

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

**федеральное государственное автономное
образовательное учреждение высшего образования
«Национальный исследовательский Нижегородский государственный университет
им. Н.И. Лобачевского»**

Арзамасский филиал ННГУ - Историко-филологический факультет

УТВЕРЖДЕНО

решением президиума Ученого совета ННГУ

протокол № 1 от 16.01.2024 г.

Рабочая программа дисциплины

Введение в анализ данных и искусственный интеллект

Уровень высшего образования

Бакалавриат

Направление подготовки / специальность

44.03.05 - Педагогическое образование (с двумя профилями подготовки)

Направленность образовательной программы

Русский язык и литература

Форма обучения

очная

г. Арзамас

2024 год начала подготовки

1. Место дисциплины в структуре ОПОП

Дисциплина ФТД.02 Введение в анализ данных и искусственный интеллект является факультативом в образовательной программе.

2. Планируемые результаты обучения по дисциплине, соотнесенные с планируемыми результатами освоения образовательной программы (компетенциями и индикаторами достижения компетенций)

Формируемые компетенции (код, содержание компетенции)	Планируемые результаты обучения по дисциплине (модулю), в соответствии с индикатором достижения компетенции		Наименование оценочного средства	
	Индикатор достижения компетенции (код, содержание индикатора)	Результаты обучения по дисциплине	Для текущего контроля успеваемости	Для промежуточной аттестации
ОПК-9: Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности	ИОПК 9.1: Знает принципы работы современных информационных технологий. ИОПК 9.2: Умеет осуществлять отбор современных информационных технологий для решения стандартных задач профессиональной деятельности. ИОПК 9.3: Владеет навыками использования современных информационных технологий для решения задач профессиональной деятельности	ИОПК 9.1: Знать принципы работы современных информационных технологий ИОПК 9.2: Уметь решать практические задачи с помощью современных информационных технологий и классификации программных средств ИОПК 9.3: Владеть навыками применения современных информационных технологий и классификации программных средств для решения практических задач, в том числе отечественного производства при решении задач профессиональной деятельности	Индивидуальное устное собеседование Практическое задание Тест	Зачёт: Контрольные вопросы

3. Структура и содержание дисциплины

3.1 Трудоемкость дисциплины

	очная
Общая трудоемкость, з.е.	1
Часов по учебному плану	36
в том числе	
аудиторные занятия (контактная работа):	

- занятия лекционного типа	8
- занятия семинарского типа (практические занятия / лабораторные работы)	8
- КСР	1
самостоятельная работа	19
Промежуточная аттестация	0 Зачёт

3.2. Содержание дисциплины

(структурированное по темам (разделам) с указанием отведенного на них количества академических часов и виды учебных занятий)

Наименование разделов и тем дисциплины	Всего (часы)	в том числе			
		Контактная работа (работа во взаимодействии с преподавателем), часы из них			Самостоятельная работа обучающегося, часы
		Занятия лекционного типа	Занятия семинарского типа (практические занятия/ лабора- торные работы), часы	Всего	
	о ф о	о ф о	о ф о	о ф о	о ф о
Тема 1. Постановки и примеры задач	6	2	0	2	4
Тема 2. Введение в язык Python	6	0	2	2	4
Тема 3. Описательная статистика и разведочный анализ данных	8	2	2	4	4
Тема 4. Задачи классификации и регрессии	8	2	2	4	4
Тема 5. Задачи обучения без учителя	7	2	2	4	3
Аттестация	0				
КСР	1			1	
Итого	36	8	8	17	19

Содержание разделов и тем дисциплины

Тема 1. Постановки и примеры задач.

Введение: основные понятия и задачи машинного обучения. Постановка задачи: определение целей и требований к анализу данных. Сбор данных: выбор источников, методов сбора и очистки данных. Обработка данных: преобразование и структуризация данных для анализа. Визуализация данных: представление данных в графическом виде для лучшего понимания. Анализ данных: применение статистических методов и машинного обучения для выявления закономерностей и зависимостей. Интерпретация результатов: объяснение полученных результатов и их применимость в контексте поставленной задачи.

Тема 2. Введение в программирование на языке Python. Установка программного обеспечения и основы синтаксиса языка Python. Базовые типы данных: числа, строки, списки, кортежи, словари, множества. Ввод и вывод в Python. Функции и итераторы в Python. Элементы функционального программирования. Классы и объекты в Python.

Тема 3. Описательная статистика и разведочный анализ данных. Табличный метод. Графический метод. Расчёт статистических показателей.

Тема 4. Задачи классификации и регрессии.

Задача классификации: определение принадлежности объекта к одному из нескольких классов на основе его характеристик. Логистическая регрессия. Деревья решений. Нейронные сети. Задача регрессии: прогнозирование значения непрерывной величины на основе входных данных. Линейная регрессия. Тема 5. Задачи обучения без учителя.

Обучение без учителя: определение и принципы работы. Преимущества и недостатки обучения без учителя. Задачи, решаемые с помощью обучения без учителя. Методы обучения без учителя: кластерный анализ; выявление аномалий; нахождение ассоциаций.

4. Учебно-методическое обеспечение самостоятельной работы обучающихся

Самостоятельная работа обучающихся включает в себя подготовку к контрольным вопросам и заданиям для текущего контроля и промежуточной аттестации по итогам освоения дисциплины приведенным в п. 5.

Для обеспечения самостоятельной работы обучающихся используются:

Электронные курсы, созданные в системе электронного обучения ННГУ:

Введение в анализ данных и искусственный интеллект, .

Иные учебно-методические материалы:

Учебно-методические документы, регламентирующие самостоятельную работу

адреса доступа к документам:

<https://arz.unn.ru/sveden/document/>

https://arz.unn.ru/pdf/Metod_all_all.pdf

5. Фонд оценочных средств для текущего контроля успеваемости и промежуточной аттестации по дисциплине (модулю)

5.1 Типовые задания, необходимые для оценки результатов обучения при проведении текущего контроля успеваемости с указанием критериев их оценивания:

5.1.1 Типовые задания (оценочное средство - Индивидуальное устное собеседование) для оценки сформированности компетенции ОПК-9:

1. Что такое искусственный интеллект?
2. Что такое слабый (специализированный или прикладной) искусственный интеллект?
3. Что такое сильный искусственный интеллект?
4. Что такое глубокое обучение?
5. Что такое экспертная система?
6. Что такое машинное обучение?
7. Что такое анализ данных?
8. Что такое обучение с учителем?
9. Что такое обучающая выборка?
10. Что такое обучение без учителя?
11. Что такое классификация?
12. Что такое регрессия?

13. Что такое кластеризация?
14. Что такое иерархическая кластеризация?
15. Основные библиотеки и программные средства для решения задач анализа данных и машинного обучения.

Критерии оценивания (оценочное средство - Индивидуальное устное собеседование)

Оценка	Критерии оценивания
зачтено	Студент дал развернутый ответ на все вопросы
не зачтено	Студент ответил только на часть вопросов или дал неразвернутый ответ на все вопросы

5.1.2 Типовые задания (оценочное средство - Практическое задание) для оценки сформированности компетенции ОПК-9:

Практические задания выполняются в виде Jupyter-ноутбука и представляет собой документ, содержащий описание всех этапов выполнения работы, код, результаты и выводы.

1. Эссе “Мое любимое приложение, использующее технологию искусственного интеллекта”

2. Практическое задание “Программа на Python “Угадай число”.

- a. Написать программу, которая умеет играть с человеком в игру “Угадай число”. Программа загадывает случайно число от 1 до 100. Человек должен угадать его. Для этого он делает несколько (не более 7) попыток, называя какое-то число в этом диапазоне. На каждую попытку человека компьютер говорит, угадал человек или нет, и в случае, если человек не угадал, больше или меньше загаданное число числа, названного человеком. Напишите такую программу. Для ввода информации человеком используйте функцию `input()`. Для вывода - `print()`.
- b. Реализуйте программу, играющую в игру "Угадай число", но человек и компьютер меняются ролями.

3. Практическое задание “Разведочный анализ данных”.

- a. Найдите интересные данные, с которыми вы хотите работать. Рекомендуемые ресурсы: <https://www.kaggle.com/datasets>, <https://www.kaggle.com/competitions>, <https://archive.ics.uci.edu/ml/index.php>. Рекомендуется работать с табличными данными в несколько десятков признаков (столбцов), несколько тысяч (строк).
- b. Опишите, что представляют из себя эти данные, какие признаки, почему они вам интересны.
- c. Примените к данным подходящие методы очистки (борьбы с выбросами и пропущенными значениями) и визуализации. Сделайте выводы.
- d. Сделайте выводы

4. Практическое задание “Решение задачи классификации и регрессии”

- a. Для данных из задания 3 определите, какой признак вы будете определять по остальным. Что это за задача: классификации или регрессии?
- b. Подготовьте данные (отмасштабируйте количественные и бинаризируйте категориальные признаки).
- c. Разбейте выборку на обучающую и тестовую.
- d. Обучите несколько методов машинного обучения (не менее трех, например, kNN, RandomForests, MLP). Сравните качество этих методов на обучающей и тестовой выборках.
- e. Сделайте выводы

5. Практическое задание “Решение задачи кластеризации”.

- a. Для данных из задания 3 сформулируйте задачу кластеризации.
- b. Обучите несколько методов кластеризации (не менее двух, например, k-means, DBSCAN). Объясните ваш выбор. Сравните результаты работы алгоритмов, а также полученные результаты с результатами работы методов обучения с учителем из задания 4.
- c. Сделайте выводы

Критерии оценивания (оценочное средство - Практическое задание)

Оценка	Критерии оценивания
зачтено	Все практические задания выполнены в полном объеме и в срок. Описание всех этапов выполнения заданий, код и результаты работы представлены преподавателю
не зачтено	Выполнены не все практические задания или выполнены не в полном объеме (представлено не полное описание этапов выполнения заданий, код работает некорректно, результаты работы не представлены преподавателю)

5.1.3 Типовые задания (оценочное средство - Тест) для оценки сформированности компетенции ОПК-9:

1. Что такое искусственный интеллект?

- Искусственный интеллект — это программа или устройство, имитирующее интеллектуальные способности человека.
- Искусственный интеллект — это кибер-механическое устройство, наполовину — человек, наполовину — машина. Под ним — боевое шасси из гиперсплава, управляемое микропроцессором, полностью бронированное, очень прочное. Но снаружи — живая человеческая ткань: плоть, кожа, волосы, кровь, выращенные для киборгов.
- Искусственный интеллект — это искусственная нейронная сеть, способная к обучению и самообучению.

2. Что такое слабый искусственный интеллект?

- Слабый искусственный интеллект - это программа, устройство или программно-аппаратный комплекс, умеющая решать только некоторый класс интеллектуальных задач.
- Слабый искусственный интеллект - это киборг модели от T-800 до T-888.
- Слабый искусственный интеллект - это однослойная искусственная нейронная сеть.

3. Что такое сильный искусственный интеллект?

- Сильный искусственный интеллект - это программа, устройство или программно-аппаратный комплекс, по своим интеллектуальным способностям ни в чем не уступающая человеку.
- Сильный искусственный интеллект - это киборг модели от T-900 и выше.
- Сильный искусственный интеллект - это любая глубокая нейронная сеть, умеющая решать задачи классификации и регрессии.

4. Что такое экспертная система?

- Экспертная система - это программная система позволяющая решать задачи искусственного интеллекта с помощью правил, сформулированных экспертом, составляющих базу знаний о предметной области.
- Экспертная система - это коллектив экспертов в данной предметной области, решающих интеллектуальные задачи.
- Экспертная система - это искусственная многослойная нейронная сеть.

5. Что такое машинное обучение?

- Машинное обучение - это подход в искусственном интеллекте, основанный на построении моделей (решающих правил) по данным.
- Машинное обучение - это система обучения при помощи информационных и электронных технологий.
- Машинное обучение - это целенаправленное объективное отражение действительности

6. Что такое анализ данных?

- Анализ данных - это извлечение закономерностей и знаний из данных.
- Анализ данных - это подход в машинном обучении, основанный на использовании глубоких нейронных сетей.
- Анализ данных - это извлечение закономерностей с помощью знаний, сформулированных экспертом.

7. Что такое обучающая выборка?

- Обучающая выборка - это набор объектов (входов) с известным для каждого объекта значением ответа (выхода, метки, целевой переменной).
- Обучающая выборка - это набор объектов, которые пользователь выбирает для своего обучения.
- Обучающая выборка - это набор объектов, которые человек удаляет из базы данных для эффективного обучения.

8. Что такое обучение с учителем?

- Обучение с учителем - это построение модели по данным.
- Обучение с учителем - это решение задач с помощью интеллектуального помощника.
- Обучение с учителем - это задача разбиения данных на группы похожих друг на друга объектов.

9. Что такое обучение без учителя?

- Обучение без учителя - это установление структурных свойств набора объектов.
- Обучение без учителя - это обучение человека или робота полезным навыкам без интеллектуальных помощников.
- Обучение без учителя - это задача определения по данным категориальных значений.

10. Что такое классификация?

- Классификация - это задача обучения с учителем, в которой ответ - это номер класса.
- Классификация - это задача определения по данным числовых значений.
- Классификация - это задача обучения без учителя, в которой необходимо разбить данные на группы похожих объектов.

11. Что такое регрессия?

- Регрессия - это задача обучения с учителем, в которой ответы принимают количественные (числовые) значения.
- Регрессия - это задача, в которой необходимо разбить множество объектов на группы похожих друг на друга объектов.
- Регрессия - это задача обучения с учителем, в которой необходимо определить по данным номер класса.

12. Что такое кластеризация?

- Кластеризация - это задача обучения без учителя, в которой необходимо разбить данные на группы похожих объектов.
- Кластеризация - это задача обучения с учителем, в которой необходимо определить по данным номер класса.
- Кластеризация - это задача обучения с учителем, в которой ответы принимают количественные значения.

13. Что такое иерархическая кластеризация?

- Иерархическая кластеризация - это задача обучения без учителя, в которой необходимо построить иерархию объектов, в которой вершине соответствует «кластер», содержащий все объекты, и на каждом уровне каждый кластер разбивается на два или более подкластеров.
- Иерархическая кластеризация - это задача классификации, в которой каждый объект может иметь несколько меток.
- Иерархическая кластеризация - это метод решения задачи кластеризации с помощью иерархии нейронных сетей.

Следующие вопросы предполагают ввод числа.

Ответ следует записать в скобках.

14. Рост детей в группе детского сада задается следующими значениями (в см.):

118, 112, 116, 114, 116, 115, 119, 111, 114, 117, 118

- Найдите медиану ()
- Найдите нижний квартиль ()
- Найдите верхний квартиль ()

15. Вес детей в группе детского сада задается значениями (в кг):

23, 23, 18, 18, 20, 19, 20, 17, 22, 20, 20

- Вычислите среднее ()

- Вычислите несмещенную оценку дисперсии s^2 ()
- На основе несмещенной оценки дисперсии вычислите оценку среднеквадратического отклонения s ()

16. На тестовой выборке, содержащей 2000 изображений котиков и собачек (950 котиков, 1050 собачек), нейронная сеть совершила 202 ошибки: 76 котиков были классифицированы как собачки, а 126 собачек – как котики. Считая положительным – класс собачек, вычислите метрики:

- Accuracy ()
- Error ()
- Precision ()
- Recall ()
- F1-мера ()

Ответ округлите до 3 правильных цифр после десятичной точки.

Критерии оценивания (оценочное средство - Тест)

Оценка	Критерии оценивания
зачтено	80-100 % правильных ответов
не зачтено	0-79 % правильных ответов

5.2. Описание шкал оценивания результатов обучения по дисциплине при промежуточной аттестации

Шкала оценивания сформированности компетенций

Уровень сформированности компетенций (индикатора достижения компетенций)	неудовлетворительно	удовлетворительно	хорошо	отлично
	не зачтено	зачтено		
<u>Знания</u>	Уровень знаний ниже минимальных требований. Имели место грубые ошибки	Минимально допустимый уровень знаний. Допущено много негрубых ошибок	Уровень знаний в объеме, соответствующем программе подготовки. Допущено несколько негрубых ошибок	Уровень знаний в объеме, соответствующем программе подготовки, без ошибок
<u>Умения</u>	При решении стандартных задач не продемонстрированы основные умения. Имели место грубые ошибки	Продemonстрированы основные умения. Решены типовые задачи с негрубыми ошибками. Выполнены все задания, но не в полном объеме	Продemonстрированы все основные умения. Решены все основные задачи с негрубыми ошибками. Выполнены все задания в полном объеме, но некоторые с недочетами	Продemonстрированы все основные умения. Решены все основные задачи с отдельными незначительными недочетами, выполнены все задания в полном объеме
<u>Навыки</u>	При решении стандартных задач не продемонстрированы базовые навыки.	Имеется минимальный набор навыков для решения стандартных задач с	Продemonстрированы базовые навыки при решении стандартных задач с	Продemonстрированы навыки при решении нестандартных задач без ошибок и недочетов

	Имели место грубые ошибки	некоторыми недочетами	некоторыми недочетами	
--	---------------------------	-----------------------	-----------------------	--

Шкала оценивания при промежуточной аттестации

Оценка		Уровень подготовки
зачтено	отлично	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «отлично», при этом хотя бы одна компетенция сформирована на уровне «отлично»
	хорошо	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «хорошо», при этом хотя бы одна компетенция сформирована на уровне «хорошо»
	удовлетворительно	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «удовлетворительно», при этом хотя бы одна компетенция сформирована на уровне «удовлетворительно»
не зачтено	неудовлетворительно	Хотя бы одна компетенция сформирована на уровне «неудовлетворительно».

5.3 Типовые контрольные задания или иные материалы, необходимые для оценки результатов обучения на промежуточной аттестации с указанием критериев их оценивания:

5.3.1 Типовые задания (оценочное средство - Контрольные вопросы) для оценки сформированности компетенции ОПК-9

1. Что такое искусственный интеллект?
2. Что такое слабый (специализированный или прикладной) искусственный интеллект?
3. Что такое сильный искусственный интеллект?
4. Что такое экспертная система?
5. Что такое машинное обучение?
6. Что такое анализ данных?
7. Что такое обучение с учителем?
8. Что такое обучающая выборка?
9. Что такое обучение без учителя?
10. Что такое классификация?
11. Что такое регрессия?
12. Что такое кластеризация?
13. Что такое иерархическая кластеризация?

Критерии оценивания (оценочное средство - Контрольные вопросы)

Оценка	Критерии оценивания
зачтено	Студент дал развернутый ответ на все вопросы.
не зачтено	Студент ответил только на часть вопросов или дал неразвернутый ответ на все вопросы.

6. Учебно-методическое и информационное обеспечение дисциплины (модуля)

Основная литература:

1. Смолин Д. В. Введение в искусственный интеллект: конспект лекций / Смолин Д. В. - 2-е изд., перераб. - Москва : ФИЗМАТЛИТ, 2007. - 264 с. - Библиогр.: доступна в карточке книги, на сайте ЭБС Лань. - Книга из коллекции ФИЗМАТЛИТ - Математика. - ISBN 978-5-9221-0862-1., <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=665793&idb=0>.
2. Воронина В. В. Теория и практика машинного обучения : учебное пособие / Воронина В. В. - Ульяновск : УлГТУ, 2017. - 290 с. - Библиогр.: доступна в карточке книги, на сайте ЭБС Лань. - Книга из коллекции УлГТУ - Информатика. - ISBN 978-5-9795-1712-4., <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=750984&idb=0>.

Дополнительная литература:

1. Басалин Павел Дмитриевич. Модели и методы интеллектуальной поддержки процессов принятия решений : учебно-методическое пособие / П. Д. Басалин, К. В. Безрук, М. В. Радаева ; ННГУ им. Н. И. Лобачевского. - Нижний Новгород : Изд-во ННГУ, 2018. - 134 с. - Текст : электронный., <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=822376&idb=0>.

Программное обеспечение и Интернет-ресурсы (в соответствии с содержанием дисциплины):

1. Python: <https://www.python.org/>
2. Anaconda: The Most Popular Python Data Science Platform
<https://www.anaconda.com/download/>
3. Google colaboratory colab.research.google.com

7. Материально-техническое обеспечение дисциплины (модуля)

Учебные аудитории для проведения учебных занятий, предусмотренных образовательной программой, оснащены мультимедийным оборудованием (проектор, экран), техническими средствами обучения, компьютерами.

Помещения для самостоятельной работы обучающихся оснащены компьютерной техникой с возможностью подключения к сети "Интернет" и обеспечены доступом в электронную информационно-образовательную среду.

Программа составлена в соответствии с требованиями ОС ННГУ по направлению подготовки/специальности 44.03.05 - Педагогическое образование (с двумя профилями подготовки).

Автор(ы): Золотых Николай Юрьевич, доктор физико-математических наук, доцент
Статуев Алексей Анатольевич, кандидат педагогических наук, доцент.

Заведующий кафедрой: Золотых Николай Юрьевич, доктор физико-математических наук.

Программа одобрена на заседании методической комиссии от 10.01.2024, протокол № 1.