, Baku

**HASANGULIYEVA Metanet Muhammad**

Associate Professor, Azerbaijan State Oil and Industry University,  
Azerbaijan, Baku

## ANALYSIS OF BIG DATA AND STATISTICAL APPLICATIONS

**Abstract.** *Big Data and its statistical applications play a vital role in the modern process of large-scale data collection, storage, processing, and analysis. Big Data refers to datasets that are large in volume, rapidly changing, and diverse in format—making them difficult to handle using traditional data processing technologies. These data are collected from sources such as social media, sensors, Internet of Things (IoT) devices, business operations, and others, and have become a valuable resource for companies, organizations, and government institutions. Statistical applications are essential for fully realizing the analytical potential of Big Data. Methods of statistical analysis include processes such as data description, random variable analysis, correlation and regression analysis, clustering, and classification. Statistical applications in the context of Big Data allow for the discovery of meaningful patterns and trends, optimization of decision-making processes, and the creation of predictive models. Today, Big Data and statistical applications are used across many fields, including customer behavior analysis in business and marketing, sales forecasting, and personalized offers; medical data analysis in healthcare, disease prediction, and the development of personalized treatments; financial market analysis in the finance sector, fraud detection, and credit risk assessment; optimization of road networks and identification of efficient transport routes in the transportation industry; genomic data analysis in science and research, climate change studies, and more. Through the statistical use of Big Data, businesses and organizations can extract valuable insights, create effective decision-making processes, and support future development strategies. Moreover, the statistical analysis of Big Data enables the extraction of deep insights when combined with technologies such as artificial intelligence and machine learning.*

**Keywords:** *big data, statistical data, machine data.*

**СОКОЛОВ Иван Андреевич**

магистрант, Вологодский государственный университет, Россия, г. Вологда

*Научный руководитель – доцент кафедры автоматики и вычислительной техники*

*Вологодского государственного университета,*

*кандидат технических наук Сорокин Арсений Николаевич*

## **АРХИТЕКТУРНЫЕ ПОДХОДЫ К ОБЕСПЕЧЕНИЮ ОТКАЗОУСТОЙЧИВОСТИ И МАСШТАБИРУЕМОСТИ В OPEN-SOURCE СИСТЕМАХ КОНТРОЛЯ ВЕРСИЙ: СРАВНИТЕЛЬНЫЙ АНАЛИЗ GITLAB CE, GITEA И FORGEJO**

**Аннотация.** В статье проводится сравнительный анализ архитектурных решений современных open-source систем контроля версий (СКВ) – GitLab CE, Gitea и Forgejo – в контексте их способности обеспечить отказоустойчивость, масштабируемость и интеграцию с кластерными средами. Рассматриваются компонентная структура каждой платформы, методы реализации горизонтального масштабирования, поддержка георепликации и балансировки нагрузки. Особое внимание уделено практическим сценариям использования: от глобальных корпоративных проектов до стартапов и децентрализованных open-source сообществ. Результаты исследования демонстрируют, что выбор оптимальной СКВ зависит от масштаба проекта, требований к инфраструктуре и доступных ресурсов. GitLab CE выделяется встроенной кластеризацией и интеграцией с DevOps-инструментами, тогда как Gitea и Forgejo предлагают легковесные решения для малых команд. На основе анализа сформулированы рекомендации по внедрению систем в распределенные среды, включая совместимость с Kubernetes и стратегии минимизации затрат.

**Ключевые слова:** системы контроля версий, GitLab CE, Gitea, Forgejo, кластерные архитектуры, масштабируемость, отказоустойчивость, георепликация, горизонтальное масштабирование, DevOps, CI/CD, open-source.

### **Введение**

Современная разработка программного обеспечения требует не только эффективного управления исходным кодом, но и обеспечения высокой доступности, масштабируемости и отказоустойчивости систем контроля версий (СКВ). Open-source решения, такие как **GitLab CE**, **Gitea** и **Forgejo**, предлагают различные архитектурные подходы к реализации этих требований. В статье анализируются их архитектурные особенности, преимущества и ограничения, а также методы интеграции в кластерные среды для поддержки распределенных команд и крупных проектов:

### **1. GitLab CE: Многофункциональная платформа с встроенной кластеризацией** **Архитектурная модель**

GitLab CE (Community Edition) – это комплексная платформа, объединяющая управление репозиториями, CI/CD, мониторинг и аналитику. Его модульная архитектура разделена на несколько ключевых компонентов:

#### **1) GitLab Rails:**

- Основное веб-приложение на Ruby on Rails, обрабатывающее HTTP-запросы и API;

- Управляет пользователями, проектами и настройками.

#### **2) Gitaly:**

- Высокопроизводительный сервис на Go, отвечающий за операции с Git-репозиториями;
- Отделен от основного приложения для упрощения масштабирования и снижения нагрузки на CPU.

3) Sidekiq: фоновый обработчик задач на Ruby, используемый для асинхронного выполнения CI/CD-пайплайнов, уведомлений и других операций.

4) PostgreSQL и Redis: основная база данных (PostgreSQL) хранит метаданные проектов, а Redis используется для кэширования и очереди задач.

Кластерные возможности:

Горизонтальное масштабирование: компоненты GitLab Rails, Sidekiq и Gitaly могут быть развернуты на отдельных узлах. Например, несколько Gitaly-серверов распределяют нагрузку между репозиториями.

Интеграция с Kubernetes:

- GitLab Runner автоматически масштабирует CI/CD-задачи в Kubernetes-кластере.

- Поддержка GitLab Operator для управления жизненным циклом в облачных средах.

Георепликация (GitLab Geo): создание вторичных узлов в разных регионах для снижения задержек и обеспечения отказоустойчивости.

Преимущества:

- Полнофункциональность (встроенные DevOps-инструменты).
- Готовая поддержка кластеров и облачных инфраструктур.

Недостатки:

- Высокие требования к ресурсам (рекомендуется от 8 ГБ RAM для средних нагрузок).
- Сложность настройки HA-кластера (требует глубокого знания компонентов).

**2. Gitea: Легковесная и минималистичная альтернатива**

**Архитектурная модель**

Gitea – это упрощенная система управления Git-репозиториями, разработанная для малых команд. Ее архитектура ориентирована на минимализм и низкое потребление ресурсов:

1. Монолитное ядро на Go: веб-интерфейс, API и Git-операции выполняются в одном процессе, что упрощает развертывание.
2. База данных: поддерживает SQLite (для тестовых сред), MySQL и PostgreSQL.
3. Отсутствие зависимостей: для работы не требуются Redis или отдельные сервисы для фоновых задач.

Кластерные возможности:

- Репликация базы данных: настройка мастер-слейв репликации PostgreSQL для повышения доступности.
- Балансировка нагрузки: использование обратного прокси (Nginx, HAProxy) для распределения запросов между несколькими инстансами Gitea.
- Интеграция с внешними CI/CD: поддержка Drone, Jenkins и других инструментов через Webhooks.

Преимущества:

- Низкие системные требования (работает на 512 МБ RAM).
  - Простота установки и обслуживания.
- Недостатки:
- Отсутствие встроенной кластеризации.
  - Ограниченный функционал (нет встроенного CI/CD, мониторинга).

**3. Forgejo: Децентрализованный форк Gitea**

**Архитектурная модель**

Forgejo – это форк Gitea, созданный для обеспечения прозрачности управления и акцента на безопасности. Архитектурно он наследует черты Gitea, но включает улучшения:

1. Децентрализованное управление: разработка ведется сообществом без контроля единой организации.
2. Усиленная безопасность: регулярные аудиты кода и приоритизация исправлений уязвимостей.
3. Экспериментальные функции: поддержка Federation через ActivityPub (в разработке) для создания распределенной сети репозитория.

Кластерные возможности:

Репликация и балансировка:

- Аналогично Gitea, Forgejo полагается на внешние инструменты (Patroni для PostgreSQL, HAProxy).

Геораспределение:

- Возможность ручной настройки реплик в разных регионах через прокси.

Преимущества:

- Сообщество-ориентированная разработка.
- Акцент на децентрализацию и безопасность.

Недостатки:

- Меньшая популярность по сравнению с Gitea.
- Ограниченная документация по кластерным сценариям.

Таблица

**Сравнительный анализ архитектур**

Критерий	GitLab CE	Gitea	Forgejo
Язык разработки	Ruby (Rails), Go (Gitaly)	Go	Go
Базы данных	PostgreSQL, Redis	SQLite, MySQL, PostgreSQL	SQLite, MySQL, PostgreSQL
Встроенная кластеризация	Да (Gitaly Cluster, GitLab Geo)	Нет	Нет
CI/CD	Встроенный (GitLab CI)	Сторонние интеграции	Сторонние интеграции
Ресурсы	Высокие (от 4 ГБ RAM)	Низкие (от 512 МБ RAM)	Низкие (от 512 МБ RAM)
Целевая аудитория	Крупные предприятия, DevOps-команды	Небольшие команды, энтузиасты	Open-source сообщества, активисты

#### 4. Практические сценарии использования

##### Сценарий 1: Крупная компания с глобальной командой

Инструмент: GitLab CE; архитектура:

- Кластер Gitaly для распределения репозитория.
- GitLab Geo с узлами в ЕС, США и Азии.
- Kubernetes для автоскейлинга Runner-ов.

Результат: отказоустойчивость 99.99%, снижение задержек для удаленных команд.

##### Сценарий 2: Стартап с ограниченным бюджетом

Инструмент: Gitea; архитектура:

- Репликация PostgreSQL между двумя VPS.
- Балансировка нагрузки через Cloudflare.

Результат: годовая стоимость инфраструктуры – менее \$200.

##### Сценарий 3: Open-source проект с акцентом на безопасность

Инструмент: Forgejo; архитектура:

- Развертывание на распределенных серверах с использованием Tor-сетей.
- Интеграция с Matrix для децентрализованной коммуникации.

Результат: полная независимость от корпоративных платформ.

#### 5. Рекомендации по выбору:

1. **GitLab CE**: подходит для организаций, готовых инвестировать в мощную инфраструктуру и нуждающихся в полном DevOps-стеке.
2. **Gitea**: идеален для малых команд, стартапов и проектов с ограниченными ресурсами.

3. **Forgejo**: рекомендован для сообществ, ценящих децентрализацию, приватность и независимость.

#### Заключение

GitLab CE, Gitea и Forgejo представляют три разных подхода к построению систем контроля версий. GitLab CE выделяется встроенной поддержкой кластеризации и интеграцией с облачными экосистемами, тогда как Gitea и Forgejo предлагают простоту и минимализм. Выбор между ними зависит от масштаба проекта, требований к отказоустойчивости и доступных ресурсов. Для эффективного развертывания кластера СКВ критически важно учитывать архитектурные особенности каждого решения, их совместимость с оркестраторами (Kubernetes, Docker Swarm) и возможности кастомизации под конкретные бизнес-задачи.

#### Литература

1. Чернявский Д.А., Кузнецов М.И. GitLab: практическое руководство по DevOps. М.: ДМК Пресс, 2021. 320 с.
2. Лорш В.Р., Брайан М. Системы контроля версий: от основ к кластерным решениям. СПб.: Питер, 2019. 256 с.
3. Сидоров А.Н., Петров К.Л. Архитектура распределенных систем управления кодом // Современные информационные технологии. 2022. № 4. С. 45-58.
4. GitLab Documentation: Geo Replication Электронный ресурс: <https://docs.gitlab.com/ee/administration/geo/>.
5. Gitea: Official Guide to Clustering Электронный ресурс: <https://docs.gitea.io/en-us/clustering/>.
6. Forgejo: Community-Driven Git Hosting Электронный ресурс: <https://forgejo.org/docs/>.

**SOKOLOV Ivan Andreevich**

Master's Student, Vologda State University, Russia, Vologda

*Scientific Advisor – Associate Professor of the Department of Automation and Computer Engineering of Vologda State University, Candidate of Technical Sciences Sorokin Arseniy Nikolaevich*

## **ARCHITECTURAL APPROACHES TO ENSURING FAULT TOLERANCE AND SCALABILITY IN OPEN SOURCE VERSION CONTROL SYSTEMS: A COMPARATIVE ANALYSIS OF GITLAB CE, GITEA AND FORGEJO**

**Abstract.** *The article provides a comparative analysis of architectural solutions of modern open-source version control systems (VCS) – GitLab CE, Gitea, and Forgejo – in the context of their ability to ensure fault tolerance, scalability, and integration with cluster environments. The component structure of each platform, methods for implementing horizontal scaling, support for geo-replication, and load balancing are considered. Particular attention is paid to practical use cases: from global corporate projects to startups and decentralized open-source communities. The results of the study demonstrate that the choice of the optimal VCS depends on the scale of the project, infrastructure requirements, and available resources. GitLab CE stands out for its built-in clustering and integration with DevOps tools, while Gitea and Forgejo offer lightweight solutions for small teams. Based on the analysis, recommendations are formulated for implementing systems in distributed environments, including compatibility with Kubernetes and cost minimization strategies.*

**Keywords:** *version control systems, GitLab CE, Gitea, Forgejo, cluster architectures, scalability, fault tolerance, geo-replication, horizontal scaling, DevOps, CI/CD, open-source.*

**ШВАЛЕВ Игорь Евгеньевич**

студент, Высшая школа компьютерных наук и искусственного интеллекта,  
Балтийский федеральный университет имени Иммануила Канта, Россия, г. Калининград

*Научный руководитель – доцент Высшей школы компьютерных наук и искусственного интеллекта Балтийского федерального университета имени Иммануила Канта,  
кандидат технических наук Зеленина Лариса Ивановна*

## **ГЕНЕРАЦИЯ ЖЕСТИКУЛЯЦИИ С ПРИМЕНЕНИЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА НА ОСНОВЕ ТЕКСТОВОГО ВВОДА**

**Аннотация.** Генерация естественной жестикуляции на основе текстового ввода остается актуальной задачей в сфере искусственного интеллекта, особенно для создания реалистичных виртуальных аватаров, анимации и робототехники. В данной статье представлен подход к автоматизации синтеза жестов с использованием нейросетевой модели, обученной на датасете из 7 часов видео, обработанных в CapCut. Для извлечения данных применены инструменты Whisper (транскрипция аудио) и MediaPipe (анализ движений тела), а предложенная модель объединяет текстовые эмбединги с кинематическими последовательностями.

**Ключевые слова:** искусственный интеллект, генерация жестикуляции, текст в жестикуляцию, анализ движений тела, MediaPipe, Whisper, траектории движений.

Современные технологии человеко-машинного взаимодействия всё чаще требуют создания систем, способных воспроизводить не только речь, но и невербальные элементы коммуникации, такие как жестикуляция. Жесты играют ключевую роль в передаче эмоций, акцентов и смысловых нюансов, что особенно важно для виртуальных ассистентов, анимационных персонажей или социальных роботов. Однако автоматическая генерация естественных и контекстно-зависимых движений тела на основе текста остается сложной задачей.

Процесс сбора данных включал несколько этапов, объединенных в последовательный пайплайн.

На первом этапе исходные видеозаписи длительностью 7 часов были нарезаны в CapCut на фрагменты, фокусируясь на сценах с выраженной жестикуляцией, синхронизированной с речью; это позволило исключить паузы,

фоновые движения и шумы, сохранив только релевантные для обучения данные.

Далее для каждого видеофрагмента с помощью Whisper выполнялась транскрипция аудио в текст с точными временными метками начала и конца каждой реплики, включая обработку пауз, интонаций и частичное распознавание эмоциональной окраски речи.

Параллельно через MediaPipe (модули Pose и Holistic) извлекались ключевые точки тела в формате 3D-координат: 8 точки скелета (только туловище и предплечье), 21 точка рук (суставы пальцев), что дало детальное представление о динамике движений.

Для синхронизации текстовых транскрипций и кинематических данных использовались временные метки из текстовых данных и данных MediaPipe.

Автоматизация обеспечивалась скриптом на Python, где цикл последовательно обрабатывал каждый видеофайл.

text	Type	Landmark_ID	X	Y	Z	Frame
Друзья, добрый вечер	Pose	0	X1	Y1	Z1	0
Друзья, добрый вечер	Left_Hand	0	X2	Y2	Z2	0
Друзья, добрый вечер	Right_Hand	0	X3	Y3	Z3	0
Друзья, добрый вечер	Pose	1	X4	Y4	Z4	1
Друзья, добрый вечер	Left_Hand	1	X5	Y5	Z5	1
Друзья, добрый вечер	Right_Hand	1	X6	Y6	Z6	1

Рис. 1. Структура полученного датасета

На рисунке 1 изображена структура полученного датасета, в котором объединены текстовые и жестовые данные. Где text – текст, извлеченный из видео. Type – один из типов позы Media Pipe. Landmark\_Id – id сустава, соответствующего своему типу позы. X, Y, Z – координаты точек суставов, Frame – номер кадра в анимации.

Архитектура нейронной сети для генерации жестикуляции имеет вид, представленный на рисунке 2.

Текст подается на вход в предобученную модель BERT, где он преобразуется в эмбединги и его размерность корректируется до 512D, для возможности его дальнейшей обработки с помощью механизма внимания.

С другой стороны, данные о жестах подаются в Frame Encoder, где они кодируются в скрытое пространство и их размерность так же приводится к 512D для тех же целей.

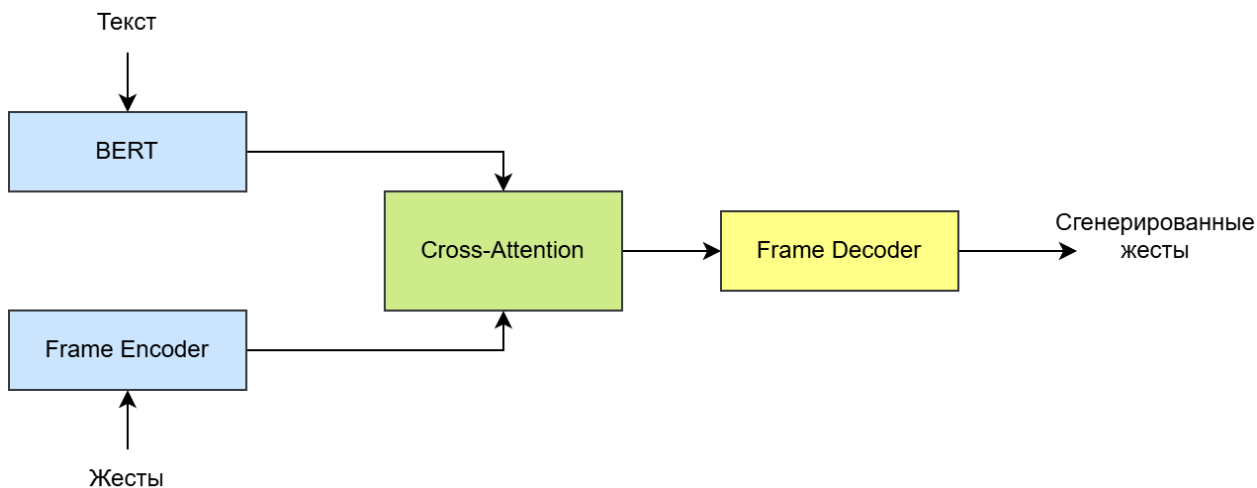


Рис. 2. Архитектура нейронной сети для генерации жестикуляции

Данные подаются в блок Cross-Attention, где между ними устанавливаются связи с помощью механизма внимания для того, чтобы модель могла лучше уловить и обучиться зависимости между текстовыми и жестовыми признаками.

После обработки данные попадают в Frame Decoder, где из них в обратном порядке получают данные о жестах.



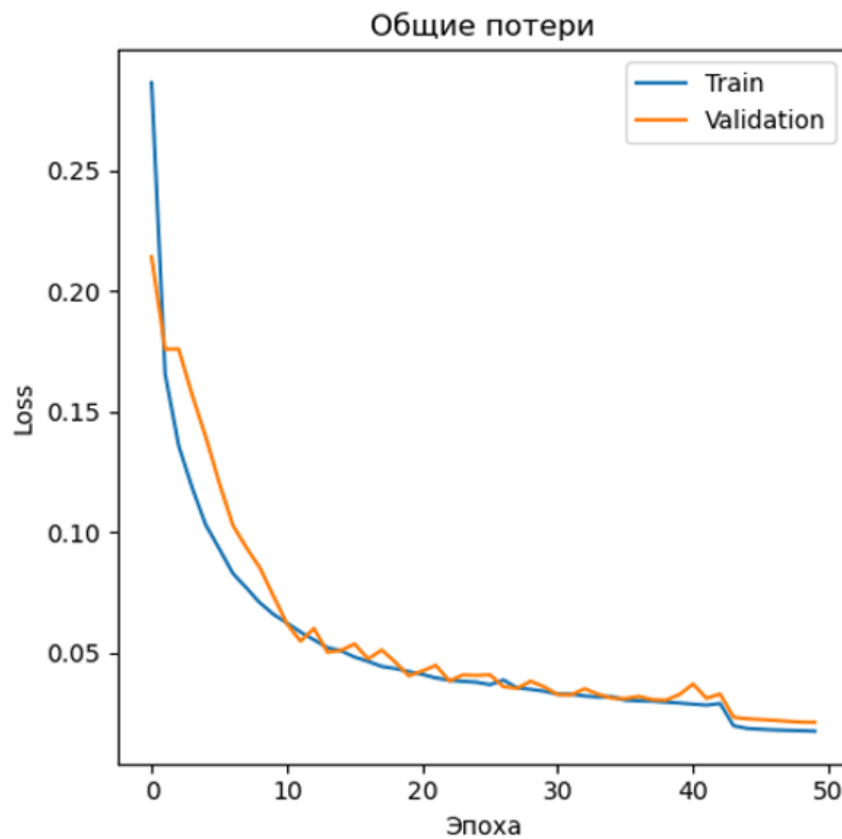


Рис. 3. Результаты обучения модели

На рисунке 3 представлены результаты обучения модели. Видно, что общие потери модели стабильно уменьшаются, что указывает на то, что модель обучается улавливать зависимости.

На рисунке 4 изображен график накопления ошибки MAE. Видно, что к 150 значение

ошибки практически стабильно. На 150 и 200 кадрах наблюдаются скачки, что говорит о том, что модель хуже справляется с длинными жестами. После 250 кадра наблюдается резкий рост ошибки, что значит, что модель не справляется с генерацией жестов такой длины.



Рис. 4. Накопление ошибки MAE

**Литература**

1. Vaswani A., et al. (2017). Attention Is All You Need. NeurIPS. arXiv:1706.03762
2. Devlin J., et al. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. NAACL. arXiv:1810.04805.
3. Goodfellow, I., et al. (2014). Generative Adversarial Networks. NeurIPS. arXiv:1406.2661.
4. Бенгфорт Б., Билбро Р., Охеда Т. Б46 Прикладной анализ текстовых данных на Python. Машинное обучение и создание приложений обработки естественного языка. – СПб.: Питер, 2019. – 368 с.: ил. – (Серия «Бестселлеры O'Reilly»).

**SHVALEV Igor Evgenievich**

Student, Graduate School of Computer Science and Artificial Intelligence,  
Immanuel Kant Baltic Federal University, Russia, Kaliningrad

*Scientific Advisor – Associate Professor of the Higher School of Computer Science  
and Artificial Intelligence of the Immanuel Kant Baltic Federal University,  
Candidate of Technical Sciences Zelenina Larisa Ivanovna*

**GENERATION OF GESTURES USING ARTIFICIAL INTELLIGENCE  
BASED ON TEXT INPUT**

**Abstract.** *Generating natural gestures based on text input remains an urgent task in the field of artificial intelligence, especially for creating realistic virtual avatars, animations, and robotics. This article presents an approach to automating gesture synthesis using a neural network model trained on a dataset of 7 hours of video processed in CapCut. To extract the data, the Whisper (audio transcription) and MediaPipe (body movement analysis) tools were used, and the proposed model combines text embeddings with mathematical sequences.*

**Keywords:** *artificial intelligence, generation of gestures, text in gestures, analysis of body movements, MediaPipe, Whisper, movement trajectories.*

# Актуальные исследования

Международный научный журнал

2025 • № 21 (256)

Часть I

ISSN 2713-1513

Подготовка оригинал-макета: Орлова М.Г.

Подготовка обложки: Ткачева Е.П.

*Учредитель и издатель:* ООО «Агентство перспективных научных исследований»

*Адрес редакции:* 308000, г. Белгород, пр-т Б. Хмельницкого, 135

*Email:* [info@apni.ru](mailto:info@apni.ru)

*Сайт:* <https://apni.ru/>

Отпечатано в ООО «ЭПИЦЕНТР».

Номер подписан в печать 02.06.2025г. Формат 60×90/8. Тираж 500 экз. Цена свободная.  
308010, г. Белгород, пр-т Б. Хмельницкого, 135, офис 40