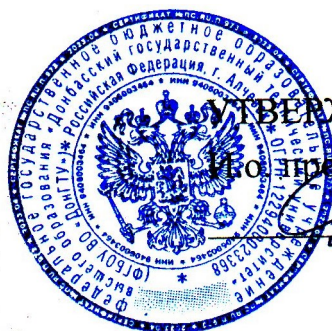


МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
(МИНОБРНАУКИ РОССИИ)

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«ДОНБАССКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»
(ФГБОУ ВО «ДонГТУ»)

Факультет	информационных технологий и автоматизации производственных процессов
Кафедра	интеллектуальных систем и информационной безопасности



УТВЕРЖДАЮ

Проректор по учебной работе
Д.В. Мулов

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Методы анализа данных

(наименование дисциплины)

10.05.03 Информационная безопасность автоматизированных систем

(код, наименование специальности)

Безопасность открытых информационных систем

(специализация)

Квалификация	специалист по защите информации
	(бакалавр/специалист/магистр)

Форма обучения	очная
	(очная, очно-заочная, заочная)

Алчевск, 2024

1 Цели и задачи изучения дисциплины

Цели дисциплины. Целью изучения дисциплины «Методы анализа данных» является изучить технологии анализа данных: OLAP, KDD, DataMining и подготовки данных; дать представление об автоматизированных моделях анализа данных, применить методы анализа данных на примере решения задач сегментации, классификации, прогнозирования.

Задачи изучения дисциплины:

- изучение понятийно-категориального аппарата в области углубленного анализа данных;
- формирование представлений об общей методологии консолидации, подготовки и анализа данных;
- обеспечение освоения современных методов OLAP, KDD, DataMining;
- формирование навыков и умений, необходимых для создания и развития корпоративных аналитических систем.

Дисциплина направлена на формирование общепрофессиональной (ОПК-5.3) компетенции выпускника.

2 Место дисциплины в структуре образовательной программы

Логико-структурный анализ дисциплины – курс входит в БЛОК 1 «Дисциплины (модули)», обязательную часть подготовки студентов по специальности 10.05.03 Информационная безопасность автоматизированных систем (10.05.03-05 Безопасность открытых информационных систем).

Дисциплина реализуется кафедрой интеллектуальных систем и информационной безопасности.

Основывается на базе дисциплин: «Основы информационной безопасности», «Основы программирования», «Информатика», «Алгебра».

Является основой для изучения следующих дисциплин: «Базы данных», «Операционные системы».

Курс является фундаментом для ориентации студентов в сфере разработки программного обеспечения информационных систем.

Общая трудоемкость освоения дисциплины составляет 4 зачетных единицы, 144 ак.ч. Программой дисциплины предусмотрены лекционные (36 ак.ч.), лабораторные (36 ак.ч.) занятия и самостоятельная работа студента (72 ак.ч.).

Дисциплина изучается на 2 курсе в 4 семестре. Форма промежуточной аттестации – зачет.

3 Перечень результатов обучения по дисциплине, соотнесённых с планируемыми результатами освоения образовательной программы

Процесс изучения дисциплины «Методы анализа данных» направлен на формирование компетенции, представленной в таблице 1.

Таблица 1 – Компетенции, обязательные к освоению

Содержание компетенции	Код компетенции	Код и наименование индикатора достижения компетенции
Способен осуществлять контроль обеспечения информационной безопасности и проводить верификацию данных в открытых информационных системах	ОПК-5.3	ОПК-5.3.2. Проводит верификацию данных в открытых информационных системах

4 Объём и виды занятий по дисциплине

Общая трудоёмкость учебной дисциплины составляет 4 зачётных единицы, 144 ак.ч.

Самостоятельная работа студента (СРС) включает проработку материалов лекций, подготовку к практическим занятиям, текущему контролю, выполнение индивидуального задания, самостоятельное изучение материала и подготовку к зачету.

При организации внеаудиторной самостоятельной работы по данной дисциплине используются формы и распределение бюджета времени на СРС для очной формы обучения в соответствии с таблицей 2.

Таблица 2 – Распределение бюджета времени на СРС

Вид учебной работы	Всего ак.ч.	Ак.ч. по семестрам
		4
Аудиторная работа, в том числе:	72	72
Лекции (Л)	36	36
Практические занятия (ПЗ)	–	–
Лабораторные работы (ЛР)	36	36
Курсовая работа/курсовой проект	–	-
Самостоятельная работа студентов (СРС), в том числе:	72	72
Подготовка к лекциям	9	9
Подготовка к лабораторным работам	18	18
Подготовка к практическим занятиям / семинарам	–	–
Выполнение курсовой работы / проекта	–	–
Расчетно-графическая работа (РГР)	–	–
Реферат (индивидуальное задание)	10	10
Домашнее задание	–	–
Подготовка к контрольным работам	–	–
Подготовка к коллоквиуму	–	–
Аналитический информационный поиск	12	12
Работа в библиотеке	17	17
Подготовка к зачету	6	6
Промежуточная аттестация –зачет (З)	3	3
Общая трудоемкость дисциплины		
ак.ч.	144	144
з.е.	4	4

5 Содержание дисциплины

С целью освоения компетенции, приведенной в п.3 дисциплина разбита на 2 темы:

- тема 1 (Основы программирования на языке Python);
- тема 2 (Модели анализа данных);

Виды занятий по дисциплине и распределение аудиторных часов для очной формы приведены в таблице 3.

Таблица 3 – Виды занятий по дисциплине и распределение аудиторных часов (очная форма обучения)

№ п/п	Наименование темы (раздела) дисциплины	Содержание лекционных занятий	Трудоемкость в ак.ч.	Темы практических занятий	Трудоемкость в ак.ч.	Тема лабораторных занятий	Трудоемкость в ак.ч.
1	2	3	4	5	6	7	8
1	Основы программирования на языке Python	Особенности языка. Синтаксис и базовые конструкции. Структуры данных. Модули. Ветвления и циклы. Элементы функционального программирования. Объектно-ориентированное программирование. Средства визуализации, библиотека matplotlib.	8	–	–	Программирование на Python	6
2	Модели анализа данных	Векторное пространство. Операции над векторами и матрицами, их реализация в Python. Библиотека numpy. Понятие числовых данных. Выборка. Статистики. Статистические распределения. Вероятность. Нормальное распределение. Теорема Байеса. Статистический вывод. Проверка статистических гипотез	6	–	–	Формулировка и проверка статистических гипотез.	4

Окончание таблицы 3

1	2	3	4	5	6	7	8
2	Модели анализа данных	Принципы сбора данных. Работа с данными с помощью pandas. Исследование и управление данными. Параметризация и математическое моделирование.	6	—	—	Извлечение данных из web-ресурсов.	8
		Метод кближайших соседей. Наивный байесовский фильтр. Регрессия. Оптимизация градиентным спуском: пакетным и стохастическим. Деревья принятия решений. Случайные леса	8	—	—	Регрессионный анализ.	8
		Задача кластеризации. Метод ксредних. Восходящая и нисходящая кластеризация. Нелинейные зависимости в данных. Метрики и трансформеры. Естественный язык. Грамматики. Тематическое моделирование. Понятие о семантическом анализе.	8	—	—	Кластеризация и поиск зависимостей в наборе данных	10
Всего аудиторных часов			36	—		36	

6 Фонд оценочных средств для проведения текущего контроля успеваемости и промежуточной аттестации студентов по дисциплине

6.1 Критерии оценивания

В соответствии с Положением о кредитно-модульной системе организации образовательного процесса ФГБОУ ВО «ДонГТУ» (https://www.dstu.education/images/structure/license_certificate/polog_kred_modul.pdf) при оценивании сформированности компетенций по дисциплине используется 100-балльная шкала.

Перечень компетенций по дисциплине и способы оценивания знаний приведены в таблице 5.

Таблица 5 – Перечень компетенций по дисциплине и способы оценивания знаний

Код и наименование компетенции	Способ оценивания	Оценочное средство
ОПК-5.3	Зачет	Комплект контролирующих материалов для зачета

Всего по текущей работе в семестре студент может набрать 100 баллов, в том числе:

– лабораторные работы – всего 100 баллов.

Зачет проставляется автоматически, если студент набрал в течении семестра не менее 60 баллов и отчитался за каждую контрольную точку. Минимальное количество баллов по каждому из видов текущей работы составляет 60% от максимального.

Зачет по дисциплине «Методы анализа данных» проводится по результатам работы в семестре. В случае, если полученная в семестре сумма баллов не устраивает студента, во время сессии студент имеет право повысить итоговую оценку либо в форме устного собеседования по приведенным ниже вопросам (п.п. 6.5), либо в результате тестирования.

Шкала оценивания знаний при проведении промежуточной аттестации приведена в таблице 6.

Таблица 6 – Шкала оценивания знаний

Сумма баллов за все виды учебной деятельности	Оценка по национальной шкале зачёт/экзамен
0-59	Не зачтено/неудовлетворительно
60-73	Зачтено/удовлетворительно
74-89	Зачтено/хорошо
90-100	Зачтено/отлично

6.2 Домашние задания

Домашние задания не предусмотрены.

6.3 Темы для рефератов (презентаций) – индивидуальное задание

- 1) Преимущества использования Python для анализа данных по сравнению с другими инструментами.
- 2) Библиотеки для работы с данными в Python.
- 3) Основы работы с dataframe: загрузка, очистка, фильтрация, группировка и агрегация.
- 4) Возможности Pandas для анализа данных.
- 5) Принципы хорошей визуализации.
- 6) Основные ошибки при визуализации данных.
- 7) Визуализация данных в Python.
- 8) Основы визуализации в matplotlib.
- 9) Оформление и настройка графиков.
- 10) Визуализации отфильтрованных и сгруппированных данных.
- 11) Создание интерактивных визуализаций в Plotly.
- 12) NumPy – основа для многих библиотек, используемых в анализе данных.
- 13) Математический аппарат (refresher) для анализа данных.
- 14) Текстовый анализ.
- 15) Виды задач машинного обучения.
- 16) Baseline модели.
- 17) Сравнение Python и других языков по синтаксису языка.
- 18) Возможности языка Python по решению задач линейной алгебры и задач интерполяции.
- 19) Вставка кода на C++ в программы Python.
- 20) Подключение динамических библиотек к программам Python.

6.4 Оценочные средства для самостоятельной работы и текущего контроля успеваемости

Тема 1. Основы программирования на языке Python.

- 1) Обязательно ли в Python объявлять тип переменной?
 - а) по желанию программиста;
 - б) обязательно;
 - в) зависит от ситуации;
 - г) объявлять тип не нужно.
- 2) Какие из предлагаемых имен допустимы в Python?
 - а) s_5; б) _5s; в) 5_s; г) Щ5.
- 3) Какая инструкция используется в Python для определения функций?
 - а) df; б) def; в) dfn; г) define.
- 4) Какой тип данных в Python представляет кортежи?
 - а) dict; б) list; в) tuple; г) set.
- 5) Какой тип данных в Python представляет словари?
 - а) dict; б) list; в) tuple; г) set.
- 6) Какие из представленных литералов относятся к типу set?
 - а) {'кот', 'ток'}; б) "{1, 2, 3}"; в) {[], (), {}, ""}; г) [{'set': {1, 2, 3}}].
- 7) Какое ключевое слово используется для импорта модуля в Python?
 - а) import; б) include; в) load; г) require.
- 8) Какой атрибут хранит путь к файлу текущего модуля?
 - а) __name__; б) ____; в) __path__; г) __file__.

Тема 2. Модели анализа данных.

- 1) За что отвечает данный фрагмент кода: `d = np.eye(5)`
 - а) удаление из массива 5 элементов;
 - б) создание пустого массива;
 - в) очищение массива;
 - г) создание единичной матрицы.
- 2) 7. Каков будет результат выполнения данного кода:


```
b = np.array([[1.5, 2, 3], [4, 5, 6]])
d = b.copy()
d is a
```

 - а) 0
 - б) 1
 - в) True
 - г) False
- 3) Какой метод отвечает за объединение массивов по вертикали?
 - а) stack()
 - б) hstack()

- в) `vstack()`
- г) `ver_stack()`

4) Как поменять местами две строки в двумерном массиве NumPy?
Поменяйте местами строки 1 и 3 массива `a`.

Ответ: `a = np.arange(9).reshape(3,3)`

5) Как загрузить файл в `numpy`?

- а) `.open(file_name, access_mode);`
- б) `.read_csv[exsel ets](file_name);`
- в) `.loadtxt(file_name, delimiter = ", skiprows =)pd.`

6) Какой метод возвращает количество раз, когда указанный элемент появляется в списке?

- а) `.count();`
- б) `.info();`
- в) `.isna().sum();`
- г) `.isna().mean().`

7) Что делает `.append()`?

- а) добавляет новый элемент в конец списка;
- б) это можно применять только с строками;
- в) удаляет элемент, указанный в скобках, из списка.

6.5 Вопросы для подготовки к зачету

1) Что такое предмет DataScience, в чем его актуальность и применимость в сфере информационной безопасности?

2) Какие есть структуры данных языка Python?

3) В чем состоит организация управляющих конструкций в языке Python?

4) Какие Вы знаете встраиваемые библиотеки языка Python – способы их создания, работы?

5) Что такое функциональное программирование на Python?

6) Что Вы знаете об объектно-ориентированном программировании на Python?

7) Что такое средство визуализации – библиотека `matplotlib` и в чем сложность алгоритмов поиска?

8) Реализация операций над векторами и матрицами в Python

9) Что такое функциональная реализация распределений и статистик в Python?

10) Как проверить статистические гипотезы?

11) Как использовать р-значения?

12) В чем состоит аппроксимация как статистическая гипотеза?

13) Для чего нужна библиотека `BeautifulSoup`: назначение и способы применения?

- 14) Для чего нужна библиотека pandas?
- 15) Как производится исследование одномерных, двумерных и многомерных данных?
- 16) Что такое управление данными: очистка, форматирование?
- 17) Что такое управление данными: шкалирование и снижение размерности?
- 18) Что Вы знаете о машинном обучении?
- 19) Что такое точность и полнота предсказаний модели; роль правильности предсказаний?
- 20) Что такое смещение и дисперсия в предсказаниях модели?
- 21) В чем состоит отбор признаков при создании модели?
- 22) Что такое байесовский статистический вывод?
- 23) Что такое обыкновенный (пакетный) градиентный спуск?
- 24) Что такое стохастический градиентный спуск?
- 25) Что такое метод k ближайших соседей: назначение, алгоритм, результат?
- 26) Что такое наивный Байесовский классификатор?
- 27) Что такое простая линейная регрессия?
- 28) Что такое множественная линейная регрессия?
- 29) Что такое логистическая регрессия?
- 30) Что такое бутстрапирование и регуляризация данных?
- 31) Что такое деревья принятия решений?
- 32) Деревья принятия решений: в чем сущность алгоритма на основе энтропии?
- 33) Что такое задача классификации: формулировка и подходы к решению?
- 34) В чем состоит задача кластеризации: формулировка и результат решения?
- 35) В чем состоит метод k средних?
- 36) Что такое восходящая кластеризация?
- 37) Что такое n-граммная лексика?
- 38) Что такое грамматики, мешки слов и их применения?
- 39) Что такое семантический анализ?
- 40) Что такое трансформеры?
- 41) В чем состоит принцип тематического моделирования?

7 Учебно-методическое и информационное обеспечение дисциплины

7.1 Рекомендуемая литература

Основная литература

1. Анализ данных : учебник для вузов / В. С. Мхитарян [и др.] ; под редакцией В. С. Мхитаряна. — Москва : Издательство Юрайт, 2023. — 490 с. <https://publications.hse.ru/books/166113174?ysclid=m2pz52wjt8316776826> (Дата обращения 26.08.2024).

Дополнительная литература

1. Уэс Маккинни. Python и анализ данных: Первичная обработка данных с применением pandas, NumPy и Jupiter / пер. с англ. А. А. Слинкина. 3-е изд. — М.: МК Пресс, 2023. — 536 с.: ил. URL: <https://znanium.ru/catalog/document?id=445331&ysclid=m2pz2nhbrc399130259> (Дата обращения 26.08.2024).

Учебно-методическое обеспечение

1. Бизянов, Е.Е. Методы анализа данных : лабораторный практикум / Е.Е. Бизянов, А.С. Закутный ; Каф. Специализированных компьютерных систем . — Алчевск : ГОУ ВО ЛНР ДонГТИ, 2023 . — 91 с. — URL: <https://library.dstu.education/download.php?rec=132246>.

7.2 Базы данных, электронно-библиотечные системы, информационно-справочные и поисковые системы

1. Научная библиотека ДонГТУ : официальный сайт.— Алчевск. —URL: library.dstu.education.—Текст : электронный.

2. Научно-техническая библиотека БГТУ им. Шухова : официальный сайт. — Белгород. — URL: <http://ntb.bstu.ru/jirbis2/>.—Текст : электронный.

3. Консультант студента :электронно-библиотечная система.— Москва. — URL: <http://www.studentlibrary.ru/cgi-bin/mb4x>.—Текст : электронный.

4. Университетская библиотека онлайн :электронно-библиотечная система.— URL: http://biblioclub.ru/index.php?page=main_ub_red.—Текст : электронный.

5. Сайт кафедры ИСИБ <http://scs.dstu.education>.

Материально-техническое обеспечение представлено в таблице 9.

Таблица 9 – Материально-техническое обеспечение

Наименование оборудованных учебных кабинетов	Адрес (местоположение) учебных кабинетов
Аудитория для проведения лекций <i>Компьютерные классы (22 посадочных места), оборудованные учебной мебелью, компьютерами с неограниченным доступом к сети Интернет, включая доступ к ЭБС:</i> ПК – 11 шт.; Интерактивная доска– 1 шт. ПК - 11 шт.; Доска– 1 шт.	ауд. <u>309</u> корп.<u>4</u> ауд. <u>208</u> корп. <u>4</u> ауд. <u>211</u> корп. <u>4</u>

Лист согласования рабочей программы дисциплины

Разработал
и.о. заведующего кафедрой
интеллектуальных систем и
информационной безопасности
(должность)


(подпись) Е.Е.Бизянов
Ф.И.О.)

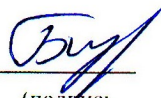
(должность)

(подпись) Ф.И.О.)

(должность)

(подпись) Ф.И.О.)

И.о. заведующего кафедрой
интеллектуальных систем и
информационной безопасности


(подпись) Е.Е.Бизянов
Ф.И.О.)

Протокол № 1 заседания кафедры ИСИБ от 27.08.2024 г

И.о. декана факультета
информационных технологий и
автоматизации производственных
процессов


(подпись) В.В. Дьячкова
Ф.И.О.)

Согласовано
Председатель методической
комиссии по специальности 10.05.03
«Информационная безопасность автоматизированных
систем»


(подпись) Е.Е. Бизянов
Ф.И.О.)

Начальник учебно-методического центра


(подпись) О.А.Коваленко
Ф.И.О.)

Лист изменений и дополнений

Номер изменения, дата внесения изменения, номер страницы для внесения изменений	
ДО ВНЕСЕНИЯ ИЗМЕНЕНИЙ:	ПОСЛЕ ВНЕСЕНИЯ ИЗМЕНЕНИЙ:
Основание:	
Подпись лица, ответственного за внесение изменений	