

**Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Сахалинский государственный университет»
Кафедра информатики**

УТВЕРЖДЕН
на заседании кафедры
«22» мая 2025 г., протокол № 9
Исполняющий обязанности
заведующего кафедрой



Осипов Г.С.

**ФОНД
ОЦЕНОЧНЫХ СРЕДСТВ
ПО ДИСЦИПЛИНЕ (МОДУЛЮ)**

Б1.В.ДВ.03.02 Введение в анализ данных

Направление подготовки
09.03.03 Прикладная информатика
профиль

Автоматизированные системы обработки информации и управления

**Уровень высшего образования
БАКАЛАВРИАТ**

Южно-Сахалинск, 2025 г.

1. Формируемые компетенции и индикаторы их достижения по дисциплине

Код компетенции	Содержание компетенции	Код и наименование индикатора достижения компетенции
ПКС-2	Способен проводить формализацию предметной области с целью создания информационной системы	ПКС-2.1 - Знает требования к компьютерному программному обеспечению; виды технической спецификации на программные компоненты и их взаимодействие; методы проектирования компьютерного программного обеспечения ПКС-2.2 – Умеет применять требования к компьютерному программному обеспечению; разрабатывать технические спецификации на программные компоненты и их взаимодействие; применять методы проектирования компьютерного программного обеспечения; ПКС-2.3 – Владеет методами разработки требований к компьютерному программному обеспечению, технических спецификаций на программные компоненты, методами проектирования компьютерного программного обеспечения.

2. Паспорт фонда оценочных средств по дисциплине (модуля)

№ п/п	Контролируемые разделы (темы) дисциплины	Код контролируемой компетенции (или ее части)	Наименование оценочного средства
1.	Математический аппарат (refresher). Введение в модуль NumPy. Основы работы с Pandas. Разведывательный анализ данных.	ПКС-2	Задания к лабораторным работам, контрольные вопросы, вопросы к зачету
2.	Визуализация данных. Представление результатов исследования.	ПКС-2	Задания к лабораторным работам, контрольные вопросы, вопросы к зачету
3.	Работа с текстовыми данными. Сбор данных из открытых источников. Предварительная обработка текстовых данных. Текстовый анализ.	ПКС-2	Задания к лабораторным работам, контрольные вопросы, вопросы к зачету
4.	Анализ сетей	ПКС-2	Задания к лабораторным работам, контрольные вопросы, вопросы к зачету
5.	Введение в машинное обучение. Модуль sklearn. Задачи классификации и линейные модели.	ПКС-2	Задания к лабораторным работам, контрольные вопросы, вопросы к зачету
6.	Деревья решений. Случайный лес. Ансамбли моделей.	ПКС-2	Задания к лабораторным работам, контрольные вопросы, вопросы к зачету
7.	зачет	ПКС-2	Задания к лабораторным работам, контрольные вопросы, вопросы к зачету

1. Оценочные средства

Тестирование является одним из средств формального контроля качества обучения – методе, основанном на стандартизированных заданиях, которые позволяют оценить знания, умения и навыки испытуемого

Требования к составлению тестового задания

Тесты должны быть построены по методике, обеспечивающей выполнение требований соответствующего федерального государственного образовательного стандарта.

В тестовых заданиях используются четыре типа вопросов:

– **закрытая** форма – является наиболее распространенной и предлагает несколько альтернативных ответов на поставленный вопрос. Тестовое задание, содержащее вопрос в закрытой форме, включает в себя один или несколько правильных ответов.

Закрытая форма вопросов используется также в тестах-задачах с выборочными ответами. В тестовом задании в этом случае сформулированы условие задачи и все необходимые исходные данные, а в ответах представлены несколько вариантов результата решения в числовом или буквенном виде.

Обучающийся должен решить задачу и показать, какой из представленных ответов он получил.

– **открытая** форма – вопрос представляет собой утверждение, которое необходимо дополнить. Данная форма может быть представлена в тестовом задании, например, в виде словесного текста, формулы (уравнения), графика, в которых пропущены существенные составляющие – части слова или буквы, условные обозначения, линии или изображения элементов схемы или графика.

Обучающийся должен по памяти вставить соответствующие элементы в указанные места («пропуски»).

– **установление соответствия** – в данном случае обучающемуся предлагают два списка, между элементами которых следует установить соответствие;

– **установление последовательности** – предполагает необходимость установить правильную последовательность предлагаемого списка слов или фраз.

Примерный тест

Задания для проверки сформированности компетенции ПКС-2 Способен проводить формализацию предметной области с целью создания информационной системы

Часть 1: Математический аппарат и NumPy

1. Выберите правильный ответ:

Укажите метод NumPy, который используется для создания единичной матрицы

1. np.ones()
2. np.identity()
3. np.eye()
4. np.matrix()

Ответ: 3

2. Запишите ответ:

Запишите результат выполнения фрагмента кода: _____

```
import numpy as np
arr = np.array([[1, 2], [3, 4], [5, 6]])
print(arr.shape, arr.ndim)
```

Ответ: (3, 2), 2

3. Запишите ответ:

Напишите код для вычисления скалярного произведения двух векторов в NumPy.

Ваш ответ:

Ответ: Код может выглядеть так:

```
np.linalg.eig() # для собственных значений и векторов
np.linalg.eigvals() # только собственные значения
```

4. Запишите ответ:

Запишите функцию, которая используется для вычисления собственных значений матрицы в NumPy: _____

Ответ: `numpy.linalg.eigvals()`

Часть 2: Pandas и разведывательный анализ

5. Выберите правильный ответ:

Укажите функцию удаления всех строк с пропущенными значениями в DataFrame

1. `df.dropna()`
2. `df.fillna()`
3. `df.remove()`
4. `df.clean()`

Ответ: 1

6. Запишите ответ:

Напишите код для группировки данных по столбцу 'category' и вычисления среднего значения 'price': _____

Ответ: `df.groupby('category')['price'].mean()`

7. Запишите ответ:

Напишите код для:

- Чтения CSV файла 'data.csv'
- Просмотра первых 5 строк
- Проверки информации о типах данных
- Проверки пропущенных значений

Ваш код:

Ответ: Код может быть таким:

```
import pandas as pd
df = pd.read_csv('data.csv')
print(df.head())
print(df.info())
print(df.isnull().sum())
```

8. Сопоставьте инструмент и его назначение:

инструмент	назначение
1. Boxplot	А. Визуализирует распределение числовых данных, группируя их в интервалы (столбцы), высота которых показывает частоту (количество) значений в каждом интервале
2. Heatmap	В. Строит матрицу диаграмм рассеяния для всех пар числовых признаков в датасете, позволяя быстро исследовать взаимосвязи и распределения данных; на диагонали матрицы отображаются гистограммы (или оценки плотности) для каждого признака, показывая их индивидуальное распределение
3. Histogram	С. Использует цвета для отображения интенсивности или плотности данных в разных областях, где теплые цвета (красный, желтый) указывают на высокую активность или значения, а холодные (синий, зеленый) — на низкие
4. Scatter plot	Д. Компактно отображает распределение данных, показывая их центр, разброс и наличие выбросов
5. Pairplot	Е. визуализирует взаимосвязь между двумя числовыми переменными, отображая каждую пару значений в виде точки на декартовой плоскости (X и Y), что помогает выявлять корреляцию (силу и направление связи), кластеризацию данных и наличие выбросов (аномальных точек)

Ответ: 1-D, 2-C, 3-A, 4-E, 5-B

Часть 4: Работа с текстовыми данными

9. Запишите ответ:

Приведите пример токенизации на Python:

Ваш код:

Ответ Результат может быть таким:

```
from nltk.tokenize import word_tokenize
text = "Hello world! This is NLP."
tokens = word_tokenize(text)
# ['Hello', 'world', '!', 'This', 'is', 'NLP', '.']
```

Часть 5: Анализ сетей (4 вопроса)

10. Выберите правильный ответ:

Укажите элементарную единицу любой сети

1. Связь (ребро)
2. Сообщество
3. Узел (вершина)
4. Граф

Ответ: 3

11. Выберите правильный ответ:

Связь (ребро) в сети, показывающая взаимную подписку двух пользователей, является:

1. Направленной и взвешенной
2. Ненаправленной и взвешенной
3. Направленной и невзвешенной
4. Ненаправленной и невзвешенной

Ответ: 4

12. Выберите правильный ответ:

Укажите меру центральности, которая оценивает узел по важности его соседей

1. Степень узла (Degree)
2. Центральность по близости (Closeness)
3. Центральность по посредничеству (Betweenness)
4. Центральность по собственному значению (Eigenvector)

Ответ: 4

13. Выберите правильный ответ:

Математическое представление сети — это...

1. Матрица
2. Граф
3. Диаграмма
4. Кластер

Ответ: 2

14. Выберите ВСЕ верные утверждения о «безмасштабных сетях» (Scale-Free):

1. Распределение степеней узлов следует степенному закону.
2. В них не бывает узлов с очень большим количеством связей (хабов).
3. Они очень уязвимы к случайному удалению узлов.
4. Они очень уязвимы к целенаправленной атаке на наиболее связанные узлы.
5. Примером может служить сеть авиарейсов между крупными хабами.

Ответ: 1, 4 (2 — неверно, хабы как раз есть; 3 — неверно, они устойчивы к случайным отказам; 5 — спорно, сеть авиарейсов часто имеет иерархическую, но не чисто безмасштабную структуру)

15. Установите соответствие между мерой центральности и её определением:

Мера	Определение
1. Degree	А. Показывает, насколько узел близок в среднем ко всем остальным узлам сети.
2. Betweenness	Б. Определяет важность узла через важность его соседей.
3. Closeness	В. Просто подсчитывает количество связей у узла.
4. Eigenvector	Г. Измеряет, какая доля кратчайших путей в сети проходит через данный узел.

Ответ: 1-В, 2-Г, 3-А, 4-Б

16. Выберите все правильные ответы:

Укажите что из перечисленного является метрикой ГЛОБАЛЬНОГО уровня (описывает всю сеть), а что — ЛОКАЛЬНОГО (описывает узел):

- А. Плотность (Density)
- В. Степень узла (Degree)
- С. Коэффициент кластеризации (Clustering Coefficient)

D. Средняя длина пути (Average Path Length)

Глобальные: ____ ; Локальные: ____.

Ответ: Глобальные: A, D ; Локальные: B, C

Часть 6: Машинное обучение

17. Выберите правильный ответ:

Определите, какая из перечисленных задач НЕ является типичной задачей обучения с учителем (supervised learning):

1. Классификация изображений (определение, изображение некоторого объекта)
2. Кластеризация клиентов по их purchasing behavior
3. Прогнозирование стоимости дома на основе его характеристик
4. Определение спама в электронных письмах

Ответ: 2. Кластеризация клиентов... (Это задача обучения без учителя, так как нет заранее известных меток кластеров)

18. Установите соответствие между типом алгоритма машинного обучения и его кратким описанием:

Алгоритм	Описание
1. Линейная регрессия	А. Алгоритм, строящий дерево решений на основе энтропии/примеси для классификации
2. Метод опорных векторов (SVM)	Б. Ансамблевый метод, где каждое следующее дерево учится на ошибках предыдущих
3. Дерево решений	В. Алгоритм, находящий гиперплоскость, максимально разделяющую классы
4. Градиентный бустинг	Г. Алгоритм для предсказания непрерывной величины, находящий линейную зависимость

Ответ:

- 1 – Г (Линейная регрессия предсказывает число)
- 2 – В (SVM ищет разделяющую гиперплоскость)
- 3 – А (Дерево решений использует критерии вроде энтропии)
- 4 – Б (Градиентный бустинг — ансамбль, исправляющий ошибки)

Критерии оценки знаний при проведении тестирования

Отметка «**отлично**» выставляется при условии правильного ответа не менее чем 85% тестовых заданий;

Отметка «**хорошо**» выставляется при условии правильного ответа не менее чем 70 % тестовых заданий;

Отметка «**удовлетворительно**» выставляется при условии правильного ответа не менее 50 %;

Отметка «**неудовлетворительно**» выставляется при условии правильного ответа менее чем на 50 % тестовых заданий.

Результаты текущего контроля используются при проведении промежуточной аттестации