

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

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

Павловский филиал ННГУ

---

УТВЕРЖДЕНО

решением президиума Ученого совета ННГУ

протокол № 1 от 16.01.2024 г.

**Рабочая программа дисциплины**

Технологии обработки больших данных

---

Уровень высшего образования

Бакалавриат

---

Направление подготовки / специальность

09.03.03 - Прикладная информатика

---

Направленность образовательной программы

Прикладная информатика в экономике и управлении

---

Форма обучения

очная, очно-заочная

---

г. Павлово

2024 год начала подготовки

## 1. Место дисциплины в структуре ОПОП

Дисциплина Б1.В.ДВ.02.02 Технологии обработки больших данных относится к части, формируемой участниками образовательных отношений образовательной программы.

## 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
ПК-9: Способен моделировать прикладные (бизнес) процессы и объекты предметной области	ПК-9.1: Демонстрирует знание методических основ моделирования процессов и объектов предметной области ПК-9.2: Демонстрирует умение применения знаний к моделированию прикладных процессов и объектов предметной области при разработке программного обеспечения ИС ПК-9.3: Имеет практический опыт моделирования процессов и объектов на примере конкретной предметной области	ПК-9.1: Знать методические основы анализа данных и нейронных сетей  ПК-9.2: Уметь применять полученные знания при анализе данных и при разработке нейронных сетей  ПК-9.3: Владеть навыками анализа данных и разрабатывать нейронные сети	Опрос	Экзамен: Тест

## 3. Структура и содержание дисциплины

### 3.1 Трудоемкость дисциплины

	очная	очно-заочная
Общая трудоемкость, з.е.	3	3
Часов по учебному плану	108	108
в том числе		
аудиторные занятия (контактная работа):		
- занятия лекционного типа	32	16
- занятия семинарского типа (практические занятия / лабораторные работы)	16	12
- КСР	2	2
самостоятельная работа	22	42
Промежуточная аттестация	36	36

	Экзамен	Экзамен
--	---------	---------

### 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/лабораторные работы), часы		Всего				
	о Ф о	о З Ф о	о Ф о	о З Ф о	о Ф о	о З Ф о	о Ф о	о З Ф о	о Ф о	о З Ф о	
Тема 1. Введение в анализ данных	5	4	2	1	1	1	3	2	2	2	
Тема 2. Типы данных и источники	5	6	2	1	1	1	3	2	2	4	
Тема 3. Сбор и очистка данных	5	6	2	1	1	1	3	2	2	4	
Тема 4. Визуализация и анализ	6	5	2	1	2	1	4	2	2	3	
Тема 5. Основы статистики для анализа	6	5	2	1	2	1	4	2	2	3	
Тема 6. Введение в машинное обучение	7	5	4	2	1	1	5	3	2	2	
Тема 7. Основы нейронных сетей	7	6	4	2	1	1	5	3	2	3	
Тема 8. Архитектуры нейронных сетей	7	8	4	2	1	1	5	3	2	5	
Тема 9. Инструменты и библиотеки	6	6	2	1	2	1	4	2	2	4	
Тема 10. Применение нейросетей	8	10	4	2	2	1	6	3	2	7	
Тема 11. Проектный анализ и кейс-стади	8	9	4	2	2	2	6	4	2	5	
Аттестация	36	36									
КСР	2	2						2	2		
Итого	108	108	32	16	16	12	50	30	22	42	

### Содержание разделов и тем дисциплины

#### Тема 1. Введение в анализ данных

Основные понятия: данные, типы данных, этапы анализа. Роль анализа данных в современном мире.

Знакомство с задачами: классификация, регрессия, кластеризация.

#### Тема 2. Типы данных и источники

Структурированные и неструктурированные данные. Таблицы, текст, изображения. Основные источники: базы данных, API, сенсоры, лог-файлы, соцсети.

#### Тема 3. Сбор и очистка данных

Методы сбора, предварительная обработка, удаление шумов, пропусков, дубликатов. Практика: использование Python (pandas, numpy) для обработки данных.

#### Тема 4. Визуализация и анализ

Способы визуализации: графики, гистограммы, тепловые карты. Инструменты: Matplotlib, Seaborn. Как находить закономерности, аномалии и тренды.

#### Тема 5. Основы статистики для анализа

Среднее, мода, медиана, дисперсия, корреляция, стандартное отклонение. Как интерпретировать статистические показатели при анализе.

Тема 6. Введение в машинное обучение

Разница между анализом данных и ML. Обзор алгоритмов: линейная регрессия, дерево решений, k-ближайших соседей. Надзорное и ненадзорное обучение.

Тема 7. Основы нейронных сетей

Что такое нейрон, активация, слои, обучение. Архитектура перцептрона. Простейшая нейросеть и её обучение. Интуитивное понимание backpropagation.

Тема 8. Архитектуры нейронных сетей

Feedforward, сверточные (CNN), рекуррентные (RNN), трансформеры. Примеры их применения: изображение, текст, временные ряды.

Тема 9. Инструменты и библиотеки

Обзор популярных библиотек: Scikit-learn, TensorFlow, PyTorch, Keras. Практика создания моделей на примерах. Основы работы с Jupyter Notebook.

Тема 10. Применение нейросетей

Распознавание лиц, голосовых команд, перевод текста, генерация изображений. Как работает ChatGPT, DALL·E, Midjourney. Этика ИИ и нейросетей.

Тема 11. Проектный анализ и кейс-стади

Работа над мини-проектом: от постановки задачи до построения и оценки модели. Работа в командах. Разбор реальных кейсов из бизнеса и технологий.

Практические занятия /лабораторные работы организуются, в том числе, в форме практической подготовки, которая предусматривает участие обучающихся в выполнении отдельных элементов работ, связанных с будущей профессиональной деятельностью.

На проведение практических занятий / лабораторных работ в форме практической подготовки отводится: очная форма обучения - 24 ч., очно-заочная форма обучения - 14 ч.

#### **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:

Технологии обработки больших данных, <https://e-learning.unn.ru>.

Открытые онлайн-курсы MOOC:

Технологии обработки больших данных, <https://mooc.unn.ru>.

Иные учебно-методические материалы:

Онлайн-библиотеки и документация:

Документация: Pandas, Scikit-learn

Онлайн - библиотеки: PyTorch Tutorials, TensorFlow Guide, NumPy User Guide

Рекомендуемый формат подключения в курс:

Google Drive + Google Colab для онлайн-работы;

GitHub для совместных проектов;

Kaggle для соревновательной практики;  
Stepik или LMS-платформы для встроенного обучения.

## **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 - Опрос) для оценки сформированности компетенции ПК-9:**

Список вопросов для опроса студентов

Раздел 1: Общие понятия и основы анализа данных

Что такое анализ данных? Назовите этапы анализа.

Какие бывают типы данных и чем отличаются структурированные от неструктурированных?

Какие источники данных используются в анализе?

Что такое DataFrame в pandas?

Для чего используется библиотека NumPy?

Раздел 2: Предобработка и визуализация

Какие бывают методы очистки и обработки данных?

Что такое пропущенные значения и как их обрабатывать?

Назовите методы визуализации данных. В каких случаях их применяют?

Чем отличается график "boxplot" от "histogram"?

Как можно выявить выбросы в данных?

Раздел 3: Статистика и базовые методы

Что такое корреляция и как она интерпретируется?

Что такое среднее, мода и медиана?

Что показывает стандартное отклонение?

В чем отличие дисперсии от среднеквадратичного отклонения?

Раздел 4: Машинное обучение

Что такое машинное обучение? Какие его виды вы знаете?

Объясните разницу между надзорным и ненадзорным обучением.

Что такое регрессия и когда она применяется?

Назовите алгоритмы классификации, которые вы изучали.

Что такое переобучение модели и как его избежать?

Как разделяют данные на обучающую и тестовую выборки?

Раздел 5: Нейронные сети

Что такое нейронная сеть и как она устроена?

Объясните работу одного искусственного нейрона.

Какие функции активации вы знаете?

Что такое обратное распространение ошибки (backpropagation)?

Назовите типы нейросетей и области их применения.

В чем особенности сверточных (CNN) и рекуррентных (RNN) сетей?

Раздел 6: Применение и практика

Какие инструменты и библиотеки используются для анализа данных?

В чем преимущества Google Colab и Kaggle для анализа?

Приведите примеры применения нейросетей в реальной жизни.

Чем занимается специалист по анализу данных?

Критерии оценивания (оценочное средство - Опрос)

Оценка	Критерии оценивания
превосходно	Полное понимание материала
отлично	Почти безошибочное выполнение
очень хорошо	Небольшие неточности, высокая уверенность
хорошо	Знание основ, понимание ключевых понятий
удовлетворительно	Понимание базовых элементов, есть пробелы

Оценка	Критерии оценивания
неудовлетворительно	Поверхностные знания, непонимание связи
плохо	Отсутствие понимания дисциплины

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

### Шкала оценивания сформированности компетенций

Уровень сформированности компетенций (индикатора достижения 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. Ошибок нет.	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	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. Выполнены все задания в полном объеме, но некоторые с недочетами.	Продemonстрированы все основные умения. Решены все основные задачи с отдельным и несущественными недочетами, выполнены все задания в полном объеме	Продemonстрированы все основные умения. Решены все основные задачи. Выполнены все задания, в полном объеме без недочетов
<u>Навыки</u>	Отсутствие базовых навыков. Невозможность оценить наличие навыков вследствие отказа обучающегося от	При решении стандартных задач не продемонстрированы базовые навыки. Имели место грубые ошибки	Имеется минимальный набор навыков для решения стандартных задач с некоторым	Продemonстрированы базовые навыки при решении стандартных задач с некоторым и	Продemonстрированы базовые навыки при решении стандартных задач без ошибок и недочетов	Продemonстрированы навыки при решении нестандартных задач без ошибок и	Продemonстрирован творческий подход к решению нестандартных задач

	ответа		и недочетами	недочетами		недочетов	
--	--------	--	-----------------	------------	--	-----------	--

### Шкала оценивания при промежуточной аттестации

Оценка		Уровень подготовки
зачтено	превосходно	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «превосходно», продемонстрированы знания, умения, владения по соответствующим компетенциям на уровне выше предусмотренного программой
	отлично	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «отлично».
	очень хорошо	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «очень хорошо»
	хорошо	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «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».
	плохо	Хотя бы одна компетенция сформирована на уровне «плохо»

### 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 - Тест) для оценки сформированности компетенции ПК-9

Экзаменационный тест

##### ВАРИАНТ 1

1. Что такое анализ данных?

- а) Ручная сортировка таблиц
- б) Изучение, интерпретация и представление данных
- в) Архивирование данных
- г) Только визуализация

2. Что такое pandas в Python?



a) Игровая библиотека

b) Библиотека машинного зрения

c) Библиотека для анализа и обработки таблиц

d) Система базы данных

3. Какой метод машинного обучения относится к надзорному обучению?

a) К-средних

b) Линейная регрессия

c) PCA

d) DBSCAN

4. Что делает функция `.groupby()` в pandas?

a) Удаляет дубликаты

b) Объединяет строки

c) Группирует данные по признаку

d) Переводит данные в числовой формат

5. Какой слой нейронной сети отвечает за итоговое решение?

a) Входной

b) Скрытый

c) Выходной

d) Активационный

6. Что такое сверточная нейронная сеть (CNN)?

a) Сеть для обработки текстов

b) Сеть для работы с изображениями

c) Алгоритм поиска

d) Система логики

7. Какой формат используется для хранения датасета в pandas?

a) string

b) list

c) DataFrame

d) Series

8. Что делает функция `train_test_split()` из Scikit-learn?

a) Обучает модель

b) Применяет модель

c) Разделяет данные на обучающую и тестовую выборки

d) Удаляет пропуски

9. Что такое переобучение?

a) Недостаточное обучение модели

b) Слишком точное запоминание обучающих данных

c) Удаление лишних признаков

d) Ошибка компиляции

10. Какой язык программирования чаще всего используется в анализе данных?

a) C++

b) Python

c) Java

d) HTML

## ВАРИАНТ 2

1. Что такое нейронная сеть?

a) Алгоритм сортировки

b) База данных

c) Модель, имитирующая работу мозга

d) Система хранения

2. Что такое регрессия в машинном обучении?

a) Разбиение данных по классам

b) Предсказание числового значения

c) Построение графика

d) Сокращение признаков

3. Какая библиотека используется для визуализации данных?

a) NumPy

b) Matplotlib

c) Pytorch

d) JSON

4. Что такое точность модели (accuracy)?

a) Отношение ложных срабатываний

b) Среднеквадратичная ошибка

c) Доля правильных предсказаний

d) Время выполнения

5. Что делает функция `dropna()` в pandas?

a) Объединяет таблицы

b) Удаляет строки с пропусками

c) Преобразует тип данных

d) Заполняет пустые значения

6. Какой тип сети обрабатывает последовательности (временные ряды)?

a) RNN (рекуррентная)

b) CNN

c) GAN

d) LSTM (неправильно как самостоятельный тип)

7. Что означает one-hot encoding?

a) Очистка от выбросов

b) Удаление категориальных признаков

c) Представление категорий в виде бинарных векторов

d) Метод регрессии

8. Какой метод используется для снижения размерности?

a) KNN

b) SVM

c) PCA (главные компоненты)

d) CNN

9. Что такое датасет?

a) Набор данных для анализа и обучения

b) Библиотека для Python

c) Интерфейс пользователя

d) Массив случайных чисел

10. Что происходит в процессе обучения нейросети?

a) Запоминаются выходные значения

b) Обновляются веса с целью уменьшения ошибки

c) Создаются новые данные

d) Записываются данные в память

### Критерии оценивания (оценочное средство - Тест)

Оценка	Критерии оценивания
превосходно	Полное понимание материала (10)
отлично	Почти безошибочное выполнение (9)
очень хорошо	Небольшие неточности, высокая уверенность (8)
хорошо	Знание основ, понимание ключевых понятий (7)
удовлетворительно	Понимание базовых элементов, есть пробелы (5-6)
неудовлетворительно	Поверхностные знания, непонимание связи (3-4)

Оценка	Критерии оценивания
плохо	Отсутствие понимания дисциплины (0-2)

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

Основная литература:

1. Алексейчук А. С. Введение в нейронные сети: модели, методы и программные средства : учебное пособие / Алексейчук А. С. - Москва : МАИ, 2023. - 105 с. - Редсовет МАИ. - Книга из коллекции МАИ - Информатика. - ISBN 978-5-6049766-0-9., <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=893017&idb=0>.
2. Алетдинова А. А. Интеллектуальный анализ больших данных : учебное пособие / Алетдинова А. А., Муртазина М. Ш. - Новосибирск : НГТУ, 2023. - 66 с. - Утверждено Редакционно-издательским советом университета в качестве учебного пособия. - Книга из коллекции НГТУ - Информатика. - ISBN 978-5-7782-4899-1., <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=896742&idb=0>.
3. Ростовцев В. С. Искусственные нейронные сети : учебник для вузов / Ростовцев В. С. - 5-е изд., стер. - Санкт-Петербург : Лань, 2025. - 216 с. - Книга из коллекции Лань - Информатика. - ISBN 978-5-507-50568-5., <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=930323&idb=0>.

Дополнительная литература:

1. Ульман Дж.Д. Анализ больших наборов данных : монография / Ульман Дж.Д.; Лесковец Ю.; Раджараман А. - Москва : ДМК-пресс, 2023. - 500 с. - ISBN 978-5-89818-304-2., <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=878866&idb=0>.
2. Терлецкий А. С. Нейронные сети и искусственный интеллект: Основы нейронных сетей на языке Python / Терлецкий А. С., Терлецкая Е. С. - Липецк : Липецкий ГПУ, 2023. - 76 с. - Книга из коллекции Липецкий ГПУ - Информатика. - ISBN 978-5-907792-40-1., <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=928777&idb=0>.
3. Анализ и обработка больших данных на платформе QuAnt : учебное пособие / Ермакова С. Г., Бухарова К. А., Мандрика О. С., Лабузов С. В. - Санкт-Петербург : ПГУПС, 2024. - 115 с. - Книга из коллекции ПГУПС - Информатика. - ISBN 978-5-7641-2002-7., <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=928834&idb=0>.
4. Интеллектуальный предиктивный мультимодальный анализ слабоструктурированных больших данных / Ярушкина Н. Г., Андреев И. А., Гуськов Г. Ю., Дударин П. В., Желепов А. С., Мошкин В. С., Наместников А. М., Романов А. А., Филиппов А. А., Эгов Е. Н. - Ульяновск : УлГТУ, 2020. - 220 с. - Библиогр.: доступна в карточке книги, на сайте ЭБС Лань. - Книга из коллекции УлГТУ - Экономика и менеджмент. - ISBN 978-5-9795-2088-9., <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=754203&idb=0>.

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

Программное обеспечение:

Python (3.x) - Язык программирования для анализа данных и ИИ. Кроссплатформенный,

открытый.

Anaconda - Дистрибутив Python с предустановленными библиотеками. Включает Jupyter, pandas и др.

Jupyter Notebook / Lab - Среда для интерактивной работы с кодом и данными. Web-интерфейс, встроен в Anaconda.

PyCharm (Community) - IDE для Python. Удобен для работы с ML и AI.

Visual Studio Code - Легкий редактор с поддержкой Python и расширений AI.

Кроссплатформенный.

Excel / Google Таблицы - Предобработка и базовая визуализация данных. Начальный уровень анализа.

Интернет-ресурсы и платформы:

Kaggle - Соревнования, наборы данных, Jupyter-блокноты.

Stepik - Онлайн-курсы по Python, ML, DL и обработке данных на русском.

Coursera - Международные курсы по анализу данных, машинному обучению, Python.

Google Colab - Онлайн-среда для написания и запуска кода, с доступом к GPU.

Hugging Face - Библиотеки и модели для нейросетей, особенно NLP.

GitHub - Хостинг проектов с открытым кодом, поиск библиотек и датасетов.

OpenDataScience (ods.ai) - Сообщество практиков и студентов по DS и ML.

AI Google Learning - Учебные материалы по ML, AI, TensorFlow от Google

## **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 09.03.03 - Прикладная информатика.

Автор(ы): Смелова Татьяна Сергеевна.

Рецензент(ы): Голубева Екатерина Александровна, кандидат физико-математических наук.

Заведующий кафедрой: Ягунова Наталья Александровна, кандидат экономических наук.

Программа одобрена на заседании методической комиссии от 27.11.2023, протокол № 5.