

АВТОНОМНАЯ НЕКОММЕРЧЕСКАЯ ОРГАНИЗАЦИЯ ВЫСШЕГО ОБРАЗОВАНИЯ
САМАРСКИЙ УНИВЕРСИТЕТ ГОСУДАРСТВЕННОГО УПРАВЛЕНИЯ
«МЕЖДУНАРОДНЫЙ ИНСТИТУТ РЫНКА»

УТВЕРЖДАЮ
Проректор по учебной работе и
качеству образования

_____ И. А. Долгова

15 апреля 2026 г.

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДИСЦИПЛИНЫ

АНАЛИЗ ДАННЫХ

Направление подготовки:	09.03.03 Прикладная информатика
Профиль подготовки:	Корпоративные информационные системы
Квалификация:	бакалавр
Форма обучения:	очно-заочная
Год начала подготовки:	2026

Самара
2026

1. ОЦЕНОЧНЫЕ СРЕДСТВА, СООТНЕСЁННЫЕ С ПЛАНИРУЕМЫМИ РЕЗУЛЬТАТАМИ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ

Код и наименование компетенции	Код и наименование индикатора достижения компетенции	Результаты обучения по дисциплине	Вид аттестации и оценочных средств
ПК-1. Способен анализировать и моделировать бизнес-процессы в организации заказчика, а также использовать типовые решения информационных систем для удовлетворения требований заказчика	<p>ПК-1.И-1. Анализирует бизнес-процессы в организации заказчика</p> <p>ПК-1.И-2. Моделирует бизнес-процессы с использованием современных методов моделирования</p> <p>ПК-1.И-3. Применяет типовые решения для внедрения информационных систем</p>	<p>ПК-1.И-1.3-1. Знает основные определения и характеристики бизнес-процессов</p> <p>ПК-1.И-1.У-1. Умеет применять основные методы анализа бизнес-процессов</p> <p>ПК-1.И-2.3-1. Знает основные методы моделирования бизнес-процессов</p> <p>ПК-1.И-2.У-1. Умеет использовать современные методы моделирования бизнес-процессов в организации заказчика</p> <p>ПК-1.И-3.3-1. Знает структуру типовых информационных систем</p> <p>ПК-1.И-3.У-1. Умеет применять типовые информационные системы для удовлетворения требований заказчика</p>	<p>Текущий контроль: устный опрос, лабораторная работа, промежуточный тест, реферат.</p> <p>Промежуточная аттестация: экзамен (вопросы на экзамене).</p>

2. ОЦЕНОЧНЫЕ СРЕДСТВА ДЛЯ ТЕКУЩЕГО КОНТРОЛЯ

2.1. Вопросы для подготовки к семинарским/практическим занятиям

Раздел 1. Введение в дисциплину «Анализ данных»

1. Дайте определение анализа данных. Какие этапы входят в анализ процесса?
2. Какие типы данных вы знаете? Приведите примеры количественных, качественных и категориальных данных.
3. Объясните разницу между описательной и прогнозной аналитикой.
4. Какие инструменты и языки программирования чаще всего используются при анализе данных?

Раздел 2. Обработка информации в Excel: корреляционный и дисперсионный анализ

1. Как представить коэффициент корреляции Пирсона в Excel? Продемонстрируйте на примере.
2. В чем разница между однофакторным и многофакторным дисперсионным анализом (ANOVA)?
3. Проведите дисперсионный анализ данных о продажах в трех регионах (пример данных прилагается). Сделайте выводы.
4. Какие графики можно использовать в Excel для визуализации пропорциональности? Постройте диаграмму рассеяния.

Раздел 3. Регрессионный анализ, расчет интегрального показателя

1. Что такое регрессионная модель? Опишите шаги линейной регрессии в Excel.
2. Как проверить оригинальность коэффициентов регрессии? Объясните причину.
3. Рассчитайте интегральный показатель на основе трех критериев (например, дохода, расходов, клиентской ценности) методом нормирования.
4. В чем недостатки использования интегральных показателей?

Раздел 4. Проверка статистических гипотез

1. Сформулируйте нулевые и альтернативные гипотезы для сравнения средних двух выборок.
2. Проведите t-тест для данных о времени доставки двух курьерских служб (пример данных прилагается).
3. Что показывает p-значение? Как его интерпретировать на уровне инновации 0,05?
4. Когда применяются критерии хи-квадрата? Проверьте гипотезу независимости (на основе данных).

Раздел 5. Интеллектуальный анализ данных, внедрение в Deductor

1. Какие задачи решаются с помощью интеллектуального анализа данных (Data Mining)?
2. Оставьте процесс подготовки данных в Deductor Studio (загрузка, очистка, трансформация).
3. Постройте дереворешения в Deductor для классификации клиентов (на основе данных об оттоке).
4. Как работает алгоритм ассоциативных правил (Априори)? Приведите пример использования.

Раздел 6. Инструменты обработки данных

1. Сравните возможности Excel и Python (Pandas) для очистки данных.
2. Напишите SQL-запрос для выбора данных по условию (например, продажа за 2023 год).
3. Как визуализировать финансовые данные с помощью библиотеки Matplotlib? Приведите пример кода.
4. Создайте сводную таблицу в Power BI и настройте интерактивную панель мониторинга.

Раздел 7. Группировка данных

1. Сгруппируйте данные о возрасте клиентов по интервалам (например, 18–25, 26–35 и т. д.) в Excel.

2. Какие методы группировки соответствуют качественным данным? Приведите пример.
3. Постройте гистограмму для сгруппированных данных о доходах и проанализируйте финансовые показатели.
4. Как группировка данных помогает в анализе поведения клиентов?

Раздел 8. Хранилище данных

1. Назовите основные компоненты структуры хранилища данных.
2. В чем разница между OLAP и OLTP?
3. Создайте схему «звезда» для хранения данных интернет-магазина (факты: продажи, измерения: товары, клиенты, время).
4. Опишите процесс ETL (Извлечение, Преобразование, Загрузка) по принципу передачи данных из CRM в хранилище.

Раздел 9. Кластерный анализ

1. Объясните алгоритм k-средних (k-средних). Как выбрать расчетное число кластеров?
2. Проведите кластерный анализ данных о покупательской активности (пример данных прилагается).
3. В чем преимущества иерархической кластеризации перед k-средними?
4. Как интерпретировать результаты кластеризации?

Критерии оценки работы на практическом занятии

Критерии	Максимальное количество баллов за занятие
Устный опрос, коллоквиум	
Основные теоретические положения по вопросу раскрыты. Имеются элементы обоснования выводов. Имеются элементы систематизации информации, факты применения профессиональной терминологии. Очевидно использование источников рекомендованной литературы.	5 баллов

2.2. Темы лабораторных работ

Раздел 1. Введение в дисциплину «Анализ данных»

Лабораторная работа №1. Знакомство с базовыми методами анализа данных.

Цель: Освоить этапы анализа данных, научиться классифицировать данные и использовать инструменты визуализации.

Вопросы для самопроверки:

1. Назовите основные типы данных. Приведите примеры для каждого.
2. Какие этапы включает процесс анализа данных?
3. В чем разница между структурированными и неструктурированными данными?
4. Создайте гистограмму для набора данных в Excel.
5. Как визуализация помогает интерпретировать результаты анализа?

Раздел 2. Обработка информации в Excel: корреляционный и дисперсионный анализ

Лабораторная работа №2. Корреляционный анализ в Excel.

Цель: Научиться рассчитывать коэффициенты корреляции и проводить дисперсионный анализ.

Вопросы для самопроверки:

1. В чем разница между коэффициентами Пирсона и Спирмена?
2. Проведите однофакторный дисперсионный анализ данных о продажах (датасет прилагается).
3. Постройте диаграмму размышления с устойчивым трендом.
4. Когда следует применять корреляционный анализ?
5. Объясните понятие «статистическая инновационность».

Раздел 3. Регрессионный анализ, расчет интегрального показателя

Лабораторная работа №3. Построение линейной регрессии.

Цель: Построить и интерпретировать регрессионную модель, вычисляющую интегральный показатель.

Вопросы для самопроверки:

1. Как проверить оригинальность коэффициентов регрессии?
2. Рассчитайте интегральный показатель на основе трех критериев (например, доход, расходы, рейтинг клиента).
3. Какие метрики качества регрессионной модели вы знаете?
4. В чем границы линейной регрессии?
5. Как нормировать данные для интегрального показателя?

Раздел 4. Проверка статистических гипотез

Лабораторная работа №4. Проверка гипотезы о среднем уровне.

Цель: Освоить применение t-теста и критериев хи-квадрат.

Вопросы для самопроверки:

1. Сформулируйте нулевую и альтернативную гипотезы для сравнения средних.
2. Проведите t-тест для двух выборок данных о времени доставки.
3. Интерпретируйте значение p при $\alpha=0,05$.
4. Как проверить гипотезу о нормальности распределения?
5. В чем суть критерия хи-квадрат?

Раздел 5. Интеллектуальный анализ данных, внедрение в Deductor
Лабораторная работа №5. Построение древесных решений в Deductor.
Цель: Освоить работу с алгоритмами классификации в Deductor Studio.

Вопросы для самопроверки:

1. Как подготовить данные для алгоритма решений дерева?
2. Подать модель для прогнозирования оттока клиентов.
3. Что такое информационный прирост (Получение информации)?
4. Как оценить точность модели?
5. В чем преимущества электронных решений перед линейной регрессией?

Раздел 6. Инструменты обработки данных

Лабораторная работа №6. Очистка данных в Python (Pandas).
Цель: Научиться обрабатывать и очищать данные с помощью библиотеки Pandas.

Вопросы для самопроверки:

1. Как удалить пропущенные значения в DataFrame?
2. Проведите агрегирование данных для расчета средних показателей по группам.
3. Напишите код для слияния двух таблиц по ключевому столбцу.
4. В чем отличия функций `map()` и `apply()`?
5. Постройте линейный график с использованием Matplotlib.

Раздел 7. Группировка данных

Лабораторная работа №7. Группировка и визуализация данных в Excel.
Цель: Создать категориальные группы данных и визуализировать их.

Вопросы для самопроверки:

1. Сгруппируйте данные по возрастным группам (например, 18–25, 26–30 лет).
2. Постройте круговую диаграмму для сгруппированных данных.
3. Как определить количество интервалов группировки?
4. Какие ошибки возможны при группировке?
5. Как группировка помогает при анализе основных датасетов?

Раздел 8. Хранилище данных

Лабораторная работа №8. Проектирование хранения данных.
Цель: Создать схему «звезда» для интернет-магазина.

Вопросы для самопроверки:

1. Назовите компоненты хранилища данных.
2. Спроектируйте таблицы фактов и измерений для анализа продаж.
3. Как реализовать ETL-процесс для загрузки данных из Excel в СУБД?
4. В чем разница между OLAP и OLTP?
5. Зачем нужны сборные таблицы в хранилище?

Раздел 9. Кластерный анализ

Лабораторная работа №9. Кластеризация методом k-средних.
Цель: Применить алгоритм k-means для сегментации данных.

Вопросы для самопроверки:

1. Как выбрать оптимальное количество кластеров?
2. Проведите кластеризацию данных о покупателях (датасет прилагается).

3. В чем отличие иерархической кластеризации от k-средств?
4. Как визуализировать результаты кластеризации?
5. Какие метрики оценки кластеров вы знаете?

Методические указания к проведению лабораторных работ

Цели лабораторных занятий:

1. Углубление и закрепление знания теоретического курса путем практического изучения в лабораторных условиях изложенных в лекциях методов и технологий;
2. Приобретение навыков в научном экспериментировании, анализе полученных результатов;
3. Формирование первичных навыков организации, планирования и проведения научных исследований.

Порядок проведения лабораторного занятия:

1. Вводная часть:
 - входной контроль подготовки обучаемого;
 - вводный инструктаж (знакомство обучающихся с содержанием предстоящей работы, краткий анализ теоретических положений и выводов, демонстрация подходов к выполнению отдельных операций, напоминание о технике безопасности, предупреждение о возможных ошибках).
2. Основная часть:
 - проведение обучаемым лабораторной работы;
 - текущее индивидуальное консультирование обучаемого;
3. Заключительная часть:
 - демонстрация результатов выполненного задания;
 - заключительный инструктаж (подведение итогов выполнения учебных задач, разбор допущенных ошибок и выявление их причин, сообщение результатов работы каждого обучаемого, объявление о том, что необходимо повторить к следующему занятию).

Особенности подготовки к проведению лабораторного занятия

Подготовка лабораторного занятия начинается с изучения теоретических положений, определения (уточнения) целей и задач данного занятия, времени, выделяемого обучаемым для подготовки.

В ходе подготовки к лабораторной работе необходимо пояснить проблематику, объем и содержание лабораторного занятия, определить, какие понятия, определения, теории могут быть иллюстрированы данным экспериментом, какие умения и навыки должны приобрести обучаемые в ходе занятия, какие знания углубить и расширить.

При этом преподавателю необходимо решить, на каком этапе обучения следует поставить задачу о подготовке к лабораторной работе, каким образом достигнуть активизации познавательной деятельности обучающихся. Задача на подготовку к лабораторной работе может быть поставлена на лекции, с таким временным расчетом, чтобы обучаемые смогли качественно подготовиться к ее проведению. Одновременно им выдаются учебно-методические материалы, иллюстрирующие круг вопросов, затрагиваемых в ходе выполнения лабораторного задания. Это могут быть методические указания по соответствующему курсу, презентации, ссылки на Интернет-источники и др. Эти материалы могут отражать учебные вопросы, краткие сведения по теории, программу выполнения работы, содержание отчета, вопросы для подготовки и литературу, рекомендуемую к изучению и т.д. В них также ставятся задачи, которые обучаемые должны решить при подготовке к работе, в процессе эксперимента и при обработке полученных результатов.

В ходе подготовки к лабораторной работе необходимо обратить внимание обучающегося на результат ее выполнения. Результат лабораторной работы должен быть четко сформулирован, приведены критерии его достижения, перечень материалов, его (результат) иллюстрирую-

щих – файлы, графики, скриншоты и т.д. Учащийся должен уметь формулировать основные выводы, опираясь на полученный на лабораторной работе результат.

В отдельных случаях, на лабораторном занятии может быть предусмотрена защита выполненной работы.

Шкала и критерии оценки лабораторной работы

Критерии	Баллы
Степень соответствия выполненного задания поставленным требованиям	25
Структурирование и комментирование лабораторной работы	25
Уникальность выполненной работы (отличие от работ коллег)	25
Ответы на контрольные вопросы	25

Лабораторная работа оценивается по 100 балльной шкале, баллы переводятся в оценки успеваемости следующим образом:

90 – 100 баллов – «отлично»;

70 – 89 баллов – «хорошо»;

50 – 69 баллов – «удовлетворительно»;

менее 50 баллов – «неудовлетворительно».

2.3. Темы рефератов

1. Основные этапы анализа данных: от сбора до визуализации.
2. Сравнение инструментов анализа данных: Excel, Python (Pandas), R.
3. Роль машинного обучения в прогнозной аналитике.
4. Классификация данных: методы и применение в бизнесе.
5. Анализ временных рядов: прогнозирование продаж с использованием ARIMA.
6. Основы работы с полученными данными: технологии Hadoop и Spark.
7. Этика анализа данных: проблемы приватности и анонимизации.
8. Использование нейронных сетей для измерения изображений.
9. Анализ социальных сетей: методы сбора и обработки данных.
10. Предварительный просмотр данных в Power BI: создание интерактивных дашбордов.
11. Применение кластерного анализа в сегментации клиентов.
12. Текстовая аналитика: обработка естественного языка (НЛП) с помощью Python.
13. А/В-тестирование: методы и интерпретация результатов.
14. Оптимизация бизнес-процессов с помощью анализа данных.
15. Анализ выживаемости (Survival Analysis) в медицинских исследованиях.
16. Методы уменьшения размерности данных: PCA и t-SNE.
17. Рекомендательные системы: коллаборативная фильтрация и алгоритмы.
18. Анализ геоданных: инструменты и применение в логистике.
19. Машинное обучение в финансах: прогнозирование курсов валют.
20. Автоматизация сети с помощью скриптов на Python.

Шкала и критерии оценки реферата (эссе)

Критерии	Показатели	Баллы
1. Новизна реферированного текста	<ul style="list-style-type: none"> – актуальность проблемы и темы; – новизна и самостоятельность в постановке проблемы, в формулировании нового аспекта выбранной для анализа проблемы; – наличие авторской позиции, самостоятельность суждений. 	20
2. Степень раскрытия сущности проблемы	<ul style="list-style-type: none"> – соответствие плана теме реферата (доклада); – соответствие содержания теме и плану; – полнота и глубина раскрытия основных понятий; – обоснованность способов и методов работы с материалом; – умение работать с литературой, систематизировать и структурировать материал; – умение обобщать, сопоставлять различные точки зрения по рассматриваемому вопросу, аргументировать основные положения и выводы. 	30
3. Обоснованность выбора источников	<ul style="list-style-type: none"> – круг, полнота использования литературных источников по теме; – привлечение новейших работ (журнальные публикации, материалы сборников научных трудов и т.д.). 	20

Критерии	Показатели	Баллы
4. Соблюдение требований к оформлению	<ul style="list-style-type: none"> – правильное оформление ссылок на используемую литературу; – грамотность и культура изложения; – владение терминологией и понятийным аппаратом; – соблюдение требований к объему работы; – культура оформления: выделение абзацев; – использование информационных технологий. 	15
5. Изложение	<ul style="list-style-type: none"> – отсутствие орфографических и синтаксических ошибок, стилистических погрешностей; – отсутствие опечаток, сокращений слов, кроме общепринятых; – литературный стиль. 	15

Реферат оценивается по 100 балльной шкале, баллы переводятся в оценки успеваемости следующим образом:

90 – 100 баллов – «отлично»;

70 – 89 баллов – «хорошо»;

50 – 69 баллов – «удовлетворительно»;

менее 50 баллов – «неудовлетворительно».

3. ОЦЕНОЧНЫЕ СРЕДСТВА ДЛЯ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

3.1. Банк контрольных заданий

1. Прочитайте текст, выберите правильные ответы и запишите аргументы, обосновывающие выбор ответов (ПК-1.1)

Какой инструмент НЕ относится к анализу данных?

- А. Power BI
- Б. Автокад
- В. Питон (Панды)
- Г. Таблица

Ответ:

Обоснование:

2. Прочитайте текст и установите соответствие (ПК-1.1)

Соотнести методы анализа с задачами:

Фактор ранжирования	Описание
1. Кластеризация	А. Прогнозирование числовых результатов (например, цены).
2. Regressia	Б. Группировка объектов по сходству.
3. Классификация	Г. Поиск закономерностей («если X, то Y»).
4. Ассоциативные правила	Г. Поиск закономерностей («если X, то Y»).

Запишите выбранные цифры под соответствующими буквами:

А	Б	В	Г

3. Прочитайте текст и установите последовательность действий (ПК-1.1)

Этапы обработки данных:

- А. Очистка от шума.
- Б. Предварительный просмотр.
- В. Сбор данных.
- Г. Построение модели.

Запишите соответствующую последовательность букв слева направо

Ответ:

4. Прочитайте текст и запишите развёрнутый обоснованный ответ (ПК-1.1)

Какие показатели используются для оценки качества классификационной модели?

Ответ:

5. Прочитайте текст, выберите правильные ответы и запишите аргументы, обосновывающие выбор ответов (ПК-1.1)

Какой тест используется для сравнения средних двух вариантов?

- А. t-критерий Стьюдента
- Б. Критерий хи-квадрат
- В. Тест Манна-Уитни
- Г. Z-тест

Ответ:

Обоснование:

6. Прочитайте текст и установите соответствие (ПК-1.2)

Соотнести термин с определениями:

Фактор ранжирования	Описание
1. Импутация	А. Заполнение пропущенных записей.
2. Стандартизация	Б. Приведение данных к диапазону $[0,1]$.
3. Нормализация	В. Фиксация изменений данных.
4. Ведение лога	Г. Приведение данных к среднему=0, std=1.

Запишите выбранные цифры под соответствующими буквами:

А	Б	В	Г

7. Прочитайте текст и установите последовательность действий (ПК-1.2)

Этапы работы алгоритма k-means:

- А. Выбор числа кластеров.
- Б. Расчет центроидов.
- В. Назначение объектов ближайшим кластерам.
- Г. Повторные шаги Б-В до сохранения.

Запишите соответствующую последовательность букв слева направо

Ответ:

8. Прочитайте текст и запишите развёрнутый обоснованный ответ (ПК-1.2)

Что такое переобучение (переобучение)? Как его избежать?

Ответ:

9. Прочитайте текст и установите соответствие (ПК-1.2)

Соотнести инструменты с задачами:

Фактор ранжирования	Описание
1. Matplotlib	А. Машинное обучение.
2. Scikit-learn	Б. Предварительный просмотр данных.
3. SQL	В. Работа с нейронными сетями.
4. TensorFlow	Г. Запросы к базам данных.

Запишите выбранные цифры под соответствующими буквами:

А	Б	В	Г

10. Прочитайте текст и установите последовательность действий (ПК-1.2)

Этапы ETL:

- А. Извлечение (извлечение данных).
- Б. Загрузка (загрузка в хранилище).
- В. Трансформация (трансформация).

Запишите соответствующую последовательность букв слева направо

Ответ:

11. Прочитайте текст и запишите развёрнутый обоснованный ответ (ПК-1.3)

Как выбрать расчетное число кластеров в алгоритме k-means?

Ответ:

12. Прочитайте текст, выберите правильные ответы и запишите аргументы, обосновывающие выбор ответов (ПК-1.3)

Какая библиотека Python используется для работы с DataFrame?

- А. NumPy
- Б. Панды
- В. SciPy
- Г. Матplotlib

Ответ:

Обоснование:

13. Прочитайте текст и установите соответствие (ПК-1.3)

Соотнести алгоритмы с задачами:

Фактор ранжирования	Описание
1. Линейная регрессия	А. Прогнозирование непрерывных измерений.
2. К-ближайших соседей (КНН)	Б. Классификация на основе расстояний.
3. Дерево решений	В. Поиск ассоциативных правил.
4. Априори	Г. Классификация с использованием иерархических условий.

Запишите выбранные цифры под соответствующими буквами:

А	Б	В	Г

14. Прочитайте текст и установите последовательность действий (ПК-1.3)

Этапы анализа данных в CRISP-DM:

- А. Понимание бизнес-задач.
- Б. Подготовка данных.
- В. Построение модели.
- Г. Интерпретация результатов.

Запишите соответствующую последовательность букв слева направо

Ответ:

15. Прочитайте текст, выберите правильные ответы и запишите аргументы, обосновывающие выбор ответов (ПК-1.3)

Какой метод визуализации подходит для категориальных данных?

- А. Гистограмма
- Б. Круговая диаграмма
- В. Линейный график
- Г. Диаграмма размышления

Ответ:

Обоснование:

3.2. Ключи к контрольным заданиям

№ задания	Верный ответ
1.	Б AutoCAD — программа для 3D-моделирования, а остальные инструменты используются для анализа данных.
2.	A2B1B3Г4
3.	ВАГБ
4.	Точность. Точность. Recall (полнота). F1-мера (гармоническое среднее). ROC-кривая.
5.	А, Г t-тест — для выбора маленьких, z-тест — для больших.
6.	A1B3B4Г2
7.	АБВГ
8.	Переобучение — когда модель запоминает шумы, а не правила. Способности избежать: использование кросс-валидации; упрощение модели; регуляризация.
9.	A2B1B4Г3
10.	АВБ
11.	Метод локтя (минимизация количества квадратов расстояний). Индекс силуэта. Кросс-валидация.
12.	Б Pandas — Основная библиотека для табличных данных.
13.	A1B2B4Г3
14.	АБВГ
15.	Б Круговая диаграмма показывает долю категорий.

Шкала и критерии оценки текущего тестирования

Число правильных ответов	Оценка
90-100% правильных ответов	Оценка «отлично»
70-89% правильных ответов	Оценка «хорошо»
50-69% правильных ответов	Оценка «удовлетворительно»
Менее 50% правильных ответов	Оценка «неудовлетворительно»

3.3. Перечень тем для проверки образовательных результатов на знания (вопросы к экзамену)

1. Типы данных: количественные, качественные, категориальные. Примеры классификации.
2. Этапы анализа данных: от сбора до значительных результатов.
3. Описательная статистика: основные показатели (среднее, медиана, мода, дисперсия).
4. Корреляционный анализ: коэффициенты Пирсона и Спирмена, интерпретация результатов.
5. Линейная регрессия: построение модели, оценка коэффициентов инновации.
6. Множественная регрессия: мультиколлинеарность и методы борьбы с ней.
7. Проверка статистических гипотез: t-тест, критерии χ^2 , ANOVA.
8. Интеллектуальный анализ данных (Data Mining): задачи и методы.
9. Кластеризация методом k-средних: алгоритм и оптимизация числа кластеров.
10. Иерархическая кластеризация: дендрограммы и их применение.
11. Анализ временных рядов: тренды, сезонность, прогнозирование.
12. Работа с пропущенными данными: методы заполнения и удаления.
13. Нормализация и стандартизация данных: цели и алгоритмы.
14. Снижение размерности данных: PCA (метод основного компонента).
15. Инструменты анализа: Excel, Python (Pandas, NumPy), R.
16. Предварительный просмотр данных: построение графиков в Matplotlib и Seaborn.
17. Создание интерактивных дашбордов в Power BI/Tableau.
18. SQL для анализа данных: агрегация, JOIN-запросы, оконные функции.
19. Большие данные: Hadoop, Spark, независимость с постоянным анализом.
20. Алгоритмы классификации: деревья решений, случайный лес, SVM.
21. Метрики оценки качества моделей: точность, полнота, F1-мера, ROC-кривая.
22. Метод воздействия соседей (k-NN): принцип работы и настройка параметров.
23. Анализ текстовых данных: НЛП, TF-IDF, векторизация.
24. Ассоциативные правила (Априори): поиск паттернов в транзакционных данных.
25. Хранилище данных: архитектура, ETL-процессы, OLAP vs OLTP.
26. Этика анализа данных: конфиденциальность, смещения в алгоритмах.
27. A/B-тестирование: дизайн экспериментов и интерпретация результатов.
28. Рекомендательные системы: совместная и контентная фильтрация.
29. Анализ социальных сетей: графовые модели и кластеризация сообществ.
30. Практические кейсы анализа данных: прогнозирование погоды, оценка рисков, маркетинговая аналитика.