

**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ**
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Воронежский государственный технический университет»

УТВЕРЖДАЮ

Декан факультета инженерных систем и
сооружений


/С.А. Яременко/
«21» иссента 2025г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

«Аналитика данных»

Направление подготовки 20.03.01 Техносферная безопасность

Профиль Пожарная безопасность в строительстве

Квалификация выпускника бакалавр

Нормативный период обучения 4 года / 4 года и 11 м.

Форма обучения очная / заочная

Год начала подготовки 2023

Автор программы _____ Н.А. Рындин

Заведующий кафедрой
Искусственного интеллекта
и цифровых технологий _____ П.Ю. Гусев

Руководитель ОПОП _____ Е.А. Сушко

Воронеж 2025

1. ЦЕЛИ И ЗАДАЧИ ДИСЦИПЛИНЫ

1.1. Цели дисциплины

Целью дисциплины является формирование у студентов системного представления об аналитике данных как области знаний и практической деятельности, овладение базовыми методами анализа данных, понимание этапов аналитического процесса, а также развитие навыков применения инструментов аналитики для решения задач в сфере искусственного интеллекта и цифровизации.

1.2. Задачи освоения дисциплины

- ознакомить студентов с основами аналитики данных, включая этапы подготовки, обработки и интерпретации данных;
- научить выявлять закономерности в данных с использованием статистических методов и алгоритмов машинного обучения;
- сформировать навыки работы с инструментами и библиотеками анализа данных;
- развить способность формулировать аналитические задачи на основе проблем профессиональной деятельности;
- обеспечить практический опыт построения аналитических моделей для естественных и искусственных систем.

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОПОП

Дисциплина «Аналитика данных» относится к дисциплинам части, формируемой участниками образовательных отношений (дисциплина по выбору) блока Б1.

3. ПЕРЕЧЕНЬ ПЛАНИРУЕМЫХ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ

Процесс изучения дисциплины «Аналитика данных» направлен на формирование следующих компетенций:

УК-1 - Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач

Компетенция	Результаты обучения, характеризующие сформированность компетенции
УК-1	знать основные этапы аналитического процесса и их взаимосвязь; основные методы и подходы анализа данных; принципы построения и валидации моделей машинного обучения; методы визуализации и интерпретации результатов анализа
	уметь формулировать аналитические задачи на основе реальных проблем; применять статистические и машинные методы анализа данных; использовать инструменты Python для обработки и анализа данных; интерпретировать

	результаты анализа в контексте прикладной задачи
	владеть навыками работы с библиотеками Python (Pandas, NumPy, Scikit-learn, Matplotlib); техниками предобработки и очистки данных; приемами построения, настройки и оценки моделей машинного обучения; методами визуализации данных и представления результатов

4. ОБЪЕМ ДИСЦИПЛИНЫ

Общая трудоемкость дисциплины «Аналитика данных» составляет 3 з.е.

Распределение трудоемкости дисциплины по видам занятий
очная форма обучения

Виды учебной работы	Всего часов	Семестры
		5
Аудиторные занятия (всего)	54	54
В том числе:		
Лекции	18	18
Лабораторные работы (ЛР)	36	36
Самостоятельная работа	54	54
Виды промежуточной аттестации - зачет	+	+
Общая трудоемкость:		
академические часы	108	108
зач.ед.	3	3

заочная форма обучения

Виды учебной работы	Всего часов	Семестры
		5
Аудиторные занятия (всего)	8	8
В том числе:		
Лекции	4	4
Лабораторные работы (ЛР)	4	4
Самостоятельная работа	96	96
Часы на контроль	4	4
Виды промежуточной аттестации - зачет	+	+
Общая трудоемкость:		
академические часы	108	108
зач.ед.	3	3

5. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

5.1 Содержание разделов дисциплины и распределение трудоемкости по видам занятий

очная форма обучения

№ п/п	Наименование темы	Содержание раздела	Лекц	Лаб. зан.	СРС	Всего, час
1	Введение в аналитику данных	Понятие и роль аналитики данных. Виды аналитики: описательная, диагностическая,	2	4	6	12

		предсказательная, предписывающая.				
2	Аналитический процесс и источники данных	Этапы аналитического цикла. Сбор данных, типы источников, открытые и закрытые данные.	2	4	6	12
3	Предобработка и очистка данных	Работа с пропущенными значениями, выбросами, типами данных. Принципы качественной подготовки данных.	2	4	6	12
4	Разведочный анализ данных (EDA)	Методы первичного анализа. Визуализация и обобщение данных.	2	4	6	12
5	Базовые статистические методы в аналитике данных	Средние значения, дисперсия, корреляция, статистические тесты.	2	4	6	12
6	Машинное обучение для аналитики данных	Основные понятия, типы задач, обучение с учителем и без учителя.	2	4	6	12
7	Классификация и регрессия	Логистическая регрессия, деревья решений, kNN, метрики качества.	2	4	6	12
8	Кластеризация и понижение размерности	K-Means, DBSCAN, PCA, t-SNE и их применение в практике.	2	4	6	12
9	Оценка и валидация моделей	Кросс-валидация, переобучение, недообучение, выбор модели.	2	4	6	12
Итого			18	36	54	108

заочная форма обучения

№ п/п	Наименование темы	Содержание раздела	Лекц	Лаб. зан.	СРС	Всего, час
1	Введение в аналитику данных	Понятие и роль аналитики данных. Виды аналитики: описательная, диагностическая, предсказательная, предписывающая.	2	2	12	16
2	Аналитический процесс и источники данных	Этапы аналитического цикла. Сбор данных, типы источников, открытые и закрытые данные.	2	2	12	16
3	Предобработка и очистка данных	Работа с пропущенными значениями, выбросами, типами данных. Принципы качественной подготовки данных.	-	-	12	12
4	Разведочный анализ данных (EDA)	Методы первичного анализа. Визуализация и обобщение данных.	-	-	16	16
5	Базовые статистические методы в аналитике данных	Средние значения, дисперсия, корреляция, статистические тесты.	-	-	16	16
6	Машинное обучение для аналитики данных	Основные понятия, типы задач, обучение с учителем и без учителя.	-	-	16	16
7	Классификация и регрессия	Логистическая регрессия, деревья решений, kNN, метрики качества.	-	-	4	4
8	Кластеризация и понижение размерности	K-Means, DBSCAN, PCA, t-SNE и их применение в практике.	-	-	4	4
9	Оценка и валидация моделей	Кросс-валидация, переобучение, недообучение, выбор модели.	-	-	4	4
Итого			4	4	96	104

5.2 Перечень лабораторных работ

1. Обзор инструментов анализа данных в Python. Установка, настройка, обзор Jupyter Notebook, библиотеки для аналитики.
2. Загрузка и первичная обработка данных. Работа с CSV, Excel, API. Очистка и преобразование данных.
3. Исследовательский анализ данных. Построение графиков, изучение

распределений, корреляции.

4. Основы статистики на практике. Реализация статистических тестов, вывод статистических закономерностей.

5. Построение моделей классификации. Реализация логистической регрессии и дерева решений на реальных данных.

6. Построение моделей регрессии. Линейная регрессия, оценка качества модели, визуализация результатов.

7. Методы кластеризации и понижения размерности. Применение K-Means, PCA и t-SNE на многомерных данных.

8. Оценка качества моделей и кросс-валидация. Метрики, confusion matrix, ROC-кривые.

9. Проектная работа: аналитика в реальном кейсе

Применение полного цикла аналитики от данных до модели.

6. ПРИМЕРНАЯ ТЕМАТИКА КУРСОВЫХ ПРОЕКТОВ (РАБОТ) И КОНТРОЛЬНЫХ РАБОТ

В соответствии с учебным планом освоение дисциплины не предусматривает выполнение курсового проекта (работы) или контрольной работы.

7. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ДЛЯ ПРОВЕДЕНИЯ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ

7.1. Описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания

7.1.1 Этап текущего контроля

Результаты текущего контроля знаний и межсессионной аттестации оцениваются по следующей системе:

«аттестован»;

«не аттестован».

Компетенция	Результаты обучения, характеризующие сформированность компетенции	Критерии оценивания	Аттестован	Не аттестован
УК-1	знать основные этапы аналитического процесса и их взаимосвязь; основные методы и подходы анализа данных; принципы построения и валидации моделей машинного обучения; методы визуализации и интерпретации результатов анализа	активная работа на практических занятиях	Выполнение работ в срок, предусмотренный в рабочих программах	Невыполнение работ в срок, предусмотренный в рабочих программах
	уметь	решение стандартных	Выполнение работ в	Невыполнение

	формулировать аналитические задачи на основе реальных проблем; применять статистические и машинные методы анализа данных; использовать инструменты Python для обработки и анализа данных; интерпретировать результаты анализа в контексте прикладной задачи	практических задач	срок, предусмотренный в рабочих программах	работ в срок, предусмотренный в рабочих программах
	владеть навыками работы с библиотеками Python (Pandas, NumPy, Scikit-learn, Matplotlib); техниками предобработки и очистки данных; приемами построения, настройки и оценки моделей машинного обучения; методами визуализации данных и представления результатов	решение прикладных задач в конкретной предметной области	Выполнение работ в срок, предусмотренный в рабочих программах	Невыполнение работ в срок, предусмотренный в рабочих программах

7.1.2 Этап промежуточного контроля знаний

Результаты промежуточного контроля знаний оцениваются в 5 семестре для очной формы обучения, 5 семестре для заочной формы обучения по двухбалльной системе:

«зачтено»

«не зачтено»

Компетенция	Результаты обучения, характеризующие сформированность компетенции	Критерии оценивания	Зачтено	Не зачтено
УК-1	знать основные этапы аналитического процесса и их взаимосвязь; основные методы и подходы анализа данных; принципы построения и валидации моделей машинного	Тест	Выполнение теста на 70-100%	Выполнение менее 70%

	обучения; методы визуализации и интерпретации результатов анализа			
	уметь формулировать аналитические задачи на основе реальных проблем; применять статистические и машинные методы анализа данных; использовать инструменты Python для обработки и анализа данных; интерпретировать результаты анализа в контексте прикладной задачи	Решение стандартных практических задач	Продемонстрирован верный ход решения в большинстве задач	Задачи не решены
	владеть навыками работы с библиотеками Python (Pandas, NumPy, Scikit-learn, Matplotlib); техниками предобработки и очистки данных; приемами построения, настройки и оценки моделей машинного обучения; методами визуализации данных и представления результатов	Решение прикладных задач в конкретной предметной области	Продемонстрирован верный ход решения в большинстве задач	Задачи не решены

7.2 Примерный перечень оценочных средств (типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности)

7.2.1 Примерный перечень заданий для подготовки к тестированию

1. Что относится к описательной аналитике?
 - A) Выявление причин аномалий
 - B) Построение прогнозной модели
 - C) Анализ текущего состояния**
 - D) Генерация оптимальных стратегий
2. Что такое выброс в данных?
 - A) Повторяющееся значение
 - B) Значение, сильно отличающееся от остальных**
 - C) Отсутствующее значение

- D) Категориальный признак
3. Какой метод используется для оценки линейной зависимости двух переменных?
A) Мода
B) Дисперсия
C) Корреляция
D) Медиана
 4. Что представляет собой EDA (Exploratory Data Analysis)?
A) Метод оптимизации моделей
B) Глубокое обучение
C) Первичный анализ и визуализация данных
D) Процесс сбора данных
 5. Что такое переобучение (overfitting)?
A) Недостаточное обучение модели
B) Модель хорошо работает на новых данных
C) Модель слишком точно запомнила обучающую выборку
D) Проблема сбора данных
 6. Какой алгоритм применяется для кластеризации?
A) Логистическая регрессия
B) K-Means
C) Линейная регрессия
D) SVM
 7. Что означает высокая дисперсия модели?
A) Модель стабильна
B) Модель склонна к переобучению
C) Модель недообучена
D) Модель работает на несбалансированных данных
 8. Что используется для уменьшения размерности данных?
A) ROC-кривая
B) PCA
C) Decision Tree
D) R-счет
 9. Какой тип данных относится к категориальным?
A) Возраст
B) Доход
C) Пол
D) Температура
 10. Что такое F1-мера?
A) Среднее значение
B) Мера разброса
C) Среднее гармоническое точности и полноты
D) Способ визуализации данных

7.2.2 Примерный перечень заданий для решения стандартных задач

1. У вас есть столбец с пропущенными значениями. Какой метод будет наименее предпочтительным для количественных данных?
A) Заполнение медианой
B) Заполнение средним
C) Удаление всех строк с пропусками
D) Интерполяция
2. Найдите среднее значение: [3, 7, 8, 5, 10]
A) 6

- В) 7
С) 5
D) 6.6
3. Какой коэффициент корреляции ближе всего к отсутствию связи?
A) 0
B) 1
C) -1
D) 0.9
4. Какой метод классификации чувствителен к масштабу признаков?
A) Дерево решений
B) kNN
C) Наивный Байес
D) Логистическая регрессия
5. Какая метрика **наилучше** подходит для несбалансированных классов?
A) Accuracy
B) Precision
C) F1-мера
D) MAE
6. У вас есть числовой признак с большими выбросами. Какой метод масштабирования предпочтительнее?
A) StandardScaler
B) MinMaxScaler
C) RobustScaler
D) OneHotEncoder
7. Метод DBSCAN определяет:
A) Кластеры равного размера
B) Иерархические группы
C) Плотностные кластеры
D) Группы с одинаковым числом точек
8. При уменьшении размерности PCA сохраняет:
A) Категории
B) Метки классов
C) Максимум дисперсии
D) Плотность распределения
9. Какой метод используется для оценки качества модели на новых данных?
A) Кросс-валидация
B) Балансировка классов
C) Стандартизация
D) Гиперпараметры
10. Какой из следующих алгоритмов подходит для задачи регрессии?
A) K-Means
B) DBSCAN
C) Линейная регрессия
D) Логистическая регрессия

7.2.3 Примерный перечень заданий для решения прикладных задач

1. Вы анализируете данные продаж по регионам. Что следует сделать в первую очередь?
A) Построить модель
B) Выполнить EDA
C) Подобрать гиперпараметры

- D) Провести кластеризацию
2. В данных о клиентах обнаружены пропущенные значения возраста. Что предпочтительнее?
- A) Удалить весь столбец
B) Заполнить медианой
C) Заполнить нулями
D) Оставить как есть
3. При анализе клиентской базы вы хотите сгруппировать клиентов по поведению. Что примените?
- A) Регрессию
B) Классификацию
C) Кластеризацию
D) Тест Стьюдента
4. Вы обучили модель, которая показывает 99% точности, но работает плохо на новых данных. В чём причина?
- A) Недообучение
B) Отсутствие выбросов
C) Переобучение
D) Хорошая генерализация
5. Вы работаете с датасетом, где целевая переменная — стоимость квартиры. Что используете?
- A) Логистическая регрессия
B) KNN-классификатор
C) Линейная регрессия
D) DBSCAN
6. Вы используете t-SNE для визуализации. Почему?
- A) Уменьшить дисперсию
B) Отобрать признаки
C) Получить 2D-проекцию сложных данных
D) Провести классификацию
7. Вы строите модель для оценки кредитного риска. Что важно при оценке качества модели?
- A) Ассигасу
B) ROC-AUC
C) MAE
D) t-SNE
8. В модели обнаружена сильная корреляция между двумя признаками. Что предпочтительно?
- A) Удалить один из признаков**
B) Добавить еще признаки
C) Повысить размерность
D) Ничего не делать
9. При сегментации клиентов модель определила 4 кластера. Что следует сделать?
- A) Проверить интерпретируемость кластеров**
B) Объединить их вручную
C) Удалить кластеры
D) Повторить классификацию
10. Для ускорения обучения большой модели вы хотите уменьшить размерность признаков. Что используете?
- A) KNN
B) PCA
C) DBSCAN

D) EDA

7.2.4 Примерный перечень вопросов для подготовки к зачету

1. Назовите основные виды аналитики данных и кратко охарактеризуйте каждый.
2. Опишите этапы аналитического цикла.
3. Какие методы используются для обработки пропущенных значений и выбросов?
4. В чём состоит цель разведочного анализа данных (EDA)?
5. Каковы основные характеристики распределения данных?
6. Какие метрики используются для оценки классификационной модели?
7. Объясните разницу между обучением с учителем и без учителя.
8. Как работают алгоритмы кластеризации K-Means и DBSCAN?
9. Что такое кросс-валидация и зачем она нужна?
10. Как понять, что модель переобучена или недообучена?

7.2.5 Примерный перечень заданий для подготовки к экзамену

Не предусмотрено учебным планом

7.2.6. Методика выставления оценки при проведении промежуточной аттестации

Зачет проводится по тест-билетам, каждый из которых содержит 20 вопросов. Каждый правильный ответ на вопрос в тесте оценивается 1 баллом. Максимальное количество набранных баллов – 20.

1. Оценка «Не зачтено» ставится в случае, если студент набрал менее 15 баллов.
2. Оценка «Зачтено» ставится в случае, если студент набрал от 15 до 20 баллов.

7.2.7 Паспорт оценочных материалов

№ п/п	Контролируемые разделы (темы) дисциплины	Код контролируемой компетенции	Наименование оценочного средства
1	Введение в аналитику данных	УК-1	тест
2	Аналитический процесс и источники данных	УК-1	тест
3	Предобработка и очистка данных	УК-1	тест
4	Разведочный анализ данных (EDA)	УК-1	тест
5	Базовые статистические методы в аналитике данных	УК-1	тест
6	Машинное обучение для аналитики данных	УК-1	тест
7	Классификация и регрессия	УК-1	тест
8	Кластеризация и понижение размерности	УК-1	тест
9	Оценка и валидация моделей	УК-1	тест

7.3. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности

Тестирование осуществляется, либо при помощи компьютерной

системы тестирования, либо с использованием выданных тест-заданий на бумажном носителе. Время тестирования 30 мин. Затем осуществляется проверка теста экзаменатором и выставляется оценка согласно методики выставления оценки при проведении промежуточной аттестации.

Решение стандартных задач осуществляется, либо при помощи компьютерной системы тестирования, либо с использованием выданных задач на бумажном носителе. Время решения задач 30 мин. Затем осуществляется проверка решения задач экзаменатором и выставляется оценка, согласно методики выставления оценки при проведении промежуточной аттестации.

Решение прикладных задач осуществляется, либо при помощи компьютерной системы тестирования, либо с использованием выданных задач на бумажном носителе. Время решения задач 30 мин. Затем осуществляется проверка решения задач экзаменатором и выставляется оценка, согласно методики выставления оценки при проведении промежуточной аттестации.

8 УЧЕБНО МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ)

8.1 Перечень учебной литературы, необходимой для освоения дисциплины

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: <https://e.lanbook.com/book/381533>. — Режим доступа: для авториз. пользователей.

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: <https://e.lanbook.com/book/408569>. — Режим доступа: для авториз. пользователей.

3. Котиков, П. Е. Анализ данных : учебно-методическое пособие / П. Е. Котиков. — Санкт-Петербург : СПбГПУ, 2019. — 48 с. — ISBN 978-5-907184-46-6. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/174498>. — Режим доступа: для авториз. пользователей.

8.2 Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень лицензионного программного обеспечения, ресурсов информационно-телекоммуникационной сети «Интернет», современных профессиональных баз данных и информационных справочных систем:

- Базы данных и поисковые системы:
- электронный фонд правовой и нормативно-технической документации <http://docs.cntd.ru/>
 - поисковая система Яндекс <https://www.yandex.ru/>
 - поисковая система Google <https://www.google.ru/>

- база знаний Энциклопедия_анализа_данных
<http://www.machinelearning.ru/>

9 МАТЕРИАЛЬНО-ТЕХНИЧЕСКАЯ БАЗА, НЕОБХОДИМАЯ ДЛЯ ОСУЩЕСТВЛЕНИЯ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА

Лекционная аудитория с проекционным оборудованием, компьютерный класс с доступом к сети «Интернет».

10. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ (МОДУЛЯ)

По дисциплине «Аналитика данных» читаются лекции, проводятся лабораторные работы.

Основой изучения дисциплины являются лекции, на которых излагаются наиболее существенные и трудные вопросы, а также вопросы, не нашедшие отражения в учебной литературе.

Лабораторные работы выполняются на лабораторном оборудовании в соответствии с методиками, приведенными в указаниях к выполнению работ.

Вид учебных занятий	Деятельность студента
Лекция	Написание конспекта лекций: кратко, схематично, последовательно фиксировать основные положения, выводы, формулировки, обобщения; пометать важные мысли, выделять ключевые слова, термины. Проверка терминов, понятий с помощью энциклопедий, словарей, справочников с выписыванием толкований в тетрадь. Обозначение вопросов, терминов, материала, которые вызывают трудности, поиск ответов в рекомендуемой литературе. Если самостоятельно не удастся разобраться в материале, необходимо сформулировать вопрос и задать преподавателю на лекции или на практическом занятии.
Лабораторная работа	Лабораторные работы позволяют научиться применять теоретические знания, полученные на лекции при решении конкретных задач. Чтобы наиболее рационально и полно использовать все возможности лабораторных для подготовки к ним необходимо: следует разобрать лекцию по соответствующей теме, ознакомиться с соответствующим разделом учебника, проработать дополнительную литературу и источники, решить задачи и выполнить другие письменные задания.
Самостоятельная работа	Самостоятельная работа студентов способствует глубокому усвоения учебного материала и развитию навыков самообразования. Самостоятельная работа предполагает следующие составляющие: - работа с текстами: учебниками, справочниками, дополнительной литературой, а также проработка конспектов лекций; - выполнение домашних заданий и расчетов; - работа над темами для самостоятельного изучения; - участие в работе студенческих научных конференций, олимпиад; - подготовка к промежуточной аттестации.
Подготовка к промежуточной аттестации	Готовиться к промежуточной аттестации следует систематически, в течение всего семестра. Интенсивная подготовка должна начаться не позднее, чем за месяц-полтора до промежуточной аттестации. Данные перед зачетом три дня эффективнее всего использовать для повторения

	и систематизации материала.
--	-----------------------------

ЛИСТ РЕГИСТРАЦИИ ИЗМЕНЕНИЙ

№ п/п	Перечень вносимых изменений	Дата внесения изменений	Подпись заведующего кафедрой, ответственной за реализацию дисциплины
----------	-----------------------------	----------------------------	---