

Министерство образования и науки Астраханской области  
Государственное автономное образовательное учреждение  
Астраханской области высшего образования  
«Астраханский государственный архитектурно-строительный  
университет»  
(ГАОУ АО ВО «АГАСУ»)



УТВЕРЖДАЮ

Проректор

И. И. Игнатова /

И. И. Игнатова /

Ф. И. О.

сентябрь 2021 г.

## РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Наименование дисциплины

Системы обработки больших данных

(указывается наименование в соответствии с учебным планом)

По направлению подготовки

09.04.02 «Информационные системы и технологии»

(указывается наименование направления подготовки в соответствии с ФГОС)

Программа Искусственный интеллект в проектировании и производстве

Направленность (профиль)

"Искусственный интеллект в проектировании городской среды"

(указывается наименование профиля в соответствии с ОПОП)

Кафедра

Системы автоматизированного проектирования и моделирования

Квалификация выпускника *магистр*

Астрахань – 2021



## Оглавление

1. Цель освоения дисциплины .....	4
2. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы .....	4
3. Место дисциплины в структуре ОПОП магистратуры .....	4
4. Объем дисциплины в зачетных единицах с указанием количества академических часов, выделенных на контактную работу обучающихся с преподавателем (по типам занятий) и на самостоятельную работу обучающихся.....	5
5. Содержание дисциплины, структурированное по разделам с указанием отведенного на них количества академических часов и типов учебных занятий .....	6
5.1. Разделы дисциплины и трудоемкость по типам учебных занятий и работы обучающихся (в академических часах) .....	6
5.1.1. Очная форма обучения .....	6
5.2. Содержание дисциплины, структурированное по разделам .....	7
5.2.1. Содержание лекционных занятий .....	7
5.2.2. Содержание лабораторных занятий .....	8
5.2.3. Содержание практических занятий .....	8
5.2.4. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине .....	8
5.2.5. Темы контрольных работ .....	9
5.2.6. Темы курсовых проектов/ курсовых работ .....	9
6. Методические указания для обучающихся по освоению дисциплины .....	9
7. Образовательные технологии .....	10
8. Учебно-методическое и информационное обеспечение дисциплины.....	10
8.1. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины .....	10
8.2. Перечень необходимого лицензионного и свободно распространяемого программного обеспечения, в том числе отечественного производства, используемого при осуществлении образовательного процесса по дисциплине .....	11
8.3. Перечень современных профессиональных баз данных и информационных справочных систем, доступных обучающимся при освоении дисциплины .....	11
9. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине .....	12
10. Особенности организации обучения по дисциплине для инвалидов и лиц с ограниченными возможностями здоровья .....	12
11. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ .....	13
Лист внесения дополнений и изменений.....	16

## **1. Цель освоения дисциплины**

Целью освоения дисциплины «Системы обработки больших данных» является формирование компетенций обучающихся в соответствии с требованиями Федерального государственного образовательного стандарта высшего образования по направлению подготовки 09.04.02 «Информационные системы и технологии»

## **2. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы**

В результате освоения дисциплины обучающийся должен овладеть следующими компетенциями:

ОПК-3. Способен анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с обоснованными выводами и рекомендациями

ОПК-3.1. Анализирует профессиональную информацию, выделяет в ней главное, структурирует, оформляет и представляет в виде аналитических обзоров с обоснованными выводами и рекомендациями.

ОПК-ЗИИП. Способен исследовать современные проблемы и методы информатики, искусственного интеллекта и развития информационного общества, цифровой экономики

ОПК-ЗИИП.1 Исследует современные проблемы информатики, искусственного интеллекта и развития информационного общества, цифровой экономики

ОПК-ЗИИП.2 Проводит анализ современных методов и средств информатики и искусственного интеллекта для решения задач профессиональной деятельности

В результате освоения дисциплин, обучающийся должен овладеть следующими результатами обучения по дисциплине:

ОПК-3. З-1. Знает принципы, методы и средства анализа и структурирования профессиональной информации

ОПК-3. У-1. Умеет анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров

ОПК-3. В-1. Иметь навыки подготовки научных докладов, публикаций и аналитических обзоров с обоснованными выводами и рекомендациями

ОПК-ЗИИП.1 З-1. 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

ОПК-ЗИИП.1 У-1. Умеет применять при решении задач профессиональной деятельности критерии эффективности функционирования информационного общества и цифровой экономики; структуру интеллектуального капитала, методы оценки эффективности

ОПК-ЗИИП.2 З-1. Знает состав современных методов и средств информатики, передовые методы искусственного интеллекта для решения задач профессиональной деятельности

ОПК-ЗИИП.2 У-1. Умеет проводить анализ современных методов и средств информатики и искусственного интеллекта для решения прикладных задач различных классов

## **3. Место дисциплины в структуре ОПОП магистратуры**

Дисциплина Б1.О.6 «Системы обработки больших данных» реализуется в рамках Блока 1 «Дисциплины (модули)» обязательной части.

Дисциплина базируется на основах обучения, полученных в рамках изучения дисциплины: «Специальные главы математики», «Технологии программирования»,

«Модели информационных процессов и систем», Технологии проектирования информационных систем и искусственного интеллекта».

**4. Объем дисциплины в зачетных единицах с указанием количества академических часов, выделенных на контактную работу обучающихся с преподавателем (по типам занятий) и на самостоятельную работу обучающихся**

Форма обучения	Очная
1	2
Трудоемкость в зачетных единицах:	4 семестр – 3 з.е.; всего - 3 з.е.
Лекции (Л)	4 семестр – 12 часов; всего -12 часа
Лабораторные занятия (ЛЗ)	учебным планом не предусмотрены
Практические занятия (ПЗ)	4 семестр – 24 часа; всего – 24 часа
Самостоятельная работа (СР)	4 семестр – 72 часа; всего - 72 часа
Форма текущего контроля:	
Контрольная работа	учебным планом не предусмотрены
Форма промежуточной аттестации:	
Экзамен	4 семестр
Зачет	учебным планом не предусмотрен
Курсовая работа	учебным планом не предусмотрен
Курсовой проект	учебным планом не предусмотрен

**5. Содержание дисциплины, структурированное по разделам с указанием отведенного на них количества академических часов и типов учебных занятий**

**5.1. Разделы дисциплины и трудоемкость по типам учебных занятий и работы обучающихся (в академических часах)**

**5.1.1. Очная форма обучения**

№ п/п	Раздел дисциплины. (по семестрам)	Всего часов на раздел	Семестр	Распределение трудоемкости раздела (в часах) по типам учебных занятий и работы обучающихся			Форма текущего контроля и промежуточной аттестации	
				контактная				СР
				Л	ЛЗ	ПЗ		
1	2	3	4	5	6	7	8	9
1.	Модуль 1. Большие данные и методы их обработки	21	4	2		4	15	Экзамен
2.	Модуль 2. Технологии обработки больших данных	21	4	2		4	15	
3.	Модуль 3. Технологии data mining	32	4	4		8	20	
4.	Модуль 4. Интеллектуальный анализ текстовой информации с использованием машинного обучения	34	4	4		8	22	
Итого:		108		12		24	72	

## 5.2. Содержание дисциплины, структурированное по разделам

### 5.2.1. Содержание лекционных занятий

№	Наименование раздела дисциплины	Содержание
1	2	3
1.	Модуль 1. Большие данные и методы их обработки	Цели, задачи и структура курса. Большие данные и информация. Способы представления, хранения и обработки. Задачи классификации, индексирования и поиска.
2.	Модуль 2. Технологии обработки больших данных	HDFS (hadoop distributed file system) системы для работы с большими данными. Hadoop-технология. Среда обработки RStudio. Библиотеки для работы с Big Data.
3.	Модуль 3. Технологии data mining	Статистическая обработка, сжатие, кластеризация, факторный анализ, парсинг и визуализация данных.
4.	Модуль 4. Интеллектуальный анализ текстовой информации с использованием машинного обучения	Принципы и методы семантико-синтаксического анализа естественно-языковых текстов. Методы кластеризации текстов. Методы автоматизации реферирования (аннотирования) текстов. Методы извлечения фрагментов для аннотации.

### 5.2.2. Содержание лабораторных занятий

Учебным планом не предусмотрены

### 5.2.3. Содержание практических занятий

№	Наименование раздела дисциплины	Содержание
1.	Модуль 1. Большие данные и методы их обработки	Способы представления, хранения и обработки. Задачи классификации, индексирования и поиска.
2.	Модуль 2. Технологии обработки больших данных	Исследование пакета обработки Hadoop Изучение парадигмы Map Reduce. Подсчет кол-ва слов, реализация алгоритма k-means в рамках парадигмы Map Reduce с использованием Hadoop.
3.	Модуль 3. Технологии data mining	Знакомство с языком программирования для статистической обработки данных в среде R-Studio (установка и настройка). Подготовка статистических данных в текстовом редакторе. Импорт данных из текстового файла. Построение графиков в среде R-Studio. Вычисление общих характеристик выборки данных. Оценка общей тенденции, выявление ошибок в данных и пропусков. Анализ временного ряда на предмет прогнозирования будущего с помощью R-Studio. Тренд и период колебаний. Анализ связей двух выборок данных с помощью R-Studio. Анализ корреляций, регрессионный анализ.
4.	Модуль 4. Интеллектуальный анализ текстовой информации с использованием машинного обучения	Использование инструментов Data Mining Client для Excel 2007 в подготовке данных Изучение методов семантического анализа. Исследование методов машинного обучения.

### 5.2.4. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине

Очная форма обучения

№	Наименование раздела дисциплины	Содержание	Учебно-методическое обеспечение
1	2	3	4
1.	Модуль 1. Большие данные и методы их обработки	Подготовка к практическому занятию №1 Подготовка к экзамену	[1-6], [9], [10]
2.	Модуль 2. Технологии обработки больших данных	Подготовка к практическому занятию №2. Чтение публикаций о распределенных хранилищах данных, их особенностях и принципах построения распределенных файловых систем. Подготовка к экзамену	[1-6], [9], [10]
3.	Модуль 3. Технологии data mining	Подготовка к практическому занятию №3 Чтение публикаций по темам: Проверка статистических гипотез Исследование методов факторного анализа. Кластеризация и визуализация данных. Подготовка к экзамену	[1-6], [9], [10]
4.	Модуль 4. Интеллектуальный анализ текстовой информации с использованием машинного обучения	Подготовка к практическому занятию №4 Чтение публикаций по темам: Методы семантического анализа. Методы машинного обучения Подготовка к экзамену.	[5], [7-12]



### 5.2.5. Темы контрольных работ

Учебным планом не предусмотрены.

### 5.2.6. Темы курсовых проектов/ курсовых работ

Учебным планом не предусмотрены.

## 6. Методические указания для обучающихся по освоению дисциплины

### Организация деятельности студента

#### Лекция.

В ходе лекционных занятий необходимо вести конспектирование учебного материала, обращать внимание на категории, формулировки, раскрывающие содержание тех или иных явлений и процессов, научные выводы и практические рекомендации. Необходимо задавать преподавателю уточняющие вопросы с целью уяснения теоретических положений, разрешения спорных ситуаций. Целесообразно дорабатывать свой конспект лекции, делая в нем соответствующие записи из литературы, рекомендованной преподавателем и предусмотренной учебной программой.

#### Практическое занятие.

Работа с конспектом лекций, подготовка ответов к контрольным вопросам, просмотр рекомендуемой литературы. Особое внимание при этом необходимо обратить на содержание основных положений и выводов, объяснение явлений и фактов, уяснение практического приложения рассматриваемых теоретических вопросов. Решение расчетно-графических заданий, решение задач по алгоритму и др.

По дисциплине «Логика и методология науки» практические занятия проводятся с использованием следующей интерактивной технологии: Работа в малых группах - это одна из самых популярных стратегий, так как она дает всем обучающимся (в том числе и стеснительным) возможность участвовать в работе, практиковать навыки сотрудничества, межличностного общения (в частности, умение активно слушать, вырабатывать общее мнение, разрешать возникающие разногласия). Все это часто бывает невозможно в большом коллективе.

#### Самостоятельная работа.

Самостоятельная работа студента над усвоением учебного материала по учебной дисциплине может выполняться в помещениях для самостоятельных работ, а также в домашних условиях. Содержание самостоятельной работы студента определяется учебной программой дисциплины, методическими материалами, заданиями и указаниями преподавателя.

Самостоятельная работа в аудиторное время может включать:

- конспектирование (составление тезисов) лекций;
- выполнение контрольных работ; решение задач;
- участие в тестировании.

Самостоятельная работа во внеаудиторное время может состоять из:

- повторение лекционного материала;
- изучения учебной и научной литературы;
- решения задач, выданных на практических занятиях;
- подготовки к контрольным работам, тестированию.

Подготовка студентов к экзамену включает три стадии:

- самостоятельная работа в течение семестра;
- непосредственная подготовка в дни, предшествующие экзамену.
- экзамен организован в виде электронного тестирования на портале ЭОС АГАСУ.

## 7. Образовательные технологии

По дисциплине «Анализ больших данных» лекционные занятия проводятся с использованием следующих образовательных технологий

### Традиционные образовательные технологии

Перечень образовательных технологий, используемых при изучении дисциплины «Логика и методология науки»:

Лекция - последовательное изложение материала в дисциплинарной логике, осуществляемое преимущественно вербальными средствами (монолог преподавателя).

Практическое занятие - занятие, посвященное освоению конкретных умений и навыков по предложенному алгоритму.

Интерактивные технологии - организация образовательного процесса, которая предполагает активное и нелинейное взаимодействие всех участников, достижение на этой основе лично значимого для них образовательного результата. Наряду со специализированными технологиями такого рода принцип интерактивности прослеживается в большинстве современных образовательных технологий. Интерактивность подразумевает субъект-субъектные отношения в ходе образовательного процесса и, как следствие, формирование саморазвивающейся информационно-ресурсной среды.

Интерактивные образовательные технологии:

Лекция-визуализация - представляет собой визуальную форму подачи лекционного материала средствами ТСО или аудиовидеотехники (видео-лекция). Чтение такой лекции сводится к развернутому или краткому комментированию просматриваемых визуальных материалов (в виде схем, таблиц, графов, графиков, моделей). Лекция-визуализация помогает студентам преобразовывать лекционный материал в визуальную форму, что способствует формированию у них профессионального мышления за счет систематизации и выделения наиболее значимых, существенных элементов.

Практические занятия проводятся с использованием интерактивной технологии: Работа в малых группах - это одна из самых популярных стратегий, так как она дает всем обучающимся (в том числе и стеснительным) возможность участвовать в работе, практиковать навыки сотрудничества, межличностного общения (в частности, умение активно слушать, вырабатывать общее мнение, разрешать возникающие разногласия).

## 8. Учебно-методическое и информационное обеспечение дисциплины

### 8.1. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины

а) основная учебная литература:

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

2. Адлер Ю.П. Статистическое управление процессами. «Большие данные» : учебное пособие / Адлер Ю.П., Черных Е.А.. — Москва : Издательский Дом МИСиС, 2016. — 52 с. — ISBN 978-5-87623-969-3. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <https://www.iprbookshop.ru/64199.html>

3. Пальмов С.В. Интеллектуальный анализ данных [Электронный ресурс]: учеб. пособие. – Самара: Поволжский государственный университет телекоммуникаций и информатики, 2017. – 127 с. – Режим доступа <http://www.iprbookshop.ru/75376>

4. Минаева Ю.В. Методы статистического и интеллектуального анализа данных: учеб. пособие. – Воронеж: ВГТУ, 2017. 90 с.

5. Чубукова И.А. Data Mining : учебное пособие / Чубукова И.А.. — Москва, Саратов : Интернет-Университет Информационных Технологий (ИНТУИТ), Ай Пи Ар Медиа, 2020. — 469 с. — ISBN 978-5-4497-0289-0. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <https://www.iprbookshop.ru/89404.html>

б) дополнительная учебная литература:

6. Новикова Н.М. Обработка экспериментальных данных: Учеб. пособие. – Воронеж: ВГТУ, 2010. – 119 с

7. Жуковский, О.И. Информационные технологии и анализ данных : учебное пособие / О.И. Жуковский ; Министерство образования и науки Российской Федерации, Томский Государственный Университет Систем Управления и Радиоэлектроники (ТУСУР). - Томск : Эль Контент, 2014. - 130 с. : схем., ил. - Библиогр.: с. 126. - - URL: <http://biblioclub.ru/index.php?page=book&id=480500>

в) перечень онлайн курсов

8. Введение в инженерии больших данных, НПОО, МИСиС, URL: <https://openedu.ru/course/misis/ABD/>

9. Наука о данных и аналитика больших объемов данных, НПОО, Политех, URL: <https://openedu.ru/course/spbstu/BIGDATA/>

## **8.2. Перечень необходимого лицензионного и свободно распространяемого программного обеспечения, в том числе отечественного производства, используемого при осуществлении образовательного процесса по дисциплине**

1. 7-Zip GNU Открытое лицензионное соглашение GNU GeneralPublicLicense Бессрочно
2. Office 365 A1 Академическая подписка. Бессрочно.
3. Adobe Acrobat Reader DC. Открытое лицензионное соглашение GNU GeneralPublicLicense.
4. Internet Explorer. Предоставляется в рамках Microsoft Azure Dev Tools for Teaching
5. Apache Open Office. Apache license 2.0 Бессрочно.
6. Google Chrome Бесплатное программное обеспечение. Бессрочно
7. VLC media player GNU Lesser General Public License, version 2.1 or later. Бессрочно
8. Azure Dev Tools for Teaching
9. Kaspersky Endpoint Security. Лицензия действует до 16.03.2022
10. MathcadEducation - UniversityEdition.
11. R Studio.

## **8.3. Перечень современных профессиональных баз данных и информационных справочных систем, доступных обучающимся при освоении дисциплины**

1. Электронная информационно-образовательная среда Университета: (<http://edu.aucu.ru>, <http://moodle.aucu.ru>).
2. Электронно-библиотечная система «Университетская библиотека» (<https://biblioclub.ru/>).

3. Электронно-библиотечная система «IPRbooks» ([www.iprbookshop.ru](http://www.iprbookshop.ru)).
4. Научная электронная библиотека (<http://www.elibrary.ru/>).
5. Консультант + (<http://www.consultant-urist.ru/>).
6. Федеральный институт промышленной собственности (<https://www1.fips.ru/>).
7. Патентная база USPTO (<https://www.uspto.gov/patents-application-process/search-patents>).

**9. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине**

№ п/п	Наименование специальных помещений и помещений для самостоятельной работы	Оснащенность специальных помещений и помещений для самостоятельной работы
1.	Учебные аудитории для проведения учебных занятий:  414056, г. Астрахань, ул. Татищева, 18, аудитории № 204, 4.	<p style="text-align: center;"><b>№ 204</b></p> Комплект учебной мебели. Стационарный мультимедийный комплект. Доступ к информационно – телекоммуникационной сети «Интернет».
2.	Помещения для самостоятельной работы:  414056, г. Астрахань, ул. Татищева, 22а, аудитории № 201, 203.	<p style="text-align: center;"><b>№ 4</b></p> Комплект учебной мебели. Переносной мультимедийный комплект. Доступ к информационно – телекоммуникационной сети «Интернет».
		<p style="text-align: center;"><b>№ 201</b></p> Комплект учебной мебели. Компьютеры – 8 шт. Доступ к информационно – телекоммуникационной сети «Интернет».
		<p style="text-align: center;"><b>№ 203</b></p> Комплект учебной мебели. Компьютеры – 8 шт. Доступ к информационно – телекоммуникационной сети «Интернет».

**10. Особенности организации обучения по дисциплине для инвалидов и лиц с ограниченными возможностями здоровья**

Для обучающихся из числа инвалидов и лиц с ограниченными возможностями здоровья на основании письменного заявления дисциплина «Системы обработки больших данных» реализуется с учетом особенностей психофизического развития, индивидуальных возможностей и состояния здоровья (далее – индивидуальных особенностей).

## **Аннотация**

к рабочей программе дисциплины «Системы обработки больших данных»  
**По направлению подготовки 09.04.02 «Информационные системы и технологии»**  
**Программа Искусственный интеллект в проектировании и производстве**  
**Направленность (профиль) "Искусственный интеллект в проектировании городской среды"**

Общая трудоемкость дисциплины составляет 3 зачетных единиц.

Форма промежуточной аттестации: зачет.

Целью учебной дисциплины «Системы обработки больших данных» является формирование компетенций обучающихся в соответствии с требованиями Федерального государственного образовательного стандарта высшего образования по направлению подготовки 09.04.02 «Информационные системы и технологии».

Дисциплина Б1.О.6 «Системы обработки больших данных» реализуется в рамках Блока 1 «Дисциплины (модули)» обязательной части.

Дисциплина базируется на основах обучения, полученных в рамках изучения дисциплины: «Специальные главы математики», «Технологии программирования», «Модели информационных процессов и систем», Технологии проектирования информационных систем и искусственного интеллекта».

Краткое содержание дисциплины:

Модуль 1. Большие данные и методы их обработки

Модуль 2. Технологии обработки больших данных

Модуль 3. Технологии data mining

Модуль 4. Интеллектуальный анализ текстовой информации с использованием машинного обучения

Лист внесения дополнений и изменений  
в рабочую программу учебной дисциплины

Системы обработки больших данных  
(наименование дисциплины)

на 2022 - 2023 учебный год

Рабочая программа пересмотрена на заседании кафедры «Систем автоматизированного проектирования и моделирования»,  
протокол № 9 от 18.04. 2022 г.

Зав. кафедрой  
К.Т.Н., доцент  
ученая степень, ученое звание

  
\_\_\_\_\_ подпись

/О.И. Евдошенко/  
И.О. Фамилия

В рабочую программу вносятся следующие изменения:

1. В п.5.2.3. внесены следующие изменения:

**5.2.3. Содержание практических занятий**

№	Наименование раздела дисциплины	Содержание
1.	Модуль 1. Большие данные и методы их обработки	Входное тестирование по дисциплине. Способы представления, хранения и обработки. Задачи классификация, индексирования и поиска.

Составители изменений и дополнений:

К.Т.Н., доцент  
ученая степень, ученое звание


  
\_\_\_\_\_ подпись

/ М.И. Шиккульский/  
И.О. Фамилия

Председатель МКН «Информационные системы и технологии» направленность (профиль)

«Искусственный интеллект в проектировании городской среды»

К.Т.Н., доцент  
ученая степень, ученое звание

  
\_\_\_\_\_ подпись

/ О.И. Евдошенко/  
И.О. Фамилия

«18» апреля 2022г.

## 11. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

### 11.1. Контрольные вопросы и задания

Оценочные средства планируемых результатов обучения представлены в виде фондов оценочных средств (ФОС), разработанных в соответствии с локальным нормативным актом университета. В целях освоения компетенций, указанных в рабочей программе дисциплины, предусмотрены следующие вопросы, задания текущего контроля:

#### База контрольных вопросов для экзамена

**ОПК-3. Способен анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с обоснованными выводами и рекомендациями**

**ОПК-3.1. Анализирует профессиональную информацию, выделяет в ней главное, структурирует, оформляет и представляет в виде аналитических обзоров с обоснованными выводами и рекомендациями**

ОПК-3. З-1. Знает принципы, методы и средства анализа и структурирования профессиональной информации

ОПК-3. У-1. Умеет анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров

ОПК-3. В-1. Иметь навыки подготовки научных докладов, публикаций и аналитических обзоров с обоснованными выводами и рекомендациями

#### Вопросы:

1. Что означает термин «Big Data» в информационных технологиях?
2. Что является основной целью обработки Big Data?
3. Кто и в каком году впервые ввел термин «Big Data»?
4. Какие главные характеристики Big Data?
5. Чем характеризуются "Большие данные"?
6. Какие цели ставит перед собой Data Science?
7. Что такое жизненный цикл аналитики данных?
8. Чем анализ больших данных отличается от традиционного анализа?
9. Какая основная цель статистического анализа?
10. Что такое генеральная совокупность?
11. Корреляция и регрессионный анализ. Коэффициент корреляции. Графическое представление. Постановка задачи регрессионного анализа.
12. Пояснить термин "Линейная регрессия". Привести примеры использования регрессионного анализа.
13. Что подразумевается под определением "статистический вывод"?
14. Что является результатом решения задачи регрессии?
15. Что такое ошибка первого рода  $\alpha$ -error?
16. Что такое ошибка второго рода  $\beta$ -error?
17. Научные проблемы больших данных. Показать значимость проблем, актуальность, связь с областями математики и инженерии

**ОПК-ЗИИП. Способен исследовать современные проблемы и методы информатики, искусственного интеллекта и развития информационного общества, цифровой экономики**

**ОПК-ЗИИП.1 Исследует современные проблемы информатики, искусственного интеллекта и развития информационного общества, цифровой экономики**

ОПК-ЗИИП.1 З-1. 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

ОПК-ЗИИП.1 У-1. Умеет применять при решении задач профессиональной деятельности критерии эффективности функционирования информационного общества и цифровой экономики; структуру интеллектуального капитала, методы оценки эффективности

#### Вопросы:

18. Что такое Apache Hadoop?
19. В чем преимущества решений на базе Hadoop?
20. Что такое MapReduce?
21. Какими достоинствами и недостатками обладает MapReduce?
22. Какому основному принципу следует HDFS?
23. Какой размер блока по умолчанию в HDFS?
24. Какие функции выполняет NameNode в HDFS?
25. Какой узел отвечает за репликацию данных в Hadoop?

26. Какие компоненты содержит Slave узел в Hadoop?
27. Какие компоненты содержит Master узел в Hadoop?
28. Какие компоненты являются частями HDFS?
29. Какое API было добавлено в Hadoop v2.0?
30. Для чего используется автономный режим Hadoop?
31. Какой режим необходим для того, чтобы на локальной машине использовать Hadoop как кластер, состоящий из одного узла?

**ОПК-ЗИИП.2 Проводит анализ современных методов и средств информатики и искусственного интеллекта для решения задач профессиональной деятельности**

ОПК-ЗИИП.2 3-1. Знает состав современных методов и средств информатики, передовые методы искусственного интеллекта для решения задач профессиональной деятельности

ОПК-ЗИИП.2 У-1. Умеет проводить анализ современных методов и средств информатики и искусственного интеллекта для решения прикладных задач различных классов

**Вопросы:**

32. Для чего нужна визуализация данных?
33. Какие традиционные виды визуализации данных вы знаете?
34. В чем основные задачи визуализации данных?
35. Какие требования предъявляются к визуализации данных?
36. Какие типы визуализации данных можно выделить?
37. Какие основные типы Data Mining?
38. Какие категории Web Mining можно выделить?
39. В чем основная задача Web Content Mining?
40. В чем основные задачи интеллектуального анализа текстов?
41. Что является целью кластеризации?
42. С помощью какого алгоритма можно найти ассоциативное правило?
43. Как оценить качество обучения модели?
44. Классификация и кластеризация – суть и назначение. Метрики. Постановка задачи кластеризации.
45. Методы кластеризации на графах. Отличие от задачи классификации. Привести примеры использования алгоритмов кластеризации.

**Типовые практические задания:**

1. Применение методов машинного обучения к большим данным
2. Применение методов Data Mining к большим данным
3. Разработка архитектуры и прототипа информационной системы обработки больших данных.
4. Применение методов классификации к большим данным
5. Применение нейронных сетей к анализу больших данных
6. Поиск ассоциативных правил в больших данных
7. Исследование и визуализация структуры Facebook, Вконтакте (других социальных сетей)
8. Программные средства анализа «больших данных» с открытым исходным кодом

**11.2 Перечень видов оценочных средств**

Наименование оценочного средства: Экзамен. Средство контроля, организованное в виде электронного тестирования на портале ЭОС АГАСУ и предназначено для выяснения объема знаний обучающегося по предмету.

Наименование оценочного средства: Практические задания. Средство контроля, организованное в виде заданий с последующей защитой выполненной работы на практических занятиях.

**11.3. Показатели и критерии оценивания компетенций, шкалы оценивания**

В рамках изучаемой дисциплины студент может демонстрировать следующие уровни овладения компетенциями.

**Повышенный уровень:** обучающийся демонстрирует глубокое знание учебного материала; способен использовать сведения из различных источников для успешного исследования и поиска решения в нестандартных ситуациях; способен анализировать, проводить сравнение и обоснование выбора методов решения практико-ориентированных заданий. **Оценка промежуточной аттестации (экзамен, зачёт с оценкой):** 5 (отлично) – 91 балл и более.

**Базовый уровень:** обучающийся способен понимать и интерпретировать освоенную информацию; демонстрирует осознанное владение учебным материалом и учебными умениями, навыками и способами деятельности, необходимыми для решения практико-ориентированных заданий. **Оценка промежуточной аттестации (экзамен, зачёт с оценкой):** 4 (хорошо) – 71-90 баллов.

**Пороговый уровень:** обучающийся обладает необходимой системой знаний и владеет некоторыми умениями; демонстрирует самостоятельность в применении знаний, умений и навыков к решению учебных заданий на репродуктивном уровне. **Оценка промежуточной аттестации (экзамен, зачёт с оценкой):** 3 (удовлетворительно) – 60-70 баллов.

**Уровень ниже порогового:** система знаний, необходимая для решения учебных и практико-ориентированных заданий, не сформирована; обучающийся не владеет основными умениями, навыками и способами деятельности. **Оценка промежуточной аттестации (экзамен, зачёт с оценкой):** 2 (неудовлетворительно) – ниже 60 баллов.

В рамках данной дисциплины используются следующие критерии оценки знаний студентов.

**Отлично**

Обучающийся демонстрирует:



- систематизированные, глубокие и полные знания по всем разделам учебной дисциплины, а также по основным вопросам, выходящим за ее пределы;
- точное использование научной терминологии, грамотное, логически правильное изложение ответа на вопросы;
- безупречное владение инструментарием учебной дисциплины, умение его эффективно использовать в постановке и решении научных и профессиональных задач;
- выраженную способность самостоятельно и творчески решать сложные проблемы в нестандартной ситуации;
- полное и глубокое усвоение основной, и дополнительной литературы, по изучаемой учебной дисциплине;
- умение свободно ориентироваться в теориях, концепциях и направлениях по изучаемой учебной дисциплине и давать им аналитическую оценку, использовать научные достижения других дисциплин;
- творческую самостоятельную работу на учебных занятиях, активное творческое участие в групповых обсуждениях, высокий уровень культуры исполнения заданий.

#### **Хорошо**

Обучающийся демонстрирует:

- систематизированные, глубокие и полные знания по всем разделам учебной дисциплины;
- использование научной терминологии, грамотное, логически правильное изложение ответа на вопросы, умение делать обобщенные выводы и обобщения;
- владение инструментарием учебной дисциплины (методами комплексного анализа, техникой информационных технологий), умение его использовать в постановке и решении научных и профессиональных задач;
- способность решать сложные проблемы в рамках учебной дисциплины;
- свободное владение типовыми решениями;
- усвоение основной и дополнительной литературы, рекомендованной рабочей программой по учебной дисциплине;
- умение ориентироваться в теориях, концепциях и направлениях по изучаемой учебной дисциплине и давать им аналитическую оценку;
- активную самостоятельную работу на учебных занятиях, систематическое участие в групповых обсуждениях, высокий уровень культуры исполнения заданий.

#### **Удовлетворительно**

Обучающийся демонстрирует:

- достаточные знания в объеме рабочей программы по учебной дисциплине;
- использование научной терминологии, грамотное, логически правильное изложение ответа на вопросы, умение делать выводы без существенных ошибок;
- владение инструментарием учебной дисциплины, умение его использовать в решении учебных и профессиональных задач;
- способность самостоятельно применять типовые решения в рамках изучаемой дисциплины;
- усвоение основной литературы, рекомендованной рабочей программой по дисциплине;
- умение ориентироваться в базовых теориях, концепциях и направлениях по дисциплине;
- работу на учебных занятиях под руководством преподавателя, фрагментарное участие в групповых обсуждениях, достаточный уровень культуры исполнения заданий.

#### **Неудовлетворительно**

Обучающийся демонстрирует:

- фрагментарные знания в рамках изучаемой дисциплины; знания отдельных литературных источников, рекомендованных рабочей программой по учебной дисциплине;
- неумение использовать научную терминологию учебной дисциплины, наличие в ответе грубых, логических ошибок;
- пассивность на занятиях или отказ от ответа, низкий уровень культуры исполнения заданий

### **11.4. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности**

Промежуточная аттестация обучающихся ведется непрерывно и включает в себя текущую аттестацию (контроль текущей работы в семестре, включая оценивание промежуточных результатов обучения по дисциплине и семестровую аттестацию (экзамен) – оценивание окончательных результатов обучения по дисциплине.

По данной дисциплине, завершающейся экзаменом, по обязательным формам текущего контроля студенту предоставляется возможность набрать в сумме не менее 60 баллов. Оценивание окончательных результатов обучения по дисциплине ведется по 100-балльной шкале, оценка формируется автоматически как сумма количества баллов, набранных обучающимся за выполнение заданий обязательных форм текущего контроля и количества баллов, набранных на семестровой аттестации (экзамене).

#### **Система оценивания**

Текущий контроль представляет собой проверку усвоения учебного материала теоретического и практического характера, регулярно осуществляемую на протяжении семестра. К основным формам текущего контроля можно отнести устный опрос, письменные задания, лабораторные работы, контрольные работы.

#### **Контрольная работа**

Контрольная работа по настоящей дисциплине представляет собой законченную работу, включающую в себя разработку модели предметной области, документа инициации проекта по созданию информационной системы, проекта инжиниринга информационной системы (в соответствии с заданием).