

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

И.о. проректора
по учебной работе

Мулов Д.В.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Методы анализа данных

(наименование дисциплины)

02.03.01 Математика и компьютерные науки

(код, наименование направления/специальности)

09.03.01 Информатика и вычислительная техника

(код, наименование направления/специальности)

10.05.03 Информационная безопасность автоматизированных систем

(код, наименование направления/специальности)

38.03.05 Бизнес-информатика

(код, наименование направления/специальности)

Квалификация	бакалавр, специалист по защите информации (бакалавр/специалист)
--------------	--

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

Алчевск, 2024

1 Цели и задачи изучения дисциплины

Цели дисциплины. Целью изучения дисциплины «Методы анализа данных» является предоставление студентам теоретических знаний и практических навыков в области современных средств и методов анализа данных.

Задачи изучения дисциплины:

- научиться использовать методы анализа данных для выявления событий информационной безопасности в открытых информационных системах;
- научиться ориентироваться в выборе современных программных средств и математического обеспечения;
- приобрести навыки применения методов анализа данных, создания программных модулей, решающих простейшие задачи сбора и обработки данных.

2 Место дисциплины в структуре ОПОП ВО

Логико-структурный анализ дисциплины – курс входит в обязательную часть БЛОКА 1 «Дисциплины (модули)» по направлениям 02.03.01 Математика и компьютерные науки, 09.03.01 Информатика и вычислительная техника, 38.03.05 Бизнес-информатика и специальности 10.05.03 Информационная безопасность автоматизированных систем.

Дисциплина реализуется кафедрой интеллектуальных систем и информационной безопасности.

Основывается на базе дисциплин: «Вычислительная математика», «Основы программирования», «Информатика», «Высшая математика».

Является основой для изучения следующих дисциплин: «Интеллектуальный анализ больших данных», «Подготовка и защита ВКР».

Для изучения дисциплины необходимы компетенции, сформированные у студента для решения профессиональных задач деятельности, связанных с разработкой программного обеспечения.

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 – экзамен.

3 Перечень результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения ОПОП ВО

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

Таблица 1 –Компетенции, обязательные к освоению

Код	Наименование специальности, направления подготовки	Компетенция (код, содержание)	Индикатор (код, наименование)
02.03.01	Математика и компьютерные науки	УК-1 Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач ОПК-1 Способен консультировать и использовать фундаментальные знания в области математического анализа, комплексного и функционального анализа алгебры, аналитической геометрии, дифференциальной геометрии и топологии, дифференциальных уравнений, дискретной математики и математической логики, теории вероятностей, математической статистики и случайных процессов, численных методов, теоретической механики в профессиональной деятельности	УК-1.1. Знает принципы сбора, отбора и обобщения информации ОПК-1.3. Имеет навыки выбора методов решения задач профессиональной деятельности на основе теоретических знаний
09.03.01	Информатика и вычислительная техника	ОПК-9 Способен осваивать методики использования программных	ОПК-9.3. Составляет алгоритмы, пишет и отлаживает коды на заданном языке программирования, тестирует

Код	Наименование специальности, направления подготовки	Компетенция (код, содержание)	Индикатор (код, наименование)
		средств для решения практических задач	работоспособность программ, интегрирует программные модули
		ПК-4 Способен разрабатывать системы анализа больших данных	ПК-4.2. Разрабатывает программные компоненты извлечения, хранения, подготовки больших данных с учетом вариантов использования больших данных, определений, словарей и эталонной архитектуры больших данных
10.05.03	Информационная безопасность автоматизированных систем	ОПК-7 Способен создавать программы на языках общего назначения, применять методы и инструментальные средства программирования для решения профессиональных задач, осуществлять обоснованный выбор инструментария программирования и способов организации программ	ОПК-7.2. Применяет методы и инструментальные средства программирования для решения профессиональных задач
38.03.05	Бизнес-информатика	ОПК-4 Способен понимать принципы работы информационных технологий; использовать информацию, методы и программные средства ее сбора, обработки и анализа для информационно-аналитической поддержки принятия управленческих решений	ОПК-4.2. Использует методы и программные средства для сбора, обработки и анализа бизнес-информации

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 экзамену.

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)	–	–
Домашнее задание	–	–
Подготовка к контрольным работам	–	–
Подготовка к коллоквиуму	–	–
Аналитический информационный поиск	9	9
Работа в библиотеке	18	9
Подготовка к экзамену	27	27
Промежуточная аттестация – экзамен (Э)	Э	Э
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 по итогам освоения дисциплины

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 работ по дисциплине и способы оценивания знаний приведены в таблице 5.

Таблица 5 – Перечень работ по дисциплине и способы оценивания знаний

Вид учебной работы	Способ оценивания	Количество баллов
Выполнение лабораторных работ	Предоставление отчетов	60-100
Итого	—	60-100

Экзамен 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.

Экзамен по дисциплине «Методы анализа данных» проводится по результатам работы в семестре. В случае, если полученная в семестре сумма баллов не устраивает студента, во время зачетной 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

Рефераты (индивидуальные задания) отсутствуют.

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 экзамену

- 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) Что такое библиотека BeautifulSoup: назначение и способы применения?
- 14) Что такое 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?

6.6 Примерная тематика курсовых работ

Отсутствует.

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>.

8 Материально-техническое обеспечение дисциплины

Материально-техническая база обеспечивает проведение всех видов деятельности в процессе обучения, соответствует требованиям ФГОС ВО.

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), оборудованный учебной мебелью, компьютерами с неограниченным доступом к сети Интернет, включая доступ к ЭБС:</i> ПК – 11шт.; 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 РПД

Разработал

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


(подпись) Е.Е. Бизянов
(Ф.И.О.)

И.о. заведующего кафедрой
интеллектуальных систем и
информационной безопасности
(наименование кафедры)


(подпись) Е.Е. Бизянов
(Ф.И.О.)

Протокол № 5 заседания кафедры

от 09.10 2024 г.

Согласовано

Председатель методической
комиссии по направлению подготовки 02.03.01
Математика и компьютерные науки


(подпись) А.Н. Баранов
(Ф.И.О.)

Председатель методической
комиссии по направлению подготовки 09.03.01
Информатика и вычислительная техника


(подпись) Е.Е. Бизянов
(Ф.И.О.)

Председатель методической
Комиссии по специальности 10.05.03
Информационная безопасность
автоматизированных систем


(подпись) Е.Е. Бизянов
(Ф.И.О.)

Председатель методической
комиссии по направлению подготовки 38.03.05
Бизнес-информатика


(подпись) Н.Н. Лепило
(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	