

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«ЛУГАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ ВЛАДИМИРА ДАЛЯ»

Краснодонский факультет инженерии и менеджмента (филиал)
Кафедра информационных технологий и транспорта



УТВЕРЖДАЮ:

Директор

Панайотов К.К.

(подпись)

«14» марта 2025 года

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

по учебной дисциплине

Интеллектуальные системы

(наименование учебной дисциплины, практики)

09.04.01 Информатика и вычислительная техника

(код и наименование направления подготовки (специальности))

«Интеллектуальные системы

в производственно-транспортных комплексах»

наименование профиля подготовки (специальности, магистерской программы); при отсутствии ставится прочерк)

Разработчик(разработчики):

доцент


(подпись)

Бихдрикер А. С.

ФОС рассмотрен и одобрен на заседании кафедры информационных технологий и транспорта от «26» февраля 2025 г., протокол № 7

Заведующий кафедрой
информационных
технологий и транспорта


(подпись)

Верительник Е. А.

Краснодон 2025

**Комплект оценочных материалов по дисциплине
«Интеллектуальные системы»**

Задания закрытого типа

Задания закрытого типа на выбор правильного ответа

1. *Выберите один правильный ответ.*

Какой из методов относится к методам классификации в интеллектуальном анализе данных?

- А) Алгоритм a priori.
- Б) Метод K-means.
- В) Метод CART.
- Г) Анализ временных рядов.

Правильный ответ: В.

Компетенции (индикаторы): ПК-3 (ПК-3.1).

2. *Выберите один правильный ответ.*

Как называется процесс обнаружения полезных закономерностей в больших объемах данных?

- А) Data Warehousing.
- Б) Data Mining.
- В) Big Data Processing.
- Г) Data Structuring.

Правильный ответ: Б

Компетенции (индикаторы): ПК-3 (ПК-3.1).

3. *Выберите один правильный ответ.*

Какой алгоритм используется для кластерного анализа?

- А) ID3.
- Б) K-means.
- В) Логистическая регрессия.
- Г) Метод опорных векторов.

Правильный ответ: Б

Компетенции (индикаторы): ПК-3 (ПК-3.1).

4. *Выберите один правильный ответ.*

Какой из следующих методов чаще всего используется для прогнозирования временных рядов?

- А) K-means.
- Б) Деревья решений.
- В) Линейная регрессия.
- Г) Байесовские сети.

Правильный ответ: В

Компетенции (индикаторы): ПК-3 (ПК-3.1).

Задания закрытого типа на установление соответствия

1. Установите соответствие. Каждому элементу левого столбца соответствует только один элемент правого столбца.

Этап анализа данных	Описание
1) Очистка данных	А) Определение цели анализа и постановка задачи.
2) Предобработка данных	Б) Приведение данных к удобному формату, кодирование категориальных переменных
3) Разработка модели	В) Выбор и обучение алгоритма машинного обучения.
4) Оценка модели	Г) Проверка модели на тестовых данных, вычисление метрик точности.

Правильный ответ: 1-Г, 2-А, 3-Б, 4-В

Компетенции (индикаторы): ПК-3 (ПК-3.1).

2. Установите соответствие. Каждому элементу левого столбца соответствует только один элемент правого столбца.

Метод	Характеристика
1) Фильтрационные методы	А) Отбирают признаки на основе их статистических свойств (корреляция, дисперсия).
2) Встроенные методы	Б) Встроены в алгоритм обучения, отбор происходит в процессе обучения модели.
3) Жадные методы	В) Последовательно добавляют или удаляют признаки, оценивая влияние каждого.

Правильный ответ: 1-В, 2-А, 3-Б

Компетенции (индикаторы): ПК-3 (ПК-3.1).

3. Установите соответствие. Каждому элементу левого столбца соответствует только один элемент правого столбца.

Алгоритм	Особенность
1) K-means	А) Использует центроиды для группировки данных.
2) DBSCAN	Б) Определяет кластеры на основе плотности данных.
3) Сети Кохонена	В) Используют нейронные сети для самоорганизации данных.

Правильный ответ: 1-Б, 2-В, 3-А

Компетенции (индикаторы): ПК-3 (ПК-3.1).

4. Установите соответствие. Каждому элементу левого столбца соответствует только один элемент правого столбца.

Метрика	Описание
1) Точность (Accuracy)	А) Доля правильно предсказанных примеров ко всем примерам.
2) Полнота (Recall)	Б) Доля правильно предсказанных положительных примеров ко всем реальным положительным примерам.
3) F1-мера	В) Среднее гармоническое точности и полноты.

Правильный ответ: 1-В, 2-А, 3-Б

Компетенции (индикаторы): ПК-3 (ПК-3.1).

Задания закрытого типа на установление правильной последовательности

1. Установите правильную последовательность действий при подготовке набора данных для построения модели машинного обучения. Запишите правильную последовательность букв слева направо:

Для успешного построения модели необходимо сначала привести данные в пригодный для анализа вид, а затем выполнить их обработку. В процессе подготовки данных могут использоваться различные методы очистки, трансформации и нормализации признаков. Запишите правильную последовательность букв слева направо.

А) Заполнение пропущенных значений и устранение выбросов.

Б) Разделение данных на обучающую, валидационную и тестовую выборки.

В) Кодирование категориальных переменных и нормализация числовых данных.

Г) Первичный анализ данных, выявление аномалий и пропусков.

Д) Очистка данных от дубликатов и устранение несоответствий.

Правильный ответ: Г, Д, А, В, Б.

Компетенции (индикаторы): ПК-3 (ПК-3.1)

2. Установите правильную последовательность этапов построения ансамблевой модели. Запишите правильную последовательность букв слева направо:

Ансамблирование моделей позволяет повысить точность предсказаний за счет комбинирования нескольких алгоритмов. Перед построением ансамбля необходимо провести предварительную обработку данных, затем обучить и объединить модели. Запишите правильную последовательность букв слева направо.

А) Определение типа ансамбля (бэггинг, бустинг или стекинг).

Б) Выбор базовых моделей и их обучение на обучающей выборке.

В) Оценка качества ансамбля на тестовой выборке.

Г) Подготовка данных и выбор целевой переменной.

Д) Комбинирование предсказаний отдельных моделей в единый результат.

Правильный ответ: Г, А, Б, Д, В.

Компетенции (индикаторы): ПК-3 (ПК-3.1)

3. Установите правильную последовательность действий при построении модели нейронной сети. Запишите правильную последовательность букв слева направо:

Перед запуском нейронной сети необходимо подготовить данные и выбрать архитектуру модели. Затем следует процесс обучения и оценки качества работы сети. Запишите правильную последовательность букв слева направо.

А) Определение количества слоев и нейронов в каждом слое.

Б) Подготовка и нормализация данных для входа в сеть.

В) Выбор функции активации и метода оптимизации.

Г) Обучение модели на тренировочных данных.

Д) Оценка качества модели на тестовой выборке и настройка гиперпараметров.

Правильный ответ: Б, А, В, Г, Д.

Компетенции (индикаторы): ПК-3 (ПК-3.1)

4. Установите правильную последовательность действий при построении модели прогнозирования временного ряда. Запишите правильную последовательность букв слева направо:

Прогнозирование временных рядов требует подготовки данных и выбора модели, которая сможет учитывать тренды и сезонность. Для получения корректных предсказаний необходимо выполнить несколько ключевых шагов. Запишите правильную последовательность букв слева направо.

А) Разделение временного ряда на обучающую и тестовую выборки.

Б) Выбор модели прогнозирования (ARIMA, экспоненциальное сглаживание и т. д.).

В) Оценка качества прогноза и оптимизация гиперпараметров.

Г) Анализ временного ряда на наличие тренда и сезонности.

Д) Обучение модели на исторических данных и построение прогноза.

Правильный ответ: Г, А, Б, Д, В.

Компетенции (индикаторы): ПК-3 (ПК-3.1)

Задания открытого типа

Задания открытого типа на дополнение

1. Напишите пропущенное слово или словосочетание.

Метод, используемый для автоматического поиска зависимостей и закономерностей в больших массивах данных, называется _____.

Правильный ответ: Data Mining/интеллектуальный анализ данных.

Компетенции (индикаторы): ПК-3 (ПК-3.1).

2. *Напишите пропущенное слово или словосочетание.*

Алгоритм, который разделяет объекты на группы на основе их схожести и используется для кластеризации данных, называется _____.

Правильный ответ: K-means/метод k-средних.

Компетенции (индикаторы): ПК-3 (ПК-3.1).

3. *Напишите пропущенное слово или словосочетание.*

Формула, используемая для измерения качества регрессионной модели и отражающая среднюю разницу между предсказанными и реальными значениями, называется _____.

Правильный ответ: Среднеквадратичная ошибка/MSE (Mean Squared Error).

Компетенции (индикаторы): ПК-3 (ПК-3.1).

4. *Напишите пропущенное слово или словосочетание.*

Процесс приведения числовых данных к одному масштабу для корректного обучения моделей машинного обучения называется _____.

Правильный ответ: Нормализация/стандартизация.

Компетенции (индикаторы): ПК-3 (ПК-3.1).

Задания открытого типа с кратким свободным ответом

1. *Дайте ответ на вопрос.* Осуществите вычисления и прокомментируйте результат.

В ходе кластерного анализа методом K-means были сформированы 4 кластера. Среднее расстояние от точек до центроидов кластеров составило 2.1, 3.5, 1.8 и 2.9 соответственно. Рассчитайте среднее внутрикластерное расстояние. Ответ округлите до одного знака после запятой.

Правильный ответ: 2.6. Чем меньше среднее внутрикластерное расстояние, тем плотнее расположены точки внутри кластеров, что свидетельствует о более качественной кластеризации.

Компетенции (индикаторы): ПК-3 (ПК-3.1).

2. *Дайте ответ на вопрос.* Осуществите вычисления и прокомментируйте результат.

Датасет содержит 1000 наблюдений. Модель классификации предсказала 450 объектов как положительные, из которых 360 оказались верными. Полнота (Recall) рассчитывается по формуле:

$$Recall = \frac{TP}{TP + FN}$$

где TP (True Positives) = 360, а FN (False Negatives) = 90. Найдите значение полноты и округлите до сотых.

Правильный ответ: 0.80 / 80%. Это означает, что модель корректно определила 80% всех положительных объектов, но 20% остались

нераспознанными, что может повлиять на общую эффективность предсказаний.

Компетенции (индикаторы): ПК-3 (ПК-3.1).

3. *Дайте ответ на вопрос.* Осуществите вычисления и прокомментируйте результат.

Имеется временной ряд продаж за 4 месяца: 120, 135, 150 и 165 единиц. Рассчитайте прогноз на 5-й месяц, используя метод скользящего среднего (moving average) с окном 3. Ответ округлите до целого.

Правильный ответ: 150. Это означает, что продажи остаются стабильными без резких изменений, но метод скользящего среднего не учитывает возможные тренды и сезонные колебания.

Компетенции (индикаторы): ПК-3 (ПК-3.1).

4. *Дайте ответ на вопрос.* Осуществите вычисления и прокомментируйте результат.

Даны два признака: $X = \{2, 4, 6, 8, 10\}$ и $Y = \{3, 6, 9, 12, 15\}$. Рассчитайте коэффициент корреляции Пирсона между X и Y . Ответ округлите до сотых.

Правильный ответ: 1.00. Это означает, что между переменными существует идеальная линейная зависимость: увеличение X приводит к пропорциональному увеличению Y .

Компетенции (индикаторы): ПК-3 (ПК-3.1).

Задания открытого типа с развернутым ответом

1. *Дайте развернутый ответ на вопрос.*

Дайте определение метрики «Точность» (Accuracy) в задаче классификации. Назовите основные преимущества и недостатки.

Время выполнения: 15 мин.

Ожидаемый результат:

Точность (Accuracy) — это метрика, которая оценивает долю правильно классифицированных объектов ко всем объектам в наборе данных.

Преимущества точности.

1. Простота интерпретации. Легко понять и объяснить значение точности. Она показывает общую долю правильно классифицированных объектов.

2. Простота вычисления. Легко вычислить, используя простую формулу.

3. Интуитивность. Хорошо отражает общую производительность модели в сбалансированных наборах данных.

Недостатки точности.

1. Неинформативность для несбалансированных классов. Основной недостаток — плохая пригодность для задач с сильно несбалансированными классами (когда один класс значительно преобладает над другими). В таких случаях, модель может достигать высокой точности, просто предсказывая преобладающий класс для всех объектов.

2. Не учитывает типы ошибок. Точность не различает ошибки первого рода и ошибки второго рода, хотя в некоторых задачах один тип ошибки может быть значительно более критичным, чем другой.

3. Не подходит для многоклассовой классификации с неравной стоимостью ошибок. Точность предполагает одинаковую «стоимость» для всех ошибок классификации, что не всегда верно.

Критерии оценивания: наличие в ответе определения метрики и перечисление преимуществ и недостатков.

Компетенции (индикаторы): ПК-3 (ПК-3.1).

2. Дайте развернутый ответ на вопрос.

Опишите пошагово алгоритм К-средних (K-means) для кластеризации данных.

Время выполнения: 20 мин.

Ожидаемый результат:

Алгоритм К-средних (K-means) – это итеративный алгоритм кластеризации, который стремится разделить n объектов на k кластеров, в которых каждый объект принадлежит кластеру с ближайшим средним значением (центроидом). Алгоритм состоит из следующих шагов:

1. Инициализация. Выбирается k начальных центроидов кластеров. Это можно сделать случайно (например, выбрав k случайных объектов из набора данных) или с использованием других методов (например, K-means++).

2. Присвоение объектов кластерам: Для каждого объекта в наборе данных вычисляется расстояние до каждого центроида. Присваивается объект кластеру с ближайшим центроидом. Обычно используется евклидово расстояние, но можно использовать и другие метрики расстояния.

3. Пересчет центроидов. Для каждого кластера вычисляется новый центроид как среднее значение всех объектов, принадлежащих этому кластеру.

4. Проверка условия остановки. Повторяются шаги 2 и 3 до тех пор, пока не выполнится одно из следующих условий остановки: центроиды больше не меняются (или меняются незначительно); объекты больше не переходят между кластерами; достигнуто максимальное количество итераций.

Результатом работы алгоритма является разделение объектов на k кластеров, где каждый кластер характеризуется своим центроидом.

Критерии оценивания: наличие в ответе определения алгоритма и каждый шаг алгоритма должен быть описан минимум одним предложением.

Компетенции (индикаторы): ПК-3 (ПК-3.1).

3. Дайте развернутый ответ на вопрос.

Дайте определение ансамблированию моделей. Объясните, зачем используются ансамбли моделей, какие преимущества они обеспечивают по сравнению с одиночными моделями.

Время выполнения: 20 мин.

Ожидаемый результат:

Ансамблирование моделей – это метод машинного обучения, который объединяет предсказания нескольких базовых моделей для получения более точного и надежного предсказания. Вместо того, чтобы полагаться на одну сложную модель, ансамблирование использует несколько простых моделей, которые работают вместе.

Цели применения ансамблей моделей.

1. Улучшение точности. Ансамбли моделей часто обеспечивают более высокую точность, чем одиночные модели. Это связано с тем, что ансамблирование позволяет снизить как смещение, так и дисперсию модели.
2. Снижение переобучения. Ансамблирование помогает снизить риск переобучения, так как усредняет ошибки различных моделей.
3. Улучшение надежности. Ансамблирование более устойчиво к шуму в данных и ошибкам отдельных моделей. Если одна или несколько базовых моделей ошибаются, ансамбль в целом все равно может выдать правильное предсказание.
4. Улучшение обобщающей способности. Ансамблирование позволяет улучшить обобщающую способность модели, то есть ее способность хорошо работать на новых, ранее не виданных данных.

Преимущества ансамблирования по сравнению с одиночными моделями.

1. Более высокая точность и надежность.
2. Снижение переобучения и улучшение обобщающей способности.
3. Устойчивость к шуму и ошибкам отдельных моделей.
4. Возможность использования различных типов базовых моделей.

Критерии оценивания: наличие в ответе трёх целей и трёх преимуществ.

Компетенции (индикаторы): ПК-3 (ПК-3.1).

4. Дайте развернутый ответ на вопрос.

Объясните, почему важно разделять данные на три отдельные выборки: обучающую, валидационную и тестовую. Опишите роль каждой из этих выборок в процессе построения и оценки модели машинного обучения.

Время выполнения: 20 мин.

Ожидаемый результат:

Разделение данных на обучающую, валидационную и тестовую выборки является критически важным шагом при построении моделей машинного обучения. Цели разделения данных:

1. Оценка обобщающей способности модели. Обучающая выборка используется для обучения модели, но оценка ее качества на тех же данных может быть обманчивой, так как модель может запомнить обучающие данные, а не выучить общие закономерности. Валидационная и тестовая выборки позволяют оценить, насколько хорошо модель обобщает на новые, ранее не виданные данные.
2. Предотвращение переобучения. Разделение на три выборки позволяет контролировать процесс обучения и предотвращать переобучение.

3. Объективной оценки качества модели. Тестовая выборка используется для окончательной оценки качества модели после того, как все параметры и гиперпараметры модели были настроены с использованием валидационной выборки.

Роль каждой выборки.

1. Обучающая выборка. Используется для обучения модели. Модель учится находить закономерности в данных и устанавливать параметры, которые минимизируют ошибку на обучающей выборке.

2. Валидационная выборка. Используется для настройки гиперпараметров модели и для выбора наилучшей модели из нескольких кандидатов. Модель оценивается на валидационной выборке после каждой итерации обучения или после изменения гиперпараметров. Результаты оценки используются для корректировки параметров модели и предотвращения переобучения. Валидационная выборка позволяет оценить, насколько хорошо модель обобщает на новые данные, не участвуя непосредственно в процессе обучения.

3. Тестовая выборка. Используется для окончательной, независимой оценки качества обученной модели. Тестовая выборка используется только один раз, в самом конце процесса построения модели, после того, как все параметры и гиперпараметры были настроены с использованием обучающей и валидационной выборок. Это позволяет получить объективную оценку обобщающей способности модели.

Критерии оценивания: каждая цель и роль разделения данных должна быть описана минимум одним предложением.

Компетенции (индикаторы): ПК-3 (ПК-3.1).

Экспертное заключение

Представленный фонд оценочных средств (далее – ФОС) по дисциплине «Интеллектуальные системы» соответствует требованиям ФГОС ВО.

Предлагаемые формы и средства текущего и промежуточного контроля адекватны целям и задачам реализации основной профессиональной образовательной программы по направлению подготовки 09.04.01 Информатика и вычислительная техника.

Оценочные средства для текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины представлены в полном объеме.

Виды оценочных средств, включенные в представленный фонд, отвечают основным принципам формирования ФОС.

Разработанный и представленный для экспертизы фонд оценочных средств рекомендуется к использованию в процессе подготовки обучающихся по указанному направлению 09.04.01 Информатика и вычислительная техника.

Председатель учебно-методической
комиссии Краснодонского факультета
инженерии и менеджмента (филиала)

 Родионова О.Ю.

Лист изменений и дополнений

№ п/п	Виды дополнений и изменений	Дата и номер протокола заседания кафедры (кафедр), на котором были рассмотрены и одобрены изменения и дополнения	Подпись (с расшифровкой) заведующего кафедрой (заведующих кафедрами)