

**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«ИНГУШСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»**

**ФИЗИКО-МАТЕМАТИЧЕСКИЙ ФАКУЛЬТЕТ
Кафедра «Информационные системы и технологии»**

СОГЛАСОВАНО

Руководитель образовательной программы

_____/М.Х. Мальсагов
«20» мая 2024г.

УТВЕРЖДАЮ

И.о. декана физико-математического
факультета

_____/Б.С.Кульбужев
«23» мая 2024г.

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДИСЦИПЛИНЫ (МОДУЛЯ)

Б1.В.17 Технология анализа и визуализации данных

Направление подготовки

09.03.02 Информационные системы и технологии

Направленность (профиль подготовки)

Перспективные информационные технологии

Квалификация выпускника

Бакалавр

Форма обучения

Очная, очно-заочная

Магас, 2024г

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

Перечень формируемых компетенций

Паспорт фонда оценочных средств

ОПК-2. Способен понимать принципы работы современных информационных технологий и программных средств, в том числе отечественного производства, и использовать их при решении задач профессиональной деятельности

ОПК-4. Способен участвовать в разработке технической документации, связанной с профессиональной деятельностью с использованием стандартов, норм и правил.

ПК-4. Способен проектировать и эксплуатировать ИС и их подсистемы.

№ п/п	Контролируемые разделы (темы) дисциплины	Код Контролируемой компетенции (или ее части)	Наименование оценочного средства
1.	Тема 1: Введение в анализ и визуализацию данных <ul style="list-style-type: none">Основные понятия и цели анализа и визуализации данных.Важность данных в современных информационных системах.Примеры успешного применения анализа и визуализации данных в различных отраслях.	ОПК-2, ОПК-4, ПК-4	Тест
2.	Тема 2: Сбор и предобработка данных Содержание темы: <ul style="list-style-type: none">Методы сбора данных из различных источников.Очистка данных: устранение дубликатов, заполнение пропусков.Нормализация и стандартизация данных.	ОПК-2, ОПК-4, ПК-4	Тест
3.	Тема 3: 3. Методы анализа данных Содержание темы: <ul style="list-style-type: none">Основные методы и алгоритмы анализа данных: статистические методы, машинное обучение.Выявление тенденций и закономерностей.Прогнозирование и моделирование на основе данных.	ОПК-2, ОПК-4, ПК-4	Коллоквиум

4.	Тема 4: . Инструменты и технологии визуализации данных Содержание темы: <ul style="list-style-type: none"> • Обзор инструментов визуализации данных (Tableau, Power BI, D3.js). • Принципы создания эффективных визуализаций. • Типы графиков и диаграмм и их применение. 	ОПК-2, ОПК-4, ПК-4	Коллоквиум
5.	Тема 5: Проектирование и реализация систем анализа и визуализации данных Содержание темы: <ul style="list-style-type: none"> • Принципы проектирования систем анализа и визуализации данных. • Архитектурные особенности таких систем. • Интеграция различных компонентов и источников данных. 	ОПК-2, ОПК-4, ПК-4	Тест, коллоквиум
6.	Тема 6 Практическое применение анализа и визуализации данных Содержание темы: <ul style="list-style-type: none"> • Примеры успешного применения анализа и визуализации данных в бизнесе, науке и других областях. • Обсуждение реальных кейсов и получение инсайтов. 	ОПК-2, ОПК-4, ПК-4	Коллоквиум

Типовой тест промежуточной аттестации

- Какие этапы включает в себя процесс анализа данных?
 - а) Сбор данных
 - б) Предобработка данных
 - в) Анализ данных
 - г) Все вышеперечисленные +
- Какие методы предобработки данных существуют?
 - а) Очистка данных
 - б) Нормализация данных
 - в) Стандартизация данных
 - г) Все вышеперечисленные +
- Какие основные типы графиков используются для визуализации данных?
 - а) Столбчатые
 - б) Линейные
 - в) Круговые
 - г) Все вышеперечисленные +
- Какие основные методы анализа данных существуют?
 - а) Статистические методы
 - б) Машинное обучение

- в) Временные ряды
 - г) Все вышеперечисленные +
5. Что такое машинное обучение?
- а) Процесс обучения компьютерных систем на основе данных и алгоритмов, позволяющий им выполнять задачи без явного программирования
 - б) Процесс создания и обработки больших объемов данных
 - в) Процесс создания моделей для визуализации данных
 - г) Ответ "а" +
6. Какие инструменты используются для визуализации данных?
- а) Tableau
 - б) Power BI
 - в) D3.js
 - г) Все вышеперечисленные +
7. Какой инструмент широко используется для анализа данных на языке Python?
- а) NumPy
 - б) pandas
 - в) Matplotlib
 - г) Все вышеперечисленные +
8. Что такое предобработка данных?
- а) Процесс преобразования данных в удобный для анализа формат
 - б) Процесс создания данных
 - в) Процесс визуализации данных
 - г) Ответ "а" +
9. Какие методы используются для анализа временных рядов?
- а) ARIMA
 - б) LSTM
 - в) Prophet
 - г) Все вышеперечисленные +
10. Какие основные этапы проектирования системы анализа данных?
- а) Архитектурное проектирование
 - б) Интеграция данных
 - в) Разработка прототипа
 - г) Все вышеперечисленные +
11. Какие алгоритмы применяются в статистическом анализе данных?
- а) t-тест
 - б) ANOVA
 - в) Кластерный анализ
 - г) Все вышеперечисленные +
12. Какие методы оценки качества модели машинного обучения существуют?
- а) Точность
 - б) Полнота
 - в) F1-мера
 - г) Все вышеперечисленные +
13. Какие этапы включает в себя процесс визуализации данных?
- а) Подготовка данных

- б) Создание визуализации
- в) Интерпретация результатов
- г) Все вышеперечисленные +

14. Какова роль визуализации данных в анализе данных?

- а) Помогает обнаруживать тенденции и паттерны в данных
- б) Облегчает интерпретацию результатов
- в) Способствует принятию управленческих решений
- г) Все вышеперечисленные +

15. Какие основные типы данных используются в анализе данных?

- а) Количественные
- б) Категориальные
- в) Временные ряды
- г) Все вышеперечисленные +

16. Что такое графовые базы данных?

- а) Тип базы данных, использующий графовые структуры для хранения и обработки данных
- б) Инструмент для визуализации данных
- в) Технология машинного обучения
- г) Ответ "а" +

17. Какие основные этапы включает в себя процесс создания модели машинного обучения?

- а) Подготовка данных
- б) Выбор алгоритма
- в) Обучение модели
- г) Все вышеперечисленные +

18. Какие методы используются для интерпретации результатов анализа данных?

- а) Построение выводов и рекомендаций на основе анализа данных
- б) Визуализация результатов
- в) Обратная связь от заказчика
- г) Все вышеперечисленные +

19. Какие ключевые навыки необходимы для успешного анализа данных?

- а) Знание языков программирования (Python, R)

б) Умение работать с данными и базами данных

- в) Навыки визуализации данных
- г) Все вышеперечисленные +

20. Что такое регрессионный анализ?

- а) Статистический метод анализа данных, используемый для изучения отношений между независимыми и зависимой переменными
- б) Метод предобработки данных
- в) Алгоритм машинного обучения
- г) Ответ "а" +

21. Какие методы используются для прогнозирования временных рядов?

- а) ARIMA
- б) LSTM
- в) Простые скользящие средние
- г) Все вышеперечисленные +

22. Какой инструмент широко используется для создания интерактивных визуализаций данных?

- а) D3.js
- б) Matplotlib
- в) Seaborn
- г) Ответ "а" +

23. Какие основные шаги включает в себя анализ данных?

- а) Загрузка данных
- б) Предобработка данных
- в) Моделирование данных
- г) Все вышеперечисленные +

24. Какие методы используются для кластерного анализа данных?

- а) K-means
- б) DBSCAN
- в) Hierarchical clustering
- г) Все вышеперечисленные +

25. Какие основные этапы включает в себя процесс подготовки данных?

- а) Очистка данных
- б) Нормализация данных
- в) Разделение данных на обучающую и тестовую выборки
- г) Все вышеперечисленные +

26. Какие методы используются для обработки пропущенных значений в данных?

- а) Удаление строк или столбцов с пропущенными значениями
- б) Заполнение пропущенных значений средним или медианным значением
- в) Использование алгоритмов машинного обучения для заполнения пропущенных значений
- г) Все вышеперечисленные +

27. Какой инструмент широко используется для анализа данных на языке R?

- а) ggplot2
- б) matplotlib
- в) seaborn
- г) Ответ "а" +

28. Какие алгоритмы машинного обучения используются для классификации данных?

- а) Логистическая регрессия
- б) Деревья решений
- в) Метод опорных векторов (SVM)
- г) Все вышеперечисленные +

29. Какой метод используется для оценки важности признаков в модели машинного обучения?

- а) Анализ компонент
- б) Перестановочная важность
- в) Отбор признаков с помощью регуляризации
- г) Ответ "б" +

30. Какие существуют методы оценки качества модели машинного обучения?

- а) Точность
- б) Полнота
- в) F1-мера
- г) Все вышеперечисленные +

Типовой вариант задания на контрольную работу

1. Объясните основные этапы анализа данных и опишите их последовательность.

Практические задания (70 баллов)

Ответьте на следующие вопросы и выполните задания:

1. Сбор и предобработка данных

- Загрузите набор данных "iris.csv" и проведите его предварительный анализ.
- Определите количество пропущенных значений в каждом столбце набора данных и заполните их медианным значением.

2. Анализ данных

- Постройте гистограмму распределения длины чашелистика для каждого вида ириса.
- Вычислите среднее, медиану и стандартное отклонение показателей "длина чашелистика" и "ширина чашелистика" для каждого вида ириса.

3. Визуализация данных

а) Создайте scatter plot, отображающий зависимость длины чашелистика от ширины чашелистика для каждого вида ириса.

б) Постройте box plot для каждого показателя ("длина чашелистика", "ширина чашелистика", "длина лепестка", "ширина лепестка") и сравните их распределения для каждого вида ириса.

Экзаменационные вопросы по дисциплине

1. Какие основные этапы включает в себя процесс анализа данных?
2. Опишите основные методы сбора данных из различных источников.
3. Какие существуют методы предобработки данных и почему они важны?
4. Что такое очистка данных и какие основные техники используются для очистки данных?
5. Какова роль нормализации и стандартизации данных в анализе данных?
6. Какие статистические методы применяются в анализе данных?
7. Объясните основные концепции машинного обучения и его роль в анализе данных.
8. Каковы основные этапы создания модели машинного обучения?
9. В чем различие между обучением с учителем и без учителя?
10. Что такое гиперпараметры модели и как они влияют на результаты машинного обучения?
11. Какие существуют инструменты для визуализации данных? Приведите примеры.
12. Каковы основные принципы создания эффективных визуализаций данных?
13. Какие типы графиков и диаграмм используются для визуализации данных и в каких случаях?
14. Опишите процесс создания интерактивной визуализации данных.
15. Каковы основные функции и возможности таких инструментов, как Tableau и Power BI?
16. В чем заключаются основные различия между различными типами графиков (столбчатые, линейные, круговые)?
17. Какие существуют методы оценки качества данных перед их анализом?
18. Как интегрировать различные источники данных для их анализа и визуализации?
19. Какие существуют методы и инструменты для обработки больших данных (Big Data)?
20. Опишите процесс проектирования системы анализа и визуализации данных.
21. Какие архитектурные особенности характерны для систем анализа данных?

22. Каковы основные шаги в разработке прототипа системы анализа данных?
23. Объясните роль и использование библиотек pandas и NumPy в анализе данных.
24. Каковы особенности работы с временными рядами в анализе данных?
25. В чем заключается процесс моделирования данных?
26. Какие существуют методы оценки результатов анализа данных?
27. Как визуализация данных помогает в принятии управленческих решений?
28. Приведите примеры успешного применения анализа и визуализации данных в бизнесе.
29. Какие ключевые навыки необходимы для эффективного анализа и визуализации данных?
30. Какие новые тенденции и технологии появляются в области анализа и визуализации данных?

Типовая практическая работа по дисциплине

Практическая работа: Исследование данных о потреблении электроэнергии

Цель работы: Проанализировать данные о потреблении электроэнергии и визуализировать результаты анализа.

Исходные данные: Набор данных "electricity_consumption.csv", содержащий информацию о потреблении электроэнергии в различных регионах за определенный период времени.

Задачи:

1. Загрузите набор данных и проведите его предварительный анализ.
2. Определите общий объем потребления электроэнергии за весь период.
3. Вычислите средний объем потребления электроэнергии в каждом регионе за весь период.
4. Постройте график изменения потребления электроэнергии по времени для каждого региона.
5. Определите месяцы с наибольшим и наименьшим объемом потребления электроэнергии.
6. Постройте круговую диаграмму, отражающую долю каждого региона в общем объеме потребления электроэнергии.

Ход работы:

1. **Предварительный анализ данных**
 - Загрузите набор данных "electricity_consumption.csv".
 - Изучите структуру данных: количество строк и столбцов, типы данных, наличие пропущенных значений.
 - Выведите первые несколько строк набора данных для ознакомления.
2. **Общий объем потребления электроэнергии**
 - Просуммируйте значения потребления электроэнергии по всем регионам за весь период.
 - Выведите полученный результат.

3. **Средний объем потребления электроэнергии**
 - Вычислите средний объем потребления электроэнергии для каждого региона.
 - Выведите полученные значения с указанием соответствующих регионов.
4. **График изменения потребления электроэнергии по времени**
 - Постройте график изменения потребления электроэнергии для каждого региона по времени.
 - Добавьте подписи к осям и легенду для каждого региона.
5. **Месяцы с наибольшим и наименьшим объемом потребления электроэнергии**
 - Определите месяцы с наибольшим и наименьшим объемом потребления электроэнергии.
 - Выведите полученные результаты.
6. **Круговая диаграмма доли каждого региона в общем объеме потребления электроэнергии**
 - Постройте круговую диаграмму, отражающую долю каждого региона в общем объеме потребления электроэнергии.
 - Добавьте подписи к секторам диаграммы и процентное значение для каждого региона.

Ожидаемые результаты:

- Загруженный и предобработанный набор данных.
- Общий объем потребления электроэнергии.
- Средний объем потребления электроэнергии для каждого региона.
- График изменения потребления электроэнергии по времени для каждого региона.
- Месяцы с наибольшим и наименьшим объемом потребления электроэнергии.
- Круговая диаграмма доли каждого региона в общем объеме потребления электроэнергии.