

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

УТВЕРЖДАЮ

Заведующий кафедрой  
ПиИТ



проф. Махортов С.Д.,  
подпись, расшифровка подписи  
05.03.2024 г.

**РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ**  
**Б1.В.ДВ.01.07.05 Интеллектуальный анализ производственных данных**

*Код и наименование дисциплины в соответствии с учебным планом*

**1. Код и наименование направления подготовки/специальности:**

09.03.02 Информационные системы и технологии

**2. Профиль подготовки/специализация:** Инженерия информационных систем и технологий

**3. Квалификация выпускника:** Бакалавр

**4. Форма обучения:** очная

**5. Кафедра, отвечающая за реализацию дисциплины:** Информационных систем

**6. Составители программы:** Чижов М.И., доктор технических наук, профессор  
(ФИО, ученая степень, ученое звание)

**7. Рекомендована:** НМС факультета компьютерных наук протокол № 5 от 05.03.2024 г.

*(наименование рекомендующей структуры, дата, номер протокола,*

*отметки о продлении вносятся вручную)*

**8. Учебный год:** 2027/2028

**Семестр(ы):** 8

## 9. Цели и задачи учебной дисциплины

### Цели курса:

- Формирование компетенций в области анализа данных с акцентом на применение в производственной среде.
- Развитие навыков работы с инструментами анализа и методами, необходимыми для анализа и визуализации данных.
- Применение теоретических знаний на практике теоретических знаний к реальным производственным данным для принятия обоснованных решений.

### Задачи курса:

- Изучить основные методы первичного анализа данных и их применение в производстве.
- Освоить техники определения и обработки выбросов и аномалий в данных.
- Научиться использовать библиотеку pandas для обработки и анализа данных, включая работу с пропущенными значениями.
- Изучить визуализацию данных с использованием библиотеки Matplotlib для представления результатов анализа.
- Развить навыки построения и оптимизации моделей анализа данных, а также интерпретации их результатов.
- Рассмотреть метрики оценки и тестирования моделей для обеспечения их надежности и точности.

**10. Место учебной дисциплины в структуре ООП:** Дисциплина относится к блоку Дисциплины (модули) по выбору 5 (ДВ.5).

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

Код и название компетенции	Код и название индикатора компетенции	Знания, умения, навыки
ПК-7 Способен выполнять работы по созданию (кастомизации), использованию и сопровождению систем автоматизированного проектирования и цифровой подготовки производства	ПК-7.1 Знает среды, языки и методы программирования, инструменты и методики тестирования разрабатываемых программных средств	Знать основные принципы анализа производственных данных и их значение для оптимизации производственных процессов, методы и техники первичного анализа данных, включая статистические методы и визуализацию. Применять математические методы, такие как статистические и алгоритмические модели, для решения задач в программных средствах.
ПК-7 Способен выполнять работы по созданию (кастомизации), использованию и сопровождению систем автоматизированного проектирования и цифровой подготовки производства	ПК-7.2 Знает устройство и функционирование современных систем поддержки ЖЦ, протоколы, интерфейсы и форматы обмена данными	Уметь применять методологии разработки программного обеспечения для создания систем анализа данных, использовать языки программирования и инструменты для обработки и анализа производственных данных Знать основные технологии и методы обработки данных, включая структурированные и неструктурированные данные, принципы параллельной обработки данных и её применение для повышения производительности в программных системах
ПК-7 Способен выполнять работы по созданию (кастомизации), использованию и	ПК-7.6 Инсталлирует и настраивает операционные системы, СУБД, прикладное ПО,	Уметь анализировать потребности заказчика и формулировать четкие требования к информационным системам с учетом производственных процессов,

сопровождению систем автоматизированного проектирования и цифровой подготовки производства	необходимое для функционирования программных средств поддержки ЖЦ	оценивать возможности типовых информационных систем для удовлетворения требований заказчика и адаптировать решения под конкретные нужды. Разрабатывать и адаптировать алгоритмы для обработки данных и анализа производственных процессов, обеспечивая их эффективное функционирование
--	---	---

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

**Форма промежуточной аттестации зачет с оценкой.**

### 13. Трудоемкость по видам учебной работы

Вид учебной работы		Трудоемкость	
		Всего	По семестрам
Аудиторные занятия		48	48
в том числе:	лекции	24	24
	практические		
	лабораторные	24	24
Самостоятельная работа		60	60
Курсовая работа			
Промежуточная аттестация		8	8
Часы на контроль		8	8
Всего		108	108

#### 13.1. Содержание дисциплины

п/п	Наименование раздела дисциплины	Содержание раздела дисциплины	Реализация раздела дисциплины с помощью онлайн-курса, ЭУМК *
<b>1. Лекции</b>			
1.1	Введение в анализ данных	Изучение основ анализа данных, его важности и применения в производстве. Этапы анализа данных, включая сбор, обработку и визуализацию.	
1.2	Обработка и очистка данных	Методы обработки данных, включая обработку пропущенных значений, выбросов и аномалий.	
1.3	Корреляционный и регрессионный анализ	Методы корреляционного и регрессионного анализа для выявления взаимосвязей между переменными	
1.4	Моделирование и оптимизация	Принципы построения моделей и методов их оптимизации для решения задач анализа данных	
1.5	Визуализация и интерпретация результатов	Методы визуализации и интерпретации результатов анализа данных для поддержки принятия решений	
<b>2. Практические занятия</b>			
2.1	Введение в анализ данных	Анализ данных с использованием методов первичного	

		анализа (среднее, медиана, мода) на примере производственных показателей. Визуализация данных с использованием Matplotlib: создание графиков и диаграмм для анализа производственных процессов.	
2.2	Обработка и очистка данных	Очистка данных с использованием библиотеки pandas: выявление и обработка пропусков и выбросов на реальных данных. Создание отчетов о качестве данных: анализ и визуализация результатов обработки.	
2.3	Корреляционный и регрессионный анализ	Выполнение корреляционного анализа: расчет коэффициента корреляции и интерпретация результатов. Построение регрессионной модели: анализ влияния факторов на производственные результаты	
2.4	Моделирование и оптимизация	Построение модели на основе исторических данных: использование методов машинного обучения для прогнозирования производственных показателей. Оптимизация параметров модели: применение методов кросс-валидации и подбора гиперпараметров.	
2.5	Визуализация и интерпретация результатов	Создание интерактивной визуализации данных с использованием библиотек визуализации: демонстрация результатов анализа. Подготовка и представление итогового отчета по результатам анализа данных, включая выводы и рекомендации для производства.	
<b>3. Лабораторные занятия</b>			
3.1	Введение в анализ данных	Изучение набора данных о производственных процессах. Загрузка и первичный анализ данных с использованием pandas. Визуализация данных с использованием Matplotlib: создание различных графиков для анализа производственных показателей.	
3.2	Обработка и очистка данных	Очистка реального набора данных о производственных показателях: выявление пропусков и аномалий, применение методов обработки. Анализ и визуализация качества данных: создание отчетов о пропусках и выбросах на основе реальных данных.	
3.3	Корреляционный и регрессионный анализ	Анализ взаимосвязей между переменными: применение корреляционного анализа на реальных производственных данных. Построение регрессионной модели на реальных данных: оценка влияния определенных факторов на производственные результаты.	
3.4	Моделирование и оптимизация	Построение модели прогнозирования производственных показателей на основе исторических данных: использование методов машинного обучения. Оптимизация параметров модели с использованием реальных данных: применение методов кросс-валидации и подбора гиперпараметров.	
3.5	Визуализация и интерпретация результатов	Создание интерактивной визуализации на основе результатов анализа реальных данных, использование библиотек для отображения ключевых показателей. Подготовка и представление итогового отчета по результатам анализа на реальных данных, включая выводы и рекомендации для производственных процессов.	

### 13.2. Темы (разделы) дисциплины и виды занятий

№	Наименование темы	Виды занятий (количество часов)
---	-------------------	---------------------------------

п/п	(раздела) дисциплины	Лекции	Практические	Лабораторные	Самостоятельная работа	Всего
1.	Введение в анализ данных	4		4	12	20
2.	Обработка и очистка данных	5		5	12	22
3.	Корреляционный и регрессионный анализ	5		5	12	22
4.	Моделирование и оптимизация	5		5	12	22
5.	Визуализация и интерпретация результатов	5		5	12	22
	Итого:	32	16	16	44	108

#### 14. Методические указания для обучающихся по освоению дисциплины

При использовании дистанционных образовательных технологий и электронного обучения выполнять все указания преподавателей по работе на LMS-платформе, своевременно подключаться к online-занятиям, соблюдать рекомендации по организации самостоятельной работы.

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

а) основная литература:

№ п/п	Источник
1.	Татарникова, Т. М. Интеллектуальный анализ данных : учебное пособие / Т. М. Татарникова. — Вологда : Инфра-Инженерия, 2024. — 172 с. — ISBN 978-5-9729-1772-3. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <a href="https://e.lanbook.com/book/427826">https://e.lanbook.com/book/427826</a> (дата обращения: 22.09.2024). — Режим доступа: для авториз. пользователей.

б) дополнительная литература:

№ п/п	Источник
2.	Ланских, Ю. В. Интеллектуальный анализ данных : учебное пособие / Ю. В. Ланских, В. Г. Ланских. — Киров : ВятГУ, 2023. — 240 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <a href="https://e.lanbook.com/book/408569">https://e.lanbook.com/book/408569</a> (дата обращения: 22.09.2024). — Режим доступа: для авториз. пользователей.

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

№ п/п	Ресурс
3.	<a href="https://edu.vsu.ru/course/view.php?id=2899">https://edu.vsu.ru/course/view.php?id=2899</a> Автоматизация конструкторско-технологической подготовки

#### 16. Перечень учебно-методического обеспечения для самостоятельной работы

№ п/п	Источник
1.	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: <a href="https://e.lanbook.com/book/381533">https://e.lanbook.com/book/381533</a> (дата обращения: 22.09.2024). — Режим доступа: для авториз. пользователей.

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

Дисциплина реализуется с применением электронного обучения и дистанционных образовательных технологий. Для организации занятий рекомендован онлайн-курс «Интеллектуальный анализ производственных данных (<https://edu.vsu.ru/user/index.php?id=27384>)», размещенный на платформе Электронного университета ВГУ (LMS moodle), а также Интернет-ресурсы, приведенные в п.15в.

**18. Материально-техническое обеспечение дисциплины:** Лекционная аудитория, оборудованная мультимедийным проектором. Компьютерные классы

факультета для проведения лабораторных занятий. Образовательный портал «Электронный университет ВГУ» <https://edu.vsu.ru>.

### **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 в анализ данных	ПК-7	ПК-7.1	<i>Тестовое задание</i>
2.	Обработка и очистка данных	ПК-7	ПК-7.2	<i>Лабораторные работы</i>
3.	Корреляционный и регрессионный анализ	ПК-7	ПК-7.6	<i>Практическое задание</i>
4.	Моделирование и оптимизация	ПК-7	ПК-7.2	<i>Тестовое задание</i>
5.	Визуализация и интерпретация результатов	ПК-7	ПК-7.6	<i>Лабораторные работы</i>
Промежуточная аттестация форма контроля – <i>зачет с оценкой</i>				<i>Перечень вопросов Практическое задание</i>

### **20 Типовые оценочные средства и методические материалы, определяющие процедуры оценивания**

#### **20.1 Текущий контроль успеваемости**

Контроль успеваемости по дисциплине осуществляется с помощью следующих оценочных средств:

- Тестовые задания
- Лабораторные работы

#### **Примеры вопросов к тесту:**

1 Какое из следующих утверждений лучше всего описывает анализ данных?

- A) Процесс сбора данных для отчетности
- B) Процесс преобразования данных в полезную информацию +
- C) Процесс хранения данных в базе данных
- D) Процесс создания новых данных

2 Какой из следующих этапов не является частью процесса анализа данных?

- A) Сбор данных
- B) Обработка данных
- C) Уничтожение данных +

D) Визуализация данных

3 Какой инструмент обычно используется для визуализации данных?

- A) SQL
- B) Excel
- C) Matplotlib +
- D) Python

4 Что такое среднее арифметическое?

- A) Разница между максимальным и минимальным значением
- B) Сумма всех значений, деленная на их количество +
- C) Наиболее часто встречающееся значение
- D) Значение, разделяющее данные на две равные части

5 Какой из следующих методов используется для анализа распределения данных?

- A) Корреляционный анализ
- B) Гистограмма +
- C) Регрессионный анализ
- D) Классификация

6 Какой тип графика лучше всего подходит для отображения трендов во времени?

- A) Столбчатая диаграмма
- B) Линейный график +
- C) Круговая диаграмма
- D) Диаграмма рассеяния

7 Какой из следующих факторов не влияет на качество данных?

- A) Неполнота данных
- B) Актуальность данных
- C) Время обработки данных +
- D) Ошибки в данных

#### **Описание технологии проведения**

Тестирование проводится в электронной форме и оценивается по 5-балльной шкале. Тесты содержат по 15 вопросов. При трех и более ошибочных ответах выставляется оценка «не зачетно»

При защите лабораторных работ оценивается работоспособность модели, адекватность результатов и общий подход к решению задачи. Задаются дополнительные вопросы уточняющего и теоретического характера. Каждая лабораторная работа оценивается по 5-балльной шкале.

## **20.2 Промежуточная аттестация**

### **Перечень вопросов к собеседованию**

1. Что такое анализ данных и какую роль он играет в производственных системах?

2. Опишите основные этапы процесса анализа данных.
3. Какие типы данных существуют, и как они различаются?
4. Что такое первичный анализ данных, и какие методы вы используете для его проведения?
5. Каковы основные причины возникновения пропусков в данных, и какие методы их обработки вы знаете?
6. Что такое выбросы, и как вы их можете выявить в наборе данных?
7. Опишите метод корреляционного анализа и его значение в исследовании взаимосвязей между переменными.
8. Что такое регрессионный анализ, и как он может помочь в прогнозировании производственных показателей?
9. Какие основные методы машинного обучения вы знаете, и как они применяются в анализе данных?
10. Что такое кросс-валидация и почему она важна при построении моделей?
11. Опишите, как вы бы визуализировали данные для представления их результатов заинтересованным сторонам.
12. Как вы определяете качество данных и что включает в себя этот процесс?
13. Какие инструменты и технологии вы используете для анализа и обработки данных?
14. Какова роль статистики в анализе данных, и какие статистические меры вы считаете наиболее важными?
15. Что такое «big data», и как технологии обработки больших данных влияют на анализ производственных систем?

#### Примеры вопросов к тесту:

1. Какой из следующих этапов анализа данных включает в себя сбор информации?
  - A) Обработка данных
  - B) Визуализация данных
  - C) Сбор данных +
  - D) Интерпретация данных
2. Какой из следующих инструментов чаще всего используется для визуализации данных?
  - A) R
  - B) Excel
  - C) Matplotlib +
  - D) SQL
3. Какое из следующих утверждений является верным?
  - A) Анализ данных не требуется для принятия решений
  - B) Анализ данных помогает выявлять паттерны и тренды +
  - C) Данные не нуждаются в обработке
  - D) Все данные одинаково полезны
4. Что такое выбросы в данных?
  - A) Пропущенные значения
  - B) Значения, значительно отличающиеся от остальных +
  - C) Средние значения
  - D) Неправильные форматы данных
5. Какой из методов обработки данных используется для заполнения пропущенных значений?
  - A) Регрессия +
  - B) Дисперсия
  - C) Корреляция
  - D) Стандартное отклонение
6. Какой из следующих инструментов используется для очистки данных?
  - A) Matplotlib
  - B) pandas +
  - C) TensorFlow



- D) SQL
7. Какой коэффициент используется для измерения силы связи между переменными?
    - A) Коэффициент корреляции +
    - B) Стандартное отклонение
    - C) Среднее арифметическое
    - D) Дисперсия
  8. Что представляет собой линейная регрессионная модель?
    - A) Модель, предсказывающая будущие события
    - B) Модель, которая описывает зависимость одной переменной от другой +
    - C) Модель, использующая случайные величины
    - D) Модель, которая не требует данных
  9. Какой метод используется для оценки качества модели?
    - A) Кросс-валидация +
    - B) Статистический анализ
    - C) Обработка данных
    - D) Группировка данных
  10. Какой из следующих методов машинного обучения лучше всего подходит для задач классификации?
    - A) Линейная регрессия
    - B) Метод опорных векторов +
    - C) K-средних
    - D) Линейная интерполяция
  11. Какой тип графика лучше всего подходит для сравнения категорий?
    - A) Линейный график
    - B) Круговая диаграмма
    - C) Столбчатая диаграмма +
    - D) Гистограмма
  12. Что важно учитывать при интерпретации визуализации данных?
    - A) Точность данных +
    - B) Цвета графиков
    - C) Формат файла
    - D) Размер шрифта
  13. Какой из следующих этапов не является частью процесса анализа данных?
    - A) Сбор данных
    - B) Очистка данных
    - C) Уничтожение данных +
    - D) Визуализация данных
  14. Какой из следующих методов анализа данных помогает выявить паттерны в больших объемах информации?
    - A) Статистический анализ
    - B) Машинное обучение +
    - C) Ручной анализ
    - D) Интуитивный анализ
  15. Какой из следующих факторов может повлиять на качество анализа данных?
    - A) Неполнота данных
    - B) Неправильная обработка данных
    - C) Ошибки в данных
    - D) Все вышеперечисленные +

#### **Описание технологии проведения**

Собеседование производится в форме устного ответа на заданный вопрос. При необходимости преподаватель может задавать уточняющие вопросы. Ответ оценивается по 5 бальной шкале.

Тестирование проводится в электронной форме и оценивается по 5-балльной шкале. Тесты содержат по 15 вопросов. При трех и более ошибочных ответах выставляется оценка «не зачетно»

При защите лабораторных работ оценивается работоспособность модели, адекватность результатов и общих подход к решению задачи. Задаются дополнительные вопросы уточняющего и теоретического характера. Каждая лабораторная работа оценивается по 5-балльной шкале.

#### **Требования к выполнению заданий, шкалы и критерии оценивания**

Оценка знаний, умений и навыков, характеризующая этапы формирования компетенций в рамках изучения дисциплины, осуществляется в ходе текущей и промежуточной аттестаций.

При оценивании результатов промежуточной аттестации используется количественная шкала оценок. Оценка за защиту лабораторных работ (2 балла за каждую защищенную работу) и тест (максимум 5 баллов) складываются с оценкой, полученной на собеседовании, и результат суммируется. Полученное значение определяет уровень сформированности компетенций и итоговую оценку согласно следующей шкале:

- оценка «отлично» - 19..20 баллов
- оценка «хорошо» - 16..18 баллов
- оценка «удовлетворительно» - 11..15 баллов
- оценка «неудовлетворительно» - 0..10 баллов

#### **Описание критериев и шкалы оценивания результатов обучения при промежуточной аттестации.**

Текущая аттестация проводится в соответствии с Положением о балльно-рейтинговой системе факультета компьютерных наук Воронежского государственного университета.

Текущая аттестация проводится в форме тестов по теоретической части курса, выполняемых в электронном виде в портале «Электронный университет ВГУ», и в форме решения практических задач, выполняемые в компьютерном классе (в лаборатории) факультета компьютерных наук. Критерии оценивания приведены выше.

Промежуточная аттестация проводится в соответствии с Положением о промежуточной аттестации обучающихся по программам высшего образования и Положением о балльно-рейтинговой системе факультета компьютерных наук.

При оценивании используются количественные шкалы оценок.