

**Федеральное государственное бюджетное учреждение науки
Институт машиноведения им. А.А. Благонравова
Российской академии наук
(ИМАШ РАН)**

УТВЕРЖДАЮ
Заместитель директора по вопросам
подготовки кадров высшей
квалификации ИМАШ РАН
« 01 » _____ 2026 года
к.ф.-м.н. Жданов Э.Р.
М.П.



**Комплект оценочных материалов
по дисциплине «Компьютерные информационно-аналитические системы»**

Направление подготовки:	02.04.03 Математическое обеспечение и администрирование информационных систем
Профиль подготовки:	Искусственный интеллект и интеллектуальная обработка данных
Квалификация:	Магистр
Форма обучения	Очная

Содержание

1. Общие положения.....	3
2. Результаты освоения дисциплины, подлежащие проверке.....	3
3. Контроль и оценка освоения учебной дисциплины.....	4
4. Оценочные материалы для проведения текущего контроля успеваемости.....	4
5. Оценочные материалы для проведения промежуточной аттестации.....	7

1. Общие положения

Комплект оценочных материалов (далее - ОМ) предназначен для контроля и оценки образовательных достижений обучающихся, освоивших программу дисциплины «Компьютерные информационно-аналитические системы».

ОМ включают контрольные материалы для проведения текущего контроля успеваемости и промежуточной аттестации в форме зачета.

ОМ разработаны на основании:

- основной профессиональной образовательной программы по направлению подготовки 02.04.03 Математическое обеспечение и администрирование информационных систем;
- рабочей программы дисциплины «Компьютерные информационно-аналитические системы».

2. Результаты освоения дисциплины, подлежащие проверке

В результате освоения дисциплины обучающийся должен обладать следующими компетенциями

ПК-1	Создание информационных технологий нового поколения
ПК-1.2	Управление разработкой продуктов, услуг и решений на основе больших данных
ПК-1.3	Разработка и внедрение новых методов и технологий исследования больших данных

Код индикатора компетенции	Планируемые результаты обучения по дисциплине
ПК-1.2	<i>Знать:</i> - архитектуру современных информационно-аналитических систем и их роль в управлении продуктами; - методологии проектирования хранилищ данных (DWH) и процессов ETL/ELT; - принципы управления качеством данных в больших системах; <i>Уметь:</i> - формулировать требования к сбору и интеграции данных из гетерогенных источников; - оценивать технологические стеки для реализации решений на основе больших данных; - планировать этапы разработки аналитического продукта от сбора требований до внедрения дашбордов. <i>Иметь практический опыт:</i> - проектирования концептуальной и логической моделей хранилища данных для решения конкретной бизнес-задачи; - разработки прототипа ETL-конвейера для загрузки и первичной очистки данных; - создания интерактивного отчета (дашборда) для визуализации ключевых показателей продукта.
ПК-1.3	<i>Знать:</i> - современные методы машинного обучения и их применимость к задачам исследования больших данных; - алгоритмы предобработки и трансформации больших массивов информации; - технологии распределенных вычислений (на примере SQL-подобных запросов к колоночным БД); <i>Уметь:</i> - применять алгоритмы машинного обучения для решения задач классификации, регрессии и кластеризации в рамках ИАС; - разрабатывать скрипты на Python для автоматизации исследовательского

	анализа данных (EDA); - интегрировать разработанные аналитические модели в контур информационно-аналитической системы. <i>Иметь практический опыт:</i> - решения прикладной задачи Data Mining с полным циклом: от постановки задачи до оценки модели; - реализации нестандартного метода предобработки или анализа данных с использованием библиотек Pandas/NumPy; - внедрения разработанной модели машинного обучения в BI-систему для генерации прогнозов в реальном времени.
--	--

3. Контроль и оценка освоения учебной дисциплины

Элемент учебной дисциплины	Текущий контроль		Промежуточная аттестация	
	Форма контроля	Проверяемые идк	Форма контроля	Проверяемые идк
Тема 1. Введение в информационно-аналитические системы (ИАС)	Отчет по практикуму по решению задач	ПК-1.2 ПК-1.3		
Тема 1. Введение в информационно-аналитические системы (ИАС)	Отчет по лабораторному практикуму	ПК-1.2 ПК-1.3		
Тема 2. Проектирование и реализация хранилищ данных (DWH)	Отчет по практикуму по решению задач	ПК-1.2 ПК-1.3		
Тема 2. Проектирование и реализация хранилищ данных (DWH)	Отчет по лабораторному практикуму	ПК-1.2 ПК-1.3		
Тема 3. Методы и алгоритмы интеллектуального анализа данных	Отчет по практикуму по решению задач	ПК-1.2 ПК-1.3		
Тема 3. Методы и алгоритмы интеллектуального анализа данных	Отчет по лабораторному практикуму	ПК-1.2 ПК-1.3		
Тема 4. Визуализация данных и поддержка принятия решений	Отчет по практикуму по решению задач	ПК-1.2 ПК-1.3		
Тема 4. Визуализация данных и поддержка принятия решений	Отчет по лабораторному практикуму	ПК-1.2 ПК-1.3		
			Зачет	ПК-1.2 ПК-1.3

4. Оценочные материалы для проведения текущего контроля успеваемости

Показатели и критерии оценивания компетенций на различных этапах их формирования, описание шкал оценивания

Практикум по решению задач
10-9 – работа и отчет выполнены в срок, самостоятельно, правильно определены соответствующие спецификации, использована требуемая информация, правильно выполнены

требуемые расчеты, сделаны необходимые выводы, хорошо аргументированы, даны исчерпывающие ответы на все поставленные вопросы; 8-6 – работа и отчет выполнены в срок, самостоятельно, правильно определены соответствующие спецификации, использована требуемая информация, правильно выполнены требуемые расчеты, необходимые выводы сделаны частично, хорошо аргументированы, даны ответы на все поставленные вопросы; 5-3 – работа и отчет выполнены в срок, в основном самостоятельно, выводы сделаны частично, слабо аргументированы, даны ответы не на все вопросы; 2-1 – обучающийся подготовил работу и отчет несамостоятельно и/или не завершил в срок, описание содержит ошибки, выводы и ответы на вопросы отсутствуют.
Лабораторный практикум
15-13 – работа выполнена в срок, самостоятельно, правильно поняты и использованы соответствующие формулы, правильно определены соответствующие спецификации, использована требуемая информация, правильно выполнены требуемые расчеты, правильно выбраны совместимые комплектующие, сделаны необходимые выводы, хорошо аргументированы, даны исчерпывающие ответы на все поставленные вопросы; 12-9 – работа выполнена в срок, самостоятельно, правильно поняты и использованы соответствующие формулы, правильно определены соответствующие спецификации, использована требуемая информация, правильно выполнены требуемые расчеты, правильно выбраны совместимые комплектующие, необходимые выводы сделаны частично, хорошо аргументированы, даны ответы на все поставленные вопросы; 8-5 – работа выполнена в срок, в основном самостоятельно, использованы соответствующие формулы; определены соответствующие спецификации, имеются ошибки в расчетах; выбраны совместимые комплектующие необходимые, выводы сделаны частично, слабо аргументированы, даны ответы не на все вопросы; 4-1 – обучающийся подготовил работу несамостоятельно или не завершил в срок, описание спецификации содержит незначительные ошибки, выводы и ответы на вопросы отсутствуют.

Порядок начисления баллов

Если студент набрал достаточное количество баллов ТКУ по дисциплине (более 70 баллов) в течение семестра, он может отказаться от сдачи зачета, и результат будет выставлен автоматически по его согласию.

Типовые контрольные задания и иные материалы, необходимые для процедуры оценивания знаний, умений, навыков и(или) опыта деятельности

Типовые задания для лабораторных практикумов

Тема 1. Введение в информационно-аналитические системы (ИАС)

Лабораторное занятие: «Сборка макета простейшей ИАС».

Задание:

Используя любой доступный инструмент (например, Python + SQLite или PostgreSQL), реализовать упрощенную модель ИАС. Создать базу данных (DWH) с несколькими таблицами-источниками (например, sales, customers, products).

Реализовать процесс ETL: написать скрипт, который выгружает данные из CSV-файлов, очищает их (удаляет дубликаты, заполняет пропуски) и загружает в целевые таблицы хранилища.

Построить витрину данных (Data Mart) в виде денормализованной таблицы, агрегирующей продажи по месяцам и категориям товаров.

Написать SQL-запрос, имитирующий работу OLAP-куба (например, рассчитать динамику продаж по годам и регионам).

Тема 2. Проектирование и реализация DWH

Лабораторное занятие: «Проектирование схемы "Звезда" и загрузка данных».

Задание:

На основе бизнес-требований (например, «анализировать выручку интернет-магазина») спроектировать схему «звезда». Определить центральную таблицу фактов (например, fact_sales) и окружающие ее таблицы измерений (dim_date, dim_customer, dim_product).

Создать DDL-скрипты для генерации этих таблиц в реляционной СУБД (PostgreSQL). Особое внимание уделить типам данных и первичным/внешним ключам.

Разработать и выполнить SQL-скрипт для наполнения таблиц измерений справочными данными.

(Продвинутый уровень) Сравнить скорость выполнения простого агрегатного запроса (например, SUM (sales_amount) по всем данным) на этой схеме в PostgreSQL и в колоночной СУБД ClickHouse.

Тема 3. Методы и алгоритмы интеллектуального анализа данных

Лабораторное занятие: «Решение задачи классификации».

Задание: Используя библиотеку scikit-learn в Python, решить задачу бинарной классификации (например, предсказание оттока клиентов). Загрузить датасет, провести разведочный анализ данных.

Выполнить предобработку: кодирование категориальных признаков, масштабирование числовых.

Разделить данные на обучающую и тестовую выборки.

Обучить две модели: логистическую регрессию и случайный лес.

Оценить качество моделей на тестовой выборке с помощью метрик accuracy, precision, recall, F1-score и матрицы ошибок (confusion matrix).

Сделать вывод, какая модель лучше подходит для данной задачи и почему.

Тема 4. Визуализация данных и поддержка принятия решений

Лабораторное занятие: «Разработка интерактивного дашборда».

Задание:

С использованием BI-инструмента (Power BI, Tableau или Apache Superset) создать дашборд на основе готового датасета (например, данные о COVID-19, продажах или демографии).

Подключиться к источнику данных.

Подготовить модель данных: создать необходимые связи, вычисляемые поля (например, доля от общего числа).

Спроектировать и разместить на холсте ключевые визуальные элементы: Линейная диаграмма динамики основного показателя во времени.

Карта мира/страны с цветовой дифференциацией значений.

Древовидная карта (treemap) для отображения долей категорий.

Фильтры для взаимодействия пользователя с дашбордом.

Оформить дашборд, добавить заголовки и пояснения.

Типовые задания для практикумов по решению задач

Тема 1. Введение в информационно-аналитические системы (ИАС)

Практикум по решению задач: «Анализ концепции "V" больших данных».

Задание:

Дано описание трех различных информационных потоков. Для каждого потока необходимо определить, какие из характеристик Big Data (Volume, Velocity, Variety, Veracity) являются наиболее критичными, а какие — второстепенными, и обосновать свой выбор.

Пример: Поток данных от датчиков температуры на промышленном оборудовании (высокая Velocity, умеренная Variety, высокая Veracity).

Пример: Сбор пользовательских отзывов о продукте из соцсетей (высокая Variety,

низкая Veracity, средний Volume).

Тема 2. Проектирование и реализация DWH

Практикум по решению задач: «Нормализация и денормализация».

Задание:

Дана таблица в первой нормальной форме (1НФ), содержащая избыточность и аномалии обновления. Необходимо: Привести таблицу к третьей нормальной форме (3НФ), выделив отдельные сущности и связи. Нарисовать ER-диаграмму полученной нормализованной схемы.

Объяснить, какие аномалии были устранены.

Предложить вариант денормализации одной из полученных таблиц для оптимизации конкретного аналитического запроса (например, для ускорения отчетов), указав, какую проблему производительности это решает и какой риск избыточности создает.

Тема 3. Методы и алгоритмы интеллектуального анализа данных

Практикум по решению задач: «Выбор метода Data Mining».

Задание:

Описаны различные бизнес-ситуации. Для каждой нужно выбрать наиболее подходящий метод анализа и объяснить свой выбор.

Ситуация 1: Банк хочет разделить всех заемщиков на группы по уровню кредитного риска. Ответ: Кластеризация (например, k-means) или классификация (если есть исторические данные с метками «дефолт»).

Ситуация 2: Ритейлер хочет понять, какие товары часто покупают вместе, чтобы оптимизировать выкладку на полках. Ответ: поиск ассоциативных правил (Apriori).

Ситуация 3: Агентство недвижимости хочет спрогнозировать стоимость квартиры на основе ее параметров (площадь, район, этаж). Ответ: регрессия.

Ситуация 4: Почтовая служба хочет автоматически определять, является ли входящее письмо спамом. Ответ: классификация.

Тема 4. Визуализация данных и поддержка принятия решений

Практикум по решению задач: «Критика визуализаций».

Задание:

Студентам предоставляются скриншоты нескольких графиков и диаграмм (как хорошо сделанных, так и содержащих ошибки). Задача — проанализировать каждый пример и ответить на вопросы: Какую идею пытается донести эта визуализация?

Соблюдены ли базовые принципы эффективной визуализации (соответствие типа графика данным, отсутствие «визуального мусора», корректная шкала осей)?

Какие выводы можно сделать на основе этого графика? Не вводят ли искаженные данные в заблуждение?

Как бы вы улучшили эту визуализацию, чтобы она стала более наглядной и честной?

5. Оценочные материалы для проведения промежуточной аттестации

Промежуточная аттестация по дисциплине проводится в форме зачета.

Зачет представляет собой выполнение обучающимся заданий билета, включающего в себя:

Задание №1 – теоретический вопрос на знание базовых понятий предметной области дисциплины, а также позволяющий оценить степень владения обучающегося принципами предметной области дисциплины, понимание их особенностей и взаимосвязи между ними.

Задание №2 – задание на анализ ситуации из предметной области дисциплины и выявление способности обучающегося выбирать и применять соответствующие принципы и методы решения практических проблем, близких к профессиональной деятельности.

Задания №3 – задания на проверку умений и навыков, полученных в результате освоения дисциплины.

Выполнение обучающимся заданий билета оценивается по следующей балльной шкале:

Задание 1: 0-30 баллов

Задание 2: 0-30 баллов

Задание 3: 0-40 баллов

«Зачтено»

– 90-100 – ответ правильный, логически выстроен, использована профессиональная терминология. Задания решены правильно. Обучающийся правильно интерпретирует полученный результат.

– 70 -89 – ответ в целом правильный, логически выстроен, использована профессиональная терминология. Ход решения заданий правильный, ответ неверный. Обучающийся в целом правильно интерпретирует полученный результат.

– 50 - 69 – ответ в основном правильный, логически выстроен, использована профессиональная терминология. Задание решено частично.

«Не зачтено»

– менее 50 – ответы на теоретическую часть неправильные или неполные. Задания не решены.

Типовые задания для проведения промежуточной аттестации обучающихся

Задания 1 типа

1. Дайте определение информационно-аналитической системы (ИАС) и опишите её ключевые отличия от традиционных информационных систем.

2. Какова роль ИАС в цифровой экономике и в процессе цифровой трансформации бизнеса?

3. Перечислите основные компоненты типовой архитектуры ИАС и опишите их назначение (источники данных, подсистема сбора, DWH, OLAP, витрины данных).

4. Раскройте концепцию «больших данных» и охарактеризуйте каждую из характеристик V: Volume, Velocity, Variety, Veracity.

5. В чем заключается суть моделирования измерений (Dimensional Modeling) и какова его роль при проектировании хранилищ данных?

6. Сравните схемы «звезда» и «снежинка»: назовите преимущества и недостатки каждой из них.

7. Объясните разницу между нормализацией и денормализацией данных в контексте проектирования DWH. В каких случаях применяется каждая из стратегий?

8. Сравните архитектурные особенности и области применения реляционных СУБД (PostgreSQL, Oracle) и колоночных СУБД (ClickHouse) для реализации хранилищ данных.

9. Опишите этапы процесса ETL. Какие задачи решаются на каждом из этапов (извлечение, преобразование, загрузка)?

10. В чем заключается принципиальное отличие подхода ELT от ETL и как он связан с современными технологиями обработки больших данных?

11. Классифицируйте основные задачи Data Mining: приведите примеры бизнес-задач для классификации, регрессии, кластеризации и поиска ассоциативных правил.

12. Опишите принцип работы алгоритма линейной регрессии. В чем заключается задача логистической регрессии и как она отличается от линейной?

13. Как устроен алгоритм дерева решений? Каким образом случайный лес (Random Forest) повышает качество предсказаний по сравнению с одним деревом?

14. Опишите алгоритм кластеризации k-means. Как определяется оптимальное число кластеров и каковы основные метрики оценки качества кластеризации?

15. Какие методы используются для оценки качества моделей машинного обучения (метрики)? Приведите примеры метрик для задач классификации и регрессии.

16. Сформулируйте основные принципы эффективной визуализации данных (например, по Тафти). Почему важно избегать визуального мусора на графиках?

17. Сравните функциональные возможности и сценарии использования BI-инструментов: Tableau, Power BI, Apache Superset.
18. Опишите процесс разработки интерактивного дашборда: от постановки бизнес-задачи до выбора визуализаций и реализации.
19. Как осуществляется интеграция аналитических моделей (Data Mining) в бизнес-процессы предприятия? Приведите примеры использования результатов анализа.
20. Каким образом системы поддержки принятия решений (СППР) используют данные из хранилищ и результаты интеллектуального анализа для помощи руководителям?

Задания 2 типа

1. Ритейлер собирает данные о продажах из кассовых аппаратов, информацию о клиентах из программы лояльности и данные о посещаемости сайта. Опишите, как вы спроектируете архитектуру ИАС для этого бизнеса, какие источники данных будут ключевыми и как организовать их интеграцию.
2. В крупной логистической компании данные о маршрутах и грузах поступают в режиме реального времени (высокая Velocity). Предложите технологический стек для сбора, хранения и оперативной аналитики этих данных, обоснуйте выбор.
3. У вас есть задача построить хранилище данных для интернет-магазина. Опишите, как вы примените методологию моделирования измерений: какие факты и измерения выделите, какую схему («звезда» или «снежинка») выберете для справочника «Товары» и почему.
4. Руководство банка требует построить отчет по кредитным заявкам за 5 лет. Данные хранятся в нескольких разрозненных OLTP-системах. Опишите последовательность действий в рамках ETL-процесса для подготовки этих данных к загрузке в DWH.
5. В компании принято решение использовать колоночную СУБД (например, ClickHouse) вместо реляционной (PostgreSQL) для построения аналитических отчетов. Сформулируйте аргументы для защиты этого решения перед техническим директором.
6. В процессе загрузки данных в хранилище вы обнаружили, что 15% записей о транзакциях содержат ошибки в коде валюты. Опишите шаги по обработке (трансформации) этих данных в рамках ETL-процесса.
7. Маркетологи просят создать витрину данных для анализа эффективности рекламных кампаний. Опишите, какие данные из центрального DWH вы включите в эту витрину и как обеспечите её актуальность.
8. Опишите практическую ситуацию, когда применение подхода ELT (вместо классического ETL) является более предпочтительным. Какую роль здесь играют облачные хранилища данных?
9. Отдел кадров просит вас разработать модель для прогнозирования увольнений сотрудников. Какую задачу Data Mining (классификация или регрессия) вы выберете и почему? Какие признаки (данные) будут наиболее важны?
10. Для задачи кредитного скоринга (оценка вероятности невозврата кредита) вам нужно выбрать алгоритм. Сравните применимость логистической регрессии и случайного леса. Какой алгоритм вы выберете, если приоритетом является интерпретируемость результата для регулятора?
11. Вам необходимо сегментировать 2 миллиона абонентов сотовой связи на группы со схожим поведением. Какой алгоритм вы выберете? Как вы определите оптимальное количество кластеров (k) и как оцените качество полученной сегментации?
12. Аптечная сеть хочет выявить товары, которые часто покупают вместе (анализ рыночной корзины). Какую задачу Data Mining вы решите? Какой алгоритм примените и как будете оценивать значимость найденных правил?
13. Финансовый аналитик жалуется, что график динамики прибыли за год выглядит как «спагетти» из-за слишком большого количества линий категорий товаров. Примените принципы эффективной визуализации для переработки этого графика.

14. Руководителю отдела продаж нужен инструмент для мониторинга выполнения плана в реальном времени. Опишите структуру интерактивного дашборда: какие виджеты (графики, таблицы) вы разместите на нем и почему.

15. В системе Power BI данные из источника обновляются слишком долго, что блокирует работу пользователей. Предложите архитектурное решение (например, использование промежуточных витрин или изменение процесса), чтобы ускорить обновление отчетов.

16. Вы внедрили модель машинного обучения для прогнозирования спроса на хлебобулочные изделия. Опишите пошагово, как результаты работы этой модели (прогнозы) будут интегрированы в ежедневный операционный процесс пекарни.

17. При построении отчета по продажам в Tableau данные из разных таблиц не объединяются корректно, что приводит к неверным суммам. Проанализируйте вероятную причину (проблему с ключами или связями) и предложите решение.

18. Бизнесу требуется система поддержки принятия решений (СППР) для выбора оптимального места открытия нового филиала банка на основе демографических данных города. Какие данные и методы анализа вы будете использовать?

19. Данные о поведении пользователей на сайте (клики, просмотры страниц) хранятся в сыром виде в виде логов объемом несколько терабайт в день (высокий Volume). Предложите архитектуру системы для их предварительной обработки перед загрузкой в основное DWH.

20. Вы разработали дашборд в Apache Superset для мониторинга ИТ-инфраструктуры. Пользователи говорят, что он перегружен информацией. Какие принципы проектирования интерфейсов вы примените для редизайна дашборда?

Задания 3 типа

1. Вам даны три бизнес-сущности: «Продажи», «Товар» и «Магазин». За 20 минут нарисуйте на бумаге или в любом редакторе схему «звезда» для витрины данных отдела продаж. Обозначьте центральную таблицу фактов и связанные с ней таблицы измерений. Укажите 2–3 ключевых атрибута для каждого измерения.

2. Дан список строк, представляющих цену товара: ["1000", "1 500 руб.", "2000.50", "ошибка", "3 000"]. Напишите короткий скрипт на Python, который очистит этот список, превратив его в список чисел с плавающей точкой (например, [1000.0, 1500.0, 2000.5, 3000.0]). Игнорируйте некорректные значения.

3. У вас есть одна плоская таблица orders_flat с полями order_id, customer_name, customer_city, product_name, product_price. За 25 минут нормализуйте эту структуру до 3-й нормальной формы (3NF), написав SQL-код для создания таблиц customers, products, orders и связей между ними.

4. Дана таблица sales (продажи) с полями: date, product_id, quantity, price. Напишите SQL-запрос, который выведет отчет по продажам за последний месяц в разрезе месяцев и товаров, показав общую выручку (quantity * price) для каждой пары «месяц-товар».

5. Опишите, какой алгоритм машинного обучения вы бы выбрали для решения двух задач: Предсказать, уйдет ли клиент из банка (да/нет).

Спрогнозировать сумму чека следующего заказа.

Объясните свой выбор в 3–4 предложениях для каждой задачи.

6. Опишите структуру дашборда для руководителя отдела продаж. Назовите 4 ключевых визуализации (например, график, карта, таблица, индикатор), которые вы бы разместили на нем, и объясните, какую метрику или инсайт показывает каждая из них.

7. У вас есть модель, предсказывающая отток клиентов. Матрица ошибок показывает: 80 истинно положительных, 10 ложноотрицательных, 5 ложноположительных и 105 истинно отрицательных результатов. Рассчитайте точность (Precision) и полноту (Recall) для класса «Ушедший клиент».

8. У вас есть график, показывающий динамику продаж 50 категорий товаров на одном полотне — линии пересекаются, ничего не видно. Предложите два разных способа

визуализации этих же данных, чтобы сделать отчет понятным.

9. Опишите кратко (в виде шагов), как работает процесс ELT при загрузке логов веб-сервера в облачное хранилище данных (например, в ClickHouse или Google BigQuery). В чем ключевое отличие этого процесса от ETL?