

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ  
Федеральное государственное бюджетное образовательное учреждение  
высшего образования  
«КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»  
Факультет компьютерных технологий и прикладной математики

УТВЕРЖДАЮ:

Проректор по учебной работе,  
качеству образования – первый  
проректор

Хагуров Т.А.

*подпись*

« 29 » августа 2025 г.

## **РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)**

**Б1. О.26 Аналитика данных**

Направление подготовки 02.03.03 Математическое обеспечение и администрирование информационных систем

Профиль Искусственный интеллект и аналитика данных

Форма обучения очная

Квалификация бакалавр

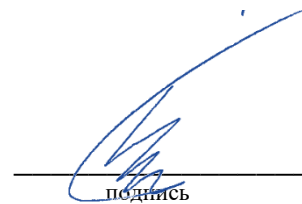
Краснодар 2025

Рабочая программа дисциплины Аналитика данных составлена в соответствии с федеральным государственным образовательным стандартом высшего образования (ФГОС ВО) по направлению подготовки 02.03.03 Математическое обеспечение и администрирование информационных систем

Программу составил(и):

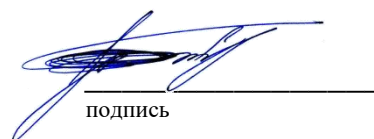
Т.М. Панеш, ст. преподаватель

И.О. Фамилия, должность, ученая степень, ученое звание



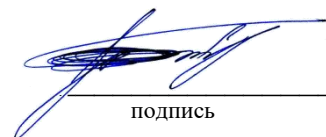
подпись

Рабочая программа дисциплины утверждена на заседании центра  
искусственного интеллекта  
протокол № 01 «28» августа 2025 г.  
Руководитель центра ИИ Коваленко А.В.



подпись

Утверждена на заседании учебно-методической комиссии факультета  
компьютерных технологий и прикладной математики  
протокол № 01 «28» августа 2025 г.  
Председатель УМК факультета Коваленко А.В.



подпись

Рецензенты:

Мостовой Евгений Викторович, генеральный директор ООО «Портал-Юг»,  
e-mail: mostovoy@portal-yug.ru

Луценко Евгений Вениаминович, доктор экономических наук, кандидат технических наук, профессор кафедры компьютерных технологий и систем Федерального государственного бюджетное образовательное учреждение высшего образования «Кубанский государственный аграрный университет имени И.Т. Трубилина», e-mail: prof.lutsenko@gmail.com

## **1 Цели и задачи изучения дисциплины (модуля)**

### **1.1 Цель освоения дисциплины**

Цели определены государственным образовательным стандартом высшего образования и соотнесены с общими целями ООП ВО по направлению подготовки «Прикладная информатика», в рамках которой преподается дисциплина.

**Цель:** Формирование у студентов системного представления о методах и инструментах аналитики данных, применяемых для исследования, интерпретации и визуализации данных, а также для поддержки принятия решений в различных предметных областях.

### **1.2 Задачи дисциплины**

- Ознакомление с современными методологиями аналитики данных;
- Применение инструментов Python для обработки, анализа и визуализации данных;
- Применение методов разведочного анализа данных (EDA) и построения признаков;
- Приобретение навыков построения и интерпретации моделей анализа данных;
- Автоматизация процессов аналитических вычислений и визуализации данных;
- Оформление результатов в виде технических отчетов и курсовой работы.

### **1.3 Место дисциплины (модуля) в структуре образовательной программы**

Дисциплина «Аналитика данных» относится к части Блока 1 «Дисциплины (модули)», часть, формируемая участниками образовательных отношений учебного плана.

Данная дисциплина тесно связана с дисциплинами: «Теория вероятностей и математическая статистика», «Машинное обучение», «Программирование на Python», «Основы анализа данных».

Материал курса является связкой между статистикой, алгоритмами машинного обучения и прикладными задачами, связанными с исследованием и интерпретацией данных в реальных предметных областях. Знания, полученные в данной дисциплине, используются в ходе изучения курсов «Интеллектуальные методы анализа данных», «Бизнес-аналитика и большие данные», «Системы поддержки принятия решений» и др.

### **1.4 Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения образовательной программы**

#### **Роль 1: Data Analyst (Аналитик данных)**

Задачи:

1. Статистический анализ, визуализация данных, предварительная обработка.
2. Создание прогнозных моделей
3. Построение аналитических моделей для поддержки бизнес-решений.

#### **Роль 2: Data Engineer (Специалист по подготовке и управлению данными)**

Задачи:

1. Сбор, очистка и предобработка данных из различных источников.
2. Организация хранения и управления данными.
3. Обеспечение качества и воспроизводимости аналитических процессов.

#### **Роль 3: AI PM (Менеджер проектов ИИ)**

Задачи:

1. Формулирование бизнес-требований к данным и постановка аналитических задач.
2. Проведение A/B-тестов и оценка гипотез.
3. Презентация результатов анализа и разработка рекомендаций для бизнеса.

Изучение данной учебной дисциплины направлено на формирование у обучающихся следующих компетенций:

Код и наименование индикатора*	Результаты обучения по дисциплине
ОПК-2 Способен применять современный математический аппарат, связанный с проектированием, разработкой, реализацией и оценкой качества программных продуктов и программных комплексов в различных областях человеческой деятельности	
ОПК-2.2 Применяет современный математический аппарат при построении моделей в различных областях человеческой деятельности	знать основные классы математических моделей и метрики их оценки; <b>уметь</b> формализовать прикладную задачу, выбирать и применять соответствующий математический аппарат с использованием современных библиотек; <b>владеть</b> навыками полного цикла построения, верификации и интерпретации моделей анализа данных в различных предметных областях для достижения практического результата
ВД-1 Способен осуществлять поиск сбор очистку и предварительный анализ данных	
ВД-1.3 Применяет методы понижения размерности для первичной интерпретации и визуализации многомерных данных.	Применяет методы снижения размерности (PCA, t-SNE и др.) для упрощения анализа и визуализации многомерных данных.
ВД-1.4 Отбирает признаки данных, значимые для исследования.	Отбирает и обосновывает выбор значимых признаков для последующего анализа и моделирования
ВД-2 Способен определять требования к наборам данных для решения задач машинного обучения проводить разметку и анализ наборов данных оценивать качество данных обеспечивать непрерывную интеграцию данных	
ВД-2.1 Определяет требования к наборам и качеству данных для решения задач машинного обучения.	Определяет критерии качества и требования к наборам данных для корректного применения методов машинного обучения
МЛ-2 Способен применять фундаментальные принципы и методы машинного обучения включая подготовку данных оценку качества моделей и работу с признаками	
МЛ-2.2 Применяет методы предварительной обработки данных и работы с признаками.	Применяет методы очистки, нормализации и кодирования признаков для подготовки данных к анализу.
МЛ-4 Способен применять методы обучения без учителя для анализа структуры данных и выявления скрытых закономерностей.	
МЛ-4.1 Применяет алгоритмы кластеризации и понижения размерности для решения практических задач.	Использует алгоритмы кластеризации и снижения размерности для группировки объектов и упрощения анализа данных.
МЛ-4.2 Выявляет аномалии и применяет методы поиска ассоциативных правил.	Выявляет аномалии в данных и применяет методы поиска ассоциативных правил для анализа взаимосвязей.
О-1 Способен осуществлять управление знаниями в том числе с применением алгоритмов интеллектуального поиска решений и формирования стратегий	
О-1.2 Способен преобразовывать неформализованные и слабо-формализованные данные предприятия в семантические единицы баз знаний.	Преобразует неструктурированные и слабо формализованные данные в семантические единицы, пригодные для хранения и анализа в базах знаний.

## 2. Структура и содержание дисциплины

### 2.1 Распределение трудоёмкости дисциплины по видам работ

Общая трудоёмкость дисциплины составляет 2 зач. ед. (72 час.), их распределение по видам работ представлено в таблице

Вид учебной работы	Всего часов	Семестры (часы)
		4
<b>Контактная работа, в том числе:</b>	<b>34,2</b>	<b>34,2</b>

<b>Аудиторные занятия (всего):</b>	<b>32</b>	<b>32</b>
Занятия лекционного типа	16	16
Лабораторные занятия	16	16
Занятия семинарского типа (семинары, практические занятия)		
<b>Иная контактная работа:</b>	<b>2,2</b>	<b>2,2</b>
Контроль самостоятельной работы (КСР)	2	2
Промежуточная аттестация (ИКР)	0,2	0,2
<b>Самостоятельная работа, в том числе:</b>	<b>37,8</b>	<b>37,8</b>
Проработка учебного (теоретического) материала	30	30
Выполнение индивидуальных заданий (типовой расчет)	5	5
Подготовка к текущему контролю	2,8	2,8
<b>Контроль:</b>	<b>35,7</b>	<b>35,7</b>
Подготовка к экзамену		
<b>Общая трудоемкость</b>	<b>час.</b>	<b>72</b>
	<b>в том числе контактная работа</b>	<b>34,2</b>
	<b>зач. ед</b>	<b>2</b>

## 2.2 Структура дисциплины

Распределение видов учебной работы и их трудоемкости по разделам дисциплины.  
Разделы (темы) дисциплины, изучаемые в 3 семестре

№	Наименование разделов (тем)	Количество часов				
		Всего	Аудиторная работа			Внеаудиторная работа СРС
			Л	ПЗ	ЛР	
1	2	3	4	5	6	7
1.	Основы аналитики данных и виды анализа (описательная, диагностическая, предсказательная, предписывающая).	8	2		2	4
2.	Сбор, очистка и подготовка данных.	8	2		2	4
3.	Разведочный анализ данных (EDA) и визуализация.	8	2		2	4
4.	Создание новых признаков, кодирование категориальных переменных, методы отбора информативных признаков.	8	2		2	4
5.	Статистический анализ данных и A/B-тестирование.	8	2		2	4
6.	Методы анализа, тренды, сезонность, прогнозирование.	8	2		2	4
7.	Кластеризация, выявление аномалий, ассоциативные правила.	10	2		2	6
8.	Создание дашбордов, storytelling с данными, подготовка аналитических отчетов.	11,8	2		2	7,8
<b>ИТОГО по разделам дисциплины</b>		<b>69,8</b>	<b>16</b>		<b>16</b>	<b>37,8</b>
Контроль самостоятельной работы (КСР)		2				
Промежуточная аттестация (ИКР)		0,2				

№	Наименование разделов (тем)	Количество часов				
		Всего	Аудиторная работа			Внеаудиторная работа
			Л	ПЗ	ЛР	СРС
1	2	3	4	5	6	7
Подготовка к текущему контролю						
<b>Общая трудоемкость по дисциплине</b>		<b>72</b>				

Примечание: Л – лекции, ПЗ – практические занятия/семинары, ЛР – лабораторные занятия, СРС – самостоятельная работа студента

## 2.3 Содержание разделов (тем) дисциплины

### 2.3.1 Занятия лекционного типа

№	Наименование раздела (темы)	Содержание раздела (темы)	Форма текущего контроля
1	2	3	4
1	Введение в аналитику данных	Основные понятия математических моделей. Классификация моделей (аналитические, численные, символьные). Обзор инструментов: MATLAB (Symbolic Math Toolbox), Maple, Wolfram Mathematica	Вопросы к зачету 1 - 5
2	Сбор и предобработка данных	Источники данных. Работа с пропусками, выбросами. Масштабирование и нормализация данных. Кодирование категориальных признаков.	Вопросы к зачету 6 - 10
3	Разведочный анализ данных (EDA) и визуализация	Методы описательной статистики. Визуализация данных: гистограммы, диаграммы рассеяния, box-plot, тепловые карты.	Вопросы к зачету 11 - 15
4	Фичеринжиниринг и отбор признаков	Создание новых признаков. Методы отбора (корреляция, дисперсионный анализ, регуляризация). Снижение размерности (PCA).	Вопросы к зачету 16 - 20
5	Статистический анализ и А/В-тесты	Проверка гипотез. Доверительные интервалы. Методы проведения А/В-тестов, интерпретация результатов.	Вопросы к зачету 21 - 25
6	Аналитика временных рядов	Тренды, сезонность, методы сглаживания. Модели прогнозирования временных рядов (ARIMA, Prophet).	Вопросы к зачету 26 - 30
7	Методы кластеризации и выявления закономерностей	Алгоритмы кластеризации (K-means, DBSCAN). Выявление аномалий. Поиск ассоциативных правил.	Вопросы к зачету 31 - 35
8	BI-инструменты и презентация результатов анализа	Создание дашбордов (Tableau, Power BI, Plotly/Dash). Подготовка аналитического отчета и storytelling с данными.	Вопросы к зачету 36 - 40

### 2.3.2 Занятия семинарского типа / лабораторные занятия

№	Наименование раздела (темы)	Наименование лабораторных работ	Форма текущего контроля
1	2	3	4
1	Работа с данными и EDA.	1. Знакомство с инструментами Python (Pandas, NumPy, Matplotlib, Seaborn). 2. Загрузка и первичная обработка наборов данных. 3. Работа с пропусками и выбросами. 4. Описательные статистики и базовые визуализации. 5. Разведочный анализ (EDA) на реальном датасете.	Проверка домашнего задания Контрольная работа
2	Фичеринжиниринг и статистический анализ.	6. Кодирование категориальных переменных. 7. Масштабирование и нормализация признаков. 8. Отбор признаков и методы снижения размерности (PCA). 9. Проверка статистических гипотез. 10. Проведение A/B-тестов.	Проверка домашнего задания Контрольная работа
3	Временные ряды и кластеризация	11. Работа с временными рядами: тренды, сезонность, скользящее среднее. 12. Построение прогнозных моделей (ARIMA, Prophet). 13. Кластеризация данных (K-means, DBSCAN). 14. Выявление аномалий и поиск ассоциативных правил.	Проверка домашнего задания Тестирование
4	BI-инструменты и итоговый проект	15. Создание интерактивных визуализаций (Plotly, Dash). 16. Построение простого дашборда (Power BI / Tableau / Dash). 17. Мини-проект: полный цикл анализа данных на реальном наборе.	

Примечание: ЛР – отчет/защита лабораторной работы, КП - выполнение курсового проекта, КР - курсовой работы, РГЗ - расчетно-графического задания, Р - написание реферата, Э - эссе, К - коллоквиум, Т – тестирование, РЗ – решение задач.

### 2.3.4 Примерная тематика курсовых работ (проектов)

Не предусмотрены

### 2.4 Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине (модулю)

№	Вид СРС	Перечень учебно-методического обеспечения дисциплины по выполнению самостоятельной работы
1	2	3
1	Изучение теоретического материала	Методические указания по организации самостоятельной работы студентов, утвержденные УСФ, протокол №1 от 30.06.2025

2	Решение задач	Методические указания по организации самостоятельной работы студентов, утвержденные УСФ, протокол №1 от 30.06.2025
---	---------------	--

Учебно-методические материалы для самостоятельной работы обучающихся из числа инвалидов и лиц с ограниченными возможностями здоровья (ОВЗ) предоставляются в формах, адаптированных к ограничениям их здоровья и восприятия информации:

Для лиц с нарушениями зрения:

- в печатной форме увеличенным шрифтом,
- в форме электронного документа,
- в форме аудиофайла,
- в печатной форме на языке Брайля.

Для лиц с нарушениями слуха:

- в печатной форме,
- в форме электронного документа.

Для лиц с нарушениями опорно-двигательного аппарата:

- в печатной форме,
- в форме электронного документа,
- в форме аудиофайла.

Данный перечень может быть конкретизирован в зависимости от контингента обучающихся.

### **3. Образовательные технологии, применяемые при освоении дисциплины (модуля)**

В соответствии с требованиями ФГОС программа дисциплины предусматривает использование в учебном процессе следующих образовательные технологии: чтение лекций с использованием мультимедийных технологий; лабораторные занятия.

С точки зрения применяемых методов используются как традиционные информационно-объяснительные лекции, так и интерактивная подача материала с мультимедийной системой. Компьютерные технологии в данном случае обеспечивают возможность разнопланового отображения алгоритмов и демонстрационного материала. Такое сочетание позволяет оптимально использовать отведенное время и раскрывать логику и содержание дисциплины.

Лабораторное занятие позволяет научить студента применять теоретические знания при решении и исследовании конкретных задач. Лабораторные занятия проводятся в компьютерных классах. Подход разбора конкретных ситуаций широко используется как преподавателем, так и студентами при проведении анализа результатов самостоятельной работы. Это обусловлено тем, что в процессе исследования часто встречаются задачи, для которых единых подходов не существует. Каждая конкретная задача при своем исследовании имеет множество подходов, а это требует разбора и оценки целой совокупности конкретных ситуаций.

– Для лиц с ограниченными возможностями здоровья предусмотрена организация консультаций с использованием электронной почты.

– Информационно-коммуникационные технологии (ИКТ) - расширяют рамки образовательного процесса, повышая его практическую направленность, способствуют интенсификации самостоятельной работы учащихся и повышению познавательной активности. В рамках ИКТ выделяются 2 вида технологий:

– Технология использования компьютерных программ – позволяет эффективно дополнить процесс обучения языку на всех уровнях.

– Интернет-технологии – предоставляют широкие возможности для поиска информации, разработки научных проектов, ведения научных исследований.

- проектная технология - индивидуальная или коллективная деятельность по отбору, распределению и систематизации материала по определенной теме, в результате которой составляется проект;
- анализ конкретных ситуаций - анализ реальных проблемных ситуаций, имевших место в соответствующей области профессиональной деятельности, и поиск вариантов лучших решений;
- развитие критического мышления – образовательная деятельность, направленная на развитие у студентов разумного, рефлексивного мышления, способного выдвинуть новые идеи и увидеть новые возможности.

Подход разбора конкретных задач и ситуаций широко используется как преподавателем, так и студентами во время лекций, лабораторных занятий и анализа результатов самостоятельной работы. Это обусловлено тем, что при решении каждой конкретной задачи имеется, как правило, несколько методов, а это требует разбора и оценки целой совокупности конкретных ситуаций.

#### **4. Оценочные и методические материалы**

##### **4.1 Оценочные средства для текущего контроля успеваемости и промежуточной аттестации**

Оценочные средства предназначены для контроля и оценки образовательных достижений обучающихся, освоивших программу учебной дисциплины «Аналитика данных».

Оценочные средства включает контрольные материалы для проведения **текущего контроля** в форме тестовых заданий, разноуровневых заданий, типовых расчетов и **промежуточной аттестации** в форме вопросов и заданий к экзамену.

Оценочные средства для инвалидов и лиц с ограниченными возможностями здоровья выбираются с учетом их индивидуальных психофизических особенностей.

– при необходимости инвалидам и лицам с ограниченными возможностями здоровья предоставляется дополнительное время для подготовки ответа на экзамене;

– при проведении процедуры оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья предусматривается использование технических средств, необходимых им в связи с их индивидуальными особенностями;

– при необходимости для обучающихся с ограниченными возможностями здоровья и инвалидов процедура оценивания результатов обучения по дисциплине может проводиться в несколько этапов.

Процедура оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья по дисциплине (модулю) предусматривает предоставление информации в формах, адаптированных к ограничениям их здоровья и восприятия информации:

Для лиц с нарушениями зрения:

- в печатной форме увеличенным шрифтом,
- в форме электронного документа.

Для лиц с нарушениями слуха:

- в печатной форме,
- в форме электронного документа.

Для лиц с нарушениями опорно-двигательного аппарата:

- в печатной форме,
- в форме электронного документа.

Данный перечень может быть конкретизирован в зависимости от контингента обучающихся.

##### **Структура оценочных средств для текущей и промежуточной аттестации**

№ п/п	Контролируемые разделы (темы) дисциплины	Код контролируемой компетенции	Наименование оценочного средства	Текущий контроль
1	Разведочный анализ данных (EDA)	BD-1.3, ОПК-2.2	ЛР №1	Проверка отчетов, вопросы к зачету 1–5
2	Отбор и создание признаков (фичеринжиниринг)	BD-1.4	ЛР №2	Проверка отчетов, вопросы к зачету 6–10
3	Оценка качества данных и требований к выборкам	BD-2.1 ОПК-2.2	ЛР №3	Проверка отчетов, вопросы к зачету 11–15
4	Предобработка данных (нормализация, кодирование)	ML-2.2, ОПК-2.2	ЛР №4	Проверка отчетов, вопросы к зачету 16–20
5	Кластеризация и снижение размерности	ML-4.1	ЛР №5	Проверка отчетов, вопросы к зачету 21–25
6	Выявление аномалий и ассоциативные правила	ML-4.2	ЛР №6	Проверка отчетов, вопросы к зачету 26–30
7	Преобразование неструктурированных данных в базы знаний	О-1.2	ЛР №7	Проверка отчетов, вопросы к зачету 31–35
8	Итоговый мини-проект по аналитике данных	BD-1.3, BD-1.4, ML-4.1	КП	Защита проекта, вопросы к зачету 36–40

### Показатели, критерии и шкала оценки сформированных компетенций

№ п/п	Код и наименование индикатора	Результаты обучения	Текущий контроль	
1	BD-1.3 Применяет методы снижения размерности для первичной интерпретации и визуализации многомерных данных.	Применяет базовые методы снижения размерности для анализа и визуализации данных.	ЛР, КП	Вопросы к зачету 1–5
2	BD-1.4 Отбирает признаки данных, значимые для исследования.	Выполняет отбор простых признаков для улучшения интерпретируемости моделей.	ЛР, КП	Вопросы к зачету 6–10

3	BD-2.1 Определяет требования к наборам и качеству данных для решения задач машинного обучения.	Определяет основные требования к данным и их качеству.	ЛР, КП	Вопросы к зачету 11–15
4	ML-2.2 Применяет методы предварительной обработки данных и работы с признаками.	Выполняет очистку и нормализацию данных.	ЛР, КП	Вопросы к зачету 16–20
5	ML-4.1 Применяет алгоритмы кластеризации и понижения размерности для решения практических задач.	Применяет простейшие алгоритмы кластеризации и снижения размерности.	ЛР, КП	Вопросы к зачету 21–25
6	ML-4.2 Выявляет аномалии и применяет методы поиска ассоциативных правил.	Находит аномалии и строит базовые ассоциативные правила.	ЛР, КП	Вопросы к зачету 26–30
7	О-1.2 Способен преобразовывать неформализованные и слабоформализованные данные предприятия в семантические единицы баз знаний.	Преобразует простые неструктурированные данные в таблицы и базы.	ЛР, КП	Вопросы к зачету 31–35
8	BD-1.3 Применяет методы снижения размерности для первичной интерпретации и визуализации многомерных данных.	Сравнивает методы снижения размерности и обосновывает их выбор.	ЛР, КП	Вопросы к зачету 1–5
9	BD-1.4 Отбирает признаки данных, значимые для исследования.	Анализирует значимость признаков и формирует наборы данных.	ЛР, КП	Вопросы к зачету 6–10
10	BD-2.1 Определяет требования к наборам и качеству данных для решения задач машинного обучения.	Проводит оценку качества данных и формулирует требования к выборкам.	ЛР, КП	Вопросы к зачету 11–15
11	ML-2.2 Применяет методы предварительной обработки данных и работы с признаками.	Применяет кодирование и масштабирование признаков.	ЛР, КП	Вопросы к зачету 16–20
12	ML-4.1 Применяет алгоритмы кластеризации и понижения размерности для решения практических задач.	Реализует алгоритмы кластеризации и интерпретирует сегментацию.	ЛР, КП	Вопросы к зачету 21–25
13	ML-4.2 Выявляет аномалии и применяет методы поиска ассоциативных правил.	Применяет методы выявления аномалий и закономерностей.	ЛР, КП	Вопросы к зачету 26–30
14	О-1.2 Способен преобразовывать неформализованные и слабоформализованные данные предприятия в семантические единицы баз знаний.	Преобразует неструктурированные данные предприятия в семантические единицы.	ЛР, КП	Вопросы к зачету 31–35

15	BD-1.3 Применяет методы снижения размерности для первичной интерпретации и визуализации многомерных данных.	Разрабатывает аналитические пайплайны с применением сложных методов снижения размерности.	ЛР, КР	Вопросы к зачету 1–5
16	BD-1.4 Отбирает признаки данных, значимые для исследования.	Разрабатывает методологии отбора признаков для повышения эффективности моделей.	ЛР, КР	Вопросы к зачету 6–10
17	BD-2.1 Определяет требования к наборам и качеству данных для решения задач машинного обучения.	Проектирует схемы сбора и хранения данных, формулирует комплексные требования к качеству.	ЛР, КР	Вопросы к зачету 11–15
18	ML-2.2 Применяет методы предварительной обработки данных и работы с признаками.	Проектирует сложные процессы предобработки данных и автоматизирует работу с признаками.	ЛР, КР	Вопросы к зачету 16–20
19	ML-4.1 Применяет алгоритмы кластеризации и понижения размерности для решения практических задач.	Применяет современные методы кластеризации и факторного анализа.	ЛР, КР	Вопросы к зачету 21–25
20	ML-4.2 Выявляет аномалии и применяет методы поиска ассоциативных правил.	Разрабатывает методы выявления сложных аномалий и закономерностей.	ЛР, КР	Вопросы к зачету 26–30
21	O-1.2 Способен преобразовывать неформализованные и слабо-формализованные данные предприятия в семантические единицы баз знаний.	Создает модели преобразования слабо формализованных данных в семантические структуры знаний.	ЛР, КР	Вопросы к зачету 31–35

**Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций в процессе освоения образовательной программы**

Лабораторная работа №1: Разведочный анализ данных (EDA)

Кейс: «Исследование факторов успеваемости студентов»

Задача: Провести анализ набора данных с результатами студентов (оценки, посещаемость, активность, демографические параметры).

Что нужно сделать:

Загрузить датасет (например, StudentsPerformance.csv или аналогичный);

Выполнить анализ пропусков, выбросов и распределений признаков;

Построить визуализации (гистограммы, коробчатые диаграммы, корреляционные матрицы);

Сформулировать предварительные выводы о взаимосвязях признаков.

Инструменты: Python (pandas, matplotlib, seaborn), Jupyter Notebook, Excel, Power BI.

Лабораторная работа №2: Предобработка и очистка данных

Кейс: «Качество данных в финансовой аналитике»

Задача: Подготовить набор данных по транзакциям клиентов для последующего анализа рисков.

Что нужно сделать:

Обнаружить и обработать пропущенные значения, выбросы, дубликаты;  
 Провести кодирование категориальных признаков (Label, One-Hot);  
 Масштабировать числовые данные;  
 Сравнить влияние разных стратегий очистки на итоговые статистики.  
 Инструменты: Python (pandas, numpy, scikit-learn), Excel, Jupyter Notebook.

Лабораторная работа №3: Отбор признаков и понижение размерности  
 Кейс: «Оптимизация набора признаков для модели предсказания цен недвижимости»  
 Задача: Определить наиболее значимые признаки, влияющие на стоимость объектов недвижимости.

Что нужно сделать:

Провести корреляционный анализ и визуализацию взаимосвязей признаков;  
 Применить методы отбора признаков (SelectKBest, LASSO);  
 Использовать PCA для понижения размерности;  
 Сравнить качество моделей до и после отбора признаков.  
 Инструменты: Python (scikit-learn, pandas, seaborn), Jupyter Notebook.

### В результате студент получит:

- Навыки работы с современными инструментами анализа данных (Python, библиотеки Pandas, NumPy, scikit-learn, а также BI-системы).
- Понимание, как применять методы статистики, визуализации и машинного обучения для анализа и интерпретации данных.
- Умение проводить предобработку, отбор признаков, снижение размерности и построение аналитических моделей.
- Опыт разработки аналитических пайплайнов и применения их к практическим задачам.
- Базу для дальнейшего углублённого изучения методов машинного обучения, MLOps и продвинутой аналитики.

### Зачетно-экзаменационные материалы для промежуточной аттестации

#### Примерный перечень вопросов для зачета

Вопрос	Компетенции
1. Объясните роль разведочного анализа данных (EDA). Какие методы визуализации и статистического анализа вы знаете?	BD-1.3, ОПК-2.2
2. Какие методы снижения размерности данных применяются для визуализации и анализа?	BD-1.3 ОПК-2.2
3. Как интерпретировать результаты PCA?	BD-1.3
4. В чём преимущества t-SNE и UMAP для визуализации данных?	BD-1.3 ОПК-2.2
5. Какие ошибки можно допустить при применении методов снижения размерности?	BD-1.3
6. Что такое отбор признаков? Какие задачи он решает?	BD-1.4 ОПК-2.2
7. Какие методы отбора признаков вы знаете (Filter, Wrapper, Embedded)?	BD-1.4

8. Как оценить значимость признаков?	BD-1.4
9. Как фичеринжиниринг влияет на качество модели?	BD-1.4
10. Приведите пример создания новых признаков на основе существующих.	BD-1.4
11. Какие требования предъявляются к данным для машинного обучения?	BD-2.1, ОПК-2.2
12. Как определить наличие пропусков и выбросов в данных?	BD-2.1
13. Какие метрики используются для оценки качества данных?	BD-2.1
14. Как справляться с несбалансированными данными?	BD-2.1
15. Почему важна репрезентативность выборки?	BD-2.1
16. Какие методы нормализации и стандартизации данных вы знаете?	ML-2.2
17. В чём разница между One-Hot Encoding и Label Encoding?	ML-2.2
18. Какие методы применяются для обработки пропусков в данных?	ML-2.2
19. Почему важна предобработка данных перед обучением модели?	ML-2.2 ОПК-2.2
20. Приведите пример пайплайна предобработки данных в scikit-learn.	ML-2.2
21. Какие алгоритмы кластеризации вы знаете (K-means, DBSCAN, и др.)?	ML-4.1
22. Как выбрать оптимальное количество кластеров в K-means?	ML-4.1
23. Объясните принцип работы и применения иерархической кластеризации.	ML-4.1
24. Что такое метод локтя и силуэтный коэффициент?	ML-4.1
25. Как сочетать методы снижения размерности и кластеризации?	ML-4.1
26. Что такое аномалии в данных?	ML-4.2
27. Какие методы обнаружения аномалий вы знаете?	ML-4.2
28. Как применяются ассоциативные правила в анализе транзакций?	ML-4.2
29. Объясните принцип работы алгоритма Apriori.	ML-4.2
30. Какие задачи решаются с помощью поиска ассоциативных правил?	ML-4.2
31. Что такое неструктурированные данные?	O-1.2
32. Какие подходы применяются для преобразования текстов в численные представления?	O-1.2
33. Объясните разницу между Bag-of-Words и TF-IDF.	O-1.2
34. Что такое семантические единицы баз знаний?	O-1.2

35. Какие инструменты применяются для работы с базами знаний?	O-1.2
36. Разберите пример полного аналитического пайплайна: от сбора данных до визуализации результатов.	BD-1.3, BD-1.4, ML-2.2
37. Какие этапы включает процесс подготовки данных для машинного обучения?	BD-2.1, ML-2.2
38. Какие ошибки чаще всего совершают начинающие аналитики данных?	BD-1.3, BD-2.1
39. Как связаны разведочный анализ и последующее моделирование?	BD-1.3, BD-1.4
40. Какую роль играет визуализация данных в аналитике?	BD-1.3

#### 4.2 Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций

##### *Методические рекомендации, определяющие процедуры оценивания ответа на зачете:*

Шкала оценивания зачета

«Зачтено»: полный, структурированный ответ с примерами, умение связать теорию с практикой (например, «SVD применяется в PCA для уменьшения размерности данных»), даны не только определения, но и примеры применения, для практических вопросов разобраны код из лабораторных, может показать связь тем.

«Не зачтено»: Отсутствие ключевых понятий или грубые ошибки; неспособность объяснить применение метода в ИИ; не может разобрать коды из лабораторных, проблемы найти связь в темах.

##### *Методические рекомендации, определяющие процедуры оценивания лабораторных работ:*

Оценка зависит от:

- **Качества реализации** (работоспособность кода, эффективность)
- **Анализа результатов** (интерпретация, сравнение методов)
- **Оформления и отчетности** (четкость, соответствие стандартам)

Уровень	Требования
<b>Теоретическая подготовка (20%)</b>	
<b>Пороговый (3/5)</b>	Приведены базовые формулы, но без глубокого объяснения.
<b>Базовый (4/5)</b>	Дано обоснование выбранных методов, ссылки на источники.
<b>Продвинутый (5/5)</b>	Сравнение альтернативных подходов, критика ограничений.
<b>Практическая реализация (40%)</b>	

<b>Пороговый (3/5)</b>	Код работает, но без обработки ошибок и оптимизации.
<b>Базовый (4/5)</b>	Чистый код с комментариями, тестами и визуализацией.
<b>Продвинутый (5/5)</b>	Использование векторных операций, сравнение нескольких методов.
<b>Анализ результатов (30%)</b>	
<b>Пороговый (3/5)</b>	Простые графики без интерпретации.
<b>Базовый (4/5)</b>	Сравнение с теоретическими ожиданиями (например, сходимость градиентного спуска).
<b>Продвинутый (5/5)</b>	Анализ влияния гиперпараметров, статистические выводы.
<b>Оформление отчета (10%)</b>	
<b>Пороговый (3/5)</b>	Есть введение, код и выводы.
<b>Базовый (4/5)</b>	Структурированный отчет с графиками и таблицами.
<b>Продвинутый (5/5)</b>	LaTeX-документ или Jupyter Notebook с интерактивными визуализациями.

*Методические рекомендации, определяющие процедуры оценивания курсовых работ (проектов):*

#### **Общие критерии курсовых работ (проектов)**

<b>Уровень</b>	<b>Требования</b>
<b>Теоретическая часть (математическое обоснование)</b>	
<b>Пороговый</b>	Приведены базовые формулы и определения без глубокого анализа
<b>Базовый</b>	Дано развернутое объяснение методов, их плюсы/минусы, ссылки на источники.
<b>Продвинутый</b>	Собственный анализ методов, критика существующих подходов, предложения улучшений.
<b>Практическая реализация</b>	
<b>Пороговый</b>	Работающий код на Python/MATLAB с минимальными комментариями.
<b>Базовый</b>	Чистый, модульный код с пояснениями, тестами и визуализацией.

<b>Продвинутый</b>	Оптимизированный код, сравнение нескольких методов, обработка edge-cases.
<b>Эксперименты и анализ результатов</b>	
<b>Пороговый</b>	Простые тесты на синтетических данных без глубокого анализа.
<b>Базовый</b>	Сравнение с baseline-методами, интерпретация метрик (MSE, Accuracy).
<b>Продвинутый</b>	Анализ устойчивости, влияние гиперпараметров, статистическая значимость.
<b>Оформление и структура</b>	
<b>Пороговый</b>	Наличие введения, основной части и выводов.
<b>Базовый</b>	Четкая структура (оглавление, список литературы, графики).
<b>Продвинутый</b>	Публикация кода на GitHub, оформление в LaTeX, видео-презентация.

### Шкала оценивания

Балл	Уровень выполнения	Соответствие критериям
5 (Отлично)	Продвинутый	Все критерии выполнены на высоком уровне + инновации.
4 (Хорошо)	Базовый	Полное соответствие без серьезных недочетов.
3 (Удовл.)	Пороговый	Минимальные требования выполнены, но есть недочеты.
2 (Неуд.)	Ниже порога	Критические ошибки в теории/практике.

### Дополнительные критерии по тематикам

#### Математические основы ИИ (MF-1.1, MF-1.2)

- **Пороговый:** Решение типовых задач в MATLAB/Maple без модификации методов.
- **Базовый:** Сравнение аналитического и численного решений.
- **Продвинутый:** Доказательство сходимости/устойчивости метода.

#### Оптимизация (MF-3.1)

- **Пороговый:** Реализация градиентного спуска для линейной регрессии.
- **Базовый:** Сравнение SGD, Adam, RMSprop на реальных данных.
- **Продвинутый:** Модификация алгоритма (например, адаптивный learning rate).

#### Теория глубокого обучения (FC-1)

- **Пороговый:** Обучение готовой модели на датасете (MNIST/CIFAR).
- **Базовый:** Анализ влияния архитектуры на точность.
- **Продвинутый:** Предложение новой архитектуры/метода обучения.

## 4.3 Методические указания по организации лабораторных работ по дисциплине "Аналитика данных"

---

### 1. Общие сведения

**Образовательная программа:** «Искусственный интеллект и аналитика данных»

**Дисциплина:** «Аналитика данных»

**Вид обеспечения:** Проведение лабораторных работ

**Условия применения:**

Для успешного выполнения лабораторных работ необходимо обеспечить:

### 1. Программное обеспечение:

Python (с библиотеками: NumPy, pandas, SciPy, scikit-learn, matplotlib, seaborn, statsmodels)

Jupyter Notebook / Google Colab (для выполнения и документирования анализа данных)

Power BI или Tableau (для визуализации и дашбордов)

Excel (для базовой обработки и представления данных)

### 2. Аппаратное обеспечение:

Компьютеры с ОС Windows/Linux, поддерживающие Python и инструменты визуализации

Достаточные вычислительные ресурсы (CPU, рекомендуется наличие GPU при работе с большими наборами данных)

### 3. Облачная инфраструктура (опционально):

Google Colab Pro / Kaggle Notebooks для ускоренных вычислений

Облачные хранилища (Google Drive, Yandex Disk) для совместной работы над проектами

## 2. Цели, задачи и ожидаемые результаты выполнения лабораторных работ

### 2.1. Обоснование необходимости лабораторных работ

Лабораторные работы в данной дисциплине необходимы, поскольку:

а) **Практическая направленность:** Аналитика данных требует не только знания теории статистики и машинного обучения, но и навыков обработки, визуализации и интерпретации реальных наборов данных с использованием современных инструментов (Python, Power BI, Tableau).

б) **Формирование аналитического мышления:** Студенты учатся применять методы статистического анализа, кластеризации, регрессии и визуализации данных для принятия обоснованных решений.

в) **Подготовка к реальным проектам:** Работа с большими и неструктурированными данными, очистка и построение аналитических моделей развивают навыки, востребованные в индустрии Data Science и бизнес-аналитике.

### 2.2. Задачи лабораторных работ

#### 1. Освоение инструментов аналитики:

- Работа с Python (pandas, numpy, scikit-learn, matplotlib, seaborn).
- Использование визуализационных инструментов Power BI и Tableau.

## 2. Применение статистических и аналитических методов:

- Выполнение разведочного анализа данных (EDA).
- Применение методов предобработки, очистки и отбора признаков.
- Применение алгоритмов машинного обучения для классификации, регрессии и кластеризации.

## 3. Интерпретация и представление результатов:

- Формирование аналитических отчётов и визуализаций.
- Разработка дашбордов и презентаций для обоснования выводов.

### 2.3. Ожидаемые результаты

После выполнения лабораторных работ студенты смогут:

- Выполнять предобработку и очистку данных с применением современных инструментов.
- Применять методы разведочного анализа данных для выявления закономерностей и аномалий.
- Использовать алгоритмы машинного обучения для анализа и прогнозирования.
- Применять методы отбора признаков и снижения размерности (PCA, LASSO, корреляционный анализ).
- Создавать визуализации и аналитические дашборды в Power BI или Tableau.
- Представлять результаты анализа в структурированной и наглядной форме

### 2.4. Что необходимо для реализации лабораторных работ?

#### 1. Методические материалы:

- Пошаговые инструкции по каждой теме.
- Примеры кода и шаблоны ноутбуков (Jupyter/Colab).

#### 2. Техническая поддержка:

- Настройка и установка необходимых библиотек Python.
- Доступ к Power BI/Tableau или аналогичным инструментам визуализации.

#### 3. Контроль выполнения:

- Отчёты по каждой работе в формате Jupyter Notebook или PDF.
- Проверка качества визуализации и интерпретации данных.

## 3. Порядок реализации лабораторных работ

### 3.1. Задача №1: Освоение инструментов

**Цель:** Научиться работать с основными инструментами анализа данных, выполнять загрузку, предварительную обработку и визуализацию наборов данных.

**Тема 1: Введение в аналитические инструменты (Python / pandas / matplotlib / seaborn)**

#### Пошаговая инструкция:

##### 1. Знакомство с интерфейсом:

- Запуск Jupyter Notebook или Google Colab.

- Создание нового Python-ноутбука и установка библиотек (pip install pandas matplotlib seaborn).

## 2. Загрузка и просмотр данных:

```
python

import pandas as pd
data = pd.read_csv("students_performance.csv")
data.head()
data.info()
```

## 3. Очистка данных:

```
python

data = data.dropna()
data = data[data["math score"] > 0]
```

## 4. Визуализация признаков:

```
python

import seaborn as sns
import matplotlib.pyplot as plt
sns.histplot(data["math score"], bins=20)
plt.show()
```

### Задания:

1. Загрузите любой открытый набор данных (например, Titanic, Iris, StudentsPerformance).
2. Определите наличие пропусков, выбросов, дубликатов.
3. Постройте три различных визуализации: гистограмму, диаграмму рассеяния и boxplot.
4. Сформулируйте выводы о структуре данных и возможных проблемах качества.

---

## Задача №2: Корреляционный анализ, регрессия и оптимизация моделей

**Цель:** Освоить основные методы статистического анализа и оптимизации при работе с данными.

### Пошаговая инструкция:

#### 1. Корреляционный анализ:

```
python

corr = data.corr()
sns.heatmap(corr, annot=True, cmap='coolwarm')
plt.show()
```

#### 2. Построение модели линейной регрессии:

```
python

from sklearn.linear_model import LinearRegression
X = data[["reading score", "writing score"]]
y = data["math score"]
model = LinearRegression().fit(X, y)
print(model.coef_, model.intercept_)
```

### 3. Оптимизация модели (GridSearchCV):

```
python

from sklearn.model_selection import GridSearchCV
from sklearn.tree import DecisionTreeRegressor

params = {'max_depth': [2, 4, 6, 8]}
grid = GridSearchCV(DecisionTreeRegressor(), params, cv=3)
grid.fit(X, y)
print(grid.best_params_)
```

#### Задания:

1. Провести корреляционный анализ и выделить 3 наиболее связанных признака.
2. Построить модель линейной регрессии для прогноза целевой переменной.
3. Провести подбор гиперпараметров (GridSearchCV) для дерева решений.
4. Оценить качество модели ( $R^2$ , MAE, RMSE) и сделать выводы

#### Итоговая структура каждой лабораторной работы:

1. **Теоретическая часть** (краткое описание методов).
2. **Пошаговые инструкции** (код + пояснения).
3. **Индивидуальные задания** (3–5 задач на самостоятельное выполнение).
4. **Контрольные вопросы**

Такой подход обеспечит **постепенное усложнение задач и интеграцию знаний** из разных тем.

#### Соответствие лабораторных работ и индикаторов компетенций

В таблице ниже представлено, как каждая лабораторная работа способствует формированию заявленных компетенций.

Компетенция (код и наименование)	Соответствующие лабораторные работы	Обоснование
BD-1.3 Применяет методы снижения размерности для первичной интерпретации и визуализации многомерных данных	ЛР: Разведочный анализ данных (EDA); ЛР: PCA/UMAP + визуализация кластеров	<ul style="list-style-type: none"> <li>• Снижение размерности (PCA/UMAP) позволяет визуально интерпретировать многомерные данные.</li> <li>• EDA выявляет закономерности и выбросы до построения моделей.</li> </ul>
BD-1.4 Отбирает признаки данных, значимые для исследования	ЛР: Фичеринжиниринг и отбор признаков (Filter/Wrapper/Embedded)	<ul style="list-style-type: none"> <li>• Осознанный выбор признаков повышает качество и интерпретируемость</li> </ul>

		<p>моделей.</p> <ul style="list-style-type: none"> <li>• Регуляризация и важность признаков помогают избавиться от шума.</li> </ul>
<p>BD-2.1 Определяет требования к наборам и качеству данных для решения задач машинного обучения</p>	<p>ЛР: Оценка качества данных (пропуски, выбросы, дисбаланс); ЛР: Формирование выборок и валидация</p>	<ul style="list-style-type: none"> <li>• Полнота, корректность и репрезентативность определяют надёжность выводов.</li> <li>• Критерии качества данных задают границы применимости моделей.</li> </ul>
<p>ML-2.2 Применяет методы предварительной обработки данных и работы с признаками</p>	<p>ЛР: Нормализация/стандартизация; кодирование категорий; пайплайны scikit-learn</p>	<ul style="list-style-type: none"> <li>• Единый пайплайн предобработки обеспечивает воспроизводимость.</li> <li>• Масштабирование и кодирование устраняют артефакты в обучении.</li> </ul>
<p>ML-4.1 Применяет алгоритмы кластеризации и понижения размерности для решения практических задач</p>	<p>ЛР: K-means/DBSCAN; ЛР: PCA + кластеризация; оценка качества кластеров</p>	<ul style="list-style-type: none"> <li>• Ненадзорные методы выявляют сегменты и скрытую структуру.</li> <li>• Совмещение PCA и кластеризации улучшает разделимость классов.</li> </ul>
<p>ML-4.2 Выявляет аномалии и применяет методы поиска ассоциативных правил</p>	<p>ЛР: Обнаружение аномалий (IQR, Isolation Forest); ЛР: Ассоциативные правила (Apriori)</p>	<ul style="list-style-type: none"> <li>• Детекция аномалий важна для безопасности и качества данных.</li> <li>• Ассоциативные правила раскрывают закономерности поведения.</li> </ul>
<p>O-1.2 Способен преобразовывать неформализованные и слабо-формализованные данные предприятия в семантические единицы баз знаний</p>	<p>ЛР: Извлечение сущностей из текстов; построение простых онтологий/семантических схем</p>	<ul style="list-style-type: none"> <li>• Преобразование текстов и логов в структурированное представление создаёт основу для аналитики.</li> <li>• Семантические единицы повышают совместимость и повторное использование данных.</li> </ul>

## 5. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины (модуля)

### 5.1 Литература

1. Бизнес-статистика : учебник и практикум для вузов / под ред. И. И. Елисеевой. — 2-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2022. — 444 с. — (Высшее образование). — ISBN 978-5-534-14822-0. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/490172> (дата обращения: 01.11.2025).

2. Маккини, У. Python и анализ данных : первичная обработка данных с применением pandas, NumPy и Jupyter / У. Маккини ; пер. с англ. 2-е изд. — Москва : ДМК Пресс, 2020. — 539 с. — ISBN 978-5-97060-590-5. — Текст : электронный. — URL: <https://dokumen.pub/python-2nbsped-9785940745905-9781491957660.html> (дата обращения: 01.11.2025).
3. Подкорытова, О. А., Соколов, М. В. Анализ временных рядов : учебное пособие для вузов. — 2-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2022. — 267 с. — (Высшее образование). — Текст : электронный. — URL: (дата обращения: 01.11.2025).
4. Колоколов, А. Азбука визуализации Power BI : практическое руководство по построению дашбордов и аналитических отчётов в Power BI. — 2024. — Текст : электронный // Литрес [сайт]. — URL: <https://www.litres.ru/book/aleksey-kolokolov-33195442/azbuka-vizualizacii-power-bi-71183659/chitat-onlayn/> (дата обращения: 01.11.2025).
5. Платонов, А. В. Машинное обучение : учебное пособие для вузов / А. В. Платонов. — 2-е изд. — Москва : Издательство Юрайт, 2025. — 89 с. — (Высшее образование). — ISBN 978-5-534-20732-3. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/558662> (дата обращения: 01.11.2025).
6. Sun, X., Li, J., Kovalenko, A.V., Feng, W., Ou, Y. Integrating Reinforcement Learning and Learning From Demonstrations to Learn Nonprehensile Manipulation //IEEE Transactions on Automation Science and Engineering, 2023, 20(3), 1735–1744, DOI: 10.1109/TASE.2022.3185071, Q1
7. Petukhova, A.V.; Kovalenko, A.V.; Ovsyannikova, A.V. Algorithm for Optimization of Inverse Problem Modeling in Fuzzy Cognitive Maps. Mathematics 2022, 10, 3452. DOI: 10.3390/math10193452, Q1
8. Kirillova, E.; Kovalenko, A.; Urtenov, M. Study of the Current–Voltage Characteristics of Membrane Systems Using Neural Networks. AppliedMath 2025, 5, 10. <https://doi.org/10.3390/appliedmath5010010>
9. Kadurin, Artur, et al. "The cornucopia of meaningful leads: Applying deep adversarial autoencoders for new molecule development in oncology." Oncotarget 8.7 (2016): 10883.
10. Kadurin, Artur, et al. "druGAN: an advanced generative adversarial autoencoder model for de novo generation of new molecules with desired molecular properties in silico." Molecular pharmaceutics 14.9 (2017): 3098-3104.
11. Polykovskiy, Daniil, et al. "Molecular sets (MOSES): a benchmarking platform for molecular generation models." Frontiers in pharmacology 11 (2020): 565644.
12. Khrabrov, Kuzma, et al. "\$\nabla^2\$ DFT: A Universal Quantum Chemistry Dataset of Drug-Like Molecules and a Benchmark for Neural Network Potentials." Advances in Neural Information Processing Systems 37 (2024): 36869-36889.
13. Polykovskiy, Daniil, et al. "Entangled conditional adversarial autoencoder for de novo drug discovery." Molecular pharmaceutics 15.10 (2018): 4398-4405.
14. Николенко, Сергей, Кадури, Артур и Архангельская Екатерина. Глубокое обучение. Издательский дом "Питер", 2017..

## 5.2. Периодические издания и конференции (А\*):

1. IEEE Transactions on Big Data – научные статьи по обработке больших данных.
2. Journal of Big Data (SpringerOpen) – открытый журнал с исследованиями в области Big Data.
3. Big Data Research (Elsevier) – публикации по анализу, управлению и визуализации данных.
4. Data Science Journal (CODATA) – междисциплинарные исследования данных.

5. ACM Transactions on Knowledge Discovery from Data (TKDD) – методы извлечения знаний из больших данных.
6. <https://openreview.net/forum?id=FMMF1a9ifL>
7. <https://openreview.net/forum?id=EIUrNM9U8c#discussion>
8. <https://openreview.net/forum?id=JoO6mtCLHD>
9. <https://aclanthology.org/2024.findings-emnlp.760/>
10. <https://aclanthology.org/2020.coling-main.588/>
11. [https://link.springer.com/chapter/10.1007/978-3-030-72113-8\\_30](https://link.springer.com/chapter/10.1007/978-3-030-72113-8_30)
12. [https://link.springer.com/chapter/10.1007/978-3-031-42448-9\\_10](https://link.springer.com/chapter/10.1007/978-3-031-42448-9_10)
13. <https://aclanthology.org/2024.findings-naacl.288/>

### **5.3. Интернет-ресурсы, в том числе современные профессиональные базы данных и информационные справочные системы**

*Электронно-библиотечные системы (ЭБС):*

1. ЭБС «ЮРАЙТ» <https://urait.ru/>
2. ЭБС «УНИВЕРСИТЕТСКАЯ БИБЛИОТЕКА ОНЛАЙН» <http://www.biblioclub.ru/>
3. ЭБС «BOOK.ru» <https://www.book.ru>
4. ЭБС «ZNANIUM.COM» [www.znanium.com](http://www.znanium.com)
5. ЭБС «ЛАНЬ» <https://e.lanbook.com>

*Профессиональные базы данных*

1. Scopus <http://www.scopus.com/>
2. ScienceDirect <https://www.sciencedirect.com/>
3. Журналы издательства Wiley <https://onlinelibrary.wiley.com/>
4. Научная электронная библиотека (НЭБ) <http://www.elibrary.ru/>
5. Полнотекстовые архивы ведущих западных научных журналов на Российской платформе научных журналов НЭИКОН <http://archive.neicon.ru>
6. Национальная электронная библиотека (доступ к Электронной библиотеке диссертаций Российской государственной библиотеки (РГБ) <https://rusneb.ru/>
7. Президентская библиотека им. Б.Н. Ельцина <https://www.prlib.ru/>
8. База данных CSD Кембриджского центра кристаллографических данных (CCDC) <https://www.ccdc.cam.ac.uk/structures/>
9. Springer Journals: <https://link.springer.com/>
10. Springer Journals Archive: <https://link.springer.com/>
11. Nature Journals: <https://www.nature.com/>
12. Springer Nature Protocols and Methods: <https://experiments.springernature.com/sources/springer-protocols>
13. Springer Materials: <http://materials.springer.com/>
14. Nano Database: <https://nano.nature.com/>
15. Springer eBooks (i.e. 2020 eBook collections): <https://link.springer.com/>
16. "Лекториум ТВ" <http://www.lektorium.tv/>
17. Университетская информационная система РОССИЯ <http://uisrussia.msu.ru>

*Информационные справочные системы*

1. Консультант Плюс - справочная правовая система (доступ по локальной сети с компьютеров библиотеки)

*Ресурсы свободного доступа*

1. КиберЛенинка <http://cyberleninka.ru/>;
2. Американская патентная база данных <http://www.uspto.gov/patft/>
3. Министерство науки и высшего образования Российской Федерации <https://www.minobrnauki.gov.ru/>;
4. Федеральный портал "Российское образование" <http://www.edu.ru/>;

5. Информационная система "Единое окно доступа к образовательным ресурсам" <http://window.edu.ru/>;
6. Единая коллекция цифровых образовательных ресурсов <http://school-collection.edu.ru/> .
7. Проект Государственного института русского языка имени А.С. Пушкина "Образование на русском" <https://pushkininstitute.ru/>;
8. Справочно-информационный портал "Русский язык" <http://gramota.ru/>;
9. Служба тематических толковых словарей <http://www.glossary.ru/>;
10. Словари и энциклопедии <http://dic.academic.ru/>;
11. Образовательный портал "Учеба" <http://www.uceba.com/>;
12. Законопроект "Об образовании в Российской Федерации". Вопросы и ответы [http://xn--273--84d1f.xn--plai/voprosy\\_i\\_otvety](http://xn--273--84d1f.xn--plai/voprosy_i_otvety)

#### *Собственные электронные образовательные и информационные ресурсы КубГУ*

1. Электронный каталог Научной библиотеки КубГУ <http://megapro.kubsu.ru/MegaPro/Web>
2. Электронная библиотека трудов ученых КубГУ <http://megapro.kubsu.ru/MegaPro/UserEntry?Action=ToDb&idb=6>
3. Среда модульного динамического обучения <http://moodle.kubsu.ru>
4. База учебных планов, учебно-методических комплексов, публикаций и конференций <http://infoneeds.kubsu.ru/>
5. Библиотека информационных ресурсов кафедры информационных образовательных технологий <http://mschool.kubsu.ru;>
6. Электронный архив документов КубГУ <http://docspace.kubsu.ru/>
7. Электронные образовательные ресурсы кафедры информационных систем и технологий в образовании КубГУ и научно-методического журнала "ШКОЛЬНЫЕ ГОДЫ" <http://icdau.kubsu.ru/>

#### **6. Методические указания для обучающихся по освоению дисциплины (модуля)**

По курсу предусмотрено проведение лекционных занятий, на которых дается основной систематизированный материал. В ходе лекционных занятий разбираются элементы теории и практики дискретной математики, приводятся примеры решения задач, проводится анализ наиболее распространенных ошибок. После прослушивания лекции рекомендуется выполнить упражнения, приводимые в аудитории для самостоятельной работы.

По курсу предусмотрено проведение лабораторных занятий, на которых дается прикладной систематизированный материал. В ходе занятий разбираются методы решений задач по темам. После занятия рекомендуется выполнить упражнения, приводимые для самостоятельной работы.

При самостоятельной работе студентов необходимо изучить литературу, приведенную в перечнях выше, для осмысления вводимых понятий, анализа предложенных подходов и методов дискретной математики. При решении новой задачи студент должен уметь выбрать метод решения и его обоснование.

Важнейшим этапом курса является самостоятельная работа по дисциплине. В процессе самостоятельной работы студент приобретает навыки работы с дискретными объектами.

Используются активные, инновационные образовательные технологии, которые способствуют развитию общекультурных, общепрофессиональных компетенций и профессиональных компетенций обучающихся:

- проблемное обучение;
- разноуровневое обучение;
- проектные методы обучения;
- исследовательские методы в обучении;

- обучение в сотрудничестве (командная, групповая работа);
- информационно-коммуникационные технологии.

Для лиц с ограниченными возможностями здоровья предусмотрена организация консультаций с использованием электронной почты.

Учебно-методическим обеспечением курсовой работы студентов являются:

1. учебная литература;
2. нормативные документы ВУЗа;
3. методические разработки для студентов.

Самостоятельная работа студентов включает:

- оформление итогового отчета (пояснительной записки).
- анализ нормативно-методической базы организации;
- анализ научных публикации по заранее определённой теме;
- анализ и обработку информации;
- работу с научной, учебной и методической литературой,
- работа с конспектами лекций, ЭБС.

Для самостоятельной работы представляется аудитория с компьютером и доступом в Интернет, к электронной библиотеке вуза и к информационно-справочным системам.

Перечень учебно-методического обеспечения:

1. Основная образовательная программа высшего профессионального образования федерального государственного бюджетного образовательного учреждения высшего образования «Кубанский государственный университет» по направлению подготовки.
2. Положение о проведении текущего контроля успеваемости и промежуточной аттестации в федеральном государственном бюджетном образовательном учреждении высшего образования «Кубанский государственный университет».
3. Общие требования к построению, содержанию, оформлению и утверждению рабочей программы дисциплины Федерального государственного образовательного стандарта высшего профессионального образования.
4. Методические рекомендации по содержанию, оформлению и применению образовательных технологий и оценочных средств в учебном процессе, основанном на Федеральном государственном образовательном стандарте.
5. Учебный план основной образовательной программы по направлению подготовки.
6. Федеральный государственный образовательный стандарт высшего профессионального образования по направлению подготовки.

В освоении дисциплины инвалидами и лицами с ограниченными возможностями здоровья большое значение имеет индивидуальная учебная работа (консультации) – дополнительное разъяснение учебного материала.

Индивидуальные консультации по предмету являются важным фактором, способствующим индивидуализации обучения и установлению воспитательного контакта между преподавателем и обучающимся инвалидом или лицом с ограниченными возможностями здоровья.

Важнейшим компонентом курса является самостоятельная проектная работа, в ходе которой студент разрабатывает законченное решение для решения задач (кейсов) индустриальных партнеров. Допускается выполнение проектов в командах.

Подход, определяющий установление соответствия кейсов ИП и УГТ (5-7), позволяет четко соотносить этапы развития технологии с вовлеченностью партнера и снижать риски при переходе от лабораторных испытаний к промышленному внедрению.

## **А. Применение инструментальных средств нейросетевого моделирования в кейсах ПАО «Сбербанк»**

Кейс 1. Анализ факторов оттока клиентов и сегментация пользователей

Описание:

Компания хочет снизить отток клиентов и понять, какие типы пользователей наиболее склонны уйти в ближайшие месяцы. Требуется провести анализ транзакционной активности, обращений в поддержку и частоты использования сервиса.

Цель:

Разработать аналитический прототип на Python, который:

- Выполняет предобработку и очистку данных о клиентах (пропуски, выбросы, категоризация событий).
- Проводит разведочный анализ данных (EDA) с визуализацией ключевых метрик активности.
- Выполняет кластеризацию пользователей (например, k-means / DBSCAN) для выявления групп риска.
- Оценивает вклад признаков (поведение, обращения, метрики лояльности) в вероятность оттока. Технологии:
- Python (pandas, scikit-learn, matplotlib, seaborn).
- Методы кластеризации и понижения размерности (PCA для визуализации сегментов).
- BI-инструменты (Power BI / Tableau) для итогового дашборда.

Реализация:

- a) Сбор и подготовка набора данных: история транзакций, обращения в техподдержку, активность в приложении.
- b) Очистка и нормализация признаков, кодирование категориальных полей.
- c) Построение сегментов клиентов и выделение кластеров с повышенной вероятностью ухода.
- d) Построение дашборда с ключевыми метриками удержания.

Результат:

- Обоснованные сегменты пользователей (“лояльные”, “на грани ухода”, “низкоактивные новички”).
- Рекомендации по снижению оттока (персональные предложения, приоритизация обращений).

Кейс 2. Обнаружение подозрительных транзакций и аномалий в платёжных операциях

Описание:

Необходимо выявлять потенциально мошеннические операции или нетипичное поведение клиентов в финансовой системе.

Цель: Создать аналитическую модель на Python, которая:

- Определяет аномальные транзакции на основе величины платежа, геолокации, времени и устройства.
- Выделяет поведенческие паттерны (частота переводов, ночная активность, новые получатели).
- Формирует сигналы для службы риска.

Технологии:

- Выявление аномалий: Isolation Forest, Local Outlier Factor, One-Class SVM (scikit-learn).
- Ассоциативные правила для анализа последовательностей действий клиента (mlxtend / apriori).
- Визуализация распределений и временных паттернов (seaborn, plotly).

Реализация:

- а) Подготовка обезличенного набора транзакций с метками “норма” / “подозрительно”.
- б) Масштабирование числовых признаков и кодирование категориальных полей.
- в) Обучение и сравнение нескольких методов детекции аномалий.
- д) Формирование отчёта с показателями качества (precision/recall, доля ложных срабатываний).

Результат:

- Снижение доли ложных тревог за счёт адаптивной настройки порогов.
- Возможность раннего выявления подозрительной активности до финансовой потери.

Кейс 3. Прогнозирование спроса и динамики показателей по временным рядам

Описание:

Бизнесу нужно прогнозировать ключевые метрики (например, количество активных пользователей в день, количество покупок, загрузку инфраструктуры) на горизонте 1–4 недели.

Цель:

Построить модель прогнозирования временных рядов, которая:

- Учитывает тренд, сезонность и всплески активности.
- Даёт количественный прогноз на заданный период.
- Позволяет визуально сравнивать факт и прогноз.

Технологии:

- Анализ временных рядов: ARIMA / SARIMA, Prophet, экспоненциальное сглаживание.
- Обработка временных данных: ресемплирование, скользящее среднее, генерация лаговых признаков.

- Визуализация динамики и ошибок прогноза (matplotlib / plotly).

Реализация:

- a) Загрузка и очистка исторических данных (дневные/часовые ряды).
- b) Исследование сезонности и трендов, выделение аномальных всплесков.
- c) Обучение и сравнение как минимум двух моделей прогноза.
- d) Оценка точности прогноза (MAE, RMSE, MAPE) и интерпретация расхождений.

Результат:

- Прогноз ключевых метрик с заданным горизонтом.
- Выявление периодов риска (пики нагрузки, спады спроса) для планирования ресурсов.

Кейс 4. Оптимизация маркетинговых кампаний на основе A/B-тестирования

Описание:

Компания проводит рекламные кампании и хочет доказуемо оценивать, какие акции реально влияют на конверсию и удержание.

Цель:

Разработать процедуру аналитической оценки эффективности кампаний, которая:

- Планирует эксперимент (рандомизация аудиторий, контрольная и тестовая группы).
- Проверяет статистическую значимость результата.
- Формирует рекомендации по запуску/отмене кампаний.

Технологии:

- Проверка гипотез (t-test,  $\chi^2$ -тест, U-критерий Манна–Уитни).
- Расчёт доверительных интервалов, визуализация результатов.
- Построение отчёта в формате управленческой аналитики.

Реализация:

- a) Подготовка набора данных по тесту (пользователь → попал/не попал в акцию → конверсия).
- b) Расчёт метрик (CTR, CR, ARPU, удержание через 7/30 дней).
- c) Проверка статистической значимости различий между группами.
- d) Формирование чётких рекомендаций для маркетинга.

Результат:

- Прозрачная оценка влияния кампаний.
- Минимизация неэффективных расходов.

- Обоснованные рекомендации по масштабированию удачных механик.

Кейс 5. Построение управленческого дашборда и визуализация ключевых метрик

Описание:

Руководству нужен наглядный дашборд для мониторинга продукта: активность пользователей, конверсия, выручка, отток, нагрузка.

Цель: Построить дашборд, который:

- Обновляется из заданного источника данных.
- Показывает ключевые KPI по сегментам пользователей и по времени.
- Позволяет выделять зоны риска (падение вовлечённости, рост отказов).

Технологии:

- Power BI / Tableau / Plotly Dash.
- Подключение к CSV/SQL/логам событий.
- Построение интерактивных графиков (фильтры, drill-down).

Реализация:

- Импорт данных (например, продуктовая аналитика или финпоказатели по неделям).
- Построение набора визуализаций: воронка конверсии, распределение выручки, динамика оттока.
- Настройка фильтров (период, регион, сегмент пользователей).
- Подготовка пояснительной записки с интерпретацией метрик для управленцев.

Результат:

- Централизованный источник истины по ключевым показателям.
- Возможность оперативно выявлять проблемы и защищать решения на основе данных.

Итог:

Каждый кейс сочетает:

- Аналитические методы (EDA, кластеризация, прогнозирование, A/B-тестирование, анализ аномалий).
- Техническую реализацию (Python, scikit-learn, Power BI / Tableau, Jupyter / Colab).
- Прикладную пользу для бизнеса (снижение оттока, борьба с фродом, повышение эффективности маркетинга, поддержка управленческих решений).

**Б. Применение инструментальных средств аналитики данных кейсах компании AVA Lab**

**Кейс 1: Кейс 1. Оптимизация эффективности образовательных онлайн-курсов.**

**Описание:** AVA Lab анализирует эффективность онлайн-курсов и хочет определить, какие темы и форматы обучения дают максимальный прирост знаний и удержание студентов.

**Цель: Разработать аналитический прототип, который:**

- собирает и обрабатывает события студентов (просмотры, тесты, время в системе, переходы),
- выявляет факторы, влияющие на успеваемость,
- формирует рекомендации по улучшению учебных модулей.

**Технологии:**

Python (pandas, numpy), Power BI/Tableau, визуализация (seaborn/plotly), кластеризация (k-means), ассоциативные правила.

**Реализация:**

- a) Сбор логов прохождения курсов и результатов тестов
- b) Очистка данных, нормализация, выделение учебных паттернов.
- c) Кластеризация студентов (например: “быстрые”, “пассивные”, “медленно прогрессирующие”).
- d) Дашборд с проблемными уроками и показателями усвоения.

**Результат:**

- Повышение вовлечённости на 15–25%.
- Сокращение числа «брошенных модулей».
- Рекомендации по оптимизации структуры курса

**Кейс 2: Детекция аномалий в студенческих тестах и анти-чит система**

**Описание:** AVA Lab фиксирует случаи аномально быстрых прохождений тестов, использования подсказок и совпадений ответов. Требуется аналитическая система обнаружения подозрительного поведения.

**Цель:** Создать модель, которая:

- анализирует время ответов, сложность вопросов, паттерны ошибок.
- выделяет студентов с “нетипично высокой” скоростью или повторяющимся поведением.
- формирует отчёт для службы контроля качества.

**Технологии:**

Isolation Forest, Local Outlier Factor, статистические метрики, визуализация распределений.

**Реализация**

- a) Построение набора признаков: скорость, частота подсказок, совпадения ответов.
- b) Обучение моделей поиска аномалий и ранжирование студентов по уровню риска.
- c) Формирование списка подозрительных кейсов и пояснений (explainable AI).

**Результат**

- Повышение объективности оценки знаний
- Честная и прозрачная система контроля.

**Кейс 3: Прогнозирование нагрузки на серверы и пиковой посещаемости платформы**

**Описание:** Во время запусков новых курсов серверы AVA Lab перегружаются. Нужно предсказывать будущую нагрузку, чтобы масштабировать инфраструктуру.

**Цель:** Построить модель прогнозирования временных рядов, которая:

- учитывает сезонность и всплески трафика,
- прогнозирует нагрузку на день/неделю вперёд,
- предупреждает о пиковых периодах.

**Технологии**

ARIMA/SARIMA, Prophet, экспоненциальное сглаживание, анализ временных рядов, Plotly.

### **Реализация**

- a) Загрузка исторических данных посещений по часам/дням
- b) Очистка аномалий, пропусков, агрегация
- c) Обучение модели, сравнение нескольких алгоритмов
- d) Визуализация прогноза и построение уведомлений

### **Результат**

- Снижение количества перегрузок и отказов.
- Экономия за счёт оптимизации аренды серверов.
- Планирование технических работ без простоев.

### **Кейс 4: Персонализация рекомендаций учебных материалов**

**Описание:** AVA Lab хочет рекомендовать каждому студенту индивидуальные задания, задания по слабым темам и видео с подходящим уровнем сложности.

**Цель:** Разработать систему рекомендаций, которая:

- Выделяет ключевые категории (сенсоры, ПО, механические сбои).
- Определяет срочность решений.

### **Технологии**

Content-based filtering, cosine similarity, TF-IDF, кластеризация пользователей, Python.

### **Реализация:**

- a) Анализ выполненных заданий, ошибок, просмотренных материалов
- b) Выделение тематических профилей и уровней сложности
- c) Формирование персональных рекомендаций
- d) A/B-тест эффективности (рост вовлечённости и оценок)

### **Результат:**

- Повышение успеваемости и завершения курсов.
- Рост мотивации студентов.
- Улучшение качества адаптивного обучения.

### **Кейс 5: Аналитический дашборд для мониторинга эффективности платформы**

**Описание:** У руководства AVA Lab нет единого источника аналитики по курсам, студентам, тестам, отзывам и выручке.

**Цель:** Создать управленческий дашборд, который:

- собирает данные из разных источников (CRM, LMS, платежи),
- показывает ключевые KPI,
- помогает принимать решения на основе данных.

### **Технологии**

Power BI, Tableau или Plotly Dash; подключение к SQL/CSV; интерактивные графики.

### **Реализация**

- a) Интеграция данных (студенты, курсы, оплата, успеваемость)
- b) Построение метрик: удержание, воронка обучения, вовлечённость
- c) Настройка фильтров по курсам, регионам, типам пользователей
- d) Подготовка отчёта и сценариев интерпретации

### **Результат**

- Руководство видит реальные показатели бизнеса.
- Уменьшается количество “слепых” управленческих решений.

### **Итог**

Каждый кейс решает конкретную задачу AVA Lab:

## **7. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю)**

### **7.1 Перечень информационно-коммуникационных технологий**

1. Электронная почта mail.ru, yandex.ru

2. Yandex Browser
3. Система управления обучением Moodle – сдача работ

## 7.2 Перечень лицензионного и свободно распространяемого программного обеспечения

1. OpenOffice
2. GIT
3. Yandex Browser
4. Mozilla Firefox
5. Google Chrome
6. Python + Jupyter + Google Colab
7. SymPy/SageMath
8. Octave (аналог MATLAB)

## 8. Материально-техническое обеспечение по дисциплине (модулю)

№	Вид работ	Наименование учебной аудитории, ее оснащенность оборудованием и техническими средствами обучения
1.	Лекционные занятия	Аудитория, укомплектованная специализированной мебелью и техническими средствами обучения
2.	Лабораторные занятия	Аудитория, укомплектованная специализированной мебелью и техническими средствами обучения, компьютерами, проектором, программным обеспечением
3.	Групповые (индивидуальные) консультации	Аудитория, укомплектованная специализированной мебелью и техническими средствами обучения, компьютерами, программным обеспечением
4.	Текущий контроль, промежуточная аттестация	Аудитория, укомплектованная специализированной мебелью и техническими средствами обучения, компьютерами, программным обеспечением
5.	Самостоятельная работа	Кабинет для самостоятельной работы, оснащенный компьютерной техникой с возможностью подключения к сети «Интернет», программой экранного увеличения и обеспеченный доступом в электронную информационно-образовательную среду университета.

№	Продукт	Параметры продукта	Кол-во	Кол-во конфигураций	Ед. изм.		
1	Виртуальная машина	Виртуальная машина 10% vCPU	1	60	Шт		
		2 vCPU 4 RAM					
		ОС Ubuntu 22.04	1			Шт	
		Системный диск SSD	1			Шт	
			10			Гб	
	Аренда публичного IP	1	Шт				
2	Виртуальная машина с GPU	Виртуальная машина с GPU	1	1	Шт		
		NVIDIA® Tesla® V100 2 GPU 8					
		vCPU 128 ГБ RAM					
		ОС Ubuntu_24.04				1	Шт
		Системный диск SSD				1	Шт
						2000	Гб
	Диск SSD	1	Шт				
		4096	Гб				
	Диск SSD	1	Шт				

			4096		Гб
		Аренда публичного IP	1		Шт
3	K8S	Master node 8 vCPU 16 RAM	1	1	Шт
		Worker node 10% доля 4 vCPU 32 RAM	5		Шт
		Worker node SSD-NVME	64		Гб
		Аренда публичного IP	1		Шт
4	ML Inference Instance Type GPU	Время работы в месяц	40	1	Ч
		Инстанс 8 x NVIDIA® H100 NVLink PCIe 160 vCPU 1520 GB RAM	1		Шт
		Количество запросов к ML-моделям	1		Млн. Шт
		Кэш ML-моделей	160		Гб
5	LLM	Токены GigaChat 2 Max	50		Млн. Шт
		Токены Embeddings	400		Млн. Шт

Дополнительные облачные ресурсы предоставляются технологическим партнером Yandex Cloud.

Примечание: Конкретизация аудиторий и их оснащение определяется ОПОП.