

ДОПОЛНЕНИЯ И ИЗМЕНЕНИЯ К УЧЕБНОЙ ПРОГРАММЕ ПО ИЗУЧАЕМОЙ УЧЕБНОЙ ДИСЦИПЛИНЕ

Анализ данных
для специальности

1-31 03 09 Компьютерная математика и системный анализ
от 5.07.2023, № УД – 12911/уч
на 2025/2026 учебный год

№ п/п	Дополнения и изменения	Основание
1	Изменение содержания учебного материала. Содержание учебного материала для программы № УД – 12911/уч на 2025/2026 учебный год прилагается.	Актуализация содержания учебной дисциплины.
2	Изменения учебно-методической карты. Учебно-методическая карта для программы № УД – 12911/уч на 2025/2026 учебный год прилагается.	Актуализация содержания учебной дисциплины.

Учебная программа пересмотрена и одобрена на заседании кафедры дифференциальных уравнений и системного анализа (Протокол № 15 от 23.06.25)

Заведующий кафедрой дифференциальных
уравнений и системного анализа
кандидат.ф.-м.наук., доцент

 Л.Л.Голубева

УТВЕРЖДАЮ

Декан факультета
доктор.ф.-м.наук., профессор

 С.М.Босяков

СОДЕРЖАНИЕ УЧЕБНОГО МАТЕРИАЛА

Раздел 1. Анализ данных

Тема 1.1. Введение анализа данных

Основные понятия и принципы анализа данных. Инструменты и методы анализа данных. Анализ данных как процесс исследования, очистки, преобразования, интерпретации и моделирования данных с целью выявления закономерностей, трендов, корреляций. Использование в различных областях: бизнес, наука, медицина, финансы, маркетинг.

Основные этапы: сбор данных, предварительная обработка, визуализация, статистический анализ, построение моделей и проверка гипотез. Инструменты анализа данных: программные пакеты, статистические методы, машинное обучение. Цели: получение информации для принятия решений, прогнозирования будущих событий, выявления закономерностей.

Тема 1.2. Python для анализа и визуализации данных

Общие положения языка Python. Jupyter Notebook. Построение и синтаксис. Модули и импорт модулей. Функции. Сложность алгоритмов.

Структуры и типы данных.

Управляющие конструкции и общие методы. Работа с логическими значениями. Сортировка. Циклы. Генераторы последовательностей.

Инструменты функционального и объектно-ориентированного программирования. Определение класса и его структура. Модули random, re. Встроенные функции map, reduce и filter.

Тема 1.3. Визуальный анализ данных

Средства визуализации данных. Библиотеки Matplotlib, Seaborn. Графики, столбчатые диаграммы, диаграммы рассеяния (scatterplot). Дополнительные инструменты: Tableau, Power BI. Статистика и отчеты по исследованиям данных. Структура, связи и закономерности в данных. Принятие решений на основе визуального анализа данных.

Тема 1.4. Математические основы анализа данных (основы статистики и теории вероятности)

Линейная алгебра. Действия над векторами и матрицами. Специальные типы матриц. Решение задач линейной алгебры при помощи библиотеки NumPy. Функция gauss, решение СЛАУ.

Математическая статистика и вероятность. Комбинаторика. Случайные величины, выборки. Вариация, ковариация, корреляция. Размах, дисперсия, стандартное отклонение, интерквартильный размах. Условная вероятность и формула Байеса. Теорема Байеса. Практическое применение.

Случайные величины и распределение. Математическое ожидание. Плотность вероятности случайной величины. Функция распределения. Функция ошибок.

Тема 1.5. Математические основы анализа данных

Проверка статистических гипотез. Схема Бернулли. Критерий уровня значимости. Ошибки 1-го и 2-го рода. Метод проверки гипотез на основе вероятности

ошибки первого рода – p -значение. Доверительный интервал и подгонка p -значения (p -hacking).

Функциональные зависимости, аппроксимация как статистическая гипотеза. Пример с испанкой в Нью-Йорке (экспоненциальная аппроксимация). Байесовский статистический вывод. Априорное, и апостериорное распределение. Параметры распределения.

Тема 1.6. Сбор данных

Извлечение данных из веб-ресурсов. Библиотека `requests` для `http`-запросов. Скрейпинг. Библиотека `BeautifulSoup` для создания в памяти структур данных. Кодировки `Unicode` и др. Запрос `requests.get()`.

Тема 1.7. Обработка данных

Работа с табличными данными из файлов с помощью библиотеки `Pandas`. Чтение, обработка и агрегация данных. Структуры данных и методы `Pandas`. Преобразование данных в/из структуры `Pandas`.

Исследование данных и управление ими. Многомерные данные. Очистка и форматирование. Управление данными. Шкалирование. Проблемы качества и размерности данных. Снижение размерности.

Тема 1.8. Простые прикладные модели

Машинное обучение модели для предсказания поведения систем. Переобучение и недообучение. Правильность предсказаний модели. Смещение и дисперсия. Извлечение и отбор признаков.

Прикладные модели. Метод k ближайших соседей. Наивный байесовский классификатор.

Тема 1.9. Оптимизация градиентным спуском

Оптимизация градиентным спуском. Виды градиентного спуска и его оптимизация. Стохастический градиентный спуск. Функции оптимизации. Проведение исследования для задачи безусловной оптимизации и тестовых целевых функций Розенброка, Экли, Швевеля.

Тема 1.10. Модели, методы и алгоритмы работы с данными.

Регрессионный анализ

Регрессионный анализ. Линейная регрессия и метод наименьших квадратов. Сведение к задаче оптимизации. Множественная регрессия. Полиномиальная регрессия. Бутстрапирование данных. Регуляризация. Логистическая регрессия. Метод опорных векторов. Метрики качества для алгоритмов регрессии. Функция потерь (`loss function`). Создание и обучение модели.

Деревья принятия решений. Понятие дерева принятия решений. Энтропия. Алгоритм `ID3`.

Тема 1.11. Классификация данных

Общие сведения. Постановка задачи классификации. Классификация данных с учителем и без учителя. Бинарная классификация. Множественная классификация.

Качество классификации. Простейшие алгоритмы классификации: дерево решений, метод ближайших соседей. Метод снижения размерности. Автокодировщики (приведение многомерного вектора к меньшей размерности и его восстановление). Нормирование данных. Эффекты недообученности и переобученности.

Тема 1.12. Кластерный анализ

Модель и примеры кластеризации. Задача кластеризации. Метод k средних. Выбор числа k. Иерархическая кластеризация. Восходящий метод.

Раздел 2. Современные технологии обработки, хранения и поиска данных

Тема 2.1. Введение. Общий обзор актуальных СУБД

Реляционные СУБД на примере Oracle и PostgreSQL и NoSQL СУБД на примере MongoDB. Их достоинства и недостатки при использовании в задачах, требующих обработки большого количества разнородных данных.

Тема 2.2. Правила нормализации данных

Пять форм нормализации. Преимущества и недостатки нормализации. Использование единой универсальной таблицы как решение проблемы нормализации. Структура единой универсальной таблицы.

Тема 2.3. Международный стандарт ISAD(G) архивного описания данных

Основные положения международного стандарта ISAD(G) архивного описания данных. Примеры.

Тема 2.4. Распространение стандарта ISAD(G) на любые базы данных

Способы описания любых объектов из любой предметной области при помощи набора полей, предлагаемых стандартом ISAD(G).

Тема 2.5. Особенности индексации данных

Индексы. Уникальные индексы. Полнотекстовый индекс. Применение индексов в реальных проектах БД. Преимущества полнотекстового индекса при работе с большими данными.

Тема 2.6. Метод введения «искусственных» объектов для упрощения работы с большим объёмом данных

Кластеризация. Понятие искусственного объекта в контексте баз данных. Промежуточные, абстрактные или служебные структуры.

Тема 2.7. Диапазонные индексы для повышения производительности

Диапазонные индексы и методы их использования для упрощения и ускорения поиска данных. Использование диапазонных индексов при проектировании оптимальных групп для эффективного поиска.

Тема 2.8. Особенности программирования при взаимодействии с хранилищем данных.

Универсальная таблица. Запись универсальной таблицы. Сравнение по всем свойствам и характеристикам подходов к проектированию БД.

УЧЕБНО-МЕТОДИЧЕСКАЯ КАРТА УЧЕБНОЙ ДИСЦИПЛИНЫ

Очная форма получения высшего образования с применением дистанционных образовательных технологий (ДОТ)

Номер раздела, темы	Название раздела, темы	Количество аудиторных часов					Количество часов УСР	Форма контроля знаний
		Лекции	Практические занятия	Семинарские занятия	Лабораторные занятия	Иное		
1	2	3	4	5	6	7	8	9
1.	Анализ данных	34			28		6	
1.1	Введение в анализ данных	2			2			Письменный отчет по лабораторной работе.
1.2	Python для анализа и визуализации данных	2			2			Письменный отчет по лабораторной работе.
1.3	Визуальный анализ данных	2			2			Письменный отчет по лабораторной работе.
1.4	Математические основы анализа данных (основы статистики и теории вероятности)	2			2			Письменный отчет по лабораторной работе. Контрольная работа.
1.5	Математические основы анализа данных	4			4			Письменный отчет по лабораторной работе.
1.6	Сбор данных	2			2			Письменный отчет по лабораторной работе.
1.7	Обработка данных	4			2		2	Письменный отчет по лабораторной работе. Контрольная работа.
1.8	Простые прикладные модели	4			4			Письменный отчет по лабораторной работе.
1.9	Оптимизация градиентным спуском	2			2			Письменный отчет по лабораторной работе.
1.10	Модели, методы и алгоритмы работы с данными. Регрессионный анализ	4			2			Письменный отчет по лабораторной работе.
1.11	Классификация данных	2			2			Письменный отчет по лабораторной работе.
1.12	Кластерный анализ	4			2		4	Контрольная работа.
2.	Современные технологии обработки, хранения и поиска данных	18			14		4	

2.1	Введение. Общий обзор актуальных СУБД	2			2			
2.2	Правила нормализации данных	2			2			Отчет по лабораторной работе с устной защитой. Контрольная работа
2.3	Международный стандарт ISAD(G) архивного описания данных	2			2			Отчет по лабораторной работе с устной защитой
2.4	Распространение стандарта ISAD(G) на любые базы данных	2			2			Отчет по лабораторной работе с устной защитой
2.5	Особенности индексации данных	2			2			Отчет по лабораторной работе с устной защитой
2.6	Метод введения «искусственных» объектов для упрощения работы с большим объемом данных	2			2			Отчет по лабораторной работе с устной защитой
2.7	Диапазонные индексы для повышения производительности	2			2			Отчет по лабораторной работе с устной защитой. Контрольная работа
2.8	Особенности программирования при взаимодействии с хранилищем данных.	4					4	Контрольная работа