

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
ПОВОЛЖСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ



УТВЕРЖДАЮ
Декан ФИиВТ

УТВЕРЖДАЮ /А.А. Кречетов/
(Ф.И.О. декана (директора института))

14.02.2025 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

М.1.2.1 Анализ больших данных

(код и наименование дисциплины по учебному плану)

Направление подготовки
(специальность)

09.04.01 Информатика и вычислительная техника

Квалификация выпускника

Магистр

(бакалавр/магистр/специалист)

Программа магистратуры

Интеллектуальные системы

Курс 1
Триместр 3

Распределение учебного времени

Трудоемкость по учебному плану	<u>144 / 4</u>	часов/зачетных единиц
Лекции	<u>12</u>	часов
Лабораторные работы	<u>-</u>	часов
Практические занятия	<u>36</u>	часов
Иная контактная работа	<u>-</u>	часов
Всего контактной работы (без учета экз.)	<u>48</u>	часов
Контактная работа по экзамену	<u>-</u>	часов
Курсовой проект (работа)	<u>-</u>	триместр
Самостоятельная работа обучающихся (без учета экз.)	<u>96</u>	часов
Самостоятельная работа по подготовке к экзамену	<u>-</u>	часов
Экзамен	<u>-</u>	триместр
Зачет	<u>-</u>	триместр
БРК, ДЗ	<u>3</u>	триместр

(год)

Программа составлена в соответствии с требованиями ФГОС ВО направления подготовки (специальности) 09.04.01 Информатика и вычислительная техника

Программу составили:

доцент с ученой степенью кандидата наук	ИВС	СОГЛАСОВАНО	В.Б. Малашкевич
(должность)	(кафедра)		(И.О. Фамилия)

РАССМОТРЕНА и ОДОБРЕНА на заседании кафедры, за которой закреплена дисциплина
Кафедра информационно-вычислительных систем

(наименование кафедры)			
14.01.2025	протокол №	19	
(дата)			
Заведующий кафедрой	СОГЛАСОВАНО	Д.В. Морохин	
		(И.О. Фамилия)	

Рабочая программа СОГЛАСОВАНА с факультетом (институтом), выпускающей(ими)
кафедрой(ами).
СООТВЕТСТВУЕТ действующей ОП.

Заведующий кафедрой	СОГЛАСОВАНО	Д.В. Морохин
		(И.О. Фамилия)

Председатель методической комиссии факультета (института), в который входит
выпускающая кафедра

СОГЛАСОВАНО	А.А. Кречетов
	(И.О. Фамилия)

Эксперт(ы): Усков Юрий Викторович, Генеральный директор ООО «Ричмедиа»

Рабочая программа проверена и зарегистрирована в УМЦ 17.02.2025 г.
Специалист учебно-методического центра СОГЛАСОВАНО /Т.А. Смирнова/

Раздел 1. ЦЕЛЬ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Целью освоения дисциплины является достижение планируемых результатов обучения, соответствующих установленным в ОПОП индикаторам достижения компетенций:

Код и наименование компетенции	Код и наименование индикатора достижения компетенции	Результаты обучения
1. ПК-4 Способен разрабатывать и применять методы и алгоритмы машинного обучения для решения задач искусственного интеллекта	ПК-4.1. Ставит задачи по разработке или совершенствованию методов и алгоритмов для решения комплекса задач предметной области	знания: Знает классы методов и алгоритмов машинного обучения умения: Умеет ставить задачи и разрабатывать новые методы и алгоритмы машинного обучения навыки: Владеет навыками разработки алгоритмов машинного обучения
2. ПК-5 Способен руководить проектами по созданию комплексных систем на основе аналитики больших данных в различных отраслях	ПК-5.1. Применяет варианты использования больших данных, определений, словарей и эталонной архитектуры больших данных при руководстве проектами по построению комплексных систем на основе аналитики больших данных в различных отраслях	знания: Знает методы и критерии оценки качества моделей машинного обучения умения: Умеет определять критерии и метрики оценки результатов моделирования при построении систем искусственного интеллекта в исследуемой области навыки: Владеет навыками руководства проектами по разработке алгоритмов машинного обучения

Раздел 2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОПОП

Дисциплина относится к части, формируемой участниками образовательных отношений ОПОП.

Дисциплина является обязательной

Для продолжения формирования заявленных компетенций необходимы знания предшествующих практик: Производственная практика. Научно-исследовательская работа (рассредоточенная) (ПК-4), Производственная практика. Научно-исследовательская работа (рассредоточенная) (ПК-5)

Изучаемая дисциплина является основой для продолжения формирования указанных компетенций в следующих дисциплинах: Инструментальные средства разработки систем ИИ (ПК-4), Прикладные интеллектуальные системы (ПК-4), Интеллектуальные САПР (ПК-4), Параллельные вычисления (ПК-5); государственной итоговой аттестации в форме: Выполнение и защита выпускной квалификационной работы (ПК-4), Выполнение и защита выпускной квалификационной работы (ПК-5)

Раздел 3. ОПИСАНИЕ ОБРАЗОВАТЕЛЬНЫХ ТЕХНОЛОГИЙ

Для формирования заявленных компетенций используются методологические технологии, реализующие деятельностный, личностно-ориентированный, практико-ориентированный подходы.

Основными стратегическими технологиями являются: лекционные занятия, практика

На достижение конкретных целей обучения направлены применяемые тактические технологии: классическая лекция

Раздел 4. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

3 триместр

Виды и темы занятий	Количество часов	Формируемые компетенции
Введение в Большие Данные	18	ПК-4, ПК-5
Лекция. Предмет курса. Определение Больших Данных. Области применения и примеры использования. Сосредоточенные и распределенные базы данных. CAP-теорема и ее следствия для бизнеса. Серверные кластеры. Репликация и шардинг данных. Программные продукты для анализа Больших Данных.	1	ПК-4, ПК-5
Практическое занятие. Архитектура систем Баз Данных. Распределенные реляционных Базы Данных.	2	
Практическое занятие. Операторы языка SQL. Разработка SQL-запросов к реляционным сосредоточенным БД.	2	
Практическое занятие. Разработка SQL-запросов к реляционным распределенным БД.	2	
Лекция. Модели данных. NoSQL, базы данных. Сравнение NoSQL и реляционных баз данных.	1	
Задания для самостоятельной работы, в том числе выполнение Проработка лекций Подготовка к практическим работам Операторы языка SQL	10	
Методы хранения Больших Данных.	17	
Лекция. Понятие хранилища данных. Технологии и архитектура систем хранения данных. Системы DAS, SAN и NAS. RAID-массивы. Многоуровневое хранение данных. Распределенные хранилища данных. Облачные хранилища.	1	
Практическое занятие. Расчет дисковой подсистемы хранилища данных.	2	
Практическое занятие. Выбор конфигурации дисковой подсистемы хранилища данных	2	ПК-4, ПК-5
Задания для самостоятельной работы, в том числе выполнение Проработка лекций Подготовка к практическим работам Системы дисковой памяти	12	
Методы сбора и очистки Больших Данных	17	
Лекция. Источники данных – Web-сайты, социальные сети, мобильные устройства, данные с измерительных устройств, видео- и аудио-оборудование, бизнес-информация и корпоративные БД. Методы сбора данных. Подготовка данных к анализу (алгоритмы выборки, очистки, сортировки); Введение в распределенную файловую систему Hadoop.	1	ПК-4, ПК-5

Технологии распределенной файловой системы HDFS. Концепции и терминология базы данных Oracle NoSQL. Введение в Oracle NoSQL - архитектура и компоненты.		
Практическое занятие. Алгоритмы передискретизации данных,	2	
Практическое занятие. Алгоритмы цифровой фильтрации данных	2	
Задания для самостоятельной работы, в том числе выполнение Проработка лекций Подготовка к практическим работам Основы баз данных. Цифровая обработка сигналов.	12	
Методы вычислений над Большими Данными	22	ПК-4, ПК-5
Лекция. Модель вычисления Map Reduce. Этапы вычислений.	2	
Практическое занятие. Разработка схемы БД Oracle NoSQL MongoDB	1	
Практическое занятие. Разработка приложения для БД NoSQL MongoDB	2	
Практическое занятие. Разработка приложения для вычислений Map Reduce БД NoSQL MongoDB.	2	
Практическое занятие. Работа с NoSQL СУБД Redis	1	
Задания для самостоятельной работы, в том числе выполнение Проработка лекций Подготовка к практическим работам Реляционная модель данных.	14	
Исследовательский анализ данных	22	ПК-4, ПК-5
Лекция. Поиск закономерностей в данных. Визуализация данных. Статистический анализ данных. Кластерный анализ. Модели временных рядов.	2	
Практическое занятие. Алгоритмы поиска данных.	1	
Практическое занятие. Алгоритмы кластерного анализа.	1	
Практическое занятие. Модели временных рядов. Прогнозирование.	2	
Задания для самостоятельной работы, в том числе выполнение Проработка лекций Подготовка к практическим работам Теория вероятностей и мат. статистика	16	
Методы машинного обучения	22	ПК-4, ПК-5
Лекция. Алгоритмы машинного обучения для решения задач регрессии, классификации и кластеризации. Алгоритмы работы рекомендательных систем и прогнозирования. Методы бустинга в задачах распознавания. Методы видеоаналитики. Нейронные сети в задачах классификации. Задача классификации изображений. Оценка качества классификации.	2	
Практическое занятие. Распознавание объектов и детектирование событий по данным видеонаблюдения.	4	
Задания для самостоятельной работы, в том числе выполнение Проработка лекций Подготовка к практическим работам Теория вероятностей и мат. статистика	16	
Интеллектуальный анализ данных	26	ПК-4, ПК-5
Лекция. Алгоритмы Data mining — глубинный анализ данных.	2	

Классификация. Ассоциация. Регрессионный анализ.		
Практическое занятие. Байесовский классификатор.	1	
Практическое занятие. Метод k ближайших соседей.	1	
Практическое занятие. Решающие деревья и случайный лес.	2	
Практическое занятие. Сокращения размерности многомерного пространства признаков	2	
Практическое занятие. Логистическая регрессия и дискриминационный анализ.	2	
Задания для самостоятельной работы, в том числе выполнение Проработка лекций Подготовка к практическим работам Интеллектуальный анализ	16	
Иная контактная работа: дифференцированный зачет (БРК)	0	

Раздел 5. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

Изучение дисциплины "Анализ больших данных" рекомендуется начать с ознакомления с рабочей программой, ее структурой и содержанием разделов. Учебный материал структурирован, изучение дисциплины осуществляется в тематической последовательности. **Занятия лекционного типа** дают систематизированные знания по дисциплине "Анализ больших данных", концентрируют внимание на наиболее сложных и важных вопросах. Во время лекционных занятий рекомендуется вести конспектирование учебного материала; обращать внимание на формулировки и категории, раскрывающие суть проблемы, явления или процесса; зафиксировать выводы и практические рекомендации. Подготовка к **занятиям семинарского типа** включает ознакомление с планом практического занятия; работу с конспектом лекций, выполнение домашнего задания, работу с учебной и учебно-методической литературой, научными изданиями и электронными образовательными ресурсами, рекомендованными рабочей программой дисциплины "Анализ больших данных".

При выполнении практических заданий обучающимся целесообразно использовать следующее программное обеспечение: среда разработки Jupyter Notebook, аналитическая платформа Loginom, сервис визуализации и анализа данных Yandex DataLens, аналитическая платформа Полиматика, система визуализации Constru.

Содержание **самостоятельной работы** определяется рабочей программой дисциплины "Анализ больших данных", оценочными и методическими материалами, заданиями и указаниями преподавателя. Самостоятельная работа может осуществляться в аудиторной и внеаудиторной формах. Эффективным средством осуществления самостоятельной работы является электронная информационно-образовательная среда университета, которая обеспечивает доступ к образовательной программе, рабочей программе дисциплины "Анализ больших данных", к электронным библиотечным системам, профессиональным базам данных и информационным справочным системам.

В соответствии с меморандумом о сотрудничестве с компанией Яндекс доступны для изучения электронные курсы и проверочные материалы на платформе Coursera. Периодичность проведения, формы текущего контроля успеваемости, система оценивания хода освоения дисциплин представлены в рабочей программе. Условия аттестации приведены в технологической карте, входящей в состав рабочей программы дисциплины "Анализ больших данных". Формой промежуточной аттестации по дисциплине "Анализ больших данных" является балльно-

рейтинговый контроль.

Раздел 6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ И УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

6.1. Учебно-методическое обеспечение

№№ п/п	Список используемой литературы	Количество экземпляров печатных изданий, имеющихся в библиотеке, или электронный адрес издания (ресурса) в сети Интернет
УЧЕБНЫЕ, УЧЕБНО-МЕТОДИЧЕСКИЕ И НАУЧНЫЕ ИЗДАНИЯ		
1.	Методы и модели анализа данных: OLAP и Data Mining [Текст] : учеб. пособие по специальности 071900 "Информ. системы и технологии" направления 654700 "Информ. системы" / А. А. Барсегян, М. С. Куприянов, В. В. Степаненко, И. И. Холод. СПб.: БХВ-Петербург, 2004. - 336 с. ISBN 5-94157-522-X. Экземпляры: всего 32.	32
2.	Усков, Андрей Александрович. Интеллектуальные технологии управления [Текст] : искусственные нейронные сети и нечеткая логика / А. А. Усков, А. В. Кузьмин. М.: Горячая линия - Телеком, 2004. - 143 с. ISBN 5-93517-181-3. Экземпляры: всего 10.	10
3.	Сидоркина, Ирина Геннадьевна. Системы искусственного интеллекта [Текст] : [учеб. пособие для вузов по направлению 230100 "Информатика и вычисл. техника"] / И. Г. Сидоркина. М.: Кнорус, 2011. - 245 с. ISBN 978-5-406-00449-4. Экземпляры: всего 92.	92
4.	Кутузов, О. И. Инфокоммуникационные системы и сети [Электронный ресурс] / Кутузов О. И., Татарникова Т. М., Цехановский В. В. 3-е изд., стер. Санкт-Петербург: Лань, 2022. - 244 с. ISBN 978-5-507-44763-3.	https://e.lanbook.com/book/242858
5.	Цехановский, В. В. Управление данными [Электронный ресурс] : учебник / В. В. Цехановский, В. Д. Чертовской. Санкт-Петербург: Лань, 2022. - 432 с. ISBN 978-5-8114-1853-4.	https://e.lanbook.com/book/212084

6.2. Материально-техническая база и программное обеспечение

№№ п/п	Аудитории для проведения учебных занятий, самостоятельной работы и проведения государственной итоговой аттестации	Перечень основного оборудования	Программное обеспечение
1.	518 (III)	Системный блок CEL D-341 FAN/ASUS S-775/512 M/160.0G/DVD+-RW (1), ПК 5 - ICL RAY P222.3 ,клавиат.,мышь.,монитор LG E2251T-BN (14), Сист. блок CE 331/256*2/PC 3200/80 Gb/FDD/DVD-	Microsoft Windows Enterprise, Microsoft Office Standard, Агент Dr.Web, Microsoft Access, Microsoft Visio Professional, Microsoft Project Professional, Microsoft Visual Studio

		Комплект учебной мебели (1)	Enterprise, Комплект ПО для решения основных пользовательских задач
--	--	-----------------------------	---

Раздел 7. ФОРМЫ КОНТРОЛЯ ОСВОЕНИЯ ДИСЦИПЛИНЫ/ ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

Критерии оценивания индикаторов достижения компетенций направлены на:

- усвоение теоретического материала (объем знаний, глубина усвоения), предусмотренного рабочей программой;
- умение излагать материал (четкость, грамотность изложения материала, точность и полнота воспроизведения учебного материала);
- умение применять теоретические знания при решении практических заданий.

Шкала оценивания представлена ниже.

Уровень сформированности элементов компетенции	Критерии оценивания	Шкала оценивания
Пороговый уровень	Обучающийся имеет знания основного материала, проявляет умение логично его излагать, но может допускать неточности в изложении материала, недостаточно правильные формулировки, испытывает затруднения в выполнении практических заданий.	удовлетворительно
Продвинутый уровень	Обучающийся твердо знает программный материал, излагает его грамотно и по существу, не допускает существенных неточностей в ответе на вопрос, правильно применяет теоретические положения при решении практических вопросов и задач, владеет необходимыми навыками и приемами их выполнения	хорошо
Высокий уровень	Обучающийся глубоко и прочно усвоил программный материал, грамотно и логически стройно его излагает, дает исчерпывающие ответы на поставленные вопросы. В ответе тесно увязывается теория с практикой, при этом обучающийся не затрудняется с ответом при видоизменении задания, свободно справляется с задачами, вопросами и другими видами применения знаний, показывает знакомство с монографической литературой, периодическими изданиями, правильно обосновывает принятые решения, свободно владеет разносторонними навыками, приемами выполнения практических работ	отлично

7.1. Текущий контроль успеваемости

Текущий контроль успеваемости обеспечивает оценивание хода освоения дисциплины (модуля) и производится с применением технологии рейтингового контроля в соответствии с технологической картой дисциплины. Порядок составления технологической карты и алгоритм проведения процедуры оценивания видов деятельности обучающихся, направленных на освоение знаний, умений, навыков и/ или опыта деятельности, по накопительной системе в баллах устанавливается положением о системе РИТМ в ФГБОУ ВО «ПГТУ»

7.2. Промежуточная аттестация обучающихся

Промежуточная аттестация обучающихся направлена на оценивание результатов обучения по дисциплине (модулю) и проводится с использованием фондов оценочных средств.

Примеры типовых контрольных заданий из базы фонда оценочных средств по образовательной программе.

1. Дисковые системы RAID. Способы организации и обеспечения доступности.
2. Принципы информационного поиска.

Перечень вопросов для проведения промежуточной аттестации

3.
 1. Какие технологии совместного использования дисков применяются в RAID-массивах?
 2. Какие технологии обеспечивают доступность данных?
 3. Как обеспечивается доступность данных в RAID-массивах?
 4. В чем состоит принцип распределения в дисковых массивах?
 5. В чем состоит принцип зеркалирования в дисковых массивах?
 6. Для каких целей используется принцип контроля четности?
 7. Объясните принципы работы RAID-массивов базового уровня 0.
 8. Объясните принципы работы RAID-массивов базового уровня 1.
 9. Объясните принципы работы RAID-массивов базового уровня 10 и 01.
 10. Объясните принципы работы RAID-массивов базового уровня 5.
 11. Объясните принципы работы RAID-массивов разных уровней при авариях дисков.
 12. Как влияет на конфигурацию RAID-массива контроллер RAID?
 13. Опишите архитектуру распределенного хранилища данных.
 14. Что такое кластер?
 15. Какие системы называются масштабируемыми?
 16. Что такое «вертикальное» масштабирование.
 17. Что такое «горизонтальное» масштабирование.
 18. Что такое репликация данных?
 19. Опишите технологию репликации «Master-Slave».
 20. Опишите технологию репликации «Peer-to-Peer».
 21. Что называется «шардингом»?
 22. В чем различие между «вертикальным» и «горизонтальным» шардингом?
 23. Объясните термин NoSQL.
 24. В чем заключается «горизонтальное» масштабирование?

25. Какая система хранения данных называется распределенной?
26. Что утверждает теорема CAP и для каких систем хранения она справедлива?
27. В чем состоит состоятельность данных?
28. Как в системах BigData компенсируется возможная несостоятельность данных?
29. Охарактеризуйте модель данных «Ключ-Значение».
30. Укажите основные этапы вычислений технологии Map-Reduce.
31. Какие действия выполняются на этапе Map?
32. Какова цель этапа Shuffle в технологии Map-Reduce?
33. Какие задачи решаются на этапе Reduce?
34. Какую роль играет Master-node в технологии Map-Reduce?
35. Каковы функции Worker-node в технологии Map-Reduce?
36. Какие задачи решаются методами DataMining?
37. Как определяется задача информационного поиска?
38. Что такое «Инвертированный список»?
39. В чем заключается принцип работы алгоритма «грубой силы» при поиске в тексте заданной строки?
40. Объясните принцип работы алгоритма Бойера — Мура — Хорспула поиска в тексте заданной строки
41. Сформулируйте задачу нечеткого поиска в тексте заданной строки.
42. Как вычисляется расстояние Левенштейна между двумя строками?
43. Объясните принципы логистической регрессии данных.
44. Что такое векторизация слов?
45. Как работают деревья принятия решения?
46. Какую структуру может иметь элемент «Значение» в паре «Ключ-Значение»?
47. Охарактеризуйте графовую модель данных.
48. В каких приложениях целесообразно использовать графовую модель данных?
49. Объясните структуру хранилища данных СУБД MongoDB и соответствие элементов этой структуры элементам традиционных реляционных СУБД.
50. Поясните принципы моделирования данных в документарной модели.
51. Аналогом какой структуры данных реляционной модели является коллекция?
52. Сравните ссылочную модель и модель вложения с реляционной моделью данных.
53. Какие операции понимаются под аббревиатурой CRUD?
54. Каким образом приложения, написанные на разных языках программирования, получают доступ к документам СУБД MongoDB?

55. Как выполнить подключение к серверу СУБД MongoDB из приложения?
56. Как из приложения получить доступ к коллекции документов?
57. Каким образом обеспечивается выбор документов, соответствующих заданным критериям отбора (на примере MongoDB)?
58. Охарактеризуйте байесовские алгоритмы классификации.
59. Что такое "бустинг"?
60. Опишите метод ближайшего соседа.

An error has occurred while processing HtmlTextBox 'htmlTextBox3': List item () must be in a list (or).