

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
СТАВРОПОЛЬСКИЙ ГОСУДАРСТВЕННЫЙ АГРАРНЫЙ УНИВЕРСИТЕТ**

УТВЕРЖДАЮ

Директор/Декан
факультета цифровых технологий
Шлаев Дмитрий Валерьевич

«__» _____ 20__ г.

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ (ОЦЕНОЧНЫХ МАТЕРИАЛОВ)

Б1.В.04 Анализ и визуализация данных

09.03.02 Информационные системы и технологии

Инженерия информационных систем

бакалавр

очная

1. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы

Процесс изучения дисциплины направлен на формирование следующих компетенций ОП ВО и овладение следующими результатами обучения по дисциплине:

Код и наименование компетенции	Код и наименование индикатора достижения	Перечень планируемых результатов обучения по дисциплине
--------------------------------	--	---

2. Перечень оценочных средств по дисциплине

№	Наименование раздела/темы	Семестр	Код индикаторов достижения компетенций	Оценочное средство проверки результатов достижения индикаторов компетенций
1.	1 раздел. Раздел 1. Основы анализа данных			
1.1.	1. Введение в науку о данных	7		
1.2.	2. Инструменты обработки данных	7		
1.3.	3. Визуализация данных	7		Задачи
2.	2 раздел. Раздел 2. Трансформация данных			
2.1.	4. Анализ и преобразование данных	7		
2.2.	5. Работа с временными рядами	7		
2.3.	6 Системы управления базами данных	7		
2.4.	7 Проектирование структурированных данных	7		
3.	3 раздел. Раздел 3. Сбор и обработка больших данных			
3.1.	8. Запросы на языке SQL	7		
3.2.	9. Объекты базы данных	7		
3.3.	10. NoSQL хранилища	7		
4.	4 раздел. Экзамен			
4.1.	Тема 11	7		
	Промежуточная аттестация			За

3. Оценочные средства (оценочные материалы)

Примерный перечень оценочных средств для текущего контроля успеваемости и промежуточной аттестации

№ п/п	Наименование оценочного средства	Краткая характеристика оценочного средства	Представление оценочного средства в фонде (Оценочные материалы)
-------	----------------------------------	--	---

Текущий контроль			
Для оценки знаний			
1	Задачи	Задачи репродуктивного уровня, позволяющие оценивать и диагностировать знание фактического материала (базовые понятия, алгоритмы, факты) и правильное использование специальных терминов и понятий, узнавание объектов изучения в рамках определенного раздела дисциплины;	Комплект задач минимального уровня
Для оценки умений			
Для оценки навыков			
Промежуточная аттестация			
2	Зачет	Средство контроля усвоения учебного материала практических и семинарских занятий, успешного прохождения практик и выполнения в процессе этих практик всех учебных поручений в соответствии с утвержденной программой с выставлением оценки в виде «зачтено», «незачтено».	Перечень вопросов к зачету

4. Примерный фонд оценочных средств для проведения текущего контроля и промежуточной аттестации обучающихся по дисциплине (модулю) "Анализ и визуализация данных"

Примерные оценочные материалы для текущего контроля успеваемости

1. Что такое наука о данных и как она отличается от статистики?
2. Какие ключевые этапы включает процесс анализа данных?
3. Каковы основные задачи науки о данных (кластеризация, классификация, регрессия и др.)?
4. В чем заключается разница между описательной и предсказательной аналитикой?
5. Каковы преимущества и вызовы работы с большими данными?
6. Что такое машинное обучение и как оно связано с анализом данных?
7. Какие роли и навыки требуются специалисту по анализу данных?
8. Какие категории инструментов используются для анализа данных (настольные программы, языки программирования и платформы)?
9. Чем различаются инструменты для обработки данных, такие как Python и R?
10. Какие библиотеки Python наиболее популярны для анализа данных и зачем они используются?
11. Какова роль SQL в обработке данных?
12. В чем преимущества облачных решений для анализа данных?
13. Какие инструменты наиболее подходят для визуализации данных и почему?
14. Как сравнить популярные платформы для работы с большими данными (Hadoop, Spark)?
15. Какие типы визуализаций данных существуют и когда их следует использовать?
16. Что такое диаграмма рассеяния и в каких ситуациях она полезна?
17. В чем разница между столбчатой диаграммой и гистограммой?

18. Как строить временные ряды и какие особенности их визуализации нужно учитывать?
19. Что такое интерактивная визуализация данных и каковы ее преимущества?
20. Какие инструменты используются для создания сводных таблиц и диаграмм?
21. В чем ключевые отличия между Tableau, Power BI и Google Data Studio?
22. Что такое очистка данных и почему это важный этап анализа?
23. Как определить и устранить пропуски в данных?
24. Какие методы нормализации данных существуют и когда их нужно применять?
25. Что такое агрегация данных и как она используется в аналитике?
26. Как категоризировать данные и какие методы преобразования категориальных данных существуют?
27. Что такое масштабирование данных и зачем оно нужно в анализе данных?
28. Как выполняется кодирование категориальных данных (например, one-hot encoding)?
29. Что такое временные ряды и где они применяются в анализе данных?
30. Каковы основные компоненты временных рядов (тренд, сезонность, шум)?
31. Что такое скользящее среднее и как его использовать для сглаживания временных рядов?
32. Какие методы прогнозирования временных рядов существуют?
33. Как проводится анализ сезонности во временных рядах?
34. В чем заключается разложение временного ряда и для чего оно нужно?
35. Каковы основные методы оценки модели временных рядов?
36. Что такое реляционная база данных и какие принципы лежат в ее основе?
37. Чем реляционные базы данных отличаются от нереляционных?
38. Каковы основные компоненты любой СУБД?
39. Какие типы данных поддерживаются в реляционных базах данных?
40. Что такое первичные и внешние ключи и зачем они нужны?
41. Какие популярные СУБД существуют и каковы их ключевые особенности (MySQL, PostgreSQL, Oracle)?
42. В чем разница между OLTP и OLAP системами?
43. Какие этапы включает проектирование базы данных?
44. Что такое нормализация данных и зачем она нужна?
45. Что такое первая, вторая и третья нормальные формы?
46. Как разрабатывать ER-диаграмму для базы данных?
47. Что такое денормализация данных и в каких случаях ее применяют?
48. Как спроектировать схему базы данных для интернет-магазина?
49. Какие принципы нужно учитывать при проектировании базы данных для больших объемов данных?
50. Как выполнить базовый SELECT-запрос для выборки данных?
51. Что такое оператор WHERE и как использовать его для фильтрации данных?
52. Как выполнять сортировку данных с помощью оператора ORDER BY?
53. Как применять оператор JOIN для объединения данных из нескольких таблиц?
54. Какие существуют агрегатные функции в SQL и как они работают?
55. В чем заключается использование оператора GROUP BY и HAVING?
56. Как создавать подзапросы в SQL и в чем их преимущества?
57. Что такое таблицы в базе данных и как они структурированы?
58. Какова роль индексов в базах данных и как они помогают ускорить запросы?
59. Что такое представления (VIEW) в базе данных и зачем их использовать?
60. Как создавать и использовать хранимые процедуры в базе данных?
61. Что такое триггеры в базе данных и как они работают?
62. Какие типы ограничений могут быть наложены на данные в таблицах?
63. В чем различие между постоянными и временными таблицами в базе данных?
64. Что такое NoSQL базы данных и чем они отличаются от реляционных баз данных?
65. Какие типы NoSQL баз данных существуют (документные, key-value, графовые, столбцовые)?
66. В чем преимущества и недостатки NoSQL баз данных по сравнению с реляционными?
67. Как работает база данных MongoDB и для каких задач она используется?
68. Что такое Redis и какие случаи его использования наиболее популярны?

69. В чем особенности работы с данными в базах данных типа Cassandra?
70. Какой подход к горизонтальному масштабированию используют NoSQL базы данных?

**Примерные оценочные материалы
для проведения промежуточной аттестации (зачет, экзамен)
по итогам освоения дисциплины (модуля)**

1. Что такое наука о данных?
а) Метод анализа текстов
б) Область знаний, основанная на обработке и анализе данных
в) Исключительно машинное обучение
г) Управление базами данных
2. Какой из этапов анализа данных включает очистку и подготовку данных?
а) Сбор данных
б) Моделирование
в) Обработка данных
г) Визуализация данных
3. Что из перечисленного не является задачей науки о данных?
а) Классификация
б) Регрессия
в) Агрегация данных
г) Линейная алгебра
4. Какой метод чаще всего используется для поиска скрытых закономерностей в данных?
а) Кластеризация
б) Ранжирование
в) Анализ выбросов
г) Анкеты и опросы
5. Что отличает анализ данных от традиционной статистики?
а) Только работа с большими данными
б) Широкий спектр методов и подходов
в) Использование исключительно числовых данных
г) Отсутствие необходимости в визуализации
6. Какая задача лучше всего подходит для классификации?
а) Определение будущих продаж
б) Определение пола по фотографии
в) Поиск ключевых слов в тексте
г) Вычисление средней заработной платы
7. Что является конечной целью анализа данных?
а) Построение графиков
б) Обнаружение инсайтов для принятия решений
в) Хранение данных
г) Создание таблиц
8. Какую задачу решает метод регрессии?
а) Определение категорий данных
б) Прогнозирование количественных значений
в) Обнаружение аномалий
г) Сегментация данных
9. Что подразумевается под термином "большие данные"?
а) Данные, которые невозможно хранить в реляционных базах данных
б) Большие файлы данных
в) Данные объемом свыше 1 ГБ
г) Данные, которые требуют специальных методов обработки
10. Какие основные этапы анализа данных?
а) Визуализация и моделирование
б) Сбор, обработка, анализ, визуализация
в) Сбор и анализ
г) Моделирование и прогнозирование
11. Какой язык программирования наиболее популярен для анализа данных?

- a) Python
- b) C++
- c) Java
- d) PHP

12. Какая библиотека Python предназначена для работы с массивами данных?

- a) Matplotlib
- b) NumPy
- c) Pandas
- d) Seaborn

13. Что такое Jupyter Notebook?

- a) Язык программирования
- b) Среда для написания и выполнения кода
- c) Веб-сервер
- d) База данных

14. Для чего используется библиотека Pandas?

- a) Для работы с многомерными массивами
- b) Для работы с таблицами и временными рядами
- c) Для работы с нейронными сетями
- d) Для визуализации данных

15. Какую роль выполняет инструмент Apache Spark?

- a) Управление базами данных
- b) Обработка больших объемов данных
- c) Машинное обучение
- d) Визуализация данных

16. Какой формат файлов чаще всего используется для хранения данных?

- a) .xls
- b) .docx
- c) .json
- d) .png

17. Что позволяет делать язык R?

- a) Хранить данные
- b) Строить интерактивные веб-сайты
- c) Проводить статистический анализ
- d) Создавать базы данных

18. Какая библиотека Python используется для визуализации данных?

- a) Pandas
- b) Matplotlib
- c) NumPy
- d) TensorFlow

19. Что из перечисленного является инструментом для работы с большими данными?

- a) MongoDB
- b) Apache Hadoop
- c) SQLite
- d) Redis

10. Какова цель визуализации данных?

- A) Сжатие данных
- B) Представление данных в наглядной форме
- C) Хранение данных
- D) Анализ алгоритмов

11. Какой тип диаграммы используется для отображения частей целого?

- A) Столбчатая диаграмма
- B) Линейная диаграмма
- C) Круговая диаграмма
- D) Точечная диаграмма

12. Что из перечисленного НЕ является инструментом визуализации данных?

- A) Tableau

- B) Power BI
- C) Jupyter Notebook
- D) PostgreSQL

13. Какая библиотека Python чаще всего используется для построения графиков?

- A) NumPy
- B) pandas
- C) Matplotlib
- D) SQLAlchemy

14. Что из следующего лучше всего подходит для представления временных рядов?

- A) Линейный график
- B) Тепловая карта
- C) Гистограмма
- D) Круговая диаграмма

15. Какой тип диаграммы лучше всего использовать для отображения распределения данных?

- A) Точечная диаграмма
- B) Гистограмма
- C) Линейный график
- D) Столбчатая диаграмма

16. Как называется процесс уменьшения количества точек на диаграмме без потери сути?

- A) Нормализация
- B) Агрегация
- C) Интеграция
- D) Уменьшение размерности

17. Что такое интерактивная визуализация данных?

- A) Визуализация с возможностью взаимодействия пользователя с данными
- B) Статическая визуализация
- C) Программирование визуальных эффектов
- D) Использование 3D графики

18. Какая диаграмма лучше всего подходит для анализа зависимости между двумя переменными?

- A) Точечная диаграмма
- B) Круговая диаграмма
- C) Линейный график
- D) Гистограмма

19. Для чего используется тепловая карта?

- A) Для отображения корреляций между данными
- B) Для представления временных данных
- C) Для анализа сезонности
- D) Для отображения категорий

20. Что означает термин "очистка данных"?

- A) Удаление данных
- B) Устранение ошибок и пропусков в данных
- C) Создание новых переменных
- D) Преобразование форматов данных

21. Какой метод используется для обработки пропущенных данных?

- A) Удаление строк с пропусками
- B) Интерполяция данных
- C) Заполнение средними значениями
- D) Все перечисленные

22. • Что такое нормализация данных?

- A) Приведение данных к диапазону от 0 до 1
- B) Уменьшение объема данных
- C) Увеличение точности данных
- D) Удаление выбросов

23. • Какой метод применяется для уменьшения размерности данных?

A) Метод главных компонент (PCA)

B) Кластеризация

C) Регрессия

D) Нормализация

24. • Что такое категоризация данных?

A) Преобразование числовых данных в категориальные

B) Сортировка данных

C) Упорядочивание данных

D) Объединение данных

25. • Что такое выбросы в данных?

A) Нормальные значения в данных

B) Исключительные или аномальные значения

C) Упущенные данные

D) Минимальные значения

26. • Для чего используется стандартизация данных?

A) Для приведения данных к одному масштабу

B) Для удаления лишних данных

C) Для объединения данных

D) Для создания моделей

27. • Какой метод используется для выявления скрытых закономерностей в данных?

A) Регрессия

B) Кластеризация

C) Тестирование гипотез

D) Вычисление средних

28. • Какой формат данных чаще всего используется для работы с большими объемами

данных?

A) JSON

B) CSV

C) Excel

D) TXT

29. • Что такое фича-инжиниринг?

A) Процесс создания новых признаков на основе существующих данных

B) Процесс удаления лишних данных

C) Процесс агрегирования данных

30. Для чего используется индекс в СУБД?

A) Для улучшения производительности запросов

B) Для хранения данных

C) Для сжатия данных

D) Для объединения таблиц

Темы письменных работ (эссе, рефераты, курсовые работы и др.)

"Разработка информационной системы для обработки, трансформации и визуализации данных (название набора данных)"

Наборы данных (варианты курсовой работы)

1. Малые и средние предприятия в сфере лёгкой промышленности за 2016–2023 гг.
2. Загрязнение поверхностных вод в России: ежемесячные данные о высоком и экстремально высоком загрязнении водных объектов за 2008—2021 гг.
3. Показатели результативности российских вузов: объединенные данные за 2013-2017 годы
4. Загрязнение атмосферы в России: объем выбросов в воздух в разрезе муниципальных образований за 2019-2020 гг.
5. Населенные пункты России: численность населения и географические координаты
6. Федеральный реестр медицинских организаций
7. Статистические данные о занятости и безработице среди населения по возрастным группам
8. Демографическая ситуация по субъектам РФ
9. История изменений муниципалитетов в России: наименования, коды ОКТМО и входящие в состав населенные пункты
10. Пожары в России: данные о местах и типах природных пожаров за 2012–2021 гг.
11. Показатели валового регионального продукта в России: объединенные данные за 1998—2019 гг.
12. «Зеленые» инвестиции: затраты на охрану окружающей среды в регионах РФ по классификации СЕРА
13. Площадки для выгула (дрессировки) собак
14. Спортивные площадки
15. Залы тренажерные