

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
ПОВОЛЖСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ



УТВЕРЖДАЮ
Декан ФУП

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

02.02.2023 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

М.1.2.1 Анализ больших данных

(код и наименование дисциплины по учебному плану)

Направление подготовки
(специальность)

38.04.04 Государственное и муниципальное управление

Квалификация выпускника

Магистр

(бакалавр/магистр/специалист)

Программа магистратуры

Искусственный интеллект в архитектуре информационных
систем цифрового государства

Курс 2
Семестр 3, 4

Распределение учебного времени

Трудоемкость по учебному плану	144 / 4	часов/зачетных единиц
Лекции	2	часов
Лабораторные работы	-	часов
Практические занятия	4	часов
Иная контактная работа	-	часов
Всего контактной работы (без учета экз.)	6	часов
Контактная работа по экзамену	6	часов
Курсовой проект (работа)	-	семестр
Самостоятельная работа обучающихся (без учета экз.)	102	часов
Самостоятельная работа по подготовке к экзамену	30	часов
Экзамен	4	семестр
Зачет	-	семестр
БРК, ДЗ	-	семестр

(год)

Программа составлена в соответствии с требованиями ФГОС ВО направления подготовки (специальности) 38.04.04 Государственное и муниципальное управление

Программу составили:

заведующий кафедрой с ученой степенью кандидата наук	УиП	СОГЛАСОВАНО	Т.В. Ялялиева
(должность)	(кафедра)		(И.О. Фамилия)

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

(наименование кафедры)		
25.01.2023	протокол №	7
(дата)		

Заведующий кафедрой	СОГЛАСОВАНО	И.Г. Сидоркина
		(И.О. Фамилия)

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

СООТВЕТСТВУЕТ действующей ОП.

Заведующий кафедрой	СОГЛАСОВАНО	Т.В. Ялялиева
		(И.О. Фамилия)

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

СОГЛАСОВАНО	С.В. Краснова
	(И.О. Фамилия)

Эксперт(ы): Мельник Оксана Николаевна , директор АУ Республики Марий Эл «Дирекция
многофункциональных центров предоставления государственных и муниципальных услуг в
Республике Марий Эл»

Рабочая программа проверена и зарегистрирована в УМЦ 09.02.2023 г.

Специалист учебно-методического центра СОГЛАСОВАНО /М.Л. Бойкова/

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-6 Способен руководить проектами по созданию комплексных систем на основе аналитики больших данных в различных отраслях со стороны заказчика	ПК 6.1 Руководит проектами по построению комплексных систем на основе аналитики больших данных в различных отраслях со стороны заказчика	знания: знает правила управления проектами по построению комплексных систем на основе аналитики больших данных в различных отраслях со стороны заказчика умения: умеет руководить проектами по построению комплексных систем на основе аналитики больших данных в различных отраслях со стороны заказчика навыки: имеет навыки руководства проектами по построению комплексных систем на основе аналитики больших данных в различных отраслях со стороны заказчика

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 построения и эволюции систем искусственного интеллекта (ПК-6)

Изучаемая дисциплина является основой для продолжения формирования указанных компетенций в следующих дисциплинах: Прикладные системы искусственного интеллекта в государственном и муниципальном управлении (ПК-6); практика: Преддипломная практика (ПК-6); государственной итоговой аттестации в форме: Подготовка к процедуре защиты и защита выпускной квалификационной работы (ПК-6)

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, процедуры самообучения

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

Раздел 4. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

Виды и темы занятий	Количество часов	Формируемые компетенции
Большие данные и хранилища данных	36	ПК-6
Лекция. Общая характеристика сквозных цифровых технологий	2	
Задания для самостоятельной работы, в том числе выполнение Нейронные сети на Python	34	
Иная контактная работа:	0	

4 семестр

Виды и темы занятий	Количество часов	Формируемые компетенции
Технологии и системы работы с большими данными	72	ПК-6
Практическое занятие. Технологии и системы работы с большими данными	2	
Практическое занятие. Общая характеристика сквозных цифровых технологий	2	
Задания для самостоятельной работы, в том числе выполнение Нейронные сети на Python	68	
Иная контактная работа:	0	
Подготовка к экзамену	30	
Проведение экзамена	6	

Раздел 5. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

Изучение дисциплины 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, 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.

Содержание **самостоятельной работы** определяется рабочей программой дисциплины, 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, 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.

* ..

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.	Макшанов, А. В. Большие данные. Big Data [Электронный ресурс] : учебник для вузов / Макшанов А. В., Журавлев А. Е., Тындыкарь Л. Н.; Журавлев А. Е., Тындыкарь Л. Н. 3-е изд., стер. Санкт-Петербург: Лань, 2023. - 188 с. ISBN 978-5-507-46866-9.	https://e.lanbook.com/book/322664
2.	Макшанов, А. В. Системы поддержки принятия решений [Электронный ресурс] / Макшанов А. В., Журавлев А. Е., Тындыкарь Л. Н. 2-е изд., стер. Санкт-Петербург: Лань, 2022. - 108 с. ISBN 978-5-8114-8489-8.	https://e.lanbook.com/book/176903
3.	Чубукова, И. А. Data Mining [Электронный ресурс] / Чубукова И. А. 2-е изд. Москва: ИНТУИТ, 2016. - 470 с. ISBN 978-5-94774-819-2.	https://e.lanbook.com/book/100582
4.	Интеллектуальный анализ данных средствами MS SQL Server 2008 [Электронный ресурс]. 2-е изд. Москва: ИНТУИТ, 2016. - 337 с.	https://e.lanbook.com/book/100609
ЭЛЕКТРОННЫЕ ОБРАЗОВАТЕЛЬНЫЕ РЕСУРСЫ		
1.	Научная электронная библиотека eLIBRARY.RU	http://elibrary.ru
2.	Научная электронная библиотека «Киберленинка»	http://cyberleninka.ru
ПРОФЕССИОНАЛЬНЫЕ БАЗЫ ДАННЫХ И ИНФОРМАЦИОННЫЕ СПРАВОЧНЫЕ СИСТЕМЫ		
1.	Справочно-правовая система Консультант+	http://www.consultant.ru
2.	Информационно-правовой портал Гарант	http://www.garant.ru

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.	515 (I)	Мультимедийный проектор Hitachi CP-X440 (1), ПК Моноблок RAMEC GALE Custom 21,5"/i3-3240/H61M/4DDR3/500SATA3/клав. ,мышь (29), Комплект учебной мебели (1)	Microsoft Windows Enterprise, Справочная правовая система "Консультант Плюс", Microsoft Office Standard, Агент Dr.Web, Комплект ГАРАНТ-Мастер, 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, периодическими изданиями, правильно обосновывает принятые решения, свободно владеет разносторонними навыками, приемами выполнения практических работ	отлично

7.1. 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.

Кейс 1. В файле ГосСлужба.txt приведены данные по стажу работы, стажу в должности и возрасту в

государственной службе. Построить гистограммы распределения случайных величин. Оценить выборочные характеристики. Проверить статистические гипотезы о значимом отличии стажа в должности, стажа работы на гос. службе и возраста для мужчин и женщин с использованием t-критерия и критерия Манна-Уитни. Построить диаграммы размаха для случайных величин: возраст, стаж службы. Задачу решить в SPSS.

Кейс 2. Проверить гипотезу о значимом отличии среднего балла за экзамены в десятом и одиннадцатом классах, используя критерий Стьюдента и критерий Манна-Уитни. Построить диаграммы «ящик с усами» для школьников, имеющих разные хобби. Построить диаграмму «дерево-листья». Данные находятся в файле тесты Школа.txt. Задачу решить в R и в SPSS.

Построить задачу классификации хобби в зависимости от результатов тестирования.

Задачу классификации решить с помощью деревьев решений в R.

Кейс 3. Решить задачу кластерного анализа для файла Семейное положение.txt. При решении задачи кластерного анализа: - определить склонность к кластеризации; - определить лучшую метрику иерархической кластеризации; - выполнить иерархическую кластеризацию; - определить состав и центроиды кластеров; - Решить задачу кластеризации методом k-средних; - Выполнить интерпретацию полученных кластеров; - Визуализировать полученную кластеризацию; - Задачу решить в RStudio и в SPSS

Перечень вопросов для проведения промежуточной аттестации

Билет 0

1. Характеризовать стандарты Data Mining, стандарты CWM, CRISP, PMML.
2. Характеризовать базовые алгоритмы кластеризации. Рассмотреть иерархическую кластеризацию, понятие дендрограммы, организацию ее использования.
3. Дать определение понятия «большие данные». Перечислить свойства больших данных. Перечислить основные сквозные цифровые технологии.
4. Указать роль аналитика по данным (Data Scientist). Сформулировать ключевые компетенции аналитика. Выделить отличия BI от Data Science.
5. Сформулировать парадигма Map Reduce. Описать принцип работы. Нарисовать диаграмму. Перечислить слабые и сильные стороны. Обозначить области применимости. Привести примеры использования.
6. Перечислить задачи систем поддержки принятия решений. OLTP и OLAP-системы. Принципы построения информационных хранилищ. Модели информационных хранилищ. Многомерная модель данных. Правила Кодда.
7. Дать характеристику размерностных моделей, моделей MOLAP, ROLAP, HOLAP системы.
8. Характеризовать витрины данных
9. Сделать обзор ETL (Extract Transform Load) технологии.
10. Характеризовать стандарты Data Mining, стандарты CWM, CRISP, PMML.
11. Рассмотреть содержание и этапы методология KDD. Дать характеристику задачи предобработки данных.
12. Рассмотреть модель распределенных вычислений. Map Reduce. Реализацию Map Reduce на Python
13. Сделать обзор языков Python. Описать основы языка Python. Привести примеры среды разработки. Сделать обзор платформы Anaconda.
14. Продемонстрировать организацию использования пакета Anaconda.
15. Характеризовать основы синтаксиса языка, переменных, ключевых слов языка, основ программирования на языке Python
16. Характеризовать типы данных языка Python, функциональности для работы с данными. Описать организацию установки пакетов научных вычислений на Python, установки Keras.
17. Сделать обзор организации решения задач анализа данных. Дать определение понятию тензора, скаляров, векторов, матриц, тензоров третьего и высшего рангов, временных рядов или последовательностей. Рассмотреть операции над тензорами.
18. Характеризовать фреймворк Keras, библиотеки Google TensorFlow
19. Определить содержание задач кластерного анализа, понятие кластера, параметров кластера. Объяснить меры близости между кластерами, используемые метрики кластерного анализа.
20. Характеризовать базовые алгоритмы кластеризации. Рассмотреть иерархическую кластеризацию,

понятие дендрограммы, организацию ее использования.

19. Характеризовать метод К-средних. Раскрыть понятие центроида, профили кластеров. Связать понятия кластерного и регрессионного анализа. Описать организацию решения задач кластерного анализа на Python.

20. Дать определение задачи классификации. Перечислить методы классификации. Дать общую характеристику классификационного анализа с обучением. Продемонстрировать решение задачи классификации в Orange.

21. Объяснить основное содержание метода деревьев решений. Привести примеры деревьев решений. Характеризовать алгоритмы построения деревьев решений. Показать организацию решения задачи классификации в Python.

22. Характеризовать метод логистической регрессии, организацию ее решения в Python. Привести пример решения задачи различными методами классификации с помощью Orange. Исследовать результаты классификации различными методами.

23. Объяснить содержание метода k-ближайших соседей. Продемонстрировать решение задачи и сравнить с результатами решения другими методами.

24. Сделать обзор инструментов, используемых при оценке качества задач классификации. Таблица сопряженности (матрица ошибок). Дать определение понятий чувствительности и специфичности, ROC-кривой, критериев оценки качества классификации.

25. Дать определение понятия нейронной сети. Характеризовать архитектуру нейронной сети.

26. Сделать обзор основ машинного обучения. Характеризовать организацию оценки моделей машинного обучения, предназначение тренировочных, проверочных и контрольных данных, понятие функции потерь.

27. Характеризовать механизм нейронных сетей на основе обучения, метод обратного распространения ошибки.

28. Продемонстрировать решение задач классификации, регрессии, прогнозирования с помощью нейронных сетей в Python, Orange.

29. Дать определения понятий поверхностного и глубокого обучения, глубокого обучения в технологиях компьютерного зрения.

30. Характеризовать содержание глубокого обучения для текста и последовательностей. Сделать обзор рекуррентных нейронных сетей, слоев LSTM, GRU.