

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
ПОВОЛЖСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ



УТВЕРЖДАЮ
Директор ИЛП

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

29.02.2024 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

Б.1.1.25 Введение в анализ данных

(код и наименование дисциплины по учебному плану)

Направление подготовки
(специальность)

19.03.01 Биотехнология

Квалификация выпускника

Бакалавр

(бакалавр/магистр/специалист)

Направленность

Биотехнология

Курс 3
Семестр 5

Распределение учебного времени

Трудоемкость по учебному плану	108 / 3	часов/зачетных единиц
Лекции	18	часов
Лабораторные работы	54	часов
Практические занятия	-	часов
Иная контактная работа	-	часов
Всего контактной работы (без учета экз.)	72	часов
Контактная работа по экзамену	-	часов
Курсовой проект (работа)	-	семестр
Самостоятельная работа обучающихся (без учета экз.)	36	часов
Самостоятельная работа по подготовке к экзамену	-	часов
Экзамен	-	семестр
Зачет	5	семестр
БРК, ДЗ	-	семестр

(год)

Программа составлена в соответствии с требованиями ФГОС ВО направления подготовки (специальности) 19.03.01 Биотехнология

Программу составили:

доцент с ученой степенью кандидата наук	Информатики	СОГЛАСОВАНО	Ю.А. Ипатов
(должность)	(кафедра)		(И.О. Фамилия)

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

		(наименование кафедры)	
12.02.2024	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.И. Мухортов
		(И.О. Фамилия)

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

СОГЛАСОВАНО	Д.И. Мухортов
	(И.О. Фамилия)

Эксперт(ы): Чикилев Виталий Алексеевич, Директор ООО «Казанское»

Рабочая программа проверена и зарегистрирована в УМЦ 12.03.2024 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. ОПК-2 Способен осуществлять поиск, хранение, обработку и анализ профессиональной информации из различных источников и баз данных, представлять её в требуемом формате с использованием информационных, компьютерных и сетевых технологий, включая проведение расчетов и моделирование, с учетом основных требований информационной безопасности	ОПК-2.1 Знает способы и методы поиска, хранения, обработки и анализа информации, представления информации в требуемом формате с использованием информационных компьютерных и сетевых технологий	знания: методологические подходы извлечения знаний из данных и основы применения соответствующих инструментальных средств умения: умеет выбрать соответствующие инструментальные средства и платформы навыки: имеет навык выбора инструментальных средств и платформ, основанных на извлечении знаний, для решения поставленных задач
	ОПК-2.2 Умеет применять способы и методы поиска, хранения, обработки и анализа информации, представления информации в требуемом формате с использованием информационных компьютерных и сетевых технологий	знания: принципы и методы машинного обучения, типы и классы задач машинного обучения умения: сопоставлять задачам предметной области классы задач машинного обучения навыки: постановки задач по разработке или совершенствованию методов и алгоритмов для решения комплекса задач
	ОПК-2.3 Владеет навыками поиска, хранения, обработки и анализа информации, представления информации в требуемом формате с использованием информационных компьютерных и сетевых технологий	знания: функциональные возможности современных инструментальных средств и систем программирования в области создания моделей и методов машинного обучения умения: проводить сравнительный анализ и осуществлять выбор инструментальных средств для решения задач машинного обучения, в частности для решения задач распознавания объектов по звуку и по изображениям навыки: разработки требований к архитектуре комплексных систем искусственного интеллекта со стороны заказчика

Раздел 2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОПОП

Дисциплина относится к обязательной части ОПОП.

Дисциплина является обязательной

Для продолжения формирования заявленных компетенций необходимы знания предшествующих дисциплин: Информационные технологии (ОПК-2)

Изучаемая дисциплина является основой для продолжения формирования указанных

компетенций в следующих государственной итоговой аттестации в форме: Подготовка к процедуре защиты и защита выпускной квалификационной работы (ОПК-2)

Раздел 3. ОПИСАНИЕ ОБРАЗОВАТЕЛЬНЫХ ТЕХНОЛОГИЙ

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

Основными стратегическими технологиями являются: имитационное моделирование, исследовательские, лекционные занятия, практические занятия, процедуры самообучения

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

Раздел 4. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

5 семестр

Виды и темы занятий	Количество часов	Формируемые компетенции
Введение в анализ данных на Orange Data Mining	30	ОПК-2
Лекция. Лекции №1 (Предобработка и подготовка данных)	2	
Лекция. Лекции №2 (Вычисление и сравнение статистических характеристик различных массивов данных)	2	
Лекция. Лекции №3 (Визуализация данных)	2	
Лабораторная работа. Лабораторная работа №1 (Инструменты предварительной обработки данных)	3	
Лабораторная работа. Лабораторная работа №2 (Форматы входных данных и операция импортирования)	3	
Лабораторная работа. Лабораторная работа №3 (Вычисление и сравнение статистических характеристик различных массивов данных)	3	
Лабораторная работа. Лабораторная работа №4 (Визуализация данных)	3	
Задания для самостоятельной работы, в том числе выполнение Предварительная обработка Визуализация данных Меры центральной тенденции	12	
Регрессионный анализ	38	ОПК-2
Лекция. Лекции 4 (Задача регрессионного анализа данных)	2	
Лекция. Лекции 5 (Линейная модель регрессии)	2	
Лекция. Лекции 6 (Логистическая модель регрессии)	2	
Лабораторная работа. Лабораторная работа №5 (Подготовка базы данных для задачи регрессионного анализа)	4	
Лабораторная работа. Лабораторная работа №6 (Линейная модель регрессии)	4	
Лабораторная работа. Лабораторная работа №7 (Логистическая модель регрессии)	4	
Лабораторная работа. Лабораторная работа №8 (Корреляционный анализ данных)	4	
Лабораторная работа. Лабораторная работа №9 (Мини проект по индивидуальному варианту)	4	

Задания для самостоятельной работы, в том числе выполнение Понятие временного ряда Интерполяция функций Экстраполяция функций (регрессионные модели) Авторегрессионные модели Корреляция	12	
Кластеризация и классификация	40	ОПК-2
Лекция. Лекции 7 (Задача классификации данных)	2	
Лекция. Лекции 8 (Модели классификации данных)	2	
Лекция. Лекции 9 (Кластеризация данных. Меры расстояния)	2	
Лабораторная работа. Лабораторная работа №10 (Классификация данных)	3	
Лабораторная работа. Лабораторная работа №11 (Выбор моделей для классификации)	3	
Лабораторная работа. Лабораторная работа №12 (Оценка качества обученных моделей)	3	
Лабораторная работа. Лабораторная работа №13 (Метод К- средних)	3	
Лабораторная работа. Лабораторная работа №14 (Метод иерархической кластеризации)	3	
Лабораторная работа. Лабораторная работа №15 (Мини проект на классификацию данных)	3	
Лабораторная работа. Лабораторная работа №16 (Мини проект на кластеризацию данных)	4	
Задания для самостоятельной работы, в том числе выполнение РГР Методы решения задачи классификации Классификация с подкреплением Кластеризация с подкреплением	12	
Иная контактная работа:	0	

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

системам, профессиональным базам данных и информационным справочным системам. Изучение дисциплины включает выполнение **расчётно-графической работы**. Периодичность проведения, формы текущего контроля успеваемости, система оценивания хода освоения дисциплин представлены в рабочей программе. Формой промежуточной аттестации по дисциплине является **зачет**.

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.	Кревецкий, Александр Владимирович. Основы технологий искусственного интеллекта [Электронный ресурс] : учебное пособие / А. В. Кревецкий, Ю. А. Ипатов, Н. И. Роженцова ; под общей редакцией А. В. Кревецкого; Министерство науки и высшего образования Российской Федерации, ФГБОУ ВО "Поволжский государственный технологический университет". Йошкар-Ола: ПГТУ, 2023. - 272 с. ISBN 978-5-8158-2358-	https://portal.volgatech.net/books/Krevetskiy_Osnovy_tekhnologii_iskusstvennogo_intellekta_2023.pdf
2.	Рутковская, Данута. Нейронные сети, генетические алгоритмы и нечеткие системы [Текст] / Д. Рутковская, М. Пилиньский, Л. Рутковский ; пер. с пол. И. Д. Рудинского. М.: Горячая линия - Телеком, 2007. - 383 с. ISBN 5-93517-103-1. Экземпляры: всего 4.	4
3.	Барский, Аркадий Бенционович. Нейронные сети [Текст] : распознавание, управление, принятие решений / А. Б. Барский. М.: Финансы и статистика, 2004. - 175 с. ISBN 5-279-02757-X. Экземпляры: всего 9.	9
4.	Макшанов, А. В. Технологии интеллектуального анализа данных [Электронный ресурс] : учебное пособие / А. В. Макшанов, А. Е. Журавлев. 2-е изд., стер. Санкт-Петербург: Лань, 2022. - 212 с. ISBN 978-5-8114-4493-9.	https://e.lanbook.com/book/206711

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.	120 (II)	Доска классная 1.0*1.5 (1), ПК S404,2 400W/Intel Core i3 540/клав.,мышь,монит. 21,5" VA2248-LED (15), 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	Зачтено

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. Машинное обучение - это ...

- научное направление, задачей которого является создание интеллектуальных систем, лежит на стыке информатики, статистики и анализа данных, а также занимается вопросами, связанными с философией и этичностью использования интеллектуальных систем
- объединение ряда научных областей, занимающихся построением систем анализа и обработки

данных

- математическая область, связанная с построением предсказательных алгоритмов (как правило представленных статистическими моделями) на основе данных
- прикладное направление, развивающее методы построения баз знаний и правил, явным образом описывающих знания экспертов

2. Какие задачи из ниже перечисленных относятся к задачам классификации?

- определение наиболее целесообразного способа лечения
- определение длительности и исхода заболевания
- задачи поискового вывода
- оценивание кредитоспособности заёмщика

3. Определения дохода клиента банка является задачей ...

- регрессии
- кластеризации
- классификации

4. В чём заключается задача кластеризации?

- задача машинного обучения, в которой метки объектов принимают ограниченное число значений, например, город проживания, пол клиента
- задача машинного обучения, в которой метки объектов принимают любое численное значение, например, стоимость квартиры, сумма кредита
- задача машинного обучения, заключающаяся в объединении похожих объектов в однородные группы

5. Что называют обучением с подкреплением?

- обучение с контролирующими параметрами
- обучение с большим количеством учителей
- обучение основанное на собственном опыте

6. У машинного обучения есть ряд задач. Как называется та, что направлена на предсказание значения той или иной непрерывной числовой величины для входных данных?

- Переобучение
- Классификация
- Регрессия
- Кластеризация

7. Нейросети хорошо проявляют себя не только в распознавании, но и в генерации изображений. Но кое с чем у них все-таки возникают проблемы. С чем именно?

- Цвет
- Текстуры

- Формы

- Глубина, количество пикселей

8. Какой из видов машинного обучения основывается на взаимодействии обучаемой системы со средой?

- Обучение с подкреплением

- Глубинное обучение

- Обучение с учителем

- Обучение без учителя

9. Когда говорят о нейронных сетях и машинном обучении, часто упоминают закон Мура. В чем его суть?

- 20% усилий дают 80% результата, а остальные 80% усилий — лишь 20% результата

- Если все слова языка или длинного текста упорядочить по убыванию частоты их использования, то частота n -го слова в таком списке окажется приблизительно обратно пропорциональной его порядковому номеру n

- Не следует множить сущее без необходимости

- Каждое следующее поколение компьютеров работает в 2,5 раза быстрее

10. Какой тип искусственной нейронной сети представлен на картинке?

- Нейронная сеть Джордана

- Рекуррентная нейронная сеть

- Простая нейронная сеть

- Сверточная нейронная сеть

Перечень вопросов для проведения промежуточной аттестации

1. Найдите значения коэффициентов c_1 и c_2 для функции $y=c_1+c_2x$, если известно, что функция экстраполирует данные, приведенные в таблице.

2. Постройте экстраполяционную модель с помощью кривых Гомперца и Перла, описывающую данные Росстата о количестве автомобилей на тысячу человек при условии, что значение должно приближаться к значению 1000.

3. С помощью теории автоматов опишите работу вендинговой машины (автоматапродавца) по продаже двух товаров по цене 5 и 10 рублей, при условии, что машина принимает монеты номиналов 1, 2, 5 и 10 рублей. Учтите, что выдача товара должна происходить при нажатии кнопки выбора соответствующего товара. Усложните автомат таким образом, чтобы он выдавал сдачу.

4. С помощью нейронной сети опишите работу вендинговой машины (автоматапродавца) по продаже двух товаров по цене 5 и 10 рублей, при условии, что машина принимает монеты номиналов 1, 2, 5 и 10 рублей. Учтите, что выдача товара должна происходить при нажатии кнопки выбора соответствующего товара. Усложните автомат таким образом, чтобы он выдавал сдачу.

5. На рисунке приведены объекты, относящиеся к разным классам (треугольники и круги), определяемые на основе значений пары параметров y и x . Запишите функцию $y = f(x)$, которая

позволит разделять приведенные данные на классы.

6. Постройте матрицу признаков для экспертной системы матричного типа, которая позволит идентифицировать несколько простудных заболеваний (например, ОРЗ, грипп, ангина) по таким признакам, как температура тела, головная боль, кашель, ломота в суставах, сонливость, насморк, боль в горле.

7. Проверьте модели, полученные в результате выполнения заданий, с помощью критериев согласия Фишера, Пирсона, Стьюдента и Колмогорова — Смирнова. Сравните полученные результаты, сделайте выводы о применимости критериев.

8. Проверьте модели, полученные в результате выполнения заданий, с использованием Accurasy, F-меры и логистической функции потерь. Сравните результаты, сделайте выводы о применимости критериев оценки моделей.

