

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

УТВЕРЖДАЮ

Проректор по учебной работе и
образовательным инновациям

О.Г. Прохоренко

«05» июля 2023 г.

Регистрационный № УД – 12911/уч.

АНАЛИЗ ДАННЫХ

Учебная программа учреждения высшего образования
по учебной дисциплине для специальности:

1-31 03 09 Компьютерная математика и системный анализ

2023 г.

Учебная программа составлена на основе ОСВО 1-31 03 09-2021, утверждён постановлением № 98 от 25.04.2022 г., типового учебного плана, регистрационный № G31-1-021/пр-тип. от 21.04.2021 г., учебных планов: № G31-1-019/уч. от 25.05.2021, № G 31-1-004/уч. ин. от 31.05.2021, № G31-1-222/уч. от 22.03.2022, №G31-1-226/уч. ин. от 27.05.2022.

СОСТАВИТЕЛИ:

М.Н. Василевич, доцент кафедры дифференциальных уравнений и системного анализа механико-математического факультета Белорусского государственного университета, кандидат физико-математических наук, доцент;

РЕЦЕНЗЕНТЫ:

Буряков Александр Андреевич, инженер-программист, ИООО "ЭКСАДЕЛ"

РЕКОМЕНДОВАНА К УТВЕРЖДЕНИЮ:

Кафедрой дифференциальных уравнений и системного анализа БГУ
(протокол № 13 от 14.06.2023);

Научно-методическим советом БГУ
(протокол № 9 от 29.06.2023)

Зав. кафедрой дифференциальных уравнений
и системного анализа



Л.Л. Голубева

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

Цели и задачи учебной дисциплины

Учебная программа по учебной дисциплине «Анализ данных» разработана и утверждена для обучающихся по специальности 1-31 03 09 Компьютерная математика и системный анализ. В учебной дисциплине изучаются теоретические и практические основы анализа данных.

Анализ данных – один из важнейших разделов математической статистики и информатики, представляющий собой комплекс методов и средств, позволяющих получить из определенным образом организованных данных информацию для принятия решений.

Изучение дисциплины должно способствовать формированию у студентов основ алгоритмического мышления и представления о современных подходах к программному решению научных и прикладных задач, позволяет применить знания, полученные при изучении анализа данных, для практического решения задач проектирования и программирования.

Цель учебной дисциплины «Анализ данных» — освоение теоретических основ и методов анализа данных, применяемых при решении прикладных задач. Формирование у студентов системного взгляда на мир, познание сущности анализа данных как методологии исследования сложных систем и процессов, приобретение знаний и навыков, соответствующих квалификации «системный аналитик».

Образовательная цель: обучение студентов приемам, методам и средствам для организации проектов в области информационных систем, анализа требований к информационным системам, построения моделей сложных систем и процессов, анализа данных и машинного обучения.

Развивающая цель: приобретение практических навыков использования анализа данных и творческого мышления при оценке возникающих ситуаций и поиске приемлемых решений проблем организационного характера в различных сферах повседневной деятельности человека.

Задачи учебной дисциплины:

- формирование теоретических знаний и практических навыков для проведения полного цикла анализа данных с применением машинного обучения.
- формирование навыков обработки, обобщения и анализа информации для оценки состояния, и выявления тенденций, закономерностей и конкретных особенностей различных процессов.
- изучение и обсуждение со студентами основ и принципов анализа данных;
- формирование системного взгляда на мир и понятийного аппарата в области анализа данных;
- получение представления о методике и особенностях работы аналитиком данных в проектах в области информационных технологий.

Место учебной дисциплины в системе подготовки специалиста с высшим образованием.

Учебная дисциплина «Анализ данных» относится к модулю «Анализ данных» компонента учреждения высшего образования.

Связи с другими учебными дисциплинами, включая учебные дисциплины компонента учреждения высшего образования, дисциплины специализации и др.

Учебная дисциплина «Анализ данных» взаимосвязана с учебными дисциплинами «Основы машинного обучения», «Методы программирования» и «Теория вероятностей и математическая статистика».

Требования к компетенциям

Освоение учебной дисциплины «Анализ данных» должно обеспечить формирование у студентов следующих компетенций:

универсальные компетенции:

УК-1. Владеть основами исследовательской деятельности, Осуществлять поиск, анализ и синтез информации.

УК-2. Решать стандартные задачи профессиональной деятельности на основе применения информационно-коммуникационных технологий.

специализированные компетенции:

СК-3. Осуществлять полный цикл анализа данных с применением машинного обучения.

В результате изучения дисциплины «Анализ данных» студент должен **знать:**

- методы анализа данных и машинного обучения, необходимые для успешного решения математических задач, связанных с анализом и обработкой результатов наблюдений;
- популярные архитектуры и методологии анализа данных;
- основные понятия и методы теории вероятностей и математической статистики, основные математические методы в контексте анализа данных.
- специализированные языки, программные средства для анализа данных;
- методы организации командной работы;
- методы сбора требований и анализа предметной области;

уметь:

- строить модели сложных систем и описывать их на языке python с применением современных средств анализа данных;
- решать типовые задачи машинного обучения и самостоятельно разрабатывать алгоритмы для решения подобных задач;
- применять основные математические методы и инструментальные средства для решения прикладных задач и исследования объектов профессиональной деятельности;
- использовать математические инструментальные средства для обработки, анализа и систематизации информации по теме исследования.
- решать типовые теоретико-вероятностные и статистические задачи, применять соответствующие методы и знания высшей математики и теории вероятностей для решения математических задач, интерпретировать результаты решения задач,

- применять творческое мышление и методы анализа данных при решении технических задач;
- документировать и представлять отчеты о ходе аналитической работы.

владеть:

- методами и инструментами для проведения полного цикла анализа данных с применением машинного обучения;
- методами и средствами анализа данных при построении моделей сложных систем;
- методами сбора и анализа требований к информационным системам;
- методами построения, анализа и применения статистических моделей для оценки состояния и прогноза развития анализируемых явлений и процессов;
- навыками применения инструментов математического моделирования, методами статистического анализа и прогнозирования случайных процессов.

Структура учебной дисциплины

Дисциплина «Анализ данных» изучается в 6-м и 7-м семестрах. Форма получения высшего образования очная (дневная).

Всего на изучение учебной дисциплины «Анализ данных» отводится 200 часов, в том числе 104 аудиторных часа, из них: лекции — 52 часа, лабораторные занятия — 42 часа, управляемая самостоятельная работа — 10 часов.

– 6 семестр – всего 110 часов, в том числе 68 аудиторных часов, из них: лекции – 34 часа, лабораторные занятия – 28 часов, управляемая самостоятельная работа – 6 часов.

Трудовое количество учебной дисциплины составляет 3 зачетные единицы.

Форма промежуточной аттестации— зачет.

– 7 семестр – всего 90 часов, в том числе 36 аудиторных часов, из них: лекции - 18 часов, лабораторные занятия - 14 часов, управляемая самостоятельная работа - 4 часа.

Трудовое количество учебной дисциплины составляет 3 зачетные единицы.

Форма промежуточной аттестации— зачет.

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

1. Общие положения языка Python. Jupyter Notebook. Построение и синтаксис. Модули и импорт модулей. Функции. Сложность алгоритмов.
2. Структуры и типы данных.
3. Управляющие конструкции и общие методы. Работа с логическими значениями. Сортировка. Циклы. Генераторы последовательностей.
4. 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)

- 1 Линейная алгебра. Действия над векторами и матрицами. Специальные типы матриц. Решение задач линейной алгебры при помощи библиотеки NumPy. Функция gauss, решение СЛАУ.
- 2 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.
3. Случайные величины и распределение. Математическое ожидание. Плотность вероятности случайной величины. Функция распределения. Функция ошибок.

Тема 1.5. Математические основы анализа данных

1 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).

2 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

1 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`.

2 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

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

2 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

1 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.

2 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

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

Раздел 2. Анализ данных и прикладные задачи

Тема 2.1. Введение

Данные и информация. Информационное обогащение данных. Информация для принятия решений. Полезность информации. Извлечение информации из данных. Интерполяция, регрессия, экстраполяция. Информация для прогноза данных.

Тема 2.2. Задачи анализа данных

Анализ данных в стационарных и нестационарных условиях (эволюционирующие системы). Задача анализа функциональных данных. Задача анализа временных данных. Задача классификации данных. Представление данных (нейронные сети, карты Кохонена). Задача распознавания образов.

Тема 2.3. Анализ временных рядов

Введение во временные ряды. Метод Надарая-Ватсона. Метод Хольта. Прогнозирование временных рядов: стационарных, трендовых, сезонных. Основные алгоритмы и их применение. Модели ARIMA, SARIMAX и GARCH. Выявления аномалий во временных рядах. Оценка качества метода. Наведенная случайность.

Тема 2.4. Представление данных

Представление данных. Визуализация большого количества многомерных данных. Самоорганизующиеся карты Кохонена. Реализация алгоритма. Наглядность карт Кохонена.

Тема 2.5. Обработка естественного языка

Введение в обработку естественного языка (Natural Language Processing, NLP). Основные задачи обработки естественного языка: распознавание и анализ текста, извлечение информации, факты, машинный перевод, генерацию текста, анализ настроений и мнения, а также создание чат-ботов и виртуальных ассистентов. Пакеты для обработки текстовой информации на языке Python. Морфологический и синтаксический анализ языка. Векторная модель и методы снижения размерности. Информационный поиск. Классификация текста. Генерация текста. Системы автоматической обработки текстов, поисковые системы, медицинскую диагностику, финансовый анализ.

Тема 2.6. Обработка изображений

- 1 Форматы данных изображений, обработка данных для задач машинного зрения. Распознавание изображений с использованием традиционных методов компьютерного зрения. Гистограмма направленных градиентов. Пример кода для распознавания изображений.
- 2 Выбор свёрточной нейронной сети (CNN), архитектуры глубокого обучения для машинного зрения. Обучение и тестирование собственного детектора объектов OpenCV
- 3 Распознавание изображений с использованием свёрточных нейронных сетей. Обнаружение объектов с использованием глубокого обучения. Генерация изображений.

Тема 2.7. Общие понятия и схема алгоритмов оптимизации

Понятие фитнес-функции, агента. Общая схема алгоритмов. Инициализация популяции. Миграция агентов популяции. Завершение поиска. Алгоритм роевой оптимизации (PSO). Проведение исследования для задачи безусловной оптимизации и тестовых целевых функций Розенброка, Экли, Швепеля.

Тема 2.8. Применение эволюционных алгоритмов

Алгоритм роя частиц:

- Задачи оптимизации функций многих параметров, форм, размеров и топологий;
- Область проектирования;
- Биоинженерия, биомеханика, биохимия.

Муравьиный алгоритм:

- Расчеты компьютерных и телекоммуникационных сетей;
- Задача коммивояжера;
- Задача раскраски графа;
- Задача оптимизации сетевых трафиков.

Алгоритм пчелиной колонии:

- Оптимизация управления;
- Оптимизация классификаторов.

Практические задачи. Эксперименты на реальных данных и модельных данных.

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 УСП	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 индивидуальному проекту с устной защитой.
1.12	Кластерный анализ	4			2		4	Контрольная работа.
2.	Анализ данных и прикладные задачи	18			14		4	
2.1	Введение	2			2			Письменный отчет по лабораторной работе.
2.2	Задачи анализа данных	2			2			Письменный отчет по лабораторной работе.
2.3	Анализ временных рядов	2			2			Письменный отчет по лабораторной работе.
2.4	Представление данных	2					4	Контрольная работа.
2.5	Обработка естественного языка	2			2			Письменный отчет по лабораторной работе.
2.6	Обработка изображений	2			2			Письменный отчет по лабораторной работе.
2.7	Общие понятия и схема алгоритмов оптимизации	2			2			Письменный отчет по лабораторной работе.
2.8	Применение эволюционных алгоритмов	4			2			Письменный отчет по лабораторной работе.

ИНФОРМАЦИОННО-МЕТОДИЧЕСКАЯ ЧАСТЬ

Перечень основной литературы

1. Трушков, А. С. Статистическая обработка информации. Основы теории и компьютерный практикум : учебное пособие / А. С. Трушков. - Санкт-Петербург ; Москва ; Краснодар : Лань, 2020. - 150 с. - URL: <https://e.lanbook.com/book/126947>.
2. Стефанова, И. А. Обработка данных и компьютерное моделирование : учебное пособие / И. А. Стефанова. - Санкт-Петербург ; Москва ; Краснодар : Лань, 2020. - 109 с. - URL: <https://e.lanbook.com/book/126939>.
3. Дюк, В. А. Логический анализ данных : учебное пособие / В. А. Дюк. - Санкт-Петербург ; Москва ; Краснодар : Лань, 2020. - 77 с. - URL: <https://e.lanbook.com/book/126935>.
4. Яцков, Н. Н. Интеллектуальный анализ данных [Электронный ресурс] : электронный учебно-методический комплекс для специальностей: 1-98 01 01 «Компьютерная безопасность» (по направлениям), 1-31 04 02 «Радиофизика», 1-31 04 03 «Физическая электроника» / Н. Н. Яцков ; БГУ, 2023. - URL: <https://elib.bsu.by/handle/123456789/306842>.

Перечень дополнительной литературы

1. Грас Д. Data Science. Наука о данных с нуля. Перевод с английского. – СПб: БХВ-Петербург, 2017. – 336 с.
2. Силен Д., Мейсман А., Али М. Основы Data Science и Big Data. Python и наука о данных. СПб.: Питер, 2017. –336 с.
3. Воронина В.В. и др. Теория и практика машинного обучения. Учебное пособие. / В.В. Воронина, А. В. Михеев, Н. Г. Ярушкина, К. В. Святков. – Ульяновск : УлГТУ, 2017. — 290 с.
4. Плас, Джейк Вандер Python для сложных задач. Наука о данных и машинное обучение. Руководство / Плас Джейк Вандер. - М.: Питер, 2018. - 759 с.
5. Шолле Ф. Глубокое обучение на Python = Deep Learning with Python / Франсуа Шолле ; [пер. с англ. А. Киселева]. - Санкт-Петербург [и др.] : Питер, 2018.
6. Плас Д. В. Python для сложных задач: наука о данных и машинное обучение = Python Data Science Handbook. Essential Tools for Working with Data / Дж. Вандер Плас ; [пер. с англ. И. Пальти]. - Санкт-Петербург [и др.] : Питер, 2020.
7. Бринк Х. Машинное обучение = Real-World Machine Learning / Хенрик Бринк, Джозеф Ричардс, Марк Феверолф ; [пер. с англ. И. Рузмайкиной]. - Санкт-Петербург [и др.] : Питер, 2018.
8. Николенко С. Глубокое обучение. Погружение в мир нейронных сетей / С. Николенко, А. Кадури, Е. Архангельская. - Санкт-Петербург [и др.] : Питер, 2020.

9. Мюллер А. Введение в машинное обучение с помощью Python : руководство для специалистов по работе с данными : [полноцветное издание] / А. Мюллер, С. Гвидо ; [гл. ред. С. Н. Тригуб ; пер. с англ. А. В. Груздева]. - Москва ; Санкт-Петербург ; Киев : Диалектика, 2018.
10. Флах, П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных / П. Флах . – М.: ДМК Пресс, 2015 – 400 с.

Перечень электронных источников

1. Воронцов, К.В. Машинное обучение. Курс лекций. <http://www.machinelearning.ru>
2. Daumé, Н. A course in Machine Learning. http://ciml.info/dl/v0_9/cimlv0_9-all.pdf
3. Учебник по машинному обучению. [Электронный ресурс] – Режим доступа – <https://ml-handbook.ru/>
4. Ковалева М.А., Волошин С.Б. Анализ данных. Учебное пособие – М.: Мир науки, 2019 – Сетевое издание. Режим доступа: <https://izd-mn.com/PDF/32MNNPU19.pdf>

Рекомендуемое учебно-лабораторное оборудование

Для проведения занятий требуются следующее программное обеспечение: операционные системы Microsoft Windows или MacOS, офисные пакеты Microsoft Office или Google Docs, пакет Python-дистрибутив Anaconda, сервис для управления проектом Trello.

Перечень рекомендуемых средств диагностики и методика формирования итоговой отметки

Оценка текущего контроля знаний студента по дисциплине «Анализ данных» формируется в результате регулярной и систематической проверки знаний студентов во время занятий и по итогам их самостоятельной работы. Текущий контроль знаний проходит в форме контрольной работы и опроса во время устной защиты отчета по лабораторным работам. Задания к лабораторным и контрольным работам составляются согласно содержанию учебного материала. При защите лабораторных работ оценивается полнота ответа, аргументация выбранных решений, последовательность и оригинальность изложения материала, оригинальность кода, корректность оформления, самостоятельность выполнения заданий.

Формой промежуточной аттестации по дисциплине «Анализ данных» учебным планом предусмотрен **зачет**

Зачет по дисциплине проходит в форме устной защиты отчетов по лабораторным работам. Если студент успешно, не менее чем на 4 балла защитил все лабораторные работы, то допускается определение результатов

промежуточной аттестации по дисциплине без проведения дополнительного опроса на зачете. При этом явка обучающегося на зачет является обязательной.

При формировании итоговой отметки используется рейтинговая система оценки знаний студента, дающая возможность проследить и оценить динамику процесса достижения целей обучения.

Рейтинговая система предусматривает использование весовых коэффициентов в ходе проведения контрольных мероприятий текущей аттестации.

Примерные весовые коэффициенты, определяющие вклад текущей аттестации в отметку при прохождении промежуточной аттестации:

Формирование отметки за текущую аттестацию:

- отчеты по лабораторным работам — 60%;
- контрольные работы — 40%.

Итоговая отметка по дисциплине рассчитывается на основе отметки текущей аттестации и отметки на зачёте с учетом их весовых коэффициентов. Вес отметки по текущей аттестации составляет 40%, отметки на зачёте – 60%.

Примерный перечень заданий для управляемой самостоятельной работы студентов

Тема 1.7 Обработка данных (2 часа)

Студенту предлагается провести подготовку к дальнейшему анализу реального набора данных. В ходе работы предполагается выполнение следующих заданий:

- Поиск аномалий в данных
- Заполнение пропусков в данных
- Проверка распределения в целевых признаках
- Нормализация и стандартизация данных.

Форма контроля – контрольная работа.

Тема 1.12 Кластерный анализ (4 часа)

Решение группой подготовленной преподавателем задачи на классификацию, регрессию или кластерный анализ в режиме соревнования «кто лучше». Студенты могут объединяться в команды по два человека. Задача предполагает применение навыков из большинства тем дисциплины.

Задания выполняются на основе методических указаний к лабораторной работе.

Форма контроля – контрольная работа.

Тема 2.4 Представление данных (4 часа)

Рассмотреть техническое устройство (динамическую систему) описываемое в каждый момент времени n -мерным вектором его параметров. Получить эмуляцию некоторых режимов работы устройства. Построить карту Кохенена. В каждый момент времени пересчитать местоположение нашей системы на карте. Показать дрейф динамической системы в зеленую и красную зоны. Научиться управлять процессом. Получить наглядную динамику.

Задания выполняются на основе методических указаний к лабораторной работе.

Форма контроля – контрольная работа.

Описание инновационных подходов и методов к преподаванию учебной дисциплины

При организации образовательного процесса используется *эвристический подход*, который предполагает демонстрацию многообразия решений большинства профессиональных задач и жизненных проблем.

При организации образовательного процесса используется *практико-ориентированный подход*, который предполагает освоение содержания через решения практических задач.

При организации образовательного процесса *используются методы и приемы развития критического мышления*, которые представляют собой систему, формирующую навыки работы с информацией в процессе чтения и письма; понимания информации как отправного, а не конечного пункта критического мышления.

Методические рекомендации по организации самостоятельной работы обучающихся

Для организации самостоятельной работы студентов по учебной дисциплине рекомендовано разместить на образовательном портале или сайте кафедры учебно-методические материалы: курсы лекций и лабораторные практикумы, методические указания к лабораторным занятиям, вопросы для подготовки к зачету, перечень рекомендуемой литературы, информационные ресурсы.

Самостоятельная работа студента включает в себя работу с учебной литературой по заданным разделам дисциплины, поиск в Интернете новейшей учебной и научной информации в указанных областях знаний и знакомство с ней, а также выполнение задач, поставленных на занятиях. Во время самостоятельной работы студент выполняет задания, полученные на лабораторных занятиях, а также изучает рекомендуемую литературу. Для совершенствования способностей учиться самостоятельно студентам могут выдаваться темы докладов, с которыми они выступают на занятиях.

Примерный перечень вопросов к зачёту

1. Введение в машинное обучение. Основные понятия МО. Обучение с учителем. Обучение без учителя.
2. Тестовая и обучающая выборки. Понятие эстиматора. Метрики качества моделей.
3. Python библиотеки numpy, matplotlib, sklearn, seaborn. Краткий обзор функционала, применение в контексте МО.
4. Работа с признаками. Нормализация данных и шкалирование. Поиск аномалий в данных.
5. Числовые и категориальные признаки. Корреляция. Описательная статистика.
6. Алгоритмы регрессии. Постановка задачи регрессии. Линейная регрессия.
7. Полиномиальная регрессия. Регрессия с использованием гауссовских базисных функций.
8. Линейная регрессия по произвольному базису. Графическая визуализация задачи регрессии.
9. Проблема переобучения. Регуляризация. Ridge – регуляризация.
10. Проблема переобучения. Регуляризация. Lasso – регуляризация.
11. Алгоритмы классификации. Постановка задачи классификации.
12. Простейшие алгоритмы классификации: дерево решений.
13. Простейшие алгоритмы классификации: метод ближайших соседей.
14. Гиперпараметры моделей. Подбор гиперпараметров. Валидационные кривые и кривые обучения.
15. Линейный классификатор. Линейная классификация. Разделяющая гиперплоскость. Логистическая регрессия.
16. L2-регуляризация логистической функции потерь.
17. Бэггинг. Ансамбли оценщиков. Бутстреп.
18. Случайные леса и подбор гиперпараметров для них. Сверхслучайные леса.
19. Оценка качества модели. Метрики качества модели. Кросс-валидация. Дисперсия и отклонение в задачах МО.
20. Проблема переобучения. Выбор оптимальной сложности модели оценщика.
21. Преобразование признаков. Удаление аномальных данных.
22. Обучение без учителя. Понижение размерности. Метод главных компонент.
23. Обучение без учителя. Понижение размерности. Метод TSNE.
24. Обучение без учителя. Кластеризация. Алгоритм k-средних.

25. Метрики оценки качества кластеризации.
26. Обучение без учителя. Кластеризация. Алгоритм DBSCAN. Вычислительная сложность.
27. Визуализация в задача МО. Библиотека matplotlib. Библиотека seaborn.

ПРОТОКОЛ СОГЛАСОВАНИЯ УЧЕБНОЙ ПРОГРАММЫ УВО

Название учебной дисциплины, с которой требуется согласование	Название кафедры	Предложения об изменениях в содержании учебной программы УВО по учебной дисциплине	Решение, принятое кафедрой, разработавшей учебную программу (с указанием даты и номера протокола) ¹
Вейвлет-анализ	Кафедра дифференциальных уравнений и системного анализа	нет	Вносить изменения не требуется (протокол № 13 от 14.06.2023)

¹ При наличии предложений об изменениях в содержании учебной программы УВО.

**ДОПОЛНЕНИЯ И ИЗМЕНЕНИЯ К УЧЕБНОЙ ПРОГРАММЕ ПО
ИЗУЧАЕМОЙ УЧЕБНОЙ ДИСЦИПЛИНЕ**
на 2024 / 2025 учебный год

№ п/п	Дополнения и изменения	Основание

Учебная программа пересмотрена и одобрена на заседании кафедры
 _____ (протокол № _____ от _____ 200____ г.)
 (название кафедры)

Заведующая кафедрой
 кандидат физ.-мат. наук, доцент _____
 (ученая степень, ученое звание) (подпись)

Л.Л. Голубева
 (И. О.Фамилия)

УТВЕРЖДАЮ
 Декан факультета
 доктор физ.-мат. наук, доцент _____
 (ученая степень, ученое звание) (подпись)

С. М. Босяков
 (И. О.Фамилия)