

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

Юридический факультет

(факультет / институт / филиал)

УТВЕРЖДЕНО
президиумом Ученого совета ННГУ
протокол от
«14» декабря 2021 г. № 4

Рабочая программа дисциплины

**Введение в анализ данных и
искусственный интеллект**

(наименование дисциплины (модуля))

Уровень высшего образования
бакалавриат

(бакалавриат / магистратура / специалитет)

Направление подготовки / специальность

46.03.02 Документоведение и архивоведение

(указывается код и наименование направления подготовки / специальности)

Направленность образовательной программы
Документационное обеспечение управления

(указывается профиль / магистерская программа / специализация)

Форма обучения

Очная

(очная / очно-заочная / заочная)

Нижний Новгород

2022 год

1. Место дисциплины в структуре ОПОП

№ варианта	Место дисциплины в учебном плане образовательной программы	Стандартный текст для автоматического заполнения в конструкторе РПД
3	ФТД. Факультативы	Дисциплина <i>ФТД.04 Введение в анализ данных и искусственный интеллект</i> является факультативом в ООП специальности 46.03.02 Документоведение и архивоведение

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	
ОПК-4 - Способен использовать базовые знания в области информационно-коммуникативных технологий в сфере своей профессиональной деятельности;	ОПК-4.1. Владеет базовыми знаниями в области информационно-коммуникационных технологий в сфере документационного обеспечения управления и в архивном деле. ОПК-4.2. Использует программные продукты в управлении документами, применяет навыки ведения электронного документооборота, владеет технологиями сканирования документов. ОПК-4.3. Быстро адаптируется в использовании программ СЭД и их обновлении	Знать: основные методы, способы и средства поиска, получения, хранения (накопления), переработки информации Уметь: совершать различные манипуляции с информацией разными методами и способами, делать выводы и прогнозирование в рамках проанализированной информации Владеть: способностью анализировать, синтезировать, систематизировать информацию и прогнозировать дальнейшую деятельность учреждения	Собеседование Тест

3. Структура и содержание дисциплины

3.1. Трудоемкость дисциплины

	Очная форма обучения	Заочная форма обучения
Общая трудоемкость	1 ЗЕТ	1 ЗЕТ
Часов по учебному плану	36	36
в том числе		
аудиторные занятия (контактная работа)		
- занятия лекционного типа	8	8
- занятия семинарского типа	8	8
- текущий контроль (КСР)	1	1
самостоятельная работа	19	15
контроль		4
Промежуточная аттестация – зачет		

3.2. Содержание дисциплины

Наименование и краткое содержание разделов и тем дисциплины	Всего (часы)	В том числе				Самостоятельная работа обучающегося, часы	Самостоятельная работа обучающегося, часы
		Контактная работа (работа во взаимодействии с преподавателем), часы. Из них					
		Занятия лекционного типа	Занятия семинарского типа	Занятия лабораторного типа	Всего		
	Очная	Очная, заочная	Очная, заочная	Очная	Очная	Очная	Заочная
1. Постановки и примеры задач		2				4	3
2. Введение в язык Python			2			4	3
3. Описательная статистика и разведочный анализ данных		2	2			4	3
4. Задачи классификации и регрессии		2	2			4	3
5. Задачи обучения без учителя		2	2			4	3
Итого	36	8	8			20	15

Занятия семинарского типа организуются, в том числе в форме практической подготовки.

Практическая подготовка направлена на формирование и развитие знаний, умений и навыков применения современных информационных технологий, в том числе отечественного производства, при решении задач профессиональной деятельности.

Текущий контроль успеваемости реализуется в форме собеседования, тестирования и практических заданий.

Промежуточная аттестация проходит в традиционной форме (зачет).

4. Учебно-методическое обеспечение самостоятельной работы обучающихся

Самостоятельная работа студента при изучении дисциплины «Введение в анализ данных и искусственный интеллект» включает выполнение практических заданий под контролем преподавателя и подготовку к зачету.

Контрольные и тестовые вопросы, практические задания для проведения текущего контроля и промежуточной аттестации по итогам освоения дисциплины приведены в п. 5.2.

5. Фонд оценочных средств для промежуточной аттестации по дисциплине (модулю), включающий:

5.1. Описание шкал оценивания результатов обучения по дисциплине

Уровень сформированности компетенций (индикатора достижения компетенций)	Шкала оценивания сформированности компетенций						
	плохо	неудовлетворительно	удовлетворительно	хорошо	очень хорошо	отлично	превосходно
	Не зачтено		Зачтено				
<u>Знания</u>	Отсутствие знаний теоретического материала. Невозможность оценить полноту знаний вследствие отказа обучающегося от ответа	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. Допущено несколько несущественных ошибок	Уровень знаний в объеме, соответствующем программе подготовки, без ошибок.	Уровень знаний в объеме, превышающем программу подготовки.

<u>Умения</u>	Отсутствие минимальных умений. Невозможность оценить наличие умений вследствие отказа обучающегося от ответа	При решении стандартных задач не продемонстрированы основные умения. Имели место грубые ошибки.	Продемонстрированы основные умения. Решены типовые задачи с негрубыми и ошибками. Выполнены все задания, но не в полном объеме.	Продемонстрированы все основные умения. Решены все основные задачи с негрубыми ошибками. Выполнены все задания, в полном объеме, но некоторые с недочетами.	Продемонстрированы все основные умения. Решены все основные задачи. Выполнены все задания, в полном объеме, но некоторые с недочетами.	Продемонстрированы все основные умения, решены все основные задачи с отдельными несущественными недочетами, выполнены все задания в полном объеме.	Продемонстрированы все основные умения, решены все основные задачи. Выполнены все задания, в полном объеме без недочетов
<u>Навыки</u>	Отсутствие владения материалом. Невозможность оценить наличие навыков вследствие отказа обучающегося от ответа	При решении стандартных задач не продемонстрированы базовые навыки. Имели место грубые ошибки.	Имеется минимальный набор навыков для решения стандартных задач с некоторыми недочетами.	Продемонстрированы базовые навыки при решении стандартных задач с некоторыми недочетами	Продемонстрированы базовые навыки при решении стандартных задач без ошибок и недочетов.	Продемонстрированы навыки при решении нестандартных задач без ошибок и недочетов.	Продемонстрирован творческий подход к решению нестандартных задач.

Шкала оценки при промежуточной аттестации

Оценка		Уровень подготовки
зачтено	Превосходно	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «превосходно»
	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»
	Очень хорошо	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «очень хорошо», при этом хотя бы одна компетенция сформирована на уровне «очень хорошо»
	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», ни одна из компетенций не сформирована на уровне «плохо»
	Плохо	Хотя бы одна компетенция сформирована на уровне «плохо»

Критерий оценивания ответов на типовые контрольные вопросы для собеседования и вопросы к зачету

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

Критерий оценивания практических заданий

Результаты работы	Оценка
Все практические задания выполнены в полном объеме и в срок. Описание всех этапов выполнения заданий, код и результаты работы представлены преподавателю.	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).	не зачтено

Шкала оценки результатов тестирования

Баллы, %	Оценка сформированности компетенции
80-100	зачтено
0-79	не зачтено

5.2. Типовые контрольные задания или иные материалы, необходимые для оценки результатов обучения

5.2.1. Вопросы к зачету по дисциплине «Введение в анализ данных и искусственный интеллект» для оценки сформированности компетенции ОПК-4

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

5.2.2. Типовые контрольные вопросы для собеседования для оценки сформированности компетенции _ОПК-4 (индикатор компетенции ОПК – 4.1, ОПК – 4.2, ОПК – 4.3,)

На собеседовании проверяется, как студент освоил основные понятия. Он должен уметь давать развернутый ответ на следующие вопросы.

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.

5.2.3. Типовые практические задания для оценки сформированности компетенции _ОПК-4

Практические задания направлены на формирование и контроль у обучающихся индикаторов компетенций ОПК-4.2, ОПК-4.3. 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”.
 1. Найдите интересные данные, с которыми вы хотите работать. Рекомендуются 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).
 2. Опишите, что представляют из себя эти данные, какие признаки, почему они вам интересны.
 3. Примените к данным подходящие методы очистки (борьбы с выбросами и пропущенными значениями) и визуализации. Сделайте выводы.
 4. 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

5.2.4. Типовые тестовые вопросы для оценки сформированности компетенции _ОПК-4

Все вопросы альтернативные

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 коты. Считая положительным – класс собачек, вычислите метрики:

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

Ответ округлите до 3 правильных цифр после десятичной точки.

6. Учебно-методическое и информационное обеспечение дисциплины

а) основная литература:

1. Смолин, Д. В. Введение в искусственный интеллект: конспект лекций: учебное пособие / Д. В. Смолин. — 2-е изд., перераб. — Москва: ФИЗМАТЛИТ, 2007. — 264 с. — ISBN 978-5-9221-0862-1. — Текст: электронный // Лань: электронно-библиотечная система. — URL: <https://e.lanbook.com/book/2325> (дата обращения: 20.05.2022). — Режим доступа: для авториз. пользователей.
2. Воронина, В. В. Теория и практика машинного обучения: учебное пособие / В. В.

Воронина. — Ульяновск: УлГТУ, 2017. — 290 с. — ISBN 978-5-9795-1712-4. — Текст: электронный // Лань: электронно-библиотечная система. — URL: <https://e.lanbook.com/book/165053> (дата обращения: 23.05.2022). — Режим доступа: для авториз. пользователей.

б) дополнительная литература:

1. Басалин П.Д., Безрук К.В., Радаева М.В. Модели и методы интеллектуальной поддержки процессов принятия решений: Учебное пособие. – Нижний Новгород: Нижегородский госуниверситет, 2018. – 129 с. Фонд образовательных электронных ресурсов. Рег. № 1703.18.06. URL: <http://www.unn.ru/books/resources.html> – доступ свободный.

в) программное обеспечение и Интернет-ресурсы:

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

Помещения представляют собой учебные аудитории для проведения учебных занятий, предусмотренных программой (лекционного типа), оснащенные оборудованием и техническими средствами обучения.

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 специальности 46.03.02 Документоведение и архивоведение

Составитель: Н.Ю. Золотых

Заведующий кафедрой АГДМ: Н.Ю. Золотых

Программа одобрена на заседании методической комиссии юридического факультета от 13.12.2021 года, протокол № 4.