

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.

«14» марта 2025 года

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ
по учебной дисциплине
Интеллектуальный анализ данных
(наименование учебной дисциплины, практики)

38.03.05 Бизнес-информатика
(код и наименование направления подготовки (специальности))
«Информационная бизнес-аналитика»

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

Разработчик(разработчики):
доцент

Попова Н.Н.
(подпись)

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

Заведующий кафедрой
информационных
технологий и транспорта

Верительник Е.А.
(подпись)

Краснодон 2025

Комплект оценочных материалов по дисциплине (практике)
«Интеллектуальный анализ данных»

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): ОПК-6 (ОПК-6.2)

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): ОПК-6 (ОПК-6.2)

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): ОПК-6 (ОПК-6.2)

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): ОПК-6 (ОПК-6.2)

5. Выберите один правильный ответ.

Какой алгоритм машинного обучения основан на деревьях решений?

- А) Метод KNN
- Б) Алгоритм CART
- В) Байесовская классификация
- Г) Метод главных компонент

Правильный ответ: Б

Компетенции (индикаторы): ОПК-6 (ОПК-6.2)

6. Выберите один правильный ответ.

Какой метод относится к ансамблевым моделям?

- А) Бэггинг
- Б) Кросс-валидация
- В) K-means
- Г) Карты Кохонена

Правильный ответ: А

Компетенции (индикаторы): ОПК-6 (ОПК-6.2)

7. Выберите один правильный ответ.

Какая из следующих моделей используется для предсказания категориальных переменных?

- А) Линейная регрессия
- Б) Логистическая регрессия
- В) K-means
- Г) Метод главных компонент

Правильный ответ: Б

Компетенции (индикаторы): ОПК-6 (ОПК-6.2)

8. Выберите один правильный ответ.

Какой алгоритм машинного обучения относится к нейронным сетям?

- А) Метод CART
- Б) K-means
- В) Алгоритм обратного распространения ошибки
- Г) Байесовская классификация

Правильный ответ: В

Компетенции (индикаторы): ОПК-6 (ОПК-6.2)

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): ОПК-6 (ОПК-6.2)

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): ОПК-6 (ОПК-6.2)

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): ОПК-6 (ОПК-6.2)

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): ОПК-6 (ОПК-6.2)

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

Тип регрессии	Применение
1) Линейная регрессия	А) Оценка зависимости одной переменной от другой в числовых данных
2) Логистическая регрессия	Б) Прогнозирование бинарного результата (да/нет)
3) Полиномиальная регрессия	В) Используется, если зависимость переменной нелинейна

Правильный ответ: 1-Б, 2-В, 3-А

Компетенции (индикаторы): ОПК-6 (ОПК-6.2)

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): ОПК-6 (ОПК-6.2)

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): ОПК-6 (ОПК-6.2)

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): ОПК-6 (ОПК-6.2)

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): ОПК-6 (ОПК-6.2)

5. Установите правильную последовательность действий при оценке качества модели машинного обучения. Запишите правильную последовательность букв слева направо:

После обучения модели важно проверить ее способность к генерализации и выявить возможные проблемы переобучения. Оценка качества модели включает проверку ее работы на тестовых данных и анализ метрик. Запишите правильную последовательность букв слева направо.

- А) Вычисление метрик (точность, полнота, F1-мера, MSE, R² и др.).
- Б) Разделение данных на обучающую и тестовую выборки.
- В) Анализ ошибок модели и поиск возможных улучшений.
- Г) Проверка модели на тестовых данных.
- Д) Обучение модели на тренировочной выборке.

Правильный ответ: Б, Д, Г, А, В

Компетенции (индикаторы): ОПК-6 (ОПК-6.2)

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): ОПК-6 (ОПК-6.2)

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): ОПК-6 (ОПК-6.2)

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): ОПК-6 (ОПК-6.2)

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): ОПК-6 (ОПК-6.2)

5. Напишите пропущенное слово или словосочетание.

Модель машинного обучения, имитирующая работу человеческого мозга и состоящая из искусственных нейронов, называется _____.

Правильный ответ: Нейронная сеть/искусственная нейронная сеть (ИНС).

Компетенции (индикаторы): ОПК-6 (ОПК-6.2)

6. Напишите пропущенное слово или словосочетание.

Метод уменьшения размерности данных, который используется для выделения главных компонент и удаления избыточной информации, называется _____.

Правильный ответ: Метод главных компонент/PCA (Principal Component Analysis).

Компетенции (индикаторы): ОПК-6 (ОПК-6.2)

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): ОПК-6 (ОПК-6.2)

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): ОПК-6 (ОПК-6.2)

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): ОПК-6 (ОПК-6.2)

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): ОПК-6 (ОПК-6.2)

5. Дайте ответ на вопрос. Осуществите вычисления и прокомментируйте результат.

В логистической регрессии вероятность принадлежности объекта классу 1 рассчитывается по формуле:

$$P = \frac{1}{1 + e^{-z}}$$

где $z=2.3$. Найдите вероятность (P), округлите до сотых.

Правильный ответ: 0.91. Это означает, что объект с высокой вероятностью принадлежит классу 1, и модель уверена в своем предсказании.

Компетенции (индикаторы): ОПК-6 (ОПК-6.2)

Задания открытого типа с развернутым ответом

1. Прочтайте текст задания. Продумайте логику и полноту ответа. Запишите развернутый и обоснованный ответ. В вашей компании возникла необходимость организовать удаленный доступ к внутренним ресурсам для сотрудников, работающих из дома. Опишите пошагово, как бы вы настроили VPN-сервер на базе Windows Server, начиная с установки необходимых ролей и заканчивая настройкой клиентских подключений. Какие протоколы туннелирования вы бы использовали и почему? Какие меры безопасности вы бы предприняли для защиты VPN-соединения?

Время выполнения -15 мин.

Ожидаемый результат:

Установка роли "Удаленный доступ" на Windows Server. Настройка VPN-сервера с использованием протоколов PPTP, L2TP/IPsec или SSTP (с обоснованием выбора). Настройка правил брандмауэра для разрешения VPN-трафика. Создание учетных записей пользователей с правами удаленного доступа. Настройка параметров шифрования и аутентификации. Установка и настройка VPN-клиента на компьютерах сотрудников. Рекомендации по использованию надежных паролей, многофакторной аутентификации и регулярному обновлению программного обеспечения.

Критерии оценивания: Полнота и корректность описания процесса настройки VPN-сервера, обоснованность выбора протоколов туннелирования, учет аспектов безопасности.

Компетенции (индикаторы): ОПК-6 (ОПК-6.2)

2. Прочтайте текст задания. Продумайте логику и полноту ответа. Запишите развернутый и обоснованный ответ.

Компания анализирует продажи товаров в трех регионах, чтобы определить динамику спроса. Вам даны данные о средних ежемесячных продажах за последние шесть месяцев:

- Регион А: 120, 130, 140, 150, 155, 165
- Регион В: 80, 95, 110, 120, 130, 140
- Регион С: 150, 160, 175, 185, 200, 220

Для прогнозирования продаж используется метод скользящего среднего с окном 3. Вам необходимо рассчитать значения скользящего среднего для каждого региона и найти разницу между последним фактическим значением и предсказанным скользящим средним. Затем определите, в каком регионе продажи растут наиболее равномерно. В выводе укажите, как можно использовать полученные данные для маркетинговых решений.

Время выполнения - 15 мин.

Ожидаемый результат:

- Верно рассчитанные значения скользящего среднего.
- Корректные вычисления отклонений фактических значений от предсказанных.
- Обоснованный вывод о стабильности роста продаж.

Правильное решение:

Шаг 1. Вычисляем скользящее среднее с окном 3

Для региона А:

- $(120 + 130 + 140) / 3 = \mathbf{130}$
- $(130 + 140 + 150) / 3 = \mathbf{140}$
- $(140 + 150 + 155) / 3 = \mathbf{148.3}$
- $(150 + 155 + 165) / 3 = \mathbf{156.7}$

Для региона В:

- $(80 + 95 + 110) / 3 = \mathbf{95}$
- $(95 + 110 + 120) / 3 = \mathbf{108.3}$
- $(110 + 120 + 130) / 3 = \mathbf{120}$

- $(120 + 130 + 140) / 3 = \mathbf{130}$

Для региона С:

- $(150 + 160 + 175) / 3 = \mathbf{161.7}$
- $(160 + 175 + 185) / 3 = \mathbf{173.3}$
- $(175 + 185 + 200) / 3 = \mathbf{186.7}$
- $(185 + 200 + 220) / 3 = \mathbf{201.7}$

Шаг 2. Рассчитываем отклонения фактических значений от предсказанных

- Регион А: $165 - 156.7 = \mathbf{8.3}$
- Регион В: $140 - 130 = \mathbf{10}$
- Регион С: $220 - 201.7 = \mathbf{18.3}$

Шаг 3. Вывод

Продажи в регионе А наиболее стабильны, так как отклонение минимальное. В регионе С отклонение максимальное, что может говорить о резком росте спроса.

Практическое применение:

Стабильные продажи позволяют более точно прогнозировать будущий спрос. В регионе С, где наблюдается резкий рост, может потребоваться дополнительный анализ факторов (сезонность, акции, конкурентная активность) и адаптация маркетинговой стратегии.

Критерии оценивания:

- Корректность вычислений.
- Логичность интерпретации полученных данных.
- Грамотность вывода о стабильности продаж.

Компетенции (индикаторы): ОПК-6 (ОПК-6.2)

Экспертное заключение

Представленный фонд оценочных средств (далее – ФОС) по дисциплине «Численные методы в экономических расчетах» соответствует требованиям ФГОС ВО.

Предлагаемые формы и средства текущего и промежуточного контроля адекватны целям и задачам реализации основной профессиональной образовательной программы по направлению подготовки 38.03.05 Бизнес-информатика.

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 38.03.05 Бизнес-информатика.

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

Ольга 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)