

**Комплект оценочных материалов по дисциплине
«Методы и технологии интеллектуального анализа данных»**

Задания закрытого типа

Задания закрытого типа на выбор правильного ответа

1. Выберите один правильный ответ.

Что является основной целью интеллектуального анализа данных (ИАД)?

- А) Создание баз данных
- Б) Извлечение скрытых закономерностей и знаний
- В) Разработка веб-приложений
- Г) Сжатие информации

Правильный ответ: Б

Компетенции (индикаторы): ПК-5

2. Выберите один правильный ответ.

Какой метод ИАД используется для разделения данных на заранее известные категории?

- А) Кластеризация
- Б) Классификация
- В) Ассоциативные правила
- Г) Метод главных компонент

Правильный ответ: Б

Компетенции (индикаторы): ПК-5

3. Выберите один правильный ответ.

Какой алгоритм используется для кластеризации данных?

- А) Логистическая регрессия
- Б) Метод опорных векторов (SVM)
- В) K-средних (k-means)
- Г) Нейронные сети

Правильный ответ: В

Компетенции (индикаторы): ПК-5

4. Выберите один правильный ответ.

Что такое «разведочный анализ данных» (exploratory data analysis, EDA)?

- А) Применение моделей машинного обучения к данным
- Б) Исследование и визуализация данных для выявления паттернов и аномалий
- В) Процесс оценки производительности моделей
- Г) Обработка данных перед обучением моделей

Правильный ответ: Б

Компетенции (индикаторы): ПК-5

Задания закрытого типа на установление соответствия

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

Методы ИАД	Категории
1) Деревья решений	А) Классификация
2) Метод K-means	Б) Уменьшение размерности
3) Метод главных компонент (PCA)	В) Кластеризация
4) Алгоритм Apriori	Г) Поиск ассоциативных правил

Правильный ответ:

1	2	3	4
А	В	Б	Г

Компетенции (индикаторы): ПК-5

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

Ключевые этапы ИАД	Описание
1) Подготовка данных	А) Проверка точности и корректности модели
2) Выбор модели	Б) Очистка, трансформация и нормализация данных
3) Обучение модели	В) Определение подходящего алгоритма для решения задачи
4) Оценка модели	Г) Использование данных для обучения модели

Правильный ответ:

1	2	3	4
Б	В	Г	А

Компетенции (индикаторы): ПК-5

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

Алгоритм	Применение
1) Линейная регрессия	А) Классификация данных с использованием гиперплоскости
2) Дерево принятия решений	Б) Классификация и регрессия с разветвленной структурой
3) DBSCAN	В) Выявление кластеров без заранее заданного количества
4) SVM	Г) Прогнозирование числовых значений

Правильный ответ:

1	2	3	4
Г	Б	В	А

Компетенции (индикаторы): ПК-5

Задания закрытого типа на установление правильной последовательности

1. Расположите этапы предварительной обработки данных в правильном порядке:

- А) Очистка данных
- Б) Нормализация данных
- В) Заполнение пропущенных значений
- Г) Кодирование категориальных признаков

Правильный ответ: А, В, Г, Б

Компетенции (индикаторы): ПК-5

2. Расположите этапы поиска ассоциативных правил методом Apriori в правильном порядке:

- А) Генерация частых наборов элементов
- Б) Определение кандидатов на частые наборы
- В) Подсчет поддержки наборов
- Г) Формирование правил ассоциации

Правильный ответ: Б, В, А, Г

Компетенции (индикаторы): ПК-5

3. Расположите этапы обработки больших данных в правильном порядке:

- А) Сбор данных
- Б) Обработка данных
- В) Анализ данных
- Г) Визуализация и интерпретация

Правильный ответ: А, Б, В, Г

Компетенции (индикаторы): ПК-5

Задания открытого типа

Задания открытого типа на дополнение

1. Напишите пропущенное словосочетание.

_____ – это направление в машинном обучении, посвящённое распознаванию, генерации и обработке устной и письменной человеческой речи.

Правильный ответ: Обработка естественного языка.

Компетенции (индикаторы): ПК-5

2. Напишите пропущенное словосочетание.

_____ – это область искусственного интеллекта, которая занимается разработкой алгоритмов и моделей, позволяющих компьютерам обучаться и делать прогнозы или принимать решения на основе имеющихся данных.

Правильный ответ: Машинное обучение

Компетенции (индикаторы): ПК-5

3. Напишите пропущенное словосочетание.

_____ – это один из способов машинного обучения, в ходе которого испытуемая система принудительно обучается с помощью примеров «стимул-реакция».

Правильный ответ: Обучение с учителем.

Компетенции (индикаторы): ПК-5

4. Напишите пропущенное словосочетание.

_____ – это алгоритм машинного обучения, используемый для решения задач классификации, регрессии и обнаружения выбросов.

Правильный ответ: Метод опорных векторов.

Компетенции (индикаторы): ПК-5

5. Напишите пропущенное словосочетание.

Метод главных компонент (PCA) применяется для уменьшения _____, сохраняя при этом наибольшее количество информации.

Правильный ответ: размерности данных.

Компетенции (индикаторы): ПК-5

6. Напишите пропущенное словосочетание.

_____ – это метод оценки качества модели, делящий данные на обучающую и тестовую выборки

Правильный ответ: Кросс-валидация

Компетенции (индикаторы): ПК-5

7. Напишите пропущенное словосочетание.

_____ – это программные библиотеки и инструменты для разработки и обучения моделей (например, TensorFlow, PyTorch, Scikit-learn).

Правильный ответ: Фреймворки машинного обучения

Компетенции (индикаторы): ПК-5

8. Напишите пропущенное словосочетание

_____ – это математические модели прогнозирования, которые стремятся найти зависимость будущего значения от прошлого внутри самого процесса и на этой зависимости вычислить прогноз.

Правильный ответ: Модели временных рядов

Компетенции (индикаторы): ПК-5

Задания открытого типа с кратким свободным ответом

1. Напишите пропущенное слово (словосочетание)

_____ – это алгоритм кластеризации данных, который используется для объединения точек в группы, основываясь на плотности их расположения.

Правильный ответ: DBSCAN / Density-Based Spatial Clustering of Applications with Noise / алгоритм DBSCAN / алгоритм Density-Based Spatial Clustering of Applications with Noise / метод DBSCAN / метод Density-Based Spatial Clustering of Applications with Noise /

Компетенции (индикаторы): ПК-5.

2. Напишите пропущенное словосочетание.

_____ – это технология многомерного статистического анализа, используемая для сокращения размерности пространства признаков с минимальной потерей полезной информации.

Правильный ответ: PCA / Principal Component Analysis / Метод главных компонент / Алгоритм PCA / Алгоритм Principal Component Analysis / Метод PCA / Метод Principal Component Analysis /

Компетенции (индикаторы): ПК-5

3. Напишите пропущенное словосочетание.

_____ – это метод машинного обучения, который объединяет несколько базовых моделей для создания одной оптимальной прогностической модели.

Правильный ответ: Ансамблирование / Ensembling / Метод ансамблирования / Метод Ensembling /.

Компетенции (индикаторы): ПК-5

4. Напишите пропущенное словосочетание

_____ – это процесс устранения пропусков в данных, который является ключевой задачей в машинном обучении и анализе данных.

Правильный ответ: Обработка пропущенных данных / Missing Data Handling

Компетенции (индикаторы): ПК-5

Задания открытого типа с развернутым ответом

1. Разобрать код на Python: распознать какой алгоритм интеллектуального анализа приведен в коде на Python, определить его назначение, дать пояснения, перечислить выполняемые действия и использованные возможности.

```
from sklearn.cluster import KMeans
data = [[1, 2], [3, 4], [5, 6], [8, 9]]
model = KMeans(n_clusters=2)
model.fit(data)
print(model.labels_)
```

Время выполнения 30 минут.

Ожидаемый результат:

Пошаговое объяснение кода:

1) Импорт библиотеки:

```
from sklearn.cluster import KMeans
```

– здесь импортируется класс KMeans из модуля sklearn.cluster, который реализует алгоритм К-средних.

2) Определение данных:

```
data = [[1, 2], [3, 4], [5, 6], [8, 9]]
```

– создаётся список из четырёх двумерных точек, которые будут использоваться для кластеризации.

3) Создание модели KMeans:

```
model = KMeans(n_clusters=2)
```

– инициализируется модель KMeans с параметром n_clusters=2, что означает разделение данных на два кластера.

4) Обучение модели:

```
model.fit(data)
```

– модель обучается на предоставленных данных, определяя центроиды кластеров и распределяя точки по кластерам.

5) Вывод меток кластеров:

```
print(model.labels_)
```

– выводятся метки кластеров для каждой точки данных. Метки представляют собой индексы кластеров, к которым отнесена каждая точка.

6) Результат выполнения кода:

```
[1 1 0 0]
```

– это означает, что точки [1, 2] и [3, 4] отнесены к кластеру 1, а точки [5, 6] и [8, 9] — к кластеру 0.

Представленный код на Python демонстрирует применение алгоритма кластеризации К-средних (K-Means) из библиотеки scikit-learn. Этот алгоритм относится к методам неподконтрольного обучения и используется для разделения набора данных на к кластеров на основе схожести между объектами.

Критерии оценивания:

- точность определения алгоритма (в коде используется алгоритм К-средних (K-Means)).
 - понимание процесса работы алгоритма;
 - корректная интерпретация результатов;
 - логичность и последовательность изложения;
 - использование корректной терминологии.

Компетенции (индикаторы): ПК-5

2. Прочитайте текст задания. Продумайте логику и полноту ответа. Запишите ответ, используя четкие компактные формулировки.

Вопрос: Какие ключевые проблемы возникают при выборе метода нормализации данных для анализа временных рядов в маркетинге, и как они влияют на точность прогнозов?

Время выполнения 30 минут.

Ожидаемый результат:

При выборе метода нормализации данных в анализе временных рядов маркетинговых показателей необходимо учитывать ряд проблем, которые могут значительно повлиять на точность прогнозов.

1) Сезонность и тренды:

- маркетинговые данные часто имеют сезонные колебания (например, рост продаж перед праздниками) и долгосрочные тренды;
- если нормализация не учитывает эти факторы, важные закономерности могут быть искажены;
- пример: применение минимально-максимальной нормализации без предварительного удаления сезонности может скрыть важные пики и спады.

2) Выбросы и аномалии:

- в данных могут встречаться резкие всплески (например, эффект рекламной кампании или флешмобов);
- стандартные методы, такие как min-max scaling, чувствительны к выбросам и могут неправильно интерпретировать масштабы изменений;
- пример: если промо-акция резко увеличила продажи, нормализация может создать иллюзию общей тенденции роста. В этом случае лучше использовать медианную нормализацию или логарифмическую трансформацию.

3) Автокорреляция и временные зависимости:

- временные ряды маркетинга обладают зависимостью от предыдущих значений;
- применение методов, не учитывающих автокорреляцию, может исказить характеристики данных и снизить точность прогнозов;
- пример: если не учесть временные лаги или скользящее среднее, нормализация может привести к потере информации о закономерностях изменения показателей.

Вывод:

Выбор метода нормализации в маркетинговом анализе временных рядов требует индивидуального подхода. Необходимо тестировать различные методы и оценивать их влияние на точность прогнозов, чтобы выбрать оптимальный вариант.

Критерии оценивания:

- корректность определения ключевых проблем (минимум три основные проблемы, связанные с нормализацией данных);
- объяснение того, как каждая проблема влияет на точность прогнозов;
- приведены практические примеры ситуаций, в которых ошибки нормализации могут исказить результаты анализа;
- структурированность ответа: изложение логично организовано, представлены четкие аргументы и вывод;
- ясность и лаконичность: используются четкие и корректные формулировки.
- ответ может варьироваться по деталям, допустимы разные формулировки, однако общий смысл должен быть верным.

Компетенции (индикаторы): ПК-5

Экспертное заключение

Представленный комплект оценочных материалов по дисциплине «Методы и технологии интеллектуального анализа данных» соответствует требованиям ФГОС ВО.

Предлагаемые оценочные материалы адекватны целям и задачам реализации основной профессиональной образовательной программы по направлению подготовки 09.04.03 Прикладная информатика.

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

Разработанные и представленные для экспертизы оценочные материалы рекомендуются к использованию в процессе подготовки обучающихся по указанному направлению.

Председатель учебно-методической комиссии
института компьютерных систем
и информационных технологий



Ветрова Н.Н.

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)
1.	Дополнен комплектом оценочных материалов	протокол заседания кафедры компьютерных систем и сетей № <u>8</u> от <u>10.03.2025</u>	 С.В. Попов