

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Макаренко Елена Николаевна

Должность: Рector

Дата подписания: 15.08.2024 17:28:05

Уникальный программный ключ:

c098bc0c1041cb2a4cf926cf171d6715d99a6ae00adc8e27b55cbe1e2dbd7c78

Министерство науки и высшего образования Российской Федерации

Федеральное государственное бюджетное образовательное учреждение высшего образования «Ростовский государственный экономический университет (РИНХ)»

УТВЕРЖДАЮ

Директор Института магистратуры

Иванова Е.А.

«03» июня 2024 г.

**Рабочая программа дисциплины  
Системы аналитики больших данных**

Направление 09.04.03 Прикладная информатика

магистерская программа

09.04.03.03 Машинное обучение и технологии больших данных

Для набора 2024 года

Квалификация

магистр

Кафедра Информационных систем и прикладной информатики

**Составители рабочей программы:**

д.э.н., зав.каф. Щербаков Сергей Михайлович

## СОДЕРЖАНИЕ

I. Цели и задачи освоения дисциплины .....	4
II. Место дисциплины в структуре образовательной программы.....	4
III. Требования к результатам освоения дисциплины .....	8
IV. Содержание и структура дисциплины.....	9
4.1. Содержание дисциплины, структурированное по темам.....	9
4.2. План внеаудиторной самостоятельной работы.....	10
4.3. Содержание учебного материала.....	12
V. Образовательные технологии .....	13
VI. Учебно-методическое обеспечение дисциплины .....	13
6.1. Основная литература .....	13
6.2. Дополнительная литература.....	13
6.3. Периодические издания.....	14
6.4. Перечень ресурсов сети Интернет.....	14
VII. Материально-техническое обеспечение дисциплины .....	14
VIII. Методические указания для обучающихся по освоению дисциплины .....	15
IX. Учебная карта дисциплины .....	16
X. Фонд оценочных средств.....	17
10.1. Паспорт фонда оценочных средств .....	17
10.2. Практическая работа №1 .....	17
10.3. Практическая работа №2 .....	18
10.4. Практическая работа №3 .....	18
10.4. Экзаменационные вопросы и билеты.....	20

## I. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Цели освоения дисциплины: формирование у обучающихся способности осуществлять руководство проектами по созданию, поддержке и использованию комплексных систем на основе аналитики больших данных с применением новых методов и алгоритмов машинного обучения.

Задачи освоения дисциплины:

- формирование у обучающихся знаний о методологии и принципах руководства проектами по созданию, поддержке и использованию комплексных систем на основе аналитики больших данных со стороны заказчика;
- развитие у обучающихся умения применять современные инструментальные средства и системы программирования для разработки новых методов и моделей машинного обучения;
- развитие у обучающихся умения решать задачи по руководству коллективной проектной деятельностью для создания, поддержки и использования комплексных систем на основе аналитики больших данных со стороны заказчика.

## II. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Дисциплина относится к модулю обязательных профессиональных дисциплин обязательной части образовательной программы.

Для изучения данной дисциплины необходимы знания, умения и навыки, формируемые предшествующими элементами образовательной программы:

Наименование дисциплины (модуля), практики	Требуемые знания, умения, навыки
Методология научной деятельности	<p><i>Знания:</i></p> <ul style="list-style-type: none"><li>– Основы методологии научной и проектной деятельности.</li><li>– Стратегия действий для достижения поставленной цели.</li><li>– Информационно-аналитические системы для оперативного обеспечения научных исследований актуальной справочно-библиографической информацией и осуществления оценки результативности и эффективности деятельности научно-исследовательских организаций, ученых.</li><li>– Наукометрические инструменты, сервисы, платформы.</li><li>– Основные понятия научного коллектива, научной школы и невидимого колледжа как форм академического взаимодействия.</li><li>– Знает правовую базу информационного законодательства, правовые нормы и стандарты в области искусственного интеллекта и смежных областей.</li><li>– Знает содержание нормативно-правовых документов в сфере информационных технологий, искусственного интеллекта и информационной безопасности.</li><li>– Новые и существующие достижения науки для анализа и постановки задачи исследования.</li><li>– Работа с информационными системами анализа данных.</li><li>– Методы решения нестандартных задач в профессиональной деятельности.</li><li>– Математические основы наукометрии.</li><li>– Специфика регистрации интеллектуальной собственности источники и типы трендов, в том числе научной деятельности.</li><li>– Принципы представления результатов НИР.</li><li>– Принципы подготовки отчетов о НИР в соответствии с ГОСТ 7.32–2017.</li></ul>

Наименование дисциплины (модуля), практики	Требуемые знания, умения, навыки
	<ul style="list-style-type: none"> <li>– Правила оформления библиографического описания и библиографических ссылок в соответствии с ГОСТ Р 7.0.100–2018.</li> <li>– Принципы эффективной подготовки отчётов о НИР в офисном пакете Microsoft Word и издательской среде LaTeX.</li> <li>– Знает содержание, объекты и субъекты информационного общества и цифровой экономики, критерии эффективности функционирования информационного общества, теоретические проблемы информатики, искусственного интеллекта, современные методы, средства, стандарты информатики для решения прикладных задач различных классов; правовые, экономические, социальные и психологические аспекты информатизации деятельности организационно-экономических систем.</li> <li>– Знает состав современных методов и средств информатики, передовые методы искусственного интеллекта для решения задач профессиональной деятельности.</li> </ul> <p><i>Умения:</i></p> <ul style="list-style-type: none"> <li>– Применять системный подход в научно-исследовательской и проектной деятельности.</li> <li>– Вырабатывать стратегию действий при выполнении научно-исследовательских работ и проектов.</li> <li>– Работать с ведущими базами данных научных статей и патентов.</li> <li>– Применять наукометрические инструменты для анализа сетей научных коммуникаций.</li> <li>– Умеет применять правовые нормы и стандарты в области искусственного интеллекта при создании систем искусственного интеллекта.</li> <li>– Умеет применять этические нормы и стандарты в области искусственного интеллекта при создании систем искусственного интеллекта.</li> <li>– Умеет использовать нормативно-правовые документы в сфере информационных технологий, искусственного интеллекта и информационной безопасности при разработке стандартов, норм и правил.</li> <li>– Работать и российскими и зарубежными научными и патентными базами данных.</li> <li>– Применять наукометрические методы картирования науки и анализа динамики предметной области.</li> <li>– Осуществлять постановку и формализацию задач в профессиональной сфере.</li> <li>– Самостоятельно анализировать информацию и получать обоснованные выводы.</li> <li>– Осуществлять поиск и анализ информации в базах данных ведущих патентных ведомствах мира.</li> <li>– Применять наукометрические методы и аналитические инструменты для самостоятельного анализа научных трендов.</li> <li>– Структурировать результаты научно-исследовательской деятельности в соответствии принципами подготовки отчётов по</li> </ul>

Наименование дисциплины (модуля), практики	Требуемые знания, умения, навыки
	<p>НИР.</p> <ul style="list-style-type: none"> <li>– Составлять аналитические обзоры с обоснованными выводами.</li> <li>– Оформлять результаты научно-исследовательской деятельности в соответствии со стандартами, нормами и правилами.</li> <li>– Умеет применять при решении задач профессиональной деятельности критерии эффективности функционирования информационного общества и цифровой экономики; структуру интеллектуального капитала, методы оценки эффективности.</li> <li>– Умеет проводить анализ современных методов и средств информатики и искусственного интеллекта для решения прикладных задач различных классов.</li> </ul> <p><i>Навыки:</i></p> <ul style="list-style-type: none"> <li>– Критического анализа при наличии проблемных ситуаций.</li> <li>– Разработки идеологии жизненного цикла проекта.</li> <li>– Поиска, обработки и анализа информации из ведущих баз данных научных статей и патентов России и мира.</li> <li>– Визуализации анализа сетей научных коммуникаций на основе карты компетенций ведущих ученых и организаций в предметной области.</li> <li>– Работы с информационными системами научных баз данных и патентов.</li> <li>– Проведения аналитических обзоров научных результатов в междисциплинарном контексте задач профессиональной деятельности.</li> <li>– Работы с патентно-информационными ресурсами.</li> <li>– Визуализации анализа глобальных научных трендов с использованием наукометрических систем.</li> <li>– Подготовки отчётов о результатах НИР в виде аналитического обзора.</li> <li>– Подготовки отчётов о результатах научно-исследовательской работы.</li> </ul>
Математические методы анализа больших данных	<p><i>Знания:</i></p> <ul style="list-style-type: none"> <li>– Знает классы методов и алгоритмов машинного обучения.</li> <li>– Знает математические модели, методы и алгоритмы для обработки и анализа больших данных.</li> <li>– Знает принципы построения моделей глубоких нейронных сетей и глубокого машинного обучения.</li> <li>– Знает подходы к применению моделей на основе нечеткой логики в системах искусственного интеллекта.</li> </ul> <p><i>Умения:</i></p> <ul style="list-style-type: none"> <li>– Умеет ставить задачи и адаптировать методы и алгоритмы машинного обучения.</li> <li>– Умеет выбирать и применять математические модели, методы и алгоритмы для решения прикладных задач анализа больших данных.</li> <li>– Умеет руководить выполнением коллективной проектной деятельности для создания, поддержки и использования систем</li> </ul>

Наименование дисциплины (модуля), практики	Требуемые знания, умения, навыки
	искусственного интеллекта на основе моделей глубоких нейронных сетей и нечетких моделей и методов.
Методология проектирования и управления информационными системами	<p><i>Знания:</i></p> <ul style="list-style-type: none"> <li>– Актуальные проблемы разработки сложных программных систем.</li> <li>– Эволюции моделей жизненного цикла информационных систем.</li> <li>– Онтологический подход концептуального моделирования предметной области.</li> <li>– Теории моделирования систем из объектов.</li> <li>– Парадигмы программирования.</li> <li>– Основы управления ИТ-инфраструктурой, базирующимся на понятии информационного сервиса, моделях управления информационными системами (ITSM), библиотеках ITIL.</li> <li>– Методологии ведения программных проектов: структурное, функциональное, объектно-ориентированное и унифицированное моделирование.</li> </ul> <p><i>Умения:</i></p> <ul style="list-style-type: none"> <li>– Сравнить процессы проектирования и управления, принятые в различных парадигмах программирования.</li> <li>– Применять методы патентных исследований при создании инновационных продуктов в области профессиональной деятельности.</li> <li>– Разрабатывать оригинальные программные средства, в том числе с использованием современных информационно-коммуникационных и интеллектуальных компьютерных технологий, для решения задач в области создания и применения искусственного интеллекта.</li> <li>– Формировать и анализировать модели представления знаний.</li> <li>– Определять критерии и метрики оценки результатов моделирования при построении систем искусственного интеллекта в исследуемой области.</li> <li>– Применять методы многомерного анализа данных.</li> </ul> <p><i>Навыки:</i></p> <ul style="list-style-type: none"> <li>– Обоснования выбора модели жизненного цикла ИС.</li> <li>– Модернизации программного и аппаратного обеспечения информационных и автоматизированных систем для решения профессиональных задач.</li> <li>– Проектирования процессов и практик методологии Rapid Application Development, Unified Process, OpenUP.</li> </ul>

Знания, умения и навыки, формируемые данной дисциплиной, потребуются при освоении следующих элементов образовательной программы:

- производственная практика, проектно-технологическая практика;
- производственная практика, преддипломная практика.

### III. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Освоение дисциплины направлено на формирование следующих компетенций в соответствии с образовательной программой:

#### Перечень планируемых результатов обучения по дисциплине, соотнесённых с индикаторами достижения компетенций

Компетенция	Индикаторы достижения компетенции	Результаты обучения
ПК-2. Способен руководить проектами по созданию систем искусственного интеллекта с применением новых методов и алгоритмов машинного обучения со стороны заказчика	ПК-2.2. Осуществляет руководство созданием комплексных систем искусственного интеллекта с применением новых методов и алгоритмов машинного обучения	<i>Знания:</i> <ul style="list-style-type: none"><li>– Знает функциональность современных инструментальных средств и систем программирования в области создания моделей и методов машинного обучения.</li><li>– Знает принципы построения систем искусственного интеллекта, методы и подходы к планированию и реализации проектов по созданию систем искусственного интеллекта, методы интеллектуального планирования экспериментов.</li></ul> <i>Умения:</i> <ul style="list-style-type: none"><li>– Умеет применять современные инструментальные средства и системы программирования для разработки новых методов и моделей машинного обучения.</li><li>– Умеет руководить выполнением коллективной проектной деятельности для создания, поддержки и использования систем искусственного интеллекта.</li></ul>
ПК-4. Способен руководить проектами по созданию комплексных систем на основе аналитики больших данных в различных отраслях со стороны заказчика	ПК-4.1. Руководит проектами по построению комплексных систем на основе аналитики больших данных в различных отраслях со стороны заказчика	<i>Знания:</i> <ul style="list-style-type: none"><li>– Знает методологию и принципы руководства проектами по созданию, поддержке и использованию комплексных систем на основе аналитики больших данных со стороны заказчика.</li><li>– Знает специфику сфер и отраслей, для которых реализуется проект по аналитике больших данных</li></ul> <i>Умения:</i> <ul style="list-style-type: none"><li>– Умеет решать задачи по руководству коллективной проектной деятельностью для создания, поддержки и использования комплексных систем на основе аналитики больших данных со стороны заказчика.</li><li>– Умеет выявлять небольшие по масштабу проекты аналитики, которые потенциально могут представлять интерес для ряда подразделений / служб или для организации в целом.</li><li>– Умеет выявлять области деловой деятельности, которые потенциально могут получить отдачу от аналитики.</li></ul>



## IV. СОДЕРЖАНИЕ И СТРУКТУРА ДИСЦИПЛИНЫ

Трудоёмкость дисциплины составляет 5 зачётных единиц, 180 часов.  
в том числе 1 зачётная единица, 36 часов на экзамен

Форма промежуточной аттестации: экзамен.

### 4.1. Содержание дисциплины, структурированное по темам

№ п/п	Темы дисциплины	Семестр	Виды учебной работы и их трудоёмкость, часы (в том числе с использованием онлайн-курсов)				Наименования оценочных средств
			Контактная работа			Самостоя- тельная работа	
			Лекции	Практические занятия	Лабораторные занятия		
<b>Модуль 1. Основы построения и работы с системами аналитики больших данных</b>							
1	<b>Архитектура систем аналитики больших данных.</b>	3	2	-	-	6	Практическая работа №1
2	<b>Хранение больших данных в облаке.</b> 1. Хранилища общего назначения. Форматы хранения данных. Облачное хранилище Microsoft Azure Storage. Облачное хранилище AWS. 2. Реляционные базы данных. Azure SQL. AWS RDS. 3. Нереляционные базы данных. Сервисы нереляционных баз данных от Azure и AWS. 4. Корпоративные хранилища данных (DWH). Azure SQL DWH. AWS RedShift. 5. Хранилища данных типа Data Lake. Azure Data Lake Store. AWS Data Lake Solutions.	3	4	12	-	28	Практическая работа №1

Модуль 2. Разработка систем аналитики больших данных							
3	<p><b>Доставка больших данных в облако.</b></p> <p>1. Прямая загрузка данных. Доставка данных в облачное хранилище общего назначения. Доставка данных в реляционные БД и хранилища. Доставка данных в нереляционные базы данных. Доставка данных в HDFS-совместимые хранилища.</p> <p>2. Прямая загрузка потоковых данных. Azure Event Hub. AWS Kinesis Data Streams. Облачные сервисы развертывания кластерных систем.</p> <p>3. Облачные сервисы копирования и трансформации данных. Azure Data Factory. AWS Data Pipeline. AWS Glue.</p>	3	6	12	-	28	Практическая работа №2
4	<p><b>Аналитика больших данных в облаке.</b></p> <p>1. Интерактивный анализ данных. Анализ реляционных данных. Azure Data Lake Analytics. AWS Athena. Apache Spark. Встроенные редакторы запросов сервиса CosmosDB</p> <p>2. Потоковый анализ данных. Общие сведения. Azure Stream Analytics. Amazon Kinesis Analytics. Apache Storm.</p> <p>3. Пакетный анализ данных. Hadoop. Apache Pig. Apache Hive.</p>	3	6	12	-	28	Практическая работа №3
Промежуточная аттестация (для дисциплин с экзаменом)		3	–	–	–	36	Экзаменационные вопросы и билеты
<b>Итого часов</b>			<b>18</b>	<b>36</b>	<b>-</b>	<b>126</b>	<b>–</b>

#### 4.2. План внеаудиторной самостоятельной работы

№ п/п	Темы дисциплины	Семестр	Вид самостоятельной работы	Сроки выполнения (нед.)	Затраты времени (часы)	Учебно-методическое обеспечение
Модуль 1. Основы построения и работы с системами аналитики больших данных						
1	<b>Архитектура систем аналитики больших данных.</b>	3	– проработка и повторение материала лекционных занятий;	1–2	6	[1]-[4]

№ п/п	Темы дисциплины	Семестр	Вид самостоятельной работы	Сроки выполнения (нед.)	Затраты времени (часы)	Учебно-методическое обеспечение
2	<p><b>Хранение больших данных в облаке.</b></p> <p>1. Хранилища общего назначения. Форматы хранения данных. Облачное хранилище Microsoft Azure Storage. Облачное хранилище AWS.</p> <p>2. Реляционные базы данных. Azure SQL. AWS RDS.</p> <p>3. Нереляционные базы данных. Сервисы нереляционных баз данных от Azure и AWS.</p> <p>4. Корпоративные хранилища данных (DWH). Azure SQL DWH. AWS RedShift.</p> <p>5. Хранилища данных типа Data Lake. Azure Data Lake Store. AWS Data Lake Solutions.</p>	3	<ul style="list-style-type: none"> <li>– проработка и повторение материала лекционных занятий;</li> <li>– подготовка к практическим занятиям</li> </ul>	3-10	28	[1]-[4]
<b>Модуль 2. Разработка систем аналитики больших данных</b>						
4	<p><b>Доставка больших данных в облако.</b></p> <p>1. Прямая загрузка данных. Доставка данных в облачное хранилище общего назначения. Доставка данных в реляционные БД и хранилища. Доставка данных в нереляционные базы данных. Доставка данных в HDFS-совместимые хранилища.</p> <p>2. Прямая загрузка потоковых данных. Azure Event Hub. AWS Kinesis Data Streams. Облачные сервисы развертывания кластерных систем.</p> <p>3. Облачные сервисы копирования и трансформации данных. Azure Data Factory. AWS Data Pipeline. AWS Glue.</p>	3	<ul style="list-style-type: none"> <li>– проработка и повторение материала лекционных занятий;</li> <li>– подготовка к практическим занятиям</li> </ul>	11-14	28	[1]-[4]
5	<p><b>Аналитика больших данных в облаке.</b></p> <p>1. Интерактивный анализ данных. Анализ реляционных данных. Azure Data Lake Analytics. AWS Athena. Apache Spark. Встроенные редакторы запросов сервиса CosmosDB</p> <p>2. Поточковый анализ данных. Общие сведения. Azure Stream Analytics. Amazon Kinesis Analytics. Apache Storm.</p> <p>3. Пакетный анализ данных. Hadoop. Apache Pig. Apache Hive.</p>	3	<ul style="list-style-type: none"> <li>– проработка и повторение материала лекционных занятий;</li> <li>– подготовка к практическим занятиям</li> </ul>	15-18	28	[1]-[4]
<b>Подготовка к экзамену (для дисциплин с экзаменом)</b>					<b>36</b>	<b>[1]-[4]</b>
<b>Общая трудоёмкость самостоятельной работы по дисциплине</b>					<b>126</b>	<b>–</b>

### 4.3. Содержание учебного материала

#### Модуль 1. Основы построения и работы с системами аналитики больших данных

Архитектура систем аналитики больших данных.

Хранение больших данных в облаке.

1. Хранилища общего назначения. Форматы хранения данных. Облачное хранилище Microsoft Azure Storage. Облачное хранилище AWS.
2. Реляционные базы данных. Azure SQL. AWS RDS.
3. Нереляционные базы данных. Сервисы нереляционных баз данных от Azure и AWS.
4. Корпоративные хранилища данных (DWH). Azure SQL DWH. AWS RedShift.
5. Хранилища данных типа Data Lake. Azure Data Lake Store. AWS Data Lake Solutions.

#### Модуль 2. Разработка систем аналитики больших данных

Доставка больших данных в облако.

1. Прямая загрузка данных. Доставка данных в облачное хранилище общего назначения. Доставка данных в реляционные БД и хранилища. Доставка данных в нереляционные базы данных. Доставка данных в HDFS-совместимые хранилища.
2. Прямая загрузка потоковых данных. Azure Event Hub. AWS Kinesis Data Streams. Облачные сервисы развертывания кластерных систем.
3. Облачные сервисы копирования и трансформации данных. Azure Data Factory. AWS Data Pipeline. AWS Glue.

Аналитика больших данных в облаке.

1. Интерактивный анализ данных. Анализ реляционных данных. Azure Data Lake Analytics. AWS Athena. Apache Spark. Встроенные редакторы запросов сервиса CosmosDB
2. Потоковый анализ данных. Общие сведения. Azure Stream Analytics. Amazon Kinesis Analytics. Apache Storm.
3. Пакетный анализ данных. Hadoop. Apache Pig. Apache Hive.

#### Перечень тем практических занятий

№ п/п	Тема практического занятия	Количество часов
<b>Модуль 1. Основы построения и работы с системами аналитики больших данных</b>		
1	Хранение больших данных в облаке. Развертывание облачного хранилища общего назначения	12
<b>Модуль 2. Разработка систем аналитики больших данных</b>		
2	Доставка больших данных в облако. Развертывание облачных сервисов копирования и трансформации данных	12
3	Аналитика больших данных в облаке. Развертывание облачных сервисов для потокового анализа данных	12
<b>Всего часов</b>		<b>36</b>

## V. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ

Наряду с традиционными образовательными технологиями, для реализации дисциплины будут использоваться технологии электронного обучения и дистанционные образовательные технологий в электронной информационно-образовательной среде университета.

Активные формы обучения, применяемые на практических занятиях, способствуют разнообразному (индивидуальному, групповому, коллективному) изучению (усвоению) учебных вопросов (проблем), активному взаимодействию обучающихся и преподавателя, живому обмену мнениями между ними, нацеленному на выработку правильного понимания содержания изучаемой темы и способов ее практического использования.

Аудиторные занятия и другие формы контактной работы обучающихся с преподавателем могут проводиться с использованием платформ Microsoft Teams, Moodle (BigBlueButton) и др., что позволяет обеспечить онлайн и офлайн взаимодействие преподавателя с обучающимися в рамках дисциплины

Основными методами текущего контроля являются электронный учёт и контроль учебных достижений студентов (использование средств сервиса балльно-рейтинговой системы; ведение электронного журнала успеваемости, проведение электронного тестирования и применение других средств контроля с использованием системы электронного обучения).

## VI. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

### 6.1. Основная литература

1. Прокопенко, Н. Ю. Аналитические информационные системы поддержки принятия решений : учебное пособие / Н. Ю. Прокопенко. — Нижний Новгород : Нижегородский государственный архитектурно-строительный университет, ЭБС АСВ, 2020. — 143 с. — ISBN 978-5-528-00395-5. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <https://www.iprbookshop.ru/107361.html>

2. Алексеев Д. С. Технологии интеллектуального анализа данных [Электронный ресурс]: учебное пособие / Алексеев Д. С. – Кострома: КГУ им. Н.А. Некрасова, 2020. - 141 с. – URL: <https://e.lanbook.com/book/160082>.

3. Железнов М. М. Методы и технологии обработки больших данных [Электронный ресурс]: учебно-методическое пособие / Железнов М. М. - Москва: МИСИ – МГСУ, 2020. - 46 с. – URL: <https://e.lanbook.com/book/145102>.

4. Воронов, В. И. Data Mining - технологии обработки больших данных : учебное пособие / В. И. Воронов, Л. И. Воронова, В. А. Усачев. – Москва : Московский технический университет связи и информатики, 2018. – 47 с. – ISBN 2227-8397. – Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. – URL: <https://www.iprbookshop.ru/81324.html>

### 6.2. Дополнительная литература

5. Интеллектуальный предиктивный мультимодальный анализ слабоструктурированных больших данных / Н. Г. Ярушкина, И. А. Андреев, Г. Ю. Гуськов [и др.]. – Ульяновск : Ульяновский государственный технический университет, 2020. – 221 с. – ISBN 978-5-9795-2088-9. – Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. – URL: <https://www.iprbookshop.ru/106136.html>

6. Билл, Фрэнкс Укрощение больших данных : как извлекать знания из массивов информации с помощью глубокой аналитики / Фрэнкс Билл ; перевод А. Баранов. – Москва : Манн, Иванов и Фербер, 2014. – 340 с. – ISBN 978-5-00057-146-0. – Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. – URL: <https://www.iprbookshop.ru/39433.html>

7. Крутиков В. Н. Анализ данных / В.Н. Крутиков; В.В. Мешечкин - Кемерово: Кемеровский государственный университет, 2014. – 138 с. <http://biblioclub.ru/index.php?page=book&id=278426>

8. Жуковский О. И. Информационные технологии и анализ данных: учебное пособие / О.И. Жуковский – Томск: Эль Контент, 2014. – 130 с. <http://biblioclub.ru/index.php?page=book&id=480500>

9. Е.И. Николаев. Базы данных в высокопроизводительных информационных системах : учебное пособие / авт.-сост. Е. И. Николаев ; Северо-Кавказский федеральный университет. – Ставрополь : Северо-Кавказский Федеральный университет (СКФУ), 2016. – 163 с. – URL: <https://biblioclub.ru/index.php?page=book&id=466799>

10.Б. Фрэнкс. Революция в аналитике: Как в эпоху Big Data улучшить ваш бизнес с помощью операционной аналитики / Фрэнкс Билл ; перевод И. Евстигнеева ; под редакцией В. Мылова. — Москва : Альпина Паблишер, 2020. — 320 с. — ISBN 978-5-9614-5302-7. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <https://www.iprbookshop.ru/93032.html>

### 6.3. Периодические издания

- IEEE Spectrum <https://spectrum.ieee.org/>
- Journal of Big Data <https://journalofbigdata.springeropen.com/>
- Научный журнал «Машинное обучение и анализ данных» <http://jmla.org/ru/journal>

### 6.4. Перечень ресурсов сети Интернет

- ЭБС IPR Books <http://www.iprbookshop.ru/>
- ЭБС «Университетская библиотека онлайн» <http://biblioclub.ru>.
- Образовательная платформа Юрайт <https://urait.ru/>
- IBM Academic Initiative [http://ictis.sfedu.ru/ibm\\_academic\\_initiative/](http://ictis.sfedu.ru/ibm_academic_initiative/) (учебные материалы)
- <http://github.com/>
- <http://habr.com/>
- <http://www.kdnuggets.com/>
- Python, Свободное ПО, <https://www.python.org/>
- <https://www.jetbrains.com/pycharm/>

## VII. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

При реализации дисциплины используются следующие помещения, оборудование и программное обеспечение:

*Лаборатория машинного обучения и технологий больших данных*

Персональные компьютеры (8 шт.), проектор, экран. Windows 10, Microsoft Office 365, Adobe Acrobat Reader (Бесплатное проприетарное ПО, <https://acrobat.adobe.com/ru/ru/acrobat/pdf-reader/volume-distribution.html>), Google Chrome (Свободное ПО, <https://google.com/chrome/browser/>), Mozilla Firefox, Бесплатное ПО (GNU GPL), <https://firefox.com/>, Foxit (Бесплатное проприетарное ПО, <https://www.foxitsoftware.com/ru/>), i2 Analyst's Notebook (Бесплатная лицензия для образовательных целей), <https://developer.ibm.com/academic/>), Notepad++, Бесплатное ПО (GNU GPL 2), <https://notepad-plus-plus.org/>, Total Commander 7.x , WinRAR, XAMPP, Бесплатное ПО (GNU GPL), <http://www.apachefriends.org/en/xampp.html>, Team Foundation Server 2015, Visual Studio 2015, Android Studio, Операционная система на базе Linux; Офисный пакет Open Office, актуальные версии браузеров Google Chrome (Свободное ПО, <https://google.com/chrome/browser/>), Mozilla Firefox, Бесплатное ПО (GNU GPL), <https://firefox.com/>, Edge, Safari с поддержкой протокола WebRTC, PyCharm 2017.1.2 <https://www.jetbrains.com/pycharm/> Свободное ПО, <https://www.python.org/>, Evolus Pencil, Свободное ПО (GNU GPL 2), <https://pencil.evolus.vn/>

## **VIII. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ**

Дисциплина включает в себя лекционные и практические занятия, а также самостоятельную работу обучающихся.

Организация образовательного процесса по дисциплине осуществляется с использованием системы электронного обучения.

Все лекционные занятия проводятся с визуализацией учебного материала в форме презентаций лекционного материала, которые доступны в системе электронного обучения.

Лекционная часть курса включает следующие компоненты системы знаний учебной дисциплины: понятийный аппарат (тезаурус курса), теоретические утверждения, разъяснения и комментарии; междисциплинарные точки зрения; описание рассматриваемых разделов; ретроспективный и перспективный взгляды на изучаемую проблематику.

Практические занятия по всем модулям дисциплины требуют предварительной теоретической подготовки по соответствующим темам: проработка лекционного материала, ознакомление и изучение отдельных источников основной и дополнительной литературы.

Лекционные и практические занятия могут проводиться с применением дистанционных образовательных технологий с использованием платформ Microsoft Teams, Cisco, Moodle (BigBlueButton) и др.

Проведение лекционных и практических занятий осуществляется с постановкой проблемных вопросов, допускающих возникновение дискуссий, что предполагает активное включение студентов в образовательный процесс.

В организации процесса обучения используются как традиционные, характерные лекционно-семинарской форме обучения, так и инновационные (интерактивные, имитационные, проектные) технологии.

Используемые технологии обеспечивают:

- формирование компетенций, осознанное усвоение знаний, качественное освоение умений их применять и формирование заинтересованного отношения к изучаемым объектам в единстве;

- продуктивность познавательной деятельности, научный поиск, создание субъективно и объективно новых знаний или других продуктов;

- ориентацию на студентов, стимулирование их активности, самостоятельности, инициативы и ответственности;

- контекстный характер обучения, то есть привязку к реальным профессиональным задачам;

- вовлеченность студентов в выполняемую деятельность, возможность проявить и развить свой интеллектуальный, творческий, личностный, деловой потенциал.

Самостоятельная работа направлена на повышение качества обучения, углубление и закрепление знаний студента, развитие аналитических навыков по проблематике учебной дисциплины, активизацию учебно-познавательной деятельности студентов и снижение аудиторной нагрузки.

Максимальное количество баллов по каждому виду контрольных мероприятий указано в учебной карте дисциплины.

## IX. УЧЕБНАЯ КАРТА ДИСЦИПЛИНЫ

**Курс 2, семестр 3, очная форма обучения**

№ п/п	Виды контрольных мероприятий (наименования оценочных средств)	Количество баллов	
		Текущий контроль	Рубежный контроль
<b>Модуль 1. Основы построения и использования систем аналитики больших данных</b>			
1	Практическая работа № 1	20	–
<b>Модуль 2. Разработка систем аналитики больших данных</b>			
2	Практическая работа № 2	20	–
3	Практическая работа № 3	20	–
<b>Всего</b>		<b>60</b>	–
Бонусные баллы		<b>Не предусмотрены</b>	
<b>Промежуточная аттестация в форме экзамена</b>		<p style="text-align: center;"><b>40 баллов</b></p> <p>Экзамен считается сданным при получении не менее 22 баллов, для допуска к экзамену необходимо набрать не менее 38 баллов по сумме текущего и рубежного контроля.</p> <p>Оценка по дисциплине выставляется по сумме баллов за текущий контроль, рубежный контроль и экзамен:</p> <ul style="list-style-type: none"> <li>– 85–100 баллов – оценка «отлично»;</li> <li>– 71–84 балла – оценка «хорошо»;</li> <li>– 60–70 баллов – оценка «удовлетворительно»;</li> <li>– менее 60 баллов – оценка «неудовлетворительно»</li> </ul>	



## Х. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

### 10.1. Паспорт фонда оценочных средств

№ п/п	Индикатор достижения компетенции	Наименование оценочного средства
1	ПК-2.2. Осуществляет руководство созданием комплексных систем искусственного интеллекта с применением новых методов и алгоритмов машинного обучения	– практические работы №1-3;
2	ПК-4.1. Руководит проектами по построению комплексных систем на основе аналитики больших данных в различных отраслях со стороны заказчика	– практические работы №1-3;

### 10.2. Практическая работа №1

#### *Развертывание облачного хранилища общего назначения*

Облачное хранилище Microsoft Azure Storage состоит из четырех сервисов: BLOB Storage, Queue Storage, Table Storage и File Storage. Непосредственно хранение информации осуществляется в сервисах BLOB, Table и File Storage. Queue Storage — это облачный сервис обмена сообщениями и синхронизации распределенных приложений. Сервис Table Storage — база данных NoSQL типа «ключ — значение».

1. Разверните облачное хранилище Microsoft Azure Storage. Получите доступ к панели мониторинга и управления. Организуйте работу с хранилищем BLOB на веб-портале.

В отличие от Azure у AWS хранилища не объединяются логически в Storage Account, а представляют собой набор отдельных сервисов. Хранение двоичных объектов обеспечивается сервисом AWS S3 – Amazon Simple Storage Service. Экземпляр сервиса AWS S3 может быть логически разбит на бакеты, которые будут определять видимость объектов в них, шифрование, жизненный цикл и т.д.

Разверните сервис AWS S2. Создайте бакеты и настройте права доступа к ним. Разверните сервис облачного сетевого хранилища Amazon Elastic File. Разверните сервис AWS EBS (Elastic Block Storage) для создания, размещения, шифрования, подключения к виртуальным машинам, создания бэкапов в виде мгновенных снимков (snapshots).

Сравните возможности и ограничения облачных хранилищ общего назначения Azure BlobStorage и AWS S3 для решения прикладных задач.

#### **Критерии оценки:**

17-20 баллов выставляется студенту, если он своевременно выполнил все задачи, предусмотренные в практической работе, подготовил отчет в соответствии с требованиями преподавателя и в процессе защиты продемонстрировал наличие теоретических знаний в объеме содержания учебной дисциплины, относящейся к практической работе. Сумел ответить на дополнительные вопросы, связанные не только с процессом выполнения практической работы, но и с пониманием совершенных действий и решенных задач.

14-16 баллов выставляется студенту, если он выполнил от 70% задач, предусмотренных в практической работе, подготовил отчет в соответствии с требованиями преподавателя и в процессе защиты продемонстрировал наличие теоретических знаний в объеме содержания учебной дисциплины, относящейся к практической работе. Сумел ответить на вопросы, связанные с процессом выполнения практической работы.

12-13 баллов выставляется студенту, если он более чем на 50% выполнил поставленные в практической работе задачи, способен ответить на вопросы, касающиеся теоретической составляющей в объеме содержания учебной дисциплины, относящейся к практической работе.

0 баллов выставляется студенту, если он более чем на 50% не выполнил поставленные в практической работе задачи, не способен ответить на вопросы, касающиеся теоретической составляющей в объеме содержания учебной дисциплины, относящейся к практической работе.

### **10.3. Практическая работа №2**

#### *Развертывание облачных сервисов копирования и трансформации данных*

Сервисы трансформации и копирования играют важную роль в построении информационных систем анализа больших данных. Требования к сервисам хранения и анализа противоречивы и не могут быть удовлетворены в рамках одного хранилища, особенно когда данных много. Чтобы удовлетворить эти требования, необходимо использовать хранилища разных типов и сервис, который связывает их информационно.

Сервис AWS Glue представляет собой удобное средство каталогизации, трансформации и копирования данных, что позволяет одновременно создавать каталоги данных и использовать их для ETL.

Задание:

1. На веб-портале Azure создайте и сконфигурируйте сервис Azure Data Factory. Предложите решение задачи копирования данных из таблицы Azure Table Storage в реляционную БД Azure SQL.

2. Разверните сервис AWS Data Pipeline для автоматизации копирования и трансформации данных.

3. Разверните бессерверный сервис AWS Glue для категоризации данных в каталог данных и выполнения задач ETL. Предложите решение задачи копирования в AWS Redshift данных в виде файла JSON, хранящегося в AWS S3

#### **Критерии оценки:**

17-20 баллов выставляется студенту, если он своевременно выполнил все задачи, предусмотренные в практической работе, подготовил отчет в соответствии с требованиями преподавателя и в процессе защиты продемонстрировал наличие теоретических знаний в объеме содержания учебной дисциплины, относящейся к практической работе. Сумел ответить на дополнительные вопросы, связанные не только с процессом выполнения практической работы, но и с пониманием совершенных действий и решенных задач.

14-16 баллов выставляется студенту, если он выполнил от 70% задач, предусмотренных в практической работе, подготовил отчет в соответствии с требованиями преподавателя и в процессе защиты продемонстрировал наличие теоретических знаний в объеме содержания учебной дисциплины, относящейся к практической работе. Сумел ответить на вопросы, связанные с процессом выполнения практической работы.

12-13 баллов выставляется студенту, если он более чем на 50% выполнил поставленные в практической работе задачи, способен ответить на вопросы, касающиеся теоретической составляющей в объеме содержания учебной дисциплины, относящейся к практической работе.

0 баллов выставляется студенту, если он более чем на 50% не выполнил поставленные в практической работе задачи, не способен ответить на вопросы, касающиеся теоретической составляющей в объеме содержания учебной дисциплины, относящейся к практической работе.

### **10.4. Практическая работа №3**

#### *Развертывание облачных сервисов для потокового анализа данных*

Потоковый анализ состоит в анализе потока в виде последовательности сообщений, то есть в применении алгоритма анализа к каждому сообщению потока. Этот анализ может состоять в выделении из всего потока сообщений, соответствующих определенным признакам, например, сообщений об ошибках, которые должны быть направлены в отдельное хранилище или сервис, занимающийся обработкой ошибок. Пример прикладной задачи – анализ банковских транзакций и определение в этом потоке потенциально подозрительных (выдача большой суммы в нетипичной для плательщика стране или выдача мелких сумм в разных банкоматах в течение малого интервала времени).

1. Разверните облачный сервис Azure Stream Analytics для потоковой обработки Microsoft-сообщений. Создайте экземпляр потокового задания на веб-портале. Обеспечьте интеграцию с сервисом Azure ML: для этого обучите модель с помощью сервиса Azure ML, а затем, после ее размещения в качестве REST-сервиса, получите ее адрес и используйте для интеграции с Azure Stream Analytics.

2. Разверните сервис потокового анализа данных AWS Kinesis Analytics. Подключите входной источник. Определите схему сообщений, поступающих на вход сервиса. Сконфигурируйте приложение.

Сравните возможности и ограничения облачных сервисов для потокового анализа данных Azure Stream Analytics и AWS Kinesis Analytics для решения прикладных задач.

### **Критерии оценки:**

17-20 баллов выставляется студенту, если он своевременно выполнил все задачи, предусмотренные в практической работе, подготовил отчет в соответствии с требованиями преподавателя и в процессе защиты продемонстрировал наличие теоретических знаний в объеме содержания учебной дисциплины, относящейся к практической работе. Сумел ответить на дополнительные вопросы, связанные не только с процессом выполнения практической работы, но и с пониманием совершенных действий и решенных задач.

14-16 баллов выставляется студенту, если он выполнил от 70% задач, предусмотренных в практической работе, подготовил отчет в соответствии с требованиями преподавателя и в процессе защиты продемонстрировал наличие теоретических знаний в объеме содержания учебной дисциплины, относящейся к практической работе. Сумел ответить на вопросы, связанные с процессом выполнения практической работы.

12-13 баллов выставляется студенту, если он более чем на 50% выполнил поставленные в практической работе задачи, способен ответить на вопросы, касающиеся теоретической составляющей в объеме содержания учебной дисциплины, относящейся к практической работе.

0 баллов выставляется студенту, если он более чем на 50% не выполнил поставленные в практической работе задачи, не способен ответить на вопросы, касающиеся теоретической составляющей в объеме содержания учебной дисциплины, относящейся к практической работе.

### **Методические рекомендации по выполнению практических работ**

Целью практических работ является приобретение практических навыков использования математических моделей, методов и алгоритмов в области технологий анализа больших данных; усвоение полученных знаний студентами, а также формирование у них мотивации к самообразованию за счет активизации самостоятельной познавательной деятельности.

Все работы выполняются студентами в рамках 4-х академических часов, которые отведены учебным планом.

Итогом работы является защита полученных результатов. Защита проводится индивидуально в форме собеседования и проверки полученных навыков работы с системой на компьютере.

#### 10.4. Экзаменационные вопросы и билеты

1. Хранилища общего назначения. Форматы хранения данных. Облачное хранилище Microsoft Azure Storage. Облачное хранилище AWS.
2. Реляционные базы данных. Azure SQL. AWS RDS.
3. Нереляционные базы данных. Сервисы нереляционных баз данных от Azure и AWS.
4. Корпоративные хранилища данных (DWH). Azure SQL DWH. AWS RedShift.
5. Хранилища данных типа Data Lake. Azure Data Lake Store. AWS Data Lake Solutions.
6. Прямая загрузка данных. Доставка данных в облачное хранилище общего назначения.
7. Прямая загрузка данных. Доставка данных в реляционные БД и хранилища.
8. Прямая загрузка данных. Доставка данных в нереляционные базы данных.
9. Прямая загрузка данных. Доставка данных в HDFS-совместимые хранилища.
10. Прямая загрузка потоковых данных. Azure Event Hub. AWS Kinesis Data Streams. Облачные сервисы развертывания кластерных систем.
11. Облачные сервисы копирования и трансформации данных. Azure Data Factory.
12. Облачные сервисы копирования и трансформации данных. AWS Data Pipeline. AWS Glue.
13. Интерактивный анализ данных. Анализ реляционных данных. Azure Data Lake Analytics. AWS Athena. Apache Spark. Встроенные редакторы запросов сервиса CosmosDB.
14. Потоковый анализ данных. Общие сведения. Azure Stream Analytics. Amazon Kinesis Analytics. Apache Storm.
15. Пакетный анализ данных. Hadoop. Apache Pig. Apache Hive.

#### ОБРАЗЕЦ ЭКЗАМЕНАЦИОННОГО БИЛЕТА

Направление/специальность 09.04.03 Прикладная информатика

1. Прямая загрузка потоковых данных. Azure Event Hub. AWS Kinesis Data Streams. Облачные сервисы развертывания кластерных систем.
2. Хранилища общего назначения. Форматы хранения данных. Облачное хранилище Microsoft Azure Storage. Облачное хранилище AWS.

« \_\_\_\_ » \_\_\_\_\_ 20 г.

#### Критерии оценивания:

**40-34 балла** выставляется студенту, если:

- ответ на вопросы экзаменационного билета дан в полном объеме;
- ответ характеризуется связностью, логичностью, достаточным объёмом, с приведением примеров и объяснением алгоритмов выполнения операций;
- допустимы единичные ошибки, которые исправлялись в процессе ответа самим экзаменуемым.

**33-28 баллов** выставляется студенту, если:

- экзаменационное задание в основном выполнено;
- ответ характеризуется достаточной связностью и логичностью, но небольшим объёмом изложенного материала;
- допустимы единичные ошибки, которые исправлялись экзаменаторами в процессе ответа и отсутствие практических примеров.

**27-22 балла** выставляется студенту, если:

- экзаменационное задание выполнено частично.

- ответ характеризуется недостаточной связностью, логичностью, небольшим объёмом изложенного материала;
- при ответе студент допускает ошибки, которые не в состоянии самостоятельно исправить.

**0 баллов** выставляется студенту, если:

- экзаменационное задание выполнено фрагментарно;
- в ответе отсутствует связность и логичность;
- при ответе студент допускает грубые ошибки, которые не в состоянии исправить самостоятельно.