

МИНОБРНАУКИ РОССИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ
«ВОРОНЕЖСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
(ФГБОУ ВО «ВГУ»)

УТВЕРЖДАЮ

Заведующий кафедрой
цифровых технологий



/ Кургалин С.Д.

28.02.2022 г.

РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ

Б1.В.01 ВВЕДЕНИЕ В МАШИННОЕ ОБУЧЕНИЕ И АНАЛИЗ ДАННЫХ

1. Код и наименование направления подготовки/специальности:

02.03.01 Математика и компьютерные науки

2. Профиль подготовки/специализация:

Квантовая теория информации

Распределенные системы и искусственный интеллект

3. Квалификация выпускника: бакалавр

4. Форма обучения: очная

5. Кафедра, отвечающая за реализацию дисциплины: цифровых технологий

6. Составители программы:

Крыловецкий Александр Абрамович, к.ф.-м.н., доцент

7. Рекомендована: НМС ФКН (протокол № 3 от 25.02.2022)

8. Учебный год: 2024-2025 Семестр: 5

9. Цели и задачи учебной дисциплины

Целью дисциплины является формирование знаний, умений и компетенций в области машинного обучения и анализа данных.

Задачи дисциплины:

- изучение различных математических моделей данных и алгоритмов анализа данных;
- формирование практических навыков реализации алгоритмов машинного обучения на языке программирования Python с использованием пакетов NumPy, Pandas, Matplotlib и Scikit-Learn.

10. Место учебной дисциплины в структуре ООП:

Дисциплина относится к части блока Б1, формируемой участниками образовательных отношений. Для успешного освоения дисциплины требуется предварительное изучение математического анализа, линейной алгебры, аналитической геометрии.

11. Планируемые результаты обучения по дисциплине/модулю (знания, умения, навыки), соотнесенные с планируемыми результатами освоения образовательной программы (компетенциями) и индикаторами их достижения:

Код	Название компетенции	Код(ы)	Индикатор(ы)	Планируемые результаты обучения
ПК-1	Способен демонстрировать базовые знания математических и естественных наук, основ программирования и информационных технологий.	ПК-1.1	Обладает базовыми знаниями, полученными в области математических и (или) естественных наук, программирования и информационных технологий.	Знать: особенности различных областей приложения методов машинного обучения.
		ПК-1.2	Умеет находить, формулировать и решать стандартные задачи в собственной научно-исследовательской деятельности в математике и информатике.	Уметь: применять алгоритмы машинного обучения для анализа данных.
		ПК-1.3	Имеет практический опыт научно-исследовательской деятельности в математике и информатике.	Владеть: методами машинного обучения и подходами к их верификации.
ПК-3	Способен создавать и исследовать новые математические модели в естественных науках, промышленности и бизнесе, с учетом возможностей современных информационных технологий и программирования	ПК-3.1	Знает основные методы проектирования и производства программного продукта, принципы построения, структуры и приемы работы с инструментальными средствами, поддерживающими создание программных продуктов и программных комплексов, их сопровождения, администрирования и развития (эволюции).	Знать: основные алгоритмы машинного обучения.
		ПК-3.2	Умеет использовать методы проектирования и	Уметь: анализировать применимость алгоритмов

	ия и компьютерной техники.		производства программного продукта, принципы построения, структуры и приемы работы с инструментальными средствами, поддерживающими создание программного продукта.	машинного обучения к конкретной задаче.
		ПК-3.3	Имеет практический опыт применения указанных выше методов и технологий.	Владеть: технологиями применения методов машинного обучения в различных областях.
ПК-4	Способен использовать современные методы разработки и реализации конкретных алгоритмов математических моделей на базе языков программирования и пакетов прикладных программ моделирования.	ПК-4.1	Знает современные методы разработки и реализации алгоритмов математических моделей на базе языков и пакетов прикладных программ моделирования.	Знать: математические основы алгоритмов классификации, кластеризации и регрессии.
		ПК-4.2	Умеет разрабатывать и реализовывать алгоритмы математических моделей на базе языков и пакетов прикладных программ моделирования.	Уметь: формализовывать данные для разработки приложений.
		ПК-4.3	Имеет практический опыт разработки и реализации алгоритмов на базе языков и пакетов прикладных программ моделирования.	Владеть: методами разработки моделей конкретной задачи для применения методов машинного обучения.

12. Объем дисциплины в зачетных единицах/час. — 3/108.

Форма промежуточной аттестации 5 семестр – экзамен.

13. Трудоемкость по видам учебной работы

Вид учебной работы		Трудоемкость	
		Всего	По семестрам
			5 сем.
Аудиторные занятия		50	50
в том числе:	лекции	34	34
	практические		
	лабораторные	16	16
Самостоятельная работа		22	22
Экзамен		36	36
Итого:		108	108

13.1. Содержание дисциплины

№ п/п	Наименование раздела дисциплины	Содержание раздела дисциплины	Реализация раздела дисциплины с помощью онлайн-курса, ЭУМК*
1. Лекции			
1.1	Введение. Оболочка IPython, Jupiter	Введение. Горячие клавиши, магические команды. Ошибки и отладка	https://edu.vsu.ru/course/view.php?id=7044

1.2	Библиотека NumPy	Типы данных, массивы, вычисления над массивами. Агрегирование, транслирование, сравнение, булева логика. Индексация, сортировка. Структурированные данные.	https://edu.vsu.ru/course/view.php?id=7044
1.3	Библиотека Pandas	Объекты. Индексация и выборка данных. Операции над данными. Индексация. Объединение. Агрегирование. Векторизованные операции. Временные ряды.	https://edu.vsu.ru/course/view.php?id=7044
1.4	Библиотека Matplotlib	Графики и диаграммы. Гистограммы. Погрешности. Настройки легенд и шкал цветов. Трехмерные графики.	https://edu.vsu.ru/course/view.php?id=7044
1.5	Библиотека Scikit-Learn и машинное обучение	Гиперпараметры и проверка модели. Проектирование признаков. Наивная байесовская классификация. Линейная регрессия. Метод опорных векторов. Деревья решения и случайные леса. Метод главных компонент. Метод k-средних. Смеси Гауссовых распределений. Ядерная оценка плотности распределения. Конвейер распознавания лиц.	https://edu.vsu.ru/course/view.php?id=7044
2. Лабораторные занятия			
2.1	Введение. Оболочка IPython, Jupiter	Введение. Горячие клавиши, магические команды. Ошибки и отладка	https://edu.vsu.ru/course/view.php?id=11173
2.2	Библиотека NumPy	Типы данных, массивы, вычисления над массивами. Агрегирование, транслирование, сравнение, булева логика. Индексация, сортировка. Структурированные данные.	https://edu.vsu.ru/course/view.php?id=11173
2.3	Библиотека Pandas	Объекты. Индексация и выборка данных. Операции над данными. Индексация. Объединение. Агрегирование. Векторизованные операции. Временные ряды.	https://edu.vsu.ru/course/view.php?id=11173
2.4	Библиотека Matplotlib	Графики и диаграммы. Гистограммы. Погрешности. Настройки легенд и шкал цветов. Трехмерные графики.	https://edu.vsu.ru/course/view.php?id=11173
2.5	Библиотека Scikit-Learn и машинное обучение	Гиперпараметры и проверка модели. Проектирование признаков. Наивная байесовская классификация. Линейная регрессия. Метод опорных векторов. Деревья решения и случайные леса. Метод главных компонент. Метод k-средних. Смеси Гауссовых распределений. Ядерная оценка плотности распределения. Конвейер распознавания лиц.	https://edu.vsu.ru/course/view.php?id=11173

13.2. Темы (разделы) дисциплины и виды занятий

№ п/п	Наименование темы (раздела) дисциплины	Виды занятий (количество часов)				Всего
		Лекции	Практические	Лабораторные	Самостоятельная работа	
1	Введение. Оболочка IPython, Jupiter	2		2	2	6
2	Библиотека NumPy	4		2	2	8
3	Библиотека Pandas	6		2	2	10
4	Библиотека Matplotlib	4		2	2	8
5	Библиотека Scikit-Learn и машинное обучение	18		8	14	40
	Итого:	34		16	22	72

14. Методические указания для обучающихся по освоению дисциплины:

Освоение дисциплины складывается из аудиторной работы (учебной деятельности, выполняемой под руководством преподавателя) и внеаудиторной работы (учебной деятельности, реализуемой обучающимся самостоятельно).

Аудиторная работа состоит из работы на лекциях и выполнения практических (или лабораторных) заданий в объеме, предусмотренном учебным планом. Лекция представляет собой последовательное и систематическое изложение учебного материала, направленное на знакомство обучающихся с основными понятиями и теоретическими положениями изучаемой дисциплины. Лекционные занятия формируют базу для практических (или лабораторных) занятий, на которых полученные теоретические знания применяются для решения конкретных практических задач. Обучающимся для успешного освоения дисциплины рекомендуется вести конспект лекций и практических (лабораторных) занятий.

Самостоятельная работа предполагает углублённое изучение отдельных разделов дисциплины с использованием литературы, рекомендованной преподавателем, а также конспектов лекций, презентационным материалом (при наличии) и конспектов практических (лабораторных) занятий. В качестве плана для самостоятельной работы может быть использован раздел 13.1 настоящей рабочей программы, в котором зафиксированы разделы дисциплины и их содержание. В разделе 13.2 рабочей программы определяется количество часов, отводимое на самостоятельную работу по каждому разделу дисциплины. Больше количество часов на самостоятельную работу отводится на наиболее трудные разделы дисциплины. Для самостоятельного изучения отдельных разделов дисциплины используется перечень литературы и других ресурсов, перечисленных в пунктах 15 и 16 настоящей рабочей программы.

Успешность освоения дисциплины определяется систематичностью и глубиной аудиторной и внеаудиторной работы обучающегося.

При использовании дистанционных образовательных технологий и электронного обучения выполнять все указания преподавателей, вовремя подключаться к online занятиям, ответственно подходить к заданиям для самостоятельной работы.

15. Перечень основной и дополнительной литературы, ресурсов интернет, необходимых для освоения дисциплины

а) основная литература:

№ п/п	Источник
1	Баланов, А. Н. Машинное обучение и искусственный интеллект : учебное пособие для вузов / А. Н. Баланов. — Санкт-Петербург : Лань, 2024. — 172 с. — ISBN 978-5-507-49194-0. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/414920
2	Митина, О. А. Технологии и инструментарий машинного обучения : учебное пособие / О. А. Митина, В. В. Жаров. — Москва : РТУ МИРЭА, 2023. — 203 с. — ISBN 978-5-7339-1758-0. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/368633
3	Вольфсон, М. Б. Анализ данных : учебно-методическое пособие / М. Б. Вольфсон. — Санкт-Петербург : СПбГУТ им. М.А. Бонч-Бруевича, 2023. — 69 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/381533

б) дополнительная литература:

№ п/п	Источник
1	Котиков, П. Е. Анализ данных : учебно-методическое пособие / П. Е. Котиков. — Санкт-Петербург : СПбГПМУ, 2019. — 48 с. — ISBN 978-5-907184-46-6. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/174498
2	Хачумов, М. В. Введение в интеллектуальный анализ данных : учебное пособие / М. В. Хачумов. — Москва : РТУ МИРЭА, 2023. — 123 с. — ISBN 978-5-7339-2073-3. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/398240

в) информационные электронно-образовательные ресурсы (официальные ресурсы интернет)*:

№ п/п	Ресурс
1	ЗНБ ВГУ: https://lib.vsu.ru/
2	Электронно-библиотечная система "Университетская библиотека online": http://biblioclub.ru/
3	Электронно-библиотечная система "Лань": https://e.lanbook.com/

16. Перечень учебно-методического обеспечения для самостоятельной работы

№ п/п	Источник
1	Баланов, А. Н. Машинное обучение и искусственный интеллект : учебное пособие для вузов / А. Н. Баланов. — Санкт-Петербург : Лань, 2024. — 172 с. — ISBN 978-5-507-49194-0. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/414920
2	Вольфсон, М. Б. Анализ данных : учебно-методическое пособие / М. Б. Вольфсон. — Санкт-Петербург : СПбГУТ им. М.А. Бонч-Бруевича, 2023. — 69 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/381533

17. Образовательные технологии, используемые при реализации учебной дисциплины, включая дистанционные образовательные технологии (ДОТ, электронное обучение (ЭО), смешанное обучение):

При реализации дисциплины могут использоваться технологии электронного обучения и дистанционные образовательные технологии на базе портала edu.vsu.ru, а также другие доступные ресурсы сети Интернет.

18. Материально-техническое обеспечение дисциплины:

Аудитория для лекционных занятий: мультимедиа-проектор, экран для проектора, компьютер с выходом в сеть «Интернет». Специализированная мебель (столы ученические, стулья, доска). Программное обеспечение: LibreOffice v.5-7, программа для просмотра файлов формата pdf, браузер.

Компьютерный класс: специализированная мебель, персональные компьютеры на базе i3-9100-3,6ГГц, мониторы ЖК 19» (30 шт.), мультимедийный проектор, экран.

ПО: ОС Windows v.7, 8, 10, Набор утилит интерпретатор языка CPython, интерпретатор языка Anaconda, IDE PyCharm, редактор Jupiter.

19. Оценочные средства для проведения текущей и промежуточной аттестаций

Порядок оценки освоения обучающимися учебного материала определяется содержанием следующих разделов дисциплины:

№ п/п	Наименование раздела дисциплины (модуля)	Компетенция(и)	Индикатор(ы) достижения компетенции	Оценочные средства
1	Введение. Оболочка IPython, Jupiter	ПК-1 ПК-3 ПК-4	ПК-1.1, ПК-1.2, ПК-1.3 ПК-3.1, ПК-3.2, ПК-3.3 ПК-4.1, ПК-4.2, ПК-4.3	Лабораторные работы
2	Библиотека NumPy	ПК-1 ПК-3 ПК-4	ПК-1.1, ПК-1.2, ПК-1.3 ПК-3.1, ПК-3.2, ПК-3.3 ПК-4.1, ПК-4.2, ПК-4.3	Лабораторные работы
3	Библиотека Pandas	ПК-1 ПК-3 ПК-4	ПК-1.1, ПК-1.2, ПК-1.3 ПК-3.1, ПК-3.2, ПК-3.3 ПК-4.1, ПК-4.2, ПК-4.3	Лабораторные работы

№ п/п	Наименование раздела дисциплины (модуля)	Компетенция(и)	Индикатор(ы) достижения компетенции	Оценочные средства
4	Библиотека Matplotlib	ПК-1 ПК-3 ПК-4	ПК-1.1, ПК-1.2, ПК-1.3 ПК-3.1, ПК-3.2, ПК-3.3 ПК-4.1, ПК-4.2, ПК-4.3	Лабораторные работы
5	Библиотека Scikit-Learn и машинное обучение	ПК-1 ПК-3 ПК-4	ПК-1.1, ПК-1.2, ПК-1.3 ПК-3.1, ПК-3.2, ПК-3.3 ПК-4.1, ПК-4.2, ПК-4.3	Лабораторные работы
Промежуточная аттестация форма контроля – экзамен				Перечень вопросов к экзамену

20. Типовые оценочные средства и методические материалы, определяющие процедуры оценивания

20.1. Текущий контроль успеваемости

Контроль успеваемости по дисциплине осуществляется с помощью лабораторных работ.

Перечень лабораторных работ

1. Библиотека NumPy
2. Библиотека Pandas
3. Библиотека Matplotlib
4. Библиотека Scikit-Learn
5. Наивная байесовская классификация.
6. Линейная регрессия.
7. Метод опорных векторов.
8. Деревья решения и случайные леса.
9. Метод главных компонент.
10. Метод k-средних.
11. Ядерная оценка плотности распределения.
12. Конвейер распознавания лиц.

Типовое задание для лабораторной работы

Лабораторная работа № 1 «Библиотека NumPy»

Цель работы: получение навыков работы с библиотекой NumPy.

Требования к выполнению работы: выполнение лабораторной работы предусматривает написание программы. Отчёт о работе проводится в виде собеседования и заключается в демонстрации работы программы, объяснении принципов работы алгоритма и ответов на дополнительные вопросы.

Критерии оценки: для получения оценки «зачтено» необходимо показать высокий уровень владения теоретическим материалом, уметь объяснить принцип работы написанной программы, верно ответить на дополнительные вопросы.

Задание: в одномерном массиве, состоящем из n вещественных элементов, вычислить:

- 1) Сумму отрицательных элементов.
- 2) Произведение элементов, расположенных между максимальным и минимальным элементами.

Упорядочить элементы массива по возрастанию.

20.2. Промежуточная аттестация

Промежуточная аттестация по дисциплине осуществляется с помощью следующих оценочных средств: перечень вопросов к экзамену.

Перечень вопросов к экзамену

Раздел 1. Введение.

1. Оболочка IPython, Jupiter.
2. Горячие клавиши, магические команды.
3. Ошибки и отладка

Раздел 2. Библиотека NumPy

1. Типы данных, массивы, вычисления над массивами.
2. Агрегирование, транслирование, сравнение, булева логика.
3. Индексация, сортировка.
4. Структурированные данные.

Раздел 3. Библиотека Pandas

1. Объекты. Индексация и выборка данных. Операции над данными.
2. Объединение. Агрегирование.
3. Векторизованные операции.
4. Временные ряды.

Раздел 4. Библиотека Matplotlib

1. Графики и диаграммы.
2. Гистограммы.
3. Погрешности.
4. Настройки легенд и шкал цветов.
5. Трёхмерные графики.

Раздел 5. Библиотека Scikit-Learn и машинное обучение

1. Гиперпараметры и проверка модели.
2. Проектирование признаков.
3. Наивная байесовская классификация.
4. Линейная регрессия.
5. Метод опорных векторов.
6. Деревья решения и случайные леса.
7. Метод главных компонент.
8. Метод k-средних.
9. Смеси Гауссовых распределений.
10. Ядерная оценка плотности распределения.
11. Конвейер распознавания лиц.

Для оценивания результатов обучения на экзамене используется 4-балльная шкала: «отлично», «хорошо», «удовлетворительно», «неудовлетворительно».

Критерии оценивания компетенций	Уровень сформированности компетенций	Шкала оценок
Дан полный, развёрнутый ответ на поставленный вопрос (вопросы), обучающийся свободно оперирует основными понятиями дисциплины, ориентируется в предметной области. Изложение материала не содержит ошибок, отличается последовательностью, грамотностью, логической стройностью.	Повышенный уровень	Отлично
Дан развёрнутый ответ на поставленный вопрос (вопросы), обучающийся свободно оперирует основными понятиями дисциплины, ориентируется в предметной области. Материал	Базовый уровень	Хорошо

изложен в целом последовательно и грамотно, отсутствуют грубые ошибки, однако имеются отдельные неточности в определениях, вычислениях, доказательствах, изложениях положений теории.		
Ответ на поставленный вопрос (вопросы) содержит изложение только базового теоретического материала, имеются ошибки в определениях, вычислениях, доказательствах, формулировках положений теории. Нарушена логическая последовательность в изложении материала.	Пороговый уровень	Удовлетворительно
Ответ на поставленный вопрос (вопросы) отсутствует, либо содержит грубые ошибки в определениях, вычислениях, доказательствах, формулировках положений теории. Обучающийся не владеет основными понятиями дисциплины. Отсутствует логическая последовательность в изложении материала.	–	Неудовлетворительно