


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

Кафедра информатики

УТВЕРЖДЕН  
на заседании кафедры  
«20» сентября 2024 г., протокол № 1  
Исполняющий обязанности  
заведующего кафедрой

 Осипов Г.С.

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

*Б1.В.ДВ.01.02 Технологии организации и управления инфраструктурой  
больших данных*

Направление подготовки  
09.04.03 Прикладная информатика

профиль  
*Искусственный интеллект и анализ данных*

Уровень высшего образования  
МАГИСТРАТУРА

Южно-Сахалинск  
2024 г.

## 1. Формируемые компетенции и индикаторы их достижения по дисциплине

Код компетенции	Содержание компетенции	Код и наименование индикатора достижения компетенции
ПКС-1.	ПКС-1 - Способен решать задачи профессиональной деятельности с учетом текущего состояния и тенденций развития информационных технологий	ПКС-1.1 - Знает виды моделей бизнес-процессов, требования к информационной системе, виды архитектур ИС; технологии программирования, тестирования и внедрения ИС; ПКС-1.2 - Умеет разрабатывать модели бизнес-процессов, требования к информационной системе, архитектуру ИС, применять технологии программирования, тестирования и внедрения ИС; ПКС-1.3 – Владеет методами разработки модели бизнес-процессов, требований к информационной системе, архитектур ИС, технологиями программирования, тестирования и внедрения ИС
ПКС-2	ПКС-2. Способен управлять проектированием, процессом, разработки компьютерного программного обеспечения, конфигурациями и выпусками программного продукта	ПКС-2.1 - Знает методы управления проектированием, процессом, разработки компьютерного программного обеспечения, конфигурациями и выпусками программного продукта ПКС-2.2 – Умеет применять методы управления проектированием, процессом, разработки компьютерного программного обеспечения, конфигурациями и выпусками программного продукта; ПКС-2.3 – Владеет методами управления проектированием, процессом, разработки компьютерного программного обеспечения, конфигурациями и выпусками программного продукта.
ПКС-3	ПКС-3 - Способен осуществлять организацию взаимодействия с заказчиком, планирования проекта ИС; руководить разработкой программного кода, верификацией и тестированием ИС	ПКС-3.1 - Знает методы организации взаимодействия с заказчиком, планирования проекта, разработки, верификации и тестирования ИС; ПКС-3.2 - Умеет применять методы организации взаимодействия с заказчиком, планирования проекта, разработки, верификации и тестирования ИС; ПКС-3.3 - Владеет методами организации взаимодействия с заказчиком, планирования проекта, разработки, верификации и тестирования ИС.

## 2. Паспорт фонда оценочных средств по дисциплине (модулю)

№ п/п	Контролируемые разделы (темы) дисциплины	Код контролируемой компетенции (или ее части)	Наименование оценочного средства
1.	Тема 1. Анализ систем обработки больших данных	ПКС-1 ПКС-2 ПКС-3	задания к лабораторным работам
2.	Тема 2. Современные методы визуализации информации Big Data	ПКС-2 ПКС-3	задания к лабораторным работам
3.	Тема 3 Детектирование выбросов и аномалий. Очистка данных и технологии регуляризации	ПКС-1 ПКС-2	задания к лабораторным работам
4.	Тема 4 Технологии кластеризации	ПКС-1	задания к лабораторным работам

	и классификации. Нейронные сети. Генетические алгоритмы	ПКС-2 ПКС-3	
5.	Тема 5 Выделение особенностей (Feature detection); нормализация данных. Нечеткие множества. Байесовы сети	ПКС-1 ПКС-3	задания к лабораторным работам
	зачет	ПКС-1 ПКС-2 ПКС-3	вопросы к зачету.

### Лабораторное занятие №1 (6 ч.)

Тема **Анализ систем обработки больших данных.**

Вопросы для обсуждения:

1. Достоверность собранных данных
2. Обучение с учителем; регрессия и классификация; обучение без учителя; кластеризация, снижение размерности; рекомендательные системы;
3. обработка текстов; обработка изображений; обучение с подкреплением.
4. Примеры задач.
5. Программный инструментарий Data Mining и Machine Learning
6. Примеры реализации.

#### Примерные варианты индивидуальных заданий

**Задача.** Классификация текстовых сообщений

- 1) Загрузите исходные данные
- 2) Разбейте загруженные данные на обучающее (train) и тестовое подмножества (test)
- 3) Преобразуйте текстовые данные в векторный вид, TF-IDF
- 4) Постройте на обучающем подмножестве следующие модели классификации:
  - К-ближайших соседей
  - Логистическая регрессия
  - Наивный байесовский классификатор: модель Бернулли
  - Наивный байесовский классификатор: полиномиальная модель
- 5) Определите качество классификации на тестовом подмножестве
- 6) Определите с помощью timeit время обучения и предсказания (на тестовом подмножестве)
- 7) Повторите процесс для различных гиперпараметров моделей
- 8) Сравните полученные результаты и сделайте выводы

### Лабораторное занятие №2 (6 ч.)

Тема **Современные методы визуализации информации Big Data**

Вопросы для обсуждения:

1. Поточковый график, матричная диаграмма, диаграмма рассеяния
2. Задача регрессии. Минимизация квадрата отклонения. Регрессионная функция: условное мат.ожидание.
3. Линейная регрессия и метод k ближайших соседей.
4. Переобучение и недообучение Особенности реализации.

#### Примерные варианты индивидуальных заданий

**Задача** Кластеризация данных

- 1) Загрузите исходные данные
- 2) Постройте графики исходных данных
- 3) Произведите кластеризацию данных с использованием заданных методов и их параметров
- 4) Выберите подходящее количество кластеров
- 5) Постройте результирующие графики

### **Лабораторное занятие №3 (6 ч.)**

Тема **Задачи параллельной обработки данных в научных исследованиях**

Вопросы для обсуждения:

1. Проблемные точки процесса реализации программы
2. Очистка данных и технологии регуляризации.
3. Основные виды регуляризации.
4. Метод редукции размерности.
5. Методы отбора признаков
6. Особенности реализации.
7. .

**Примерные варианты индивидуальных заданий**

15 вариантов индивидуальных заданий по теме **Задачи параллельной обработки данных в научных исследованиях**

### **Лабораторное занятие №4 (6 ч.)**

Тема **Технологии кластеризации и классификации. Нейронные сети. Генетические алгоритмы**

Вопросы для обсуждения:

1. Логистическая регрессия, максимизация правдоподобия.
2. Нейронные сети: общая архитектура.
3. Многослойные сети.
4. Обратное распространение ошибки.
5. Стохастический градиентный спуск.
6. Генетические алгоритмы.
7. Особенности реализации.

**Примерные варианты индивидуальных заданий**

15 вариантов индивидуальных заданий по теме **Кластеризация и классификация**

1. Построить нейронную сеть в среде Deductor (Loginom) для решения задачи прогнозирования
2. Решить задачу классификации в среде Wolfram Mathematica

### **Лабораторное занятие №5 (6 ч.)**

Тема **Выделение особенностей (Feature detection); нормализация данных. Нечеткие множества. Байесовы сети**

Вопросы для обсуждения:

1. Методы нормализации данных.
2. Нормализация по методу минимакса.
3. Нормализация по Z-показателю.
4. Десятичное масштабирование.
5. Нечеткие множества
6. Особенности реализации.

**Примерные варианты индивидуальных заданий**

15 вариантов индивидуальных заданий по теме **Нечеткие множества**

1. Создайте лингвистическую шкалу для переменной «Скорость автомобиля» при условии, что минимальная скорость равна 0 км/ч, а максимальная 120 км/ч. Используйте не менее 4-х термов. Изобразите полученную шкалу.

2. Постройте функцию принадлежности нечеткого множества «высокий мужчина» на универсальном множестве {170, 175, 180, 185, 190, 195}.

Парные экспертные сравнения заданы следующей матрицей:

$$A = \begin{matrix} & \begin{matrix} (170 & 175 & 180 & 185 & 190 & 195) \end{matrix} \\ \begin{pmatrix} 170 \\ 175 \\ 180 \\ 185 \\ 190 \\ 195 \end{pmatrix} & \begin{bmatrix} 1 & 1/2 & 1/4 & 1/6 & 1/8 & 1/9 \\ 2 & 1 & 1/3 & 1/5 & 1/7 & 1/8 \\ 4 & 3 & 1 & 1/4 & 1/4 & 1/5 \\ 6 & 5 & 4 & 1 & 1/3 & 1/3 \\ 8 & 7 & 4 & 3 & 1 & 1 \\ 9 & 8 & 5 & 3 & 1 & 1 \end{bmatrix} \end{matrix}.$$

Нормализуйте полученное нечеткое множество, постройте графики функций принадлежности.

Построить интеллектуальную систему поддержки решения по оценке стоимости квартиры

Форма контроля – *зачет*

**Примерные вопросы к зачету**

1. Что такое большие данные
2. Принципы организации и управления инфраструктурой больших данных
3. Задачи машинного обучения, классификацию задач, математическое определение, примеры.
4. Основные этапы построения моделей в задачах регрессии и классификации.
5. Случайное событие, случайная величина, закон распределения, математическое ожидание, дисперсия, мода, медиан.
6. Сложение и умножение вероятностей. Формула полной вероятности, формула Байеса.
7. Основные законы распределения и их параметры.
8. Построение гипотез и доверительный интервал.
9. Задача оптимизации. Градиентный спуск.
10. Алгебраические модели представления данных. Гистограммная модель.
11. Регрессия. Метод наименьших квадратов и градиентный спуск.
12. Регрессия. Градиентный спуск и стохастический градиентный спуск.
13. Регрессия и классификация. Метод ближайших соседей.
14. Регрессия. Байесовская линейная регрессия.
15. Регрессия. Деревья решений. Критерии деления

**Критерии оценивания**

**Оценка «зачтено»** выставляется:

- студенту глубоко и прочно усвоившему программный материал, исчерпывающе, последовательно, грамотно и логически стройно его излагающему, в ответе которого увязывается теория с практикой, он показывает знакомство с литературой, правильно обосновывает и использует рациональные и современные средства решения поставленной проблемы.
- студенту, твердо знающему программный материал, грамотно и по существу, излагающему его, который не допускает существенных неточностей в ответе на вопрос, правильно применяет теоретические положения при решении поставленной задачи.

- студенту, который знает только основной программный материал, но не усвоил особенностей, допускает в ответе неточности, некорректно формулирует основные законы и правила, затрудняется в выполнении практических задач.

**Оценка «не зачтено»** выставляется студенту, который не знает значительной части программного материала, допускает в ответе существенные ошибки, с затруднениями практические задания.

Форма контроля	За одну работу		Всего	
	Мин. баллов	Макс. баллов	Мин. баллов	Макс. баллов
Текущий контроль:				
Активная работа на занятии	0,25	0,5	9	18
Выполнение домашнего задания	0,75	0,75	27	27
Выполнение заданий самостоятельной работы	1	3	1	3
Коллоквиум	1	3	3	9
Промежуточная аттестация (зачет)			20	43
<b>Итого за семестр</b>			<b>60</b>	<b>100</b>

Составитель \_\_\_\_\_



Г.С. Осипов,  
профессор кафедры информатики

«19» сентября 2024 г.