

Федеральное государственное бюджетное образовательное учреждение высшего образования

«Сахалинский государственный университет»

Кафедра информатики

УТВЕРЖДЕН

на заседании кафедры

«22» мая 2025 г., протокол № 9

И.О. заведующего кафедрой



Осипов Г.С.

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

**Б1.В.ДВ.07.02 Технологии машинного обучения и анализа больших
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 проектирование 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 (модулю)

№ п/п	Контролируемые разделы (темы) дисциплины	Код контролируемой компетенции (или ее части)	Наименование оценочного средства
1.	Тема 1. Введение в машинное обучение и обработку данных. Программный инструментарий Data Mining и Machine Learning	ПКС-2	задания к лабораторным работам
2.	Тема 2. Основной аппарат комбинаторики и мат. статистики. Регрессионный анализ и сжатие данных	ПКС-2	задания к лабораторным работам
3.	Тема 3 Детектирование выбросов и аномалий. Очистка данных и технологии регуляризации	ПКС-2	задания к лабораторным работам
4.	Тема 4 Технологии кластеризации и классификации. Нейронные сети. Генетические алгоритмы	ПКС-2	задания к лабораторным работам
5.	Тема 5 Выделение особенностей (Feature detection); нормализация данных. Нечеткие множества. Байесовы сети	ПКС-2	задания к лабораторным работам
	зачет	ПКС-2	вопросы к зачету.

3. 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 компетенций:

ПКС-2 Способен проводить формализацию предметной области с целью создания информационной системы

Множественный выбор (один верный)

1 Выберите правильный ответ:

Что такое «машинное обучение» в контексте работы с большими данными?

1. Ручной анализ данных экспертом
2. Раздел искусственного интеллекта, позволяющий системам автоматически обучаться и улучшаться на основе опыта (данных)
3. Процесс хранения данных на разных серверах
4. Простая статистическая выборка

Ответ: 2

2 Выберите правильный ответ:

Что означает свойство "Velocity" в контексте Big Data?

1. Точность поступающих данных
2. Скорость генерации и обработки данных
3. Местонахождение данных
4. Ценность данных

Ответ 2

3 Выберите правильный ответ:

Какой тип баз данных наиболее подходит для хранения неструктурированных данных, таких как сообщения из социальных сетей или логи веб-сервера?

1. Реляционные базы данных (SQL)
2. Нереляционные базы данных (NoSQL)
3. Табличные процессоры
4. Иерархические базы данных

Ответ 2

4 Выберите правильный ответ:

Что такое "data lake" (озеро данных)?

1. База данных с жесткой, предопределенной схемой
2. Централизованное хранилище, позволяющее хранить все структурированные и неструктурированные данные любого масштаба
3. Инструмент для визуализации данных
4. Система для очистки данных

Ответ 2

5 Выберите правильный ответ:

Какой процесс заключается в обнаружении и исправлении ошибок и несоответствий в данных?

1. Интеграция данных
2. Очистка данных (Data cleaning)
3. Визуализация данных
4. Агрегация данных

Ответ 2

6 Выберите правильный ответ:

Что из перечисленного является примером инструмента для потоковой обработки данных?

1. Apache Hive
2. Apache Kafka
3. Apache HBase
4. Tableau

Ответ 2

7 Выберите правильный ответ:

Какой язык программирования является одним из самых популярных для задач анализа больших данных и data science?

1. JavaScript
2. Python
3. C++
4. PHP

Ответ 2

8 Выберите правильный ответ:

Какой подход к анализу данных основывается на изучении прошлых данных для выявления тенденций?

1. Пресцитативная аналитика
2. Дескриптивная аналитика
3. Предиктивная аналитика
4. Предписывающая аналитика

Ответ 2

9 Выберите правильный ответ:

Что изучает дисциплина "Data Science" (наука о данных)?

1. Только разработку баз данных
2. Извлечение знаний из структурированных и неструктурированных данных с помощью научных методов
3. Проектирование компьютерных сетей
4. Создание аппаратного обеспечения для серверов

Ответ 2

10 Выберите правильный ответ:

Что такое "предиктивная аналитика"?

1. Описание того, что произошло в прошлом
2. Использование данных и алгоритмов для прогнозирования будущих результатов на основе исторических данных
3. Рекомендация действий для достижения желаемых результатов
4. Поточковая обработка событий в реальном времени

Ответ 2

11 Выберите правильный ответ:

Какой термин описывает данные, которые имеют определенную структуру и могут быть легко представлены в таблицах с строками и столбцами?

1. Неструктурированные данные
2. Полуструктурированные данные
3. Структурированные данные
4. Сырые данные (raw data)

Ответ 3

12 Выберите правильный ответ:

Какой аспект Big Data вызывает наибольшие опасения с точки зрения конфиденциальности и безопасности?

1. Сложность алгоритмов
2. Риск несанкционированного доступа и неправильное использование персональной информации
3. Высокая стоимость хранения
4. Необходимость в мощном оборудовании

Ответ 2

13 Выберите правильный ответ:

Что такое логистическая регрессия?

1. Алгоритм для бинарной классификации
2. Алгоритм для кластеризации
3. Алгоритм для регрессии

Ответ 1

14 Выберите правильный ответ:

Что такое one-hot encoding?

1. Группировка данных
2. Удаление данных
3. Преобразование категориальных данных в бинарные векторы

Ответ 3

15 Выберите правильный ответ:

Что такое тестовая выборка?

1. Данные для обучения модели
2. Данные для оценки модели
3. Данные для настройки гиперпараметров

Ответ 2

16 Выберите правильный ответ:

Что такое F1-мера?

1. Среднее гармоническое точности и полноты
2. Мера сложности модели

3. Доля ошибок модели
Ответ 1

17 Выберите правильный ответ:

Что такое обучающая выборка?

1. Данные для валидации модели
2. Данные для обучения модели
3. Данные для тестирования модели

Ответ 2

18 Выберите правильный ответ:

Что такое pipeline в машинном обучении?

1. Метод визуализации данных
2. Метод кластеризации
3. Последовательность шагов для обработки данных и обучения модели

Ответ 3

19 Выберите правильный ответ:

Что такое random search?

1. Метод кластеризации
2. Метод случайного поиска гиперпараметров
3. Метод визуализации данных

Ответ 2

20 Выберите правильный ответ:

Что такое бэггинг (bagging)?

1. Обучение модели на разных подвыборках данных
2. Удаление данных
3. Обучение модели последовательно

Ответ 1

21 Выберите правильный ответ:

Что такое ансамбль моделей?

1. Удаление данных
2. Комбинация нескольких моделей для улучшения предсказаний
3. Упрощение модели

Ответ 2

22 Выберите правильный ответ:

Что такое нормализация данных?

1. Группировка данных
2. Приведение данных к одному масштабу
3. Удаление данных

Ответ 2

23 Выберите правильный ответ:

Что такое недообучение (underfitting)?

1. Модель идеально подходит
2. Модель слишком сложная
3. Модель слишком простая и не обучается

Ответ 3

24 Выберите правильный ответ:

Что такое целевая переменная?

1. Переменная для визуализации данных
2. Переменная для обучения модели
3. Переменная, которую модель предсказывает

Ответ 3

25 Выберите правильный ответ:

Что такое метод главных компонент (РСА)?

1. Метод снижения размерности данных
2. Метод классификации
3. Метод кластеризации

Ответ 1

26 Выберите правильный ответ:

Что такое кластеризация?

1. Группировка похожих объектов
2. Прогнозирование значений
3. Классификация данных

Ответ 1

27 Выберите правильный ответ:

Что такое кросс-валидация?

1. Метод обучения на одной выборке
2. Метод оценки модели на нескольких подвыборках
3. Метод визуализации данных

Ответ 2

28 Выберите правильный ответ:

Что такое confusion matrix (матрица ошибок)?

1. Таблица для оценки качества классификации
2. Метод кластеризации
3. Метод визуализации данных

Ответ 1

29 Выберите правильный ответ:

Когда говорят о нейронных сетях и машинном обучении, часто упоминают закон Мура. В чем его суть?

1. Не следует множить сущее без необходимости
2. Каждое следующее поколение компьютеров работает в 2,5 раза быстрее
3. Если все слова языка или длинного текста упорядочить по убыванию частоты их использования, то частота n-го слова в таком списке окажется приблизительно обратно пропорциональной его порядковому номеру n
4. 20% усилий дают 80% результата, а остальные 80% усилий — лишь 20% результата

Ответ 2

Множественный выбор (несколько верных)

30 Выберите все правильные ответы:

Какие виды искусственных нейронных сетей существуют?

1. Рекуррентные
2. Наивные
3. Противоборствующие
4. Импульсные

Ответ: 1, 3, 4

31 Выберите все правильные ответы:

Какие из видов машинного обучения НЕ основываются на взаимодействии обучаемой системы со средой?

1. Обучение с подкреплением
2. Глубинное обучение
3. Обучение с учителем
4. Обучение без учителя

Ответ 2, 3, 4

32 Выберите все правильные ответы:

Нужно рассчитать необходимые параметры для создания обшивки самолета. Какая из областей машинного обучения нам в этом НЕ пригодится?

1. Латентная модель
2. Предсказательное моделирование
3. Обучение ранжированию
4. Компьютерное зрение

Ответ 1, 3, 4

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