

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
ПОВОЛЖСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ



УТВЕРЖДАЮ
Директор ИММ

УТВЕРЖДАЮ /Н.П. Сютлов/
(Ф.И.О. декана (директора института))

01.03.2023 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

Б.1.2.3 Анализ больших данных

(код и наименование дисциплины по учебному плану)

Направление подготовки
(специальность)

35.03.06 Агроинженерия

Квалификация выпускника

Бакалавр

(бакалавр/магистр/специалист)

Направленность

Искусственный интеллект в агроинженерии

Курс 3
Семестр 5

Распределение учебного времени

Трудоемкость по учебному плану	180 / 5	часов/зачетных единиц
Лекции	16	часов
Лабораторные работы	-	часов
Практические занятия	64	часов
Иная контактная работа	-	часов
Всего контактной работы (без учета экз.)	80	часов
Контактная работа по экзамену	6	часов
Курсовой проект (работа)	-	семестр
Самостоятельная работа обучающихся (без учета экз.)	64	часов
Самостоятельная работа по подготовке к экзамену	30	часов
Экзамен	5	семестр
Зачет	-	семестр
БРК, ДЗ	-	семестр

(год)

Программа составлена в соответствии с требованиями ФГОС ВО направления подготовки (специальности) 35.03.06 Агроинженерия

Программу составили:

доцент с ученой степенью кандидата наук	Информатики	СОГЛАСОВАНО	Ю.А. Ипатов
(должность)	(кафедра)		(И.О. Фамилия)
доцент, кандидат наук	ИВС	СОГЛАСОВАНО	В.Б. Малашкевич
(должность)	(кафедра)		(И.О. Фамилия)

РАССМОТРЕНА и ОДОБРЕНА на заседании кафедры, за которой закреплена дисциплина
Кафедра информатики

	(наименование кафедры)		
20.02.2023	протокол №	7	
(дата)			
Заведующий кафедрой	СОГЛАСОВАНО	А.В. Кревецкий	
		(И.О. Фамилия)	

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

СООТВЕТСТВУЕТ действующей ОП.

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

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

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

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

Рабочая программа проверена и зарегистрирована в УМЦ 06.03.2023 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-2 Способен классифицировать и идентифицировать задачи искусственного интеллекта, выбирать адекватные методы и инструментальные средства решения задач искусственного интеллекта	ПК-2.1 Классифицирует и идентифицирует задачи систем искусственного интеллекта в зависимости от особенностей проблемной и предметной областей	знания: Методологические подходы к выбору и применению методов обработки и распределения знаний умения: Применение методов обработки и распространения знаний с системами, основанных на знаниях, для решения задач профессиональной деятельности навыки: Организации решения задач профессиональной деятельности на основе использования систем, основанных на знаниях
2. ПК-5 Способен осуществлять сбор и подготовку данных для систем искусственного интеллекта	ПК-5.1 Осуществляет поиск данных в открытых источниках, специализированных библиотеках и репозиториях.	знания: Методологические подходы к выбору и применению методов обработки и распределения знаний умения: Применение методов обработки и распространения знаний с системами, основанных на знаниях, для решения задач профессиональной деятельности навыки: Организации решения задач профессиональной деятельности на основе использования систем, основанных на знаниях
3. ПК-6 Способен выполнять анализ больших данных	ПК-6.1 Использует знания о вариантах использования больших данных, определениях, словарях и эталонной архитектуре больших данных для эффективного извлечения, хранения и подготовки больших данных.	знания: Методологические подходы к выбору и применению методов структурирования знаний для предметных областей в виде ментальных карт, таксономий, деревьев целей и решений умения: Применение методов структурирования знаний для построения моделей знаний (онтологий знаний) навыки: Участие в процессе концептуального моделирования и структурирования знаний

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 дисциплин: Основы программирования систем искусственного интеллекта на Python (ПК-5)

Изучаемая дисциплина является основой для продолжения формирования указанных компетенций в следующих государственной итоговой аттестации в форме: Выполнение и защита выпускной квалификационной работы (ПК-5), Выполнение и защита выпускной квалификационной работы (ПК-6), Выполнение и защита выпускной квалификационной работы (ПК-2)

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, практические занятия

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

Раздел 4. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

5 семестр

Виды и темы занятий	Количество часов	Формируемые компетенции
Основы Больших Данных	36	ПК-2, ПК-5, ПК-6
Лекция. Предмет курса. Основные понятия, области применения и примеры использования Больших Данных. Архитектура хранилищ Больших Данных. Распределенные базы данных (БД). CAP-теорема и ее следствия для бизнеса. Введение в экосистему Hadoop.	2	
Практическое занятие. Операторы языка SQL. Разработка SQL-запросов к реляционным сосредоточенным и распределенным БД.	8	
Лекция. Модели данных. NoSQL, базы данных. Операции CRUD в реляционных и NoSQL БД. Технология MapReduce для работы с BigData. Программные продукты для анализа Больших Данных. Аналитическая платформа Loginom. Язык Python как инструмент анализа данных.	2	
Практическое занятие. Введение в Аналитическую платформу Loginom.	8	
Задания для самостоятельной работы, в том числе выполнение Проработка лекций Подготовка к практическим работам Операторы языка SQL	16	
Технологии анализа данных	52	ПК-2, ПК-5, ПК-6
Лекция. Структурированные и неструктурированные данные. Подготовка данных к анализу. Алгоритмы выборки, очистки, сортировки;	2	
Лекция. Поиск закономерностей в данных. Визуализация	2	

данных. Статистический анализ данных. Кластерный анализ. Модели временных рядов.		
Лекция. Основы языка Python. Библиотеки для анализа данных (NumPy, Matplotlib, scikit-learn)	2	
Практическое занятие. Алгоритмы цифровой фильтрации данных	8	
Практическое занятие. Алгоритмы поиска данных. Алгоритмы кластерного анализа.	8	
Лекция. Смешение и интеграция данных - набор техник интегрировать разнородных данных. Имитационное моделирование – построение моделей имитирующих реальные процессы, разновидность экспериментальных испытаний. Пространственный анализ - класс методов, использующих топологическую, геометрическую и географическую информацию в данных.	2	
Практическое занятие. Методы сравнения наборов данных: Визуальное сравнение гистограмм Метод актуарного анализа Метод регрессионного анализа	8	
Задания для самостоятельной работы, в том числе выполнение Проработка лекций Подготовка к практическим работам Системы дисковой памяти	20	
Алгоритмы интеллектуального анализа данных	56	ПК-2, ПК-5, ПК-6
Лекция. Алгоритмы машинного обучения для решения задач регрессии, классификации и кластеризации. Алгоритмы работы рекомендательных систем и прогнозирования. Методы бустинга в задачах распознавания и видеоаналитики. Нейронные сети в задачах классификации. Задача классификации изображений. Оценка качества классификации.	2	
Лекция. Алгоритмы Data mining — глубинный анализ данных. Классификация. Ассоциация. Регрессионный анализ. Задача и алгоритмы видеоаналитики.	2	
Практическое занятие. Статистическая проверка гипотез.	8	
Практическое занятие. Байесовский классификатор.	8	
Практическое занятие. Представление и визуализация результатов анализа.	8	
Задания для самостоятельной работы, в том числе выполнение Проработка лекций Подготовка к практическим работам Теория вероятностей и мат. статистика	28	
Иная контактная работа:	0	
Подготовка к экзамену	30	
Проведение экзамена	6	

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". 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. Периодичность проведения, формы текущего контроля успеваемости, система оценивания хода освоения дисциплин представлены в рабочей программе. Формой промежуточной аттестации по дисциплине "Анализ больших данных" является экзамен.

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

6.1. Учебно-методическое обеспечение

№№ п/п	Список используемой литературы	Количество экземпляров печатных изданий, имеющих 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 16.	16
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.	Цехановский, В. В. Распределенные информационные системы [Электронный ресурс] / Цехановский В. В., Чертовской В. Д. 2-е изд., стер. Санкт-Петербург: Лань, 2021. - 240 с. ISBN 978-5-8114-8732-5.	https://e.lanbook.com/book/179622

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-ROM/КЛАВ+МЫШЬ+коврик (1), Комплект учебной мебели (1)	Microsoft Windows Enterprise, Microsoft Office Standard, Агент Dr.Web, Microsoft Access, Microsoft Visio Professional, Microsoft Project Professional, Microsoft Visual Studio 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,	отлично

	периодическими изданиями, правильно обосновывает принятые решения, свободно владеет разносторонними навыками, приемами выполнения практических работ	
--	--	--

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.

Билет № 0

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

Перечень вопросов для проведения промежуточной аттестации

3.

1. В чем состоит принцип распределения в дисковых массивах?
2. В чем состоит принцип зеркалирования в дисковых массивах?
3. Объясните принципы работы RAID-массивов базового уровня 10 и 01.
4. Объясните принципы работы RAID-массивов базового уровня 5.
5. Опишите архитектуру распределенного хранилища данных.
6. Что такое кластер?
7. Какие системы называются масштабируемыми?
8. Что такое «вертикальное» и "горизонтальное" масштабирование.
9. Что такое репликация данных?
10. Опишите технологии репликации «Master-Slave», «Peer-to-Peer».
11. Что называется «шардингом»?
12. В чем различие между «вертикальным» и «горизонтальным» шардингом?
13. Объясните термин NoSQL.
14. Какая система хранения данных называется распределенной?
15. Что утверждает теорема CAP и для каких систем хранения она справедлива?

16. В чем состоит состоятельность данных?
17. Как в системах BigData компенсируется возможная несостоятельность данных?
18. Охарактеризуйте модель данных «Ключ-Значение».
19. Укажите основные этапы вычислений технологии Map-Reduce.
20. Какие действия выполняются на этапе Map?
21. Какова цель этапа Shuffle в технологии Map-Reduce?
22. Какие задачи решаются на этапе Reduce?
23. Какую роль играет Master-node в технологии Map-Reduce?
24. Каковы функции Worker-node в технологии Map-Reduce?
25. Какие задачи решаются методами DataMining?
26. Как определяется задача информационного поиска?
27. Что такое «Инвертированный список»?
28. В чем заключается принцип работы алгоритма «грубой силы» при поиске в тексте заданной строки?
29. Объясните принцип работы алгоритма Бойера — Мура — Хорспула поиска в тексте заданной строки
30. Сформулируйте задачу нечеткого поиска в тексте заданной строки.
31. Как вычисляется расстояние Левенштейна между двумя строками?
32. Объясните принципы логистической регрессии данных.
33. Что такое векторизация слов?
34. Как работают деревья принятия решения?
35. Какую структуру может иметь элемент «Значение» в паре «Ключ-Значение»?
36. Охарактеризуйте графовую модель данных.
37. В каких приложениях целесообразно использовать графовую модель данных?
38. Объясните структуру хранилища данных СУБД MongoDB и соответствие элементов этой структуры элементам традиционных реляционных СУБД.
39. Поясните принципы моделирования данных в документарной модели.
40. Сравните ссылочную модель и модель вложения с реляционной моделью данных.
41. Какие операции понимаются под аббревиатурой CRUD?
42. Каким образом приложения, написанные на разных языках программирования, получают доступ к документам СУБД MongoDB?
43. Как выполнить подключение к серверу СУБД MongoDB из приложения?
44. Как из приложения получить доступ к коллекции документов?
45. Каким образом обеспечивается выбор документов, соответствующих заданным критериям отбора (на примере MongoDB)?

46. Охарактеризуйте байесовские алгоритмы классификации.
47. Что такое "бустинг"? Укажите основные принципы бустинга.
48. Опишите метод ближайшего соседа.
49. По каким алгоритмам строятся решающие деревья?
50. Охарактеризуйте алгоритм ID3.
51. Охарактеризуйте алгоритм CART.
52. Какова задача регрессионного анализа?
53. Что такое "множественная регрессия"?
54. С какой целью выполняется тест Дарбина-Уотсона?
55. Как оценить значимость модели регрессии?
56. Перечислите виды дискриминантного анализа.
57. Какие задачи решает дискриминантный анализ?
58. Охарактеризуйте виды дискриминантного анализа.
59. Алгоритм поиска дискриминантной функции и критерий Фишера. Какую функцию расстояния рекомендуется использовать для оценки качества дискриминантного анализа?

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